9 Colour

Having established which charts you will use, the potential interactive functions that might be required and the annotation features that will be especially useful, you have effectively determined all the visible elements that will be included in your project. The final two layers of design concern not *what* elements will be included or excluded, but *how* they will appear. After this chapter you will look at issues on composition, but before that the rather weighty matter of colour.

As one of the most powerful sensory cues, colour is a highly influential visual property. It is arguably the design decision that has the most immediate impact on the eye of the viewer. All the design features of your visualisation display hold some attribute of colour, otherwise they are invisible:

- Every mark and item of apparatus in your *charts* will be coloured; indeed colour in itself may be an attribute that represents your data values.
- *Interactive* features do not always have an associated visible property (some are indeed invisible and left as intuitively discoverable). However, those features that involve buttons, menus, navigation tabs and value sliders will always have a colour.
- *Annotation* properties such as titles, captions and value labels will all be coloured.
- *Composition* design mainly concerns the arrangement of all the above features, though you might use colour to help achieve a certain design layout. As you will see, emptiness is a useful organising device – leaving something blank is a colour choice.

Thankfully, there is a route through all of this potential complexity relying on just a little bit of science mixed in with lots of common sense. By replacing any arbitrary judgements that might have been previously based on taste, and through increasing the sensitivity of your choices, colour becomes one of the layers of visualisation design that can be most quickly and significantly improved.

> 'Colors are perhaps the visual property that people most often misuse in visualization without being aware of it.' Robert Kosara, Senior Research Scientist at Tableau Software

The key factor in thinking about colour is to ensure you establish meaning first and decoration last. That is not to rule out the value of certain decorative benefits of colour, but to advise that these should be your last concern. Besides, in dealing with meaningful applications of colour you will already have gone a long way towards establishing the 'decorative' qualities of your project's aesthetic appearance.

This chapter begins with a look at some of the key components of colour science, offering a foundation for your understanding about this topic. After that you will learn about the ways and places in which colour could be used. Finally, you will consider the main factors that influence colour decisions.

COLOUR thinking begins from inside the chart(s), working outwards across the rest of the visualisation anatomy:

- Data legibility.
- Editorial salience.
- Functional harmony.

9.1 Overview of Colour Theory

COLOUR in visualisation is something of a minefield. As with many of these design layer chapters, an introduction to colour involves judging the right amount of science and the right amount of practical application. What does justice to the essence of the subject and gives you the most relevant content to work with is a delicate balance.

When you lift the lid on the science behind colour you open up a world of brain-ache. When this chapter is finalised I will have spent a great deal of time agonising over how to explain this subject and what to leave in or leave out because there is so much going on with colour. And it is tricky. Why? Because you almost come face to face with philosophical questions like 'what is white?' and the sort of mathematical formulae that you really rather hoped had been left behind at school. You learn how the colours you specify in your designs as X might be perceived by some people as Y and others as Z. You discover that you are not just selecting colours from a neat linear palette but rather from a multi-dimensioned colour space occupying a cubic, cylindrical or spherical conceptual shape, depending on different definitions.

The basis of this topic is the science of optics – the branch of physics concerned with the behaviour and properties of light – as well as colorimetry – the science and technology used to quantify and describe human colour perception. Two sciences, lots of maths, loads of variables, endless potential for optical illusions and impairment: that is why colour is tricky and why you need to begin this stage of thinking with an appreciation of some colour theory.

The most relevant starting point is to recognise that when dealing with issues of colour in data visualisation you will almost always be creating work on some kind of computer. Unless you are creating something by hand using paints or colouring pencils, you will be using software viewed through an electronic display.

This is important because a discussion about colour theory needs to be framed around the RGB (Red, Blue, Green) colour model. This is used to define the combination of light that forms the colours you see on a screen, conceptually laid out in a cubic space based on variations across these three attributes.

The output format of your work will vary between screen display and print display. If you are creating something for print you will have to shift your colour output settings to CMYK (Cyan, Magenta, Yellow and Black). This is the model used to define the proportions of inks that make up a printed colour. This is known as a subtractive model, which means that combining all four inks produces black, whereas RGB is additive as the three screen colours combine to produce white.

> When you are creating work to be consumed on the Web through screen displays, you will often program using HEX (Hexadecimal) codes to specify the mix of red, green and blue light (in the form #RRGGBB using codes 00 to FF).

While CMYK communicates from your software to a printer, telling it what colours to print as an output, it does not really offer a logical model to think about the input decisions you will make about colour. Neither, for that matter, does RGB: it just is not realistic to think in those terms when considering what choices are needed in a visualisation design. There are different levers to adjust and different effects being sought that require an alternative model of thinking.

Figure 9.1 HSL Colour Cylinder

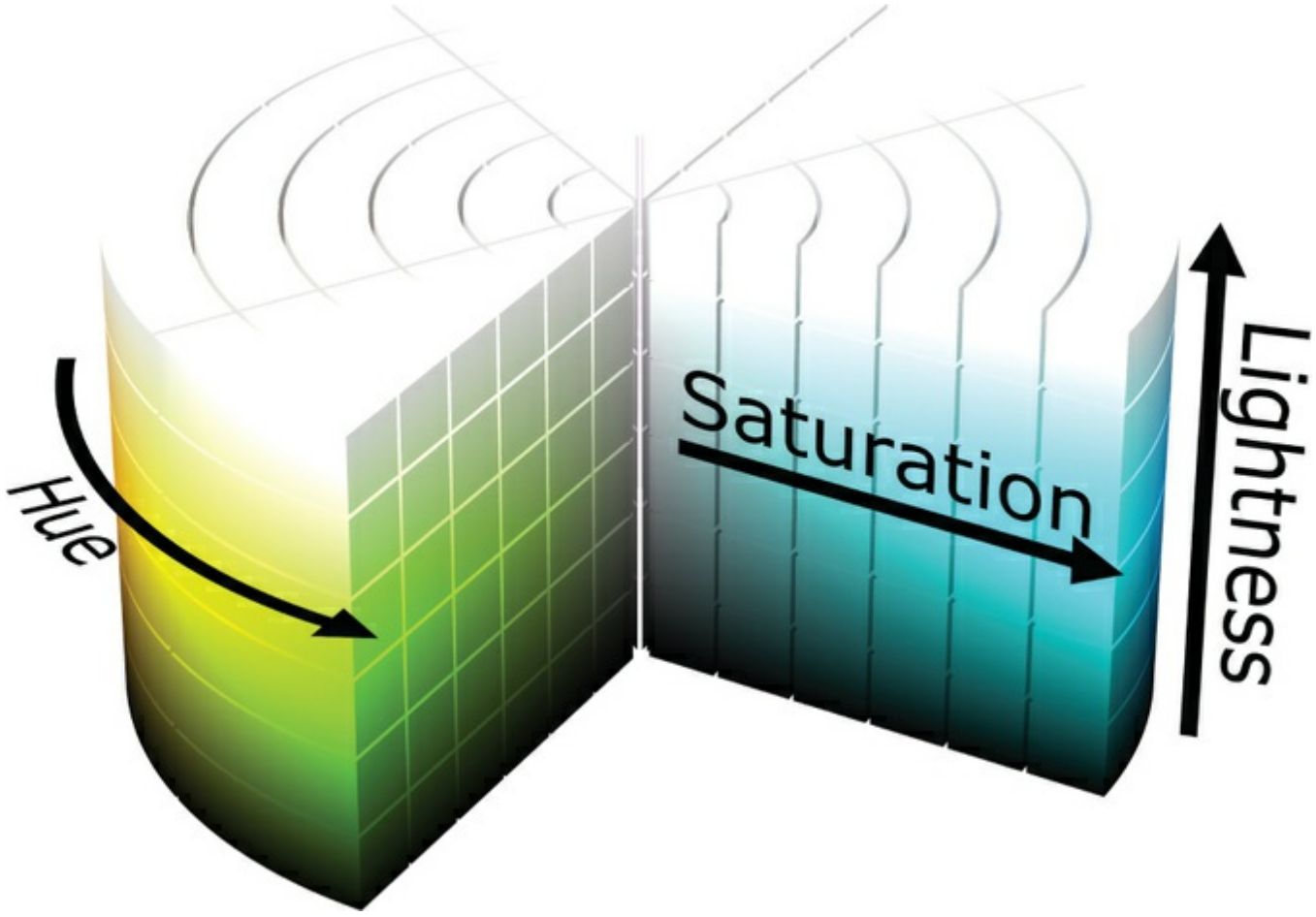


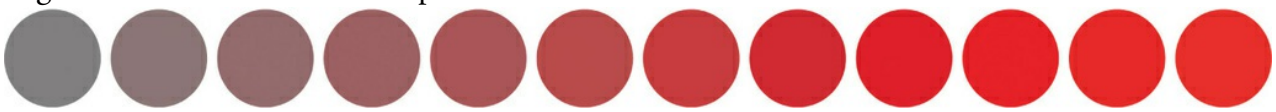
Figure 9.2 Colour Hue Spectrum

I share the belief with many in the field that the most accessible colour model – in terms of considering the application of colour in data visualisation – is HSL (Hue, Saturation, Lightness), devised by Albert Munsell in the 1980s. These three dimensions combine to make up what is known as a cylindrical-coordinate colour representation of the RGB colour model (I did warn you about the cylinders).

Hue is considered the *true* colour. With hue there are no shades (adding black), tints (adding whites) or tones (adding grey) – a consideration of these attributes follows next. When you are describing or labelling colours you are most commonly referring to their hue: think of the colours of the rainbow ranging through various mixtures of red, orange, yellow, green, blue, indigo and violet. Hue is considered a qualitative colour attribute because it is defined by difference and not by scale.

Saturation defines the purity or colourfulness of a hue. This does have a scale from intense pure colour (high saturation) through increasing tones (adding grey) to the no-colour state of grey (low saturation). In language terms think *vivid* through to *muted*.

Figure 9.3 Colour Saturation Spectrum



Lightness defines the contrast of a single hue from dark to light. It is not a measure of brightness – there are other models that define that – rather a scale of light tints (adding white) through to dark shades (adding black). In language terms I actually think of lightness more as degrees of darkness, but

that is just a personal mindset.

Figure 9.4 Colour Lightness Spectrum

Technically speaking, black, white and grey are not considered colours.

I have deliberately described these dimensions separately because, as you will see when looking at the applications of colour in visualisation, your decisions will often be defined by how you might employ these distinct dimensions of colour to form your visual display. The main choices tend to fall between employing difference in hue and variation in lightness, with the different levels of saturation often being a by-product of the definitions made for the other two dimensions.

Alternative models exist offering variations on a similar theme, such as HSV (Hue, Saturation, Value), HSI (Hue, Saturation, Intensity), HSB (Hue, Saturation, Brightness) and HCL (Hue, Chroma, Luminance).

These are all primarily representations of the RGB model space but involve differences in the mathematical translation into/from RGB and offer subtle differences in the meaning of the same terms (local definitions of hue and saturation vary). The biggest difference relates to their emphasis as a means of specifying either a colour quality (in an input, created sense) or a colour perception (in how a colour is ultimately experienced).

Pantone is another colour space that you might recognise. It offers a proprietary colour-matching, identifying and communicating service for print, essentially giving 'names' to colours based on the CMYK process.

The argument against using the HSL model for defining colour is that, while it is fine for colour setting (i.e. an intuitive way to think about and specify the colours you want to set in your visualisation work), the resulting colours will not be uniformly perceived the same, from one device to the next. This is because there are many variables at play in the projection of light to display colour and the light conditions present in the moment of perception. That means the same perceptual experience will not be guaranteed. It is argued that more rigorous models (such as CIELAB) offer an absolute (as opposed to a relative) definition of colour for both input and output. My view is that they are just a little bit too hard to easily translate into visualisation design thinking. Furthermore, trying to control for all the subtleties of variation in consumption conditions is an extra burden you should ideally avoid.

At this stage, it is important to be pragmatic about colour as much as possible. The vast majority of your colour manipulating and perceptual needs should be nicely covered by the HSL model. As and when you develop a deeper, purist interest in colour you should then seek to learn more about the nuances in the differences between the definitions of these models and their application.

9.2 Features of Colour: Data Legibility

Data legibility concerns the use of the attribute of colour to encode data values in charts. The objective here is to make the data being represented by differences in colour as clearly readable and as meaningful as possible.

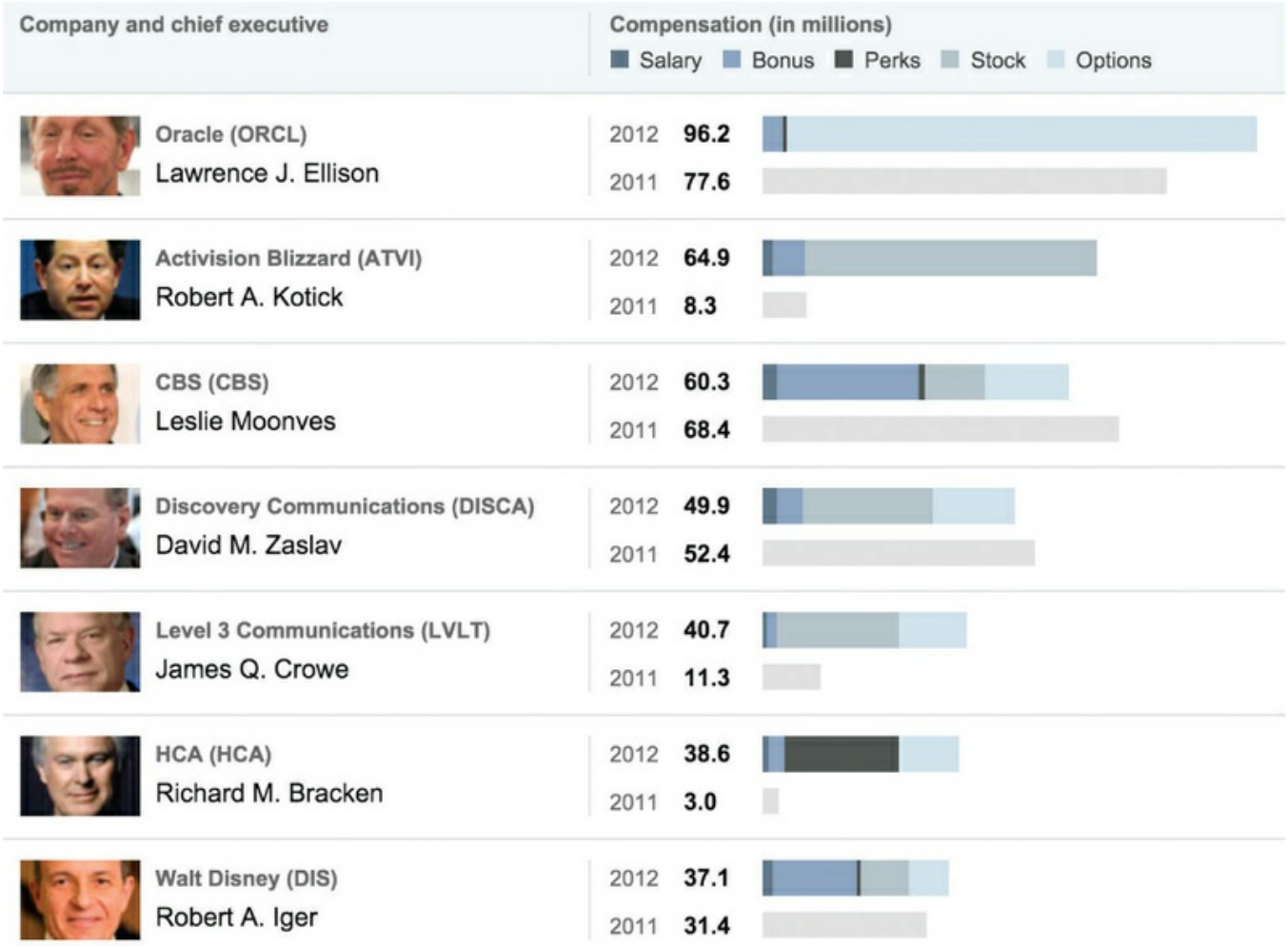
While you have probably already decided by now the chart or charts you intend to use, you still need to take think carefully – and separately – about how you will specifically employ colour. To do this we first need to revisit the classification of data types and consider how best to use colour for representing each different type.

Nominal (Qualitative)

With nominal data colour is used to classify different categorical values. The primary motive for the choice of colour is to create a visible distinction between each unique categorical association, helping the eye to discern the different categories as efficiently and accurately as possible.

Creating contrast is the main aim of representing nominal data. What you are *not* seeking to show or even imply is any sense of an order of magnitude. You want to help differentiate one category from the next – and make it easily identifiable – but to do so in a way that preserves the sense of equity among the colours deployed.

Figure 9.5 Excerpt from 'Executive Pay by the Numbers'



Variation in hue is typically the colour dimension to consider using for differentiating categories. Additionally, you might explore different tones (variations in saturation across the hues). You should not, though, consider using variations in the lightness dimension. That is because the result is insufficiently discernible. As you can see demonstrated in Figure 9.5, the lightness variation of a blue tone makes it quite hard to connect the colour scale presented in the key at the top with the colours displayed in the stacked bars underneath. With the shading in the column header and the 2011 grey bar also contributing similar tones to the overall aesthetic of the table our visual processing system has to work much harder to determine the associations than it should

need to do.

Often the categories you will be differentiating with colour will be relatively few in number, maybe two or three, such as in the separation between political parties or plotting different values for gender, as seen in Figure 9.6.

Figure 9.6 How Nations Fare in PhDs by Sex

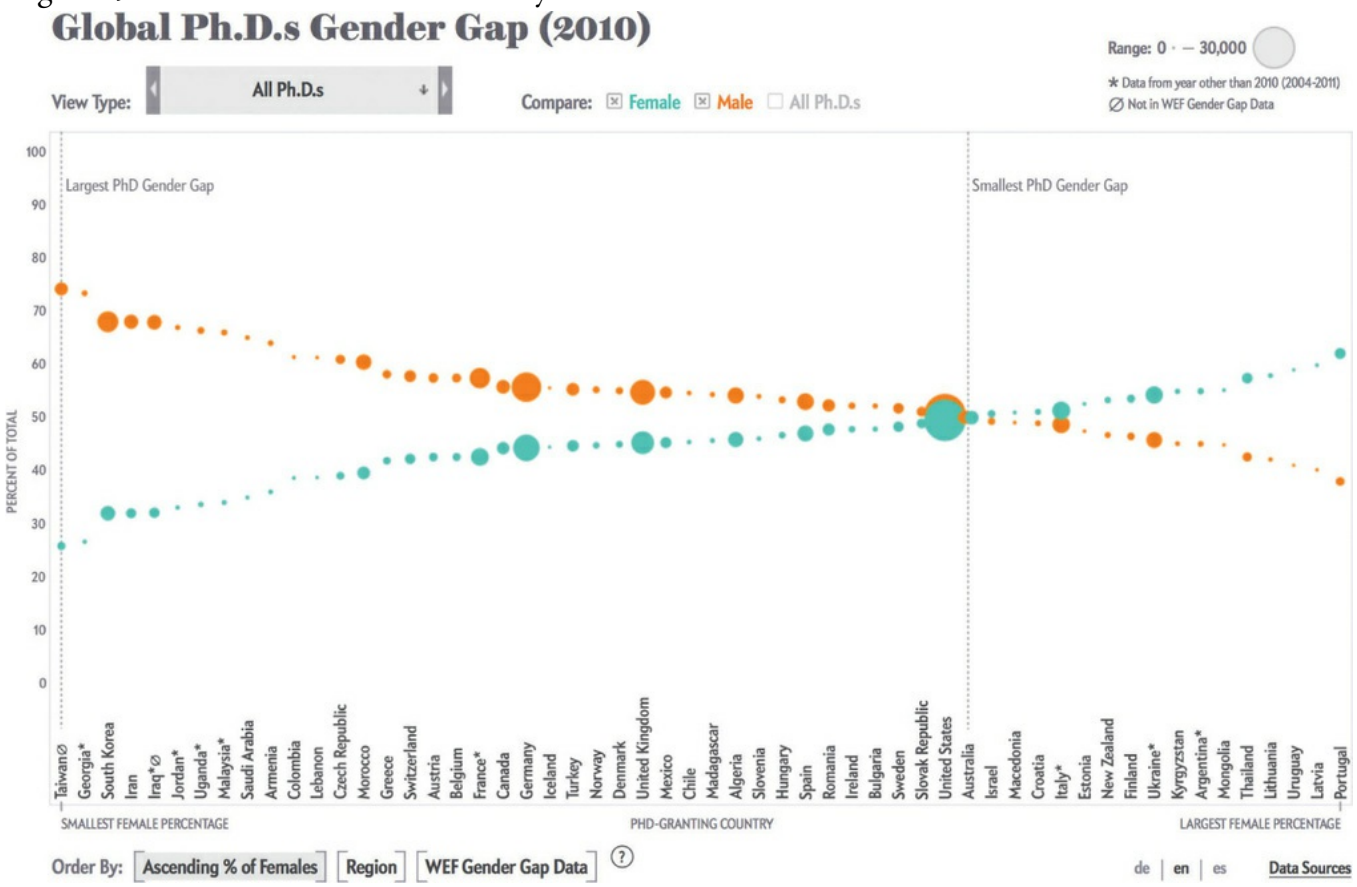
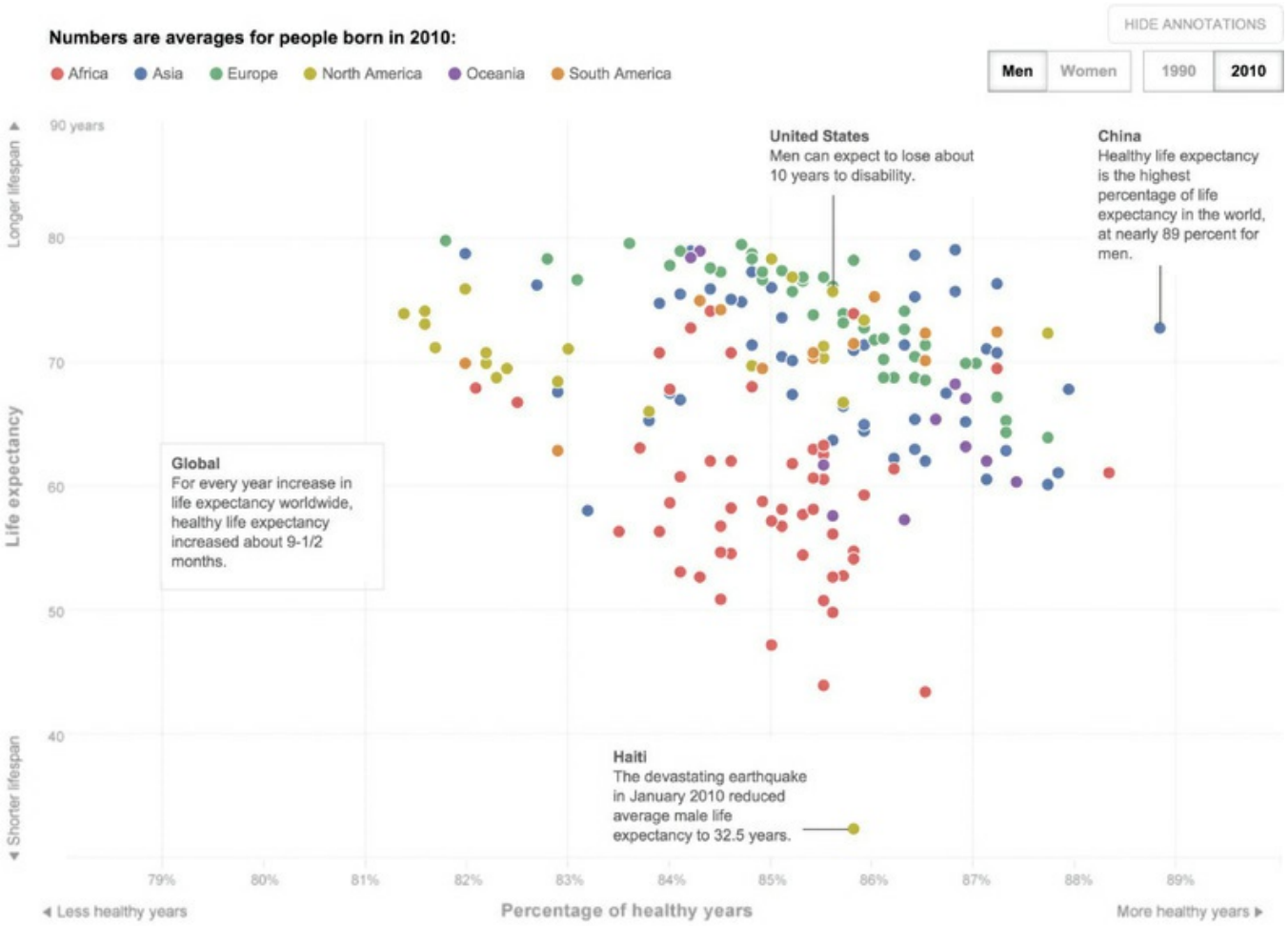


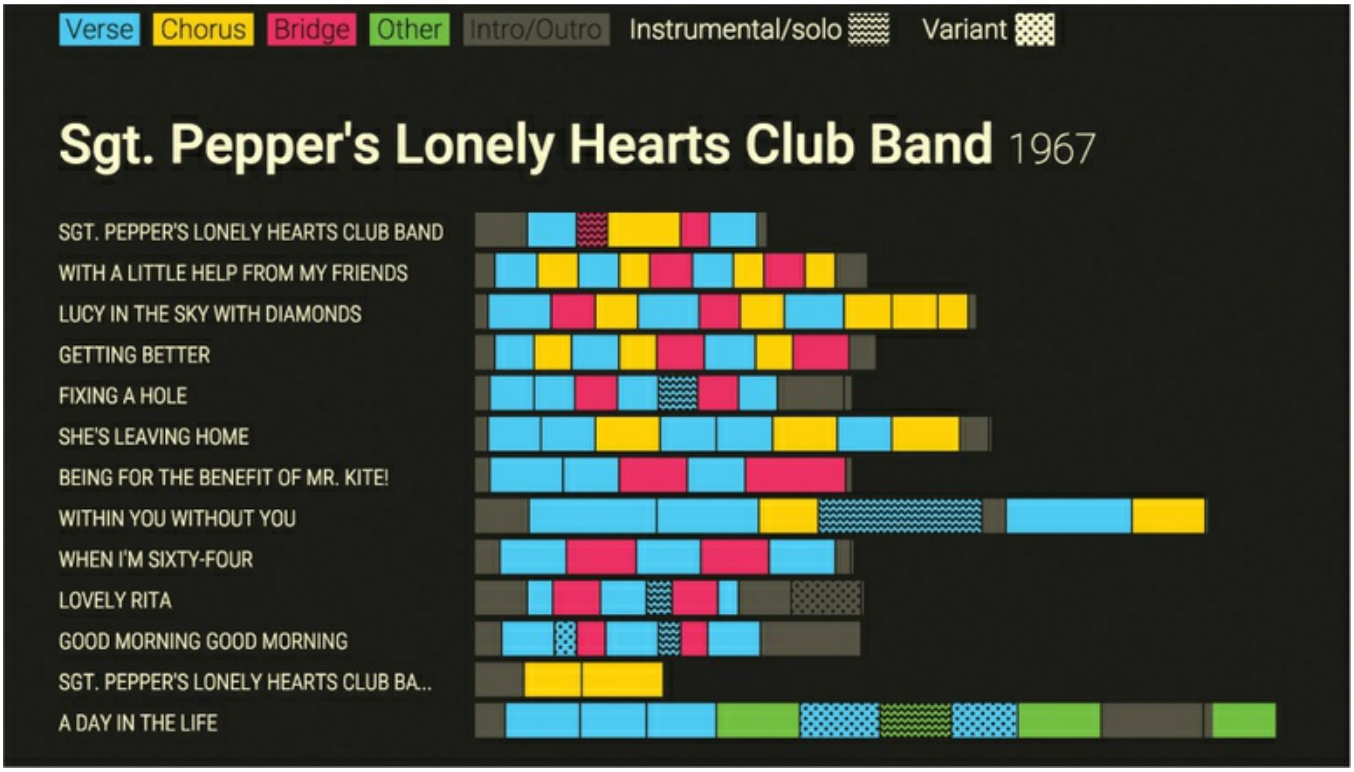
Figure 9.7 How Long Will We Live – And How Well?



Beyond these small numbers, you still typically might only need to contend with assigning colours to around four to six categories, perhaps in analysis that needs to visually distinguish values for different continents of the world, as seen in the scatter plot in Figure 9.7.

As the range of different categories grows, the ability to preserve clear differentiation becomes harder. In expanding your required palette, the colours used become decreasingly unique. The general rule of thumb is that once you have more than 12 categories it will not be possible to find a sufficiently different colour to assign to categories from 13 upwards. Additionally, you are really increasing the demands of learning and recognition for viewers. This then becomes quite a cognitive burden and delays the process of understanding.

Figure 9.8 Charting the Beatles: Song Structure



Two approaches for dealing with this. Firstly, consider offering interactive filters to modify what categories are displayed in a visualisation – thus potentially reducing the impact of so many being available. Secondly, think about transforming your data by excluding or combining categories in to a reduced number of aggregate groupings.

Depending on the subject of your data, sometimes you can look to supplement the use of colour with texture or pattern to create further visible distinctions. In Figure 9.8 you can see two patterns being used occasionally as additive properties to show the structure of tracks on The Beatles' album.

Ordinal (Qualitative)

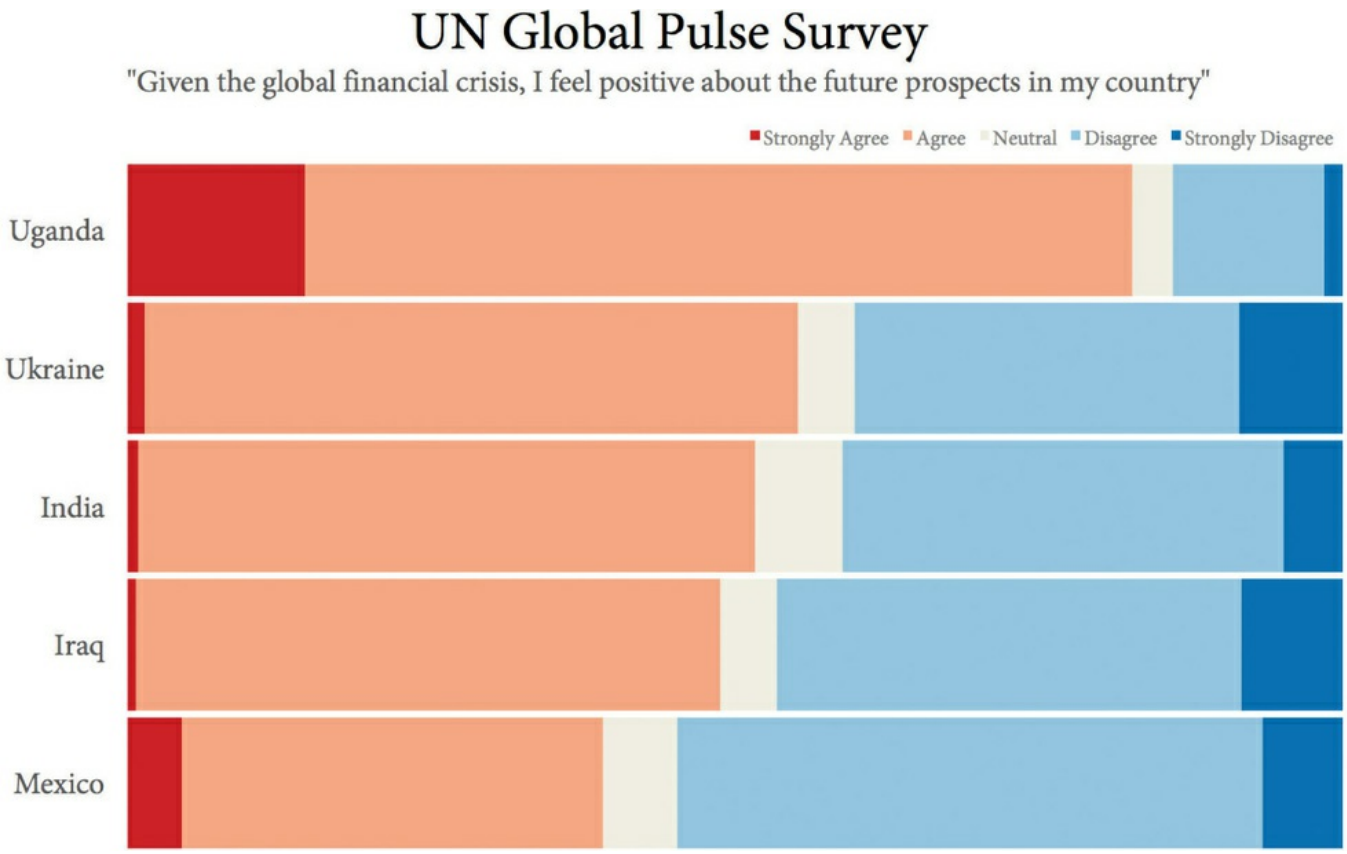
With ordinal data you are still dealing with categories but now they have a natural hierarchy or ordering that can be exploited. The primary motive for using colour in this case is not only to create a visible distinction between each unique category association but also to imply some sense of an order of magnitude through the colour variation. The colour dimensions used to achieve this tend to employ variations of either the saturation or the lightness (or a combination of both). You might also introduce different hues when dealing with *diverging* (dual-direction) scales rather than simply *converging* (single-direction) ones.

Figure 9.9 displays a simple example of colour used to display a converging ordinal variable. This is the teacup that I use in my office. On the inside you can see it has a colour guide to help ascertain how much milk you might need to add: going through Milky, Classic British, Builder's Brew, and finally Just Tea (zero milk).

Figure 9.9 Photograph of MyCuppa Mug

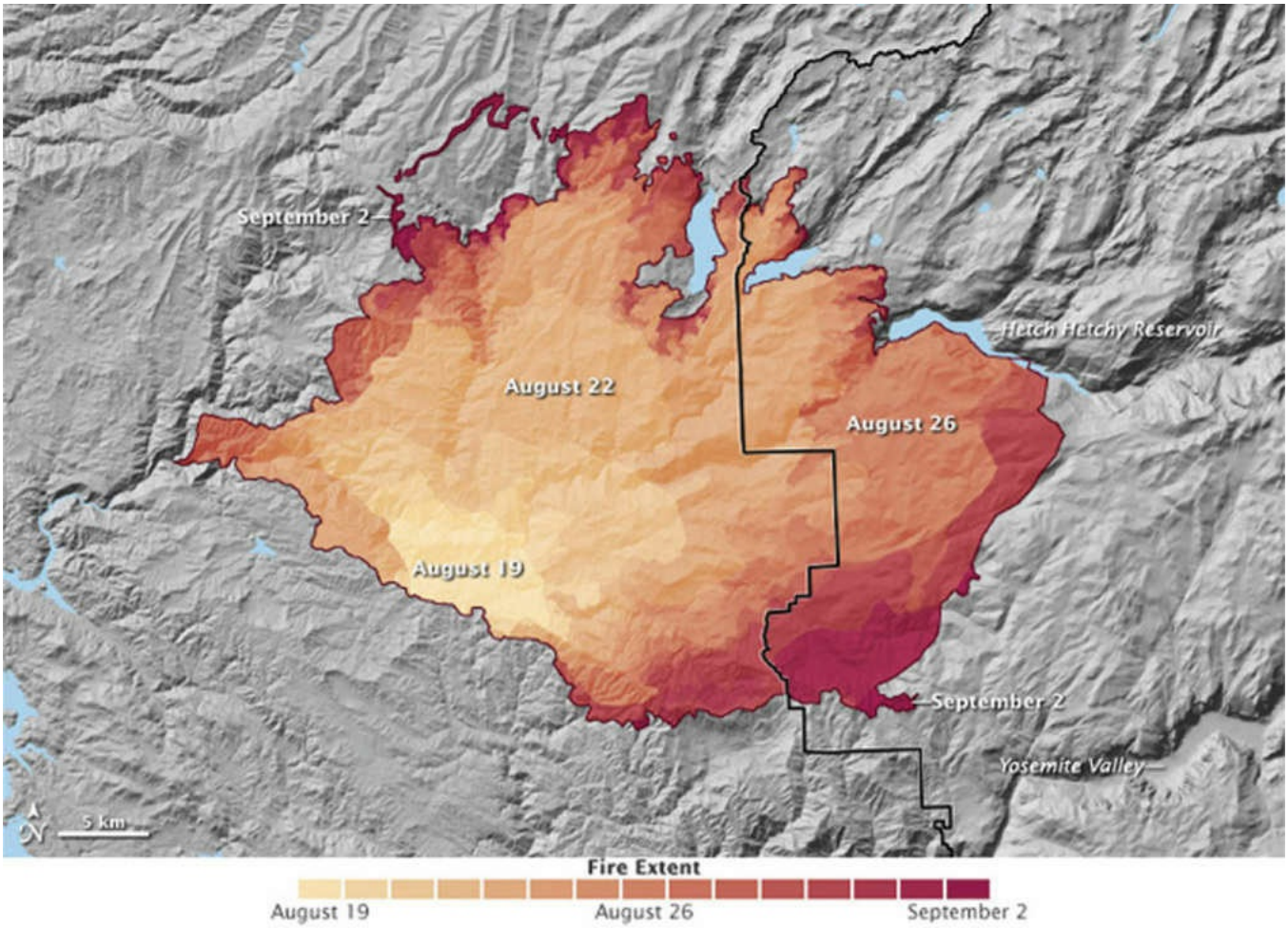
A typical example of a diverging ordinal scale might be seen in the stacked bar chart showing the results of a survey question (Figure 9.10). The answers are based on the strength of feeling: strongly agree, agree, neutral, disagree, strongly disagree. By colouring the agreement in red ('hot' sometimes used to represent 'good') and the disagreement in blue ('cold' to mean 'bad') means a viewer can quickly perceive the general balance of feelings being expressed.

Figure 9.10 Example of a Stacked Bar Chart Based on Ordinal Data



Another example of ordinal data might be to represent the notion of recency. In Figure 9.11 you see a display plotting the 2013 Yosemite National Park fire. Colour is used to display the recorded day-by-day progress of the fire's spread. The colour scale is based on a recency scale with darker = recent, lighter = furthest away (think faded memory).

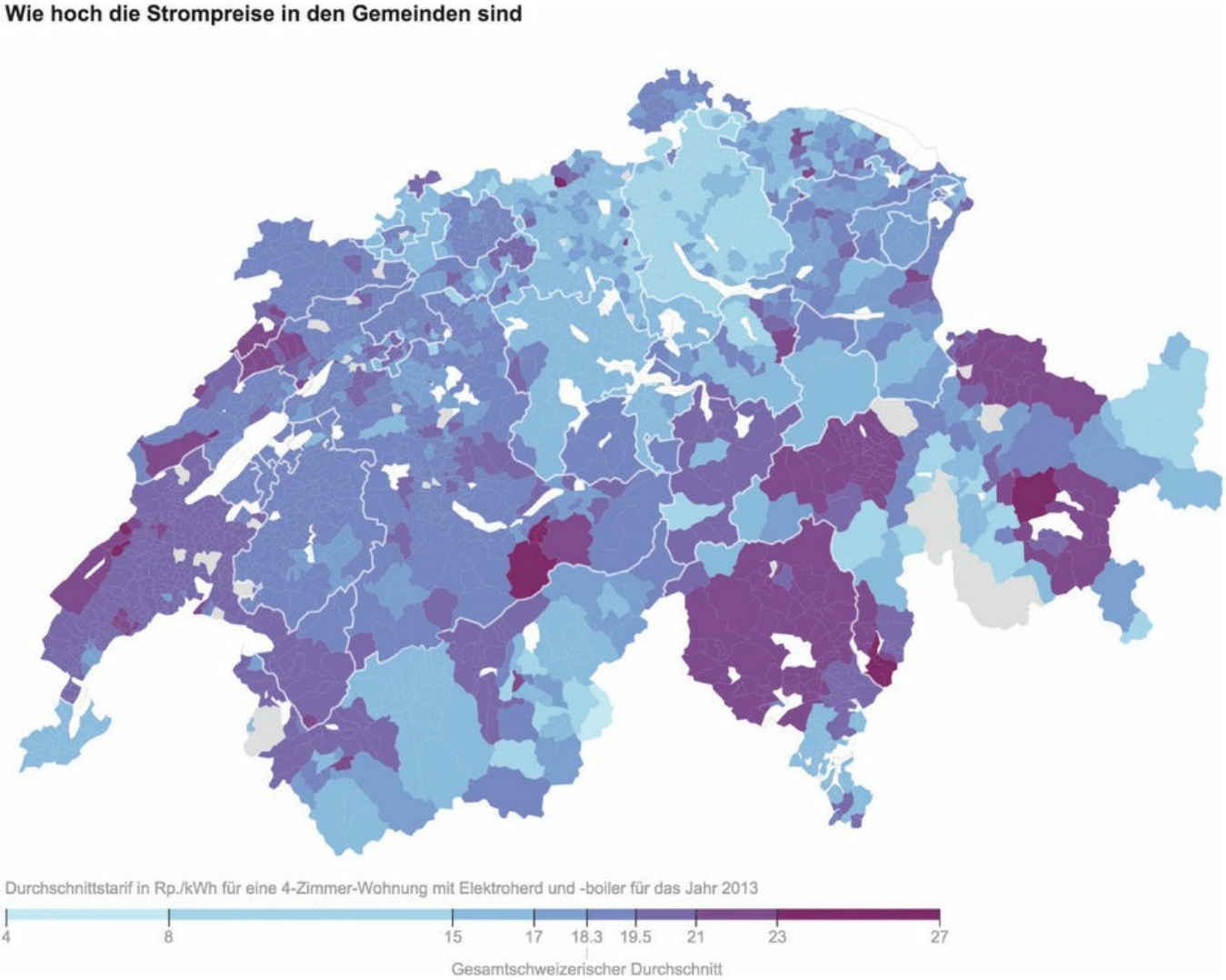
Figure 9.11 The Extent of Fire in the Sierra Nevada Range and Yosemite National Park, 2013



Interval and Ratio (Quantitative)

With quantitative data (ratio and interval) your motive, as it is with ordinal data, is to demonstrate the difference between and of a set of values. In the choropleth map in Figure 9.12, showing the variation in electricity prices across Switzerland, the darker shades of blue indicate the higher values, the lighter tints the lower prices. This approach makes the viewer's perception of the map's values immediate – it is quite intuitive to recognise the implication of the general patterns of light and dark shades.

Figure 9.12 What are the Current Electricity Prices in Switzerland [Translated]

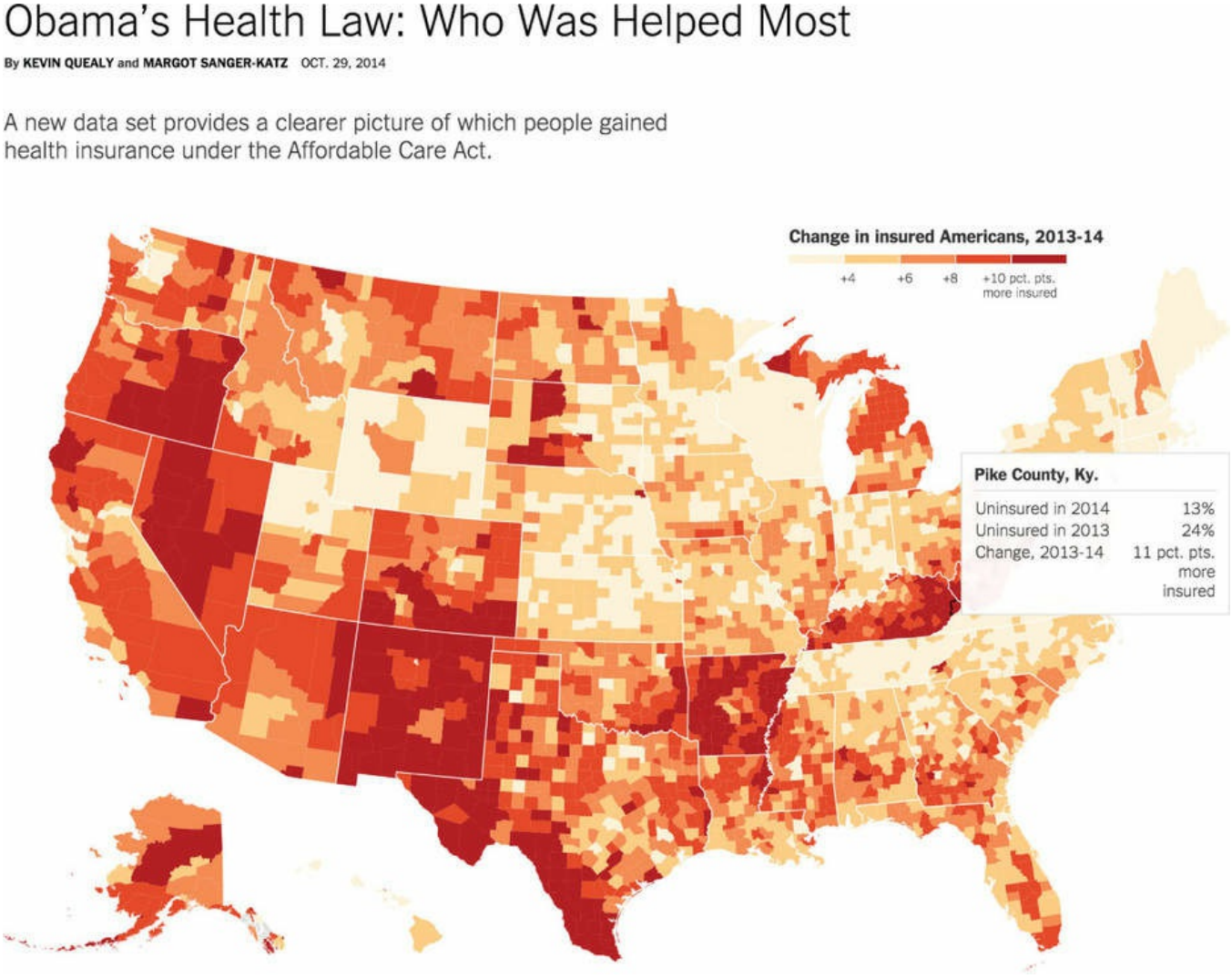


Typically, using colour to represent quantitative data will involve breaking up your data values into discrete classifications or 'bins'. This makes the task of reading value ranges from their associated colour shade or tone a little easier than when using a continuous gradient scale. While our capacity to judge exact variations in colour is relatively low (even with a colour key for reference), we are very capable of detecting local variations of colour through differences in tint, shade or tone. Assessing the relative contrast between two colours is generally how we construct a quantitative hierarchy.

Look at the fascinating local patterns that emerge in the next map (Figure 9.13), comparing increases in the percentage of people gaining health insurance in the USA (during 2013–14). The data is broken down to county level detail with a colour scale showing a darker red for the higher percentage increases.

> Some of the most relevant colour practices for data visualisation come from the field of cartography (as do many of the most passionate colour purists). Just consider the amount of quantitative and categorical detail shown in a reference map that relies on colour to differentiate types of land, indicate the depth of water or the altitude of high ground, present route features of road and rail networks, etc. The best maps pack an incredible amount of detail into a single display and yet somehow they never feel disproportionately overwhelming.

Figure 9.13 Excerpt from 'Obama's Health Law: Who Was Helped Most'



Aside from the big-picture observations of the darker shades in the west and the noticeably lighter tints to the east and parts of the mid-west, take a closer look at some of the interesting differences at a more local level. For example, notice the stark contrast across state lines between the dark regions of southern Kentucky (to the left of the annotated caption) and the light regions in the neighbouring counties of northern Tennessee. Despite their spatial proximity there are clearly strong differences in enrolment on the programme amongst residents of these regions.

Both of these previous examples use a *convergent* colour scale, moving through discrete variations in colour lightness to represent an increasing scale of quantitative values, from zero or small through to large. As illustrated with the stacked bar chart example shown earlier, portraying the range of feelings from an ordinal dataset, sometimes you may need to employ a *divergent* colour scale. This is when you want to show how values are changing in two directions either side of a set breakpoint.

Figure 9.14 Daily Indego Bike Share Station Usage

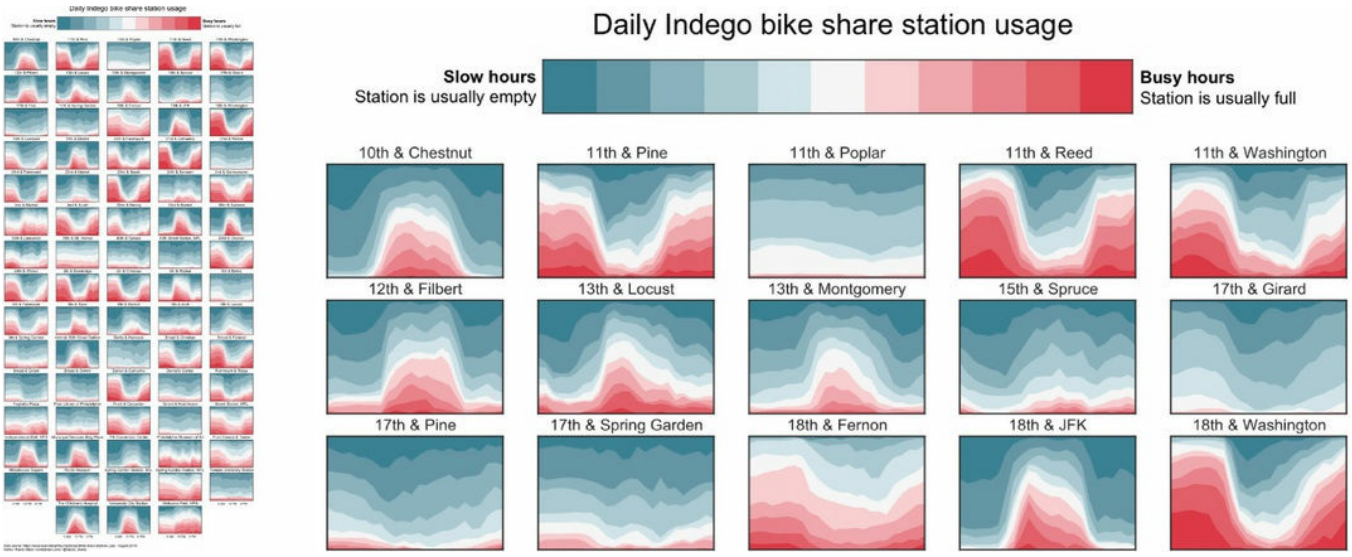


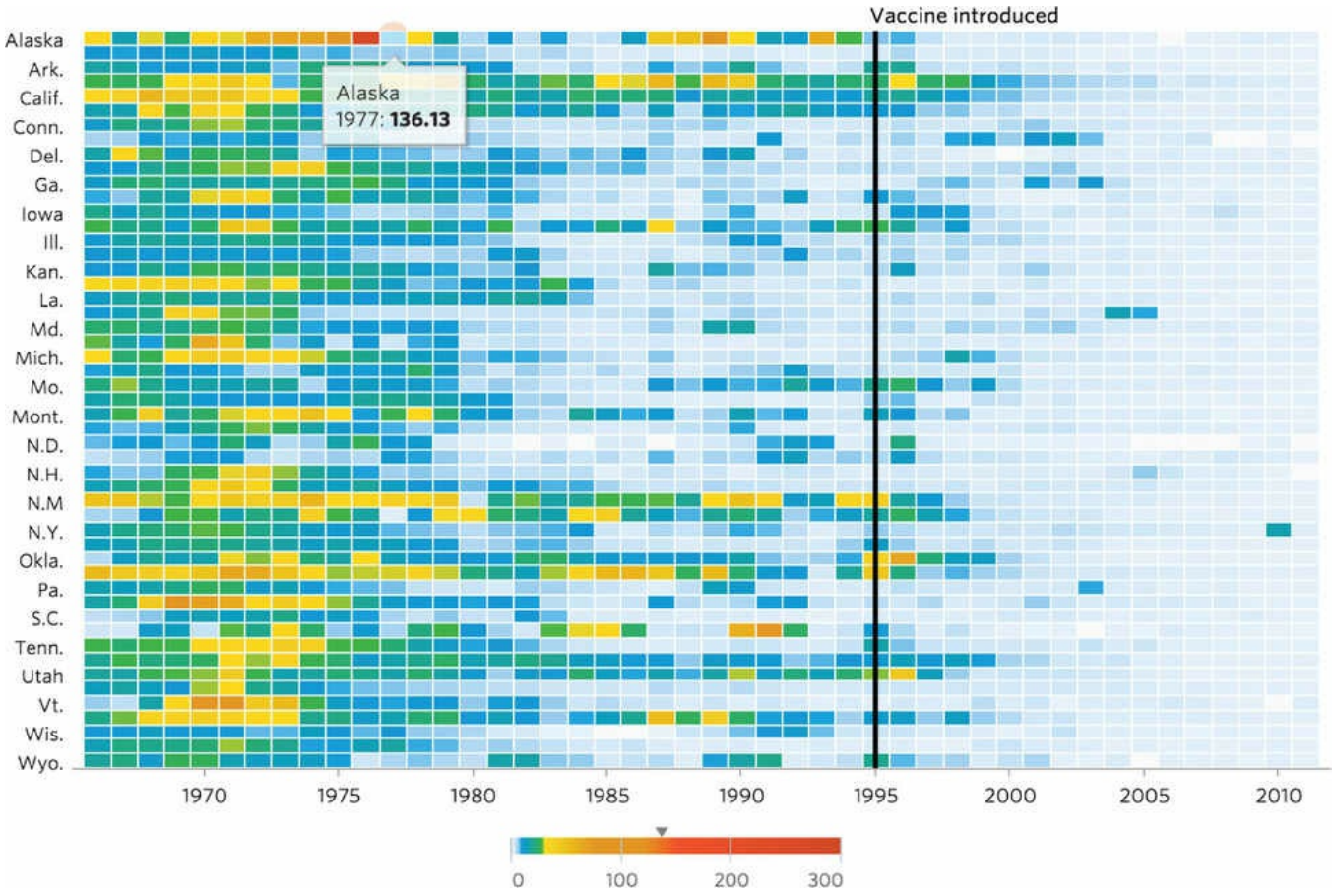
Figure 9.14 shows a cropped view of a larger graphic comparing the relative peaks and troughs of usage across all bike share stations in Philadelphia over a 24-hour period. The divergent colour scale uses two hues and variations in lightness to show the increasingly busy and increasingly slow periods of station activity either side of a breakpoint, represented by a very light grey to indicate the average point. The darkest red means the station is full, the darkest blue means the station is empty.

Regardless of whether you are plotting a converging or diverging scale, judging how you might divide up your colour scales into discrete value bins needs careful thought. The most effective colour scales help viewers perceive not just the relative order of magnitude – higher or lower – but also a sense of the absolute magnitude – how different a value might be compared to another value.

There is no universal rule about the number of value bins. Indeed, it is not uncommon to see entirely continuous colour scales. However, a general rule of thumb I use is that somewhere between between four and nine meaningful – and readable – value intervals should suffice. There are two key factors to consider when judging your scales:

- Are you plotting *observed* data or *observable* data? You might only have collected data for a narrow range of quantities (e.g. 15 to 35) so will your colour classifications be based on this observed range or on the potentially observable data range i.e. the values you know would/could exist with a wider sample size or on a different collection occasion (e.g. 0 to 50)?
- What are the range and distribution of your data? Does it make sense to create equal intervals in your colour classifications or are there more meaningful intervals that better reflect the shape of your data and the nature of your subject? Sometimes, you will have legitimate outliers that, if included, will stretch your colour scales far beyond the meaningful concentration of most of your data values.

Figure 9.15 Battling Infectious Diseases in the 20th Century: The Impact of Vaccines



You can see this effect in Figure 9.15, showing the incidence of Hepatitis A per 100,000 population. There are only three values that exceed 100 (you can see them on the top line for Alaska in the late 1970s). To accommodate these outliers the colour scale becomes somewhat stretched-out, with a wide range of potential values being represented by a dark yellow to red colour. With 99.9% of the values being under 100 there is little discernibly in the blue/green shades used for the lower values. If outliers are your focus, it makes sense to include these and colour accordingly to emphasise their exceptional quality. Otherwise if they risk compromising the discrete detail of the lower values you might look to create a broad classification that uses a single colour for any value beyond a threshold of maybe 75, with even value intervals of maybe 15 below that help to show the patterns of smaller values.

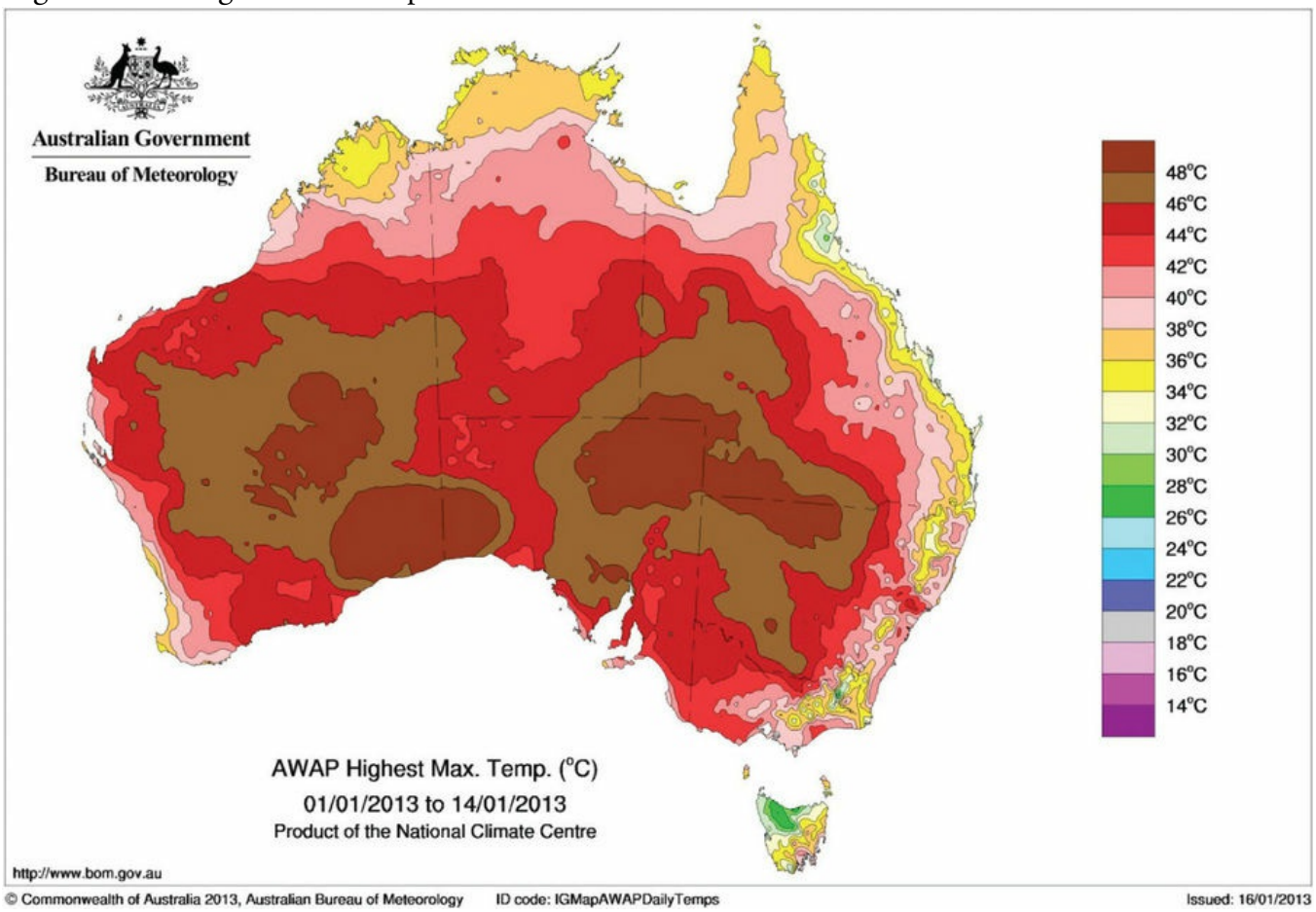
For diverging scales, the respective quantitative shades either side of a breakpoint need to imply parity in both directions. For example, a shade of colour that means +10% one side of the breakpoint should have an equal shade intensity in a different hue on the other side to indicate the same interval, i.e. –10%. Additionally, the darkest shades of hues at the extreme ends of a diverging scale must still be discernible. Sometimes the darkest shades will be so close to black that you will no longer be able to distinguish the differences in their underlying hues when plotted in a chart or map.

As well as considering the most appropriate discrete bins for your values, for diverging scales one must also pay careful attention to the role of the breakpoint. This is commonly set to separate values visually above or below zero or those either side of a meaningful threshold, such as target, average or median.

One of the most common mistakes in using colour to represent quantitative data comes with use of the much-derided rainbow scale. Look at Figure 9.16, showing the highest temperatures across Australia during the first couple of weeks in 2013. Consider the colour key to the right of the map and ask yourself if this feels like a sufficiently intuitive scale. If the key was not provided, would you be able to perceive the order of magnitude

relationship between the colours on the map? If you saw a purple colour next to a blue colour, which would you expect to mean hotter and which colder?

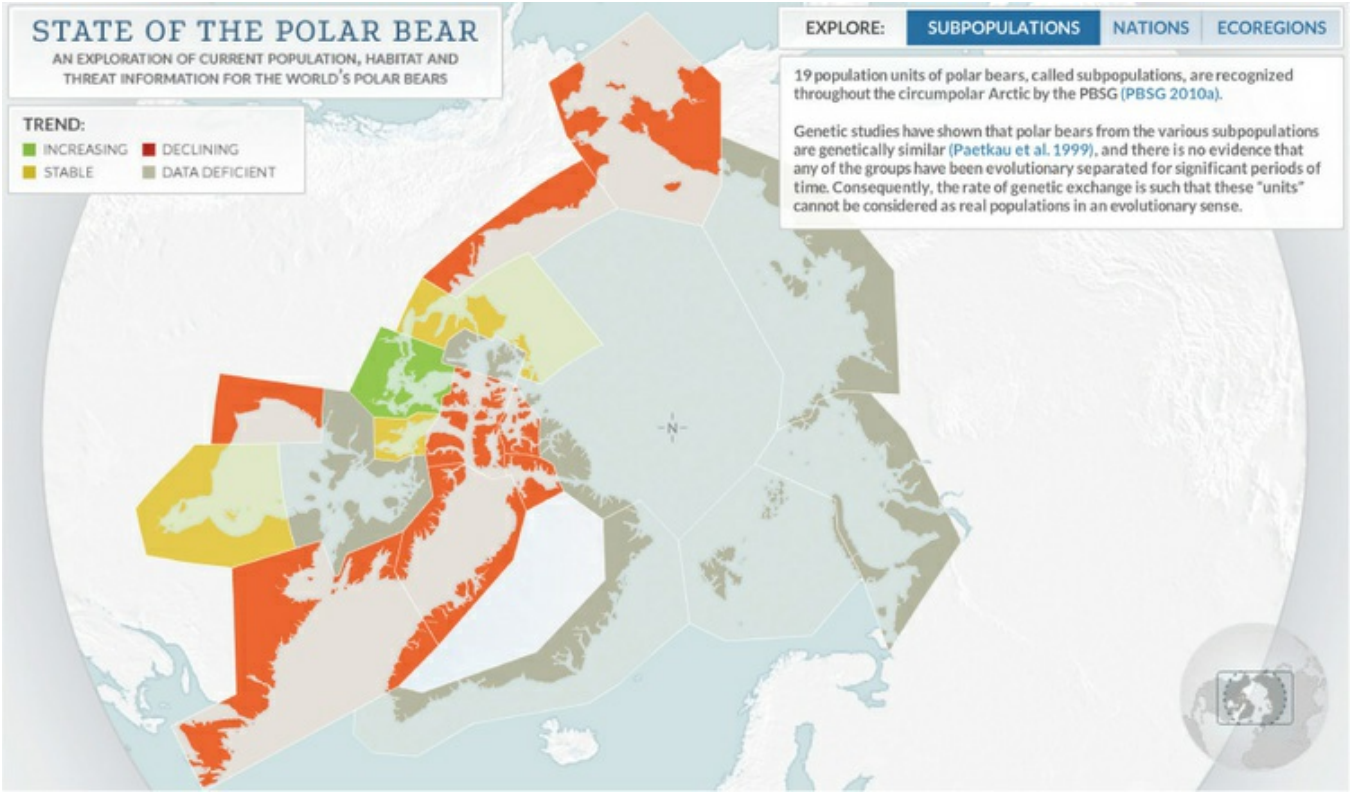
Figure 9.16 Highest Max Temperatures in Australia



While the general implication of blue = 'colder' through to red = 'hotter' is included within sections of this temperature colour scale, it is the presence of many other hues that obstructs the accessibility and creates inconsistency in logic. For instance, do the colours used to show 24°C (light blue) jumping to 26°C (dark green) make sense as a means for showing an increasing temperature? How about 18°C (grey) to 20°C (dark blue), or the choice of the mid-brown used for 46°C which interrupts the increasingly dark red sequence? If you saw on the map a region with the pink tone as used for 16°C would you be confident that you could easily distinguish this from the lighter pink used to represent 38°C? Unless there are meaningful thresholds within your quantitative data – justifiable breakpoints – you should only vary your colour scales through the lightness dimension, not the hue dimension.

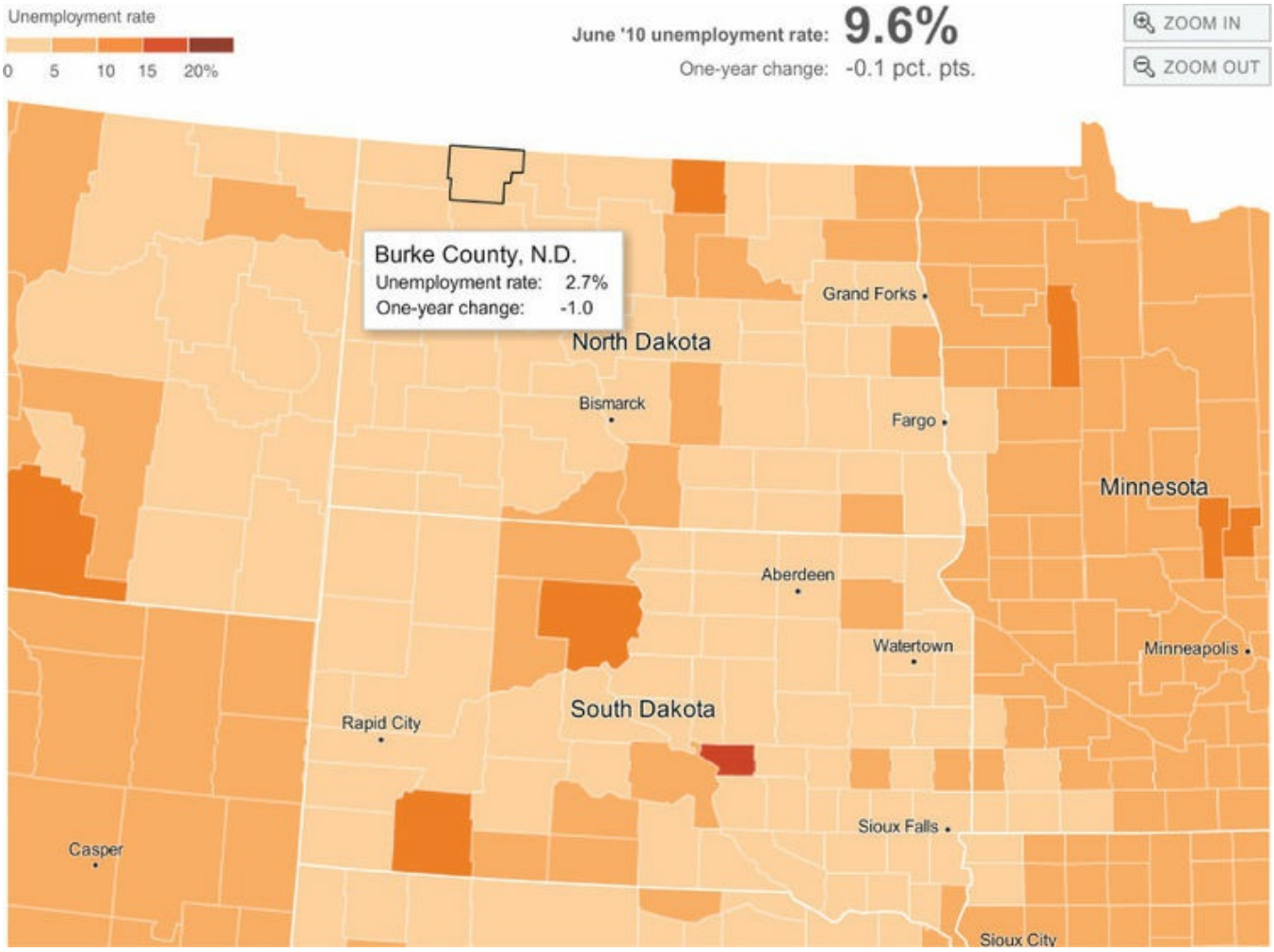
One of the interesting recurring challenges faced by visualisers is how to represent *nothing*. For example, if a *zero* quantity or *no* category is a meaningful state to show, you still need to represent this visually somehow, even though it might possess no size, no position and no area. How do you distinguish between no data and a zero value?

Figure 9.17 State of the Polar Bear



Typically, using colour is one of the best ways to portray this. Figure 9.17 shows one solution to making 'no data' a visible value. This map displays the population trends of the polar bear. Notice those significant areas of grey representing 'data deficient'. A subtle but quite effective political point is being made here by including this status indicator. As I mentioned before, sometimes the absence of data can be the message itself.

Figure 9.18 Excerpt from 'Geography of a Recession'



When considering colour choices for quantitative classifications, you will need to think especially carefully about the lowest value grouping: is it to be representative of zero, an interval starting from zero up to a low value, or an interval starting only from the minimum value and never including zero? In this choropleth map (Figure 9.18) looking at the unemployment rate across the counties of the USA, no value is as low as zero. There might be value that are close, but nowhere is the unemployment rate at 0%. As you can see, the lowest tint used in this colour key is not white, rather a light shade of orange, so as not to imply zero. Whilst not relevant to this example, if you wanted to create a further distinction between the lowest value interval and the 'null' or 'no data' state you could achieve this by using a pure white/blank.

9.3 Features of Colour: Editorial Salience

Having considered options for the application of colour in facilitating data legibility, the next concern is colour used for *editorial salience*. Whereas data legibility was concerned with helping to represent data, using colour for editorial salience is about drawing the viewer's attention to the significant or meaningful features of your display. Colour offers such a potent visual stimulus and an influential means for drawing out key aspects of your data and project that you might feel are sufficiently relevant to make prominent.

Consider again the idea of photography and the effect of taking a photograph of a landscape. You will find the foreground objects are darker and more prominent than the faded view of the background in the distance as light and colour diminish. Using colour to achieve editorial salience involves creating a similar effect of depth across your visualisation's contents: if everything is shouting, nothing is heard.

The goal of using colour to facilitate editorial salience is a suitable contrast. For things to stand out, you are in turn determining which other things will not.

The degree of contrast you might seek to create will vary. Often you will be seeking to draw a significant contrast, maximising the emphasis of a value or subset of values so the viewer can quickly home in on what you have elevated for their attention relative to everything else.

For this reason, grey will prove to be one of your strongest allies in data visualisation. When contrasted with reasonably saturated hues, grey helps to create depth. Elements coloured in greyscale will sit quietly at the back of the view, helping to provide a deliberately subdued context that enables the more emphasised coloured properties to stand proudly in the foreground.

In Figure 9.19, the angle of analysis shows a summary of the most prevalent men's names featuring among the CEOs of the S&P 1500 companies. As you can see there are more guys named 'John' or 'David' than the percentage of *all* the women CEOs combined. With the emphasis of the analysis on this startling statement of inequality the bar for 'All women' is emphasised in a burgundy colour, contrasting with the grey bars of all the men's names. Notice also that the respective axis and bar value labels are both presented using a bold font, which further accentuates this emphasis. It is also editorially consistent with the overriding enquiry of the article. As discussed in Chapter 3, bringing to the surface key insights from data displays in this way contributes towards facilitating an 'explanatory' experience.

Figure 9.19 Fewer Women Run Big Companies Than Men Named John

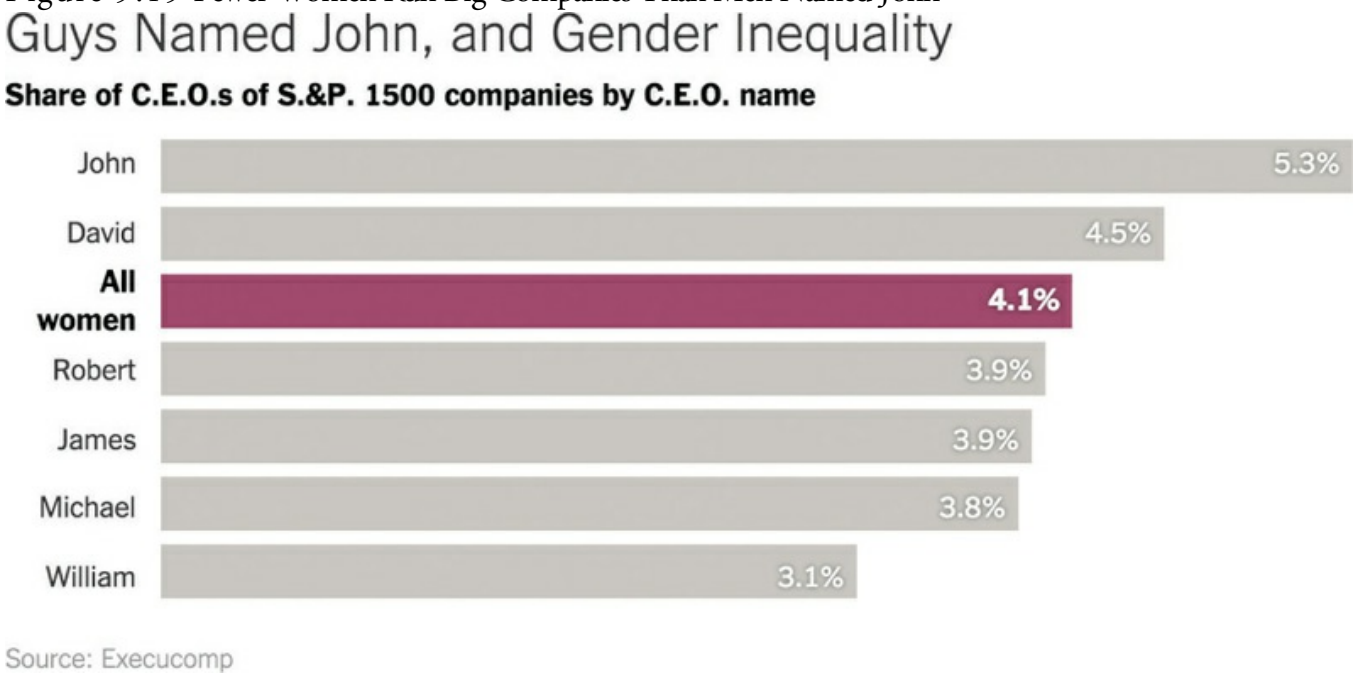
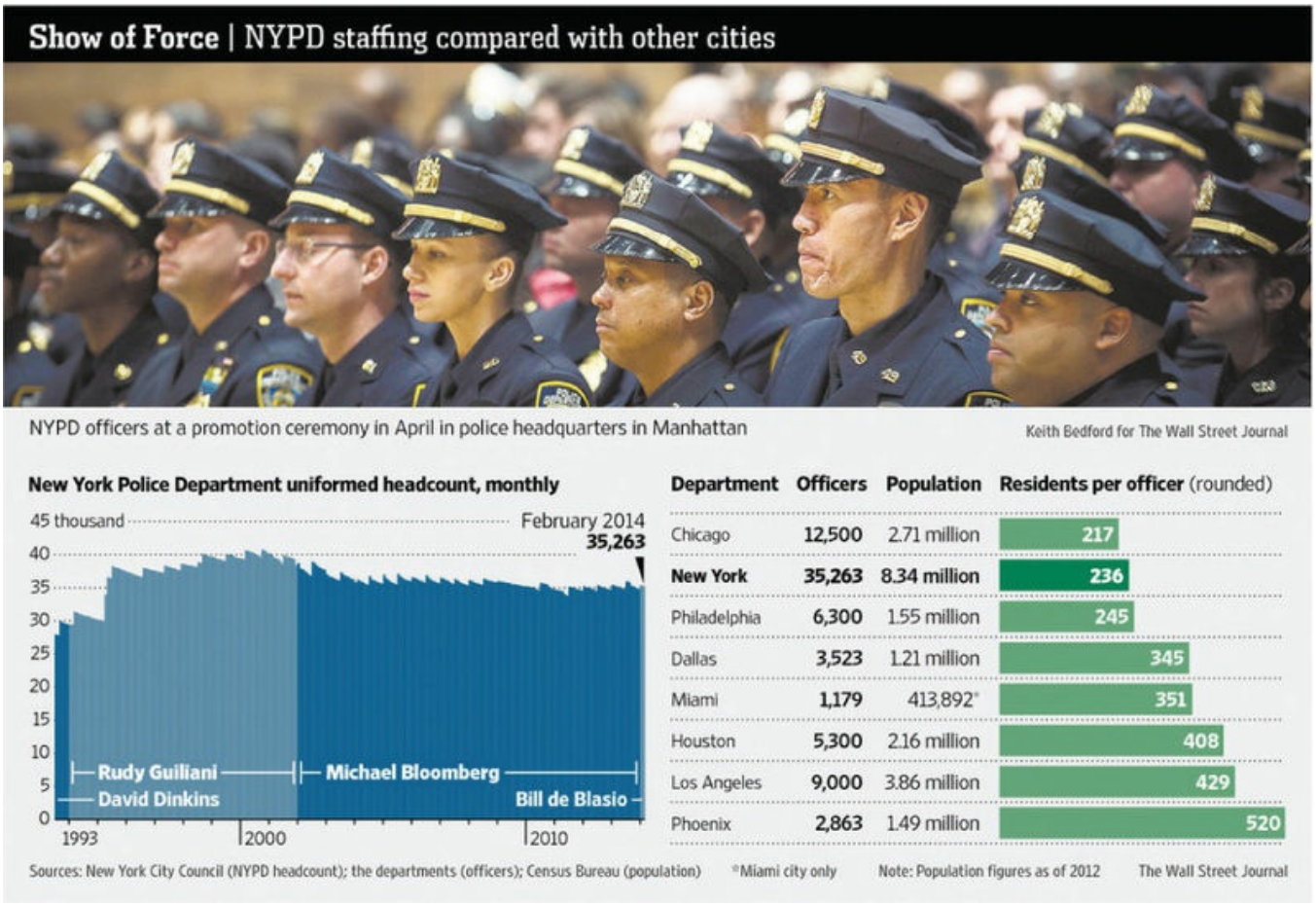


Figure 9.20 NYPD, Council Spar Over More Officers



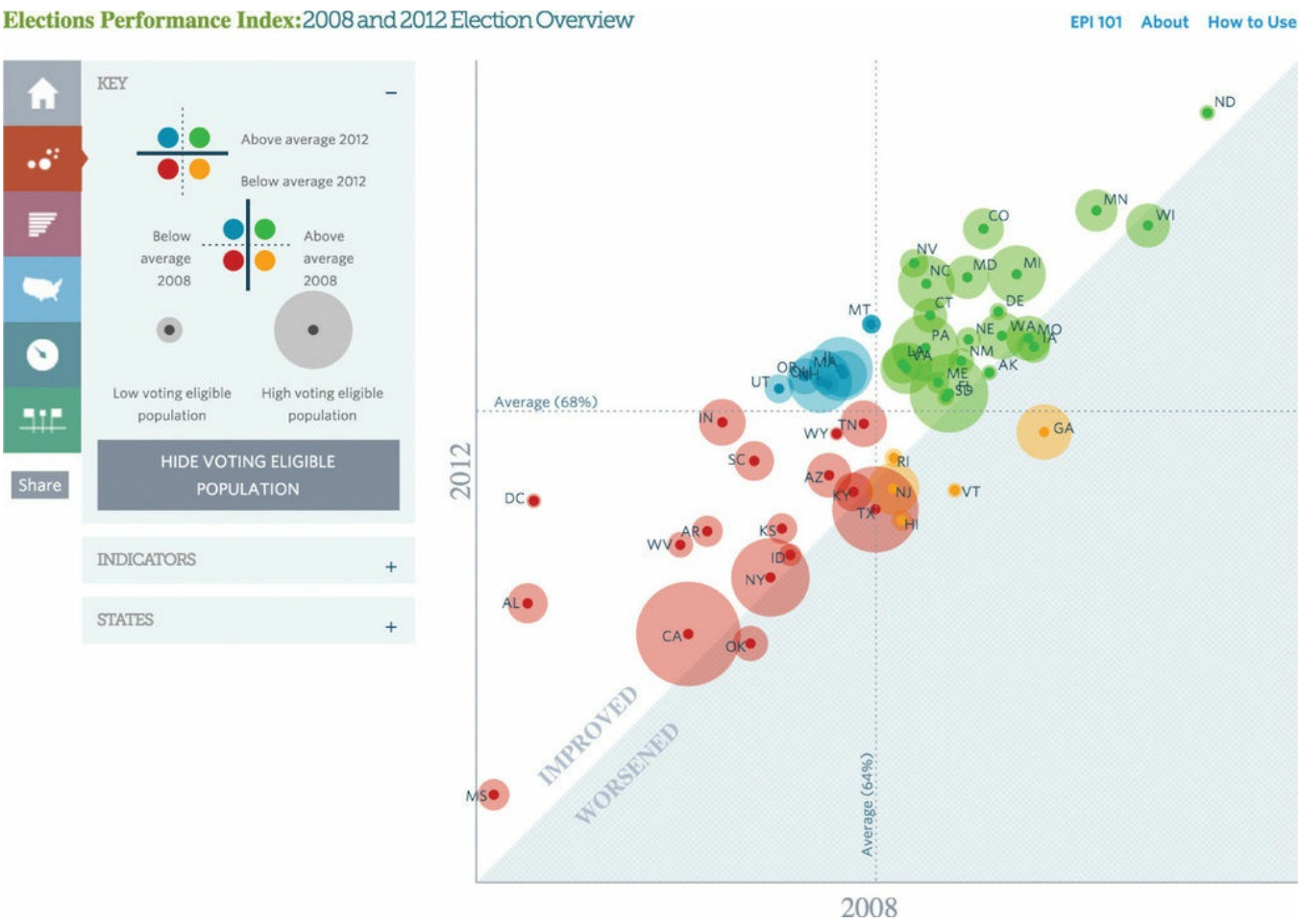
Sometimes, only noticeable contrast – not shouting, just being slightly more distinguishable – may be appropriate. Compared with the previous bar chart example, Figure 9.20 creates a more subtle distinction between the slightly darker shade of green (and emboldened text) emphasising the New York figures compared to the other listed departments in a slightly lighter green. As with the CEOs' example, the object of our attention is the subject of focus in the analysis, in this case regarding a drive for more NYPD officers. This does not need to be any more contrasting; it is just as sufficiently noticeable as the visualiser wishes it to be.

Sometime you will seek to create several levels of visual 'urgency' in the relative contrast of your display. The colour choices in Figure 9.21 gives foreground prominence to the yellow coloured markers and values (the dots are also larger) and then mid-ground/secondary prominence to the slightly muted red markers. In perceiving the values of the yellow markers, the viewer is encouraged to concentrate on primarily comparing these with the red markers. The subtle grey markers are far less visible – closer in shade to the background than the foreground – and deliberately relegated to a tertiary level so they do not clutter up the display and cause unwarranted attention. They provide further context for the distribution of the values but do not need to be any more prominent in their relationship with the foreground and mid-ground colours.

Figure 9.21 Excerpt from a Football Player Dashboard

I touched on the use of encoded overlays earlier where coloured areas or bandings can be used to help separate different regions of a display in order to facilitate faster interpretation of the meaning of values. In the bubble plot in Figure 9.22, you can see the circle markers are colour coded to help viewers quickly ascertain the significance of each location on the chart according to the quadrants in which they fall. Notice how in the background the diagonal shading further emphasises the distinction between above the line 'improvement' and below the line 'worsening', a very effective approach.

Figure 9.22 Elections Performance Index



9.4 Features of Colour: Functional Harmony

After achieving data legibility and editorial salience through astute colour choices, functional harmony is concerned with ensuring that any remaining colour choices will aid, and not hinder, the functional effectiveness and elegance of the overall visualisation.

> 'When something is not harmonious, it's either boring or chaotic. At one extreme is a visual experience that is so bland that the viewer is not engaged. The human brain will reject under-stimulating information. At the other extreme is a visual experience that is so overdone, so chaotic, that the viewer can't stand to look at it. The human brain rejects what it can not organise, what it cannot understand.' Jill Morton, Colour Expert and Researcher

You must judge the overall balance of and suitability of your collective colour choices and not just see these as isolated selections. This is again primarily a judgement about contrast – what needs to be prominent and what needs to be less so. Such an apparent calming quality about a well-judged and cohesive colour palette is demonstrated by Stefanie Posavec's choices in visualising the structure of Walter Benjamin's essay 'Art in the age of mechanical reproduction' (Figure 9.23). There is effortless harmony here between the colour choices extending across the entire anatomy of design: the petals, branches, labels, titles, legend, and background.

A reminder that any and every design feature you incorporate into your display will have a property of colour otherwise they will be invisible. In looking at data legibility and editorial salience you have considered your colour choices for representing data. A desire to achieve functional harmony means considering further colour decisions that will help establish visual relationships across and between the rest of your visualisation's

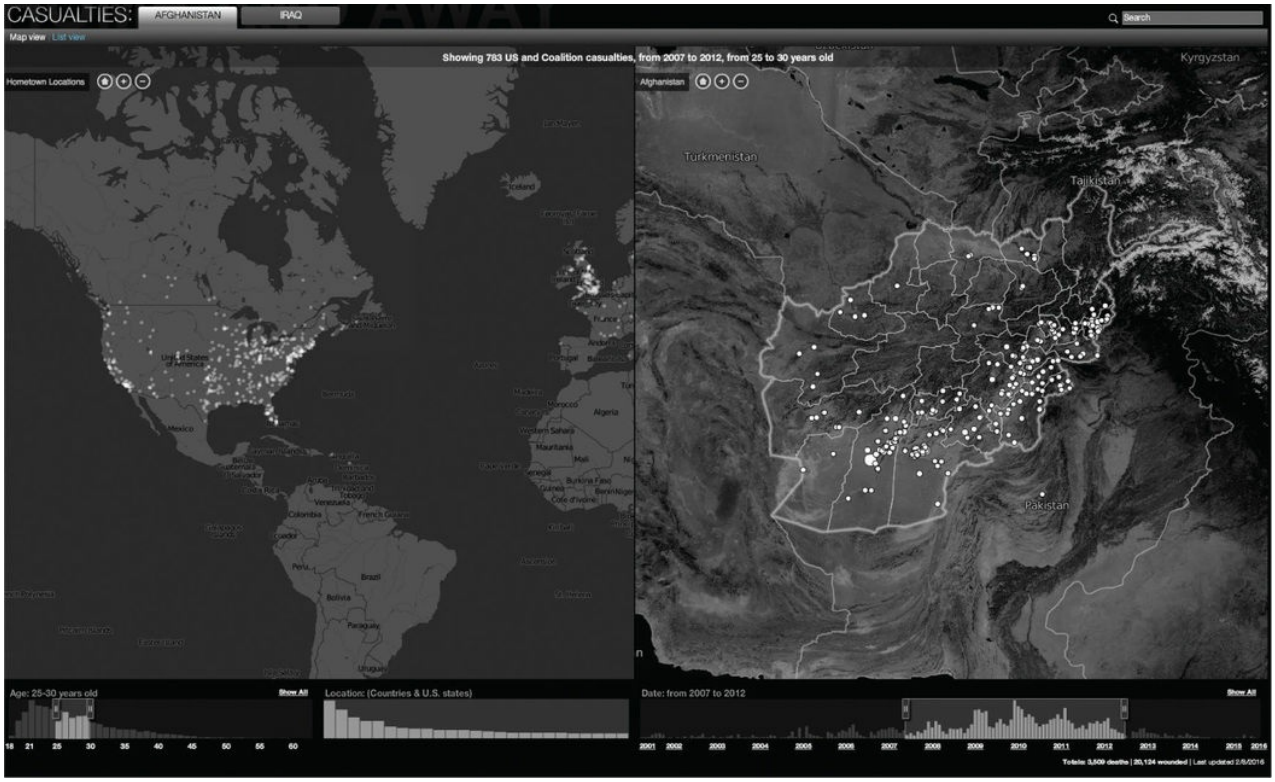
anatomy: its interactive features, annotations and composition.

Figure 9.23 Art in the Age of Mechanical Reproduction: Walter Benjamin

Interactive features: Visible interactive features will include controls such as dropdown menus, navigation buttons, time sliders and parameter selectors. The colour of every control used will need to be harmonious with the rest of the project but also, critically, must be functionally clear. How you use colour to help the user discern what is selected and what is not will need to be carefully judged.

To illustrate this, Figure 9.24 shows an interactive project that examines the connected stories of the casualties and fatalities from the Iraqi and Afghan conflicts. Here you can see that there are several interactive features, all of which are astutely coloured in a way that feels both consistent with the overall tone of the project but also makes it functionality evident what each control's selected status or defined setting is. This is achieved through very subtle but effective combinations of dark and light greys that help create intuitive clarity about which values the user has selected or highlighted. When a button has a toggle setting (on/off, something/something else), such as the 'Afghanistan' or 'Iraq' tabs at the top, the selected tab is highlighted in bright grey and the unselected tab in a more subdued grey. Filters can either frame (include/exclude) or focus (highlight/relegate) the data. The same approach to using brighter greys for the selected parameter values makes it very clear what you have chosen, but also what you have excluded (while making evident the other currently-unselected values from which you can potentially choose).

Figure 9.24 Casualties



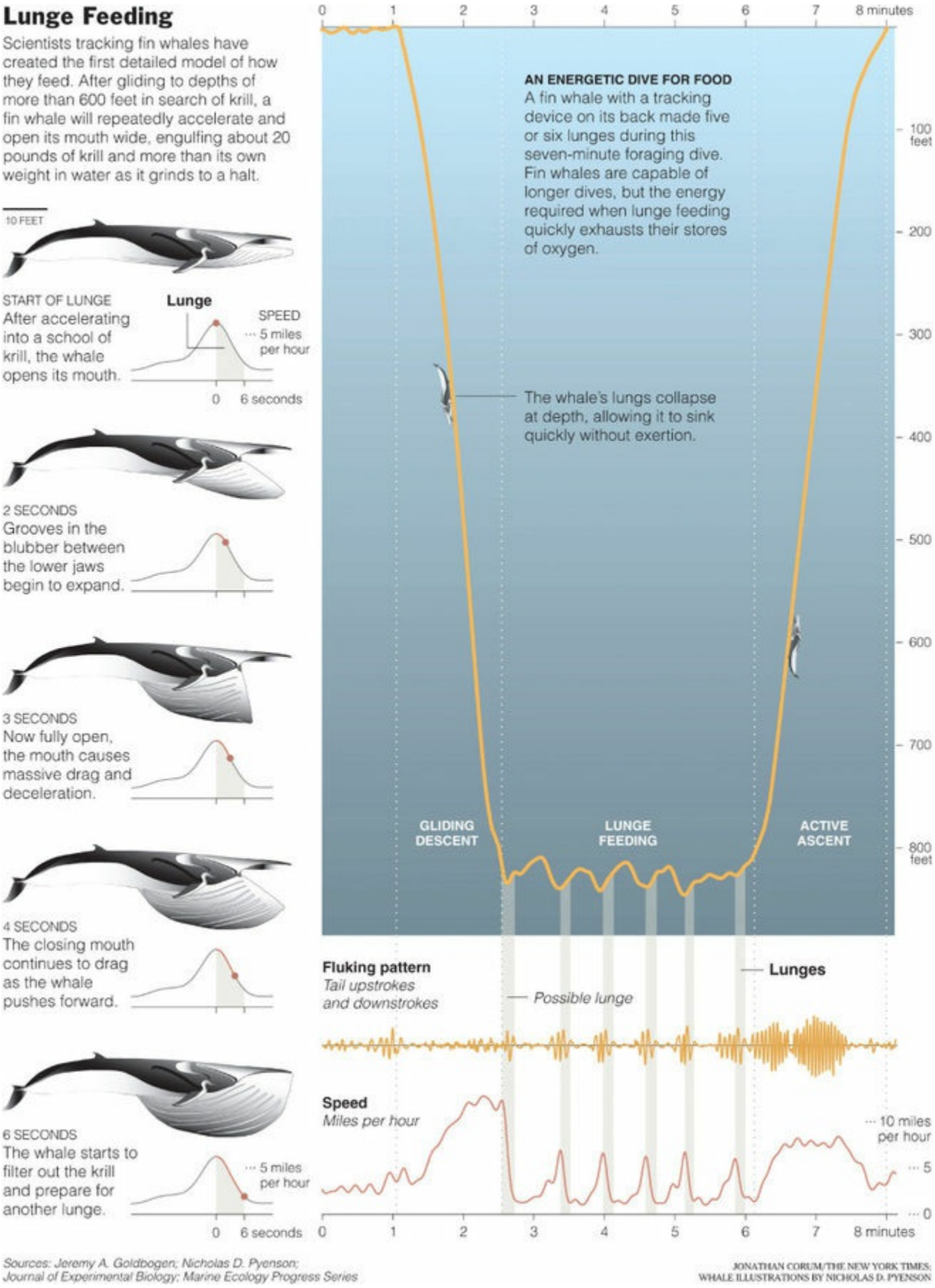
Annotations: Chart annotations such as gridlines, axis lines and value labels all need colouring in a way that will be sympathetic to the colour choices already made for the data representation and, possibly, editorial contrasting. As mentioned in the last chapter, many annotation devices exist in the form of text and so the relative font colour choices will need to be carefully considered. For any annotation device the key guiding decision is to find the level at which these are suitably prominent. Not loud, not hidden, just at the right level. This will generally take a fair amount of trial and error to get right but once again, depending on your context, your first thought should be to consider the merits offered by different shades of grey.

You might be starting to suspect I'm a lobbyist for the colour grey. Nobody wants to live in a world of only grey. The point is more about how its presence enables other colours to come alive. The great Bill Shankly once said 'Football is like a piano, you need 8 men to carry it and 3 who can play the damn thing'. In data visualisation, grey does the heavy lifting so the more vibrant colours can bring the energy and vibrancy to your design.

Figure 9.25 First Fatal Accident in Spain on a High-speed Line [Translated]

Another example of the role of greyscale is demonstrated by Figure 9.25, illustrating key aspects of the tragic rail crash in Spain in 2013. The sense of foreground and background is clearly achieved by the prominence of the scarlet-coloured annotations and visual cues offset against the backdrop of an otherwise greyscale palette.

Figure 9.26 Lunge Feeding



There are other features of annotation that will have an impact on functional harmony through their colouring. Multimedia assets like photos, embedded videos, images and illustrations need to be consistent in tone according to their relative role on the display. If they are to dominate the page then unleash the vibrancy of their colours to achieve this; if they are playing more of a secondary or

supporting role then relegate their constituent colours to allow other primary features due prominence. Figure 9.26 includes small illustrations of a whale, showing how it goes through the stages of lunge feeding. The elegance of the colours used in these illustrations is entirely harmonious with the look and feel of the overall piece. They are entirely at one with the rest of the graphic.

Composition: The clarity in layout of a project will often be achieved by the use of background colour to create logical organisation. In the 'Lunge Feeding' graphic the shading of the blue sea getting darker as it moves down is not attempting to offer a precise representation of the sea, but it gives a sense of depth and draws maximum attention to that panel. It is also naturally congruent with the subject matter.

Figure 9.27 Examples of Common Background Colour Tones

In general, there are no fixed rules on the benefits of any particular colour for background shading. Your choices will depend mostly on the circumstances and conditions in which your viewers are consuming the work. Usually, when there is no associated congruence for a certain background colour, your options will tend to come from one of the selection of neutral and/or non-colours (Figure 9.27). This is because they particularly help to aid accentuation in combination with foreground colours.

Typically, though, a white background (at least for your chart area) gives viewers the best chance of being able to accurately perceive the different colour attributes used in your data representation and the contrasting nature of your editorial contrast.

White – or more specifically emptiness – is one of your most important options for creating functional meaning for *nothingness*, something I touched on earlier. The emptiness of uncoloured space can be used very effectively to direct the eye's attention. It organises the relationship between space on a page without the need for visible apparatus, as seen in the left hand column of the lunge feeding graphic. It can also be used to represent or emphasise values that might have the state of 'null' or 'zero' to maximise contrast.

'The single most overlooked element in visual design is emptiness. Space must look deliberately used.' Alex

9.5 Influencing Factors and Considerations

Having mapped out the ways and places where colour *could* be used, you will now need to consider the factors that will influence your decisions about how colour *should* be used.

Formulating Your Brief

Format: This is a simple concern but always worth pointing out: if you are producing something for screen display you will need to set your colour output to RGB; if it is for print you will need CMYK. Additionally, when you are preparing work for print, running off plenty of proofs before finalising a design is imperative. What you are preparing digitally is a step away from the form of its intended output. What looks like a perfect colour palette on screen may not ultimately look the same when printed.

Print quality and consistency is also a factor. Graphics editors who create work for print newspapers or magazines will often consider using colours as close in tone as possible to pure CMYK, especially if their work is quite intricate in detail. This is because the colour plates used in printing presses will not always be 100% aligned and thus mixtures of colours may be slightly compromised.

As black and white printing is still commonplace, you need to be aware of how your work might look if printed without colour. If you are creating a visualisation that might possibly be printed by certain users in black and white, the only colour property that you can feasibly utilise will be the lightness dimension. Sometimes, as a designer or author, you will be unaware of this intent and the colourful design that you worked carefully towards will end up not being remotely readable.

We all refer to black and white printing, but technically printers do not actually print using white ink, it is just less black or no black.

Furthermore, there is an important difference in how colours appear when published in *colour* and how they appear when published in *black and white*. Hues inherently possess different levels of brightness: the purest blue is darker than the purest yellow. If these were printed in black and white, blue would therefore appear a darker, more prominent shade of grey. If your printed work will need to be compatible for both colour and black and white output, before finalising your decisions check that the legibility and intended meaning of your colour choices are being maintained across both forms.

Setting: For digital displays, the conditions in which the work will be consumed will have some influence over the choice between light and dark backgrounds. The main factor is the relative contrast and the stresses this can place on the eye to adjust against the surroundings. If your work is intended for consumption in a light environment, lighter backgrounds tend to be more fitting; likewise darker backgrounds will work best for consuming in darker settings. For tablets/smartphones, the bordering colour of the devices can also influence the most suitable choice of background tone to most sympathetically contrast with the surroundings.

Colour rules and identities: In some organisations there are style guidelines or branding identities that require the strict use of only certain colour options. Similar guidelines may exist if you are creating work for publication in a journal, magazine, or on certain websites. Guidelines like these are well

intended, driven by a desire to create conformity and consistency in style and appearance. However, in my experience, the basis of such colour guides rarely incorporates consideration for the subtleties of data visualisation. This means that the resulting palettes are often a bad fit for ideal visualisation colour needs, providing limited scope for the variation and salience you might seek to portray.

Your first task should always be to find out if there is any compromise – any chance of not having these colour restrictions imposed. If there is no flexibility, then you will just have to accept this and begin acquainting yourself with the colours you do have to work with. Taking a more positive view, achieving consistency in the use of colour for visualisation within an organisation does have merits if the defined palettes offer suitably rich variety. Developing a recognisable 'brand' and not having to think from scratch about what colours to use every time you face a new project is something that can be very helpful, especially across a team.

Purpose map: Does it need to be utilitarian or decorative? Should it be functional or appealingly seductive? Does it lend itself to being vivid and varied in colour or more muted and distinguished? Colour is the first thing we notice as viewers when looking at a visualisation, so your choices will play a huge part in setting the visible tone of voice. How you define your thinking across the vertical dimension of your purpose map will therefore have an influence on your colour thinking.

Along the horizontal dimension, the main influencing consideration will be a desire to offer an 'explanatory' experience. As mentioned, some of the tactics for incorporating editorial salience will be of specific value if you are seeking to emphasise immediately apparent, curated insights.

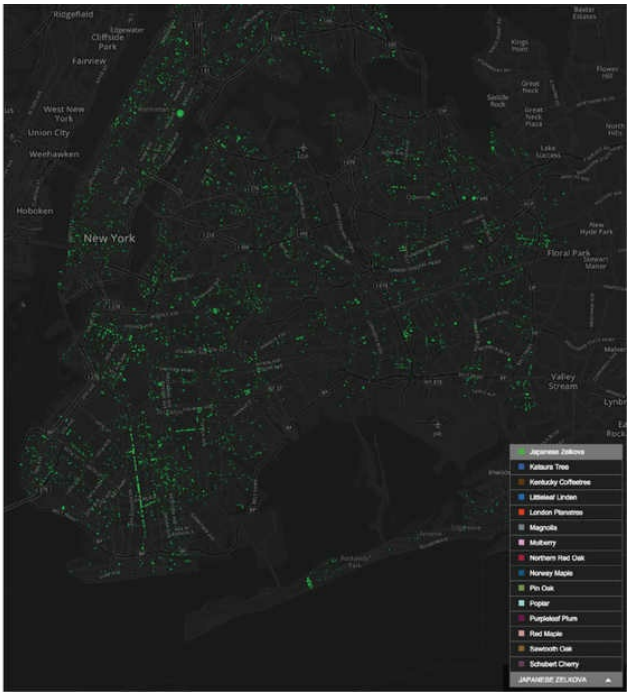
Ideas and inspiration: In the process of sketching out your ideas and capturing thoughts about possible sources of influence, maybe there were already certain colours you had identified as being consistent with your thinking about this subject? Additionally, you might have already identified some colours you wish to avoid using.

Working With Data

Data examination: The characteristics of your data will naturally have a huge impact, on the decisions you make around data legibility. Firstly, the type of data you are displaying (primarily nominal vs all other types) will require a different colour treatment, as explained. Secondly, the range of categorical colour associations (limits on discernible hues) and the range and distribution of quantitative values (numbers of divisions and definition of the intervals across your classification scale) will be directly shaped by the work you did in the examination stage.

In Figure 9.28 you can see a census of the prevalence and species of trees found around the boroughs of New York City. This initial big-picture view creates a beautiful tapestry made up of tree populations across the region (notice the big void where JFK Airport is located).

Figure 9.28 Excerpt from 'NYC Street Trees by Species'



To observe patterns for individual tree types is harder: with 52 different tree species there are simply too many classifications to be able to allocate sufficiently unique colours to each. To overcome this, the project features a useful pop-up filter list which then allows you to adjust the data on view to reveal the species you wish to explore.

It is often the case when thinking about colour classifications that you may need to revisit the data transformation actions to find new ways of grouping your data to create better-fit quantitative value classifications or to look at ways of grouping your categories. For the latter, actions such as combining less important categories in an 'other' bin to reduce the variability or eliminating certain values from your analysis may be necessary.

'If using colour to identify certain data, be careful to not accidentally apply the same identity to a nearby part of the graphic. Don't allow colour to confuse just for the sake of aesthetics. I also like to use colour to highlight. A single colour highlight on a palette of muted colours can be a strong way to draw attention to key information.' Simon Scarr, Deputy Head of Graphics, ThomsonReuters

Establishing Your Editorial Thinking

Focus: When considering the perspective of 'focus' in the editorial thinking stage, you were defining which, if any, elements of content would merit being emphasised. Are there features of your analysis that you might wish to accentuate? How might colour be used to accentuate key insights in the foreground and push other (less important) features into the background? What are the characteristics of your data that you might want to emphasise through changes in colour? For example, are there certain threshold values that will need to be visually amplified if exceeded? Your decisions here will directly influence your thinking about using colour to facilitate editorial salience.

Data Representation

Chart type choice: Specifically in relation to data legibility, depending on which chart type you selected to portray your data, this may have attributes requiring decisions about colour. The heat map and choropleth map are just two examples that use variation in colour to encode quantitative value.

Almost every chart has the potential to use colour for categorical differentiation.

Trustworthy Design

Data classification: The decisions you make about how to encode data through colour have a great bearing on the legibility and accuracy of your design, especially with quantitative data. You will need to ensure the classifications present a true reflection of the shape and characteristics of your data and do not suppress any significant interpretations.

'Start with black and white, and only introduce color when it has relevant meaning. In general, use color very sparingly.' Nigel Holmes, Explanation Graphic Designer

Meaningful: Eliminating arbitrary decisions is not just about increasing the sophistication of your design thinking, it is also an essential part of delivering a trustworthy design. If something looks visually significant in its data or editorial colouring it will be read as such, so make sure it is significant, otherwise remove it. You especially want to avoid any connotation of significant meaning across your functional or decorative colour choices. This will be confusing at best, or will appear deceptive at worst.
Do not try to make something look more interesting than it fundamentally is. Colour should not be used to decorate data. You might temporarily boost the apparent appeal of your work in the eye of the viewer but this will be short-lived and artificial.
Illusions: The relationship between a foreground colour and a background one can create distorting illusions that modify the perceived judgement of a colour. You saw an effect of this earlier with the inverted area chart showing 'Gun deaths in Florida', whereby the rising white mountain was seen by some as the foreground data, when in fact it was the background emptiness framed by the red area of data and the axis line. Illusions can affect all dimensions of colour perception. There are simply too many to mention here and they are hard to legislate for entirely; it is really more about mentioning that you need to be aware of these as a consequence of your colour choices.

Accessible Design

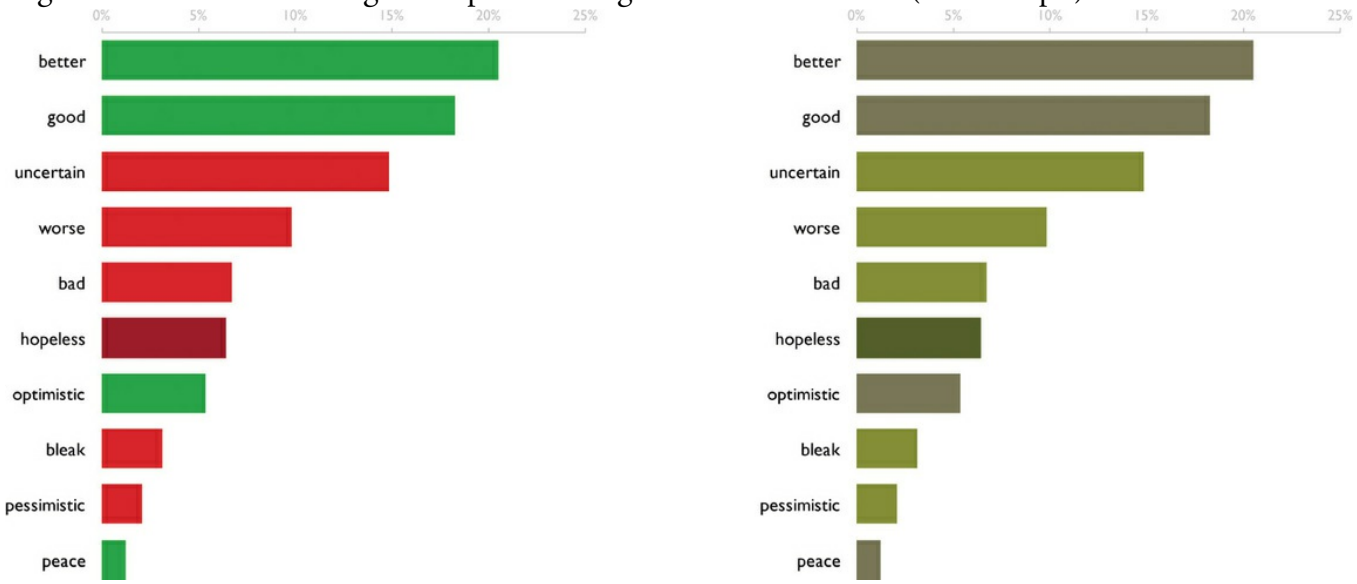
Consistency: Consistency in the use of colours helps to avoid visual chaos and confusion and minimises cognitive effort. When you establish association through colour you need to maintain that meaning for as long as possible. Once a viewer has allocated time and effort to learn what colours represent, that association becomes locked down in the eye and the mind. However, if you then allocate the same colour(s) to mean something different (within the same graphic or on a different page/screen view) this creates an additional cognitive burden. The viewer has almost to disregard the previous association and learn the new one. This demands effort that undermines the accessibility of your design.

Sometimes this can prove difficult, especially if you have a restricted colour palette. The main advice here is to try to maximise the 'space' between occasions of the same colour meaning different things. This space may be physical (different pages, interactive views), time (the simple duration of reading between the associations being changed) or editorial (new subject matter, new angle of analysis). Such space effectively helps to clean the palate (pun intended). Of course, at the point of any new assignment in your colour usage, clear explanations are mandatory.

Visual accessibility: Approximately 5% of the population have visual impairments that compromise their

ability to discern particular colours and colour combinations. Deuteranopia is the most common form, often known as red–green colour blindness, and is a particular genetic issue associated with men. The traffic light scheme of green = 'good', red = 'bad' is a widespread approach for using colour as an indicator. It is a convenient and common metaphor and the reasons for its use are entirely understandable. However, as demonstrated in the pair of graphics in Figure 9.29, looking at some word-usage sentiment analysis, the reds and greens that most of us would easily discern (from the left graphic) are often not at all distinguishable for those with colour blindness (simulated on the right).

Figure 9.29 Demonstrating the Impact of Red-green Colour Blindness (deuteranopia)



Of course, if you have a particularly known, finite and fixed audience then you can easily discover if any colour-blindness issues do in fact exist. However, if your audience is much larger and undefined you are going to alienate potentially 1 person in every 14, in which case the use of the default red–green colour combination is not acceptable. Be more sensitive to your viewers by considering other options:

Figure 9.30 Colour-blind Friendly Alternatives to Green and Red

Similar metaphor of green (good) and red (bad) but with a lighter green and pinker red	#4DAC26 #D01C8B
Similar metaphor of green (good) but replacing red with a purple tone (for bad)	#008837 #7B3294
Similar metaphor but using a teal – blue-green – colour (good) to replace the standard green and a brown tone to replace red (for bad)	#018571 #A6611A
Switching meaning completely, this option uses red (hot = good) and blue (cold = bad)	#D7191C #2C7BB6
An alternative to only using colour is to supplement your choices with symbols, to introduce useful 'redundancy'. In this case the default reds and greens would be okay as the shapes would encode the same meaning.	

If you are working on an interactive solution, you may consider having a toggle option to switch between different colour modes. For print outputs you might normally have reduced flexibility, but in certain circumstances the option of creating dual versions (second output for colour-impaired viewers) may be legitimate.

Connotations and congruence: Whether it is in politics, sport, brands or in nature, there are many subjects that already have established colour associations you can possibly look to exploit. This association may sit directly with the data, such as the normal colour associations for political party categories, or more through the meaning of the data, such as perhaps through the use of green to present analysis about ecological topics.

In support of accessible design, exploiting pre-existing colour associations in your work can create more immediacy in subject recognition. You might also benefit from the colour learning experiences your viewers may already have gone through. This provides a shortcut to understanding through familiarity.

However, while some colour connotations can be a good thing, in some cases they can be a bad thing and possibly should be avoided. You need to be considerate of and sensitive to any colour usage to ensure that you do not employ connotations that may have a negative implication and may evoke strong emotions and reactions from people.

Sometimes a colour is simply incongruent with a subject. You would not use bright, happy colours if you were portraying data about death or disease. Earlier, in the 'Vision: Ideas' section, I described a project context where I knew I wanted to avoid the use of blue colours in a particular project about psychotherapy treatment in the Arctic, because it would carry an unwelcome clichéd association given the subject matter. The use of 'typical' skin colours to represent ethnic groups in a visualisation is something that would be immediately clumsy (at best) and offensive (at worst).

Cultural sensitivities and inconsistencies are also important to consider. In China, for example, red is a lucky colour and so the use of red in their stock market displays, for example, indicates the rising values. A sea of red on the FTSE or Dow Jones implies the opposite. In Western society red is often the signal

for a warning or danger.

Occasionally established colour associations are out of sync with contemporary culture or society. For example, when you think about colour and the matter of gender, because it has been so endlessly utilised down the years, it is almost impossible not to think instinctively about the use of blue (boys) and pink (girls). My personal preference is to avoid this association entirely. I agree with so many commentators out there that the association of pink to signify the female gender, in particular, is clichéd, outdated and no longer fit for purpose. It is not too much to expect viewers to learn the association of – at most – two new colours for representing gender.

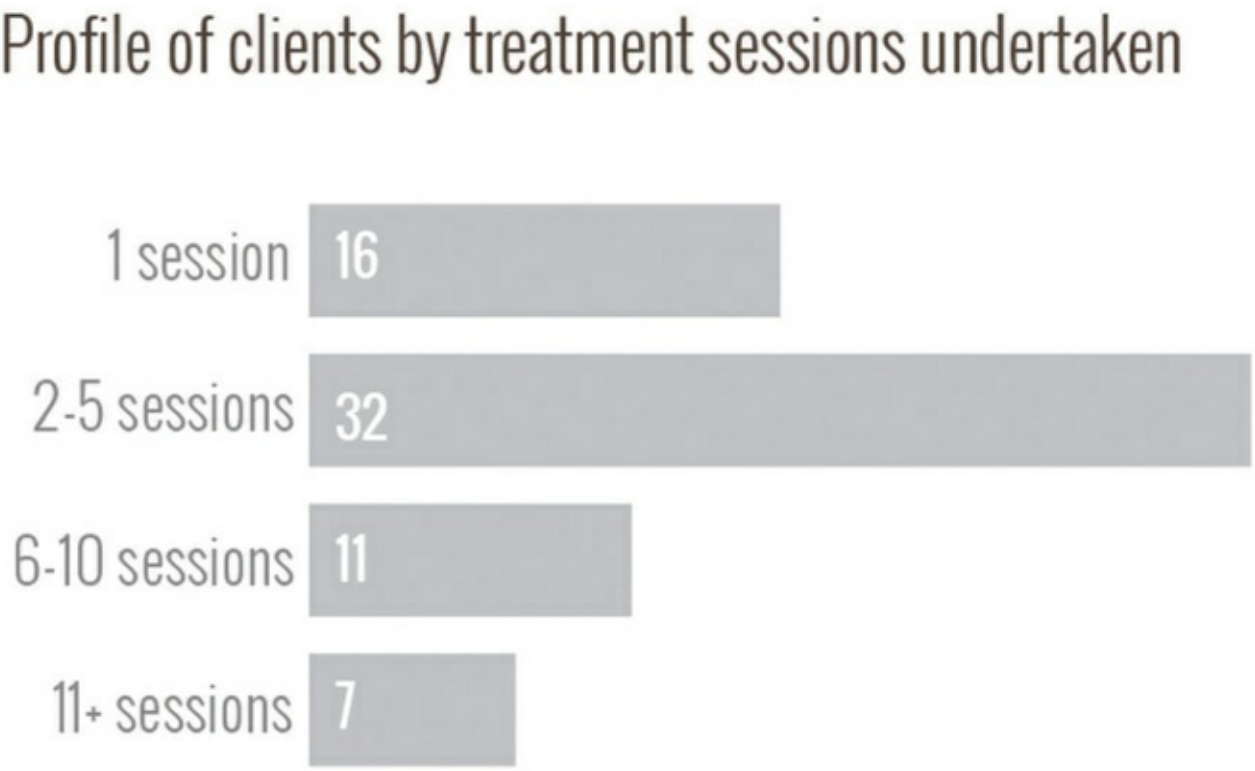
Elegant Design

Unity: As I alluded to in the discussion about using colours for editorial salience, colour choices are always about contrast. The effect of using one colour is not isolated to just that instance of colour: choosing one colour will automatically create a relationship with another. There is always a minimum of two colours in any visualisation – a foreground and background colour – but generally there are many more.

We notice the impact of colour decisions more when they are done badly. Inconsistent and poorly integrated colour combinations create jarring and discordant results. If we do not consciously notice colour decisions this probably means they have been seamlessly blended into the fabric of the overall communication.

Neutral colouring: Even if there is no relevance in the use of colour for quantitative or categorical classifications, you still have to give your chart some colour, otherwise it will be invisible. The decision you make will depend again on the relative harmony with other colour features but should also avoid unnecessarily 'using up' a useful colour. Suppose you colour your bars in blue but then elsewhere across your visualisation project blue would have been a useful colour to show something meaningful; you then have unnecessarily taken blue out of the reckoning. My default choice is to go with grey to begin with (Figure 9.31) and only use a colour if there is a suitable and available colour not used elsewhere or if it needs to be left as a back- or mid-ground artefact to preserve prominence elsewhere in the display.

Figure 9.31 Excerpt from 'Pyschotherapy in The Arctic'

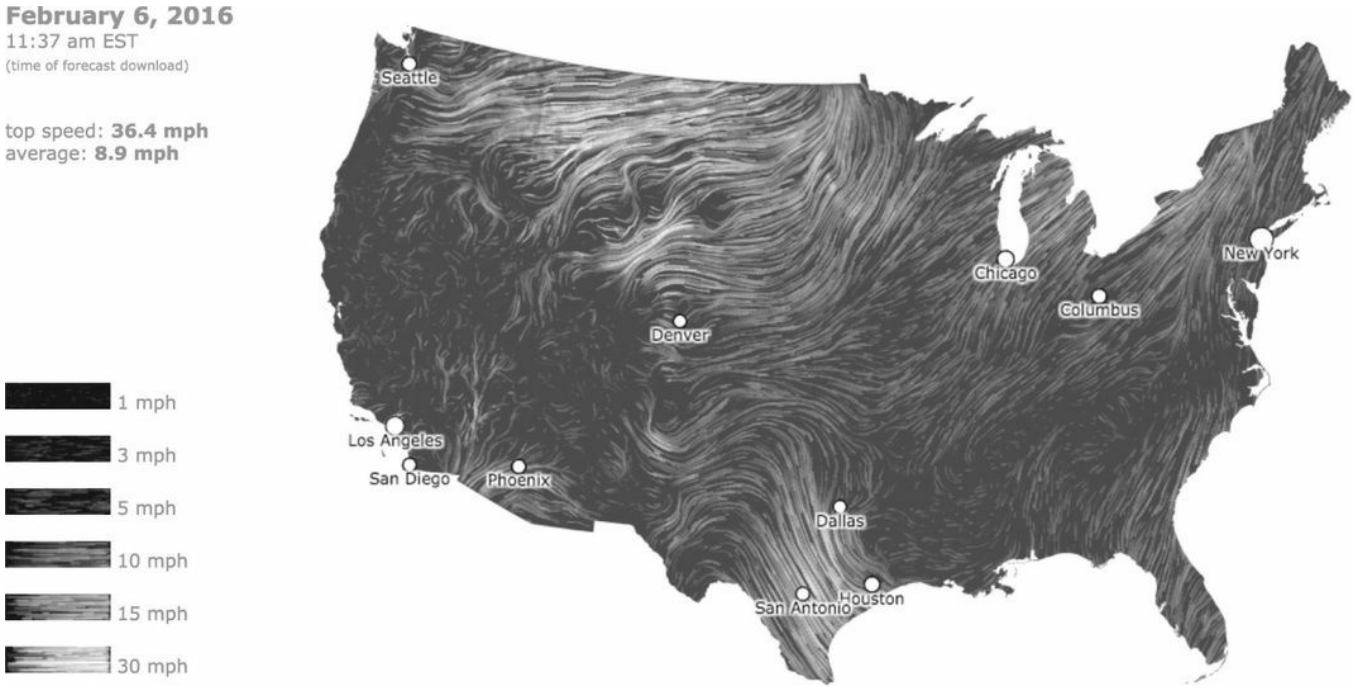


Justified: Achieving elegant design is about eliminating the arbitrary. In thinking about colour usage I often get quite tough with myself. If I want to show any feature on my visualisation display I have to seek permission from myself to unlock access to the more vibrant colours by justifying why I should be allowed to use and apply that colour (I know what you're thinking, 'what a fun existence this guy leads'). Elegance in visualisation design is often about using only the colours you need to use and avoiding the temptation to inject unnecessary decoration. The Wind Map project (Figure 9.32) demonstrates unquestionable elegance and yet uses only a monochromatic palette. There is no colouring of the sea, no topographic detail, no emphasising of any extreme wind speed thresholds being reached. The resulting elegance is quite evident: the map has artistic *and* functional beauty.

To emphasise again, I am not advocating a need to pursue minimalism: while you can create incredibly elegant and detailed works from a limited palette of colours, justifying the use of colours is not the same as unnecessarily restricting the use of colour.

Feels right: The last component of influence is yourself. Sometimes you will just find colours that feel right and look good when you apply them to your work. There is maybe no underlying science behind such choices, and as such you will simply need to back your own instinctive judgement as an astute visualiser and know when something looks good. Creating the right type of visual appeal, something that is pleasing to the eye and equally fit for purpose in all the functional ways I have outlined, is a hard balance to achieve, but you will find that weighing up all these different components of influence alongside your own flair for design judgement will give you the best chance of getting there.

Figure 9.32 Wind Map



Summary: Colour

Data legibility involves using colours to represent different types of data. The most appropriate colour association or scale decisions will depend on the data type: nominal (qualitative), ordinal (qualitative), interval and ratio (quantitative).

Editorial salience is about using colour to direct the eye. For which features and to what degree of emphasis do you want to create contrast?

Functional harmony concerns deciding about every other colour property as applied to all interactive features, annotations and aspects of your composition thinking.

Influencing Factors and Considerations

- Formulating the brief: format, setting, colour rules and imposed guidelines all have a significant impact. Your definitions about both tone and and experience, on the purpose map, will lead to specific choices being more suitable than others. What initial ideas did you form? Have any sources of inspiration already implanted ideas inside your head about which colours you could use?
- Working with data: what type of data and what range of values/number of classifications have you got?
- Establishing your editorial thinking: what things do you want to emphasise or direct the eye towards (focus)?
- Data representation: certain chart type choices will already include colour as an encoded attribute.
- Trustworthy design: ensure that your colour choices are faithful to the shape of your data and the integrity of your insights. If something looks meaningful it should be, otherwise it will confuse or deceive.
- Accessible design: once you've committed colour to mean something preserve the consistency of association for as long as possible. Be aware of the sensitivities around visual accessibility and positive/negative colour connotations.
- Elegant design: the perception of colours is relative so the unity of your choices needs to be upheld. Ensure that you can justify every dot of colour used and, ultimately, rely on your own judgment to

determine when your final palette feels right.

Tips and Tactics

- Use the squint test: shrink things down and/or half close your eyes to see what coloured properties are most prominent and visible – are these the right ones?
- Experimentation: trial and error is still often required in colour, despite the common sense and foundation of science attached to it.
- Developing a personal style guide for colour usage saves you the pain of having to think from scratch every time and will help your work become more immediately identifiable (which may or may not be an important factor).
- Make life easier by ensuring your preferred (or imposed) colour palettes are loaded up into any tool you are using, even if it is just the tool you are using for analysis rather than for the final presentation of your work.
- If you are creating for print, make sure you do test print runs of the draft work to see how your colours are looking – do not wait for the first print when you (think you) have finished your process.