

Effective Data Visualization

How to design impactful and aesthetically pleasing charts

Dr. Cédric Scherer

bespokeDS | September 28th 2020

Photo by Richard Strozyński

Scientist by

→ population and community dynamics, movement ecology, wildlife diseases

- Ph.D. in *Ecology*
- PostDoc in *Computational Ecology*

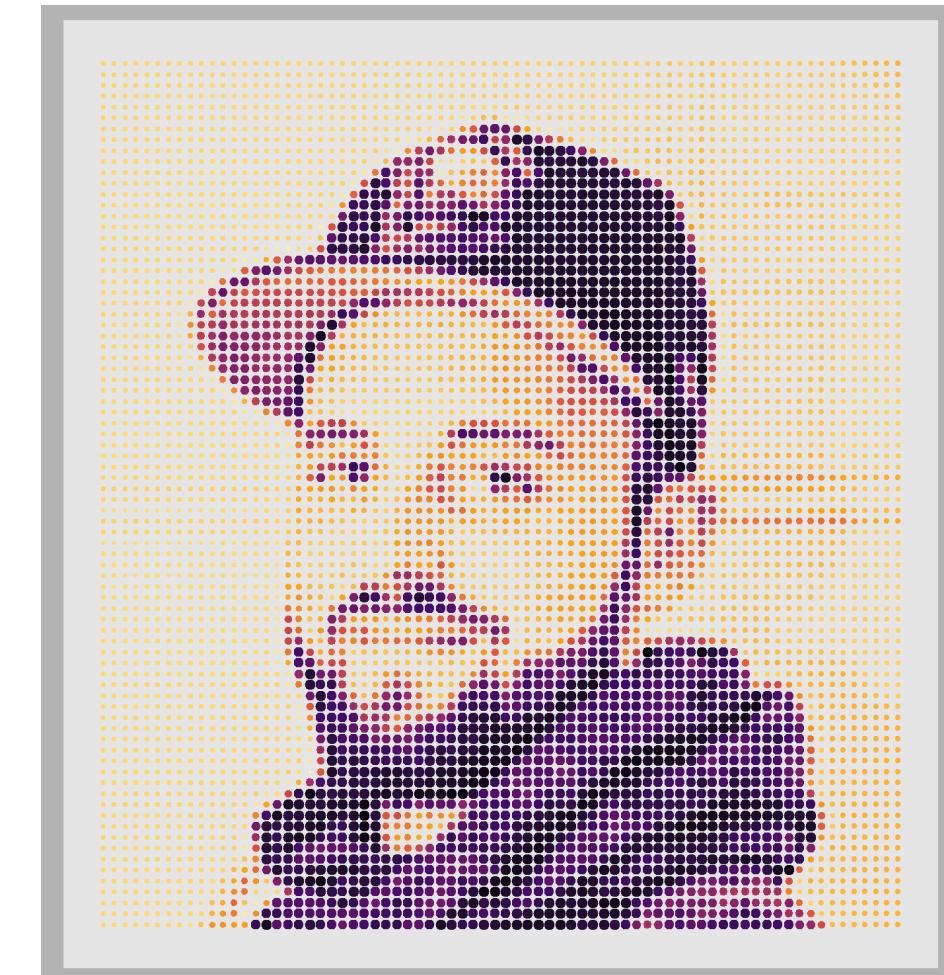
@ Leibniz Institute for Zoo and Wildlife Research (IZW)

DataViz Designer by

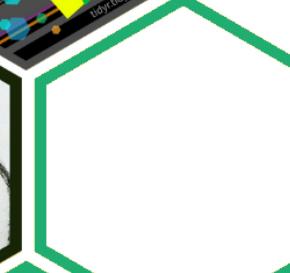
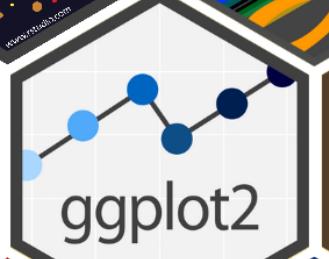
→ analyst, designer, consultant, lecturer

- Freelancing Data Visualization Specialist
- PostDoc in *Computational Ecology*
- Data Challenges and Personal Projects

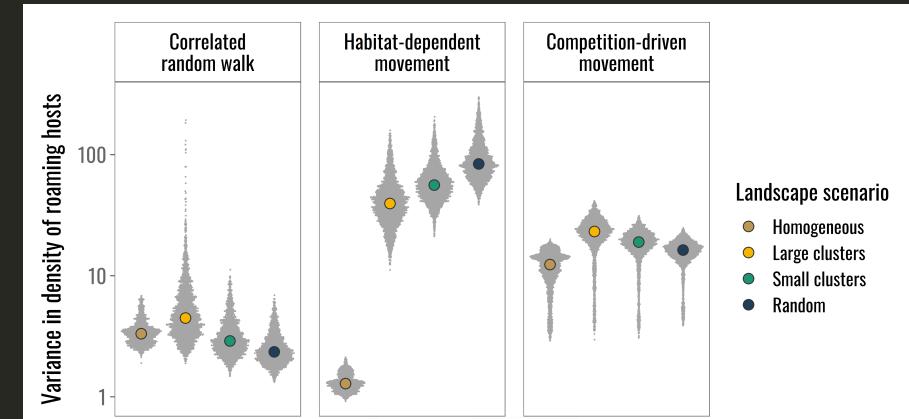
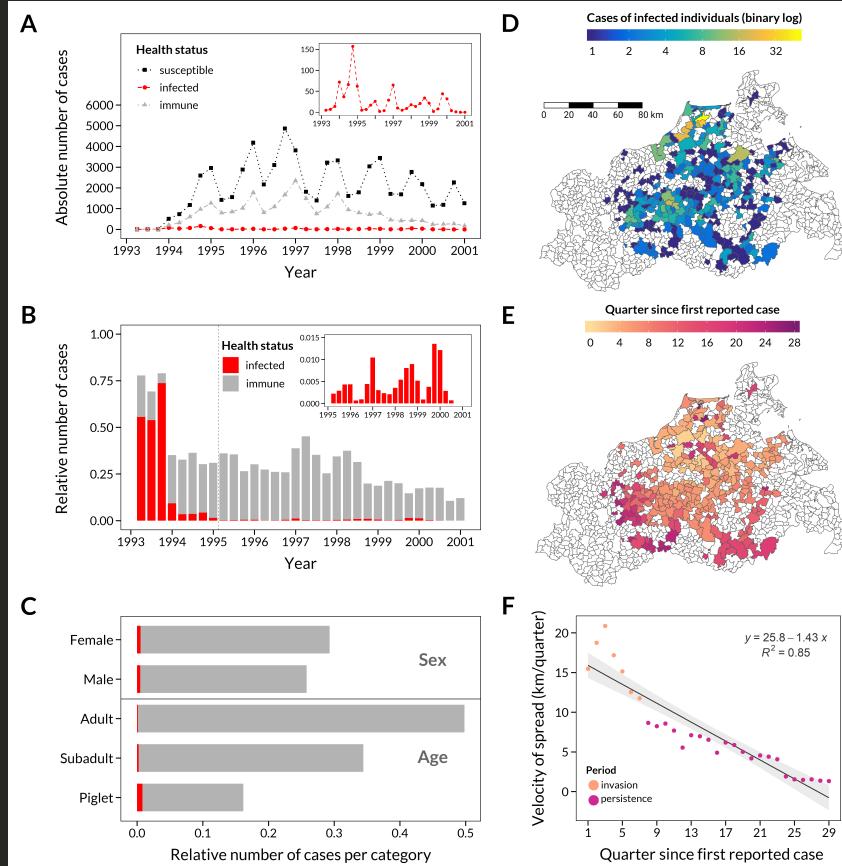
#TidyTuesday, #30DayMapChallenge, #MakeoverMonday, #SWDchallenge



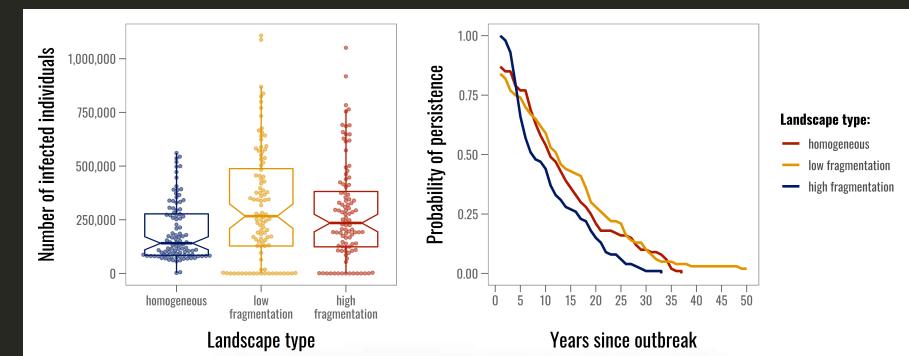
Scientist by  DataViz Designer by



Data Visualizations for Scientific Publications & Talks



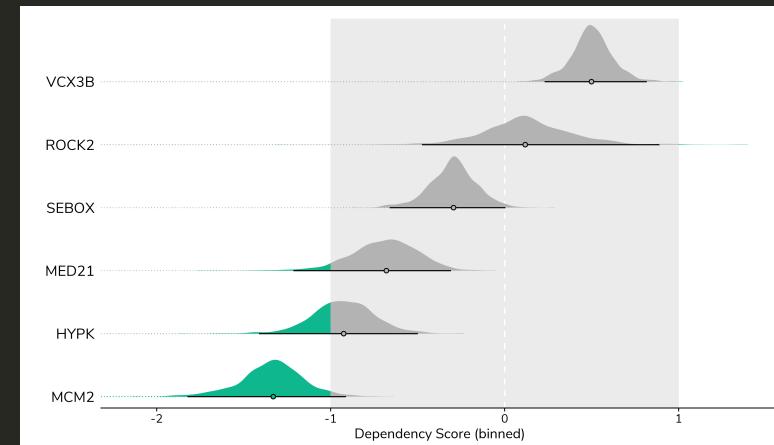
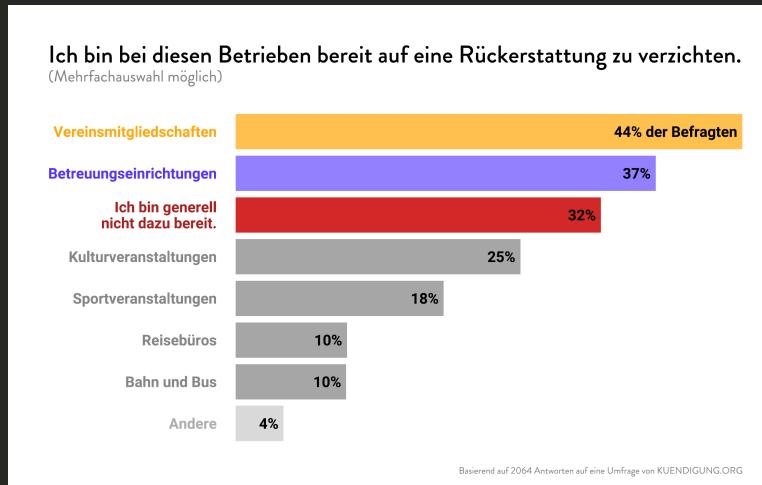
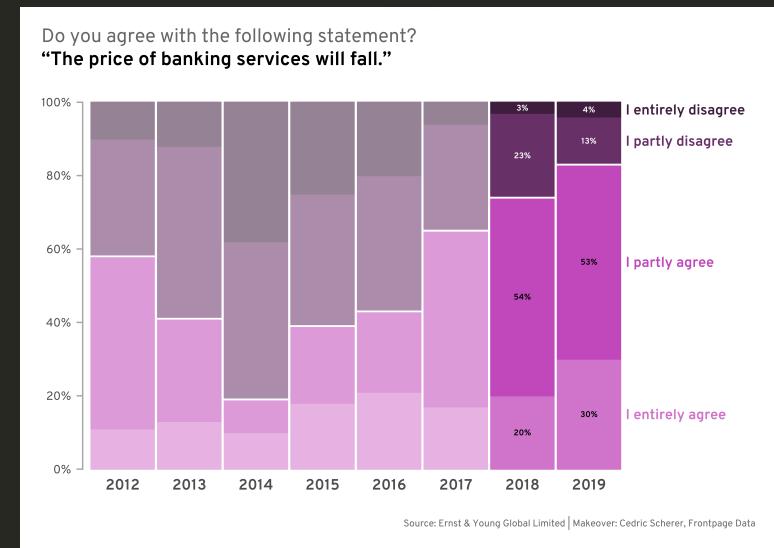
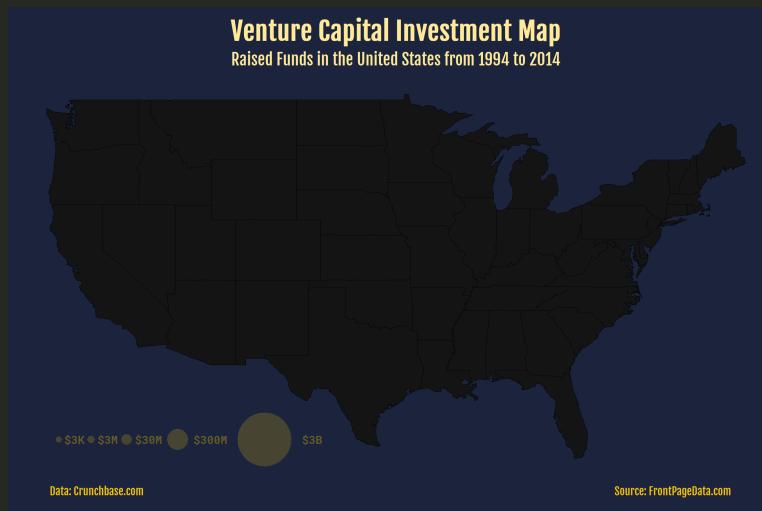
Scherer et al. 2020 *Oikos*



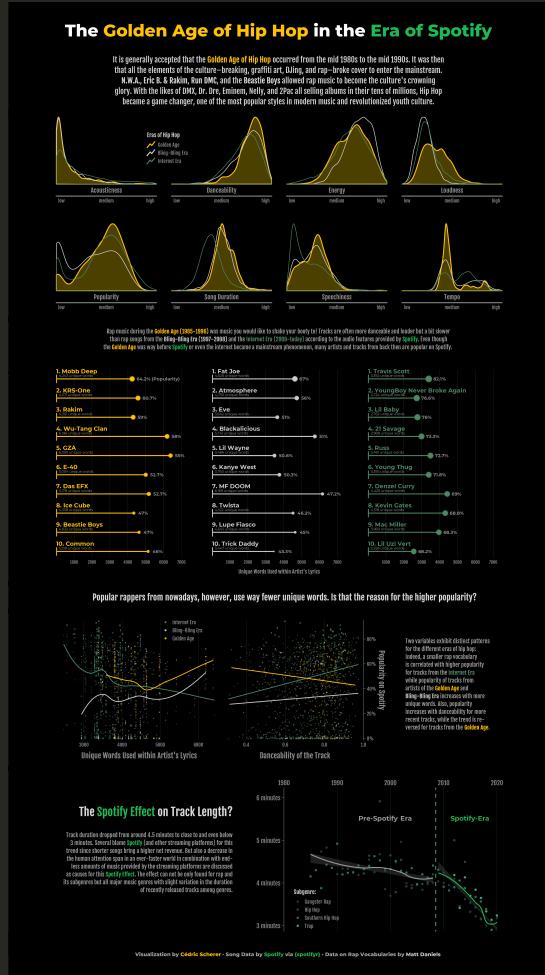
Sciaini et al. 2019 *Methods in Ecology & Evolution*

Scherer et al. 2019 *Journal of Animal Ecology*

Data Visualizations for Client Projects



Data Visualizations as Challenge Contributions



Contribution to #TidyTuesday



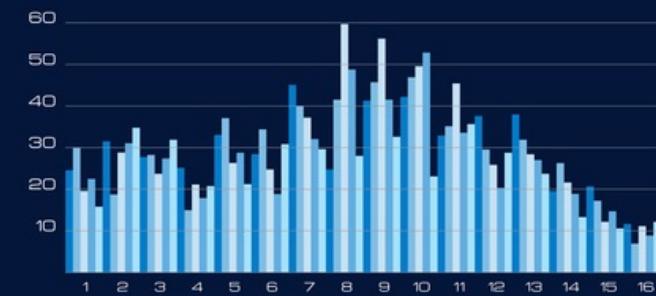
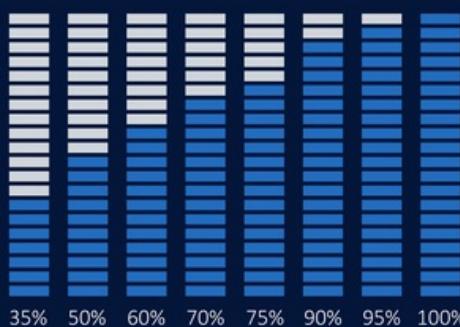
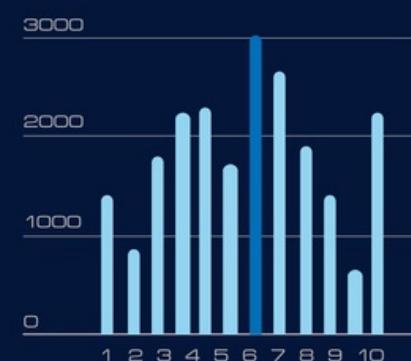
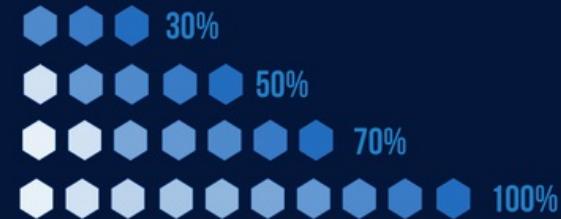
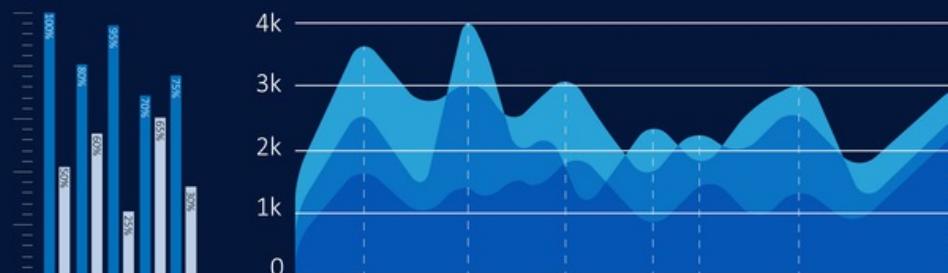
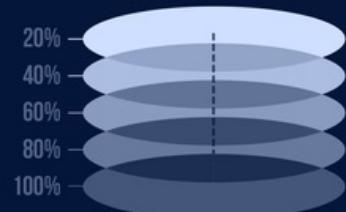
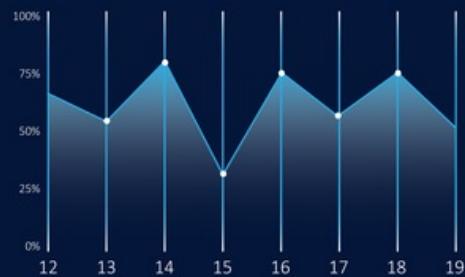
Contribution to #30DaymapChallenge

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.



hanging (strangulation)



taking drugs/alcohol/medicines



in front of train or metro



drowning



jumping from height



other method*



unknown method

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Précise 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 entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. Chiers, de Léger, de Fezensac, de Chambray et le journal médical 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 Sébastien et du Maréchal Davout qui avaient été délocalisés sur Minsk et Malibow et qui rejoignirent Orléans et Witebsk, avaient toujours marché avec l'armée.

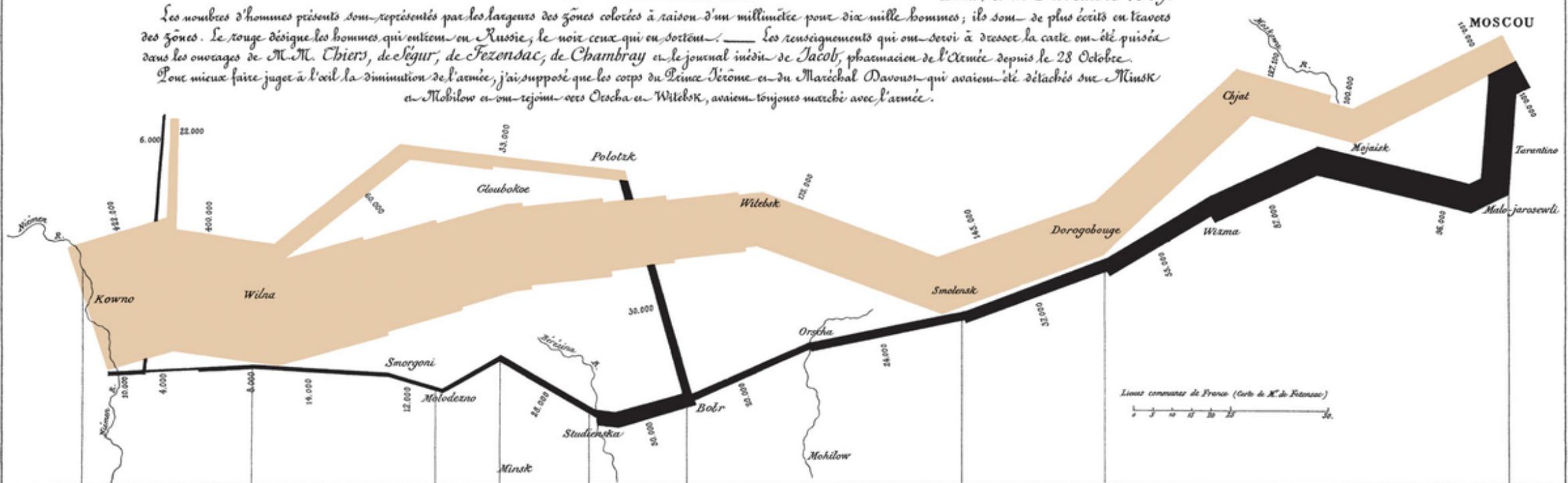
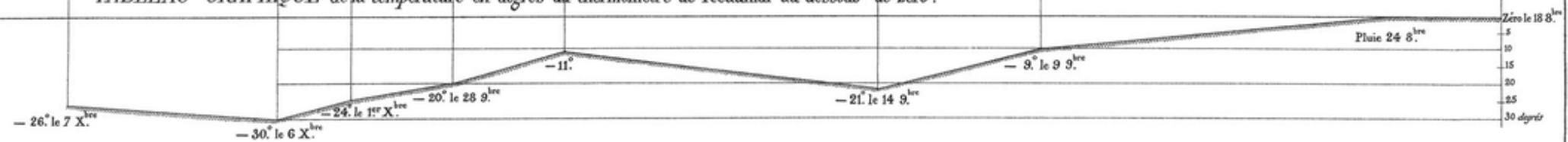
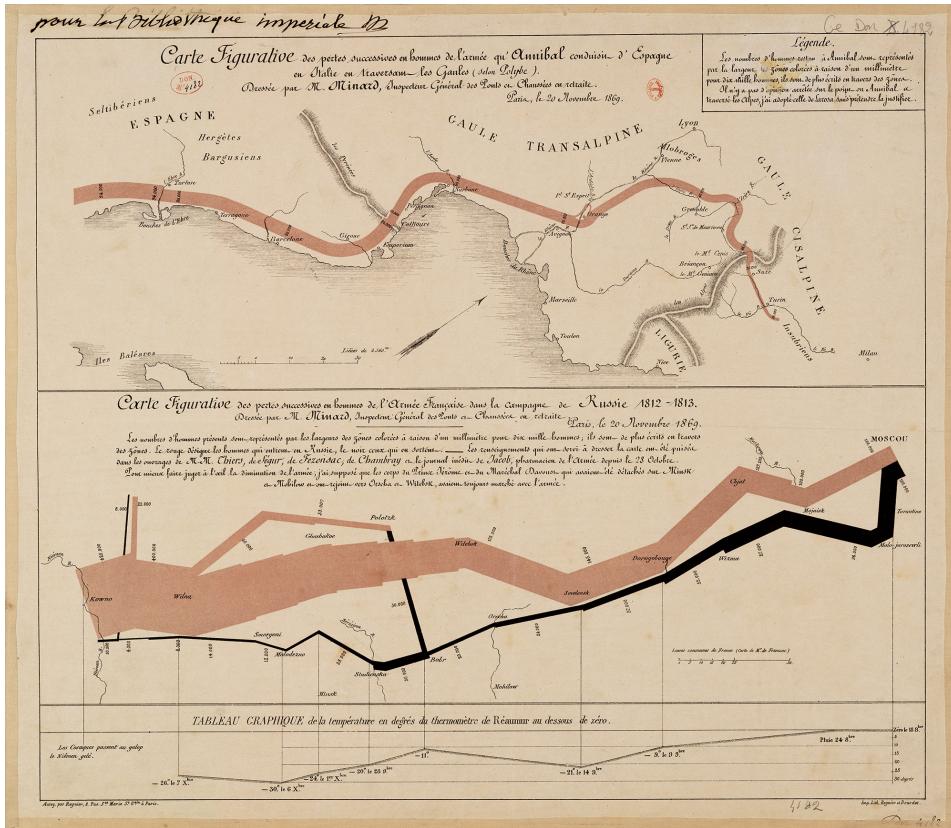


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niemen gelé.





What Makes It a Bad Data Visualization

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

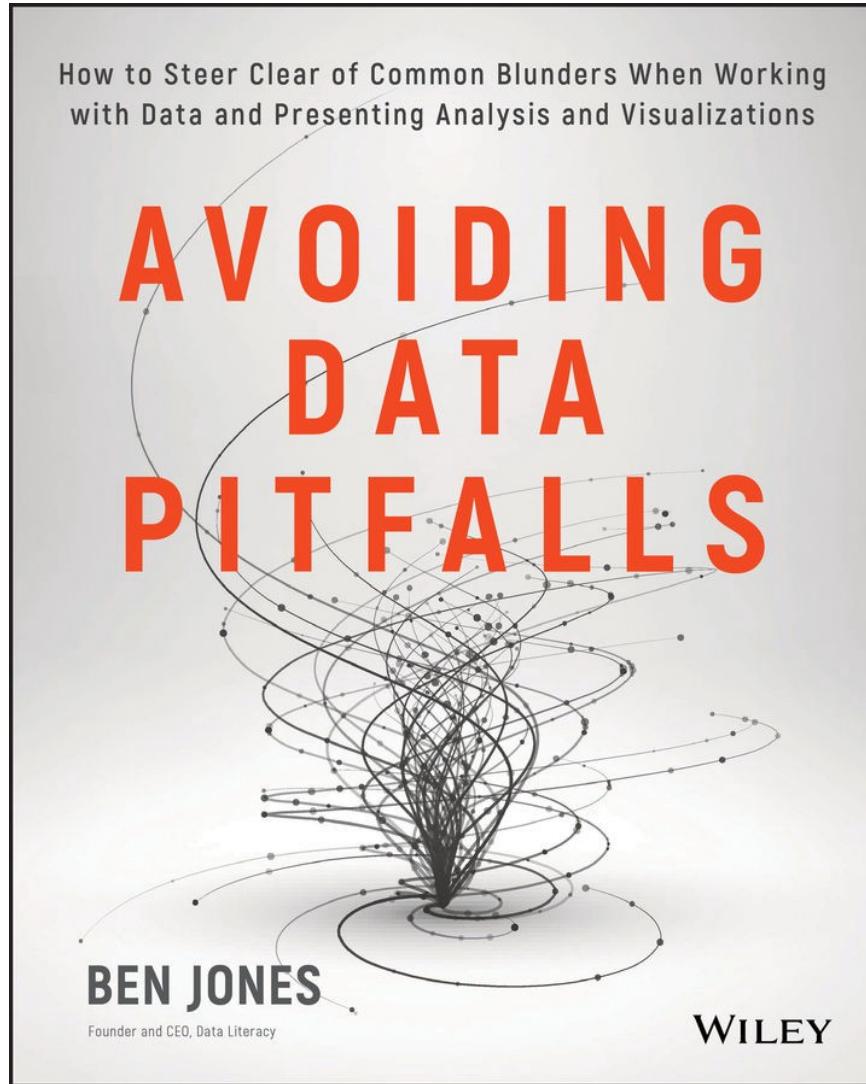
What Makes It a Good Data Visualization

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

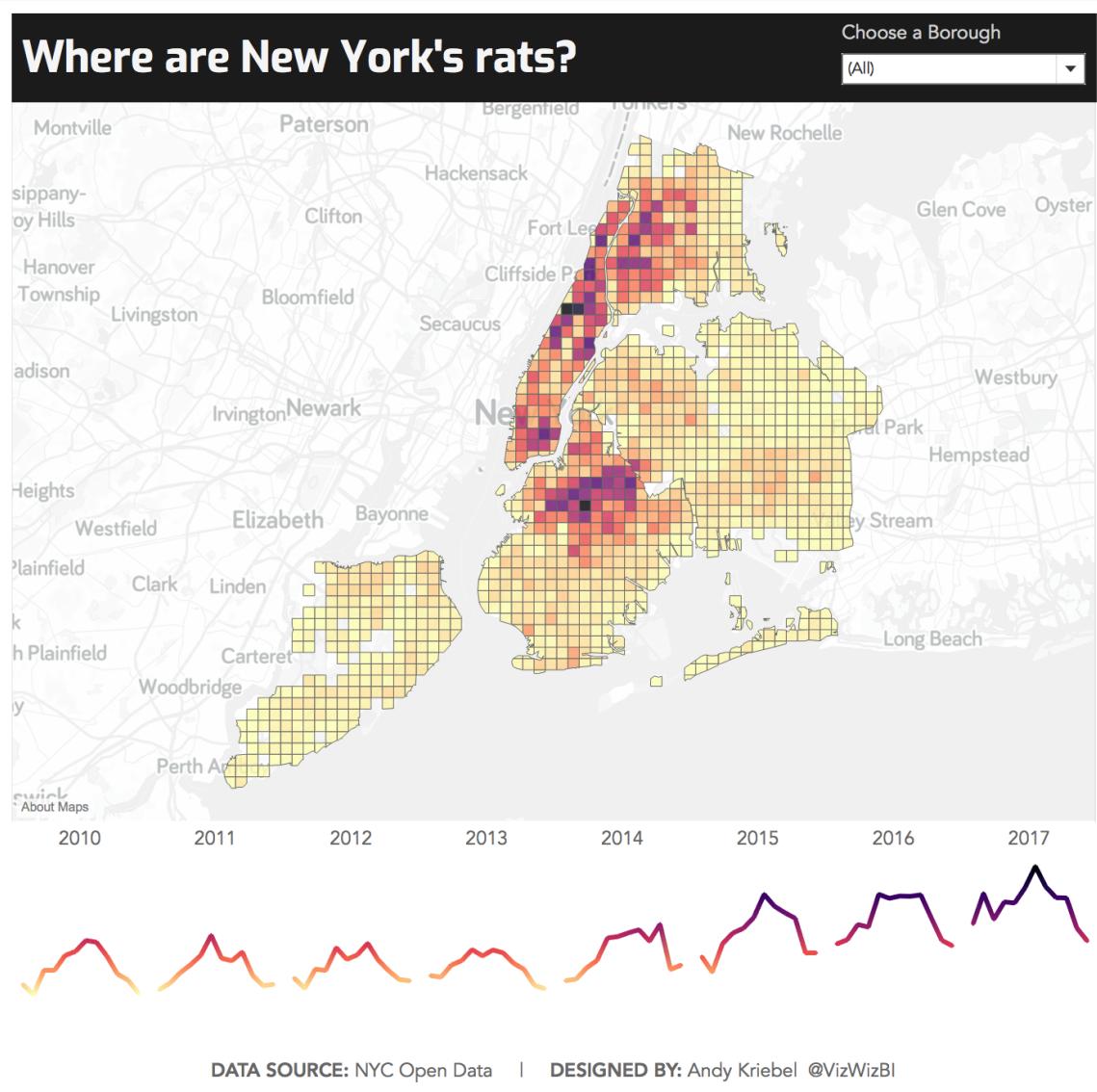


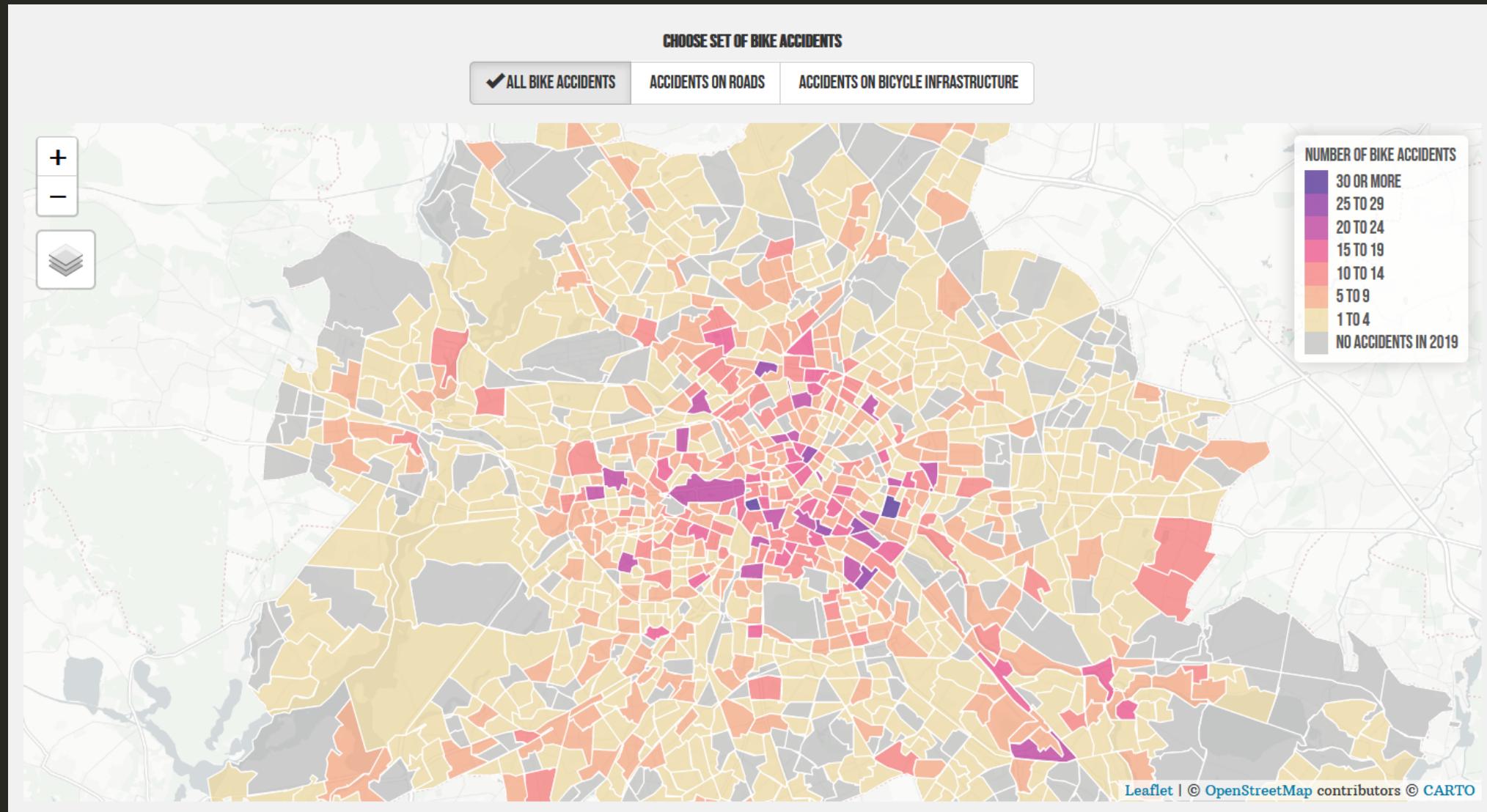
INFORMATION

Understand your data and be accurate



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





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



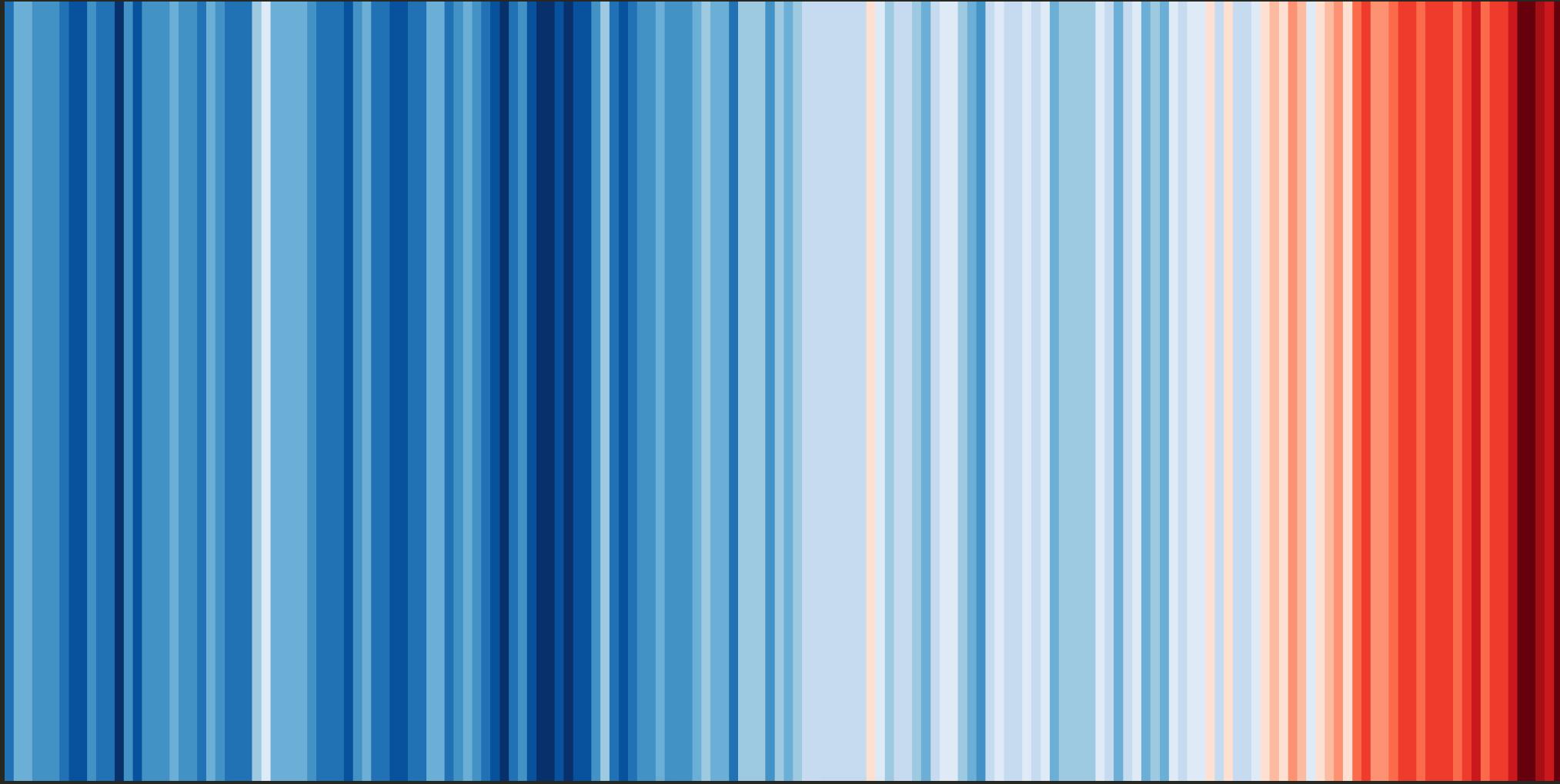
Source: inhomelandsecurity.com/risk-management-and-black-swan-events

STORY

Be clear about the message of your visualization

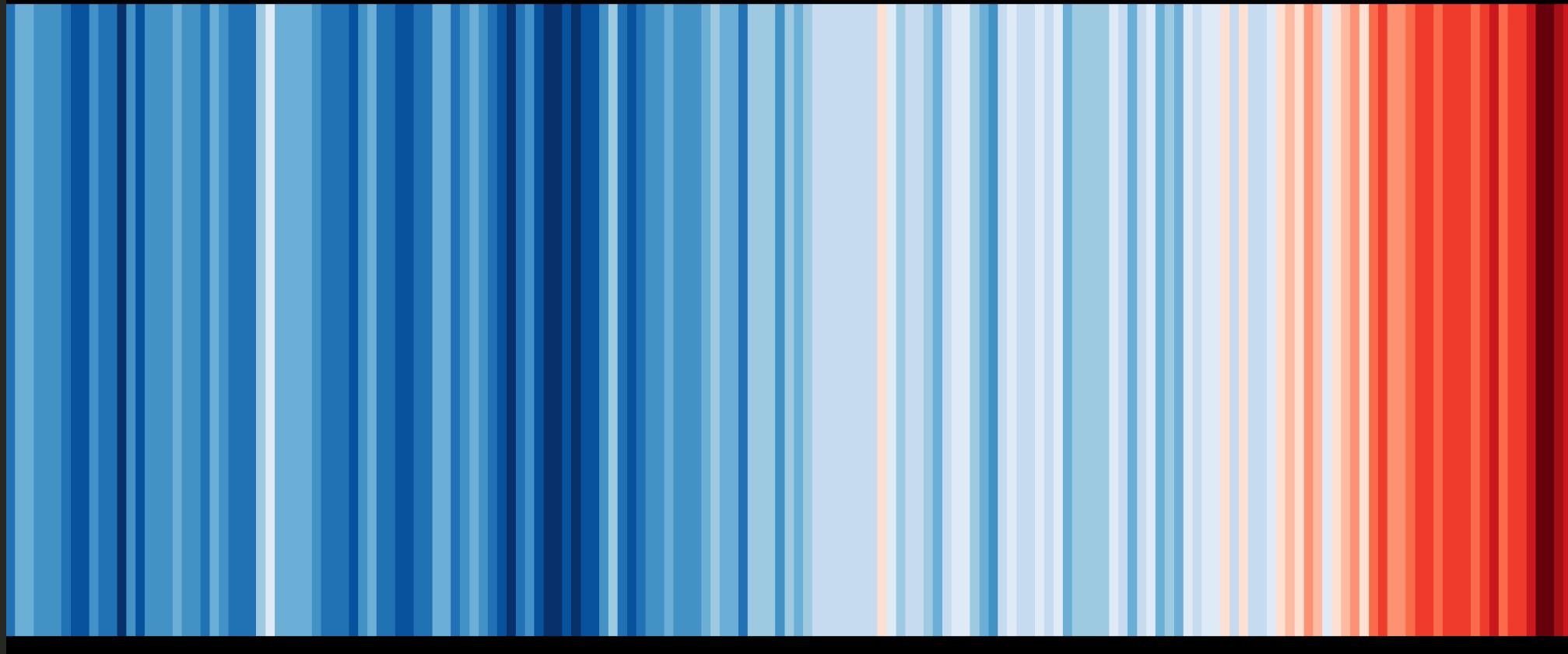
Who is my audience?

What am I trying to achieve with the visualization?

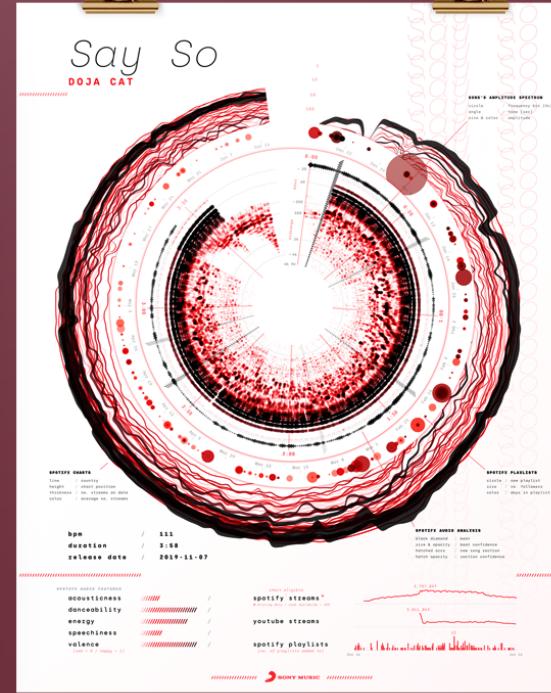
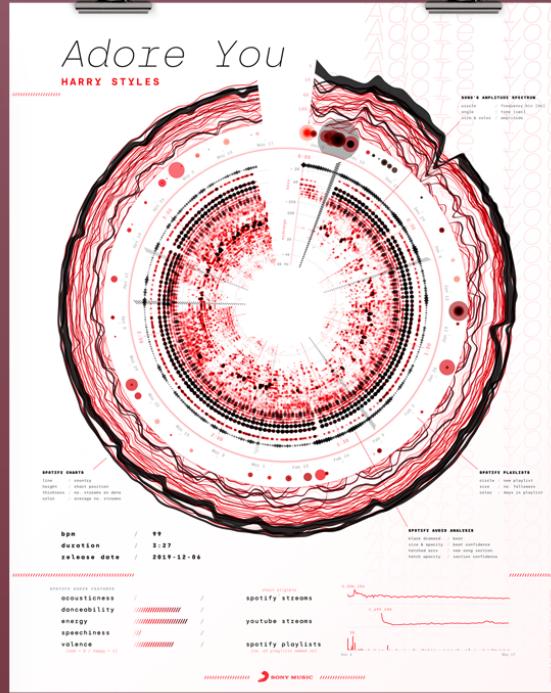
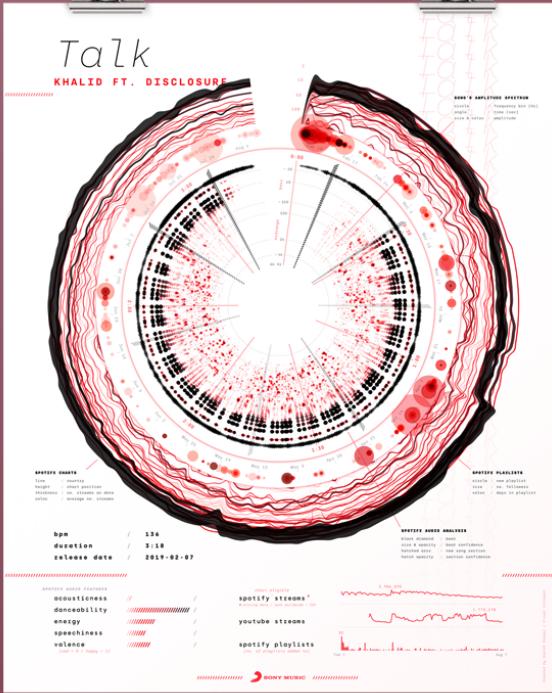


Warming Stripes by Ed Hawkins

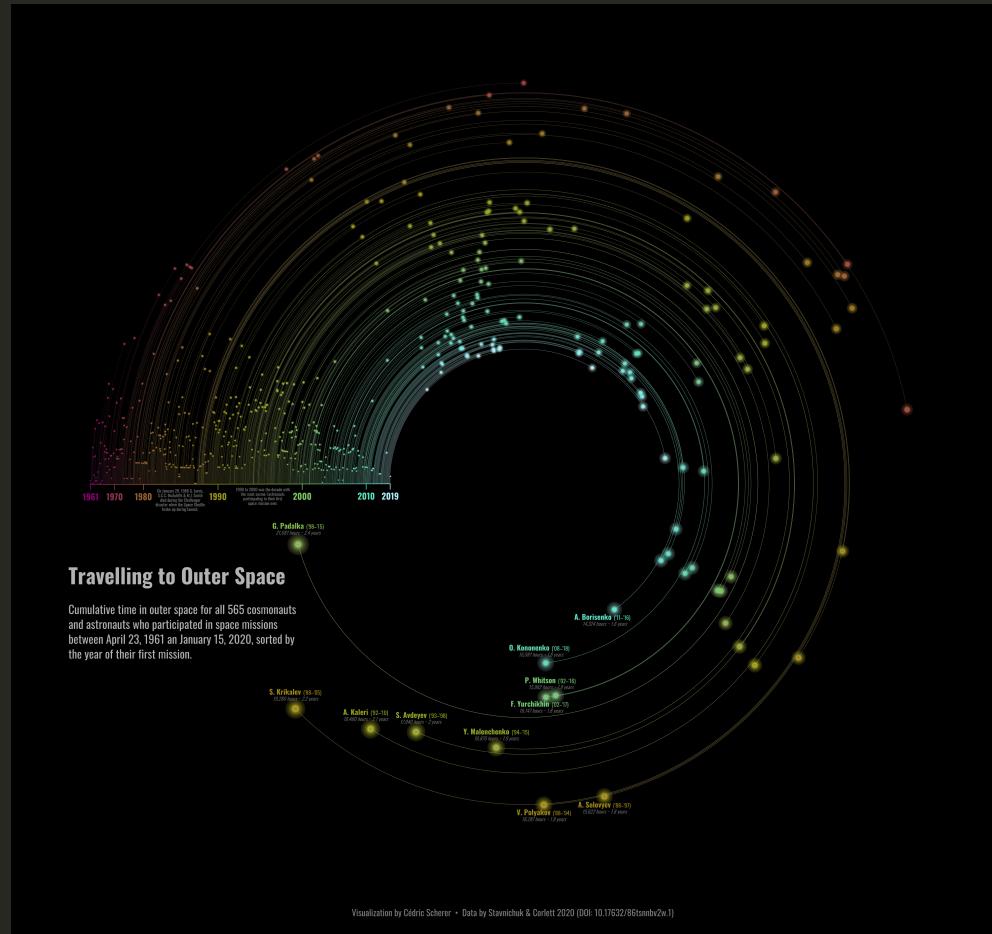
Global temperature change (1850-2019)



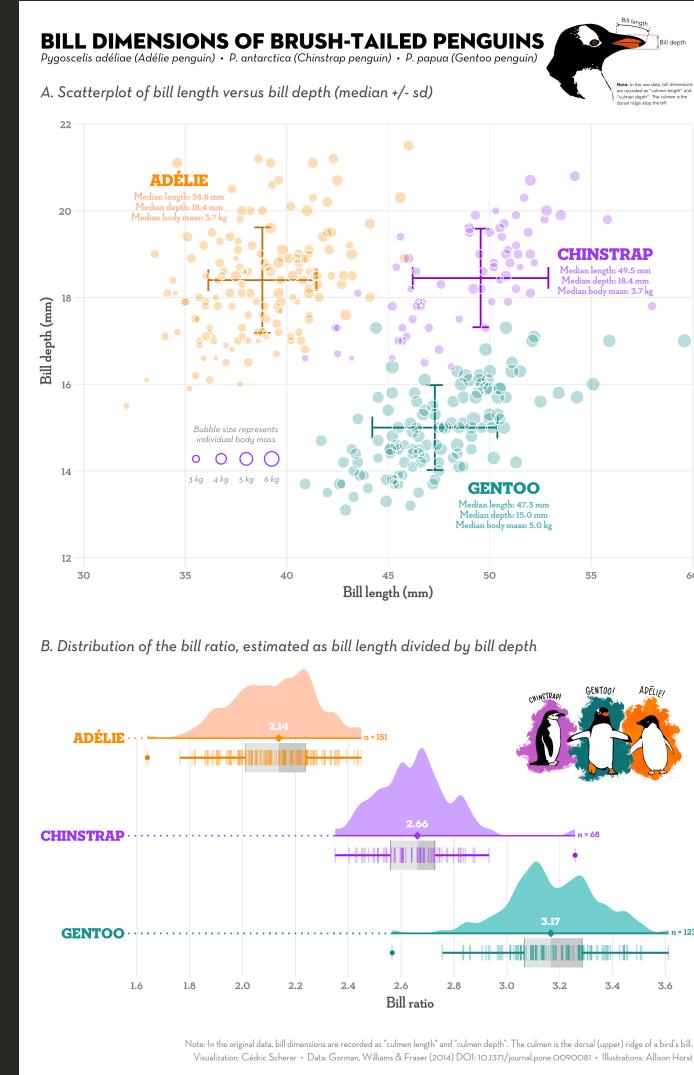
Warming Stripes by Ed Hawkins



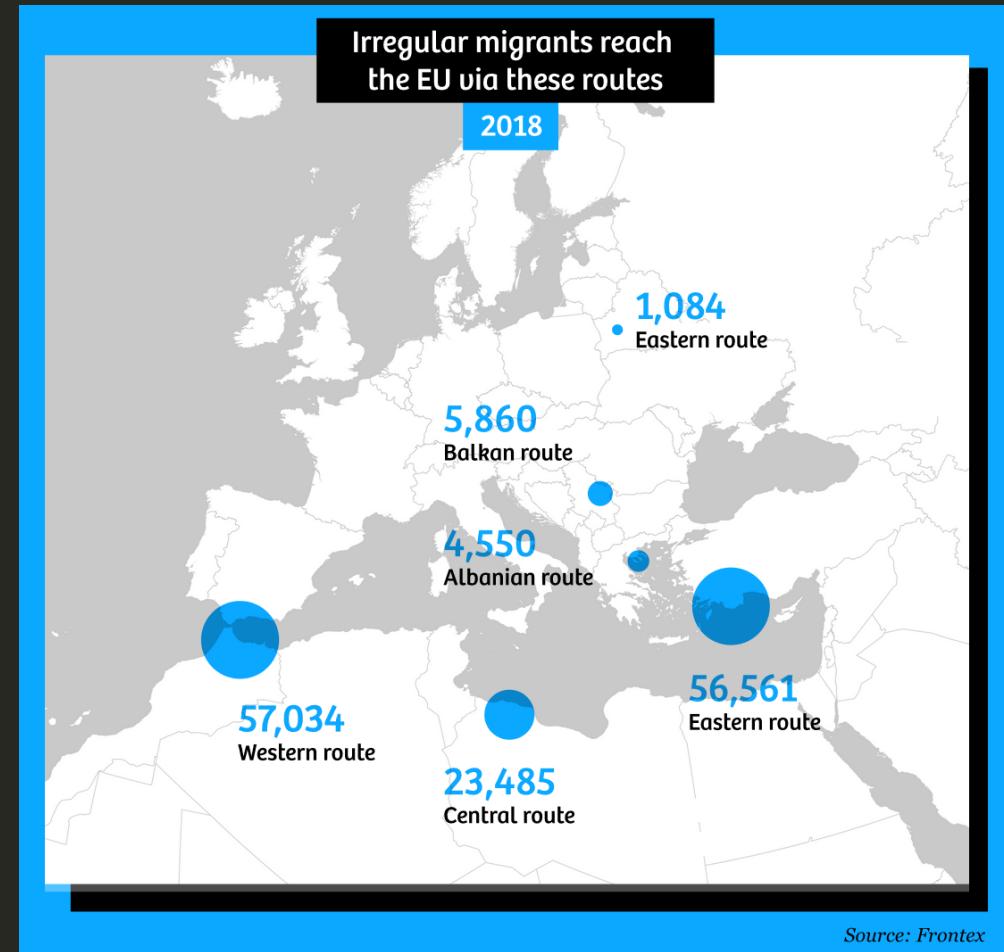
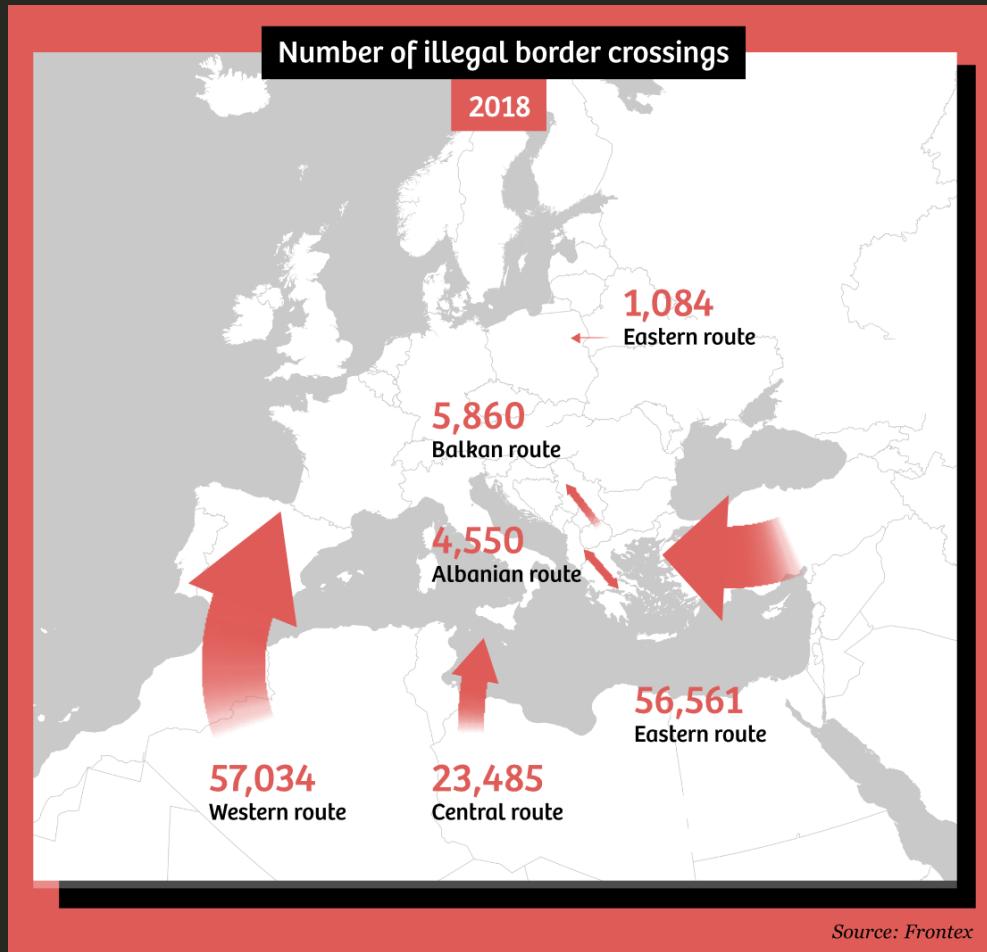
Sony Music Song Posters by Nadieh Bremer



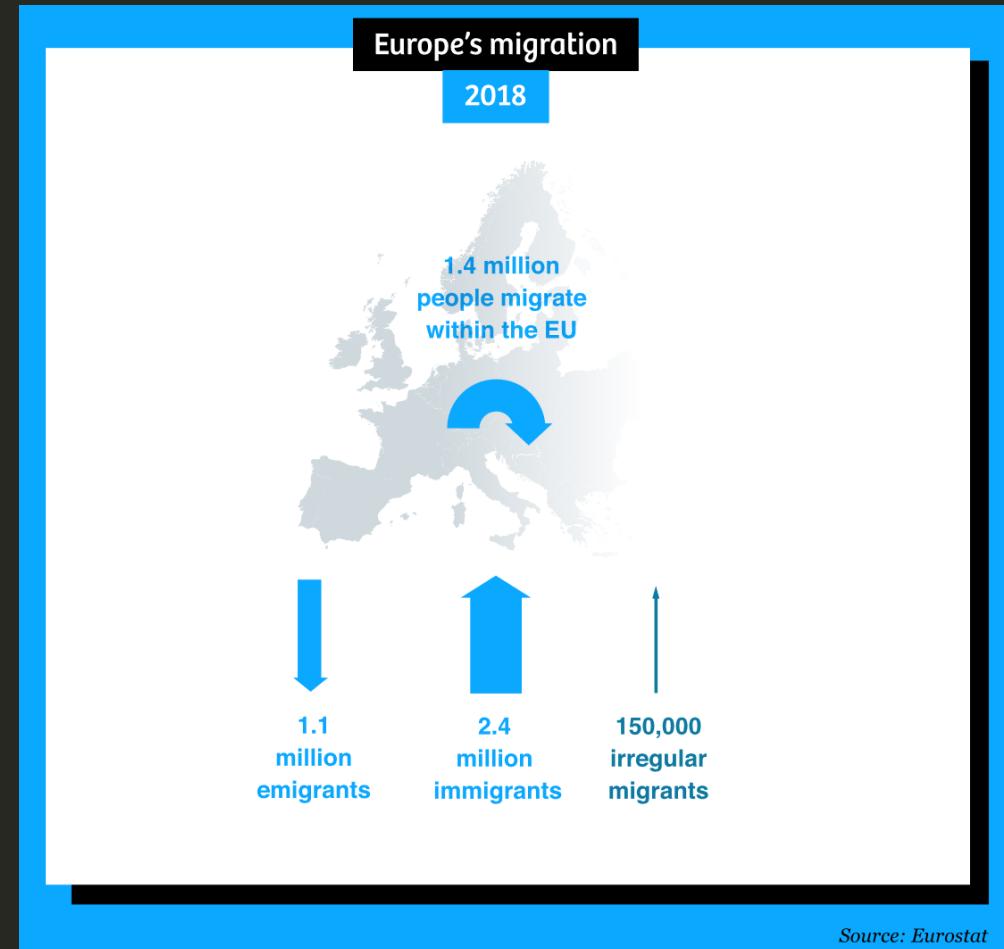
#TidyTuesday contribution Week 2020/29



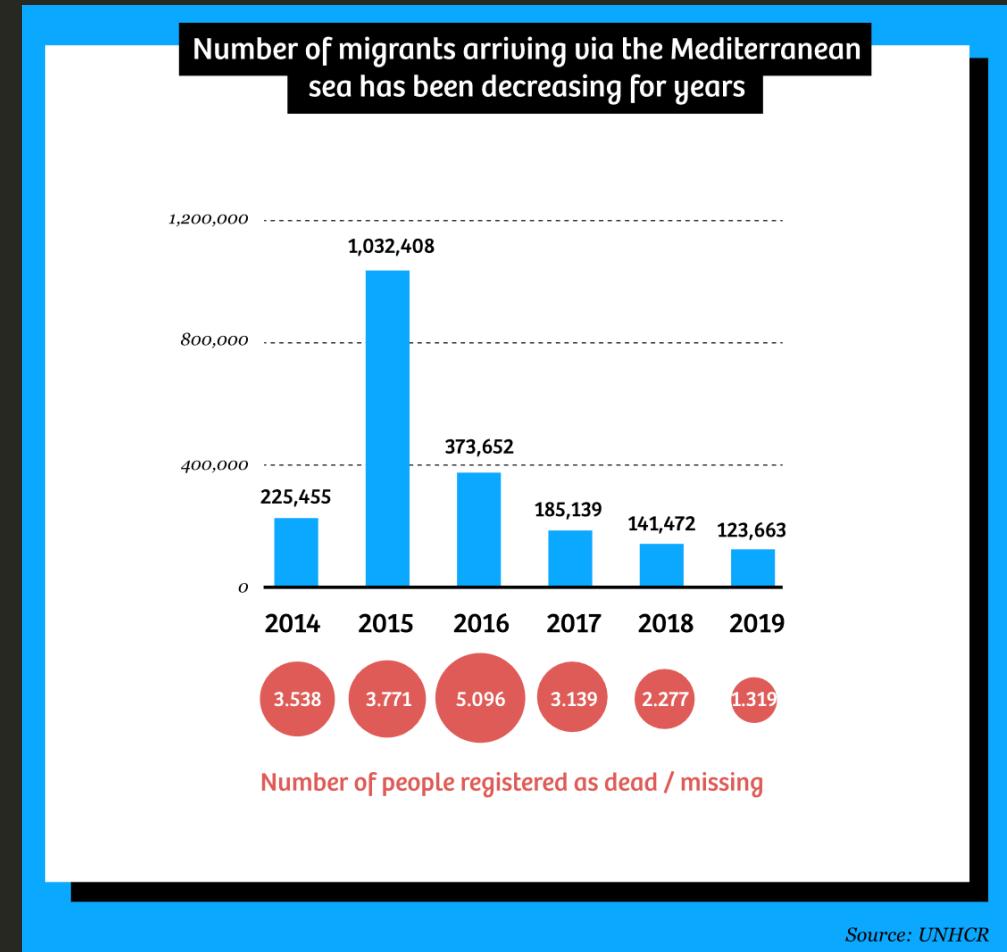
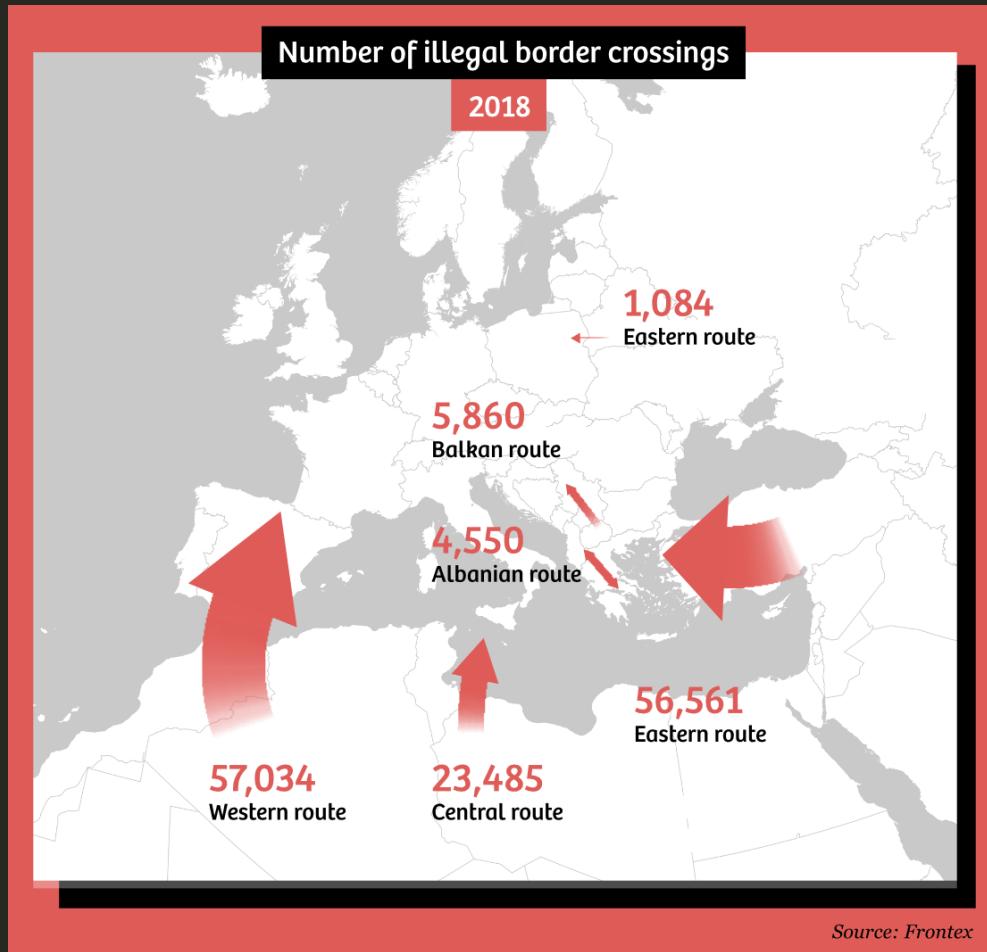
#TidyTuesday contribution Week 2020/31



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

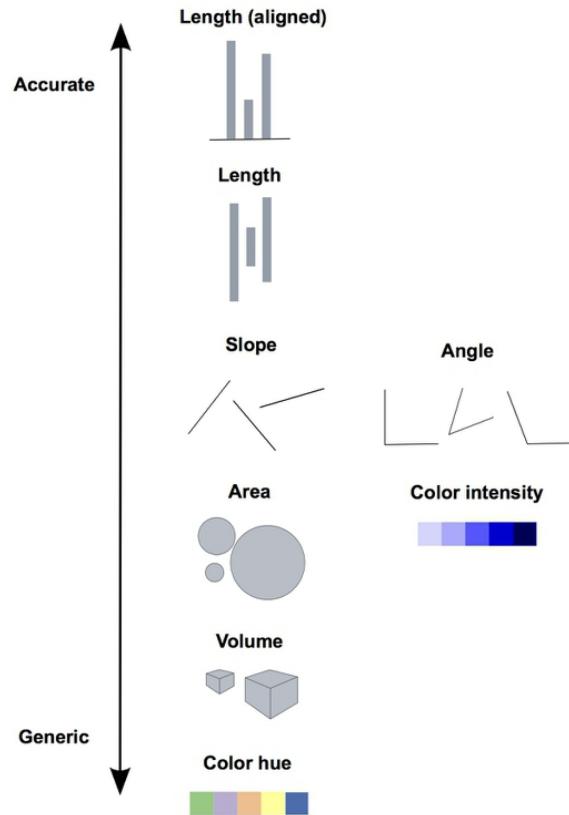


How maps in the media make us more negative about migrants by Maite Vermeulen, Leon de Korte & Henk van Houtum

GOAL

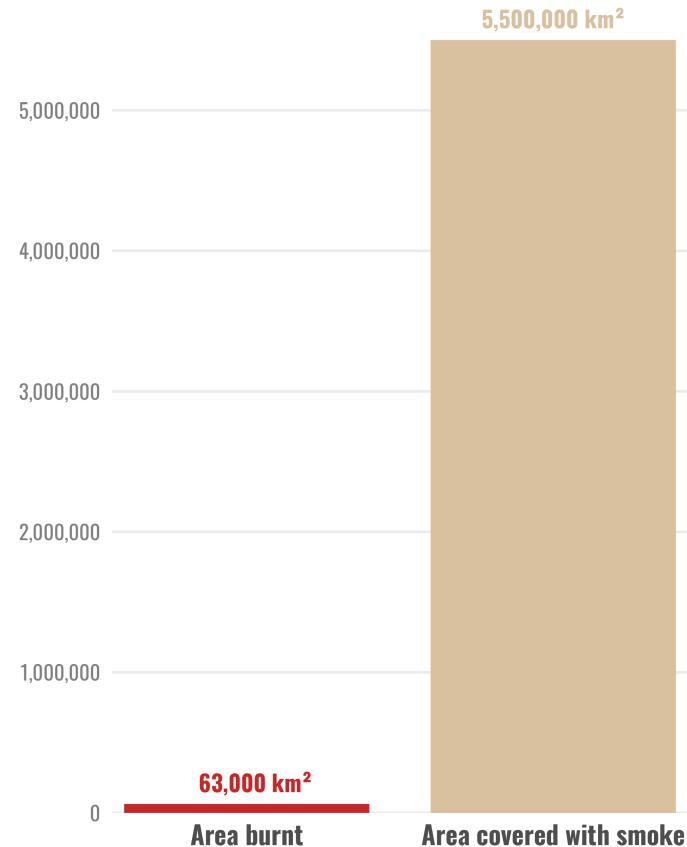
Select charts that successfully transport your story

Data visualizations map values into quantifiable features (aesthetics)

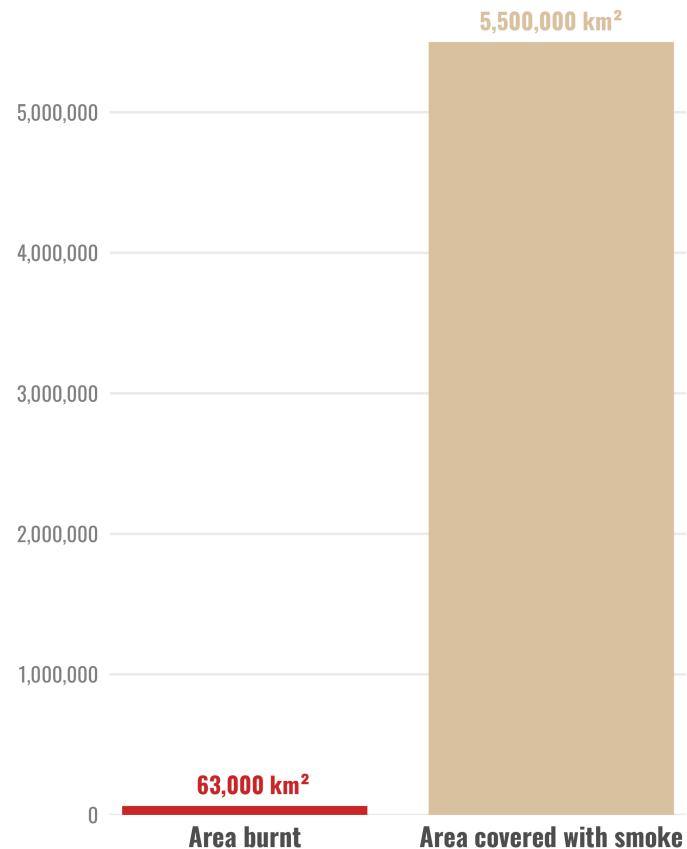


Peter Aldhous based on experiments by William Cleveland and Robert McGill

**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)



**Burnt Land and Plume of Smoke Caused
by the Australian Bushfires in 2019/20
in Comparison to European Countries**

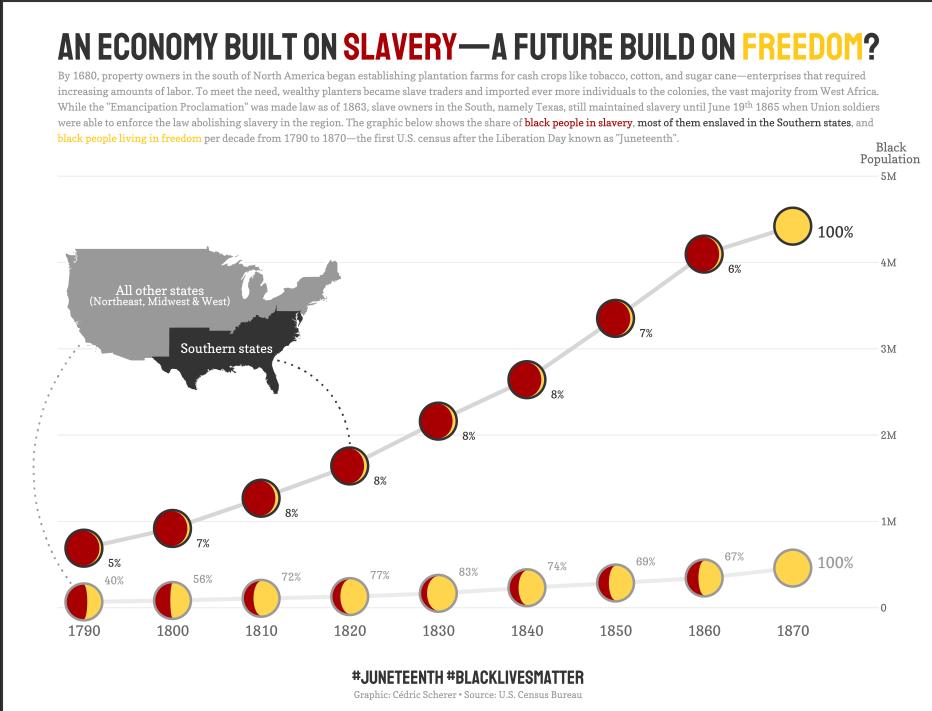
An area of bush, forest and parks larger than Latvia have been burned during a series of massive bushfires across Australia. Thick smoke covering about the area of Denmark (incl. Greenland and the Faroe Islands), France, Spain, Sweden, Norway, Germany, Finland, Poland & Italy has started drifting over the Pacific Ocean towards New Zealand.

Status: 6th of January 2020

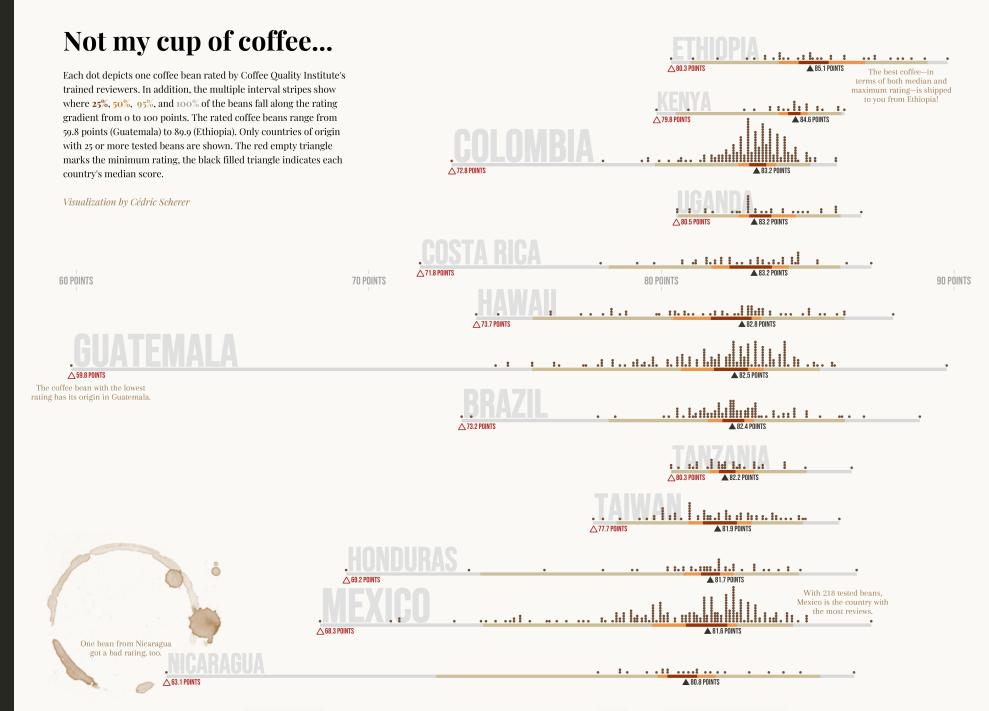
Visualization by Cédric Scherer - Data by NASA FIRMS & The Independent



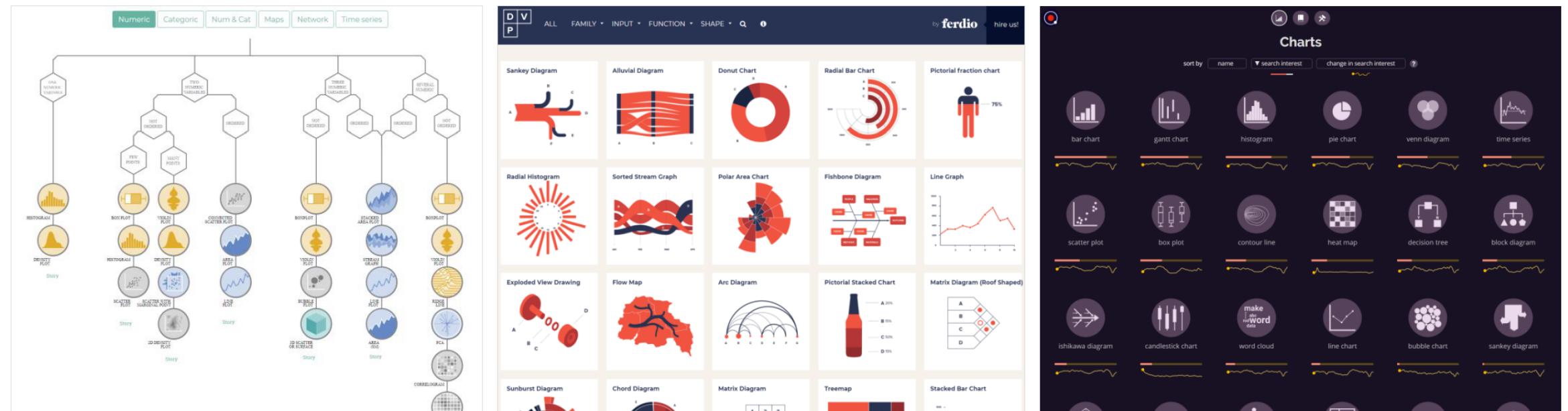
#TidyTuesday contribution Week 2020/02



#TidyTuesday contribution Week 2020/25



#TidyTuesday contribution Week 2020/28



data-to-viz.com

datavizproject.com

visualizationuniverse.com/charts

VISUAL FORM

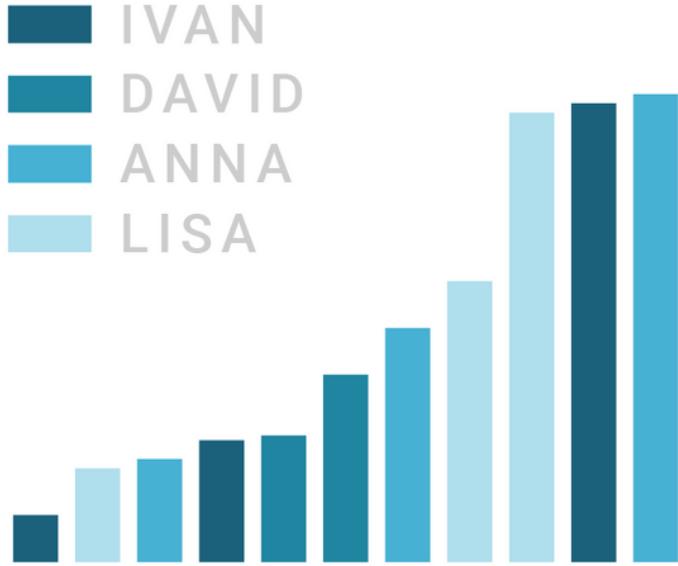
*Follow design rules and data visualization
principles*

Colors

and Common Pitfalls

Color Choice & Accessibility

NOT IDEAL



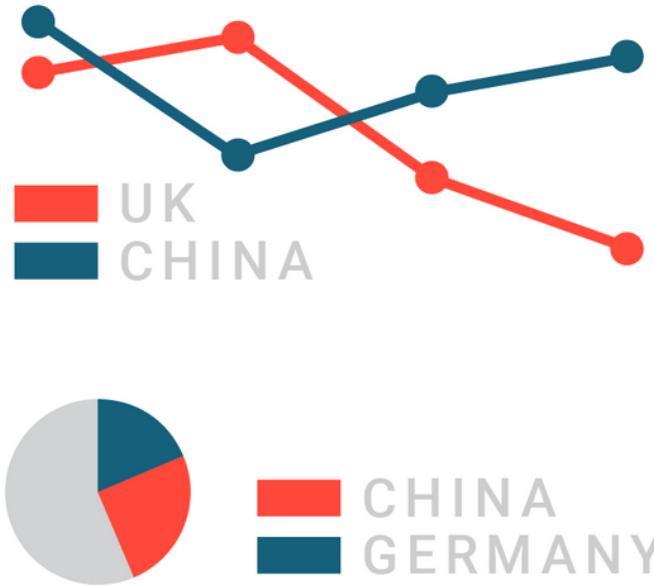
BETTER



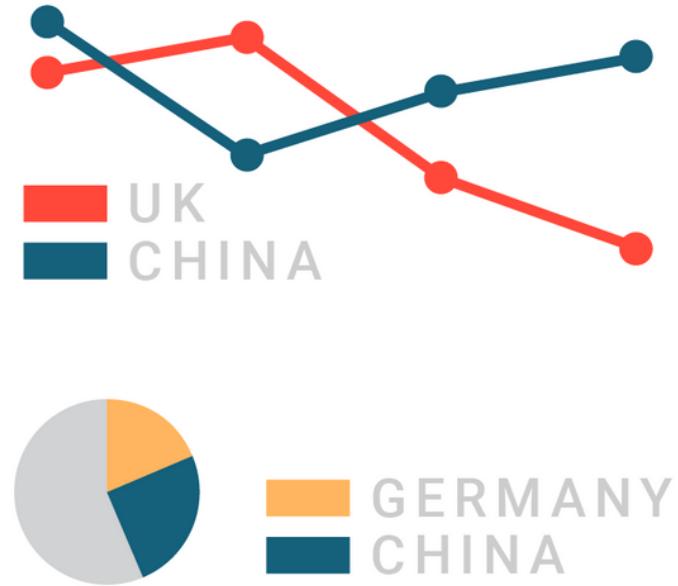
What to consider when choosing colors for data visualization by Lisa Charlotte Rost/DataWrapper

Color Choice & Accessibility

NOT IDEAL

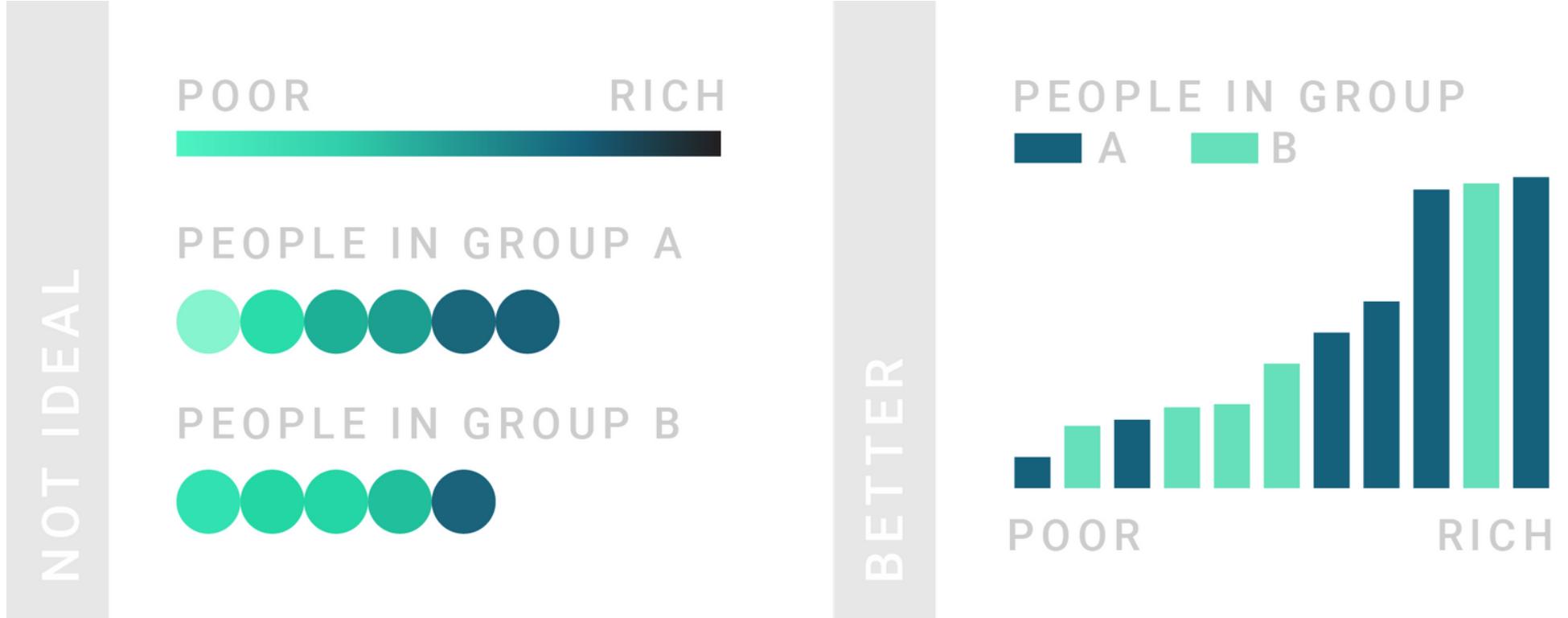


BETTER



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

Rainbow Color Map (Still) Considered Harmful

Publisher: IEEE

2 Author(s)

David Borland ; Russell M. Taylor li View All Authors

172
Paper
Citations

3
Patent
Citations

9091
Full
Text Views



Medical Physics

Current Issue | Authors | Submissions | Advertise | Search

Med Phys. 2015 Jun; 42(6): 2942–2954.

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

PMCID: PMC5148121

PMID: [26127048](https://pubmed.ncbi.nlm.nih.gov/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](#)

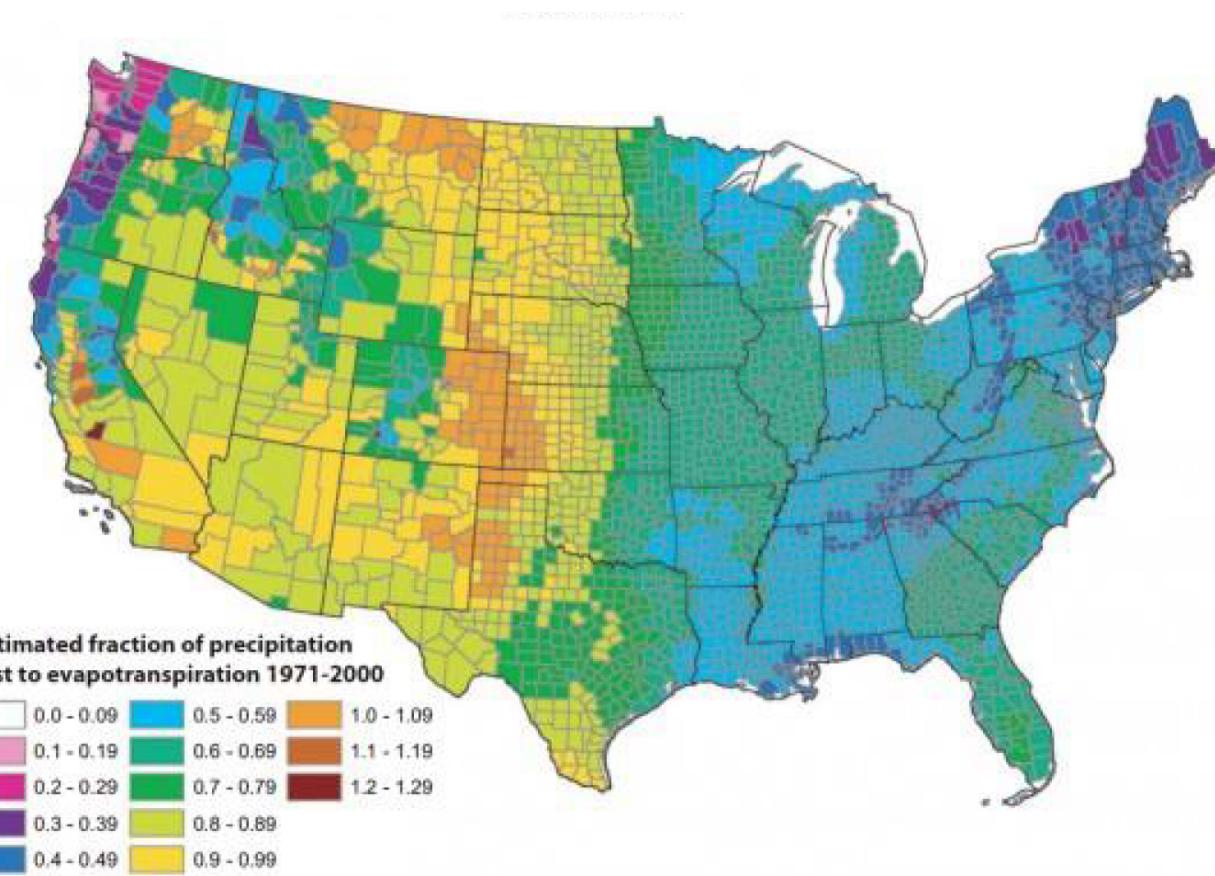


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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

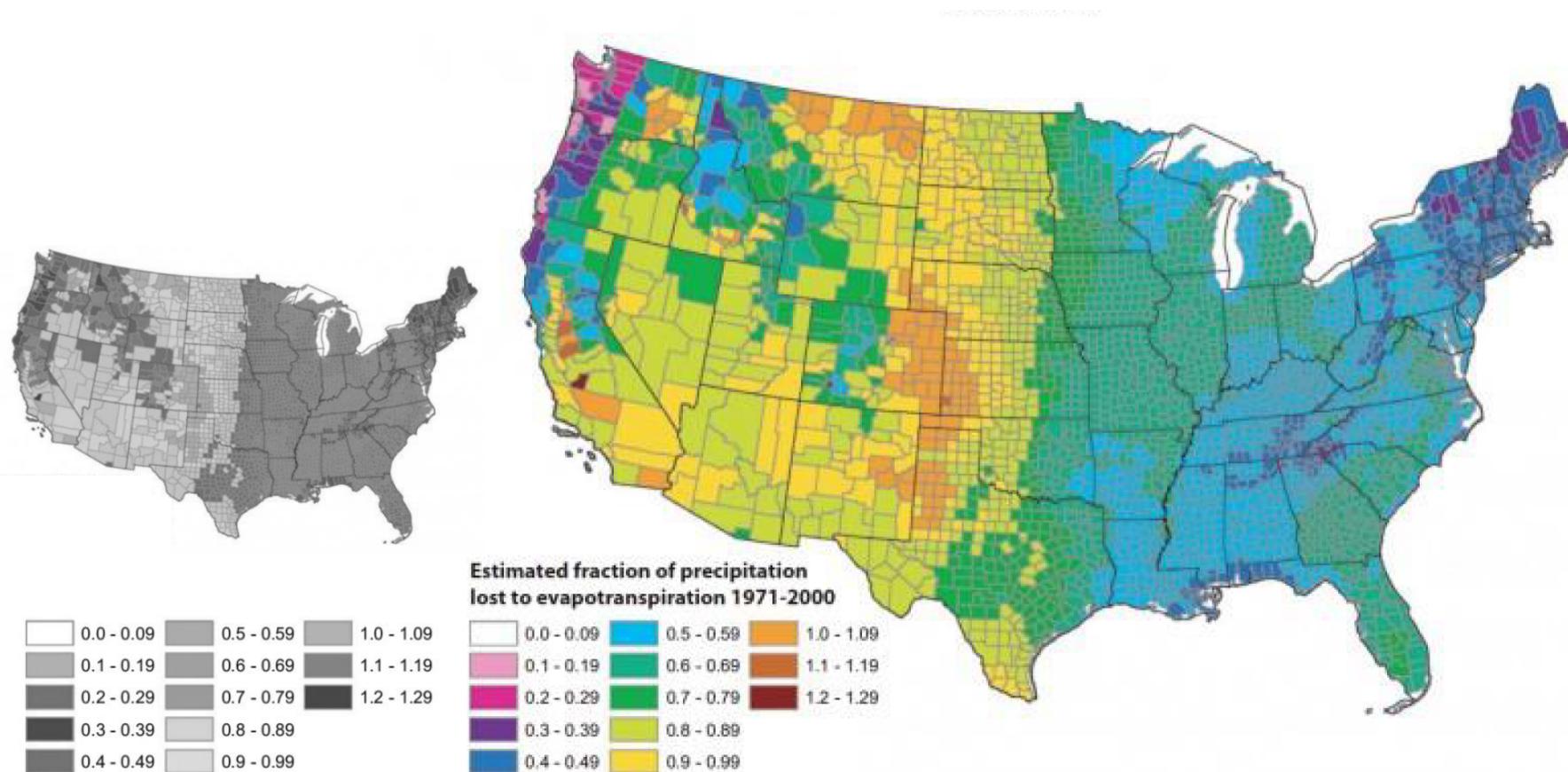


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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

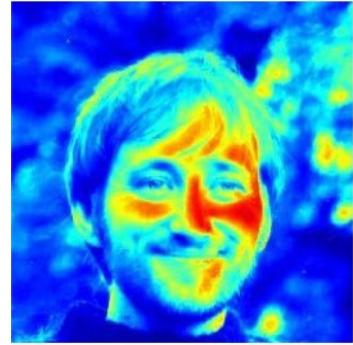


true-colour Phil

Source: fabiocramerich/batlow



true-colour Phil



rainbow Phil
is distorted



batlow Phil
is flawless

Source: fabiocramerich/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 Colorgorical

Use ColorBrewer

EDIT

7 Colors

1 ● #ffd700	<input type="button" value="x"/>
2 ● #ffb14e	<input type="button" value="x"/>
3 ● #fa8775	<input type="button" value="x"/>
4 ● #eaaf94	<input type="button" value="x"/>
5 ● #cd34b5	<input type="button" value="x"/>
6 ● #9d02d7	<input type="button" value="x"/>
7 ● #0000ff	<input type="button" value="x"/>

hex rgb
 hsl

GET

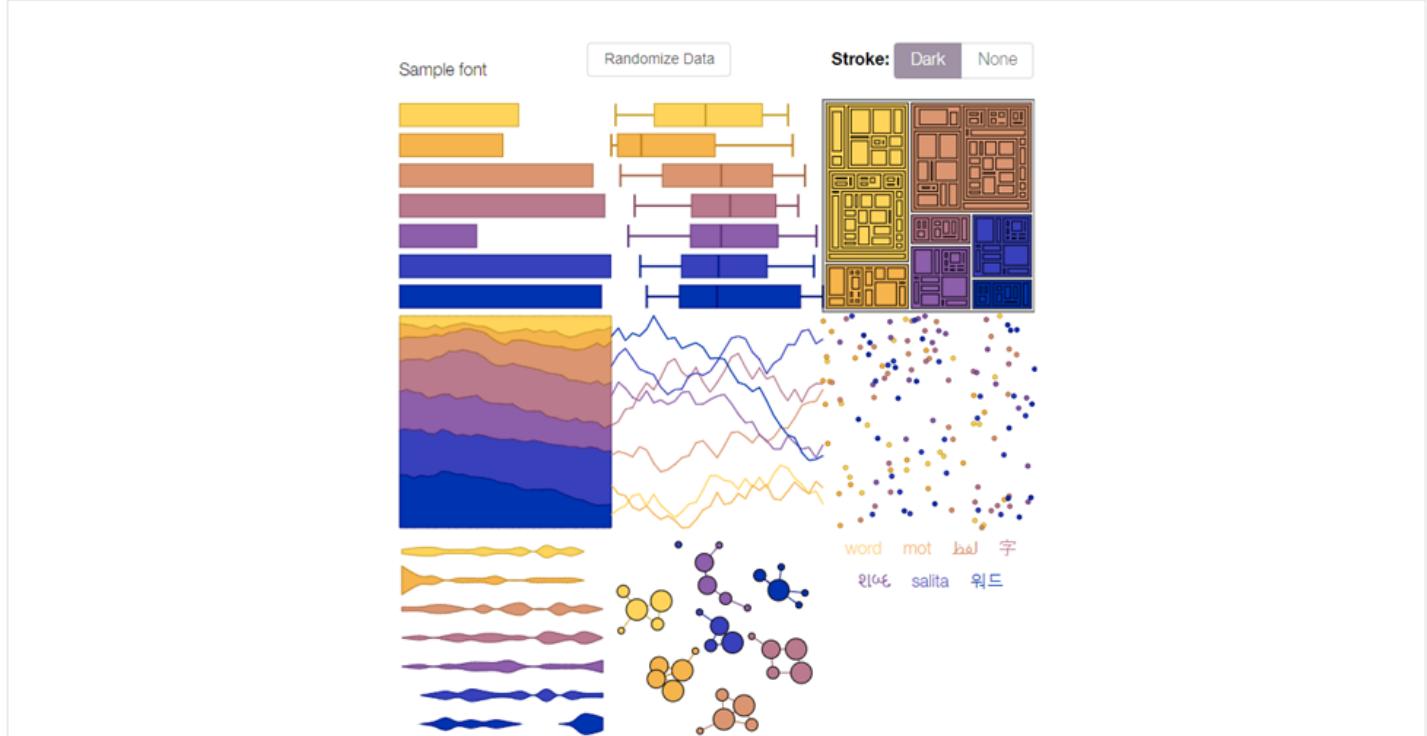
String quotes
 Object with metadata

```
[ "#ffd700",
  "#ffb14e",
  "#fa8775",
  "#eaaf94",
  "#cd34b5",
  "#9d02d7",
  "#0000ff" ]
```

COLORS IN ACTION

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

Sample font Stroke: Dark None



The image is a dense collage of typography-related text elements on a solid black background. At the top, the word 'TYPOGRAPHY' is repeated in large, white, serif capital letters. Below this, the word 'TYPOGRAPHY' appears again in a larger, bold, sans-serif font. To the left, the word 'TYPOGRAPHY' is written vertically in a smaller, white, serif font. In the center, there is a large, bold, yellow sans-serif word 'Typography'. Above this central word, there is a row of decorative, ornate letterforms, including 'C', 'H', 'D', 'E', 'B', 'J', 'K', 'F', and 'H'. To the right of these decorative letters, the word 'TYPOGRAPHY' is repeated in a small, white, serif font. The overall composition is a study of different typefaces and their variations.

Typography

The Choice of the Font(s)

- The font(s) should fit the topic and audience - context matters.
- Avoid fancy fonts and squiggle 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 squiggle 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 **1I1** Test

1II Calibri

1II Bitter

1II Open Sans

1II Monda

1II Roboto

1II Chivo

1II Avenir Next Condensed

1II Fira Sans

1II Lato

1II Noto Sans

1II Oswald

1II Bahnschrift

Tabular (Monospaced) Numbers

32154

Montserrat
proportional numbers

32154

Open Sans
tabular numbers

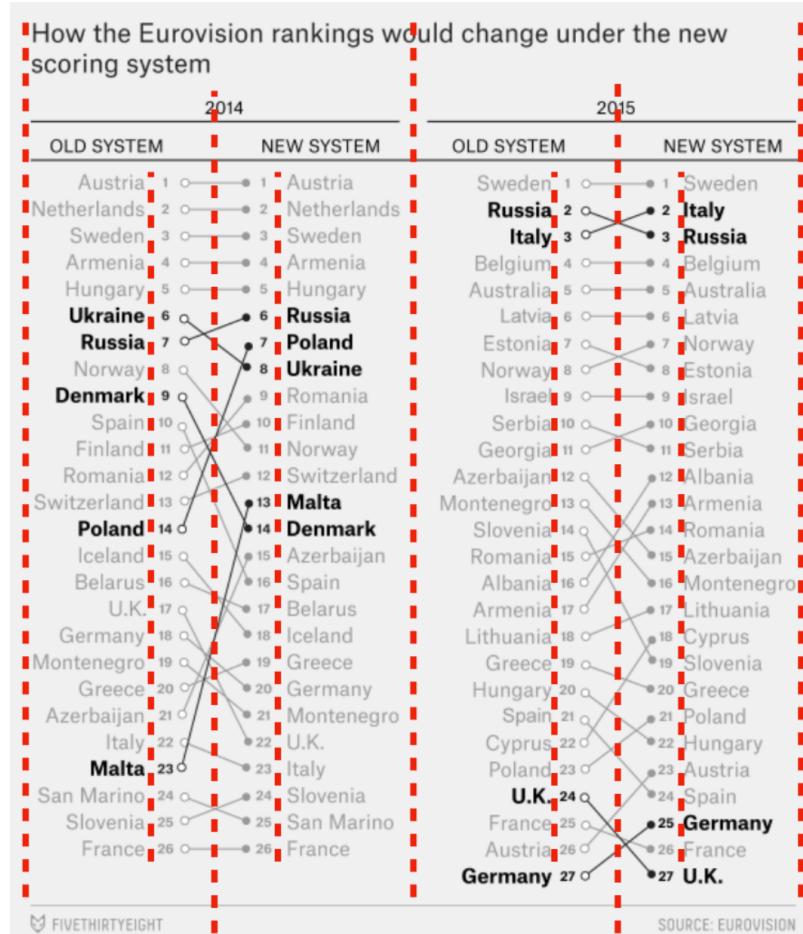
32154

Lato
tabular numbers

Source: Tiffany France on Medium.com

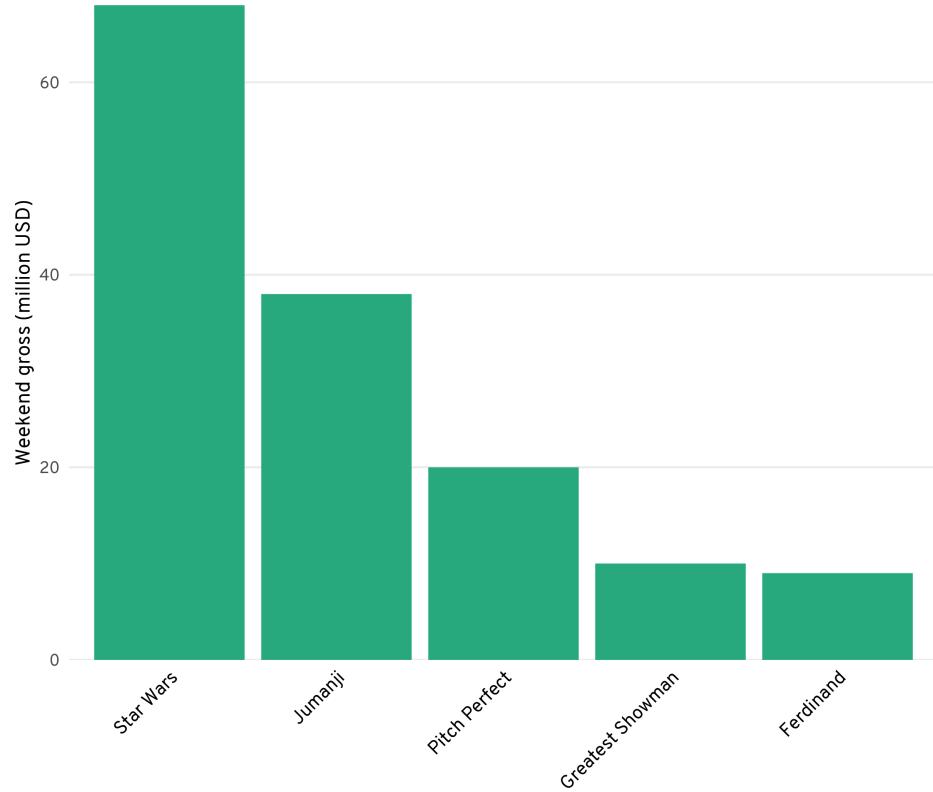
Allign Your Text

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

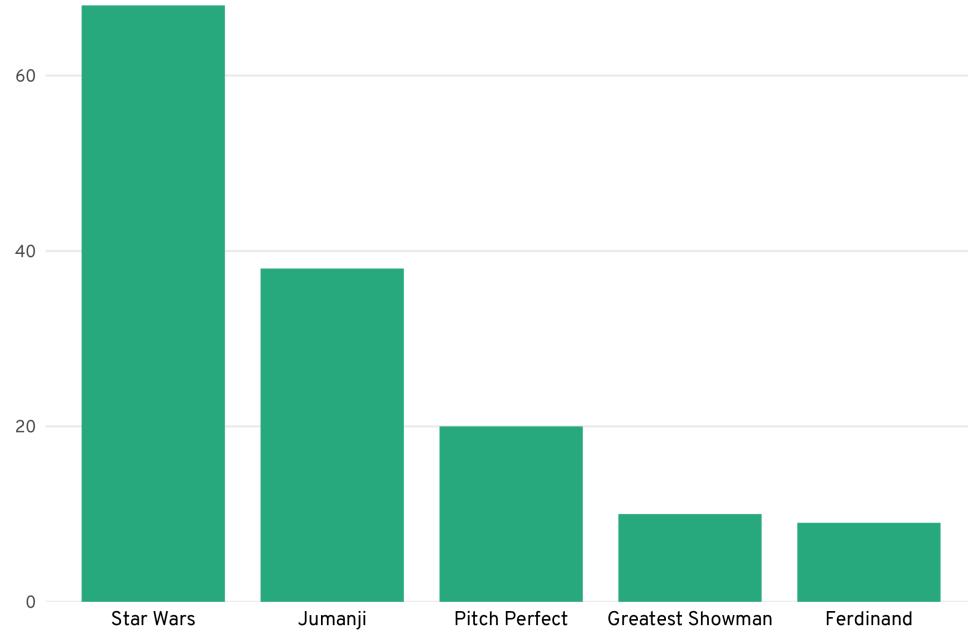


Source: Will R. Chase

(Don't) Rotate Your Text

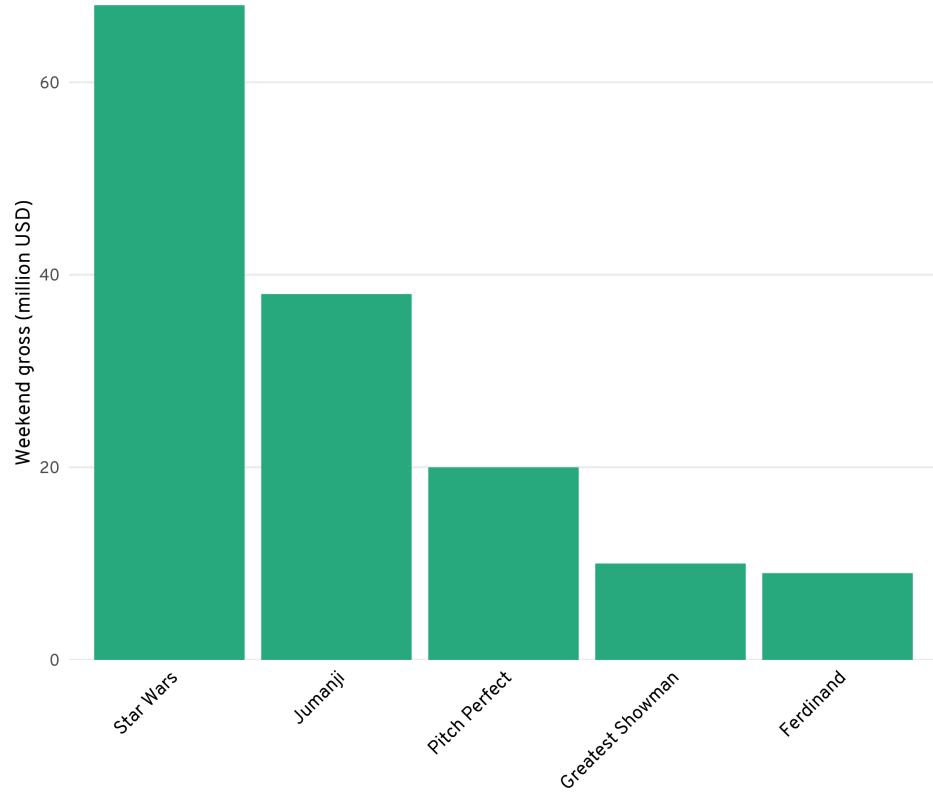


Weekend gross in million USD of popular blockbusters

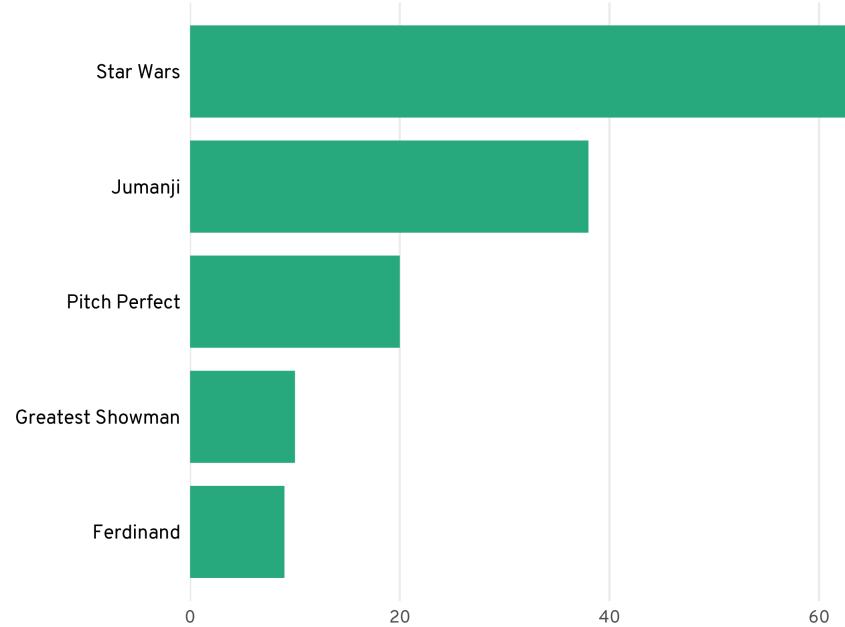


Modified from *Fundamentals of Data Visualization* by Claus Wilke

(Don't) Rotate Your Text

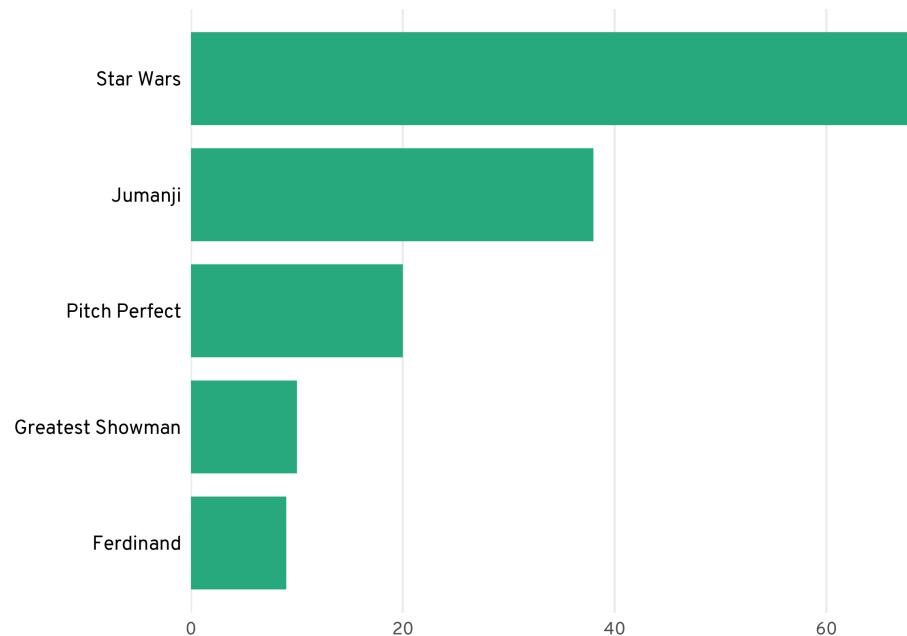


Weekend gross in million USD of popular blockbusters

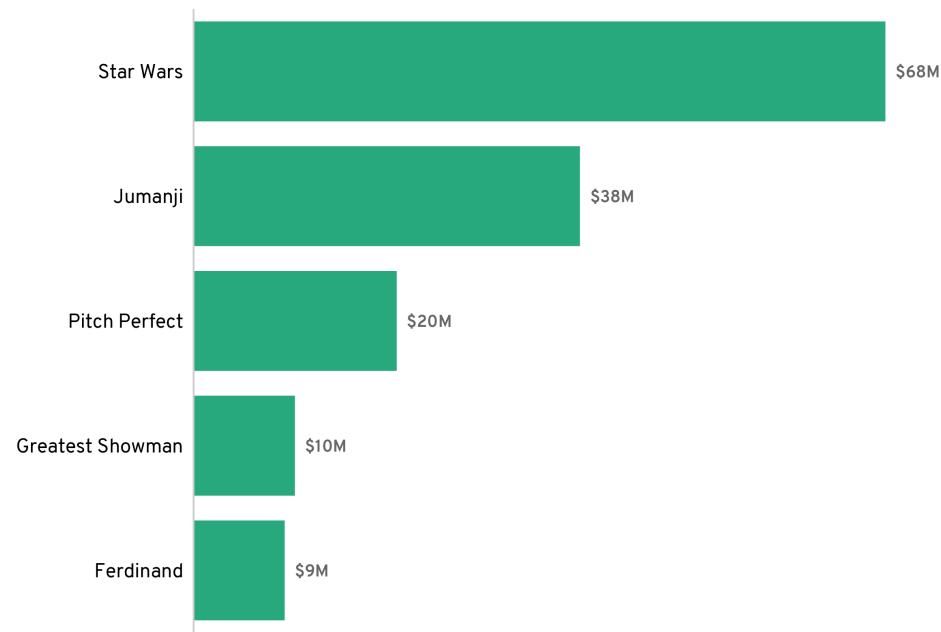


Use Annotations

Weekend gross in million USD of popular blockbusters

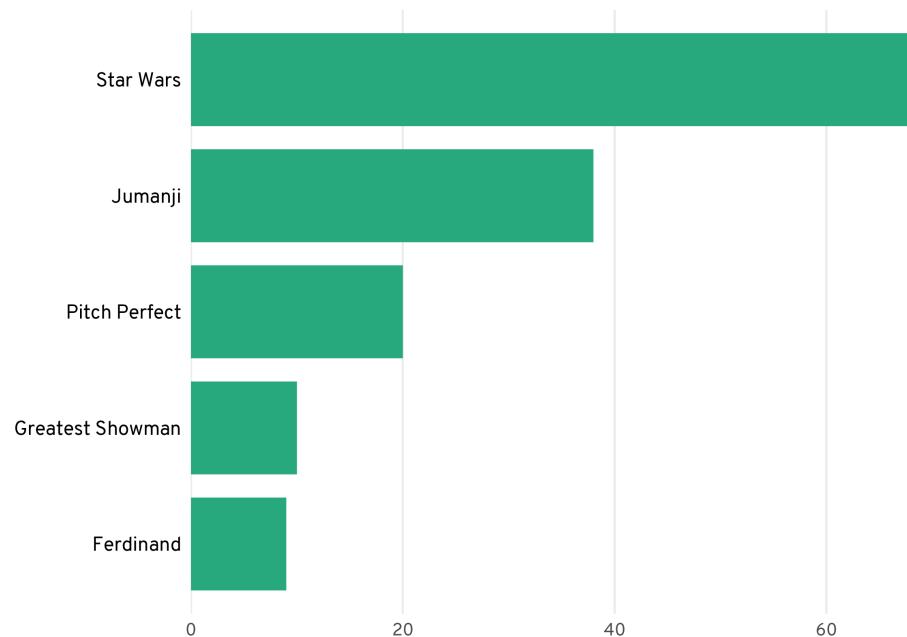


Weekend gross in million USD of popular blockbusters

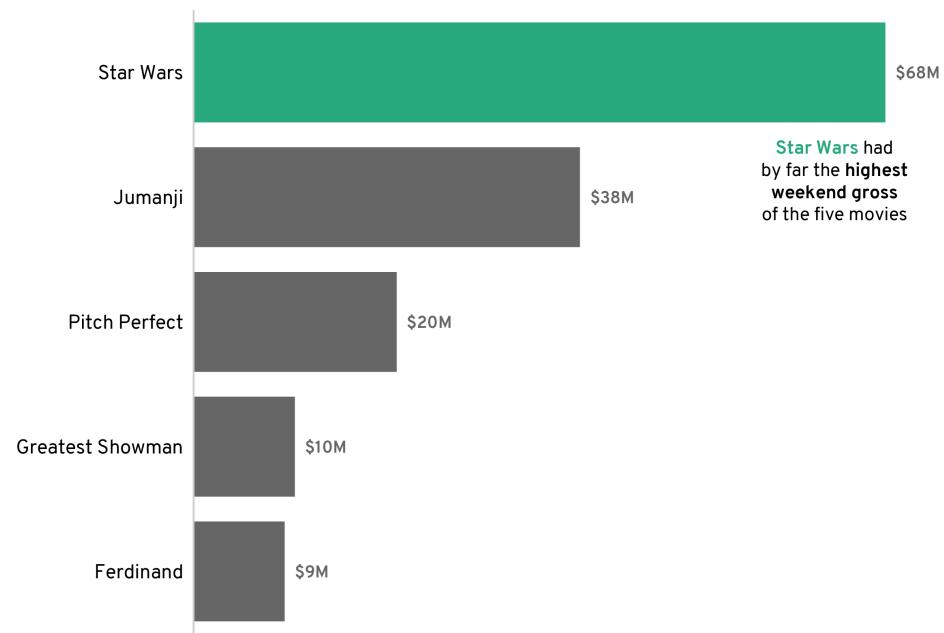


Use Annotations

Weekend gross in million USD of popular blockbusters



Weekend gross in million USD of popular blockbusters



Cédric Scherer
Z3tt
Computational Ecologist & Data Visualization Designer

Highlights

- Arctic Code Vault Contributor
- RIOT

Organizations

- Z3tt

Pinned

- TidyTuesday
- responsiblyMLR
- ggplot-courses
- DataViz-Teaching
- 30DayMapChallenge
- Corona-Worst-Days

1,879 contributions in the last year

Contribution settings: 2019

Learn more about contributions.

Contributors: @davidmccall, @FrontpageData, @CedricScherer, More

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, 2018

ABOUT ME

Computational Ecologist • Data Visualization Designer • Proud Dad

[Email](mailto:cedricscherer.netlify.app) [Twitter](https://twitter.com/cedricscherer) [GitHub](https://github.com/Z3tt) [LinkedIn](https://linkedin.com/in/cedricscherer)

FEATURED TAGS

- BERLIN
- CORRELATION
- TIDYTUESDAY
- ANIMATIONS
- GGPLOT2
- MAPS
- GGVIS
- 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

BLOG ABOUT ME PUBLICATIONS VISUALIZATIONS LINKS

[DATAVIS](#) [TUTORIAL](#) [B](#) [TINYVERSE](#) [GGPLOT2](#)

A GGPLOT2 TUTORIAL FOR BEAUTIFUL PLOTTING IN R

POSTED BY CÉDRIC ON MONDAY, AUGUST 5, 2019

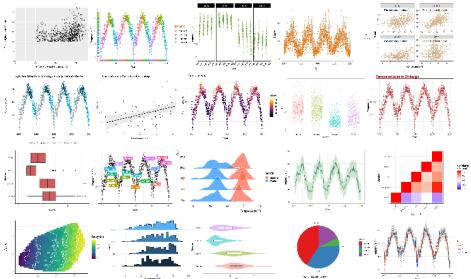
Last update: 2019-11-01

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

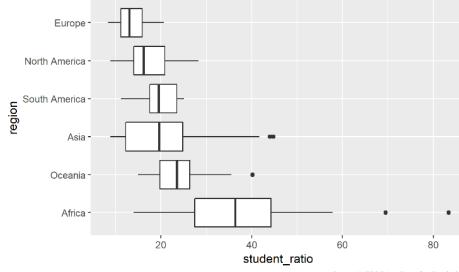
BLOG ABOUT ME PUBLICATIONS VISUALIZATIONS LINKS

[DATAVIS](#) [TUTORIAL](#) [ANIMATIONS](#) [GGPLOT EVOLUTION](#) [B](#) [GGPLOTS](#) [TINYVERSE](#) [TINYTHURSDAY](#)

THE EVOLUTION OF A GGPLOT (EP. 1)

POSTED BY CÉDRIC ON FRIDAY, MAY 17, 2019

The Evolution of a ggplot



Data: UNESCO Institute for Statistics
Visualization by Cédric Scherer

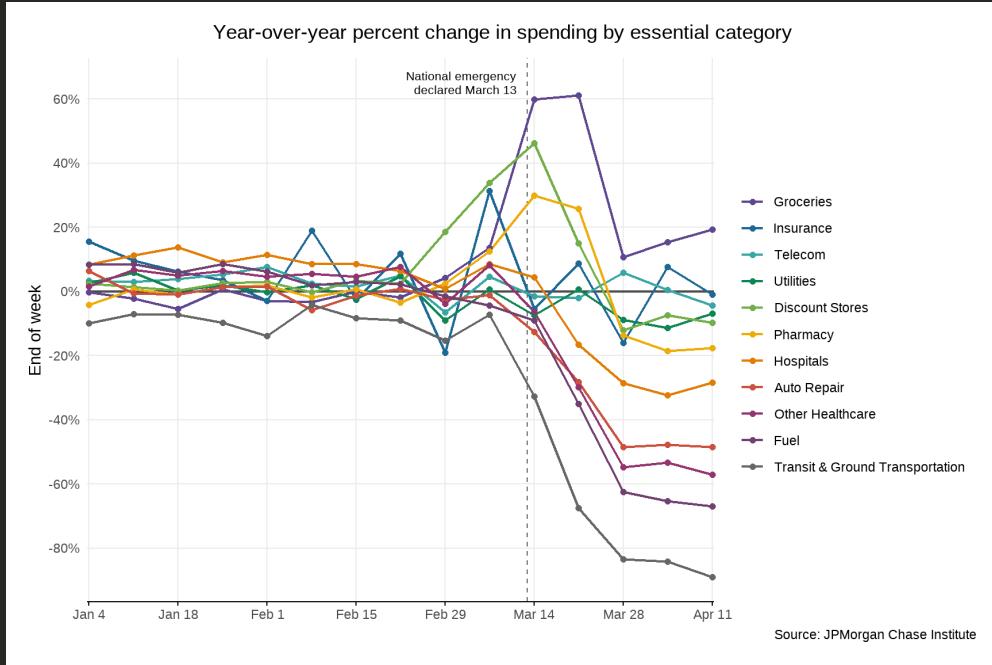
- Aim of this Tutorial
- Data Preparation
- The Default Boxplot
- Sort Your Data!
- Let Your Plot Shine—Get Rid of the Default Settings
- The Choice of the Chart Type
- More Geoms, More Fun, More Info!
- Add Text Boxes to Let The Plot Speak for Itself
- Bonus: Add a Tile Map as Legend
- The Final Evolved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

AIM OF THIS TUTORIAL

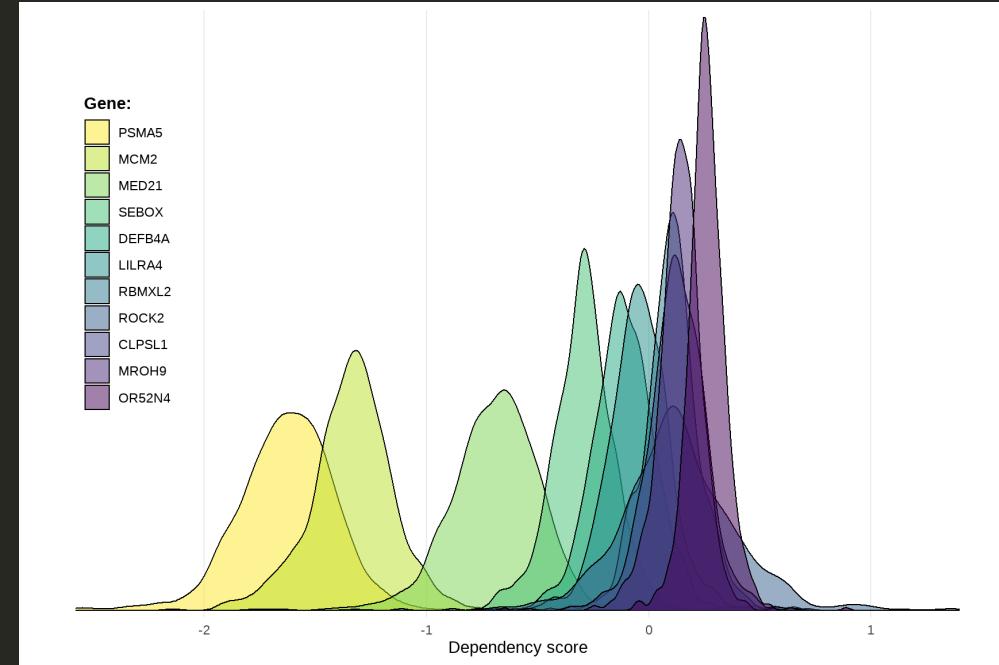
In this series of blog posts, I aim to show you how to turn a default ggplot into a plot that visualizes information in an appealing and easily understandable way. The goal of each blog post is to provide a step-by-step tutorial explaining how my visualization have evolved from a typical basic ggplot. 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