Thinking About Data Visualisation

Christopher Ball
Christopher.Ball@rbnz.govt.nz
Reserve Bank of New Zealand
2021 Economics Department Seminar Series

Outline

1. Introduction
2. Fundamentals
3. Design philosophy
4. Special charts
5. Animation
6. Interactivity
7. Colour
8. Storytelling with Data

Introduction

Fundamentals

Design philosophy

Special charts

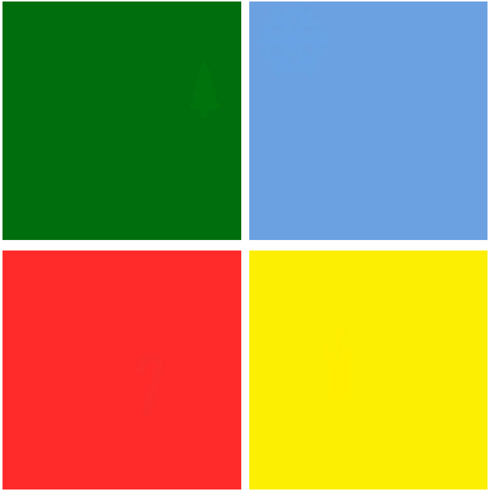
Animation

Interactivity

Colour

Reality leaves a lot to the imagination.

Colour



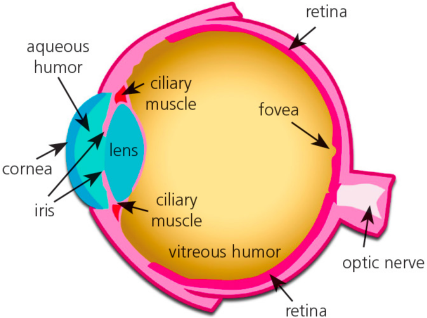
Source: https://www.buzzfeed.com/jonmichaelpoff/only-1-in-20-people-can-ace-this-color-vision-test

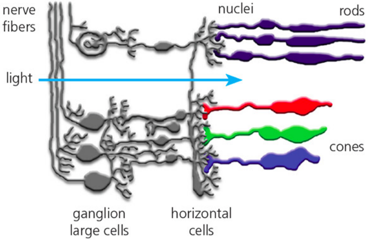
Colour throughout history

- In an article on baby clothes in the New York Times in 1893 the rule stated was that you should "always give pink to a boy and blue to a girl." In 1918 a trade publication affirmed that this was the "generally accepted rule" because pink was the "more decided and stronger color," while blue was "more delicate and dainty."
- In the West green started becoming visually associated with the devil and demonic creatures from the twelfth century, possibly as a result of the Crusades and the increasing antagonism between Christians and Muslims, for whom the color was sacred. In Shakespeare's day green costumes were considered bad luck onstage, a belief that persisted into the nineteenth century.

Source: https://www.amazon.com/Secret-Lives-Color-Kassia-Clair/dp/0143131141

The all-seeing eye

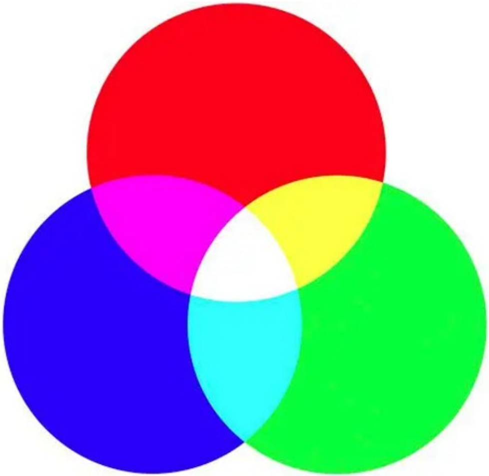




Rods can't distinguish colours - they detect changes in light and adapt the eye to help us see, even when it's dark. Cones allow us see colour. There are three kinds, each responsible for seeing red, green or blue.

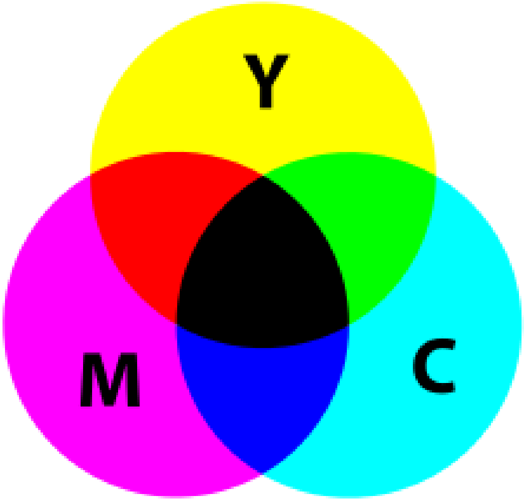
Source: https://www.xrite.com/blog/color-perception-part-3

Colour models – sRGB



The RGB colour model is an "additive" model. When 100% of each colour is mixed together, it creates white light. When 0% of each colour is combined, no light is generated, creating black. Since each colour has 256 possible values, it can also be represented using two hexadecimal values (16 x 16), as shown above. The standard way to display an RGB value is to use the hexadecimal values for red, green, and blue, preceded by a number symbol, or hashtag.

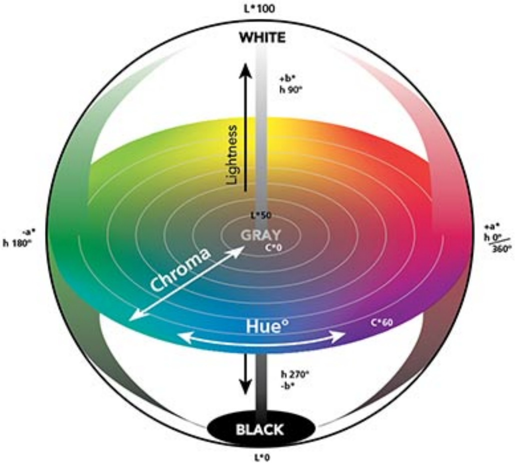
Colour models – CYMK



The CMYK colour model is a subtractive color model. CMYK refers to the four ink plates used in some colour printing: cyan, magenta, yellow, and key (black).

The CMYK model works by partially or entirely masking colours on a lighter, usually white, background. The ink reduces the light that would otherwise be reflected. Such a model is called subtractive because inks "subtract" the colours red, green and blue from white light. White light minus red leaves cyan, white light minus green leaves magenta, and white light minus blue leaves yellow.

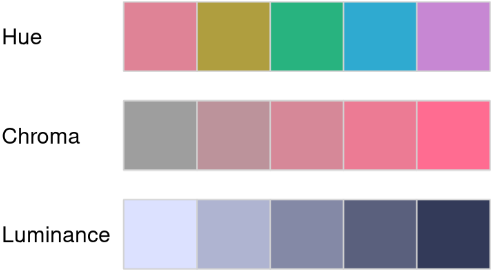
Colour models – CIE Lch



In this model, L for stands for lightness, C for chroma, and H for hue. Hue moves in a circle around the "equator" to describe the color family – red, yellow, green, and blue – and all of the colors that fall in-between. The numbers in the hue circle range from 0 to 360, starting with red at zero degrees, then moving counter-clockwise through yellow, green, blue, then back to red.

Source: https://www.xrite.com/blog/tolerancing-part-3

Colour models – LCH example



Hue: Only the hue (= type of color) changes from H = 0 (red) via 60 (yellow), etc. to 300 (purple) while chroma and luminance are fixed to moderate values of C = 60 and L = 65, respectively.
Chroma: Only the chroma (= colorfulness) changes from C = 0 (gray) to 100 (colorful) while hue and luminance are fixed to H = 0 (red) and L = 65, respectively.
Luminance: Only the luminance (= brightness) changes from L = 90 (light) to 25 (dark) while hue and chroma are fixed to H = 260 (blue) and C = 25 (low, close to gray), respectively.

Who would pay $999USD for this?

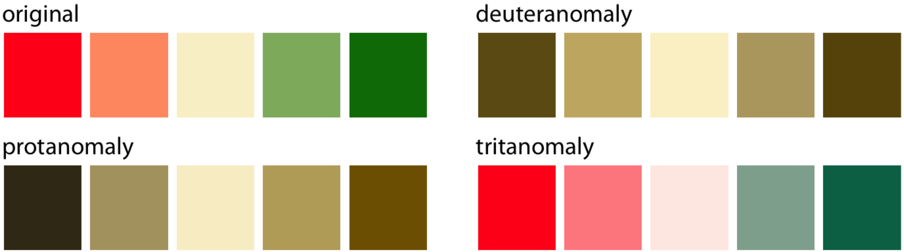
RESERVE BANK OF NEW ZEALAND
TE PŪTEA MATUA

Color Vision Deficiency

- Whenever we are choosing colors for a visualization, we need to keep in mind that a good proportion of our readers may have some form of color-vision deficiency (i.e., are colorblind).
- These readers may not be able to distinguish colors that look clearly different to most other people. People with impaired color vision are not literally unable to see any colors, however. Instead, they will typically have difficulty to distinguish certain types of colors, for example red and green (red–green color-vision deficiency) or blue and green (blue–yellow color-vision deficiency).
- Sequential scales are generally the easiest to modify for CVD, diverging scales may cause issues with popular end-point choices and qualitative scales are the most challenging.

Source: https://clauswilke.com/dataviz/color-pitfalls.html

Colour blind example



Source: https://clauswilke.com/dataviz/color-pitfalls.html

Types of basic colour palettes

- Qualitative palettes - different categories with equal prominence.
- Sequential palettes - ordinal or numeric variables (low to high).
- Diverging palettes - numeric variables with a “significant” central value
- Cyclical palettes - numeric variable with cyclical values

Qualitative colour palettes

- When you want to represent multiple categories in a plot, you typically should vary the colour of the elements… Hue is useful for representing categories: most people can distinguish a moderate number of hues relatively easily, and points that have different hues but similar brightness or intensity seem equally important.
  Source: https://seaborn.pydata.org/tutorial/color_palettes.html
- In this case, we use a qualitative color scale. Such a scale contains a finite set of specific colors that are chosen to look clearly distinct from each other while also being equivalent to each other… the colors should not create the impression of an order, as would be the case with a sequence of colors that get successively lighter.
  Source: https://clauswilke.com/dataviz/color-basics.html
- Avoid having two colors with the same hue, but different lightness and saturation, unless the values associated with those colors are related.

Source: https://medium.com/nightingale/how-to-choose-the-colors-for-your-data-visualizations-50b2557fa335

Qualitative colour palettes II

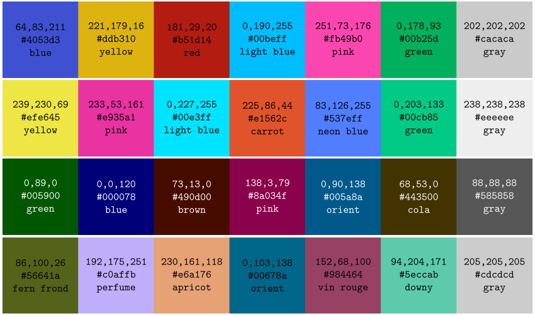
- Paul Tol - variation in hue.
  https://personal.sron.nl/~pault/
- Adobe - variation in hue.
  https://spectrum.adobe.com/page/color-for-data-visualization/
- Colorbrewer - variation in hue.
  https://colorbrewer2.org/#type=qualitative&scheme=Accent&n=3
- Okabe-Ito - variation in hue.
  https://jfly.uni-koeln.de/color/
- Tableau - variation in hue.
  https://www.tableau.com/about/blog/2016/7/colors-upgrade-tableau-10-56782

Qualitative colour palettes III

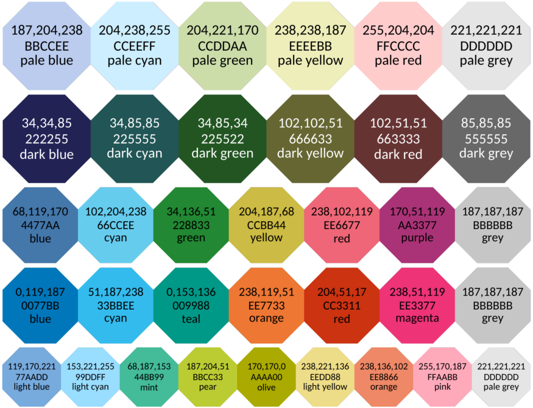
- It is best to avoid large areas of high-chroma colours in graphs which must hold a users attention for extended periods. This is because such colours tend to produce after-image effects which are can be distracting.
- If the size of the areas presented in a graph is important, then the areas should be rendered with colours of similar luminance. This is because lighter colours tend to make areas look larger than darker colours.
- When colours are used to indicate group membership, the colours should be easy to distinguish.
- There is a simple way of choosing colours which meets all the constraints above. In a perceptually uniform space, choose colours which have equal luminance and chroma and correspond to set of evenly spaced hues. <u>This is not great for color-blind/greyscale.</u>

Source: https://www.r-project.org/conferences/DSC-2003/Proceedings/Ihaka.pdf

Qualitative palette examples



Source: http://tsitsul.in/blog/coloropt/



Source: https://personal.sron.nl/~pault/

RESERVE BANK OF NEW ZEALAND
TE PŪTEA MATUA

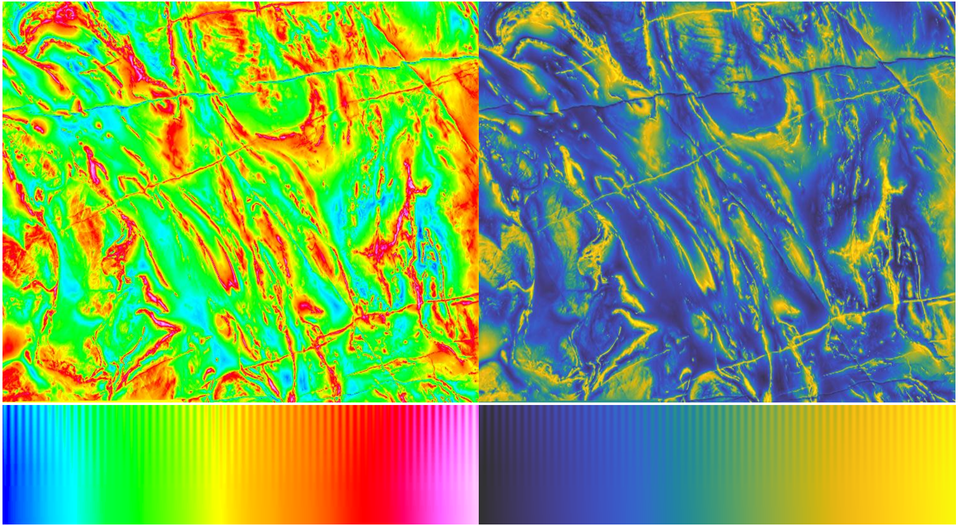
Helmholtz–Kohlrausch effect

Perceptual phenomenon wherein the intense saturation of spectral hue is perceived as part of the color's luminance.

Source: https://en.wikipedia.org/wiki/Helmholtz%E2%80%93Kohlrausch_effect

RESERVE BANK OF NEW ZEALAND
TE PŪTEA MATUA

Sequential palettes?



Source: https://colorcet.com/

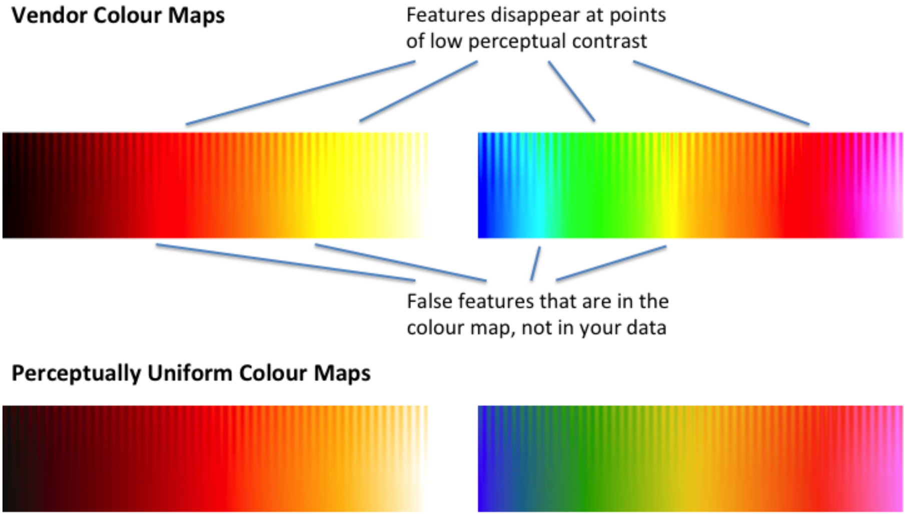
Sequential palettes?

- The previous slide shows an aeromagnetic image of the Yilgarn region in Western Australia. The left side shows the default colour map of a widely used geophysics package. On the right is the same data rendered with the perceptually uniform colour map CET-L20.

- The left map exhibits a large perceptual dead zone at green, a smaller dead zone at red, and false features at cyan and yellow.

Source: https://colorcet.com/

Sequential palettes?



Source: https://colorcet.com/

Grayscale?

- One possible solution for representing data is to simply switch to grayscale colormaps. Grayscale maps avoid the issues associated with color perception and can provide linear luminance with relation to underlying values.
- However, grayscale suffers from a small dynamic range due to the lower discernibility of the shades of gray to the human visual system. Humans without CVD can distinguish around ten million different colors versus only about thirty shades of gray.
- Even limited color vision can make use of the colorspace available from non-grayscale colormaps when they are optimized correctly. This makes color an indispensable component, as it enables the ability to see subtle changes in the underlying data (i.e. increases our visual perception precision due to the larger dynamic range).

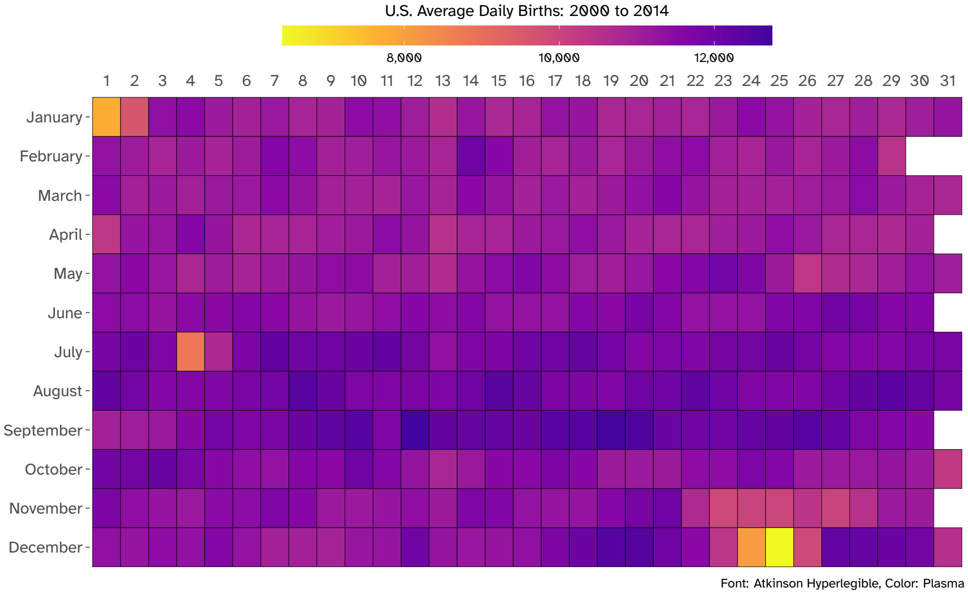
Source: https://arxiv.org/ftp/arxiv/papers/1712/1712.01662.pdf

Sequential palette design

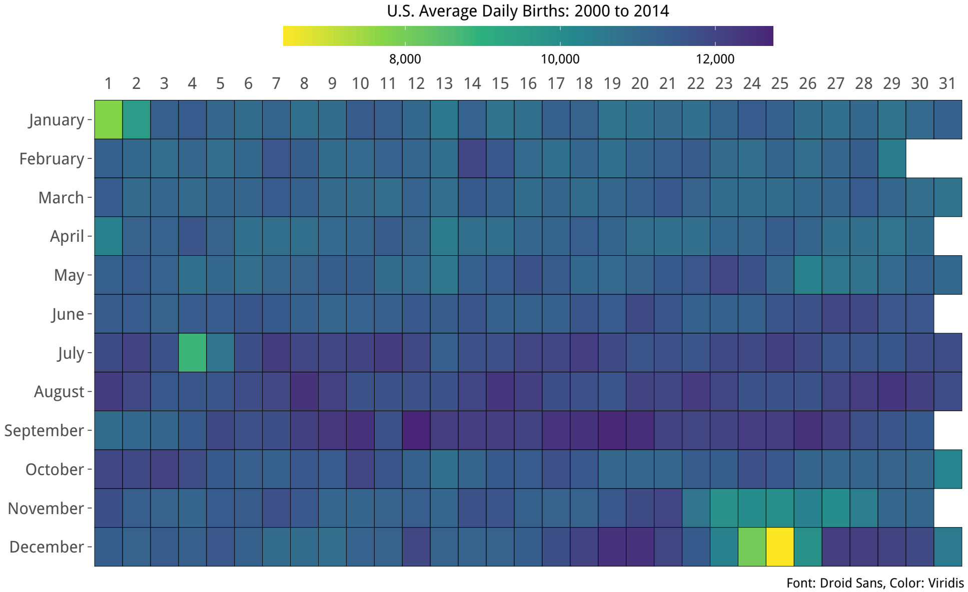
- Perceptually uniform – The distinguishing feature of these colour maps is that the lightness values vary in a linear manner even though the colour map path itself may be curved. This linear variation of lightness induces a clear ordering of colours making interpretation of data straightforward.
- It can be useful to constrain the lightness values to, say, 10 to 95 rather than using the full range of 0 to 100. The overall image contrast will be reduced slightly but the ability to identify features in the data may be better.
- To minimise the effect of scale a colour map path should be dominated either by lightness changes or by chromatic changes, but not a mixture of both. We want to avoid any perceptual contrast equalization process requiring luminance contrast to be played off against chromatic contrast over different sections of the colour map

Source: https://arxiv.org/pdf/1509.03700.pdf

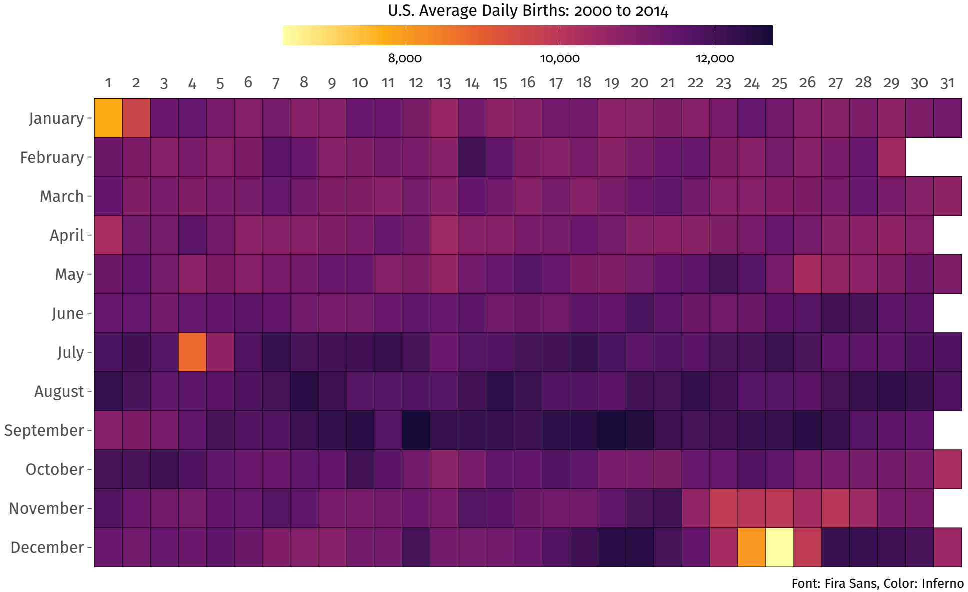
Colour scales – practical examples



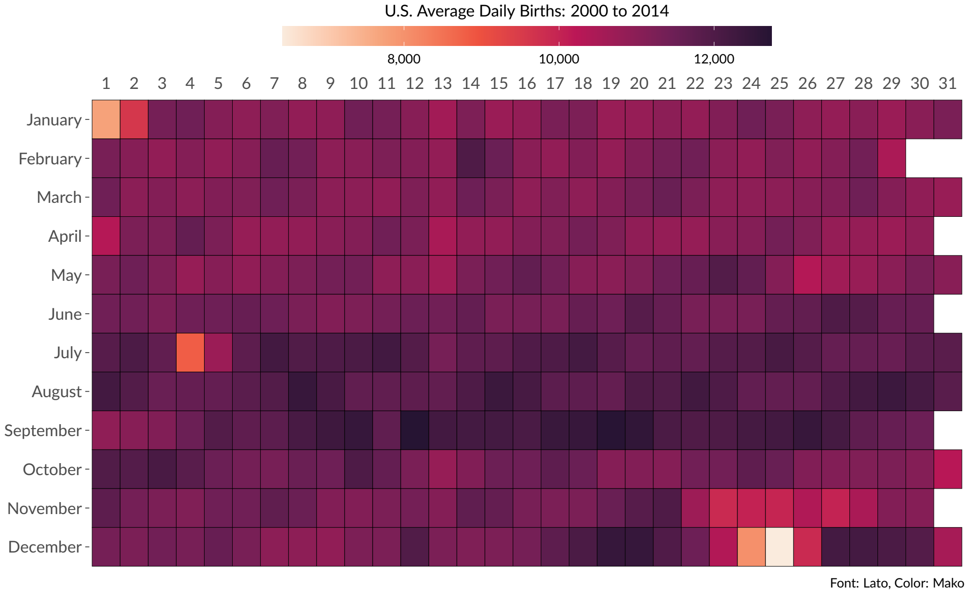
Colour scales – practical examples



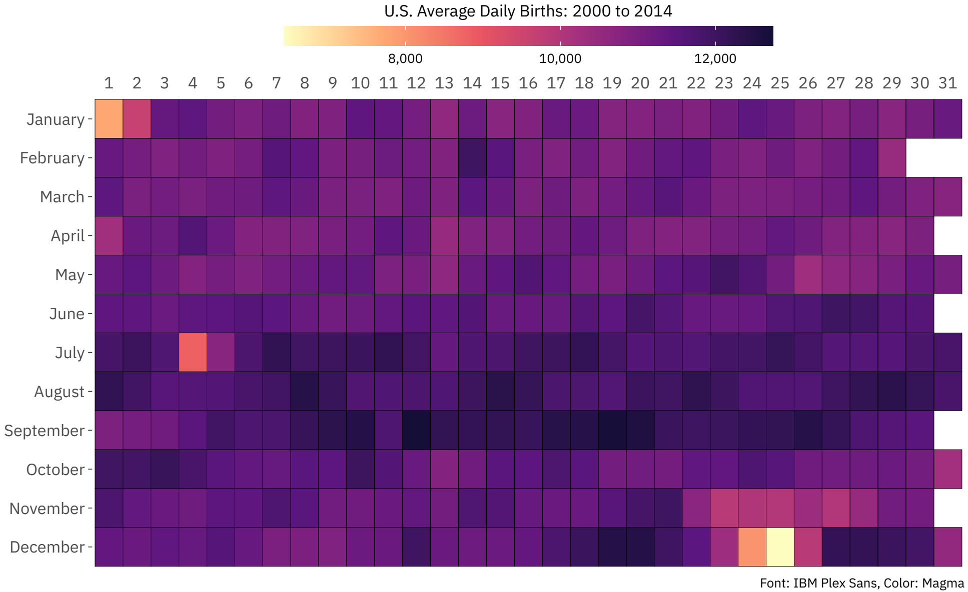
Colour scales – practical examples



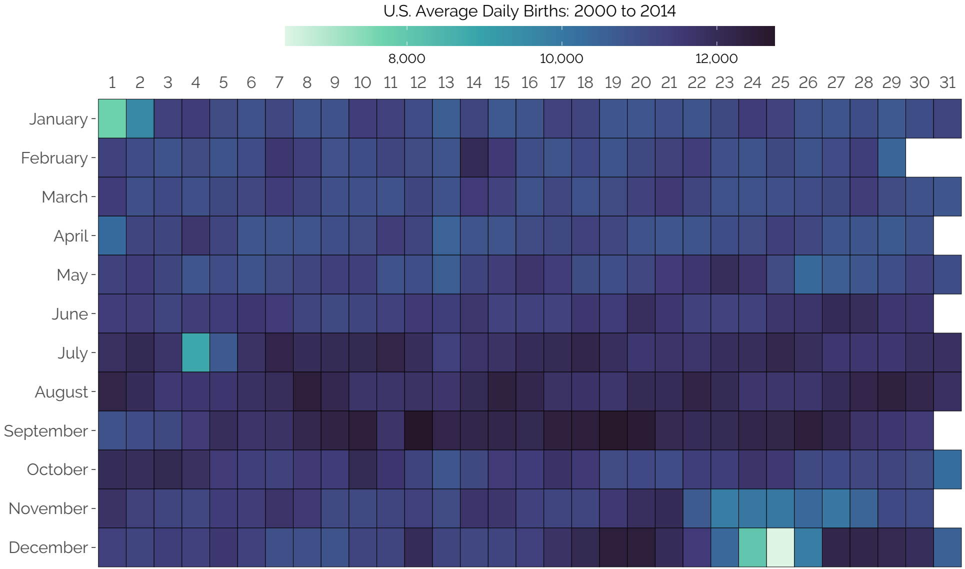
Colour scales – practical examples



Colour scales – practical examples

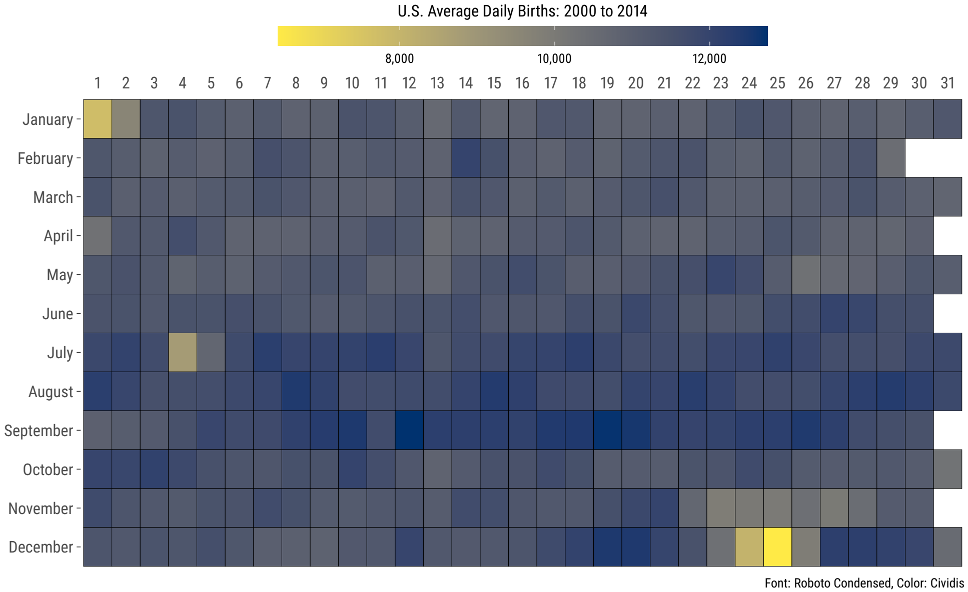


Colour scales – practical examples

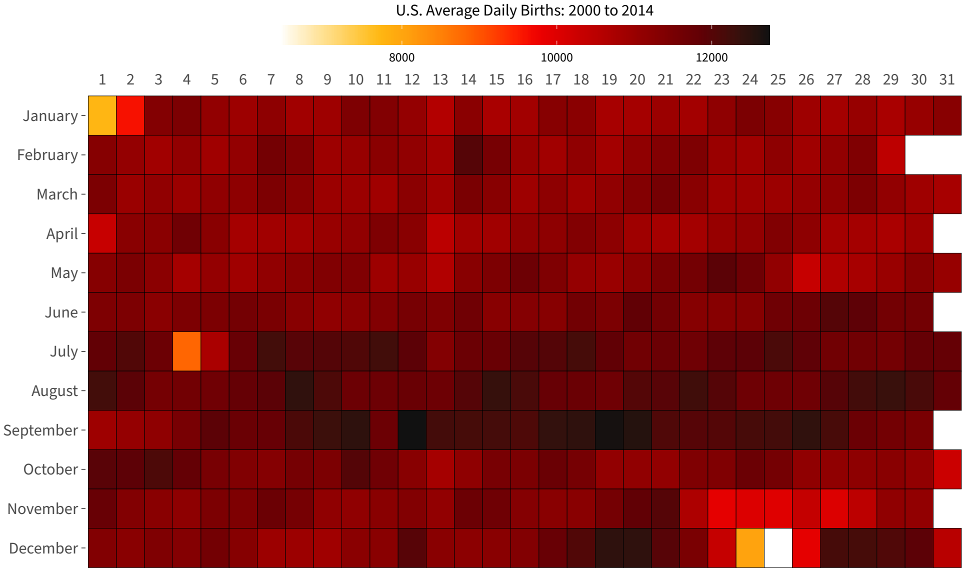


Font: Raleway, Color: Rocket

Colour scales – practical examples

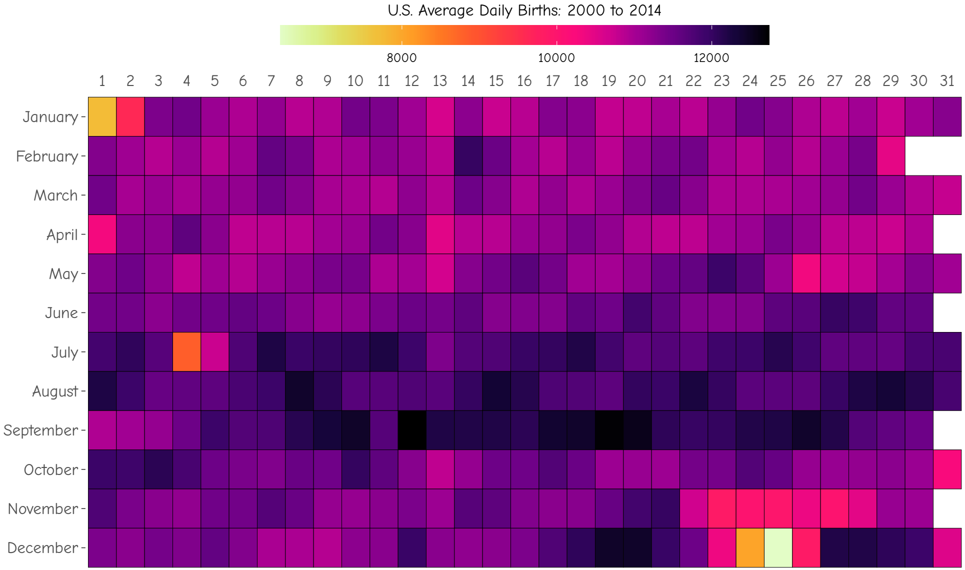


Colour scales – practical examples



Font: Source Sans Pro, Color: Fire

Colour scales – practical examples



Font: Comic Sans, Color: Cubehelix

Diverging palette example

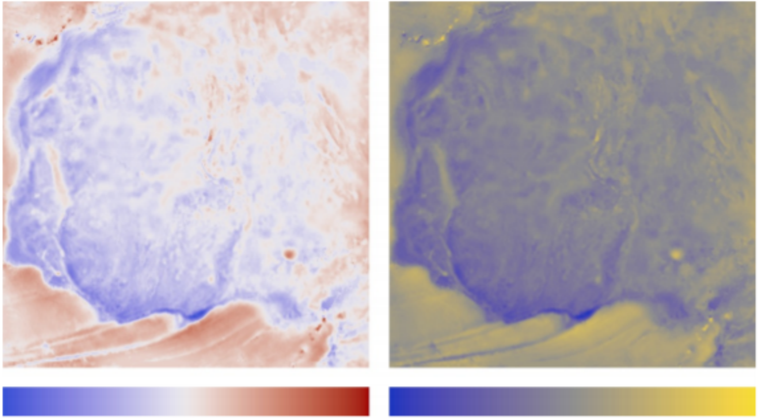


Figure 10: Residual gravity data of West Africa displayed with a red-white-blue colour map and a blue-grey-yellow linear-diverging map. Note the easier interpretation of features near zero with the blue-grey-yellow map.

Source: https://arxiv.org/pdf/1509.03700.pdf

Diverging palette design

- Diverging colour maps are intended for the display of data having a well defined reference value where we are interested in differentiating values that lie above, or below, this reference point. Within the colour map the reference value is typically denoted by a neutral colour, white, black or grey.
- A lightness gradient reversal will induce a perceptual flat spot even if no smoothing is applied. Structures in the data with values that straddle the central reference point in the map will be represented by colours that are effectively isoluminant. Thus, structures in this data range will be harder to resolve.
- To ensure perceptual symmetry the end points of a diverging map should have the same chroma, and if the lightness values reverse then the end points should also have the same level of lightness.

Source: https://arxiv.org/pdf/1509.03700.pdf

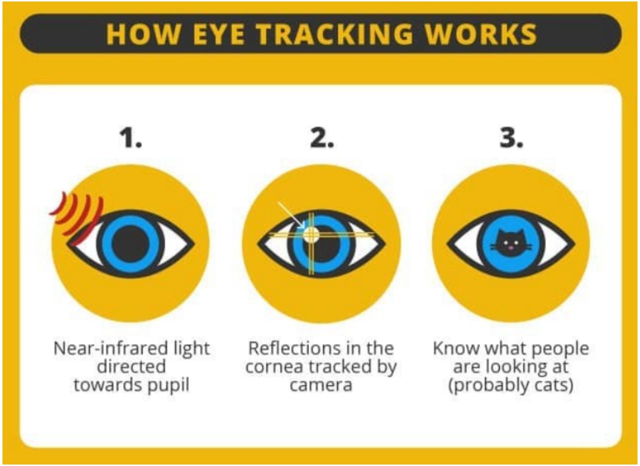
Semantically-resonant colours

- Colors are charged with rich associations. Concepts can invoke colors, and vice versa. Common associations in the United States include **bananas ↔ yellow**, **anger ↔ red**, and **money ↔ green**. These associations may be grounded in the physical appearance of objects, common metaphors, or other linguistic or cultural conventions.
- Semantically-resonant colours may improve chart reading by aiding understanding through semantic facilitation, i.e., they may allow people to use more automated pathways to process value-color associations and require less conscious thought.
- Using resonant colors may improve memory. Practically, improved recognition of category values may reduce the need to consult a legend and may promote future recall.

Source: https://idl.cs.washington.edu/files/2013-SemanticColor-EuroVis.pdf

Semantic colour associations

Eye tracking



Source: https://www.grahamjones.co.uk/2016/multimedia/infographics/the-basics-of-eye-tracking.html

Eye tracking example

In the six seconds they spend on a resume, recruiters focus on name, current and past position titles and dates, and education.

Source: https://www.linkedin.com/pulse/eye-tracking-study-among-recruiters-how-pass-initial-resume-eekman

Eye tracking design recommendations

- Charts are usually better for complex tasks when very large amounts of data are involved. Charts also attract more attention from the reader, and readers remember them better.
- The best report format sometimes depends on the personality of the reader:
    - Conscientious people prefer tables, yet they can perform equally well with both tables and charts. They usually take longer to analyze visual aids because they tend to be more detail oriented and accurate.
    - Extroverted people say they want more visuals because they tend to think more broadly (“the big picture”), but they may actually perform better with tables because it forces them to look at the details.

Source: https://sfmagazine.com/post-entry/october-2017-eye-tracking-for-better-reports/

Eye tracking design recommendations

- The optimal format might also depend on culture. Our experiments show differences between Asian and Western cultures. For example, in the Asia-Pacific region, tables are often more effective while, in Europe, charts are more favored. Asians generally also prefer more colors in their reports than European readers do.
- Familiarity with the respective reporting formats has a significant influence on comprehension and the resulting ability to make decisions. For example, the research clearly showed a "reporting experience curve." Readers are able to analyze reports faster over time if the visual representation (table or chart) doesn't change. <u>Note that this is not a free pass to continue with bad visualisations – this only applies in the context of tables versus charts.</u>

Source: https://sfmagazine.com/post-entry/october-2017-eye-tracking-for-better-reports/

Storytelling with Data