One
Find, design, make, refine

Adam Frost

Our approach to data visualisation

Tobias and I have been running design agencies for the last two decades, most notably between 2011 and 2014, when we were part of a digital agency set up by the *Guardian* newspaper in the UK. Today we head up our own agency, Add Two Digital. For those of you who are unaware of the agency model, it's slightly different from a conventional design team set-up, in that you are not just working for one company or focusing on one subject. Even when we were part of the *Guardian*, our role wasn't usually to make graphics for the newspaper. We were set up to generate revenue for the organisation by offering *Guardian* skills to anyone outside of the business who might need them. These clients could be big corporates, charities, academic institutions or organisations such as the World Trade Organization or the European Union.

Most of the advice in this book will be drawn from this aspect of our careers. Although we both produce our own work, these personal projects tend to be less useful for teaching purposes, just because they are fuelled by our own obsessions and opinions, and we can freely choose what to make, how to make it and how much time to spend on it. Our client work is the opposite: the data differs with every project (simple or complex, qualitative or quantitative), the subject-matter varies (from trade treaties to toilet paper), the output can be static or interactive, it can be a PowerPoint presentation or a social media gif, it can be for experts or for everyone, the deadline can be tomorrow or next year, it can cost a little or a lot. No two projects ever look the same (see Figure 1.1).

However, the way we approach each of these projects *does* look the same. Explaining our working processes for client projects will form the basis of
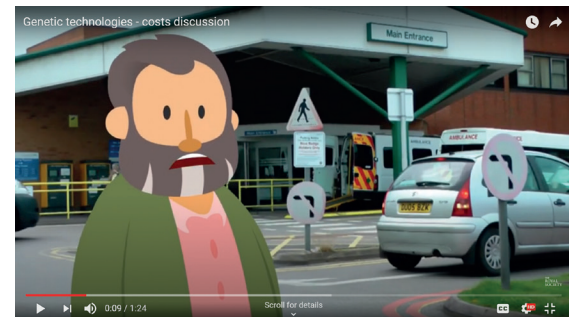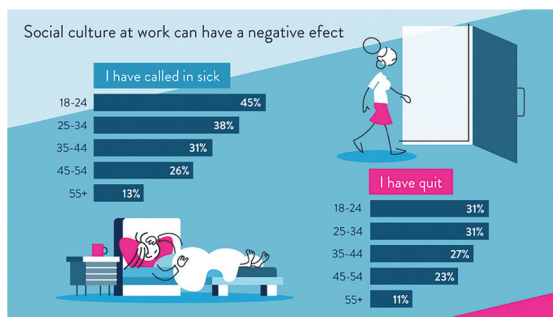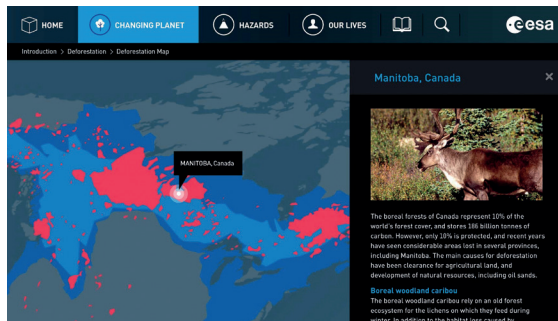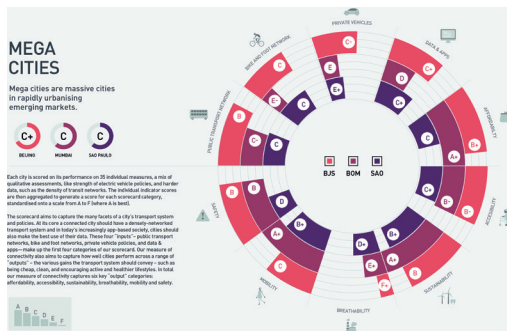

Figure 1.1   A range of client projects

this book because, in our experience, it is this framework that has proved most useful to our students, regardless of job role or previous experience.

Defining data visualisation

It will also serve as our definition of data visualisation (Figure 1.2). There appears to be widespread disagreement about what data visualisation is, whether it overlaps (or not) with infographics, whether it must (or need not) contain interactive elements, whether figurative graphics (illustrations and maps) count, or whether it is something else entirely.[1] For us, any work that adheres to the process we describe – selecting relevant data for an audience and turning it into its most compelling verbal and visual form – can happily be described as data visualisation. If any element is missing – for example, the data is flawed or the charts are unintelligible – then it isn't data visualisation. Because what you're visualising isn't data, or what you're doing to the data isn't visualising it.

One other side point: there is often disagreement about whether data visualisation is 'new' or 'old'. In other words, is it a response to the modern 'big data' and 'open data' revolutions in which unprecedented amounts of information are being collected and therefore new visual methods are required to interrogate and represent this data? Or is it simply a continuation of a discipline that has existed for at least two centuries, since the engineer, spy and blackmailer William Playfair invented the bar chart in 1786? (Or maybe since the Imago Mundi was created in the fifth century BC?) Both points of view have merit. You could compare it to the distinction between cinema and stage plays prior to 1900, or computer games, and puzzles and games invented before the 1970s. A lot of the conventions are the same, and the old and new media both satisfy a similar emotional need in their audience.

Likewise, anyone who makes charts today is performing a similar function to those early data visualisation pioneers. They are clarifying complex information using memorable text and visuals. This is no different to John Snow in the 1850s creating his seminal cholera map, or Otto and Marie Neurath in the 1930s designing the first isotype charts. However, at the same time, there is no question that technology has changed both the chart-making process and the end product in several important ways, enabling anyone with even basic software to make perfectly functional charts and anyone with more technical ability to add interactive elements that can dramatically expand the possibilities of what charts mean, in some cases delivering experiences that resemble the feature films and computer games mentioned above. So yes, data visualisation owes a lot to the past, and this continuity of purpose is important, but at the same time, it is easy to see why a new term has materialised to describe what feels like a rapid evolution in the form's potential.

But let's get back to our process – which, I should add, requires no prior knowledge of (or even interest in) the history of data visualisation. It only requires a willingness to study each part of the process, to understand that they are all of equal importance and to realise that each element adds an additional positive attribute to the final product.

---

[1]Valinda Chan's view is fairly representative: 'Infographics tell a premeditated story to guide the audience to conclusions (subjective). Data visualizations let the audience draw their own conclusions (objective)' (Prototypr.io, 6 June 2017). Kim Rees of the design agency Periscopic believes that an infographic is always static and data visualisation is 'most often' interactive (https://readwrite.com/2011/01/07/difference-between-datavisualization-infographics/).

	THE BIG QUESTION	WHO'S DOING IT?
FIND	What is the story?	Researcher, Analyst, Editor
DESIGN	Who is the story for?	Editor, Copywriter, Producer, Designer
MAKE	How do I put this story together?	Designer, Illustrator, Animator, Developer
REFINE	How do I make the best possible version of this story?	Editor, Designer, QA, Illustrator, Animator, Developer

REMEMBER **1** **Move through each stage one by one.**   **2** **You can move quickly through stages, but never skip them.**

Figure 1.2 Our process

WHAT'S HAPPENING?	WITH WHICH TOOLS?	TO CREATE SOMETHING...?	
Find all relevant datasets. Check the provenance of your data. Find all potential stories in your data.	• Excel • R • Tableau • Google Sheets	Trustworthy	
Find the right story for your audience. Use comparison and context to augment your story. Wireframe your information hierarchy.	• Pen and paper • Google Docs • Photoshop • Pencil Project	Relevant	**BEAUTIFUL**
Choose visuals that help people see the story in the data. Use a consistent visual language. Use design elements to guide people through the story.	• Illustrator • Powerpoint • After Effects • D3	Clear	
Hone your story based on testing and/or feedback. Ensure the copy is as clear and the visuals are as engaging as possible. If necessary, adapt for different channels and use cases.	• Illustrator • Powerpoint • After Effects • D3	Useful	

3 Go back a stage if something isn't working.

4 If you need to go back two or three stages, work through the intervening stages again.

Find

This is where we look for stories in the datasets that we have found.

The first task is always to verify the provenance of the data. Few datasets are perfect of course – most have gaps or inconsistencies – so we often have to use our professional judgement to decide whether these flaws will undermine the validity of our work or sabotage any chance of creating readable charts. If our data passes this test (viability, and visualisation is possible), then we begin to analyse the data we have.

As we work, we try to keep as open a mind as possible. Sometimes we will have a hypothesis that we want to test or a client question which determines the order in which we examine the data, but, even in these cases, we try not to be restricted by these parameters, and to document anything interesting or unusual that we find. A journalist researching a story may have an angle in mind, but it doesn't stop them reading widely on the subject, and taking copious notes that may be of tangential relevance to the final piece. This kind of curiosity is a critical tool when you're analysing a dataset too.

Sometimes we visualise what we find in order to examine its significance. However, it is worth stressing that the visualisation tools we use when we analyse data (e.g. R, Excel, Power BI) and the charts we use (e.g. scatter charts, box plots, histograms) are rarely helpful when it comes to communicating what we have discovered to an audience. When we make charts for analytical purposes, there is usually greater information density, our language (textual and visual) is more technical and we're often using software that won't let us change design and interface elements even if we wanted to. Most importantly, these charts have an audience of one – the person making them. For this reason, they will play a very limited role in this book. For us, data visualisation is always motivated by the desire to make something clear and interesting for someone else. Without this element, it is

like a journalist publishing the jottings in their notebook and calling it journalism. It is a fundamental part of the job but it doesn't have much in common with the finished product.

It tends to be analysts who carry out this work. If it is a simple dataset, then editors and journalists can usually manage it too (there are some examples of simple datasets in Figure 1.3). We use Google Sheets most of the time, although Excel is still useful, and Tableau is indispensable for large or dynamic datasets. Finding a solid dataset with a range of potential stories is the rock on which you build everything else, giving your work depth and integrity.

Design

Although the four stages of this process are all equal, this stage is more equal than the others. Perhaps because of our background at the *Guardian*, we believe that it is this part – designing a story that will resonate with your audience – that is at the heart of any successful data visualisation project. When a chart or graphic misfires, it is usually because the data has been visualised but little else. No thought has been given to removing superfluous data or arranging the data into a hierarchy of importance. There is often a sensible reason for this: designing a visual story is difficult and takes ages. It requires empathy (putting your audience's needs first), ruthlessness (deleting datapoints that you might find interesting but your audience won't) and creative ability (arranging information so that audience interest is sustained). When you finally get to the end, you often find that you need to start again. A quote attributed to Terry Pratchett feels relevant here: 'The first draft is just you telling yourself the story.' In other words, you need patience and persistence too, as you rework the same material over and over again until only the indispensable elements remain.

Number of stolen objects used to make the tunnels in Stalag Luft III in 1944 ("The Great Escape")

Knives	1219
Spoons	478
Forks	582
Pillow cases	161
Mess hall tables	52
Single tables	10
Chairs	34
Benches	76
Bed bolsters	1212
Bed boards	4000
Wall/ floor boards	1370
Double bunk beds	90
Mattresses	635
Bed covers	192
Lamps	69
Water cans	246
Shovels	30
Towels	3424
Blankets	1700
Tin cans	1400

Number of insect fragments permitted in foods (per 100g, US FDA, 2019)

Oregano (ground)	12499
Marjoram (ground)	11749
Marjoram (unground)	2499
Sage (ground)	1999
Nutmeg (ground)	999
Pepper (ground)	949
Cinnamon (ground)	799
Curry powder	399
Paprika (ground)	299
Wheat flour	149
Cocoa powder	149
Corn flour	99
Noodles/ Macaroni	99
Chocolate	59
Broccoli	59
Spinach (canned)	49
Asparagus (canned)	39
Brussels sprouts (frozen)	29
Peanut butter	29

Most common first names of female babies in England and Wales (1924–2014)

	2014	1984	1954	1924
1	Amelia	Sarah	Susan	Margaret
2	Olivia	Laura	Linda	Mary
3	Isla	Gemma	Christine	Joan
4	Emily	Emma	Margaret	Joyce
5	Poppy	Rebecca	Janet	Dorothy
6	Ava	Claire	Patricia	Kathleen
7	Isabella	Victoria	Carol	Doris
8	Jessica	Samantha	Elizabeth	Irene
9	Lily	Rachel	Mary	Elizabeth
10	Sophie	Amy	Anne	Eileen

Sources: *Daily Telegraph*, FDA *Food Defect Levels Handbook*, Office for National Statistics

Figure 1.3 It always starts with a dataset

Of course, some of you might be resistant to the word 'story'. In an exchange on Twitter, Moritz Stefaner warned of the 'power and danger of story-telling', and Scott Murray responded by referring to the 'toxic nature of the "storytelling" fad'.[2] These are two of the most celebrated information designers in the world, so if you'd prefer to think of data visualisation as something other than storytelling, then you are in good company.

However, all we can say is that our work became more focused and more effective when we began to think of ourselves as visual storytellers. Even when we are creating interactive explorers – which is what both Moritz Stefaner and Scott Murray primarily do – we ensure that we are aware of the key stories in our dataset, so we can make intelligent choices about the best charts to use, the default state of each chart, which datasets to exclude, and the interface elements that will guide our audience to the most interesting stories. We cede more control to our users than in a static chart, but this is still tightly circumscribed to ensure that they don't get confused or bored.

Before we thought of ourselves as storytellers, we would often create interactive charts that weren't too different from the charts we made during the **Find** (data analysis) stage of the process. We would argue that 'the user was the storyteller' and the audience could 'discover their own stories'. Which was convenient for us, as it involved a lot less work. And guess what – we discovered that, unless they *really* need the data, users tend to abandon dash-boards that don't meet them halfway.

We found that we achieved better results when we thought of interactive data visualisation as allowing an audience to explore the stories that *we* had found, in an order that made sense to them. Of course it is a mathematical certainty that your audience will find stories that you weren't aware of (quality assurance can only test so many outcomes), but that's true of any storytelling medium. Authors are always pleasantly surprised at the meanings that readers find in their stories.[3] We will have more to say about the relationship between static and interactive storytelling in later chapters.

Although we call this stage 'design', it is not usually led by designers. It tends to be coordinated by journalists, producers or project managers, although sometimes graphic designers and user experience (UX) designers are involved. What is happening here is usually sketching and wireframing, defining the building blocks of our story, and arranging them so that information is released in a logical order. We stay focused on what is most relevant for our audience and the intellectual and emotional impact that we want our story to have on them. In terms of tools, for static graphics, we use pen and paper (as in Figure 1.4); for interactives, we start with pen and paper before moving on to wireframing software. For motion graphics, we start with sketched storyboards, before creating more worked-up versions in Illustrator.

The **Design** stage makes your work relevant and meaningful, creating an emotional bond with your audience.

---

[2]To see the full thread, go to: https://twitter.com/moritz_stefaner/status/550607190378414080

[3]Video game design is another example of this. If you play a sandbox video game such as The Legend of Zelda: Breath of the Wild, you are encouraged to explore and experiment. As any search of gaming videos on YouTube will demonstrate, gamers pride themselves in finding combinations of tools and tricks that the game designers couldn't possibly have imagined. However, there is always a narrative to follow (if you choose). And it is clear that most of the key player choices and pathways have been tested. It's the same when you design interactive data visualisation: users will have less fun if you haven't set up, signposted and tested the main pathways into the content.

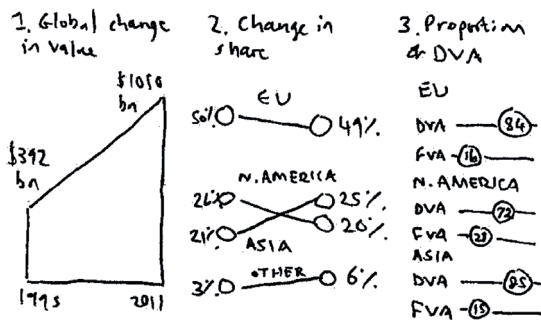20% of the country received food assistance.
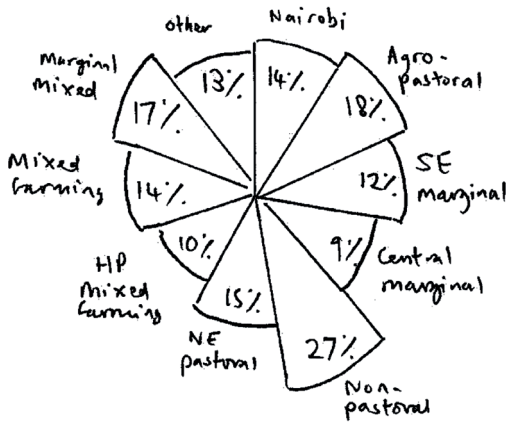
Figure 1.4   Example chart designs

Make

Here we make what we've designed in the previous stage. We don't open any design or animation software or code anything until we've got a clear sense of what we are making. Like most digital agencies, we use agile methodology – which means that we accept change as an inevitable part of any project – so it's not as if our designs are set in stone. We share our work regularly with our client, and we accept that seeing a concept worked up into something more concrete can lead to us going back a stage and reworking aspects of the design.

However, it is best to see the **Make** stage as a separate part of the project as you are almost always

using different software and different team members. Your key focus is now 'Am I telling this story effectively?' rather than 'Is this story worth telling?' (You will have answered the second question already.)

It is in this phase that we agonise over the specifics. For example, let us say that we are creating a presentation and in the **Design** phase we have sketched something like Figure 1.5.

It is here that we will be thinking hard about what that icon at the top should be (see Figure 1.6). An ear of wheat? A farmer with a pitchfork? A tractor? A spade? Or should we just lose the icon? (Probably, yes.)

Should this stay as a column chart? Or should we make it a horizontal bar chart so we can read the labels? And rank it largest to smallest? Add labels to the bars? Leave off the *y*-axis and the gridlines, since we have labels? (See Figure 1.7.)
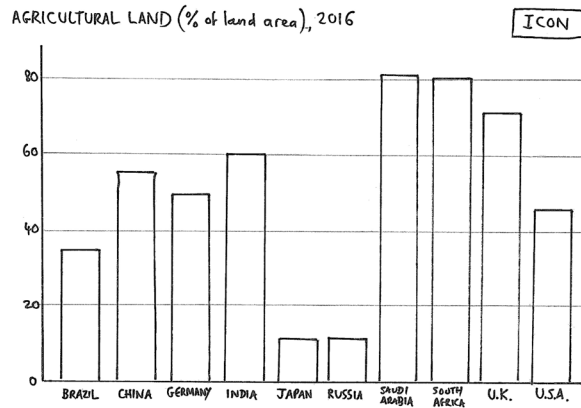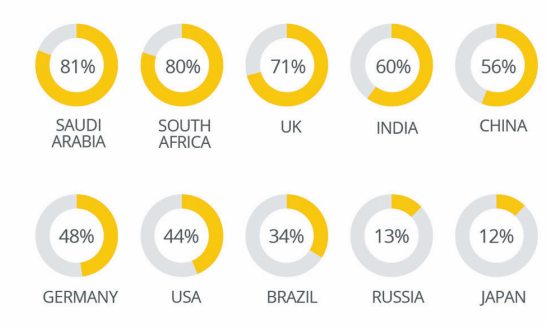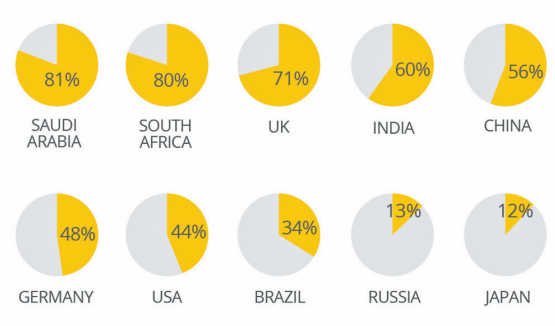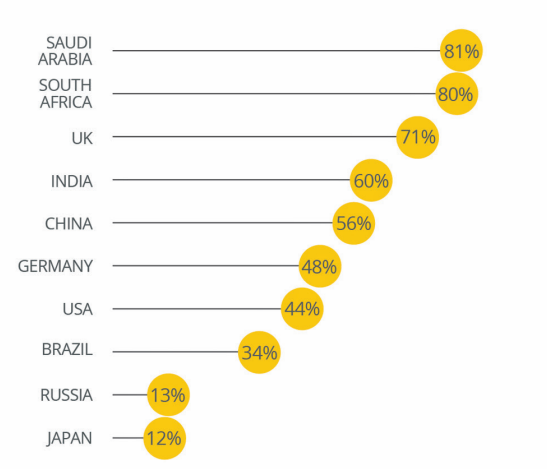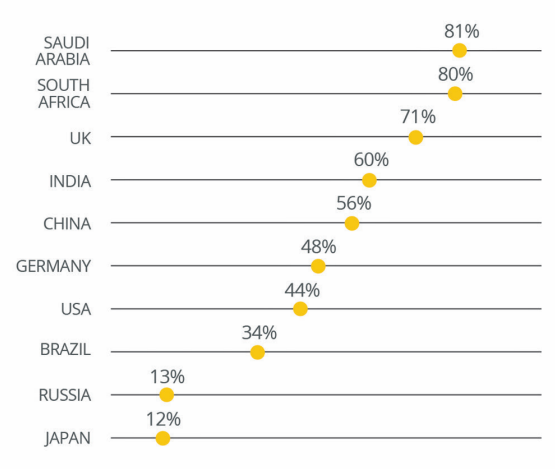

Figure 1.5   Percentage of land in each country used for agriculture, 2016

Figure 1.6   Potential agricultural icons
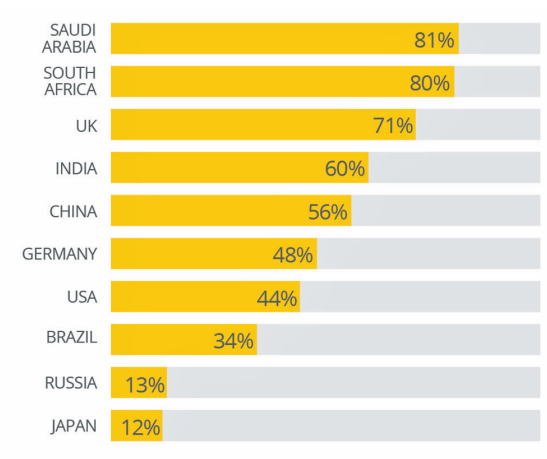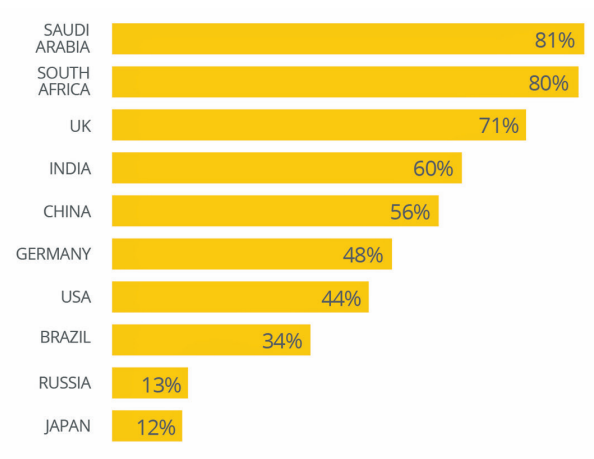
Figure 1.7 Agricultural land (% of land area), 2016, in 12 different charts
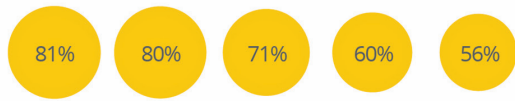
Source: World Bank
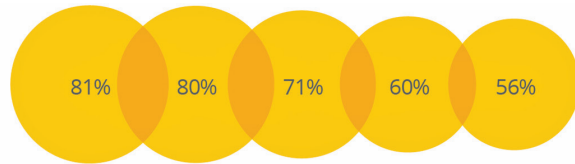
Top-left (circles sized by value):

81%	80%	71%	60%	56%
SAUDI ARABIA	SOUTH AFRICA	UK	INDIA	CHINA

48%	44%	34%	13%	12%
GERMANY	USA	BRAZIL	RUSSIA	JAPAN

Top-right (overlapping circles):

81%	80%	71%	60%	56%
SAUDI ARABIA	SOUTH AFRICA	UK	INDIA	CHINA

48%	44%	34%	13%	12%
GERMANY	USA	BRAZIL	RUSSIA	JAPAN

Middle-left (radial chart):
SAUDI ARABIA 81%, SOUTH AFRICA 80%, UK 71%, INDIA 60%, CHINA 56%, GERMANY 48%, USA 44%, BRAZIL 34%, RUSSIA 13%, JAPAN 12%

Middle-right (radial chart):
UK 71%, GERMANY 48%, RUSSIA 13%, INDIA 60%, JAPAN 12%, CHINA 56%, SAUDI ARABIA 81%, SOUTH AFRICA 80%, BRAZIL 34%, USA 44%

Bottom-left (bars):

81%	80%	71%	60%	56%
SAUDI ARABIA	SOUTH AFRICA	UK	INDIA	CHINA

48%	44%	34%	13%	12%
GERMANY	USA	BRAZIL	RUSSIA	JAPAN

Bottom-right (world map):
71%, 13%, 12%, 44%, 48%, 81%, 56%, 34%, 60%, 80%
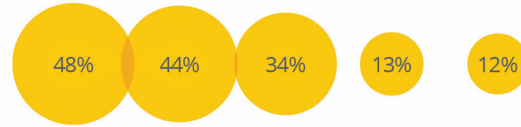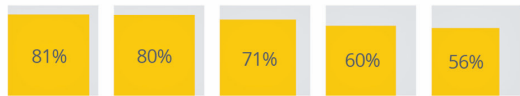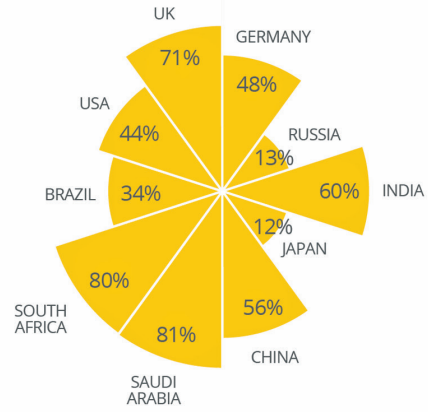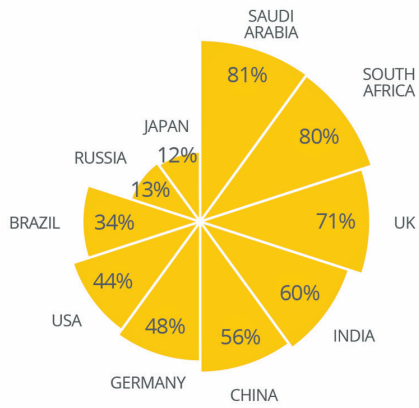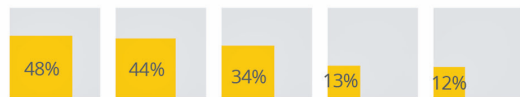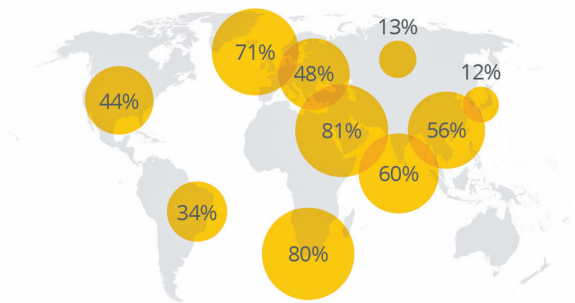
We are talking about percentages, so we could make them 100%-width bars. That gives us a better sense of the proportion of land that is taken up by farming, but it makes the differences less dramatic.

An alternative to this might be a dot plot (effectively, an abacus) which makes the end points stand out, although there is usually no room for the labels. If you want the labels to stay, a variant of a dot plot – sometimes called a lollipop chart – keeps the actual numbers visible.

Since we are charting composition, we could use pie charts. (We'll talk more about pie charts in the next chapter.) They are harder to read than bars, but they catch the eye, like those rows of clocks that show the different times in different cities. If we change the pies into doughnuts, this makes them less legible as charts, but we can add the number in the centre, which aids comprehension.

We could hark back to our *Guardian* days and go for rows of bubbles. (Some designers like making the bubbles overlap to signal the largeness of the largest value – I don't.)

If we wanted to add more visual drama, we could try a polar area chart. Maybe arranged largest to smallest? Or oriented so it matches each country's approximate geographical location?

But this is about agricultural land, so how about using squares to resemble fields seen from above? That might work better as a visual metaphor.

This is global data and people love maps, so we could also experiment with a cartogram.

Throughout this process, we will be thinking about who our audience is (will they tolerate a more 'creative' chart?), what our story is (is ranking the countries helpful or distracting?) and where in our story this chart occurs (is it at the beginning or the end? Have we just had ten bar charts in a row?).

The story also affects any colour choices we make and annotations we add. Here we have assumed our presentation is a general introduction to the subject for a global institution (say, the World Trade Organization), so we haven't highlighted any specific country. But if it were for OPEC and it was about Saudi Arabia's successful programme to 'make the desert bloom', we would be highlighting Saudi Arabia both here and in every chart in our deck.

There are limits to our experimentation, however. You will notice that none of the charts in Figure 1.7 has any 3D elements. In nearly every case, adding 3D to a chart renders it unattractive and confusing. The same applies to drop shadows and gradients: they don't help (see Figure 1.8). Although we are charting different countries, we have resisted the urge to add flags, because the flag colours clash with each other, they distract from the data, and few people know more than a handful of flags so we would have to add labels anyway. Of course, if a client absolutely insists that we include these elements, then in they go, but we would never include any of these design elements without being told to.

It is also worth saying that all of these charting options suit a static presentation, but not an interactive one. If this were a dashboard about land use, particularly one that had to be usable on mobile devices, the number of potential charts would be more limited (probably to horizontal bars, line charts, maps and tables) and we would be worrying less about making the charts immersive and more about making the user journeys *between* charts immersive.

The **Make** stage typically involves teams of designers, illustrators, animators, front-end developers, software developers and quality assurance managers. There is usually an internal project manager, and an account manager who interacts with the client – but for small projects, you can usually merge these roles. The tools used vary hugely: for static graphics, it is usually Inkscape, Illustrator, InDesign or PowerPoint. For interactive charts we use JavaScript libraries such as Highcharts and D3, and for animations we use After Effects. But new tools and charting libraries appear all the time, so it is worth keeping your

Percentage of land in each country used for agriculture
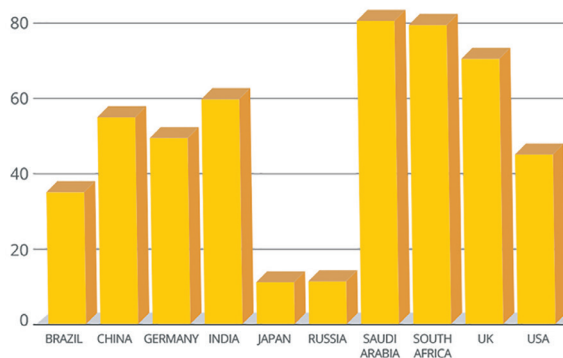
Source: World Bank

Percentage of land in each country used for agriculture

Source: World Bank

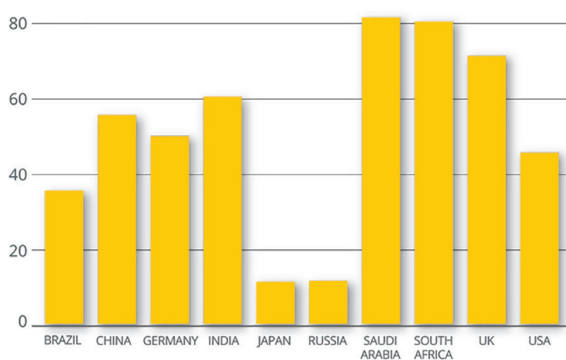## Percentage of land in each country used for agriculture
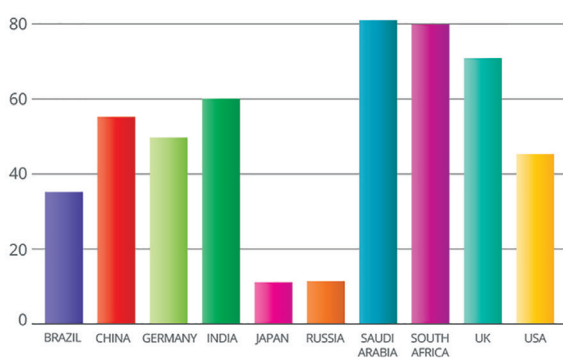
Source: World Bank

Figure 1.8   Experiments to avoid

eye on data viz blogs to see what practitioners are currently recommending.

This part of the process brings clarity and comprehensibility to your work. The objective is to enlighten your audience.

Refine

Last, but not least, you refine your work. This means correcting any spelling mistakes and double-checking the data, testing it for bugs (if it is interactive), sharpening up the visuals, and running it past key readers or users.

Every member of the team is involved in this part of the process. Quality assurance teams test for bugs; developers fix errors; editors and producers correct and rework copy; designers and front-end developers make sure any graphical elements are as engaging as possible, finalising font choices, colour palettes and layout. This last aspect is more important than it has ever been, just because the amount of competition is growing all the time. Thousands of charts and graphics get published every day, and if your work looks distinctive, then people are more likely to stop and listen to what you're telling them. Certainly we have seen many important stories sink without trace because of flat copy or unappealing visuals.

It is also during this stage that we process most of the client feedback. Let joy be unconfined.

Of course clients will have fed back at earlier stages too, but this is where the senior people get involved, with reliably hilarious consequences. 'The Eiffel Tower just doesn't say "Paris" to me', 'Surely you can rework that web page over the weekend?', 'Could you make us an icon for creative synergies?' and the timeless classic 'Make the logo bigger'.

The thing is, even if it's awful, most client feedback is a *good thing*. Your client is usually much closer to the intended audience of your

chart or interactive than you are. Their solutions might be wrong, but they are sometimes putting their finger on a weakness in your work. Even when we are making our own graphics, uncommissioned by anyone, we draft in friends or colleagues to act as proxy clients, because that cycle of feedback and refinement is so useful in creating a piece of work that connects with others, that pre-empts where an audience might need more or less information, and that anticipates where design elements can speed up understanding.

I'm not saying that doing this is easy; it's not. It's painful to throw work away. But unfortunately an audience doesn't care how long it took you, or how hard you worked. Do *you* care how long it took your favourite artist to make that incredible painting or song or novel? They either got it right, or they didn't. Having access to a critical friend – which is what the best clients are – can give you a valuable competitive edge, and is usually a cause for celebration rather than dread.

And I should stress, although I'm using the word 'client' here, I mean anyone who has oversight of your work: it could be someone editing your news story, or an academic peer-reviewing your journal article. They are performing the same role: acting as a test audience. The refinement stage makes your work truly useful. If you don't go through this part of the process, you might have a solid story, but few people will notice it and fewer still will share it. We often compare it to the difference between *The Spanish Tragedy* by Thomas Kyd and Shakespeare's *Hamlet*.

So – have you read Kyd's *Spanish Tragedy*? Do you know any of the characters' names? Are you aware of the basic story? Could you quote any of the lines?

What about *Hamlet*? Do you know the plot of Shakespeare's play? It's likely that you do. Maybe you know a few of the lines too: 'To be or not to be', 'Alas, poor Yorick', 'Get thee to a nunnery', 'To thine own self be true'.

Well, *The Spanish Tragedy* and *Hamlet* are essentially the same play. Similar plot, similar characters.

Shakespeare stole the whole thing, just like he stole the plots of all of his other plays except (possibly) *The Tempest*. But we remember *Hamlet* and we have forgotten Kyd's earlier tragedy because of the way in which Shakespeare refined the story. He changed and rearranged scenes to make the plot more exciting, and gave the characters psychological depth to make their actions more credible. Above all, he added *style*, the poetry that makes his characters' words so memorable.

Refining has another meaning too. As a culinary term, it means breaking things down, getting rid of the lumps. When you're making a graphic, those lumps are *you* – your taste, your interests, your *stuff* getting in between the story and the audience. Of course you have interpreted a dataset and crafted a story for your audience; it wouldn't exist without you. But for an audience to lose themselves in your story, you have to get lost too. This is what happens during the refinement process: you are erasing as many traces of your presence as possible, so the story can be experienced clearly and cleanly, as if your audience were looking through a window.

This is ultimately the purpose of refinement. It is getting to the best possible expression of your story. That's why people will share it, quote it and use it in their own work, because they are unlikely to find a better version.

There is one final reason why refinement is so important. Although you build your story from the ground up – working with the data, building a story, making it look and sound distinctive – your audience will be experiencing it in the other direction. They will often notice the superficial features first – the snappy title and the attractive charts – and only if they find these appealing will they spend longer with your work, getting drawn into the story and finally acquiring a deeper understanding of the data. If you haven't spent time perfecting your work, then your audience may never get past these superficial elements, and

all that work you put into finding, designing and making your story will be wasted.

To put it in architectural terms, your house might be built on solid foundations but if it's a dump, no one is going to come and visit.

That first impression of your work is also linked to one of our most fundamental emotions: approach–avoidance conflict. Should I get closer to this? Or should I move further away? Any information source is going to require effort to understand, and it is impossible to predict in advance whether it will be worth the time you invest in it. But an intriguing appearance can overcome some of this ambivalence, because it gives pleasure in itself, and so offers emotional rewards before any of the information has even been assimilated.

Of course, refining a bad story is unwise. In fact, a story that looks and sounds beautiful will get noticed, which means that experts will start to pay attention, because they are usually very wary of newcomers (especially *designers*) trespassing on their particular area of expertise. It is at this point that any shortcuts or errors you made during the **Find**, **Design** and **Make** stages will be uncovered. Therefore it is a mistake to think that beautification can save a bad story. On the contrary, it will only make any shortcomings in your story more visible.

One example: in 2017, I made a graphic on Jane Austen for the *Guardian*, along with designer Jim Kynvin and illustrator Amy Watt. I read all the books twice and checked the data dozens of times. But I got the first name of one of the minor characters wrong, the sub-editors didn't pick it up, and that is now reflected in the (furious) comments for ever, even though we corrected the mistake within half an hour of the graphic being published. Without the beauty of Jim Kynvin's designs, the *Guardian* probably wouldn't have published it, and if they had, few people would have noticed the mistake. I'm glad these people did notice, but the point is, augmenting and amplifying your story also amplifies its errors.

Why teamwork matters

So that is our process. We do not pretend that it is unique; most agencies will have created a similar diagram of their working practices, if only so that clients don't think they are making it up as they go along (even though everyone is, a bit). However, we do find that it helps us to keep projects on track, provided we have the right team in place.

Which leads us to the next point. You might have noticed that there are a range of different job roles, skillsets and software options included in the table of our process (Figure 1.2). We are often asked by clients to recommend a 'data visualiser' to them, because they want to build up their skills in this area. We tend to reply by saying we don't really know any. There may well be a modern da Vinci who can handle all the different aspects of the data visualisation process: a statistician, data analyst, editor, copywriter, project manager and graphic designer (as a minimum), but also an illustrator, animator, front-end developer, back-end developer, UX designer and quality assurance manager. To be honest, we're happy if we can find people who can combine two of these skillsets: a designer who can work with Excel, or a coder who can also make charts look attractive. But these people don't come along very often. So that is our other key piece of advice: work with other people. It is unusual for good data viz to be produced by individuals working alone. Writing, data analysis, designing, coding – it can take a lifetime to master just one of these disciplines. So finding people to work with – people who complement your own abilities – is important.

I was trying to think of a school subject that data visualisation doesn't draw on, but I think they are all relevant. In a typical week, I will be using skills I learnt in English, maths, history, geography, computer science, art and design, modern foreign languages and more. It is why

there is no single route into data visualisation, no area of expertise shared by all practitioners.

Having said all this, it is still worth learning about each part of the process above. After all, you might turn out to be a modern da Vinci who does excel in multiple fields. (They do seem to be out there: Nathan Yau, Maarten Lambrechts, Nadieh Bremer and others.) However, even if this isn't the case, learning about the key tools and techniques used by other members of the team makes the process of collaboration so much easier. If you are an analyst and you know a little about Adobe Illustrator, it means that when you're putting together a brief for a designer, you'll be speaking a similar language. Likewise, if you're an editor talking to developers, understanding the basics of a JavaScript library like D3 means you'll know more about what it can and can't do. A little learning *isn't* a dangerous thing in data visualisation, it's a necessary precondition for effec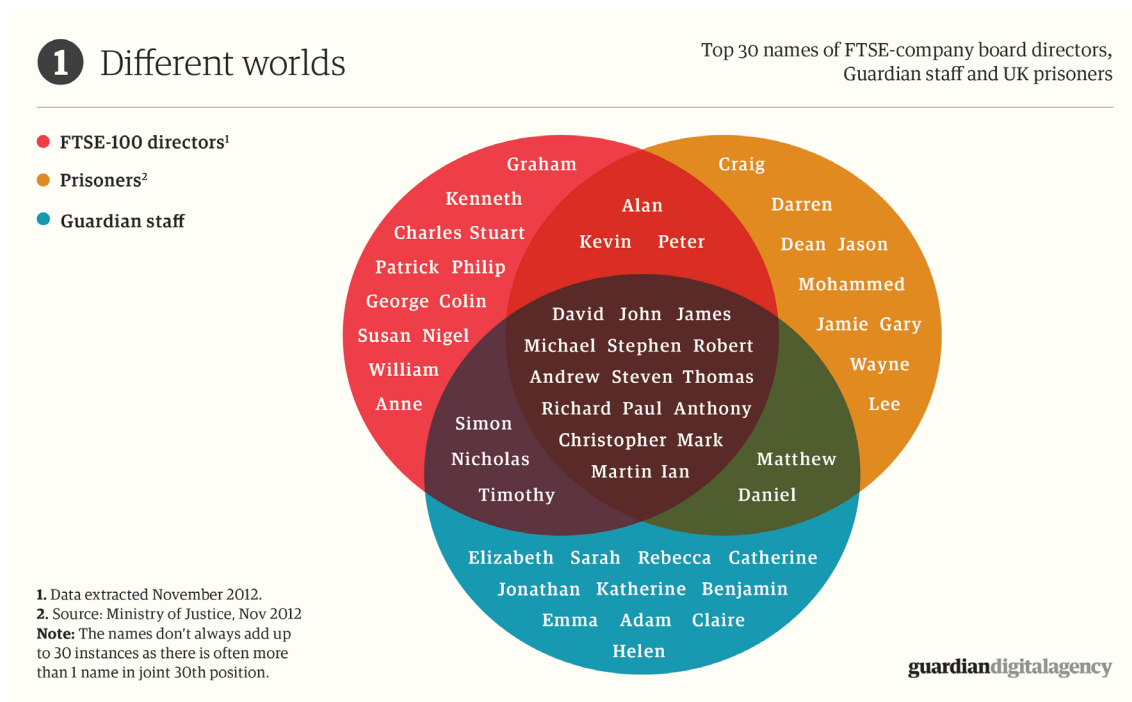tive working relationships. In fact, the technology develops so quickly that a continual willingness to try out new tools, approaches and visual effects ought to come as naturally as breathing.

The process in action

Here is one example of what the whole process looks like in action (Figure 1.9). 'What's in a Name?' is a piece that I created with the designer Zhenia Vasiliev and published on the *Guardian*'s Datablog in 2013. It's about people's names in the UK and how they tend to be a marker of age, ethnicity and social class.

Find

To find this story, I started by asking a range of professional bodies to supply me with the most

1 Different worlds — Top 30 names of FTSE-company board directors, Guardian staff and UK prisoners

- FTSE-100 directors[1]
- Prisoners[2]
- Guardian staff

Graham
Kenneth
Charles Stuart
Patrick Philip
George Colin
Susan Nigel
William
Anne
Alan
Kevin Peter
Craig
Darren
Dean Jason
Mohammed
Jamie Gary
Wayne
Lee

David John James
Michael Stephen Robert
Andrew Steven Thomas
Richard Paul Anthony
Christopher Mark
Martin Ian

Simon
Nicholas
Timothy

Matthew
Daniel

Elizabeth Sarah Rebecca Catherine
Jonathan Katherine Benjamin
Emma Adam Claire
Helen

1. Data extracted November 2012.
2. Source: Ministry of Justice, Nov 2012
Note: The names don't always add up to 30 instances as there is often more than 1 name in joint 30th position.

guardiandigitalagency

Figure 1.9 An example of **Find, Design, Make, Refine**

Figure 1.10 The original spreadsheet for 'What's in a name?'

common first names of practitioners in their field. I supplemented this with publicly available datasets (e.g. the names of company directors). I also contacted government bodies such as the UK

Home Office. After analysing dozens of datasets, we decided to focus on the six groups that contained the most interesting stories – doctors, company directors, footballers, prisoners, Oxford University students and *Guardian* employees (Figure 1.10).

I looked for any interesting patterns and relationships in these datasets (Figure 1.11). I isolated the female names (there were only two female names in the top 30 company directors), and those commonly associated with an ethnic minority group (there was one – Mohammed – in the list of most common prisoners' names) and explored how names fall in and out of fashion (lots of doctors and FTSE company directors called John, but no Johns in the top 30 Oxford undergraduate names).

Figure 1.11 Mapping some of the key stories in the data

Figure 1.12 Names that only appeared in one of the datasets

However, what soon became apparent was that the real story lay in what these lists *didn't* have in common. There were names that only appeared in one of the top 30s but not in any of the others (Figure 1.12). This would become our story.

Design

Next, we needed to come up with a visualisation that showed our audience which names appeared to *impede* you from passing freely from one profession
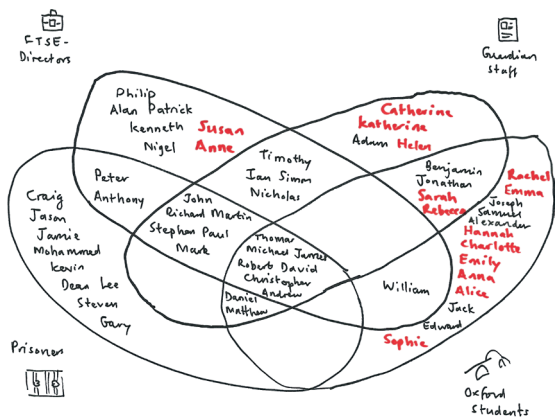
to another. We needed a visual language of inclusion and exclusion, which led us towards the overlapping circles of a Venn diagram. The **Design** stage involved dozens and dozens of sketches. At first, we wondered if we could have more than three circles (Figure 1.13) and include four different professions, but this didn't tell a clear story at all. This is why sketching is a good idea – because your first ideas are usually wildly overambitious.

After we accepted that three overlapping circles would be easiest for our audience to process, we selected the three professions that told the most dramatic story – FTSE directors, prisoners and *Guardian* staff. Then we moved on to the question of how many names to show. We sketched Venns with the top 50 names in each group (adding more evidence for our case, but harder to read), top 10 (too few names undermined our point because the interesting variations started lower down the list) or top 30 (which is where we ended up, Figure 1.14). We couldn't have a hard cut-off at 30 though. There were five FTSE director names in joint 30th place, so we had to have 34 names in that circle.

After we had finished the sketch of the first Venn diagram, and established our design language, we decided to create three more Venns
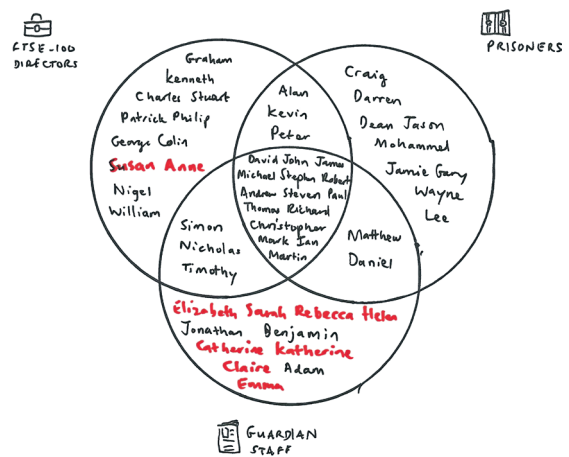

Figure 1.13 No. Just no.

Figure 1.14 The final sketch of 'What's in a name?'

to show interesting relationships between other groups in our dataset – just two circles this time. We sketched FTSE directors v. doctors (to tell the gender story), male Oxford undergraduates v. footballers (focusing on class), and female Oxford undergraduates v. most common female names for that birth cohort (again, a class focus). At the end, to provide wider context about how the meaning of names changes over time, we looked at historical baby names data from the Office for National Statistics. For this, we devised two Venn diagrams, both with three circles, to show how there was some overlap for boys' names between generations (for example, James has been a permanent feature in the top 15), but none at all for female names (Figure 1.15).

Note that we continued to look for contextual and supplementary data even when we had started to design our story. Although it's ideal if you find and analyse all relevant data in the first part of your project, it is frequently the case that you won't understand which dataset will provide the best context until you know for sure the final focus of your story. Each of our four stages – **Find**, **Design**, **Make**, **Refine** – overlap to some degree.

At the end of the **Design** stage, we had a series of five charts: a story with a beginning, middle and end, covering the key stories in our dataset (gender, class, ethnicity and change over time) in a format that the *Guardian*'s Datablog would like (the editor preferred individual charts to single long, scrolling jpegs), legible for readers on both mobile and desktop, and on a subject that we thought *Guardian* readers would find interesting.
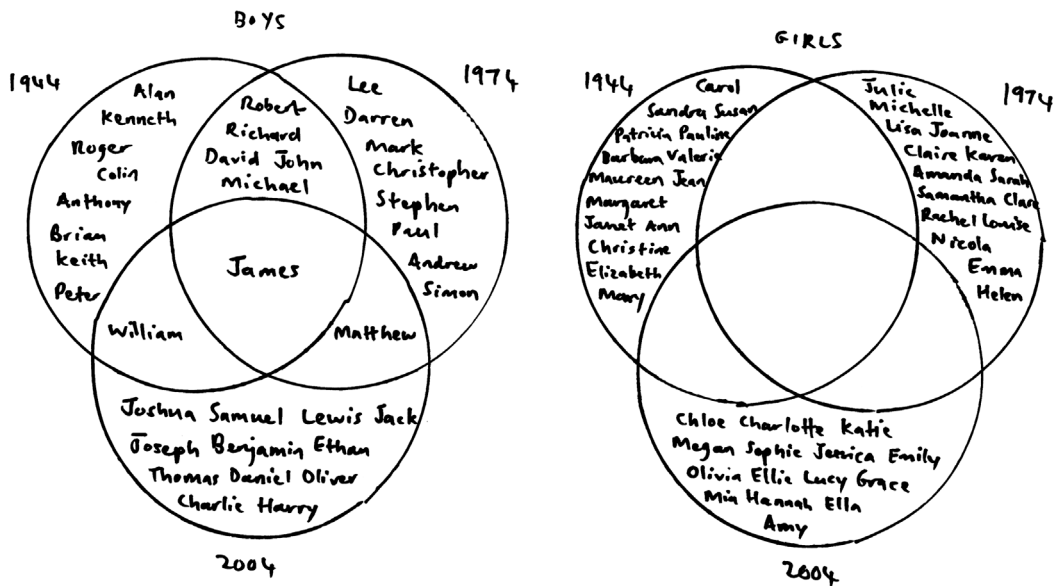
Figure 1.15 Boys' and girls' names – change over time

Make

After I had designed the charts, it was time for Zhenia to make them. I did the first part, transferring my sketches into PowerPoint slides, so Zhenia would have a digital version with all the names correctly spelt and located in the relevant section of the Venn. This is where errors creep in, so the more you can do to ensure consistency across iterations, the better. Suggesting that a designer retypes text from a hand-drawn sketch directly into Illustrator is asking for trouble.

Zhenia started with the first Venn, the hardest one, where we had over 50 names to fit into three circles. His first version (Figure 1.16) looked nothing like the finished version.

He used Arial as a typeface; he chose light shades of primary colours; he also pulled the names in the central overlapping section out of the chart because he was convinced he couldn't fit them in. But doing this meant we lost the central plank of our story: it was no longer possible to clearly see the difference between names that sat in *all* the circles and names that only sat in *one* of them. Believe it or not, this is quite a common issue. When you make a chart, your first attempt can make the wireframe *worse*, because you've started to tease out some of its structural problems without yet alighting on solutions. It's why a lot of PowerPoint presentations are hard going, because the presenter has never moved beyond a first iteration of their chart, and so hasn't ironed out any kinks in the translation from page (or brain) to screen.
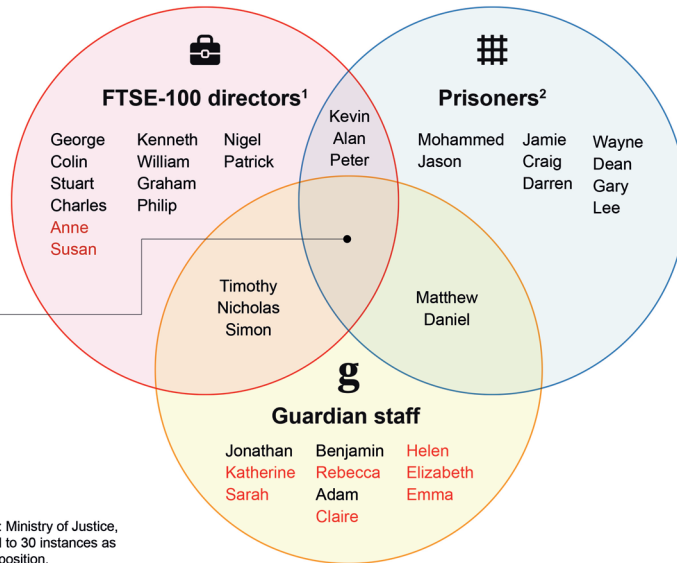
The next iteration was much sharper (Figure 1.17). We lost Arial and used Guardian Egyptian – a less functional font.[4] Zhenia also began to bend the rules

1. Different worlds

Top 30 names of FTSE-company board directors, Guardian staff and UK prisoners

FTSE-100 directors[1]

George	Kenneth	Nigel
Colin	William	Patrick
Stuart	Graham	
Charles	Philip	
Anne		
Susan		

David	Anthony
John	James
Michael	Thomas
Andrew	Steven
Richard	Paul
Christopher	Robert
Martin	Mark
Stephen	Ian

Kevin
Alan
Peter

Prisoners[2]

Mohammed	Jamie	Wayne
Jason	Craig	Dean
	Darren	Gary
		Lee

Timothy
Nicholas
Simon

Matthew
Daniel

g

Guardian staff

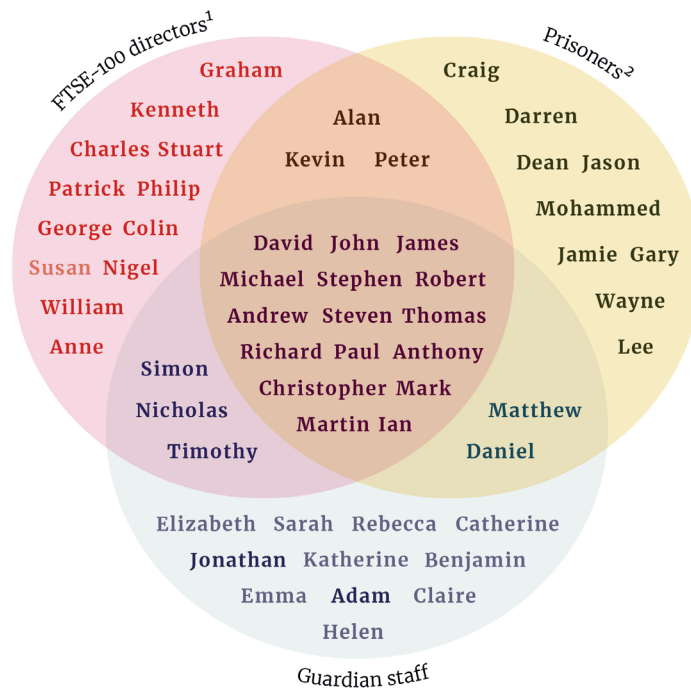|		
Jonathan	Benjamin	Helen
Katherine	Rebecca	Elizabeth
Sarah	Adam	Emma
	Claire	

1. Data extracted November 2012. **2.** Source: Ministry of Justice, Nov 2012 **Note:** The names don't always add to 30 instances as there is often more than 1 name in joint 30th position.

Figure 1.16   The first version of 'What's in a name?'

---

[4]Note that for licensing reasons, we've used Merriweather here. This is a clear and elegant free font that looks a lot like Guardian Egyptian, but you can download it and try it in your own designs for no cost.

1 Different worlds

Top 30 names of
FTSE-company board
directors, Guardian staff
and UK prisoners

FTSE-100 directors[1]

Prisoners[2]

Graham

Craig

Kenneth

Darren

Alan

Charles Stuart

Kevin   Peter

Dean Jason

Patrick Philip

Mohammed

George Colin

David   John   James

Jamie Gary

Susan Nigel

Michael Stephen Robert

Wayne

William

Andrew Steven Thomas

Lee

Anne

Richard Paul Anthony

Simon

Christopher Mark

Nicholas

Martin Ian

Matthew

Timothy

Daniel

Elizabeth   Sarah   Rebecca   Catherine

Jonathan   Katherine   Benjamin

Emma   Adam   Claire

Helen

Guardian staff

1. Data extracted November 2012.
2. Source: Ministry of Justice, Nov 2012
Note: The names don't always add up
to 30 instances as there is often more
than 1 name in joint 30th position.

Figure 1.17   The second version of 'What's in a name?'

a little. This is a vital part of any second iteration,
even if you end up bending the rules back again. As
any information designer will tell you, rules on
charting are only guidelines: what accelerates
information transfer for one audience can impede
it for another.

Here Zhenia tried *squashing* the circles. After all,
this wasn't a *scaled* Venn, where the size of the
circles had a mathematical link to the number of
elements contained within; we were under no obli-
gation to use perfect circles. By distorting the
circles slightly, Zhenia could put all the names back
in the central section. As a side benefit, there was
more horizontal space for the other names, so they
became more legible too.

Zhenia also thought hard about the use of ico-
nography. In my sketches, I'd placed small icons
with each of my diagram labels, probably out of
fear that there wasn't much visual interest other-
wise. In the first iteration Zhenia dutifully kept
these, but in the second the icons are gone. We

both sensed they were now surplus to
requirements. Icons only work if they serve the
story, and by 'serve' I mean that they save time,
save space, increase visual interest or speed up
understanding. Think of how icons are used on,
say, an online video player – Play, Pause, Rewind,
Full Screen, Minimise and so on. They are
sensible choices, simplifying the interface. In the
first iteration of our Venn diagrams, the icons do
not save us time (we have to read the labels to
understand them), they don't save any space
(they cannot replace the labels), they don't
increase interest (they distract from the circles)
and they don't enhance understanding (the
meaning of the labels isn't difficult). So Zhenia
put the icons out of their (and everyone else's)
misery. It's not a bad habit to get into – deleting
icons. If only to see if your chart actually works
better without them.

In the final iteration (Figure 1.18), everything
came together.

❶ Different worlds

- 🔴 **FTSE-100 directors[1]**
- 🟠 **Prisoners[2]**
- 🟢 **Guardian staff**
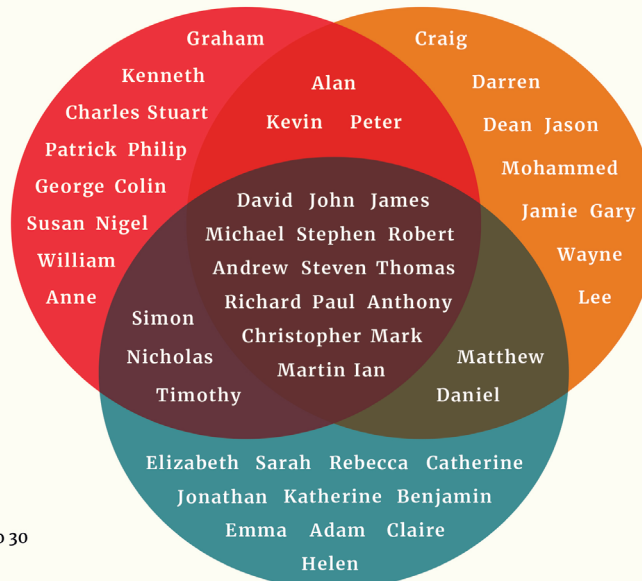
1. Data extracted November 2012.
2. Source: Ministry of Justice, Nov 2012
Note: The names don't always add up to 30 instances as there is often more than 1 name in joint 30th position.

Figure 1.18 The final version of 'What's in a name?'

The big breakthrough here was colour – those beautiful solid discs of red, orange and blue – creating four new colours at the points where the circles overlapped.

It was my fault that this took Zhenia a few iterations to get right. In my sketch and in my PowerPoint slide, I had put female names in a different font colour – as this felt like an important layer of the story. In the first two iterations, Zhenia preserved this – male names were in a different colour (iteration 1) or different shade (iteration 2) than female names. But this aspect of the wireframe was tying Zhenia's hands, preventing him from using bold colours elsewhere. More critically, it was preventing him from supporting our main story. Colour is one of the most powerful weapons we have for signalling groups or categories. Here our main categories are FTSE directors, prisoners and *Guardian* staff. Within these groups, some names are male and some names are female, but this is a secondary grouping. We need to understand those top-level groups first and so that is where our colours should be brightest and boldest. By trying to share colour meanings too widely (as we did in Figures 1.16 and 1.17), *all* the layers of our story end up weaker. By using colour to support our main story *only* (Figure 1.18), every component of our story benefits, because a stronger visual engages people for longer, and they see not just our main story, but secondary stories, and stories that are relevant only to them.

So the font colour in our circles became one colour – white – and our brightly coloured circles became the clear heroes of our graphic.

The other reason I found Zhenia's breakthrough so inspiring is that it brought home the degree of social breakdown so clearly. In a fair society, there should be almost complete overlap: the same names in the centre of those three circles, the same life chances regardless of gender, class, age and ethnicity. Because the contrast in that central section – white font on a dark crimson

background – was stronger than anywhere else in the graphic, this increased the prominence of these names, and made them almost push in front of the others, just as they had pushed past them in life.

The new colours had another beneficial side-effect. Because they were so distinct, we could move the circle labels into a key. Before, these labels had been inside the circles (Figure 1.16), taking up valuable space, or they'd curled round the outside, almost apologetically (Figure 1.17). Now they were out of the way, but still close enough to the chart.

In fact, the information hierarchy improves overall: the title and subtitle are at the top, ruled off from the graphic, the key is top left, the footnotes are bottom left; the relative importance of each element is clearly signalled.

Once Zhenia had set up this design language, he was able to use it in the other four Venn diagrams in the series (Figure 1.19). Note that we 'solved' the first graphic before starting any of the others, as this meant that any edits only had to be done in one place, not five.

Refine

So we have worked through **Find**, **Design** and **Make**. Fortunately for these charts, the **Refine** stage was fairly speedy. The Datablog editor, Simon Rogers, accepted the graphics with minimal edits. The only real changes were reformatting the charts for different use cases: creating crops for social media and for specific image slots on the *Guardian* website. We also prepared a single long jpeg version so that we could publish it on specialist infographic sites and enter it for competitions that stipulated single-image submissions (Figure 1.20).

This is a dull but vital part of the job. When the graphic is finished, it isn't finished – as you are rarely just publishing that image in one place. If you don't create custom versions, then a bored sub-editor or an algorithm will, and people may never click through if what they see is confusing, blurry or doesn't match the supporting text.
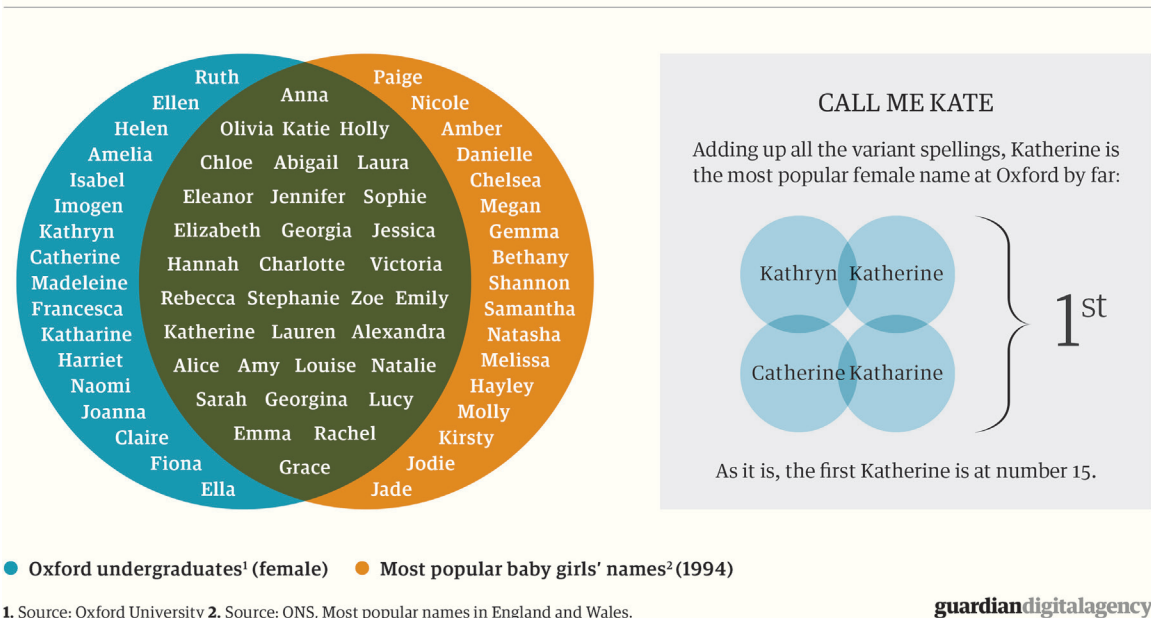

Figure 1.19 Another chart in the 'What's in a name?' series

One increasingly important part of the **Refine** stage is creating animated versions of graphics, particularly for social media. We didn't create one in this instance (it was 2012, life was simpler then) and by the time you read this, perhaps animated gifs will have been replaced by something even more insufferable. But what is unlikely to have changed is the need to think about different audiences discovering your work in different places.

The importance of each stage

Each stage – **Find**, **Design**, **Make**, **Refine** – is essential (Figure 1.21). With our Venn diagrams, if we had rushed or skipped a stage, we would have compromised on accuracy, interest, clarity or usefulness, marring the beauty of the whole. For example, refining a design is often seen as 'adding the finishing touches'. But if you look at our Venn diagram, you'll see that *at no point* before the final iteration (Figure 1.18) was the graphic finished. Even going one step back (Figure 1.17), the colours are washed out and the meaning is less clear. If we had stopped at this point, it is unlikely that the chart would have had the same impact, and all the time and effort we had spent finding and designing the story would have been wasted. Colour, font and layout are not 'finishing touches'; they are your audience's starting point, and an integral part of how your story will be understood.

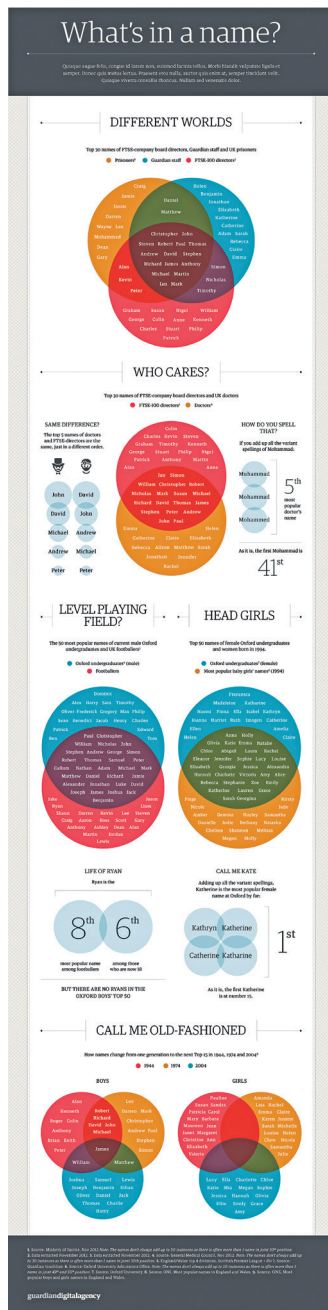Although this was a series of static graphics, we go through the same process whatever we are

Figure 1.20   Cropping and customising your graphic
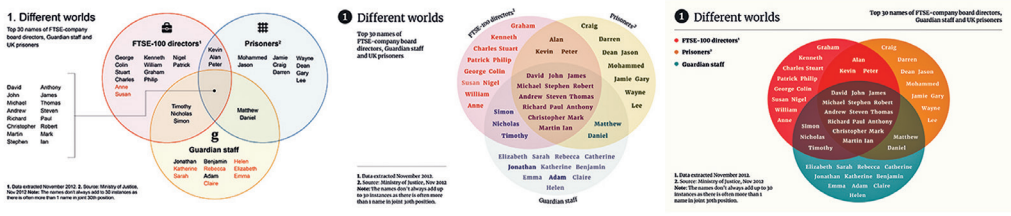
Figure 1.21 **Find**, **Design**, **Make**, **Refine**

making: an animation, a PowerPoint presentation, a dashboard or a physical installation. The media, audience and subject-matter changes, but there is always a story to be found, designed, built and refined.

The only further point to make is that sometimes the story and media type are fixed by our client and this limits the work we can do in the first three stages – **Find**, **Design** and **Make**. For one example of how we approach projects like these, see Appendix A, which you can find online here: www.addtwodigital.com/book

---

Try it yourself!

Whatever your data viz dilemma, you can bet that someone else has already encountered it, and probably come up with a pretty good solution. The more you're aware of data visualisation's past, the easier you'll find it to come up with your own answers. Try the following:

- Monitor the best data viz sites and practitioners and create a (real or virtual) scrapbook of the

(Continued)

- work that you find most inspirational. Some suggested starting points: *The Economist*, the *Financial Times*, the *New York Times*, the *Washington Post* and *National Geographic*.
- Have a look at the recent shortlisted entries for a data viz award, such as the Information is Beautiful Awards or Malofiej. Do you think the judges made the right decision? Would your winner have been different? Have a look at previous winners too.
- Is there a particular subject or theme that you care about, or that is important for your work? Identify the graphics or interactives on this topic that you think are particularly successful. What makes them work so well? Consider writing a blog post about what you've discovered: 'The five most inspirational graphics about crime/health/outer space' (or whatever it might be).