



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. Colour and perception
6. Interactivity
7. Storytelling with Data – Practical Examples

INTRODUCTION

Data Visualisation

What you must learn is that these rules are no different than the rules of a computer system. Some of them can be bent. Others can be broken.

Types of data visualisation

- Scientific visualisation:
 - 1961 - Elizabeth Waldram wrote code to display radio-astronomy maps on a cathode ray tube.
 - Early 1960s - Pierre Bézier developed 3D modeling techniques for Renault car bodies.
 - 1963 - Boeing Aircraft created a film called *Vibration of an Aircraft*.
- Exploratory visualisation - identify outliers, trends and patterns in data that merit further study. See John W. Tukey (1977) *Exploratory Data Analysis* for an early example (**S**).
- Information visualisation - representing data in a visual and meaningful way so that a user can better understand it.

Source: <https://www.igi-global.com/chapter/visualization-future-technology-practices-computational/60368>

Source: https://en.wikipedia.org/wiki/Computer_graphics

Source: <https://www.interaction-design.org/literature/topics/information-visualization>

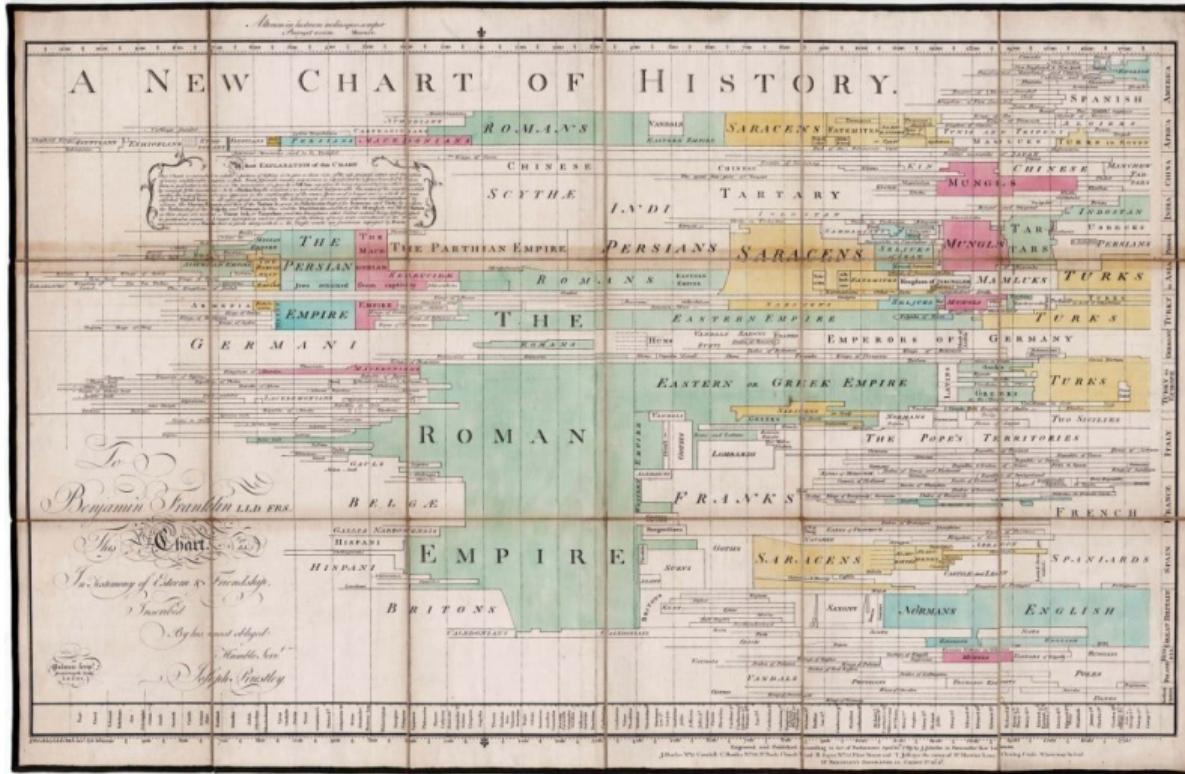
Early examples



Possibly the oldest surviving map has been engraved on this mammoth tusk, dated to 25,000 BC, found from Pavlov in the Czech Republic.

Source: https://en.wikipedia.org/wiki/History_of_cartography

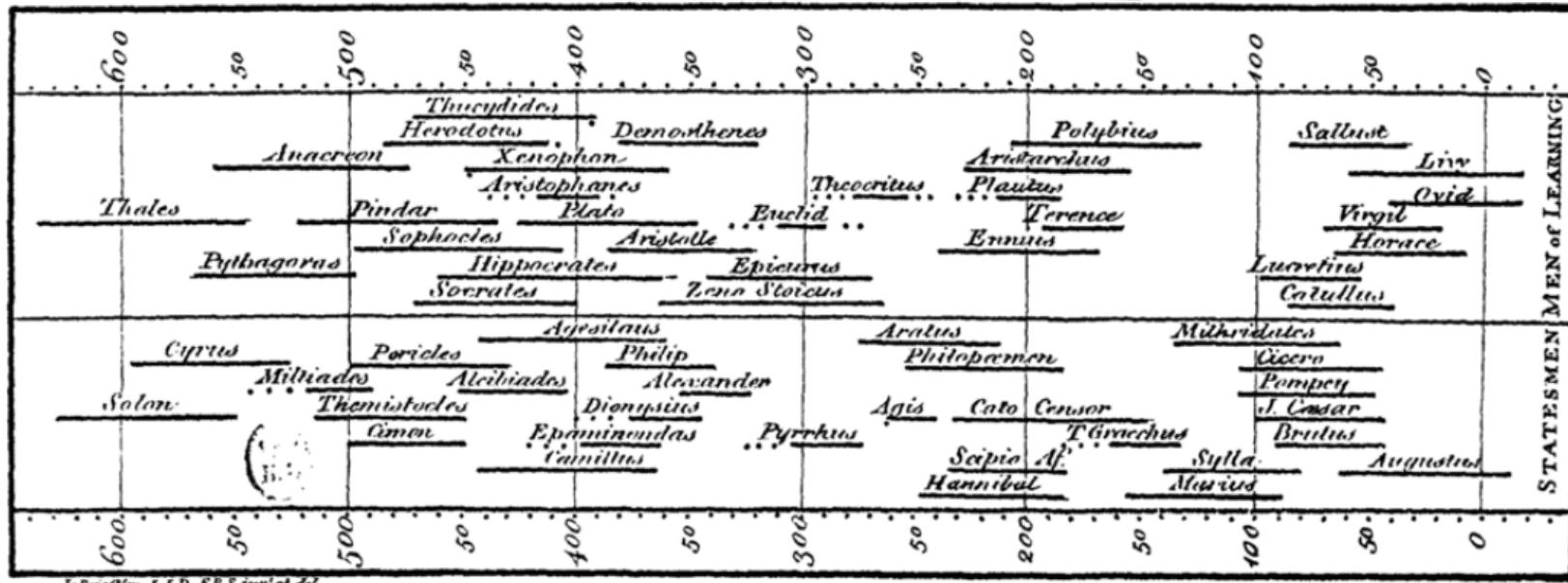
Early examples



Source: <https://bostonraremaps.com/inventory/joseph-priestley-chart-of-history/>

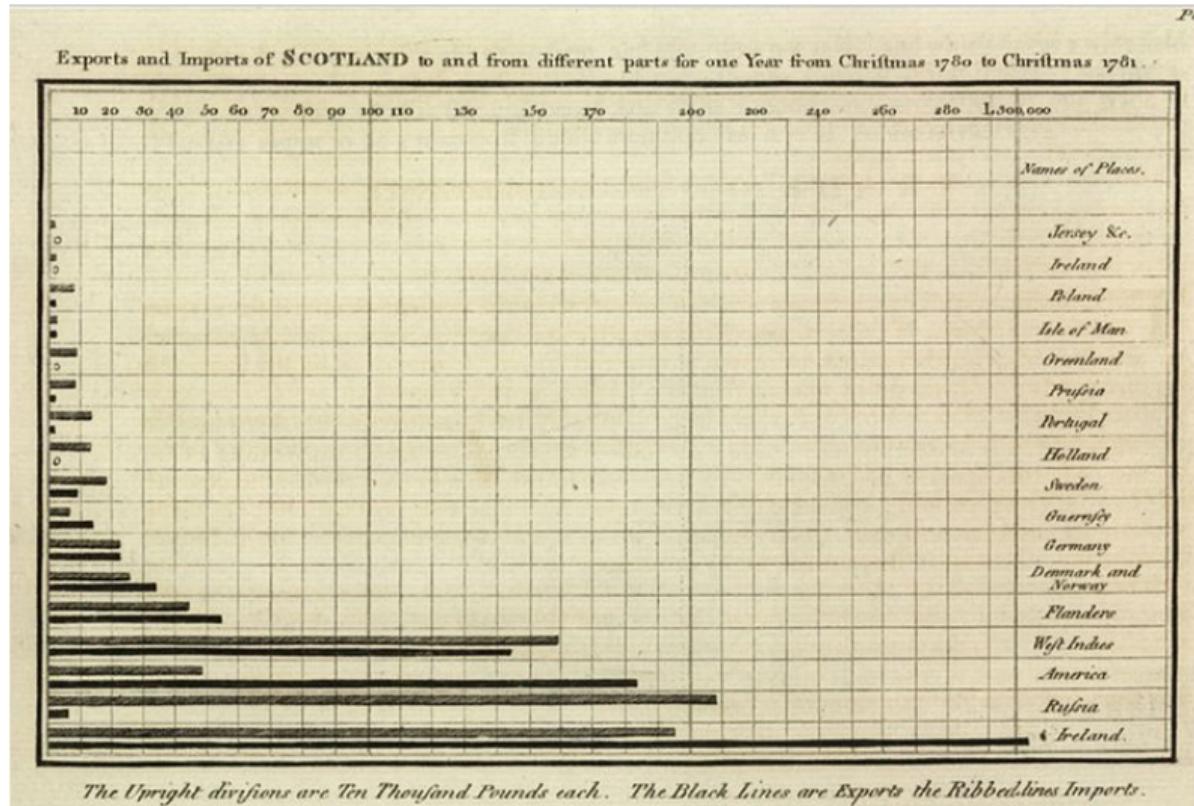
Early examples

A Specimen of a Chart of Biography.



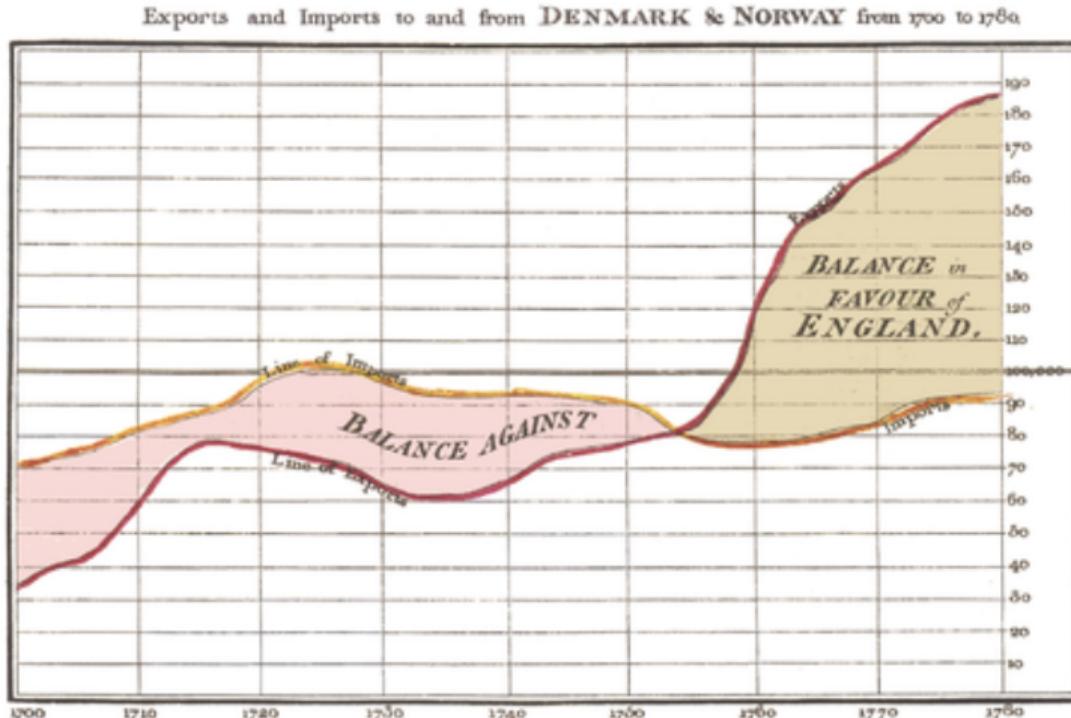
J. Priestley L.L.D. F.R.S. invent'd & draw.

Early examples



Source: https://en.wikipedia.org/wiki/William_Playfair

Early examples



Source: https://en.wikipedia.org/wiki/William_Playfair

Inspirational quotes

- The purpose of visualization is insight, not pictures.¹
- The graphic is only as useful as the audience finds it.²
- Data is personal. The best way to stakeholder's mind is through their heart³
- Data visualisation make critical information useful to those who need it most.⁴

¹ Source: <https://interactions.acm.org/blog/view/the-purpose-of-visualization-is-insight-not-pictures-an-interview-with-ben>

² Source: <https://medium.com/nightingale/data-visualization-for-audiences-in-low-middle-income-countries-ed722d161313>

³ Source: <https://medium.com/multiple-views-visualization-research-explained/data-is-personal-what-we-learned-from-42-interviews-in-rural-america-93539f25836d>

⁴ Source: <https://medium.com/nightingale/barriers-to-a-thriving-data-visualization-culture-75375a7ce4bf>

Software

If you are wondering about software for this seminar series...

Software

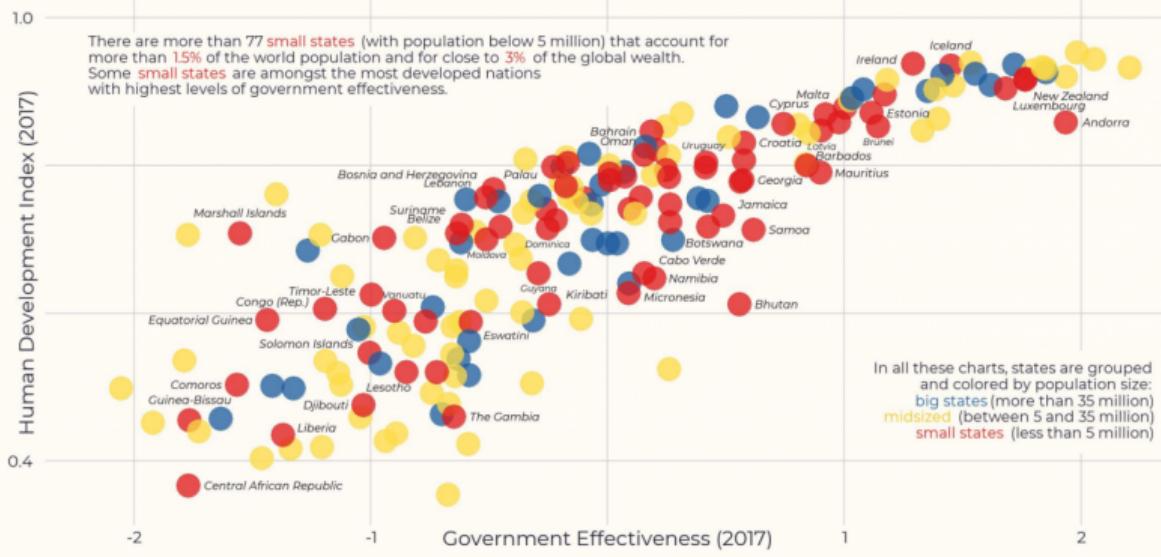
At no point do I intend to teach you how to make a specific graphic in a specific software. I don't know what software might be applicable to your needs in the future, or what visualisations you'll need to formulate when – and quite frankly, Google exists – so this isn't a cookbook with step-by-step instructions.

The goal here is not to provide you with recipes for future use, but rather to teach you what flour is – to introduce you to the basic concepts and building blocks of effective data visualisations.

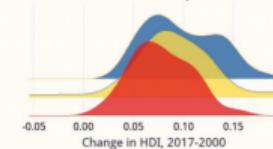
Source: <https://towardsdatascience.com/the-art-and-science-of-data-visualization-6f9d706d673e>

Good data visualisations?

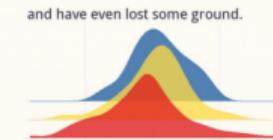
Small States Can Be Big Players in Development and Good Governance



But small states have improved less

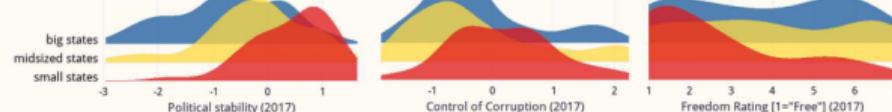


In all these charts, states are grouped and colored by population size:
big states (more than 35 million)
mid-sized (between 5 and 35 million)
small states (less than 5 million)



and have even lost some ground.

These charts show the distribution densities of different variables by three groups based on state size:

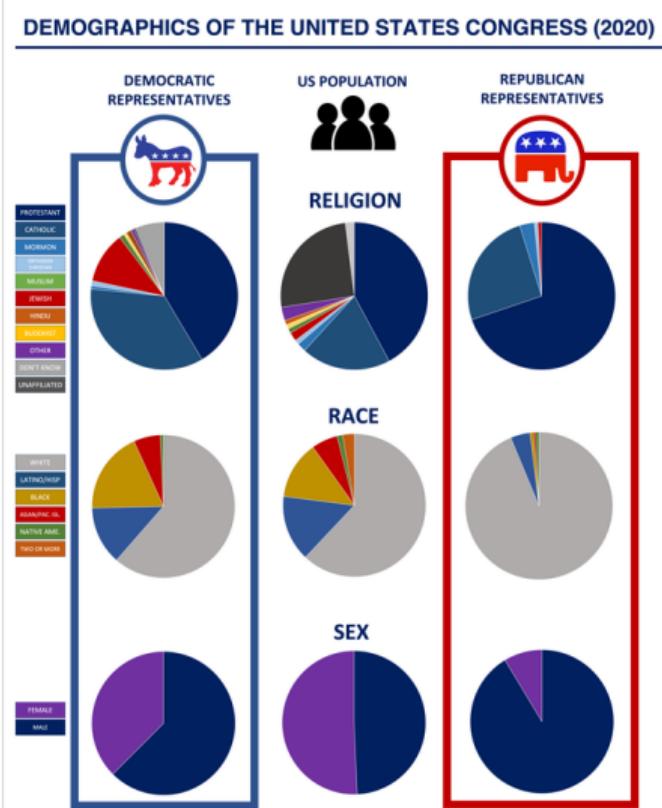


Data: Human Development Index from Human Development Reports. Government Effectiveness, Control of Corruption, and Political Stability from Worldwide Governance Indicators. Freedom Rating from Freedom House.

Source: <https://informationisbeautiful.net/2019/winners-of-the-world-data-visualization-prize/>



Good data visualisations?

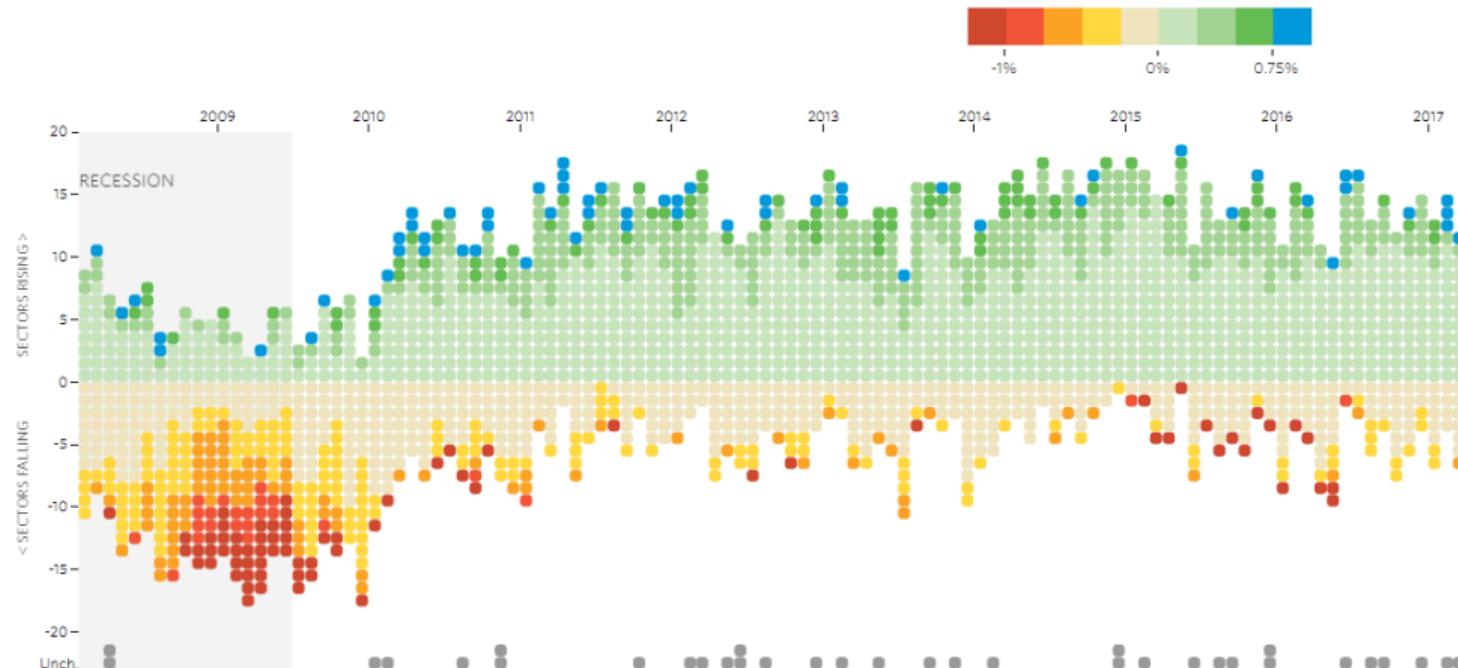


Source: <https://www.reddit.com/r/dataisbeautiful/comments/ih0046/howRepresentativeAreTheRepresentativesThe/>

Good data visualisations?

Winners and Losers: Job Gains and Losses [Jump to National Unemployment](#)

Track the number of sectors gaining or losing jobs each month. Boxes are shaded based on percentage change from the previous month in each sector's payrolls.



Source: <http://graphics.wsj.com/job-market-tracker/>

Infographics vs Data visualisations

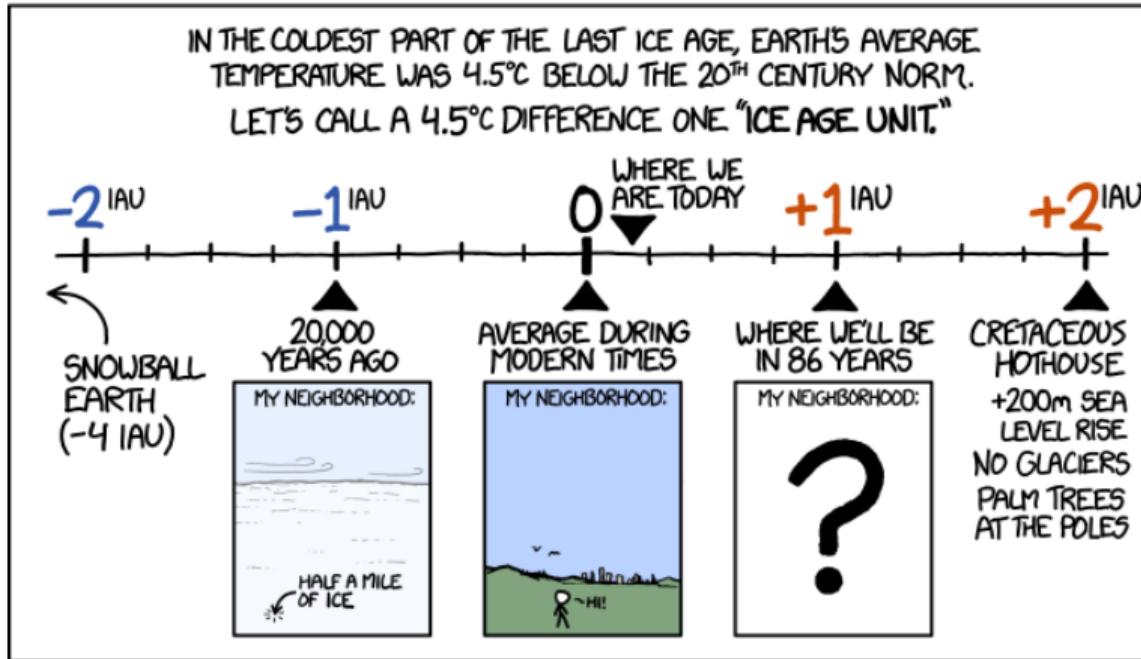
	Infographics	Data visualisations
Complexity	Generally simple	Varies
Narrative	Usually	Rarely
Meta data	Sometimes	Included, but usually external
Graphics	Generally to add visual appeal	Sometimes, considered “chartjunk”
Interactive	Rarely. Interactive infographic are called “interactive data stories”	Progressively more common
Production time	Generally quite time consuming	Typically very little time; automatically produced

Source: <https://www.statsilk.com/blog/real-difference-between-infographics-and-data-visualizations>

Data visualisation or Infographic?

WITHOUT PROMPT, AGGRESSIVE LIMITS ON CO₂ EMISSIONS, THE EARTH WILL LIKELY WARM BY AN AVERAGE OF 4°-5°C BY THE CENTURY'S END.

HOW BIG A CHANGE IS THAT?



Source: <https://xkcd.com/1379/>

Data visualisation or Infographic?



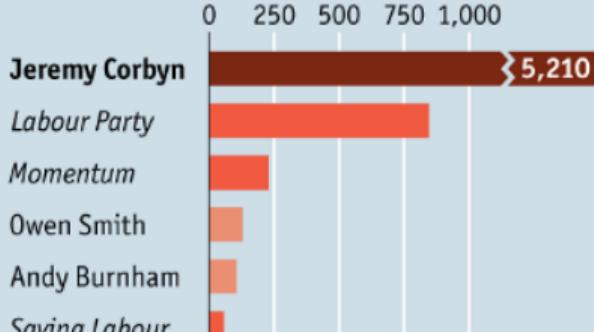
Source: <http://vis.csail.mit.edu/classes/6.859/readings/pdfs/Bateman-UsefulJunk.pdf>

Mistakes – Truncating the scale

Original

Left-click

Average number of likes per Facebook post
2016



Source: Facebook

Better

Left-click

Average number of likes per Facebook post
2016, '000



Source: Facebook

Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – Truncating the scale

- The original chart not only downplays the number of Mr Corbyn's likes but also exaggerates those on other posts. In the redesigned version, we show Mr Corbyn's bar in its entirety. All other bars remain visible.
- Another odd thing is the choice of colour. In an attempt to emulate Labour's colour scheme, we used three shades of orange/red to distinguish between Jeremy Corbyn, other MPs and parties/groups. While the logic behind the colours might be obvious to a lot of readers, it perhaps makes little sense for those less familiar with British politics.

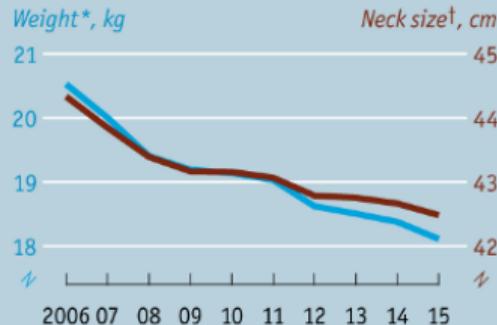
Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – Cherry-picking scales

Original

Fit as a butcher's dog

Characteristics of dogs registered with the UK's Kennel Club, average when fully grown



Sources: Kennel Club;
The Economist

*Where at least 50 are
registered per year †Where at
least 100 are registered per year

Better

Fit as a butcher's dog

Characteristics of dogs registered with the UK's Kennel Club, average when fully grown



Sources: Kennel Club;
The Economist

*Where at least 50 are registered per year
†Where at least 100 are registered per year

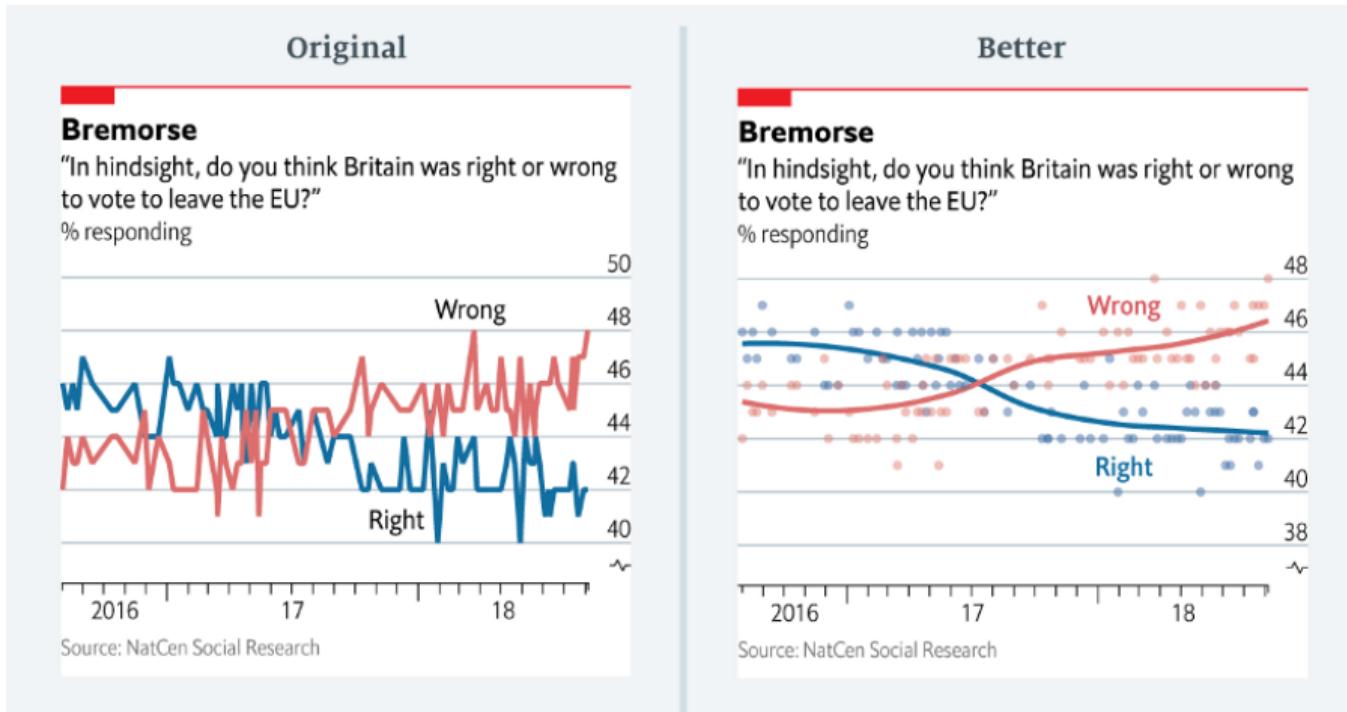
Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – Cherry-picking scales

- In the original chart, both scales decrease by three units (from 21 to 18 on the left; from 45 to 42 on the right). In percentage terms, the left scale decreases by 14% while the right goes down by 7%. In the redesigned chart, I retained the double scale but adjusted their ranges to reflect a comparable proportional change.
- Considering the jolly topic of this chart, this mistake may seem relatively minor. The message of the chart, after all, is the same in both versions. But the takeaway is important: if two series follow each other too closely, it is probably a good idea to have a closer look at the scales.

Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – wrong visualisation method



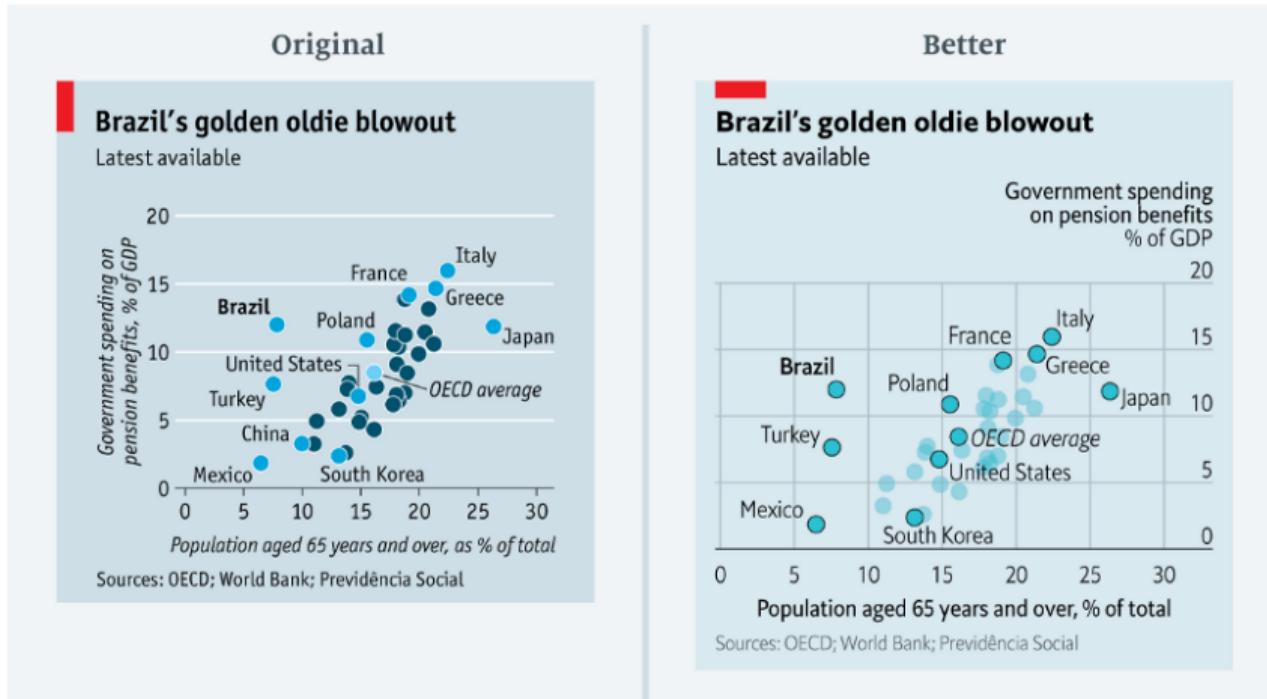
Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – wrong visualisation method

- Looking at the first chart, it appears as if respondents had a rather erratic view of the referendum result — increasing and decreasing by a couple of percentage points from one week to the next.
- Instead of plotting the individual polls with a smoothed curve to show the trend, we connected the actual values of each individual poll. This happened, because our previous in-house charting tool does not plot smoothed lines.
- In the redesigned version, there is more space between the start of the scale and the smallest data point. Francis Gagnon has put together a nice formula for this: aim for leaving at least 33% of the plot area free under a line chart that doesn't start at zero.

Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – confusing colour



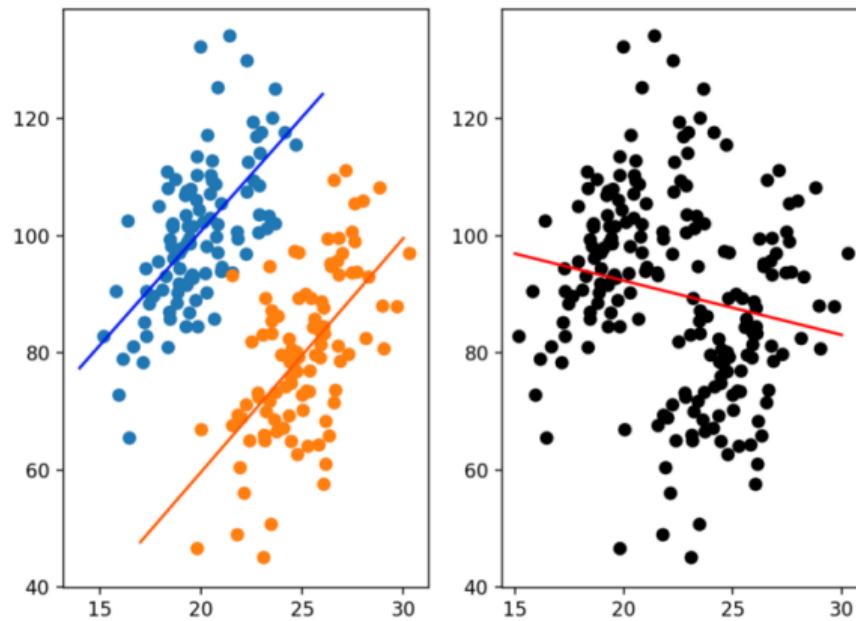
Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Mistakes – confusing colour

- To keep the chart small, the visualiser only labelled a selection of countries and highlighted those in electric blue. The OECD average is highlighted in pale blue.
- The visualiser ignored the fact that a change of colour often implies a categorical change. On first glance, this seems to be the case in this chart too — all electric blue colours seem to belong to a different grouping to the dark blue ones. This is not the case. The only thing they have in common is that they were chosen to be labelled.
- In the redesigned version, the colour of the circles remains the same for all countries. The opacity changes for those that aren't labelled to make the others stand out. Typography does the rest: Brazil, the focus country, is written in bold text, and the OECD average in italics.

Source: <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>

Simpson's paradox



Source: <https://towardsdatascience.com/what-is-simpsons-paradox-4a53cd4e9ee2>

More resources

- Storytelling with Data - <http://www.storytellingwithdata.com/blog>
- Flowing Data - <http://flowingdata.com/>
- Visualising Data - <http://www.visualisingdata.com/>
- Junk Charts - <http://junkcharts.typepad.com/>
- The Pudding - <https://pudding.cool/>
- The Atlas - <https://www.theatlas.com/>
- Graphic Detail - <https://www.economist.com/blogs/graphicdetail>
- Tableau Blog - <https://www.tableau.com/about/blog>
- FiveThirtyEight - <http://fivethirtyeight.com/>

Source: <https://www.tableau.com/learn/articles/best-data-visualization-blogs>

What is this based on? Part 1

New York University – Certificate in Data Visualisation

Source: <https://www.sps.nyu.edu/professional-pathways/certificates/technology/data-visualization.html>

1. Statistical Foundations of Data Visualization
2. Data Visualization for Business
3. Interactive Data Visualization
4. The Art of Data Visualization
5. Visual Analytics with Tableau
6. Designing Infographics

What is this based on? Part 2

Harvard University – Communicating Data and Complex Ideas Visually

Source: <https://online-learning.harvard.edu/course/data-visualization-communicating-data-and-complex-ideas-visually?delta=1>

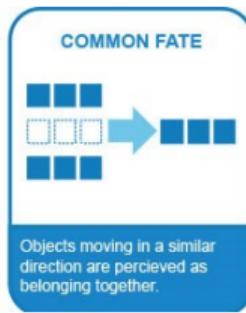
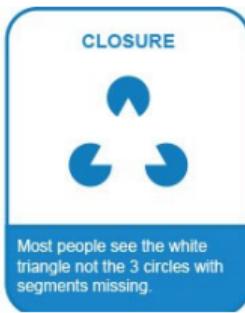
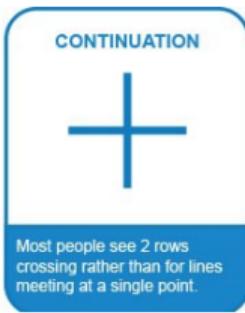
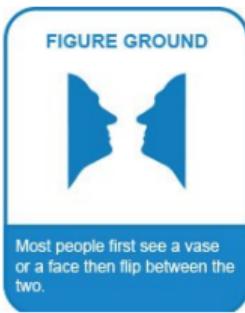
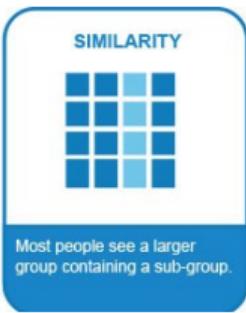
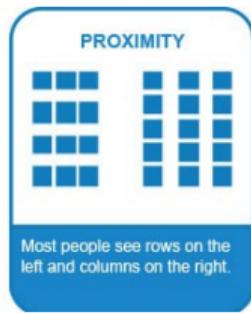
1. Critically evaluate visualizations and suggest improvements and refinements
2. Avoid creating misleading visual representations of data—and being misled by others
3. Apply a structured design process to create effective visualizations
4. Conceptualize ideas using sketching and prototyping
5. Use principles of perception and cognition in visualization design
6. Learn how to tell data stories with visualizations and communicate more precisely by pinpointing the most relevant information

FUNDAMENTALS

Data Visualisation

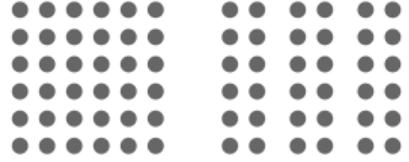
We don't rise to the level of our expectations, we fall to the level of our training.

Gestalt principles of design



Source: <https://bloggotype.blogspot.com/2016/08/holiday-notes2-grammar-of-graphics.html>

Proximity



- Proximity refers to how close elements are to one another. The strongest proximity relationships are those between overlapping subjects, but just grouping objects into a single area can also have a strong proximity effect.
- The opposite is also true, of course. By putting space between elements, you can add separation even when their other characteristics are the same.

Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>

Similarity



- It's human nature to group like things together. In gestalt, similar elements are visually grouped, regardless of their proximity to each other. They can be grouped by color, shape, or size. Similarity can be used to tie together elements that might not be right next to each other in a design.
- Of course, you can make things dissimilar if you want to make them stand out from the crowd. It's why buttons for calls to action are often designed in a different color than the rest of a page – so they stand out and draw the visitor's attention to the desired action.

Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>

Figure/ground



- The figure/ground principle takes advantage of the way the brain processes negative space. You've probably seen examples of this principle floating around in memes on social media, or as part of logos (like the arrow in the FedEx logo).
- Your brain will distinguish between the objects it considers to be in the foreground of an image (the figure, or focal point) and the background (the area on which the figures rest). Where things get interesting is when the foreground and background actually contain two distinct images, like the example in the top corner of this slide.
- In general terms, your brain will interpret the larger area of an image as the ground and the smaller as the figure. You can see that lighter and darker colors can influence what is viewed as the figure and what is viewed as the ground.

Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>

Continuation



- The law of continuity posits that the human eye will follow the smoothest path when viewing lines, regardless of how the lines were actually drawn.
- This continuation can be a valuable tool when the goal is to guide a visitor's eye in a certain direction. They will follow the simplest path on the page, so make sure the most vital parts they should see fall within that path.
- Since the eye naturally follows a line, placing items in a series in a line will naturally draw the eye from one item to the next. Horizontal sliders are one such example, as are related product listings on sites like Amazon.

Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>

Closure



- Closure is the idea that your brain will fill in the missing parts of a design or image to create a whole.
- In its simplest form, the principle of closure allows your eye to follow something like a dotted line to its end. But more complex applications are often seen in logos, like that for the World Wildlife Fund. Large chunks of the outline for the panda are missing, but your brain has no problem filling in the missing sections to see the whole animal.

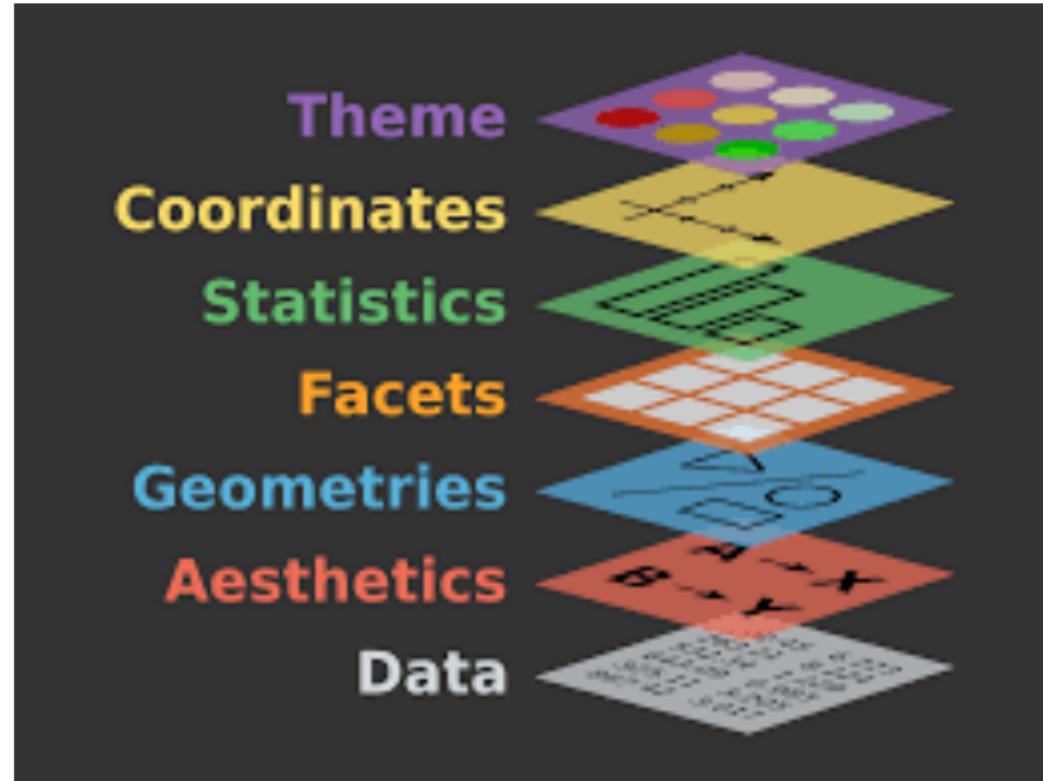
Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>

Common fate

- Humans tend to perceive elements moving in the same direction as being more related than elements that are stationary or that move in different directions.
- The principle of common fate serves us just as well by how the world around us conflicts with it as it does by how the world around us follows in accordance to it.
- The clear patterns of the synchronized movement of vehicle groups on the road inform us strongly on how to blend harmoniously with them. But just as important, things that conflict with our vehicular common fate capture our attention.

Source: <https://www.andyrutledge.com/common-fate.html>

Grammar of graphics



Source: <https://bloggotype.blogspot.com/2016/08/holiday-notes2-grammar-of-graphics.html>

Grammar of graphics



Source: <https://rfortherestofus.com/2019/07/a-short-overview-of-the-grammar-of-graphics/>

Grammar of graphics

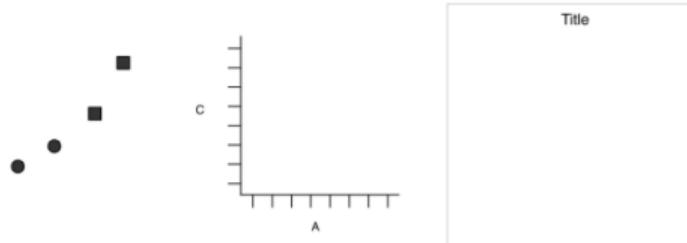


Figure 1. Graphics objects produced by (from left to right): geometric objects, scales and coordinate system, plot annotations.

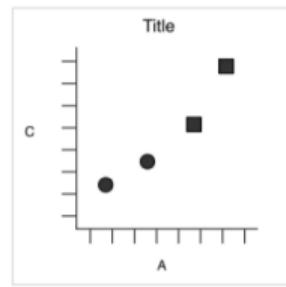


Figure 2. The final graphic, produced by combining the pieces in Figure 1.

Source: <https://vita.had.co.nz/papers/layered-grammar.pdf>

Insight needs

Table 1. Typology of the DVL-FW

Insight needs	Data scales	Analyses	Visualizations	Graphic symbols	Graphic variables
Categorize/cluster	Nominal	Statistical	Table	Geometric symbols	Spatial
Order, rank, sort	Ordinal	Temporal	Chart	Point	Position
Distributions (also outliers)	Interval	Geospatial	Graph	Line	Retinal
Comparisons	Ratio	Topical	Map	Area	Form
Trends (process and time)		Relational	Tree	Surface	Color
Geospatial			Network	Volume	Optics
Compositions (also of text)				Linguistic symbols	Motion
Correlations/relationships				Text Numerals Punctuation marks Pictorial symbols Images Icons Statistical glyphs	

Source: <https://www.pnas.org/content/pnas/116/6/1857.full.pdf>

Tools to present insights

		Geometric Symbols		Linguistic Symbols	Pictoral Symbols	
		Point	Line			
Spatial	Position	X Y	 			
	From	Size	  	  	  	
	Shape	  	  	  	  	
	Retinal	Color	Value	  	  	  
		Hue	  	  	  	  
		Saturation	  	  	  	  
	Texture	Granularity	       	  		
		Pattern	        	  		
	Optics	Blur	  	  	  	
Motion	Speed	  	  	  	  	

Fig. 3. Four graphic symbols and 11 graphic variables from full 11 graphic symbols by 24 graphic variables set in ref. 34. Qualitative nominal variables (shape, color hue, and pattern) have a gray mark.

Fonts – X-height

X-height is the height of the body of the lower case letters minus any ascenders or descenders. Because lowercase letters sometimes have different heights, this height is measured using the letter “x”.

The x-height of a typeface affects readability at small sizes.

Gill Sans 10px

The x-height of a typeface affects readability at small sizes.

Athelas 10px

The x-height of a typeface affects readability at small sizes.

Open Sans 10px

The x-height of a typeface affects readability at small sizes.

Noto Sans 10px

The x-height of a typeface affects readability at small sizes.

Lato 10px

When choosing a font for data visualization, opt for a typeface with a large x-height.

Source: <https://medium.com/nightingale/choosing-a-font-for-your-data-visualization-2ed37afea637>

Fonts – Counter

The counter is the empty space in letters such as “p” and “o”.

The shape of the counter affects readability at small sizes.

Marker Felt 10px

The shape of the counter affects readability at small sizes.

League Gothic 10px

The shape of the counter affects readability at small sizes.

Open Sans 10px

The shape of the counter affects readability at small sizes.

Futura Condensed 10px

The shape of the counter affects readability at small sizes.

Lato 10px

The shape of the counter affects readability at small sizes. Your viewer should not have to spend time figuring out if the letter is an o or an e. Fonts with distorted counters render poorly in small sizes. When choosing a font for your visualization, look to use a font with a stable, open counter.

Source: <https://medium.com/nightingale/choosing-a-font-for-your-data-visualization-2ed37afea637>

Fonts – Serifs

Serifs are the little swooshes or feet around a letter form.

The x-height of a typeface affects readability at small sizes.

Baskerville 10px

The x-height of a typeface affects readability at small sizes.

Athelas 10px

The x-height of a typeface affects readability at small sizes.

Playfair Display 10px

The x-height of a typeface affects readability at small sizes.

Noto Sans 10px

The x-height of a typeface affects readability at small sizes.

Lato 10px

Serifs are often avoided in labeling content because the extra flourishes sometimes muddle the view of the letterform. While some serif fonts are passable at smaller sizes, if you are unsure, then skip them and opt to use a sans-serif typeface.

Source: <https://medium.com/nightingale/choosing-a-font-for-your-data-visualization-2ed37afea637>

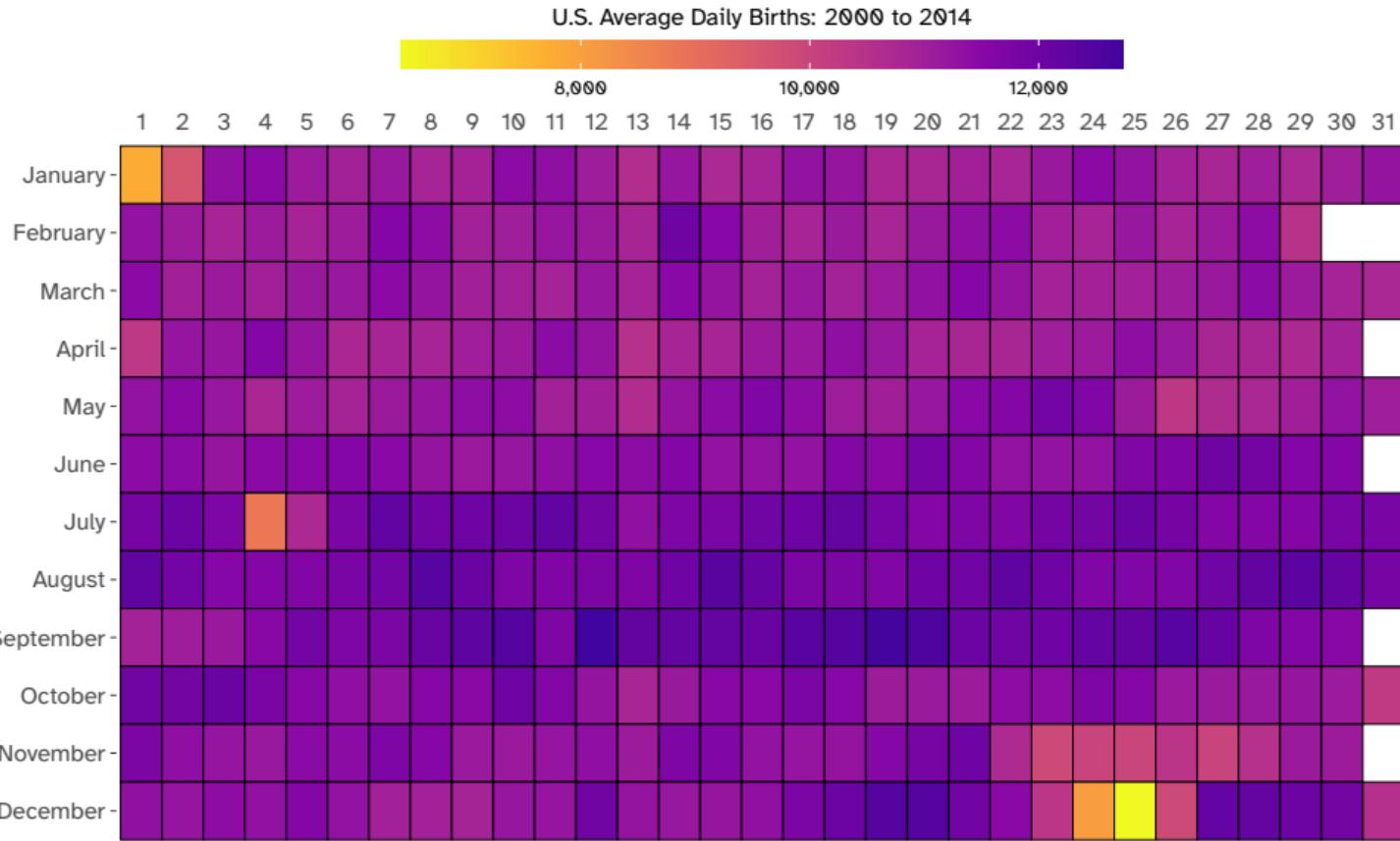
Fonts – accessibility

Key questions:

- Is there a difference between capital I, lowercase l, and the number 1?
E.g. Illinois.
- Compare letters b and d, p and q — are they mirror images or distinguished?
- Compare letters g, a, and o — are they distinguished?
- Do the letters rn look like the letter m?

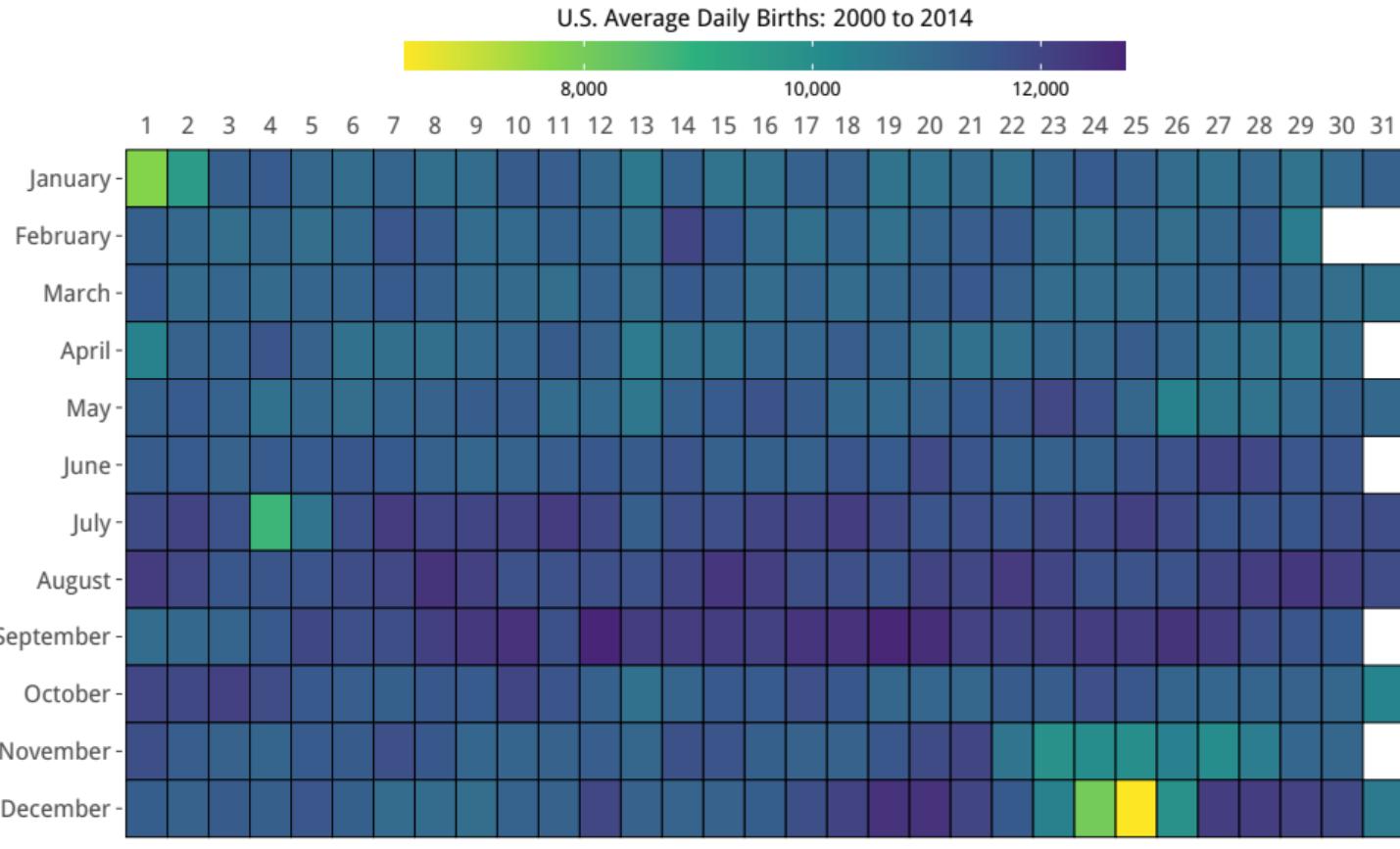
Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Fonts – practical examples



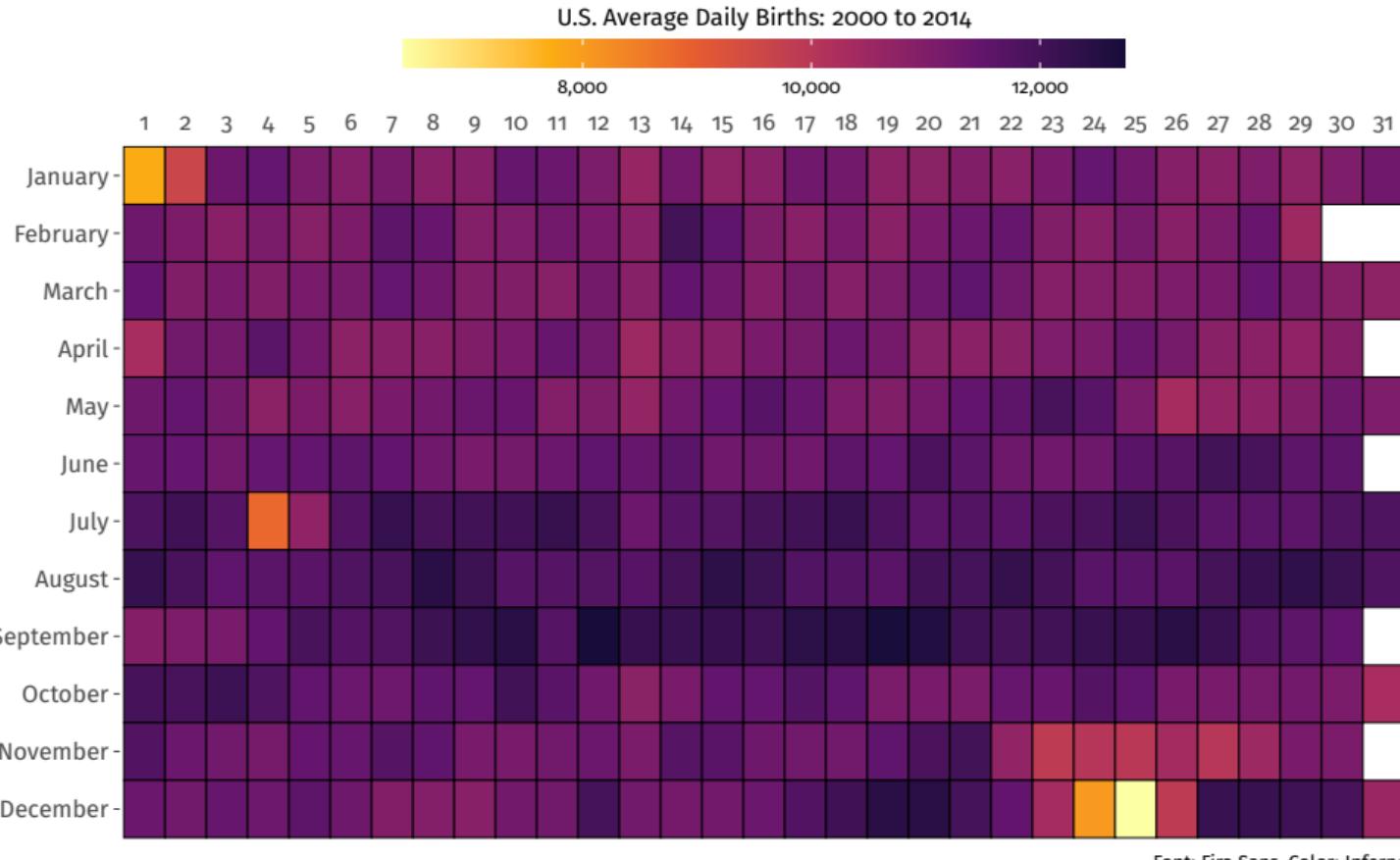
Font: Atkinson Hyperlegible, Color: Plasma

Fonts – practical examples



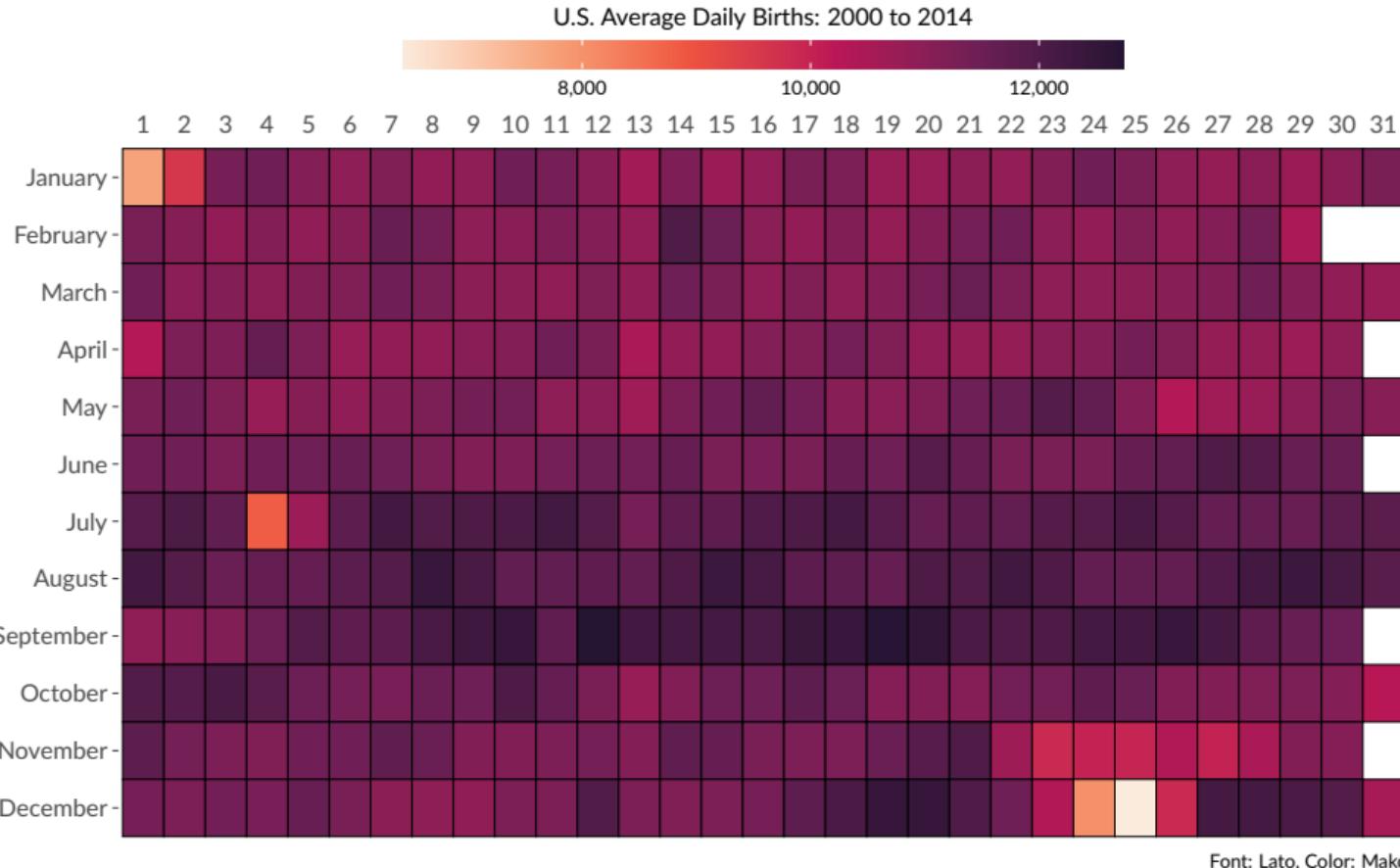
Font: Droid Sans, Color: Viridis

Fonts – practical examples

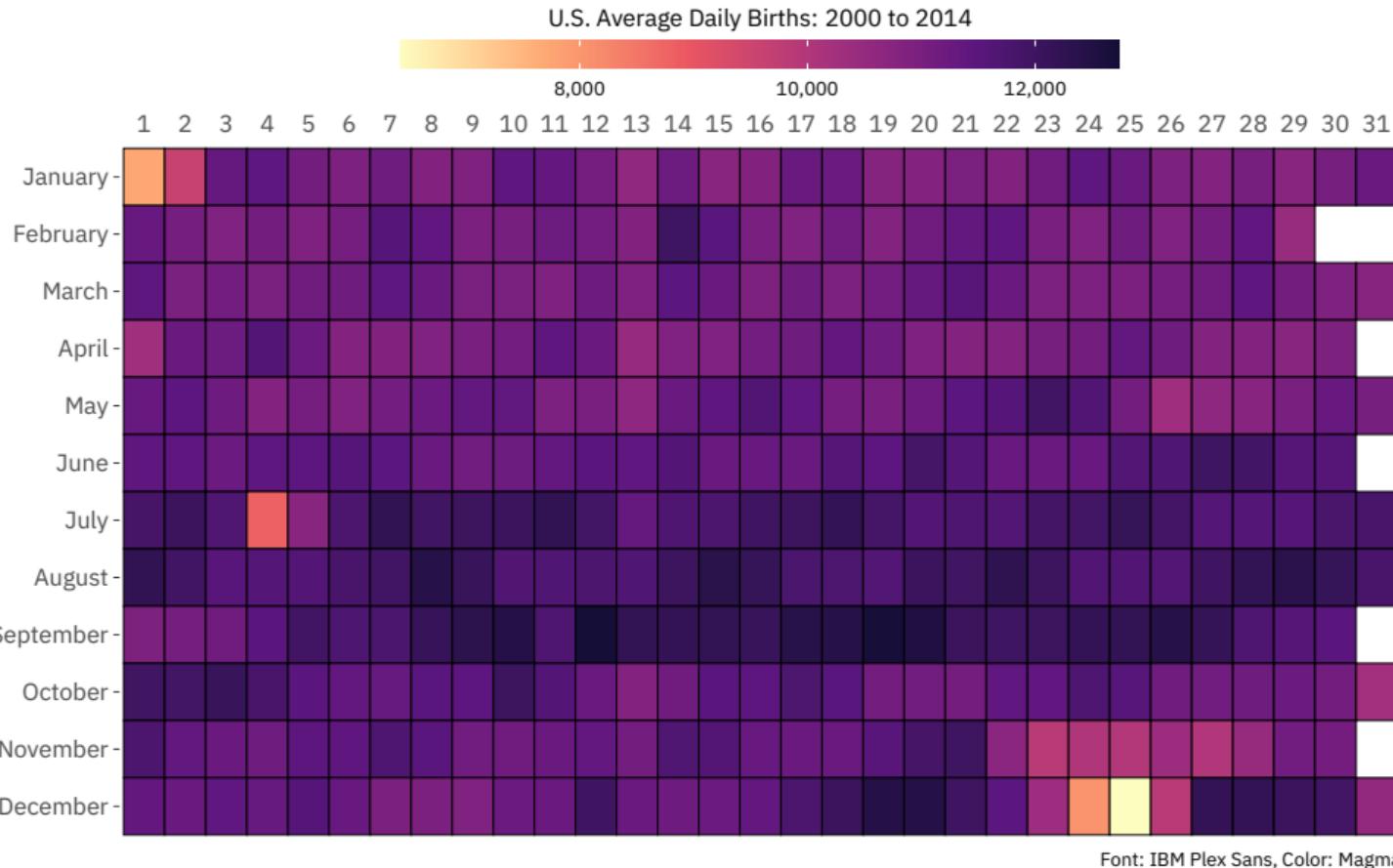


Font: Fira Sans, Color: Inferno

Fonts – practical examples

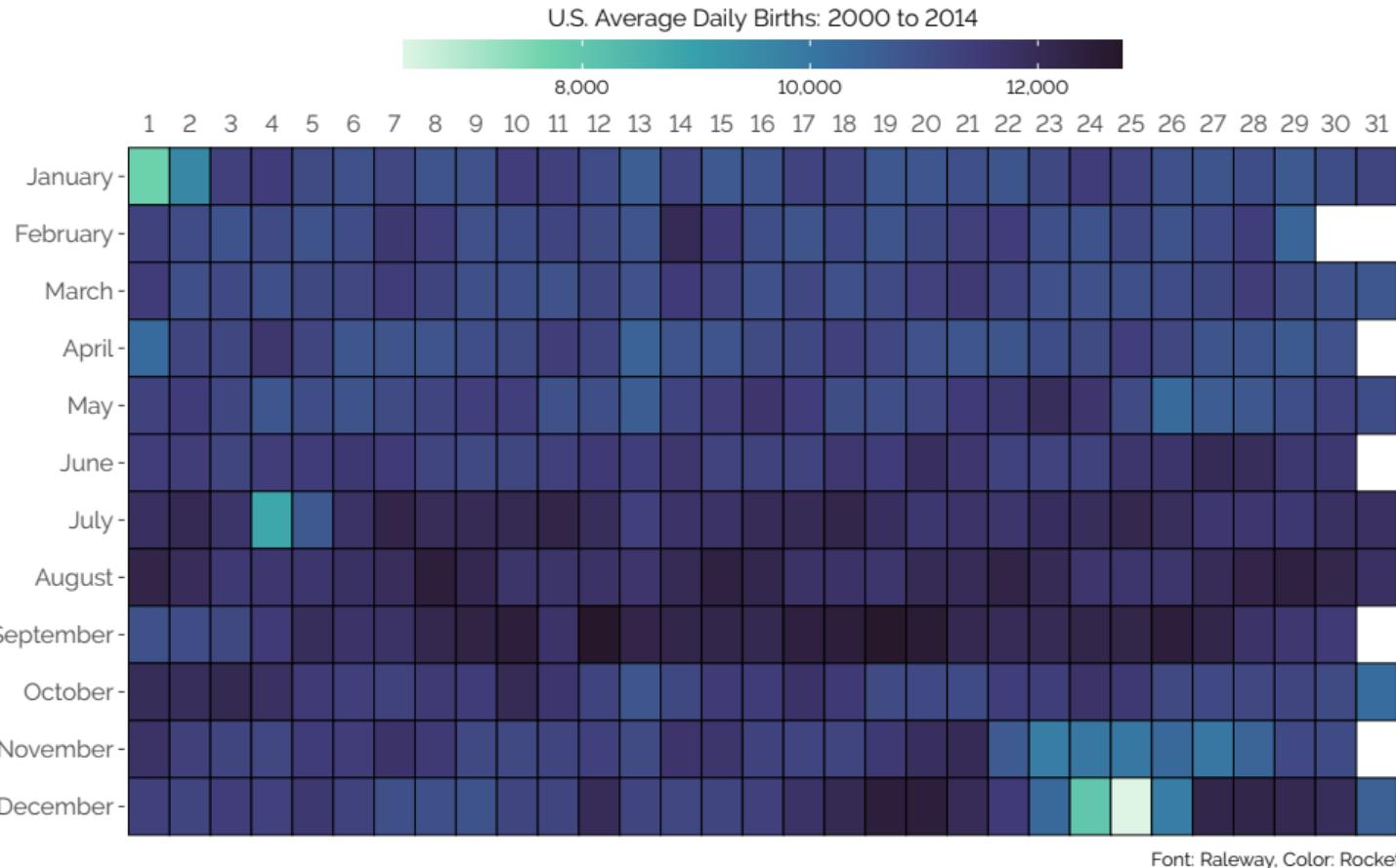


Fonts – practical examples



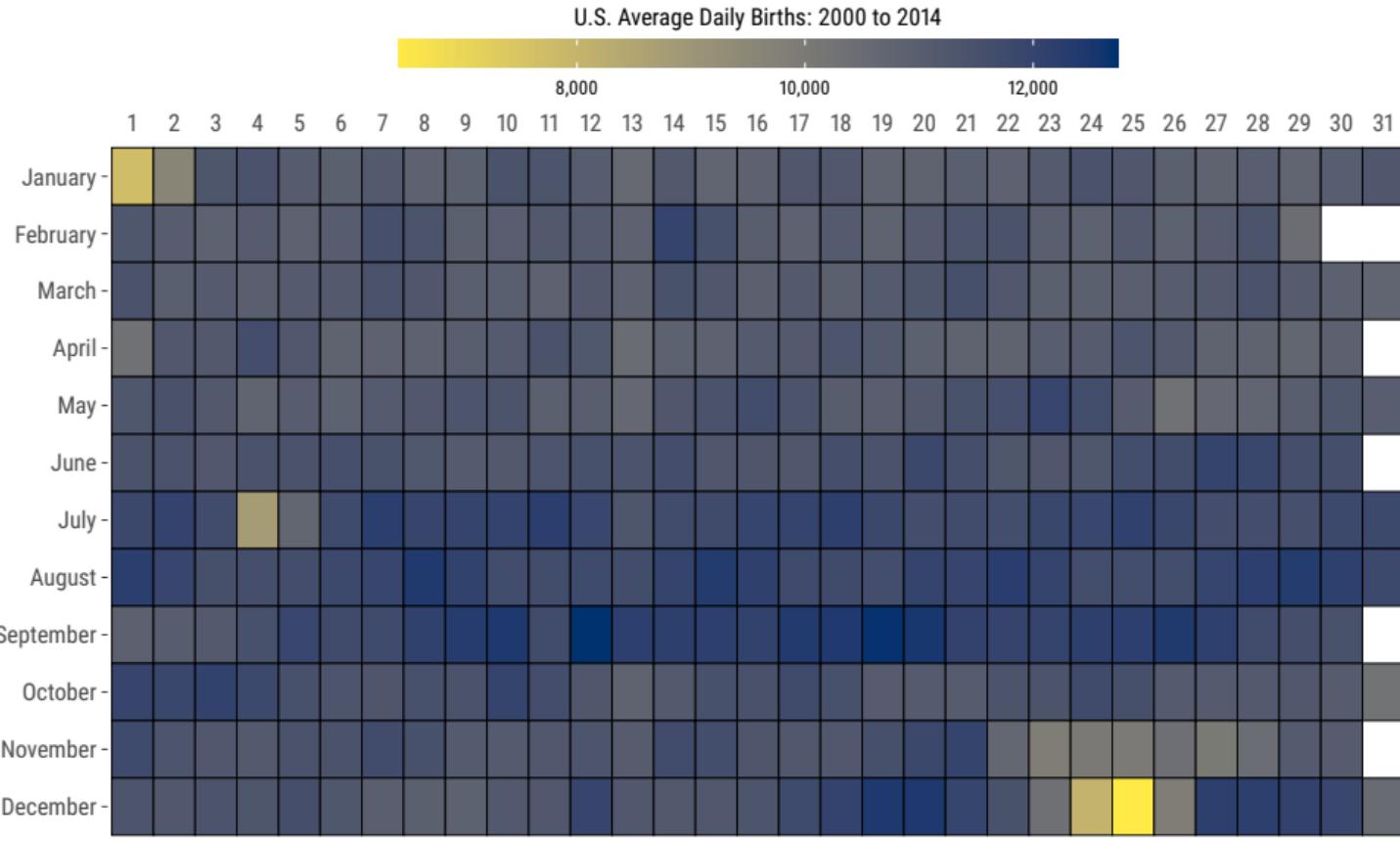
Font: IBM Plex Sans, Color: Magma

Fonts – practical examples



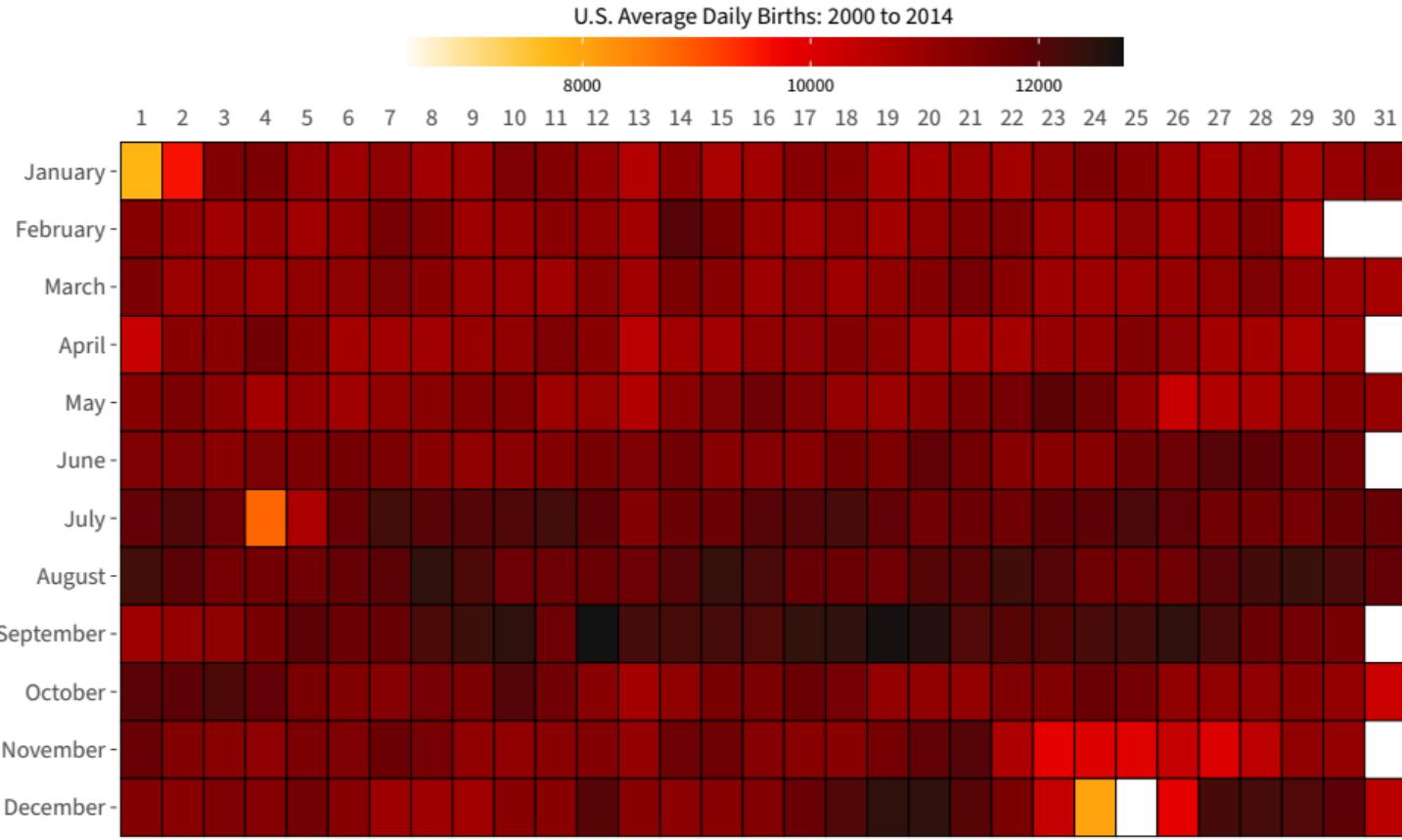
Font: Raleway, Color: Rocket

Fonts – practical examples



Font: Roboto Condensed, Color: Cividis

Fonts – practical examples

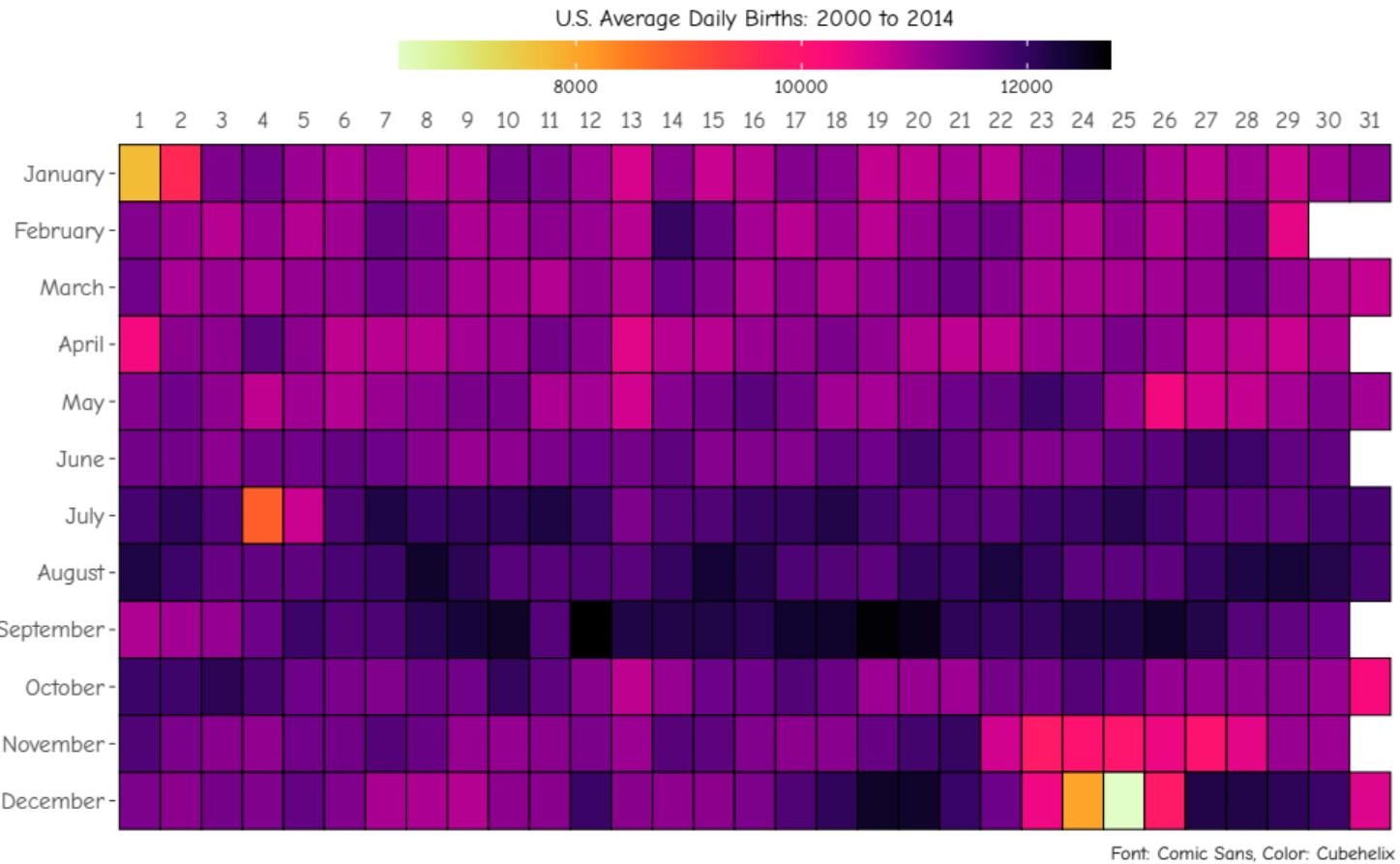


Font: Source Sans Pro, Color: Fire

Fonts – suggestions

- IBM Plex Sans
- Fira Sans – the font for this presentation!
- Tahoma/Segoe UI (old version)
- Atkinson Hyperlegible
- Clear Sans
- B612
- Source Sans Pro, Roboto, Raleway, Lato, Droid Sans, ...

Comic Sans?



Font: Comic Sans, Color: Cubehelix

You're a

COMIC SANS CRIMINAL

*but we're here to
help you*

CLICK ANYWHERE ON THIS PAGE TO CONTINUE

Source: <http://www.comicsanscriminal.com/>

Comic Sans and Accessibility

- Almost everything has been created by non-disabled people, forming barriers and preventing disabled people from participating.
- Accessibility is the process of removing barriers to meet people's access needs. It goes hand in hand with usability, affordability, availability, and overall good design.
- The general consensus amongst graphic designers is that sans serif is more accessible because it's more clean and less distracting.
- The overarching message in accessibility is: listen to disabled people. Many people in the disability community prefer fonts like Comic Sans. Comic Sans is more legible because it has bulbous curves and disproportionate lines. All these characteristics help people differentiate letters more when they read.

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Accessibility - alignment



Accessible

Space, the final frontier. These are the voyages of the Starship Enterprise.



Not accessible

Space, the final frontier. These are the voyages of the Starship Enterprise.

The fully justified text, while it creates a clean overall text block, creates uneven “rivers of white” spacing that makes it difficult to read.

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Accessibility - font size



Accessible

Pokemon! Gotta
catch 'em all

A heart so true

Our courage will pull
us through



Not accessible

Pokemon! Gotta catch 'em all
A heart so true
Our courage will pull us through

Use large font sizes instead of tiny text. It's best practice for body text to be at least 16px for web.

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Accessibility - font style



Accessible



Not accessible

Hello world

Use the `` tag instead of the `` tag, that's better for *screen readers*.

HELLO WORLD

Use the `` tag instead of the `` tag, that's better for screen readers.

Use different weights and styles sparingly to make text simple and clean. Use sentence case, not all caps. Don't combine bold and italics. Don't create blocks of text with italics.

Accessibility - type face



Accessible



Not accessible

Comic Sans:

bad dad on no Illinois

Verdana:

bad dad on no Illinois

Helvetica:

bad dad on no Illinois

AquaFina Script:

bad dad on no Illinois

Official Antiqua:

bad dad on no Illinois

Henny Penny:

bad dad on no Illinois

Use a typeface with clean, distinguishable letters instead of cursive, themed, and handwritten styles.

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Accessibility - spacing



Accessible

Everything you own
in the box to the left

In the closet that's
my stuff, yes



Not accessible

Everything you own
in the box to the left
In the closet that's
my stuff, yes

Use accessible spacing to differentiate lines of text and make reading easier. Best practices are 1.5x font size for line spacing and 2x font size for paragraph spacing.

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Accessibility - contrast



Accessible



Not accessible

Let it be

Let it be

Light

Let it be

30%

Let it be

1.90:1

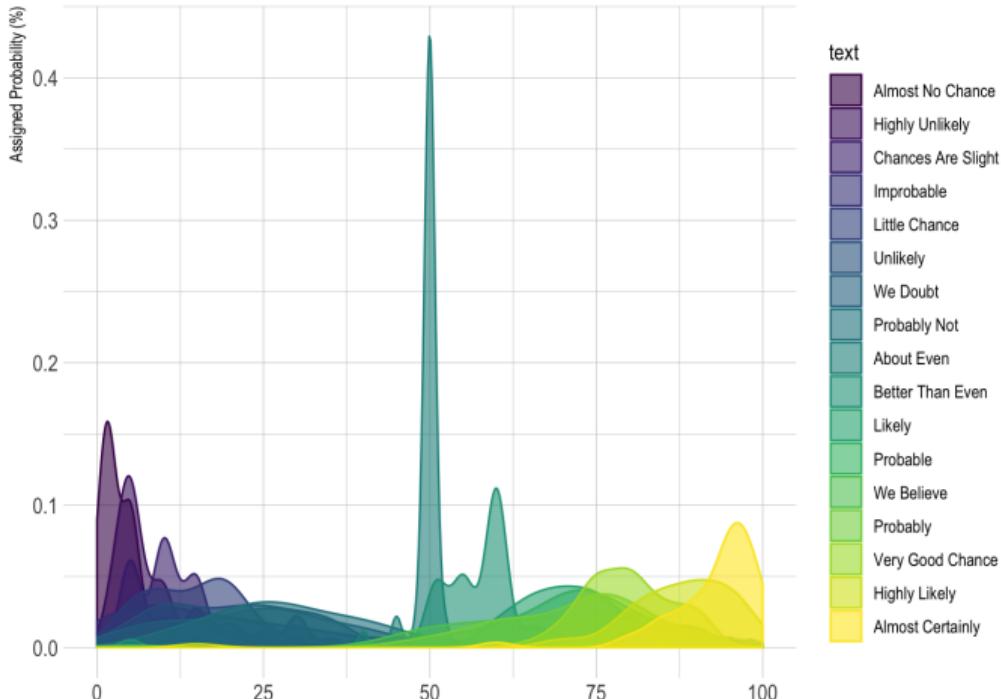
Let it be

1.62:1

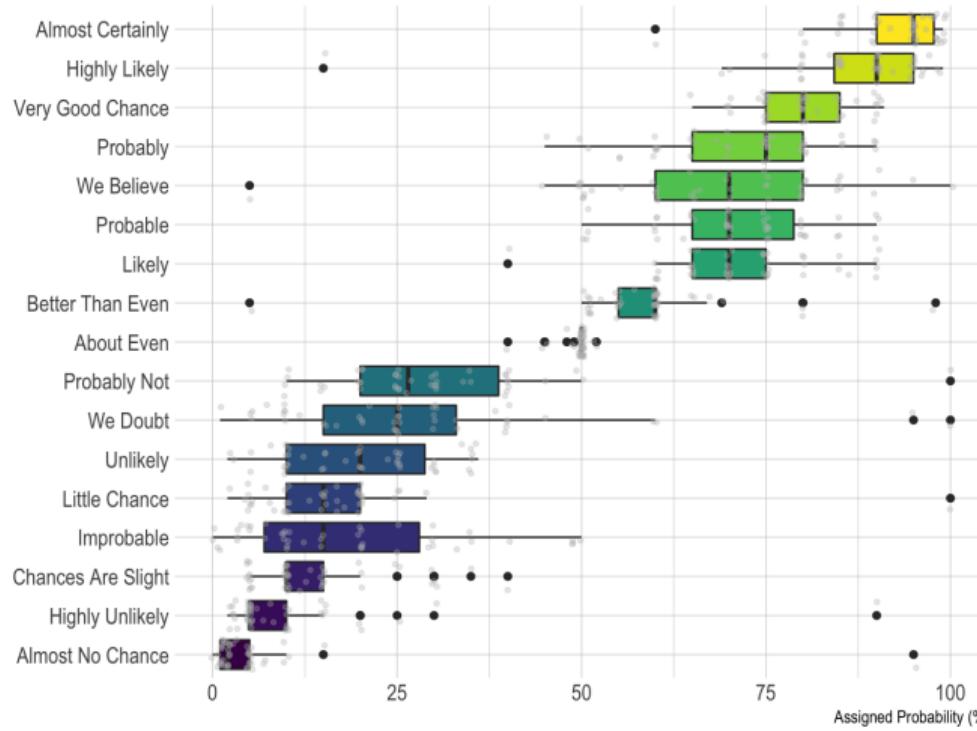
Use high contrast to make text stand out. The text on the left is easy to read because: It uses regular instead of light weight; It's at 100% instead of 30% opacity; and It's at 21:1 contrast (aim for at least 7:1).

Source: <https://medium.com/queer-design-club/the-controversy-of-accessible-type-8def04eb8808>

Common issues I - too many distributions

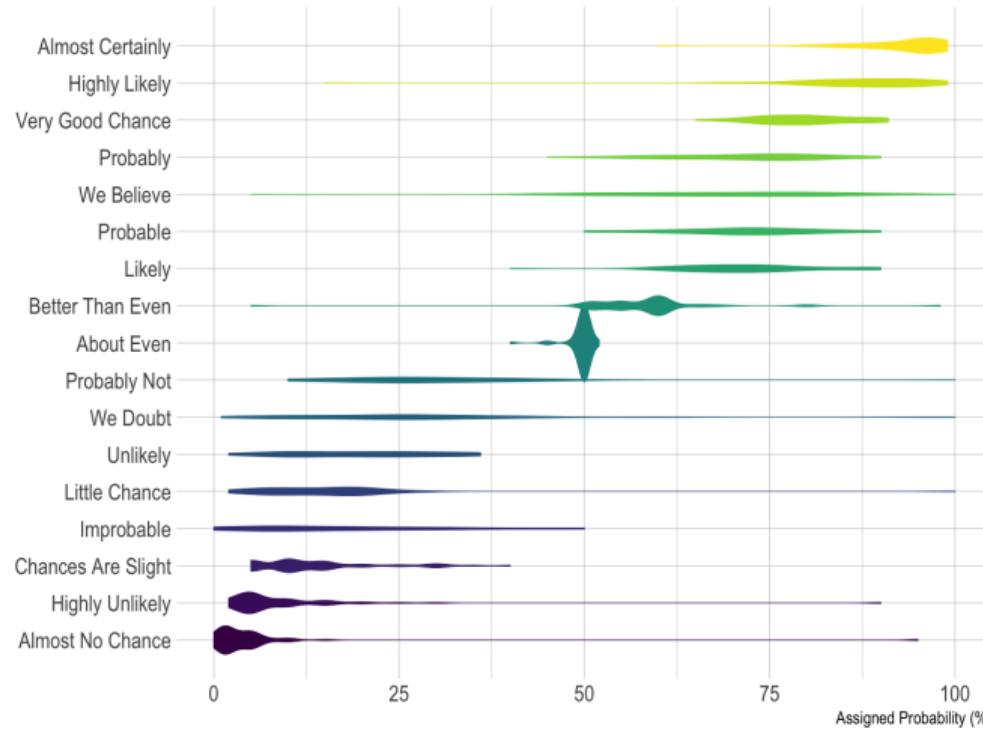


Common issues I - Box Plots



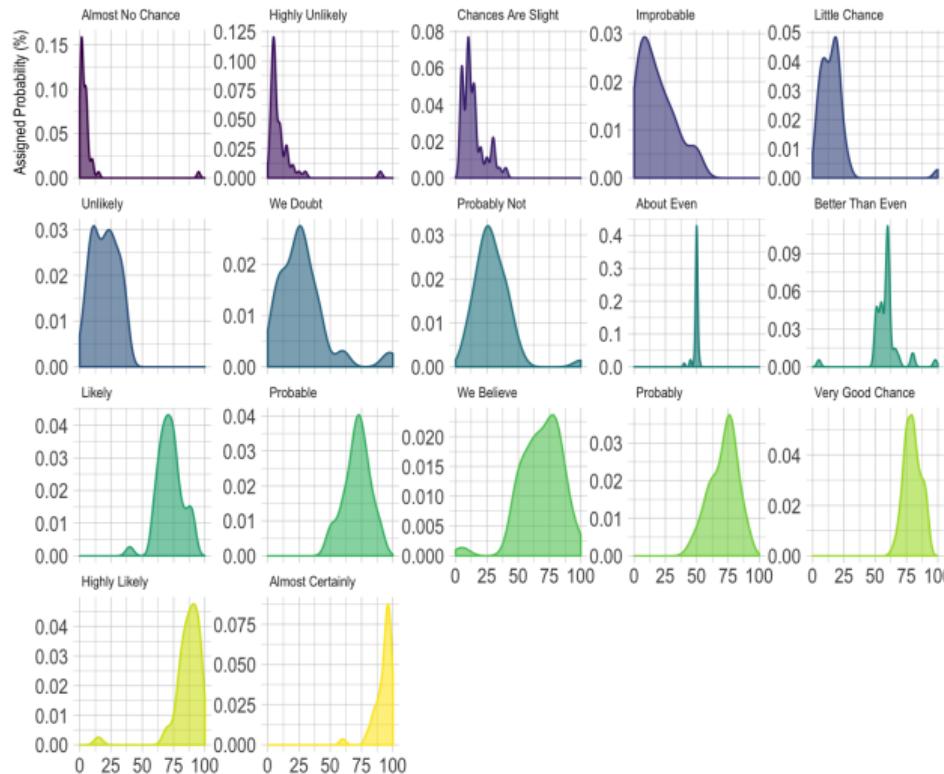
Source: https://www.data-to-viz.com/caveat/multi_distribution.html

Common issues I - Violin Plots



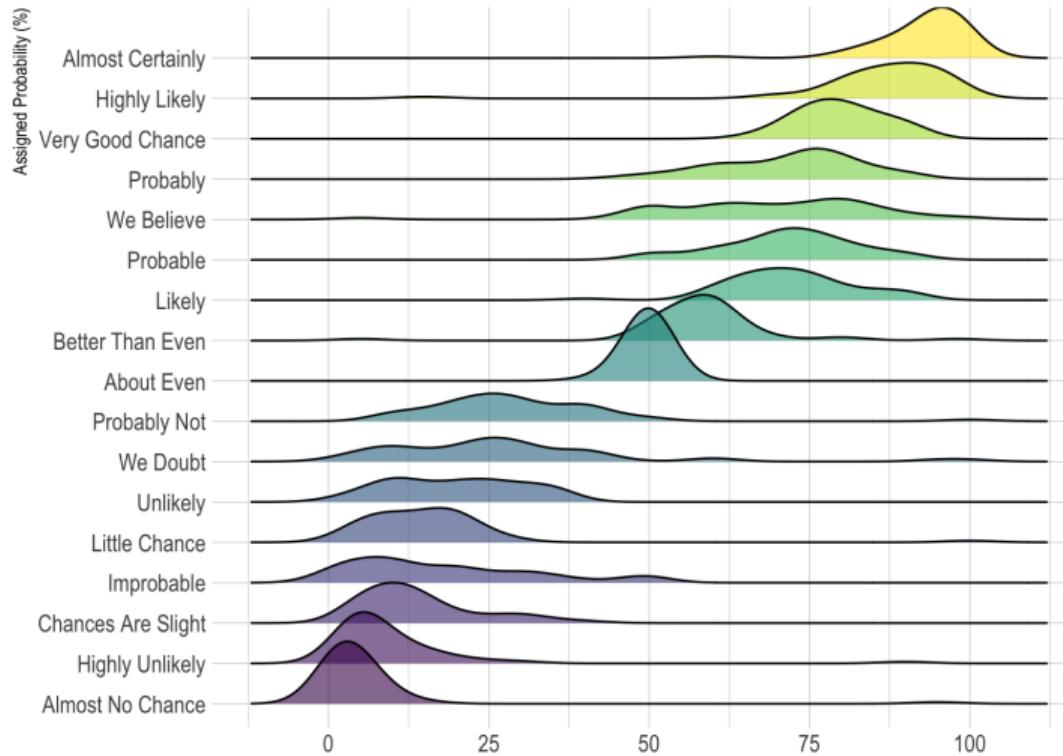
Source: https://www.data-to-viz.com/caveat/multi_distribution.html

Common issues I - Small Multiples



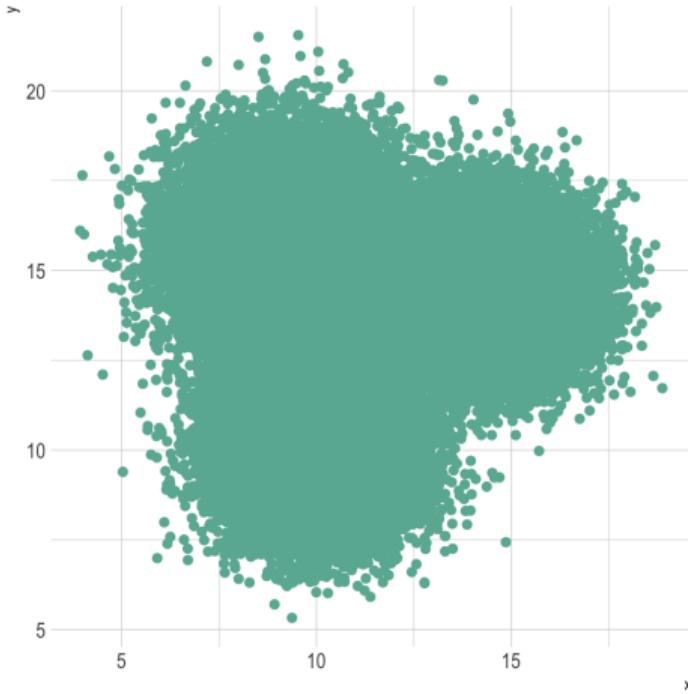
Source: https://www.data-to-viz.com/caveat/multi_distribution.html

Common issues I - Ridgelines



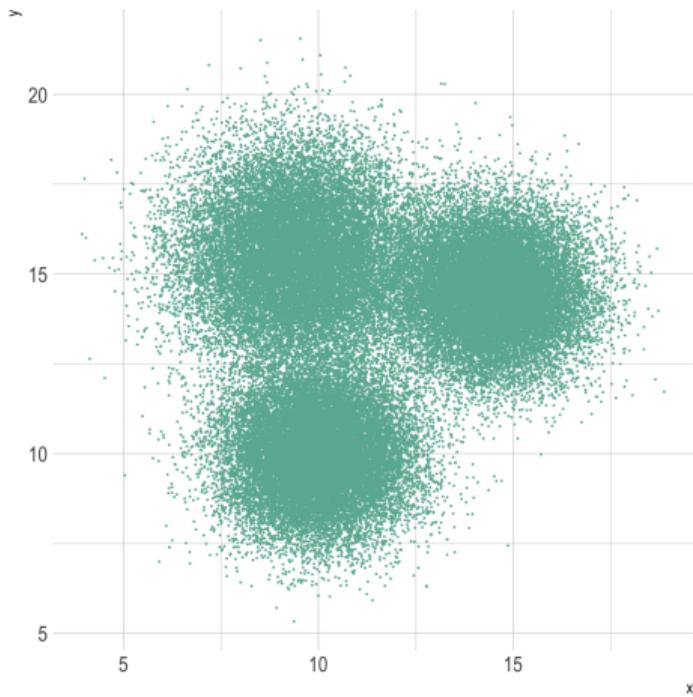
Source: https://www.data-to-viz.com/caveat/multi_distribution.html

Common issues II - Overplotting



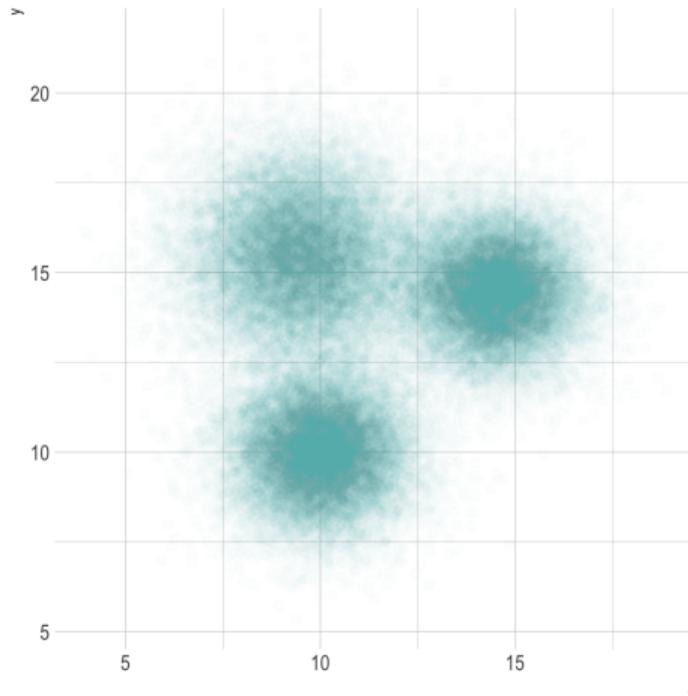
Source: <https://www.data-to-viz.com/caveat/overplotting.html>

Common issues II - Small dots



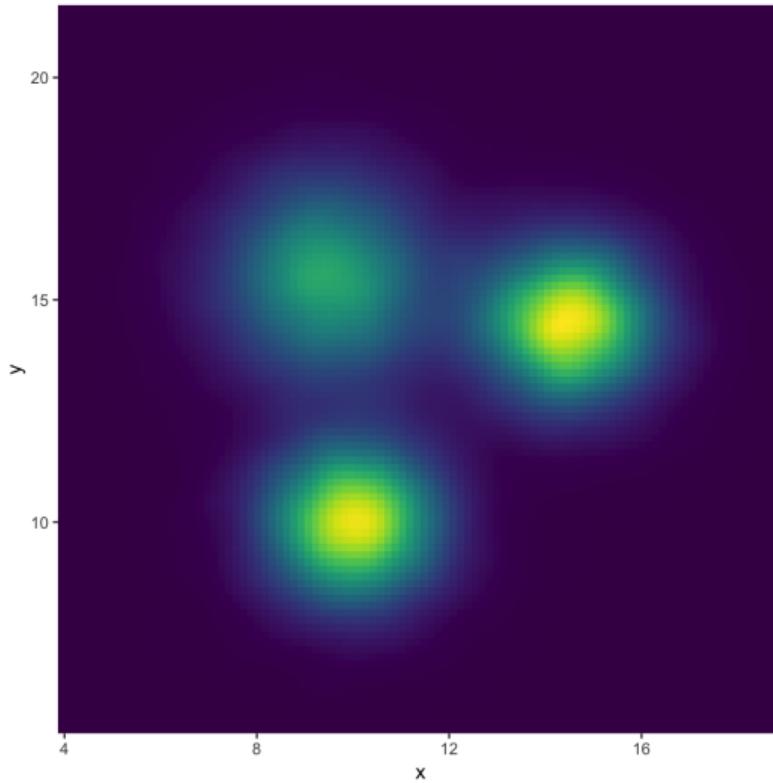
Source: <https://www.data-to-viz.com/caveat/overplotting.html>

Common issues II - Transparency



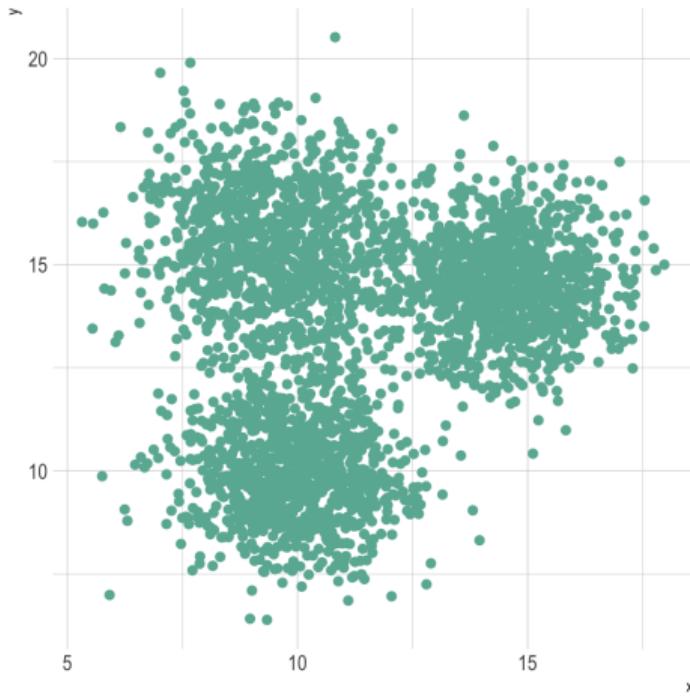
Source: <https://www.data-to-viz.com/caveat/overplotting.html>

Common issues II - 2D Density



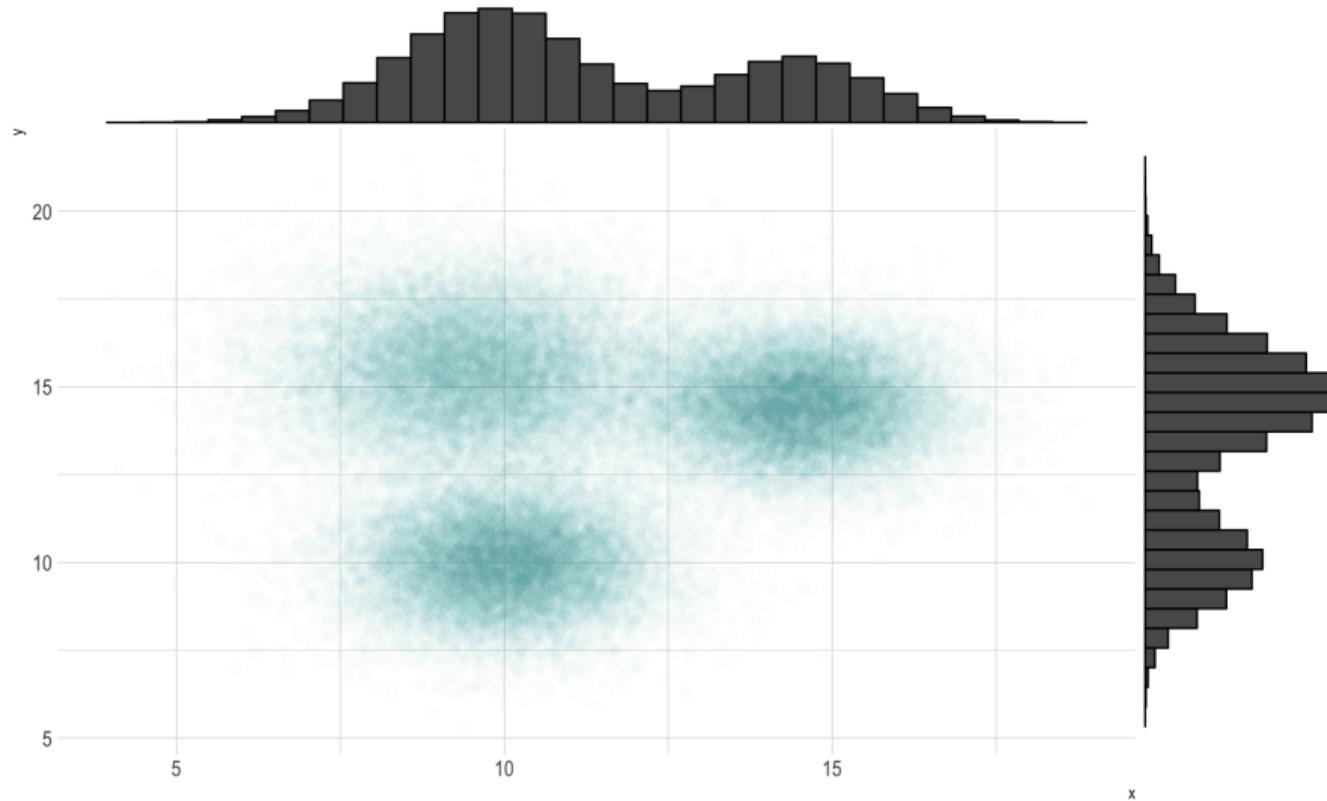
Source: <https://www.data-to-viz.com/caveat/overplotting.html>

Common issues II - Sampling



Source: <https://www.data-to-viz.com/caveat/overplotting.html>

Common issues II - Marginal plots



Source: <https://www.data-to-viz.com/caveat/overplotting.html>

DESIGN PHILOSOPHY

Data Visualisation

I have the simplest tastes. I am always satisfied with the best.

Gestalt design - revisited

- Proximity – objects that are close together are perceived as a group.
- Similarity – objects that share similar attributes (e.g., color or shape) are perceived as a group.
- Enclosure – objects that appear to have a boundary around them (e.g., formed by a line or area of common color) are perceived as a group.
- Closure – open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.
- Continuity – objects that are aligned together or appear to be a continuation of one another are perceived as a group.
- Connection – objects that are connected (e.g., by a line) are perceived as a group.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

What are we trying to achieve?

- The effectiveness of data visualization is its ability to represent information in a manner that our eyes can recognize and our brains can understand.
- Getting this right is more of a science than an art, which we can achieve by studying human perception.
- The goal is to translate abstract information into visual representations that can be easily, efficiently, accurately, and meaningfully understood.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

A good data visualisation...

- Clearly indicate how values relate to one another.
- Represents the values accurately.
- Makes it easy to compare values.
- Makes it easy to rank and order the values.
- Makes it obvious how people should use this information, what they should use it to accomplish, and encourage them to follow through.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

How are data visualisations processed

- Data visualisations are so effective because they shift the balance between perception and cognition to take better advantage of the brain's abilities.
- Seeing (visual perception), which is handled by the visual cortex located in the rear of the brain, is extremely fast and efficient.
- Thinking (cognition), on the other hand, which is handled predominantly by the cerebral cortex in the front of the brain, is much slower and less efficient.
- This means traditional data sense-making and presentation methods require conscious thinking for almost all of the work. However, data visualization shifts the balance towards significant use of visual perception, taking advantage of our eyes and speed of seeing.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

Why data visualisations are faster

- Our brain has the capability of detecting shape, color and motion separately which allows our audience to understand, make associations and identify dimensions in data visualizations almost instantaneously.
- Designing data visualizations that incorporate pre-attentive features make them much more effective thanks to pre-attentive processing, which is the subconscious accumulation of information from the environment, where all the available information is pre-attentively processed. Then, the brain filters and processes what is important.
- Information that stands out the most or is of most relevance to what a person is thinking about is selected for further and more complete analysis by attentive processing.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

Why data visualisations are faster

- Pre-attentive features are processed very quickly, within around 10 milliseconds. To put it simply, the better designed the data visualization, the more understandable it is, the more information it can represent and the faster it is to read and register.
- Considering the speed at which people can understand multiple data points through the dual processing theory and pre-attentive processing, it is wise to take advantage of data visualization.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

Why data visualisations are faster

- Pre-attentive features are processed very quickly, within around 10 milliseconds. To put it simply, the better designed and thoughtfully created the data visualization, the more understandable it is, the more information it can represent and the faster it is to read and register.
- Considering the speed at which people can understand multiple data points through the dual processing theory and pre-attentive processing, it is wise to take advantage of data visualization.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

Data visualisation qualities

- Truthful – Based on thorough and objective research.
- Functional – Accurate and allow your users to act upon your information.
- Beautiful – Well-designed and draw the user's attention through an aesthetically pleasing display of information.
- Insightful – Provide information like trends, insights, and inferences that would be difficult to see otherwise.
- Enlightening – Highlight your evidence, tell a story and enlighten users with your information in a way that is easy to understand.

Source: <https://treehousetechgroup.com/the-psychology-behind-data-visualization/>

Edward Tufte (VDQI, 1983)

- The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new... Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often chartjunk.
- Fortunately most chartjunk does not involve artistic considerations. It is simply conventional graphic paraphernalia like: over-busy grid lines and excess ticks, redundant representations of the simplest data, the debris of computer plotting...
- Chartjunk can turn bores into disasters, but it can never rescue a thin data set. The best designs... are intriguing and curiosity-provoking, drawing the viewer into the wonder of the data, sometimes by narrative power, sometimes by immense detail, and sometimes by elegant presentation of simple but interesting data.

Source: https://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

Useful junk?

- We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Source: <http://www.stat.columbia.edu/~gelman/communication/Bateman2010.pdf>

- Although our results largely corroborate [Bateman (2010)] that visual embellishments do not significantly hinder the interpretability of a chart, we did not find evidence to corroborate Bateman et al.'s result that embellished charts lead to better long-term recall. Source:
<https://osf.io/fkwav/download>
- Our findings agree with the previous studies regarding interpretability of the charts but disagree with the findings regarding long and short-term recall.

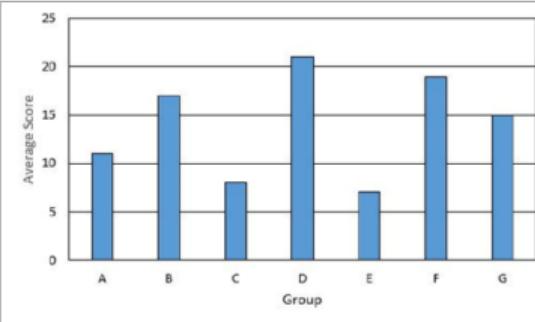
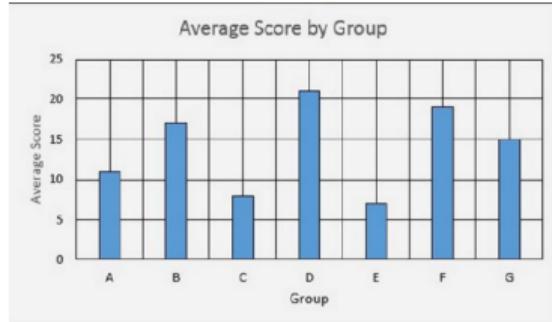
Source: <https://journals.sagepub.com/doi/abs/10.1177/1541931214581316?journalCode=proe>

Hill et al (2018)

In this work, perceptions of minimalist approaches to data visualization were examined via a survey. Respondents were presented with two sets of four plots and were asked to rate each plot on perceived Beauty, Clarity, Effectiveness, and Simplicity. The plots presented similar data, but were varied in their data-ink ratio.

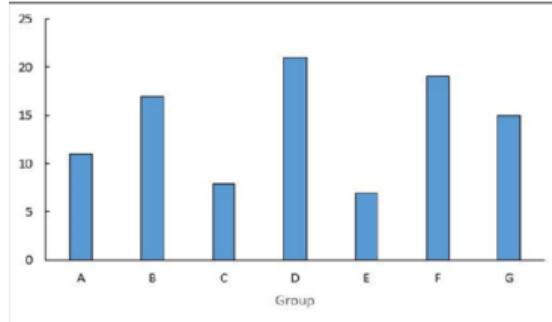
Source: <https://www.semanticscholar.org/paper/Minimalism-in-Data-Visualization%3A-Perceptions-of-Hill-Wray/c865133ed0419e9614f69277a04c292aa19b98>

Hill (2018) Bar Plots

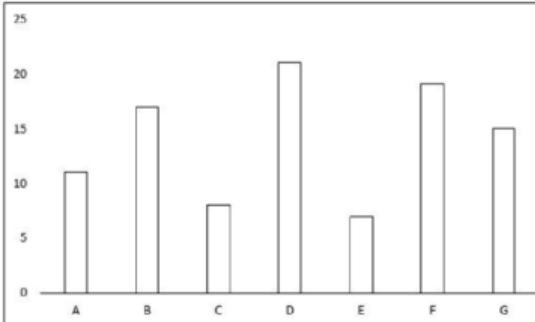


Plot A

Plot B

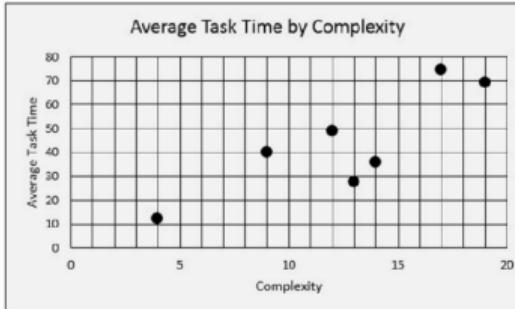


Plot C

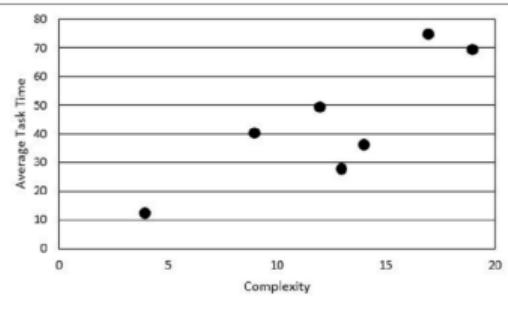


Plot D

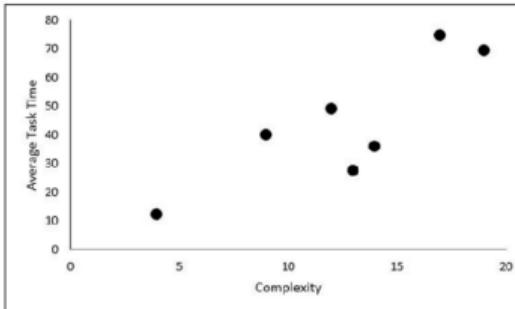
Hill (2018) Scatter Plots



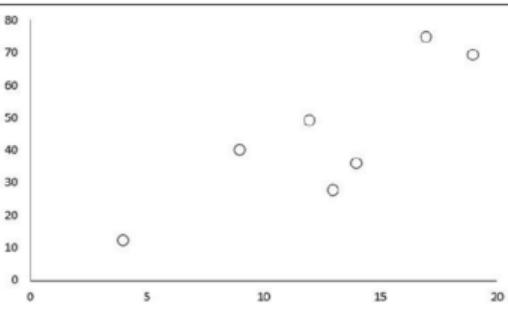
Plot A



Plot B



Plot C



Plot D

Hill (2018) Results

- The results of the survey response analysis suggest that plots with lower data-ink ratios (less minimalist) are perceived as superior on each of the four dimensions. This held true whether the plots presented were barplots or scatterplots...The results are contrary to the philosophy of highly-regarded visualization expert Edward Tufte.
- The authors of this work have formed a post-analysis hypothesis. This hypothesis suggests that the respondents are “used” to seeing default plots as produced by Excel. These plots feature a moderately low data-ink ratio.
- A lack of familiarity with minimalist visualizations with high data-ink ratios may drive respondents to view these visualizations less favorably...Also, survey respondents were not provided definitions for the terms “beauty”, “clarity”, “effectiveness”, and “simplicity”.

Attitudes towards high data-ink charts

Motivation To evaluate people's acceptance of the minimalist approach to information visualization.

Research approach Eighty seven students, divided into three experimental conditions, rated their preference for two different graphs displaying identical information – a standard bar-graph and a minimalist version from Tufte (1983).

Findings/Design The results indicate a clear preference of non-minimalist bar-graphs, suggesting low acceptance of minimalist design principles such as high data-ink ratio.

Research limitations/Implications Subjects had no prior experience with the minimalist graph.

Take away message People did not like Tufte's minimalist design of bar-graphs; they seem to prefer "chartjunk" instead.

Source: <https://dl.acm.org/doi/abs/10.1145/1362550.1362587>

Progressively more high data-ink charts?

With regard to Tufte's claim that his high data-ink designs would be accepted with time, interview feedback indicated that high data-ink designs are not encountered or accepted by frequent users of graphs. Although models of graph comprehension and the results of the present study do seem to support the claim that viewers would be accustomed to high data-ink ratio designs, it does not seem that they have started to "catch on" in the years since Tufte published the data-ink concept.

Source: <https://scholarworks.rit.edu/cgi/viewcontent.cgi?article=9824&context=theses>

More thoughts

Graphical embellishments can support the effectiveness of a data visualization in three potential ways:

- by engaging the interest of the reader (i.e., getting them to read the content),
- by drawing the reader's attention to particular items that merit emphasis, and
- by making the message more memorable.

Embellishments only enhance effectiveness, however, if they refrain from undermining the message by significantly distracting from it or misrepresenting it.

Source: https://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

Declutter and focus

- The declutter guidelines suggest removing non-critical gridlines, excessive labeling of data values, and color variability to improve aesthetics and to maximize the emphasis on the data relative to the design itself.
- The focus guidelines for explanatory communication recommend including a clear headline that describes the relevant data pattern, highlighting a subset of relevant data values with a unique color, and connecting those values to written annotations that contextualize them in a broader argument.
- Decluttering designs led to higher ratings on professionalism, and adding focus to the design led to higher ratings on aesthetics and clarity. They also showed better memory for the highlighted pattern in the data, as reflected across redrawings of the original visualization and typed free-response conclusions...

Embellishments in Infographics

- By analyzing questionnaires, interviews, and eye-tracking data simplified by bundling, we show that, within bounds, embellishments have a positive effect on how users get engaged in understanding an infographic, with very limited downside.

Source: <https://www.semanticscholar.org/paper/Interpreting-the-Effect-of-Embellishment-on-Chart-Andry-Hurter/95dcefcb5fa362a285f141400754014ffff781b6>

More than you ever wanted to know about general design principles

Accessibility There are four characteristics of accessible designs: perceptibility, operability, simplicity, and forgiveness.

Advance organisers These are brief chunks of information presented prior to new material to help facilitate learning and understanding - they present the “big picture” prior to the details.

Aesthetic-Usability Effect Aesthetic designs look easier to use and have a higher probability of being used, whether or not they actually are easier to use.

Alignment Elements in a design should be aligned with one or more other elements. This creates a sense of unity and cohesion, which contributes to the design’s overall aesthetic and perceived stability. Alignment can also be a powerful means of leading a person through a design.

Anthropomorphic Form The classic 1915 Coca-Cola “contour” bottle, often referred to as the “Mae West” bottle due to its distinctly feminine proportions, was a break with the straight and relatively featureless bottles of its day. In addition to its novelty, however, the bottle benefited from a number of anthropomorphic projections such as health, vitality, sexiness, and femininity, attributes that appealed to the predominantly female buyers of the time.

SPECIAL CHARTS

COLOUR AND PERCEPTION

INTERACTIVITY

STORYTELLING WITH DATA – PRACTICAL EXAMPLES