

# Key Concepts in Data Visualization

*How to Design Effective & Engaging Charts*

**Dr. Cédric Scherer**

MSC Data Visualization Training | March 9 and 11 2021

Photo by Richard Strozyński

**HELLO**  
my name is

Cédric

[cedricscherer.com](http://cedricscherer.com)



@CedScherer



@Z3tt

# Scientist by



→ population and community dynamics, movement ecology, wildlife diseases

- **M.Sc. in *Ecology, Evolution & Nature Conservation***

*University of Potsdam 2011–2014*

- **Ph.D. in *Ecology***

*Research Training Group "BioMove" 2015–2019*

- **PostDoc in *Computational Ecology***

*Leibniz Institute for Zoo and Wildlife Research (IZW) 2019–*



# Scientist by



→ population and community dynamics, movement ecology, wildlife diseases

- **M.Sc. in Ecology, Evolution & Nature Conservation**

*University of Potsdam 2011–2014*

- **Ph.D. in Ecology**

*Research Training Group "BioMove" 2015–2019*

- **PostDoc in Computational Ecology**

*Leibniz Institute for Zoo and Wildlife Research (IZW) 2019–*



# DataViz Specialist by



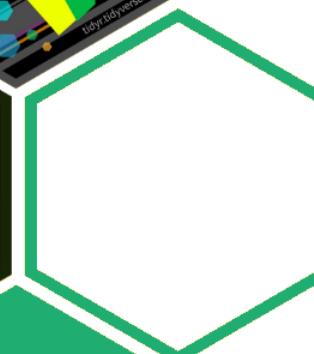
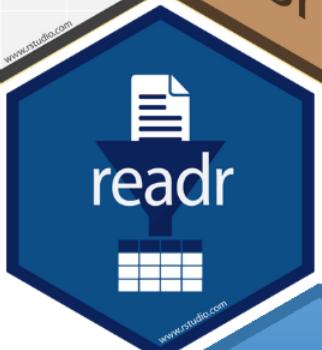
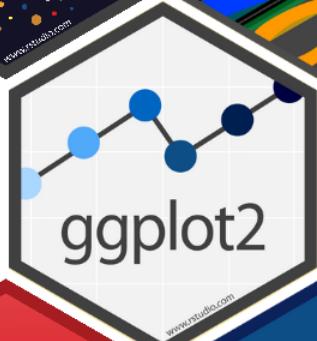
→ analyst, designer, consultant, workshop instructor

- **Freelancing Data Visualization Designer**

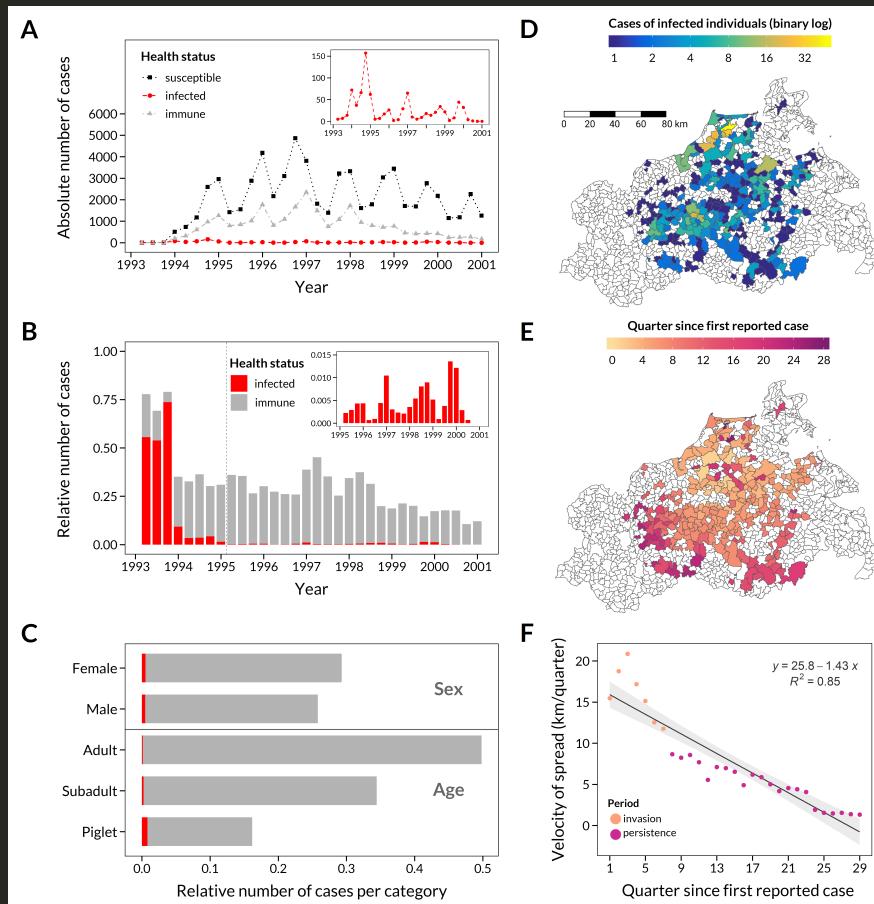
- **Data Challenges and Personal Projects**

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

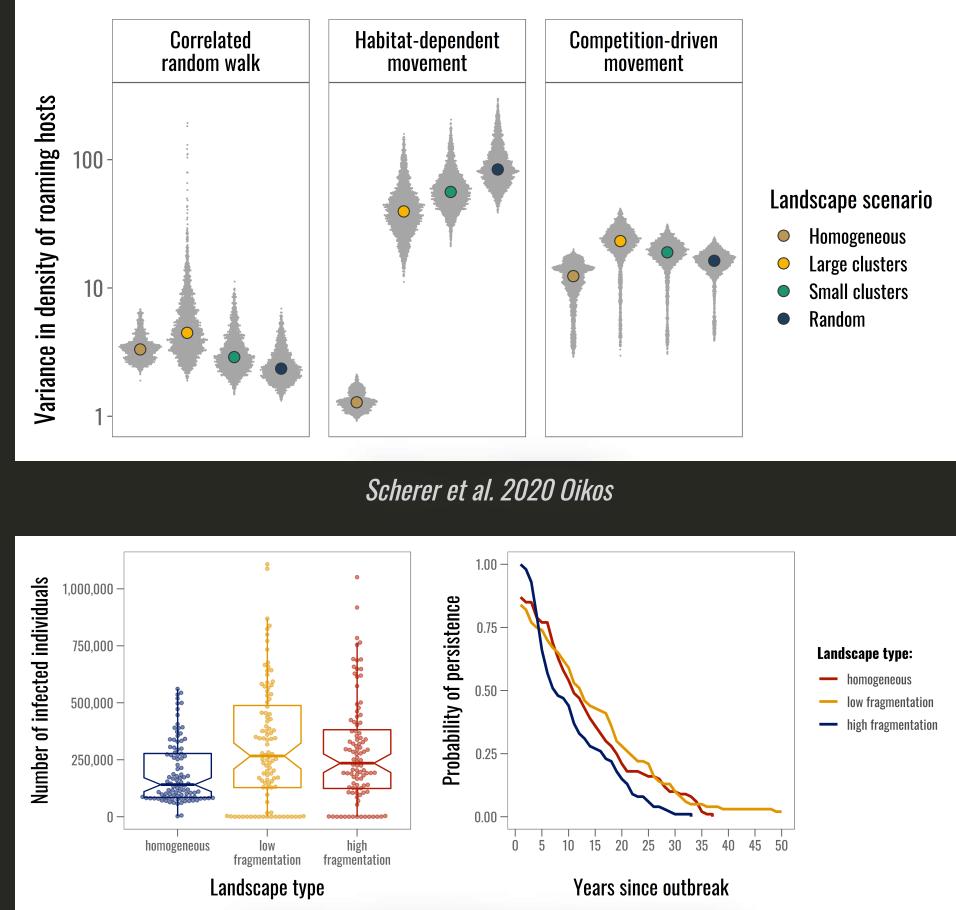
Scientist by 🎓 DataViz Specialist by ❤



# Data Visualizations for Scientific Publications & Talks

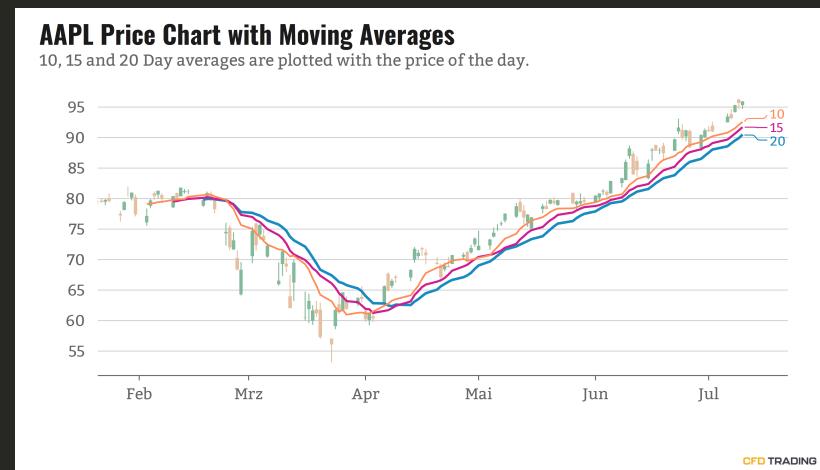
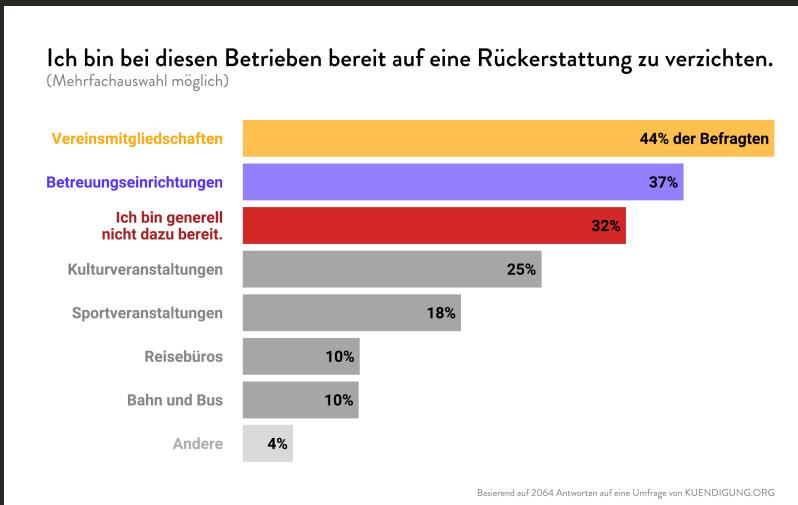
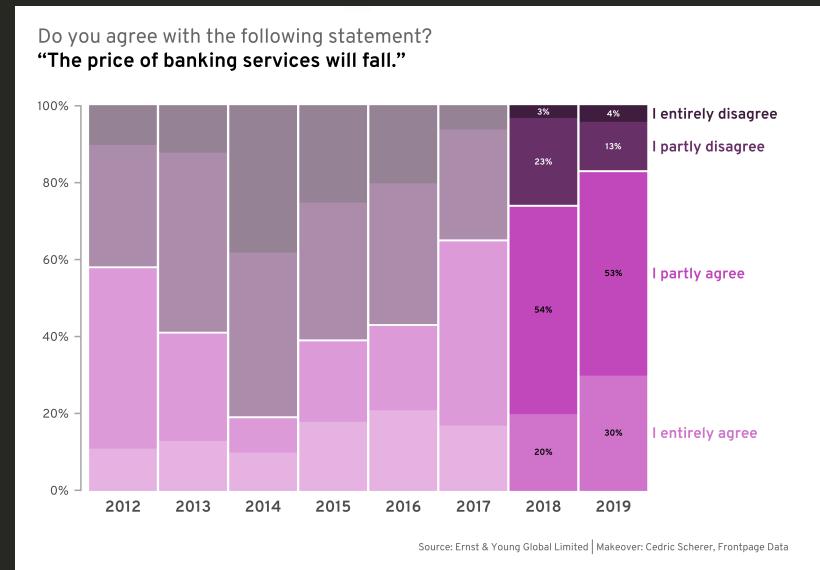
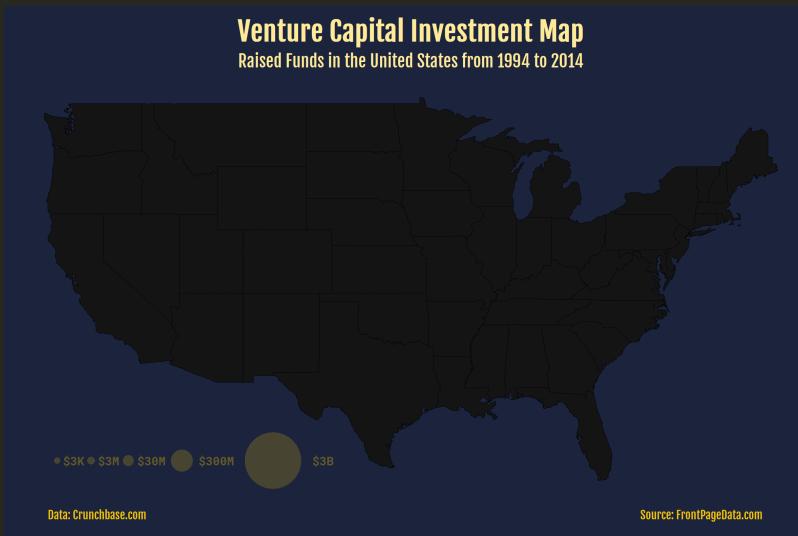


Scherer et al. 2019 *Journal of Animal Ecology*

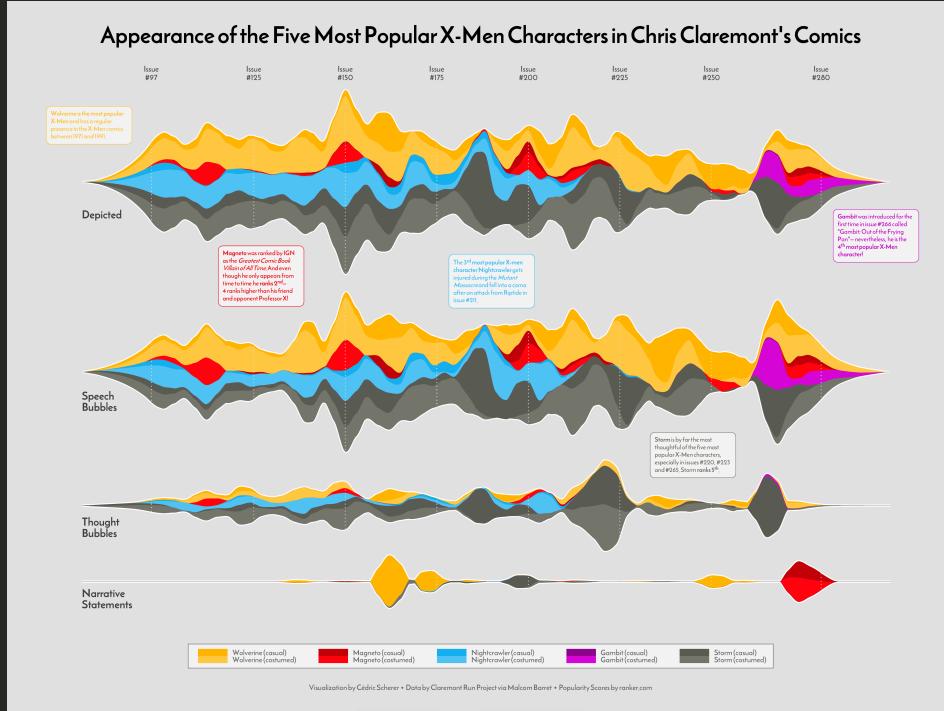


Sciaini et al. 2019 *Methods in Ecology & Evolution*

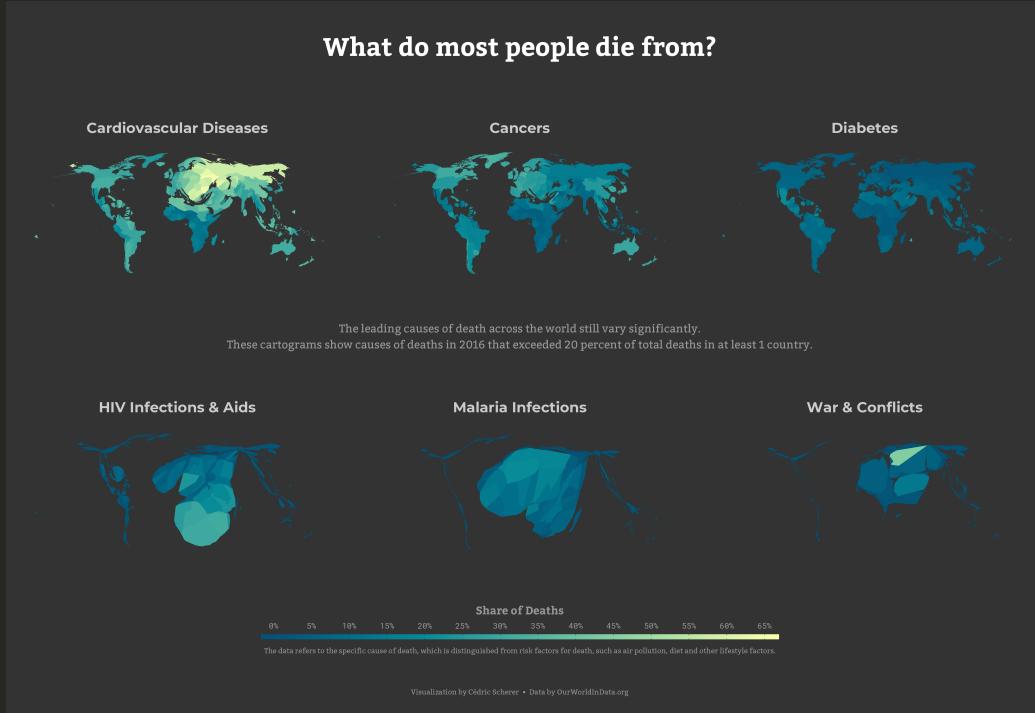
# Data Visualizations for Client Projects



# Data Visualizations as Challenge Contributions



Contribution to #TidyTuesday



Contribution to #30DayMapChallenge

# 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 9, 2020

## ABOUT ME



Computational Ecologist •  
Data Visualization Designer  
• Proud Dad



**HELLO**  
my name is

*Your Turn!*

# Data Visualization

is any graphical representation  
of information and data.

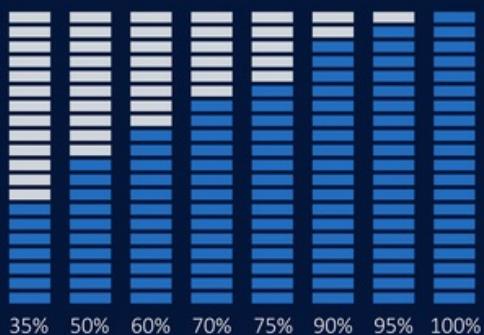
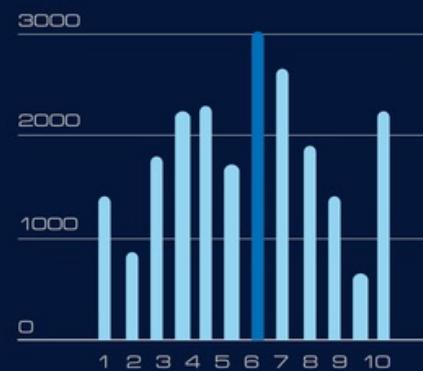
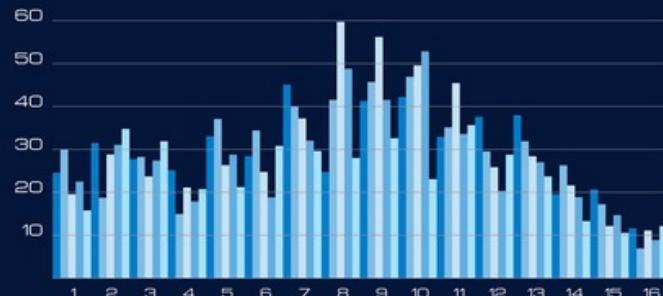
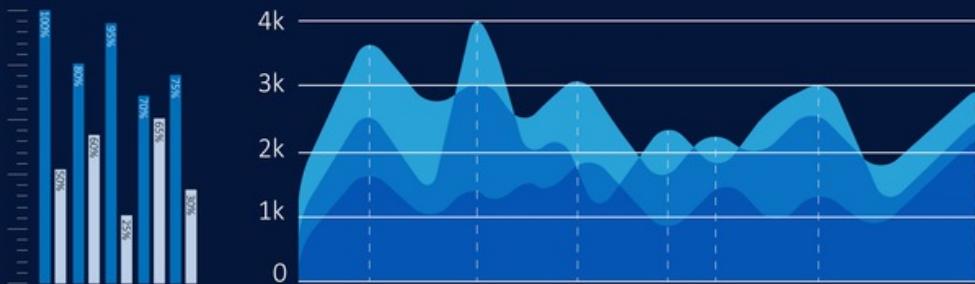
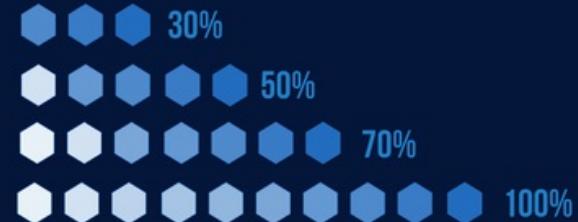
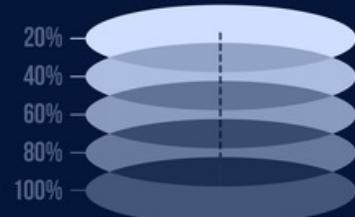
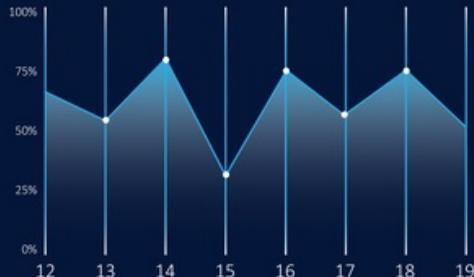
# Data Visualization

helps to amplify cognition, gain insights,  
discover, explain, and make decisions.

# Data Visualization is part art and part science.

*Claus O. Wilke, "Fundamentals of Data Visualization"*





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

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite  
Paris, le 20 Novembre 1869.

Les nombres d'hommes perdus sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à desser la carte ont été puisés dans les ouvrages de M. Chiers, de Séguir, 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 Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mogilow et se rejoignaient vers Osscha et Witelisk, avaient toujours marché avec l'armée.

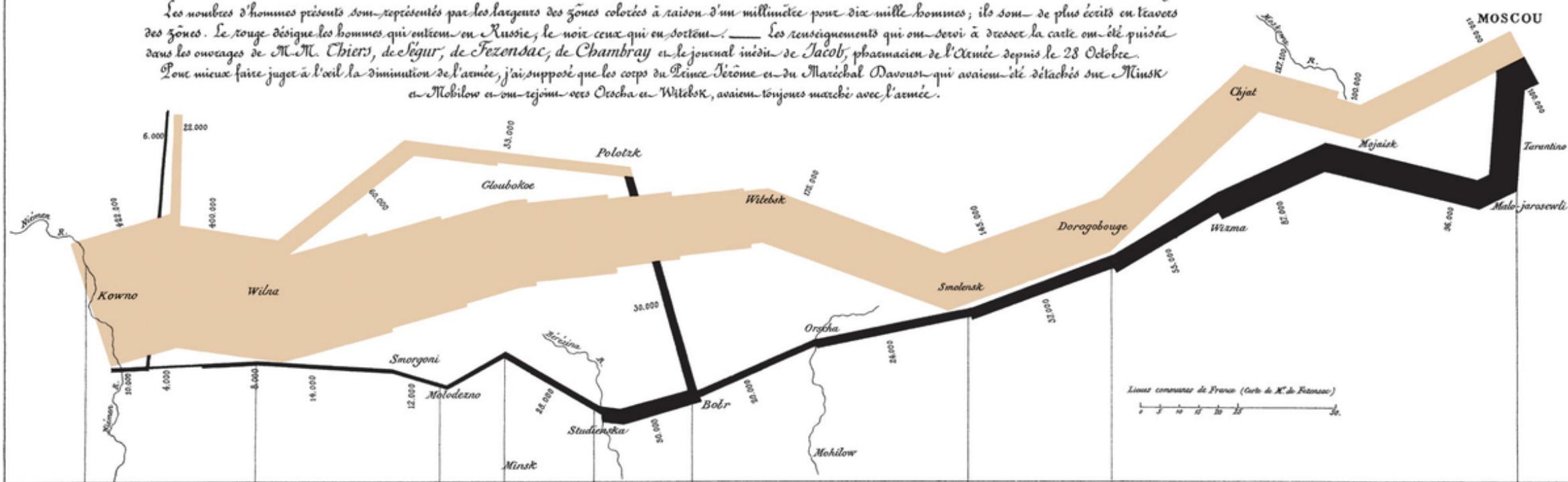
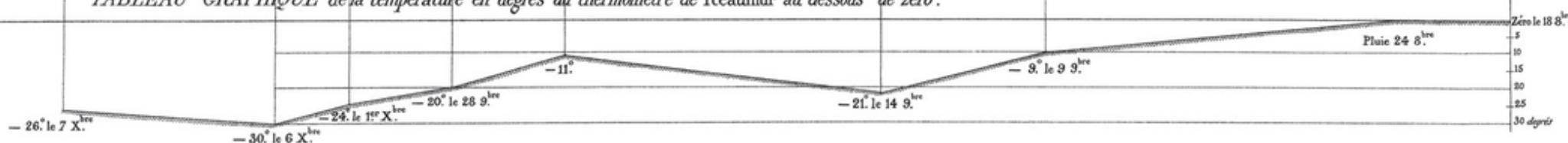
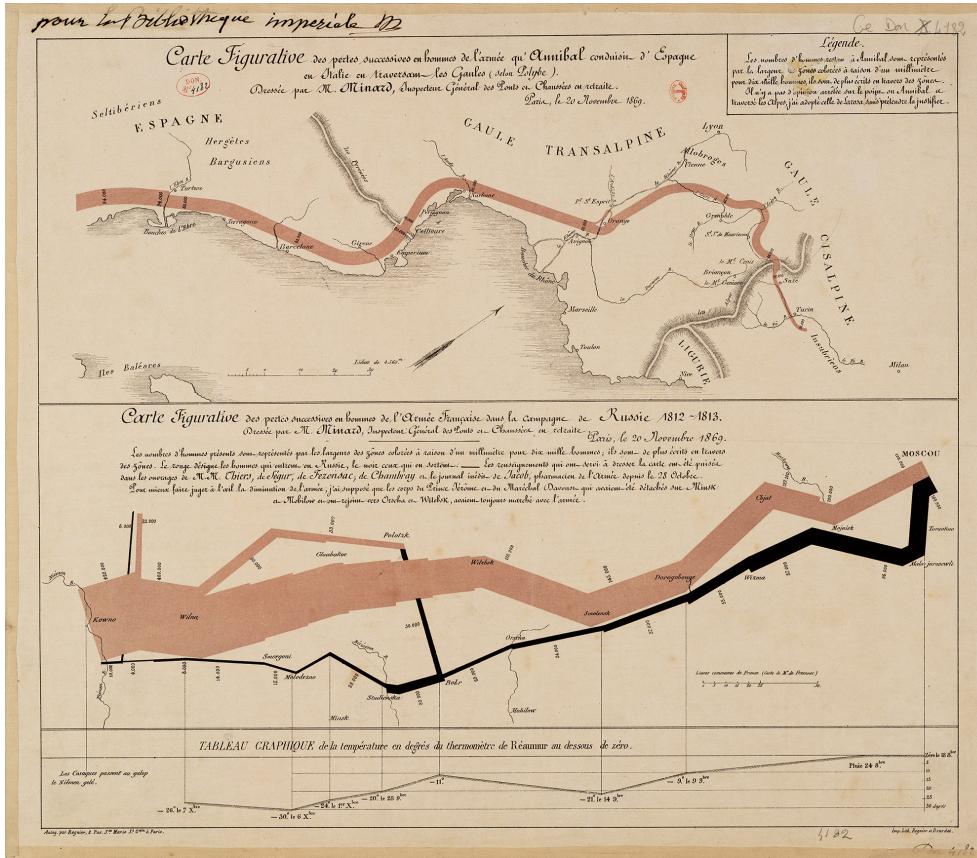


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop  
le Niemen gelé.

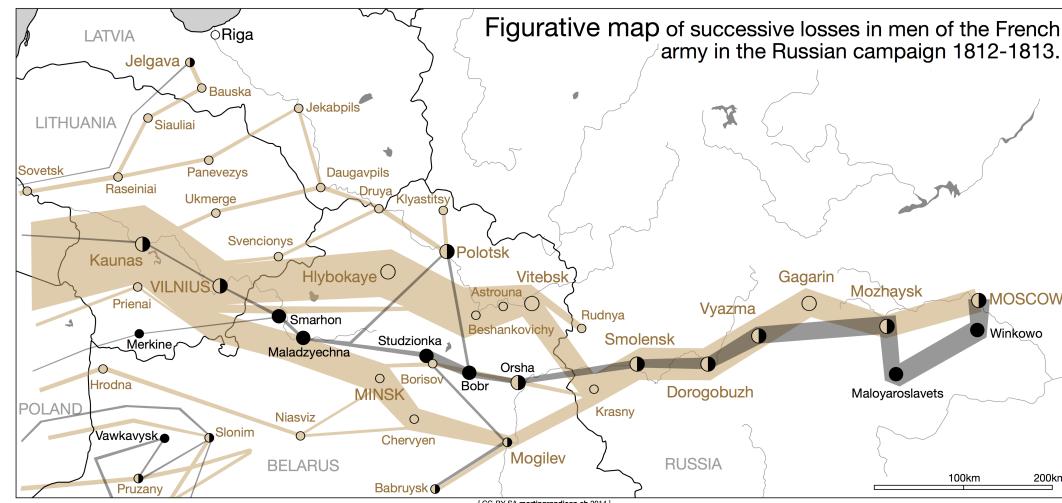
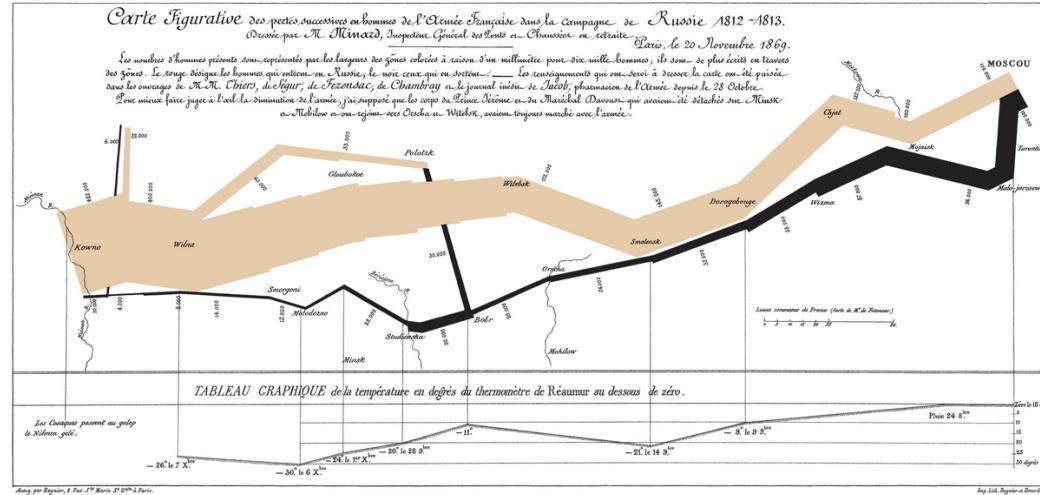




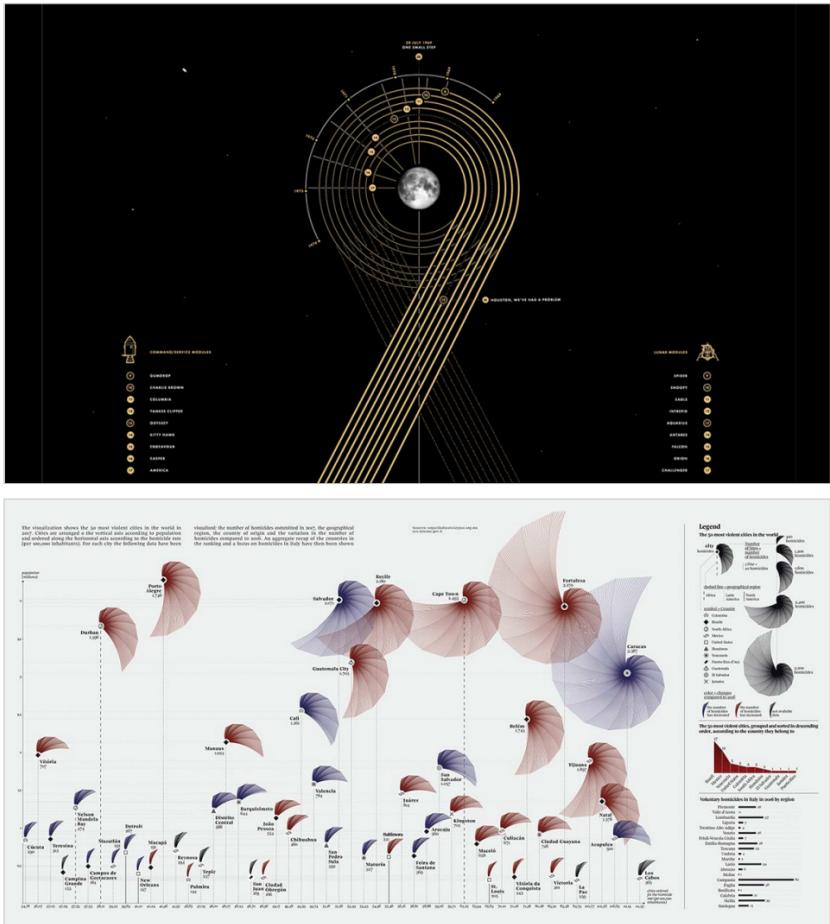
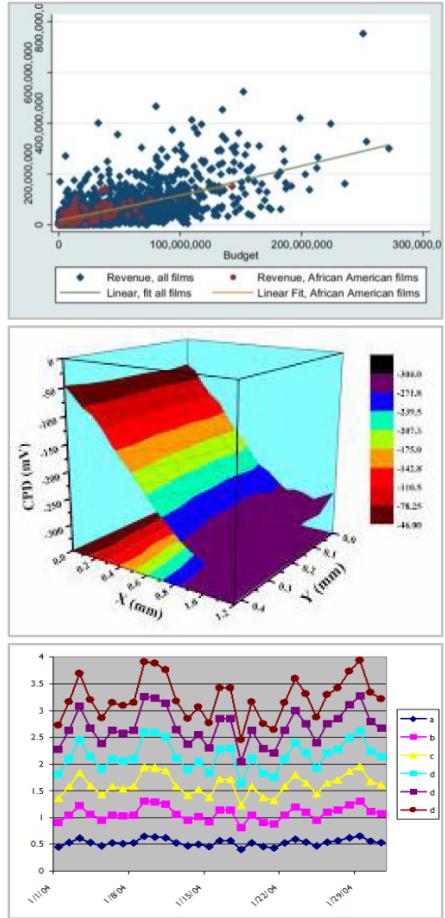
*Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813 and*

*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**



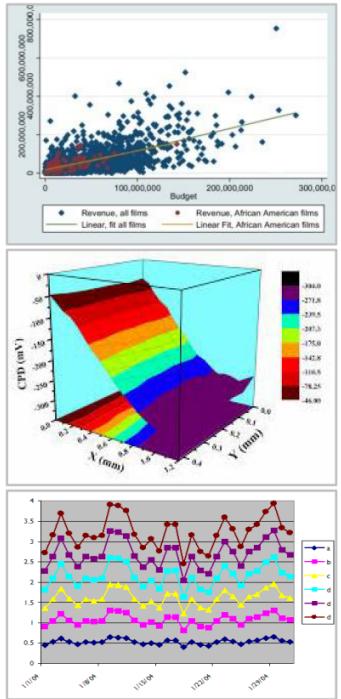
The map created by Charles Joseph Minard projected in the geographical reality with the most accurate information on the actual route of different corps by Martin Grandjean



Anonymous

Sonia Kuijpers

Upper: Paul Button  
Lower: Frederica Fragapane



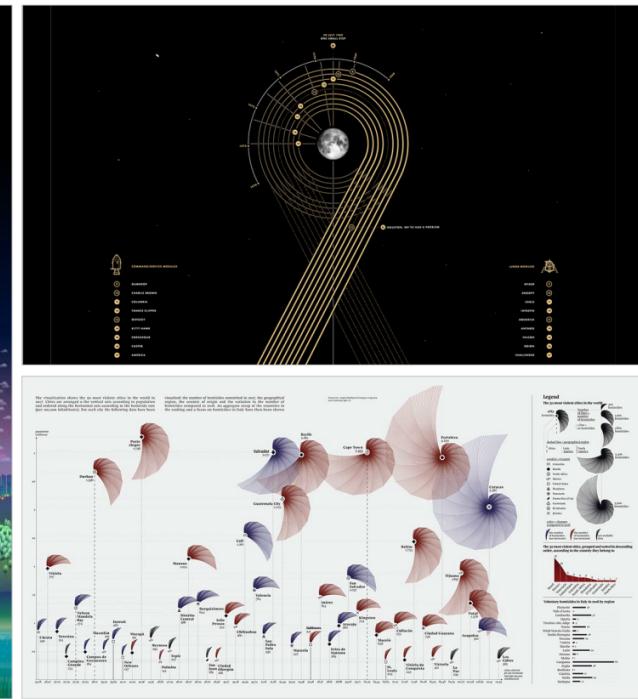
Anonymous

## We aim for DataViz that:

- is informative
- is truthful
- is easy to grasp
- is visually appealing
- draws attention



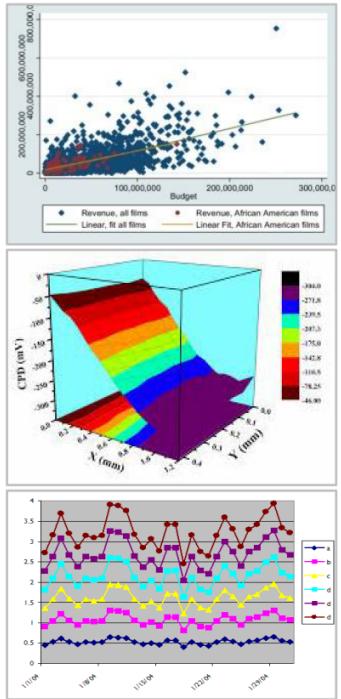
Sonia Kuijpers



Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from uninformative & poorly designed data visualizations to data art



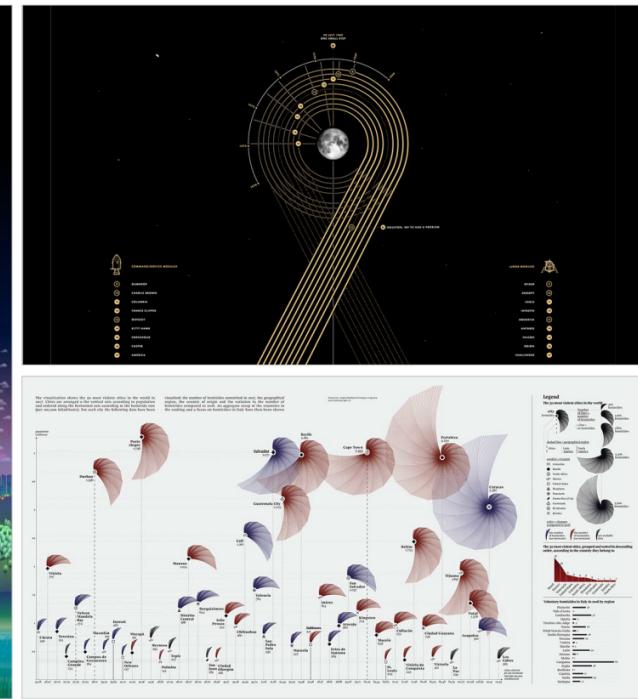
Anonymous

## We aim for DataViz that:

- is informative
  - is truthful
  - is easy to grasp
  - is visually appealing
  - draws attention
- but:
- is not abstract
  - is not too unusual



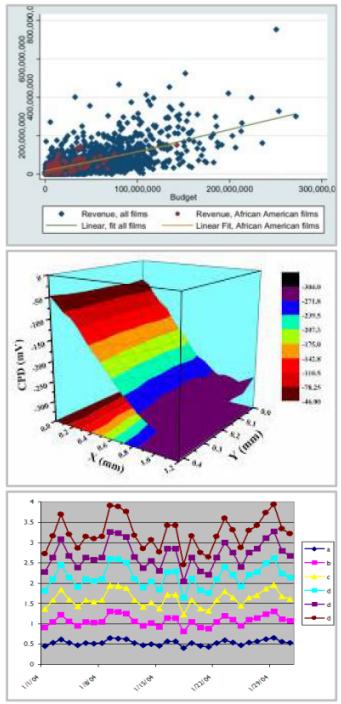
Sonia Kuijpers



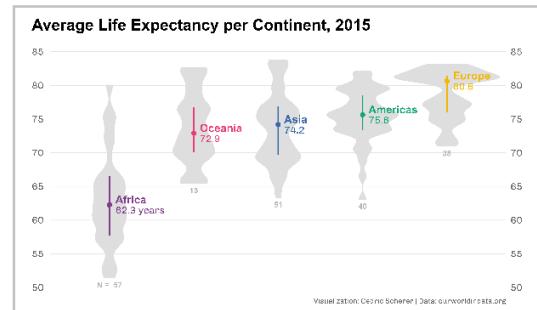
Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from uninformative & poorly designed data visualizations to data art

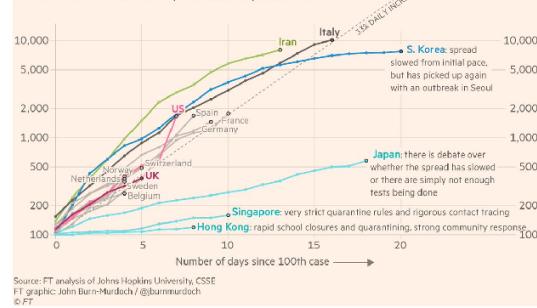


Anonymous



Most western countries are on the same coronavirus trajectory. Hong Kong and Singapore have managed to slow the spread

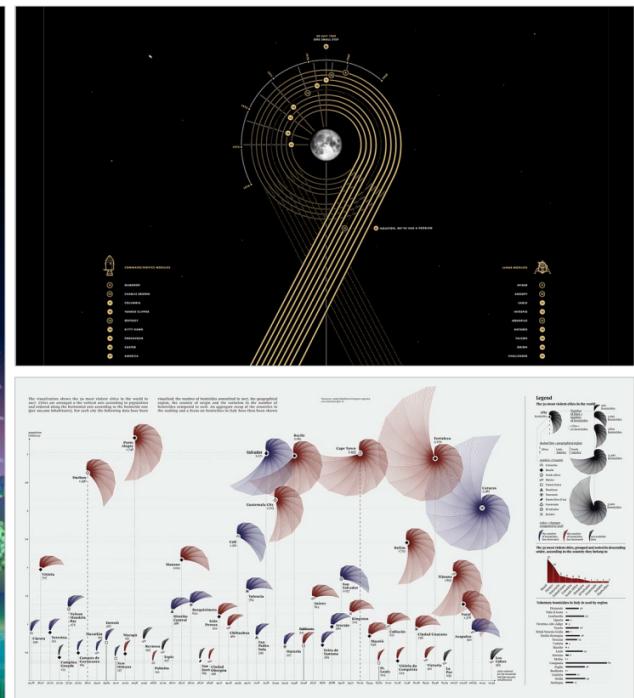
Cumulative number of cases, by number of days since 100th case



Upper: Cédric Scherer  
Lower: John Burn-Murdoch



Sonia Kuijpers

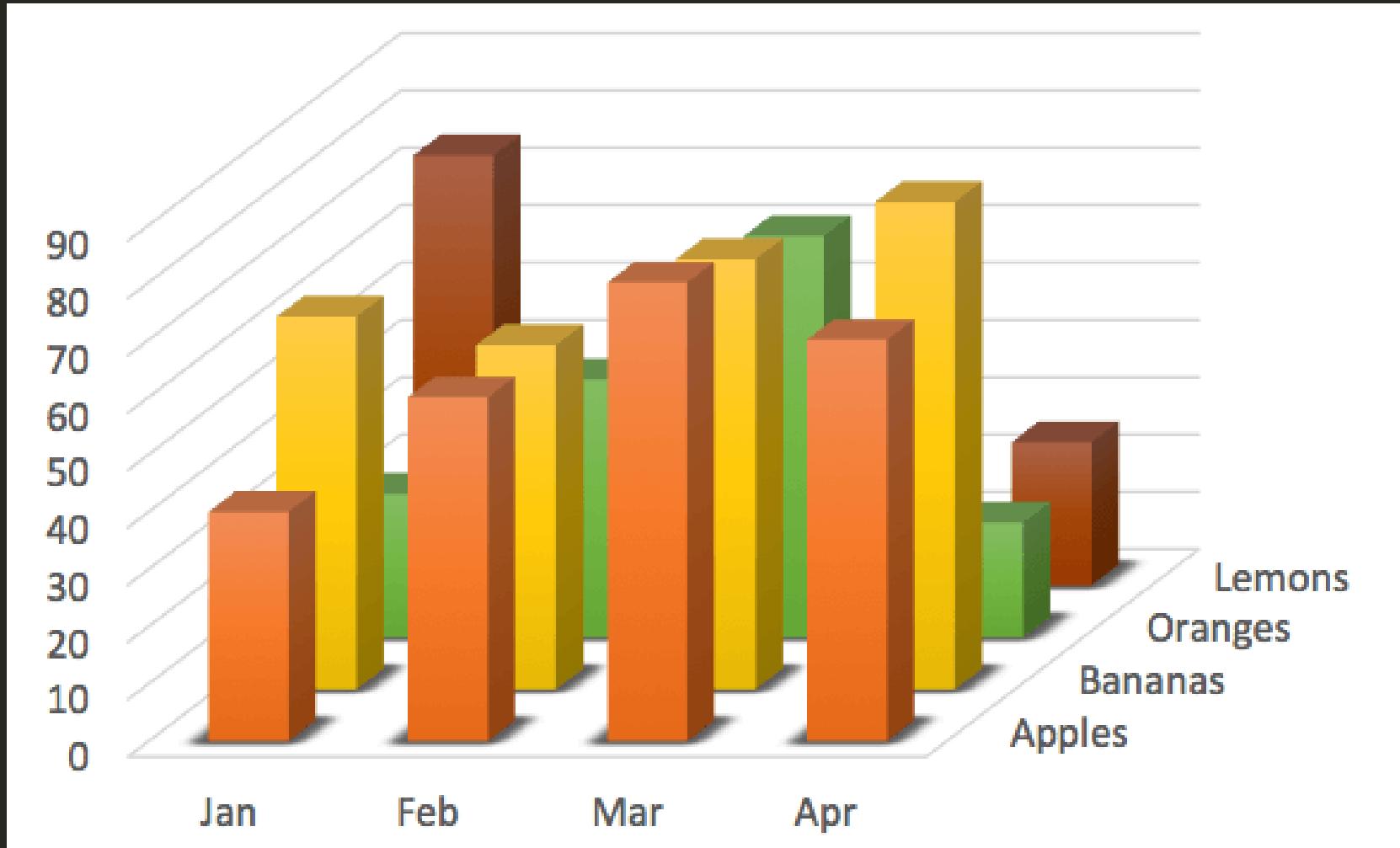


Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from uninformative & poorly designed data visualizations to data art

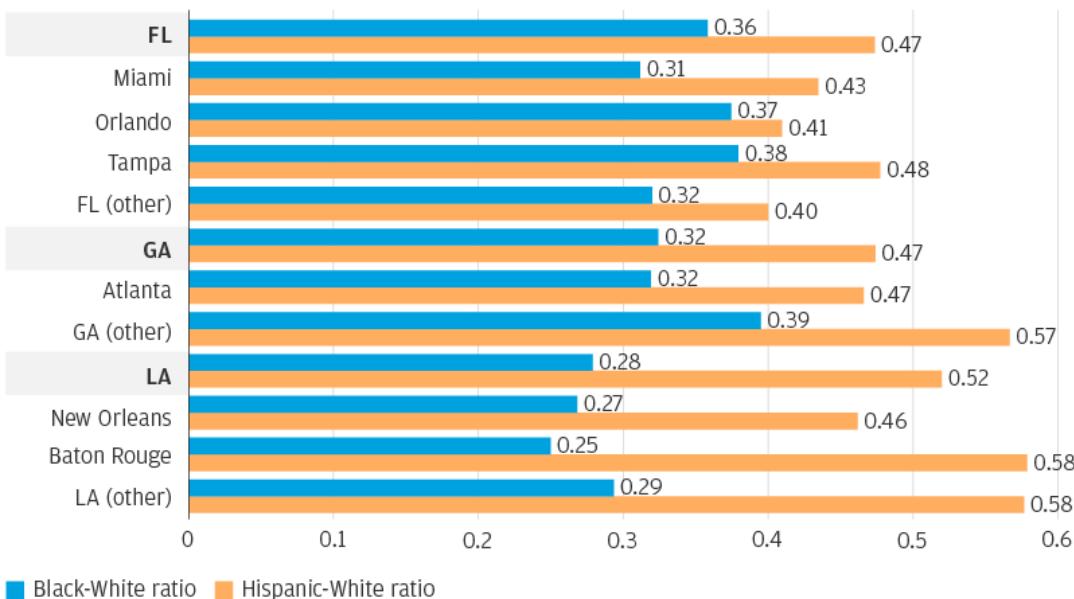
# What Makes It a Bad Data Visualization?



# What Makes It a Bad Data Visualization?

**Finding Four:** Across geographies, the financial outcomes of Hispanic families vary the most, while the financial outcomes of Black families vary the least. Black-White gaps in financial outcomes are largest in Louisiana, while Hispanic-White gaps are largest in Florida.

Black-White and Hispanic-White ratios of annual median liquid assets (2018), by geography



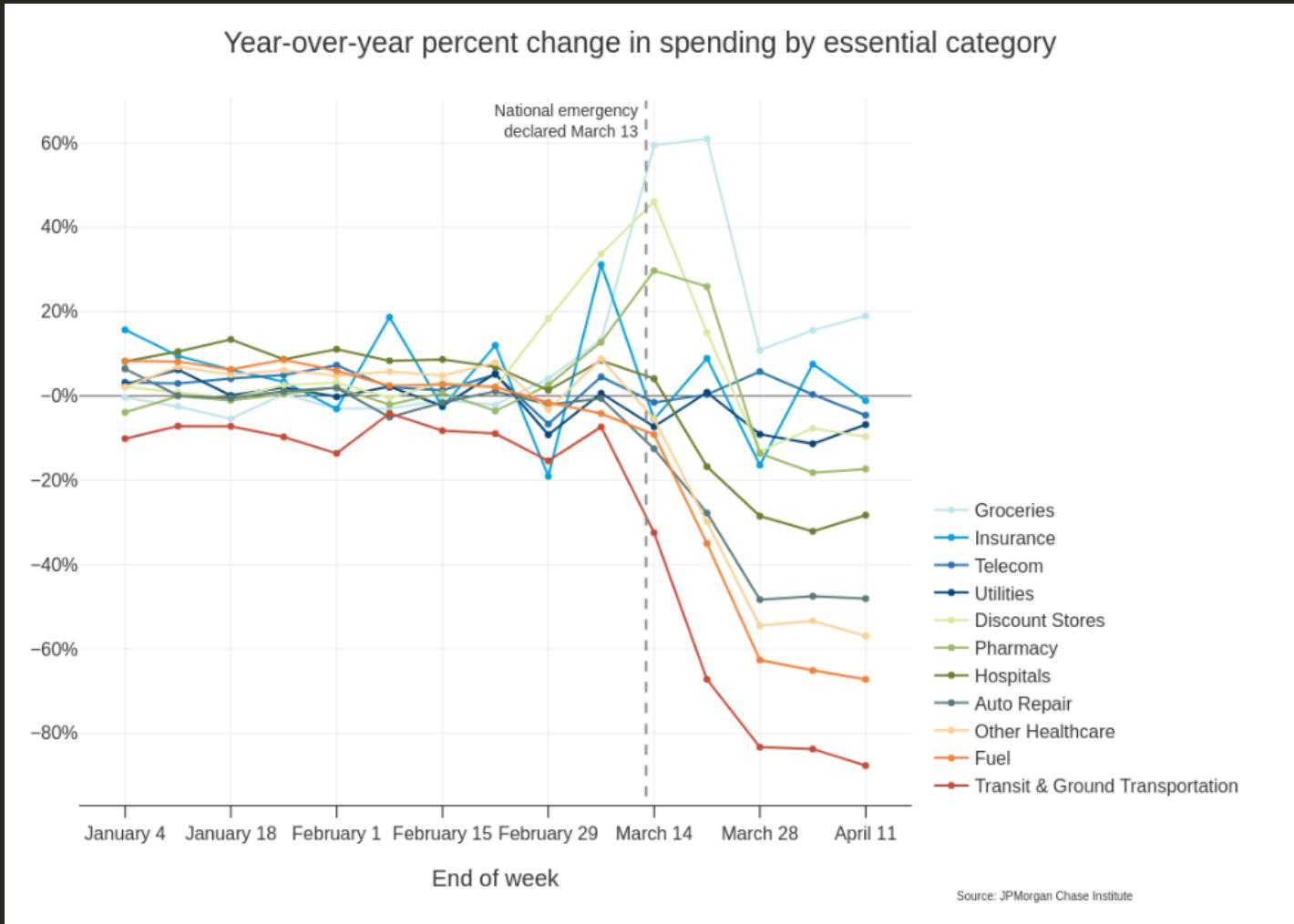
■ Black-White ratio ■ Hispanic-White ratio

Note: Liquid assets is the sum of balances in one's checking, prepaid debit cards, savings, money market, and certificates of deposit accounts. Cities refer to CBSAs (e.g., Miami refers to the Miami-Fort Lauderdale-West Palm Beach CBSA).

Source: JPMorgan Chase Institute

[View the Text Version >](#)

# What Makes It a Bad Data Visualization?



# **What Makes It a Bad Data Visualization**

# **What Makes It a Bad Data Visualization**

→ **Substantive problems** (bad data/story)

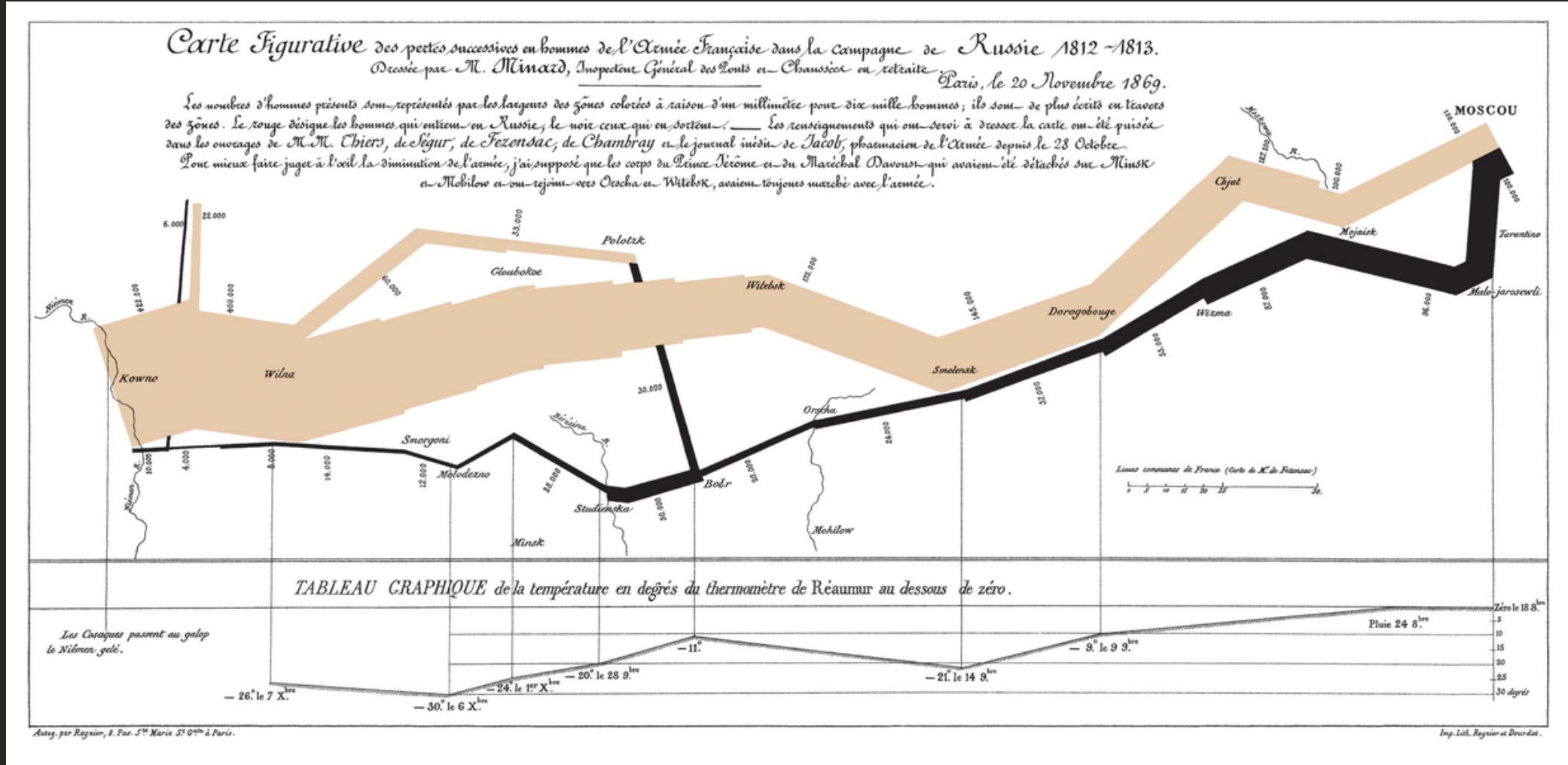
# **What Makes It a Bad Data Visualization**

- **Substantive problems** (bad data/story)
- **Aesthetic problems** (bad design)

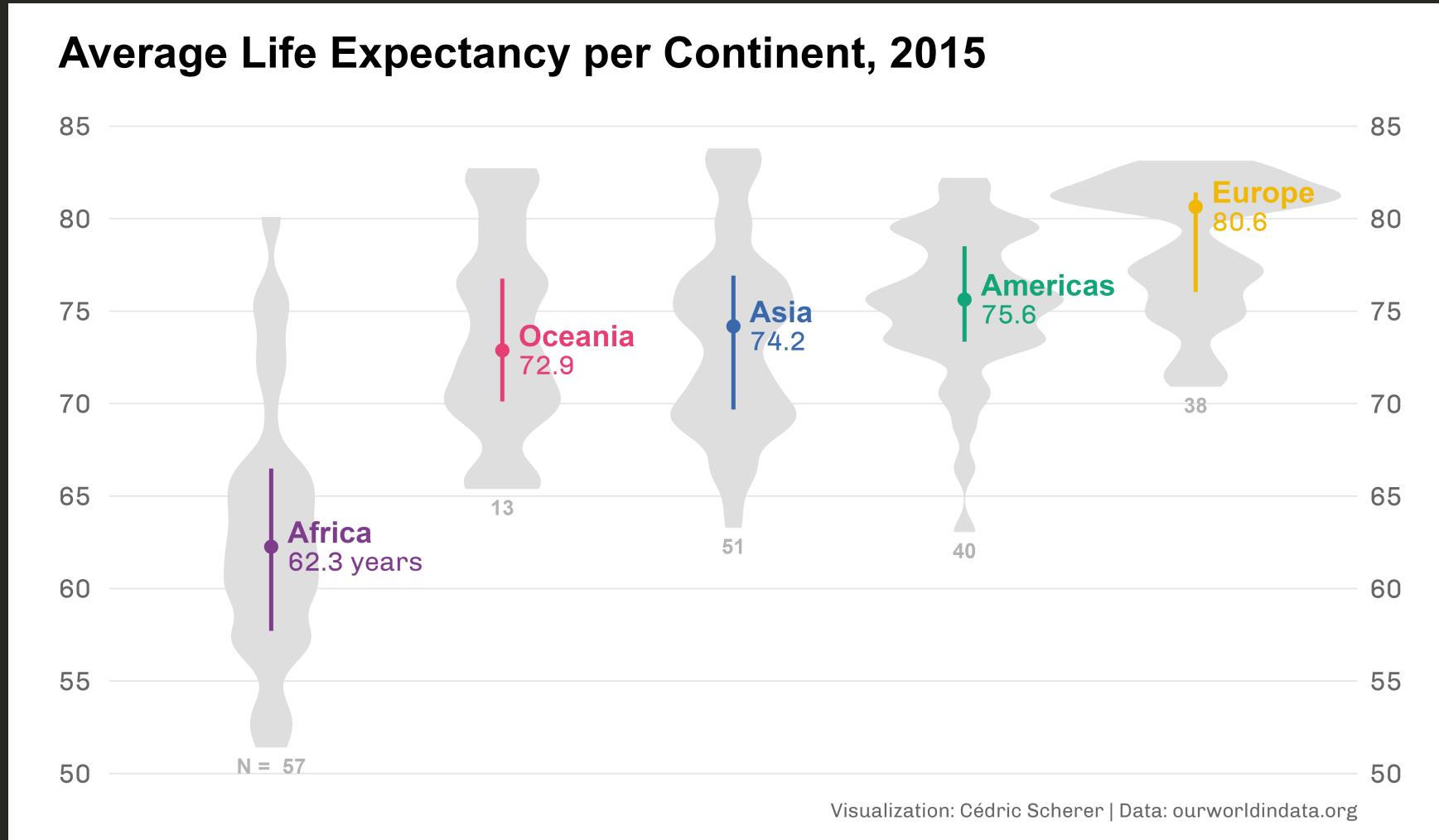
# **What Makes It a Bad Data Visualization**

- **Substantive problems** (bad data/story)
- **Aesthetic problems** (bad design)
- **Perceptual problems** (bad perception)

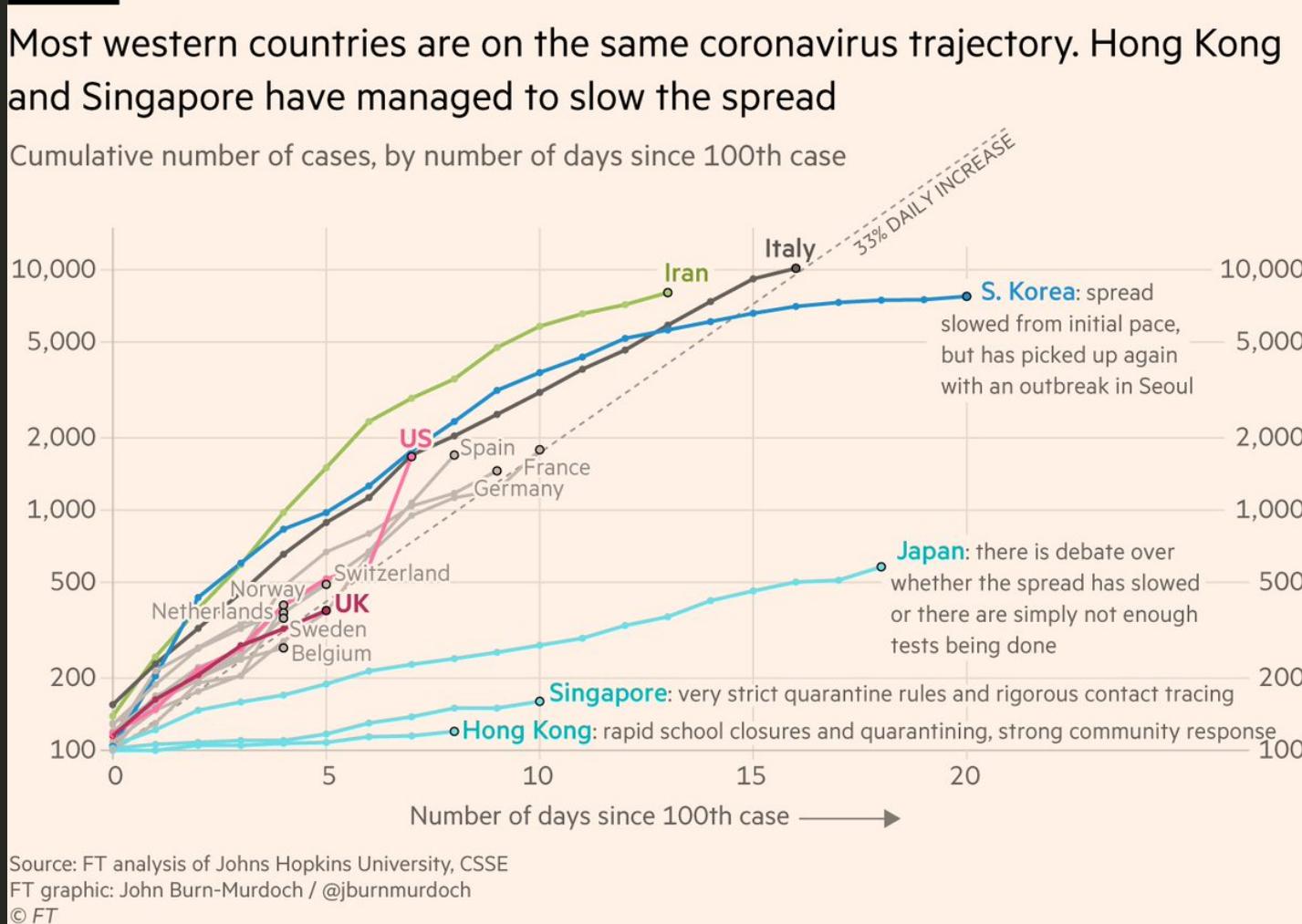
# What Makes It a Good Data Visualization?



# What Makes It a Good Data Visualization?



# What Makes It a Good Data Visualization?



# What Makes It a Good Data Visualization

# What Makes It a Good Data Visualization

→ **Information** (Integrity)

# What Makes It a Good Data Visualization

- **Information** (Integrity)
- **Story** (Interestingness)

# What Makes It a Good Data Visualization

- **Information** (Integrity)
- **Story** (Interestingness)
- **Goal** (Usefulness)

# What Makes It a Good Data Visualization

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



# INFORMATION

*Understand your data and be accurate*



How to Steer Clear of Common Blunders When Working  
with Data and Presenting Analysis and Visualizations

# AVOIDING DATA PITFALLS

BEN JONES

Founder and CEO, Data Literacy

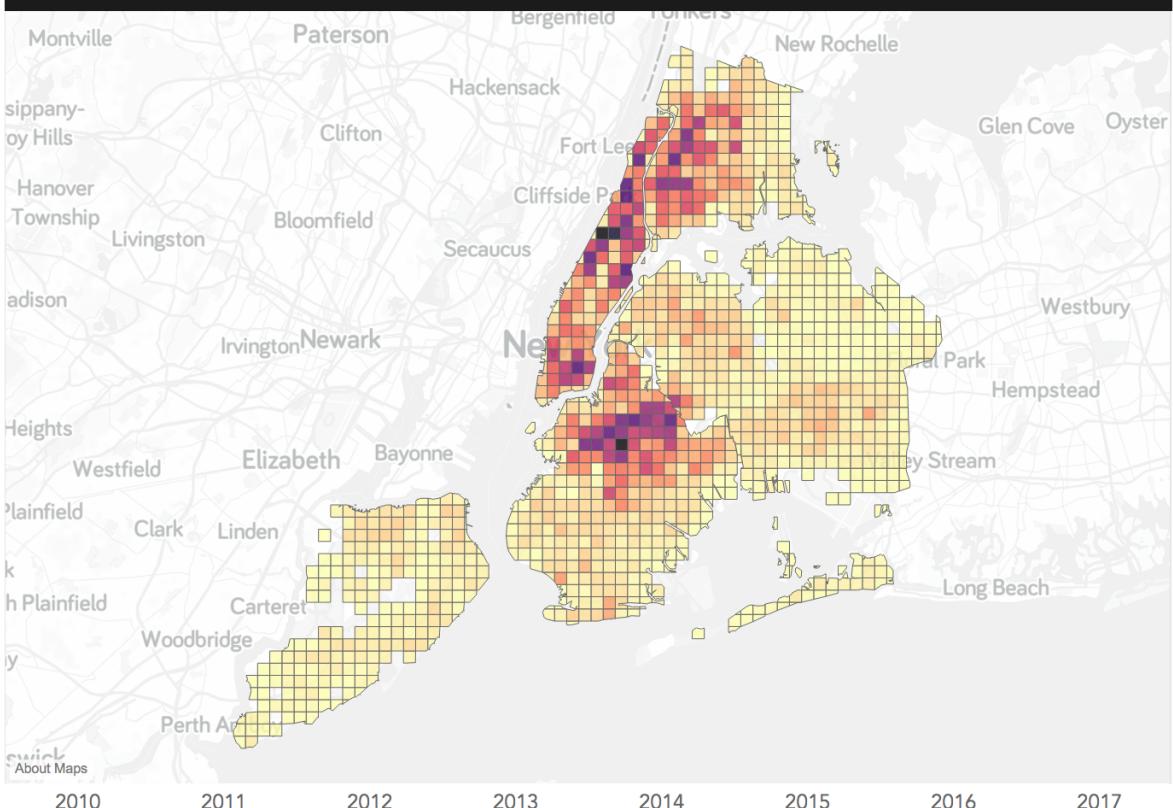
WILEY

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

# Where are New York's rats?

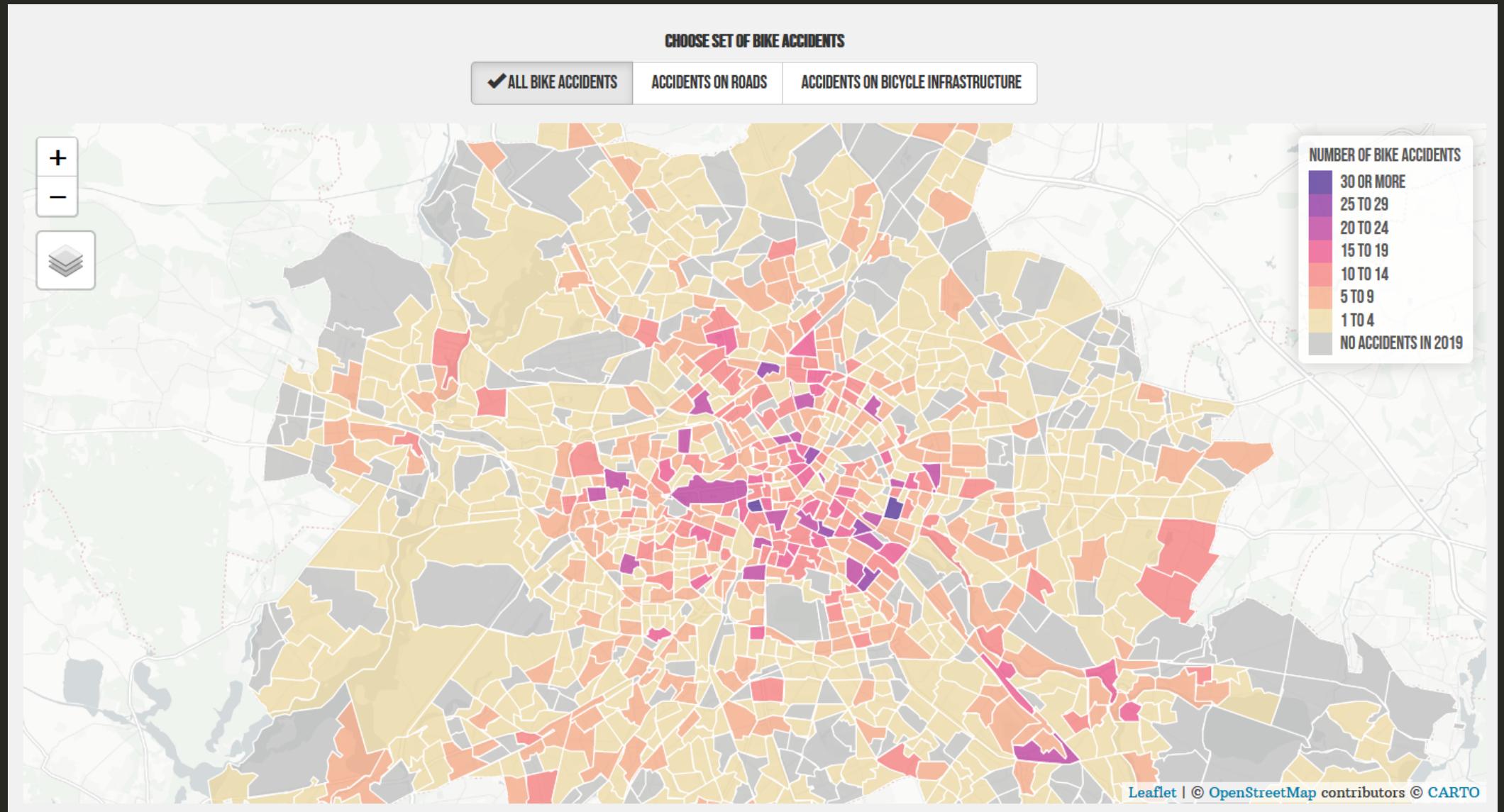
Choose a Borough

(All)



2010 2011 2012 2013 2014 2015 2016 2017

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



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



Source: [inhomelandsecurity.com/risk-management-and-black-swan-events](http://inhomelandsecurity.com/risk-management-and-black-swan-events)



→ "The swan is white" (singular statement)  
→ "All swans are white" (universal statement)

# Know Your Types of Data

# Types of Data

- Quantitative (numerical) versus qualitative (categorical) data

# Types of Data

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data

# Types of Data

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female"
  - 2019/09/26 "17:01:35"
  - 1

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35"
  - 1

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35"
  - 1

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35" → quantitative + continuous + ordered
  - 1

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35" → quantitative + continuous + ordered
  - 1

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35" → quantitative + continuous + ordered
  - 1 → quantitative + continuous + ordered

# Types of Data – Your Turn!

- Quantitative (numerical) versus qualitative (categorical) data
- Ordered versus unordered data
- Continuous versus discrete data
- What are the data types of:
  - "female" → qualitative + discrete + unordered
  - 2019/09/26 "17:01:35" → quantitative + continuous + ordered
  - 1 → quantitative + continuous + ordered  
or: quantitative + discrete + ordered  
or: qualitative + discrete + ordered  
or: qualitative + discrete + unordered

## NOMINAL

UNORDERED DESCRIPTIONS



## ORDINAL

ORDERED DESCRIPTIONS



## BINARY

ONLY 2 MUTUALLY EXCLUSIVE OUTCOMES



@allison\_horst

## CONTINUOUS

measured data, can have  $\infty$  values within possible range.



I AM 3.1" TALL

I WEIGH 34.16 grams

## DISCRETE

OBSERVATIONS CAN ONLY EXIST  
AT LIMITED VALUES, OFTEN COUNTS.



I HAVE 8 LEGS  
and  
4 SPOTS!

@allison\_horst

# STORY

*Be clear about the message of your visualization*

*It is hard to effectively design for others.  
It is harder if you don't even care to try.*

**Andy Kirk**

# Who is my audience?

# Who is my audience?

Which story is **interesting** for my audience?

# Who is my audience?

Which story is **interesting** for my audience?

What are **relevant** details to include?

# Who is my audience?

Which story is **interesting** for my audience?

What are **relevant** details to include?

Which variables are **meaningful** to my audience?

# Who is my audience?

Which story is **interesting** for my audience?

What are **relevant** details to include?

Which variables are **meaningful** to my audience?

How will they **encounter** the visualization?

# **Who is my audience?**

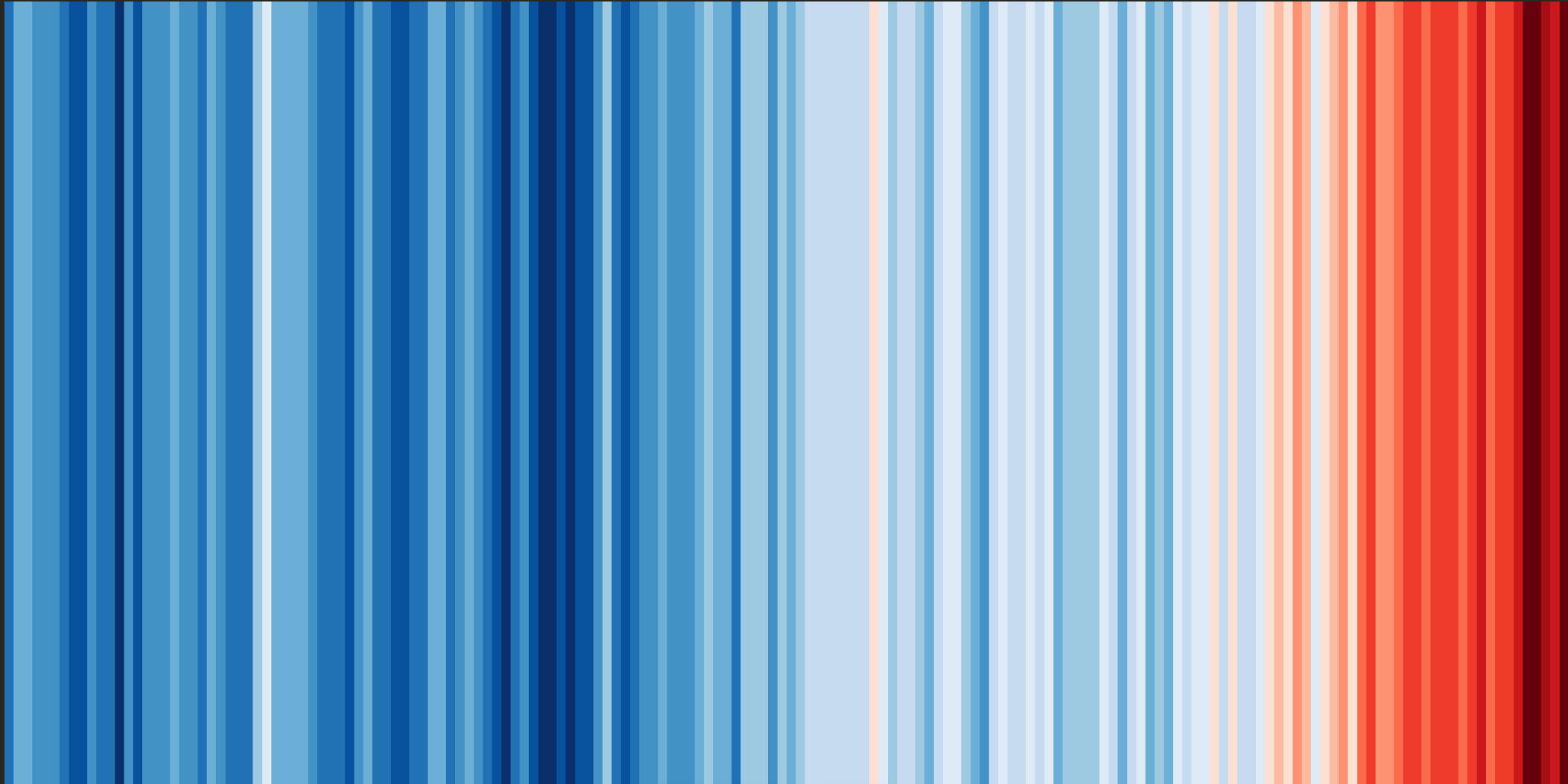
Which story is **interesting** for my audience?

What are **relevant** details to include?

Which variables are **meaningful** to my audience?

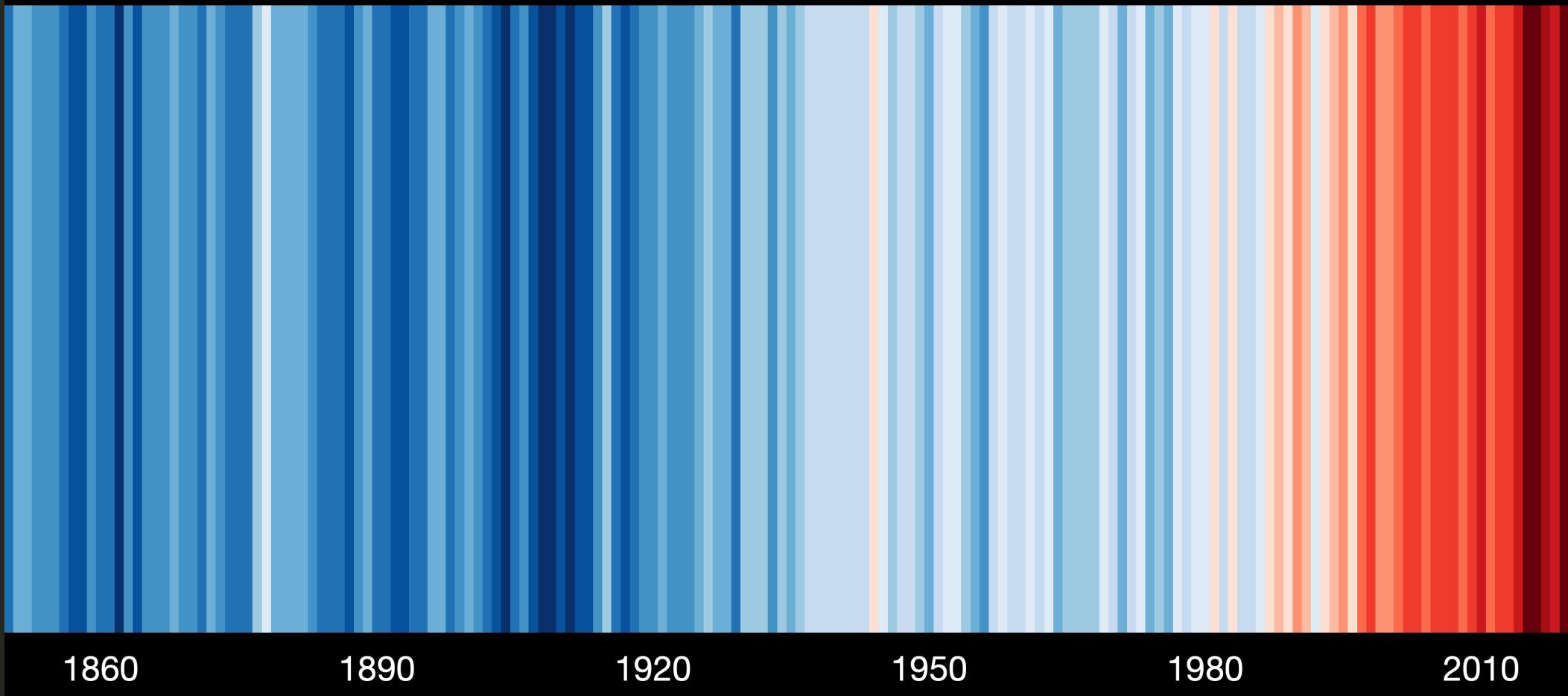
How will they **encounter** the visualization?

**Do I need a visualization at all??**

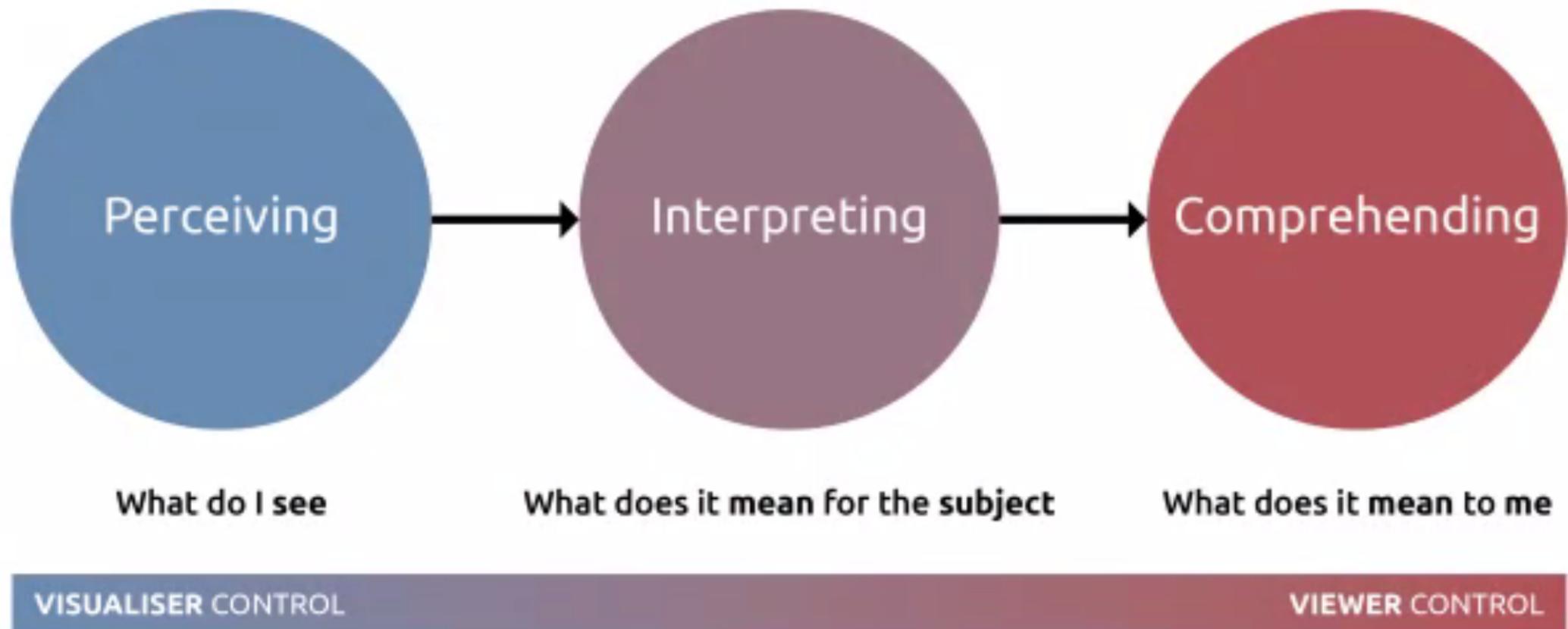


**Warming Stripes** by Ed Hawkins

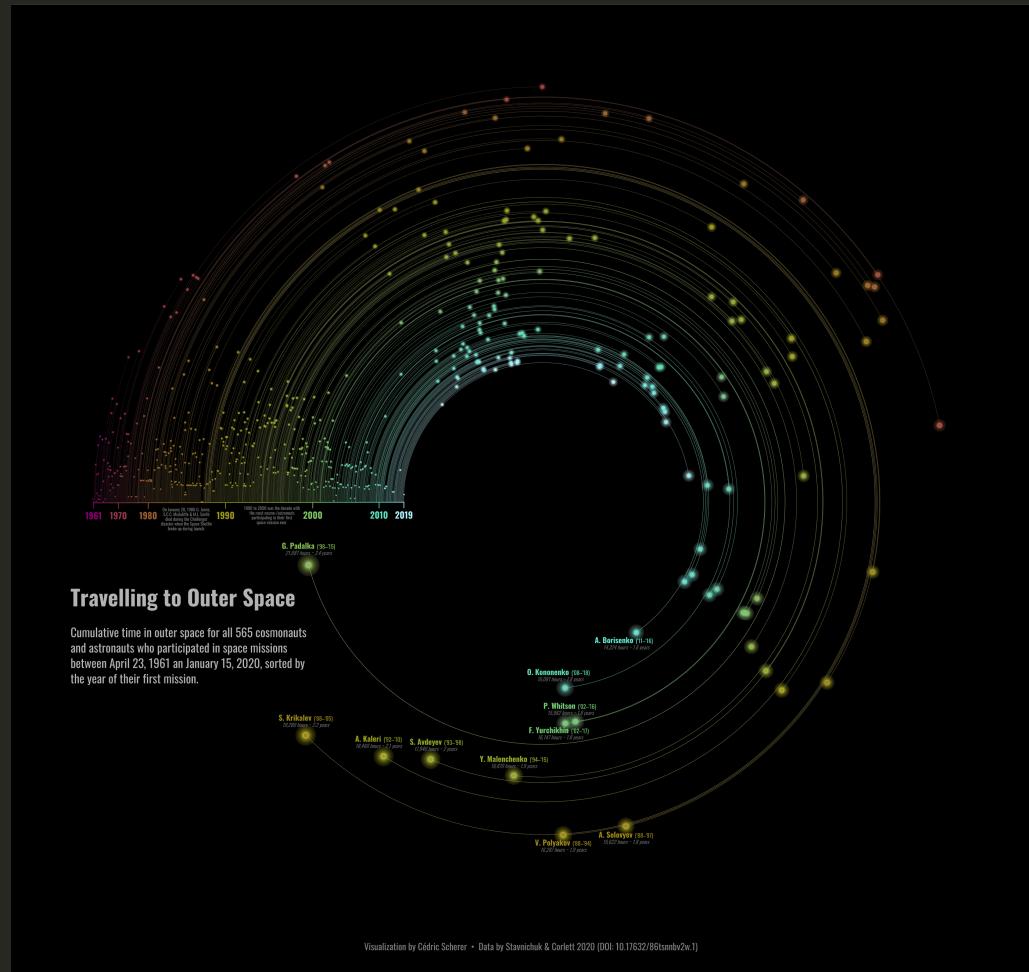
# Global temperature change (1850-2019)



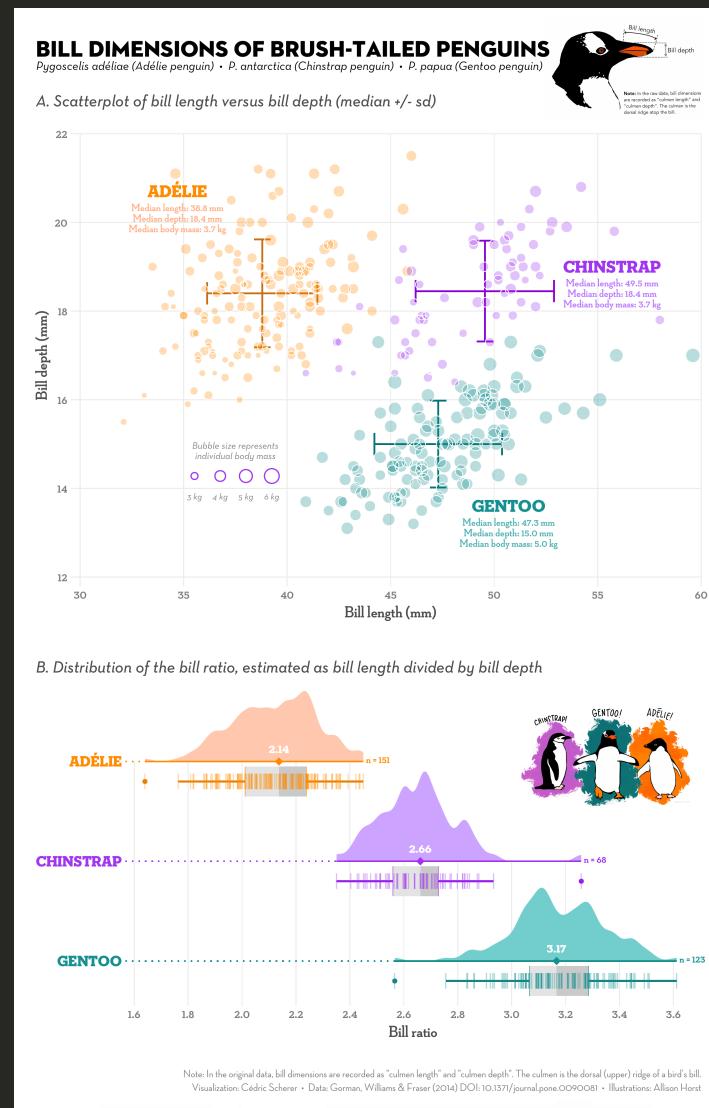
*Warming Stripes* by Ed Hawkins



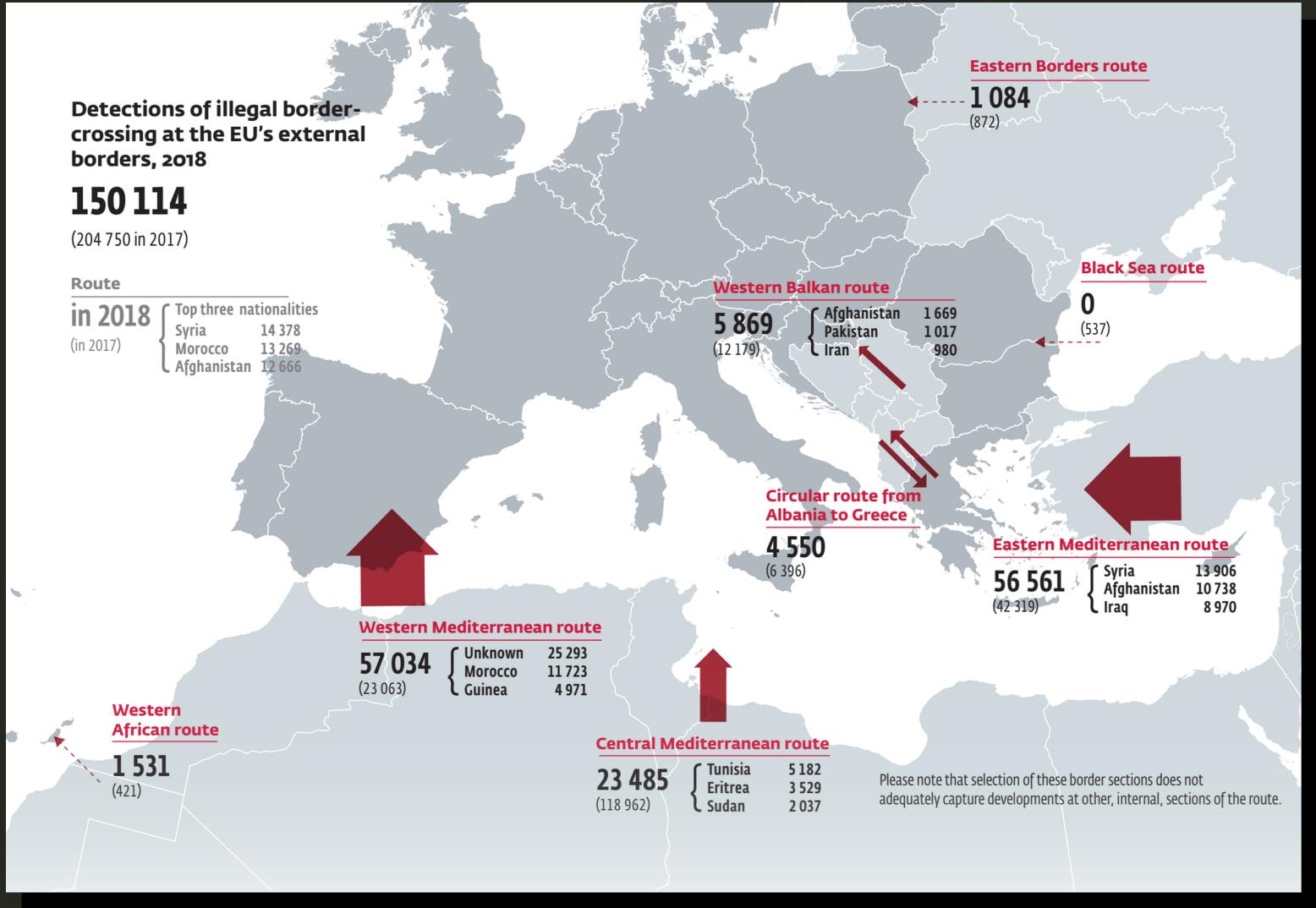
*Source: Andy Kirk, S-H-O-W Feb 2021*

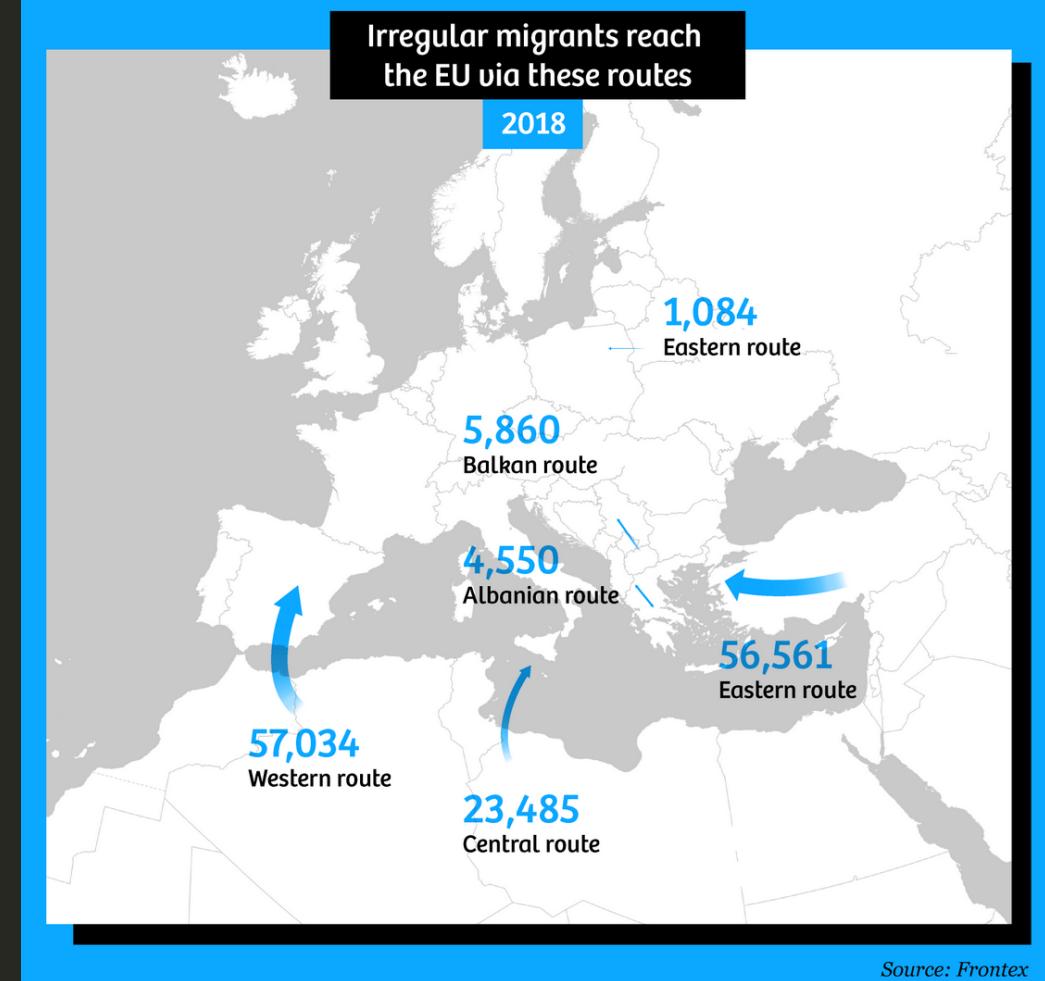
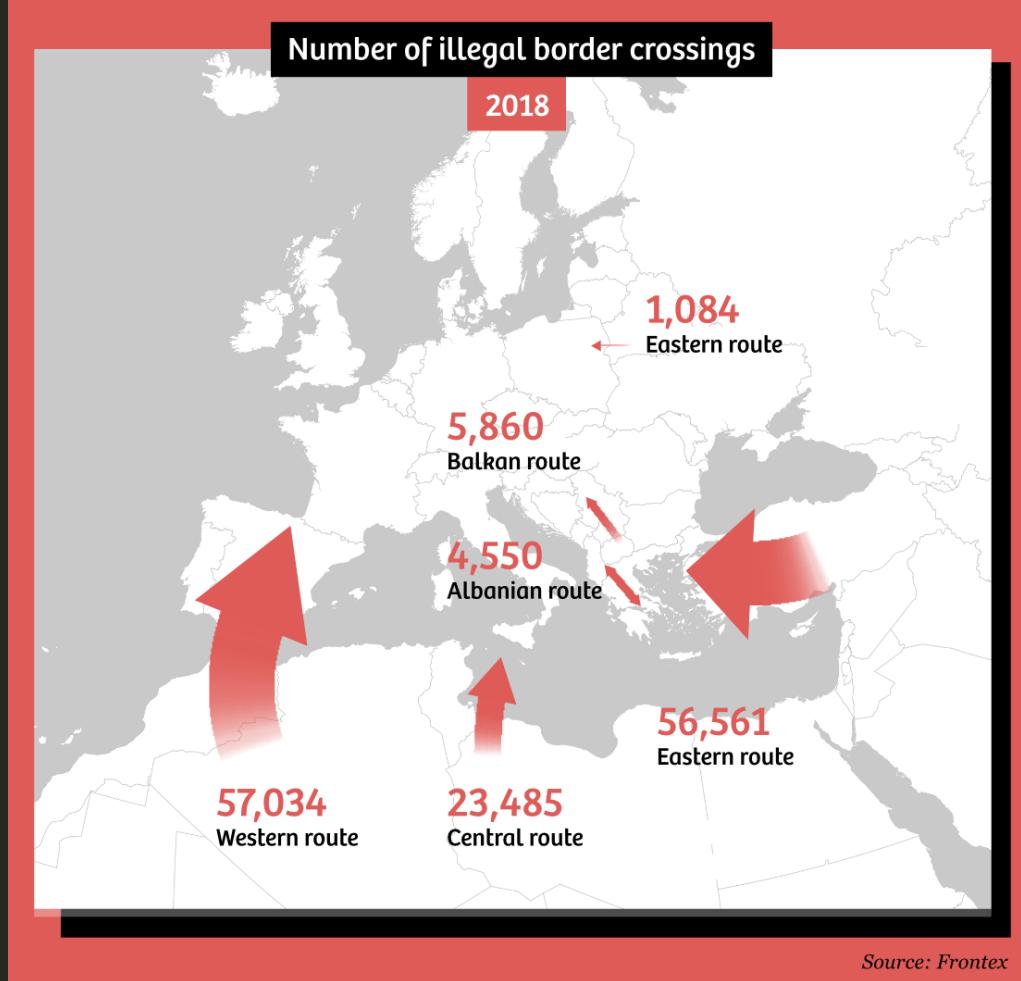


#TidyTuesday contribution Week 2020/29

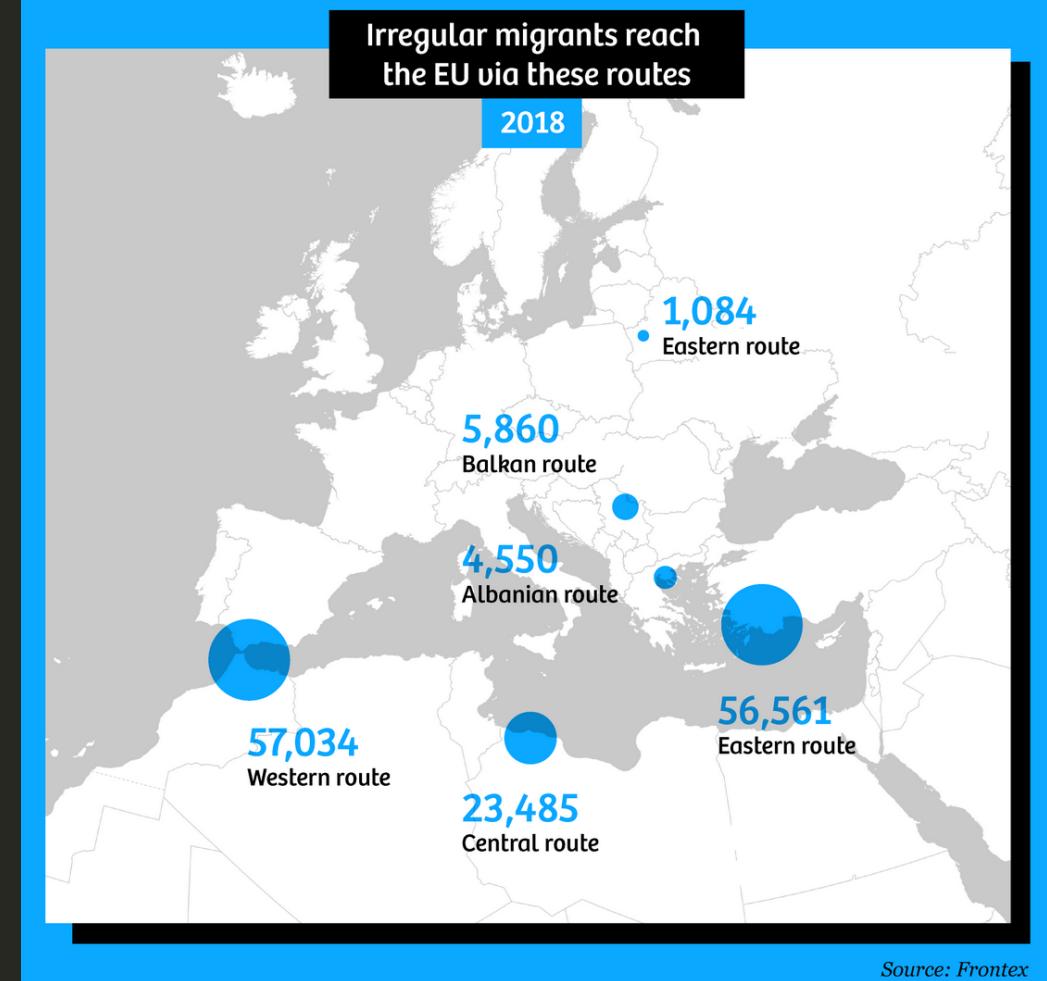


#TidyTuesday contribution Week 2020/31

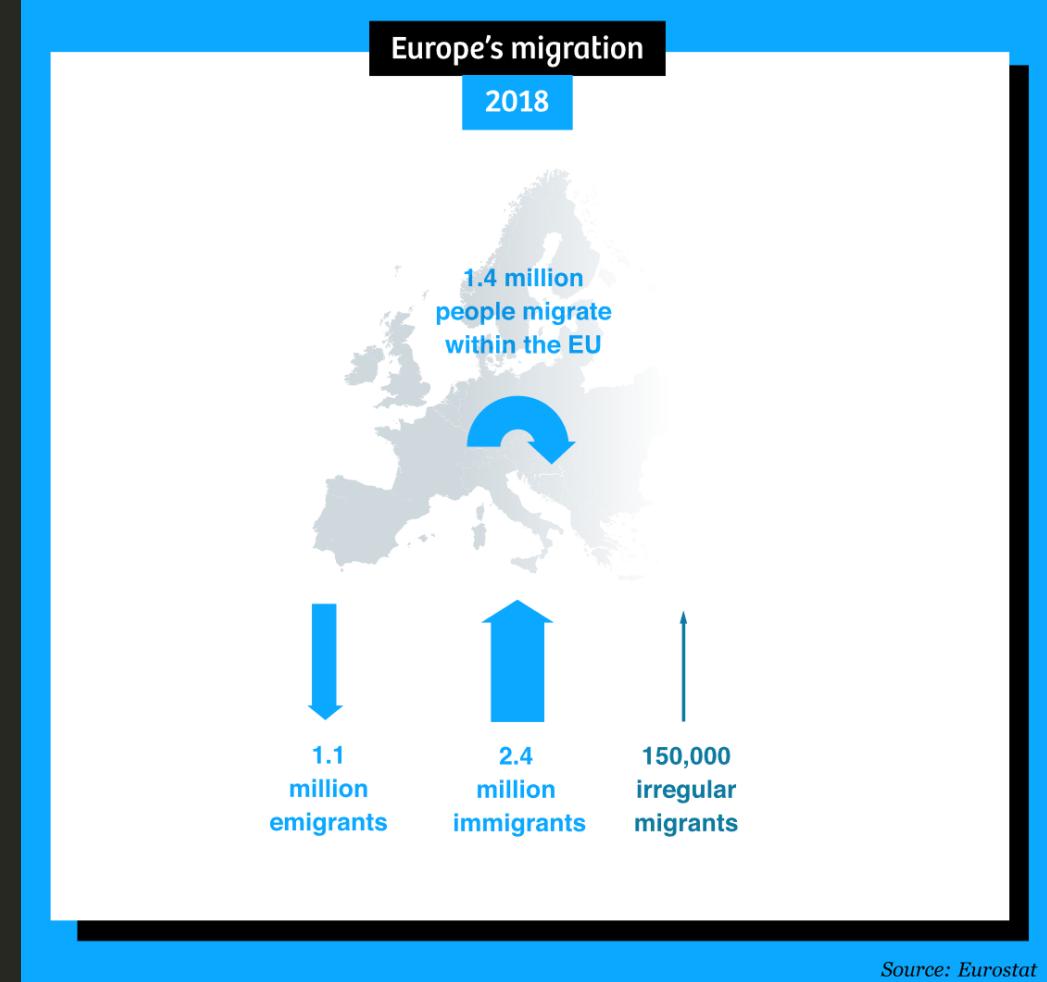




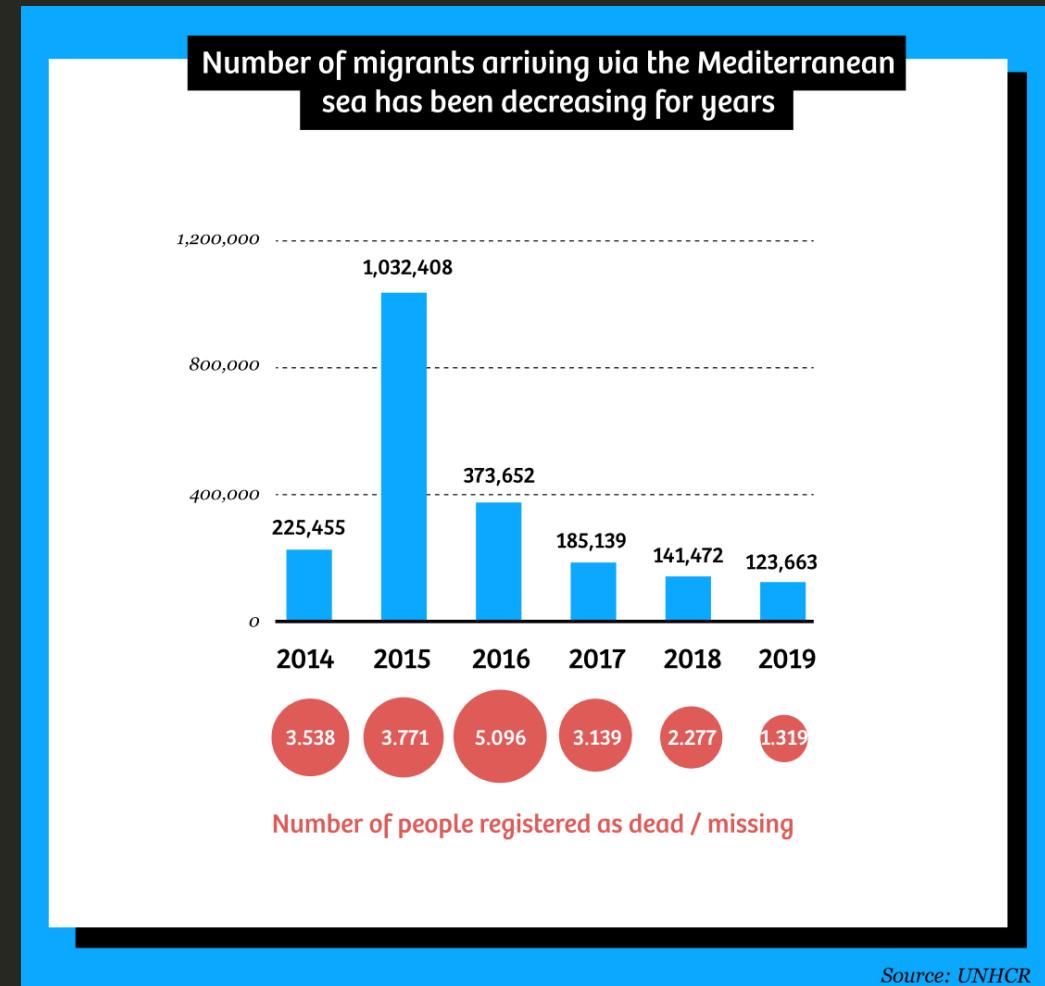
*How maps in the media make us more negative about migrants* by Maite Vermeulen, Leon de Korte & Henk van Houtum



*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

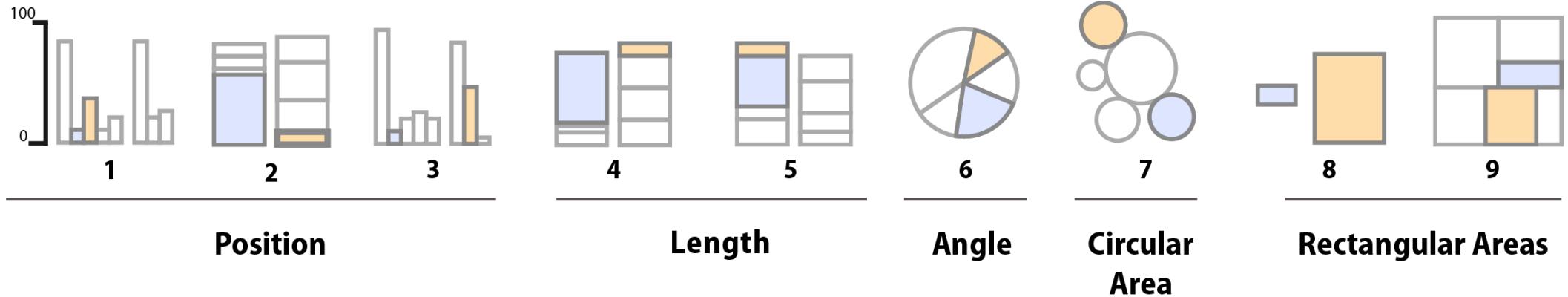


*How maps in the media make us more negative about migrants* by Maite Vermeulen, Leon de Korte & Henk van Houtum

# GOAL

*Select charts that successfully transport your story*

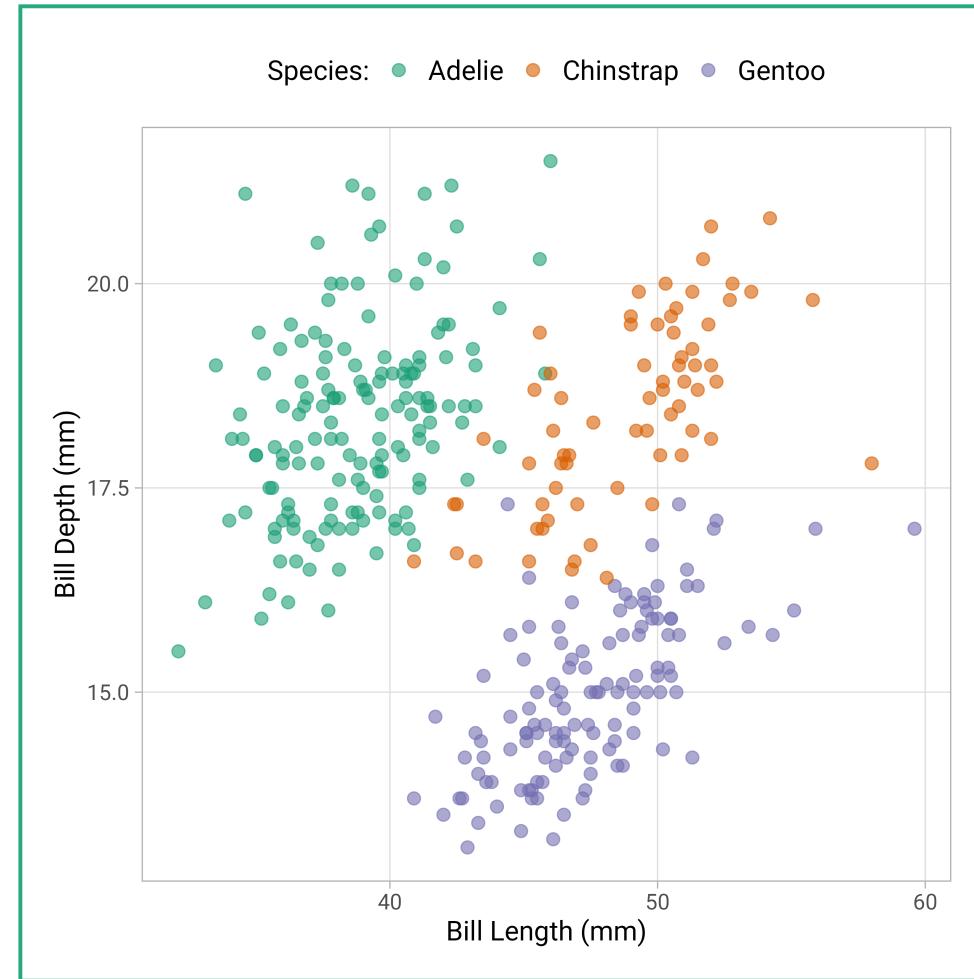
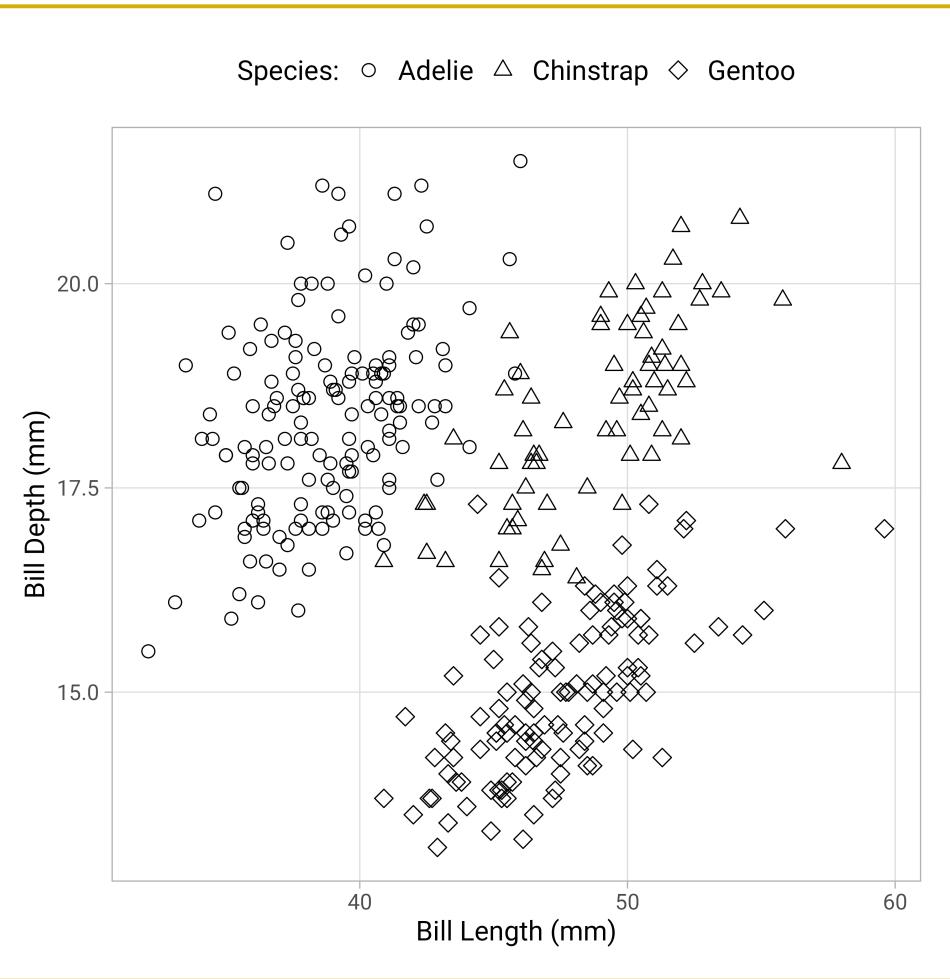
# Data visualizations map values into quantifiable features

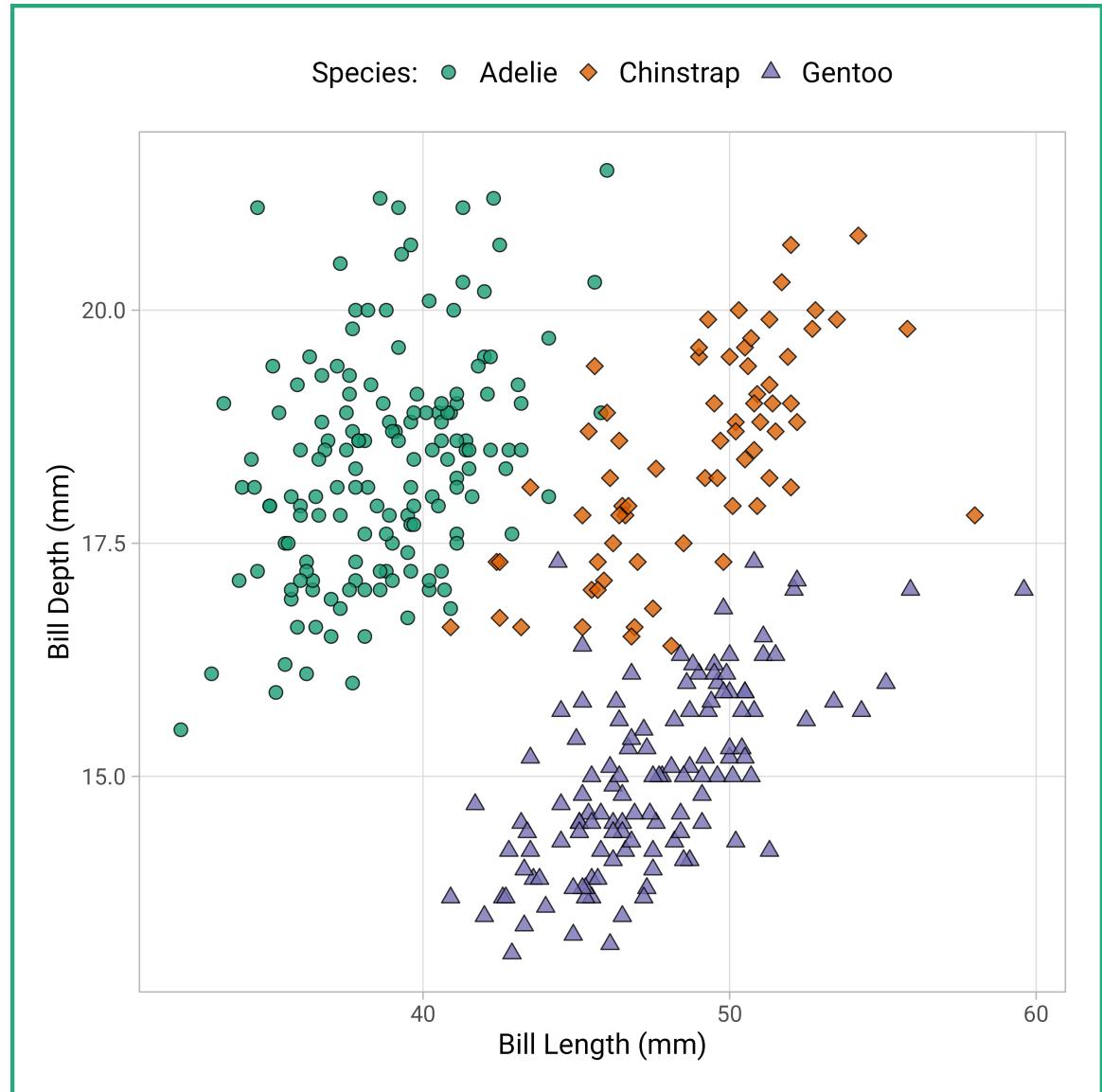


Source: [Kieran Healy](#) based on Heer and Bostock, following Cleveland and McGill

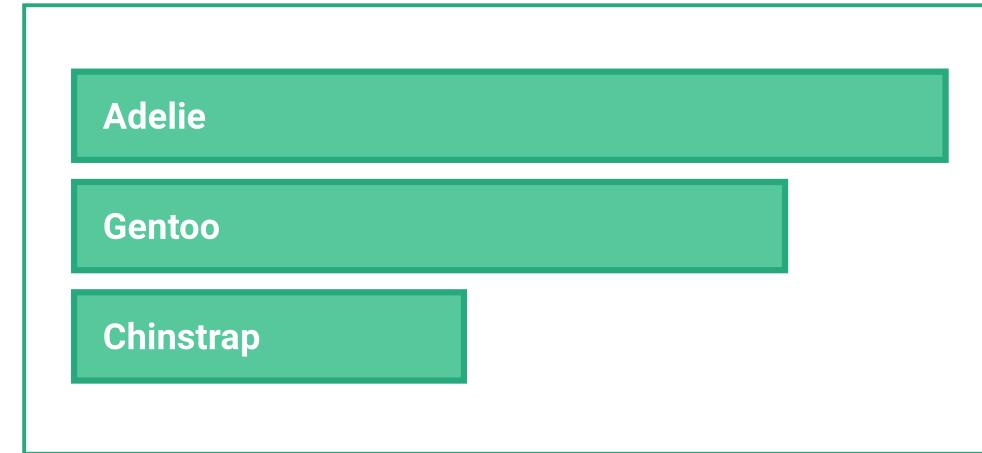
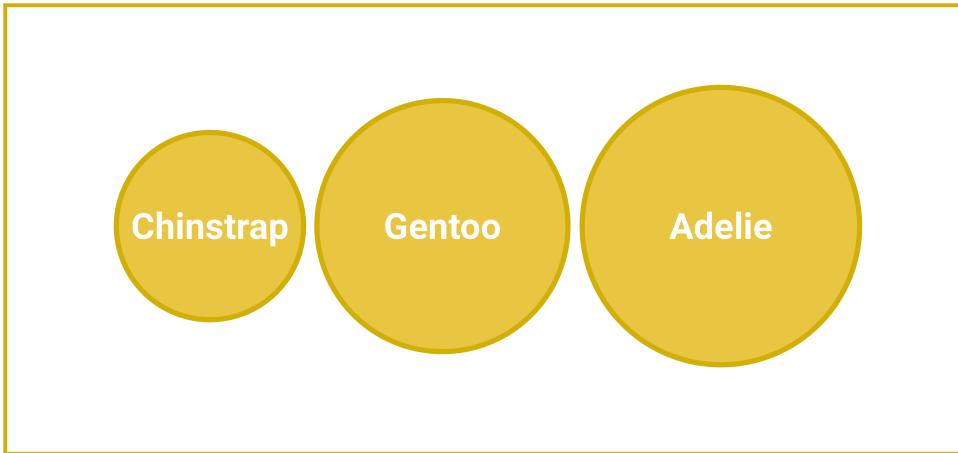
<b>Qualitative Nominal</b>	<b>Qualitative Ordinal</b>	<b>Quantitative Interval/Ratio</b>
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol

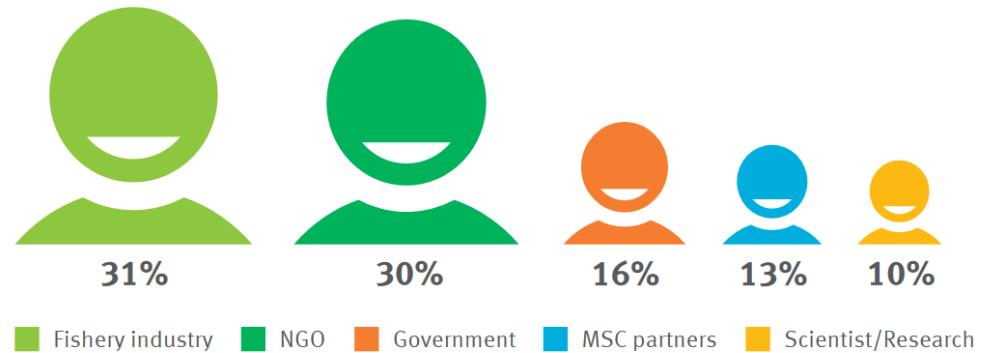
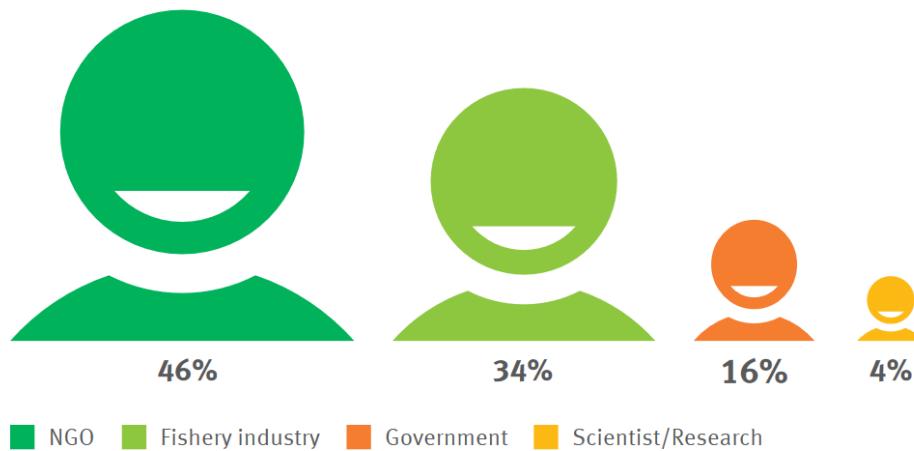
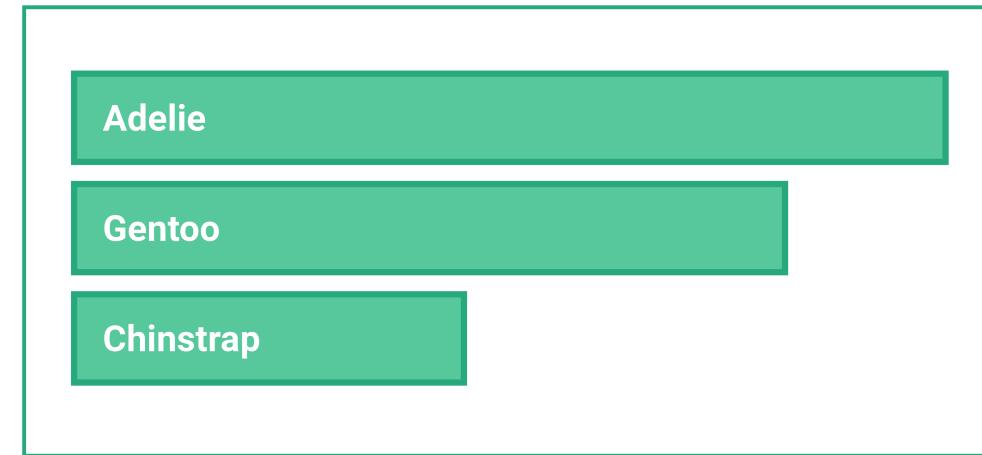
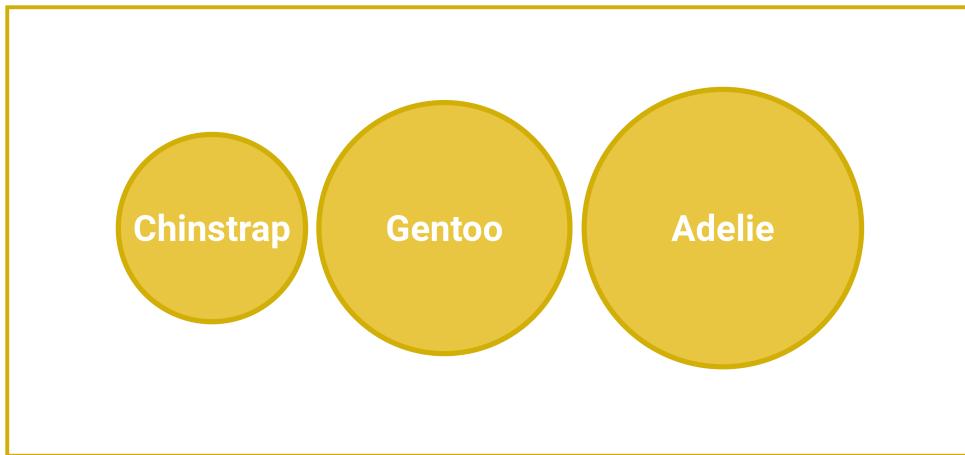
<b>Qualitative Nominal</b>	<b>Qualitative Ordinal</b>	<b>Quantitative Interval/Ratio</b>
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol



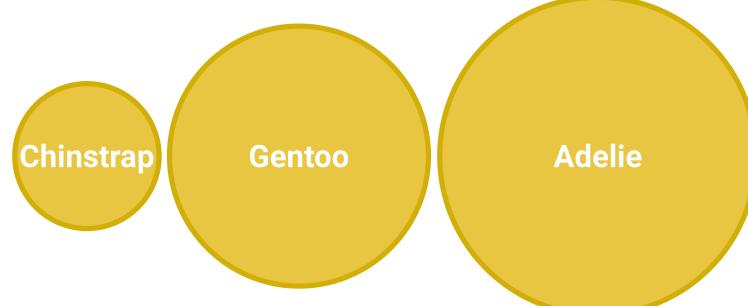
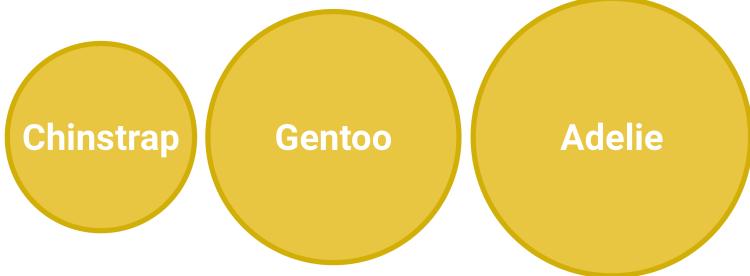


<b>Qualitative Nominal</b>	<b>Qualitative Ordinal</b>	<b>Quantitative Interval/Ratio</b>
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Pattern (Density)	Pattern (Texture)	Size (Volume)
Colour (Lightness)	Connection	Pattern (Density)
Symbol	Size (Length)	Colour (Lightness)
Size (Length)	Angle	Colour (Hue)
Angle	Size (Area)	Pattern (Texture)
Size (Area)	Size (Volume)	Connection
Size (Volume)	Symbol	Symbol





# Always use area. Never use radius!



Adelie

Gentoo

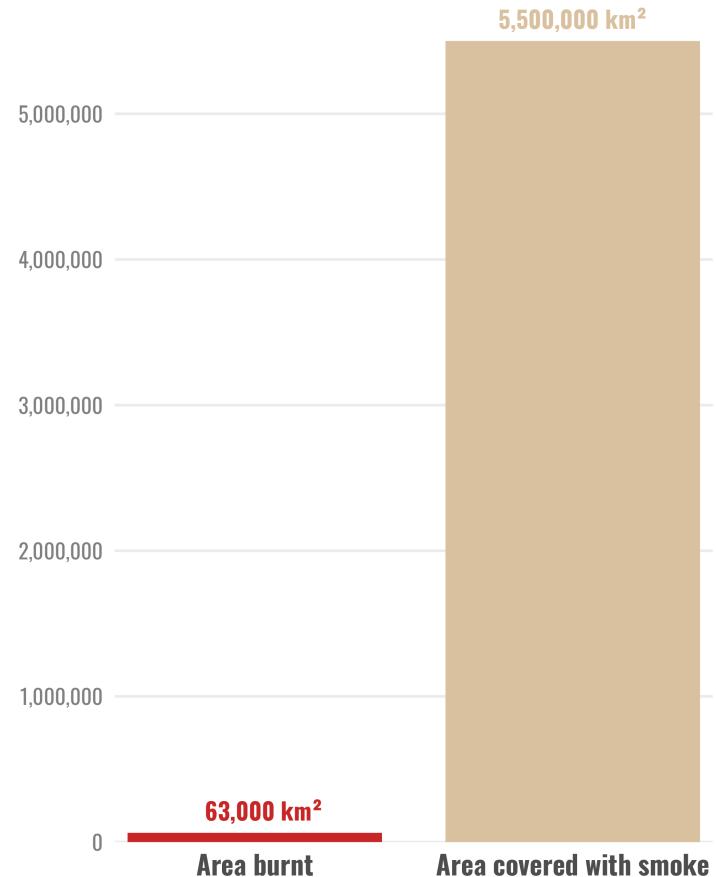
Chinstrap

Adelie

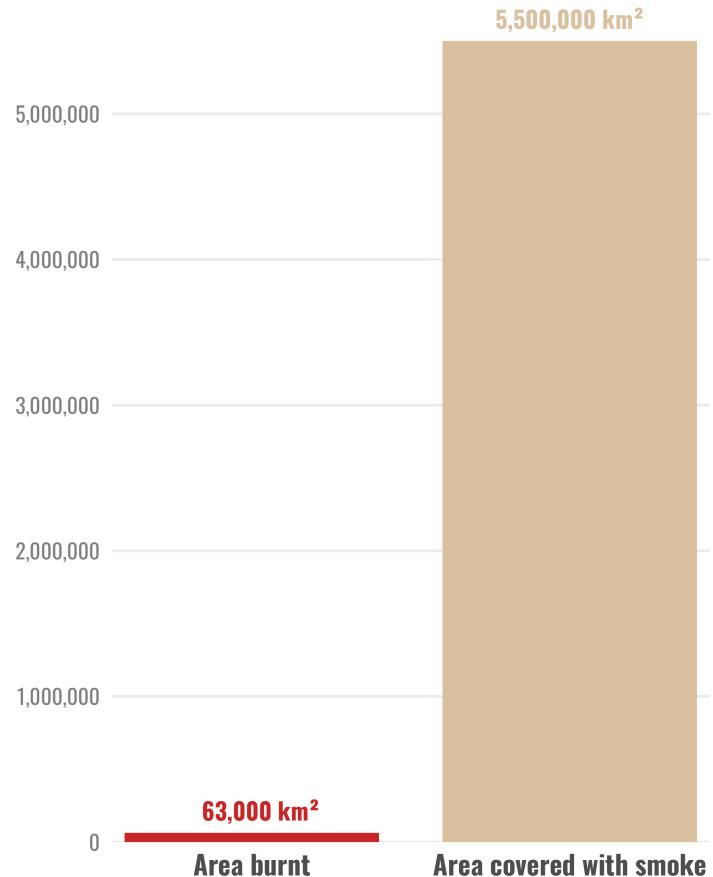
Gentoo

Chinstrap

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



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

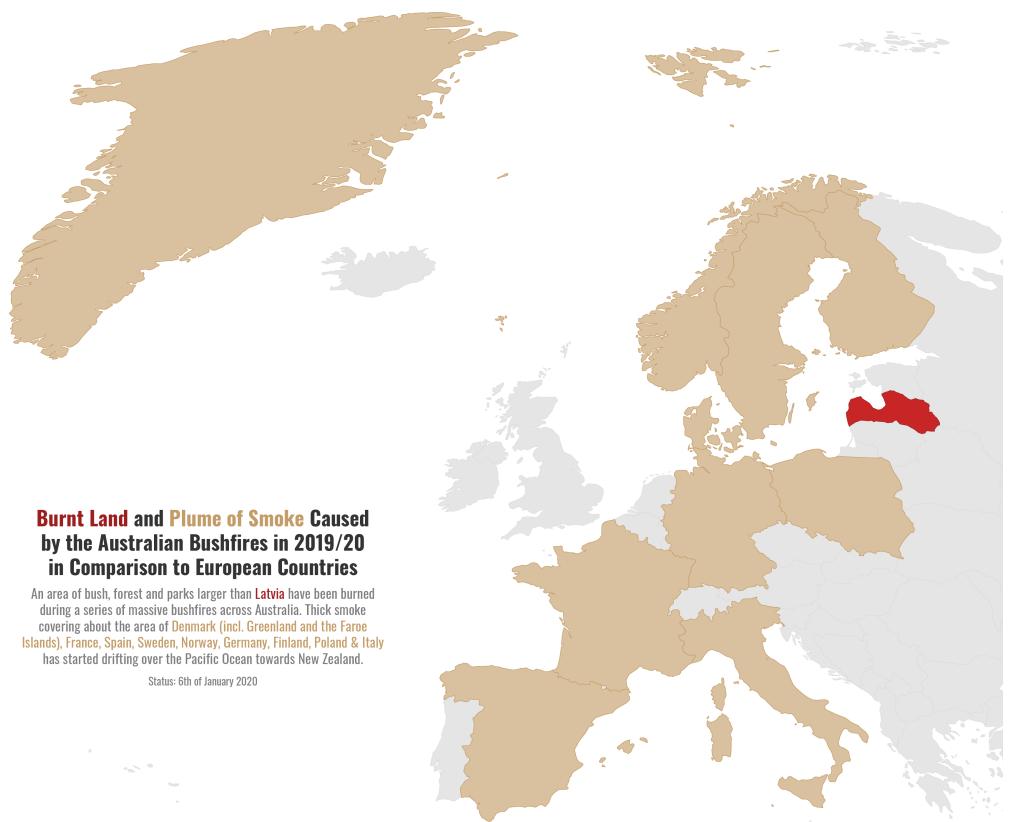


**Burnt Land and Plume of Smoke Caused  
by the Australian Bushfires in 2019/20  
in Comparison to European Countries**

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 Gédric Scherer · Data by NASA FIRMS & The Independent



#TidyTuesday contribution Week 2020/02

# Typology of Charts

by Scott Berinato, "Good Charts" (2016), pp. 54–63

Is the information **conceptual** or **data-driven**?

Is the purpose to **declare** or to **explore** the information?

# Typology of Charts

by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25

Is the information **conceptual** or **measurable**?

Is the purpose to **explain** or to **explore** the information?

# Typology of Charts

by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25

Is the information **conceptual** or **measurable**?

→ Type of Information

Is the purpose to **explain** or to **explore** the information?

→ Purpose of the Graphic

# Typology of Charts

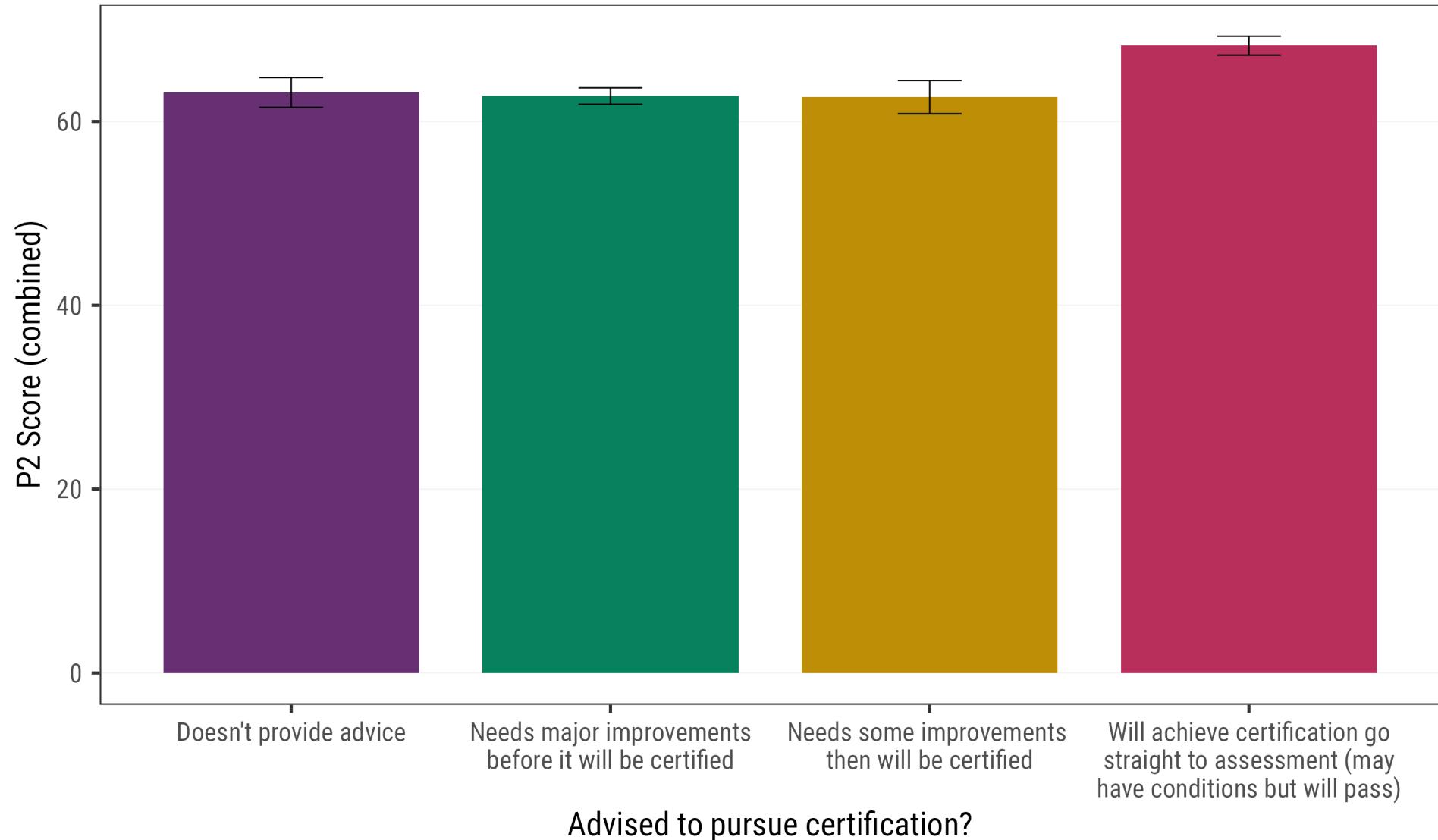
by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25

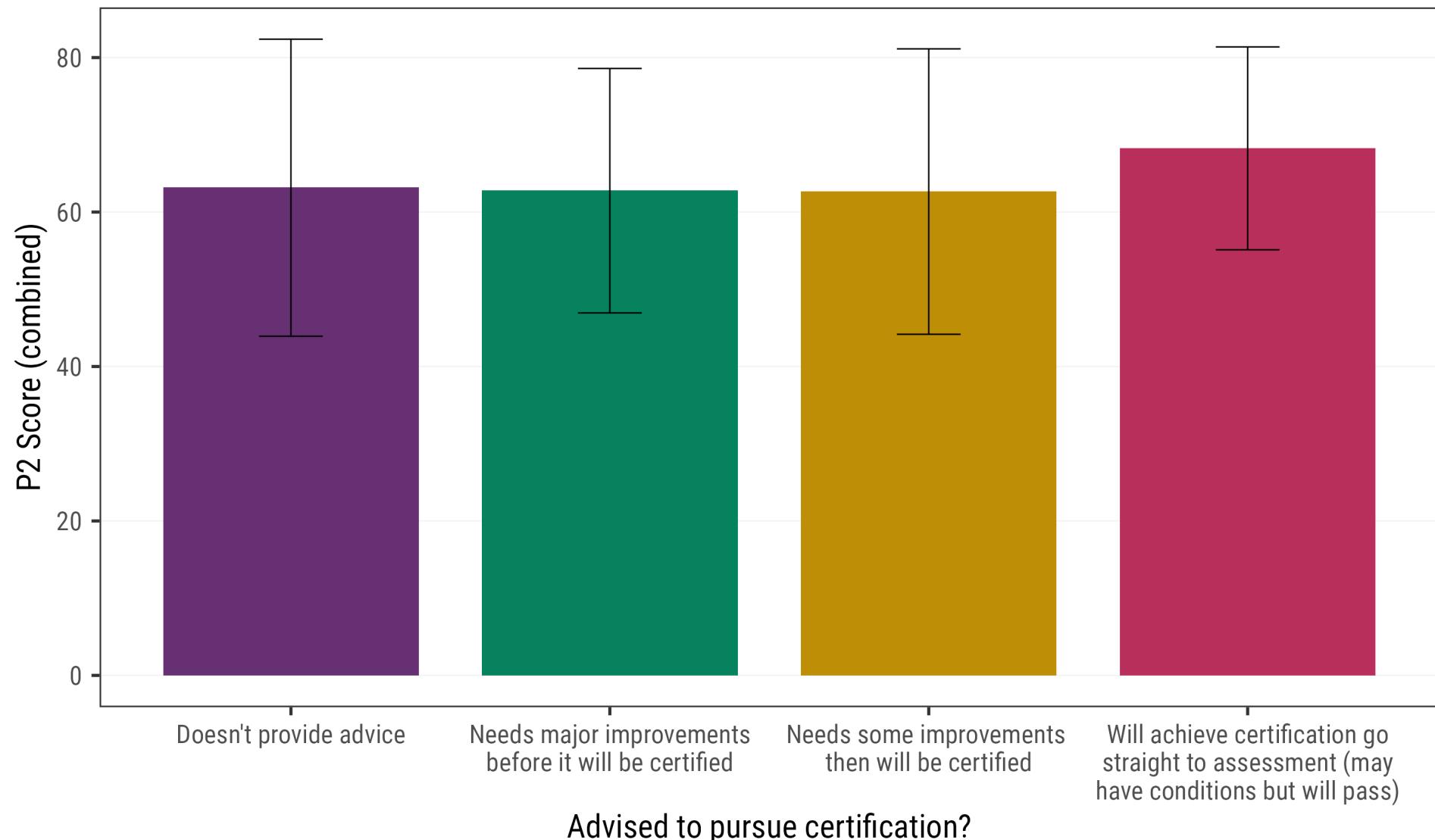
Is the information **conceptual** or **measurable**?

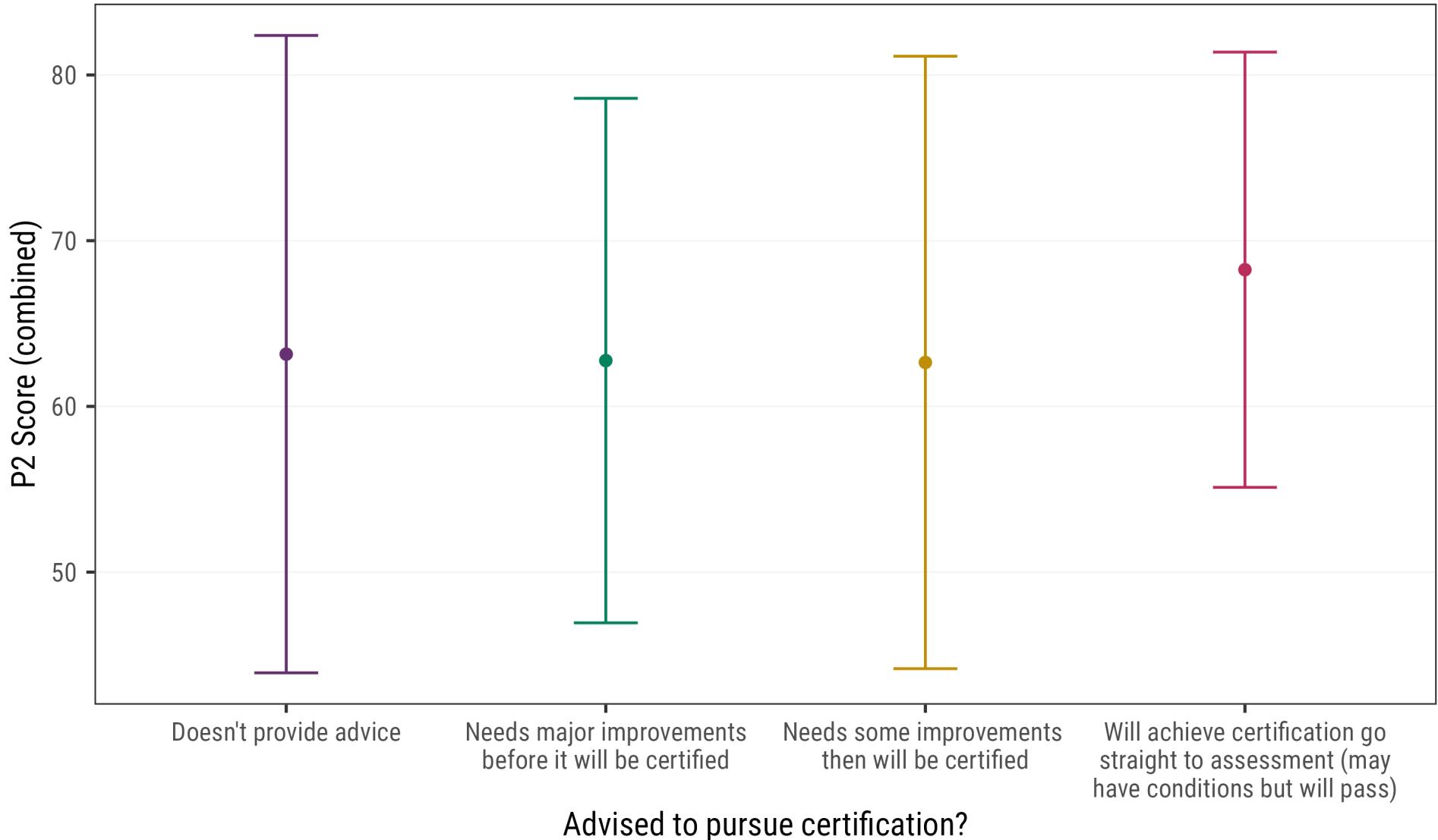
→ conceptual information *versus* convert information into visual forms

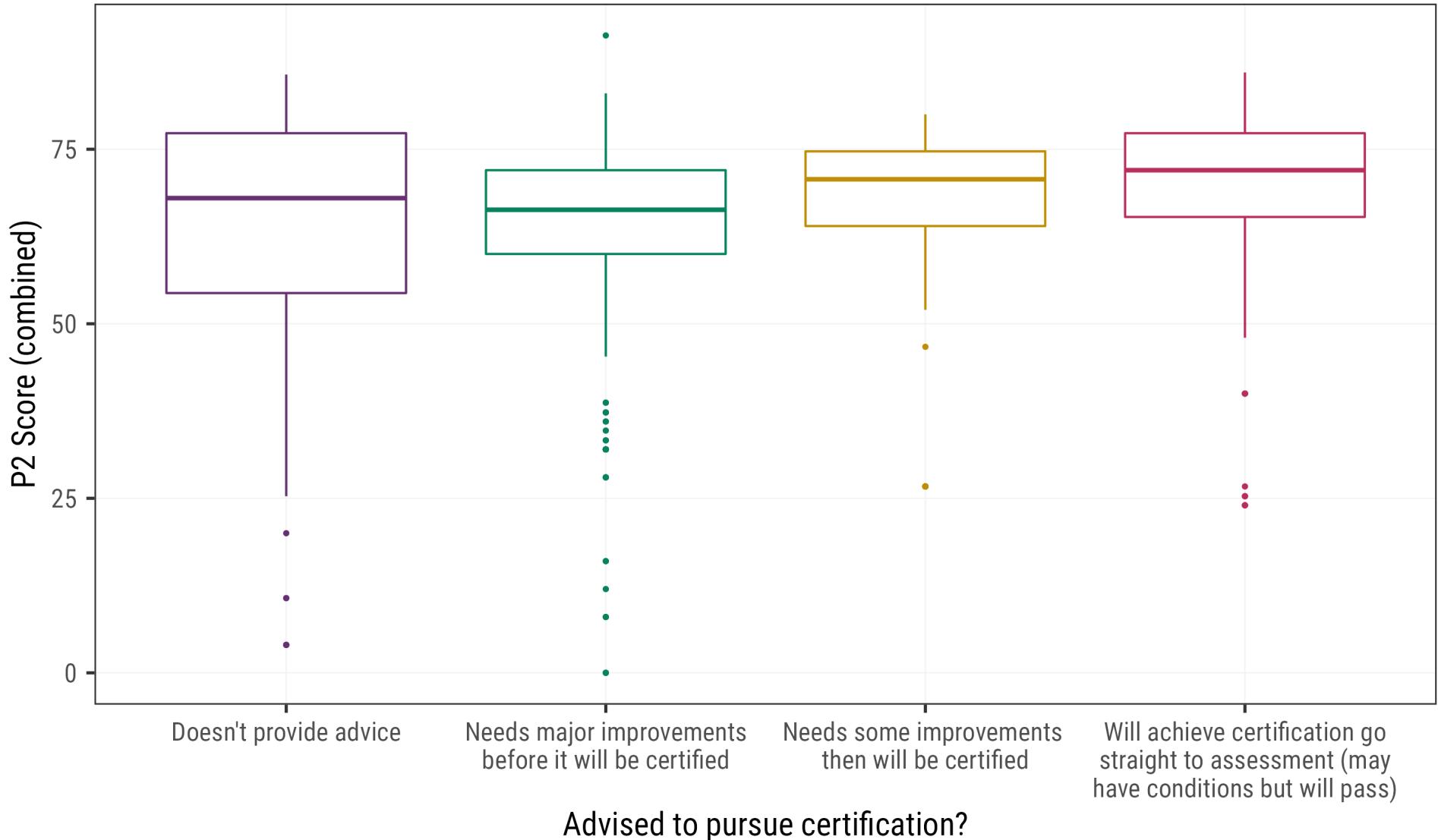
Is the purpose to **explain** or to **explore** the information?

→ Communicate information *versus* facilitate discovery

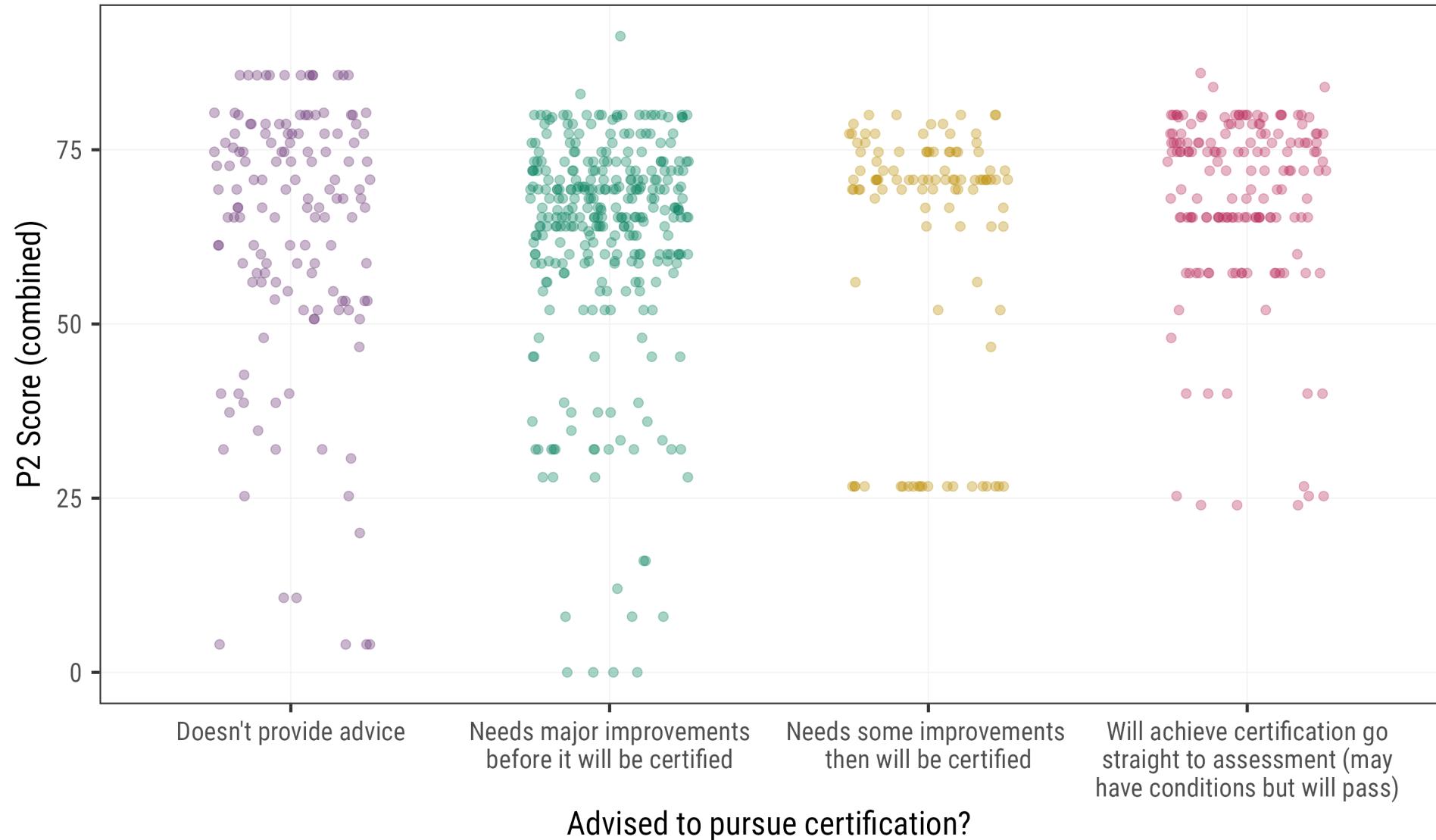


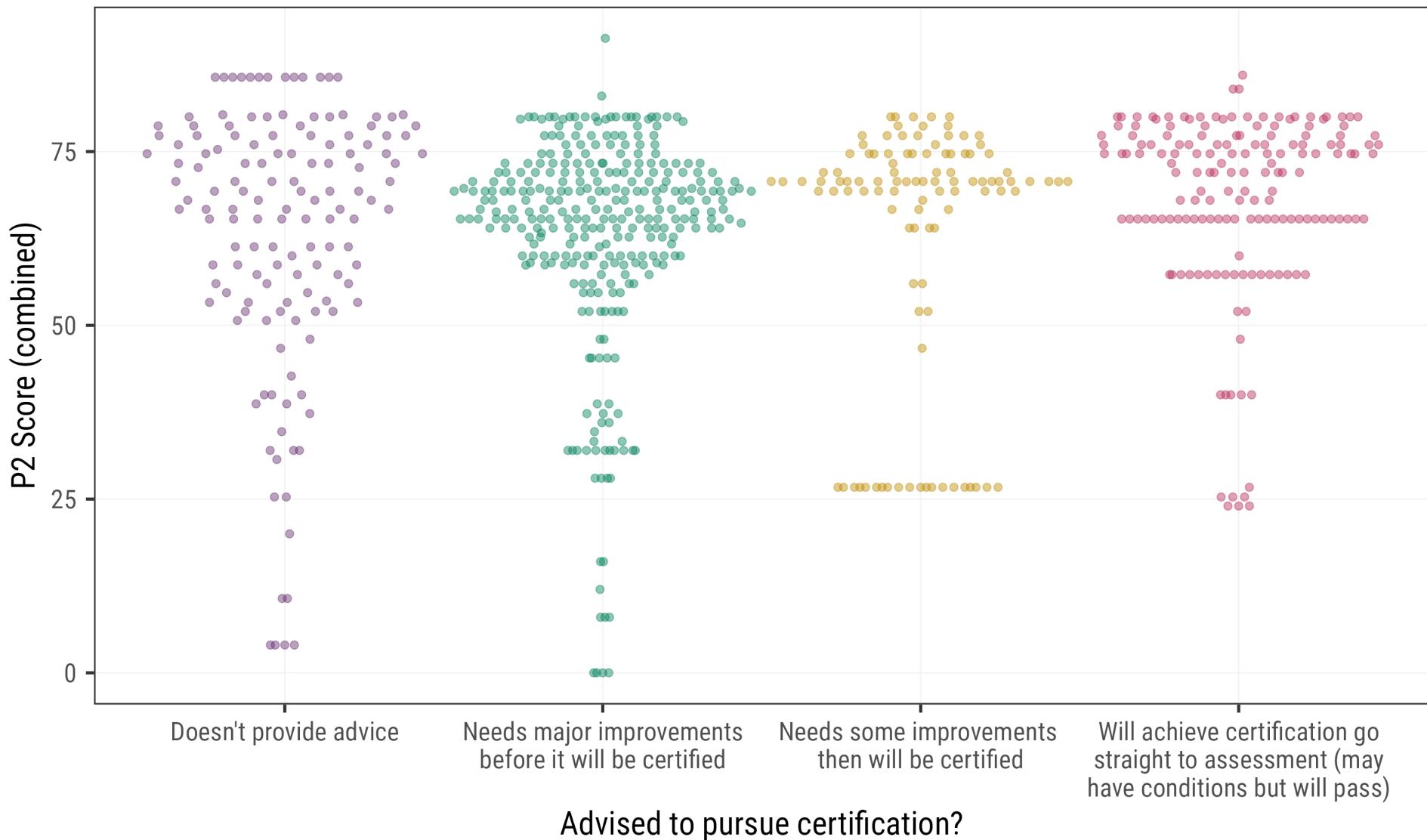


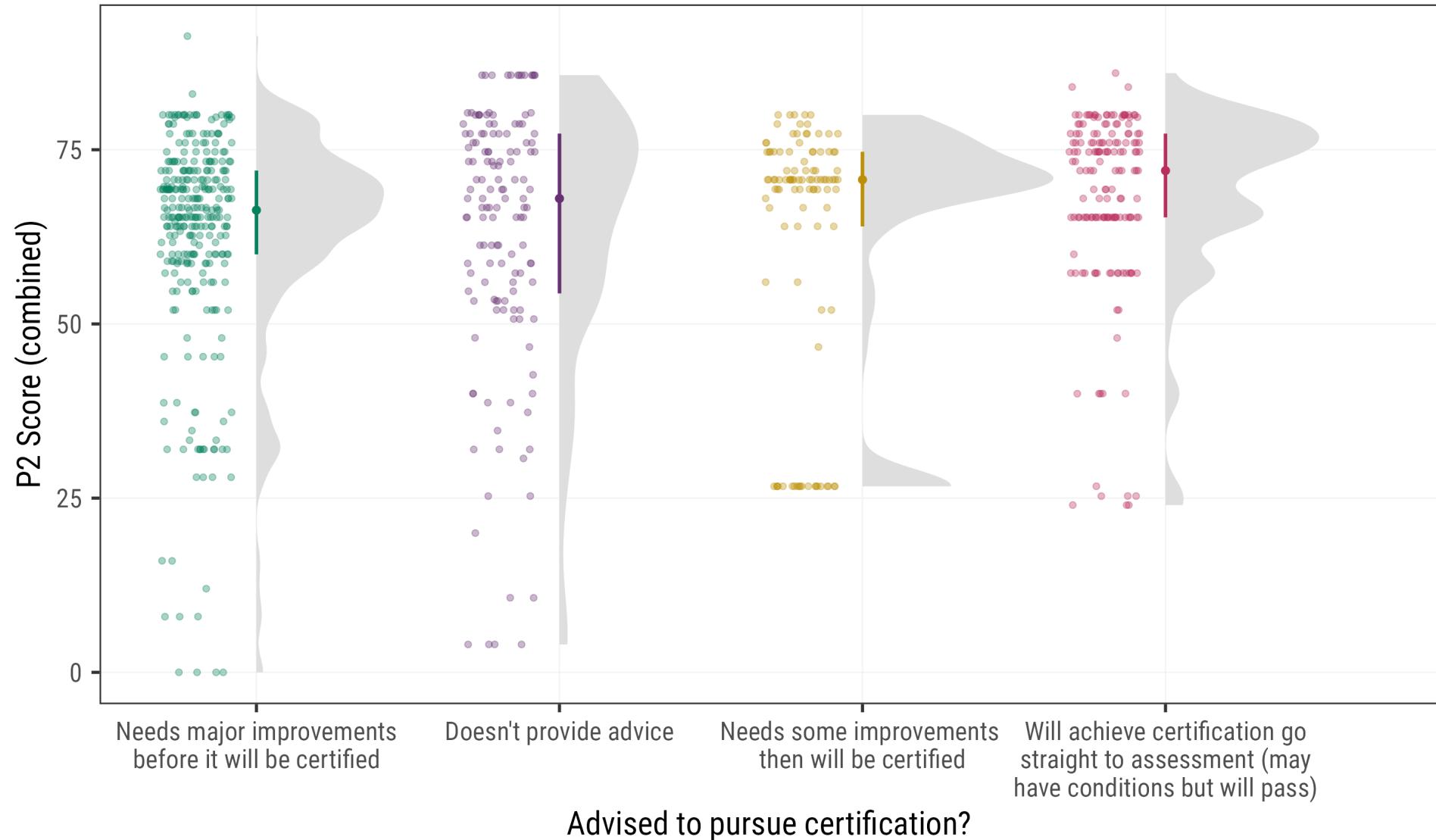


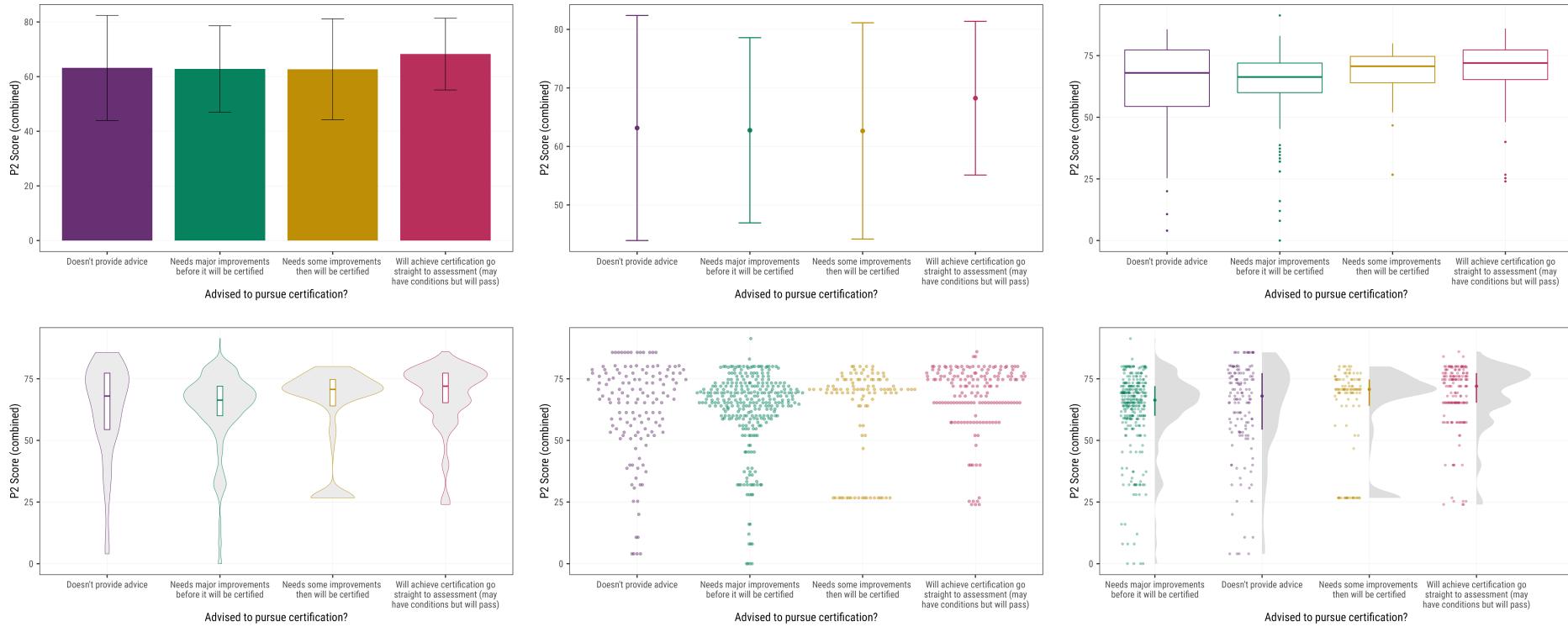












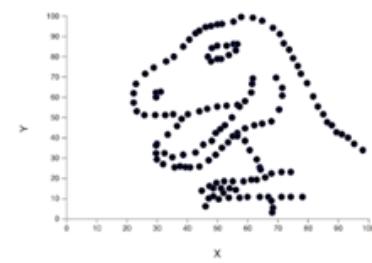
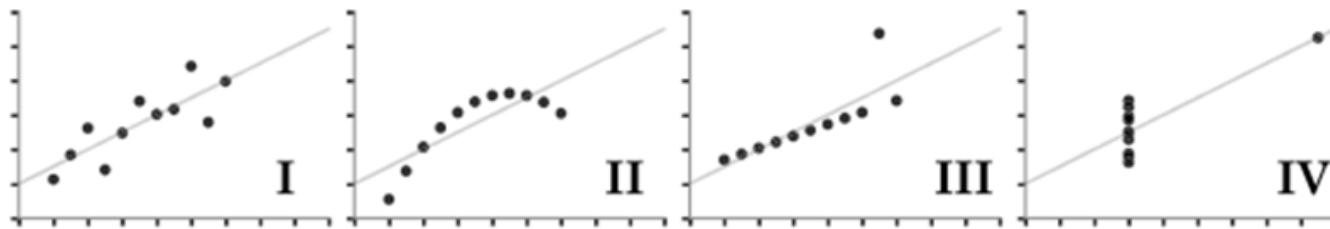
- Always check raw data and sample size
- Try several chart types
- Be open to combine chart types
- Choose chart type with your audience in mind

# Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing

Justin Matejka and George Fitzmaurice  
Autodesk Research, Toronto Ontario Canada

## "The Datasaurus Dozen"

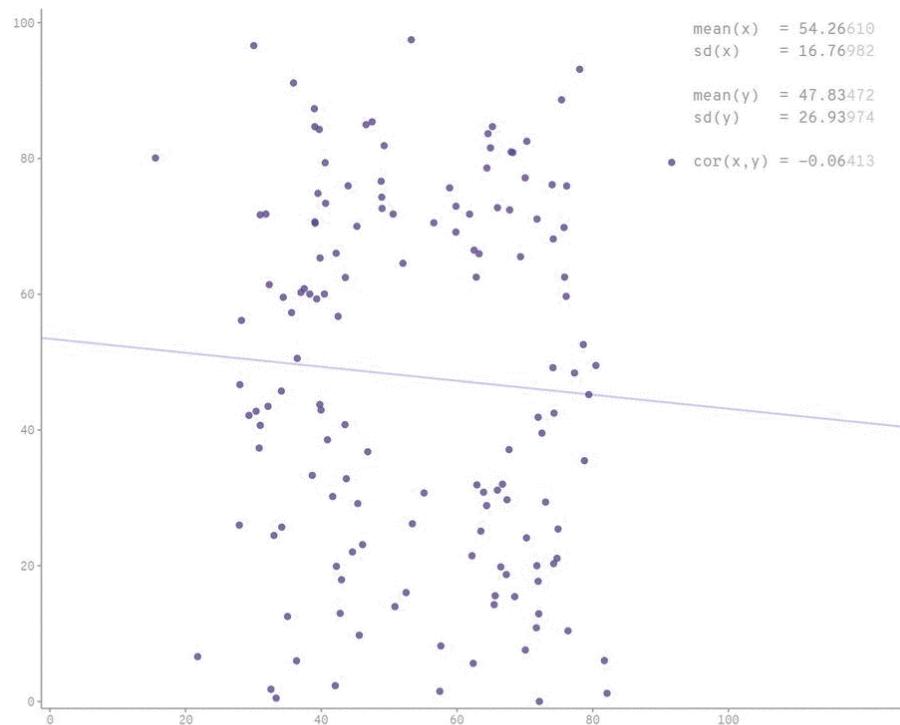
based on Anscombe's Quarter and Alberto Cairo's "Datasaurus" (or "Anscombosaurus")



# Same Stats. Different Graphs.

## The Famous Datasaurus (Anscombusaurus spec.) Dozen

is a set of 13 different datasets with high-identical summary statistics, which could lead one to believe the datasets are quite similar. After visualizing the data, it becomes clear that the datasets are markedly different. It is based on [Anscombe's Quartet](#) which was developed by F.J. Anscombe in 1973 to demonstrate the importance of data visualization. In 2016, Alberto Cairo created the Datasaurus dataset which urges people to never trust summary statistics alone; always visualize your data. This dataset was published together with a dozen other datasets with almost the same summary statistics by Justin Matejka & George Fitzmaurice in 2017.



Idea by Francis Anscombe, Alberto Cairo, Justin Matejka & George Fitzmaurice  
Visualization by Cédric Scherer based on code by Tom Westlake

PERSPECTIVE

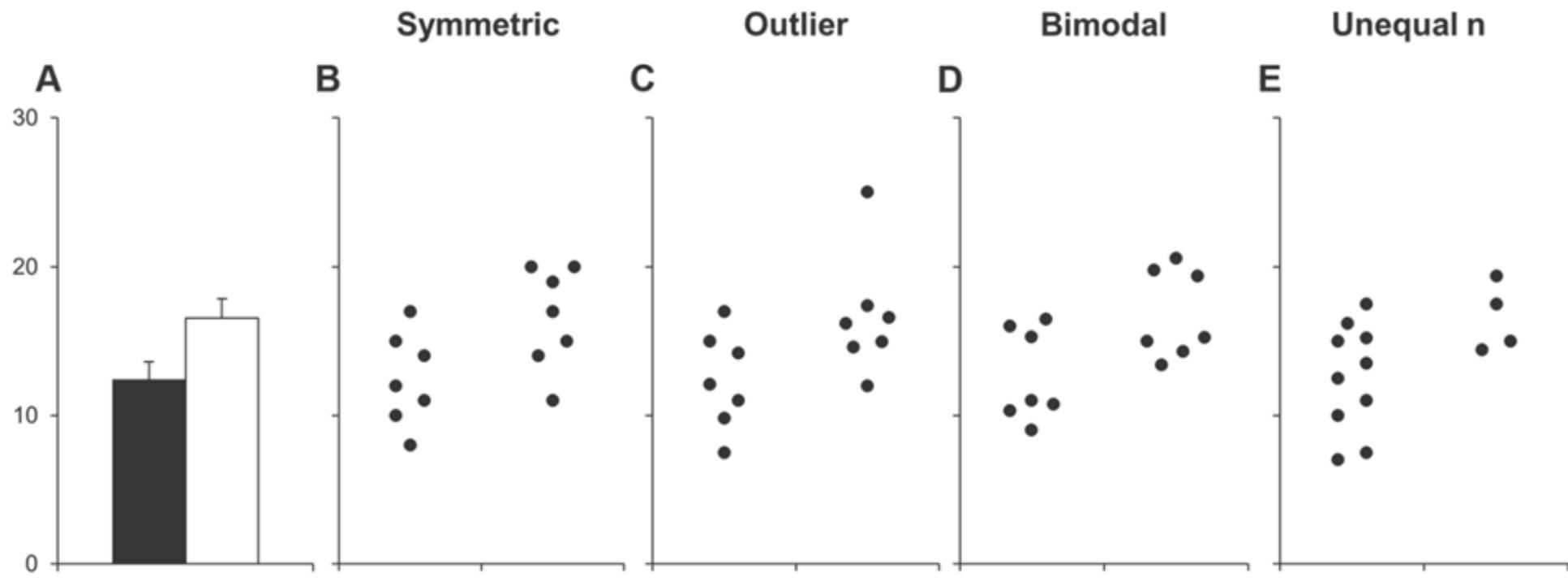
# Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

Tracey L. Weissgerber<sup>1\*</sup>, Nataša M. Milic<sup>1,2</sup>, Stacey J. Winham<sup>3</sup>, Vesna D. Garovic<sup>1</sup>

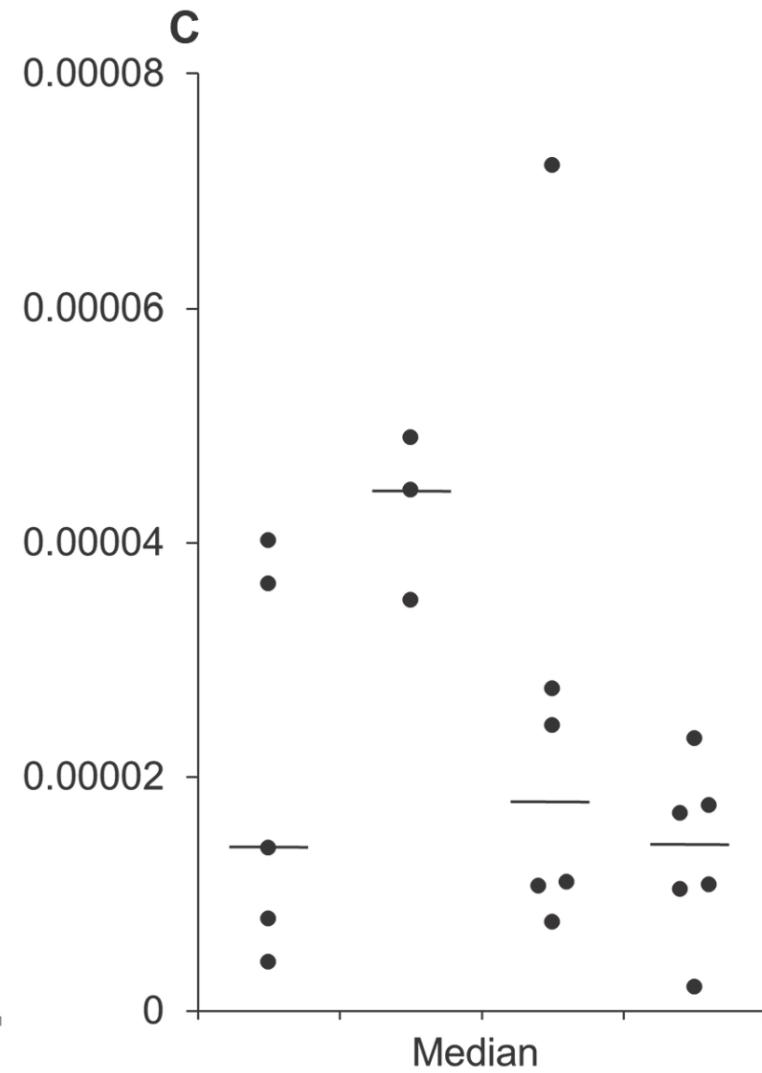
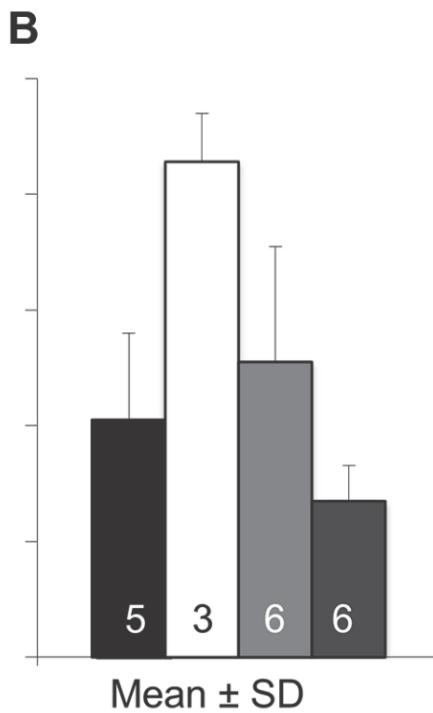
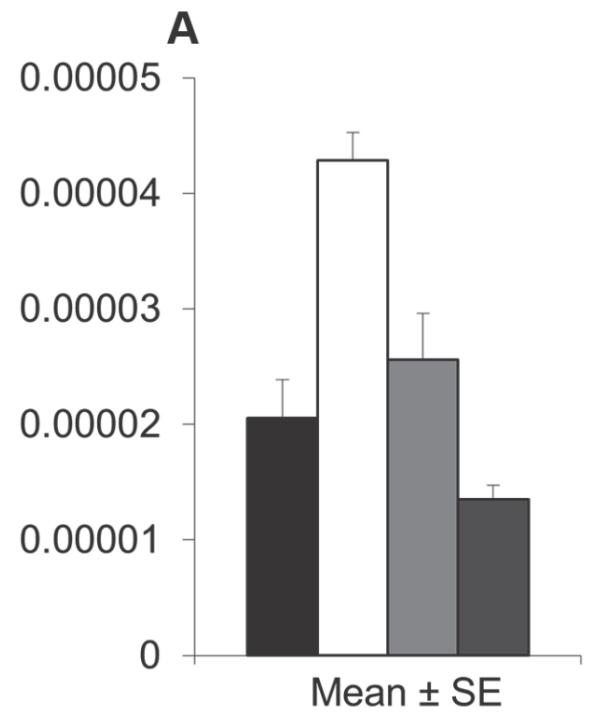
**1** Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America,

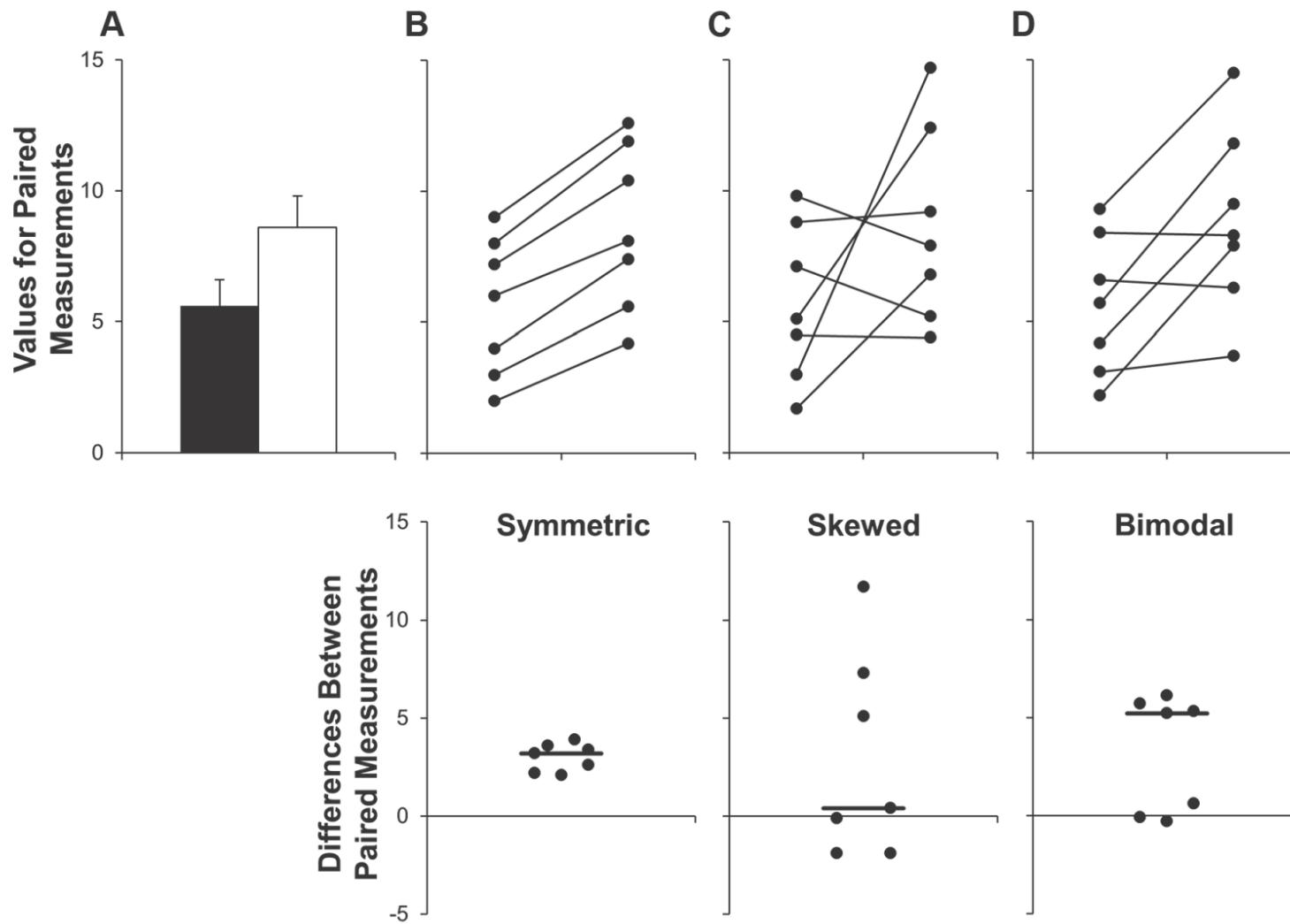
**2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, United States of America

\* [weissgerber.tracey@mayo.edu](mailto:weissgerber.tracey@mayo.edu)



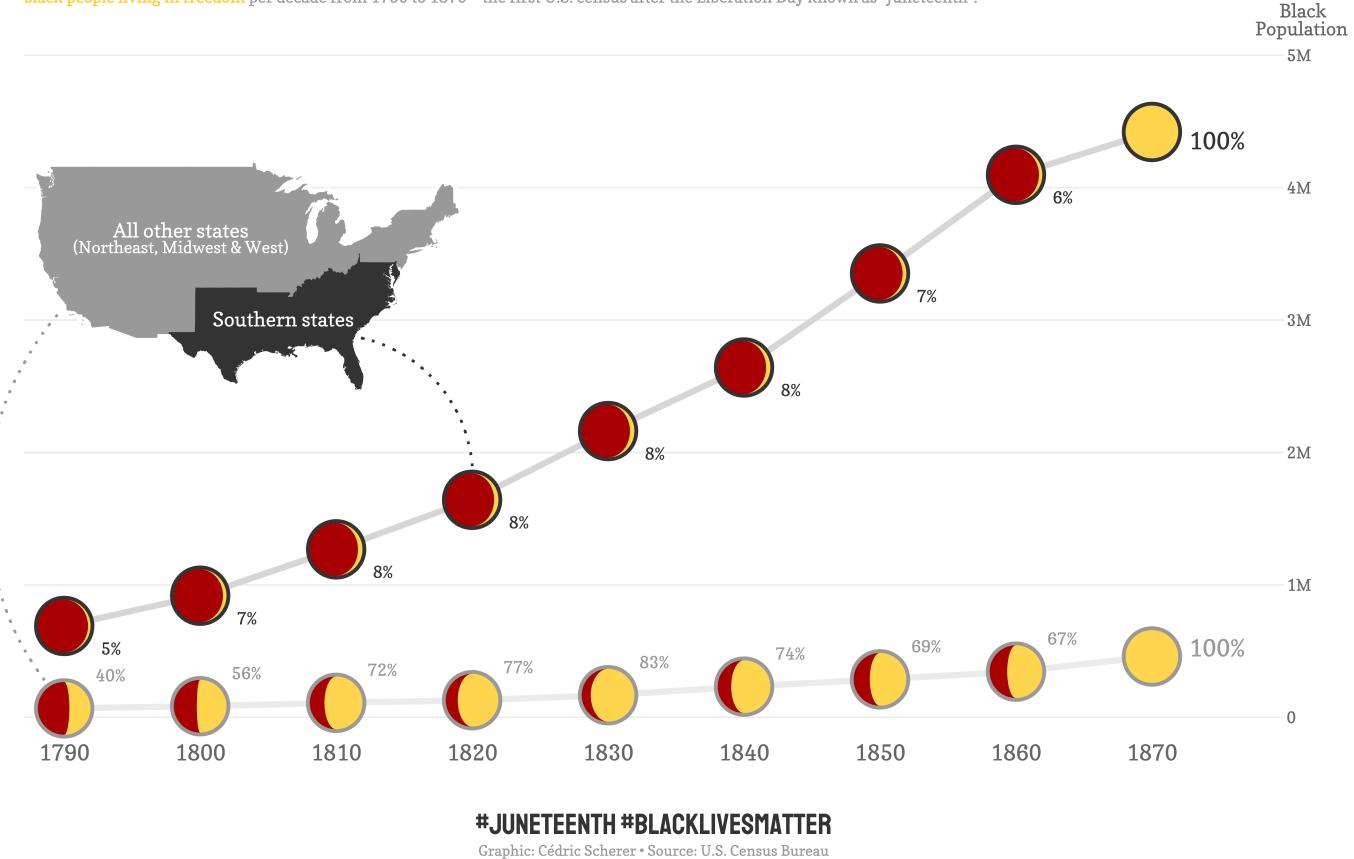
Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103





# AN ECONOMY BUILT ON SLAVERY—A FUTURE BUILD ON FREEDOM?

By 1680, 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 19<sup>th</sup> 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".



## 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 89.1 points (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

60 POINTS

GUATEMALA  
△59.8 POINTS

The coffee bean with the lowest rating has its origin in Guatemala.



One bean from Nicaragua got a bad rating, too.

70 POINTS

HONDURAS  
△69.2 POINTS

MEXICO  
△68.3 POINTS

NICARAGUA  
△63.1 POINTS

COSTA RICA  
△71.8 POINTS

HAWAII  
△73.7 POINTS

BRAZIL  
△73.2 POINTS

TAIWAN  
△77.7 POINTS

HONDURAS  
△81.7 POINTS

MEXICO  
△81.6 POINTS

NICARAGUA  
△80.8 POINTS

ETHIOPIA  
△80.3 POINTS  
▲85.1 POINTS

The best coffee—in terms of both median and maximum rating—is shipped to you from Ethiopia!

KENYA  
△79.8 POINTS  
▲84.6 POINTS  
COLOMBIA  
△72.8 POINTS  
▲83.2 POINTS

UGANDA  
△80.5 POINTS  
▲83.2 POINTS

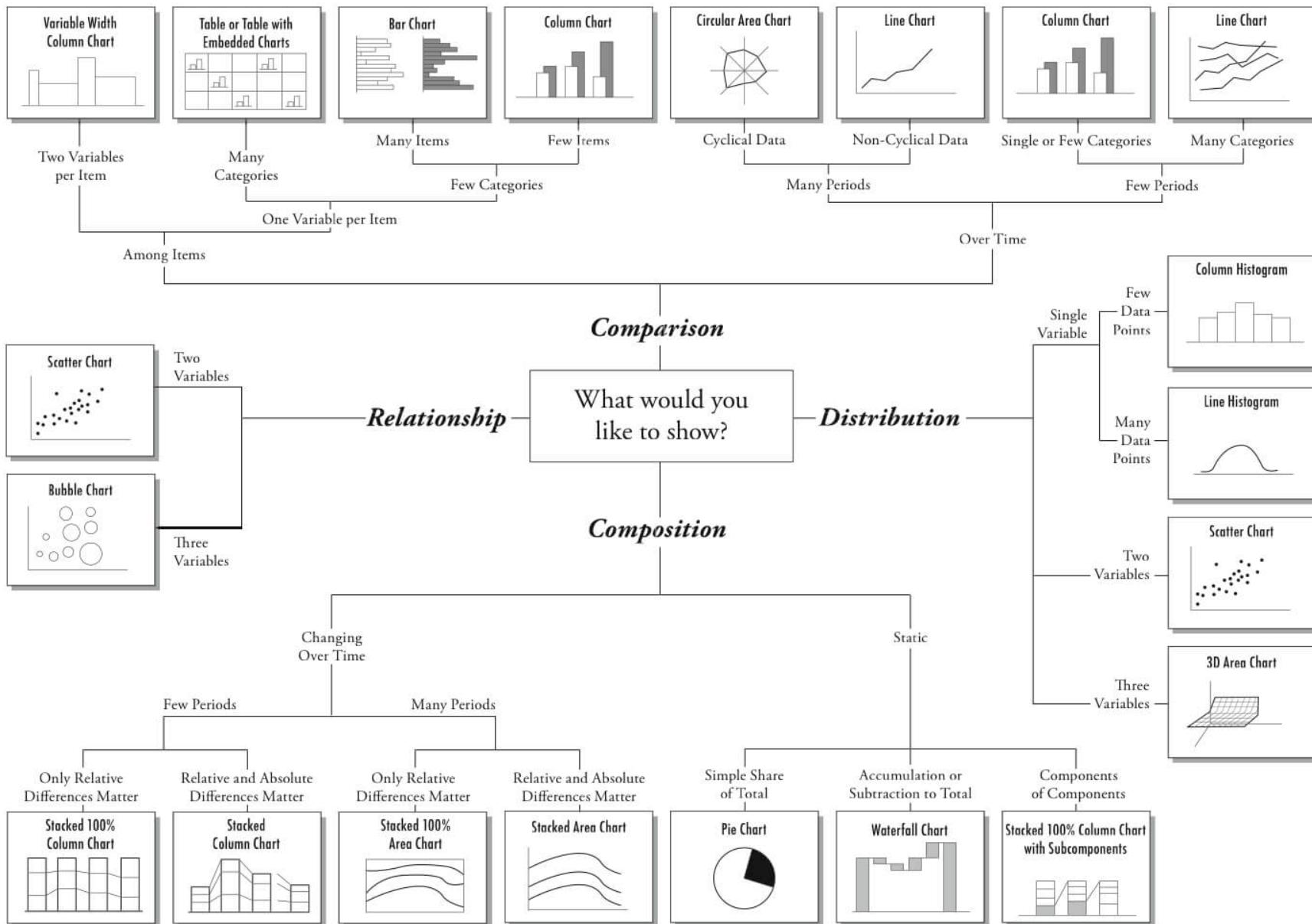
90 POINTS

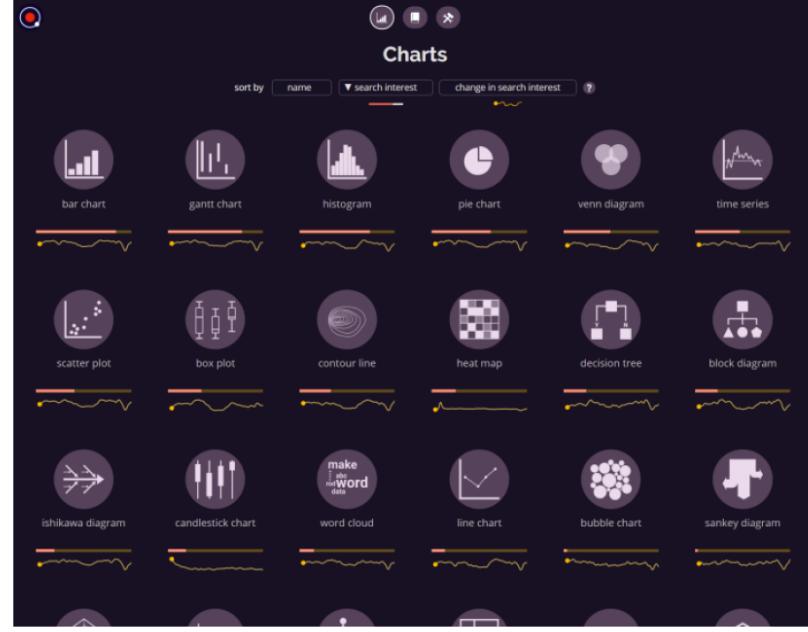
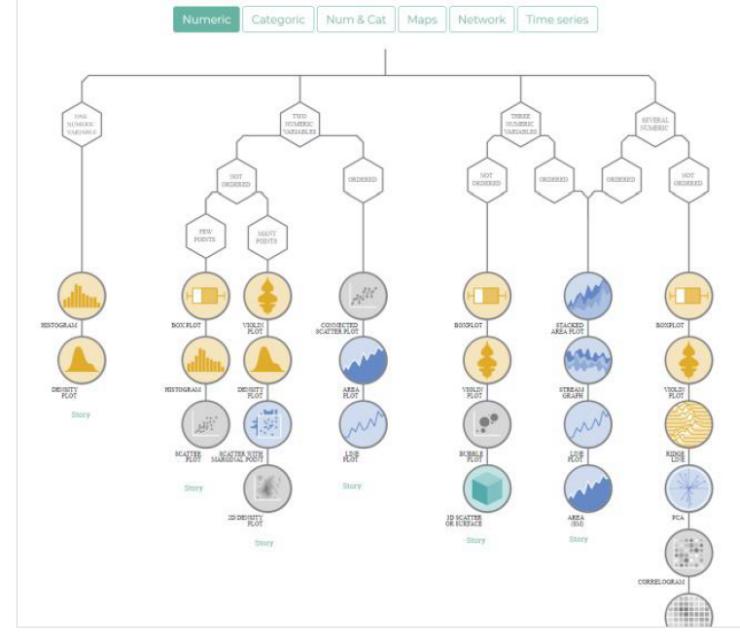
TANZANIA  
△80.3 POINTS  
▲82.2 POINTS

With 218 tested beans, Mexico is the country with the most reviews.

#TidyTuesday contribution Week 2020/28

# Chart Suggestions—A Thought-Starter







# from Data to Viz

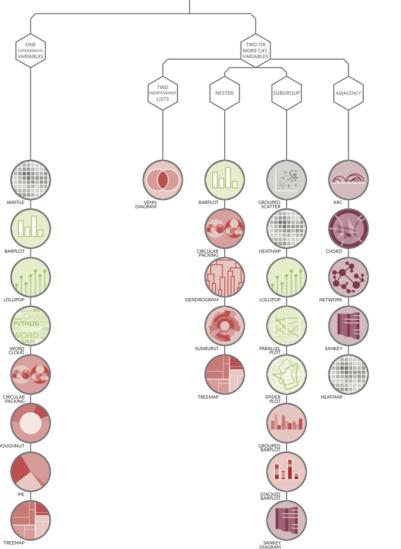
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- 1** Identify what type of data you have.
- 2** Go to the corresponding decision tree and follow it down to a set of possible charts.
- 3** Choose the chart from the set that will suit your data and your needs best.

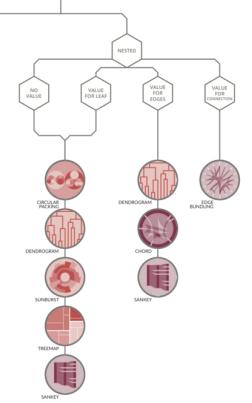
DataViz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

[data-to-viz.com](http://data-to-viz.com)

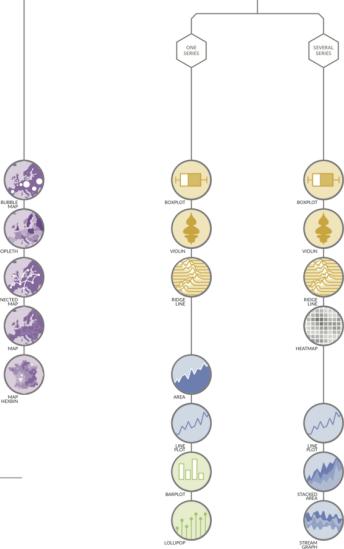
## CATEGORIC



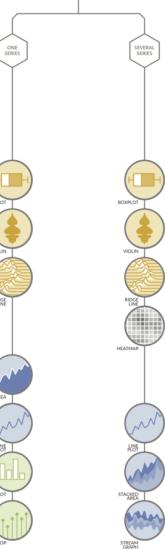
## RELATIONAL



## MAP



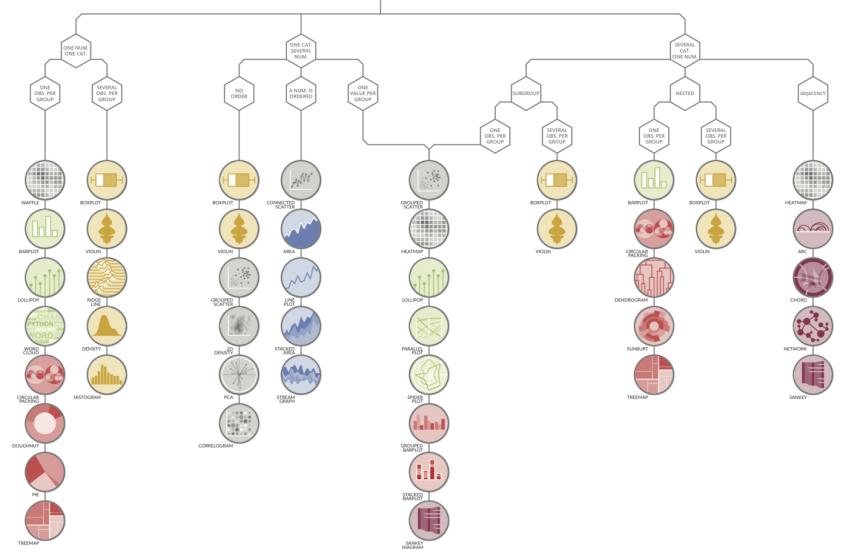
## TIME SERIES



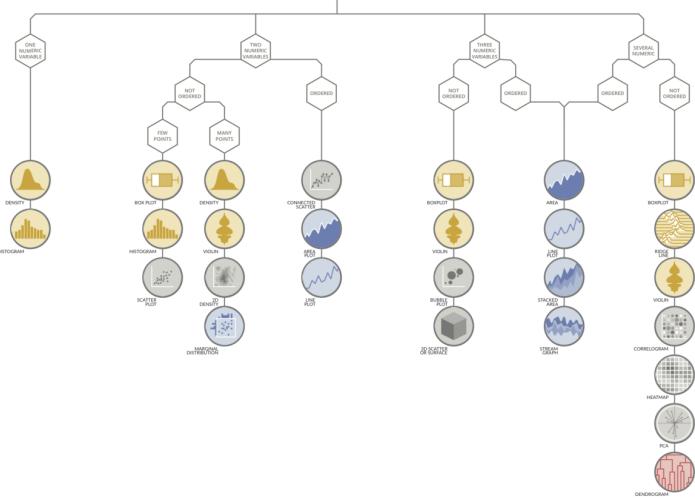
### WHAT DO YOU WANT TO SHOW ?

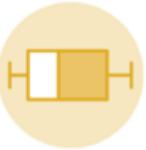
- |   |   |
|---|---|
| <span style="color: gold;">●</span> Distribution    | <span style="color: blue;">●</span> Evolution |
| <span style="color: darkgrey;">●</span> Correlation | <span style="color: purple;">●</span> Maps    |
| <span style="color: green;">●</span> Ranking        | <span style="color: maroon;">●</span> Flow    |

## CATEGORIC AND NUMERIC



## NUMERIC





## BOXPLOT

Summarize the distribution of numeric variables

[About](#)

A boxplot gives a nice summary of one or several numeric variables. The line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers.

[Common Mistakes](#)

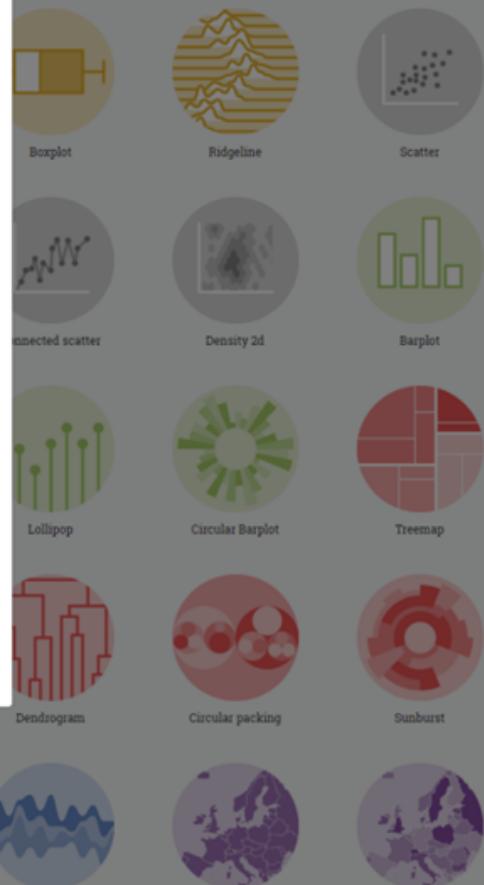
- Boxplot hides the sample size of each group, show it with annotation or box width.
- Boxplot hides the underlying distribution. Use jitter if low number of data points, or use violin with bigger data.
- Order your boxplot by median can make it more insightful.

[Code](#)

[R graph gallery](#) [Python gallery](#) [D3.js gallery](#) [Flourish](#)

[Read More](#)

See the [dedicated page](#).

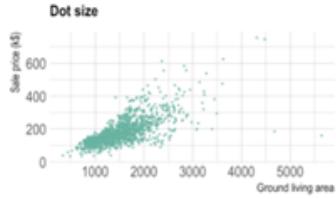


**VISIBILITIES**  
presented in this website.

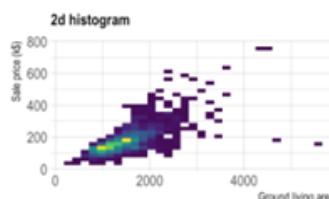
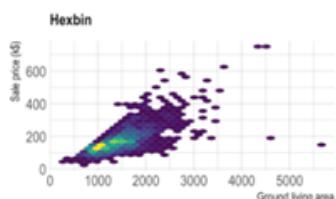
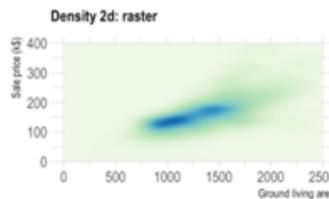
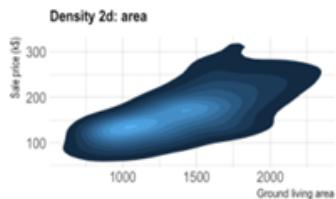
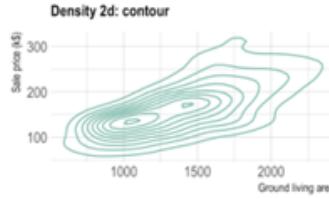
[Part of a whole](#) [Evolution](#) [Map](#) [Flow](#)

## Overplotting

The most common pitfall with scatterplot is overplotting: when the sample size gets big, dots are plotted on top of each other what makes the chart unreadable. There are several work around to avoid this issue as describe in this [specific post](#). Here is a summary of the different offered techniques:



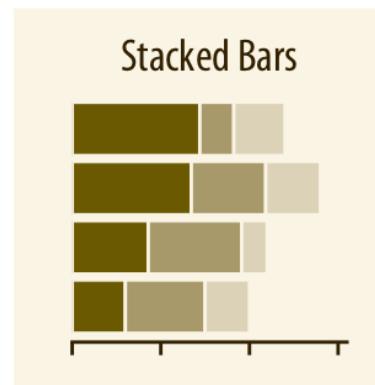
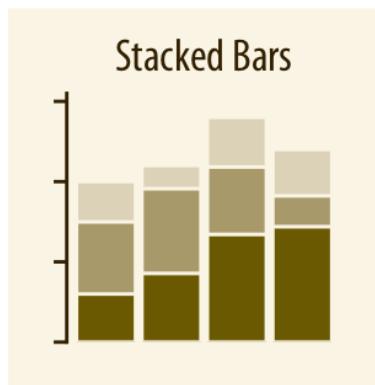
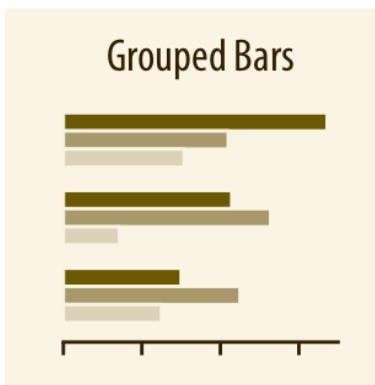
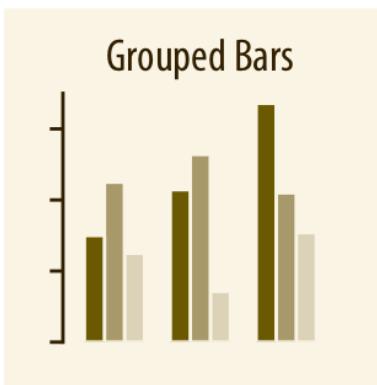
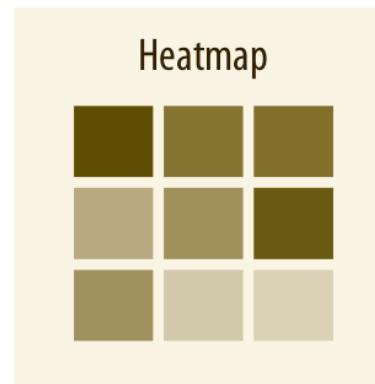
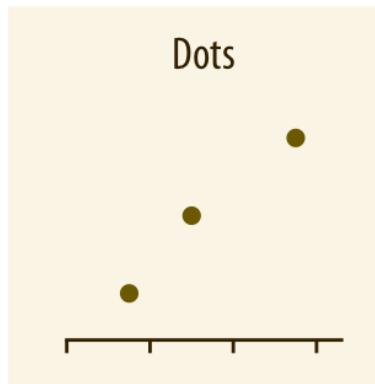
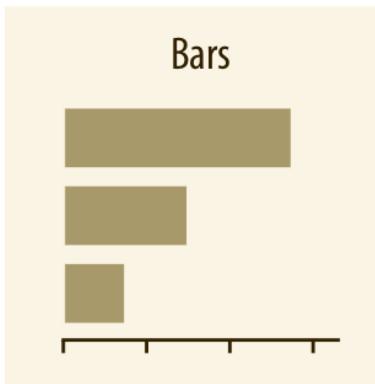
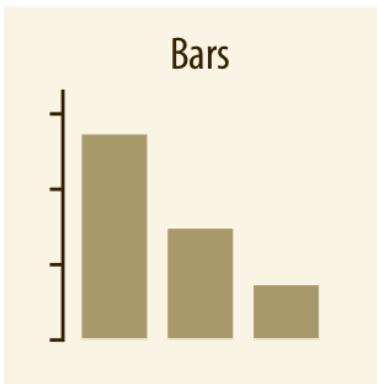
CODE



## Going further

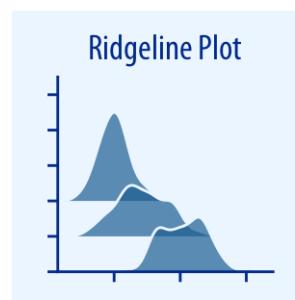
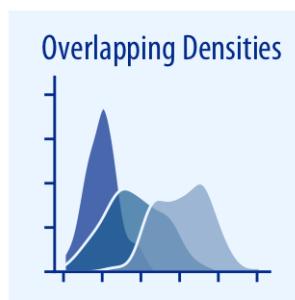
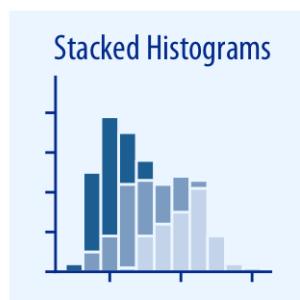
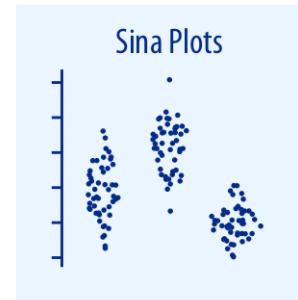
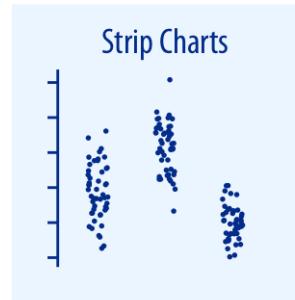
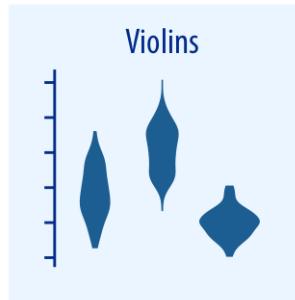
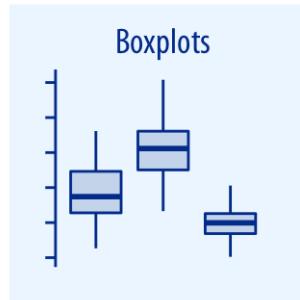
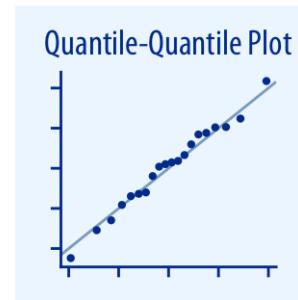
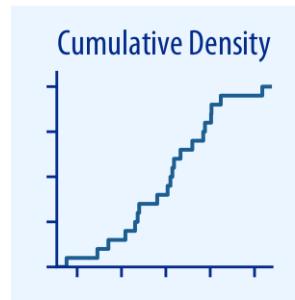
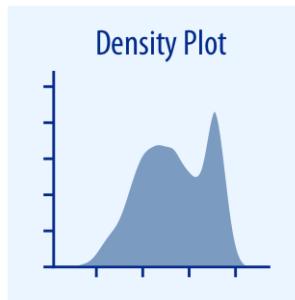
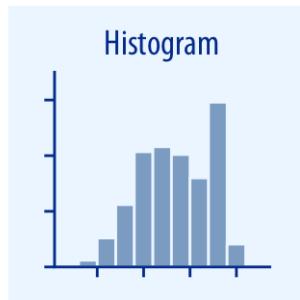
You can learn more about each type of graphic presented in this story in the dedicated

# Charts to Visualize Amounts



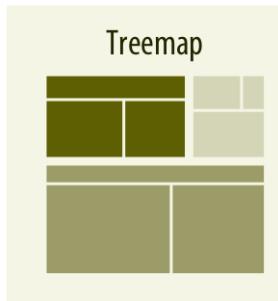
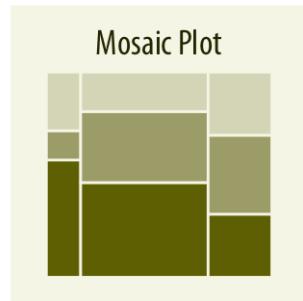
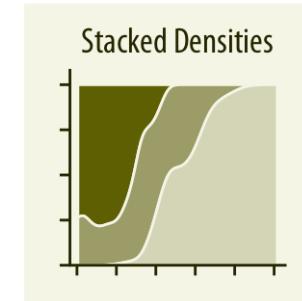
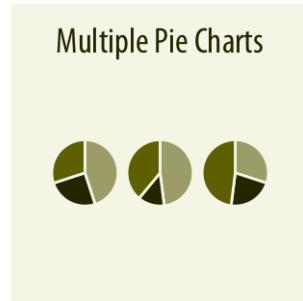
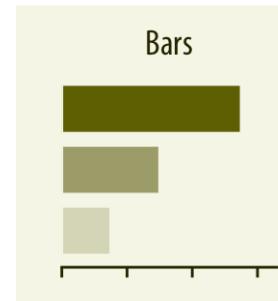
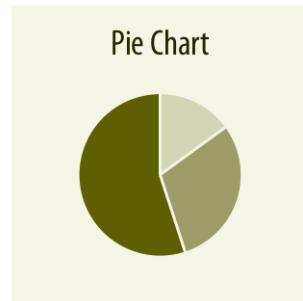
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Charts to Visualize Distributions



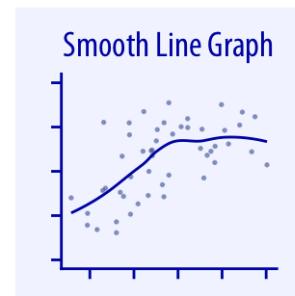
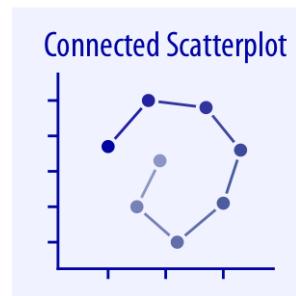
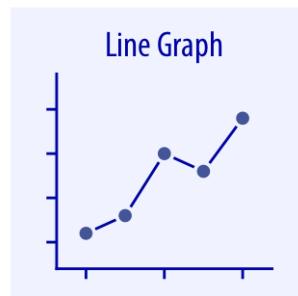
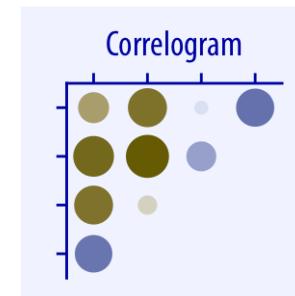
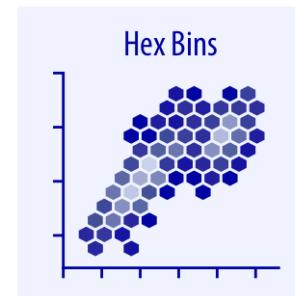
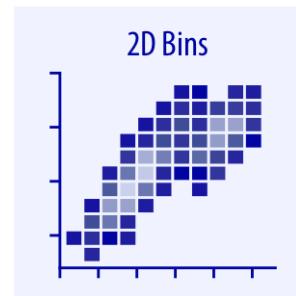
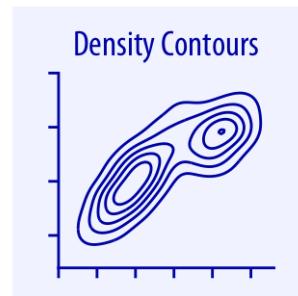
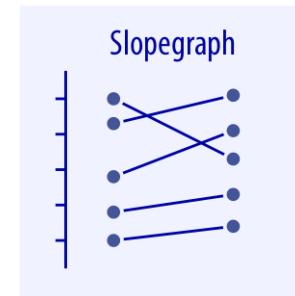
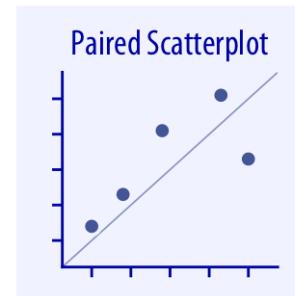
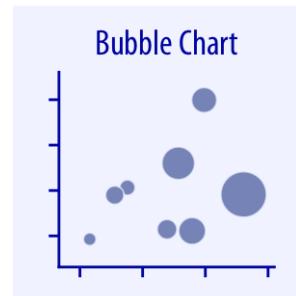
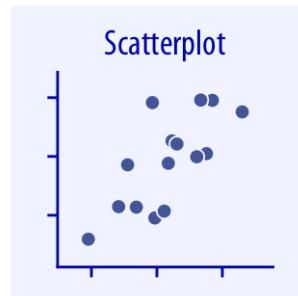
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Charts to Visualize Proportions



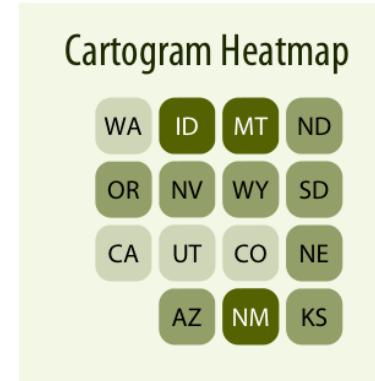
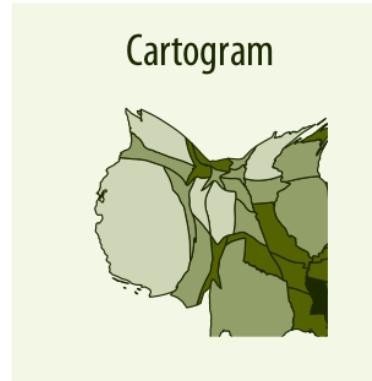
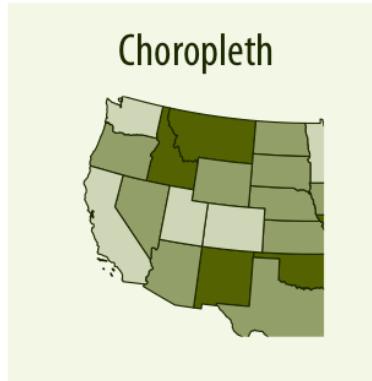
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Charts to Visualize x-y Relationships



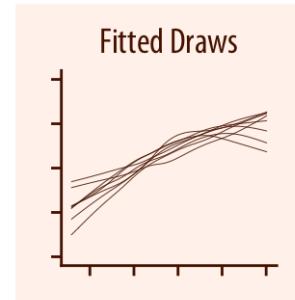
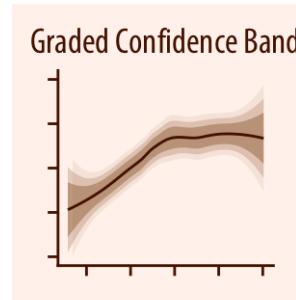
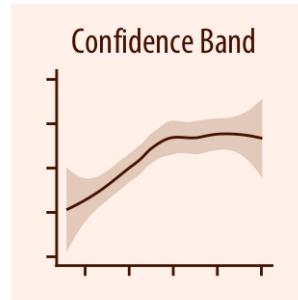
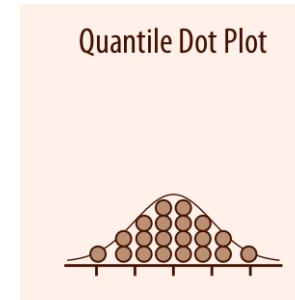
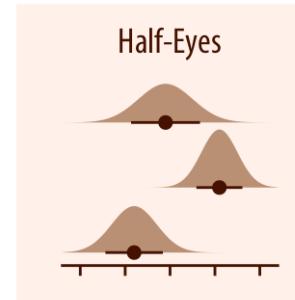
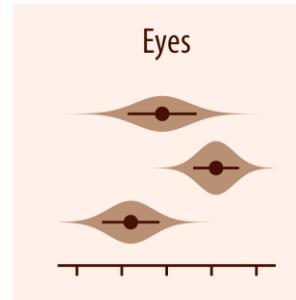
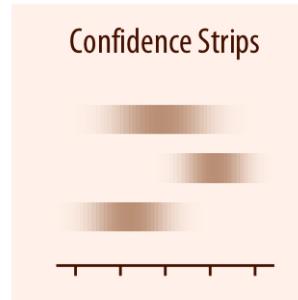
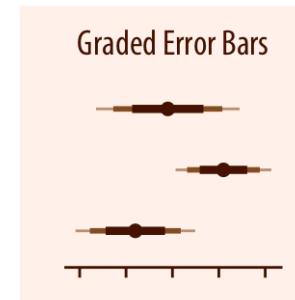
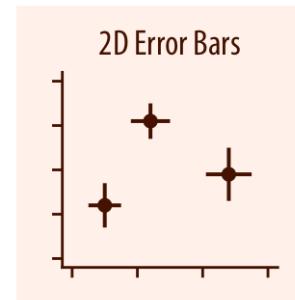
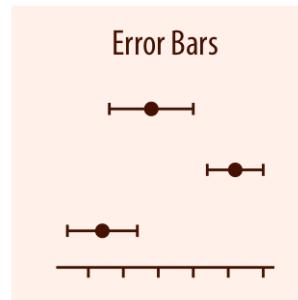
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Charts to Visualize Geospatial Data



*Source: "Fundamentals of Data Visualization" by Claus Wilke*

# Charts to Visualize Uncertainty



Source: "Fundamentals of Data Visualization" by Claus Wilke

# VISUAL FORM

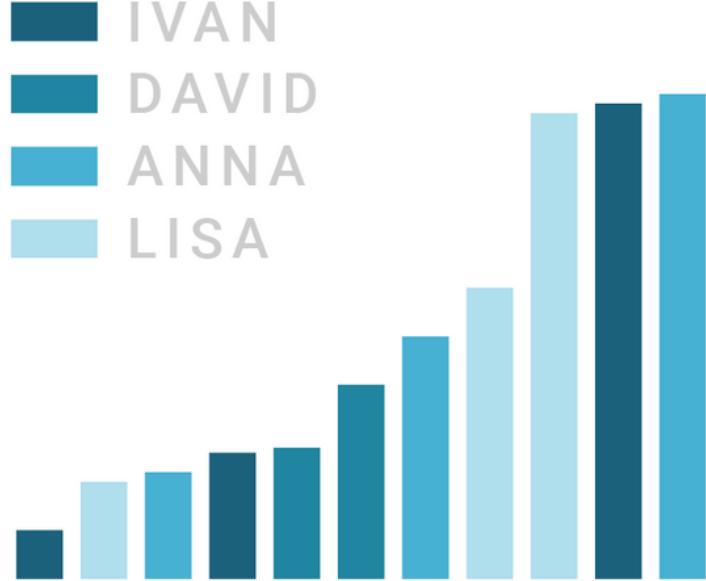
*Follow design rules and data visualization principles*



# Colors and Common Pitfalls

# Color Choice & Accessibility

NOT IDEAL



*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 Palette Types

Categorical



# Color Palette Types

Categorical



Sequential: Single-Hue



# Color Palette Types

Categorical



Sequential: Single-Hue



Sequential: Multi-Hue



# Color Palette Types

Categorical



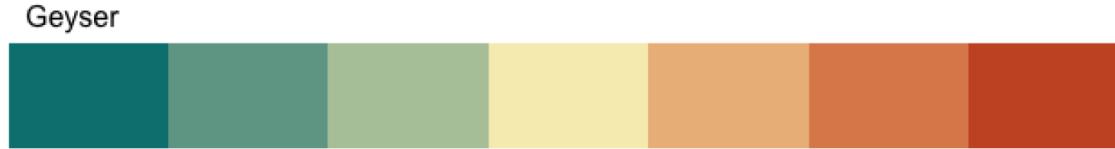
Sequential: Single-Hue



Sequential: Multi-Hue



Diverging



# Color Palette Types

Categorical



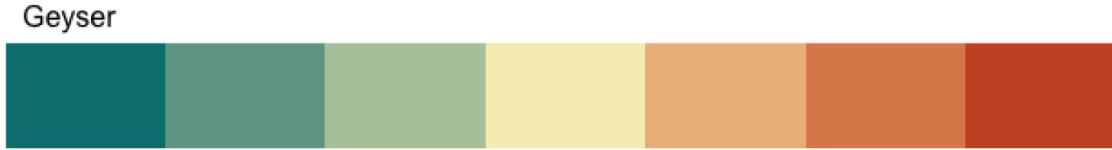
Sequential: Single-Hue



Sequential: Multi-Hue



Diverging



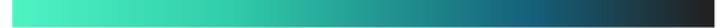
Cyclical



# Color Choice & Accessibility

NOT IDEAL

POOR                    RICH



PEOPLE IN GROUP A

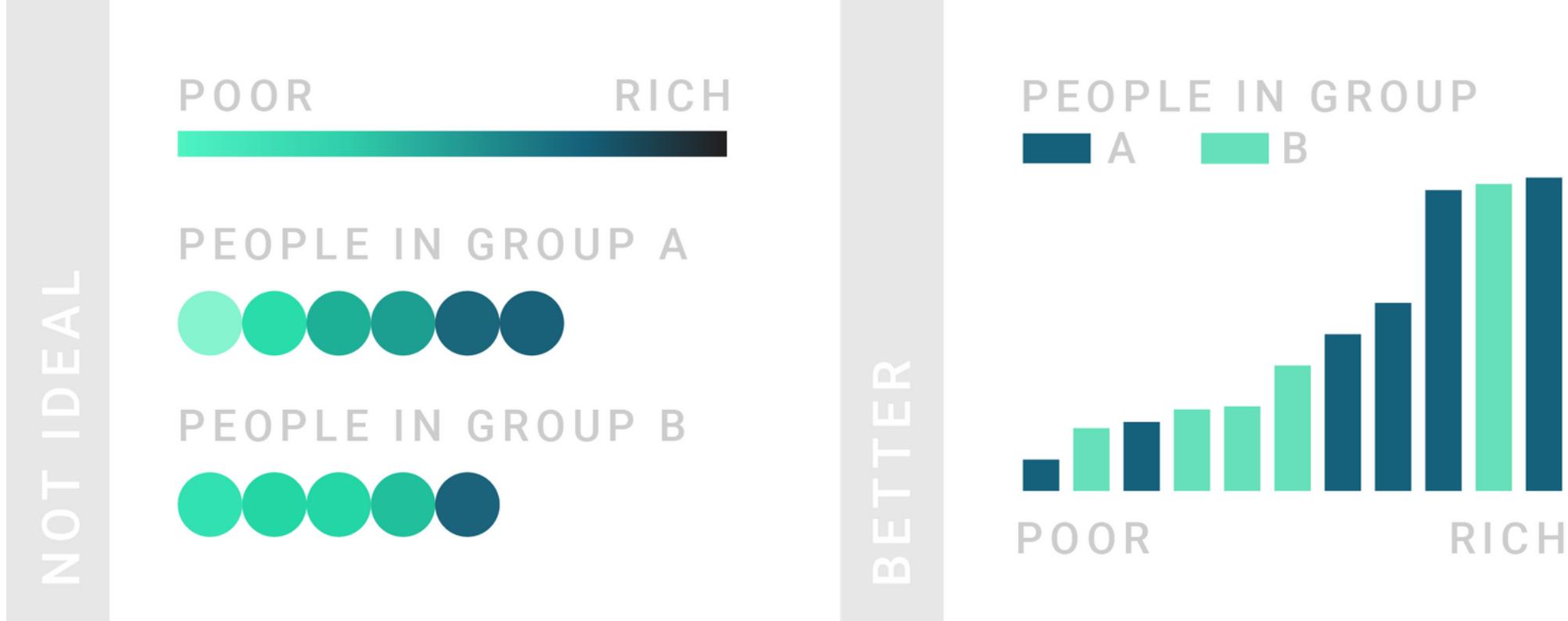


PEOPLE IN GROUP B



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

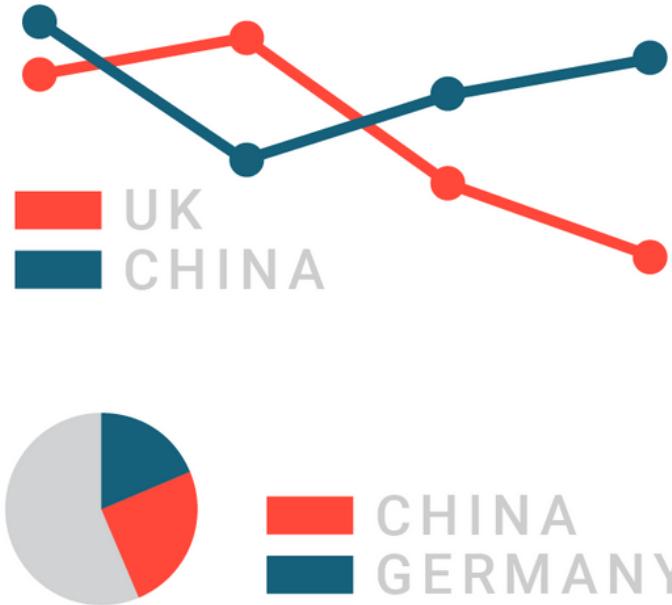
# Color Choice & Accessibility



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

# Color Choice & Accessibility

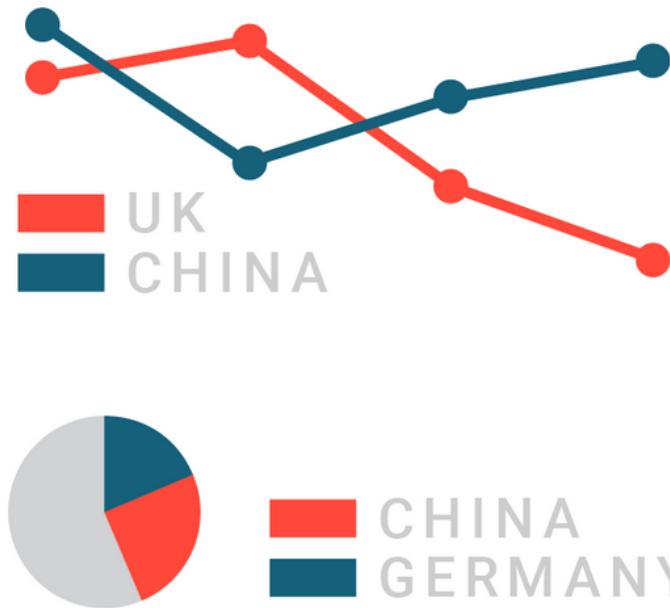
NOT IDEAL



*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

## 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

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](#)<sup>a)</sup>

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](#)

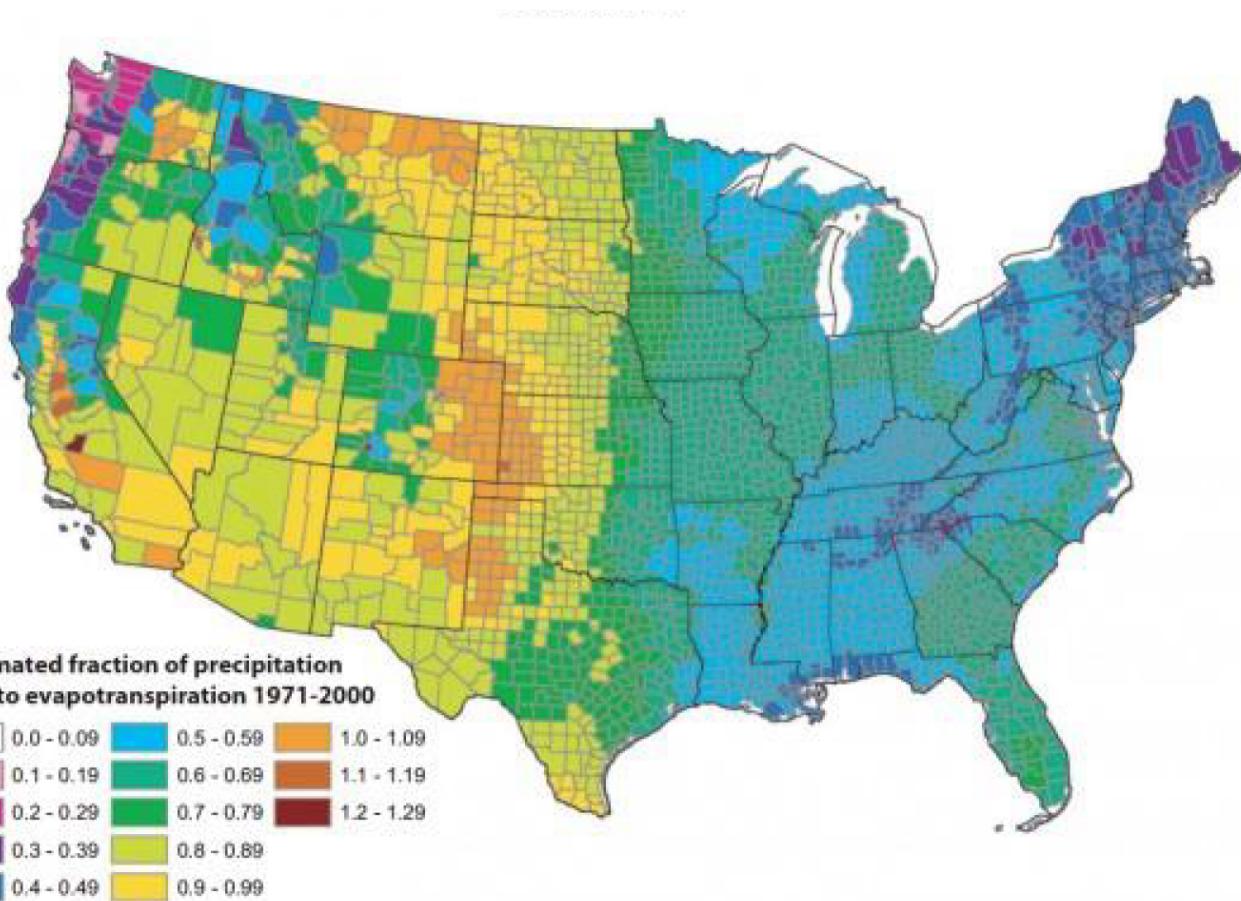


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of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

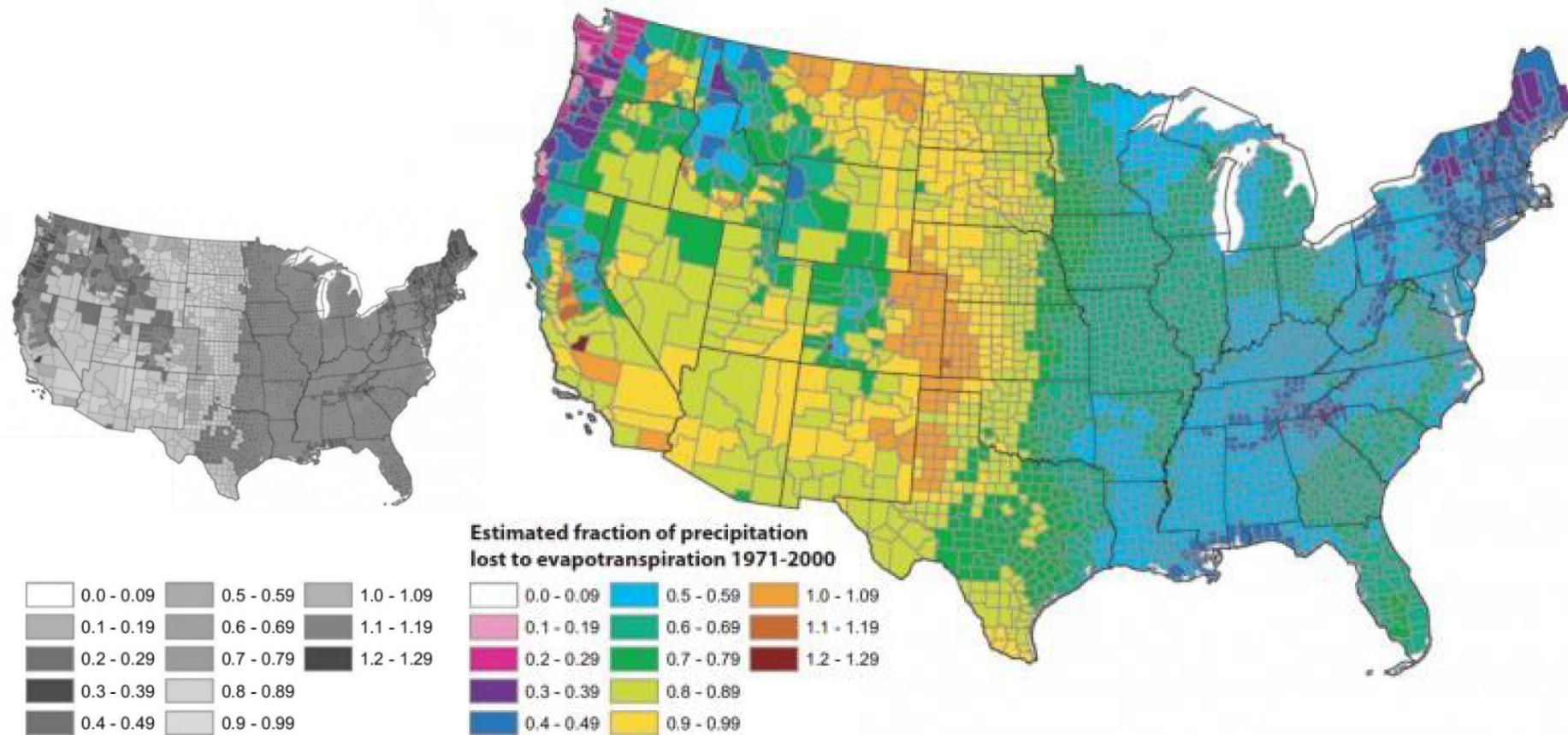


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of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

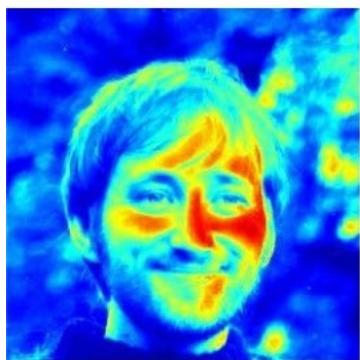


**true-colour Phil**

*Source: [Fabio Cramer](#)*



**true-colour Phil**

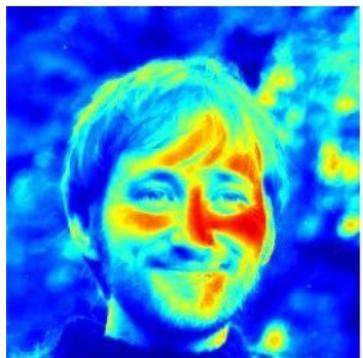


**rainbow Phil**  
*is distorted*

*Source: Fabio Cramer*



**true-colour Phil**



**rainbow Phil**  
*is distorted*



**batlow Phil**  
*is flawless*

*Source: Fabio Cramer*

# Color Choice & Accessibility



*Source: cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html*

# Color Choice & Accessibility



**Deutanopia:** present in 6% of males



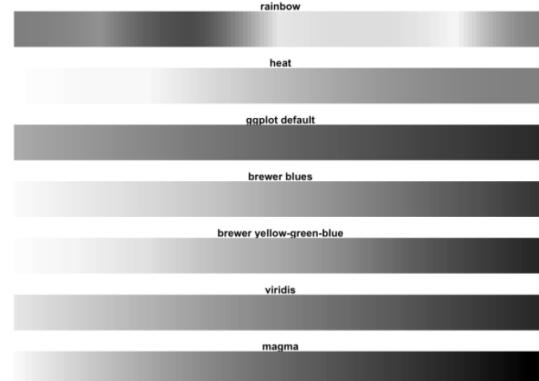
**Protanopia:** present in 1% of males



**Tritanopia:** present in 0.008% of humans



**Monochromacy:** present in 0.001% of humans



... and present in ~75% of university printers! ☺

Modified from [cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html](http://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html)

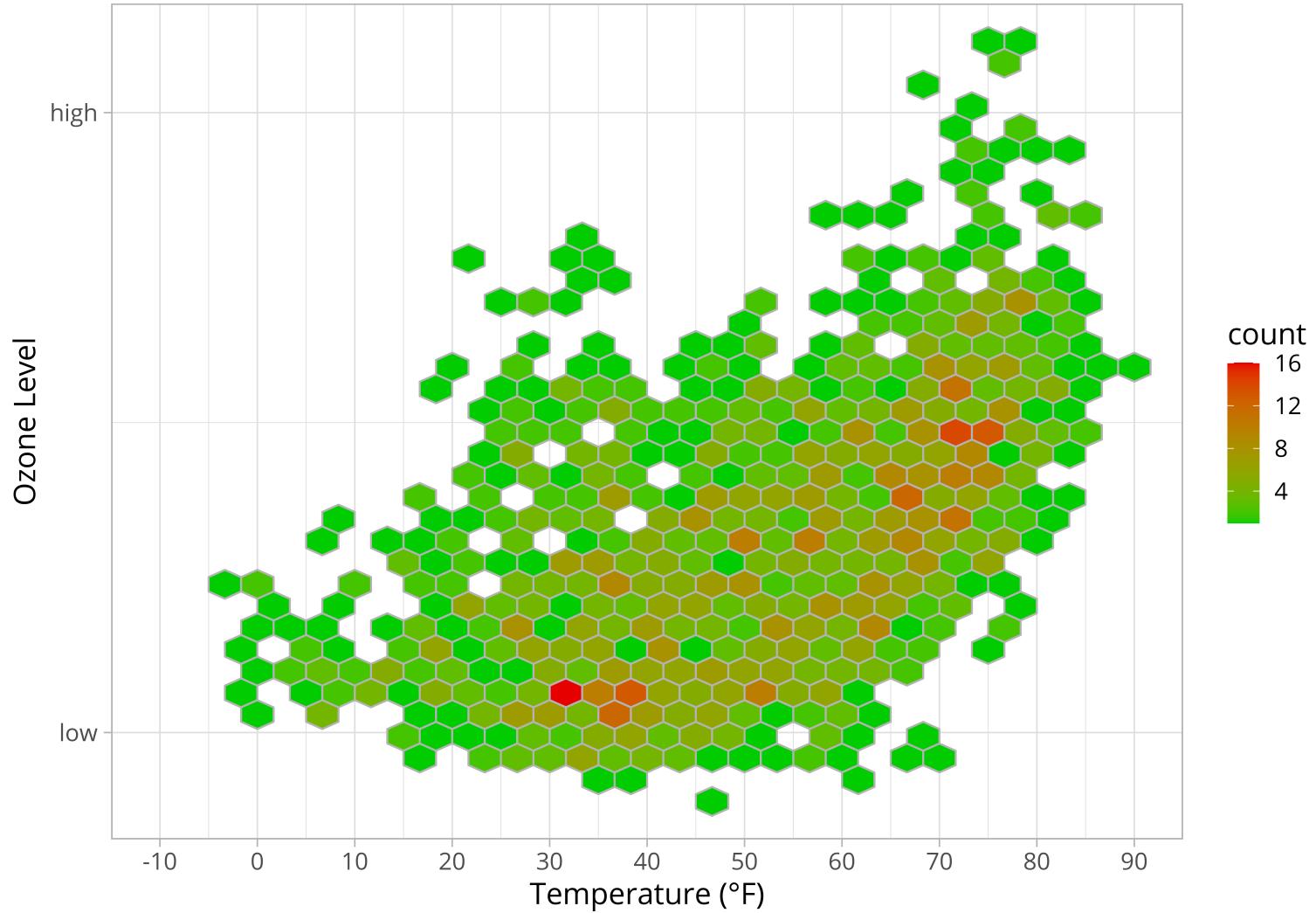
# Color Choice & Accessibility

Choose color-blind friendly palettes:  
[projects.susielu.com/viz-palette](http://projects.susielu.com/viz-palette)

Test your final visualization:  
[color-blindness.com/coblis-color-blindness-simulator](http://color-blindness.com/coblis-color-blindness-simulator)

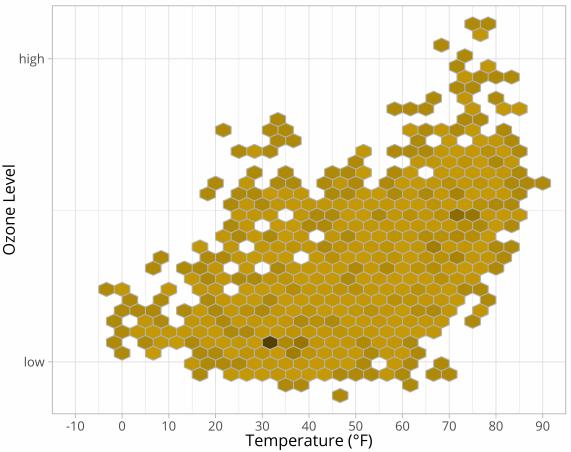
Create a CVD-version of your ggplot in R:  
[github.com/clauswilke/colorblindr](https://github.com/clauswilke/colorblindr)

# Color Choice & Accessibility

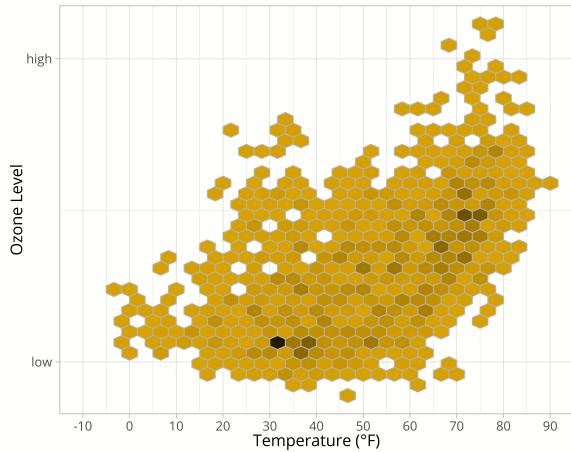


# Color Choice & Accessibility

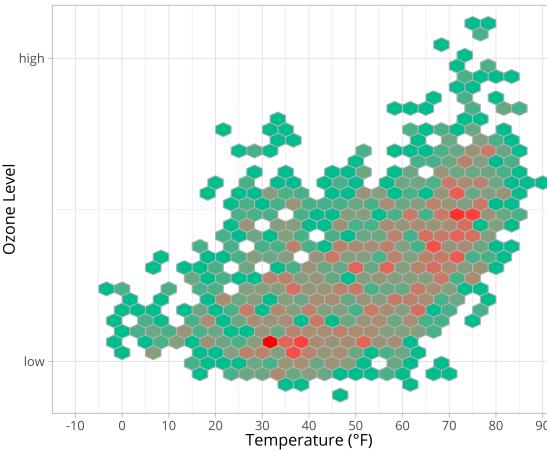
Deutanomaly



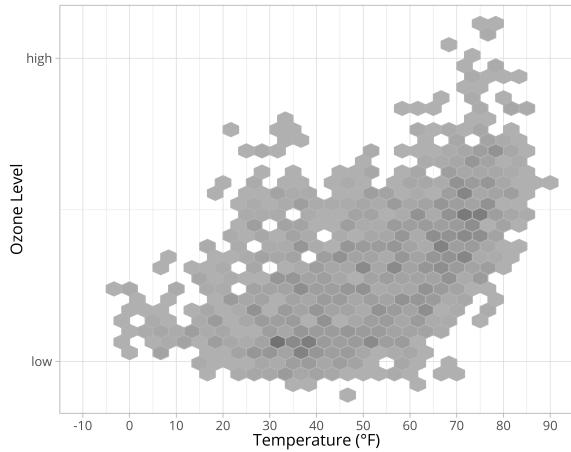
Protanomaly



Tritanomaly



Desaturated



Created with [colorblindr](#)

# Color Choice & Accessibility

## VIZ PALETTE

By: Elijah Meeks & Susie Lu

### PICK

Use Chroma.js

Use Colorgorical

Use ColorBrewer

### EDIT

7 Colors

- 1 ● #ffd700
- 2 ● #ffb14e
- 3 ● #fa8775
- 4 ● #ea5f94
- 5 ● #cd34b5
- 6 ● #9d02d7
- 7 ● #0000ff

hex  rgb

hsl

### GET

String quotes  
 Object with metadata

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

Background color: #ffffcc

Font color: ● #000000

Charts made with [Semiotic](#)

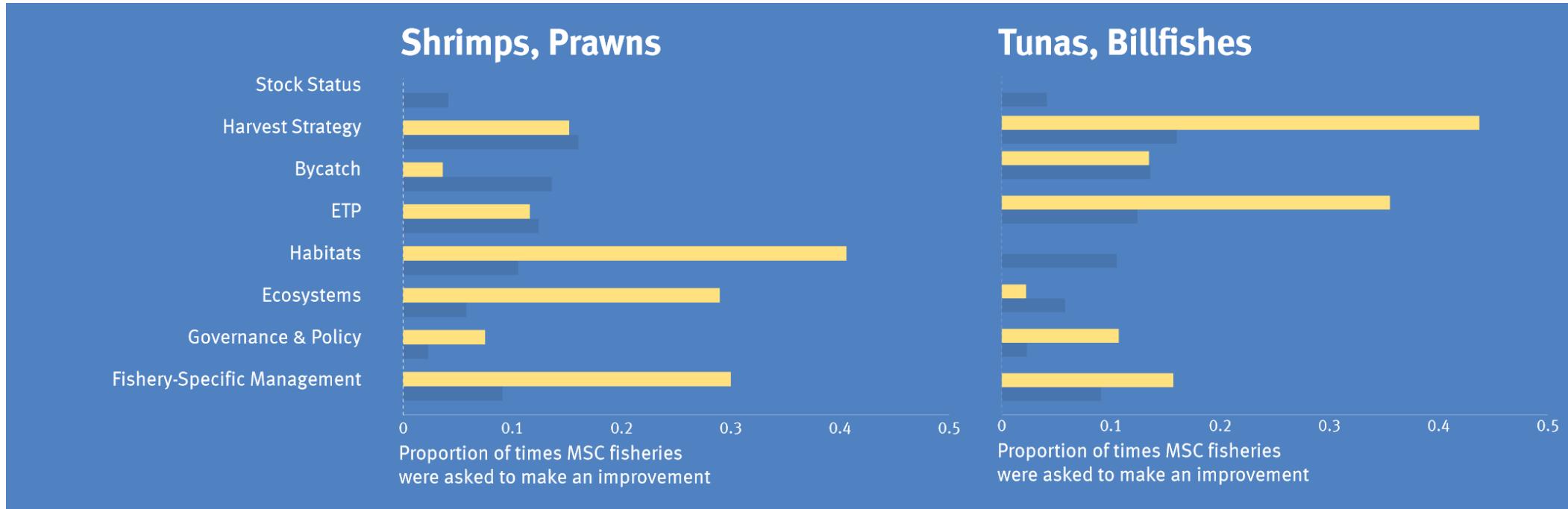
## COLORS IN ACTION

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

Sample font | Randomize Data | Stroke: Dark | None



# Color Choice & Accessibility





# 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



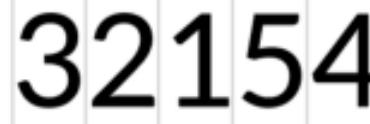
32154

Montserrat  
proportional numbers



32154

Open Sans  
tabular numbers



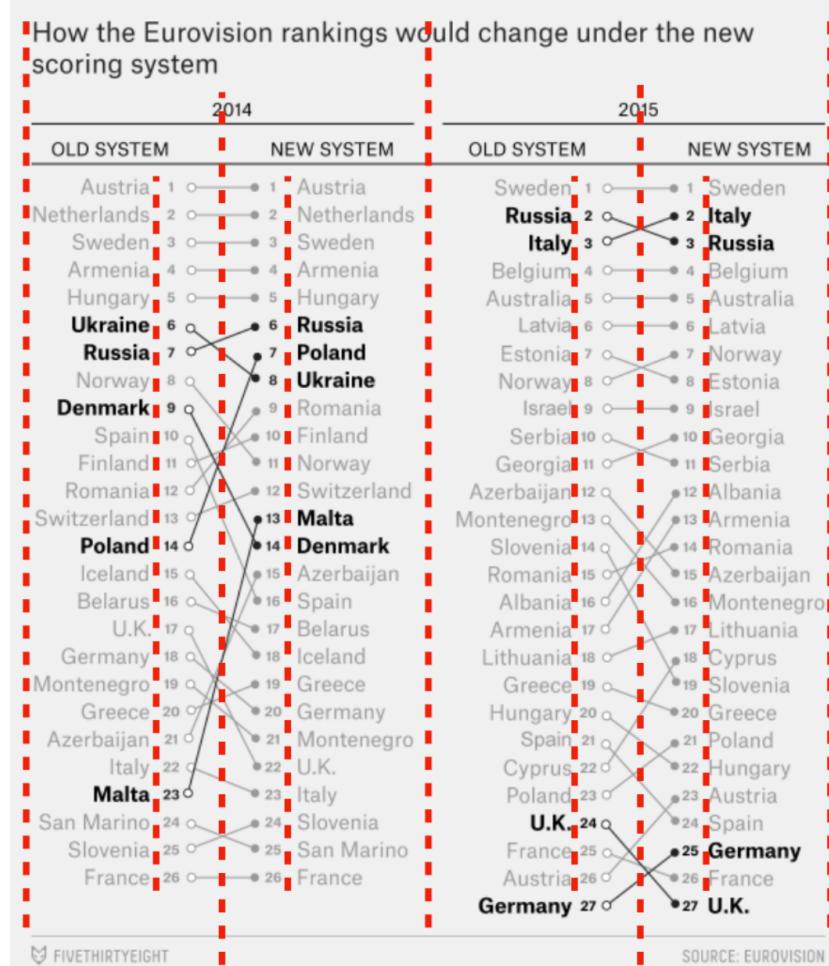
32154

Lato  
tabular numbers

Source: *Choosing Fonts for Your Data Visualization* by Tiffany France

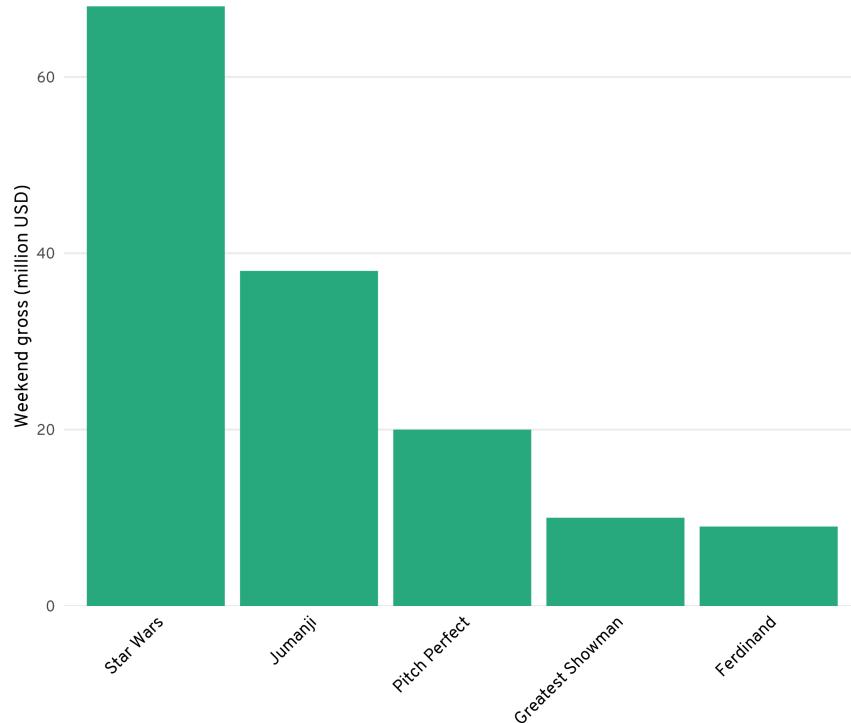
# Allign Your Text

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

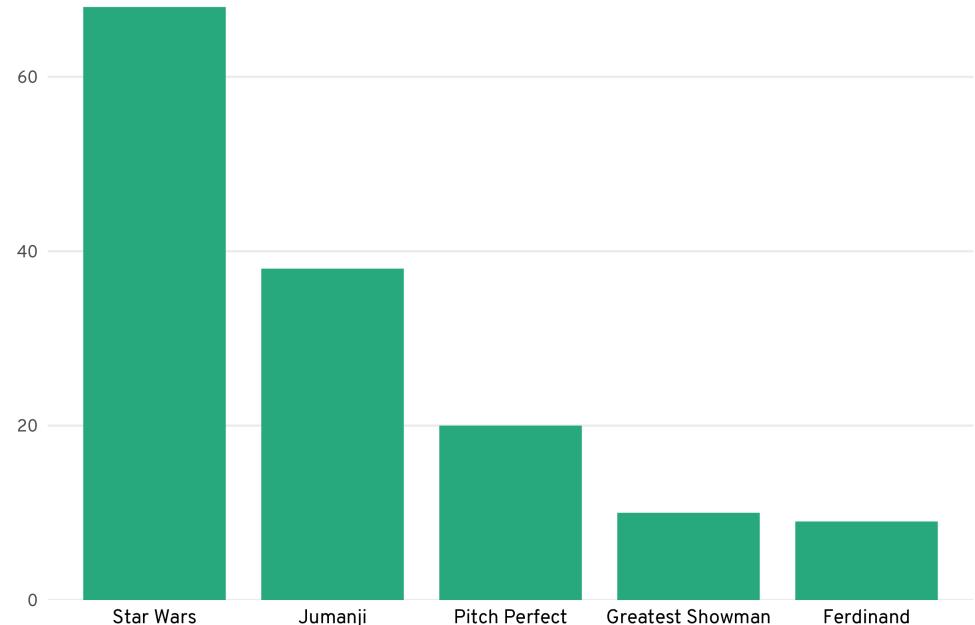


Source: Will R. Chase

# (Don't) Rotate Your Text

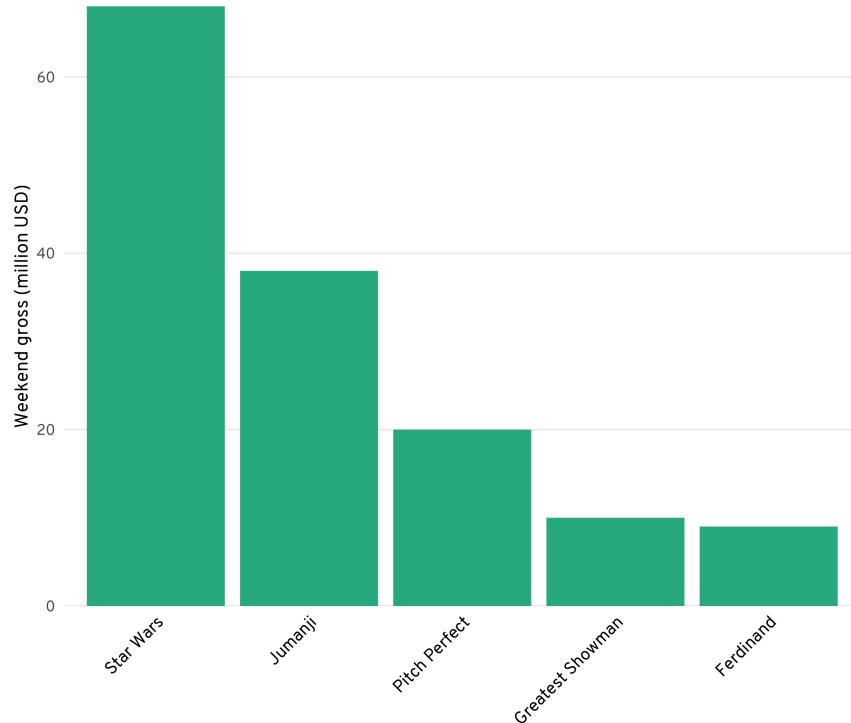


Weekend gross in million USD of popular blockbusters

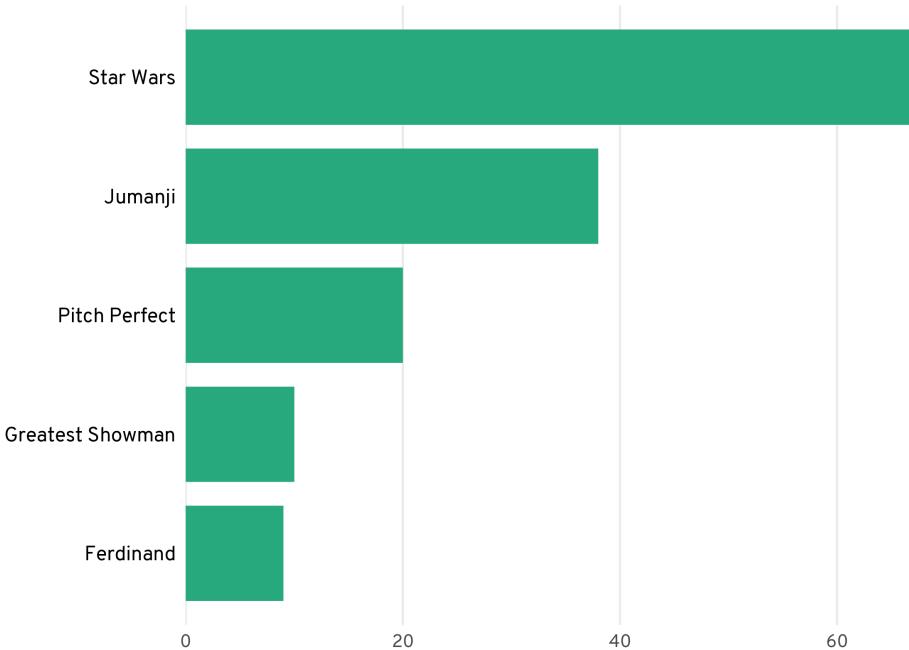


Modified from [Fundamentals of Data Visualization](#) by Claus Wilke

# (Don't) Rotate Your Text

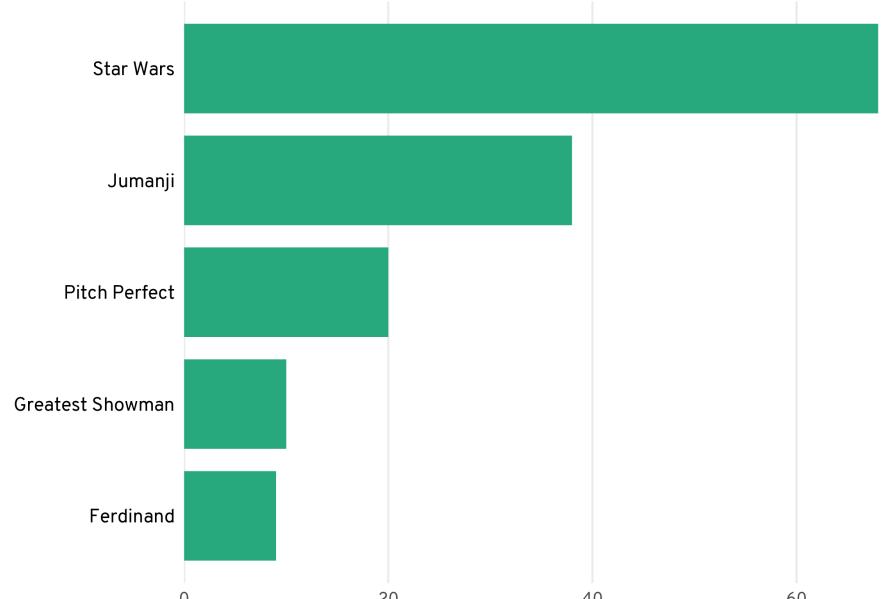


Weekend gross in million USD of popular blockbusters

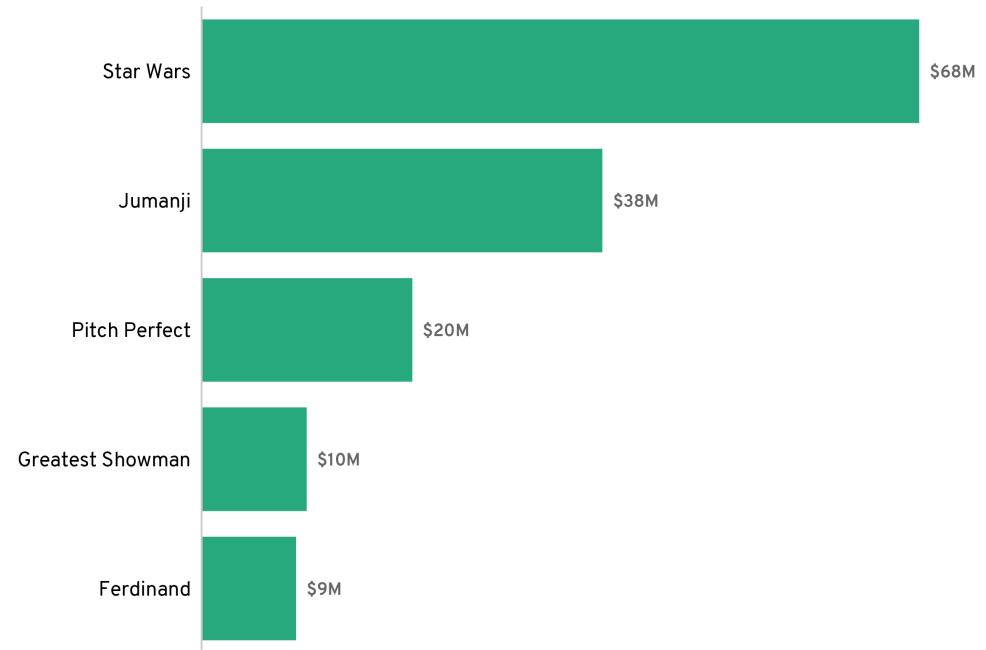


# Use Annotations

Weekend gross in million USD of popular blockbusters

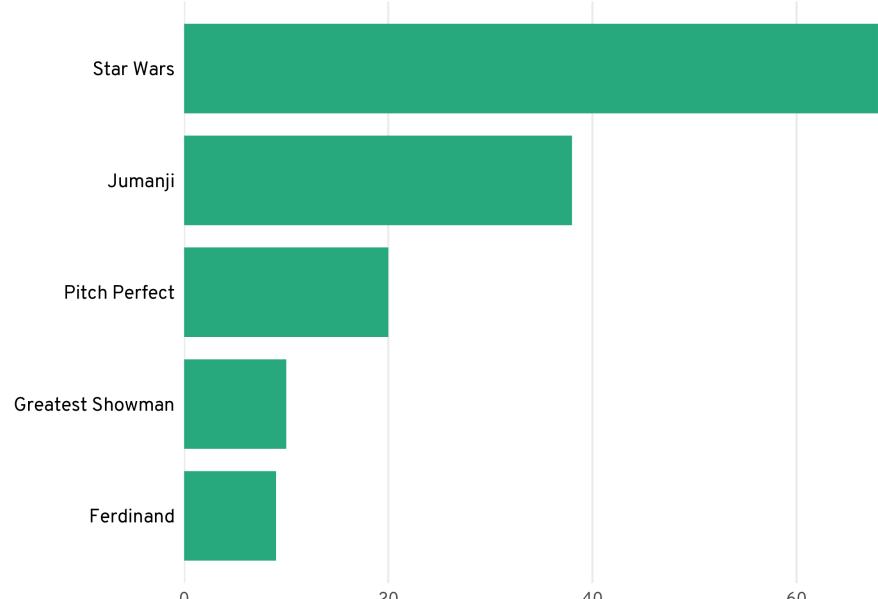


Weekend gross in million USD of popular blockbusters

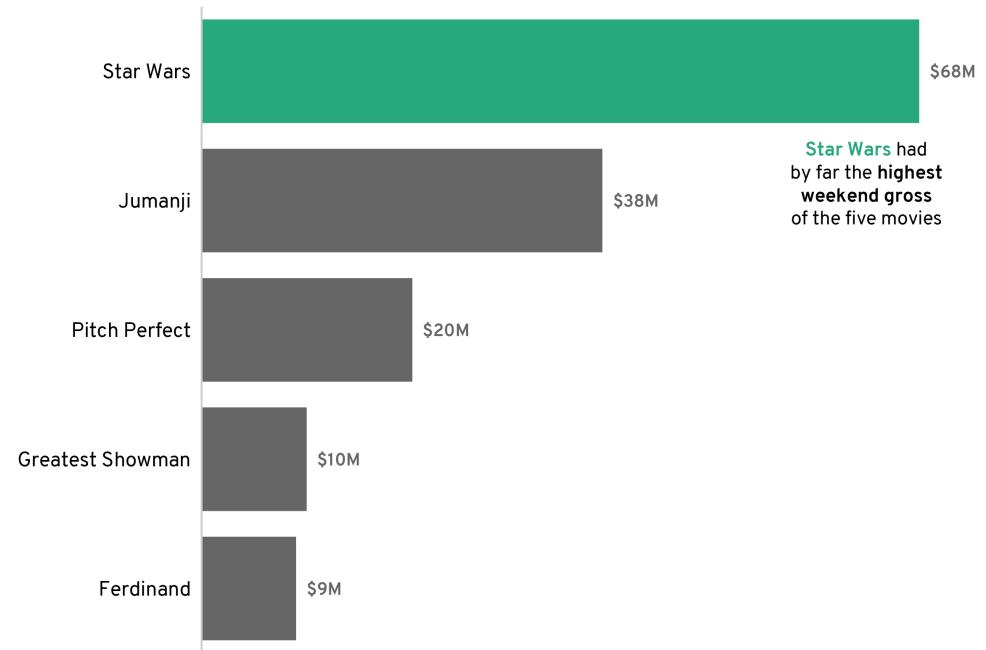


# Use Annotations

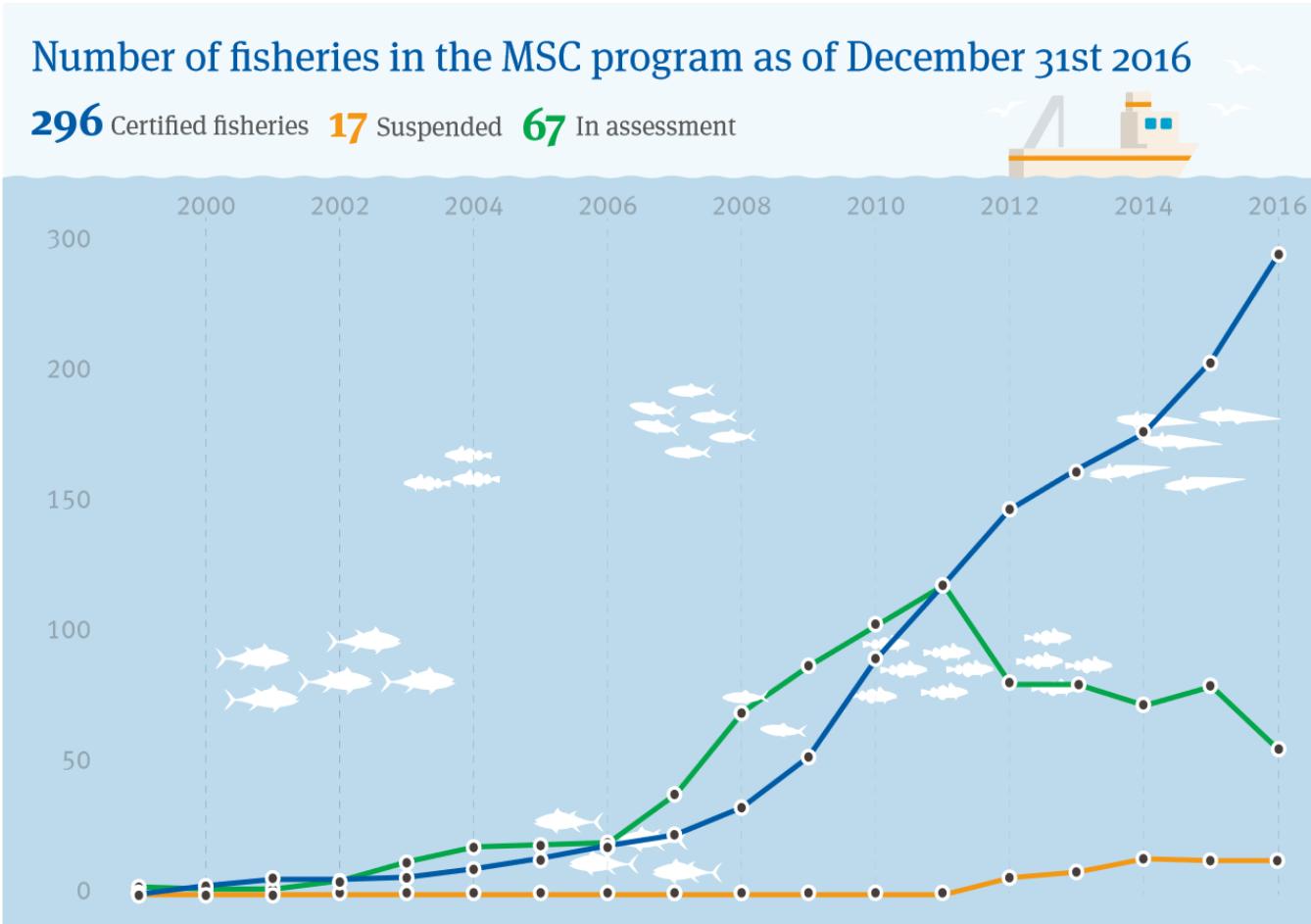
Weekend gross in million USD of popular blockbusters



Weekend gross in million USD of popular blockbusters



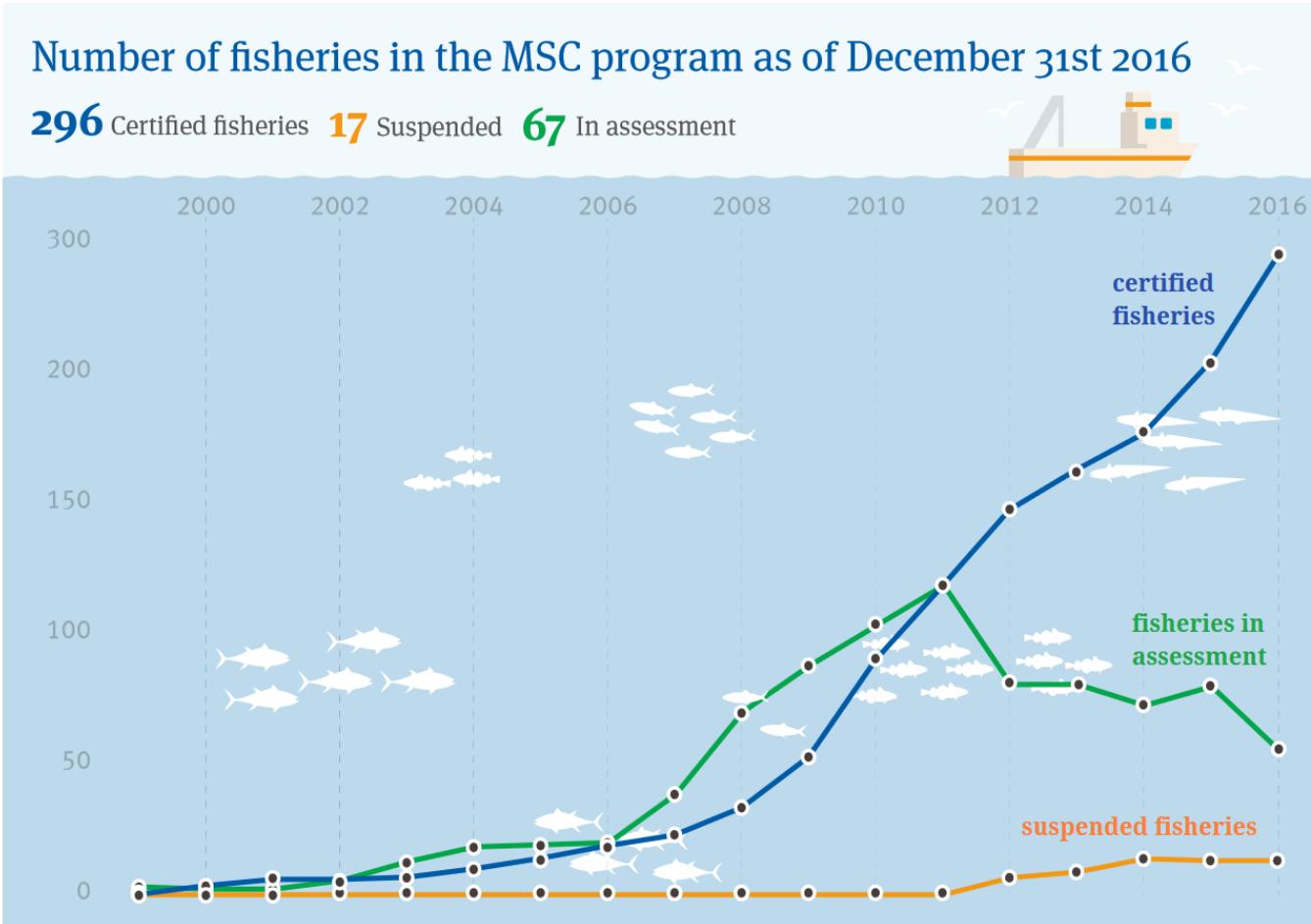
# Use Annotations



**100s**

of fisheries are  
not yet ready for  
assessment and  
are engaged in  
pre-assessment  
activities and FIPs

# Use Annotations

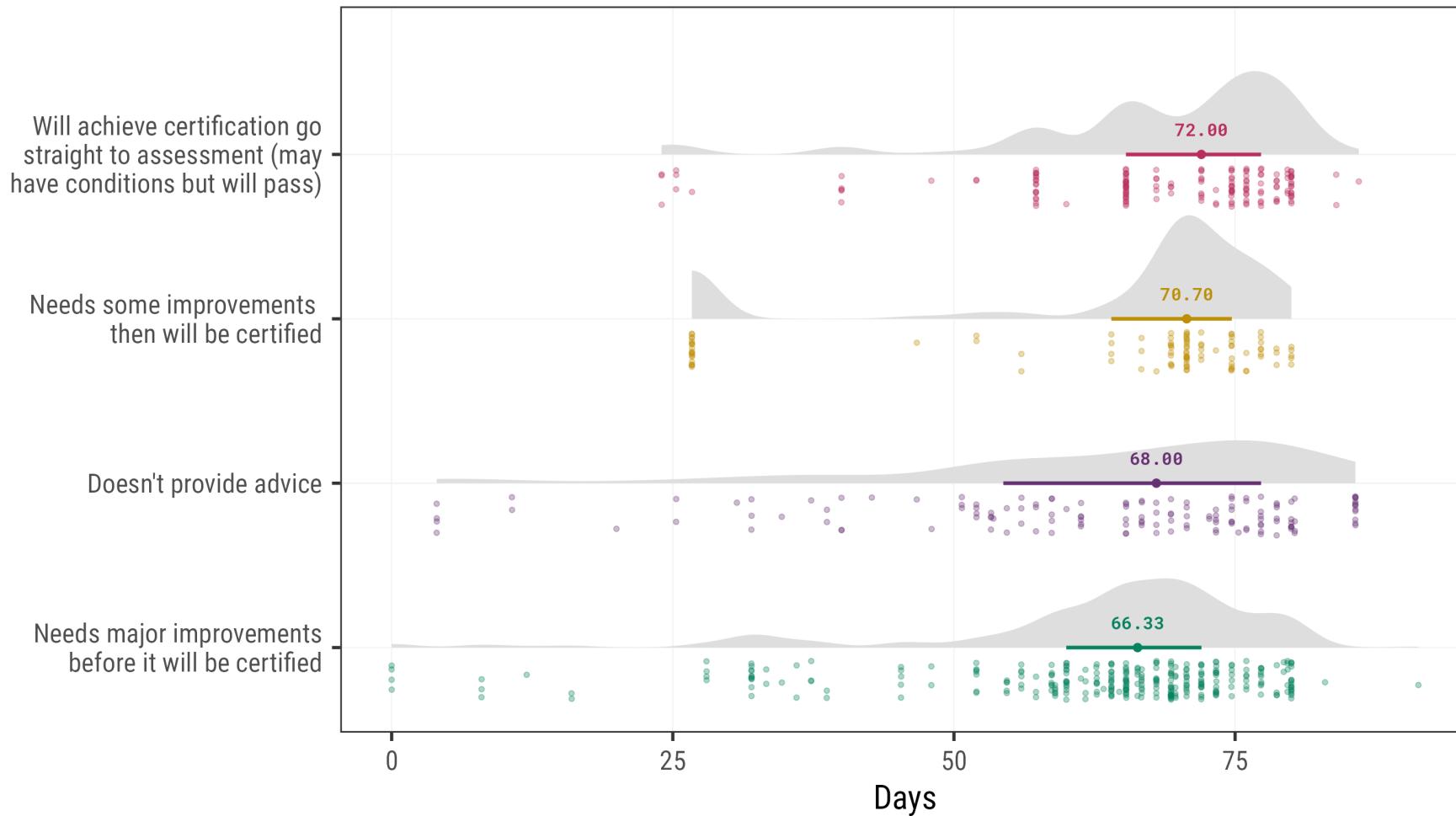


**100s**

of fisheries are  
not yet ready for  
assessment and  
are engaged in  
pre-assessment  
activities and FIPs

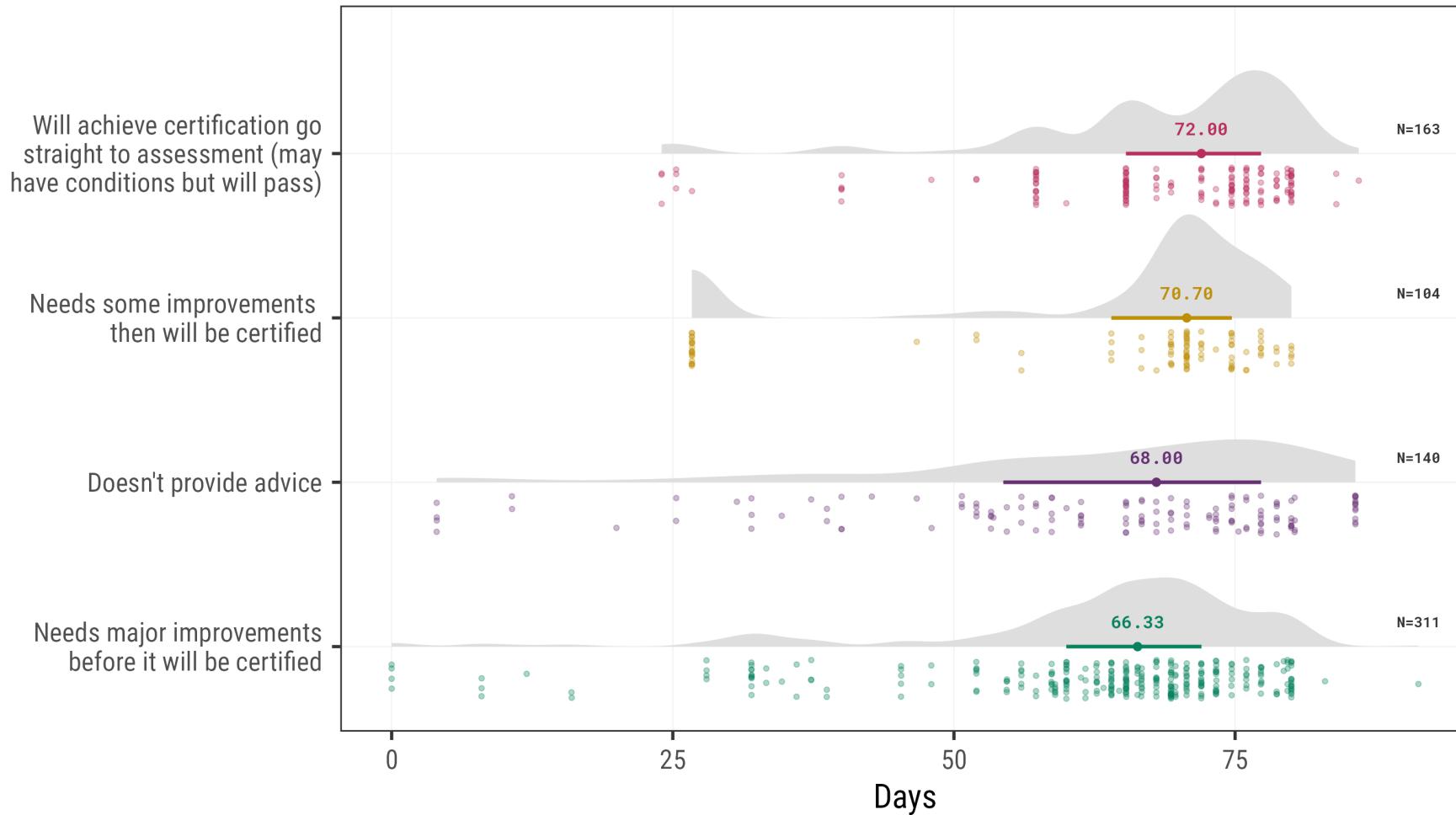
# Use Annotations

P2 Score (combined)



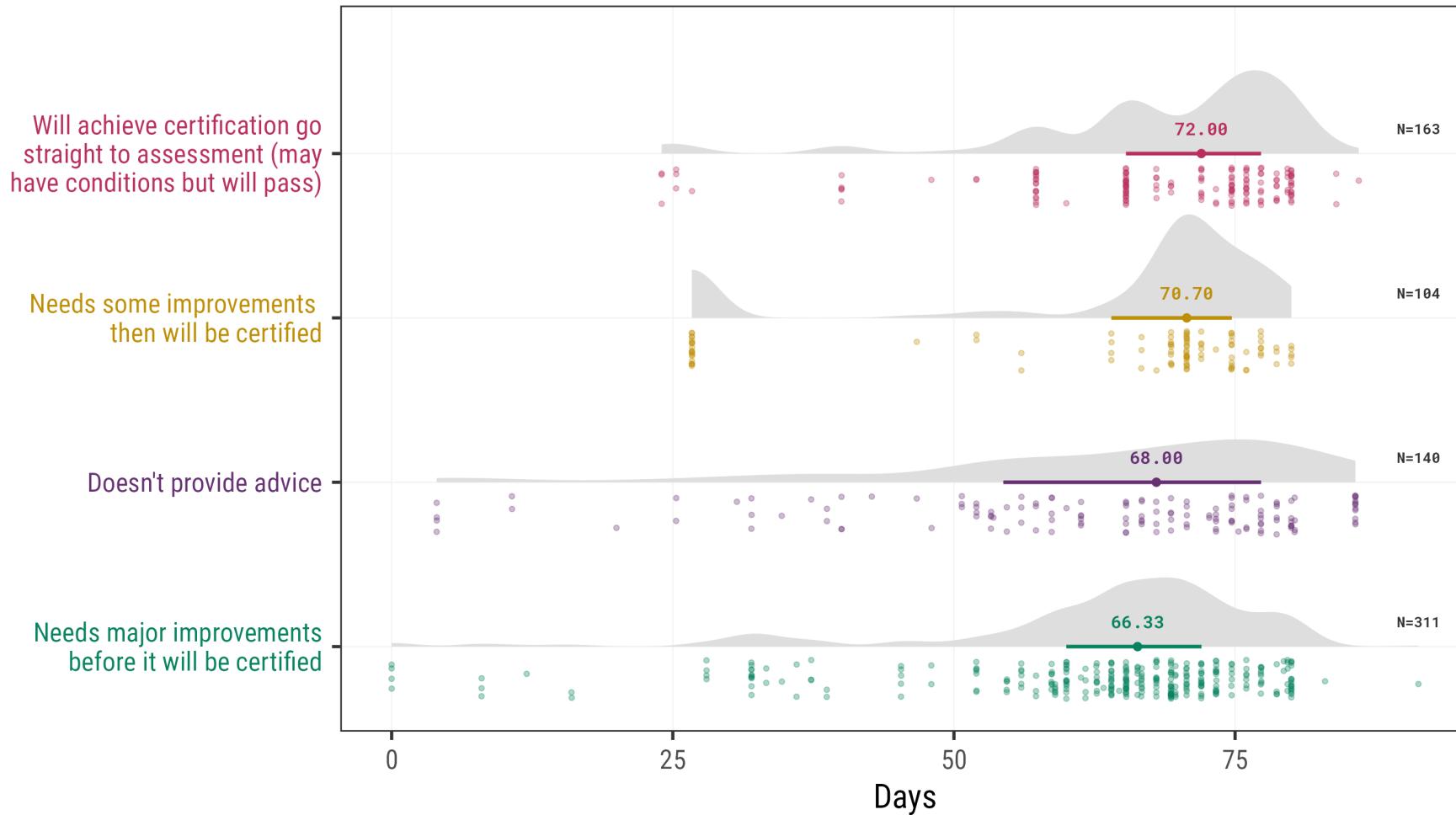
# Use Annotations

P2 Score (combined)



# Use Color Consistency in Labels

P2 Score (combined)

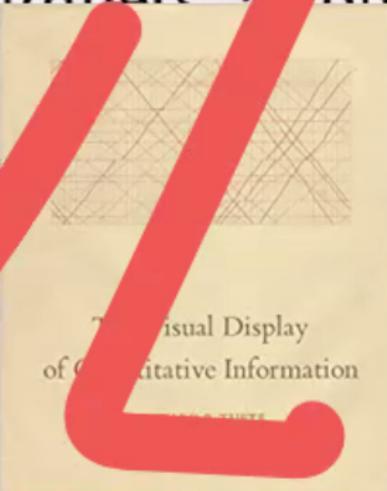


# Chart Design Principles

Graphical integrity is more likely to result if these six principles are followed:

The presentation of numbers or physically measured on the surface of the graphic itself should be directly proportional to the numerical quantities.

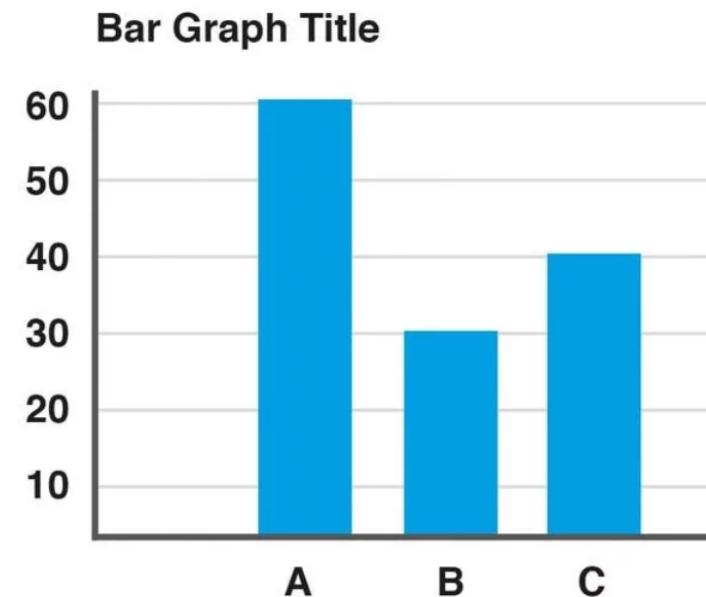
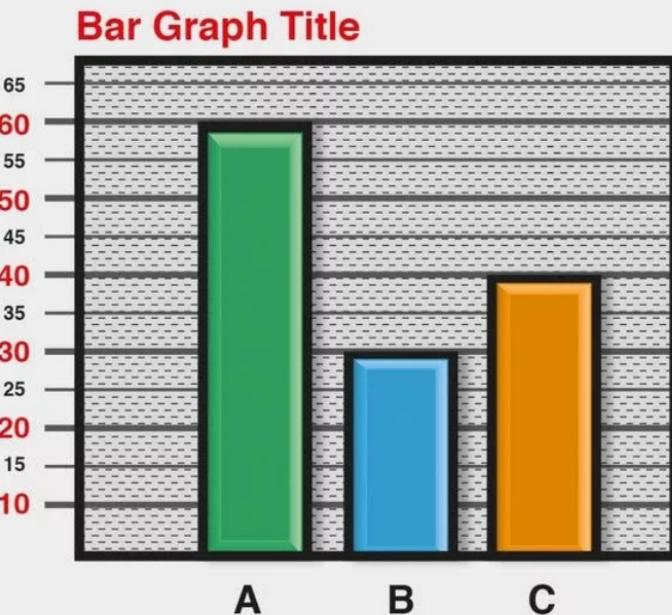
Clear, detailed, and color-coded legends should be used to defeat graphical distortion and the data on the graphic itself. Label important events in the data.

The image shows the front cover of a book titled "The Visual Display of Quantitative Information" by Edward Tufte. The cover is light yellow with a grid pattern. The title and author's name are printed in a serif font at the top. Below the title, there is smaller text and a small graphic of a bar chart.

Show data variation, not design variation.

In time-series displays of money, deflated and standardized units

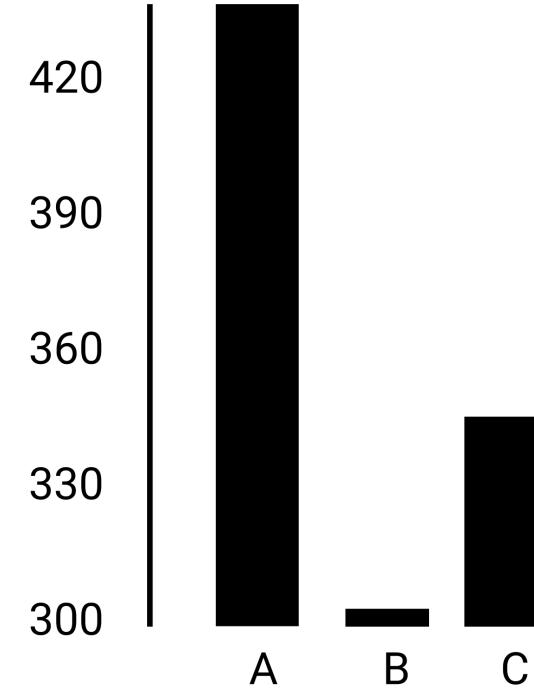
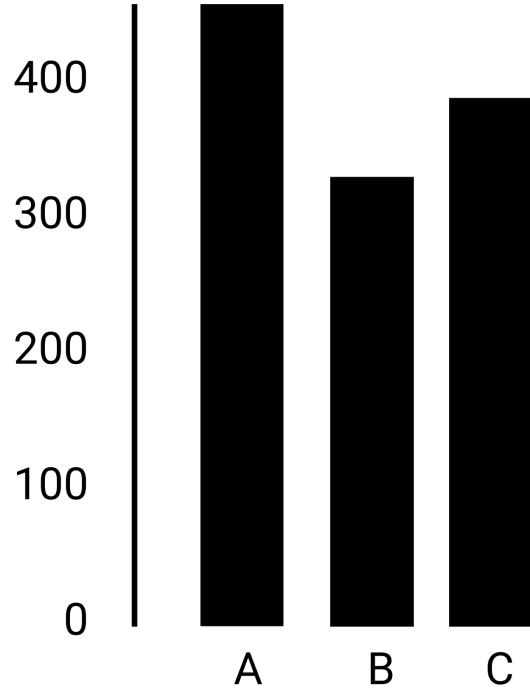
# Don't Clutter Your Plot



*Source: canva.com*

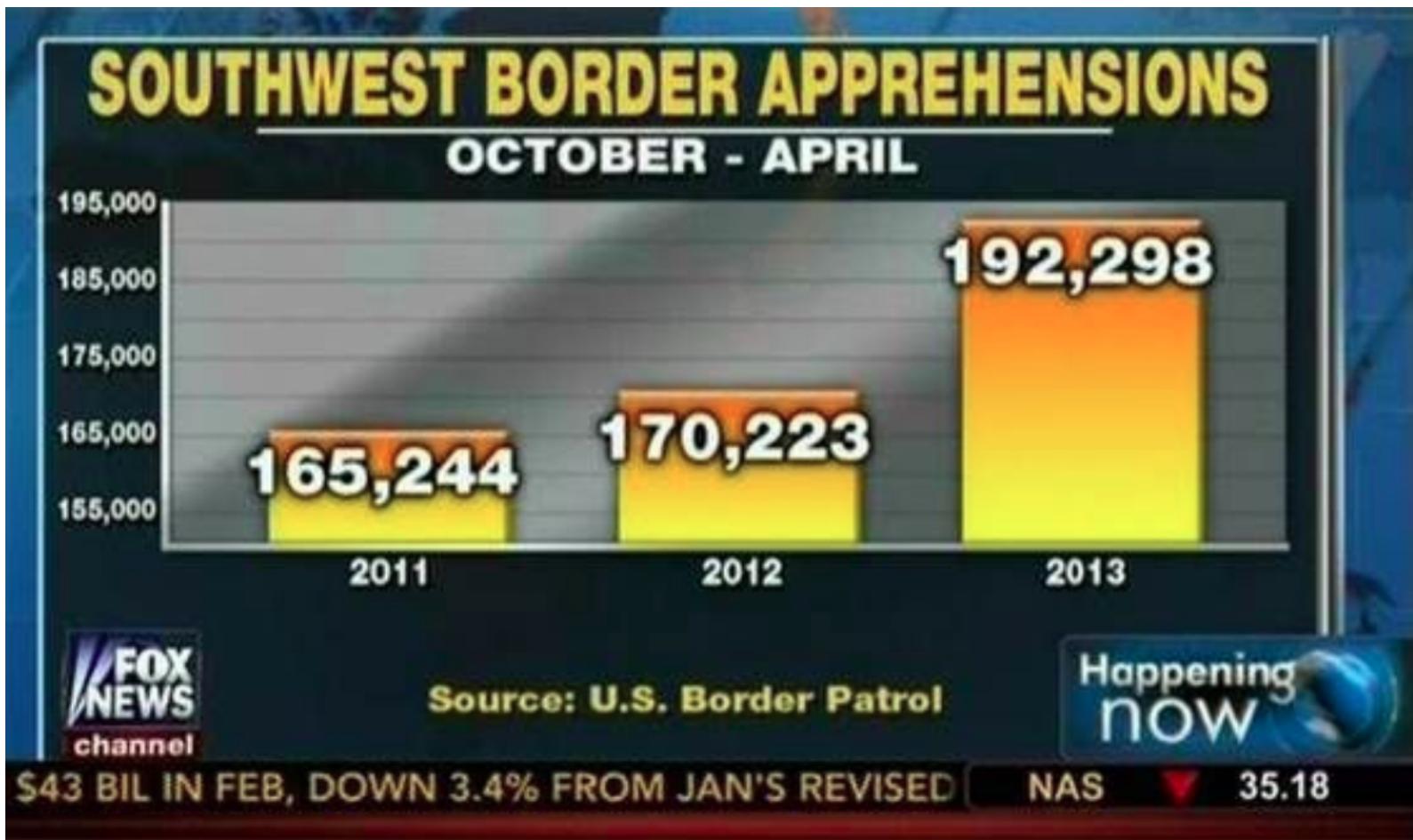
**Remove**  
to improve  
(the **data-ink** ratio)

# Start at Zero



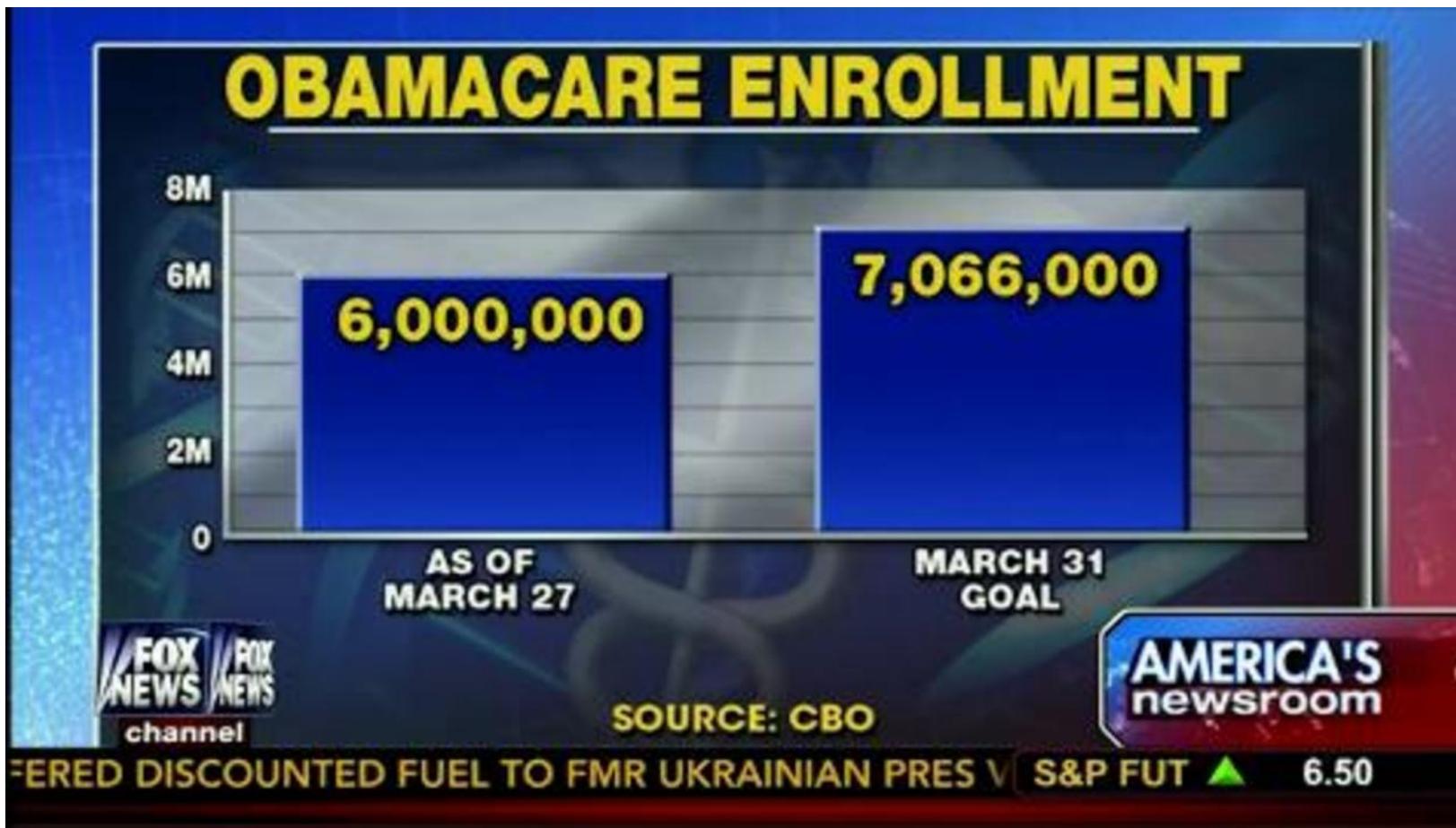
*“Hands-On Data Visualization” by Jack Dougherty & Ilya Illyankou*

# Always Start at Zero



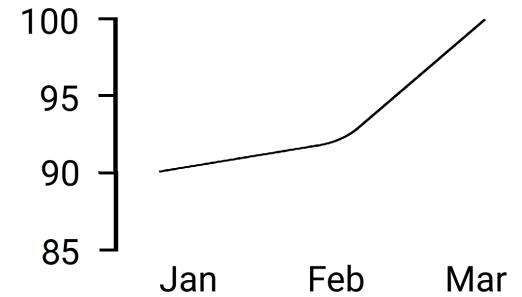
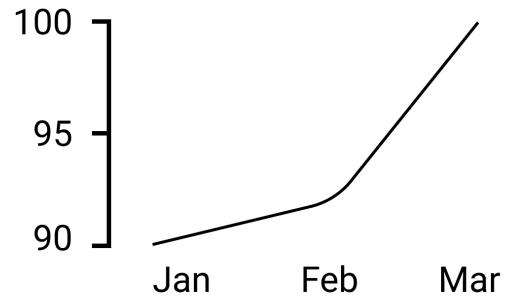
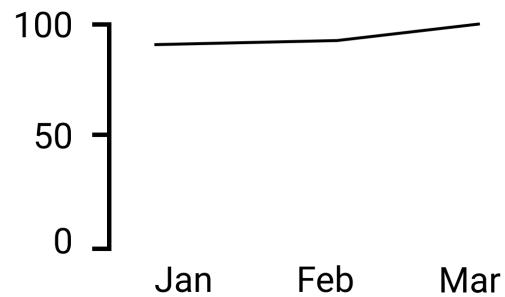
Fox News

# Always Start at **Zero**?



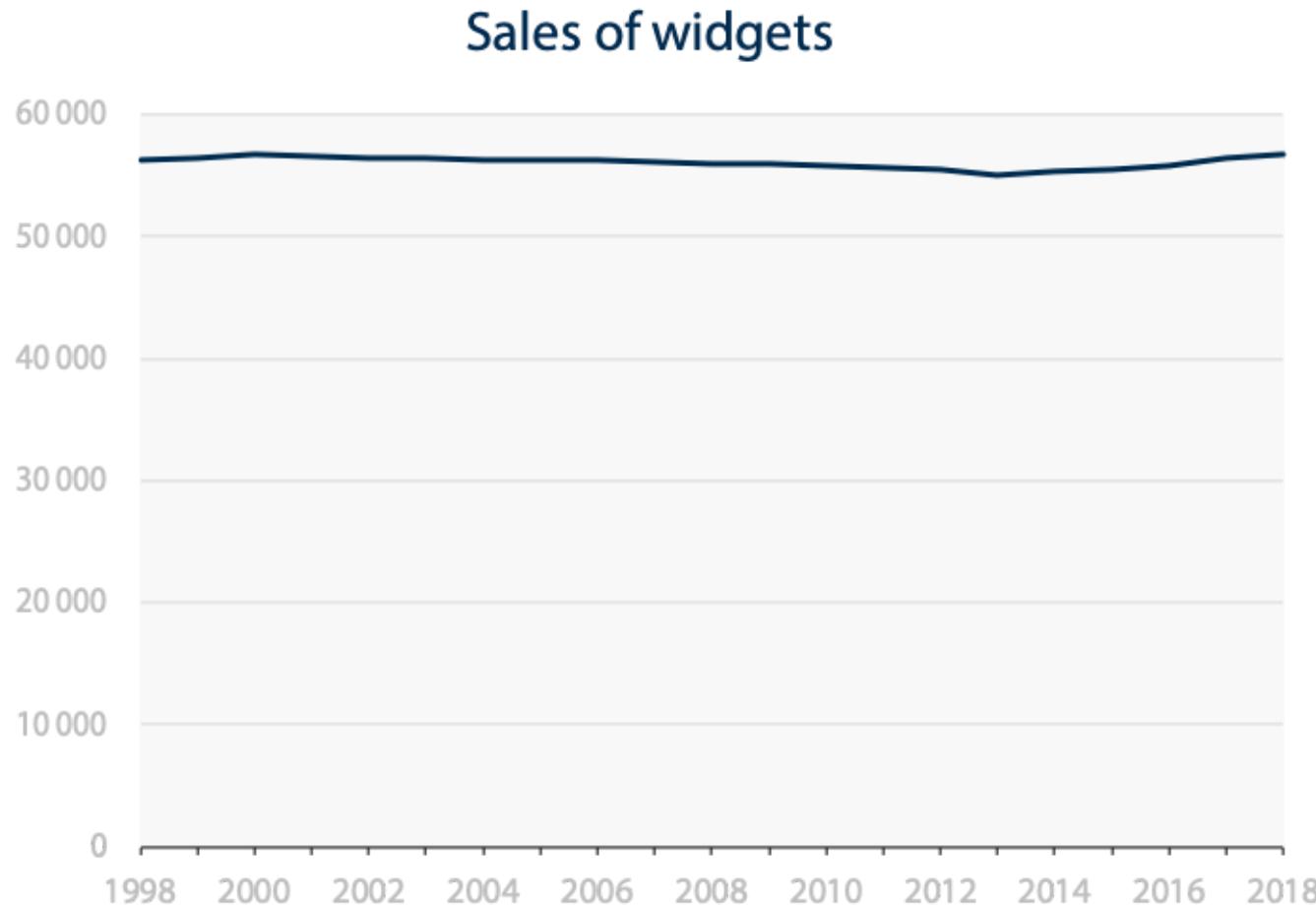
Fox News

# (Always) Start at Zero?



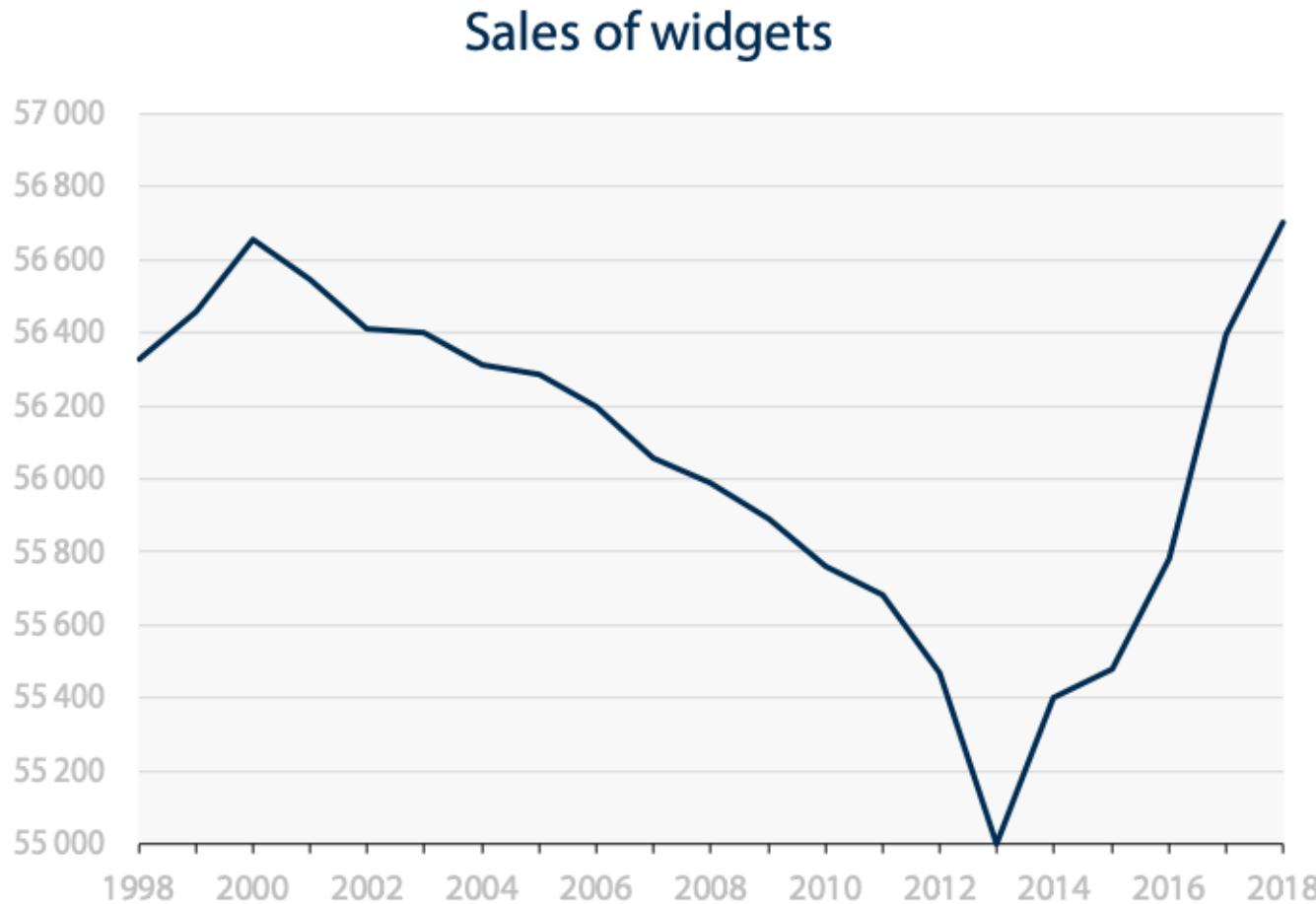
*“Hands-On Data Visualization” by Jack Dougherty & Ilya Illyankou*

# (Always) Start at Zero?



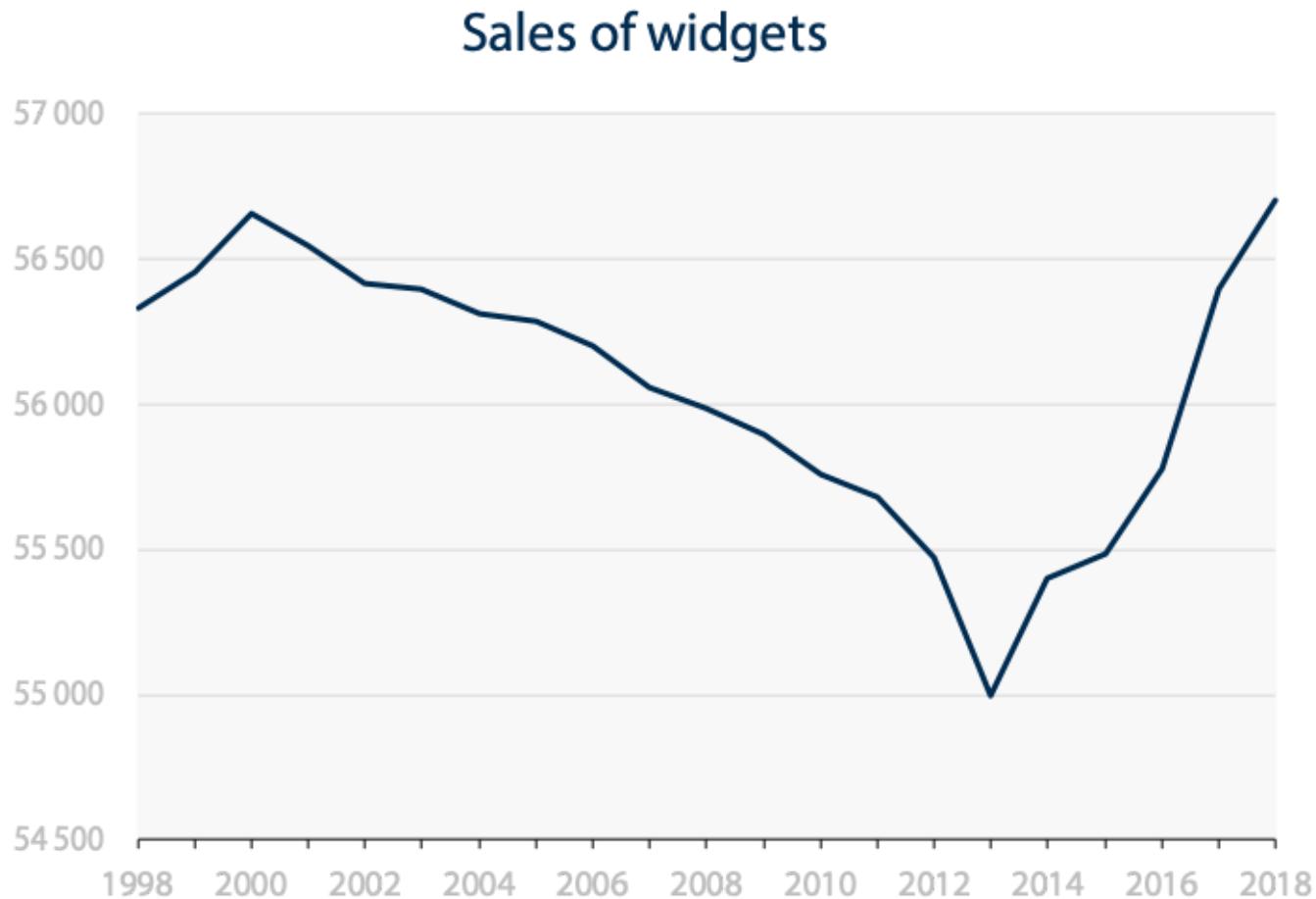
*Francis Gagnon, Voilà*

# (Always) Start at Zero?



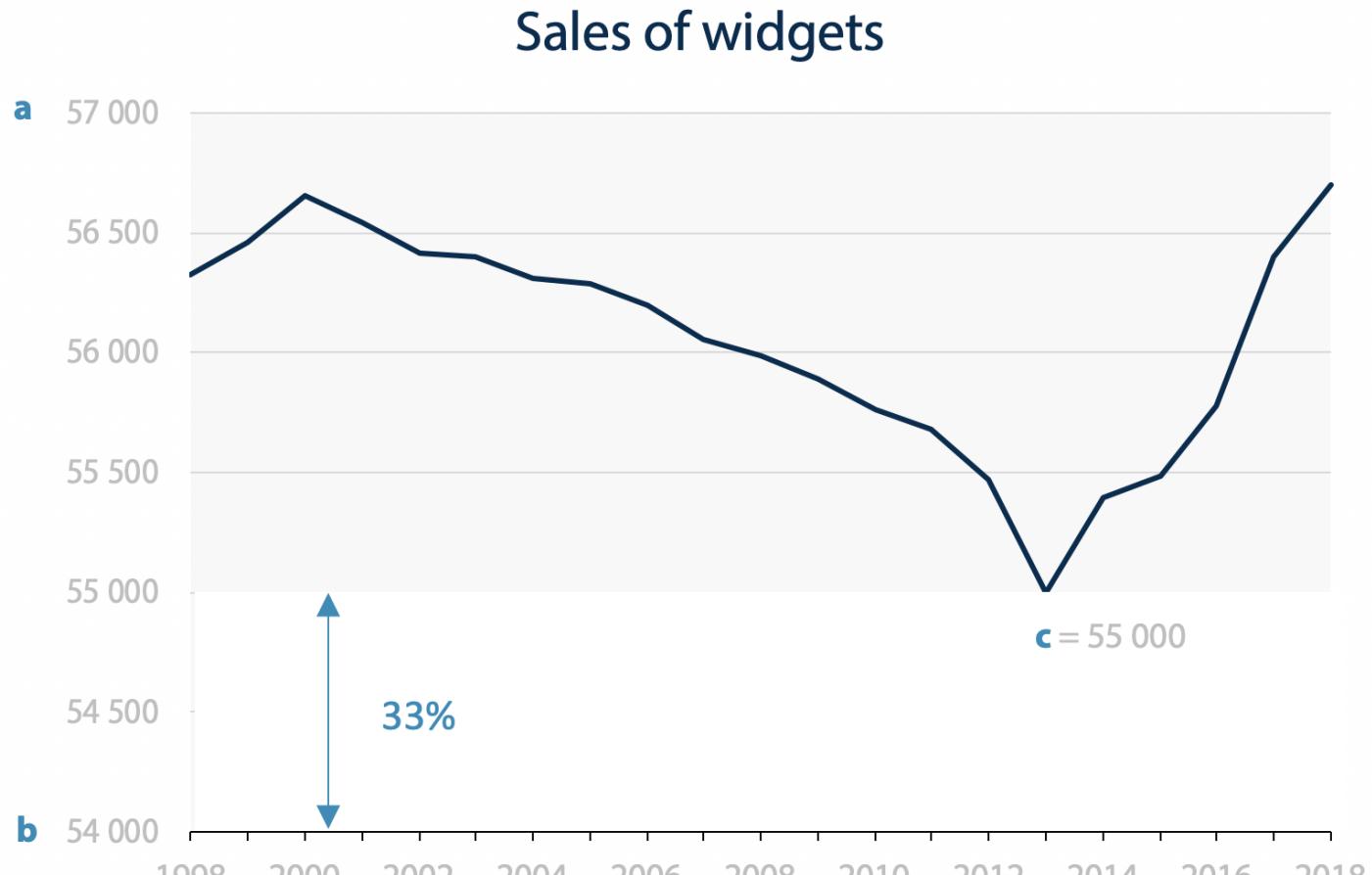
Francis Gagnon, Voilà

# The Golden Ratio



*Francis Gagnon, Voilà*

# The Golden Ratio



*Francis Gagnon, Voilà*

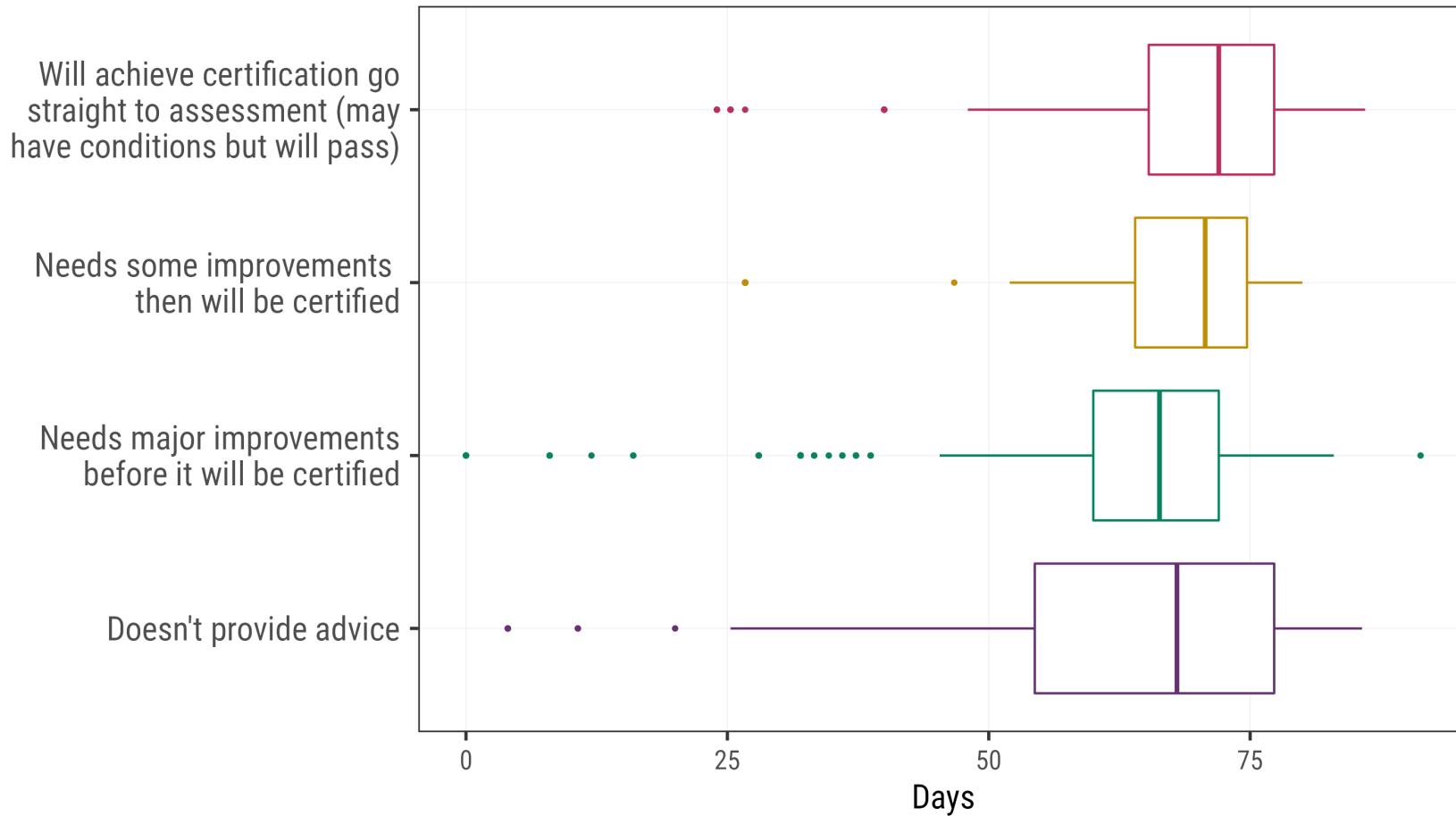
# Order Your Data



*“Hands-On Data Visualization” by Jack Dougherty & Ilya Illyankou*

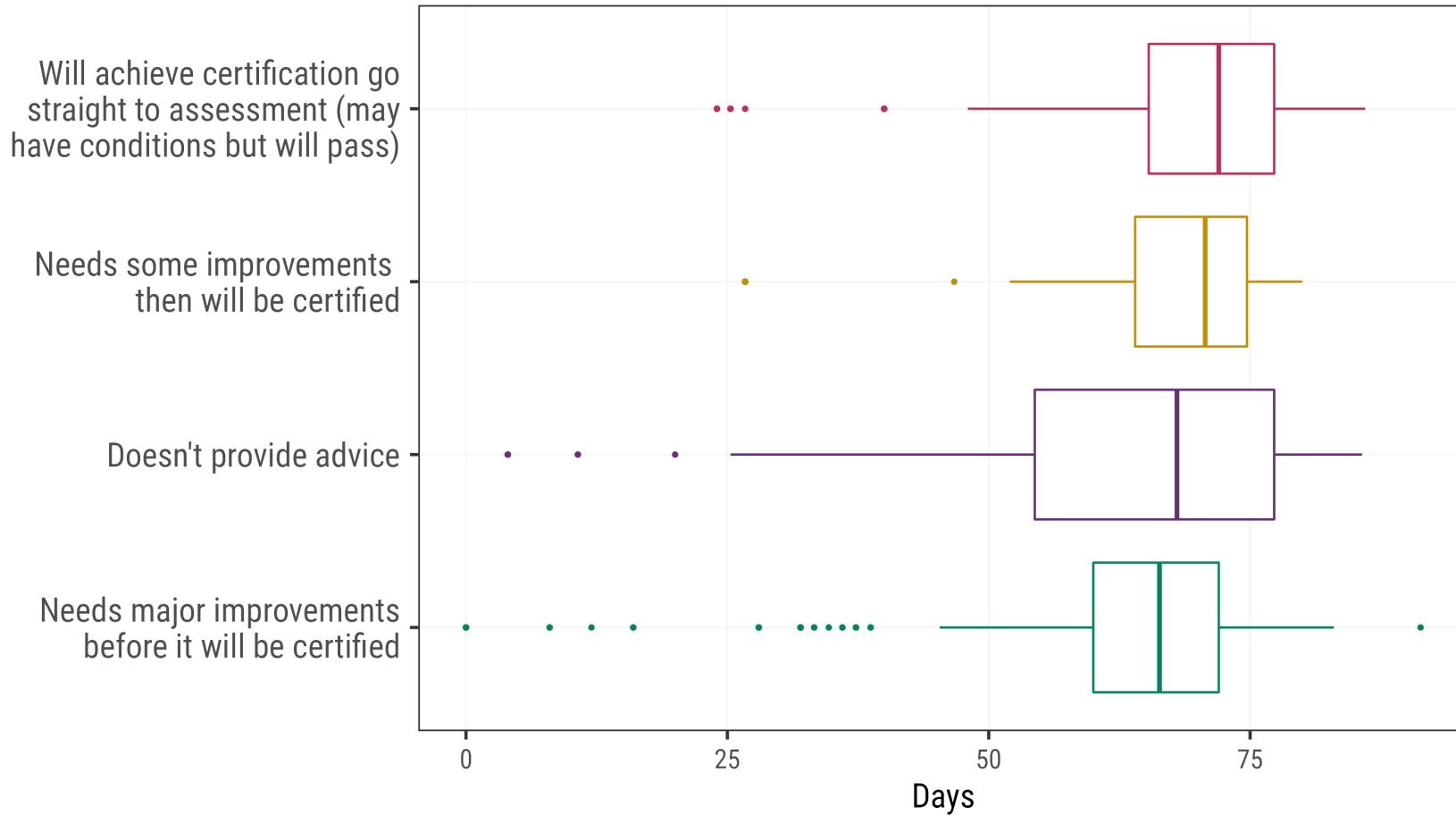
# Order Your Data

P2 Score (combined)



# Order Your Data

P2 Score (combined)



## Design for Your Audience

- Choose charts based on your goal not tradition or novelty (only).
- Make sure your visualization is accessible for everyone (colors, readability).
- Provide meaningful labels and titles.

## Be Honest

- Show raw data if possible.
- Don't truncate bar charts. Add spacing to truncated axes.
- Be consistent with axis scaling.

## Lend A Helping Hand

- Use annotations and direct labels instead of/in addition to text and legends.
- Order your data, either by value or intrinsic ranking.
- Reveal information step by step (if applicable).

## Books

- “[Fundamentals of Data Visualization](#)”, an open-access book by Claus Wilke
- “[Data Visualization](#)”, an open-access book by Kieran Healy
- “[Avoiding Data Pitfalls](#)” by Ben Jones
- “[The Functional Art](#)”, “[The Truthful Art](#)”, and “[How Charts Lie](#)” by Alberto Cairo
- “[Storytelling with Data](#)” by Cole Nussbaumer Knaflic
- “[Data Visualization Handbook](#)” by Juuso Koponen & Jonatan Hildén
- “[Info We Trust: How to Inspire the World with Data](#)” by RJ Andrews

## Blogs

- “[Nightingale](#)”, the journal of the Data Visualization Society
- “[Chartable](#)”, the blog by DataWrapper

## Chart Choice Helpers

- [From Data to Viz](#)
- [DataViz Project](#)
- [Visualizaiton Universe](#)
- [Material Design](#)

## Color Choice Helpers

- [Viz Palette](#) by Elijah Meeks & Susie Lu
- “[How to pick more beautiful colors for your data visualizations](#)” by Lisa Charlotte Rost