Meaning making in the design and interpretation of data visualisations

Margaret Jill Simpson

PhD

The University of York

Sociology

September 2020

Abstract

This thesis examines how meaning is made in the design and interpretation of data visualisations. As data visualisations proliferate, they have become the main way that non-experts access data (Kennedy, Hill, Allen and Kirk, 2016). Therefore, analysing *how* meaning is made in their production and consumption contributes to understandings of how people experience data through visualisation. Using *encoding/decoding* (Hall, 1980) as a theoretical framing, this thesis focuses on the meaning making practices that occur before and after the final image of a graph or chart is determined. This includes the data and design practices of visualisation designers and the interpretive strategies of ordinary users when they engage with data visualisations. Therefore, research was conducted in two stages; a *short-term ethnography* (Pink and Morgan, 2013) at a data visualisation design agency, followed by interviews with 'ordinary users' (Kennedy, 2018). Analysis of the research data identified three ways in which meaning is made in the design and interpretation of data visualisations; *encoding meaning*, *imaginary figures*, and *contextualising and re-contextualising data*. This builds upon and extends understandings of how encoding and decoding works, particularly in relation to how users are *imagined* by designers and the various strategies designers and users employ to make meanings from data through and from its visualisation. These findings lead the thesis to argue that the meanings of data visualisations are determined by contextual factors that exist beyond the dataset being visualised. These include what the data and its visualisation will be used for, how the designers envision the user, and the discourses designers and users draw on to situate data within broader contextual frameworks. This thesis contributes knowledge to both the academic field of data studies and to data visualisation design.

Table of Contents

List of tables and figures

List of appendices

Acknowledgements

I would like to extend my deep gratitude to all those who have contributed to the completion of this thesis. A special thanks to all my participants who so generously gave their time and to Evoke for inviting me into their organisation and for making me feel so welcome.

I would like to thank my wonderful PhD supervisors, Professor David Beer and Dr Rosemary Hill. Dave, your unwavering support has been a lifeline to this thesis and my academic career. Without your guidance, encouragement, and your belief in me I would not have made it this far. Thank you for keeping my feet on the ground in moments of anxiety and self-doubt and for making me see the funny side of my worries. I will be forever grateful to you. Rosey, thank you for your encouragement, for sharing your early career experiences with me, and for challenging me on my research. I will miss your sense of humour and seeing your cat and impressive guitar collection during our Skype supervisions.

I would also like to thank Professor Helen Kennedy, without whom the 'relating to data through visualisation' research network would not exist. This interdisciplinary network provided me with the opportunity to undertake this thesis and thank you to the ESRC, White Rose DTC and University of York who funded my studentship. Thank you also to the academics and commercial partners who formed the network and who supported me and challenged me on my research. Finally, thank you to my fellow network PhD students, Lulu and Arran. It has been a pleasure to go through this journey with you both.

To my amazing Sociology PhD friends, you are such a talented, funny and generous group of people and I have missed you enormously during lockdown. A special thanks to my office mate, Dan, for understanding when I have felt overwhelmed and for always believing in me.

To my brilliant family and friends, I know you are bemused by my career choice and baffled by what I have been doing for the last four years! However, when the thesis has felt all encompassing, your support and good-natured teasing has helped me to keep my feet on the ground. A special thanks to Ruth and Ben who have provided me with a second home in York, always full of fun and laughter.

Finally, to my husband Josh. Thank you. I could never have completed this thesis without your support. We have sacrificed a lot for me to do this PhD. I hope I have done us both proud and I am so excited for the next chapter of our lives together.

Author’s declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1: Introduction

1.1 Research introduction

Data is permeating "civic, social and cultural life", requiring people who may not be experts in data to engage with it in order to be included in these arenas (Kennedy and Hill, 2018, p. 831). Gray uses the term "public data cultures" to describe the increasing ways in which people are persuaded to engage with data and make sense of it (2020, p. 315). Most people who are not experts in data encounter and experience it through its visualisation in graphs and charts (Kennedy, Hill, Allen and Kirk, 2016b; Kennedy and Hill, 2018; Kennedy, 2018). This is because access to datasets and the skills to analyse and interpret them are distributed unequally in society (boyd and Crawford, 2012). However, visualisations are regarded as a means of making data accessible, comprehendible and transparent to non-experts (Kennedy and Hill, 2018). Indeed, Kennedy and Hill (2018) describe how, for most people, data is not perceptible outside of its visual representation. This makes the production and consumption of data visualisations an important area of research in relation to how people experience data in their everyday lives. This thesis contributes to debates around how people *experience* data, by explaining how meaning is made in the design and interpretation of data visualisations. To do so, it draws on observation and interview data from a short-term ethnography (Pink and Morgan, 2013) conducted in a data driven design agency as well as data from interviews with 'ordinary users.' The latter incorporated examples of data visualisations which I asked participants to interpret.

The prevalence of data visualisations in people's everyday lives can be linked to the enjoyment people find in engaging with them (Beer, 2013). Indeed, one way in which data is circulated and becomes part of the everyday is through its incorporation into popular culture (Beer, 2013). Beer argues that data visualisations are fun. They are shared, tweeted, tagged and re-tweeted using social media in a form of *data play* (Beer, 2013, see p.110). The proliferation of data visualisations within popular culture, particularly within journalism, has led Beer to suggest there is a "culture of visualisation" (2013, p.113). Not only do people wish to engage in play with data, they also want to reveal what the data has to say (Beer, 2013) and visualisation is one way of achieving these aims. This enthusiasm for exploring the insights 'hidden' within data has resulted in a wealth of digital platforms, which enable both expert and non-expert users to create and share data visualisations (Beer, 2013). Despite this, there has been little research into how non-expert users engage with data in their everyday lives (Masson and van Es, 2020; Kennedy,

2018). Being attentive to how users make meanings from data through visualisation provides a means through which to understand how ordinary people engage with data and information.

Kennedy draws on cultural studies when using the term " "ordinary people" " who, she argues, are missing from research on non-experts' experiences of datafication (2018, p. 20). By using this term, Kennedy (2018) directs research attention away from elites who have the power, access and expertise to utilise data, towards how ordinary people experience data as part of their everyday lives. Drawing on Kennedy's use of this term, this thesis refers to *ordinary users* to represent people who may encounter data visualisations as part of their everyday life, for example at work, in news media or through self-tracking apps. Furthermore, as data and its visualisation permeate more aspects of daily life (Kennedy and Hill, 2018), the term 'ordinary user' reflects the varied experience and skills people possess in engaging with graphs and charts and the diverse settings in which they may encounter them. However, unlike the term 'non-expert users,' it makes no judgement about the skills and experience different users bring to their interpretation of graphs and charts.

As well as being an important way through which publics encounter and experience data, graphs and charts make it both perceivable and actionable in the social world (Gitelman and Jackson, 2013; Espeland and Stevens, 2008). In doing so, they tend to reinforce existing power relations by reproducing very particular ways of seeing the world (Kennedy and Hill, 2017). In this sense, Kennedy and Hill argue that data visualisations do "ideological work" (2017, p.773). Despite their rhetorical power, Drucker (2014) explains how visual representations of data are particularly good at hiding the way in which they function as argument. This "aura of objectivity" makes the claims made through data visualisations so convincing (Kennedy, Hill, Aiello, and Allen, 2016a, p. 723). Therefore, those who possess access to data and the skill set to tell stories with it occupy a position of power (Beer, 2019). This is important because of the ways in which data is implicated in social ordering and governance while appearing to produce neutral forms of knowledge (Beer, 2016a; 2016b). Indeed, Drucker poetically describes data visualisations as "acts of interpretation masquerading as presentation" (2014, n.p). While this thesis does not focus on the politics of data visualisations, it is important to understand how data visualisations are produced and consumed in this critical context.

Although attentive to how ordinary users experience data through its visual representation, this thesis is weighted towards how designers make meanings from data through the design and production of data visualisations. It is important to analyse the production of data visualisations for three reasons. Firstly, their production frames how ordinary users interpret them. Secondly, the claims made through data visualisations are convincing (Kennedy et al., 2016a) therefore, it

is important to understand how these rhetorical devices are produced. Finally, directing attention towards the production of data visualisation may reveal more about how the politics of data play out through its visual representation. Therefore, researching the production of data visualisations will contribute to ongoing debates within data studies, particularly those pertaining to the subjectivities of data, as outlined in Chapter two.

This chapter will introduce the aims of the project and the research questions before explaining how the thesis will address these. It will go on to discuss various ways in which meaning making has been conceptualised in relation to images. Kennedy at al., argue that "media audience research" is useful to draw on when exploring how social and cultural factors play out in user engagement with data visualisations (2016b, n.p.). The thesis draws upon and develops Stuart Hall's (1973; 1980) *encoding/decoding* model as a theoretical framing from which to approach meaning making, as will be discussed in more depth in section 1.3.2. Finally, the chapter will provide an overview of the thesis and its findings.

1.2 Research aims and questions

This project forms part of an interdisciplinary research network called '*Relating to data through visualisation,'* led by Professor Helen Kennedy. The network comprises three PhD students (each working on a separate project about data visualisation), their supervisors, and two commercial partners who are both considered experts in the field of data visualisation design. The contributions from the commercial partners helped to ensure the three PhD projects not only contributed to academic debate but also produced knowledge that has 'real-world' impact among design practitioners. The following section of this introduction outlines the aims and research questions of this thesis, which explores meaning making in the design and interpretation of data visualisations.

A review of the literature, as outlined in Chapter two, revealed how important moments in the meaning making process are neglected in existing research into the design of data visualisations. A tendency to focus analysis on the image of graphs and charts to explain how they communicate meaning neglects the meaning making practices that happen before and after the image is finalised. For example, the data and design practices of the makers of data visualisations and the interpretive practices of those who engage with them. Furthermore, existing research has tended to focus on the visual elements of data visualisations while ignoring the role of data altogether. In order to understand how the data itself shapes the meaning of data visualisations, this thesis aims to analyse the meaning making practices that take place during the design process to assess how designers encode visualisations with

meaning. Furthermore, by drawing on Kitchin's (2014) concept of data assemblages, it aims to understand how the socio-technical contexts in which data are embedded frame how meaning is made from data and encoded into data visualisations.

Drawing on a cultural perspective, this thesis considers the users of data visualisations to be active participants in the meaning making process (see Hall, 1997 and du Gay, Hall, Janes, Macay, Negus, 1997). Much of the existing user research is from a Human Computer Interaction studies perspective and often focuses on issues around the effectiveness of different data visualisations to communicate information. While there is some limited user research from a sociological perspective (see for example Kennedy et al., 2016b), this has focused on the issue of engagement with data visualisations. This has resulted in a gap in social and cultural understandings of how users interpret data visualisations and the meaning making strategies they employ to make sense of them. While weighted towards research conducted into the design of data visualisations, this thesis aims to address this gap in existing research by analysing how ordinary users make meanings through their interpretation of graphs and charts. Indeed, this project is unusual in that it aims to track meaning making in *both* the design and interpretation of data visualisations whereas existing research has predominantly approached these as separate areas of research.

1.2.1 Research Questions:

The following subsection will introduce three research questions, explaining how they address the aims of this thesis.

Q 1. *How do designers encode meaning into data visualisations?*

This question responds to the thesis' aim to understand how designers encode data visualisations with meaning. Rather than directing analytic attention to the image of graphs and charts, it focuses on the practices of the designers in the design and production of data visualisations. In answering this question, the thesis will be attentive to the role data plays in the meaning making process. It will do so by focusing on the discourse, design and production (Kress and van Leeuwen, 2011) practices that take place in a data driven design studio.

Q 2. *How do ordinary users do meaning making when engaging with data visualisations?*

This question addresses the thesis' aim to understand how ordinary users make meanings from data visualisation through their interpretation of them. Furthermore, asking this question allows

the thesis to track meaning making in both the design and interpretation of data visualisations, something the existing literature does not tend to do. While the thesis is sensitive to how socio-cultural categories such as age, gender, ethnicity and level of education may be influential in determining how people make meanings from data visualisations, analysis will not be limited to these factors. Instead, in asking this research question, the thesis will identify the diverse meaning making strategies different users employ in order to make meanings from graphs and charts.

Q 3. *What is the relationship between meaning making and data assemblages?*

This question responds to the thesis' aim to understand how the socio-technical contexts in which data are embedded frame the knowledge that is produced from them through visualisation. In order to direct attention to the role data plays in how meaning is made in the design of data visualisations, the thesis will employ Kitchin's (2014) concept of data assemblages to the domain of data visualisation. In order to answer this question the research will analyse how meaning is made through the data and design practices of designers working in a data driven design studio.

This section of the chapter has introduced the aims of the thesis and unpacked the research questions it sets out to answer. In order to address these aims and questions it is important to examine how meaning making has been conceptualised and introduce the theoretical framing that will form the foundation of this thesis' approach to researching meaning making in relation to the design and interpretation of data visualisations.

1.3 Conceptualising meaning making

This section of the introduction provides a brief overview of different approaches to theorising meaning making in relation to images that this thesis will build upon. This field of literature is broad therefore this discussion will draw on several perspectives including art history, STS, literary theory and audience studies in order to consider a range of theoretical approaches. It is organised into two main sections: *Reading images* and *Encoding/decoding*, the latter of which will form the theoretical framing for this thesis. Firstly, however, I will outline my broader theoretical understanding of meaning making as being culturally and discursively constructed.

I consider meaning to be culturally constructed, that is, meanings are not "'found' within things", they are formed through the "cultural practices," within which we are immersed (du Gay et al., 1997, p.14). Therefore, a graph or chart only becomes meaningful when those

designing or interpreting it apply broadly shared interpretive frameworks in order to make sense of it. As Nafus (2014) explains in her article on sensor data, data often does not resemble the phenomenon it represents. Therefore, in order to make meanings from data those interpreting it must draw on "language and concepts" which are socially and culturally dependent (du Gay et al., 1997, p.14). Nafus (2014) describes the labour required to make data meaningful, acknowledging the cultural practices that allow people to make sense of things. Much of this meaning making *work* is done through language and discourse, through which it is possible to gauge the relation between one thing and another thing (du Gay et al., 1997). Nafus uses the example of weather predictions communicated as percentages, for example, "a 20% chance of rain," to explain how on its own the number makes little sense (2014, p. 210). Yet such percentages become meaningful through cultural practices (Nafus, 2014), that is, through the shared "interpretive framework" (du Gay, et al., 1997, p.18) and discourse of weather forecasting. However, this is not to say that meanings are fixed, they shift and change over time and they are contested (du Gay et al, 1997). This broad understanding frames the literature I have drawn on in the next sections, which consider different theoretical approaches to meaning making in relation to the aims of this thesis.

1.3.1 Reading images

The following discussion explores approaches to meaning making from a reader centred perspective, that is, how interpreters make meanings through the design and interpretation of images. Although recognising the importance of the image in meaning making, Berger (1972) draws attention to the role of the reader in making images meaningful. Berger argues, "The relation between what we see and what we know is never settled" (1972, front cover). This simple statement sums up beautifully the subtle distinction between seeing an image and making sense of it. While Berger does not set out a theoretical framework from which to understand meaning making from images, his book *Ways of Seeing* (1972) provides provocations from which to think about how ordinary users may interpret data visualisations. For example, Berger (1972) suggests that how an individual makes sense of what they see is related to that person's worldview, in terms of their knowledge and beliefs. This supports Kennedy et al's (2016b) assertion that research into user engagement with data visualisations ought to be attentive to socio-cultural factors, which may shape how different people experience them. Placing the viewer of the image at the centre of the meaning making process, Berger states, "we are always looking at the relation between things and ourselves" (1972, p. 1). This further suggests that an individual's socio-cultural position frames how they make sense of images such as data visualisations. Similarly, Haraway (1988) recognises how an individual's social and cultural context shapes their view of the world.

In her critique of scientific understandings of objectivity, Haraway describes visualisation as "the god trick of seeing everything from nowhere" (1988, p. 581). While technologies of visualisation may seem to present a disembodied view, Haraway (1988) argues that all knowledge is situated. That is, both the production of images and individuals' interpretations of them are always positioned within social and cultural contexts (Haraway, 1988). Therefore, Haraway's concept of 'situated knowledge' explains how vision is both partial and embodied (1988, see pp.581-583). This view challenges the "aura of objectivity" that surrounds data visualisations as legitimate forms of knowledge (Kennedy et al., 2016a, p. 723) and points towards the significant power held by those who have the access and skills to interpret and visualise data (Beer, 2019; also see boyd and Crawford, 2012). Furthermore, understanding knowledge as embodied opens up discussions around how emotions and feelings may play a role in how ordinary users experience and make meanings from data visualisations (see Kennedy and Hill, 2018 and D'Ignazio and Klein, 2016). Berger (1972) and Haraway's (1988) work suggests that meanings are framed by the social and cultural contexts of both those who interpret visual representations of data and of those who produce them.

Visual images, including data visualisations, are often made up of more than one semiotic resource (Aiello, 2020). These resources might include a combination of sound, text, animation, and images which work together to frame the meaning of the image (Rose, 2016). Barthes (1977) draws on a semiotic framework of codes and signs to explain how meaning is communicated through messages, using the example of a printed advertisement for tinned tomatoes. Firstly, Barthes (1977) describes the linguistic message, which is expressed in the caption and the label on the tin. The linguistic message anchors the meaning of the image in order to focus and limit the viewers' identification (the denotive message) and interpretation (the symbolic or connotative message) of the advertisement. Secondly, Barthes' describes the coded iconic message; this is a *connoted* meaning and a symbolic message, which requires cultural knowledge to decode. In his example, the reader must draw on tourist stereotypes of Italy and Italian food as understood by a French audience to decode the advert (Barthes, 1977). Thirdly, Barthes describes the non-coded iconic message. This he refers to as the *denoted* meaning, a literal message made up of the elements or objects within the image, in this example the fresh vegetables, shopping bag, and tin (Barthes, 1977). This layered messaging is received by the reader at the same time and Barthes argues that there is no literal message without a symbolic message. That is, the denoted image and connoted image are always working together to construct meaning (Barthes, 1977). This suggests that both the literal (denotive) meaning of its visual elements and the social and cultural context of the reader determine the meaning of an image.

Although Barthes' semiotic framework provides a theoretical tool to unpick how meaning is communicated through the image, it pays little attention to the producers of that image. Nor does it allow for subtle differences between readers in their interpretation of it, other than in broad cultural terms. The image itself is the focus of the analysis, which does not take into account the meaning making practices that occur before and after the image is complete, through processes of design, production and interpretation. Barthes also neglects to explore how the context in which the image was produced shapes its meaning. In contrast, a social semiotic approach not only investigates the semiotic resources used to produce images and communicate meaning but also aims to understand how these resources are embedded in broader histories and conventions of data visualisation design (Aiello, 2020). Indeed, Aiello (2006) suggests that social semiotics requires a critical focus on who in society governs systems of semiotic resources. Therefore, this approach not only requires the systematic inventorying of the semiotic resources used in the broad field of data visualisation design, but also situates these "in their social and cultural contexts" (Aiello, 2020, p. 56). Aiello (2020) advocates empirical research that does not stop at collecting data in the form of visual representations of data visualisations, but also involves engaging with designers and users to understand how semiotic resources are used and understood by different groups. This, she argues, requires questions about what semiotic resources *mean* to those individuals (Aiello, 2020).

Social semiotics can reveal the politics and power relations (Aiello, 2020) inherent in data visualisations by identifying the semiotic resources that make up their design conventions and interrogating their origins as a system of signs (see for example, Kennedy et al., 2016a). However, despite Aiello's assertion that a social semiotic approach might involve talking to users of data visualisations, such work tends to focus on the "meaning potential" of semiotic resources and the social context in which they are deployed (Aiello, 2020, p. 53). This does not account for how social and cultural differences between users may frame how individuals make meanings from data visualisations and presupposes that meaning can be reduced to the arrangement of semiotic modes. Furthermore, focusing on semiotic resources neglects to consider how meaning is made from data prior to it being encoded into visualisations, for example through data and design practices. In this sense, the semiotic approaches discussed above have focused on how meaning is communicated through visual sign systems, rather than how it is *made* through the social and cultural practices involved in the design and interpretation of images.

Kress and van Leeuwen (2001) write about multimodal forms of communication, that is, semiotic artefacts that are made up of more than one mode. Multimodal data visualisations, for example, often incorporate images, text and numbers (Weber, Engebretsen, and Kennedy, 2018)

alongside their *graphic modes* (Engebretsen and Weber, 2017). In theorising multimodal communication, Kress and van Leeuwen (2001) focus on the practices and resources through which meaning is made. In doing so, they map how meaning is produced through four domains, "discourse, design, production and distribution" (2001, p. 4). The first concerns the socially constructed *discourses* that represent a particular thing or phenomenon. *Design* refers to the different ways semiotic resources are brought together to communicate particular discourses. The third concerns the *production* of the "semiotic artefact," which the authors note may require particular skill sets in relation to the media of its production (Kress and van Leeuwen, 2001, p. 6). Finally, the fourth refers to the ways in which "semiotic products" are re-produced and *distributed* (Kress and van Leeuwen, 2001, p. 21). While appearing to foreground producers, Kress and van Leeuwen (2001) argue that these domains are implicated in meaning making in both production and interpretation. They explain how those interpreting artefacts need to recognise the discourse being used and apply their "semiotic knowledge" in relation to principles of design, production and distribution to make meanings from it (Kress and van Leeuwen, 2001, p. 8).

By distinguishing these layers in meaning making, Kress and van Leeuwen (2001) draw attention to the ways in which meaning is made both before and after the semiotic artefact has come into being. Furthermore, by drawing a distinction between discourse, design, and production, the authors imply the various meaning making practices at work in the design of an image. These indicate important moments of meaning making in the design and interpretation of data visualisations, which are underexplored in existing literature (as outlined in Chapter two). Hall's (1980; 1973) concept of *encoding/decoding* provides a useful theoretical framing from which to move beyond analysis of the image of graphs and charts to the meaning making practices at work in their design and interpretation.

1.3.2 Encoding/decoding

This section will introduce Hall's (1980) model of *encoding/decoding*, a particularly influential theory within the broader fields of audience studies. As this section will explain, this particular theoretical approach to meaning making offers a framing from which to research how meaning is made through both the production and consumption of data visualisations. It allows analytical attention to move beyond the image of the graph and chart to the moments in which meaning is made during their design and interpretation. The following section will critically discuss the different elements of *encoding/decoding* in relation to the aims of this thesis.

Using a semiotic framework, Hall (1980) introduced *encoding/decoding* to explain how meanings are *encoded* into television news programmes through processes of production, before being *decoded* by their audiences. In order for a message to be effective, Hall argues that the audience must first decode it as being meaningful and this relies on the audience recognising it as a "meaningful discourse" (1980, p. 130). Describing how television news represents events, Hall (1980) explains that they must be turned into a story before they can be communicated to audiences. Similarly, in defining *discourse* in the context of multimodal communication, Kress and van Leeuwen (2001) describe how a journalist might draw on a discourse of war to explain conflict events in different countries. Stressing the importance of language in mediating reality, Hall explains that what it is possible to "know" or to "say" is a product of discourse (1980, p. 131). Therefore, transforming events into meaningful discourses underpins both the encoding and decoding processes. When this is considered in relation to the design and interpretation of data visualisations, it reinforces the need to direct research towards the moments before and after the image of the chart comes into being to understand how meaning is made from data through and from its visualisation.

While Hall (1980) recognises that audiences may not decode messages in the way encoders intended, he is cautious to state that this does not mean encoded messages have a plurality of meanings. While it is not possible for the meaning of encoded messages to be entirely fixed (Morley, 1980), Hall (1980) explains that encoders construct the limits within which an audience can decode them. They do this, in part, through the "discursive domains" that enable viewers to make sense of events in the social world (Hall, 1980, p 134). Hall proposes four codes to explain how audiences decode meanings from television news (Steiner, 2018), which I will briefly describe here. The first is the dominant code, or what Hall (1980) also refers to as the *preferred reading*. This describes when audiences decode the meaning of the news content as the encoders intended it to be read. The second is the *oppositional* code which describes audiences who recognise the dominant code yet reject the intended meaning and instead come up with a different meaning based on an "alternative frame of reference" (Hall, 1980, p. 138). The third is the *negotiated code*, this describes how an audience may agree with the dominant code but employ a "mixture of adaptive and oppositional elements" to determine its meaning in relation to their particular context or situation (Hall, 1980, p, 137). The latter two codes accept the possibility of disjuncture between the intended messages producers encode into media and how members of the audience decode them.

Hall describes a fourth code, the *professional* code, which works to support the dominant code through "apparently neutral techniques" such as the style and format of news programmes and the stories chosen (1980, p. 136). These, he argues, form the "presentational values" of

professional news production (Hall, 1980, p. 136). Steiner (2016) argues that the professional code tends to be neglected in research and suggests that this may be because Hall himself did not focus his attention towards encoding. However, it is possible to see parallels between Hall's professional code and the conventions of data visualisation design, which frame how meaning is made through graphs and charts (see Kennedy et al., 2016a). Indeed, Hall recognises how the production of television media "is framed throughout by meanings and ideas," which originate from within the industry, including industry-accepted knowledge, technical skills and professional ideologies (1980, p. 129). In the context of this thesis, it is possible to relate these industry specific systems to the data assemblages (Kitchin, 2014) within which data, and thus its visualisation, is embedded. Therefore, Hall's work highlights the importance of paying attention to the practices that take place during the production of data visualisations and how they frame how meaning is made (see Chapter four). This not only includes data and design practices, but also how the designers envision the users of the data visualisation they are producing (see Chapter five). As Hall (1980) suggests, encoding requires the encoders to make assumptions about the audience and these assumptions shape production.

Encoding/decoding (Hall, 1980) provides a useful theoretical framing from which to explore how meaning is made in both the design and interpretation of data visualisations. Indeed, Aiello suggests that a cultural studies approach is "decentered from the text," thus allowing the analytical focus to shift towards the practices involved in the text's production and consumption (2006, p. 97). However, it is important to reflect on the cultural context in which Hall was writing and how it has changed in a digital era. For example, how people watch television and engage with news content has altered dramatically since Hall wrote about the *encoding/decoding* model. During the nineteen seventies and nineteen eighties there were significantly fewer television channels and what viewers could watch at any given time was determined by television scheduling. Today, streaming services such as Amazon Prime and Netflix mean that individuals have access to a huge amount of television programmes on their computer, tablet or mobile phone, which they can watch whenever and wherever they like. This substantial shift in the culture of television viewing is reflected in the following quote from Dahlgren's 1988 research on how people make sense of news on television:

"*the daily recurrence and readily recognisable features of the programmes serve to link the viewer and his/her everyday life to the larger world in a manner which is ritualistic, symbolic and ultimately mythic [..]*" (Dahlgren, 1988, p. 289).

At the time Dahlgren was writing, people would have accessed the majority of their news through newspapers and television programmes. The limited number of organisations producing

news content would be known and familiar to audiences and acquiring a newspaper or watching a news programme would be embedded in many of the audience's routines. In this context, Dahlgren's (1988) reference to the way in which news programs connected their audiences to the world in ritualistic ways makes sense. Today, however, news content is produced and distributed through new forms of media, including apps, websites and social media platforms. Furthermore, Bødker describes how journalistic content is being circulated through these new media in ways that detach it from "their context of production" (2016, p.414). This removes important cultural information that may have previously shaped how audiences made meanings from the content, such as who produced it and why. This form of de-contextualised news content is similar to the ways in which data visualisations circulate online (Beer, 2013) and raises questions around how audiences decode them when they are detached from the context within which they were produced.

In contemporary Western societies, media content is embedded in peoples' daily routines in ever more individualised ways, which would have been unfamiliar to Hall when he developed *encoding/decoding* (Bødker, 2016). Furthermore, social media is increasingly where news "circulates and attains meaning" as more people access journalistic content through these platforms (Bødker, 2016, p.410). The functionality built into social media allows news content to be shared, 'liked', and commented on by users, which in turn produces new layers of discursive content to be decoded by those interpreting them (Bødker, 2016). This is in contrast to the context of mass media televised news Hall wrote about, in which he considered news audiences as "relatively isolated individuals" (Bødker, 2016 p. 415). This thesis does not set out to explore how the circulation of data visualisations on social media platforms affects how people consume and make meanings from them. Nevertheless, this is representative of the broader shift towards a "culture of visualization," in which data visualisations are produced and circulated online (Beer and Burrows, 2013, p. 62, also see Beer, 2013). Furthermore, as data visualisation is increasingly incorporated into journalistic content, the points Bødker (2016) raises are indicative of the broader cultural context within which ordinary users encounter graphs and charts.

In applying Hall's theoretical model to a contemporary context, Bødker argues that *encoding/decoding* is still useful when researching the circulation of news media online. Moreover, he argues that Hall intended for the model to be developed and highlights the importance Hall placed on developing theory in light of contemporary contexts (Bødker, 2016). This, Bødker (2016) argues, justifies researchers developing the *encoding/decoding* model in relation to contemporary forms of news production and circulation. The invitation to develop the theory of encoding and decoding in relation to new contexts and in this case, a new domain,

makes it a particularly attractive framing from which to research meaning making in the context of data visualisations. Therefore, rather than simply applying the *encoding/decoding* model to the domain of data visualisation, I intend to use it as a provocation from which to research and analyse meaning making. Drawing on Kress and van Leeuwen (2001), this framing will allow me to focus on the moments and practices of meaning making that take place before and after the image of the visualisation comes into being. This includes how meaning is made from data before it is encoded into graphs and charts and the discursive resources designers and ordinary users draw on in making meanings from data through visualisation. It is these moments of meaning making that are overlooked in existing data visualisation research, as outlined in Chapter 2.

1.4 Thesis Outline: Answering the research questions

In conclusion, this thesis analyses how meaning is made in both the design and interpretation of data visualisations. Using *encoding/decoding* (Hall, 1980) as a framing for researching meaning making in the context of data visualisations, it directs attention to the meaning making practices that occur before and after the image of the graph or chart comes into being. This includes the data and design practices of designers in the studio and the interpretive strategies of ordinary users when engaging with data visualisations. In doing so, it identifies three ways in which meaning is made in the design and interpretation of data visualisations; *encoding meaning*, *imaginary figures*, and *contextualising and re-contextualising data*. This builds upon and extends understandings of how encoding and decoding works, particularly in relation to how users are *imagined* and the various strategies through which designers and users make meanings through and from data visualisation. The findings lead the thesis to argue that the meanings of data visualisations are determined by contextual factors that exists beyond the dataset being visualised. These include the intended use of the data, the designers' perception of the end user, and the discourses designers and users draw on to situate the data within broader contextual frameworks. This reinforces Kennedy and Hill's argument that "data visualisations are not neutral windows onto data" (2017, p. 773). Instead, they are a product of the subjectivities, which frame them, and the socio-technical contexts in which they are embedded.

An outline is given of the subsequent chapters of the thesis, which also sets out how the analysis chapters answer the research questions posed above.

Chapter 2 *Literature Review*

This chapter sets out the critical context of the project, situating it within the interdisciplinary field of data studies, specifically debates around data subjectivities. The chapter reviews existing research into data visualisations in three sections, *data visualisation conventions; story telling in the design of data visualisations;* and *user engagement with data visualisations.* By identifying gaps in understanding, this chapter positions the thesis and its research questions among the existing literature.

Chapter 3 *Methodology*

This chapter sets out the research methodology, explaining how the research took place in two stages. Firstly, a short-term ethnography (Pink and Morgan, 2013) was undertaken at a data driven design agency, which involved observing and interviewing designers of data visualisations. Secondly, interviews were conducted with ordinary users, which incorporated visual resources in the form of examples of data visualisations. The chapter addresses the practicalities and ethics of conducting the research and critically reflects on the methods used and how these shaped the research findings.

Chapter 4 *Encoding meaning in the design of data visualisations*

This chapter draws on fieldwork data in the form of observations and interviews with designers at a data driven design agency. In doing so it answers research question one, how do designers encode meaning into data visualisations, and research question three, what is the relationship between meaning making and data assemblages? The chapter begins by explaining how the designers made meanings from data prior to the design of its visual representation. This involved decoding the data in order to transform it into a "meaningful discourse" (Hall, 1980, p. 130) that could be decoded by users. However, how the designers decoded the meaning of the data was dependent on contextual factors beyond the dataset being visualised. These included what the data meant to the client, the intended use of the visualisation and how the designers perceived the needs of the intended users. Once meaning had been made from data it was encoded into the design of digital data visualisations through a *back and forth* process. This required designers with different skill sets (visual design and front end development) to work together to produce the data visualisation within the confines of the project and the limitations of the software within which they were working. Therefore, the chapter argues that how meaning is encoded into the design of data visualisations is a product of this assemblage of "*actors and actants"* (Kitchin, Maalsen, and McArdle, 2016, p. 98), operating within the particular context and constraints of each project. It argues that data visualisations are not merely representations of data but of the decisions of multiple actors, which are framed by the

context of the project. Furthermore, apparatuses of the data assemblage shaped how meanings emerged from data and were encoded into the design of data visualisations. This formed part of the broader context within which data visualisations were being designed and produced and this was made visible through the *back and forth* process of design and production.

Chapter 5 *Designing visualisations for imaginary figures*

Drawing on interviews and observations with designers this chapter introduces the concept of *imaginary figures,* to explain how the designers envisioned the intended user shaped how they encoded meaning into the design of data visualisation. In doing so, it responds to research question one, how do designers encode meaning into data visualisations? While Hall (1980) recognised that encoding requires producers to make assumptions about audiences, he did not elaborate on what this involved. Chapter five addresses this by explaining how the shared assumptions the designers possessed about different types of user were personified in three imaginary figures; *the grandma, the public,* and *the professional.* The designers drew on these figures to explain and account for the design decisions they made, revealing how they encoded meaning differently depending on how they imagined the intended user. This might determine the type of data visualisation used, the level of interactivity it provided the user, and how the user was guided through the data. The chapter explains how each imaginary figure represented a homogenous group of users and there was very little sense of diversity among users. Indeed, the designers working understanding of diversity was in relation to different users' perceived level of ability to engage with or make sense of data visualisations. Therefore, rather than thinking about the designers' understanding of diversity in relation to socio-cultural categories, it suggests that it might be more productive to think about it in terms of the *limits of meaning making*.

Chapter 6 *Contextualising and re-contextualising data*

This chapter is in two parts. The first draws on fieldwork interviews with designers to explain how they contextualise data in order to make meanings from it prior to the design of its visual representation. This section of the chapter responds to research question one. The second part of the chapter draws on interviews conducted with ordinary user participants in which I showed them examples of data visualisations. This part of the chapter explains how ordinary users re-contextualised the data within the visualisation in order to make meanings from it, thus answering research question two, how do ordinary users do meaning making when engaging with data visualisations? The chapter reveals how designers and ordinary users drew on some of the same strategies to add context to the data, situating it in broader frames of reference. These

included identifying patterns in the data and drawing on their prior knowledge of what the data represented to account for what the data showed. While the designers had the opportunity to conduct research beyond the dataset to add context, during interviews ordinary users tended to draw on their *prior knowledge*, *personal experience*, and even *imagination* to situate the data in frames of reference that were familiar to them. In doing so, they constructed narratives of the data that went beyond what the graph or chart could show, thus extending the data's reach (Dourish and Gómez Cruz, 2018).

Chapter 7 *Conclusion*

This chapter concludes the thesis by outlining its key findings and argument before stating its contribution to academic debate and to the field of data visualisation design.

Chapter 2: Literature Review

2.1 Introduction

To set out the critical context of this thesis and to situate its questions among existing research into data visualisations, this chapter will analyse the literature in two parts. The first part briefly defines the concepts of big data and datafication as *socio-technical phenomena* (boyd and Crawford, 2012) and introduces the politics of data. It goes on to review existing research on making sense of data. This first section addresses the critical context of this thesis through the lens of data subjectivities and positions it within the interdisciplinary field of data studies, to which this thesis contributes. In doing so, it sets out how an understanding of how data visualisations are produced and consumed could further debates around data subjectivities. It also introduces Kitchin's (2014) notion of *data assemblages*, which concerns my third research question. This will be used as a conceptual tool to explore how the complex socio-technical structures that data are embedded in frame how knowledge is produced through visualisation. The second part of the chapter analyses the existing research on the design and interpretation of data visualisation in three distinct but interrelated themes: *Data visualisation conventions*; *Story telling in the design of data visualisations*; and *User engagement with data visualisation*. This second section of the chapter reveals gaps in the literature that this thesis will address through research Questions one and two. It also demonstrates a propensity within data visualisation research to approach the design and interpretation of data visualisations as two separate issues. As there is limited data studies research into data visualisations from a Sociological perspective, the chapter also draws on research from Human Computer Interaction studies and Media Studies.

The second part of the chapter analyses existing research into the design of data visualisations and users' engagement with them, through three themes identified in the literature. The first, *Data visualisation conventions*, highlights how existing research into the production of data visualisation has tended to focus on the image of graphs and charts as the site of analysis. Much of this research takes a semiotic approach to analyse how visualisations communicate meaning through various visual signs (Engebretsen and Weber, 2017). This section argues how the literature neglects the meaning making practices that take place before and after the image of the visualisation comes into being. Drawing on Kress and van Leeuwen (2001), I consider these to be important moments of meaning making and this thesis directs attention to these moments by interrogating how designers *encode* data visualisations with meaning and how ordinary users make meanings from them. The second theme, *Story telling in the design of data visualisations,*

reviews existing literature that has focused on the narrative and rhetorical techniques employed in data visualisation design in order to communicate information and meaning to users. Reviewing this research identified two gaps in the literature. Firstly how rhetorical strategies shape how users engage with and make meanings from data visualisations. Secondly, an understanding of the practices that shape how designers explore the data in order to find a *story* to visualise. The third theme, *User engagement with data visualisation*, explores research that moves beyond the image of graphs and charts to understand how users engage with them. This revealed how existing literature has focused on effectiveness and engagement, identifying a gap in understanding in relation to how users *make meanings* from data visualisation through their interpretive practices.

2.2 Data Subjectivities

The first part of this literature review situates the project within the interdisciplinary field of data studies, which is the critical context of this thesis. By introducing the concept of big data and datafication it will explain the socio-cultural context which has resulted in what Beer and Burrows describe as a "culture of visualization", in which data visualisations increasingly circulate (2013, p. 62). The field of data studies is broad and therefore the literature is discussed through the lens of data subjectivities. By this, I mean the many ways in which subjectivities permeate data through the socio-technical processes involved in their generation, collection, analysis and visualisation (boyd and Crawford, 2012, also see Kitchin, 2014). This is in contrast to the mythology that surrounds big data, which promotes its ability to produce new, accurate and objective insights (Crawford, Miltner and Gray, 2014). Situating the project among literature on data subjectivities is important. The claims made through graphs and charts are persuasive because of the "aura of objectivity" that surrounds data and their visualisation (Kennedy et al., 2016a, p. 723). Being attentive to the many ways data is subjective challenges data visualisations' perceived neutrality. The section begins by explaining the concepts of big data and datafication before moving on to introduce the politics of data. Finally, the section will discuss literature that relates to the various ways it is possible to *make sense* of data.

2.2.1 Big data and datafication

In order to research data visualisations from a data studies perspective, it is first important to think critically about the big data phenomenon and datafication to understand the context in which data visualisations are proliferating. Kitchin describes data as "raw material produced by abstracting the world into categories, measures and other representational forms" (2014, p. 1). Data can take many forms, including numbers, text, and images, and can be analysed to

generate knowledge and insights that not only tell us about the world but are used to shape it (Kitchin, 2014). Although the term big data began to appear in scientific circles during the 2000s (Mayer-Schönberger and Cukier, 2013), large datasets have been used to generate insights and knowledge within academic, commercial and institutional arenas for a long time (Kitchin, 2014). However, boyd and Crawford argue that it is not the size of datasets that define 'big data,' rather it is the "capacity to search, aggregate and cross reference large datasets" (2012, p. 663). There is no consensus within academia about the exact definition of big data, however, I will draw on characteristics compiled by Kitchin (2014). Drawing on existing literature, Kitchin describes big data as possessing, "volume, velocity and variety," so that it is exhaustive in reach, fine grained in detail, relational and indexical in structure and scalable in functionality (2014, p. 68). Kitchin (2014) also credits developments in technology, which have allowed for the digitisation of data as well as real time data generation and collection, for enabling the formation of vast infrastructures of data.

The generation and collection of big data has been fuelled by technological developments such as networked computing, which has enabled the routine and indiscriminate collection of data about all aspects of people's online lives (Couldry, 2016). This has enabled qualitative elements of people's lives to be quantified through processes of *datafication,* by which social phenomenon of all kinds is transformed into data (Mayer-Schönberger and Cukier, 2013). Kennedy defines datafication as, "the quantification of aspects of life previously experienced in qualitative, non-numeric forms, such as communication, relationships, health and fitness, transport and mobility, democratic participation, leisure and consumption" (2018, p. 18). This definition provides a sense of the scale of the data being generated and gathered from people's everyday actions (Kennedy, 2018) and reveals how most big data is actually *social data,* produced through and about the social world (Uprichard, 2013). Datafying previously qualitative information allows it to be used "in new ways," primarily through processes of analysis which aim to extract value from data through the generation of insights (Mayer-Schönberger and Cukier, 2013, p. 15).

Although it is a significant factor, big data has not emerged only in response to developments in computational technology (Crawford, Miltner and Gray, 2014). Indeed, it can be useful to think about big data as a "mythology" (Crawford, Miltner and Gray, 2014, p. 1663), as this allows researchers to consider the "economic, political and cultural forces at work", which promote and legitimise it as a means of knowledge production (ibid, p. 1664). boyd and Crawford consider big data itself to be an emergent "system of knowledge" which is transforming research and how we understand the social world (2012, p. 665). The mythology of big data centres on the idea that with large datasets, computational power and new methods of analysis comes hitherto

unattainable knowledge and insights, which are both objective and accurate (boyd and Crawford, 2012). However, the powerful rhetoric of objectivity that surrounds big data (Crawford, Miltner and Gray, 2014) conceals how data are never neutral or free from human bias (boyd and Crawford, 2012). As Crawford states, data are "creations of human design" and they rely on human interpretation to give them meaning (2013, n.p.).

In explaining the aggregative quality of data, Gitelman and Jackson describe how data "pile up" (2013, p.8). This is a useful metaphor that helps to conceptualise data as a material resource; masses of individual fragments of information, each possessing latent potential that can be harnessed through their aggregation (Gitelman and Jackson, 2013). Of course, data's capacity to reveal new insights hinges on its ability to reveal the hidden connections that exist between these individual fragments, thus highlighting the complex ways in which data hovers on the dichotomy of singular and plural (Gitelman and Jackson, 2013, p.8). Indeed, visualisation is one method of revealing patterns and relationships concealed in large datasets (Manovich, 2011). Although Gitelman and Jackson (2013) are highlighting the materiality of the data itself, it is also important to consider the material infrastructures in which data are embedded (Kitchin, 2014) as well as the material outcomes that their circulation bring into being (Beer, 2016).

Kitchin describes big data as a socio-technical phenomenon, arguing that big data are generated, framed and used within a socio-technical data assemblage, shaping "what is possible, desirable and expected of data" (2014, p. 24). This assemblage is made of up interconnected apparatuses, which include material infrastructures, systems of thought, forms of knowledge, regulatory frameworks, finance, people, places, and practices (Kitchin, 2014, see p.25). This demonstrates how data does not simply emerge in the world as a raw material ready to be mined for insights; actors working within specific contexts and structures generate it (Kitchin, 2014). Indeed, the concept of the data assemblage is particularly useful as it connects the social, political, economic and technological dimensions of data to demonstrate how data cannot be uncoupled from the contexts in which it is generated, stored and used (Kitchin, 2014). The context specific nature of data assemblages and the apparatuses that make them up means that they differ between spheres (Kitchin, 2014). Therefore, Kitchin (2014) argues that further research is required to assess how existing and new data assemblages are formed and function. However, this thesis is interested in understanding how data assemblages frame the production of knowledge through visualisation. This is the basis of my third research question, which asks, *what is the relationship between meaning making and data assemblages?*

This section has briefly introduced the concept of big data as a socio-technical phenomenon (boyd and Crawford, 2012; Kitchin, 2014) and explained how advances in technology have led

to the datafication (Couldry, 2016) of aspects of the social world that were previously experienced qualitatively (Kennedy, 2018). It has argued that Kitchin's notion of the data assemblage provides a conceptual tool through which to consider the complex socio-technical structures that data is embedded within and which shape the production of knowledge through data. The following section builds on this introduction to discuss the politics of data in order to consider how this frames its visualisation.

2.2.2 The politics of data

Beer highlights the inherent politics and power of data by focusing specifically on metrics, which he describes as data that represents a "measure of the world" (2016a, p. 3). Big data has brought about an "intensification of measurement," while also being a significant way by which metrics are circulated and become embedded in the social world (Beer, 2016a, p. 9). Espeland and Stevens explain how measurements are "reactive;" they not only represent aspects of the social world, they shape social worlds and the behaviours of those within them (2008, p. 412). They can do this because they make visible and reinforce existing boundaries between "categories of people and things" and even create new ones (Espeland and Stevens, 2018, p. 414). Therefore, metrics are an important part of how the world is ordered and how value is judged (Beer, 2016a). Focusing attention towards the circulation of metrics reveals how data is not passive, nor can it be regarded as neutral fragments of information, instead it is implicit in social ordering and governing (Beer, 2016a). Metrics have purpose and intent, they not only describe what is going on but they also make things happen as they shape the world and our understanding of it (Beer, 2016a). It is possible to view data visualisations as the materialisation of metrics. Indeed, Espeland and Stevens (2008) note how graphical depictions of data are part of its aesthetic, which is itself carefully governed. Visualisation is embedded in data's politics because it is an important way through which complex quantitative information becomes comprehendible and therefore, actionable in the social world (Espeland and Stevens, 2008).

Despite data being social, political and cultural constructions (Beer, 2016a), numbers are persuasive and have authority (Espeland and Stevens, 2008). This is reflected in the way that data and quantification has permeated everyday aspects of the social world including through public administration and policymaking (Rieder and Simon, 2016). Part of big data's appeal in these spheres is its perceived objectivity, which fosters a "trust in numbers" among the public (Rieder and Simon, 2016, p. 3). This trust, combined with the perceived mechanical objectivity associated with automated data collection and analysis, works to de-politicise decision-making (Rieder and Simon, 2016). As Porter explains, "quantification is a technology of distance" which make bureaucratic decisions appear fair and unbiased by concealing behind the numbers

the very act of deciding (1995, p. ix). Beer (2016b) argues that in order to understand its power, researchers need to direct attention towards the discourse that surrounds big data and the work that this does. The concept of big data frames how data is imagined and the possibilities it is afforded through a powerful discourse, which presents data as an objective, neutral and accurate source of knowledge (Beer, 2016b). Therefore, Beer argues that it is impossible to decouple the concept of big data from its "material phenomenon" (2016b, p. 9). As such, it is possible to consider the discourse that surrounds the concept of big data as part of the socio-cultural context in which data visualisations are produced, interpreted and valued as representations of the social world (Kennedy et al., 2016a).

Despite the claims made about big data's accuracy, biases permeate large datasets at various points during their creation, analysis and visualisation. Crawford (2013) explains how communities who do not emit digital signals are often excluded from digital datasets. Yet important decisions are made from the analysis of these datasets, including the allocation of public resources (Crawford, 2013). Crawford (2013) describes this as a form of hidden bias, which occurs during both the collection and analysis of data. While Nafus (2014) explains how coding phenomena into numbers provides a means to materialise abstract elements of the social world, Crawford (2013) reminds us that data itself is an abstraction of these phenomena, concealing data's links to people and places. Indeed, boyd and Crawford (2012) argue that in order to accurately analyse big data it is crucial to understand the context in which the data was generated (also see, D'Ignazio and Klein, 2020). Without an understanding of how data is generated, collected and analysed, pre-existing material and social inequalities may become further entrenched through its use (Crawford, 2013; D'Ignazio and Klein, 2020). Indeed, Kitchin (2017) argues that computer technology is accelerating classification and sorting practices that reinforce existing ways of thinking. While data visualisations are an important way in which data are mobilised (Gitelman and Jackson, 2013) they stand accused of simplifying and de-contextualising the complex issues they represent, while at the same time seeming to present the facts (Boehnert, 2015). This, Boehnert (2015) argues, allows data visualisation to do ideological work (also see, Kennedy and Hill, 2017). This demonstrates how both data and their visual representation are embedded in politics, a point which Hill (2017) examines in her research on data visualisations representing abortion.

Unintentional biases in datasets can be attributed to both the technical processes through which the data were generated and manipulated and the ideology and intent of the analyst (Kitchin, 2014). Indeed, boyd and Crawford (2012) highlight the role of human judgement in the analysis of data, arguing that designing methods of analysis and attempting to understand what the data means are both acts of interpretation. These interpretative acts are framed by individuals'

subject positions and particular ways of viewing the world (Crawford, Miltner and Gray, 2014). This demonstrates how data "is never ontologically prior to interpretation" (Boelstorff, 2013, n.p.). This is significant when one considers how access to data and the skills to analyse it is skewed towards well-educated men (boyd and Crawford, 2012). This narrow subsection of society are disproportionately influencing the knowledge produced from data (boyd and Crawford, 2012) and this matters in relation to class, gender, sexuality and race (Crawford, Miltner and Gray, 2014, p. 1665).

Although all analysis is an act of interpretation, only some people have the access and skills required to make legitimate claims from data (see Gabrys, Pritchard and Barratt, 2016). This is also true of data visualisations. Visual representations are governed by a network of gatekeepers (Espeland and Stevens, 2008) and established design conventions (Kennedy et al., 2016a). The ability to produce and interpret them requires specific skills that can vary between domains (Espeland and Stevens, 2008). Therefore, those in society with such skills possess more power to produce visual representations of data than those without (Boehnert, 2015). As Beer (2019) argues, those who possess the authority to speak with and tell stories with data wield power over what data is used to reveal. Therefore, it is important to understand how data visualisations are both produced and consumed within this context.

This section has introduced the politics inherent in the generation, collection and analysis of data, which frame the meanings made from data and its visual representation. These politics are part of the data landscape in which data visualisations are designed, circulate, and are interpreted. By demonstrating the inherent subjectivities built into datasets and their analysis, this section has begun to unpick the perceived neutrality of data visualisations (Kennedy et al., 2016a). Furthermore, the literature prompts questions around how the politics of data play out in how meaning is made in the design and interpretation of data visualisations. Although the politics of data does not relate to a specific aim of my research, it forms part of the social and cultural context within which the meaning of data visualisations are encoded and decoded. Furthermore, an understanding of how meaning is made through their design and interpretation will contribute to existing debates around the politics of data.

2.2.3 Making sense of data

While powerful claims of mechanical objectivity separates the knowledge produced through quantification from "the particular people who make it" (Porter, 1995, p. ix), there is human labour involved in making meaning from data (Nafus, 2014). Nafus (2014) uses the example of open source citizen sensor data to make visible the labour required among her participants to

make *their* data meaningful. While the designers of sensors imagine how the data they bring into being might be put to use, this is often out of kilter with the "lived contexts" of the users of these devices (Nafus, 2014, p. 210). This disjuncture means that for the users of home energy sensors, the social meaning of the data they generated and uploaded to open source data sites was uncertain (Nafus, 2014). It is in such instances, Nafus (2014) argues, that the labour required to make data meaningful becomes visible.

During interviews, Nafus' (2014) participants reflected on the sensor data they collected about their home energy consumption. Despite having a personal connection to the contexts in which the data was generated, they struggled to make meanings from it, in part, because the contexts within which the data would have meaning were "always out of reach" (Nafus, 2014, p. 216). The participants were disappointed to find that they could know far less from the sensor data than they had expected, instead it tended to trigger new questions relating to context, such as which of their home appliances used the most energy. Furthermore, despite sharing their sensor data on an open source platform, the participants did not possess the skills required to aggregate their data with other datasets on the platform to generate new meanings (Nafus, 2014). Nafus explains that unless such open data can be "re-contextualized and re-interpreted" it remains meaningless (2014, p. 221).

By drawing attention to how home sensor data is purposefully vague, Nafus' (2014) article demonstrates how social and cultural practices are key in forming both users' expectations of what can be known through sensor data and the human effort required to make it meaningful. However, there is a politics to making sense of data. Without the skills to collect and analyse quantitative data it can be hard for communities to use data to make legitimate evidence claims (Gabrys et al., 2016). Gabrys et al., (2016) conducted a participatory research project with a community in Pennsylvania, working with them to collect, analyse and present data on a local environmental issue. The project produced quantitative sensor data but also qualitative data in the form of citizen accounts, images and maps. While this was rich contextual data, Gabrys et al., (2016) explain the significant labour involved in combining and transforming these diverse datasets into a coherent narrative that would be accepted as evidence by regulatory bodies. The authors describe how they had to balance the need to have the data taken seriously with their wish to preserve the alternative perspectives and contextual nuances afforded by the qualitative forms of data (Gabrys et al, 2016). This struggle not only highlights the narrow perception of 'legitimate data' held by some scientific and regulatory fields but also suggests a broader politics of data collection, in terms of *what* counts as data and *who* is authorised to collect and make sense of it.

Dourish and Gómez Cruz (2018) argue that 'we' make sense of data through the narratives we construct around it. They explain that the process of narrating data relies on an understanding of it as representing an event, person or phenomena in the world (Dourish and Gómez Cruz, 2018). This prompts those interpreting what the data means to call on "narrative elements," such as "actors, motives, expectations […] [and] histories" in order to tell stories from the data (2018, p.4). For example, the authors explain how data from devices associated with the internet of things might evoke wider narratives around domesticity and routine that can be used to contextualise the data (Dourish and Gómez Cruz, 2018). Therefore, meaning is made from data through an entanglement of "data practices and narrative practices" (Dourish and Gómez Cruz, 2018, p. 4). Espeland (2015) explains that quantification simplifies the things that data represents. This prompts people to come up with narratives that explain the data, both in relation to what the data represents and how the data is being used to make knowledge (Espeland, 2015). However, Nafus' participants were unable to invoke such narratives to make meanings about their own home sensing data, suggesting that this strategy is not always successful.

As quantitative data circulates, it provokes new narratives that explain or account for what it means (Espeland, 2015). Espeland's chapter suggests that by being attentive to these narratives, it is possible to understand how individuals make meanings from data. However, the source of narratives used to explain data are not always human. Veel's research explains how automated technology such as "Natural Language Generation software" can transform data into stories (2018, p. 2). Veel (2018) explains how narratives have become an important discourse within big data analytics and this relies on the assumption that all datasets contain stories that are just waiting to be found. Both Dourish and Gómez Cruz (2018) and Espeland's (2015) work suggest that narratives are an important way through which data is contextualised in order to make meanings from it, while Veel's (2018) work highlights the ways in which commercial analytics companies are tapping into the rhetoric of narratives to generate value from data. In aiming to understand how ordinary users make sense of data visualisations this project will be attentive to how narratives might provide a strategy through which participants make meanings from them.

Although they all refer to meaning, neither Nafus (2014), Dourish and Gómez-Cruz (2018) or Espeland (2015) conceptualise meaning making in their discussions. While Nafus' article suggests that it relates to how data is made useful or operational in the social world, the latter authors imply that meaning is also discursive. However, data can also be considered relational, that is, it becomes meaningful to people through their relation to what it represents. In researching how people engage with data and make sense of it when presented with an opportunity to use data to address their concerns and to affect change in their communities, Taylor, Lindley, Regan and Sweeney (2014) reveal how data is materialised through their

connection to people and place. Furthermore, like Gabrys et al., (2016) they highlight the contested nature around *what* data is collected and *how* it is interpreted (Taylor et al., 2014). Their participants, members of the community of Tennison Road in Cambridge, prioritised different sources of data and different interpretations of them depending on their interests or feelings towards particular local issues (Taylor et al., 2014). This draws attention to the difficulty of using big datasets to speak for whole communities (Crawford, Miltner and Gray, 2014), and the way in which even *small data* are embedded in social relations and lived experiences (Taylor et al., 2014). Importantly in the context of this project, it also implies that the meanings made from data, and perhaps data visualisation, are relational in terms of how the data connects to individual people and places.

In order to make sense of large volumes of diverse and relational data, new forms of *visual analytics* have been developed and deployed (Kitchen, Maalsen, and McArdle, 2016). This is because visualisation is viewed, by the growing analytics industry, as a tool through which insights from data can be made accessible to users, regardless of their data analysis skills (Beer, 2019). These range from simple graphs and chart types to complex digital models that users can interactively engage with (Kitchin, Maalsen and McArdle, 2016). Indeed, Amoore and Piotukh (2015) argue that analytics play an important role in sense-making, by identifying an object of interest among the dataset and singling it out for attention (2015, see p. 349). In doing so, analytic tools strip the data of its context and create particular vantage points through which to view the data (Amoore and Piotukh, 2015). Often this is through processes of visualisation (Beer, 2019). Indeed, Manovich (2011) describes how data visualisation reduces data sets to particular spatial relationships in much the same way as Amoore and Piotukh (2015) describe the work of analytics. This, Amoore and Piotukh argue, distances the data from its context and narrative as the focus becomes identifying "people, objects or patterns of interest" within the dataset (2015, p. 359). This suggests that making sense of large datasets will only ever produce partial meanings, which are at the same time both selective and abstract. This because the data is distanced from the contexts in which it was generated and from whom it produces knowledge about. Amoore and Piotukh's (2015) article also implies that making sense of data is a political endeavour, as data is *viewed* from particular perspectives. This relates to Haraway's (1988) argument that all knowledge is situated. Therefore, it is important that the thesis remains attentive to how politics unfold in the ways designers and ordinary users make meanings from data through and from visualisations.

The authors covered in this section do not define what they mean by sense making or meaning. However, they still offer important insights that may be useful in understanding how people make meanings from engagements with data visualisations. Their work also reveals the human

labour and therefore, the inherent politics involved in making sense of data, which is so often portrayed as transparent and objective (Porter, 1995). In summing up this section on data subjectivities, by introducing the concepts of big data and datafication, this part of the chapter has set out the social, cultural and technological context within which data visualisations are proliferating. As explained, Kitchin's (2014) notion of the data assemblage will be used as a conceptual tool to consider the broader socio-technical context within which data visualisations are made and how this shapes the knowledge they produce. By introducing the politics of data this section has also begun to set out why research into how designers and users make meanings from data through visualisation is important. Although it is not the aim of this thesis to unpick the politics of meaning making, as this section has demonstrated, politics frame how data is thought about and made sense of. Therefore, developing an understanding of how meanings are made from data through visualisation may contribute to this debate by challenging the rhetoric of objectivity that surrounds data visualisation and revealing the subjectivities inherent in all interpretations of data.

2.3 Data Visualisation

Data visualisations are the medium by which most people who are not experts access data (Kennedy et al., 2016b; Kennedy and Hill, 2018; Kennedy, 2018). Indeed, Kennedy and Hill (2018) argue that data become accessible and perceptible to non-experts through their visualisation in graphs and charts. Therefore, understanding how meaning is encoded into data visualisations and then interpreted by users provides a way to explore how non-experts experience and engage with data. This section of the chapter reviews existing literature on data visualisations. It discusses research that focusses on the image as a text for analysis, the design of data visualisations, and engagement with graphs and charts by non-expert users. In doing so, it points out a tendency to consider the production of data visualisations and ordinary people's engagement with them as two separate issues. Furthermore, this section highlights how much research on meaning and data visualisation focuses on the image of the graph or chart as the site of analysis. This neglects to explore the meaning making practices at work in the design of visualisations or during their interpretation. As Kress and van Leeuwen (2001) argue, meaning is made during the discourse, design, production and distribution phases; therefore, these are important sites of analysis when understanding how meaning is made from data through visualisation.

By situating this project in relation to existing empirical literature on data visualisation, this part of the chapter will demonstrate the need for research that tracks the meaning making practices at work in *both* the design and interpretation of graphs and charts. It argues that by exploring both

design and interpretation practices, this thesis will provide an account of *how* meaning is made from data through and from its visualisation. This will help to drive debates within data studies forward in relation to how people engage with and experience data.

The section begins by discussing literature on the conventions that exist within the field of data visualisation design. This literature focuses on research about designers' practices, predominantly researched through the analysis of graphs and charts as a visual text. The second part of the section discusses research that explores the role of storytelling and narration in the design of data visualisations. This focuses on the rhetorical strategies designers use to communicate meaning from data when producing graphs and charts. Although some of this research draws on interviews with designers, the literature is predominantly based on analysis of images of graphs and charts. The final part of this section directs attention towards existing research on how users engage with and make sense of data visualisations. By discussing these three areas of research on data visualisations, this chapter will identify the gaps in the existing literature that this thesis addresses.

2.3.1 Data visualisation conventions

The conventions of data visualisation refer to the shared organisational principles, practices and standards that exist in the field of data visualisation design (Kennedy et al., 2016a). To understand how meaning is made in the production of data visualisations it is important to look at literature that examines the conventions that frame their design. The section will begin by defining what a data visualisation is by drawing on academic and design literature. It will then go on to discuss research that has predominantly analysed images of graphs and charts in order to examine the conventions that shape their production. In doing so, it is mindful of the politics of data visualisation by highlighting the subjectivities involved in visualising data, which are so often concealed behind their design (D'Ignazio and Klein, 2016). In reviewing this existing literature on the design conventions of data visualisations, this section reveals how such an approach tends to focus attention on the image of the graph or chart as the site of analysis. This neglects potentially important moments of meaning making that occur before and after the visualisation comes into being. For example, the data and design practices of the designers who produce data visualisations and the interpretive practices of users when engaging with graphs and charts.

Kennedy et al., define data visualisations as, "visual representations of data and datasets which communicate precise information and values" (2016a, pp. 715-716). They argue that the visualisation of precise values separates data visualisations from infographics, which use visual

elements to provide the user with a "sense of" the data they represent (p. 716). Engebretsen and Weber extend Kennedy et al's definition by defining data visualisations as, "a visual representation of data created to amplify the cognitive processing and the social application of the data represented" (2017, p. 278). While Kennedy et al., (2016) refer to the communication of data, Engebretsen and Weber (2017) foreground data visualisation's ability to aid users' cognitive processing and perception of data. Data visualisation design practitioner, Andy Kirk, defines them as, "The representation and presentation of data to facilitate understanding" (2016, p. 19). Kirk outlines the distinction between representation, which are the visual forms and design choices used to *represent the data*, and presentation, which are the other design choices and visual elements that make up the design of the visualisation (2016, see p. 21). In their definition, Kirk (2016) focuses on data visualisation's role in facilitating users *understanding* of data. Although they differ, all three of these definitions involve two aspects, the visual representation of data and users' perception of data.

Definitions of data visualisation and definitions of information visualisation occasionally overlap (Engebretsen and Weber, 2017). The term information visualisation is associated with computer science and human computer interaction studies, while data visualisation tends to be associated with various fields including statistics, information design, and journalism (Engebretsen and Weber, 2017 see p. 279). Manovich defines information visualisation "as mapping between discrete data and a visual representation" (2011, p. 37). While this definition does not foreground how visualisation can aid users' perception of data, Manovich's (2011) discussion of the visual elements and processes of reduction used to represent the structure of the data are extremely relevant to data visualisation. Therefore, the following literature includes authors who write about data visualisations and information visualisations. Drawing on the various definitions discussed above, I define data visualisation as, *the visual representation of data to communicate information.* I believe that this clear definition will allow the thesis to consider the varied ways in which datasets are represented visually in graphs and charts. Including the communication of information maintains a focus on how users perceive data through visualisation.

Manovich (2011) identifies two main principles of information visualisation; reduction and spatial variables. Visualisation reduces the phenomenon, events, or people the dataset represents to, "graphical primitives such as points, straight lines, curves, and simple geometric shapes" (Manovich, 2011, p. 38). These graphical primitives are used to show the relationship between data and enables the user to see patterns. However, Manovich (2011) explains that this requires designers to ignore the vast majority of the data in order to expose patterns across a small percentage of it. Boehnert (2016) argues that data visualisations are only ever partial, and this

form of reductionism can result in an oversimplification of the complex issues they often represent. Manovich's (2011) second principle refers to the way in which spatial variables are given greater significance than other visual elements in the design of visualisations. The spatial layout of the visualisation represents the most important relationships between the data while other visual elements, such as colour, aids readability (Manovich, 2011). Drucker argues that visualisations make "spatial relations meaningful" by drawing on conventions which make assumptions about how people perceive and make sense of visual information (2014, p. 66). While Manovich was not explicitly concerned with how *meaning* is made from data through visualisation, his article implies that meaning emerges from large datasets through the identification of patterns and relationships between data. However, his article raises questions pertaining to the politics of data visualisations in relation to *which data* is prioritised in the reduction process, how this frames the meanings that 'emerge,' and to whose benefit and expense (Boehnert, 2016).

Engebretsen and Weber (2017) provide a discussion of different analytical approaches to data visualisation in relation to their *visual language*. They draw on a range of perspectives including semiotics, Gestalt theory, and statistics to provide an interdisciplinary overview of how data visualisations produce meaning through a system of visual signs (Engebretsen and Weber, 2017). The semiotic and linguistic perspectives covered in their chapter focus on the visual elements of data visualisations rather than the practices of designers or their interpretation by users. The authors partially identify this gap in the literature, arguing that much more research is required into how different users interpret graphs, maps, and charts (Engebretsen and Weber, 2017). This thesis argues that to do so requires a shift in attention from the image as the site of analysis towards the interpretive practices of users. Nevertheless, semiotic approaches have proved useful among researchers who want to unpick the rhetoric of objectivity that data visualisations possess (see Kennedy et al., 2016a).

Kennedy et al., use a social semiotic approach to explore data visualisation in relation to the "contexts in which they are embedded and the structures that shape them" (2016a, p. 716). The authors combined interviews with visualisers with semiotic analysis of data visualisations to reveal the conventions that work to present data as objective (Kennedy et al., 2016a). Kennedy et al., (2016a) are careful to draw a distinction between the designers who create graphs and charts and the conventions within which they work. They argue it is the conventions that produce the appearance of objectivity rather than the data visualisation designers (Kennedy et al., 2016a). The authors define a convention as a "symbolic or social practice that is shared, readily understood and widely accepted by members of a cultural group," in this instance the community of data visualisers (2016a, p. 717). Their article identifies four conventions: "two-

dimensional viewpoints;" "clean layouts;" "geometric shapes and lines;" and "the inclusion of data sources" (Kennedy et al., 2016a, p. 716). These conventions frame the design of graphs and charts and therefore shape how meaning is encoded through their production. However, Kennedy et al's., (2016a) research does not address how these conventions play out in how different ordinary users engage with and make sense of data visualisations.

Hill (2017) builds on the work of Kennedy et al (2016a) to research how the conventions of data visualisation and their rhetoric of objectivity are being used to potentially limit women's rights to abortion services. Hill (2017) uses a social semiotic approach to analyse a sample of graphs and charts obtained through google images, which represent the topic of abortion. As well as analyzing the visual elements of the visualisations, Hill's analysis includes the context in which they would be viewed by including a *reading* of the web pages where they were published (Hill, 2017). Hill found that anti-abortion groups often presented data without providing sufficient context about how that data was generated or how the visualisation was produced. However, by citing credible sources of data, an important convention of data visualisation design (Kennedy et al., 2016a), the visualisations had credence (Hill, 2017). Despite this lack of information, the visualisations Hill (2017) highlights in her article tended to present the data as facts, inviting users to *see* what the visualisation shows for themselves, which made them very persuasive. Furthermore, Hill (2017) argues the visualisation of abortion data creates distance between the viewer and the real life experiences of women through their rhetoric of objectivity. This enables such visualisations to make powerful arguments that have political implications for women's rights.

Hill's (2017) research demonstrates the powerful work that data visualisation conventions do in shaping how viewers perceive data. Furthermore, it demonstrates the importance of researching how meaning is made in the design and interpretation of data visualisations in order to better understand *how* graphs and charts do political work. However, by focusing her analysis on the images of visualisations, Hill's paper presents a somewhat deterministic view that neglects to explore users' agency in how they made meanings from them. By making the image of graphs and charts their analytical focus, Kennedy et al., (2016a) revealed the structural conventions that shape how data is visualised and how these conventions produce a powerful rhetoric of objectivity. Hill (2017) explains how these conventions allow data visualisations to do political work. However, research that focuses on the image as the site of analysis neglects the meaning making work that is done before and after the graph or chart comes into being. By this, I mean the data and design practices of the designers who produce visualisations and the interpretive practices of users who engage with them. This research project addresses this gap in understanding by directing attention to these moments in the construction of meaning.

The conventions of data visualisation design are part of a longer history of visualisation principles dating back to the Enlightenment (Kennedy et al., 2016a). These principles are embedded in notions of rational knowledge, reason, and objectivity, while rejecting emotion, passion and subjectivity (Kennedy and Hill, 2018). Indeed, information designer Edward Tufte describes data visualisations as, "instruments for reasoning" (2001, Introduction). However, Kennedy and Hill (2018) argue that emotions are an important aspect of how users engage with data visualisations. They found that emotions influenced if a user engaged with a visualisation and the information they acquired from it (Kennedy and Hill, 2018). The authors argued that despite the rational discourse that surrounds data, their participants also experienced emotional responses to data visualisations that shaped their engagement with them (Kennedy and Hill, 2018). This demonstrates how focusing on design conventions as the site of analysis neglects the perhaps unexpected ways that ordinary users engage with and make meanings from data visualisations.

Kennedy and Hill (2018) note how the rhetoric of rationality and reason associated with data privileges particular groups in society. Arguably, those who are more likely to have education or employment in maths, science, and computational subjects - usually "white middle-class men" (Kennedy and Hill, 2018, p. 845). Their argument suggests that not only are legitimate representations of data (boyd and Crawford, 2012; Boehnert, 2015) shaped by this rhetoric but that *legitimate interpretations* of data are also limited to the rational (Kennedy and Hill, 2018). Indeed, Halpern (2014) argues that the act of observing relies on historically situated systems of conventions and affordances. While visualisation can make what was once beyond human perception perceivable, how one *sees* the image is part of a broader historical, cultural, and technological context (Halpern, 2014, see pp. 22-23). Therefore, the conventions of data visualisation design combined with those that govern vision may work to de-legitimise alternative ways of making meanings from graphs and charts, and exclude certain groups from engaging with them (Kennedy and Hill, 2018). As such, it is important to be attentive to the different ways people make meanings from data visualisations, which may not be restricted to the rational, including the resources and practices they employ to do so.

The conventions that frame the production of data visualisations help to conceal the politics inherent within them. Despite their rhetoric of objectivity, "data visualisations are not neutral windows onto data" (Kennedy and Hill, 2017, p. 773). Graphs and charts reflect and reproduce particular ways of looking at the world, which often work to reinforce the dominant power relations that exist in society (Kennedy and Hill, 2017). Indeed, Drucker describes information visualisations as "acts of interpretation masquerading as presentation" (2014, n.p.). While their conventions help make data visualisations appear as though they allow one to simply *see* the

facts (Kennedy et al., 2016a), Drucker argues they are in fact "arguments made in graphical form" (2014, n.p.). Drucker describes how visualisations conceal both the interpretive processes that are encoded into graphs and charts, and the ways in which their image operates as argument. Indeed, in making the case for data visualisations that foreground the data over aesthetics, Tufte calls for transparent designs that are "so good they are invisible" (1990, p. 33). Drucker credits the geometric forms, which are one of the conventions of data visualisation design, for providing them with an air of simplicity and transparency. Not only does this conceal the design decisions that shaped the visualisation but it also masks the way data is the product of interpretive constructions rather than a natural resource (Drucker, 2014). This demonstrates how the politics of data visualisation are not only embodied in the visual design of a graph or chart but also in the generation and selection of the data to be visualised (Boehnert, 2016). This suggests that to understand how meaning is made from data through its visualisation, research must pay attention to the data practices that take place prior to its visualisation as well as the context in which the visualisation comes into being.

Before data can be visualised, it must be generated, collected and cleaned (Kennedy and Hill, 2017). Cleaning data imposes structure on 'messy' datasets, allowing them to be analysed and visualised (D'Ignazio and Klein, 2020). However, D'Ignazio and Klein (2020) argue that by cleaning data, analysts remove much of its context, thus imposing their own perspectives on the data at the cost of other perspectives that may exist in the dataset. Once the dataset is ready to be visualised, further choices are required in relation to the software used to visualise it, the chart type used to represent it, and what the visualisation is being designed to do (Kennedy and Hill, 2017). Kennedy and Hill (2017) explain how this includes decisions on aesthetic factors, such as colour, scales and graphical forms, which they argue are used to communicate meanings which have already been gleaned from the data. These various processes highlight the ways in which the meaning of data is framed prior to its visualisation. This presents a significant gap in current understandings of how meaning is made from data through visualisation, as literature has tended to focus on design conventions through the visual elements of graphs and charts. This gap supports this thesis' analytical focus on the data and design practices of the designers who make them, how these practices shape how meaning is made from data, and how it is encoded into visualisations.

D'Ignazio and Klein describe their vision for a feminist approach to designing data visualisations, in which the "situated nature of knowledge" is acknowledged in both the production and interpretation of graphs and charts (2016, 2.1). Several theoretical and practical fields have informed their ideas, including Feminist STS, which the authors credit with challenging the objectivity of science and technology through emphasising how knowledge is

embedded in social, political, cultural, and economic contexts (D'Ignazio and Klein, 2016). D'Ignazio and Klein (2016) suggest that there is a need for more creative ways of collecting and presenting data that allow for more fluidity and multiplicity. This links to Boehnert's (2015) argument that data visualisations often oversimplify the issues they represent. Furthermore, Boehnert argues that data visualisations are the products of "the explicit intentions and implicit assumptions of designers" (2016, n.p). Indeed, D'Ignazio and Klein's vision for a feminist data visualisation proposes that the subjectivities of data visualisation designers be made explicit by acknowledging how their work is influenced by conscious and unconscious decision-making. They go on to argue that designers should explore the specific social, material and cultural contexts in which data visualisations are produced and received by users, and the embodied human labour that is central to data practices (D'Ignazio and Klein, 2016).

D'Ignazio and Klein's (2016; see also 2020) provocations encourage researchers to look beyond the role of visual designers. Their principles of feminist data visualisation prompt researchers to draw connections between communities of users, data visualisation design teams and those who have collected, cleaned, and been custodians of the data prior to its visualisation. By drawing on Kitchin's (2014) apparatuses of data assemblages it is possible to expand this idea further to explore how the human labour of users, designers, and data scientists is embedded in a wider assemblage of people, places, thought systems and infrastructures. By building up a picture of how data visualisations are connected to a varied network of people, places and technology it might be possible to make visible the social, cultural, political and economic contexts that shape the meanings of data visualisations. Mapping these connections offers a way of revealing the political and situated nature of knowledge in relation to the production and interpretation of data visualisations (D'Ignazio and Klein, 2016). As Aiello (2020) states, D'Ignazio and Klein (2016) are focusing on the practice of visualising data rather than on how meaning is *made* through their production. Nevertheless, as set out in Chapter one, the discourse, design, and production practices involved in the design of data visualisations are important sites of analysis in understanding how meaning is constructed (Kress and Van Leeuwen, 2001). By focusing on these moments, it will be possible to analyse the meaning making practices at work prior to the image of the graph or chart coming into being.

This section has introduced research on the conventions of data visualisation, a significant approach to the study of how meanings are communicated through visualisations. In doing so, it has also considered the politics inherent in the design of graphs and charts. However, much of this literature focuses analytical attention on the image of data visualisations. This reveals a gap in understanding in relation to the data and design practices that happen prior to the design of the visualisation and the interpretive practices employed by those who engage with graphs and

charts. These moments, before and after the visualisation is produced are important sites of meaning making. This thesis directs attention to these moments by asking how designers encode data visualisations with meaning and how ordinary users make meanings from data visualisations through their interpretation of them. The following sections focus on each of these gaps in the literature in turn, beginning by reviewing existing research into the design of data visualisation in relation to their narrative qualities.

2.3.2 Story telling in the design of data visualisations

This section discusses literature that explores the narrative and rhetorical elements of data visualisation. As Weber argues, "Stories help us to make sense of the world" (2020, p. 295). Furthermore, Dourish and Gómez Cruz (2018) highlight the importance of context and narrative in how people make sense of data. Therefore, it is important to consider how narrative and storytelling strategies in the design of data visualisations might shape how designers and users make meanings through graphs and charts. Indeed, the narrative elements of data visualisation emerged as a theme of research when reviewing the existing literature. Much of the literature covered in this section explores the narrative qualities of data visualisation through research on data journalism, which increasingly combines data and storytelling (Weber, 2020). Similar to research conducted around the conventions of data visualisation, it tends to focus analysis on the image of the graph or chart in order to unpick their narrative elements. This reveals two significant gaps in the literature: firstly an understanding of how different users make meanings from data visualisations, and secondly, research on the data practices that take place prior to the design of visualisations, including exploring the data and finding a story (Lee, Henry Riche, Isenberg, and Carpendale, 2015). Both of these are important moments where meaning is made and yet are currently underexplored in existing literature.

Commercial practitioner Brent Dykes (2016) published an article on the Forbes website in which he argued that data stories are a crucial strategy in communicating insights from data effectively. Dykes (2016) argues that data stories make data more memorable, persuasive and engaging and thus more likely to help drive change through compelling communication of insights. Dykes says, "People hear statistics, but they feel stories," (2016, n.p.). This links to the work of Kennedy and Hill (2018) who identified how feelings were significant in their participants' engagements with data visualisations. Computer scientists, Ma, Liao, Frazier, Hauser and Kostis helpfully unpick the components of *stories* which they define as "sequences of causally related events" (2012, p. 12). They identify the mechanisms that make stories particularly engaging, which include, pacing, settings, plots and characters and explain how visualisations can incorporate these same mechanisms to engage audiences (Ma, et al., 2012).

Both Dykes (2016) and Ma et al, (2012) are talking about visualising data more broadly than through graphs and charts. Nevertheless, they identify how storytelling with data has the potential to engage audiences. Indeed, data visualisations are increasingly used in news production to tell stories and communicate insights through data journalism (Weber, Engebretsen, and Kennedy, 2018).

There has been an increase in the use of data and data visualisations within the field of journalism, so much so that "data stories" has become a tagline in the industry (Weber, Engebretsen, and Kennedy, 2018, p. 191). Although the term data stories is not clearly defined among data journalists, Weber, Engebretsen and Kennedy explain that they tend to "weave together numbers, words, images and design" and that data visualisation has an important role in the design of data stories (2018, p. 192). Their research focuses on the perspectives of those who work in data journalism and involved sixty interviews over twenty-six news organisations. While the narrative elements of data visualisation design was their point of departure, it primarily identifies the main characteristics of data stories. The authors concluded that these characteristics included several "textual practices," beyond narration, which troubled simple definitions of narratives and stories in the context of journalistic practice (Weber, Engebretsen and Kennedy, 2018, p. 204). Notably, this included the potential for users to explore the data themselves, thus finding their own stories in the data. Although this recognises the potential for multiple meanings to be made from a single graph or chart, this was discussed from the perspective of data journalism practitioners. It did not explore *how users* do this kind of meaning making work when interpreting graphs and charts.

The authors published another article drawing on data from the same project, this time focusing on the data visualisation practices within ten Scandinavian news organisations (see Engebretsen, Kennedy and Weber, 2018). Drawing on interview data, their article highlighted factors that shaped visualisation practices in newsrooms, such as motivations for data visualisation, the recruitment of those with data visualisation skills, their working practices, and how much they know about their readers. It discussed these in relation to the development of the genre of journalistic data visualisation (Engebretsen, Kennedy and weber, 2018). Although the article claims to focus on visualisation practices, it did not explain the *design practices* that take place in the production of data visualisations. Nor how the designers' visualisation practices frame how meaning is made in their design. Significantly though, in the context of this thesis, Weber, Engebretsen, and Kennedy (2018) acknowledge the role data plays in shaping the story. They mention how data journalists have to search through the data for a coherent narrative and that "a question, a problem, or an idea" can influence that search (Weber, Engebretsen and Kennedy, 2018, p. 197). Despite this brief discussion, they do not interrogate *how* designers of

visualisations make meanings *from data* in order to find a story prior to its visualisation. In contrast, my own project aims to provide a detailed analysis of the meaning making practices at work in the design of data visualisations, including how designers determine the meaning of data.

Weber, Engebretsen, and Kennedy define design in the context of data stories, as, "all the visual elements that create the visual performance of the data story: *how* the data is presented" (2018, p. 199, emphasis in original). The following literature focuses on these visual elements in their discussion of narrative data visualisation. For example, Weber (2020) argues that the increasing use of data visualisation in journalism is because graphs and charts make data both visible and accessible. Journalists are not only employing standard forms of graphs and charts but are also making bespoke visualisations, designed to engage readers and to tell stories with data that are accessible (Weber, 2020). Weber (2020) argues that by exploring *narrativity* in data visualisations it is possible to gain a greater understanding of how meaning is made from data and how the visualisation process shapes these meanings. These are aspects of data visualisation design that my project is particularly interested in understanding. Weber defines narrative "as a textual, visual or multimodal representation that presents a story" (2020, p. 297), before outlining four main elements of narrativity; spatial and temporal dimensions; characters and events; sequentiality; and tellability (see p. 299). The chapter goes on to explore to what extent these elements of narrativity exist in journalistic data visualisations through the analysis of seventy-three data visualisations.

In her analysis, Weber (2020) explains how text, sound, and design techniques are used to enact the presence of a narrator in the visualisation, while the use of animation or the act of scrolling can provide sequentiality to the visualisation. In the chapter, Weber (2020) focuses on the image of the visualisations to identify the design techniques used to convey a narrative, yet how these techniques shape how ordinary users make meanings from data visualisations is neglected in her analysis. Nor does she address how designers make meanings from data or the practices they use to encode these meanings into graphs or charts. I consider these to be important and underexplored elements of the meaning making process, which this thesis aims to understand.

Gray explores the role of narration in how data is made meaningful for audiences through the concept of the "data epic" (2020, p.315). Gray's descriptions of two "'cinematic data visualizations'" (2020, p. 314) suggest some of the same techniques of *narrativity* that Weber (2020) identified in her research, namely, temporal dimensions, sequentiality and the presence of a narrator. However, Gray focuses his analysis on how the two data visualisations employed *distance* in order to narrate life and death (2020, p. 314). Gray (2020) notes that although

distance is associated with perpetuating the sense of objectivity that surrounds data and its visualisation, distance in an aesthetic sense can also be utilised to generate feelings among audiences. By distance, Gray (2020) is referring to the scale of the issues the data visualisations he analysed represented; war deaths during World War Two and the threat of nuclear weapons. Both of these topics are global issues involving large populations that cross countries and centuries (Gray, 2020). The visualisations used narrative techniques, which allowed them to tell stories at scale and close up, providing a sense of how war and the threat of nuclear weapons affects individuals as well as populations (Gray, 2020). However, Gray's focus on the narrative aesthetics of the data visualisations neglects the varied ways in which ordinary users might experience and make sense of them, particularly as these grand and complex visualisations incorporated various audio and visual elements, including animation, music, and photographs. Furthermore, although their chapter refers to quotes from the designers of these visualisations, this is in relation to various design choices and does not explore the meaning making practices that informed these decisions.

Dick (2015) looks more critically at the way in which visual representations of data can support particular narratives in his research on infographics that appeared in the Daily Express newspaper from 1956-59. Dick defines infographics as "visual representations of concepts, whether in the form of information or data" (2015, p. 152) but analysed only those that were "statistical in nature" (Dick, 2015, p. 157). After collecting a sample of 170 statistical infographics published between 1956 and 1959, Dick (2015) used content analysis and structural semiotic analysis to investigate bias in their representation of data. Dick (2015) identified five narratives of propaganda about various issues the newspaper took an interest in and wanted to influence public opinion about. He argues that each of the infographics could be attributed to one or more of these narratives before highlighting the design choices through which the infographics were made to be bias towards them. These included improper use of scale, axis, labelling and symbols (Dick, 2015).

Dick's research highlights the persuasive nature of data visualisations and the ways in which they can be deployed for political ends. This is very relevant to contemporary concerns about the integrity of news in an internet age. However, by basing his analysis on the infographics, Dick does not address how readers make meanings from these images and neglects the views and practices of the designers who make graphs and charts for news audiences. While critical of the practices of the Daily Express, Dick does not think critically about the agency of the newspaper's readers. Drawing on Hall's (1980) work on meaning making in the production and consumption of television news, it seems plausible that readers may have rejected the newspaper's *preferred reading* of the images to make alternative meanings from the

infographics. Therefore, a more detailed understanding of *how* users make meanings from data visualisation is important in relation to gauging the political power of graphs and charts.

Segel and Heer's (2010) influential article on narrative visualisation distinguishes contemporary forms of data driven story telling from the use of static graphs and charts. Images of simple charts have long accompanied text-based articles and acted as a form of evidence to back up the story in the text (Segel and Heer, 2010). However, the authors identify a new form of storytelling in which narrative elements are combined with interactive graphics to produce data stories or *narrative visualisations* (Segel and Heer, 2010). Segel and Heer's article researches the design of narrative visualisations through the analysis of graphs or charts that included 'narrative elements.' They set out to identify the design features that produce visualisations that tell a story with data and their analysis led them to define several narrative *genres* (Segel and Heer, 2010). However, it also revealed how using these genres required the designer to balance the intended narrative of the visualisation with the opportunity for users to explore it with a view to "self-discovery" (Segel and Heer, 2010, p. 1146). Thus, they argue that narrative visualisations sit on "a spectrum of author-driven and reader-driven approaches" (p 1146). While author-driven approaches tend to produce linear visualisations with little interactivity and a large amount of "messaging," reader-driven approaches produce visualisations that are predominantly interactive, with little or no messaging (Segel and Heer, 2010, p.1146).

Although the authors acknowledge that author-driven approaches lend themselves to storytelling, they explain how many narrative visualisations draw on both of these approaches in their design (Segel and Heer, 2010). To demonstrate this they discuss three models that occupy different positions on the author/reader driven spectrum. The most commonly used model among the visualisations they analysed was the martini glass structure, in which the narrative author-driven elements of the visualisation form the stem, which then opens up to provide the user an opportunity to interact with the data in a more reader-driven way (Segel and Heer, 2010). Segel and Heer's (2010) article provides an account of the design features that produce narrative visualisations, and how these produce different degrees of author/reader-driven elements. In this sense, they introduce ways of thinking about how the design of a data visualisation shapes user engagement and the different ways users are invited to interact with the data. Indeed, Hullman and Diakopoulos (2011) argue that Segel and Heer's article opens up the opportunity for research into how users interpret narrative visualisations. Segel and Heer (2010) acknowledge that their research does not take into account the user of the visualisation in terms of their emotional and cognitive engagement. Yet they also seem to assume that users are homogenous. Their article does not take into account social or cultural differences that exist among users and how these might shape their engagement. Indeed, all the literature discussed in

this section so far has focused on the story telling potential of data visualisations through a focus on their design rather than how users interpret them.

Drawing on Segel and Heer (2010) Hullman and Diakopoulos define narrative visualisations as employing both "explorative and communicative" elements (2011, p. 2231). However, instead of focusing on their design, they direct their attention towards how users interpret narrative visualisations using the concept of *visualization rhetoric* (Hullman and Diakopoulos, 2011). The authors use the term visualisation rhetoric to describe how the intended meanings represented in data visualisations are framed by the designer's choices but also by the characteristics of the end user and the social and cultural contexts within which they engage with the visualisation (see p. 2232). While Hullman and Diakopoulos' (2011) approach appears to take into account both the designer and the user in the meaning making process, the authors do not conduct empirical research with users. Instead, they identify various visualisation rhetoric techniques based on their analysis of a sample of narrative data visualisations. The authors even introduce the concept of "viewing codes" to discuss the cultural, cognitive and psychological factors that determine how users interpret visualisation rhetoric (Hullman and Diakopoulos, 2011, p. 2237). However, without conducting user research to test this, this feels somewhat arbitrary. Nevertheless, their article draws on the distinction between denotation and connotation to highlight the ways in which the perceived objectivity and transparency of narrative visualisations are in tension with their rhetoric. That is, the persuasive techniques that influence the connotated meanings that users interpret from the visualisation (Hullman and Diakopoulos, 2011). Yet, empirical research is required to understand *how* this tension unfolds and shapes meaning making in the interpretation of data visualisations by ordinary users.

Lee, Henry Riche, Isenberg, and Carpendale (2015) argue for a narrower definition of data stories that distinguishes them from data visualisation. In doing so they foreground the processes involved in transforming data into a visual story. Their discussion reveals important sites where meaning making happens, which are neglected in existing research on the design of data visualisations. While the literature discussed in this section has tended to focus on the design techniques used to *tell* the story, Lee et al's., article highlights the preceding steps which include exploring the data, and making a story (2015, see p. 86). As stated previously, I consider these to be important moments in the meaning making process, which require closer investigation. Particularly as Hall argues encoding meaning requires information becomes a "meaningful discourse," in order for it to be decoded by the intended user (1980, p. 130). Therefore, exploring the data to identify insights and transforming these insights into a story are likely to be important steps in the encoding process. Despite their interest in designers' data practices, Lee et al., do not explore the "visual data storytelling process" (2015, p. 86)

empirically, nor do they discuss it in relation to meaning making. Furthermore, Lee et al, (2015), along with Segel and Heer (2010) and Hullman and Diakopoulos (2011), do not *critically* explore the conventions of narrative data visualisations. Indeed, Hullman and Diakopoulos (2011) end their article by suggesting that an understanding of the rhetorical techniques that communicate meaning will enable analysts to communicate insights *more* successfully. This reflects the fields of study in which they are publishing, which are more aligned with human computer interaction studies than data studies.

The literature discussed in this section has predominantly explored the narrative techniques that allow designers to tell stories and communicate meanings through data visualisations. However, much of the existing research discussed here has done so through the analysis of the image of graphs and charts in relation to the rhetorical techniques deployed in their design. This neglects the perspectives of the designers of data visualisations and the audiences who interpret them. Where the literature did focus attention towards the makers of visualisations, it tended to concentrate on the narrative elements of visualisation design rather than the data practices that framed their production. However, Lee et al's., (2015) paper directs attention towards the process involved in telling stories with data, including the steps which precede the design of the final visualisation. This highlighted potential sites of meaning making during the data visualisation design process that are underexplored in the literature; namely exploring the data for insights and transforming these insights into a story (Lee et al., 2015). The thesis addresses this gap in understanding through the first research question, which asks how meanings are encoded into data visualisations. It will address this by shifting the focus beyond the image of the visualisation to explore the meaning making practices of designers during the design and production of data visualisations.

This section has also highlighted how users are neglected in research on the narrative techniques employed in data visualisation design, particularly in relation to how they interpret these rhetorical devices. Where users are discussed in the literature, this was not based on empirical research and thus it tended to treat users as a homogenous group. This disregards the potential importance of social and cultural differences among users that may impact how they make meanings from data visualisations. Although this is not the focus of my research, it is something my thesis is attentive to. The following section directs this chapter's attention towards research that has analysed how users engage with data visualisations.

2.3.3 User engagement with data visualisation

Data visualisations envisage the world in very particular ways (Beer, 2013). Beer suggests we ought to recognise them as "social actors," due to the way they influence how we perceive and make sense of the world (2013, p. 118). Therefore, it is important that research explores both the meaning making processes at work in the interpretation of data visualisations and how their very particular ways of viewing the world is performed through those interpretative and meaning making processes. This requires researchers to move the debate beyond the image as the site of analysis to explore how ordinary users interpret them. With little user research around data visualisations within the field of Sociology, this section draws heavily on four articles which originate from a project called *Seeing Data* (Kennedy et al., 2016b), whose work this thesis builds upon. The following literature directs attention towards the users of data visualisations in order to understand the varied ways in which people engage with and make sense of them. Reviewing this work identifies how issues around effectiveness and user engagement are prioritised in existing research, revealing a gap in the literature in relation to how users *make meanings* from data visualisations.

Research conducted in the field of Human Computer Interaction (HCI) studies has tended to focus on elements of human cognition and issues of effectiveness when exploring how users engage with data visualisations. For example, Borkin, Vo, Bylinskii, Isola, Sunkavalli, Oliva and Pfister (2013) researched the memorability of different types of data visualisations as a foundation for understanding what makes visualisations effective and engaging. They conducted a large-scale study, involving over four hundred individual visualisations as well as the employment of Amazon mechanical turk workers, in order to quantify how memorable different types of data visualisations were. Their findings showed consistency across the mechanical turk workers in terms of which visualisations were most memorable. The authors identified factors such as colour, as well as "the inclusion of a human recognizable object" increased memorability (Borkin et al., 2013, p. 2314). Borkin et al., (2013) accept that memorability does not equate to comprehension, yet they still associate memorability with *understanding*. However, their method did not explore if participants *understood* the chart. Nor did it explore differences in responses between their participants, which may have provided a more nuanced understanding of how memorability differs between users and how his affects their engagement with visualisations.

Huang, Eades and Hong argue that evaluating the effectiveness of graphs by measuring task based performance does not take into account the "cognitive load" required to perform such tasks (2009, p. 139). By cognitive load, the authors mean how much effort is required by the

user to accurately interpret the graph (Huang, Eades and Hong, 2009). They explain how some charts may require little effort on behalf of the user while others may require much more interpretive work, yet the speed or accuracy with which they interpret them may remain the same for both. Therefore, Huang, Eades and Hong (2009) introduce a visualisation efficiency measure with which to assess the cognitive effort required to interpret a visualisation and therefore, its effectiveness. The participants they used to test their measure were a small sample of thirty computer literate IT students who were familiar with complex visualisations. While this was a user study, diversity among participants was ignored and the focus remained on how elements of the visualisation and performance task impacts engagement and effectiveness. Similarly, Anderson, Potter, Matzen, Shepherd, Preston and Silva (2011) also focus on the relationship between cognitive load and the effectiveness of data visualisation. However, in their study they measure cognitive load using electroencephalography (EEG) to monitor their participants' brain activity while they engage with data visualisations. Apart from stating their user research included seventeen participants, ten male and seven female, Anderson et al., (2011) provide no further demographic information about them. Indeed, all of the HCI user research discussed above does not take into account the social and cultural differences of users, nor does it critically engage with data visualisation as a medium. However, more recent studies in the social sciences have been attentive to how social and cultural contexts affect non-expert users' engagement with visualisations.

Recognising the limitations of HCI research, Kennedy, Hill, Allen and Kirk argue that "contextual, social and cultural factors matter" when understanding user engagement with data visualisation (2016b, n.p.). Their project, *Seeing Data*, included focus groups and interviews with members of the public to understand the factors that affect how people engage with data visualisations. Their analysis identified six factors influencing user engagement, including subject matter, personal beliefs and opinions, and emotions (Kennedy et al., 2016b, n.p). By drawing on audience research, in particular Stuart Hall's work on *encoding/decoding*, the authors were attentive to the socio-cultural contexts of the their participants and how this shaped their engagement with the graphs and charts they encountered. Kennedy et al., note how audience research pays attention to factors beyond the "textual and technical" of cultural artefacts to consider factors like "class, gender, race, age, location […] and education of audience members" (2016b, n.p.). By being attentive to diversity among users, the research looked beyond the "visualisation text," and in doing so developed a broader understanding of user engagement in relation to data visualisation (Kennedy et al., 2016b, n.p). My own project aims to build on this work by being attentive to how different users *make meanings* from data visualisations.

Kennedy et al, (2016b) highlight how the factors of engagement they identified centered on the decoding process, however this was in relation to engagement rather than meaning making. Nevertheless, there are points from their findings that are pertinent to research which looks at meaning making among users of data visualisations. Although the factors they identified related to the interpretation of the visualisation text, Kennedy et al., identified how they were "outside the control of [the] visualisers" who designed them (2016b, n.p.). This recognises the user's agency in their response to data visualisations, even if this is mediated by the socio-cultural context of their circulation and the rhetoric of the visualisation (Hullman and Diakopoulos, 2011). Kennedy et al, (2016b) describe their research as a grounded and bottom-up approach. This proves fruitful in pushing at the boundaries of definitions of user engagement by developing new ways of thinking about it in the context of data visualisations, ways that were attentive to the social and cultural factors that affect user engagement. Just as Kennedy et al, (2016b) broadened the definition of user engagement in the field of data visualisation research, this thesis aims to open up different ways of thinking about meaning making beyond those that can be achieved through analysis of the visualisation text.

Shah and Hoeffner (2002) undertook a review of literature that addressed users' comprehension of graphs, paying particular attention to the uses of graphs in educational settings. From this, they note how a user's ability to assign meaning to different visual features of a graph or chart might be dependent on their level of *experience* (Shah and Hoeffner, 2002). They identify three factors, which are significant in how people interpret visualised data. These are the visual characteristics of the graph, the users' own knowledge of graphs and what the chart represents "(e.g., age versus height)" (p. 51). The first two factors are interdependent, in that the semantic decisions visualisers make rely on there being a shared understanding between the user and the designer in how to interpret these (Shah and Hoeffner, 2002). As Hullman and Diakopoulos (2011) argue, designers have to assume there will be some shared knowledge between them and the user in order to be able to communicate information from data through visualisation. However, the third factor implies that for some users, their prior knowledge about the subject of the graph influences how they interpret it (Shah and Hoeffner, 2002).

Shah and Hoeffner (2002) explain that when users perceive that there should be a correlation between two variables they will over emphasise this relatedness in their interpretation of the data. Users' expectations about the relationship between data not only influence their "quantitative estimates, but also their qualitative interpretations" by means of inferences about broader trends (Shah and Hoeffner, 2002, p. 61). Shah and Hoeffner (2002) also suggest that non-experts are more likely to rely on their prior knowledge of the subject the visualisation represents when interpreting graphs. The authors' article implies that how non-expert users

make meanings from data visualisations is influenced by factors beyond the semiotic resources in the text. This supports the need for research to move beyond the image of the visualisation to take into account how individual users make sense of data visualisations. This adds weight to the importance of my third research question, which asks, *how* ordinary users make meanings from data visualisations.

Masson and van Es (2020) conducted research into how users interpret weather based data visualisations, specifically rainfall prediction visualisations from one particular website and app. Although the study was small in scale, comprising sixteen users of the service, their interviews revealed how users interpreted the visualisations differently based on somewhat intuitive, "individualized reading strategies" (Masson and van Es, 2020, p.87). They explain how, for example, despite there being a clear key, the interviewees interpreted the colours on the map visualisation in various ways. Some users adopted simplified readings of the graphs, for example, interpreting a chart that represented a measure of rainfall in terms of if it was due to rain or not. Participants also varied in terms of their depth of engagement with the chart (Mason and van Es, 2020). Despite these differences, their participants tended to consult these graphs as part of their daily routines, often making decisions about day-to-day activities based on their predictions. However, Masson and van Es argue that they did so despite their skepticism of them, which the authors describe as suspicion of the information based on the user's lay knowledge of the unpredictability of weather (2020, see p. 89). Their project reveals various factors that influence how their participants engage with the data visualisations, including the everyday contexts in which they were viewing them, in order to make decisions. It also demonstrates how different users adopted different reading strategies and drew on their own lay knowledge to add context to their interpretations of the graphs and maps (Mason and van Es, 2020). These factors might also be influential in how users make meanings from other types of data visualisations and highlight the importance of acknowledging differences and nuances in the way users interpret graphs and charts.

Hill, Kennedy and Gerrard (2016) analysed their participants' focus group responses to a specific chart to explore how socio-cultural assumptions influence how people engage with data visualisations. The visualisation they focused on was published in a free tabloid newspaper and was about the number of social media followers the popstar Shakira had amassed (Hill, Kennedy and Gerrard, 2016). It incorporated large images of Shakira and another singer, Rihanna, and the predominant colours used in the visualisation were pink and orange. The visualisation was notably unpopular with focus group participants, evoking a range of negative responses (Hill, Kennedy and Gerrard, 2016). Although the chart was criticized by participants, "for not meeting the standards of a "good" data visualisation," these criticisms were intertwined

with socio-cultural discourses around gender, age, and class (Hill, Kennedy and Gerrard, 2016, see p. 346). Hill, Kennedy and Gerrard's (2016) paper reveals how the subject matter of a data visualisation is important in shaping people's judgements about it. However, it also shows how sexist, ageist, and classist assumptions can frame user interpretations. This suggests that cultural discourses and personal political beliefs (Hill, Kennedy and Gerrard, 2016) may be significant in how users make meanings from data visualisations. While the politics inherent within data visualisations have been recognized (Boehnert, 2016), less attention has been paid to how politics permeate user interpretations of graphs and charts. In this instance, the participants were using socio-cultural discourses to make a judgement about the value of the visualisation. However, this further demonstrates how factors beyond the image of the visualisation shape how users make meanings from graphs and charts, in this instance, the socio-cultural discourses users draw on to interpret them.

A further article published from the *Seeing Data* project explores participants' emotional engagements with data visualisations. Kennedy and Hill (2018) describe the emotional responses their non-expert user participants experienced when they engaged with data visualisations during focus groups. The ways in which visualisations prompted emotional responses were varied with Kennedy and Hill (2018) reporting emotional responses to both the data and the design of the visualisation. This suggests that both the subject of the data and the visual elements of the visualisation were influential (Kennedy and Hill, 2018). However, factors beyond the visualisation also prompted emotional responses, such as the "source or original location" of the visualisation and the skills the participants possessed to interpret it (Kennedy and Hill, 2018, p. 836). Kennedy and Hill (2018) identified both positive and negative feelings among their participants, which ranged from frustration to excitement, sadness to love. Their article highlights how subjective elements play an important role in peoples' engagement with data visualisations. Indeed, Kennedy and Hill (2018) argue that how people make sense of data visualisations is not limited to reason and rational knowledge. In fact Kennedy and Hill argue that it is through the visualisation of data that it becomes "possible to *feel* numbers" (2018, p. 844, emphasis in original). This provides an important contribution to data studies, further challenging the "renewed dominance of rationality, objectivity" and trust in numbers (Kennedy and Hill, 2018, p. 844).

This section of the literature review has demonstrated the limitations of user research conducted in HCI studies that focuses on improving the effectiveness of data visualisation in relation to reducing the cognitive function required to interpret them. Such research has tended to focus on factors such as memorability and cognitive load, neglecting to consider how social and cultural differences among users might shape user interactions with data visualisations. It has also

introduced literature that has attempted to address the shortcomings of this HCI research by being much more attentive to the social and cultural factors that influence how users engage with data visualisations. The articles from the *Seeing Data* project were particularly useful in demonstrating how taking notice of social and cultural factors can expand understandings of user engagement with data visualisations (Kennedy et al., 2016b; Hill, Kennedy and Gerrard, 2016; Kennedy and Hill 2018). Despite providing insights that point towards factors that might influence how users make meanings from graphs and charts, none of the literature covered here sought to understand *how* this happens or questioned what meaning making *means* in relation to data visualisation. My third research question addresses these gaps in the literature, by asking how ordinary users make meanings from data visualisations.

2.4 Conclusion

This chapter began by setting out the critical context of this thesis. It demonstrated how forging an understanding of how meaning is made in the design and interpretation of data visualisations is important in relation to existing debates in data studies, particularly those around data subjectivities. The second part reviewed existing research on data visualisation design and engagement through three distinct but interrelated themes. This revealed a tendency within research to treat the production and consumption of data visualisations as two separate issues, with little research that addresses both and none that did so empirically. This thesis addresses this gap by tracking meaning making in both the design and interpretation of data visualisations, albeit weighted towards the former. This chapter also identified the need for research that looks beyond the image of graphs and charts as the site of analysis. This is in order to understand the design practices of designers and the interpretive strategies of users when making meaning from data through visualisation. In analysing the literature this chapter situated the following research questions among the existing research:

1. How do designers encode meaning into data visualisations?
2. How do ordinary users do meaning making when engaging with data visualisations?
3. What is the relationship between meaning making and data assemblages?

Reviewing existing research into data visualisations identified gaps in understanding that informed these research questions. How data visualisations communicate information and meaning is predominantly explored through analysis of the image of graphs and charts, even when the research is in relation to designer practices. Focusing on the image as the site of analysis neglects how meaning is made before and after the final visualisation is produced. The research questions of this thesis are attentive to these moments as important sites of meaning

making in the production and consumption of data visualisations. This includes how the socio-technical contexts in which data is embedded frame the knowledge produced, as well as the data and the design practices that shape how meaning is made from data and encoded into visualisations. Although user research has been conducted in relation to data visualisations, HCI literature often focuses on effectiveness in relation to knowledge transfer, while the limited research conducted from a more Sociological perspective has focused on engagement. This has resulted in a gap in understanding in relation to how users *make meanings* from data visualisations when interpreting them. These gaps in the literature have informed the research questions this thesis will go on to address. The following chapter will explain the methodology used to answer them.

Chapter 3: Methodology

3. 1 Introduction:

It is a dark, chilly, spring evening and I am standing outside on a floodlit astro-turf pitch in the North East of England. I look back towards the side of the pitch where twenty to thirty men and women have formed a line to face me. I have only met one of them before, my friend who cleared it for me to come along tonight. They are all wearing sports clothing, holding hockey sticks, impatient to start their evening's training. They listen as I introduce myself and explain that I just want to take a couple of minutes of their time to tell them about some research I am doing. They blankly stare back at me. 'They don't want me imposing on their evening like this,' I think, as I feel my heart beating faster. As I talk about my research, my voice dissipates into the night air, so I speak unnaturally loudly to make sure that I can be heard. In this moment, I feel so self-conscious, yet I try to appear confident, as the sound of my own heartbeat reverberates in my ears. One man looks at me with a very confused expression on his face. 'I expect he is wondering what any of this has to do with him!' I think. As I look back at the confused man I feel the urge to laugh out-loud, at how surreal this situation is, but I stop myself. I continue speaking, explaining that I am looking for people to take part in my research and briefly say what this will involve. I had imagined people would nod enthusiastically at this point, keen to sign up. That was the plan, anyway! Yet, their expressions do not change. I invite anyone who might be interested in taking part in my research to take an information sheet from me and then I thank everyone for their time. As I walk back to the sideline, they jog past me and onto the pitch.

Before beginning my PhD, I had never conducted research with more than one human participant, having previously chosen to draw on literature, or even art, to investigate the social world. Therefore, my PhD has been an apprenticeship in designing and conducting empirical social research that explores how people make sense of their worlds and specifically their interactions with data through visualisation. This has placed me in situations so out of my comfort zone that I could never have imagined them before beginning this traineeship; the anecdote above is just one example. The methods textbooks do not prepare you for the messy and complex nature of research interactions with people, which are always unpredictable. Furthermore, the complicated realities of research are concealed in empirical articles where a "coherent narrative" of the research is presented (Nairn, Munro and Smith, 2005, p. 236; also see Stanley and Wise, 1983). I found any attempt to emulate the robotic and linear approaches to methods, which some textbooks seem to suggest is possible, to be futile. Instead, I have learnt

the value of being human when conducting research. Being kind, thoughtful and accepting, both towards my participants and to myself, has been invaluable in conducting good and ethical research. As I use this chapter to set out the methodology of this project I am also reflecting on all that I have learnt during this experience.

This chapter will set out the project's methodology, whilst critically reflecting on the methods used and how these, as well as other decisions I took, shaped the research and findings. It begins by discussing my approach to researching meaning making in the production and consumption of data visualisations. It then goes on to discuss the research design in two parts. Firstly, a *short-term ethnography* (Pink and Morgan, 2013) at a data visualisation design agency, which addresses research question one and three. Secondly, interviews with ordinary users, which addresses research question two. The chapter ends with a discussion of the ethical considerations of both stages of the research design.

3.2 Researching meaning making in the production and consumption of data visualisations

In order to explore meaning making in the production and consumption of data visualisations empirically, it is important to first interrogate my own theoretical position with regard to what meaning making is and how it occurs. This critical analytical thinking is necessary to inform the methodological approach to the research. As set out in Chapter one, I consider meaning to be discursive and formed through language, be that text, conversation, or images (du Gay, Hall, Janes, Macay, Negus, 1997). Meaning is also relational and dependent on social and cultural contexts (du Gay et al., 1997). As such, the thesis draws on Stuart Hall's (1980) *Encoding/decoding* model as its theoretical framing, in order to analyse meaning making in both the production and consumption of data visualisations. Therefore, to understand how meaning making unfolds in the design of data visualisations, it is important to observe the practices of designers at work, and to obtain their own accounts of their work by talking to them. Furthermore, to explore how ordinary users decode graphs and charts, it is necessary to obtain data pertaining to the lived experiences of the participants as they make meanings from graphs and charts.

Adopting this methodological approach has enabled the research to produce "situated knowledge" that is reflexive about the context in which it was generated (Mason, 2002, p.65). Mason argues that qualitative methods, "constitute compelling arguments about *how things work in particular contexts"* (2002, p1, emphasis authors own). Indeed, observations and semi-structured interviews produced data that was analysed to explain the meaning making processes

in the design and interpretation of data visualisations. Furthermore, qualitative methods, such as these, have the capacity to produce data that is rich, nuanced, contextual and that reveals the complexity of the social world and how people interpret and experience it (Mason, 2002, see p. 1). Qualitative methods are appropriate in relation to the aims of this thesis, which does not seek to make universal generalisations about meaning making but to provide an explanation of how meaning making works through and from data visualisation in particular contexts.

To research the meaning making processes at work in *both* the production and consumption of data visualisations, the project was designed to include two stages of data collection. The first stage involved a period of fieldwork at a data visualisation design studio, where I conducted a short-term ethnography involving observations and qualitative interviews with designers. The second stage involved conducting qualitative interviews with members of the public, who I refer to as 'ordinary users' (see Chapter one for a definition). These interviews incorporated visual resources in the form of seven examples of data visualisations. The following sections of this chapter will draw on methodological literature to explore these different methods, justifying their selection in terms of both theoretical and practical considerations. In doing so it will reflect critically on how the research design shaped the project including the data collected and the findings. It will begin by discussing the methods used in the two distinct research stages, as set out above.

3.2.1 Research design stage one: Fieldwork at a data driven design studio

My PhD project formed part of an interdisciplinary research network interested in how people relate to data through its visualisation. The ESRC and White Rose doctoral training centre funded the network. It included two other PhD students, our academic supervisors and two commercial partners, both of whom are considered experts in the field of data visualisation design. Prior to me commencing the project, the network had already negotiated fieldwork access to a renowned data visualisation design agency, which is referred to throughout the thesis using the pseudonym 'Evoke.' It can be difficult for researchers to gain access to people who are experts in their field (Brinkman and Kvale, 2015). Yet, the network had secured a valuable opportunity for me to be embedded in a design studio and observe the practices of designers as they produced data visualisations. Evoke is based outside of the UK in a city in Western Europe; however with both an international workforce and clients from around the world, the language the company used was English. Having access to conduct fieldwork at Evoke allowed me to take an ethnographic approach to the research in which I conducted observations in the design studio as well as semi-structured interviews with the designers. This phase of the

research was designed with the aim of generating situated and multi-dimensional knowledge about the designers' practices, observed and recorded as they were happening (Mason, 2002).

By conducting observations at the design studio, I set out to:

- Unpick how designers encoded meanings into the design of data visualisations.
- Identify individual apparatuses of the data assemblage.
- Identify the socio-technical processes within which these data and design practices were embedded.

Immersing myself in reflexive, fieldwork based qualitative research allowed for research that positioned the theoretical framing of *encoding/decoding* (Hall, 1980) into context, in the domain of data visualisation design. It did so by generating data that could be analysed to explain *how* meaning is made from data and encoded into the design of graphs and charts. This required contextual data that was rich in depth and nuance, something observation methods are particularly good at generating (Mason, 2002). Observing designers' working practices in their everyday contexts generated data that might not be revealed through artificial encounters manufactured by the researcher, such as interviews. This is because not everything can be articulated in an interview (Mason, 2002) particularly aspects of design practices that the participants take for granted or perform habitually and with little reflection. The data generated from observational methods was analysed for the meanings of the designers' social actions in relation to how they made meaning from data through visualisation. This made it possible to unpick the meaning making processes at work in the design studio.

The observational data was complemented by qualitative, semi-structured interviews with six designers, which allowed an opportunity for the designers to provide their own accounts of their design practices and allowed me to ask them questions about what I had observed. By using a semi-structured approach, the one-to-one interviews were designed to generate discussion around:

- How designers encoded meaning into the design of data visualisations.
- How their design tools (specifically software) shaped their work.

- How they imagined the users of their data visualisations and how this shaped their design practices.

Qualitative interviews are a useful method for generating data relating to how people experience and make sense of their world (Brinkman and Kvale, 2015). In the context of this research, interviews allowed the designers to give their own point of view (Brinkman and Kvale, 2015) on their design practices. This is important in providing situated accounts of how meaning making happens in the design studio. However, the knowledge generated from interviews is a "co-construction of knowledge" between the researcher and participants, formed through the research interview interaction (Brinkman and Kvale, 2015, p.22). Furthermore, "discourses, power relations and ideologies" (Brinkman and Kvale, 2015, p. 3) that exist in *their* world of data-driven design shape the *reconstructed accounts* (Mason, 2002) that designers provide in interviews. Therefore, the knowledge generated from qualitative interviews is both deeply contextual and situated (Mason, 2002). This is compatible with the thesis' aims to explain how meaning is made in the specific context of one data visualisation design studio.

The following subsection provides a reflexive account of what happened in the field.

3.2.1 In the field

It was a mild morning on the 16th October 2017 as I walked to the data driven design agency where I would be spending three weeks conducting fieldwork. Their premises were a short walk from my apartment and, as I set out that morning, I felt apprehensive, yet also excited to get there. On arrival I was welcomed by the office manager who showed me into the building. They explained where to leave my coat - hung up on a high clothes rail on the ground floor - before showing me to the studio space on the first floor. I followed the office manager up a wide black industrial style staircase. On reaching the top, the stairs opened out onto a large landing, which was home to an open plan studio space. The studio was a long rectangular room, light and airy, with six banks of spacious white desks at either side of the staircase. This was where the designers sat or stood to work. There were also two smaller rooms off this main space (a designated 'quiet' room and the print room) and a narrow staircase to a second floor office. The studio reminded me of a trendy industrial loft apartment. It had exposed wooden beams, industrial steel pipes, steel ventilation fittings, and strip lighting hanging from the ceiling. The walls and structural woodwork were painted in a bright chalky white, while the visible steel girders supporting the roof structure were painted black. There were also large black planters situated around the staircase, home to lush green plants whose finger-like green leaves leaned into the room. The green of the many plants dotted around the studio popped against the stark

black and white colour pallet. There were only a few windows in the studio, which had been frosted so that it was impossible to see in or out. However, natural light flooded into the room from the roof windows above.

While conducting fieldwork in the studio, I planned to observe designers in order to track their data and visualisation design practices. I arrived with a large rucksack carrying information sheets, consent forms, a dictaphone, sketch book, laptop, spiral bound note pads, and a safety deposit box in which to securely store confidential material (including consent forms) in a locked cupboard at the premises. I had also brought with me an interview schedule and an observation guide to help focus my field notes. The printed observation guide was brief and prompted me to focus on the following topics: *(Socio-) technical and production processes; intended meaning; encoding*; and the *apparatuses of the data assemblage* (see appendix a). Mason (2002) argues that when conducting observations it is important to enter the field knowing what you want to find out. This is particularly the case in short-term ethnographies where a more purposeful approach to data collection is required (Pink and Morgan, 2013). The observation guide was based on themes I had identified for exploration and was informed by the research questions and aims of the project (Mason, 2002). Therefore, creating an observation guide helped to provide a strategic focus to the observations that was theoretically informed. Although the observation guide was a useful tool to keep the research grounded in the aims of the project, it did not prevent the flexibility to follow new and unexpected leads in understanding how meaning was being constructed in the design studio. After all, taking an ethnographic approach to research requires an "orientation [that] is exploratory" (Hammersley and Atkinson, 2019, p. 4). Nevertheless, the research aims of the thesis undoubtedly framed the data I recorded in my field notes (Mason, 2002).

On my first day in the studio I was shown to a desk near the entrance to the staircase. The office manager explained that the member of staff who usually sat here was away for the week, so I could make use of their desk. The location was towards the middle of the studio and from there I was able to observe how people worked at their desks, albeit at a distance, and how they moved around the space. The open plan studio allowed me to produce general observations through which I made field notes about the conversations and social interactions taking place between the designers as they moved around the studio to sit or stand at one another's desks. Before commencing the fieldwork, I had envisaged that I would be free to shadow designers as they worked. However, on arrival I was informed that because the studio was very busy, project managers would need to schedule my observations with designers around the designers' workloads. On reflection, I was being managed and my time diarised by a project manager, in a similar way to how they managed the designers' time through diary systems. This reveals

something about the commercial nature of the research setting, in which time spent on design was carefully managed and quantified by project managers, who oversaw the projects to completion.

Although I felt disappointed and constrained by having my observations with designers *scheduled*, this approach added some structure to my observations and field notes. Therefore, as well as general observations in the studio I also conducted scheduled observations, where I would sit with a designer for up to an hour, observing them working and asking them questions. These allowed me to quickly position myself "at the center of the action," which is necessary when undertaking short-term ethnographies (Pink and Morgan, 2013, p. 355). However, as a project manager scheduled the latter rather than myself, on several occasions the participant I was observing was not working on a data visualisation project. In these instances, a number of things happened; sometimes participants stopped what they were doing to show me previous data visualisation projects they had worked on, explaining what they did and answering my questions about them. Other times, participants would suggest that I come back at another time when they would be working on a data visualisation. On two occasions, the latter resulted in participants inviting me to observe meetings between designers about an ongoing dashboard project. These generated interesting data about how designers with different expertise collaborated in the design of visualisations. Nevertheless, I recorded moments of deep frustration in my fieldwork diary at having to rely on a project manager to schedule observations, which I felt held up the research.

Each designer had their own white desk with a metal stand on which to perch their slim apple MacBook. The stand allowed the laptop screen to sit at roughly the same height as the very large Dell computer screen that adorned every desk. Once perched on the stands, the MacBook was connected to the computer screen, a slim narrow keyboard, and a mouse. In this clean and modern studio, I was struck by the amount of cables required to make this system work. I regularly counted at least five cables emerging from the designers' laptops, snaking their way across the back of the desks. On several of the desks an oversized pair of headphones was slung over the computer monitor, while most also had a small potted plant. The different scales of the items on the desks was perplexing. The slim laptops, large screens, oversized headphones, small keyboards and little plants all implied a prioritisation of the visual and audio in the studio. Furthermore, in this clean digital space, there still appeared to be a desire for analogue technology. The chalky white walls were punctuated by neon bright post it notes and pieces of paper suspended on the wall with masking tape. While a large bookcase at the far end of the room was home to rows of brightly coloured books - many of them art or design related – that had been carefully organised into categories.

During scheduled observations, I would take my notepad and pen and pull up a chair next to the designer I was observing. At the beginning of the observation, I explained that I wanted to watch them working and ask them a few questions about what they were doing. This usually prompted the designer to explain to me what they were working on and, with my notepad resting on my knees, I would begin to scribble down notes as I watched them flick between different views on their computer screens. When I asked a question about what they were doing, I tried to record what they said as accurately as possible. I found that during the observations, and sometimes in answering my questions, the designers would pause their work to show me other things. This included other data visualisation designs, previous projects they had worked on, or demonstrating different elements of the software they were using. On one occasion, a participant stopped what they were working on for the duration of the observation so that they could show me examples of data visualisation projects they had previously worked on. The designers knew that I was interested in researching how they designed data visualisations. This makes what the participants *chose* to show me during interviews interesting data in itself and demonstrates how the knowledge generated from these scheduled observations was co-produced during our interactions. It was not only shaped by the particular questions I asked participants during the observation, but also by what the designers felt was important to show me to communicate *their world* of data driven design.

Without digital design or development skills, I was unable to take a participatory role in the observations. In this sense, I never felt part of the setting but was aware that I was participating in it. Therefore, I used my difference to my advantage by being open to my participants about my lack of knowledge of different design and coding software. This made it easier to ask people to show me things and to explain to me what they were doing, and legitimised them helping me to understand their practices. In this sense, I adopted what Lofland and Lofland term an, "Acceptable incompetence" (1995, p.56) role as researcher within the field. Keen to learn and to be taught, I encouraged participants to show me their world within the design studio. They were the experts and I became the apprentice, allowing them to take the lead in showing me their practices (Lofland and Lofland, 1995). Through this, what they *chose* to show me during scheduled observations and follow up interviews added nuance to the data.

It is possible to relate this approach to Reyes' concept of an "ethnographic toolkit" (2020, p. 225). Reyes (2020) explains that researchers in the field draw on both visible and invisible identity traits in order to gain access to, and understand, the research setting. While my position as a young, white woman was visible, my social capital was less so. Nevertheless, Reyes suggests that researchers can draw on their social capital as a resource, or tool, to alleviate difference and power imbalances in the field. While I may have drawn on my own social capital

as a middle class, educated person to alleviate differences between the participants and myself, I also used my position as a researcher and 'outsider' to bring differences to the fore. This was in terms of my lack of digital design skills, which enabled me to adopt a role in which I could be taught about the designers' practices. While Reyes (2020) notes that strategically drawing on their personal ethnographic toolkit affects how the researcher is viewed by those in the research setting, it also frames the researcher's view of that setting. This suggests that my positioning of the designers as experts and myself as a student will have framed my view of their practices while in the field.

The relationship between student and teacher is one that people recognise and is a strategy that, in the appropriate research setting, can aid "the flow of information" (Lofland and Lofland, 1995, p. 56). Although Reyes' (2020) work suggests that utilising various aspects of our identities can be strategic, at first I did not realise that I had drawn on this social dynamic to *get on* in the field (Hammersley and Atkinson, 2019). Indeed, Coffey (1999) argues that the notion that ethnographers can simply pick and choose their identity in order to negotiate fieldwork is too simplistic. She argues that to do so implies a belief that a separation between the "'self' and identity" is possible (Coffey, 1999, p 4). Coffey (1999) criticizes Lofland and Lofland and Hammersely and Atkinson for limiting their discussion of self and identity to the conscious performance of a persona in order to achieve successful research. This, she argues, neglects the complex identity and bodywork required in the field. Furthermore, advocating the role of "ignorant outsider" fails to recognise the ways in which ethnographers engage in "complex and delicate processes of investigation, exploration and negotiation" (Coffey, 1999, p. 22). Instead, Coffey recognises how identity work happens in relation to others in the research setting, meaning our identities shift and change through the experience of fieldwork "in ways that go beyond this enactment of a work process" (1999, p 26).

My adoption of a student role during fieldwork certainly did not feel conscious initially, although as the fieldwork progressed I became aware that this dynamic worked and it became more of a choice. Therefore, in adopting this role I was also responding to the social relations I was experiencing which were shaped by my own social position as a young woman in an environment predominantly formed of male designers. Therefore, it was perhaps less of a choice than Lofland and Lofland (1995) and Hammersely and Atkinson's (2019) work might suggest. Nevertheless, this role shaped my interactions with designers, whom I perceived to be experts, and shaped the data as I encouraged them to teach me about their design practices.

Although I often let the participants take the lead in showing me their world, I was not passive in our interactions. I participated through the questions I asked them and what I recorded in my

field notes. Therefore, the data produced through the fieldwork methods was co-constructed between the participants and myself. Indeed, social researchers produce knowledge about the social world through their participation in it. Producing an explanation of social phenomenon through ethnographic methods involves writing field notes on "selective observations" and transcribing interviews in which the researcher has asked "particular questions" (Hamersley and Atkinson, 2019, p. 17). That the researcher co-produces the knowledge generated in research interactions and affects participants in those interactions does not undermine the accuracy of their accounts of the social world (Hammersley and Atkinson, 2019). However, it does require researchers to be reflexive about how they shaped the production of knowledge (Hammersely and Atkinson, 2019).

During observations, I wrote field notes in spiral bound reporters note pads. These comprised a mixture of words, phrases and sentences to aid recollection when I came to write up my field notes in the evening. These notes were about how the designers were making meaning from data and coding it into their data visualisations through their design practices. I also carried a small sketchbook with me so that I could make simple drawings of the space and some of the things in it that would help me to explain the setting. Every evening when I returned to my temporary home - a top floor studio apartment in a central area of the city - I would prepare a meal before writing up my field notes on a laptop. I sat at a small dining table with my notepad next to me and read my field notes, using them to recall the situations and conversations that took place that day. A word or a short sentence would be enough to spark vivid memories that allowed me to form detailed field notes, which often took a couple of hours to type up. As Coffey argues, "ethnography is about experiencing and remembering" and in writing up the field notes I was adding structure to my memories (1999, p. 127). However, I was not only writing up my own memories but those of my participants (Coffey, 1999) who shared their practices and experiences with me through observations and interviews. Therefore, my field notes can only ever provide a partial account, reconstructed from biography and memories, limited to what it is possible to articulate through text (Coffey, 1999). Incidentally, my apartment was close to the university in the city. There was a bar on the other side of the road and, at night, as I typed up my field notes alone, I could hear students outside socialising and having fun.

Whilst writing up my descriptions, I sometimes included methodological notes to add important contextual and reflexive details about my role or impact as a researcher within the particular situation I was recounting. Other times I would make theoretical notes, comprising analytical ideas or questions that came to mind as I wrote about what I had encountered that day. I intended to treat my field notes as "raw data," which over the days and weeks would become a

rich dataset to be analysed later (Mason, 2002, p.99). Used in this way, field notes are intended to be "concretely descriptive," (Hammersley and Atkinson, 2019, p. 4). As such, I was careful to make sure my field notes did not include any deliberate analysis of either the setting or what I observed. Particularly as a researcher's job is not to pass "moral judgement nor immediate reform, but understanding" (Lofland and Lofland, 1995, p. 44). Of course, it is impossible to write field notes without some level of analysis; what the researcher already knows and thinks affects how they describe the setting (see Lofland and Lofland, 1995, p.68). Therefore, my unconscious thoughts and ideas framed my understandings of what I observed in the field. Nevertheless, when I wrote my own reflections into the field notes I was explicit about this, distinguishing methodological and analytical notes from the rest of the text by using the acronyms 'MN' (methodological notes) and 'TN' (theoretical notes). This made it clear during analysis where observations had shifted into deliberate analytical thinking. I also kept a fieldwork diary, which was a reflexive record of conducting research in the field, including my thoughts, feelings and reflections on how my position in the studio shaped the fieldwork.

Unfortunately, after a week or so, the desk I had been using in the studio was no longer available and for a few days I sat in the small meeting room situated off the main studio space. The notice on the door indicated that this was a quiet space, yet it seemed to be used for meetings or as an overflow when additional desk space was required in the studio. When accessing the room I stepped down into it and, although I left the door open, I felt as though I might as well have been in another building. Although I was still conducting scheduled observations and interviews, from this location it became harder to conduct general observations of the studio. This was particularly frustrating as I had already identified the importance of collaboration between the designers in the studio and the general observations I was conducting had allowed me to collect data about how the studio space facilitated this. While in the quiet meeting room, I began to feel frustrated and isolated as feelings of self-doubt increasingly crept into my mind and the days began to drag. I decided to speak to the office manager about the availability of other desks in the studio and they kindly let me know on what days certain designers would be out of the office so that I could borrow their desks. On days the studio was very busy, I utilised a table situated at one end of the room. Although it was not a desk as such, it provided me with a base in the studio from which to observe but also to be visible so that I remained part of the setting.

Although people in the studio were generally friendly and welcoming towards me, there were often moments during the fieldwork where I felt like an outsider. Unsurprisingly, these feelings were more pronounced at the beginning of the research and were made worse by some of the initial reactions of the designers to my presence. I approached one of the designers to provide

them with an information sheet and consent form. On seeing me approach their desk they turned off their screen so that I could not see what was on it. On my second day at the studio, another jovially remarked, '*we're still jumpy around you but we'll get used to you in a few weeks*.' Their comment reveals a limitation of short-term ethnographies, which may not last long enough for researchers to build a rapport with participants. Although their initial unease at my presence is understandable, it made me feel somewhat marginalised, particularly as the studio was a space that facilitated social interactions between the staff. This feeling was reinforced by my first scheduled observation with a visual designer, who afterwards said that they had found the experience a little uncomfortable. These incidents suggest that some of the designers regarded me with some suspicion, initially at least. It also demonstrates how my presence in the studio *affected* my participants.

It is not possible for the researcher to remove themselves from the social world while studying it and, therefore, it is inevitable that the researcher shapes the social phenomena which they study (Hammersley and Atkinson, 2019). It is likely that participants reflected on and even adapted their behaviour, including what they said, because of my presence there. Another response available to many of the participants was to revert to their first language. On the few occasions this happened in my presence I felt excluded and wondered if they had chosen not to speak English in order to exclude me from their conversation. Despite these early experiences, it did not take long for me to become more integrated among the designers. Everyone taking their lunch break at the same time and eating together at the same table, thus offering an opportunity to chat and build rapport, undoubtedly helped this.

Ethnographic descriptions are generally "one person's construction" of what a setting is like, usually produced by someone who is an outsider to that world and who may not speak the language (Stanley and Wise, 1983, p. 159). The language of the research setting not only refers to the country in which the research takes place but also the terminologies and lexicon used by the field in which she is researching. As Stanley and Wise (1983) argue, it is possible to feel a foreigner at home as well as further afield. As I was conducting fieldwork abroad, I was a foreigner on two counts: in a different country and in an unfamiliar setting. I was not accustomed to all the terminology the designers used, the names of their software, what it did, or how to use it. Even my laptop set me apart. Almost every designer had a slim apple MacBook. Not having my own laptop, I had borrowed a Lenovo Thinkpad laptop from the University's Sociology department for the duration of my fieldwork. One day, someone in the studio joked about my laptop, saying they had not seen someone with a Lenovo Thinkpad in years! This highlights the embodied nature of fieldwork, that goes beyond the presentation of the body to the props we possess which help researchers to present a *legitimate body* within the

research setting (Coffey, 1999). Although Reyes argues that there is no such thing as "true 'insiders'" (2020, p. 226), my laptop worked to reinforce my outsider identity.

The feeling of being a foreigner was amplified during periods of near silence in the studio. This occurred only occasionally, when the designers worked quietly at their desks, yet I found these periods of silence particularly uncomfortable. I wrote in my fieldwork diary how I found them *oppressive.* An occasional cough or the sound of footsteps across the studio were welcome breaks to the stillness. In these instances, I felt I was somehow more visible to everyone and I became very aware of my own presence in the space. Mason (2002) describes how, when conducting observations, the researcher's role is "less clear-cut" than in other methods such as interviews (2002, p.87). Silence brought this to the fore and in those moments, I felt lost in my role as researcher. The doubts I felt in terms of, if I was doing the research 'right' and even about how to recognise data as it happened (Stanley and Wise, 1983, see p.156) amplified. In the writing up of research, academics tend not to reflect on these problems (Stanley and Wise, 1983), and periods of silence during research interactions are rarely discussed or reflected on in qualitative research (Poland and Pederson, 1998).

Silence in the studio revealed the ways in which it was not always possible for me, as an observer, to know what was going on between the designers and their screens. It was difficult to capture the designers' practices as they worked alone with their materials. Geertz argues that ethnographers do not have "direct access" to social phenomena and must rely on what their participants show them to develop their understanding of what is going on (1973, p. 20). During scheduled observations, the designers tended to want to explain things to me and my own questions about their practices generated talk. Had I been embedded in the studio longer, it may have become easier for the designers and I to sit quietly as they worked. However, the challenge of not being able to see what they *see*, through their eyes, would have remained. This does not undermine the data or findings of this thesis, as Geertz argues, "it is not necessary to know everything in order to understand something" (1973, p. 20). Nevertheless, silence brought my participation in both the setting and the co-construction of data to the fore, in a way that was unexpected and uncomfortable.

The following subsection provides a reflexive account of conducting interviews with designers in the field.

3.2.2 Interviewing in the field

Whilst conducting observations at Evoke, I carried out semi-structured interviews with six data visualisation designers. The interviewees were selected because they had knowledge of different elements of the design process (including visual design, development, and data analysis) and were therefore, *expert informants.* As the research was taking place in a busy working environment, I used a combination of convenience sampling and selection by others to recruit participants (Hammersley and Atkinson, 2019). The former relates to recruiting those who were willing and had time to participate, while the latter describes when a participant identifies someone else who might take part (Hammerlsey and Atkinson, 2019). The interviews took place in various meeting rooms and working spaces at the company's premises during office hours. They were solely on the topic of data visualisation practices and also afforded an opportunity for me to ask follow up questions about the practices I had observed in the studio.

The interviews were not about the participants' relationships with clients, the company (their employer), or their colleagues. Although I could not guarantee it, for ethical reasons I tried to ensure the interviews did not stray off topic towards issues beyond data visualisation practices, thus limiting the possibility of the sharing of information that was more personal. I had prepared an interview schedule (see appendix b) to guide the discussions, while maintaining the flexibility to ask follow up questions about things the participants said. After conducting three interviews, I amended my interview schedule (see appendix c), in light of what I had learned in previous interviews and what I had observed in the studio. I was also motivated to amend the schedule in order to re-craft my questions so that they were easier for participants to understand. Both schedules included questions around:

- The design techniques designers used to encode meaning into the design of data visualisations.

- How designers imagined the end users of the data visualisations they produced.

- What data visualisation tools they had and how they used them.

Because the designers were very busy, Evoke's office manager told me that interviews could only last around half an hour, although multiple interviews with the same participants were possible. I was also told that the interviews would need to be scheduled through a project manager. This meant that I lost some control over my recruitment strategy, which resulted in unexpected issues. For example, a project manager who had been scheduling observations and

interviews for me had, in good will, scheduled interviews with every designer in the studio. Although this might appear fortuitous, I risked losing control of how the research and recruitment of participants unfolded. At that time, I was not yet sure what everyone's role was in the design process, therefore I had not made a decision as to who I would like to interview. Furthermore, as the interviews were intended to be supplementary to the observations, I had only set out to recruit half a dozen interview participants. As such, I had not broached the topic of an interview with all the designers. I was concerned that when they discovered a project manager had scheduled an interview in their diary, they may feel obliged to take part. Hammersley and Atkinson argue that researchers need to, "engage in the strategic search for data that is essential to a reflexive approach" (2019, p. 109). In order to reclaim control and maintain a reflexive, ethical, and strategic approach to data collection, I thanked the project manager but asked them to remove several of the interviews from participants' diaries.

Before commencing the fieldwork, I had presumed that the participants would predominantly be working on data visualisation projects. After arriving, it soon became apparent that many designers were working on the same large dashboard project for a big commercial client. However, the dashboard incorporated graphs and charts and I was able to attend meetings about its development and observe some of the visual elements being designed. I also observed a small-scale data visualisation design job for an existing client and a meeting about the direction of a new data visualisation project that had recently been commissioned. In fact, the types of work the studio was engaged in was revealing in itself. Nevertheless, as I only had three weeks at the studio it was important to establish who would be the best people to speak to about data visualisation design work. I had a meeting with one of the directors of the company and used the opportunity to ask who in the studio had worked on data visualisation design projects. They suggested two people I should talk to, both of whom I had already identified as possible interview participants. They were each involved in different parts of the design process and so I was happy to take the director's advice, approaching them both for an interview. I also decided to conduct three follow up interviews, asking those participants to bring with them a data visualisation they had worked on to discuss. This allowed me to ask them questions and get them to reflect on the design process in relation to a chart or graph they helped to produce. Despite there being few 'purely' data visualisation design projects taking place during the fieldwork period, I was still able to investigate the data visualisation design process through observations and the accounts of the designers.

As a researcher, it is important to consider the imbalance of power within the interview interaction and how that impacts on the production of knowledge (Brinkman and Kvale, 2015). While it is common to assume the researcher holds the power during interviews, Mason (2002)

explains that this is not always the case. The researcher is not always the one who exerts power over the interaction, particularly when interviewing *elites* (Brinkman and Kvale, 2015). The designers I interviewed can be considered *elites;* they were experts in their profession and worked for a company who were renowned in their field (Brinkman and Kvale, 2015). Brinkman and Kvale argue that elite interviewees are experienced at answering questions about "their opinions and thoughts" and may even enjoy the research interview (2015, p. 171). They suggest that interviewers ought to prepare for these interactions by ensuring they are knowledgeable on the subject they wish to speak to the elite person about. Although I had few practical skills in the design of data visualisations, the interview questions were formed, in part, from the data studies literature I had read. During the fieldwork, they were developed in light of what I had already learnt while being embedded in the design agency (see appendix c). They also incorporated follow up questions about what I had observed and in response to what the designers said in the interviews. Interestingly, as Brinkman and Kvale (2015) suggested, several of the participants expressed to me that they had enjoyed the interview experience and the chance to reflect on their work.

Brinkman and Kvale warn that because elite people are experienced at interviews it is likely they will have stock responses to questions which "promote the viewpoints they wish to communicate" (2015, p.172). This, they argue, requires the researcher to use their skills to elicit responses that go beyond prepared answers. Although participants referred to similar things during interviews, I never got the impression that they were drawing on stock responses as they spoke about their design practices in different ways. Furthermore, I was not asking questions about the values of the company they worked for but rather their own individual design practices. This made it less likely for them to elicit the viewpoints of their employer. However, the issue of power in the research interaction extends beyond the professional status of the interviewee to the socially ascribed aspects of the researcher's persona (Hammersley and Atkinson, 2019; Mason, 2002). These social identities are difficult to conceal and include gender (Harvey, 2010). I was a young woman and an inexperienced researcher interviewing mainly confident young men who were experts in their profession. I was also conducting these interviews in *their* work place - *their* world. Although this meant that interviewees often spoke with some authority, I never felt unable to ask follow up questions or ask for clarification. Perhaps this was because they were a similar age to me and because, over the weeks that I was at the studio, I developed a good rapport with the people I interviewed. Furthermore, the "acceptable incompetence" (Lofland and Lofland, 1995, p.56) researcher role I had, at first, unwittingly adopted, appeared to be disarming as I posed no challenge to their position as experts.

On my return to the UK, I transcribed the interviews I had conducted with the designers and began to read through the transcripts and my field notes to identify themes relating to how meaning was made from data through visualisation. I began by using a piece of coding software called Nvivo. Although I found this tool enabled me to code the data fairly quickly, I felt that it led me to produce a surface level analysis of the data. Therefore, I chose to analyse the data manually. This involved reading through the interview transcripts and field notes again and again, interpreting and re-interpreting what the data meant and noting down analytical themes within it. Slowing down the process of analysis let me engage more thoroughly with the data. I found it useful to paste quotes from the transcripts and field notes into word documents relating to themes I had identified. Writing about these quotes helped me to develop my ideas and depth of analysis. There were certain themes that jumped out at me during the fieldwork. These included the social and collaborative nature of data visualisation design and the way designers imagined older women and their ability to engage with data visualisations. I have subsequently written about these themes in Chapters four and five, albeit with more depth and nuance after having analysed the data. Despite having identified some points of interest during fieldwork, while conducting analysis I made sure to keep an open mind about what other themes might emerge from the data.

The second stage of the research involved interviews with 'ordinary users' of data visualisation. These produced rich contextual data, which revealed important themes for analysis that form the narrative of Chapter six on re-contextualising data. The following part of this chapter provides a reflexive account of this stage of the research.

3.3 Research design stage two: Interviews with ordinary users

This stage of the research design directed attention towards the consumption of data visualisations. In order to generate research data that would explain how ordinary users make meanings from data visualisations, I designed semi-structured interviews incorporating visual resources. As set out in Chapter one, 'ordinary users' represents anyone who might encounter data visualisations as part of their everyday life, for example at work, in news media or through self-tracking apps. The term also takes into account the different experiences and skills people possess in relation to interpreting graphs and charts, and the diverse settings in which they might encounter them. In designing the interviews, I drew on a photo elicitation approach. Although this tends to involve the use of photographs as visual resources through which to elicit responses from participants, Harper explains that they can incorporate "virtually any visual image" (2002, p. 13). Indeed, when choosing images to elicit focus group discussions around the changing social, cultural and visual landscape of County Durham after the demise of the

coal mining industry, Byrne and Doyle (2004) included simple graphical depictions. The face-to-face interviews I conducted included seven examples of data visualisations, which were intended to facilitate discussion and elicit responses from participants that showed how they were making meanings from them. The graphs and charts covered various substantive topics such as climate change, wildlife, and the gender pay gap.

During the interviews I showed participants the charts, one by one, letting them engage with the visualisation for as long as they needed before asking them to tell me what it was about. It is likely that Sociological terms, such as 'meaning making,' make little sense to ordinary users, yet the use of images in interviews can form a common ground between the interviewer and participant (Harper, 2002). Meaning making is an abstract term and as Chapter one illustrated, attempts to explain it have resulted in different ideas and concepts. Its abstractness might make it difficult for individuals to articulate how they did this. Furthermore, how they explained it would relate to how the individual *perceived meaning*. Therefore, I decided not to ask the participants explicitly about how they made meanings from the data visualisations I showed them. Instead, I designed the interviews and questions in a way that would allow participants to *show me* how they did this, through the accounts they gave me concerning what the different data visualisations told them. As intended, participants drew on the visualisation to discuss the substantive issue that it represented. This generated rich data about how participants engaged with the visualisations and the strategies and resources they drew on to make meanings from them. The interview interaction allowed the participants to provide their own situated accounts of their experience (Mason, 2002) of making sense of data visualisations, as they engaged with them.

Data visualisations are increasingly multi modal, meaning they include more elements than the visual image, for example audio, animation, interactive elements, or a combination of all of these (Engebretsen and Weber, 2017). As such, it seemed plausible that ordinary users are likely to encounter increasingly multi-modal charts and graphs in their everyday lives. Therefore, I selected a range of digital and analogue graphs and charts, including some with interactive, audio, and animated elements and some that were static (see table 1). These were acquired from a range of sources, including journalism and public science. The digital examples were viewed using an iPad, while the static ones were printed onto white paper. I decided to include these analogue examples to examine if the medium in which people encountered data visualisations shaped how they engaged with them, but also because people do not only engage with data visualisations in digital formats. As the literature review revealed, data visualisations are increasingly found in journalism (Weber, Engebretsen, and Kennedy, 2018), for which content is produced in both digital and print formats.

Table 1: The format of the data visualisations used in interviews

Data Visualisation	**Digital**	**Printed**	**Interactive**	**Animation**	**Audio**
Lion tracking	**X**			**X**	**X**
Antibiotic resistant E.coli	**X**		**X**		
Global temperature change	**X**			**X**	
Gender pay gap		**X**			
Fitbit dashboard		**X**			
Carbon emissions in Indianapolis		**X**			
Traffic congestion in Washington DC	**X**		**X**		

With countless charts and graphs to choose from, sampling required a practical and pragmatic approach. When designing qualitative research and collecting visual material, Mason (2002) argues researchers need to know what they are looking for. Research Question two, which asks how ordinary users make meanings from data visualisations, informed the selection of the visual resources. The visualisations that I selected for the interviews were all publicly available on the internet and aimed at a public audience. Therefore, it was possible that participants might encounter them, or similar visualisations, in their everyday lives. The project set out with the

aim of generating knowledge that would drive debates forward within data studies, but also to provide insights for data visualisation practitioners. Therefore, a commercial partner in the network designed three of the data visualisations I selected. This was with the intention that the interviews would provide contextual insights that might shape their future working practices. Existing research suggests that a users' prior knowledge of the subject that a graph represents might be significant in their comprehension of it (Shah and Hoeffner, 2002). Therefore, the graphs and charts I selected each represented a contemporary substantive topic. Incorporating a range of topics increased the likelihood that participants would be interested in the subject matter of at least one of the visualisations, while other topics might be less familiar to them (see table 2). This proved useful in exploring what participants drew on when making meanings from data visualisations when they are unfamiliar with the subject matter they represented.

The four data visualisations that were not from the commercial partner were sourced through google and google image searches for data visualisations on substantive topics, for example, by googling the phrase 'data visualisations about climate change.' Deciding what topics to select also required a practical approach and I decided to focus on contemporary issues that featured in the mainstream media (climate change, the gender pay gap, air pollution, antibiotic resistance) or topics that had broad appeal (wildlife, health and fitness), or relevance to peoples' everyday lives (traffic congestion). I also tried to choose topics that I did not deem to be sensitive and that were not likely to cause distress to my participants. The sampling method for selecting the visual resources for the interviews was grounded in my research aims and informed by existing literature. The combined requirements for the final set of visualisations comprised; *a range of substantive topics, a range of chart types, some multi-modal and some static designs*, and *three visualisations designed by one of the commercial partners in the network*. Furthermore, all the visualisations had to be publicly accessible and aimed at a public audience, which further narrowed the search criteria.

Of course using images hosted on the internet leaves the researcher with little control over their continued accessibility. One of the interactive visualisations I used in the first interview I conducted was about how public taxes were spent in Yorkshire and the Humber. By the time I conducted the second interview, this visualisation was no longer available through the website that had hosted it. For the remainder of the interviews I replaced it with a static data visualisation produced by the Guardian newspaper about the gender pay gap. The final seven visualisations used for the remainder of the interviews are detailed in tables one and two. While Chapter six provides a more comprehensive description of the data visualisations, table two provides a brief description of each chart.

Table 2: Data visualisation description and substantive topic it represents

Data visualisation	**Chart Description**	**Substantive topic**
Lion tracking	An animated data visualisation representing the movement of two lions. Each lion is represented by a simple yellow dot overlaid onto a satellite image of their territory. A narrator explains what is happening as the dot moves across the screen.	Wildlife and conservation
Antibiotic resistant E.coli	The visualisation resembles a pie chart, however each segment of the circle is the same size and each represents a different country in Europe. Density of colour is used to indicate rates of antibiotic resistant e-coli in each country over four years. The chart is interactive, allowing the user to click on a segment to bring up that country's annual percentages of resistance from 2012 to 2015.	Public health
Global temperature change	An animated spiral chart representing global temperature since 1850. The chart has a black background and a single line spirals outwards from zero Degrees Celsius in the centre, towards 2.0 degrees at the outer circle. The line is blue to begin with but becomes yellow as it spirals outwards.	Climate change

Table 2: Data visualisation description and substantive topic it represents

Data visualisation	**Chart Description**	**Substantive topic**
Gender pay gap	A static chart in which individual companies are represented by a dot. Their position on the chart, as well as the colour of the dots, indicates if the company pays women more or less than men.	Gender pay gap
Fitbit dashboard	A range of small charts depicting one person's Fitbit data over the course of twenty-four hours.	Health and fitness
Carbon emissions in Indianapolis	A bubble chart overlaid onto a grid map of Indianapolis. Each bubble represents a source of emissions and the size of emissions from that source.	Climate change/Pollution
Traffic congestion in Washington DC	An interactive map visualisation showing traffic congestion in Washington DC. A toggle allows the user to adjust the time of the day and days of the week to see how congestion changes. Interactive overlays provide additional quantitative information in the form of the traffic time index for particular roads.	Transport and traffic congestion

The seeing data project (see Kennedy et al., 2016b) recruited focus group participants by making contact with existing community groups. This seemed like a practical way of recruiting a diverse range of people by targeting different types of community groups. I conducted an internet search of community groups in the York area and made contact with several to ask if I could attend one of their meetings to introduce myself and my research. Many of these groups did not reply to my email and one who agreed to me attending their meeting later cancelled. Therefore, I found recruitment to be slow at first. However, I soon found that gaining access to a community group was much easier when I already knew one of their members who could act as a gatekeeper. Indeed, Hammersely and Atkinson (2019) note the potential significance of utilising personal social networks and contacts in order to gain access to different groups. This led me to expand my search area to include the whole of the North East and Yorkshire so that I could draw on my social network to gain access to different community groups.

I spoke to nine groups in total. These comprised of two sports clubs, an art group, a countryside sporting group, a craft group, a social justice group, a conservation group, a breastfeeding group, and a mother and toddler group. I also gained two participants through snowball sampling; one had heard about the research through a friend while the other had been given the information sheet by their wife. I had met the latter participant's wife at a community group she attended but she had decided not to take part in my research. Interestingly, the participant told me that she had given my information sheet to him because she thought it was '*more their kind of thing*.' *Who* my research attracted in terms of participants is significant and will be discussed more in the following paragraphs.

Drawing on my social network to gain entry to community groups became crucial to my recruitment strategy, perhaps because of my research topic. Firstly, data visualisation is a clunky term and not everyone knows what it means. Secondly, people can be put off from engaging with graphs and charts because they lack confidence in their ability to interpret them (Kennedy and Hill, 2018). Finally, it might be difficult for people to see how research about data visualisation is relevant to *their* lives. These issues were reflected in encounters I had with people during the recruitment process. For example, I attended a working men's club in the North East of England where the members of a local running club met for a drink after they had trained. My friend was a member of the club and had negotiated access for me to attend that evening. As the runners sat drinking and chatting in the bar, my friend called out to say if anyone was interested in hearing about my research to head over to the table where we were sitting. One of the other members shouted, "*all the clever people go over there,*" directing people to our table while they remained seated. This person appeared to have decided my research was not for them. This was frustrating because I wanted to include people in my

sample who might encounter data visualisations in their everyday life but who had little interest or knowledge of them beyond that. However, such people were hard to recruit.

Despite the difficulties of gaining access to community groups, they offered the chance to speak to people face-to-face about my research in lay terms. I could explain the term data visualisation and make it clear to them that I was not interested in how accurately they interpreted graphs and charts. Instead, I wanted to understand how different people made sense of them. It also provided potential participants with an opportunity to meet me, ask questions, and for me to convince them that the research was interesting and important! I only spent a few minutes speaking to people at the beginning or end of their meetings so as not to take up too much of their time. Afterwards, I asked anyone who might be interested in taking part in the research to take an information sheet, which included my contact email address. This was to give people an opportunity to read about my project in their own time. Knowing what taking part would involve would allow them to make an informed decision about participating. In reality, some people agreed to and even scheduled an interview during the meeting. In fact, I found email to be a poor medium for participants to contact me. At the running club, a potential participant asked if he could take a phone number instead. I agreed and this prompted some of the other people there to ask the same. I got a higher rate of response from that group, mainly through text messages, than I had been achieving from other groups. Before conducting the research, I would have considered text messaging to be too informal to use to contact participants. However, I came to realise that it was precisely because of its informality that people preferred to communicate this way. Therefore, I began to write my mobile number on the back of the information sheets.

I set out to recruit a socially diverse sample of participants, both in terms of interests and socio-cultural categories such as gender, age, ethnicity, education, employment, and disability (specifically dyslexia and dyspraxia). While I could attend a range of community groups with the aim of recruiting participants with varied interests, as this is a small-scale qualitative study, I decided not to sample for socio-cultural categories during recruitment. Instead, I asked participants to complete a short demographic questionnaire at the beginning of the interview. After conducting twenty-five interviews my recruitment strategy had not resulted in any ethnic diversity, so I made contact with a mother and baby group at a mosque where I spoke to their members about my research. I recruited 28 participants in total for the user interviews; the youngest participant was a 26-year-old woman and the oldest was a 76-year-old woman. I interviewed 20 women and 8 men and, although the majority could be considered ‘non-experts’ in data visualisations, they had varied experience of interpreting graphs and charts. Indeed, a small number of participants were experienced in interpreting visualisations, as they engaged

with them at work. I stopped recruitment when I saw the same themes reappearing during the interviews and I recognised participants tended to be drawing on the same strategies to make meanings from the data visualisations, as outlined in Chapter six.

I reflected on the interview sample as the research progressed and I was surprised to see a disproportionate amount of teachers taking part. Out of 28 participants, 10 were current or retired teachers or teaching assistants, one of whom was male and the rest female. As I was not sampling for socio-demographic categories during recruitment, their job role only became apparent when they completed the demographic questionnaire at the beginning of the interview. I have considered why the research attracted so many participants who were teachers and although it is impossible to be sure, I can surmise that it might be because they have a general interest in education and research. However, as I reflected further on the sample I began to realise that I was attracting participants who were like me; white, educated, middle class women. The strategy of recruiting participants from different community groups was to achieve a diverse sample in terms of socio-demographic categories and interests. However, relying on my social network to gain access to different groups may have played a role in narrowing diversity. Furthermore, five of the community groups I attended took place during working hours on a weekday. Therefore, many of the people I spoke to were able to attend interviews because they did not work full time, were retired, on annual/maternity leave, or worked shifts, which will also have shaped my sample.

The interviews were designed to last no more than an hour, plus an additional ten to fifteen minutes for briefing and debriefing participants. In reality, they varied between thirty minutes long and an hour and fifteen minutes long. Twenty-three took place in a public place, usually in a café but sometimes at the community groups they attended if this was more convenient for the participant. Three took place in my home as they were with people I knew and two took place at the homes of the participants, whom I also knew. In these instances, meeting in our homes was more convenient for the participants. The interviews were audio recorded, enabling me to listen back to the interviews and to make additional notes to those I took during the interaction. While the recording is still a reconstruction of the interview interaction, it provides the possibility to revisit those conversations (Brikman and Kvale, 2015) to identify the moments in which the participants revealed how they made meanings from the data visualisations. The initial interviews were transcribed in their entirety. In later interviews, having the audio recordings enabled me to transcribe the data that related to *how participants made meanings* from the data visualisations I showed them for more in depth analysis (Brinkmand and Kvale, 2015). During the transcription, the data was anonymised through the use of pseudonyms and the omission of any personal information that the participant revealed, which might have identified them.

The majority of interviews took place in cafés because they are an informal setting where I hoped participants would feel comfortable. At the beginning of the interview, I would lay out my materials on the table, squeezing them in between our mugs of tea or coffee. These interview materials comprised of an iPad; a touch screen pen; a small square mobile wifi device; a file containing the information sheet (see appendix d) and consent form (see appendix e), questionnaire (see appendix f), interview schedule (see appendix g) and data visualisations in print; my Dictaphone; a small green notepad; and a pen and pencil. The participant and I would drink our tea or coffee as we went through the different visualisations. Occasionally the background noise of the café made it difficult for participants to hear the audio element of the wildlife data visualisation and on a number of occasions we had to skip that one. Sometimes female participants would bring their young (pre-school) children with them. This happened five times and it required the interview to be more flexible, to include pauses and breaks so that the children could be looked after. While *mum* focused on the data visualisation, I would keep an eye on her child/children, or entertain them for a few minutes. This made it difficult for both her and I to immerse ourselves fully in the interview interaction and I often made less notes during interviews where there were children present. Although this may seem detrimental to the interviews, when people are making sense of data visualisations in their everyday lives it is possible they will be distracted while doing so. Therefore, by being flexible and allowing for pauses in the interaction, these interviews still generated data that provided insights into how people make meanings from graphs and charts.

Accessing groups through my contacts meant that sometimes I would interview people I already knew. This is because my contact would often volunteer themselves, or in two cases because the group had links to my wider social circle. Out of twenty-eight participants, I already knew eight of them in some capacity prior to the interview. Interviewing people I knew brought about a challenge as during the interaction the dynamic of our relationship shifted to that of an interviewer and an interviewee. Braun and Clarke refer to this as a "'dual relationship'" in which the interviewee is both friend and participant (2013 p. 87). Interviewing participants that are known to the researcher also raises the ethical risk of the interviewer encouraging the participant to reveal something about themselves that they might not have wished to disclose (Braun and Clarke, 2013). As the interviews were structured around the visual resources and were restricted to how participants made meanings from data visualisations, the risk of this happening was low. Still, I found these interview interactions more difficult due to the change in our social dynamic and I preferred interviewing people I did not know. When interviewing strangers, our interactions were not loaded with prior expectations or the social history associated with our knowing one another.

Mason (2002) suggests that when preparing for qualitative interviews, the researcher ought to consider how they want to come across during the interaction. However, as with my fieldwork experience, I found that choosing a *demeanour* (Mason, 2002) was not always a conscious decision. As Mason (2002) argues, the interview interaction is shaped by *both* parties and I found my 'interviewer style' changed depending on the person I was interviewing. As I listened back to the audio recordings of the user interviews, I discovered that when a participant presented himself or herself as a confident interviewee, I respond by appearing more confident and direct when phrasing my questions. However, when a participant appeared to be nervous or unsure I played down my knowledge, I used the word *'like'* to fill pauses in my speech and I was less likely to allow for extended periods of silence. I realised that in empathising with the participant I adapted my interview style without being consciously aware of it. Although driven by my desire to make the participants feel comfortable it shaped and altered the interaction.

The prevalence of interviews in the media mean that publics are familiar with the concept of the interview interaction (Brinkman and Kvale, 2015). Therefore, in the context of the *interview society* (Atkinson and Silverman, 1997), it is likely that participants are aware of the "social expectations" of the interview interaction and what it means to be a *good* interviewee (Poland and Pederson, 1998, p. 301). Indeed, Poland and Pederson argue that participants might "seek to present themselves as competent social agents" during interviews (1998, p. 301). It is possible that during my interviews with ordinary users, participants were performing their understanding of the data visualisation in ways that they might not do in their everyday consumption of graphs and charts. The anxieties that data visualisations can provoke around numerical literacy may have exacerbated this, causing participants to over-explain their understanding in order to be a *good interviewee*. Furthermore, the participants were viewing the visualisations *out of context*. By this I mean that they were not viewing a graph embedded in a newspaper article nor were they exploring the website in which the graph or chart was hosted. However, it is not unusual for media content, and images in particular, to be removed from their "context of production" as they are shared on social media (Bødker, 2016, p 414). Furthermore, data visualisations are standalone images and it is possible for people to engage with them without engaging with the additional text that might surround them.

The participants 'performance' of their understanding of data visualisations and the fact they were viewing them 'out of context' frames the knowledge produced in Chapter six. While this could be argued to be a limitation of the research, qualitative interviews are always artificial encounters in which knowledge is co-produced through the social interaction between the researcher and participant (Brinkman and Kvale, 2015, see p. 63). This does not mean that the knowledge and insights produced through this method are without merit. Rather it requires the

researcher to provide an open and reflexive account of the context in which that knowledge was produced and how that knowledge continued to be constructed during the "analysis, and reporting" of the interview (Brinkman and Kvale, 2015, p.63). This requires an acknowledgement of how the analysis, through which I identified themes and patterns in the data and developed concepts to explain them (Brinkman and Kvale, 2015), was shaped in part by my "ideas, assumptions and beliefs" (Brown, 2019, p. 33). Indeed, I found that the analysis of the ordinary user interview data required me to interpret the same data visualisations I had shown my participants. Therefore, my analysis is shaped by my own interpretation of the graphs and charts I asked them to interpret. As such, where possible, I have included images of the graphs or charts so that the reader can interpret my analysis of how participants interpreted them for themselves.

I found the interview interactions to be rich in terms of the data they were generating and after conducting several interviews I began to identify different meaning making strategies used by the participants. To help me think analytically I started to take a notebook with me to the following interviews, writing down things participants said that related to the strategies they seemed to be using to make meanings from the visualisations. The act of writing these notes was the first step in identifying interpretative themes in the analysis of the data. I analysed the interview transcripts manually, reading and re-reading them to interpret their meaning. I pasted quotes from the transcripts into a word document and tentatively organized them into the rough themes I had identified, whilst keeping an open mind about alternative meaning making strategies that might appear in the data. I found the same three strategies of meaning making kept appearing in the interview data. While not every participant used all three strategies, most participants utilised at least one in their interpretation of the data visualisations. With the quotes sorted into interpretive themes, I was able to analyse them more closely for how they explained meaning making and to develop a stronger analytical foundation for the different strategies I had identified. The analysis of the 'ordinary user' data is presented in Chapter six.

The following section of this chapter will discuss the ethical considerations of this research project.

3.4. Ethical considerations

The project was subject to ethical approval from the University of York's Economics, Law, Management, Politics and Sociology (ELMPS) ethics committee. Although I was not researching a particularly sensitive topic, the ethical considerations for this project were diverse.

The following section will reflect on some of the ethical issues that arose before, during and after the research began.

3.4.1 Ethics of fieldwork

Conducting research in another country raises concerns about how cultural differences might make it more difficult for the researcher and participants to understand one another. However, I was conducting fieldwork in a Western European country and, although there were subtle cultural differences, they did not make it difficult to conduct the fieldwork. Furthermore, the language spoken in the studio was English. Doing ethnographic research is a personal and emotional experience as well as an intellectual one (Coffey, 1999). Conducting fieldwork abroad heightens feelings of "loneliness, anxiety and perhaps even alienation," all of which are associated with conducting participant observation - at home or away (Lofland and Lofland, 1995, p.52). To manage the emotional and intellectual pressures of fieldwork I drew on a strong support network, including my fellow PhD students, and my supervisors whom I kept in contact with so that I could talk through concerns as they arose. My husband, family, and friends offered invaluable support and my conversations with them on evenings and weekends helped to counter feelings of loneliness. However, the intensity of the fieldwork was exhausting and, in practice, I had little time or energy to feel alone. In fact, *being alone*, enabled me to immerse myself in the kind of intensive data gathering required when conducting short-term ethnographies (Pink and Morgan, 2013). I was present in the studio from 9am to 5pm most weekdays, conducting observations and interviews, before writing up my field notes during the evening.

All the interviews I conducted were audio recorded and on their transcription the data was anonymised using pseudonyms. As Evoke undertook projects for a range of commercial clients, I also anonymised the names of different projects or clients that the participants referred to in our conversations. As the purpose of the observations was to observe and track data visualisation design practices, I had initially planned not to record names of individuals in my field notes. Instead, I would describe their data visualisation practices in terms of job roles. However, after entering the field it quickly became apparent that this was not practical; not only would it make the field notes confusing, but I quickly realised that the participants' roles were not as clearly defined as I had expected. Therefore, I used pseudonyms when writing my field notes and my field diary.

Whilst planning the research, anonymising the data appeared to be a straightforward process. However, as I discovered, the context of the fieldwork site made this more complicated than I

had anticipated. Despite my attempts to anonymise all the participants, it is a relatively small company and as participants' roles are described through the field notes and interview data, it could be possible for colleagues to identify one another in the data. Furthermore, although the language used in the studio is English, the particular idiosyncrasies in the way people use English, most often as a second language, is another way by which the participants' colleagues or employers may be able to identify them. To protect participants I have tried to use shorter excerpts from my field notes and interview transcripts. Where this is not possible, I have paraphrased certain extracts where I think direct quotes or certain phrases may make participants particularly identifiable. Finally, the designers were nearly all male so to refer to their gender when writing up the research would make certain members of staff easily identifiable to their colleagues. To mitigate this, when writing up the research I have referred to the designers using gender neutral pronouns and pseudonyms.

Writing about the fieldwork participants' roles in the office and the projects they have worked on may also make them more easily identifiable to one another. Furthermore, the company had been publicly linked with the project prior to the fieldwork commencing. Therefore, despite anonymising the fieldwork site and using gender neutral pseudonyms to refer to my participants, it was not possible to guarantee the participants' anonymity. I explained this in the information sheet and consent form provided to all participants. I provided participants with separate information sheets and consent forms for the observations and the interviews. However, as they contain information which would reveal the identity of the fieldwork site I have not included examples in the appendices. Tilley and Woodthorpe (2011) explain how anonymity in the information age is becoming increasingly difficult as online search engines can throw up information that may make fieldwork sites identifiable. Furthermore, researchers are encouraged to disseminate widely and to make sure their work has impact, which makes it harder to protect the anonymity of fieldwork sites and participants (Tilley and Woodthorpe, 2011). To mitigate the risk that participants become identifiable to their colleagues I have carefully considered how I use data in the analysis chapters, so as not to make value judgements about the individual participants' practices that may cause them harm or embarrassment. This required striking a balance between maintaining research integrity and the protection of my participants (Hammersely and Atkinson, 2019), to ensure they would not experience negative repercussions from their decision to take part in the research.

Embedding myself as a researcher in my participants' place of work raised further ethical issues around the potential harms my research may cause them. The founders of the company had given the network their permission for me to conduct fieldwork in their company. Therefore, before I began the fieldwork, I had identified the risk that employees may feel obliged to take

part in the research. To alleviate this, I made it clear in the information sheet I gave them that it was their decision as to whether or not they take part in the observations or interviews and they were free to refuse. I reiterated this verbally to potential participants when handing them the information sheets and consent forms. I also explained in the information I gave them that they had the right to withdraw from the research process at any point before submission and dissemination, should they change their mind about participating.

The requirement to schedule observations and interviews through one of the project managers and for these to be diarised meant it was impossible to maintain the confidentiality of those who participated within the company. This was further complicated by the fact that the studio space was open plan and all but one of the meeting rooms on the ground floor were built with clear glass walls, meaning the occupants were always visible. Although it was not possible to maintain the confidentiality of participants within the design agency, the risk to the participants was mitigated by the project's narrow research focus. Nevertheless, I had to be mindful of how the participants' views and practices might not reflect those of their employer. This is made more important by my agreeing to feedback the research findings to the company, in return for them allowing me access to conduct the research. However, I only collected data on the processes involved in the production of data visualisations, therefore, limiting any personal harm to participants who may have discussed sensitive or personal issues in my presence.

While in the field, researchers are required to make ethical decisions in the moment about what to record, what not to record, and how to respond to what participants say or do (Mason, 2002). Having a clear focus in terms of my research and a strong sense of my ethical obligations to both my participants and myself enabled me to negotiate the complex ethics of fieldwork. Although it is impossible to cover every ethical consideration undertaken, this subsection has provided a reflexive account of my approach to conducting ethical fieldwork based research. The following subsection will discuss the ethical considerations relating to the interviews I conducted with ordinary users.

3.4.2 Ethics of ordinary user interviews

Using the ethical approval I had obtained from the ELMPS committee, I employed various strategies to conduct this stage of the research ethically. This began with asking community group leaders for their consent to attend one of their meetings to introduce my research and myself to their members. Either this was obtained by contacting them directly or through a contact I had within a particular group. I sent a copy of my project information sheet to the group leader prior to attending the meeting so that they could read it beforehand. This gave

them an opportunity to reflect on their decision and cancel my attendance if they wished to. Information sheets were provided to anyone at the meeting who expressed an interest in taking part in the research. Still, before every interview began, I provided participants with a copy of the information sheet with the consent form to ensure they had the opportunity to ask me questions about taking part. The consent form, which explained how the interview and demographic data would be used, was signed prior to their completion of the demographic questionnaire and the interview commencing. For data security reasons the participant's name was not on the demographic questionnaire so that it could remain anonymous. Instead, a code was used so that I could identify the participant who completed the questionnaire later, for the purposes of analysis.

I aimed to select non-sensitive data visualisations as visual resources for the interviews. However, what people find sensitive is related to their personal contexts. At the beginning of the interviews I explained to participants that they could skip a visualisation they did not want to discuss and pause or stop the interview at any time. Data visualisations can be complicated and interpreting them often relies on a level of numerical literacy. As Kennedy and Hill (2018) found, difficulty in interpreting the data visualisations may lead to confusion, frustration and/or embarrassment. It was not important to the aims of my research that participants interpreted what the visualisation meant 'accurately.' Instead, I was interested in how they drew on the visualisation to talk about the topic it represented for them. Therefore, whilst briefing the interview participants, I made it clear that the interview was not a test of their understanding of data visualisations. I explained that I was more interested in the different ways people make sense of them and furthermore, I was not an expert myself in interpreting graphs or charts. This was important for making participants feel at ease during the interview and to mitigate unpleasant feelings associated with any difficulty they experienced in engaging with the visual resources I showed them.

While conducting this research project, I discovered that ethics unfold in research interactions in ways that cannot always be anticipated. The University of York's ethical approval process prompted me to think about the ethical issues that might emerge during the research and how I would approach them. However, completing research with human participants is unpredictable and it is impossible to pre-empt every ethical dilemma that might arise. Nevertheless, I possessed a strong sense of my ethical responsibilities to my participants and I have used this to negotiate the ethical issues that arose throughout the fieldwork, data collection, analysis, and writing up of the thesis.

3.5 Conclusion

The thesis draws on encoding/decoding (Hall, 1980) as a theoretical framing. This informed the methodology, which required an approach that could explore meaning making in *both* the production and consumption of data visualisations. In response to the thesis' research questions and theoretical framing, the research was designed in two stages. The first addressed research Question one and three by conducting a short-term ethnography at a data driven design agency. This involved conducting observations of the designers working in the studio and interviewing them about their practices. Observations produced situated knowledge about their meaning making practices as they were unfolding (Mason, 2002). Interviews allowed the designers to explain these practices to me in their own words (Brinkman and Kvale, 2015), providing insights into their world of data visualisation design. Combining these methods enabled the kind of purposeful data gathering that is required when undertaking short-term ethnographies (Pink and Morgan, 2013). The fieldwork produced data that was analysed to explain how the designers encoded meaning into data visualisations within this particular socio-technical context.

The second stage of the research addressed research Question two and involved conducting interviews with ordinary users to understand how they made meanings from data visualisations through their interpretative practices. Drawing on a photo elicitation method, the interviews incorporated seven examples of data visualisations which I asked participants to tell me about during our interaction. Harper argues that researchers from a cultural studies perspective are often criticized for making assumptions about how audiences perceive messages through images (2002, see p. 19). However, *encoding/decoding* (Hall, 1980) pays close attention to how audiences decode messages. By incorporating images into the interview interaction I was able to identify the strategies ordinary users drew on to make meanings from the graphs and charts I showed them. Meaning making is an abstract concept and one that may have made little sense to my participants if I had asked them to explain how they did this. Therefore, using graphs and charts in the interviews allowed the participants to *show me* how they made meanings from them through their interpretive practices. The data from both phases of the research was analysed for themes and patterns, which were then interpretively analysed for their meaning (Brinkman and Kvale, 2015), in relation to how meaning is made through the production and consumption of data visualisations.

This chapter has set out the project's methodology, whilst critically reflecting on the methods used. It has provided a reflexive account of the research by being open about how the decisions I took shaped the research and findings. The chapter has also discussed the ethical

considerations of conducting both fieldwork with designers of data visualisations and interviews with ordinary users. The following three chapters are based on the analysis of the data collected during this research.

Chapter 4: Encoding meaning in the design of data visualisations

4.1 Introduction

This chapter draws on observations and interviews I conducted at *Evoke*, a data driven design agency who produce data visualisations and dashboards, the latter of which encourage ongoing visual engagement with data (Beer, 2019). During fieldwork there, I was able to observe designers working on visualisation projects and talk to them about their practices. This provided me with a unique insight into how meaning was made from data, before being encoded into the design of data visualisations. The large open plan studio where the designers worked facilitated a collaborative way of working. I often observed designers standing or sitting at one another's desks in twos or threes, informally discussing projects and developing designs. While observing more formal meetings, I encountered visual designers and front end developers working together to *translate* visual designs into code, in order to produce digital and interactive data visualisations. Translating the design into code required collaboration and compromise between the designers and they often described this as a "*back and forth*" process. These interactions between designers during the design process were nearly always mediated by technology. These forms of technology ranged from computer hardware and design software to white walls (to draw and write on), post-it notes, and paper. This technology acted as a focal point for the designers' collaboration, through which they designed and built data visualisations.

Rather than being the product of a lone designer, it was often the case that the design of data visualisations involved a "socio-technical assemblage of actors and actants" (Kitchin, Maalsen, and McArdle, 2016, p. 98). Dominant discourses surrounding big data suggest that insights can simply emerge from data (see Anderson, 2008). However, in the studio, meaning was *made* from data through the designers' knowledge, expertise, and ideas, which unfolded, "within a set of social and economic constraints [and] existing technologies and systems" (Kitchin, Maalsen, and McArdle, 2016, pg. 98). This chapter draws on the research data gathered at Evoke to unpick this socio-technical assemblage and explain how it frames meaning making in the design of data visualisations. In doing so, it will highlight how a range of contextual factors, originating beyond the dataset, shape how meaning is made from data. The chapter will draw on Hall's (1980) notion of *encoding* as a conceptual tool to explain how the designers negotiate how meaning from data is communicated through the design of visualisations.

Borrowing and building on Kitchin's (2014) notion of the *data assemblage*, this chapter explains how components of the assemblage combine to shape how meaning is made from data,

as well as how that meaning is communicated through the design of visualisations. Applying Kitchin's concept to the domain of data visualisation design reveals how socio-technical assemblages frame how meaning is made from data in the pursuit of value. In his work, Kitchin (2014) does not provide an explicit definition of the term *assemblage*. However, Beer (2013) provides a useful discussion of the term in his book *Popular Culture and New Media* and I have drawn on this to develop a working definition of assemblages. Therefore, this thesis understands assemblage to mean the parts and the relationship between those parts, which form a whole. This allows for research that shifts in scale from the whole, which in this instance is the data visualisation, to the relational parts within which it came into being and how they constitute one another. Like Kitchin (2014), I consider these parts to include people, technology, regulations, conventions, finance, and infrastructures. Indeed, Beer explains how the concept of assemblage is useful because it "brings together objects and infrastructures with bodies, culture and data" (2013, p. 36). Therefore, Kitchin's concept of the data assemblage complements this thesis' theoretical framing, *encoding/decoding* (Hall, 1980).

The chapter will explore how meaning is made from data and encoded into the design of data visualisations in three ways. This forms the three sections to the chapter, which address meaning making through discourse, design, and production practices (Kress and van Leeuwen, 2001). The first section reveals the social and regulatory apparatuses of the socio-technical assemblage that shapes how meaning is made from data in the design studio. This relates to the **discourse** stage of the design process and explains how the aims of the client and the level of access designers have to the dataset shape the meaning of data. The second section analyses how meaning is encoded into the **design** of data visualisations through the knowledge and ideas of the designers and how the studio facilitates collaboration between designers with different expertise. The third section explains the *back and forth* process through which designers work together in order to translate visual designs into code during the **production** of digital data visualisations. In doing so, it explains how the data assemblage frames collaboration between designers and thus shapes how meaning is encoded into the design of graphs and charts.

In analysing meaning making in the production of data visualisation design in these three ways, this chapter will argue that it is useful to conceptualise how meaning is encoded into data visualisations in two stages. The first requires the designers to decode the meaning of the data prior to its visual representation, while the second part involves encoding these meanings into the design and production of data visualisations. The interviews and observations this chapter draws on involved participants with different skill sets including visual design, coding, user experience design, and data analysis. Although most possessed skills in more than one of these specialisms, this assemblage of actors involved a demarcated division of labour, which I will

return to in section 4.4. However, throughout the thesis all fieldwork participants will be referred to simply as 'designers,' with their specific skill set discussed only when relevant to the analysis.

4.2 Making meaning from data

This section of the chapter explains how data is made meaningful prior to its visualisation. It argues that meaning does not simply emerge from the data but is made in relation to the context in which the data is to be used. Therefore, what the data means relates to what the clients who commission visualisation projects, as well as the designers, believe it is possible to do with data. This is particularly true when the designers do not have access to the *real* dataset being visualised. It will discuss how meaning is made from data in three ways. Firstly, how the particular context of the project frames how meaning is made from the data being visualised. Secondly, how designers make data meaningful when designing dashboards without access to the dataset that will eventually flow through it. Finally, how the designers made meaning from data by finding a narrative within the dataset. The latter relates to more stable data visualisation projects where the designers have access to the dataset being visualised.

4.2.1 Understanding the context

Data visualisation design is embedded within a socio-technical assemblage, whose social and regulatory apparatuses frame how meaning is made from data. This section of the chapter begins to explore how these apparatuses present opportunities and constraints for how meaning is made from data, which are context specific, vary between projects, and are dependent on the format of the visualisation. By format, I am referring to either a dashboard application or a more stable data visualisation project where the dataset is less changeable. The context of a project includes the aims of the client who commissions it, the level of access the designers have to the dataset being visualised and the perceived needs of the users of the visualisations. (I will return to how designers imagine the users of data visualisations in Chapter five). This section will argue that context specific socio-technical assemblages frame the meanings of data visualisations by shaping what they reveal. This section will also draw on Beer's concept of the "data imaginary" (2019, p.15) to analyse how meaning is made from data in relation to what people imagine it is possible to do with data.

Evoke design what I will refer to as stable data visualisations, that is, stand-alone images of graphs and charts that might be embedded in websites or news articles. Some of these are static images, published in a print or digital format, whilst others are multi layered, interactive, digital

data visualisations, which are usually made accessible to the public through websites. Evoke also produce dashboards, which have become ubiquitous in both public and private sector organisations, where they are used to facilitate "data-driven decision making" (Sarikaya, Correll, Bartram, Tory, and Fisher, 2019, p. 682). Dashboards can incorporate both visual and functional components and range from a "single view reporting screen" made up of several basic charts, to dynamic "interactive interfaces" which process data in real time (Sarikaya et al., 2019, p. 683). While I was conducting fieldwork, many of the designers were working on elements of the same large-scale dashboard project for a commercial client who wanted to monitor their global sales. That so many of the designers were working on this project is a reflection of the increasing importance of dashboards within organisations. As companies gather more and more data they are compelled to find ways of generating value from it (Beer, 2019). This has led to the rise of the analytics industry, which promises to help companies find insights in their data (Beer, 2019). Visualisation is used as a means of making insights from data accessible to those without expert knowledge of data analysis, therefore it involves simplifying insights so that they can be understood easily and quickly (Beer, 2019). Despite many of the designers being occupied with the dashboard project, during interviews and observations they talked about and reflected on other, more stable, data visualisation projects they had worked on.

The design process at Evoke, for large projects at least, appeared to be split into a number of distinct phases and I will discuss three of these phases in this chapter. Firstly, *phase one,* which was an initial research phase where the designers learned more about the aims of a new project. Secondly, *phase two,* which was a concept phase during which the 'look and feel' of the project was determined. Finally, the *development phase,* which was when the visualisation product was built and this involved collaboration between visual designers and front end developers who constructed the visualisation in code. These distinct yet interrelated phases map onto the discourse, design and production practices in which Kress and van Leeuwen (2001) argue that meaning is made in the design of multimodal texts. Although it is possible to distinguish between these phases, the overall design process was an iterative one in which elements of the design were revisited and reworked and this was particularly so during the development phase. This section of the chapter will explain how designers made meaning from the data in phases one and two, before encoding this meaning into the design of data visualisations.

Wright, one of the founders of the company explained to me that phase one is client facing and aims to establish why the client has approached them with *this* dataset and *how* the client intends to use the data. This phase, Wright told me, required designers with skills in user experience design but might also involve those with data analysis skills in order to explore the data. They explained that during phase two, visual designers sketch out some concepts based on

the information gathered in phase one and begin to prototype some design ideas. Wright explained that these are not finished or polished designs and that the designers will need to consult with the front end developers to discuss how their prototypes will work in *reality.* In explaining phases one and two, Wright revealed the range of expertise utilised in the design of data visualisations. However, Wright's use of the word *reality* eludes to the ways in which the design of data visualisations was constrained by contextual factors beyond both the dataset and the ideas of the designers, which this section will begin to unravel.

Evoke's website lists the services they provide to support the design of data driven products including those relating to the first phase of a project. These include workshops and interviews with users and stakeholders, which aim to provide Evoke with a better understanding of the needs and aims of their client. It seems likely that these services relate more to the professional dashboard interfaces Evoke design and produce for companies, for which it would be possible to gauge how the product would be used within a particular setting. This view was reinforced during my interview with Clarke. Clarke explained to me that they tend to work on projects designing applications that deal with data rather than more stable data visualisations. It was my understanding that Clarke was referring to dashboard applications, so I asked if they design the kind of graph that appears when a user wants to find out something from the data or compare data, they said:

> Clarke: "*Exactly […] my job is to first have all the interviews with potential users and with the business to hear what are the things they are generally interested in. So, what do you want to see? And then what type of decisions do you make from that? […] And by knowing what data it is that they are looking for and what potential patterns they are looking for, I can choose the right visualisation that will help them […] that plots the data in the most efficient manner for them to get that insight."*

This extract implies that data is made meaningful in relation to what people want to use the data to do. This, it seems, also determines the style of graph or charts used to visualise it. As well as making meaning from data through methods of data analysis, it is also the case that data's meaning is made through discussions between visualisation designers, their clients, and potential users. In this sense, the meaning of data is already framed by the context in which it will be used and shaped by the aims of the client. When thinking about how meaning can be attributed to data prior to its analysis or visualisation, it is useful to draw on Beer's concept of the "data imaginary" (2019, p. 15). This describes how the social world imagines data and "the possibilities they have been ascribed" (Beer, 2019, p. 6), that is, what people imagine data can do and therefore what they can do with data. The possibilities afforded to data through the data

imaginary drives a compulsion to generate insights. The analytics industry has grown with the promise of providing these insights quickly and accessibly, largely through visualisation (Beer, 2019). Perhaps though, this imaginary also plays a role in how meaning is made from data as peoples' understandings of what can be done with data frames how it is made meaningful in particular contexts and for particular purposes. Indeed, in his work on metrics, Beer (2016a) explains how data already has purpose and intent, therefore neither data nor its visualisation are ever neutral.

Clarke's quote suggests that making meaning from data in the context of dashboards is linked to the use of data for decision-making. Indeed, Gitelman and Jackson describe how "*data are mobilized"*, that is, they make things happen in the world, through graphs and charts (2013, p. 12, emphasis in original). Dashboards then, are an interface through which data become actionable through their visualisation. Although the design of visualisations relates to what the client wants to achieve from the data, Clarke suggested that the user determines the meaning of the data, when they said:

> Clarke: "*So I don't design a specific meaning into the visualisation. I design for the visualisation to show the data points or parameters that will allow them* [the user] *to make an assessment."*

Here Clarke explains that the meaning of the data is not determined by the design of the visualisation, but is left for the user to interpret. As already mentioned, I inferred from our conversation that Clarke was referring to their work on dashboard projects. In contrast to Clarke's view, Kitchen, Maalsen and McArdle argue that dashboards "actively produce meaning and do work in the world" (2016, p. 94). Yet, this can only happen through their interpretation, which is situated in particular social and cultural contexts. As this section will continue to argue, meaning does not emerge from data in isolation. The meaning of data is already embedded in the broader context of what the data means to the client and what the client wants to use it to do. Dashboards are a platform through which to communicate this meaning through visualisations. Furthermore, it is likely that how the user makes meaning from the visualisations will be framed by the context in which they are using the dashboard, particularly in the case of data driven decision-making. By listening to the designers' accounts of phase one, the way in which the meaning of data is embedded in the particular contexts of a visualisation project has been made visible.

4.2.2 Designing without data

Prior to undertaking fieldwork, I had assumed that the design process would always start with the dataset being visualised. While for some projects, the designers had access to full datasets that could be analysed and explored in search of *a story,* in other cases there was little or no access to the *real* dataset. This meant that sometimes designers began designing the concept of the visualisation project using "*dummy data."* That data visualisations could be designed without the dataset was surprising to me. Yet this scenario became apparent during my first scheduled observation with a visual designer called Turner. Turner was working on what they described as a small job, visualising some data for an existing client. Using some sample data, they explained they had been working on the design of a chart, using *bubbles* (circles) to represent the numbers. However, they explained that on inputting the *real* data they realised that the design would not work:

> *You can't see the differences between the size of the circles, they explain, as the numbers in the dataset are very similar. They had already realised this before I had come over to observe and had been thinking about a different style of graph. Perhaps a line chart, they say, "it's about time so it makes sense."*
>
> *Turner explains to me that sometimes an idea does not work with the real numbers and you have to go back to the sketch book. They say that their job is not just about making a graph, it is "about making sure it makes sense."*
>
> (Paraphrased field note from scheduled observation, 18. 10. 2017)

During the observation Turner demonstrated how, with little variance between the numbers in the *real* dataset, the bubbles appeared to be very similar in size. Thus making it difficult for a user to assess the difference between the numbers being visualised. Although in the end the data determined Turner's design, it seemed strange to me that they would begin sketching and prototyping design concepts without the dataset being visualised. The visual design of graphs and charts communicates meanings from data, however, I came to realise that designing how a visualisation would look without possession of the dataset was not unusual. This was particularly the case in dashboard projects where access to *real* datasets was less likely. During an interview with Bailey, they explained to me that despite not having the coding skills to analyse data they still considered themselves to be a visualisation designer, saying:

> Bailey: *"And the reason is because most of the time you go first by concept […] So we think about, ok, what do you want to visualise? What do you want to achieve? What can we do with it? And we know with the type of information we have, we don't know how it's structured or we don't know a lot about the numbers, but that's enough to have a sketch about what you can visualise and how."*

As this quote from Bailey illustrates, the concept for a visualisation does not always emerge from insights found through analysis of the data. Instead, it is dependent on a range of contextual factors particular to each project, and unearthed through discussions with clients during phase one. These factors relate to what the client wants to achieve through engaging visually with their data, but also *what* from the data *can* be visualised, which is an issue I will return to shortly. It seems that these questions can be answered without knowing a lot about the data or its structure. This is because it is the purpose of the visualisation and what can be achieved with the type of data being visualised that shapes how the concept develops. This strongly suggests that the meaning of data visualisations are not made from the data alone. Instead, meaning is made from data through what people want to do with it and how they want to use it to achieve certain aims. This frames and limits the meanings that are made from data. Therefore, the designers' discussions with clients and users in phase one should be considered as the first step in making meaning from data through visualisation. This determines the discourse within which the designers *decode* the data in order to make meaning from it, prior to its visualisation.

Before commencing the fieldwork, I had expected to observe designers working with large datasets and using data analysis software to analyse it. Despite observing visualisation design, after several days in the studio I had not seen any 'raw' data. During a meeting with Wright, I asked them, "*where is all the data?"* Wright laughed, before explaining that Evoke do not host the data on their servers and that if they did they would be an IT solutions business. They tell me that sometimes they access data through API (application program interface) feeds and other times they use "*dummy data"* or "*demo data"* which is stored in a file on the project's folder or in the company's *Dropbox* (cloud based storage). This demonstrates how the designers' access to data varies between projects while also revealing the broader apparatuses of the data assemblage (Kitchen, 2014), within which data and the design of its visualisation is embedded. Kitchen describes the "marketplace," as an apparatus of the data assemblage, which includes analysts who aid companies in generating value from their data (2014, p. 25). Data driven design agencies like Evoke are part of this market place, supported by infrastructure including IT solutions companies who store masses of data on behalf of their clients. As I will go on to discuss, these apparatuses intersect with others, including government and legal frameworks

(Kitchin, 2014) to produce regulatory constraints determining the designers' access to *real* datasets.

Still bemused at what I perceived to be a lack of data in the design of data visualisations, during our interview I pressed Bailey about how data fits into the design process, asking:

> Interviewer: "*I'm kind of struggling to understand when the data comes in. So, is there some data analysis that goes on in the first phase? To see what is in the data and"* [Bailey interrupts me before I finish my sentence].
>
> Bailey: *"That, there is a lot. Depending on the client, maybe. I think everybody here would agree that it would be beneficial to have this dataset as early as possible because at least you can start to play around with it a little bit. But most of the time this is confidential material, especially when it's for big clients. In those cases, they tend not to give* [us] *the real datasets, or they give* [us a] *very limited sample."*

This quote from Bailey suggests that there is a desire among the designers to have access to the data and to engage with it in the design process. However, visualising data is embedded in wider legal and regulatory frameworks surrounding the sharing of data with third parties as well as the commercial sensitivities associated with sharing confidential data. By highlighting how these apparatuses intersect to frame how meaning is made from data in the specific domain of data visualisation design, this chapter builds on Kitchin's (2014) notion of data assemblages and responds to his call to understand how they work in different contexts. In this example, the assemblage determines the extent to which the designers have access to the datasets they are visualising. If designers are working on visualising data without the dataset, then it must be assumed that the meaning of that data is, at least in part, made prior to the visualisation of the *real* dataset. Indeed, as this chapter will continue to explain, meanings from data are contextual and relate to social, economic, technical and regulatory constraints that emerge through data assemblages.

A lack of data might be perceived to present challenges in the design of data visualisations. However, in some contexts it may not be practical or possible to work with the *real dataset*, particularly if the project is working towards a dashboard that can process data in real-time. In such instances, the visualisations will continuously update and change as the most up to date data feeds into the dashboard. Bailey explained that when designing dashboards it was important to take into account the uncertainty that occurs when the designers are not working with a static database. When talking about not having access to the real dataset, Bailey says:

> Bailey: *"It's mostly fine when it comes to dashboards because we, as data visualisation designers, we should keep in consideration any possible scenario. So, little data, a lot of data, missing data. How do we visualise that in a way that conveys the right information and it doesn't mislead opinions or decision making, which is even more important."*

Bailey's quote reflects the argument made by analytics companies that visualisations can facilitate data driven decision-making (Laaksonn and Pääkkönen, 2020; Beer, 2019). It also reveals how, without access to the dataset being visualised, the design of the visualisations for dashboards is driven by how the data will be used to make decisions. Therefore, it is useful to conceptualise dashboards as a kind of data architecture, providing a structure that makes data perform in particular ways. As data feeds into the dashboard, its architecture forces it down different avenues, making some data visible through graphs and charts and concealing the rest behind the dashboard interface. This architecture is designed in relation to what insights the users want to obtain from the data and the data-driven decisions they intend to make, rather than the analysis of whole datasets. This is why Bailey explains that dashboards for dynamic datasets need to be designed to allow for uncertainty with regard to the data that will populate them. Therefore, in the context of dashboard design, designers make meaning from data in relation to how they can be used to generate insights and to inform decision-making. It is data's perceived ability to generate insights (Beer, 2019) that informs the discourses through which the designers decode the meaning of the data in relation to the context of their use.

4.2.3 Finding a story in the data

In some data visualisation projects engagement with and analysis of the dataset was very important. This was particularly the case in the design of stable data visualisations aimed at a public audience. By exploring the dataset, the designers found stories in the data that they could communicate through visualisation. Although I did not *observe* designers working on such projects during fieldwork, in interviews the designers spoke about the importance of finding a story in the data. During an interview with Bailey, they described to me a project in which data analysis was integral to visualising the dataset. This particular project was for a public audience and visualised data collected about wildlife. They said:

> Bailey: "*In that case the dataset was very important because to make the story work you need to work with the actual data, right from the start. So in that case, we got the data right from the start, the whole bunch of data, we analysed it and we tried to find a story in the dataset."*

How the data was made meaningful in this context is markedly different to the examples of dashboard projects. Meaning was made from the data visualised in dashboards through the context of their use and the goal of data driven decision-making. However, the data in this visualisation project was made meaningful through narrative. This required data analysis and human interpretation to actively look for a story in the wildlife data that could be told through visualisation. Another designer, Reed, told me that the project had been commissioned by the organisation who had collected the data and wanted to use it. They were also hoping to secure funding for their work. Having seen the visualisations they created for this project, the stories they told about the data through visualisation were emotive and powerful. There exists a plurality of insights and stories within datasets (D'Ignazio and Klein, 2016) and it is not unreasonable to assume the aims of their client will have shaped which stories from the dataset the designers told. Indeed, Weber, Engebretsen and Kennedy (2018) suggest that the search for narratives within data can be guided by a task or idea. In this case, the goal appears to have been public engagement, reflected through the emotive discourse the designers drew on to visualise the data. Therefore, the meanings made from this dataset were situated in the context of its use in the social world. They were also influenced by the subjectivities of the designers, as well as those who own the data being visualised. Both of whom exert power in shaping what insights from the data are revealed through visualisation.

During my interview with Clarke, they explained to me the importance of designers not becoming so enthusiastic about the dataset that they lose their focus on what users will take away from it. I took this to mean that instead of getting lost in the analysis of the data, they must undertake exploration of the data with a view to finding insights to visualise that users can make sense of and remember. This not only requires the designers to identify patterns in the data but to interpret meaning from those patterns. In doing so, they identify a story, or recognisable discourse, that will be meaningful to users when communicated through a graph or chart. Clarke went on to tell me that another designer, Jones, not only understands the data but also understands the need to come up with a "*narrative*." In describing how they work together, Clarke says:

> Clarke: *"[...] they make all these plots and we look at them, asking, how is this interesting? What are the follow up questions? I ask [Jones] questions and they ask me questions back and then we interrogate the data to see what's there."*

Here Clarke is describing a scenario where they have access to the dataset being visualised and are able to analyse the data in order to search for *interesting* insights. This collaborative process is indicative of the way designers worked together at Evoke to encode visualisations with

meaning. However, it is also an example of how human subjectivities are at the heart of the interpretation of data (Crawford, Miltner and Gray, 2014; Crawford, 2013). The way Clarke describes how they and Jones make meaning from data suggests that this involved judgements about what questions to ask of the data and which insights they deemed *interesting.* This contradicts the idea that data can speak for itself (Anderson, 2008) or that with vast amounts of data "human expertise" is less important (Rieder and Simon, 2016, p. 4). On the contrary, visualising datasets requires *active* human interpretation (Crawford, Miltner and Gray, 2014). Stories do not simply emerge from the data in isolation, they are sought out in a process that leads designers to prioritise which insights from the data are revealed through visualisation and which remain hidden. As this section will continue to explain, how the designers decide which stories in the data are important relates to the particular context of the project, the client's aims, and the intended user of the visualisation.

Existing literature on the storytelling qualities of data visualisations often focuses analysis on the image to identify the narrative techniques employed in their design (see for example, Weber, 2020; Hullman and Diakopoulos, 2011; and Segel and Heer, 2010). However, by focusing on the meaning making practices of the designers, this chapter develops an understanding of *how* narratives are formed from data, prior to its visualisation. Indeed, the need to find a story in the data kept coming up in the interviews I conducted with designers. In one interview, Jones said:

> Jones: "[…] *what I learned is that what you have to keep in your mind when you [are] searching for something is the story you want to tell. […] If you don't have a story to tell and you're just showing data vis, people will not be interested in it and they will not read it* […]"

Ma et al (2012) note how narrative techniques in data visualisations are often used to engage audiences. Similarly, Jones appears to correlate telling a story from the data with the users' engagement with the visualisation. Drawing on the *encoding/decoding* model, in finding a story the designers are decoding the data in order to translate it into a "meaningful discourse" that users will recognise (Hall, 1980, p. 130). Just as newspapers interpret social events for their readers (Steiner, 2016), the designers appear to be interpreting the data for stories that people will find engaging when visualised. The pursuit of an interesting story, or *meaningful discourse*, in the dataset will surely inform what insights from the data are revealed and what remains hidden. It also relies on the judgement of the designers who actively interpret the data and therefore determine which stories are told through visualisation. As Beer argues, those with the ability to "interpret or tell stories with data" hold a powerful position in data driven societies, because data become powerful through the stories they are used to tell (2019, p. 15).

Dourish and Gómez Cruz (2018) argue that narrating data is an important way through which people make sense of it. They suggest that people draw on "narrative elements" in order to construct stories which help them make sense of what the data means (2018, p. 4). However, the stories made from the data in this context are not only based on what the designers deem interesting but also their perception of what the users of that visualisation will find *engaging*. Therefore, when making meanings from data the designers encode data visualisations with their own subjectivities as well as the perceived subjectivities of the user. I will return to how the designers imagined users and how this shaped how meaning was encoded into data visualisations in Chapter five. However, in relation to how they decoded the data, Jones confirmed the user of the visualisation shapes the stories told from data through visualisation, when they said:

> Jones: *"I think actually that is the first thing you should know when you're picking a data vis, who are* [you] *talking to, because that really defines the story you want to tell* […]*"*

This quote from Jones illustrates how factors beyond the dataset, in this case the user of the visualisation, frames how meaning is made from the data prior to its visualisation. Their quote suggests that their perceptions of the intended user of the visualisation *defines* the stories that are told from the data. This relates to Clarke's claim that when exploring the data, designers need to think about what the user will take from the visualisation. The designers' understanding of the users of the visualisation is formed, in part, in the first phase of the project. This happens through discussions with their clients and sometimes the intended users but also the assumptions designers make about different users. However, my understanding is that user research was much more likely to take place in professional dashboard projects, where the users of the visualisations can be easily defined and accessed. This is in contrast to more stable data visualisation projects, which are often aimed at a general audience. Nevertheless, the way in which the designers appear to decode the data in order to translate it into a "meaningful discourse" (Hall, 1980, p. 130) demonstrates how the meaning of data is situated within the broader context of its use. It is also shaped by the "ways of seeing" (Crawford, Miltner and Grey, 2014, p. 1668) embodied by the relatively few who possess the abilities and expertise to interpret it (boyd and Crawford, 2012).

Despite the broader contextual factors framing how meaning is made from data, Jones argued the data *should* determine the stories told. During an interview I asked them if the story always comes out of the data or if it is sometimes driven by what the client wants to communicate using their data. They replied:

> Jones: *"So I would say that the data should drive the story. That's the correct way of doing things. You're using a dataset, you explore it, find something interesting, build a story about what you found and shape your data visualisation based on that."*

This quote demonstrates the nuance involved in making meaning from data prior to its final visualisation. Although the data limits *what* meanings can be made, the designer determines *how* they explore the data, which insights they decide are interesting, and how to build a story around those insights. In doing so, they are effectively decoding the data in order to translate it into a "meaningful discourse," that can be visualised and then decoded by the intended user of the visualisation (Hall, 1980, p.130). The insights that are considered *interesting* and the story that is built around them relates to the context of the project and how the designers perceive the intended users. Therefore, data visualisations are not only visual representations of a dataset. They are visual representations of the decisions made by designers (Hullman and Diakopoulos, 2011) and the contextual factors relating to the project, which determine how meaning is made from the data. Yet such conscious and unconscious decisions made by designers (D'Ignazio and Klein, 2016) are hidden behind the "aura of objectivity" that the conventions of data visualisation design work to produce (Kennedy at el., 2016a, p. 723).

To summarise this part of the chapter, meaning does not emerge from data in isolation, it is made meaningful in relation to the particular context of the project. This includes the aims of the client, the perceived needs of the user, and access to the dataset. In the case of professional dashboard applications, the designers *decode* the meaning of the data in relation to its use for data-driven decision-making. This explains how designers were able to design visualisations without access to the *real* datasets that would flow through the dashboard architecture. By drawing on what the client intended to do with the data, designers were able to draw on the context of the data's use to determine what it meant. Drawing on Hall, I argue that in doing so, they translated the data into a "meaningful discourse" by situating it in "frameworks of knowledge" that the intended user would understand (1980, p. 130). These frameworks of knowledge relate directly to how the data would be used and the decisions that would be made through the interpretation of its visualisation. However, when designing more stable visualisations for general audiences, the designers were more likely to explore the dataset in order to find an interesting story. In doing so, they used narratives to translate the data into a meaningful discourse that users could decode (Hall, 1980). In this case, while the story is limited by the data, it is also shaped by the aims of the project and the designers' perception of the intended user.

In the case of both dashboards and more stable data visualisation projects, the intended user framed how meaning was made and what insights were revealed from the data. In Chapter five, I will explain how the designers imagined the user and how this framed how meaning was *encoded* into the design of data visualisations. However, this chapter suggests that it is useful to think of encoding in relation to data visualisation in two stages. The first stage of encoding, discussed above, involves making meaning from the data prior to its visualisation. This requires the designers to *decode* the data, translating it into a meaningful discourse that the intended user can decode. The second part involves encoding these meanings into the *design* of visualisations, which the following section will go on to explore.

4.3 Encoding meaning in the design of data visualisations

The following part of this chapter explains how the designers at Evoke encoded meaning into the design of data visualisations. By this, I mean how they communicated the meaning of the data through the design of graphs and charts. It does so by drawing on fieldwork notes and interview data to analyse designers' working practices in the studio. It pays particular attention to how the designers worked collaboratively, as well as the importance of designers with different expertise working together. It argues that these collaborative practices shaped how meaning was encoded into the design of data visualisations. It argues that rather than being neutral representations of data, data visualisations are often the product of the ideas, knowledge and skills of a network of designers.

While conducting fieldwork at Evoke, I regularly observed the designers working collaboratively to design data visualisations. Sometimes this was during formal meetings between designers who were working on the same project but more often I observed this happening informally in the studio. During the course of the day, the designers would move around the open plan studio to stand or sit at one another's desks. In twos or sometimes threes, the designers would position themselves around a computer screen to discuss an element of the design. Therefore, rather than being the product of a lone designer, it was often the case that the design of data visualisations at Evoke involved a "socio-technical assemblage of actors and actants" (Kitchin, Maalsen, and McArdle, 2016, p.98). The actors in the studio included project managers and designers with different expertise. The actants included data, software (Kitchen, Maalsen, and McArdle, 2016), systems of knowledge, and even the studio space itself. This section of the chapter will explore how this assemblage framed how meaning was encoded into the design of data visualisations at Evoke. In doing so, it reveals how data visualisations are not straightforward translations of data. Instead, they are the product of interpretive acts framed by the socio-technical assemblages particular to their creation.

While sitting at a desk in the studio I observed two visual designers on the bank of desks opposite me. Bailey and Cooper were sat together at Cooper's desk looking at some graphs on the large screen in front of them. As I watched and listened, their attention turned to a particular graph as they tried to understand how one axis related to the other. They call out to Jones, who was sat opposite them, to come over. Jones walked over and stood behind the other two, all three of them were now looking at the graph on the screen. Bailey asked Jones how the scale of one of the axes related to the bar at the top of the graph. Jones explained that with the type of graph they were looking at you would not usually have "*these bars.*" Bailey said they wanted to show how gender distributes across age groups in the graph and Jones talked them through how to read the graph on the screen. "*I don't think we should use these bars*," they said, as you will not always see the difference, and Bailey agreed. Jones continued, "*what matters here is to know which has the majority, that is all you need to know here*." The three briefly talked some more before Jones returned to their desk. (Paraphrased from field note from general observation, 17. 10. 17).

As far as I could tell, the graph the designers were discussing related to a dashboard project, but I cannot be certain. Nevertheless, this brief account from my field notes illustrates how designers with different expertise worked together to encode meaning from the data into the design of graphs and charts. By which I mean, how the designers work to communicate the meaning from the data accurately through a visualisation. As the field note extract suggests, this involved selecting the most appropriate graph to communicate that meaning. However, what Jones said implied it also involves deciding what the user *needs to know* from the data to decode that meaning. This might involve, for example, making decisions about what data to make visible through visualisation to enable users to decode its meaning in the way the designers intended.

During an interview with Clarke, they explained to me that designers make decisions about which graph to use, which data to combine, and which variables to plot. However, they also said that not all designers are proficient in the use of 'r,' a piece of analysis software. In such cases, Clarke said, visual designers can work with designers who are skilled in data analysis. This certainly supports what I observed here and further demonstrates how encoding meaning involves an assemblage of *actors* - designers with different expertise - and *actants* - the studio space, data, software, and visualisation conventions. This relates to what Kitchin, Maalsen, and McArdle (2016) found in their research on the design and development of an urban dashboard. However, it also demonstrates how technology, in this case the computer screen and the software used to create the graph, is a focal point for collaboration between the designers in the studio.

During an interview with Clarke, they talked about how the designers worked together to create visualisations and the importance of the design studio in facilitating this. I asked them if the roles of visual designer, front end developer, and UX (user experience) designer are considered to be very distinct roles at Evoke. They explained to me that some designers might have more expertise in one area of the design processes than another, yet they often have backgrounds in elements of the design process other than the one they currently work in. For example, Clarke said that some of the front end developers at Evoke are trained visual designers and many of the visual designers possess coding skills too. In relation to how this plays out when designers work together on projects, they said:

> Clarke: *"Sometimes that makes people think 'argh I really want to change that,' or 'we should do it that way' but it's actually not their role within the projects at that point in time. Which in a lot of cases works out fine because it's a good suggestion or it's at the right moment, or everyone's in agreement, like, "hey – great suggestion – we'll change that or do that." And sometimes it doesn't work out, and it's like, "yeah ok, sorry but, you know, they're not going to change it. Thanks, but no.""*

This data extract reveals more about the collaboration that takes place between designers when encoding meaning into the design of data visualisations. This collaborative ethos allows designers to comment on aspects of a design that are not directly related to their role in a project. Although many of my other participants did see a clearer distinction between roles in the studio, Clarke's quote suggests that it is better to think about the designer's specialisms more loosely than their job titles suggest. Their quote also reveals how the design of visualisations emerge through discussions and compromise between the designers. How meaning is encoded into the design of data visualisations at Evoke is often not the product of one lone designer, but involves collaboration between designers with a range of expertise. Although many of the designers were proficient in more than one skill set in the data visualisation design process, the notion of expertise appeared to be important. When responding to questions I asked about how the designers use software in their work, Clarke said:

> Clarke: *"[…] you really need a set of experts […] I mean there are a few unicorns out there in the world that can do it all – that can do very good design, very good data analysis, very good programming."* […]
>
> *"So you need to focus on the communication between the people rather than finding one person that does it all."*

Here, Clarke suggests that not many designers are proficient in every element of data visualisation design. Of course, their statement needs to be understood in the particular context of the research site and the type of projects undertaken at Evoke. There are many data visualisation designers who predominantly work alone to analyse and visualise data and who are considered to be experts in their field. However, this is not how data visualisations were made at Evoke. Instead, designers with different expertise collaborated to make meaning from data and then encode those meanings into the design of visualisations. Therefore, rather than emerging in isolation, the meaning of data is encoded into visualisations within an assemblage of actors with different expertise. Clarke appeared to acknowledge that collaboration between these actors is important in the design process when they added that Evoke needed to "*focus on the communication between people.*" Indeed, my data suggests that collaboration between designers was central to how the meaning of data was encoded into the design of visualisations. This was often mediated by the use of technology, in the form of the various software the designers needed to be "*proficient"* in using to design and produce data visualisations. Clarke explained:

> Clarke: *"*[…] *I think that we operate at a professional level that requires a very high level of skill in order to be fast and proficient with the tools that we need. So you really need some multi skilled individuals as well as a group of people working together in order to create good stuff."*

Clarke does not elaborate on what they mean by "*good stuff*." However, in the context of their research on social media analytics, Laaksonn and Pääkkönen, state that, "Good visualizations are easily comprehended and effectively work as descriptions of the situation or, preferably, decision making devices for the customer" (2020, p. 101). This definition certainly resonates with the data driven application projects that Clarke said they tended to work on. Clarke's quote demonstrates the importance Evoke places on expertise and professionalism in their work. Just like data analytics companies, when providing a visualisation service to their clients, the designers at Evoke are selling their expertise (Laaksonen and Pääkkönen, 2020). Therefore, it is not difficult to see why Clarke might reinforce this in our interview. However, what is most interesting in this quote is Clarke's emphasis on designers with different skills working together. D'Ignazio and Klein (2020) note how data visualisation projects tend to involve a network of people with various skills, yet this human labour is rarely made visible. They suggest that "*data production studies"* provides a way in which to make visible the people and contexts through which data projects are produced (2020, p. 184, emphasis in original). Conducting fieldwork at Evoke provided an opportunity to focus on the production of data visualisations, revealing the network of people and the collaborative practices involved in encoding meaning into their design.

Design studios are associated with group working and collaboration and both are viewed as "critical to a design studio's creativity" (Vyas, van der Veer and Nijholt, 2013, p. 415). The role of the studio in facilitating this collaboration and creativity was reinforced by Clarke in response to a comment I made about the collaboration I had observed. Clarke responded by suggesting that the studio is one of the reasons Evoke delivers, "*the quality that we do,"* saying:

> Clarke: *"Often we get requests from clients, like, 'can you get a few people into my office and work on location?' And we always say no. Because, well, you see what's happening in that, that is how we create this, and that might be because of one person on the project dropping in, saying 'hey, if you looked at that,' and you don't know who's got the relevant knowledge."*

What Clarke says here suggests that it is not only the expertise of the designers that is important in the design of data visualisations but the studio space itself. Their quote reinforces the idea that the studio space is a facilitator in the collaboration between designers, allowing cooperation to unfold *organically*. Vyas, van der Veer and Nijholt argue that the physical space of design studios are arranged to "emphasize and stimulate communication, [and] collaboration" between designers (2013, p. 417). The open plan studio at Evoke and the glass fronted meeting rooms on the ground floor made it easy for the designers to see who was working where at any given moment. This allowed them to move around the room to talk to colleagues and to sit with one another to discuss designs on the large computer screens that dominated their desks. Indeed, Vyas, van der Veer and Nijholt state that collaborations between designers not only rely on verbal communication but also "visual and physical aspects" (2013, p. 415). At Evoke it would seem that the physical space of the design studio enabled the technology (both hardware and software) to become focal points for collaboration between designers as they moved around to meet at one another's screens. Therefore, the studio is an important component in the assemblage within which meaning is encoded into data visualisations.

Some of the discussions that took place between designers related to how the visual design of a graph or chart could be *translated* into code in order to produce a digital data visualisation. Communicating the concept of a design to front end developers who would build it was an important part of the design process and something that intrigued me. During interviews, I asked the designers about how this happened and when I asked Clarke they explained:

> Clarke: "[...] *ideally they are next to each other or at least in the same room. So when its like,* […] *'so we have this as a starting point and we were thinking something like that, or that,' and then the developer* [is] *going to bring his perspective, say, 'well*

> *actually,* [I'm] *not sure that I understand it,'* […] *or it's like, 'hey this is really easy, that's really difficult so shall we do it that way?' So in the ideal scenario they work together on those types of things."*

Here, Clarke describes an imagined conversation between a visual designer and a front end developer and in doing so highlights the importance of the designers' proximity to one another in the studio. Clarke's quote also demonstrates how designs evolve through discussions between designers as they bring their own expertise and perspectives to the visualisation. During another interview, Taylor, a visual designer, also stressed the importance of sitting next to a developer so that they could provide instant feedback as the visualisation is developed. However, this needs to be understood in the context of this particular fieldwork setting, which employs designers with a range of expertise. Nevertheless, it reveals the collaborative practices involved in determining the design of data visualisations at Evoke.

Hall (1980) suggests that the "professional code" (1980, see p. 136) frames the presentation and format of news media, yet Steiner (2016) argues that this aspect of encoding has been ignored in empirical research. As argued in Chapter one, it is possible to consider the conventions of data visualisation design as being the materialisation of the *professional code*. By making visible the human labour involved in encoding meaning into the design of data visualisations, this chapter challenges the notion of objectivity and neutrality that these conventions - part of the professional code - work to produce (Kennedy et al., 2016a). Instead, data visualisations are the product of compromise between designers and their tools, as I will discuss in more depth in the next part of this chapter.

The collaborative working practices undertaken at Evoke were supported by ways of working adopted in the studio. Weekly 'stand ups' on a Monday morning brought the designers together to discuss what they had worked on last week and what they would be working on this week. Furthermore, regular 'stand ups' within design teams working on particular projects ensured designers were kept up to date on one another's work. The studio played an important role in facilitating collaborative working practices, however, as well as the physical space there were also virtual spaces that supported this. During an observation with Cooper, they showed me how they updated '*Trello*,' a project management software that other designers working on the same project could access. This allowed the other designers to see what Cooper had been working on and what stages of the design they had completed. The office also used '*Slack*,' an internal messaging platform in which it is possible to create channels relating to different topics and even projects. Both Slack and Trello describe themselves as software that facilitates collaboration between teams. This further demonstrates the importance Evoke placed on

fostering a collaborative atmosphere in the studio and strongly suggests that the company viewed this as important in the design of data visualisations.

This part of the chapter has revealed the collaborative working practices that take place in the design of data visualisations. Rather than being the product of lone designers, data visualisation design at Evoke often involved a network of designers with different expertise. The designers brought their own perspectives and ideas to the design. They collaborated, compromised and worked within the limits of their tools, which shaped how meaning was encoded into data visualisations. Clarke's descriptions of the kind of discussions that take place between designers demonstrated how there exists a plurality of ways in which the meaning of data can be encoded into graphs and charts. Therefore, *how* meaning is encoded into the design of data visualisations is dependent on the skills and ideas of a network of designers who work together to determine their final design. Yet, their labour and subjectivities are concealed by the conventions of data visualisation that make them appear neutral (Kennedy et al., 2016a) and their design *naturalized*, as though it could never have been another way (Akrich, 1994).

The following section introduces the concept of *back and forth* to explain how software becomes a focal point in how meaning is encoded into the design and production of data visualisations. It also reveals how visual designs are translated into code through collaborative processes, involving discussion, negotiation, and compromise. This socio-technical process is shaped by the context of each project and the limitations of the software used to build digital data visualisations.

4.4 "*Back and forth*" in the production of data visualisations

This section of the chapter explores how the contextual factors relating to a visualisation project and the collaboration between designers combine when visual designs are *"translated"* into code. Coding is an important part of the production of digital data visualisations, when interactive and/or real time data visualisations and dashboards are built by front end developers. However, D'Ignazio and Klein (2020) describe how this kind of work is often invisible. In cases where the graph or chart remains a static or printed design, the development phase is not necessary. However, Evoke are known for their interactive and data driven applications and therefore development was a significant element within their design practices.

As discussed earlier in this chapter, visual designers at Evoke often start with a concept that leads to the visual design. The visual design of digital and interactive graphs and charts is then built by front end developers who code the design in html, css, and java script. However, my

fieldwork revealed how translating the visual image of a graph or chart into code was not always a straightforward process. Instead, it often required compromises between the visual designers and the developers to determine the final design. Such compromises were informed by the specific constraints of the visualisation project. This section of the chapter borrows a term that many of the designers used to describe this collaborative process, *"back and forth,"* to explain how data visualisations were produced. It refers to the way meanings are encoded into visualisations through *back and forth* discussions between visual designers and front end developers, through which the final design is determined.

Turner introduced me to the idea of *translating* a visual design into code and was the first of my participants to use the term "*back and forth"* during an interview. While discussing their use of software they described how when designing for the web, they make the visual design before it is "*translated into code*." Intrigued by the idea of the development process being a translation of the visual design into code, I asked:

> *Interviewer: "So, how does that translation bit work? So you've got your static image?" ["Yeah"] "And then you give it to the front end developers?" ["Yeah"] "So do you just hand it over?"*
>
> *Turner: "Sometimes, but that's not the idea. Like you sit together and discuss it and the front end team here are also half designers so they also influence the design and it's like a collaboration. They say, 'you have this graph design but that's really hard, maybe we can do it a bit like this or we have this plug in that makes it easier' and we go back and forth and define the final thing together."*

Turner described how they sit alongside the front end developers who will build the design in code to "*define*" the final design of the graph or chart together. During the interview Turner had explained that the initial visual design is made by the visual design team. However, their quote reveals how collaboration and compromise are involved in subsequent iterations of the design. When Turner says the developers might describe the graph design as "*hard*," I understood that to mean difficult to build in code. This suggests that the back and forth discussions between front end developers and visual designers are shaped by the software they are working with to produce it. How meaning is encoded into visualisations does not simply emerge from the data, as though merely revealing its structure. Nor are the visualisations that communicate the meaning of the data necessarily the ideal representations of that data from a visual design perspective. Indeed, Turner's quote suggests that how meaning is encoded into visualisations is not purely down to the visual design of the graph or chart but what it is feasible to code within

the confines of a project. As the chapter will go on to explain, how meaning was encoded into data visualisations was framed by the technical, temporal, and financial constraints of each project. However, first it will demonstrate how the *back and forth* process played out in a meeting I observed between visual designers and front end developers.

During fieldwork, I observed a meeting between two visual designers and two front end developers who were working on a responsive table. The table included a large amount of data and although it worked well on a desktop computer screen, it did not work well on a mobile device. The team of designers had gathered in one of the meeting rooms on the ground floor. Bailey, one of the visual designers, began the meeting by describing the table to the group. Bailey explained that the table had seven columns and described the first, which was a list of countries, as being the most important. They also said that at least one of the columns showed *ranking*. As a possible iteration to the design, Bailey suggested that the first column could appear in a list form with a drop down to show the information for that particular row. Addressing the rest of the designers, Bailey said that they would like to know if their idea is *"do-able…feasible…or* [if there is] *another way to access the table?"* (Paraphrased field note from scheduled observation, 19. 10. 2017)

Although this meeting was about a table not a graph or chart, it is a good example of how designers with different expertise collaborate in the production of data driven applications, including data visualisations. Furthermore, tables can be considered visual representations of data and like data visualisations, they "communicate precise information and values" (Kennedy et al., 2016a, p. 716). As my account of this meeting unfolds, it will demonstrate how designers worked together to translate designs into code and how this required compromise between the different actors. It is notable that Bailey wanted to use the meeting to check if their idea to amend the table design was *feasible*. By this, I understood they wanted to know if it could be produced, and work, in code. This highlights how encoding meaning was not just about the visual image but also the practicalities of producing data visualisations within the constraints of the software the designers worked with. This is why collaboration between visual designers and front end developers was so important in determining the final design of a visualisation.

Wilson, a front end developer, plugged a cable into their laptop which projected the screen of their Macbook onto a large Philips monitor positioned on the wall of the meeting room. As the designers looked at the wall-mounted screen, Wilson showed them a page of code to explain how tables are structured in html. They then cut back to the image of the table as it would currently appear when viewed by the user on a desktop computer. They flicked back and forth between the code and the table to show the design team how elements of the code made the

table look a particular way. Wilson then demonstrated how, when the user tried to resize the table for a mobile device, it would only resize to a point before the columns started to disappear. This, it seemed, was the problem they had met to solve. Wilson started to go through possible solutions by showing them a website that included a number of different types of table. From these options, Wilson suggested a *stack* table design, however Bailey explained that this was not a good option as, *"you can't compare horizontally."* By this, they meant that the user could not compare data from different countries and Bailey said *"that makes it very difficult."* Wilson returned to the different types of table, disregarding another option as the table contained too much data. (Paraphrased field note from scheduled observation, 19. 10. 2017)

Wilson's demonstration of how the code related to the image of the table hints at the imperfect translation between image and code. How the image is built using code not only affects the look and feel of the table but how it behaves when the user attempts to resize it. Therefore, what is "*feasible*" is determined by the constraints of the technology and the solutions the designers reach through collaboration and compromise. This requires designers with different expertise to work together to debate the pros and cons of possible solutions. As Wilson and Bailey's discussion of the stack table demonstrated, this can involve a trade-off between design and functionality. Interestingly, Beer describes how the data imaginary is "caught between a back and forth between vision and practice" (2019, p.18). By this, he is referring to the range of possibilities that are ascribed to data and how these relate to what it is possible to do with data in practice. The concept of *back and forth* I have introduced in this chapter emerged through the discussions I had with the designers at Evoke. Nevertheless, there are parallels between what Beer says about the data imaginary and the design of data visualisations. Namely, the discord between the 'ideal' visual design and the compromises required to that design in the production of digital data visualisations.

Over the course of the meeting the designers discussed different possible solutions to the table problem. The visual designers, Bailey and Cooper, started to draw ideas on the wipe-able wall of the meeting room with marker pens, while Wilson explained the complexities of making their ideas work. Seemingly frustrated, Cooper asked Wilson, "*can we make our own tables?*" Wilson reply implied that this would affect how people could use the tables. Cooper conceded that the user would not be able to copy and paste the content from the table anymore but that they could download the information into a csv file. As I listened to Cooper saying this, I remembered Bailey mentioning at the beginning of the meeting how much they liked that the user could copy and paste content across the different columns, from left to right. Bailey had obviously regarded this functionality as useful. As the discussions continued, Robins, another developer, entered the meeting room and reminded the team that they probably would not find a

solution to the table today. They suggested that the designers go back to do some research and try some different options. They explain that it is worth putting the time and effort in to finding the *right* solution. (Paraphrased field note from scheduled observation, 19. 10. 2017)

In the case of digital interactive graphs and dashboard applications, visual design decisions must be translated into code. Therefore, their ultimate design is shaped by what it is *feasible* to code. The account of the meeting illustrates how the visual design team wanted to know what it was possible to achieve in the development phase. While Wilson tried to work out how they could translate the designers' ideas into code, taking into account the constraints within html, css and javascript. Perhaps inevitably this collaborative process was not without resistance from either side. The visual designers and front end developers in this account were working together, yet at times they seemed frustrated that the designers' ideas did not work for the developers, and vice a versa. Kitchin, Maalsen, and McArdle describe the relations between the actors involved in the design and build of an urban dashboard, arguing that:

> *"solutions are enacted through the performativity of actors who have varying subjectivities, personalities, knowledges and agendas, and are working together within specific social and institutional relations and settings, and are shaped by the capacities of other actants."* (2016, p. 99).

Although the context of this study is different to Kitchin, Maalsen, and McArdle's (2016) research, their argument relates to my own findings in relation to the collaboration between the designers in the studio. They too were bringing their own perspectives, expertise and subjectivities to design through the performance of their roles as visual designers or front end developers. Although, as Clarke alluded to, these roles were often more fluid than their job titles might imply. D'Ignazio and Klein suggest that by directing attention to the "subject positions" of designers it is possible to uncover the conscious and unconscious decision-making that frames the design of data visualisations (2016, 3.2). At Evoke, these subject positions and design decisions became visible during the *back and forth* process, through which the designers negotiated the final design of a visualisation.

The meeting between the designers revealed how the design of a visualisation moves back and forth between the visual designer and the front end developer. This required designers with different expertise and knowledge to discuss, debate and compromise in order to translate the visual design into code. In the meeting there appeared to be a desire to strike a balance between the look and feel of the table, its functionality, and the practicalities involved in building it in code. Bailey's concern about the copy and paste function of the table implies that these

compromises relate to earlier phases of projects, during which the designers determine what the data means in relation to how it will be used. This demonstrates how encoding meaning into the design of digital data visualisations is not just about the meaning of the data, or the design of the visual image but how these things are translated, or (en)coded, in the production of the final product. This appears to be an imperfect art; constrained by the technological affordances of the coding software and determined by the subjectivities of the designers who work together to determine the final design. While Hullman and Diakopoulos (2011) recognise that data visualisations are a product of the choices of designers and framed by the intended user, they do not take into account the way in which the use of software shapes these decisions. Nor do they factor in the particular context of the visualisation project and the various ways this may shape the final design. How the context of a project influences the back and forth process is something this chapter will return to shortly.

The meaning of the data was established in the first phases of a project. Drawing on the *back and forth* it is possible to see how that meaning was then encoded into the design of data visualisations and translated into code. Collaboration and compromise between the designers was integral to how this happened, particularly between visual designers and front end developers. Their discussions shaped how meaning was encoded into the design and production of data visualisations. During an interview with Bailey, I asked them how they explained a visual design to the front end developer. They told me how they try to involve development before the design is finalised, in part to establish what is feasible. They went on to say:

> Bailey: *"But that's also why, like, you really have to show the graph, printed maybe even. And say, ok, the size of the bubble it's this data, the colour it's something else."*

Bailey's quote suggests that they communicate the meaning of a visualisation to the developer by *showing* them the visual image of the graph and explaining what the different visual elements relate to. In this case, Bailey highlights the size of the bubble and the colour as being important visual components that convey the meaning of the graph. This suggests that to translate the visual design of a graph or chart into code, the developer requires an understanding of how the various visual elements relate to its meaning. If the developer directly *translates* the visual image into code, why would they need to know the graph or chart's meaning? This is necessary because encoding does not stop at the visual design but continues into the development phase where the visual components are built in code. This explains why the developer requires an understanding of the meaning of the graph to ensure that any compromises required during this phase do not detrimentally affect its intended meaning.

Wilson, a front end developer described how they needed to understand the "*reasoning behind*" the design in order for the visual designers to communicate it to them:

> Interviewer: "*Ok. So when they're explaining their design to you, it's not just that they* [the visual designers] *give you a sheet of paper with the design on and walk away?* [*"No"*]*. You have to talk about it?"*
>
> Wilson: "*Yeah, yeah, like I wouldn't be able to understand it. Yeah you have to hear that train of thought, the reasoning behind it, the rationale, down to the details of why this button looks like this. Why this arrow is green."*

Just as Bailey described, Wilson explains that they need to know the meaning behind the visual components of the design in order to be able to understand it. This is contradictory to the notion that visualisations make data easily accessible or that they allow non-experts to gain insights through analytics (Beer, 2019). Wilson is an expert, in that they are a professional data visualisation designer. Therefore, the need to understand the meaning of the graph or chart implies that it is important to understand the context in which the data visualisation will be decoded. I will return to the significance of context in relation to decoding graphs and charts in Chapter six. However, Wilson's need to understand the graph is because at the development stage, the visual design of the graph or chart is not fixed. As the earlier accounts demonstrated, the technical constraints of the software can still shape the design as the designers grapple with how to translate it into code.

During my interview with Bailey, they went on to explain how the design can change through discussions with the developers. Again, they relate this to what is "*feasible"* within the confines of the software they are using. Feasibility appeared to be a catalyst for the *back and forth* between the visual designers and developers. In this sense, the design moves back and forth between the designers and between the limits of feasibility. The way Bailey describes this contradicts their own view of the software in relation to how it influenced their work. During interviews, when I asked designers about how the software they used shaped their work, they repeatedly referred to it as merely a *"tool"* that relied on their own ideas and creativity. The designers, including Bailey, tended to talk about its limitations in terms of their own technical skills rather than limitations built into the software. However, as the following quote demonstrates, the constraints of the technology became a focal point for the *back and forth* between designers. In response to what Bailey had told me about showing the developer an image of the graph in order to explain its meaning, I replied by saying that I expected this gave the developer an opportunity to say if they did not think the design would work. Bailey replied:

> Bailey: *"Yeah. But in my experience usually that comes not from* [a] *design perspective, it's often sort of like a technical feasibility thing. Which is an interesting comment because that means that you kind of have a sort of like give and take."*

Although, Bailey describes the *back and forth* conversations between the designers as "*give and take*," they are referring to the same thing. These are the discussions that take place between visual designers and front end developers in order to translate the visual image into code. Bailey's use of the term "*give and take"* indicates the decisions and compromises made between these actors to negotiate the technical constraints of coding the design. This is done while trying to maintain the integrity of the visual image, which has been designed to communicate the meaning of the data. Therefore, the final digital data visualisation is the product of this *back and forth* process that the designers are engaged in in the production of data visualisations. D'Ignazio and Klein (2016) argue that the labour involved in data visualisation design should be made visible in order to credit those involved. However, crediting the work of the designers may also reveal the various ways in which the subjectivities of designers frame the design of data visualisations. During our interview, Bailey highlighted the different starting positions and priorities of the visual designer and the developer when negotiating the production of a graph or chart during the development phase:

> Bailey: *"The designer wants to push to "ok this was the best design,* [it] *has been thought out for long enough so this is how it should look" and the developer says," yes but this is like 3 times the amount of work* […] *for just this thing." […] In that case you have a discussion about* [if there] *is a way we can make the design easier to implement without harming the user or the visualisation. Sometimes it's possible. Sometimes it's not and that's when we decide to invest more budget in that specific component or thing."*

Although there are different stages in the design process, the *back and forth* reveals how the design of data visualisations is not linear. The quote above highlights how designs can be contested and altered in response to the work involved in their production. This happens through discussions between the visual designer and the front end developer about what design elements can or cannot be translated into code within the confines of the project. Here, Bailey refers to the budget of the project. However, during an interview with Taylor, they explained how designs sometimes do not work in development because of time constraints, saying:

> Taylor: "*And sometimes you have a really cool idea and then they* [the developers] *say, "well that's not going to work, because it's going to take us 5 hours and we have a week for everything.""*

Time and budgetary constraints are of course interlinked. These two quotes demonstrate how the feasibility of what can or cannot be built in code is not always down to the limitations of the software but due to the particular constraints of a project. Indeed, the technical constraints related to the production of particular designs do not emerge in isolation (Kitchin, Maalsen and McArdle, 2016) but within the wider context of the project. Bailey and Taylor's quotes indicate the different factors the designers have to juggle in negotiating the translation of the visual design of a graph or chart into code. This chapter has revealed how these factors include the visual design, the user, the client, the limitations of technology and 'technical feasibility.' The latter of which brings together the use of software to code designs and the financial and time constraints of a project. Addressing the feasibility of the design requires the designers to work together, drawing on their various subjectivities, expertise and knowledge to make design decisions together that determine the final design of the data visualisation. Research that addresses the meaning of data visualisations through the image (see Kennedy et al., 2016a; Hullman and Diakopoulos, 2011) not only fail to consider the role data plays in determining their design, but neglects to consider how the particular contexts of visualisation projects also shape them. By paying attention to the practices of designers, this chapter has demonstrated how design decisions are not formulaic but are determined by a range of contextual factors as well as the subjectivities of those who work together to make them.

As Bailey's quote revealed, while a designer may push for their visual design, the time or cost involved in building it may require compromise. This might be in terms of the visual design, the functionality or the budget. Furthermore, as the meeting about the responsive table suggested, the need for compromise might also relate to the limitations built into the type of browser used to develop the visualisation. Therefore, the translation of the design into code involves collaboration, negotiation and compromise between the visual designers and front end developers. As this section has set out, this takes place through *back and forth* discussions between the different actors, which are mediated by various actants. It is through this socio-technical assemblage that meaning is encoded into the design and production of data visualisations. Drawing on Kitchin (2014), the assemblage includes the clients who commission visualisation projects, finance, regulatory frameworks, the designers, their subjectivities, ideas and expertise, the studio, and the software used to design and produce data visualisations. By focusing on the designers' practices, this assemblage becomes visible, particularly through the *back and forth* process in the production of digital data visualisations.

Although the visual designers design a visualisation so that it communicates the meaning of the data to the user, fieldwork revealed how the design is subject to change during the development phase. Therefore, the digital data visualisations built at Evoke were not necessarily 'perfect' representations of data, at least from a visual design perspective. Rather, they were often a product of the *back and forth* process through which designers drew on their expertise and ideas to collaborate, negotiate and compromise to reach the final design. This was with the aim of building digital data visualisations within the confines of individual projects and the limitations of their software. Therefore, rather than viewing data visualisations as neutral representations of data, they must be understood as a product of compromise, reached within a socio-technical assemblage involving people, practices, and "technical fixes" (Kitchin, Maalsen and McArdle, 2016, p. 98). Indeed, encoding meaning into the design of data visualisations is the materialisation of design decisions made by a network of people who draw on their expertise to work within the particular confines of visualisation projects.

4.5 Conclusion

Drawing on Hall's (1980) model of *encoding/decoding*, this chapter has discussed how meaning is encoded into the design of data visualisations in three sections. It has done so by focusing on meaning making in the discourse, design and production practices (Kress and van Leeuwen, 2001) of data visualisation designers. By bringing Hall's (1980) concept into the domain of data visualisation design, it has built on his model to argue that in this context it is useful to think of encoding in two stages. Firstly, the designers make meanings from the data. They do this by *decoding* the data in order to translate it into *meaningful discourses* (Hall, 1980) that can be recognised and decoded by users. Secondly, the designers encode this meaning through the design and production of data visualisations. This involves collaboration between designers with different expertise and, in the design of digital data visualisations, a *back and forth* process through which visual designers and front end developers work together to translate the visual design into code. This suggests that there are various stages of encoding and decoding in the interpretation of data, the design of its visual representation, and the production of data visualisations.

How designers made meaning from the data in the first stage of encoding was determined by the context of the project and the level of access they had to the dataset being visualised. In the case of visualisations designed for professional dashboard applications, there was often only limited access to the data. Therefore, they drew on how the data would be used, often for data driven decision-making, in order to situate the data in ways the users would recognise and could decode. In the case of visualisation projects where the dataset was more stable, the designers

explored the data to find a story that the intended users would find interesting, translating it into a "meaningful discourse" (Hall, 1980, p. 130) for the intended user to decode. In both instances, this required the designers to consider the aims of the project and the intended user of the visualisation. Therefore, the meaning of data within visualisations is already framed by the intended *use* and intended *user*, prior to its visualisation. While making meaning from data is the first step in encoding data visualisations with meaning, the second stage of encoding occurred in the visual design and production of data visualisations. It was during these moments that meanings were encoded into the design of the final visualisation. At Evoke, this involved collaboration and compromise between designers with different expertise who brought their own perspectives and ideas to the design process. However, the findings of this chapter must be understood in the context of the fieldwork site. There are many designers who work alone to analyse and visualise data, therefore, more research is required to understand how the two stages of encoding relate to the working practices of lone designers.

In analysing the production of data visualisations, this chapter introduced the concept of *back and forth* to explain how the design of a visualisation moves between the visual designers and front end developers in the production of digital data visualisations. This *back and forth* process centered around the technical feasibility of translating a visual design into code. This brought together the limitations of the software used to build visualisations and the contextual factors and constraints that were particular to individual projects. This required the designers to work collaboratively and to make compromises that shaped how meaning was encoded into the design of the visualisation. For example, the most appropriate representation of the data from a visual perspective might not be technically *feasible,* in relation to the limitations of the technology or the temporal or financial constraints of a project. Therefore, the final graph or chart is often a compromise between visual design and technical feasibility.

The following chapter draws on fieldwork data to explain how the designers imagined the users of data visualisations and how this shaped the way that they encoded meaning into their designs.

Chapter 5: Designing visualisations for imaginary figures

5.1 Introduction

Hall (1980) argues that encoding requires producers to make certain assumptions about their audience. Drawing on field notes and interview data, **t**his chapter explores how the designers at Evoke *imagined* the intended users of data visualisations and how they drew on these imaginary users to encode meaning into the design of graphs and charts. During fieldwork, I listened to how the designers spoke about the users who would be engaging with their designs. This revealed how they imagined different types of user and how the assumptions they made about those users determined the design choices they made. These included assumptions about what different types of users *needed* and *wanted* when engaging with a graph or chart, as well as their ability to engage with and make sense of a visualisation. These assumptions informed their practices, as the designers made design choices based on the perceived needs or desires of their imagined intended user. How designers of technological artefacts perceive the user and the context in which the technology will be used are an important part of the design and development process (Akrich, 1994). As discussed in the previous chapter, designers talked to clients and sometimes the intended users of a visualisation in the early stages of a project. However, Oudshoorn, Rommes and Stienstra argue that to understand how designers configure users it is necessary to look beyond user research to how users are *imagined* by designers (2004, see p. 31). This chapter builds on their argument by introducing the concept of *imaginary figures* to explain how the designers personified different types of imagined user in three homogenous categories. These imaginary figures shaped how meaning was encoded into the design of data visualisations.

Woolgar explains that *configuring user*s involves both defining and *constructing* the identity of the user and entails "setting constraints upon their likely future actions" (1991, p. 59). These restrictions "define, enable and constrain" the users' actions, thus configuring the user in relation to how they interact with the machine (Woolgar, 1991, p. 69). This affects their actions both in the present and into the future, as designers imagine and shape their future needs (Woolgar, 1991). Woolgar was writing about the design and production of machines, specifically microcomputers. However, his definition of the configuration process seems particularly relevant to how the designers at Evoke talked about imaginary users in terms of how they would interact with visualisations. The designers appeared to be drawing on shared assumptions about different types of users and their varied abilities to engage with data visualisations. Therefore, designers often talked about what different users *liked*, *wanted*, or

needed in order to engage with a graph or chart, making design decisions that reflected these perceptions. This is significant in understanding how designers encoded meaning into visualisations, because these design choices shape the visual elements through which the chart or graph communicates meaning. Analysing how the designers imagined users revealed that the assumptions the designers made about different types of users were personified in three distinct imaginary figures; *the grandma, the public*, and *the professional*. I will briefly introduce these figures before explaining them in more detail as the chapter unfolds. In doing so, I will pay particular attention to how they framed how meaning was encoded into the design of data visualisations.

The grandma

The grandma personified those users who the designers imagined possessed the lowest levels of ability in relation to engaging with and making sense of data visualisations. This category of user comprised of elderly users who were discussed in relation to age related needs and the accessibility of a design. However, the numerous references designers made to grandmothers and even mothers, revealed a gendered element to this imaginary figure. Older women were imagined to possess the lowest levels of ability in relation to engaging with graphs and charts, therefore they were talked about in relation to their *needs*. So much so, *the grandma* was the benchmark for accessibility.

The public

The public personified those users who were imagined to represent the general public, or 'everybody else.' Unlike the grandma, neither age nor gender were explicitly associated with this user. The designers imagined this user to encounter data visualisations in newspapers for example, yet they were not expected to be experts at interpreting them. As this user was not necessarily motivated to engage with the visualisations they encountered, the designers' priority was to encourage engagement with the graph or chart. Therefore, designers tended to talk about this imaginary figure in relation to what they would *like*.

The professional

The professional personified those users who the designers imagined engaged with data visualisations at work, probably through a dashboard interface or other form of data driven application. Like the public, this imaginary figure was not gendered and there was no *explicit* reference to their age. The designers imagined this user as being competent in both their

technical ability and their capacity for making meaning from visualisations. As these users were already motivated to engage with the graphs and charts the designers produced, they talked about this user in relation to what they would *want* when engaging with data visualisations.

5.1.1 Introducing imaginary figures

It is important to make clear that imaginary figures were not a physical presence in the design studio. Instead, they were "discursively constructed" (Clarke, 2005, p. 46) through how the designers talked about the intended users associated with different design projects. It is also worth noting how imaginary figures are distinct from broader definitions of user representations. User representations refer to identities that designers create for their users, which emerge during the design process (Rommes, 2002, see p. 44). Akrich (1995) considers the construction of user representations to be an integral part of the design process. Indeed, she states that a design's success relies on the designer's ability to develop user representations and assimilate them into the product or artefact through the design choices they make (Akrich, 1995). However, rather than being formed in relation to particular design projects, the imaginary figures identified in the fieldwork data transcended different projects. Although it is possible the designers imagined these imaginary figures slightly differently to one another, they were a constant, taken for granted resource from which designers could draw on to inform and explain their design decisions.

In order for *imaginary figures* to be a useful concept in understanding how designers encode meaning into data visualisations, it is important to consider how they are similar to, yet different from, other forms of user representation including personas. Personas are fictional accounts of individual users who stand in for the unknown user during the design process (Turner and Turner, 2011). These fictional users are formed through descriptions of varying depth and rigour, which allow designers to feel a connection to the user (see Floyd, Jones, and Twidale's typography of personas, 2008). Individual personas tend to include the following characteristics; name, age, gender, interests, occupation, and sometimes socio-cultural categories such as ethnicity, level of education, and socio-economic position (Turner and Turner, 2011, see p. 34). This method of 'knowing the user' was originally intended to generate personas which were grounded in thorough ethnographic research (Turner and Turner, 2011; Floyd, Jones and Twidale 2008). However, Floyd, Jones, and Twidale (2008) note how personas are not always grounded in empirical research and are sometimes formed from a designer's own intuition and design experience. Furthermore, they concede that designers may also draw on stereotypes about different users to construct their personas (Floyd, Jones and Twidale, 2008).

Although there are similarities between personas and imaginary figures, they differ in several ways. Imaginary figures are not individual characters but are homogenous categories of nameless users who are less clearly defined than personas. Although the designers drew on imaginary figures to make design choices, there were no physical trace of them in the studio in the form of written accounts or definitions. Furthermore, despite imaginary figures being a discursive presence, their role in shaping the design of data visualisations appeared to be taken for granted among designers. When designing for a persona it is reasonable to assume that a designer would possess an awareness of how that very particular view of the user would shape the design. However, imaginary figures were formed through assumptions the designers made about different types of user which, unchecked, might perpetuate the very stereotypes they may be drawing on to make these assumptions. Although imaginary figures are different to personas in these important ways, they share the same "generative power" in how they shaped design decisions to meet the requirements of the imagined intended user (Floyd, Jones and Twidale, 2008, p. 14).

Imaginary figures is a concept that is grounded in empirical research and was formed specifically in relation to the design of data visualisations at Evoke. It makes visible the assumptions designers made about different types of user and explains how these assumptions shaped how they encoded meaning into the design of data visualisations. This chapter introduces the concept through the discussion of the three imaginary figures identified in the fieldwork data. It explains how *imaginary figures* is useful when thinking about how designers encode meaning into data visualisations for different users. The first section of the chapter will briefly discuss the conversations I had with designers about user research to provide some context in terms of what the designers might be drawing on when imagining the user. The chapter will go on to discuss each imaginary figure, revealing the assumptions the designers made about them and how this shaped how meaning was encoded into the design of data visualisations. This chapter will also touch on how the imaginary figure of the grandma represented the designers' understanding of diversity among users.

5.2 Configuring the user: User research

During fieldwork, I did not observe user testing or user research taking place. Therefore, I am drawing on the conversations I had with the designers and what they told me about these practices to explore how they may have influenced their ideas about different users. Analysis of fieldwork data identified that the designers spoke about three forms of user research. I have termed these, *project specific user research*, *convenience user testing*, and *colleague validation*. This is not to say that other forms of user testing and research did not take place at Evoke.

However, the following discussion of the user research designers talked about provides important contextual information, which offers an explanation as to how shared assumptions about different types of users were formed within the studio.

During an interview with Turner they talked about how they would approach the design of a visualisation for a "*very old person*" with little experience of computers. I asked how they knew to design differently for this type of user. They said that sometimes there is user research to draw on but not always. I went on to ask them:

> Me: *"So is that* [user research] *built into the projects? A certain amount of time spent talking to users about how they're going to.."*
>
> Turner: *"Yeah. Yeah, that's all in the first phase. Sometimes we do workshops with users and talk to them about* [the] *subject, get to know what they want out of it, what their end goals are."*

User research was not built into every project and Turner implied that it tended to be part of the bigger projects Evoke undertakes. Therefore, it seems reasonable to assume that conducting user research depends on the scale of the project and the desire of the client to budget for the costs involved. Nevertheless, Turner explained that this kind of user research tended to happen in the first phase of a project and that it sometimes involved talking to the intended users about what they want to achieve through visualising their data. Turner went on to explain that sometimes they create personas from this user research, prompting me to ask if creating personas was not only about *who* the end user would be but also about imagining *how* they will use the visualisation. Turner agreed but said that without knowing what you are going to make, that can be hard. So those working on the project might also ask about how the user is currently using "*similar things."* This suggests that developing an understanding of how intended users are already using visualisation tools allows the designers to imagine how they will engage with the visualisations they are designing for them.

Akrich (1995) found that one way in which designers invoked representations of users in the design of new products was to compare them to the user representations elicited by existing ones. Akrich (1995) describes this as an *implicit technique* through which an understanding of the user is formed by focusing on existing products rather than talking to actual users. However, Turner's quote suggests the designers do have some contact with the user or the client to determine how they use existing data driven tools. Nevertheless, asking how users are working with similar products to inform the design of visualisations implies that existing technology

plays a role in configuring the user and shaping their potential "future actions" (Woolgar, 1991, p. 59).

Turner's account implied that this kind of user research applies to the products Evoke designs for professional settings, where it would be possible to access the intended users fairly easily and where those users might already be using similar data driven applications. This view was reinforced during my interview with Clarke. Clarke had just explained to me that they tended to work on the larger scale projects at Evoke, which involved designing applications that "*deal with data*," such as dashboards, and that these are usually "*professional applications.*" I took this to mean that these were data driven products designed for business or organisational settings. As quoted in Chapter four, Clarke went on to explain how they are involved in interviews with potential users and the client to find out what they are interested in, what they want to s*ee*, and the decisions they want to make from their data. This, Clarke suggested, allowed them to "*choose the right visualisation that will help them*" (to see the full quote refer to Chapter four, p. 96).

This echoes what Turner said about user research during the first phase of projects and reveals the designers' desire to understand what the user wants to achieve by visualising their data. As Clarke implied, in the case of dashboards it is likely to be the goal of revealing insights from the data to inform data driven decisions. Interestingly, both Turner and Clarke regard the user as knowing what they *want* from a visualisation tool. The user seems to be afforded a level of expertise in the process because it is they who know what they are "*looking for*" in the data. During the interview, Clarke alluded to how in this scenario, the designers' expertise is in designing the *"right"* visualisation to provide the user with those insights. The imagination of the user as capable and knowledgeable is reflected in the imaginary figure of *the professional*, who is imagined to *know* what they *want* from data visualisations. As this chapter will go on to reveal, this is in direct contrast to how the designers imagined *the grandma* and *the public*.

As well as conducting user research, user testing is another strategy designers can use to determine if their intended or *imagined* user will be able to make meanings from the data visualisation they have designed. During interviews and observations, the designers spoke about asking people who were close to them either geographically or personally to test their visualisations. Rather than being systematic, this informal kind of user testing appeared to be built around *convenience*. During an interview with Wilson, I asked if they thought about how users would read and interpret the data visualisations they worked on. They replied:

> Wilson: *"Yeah, and that really depends on who they are. So, ideally, when you're user testing it you look for someone in the street who is close or in that target group you're designing for and ask them if they understand what it is. Because that gives you the best* [inaudible], *if the real people you are designing for are going to get it."*

Wilson's response suggests that they believe that how a user engages with a data visualisation is dependent on "*who they are*" and that user testing offers a way to test how users would read and interpret their visualisation design. However, their view that you can get access to the *"real"* users by simply asking people on the street suggests they perceive groups of users to be fairly homogenous. Although the designers talked about user research and the importance of designing for the user, examples of diversity tended to focus on age, gender, and to a lesser degree age related disability. As the chapter will go on to explain, this understanding of diversity among users is reflected in the imaginary figure of *the grandma*. The designers did not talk about how social class, ethnicity, level of education or other socio-demographic factors might affect a person's ability to engage with data visualisations. Similarly, Oudshoorn, Neven and Stienstra (2016) found that although the designers in their study considered diversity among users to be important, they did not consider factors such as social class or ethnicity in the selection of user testing participants. This suggests a narrow working view of diversity in relation to the testing of their participants' designs. If the designers at Evoke view age and gender as the main characteristics that represent difference among users, it becomes easier to see how designers might feel it is possible to identify their target users "*in the street*."

Another designer, Taylor, also mentioned asking people *"off the street"* to conduct user testing. However, the street where Evoke is situated is close to the city's university and appeared to be a fairly affluent area, surrounded by shops and businesses. Therefore, it is likely that this form of user testing would capture a particular demographic who might be very similar to that of the designers – young, well educated, and computer literate. It is possible to see how this form of convenience user testing may well reinforce the designers' somewhat narrow assumptions about how 'different' users engage with data visualisations. However, Jones talked about asking those who were close to them personally to test their designs. During an interview, I asked Jones if they imagined how people would use or consume the visualisations they created. In response, Jones said that it is good to ask other people what they get from the visualisation, explaining:

> Jones: *"*[…] *you can try to imagine how people will consume it, but I think the best test is still to ask people around. Just take, I don't know, your* [partner], *mother, whatever, and just show them the visualisation and ask them, within 30 seconds, what do you get from it? And if they don't get anything from it, then you're doing it wrong."*

The suggestion that you can test the design by *"ask*[ing] *people around"* defines the convenience user testing strategy and seems to be used to gauge how 'users' will make sense of a design. Here, Jones said that it is best to test the visualisation rather than try to imagine how people will engage with it. Drawing on their partner and mother they suggest that you can show the design to people who are *"around"* to test if they can draw meanings from it. Jones' quote demonstrates how they do not presume that the user will be able to decode the visualisation they designed as they had intended. Nevertheless, it also seems reasonable to think that these kind of informal testing interactions may shape Jones' understanding of users more generally. Although this is a rather pragmatic approach to user testing, it does not easily allow the designers to build diversity among users into their user testing. Woolgar (1991) argues that it is hard for producers of technology to *know* the user from their position as an insider in a design company. The user represents an *outsider* (Woolgar, 1991) therefore, asking people outside of the studio, on the street or even those close to the designer, provides the designers with a means of accessing 'the user.' However, this kind of convenience user testing may also account for why the imaginary figures I identified were broad and homogenous categories of user, with only the grandma possessing socio-demographic characteristics.

During our interview, Wilson suggested that testing the visualisation with others was a good way to validate their design, explaining that, "*what I like to do most is just walk up to someone, show them something and ask them, so, what's this?"* It is not clear if Wilson is referring to someone 'off the street' or a colleague in the design studio. Either one would raise questions about the diversity of those Wilson was testing their designs with and if they were more socially diverse than the predominantly young, male, educated, and tech-savvy workforce in the studio. However, during an interview with Jones, they explained how testing designs might involve *colleague validation*. I had asked them about how they make the design of a data visualisation visually appealing but in a way that it still makes sense to the user, in this case the general public, they replied:

> Jones: "*Yes that's really hard* […]. *So sometimes you plot it and you go see your colleague and say, 'hey, you have five seconds to tell me what you get from this graph.' And when they don't* [get anything from it] *that means it's too complicated, too abstract, too* [attractive], *you're losing the core information there. It gets a bit difficult to do it here, because most of the time you ask designers and then they tell you, 'it's boring.'"*

This form of user testing relies on the judgement of other designers to determine if the visualisation can be interpreted by the user. As with convenience user testing, Jones is looking

for their colleagues to tell them what their interpretation of the graph is in order to gauge if they can make sense of it in the way the designer intended. Although testing the design of technology with colleagues is not uncommon, designers are rarely representative of the end user (Oudshoorn, Rommes and Stienstra, 2004). While imaginary figures might stand in for the user, Akrich (1995) describes how designers often attempt to step back from their role as designer to adopt a layman's perspective in order to form a representation of the user. This is referred to as the *I Methodology*, which describes how designers draw on their own preferences when making design choices (Sørensen, Faulkener and Rommes, 2011, see p 139). As mentioned above, the workforce at Evoke represents a fairly narrow and predominantly male demographic. Therefore, the way they engage with and make sense of the visualisation is unlikely to be representative of the experiences of all users. Furthermore, their position as an *insider* in the design process makes it extremely unlikely that they will share the same *worldview* as the users (Sørensen, Faulkener and Rommes, 2011) when engaging with data visualisations.

Jones' quote suggests that when seeking a colleague's validation for their designs, other designers not only respond in terms of what they *"get from"* the graph but also if they find it *"boring."* At first I assumed that the designers would be referring to the aesthetic design of the visualisation, however Jones went on to explain that designers might also deem the *topic* of their visualisation to be boring. They talked about when they found something they thought was interesting in a dataset but when they showed their finding to some of the other designers, they told them it was "*boring*." Jones seemed to view their colleagues' validation as an indicator of the success of a design, saying:

> Jones: *"So that's actually good proof because when you manage to convince a designer then you're like, 'yes! I did it.'"*

This raises questions around *who* decides what is interesting in the data and *who* decides what should be made visible. As Beer (2019) argues, those with the ability to tell stories with data are in a position of power. Indeed, those experts with access to data and the skills to analyse it set the agenda in terms of what kind of research is conducted and what insights are revealed (boyd and Crawford, 2012). In the context of how colleague validation may influence how the designers imagine users, Jones' quotes suggest that the designers assume that their own aesthetic and subject preferences reflect those of the user. It is impossible to know for certain, but it could be that the designers are drawing on these preferences when determining what the imaginary figure of *the public* will find engaging.

Project specific user research, convenience user testing, and colleague validation helped the designers at Evoke to form their ideas about how users will engage with and make meanings from the visualisations they design. From the way the designers spoke about these different strategies, it appears that project specific user testing helped the designers configure the imaginary figure of *the professional*, who is perceived to know what they *want*. *The grandma* represents the designers' limited understanding of diversity among users. Therefore, it seems likely that this is framed by convenience user testing, which does not easily allow for a diverse sample of users. Finally, it seems likely that *the public* are formed through convenience user testing and colleague validation, through which the designers determine what the user will find engaging, both aesthetically and in terms of the topic of the visualisation. This argument will become clearer as this chapter continues into a discussion of the different imaginary figures identified in the studio. Yet this section has provided some context as to what the designers may be drawing on to configure these different imaginary figures.

The following sections will discuss each of the imaginary figures, explaining how they shaped how designers encoded meaning into visualisations. It will do so by focusing on how imaginary figures prompted designers to prioritise different aspects within the design, how they framed how diversity among users was constructed in the studio, and the different design values they prompted the designers to enact.

5.3 The Grandma

The Open Data Institute (ODI) published a guide on their website called, "*What is 'open data' and why should we care?*" (Scott, 2017). The article was split into ten "need to know" points, which described open data and provided examples of ways that it can be useful. The first two points began:

> *"1. Open data is data that's available to everyone to access, use and share. Yep, even your nan."*
>
> "*2. But open data should be easy to access, especially for your nan."*

Positioned just below the second point was an animated gif of the cartoon character Homer Simpson. In the animated image, Homer is sat in front of a computer screen displaying the statement, '*to start press any key*.' Homer looks dismayed as his hand hovers back and forth over the keyboard. The gif is captioned "Where's the "any" key?" The way the ODI (Scott, 2017) worded points one and two, combined with the gif ridiculing Homer Simpson's lack of

computer literacy draws on cultural stereotypes about older women's perceived lack of ability to engage with and understand data. The ODI is playing on a cultural assumption that older women are less likely to possess the technical skills required to access, use, and share data. 'Nans' are highlighted at the beginning of the guide as an exceptional case to drive home the ODI's point that open data is easy to use and accessible. The message is, if your grandma can access open data then anyone can. The ODI's blog (Scott, 2017) demonstrates how older users are not only configured by designers but also by organisations who *advocate* for users (Oudshoorn and Pinch, 2005). The media, policy makers, and even public sector organisations all do important work in configuring users (Oudshoorn and Pinch, 2005, see p. 9). Therefore, it is important to understand that how the designers at Evoke imagined *the grandma* is part of the broader social and cultural context within which older users are viewed as "other" within technology design (Oudshoorn, Neven and Stienstra, 2016, p. 181).

During fieldwork I was struck by the designers' numerous references to grandmothers and the elderly. In interviews when asked about how they imagine users, designers often drew on the elderly, mothers and more often grandmothers, to explain how they designed for '*different'* users. They referred to these figures to explain how the intended user influenced the design decisions they made. This, in turn framed how meanings were encoded into data visualisations through the visual and technical elements of their design. As this section will explain, these imagined users made up the imaginary figure of *the grandma,* whom the designers called upon to explain to me how they made design decisions for different users. Analysis of the data suggests that the grandma actually represented all those users who possessed the lowest levels of meaning making ability in relation to data visualisations, often due to their perceived lack of technical ability.

Despite incorporating the designers' references to the elderly more generally, I have chosen to refer to this imaginary figure as *the grandma*. This is because there was a gendered element to the way in which the designers imagined this type of older user. While fathers and grandfathers were missing from the designers' accounts, references to grandmothers were made repeatedly, as well as one reference to a designer's mother. This was despite never asking the designers about how the age or gender of a user influenced the design choices they made. Yet it seems these two socio-cultural factors represented designers' working understanding of diversity among users. This section of the chapter will mainly draw on interview data to explain how designers configured the imaginary figure of *the grandma*. This will be explored in relation to the prioritisation of accessibility in design, the perceived limits of meaning making, and expressions of empathy from the designers towards this user. Importantly, it will also describe

how the designers drew on the grandma to encode meaning into the design of data visualisations for those users deemed less able to make meanings from visualisations.

Kennedy et al., (2016b) addressed the issue of diversity in their study, which was attentive to how socio-cultural categories such as age, gender, social class, and ethnicity affect how users *engage* with data visualisations. Hill, Kennedy and Gerrard (2016) wrote about how users interpret data visualisations by drawing on cultural discourses that are gendered, aged, and classed. However, neither of these papers explored how assumptions around the diversity of users are encoded into data visualisations. Oudshoorn, Rommes and Stienstra argue that, "To understand how technological artefacts come to incorporate barriers against specific groups of users we need to shift the analysis to users in the semiotic sense: users as imagined by the designers of technology" (2004, p. 31). This section of the chapter addresses this gap in understanding and not only shifts attention towards how diversity is being constructed in the studio but explores how designers' understandings of diversity are encoded into the design of visualisations. As this section will argue, the imaginary figure of the grandma reveals how designers perceive diversity among users and demonstrates how they encode meaning differently for different people. Diversity appeared to relate only to age, gender and predominantly age-related disability, all of which were considered in terms of the users' ability to engage with and make sense of data visualisations.

In configuring the imaginary figure of the grandma, some of the designers at Evoke imagined their own mother or grandmother when thinking about how *different* users engage with their products. Below is an extract from my field notes, recording a scheduled observation I conducted with Reed:

> *Interestingly, when I ask Reed if they test the product with the same people with which they conduct user research in phase one, they say no. They go on to say that sometimes it's useful to test a product with your mum, because there's a divide there, but usually it's more complex here*. (Paraphrased field note from scheduled observation, 25. 10. 2017).

Here Reed seems to be suggesting that there is a difference, or as they put it, "*a divide"* between how the intended user will use and interpret a data driven product and how their mum might engage with it. Reed then suggested that much of the work being done at Evoke is "*more complex"* than the type of products they would show to their mum. It is not clear if they mean that the work they are doing is too industry specific, too difficult for their mum to understand, or perhaps just not interesting to her. Nevertheless, analysing this field note alongside the

following data extracts have led me to infer that the divide Reed mentioned is about accessibility and specifically the designers' perception of older women's technical ability. It is also interesting that Reed would refer to their mum in relation to user testing as this supports the kind of convenience user testing discussed in the previous section. It also implies that Reed is drawing on their own relationships with older women to configure this type of user and to inform their understanding of accessibility.

Understanding how designers imagine older users is important because their perceptions of ageing can become materialised in the technology they create (Oudshoorn, Neven and Stienstra, 2016). As the chapter goes on to explain, the way the designers imagined the grandma influenced their design decisions, potentially *scripting* (Akrich, 1994) this imaginary figure into the design of data visualisations through the choices they made around chart types and interactivity. During fieldwork, I began to realise that older users were being imagined by the designers in limited ways, which centred on their perceived need for accessibility in design. As the following data extracts show, two designers also talked about children, however on both occasions this was followed up with references to older users. In fact, it seemed that these two loosely defined groups of user were imagined as possessing similar needs in terms of accessibility. This, it can be argued, reflects a broader cultural trend in Western society in which elderly people are infantilised (Hepworth, 1996). Talking about very young and older users in the same context also seemed to provide a means through which designers could express empathy towards users.

Discussing the design decisions made in relation to both children and older users is demonstrated in the below extract from an interview that I conducted with Wilson:

> Interviewer: *"So, when you are designing data vis, or making data vis, are you conscious or thinking about the end user of the visualisation when you are designing them?" [Yes, yeah] "Do you imagine who they are?"*
>
> Wilson: *"I like to have personas, either that or see some actual interviews or hear some actual interviews. Yeah just a grasp of who I'm designing for or building for is very useful. Because you can change certain things based on who it's for. Like, if it's for children, it might be more colourful, or have bigger buttons, and the same goes for elderly people. Or, you know, it's a little bit more serious for people aged in their forties. Yeah, there's all kinds of choices."*

This quote demonstrates how designers felt it was important to have a sense of *who* they were designing for. Furthermore, it shows how design choices are informed by the designers' perceptions of the intended user. Wilson explains how they might make the design "*more colourful*" or use "*bigger buttons*" when designing for children, before adding that these same aesthetic design choices would also apply for elderly users. This lumping together of very young users and elderly users is indicative of how technology design often others and infantilises older users, while at the same time not taking their views seriously (see Oudshoorn, Neven and Stienstra, 2016). The way in which older users are 'different' is represented in Wilson's quote when they say that designs are "*more serious for people aged in their forties.*" Users in this age group are more likely to be economically active and although it is impossible to be certain, I have inferred that here, Wilson was referring to the imaginary figure of *the professional*, for whom Evoke design and build dashboard applications.

Wilson's quote demonstrates how design choices are based on how designers imagine the end user, thus encoding the perceived subjectivities of these users into the design of data visualisations. Intrigued, I asked Wilson how they *knew* how to design this way for these types of user. When explaining how they design for elderly users, they said:

> Wilson: "*The elderly are often shaky, or they have Parkinson's or Alzheimer's, and you want to keep that in mind while you're designing because it's a vulnerable target group. Yeah, you've got to pay attention to stuff like that. But the rest, the rest is just assumptions, yeah."*

Here, Wilson draws on disabilities associated with elderly people to explain why they would make the design choices they had mentioned. They even refer to the elderly as a "*vulnerable*" group of users. This links to Oudshoorn, Neven and Stienstra's research into the design of technology for older users, in which they argued that among designers, "older users are depicted in terms of illness and dependence" (2016, p. 179). Wilson includes the caveat that they are also making assumptions about what design choices will work best for the different types of user they mentioned. This suggests that their knowledge is not only based on user testing and research, but also on wider held social and cultural assumptions about the level of ability older people possess to engage with technology and thus data visualisations.

Wilson's quote reveals how they imagined the world of their users in order to make design choices. Akrich argues that producers of technology imagine the world in which their products will be mobilised and ascribe "this vision" into their designs (1994, p. 208). Wilson's quote implies that they imagine the world of older users in terms of disability and vulnerability. That

this vision shapes the design choices they make suggests that the elderly are likely to be excluded from certain forms of data visualisation. While graphs and charts that represent the complexity of data or complexity in design are more likely to be aimed at users who are perceived to possess greater levels of ability to interpret them. This offers some explanation as to why Wilson suggested that more *"serious"* data visualisations are for those users who are aged "*in their forties*." Oudshoorn, Rommes and Steinstra (2004) argue that exploring how designers imagine users will lead to a greater understanding of how technology excludes certain users. Building on this argument and the work of Akrich (1994), my research suggests that it would be prudent to consider how the imaginary figure of the grandma is encoded into the design of data visualisations. As well as if and how she is decoded by older users in ways that might perpetuate inequalities in access to data. As already suggested in this chapter, the encoding and decoding of imaginary figures is one way in which data visualisations may materialise and reproduce the very stereotypes that inform their design.

Unfortunately, Wilson does not identify exactly what they are drawing on to configure older users and very young users. Wilson does say, however, that some of their ideas are based on assumptions. Even when user research is available, taken for granted assumptions or stereotypes about different types of user might still creep into the designs of data visualisations. Our conversation was not in the context of specific users or a specific project, so perhaps this also played a part in Wilson making some assertions. Nevertheless, this data extract provides an insight into how Wilson imagined the worlds and the bodies of elderly users as being characterised by disability and vulnerability. Buse, Nettleton, Martin and Twigg argue that imagining the bodies of elderly users within the design process requires "various forms of empathetic engagement" (2017, p. 1440). Perhaps then, the imaginary figure of the grandma allows the designers to perform empathy towards users who are deemed to need the most help to engage with and make sense of data visualisations. Having empathy towards both the elderly and very young users is reflected in an interview I conducted with Jones. The following extract from that interview was the first part of their response to a question I asked about how they thought different people used the visualisations they designed, they said:

> Jones: *"Yeah. I think what comes most often is trying to think about your grandma or your kid. Try to put yourself in their shoes and decide if you can understand the visualisation without all the background information."*

It is interesting that asking Jones a question, that prompted them to consider how they think about 'different' users, led them to talk about the grandma. This supports the idea that this imaginary figure represents diversity among users. In this quote, Jones suggests that when

designing they try to place themselves in the "*shoes*" of a grandma or a child in order to decide if the user will understand the visualisation. This requires them to imagine themselves in the world of the grandma, which involves a level of empathetic engagement with them. The idea of designers putting themselves in the shoes of their user is not unique to data visualisation or technology design (see Buse et al., 2017). However, Akrich (1994) argues that imagining the worlds of particular users involves making many assumptions, not least about their "competencies" (Akrich, 1994, p.208). In this quote, Jones seems to be drawing on the imaginary figure of the grandma (as well as children) to assess the accessibility and usability of their design. This supports this chapter's argument that for the designers, the grandma represents all those who possess the lower levels of meaning making ability, in terms of their aptitude for engaging with and making sense of data visualisations. In this sense, she is the benchmark for accessibility. Jones' quote also reinforces the view that empathy plays a role in how meaning is encoded into the design of data visualisations. The imaginary figure of the grandma provides a means through which designers could express empathetic engagement with the user.

The designers appeared to draw on the grandma as an exceptional user who required different design choices than other users. This is demonstrated in a quote from Turner, who was responding to a question I had asked them about if they imagined the end user during the design process. Turner asked me to explain what I meant by imagine, so I clarified by asking them:

> Interviewer*: "So do you think about who that end user is going to be when you are designing at the start, does it affect your choices?"*
>
> Turner: *"Yeah, sure, yeah. If you think it's a - if it's a very old person who doesn't have much experience with computers you should make the thing easier or more readable, or."*
>
> Interviewer: *"OK."*
>
> Turner: *"Yeah, you always think about that. It also affects the type of graph and the level of interactivity and maybe the font sizing."*

Turner appears to be making an assumption here, that older people do not have much experience with computers. This is a fairly common stereotype about older people and one that Floyd, Jones and Twidale (2008) argue is actually more productive to designers than imagining an older user who adopts new technology with ease and enthusiasm. Turner draws on this

assumption to make design decisions, prioritising accessibility in their design by making the visualisation easier to read. Turner goes on to explain that their perception of the user frames fundamental design decisions including the chart type, level of interactivity and font sizing, echoing the kind of design choices Wilson spoke about when designing for young children and elderly users. This demonstrates how designers draw on their imagination of users to encode meaning into the design of data visualisations differently, depending what category of imaginary figure they fall into. The way the designers configured the imaginary figure of the grandma in terms of accessibility positioned her at the lower levels of meaning making ability. The data suggests that how designers encoded meaning into visualisations for this user was limited to the *accessibility* of the design, rather than what they might *like* or *want* from a visualisation.

Turner's quote prompted the question of how they *knew* which graphs were easier to read. Wilson had struggled to answer a similar question I asked them in relation to how they *knew* how to design for older people. When I asked Turner how they knew which graphs would be simple to read their response also suggested they were drawing on assumptions and design intuition:

> Turner: *"It's more about experience.* [If it's] *a very complicated thing with lots of interactivity and I see my grandmother using it, I think no, this is not going to work."*

As Floyd, Jones and Twidale's (2008) article on personas argues, it is not unusual for designers to draw on their design intuition when thinking about users. This is reflected in what Turner says when they emphasise their own experience in informing their knowledge about which chart types are more accessible to older users. However, Turner also draws on their grandmother to imagine how older users would engage with different types of data visualisation to make decisions about which graphs to use. Buse et al., (2017) note how architects working on the design of care spaces for older users drew on what they perceived to be the needs of their own older relatives to inform their designs. It is interesting that Reed, Jones and Turner all drew on older female relatives when thinking about the user in relation to the accessibility of the design. While, fathers and grandfathers remained conspicuous by their absence.

Many mothers are aged in their twenties and grandmothers in their thirties and forties and are therefore not 'old' in a Western context. Therefore, it appears that in talking about mothers and grandmothers, the designers may be drawing on broader cultural stereotypes around older women in terms of care giving, the home, and economical inactivity. Indeed, D'Ignazio and Klein (2020) describe how paid work is more culturally valued because it is economically

productive, whilst the unpaid labour that takes place in the home is not. Forms of unpaid domestic labour are disproportionately done by women (D'Ignazio and Klein, 2020). It is important to remember that the designers at Evoke were not working in a "vacuum" (Stewart and Williams, 2005, p. 211). They were exposed to and were probably influenced by cultural representations and stereotypes that exist in the media and wider public imagination (Stewart and Williams, 2005). It is not clear from the data why mothers and grandmothers were imagined as somehow less able to engage with and make sense of data visualisations successfully, yet it could be linked to their perceived economic inactivity. This prompts the question, what is it about women associated with care giving roles that appears to exclude them from complex data visualisation designs? Furthermore, could this be considered as another form of data divide (boyd and Crawford, 2012) through which inequalities in access to data are built into the design of data visualisations aimed at older female users?

Significantly, the grandma not only featured in references to designs made specifically for older users but was also a consideration in projects aimed at a broader user base. During a follow up interview with Jones, they brought with them some examples of data visualisation projects they had worked on. One of these was a chart that our conversation led me to believe was for a mass audience. They explained to me:

> Jones: *"It was not an easy project because we had changes in the direction of where we wanted to go, and, it's very difficult to make something that's both understandable by your grandmother and something* [that] *like, looks* [attractive] *design wise so, you have this balance [...]."*

From Jones' quote, there appears to be a tension between designing something aesthetically pleasing but that is also accessible to a wide range of users including "*your grandmother."* It suggests that the design choices the designers make - the visual elements through which meaning is encoded into the visualisation - are a balance between aesthetics and accessibility. It appears that Jones drew on the imaginary figure of the grandma to negotiate this design process, reinforcing the argument that the grandma is the benchmark for accessibility. How the grandma became the user who represents those with the lowest levels of meaning making ability in relation to data visualisations is less clear. During fieldwork, the designers were mainly working on elements of a large dashboard project for a commercial client. Although the designers spoke about designing for older users I did not observe this taking place and there were no examples of such projects showcased on Evoke's website. I was left wondering *who* the older users they kept referring to were. Analysis of the data suggests that older people and grandmothers in particular have come to represent anyone who is somehow less able to engage with data

visualisations. Therefore, when the designers draw on the grandma, they were designing for any user who they perceived to possess low levels of ability in relation to making meaning from data visualisations.

Age and gender were important factors when imagining the user. During an interview, Turner explained that sometimes the designers would create personas from user research, using the examples of a young woman and an older woman they explained:

> Turner: *"As designers we read the research* [that has been completed about the users]. *Sometimes we create personas; so we know this is Caroline, she's 24 years old, she's just out of school, and blah, blah, blah. Or we have Margaret who is 80 years old, and then.*
>
> Interviewer: *"and then you think about those as you're designing?"*
>
> Turner: *"Yeah."*
>
> Interviewer: *"Yeah, is Caroline going to like this, or can Margaret understand this?"*
>
> Turner: *Yeah like, or use, yeah, yeah.*

During our exchange, I attempted to clarify my understanding of how Turner imagines the user as they design. In doing so, I inadvertently projected the same assumptions about how age and gender relate to accessibility and engagement that I had picked up on while embedded in the design studio. Significantly, though, Turner did not correct me and actually enthusiastically agreed with what I had said by repeating it back to me. Notably, when doing so Turner changed the word *"understand"* to *"use"* in relation to "*Margaret*." This reflects their concerns about their own grandmother not being able to use complex interactive data visualisations. In contrast to how Margaret was perceived in terms of accessibility, Caroline was imagined in terms of what she would "*like."* Oudshoorn, Neven and Stienstra suggest that when designing for younger users designers are concerned with notions of "choice, lifestyle and experience" (2016, p. 172). Yet, when designing technology for older people, designers' think about ill health and dependency (Oudshoorn, Neven and Stienstra, 2016). The empirical research the authors are drawing on is not in the context of data visualisation design. However, Turner's quote and those of the other designers in this section demonstrate how older users are being reduced to needs and accessibility over what they might *like* or *want* from their engagement with a data

visualisation. Drawing on such hegemonic stereotypes of ageing strips away any sense of diversity that might exist among this group of imagined users.

Personas encourage designers to feel a connection and empathy to unknown users (Turner and Turner, 2011), as well as preventing the tendency among designers to design for themselves (Floyd, Jones and Twidale, 2008). The design team at evoke was predominantly young and male and it is telling that Turner prioritised age and gender in their examples of personas – users who were different to the demographic of the designers. The *Seeing Data* project found that social class was significant in people's engagement with data visualisation (Kennedy et al., 2016b). However, during fieldwork I never heard the designers talk about users in relation to social class or other socio-demographic categories such as ethnicity or level of education. This finding is echoed in the work of Oudshoorn, Neven and Stienstra (2016, see p.170), who found that with the exception of age and gender, other socio-cultural categories were neglected by the design teams they studied. Indeed at Evoke, age and gender, and to an extent, age related disability, appear to represent the limits of the designers working understanding of diversity among users. These understandings of diversity were personified in the imaginary figure of the grandma, who represented a homogenous category of users possessing the lowest levels of meaning making ability in relation to data visualisations. This imaginary figure shaped how designers encoded meaning into visualisations for these users, through the design choices they made.

The imaginary figure of the grandma was configured from assumptions the designers' made about the ability of older users to engage with and make sense of data visualisations. However, during interviews the designers referred to grandmothers but not grandfathers, demonstrating the gendered element of this imaginary figure. As a benchmark for accessibility, the grandma informed design decisions including the choice of chart type, level of interactivity, font, and colours. These important visual elements communicate the meaning of the visualisation to the user. Hence, the grandma shaped how designers encoded meaning into the design of data visualisations intended for older users as well as other users positioned at the lower limits of meaning making. The grandma also represented the designers' working understanding of diversity, which appeared to be limited to age and gender. However, how the grandma is imagined leaves little room for diversity among those users who she represents. The following section will introduce the imaginary figures of the public and the professional, explaining how they shaped how designers encoded meaning into data visualisations for these types of user.

5.4 The public and the professional

During fieldwork, designers at Evoke talked about designing for *the public* and *the professional*. Like *the grandma*, they drew on these imaginary figures to explain their design decisions in relation to different types of user. These design choices were based on assumptions about what the public or the professional user would want in terms of aesthetics, usability, and how these different users should be guided through the data. Rather than discussing these imaginary figures separately, it is important to analyse them together because they are formed in relation to one another. The public and the professional are a dichotomy who come to life when viewed in opposition.

While *the grandma* represented those users who required the most help in terms of accessibility to be able to engage with and make sense of data visualisations, *the public* and *the professional* served other purposes. The professional represented engaged and competent users who could interpret the meaning of data for themselves, through its visualisation. The public were not assumed to possess the same levels of experience or engagement in interpreting data visualisations and therefore the meaning of visualisations for them was 'pre-determined.' While the designers focused on what the professional wanted from the visualisation, they concentrated on engaging the public. This section of the chapter will explain how designers imagined these two imaginary figures. Crucially, it will demonstrate how the designers at Evoke drew on the public and the professional to encode meaning into the design of data visualisations differently for these users. It will begin by highlighting some of the assumptions the designers made about users' preference for receiving information visually rather than through text, which framed how they configured the public and the professional.

I asked every designer I interviewed if they felt that data visualisations had the power to make people care about the issues they represented and they tended to agree that they did. This reflects the findings of Kennedy et al., (2016a) who report how the designers they spoke to felt that data visualisation could be used to empower and to raise public consciousness about issues. However, asking this question in my interviews also revealed some of the assumptions the designers made about the meaning making superiority of the visual over text. Jones, spoke about this in terms of engagement and speed of comprehension when they said:

> Jones: "[...] *for me it's very clear that if you communicate something with video or with text, people will go for the video because [...] nobody likes to read. I do like to read, but I'm a bit of a nerd. I can understand that some people don't like to do this, and they want to get information as quickly as possible."*

Jones makes the assumption that most people don't like to read and would prefer to receive information in a visual format such as a video. They relate the preference for video over text to the speed at which people can obtain information visually. Interestingly, Turner gave a similar response to the same question, when suggesting that people prefer to look at a visual representation of numbers such as a graph or chart, rather than viewing them in a table, saying:

> Turner*: "It* [data visualisation] *gives people insight into problems with numbers. You can give people tables with numbers but they don't look at those. If you give them graphs, people understand graphs."*

In this quote, Turner states that visualising data allows people to get an insight into a problem through numbers. This view reflects the wider rhetoric of the data analytics industry, which is built around the promise of providing accessible insights at speed, often in the form of data visualisations (Beer, 2019). However, Turner goes on to make a considerable assumption by suggesting that *people* do not look at tables, yet they possess the skills to understand graphs. It is not clear why they make this assumption but it could be that Turner is drawing on their own preferences when imagining users. During their study of the design of virtual cities, Oudshoorn, Rommes and Stienstra found that the designers often "assumed that their own preferences and skills were representative of those of the user" (2004, p.41). However, my own interviews with ordinary users revealed an instance where a participant, Glen, stated that the data would have been easier to comprehend in a table than in the visualisation I had shown them. Of course, Turner is talking here in general terms but the kind of assumptions that they and Jones make about user preferences reveals something about how they imagine users in relation to their own preferences.

While accessibility characterised the grandma, levels of expertise appeared to be an important factor in distinguishing the public from the professional. Indeed, users' expertise was an important theme, which emerged through my discussions with the designers. Visualisations can be more or less complex, for example by incorporating layers of data through interactive elements. The complexity of the visualisation and the level of access to the data it provided appeared to be, in part, dependent on how the designer imagined the intended user. While the professional might value and make use of such complexity, it was deemed unwanted and even unnecessary for the public. During an interview with Jones, they explained to me how the general public do not want to delve into the data in this way, saying:

> Jones: *"*[...]*if you are talking to the general public, who just never want to dig into a complex graph, they want to have the information right away, then you are going to*

remove the, not remove the precision, but, make the precision very easy to understand for normal people"

Here Jones suggests that the public have no desire to "*dig into a complex graph*" and that designing for this user might involve simplifying the meaning of the data so that they can interpret the visualisation quickly. It is interesting that they refer to "*normal people*" and although it is impossible to be certain, analysis of the dataset as a whole suggests they are referring to non-specialist users, or 'everybody else.' Therefore, Jones' use of this phrase draws a distinction between the public and professional. It is not clear why Jones believes that the public do not want to engage with a complex graph. Yet what they say suggests that when encoding meaning for this imaginary figure, the focus is less on layering data and interactive functions and more on communicating meaning from the data in a way that is simple and quick for users to grasp. Jones referred to the public again when they said:

Jones: *"*[.. .] *if you are talking to the general public you will not let them explore your graph. You will just show them the result and say, 'see, that's what you should understand.' And if you're talking to someone who knows a lot, maybe that person wants to manipulate it, maybe he wants to explore more because he has comprehension, and he can go a step deeper."*

Jones suggests that graphs designed for the imaginary figure of the public do not allow the user to "*explore*" the graph and instead communicate what the user "*should understand*" from the data. This suggests that for the public, the designers determine the meaning of the data before encoding it into the visualisation. In this sense, the meaning of the visualisation is made explicit in its design and there is seemingly less room for the user to make alternative meanings from the data. However, users who are imagined to be more knowledgeable are afforded the opportunity to explore the graph and even manipulate it to make their *own meanings* from the data. It is unclear if Jones is referring here to people who *know a lot* about visualisations, the data, or the topic the data in the graph represents. Nevertheless, they seem to be making a clear distinction between the non-expert user and the more expert user. While the public are not trusted to decode the meaning of the data independently, the expert user is provided with a visualisation that allows them to explore the data and make their own meanings. Jones seemed to be either adopting a more author driven or more reader driven approach to their designs (Segel and Heer, 2010) depending on how they imagined the user. This supports the argument that the designers encoded meaning into the design of data visualisations differently depending on the imagined user.

Drawing on Woolgar's (1991) definition of configuring users, imaginary figures allowed the designers to imagine the *identity* of the user and to draw on this to set limits on how they engaged with data visualisations. In removing complexity and the ability to interact with the chart or graph, Jones restricted the ways in which *the public* could engage with it and thus make sense of it. While the affordances (Erofeeva, 2019), that is, the possibilities, that are built into data visualisations designed for those users who were imagined to have more knowledge, allowed them the opportunity to explore the data for themselves. Indeed, Erofeeva argues that objects are inscribed, or encoded, in ways that dictate how the user should interact with them and this is determined by the designers' imagination of the "'ideal' user" (2019, p. 593). The public and the professional represent homogenous groups of users rather than an *ideal* type of user. Nevertheless, it is possible to see how imaginary figures allow the designers to encode meaning differently for the public and the professional, thus shaping how users engage with and make meanings from data visualisations.

Although it is not clear *whom* Jones is referring to when they talk about users who know a lot, the fact that this user is imagined to be able to make their own meanings from the data through visualisation suggests they belong to the imaginary figure of *the professional*. This conclusion is supported by what Clarke said about designing visualisations for professional applications, such as dashboards. I had asked Clarke how they design a data visualisation so that it meant something to someone else. They replied:

> Clarke: *"Well, so this is the part where I have to slightly disappoint you. In a lot of cases, at least the bigger projects we do* […] *are actually the ones where we are designing applications that deal with data. And are designs for people to derive meaning but are more in the realm of what you'd call dashboards. Or they really become professional applications* […]. *So I don't design a specific meaning into the visualisation. I design for the visualisation to show the data points or the parameters that will allow them* [the user] *to make an assessment."*

Here Clarke seems to be making a distinction between different types of visualisation in relation to how meaning is encoded into their design. They explain that dashboard applications do not require them to "*design a specific meaning into the visualisation*." Instead, the meaning of the graph is determined by the user. Clarke suggests that for professional applications they design visualisations to allow the user to make an assessment about what the data means. This reflects the way in which designers prioritised speed and comprehension in the design of visualisations for the professional user. This also links to how the designers made meaning from data intended for professional dashboards in relation to how the data would be used (see Chapter four). While

Jones suggested graphs for the public show the user what they "*should understand*" from the data, Clarke's quote suggests that the professional determines the meaning of the data visualisations designed for them. It is not clear if this is because the professional is imagined to be more skilled at interpreting graphs and charts or if it is because they are likely to be familiar with what the data represents. Either way, meaning is encoded into the design of visualisations differently depending on both the imaginary figure and the purpose of the visualisation. It is also possible to identify from Jones and Clarke's quotes, a distinction between how *the public* and *the professional* are imagined.

During an interview, Jones talked me through their work on a project involving visualising a dataset of around 9,000 cases. The data was the results of a survey conducted by a newspaper website. The aim of this project was to present to their client how Evoke could visualise the data for the newspaper's readers. Jones explained how they created simple graphs to analyse the data before transforming these graphs into "*the data vis thing*," saying:

> Jones: "*But I can give you an idea of how we go from raw analysis like this one, where it's only a very simple graph to something that's more to the data vis thing. Where you make it more* [attractive], *less words and more visual meaning so people can understand* [it] *right away - what it is about. Especially for a newspaper where people, you know, they're not here to spend two hours reading an article."*

Here Jones seems to be suggesting that to elevate the visualisations they made while exploring the data to something that resembles a "*data vis,*" required them to make it look visually appealing. In the context of data visualisations produced in newsrooms, Weber, Engebretsen and Kennedy found that designers stressed the importance of the visualisation being "attractive and engaging" and that this was important in terms of securing readers' attention (2018, p. 199). Jones explained how they prioritised the visual over text when encoding the data visualisation with meaning, suggesting that this helps the user to understand it quickly. This visualisation was aimed at the public rather than the professional, yet Jones still drew on the rhetoric of speed of comprehension, which is integral to the analytics industry (Beer, 2019). However, Jones' reference to how long people spend reading an article suggests that this has more to do with encouraging engagement among the public than delivering insights for decision-making. While the designers prioritised speed and comprehension when designing for the professional, they prioritised engagement with the visualisation among the public. As Jones' quote suggests, this was through aesthetic design choices, which made the graph more visually attractive to users. As the next part of this section will demonstrate, simple graphs and chart types were reserved for the professional user, for whom the ability to make decisions from the data was given

precedence. Whereas designing for the public allowed the designers to be more creative with the aim of persuading users to engage with the visualisation.

During our interview, Clarke talked about the differences between designing data visualisations for the public and for professionals. Despite having told me that "*Simpler is always better*" when designing data visualisations, Clarke suggests that when designing graphs for the media it is important to "*lure*" people into engaging with them, saying:

> Clarke: *"[….] in the press or in media [...] you need to engage people, show that it's cool, to lure them in* […]*"*

Their reference to the press and media implies that Clarke is referring to graphs and charts designed for the public. They suggested here that the design needs to be "*cool*" in order to "*lure*" people into engaging with the visualisation. Literature in critical data studies has highlighted the ways in which data visualisations can be persuasive in terms of the messages they communicate (Kennedy et al., 2016a; Hill, 2017; Hullman and Diakopoulos, 2011). However, Clarke's quote suggests that in certain contexts the design of data visualisations need to be *persuasive* to encourage users to engage with them. Yet, they went on to explain that for other projects, they did not need to persuade the user to engage with their data visualisations because the user was already driven to do so. A few moments before they tell me this, Clarke had been explaining that much of the work they did within the company was around professional applications and dashboards. The context of this conversation implied that what they said next was in relation to the imaginary figure of the professional, who engages with data visualisations at work:

> Clarke: *"In a lot of cases I don't have to do it because people rely on it* [the data] *so I don't need to. So that means that I can stick to simpler ones* [graphs and charts] *because* […] *it doesn't need to be the flashiest tool in their tool box, it needs to be the best tool in their tool box".*

Clarke explains that the projects they tend to work on do not require the designer to lure the user into engaging with the graphs and charts, because they are already motivated to use the visualisations. These users are more likely to "*rely"* on the data, therefore, Clarke explains they can use "*simpler*" chart types to represent it. This, along with Clarke's comment about the visualisation needing to be the user's "*best tool*" rather than the "*flashiest tool*" seems to downplay the importance of aesthetics in the design of visualisations for *the professional*. Looking across the interview data, it is possible to conclude that the more creative and attractive

elements of visualisation design are less about the efficient communication of meaning and more about encouraging user engagement. Interestingly, this argument links to what Wilson said in the previous section about the design of the visualisation being more "*serious*" when aimed at people aged in their forties. Analysis of the data led me to argue that Wilson was drawing on *the professional* when they said this, and Clarke's quote supports this. It could be that simple chart types are used for the professional because they are expected to interpret what the data means for themselves, through the visualisation. Whereas the meaning of the data is to some extent 'predetermined' for the public and encoded into visualisations through aesthetic design choices. Nevertheless, the distinction between the public and the professional frames how designers encode meaning into the design of data visualisations for these different users.

While talking me through the process of visualising a dataset for the public, Jones had implied that it was important to move away from the simple graphs and charts they plotted when analysing the data, in order to make the final design of the data visualisation more attractive to the user. However, when designing for the professional, Jones suggests that the design needs to be *"efficient"* rather than visually attractive, explaining:

> Jones: *[...] "because it's going to be a dashboard for commercial people so you're not trying to do anything that's* [really attractive]*. On the spectrum of art you're doing analysis, you're not at the academic level, but"*
>
> Interviewer: *"but you're nearer that side than the art side?"*
>
> Jones: *"Yeah. You want to do very efficient charts that just tell you the information you want, right away. You're not trying to make it fancy [..]. If people want to see the number of* [sales]*, they want to see a line chart."*

Here, Jones explains how they would approach the design of a visualisation for a professional dashboard. While Clarke talked about "*simpler*" charts, Jones talked about "*efficient*" charts, although they both imply the same thing. That is, that for the professional, the priority is to produce graphs that the user can interpret the meaning of quickly. The aesthetic elements that appear to characterise visualisation design for the public are seemingly not necessary for this engaged user, who knows what they "*want*" from a graph or chart. D'Ignazio and Klein note how "plainness" in visualisation design is associated with neutrality and a lack of emotion, which allows the user to interpret what the graph or chart means, "for themselves" (2020, p. 76). Indeed, when encoding meaning into data visualisations for the professional, the focus was on selecting the right chart type that would enable the user to determine the meaning of the data.

Jones' focus on producing efficient graphs that reveal insights quickly reflects the use of visualisations in the growing analytics industry, which often uses graphs and charts to provide their customers with fast and accessible insights (Beer, 2019). This appears to contrast with the style of visualisations Evoke produce for the public, as Jones explained:

> Jones: "*So if I'm plotting something in an academic science paper I'm going to use a bar chart. Now if you're talking to a newspaper where it's more about the format and the fact of attracting readers rather than the content itself then you want to go here or here."* [Jones is pointing to an invisible scale between academic data visualisation at one end and abstract data art at the other].

This quote demonstrates how Jones makes design choices based on the imagined user of the visualisation. Seemingly, knowing if the chart is for the public or for an academic audience informs how abstract or precise the graph or chart needs to be. Jones suggests that data visualisations designed for inclusion in newspaper articles are less about the content and more about how the format can attract readers to the article. During the same interview, Jones had spoken about a data visualisation project they worked on for a tabloid style newspaper and they may have been thinking about that project when they made this comment. Nevertheless, this suggests that Jones perceives the public as being less concerned by the content than the design. This is another example of how "*simple*" graphs, in this case a bar chart, appear to be reserved for the professional user, while designs that are more visually attractive and presumably visually complex are designed for the public. This prompts the question, are data visualisations designed for the public more about entertainment and engagement than the communication of knowledge from data? If so, this would imply that many large scale and visually striking big data visualisations that circulate in popular culture are more about the spectacle of making data beautiful than the communication of "precise information and values" (Kennedy et al., 2016a, p. 716).

This is not to say that the designers totally disregarded the importance of making visualisations engaging for the professional. Clarke explained that they too, "*like to be engaged from time to time."* Therefore, Clarke said they would not design a dashboard consisting solely of bar charts. Similarly, after saying they would use bar charts to represent data in an academic article, Jones paused for a few seconds before adding:

> Jones: *"Although some people in this design company would argue that bar charts are boring, and I agree, bar charts are very boring."*

The perception that bar charts are "*boring*" and users do not find them engaging may reflect the designers' own preferences rather than those of the user. Indeed, during ordinary user interviews for this thesis, some of my participants referred to bar charts because they were familiar with them. As already argued, designers often assume that users will share their own preferences and interests (Massanari, 2010; Oudshoorn, Rommes and Stienstra, 2004). However, it is widely recognised that the assumptions designers make about users are unlikely to reflect their reality, design preferences, and technical ability (See Massanari, 2010 and Oudshoorn, Rommes and Stienstra, 2004). One reason why the designers may fall back on their own preferences is that the user is hard to know from within the studio (Woolgar, 1991). Arguably, this is why imaginary figures were so prominent in how the designers explained their design decisions. They provided a means for the designers to 'know' the user and to 'know' users who were different to themselves. Like the grandma, the public and the professional were productive, shaping how the designers encoded meaning into visualisations for these imaginary figures.

Meaning was not only encoded differently through the type of chart used for the public and professional, but in how the visualisation *led* them through the data. During our interview, I asked Clarke if they thought about the way a person will approach the graph they are designing in order to make sense of it. In their response, they made a clear distinction between how they would design a visualisation for the professional and for the public. In doing so they reveal more about how they drew on these imaginary figures to configure different users, in terms of setting limits on how they engaged with the visualisation. Using the metaphor of a *journey* they explained that professionals require a "*top to bottom*" design, which begins with a high level metric before leading the user to the data that created that metric, saying:

> Clarke: *"Interestingly, when it's about professionals, they are more top to bottom. So they want to see the core metric first; "Show me the net promoter score."* […] *So they want to see the core net promoter score at that point and how it's changed, you know, from last week or something. Then the next step they want to see is what countries, or what parts of the organisation, then they want to see how the longer trend* [has] *been.* […] *So they are generally more a top to bottom, I call it.*
>
> […] *And we try to do sort of the content based navigation for that type of stuff, so you see the metric – bang – you click on it you get the next level – bang – you get the next level."*
>
> Interviewer: *"And that's layering?"*

Clarke: *"Yeah, exactly. Yeah, gradual disclosure or progressive disclosure."*

Here Clarke articulated how they encoded meaning into the design of data visualisations for the imagined professional user. Segel and Heer (2010) talk about *author driven* and *reader driven* approaches to the design of data visualisations. Although their article is in the context of narrative data visualisations, they describe how allowing the user to drill into the data lends itself to a more reader driven approach. In this quote, Clarke talked confidently about what the professional user *wants* from the visualisation and explained how they layer the design accordingly to allow them to drill down into the data. Interestingly, they even imagine what the user would say they wanted from the visualisation, "*show me the net promoter score*." This illustrates how the imaginary figure of the professional is well formed in Clarke's mind and how they drew on them to make and account for design decisions. This quote also reveals how Clarke has configured the professional as a confident user who knows what they want from the visualisation and who is looking to get that information in a quick and efficient manner. This appears to be in contrast to how Clarke imagined the public. They went on to explain how they would design a visualisation for the public differently, saying:

> Clarke: "*The public you generally do it the other way round because they are not used to looking at the high level aggregation numbers. So they start with something that people know and then [go] into the trend. Like this thing recently about hurricanes, you see them on the map first, you know - just their track on a map, because people can relate to that, "Oh that was that hurricane on the map." Then the same lines then turn into a trend line that's more on time and* not *strength, and then it gets colour, so it's exactly [the] reverse."*
>
> Interviewer: *"Ok, that's very interesting. So you're kind of leading them into the harder factual information, I suppose? The finely grained information."*
>
> Clarke: *"Exactly, yeah."*

Clarke makes the assumption that the public are not used to looking at "*high level"* numbers and therefore the design of the visualisation for this user requires a different approach. This might be considered more of an author driven approach, which guides the user through the data using messaging and minimal interactivity (Segel and Heer, 2010). Using the example of a visualisation on hurricanes, Clarke describes how they would engage the user in the topic first by showing them something they might recognise and that they can relate to. This relates to the formation of a "meaningful discourse" (Hall, 1980, p. 130) that the user can decode (see

Chapter four). Of course, unlike the professional who is likely to be working with data that is familiar to them and which they can contextualise through their work, when the public encounter data visualisations they may be less acquainted with what the data represents. Just as they did when describing the design of visualisations for professional users, Clarke imagines how the public will engage with the chart. When they say, *"Oh that was that hurricane on the map,"* Clarke seems to be picturing how the imaginary figure of the public will draw meanings from the visualisation by identifying the hurricanes positioned on the map. As Clarke continues to explain how the visualisation would be layered with representations of the data, they talk about the visual elements of the design rather than the metrics. This contrasts with how they described designing for the professional but links to what Jones said about making visualisations for the public more visual than text based.

When Clarke explains how they would design visualisations differently for the public and the professional, they illustrate how they imagine these two types of user differently. The professional is imagined as knowing what they *want* to see when they are presented with a data visualisation and take the lead in drilling down into the data in order to derive meaning from it. In contrast, the public are imagined to be less used to engaging with data in this way and instead need to be *lured* into engaging with the visualisation by showing them something they recognise or can relate to. Indeed, making data visualisations engaging may well be about making them relatable to the user and this will be discussed in Chapter six. However, Clarke's quote also demonstrates how the public is not trusted to explore the data unaided. Instead, they discover the meaning of the data through a range of visual elements arranged onto the chart. Of course, all users of data visualisations will undertake meaning making through the act of interpreting them. Nevertheless, in Clarke's examples, it appears that the professional and the public are guided through the data in different ways, which relate to how the designers configured them. This further demonstrates how designers draw on imaginary figures to encode meaning into the design of visualisations for different users.

Data is used by states and commercial sectors to imagine the possible futures of citizens (Amoore, 2013) however, the designers' *lines of sight* (Amoore, 2009) reveal how they were imagining the possible futures of the users of data visualisations. This is in relation to how they perceive different people's ability to engage with and make sense of data visualisations. They encode these possible futures into the design of data visualisations through the affordances (Erofeeva, 2019) that are built into the design of visualisations for imaginary figures. Therefore, it is possible that imaginary figures are self-fulfilling through the reproduction of these assumed user types in the production and consumption of data visualisations. Returning to the grandma for a moment, Twigg writes about how the cut in the design of older women's

clothes is inscribed with "information about the sort of body that is meant to inhabit it" (2012, p. 1042). Furthermore, Sørensen, Faulkner and Rommes warn that focusing on gender differences in design runs the risk of reproducing traditionally gendered products (2011, see p. 142). These examples show how powerful imaginary figures may be in configuring the user of data visualisations in very particular ways. This involves imagining and setting limits upon their future use of visualisations in ways that might reproduce the very stereotypes and assumptions from which the imaginary figures emerged.

It is important to acknowledge how at times the professional will also be the public, for example when engaging with visualisations in newspapers. Therefore, a proportion of *the public* will be made up of users whom the designers imagined very differently to the homogenous group of general users this imaginary figure represents. This raises the importance of the various contexts in which people encounter and engage with data visualisations. These might include at work, at home, while watching the news, and so on. The designers' use of imaginary figures does not account for how users might move between being the public and the professional as they shift between these contexts in their everyday lives. Indeed, the designers' use of imaginary figures does not allow for the kind of nuance that issues around context raise. Massanari argues that personas are "political tools which oversimplify difference" (2010, p. 407) and the same argument can be made about imaginary figures. The public and the professional come into being in relation to one another and whilst this dichotomy helps the designers configure these imaginary figures, it flattens any sense of diversity within them. This lack of diversity reflects the tension designers have to navigate when designing for particular users while also ensuring the product remains useable by a broad range of people (Stewart and Williams, 2005). Therefore, the designers drew on these imaginary figures to make design decisions about how to encode meaning into the design of data visualisations for different, if not diverse, users.

5.5 Conclusion

As this chapter has explained, the designers at Evoke imagined how different types of user would engage with data visualisations and made assumptions about their ability to make meanings from graphs and charts. How they configured these different users shaped the design decisions they made and limited the possibilities of those whom they were intended for. The assumptions the designers made about different types of user were personified in three imaginary figures: *the grandma*, *the public*, and *the professional*. The grandma represented those users who needed the most help when engaging with data visualisations and were the benchmark for accessibility. The professional represented a competent and motivated user who knew what they wanted from data visualisations. This user was trusted to determine the

meaning of the data themselves, through its visual representation. The public represented 'everybody else' and designers prioritised engagement in the design of data visualisations for this user. The meaning of the data was determined by the designers and communicated to the public through aesthetic design choices. As this chapter has demonstrated, the designers encoded meaning into the design of data visualisations differently for these different types of user.

Although a constant presence in the studio, there was no physical trace of these imaginary figures, instead they were "discursively constructed" (Clarke, 2005, p. 46) through the way the designers talked about different users. Drawing on imaginary figures allowed the designers to explain and account for the design decisions they made. Relying on how the designers spoke about how they imagined users makes it difficult to assess how central they were to their design practices. Furthermore, there are limitations to analysing what designers explicitly say about users, as the more implicit ways they design for different users remains concealed. For example, Rommes (2002) explains how an awareness of gender equality makes it unlikely that designers will *explicitly* design differently for men and women. Drawing on Akrich (1994), I regarded anytime designers spoke about users as useful data. This revealed how the designers imagined users and how they drew on imaginary figures to make design decisions, shaping how meaning was encoded into data visualisations for different users. This might be through the selection of chart types, the level of interactivity, or how the visualisation led users through the data. In doing so, the designers configured the user (Woolgar, 1991), constructing limits on how different imaginary figures engaged with and made meanings from data visualisations. Therefore, despite their lack of physical presence, imaginary figures were productive in shaping the designers' practices.

During fieldwork, many of the designers at Evoke spoke passionately about the importance of designing for the user. However, their working understanding of diversity among users seemed limited to age, gender, and to a lesser degree, age-related disability. Analysis of the data revealed how diversity among users is constructed in the studio, not in terms of socio-cultural categories, but in terms of perceived *differences* between users' ability to engage with and make sense of data visualisations. Therefore, it is more useful to think about 'diversity' in data visualisation design in terms of the *limits of meaning making*. By limits of meaning making, I mean, how designers imagined different users' ability to engage with data visualisations and their capacity for understanding and making sense of graphs and charts. As argued, these different users were personified in imaginary figures. Based on the assumptions they made about them, the designers shaped their practices by encoding meaning into visualisations based on what *the public* will *like*, what *the professional wants*, and what *the grandma needs*. How the

designers construct these limits of meaning making can be mapped onto their different design priorities in response to imaginary figures. For example, engaging *the public* is related to what the user will *like*, speed and comprehension is what *the professional* wants, and accessibility is what *the grandma* needs. This demonstrates how designing data visualisations for imaginary figures shaped how meaning was made and encoded into the design of data visualisations at Evoke.

Chapters four and five have focused exclusively on how meaning is *encoded* into the design of data visualisations. However, the following chapter directs attention towards the meaning making practices of designers *and* users. In doing so, it addresses how ordinary users decode data visualisations through processes of *re-contextualisation*.

Chapter 6: Contextualising and re-contextualising data

6.1 Introduction

This chapter explains how data is contextualised during the design process before being re-contextualised by users during their interpretation of various data visualisations. Gitelman and Jackson (2013) argue that the misconception that data can be raw, that is objective and free from bias, frames data as somehow de-contextualised. This chapter draws on fieldwork data and interviews with ordinary users to demonstrate how both designers and users employ various strategies to contextualise the data, situating it in broader frames of reference in order to make meanings from it. This not only relates to the context within which the data was generated, collected and analysed, but involves situating the data in broader frames of reference to determine what it can tell us about the world. Reducing people, events, or phenomenon to numbers inevitably creates gaps in our understanding of them and removes any narrative from the phenomenon the data represents (Espeland, 2015). Contextualisation allows designers and users of visualisations to fill in these gaps in order to make meanings through and from data visualisations.

Designers at Evoke contextualised the data in order to make sense of what the data meant, to find a narrative in the data, and to account for why the data showed particular trends. They did this by drawing on resources beyond the dataset in order to flesh out the numbers and make them meaningful in relation to specific contexts. This included conducting research around the data, incorporating additional datasets into their analysis, speaking to the owners of the data to get clarity about what the data showed, or drawing on their own knowledge of a topic in order to find a narrative in the data. By contextualising the data in these ways, the designers made meanings from the data prior to its visualisation (also see Chapter four). Data from the user interviews revealed how participants used their own strategies to re-contextualise the data in order to make meanings from the visualisations I showed them. Without access to the dataset or the means to conduct research on the topic during the interviews, users tended to draw on their own knowledge to add context to the data. The strategies they employed included problem solving, imagining, and drawing on their personal experience of a topic to situate the data within broader frames of reference.

I have referred to the users' interpretive practices as *re-contextualisation* because of the contextualisation work already done by the designers of the visualisation. It is through these processes of *contextualisation* and *re-contextualisation* that designers and users construct

"meaningful discourses*"* (Hall, 1980, p. 130) through which to encode and decode data visualisations. By re-contextualising the data, the user interview participants situated it in frames of reference that were familiar to *them,* and which varied between individuals. Adding context also enabled the participants to account for why the visualisation showed particular trends or patterns in ways that supported their interpretation. This led participants to construct narratives that *reached beyond* (Dourish and Gómez Cruz, 2018) the information provided in the visualisation. Indeed, re-contextualising the data and accounting for what the visualisation showed was an important way in which participants made meanings from the visualisations I showed them. Therefore, for some participants, an inability to re-contextualise the data presented a barrier to meaning making.

There were notable similarities between how the designers contextualised the data and how users re-contextualised it. This underpins the chapter's argument that being able to put data into context is extremely important when making meanings from data through visualisation. It also challenges notions that visualisations could ever be "neutral windows onto data" (Kennedy and Hill, 2017, p.773), as subjectivities are layered into both the design and interpretation of data visualisations through these processes of contextualisation and re-contextualisation. This chapter addresses this in two parts. It begins by exploring how the designers at Evoke contextualised the data in order to make meanings from it as part of the visualisation design process. This discussion is framed around an interview with Jones, a designer who talked me through the visualisation of one particular dataset. The second part of the chapter explains how user interview participants re-contextualised the data in order to make meanings from the visualisations I showed them. It does so by focusing on three re-contextualisation strategies identified in the data: *problem solving, imagining,* and drawing on *personal experience*. The chapter argues that adding context to data is an important meaning making strategy in both the design and interpretation of data through visualisation. Therefore, understanding how different people add context to data is crucial to developing knowledge on how meanings are made from data through and from its visualisation.

6.2 Contextualising data in the design of visualisations

This section of the chapter draws on fieldwork data from Evoke to explain how designers add context to data in order to make meanings from it as part of the visualisation design process. During fieldwork, I was struck by how little data I saw during my observations. My own preconceived ideas about what visualising data might entail had led me to assume that the designers would regularly work 'hands on' with the data. This may be the case for lone designers, however, at Evoke, only a few people in the studio who possessed the specialist skills

to do this work performed data analysis. This meant that often designers worked on concepts for data visualisations without conducting analysis of the data themselves. Knowing the structure of the data, what the client wanted to know, or the story they wanted to tell was sufficient knowledge for the designers to begin designing a data visualisation. Despite what I perceived to be a *lack* of data in the studio, I observed designers attempting to contextualise the data they were designing a visualisation to represent. Indeed, the issue of "*context*" came up during conversations with designers in interviews and observations.

During an interview with Jones I was still unsure about how the data featured in the design process, I asked when starting a new project if they first receive some data. They explained to me that the initial concept design phase is "*data agnostic in a way*" because it focuses on what the client wants and what kind of data they are working with rather than "*drilling into the detail*" of the data, which comes later. They then explained how their knowledge of the dataset differed depending on the project, saying:

> Jones: "*I'd say there's like two cases, there's one case where the data is quite simple and it's already in use by the client, so you don't need to explore that data because you know it already means something, for example, a simple dashboard for a company that tracks real-time traffic. I don't need to prove that knowing how many cars are driving on a road is a meaningful metric. We already know that."*

Jones explained that, in this case, there is no need to "*explore the data*," they just need to know what it looks like in terms of the structure of the dataset. Knowing this allows the designer to design a visualisation that will represent it. In this example, the meaning of the data is already determined by how the client will use it; to track real-time traffic. This supports Gitelman and Jackson's (2013) argument that data is *framed* in relation to how it is used, and furthermore, demonstrates how the context of its use also determines its meaning. However, as Jones explained, without knowing how the data will be used it is necessary to *find meaning* in the data, saying:

> Jones: *"And the second case is when the client has some data which nobody has explored yet. We don't know what we want to say from it, so in that case, you have exploration. You have to try to find meaning from it* […] *and figure out what's interesting."*

In this second scenario, the designers possess less knowledge about how the dataset will be used or the story the client intends to tell from it. Jones explains that, in this case, they must explore

the dataset in order to "*find meaning*" in it. In this example, the context of the data is not anchored in its intended use and therefore its meaning is still to be determined. As this chapter will go on to explain, "*finding meaning*" and visualising it required the designers to flesh out the numbers by adding context to the data through various strategies. These included, conducting research around the data, incorporating additional datasets into their analysis, speaking to the owners of the data to get clarity about what the data shows, or drawing on their own knowledge of what the data represents. These strategies allowed designers to construct a narrative of the data in order to make meanings from it.

During a follow up interview, I asked Jones to bring a data visualisation they had worked on to discuss. During this interview, they began by telling me about a project they had worked on where the client had not already explored the dataset. Jones explained how working on this project had required them to explore the data to find meaning before they visualised it. As they explained the process, from exploring the dataset to designing a visual representation of their findings, they revealed the various ways in which they put the data into context. The following narrative from that interview is used to frame a discussion explaining how designers at Evoke contextualised data during the visualisation design process. This supports the argument made in Chapter four, that meaning is encoded into data visualisations in two parts, firstly through decoding the meaning of the data in order to translate it into a "meaningful discourse" (Hall, 1980, p. 130) and, secondly, through the design's visual elements which are determined during the design and production of data visualisations. Although much of the following discussion relates to scenarios where the designers were working with 'unexplored' datasets, it also draws on an example where contextualising the data was necessary, despite its meaning being pre-determined by its intended use.

During the follow up interview with Jones, they took the lead by talking me through some projects they had worked on. Although I asked questions about what Jones told me, to clarify points and to make sure I understood, I let them tell me about these projects in their own way. The first thing Jones wanted to talk about was the basic analysis they had completed at the start of a 'small' data visualisation project. Evoke had been provided with a dataset by a newspaper website who had surveyed their readers about their favourite celebrities. The newspaper wanted to know what Evoke could do with this data. Jones had been asked to look at the dataset to see if they could find anything *"interesting"* in it. Jones told me the data was static and stored in a basic table format, "*like an excel sheet."* They described the dataset, telling me the name of the columns in the table and the number of responses to the survey. Then, they explained:

> Jones: *"So the first thing you want to do is understand the data, and that means just getting a rough overview of what's in there."*

The survey had asked people to identify their favourite celebrities in five categories including singer and athlete. It had also asked for the respondent's age by year of birth and the dataset included the variable "*generation*." Jones explained that the first thing they had done on receiving the dataset was to print off the first five rows of data so that they could get a feel for what was inside. The data contained numbers, for example the respondent's year of birth, and text, for example, the names of celebrities. On reviewing the printed data, Jones said they had realised that it was not "*perfectly clean.*" Providing me with examples of why the dataset was not clean, they explained:

> Jones: *"Maybe someone has written Donald Trump with a T at the end, or a D at the end. Maybe there is some data missing. Maybe* […] *a birth date does not always match the correct generation."*

D'Ignazio and Klein (2020) explain that data cleaning has a mythical status, with data scientists lauded for their ability to take an untidy dataset and impose structure on it so that it can be analysed for insights. This view relies on the belief that "data are inherently messy and need to be tidied and tamed" before they can be analysed (D'Ignazio and Klein, 2020, p. 130). Although Jones was unable to correct all the typos, explaining that to do this would require them to go through the full dataset manually, they were able to address the incorrect categorisation of some of the respondents' generation. Interestingly Jones said they were uncertain how this variable had been generated, for example, if people had self-selected the generation they identified with. Those who had collected and generated the data would know this information and this highlights the way in which the designers at Evoke were often a few steps removed from the context of the data's collection (D'Ignazio and Klein, 2020). On reviewing the generation categories in the dataset, Jones discovered there were two generation X categories and went on to explain why this was likely to be unintentional, saying:

> Jones: *"Which is probably a mistake, because 'generation Y' is supposed to be 96 onwards and 'generation X' is supposed to be 81 to 95. So I remapped these so that the generation matched the correct date range* […]*."*

At first, this appears to be a simple amendment to the dataset to correct a mistake that occurred during data collection or the initial cleaning of the dataset. However, this revision affects the results of the data's analysis and subsequently, how the data was interpreted. Jones made a

judgement about "*re-mapping"* age and generation based on their understanding of how the dataset *should* look and what it *should* include. By identifying an anomaly in the form of a duplicate category, my understanding is that they added a new category to the dataset, generation Y. It is reasonable to assume that the inclusion of this category was the intention of the persons who designed the survey and Jones was simply correcting a mistake. Yet, Jones' brief account of cleaning the data reveals how some people's perspectives might be lost from the data during this process. In this case, those respondents who may have self-identified as a particular generation in the survey that differed to their categorisation based on year of birth, or those who spelt the name of their favourite celebrity incorrectly. However, it also illustrates how the perspectives of the analyst are *imposed* on the data through the pursuit of a tidy dataset that reflects the conventions of data science (D'Ignazio and Klein, 2020). Here, Jones drew on these conventions to re-construct contextual detail within the dataset, the kind of detail that would shape the meanings and subsequent narrative they derived from the data.

During a scheduled observation with Turner, I watched as they worked on a small numeric data visualisation for an existing client. Turner was having difficulty deciding how to present the data, as they were unsure what the data meant or what the client wanted to know. They explained to me that at the start of the design process you have to think about "*what is the data?*" and "*what does it mean?*". Turner told me that they had had to research some of the terms that the numbers represented. Despite researching what the variables meant, Turner was unsure how best to visualise the data. When they received the project they were provided with some example data to work with. This was in the form of a short list of five or six numbers, which they showed me on their computer screen. However, Turner explained that they did not know if the numbers were *good or bad*, saying rhetorically *"Is 1.7 good or bad?".* As Turner mulled over the design, I asked them, what is important to the client? Is it change over time? Turner said they did not know. They had decided that in order to make progress on the design they needed to know more about what the client wants to do with the visualisation.

This short account of my observation with Turner illustrates how important it is for designers to have an understanding of the meaning of the data they are visualising. It also reveals how this meaning is contextual. Without knowing what the data 'meant,' Turner drew on various strategies to try to contextualise the data in order to make sense of it. They conducted research beyond the dataset into what the numbers represented and expressed a desire to clarify what the client wanted to know from the chart. This suggests that context can be found in both what the data will be used for and what it *means* to the client. During the observation, Turner also told me that they had consulted a colleague about whether or not the value of the numbers would remain similar over time. In these three examples, Turner is attempting to situate the data in its

broader context. This was necessary because Turner was a *stranger in the dataset* they were attempting to visualise (D'Ignazio and Klein, 2020, see p. 133). They were not a member of the community of people who collected the data or for whom it would be meaningful. Therefore, they undertook strategies of contextualisation in order to try to make meanings from the dataset. Indeed, when Turner thought about if the numbers were "*good or bad",* they were attempting to situate the data in a broader frame of reference relating to its use. This need to contextualise the data in order to make it meaningful prior to visualisation challenges the idea that data can speak for itself (Anderson, 2008). It also illustrates how data is rendered meaningless when taken out of context (boyd and Crawford, 2012) and may also be more open to interpretation and, therefore, misinterpretation.

Once the dataset was clean, Jones explained that they could begin to ask basic questions from the data, for example, who are people's favourite celebrity? They explained how they had used analysis software to quickly generate some basic graphs from the data in order to try to "*get some insight*" from it. Jones showed me how they had plotted the top twenty favourite singers as a percentage of the overall respondents. This produced a graph showing how many people in total chose each singer as their favourite and what that number was as a percentage of the total respondents. Jones explained that they were able to get some insights from this simple graph, identifying that the favourite singer among respondents was Elvis Presley, the first Dutch band was ranked eleventh, and therefore the top ten favourite singers among respondents were from "*other nationalities*." The insights Jones took from the data were framed by the fact a Dutch newspaper website had conducted the survey. This perhaps explains why they identified a pattern in the data around the nationalities of the respondents' favourite singers. After Jones outlined the insights they had taken from the data, I clarified if this was from their interpretation of the graph they had generated, they replied:

> Jones: *"Yes. I tried to take three things I could deduce from the graph that might be interesting. Always keeping in mind that this is for a newspaper, so you want to have very simple facts that talk to people. And these people are Dutch so I thought, let's try to see this from a Dutch point of view."*

Kennedy and Hill argue that data visualisations cannot be considered "neutral windows onto data" (2017, p. 773). This quote from Jones demonstrates how their interpretation of the data is from a particular perspective that relates to the visualisation's target audience. They have chosen a "*point of view*" from which to make meanings from the data by drawing on the context of the data's collection and intended use. A plurality of meanings exist within datasets (D'Ignazio and Klein, 2016; 2020), yet the insights Jones identified and the subsequent

knowledge they produced is what Haraway (1988) describes as *situated knowledge*. This refers to the way in which knowledge can only ever offer a partial perspective, situated within the social and cultural contexts of its creation. This includes Jones' individual position as a young, skilled, professional person. Although Jones may draw on their cultural understanding of a "*Dutch point of view*," they cannot represent the views of all Dutch people. Nevertheless, they draw on this perspective to help them to contextualise the data.

As Jones talked about what insights they could take from the graph that could be interesting, they suggested they could explore if there were patterns in where the favourite celebrities came from or look at the impact of the United States on the Dutch. I took the latter to mean the United States' impact on Dutch popular culture. In doing so, Jones was contextualising the data by situating it in broader frames of reference around the globalisation of culture in order to develop a narrative around what the data showed. Dourish and Gómez Cruz (2018) argue that data becomes meaningful through its narration. However, the stories people tell from the data are not wholly dependent on the dataset itself. Rather, people draw on existing "culturally available" subplots, which allow them to develop a narrative that is recognisable (Dourish and Gómez Cruz, 2018, p. 7). By contextualising the data in relation to broader debates around globalisation, Jones is identifying a narrative that is familiar to *them* and perhaps to the intended audience of the visualisation. In doing so, they situated their analysis of the data within *meaningful discourses* (Hall, 1980) that allowed them to make meanings from it that could be decoded by the intended audience.

All data visualisations have a rhetoric, which Hullman and Diakopoulos (2011) describe as their intended meanings, which are determined by the choices the designers make during the design process. They explain how these choices are also subject to the characteristics of the intended user and "societal and cultural codes" (Hullman and Diakopoulos, 2011, p. 2232). Indeed, by imposing a 'Dutch' perspective on the data, Jones is imagining what insights from the data the intended user might find interesting. This is not so much about looking into the data for meaning but about stepping back from it to speculate what someone else might find meaningful. This required Jones to situate the data in broader social and cultural contexts in order to make judgements about which insights the intended audience would be more likely to engage with. This relates to the discussion of the imaginary figures in the previous chapter (Chapter five), which indicated how engagement was prioritised in the design of data visualisations aimed at *the public*. This suggests that contextualisation not only allowed Jones to make meanings from the data, but also determined *which meanings* would be visualised based on what they thought the intended user would find engaging. Therefore, strategies of contextualisation frame the rhetoric of data visualisations. Hall (1980) argued that encoding requires producers to make

assumptions about their audience (see Chapter five). Contextualising the data with the user in mind situates the data in meaningful discourses that intended users can decode. Of course, how the intended user actually decodes a data visualisation may not always be as the designer intended (Hall, 1980).

Next, Jones incorporated the generation parameter of the survey to their analysis of the data. They explained that this allowed them to see how what they had plotted earlier was impacted by crossing it with the generations of respondents. They had plotted the top five most popular singers and then broke this down per generation. They showed me how the graph they plotted was in order of generation, pointing out the oldest and the youngest generations. By incorporating this new parameter, Jones added *characters* (Dourish and Gómez Cruz, 2018), in the form of older and younger users, to their narrative of the data. They drew on these characters to explain the spread of favourite singers among the survey's respondents. Jones explained to me how the low bars of the 'Generation Y' graph "*can*" give the impression that:

> Jones: *"the older people are more focused or more passionate about one person, whereas the younger generation are much more spread around. Probably because they hear so much about so many celebrities that it's not like they have three idols anymore like our grandparents could have."*

Jones was tentative when making this claim, however, despite their initial caution they make inferences to account for why younger people's responses to this question were more diverse than those of older respondents. They suggest that this could be because younger people are exposed to more celebrities than their grandparent's generation would have been. It is unclear what Jones was drawing on to make these inferences but it could be that they were drawing on their prior knowledge or own experience of popular culture to make sense of the differences in responses between the generations. Either way, in order to interpret the chart they had produced, Jones added context to the data by drawing on resources beyond the dataset. There was no information within the dataset to *account for why* differences existed between the responses of older and younger people. Yet, making inferences about why this might be the case was significant to how Jones made meanings from the data. By contextualising the data in this way they began to construct a narrative of the data that "extends the data's reach" (Dourish and Gómez Cruz, 2018, p. 4). Significantly, many participants of the user interviews drew on the same strategy to add context to the data in order to make meanings from the visualisations I showed them.

During my previous interview with Jones, I had asked them how they knew when they had

found something interesting in the data. In response, they said that there is "*gap*" between finding something interesting and proving that it is "*actually true*." As an example, they explained how *googling* a particular date to find out why the data showed so many people running in a location might reveal there was a marathon that day. In this example, Jones suggested by conducting research to add context to account for why there had been a spike in runners on a particular day, might confirm the accuracy of what they had found in the data. I was intrigued by the idea that designers might have to do research beyond the dataset in order to make sense of it. After asking Jones about this, they explained why they might conduct research beyond the data, saying:

> Jones: *"Because you always need to put your data in context. [….] you need to make sure that the link between what you find is true, or has meaning, or has potential for more research* […]*."*

Jones' quote illustrates how the meaning of data is not exclusively found in the dataset. Interpretation of data requires an understanding of its context. By contextualising data, designers are able to make meanings from what it shows. This might involve drawing on prior knowledge of the subject the data represents to make decisions about how to analyse and interpret it. Or, it might involve cross referencing different datasets or googling specific dates to account for anomalies in the data. The need to conduct research beyond the data in order to make it meaningful was something that I observed during fieldwork.

During an observation with Reed they showed me a data visualisation project they had worked on. The visualisation was of wildlife data provided to them by the organisation who had collected it. Reed explained that they had been given a considerable amount of data about lots of different animals and the initial response of the designers was to design something on a large scale to represent it. However, on reflection they decided that something on a smaller scale could be more powerful. Reed explained how during the project, when they found anomalies in the data, they would contact the organisation to ask if they could explain them. As the organisation was familiar with the data, they could look back at their records to find out what happened on the days the anomalies occurred. I asked Reed if doing this had brought context to the data. They agreed, adding a few moments later that it is all about context. Reed said that quantitative data can tell you what is happening but not *why*. Indeed, accounting for why the data showed what it did was an important contextualisation strategy that was employed by both designers and user interview participants when making meaning from data. By contextualising and re-contextualising data, both the designers and the users were adding qualitative detail to aid their understanding of the quantitative data. This is necessary because reducing people,

events, or phenomenon to numbers simplifies what the data represents and in doing so erases its narrative (Espeland, 2015).

Jones plotted another graph to see if they could "*prove*" that older generations were more likely to identify fewer favourite celebrities than younger generations. This chart explored the different categories of celebrity in the survey by generation. It also explored how many different celebrities were reported in each category. Jones showed me the graph and pointed out the peaks and troughs, before saying that the graph "*can mean two things*." Firstly, that the younger generation were not well represented in the survey; with less data on younger peoples' preferences, the concentration around a few specific celebrities did not occur. An imbalance in the age of respondents had been a concern that Jones had raised earlier in the interview. I do not know if Jones had checked for this in the data when they had initially worked on the project and regrettably, I did not ask. Despite their concerns about the sample, Jones said they were not satisfied with that conclusion, saying:

> Jones: *"But I wasn't really happy with this way of thinking, I wanted to think that what's interesting here is that the younger generation have more celebrities because they're exposed to many more distinct celebrities these days, because of the internet, because of communication, because of everything we're bombarding them with, new stars all the time. So that's the first thing."*

Here Jones is repeating their earlier explanation for why the data indicated younger generations reported a greater spread of favourite celebrities than older generations. While the data had indicated this, it did not include information to explain *why* this was the case. Therefore, to account for what the data showed Jones made hypotheses. In Chris Anderson's (2008) widely cited and critiqued article proclaiming the end of theory, he argues that with enough data correlation can be meaningful without causation. Whereas the previous conventions of science involved hypotheses, modelling, and testing, Anderson (2008) argues that big data science could reveal what was happening without the need to explain *why* it was happening. However, as Jones and Reed demonstrated, accounting for why the data showed particular peaks, troughs, or spikes is an important way in which they made the data meaningful prior to its final visualisation. In order to do this, Jones contextualised the data by situating it within broader debates about the rise of celebrity and how young people experience popular culture through the internet. These are arguably *meaningful discourses* (Hall, 1980) that Jones constructed in order to make meanings from the data that could be decoded by the intended users. In doing so, they went on to make an alternative hypothesis to explain the data:

> Jones: *"And the second thing I thought is, the younger generation are young so it's not very clear to them yet who their role models are, or who their favourite singer is, because when you're young you like many things."*

It is unclear what Jones is drawing on to make this inference from the data. It could be their own personal experience, or they could be drawing on their knowledge of young people they know personally, or through media portrayals. Nevertheless, what is significant in both of these hypotheses is the narrative that Jones is building around what the data is showing them, which in turn "extends the data's reach" (Dourish and Gómez Cruz, 2018, p. 4). It did not appear to be sufficient to say the data shows X or Y. Instead, there was a desire to explain *why* the data shows this. Jones made these hypotheses in order to make meanings from what the data showed in the graphs they produced during their analysis. However, they are also keeping in mind the audience for whom they are designing the final data visualisation. This represents a layering of subjectivities in their analysis, in which they drew on their own knowledge, broader social and cultural frames of reference, and their imagined user to make meanings from the data. It is these subjectivities that form the rhetoric of data visualisations (Hullman and Diakopoulos, 2011), which are then encoded into their design.

During the interview, Jones showed me several charts and graphs they had made to analyse the dataset. In interpreting these graphs, Jones drew on some of the same strategies user interview participants did when I showed *them* data visualisations. Indeed, accounting for what the visualisation showed was something many of my user interview participants did. Like Jones, in order to do this they added context to the data by situating it in broader frames of reference and I will discuss this further in the next part of this chapter. Despite Anderson's (2008) claim that with big data causation becomes redundant, it was an important part of how designers and users made meanings from data visualisations. However, it is important to note that, for Jones, the meanings they made from the dataset were interesting hypotheses that may have turned out to be inaccurate. They explained that they would be hard to prove but that this project was not for a "*scientific paper*," so they were "*allowing*" themselves to make these hypotheses with a caveat that would allow the user to *decide for themselves* if they could be true.

Contextualising the data by situating it in broader frames of reference allowed Jones, and Reed, to tell a story about what the data revealed. During my interview with Clarke, they described how, as data visualisations have become more common and prolific in the media, it is no longer enough to have large quantities of data and to visualise it. They explained that, "*the pressure's on to come up with a good story.*" This is presumably to ensure user engagement with the visualisation and is reflected in the growth of data journalism, which uses data to tell stories

(Weber, 2020). Clarke said that this means that designers have to have an interest in, or background knowledge of, the subject matter the data represents in order to "*come up with the narrative*." This, they said, means that they "*learn a lot about a lot of things*" through their interaction with diverse datasets. From what Clarke said about the difference between dashboards and more stable data visualisation projects, I have inferred that when Clarke said this they were referring to more stable data visualisation projects such as the one Jones talked me through. They explained how finding a narrative might require people who have knowledge of the topic the data represents, saying:

> Clarke: *"It also means that you'll need more people sort of interested, or that have background knowledge of a certain field in order to come up with that narrative, because I might not know what's trending or relevant within that industry."*

It is not clear from this quote if by "*people*" Clarke means designers with an interest or knowledge in the subject or if this might require the help of people with more intimate knowledge of the data. The latter was the case when Reed contacted the organisation who provided the wildlife data they were working with to explain anomalies. In that example, the explanation of an anomaly in the data led to the narrative of the visualisation Evoke produced. Nevertheless, this challenges Anderson's assertion that "With enough data, the numbers speak for themselves" (2008, n.p.). Indeed, Dourish and Gómez Cruz argue that such claims are "one of the most powerful narratives about data" (2018, p. 8). It seems that the meaning of data and their visualisation are found in contextualising the data, which requires human interpretation. Those with knowledge of the data or topic it represents are able to bring context to its interpretation, which helps to construct a narrative. Of course, this knowledge is always *situated* (Haraway, 1988) and represents the partial view of those who are in the powerful position to determine the meaning and narrative of the data (Beer, 2019; boyd and Crawford, 2012). However, contextualising the data does not only rely on the designers' prior knowledge of the dataset or topic it represents, it can also involve combining additional datasets in order make meanings.

Despite Jones' initial interest in how the respondents' favourite celebrities varied between generations, they continued to test different hypotheses. This required them to conduct research in order to incorporate additional data which they sourced by conducting research beyond the dataset they were provided with. For example, they explained that to determine if there was a pattern in the nationality of the respondents' favourite sports stars, they had googled the nationalities of the top twenty most popular sports people. To analyse if respondents' tended to prefer celebrities from the same or different generations than their own, they searched on the

internet for the date of birth of popular celebrities. Finally, Jones said they had included the gender of the respondents' favourite celebrities, to explore if and how the gender split differed between generations. In all of these examples, Jones added context by incorporating new data into the dataset. The selection of the new data was informed by culturally available discourses around age and gender that would allow Jones to make meanings from the data that the intended users would be able to *decode*. However, by changing the parameters of the data Jones added "strangers in the dataset" (D'Ignazio and Klein, 2020, p. 133). Although this strategy provided new context to the dataset, it further removes the data from the context in which it was generated. This changed the meanings that Jones could make from the data and allowed new meanings to be made that went beyond the information within the original dataset.

The practice of contextualising data in order to make meanings from it challenges notions that data can speak for itself. The designers at Evoke used various strategies to contextualise the data. This included drawing on their knowledge of a subject, conducting research into the variables, and even adding new variables to a dataset. Often, by adding context to the data, the designers were able to construct a narrative that explained why it indicated certain trends or patterns. This allowed them to flesh out the numbers and make meanings that reached beyond the information contained within the 'raw' dataset. In this sense, contextualisation "extends the data's reach" (Dourish and Gómez Cruz, 2018, p. 4). This is possible because data has no inherent narrative of its own, therefore it is repurposed as it circulates to tell *new* stories about what it means (Espeland, 2015).

The contextualisation of data also reveals the labour required to make meanings from data and the choices made during the analysis and design process that shape the rhetoric of data visualisations. However, the people, labour and decisions through which data is contextualised are rarely translated into the design of a graph or chart (D'Ignazio and Klein, 2020; 2016). Indeed, the conventions within the field of data visualisation design arguably conceal the contextualisation process through their use of "clean layouts," "geometric shapes," and "two-dimensional viewpoints" (Kennedy et al., 2016a, p. 716). It is these very conventions that give data visualisations their "aura of objectivity" (Kennedy et al., 2016a, p. 723). Yet, as this chapter has illustrated, the process of contextualisation layers the data with subjectivities and produces visualisations that are situated within the specific contexts of their creation. This includes the way the data will be used, the inevitably partial perspectives of the designer who interprets the data, and the broader contexts within which the data is situated in order to construct a narrative about what the data can tell us about the world. The following part of the chapter turns its attention towards how users decode data visualisations. It argues that user

interview participants *re-contextualised* the data in order to make meanings from the visualisations I showed them.

6.3 Re-contextualising data in the interpretation of data visualisations

This section of the chapter draws on interviews with ordinary users to analyse how they made meanings from the data visualisations I showed them. It reveals how participants made meanings from data visualisations by *re-contextualising* the data represented in the graphs, maps, and charts they interpreted. It demonstrates how, like designers, users drew on various strategies to put the data into context in order to interpret the visualisation. This involved the users situating the data within frames of reference with which they were familiar. Re-contextualising the data in this way allowed the users to make inferences and tell stories from the data that reached beyond what the content of the graph or chart showed (Dourish and Gómez Cruz, 2018). Nafus (2014) found that for her participants, the sensor data they had collected was "always pointing to something beyond" (2014, p. 220). By making these inferences, users were able to reach beyond the information in the chart to account for *why* the data visualisation showed what it did and this was an important meaning making strategy among my interviewees. Analysis of the user interview data identified three strategies of re-contextualisation used by the participants. These were:

Problem solving

In order to situate the data into broader frames of reference, the participants often identified patterns in the data and drew on their prior knowledge of the subject the data represented to explain them.

Imagining

Participants drew on their imagination in a number of ways to re-contextualise the data. These included imagining how the data was generated to account for what the visualisation showed, imagining who made the visualisation and their motivation for doing so, and imagining the intended user or intended use of the visualisation.

Personal experience

Participants made meanings from the data by drawing on their personal experience of the topic the visualisation represented. This allowed them to re-contextualise the data to make sense of what the graph or chart showed.

By drawing on one, or a combination of these re-contextualisation strategies, participants were able to make meanings from the data visualisations I showed them. I am referring to this process of meaning making as *re-contextualising data,* because as the first section of this chapter demonstrated, contextualisation work is done during the design of data visualisations. In addition, using this term recognises that in interpreting the visualisation, users may not contextualise the data using the same frames of reference as those of the designer. The following section of this chapter will address each of these strategies in turn to explain how the participants re-contextualised the data in order to decode the visualisations I showed them. It also addresses instances in the data where participants struggled to re-contextualise the data and argues how this presented an obstacle in their ability to make meanings from the visualisation.

6.3.1 Problem solving

Many of the participants used *problem solving* techniques to re-contextualise the data within the visualisations I showed them. These techniques predominantly involved identifying patterns in the data and drawing on prior knowledge of the topic the chart or graph represented to explain these patterns. As the following data extracts illustrate, this sometimes involved situating the data within broader frames of reference including social, political, economic, and historical contexts. This allowed participants to make inferences to account for their interpretation of the visualisation. However, the assumptions they made were rarely informed by the data alone but were constructed from the resources they had drawn on in re-contextualising the data. Therefore, by using problem solving techniques, participants drew on their own prior knowledge in order to make the visualisation meaningful through narratives that went beyond what could be concluded from the data represented in the graph or chart.

Ken was a retired man in his sixties who used to work in the finance sector. During our interview, I showed him a data visualisation that represented rates of antibiotic resistant E.coli in Europe. The chart resembled a pie chart, however each segment of the circle was the same size and each represented a different country in Europe. The countries were placed on the chart so that they were grouped into Northern Europe, Western Europe, Eastern Europe, and Southern Europe. The key situated to the side of the chart indicated that the density of colour within each segment represented the rate of antibiotic resistant E.coli for that country. This ranged from light pink (low rates of antibiotic resistant E.coli) to very dark pink (high rates of antibiotic

resistant E.coli). Each segment was labelled to show which country it represented and additional context was provided in the form of red or blue arcs above some of the segments, which indicated if that country's rate of resistant E.coli had increased (red) or decreased (blue) between 2012 and 2015. Finally, the chart was interactive, allowing the user to click on a segment to bring up a list of annual percentages from 2012 to 2015. When I asked Ken to tell me what this data visualisation was about, he explained:

> Ken: "*It's about the resistance of infections [..] to antibiotics. And it basically tells you which countries where it has become more prevalent* [in terms of the] *resistance to antibiotics and those that are least."*

In response to my asking him, Ken said this was not a topic he was very familiar with. Despite this, Ken was able to draw on problem solving techniques to situate the data within more familiar frames of reference. The caption accompanying the chart explained the highest percentages of resistance were found in Southern and South Eastern Europe. Ken refers to this

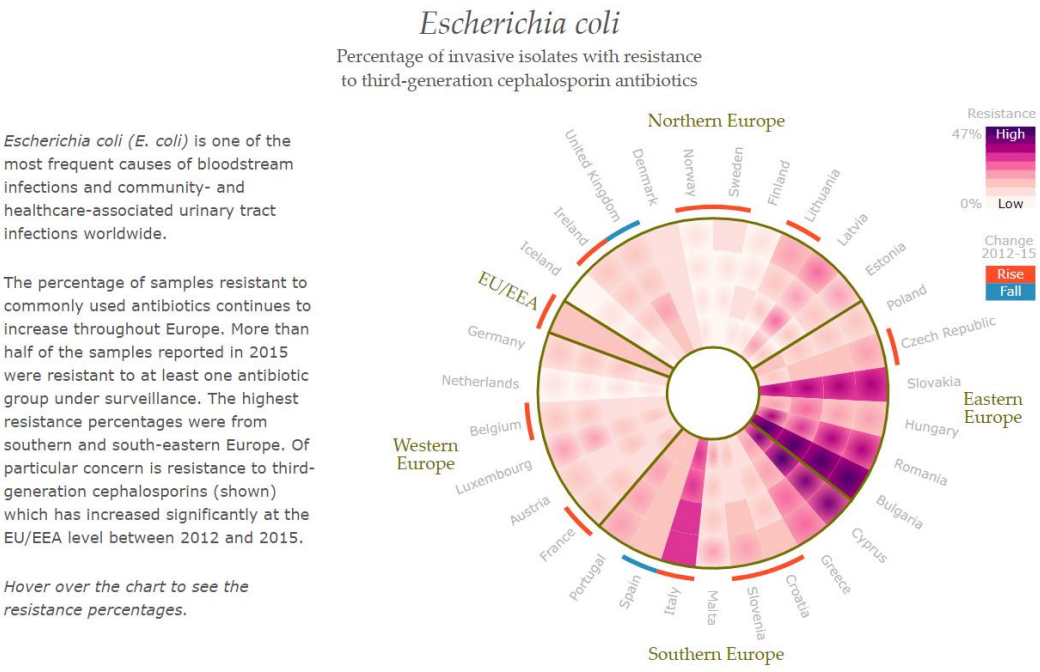

Figure 1: Screenshot of Escherichia coli from Bacteria vs. Antibiotics: Antibiotic resistance across Europe. (2016). Developed by Dr Cath Sleeman and produced by Nina Cromeyer Dieke, Nesta Longitude Prize. [online interactive data visualisation] https://s3-eu-west-1.amazonaws.com/nesta-data/amr-2016/index.html Printed with permission. CC BY-NC-SA 4.0

before saying that those countries were "*probably a bit warmer*" than one of the countries with the lowest rates of resistance, "*Iceland*." By drawing on both the caption and the chart, Ken

starts to associate the warmer climates of particular countries with overall rates of infections, stating:

> Ken: *"So there's an inference in effect that the warmer the temperature the more prevalent it is for blood infections to form. Which would lead me to believe that the infections multiply quicker in a warmer climate than a colder climate."*

The data visualisation offers no information regarding the overall rates of *infections* in different countries or why some countries have higher rates of antibiotic resistance than others. Nevertheless, by drawing on his knowledge of the climate of different European countries, Ken identified a pattern between the countries with higher rates of antibiotic resistant E.coli, namely, that they have warmer climates than those with the lowest rates of resistant infections. Although Ken is drawing on prior knowledge to re-contextualise the data, it is also possible that he associated the *warmer* pink colours, indicating higher rates of resistance, with temperature. It is impossible to know how significant the colours were to his interpretation of the chart, however using warm colours to indicate heat and cooler colours to indicate cold is a widely accepted use of colour in data visualisation design (see Kirk, 2016). Indeed, Kennedy and Hill describe colour as a "powerful sensory cue" that designers use to signify meaning (2017, p. 772). However, Ken went on to describe his thought process as he continued to make meanings from the chart. While briefly mentioning that the chart showed infections were less likely to spread in developed countries, with better healthcare, he returned to the issue of climate and in doing so, identified the importance of "*logic*":

> Ken: *"*[…] a*nd my logic tells me that if this country is warmer, then infections and germs will spread quicker and more easily than if it had been the cold countries.* [Because] *They just can't multiply, they need an ambient temperature."*

The previous two quotes demonstrate how Ken was aware that he was making inferences from the chart and recognised how "*logic*", or *problem solving* as I refer to it, played a role in how he decoded the visualisation. However, the inference he made and the "*logic*" he used to re-contextualise the data reveals a disjuncture between what the data visualisation represented and the meanings Ken made from it. The data represented in the chart refers only to rates of antibiotic resistant E.coli per country. However, by identifying patterns in the data relating to climate, Ken made sense of the visualisation in relation to the rates *and spread of infections* in different countries. This was despite the chart containing no information about overall rates of infections or *how* E.coli infections spread. Yet Ken had situated the data within broader frames of reference with which he was familiar, in order to account for *why* the data visualisation

showed countries in South and South Eastern Europe having higher rates of antibiotic resistant E.coli infections. This allowed him to make the data visualisation, which was on a topic he said he was unfamiliar with, meaningful to him.

In *Theory of Media Literacy*, Potter (2004) explains how, when constructing meaning, people will fill in missing information by drawing on their personal knowledge or by accessing other information that helps them to do this. Although not referring to data visualisations, Potter's (2004) theory explains, from a cognitive perspective, how people make meanings differently because of the individual knowledge and information they bring to a problem. Although, another participant, Rebecca, utilised a similar problem solving re-contextualisation strategy as Ken, they made different meanings from the same visualisation. Rebecca made meanings from the chart by situating the data within broader historical and political contexts to explain the higher rates of antibiotic resistant E.coli in Eastern European countries.

Rebecca was a woman in her forties and a member of a social justice group. When I showed her the antibiotic resistance chart, Rebecca identified the topic before remarking that "*Scandinavian countries*" had the "*lowest levels of resistance*," but that "*Eastern Europe's problematic*." I asked her if this is what she expected and if antibiotic resistance was a topic she knew a lot about. Rebecca explained that although she did not know a lot about the topic, some of her friends had worked abroad treating people with a multi antibiotic resistant infection, so she was aware of the problem of antibiotic resistance. Rebecca then said:

> Rebecca: *"I really can't fathom, other than the fact that potentially people born under the communist era, where healthcare wasn't great, versus people living in a high social welfare area for a long time, like Scandinavian countries. That's the only thing I can think, that people have been able to look after themselves better in the North."*

Although I never asked participants *why* there was a difference between countries' rates of antibiotic resistant infections, it was a common response among participants to try to account for why this might be. As noted earlier in the chapter, Espeland (2015) suggests that because data has no narrative it evokes new ones that allow people to explain what it means. Here, Rebecca attempted to make sense of the differences between Scandinavian countries and Eastern European countries by drawing on her knowledge of the political and social history of Europe. By doing so, she reasoned that poorer social welfare provisions in post-communist countries may have led to greater levels of antibiotic resistant E.coli. Rebecca had re-contextualised the data by situating it in a broader historical and socio-political context in order to try to make sense of what the chart showed.

Both Rebecca and Ken identified patterns in the data and drew on their own knowledge of European countries to re-contextualise it. They situated the data in a wider geographic context (in Ken's example) and in a historical and socio-political context (in Rebecca's example) in order to develop a narrative which explained what the visualisation showed. Indeed, constructing narratives provides a means through which to situate the data in particular contexts to make it meaningful (Dourish and Gómez Cruz, 2018). However, Dourish and Gómez Cruz (2018) argue that this relies on people drawing on stories that are *culturally available* to them. In this example, these are wider narratives around climate and the socio-political histories of different countries represented in the chart. This demonstrates how users may draw on very different discourses in order to re-contextualise the same data visualisation. While Hall (1980) recognised that audiences may decode messages in different ways, he argued this did not represent a plurality of meanings. While encoding might limit the meaning of data visualisations, this chapter reveals the strategies users drew on to decode them. This often involved weaving culturally available discourses with personal experiences, prior knowledge, and even their imagination.

The desire to account for why a visualisation showed what it did was reflected in some of my participants' frustration with the gender pay gap graph I showed them. This graph was originally published by The Guardian newspaper and represented data on over ten thousand companies. Each company was represented by a dot that was positioned on the chart to show if the company had reported a higher or lower median wage for women compared to men, and by what percentage. Unusually, the chart only had an x axis, and zero was positioned in the centre of the axis. If companies were positioned to the left of the zero they paid women more than men and if they were positioned to the right of the zero they paid women less than men. The further away from the zero the companies were positioned, the higher the percentage difference in pay. Without a y axis, it was unclear if the height of the dots was meaningful. The graph used colour to make the pay disparity clearer, with companies paying women more than men represented with blue dots and those paying women less than men with orange dots. Significantly, this graph included additional contextual detail by highlighting the names of fourteen of the companies and their position on the graph (four who paid women more and ten who paid women less). However, this additional *messaging* (Seer and Hegel, 2010) did not stop participants attempting to re-contextualise the data in relation to their own knowledge of the topic.

Figure 2: Image removed due to copyright. This image showed a screenshot of '10,016 companies have reported pay figures' data visualisation.

Mary was a woman in her seventies and a member of a social justice group. After engaging with the gender pay gap data visualisation, Mary said:

> Mary: *"It's a big sample, over 10,000 companies and their reports on gender pay figures, or inequalities. The first glance indicates that overall women are paid less than men. Although in a few representations of the figures women are paid more […]. But there aren't enough spotlighted names of companies within the many results that are there to indicate a category of why."*

This quote demonstrates how Mary had identified the topic of the visualisation and that the graph showed women are more likely to be paid less than men. However, it also reveals how she attempted to use the highlighted companies to account for *why* women are paid less or more in particular firms. This is a further example of how participants re-contextualised the data by identifying patterns and drawing on their prior knowledge of a topic to account for why the visualisation showed what it did. Despite Mary's initial difficulty, she then used the highlighted companies to help her re-contextualise the data in relation to different sectors of employment, saying:

> Mary: "*Although perhaps it's significant that Royal Trinity Hospice and British Heart Foundation, as caring companies or organisations are actually on the women paid more side. And so, it looks like, banking and investment and big companies are definitely in the women paid less side."*

It is interesting that Mary drew on the Royal Trinity Hospice and the British Heart Foundation, which she described as "*caring companies*", to identify a pattern among the data. It is possible that Mary was drawing on her experience of and interest in social justice to help her to re-contextualise the data in the graph. Indeed, Dourish and Gómez Cruz (2018) argue that when constructing narratives from data people do not only rely on the data but also draw on stories that are familiar to them to make sense of it. Although this graph provided some contextual detail by including the names and positions of a selection of companies, Mary still re-contextualised the data by situating it within frames of reference within which she was familiar. In doing so, she appeared to draw on a discourse around '*caring*' *versus corporate* companies, suggesting that what is a *meaningful* discourse (Hall, 1980) is personal to the user. It could be argued that the designer of the graph intended users to make the connection between charities paying women more than men and corporate businesses paying women less than men. However, a restaurant chain was also highlighted in the women paid more side of the graph, while the Guardian News and Media and Eton College were highlighted among the companies where

women were paid less. Using problem solving techniques and drawing on her knowledge of different employment sectors, Mary was able to re-contextualise the data in order to make meanings from the graph. Significantly, this involved accounting for why the data showed what it did. This reinforces the argument that re-contextualising data and constructing narratives that account for what the graph showed were important meaning making strategies when users engaged with data visualisations.

While the gender pay gap graph included some contextual detail, the carbon emissions chart I showed participants provided them with very little context. This visualisation could be described as a bubble chart and comprised of coloured circles positioned on a map of Indianapolis, a city in the United States. The colour of the circles indicated the source of the emissions, for example, residential, commercial, industrial, and transport. While the size of the circle indicated the amount of emissions; the larger the circle the greater the emissions. The circles were positioned over a simple street map of Indianapolis to indicate the location of the CO_2 source. The static, printed version of the visualisation I showed participants did not provide the unit of measure used to quantify the emissions. The lack of contextual detail provided in the visualisation led participants to make assumptions about what the source of the emissions might be, for example, suggesting a large transport emitter might be a bus station, an airport, or a train station.

Figure 3: Image removed due to copyright. The image showed a screenshot of 'Carbon Emissions in Indianapolis' data visualisation.

Some participants even used the placement of the circles to draw conclusions about the layout of the city, for example identifying commercial areas, in order to make meanings from the chart. When I showed this chart to Glen, a man in his late thirties and a member of a sports club, he seemed frustrated by the *lack of a pattern* in relation to where the industrial emitter bubbles were situated on the map. Glen had expected them to be situated in close proximity to one another whereas they appeared to be dispersed across the city. As Glen talked more about the visualisation he identified where areas of employment might be on the map, based on the positioning of the bubbles. I said that I found it interesting that he was making broader assumptions from the chart beyond what it showed. Glen replied:

> Glen: *"Well to get the full picture you have to make certain assumptions because of what's not shown. Otherwise, it could just be a child flicking dots of paint onto a piece of paper. You need to get the full understanding, so I'm making assumptions to try to*

understand why there are certain patterns, or lack of with regards to the industrial estate."

In this quote, Glen explained how he was making assumptions in order to account for patterns in the data because of "*what's not shown."* This suggests that, by attempting to re-contextualise the data, Glen and other participants were filling in perceived gaps in the information provided in the visualisation with their own knowledge. Amoore explains that visualisation results in "gaps and invisibilities" which are "filled in by the observer" (2009, p.23). Amoore was writing in the context of visualisations of people, constructed through various sources of data about them and under the guise of national security. Nevertheless, Glen's quote demonstrates how these same gaps and invisibilities are present in other forms of data visualisation. Although the context is different, the same data practices of simplification and reduction result in people filling in the gaps with their own situated knowledge and partial perspectives. The user interviews provide some insights into the re-contextualisation strategies people draw on in order to fill these gaps in knowledge. This might involve participants trying to identify patterns in the data that fit into frames of reference that are already familiar to them.

Glen held preconceived ideas that led him to believe the industrial emitters should be concentrated in a particular area or areas of the city. Yet, he could not identify a pattern among the industrial emitters in the chart that resembled his understanding of an "*industrial estate*." Glen worked as a buyer in the building industry and it is possible he was attempting to draw on his working knowledge of industrial estates to re-contextualise the data and make sense of the chart. However, this particular attempt at problem solving, by identifying a pattern and accounting for it by drawing on prior knowledge, was not successful. There was a discord between Glen's knowledge of 'industry' and what the chart showed through the placement of the industrial emissions bubbles on the map. This demonstrates how re-contextualisation strategies are not always successful. Despite this, many participants used problem solving to re-contextualise the data in order to make meanings from the visualisations I showed them.

Problem solving techniques tended to involve identifying patterns in the data and drawing on prior knowledge of the topic to explain these patterns. This included drawing on the *meaningful discourses* (Hall, 1980) available to them which, as Ken and Rebecca showed, differed between participants. In doing so, participants situated the data within frames of reference that were familiar to *them*. This allowed them to construct a narrative, which explained why the visualisation revealed certain patterns or trends, thus filling in the gaps in information presented in the data visualisations (Amoore, 2009) in order to make meanings from them. This demonstrates how the meaning of a visualisation cannot be reduced to the image and its

semiotic modes. Meaning is made in the interaction between the image and the ordinary user and is framed by the individual knowledge that a user brings to the interaction.

6.3.2 Imagining

Imagining was another strategy used by participants to re-contextualise the data in order to make meanings from the visualisations I showed them. The interview data revealed a range of ways in which participants employed their imaginations in interpreting data visualisations. These ranged from imagining the subjects of the data to imagining who designed the visualisation, the motivation for doing so, and even the intended audience. This section of the chapter explains how participants drew on their imagination to re-contextualise the data. Drawing on interview data it will demonstrate the various ways in which participants used their imagination to make sense of visualisations I showed them. As with the problem solving strategy, this often led participants to situate the data in frames of reference they were familiar with and to *account for why* the visualisation showed what it did. Both were significant factors in how many of the ordinary user participants made meanings from data visualisations.

Two of the data visualisations I showed the participants focused on the tracking data of individual subjects. These could be described as *small-data* visualisations (Simpson, 2020; Lupton 2018). The first of these was an animated visualisation representing the movements of two female lions within a national park in Kenya, and the second comprised of several visualisations representing one individual's Fitbit tracking data. When interpreting these visualisations, some participants *imagined* the subjects of the data in order to add context to what the visualisation showed. The lions in the visualisation were represented by yellow dots, which moved around the screen leaving yellow lines that showed their movements over time. The dot and line were overlaid onto an aerial image of their territory. As the animation played, a narrator explained the lions' movements and provided contextual detail about their activities. This included clarifying why the dot representing one of the lions moved back and forth over the same area many times creating a thick yellow line on the aerial background. The narrator explains how one of the lions had become stuck behind a newly repaired electric fence, which blocked her from being able to get back to her sister.

Despite there being no image of a lion in the visualisation, Cynthia imagined the lion's distress at being trapped behind the fence. Cynthia was a retired woman in her seventies, who had worked with computers and with graphs and charts in her career in logistics. After watching the animation I asked her if she had found the visualisation informative. She replied enthusiastically, saying *"Yes I could imagine it! Yeah, it was very descriptive."* I then asked

Cynthia if the visualisation would have worked without the voiceover. She explained that you would be able to get the "*gist*" but that you would have to know the fence was broken, because you would not know that from the visualisation. However, she then explained that without the voiceover, you would still know the lion was distressed, saying:

> Cynthia: *"You'd see her become stressed, because if,* [short pause] *well you just know because you'd have the visual of them sort of, because they pant a lot and things like that. And the fact she'd wanted to go back to another animal, you'd get that."*

Cynthia's response to my question is very interesting because the lions in the animated chart were represented by yellow dots and, towards the end of the animation, a simple paw icon. Therefore, the image of the distressed panting lion that Cynthia refers to is from her own imagination, not the abstract representation presented in the visualisation. This is a particularly literal example of how people draw on their imagination, informed by their prior knowledge of a subject, to make a visualisation meaningful. The audio element within this visualisation presents a strong narrative and a sense of storytelling that appears to have engaged Cynthia's imagination. However, when I asked Olivia, a primary school teacher in her twenties, about how important the audio element of this visualisation was, she said that without the audio you could not make sense of it unless you guessed what the dots meant. She explained that the narrator "*gives the story*." For Cynthia though, using her imagination was a strategy through which she was able to re-contextualise data so that she could draw meanings from visualisations I showed her. This was the case even when the sample size was much larger.

The final visualisation I showed participants was an interactive map representing traffic congestion in the city of Washington DC. The map was simple and depicted the road network of the city with thin lines. The lines, each representing individual roads, changed colour depending on the level of congestion measured through a 'travel time index.' The chart incorporated interactive toggles that the user could adjust so that the visualisation changed to represent different times of the day on weekdays or weekends. The map did not incorporate contextual details in the form of identifiable landmarks or districts. However, context was provided in the form of highlights, which drew attention to particular roads using small diamond shapes overlaid onto the map. The user could click on the diamond to find out the name of the road and its travel time index for the time of day and days of the week selected. As she engaged with the map visualisation, Cynthia used the interactive toggles to move between different times of the day and days of the week. As she did so, Cynthia drew on what the visualisation showed to imagine what a particular area of the city was used for, saying:

> Cynthia: "*Well, if you sort of click on weekends it's sort of quite congested all round there* [...] *and then if you click weekdays, look at the difference. Yes, so that to me is […] people are out and about more, and that's more going out, socialising type place, rather than working."*

As our discussion of this visualisation was ending, Cynthia returned to the area on the map that she had identified as being a place where people socialised, saying:

> Cynthia: *"[...] if you go late night, look, weekend late night, look, see? So that must be where all the theatres, restaurants and things are."*

By identifying how congestion changed at different times of the day and week, Cynthia imagined how the space that a particular area of the map represented was being used. These quotes demonstrate how Cynthia firstly identified a pattern in the data, before drawing on her knowledge of how cities are arranged to make meanings from the visualisation. Although at first this seems to be a problem-solving strategy, Cynthia does not talk about the space in abstract terms. Instead, she pointed out the area on the map where she imagined the "*theatres"* and *"restaurants"* were, adding context to what the map showed. Furthermore, Cynthia imagined "*people*," who were "*out and about*," and what they were doing – "*socialising*" – to cause this congestion. The visualisation showed congestion on the roads but did not offer an explanation as to why certain areas became congested. Yet, by using her imagination, Cynthia was able to think about the people who generated the data, what they were doing, and their motivation for travelling to an area of a city on weekend evenings. In doing so, she was able to re-contextualise the data to fill in the "gaps and invisibilities" (Amoore, 2009, p. 23) built into the visualisation. It is likely that someone who lived in the city would possess local knowledge about why those areas were congested at the weekend. Yet, despite not knowing the city, Cynthia was still able to re-contextualise the data in order to tell a story about why the visualisation showed certain patterns of congestion.

Figure 4: Image removed due to copyright. The image was a screenshot of 'Travel during the week' data visualisation.

Rebecca also imagined both the subject of the data and how the data was generated when interpreting the visualisations of an individual's Fitbit data. This visualisation was static and comprised of seven small data visualisations set out in the style of a dashboard. The data covered a twenty-four hour period and included figures as well as charts, for example the

number of steps the individual took and the hours of sleep they had. At the bottom of the 'dashboard' there was a bar chart with the time across the x axis by hour and the number of steps in increments of 200 on the y axis. In order to make meanings from this graph, Rebecca imagined the individual who had generated this data in terms of how they had spent their time over the course of the day, saying:

> Rebecca: "*Yeah, it's basically showing you what your activity is throughout the day. So obviously, that's the time you were asleep and obviously there's peaks and troughs at different times – they must have gone out for a coffee then, and gone for lunch and stuff, and then maybe gone for a walk after work.*"

Rebecca highlighted the peaks in activity represented in the bar chart and explained them by imagining what the individual might have been doing at those times. The graph provides no information on who the individual is and no clues as to what their day-to-day life might look like. Indeed, some of the participants struggled to re-contextualise the data represented in the Fitbit visualisations because they did not know *who* the data represented. Therefore, it is likely that Rebecca drew on her own experiences to imagine the activities of the individual the data represented and account for the "*peaks and troughs*" in the chart. In imagining the subjects the data represents, Rebecca and Cynthia have re-contextualised the data by adding qualitative detail to quantitative data. Qualitative elements are removed from the things data represents through the process of quantification, prompting people to reinstate this kind of detail in their interpretation of it (Espeland, 2015), in this case, through strategies of re-contextualisation. Although using their imagination also required participants to draw on their prior knowledge of a topic, through imagining they were able to add additional layers of context to the visualisation to provide a rich account of why the visualisation showed what it did. Doing this re-contextualisation work enabled Cynthia and Rebecca to make meanings from the visualisations despite knowing little about the subjects of the data.

Other participants re-contextualised the data by imagining who made the visualisation and for what purpose, demonstrating a critical interpretation of the graph or chart. Gavin was a project manager in his forties, he worked in finance and his job involved working with data and presenting it in graphs. When I showed Gavin the gender pay gap graph he said he thought it was by someone who *"is strongly trying to influence you that women are paid less than men."* Gavin looked at the graph critically, questioning if it was possible to make a judgment without knowing how pay related to job roles within the companies. When I asked Gavin if he had learnt anything from the graph, he suggested that was difficult without knowing who produced the graph and why they produced it. Gavin then gave two examples of possible scenarios,

firstly, that it might be a piece of campaign literature, in which case he would discard it, secondly, that it might have been designed as a college experiment, in which case it would be of interest. To clarify my understanding I asked Gavin:

> Interviewer: *"You're even really imagining who made it, like when you're looking at it?"*

> Gavin: *"Yeah. I want to understand the viewpoint of where it's coming from and what it's trying to do."*

It seemed that Gavin did not want to make meanings from the graph in relation to what the data represented because he felt he did not have all the information to do so. However, he was much more comfortable re-contextualising the data in terms of imagining who might have created it and for what purpose. Although this allowed him to make meanings from the graph, these were not in relation to what the data represented. Instead, these were broader assumptions about what the visualisation was trying to persuade him to believe. Gavin made similar assumptions about an animated spiral chart I showed him that represented global temperature change. In identifying the potential politics of graphs in terms of what the designers of them wanted to "*influence you*" to think, Gavin is displaying data literacy skills through his critical evaluation (D'Ignazio and Bhargava, 2020). Interestingly, employing these skills seemed to make him more reluctant to make meanings from the graphs than some other participants were. Earlier in the interview, Gavin had described himself as being cynical about graphs and when I later asked him about this, he explained:

> Gavin: *"In my view statistics is a way of formulating an argument. And I know fine well from my own work that 95% of the time you can make statistics say whatever you want it to say. And it's all in how you present it."*

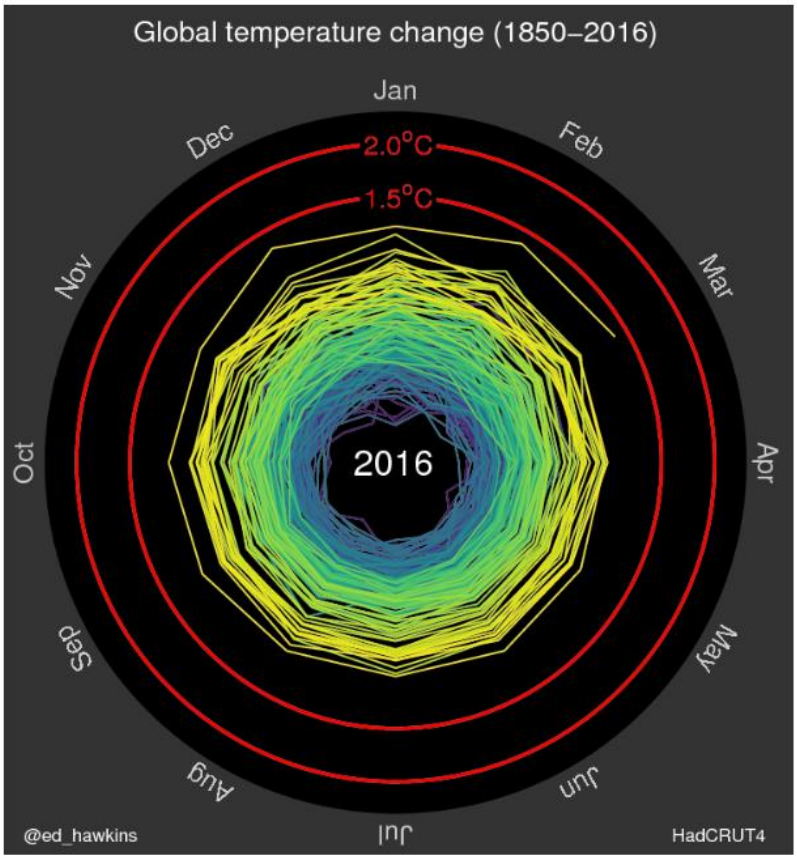

Figure 5: Screenshot of Global Temperature Change. (2016). Produced by Dr Ed Hawkins, Climate Lab Book. [online animated data visualisation]. https://www.climate-lab-book.ac.uk/2016/spiralling-global-temperatures/ Printed with permission. CC BY-SA 4.0.

Gavin's view of the potential of graphs to mislead framed how he made meanings from the visualisations I showed him. It is likely that this relates to his practical knowledge of graphs and his personal experience of working with data. This knowledge and experience shaped the frames of reference within which he re-contextualised the data. Despite not making meanings from the *data*, Gavin still accounted for why the visualisation might show what it did, albeit in relation to the designer's intent. Espeland (2015) explains that data prompts narratives not only about what the data means but about how it is used and for what purposes. This, she argues, may involve *correcting* perceived injustices in what the data is being used to claim (Espeland, 2015). This corrective narrative response is not only reflected in Gavin's response to the gender pay gap graph and global temperature change chart but also in Elizabeth's response to the carbon emissions data visualisation. For Elizabeth, it was the subject matter of the data that prompted her to imagine what the visualisation had been designed to achieve.

Elizabeth was a woman in her forties, who worked as a teacher. When I showed her the carbon emissions bubble chart, Elizabeth was perplexed that it did not provide information about the

amount of carbon emitted, nor did it provide information about the places the different circles represented. She explained she could "*have a rough guess*" that one of the transport emission circles represented an airport, but it was not clear. Elizabeth was reluctant to draw meanings from the chart because she felt she needed more information to be able to explain what it represented, stating, "*it's worthless without any further information*." Despite the lack of contextual detail, the key highlighted that a coal-fired power station was among the utilities, and the largest bubble on the chart represented a utilities emitter. Although the chart I showed them did not confirm this, this *messaging* (Segel and Heer, 2010) implied that the large utilities bubble represented emissions from the power station. Elizabeth shared this conclusion, which led her to imagine how the chart might be used to influence people, she said:

> Elizabeth: "[…] *I suppose I'm suspicious. I would suspect that someone would be using that to frighten other people possibly about what emissions"* [come from the coal-fired power station].

Elizabeth explained that the power station could be one of the other yellow bubbles on the map but that the chart seemed to be suggesting it was the largest one. I said to her that I found it interesting that she was thinking about how the visualisation might be *used* in the world. Elizabeth told me that her husband works in the power industry and therefore, seeing the power station on the map had interested her. Like Gavin, she was critical of statistics, saying that people use them to suggest things that may or may not be true. Elizabeth said she would find it interesting to know what the chart was being used for, but that her suspicion was that it was being used to highlight emissions from the power station. Her difficulty in making meanings from the data and the subject of the visualisation led Elizabeth to draw on her personal experience of her partner's work in the energy industry and her imagination to re-contextualise the data. This allowed her to make meanings from the visualisation in relation to how it might be used to "*frighten people*" about the emissions from the power station. Although these meanings were not strictly related to the data, they did offer an explanation as to why the visualisation showed what it did.

Participants drew on their imagination to add context to the data in order to make meanings from the visualisation. This allowed them to fill in the gaps in information that are inevitable when reducing complex data to pixels (Amoore, 2009). This allowed the participants to tell stories about the subjects the data represented and even the motives of those who produced the visualisation. Sometimes this required participants to draw on their prior knowledge or personal experience of the topic. Yet, their imagination allowed them to construct rich narratives from the visualisations in their pursuit of meaning. Drawing on their personal experience of a subject,

as Elizabeth also did, was another common strategy among participants when re-contextualising data, which I will explain in the following section.

6.3.3 Personal experience

As Berger (1972) suggested, how people see and interpret images tends to involve looking at them in relation to ourselves. In order to make meanings from visualisations, participants often drew on how the data related to their personal experience of the subject the chart or graph represented. They drew on this experience and the knowledge they possessed about the subject to situate the data within a frame of reference that was meaningful to them. As with the previous meaning making strategies, this sometimes involved accounting for *why* the data visualisation showed what it did. This section of the chapter will explain how people re-contextualised data by drawing on their personal experience of a topic. It will also illustrate how an inability to situate the data in relation to their personal experience could act as a barrier to engagement.

Cynthia was bemused when I showed her the antibiotic resistant E.coli data visualisation, saying "*I'm a statistic in this*." Cynthia explained that she suffers from a chronic infection, therefore, she had a personal connection to the visualisation. Although being able to relate to the subject matter personally sparked her interest in the visualisation, while engaging with it she drew on her personal experience of healthcare in Germany to re-contextualise the data. This allowed her to explain why the chart showed that some countries had lower rates of antibiotic resistant E.coli infections than others. Cynthia highlighted Germany and the Netherlands, saying:

> Cynthia: *"So that is really interesting! I would imagine Germany and the Netherlands, they will be better, and all these countries here, they won't be as good"* [referring to the chart].

I asked her why she thought Germany and the Netherlands would be better and she replied:

> Cynthia: *"Having lived there, in Germany for* [several] *years, they're far more advanced than us* [the UK]*, in most things."*

Cynthia went on to describe how the healthcare system differed in Germany and explained how on her return to the UK she had not been able to access the same medications she had had access to in Germany. Although the graph was about rates of antibiotic resistant E.coli, Cynthia drew on the graph to talk about the healthcare systems of different countries in relation to that of the UK, often identifying countries she felt had better healthcare provision than the UK.

Interestingly, Cynthia described the UK's rate of antibiotic resistant E.coli as "*rising again*," when the chart highlighted they had fallen. This suggests a disjuncture between what the chart showed and how she made sense of it. By drawing on her experience and knowledge of healthcare in different countries, Cynthia re-contextualised the data within a frame of reference that she was comfortable to discuss at length. During the discussion, Cynthia moved between the information provided in the chart and her own experiences and knowledge about healthcare systems to make meanings from the data visualisation. Similarly, Steph, a primary school teacher in her early thirties, drew on her personal experience of the Greek healthcare system to re-contextualise the data within the same visualisation.

While explaining what the visualisation was about, Steph highlighted particular countries showing higher rates of resistant E.coli, "*Bulgaria, Cyprus, Slovakia, Romania*." Despite Steph saying that she did not know much about the topic of antibiotic resistance or rates of antibiotic resistant E.coli in Europe, she was able to make meanings from the visualisation by drawing on her personal experience of healthcare in Greece. When I asked if she had learned anything from the chart, Steph explained she had been "*shocked*" by the countries that had the highest rates of resistance. When I pressed her on which countries she was referring to, she identified Romania, Bulgaria, and Slovakia again and also Italy, before going on to draw on her experience of Greek healthcare. Steph explained that in Greece, there is no national health service and therefore you can buy medicines from the chemist without prescriptions from a doctor, she said:

> Steph: "*They're* [antibiotics] *not as controlled as they are here* [the UK]. *And so, I therefore wonder if places like Romania and Bulgaria are similar. So antibiotics aren't as controlled as they are here? Maybe, but yeah that's more of a personal interest in that."*

Steph identifies that she is drawing on her personal interest in the topic in relation to her experience of healthcare in Greece and how this compared with the UK. Without referring to the rates of antibiotic resistant infections in Greece or the UK, she drew on her experience of the healthcare systems in these countries to account for why certain other countries may have high rates of antibiotic resistant E.coli. In doing so, Steph re-contextualised the data by situating it in a frame of reference she has personal experience of in order to make meanings from what the chart showed. Kennedy and Hill (2018) draw on Amoore's argument that pixilation is a technology of distance to suggest that people's emotional response to data visualisations reverses this distance. However, the data from my research suggests that re-contextualisation also helps to reverse this distance, by bringing participants closer to the data by situating it in frames of reference with which they were familiar. The way in which context brought

participants closer to the data can be seen most clearly when they drew on their personal experience. The contexts they situate the data in may not reflect the context in which the data was generated, collected or analysed, yet that can be said of all the strategies discussed in this chapter. Nevertheless, re-contextualisation allowed the participants to interpret the visualisation in relation to their own experiences. This produced somewhat partial perspectives (Haraway, 1988) that were situated in participants' individual experiences, knowledge, and/or imagination.

Although both of these examples relate to the same chart, participants drew on their personal experiences when interpreting other data visualisations I showed them. However, they often did so while shifting back and forth between their personal experience and prior knowledge of a subject, which might include identifying patterns in the data and accounting for why those patterns existed. For example, Anthony, a team manager in his fifties, re-contextualised the data represented in the gender pay gap graph by drawing on the history of traditionally male dominated jobs, during which he referred to heavy industry and coal mining, both of which have strong associations with the area of the North East in which he was living. As Anthony talked about the topic of the graph, he drew on his discussions with others about how the gender gap in heavy industry was not being reported, as well as his daughter's experience of working in a male dominated trade. Anthony was constructing a narrative from the data that had a particular history and geography while being situated in their own gendered experience and status (Dourish and Gómez Cruz, 2018). This demonstrates how although participants might draw on meaningful discourses to *decode* the data (Hall, 1980), they also weave in their personal experiences, or imagination, to construct *situated* narratives about what the data means.

This weaving together of context is indicative of how participants drew on their personal experience during the unfolding of longer narratives about what the visualisation meant. Therefore, it is difficult to illustrate this strategy through interview quotes. However, Olivia, a primary school teacher in her twenties, succinctly related the importance of being able to draw on her personal experience of a topic in order to maintain her interest in a visualisation. When I asked Olivia what the traffic congestion chart was about, the first thing she said was that the chart did not interest her and that initially she was not sure what it meant. However, despite persevering with the chart and explaining to me what it showed, she found it difficult to make meanings from it. Olivia explained that without being "*familiar*" with the place the chart represented, it was difficult to know what it related to, saying it "*kind of means nothing*" to her. This demonstrates how being able to situate the data in some kind of context that is "*familiar,*" or personal, is important for some users when making meanings from visualisations. I pressed Olivia as to why she did not find the chart interesting and she explained:

> Olivia: *"If it was like* [name of town], *if it showed* [name of town], *because I travel there to work, I would be more interested in it because I'd like to see if it matched my personal experiences and that sort of thing. But yeah, because I don't know where it is, I'm not really."*

Olivia suggests that she would have been more interested in engaging with the chart if she could have compared it with her own experience of the place it represented. However, as the visualisation represented Washington DC, she was unable to do so. Olivia did identify areas on the map where congestion was high, therefore demonstrating her comprehension of what the chart showed. Yet, she expressed difficulty in making *meanings* from the visualisation. This suggests that for Olivia, meaning was not merely about comprehension, but about being able to situate the data from the chart within a broader frame of reference. In this case, Olivia would have liked to have situated the data in relation to her personal experience. This illustrates how being able to re-contextualise the data in relation to their personal experience of a subject was an important factor in aiding participants in making meanings from graphs and charts. However, it also implies how not being able to re-contextualise the data can present a barrier to engagement.

Participants drew on their personal experience of a topic to re-contextualise the data represented by the visualisation. This not only allowed them to fill in the gaps in the visualisation but also brought them closer to the data by situating it in frames of reference with which they were familiar. Often this occurred during the unfolding of longer narratives about the data, which were embedded in particular historical and geographical contexts as well as the socially and culturally situated experience of the participant. However, as Olivia's quote indicated, re-contextualising strategies do not always work and despite being able to read a chart for what it *shows*, without putting the data into context it can be difficult to determine what it *means*. This suggests that for some, decoding data visualisations requires they identify a discourse through which to make sense of the data that is meaningful to *them.* The following section briefly explores what happens when re-contextualisation fails.

6.3.4 When re-contextualisation fails

During the interviews, there were several occasions where participants struggled to re-contextualise the data within particular graphs or charts, which limited their ability to make meanings from them. When this happened, it was common for participants to suggest they needed "*more information*" to interpret the visualisation. Some suggested they would like to see more "*numbers*" while others said they needed more information to understand what the

numbers meant. Sometimes participants were able to find ways round their difficulty in re-contextualising the data, in order to make meanings from the graph or chart. For example, both Gavin and Elizabeth employed their imagination and drew on their personal experience to re-contextualise data in relation to who may have designed the visualisation and for what purpose. However, for others, their difficulty in situating the data in familiar frames of reference led them to cease engaging with that particular graph or chart altogether.

Lisa was a self-employed woman in her forties and a member of an arts group. When I showed Lisa the antibiotic resistant E.coli data visualisation, she said that she had read the text but that it was not a topic she was interested in and so she "*didn't take it in."* Lisa said that she required more information in order to make sense of it, explaining:

> Lisa: *"And with the chart, you know, charts are always pretty to look at but I don't know what it's telling me, I don't know if it's good or bad. There's not enough information there, it's just numbers."*

I asked Lisa if by "*numbers*" she was referring to the percentages that appeared when the user clicked on a particular country's segment in the chart. Lisa agreed and said that although the chart indicated that the figures had fallen in the UK she was still not sure if it was "*good or bad*," articulating her difficulty in re-contextualising the data. Lisa conceded that the chart might tell her this somewhere but accounted her lack of interest in the topic as the reason why she was not taking in this information. I asked her if this made her not want to engage with the chart and she agreed saying no, she did not want to engage with it. Listening back to the recording of the interview, I could hear Lisa talking about the chart with disdain, suggesting an emotive response to her engagement with this visualisation. Kennedy et al., note that visualisations can prompt "strong emotions" among users and this can influence whether or not they continue to engage with a visualisation (2016b, n.p.). Lisa's response to the antibiotic resistant E.coli visualisation illustrates how both a lack of interest in the topic and difficulty re-contextualising the data led some participants to stop engaging with a chart.

Lisa's comment about not knowing if the numbers were "*good or bad*" mirrors what Turner, a designer at Evoke, said when they were struggling to contextualise the data they had been asked to represent in a chart. This demonstrates how being able to contextualise the data a chart represents is important in *both* the design and interpretation of data visualisations. Indeed, as this section will go on to show, other user interview participants expressed similar problems understanding what the data meant. For example, when I showed Jess the traffic congestion data visualisation, she found it difficult to make sense of the travel time index figures. These figures

became visible when the user clicked on white diamond shapes that hovered above some of the roads on the map. Although these highlights provided the user with additional context about the level of congestion in the city, Jess found it difficult to put these *into context*. The travel time index was shown in a key at the bottom of the visualisation as ranging from one to three and from the colour green to red. The caption to the side of the graph indicated that the higher the number of the travel time index, the slower the speed of traffic along that stretch of road. However, Jess struggled to understand what this scale meant in "*real terms,*" saying:

> Jess: *"I don't know what those numbers represent. For me it doesn't seem that much of a difference because it's all within 2.5 to 3.5 ish, but that could be an enormous difference."*

> *"*[...] *I don't know what the difference represents in real terms.* […] *it gives me the information but I don't know enough about that information."*

Jess was a woman in her early thirties who worked in social care. Her quote illustrates her inability to re-contextualise the data because she was unsure what it represented. Without being able to situate the data in a frame of reference she understood, Jess struggled to make meanings from the chart. This was despite her engaging with the interactive functions and demonstrating her understanding of what the chart showed in relation to different levels of congestion. However, Jess was unable to make sense of what these levels meant in a real-world context and this was a barrier to her ability to make meanings from the visualisation. Nafus reported similar frustrations among her participants when the context of their sensor data always seemed to be "out of reach" (2014, p. 216). The designers at Evoke spoke about accessibility and engagement as being important factors in user engagement with visualisations (see Chapter five). However, the interview data suggests that being able to situate the data in a context they are familiar with is extremely important in determining if a user is able to not only engage with data visualisations or read them, but to *make meanings* from a graph or chart.

Understanding how the data relates to the real-world situation it represents was reflected in Gavin's response to the carbon emissions bubble chart. Gavin said he did not find the visualisation interesting because there was not enough information, explaining that he wanted to see "*numbers and things*." To clarify, I asked Gavin if numbers would provide him with the information he wanted, he replied:

> Gavin: *"Numbers and percentages and also sort of what the numbers mean, so the background as in, if its 3000* CO_2 *or whatever you want to call it, the measure, what*

> *does 3000* CO_2 *mean per square metre. Does it mean that that's really bad? Are we talking people coughing and spluttering and ending up with cancer? So what does it actually mean?"*

Gavin explained his lack of interest in the graph by what he perceived to be a lack of information. In explaining that he would like to see more numbers, he added that he would also need information on what the numbers *meant*. Here, Gavin identified the kind of information he required to re-contextualise the data in relation to how the levels of carbon emissions might affect people living in Indianapolis. Without this foundation of information, Gavin struggled to make meanings from the visualisation. However, like Elizabeth, Gavin did mention that the chart might be trying to *"campaign against the coal station."* Other participants made conscious assumptions in order to fill in the gaps in information in this chart, or to identify patterns in the data, which allowed them to make other meanings from the visualisation (in relation to the layout of the city, for example). It seems likely that Gavin could have found different ways to re-contextualise the data that did not rely on him possessing the information he sought. However, he demonstrated an unwillingness to do so, which could be connected to his cynicism towards graphs and their potential to mislead.

Highlighting situations when participants were unable to re-contextualise the data demonstrates how it is not possible for the visualisations to *mean anything* that the user wanted them to. Their meanings were to some extent anchored (Barthes, 1977) through rhetorical design techniques that worked to constrain the meaning of the visualisation towards a particular message (Hullman and Diakopoulos, 2011). This supports Hall's (1980) argument that while audiences can decode messages in a number of ways, this does not imply a plurality of meaning. Although some participants constructed narratives around what the visualisation meant that extended the reach of the data (Dourish and Gómez Cruz, 2018), even they were constrained by the topic the visualisation represented. This limited the discourses they could draw on or construct, through personal experience or imagination, to make sense of them. Without enough information to re-contextualise the data within frames of reference they understood and which were familiar to them, participants were unable to make meanings from some of the visualisations. Indeed, the significance of being able to put data into context when making meaning from visualisations raises questions around what happens when attempts to re-contextualise the data do not work. Although this section has addressed how an inability to re-contextualise the data led some user interview participants to stop engaging with a visualisation, more research is required to understand how people *feel* when re-contextualisation fails, why this happens, and how users respond.

6.4 Conclusion

By focusing on the meaning making practices of designers and ordinary users this chapter has revealed how they form narratives through the *contextualisation* and *re-contextualisation* of data. While designers contextualised the data to make meanings from it prior to the design of its visualisation, users re-contextualised data in order to explain what a visualisation showed. Both used various strategies to add context in order to make meanings from data through and from its visualisation. An inability to do so presented problems in both the design and interpretation of data visualisations, leading some users to cease engaging with a graph or chart altogether.

The strategies designers used to contextualise the data included conducting research beyond the dataset, drawing on their prior knowledge of what the data represented, and incorporating new data into the dataset. Contextualisation seemed to be most important when the data's intended use or user had not predetermined the meaning of the data. Although Chapter four discussed how designers made meanings from data when designing dashboards, it is not clear what role *contextualisation* played in the production of dashboard interfaces and this requires further research. Often user interview participants drew on some of the same strategies used by designers to re-contextualise the data, namely, drawing on their prior knowledge or personal experience of the subject the data visualisation represented. However, others also drew on their imagination in order to situate the data in contexts with which they were familiar. Identifying these practices of contextualisation and re-contextualisation makes visible the layers of subjectivities that determine the design and interpretation of data visualisations.

By adding context to the data, designers and users were able to form narratives about what the data meant. It is possible to understand these narratives in relation to the meaningful discourses that Hall (1980) argues producers must identify in order to transform events into stories that can be decoded by audiences. Hall (1980) did not explain what informed these discourses or how the audience selected a discourse through which to decode messages. This chapter has shown how the narratives designers and users constructed through contextualisation and re-contextualisation were grounded in particular histories and geographies. Indeed, the narratives the users constructed to explain the data were socially and culturally available to *them* and varied between participants. Therefore, *meaningful discourses* (Hall, 1980) are not merely "culturally available" stories (Dourish and Gómez Cruz, 2018, p. 7) that users will understand and decode in similar ways. Instead, the discourses users drew on to interpret visualisations often incorporated their personal experience, prior knowledge and/or imagination. Therefore, contextualising and re-contextualising data produced knowledge from data visualisations that was both situated and partial (Haraway, 1988).

If data could really speak for itself (Anderson, 2008) the kind of contextualisation and re-contextualisation work outlined in this chapter would not be necessary. As argued, data is devoid of narrative and this prompts people to construct new narratives in order to explain what it means (Espeland, 2015). Furthermore, data and its visualisation tend to be experienced at a distance from the phenomenon, event or person they represent. By adding context to the data, designers and users were able to bring the data closer to them. As Amoore argues, visualisation results in "gaps and invisibilities" which are filled in by those who interpret them (2009, p. 23). The desire among both the designers and the users to account for why the data showed particular patterns and trends, and the inferences they made to explain them, represents a response to these gaps and invisibilities. Accounting for what the chart showed was an important factor in enabling both designers and users to make meaning from the data and its visualisation.

The data analysed in this chapter suggests there is a distinction between comprehension and meaning making when interpreting a graph. While user interview participants could identify trends in the graph, this did not always lead to meaning making. This raises the issue of misinterpretation, as users bring their own knowledge, judgements and partial perspectives to re-contextualise the data they represent. In doing so, they often construct narratives that *extend the data's reach* (Dourish and Gómez Cruz, 2018) which allow them to make claims from the visualisation that go beyond what it is possible to know from the data. Therefore, while it is known that data visualisations can mislead (see for example, Dick, 2015; Hill, 2017), how to address this becomes much more difficult when one considers the individual and highly situated ways in which users make meanings from them. Furthermore, this chapter has revealed a tension between how designers contextualise data and how users re-contextualise it through their interpretation of graphs and charts. This suggests that the contextualisation work done by designers cannot fix the meaning of the visualisation. Despite the rhetorical nature of data visualisation (Hullman and Diakopoulos, 2011), there is always the potential for a disjuncture between the contextual frameworks used to encode and decode the visualisation.

The following chapter concludes by reviewing the key findings and argument of the thesis in relation to the research questions posed in the introduction. It will also state the contribution it makes to academic debate and to data visualisation design.

Chapter 7: Conclusion

7.1 Introduction

In conclusion, this project has explored three ways in which meaning is made in the design and interpretation of data visualisations; encoding meaning, imaginary figures, and contextualising and re-contextualising data. Chapters four, five, and six revealed how context is integral to how designers and users make meanings from and through data visualisations. Therefore, this thesis argues that a range of contextual factors, which exist beyond the dataset, determine the meanings of data visualisations. boyd and Crawford (2012) and D'Ignazio and Klein (2020) highlight the importance of the context within which data is generated when interpreting its meaning. However, this thesis has revealed how the significance of context extends beyond the dataset when people make meanings through data visualisations. By contextual factors, I am referring to the intended use of the data, the perceived needs, wants and likes of the end user, and the various ways data are situated within broader contextual frameworks in order to make sense of its visualisation. These contextual factors become materialised in the design and interpretation of data visualisations; they inform how visualisations are encoded with meaning, the imaginary figures used to envision how the end user will engage with them, and how the context of the data is restored in the visualisation through processes of contextualisation and re-contextualisation. These contextual factors combine to determine the meaning of data visualisations.

Just as aggregated data cannot always be easily unpicked (Nafus, 2014), the layers of context applied to data during the design and interpretation of visualisations are entangled in ways that are hard to unravel. This required a research design that engaged with the design and production of data visualisations and the interpretive practices of those who consume them. The subjectivities and human labour (D'Ignazio and Klein, 2020; 2016) that frame design decisions are concealed behind the conventions of data visualisation design, which include clean layouts and geometric shapes, which give the impression of simplicity (Kennedy et al., 2016a, pp. 723-724). Furthermore, how ordinary users make meanings from data visualisations has been neglected in existing research. By tracking the meaning making practices at work in *both* the design and interpretation of data visualisations, this thesis has made visible the layers of context that are central to these practices.

In the design of data visualisations, designers at Evoke first made meanings from the data. How they did this was determined by the particular project and was informed by the aims of the

client, the perceived needs of the users, and the level of access the designers had to the dataset they were visualising. In the case of dashboards, the meaning of the data was made in relation to how the data would be *used* to generate insights and make data driven decisions. In more stable data visualisation projects, where the dataset was more accessible to the designers, the meaning of the data was made through identifying narratives within the data that were interesting. What the designers deemed *interesting* was in relation to their client's aims as well as their perception of the intended users of the visualisation. However, context not only influenced how meaning was made from data but how it was *encoded* into the design of data visualisations too. For example, the ways designers envisioned the end user though *imaginary figures* was materialised in how they encoded meaning into the design of data visualisations. While the *back and forth* process involved in translating a visual image into a digital data visualisation required compromise between designers with different skill sets, in order to deliver the finished design within the constraints of individual projects. This included temporal and budgetary constraints and the limitations of the software the designers were working with.

As well as being informed by the context of the visualisation project, designers and users drew on various strategies to put data *into context* in order to interpret the data through and from its visualisation. Both designers and users attempted to make meanings from data by situating it in broader contextual frameworks. For example, understanding what the data meant or finding a narrative in a dataset required designers to situate the data in frames of reference that allowed them to make meanings from the data that would be familiar to users. This relates to Hall's (1980) argument that encoding requires producers to transform events into stories - or "meaningful discourses" - before audiences can decode them (1980, p. 130). This involved conducting research beyond the dataset, drawing on prior knowledge of the subject matter it represented, or researching and aggregating datasets. In doing so, the designers restored context to the data through the processes of visualisation. When interpreting data visualisations, user interview participants used a range of strategies to *re-contextualise* the data. This involved drawing on their prior knowledge of the subject the data represented, their personal experience and even their imagination to restore context to the data in the visualisation. By situating the data in frames of reference with which they were familiar, users made meaning from the data visualisations I showed them.

Designers and users make the data *mean* certain things by drawing on particular contextual frameworks to situate it. These are varied and depend on the aims of the visualisation project, the intended user, and the frames of reference users draw on to decode them. Indeed, the user interviews demonstrated how there were multiple ways in which individuals could make meanings from the same visualisation. Therefore, the particular ways that context is deployed to

make sense of data is integral to how meaning is made in both the design and interpretation of data visualisations. However, this reveals a tension in which both designers and users are restoring context to the data being visualised. Designers attempt to contextualise the data in order to encode the visualisation with meaning that users will understand while users are drawing on their own contextual frameworks in order to decode it. Hall (1980) recognised that audiences could decode messages in different ways and that meanings could not be fixed by producers. Yet Hall (1980) claims this does not mean that it is possible to decode a plurality of meanings, because structural factors mean there is always a dominant or preferred reading of the message. While the subject of the visualisation and data might anchor its meaning (Barthes, 1977), this thesis has shown the different and individualised strategies users draw on in order to decode data visualisations. These are dependent on the contextual frameworks users draw on and deploy to situate the data in frames of reference that they understand. Espeland (2015) explains that data has no narrative, therefore it is subject to new narratives. In this case, designers and users constructed narratives that restored the data's context in order to explain what it means.

This concluding chapter will go on to review the substantive findings from the three analysis chapters in relation to the thesis' overall argument before explaining how they answer the research questions posed in the introduction. Finally, the chapter will discuss what the project has achieved in terms of its contribution to new Sociological knowledge in the field of data studies and how the findings might be useful to data visualisation design practitioners.

7.2 Findings

This subsection reiterates the key findings from the analysis chapters in three parts. Each part relates to an individual analysis chapter, addressing them in order from Chapter four to six.

7.2.1 Encoding meaning in the design of data visualisations

Chapter four explored how meaning was encoded into the design of data visualisations and showed how meanings were made from data prior to the design of its visual representation. Insights did not simply emerge from the data but were shaped by a range of factors external to the dataset. These were project specific and included what the data *meant* to the client, the intended use of the visualisation, and the perceived needs of the intended user. How meanings were made from data was also framed by the way designers and clients *imagined* data, particularly its potential for generating insights and for data driven decision-making. Indeed, this "data imaginary" (Beer, 2019, p.15) is part of the broader socio-cultural context within

which data visualisations have become so prevalent. While Anderson (2008) argued that big data can speak for itself, this chapter illustrated how those in positions of power determine what data say. This power is attained through access to data and possession of the skills to analyse, interpret and visualise it (boyd and Crawford, 2012). These powerful actors include the custodians of the data and the designers who have the expertise to interpret and visualise it. These actors are a form of gatekeeper, deciding how data is used and for what purpose, thus framing the meanings that are made from it.

Designers who worked on the concept of a design, without always having access to the dataset being visualised, demonstrated how the meaning of data was determined prior to the design of its visual representation. This was particularly the case with data visualisations designed for dashboard applications where *dummy data* stood in for the *real* dataset. The level of access the designers had to the data they were visualising was linked to the regulatory apparatuses of the data assemblage (Kitchin, 2014), which govern how data can be used and shared. While the real data flows through the final dashboard architecture, it was not always present during its design. However, in more stable data visualisation projects, where designers had access to datasets and where the meaning was not predetermined by a specific use, meaning was made by identifying an *interesting* narrative within the data. Designers explored the data for what was interesting in relation to the context of the project, this included the aims of the client and how the designers envisioned the intended user. In both the design of dashboards and data visualisations, designers decoded the data by transforming it into a *meaningful discourse* that could be understood and decoded by users (Hall, 1980). This discourse might relate to the intended use of the visualisation or an interesting narrative that users might find engaging.

Once designers had decoded the meaning of the data, these meanings were *encoded* through the design and production of data visualisations. The process of decoding and encoding the data required collaboration between designers with different expertise. Therefore, often the encoded meaning of data visualisations was not the product of a lone designer but a network of people with different skill sets. The collaborative nature of encoding was most visible during the design and production of digital and interactive visualisations. This involved a *back and forth* process through which the design of the visualisation moved between visual designers and front end developers who produce the visualisation in code. Through collaboration, negotiation and ultimately compromise, the designers worked together to determine the final design of the graph or chart within the individual constraints of a project. This included the technical limitations of the software they were using, the aims of the client, and temporal and budget constraints. The *back and forth* demonstrated how encoding does not stop at the visual design, it continues to evolve through the production of data visualisations and is framed by the context within which

the visualisation has come into being. As such, data visualisations are often a product of compromise and are a materialisation of the design decisions made in order to realise the final design.

In conclusion, data visualisations are not merely visual representations of data. They are a visual representation of the decisions of multiple actors within and outside of the studio. Design decisions are framed by the particular context of a visualisation project including the client's aims, the perceived needs or users, financial and temporal constraints, regulatory apparatuses and technical limitations. This assemblage of *actors and actants* (Kitchin, Maalsen and McArdle, 2016) converge to determine the meaning of data visualisations.

7.2.2 Designing visualisations for imaginary figures

Chapter five explained how the designers at Evoke drew on imaginary figures to make design decisions, which determined how they encoded meaning into the design of data visualisations for different users. The shared assumptions the designers made about different 'types' of user were personified in three imaginary figures: *the grandma*, *the public* and *the professional*. The public represented a general and non-expert user, while the grandma represented those users who possessed the lowest levels of ability in relation to engaging with data visualisations. However, the designers imagined the professional to engage with data visualisations at work and, therefore, they were deemed competent at interpreting graphs and charts. Imaginary figures were not a physical presence, instead, they were "discursively constructed" (Clarke, 2005, p. 46) through the way the designers spoke about 'different' users to explain and account for the design decisions they made. Drawing on *imaginary figures* allowed the designers to 'know' the user (Woolgar, 1991). They drew on them to make design decisions that framed how meaning was encoded into the design of data visualisations for different users. This informed the type of visualisation used and its level of interactivity, but also how the visualisation guided the user through the data. In *configuring* (Woolgar, 1991) imaginary figures, the designers also envisioned the contexts within which the visualisation would be viewed and interpreted.

Although imaginary figures represented three different types of user, each figure represented a homogenous category of user and there was little or no sense of diversity within categories. Indeed, identifying imaginary figures revealed how the designers' working understanding of diversity was limited primarily to gender and age. It showed how diversity among users was not understood in relation to socio-cultural categories such as social class, level of education or ethnicity. Instead, it was defined in terms of how the designers imagined different users' level of ability to engage with and make meanings from data visualisations. Indeed, the chapter argued

that *the grandma* personified all those users who were imagined to possess the lowest levels of meaning making ability, thus, she became the benchmark for accessibility. Therefore, it is more useful to think about the designers' understanding of diversity in terms of the *limits of meaning making*. By this, I mean how the designers imagined different users' ability to engage with and make sense of graphs and charts. By drawing on imaginary figures, designers encoded meaning into data visualisations by prioritising what the professional *wants*, what the public *likes*, and what the grandma *needs*. Drawing on these limits of meaning making, they encoded meaning differently for 'different,' if not diverse, users.

Relying on how the designers *spoke* about users to explain and account for design decisions makes it difficult to assess the significance of imaginary figures in more implicit design practices. Nevertheless, how the designers drew on these figures to explain their design decisions explained how their perception of different types of user shaped the way they encoded meaning into visualisations. Therefore, not only were the designers encoding their own subjectivities into the design of data visualisations through the design choices they made, they were also encoding them with the subjectivities of *imaginary figures*. This raises important questions around if and how these imaginary figures are scripted (Akrich, 1994) into the design of data visualisations and if users decode these scripts in their interpretation of them. Through the encoding and decoding of user scripts, data visualisations have the potential to perpetuate user stereotypes and reproduce unequal access to data, particularly in relation to older female users. Answering these questions will require further research. This might involve tracking the production and consumption of a data visualisation, from its design to its interpretation by the intended users.

7.2.3 Contextualising and re-contextualising data

Chapter six explained the different ways in which both designers and users added context to the data in order to make meanings through and from its visualisation. Contextualising the data allowed the designers to determine the meaning of the data prior to the design of its visual representation while re-contextualising the data enabled user interview participants to make meanings from the visualisations I showed them. The designers and users drew on some of the same strategies in order to add context to the data. Specifically, identifying patterns in the data and drawing on their knowledge of the subject it represented to make sense of what graphs or charts showed. However, designers were also able to conduct research beyond the dataset in order to contextualise the data. This involved speaking to clients to find out how the data will be used or to account for anomalies, conducting research into what variables meant, and incorporating additional datasets. Without the ability to conduct research during interviews, user

interview participants tended to draw on their own prior knowledge of the topic the visualisation represented as well as their personal experience and even their imagination to re-contextualise the data within the visualisation.

Both designers and users felt the need to add context to the data in order to make meanings from it and it is important to consider why they felt this was necessary. Firstly, as Espeland (2015) argues, data are devoid of narrative, prompting people to construct new narratives that explain what the data means. Secondly, data visualisations can only ever provide partial representations of the data which cannot show everything it has to tell (Boehnert, 2016; Manovich, 2011). Furthermore, data itself is an abstraction of the event, phenomenon or people it represents (Kitchin, 2014). Therefore, it is inevitable that graphs and charts contain gaps in information (Amoore, 2009), which both designers and users attempted to fill by restoring context to the data. For example, by situating the data in frames of reference that were familiar to them and which they felt confident to discuss, user interview participants attempted to account for what the visualisations showed. They made inferences and assumptions to explain patterns in the data, which allowed them to make meanings from the visualisations I showed them.

Narrating data (Dourish and Gómez Cruz, 2018) was an important means through which data was contextualised and re-contextualised. By adding context to the data, both designers and users constructed narratives that explained what the data or visualisation showed. In doing so, they were able to "extend[s] the data's reach" beyond the information contained in the data or visualisation (Dourish and Gómez Cruz, 2018, p. 4). Hall (1980) did not explain how producers select meaningful discourses or how users decode them. However, this chapter revealed the importance of narrative in constructing *meaningful discourses* (Hall, 1980) that helped the designers make meanings from the data and users to decode the meaning of the visualisation. In doing so, it showed how meaningful discourses are not merely "culturally available" stories (Dourish and Gómez Cruz, 2018, p. 7) that users will deploy and decode in similar ways. Instead, users interweave social and cultural discourses with their prior knowledge, personal experience and even their imagination to construct narratives that re-contextualise the data.

Chapter six acknowledged how some user interview participants struggled to re-contextualise the data within a visualisation I showed them. In some instances, this presented a barrier to meaning making. While many of these participants were able to comprehend what the graph or chart showed, they felt unable to situate the data in familiar frames of reference through re-contextualisation. This led some participants to cease engaging with a visualisation altogether. While reinforcing the argument that re-contextualisation is an important meaning making strategy when engaging with data visualisations, this finding also suggests a distinction between

meaning making and comprehension. While the user interview participant might have provided an account of what the visualisation showed, if they were not able to situate this in frames of reference that were familiar to them, they were less able to make meanings from it. Therefore, decoding a data visualisation not only involves identifying the meaningful discourse encoded into its design, but requires the user to be able to situate the data in contexts that are *meaningful to them*.

7.3 Responding to the research questions

This section of the chapter briefly responds to each of the research questions posed in the introduction and explains how the thesis has answered them.

7.3.1 How do designers encode meaning into data visualisations?

Designers at Evoke encoded meaning into data visualisations in two stages. Firstly, they made meanings from the data prior to designing a visual representation of it. This involved decoding the data to translate it into a *meaningful discourse* (Hall, 1980) that users could decode. In the case of dashboards, the meaning of the data was closely related to the context in which it would be used. When working on dashboard projects the designers often do not have access to the full dataset being visualised and design data visualisations using *dummy data.* However, in the case of more stable data visualisation projects where the meaning of the data is not already determined by its use, meaning was made by identifying an *interesting* narrative in the data. This narrative related to the client's aims and what the designer imagined the intended user would find engaging. Decoding the data to determine its meaning was an important pre-requisite to design. Without understanding what the data meant, Turner, for example, was unable to design a visualisation to represent it. In this sense, meaning does not simply emerge from the data through its visualisation, it is already framed by the intended *use* and intended *user* of the visualisation.

The second stage of encoding involves the design and production of a data visualisation that represents the meaning of the data. During this stage, the designers made decisions about the visual elements of the graph or chart, informed not only by the context of the project but also by how the designers imagined the intended user. Drawing on *imaginary figures* allowed the designers to make design decisions for different 'types' of user. The limited ways in which designers envisioned the end user framed what they prioritised in their design - accessibility, engagement or speed and comprehension - and how the visualisation would lead the user through the data. However, encoding did not stop at the visual design of the graph or chart.

When designing digital and interactive data visualisations, designers with different skill sets worked closely together to produce the final design of the visualisation. This involved a *back and forth* process in which the design moved back and forth between the visual designer and the front end developer, who was responsible for translating the visual image into code. This process involved compromise between the visual designers and developers in order to achieve the final visualisation within the limitations of the software and constraints of individual projects. This made visible the ways in which encoding meaning into the design and production of data visualisations happens within a socio-technical assemblage of actors and actants (Kitchen, Maalsen and McArdle, 2016), which converge to determine the final design.

7.3.2 How do ordinary users make meanings from data visualisations?

User interview participants added context to the data represented in the visualisation in order to make meanings from it. I have referred to this process as re-contextualisation, to recognise the contextual work done in the design of data visualisations and to acknowledge how, in their interpretation, the user may not draw on the same contextual frameworks as the designers who created it. The user interview participants drew on various strategies of re-contextualisation including *problem solving*, which involved identifying patterns in the data and explaining them by drawing on their prior knowledge of the subject. They also restored context to the data by drawing on their *personal experience* of the topic the visualisation represented and even their *imagination*. By re-contextualising the data, user interview participants situated it in frames of reference with which they were familiar and which they could draw on to make meanings. In doing so, they were able to construct narratives from the data that accounted for what the graph or chart showed.

The desire many of the participants had to re-contextualise the data can be explained by a need to fill in the gaps in information that result from the visualisation of data (Amoore, 2009). In order to make meanings from the data, user interview participants filled in these gaps by putting the data into contexts with which they were familiar. Furthermore, and as argued above, data has no inherent narrative, therefore people construct *new* narratives that restore its context in order to explain what the data means (Espeland, 2015). This often led participants to construct narratives around the data that reached beyond the information provided in the visualisation. To do so participants not only drew on social and culturally available discourses, but incorporated their prior knowledge, personal experience and imagination to make meanings that were both partial and situated (Haraway, 1988). Analysis also suggests that how users make meanings from data relates to their existing ideas and values about a subject. Although these exist outside of the visualisation they come to the fore through the ideas the visualisation provokes.

Therefore, how users make meanings from data is dependent on the individual and the ways in which they re-contextualise the data within frames of reference that are *meaningful to them*.

7.3.3 What is the relationship between meaning making and data assemblages?

The meaning of data visualisations do not emerge from the data in isolation, they are embedded in a broader assemblage of *actors and actants* (Kitchin, Maalsen, and McArdle, 2016). The data assemblage, within which data visualisation design at Evoke was embedded, included: the clients and their motivations for visualising their data; the designers and their ideas, knowledge and expertise; the studio space, which facilitated collaboration; the various hardware and software used to design and produce data visualisations; and the regulatory frameworks which govern how data is shared and used. The way in which these *apparatuses* of the assemblage (Kitchin, 2014) converged, shaped how meanings were made from the data and how meaning was encoded into the design of its visualisation. These apparatuses are part of the broader context within which data visualisations are designed and produced. The data assemblage was most visible during the *back and forth* process in the production of digital and interactive data visualisations. Here, designers with different skills and subjectivities worked together to navigate the final design of the visualisation within the constraints of the project and the limitations of the technology. Designers made design decisions based on temporal and financial constraints, the aims of the client, and the limitations of the technology they were working with. This assemblage of *actors and actants* (Kitchin, Maalsen and McArdle, 2016) framed how meaning was made with data and how it was encoded into the design of data visualisations.

7.4 Contribution to knowledge

This thesis has analysed and exposed the meaning making practices involved in the design and interpretation of data visualisations from a social and cultural perspective. It has demonstrated how *encoding/decoding* (Hall, 1973; 1980) can be a useful theoretical framing when thinking about how people make meanings from data through visualisation. However, it has also built on the concept of encoding to explain how, in the context of data visualisation design, this takes place in two stages. Furthermore, it pushes at the boundaries of Hall's model, particularly the concept of meaningful discourses, to highlight the role imagination and narratives play in how the meaning of data is encoded and decoded through visualisation. The thesis has been able to make this contribution by conducting research in two stages. Firstly, it took an ethnographic approach to the study of data visualisation design in order to analyse the data and design practices of designers. This contrasts with data visualisation research that focuses its analysis on the image of graphs and charts. Secondly, by designing innovative interviews, incorporating

visual resources, the thesis introduced the concept of *re-contextualisation* to explain how ordinary users make meanings from data visualisations.

The thesis has also contributed its own conceptual framework in the form of *imaginary figures*, which helped to unpick how designers imagined users and how diversity among users was constructed in the studio. This is significant in relation to the data divide, which explains how the skills to produce, engage with, and make sense of data visualisations are not equally distributed within communities (boyd and Crawford, 2012). It has also pointed towards areas for further research to determine if and how *imaginary figures* are scripted (Akrich, 1994) into the design of data visualisations and then decoded by users. Finally, this thesis has contributed to ongoing debates within data studies by building on the work of other scholars who have highlighted the importance of context in shaping the knowledge produced from data. boyd and Crawford (2012) and D'Ignazio and Klein (2020) highlight the importance of the context in which the data was generated and collected in relation to its *meaning*. This thesis has focused on the contexts within which data is made meaningful prior to the visualisation process and during the interpretation of its visualisation.

While the thesis is weighted towards how meaning was made during the design of data visualisations, it also addressed how users make meanings from data visualisations when engaging with them. This provides the thesis with a unique understanding of the meaning making practices at work in *both* the design and interpretation of data visualisations. This work is timely as people increasingly encounter data visualisations in their everyday lives (Kennedy et al., 2016b). This broader trend towards visual encounters with data has become more noticeable during the coronavirus pandemic. Indeed, Kennedy (2020) argues that, "simple data visualisations have become key to communicating vital information about the coronavirus pandemic to the public" (2020, n.p.). This includes the graphs, charts and maps that take centre stage in government briefings and news media. In a recent interview published in the Observer newspaper (see Watts, 2020), Latour explained how, when faced with global crises, people tend to move towards alternative narratives. Latour explained that when provided with enough contradictory information, which gives the impression of being empirical, then people are likely to fall back on what they know or believe (Watts, 2020). The findings of this thesis suggest how, when faced with a proliferation of data visualisations about various aspects of the pandemic, people may draw on frames of reference that are familiar to them in order to make sense of them. Graphs and charts can be powerfully persuasive (Kennedy et al., 2016a). However, the findings of this thesis suggest that their ability to persuade may be due, in part, to the way people decode them in relation to their existing beliefs and values.

To conclude, the involvement of commercial partners encouraged me to consider how the knowledge produced from the thesis might also be useful to the field of data visualisation design. In November 2019, myself and the two other PhD students from the '*relating to data through visualisation'* network ran a workshop attended by design practitioners from the public, private and charity sectors. During this workshop, I shared with the attendees the findings from my user interview research and explained how the participants had added context to the data in order to make sense of the visualisations I showed them. This generated discussion within the workshop and one of the public sector organisations who attended has since contacted me to request more information on these findings to help inform their practices. Therefore, it is clear that this part of the research has the potential for impact among those who design data visualisations. Furthermore, disseminating the research findings on imaginary figures may prompt reflection among designers of visualisations, in relation to how they envision different users and diversity among users.

Appendices

Appendix a: Observation guide for field notes

Date: **Time:** **Location:**

Context:

(socio-)Technical and Production Processes mapping the tools and the processes involved in the construction of data visualisations
Intended Meaning What is the intended meaning of the data visualisation being produced?
Encoding How is meaning being encoded into the data visualisation? – Through design practices.

Apparatuses of the data assemblage Infrastructure, people, places, systems of knowledge, software, regulations…and so on.
Other

Appendix b: Designer interview schedule

(Note: To aid thematic analysis continue clarifying meanings throughout interview)

Briefing:

I will briefly explain the purpose of the interview, the use of a sound recorder, and so on, and ask if the subject has any questions before, and at the start, of the interview.

Introduction and background information:

What is your job title?

How do you describe your job to people who don't work in this field?

How did you get into designing data visualisations?

What did you do before?

Design Practice:

- How are you allocated the design briefs you are working on?

  What is the first thing you do when you receive a new design brief?

- Can you tell me how the software and data visualisation tools you use shape your work?

Prompts:

1. What's really good about them?
2. Are there any limitations to what you can do?
3. Have you ever designed data visualisations any other way?

- How do you imagine the end user of the data visualisations you design and build?

- How do you think people use and make sense of the data visualisations you design?

- What design techniques do you use to ensure a data visualisation makes sense to the end user?

- Do you think data visualisations have the power to make people care about the issues they represent?

Debriefing:

- I have no further questions. Is there anything you would like to raise or add before we end the interview?

- How did you find the experience of being interviewed?

I am now turning off the tape recorder.

10 minute reflection

Set aside 10 minutes of quiet time after each interview to reflect on what has been learned from the particular interview, drawing on empathetic meanings interpreted from the interview experience. This should provide a record of immediate impressions and add valuable context for the later transcripts.

Appendix c: Amended designer interview schedule

Researcher's questions in bold/Translation into interview questions.

Intro and Background Information:

What is your job title?

How do you explain what your job is to people who don't work in design or data visualisation?

How did you get into designing data visualisations?

What did you do before?

How do designers encode their visualisations with meaning?

How do you design a data visualisation so that it means something to someone else?

Are there certain design techniques that you use to ensure a data visualisation makes sense to the end user?

What are these techniques?

How did you learn them?

People have talked about telling a story about the data - is this something you think about when you design data visualisations?

Where does the story come from?

How do you use design to tell that story?

Who are the designers designing their visualisations for?

Do you think about the end user of the data visualisation when you are designing them?

Do you imagine who they are?

Do you think about how they are going to be interacting with the visualisation?

Does this influence the design decisions you make?

Do you think about how the user will read the data visualisation, for example how they will begin to interpret it?

Do you think data visualisations have the power to make people care about the issues they represent?

Why?

What are the socio-technical processes in data visualisation design practices?

Can you explain to me how you use software in your work?

What is good about these tools?

Are there any limitations to what you can do?

Do you ever, or have you ever, designed data visualisations in any other way?

Do you work with the data to make design decisions?

How do you do this?

What tools do you use?

Do you need to explain the meanings of your design to the front end developers who are making it?

How do you explain this to them?

Alternative Questions for FED's:

How do you know what the designer was trying to communicate when you are making their designs?

Do they have to explain this to you?

How do they do this?

Do you see the roles of Designer, Front End Developer and UX Designer as very different roles?

How are these roles different?

How do these differences play out in a data visualisation project?

Appendix d: User interview information sheet

+44 (0)1904 323234
sociology@york.ac.uk
york.ac.uk/sociology

Department of Sociology
University of York, York YO10 5DD

Relating to Data through Visualisation

Interviews with Diverse Users

About the project

We live in a world driven by data. It shapes policy, drives the economy and we even use data to learn about ourselves - think about the fitness apps we use to track our day-to-day activity.

To help us make use of all this data, designers are coming up with ever more creative ways of telling stories with it. They use complex graphs, charts and maps to help us *see* what the data have to tell us. These images, these visual expressions of data, are broadly termed - *data visualisations*.

But how do people make sense of numbers in the form of pictures? This project will explore the different ways people interpret visual expressions of data. It will also investigate if there are things that make it harder for some people to make sense of certain graphs and charts than others.

My aim is that the findings of this research will be put into practice to help designers produce better, more accessible data visualisations. A summary of the research findings will also be provided to participants at the end of the project.

Participation

Your participation in the research is voluntary and you are free to withdraw at any time prior to the submission and dissemination of the project, by contacting me. If you choose to take part you will be asked to complete an informed consent form prior to the interview taking place.

What will happen?

You will be asked to complete a short demographic questionnaire at the beginning of the interview. I will then ask you to look at between 6 and 8 different types of graphs and charts, including both print and digital examples. (Any digital and interactive examples will be viewed using a tablet computer).

After viewing each example you will be asked to talk about what the graph or chart represents. This is not a test of how well you are able to understand visualisations of data. I am interested in the different ways people make sense of these data visualisations. The interviews will be audio recorded, with all recordings deleted as soon as possible after the transcripts have been written up and anonymised.

Interviews will last up to an hour, plus an additional 15 minutes to allow for briefing and debriefing. They will be held in a public location, such as a café, at a date and time and location to be agreed between the participant and myself.

What will happen to the data/findings?

The data from the questionnaires and interviews will be stored securely and anonymised. Once the research project is complete, extracts from the data may be used on the project's website, in presentations, or in articles or books in which I share my findings. It will always be used anonymously – that is, you will not be identified as the source of the data.

Once the research is finished the researcher reserves the right to store the anonymised data with the UK Data Archive, and therefore it may be used by other researchers for the purpose of their own studies. If you would prefer this not to happen to the data you have provided, you can still take part in the project, but please make this clear on your consent form using the tick boxes provided.

Who is doing the research?

This research is being undertaken by (Margaret) Jill Simpson, a Sociology PhD student within the Department of Sociology at the University of York. The project is funded by the UK Government's Economic and Social Research Council and University of York. This project has been subject to an ethical review process by the Economics, Law, Management, Politics and Sociology Ethics Committee (ELMPS) at the University of York.

Project Contact Details:

If you have any concerns about your involvement in this project you can contact:

PhD Researcher: (Margaret) Jill Simpson.
Email: j.simpson@york.ac.uk

Research Supervisor: Dr David Beer.
Email: david.beer@york.ac.uk. Tel: 01904 323049

Chair of the ELMPS Ethics Committee: Professor Tony Royle.
Email: tony.royle@york.ac.uk. Tel: 01904 325061

Economic and Social Research Council
Shaping Society

Appendix e: User interview consent form

+44 (0)1904 323234
sociology@york.ac.uk
york.ac.uk/sociology

Department of Sociology
University of York, York YO10 5DD

Diverse user interviews

Informed Consent Form

Tick boxes if you agree with the statements:

I confirm that I have read and understand the information provided above and that I have been given the opportunity to ask questions.	Yes ☐ No ☐
I understand that my participation is voluntary and I am free to withdraw from the research at any time. I can pause or stop the interview at any time and I can skip any questions or visualisations I would rather not discuss.	Yes ☐ No ☐
I understand that my responses will be anonymised and my name will not be linked to the study, research materials or any publications relating to the study.	Yes ☐ No ☐
I agree to be audio recorded and for the responses I provide to be analysed for the purpose of research.	Yes ☐ No ☐
I agree to take part in the above research.	Yes ☐ No ☐

Use of the information I provide beyond this project

I agree for the researcher to reserve the right to archive the data I provide (in the form of anonymised transcripts and demographic information) at the UK Data Archive.	Yes ☐ No ☐
I understand that by archiving the anonymised data with the UK Data Archive other authenticated researchers will have access to it, if they agree to preserve the confidentiality of the information as requested in this form.	Yes ☐ No ☐

As such, I understand that other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.

Yes ☐ No ☐

_____	_____	_____
Name of participant [printed]	Signature	Date

_____	_____	_____
Researcher [printed]	Signature	Date

Project Contact Details:

If you have any concerns about your involvement in this project you can contact:

PhD Researcher: (Margaret) Jill Simpson.
Email: j.simpson@york.ac.uk

Research Supervisor: Dr David Beer.
Email: david.beer@york.ac.uk. Tel: 01904 323049

Chair of the ELMPS Ethics Committee: Professor Tony Royle.
Email: tony.royle@york.ac.uk. Tel: 01904 325061

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Economic and Social Research Council
Shaping Society

Appendix f: User interview demographic questionnaire

+44 (0)1904 323234
sociology@york.ac.uk
york.ac.uk/sociology

Department of Sociology
University of York, York YO10 5DD

Relating to data through Visualisation

Questionnaire for Participants

You do not have to answer all, or any, of the questions if you do not wish to do so. The answers you give are confidential and will be used anonymously.

This questionnaire has been designed by Jill Simpson, a researcher in the department of Sociology at the University of York. For more information on her research project, or for contact details, please refer to the information sheet.

Biographical Details	Age: _____ Gender: ☐ Female ☐ Male ☐ Own Definition (_____) ☐ Choose Not to Answer
	What is your ethnic group? Please choose one option that best describes your ethnic group or background. (Categories taken from the ONS) *White* ☐ British (English/Scottish/Welsh/Northern Irish) ☐ Irish ☐ White Gypsy or Irish traveller ☐ Any other white background (write below) *Mixed/Multiple Ethnic Groups* ☐ White and Black Caribbean ☐ White and Black African ☐ White and Asian ☐ Any other Mixed/Multiple Ethnic Background *Asian/Asian British* ☐ Indian ☐ Pakistani ☐ Bangladeshi

☐ Chinese ☐ Any other Asian/Asian British Background

Black/African/Caribbean/Black British

☐ African

☐ Caribbean ☐ Any other Black/African/Caribbean

Other Ethnic Group

☐ Arab ☐ Any other Ethnic Group

What is your employment status?

☐ Employed (full-time) ☐ Student

☐ Employed (part-time) ☐ Retired

☐ Self-employed ☐ Other (Please write below)

_____

If employed, Self-Employed, or Retired

Sector: _____

Job Title: _____

Length of Employment: _____

If Student

Course Title:

_____

Qualification:

_____

Year of Study:

_____

Which of the following qualifications do you currently hold? (mark all that apply)

☐ O Levels (passes)/GCSEs (A*-C)

☐ AS Levels, Higher Diploma

☐ A Levels/VCE's, Higher School Certificate, Progression/Advanced Diploma

☐ Apprenticeship

☐ Degree (for example, BA or BSc)

☐ Higher Degree (For example MA, PhD, PGCE)

☐ Professional Qualifications (for example, nursing, accountancy, teaching)

☐ Other Vocational/Work-Related Qualifications (please write below)

_____

☐ Other UK Qualifications Not Listed (please write below)

_____

	☐ Foreign Qualifications (please write below) _____ ☐ No Qualifications
	Do you consider yourself to have a disability? ☐ Yes ☐ No

Economic and Social Research Council
Shaping Society

Questionnaire Code……..

Appendix g: User interview schedule

Introduction

This is not a test of how well you're able to understand visualisations of data. I'm interested in the different ways people make sense of these graphs and charts.

So I'm going to ask you to look at some data visualisations and after you've seen each one I'd just like you to tell me about what the data visualisation is about.

If there are any you don't want to talk about that's fine.

Prompts:

What is this data visualisation about?

Is this a topic you are interested in?

Have you learnt anything from it?

In response to what they say: how? / why?

Did you get all this knowledge/information from the visualisation?

Is there anything about it (the visualisation) that you don't understand?

References

Aiello, G. (2020). ‘Inventorizing, situating, transforming: Social semiotics and data visualization.’ In Engebretsen M., & Kennedy H, (Eds.). *Data visualization in society*. Amsterdam: Amsterdam University Press, pp. 49-62.

Aiello, G. (2006). Theoretical advances in critical visual analysis: Perception, ideology, mythologies and social semiotics. *Journal of Visual Literacy*. 26(2), 89-102.

Akrich, M. (1995). ‘User representation practices, methods and sociology.’ In. Rip, A., Misa, T. J., and Schot, J, (Eds.). *Managing technology in society: the approach of constructive technology assessment.* London: Pinter Publishers, pp.167-184

Akrich, M. (1994). ‘The de-scription of technical objects.’ In Bijker, W. E., and Law, J. (Eds.). *Shaping technology/Building society.* Second printing 1997. Cambridge, Massachusetts: MIT Press, pp. 205-224

Amoore, L., and Piotukh, V. (2015). Life beyond big data: governing with little analytics. *Economy and Society*. 44(3), 341-366.

Amoore, L. (2013). *The politics of possibility: risk and security beyond probability.* Durham, NC; London : Duke University Press.

Amoore, L. (2009). Lines of sight: on the visualization of unknown futures. *Citizenship Studies*. 13(1), 17-30.

Anderson, C. (2008). *The end of theory: The data deluge makes the scientific method obsolete*. [Online]. Wired. 23 June 2008. Available at: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory [Last accessed 31 08 2020]

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., and Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*. 30(3), 791-800.

Atkinson, P., and Silverman, D. (1997) Kundera's *immortality:* The interview society and the invention of the self. *Qualitative Inquiry.* 3(3), 304-325.

Barthes, R (1977). 'Rhetoric of the Image.' In Barthes, R. *Image Music Text*. London: Fontana Press, pp. 32-51.

Beer, D. (2019). *The data gaze.* London: SAGE Publications Ltd.

Beer, D. (2016a). *Metric power.* Basingstoke: Palgrave Macmillan

Beer, D. (2016b). How should we do the history of big data? *Big Data and Society.* 3(1), 1-10.

Beer, D. (2013). *Popular culture and new media: The politics of circulation.* Basingstoke: Palgrave Macmillan.

Beer, D., and Burrows, R. (2013). Popular culture, digital archives and the new social life of data. *Theory, Culture & Society.* 30(4), 47-71.

Berger, J. (1972). *Ways of seeing.* Reissued 2008. London: Hamondsworth: British Broadcasting Corporation; Penguin.

Bødker, H. (2016). Stuart Hall's encoding/decoding model and the circulation of journalism in the digital landscape. *Critical studies in media communication.* 33(5), 409-423.

Boehnert, J. (2016). *Data Visualisation Does Political Things.* [Online]. Available at: https://westminsterresearch.westminster.ac.uk/download/ca109c03787d3730c041791d71d6ca5db0de14e7c22057b73d56b1faf860540b/6726771/Data_Visualisation_Does_Political_Things_-_DRS_2016_BOEHNERT_FINAL-April.pdf [last accessed 31 08 2020]

Boehnert, J. (2015). *Viewpoint: The Politics of Data Visualisation*. [Online]. Discover Society. 03 August 2015. Available at: http://discoversociety.org/2015/08/03/viewpoint-the-politics-of-data-visualisation/ [accessed 02.05.16].

Boellstorff, T. (2013). Making big data, in theory. *First Monday,* 18(10).

Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., and Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics.* 19(12), 2307-2315.

boyd, d., and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society.* 15(5), 662-679.

Brinkmann, S. and Kvale, S., (2015). *InterViews: Learning the craft of qualitative research interviewing.* 3rd ed. London: Sage

Brown, N. (2019). "Listen to Your Gut": A reflexive approach to data analysis. *The Qualitative Report.* 24(13), 31-43.

Buse, C., Nettleton, S., Martin, D., and Twigg, J. (2017). Imagined bodies: architects and their constructions of later life. *Ageing & Society.* 37(7), 1435-1457.

Byrne, D., and Doyle, A. (2004). 'The visual and the verbal: The interaction of images and discussion in exploring cultural change.' In Knowles, C. and Sweetman, P. (Eds.). *Picturing the social landscape: Visual methods and the Sociological Imagination.* Oxon: Routledge, pp. 166-177.

Couldry, N. (2016). 'Foreword.' In Kubitschko, A. and Kaun, A. (Eds.). *Innovative methods in media and communication research.* Basingstoke: Palgrave Macmillan, pp. v-viii.

Clarke, V., and Braun, V. (2013). *Successful qualitative research: a practical guide for beginners*. London: SAGE.

Clarke, A. E. (2005). 'From Chicago ecologies to situational analysis.' In Clarke, A. E. *Situational analysis: grounded theory after the postmodern turn*. [Electronic Resource] Thousand Oaks, CA; London: SAGE Publications, pp. 38-82. DOI: https://dx.doi.org/10.4135/9781412985833 [Accessed 07 08 2020].

Coffey, A. (1999). *The ethnographic self: fieldwork and the representation of identity.* London: SAGE Publications Ltd.

Crawford, K., Miltner, K., and Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. Special Section Introduction. *International Journal of Communication.* 8, 1663-1672.

Crawford, K. (2013). *The hidden biases in big data.* [Online]. Harvard Business Review. 01 April 2013. Web article available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data [accessed 09.06.16].

Dahlgren, P. (1988). What's the meaning of this? Viewers' plural sense-making of TV news. *Media, Culture and Society.* 10(3), 285-301.

Dick, M. (2015). Just fancy that. *Journalism Studies.* 16(2), 152-174.

D'Ignazio, C., and Bhargava, R. (2020). 'Data visualization literacy: A feminist starting point.' In Engebretsen, M. and Kennedy, H. (Eds.). *Data visualization in society.* Amsterdam: Amsterdam University Press, pp. 207-222.

D'Ignazio, C., and Klein, L. F. (2020). *Data Feminism*. Cambridge, Massachusetts: The MIT Press.

D'Ignazio, C., and Klein, L. (2016). *Feminist data visualization*. Presented at and published in the workshop proceedings from the Workshop on Visualization for the Digital Humanities at IEEE VIS Conference 2016. Available at: http://www.kanarinka.com/wp-content/uploads/2015/07/IEEE_Feminist_Data_Visualisation.pdf [Accessed 13.10.16].

Dourish, P., and Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society.* 5(2), 1-10.

Drucker, J. (2014). *Graphesis: Visual forms of knowledge production.* Cambridge, MA; London: Harvard University Press.

Dykes, B. (2016). *Data storytelling: The essential data science skill everyone needs.* [Online]. Forbes. 31 March 2016. Available at: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/ [last accessed 31 08 2020]

Du Gay, P., Hall, S., Janes, L., Mackay, H., and Negus, K. (1997). *Doing cultural studies: The story of the Sony Walkman*. Sage Publications Ltd: London.

Engebretsen, M., Kennedy, H. and Weber, W. (2018). Data visualization in Scandinavian newsrooms: Emerging trends in journalistic visualization practices. *Nordicom Review.* 39(2), 3-18.

Engebretsen, M., and Weber, W. (2017). 'Graphic modes: The visual representation of data.' In D. Perrin, D., and Cotter, C (Eds.). *The Routledge Handbook of Language and Media.* London: Routledge, pp. 277–295.

Erofeeva, M. (2019). On multiple agencies: When do things matter? *Information, Communication & Society.* 22(5), 590-604.

Espeland, W. (2015). 'Narrating numbers.' In Rottenburg, R., Merry, S. E., Park, S-J. and Mugler, J. (Eds.). *The World of Indicators.* Cambridge: Cambridge University Press, pp. 56-75.

Espeland, W. N., and Stevens, M. L. (2008). A Sociology of quantification. *European Journal of Sociology*. 49(3), 401-436.

Floyd, I. R., Jones, M. C., and Twidale, M. B. (2008). Resolving incommensurable debates: A preliminary identification of persona kinds, attributes, and characteristics. *Artefact.* 2(1), 12-26.

Gabrys, J., Pritchard, H., and Barratt, B. (2016). Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society,* 3(2), 1-14.

Geertz, C. (1973). *The interpretation of cultures: Selected essays by Clifford Geertz.* 2000 edition. New York: Basic Books

Gitelman, L., and Jackson, V. (2013). 'Introduction.' In Gitelman, L. (Eds.). "*Raw Data" is an oxymoron.* Cambridge, MA; London: The MIT Press, pp. 1-14.

Gray, J. (2020). 'The data epic: Visualization practices for narrating life and death at a distance.' In Engebretsen M., and Kennedy H. (Eds.). *Data visualization in society*. Amsterdam: Amsterdam University Press, pp. 313-328.

Hall, S. (1980). ‘Encoding/Decoding*’ In Hall, S., Hobson, D., Lowe, A. and Willis, P. (Eds.). *Culture, Media, Language.* Reprinted 1992. London: Routledge, pp. 128-138.

Hall, S. (1997). ‘Introduction.’ In Hall, S. (Eds.). *Representation cultural representations and signifying practices.* London: Sage Publications Ltd in association with the Open University, p. 1-11.

Hall, S. (1973). *Encoding and decoding in the television discourse*. Birmingham: Centre for Cultural Studies, University of Birmingham.

Halpern, O. (2014). *Beautiful Data: A history of vision and research since 1945.* Durham: Duke University Press.

Hammersely, M., and Atkinson, P. (2019). *Ethnography: Principles in practice*. 4th ed. Oxon: Routledge.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies.* 14(3), 575-599.

Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies.* 17(1), 13-26.

Harvey, W. S. (2010). Methodological approaches for interviewing elites. *Geography Compass.* 4(3), 193-205.

Hepworth, M. (1996). ‘William’ and the old folks: Notes on infantilisation. *Ageing and Society.* 16(4), 423-441.

Hill, R. L. (2017). The political potential of numbers: data visualisation in the abortion debate. *Women, Gender and Research,* 26(1), 83-96.

Hill, R.L., Kennedy, H., and Gerrard, Y. (2016). Visualising junk: Big data visualizations and the need for feminist data studies. *Journal of Communication Inquiry.* 40(4), 331-350.

Huang, W., Eades, P., and Hong, S-H. (2009). Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization.* 8(3), 139-152.

Hullman, J., and Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*. 17(12), 2231-2240.

Kennedy, H. (2020). *Simple data visualisations have become key to communicating about the COVID-19 pandemic, but we know little about their impact.* [Online]. LSE Blog. 04 May 2020. Available at: https://blogs.lse.ac.uk/impactofsocialsciences/2020/05/04/simple-data-visualisations-have-become-key-to-communicating-about-the-covid-19-pandemic-but-we-know-little-about-their-impact/ [last accessed 30 08 2020]

Kennedy, H. (2018). Living with Data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis.* Issue 1, 18-30.

Kennedy, H., and Hill, R.L. (2018). The Feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology.* 52(4), 830-848.

Kennedy, H., and Hill, R. L. (2017). The pleasure and pain of visualising data in times of data power. *Television and New Media.* 18(8), 769-782.

Kennedy, H., Hill, R. L., Aiello, G., and Allen, W. (2016a). The work that visualisation conventions do. *Information, Communication and Society.* 19(6), 715-735.

Kennedy, H., Hill, R. L., Allen, W., and Kirk, A. (2016b). Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions for effectiveness. *First Monday* 21(11).

Kirk, A. (2016). *Data visualisation: A handbook for data driven design.* London: Sage.

Kitchin, R. (2017). Thinking critically about researching algorithms. *Information, Communication and Society.* 20(1), 14-29.

Kitchin, R., Maalsen, S., and McArdle, G. (2016). The praxis and politics of building urban dashboards. *Geoforum.* 77, 93-101.

Kitchin, R. (2014). *The Data Revolution: Big data, open data, data infrastructures & their consequences.* London: SAGE Publications Ltd.

Kress, G., and Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication.* London: Hodder Education.

Laaksonn, S-M., and Pääkkönen, J. (2020). 'Between automation and interpretation: Using data visualization in social media analytics companies.' In Engebretsen M. and Kennedy H. (Eds.). *Data visualization in society*. Amsterdam: Amsterdam University Press, p. 95-108.

Lee, B., Henry Riche, N., Isenberg, P., and Carpendale, S. (2015). More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*. 35(5), 84-90.

Lofland, J., and Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis.* 3rd ed. Belmont, Calif; Wadsworth Publishing Company.

Lupton, D. (2018). How do data come to matter? Living and becoming with personal data. *Big Data & Society.* 5(2), 1-11.

Ma, K-L., Liao, I., Frazier, J, Hauser, H., and Kostis, H-N. (2012). Scientific storytelling using visualization. *IEEE Computer Graphics and Applications.* 32(1), 12-19.

Manovich, L. (2011). What is visualisation? *Visual Studies*, 26(1), 36-49.

Mason, J. (2002). *Qualitative researching*. 2nd ed. London: SAGE Publications Ltd

Massanari, A.L. (2010). Designing for imaginary friends: information architecture, personas and the politics of user-centred design. *New media & society.* 12(3). P. 401-416.

Masson, E., and Van Es, K. (2020). 'Rain on your radar: Engaging with weather data visualizations as part of everyday routines.' In Engebretsen M. & Kennedy H. (Eds.). *Data visualization in society.* Amsterdam: Amsterdam University Press, pp. 77-94.

Mayer-Schönberger, V., and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think.* London: John Murray (Publishers).

Morley, D. (1980). *The Nationwide Audience: Structure and Decoding.* London: British Film Institute.

Nafus, D. (2014). Stuck data, dead data, and disloyal data: the stops and starts in making numbers into social practices. *Distinktion: Journal of Social Theory,* 15(02), 208 – 222.

Nairn, K., Munro, J., and Smith, A.B. (2005). A counter-narrative of a 'failed' interview. *Qualitative Research,* 5(2), 221-244.

Oudshoorn, N., Neven, L., and Stienstra, M. (2016). How diversity gets lost: Age and gender in design practices of information and communication technologies. *Journal or Women & Aging.* 28(2), 170-185.

Oudshoorn, N., and Pinch, T. (2005). 'Introduction: How users and non-users matter.' In Oudshoorn, N. and Pinch, T. (Eds.). *How users matter: The co-construction of users and technology.* Cambridge, Massachusetts: The MIT Press, pp. 1-25.

Oudshoorn, N., Rommes and E., and Stienstra, M. (2004). Configuring the user as everybody: Gender and design cultures in information and communication technologies. *Science, Technology and Human Values.* 29(1), 30-63.

Pink, S., and Morgan, J. (2013). Short-term ethnography: Intense routes to knowing. *Symbolic Interaction.* 36(3), 351-361.

Poland, B., and Pederson, A. (1998). Reading between the lines: Interpreting silences in qualitative research. *Qualitative Inquiry* 4(2), 293-312.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton: Princeton University Press.

Potter, J. (2004). *Theory of media literacy: A cognitive approach.* Thousand Oaks, California: Sage Publications, Inc.

Reyes, V. (2020). Ethnographic toolkit: Strategic positionality and researchers' visible and invisible tools in field research. *Ethnography.* 21(2), 220-240.

Rieder, G., and Simon, J. (2016). Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society*. 3(1), 1-6.

Rommes, E. W. M. (2002). *Gender scripts and the internet: The design and use of Amsterdam's digital city.* Enschede, the Netherlands: Twente University Press.

Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials.* 4th ed. London: SAGE Publications Ltd.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics.* 25(1), 682-692.

Scott, A. (2017). *What is 'open data' and why should we care?* [Online] Open Data Institute. 03 November 2017. Available at: https://theodi.org/article/what-is-open-data-and-why-should-we-care/ [last accessed 04 09 2020]

Segel, E., and Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*. 16(6), 1139-1148.

Shah, P., and Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review.* 14(1), 47-69.

Simpson, J. (2020). 'Visualizing data: A lived experience.' In Engebretsen, M. and Kennedy, H. (Eds.). *Data visualization in society.* Amsterdam: Amsterdam University Press, pp. 157-168.

Sørensen, K. H., Faulkner, W., and Rommes, E. (2011). *Technologies of inclusion: Gender in the information society.* Trondheim, Norway: Tapir Academic Press

Stanley, L., and Wise, S. (1983). *Breaking out: Feminist consciousness and feminist research.* London: Routledge & K. Paul.

Steiner, L. (2018). Wrestling with the angels: Stuart Hall's theory and method. *Howard Journal of Communication.* 27(2), 102-111.

Stewart, J., and Williams, R. (2005). 'The wrong trousers? Beyond the design fallacy: Social learning and the user.' In Howcroft, D. and Trauth, E. M. *Handbook of Critical Information Systems Research: Theory and Application.* Cheltenham: Edward Elgar Publishing Limited, pp. 195-221.

Taylor, A., Lindley, S., Regan, T. and Sweeney, D. (2014). Data and life on the street. *Big Data & Society.* 1(2), 1-7.

Tilley, L., and Woodthorpe, K. (2011). Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research.* 11(2), 197-212.

Tufte, E. R. (2001). *The visual display of quantitative information.* 2nd ed. Seventh printing March 2011. Cheshire, Connecticut: Graphics Press LLC.

Tufte, E, R. (1990). *Envisioning information.* Reprint 1992. Cheshire, Connecticut: Graphics Press

Turner, P., and Turner, S. (2011). Is stereotyping inevitable when designing with personas? *Design Studies.* 32(1), 30-44.

Twigg, J. (2012). Adjusting the cut: fashion, the body and age on the UK high street. *Ageing & Society*. 32(6), 1030-1054.

Uprichard, E. (2013). *Focus: Big data, little questions?* [Online] Discover Society. 01 October 2013. Available at: http://discoversociety.org/2013/10/01/focus-big-data-little-questions/ [accessed 20.07.16].

Veel, K. (2018). Make data sing: The automation of storytelling. *Big Data & Society*. 5(1), 1-8.

Vyas, D., van der Veer, G., and Nijholt, A. (2013). Creative practices in the design studio culture: collaboration and communication. *Cogn Tech Work.* 15, 415-443.

Watts, J. (2020). *Interview with Bruno Latour: 'This is a global catastrophe that has come from within.'* [Online]. The Observer. 06 June 2020. Available at: https://www.theguardian.com/world/2020/jun/06/bruno-latour-coronavirus-gaia-hypothesis-climate-crisis [last accessed 31 08 2020]

Weber, W. (2020). 'Exploring narrativity in data visualization in journalism.' In Engebretsen M. and Kennedy H. (Eds.). *Data visualization in Society*. Amsterdam: Amsterdam University Press, pp. 295-312.

Weber, W., Engebretsen, M., and Kennedy, H. (2018). Data stories: Rethinking journalistic storytelling in the context of data journalism. *Studies in Communication Sciences.* 18(1), 191-206.

Woolgar, S. (1991). 'Configuring the user: the case of usability trials.' In Law, J. (Eds.). *A Sociology of monsters: essays on power, technology, and domination.* London; New York: Routledge, pp. 57-99.