

Effective Data Visualization

How to design impactful and aesthetically pleasing charts

Dr. Cédric Scherer

bespokeDS | September 28th 2020

Photo by Richard Strozynski

Scientist by

→ population and community dynamics, movement ecology, wildlife diseases

- *Ph.D. in Ecology*
- *PostDoc in Computational Ecology*
@ Leibniz Insititute for Zoo and Wildlife Research (IZW)

DataViz Designer by

→ analyst, designer, consultant, lecturer

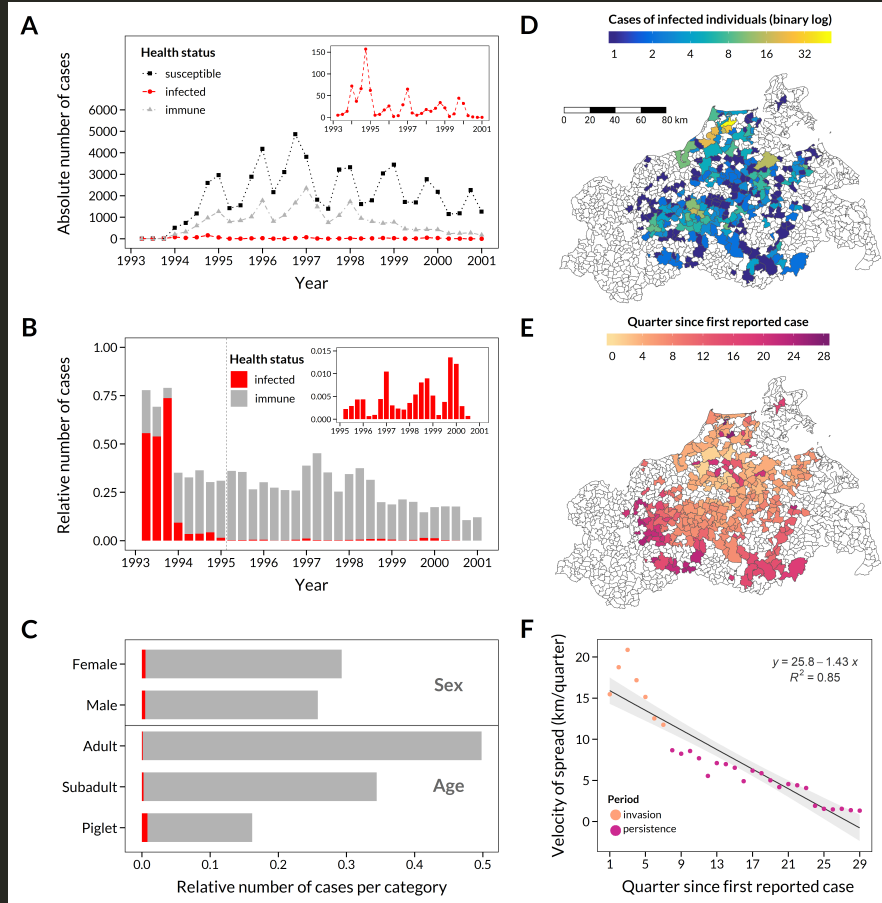
- *Freelancing Data Visualization Specialist*
- *PostDoc in Computational Ecology*
- *Data Challenges and Personal Projects*
#TidyTuesday, #30DayMapChallenge, #MakeoverMonday, #SWDchallenge



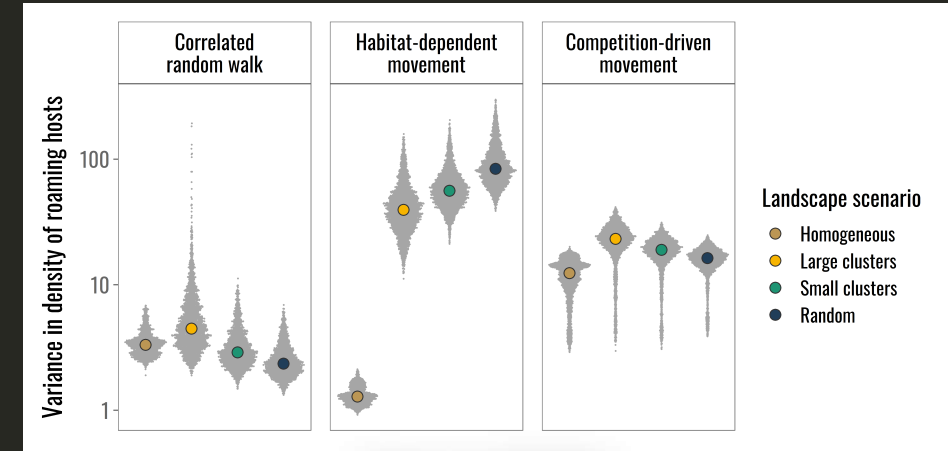
Scientist by 🎓 DataViz Designer by ♥



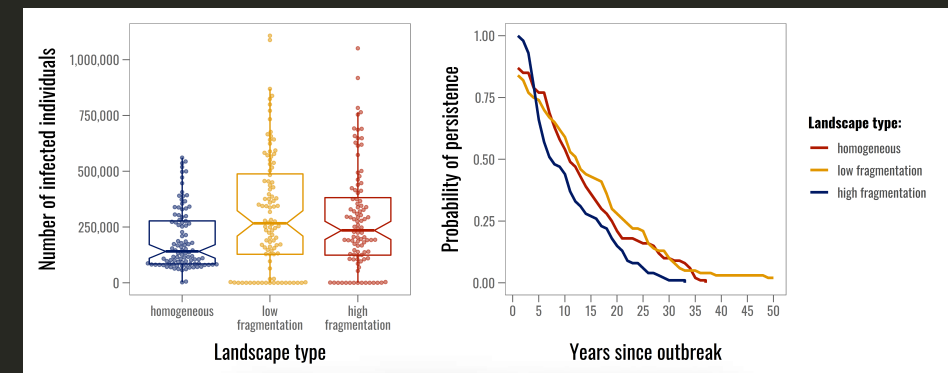
Data Visualizations for Scientific Publications & Talks



Scherer et al. 2019 Journal of Animal Ecology

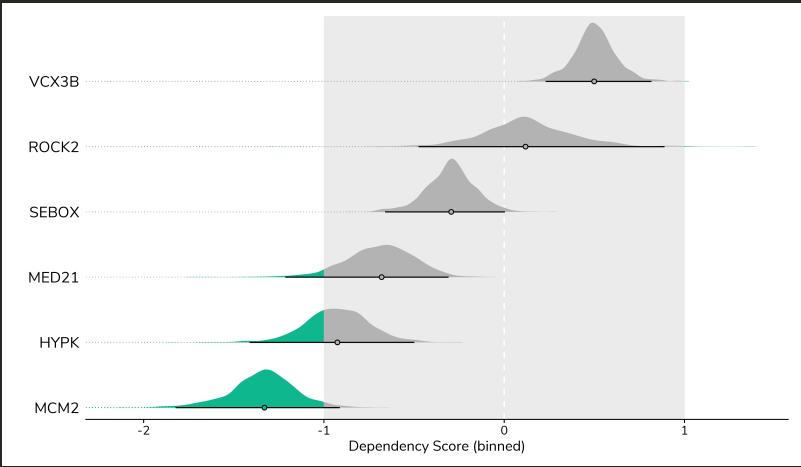
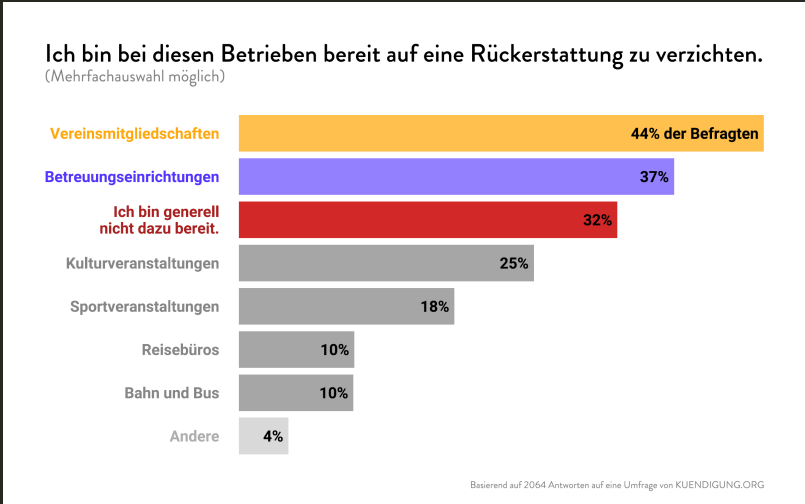
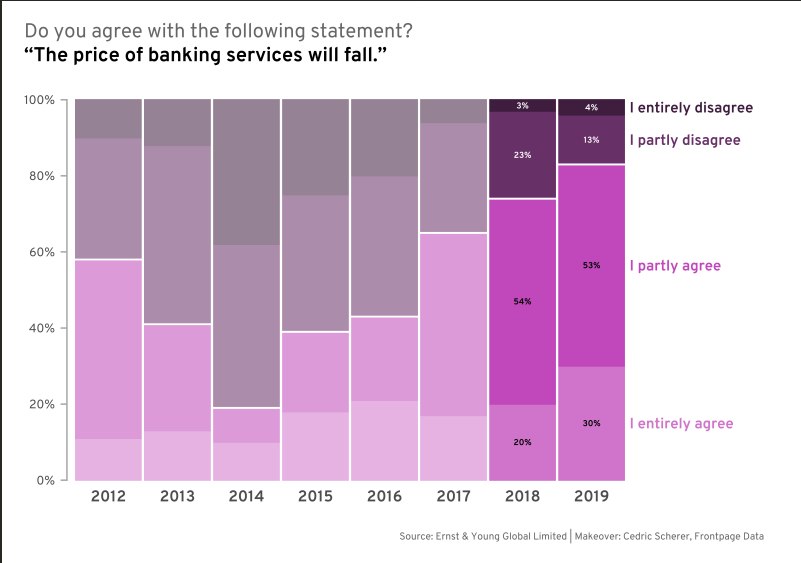
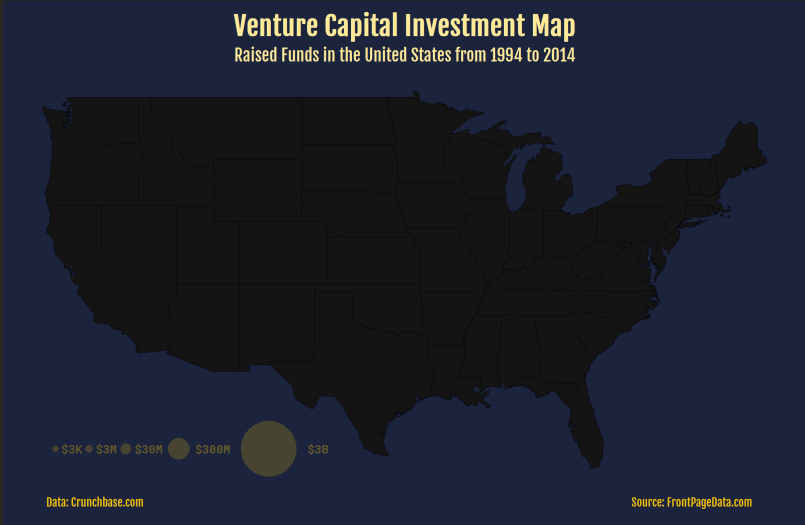


Scherer et al. 2020 Oikos

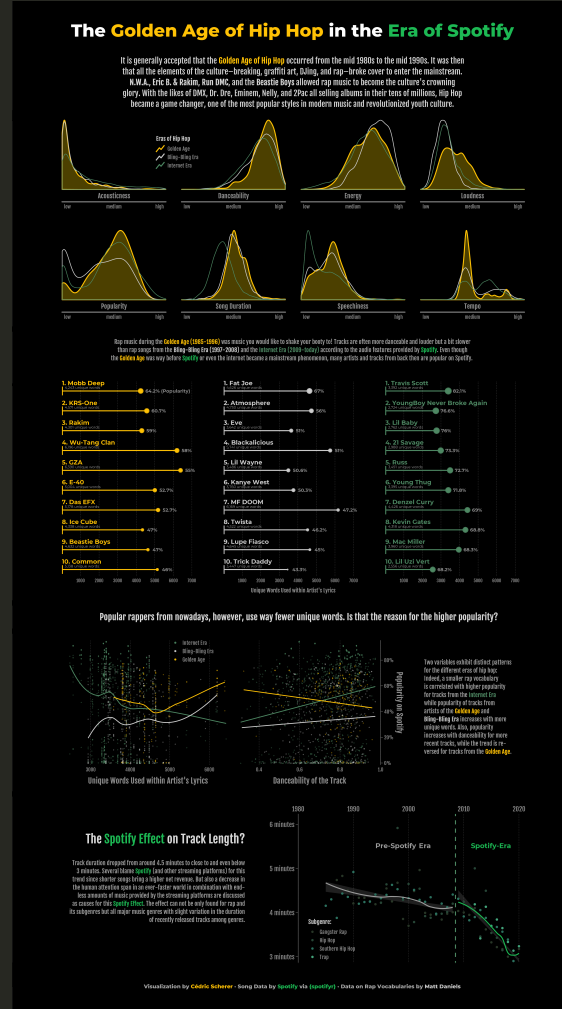


Sciaini et al. 2019 Methods in Ecology & Evolution

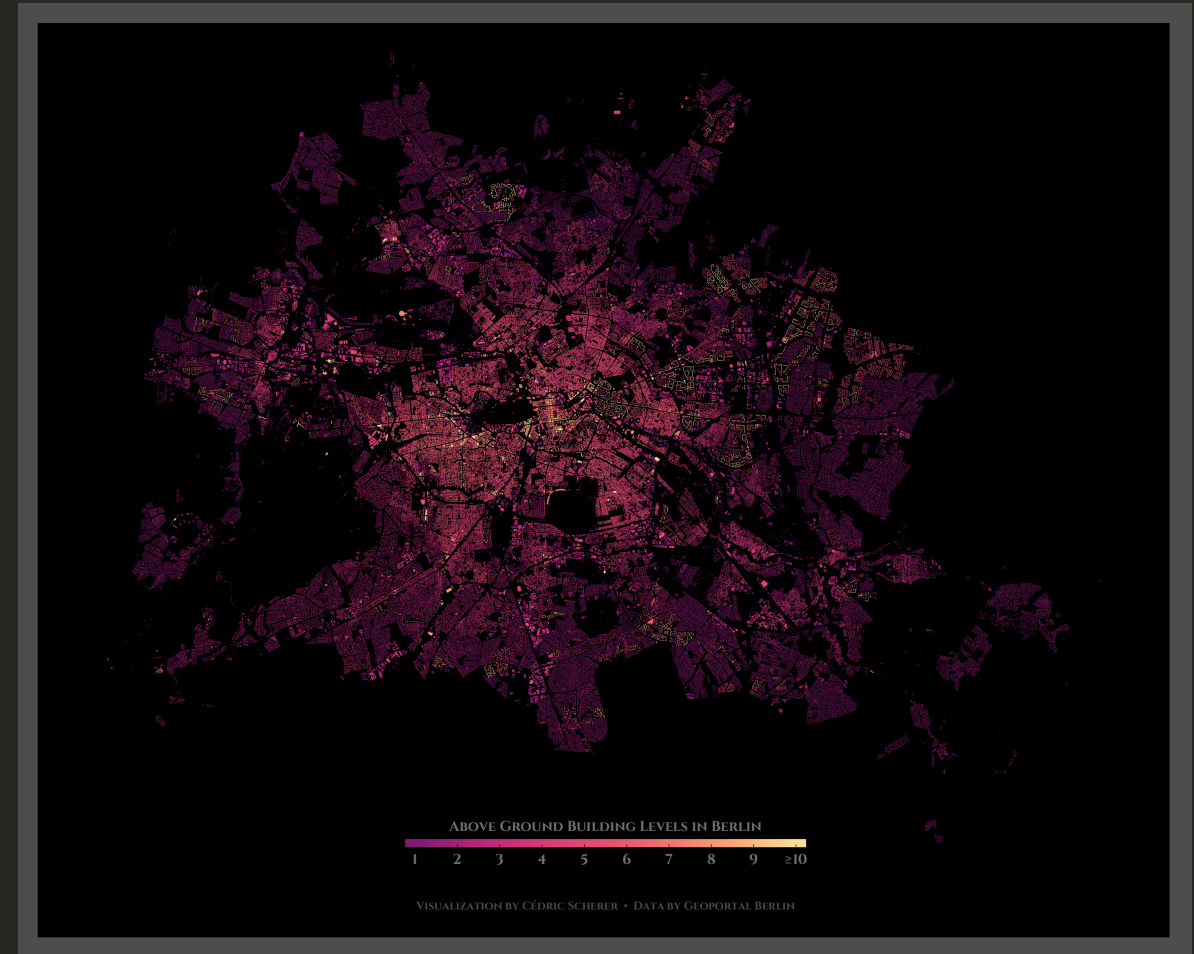
Data Visualizations for Client Projects



Data Visualizations as Challenge Contributions



Contribution to #TidyTuesday



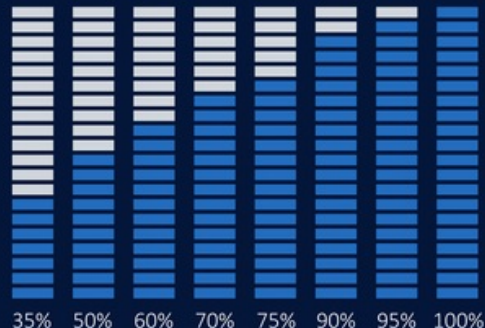
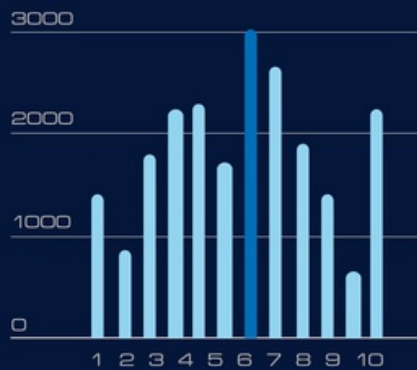
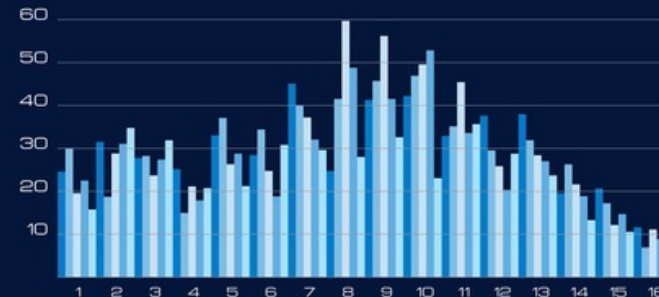
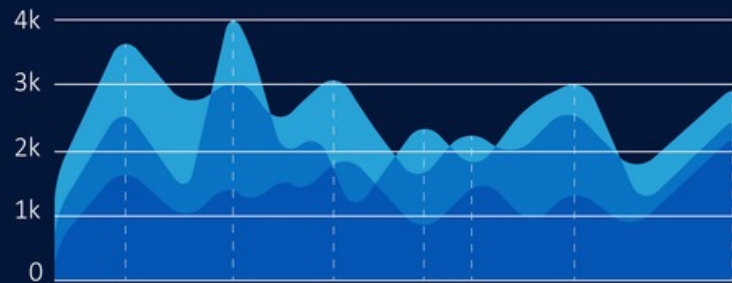
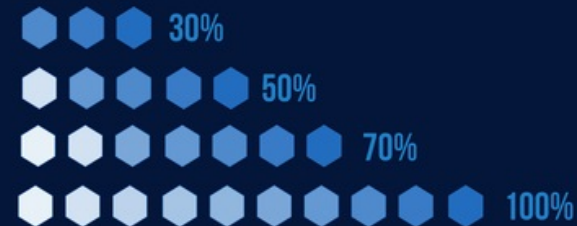
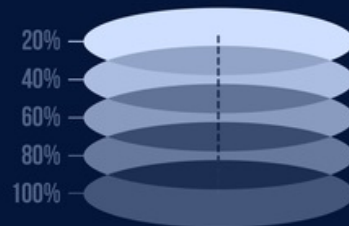
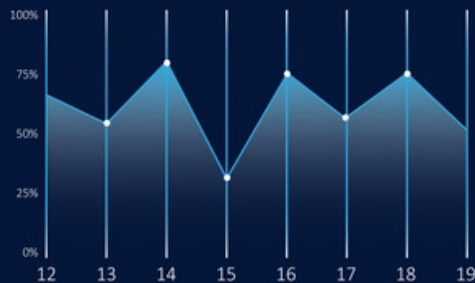
Contribution to #30DaymapChallenge

Data Visualization

is any graphical representation
of information and data.

Data Visualization
is part art and part science.

Claus O. Wilke, "Fundamentals of Data Visualization"



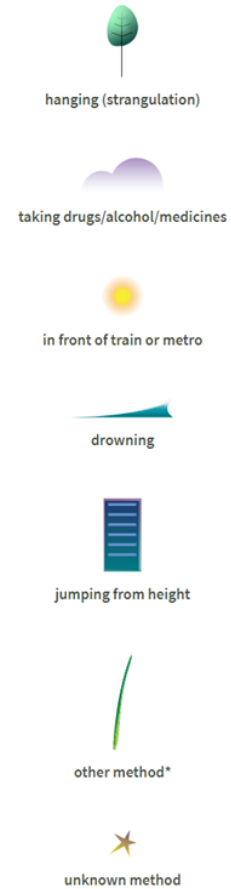
A View on Despair by Sonja Kuijpers/STUDIO TERP





A View on Despair by Sonja Kuijpers/STUDIO TERP

Each element represents a person who committed suicide in the Netherlands in 2017.



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 largeurs 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 ont été en Russie, le noir ceux qui en sont sortis. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Ségur, de Fozensac, de Chambray et le journal inédit 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 Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

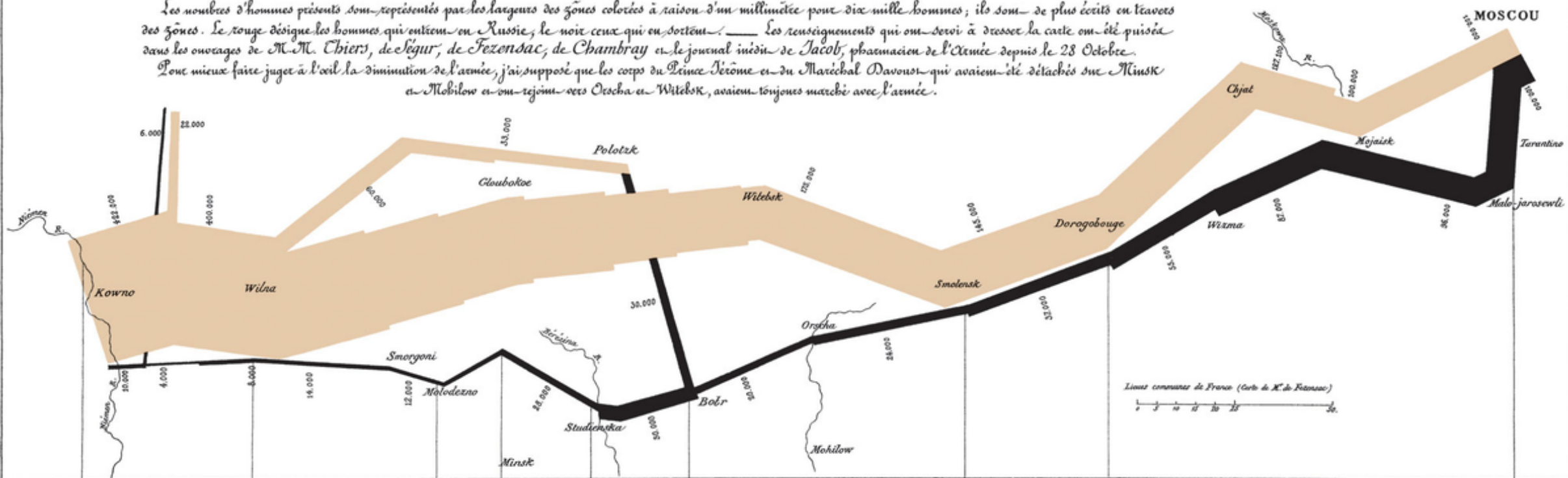
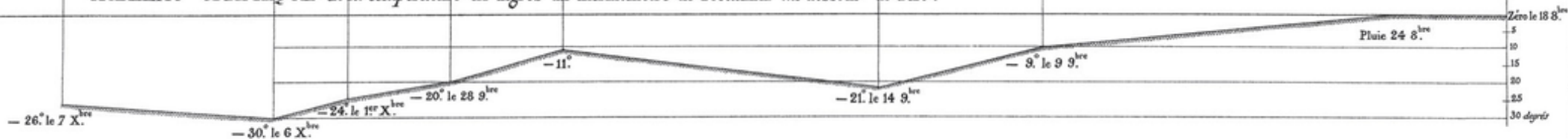
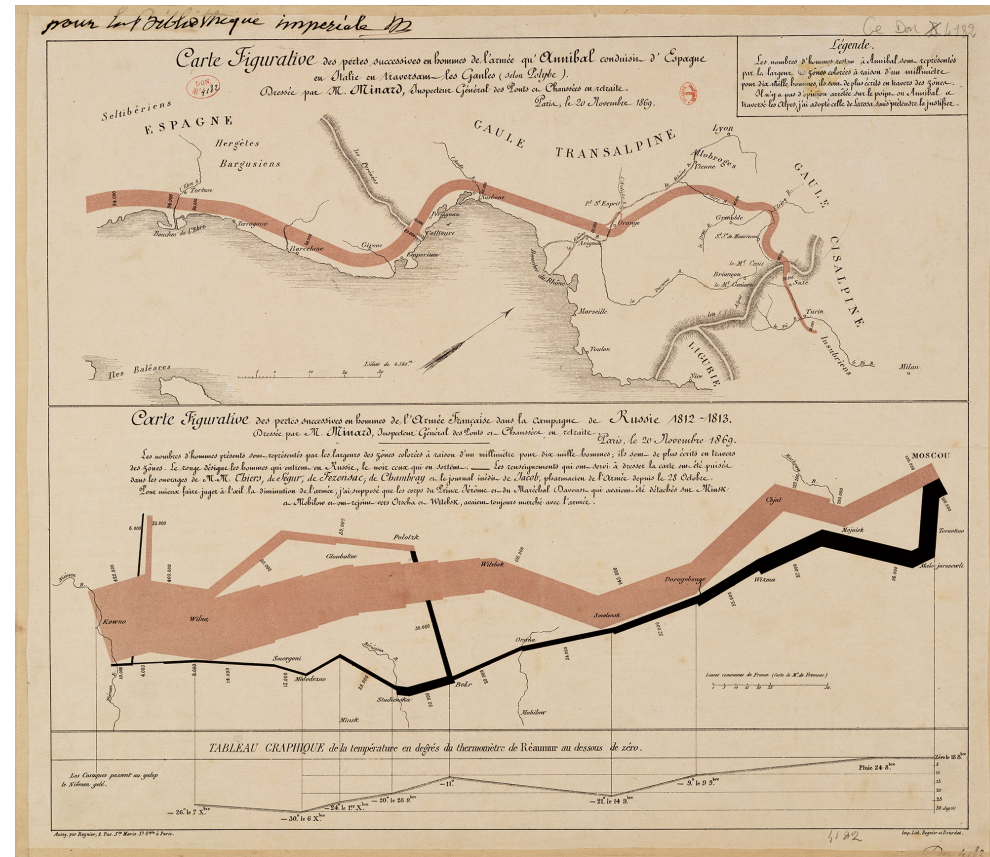


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.



Les Cosaques passent au galop le Niemen gelé.



Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 and

Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe) by Charles Joseph Minard

- displays the progress of the troops of Hannibal (218 BC) and Napoleon (1812-1813) in the form of a stream
- often considered as the best statistical graphic ever drawn

What Makes It a Bad Data Visualization

- Aesthetic problems (bad design)
- Substantive problems (bad data)
- Perceptual problems (bad perception)

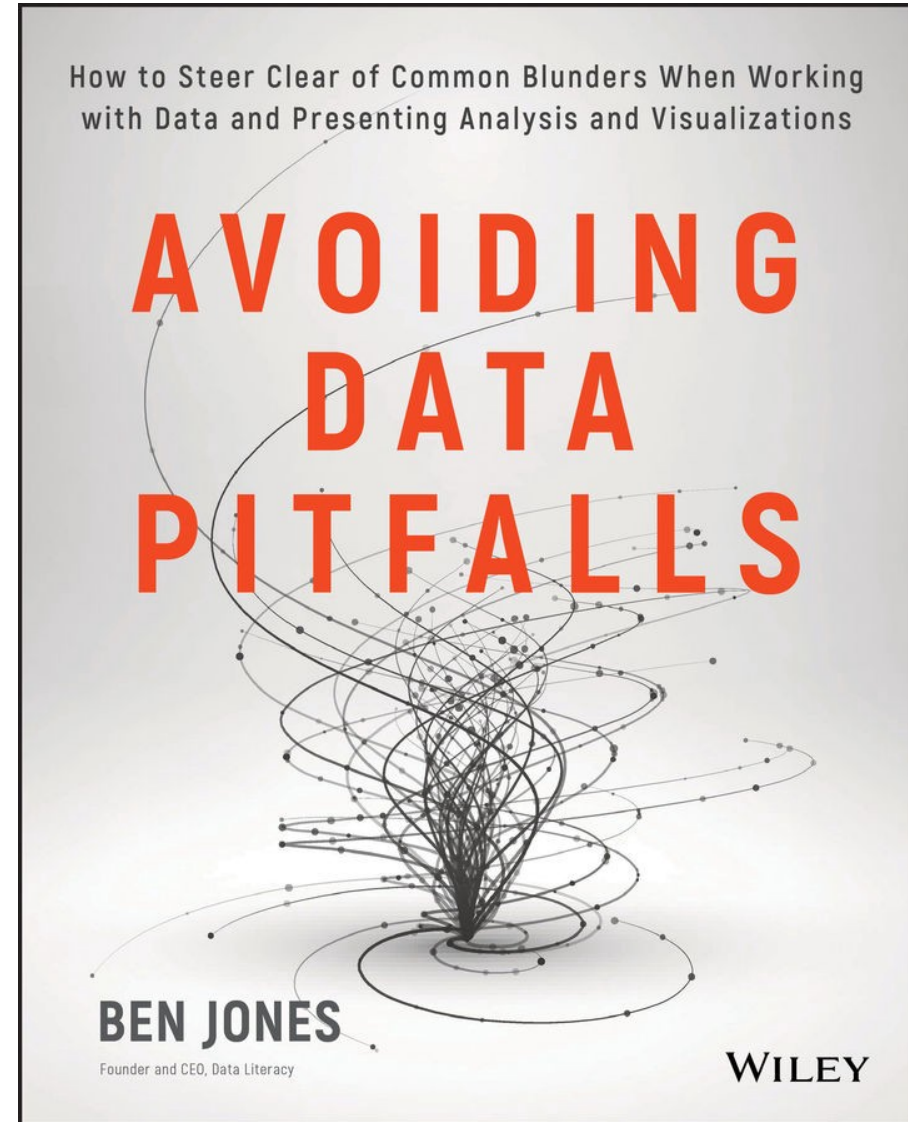
What Makes It a Good Data Visualization

- Information (Integrity)
 - Story (Interestingness)
 - Goal (Usefulness)
 - Visual Form (Beauty)
-



INFORMATION

Understand your data and be accurate

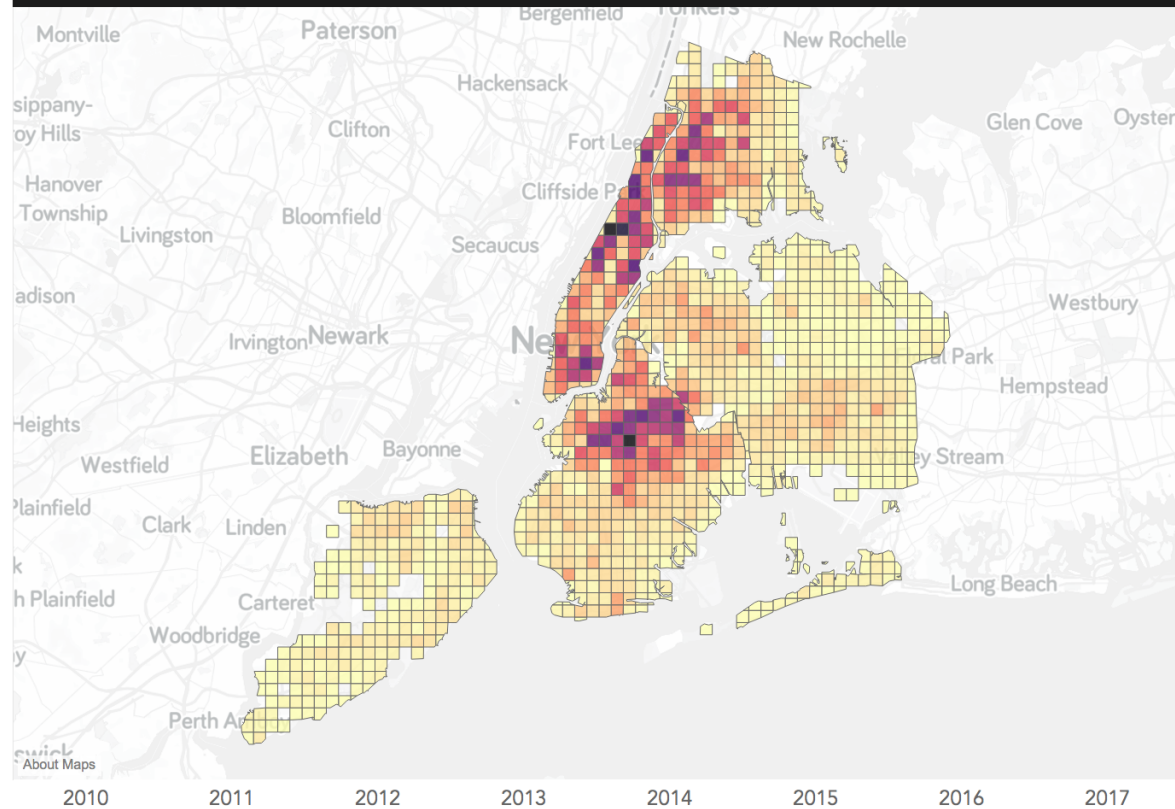


**Our data is never a perfect
reflection of the real world.**

Where are New York's rats?

Choose a Borough

(All)



DATA SOURCE: NYC Open Data | DESIGNED BY: Andy Kriebel @VizWizBI

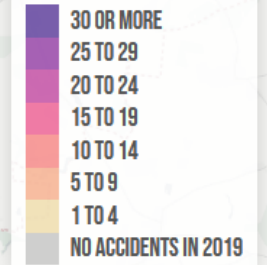
CHOOSE SET OF BIKE ACCIDENTS

☒ ALL BIKE ACCIDENTS

☐ ACCIDENTS ON ROADS

☐ ACCIDENTS ON BICYCLE INFRASTRUCTURE

NUMBER OF BIKE ACCIDENTS



**The best use of data is to
teach us what *isn't true*.**



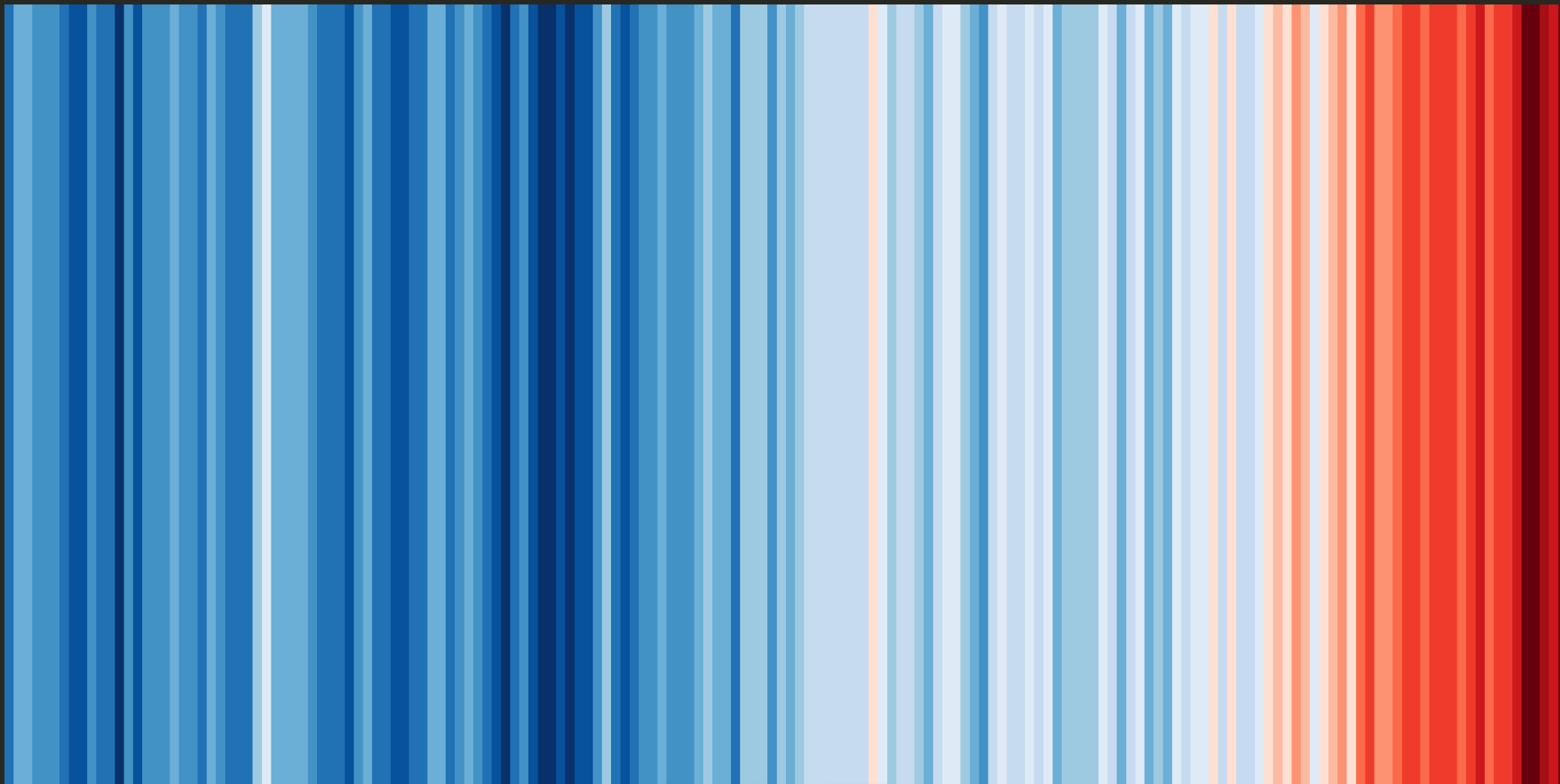
Source: inhomelandsecurity.com/risk-management-and-black-swan-events

STORY

Be clear about the message of your visualization

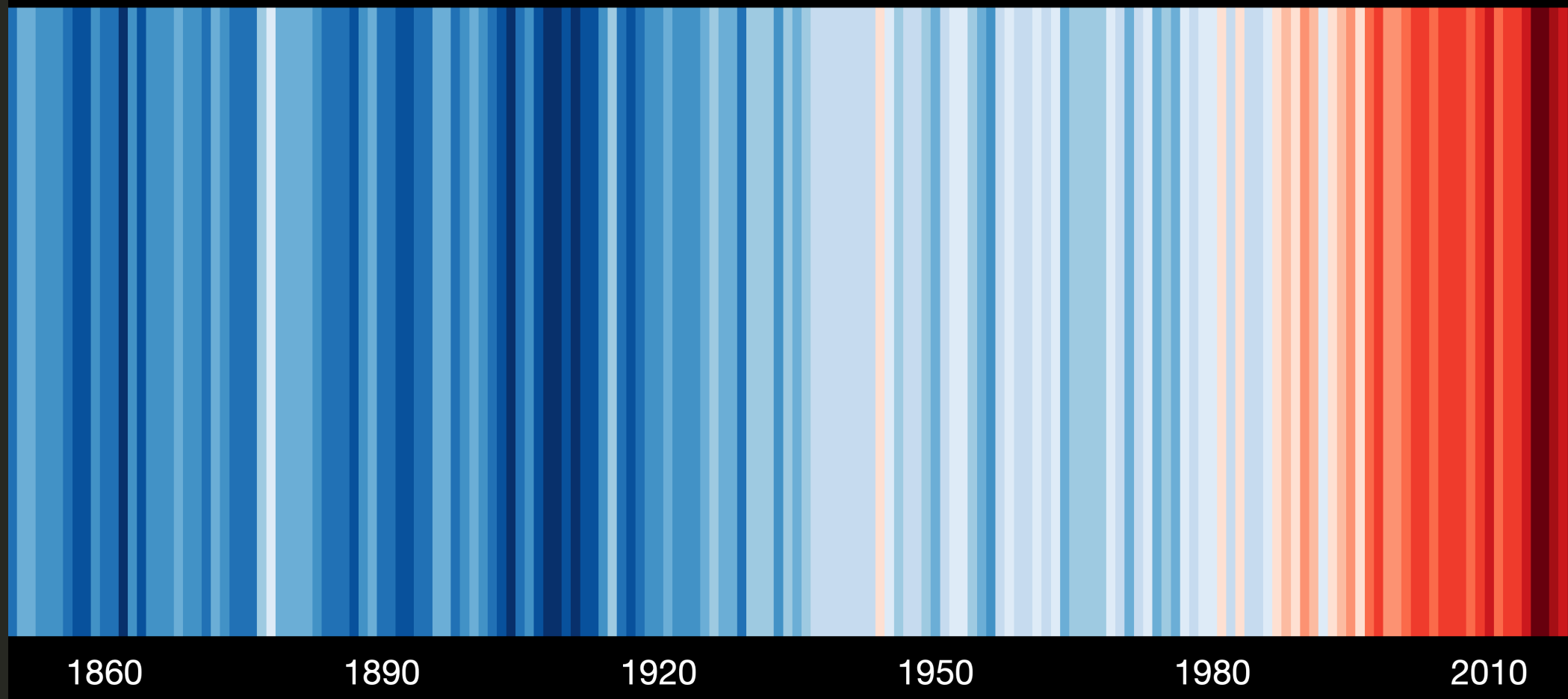
Who is my audience?

What am I trying to achieve with the visualization?

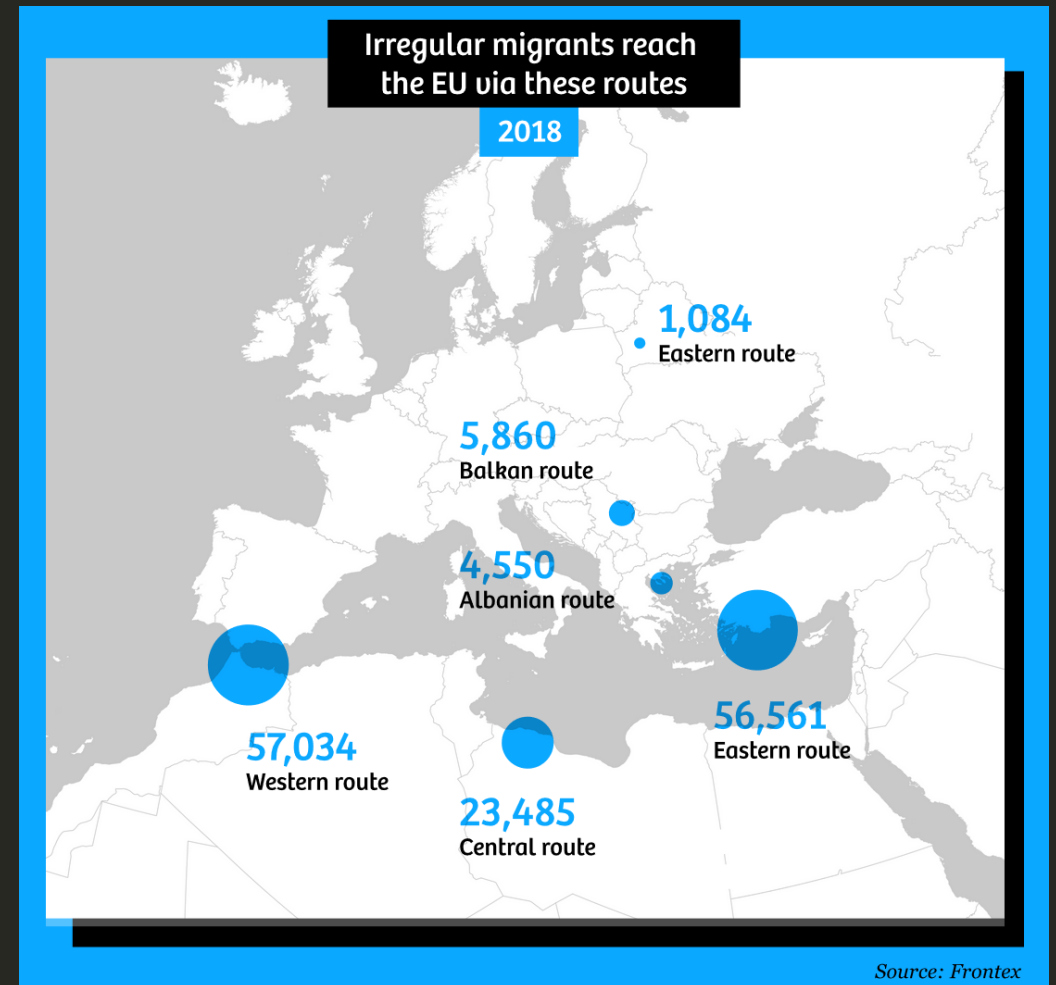
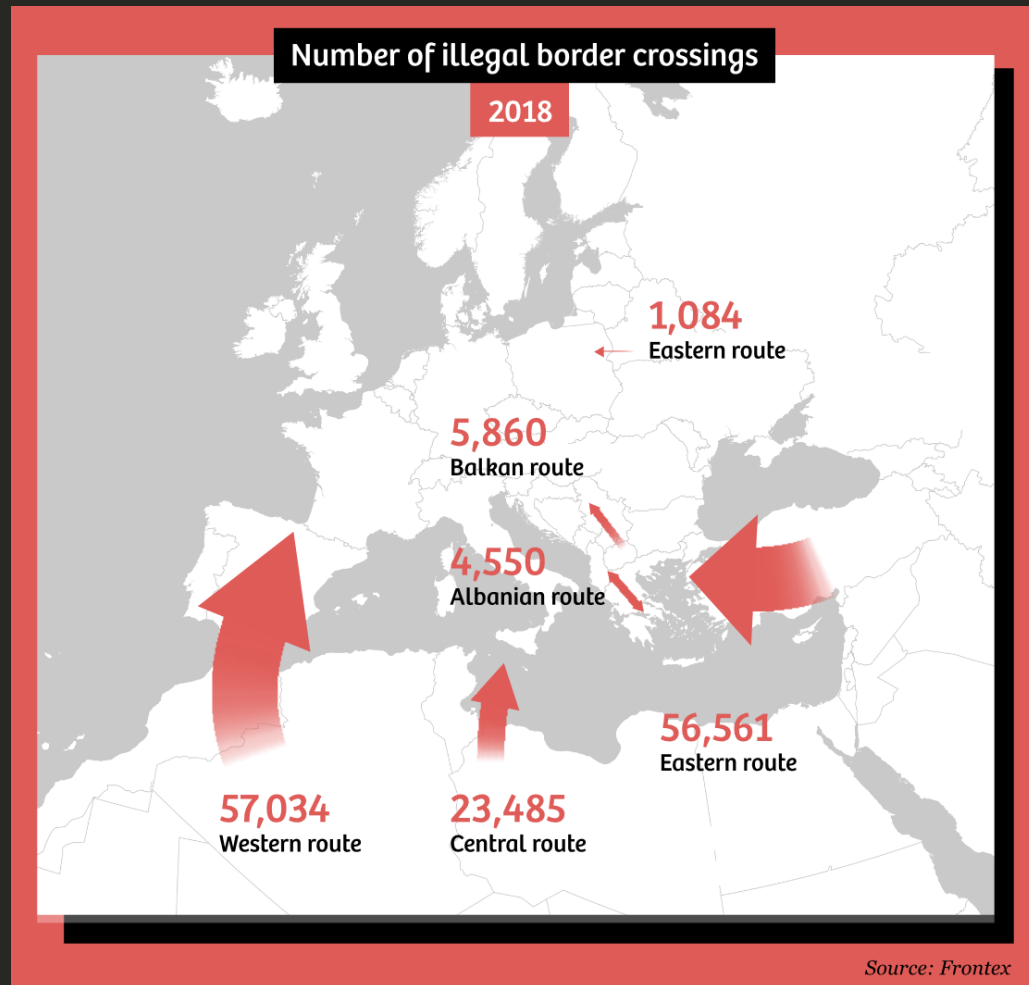


Warming Stripes by Ed Hawkins

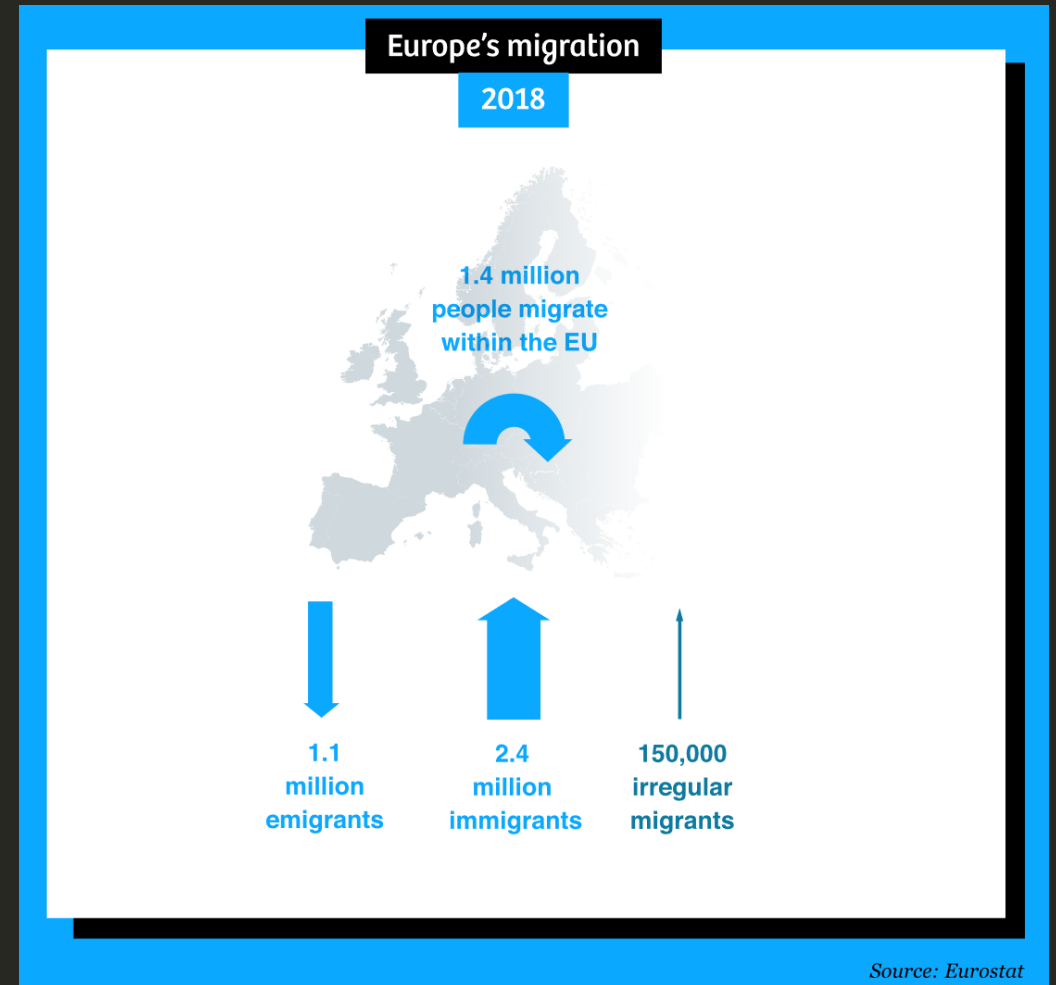
Global temperature change (1850-2019)



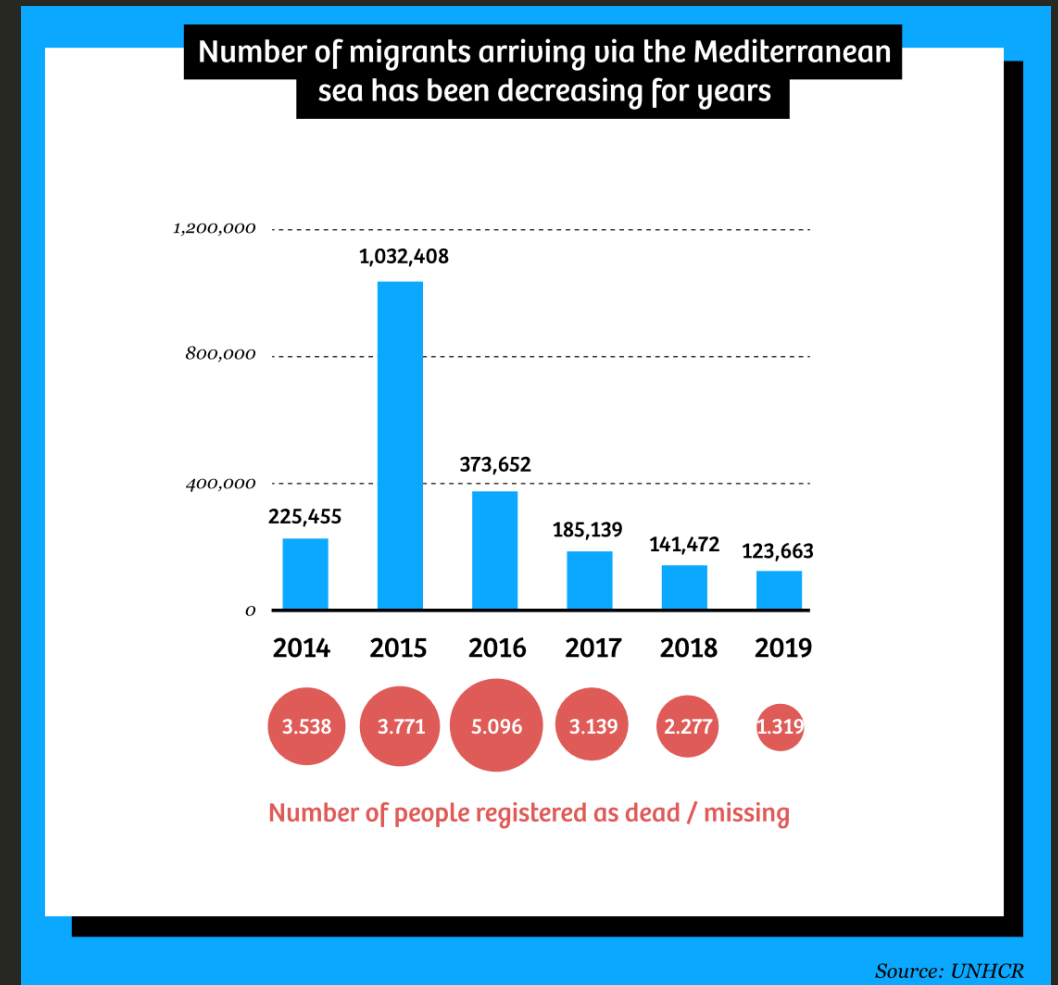
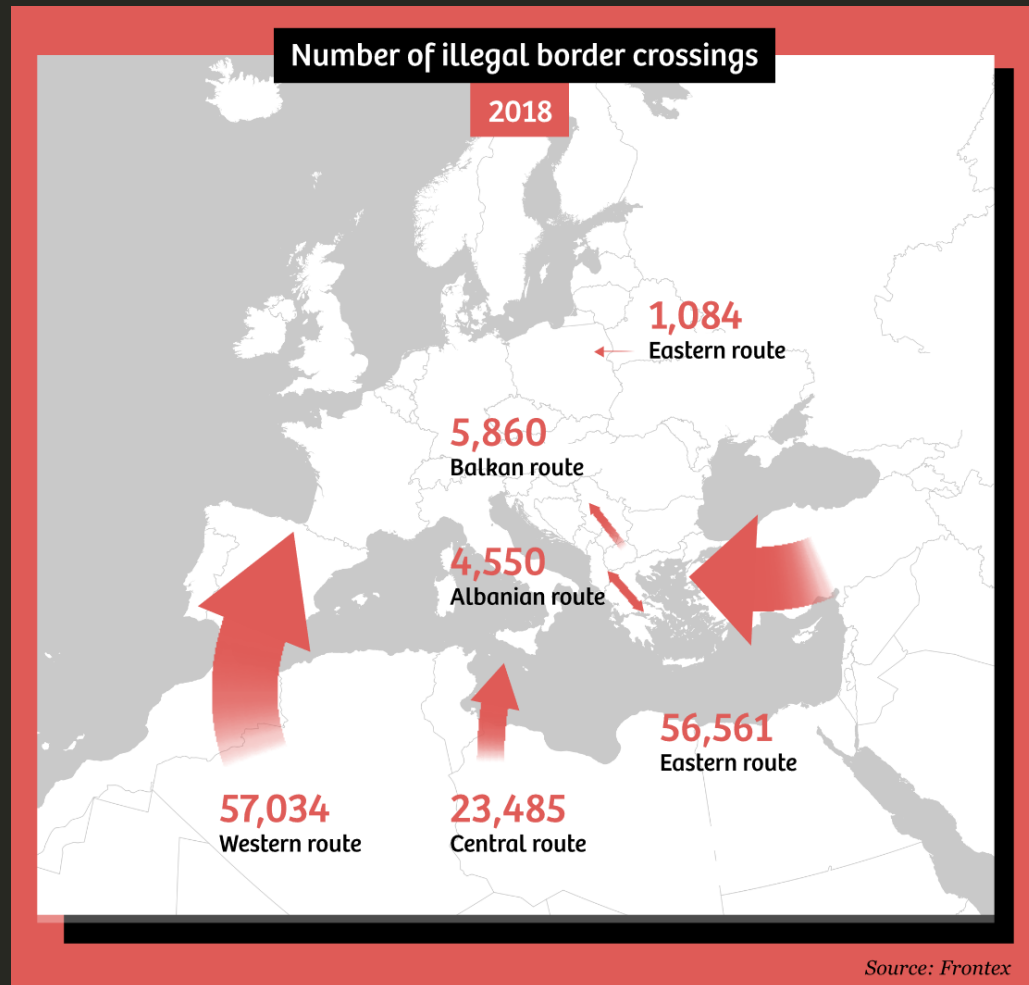
Warming Stripes by Ed Hawkins



How maps in the media make us more negative about migrants by Maite Vermeulen, Leon de Korte & Henk van Houtum



How maps in the media make us more negative about migrants by Maite Vermeulen, Leon de Korte & Henk van Houtum

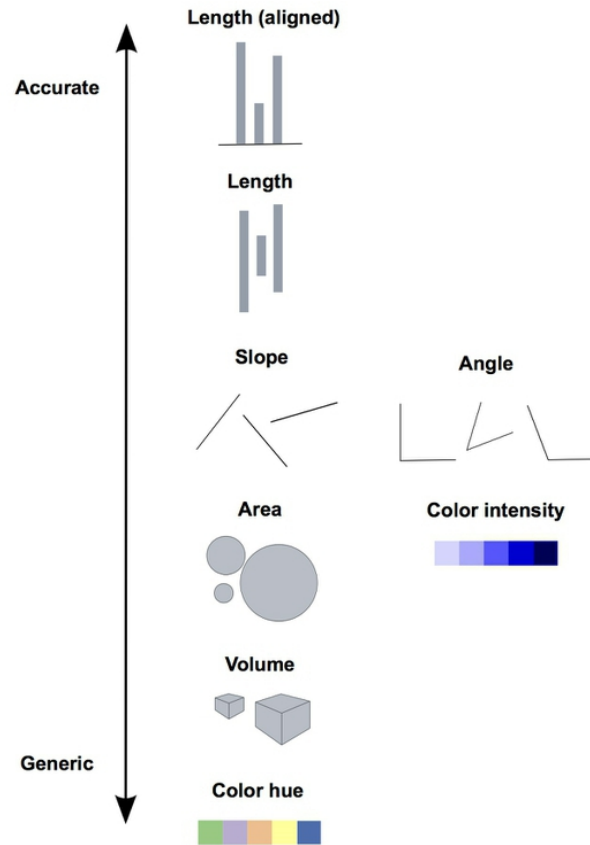


How maps in the media make us more negative about migrants by Maite Vermeulen, Leon de Korte & Henk van Houtum

GOAL

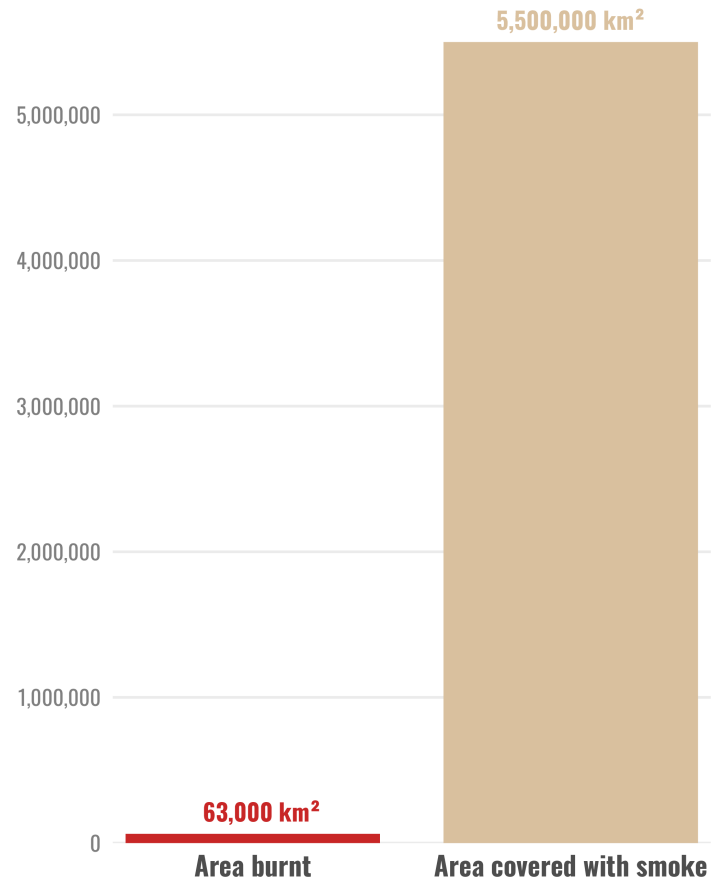
Select charts that successfully transport your story

Data visualizations map values into quantifiable features (aesthetics)

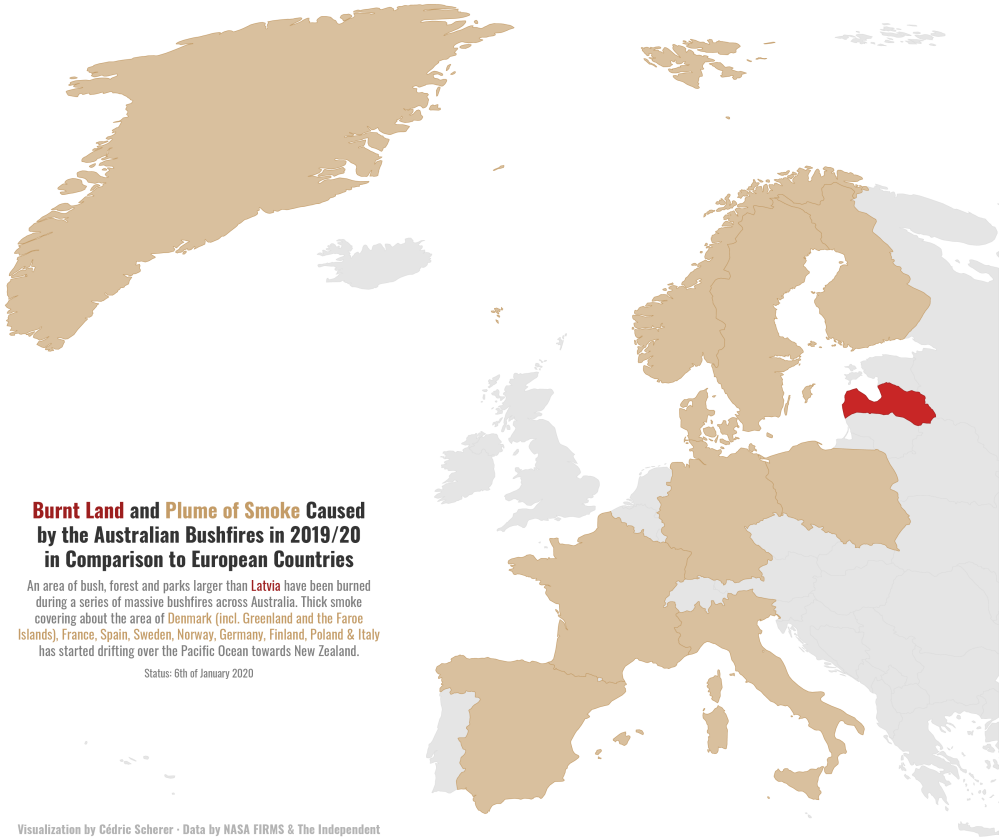
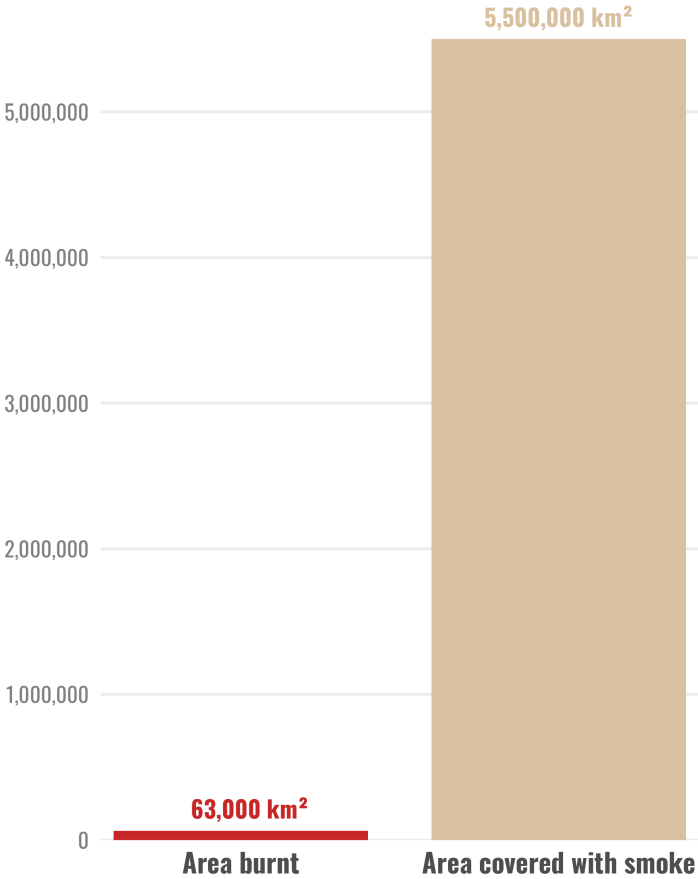


Peter Aldhous based on experiments by William Cleveland and Robert McGill

**Burnt land and plume of smoke caused
by the Australian bushfires in 2019/20**
(as of 6th of January 2020)

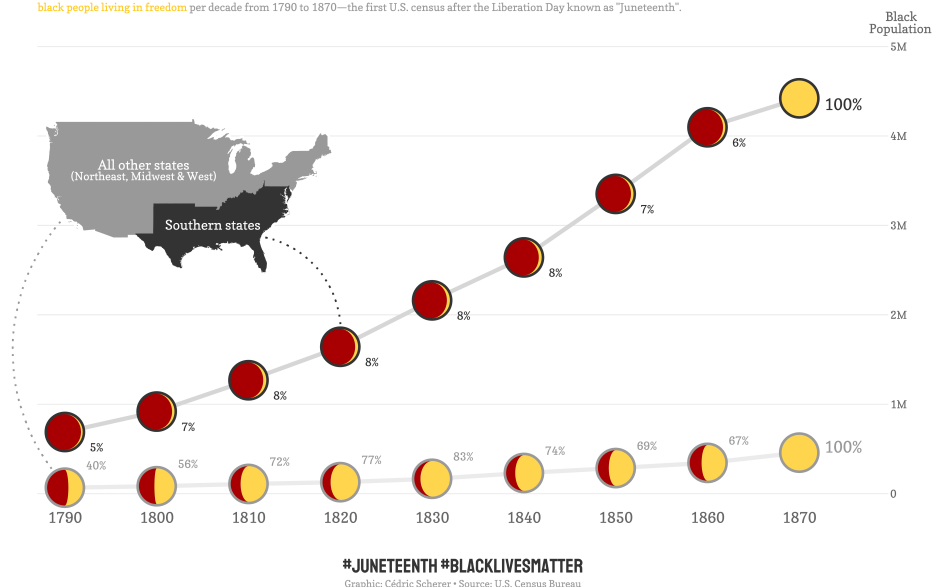


**Burnt land and plume of smoke caused
by the Australian bushfires in 2019/20**
(as of 6th of January 2020)



AN ECONOMY BUILT ON **SLAVERY**—A FUTURE BUILT ON **FREEDOM**?

By 1860, property owners in the south of North America began establishing plantation farms for cash crops like tobacco, cotton, and sugar cane—enterprises that required increasing amounts of labor. To meet the need, wealthy planters became slave traders and imported ever more individuals to the colonies, the vast majority from West Africa. While the "Emancipation Proclamation" was made law as of 1863, slave owners in the South, namely Texas, still maintained slavery until June 19th 1865 when Union soldiers were able to enforce the law abolishing slavery in the region. The graphic below shows the share of **black people in slavery**, most of them enslaved in the Southern states, and **black people living in freedom** per decade from 1790 to 1870—the first U.S. census after the Liberation Day known as "Juneteenth".

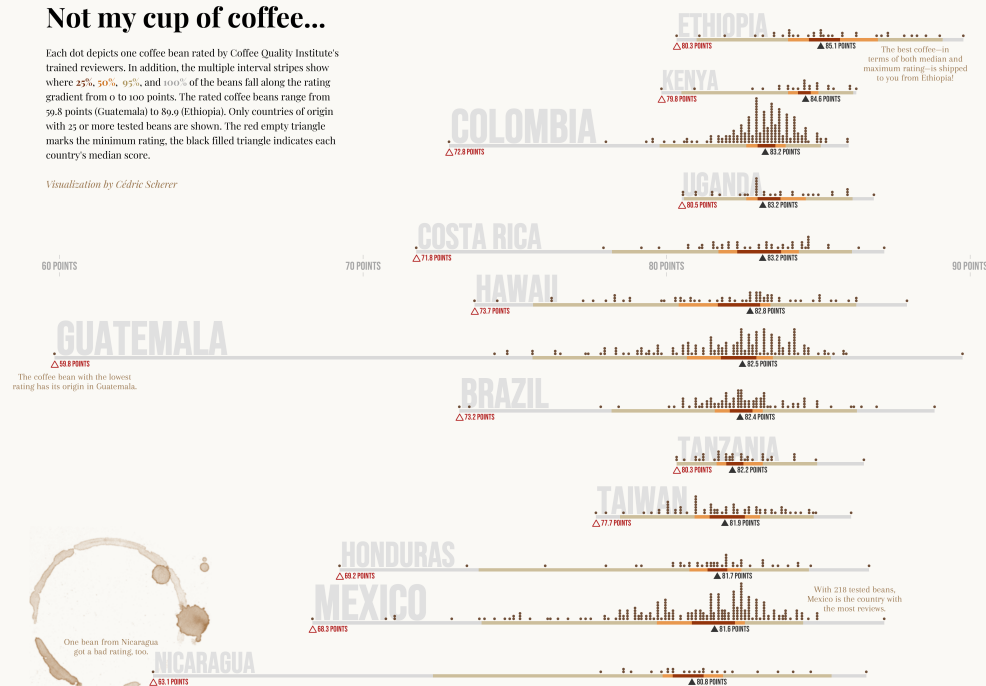


#TidyTuesday contribution Week 2020/25

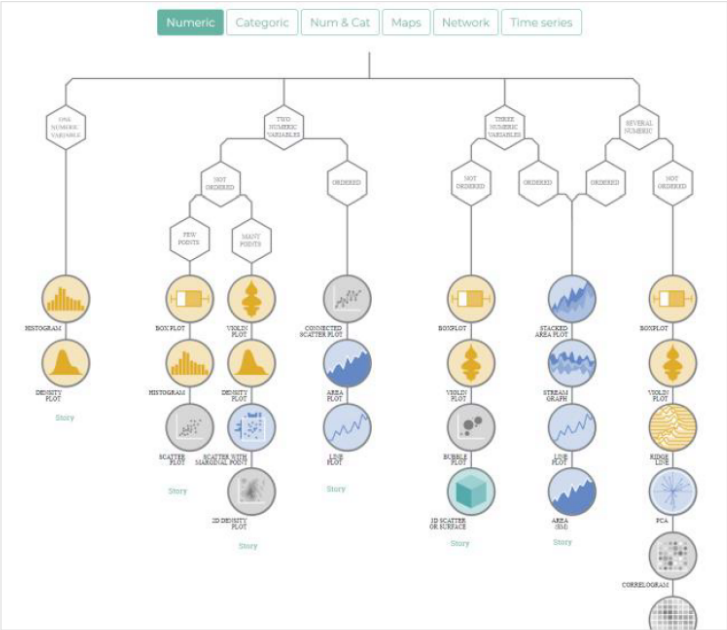
Not my cup of coffee...

Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points. The rated coffee beans range from 59.8 points (Guatemala) to 89.9 (Ethiopia). Only countries of origin with 25 or more tested beans are shown. The red empty triangle marks the minimum rating, the black filled triangle indicates each country's median score.

Visualization by Cédric Scherer



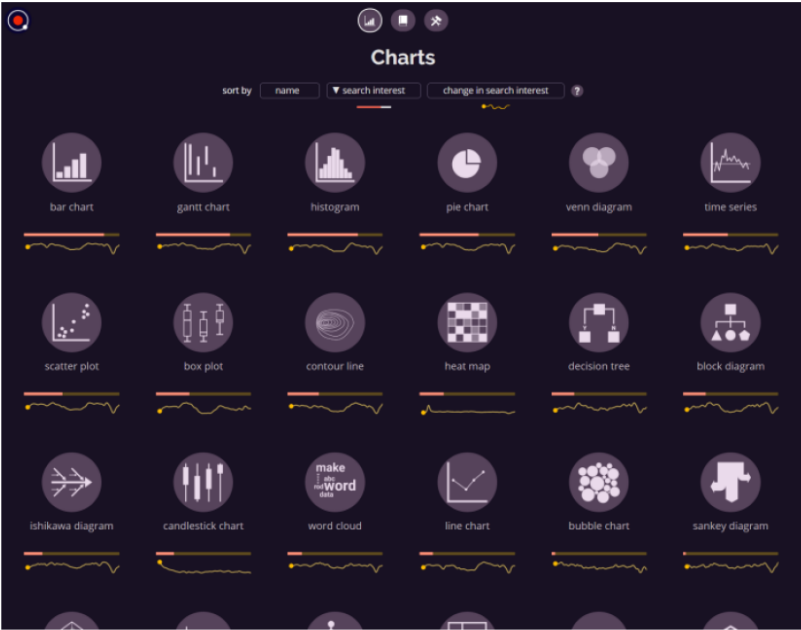
#TidyTuesday contribution Week 2020/28



data-to-viz.com



datavizproject.com



visualizationuniverse.com/charts

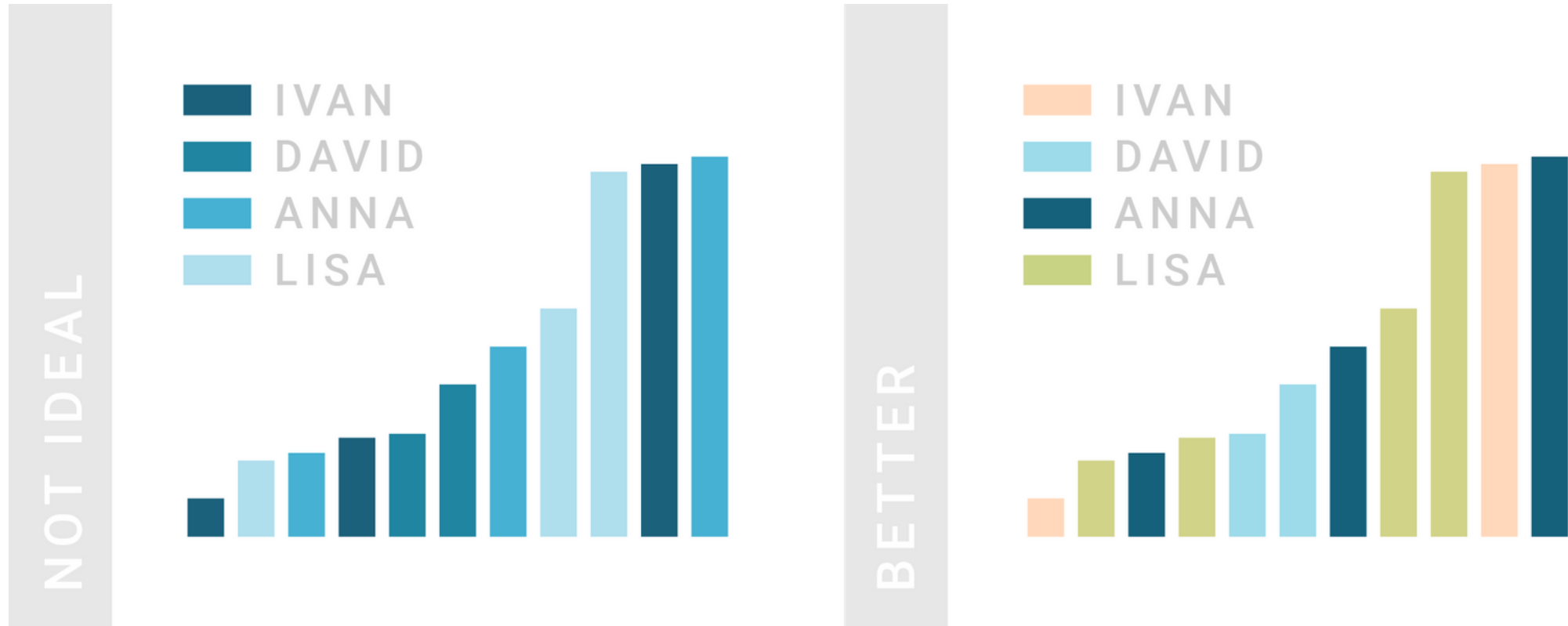
VISUAL FORM

Follow design rules and data visualization principles

Colors

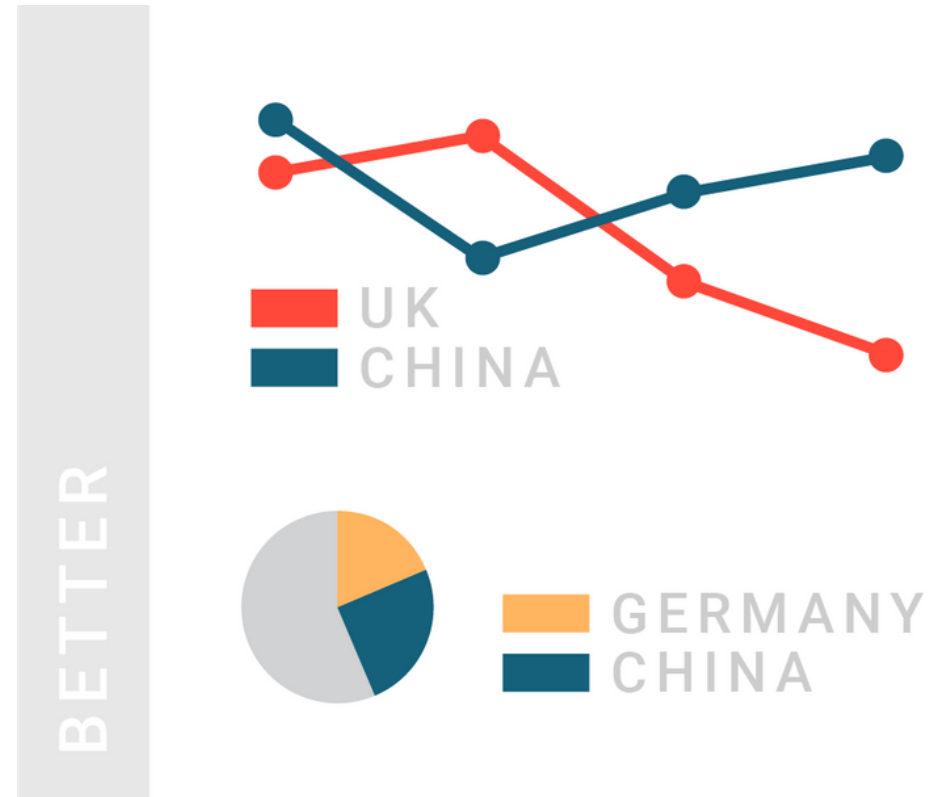
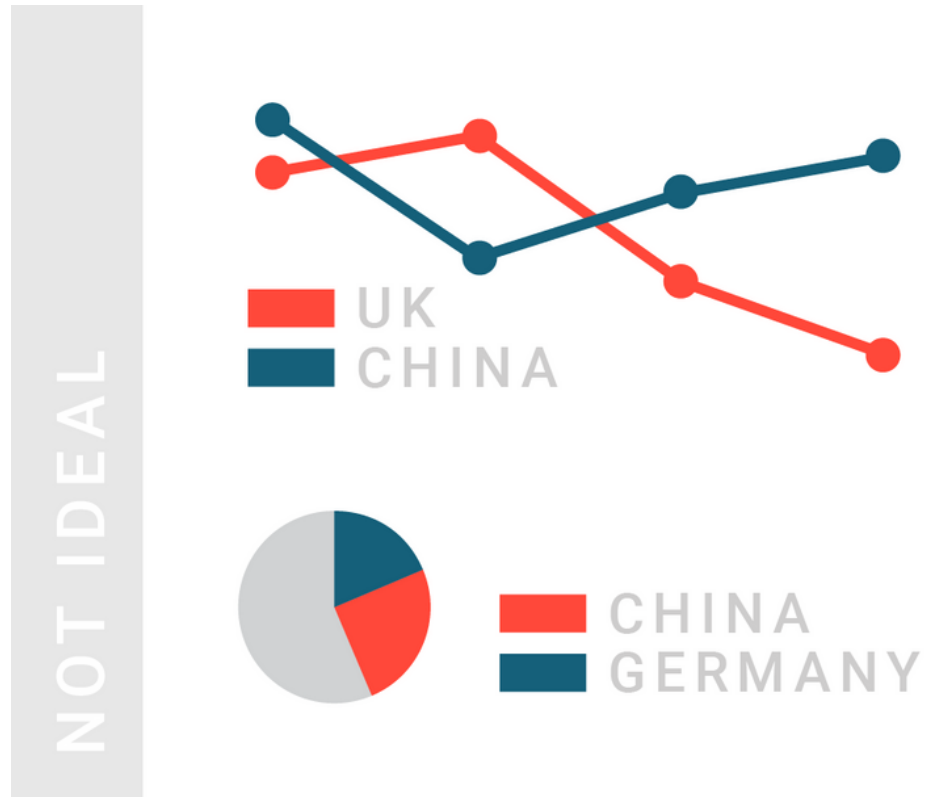
and Common Pitfalls

Color Choice & Accessibility



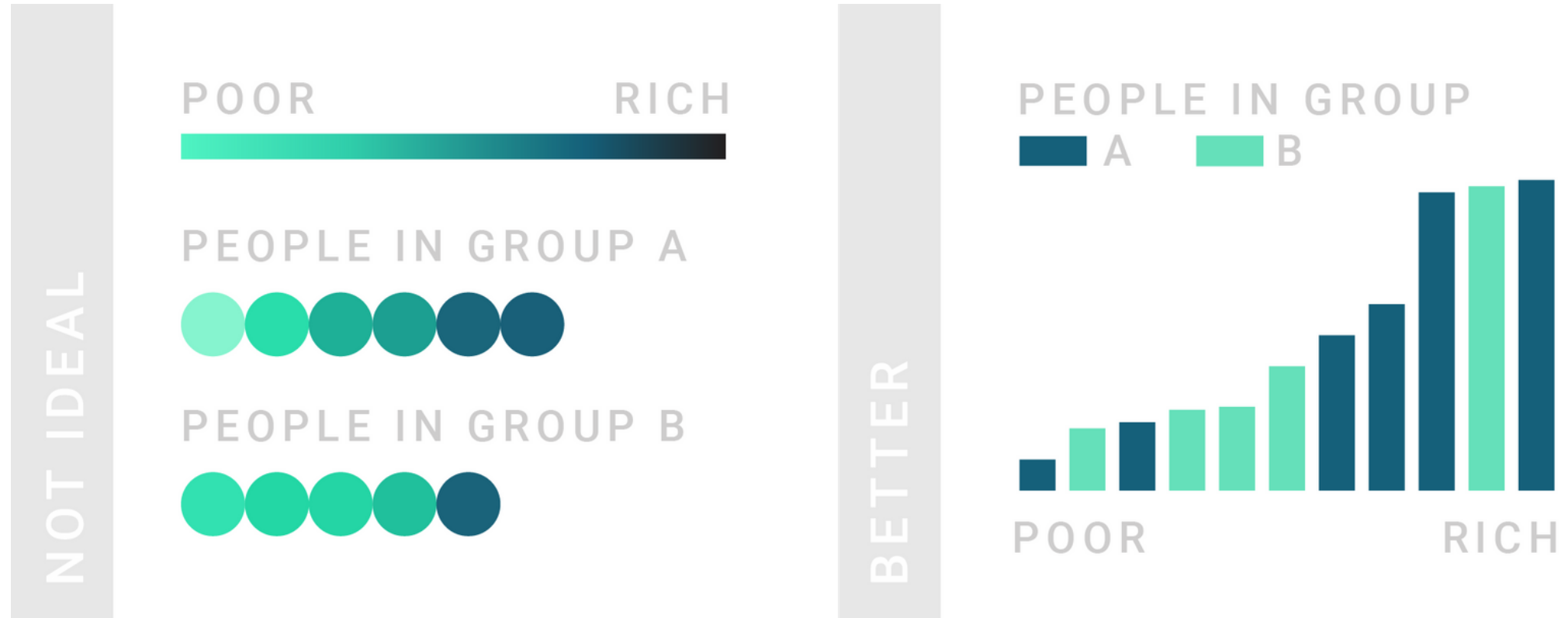
What to consider when choosing colors for data visualization by Lisa Charlotte Rost/DataWrapper

Color Choice & Accessibility



What to consider when choosing colors for data visualization by Lisa Charlotte Rost/DataWrapper

Color Choice & Accessibility



What to consider when choosing colors for data visualization by Lisa Charlotte Rost/DataWrapper

Journals & Magazines > IEEE Computer Graphics and Ap... > Volume: 27 Issue: 2 ?

Rainbow Color Map (Still) Considered Harmful

Publisher: IEEE

2 Author(s)

David Borland ; Russell M. Taylor II [View All Authors](#)

172
Paper
Citations

3
Patent
Citations

9091
Full
Text Views



[Med Phys.](#) 2015 Jun; 42(6): 2942–2954.

PMCID: PMC5148121

Published online 2015 May 20. doi: [10.1118/1.4921125](#)

PMID: [26127048](#)

Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images

[Silvina Zabala-Travers](#), [Mina Choi](#), [Wei-Chung Cheng](#), and [Aldo Badano](#)^{a)}

10 March 2017

Interpretation of the rainbow color scale for quantitative medical imaging: perceptually linear color calibration (CSDF) versus DICOM GSDF

[Frédérique Chesterman](#); [Hannah Manssens](#); [Céline Morel](#); [Guillaume Serrell](#); [Bastian Piepers](#); [Tom Kimpe](#)

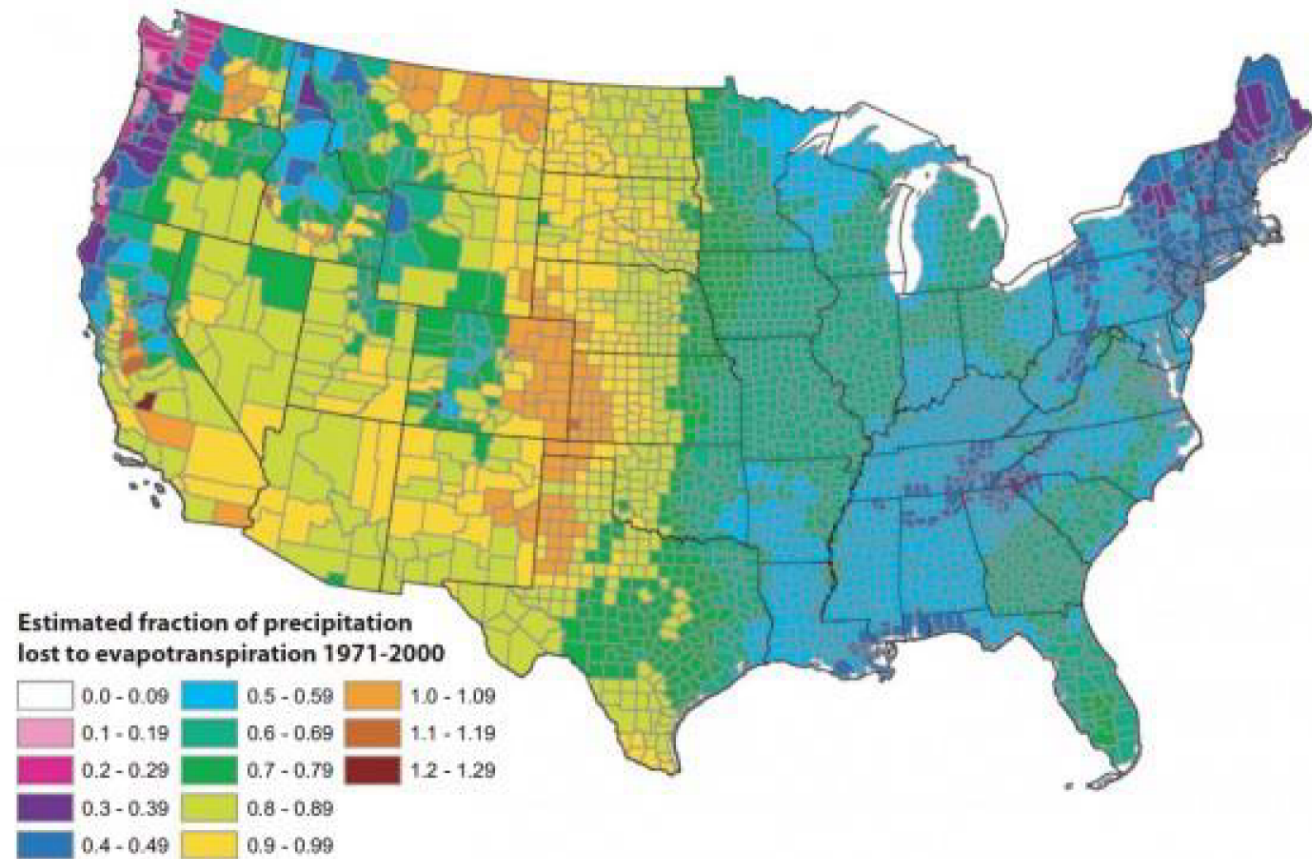


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Source: eagereyes.org/basicss/rainbow-color-map

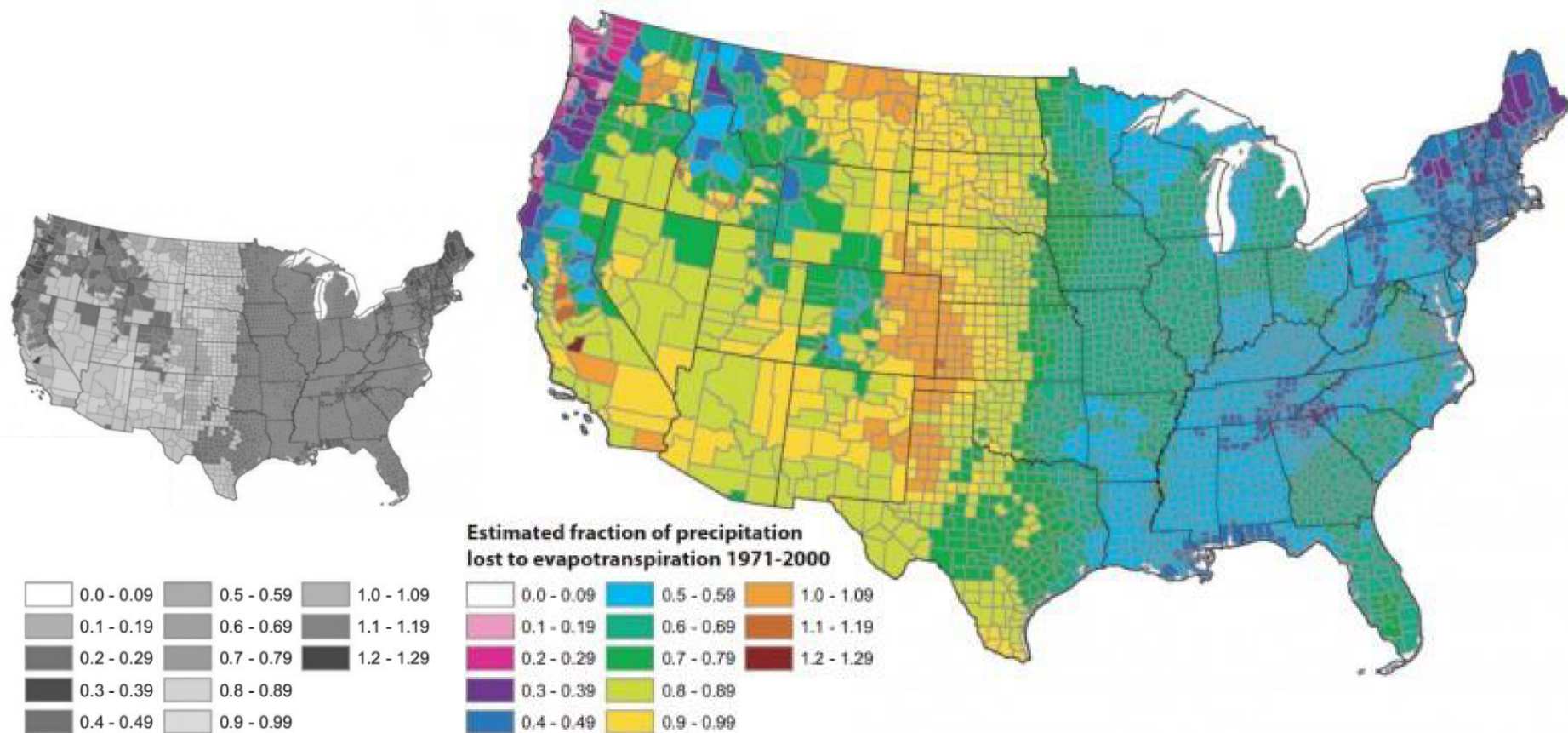


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Modified from eagereyes.org/basicss/rainbow-color-map

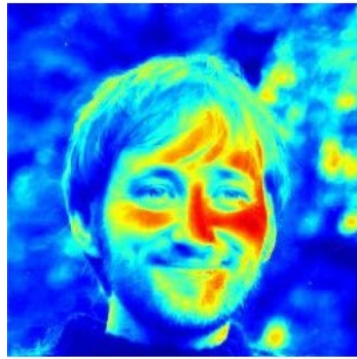


true-colour Phil

Source: fabiocrameri.ch/batlow



true-colour Phil



rainbow Phil
is distorted



batlow Phil
is flawless

Source: fabiocrameri.ch/batlow

Color Choice & Accessibility

Choose color-blind friendly palettes:

projects.susielu.com/viz-palette

Test your final visualization:

color-blindness.com/coblis-color-blindness-simulator

Create a CVD-version of your ggplot in R:

github.com/clauswilke/colorblindr

Color Choice & Accessibility

VIZ PALETTE

By: [Elijah Meeks](#) & [Susie Lu](#)

PICK

Use Chroma.js

Use Colorgical

Use ColorBrewer

Add

Replace

EDIT

7 Colors

Add

☒ #hex ☐ rgb ☐ hsl

1 ● #ffd700

2 ● #ffb14e

3 ● #fa8775

4 ● #ea5f94

5 ● #cd34b5

6 ● #9d02d7

7 ● #0000ff

×

×

×

×

×

×

×

GET

☒ #hex ☐ rgb ☐ hsl

☒ String quotes ☐ Object with metadata

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

COLORS IN ACTION

Background color: #ffffff

Font color: ● #000000

Charts made with [Semiotic](#)

Color Population: No Color Deficiency - 96% Deuteranomaly - 2.7% Protanomaly - 0.66% Protanopia - 0.59% Deuteranopia - 0.56% Greyscale

Sample font Randomize Data Stroke: Dark None

projects.susielu.com/viz-palette

The Choice of the Font(s)

- The font(s) should fit the topic and audience - context matters.
 - Avoid fancy fonts and squiggeled letters.
 - Use ways to visualize hierarchy.
 - Avoid using ALL CAPS.
 - Use a monospaced font with lining for numbers.
-

The Choice of the Font(s)

- The font(s) should fit the topic and audience - context matters.
 - Avoid fancy fonts and squiggeled letters.
 - Use ways to visualize hierarchy.
 - Avoid using ALL CAPS.
 - Use a monospaced font with lining for numbers.
 - Consistency is key!
-

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too.

Oh, hi there. Thanks for reading me...
Yeah, I know I am kinda boring. Sorry.

Keep it Simple

Using lots of fonts
can make for a design
that is cluttered,
overcomplicated,
AND JUST NOT VERY NICE

*But if you just use
a small selection,*
you can keep your
design cleaner, clearer
and just much easier
to digest

The 1Il Test

1Il Calibri

1Il Bitter

1Il Open Sans

1Il Monda

1Il Roboto

1Il Chivo

1Il Avenir Next Condensed

1Il Fira Sans

1Il Lato

1Il Noto Sans

1Il Oswald

1Il Bahnschrift

Tabular (Monospaced) Numbers

32154

Montserrat
proportional numbers

32154

Open Sans
tabular numbers

32154

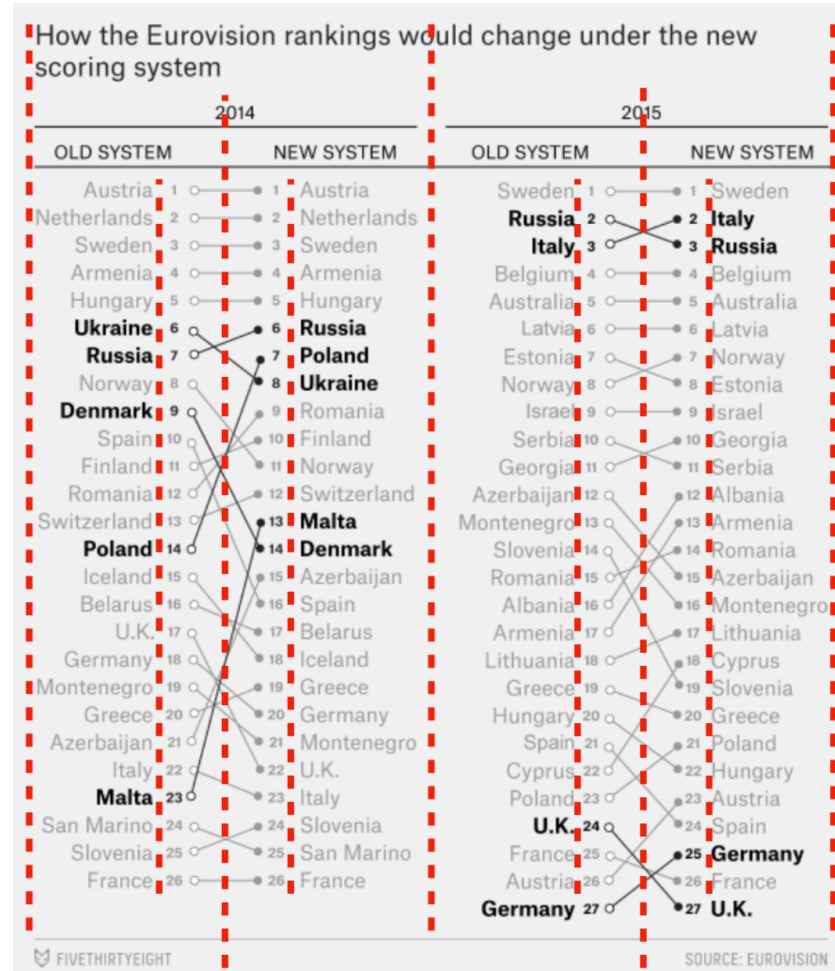
Lato
tabular numbers

So

urce: *Ti*Choosing Fonts for Your Data Visualization by ffany France on viaedium.com

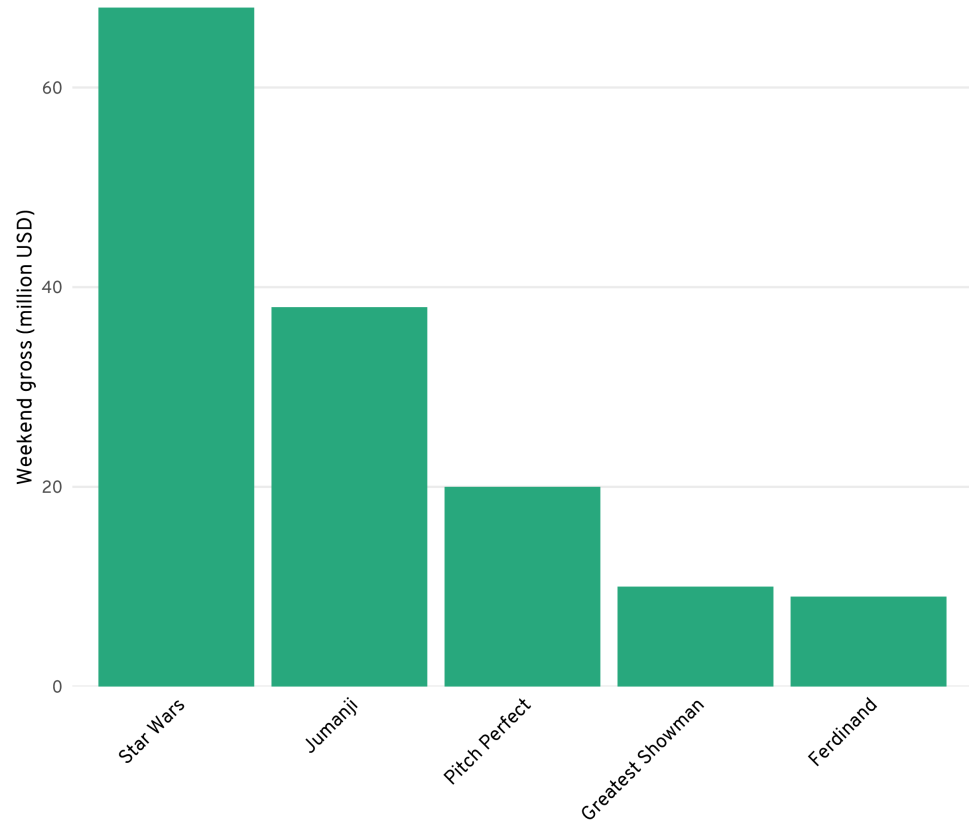
Align Your Text

- Left-align most text
- Title should be left aligned
- Labels and subtitles can be center or right aligned

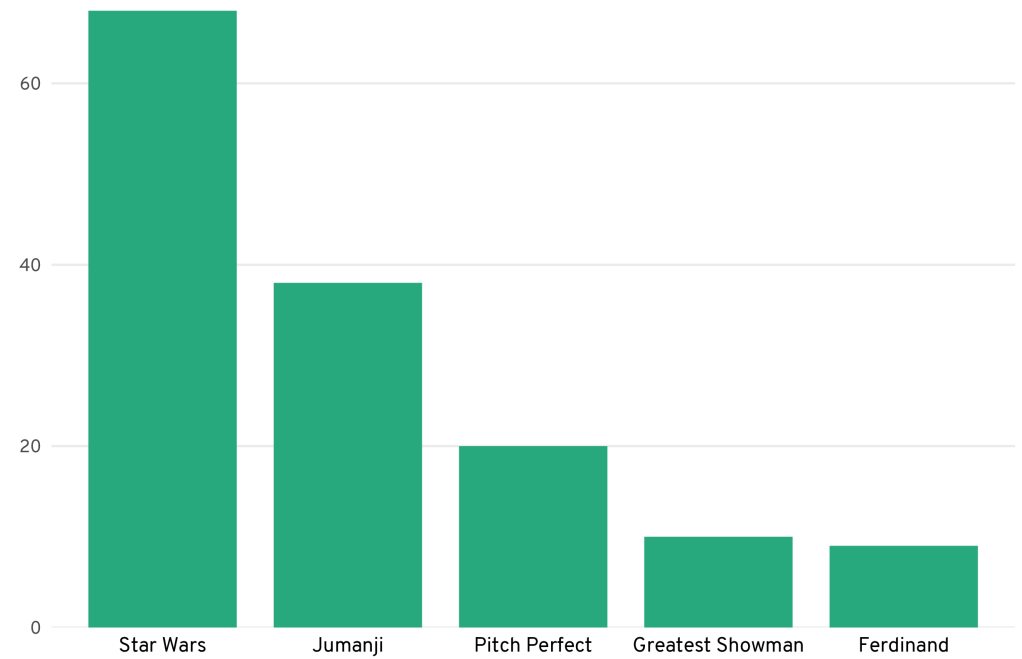


Source: Will R. Chase

(Don't) Rotate Your Text

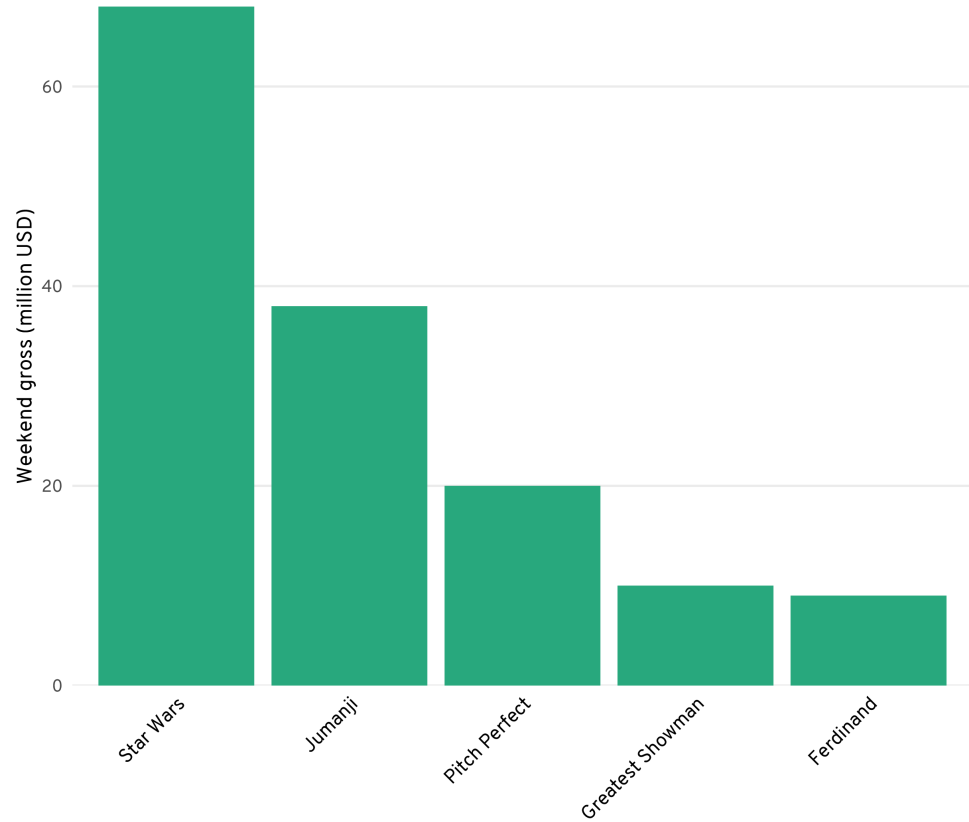


Weekend gross in million USD of popular blockbusters

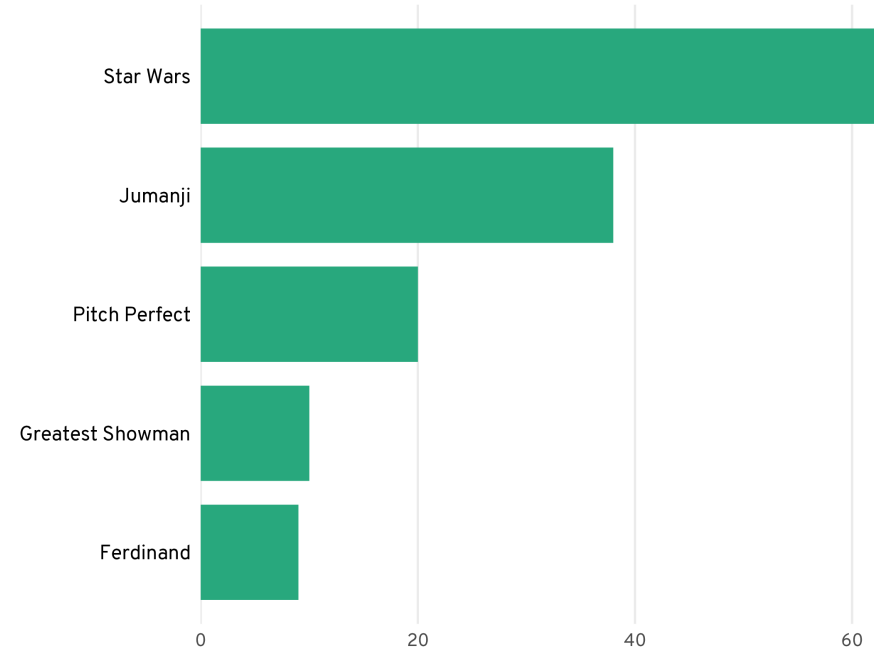


Modified from *Fundamentals of Data Visualization* by Claus Wilke

(Don't) Rotate Your Text

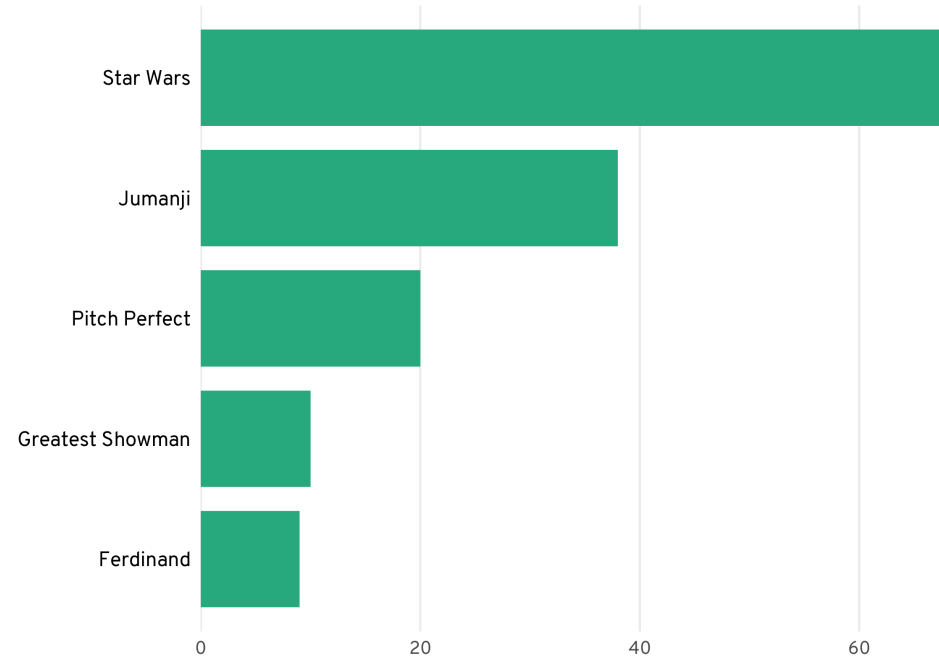


Weekend gross in million USD of popular blockbusters

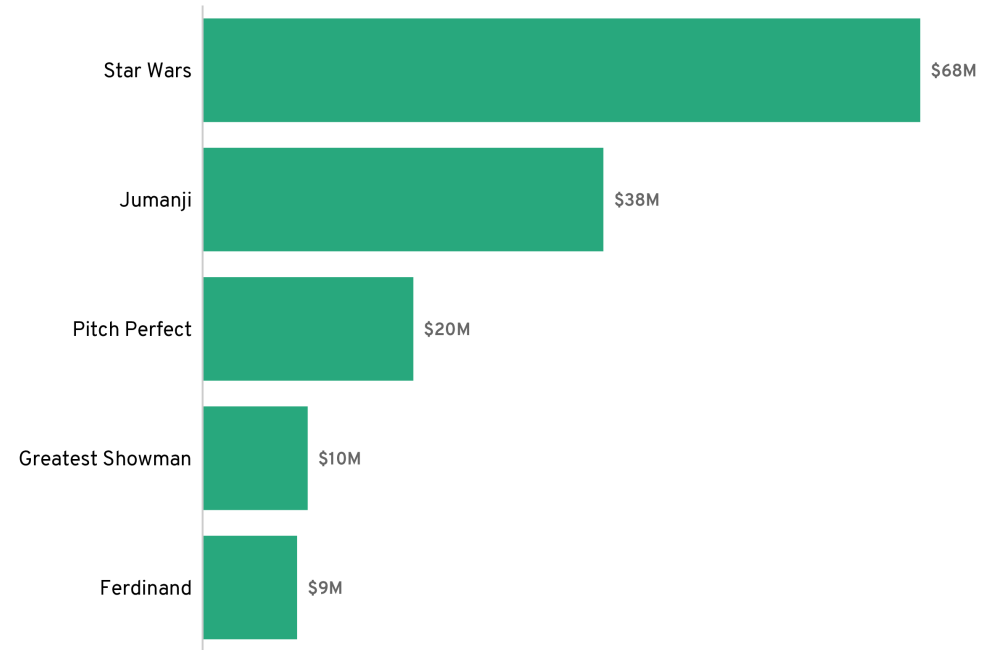


Use Annotations

Weekend gross in million USD of popular blockbusters

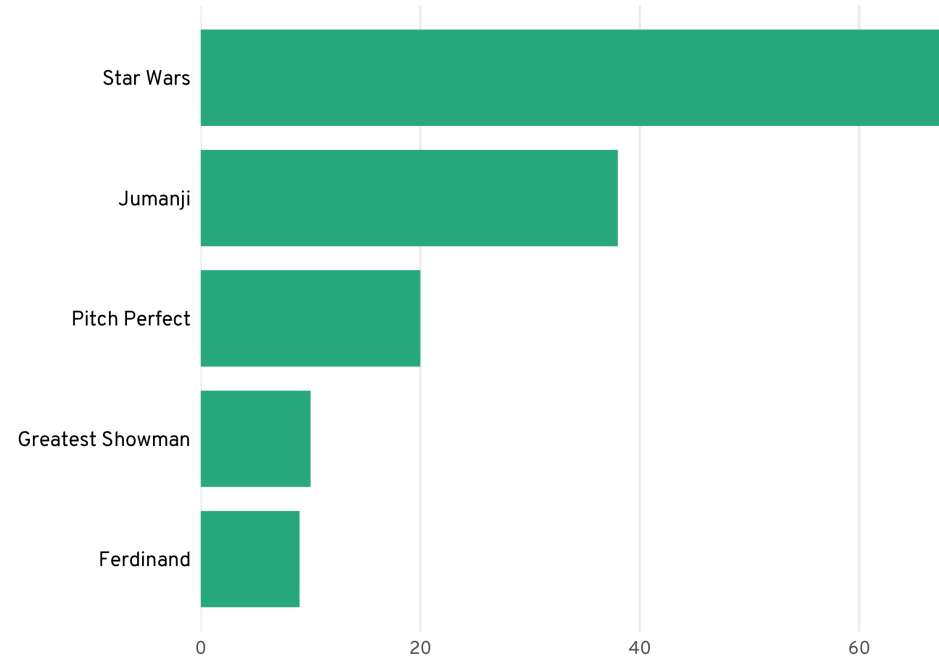


Weekend gross in million USD of popular blockbusters

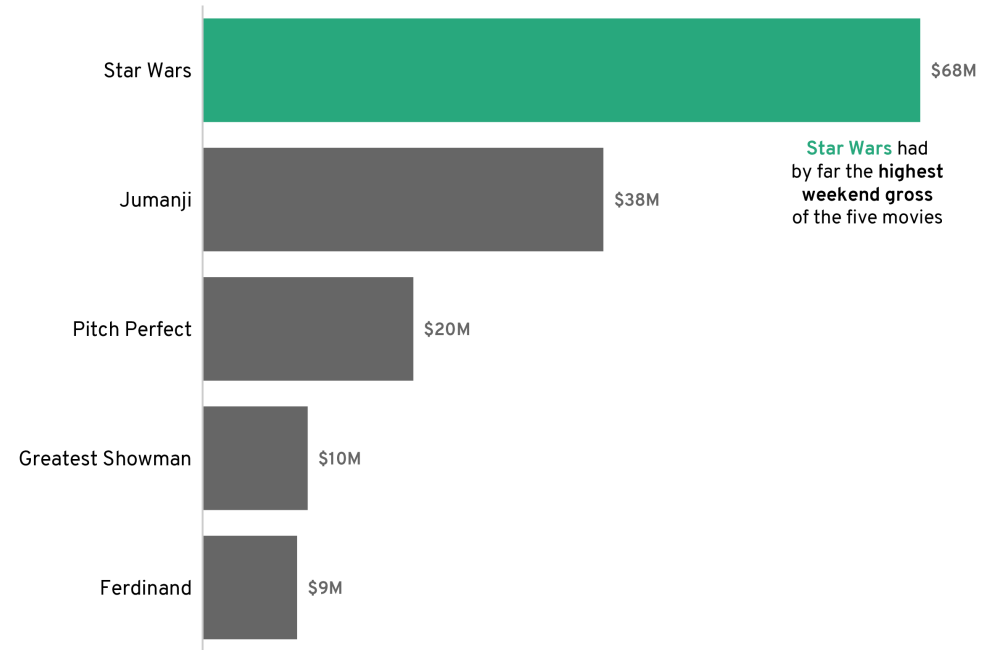


Use Annotations

Weekend gross in million USD of popular blockbusters



Weekend gross in million USD of popular blockbusters



Cédric Scherer
Z3tt

Computational Ecologist & Data Visualization Designer

Edit profile

293 followers · 82 following · 252

Home · IZW Berlin
Berlin
cedric@igggg.com
cedricscherer.netlify.app

Highlights

Arctic Code Vault Contributor
R

Organizations

igggg

Overview Repositories Projects Packages

2311 / R3D0P2 .md

Hi, I'm Cédric!

Blog Email Twitter Balance LinkedIn

I am a computational ecologist by training and a data visualization designer by heart with more than 9 years of hypothesis-driven research experience and strong skills in data wrangling, statistical analysis, model development and data visualization.

I am working as a scientific researcher (PostDoc) in the Department "Ecological Dynamics" at the Leibniz Institute for Zoo and Wildlife Research (IZW) in Berlin, Germany. Since the beginning of the year I have also been working successfully as a self-employed designer, consultant and instructor in the fields of data visualization and reproducible analysis. My favorite tool for all data-related tasks is R, an open source, highly extensible language for statistical computing and graphics techniques. To visualize data I mainly utilize the package [ggplot2](#) from the [tidyverse](#) package collection including many more that I use on a daily basis for all kinds of data preparation and analysis.

Computational Ecology

As a computational scientist, I apply analytical and mechanistic modeling approaches to answer questions related to the movement ecology of animals and the dynamics of populations, communities and diseases in space and time. By using empirical and simulated data, I investigate how disease dynamics are influenced by movement behavior, landscape structure and seasonality, how disturbance effect stability measures of ecological communities and how birds respond to global change. In 2019, I was awarded my Ph.D. degree in Ecology (Dr. rer. nat.) at the University of Potsdam as part of the [Booklove](#) research training group.

→ Read my publications and more about my projects

Data Visualization

My passion for data, design and coding is a perfect combination not only for scientific but all types of data visualization. By contributing to challenges such as [TidyTuesday](#), [MakeoverMonday](#), [Storytelling with Data](#) and the [30 Day Map Challenge](#), I am constantly strengthening my skills in design and reproducible data visualization with [ggplot2](#). Thanks to frequent contributions as well as several personal and client projects, my portfolio now includes visualizations for various purposes and is covering a wide range of topics and chart types.

→ Have a look at my [dataviz portfolio](#)

Pinned

TidyTuesday

My contributions to the #TidyTuesday challenge

133 54

Ingenieur/NLMBR

It's package to simulate neutral landscape models

30 14

igggg-courses

iggggC2 Teaching Material

38 10

DataViz-Teaching

Visualizations for DataViz Teaching

21 3

30DayMapChallenge

My contributions to the #30DayMapChallenge

25 17

Corona-Worst-Days

Visualizations showing confirmed daily deaths due to COVID-19 worldwide and per country relative to each worst day

3 3

1,879 contributions in the last year

Contribution settings

2020

2019

2018

2017

2016

2015

github.com/Z3tt

CÉDRIC SCHERER

BLOG ABOUT ME PUBLICATIONS VISUALIZATIONS LINKS

CÉDRIC SCHERER

Computational Ecology & Data Visualization

THE WORST DAYS OF THE CORONAVIRUS PANDEMIC SO FAR

Coronavirus SARS-CoV-2, COVID-19 or simply Corona—what started as an epidemic in China' has become a global pandemic. I created an animated timeseries of daily deaths relative to each country's worst day so far to visualize the first wave of COVID-19.

POSTED BY CÉDRIC TUESDAY, MARCH 31, 2020

COMPARING THE EXTENT OF THE AUSTRALIAN BUSHFIRES 2019/20

The massive bushfires in Australia are in the news worldwide. The incredible extent of burnt land and plume of smoke is hard to imagine so I have compared the areas to countries in Europe and worldwide.

POSTED BY CÉDRIC THURSDAY, JANUARY 8, 2020

BEST TIDYTUESDAY 2019

Here are my favorite visualizations of the #TidyTuesday challenge in 2019 (from those I've seen and which I remember). I present my personal top 3 in terms of design and storytelling.

POSTED BY CÉDRIC MONDAY, DECEMBER 30, 2019

MERRY (WHITE?) CHRISTMAS!

At the end of the year, I explore the history of snow cover and white Christmas in Berlin. I wish you a merry Christmas and wonderful holidays 2019!

POSTED BY CÉDRIC TUESDAY, DECEMBER 24, 2019

VISUALIZING TEMPERATURES IN BERLIN WITH BAR CHART RACES

You have seen Bar Chart Races. You have seen Bar Chart Races of Bar Chart Races. Here are some more visualizing monthly temperatures in Berlin! Yes, I jumped on the hype train. Of course, using [ggplot2](#) and [gganimate](#).

POSTED BY CÉDRIC TUESDAY, SEPTEMBER 17, 2019

ABOUT ME

Computational Ecologist •
Data Visualization Designer
• Proud Dad

FEATURED TAGS

[BERLIN](#)
[DATAVIZ](#)
[R](#)
[TIDYTUESDAY](#)
[ANIMATIONS](#)
[GGPLOT2](#)
[MAPS](#)
[TIDYVERSE](#)
[TUTORIAL](#)
[WEATHER](#)

FRIENDS & CO

[DataVizSociety](#)
[R4DS Community](#)
[CorrelAid](#)
[FrontPage Data](#)
[Dep. Ecological Dynamics](#)
[SIG Computational Ecology](#)

BOOKMARKS

[Fundamentals of DataViz](#)
[R for Data Science](#)
[Geocomputation with R](#)

cedricscherer.netlify.com

CÉDRIC SCHERER

BLOGABOUT MEPUBLICATIONSVISUALIZATIONSLINKS

DATAVIZTUTORIAL8TIDYVERSEGGPLOT2

A GGPLOT2 TUTORIAL FOR BEAUTIFUL PLOTTING IN R

POSTED BY CÉDRIC ON MONDAY, AUGUST 5, 2019

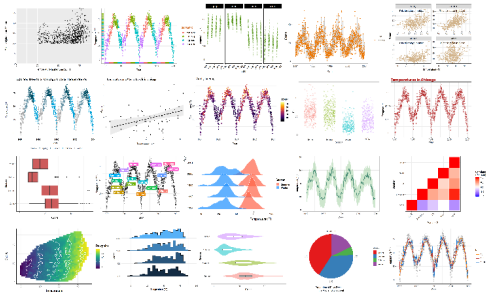
Last update: 2019-11-01

INTRODUCTORY WORDS

I don't care, just show me the content!

Begin of 2016, I had to prepare my PhD introductory talk and I started using (`ggplot2`) to visualize my data since I never liked the syntax and style of base plots in R. Because I was short on time, I plotted these figures by try'n'error and with the help of lots of googling. The resource I came always back to was a blog entry called *Beautiful plotting in R: A ggplot2 cheatsheet* by Zev Ross, posted on 4. August 2014, updated last in January 2016. After giving the talk which contained some quite beautiful plots thanks to the blog post, I decided to go through this tutorial step-by-step. I learned so much from it and directly started modifying the codes and over the time I added some additional code snippets, chart types and resources.

Since the blog entry by Zev Ross was not updated for some years, I hosted the updated version on my GitHub. Now it finds its proper place on this homepage! (Plus I added some updates, for example the fantastic (`patchwork`) and (`ggforce`) packages. And pie charts because everyone looooves pie charts!)



Some exemplary plots included in this tutorial.

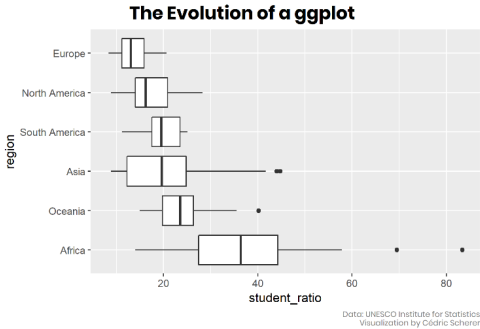
CÉDRIC SCHERER

BLOGABOUT MEPUBLICATIONSVISUALIZATIONSLINKS

DATAVIZTUTORIALANIMATIONSGGPLOT EVOLUTION8GGPLOT2TIDYVERSETIDYTUESDAY

THE EVOLUTION OF A GGPLOT (EP. I)

POSTED BY CÉDRIC ON FRIDAY, MAY 17, 2019



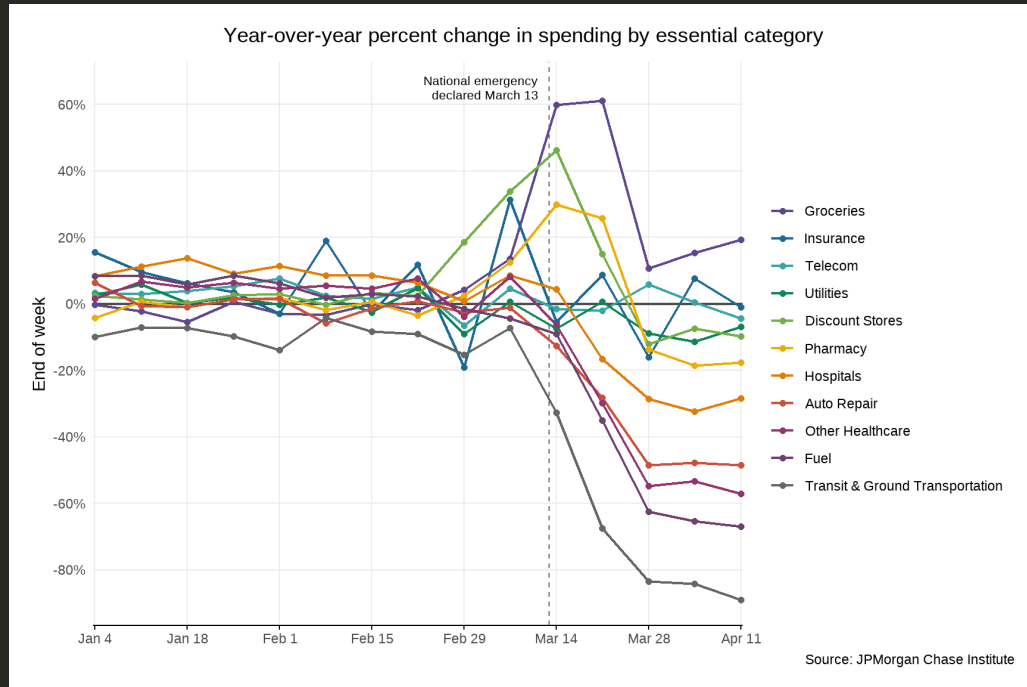
- 📖 Aim of this Tutorial
- 🔧 Data Preparation
- 📊 The Default Boxplot
- 📦 Sort Your Data!
- 💡 Let Your Plot Shine—Get Rid of the Default Settings
- 📄 The Choice of the Chart Type
- 🎨 More Geoms, More Fun, More Info!
- 💬 Add Text Boxes to Let The Plot Speak for Itself
- 🗺️ Bonus: Add a Tile Map as Legend
- 📈 The Final Evolved Visualization
- 📄 Complete Code for Final Plot
- 📄 Post Scriptum: Mean versus Median

🚩 AIM OF THIS TUTORIAL

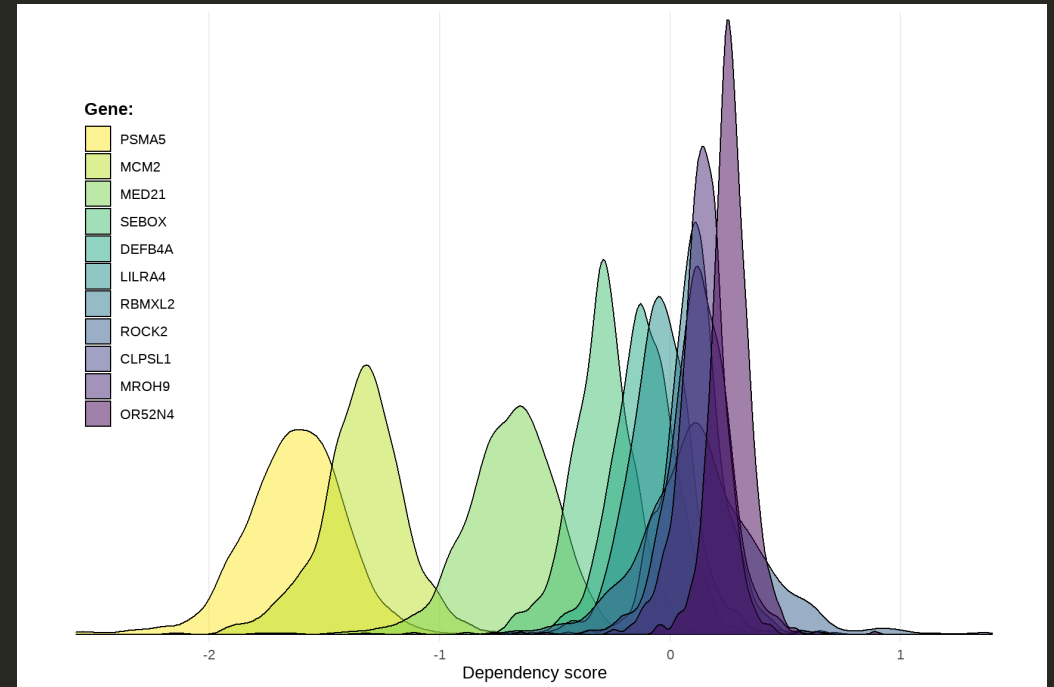
In this series of blog posts, I aim to show you how to turn a default ggplot into a plot that visualizes information in an appealing and easily understandable way. The goal of each blog post is to provide a step-by-step tutorial explaining how my visualization have evolved from a typical basic ggplot. All plots are going to be created with 100% (`ggplot2`) and 0% Inkscape.

In the first episode, I transform a basic boxplot into a colorful and self-explanatory combination of a jittered dot strip plot and a lollipop plot. I am going to use [data provided](#)

Hands-On Coding



Timeseries by JPMorgan Chase Institute



Gene Distribution by DataDrivenHypothesis

z3tt.github.io/bespokeDS