

Data Visualisation AM10

Session 1: The basics

Kostis Christodoulou
03 Nov 2021



Assignments (Individual) 25% of Total

- Dashboarding and data story telling using Tableau Due Nov 30 at 11:59pm | 20 pts
- Data cleaning Due Nov 21 at 11:59pm | 5 pts

Thought Piece (Individual) 5% of Total

- Thought piece Due Nov 9 at 11:59pm | 5 pts

Problem Sets (Individual) 20% of Total

- Problem set 1 Due Nov 9 at 11:59pm | 10 pts
- Problem set 2 Due Nov 16 at 11:59pm | 10 pts

Final Group Project (Group) 50% of Total

- Final Group Project Due Nov 24 at 8am | 50 pts

- Submit proposal for group project: Due before session 2
- Reflections: Due before session 2

Please write a short essay (max. 400 words) addressing the following:

Why do we visualize data? What makes a great visualization?

How do you choose which kind of visualization to use?

What does it mean to map data to graph aesthetics? What data was mapped to which aesthetics in Rosling's video "200 Countries, 200 Years, 4 Minutes"

- Grading visualisations: Due before session 2

A rubric to 'grade' visualizations is [The Data Visualization Checklist](#). It is a helpful set of criteria for grading the effectiveness of a graphic.

Using the Data Visualisation checklist, please 'grade' Figure 10 of the recent JPMorgan Chase Institute on [Year-over-year percent change in credit card spending by industry of employment](#). If you wanted the actual paper on which this publication was based, you can find it [here](#).

- Prepare the *worst* possible visualisation: Due before session 2

o Inspired by Allison Horst's [There's value in trying your \[dataviz\] worst](#). Inspired and The Economist's [Mistakes, we've drawn a few](#), take any of your past visualisations, or find a small dataset and please come up with the **worst** possible visualisation you can

- Problem Set 1: Due before session 2

[Stop and Search in London, part 1](#)

<https://www.met.police.uk/stopandsearch/>

<https://www.theguardian.com/law/2019/jan/26/met-police-disproportionately-use-stop-and-search-powers-on-black-people>

<https://www.met.police.uk/sd/stats-and-data/met/stop-and-search-dashboard/> (in Tableau!)

Raw data can be downloaded from <https://data.police.uk/data/> Select:

Date Range: Sep 2020

Metropolitan Police Service

Include stop and search data

← → ⌛ ⌂ 🔒 github.com/kostis-christodoulou/am10.mam2022

 Search or jump to... / Pulls Issues Marketplace Explore

🔒 [kostis-christodoulou / am10.mam2022](#) Private

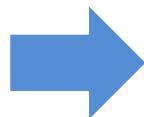
[Unwatch](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#)

[main](#) Go to file Add file Code

	kostis-christodoulou session 1 ...	21 seconds ago ⏱ 3
📁	data initial commit	16 days ago
📁	workshop_session1 session 1	21 seconds ago
📄	.gitattributes session 1	21 seconds ago
📄	.gitignore session 1	21 seconds ago
📄	README.md initial commit	16 days ago
📄	am10.mam2022.Rproj initial commit	16 days ago

Contents

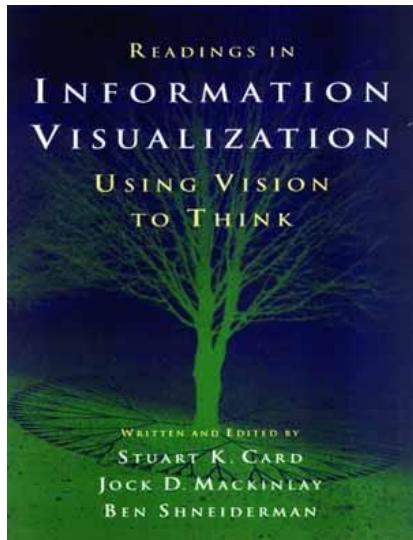


- Introduction to visualisation
- Cairo's Five Principles
- C.R.A.P. Design principles
- Workshop
- Directory of visualisations

What is a visualisation?

The use of computer-supported,
interactive visual representations of
data to amplify cognition.

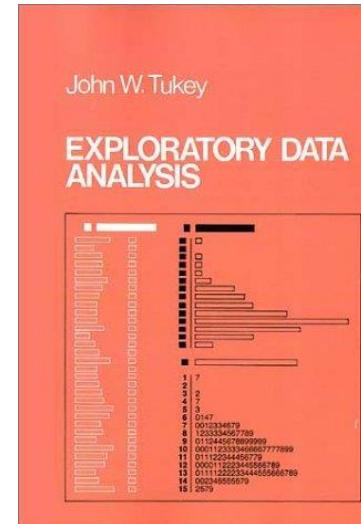
Card, Mackinley, and Shneiderman (1999)



[...] *use of visualization to discover relationships, using interactive graphics to amplify thought.*

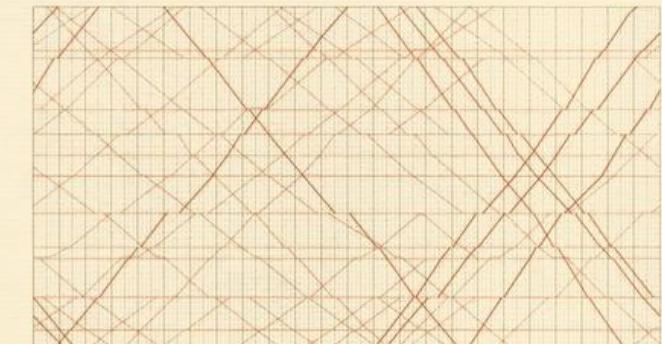
The greatest value of a picture
is when it forces us to notice
what we never expected to
see.

John Tukey (1977)



Tufte's perspective

Graphical excellence is the well-designed presentation of interesting data—a matter of substance, of statistics, and of design ... [It] consists of complex ideas communicated with clarity, precision, and efficiency. ... **[It]** is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space ... [It] is nearly always multivariate ... And graphical excellence requires telling the truth about the data. (Tufte, 1983, p. 51).



The Visual Display
of Quantitative Information

EDWARD R. TUFTE

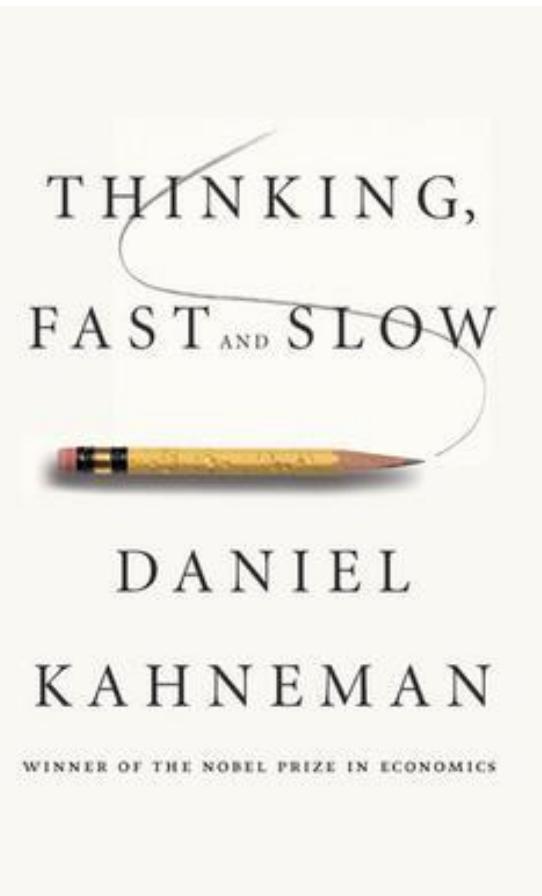
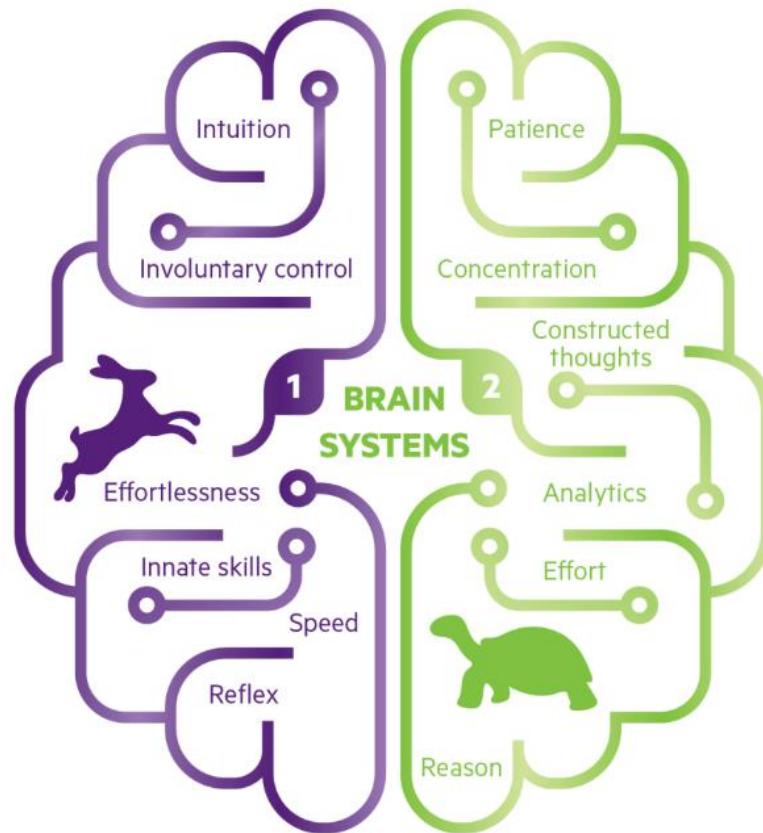
Intuition vs analysis

Why do people follow intuition? It's fast.

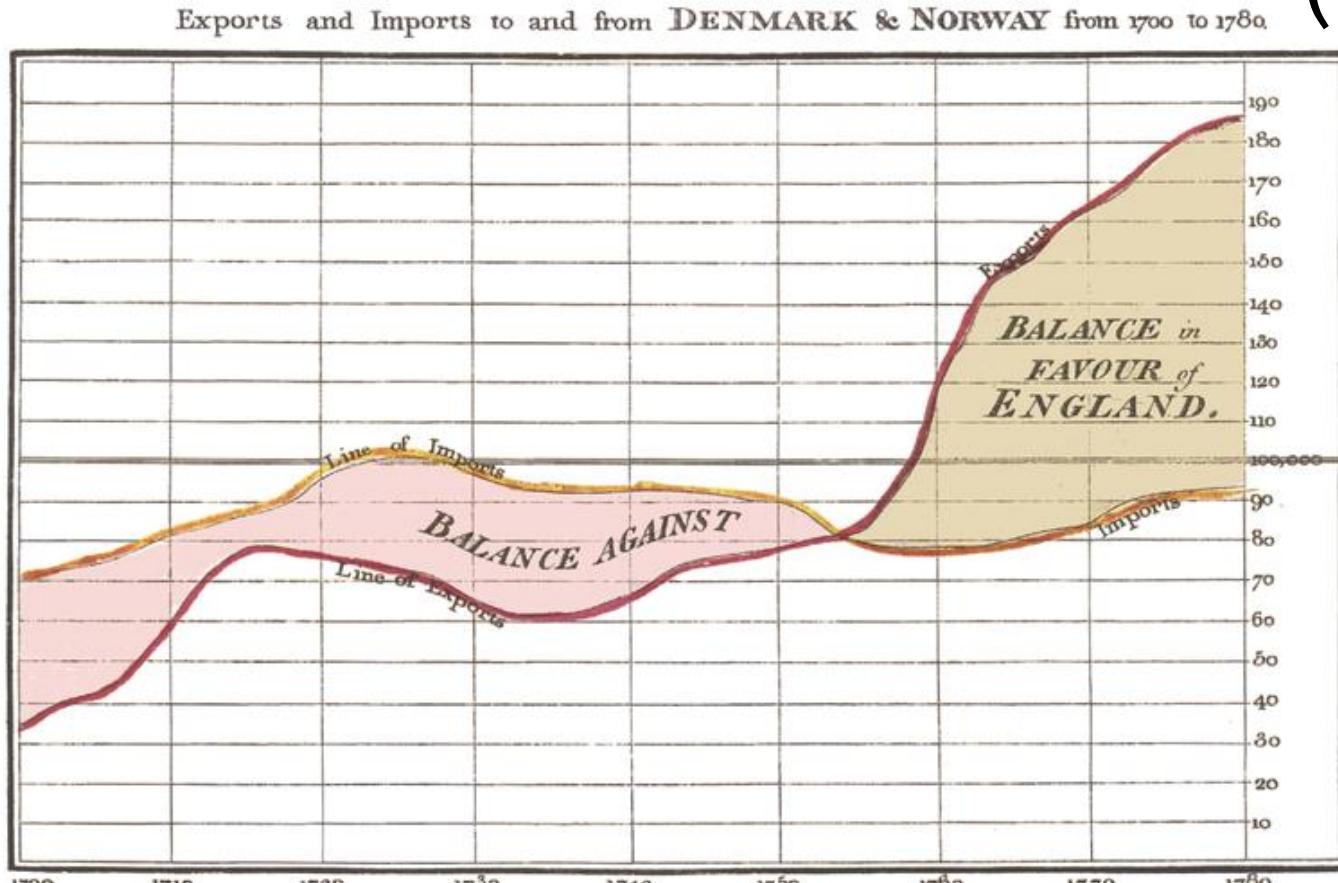
If people follow intuition that is close to their field, they do make good decisions.

System 1 – Fast Thinking: Intuition- heuristics- rule of thumb

System 2 – Slow Thinking : Analytics- data- due diligence- facts



Historical visualisations: Playfair (1786)

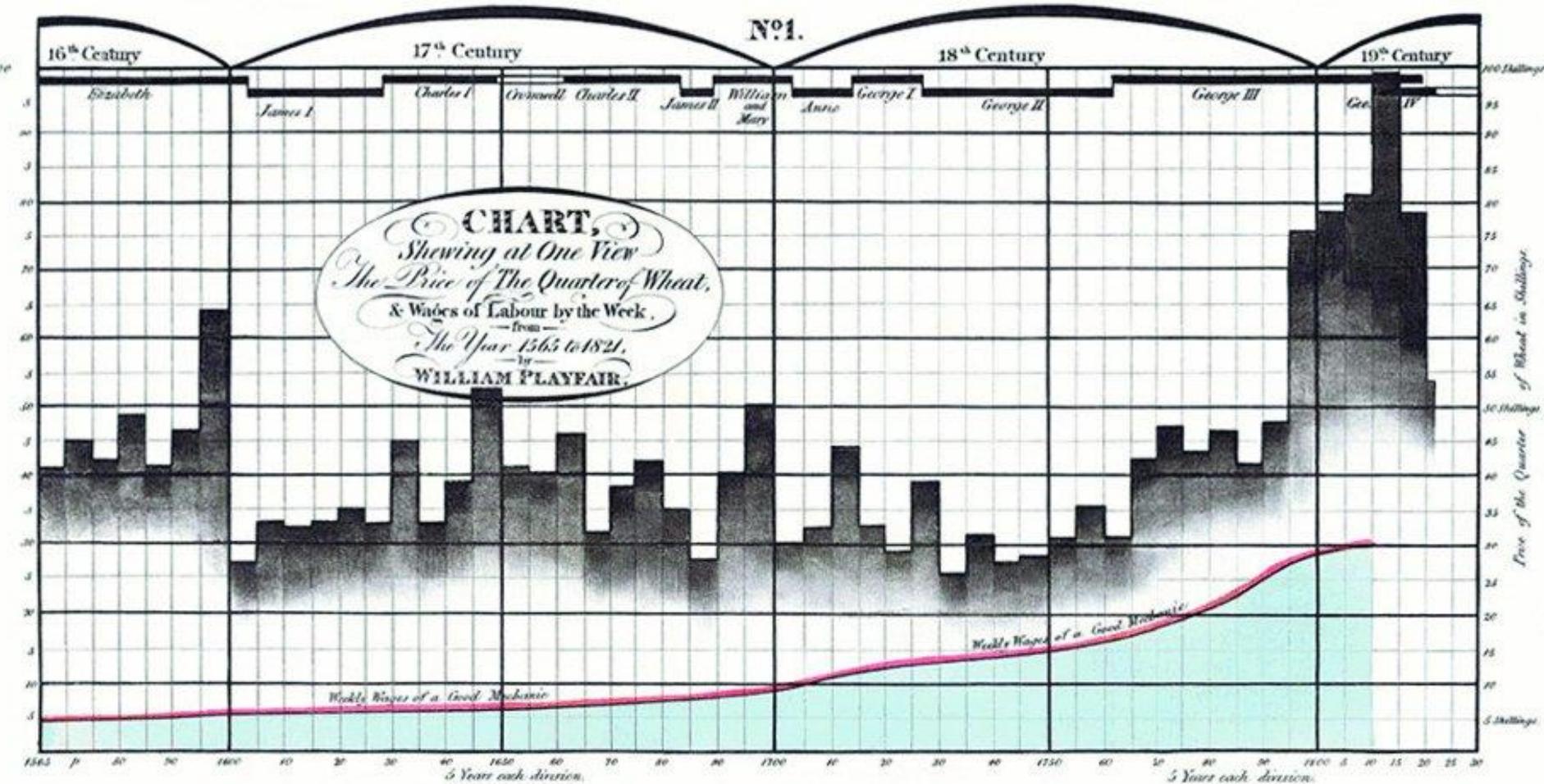


*The Bottom line is divided into Years, the Right hand line into £10,000 each.
Published as the Act directs, 1st May 1786, by W^m Playfair
Neals Yard, 352, Strand, London.*

William Playfair (22 September 1759 – 11 February 1823), a Scottish engineer and political economist, served as a secret agent on behalf of Great Britain during its war with France.^[1] The founder of graphical methods of statistics,^[2] Playfair invented several types of diagrams: in 1786 the line, area and bar chart of economic data, and in 1801 the pie chart and circle graph, used to show part-whole relations.^[3] As secret agent, Playfair reported on the French Revolution and organized a clandestine counterfeiting operation in 1793 to collapse the French currency.

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

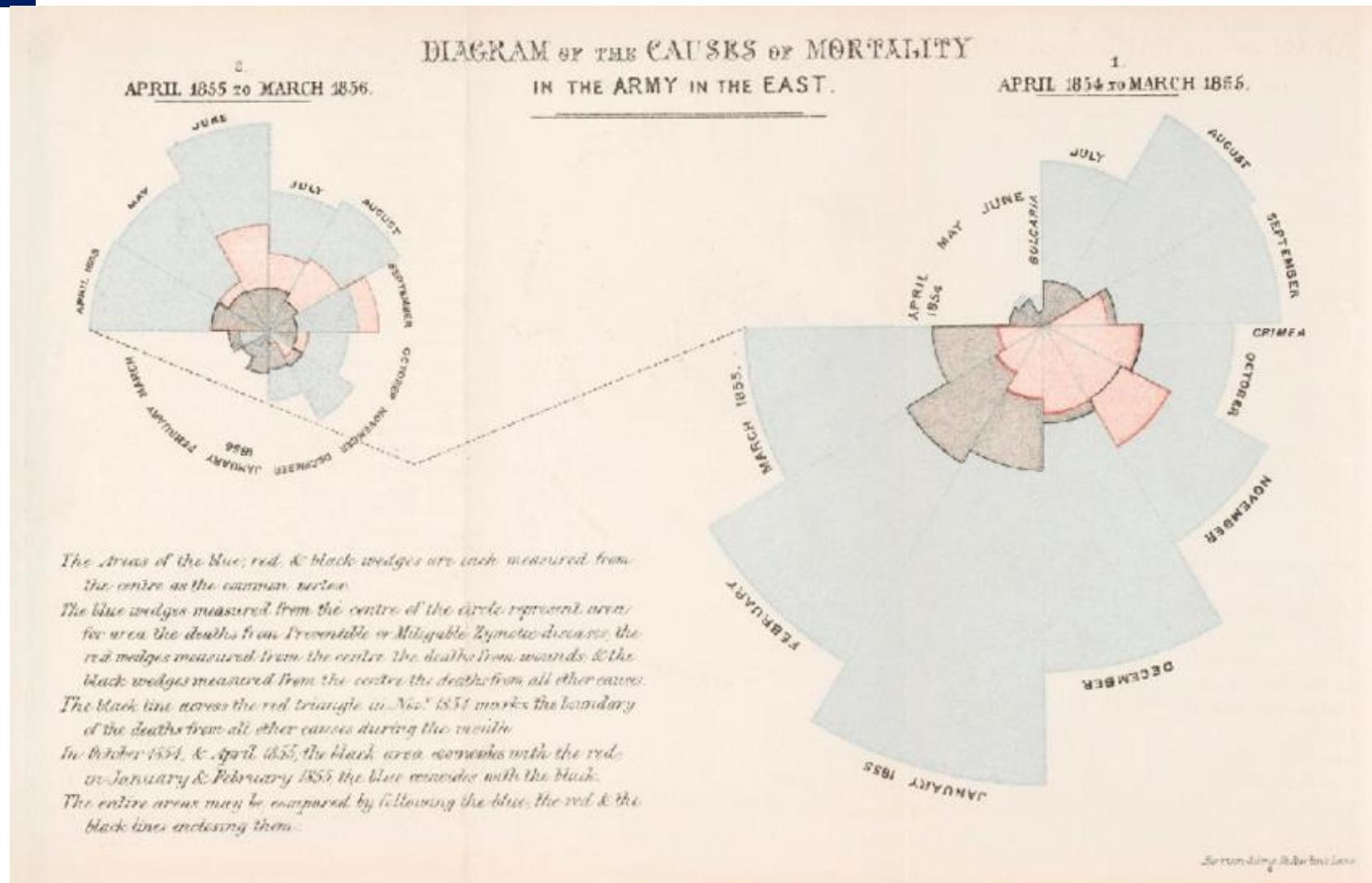
Playfair (1821) Price of Wheat



Historical visualisations: Snow (1854)



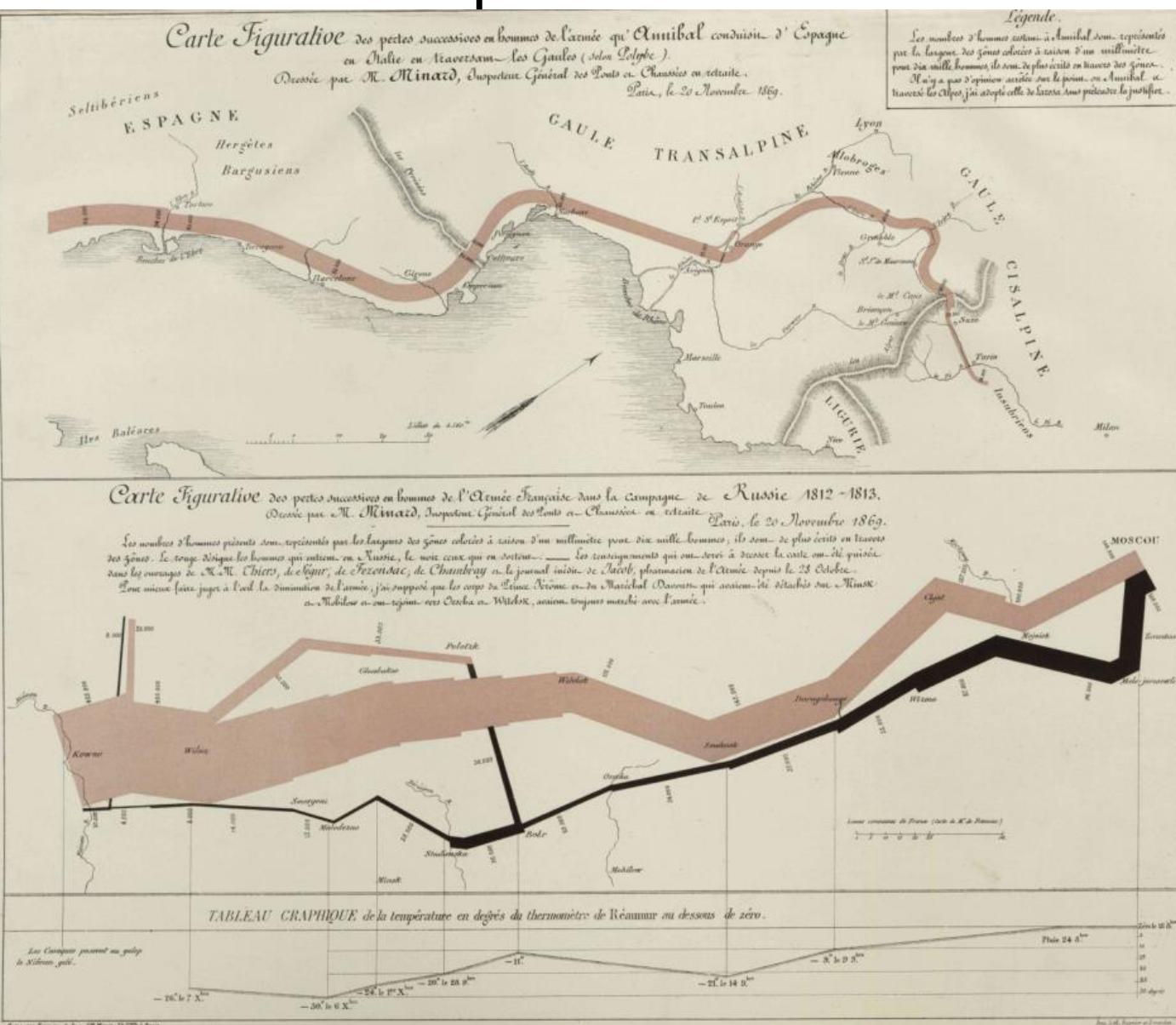
Nightingale (1856)



- Blue Deaths due to treatable diseases
Red Deaths due to war injuries
Black All other causes of death

Hannibal's crossing of the Alps (218 BC) Napoleon's invasion of Russia (1812-1813)

Flow Maps



Often considered the best statistical graphic ever drawn

Charles Minard (1869)

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.
Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les larges des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été pris dans les ouvrages de M. Chiers, de Séguir, de Fezensac, de Chambray et le journal intime de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Nérome et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent Orsha et Witelsk, avaient toujours marché avec l'armée.

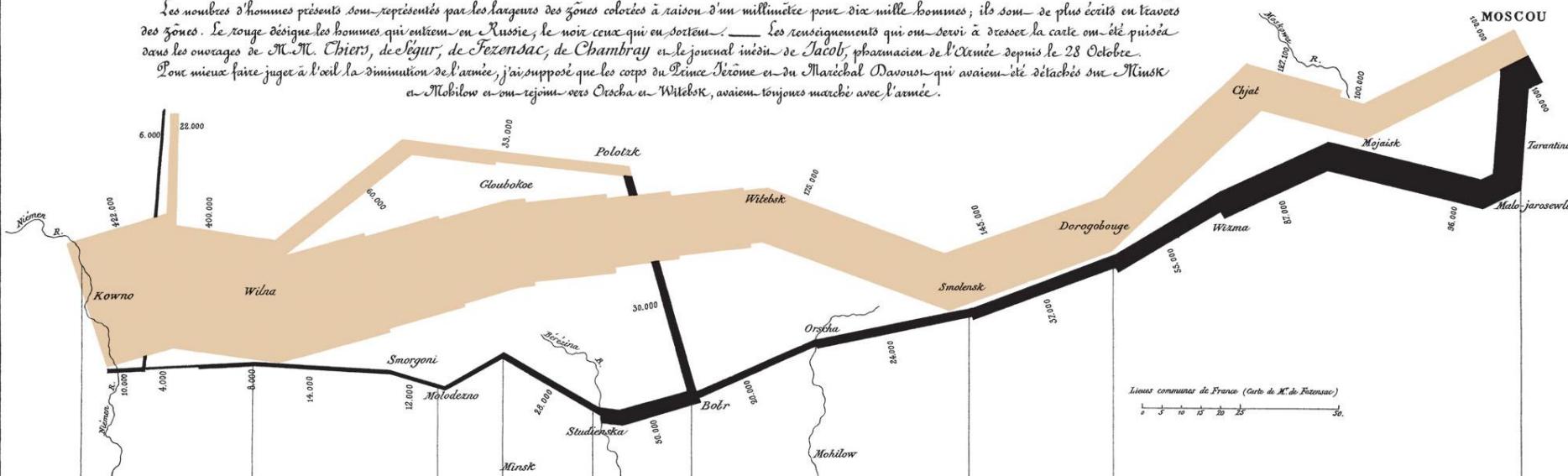
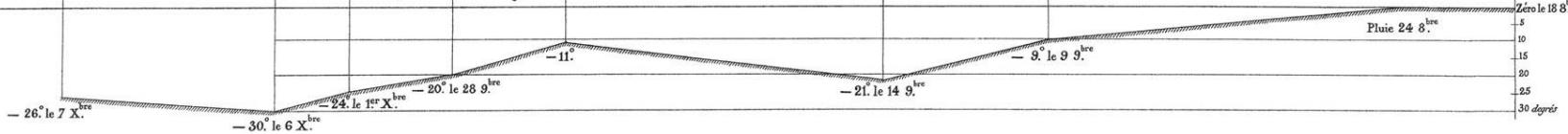


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Autog. par Regnier, S. Pas. S^e Marie S^e G^e à Paris.
Les cosaques passent au galop
le Niemen gelé.



Imp. Lith. Regnier et Duret.

Variable

- Longitude of the army's position
- Latitude of the army's position
- Size of Napoleon's Grande Armée
- Direction of the army's movement
- Date of points along retreat path
- Temperature during the army's retreat

Aesthetic

- x-axis
- y-axis
- Width of path
- Colour of path
- Text below plot
- Line below plot

Jeff Heer (2009) Brief History of Data Visualization

The screenshot shows a YouTube video player interface. At the top, there's a navigation bar with icons for refresh, home, and a lock, followed by the URL "youtube.com/watch?v=N00g9Q9stBo". Below the URL is the YouTube logo and a search bar. The main content area displays a presentation slide with a black header bar. The slide text reads:

Human-Computer Interaction Seminar
A Brief History of Data Visualization

March 6, 2009

Jeffrey Heer
Stanford University

Stanford | Center for Professional Development

At the bottom of the video player is a control bar with a progress bar showing 0:07 / 1:27:13, and standard video controls for play, volume, and settings.

A Brief History of Data Visualization

29.2 Make a figure for the generals

For the remainder of this chapter, I will discuss strategies for making individual figures and sets of figures that help your audience to connect with your story and remain engaged throughout your entire story arc. First, and most importantly, you need to show your audience figures they can actually understand. It is entirely possible to follow all the recommendations I have provided throughout this book and still prepare figures that confuse. When this happens, you may have fallen victim to two common misconceptions; first, that the audience can see your figures and immediately infer the points you are trying to make; second, that the audience can rapidly process complex visualizations and understand the key trends and relationships that are shown. Neither of these assumptions is true. We need to do everything we can to help our readers understand the meaning of our visualizations and see the same patterns in the data that we see. This usually means less is more. Simplify your figures as much as possible. Remove all features that are tangential to your story. Only the important points should remain. I refer to this concept as "making a figure for the generals."

For several years, I was in charge of a large research project funded by the U.S. Army. For our annual progress reports, I was instructed by the program managers to not include a lot of figures. And any figure I did include should show very clearly how our project was succeeding. A general, the program managers told me, should be able to look at each figure and immediately see how what we were doing was improving upon or exceeding prior capabilities. Yet when my colleagues who were part of this project sent me figures for the annual progress report, many of the figures did not meet this criterion. The figures usually were overly complex, were labeled in confusing, technical terms, or did not make any obvious point at all. Most scientists are not trained to make figures for the generals.



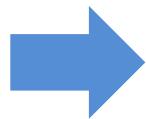
Never assume your audience can rapidly process complex visual displays.

Some might hear this story and conclude that the generals are not very smart or just not that into science. I think that's exactly the wrong take-home message. The generals are simply very busy. They can't spend 30 minutes trying to decipher a cryptic figure. When they give millions of dollars of taxpayer funds to scientists to do basic research, the least they can expect in return is a handful of clear demonstrations that something worthwhile and interesting was accomplished. This story should also not be misconstrued as being about military funding in particular. The generals are a metaphor for anybody you may want to reach with your visualization. It can be a scientific reviewer for your paper or grant proposal, it can be a newspaper editor, or it can be your supervisor or your supervisor's boss at the company you're working. If you want your story to come across, you need to make figures that are appropriate for all these generals.

Your visualisations should

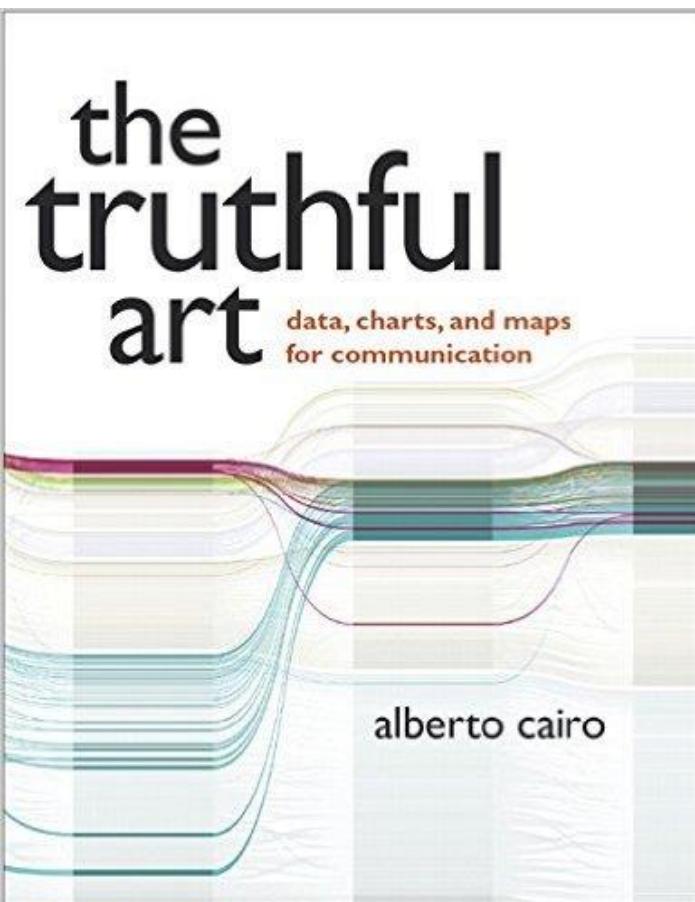
- Accurately reflect the data,
- Tell a story, and
- Look professional

Contents



- Introduction to visualisation
- Cairo's Five Principles
- C.R.A.P. Design principles
- Workshop
- Directory of visualisations

Five Principles of Visualisations



- 1. Truthful** – Be open to reasons that may contradict what you believe. Your work should be your best and most honest understanding of the reality you are trying to explain. Avoid lying to yourself and others.
- 2. Functional** – How you encode your data. What shapes, maps, techniques, etc. Be as clear as possible without oversimplifying; people are interested in complexity, so layer the information and allow your audience to satisfy their curiosity
- 3. Beautiful** – bring people's attention to relevant information and make them interested in issues that are relevant
- 4. Insightful** – Go beyond the obvious plots, and try to reveal interesting patterns in the data.
- 5. Enlightening** – Aim for helping people's understanding; change people's minds

2011 State of the Union

Truthful



2011 State of the Union

2010 Gross Domestic Product

In the 2010 Gross Domestic Product chart, the radius of each circle has been scaled based on the size of each nation's economy. This distorts the perception of the relative sizes of the circles since the radius scales linearly, but the area scales quadratically. Basically, the size of the United States economy appears much bigger than it should.

Truthful

The correct approach scales the area relative to the size of each economy, not the radius. This presents a more accurate view of the data based on the way we perceive circles. Below you can see the original chart (left) along with a corrected chart (right) which sizes each country based on the area rather than the radius.



Size = radius



Size = area

2016 US Elections results (1/3)

Truthful



Gideon Resnick

@GideonResnick

here's a pic from reuters of Trump with the electoral maps he showed to reporters yesterday



4:24 PM - 28 Apr 2017

730 Retweets 933 Likes



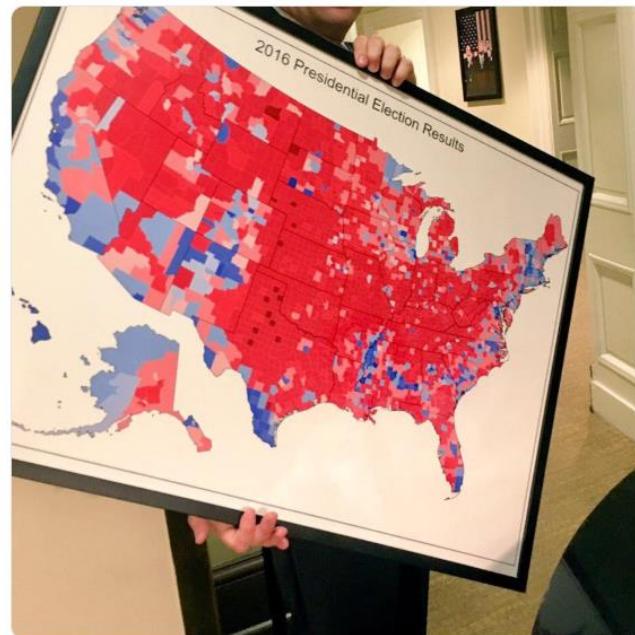
<https://twitter.com/GideonResnick/status/857948601984614405>



Trey Yingst

@TreyYingst

Spotted: A map to be hung somewhere in the West Wing



5:03 PM - 11 May 2017

3,944 Retweets 8,370 Likes



<https://twitter.com/treyyingst/status/862669407868391424>

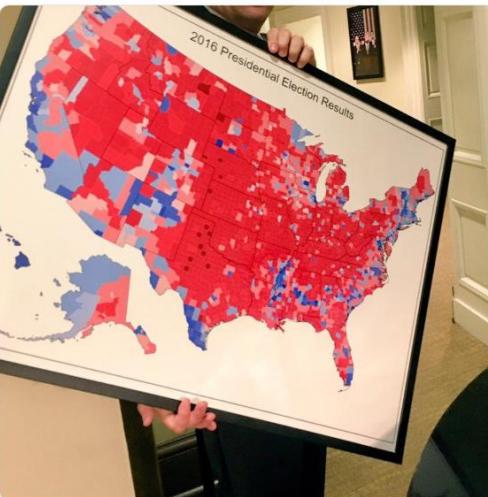
2016 US Elections results (2/3)

Truthful



Trey Yingst

Spotted: A map to be hung somewhere in the West Wing



5:03 PM - 11 May 2017

3,944 Retweets 8,370 Likes



Surface on the county-level map

Red 80%

Blue 20%

2016 United States presidential election

2016 United States presidential election		
← 2012 November 8, 2016 2020 →		
538 members of the Electoral College 270 electoral votes needed to win		
Turnout	55.7% ^[1] 0.8 pp	
Nominee	Donald Trump	Hillary Clinton
Party	Republican	Democratic
Home state	New York	New York
Running mate	Mike Pence	Tim Kaine
Electoral vote	304 ^[a]	227 ^[a]
States carried	30 + ME-02	20 + DC
Popular vote	62,984,828	65,853,514
Percentage	46.1%	48.2%

2016 US Elections results (3/3)



Kenneth Field

@kennethfield

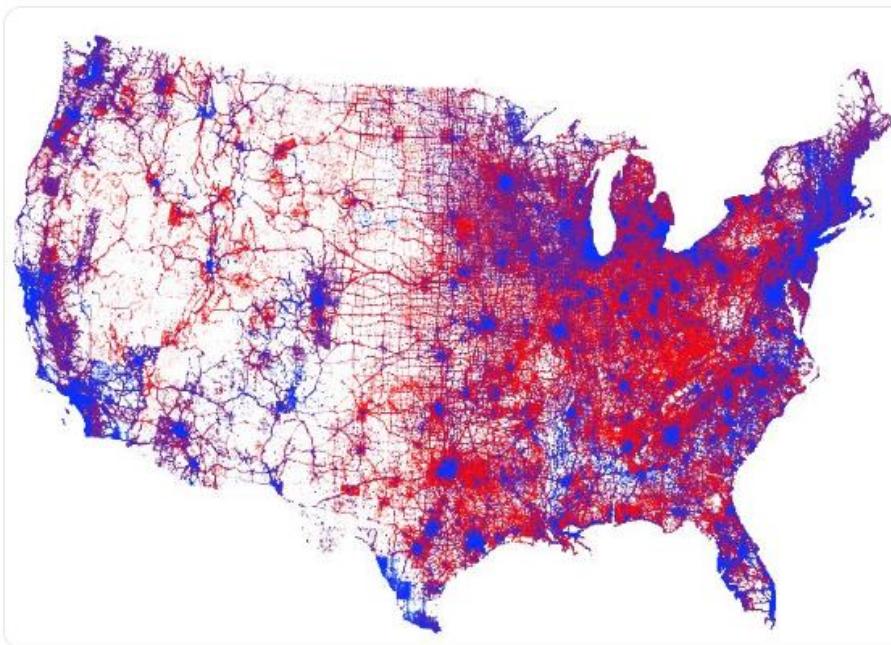
Truthful

Fair bit of data wrangling but finally...a 1 dot
= 1 vote dasymetric dot density 2016

Presidential election map from [@ArcGISPro](#).

65,844,61 blue dots. 62,979,636 red dots.

Count 'em! (note: this is just a rough
screengrab, will webify & make pretty)



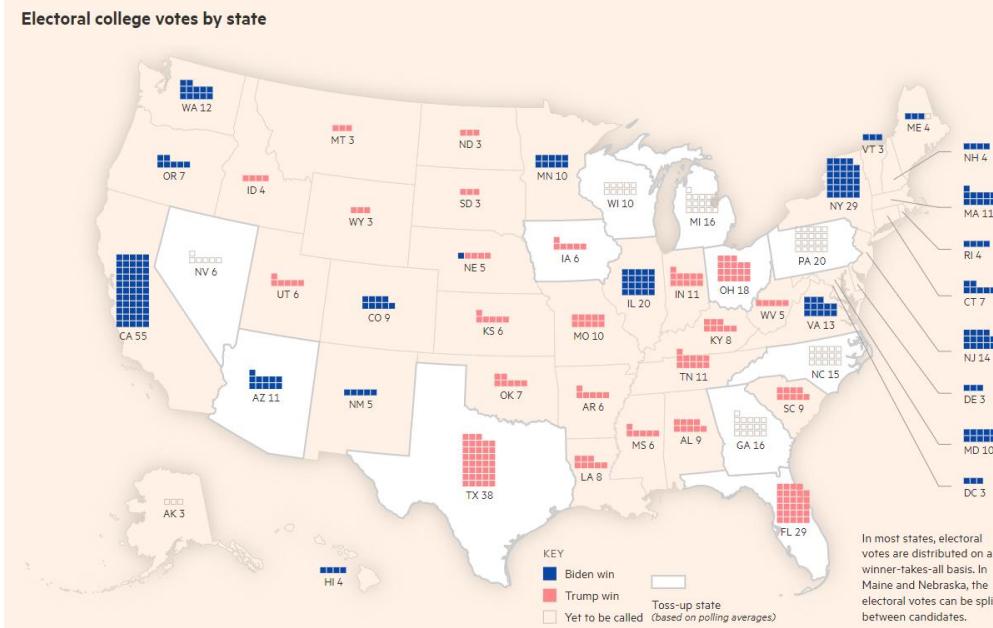
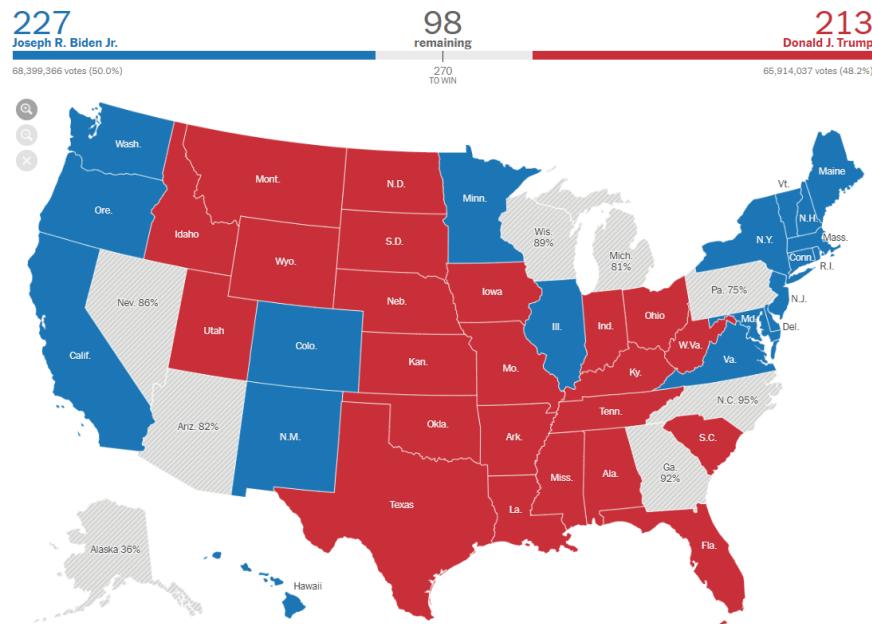
3:04 AM - 6 Mar 2018

4,336 Retweets 10,093 Likes

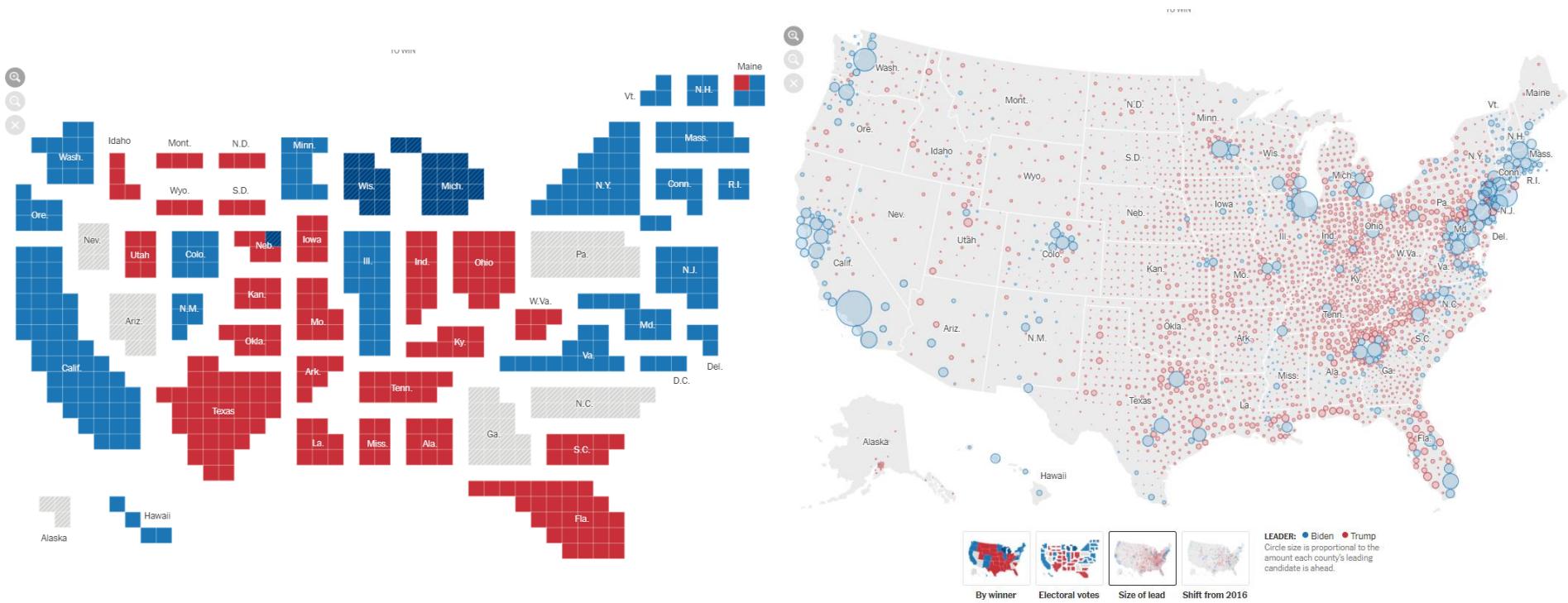


<https://twitter.com/kennethfield/status/970827334038237184>

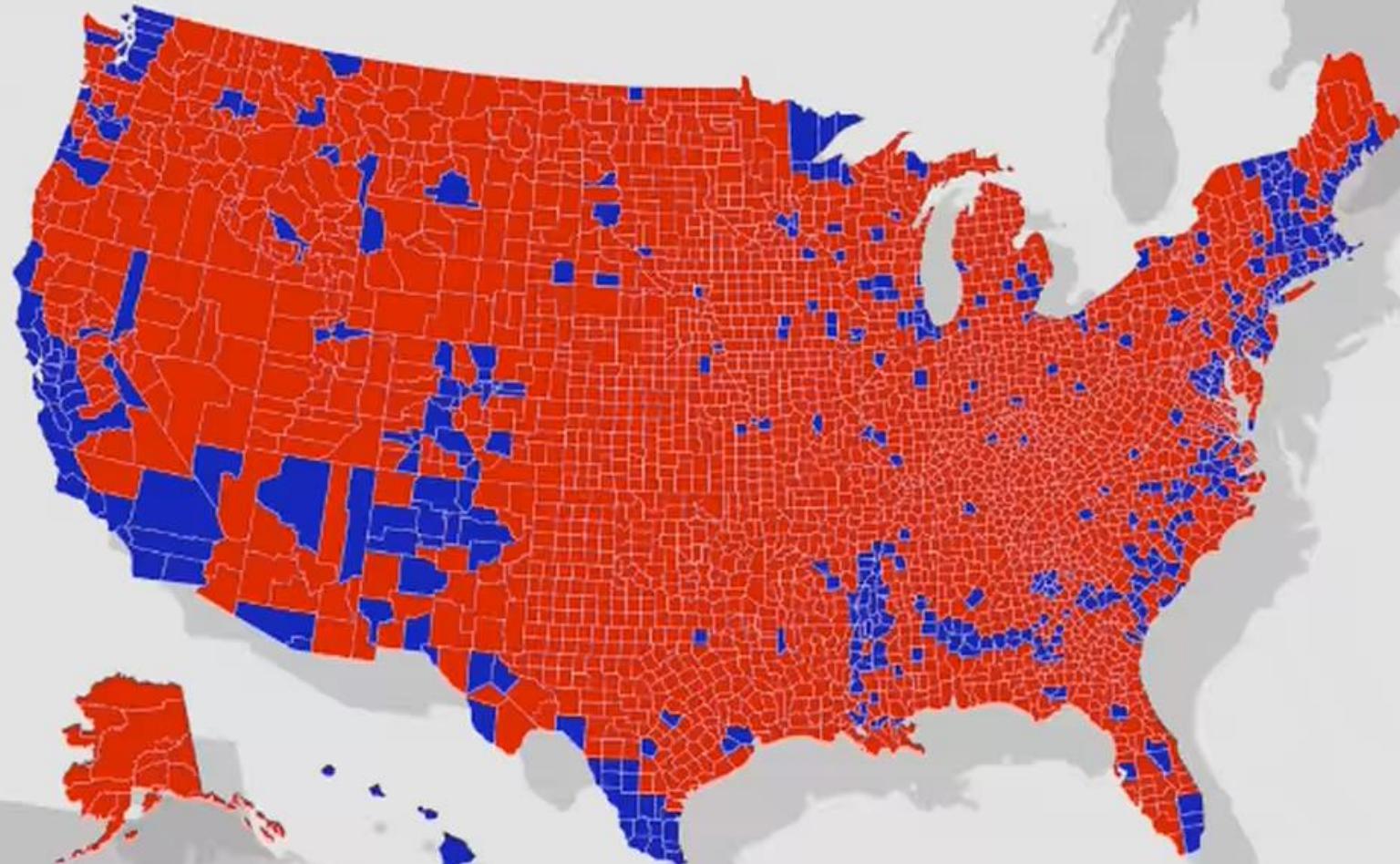
2020 US Elections results



2020 US Elections results

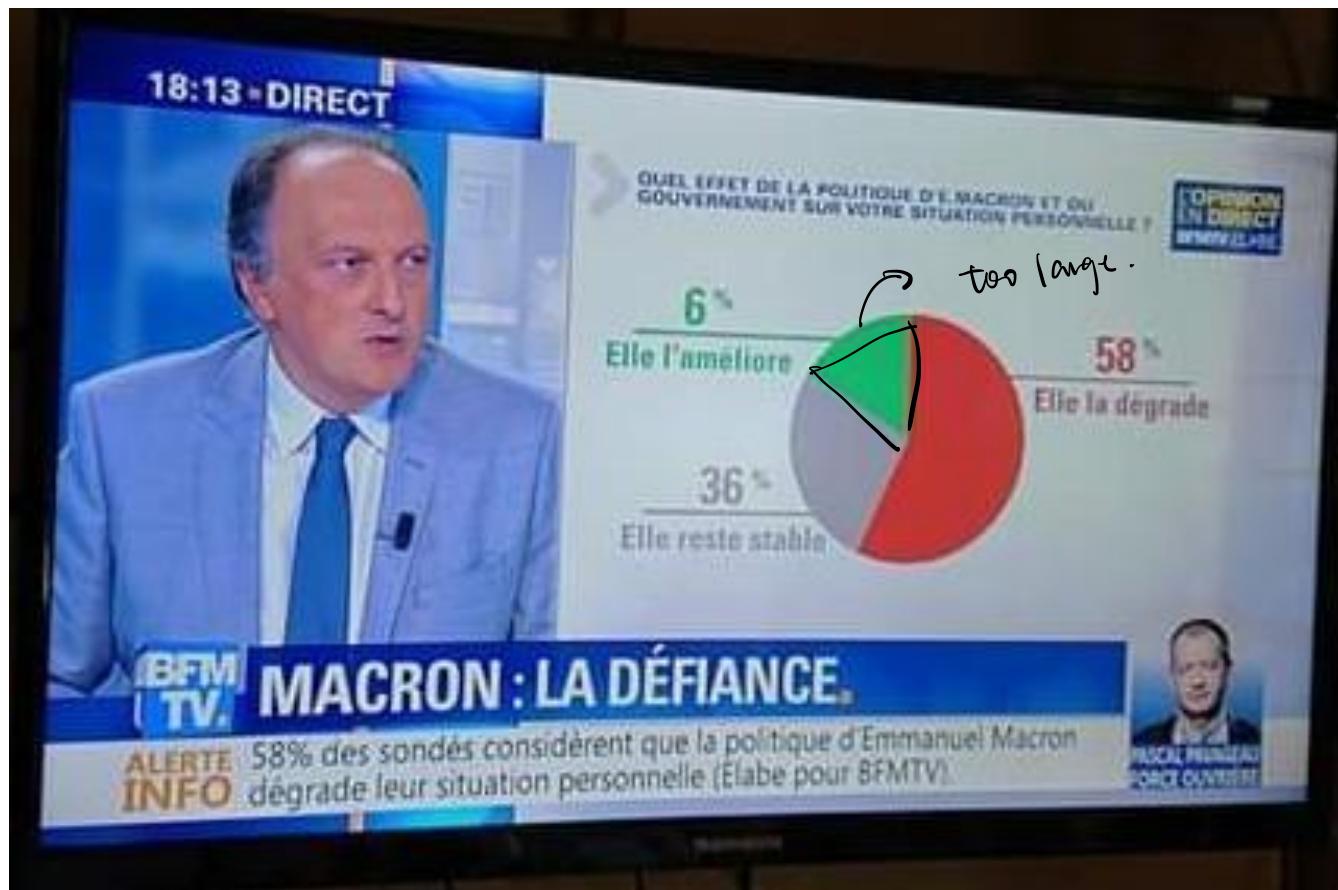


2020 US Elections results



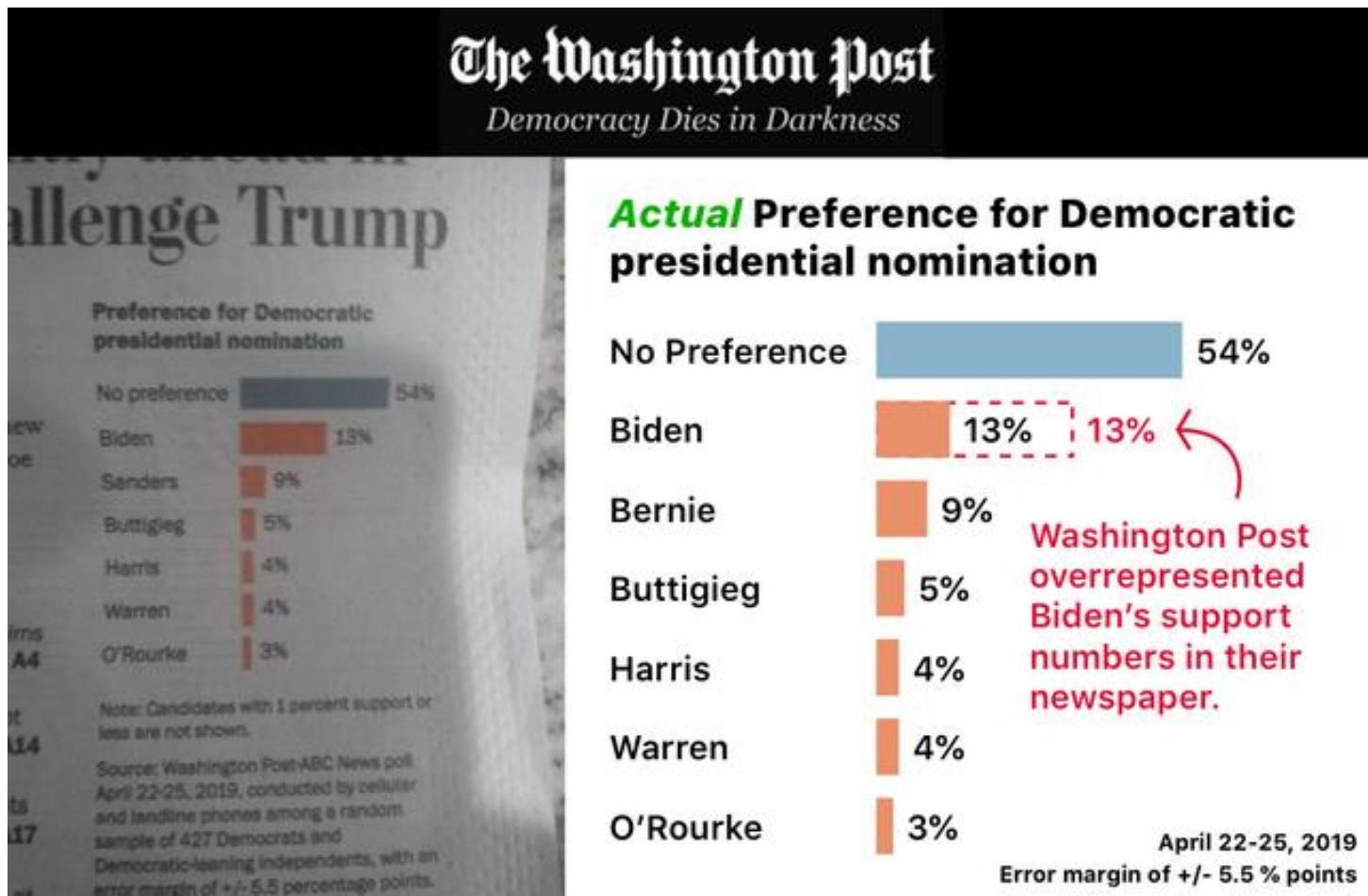
Macron's Popularity, August 2018

Truthful



Preference for Democratic presidential nomination, Apr 2019

Truthful



Lots of confusion comes from double-axis! 1. Truthful

PRODUCTION COSTS

U.S. dollars per unit

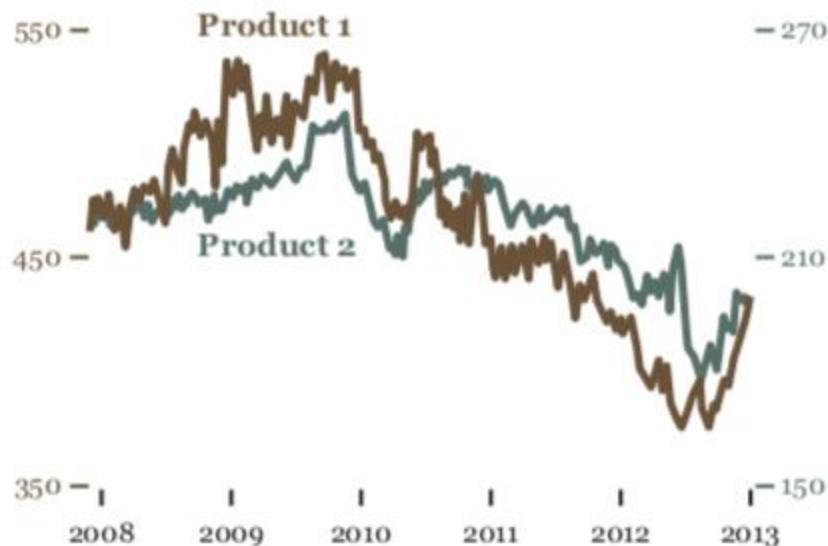


Figure 2.4 Dual-axis charts can be easily misinterpreted.

PRODUCTION COSTS

U.S. dollars per unit

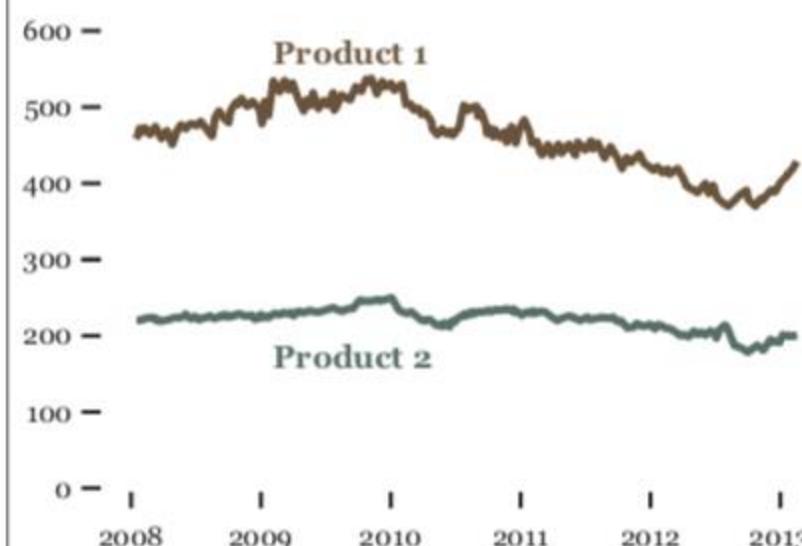


Figure 2.5 The same data, plotted on the same scale. As a general rule, avoid dual-axis charts.

should avoid !

"If you double the axes, you can double the mischief. Using two vertical axes and omitting zero from either or both opens a statistical beauty parlor with many cosmetic possibilities" (Cairo, Ch. 2)



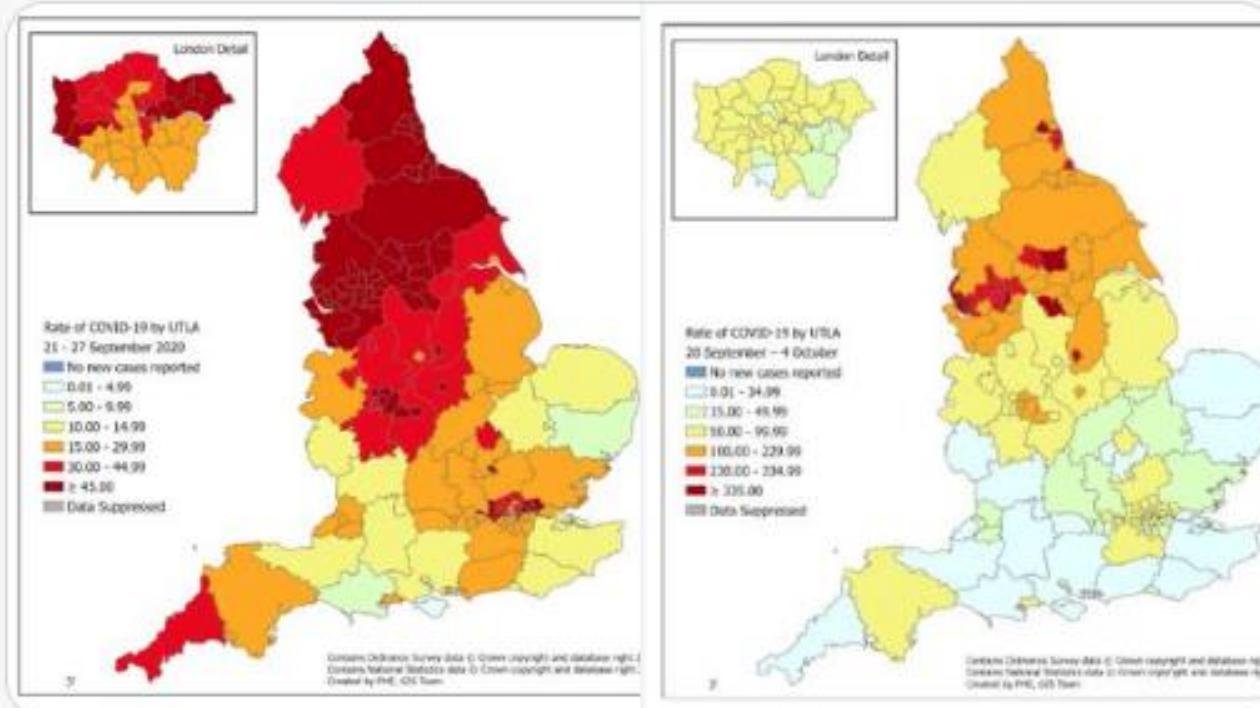
Alex Andreou @sturdyAlex · Oct 9

...

Comparing the gov't Covid19 maps for end of September and start of October, it looks as if things are getting better.

UNTIL you notice they've changed the numbers corresponding to each colour. Had they used the same ones, most of the country would be red or dark red.

#BitSneaky



444

9K

15.5K

↑

30

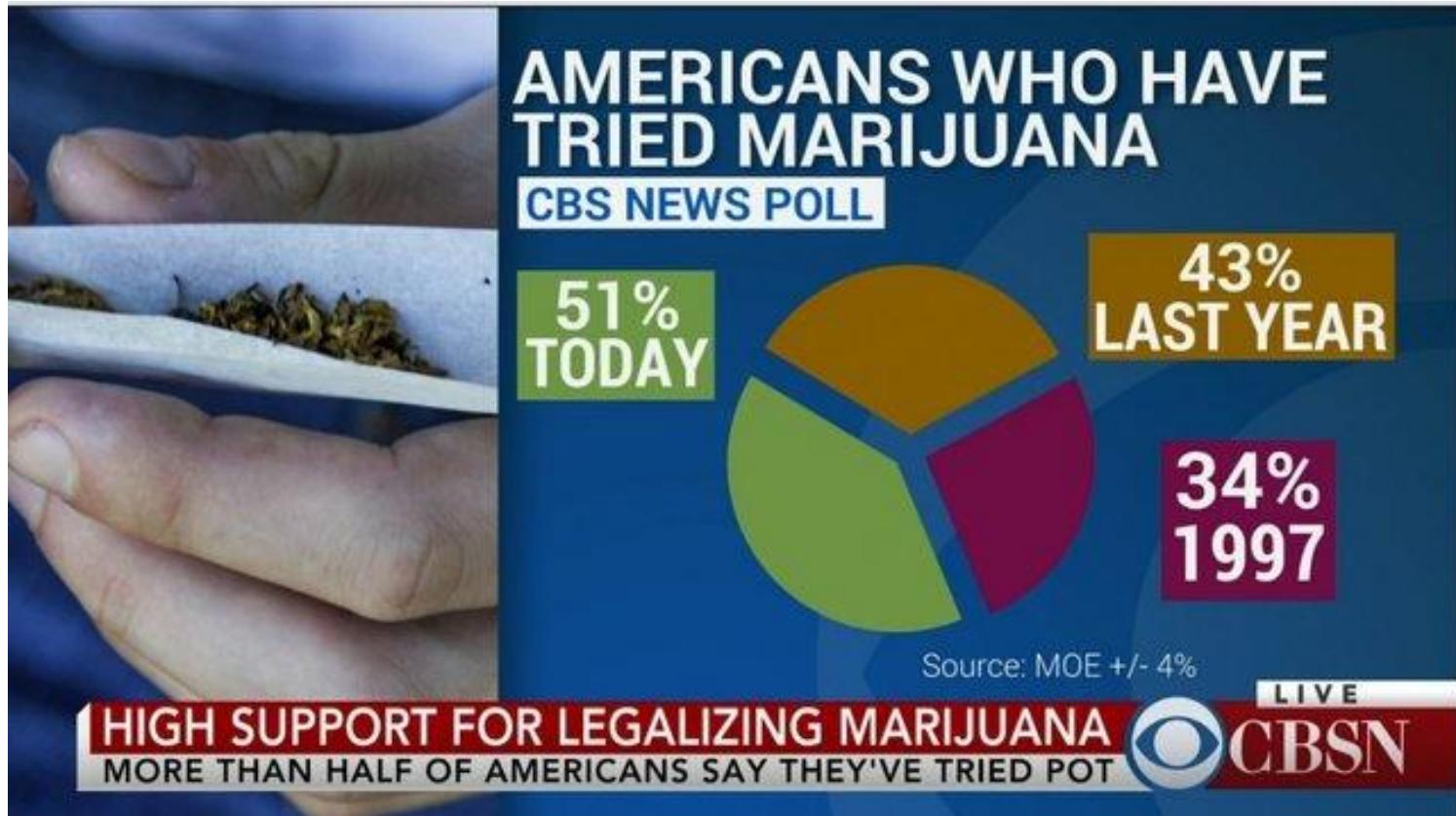
You do not always need a visualisation

Functional



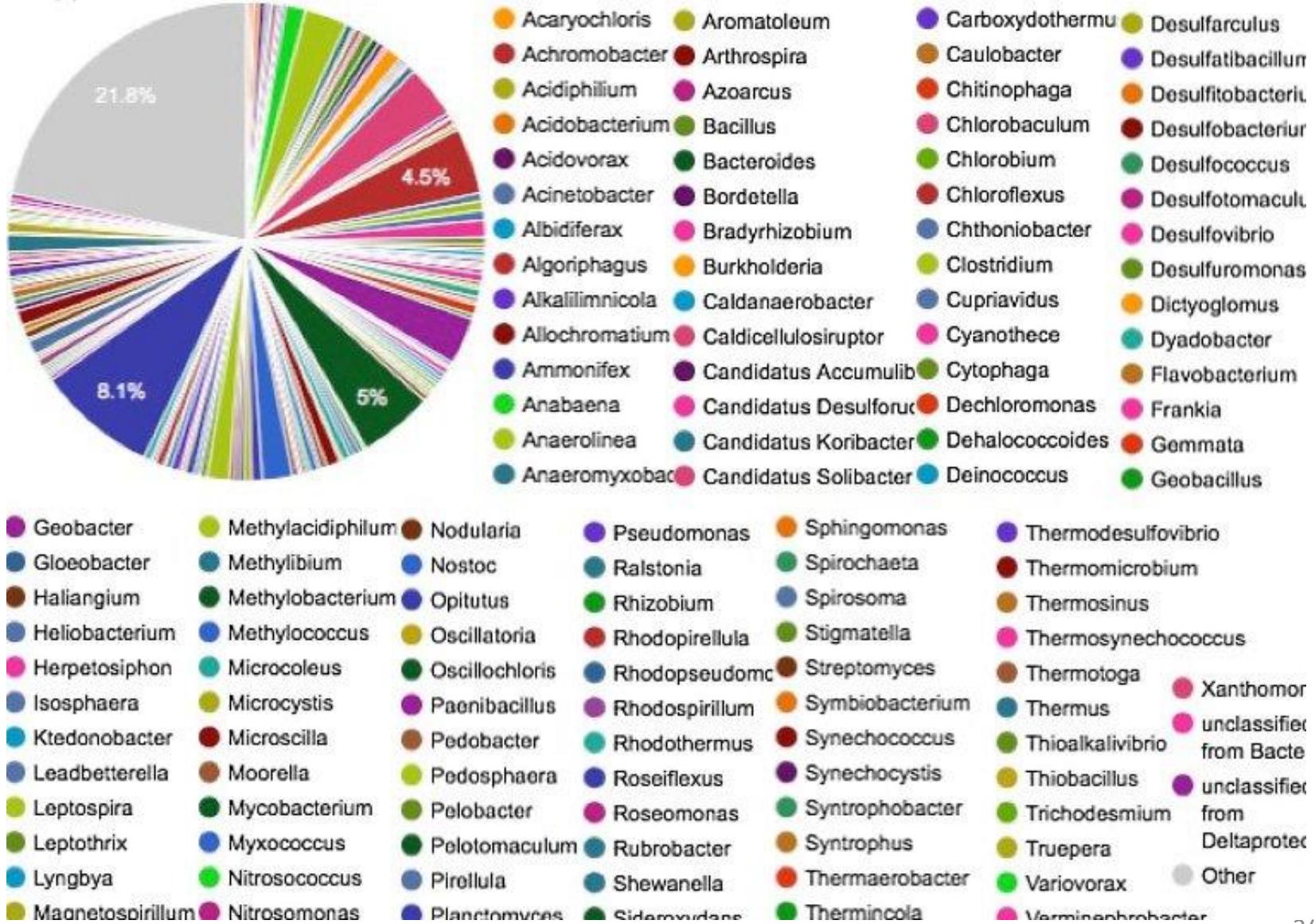
What's wrong with this pie chart?

Functional



What's wrong with this pie chart?

(f) Distribution of Genus



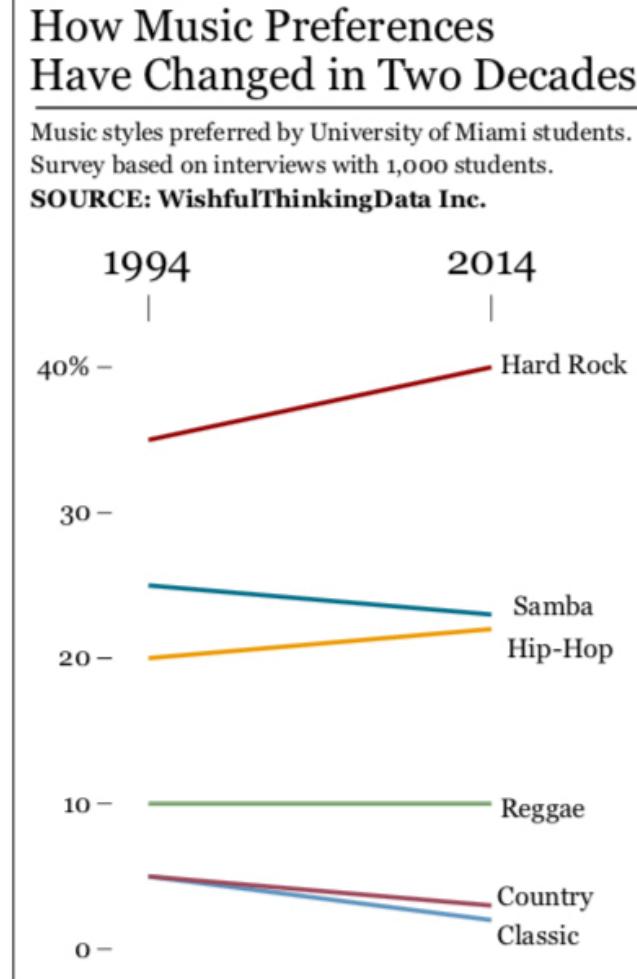
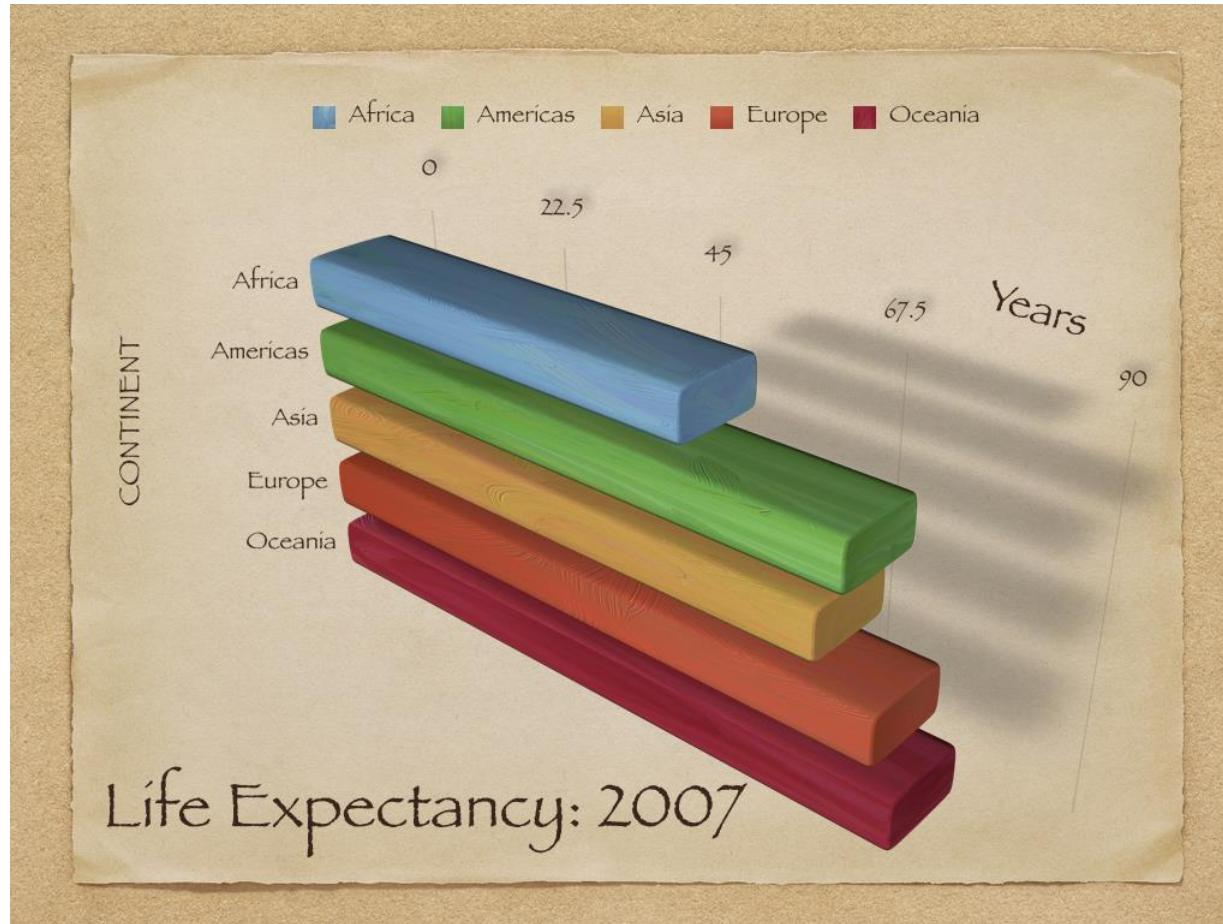


Figure 2.7 A slope chart is much better to represent change between two points in time.

Is 3D really needed?

Functional



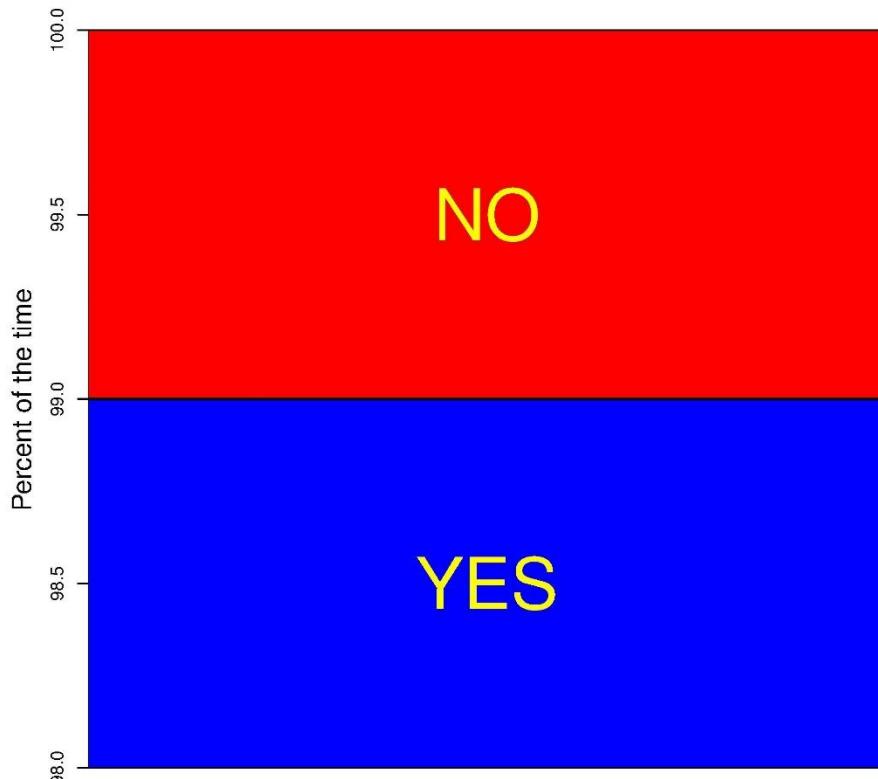
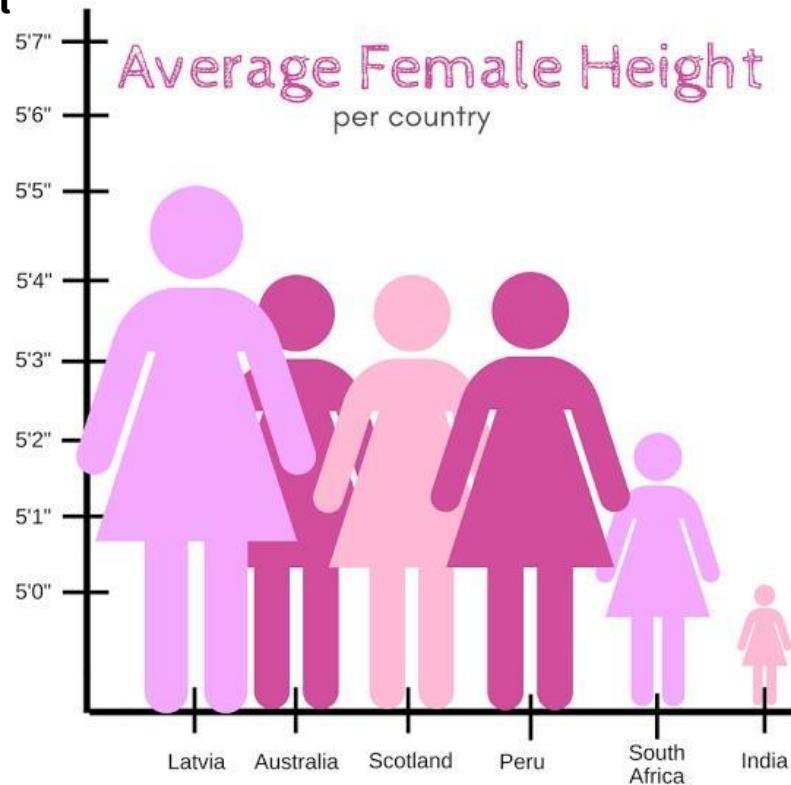
Truncating the Y-axis

But is not suggested.



Is truncating the Y-axis misleading?

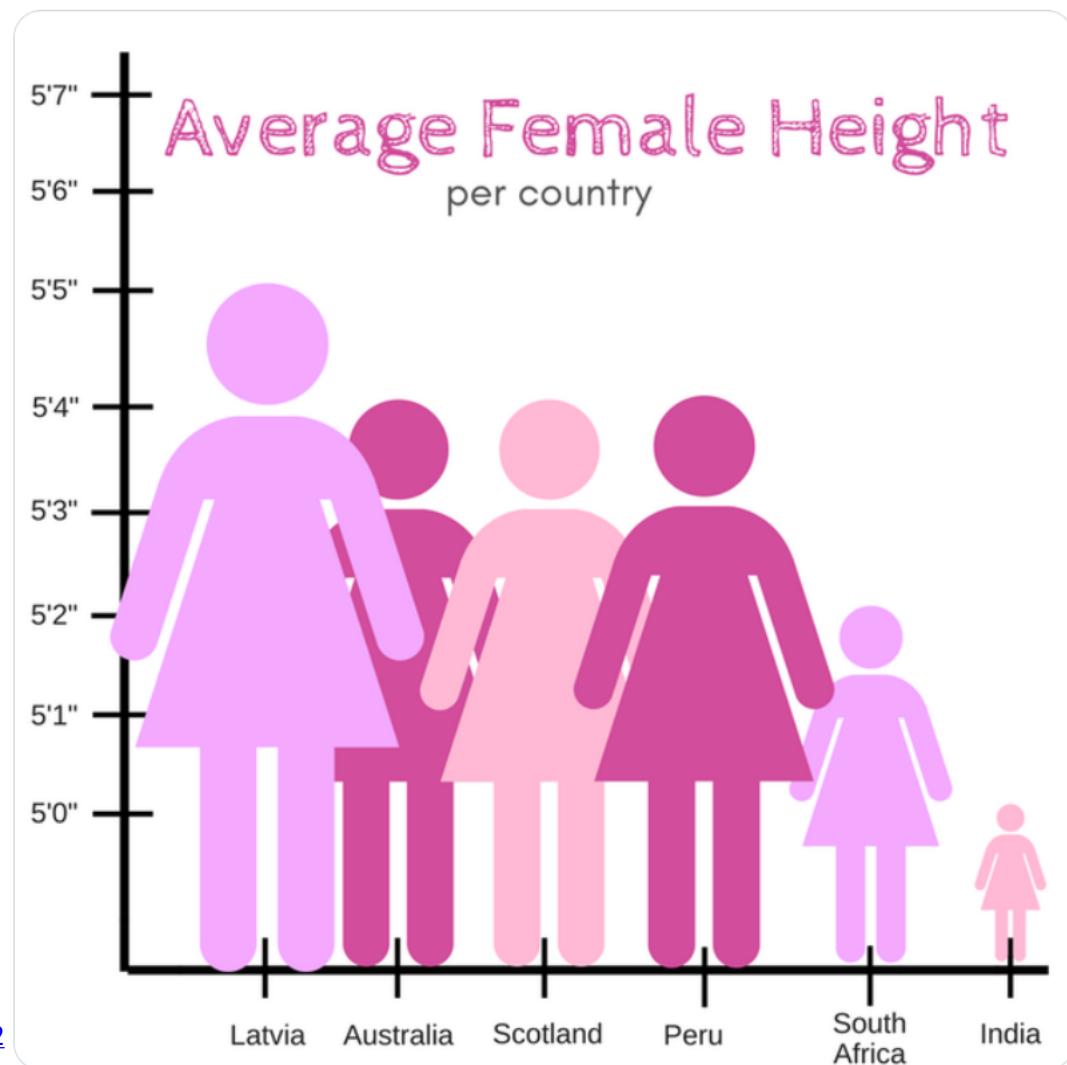
Functional





Sabah Ibrahim
@reina_sabah

As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me



3. Beautiful

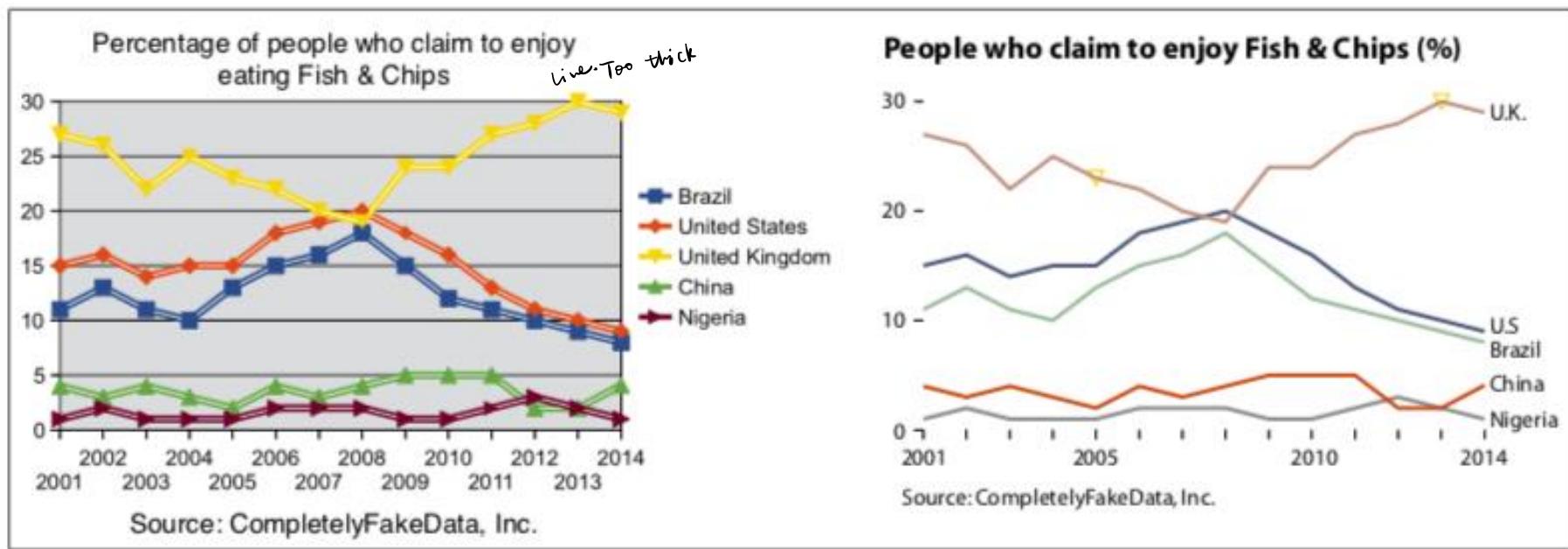


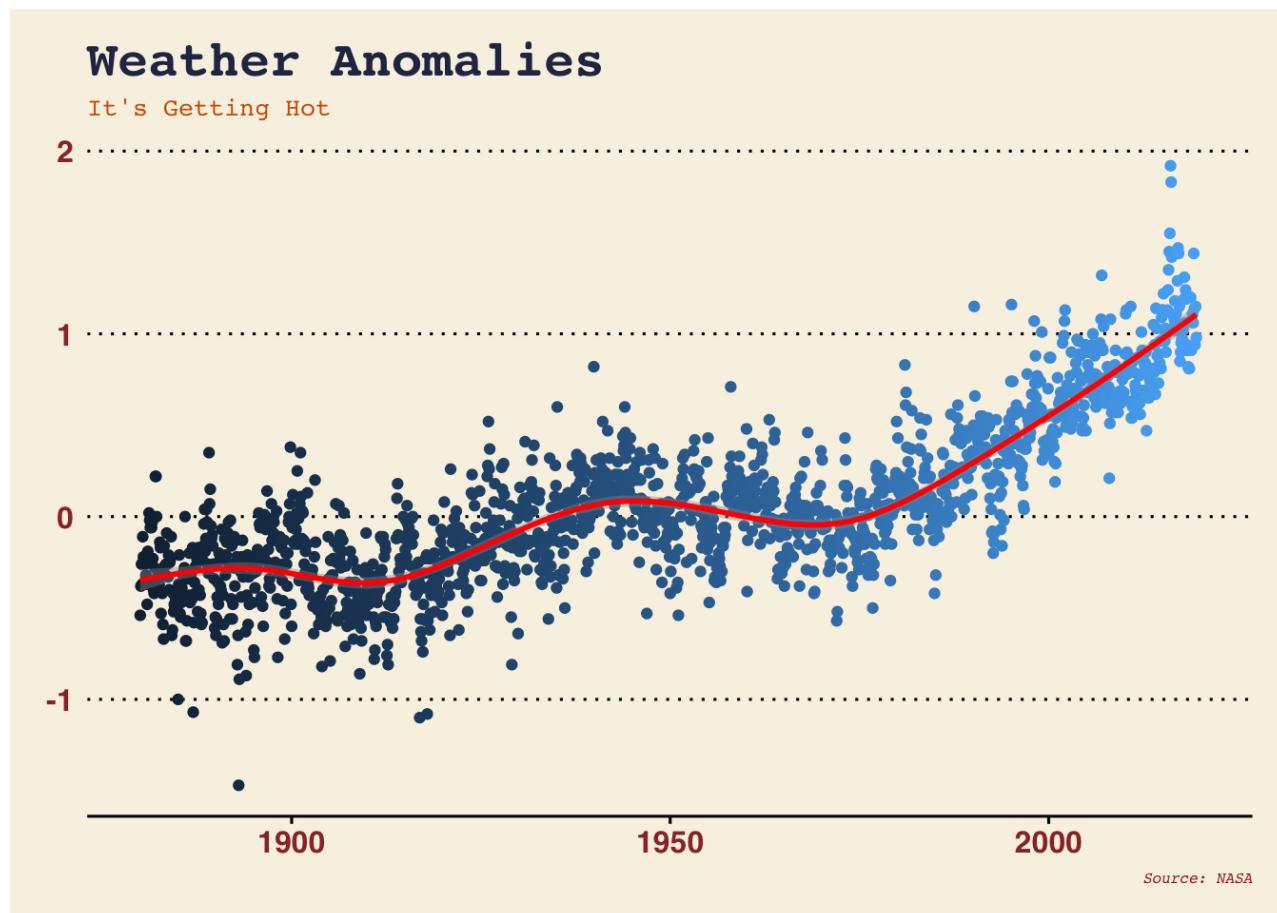
Figure 2.11 Which chart is more aesthetically pleasing?

4. Insightful

The purpose of visualization is insight, not pictures
Card, Mackinley, and Shneiderman (1999)

Two types of insight in data visualizations (Chang et al., 2009):

1. Eureka (or Aha!) moments
2. Knowledge-building



5. Enlightenment

Global Ph.D.s Gender Gap (2010)



Cairo's suggestions for visualizations

1. Think about the task(s) you want to enable
 - What is the issue?
 - What action (task) can be taken to address the issue?
 - What data do you have?
 - What are the most important data elements?
2. Try different graphic forms
3. Arrange the components of the graphic
4. Test the outcomes

Try different graphic forms

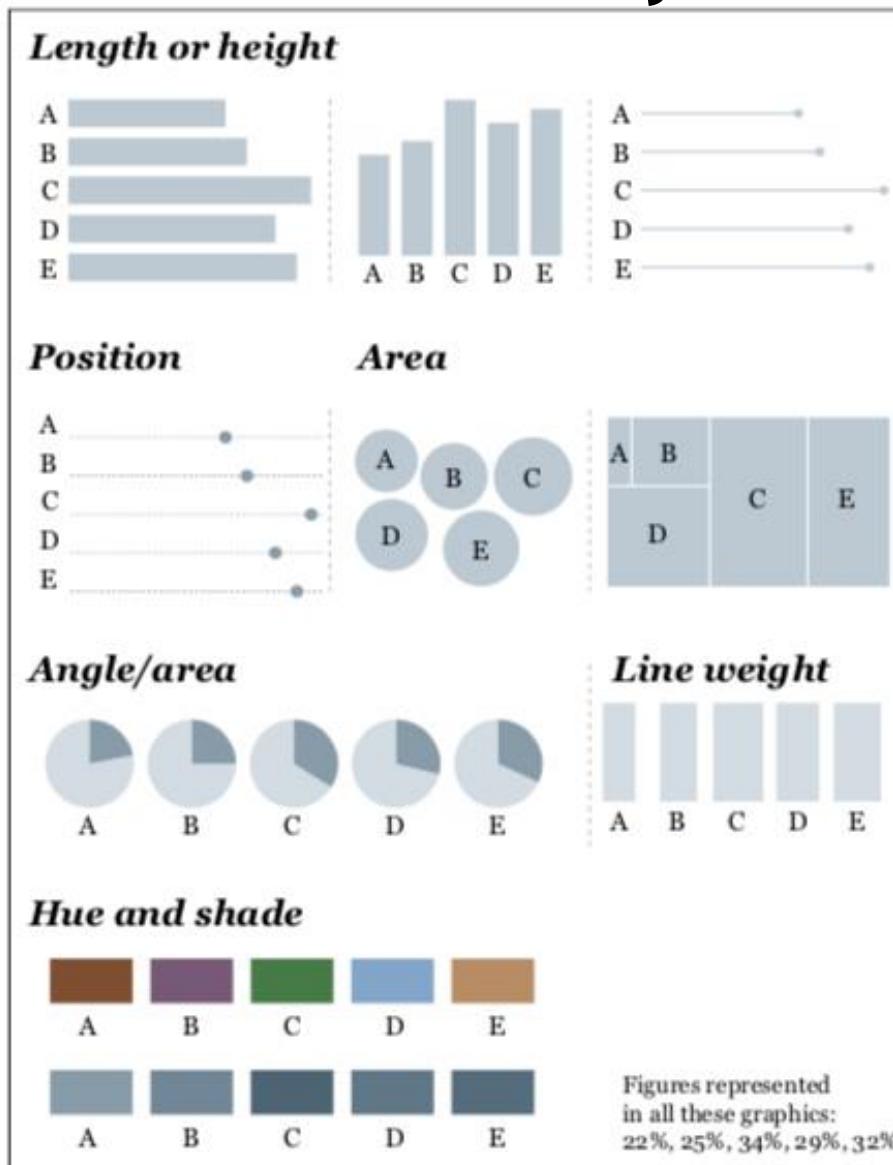
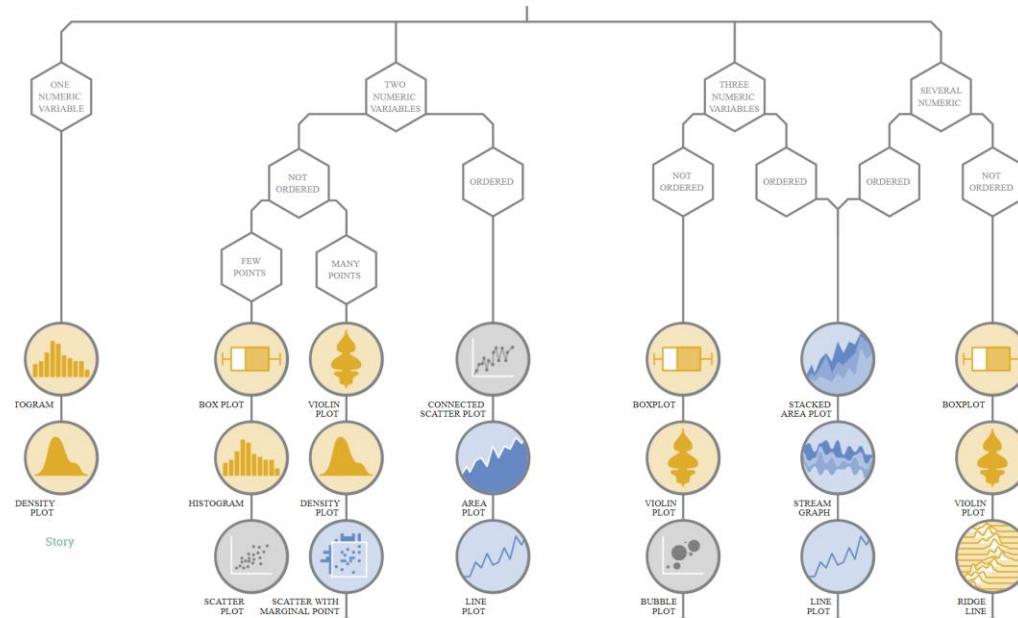
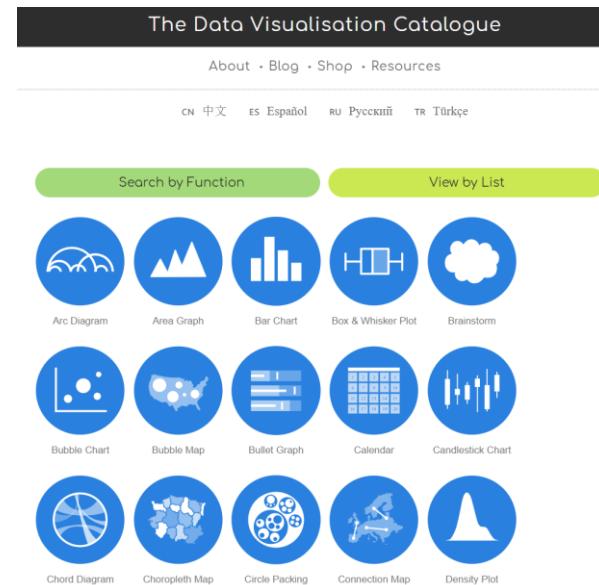


Figure 5.2 Different methods of encoding the same small data set. Remember that, perhaps because our client requested it, countries are organized alphabetically. Otherwise, it'd make more sense to arrange the figures from largest to smallest.

Choosing graphic forms



<https://www.data-to-viz.com/>



<https://datavizcatalogue.com/>

A collection of caveats



Order your data

When displaying the value of several entities, ordering them makes the graph much more insightful.



To cut or not to cut?

Cutting the Y-axis is one of the most controversial practice in data viz. See why.



The spaghetti chart

A line graph with too many lines becomes unreadable: it is called a spaghetti graph.



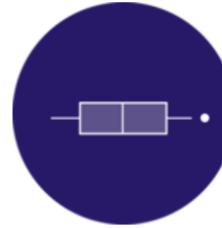
Pie chart

The human eye is bad at reading angles. See how to replace the most criticized chart ever.



Play with histogram bin size

Always try different bin sizes when you build a histogram, it can lead to different insights.



Do boxplots hide information?

Boxplots are a great way to summarize a distribution but hide the sample size and their distribution.



The problem with error bars

Barplots with error bars must be used with great care. See why and how to replace them.



Too many distributions.

If you need to compare the distributions of many variables, don't clutter your graphic.

Arrange components of chart

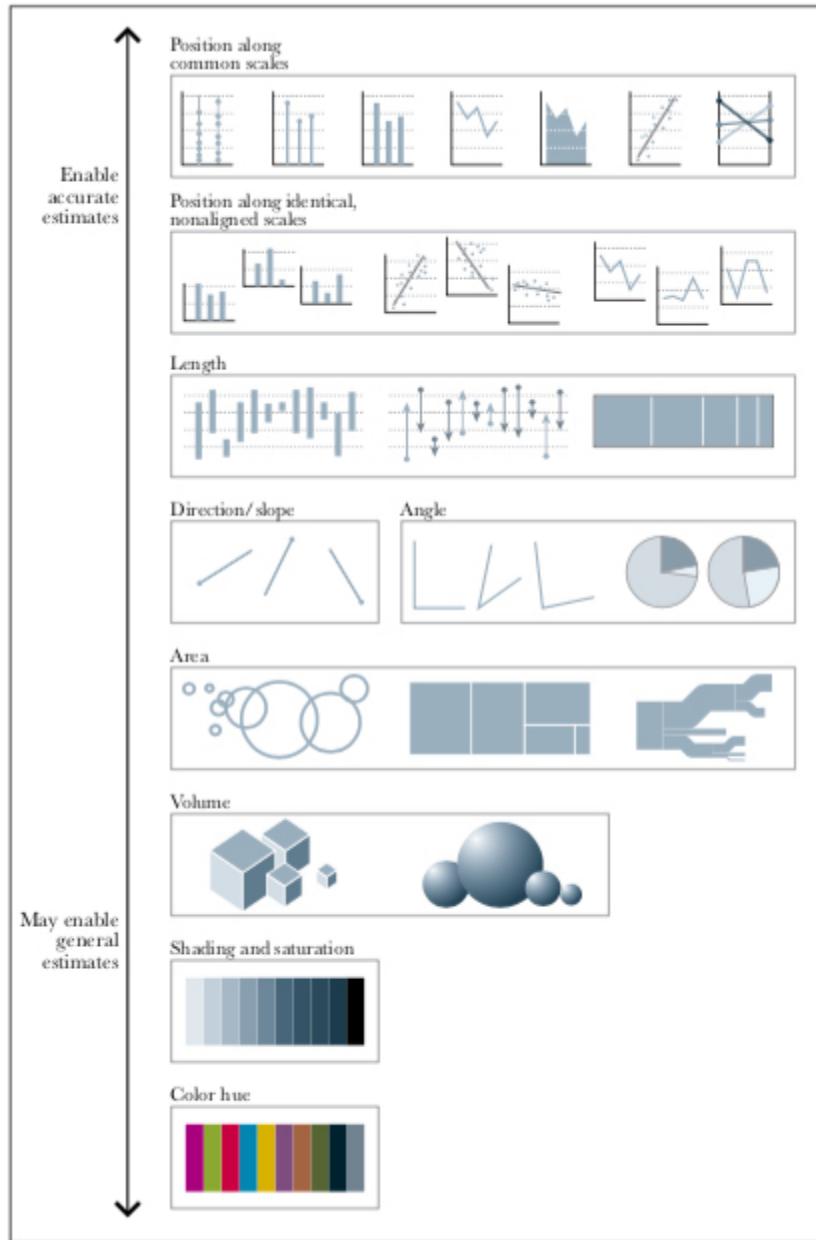


Figure 5.5 Scale of elementary perceptual tasks, inspired by William Cleveland and Robert McGill.

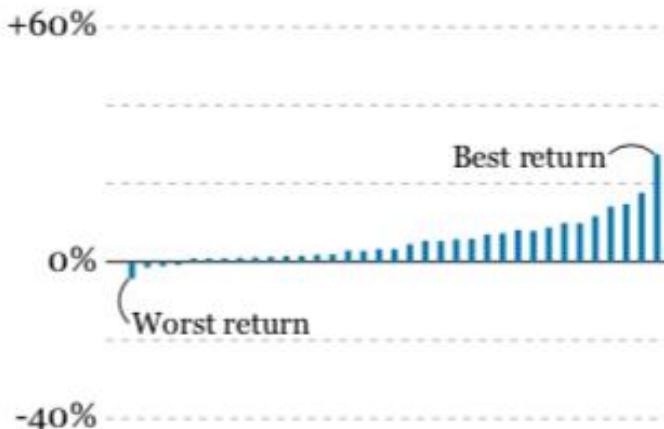
Test the Outcomes

Distribution of one-year returns

Fund A: High risk, high return

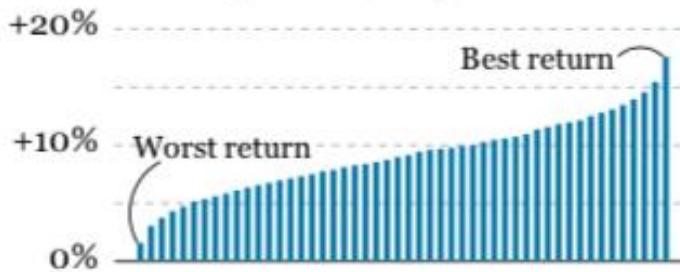


Fund B: Low risk, low return



Distribution of average annual returns over 30 years

Fund A: High risk, high return



Fund B: Low risk, low return



Figure 5.22 Charts based on Richard H. Thaler's *Misbehaving: The Making of Behavioral Economics* (2015).

How (not) to create an infographic

Step By Step Guide: How To Create An Infographic



Visually

published on September 8, 2014 in Data

Share on

Share on

Share on

Creating an infographic can be a daunting task if you've never built one before. There is a long list of things that need to be done, and it takes a varied skill set to accomplish all of them. Visually teamed up with BlockSpring and Enigma to provide an example of what the process is like.

Pick a Topic

The first step in creating an infographic is to decide on a story you want to tell. One great way to go about this is to decide why you want to tell that story. Do you want to showcase your company's capabilities? Do you simply want to share something cool or interesting? Are you reporting on accomplishments or performance? In this case we wanted to share something interesting: the aid policy of the United States.

Get the Data

The next step is to figure out the exact story and make sure you have data that supports it. Enigma is a fantastic way to navigate the world of public data – quickly searching and analyzing through billions of public records. We pulled the Creditor Reporting System simply by using the topics browser and applying "International"

Don't come up with the key message first and then try to find data that support your message. You should try to find data that can refute your message.

If you find through analysis that your message is wrong, you should change your message, not the data!

Data exploration versus data presentation

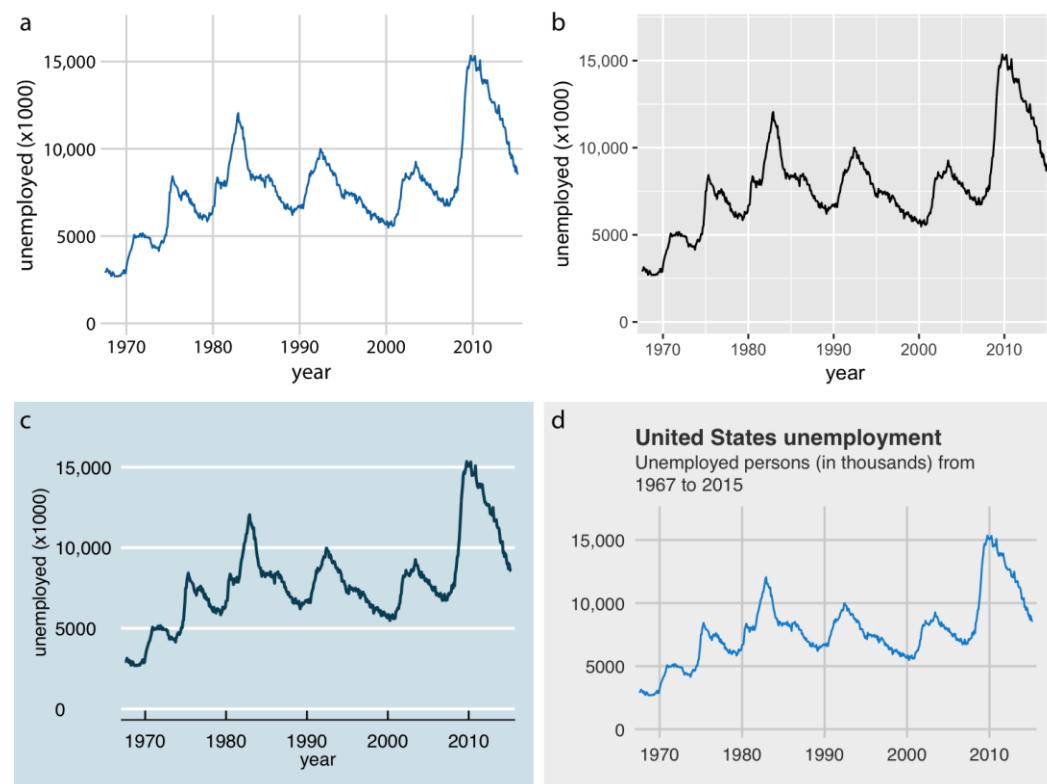
Data exploration: look at a dataset from different angles and try various ways of visualizing it. Try different visualizations, different data transformations, and different subsets of the data. It's fine if axis labels are missing, the legend is messed up, or the symbols are too small.

Data presentation. You enter presentation mode, once you know what aspects of the data you want to show to your audience. The key objective in this phase is to prepare a high-quality, publication-ready figure that can be printed in an article or book, included in a presentation, or posted on the internet.

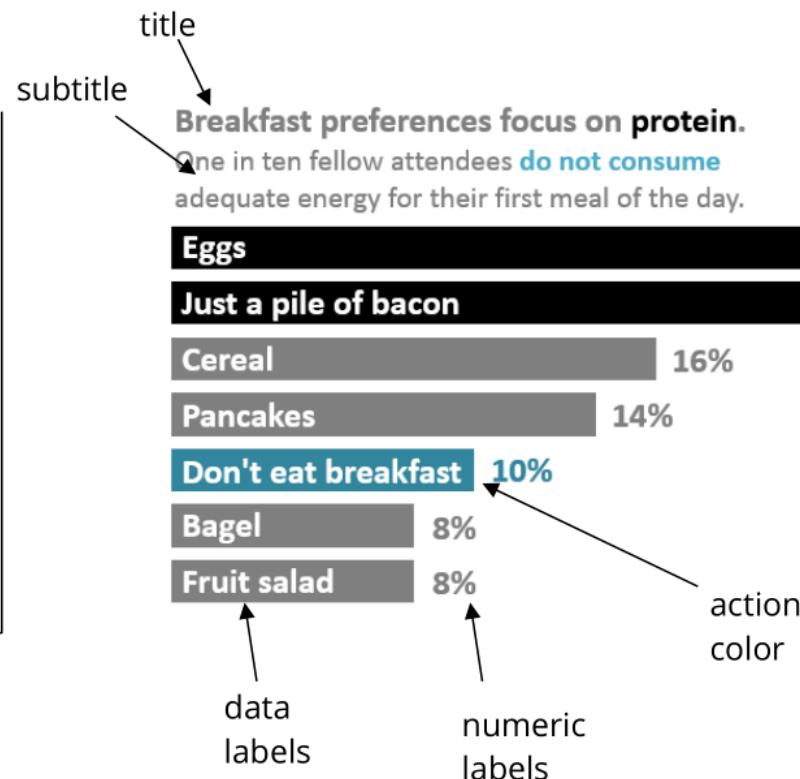
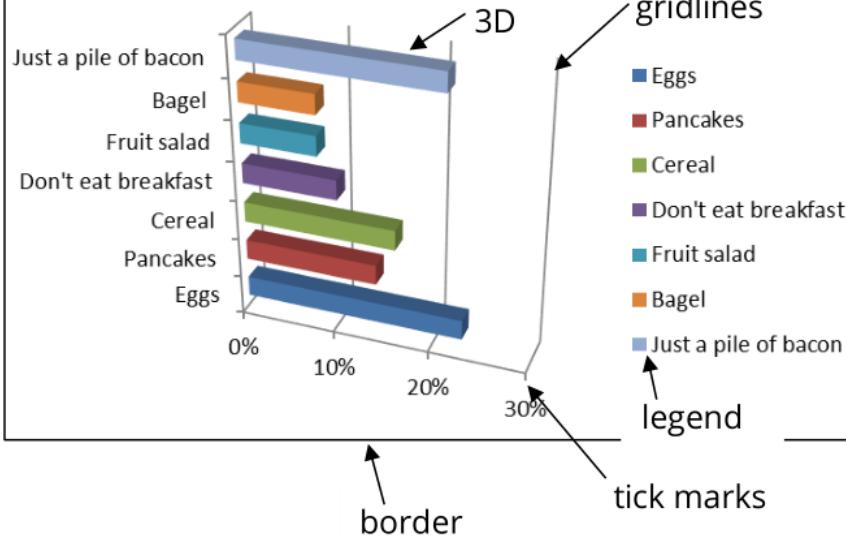
Separation of content and design

Content: Data used, data transformations applied (if any), specific mappings from data onto aesthetics, scales, axis ranges, and the type of plot (scatter plot, line plot, bar plot, boxplot, etc.)

Design. Foreground and background colors, font specifications (e.g. font size, face, and family), symbol shapes and sizes, the placement of legends, axis ticks, axis titles, and plot titles, and whether or not the figure has a background grid



Attendee Breakfast Preferences



Putting it all together

- Information
 - Understand your data and be truthful
 - Your data may never be a perfect reflection of the world
 - The best use of data is to teach us what is NOT true
- Story
 - Be clear about the message of your visualization
 - Who is my audience? What am I trying to achieve?
 - Pick visualisations that successfully make your point
- Visual Form/Design
 - Design rules and visualisation principles

twitter.com/tladeras/status/1449171307770253314

← Thread



Ted Laderas, PhD @tladeras

...

If there is one lesson I can teach future analysts:

If you want someone to draw a conclusion from a figure, state that conclusion.

Don't expect that person to "just get it" from that figure.

In fact, make that conclusion the title of your figure.

1:32 AM · Oct 16, 2021 · TweetDeck

68 Retweets 10 Quote Tweets 715 Likes



Ted Laderas, PhD @tladeras · Oct 16

...

Replying to @tladeras

Some think this is "cheating" and everything should be evident from the figure.

Nope. Everyone is tired and cranky and wants to know your point.

You are paid to do the interpreting and to support that with evidence.



3



5



95



Ted Laderas, PhD @tladeras · Oct 16

...

You're not a figure making machine. You're a conclusion generating human.

That is all.



1



4



78





Jennifer Thompson
@jent103

Following



Ratio of time spent "doing analysis" to time spent interpreting & communicating the meaning of said analysis in a way that makes sense to the audience: Currently at 1:100.

4:29 PM - 25 Aug 2018

36 Retweets 157 Likes



36



157



Ensure your audience understand, enjoy, and remember your work.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 22, NO. 1, JANUARY 2016

Beyond Memorability: Visualization Recognition and Recall

Michelle A. Borkin*, Member, IEEE, Zoya Bylinskii*, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, Senior Member, IEEE, and Aude Oliva

EXPERIMENT DESIGN

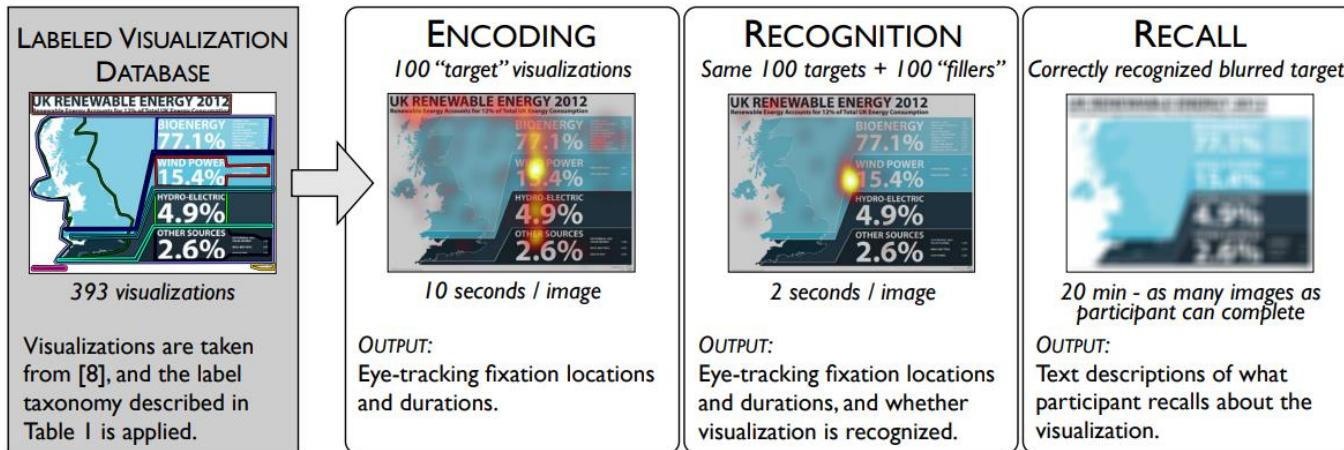


Fig. 1. Illustrative diagram of the experiment design. From left to right: the elements of the visualizations are labeled and categorized, eye-tracking fixations are gathered for 10 seconds of “encoding”, eye-tracking fixations are gathered while visualization recognizability is measured, and finally participants provide text descriptions of the visualizations based on blurred representations to gauge recall.

Abstract— In this paper we move beyond memorability and investigate how visualizations are recognized and recalled. For this study we labeled a dataset of 393 visualizations and analyzed the eye movements of 33 participants as well as thousands of participant-generated text descriptions of the visualizations. This allowed us to determine what components of a visualization attract people’s attention, and what information is encoded into memory. Our findings quantitatively support many conventional qualitative design guidelines, including that (1) titles and supporting text should convey the message of a visualization, (2) if used appropriately, pictograms do not interfere with understanding and can improve recognition, and (3) redundancy helps effectively communicate the message. Importantly, we show that visualizations memorable “at-a-glance” are also capable of effectively conveying the message of the visualization. Thus, a memorable visualization is often also an effective one.

Index Terms—Information visualization, memorability, recognition, recall, eye-tracking study

Three questions for creating charts

1. What's your point?

Headline

HYPOTHESIS

2. How can you emphasize your point in your chart?

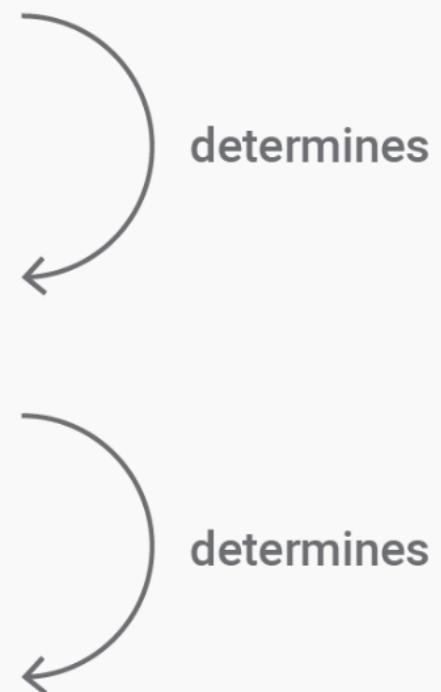
Chart type, data, color, highlights, annotations

PROOF

3. What does the final chart show exactly?

Description, legend/keys, source

EXPLAINING

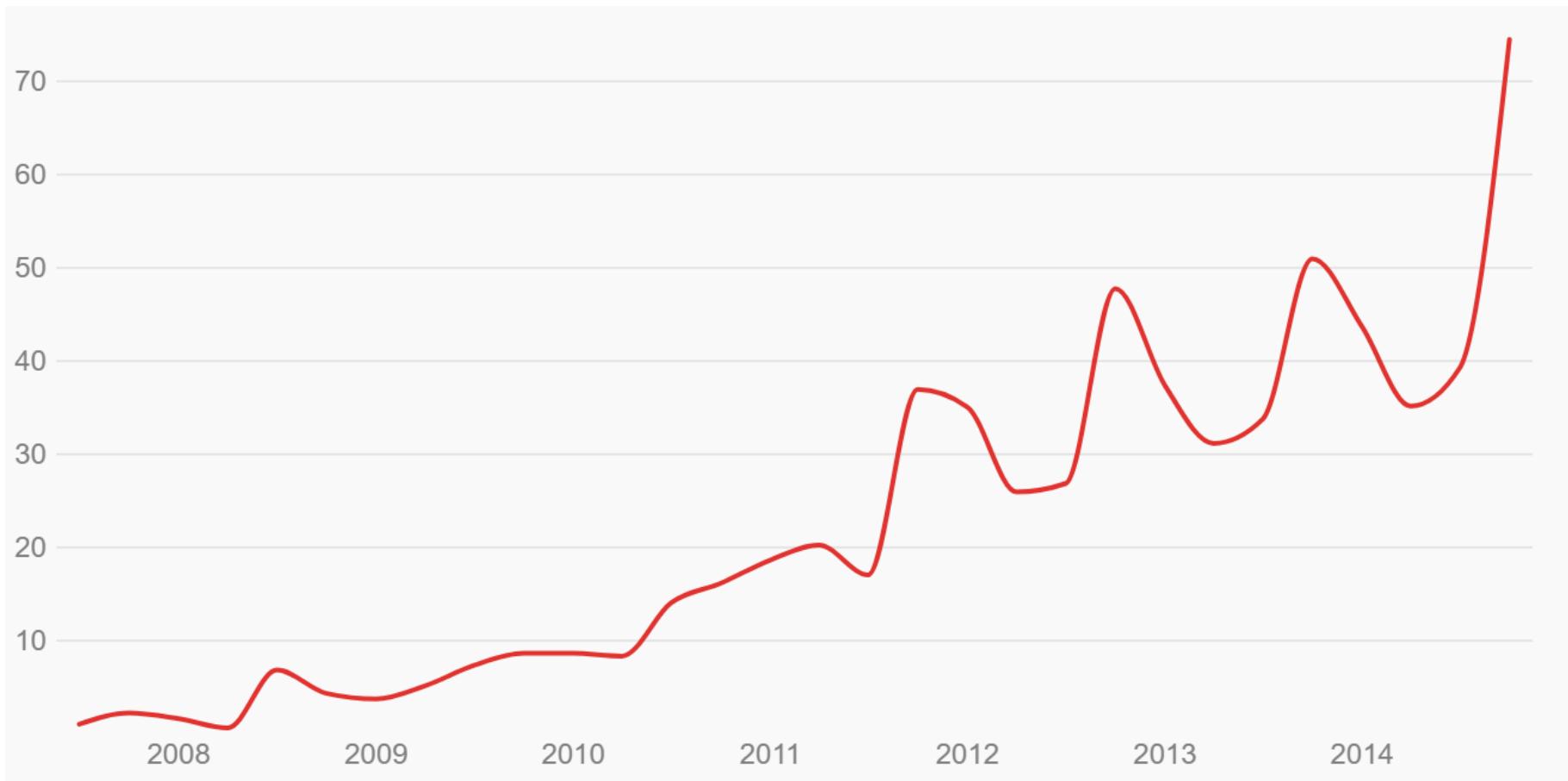


iPhone Sales 2008-2014

1. What's the point to make? What do you want to show?

iPhone Sales

1. What's the point to make? What do you want to show?



2. How can you emphasize your point? iPhone more successful than all other Apple products

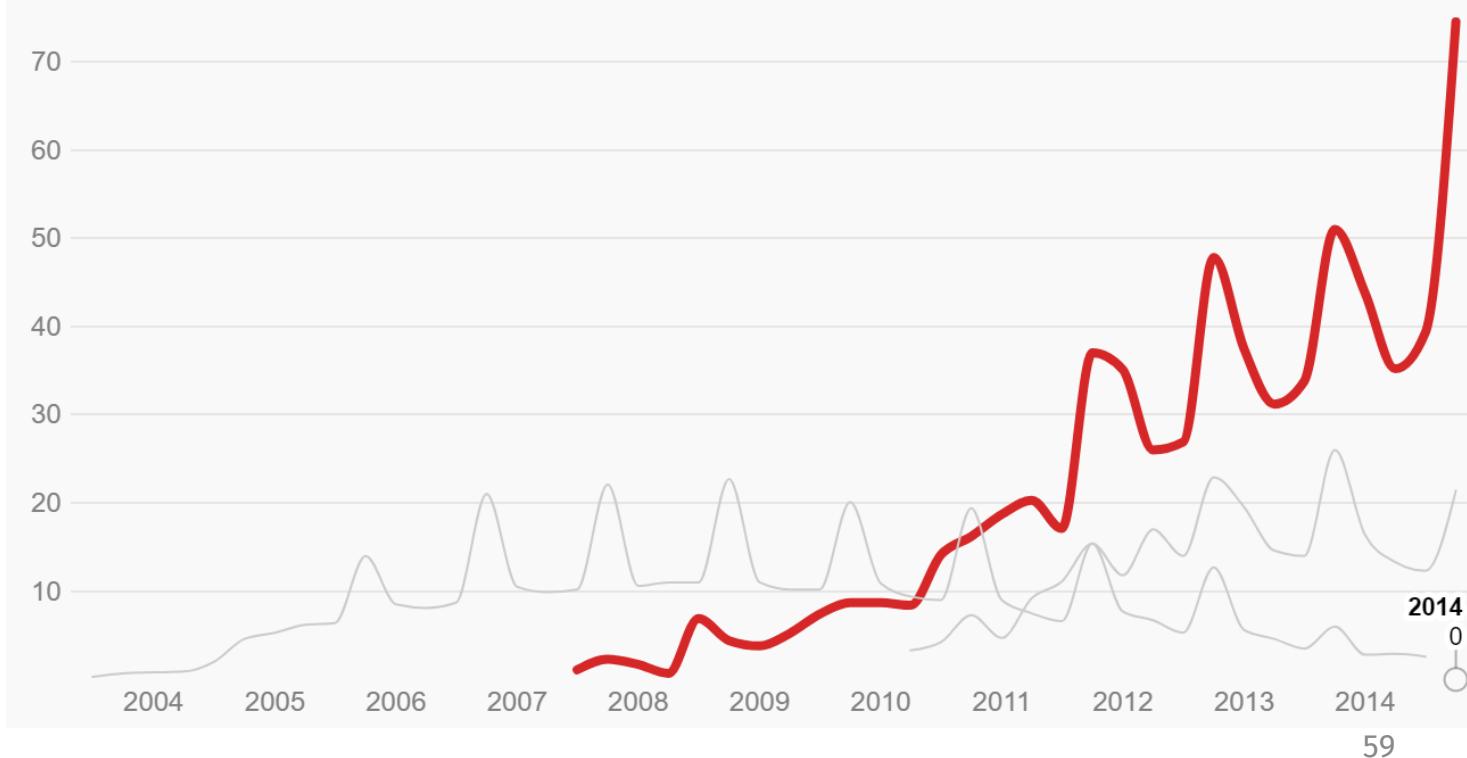
Comparisons set data into perspective. Try to compare data

- With respect to time, other industries
- in your own country, with other countries/ global average

iPhone more successful than all other Apple products

Colour comparison!

Grey is good for
comparison

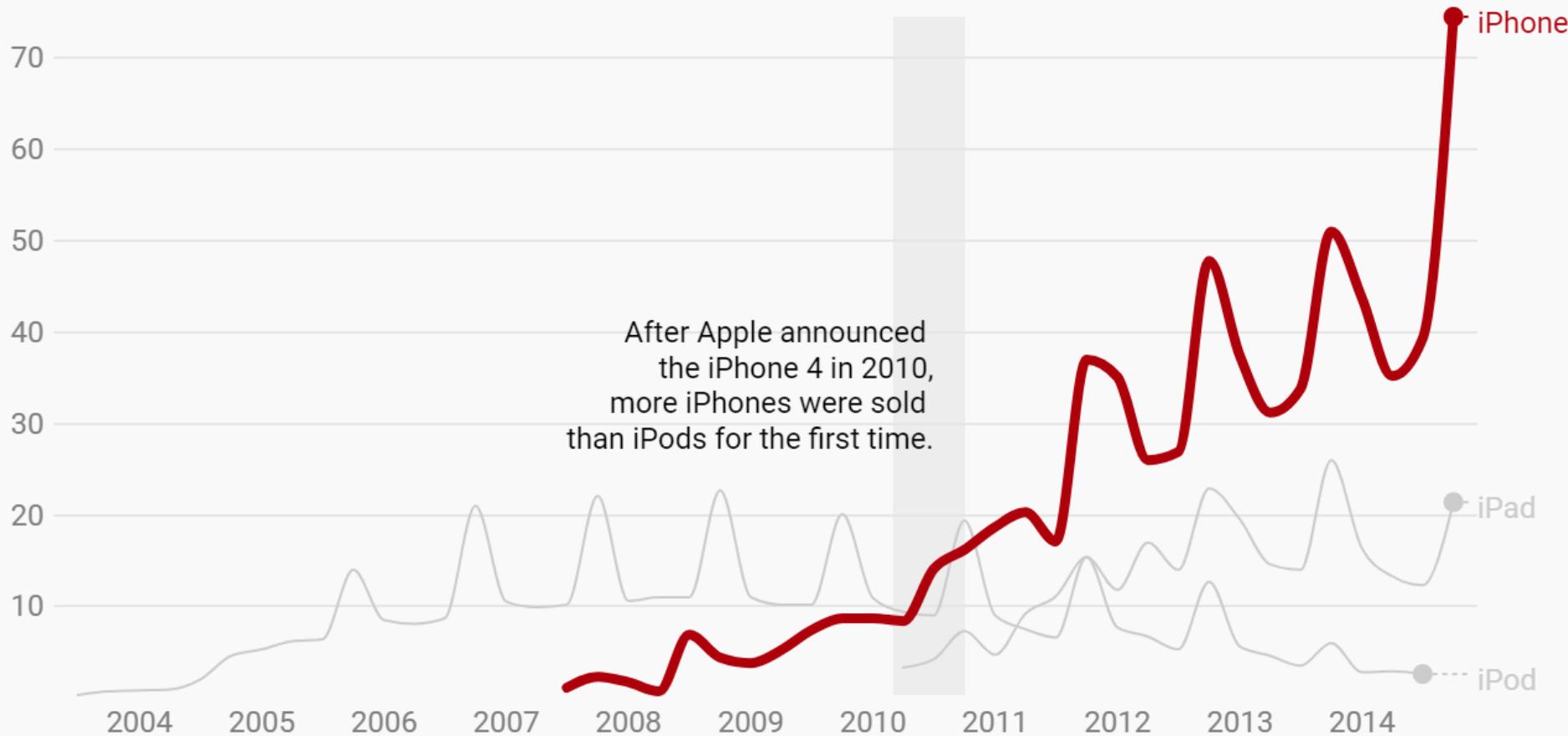


iPhone Success

3. What does the final chart show? *Description, message, legend, source, etc*

iPhone more successful than all other Apple products

Worldwide sales of selected Apple products in million, by fiscal quarter, 2000 to 2014



What is a story

A story is ...

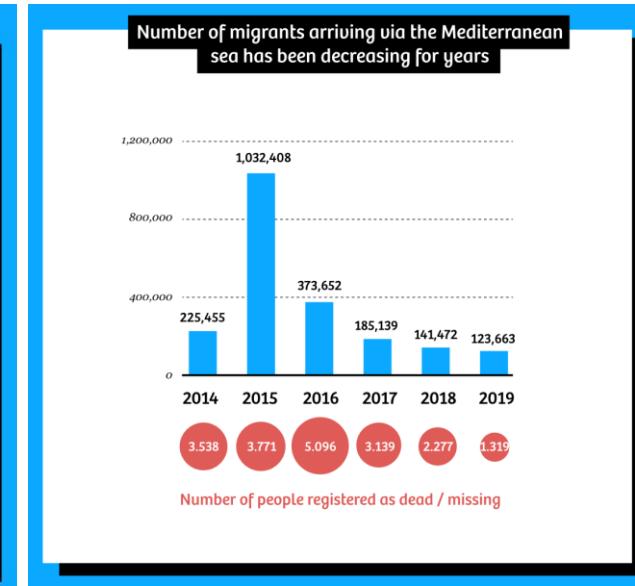
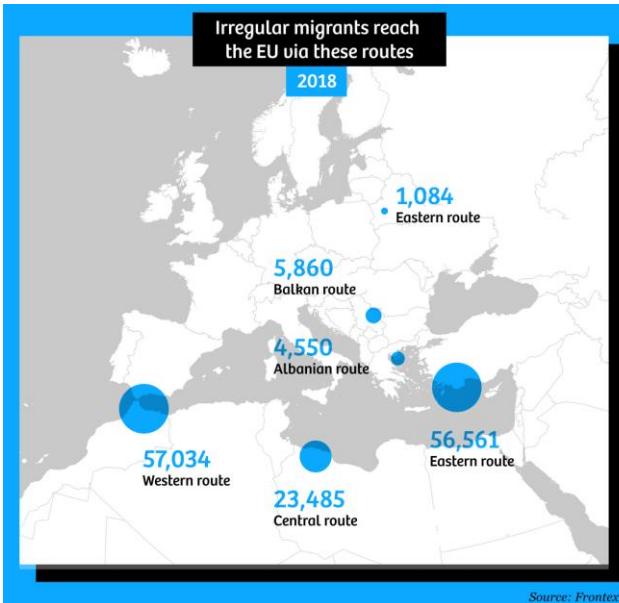
- a collection of facts, observations, or events
- presented in a specific order and fashion

content

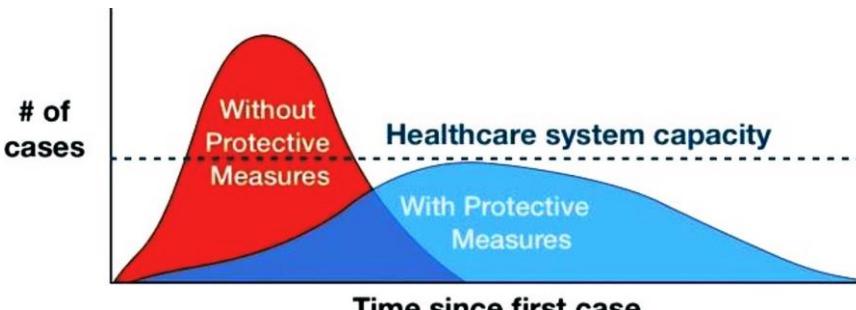
design

So that they can create an **emotional reaction**

How maps in the media make us more negative about migrants



What's right?



Adapted from CDC / The Economist

strip all the complexity.
Straightforward!



Carl T. Bergstrom ✅ @CT_Bergstrom · Mar 6

3. There is a lot of complicated epidemiological modeling behind this idea, but this graphic strips all of that away, and discards irrelevant details to provide a straightforward story that people find easy to grasp at a glance.

It *simplifies* and *highlights* what matters.

6 198 1.8K ↑

Show replies



Carl T. Bergstrom ✅ @CT_Bergstrom · Mar 6

4. I've seldom seen a piece of sci-comm matter so much. We have an opportunity to flatten the #COVID19 #coronavirus epidemic curve by aggressive social distancing and other measures.

But people don't understand what the point is, if the virus is going to circulate broadly.

8 313 2K ↑

Show replies



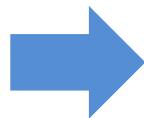
Carl T. Bergstrom ✅ @CT_Bergstrom · Mar 6

5. This graph provides the answer, powerfully and concisely.

And because of that, it has exploded across twitter and other media. I've used it myself a number of times. This graph is changing minds, and by changing minds, it is saving lives.

6 196 1.5K ↑

Contents

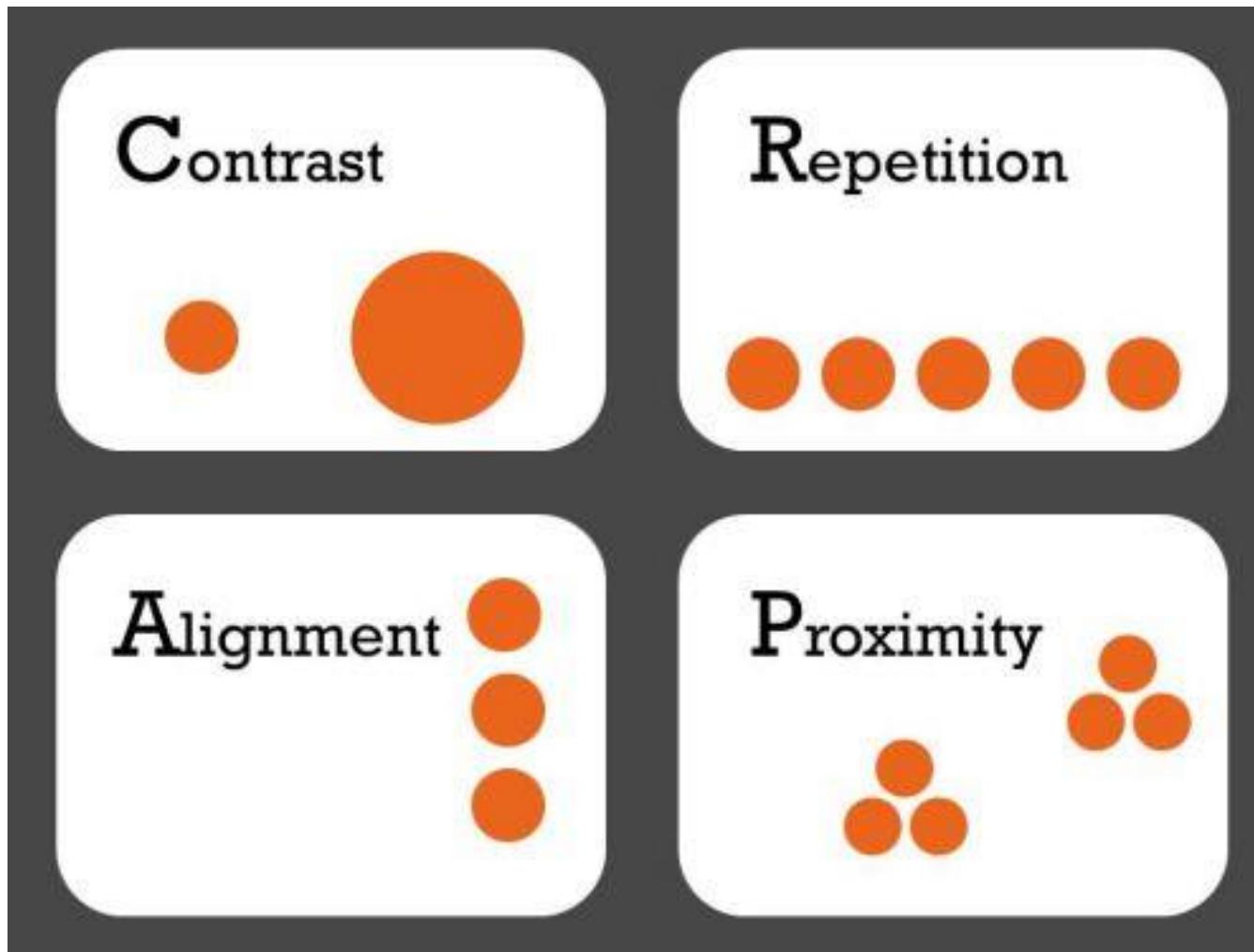


- Introduction to visualisation
- Cairo's Five Principles
- C.R.A.P. Design principles
- Workshop
- Directory of visualisations

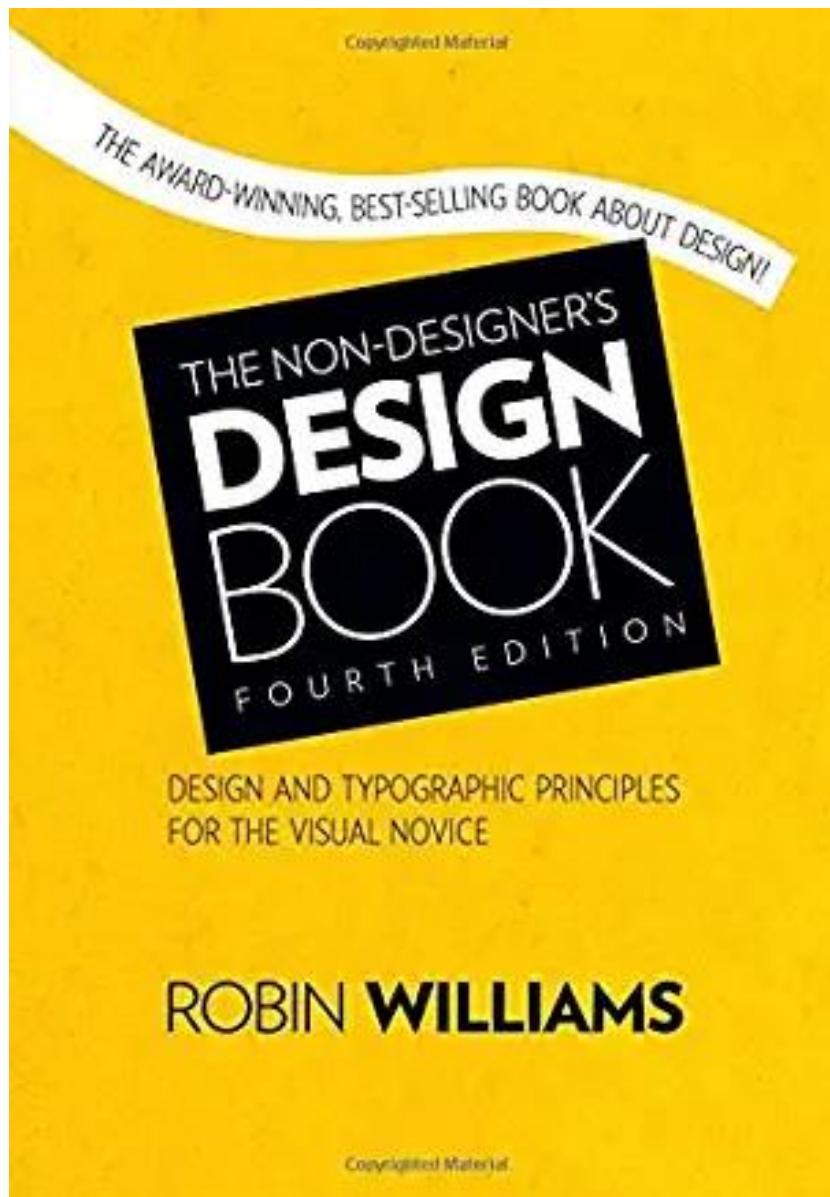
C.R.A.P. principles of design

CRAP: **C**ontrast, **R**epetition, **A**lignment, and **P**roximity.

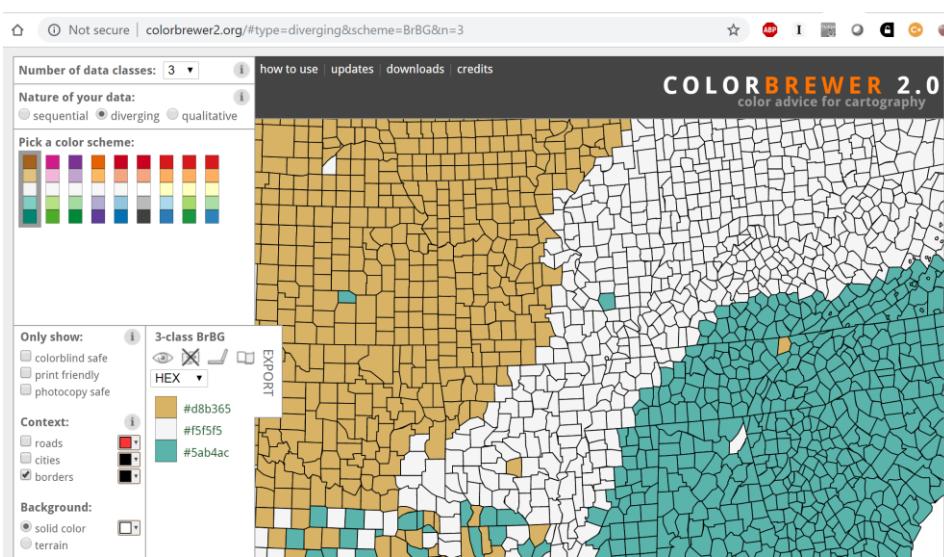
Introduced by Robin Patricia Williams, is a useful design framework



C.R.A.P. principles of design



Colour Contrast

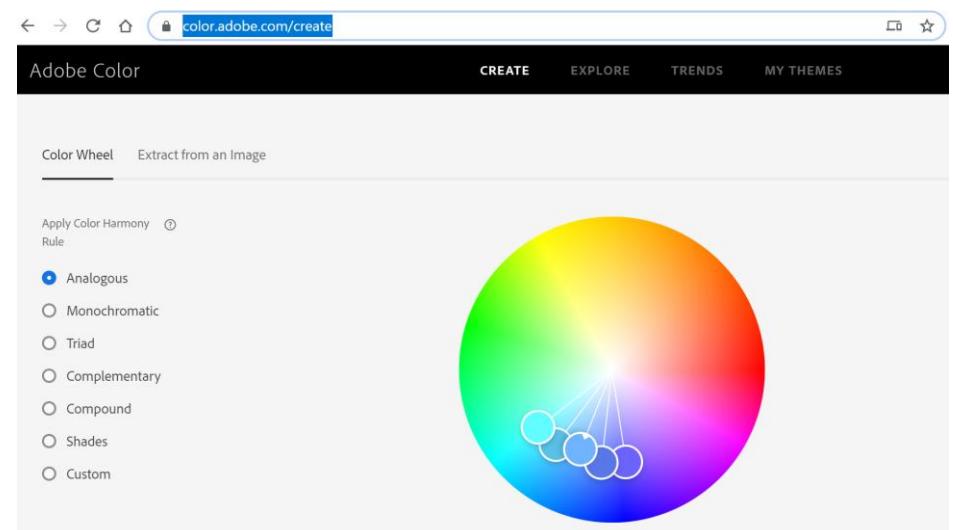


<http://colorbrewer2.org>

Contrast is used to drive a user's attention to specific elements in a design.

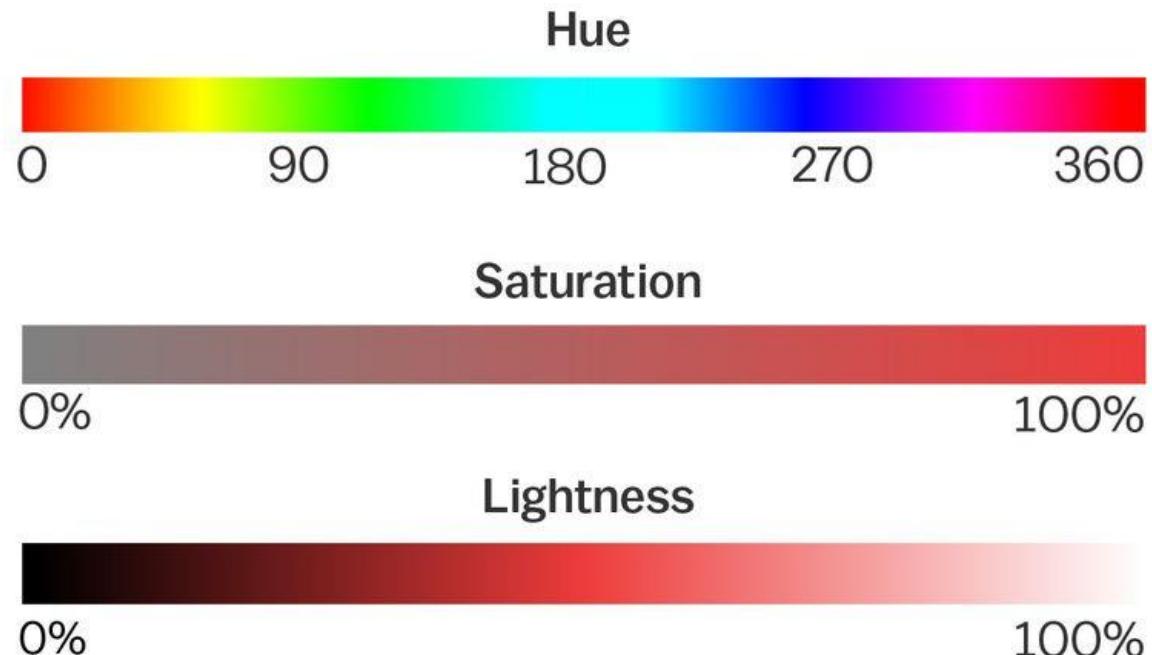
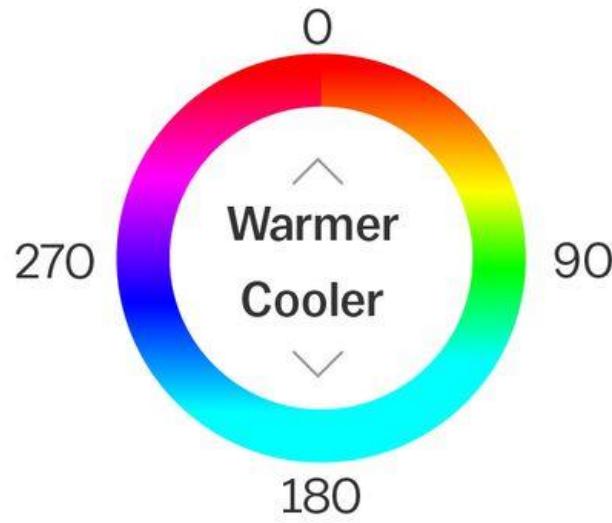
If two items are not the same, make them different. Very different

Use colour as a tool to distinguish, to highlight and contrast.



<https://color.adobe.com/create>

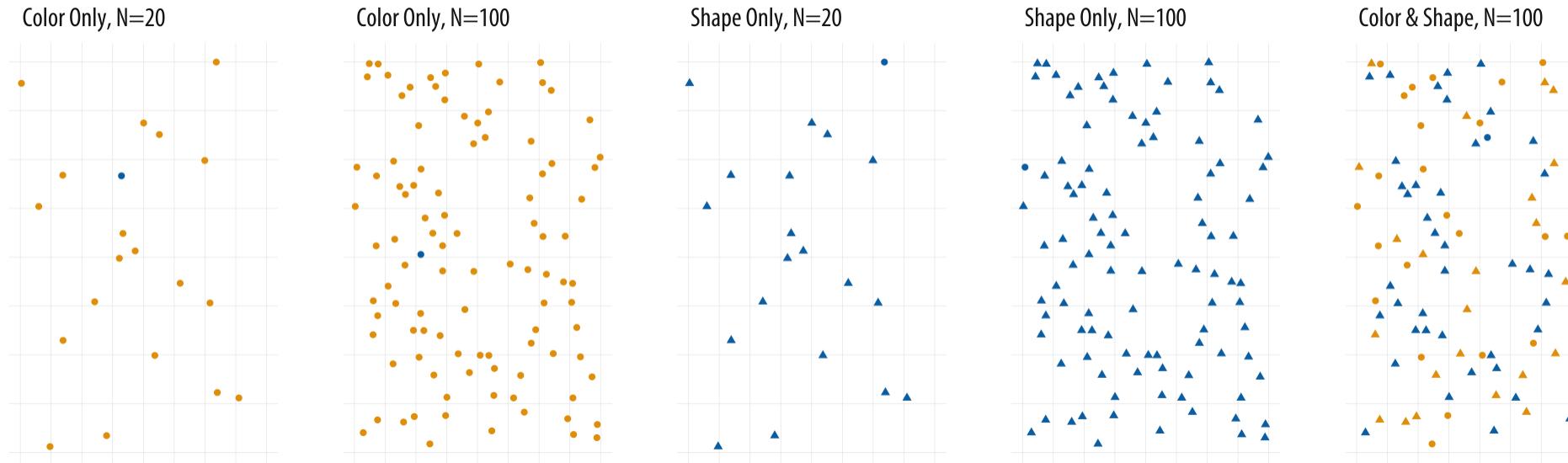
Hue, saturation, and lightness of colors



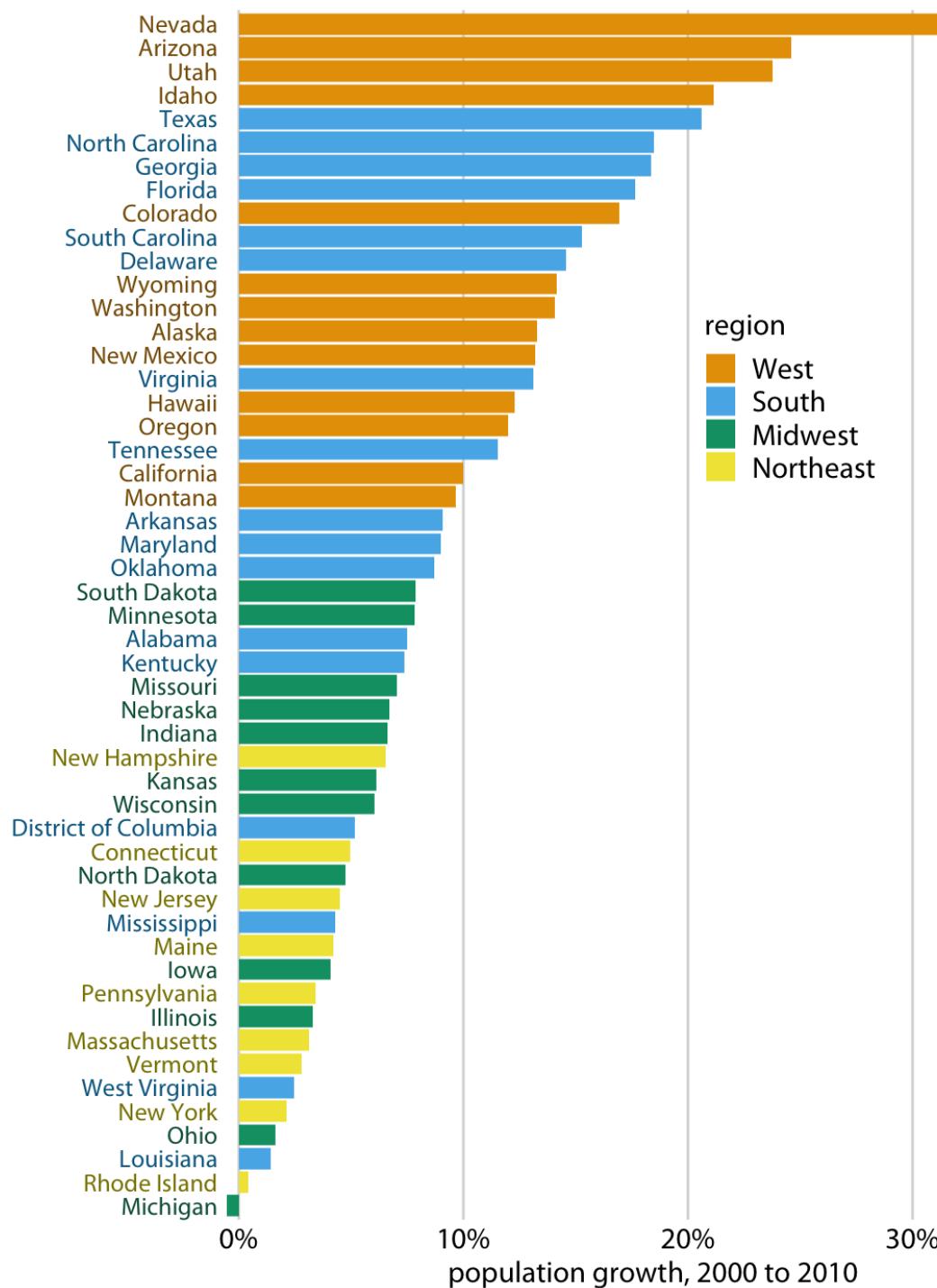
Pre-attentive search and what ‘pops’

Pop-out makes some things easier to see than others.

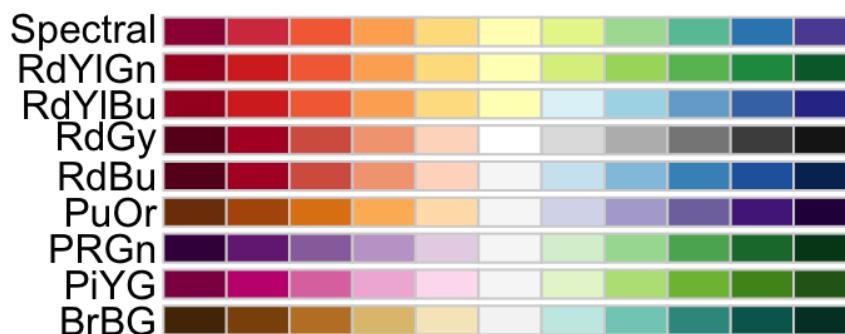
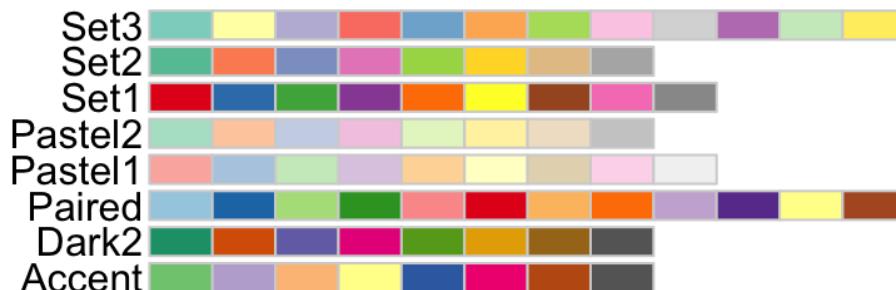
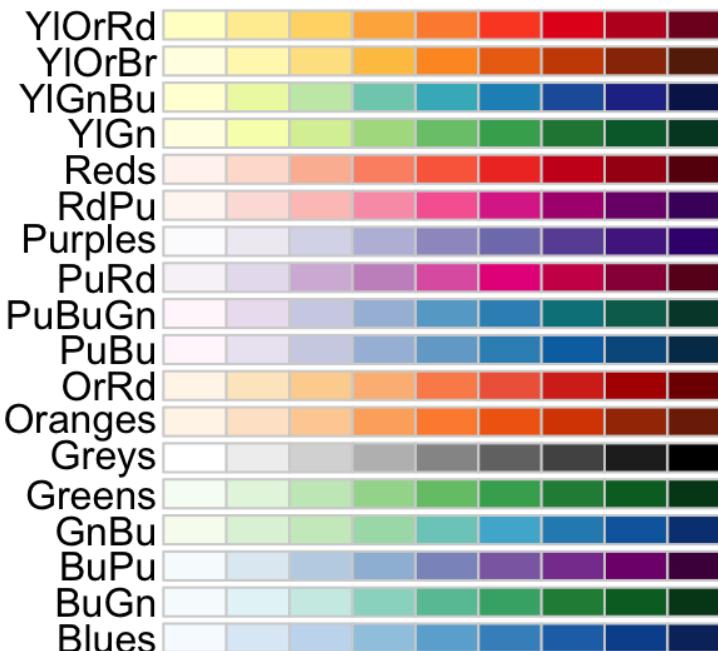
Shape and colour are two distinct channels that encode information visually



Healy(2018) Figure 1.18: Searching for the blue circle becomes progressively harder.



Colour Palettes (1/2)



```
library(RColorBrewer)  
display.brewer.all()
```

Sequential: ordered data
from low to high (gradient)

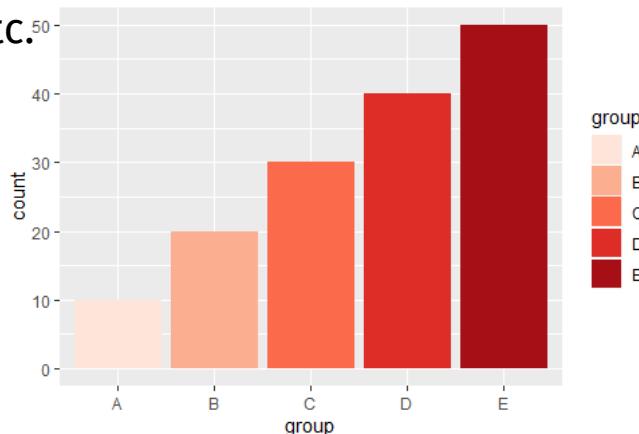
Qualitative: nominal/categorical
data. Does not imply magnitude
differences between groups

Diverging: equal emphasis on
mid-range values and extremes at
both ends of the data

Colour Palettes (2/2)

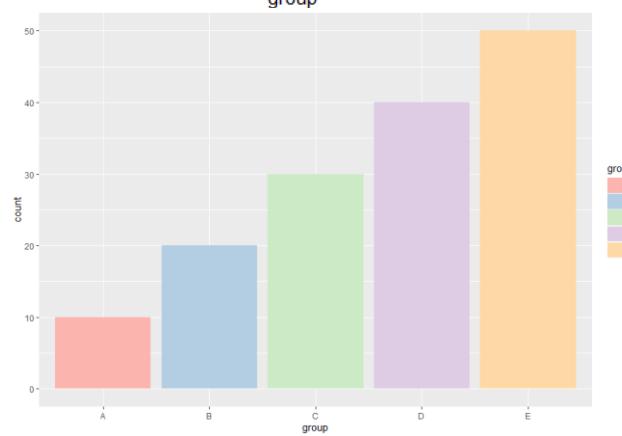
`scale_fill_brewer()` for box-, bar-, violin plot, etc.
`scale_color_brewer()` for lines and points

```
ggplot(toy_data,  
       aes(x = group, y = count, fill = group)) +  
  geom_col() +  
  scale_fill_brewer(palette = "Reds")
```



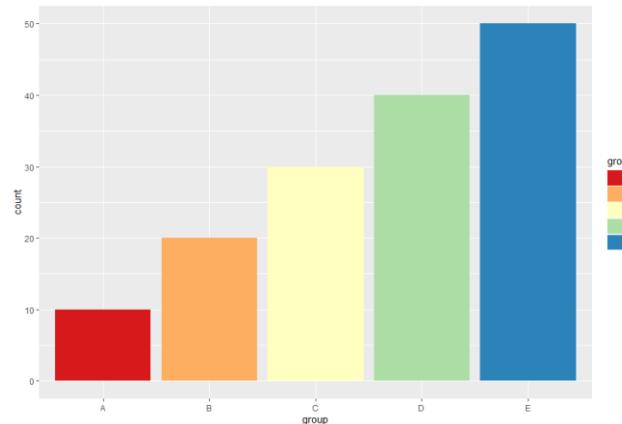
Sequential

```
ggplot(toy_data,  
       aes(x = group, y = count, fill = group)) +  
  geom_col() +  
  scale_fill_brewer(palette = "Pastel1")
```



Qualitative

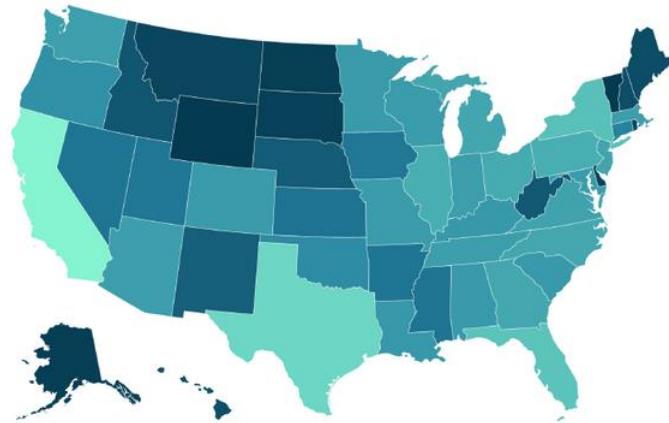
```
ggplot(toy_data,  
       aes(x = group, y = count, fill = group)) +  
  geom_col() +  
  scale_fill_brewer(palette = "Spectral")
```



Diverging

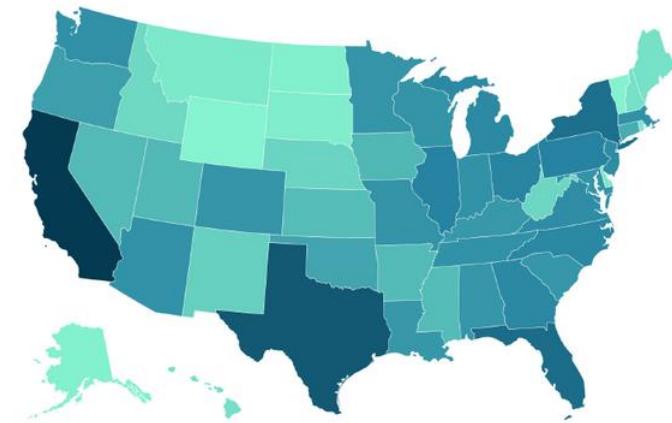
Colour

NOT IDEAL



LOW POPULATION HIGH

BETTER



LOW POPULATION HIGH

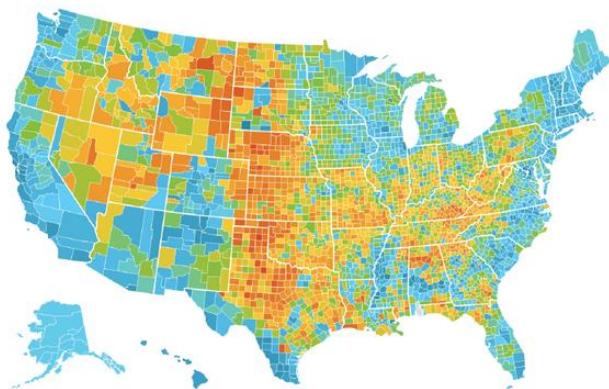
- Use light colours for low values and dark colours for high values.
- When using colour gradients, make sure that the bright colours represent low values, while the dark colours represent high values.
- This will be most intuitive for most readers



- Consider showing most important values with bars, position (like in a dot plot) or even areas. Use colour to show categories to help readers decipher your chart faster
- **Don't use a gradient colour palette for categorical data.** Most readers will associate darker colours with "more/high" and brighter colours with "less/low", such a color palette will imply a ranking of your categories.
- Use different hues (green, yellow, pink, etc.) for categories to avoid that, and to be able to talk about these colours.
- Readers might be quicker at finding specific categories if you make their colours stand out with a different lightness or saturation, but note that your chart should explain why these colors stand out. If you find your chart too colorful, consider another chart type for your data.

Colour

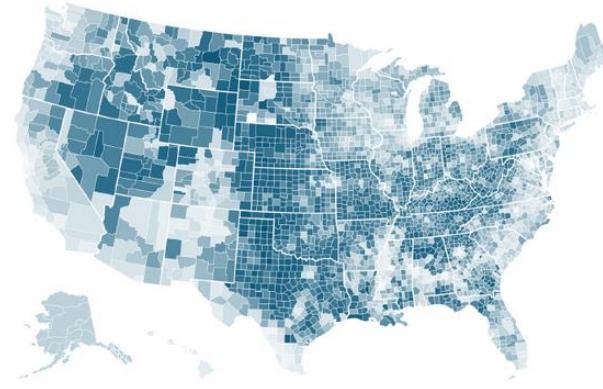
NOT IDEAL



HUE-BASED GRADIENT



BETTER

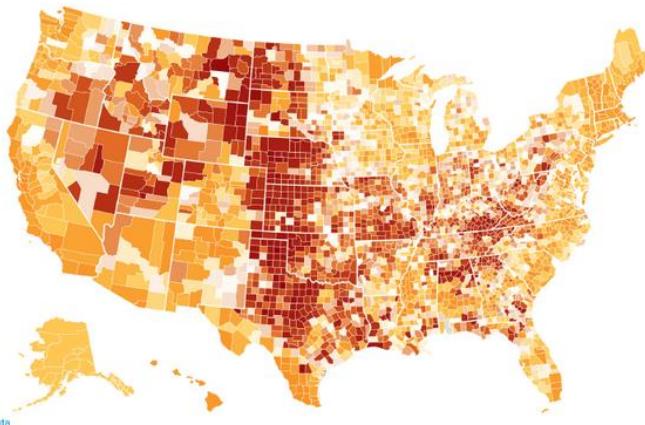


LIGHTNESS-BASED GRADIENT

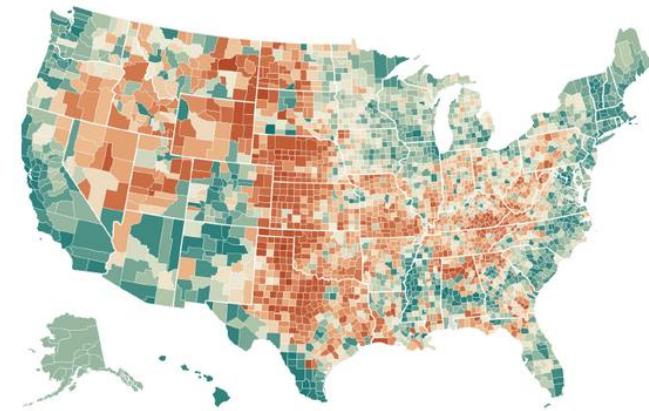


- Build gradients using **lightness**, not just hue.
- Don't place more than two hues with the same lightness in your gradient, but design it from a bright colour (e.g. white) to a dark colour (e.g. dark blue) in a consistent way.
- Your gradient should work in black and white, too.

NOT IDEAL



BETTER



- Consider using **diverging colour gradients**. If you want to emphasize how a variable diverts from a baseline (say the national average), you may want to consider using a diverging palette.
- Important to use clearly distinguishable hues for both sides of the gradient. The center color should ideally be a light grey, not white

Design colour-friendly palette

projects.susielu.com/viz-palette?colors=["#ffd700","#ffb14e","#fa8775","#ea5f94","#cd34b5","#9d02d7","#0000ff"]&backgroundColor="white"&fontColor="black"&mode="deuteranomaly"

Background color: #ffffff
Font color: ● #000000
Charts made with [Semiotic](#)

VIZ PALETTE

By: Elijah Meeks & Susie Lu

PICK

Use Chroma.js

Use Colorgorical

Use ColorBrewer

EDIT

7 Colors

Add

● #hex ○rgb

○hsl

Sample font

Randomize Data

Stroke: Dark None

word mot 詞 salita 워드

String quotes

Object with metadata

["#ffd700", "#ffb14e", "#fa8775", "#ea5f94", "#cd34b5", "#9d02d7", "#0000ff"]

Design your own colour palette

1 What kind of palette do you want to create?

Palette type: sequential diverging

Number of colors: 5

2 Select and arrange input colors

#00429d #96ffea fffffe0

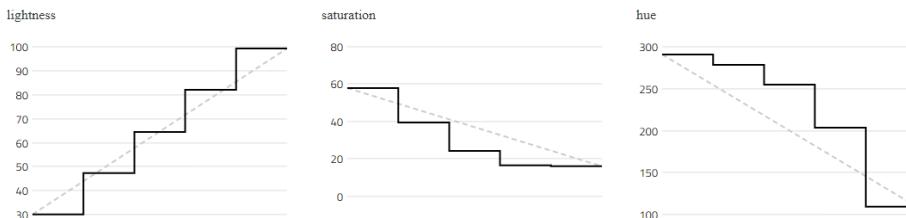
3 Check and configure the resulting palette

correct lightness

bezier interpolation

✓ This palette is colorblind-safe.

simulate: normal deut. prot. trit.



4 Export the color codes in various formats

You can also save your palette for later by bookmarking [this page](#) using **ctrl + d**.

#00429d, #4771b2, #73a2c6, #a5d5d8, #fffffe0

'#00429d', '#4771b2', '#73a2c6', '#a5d5d8', '#fffffe0'

```
ggplot(toy_data, aes(x = group, y = count, fill = group)) +
  geom_col() +
  scale_fill_manual(values=c('#00429d', '#4771b2', '#73a2c6', '#a5d5d8', '#fffffe0'))
```

<https://gka.github.io/palettes/>

1 What kind of palette do you want to create?

Palette type: sequential diverging

Number of colors: 5

2 Select and arrange input colors

#00429d #96ffea fffffe0

fffffe0 #f005e #93003a

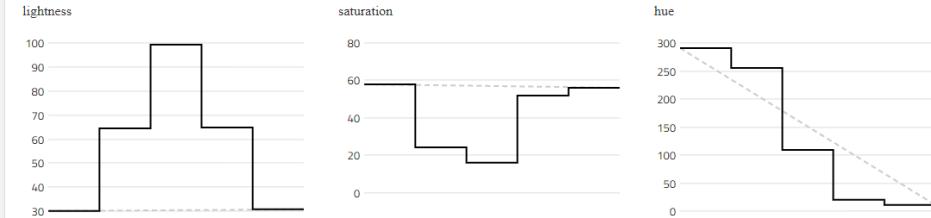
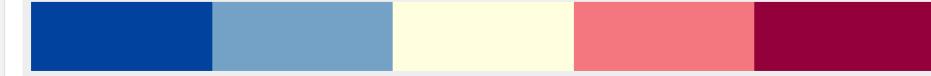
3 Check and configure the resulting palette

correct lightness

bezier interpolation

✓ This palette is colorblind-safe.

simulate: normal deut. prot. trit.



4 Export the color codes in various formats

You can also save your palette for later by bookmarking [this page](#) using **ctrl + d**.

#00429d, #73a2c6, #fffffe0, #f4777f, #93003a

'#00429d', '#73a2c6', '#fffffe0', '#f4777f', '#93003a'

Coblis — Color Blindness Simulator

If you are not suffering from a color vision deficiency it is very hard to imagine how it looks like to be colorblind. The Color BLIndness Simulator can close this gap for you. Just play around with it and get a feeling of how it is to have a color vision handicap.

As all the calculations are made on your local machine, no images are uploaded to the server. Therefore you can use images as big as you like, there are no restrictions. Be aware, there are some issues for the "Lens feature" on Edge and Internet Explorer. All others should support everything just fine.

So go ahead, choose an image through the upload functionality or just drag and drop your image in the center of our Color BLIndness Simulator. It is also possible to zoom and move your images around using your mouse – try it out, I hope you like it.

Drag and drop or paste your file in the area below or: No file chosen

Trichromatic view: Anomalous Trichromacy:

Normal

Red-Weak/Protanomaly

Green-Weak/Deutanomaly

Blue-Weak/Tritanomaly

Dichromatic view:

Red-Blind/Protanopia

Green-Blind/Deutanopia

Blue-Blind/Tritanopia

Monochromatic view:

Monochromacy/Achromatops

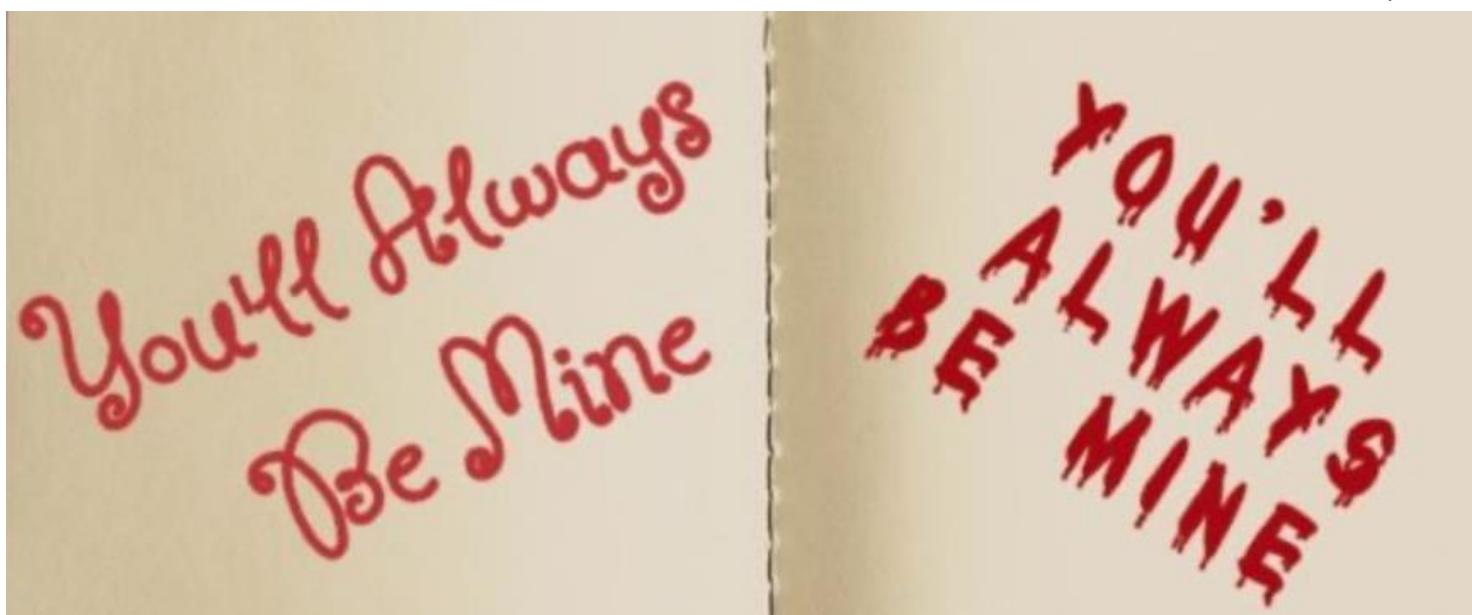
Blue Cone Monochromacy

Use lens to compare with normal view: No Lens Normal Lens Inverse Lens

[Reset View](#)



FONTS MATTER



Typographic Contrast

Serif	Lorem ipsum dolor sit amet
Sans Serif	Lorem ipsum dolor sit amet
Slab Serif	Lorem ipsum dolor sit amet
Script	<i>Lorem ipsum dolor sit amet</i>
Decorative	Lorem ipsum dolor sit amet

The screenshot shows a web page from efficientwp.com. At the top, there's a navigation bar with links for SERVICES, PLUGINS, RESOURCES, ABOUT, and CONTACT. Below the navigation, there's a search bar and some social media icons. The main content area features a heading 'GOOGLE WEB FONTS (SANS-SERIF)' and a sub-section titled 'ARIMO'. To the left of the main content, there's a sidebar with a list of sans-serif fonts: Arimo, Cabin, Didact Gothic, Dosis, Fjalla One, Lato, and Montserrat. The 'ARIMO' section contains sample text in different font weights and styles.

These are some of my preferred sans-serif Google Web Fonts. You can also see my preferred [serif](#) and [display and handwriting](#) fonts.

This is a detailed view of the ARIMO font preview section. On the left, a sidebar lists several sans-serif fonts: Arimo, Cabin, Didact Gothic, Dosis, Fjalla One, Lato, and Montserrat. The main content area is titled 'ARIMO' and shows a sample of the font in 'FONT WEIGHT 400 (NORMAL)'. The sample text includes a small sentence, a large title, and a paragraph of lorem ipsum text.

Arimo	ARIMO FONT WEIGHT 400 (NORMAL)
Cabin	The quick brown fox jumps over the lazy dog. THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.
Didact Gothic	Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec a diam lectus. Sed sit amet ipsum mauris. Maecenas congue ligula ac quam viverra nec consectetur ante hendrerit.
Dosis	Donec et mollis dolor. Praesent et diam eget libero egestas mattis sit amet vitae augue. Nam tincidunt congue enim, ut porta lorem lacinia consectetur. Donec ut libero sed arcu vehicula ultricies a non tortor.
Fjalla One	
Lato	
Montserrat	

Typographic Contrast



1

**ROBOTO
CONDENSED
BOLD** + **CABIN
REGULAR**

The life of a designer is a life of fight.

Fight against the ugliness. Just like a doctor fights against disease. For us, the visual disease is what we have around, and what we try to do is cure it somehow with design.

Massimo Vignelli

[Download Font Pair](#) [View Roboto + Cabin on Google Fonts](#)
[Copy the Code](#) [Click to open HTML + CSS](#)



2

**RALEWAY
LIGHT** + **OPEN
SANS
REGULAR**

If I was influenced by anything, it was architecture.

Structure having to do with logic. If you don't do it right, the whole thing is going to cave in. In a certain sense, you can carry that to graphic design.

Paul Rand

[Download Font Pair](#) [View Raleway + Open Sans on Google Fonts](#)
[Copy the Code](#) [Click to open HTML + CSS](#)



3

**PLAYFAIR
DISPLAY
BOLD ITALIC** + **SOURCE
SANS
REGULAR**

Design is the method of putting form and content together.

Design, just as art, has multiple definitions; there is no single definition. Design can be art. Design can be aesthetics. Design is so simple, that's why it is so complicated.

Paul Rand

[Download Font Pair](#) [View Playfair + Source Sans on Google Fonts](#)
[Copy the Code](#) [Click to open HTML + CSS](#)



4

**ROBOTO
BOLD** + **LORA
REGULAR**

So what is design all about?

So what is design all about? It [exists to] decrease the amount of vulgarity in the world ... to make the world a better place to be. It doesn't have to be one style. We're not talking about style, we're talking about quality. Style is tangible, quality is intangible. I am talking about giving to everything that surrounds us a level of quality.

Massimo Vignelli

[Download Font Pair](#) [View Roboto + Lora on Google Fonts](#)
[Copy the Code](#) [Click to open HTML + CSS](#)

Google fonts

The screenshot shows the Google Fonts homepage. At the top, there's a navigation bar with links for CATALOG, FEATURED, ARTICLES, and ABOUT. A search bar at the top left contains the placeholder "Search fonts". Below it, there are dropdown menus for "Sentence" and "Font Properties", and a text input field with the placeholder "Type something". To the right of these are font size controls (set to 40px) and a color palette. Below the search bar, there are three filter buttons: "Categories", "Language", and "Font Properties". A modal window titled "Categories" is open on the left, listing font categories with checkboxes: Serif, Sans Serif, Display, Handwriting, and Monospace, all of which are checked. On the right, the main content area displays a font named "Jomolhari" by Christopher J. Fynn, with one style listed. The text "A red flare silhouetted the jagged edge of a wing." is shown in the font. A user study message at the top encourages sign-up.

Fontbase to get all google fonts

The screenshot shows the Fontbase website interface. At the top, there's a navigation bar with links for Awesome, Learn, Blog, Updates, Contact, Download, and Account. Below the navigation is a large heading "Font management. Perfected." followed by a subtext: "All platforms. Professional features. Beautiful UI. Totally free. FontBase is the font manager of the new generation, built by designers, for designers." A prominent green button says "Download FontBase". Below this, it says "We support Mac, Windows and Linux".

The main area features a sidebar on the left with categories like All, Recent, Favorites (Active), Inactive, Collections (Serif, Sans-serif, Posters, New Project), Providers (Google), and Folders (Foundries, Pangram, Grilli Type, Colophon, TypeType, Free, Paid, Google, Text, Display). The main content area displays several font preview cards. Each card includes a font name, style, and a preview of the text "The quick fox jumps over the lazy dog". A context menu is open over the third preview card, showing options: "Activate 5 fonts" (highlighted), "Deactivate 3 fonts", "View family", "Manage tags", and "Delete from disk".

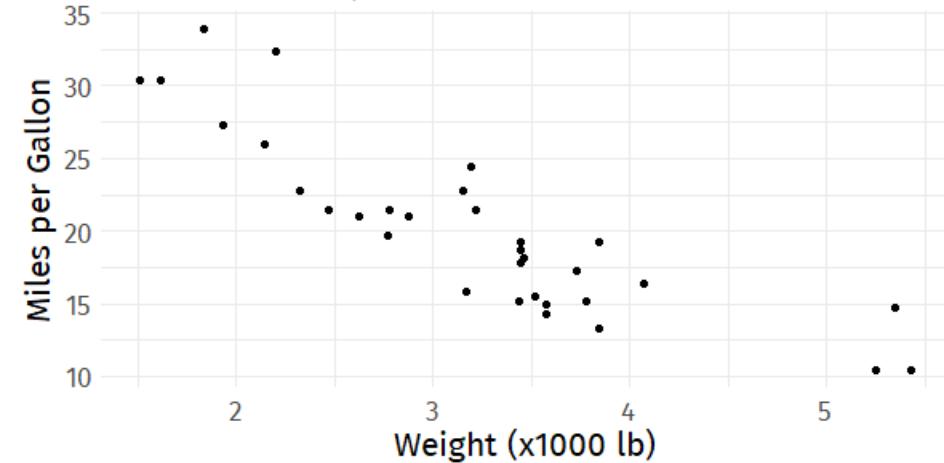
Adding fonts to R

fonts.R

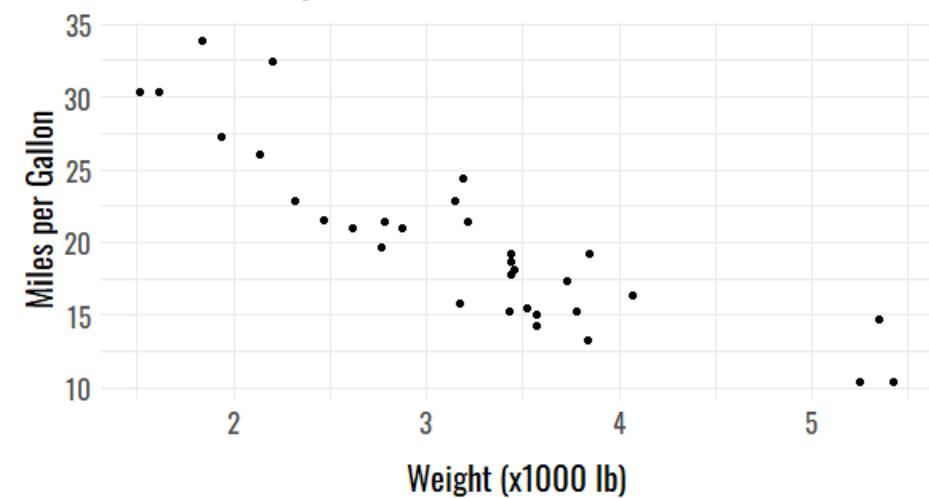
```
1 # https://github.com/wch/extrafont
2 # install.packages("extrafont")
3 library(extrafont)
4 # for windows, you may have to remotes::install_version("Rttf2pt1", version = "1.3.8")
5
6 library(tidyverse)
7 library(patchwork)
8
9 # import fonts - only once-- will look at system font folder
10 font_import()
11
12
13 # to get fonts from google, go to https://fonts.google.com/
14 # Windows: use skyfonts to download google fonts on your computer https://www.fonts.com/web-fonts/google
15 # if using skyfonts, point to the directory where google fonts live
16 # on my windows machine this is @
17 font_import( paths="C:/Users/kchristodoulou/AppData/Roaming/Monotype/skyfonts-google")
18
19 #if you use FontBase to get all Google fonts
20 font_import( paths="C:/Users/kchristodoulou/FontBase")
21
22 # Alternatively, you can download all of google fonts, by downloading and extracting
23 # https://github.com/google/fonts/archive/master.zip (over 300 Mb) from https://github.com/google/fonts
24
25 extrafont::loadfonts(device="win")
26
27
28 # Vector of font family names
29 fonts()
```

Adding fonts to R fonts.R

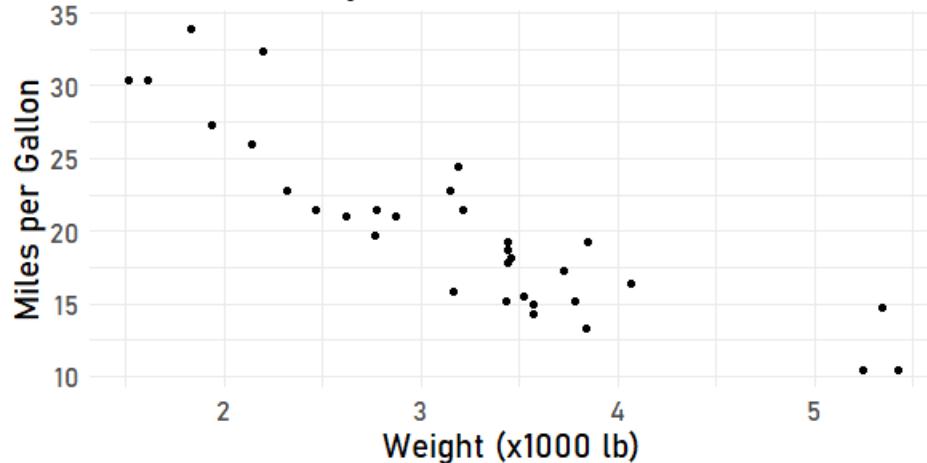
Fuel Efficiency of 32 Cars



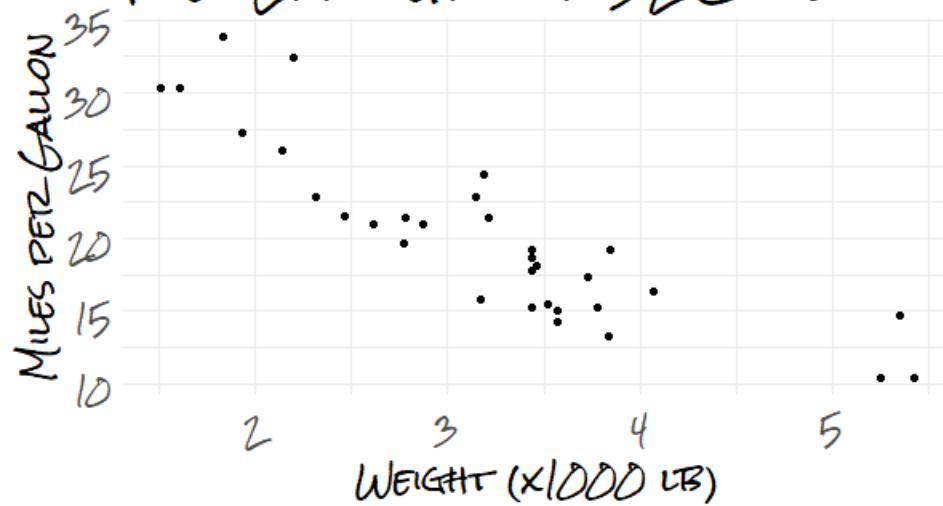
Fuel Efficiency of 32 Cars



Fuel Efficiency of 32 Cars



FUEL EFFICIENCY OF 32 CARS



1II Arial

1II Bahnschrift

1II Bitter

1II Calibri

1II Fira Sans

1II Lato

1II Montserrat

1II Noto Sans

1II Open Sans

1II Oswald

1II Times New Roman

1II Roboto

Numbers: Use tabular fonts

Proportional numbers work best when used in the text.

Tabular numbers take up the same amount of width per character and are easier to read in visualisations. If you line the figures up, you will see that they fall into equally spaced columns.



32154

Montserrat
proportional numbers



32154

Open Sans
tabular numbers



32154

Lato
tabular numbers

32154

32154

32154

C.R.A.P. - Repetition

Repeat some aspect of the design throughout the entire piece. The repetitive element may be a bold font, a thick rule (line), a certain bullet, design element, color, format, spatial relationships, etc. It can be anything that a reader will visually recognize.

Sock and Buskin

Ambrosia Sidney

109 Friday Street
Penshurst, NM
505.555.1212



Sock and Buskin

Ambrosia Sidney

109 Friday Street
Penshurst, NM
505.555.1212

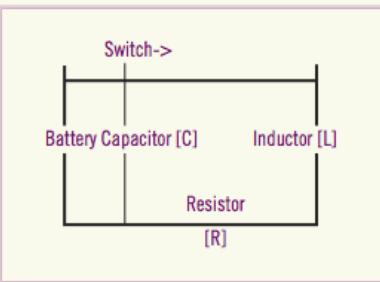


C.R.A.P. - Alignment

Every item on a page should have a visual connection with one or more items in the design, even if the connection is through an invisible line

Example 6: Value of a resistor in an electrical circuit.

Find the value of a resistor in an electrical circuit which will dissipate the charge to 1 percent of its original value within one twentieth of a second after the switch is closed.

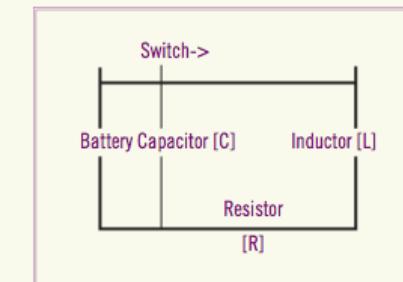


q0=	9	volts
q(t)=	0.09	volts
t=	0.05	seconds
L=	8	henrys
C=	0.0001	farads
R=	300	ohms
q(t)=	0.253889	

1/[L*C_]	1250
[R_/(2*L)]^2	351.5625
SQRT(B15-B16)	29.973947
COS(T*B17)	0.07203653
-R_*T/(2*L)	-0.9375
Q0+EXP (B 19)	3.52445064

Example 6: Value of a resistor in an electrical circuit.

Find the value of a resistor in an electrical circuit which will dissipate the charge to 1 percent of its original value within one twentieth of a second after the switch is closed.



q0=	9	volts
q(t)=	0.09	volts
t=	0.05	seconds
L=	8	henrys
C=	0.0001	farads
R=	300	ohms
q(t)=	0.253889	

1/[L*C_]	1250
[R_/(2*L)]^2	351.5625
SQRT(B15-B16)	29.973947
COS(T*B17)	0.07203653
-R_*T/(2*L)	-0.9375
Q0+EXP (B 19)	3.52445064

ggplot Resources

← → ⌂ ⌂ github.com/erikgahner/awesome-ggplot2

 Search or jump to... Pull requests Issues Marketplace Explore

erikgahner / awesome-ggplot2

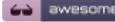
Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 0 tags Go to file Add file Code

 erikgahner add 'ggfocus' and 'ggpval' 56a55dd 7 days ago 70 commits

 README.md add 'ggfocus' and 'ggpval' 7 days ago

README.md

Awesome ggplot2 

General

- [Official website](#)
 - [Reference](#)
- [Cheat Sheet: Data Visualization with ggplot2](#)



Persons (+ twitter)

- [Hadley Wickham \(@hadleywickham\)](#)
- [Kieran Healy \(@kjhealy\)](#)
- [Claus Wilke \(@ClausWilke\)](#)
- [Thomas Lin Pedersen \(@thomasp85\)](#)
- [Winston Chang \(@winston_chang\)](#)
- [Lionel Henry \(@_lionelhenry\)](#)
- [Kara Woo \(@kara_woo\)](#)
- [Hiroaki Yutani \(@yutannihilat_en\)](#)
- [Dewey Dunnington \(@paleolimbot\)](#)

R packages

<https://github.com/erikgahner/awesome-ggplot2>

Session Overview – Next Steps

1. Review of Visualisation Guidelines
2. Individual assignment for next week
 - Thought piece (reflection) + *grading visualisations*
 - Challenge: Create the **worst** visualization you can
 - Problem set in R: Stop and Search in London, part 1
 - You will make a series of 3-4 static visualization plots about Stop and Search in London in Sep 2020
3. **Group project proposal**

- Next Session:
 - Visualising geospatial data
 - Introduction to interactive graphs: Plotly and Shiny

- Submit proposal for group project: Due before session 2

- Reflections: Due before session 2

Please write a short essay (max. 400 words) addressing the following:

Why do we visualize data? What makes a great visualization?

How do you choose which kind of visualization to use?

What does it mean to map data to graph aesthetics? What data was mapped to which aesthetics in Rosling's video "200 Countries, 200 Years, 4 Minutes"

- Grading visualisations: Due before session 2

A rubric to 'grade' visualizations is [The Data Visualization Checklist](#)  It is a helpful set of criteria for grading the effectiveness of a graphic.

Using the Data Visualisation checklist, please 'grade' Figure 10 of the recent JPMorgan Chase Institute on [Year-over-year percent change in credit card spending by industry of employment](#)  . If you wanted the actual paper on which this publication was based, you can find it [here](#). 

- Prepare the *worst* possible visualisation: Due before session 2

- Inspired by Allison Horst's [There's value in trying your \[dataviz\] worst](#)  Inspired and The Economist's [Mistakes, we've drawn a few](#)  , take any of your past visualisations, or find a small dataset and please come up with the **worst** possible visualisation you can
 -

- Problem Set 1: Due before session 2

[Stop and Search in London, part 1](#) 

<https://www.met.police.uk/stopandsearch/>

<https://www.theguardian.com/law/2019/jan/26/met-police-disproportionately-use-stop-and-search-powers-on-black-people>

<https://www.met.police.uk/sd/stats-and-data/met/stop-and-search-dashboard/> (in Tableau!)

Raw data can be downloaded from <https://data.police.uk/>  Select:

Date Range: Sep 2020

Metropolitan Police Service

Include stop and search data

Data Visualization Checklist

by Stephanie Evergreen & Ann K. Emery
February 2018

This checklist is meant to be used as a guide for the development of high impact data visualizations. Rate each aspect of the data visualization by circling the most appropriate number, where 2 points means the guideline was fully met, 1 means it was partially met, and 0 means it was not met at all. n/a should not be used frequently, but reserved for when the guideline truly does not apply. For example, a pie chart has no axes lines or tick marks to rate. If the guidelines has been broken intentionally to make a point, rate it n/a and deduct those points from the total possible. Refer to the Data Visualization Anatomy Chart on the last page for guidance on vocabulary and the Resources at the end for more details.

Text	Guideline	Rating
Graphs don't contain much text, so existing text must encapsulate your message and pack a punch.	6-12 word descriptive title is left-justified in upper left corner Short titles enable readers to comprehend takeaway messages even while quickly skimming the graph. Rather than a generic phrase, use a descriptive sentence that encapsulates the graph's finding or "so what?" Western cultures start reading in the upper left, so locate the title there.	2 1 0 n/a
	Subtitle and/or annotations provide additional information Subtitles and annotations (call-out text within the graph) can add explanatory and interpretive power to a graph. Use them to answer questions a viewer might have or to highlight specific data points.	2 1 0 n/a
	Text size is hierarchical and readable Titles are in a larger size than subtitles or annotations, which are larger than labels, which are larger than axis labels, which are larger than source information. The smallest text - axis labels - are at least 9 point font size on paper, at least 20 on screen.	2 1 0 n/a
	Text is horizontal Titles, subtitles, annotations, and data labels are horizontal (not vertical or diagonal). Line labels and axis labels can deviate from this rule and still receive full points. Consider switching graph orientation (e.g., from column to bar chart) to make text horizontal.	2 1 0 n/a
	Data are labeled directly Position data labels near the data rather than in a separate legend (e.g., on top of or next to bars and next to lines). Eliminate/embed legends when possible because eye movement back and forth between the legend and the data can interrupt the brain's attempts to interpret the graph.	2 1 0 n/a
	Labels are used sparingly Focus attention by removing the redundancy. For example, in line charts, label every other year on an axis. Do not add numeric labels *and* use a y-axis scale, since this is redundant.	2 1 0 n/a

Arrangement Improper arrangement of graph elements can confuse readers at best and mislead viewer at worst. Thoughtful arrangement makes a data visualization easier for a viewer to interpret.	Proportions are accurate A viewer should be able measure the length or area of the graph with a ruler and find that it matches the relationship in the underlying data. Y-axis scales should be appropriate. Bar charts start axes at 0. Other graphs can have a minimum and maximum scale that reflects what should be an accurate interpretation of the data (e.g., the stock market ticker should not start at 0 or we won't see a meaningful pattern).	2	1	0	n/a
	Data are intentionally ordered Data should be displayed in an order that makes logical sense to the viewer. Data may be ordered by frequency counts (e.g., from greatest to least for nominal categories), by groupings or bins (e.g., histograms), by time period (e.g., line charts), alphabetically, etc. Use an order that supports interpretation of the data.	2	1	0	n/a
	Axis intervals are equidistant The spaces between axis intervals should be the same unit, even if every axis interval isn't labeled. Irregular data collection periods can be noted with markers on a line graph, for example.	2	1	0	n/a
	Graph is two-dimensional Avoid three-dimensional displays, bevels, and other distortions.	2	1	0	n/a
	Display is free from decoration Graph is free from clipart or other illustrations used solely for decoration. Some graphics, like icons, can support interpretation.	2	1	0	n/a
	Color scheme is intentional Colors should represent brand or other intentional choice, not default color schemes. Use your organization's colors or your client's colors. Work with online tools to identify brand colors and others that are compatible.	2	1	0	n/a
Color Keep culture-laden color connotations in mind. For example, pink is highly associated with feminine qualities in the USA. Use sites like Color Brewer to find color schemes suitable for reprinting in black-and-white and for colorblindness.	Color is used to highlight key patterns Action colors should guide the viewer to key parts of the display. Less important, supporting, or comparison data should be a muted color, like gray.	2	1	0	n/a
	Color is legible when printed in black and white When printed or photocopied in black and white, the viewer should still be able to see patterns in the data.	2	1	0	n/a
	Color is legible for people with colorblindness Avoid red-green and yellow-blue combinations when those colors touch one another. Avoid using red to mean bad and green to mean good in the same chart.	2	1	0	n/a
	Text sufficiently contrasts background Black/very dark text against a white/transparent background is easiest to read.	2	1	0	n/a

Lines

Excessive lines—gridlines, borders, tick marks, and axes—can add clutter or noise to a graph, so eliminate them whenever they aren't useful for interpreting the data.

Gridlines, if present, are muted
Color should be faint gray, not black. Full points if no gridlines are used. Gridlines, even muted, should not be used when the graph includes numeric labels on each data point.

2 1 0 n/a

Graph does not have border line
Graph should bleed into the surrounding page or slide rather than being contained by a border.

2 1 0 n/a

Axes do not have unnecessary tick marks or axis lines
Tick marks can be useful in line graphs (to demarcate each point in time along the y-axis) but are unnecessary in most other graph types. Remove axes lines whenever possible.

2 1 0 n/a

Graph has one horizontal and one vertical axis
Viewers can best interpret one x- and one y-axis. Don't add a second y-axis. Try a connected scatter plot or two graphs, side by side, instead. (A secondary axis used to hack new graph types is ok, so long as viewers aren't being asked to interpret a second y-axis.)

2 1 0 n/a

Overall

Graphs will catch a viewer's attention so only visualize the data that needs attention. Too many graphics of unimportant information dilute the power of visualization.

Graph highlights significant finding or conclusion
Graphs should have a "so what?" – either a practical or statistical significance (or both) to warrant their presence. For example, contextualized or comparison data help the viewer understand the significance of the data and give the graph more interpretive power.

2 1 0 n/a

The type of graph is appropriate for data
Data are displayed using a graph type appropriate for the relationship within the data. For example, change over time is displayed as a line graph, area chart, slope graph, or dot plot.

2 1 0 n/a

Graph has appropriate level of precision
Use a level of precision that meets your audiences' needs. Few numeric labels need decimal places, unless you are speaking with academic peers. Charts intended for public consumption rarely need *p* values listed.

2 1 0 n/a

Individual chart elements work together to reinforce the overarching takeaway message
Choices about graph type, text, arrangement, color, and lines should reinforce the same takeaway message.

2 1 0 n/a

For more support, check out:

AnnKEmery.com/blog

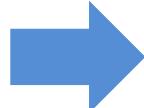
StephanieEvergreen.com/blog

Stephanie Evergreen's books, *Presenting Data Effectively* & *Effective Data Visualization*

Score: _____ / _____ = _____ %

Well-formatted data visualizations score between 90-100% of available points.
At this level, viewers are better able to read, interpret, and retain content.

Contents

- 
- Introduction to visualisation
 - Cairo's Five Principles
 - C.R.A.P. Design principles
 - Workshop
 - Directory of visualisations

Workshop

- Reduce mental burden in plots
 - Focus and declutter:
 - Use colour to draw attention
 - Have a sensible sensible ordering of your data
 - Make sure your audience does not need to turn their head
 - Simplify aesthetics and highlight
 - Make annotations and direct labels with `geom_text()` and `geom_label()`,
 - Use `ggrepel` for overlapping labels
 - Add arrows using `geom_curve()`
 - Facets and shadows to plot overlapping data
 - Combine plots with patchwork

We need to do everything we can to help our readers understand the meaning of our visualizations and see the same patterns in the data that we see. This usually means less is more. Simplify your figures as much as possible. Remove all features that are tangential to your story.

-- Claus O. Wilke

Final Project proposal

MAM 2022 Data Visualisation

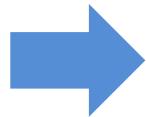
Final Group Project – Poster Presentation

In the final group project for the *Data Visualisation* course, you will prepare a poster presentation where you explore, analyse, and “tell a story” on a data set of interest to you. Your goal should be to create an original project that you would be proud to show off to your friends, a potential employer, etc. Posters will be displayed in the style of a professional conference, scheduled for the last day of classes. **A proposal is due on 10-Nov-2021.**

1. **Executive summary:** one-page summary of your questions, methods, findings, and recommendations
2. **Introduction and description of research questions:** describe the motivation for this study, outline and define what questions you are exploring and why
3. **Data and methods:** explain how the data was collected, provide basic summary statistics (tables and figures) of the main variables you're interested in, and describe what statistical tools you will use to answer your questions (i.e. regression, bootstrapped comparisons of means, etc.)
4. **Results:** answer each of your questions using statistical tools and interpret the results of the different statistical tests you use
5. **Limitations of the study:** provide caveats for your analysis and explain how confident you are in your results
6. **Recommendations and conclusion:** discuss the implications of these findings and make recommendations based on the results
7. **Appendices:** if you want to include tables of summary statistics or tables showing alternative models, you can include them in an appendix instead of in the body of the report itself.

Contents

- Introduction to visualisation
- Cairo's Five Principles
- C.R.A.P. Design principles
- Workshop
- Directory of visualisations



Directory of visualisations

serialmentor.com/dataviz/directory-of-visualizations.html

5 Directory of visualizations

This chapter provides a quick visual overview of the various plots and charts that are commonly used to visualize data. It is meant both to serve as a table of contents, in case you are looking for a particular visualization whose name you may not know, and as a source of inspiration, if you need to find alternatives to the figures you routinely make.

5.1 Amounts



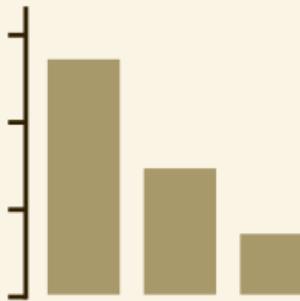
The page contains three small, square visualizations arranged horizontally. The first two are labeled 'Bars' and show different configurations of vertical bars of varying heights. The third is labeled 'Dots' and shows three individual data points plotted against a horizontal axis.

- 5.1 Amounts
- 5.2 Distributions
- 5.3 Proportions
- 5.4 x-y relationships
- 5.5 Geospatial data
- 5.6 Uncertainty
- 6 Visualizing amounts

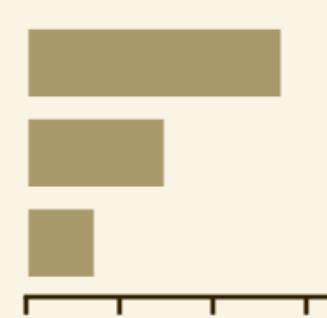
<https://serialmentor.com/dataviz/directory-of-visualizations.html>

Descriptive statistics like averages and counts by one or two categorical groups (covariates or features). These use absolute values, rather than values, therefore scale matters.

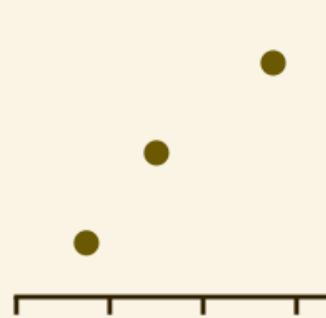
bars



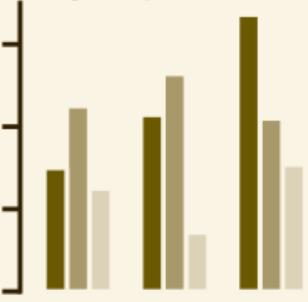
bars



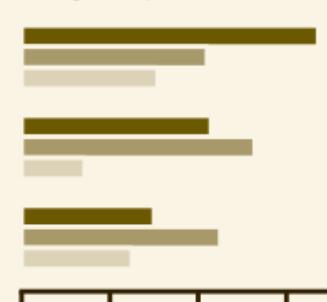
dots



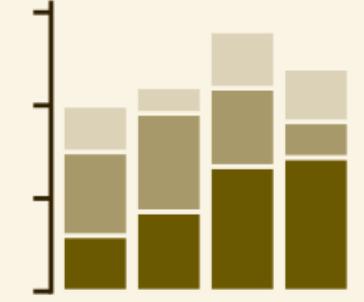
grouped bars



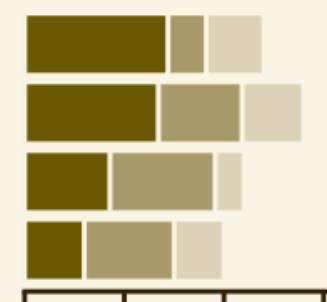
grouped bars



stacked bars



stacked bars



Amounts – General Rules

More data = better

Show actual points

Don't use bars for
summary stats

Lollipops, points,
heatmaps

The end of the bar is
often all that matters

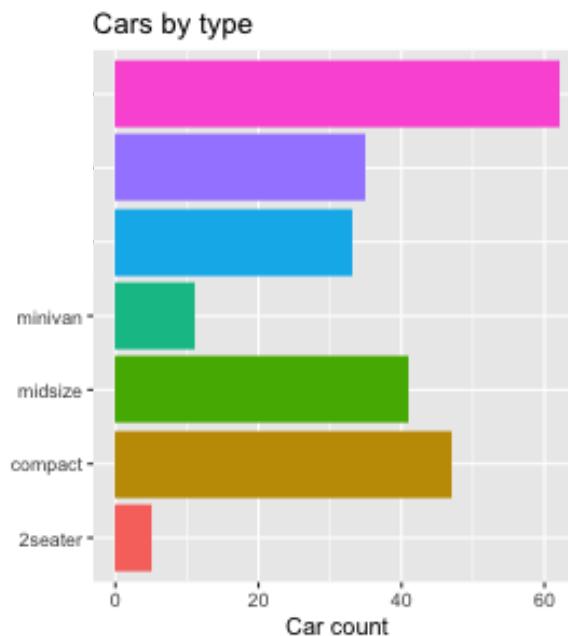
Counts okay, but there
are better solutions

Always start at zero!

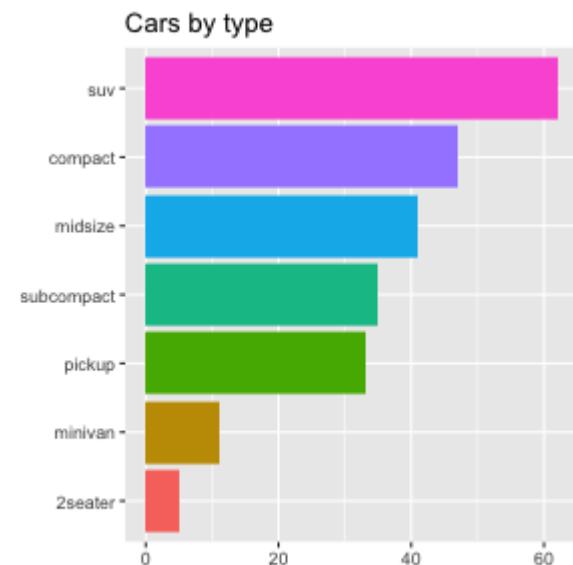
Amounts



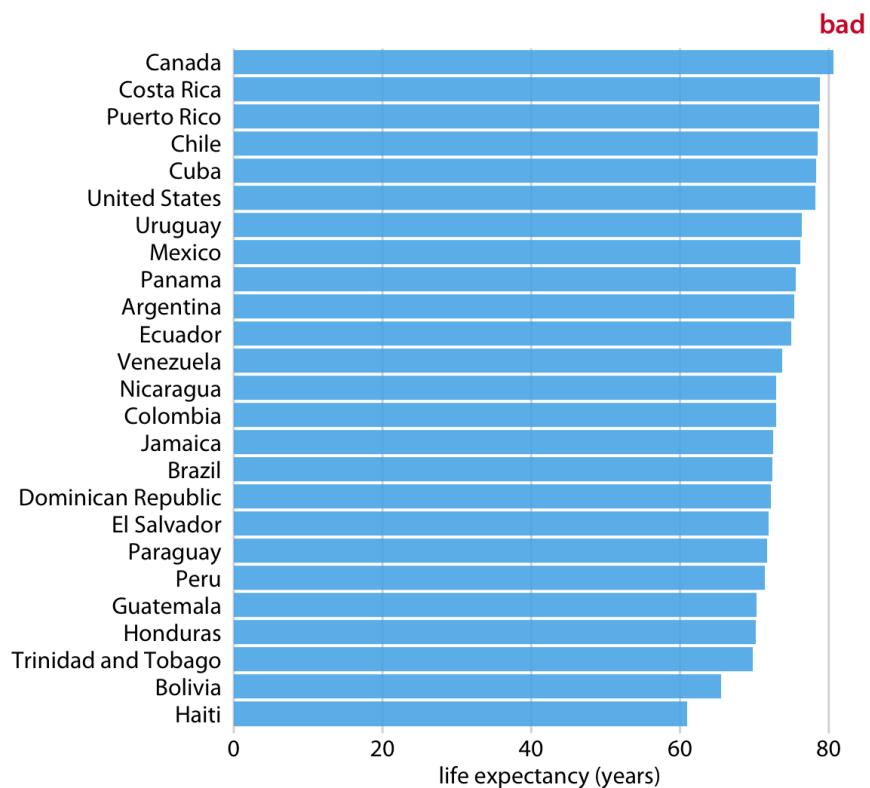
```
ggplot(mpg, aes(x = class, fill = class)) +  
  geom_bar() +  
  coord_flip() +  
  labs(x = " ",  
       y = "Car count",  
       title = "Cars by type") +  
  theme(legend.position = "none")
```



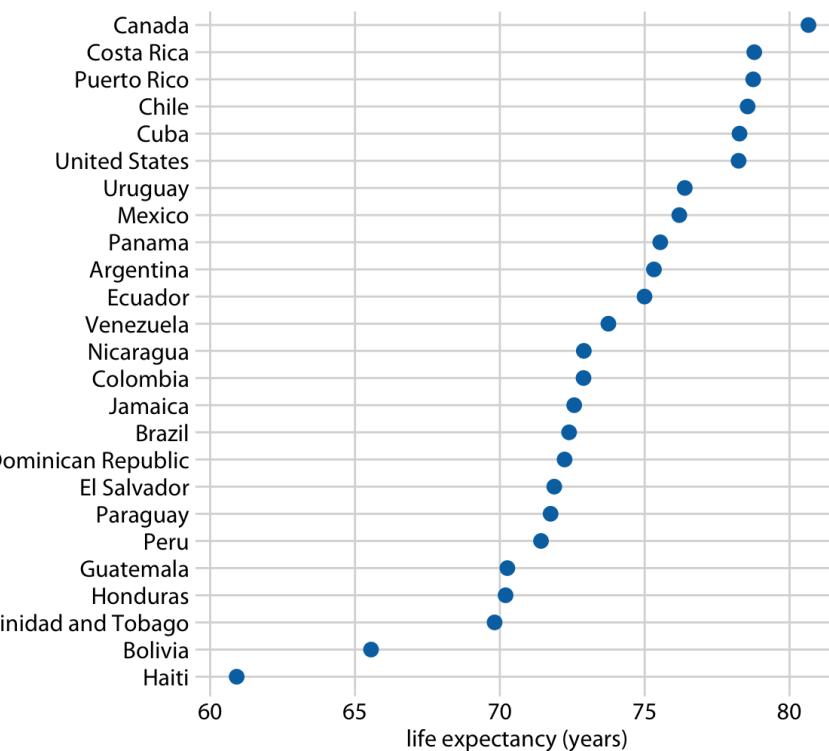
```
1 <- c("2seater", "minivan", "pickup",  
      "subcompact", "midsize", "compact", "suv")  
mpg %>%  
  mutate(class = factor(class, levels = 1)) %>%  
  ggplot(aes(x = class, fill = class)) +  
  geom_bar() +  
  coord_flip() +  
  labs(x = " ", y = "Car count",  
       title = "Cars by type") +  
  theme(legend.position = "none")
```



Points instead of bars



All bars are long in this figure and the eye is drawn to the middle of the bars rather than to their end points.
Figure fails to convey its message.



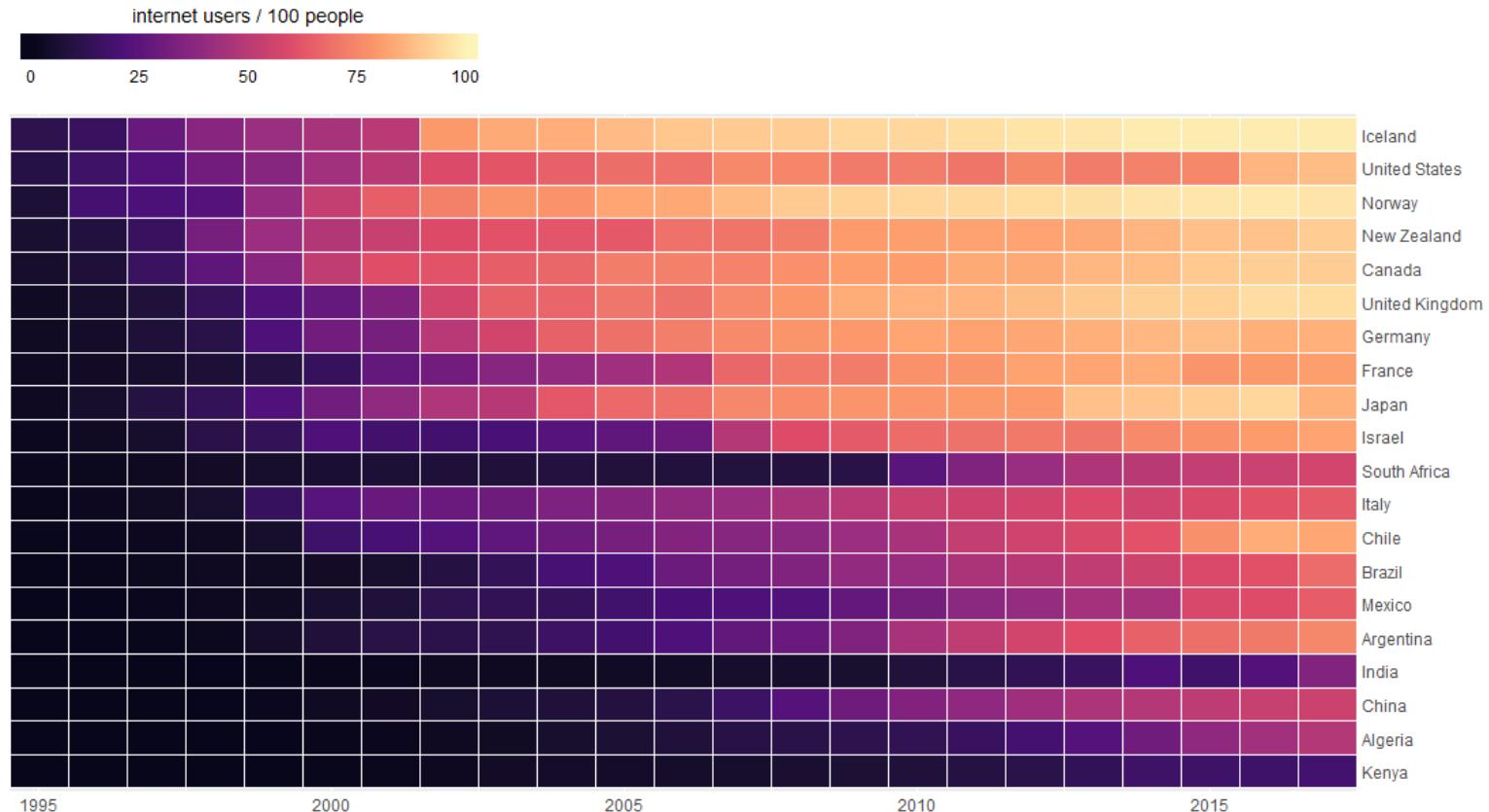
By limiting axis from 60 to 81 years, we highlight the key feature: Canada has the highest life expectancy, whereas Haiti and Bolivia have much lower life expectancies
A much more compelling figure

Heatmaps

Instead of mapping onto bars or dots, we can map onto colors.

Heatmap shows % of internet users over time in 20 countries from 1995 to 2017.

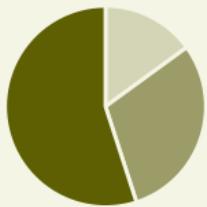
Highlights trends, but does not show exact values



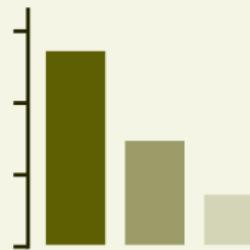
Proportions

Relative values to compare sizes of categories.

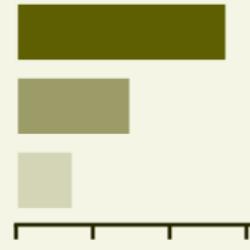
pie chart



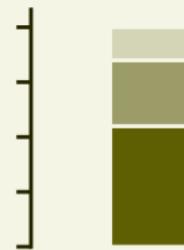
bars



bars

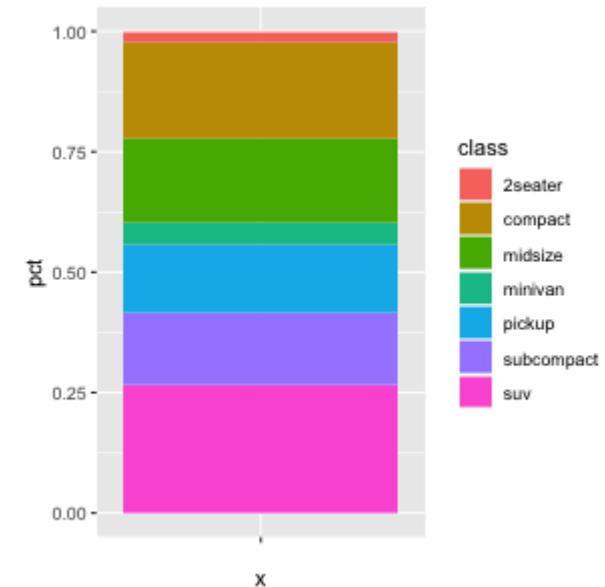


stacked bars



```
p <- mpg %>%
  count(class) %>%
  mutate(pct = n / sum(n)) %>%
  ggplot(aes(x = "", y = pct, fill = class)) +
  geom_bar(width = 1, stat = "identity")
```

p

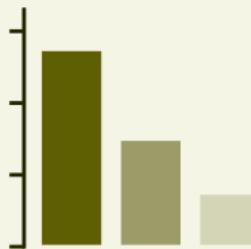


Proportions

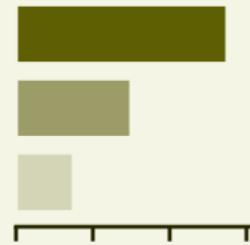
pie chart



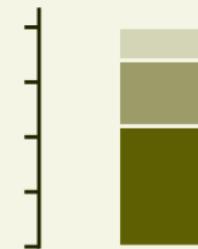
bars



bars



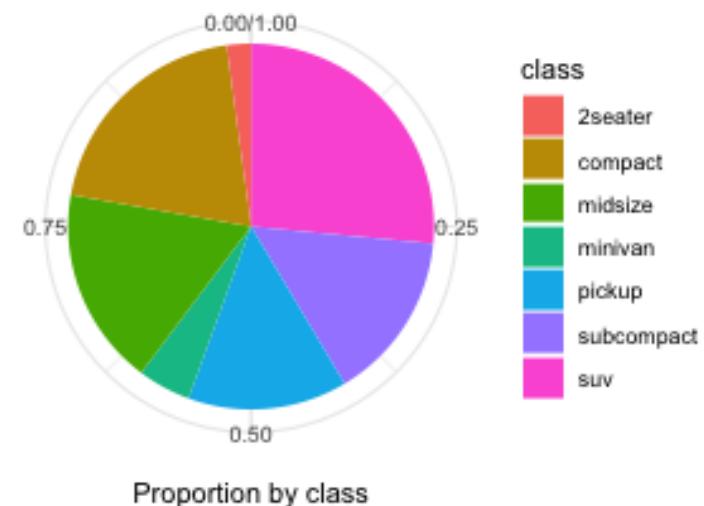
stacked bars



```
p <- mpg %>%
  count(class) %>%
  mutate(pct = n / sum(n)) %>%
  ggplot(aes(x = "", y = pct, fill = class)) +
  geom_bar(width = 1, stat = "identity")
```

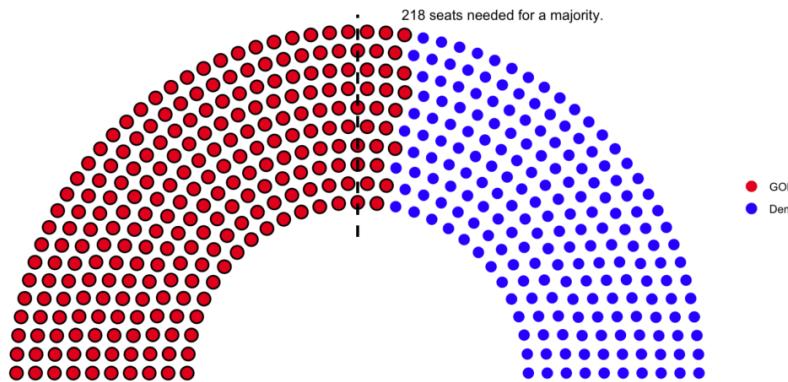
```
p + coord_polar("y", start=0) +
  theme_minimal() +
  labs(x = " ", y = "Proportion by class")
```

```
p
```

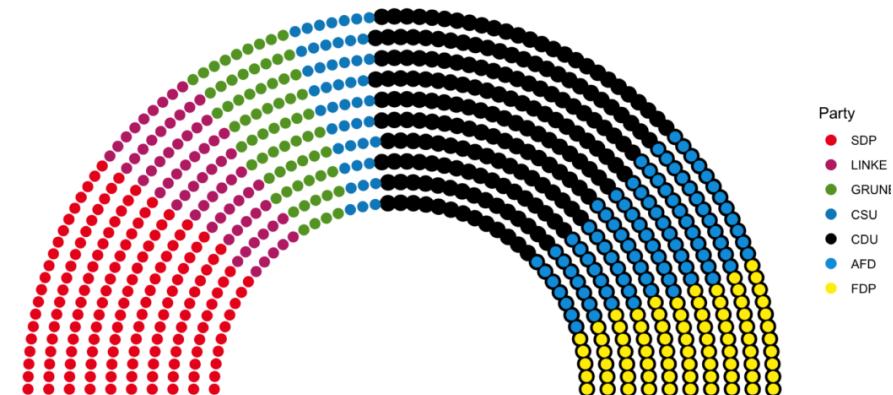


Pie chart alternatives

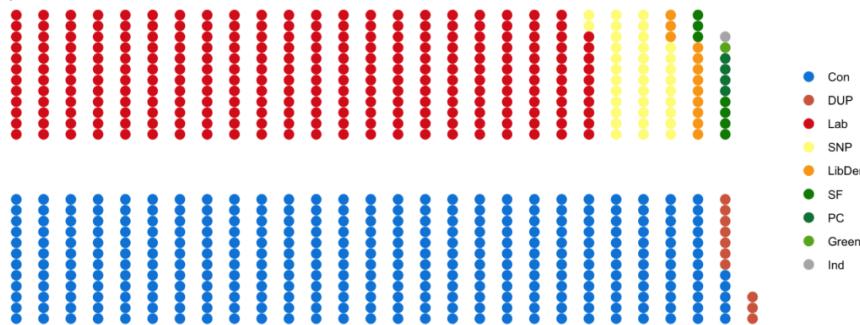
United States House of Representatives
Party that controls the House highlighted.



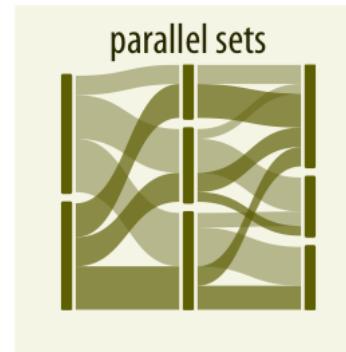
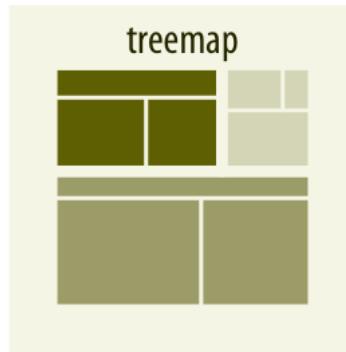
German Bundestag
Government circled in black.



UK parliament in 2017

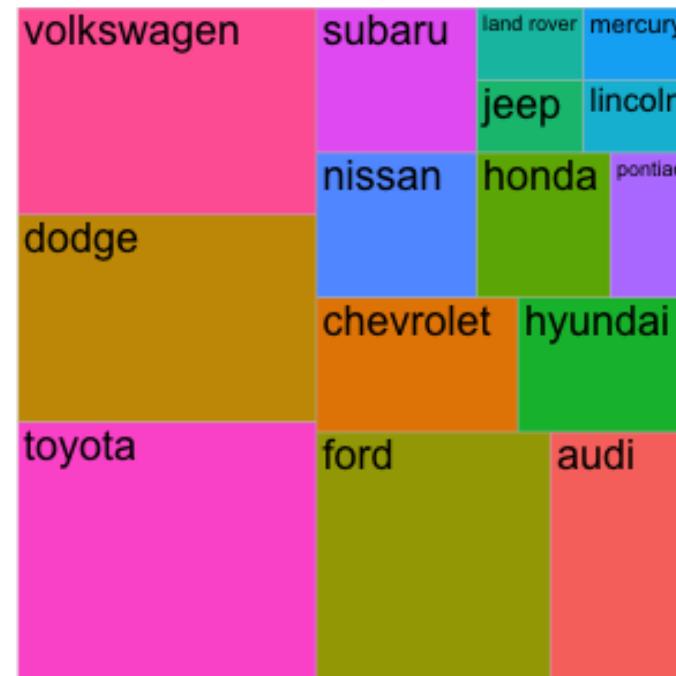


Proportions

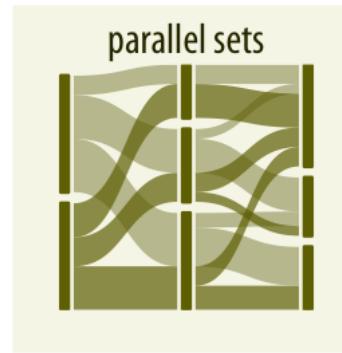
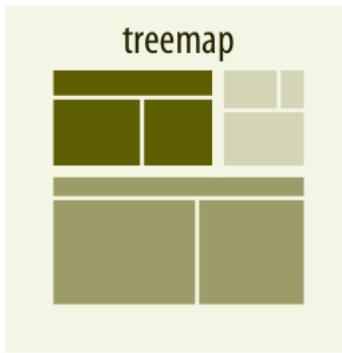


```
library(treemapify)
```

```
mpg %>%  
filter(year == 1999) %>%  
count(manufacturer) %>%  
ggplot(aes(area = n,  
           fill = manufacturer,  
           label = manufacturer)) +  
geom_treemap() +  
geom_treemap_text() +  
theme(legend.position = "none")
```

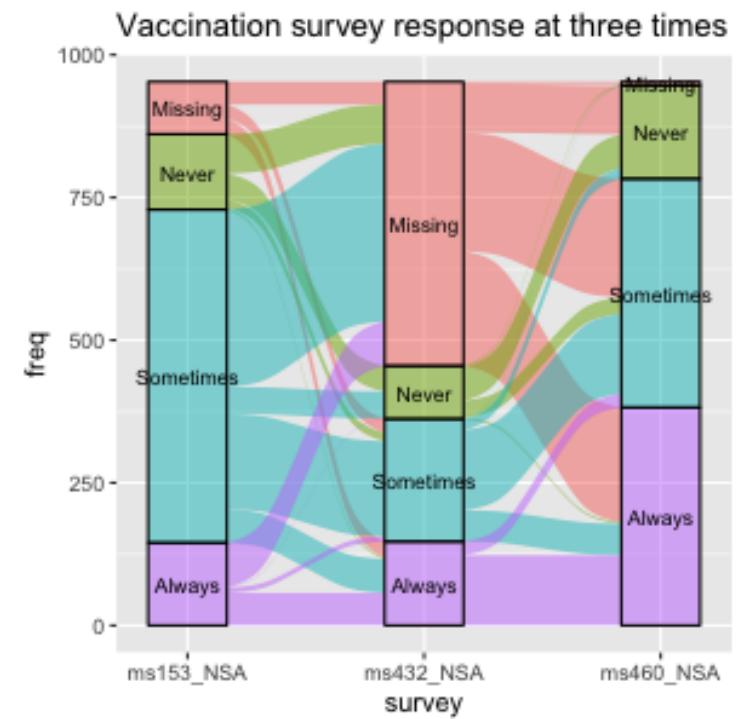


Proportions

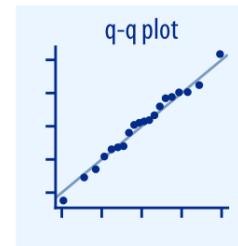
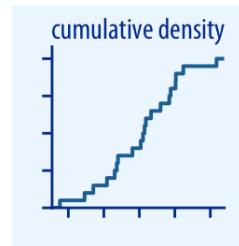
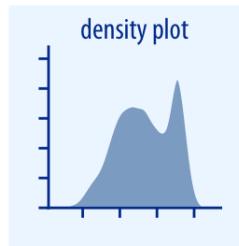
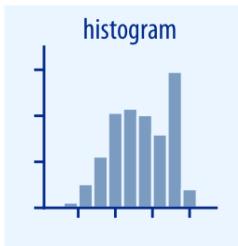


```
library(ggalluvial)  
data(vaccinations)
```

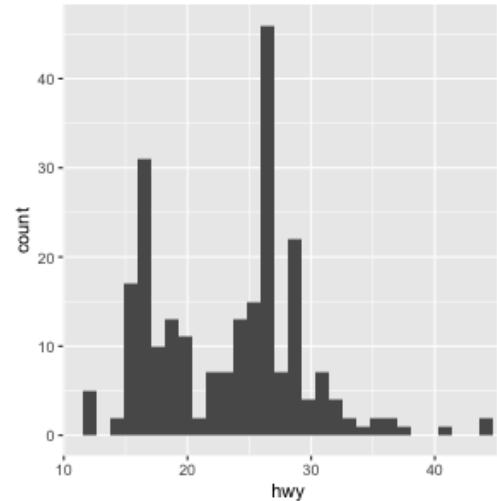
```
ggplot(vaccinations,  
       aes(x = survey, y = freq,  
            alluvium = subject, stratum = response,  
            fill = response, label = response)) +  
  scale_x_discrete(expand = c(.1, .1)) +  
  geom_flow() +  
  geom_stratum(alpha = .5) +  
  geom_text(stat = "stratum", size = 3) +  
  theme(legend.position = "none") +  
  labs(title = "Vaccination survey response at  
three times")
```



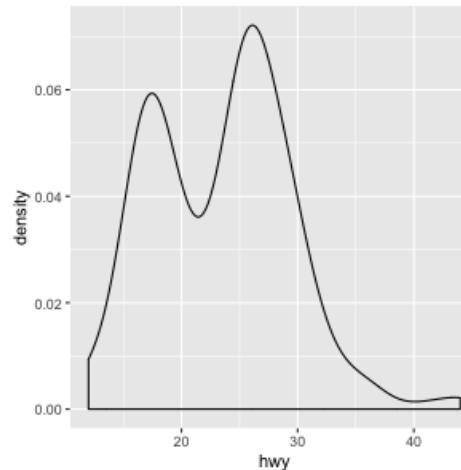
Distributions



```
ggplot(mpg, aes(x = hwy)) +  
  geom_histogram()
```

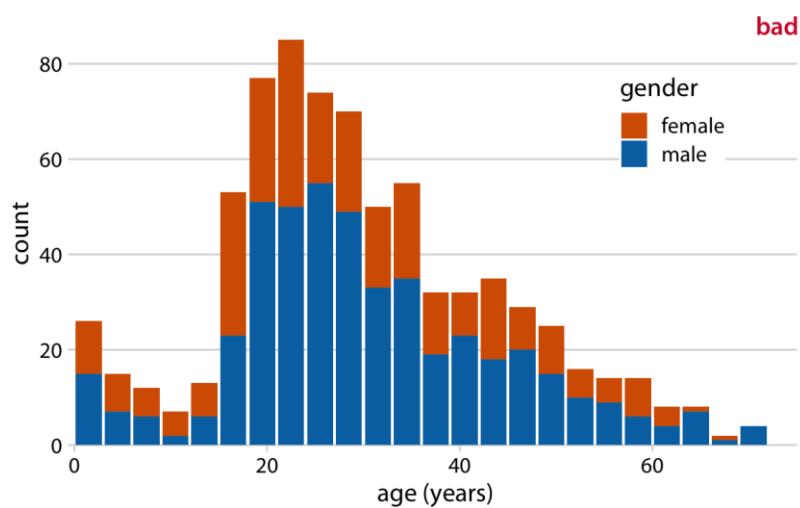


```
ggplot(mpg, aes(x = hwy)) +  
  geom_density()
```

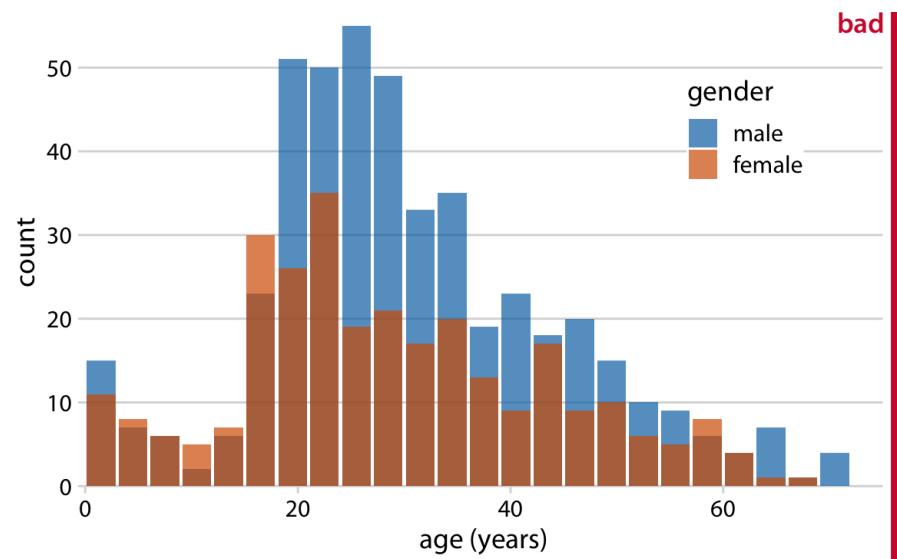


Visualising multiple distributions (1/2)

Stacked histograms are easily confused with overlapping histograms

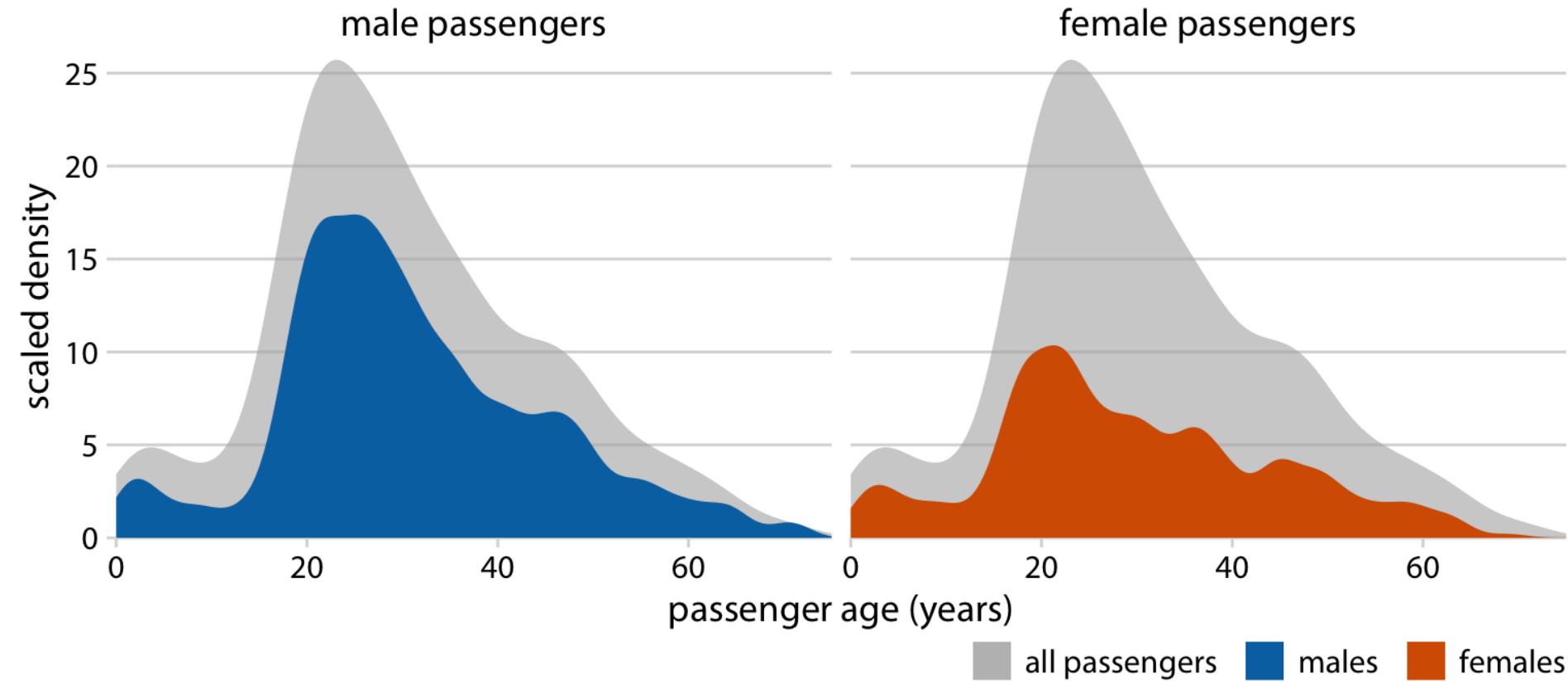


Where does each bar start and end? A semi-transparent bar drawn on top of another tends to look like like a bar drawn in a different color.



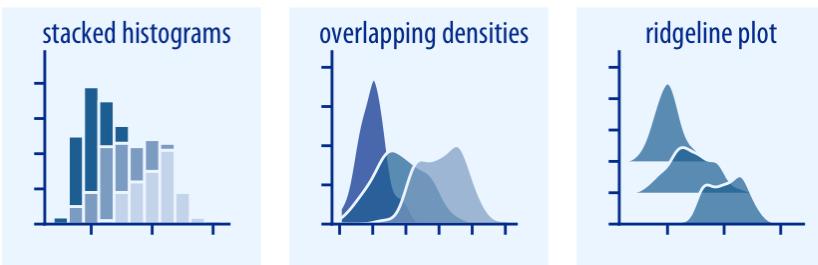
Visualising multiple distributions (2/2)

To visualize multiple distributions, density plots generally work better than histograms.

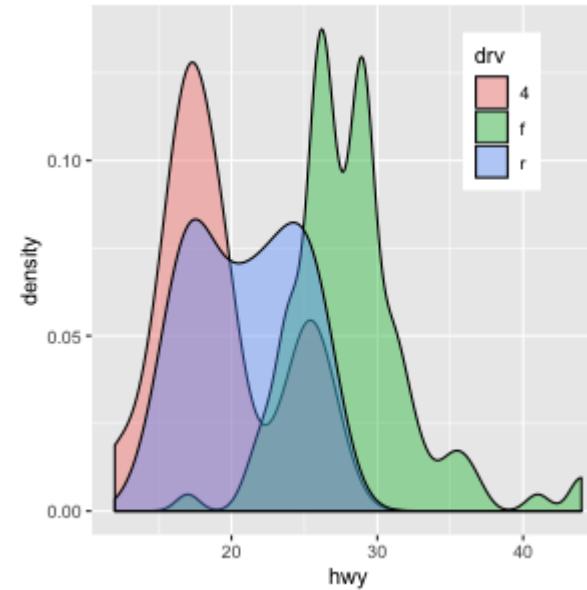


Show male and female age distributions separately, each as a proportion of the overall age distribution. This visualization shows intuitively and clearly that there were many fewer women than men in the 20–50-year age range on the Titanic.

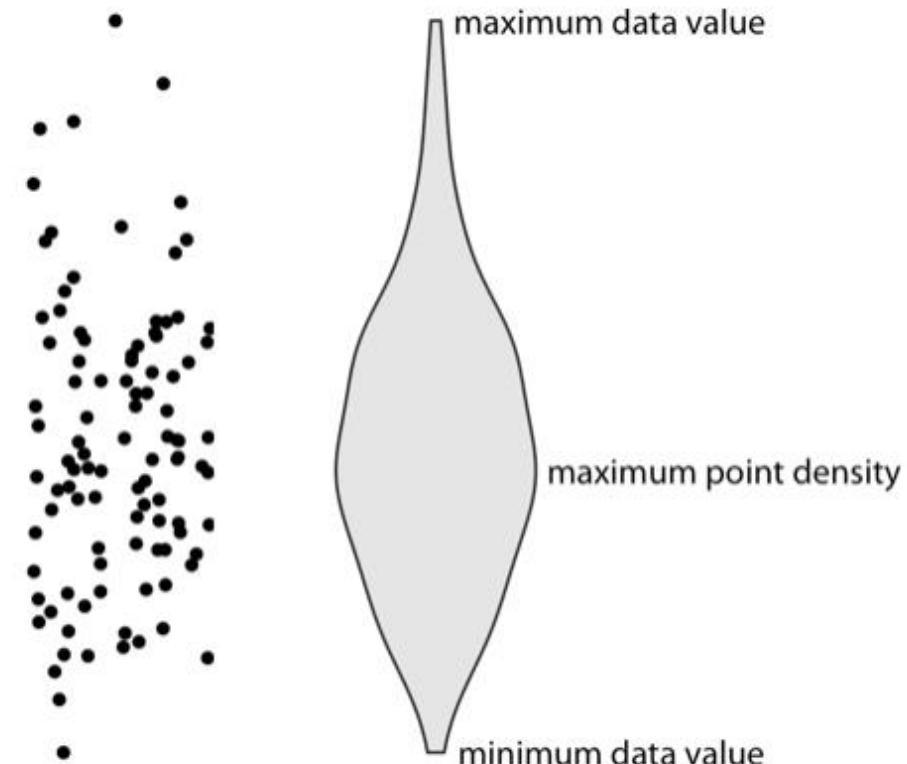
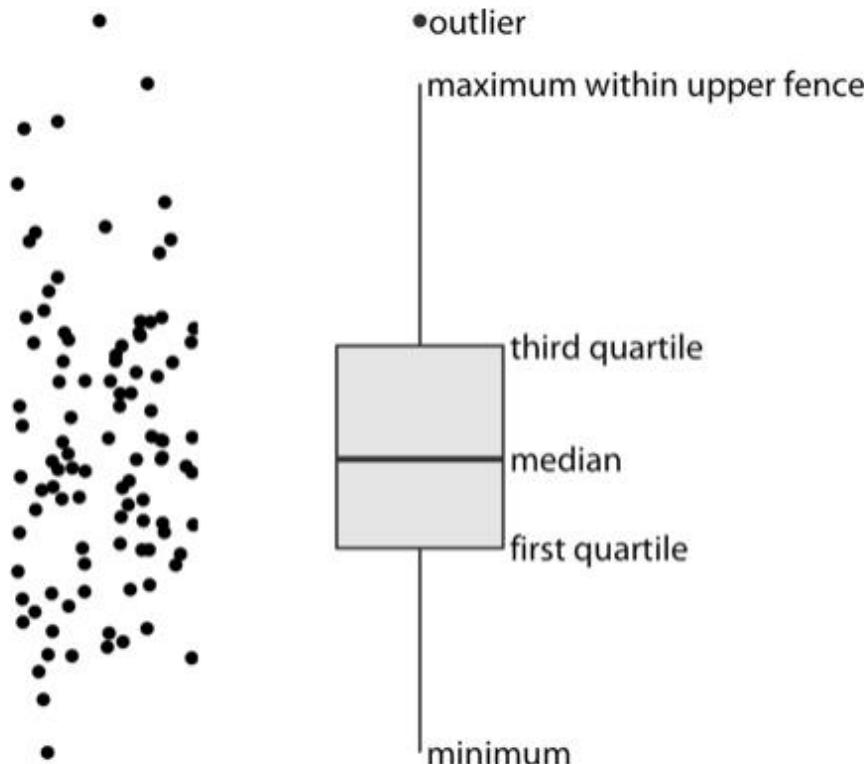
Distributions



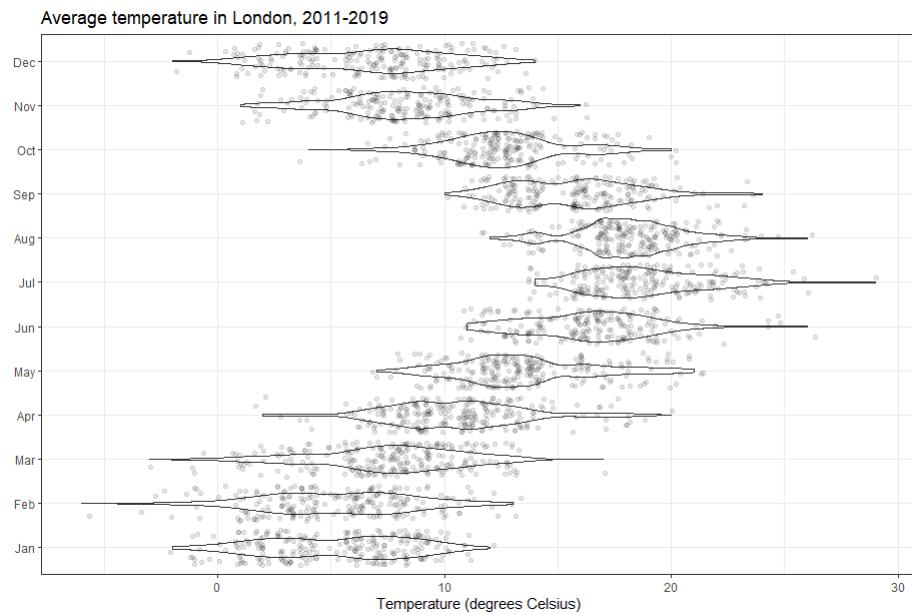
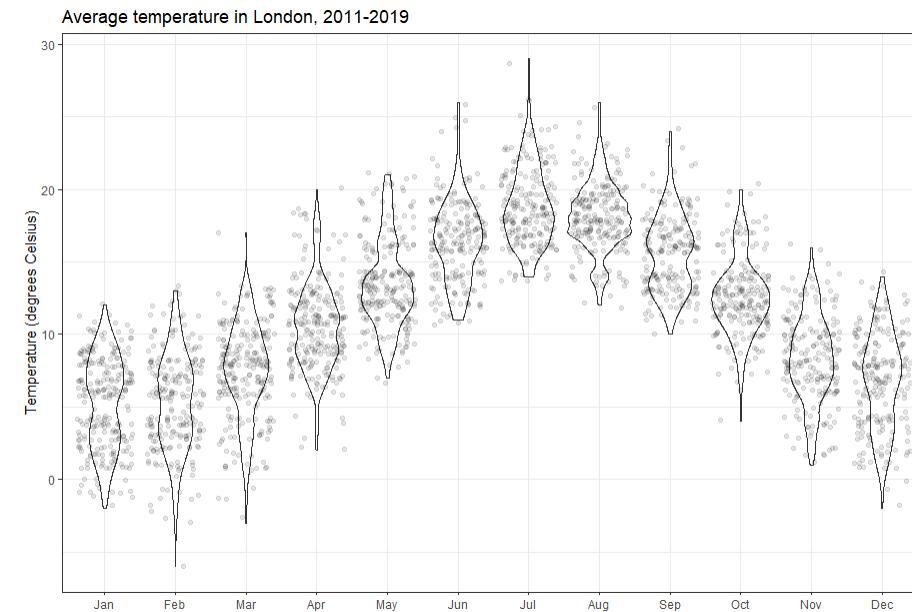
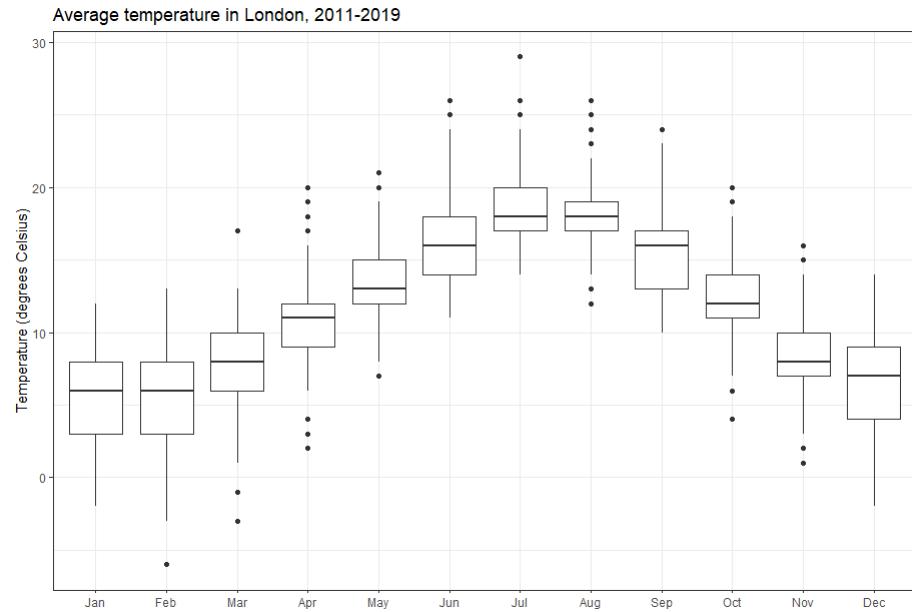
```
ggplot(mpg, aes(x = hwy, fill = drv)) +  
  geom_density(alpha = 0.4) +  
  theme(legend.position = c(0.8,0.8))
```

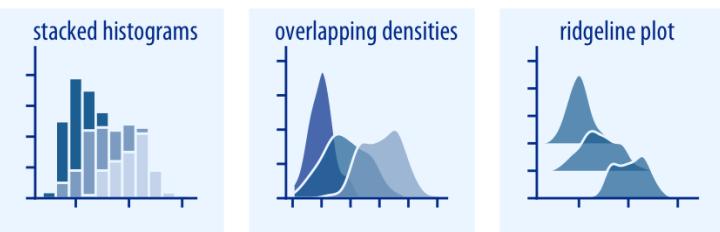


Boxplots vs violin plots



Boxplots vs violin plots

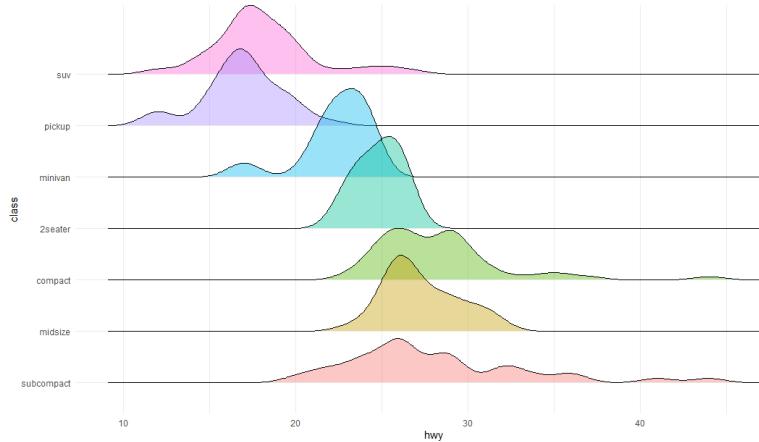




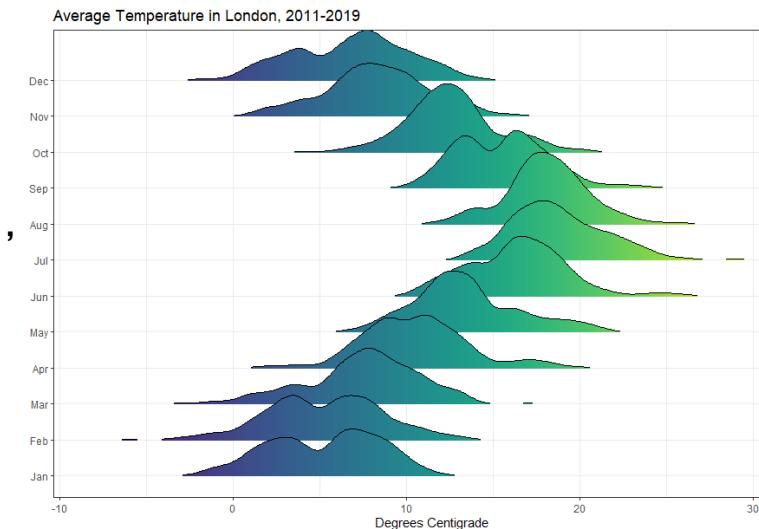
Distributions

```
library(ggrridges)
```

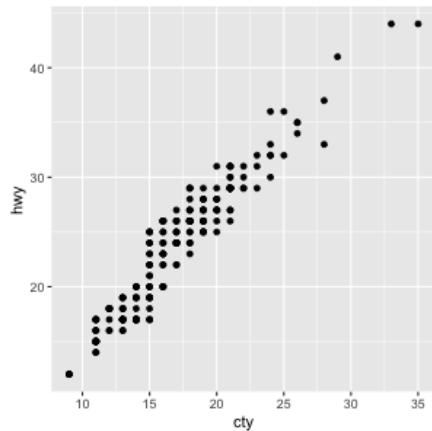
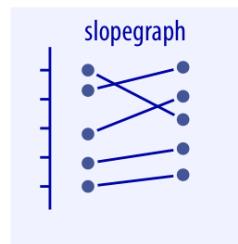
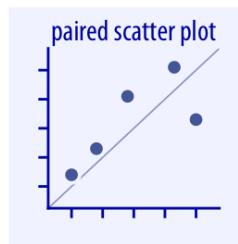
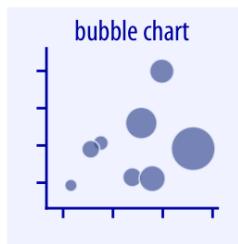
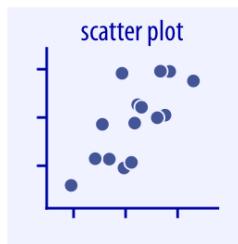
```
12 <- c("subcompact", "midsize", "compact",
      "2seater", "minivan", "pickup", "suv")
mpg %>%
  mutate(class = factor(class, levels = 12)) %>%
  ggplot(aes(x = hwy, y = class, fill = class)) +
  geom_density_ridges(alpha = 0.4) +
  theme_minimal() +
  theme(legend.position = "none")
```



```
ggplot(bike,
       aes(x=avg_temp, y=month_name, fill = . . .)) +
  geom_density_ridges_gradient(scale = 2,
                                rel_min_height = 0.01) +
  theme_bw() +
  labs(title="Average Temperature in London, 2011-2019",
       x=Degrees Centigrade', y=NULL) +
  theme(legend.position = "none")
```

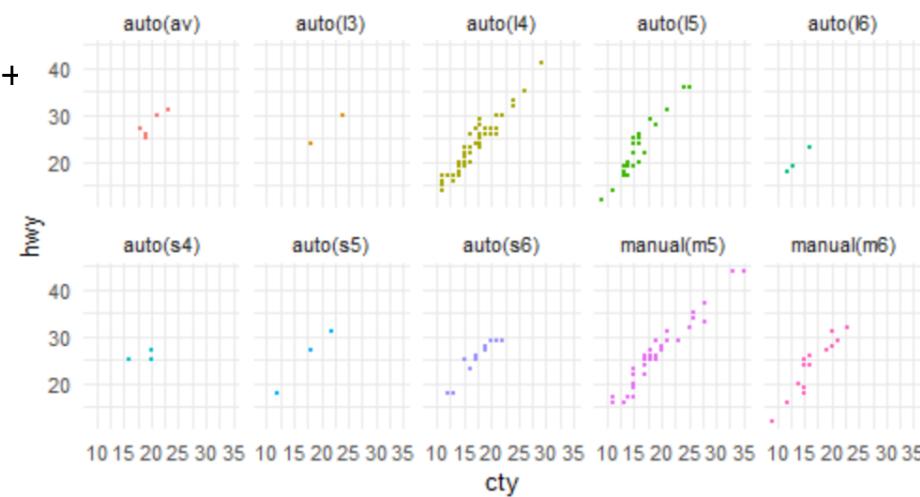


X-Y relationships

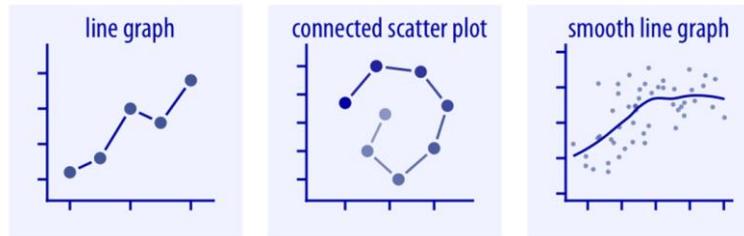


```
ggplot(mpg, aes(x = cty, y = hwy)) +  
  geom_point()
```

```
ggplot(mpg, aes(x = cty, y = hwy)) +  
  geom_point(aes(colour=trans), size = 0.5) +  
  facet_wrap(~trans, nrow=2) +  
  theme_minimal() +  
  theme(legend.position = "none",  
        text = element_text(size=10))
```



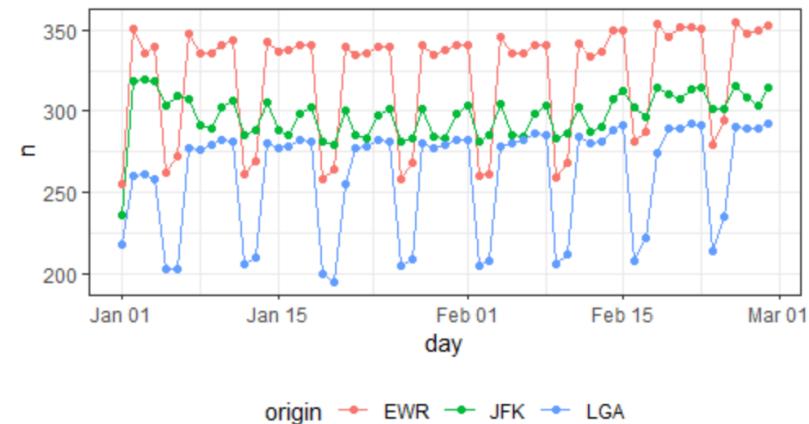
X-Y relationships



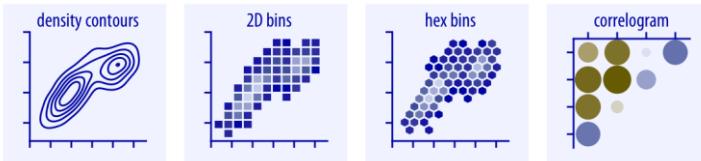
```
library(nycflights13)

# just filter to have first 2 months
df <- flights %>%
  mutate(day=as.Date(time_hour)) %>%
  filter(day < "2013-03-01") %>%
  count(day,origin)

# and plot
ggplot(df, aes(x=day, y=n, colour=origin)) +
  geom_line(aes(group=origin)) +
  geom_point() +
  theme_bw()+
  theme(legend.position="bottom")
```

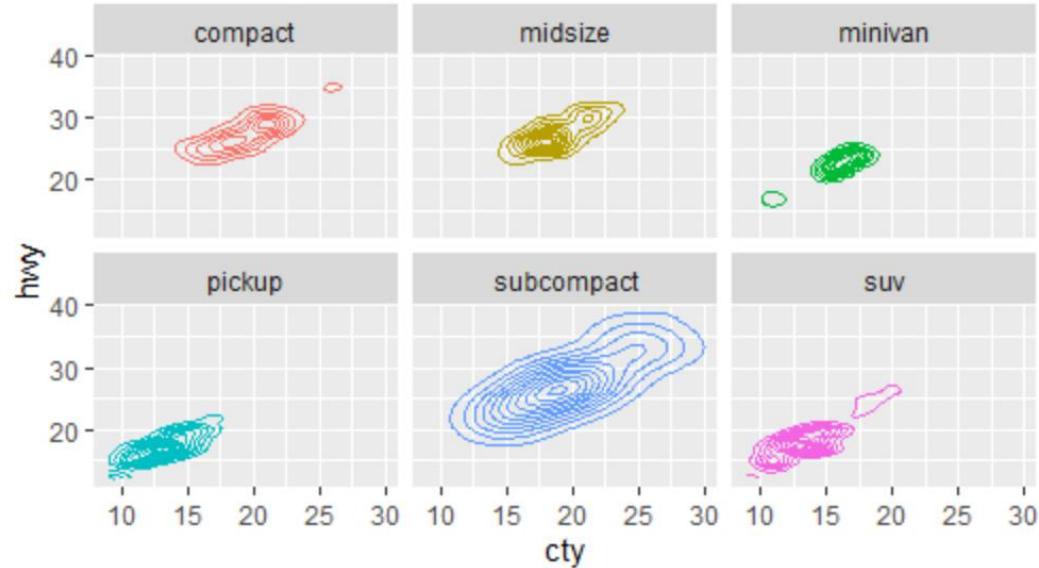


X-Y relationships



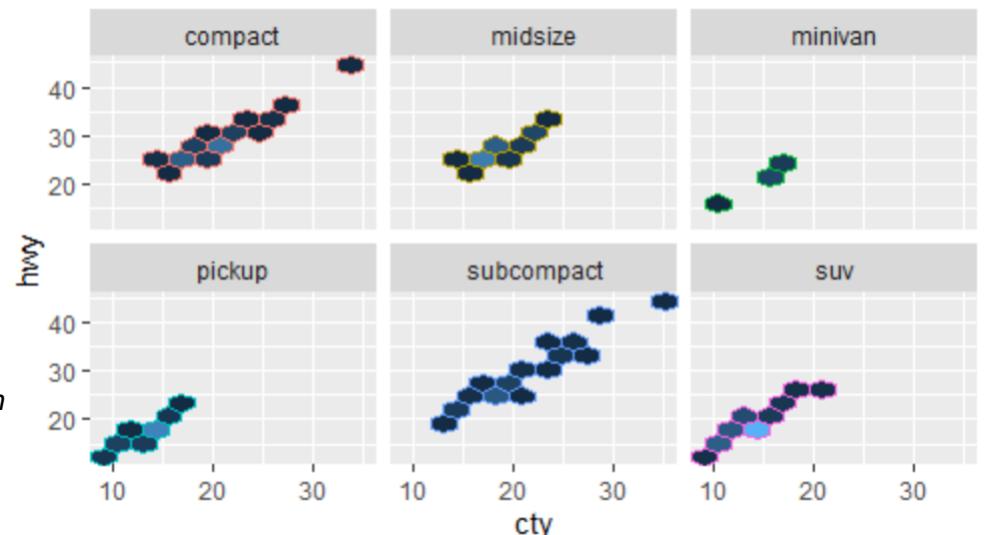
```
filter(mpg, class != "2seater") %>%  
ggplot(aes(x = cty, y = hwy)) +  
geom_density_2d(aes(colour = class)) +  
facet_wrap(~class) +  
theme(legend.position = "none")
```

Perform a 2D kernel density estimation and display the results with contours. This can be useful for dealing with overplotting. This is a 2d version of geom_density().

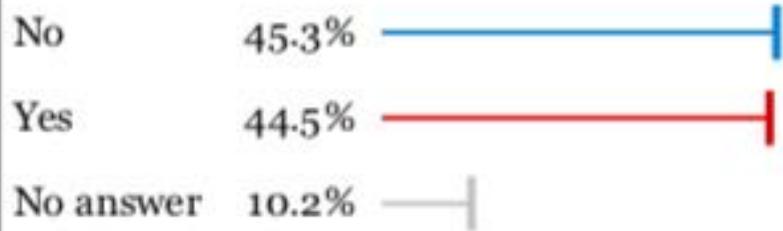


```
filter(mpg, class != "2seater") %>%  
ggplot(aes(x = cty, y = hwy)) +  
geom_hex(aes(colour=class), bins = 10)+  
facet_wrap(~class) +  
theme(legend.position = "none")
```

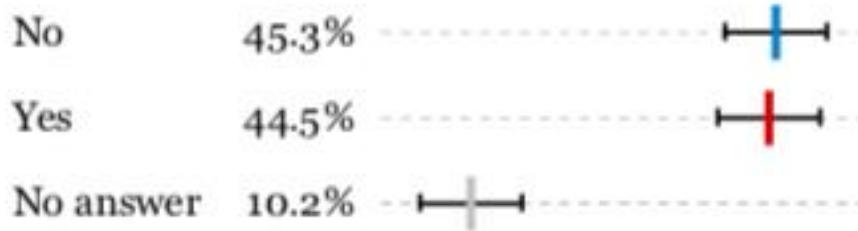
Divides the plane into regular hexagons, counts the number of cases in each hexagon, and then (by default) maps the number of cases to the hexagon fill.



*Do you want Catalonia
to become an independent state?*



*Do you want Catalonia
to become an independent state?*



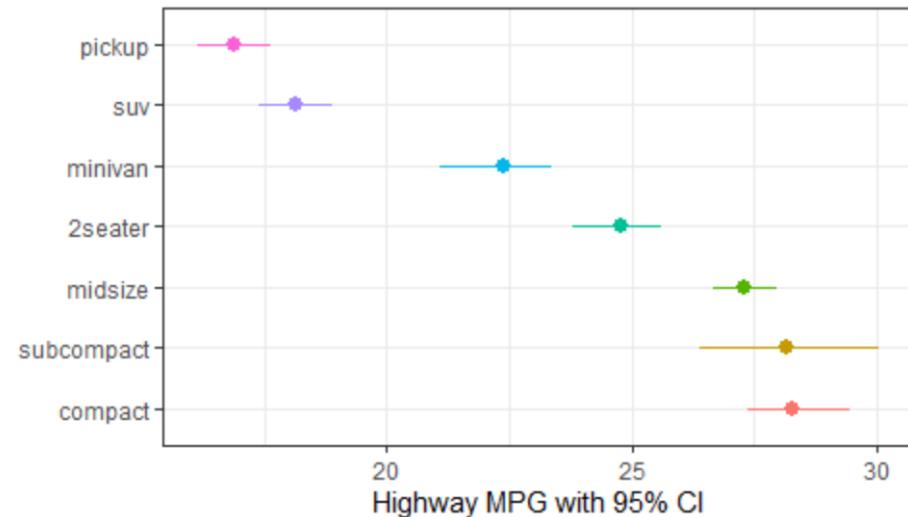
Margin of error: +/-2.95
Level of confidence = 95%

Figure 11.1 Displaying the margin of error can change your view of the data.

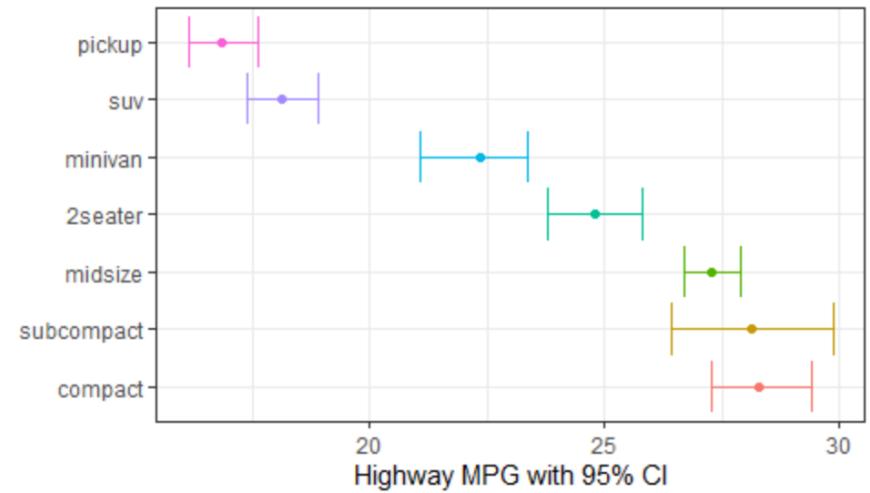
Uncertainty



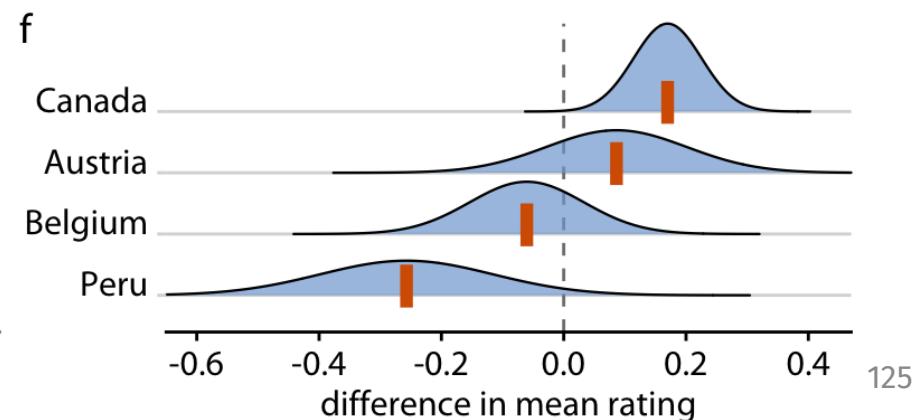
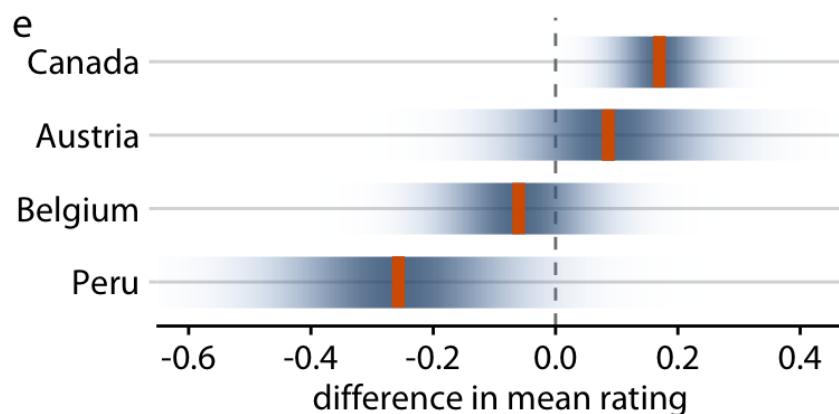
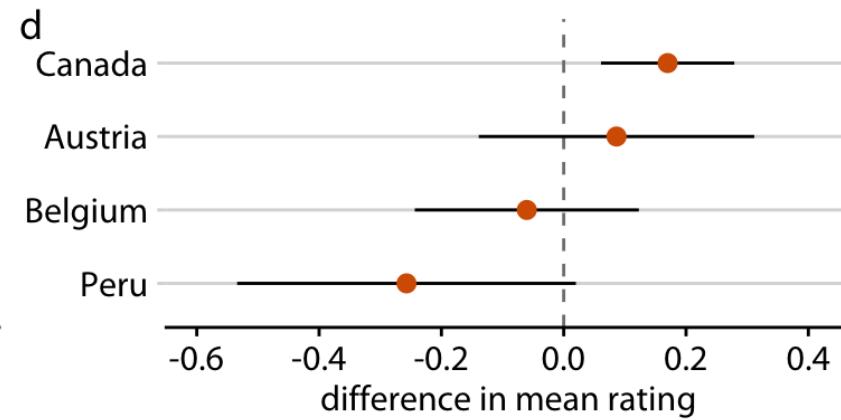
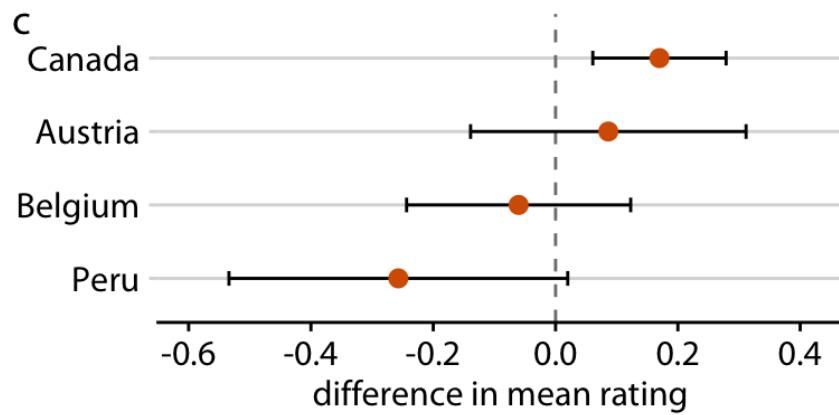
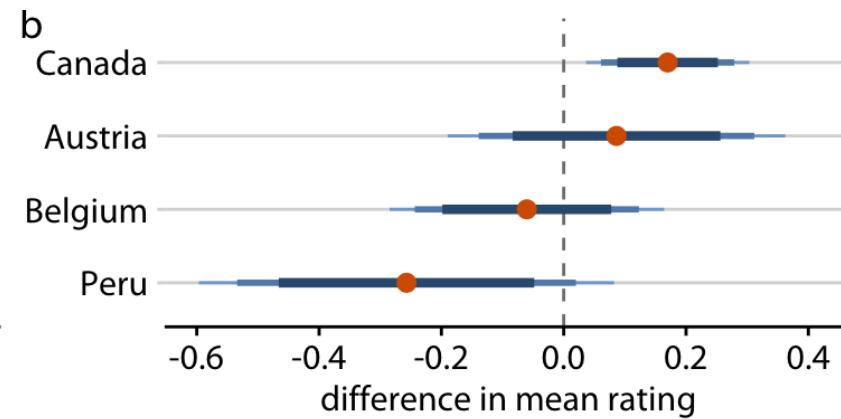
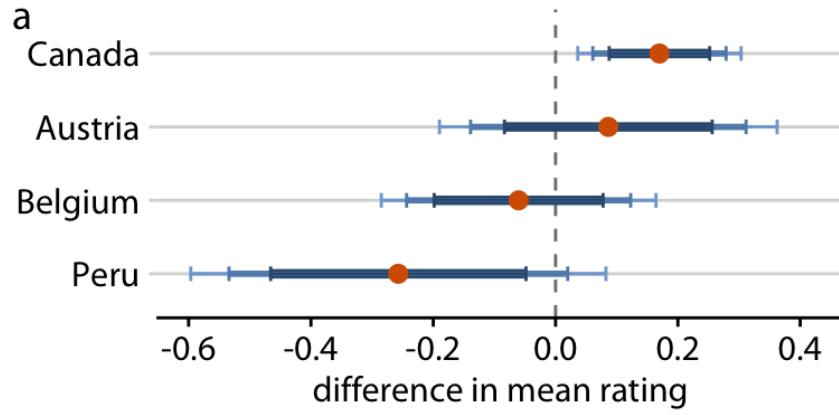
```
13 <- c("compact", "subcompact", "midsize",
      "2seater", "minivan", "suv", "pickup")
# avg highway mpg with bootstrapped 95% CI
mpg %>%
  mutate(class = factor(class, levels = 13)) %>%
  ggplot(aes(x = class, y = hwy, color = class)) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.data = mean_cl_boot,
               geom = "pointrange") +
  theme_bw() +
  coord_flip() +
  theme(legend.position = "none") +
  labs(x = " ", y = "Highway MPG with 95% CI")
```



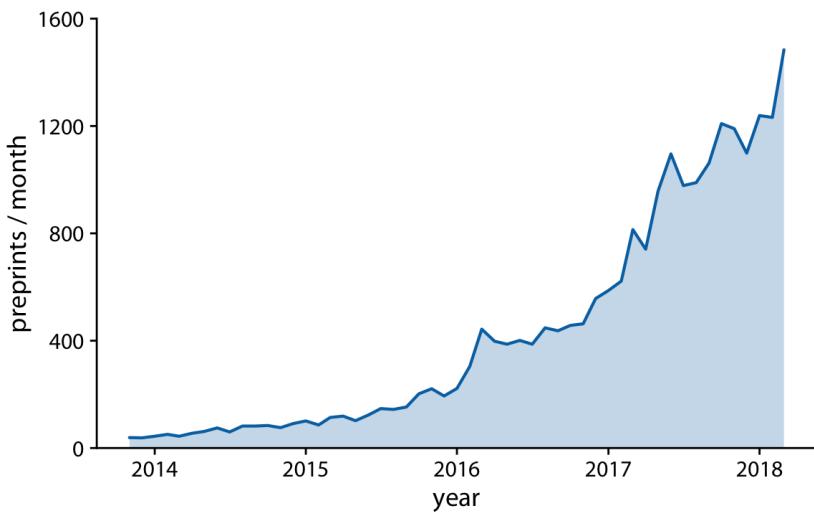
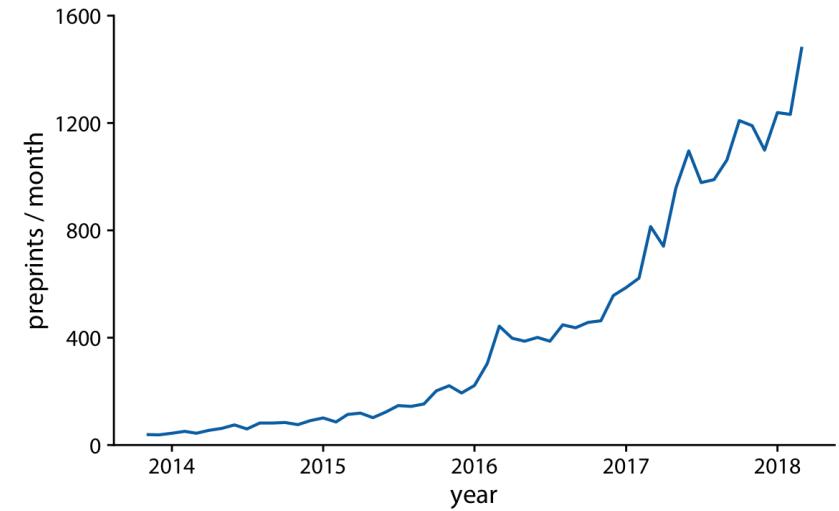
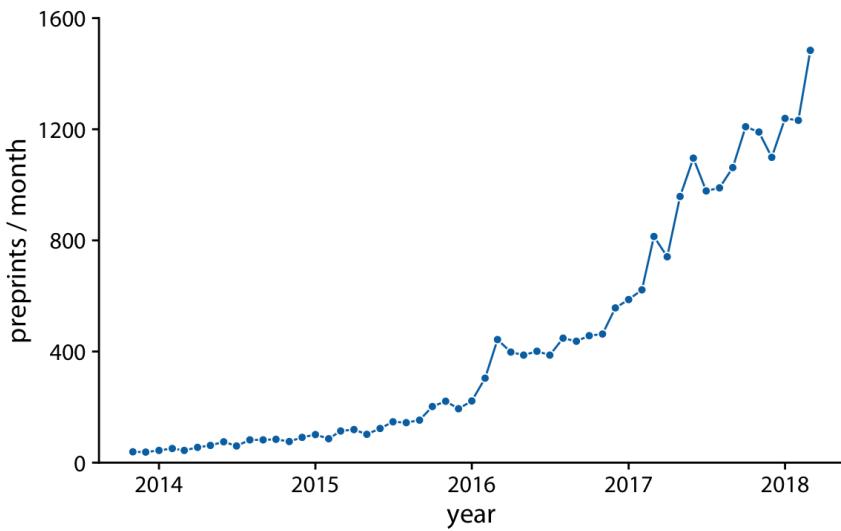
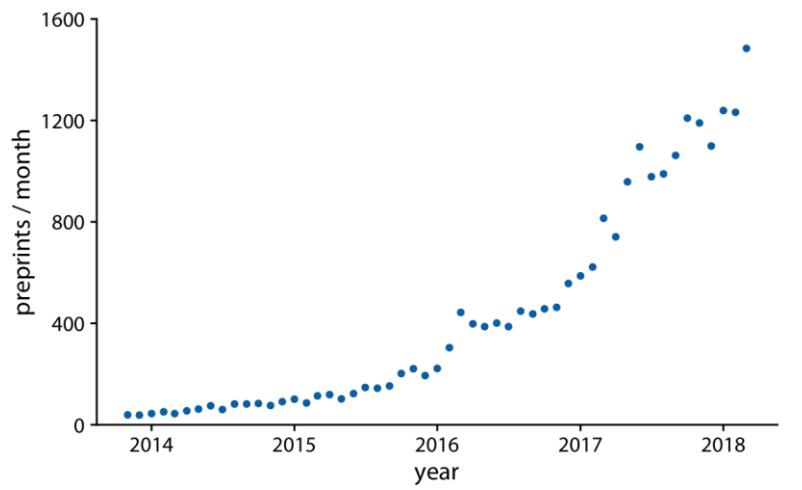
```
# avg highway mpg with bootstrapped 95% CI
mpg %>%
  mutate(class = factor(class, levels = 13)) %>%
  ggplot(aes(x = class, y = hwy, color = class)) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.data = mean_cl_boot,
               geom = "errorbar") +
  theme_bw() +
  coord_flip() +
  theme(legend.position = "none") +
  labs(x = " ", y = "Highway MPG with 95% CI")
```



Uncertainty



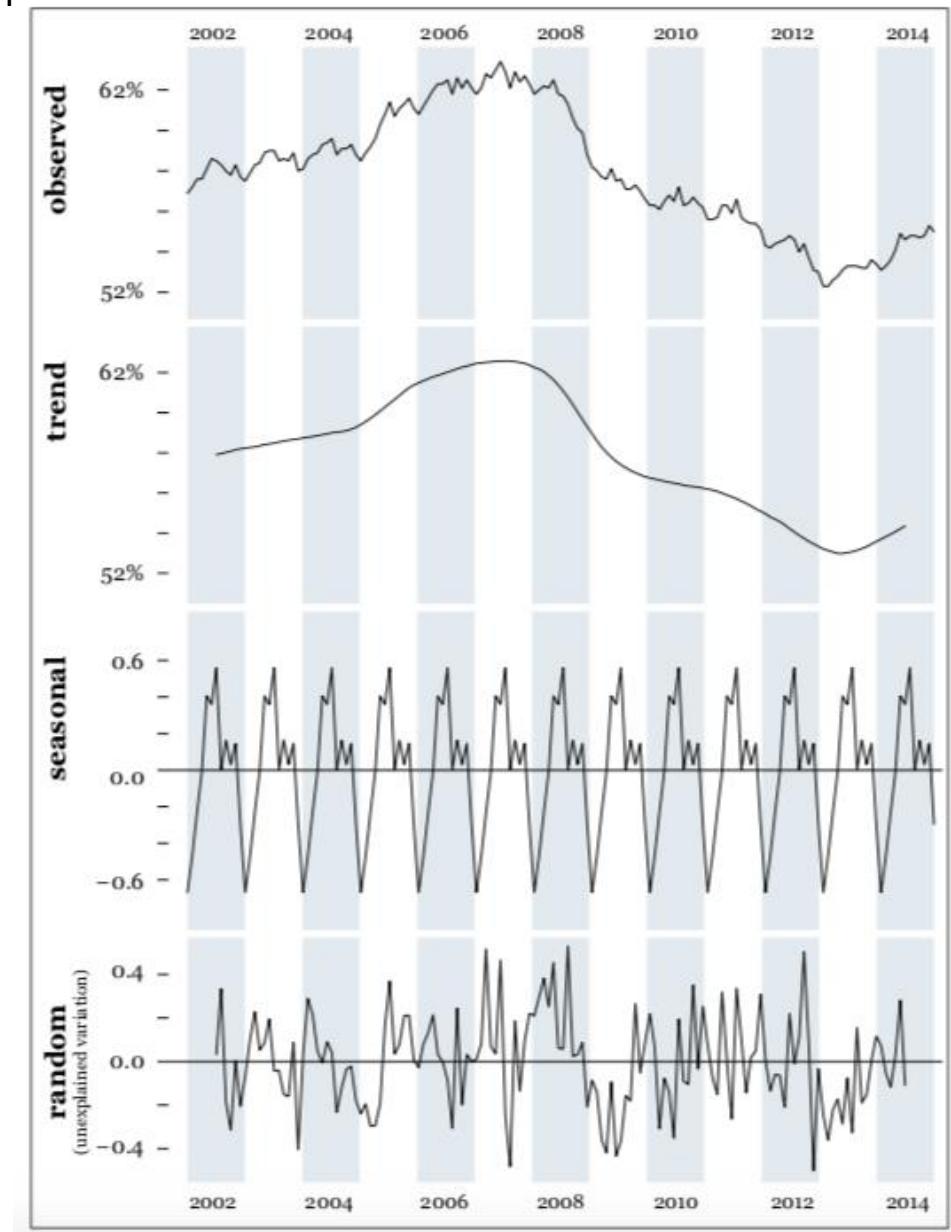
Time series



Trend +

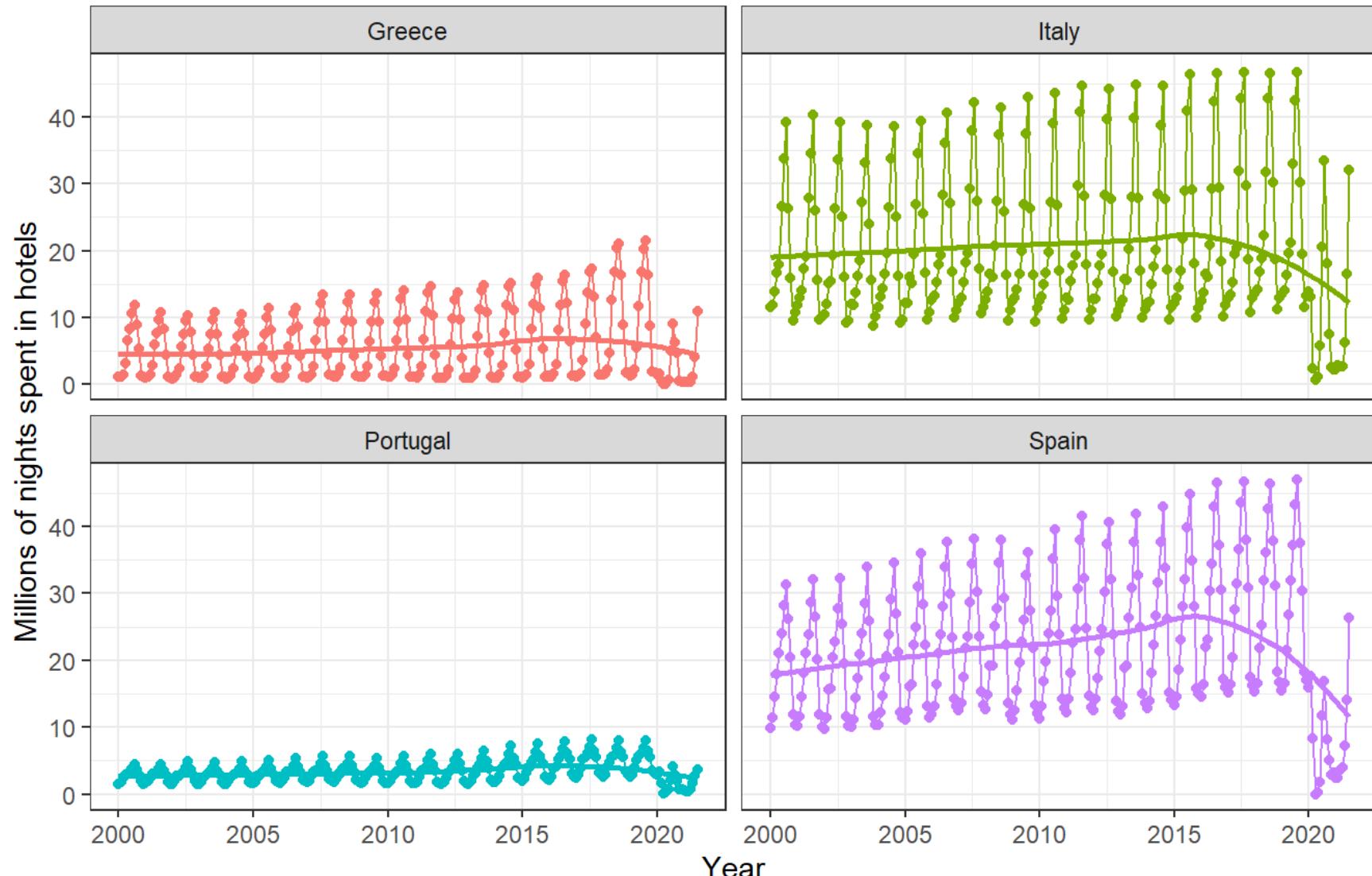
Seasonality +

Noise



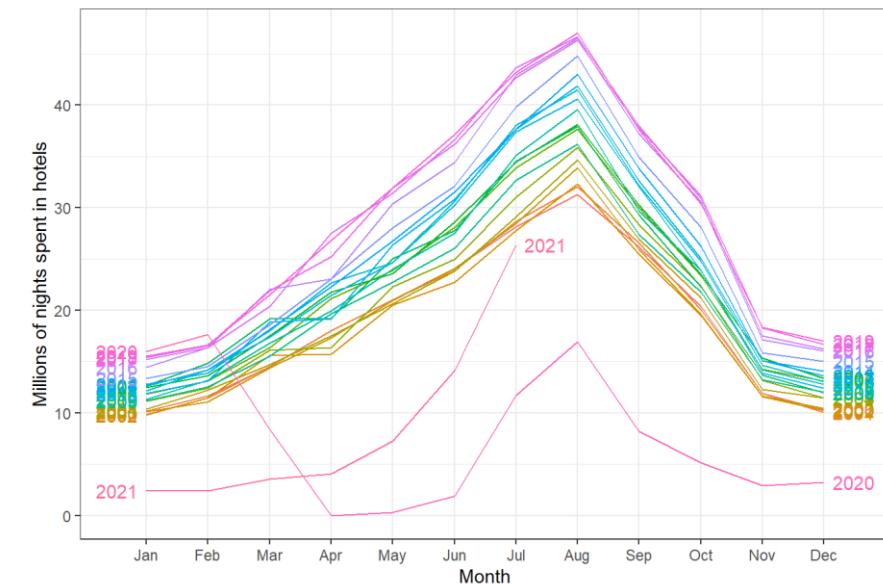
Tourism seasonality

Hotel stays in the Mediterranean, 2000-present



Tourism seasonality

Seasonal plot: Hotel stays in Spain



Seasonal plot: Hotel stays in Greece

