

1

Defining Data Visualisation

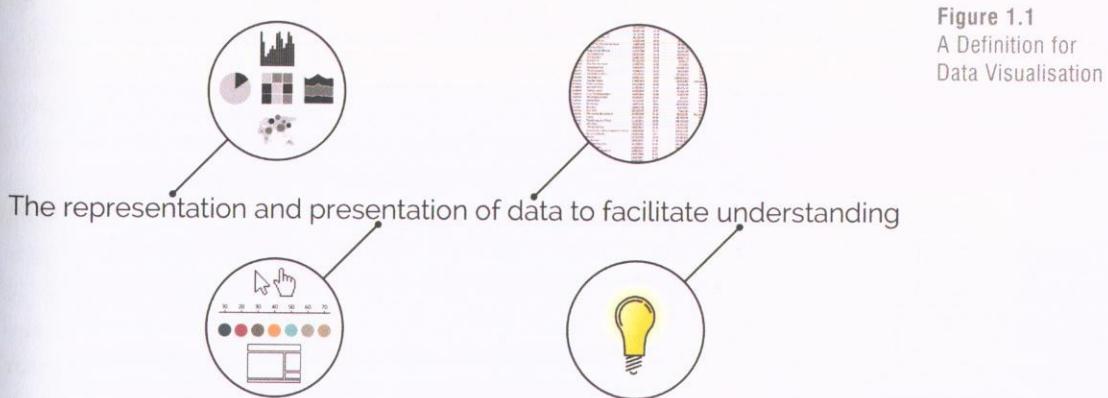
This opening chapter will introduce you to the subject of data visualisation, defining what data visualisation is and is not. It will outline the different ingredients that make it such an interesting recipe and establish a foundation of understanding that will form a key reference for all of the decision making you are faced with.

Three core principles of good visualisation design will be presented that offer guiding ideals to help mould your convictions about distinguishing between effective and ineffective in data visualisation.

You will also see how data visualisation sits alongside or overlaps with other related disciplines, and some definitions about the use of language in this book will be established to ensure consistency in meaning across all chapters.

1.1 The Components of Understanding

To set the scene for what is about to follow, I think it is important to start this book with a proposed definition for data visualisation (Figure 1.1). This definition offers a critical term of reference because its components and their meaning will touch on every element of content that follows in this book. Furthermore, as a subject that has many different proposed definitions, I believe it is worth clarifying my own view before going further:

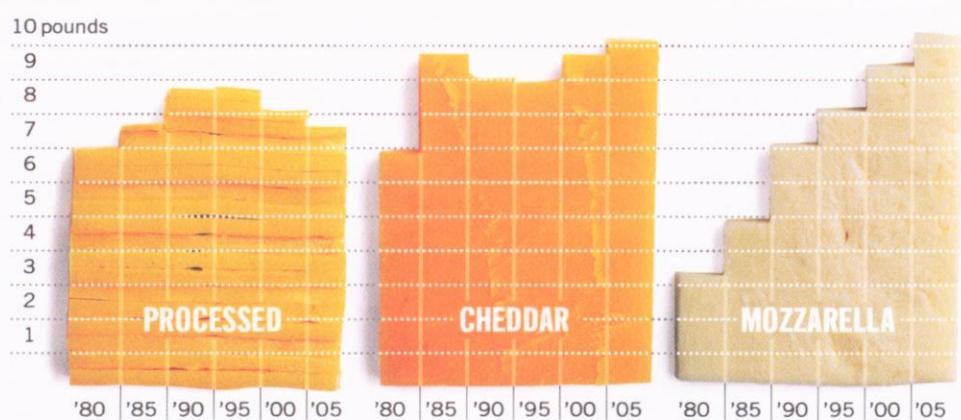


At first glance this might appear to be a surprisingly short definition: isn't there more to data visualisation than that, you might ask? Can nine words sufficiently articulate what has already been introduced as an eminently complex and diverse discipline?

I have arrived at this after many years of iterations attempting to improve the elegance of my definition. In the past I have tried to force too many words and too many clauses into one statement, making it cumbersome and rather undermining its value. Over time, as I have developed greater clarity in my own convictions, I have in turn managed to establish greater clarity about what I feel is the real essence of this subject. The definition above is, I believe, a succinct and practically useful description of what the pursuit of visualisation is truly about. It is a definition that largely informs the contents of this book. Each chapter will aim to enlighten you about different aspects of the roles of and relationships between each component expressed. Let me introduce and briefly examine each of these one by one, explaining where and how they will be discussed in the book.

Firstly, **data**, our critical raw material. It might appear a formality to mention data in the definition for, after all, we are talking about data visualisation as opposed to, let's say, cheese visualisation (though visualisation of data using cheese has happened, see Figure 1.2), but it needs to be made clear the core role that data has in the design process. Without data there is no visualisation; indeed there is no need for one. Data plays the fundamental role in this work, so you will need to give it your undivided attention and respect. You will discover in Chapter 4 the importance of developing an intimacy with your data to acquaint yourself with its physical properties, its meaning and its potential qualities.

Figure 1.2
Per Capita
Cheese
Consumption in
the US



Data is names, amounts, groups, statistical values, dates, comments, locations. Data is textual and numeric in format, typically held in datasets in table form, with rows of records and columns of different variables.

This tabular form of data is what we will be considering as the raw form of data. Through tables, we can *look* at the values contained to precisely read them as individual data points. We can look up values quite efficiently, scanning across many variables for the different records held.

However, we cannot easily establish the comparative size and relationship between multiple data points. Our eyes and mind are not equipped to translate easily the textual and numeric values into quantitative and qualitative meaning. We can look at the data but we cannot really see it without the context of relationships that help us compare and contrast them effectively with other values. To derive understanding from data we need to see it represented in a different, visual form. This is the act of **data representation**.

This word *representation* is deliberately positioned near the front of the definition because it is the quintessential activity of data visualisation design. Representation concerns the choices made about the form in which your data will be visually portrayed: in lay terms, what chart or charts you will use to exploit the brain's visual perception capabilities most effectively.

When data visualisers create a visualisation they are representing the data they wish to show visually through combinations of *marks* and *attributes*. Marks are points, lines and areas. Attributes are the appearance properties of these marks, such as the size, colour and position. The recipe of these marks and their attributes, along with other components of apparatus, such as axes and gridlines, form the anatomy of a chart.

In Chapter 6 you will gain a deeper and more sophisticated appreciation of the range of different charts that are in common usage today, broadening your visual vocabulary. These charts will vary in complexity and composition, with each capable of accommodating different types of data and portraying different angles of analysis. You will learn about the key ingredients that shape your data representation decisions, explaining the factors that distinguish the effective from the ineffective choices.

Beyond representation choices, the **presentation** of data concerns all the other visible design decisions that make up the overall visualisation anatomy. This includes choices about the possible applications of interactivity, features of annotation, colour usage and the composition of your work. During the early stages of learning this subject it is sensible to partition your thinking about these matters, treating them as isolated design layers. This will aid your initial critical thinking. Chapters 7–10 will explore each of these layers in depth, profiling the options available and the factors that influence your decisions.

However, as you gain in experience, the interrelated nature of visualisation will become much more apparent and you will see how the overall design anatomy is entirely connected. For instance, the selection of a chart type intrinsically leads to decisions about the space and place it will occupy; an interactive control may be included to reveal an annotated caption; for any design property to be even visible to the eye it must possess a colour that is different from that of its background.

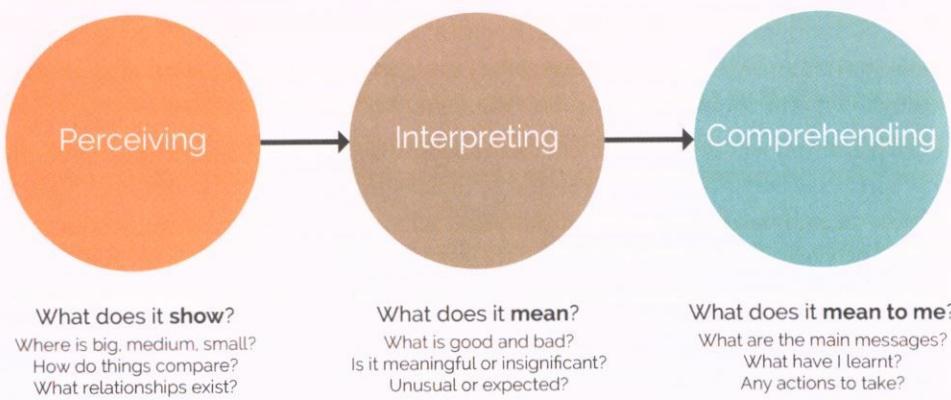
The goal expressed in this definition states that data visualisation is about **facilitating understanding**. This is very important and some extra time is required to emphasise why it is such an influential component in our thinking. You might think you know what *understanding* means, but when you peel back the surface you realise there are many subtleties that need to be acknowledged about this term and their impact on your data visualisation choices. Understanding 'understanding' (still with me?) in the context of data visualisation is of elementary significance.

These are not just synonyms for the same word, rather they carry important distinctions that need appreciating. As you will see throughout this book, the subtleties and semantics of language in data visualisation will be a recurring concern.

When consuming a visualisation, the viewer will go through a process of understanding involving three stages: *perceiving*, *interpreting* and *comprehending* (Figure 1.3). Each stage is dependent on the previous one and in your role as a data visualiser you will have influence but not full control over these. You are largely at the mercy of the viewer – what they know and do not know, what they are interested in knowing and what might be meaningful to them – and this introduces many variables outside of your control:

where your control diminishes the influence and reliance on the viewer increases. Achieving an outcome of understanding is therefore a collective responsibility between visualiser and viewer.

Figure 1.3
The Three
Stages of
Understanding



Let's look at the characteristics of the different stages that form the process of understanding to help explain their respective differences and mutual dependencies.

Firstly, *perceiving*. This concerns the act of simply being able to read a chart. What is the chart showing you? How easily can you get a sense of the values of the data being portrayed?

- Where are the largest, middle-sized and smallest values?
- What proportion of the total does that value hold?
- How do these values compare in ranking terms?
- To which other values does this have a connected relationship?

The notion of understanding here concerns our attempts as viewers to efficiently decode the representations of the data (the shapes, the sizes and the colours) as displayed through a chart, and then convert them into perceived values: estimates of quantities and their relationships to other values.

Interpreting is the next stage of understanding following on from perceiving. Having read the charts the viewer now seeks to convert these perceived values into some form of meaning:

- Is it good to be big or better to be small?
- What does it mean to go up or go down?
- Is that relationship meaningful or insignificant?
- Is the decline of that category especially surprising?

The viewer's ability to form such interpretations is influenced by their pre-existing knowledge about the portrayed subject and their capacity to utilise that knowledge to frame the implications of what has been read. Where a viewer does not possess that knowledge it may be that the visualiser has to address this deficit. They will need to make suitable design choices that help to make clear what meaning can or should be drawn from the display of data. Captions, headlines, colours and other annotated devices, in particular, can all be used to achieve this.

Comprehending involves reasoning the consequence of the perceiving and interpreting stages to arrive at a personal reflection of what all this means *to them, the viewer*. How does this information make a difference to what was known about the subject previously?

- Why is this relevant? What wants or needs does it serve?
- Has it confirmed what I knew or possibly suspected beforehand or enlightened me with new knowledge?
- Has this experience impacted me in an emotional way or left me feeling somewhat indifferent as a consequence?
- Does the context of what understanding I have acquired lead me to take action – such as make a decision or fundamentally change my behaviour – or do I simply have an extra grain of knowledge the consequence of which may not materialise until much later?

Over the page is a simple demonstration to further illustrate this process of understanding. In this example I play the role of a viewer working with a sample isolated chart (Figure 1.4). As you will learn throughout the design chapters, a chart would not normally just exist floating in isolation like this one does, but it will serve a purpose for this demonstration.

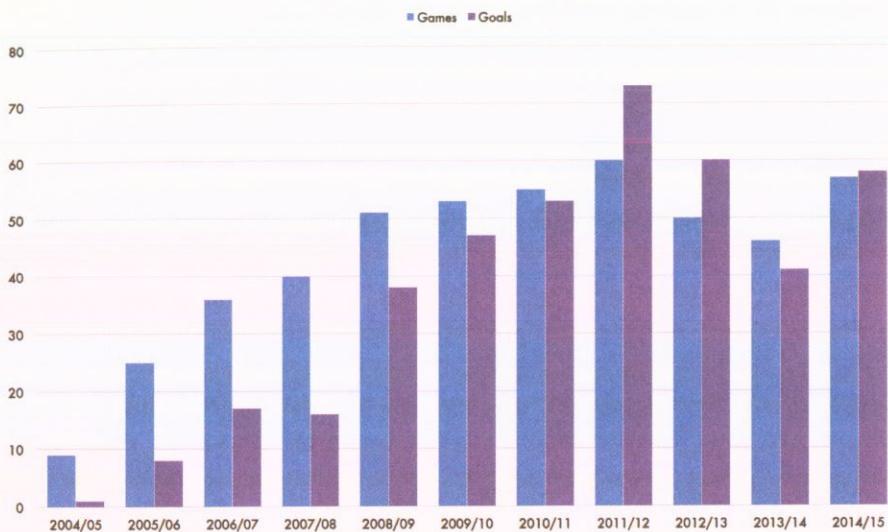
Figure 1.4 shows a clustered bar chart that presents a breakdown of the career statistics for the footballer Lionel Messi during his career with FC Barcelona.

The process commences with *perceiving* the chart. I begin by establishing what chart type is being used. I am familiar with this clustered bar chart approach and so I quickly feel at ease with the prospect of reading its display: there is no learning for me to have to go through on this occasion, which is not always the case as we will see.

I can quickly assimilate what the axes are showing by examining the labels along the x- and y-axes and by taking the assistance provided by colour legend at the top. I move on to scanning, detecting and observing the general physical properties of the data being represented. The eyes and brain are working in harmony, conducting this activity quite instinctively without awareness or delay, noting the most prominent features of variation in the attributes of size, shape, colour and position.

Figure 1.4
Demonstrating
the
Process of
Understanding

Lionel Messi: Games and Goals for FC Barcelona



I look across the entire chart, identifying the big, small and medium values (these are known as stepped magnitude judgements), and form an overall sense of the general value rankings (global comparison judgements). I am instinctively drawn to the dominant bars towards the middle/right of the chart, especially as I know this side of the chart concerns the most recent career performances. I can determine that the purple bar – showing goals – has been rising pretty much year-on-year towards a peak in 2011/12 and then there is a dip before recovery in his most recent season.

My visual system is now working hard to decode these properties into estimations of quantities (amounts of things) and relationships (how different things compare with each other). I focus on judging the absolute magnitudes of individual bars (one bar at a time). The assistance offered by the chart apparatus, such as the vertical axis (or y-axis) values and the inclusion of gridlines, is helping me more quickly estimate the quantities with greater assurance of accuracy, such as discovering that the highest number of goals scored was around 73.

I then look to conduct some relative higher/lower comparisons. In comparing the games and goals pairings I can see that three out of the last four years have seen the purple bar higher than the blue bar, in contrast to all the rest. Finally I look to establish proportional relationships between neighbouring bars, i.e. by how much larger one is compared with the next. In 2006/07 I can see the blue bar is more than twice as tall as the purple one, whereas in 2011/12 the purple bar is about 15% taller.

By reading this chart I now have a good appreciation of the quantities displayed and some sense of the relationship between the two measures, games and goals.

The second part of the understanding process is *interpreting*. In reality, it is not so consciously consecutive or delayed in relationship to the perceiving stage but you cannot get here without having already done the perceiving. Interpreting, as you will recall, is about converting

perceived ‘reading’ into meaning. Interpreting is essentially about orientating your assessment of what you’ve read against what you know about the subject.

As I mentioned earlier, often a data visualiser will choose to – or have the opportunity to – share such insights via captions, chart overlays or summary headlines. As you will learn in Chapter 3, the visualisations that present this type of interpretation assistance are commonly described as offering an ‘explanatory’ experience. In this particular demonstration it is an example of an ‘exhibitory’ experience, characterised by the absence of any explanatory features. It relies on the viewer to handle the demands of interpretation without any assistance.

As you will read about later, many factors influence how well different viewers will be able to interpret a visualisation. Some of the most critical include the level of interest shown towards the subject matter, its relevance and the general inclination, in that moment, of a viewer to want to read about that subject through a visualisation. It is also influenced by the knowledge held about a subject or the capacity to derive meaning from a subject even if a knowledge gap exists.

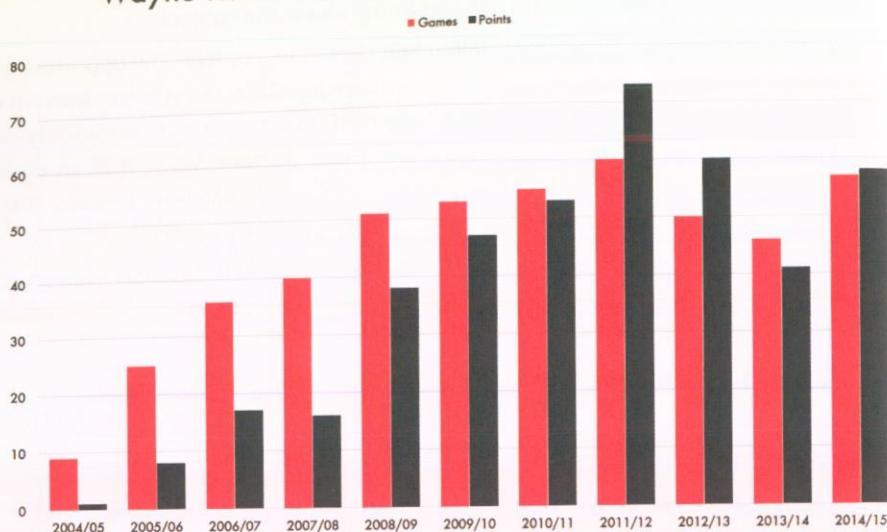
Returning to the sample chart, in order to translate the quantities and relationships I extracted from the perceiving stage into meaning, I am effectively converting the reading of value sizes into notions of *good* or *bad* and comparative relationships into *worse than* or *better than* etc. To interpret the meaning of this data about Lionel Messi I can tap into my passion for and knowledge of football. I know that for a player to score over 25 goals in a season is very good. To score over 35 is exceptional. To score over 70 goals is frankly preposterous, especially at the highest level of the game (you might find plenty of players achieving these statistics playing for the Dog and Duck pub team, but these numbers have been achieved for Barcelona in La Liga, the Champions League and other domestic cup competitions). I know from watching the sport, and poring over statistics like this for 30 years, that it is very rare for a player to score remotely close to a ratio of one goal per game played. Those purple bars that exceed the height of the blue bars are therefore remarkable. Beyond the information presented in the chart I bring knowledge about the periods when different managers were in charge of Barcelona, how they played the game, and how some organised their teams entirely around Messi’s talents. I know which other players were teammates across different seasons and who might have assisted or hindered his achievements. I also know his age and can mentally compare his achievements with the traditional football career arcs that will normally show a steady rise, peak, plateau, and then decline.

Therefore, in this example, I am not just interested in the subject but can bring a lot of knowledge to aid me in interpreting this analysis. That helps me understand a lot more about what this data means. For other people they might be passingly interested in football and know how to read what is being presented, but they might not possess the domain knowledge to go deeper into the interpretation. They also just might not care. Now imagine this was analysis of, let’s say, an NHL ice hockey player (Figure 1.5) – that would present an entirely different challenge for me.

In this chart the numbers are irrelevant, just using the same chart as before with different labels. Assuming this was real analysis, as a sports fan in general I would have the capacity to understand the notion of a sportsperson’s career statistics in terms of games played and goals

Figure 1.5
Demonstrating
the Process of
Understanding

Wayne Kane: Games and Points for Toronto Rangers



scored: I can read the chart (*perceiving*) that shows me this data and catch the gist of the angle of analysis it is portraying. However, I do not have sufficient domain knowledge of ice hockey to determine the real meaning and significance of the big–small, higher–lower value relationships. I cannot confidently convert ‘small’ into ‘unusual’ or ‘greater than’ into ‘remarkable’. My capacity to interpret is therefore limited, and besides I have no connection to the subject matter, so I am insufficiently interested to put in the effort to spend much time with any in-depth attempts at interpretation.

Imagine this is now no longer analysis about sport but about the sightings in the wild of *Winglets* and *Spungles* (completely made up words). Once again I can still read the chart shown in Figure 1.6 but now I have absolutely no connection to the subject whatsoever. No knowledge and no interest. I have no idea what these things are, no understanding about the sense of scale that should be expected for these sightings, I don’t know what is good or bad. And I genuinely don’t care either. In contrast, for those who do have a knowledge of and interest in the subject, the meaning of this data will be much more relevant. They will be able to read the chart and make some sense of the meaning of the quantities and relationships displayed.

To help with *perceiving*, viewers need the context of scale. To help with *interpreting*, viewers need the context of subject, whether that is provided by the visualiser or the viewer themselves. The challenge for you and I as data visualisers is to determine what our audience will know already and what they will need to know in order to possibly assist them in interpreting the meaning. The use of explanatory captions, perhaps positioned in that big white space top left, could assist those lacking the knowledge of the subject, possibly offering a short narrative to make the interpretations – the meaning – clearer and immediately accessible.

We are not quite finished, there is one stage left. The third part of the understanding process is *comprehending*. This is where I attempt to form some concluding reasoning that translates into

Total Sightings of Winglets and Spungles

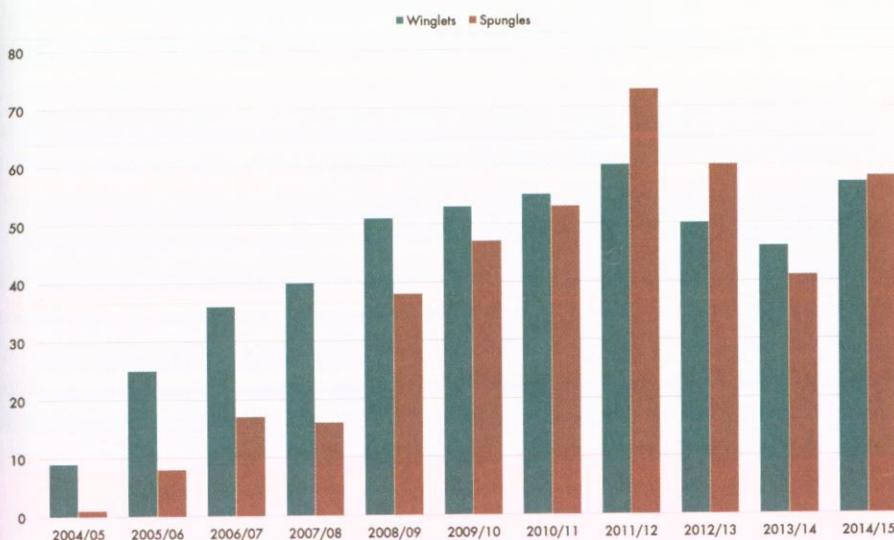


Figure 1.6
Demonstrating
the Process of
Understanding

what this analysis means for *me*. What can I infer from the display of data I have read? How do I relate and respond to the insights I have drawn out as through interpretation? Does what I've learnt make a difference to me? Do I know something more than I did before? Do I need to act or decide on anything? How does it make me feel emotionally?

Through consuming the Messi chart, I have been able to form an even greater appreciation of his amazing career. It has surprised me just how prolific he has been, especially having seen his ratio of goals to games, and I am particularly intrigued to see whether the dip in 2013/14 was a temporary blip or whether the bounce back in 2014/15 was the blip. And as he reaches his late 20s, will injuries start to creep in as they seem to do for many other similarly prodigious young talents, especially as he has been playing relentlessly at the highest level since his late teens?

My comprehension is not a dramatic discovery. There is no sudden inclination to act nor any need – based on what I have learnt. I just feel a heightened impression, formed through the data, about just how good and prolific Lionel Messi has been. For Barcelona fanatics who watch him play every week, they will likely have already formed this understanding. This kind of experience would only have reaffirmed what they already probably knew.

And that is important to recognise when it comes to managing expectations about what we hope to achieve amongst our viewers in terms of their final comprehending. One person's 'I knew that already' is another person's 'wow'. For every 'wow, I need to make some changes' type of reflection there might be another 'doesn't affect me'. A compelling visualisation about climate change presented to Sylvie might affect her significantly about the changes she might need to make in her lifestyle choices that might reduce her carbon footprint. For Robert, who is already familiar with the significance of this situation, it might have substantially less immediate impact – not indifference to the meaning of the data, just nothing new, a shrug of the

shoulders. For James, the hardened sceptic, even the most indisputable evidence may have no effect; he might just not be receptive to altering his views regardless.

What these scenarios try to explain is that, from your perspective of the visualiser, this final stage of understanding is something you will have relatively little control over because viewers are people and people are complex. People are different and as such they introduce inconsistencies. You can lead a horse to water but you cannot make it drink: you cannot force a viewer to be interested in your work, to understand the meaning of a subject or get that person to react exactly how you would wish.

Visualising data is just an agent of communication and not a guarantor for what a viewer does with the opportunity for understanding that is presented. There are different flavours of comprehension, different consequences of understanding formed through this final stage. Many visualisations will be created with the ambition to simply inform, like the Messi graphic achieved for me, perhaps to add just an extra grain to the pile of knowledge a viewer has about a subject. Not every visualisation results in a Hollywood moment of grand discoveries, surprising insights or life-saving decisions. But that is OK, so long as the outcome fits with the intended purpose, something we will discuss in more depth in Chapter 3.

Furthermore, there is the complexity of human behaviour in how people make decisions in life. You might create the most compelling visualisation, demonstrating proven effective design choices, carefully constructed with very a specific audience type and need in mind. This might clearly show how a certain decision really needs to be taken by those in the audience. However, you cannot guarantee that the decision maker in question, while possibly recognising that there is a *need* to act, will be in a *position* to act, and indeed will know *how* to act.

It is at this point that one must recognise the ambitions and – more importantly – realise the limits of what data visualisation can achieve. Going back again, finally, to the components of the definition, all the reasons outlined above show why the term **to facilitate** is the most a visualiser can reasonably aspire to achieve.

It might feel like a rather tepid and unambitious aim, something of a cop-out that avoids scrutiny over the outcomes of our work: why not aim to ‘deliver’, ‘accomplish’, or do something more earnest than just ‘facilitate’? I deliberately use ‘facilitate’ because as we have seen we can only control so much. Design cannot change the world, it can only make it run a little smoother. Visualisers can control the output but not the outcome: at best we can expect to have only some influence on it.

1.2 The Importance of Conviction

The key structure running through this book is a data visualisation design process. By following this process you will be able to decrease the size of the challenge involved in making good decisions about your design solution. The sequencing of the stages presented will help reduce the myriad options you have to consider, which makes the prospect of arriving at the *best* possible solution much more likely to occur.

Often, the design choices you need to make will be clear cut. As you will learn, the preparatory nature of the first three stages goes a long way to securing that clarity later in the design stage. On other occasions, plain old common sense is a more than sufficient guide. However, for more nuanced situations, where there are several potentially viable options presenting themselves, you need to rely on the guiding value of good design principles.

For many people setting out on their journey in data visualisation, the major influences that shape their early beliefs about data visualisation design tend to be influenced by the first authors they come across. Names like Edward Tufte, unquestionably one of the most important figures in this field whose ideas are still pervasive, represent a common entry point into the field, as do people like Stephen Few, David McCandless, Alberto Cairo, and Tamara Munzner, to name but a few. These are authors of prominent works that typically represent the first books purchased and read by many beginners.

Where you go from there – from whom you draw your most valuable enduring guidance – will be shaped by many different factors: taste, the industry you are working in, the topics on which you work, the types of audiences you produce for. I still value much of what Tufte extols, for example, but find I can now more confidently filter out some of his ideals that veer towards impractical ideology or that do not necessarily hold up against contemporary technology and the maturing expectations of people.

The key guidance that now most helpfully shapes and supports my convictions comes from ideas outside the boundaries of visualisation design in the shape of the work of Dieter Rams. Rams was a German industrial and product designer who was most famously associated with the Braun company.

In the late 1970s or early 1980s, Rams was becoming concerned about the state and direction of design thinking and, given his prominent role in the industry, felt a responsibility to challenge himself, his own work and his own thinking against a simple question: ‘Is my design *good* design?’. By dissecting his response to this question he conceived 10 principles that expressed the most important characteristics of what he considered to be good design. They read as follows:

1. Good design is innovative.
2. Good design makes a product useful.
3. Good design is aesthetic.
4. Good design makes a product understandable.
5. Good design is unobtrusive.
6. Good design is honest.
7. Good design is long lasting.
8. Good design is thorough down to the last detail.
9. Good design is environmentally friendly.
10. Good design is as little design as possible.

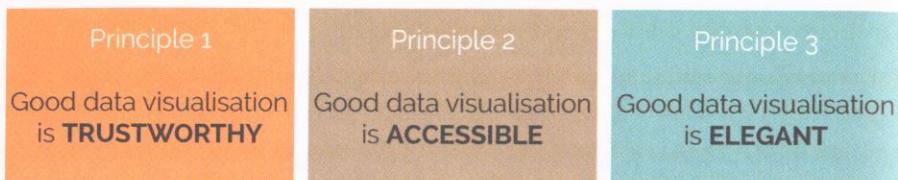
‘I say begin by learning about data visualisation’s “black and whites”, the rules, then start looking for the greys. It really then becomes quite a personal journey of developing your conviction.’

Jorge Camoes, Data Visualization Consultant

‘My key guiding principle? Know the rules, before you break them.’ Gregor Aisch, Graphics Editor, *The New York Times*

Inspired by the essence of these principles, and considering their applicability to data visualisation design, I have translated them into three high-level principles that similarly help me to answer my own question: ‘Is my visualisation design *good* visualisation design?’ These principles offer me a guiding voice when I need to resolve some of the more seemingly intangible decisions I am faced with (Figure 1.7).

Figure 1.7
The Three
Principles
of Good
Visualisation
Design



In the book *Will it Make the Boat Go Faster?*, co-author Ben Hunt-Davis provides details of the strategies employed by him and his team that led to their achieving gold medal success in the Men’s Rowing Eight event at the Sydney Olympics in 2000. As the title suggests, each decision taken had to pass the ‘will it make the boat go faster?’ test. Going back to the goal of data visualisation as defined earlier, these design principles help me judge whether any decision I make will better aid the facilitation of understanding: the equivalence of ‘making the boat go faster’.

I will describe in detail the thinking behind each of these principles and explain how Rams’ principles map onto them. Before that, let me briefly explain why there are three principles of Rams’ original ten that do not entirely fit, in my view, as *universal* principles for data visualisation.

Good design is innovative: Data visualisation does not need always to be innovative. For the majority of occasions the solutions being created call upon the tried and tested approaches that have been used for generations. Visualisers are not conceiving new forms of representation

‘I’m always the fool looking at the sky who falls off the cliff. In other words, I tend to seize on ideas because I’m excited about them without thinking through the consequences of the amount of work they will entail. I find tight deadlines energizing. Answering the question of “what is the graphic trying to do?” is always helpful. At minimum the work I create needs to speak to this. Innovation doesn’t have to be a wholesale out-of-the box approach. Iterating on a previous idea, moving it forward, is innovation.’ **Sarah Slobin, Visual Journalist**

or implementing new design techniques in every project. Of course, there are times when innovation is required to overcome a particular challenge; innovation generally materialises when faced with problems that current solutions fail to overcome. Your own desire for innovation may be aligned to personal goals about the development of your skills or through reflecting on previous projects and recognising a desire to rethink a solution. It is not that data visualisation is never about innovation, just that it is not always and only about innovation.

Good design is long lasting: The translation of this principle to the context of data visualisation can be taken in different ways. ‘Long lasting’ could be related to the desire to preserve the ongoing functionality of a digital project, for example. It is quite demoralising how many historic links you visit online only to find a project has now expired through a lack of sustained support or is no longer functionally supported on modern browsers.

Another way to interpret ‘long lasting’ is in the durability of the technique. Bar charts, for example, are the *old reliables* of the field – always useful, always being used, always there when you need them (author wipes away a respectful tear). ‘Long lasting’ can also relate to avoiding the temptation of fashion or current gimmickry and having a timeless approach to design. Consider the recent design trend moving away from *skeuomorphism* and the emergence of so-called *flat* design. By the time this book is published there will likely be a new movement. ‘Long lasting’ could apply to the subject matter. Expiry in the relevance of certain angles of analysis or out-of-date data is inevitable in most of our work, particularly with subjects that concern current matters. Analysis about the loss of life during the Second World War is timeless because nothing is now going to change the nature or extent of the underlying data (unless new discoveries emerge). Analysis of the highest grossing movies today will change as soon as new big movies are released and time elapses. So, once again, this idea of long lasting is very context specific, rather than being a universal goal for data visualisation.

Good design is environmentally friendly: This is, of course, a noble aim but the relevance of this principle has to be positioned again at the contextual level, based on the specific circumstances of a given project. If your work is to be printed, the ink and paper usage immediately removes the notion that it is an environmentally friendly activity. Developing a powerful interactive that is being hammered constantly and concurrently by hundreds of thousands of users puts an extra burden on the hosting server, creating more demands on energy supply. The specific judgements about issues relating to the impact of a project on the environment realistically reside with the protagonists and stakeholders involved.

A point of clarity is that, while I describe them as *design* principles, they actually provide guidance long before you reach the design thinking at the final stage of this workflow. Design choices encapsulate the critical thinking undertaken throughout. Think of it like an iceberg: the design is the visible consequences of lots of hidden preparatory thinking formed through earlier stages.

Finally, a comment is in order about something often raised in discussions about the principles for this subject: that is, the idea that visualisations need to be memorable. This is, in my view, not relevant as a universal principle. If something is memorable, wonderful, that will be a terrific by-product of your design thinking, but in itself the goal of achieving memorability has to be isolated, again, to a contextual level based on the specific goals of a given task and the capacity of the viewer. A politician or a broadcaster might need to recall information more readily in their work than a group of executives in a strategy meeting with permanent access to endless information at the touch of a button via their iPads.

Principle 1: Good Data Visualisation is Trustworthy

The notion of trust is uppermost in your thoughts in this first of the three principles of good visualisation design. This maps directly onto one of Dieter Rams' general principles of good design, namely that good design is honest.

Trust vs Truth

This principle is presented first because it is about the fundamental integrity, accuracy and legitimacy of any data visualisation you produce. This should always exist as your primary concern above all else. There should be no compromise here. Without securing trust the entire purpose of doing the work is undermined.

There is an important distinction to make between trust and truth. Truth is an obligation. You should never create work you know to be misleading in content, nor should you claim something presents the truth if it evidently cannot be supported by what you are presenting. For most people, the difference between a truth and an untruth should be beyond dispute. For those unable or unwilling to be truthful, or who are ignorant of how to differentiate, it is probably worth putting this book away now: my telling you how this is a bad thing is not likely to change your perspective.

If the imperative for being truthful is clear, the potential for there being multiple different but legitimate versions of 'truth' within the same data-driven context muddies things. In data visualisation there is rarely a singular view of the truth. The glass that is half full is also half empty. Both views are truthful, but which to choose? Furthermore, there are many decisions involved in your work whereby several valid options may present themselves. In these cases you are faced with choices without necessarily having the benefit of theoretical influence to draw out the *right* option. You decide what is right. This creates inevitable biases – no matter how seemingly tiny – that ripple through your work. Your eventual solution is potentially comprised of many well-informed, well-intended and legitimate choices – no doubt – but they will reflect a subjective perspective all the same. All projects represent the outcome of an entirely unique pathway of thought.

'Every number we publish is wrong but it is the best number there is.' Andrew Dilnot, Chair of the UK Statistics Authority

You can mitigate the impact of these subjective choices you make, for example, by minimising the amount of assumptions applied to the data you are working with or by judiciously consulting your audience to best

ensure their requirements are met. However, pure objectivity is not possible in visualisation.

Rather than view the unavoidability of these biases as an obstruction, the focus should instead be on ensuring your chosen path is trustworthy. In the absence of an objective truth, you need to be able to demonstrate that *your* truth is trustable.

Trust has to be earned but this is hard to secure and very easy to lose. As the translation of a Dutch proverb states, 'trust arrives on foot and leaves on horseback'. Trust is something you can build by

eliminating any sense that your version of the truth can be legitimately disputed. Yet, visualisers only have so much control and influence in the securing of trust. A visualisation can be truthful but not viewed as trustworthy. You may have done something with the best of intent behind your decision making, but it may ultimately fail to secure trust among your viewers for different reasons. Conversely a visualisation can be trustworthy in the mind of the viewer but not truthful, appearing to merit trust yet utterly flawed in its underlying truth. Neither of these are satisfactory: the latter scenario is a choice we control, the former is a consequence we must strive to overcome.

Let's consider a couple of examples to illustrate this notion of trustworthiness. Firstly, think about the trust you might attach respectively to the graphics presented in Figure 1.8 and Figure 1.9. For the benefit of clarity both are extracted from articles discussing issues about home ownership, so each would be accompanied with additional written analysis at their published location. Both charts are portraying the same data and the same analysis; they even arrive at the same summary finding. How do the design choices make you feel about the integrity of each work?

Both portrayals are truthful but in my view the first visualisation, produced by the UK Office for National Statistics (ONS), commands greater credibility and therefore far more trust than the second visualisation, produced by the *Daily Mail*. The primary reason for this begins with the colour choices. They are relatively low key in the ONS graphic: colourful but subdued, yet conveying a certain assurance. In contrast, the *Daily Mail's* colour palette feels needy, like it is craving my attention

'Good design is honest. It does not make a product appear more innovative, powerful or valuable than it really is. It does not attempt to manipulate the consumer with promises that cannot be kept.'

Dieter Rams, celebrated Industrial Designer

ultimately fail to secure trust among your viewers for different reasons. Conversely a visualisation can be trustworthy in the mind of the viewer but not truthful, appearing to merit trust yet utterly flawed in its underlying truth. Neither of these are satisfactory: the latter scenario is a choice we control, the former is a consequence we must strive to overcome.

Percentage of each age group that are home owners⁵, England, 1981 to 2012

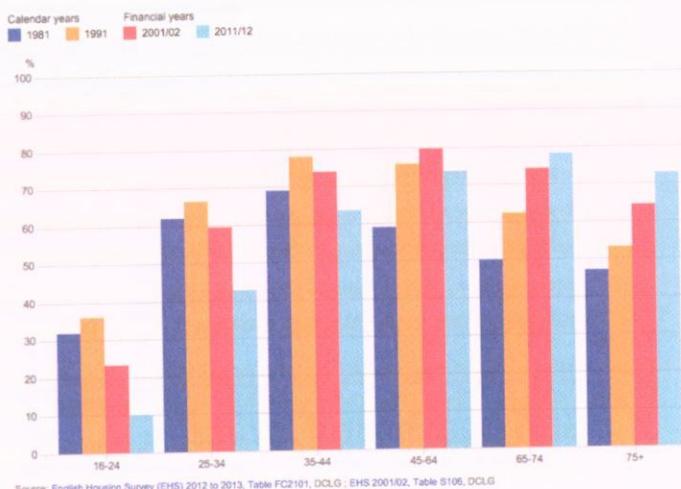


Figure 1.8 Housing and Home Ownership in the UK (ONS)

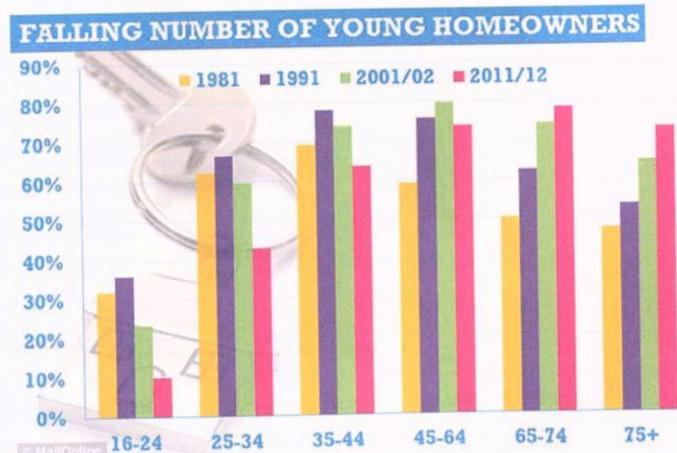


Figure 1.9 Falling Number of Young Homeowners (*Daily Mail*)

with sweetly coloured sticks. I don't care for the house key imagery in the background but it is relatively harmless. Additionally, the typeface, font size and text colour feel more gimpicky in the second graphic. Once again, it feels like it is wanting to shout at me in contrast to the more polite nature of the ONS text. Whereas the *Daily Mail* piece refers to the ONS as the source of the data, it fails to include further details about the data source, which is included on the ONS graphic alongside other important explanatory features such as the subtitle, clarity about the yearly periods and the option to access and download the associated data. The ONS graphic effectively 'shows all its workings' and overall earns, from me at least, significantly more trust.

Another example about the fragility of trust concerns the next graphic, which plots the number of murders committed using firearms in Florida over a period of time. This frames the time around the enactment of the 'Stand your ground' law in the Florida. The area chart in Figure 1.10 shows the number of murders over time and, as you can see, the chart uses an inverted vertical y-axis with the red area going lower down as the number of deaths increases, with peak values at about 1990 and 2007. However, some commentators felt the inversion of the y-axis was deceptive and declared the graphic not trustworthy based on the fact they were perceiving the values as represented by an apparent rising 'white mountain'. They mistakenly observed peak values around 1999 and 2005 based on them seeing these as the highest points. This confusion is caused by an effect known as *figure-ground* perception whereby a background form (white area) can become inadvertently recognised as the foreground form, and vice versa (with the red area seen as the background).

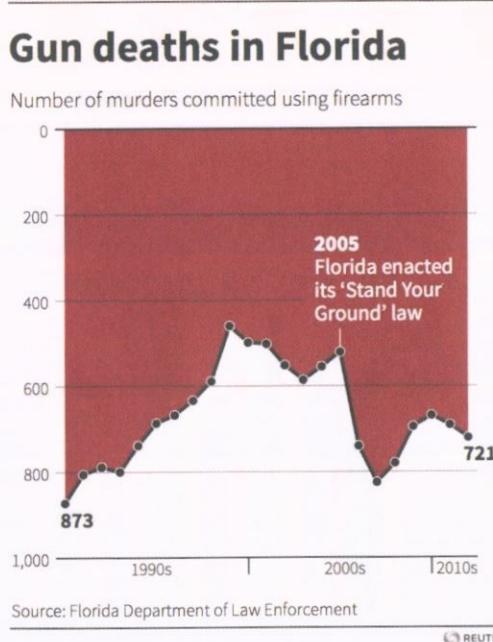


Figure 1.10 Gun Deaths in Florida

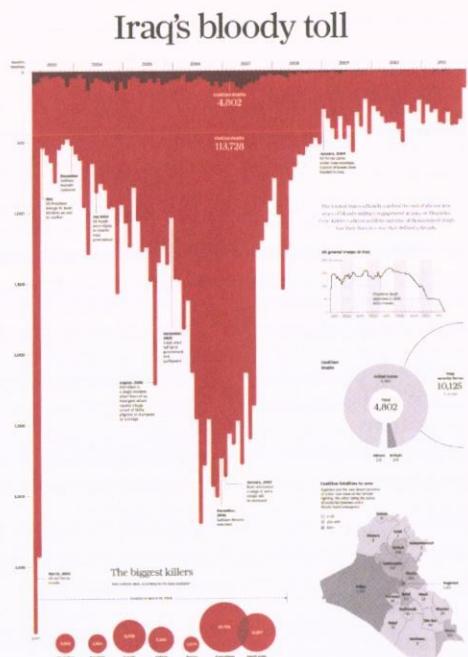


Figure 1.11 Iraq's Bloody Toll

The key point here is that there was no intention to mislead. Although the approach to inverting the y-axis may not be entirely conventional, it was technically legitimate. Creatively speaking, the effect of dribbling blood was an understandably tempting metaphor to pursue. Indeed, the graphic attempts to emulate a notable infographic from several years ago showing the death toll during the Iraq conflict (Figure 1.11). In the case of the Florida graphic, on reflection maybe the data was just too ‘smooth’ to convey the same dribbling effect achieved in the Iraq piece. However, being inspired and influenced by successful techniques demonstrated by others is to be encouraged. It is one way of developing our skills.

Unfortunately, given the emotive nature of the subject matter – gun deaths – this analysis would always attract a passionate reaction regardless of its form. In this case the lack of trust expressed by some was an unintended consequence of a single, innocent design: by reverting the y-axis to an upward direction, as shown in the reworked version in Figure 1.12, you can see how a single subjective design choice can have a huge influence on people’s perception.

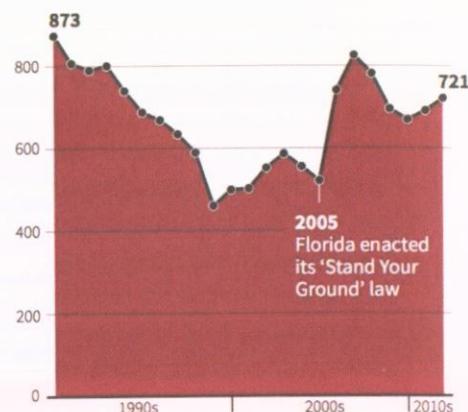
The creator of the Florida chart will have made hundreds of perfectly sound visualisations and will make hundreds more, and none of them will ever carry the intent of being anything other than truthful. However, you can see how vulnerable perceived trust is when disputes about motives can so quickly surface as a result of the design choice made. This is especially the case within the pressured environment of a newsroom where you have only a single opportunity to publish a work to a huge and widespread audience. Contrast this setting with a graphic published within an organisation that can be withdrawn and reissued far more easily.

Trust Applies Throughout the Process

Trustworthiness is a pursuit that should guide all your decisions, not just the design ones. As you will see in the next chapter, the visualisation design workflow involves a process with many decision junctions – many paths down which you could pursue different legitimate options. Obviously, design is the most visible result of your decision making, but you need to create and demonstrate complete integrity in the choices made across the entire workflow process. Here is an overview of some of the key matters where trust must be at the forefront of your concern.

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

Figure 1.12 Reworking of ‘Gun Deaths in Florida’

‘My main goal is to represent information accurately and in proper context. This spans from data reporting and number crunching to designing human-centered, intuitive and clear visualizations. This is my sole approach, although it is always evolving.’ **Kennedy Elliott, Graphics Editor, The Washington Post**

Formulating your brief: As mentioned in the discussion about the ‘Gun Crimes in Florida’ graphic, if you are working with potentially emotive subject matter, this will heighten the importance of demonstrating trust. Rightly or wrongly, your topic will be more exposed to the baggage of prejudicial opinion and trust will be precarious. As you will learn in Chapter 3, part of the thinking involved in ‘formulating your brief’ concerns defining your audience, considering your subject and establishing your early thoughts about the purpose of your work, and what you are hoping to achieve. There will be certain contexts that lend themselves to exploiting the emotive qualities of your subject and/or data but many others that will not. Misjudge these contextual factors, especially the nature of your audience’s needs, and you will jeopardise the trustworthiness of your solution. As I have shown, matters of trust are often outside of your immediate influence: cynicism, prejudice or suspicion held by viewers through their beliefs or opinions is a hard thing to combat or accommodate. In general, people feel comfortable with visualisations that communicate data in a way that fits with their world view. That said, at times, many are open to having their beliefs challenged by data and evidence presented through a visualisation. The platform and location in which your work is published (e.g. website or source location) will also influence trust. Visualisations encountered in already-distrusted media will create obstacles that are hard to overcome.

Working with data: As soon as you begin working with data you have a great responsibility to be faithful to this raw material. To be transparent to your audience you need to consider sharing as much relevant information about how you have handled the data that is being presented to them:

- How was it collected: from where and using what criteria?
- What calculations or modifications have you applied to it? Explain your approach.
- Have you made any significant assumptions or observed any special counting rules that may not be common?
- Have you removed or excluded any data?
- How representative it is? What biases may exist that could distort interpretations?

‘Data and data sets are not objective; they are creations of human design. Hidden biases in both the collection and analysis stages present considerable risks [in terms of inference].’ **Kate Crawford, Principal Researcher at Microsoft Research NYC**

Editorial thinking: Even with the purest of intent, your role as the curator of your data and the creator of its portrayal introduces subjectivity. When you choose to do one thing you are often choosing to not do something else. The choice to focus on analysis that shows how values have changed over time is also a decision to not show the same

data from other viewpoints such as, for example, how it looks on a map. A decision to impose criteria on your analysis, like setting date parameters or minimum value thresholds, in order to reduce clutter, might be sensible and indeed legitimate, but is still a subjective choice.

Data representation: A fundamental tenet of data visualisation is to never deceive the receiver. Avoiding possible misunderstandings, inaccuracies, confusions and distortions is of primary concern. There are many possible features of visualisation design that can lead to varying degrees of deception, whether intended or not. Here are a few to list now, but note that these will be picked up in more detail later:

- The size of geometric areas can sometimes be miscalculated resulting in the quantitative values being disproportionately perceived.
- When data is represented in 3D, on the majority of occasions this represents nothing more than distracting – and distorting – decoration. 3D should only be used when there are legitimately three dimensions of data variables being displayed and the viewer is able to change his or her point of view to navigate to see different 2D perspectives.
- The bar chart value axis should never be ‘truncated’ – the origin value should always be zero – otherwise this approach will distort the bar size judgements.
- The aspect ratio (height vs width) of a line chart’s display is influential as it affects the perceived steepness of connecting lines which are key to reading the trends over time – too narrow and the steepness will be embellished; too wide and the steepness is damped.
- When portraying spatial analysis through a thematic map representation, there are many different mapping projections to choose from as the underlying apparatus for presenting and orienting the geographical position of the data. There are many different approaches to flatten the spherical globe, translating it into a two-dimensional map form. The mathematical treatment applied can alter significantly the perceived size or shape of regions, potentially distorting their perception.
- Sometimes charts are used in a way that is effectively corrupt, like using pie charts for percentages that add up to more, or less, than 100%.

Data presentation: The main rule here is: if it looks significant, it should be, otherwise you are either misleading or creating unnecessary obstacles for your viewer. The undermining of trust can also be caused by what you decline to explain: restricted or non-functioning features of interactivity.

- Absent annotations such as introduction/guides, axis titles and labels, footnotes, data sources that fail to inform the reader of what is going on.
- Inconsistent or inappropriate colour usage, without explanation.
- Confusing or inaccessible layouts.
- Thoroughness in delivering trust extends to the faith you create through reliability and consistency in the functional experience, especially for interactive projects. Does the solution work and, specifically, does it work in the way it promises to do?

Principle 2: Good Data Visualisation is Accessible

This second of the three principles of good visualisation design helps to inform judgments about how best to facilitate your viewers through the process of understanding. It is informed by three of Dieter Rams’ general principles of good design:

- 2 Good design makes a product useful.
- 4 Good design makes a product understandable.
- 5 Good design is unobtrusive.

Reward vs Effort

The opening section of this chapter broke down the stages a viewer goes through when forming their understanding about, and from, a visualisation. This process involved a sequence of perceiving, interpreting and then comprehending. It was emphasised that a visualiser's control over the viewer's pursuit of understanding diminishes after each stage. The objective, as stated by the presented definition, of 'facilitating' understanding reflects the reality of what can be controlled. You can't force viewers to understand, but you can smooth the way.

To facilitate understanding for an audience is about delivering accessibility. That is the essence of this principle: to remove design-related obstacles faced by your viewers when undertaking this process of understanding. Stated another way, a viewer should experience minimum friction between the *act* of understanding (effort) and the *achieving* of understanding (reward).

This 'minimising' of friction has to be framed by context, though. This is key. There are many contextual influences that will determine whether what is judged inaccessible in one situation could be seen as entirely accessible in another. When people are involved, diverse needs exist. As I have already discussed, varying degrees of knowledge emerge and irrational characteristics come to the surface. You can only do so much: do not expect to get all things right in the eyes of every viewer.

'We should pay as much attention to understanding the project's goal in relation to its audience. This involves understanding principles of perception and cognition in addition to other relevant factors, such as culture and education levels, for example. More importantly, it means carefully matching the tasks in the representation to our audience's needs, expectations, expertise, etc. Visualizations are human-centred projects, in that they are not universal and will not be effective for all humans uniformly. As producers of visualizations, whether devised for data exploration or communication of information, we need to take into careful consideration those on the other side of the equation, and who will face the challenges of decoding our representations.' **Isabel Meirelles, Professor, OCAD University (Toronto)**

That is not to say that attempts to accommodate the needs of your audience should just be abandoned, quite the opposite. This is hard but it is essential. Visualisation is about human-centred design, demonstrating empathy for your audiences and putting them at the heart of your decision making.

There are several dimensions of definition that will help you better understand your audiences, including establishing what they know, what they do not know, the circumstances surrounding their consumption of your work and their personal characteristics. Some of these you can accommodate, others you may not be able to, depending on the diversity and practicality of the requirements. Again, in the absence of perfection optimisation is the name of the game, even if this means that sometimes the least worst is best.

The Factors Your Audiences Influence

Many of the factors presented here will occur when you think about your project context, as covered in Chapter 3. For now, it is helpful to introduce some of the factors that specifically relate to this discussion about delivering accessible design.

Subject-matter appeal: This was already made clear in the earlier illustration, but is worth logging again here: the appeal of the subject matter is a fundamental junction right at the beginning of the consumption experience. If your audiences are not interested in the subject – i.e. they are indifferent towards the topic or see no need or relevance to engage with it there and then – then they will not likely stick around. They will probably not be interested in putting in the effort to work through the process of understanding for something that might be ultimately irrelevant. For those to whom the subject matter is immediately appealing, they are significantly more likely to engage with the data visualisation right the way through.

'Data visualization is like family photos. If you don't know the people in the picture, the beauty of the composition won't keep your attention.' Zach Gemignani, CEO/Founder of Juice Analytics

Many of the ideas for this principle emerged from the Seeing Data visualisation literacy research project (seeingdata.org) on which I collaborated.

Dynamic of need: Do they need to engage with this work or is it entirely voluntary? Do they have a direct investment in having access to this information, perhaps as part of their job and they need this information to serve their duties?

Subject-matter knowledge: What might your audiences know and not know about this subject? What is their capacity to learn or potential motivation to develop their knowledge of this subject? A critical component of this issue, blending existing knowledge with the capacity to acquire knowledge, concerns the distinctions between complicated, complex, simple and simplified. This might seem to be more about the semantics of language but is of significant influence in data visualisation – indeed in any form of communication:

- *Complicated* is generally a technical distinction. A subject might be difficult to understand because it involves pre-existing – and probably high-level – knowledge and might be intricate in its detail. The mathematics that underpinned the Moon landings are complicated. Complicated subjects are, of course, surmountable – the knowledge and skill are acquirable – but only achieved through time and effort, hard work and learning (or extraordinary talent), and, usually, with external assistance.
- *Complex* is associated with problems that have no perfect conclusion or maybe even no end state. Parenting is complex; there is no rulebook for how to do it well, no definitive right or wrong, no perfect way of accomplishing it. The elements of parenting might not be necessarily complicated – cutting Emmie's sandwiches into star shapes – but there are lots of different interrelated pressures always influencing and occasionally colliding.
- *Simple*, for the purpose of this book, concerns a matter that is inherently easy to understand. It may be so small in dimension and scope that it is not difficult to grasp, irrespective of prior knowledge and experience.
- *Simplified* involves transforming a problem context from either a complex or complicated initial state to a reduced form, possibly by eliminating certain details or nuances.

Understanding the differences in these terms is vital. When considering your subject matter and the nature of your analysis you will need to assess whether your audience will be

'Strive for clarity, not simplicity. It's easy to "dumb something down," but extremely difficult to provide clarity while maintaining complexity. I hate the word "simplify." In many ways, as a researcher, it is the bane of my existence. I much prefer "explain," "clarify," or "synthesize." If you take the complexity out of a topic, you degrade its existence and malign its importance. Words are not your enemy. Complex thoughts are not your enemy. Confusion is. Don't confuse your audience. Don't talk down to them, don't mislead them, and certainly don't lie to them.' **Amanda Hobbs, Researcher and Visual Content Editor**

reasonably dilute/reduce a complex subject for kids, but generally my advice is don't underestimate the capacity of your audience. Accordingly, clarity trumps simplicity as the most salient concern about data visualisation design.

What do they need to know? The million-dollar question. Often, the most common frustration expressed by viewers is that the visualisation 'didn't show them what they were most interested in'. They wanted to see how something changed over time, not how it looked on a map. If you were them what would you want to know? This is a hard thing to second-guess with any accuracy. We will be discussing it further in Chapter 5.

Unfamiliar representation: In the final chapter of this book I will cover the issue of visualisation literacy, discussing the capabilities that go into being the most rounded creator of visualisation work and the techniques involved in being the most effective consumer also. Many people will perhaps be unaware of a deficit in their visualisation literacy with regard to consuming certain chart types. The bar, line and pie chart are very common and broadly familiar to all. As you will see in Chapter 6, there are many more ways of portraying data visually. This deficit in knowing how to read a new or unfamiliar chart type is not a failing on the part of the viewer, it is simply a result of their lack of prior exposure to these different methods. For visualisers a key challenge lies with situations when the deployment of an uncommon chart may be an entirely reasonable and appropriate choice – indeed perhaps even the 'simplest' chart that could have been used – but it is likely to be unfamiliar to the intended viewers. Even if you support it with plenty of 'how to read' guidance, if a viewer is overwhelmed or simply unwilling to make the effort to learn how to read a different chart type, you have little control in overcoming this.

Time: At the point of consuming a visualisation is the viewer in a pressured situation with a lot at stake? Are viewers likely to be impatient and intolerant of the need to spend time learning how to read a display? Do they need quick insights or is there some capacity for them to take on exploring or reading in more depth? If it is the former, the immediacy of the presented information will therefore be a paramount requirement. If they have more time to work through the process of perceiving, interpreting and comprehending, this could be a more conducive situation to presenting complicated or complex subject matter – maybe even using different, unfamiliar chart types.

immediately able to understand what you are presenting or have the capacity to learn how to understand it. If it is a subject that is inherently complex or complicated, will it need to be simplified? If you are creating a graphic about taxation, will you need to strip it down to the basics or will this process of simplification risk the subject being oversimplified? The final content may be obscured by the absence of important subtleties. Indeed, the audience may have felt sufficiently sophisticated to have had the capacity to work out and work with a complicated topic, but you denied them that opportunity. You might

Format: What format will your viewers need to consume your work? Are they going to need work created for a print output or a digital one? Does this need to be compatible with a small display as on a smartphone or a tablet? If what you create is consumed away from its intended native format, such as viewing a large infographic with small text on a mobile phone, that will likely result in a frustrating experience for the viewer. However, how and where your work is consumed may be beyond your control. You can't mitigate for every eventuality.

Personal tastes: Individual preferences towards certain colours, visual elements and interaction features will often influence (enabling or inhibiting) a viewer's engagement. The semiotic conventions that visualisers draw upon play a part in determining whether viewers are willing to spend time and expend effort looking at a visualisation. Be aware though that accommodating the preferences of one person may not cascade, with similar appeal, to all, and might indeed create a rather negative reaction.

Attitude and emotion: Sometimes we are tired, in a bad mood, feeling lazy, or having a day when we are just irrational. And the prospect of working on even the most intriguing and well-designed project sometimes feels too much. I spend my days looking at visualisations and can sympathise with the narrowing of mental bandwidth when I am tired or have had a bad day. Confidence is an extension of this. Sometimes our audiences may just not feel sufficiently equipped to embark on a visualisation if it is about an unknown subject or might involve pushing them outside their comfort zone in terms of the demands placed on their interpretation and comprehension.

The Factors You Can Influence

Flipping the coin, let's look at the main ways we, as visualisers, can influence (positively or negatively) the accessibility of the designs created. In effect, this entire book is focused on minimising the likelihood that your solution demonstrates any of these negative attributes. Repeating the mantra from earlier, you must avoid doing anything that will cause the boat to go slower.

As you saw listed at the start of this section, the selected, related design principles from Dieter Rams' list collectively include the aim of ensuring our work is useful, unobtrusive and understandable. Thinking about what *not* to do – focusing on the likely causes of failure across these aims – is, in this case, more instructive.

'The key difference I think in producing data visualisation/infographics in the service of journalism versus other contexts (like art) is that there is always an underlying, ultimate goal: to be useful. Not just beautiful or efficient – although something can (and should!) be all of those things. But journalism presents a certain set of constraints. A journalist has to always ask the question: How can I make this more useful? How can what I am creating help someone, teach someone, show someone something new?'
Lena Groeger, Science Journalist, Designer and Developer at ProPublica

Your Solution is Useless

- You have failed to focus on *relevant* content.
- It is not *deep* enough. You might have provided a summary-level/aggregated view of the data when the audience wanted further angles of analysis and greater depth in the details provided.

- A complex subject was *oversimplified*.
- It is not fit for the *setting*. You created work that required too much time to make sense of, when immediate understanding and rapid insights were needed.

Your Solution is Obtrusive

- It is visually *inaccessible*. There is no appreciation of potential impairments like colour blindness and the display includes clumsily ineffective interactive features.
- Its *format* is misjudged. You were supposed to create work fit for a small-sized screen, but the solution created was too fine-detailed and could not be easily read.
- It has too many *functions*. You failed to focus and instead provided too many interactive options when the audience had no desire to put in a lot of effort interrogating and manipulating the display.

Your Solution is not Understandable

- Complex *subject* or complex *analysis*. Not explained clearly enough – assumed domain expertise, such as too many acronyms, abbreviations and technical language.
- Used a complex *chart* type. Not enough explanation of how to read the graphic or failure to consider if the audience would be capable of understanding this particular choice of chart type.
- Absent *annotations*. Insufficient details like scales, units, descriptions, etc.

Principle 3: Good Data Visualisation is Elegant

Elegance in design is the final principle of good visualisation design. This relates closely to the essence of three more of Dieter Rams' general principles of good design:

- 3 Good design is aesthetic.
- 8 Good design is thorough down to the last detail.
- 10 Good design is as little design as possible.

What is Elegant Design?

Elegant design is about seeking to achieve a visual quality that will attract your audience and sustain that sentiment throughout the experience, far beyond just the initial moments of engagement. This is presented as the third principle for good reason. Any choices you make towards achieving 'elegance' must not undermine the accomplishment of trustworthiness and accessibility in your design. Indeed, in pursuing the achievement of the other principles, elegance may have already arrived as a by-product of trustworthy and accessible design thinking. Conversely, the visual 'look and feel' of your work will be the first thing viewers encounter before experiencing the consequences of your other principle-led thinking. It therefore stands that optimising the perceived appeal of your work will have a great impact on your viewers.

The pursuit of elegance is elusive, as is its definition: what gives something an elegant quality? As we know, beauty is in the eye of the beholder, but how do we really recognise elegance when we are confronted by it?

When thinking about what the pursuit of elegance of means, the kind of words that surface in my mind are adjectives like *stylish*, *dignified*, *effortless* and *graceful*. For me, they capture the timelessness of elegance, certainly more so than *fancy*, *cool* or *trendy*, which seem more momentary. Elegance is perhaps appreciated more when it is absent from or not entirely accomplished in a design. If something feels cumbersome, inconsistent and lacking a sense of harmony across its composition and use of colour, it is missing that key ingredient of elegance.

When it feels like style over substance has been at the heart of decision-making, no apparent beauty can outweigh the negatives of an obstructed or absent functional experience. While I'm loathe to dwell on forcing a separation in concern between form and function, as a beginner working through the design stages and considering all your options, functional judgements will generally need to be of primary concern. However, it is imperative that you also find room for appropriate aesthetic expression. In due course your experience will lead you to fuse the two perspectives together more instinctively.

In his book *The Shape of Design*, designer Frank Chimero references a Shaker proverb: 'Do not make something unless it is both necessary and useful; but if it is both, do not hesitate to make it beautiful.' In serving the principles of trustworthy and accessible design, you will have hopefully covered both the *necessary* and *useful*. As Chimero suggests, if we have served the mind, our heart is telling us that now is the time to think about beauty.

How Do You Achieve Elegance in Design?

There are several components of design thinking that I believe directly contribute to achieving an essence of elegance.

Eliminate the arbitrary: As with any creative endeavour or communication activity, editing is perhaps the most influential skill, and indeed attitude. Every single design decision you make – every dot, every pixel – should be justifiable. Nothing that remains in your work should be considered arbitrary. Even if there isn't necessarily a scientific or theoretical basis for your choices, you should still be able to offer reasons for every thing that is included and also excluded. The reasons you can offer for design options being rejected or removed are just as important in evidence of your developing eye for visualisation design.

Often you will find yourself working alone on a data visualisation project and will therefore need to demonstrate the discipline and competence to challenge yourself. Avoid going through

'When working on a problem, I never think about beauty. I think only how to solve the problem. But when I have finished, if the solution is not beautiful, I know it is wrong.' **Richard Buckminster Fuller, celebrated inventor and visionary**

'Complete is when something looks seamless, as if it took little effort to produce.' **Sarah Slobin, Visual Journalist**

"Everything must have a reason"... A principle that I learned as a graphic designer that still applies to data visualisation. In essence, everything needs to be rationalised and have a logic to why it's in the design/visualisation, or it's out.' **Stefanie Posavec, Information Designer**

the motions and don't get complacent. Why present data on a map if there is nothing spatially relevant about the regional patterns? Why include slick interactive features if they really add no value to the experience? It is easy to celebrate the brilliance of your amazing ideas and become consumed by work that you have invested deeply in – both your time and emotional energy. Just don't be stubborn or precious. If something is not working, learn to recognise when to not pursue it any further and then kill it.

Thoroughness: A dedicated visualiser should be prepared to agonise over the smallest details and want to resolve even the smallest pixel-width inaccuracies. The desire to treat your work

with this level of attention demonstrates respect for your audience: you want them to be able to work with quality so pride yourself on precision. Do not neglect checking, do not cut corners, do not avoid the non-sexy duties, and never stop wanting to do better.

Style: This is another hard thing to pin down, especially as the word itself can have different meanings for people, and especially

when it has been somewhat ‘damaged’ by the age-old complaints around something demonstrating style over substance. Developing a style – or signature, as Thomas Clever suggests – is in many ways a manifestation of elegant design. The decisions around colour selection, typography and composition are all matters that influence your style. The development of a style preserves the consistency of your strongest design values, leaving room to respond flexibly to the nuances of each different task you face. It is something that develops in time through the choices you make and the good habits you acquire.

‘I suppose one could say our work has a certain “signature”. “Style” – to me – has a negative connotation of “slapped on” to prettify something without much meaning. We don’t make it our goal to have a recognisable (visual) signature, instead to create work that truly matters and is unique. Pretty much all our projects are bespoke and have a different end result. That is one of the reasons why we are more concerned with working according to values and principles that transcend individual projects and I believe that is what makes our work recognisable.’

Thomas Clever, Co-founder CLEVER°FRANKE, a data driven experiences studio

Many news and media organisations seek to devise their own style guides to help visualisers, graphics editors and developers navigate through the choppy waters of design thinking. This is a conscious attempt to foster *consistency* in approach as well as create *efficiency*. In these industries, the perpetual pressure of tight timescales from the relentless demands of the news cycle means that creating efficiency is of enormous value. By taking away the burden of having always to think from scratch about their choices, the visualisers in such organisations are left with more room to concern

themselves with the fundamental challenge of *what* to show and not just get consumed by *how* to show it. The best styles will stand out as instantly recognisable: there is a reason why you can instantly pick out the work of the *New York Times*, *National Geographic*, Bloomberg, the *Guardian*, the *Washington Post*, the *Financial Times*, Reuters and the *South China Morning Post*.

Decoration should be additive, not negative: The decorative arts are historically considered to be an intersection of that which is useful and beauty, yet the term decoration when applied to

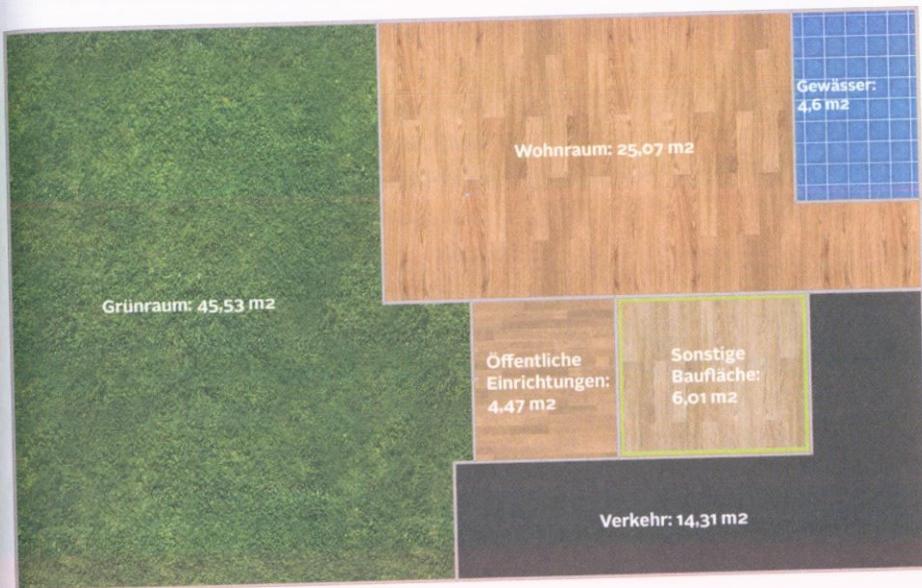


Figure 1.13
If Vienna
Would be an
Apartment

data can often suggest a negative connotation of dressing it up using superfluous devices to attract people, but without any real substance. Visual embellishments are, in moderation and when discernibly deployed, effective devices for securing visual appeal and preserving communicated value. This is especially the case when they carry a certain congruence with the subject matter or key message, such as with the use of the different ground textures in the treemap displayed in Figure 1.13. In this graphic, Vienna is reduced to an illustrative 100m² apartment and the floor plan presents the proportional composition of the different types of space and land in the city. This is acceptable gratuitousness because the design choices are additive, not negatively obstructive or distracting.

Any design choices you make with the aim of enhancing appeal through novelty or fun need to support, not distract from, the core aim of facilitating understanding. Be led by your data and your audience, not your ideas. There should, though, always be room to explore ways of seeking that elusive blend of being fun, engaging and informative. The bar chart in Figure 1.14 reflects this: using Kit Kat-style fingers of chocolate for each bar and a foil wrapper background, it offers an elegant and appealing presentation that is congruent with its subject.

Allow your personality to express itself in the times and places where such flair is supportive of the aims of facilitating understanding. After all, a singularity of style is a dull existence.

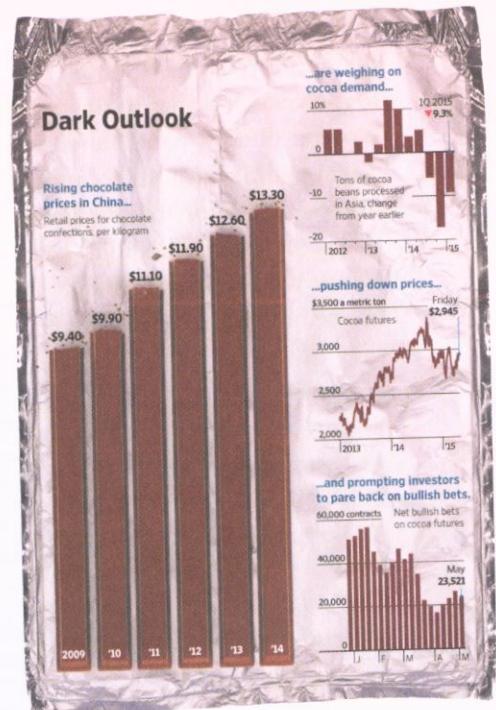


Figure 1.14 Asia Loses Its Sweet Tooth for Chocolate

As Groove Armada once sang: 'If everybody looked the same, we'd get tired of looking at each other.'

'I've come to believe that pure beautiful visual works are somehow relevant in everyday life, because they can become a trigger to get people curious to explore the contents these visuals convey. I like the idea of making people say "oh that's beautiful! I want to know what this is about!" I think that probably (or, at least, lots of people pointed that out to us) being Italians plays its role on this idea of "making things not only functional but beautiful".' **Giorgia Lupi, Co-founder and Design Director at Accurat**

Not about minimalism: As expressed by Rams' principle 'Good design is as little design as possible', elegant design achieves a certain invisibility: as a viewer you should not see design, you should see content. This is not to be confused with the pursuit of minimalism, which is a brutal approach that strips away the arbitrary but then cuts deeper. In the context of visualisation, minimalism can be an unnecessarily savage and austere act that may be incongruous with some of the design options you may need to include in your work.

In 'De architectura', a thesis on architecture written around 15 BC by Marcus Vitruvius Pollio, a Roman architect, the author declares how the essence of quality in architecture is framed by the social relevance of the work, not the eventual form or workmanship towards that form. What he is stating here is that good architecture can only be measured according to the value it brings to the people who use it. In a 1624 translation of the work, Sir Henry Wooton offers a paraphrased version of one of Vitruvius's most enduring notions that a 'well building hath three conditions: firmness, commodity, and delight', of which a further interpretation for today might be read as 'sturdy, useful, and beautiful'. One can easily translate these further to fit with these principles of good visualisation design. Trustworthy is sturdy – it is robust, reliable, and has integrity. Useful is accessible – it can be used without undue obstruction. Beautiful is elegant – it appeals and retains attraction.

1.3 Distinctions and Glossary

As in any text, consistency in the meaning of terms or language used around data visualisation is important to preserve clarity for readers. I began this chapter with a detailed breakdown of a proposed definition for the subject. There are likely to be many other terms that you either are familiar with or have heard being used. Indeed, there are significant overlaps and commonalities of thought between data visualisation and pursuits like, for example, infographic design.

As tools and creative techniques have advanced over the past decade, the traditional boundaries between such fields begin to blur. Consequently, the practical value of preserving dogmatic distinctions reduces accordingly. Ultimately, the visualiser tasked with creating a visual portrayal of data is probably less concerned about whether their creation will be filed under 'data visualisation' or 'infographic' as long as it achieves the aim of helping the audience achieve understanding.

Better people than me attach different labels to different works interchangeably, perhaps reflecting the fact that these dynamic groups of activities are all pursuing similar aims and using the same raw material – data – to achieve them. Across this book you will see plenty of references to and examples of works that might not be considered data visualisation design work in the purest sense. You will certainly see plenty of examples of infographics.

The traditional subject distinctions still deserve to be recognised and respected. People are rightfully proud of identifying with a discipline they have expertise or mastery in. And so, before you step into the design workflow chapters, it is worthwhile to spend a little time establishing clarifications and definitions for some of the related fields and activities so all readers are on the same page of understanding. Additionally, there is a glossary of the terms used that will help you more immediately understand the content of later chapters. It makes sense to position those clarifications in this chapter as well.

Distinctions

Data vis: Just to start with one clarification. While the abbreviated term of data visualisation might be commonly seen as ‘data vis’ (or ‘data viz’; don’t get me started on the ‘z’ issue), and this is probably how all the cool kids on the street and those running out of characters on Twitter refer to it, I am sticking with the full *Sunday* name of ‘data visualisation’ or at the very least the shortened term ‘visualisation’.

Information visualisation: There are many who describe data visualisation as information visualisation and vice versa, myself included, without a great deal of thought for the possible differences. The general distinction, if there is any, tends to be down to one’s emphasis on the input material (data) or the nature of the output form (information). It is also common that information visualisation is used as the term to define work that is primarily concerned with visualising abstract data structures such as trees or graphs (networks) as well as other qualitative data (therefore focusing more on relationships rather than quantities).

Infographics: The classic distinction between infographics and data visualisation concerns the format and the content. Infographics were traditionally created for print consumption, in newspapers or magazines, for example. The best infographics explain things graphically – systems, events, stories – and could reasonably be termed explanation graphics. They contain charts (visualisation elements) but may also include illustrations, photo-imagery, diagrams and text. These days, the art of infographic design continues to be produced in static form, irrespective of how and where they are published.

Over the past few years there has been an explosion in different forms of infographics. From a purist perspective, this new wave of work is generally viewed as being an inferior form of info-graphic design and may be better suited to terms like info-posters or tower graphics (these commonly exist with a fixed-width dimension in order to be embedded into websites and social media platforms). Often these works will be driven by marketing intent through a desire to get hits/viewers, generally with the compromising of any real valuable delivery of understanding. It is important not to dismiss entirely the evident – if superficial – value of this type of work,

as demonstrated by the occasionally incredible numbers for hits received. If your motive is ‘bums on seats’ then this approach will serve you well. However, I would question the legitimacy of attaching the term infographic to these designs and I sense the popular interest in these forms is beginning to wane.

Visual analytics: Some people use this term to relate to analytical-style visualisation work, such as dashboards, that serve the role of operational decision support systems or provide instruments of business intelligence. Additionally, the term visual analytics is often used to describe the analytical reasoning and exploration of data facilitated by interactive tools. This aligns with the pursuit of exploratory data analysis that I will be touching on in Chapter 5.

Data art: Aside from the disputes over the merits of certain infographic work, data art is arguably the other discipline related to visualisation that stirs up the most debate. Those creating data art are often pursuing a different motive to pure data visualisation, but its sheer existence still manages to wind up many who perhaps reside in the more ‘purest’ visualisation camps. For data artists the raw material is still data but their goal is not driven by facilitating the kind of understanding that a data visualisation would offer. Data art is more about pursuing a form of self-expression or aesthetic exhibition using data as the paint and algorithms as the brush. As a viewer, whether you find meaning in displays of data art is entirely down to your personal experience and receptiveness to the open interpretation it invites.

Information design: Information design is a design practice concerned with the presentation of information. It is often associated with the activities of data visualisation, as it shares the underlying motive of facilitating understanding. However, in my view, information design has a much broader application concerned with the design of many different forms of visual communication, such as way-finding devices like hospital building maps or in the design of utility bills.

Data science: As a field, data science is hard to define, so it is easier to consider this through the ingredients of the role of data scientists. They possess a broad repertoire of capabilities covering the gathering, handling and analysing of data. Typically this data is of a large size and complexity and originates from multiple sources. Data scientists will have strong mathematical, statistical and computer science skills, not to mention astute business experience and many notable ‘softer’ skills like problem solving, communication and presentation. If you find somebody with all these skills, tie them to a desk (legally) and never ever let them leave your organisation.

Data journalism: Also known as *data-driven journalism* (DDJ), this concerns the increasingly recognised importance of having numerical, data and computer skills in the journalism field. In a sense it is an adaption of data visualisation but with unquestionably deeper roots in the responsibilities of the reporter/journalist.

Scientific visualisation: This is another form of a term used by many people for different applications. Some give exploratory data analysis the label *scientific visualisation* (drawing out the scientific methods for analysing and reasoning about data). Others relate it to the use of visualisation for conceiving highly complex and multivariate datasets specifically concerning matters with a scientific bent (such as the modelling functions of the brain or molecular structures).

Glossary

The precision and consistency of language in this field can get caught up in a little too much semantic debate at times, but it is important to establish early on some clarity about its usage and intent in this book at least.

Roles and Terminology

Project: For the purpose of this book, you should consider any data visualisation creation activity to be consistent with the idea of a project. Even if what you are working on is only seen as the smallest of visualisation tasks that hardly even registers on the bullet points of a to-do list, you should consider it a project that requires the same rigorous workflow process approach.

Visualiser: This is the role I am assigning to you – the person making the visualisation. It could be more realistic to use a term like researcher, analyst, creator, practitioner, developer, storyteller or, to be a little pretentious, visualist. Designer would be particularly appropriate but I want to broaden the scope of the role beyond just the design thinking to cover all aspects of this discipline.

Viewer: This is the role assigned to the recipient, the person who is viewing and/or using your visualisation product. It offers a broader and better fit than alternatives such as consumer, reader, recipient or customer.

Audience: This concerns the collective group of people to whom you are intending to serve your work. Within the audience there will be cohorts of different viewer types that you might characterise through distinct personas to help your thinking about serving the needs of target viewers.

Consuming: This will be the general act of the viewer, to consume. I will use more active descriptions like ‘reading’ and ‘using’ when consuming becomes too passive and vague, and when distinctions are needed between reading text and using interactive features.

Creating: This will be the act of the visualiser, to create. This term will be mainly used in contrast with *consuming* to separate the focus between the act of the visualiser and the act of the viewer.

Data Terminology

Data is: I’m sorry ‘data are’ fans, but that’s just not how normal people speak. In this book, it’s going to be ‘data is’ all the way. Unless my editor disagrees, in which case you won’t even see this passage.

Raw data: Also known as *primary data*, this is data that has not been subjected to statistical treatment or any other transformation to prepare it for usage. Some people have a problem with the implied ‘rawness’ this term claims, given that data will have already lost its purity having been recorded by some measurement instrument, stored, retrieved and maybe cleaned

already. I understand this view, but am going to use the term regardless because I think most people will understand its intent.

Dataset: A dataset is a collection of data values upon which a visualisation is based. It is useful to think of a dataset as taking the form of a table with rows and columns, usually existing in a spreadsheet or database.

Tabulation: A table of data is based on rows and columns. The rows are the records – instances of things – and the columns are the variables – details about the things. Datasets are visualised in order to ‘see’ the size, patterns and relationships that are otherwise hard to observe. For the purpose of this book, I distinguish between types of datasets that are ‘normalised’ and others that are ‘cross-tabulated’. This distinction will be explained in context during Chapter 5.

Variables: Variables are related items of data held in a dataset that describe a characteristic of those records. It might be the names, dates of birth, genders and salaries of a department of employees. Think of variables as the different columns of values in a table, with the variable name being the descriptive label on the header row. There are different types of variables including, at a general level, quantitative (e.g. salary) and categorical (e.g. gender). A chart plots the relationship between different variables. For example, a bar chart might show the number of staff (with the size of bar showing the quantity) across different departments (one bar for each department or category).

Series: A series of values is essentially a row (or column, depending on table layout) of related values in a table. An example of a series of values would be all the highest temperatures in a city for each month of the year. Plotting this on a chart, like a line chart, would produce a line for that city’s values across the year. Another line could be added to compare temperatures for another city thus presenting a further series of values.

Data source: This is the term used to describe the origin of data or information used to construct the analysis presented. This is an important feature of annotation that can help gain trust from viewers by showing them all they need to know about the source of the data.

Big Data: Big Data is characterised by the 3Vs – high volume (millions of rows of data), high variety (hundreds of different variables/columns) and high velocity (new data that is created rapidly and frequently, every millisecond). A database of bank transactions or an extract from a social media platform would be typical of Big Data. It is necessary to take out some of the hot air spouted about Big Data in its relationship with data visualisation. The ‘Bigness’ (one always feels obliged to include a capitalised B) of data does not fundamentally change the tasks one faces when creating a data visualisation, it just makes it a more significant prospect to work through. It broadens the range of possibilities, it requires stronger and more advanced technology resources, and it amplifies the pressures on time and resources. With more options the discipline of choice becomes of even greater significance.

Visualisation

Chart type: Charts are individual, visual representations of data. There are many ways of representing your data, using different combinations of marks, attributes, layouts and apparatus: these combinations form archetypes of charts, commonly reduced to simply *chart types*. There

are some charts you might already be familiar with, such as the bar chart, pie chart or line chart, while others may be new to you, like the Sankey diagram, treemap or choropleth map.

Graphs, charts, plots, diagrams and maps: Traditionally the term *graph* has been used to describe visualisations that display network relationships and *chart* would be commonly used to label common devices like the bar or pie chart. *Plots* and *diagrams* are more specifically attached to special types of displays but with no pattern of consistency in their usage. All these terms are so interchangeable that useful distinction no longer exists and any energy expended in championing meaningful difference is wasted. For the purpose of this book, I will generally stick to the term *chart* to act as the single label to cover all visualisation forms. In some cases, this umbrella label will incorporate maps for the sake of convenience even though they clearly have a unique visual structure that is quite different from most charts. By the way, the noise you just heard is every cartographer reading this book angrily closing it shut in outrage at the sheer audacity of my lumping maps and charts together.

Graphic: The term *graphic* will be more apt when referring to visuals focused more on information-led explanation diagrams (infographics), whereas *chart* will be more concerned with data-driven visuals.

Storytelling: The term *storytelling* is often attached to various activities around data visualisation and is a contemporary buzzword often spread rather thinly in the relevance of its usage. It is a *thing* but not nearly as much a *thing* as some would have you believe. I will be dampening some of the noise that accompanies this term in the next chapter.

Format: This concerns the difference in output form between printed work, digital work and physical visualisation work.

Function: This concerns the difference in functionality of a visualisation, whether it is static or interactive. Interactive visualisations allow you to manipulate and interrogate a computer-based display of data. The vast majority of interactive visualisations are found on websites but increasingly might also exist within apps on tablets and smartphones. In contrast, a static visualisation displays a single-view, non-interactive display of data, often presented in print but also digitally.

Axes: Many common chart types (such as the bar chart and line chart) have axis lines that provide reference for measuring quantitative values or assigning positions to categorical values. The horizontal axis is known as the x-axis and the vertical axis is known as the y-axis.

Scale: Scales are marks on axes that describe the range of values included in a chart. Scales are presented as intervals (10, 20, 30, etc.) representing units of measurement, such as prices, distances, years or percentages, or in keys that explain the associations between, for example, different sizes of areas or classifications of different colour attributes.

Legend: All charts employ different visual attributes, such as colours, shapes or sizes, to represent values of data. Sometimes, a legend is required to house the 'key' that explains what the different scales or classifications mean.

Outliers: Outliers are points of data that are outside the normal range of values. They are the unusually large or small or simply different values that stand out and generally draw attention from a viewer – either through amazement at their potential meaning or suspicion about their accuracy.

Correlation: This is a measure of the presence and extent of a mutual relationship between two or more variables of data. You would expect to see a correlation between height and weight or age and salary. Devices like scatter plots, in particular, help visually to portray possible correlations between two quantitative values.

Summary: Defining Data Visualisation

In this chapter you have learned a definition of data visualisation: 'The representation and presentation of data to facilitate understanding.' The process of understanding a data visualisation involves three stages, namely:

- Perceiving: what can I see?
- Interpreting: what does it mean?
- Comprehending: what does it mean to me?

You were also introduced to the three principles of good visualisation design:

- Good data visualisation is trustworthy.
- Good data visualisation is accessible.
- Good data visualisation is elegant.

Finally, you were presented with an array of descriptions and explanations about some of the key terms and language used in this field and throughout the book.

What now? Visit book.visualisingdata.com

READING	EXERCISES	CASE STUDY
Visit the chapter's library of further reading and references to continue your learning about the foundations of data visualisation	Undertake these practical exercises to help refine your skill and understanding about the foundations of data visualisation	

2

Visualisation Workflow

Clear, effective and efficient thinking is the critical difference between a visualisation that succeeds and one that fails. You cannot expect just to land accidentally on a great solution. You have got to work for it.

In this chapter I will outline the data visualisation workflow that forms the basis of this book's structure and content. This workflow offers a creative and analytical process that will guide you from an initial trigger that instigates the need for a visualisation through to developing your final solution.

You will learn about the importance of process thinking, breaking down the components of a visualisation design challenge into sequenced, manageable chunks. This chapter will also recommend some practical tips and good habits to ensure the workflow is most effectively adopted.

2.1 The Importance of Process

As I have already established, the emphasis of this book is on better decision making. There are so many different things to think about when creating a data visualisation, regardless of whether the output will be the simplest of charts or the most ambitious of multi-faceted digital implementations.

The decisions you will face will inevitably vary in the weight of their significance. There will be some big choices – matters like defining your editorial angles and selecting the best fit chart type – and many seemingly small ones – such as picking the precise shade of grey for a chart's axis labels. The process of creating a visualisation generally follows the Pareto principle, whereby 20% of decisions made have implications for about 80% of the final visible design. However, just because some decisions will appear more significant in the final output, as visualisers we need to attend to every single decision equally, caring about detail, precision and accuracy.

To repeat, one of the main mental barriers to overcome for those new to the field is to acknowledge that the pursuit of *perfect* in data visualisation is always unfulfilled. There are better and there are worse solutions, but there is no perfect. Perfect exists in an artificial vacuum. It is free