

# Data Visualization in R with **ggplot2**

Principles of Data Visualization

**Cédric Scherer**

Physalia Courses | March 2-6 2020

Photo by Richard Strozyński

**Data Visualization**  
is any graphical representation  
of information and data

# Data Visualization

is any graphical representation  
of information and data



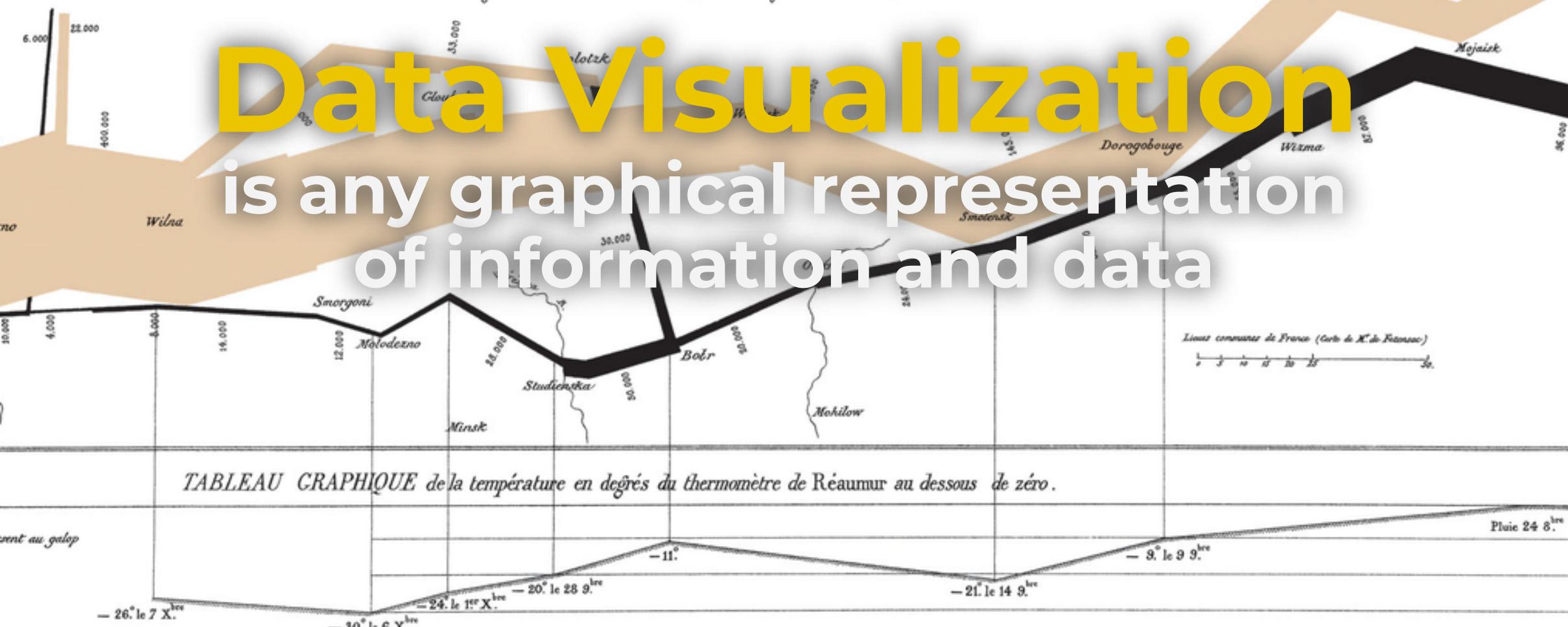
# Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

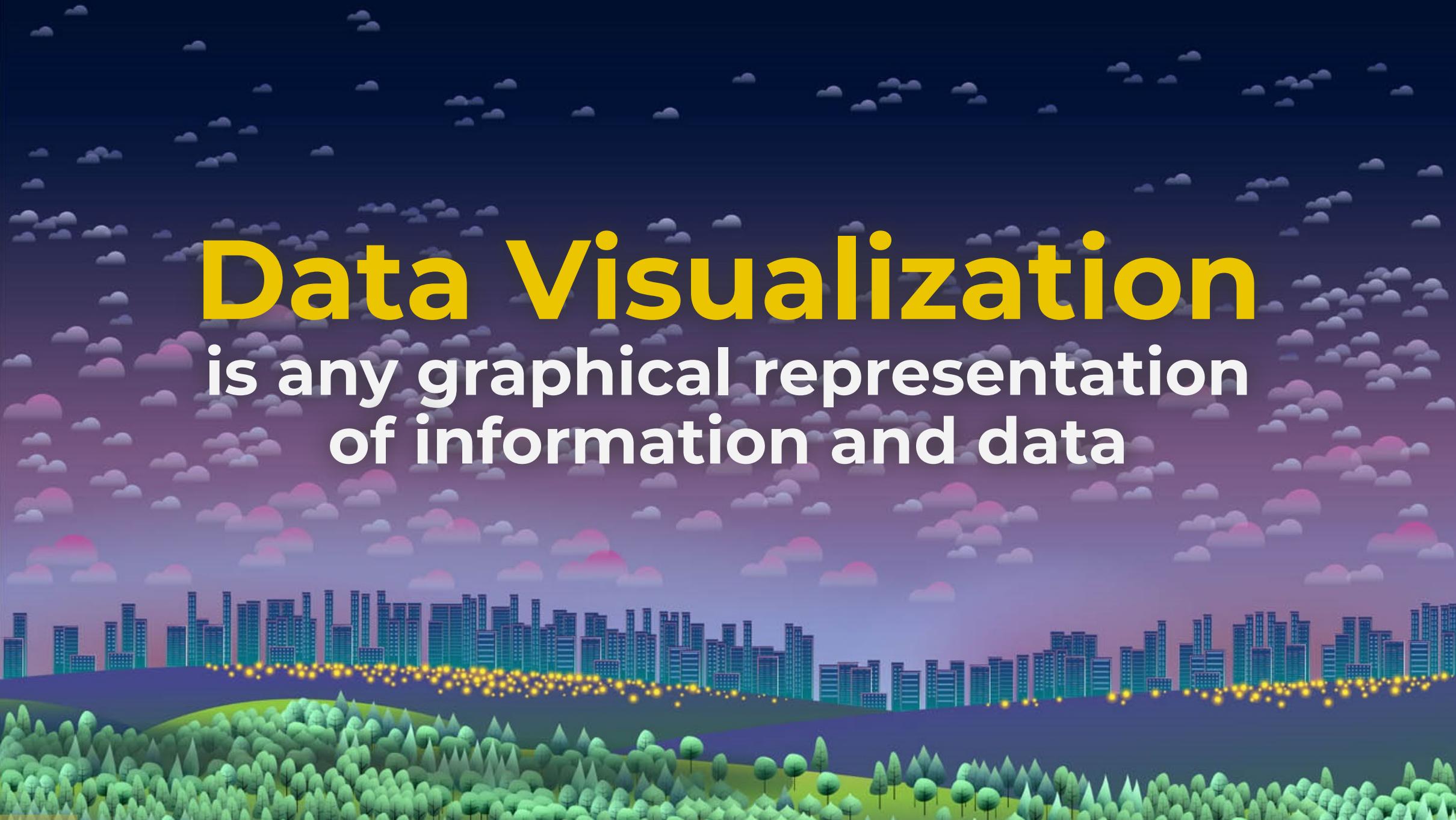
Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

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 ont été 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. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

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





**Data Visualization**  
is any graphical representation  
of information and data



*"A View on Despair"* by Sonja Kuijpers/STUDIO TERP

You might be wondering what you are viewing here.



"A View on Despair" by Sonja Kuijpers/STUDIO TERP

You might be wondering what you are viewing here.

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



"A View on Despair" by Sonja Kuijpers/STUDIO TERP

Each category/method of suicide is represented by a certain element:



hanging (strangulation)



taking drugs/alcohol/medicines



in front of train or metro



drowning



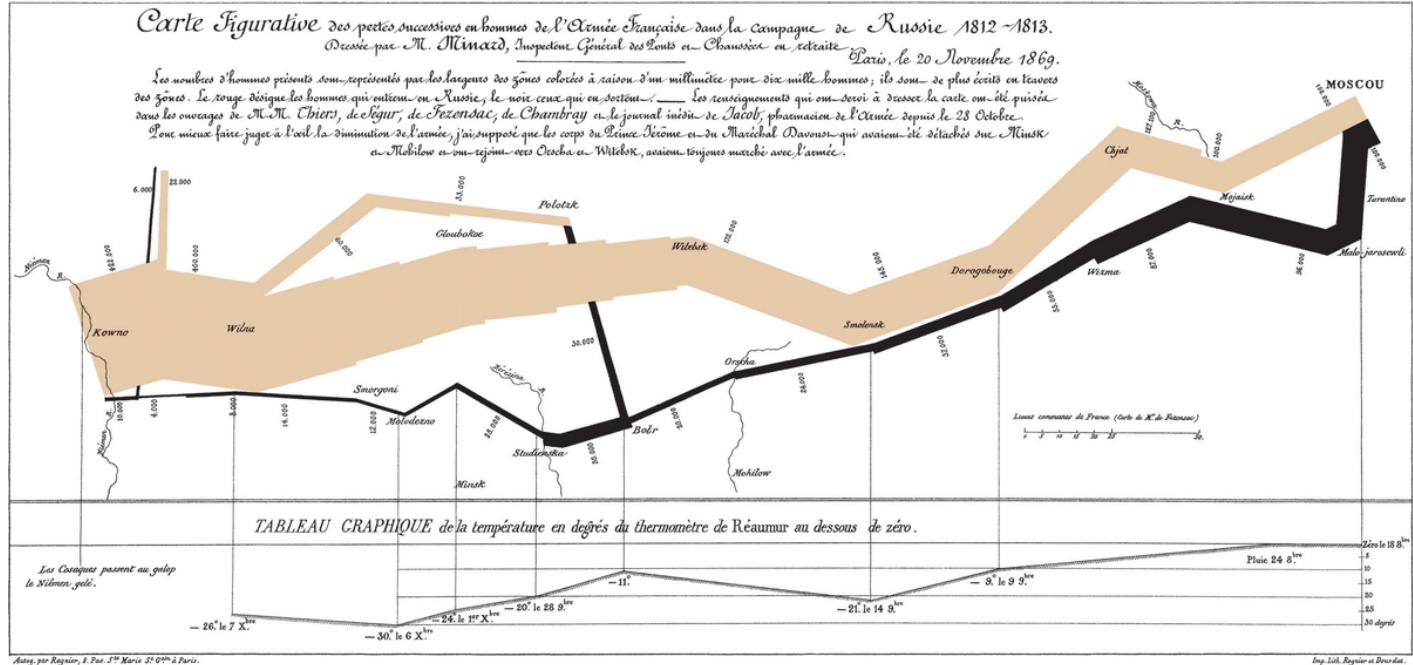
jumping from height



other method\*

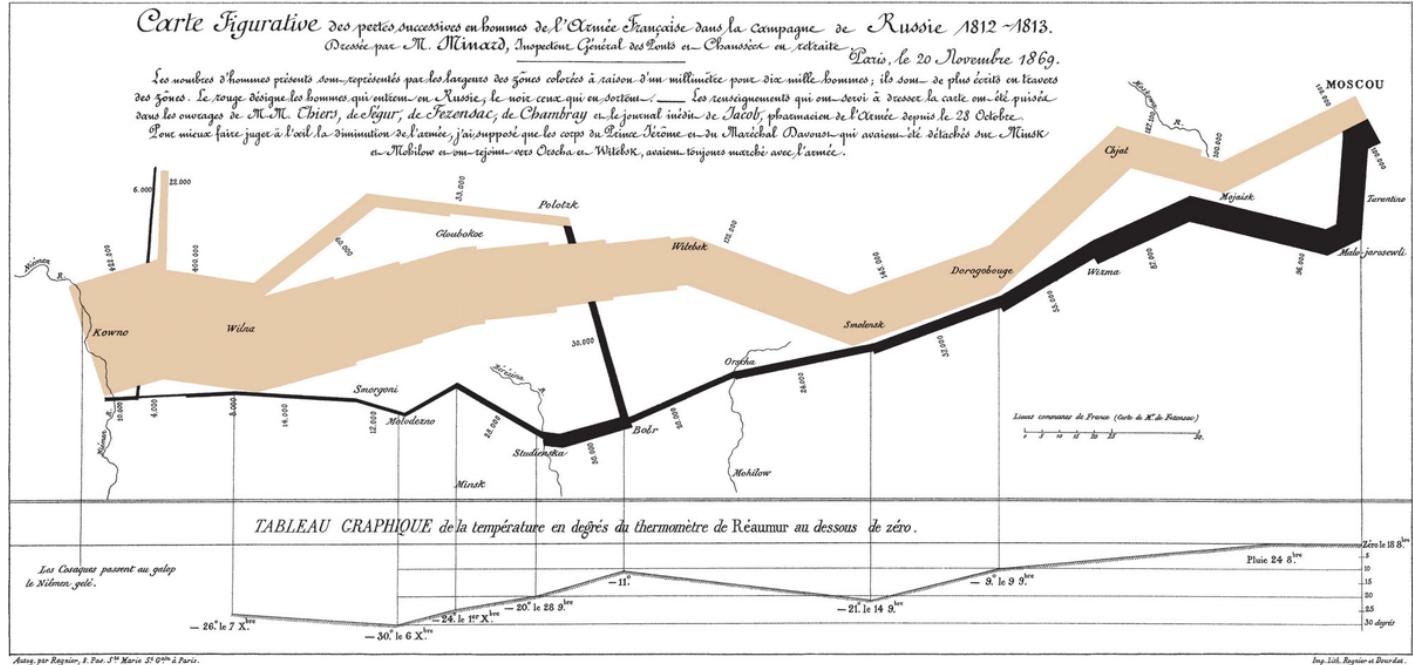


unknown method



"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

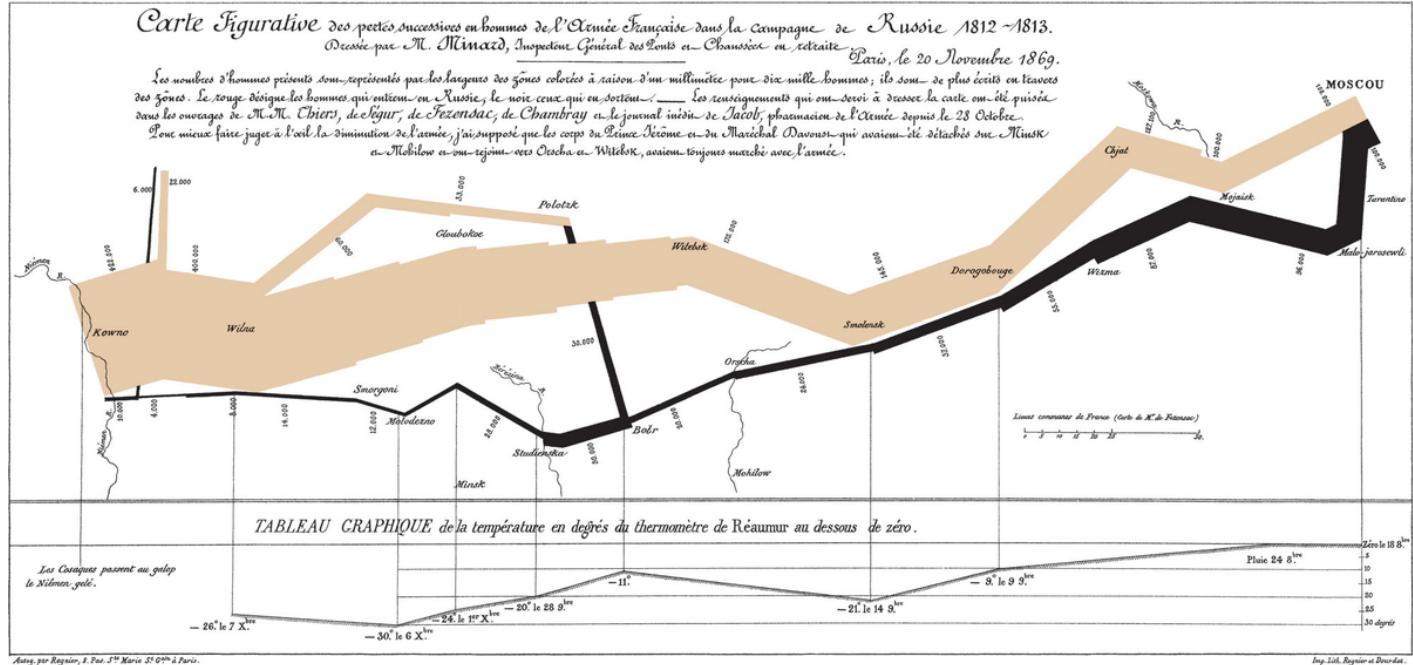
You might be wondering what you are viewing here.



"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

You might be wondering what you are viewing here.

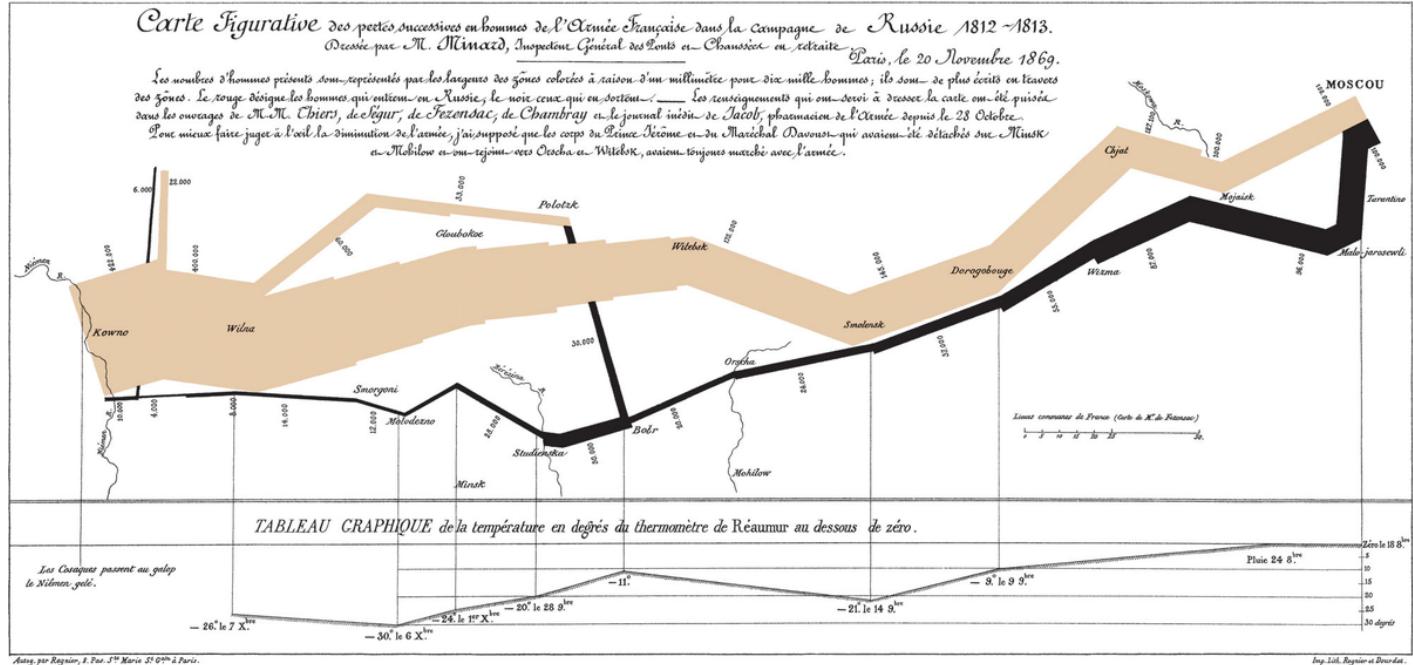
- shows the path of Napoleon's troops across the Russian Empire of Alexander I
- displays the progress of the troops in the form of a stream whose width indicates the size of the “Great Army”



"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

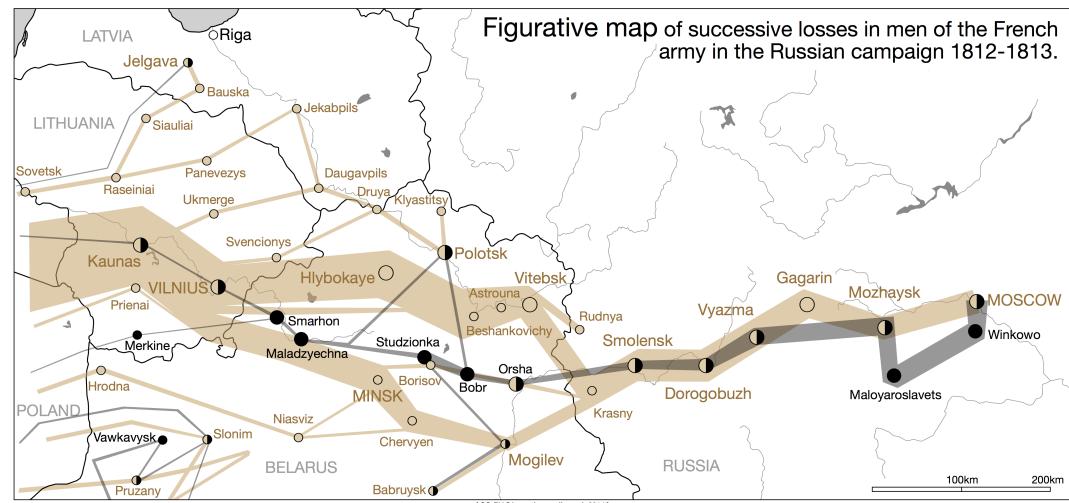
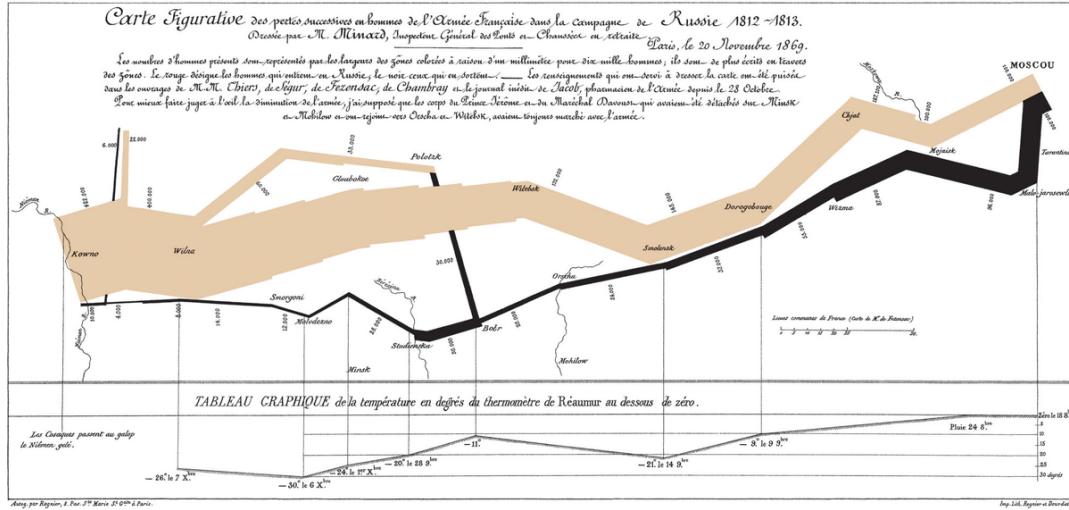
- encodes 6 variables in a simple (?) and modern way:

- width → size of Napoleon's army
- x-axis → longitude of the army's position
- y-axis → latitude of the army's position
- color → direction of the army's movement
- line chart → temperature during the army's retreat
- annotations → locations and army size (main chart) + date along retreat path (line chart)

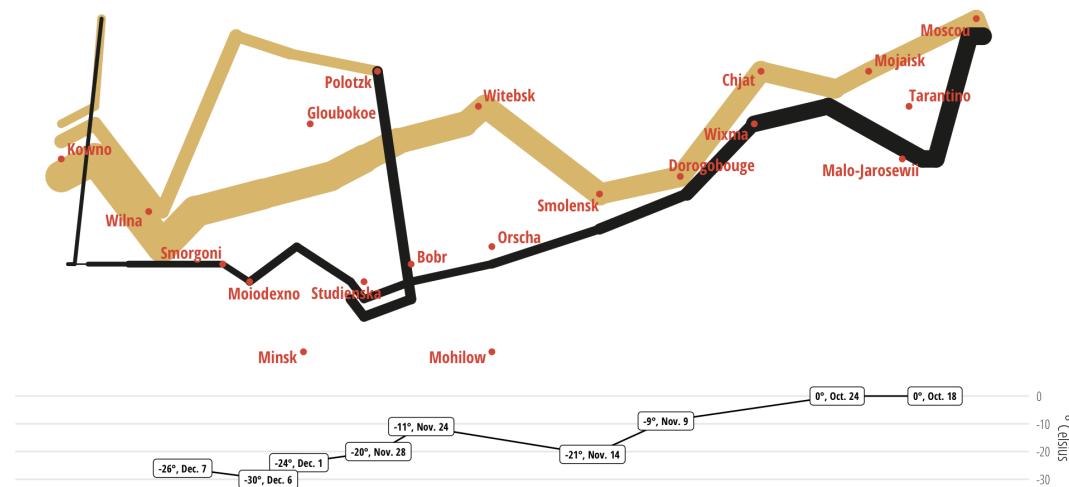
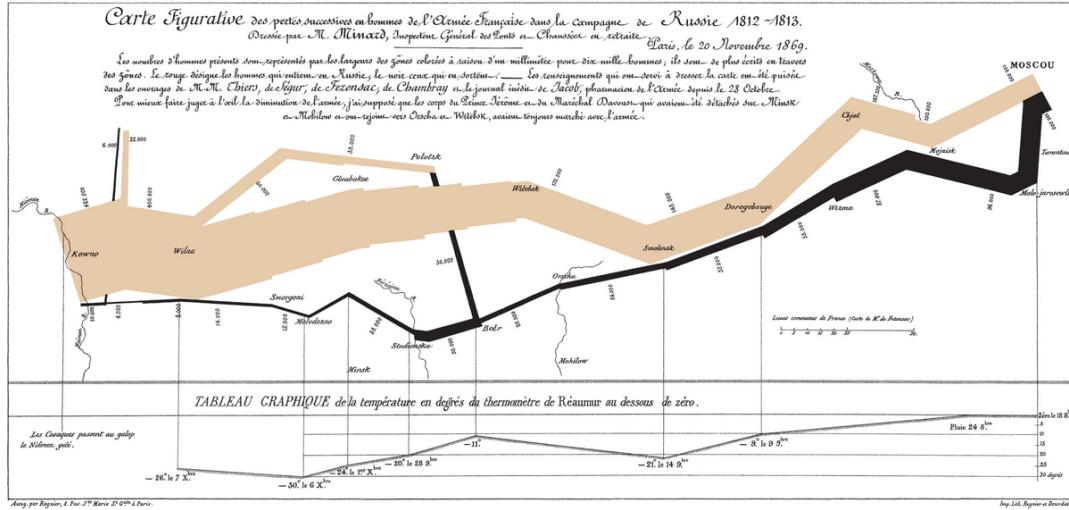


"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

- generally considered as the first data visualization (1869)
- Edward Tufte calls Minard's graphic of Napoleon in Russia one of the “best statistical drawings ever created”.



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



The map created by Charles Joseph Minard and a version coded in **ggplot2** by Andrew Heiss

The map even made it into the article "A Layered Grammar of Graphics" by Hadley Wickham that introduced [ggplot2](#)

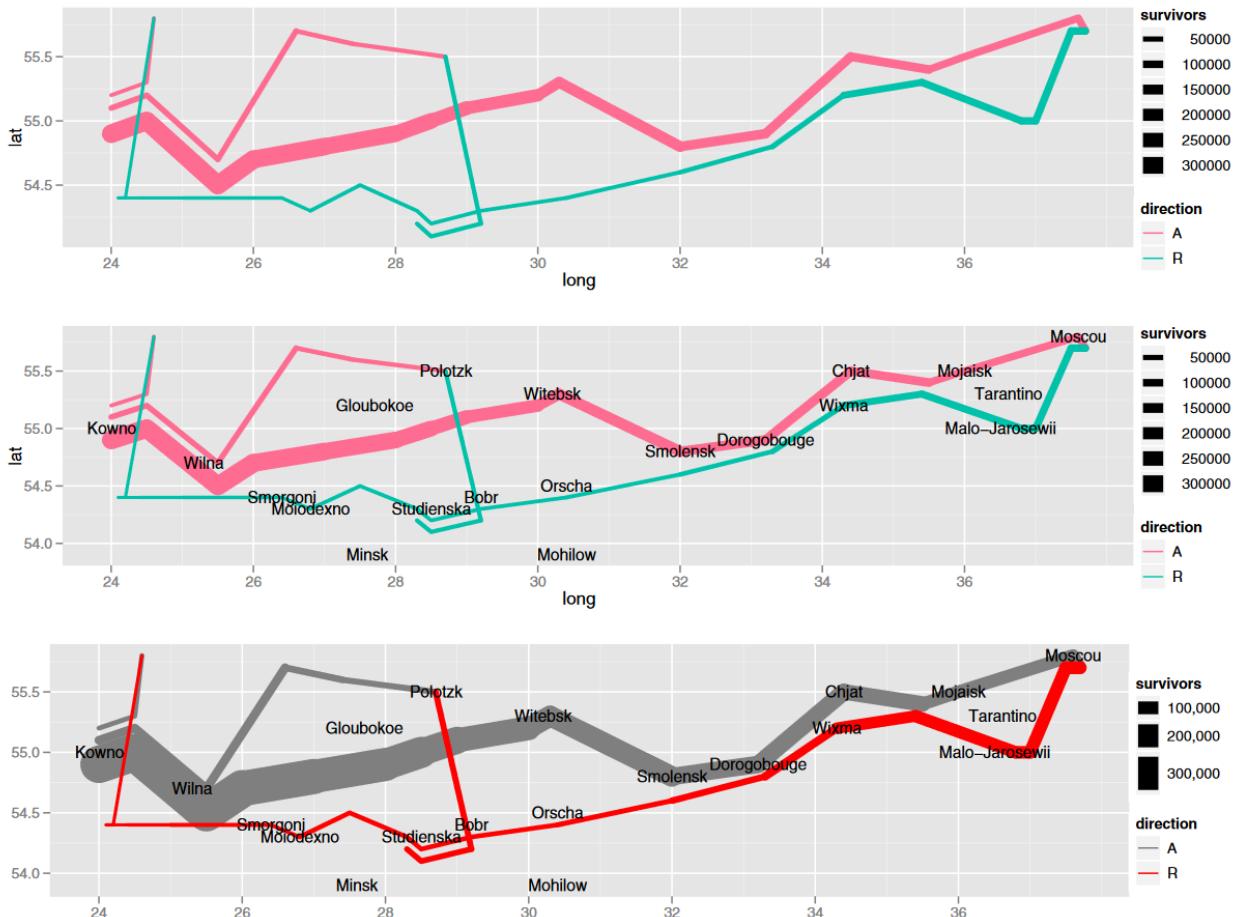


Figure 12. Iteratively reproducing the depiction of Napoleon's March by Minard. (Top) Displaying the key troop movement data. (Center) Adding town locations as reference points. (Bottom) Tweaking scales to produce polished plot.

# Data Visualization is part art and part science.

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

# Data visualization is part art and part science.

- *The challenge is to get the art right without getting the science wrong and vice versa.*
- *A data visualization first and foremost has to accurately convey the data.*
- *At the same time, a data visualization should be aesthetically pleasing.*
- *If a visualization is "good" or "bad" matters for both communication and impact!*



# **How to Develop an Eye for Good Data Visualization**

→ Information

→ Story

→ Goal

→ Visual Form

# **How to Develop an Eye for Good Data Visualization**

→ **Information (Integrity)**

→ **Story (Interestingness)**

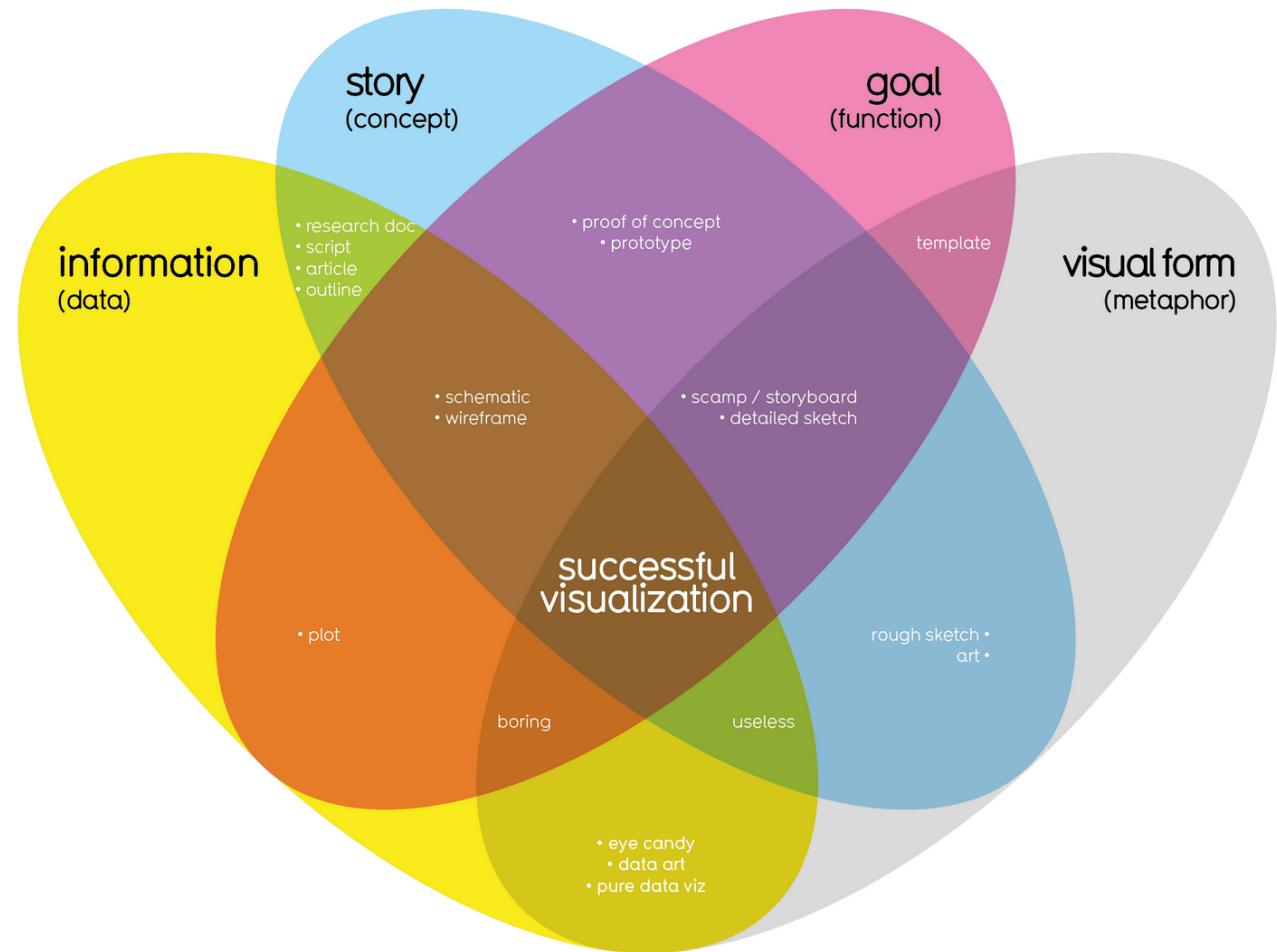
→ **Goal (Usefulness)**

→ **Visual Form (Beauty)**

# How to Develop an Eye for Good Data Visualization

- **Information** Understand your data and be accurate
- **Story** Be clear about the story of your visualization
- **Goal** Select charts that successfully transport your story
- **Visual Form** Follow design rules and data visualization principles
  - + inspiration, training and (a bit of) talent*

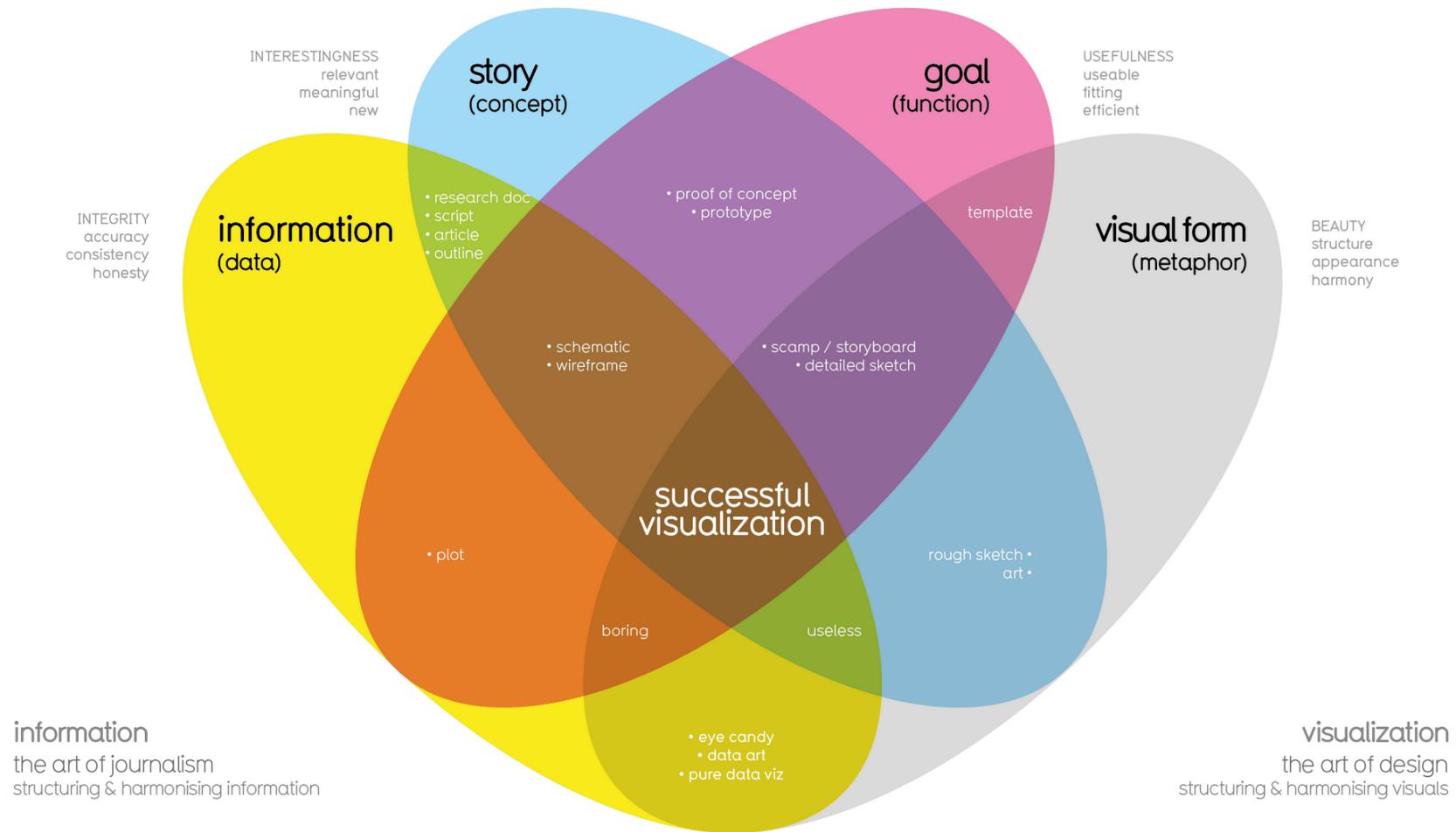
# What Makes a Good Visualization?

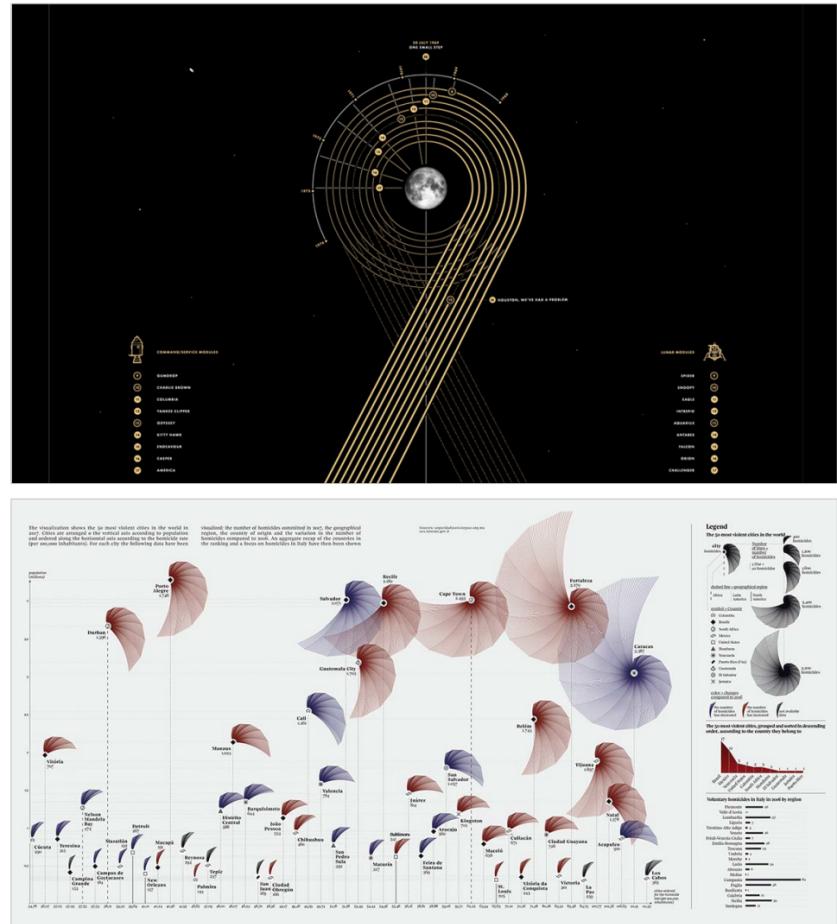
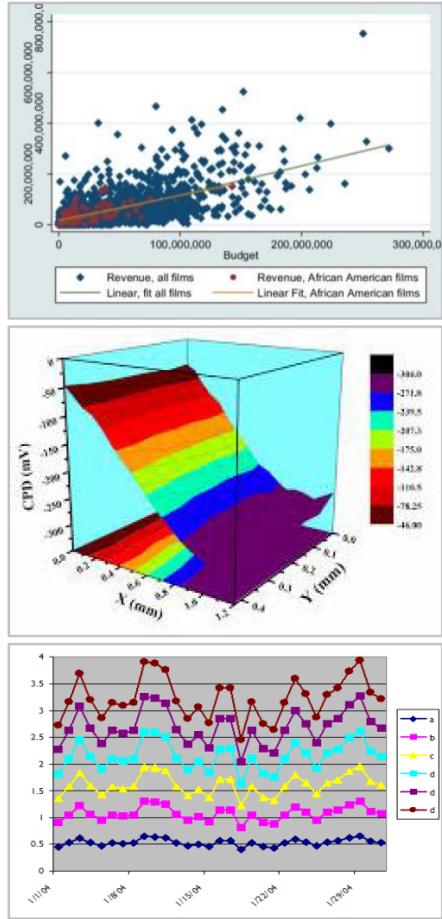


*Visualization by David McCandless ([Information is Beautiful](#))*

# What Makes a Good Visualization?

explicit (implicit)

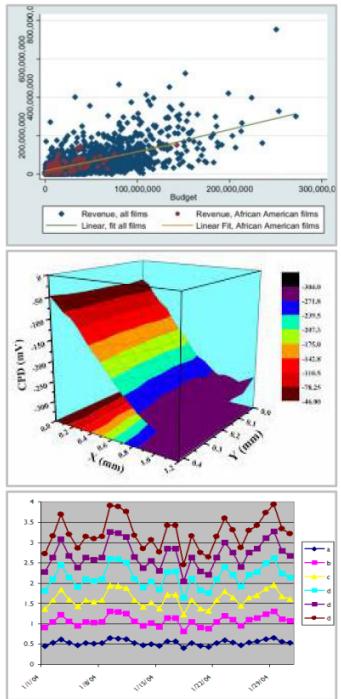




Anonymous

Sonia Kuijpers

Upper: Paul Button  
Lower: Frederica Fragapane



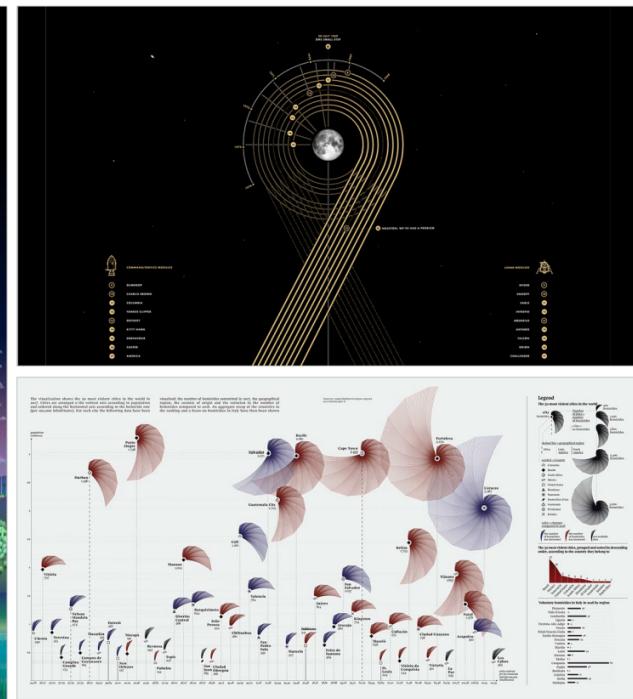
Anonymous

We aim for DataViz that:

- is informative
- reduces complexity
- is easy to grasp
- is visually appealing
- draws attention



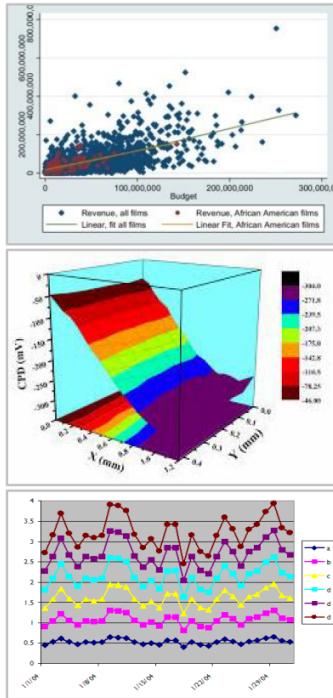
Sonia Kuijpers



Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from poorly designed & uninformative data visualization to data art



Anonymous

We aim for DataViz that:

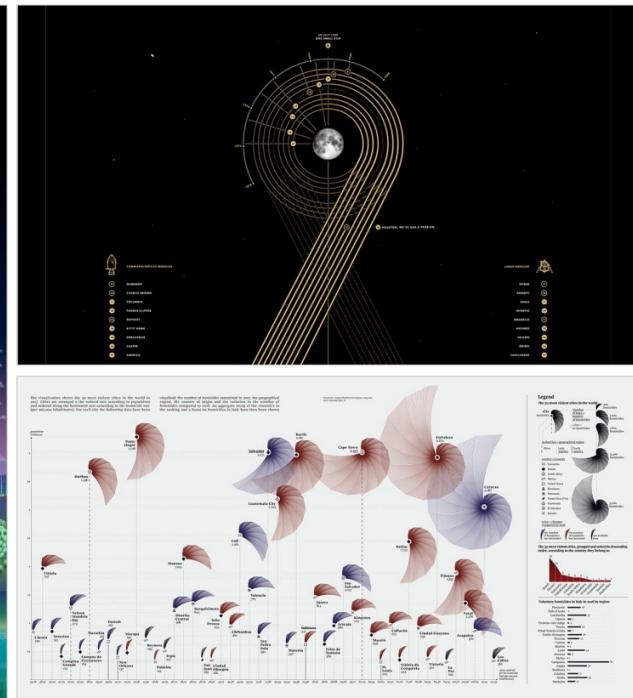
- is informative
- reduces complexity
- is easy to grasp
- is visually appealing
- draws attention

but:

- is not too abstract
- is not too “unusual”



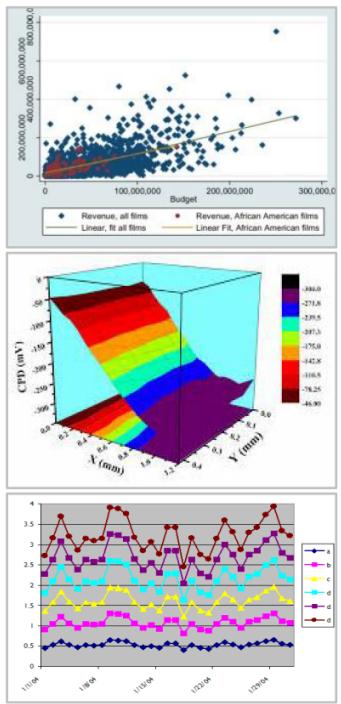
Sonia Kuijpers



Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from poorly designed & uninformative data visualization to data art



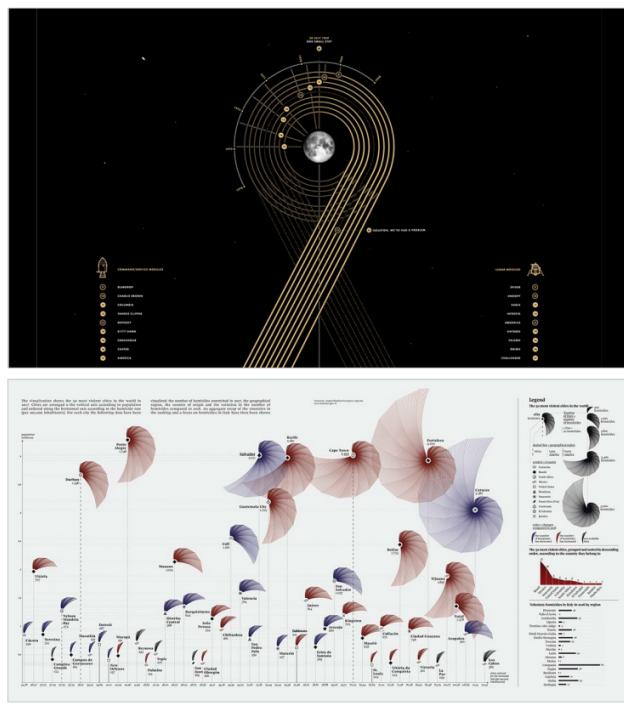
Anonymous



Upper: Cédric Scherer  
Lower: Jake Kaupp



Sonia Kuijpers



Upper: Paul Button  
Lower: Frederica Fragapane



Gradient from poorly designed & uninformative data visualization to data art

Know Your  
**Types of 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"?

# 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

# 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"?

# 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

# 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



# CONTINUOUS

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



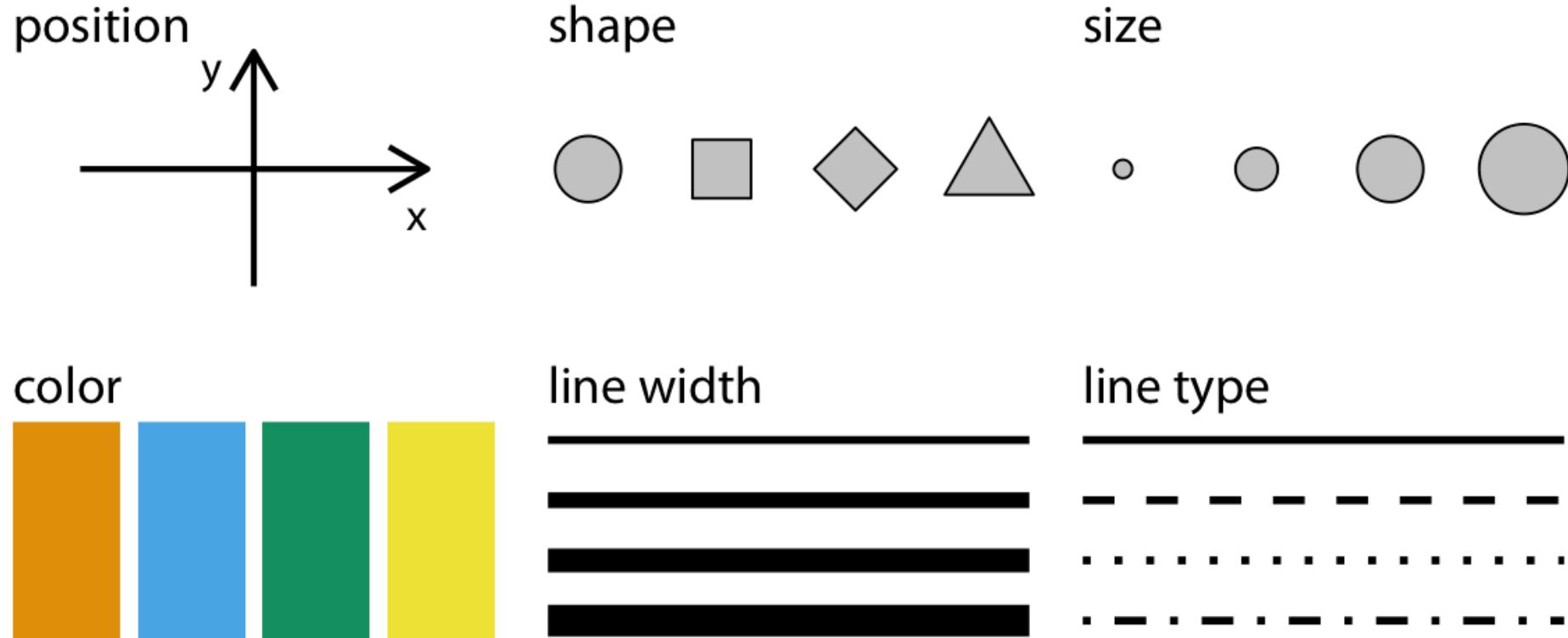
# DISCRETE

OBSERVATIONS CAN ONLY EXIST  
AT LIMITED VALUES, OFTEN COUNTS.



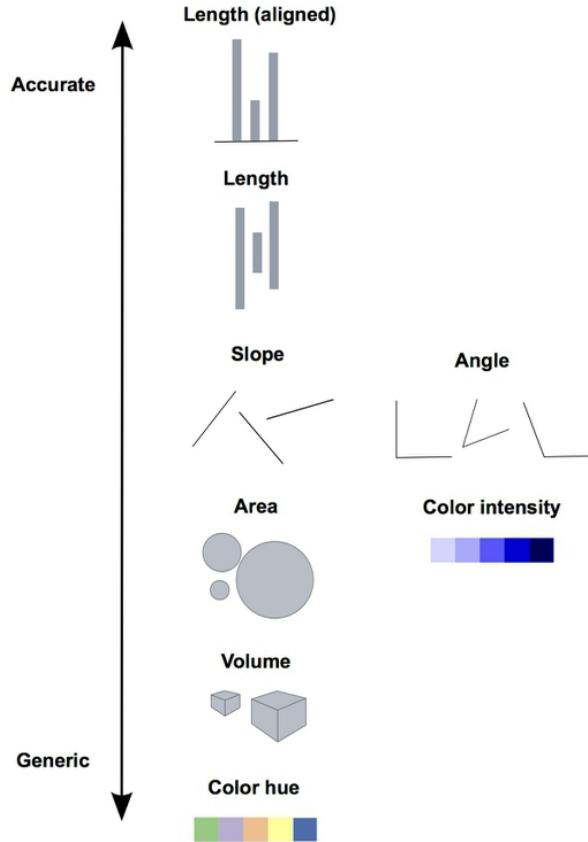
# Mapping Data to Aesthetics

# Data visualizations map values into quantifiable features (aesthetics)



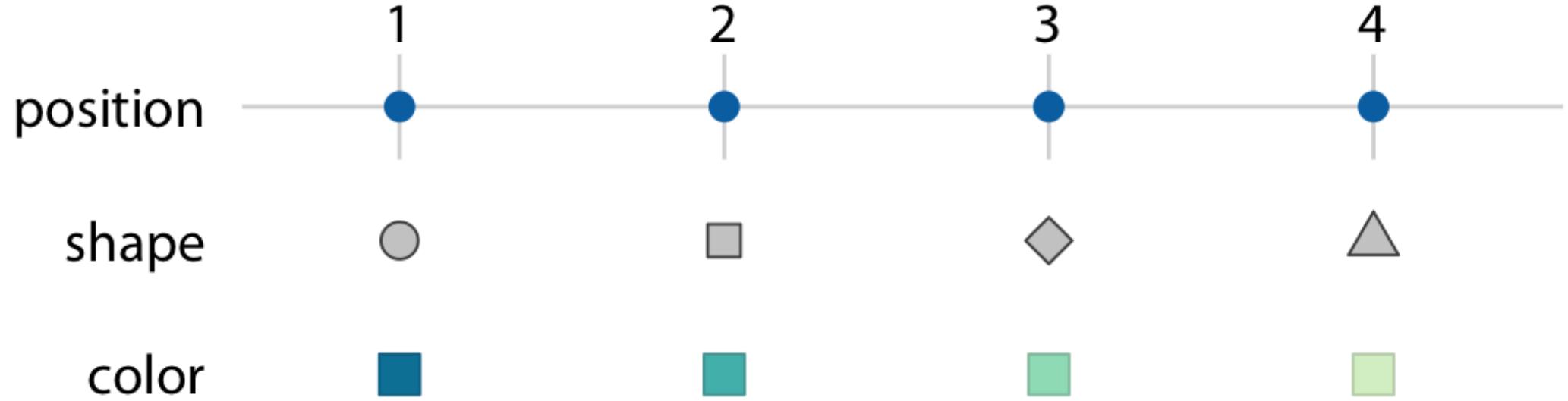
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Data visualizations map values into quantifiable features (aesthetics)



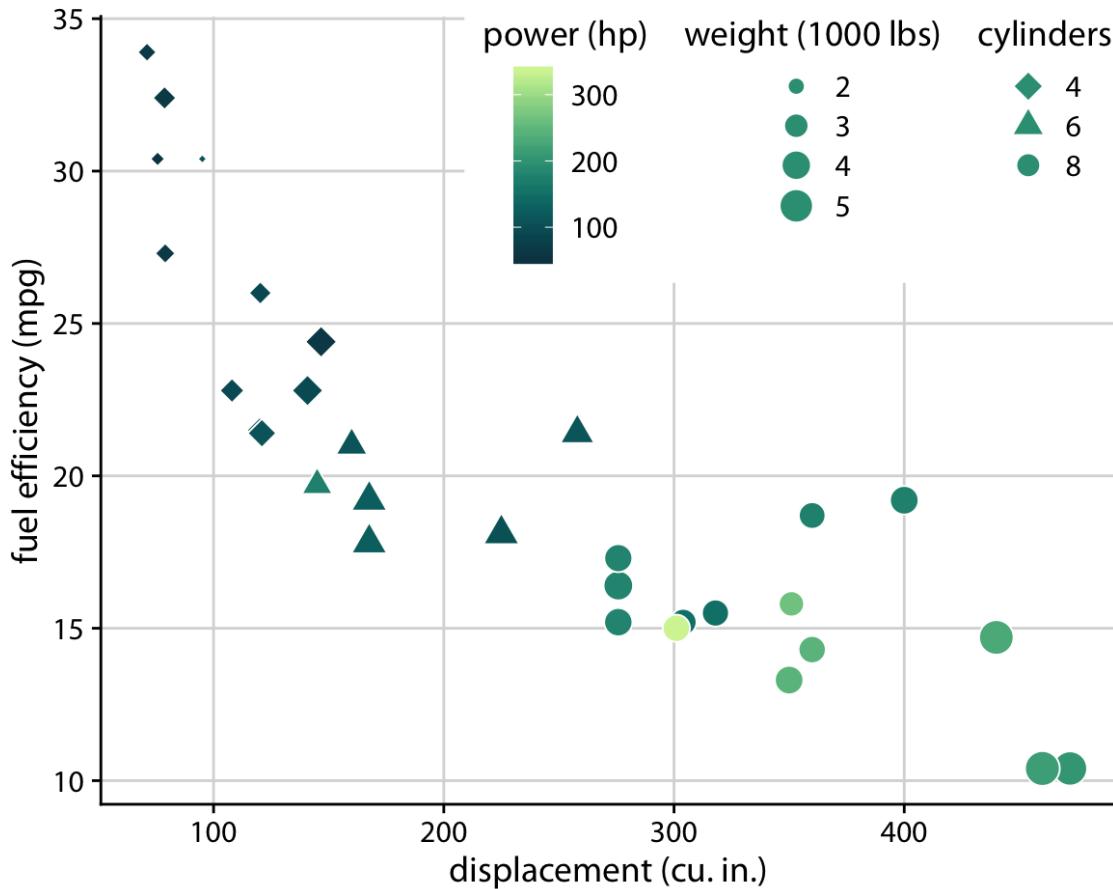
*Source: Peter Aldhous based on experiments by William Cleveland and Robert McGill*

# Scales map data values onto aesthetics



Source: "Fundamentals of Data Visualization" by Claus Wilke

# Scales map data values onto aesthetics



Source: "Fundamentals of Data Visualization" by Claus Wilke

# Colors and Common Pitfalls

# Color Terminology

**Hue:** color, like blue or red

**Chroma:** how pure a color is (saturation)

**Value:** how light or dark a color is

**Tint:** created by adding white to a hue

**Tone:** created by adding grey to a hue

**Shade:** created by adding black to a hue



# 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



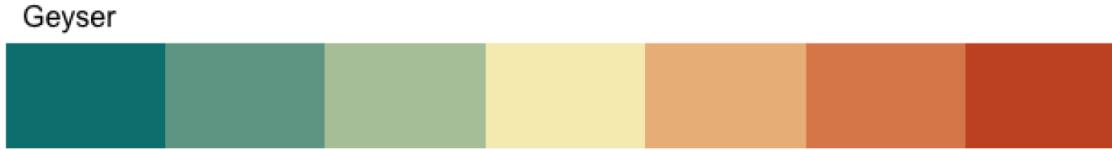
Sequential: Single-Hue



Sequential: Multi-Hue



Diverging



# Color Palette Types

Categorical



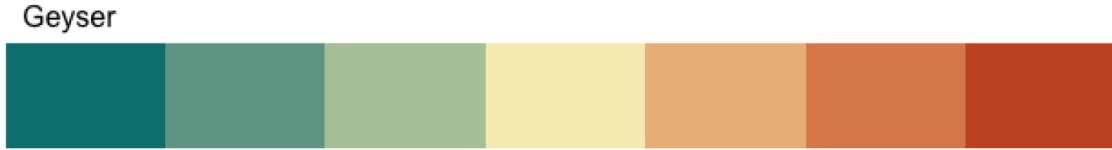
Sequential: Single-Hue



Sequential: Multi-Hue



Diverging



Cyclical



## 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

## 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](#)

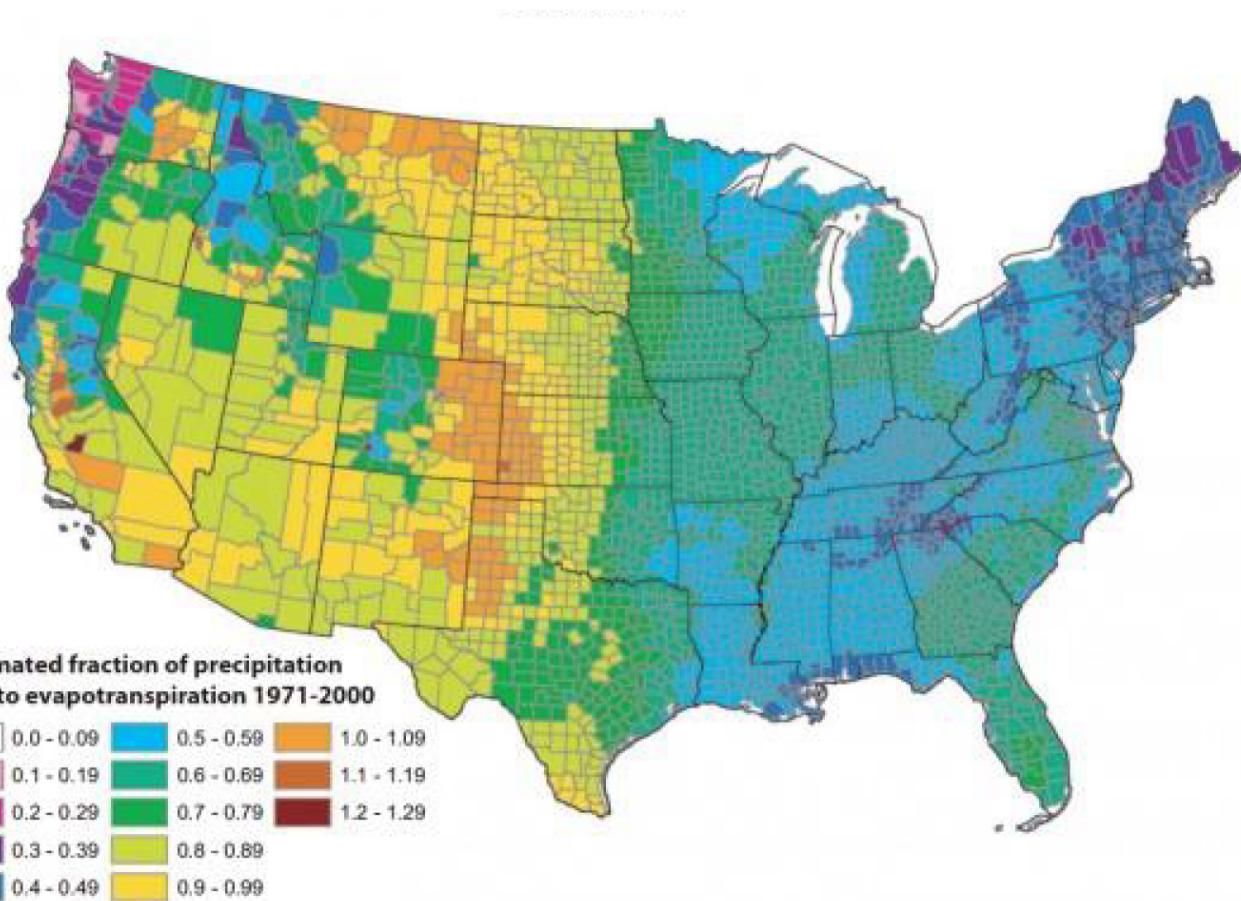


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of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

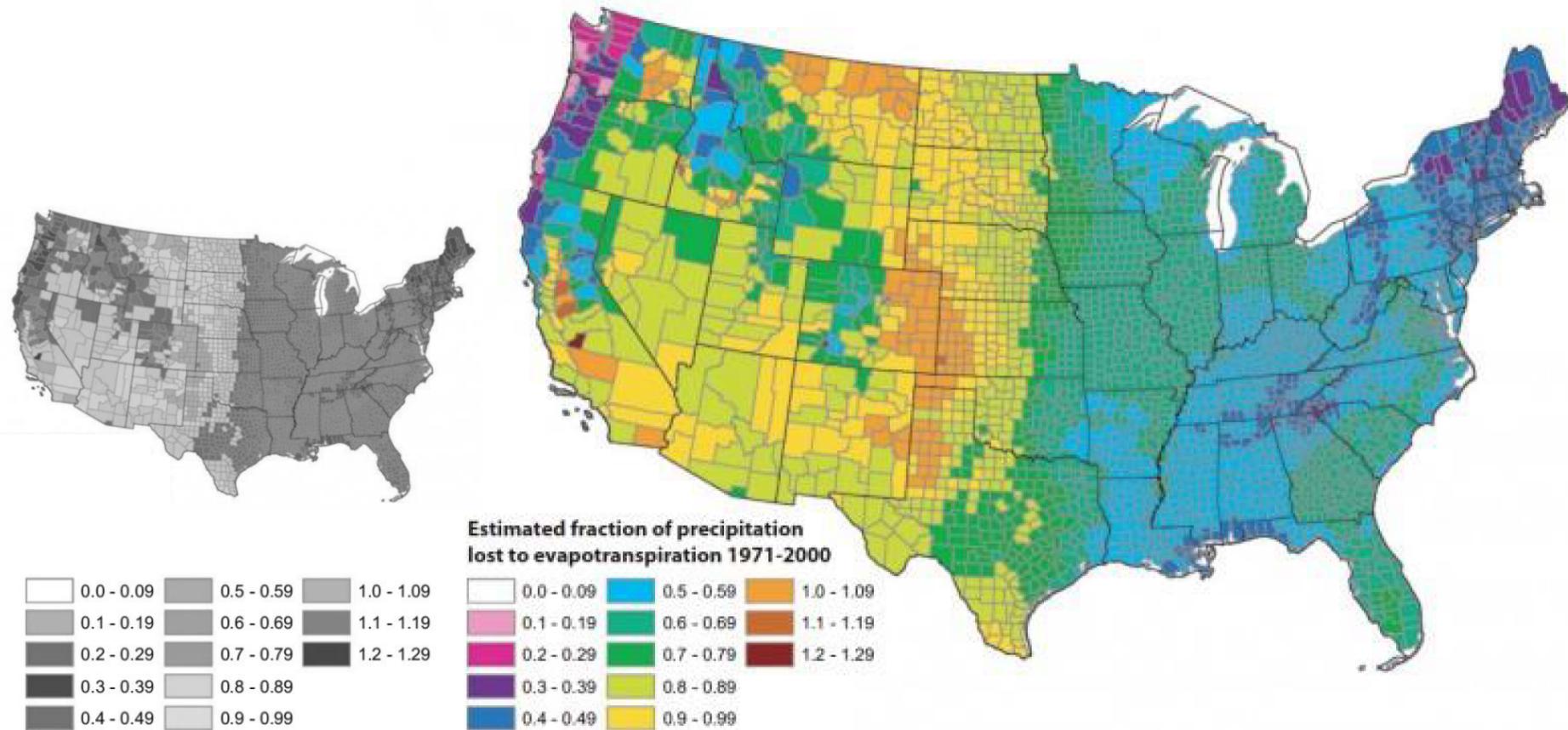


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of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

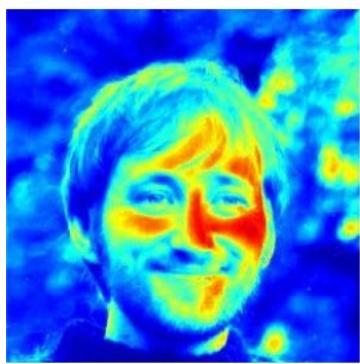


**true-colour Phil**

*Source: fabiocramerich/batlow*



**true-colour Phil**

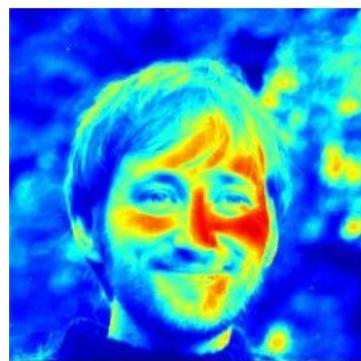


**rainbow Phil**  
*is distorted*

*Source: fabiocramerich/batlow*



**true-colour Phil**



**rainbow Phil**  
*is distorted*



**batlow Phil**  
*is flawless*

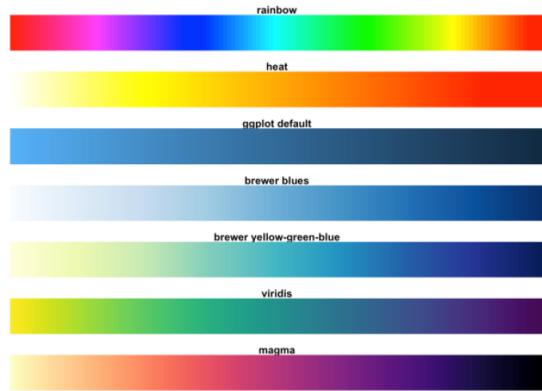
*Source: fabiocramerich/batlow*

# Choice of Color Palettes

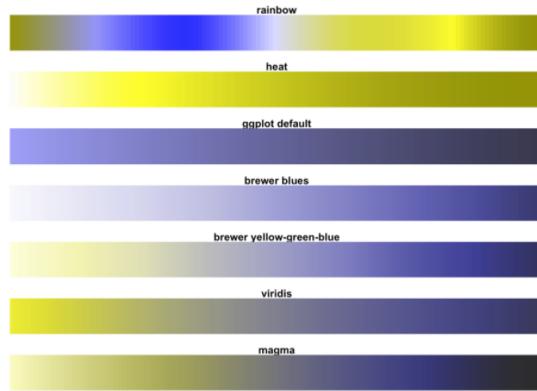


*Source: cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html*

# Choice of Colors Palettes & Color-Vision Deficiency (CVD)



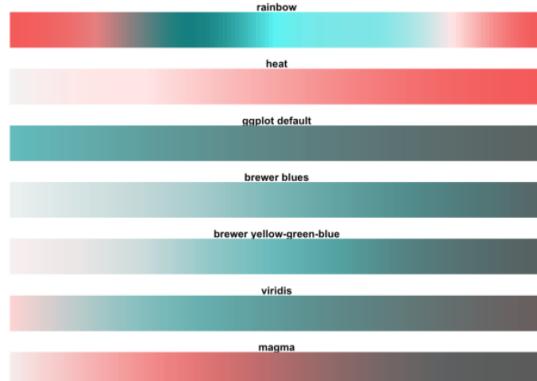
**Deutanopia:** present in 6% of males



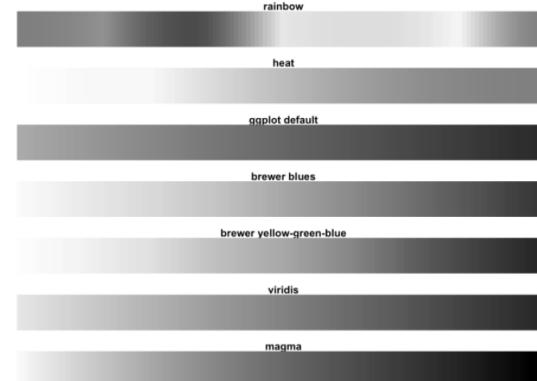
**Protanopia:** present in 1% of males



**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)

**To make sure your visualizations work for people with CVD don't just rely on provided color palettes.**

**To make sure your visualizations work for people with CVD don't just rely on provided color palettes.**

**Instead, test your figures in a color-blindness simulator!**

# Choice of the Color Palette & 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)

# Choice of the Color Palette & Accessibility

original



deuteranomaly



protanomaly

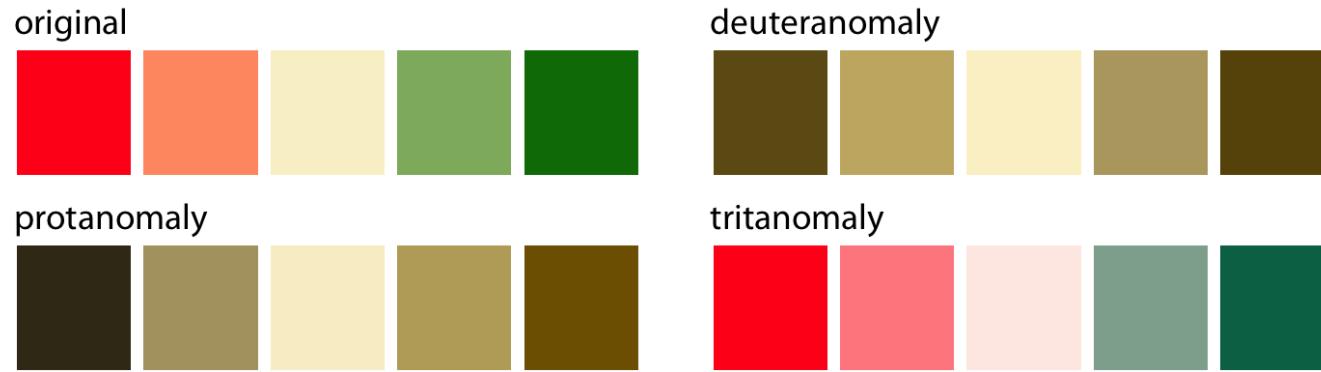


tritanomaly



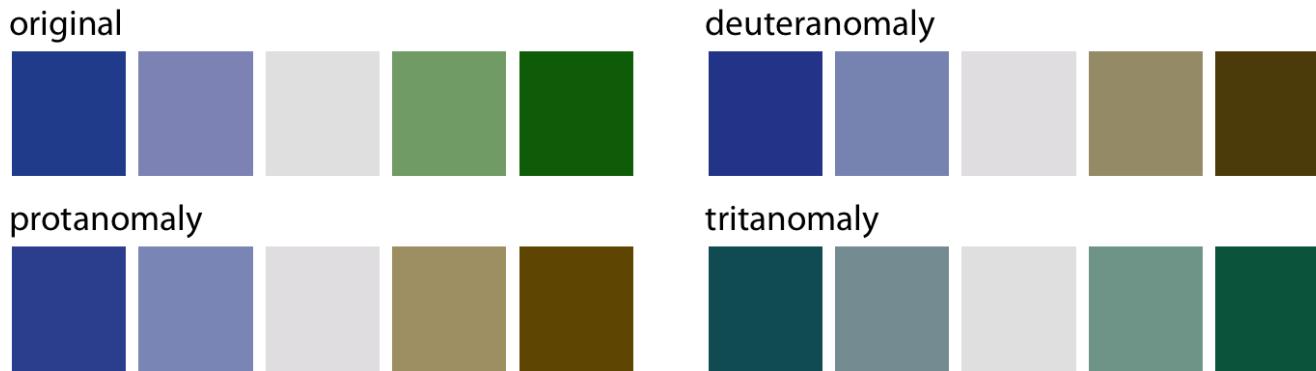
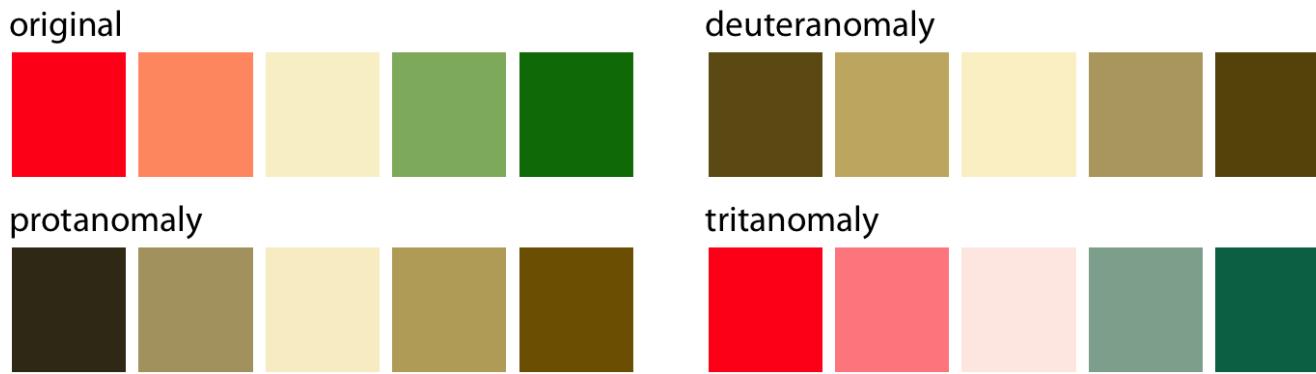
Source: "Fundamentals of Data Visualization" by Claus Wilke

# Choice of the Color Palette & Accessibility



Source: "Fundamentals of Data Visualization" by Claus Wilke

# Choice of the Color Palette & Accessibility



Source: "Fundamentals of Data Visualization" by Claus Wilke

# Choice of the Color Palette & 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

Add

Replace

### GET

String quotes

Object with metadata

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

#hex

rgb

hsl

## COLORS IN ACTION

Background color: #ffffff

Font color: #000000

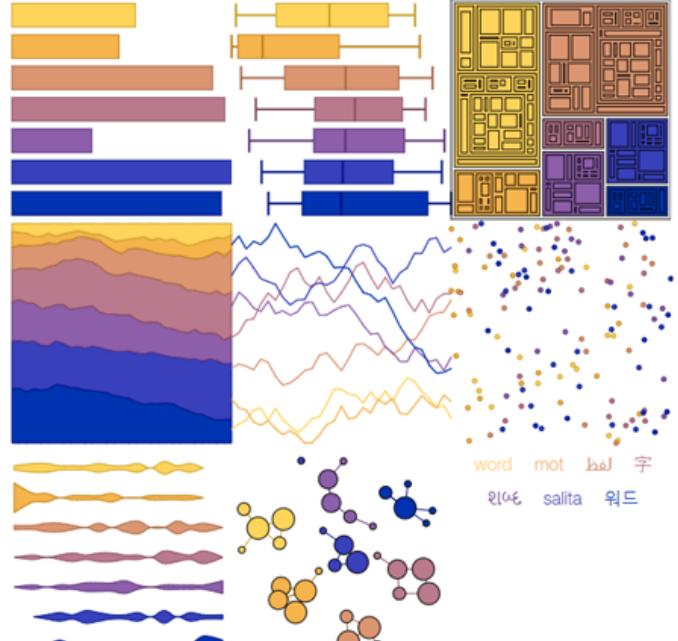
Charts made with Semiotic

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

Sample font

Randomize Data

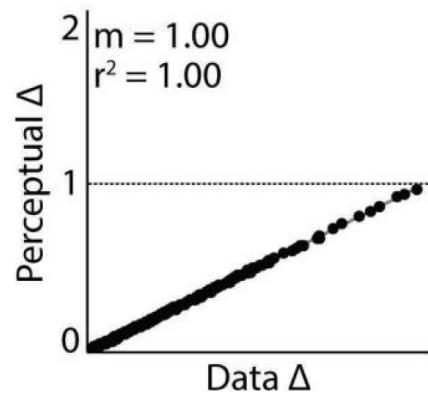
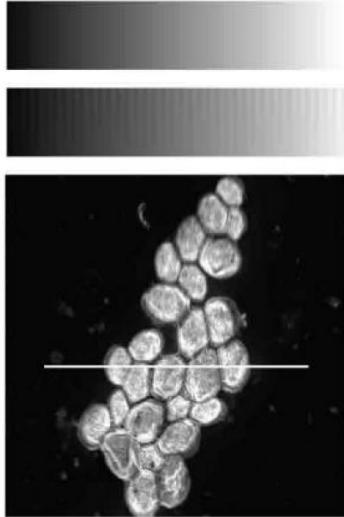
Stroke: Dark | None



word mot baj 字  
শব্দ salita 워드

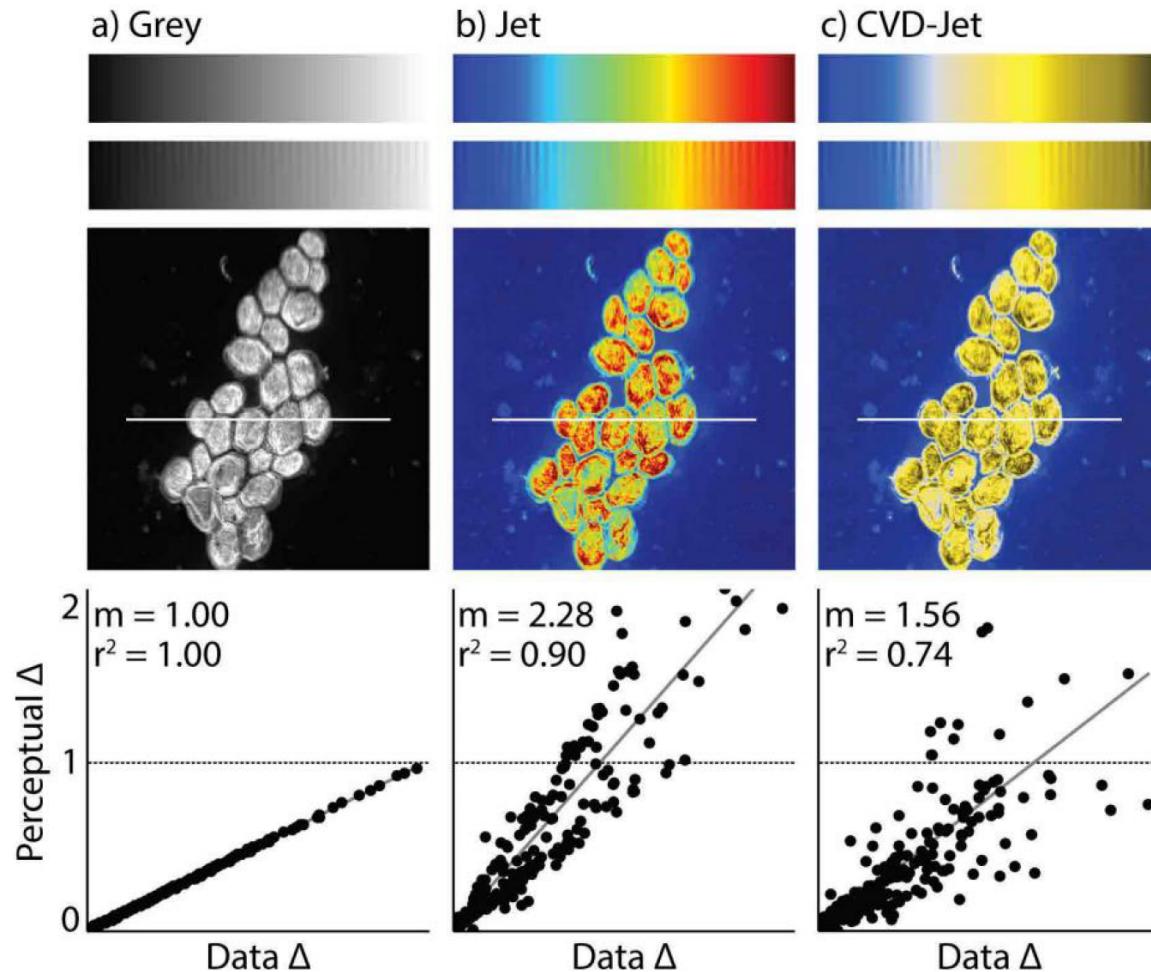
# Choice of the Color Palette & Accessibility

a) Grey



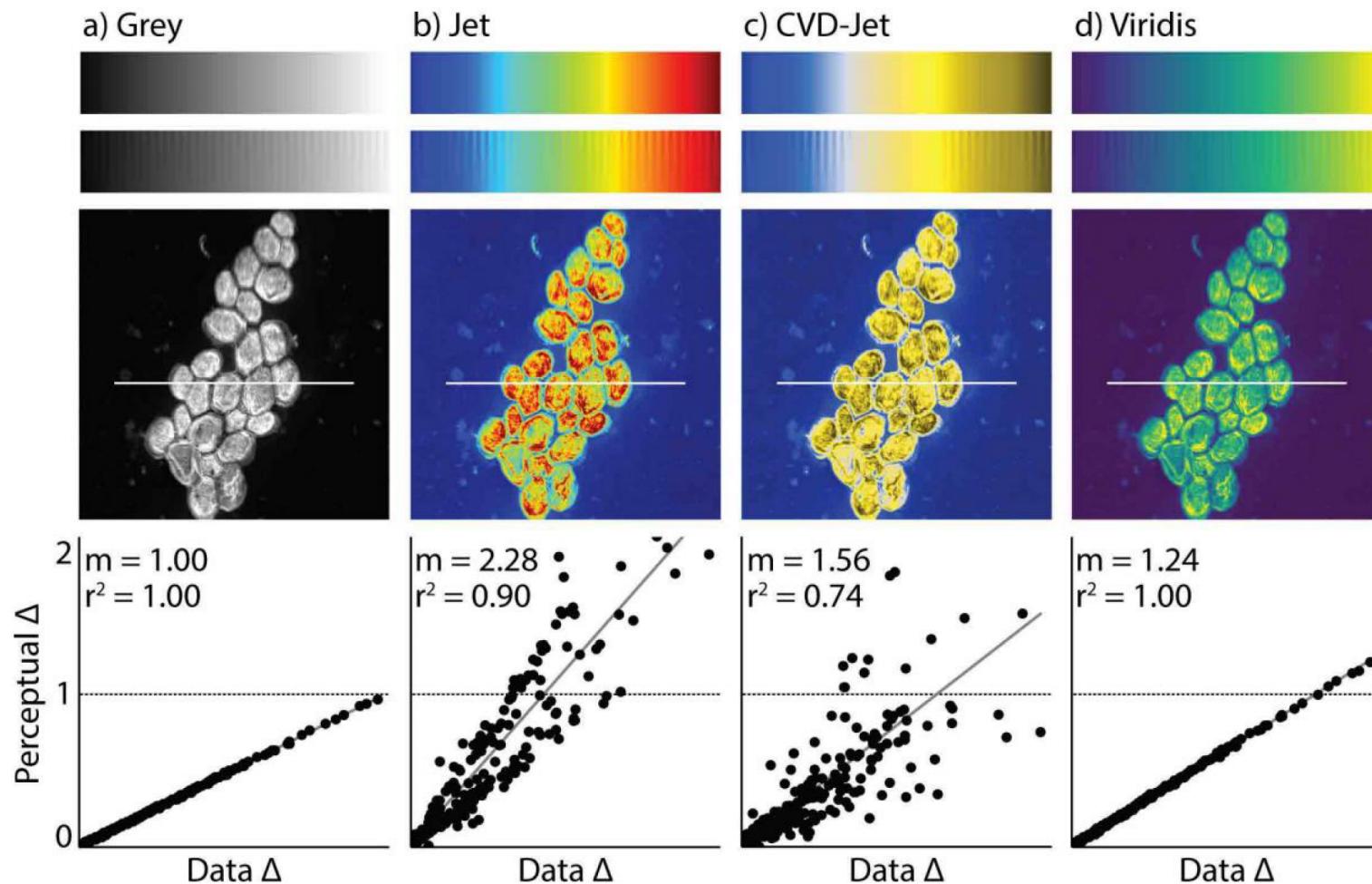
Modified from Nuñez, Anderton & Renslow (2018) PLOSone

# Choice of the Color Palette & Accessibility



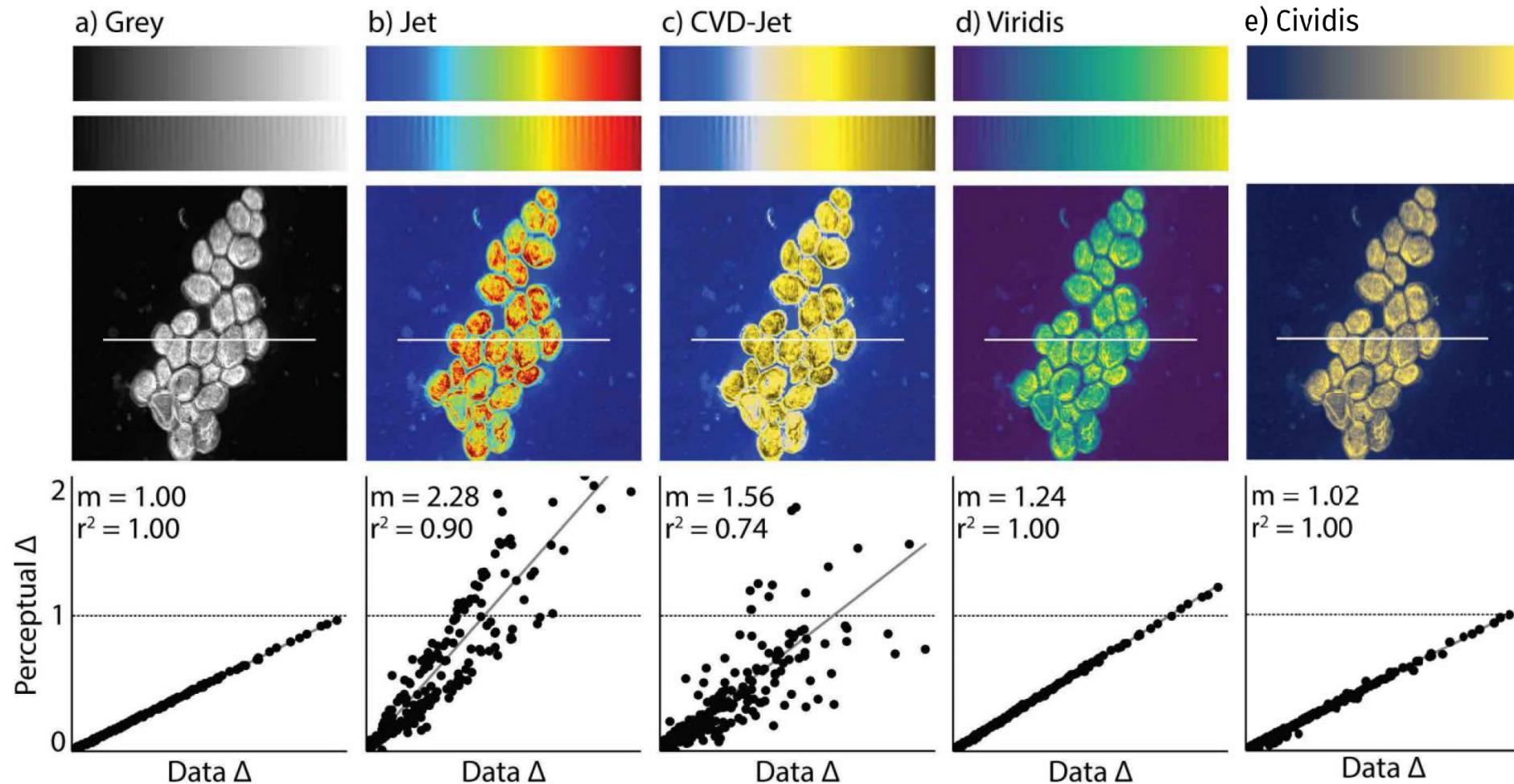
Modified from Nuñez, Anderton & Renslow (2018) PLOSone

# Choice of the Color Palette & Accessibility



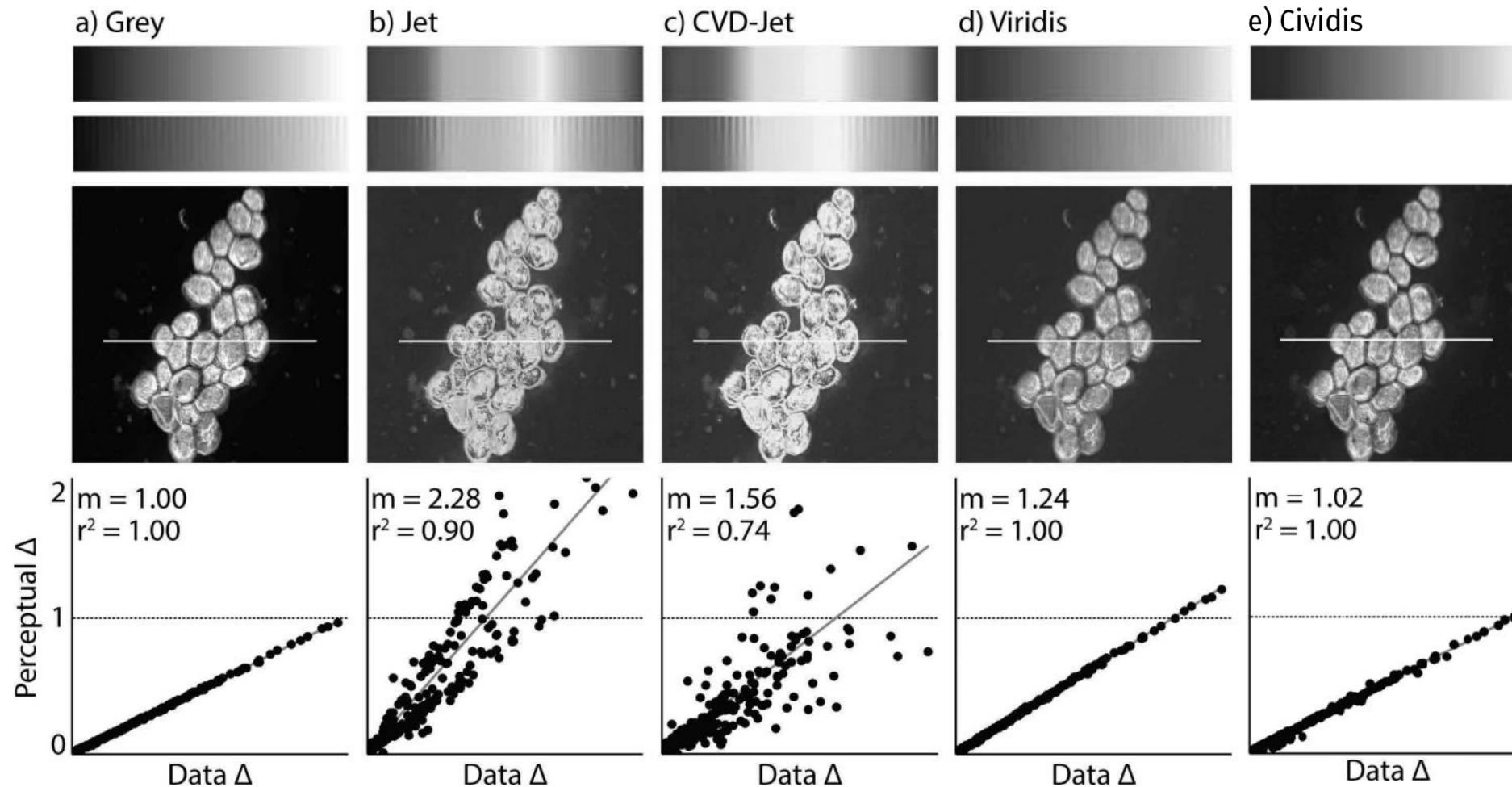
Modified from Nuñez, Anderton & Renslow (2018) PLOSone

# Choice of the Color Palette & Accessibility



Modified from Nuñez, Anderton & Renslow (2018) PLOSone

# Choice of the Color Palette & Accessibility



Modified from Nuñez, Anderton & Renslow (2018) PLOSone

The image is a graphic design centered around the word "TYPOGRAPHY". It features multiple instances of the word "TYPOGRAPHY" in different typefaces and sizes. On the left, there's a vertical stack of words: "TYPOGRAPHY" in a bold, serif font; "TYPOGRAPHY" in a smaller, sans-serif font; "PO" in a large, bold, serif font; "GRAP" in a bold, sans-serif font; and "HY" in a large, bold, serif font. To the right of these, the word "TYPOGRAPHY" is repeated in a large, bold, sans-serif font. Above this, the word "TYPOGRAPHY" is written in a large, bold, serif font. Below the main text area, there's a row of decorative ligatures: "G", "H", "P", "E", "T", "A", "R", "E", "S", "H", "W". At the bottom right, the word "Typography" is written in a large, bold, yellow sans-serif font.

# Typography

You'll always be mine! ❤

FONTS MATTER

 YOU'LL ALWAYS BE MINE!

# 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!

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

Yeah, I know I am kinda boring. Sorry.

# How to Visualize Hierarchy

I am important!

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

# How to Visualize Hierarchy

**I am important!**

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.

# Display Fonts

*A Tale of Two Cities*

Lobster Two

*A Tale of Two Cities*

Tangerine

A Tale of Two Cities

Raleway

**A Tale of Two Cities**

Abril Fatface

**A Tale of Two Cities**

Chunk

A TALE OF TWO CITIES

Cinzel

*Source: wordpress.com*

# Text Fonts

## A Tale of Two Cities

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way—in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

---

Open Sans

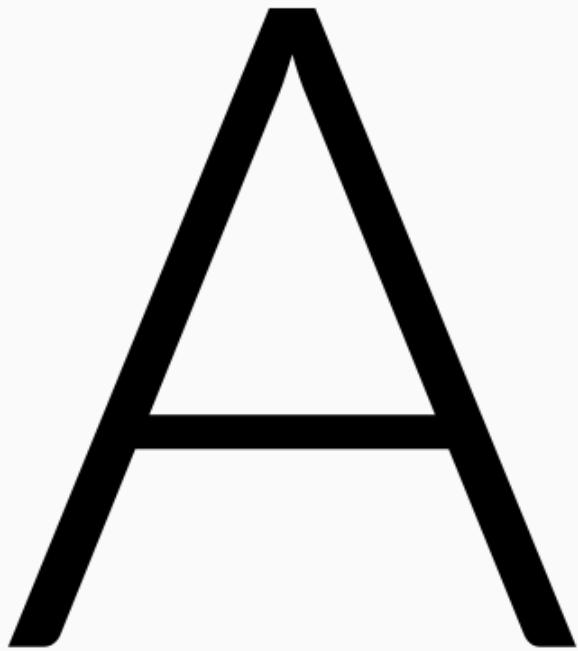
## A Tale of Two Cities

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way—in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.

---

Libre Baskerville

# Sans-Serif or Serif ?

A large, bold, black sans-serif letter 'A' centered on the page. It has a clean, modern appearance with straight lines and no decorative flourishes.

Lato

A large, bold, black serif letter 'A' centered on the page. It features traditional decorative elements like small vertical strokes at the top and bottom of the vertical stems and a distinct crossbar.

Gentium Book Basic

# 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*

# Tabular (Monospaced) Numbers

TABULAR

123.45  
678.90

PROPORTIONAL

123.45  
678.90

*Source: [invisionapp.com/inside-design/best-free-fonts-for-numbers](https://invisionapp.com/inside-design/best-free-fonts-for-numbers)*

# Number Fonts with Lining

LINING

123,456,789.0

OLDSTYLE

123,456,789.0

Source: [invisionapp.com/inside-design/best-free-fonts-for-numbers](https://invisionapp.com/inside-design/best-free-fonts-for-numbers)

# Quality of Number Symbols

NEUTON

\$123,456,789.00%

% is smaller than  
other figures

ECONOMICA

\$123,456,789.00%

\$ is smaller than  
other figures

SOURCE CODE PRO

\$123,456,789.00%

% has an  
uncommon design

MARVEL

\$123,456,789.00☒

It doesn't have the  
% symbol

Source: [invisionapp.com/inside-design/best-free-fonts-for-numbers](https://invisionapp.com/inside-design/best-free-fonts-for-numbers)

### Droid Serif

REGULAR

\$123,456,789.00%

BOLD

**\$123,456,789.00%**

### Crimson Text

REGULAR

\$123,456,789.00%

SEMIBOLD

\$123,456,789.00%

BOLD

**\$123,456,789.00%**

### Copse

REGULAR

\$123,456,789.00%

### Kameron

REGULAR

\$123,456,789.00%

BOLD

**\$123,456,789.00%**

### Open Sans

LIGHT \$123,456,789.00%  
REGULAR \$123,456,789.00%  
**BOLD** **\$123,456,789.00%**

### Roboto Condensed

LIGHT \$123,456,789.00%  
REGULAR \$123,456,789.00%  
**BOLD** **\$123,456,789.00%**

### Lato

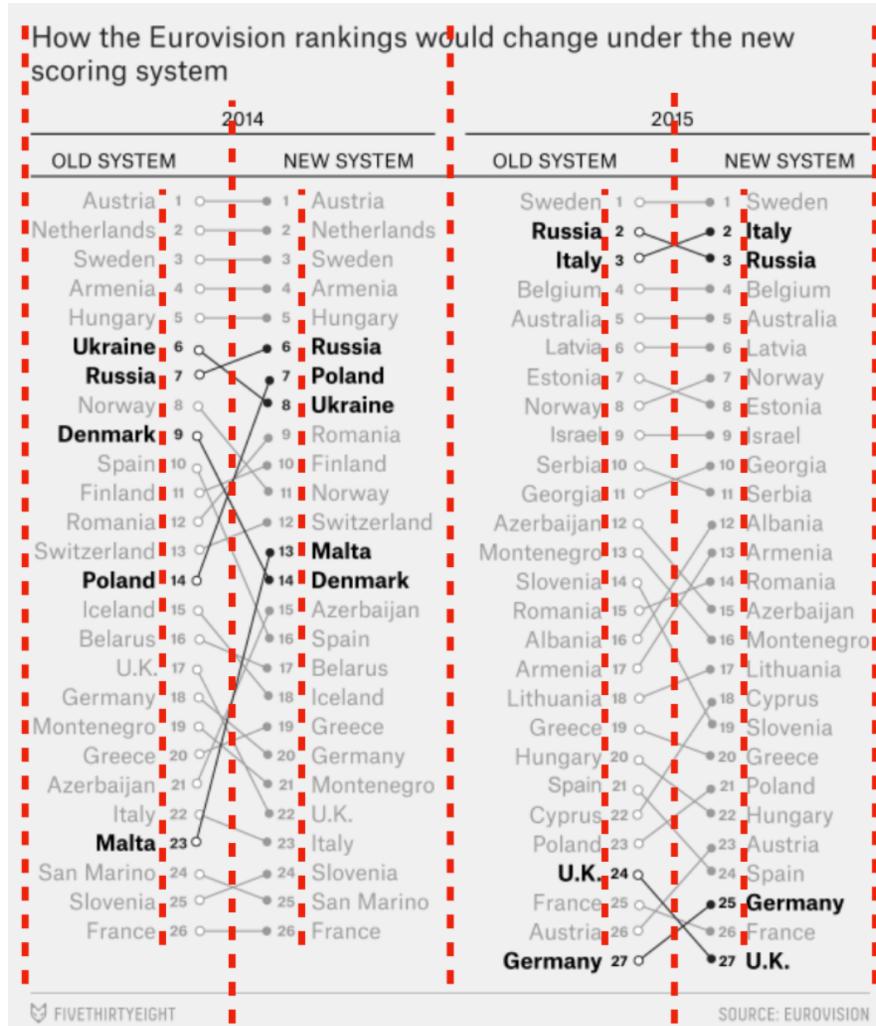
LIGHT \$123,456,789.00%  
REGULAR \$123,456,789.00%  
**BOLD** **\$123,456,789.00%**

### Varela Round

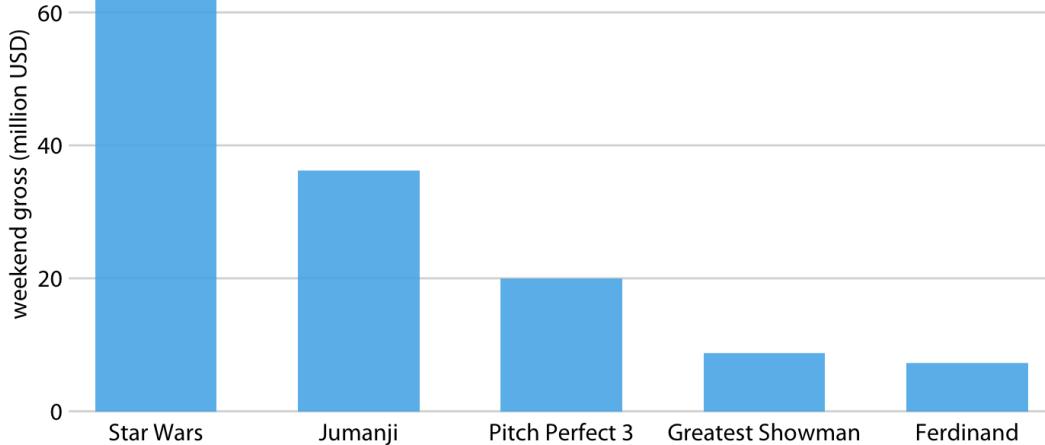
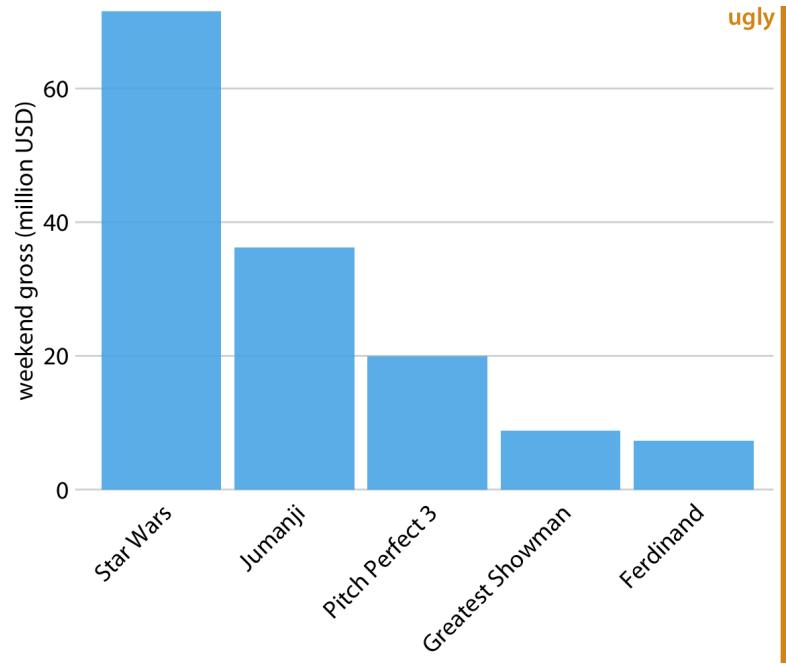
REGULAR \$123,456,789.00%

# Allign Your Text!

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



# (Don't) Rotate Your Text!



Source: "Fundamentals of Data Visualization" by Claus Wilke

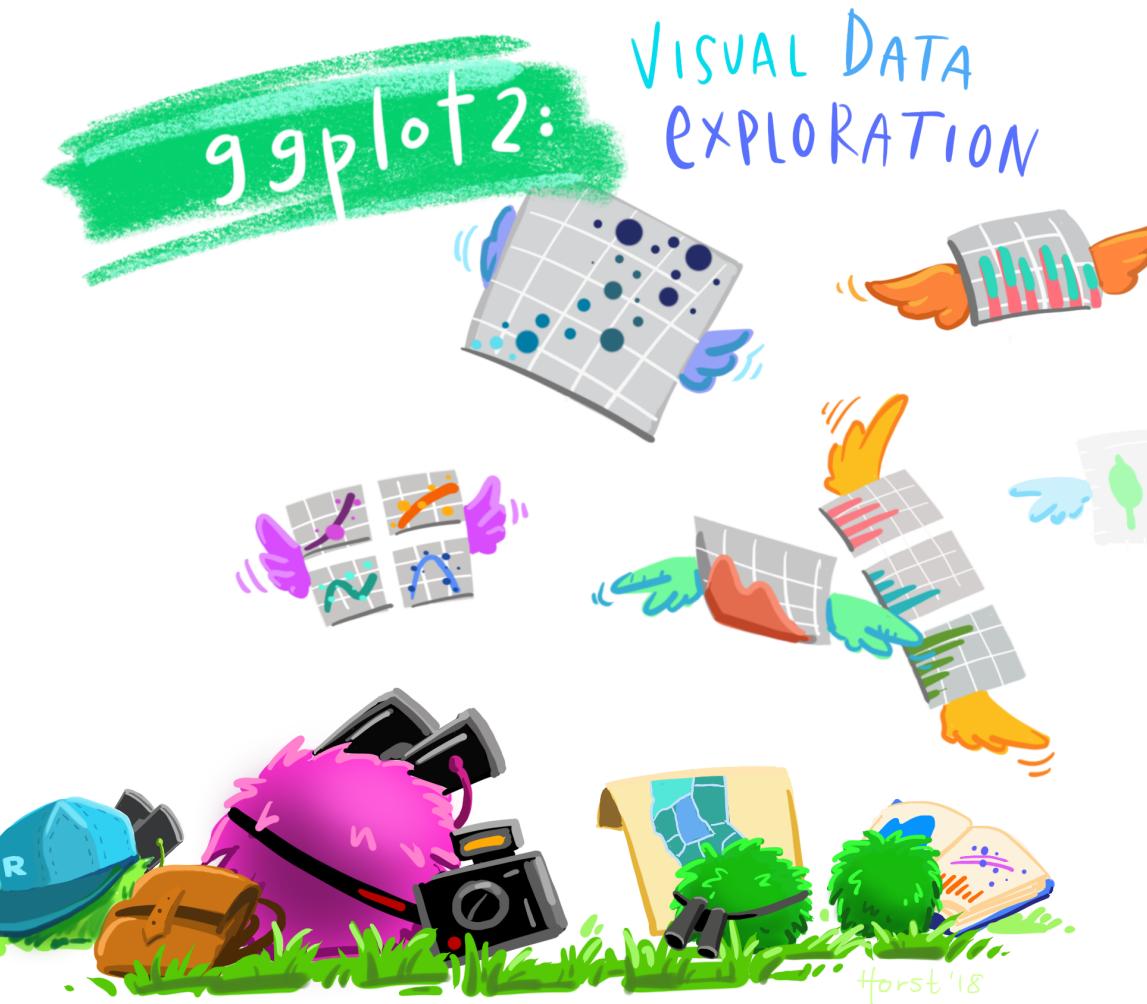


Illustration by Allison Horst ([github.com/allisonhorst/stats-illustrations](https://github.com/allisonhorst/stats-illustrations))