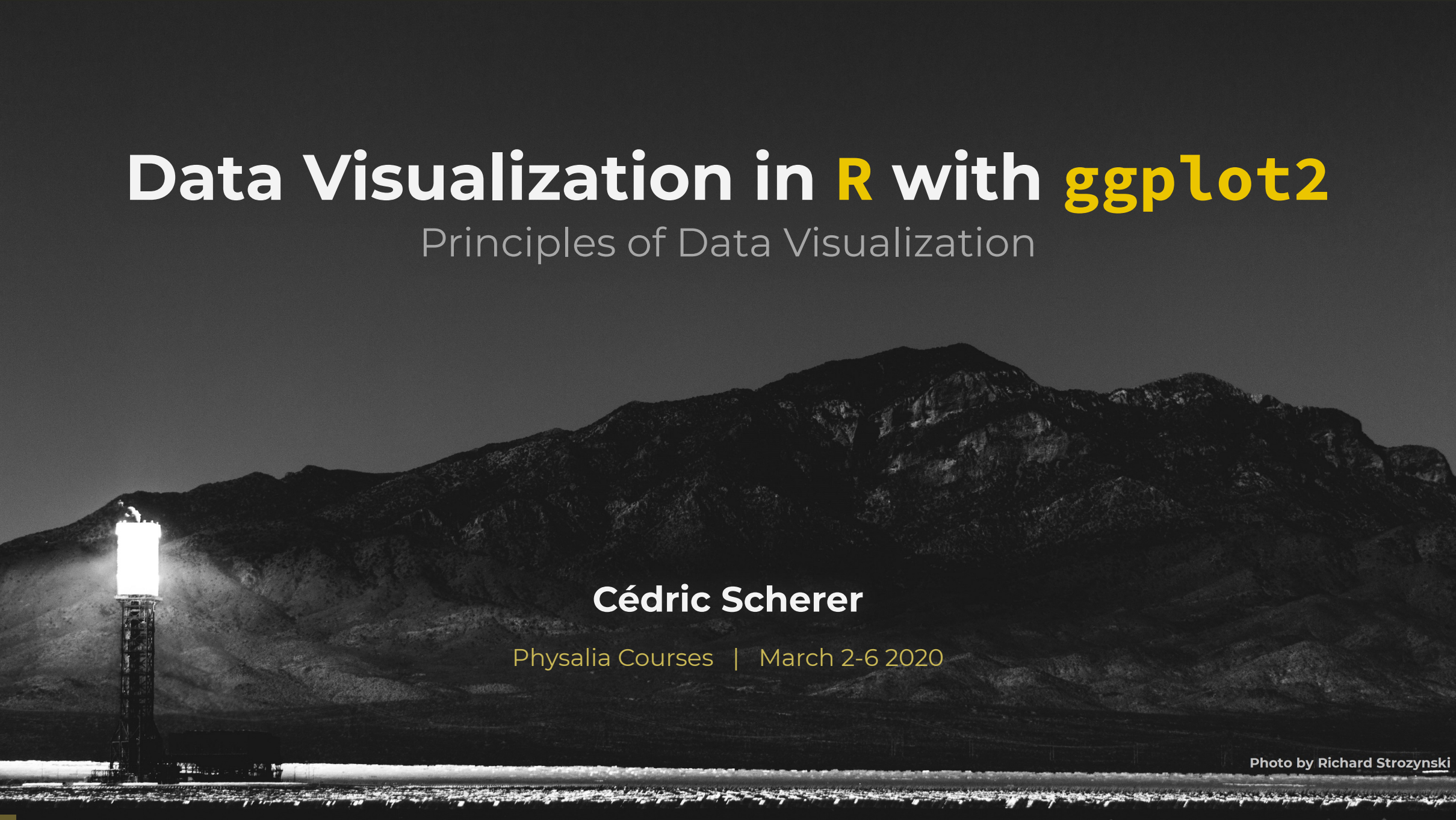
Data Visualization in R with ggplot2

Principles of Data Visualization



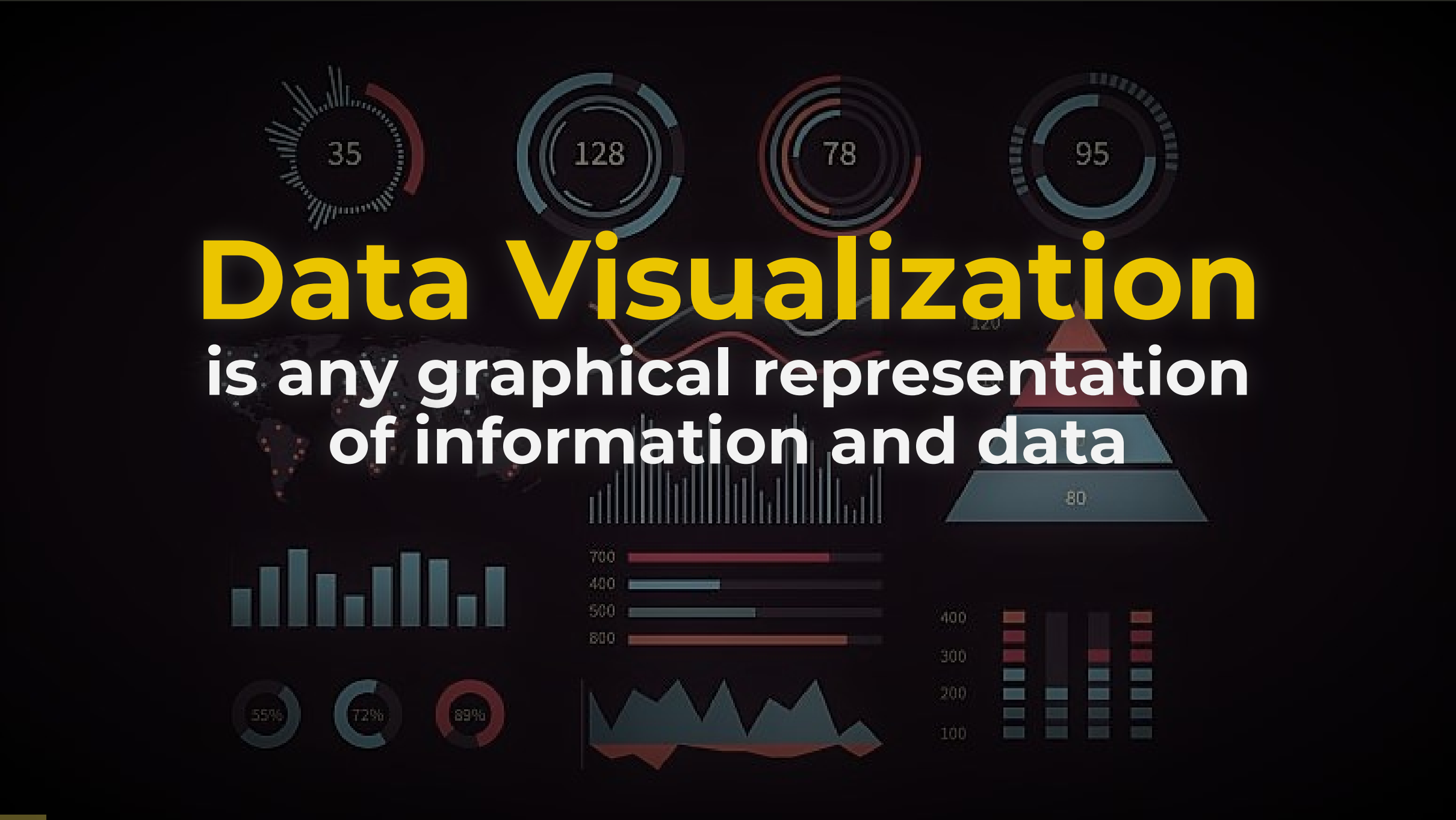
Cédric Scherer

Physalia Courses | March 2-6 2020

Photo by Richard Strozynski

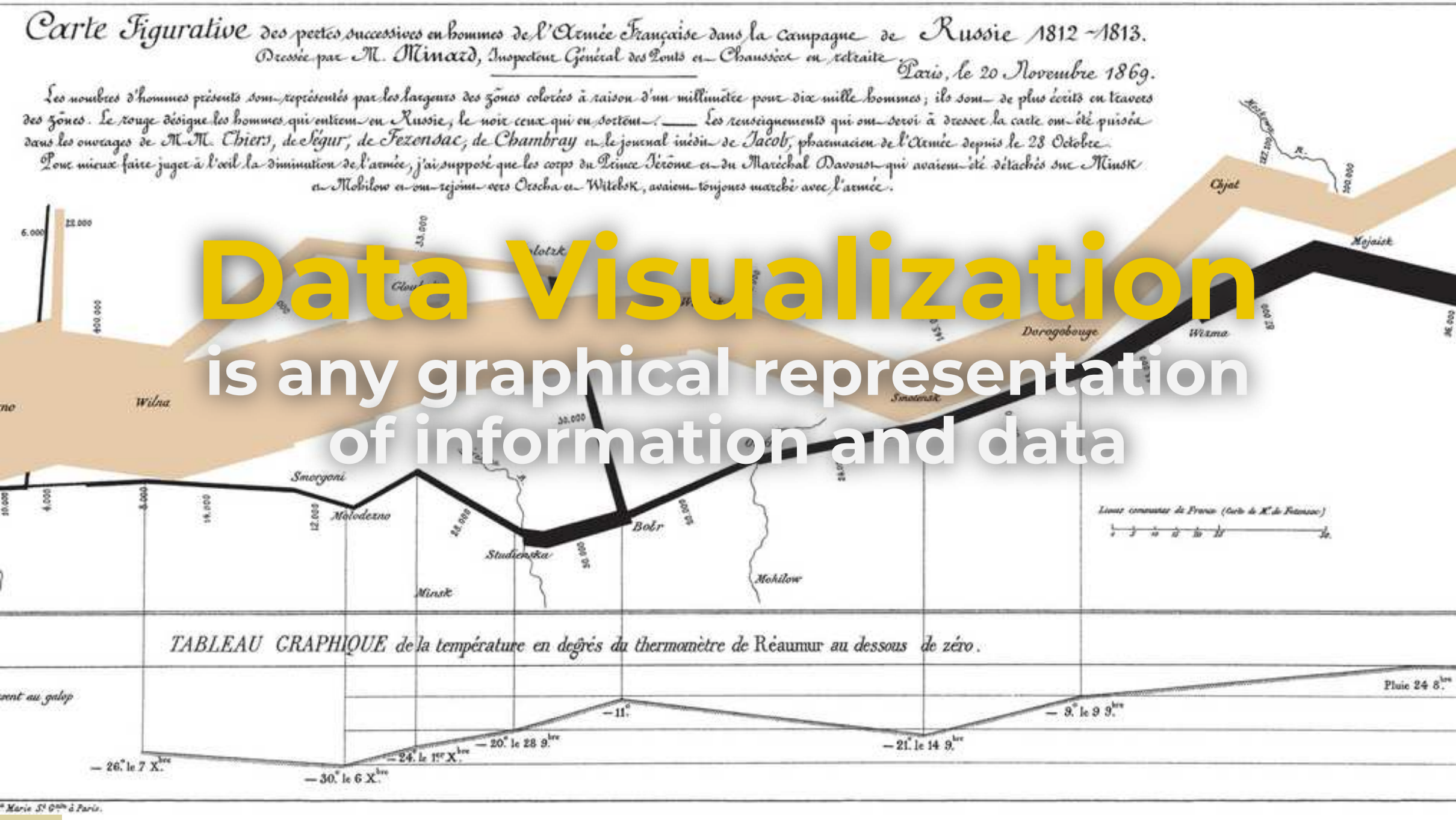
Data Visualization

is any graphical representation of information and data



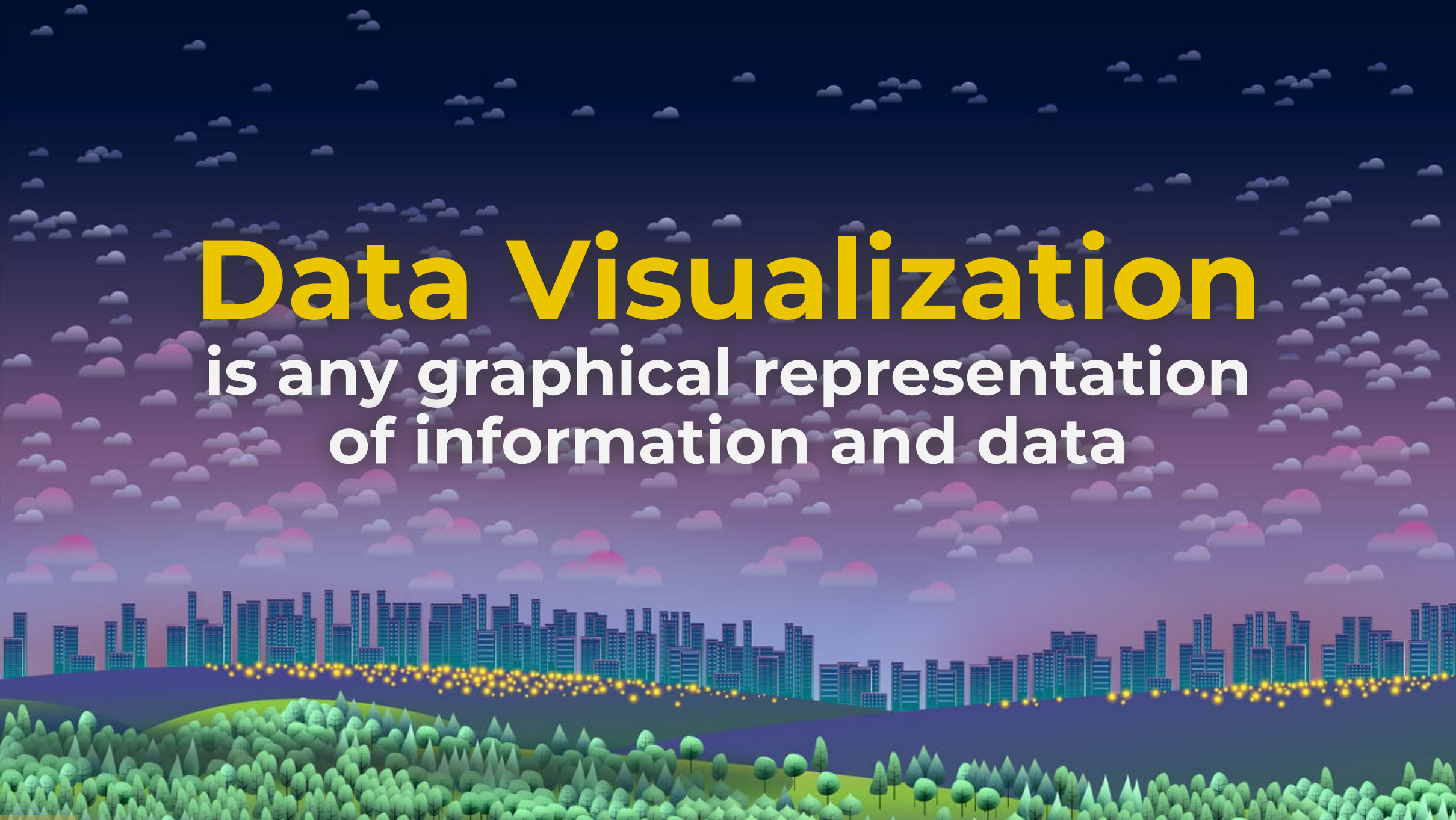
Data Visualization

is any graphical representation of information and data



Data Visualization

is any graphical representation of information and data


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.

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

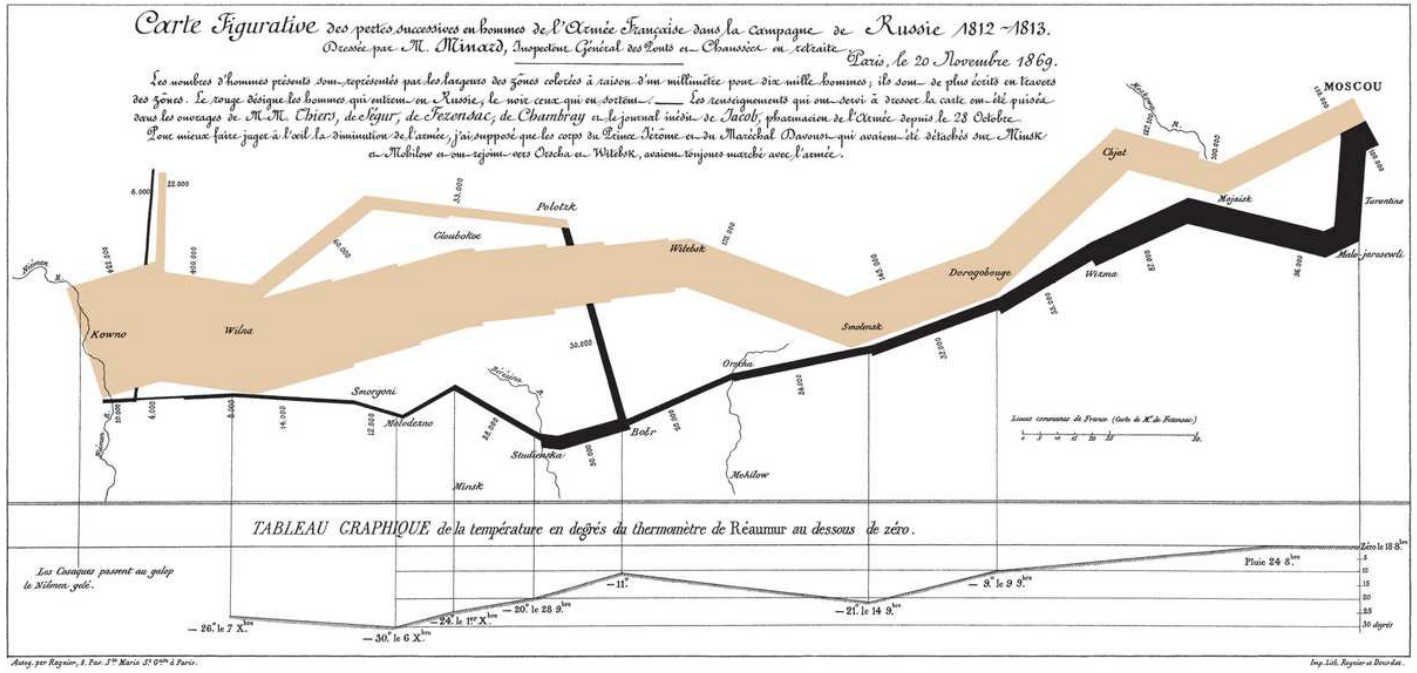
in front of train or metro

drowning

jumping from height

other method*

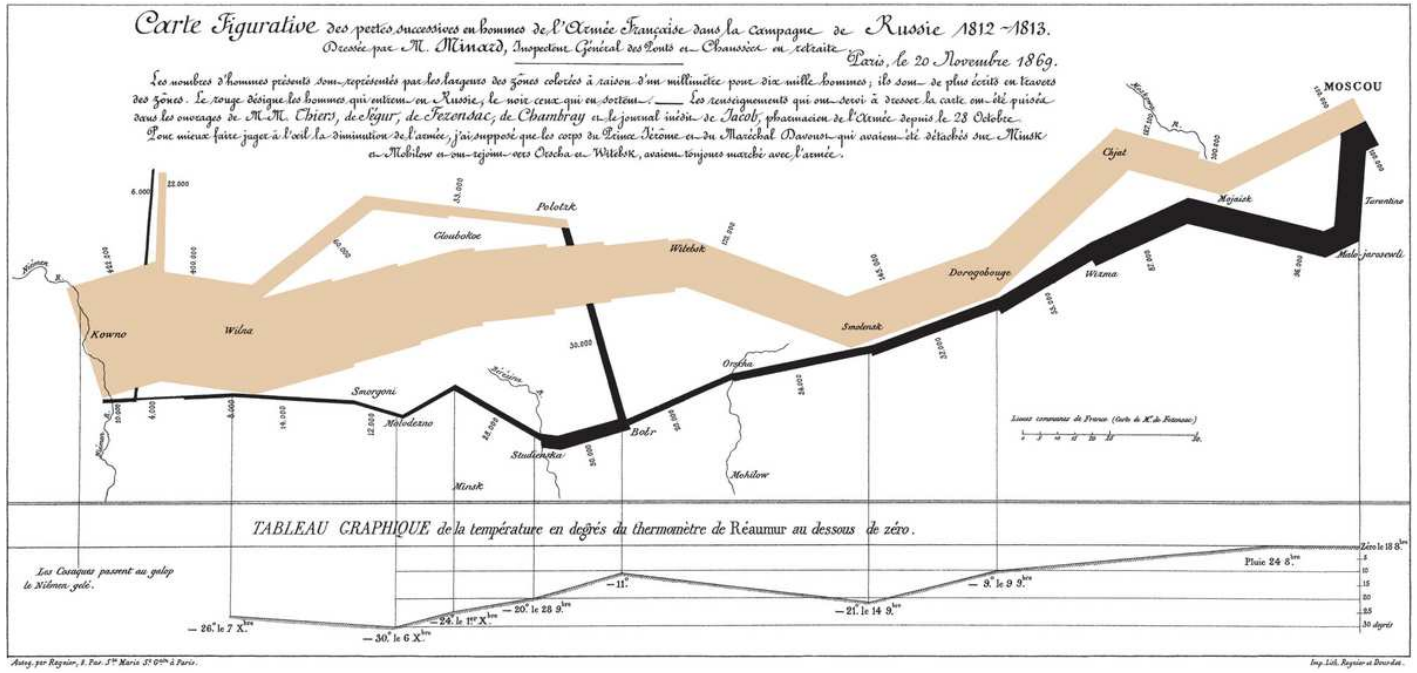
unknown method



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

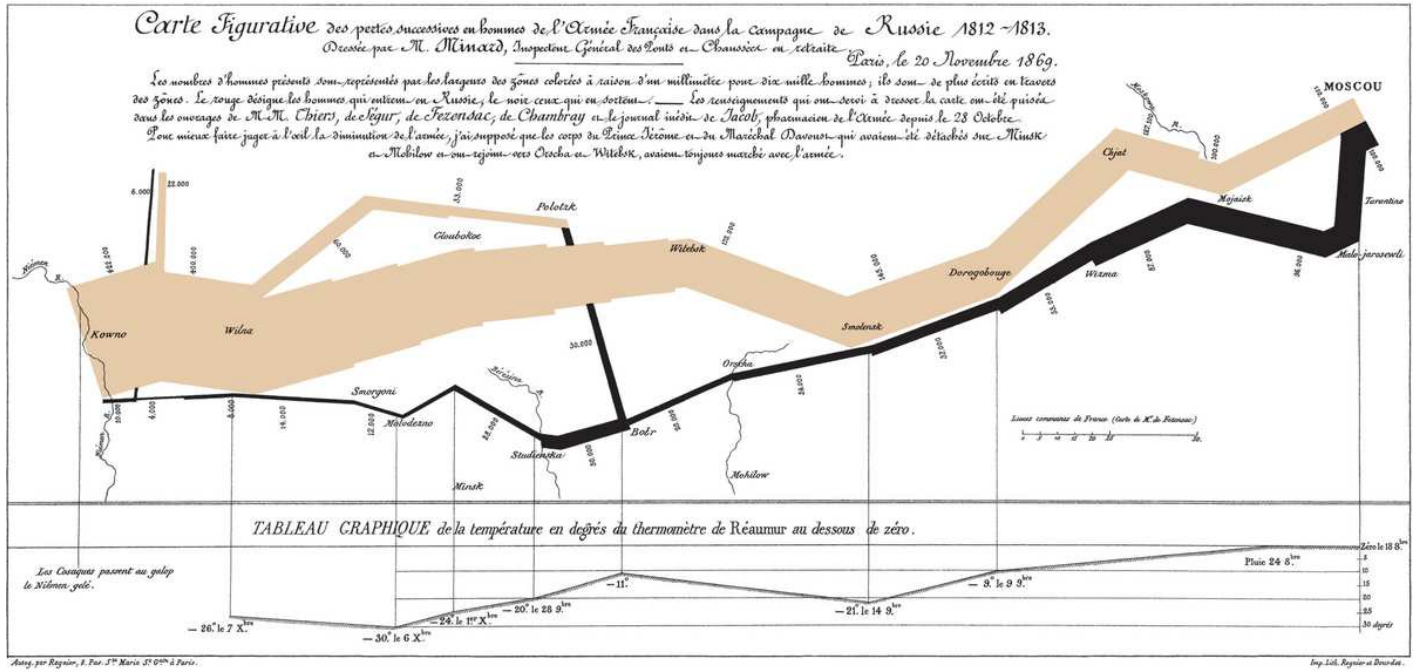
You might be wondering what you are viewing here.

- 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"



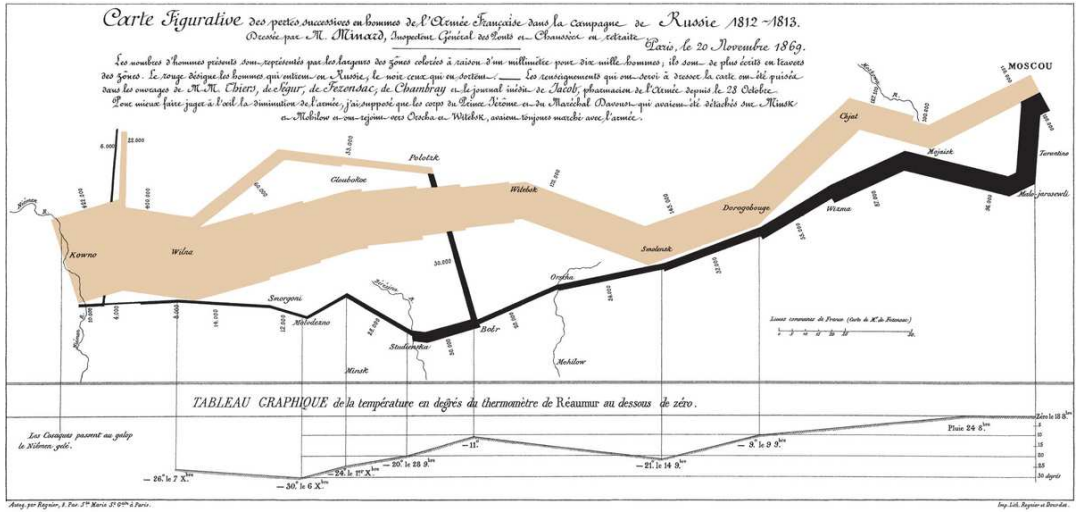
"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

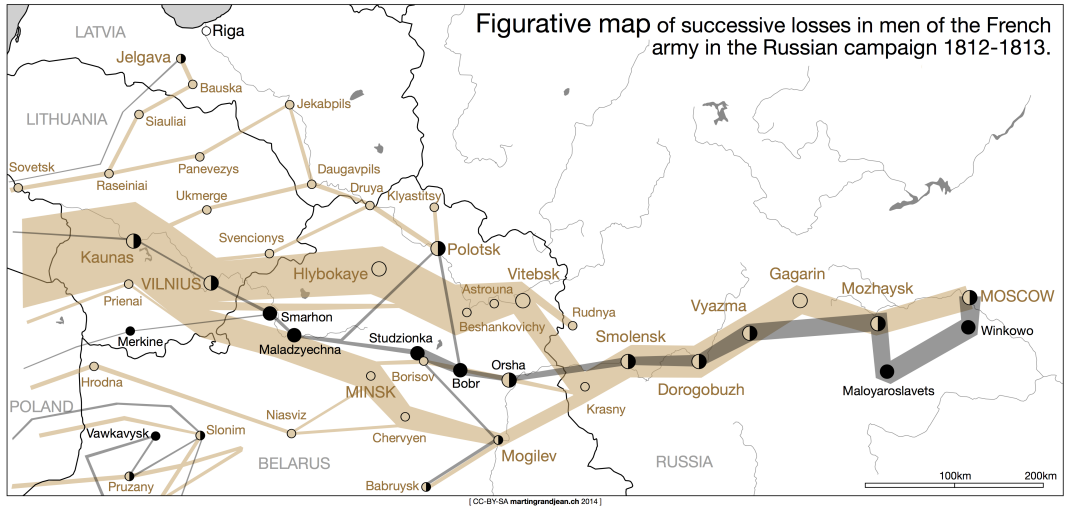
- encodes 6 variables in a simple (?) and modern way:
  - width → size of Napoleon's army
  - x-axis → longitude of the army's position
  - y-axis → latitude of the army's position
  - color → direction of the army's movement
  - line chart → temperature during the army's retreat
  - annotations → locations and army size (main chart) + date along retreat path (line chart)



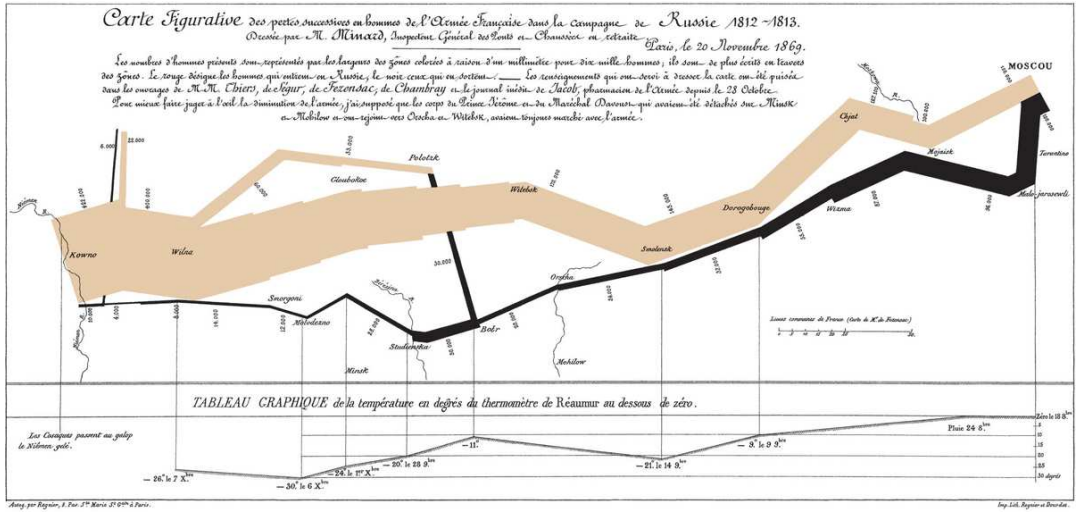
"Figurative Map of the Successive Losses in Men of the French Army in the Russian Campaign 1812–1813" by Charles Joseph Minard

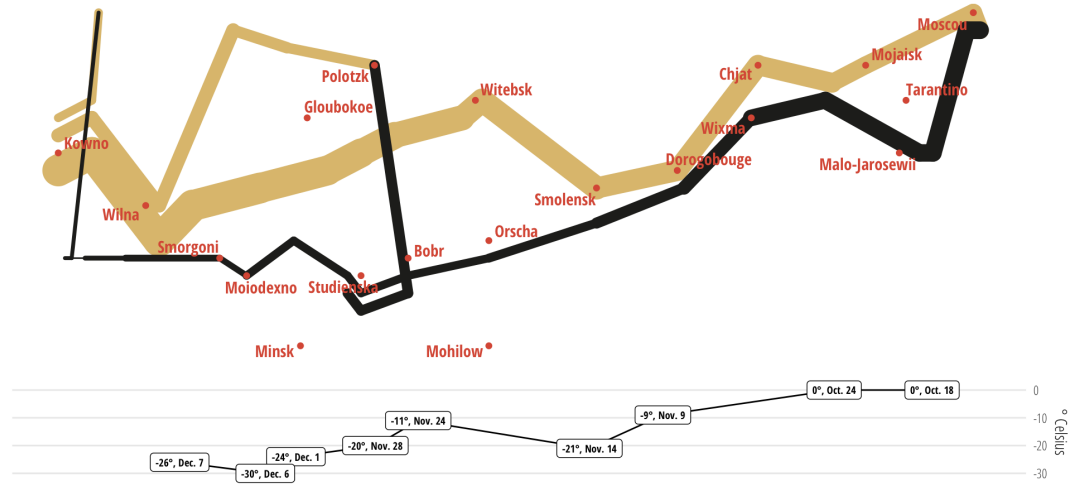
- generally considered as the first data visualization (1869)
- Edward Tufte calls Minard's graphic of Napoleon in Russia one of the "best statistical drawings ever created".





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`

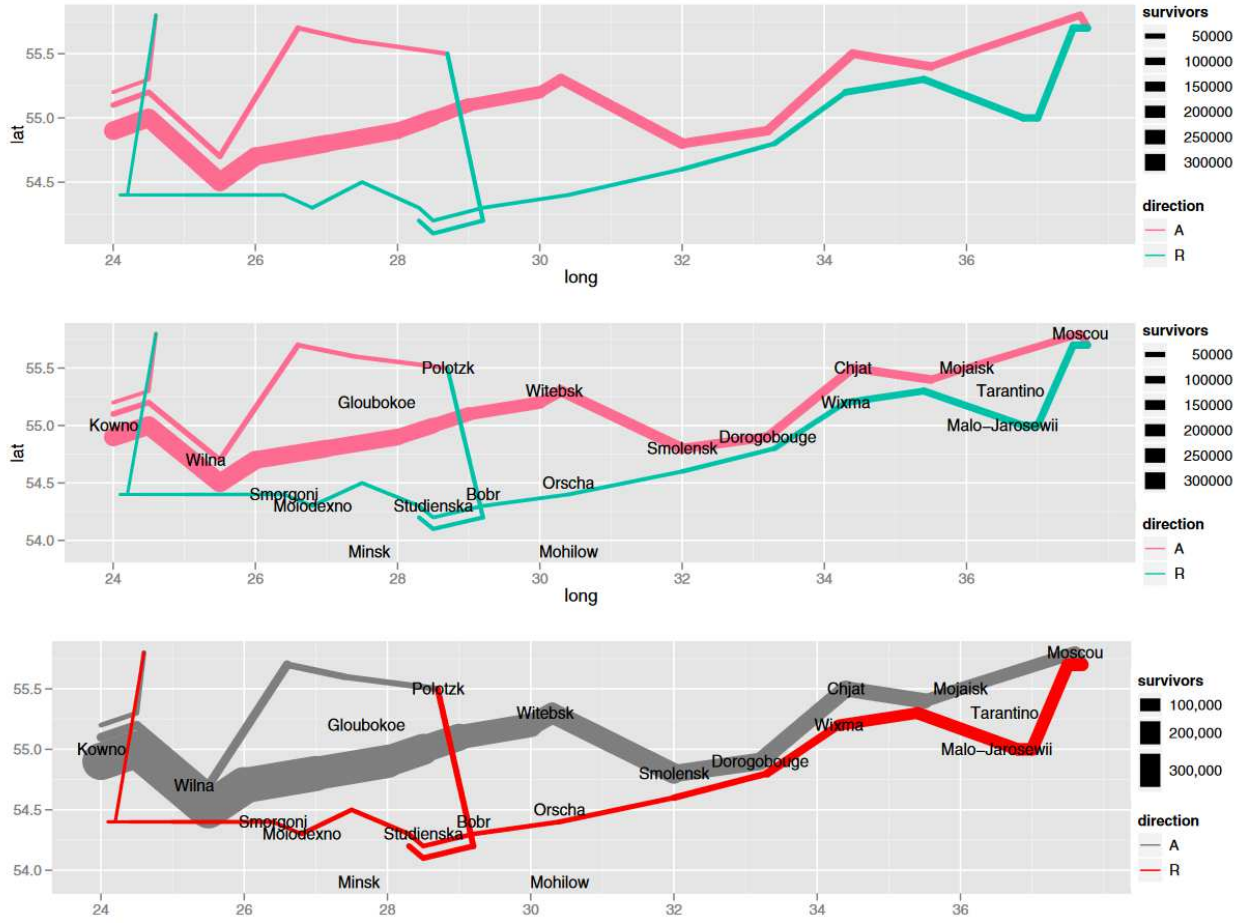


Figure 12. Iteratively reproducing the depiction of Napoleon's March by Minard. (Top) Displaying the key troop movement data. (Center) Adding town locations as reference points. (Bottom) Tweaking scales to produce polished plot.

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

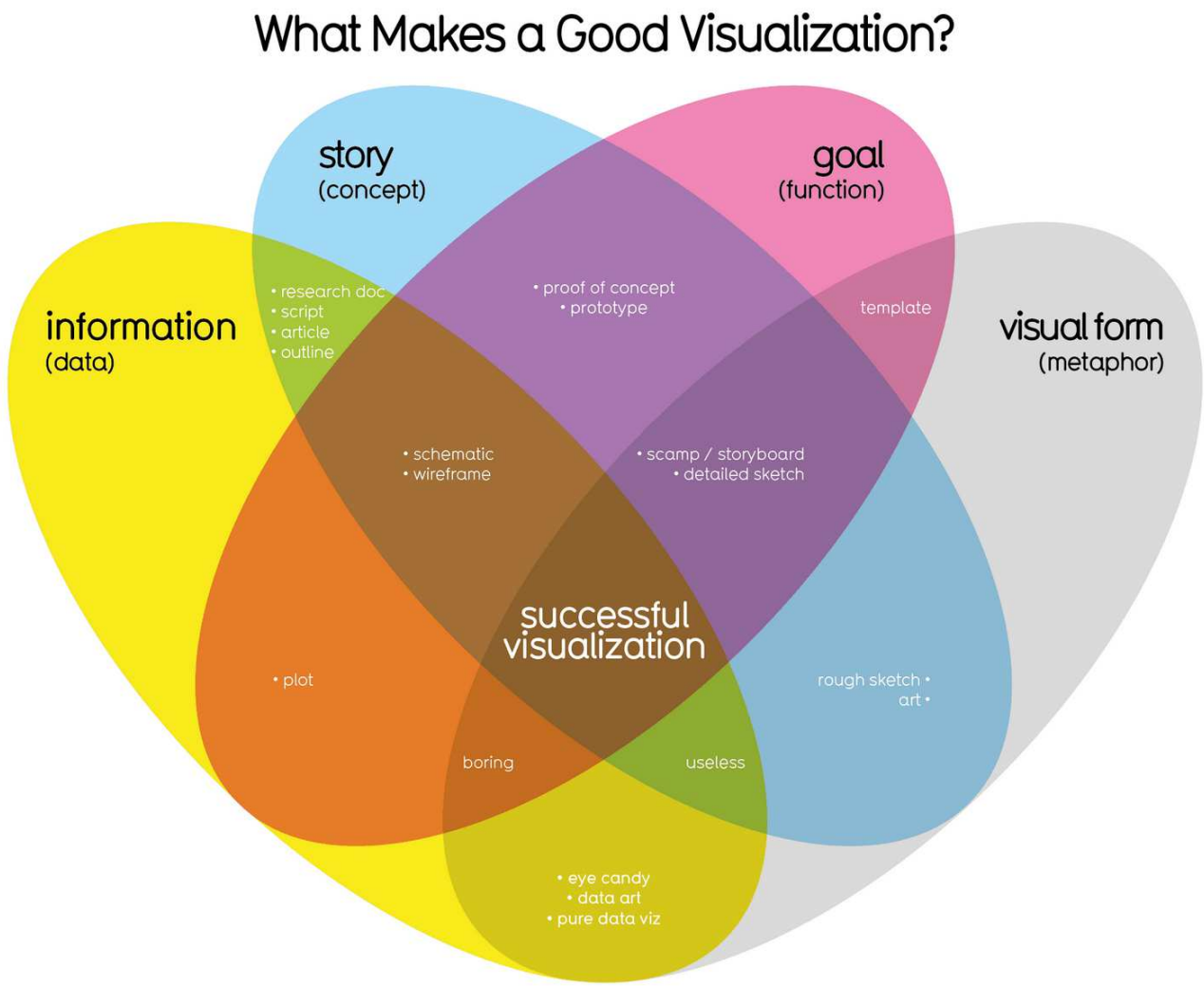
→ Information Understand your data and be accurate

→ Story Be clear about the story of your visualization

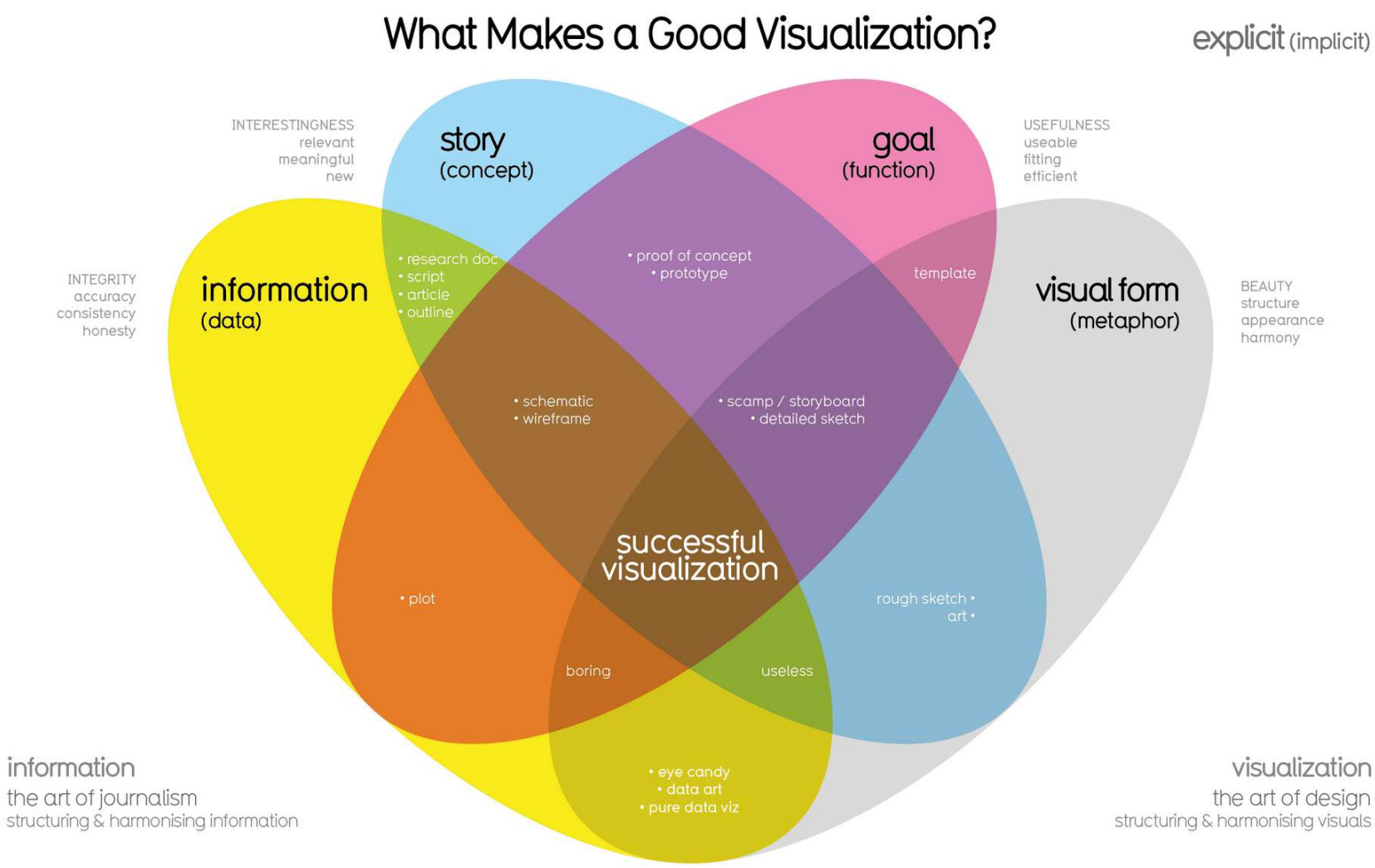
→ Goal Select charts that successfully transport your story

→ Visual Form Follow design rules and data visualization principles

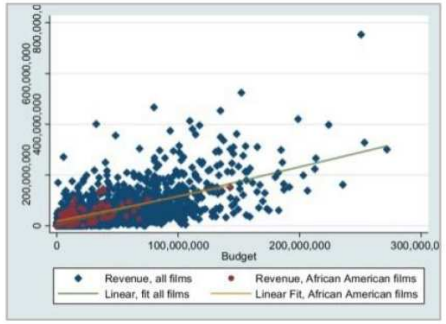
+ inspiration, training and (a bit of) talent

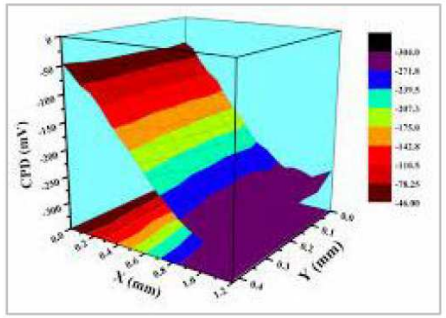


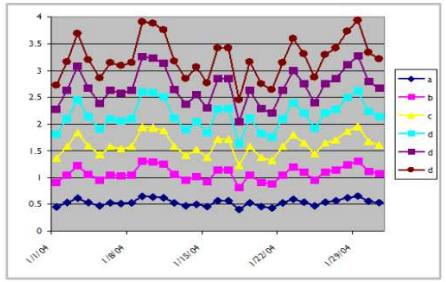
Visualization by David McCandless (Information is Beautiful)



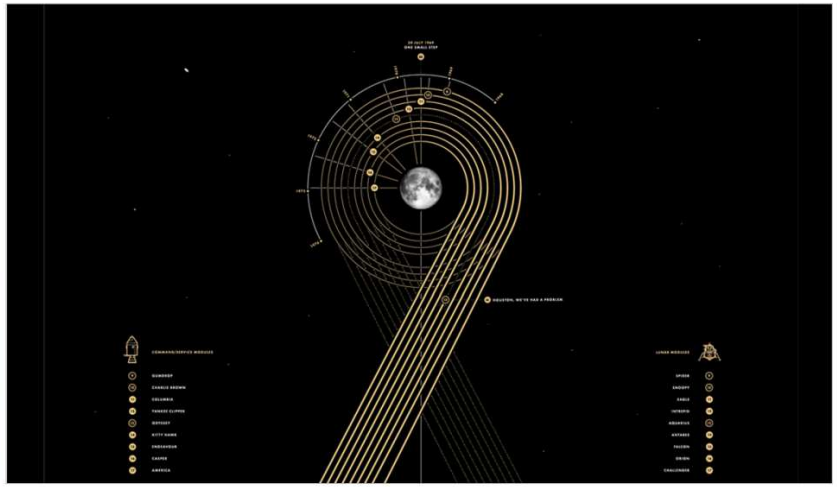
Visualization by David McCandless (Information is Beautiful)

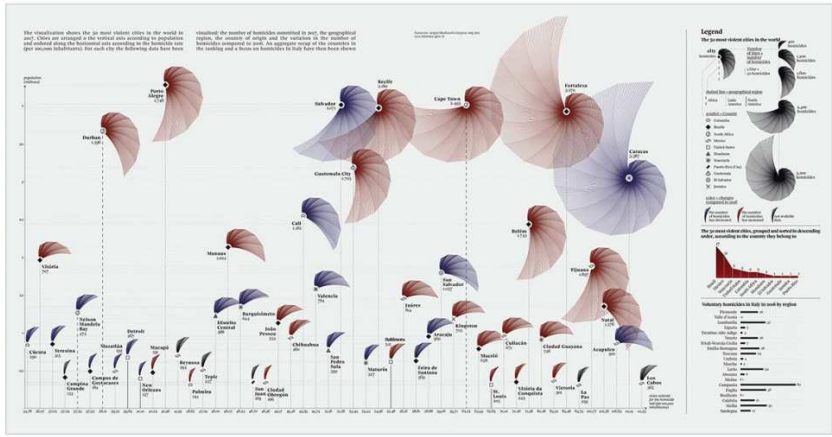






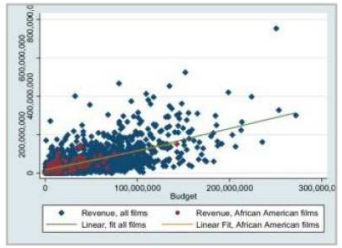
Anonymous

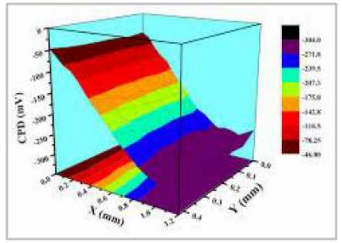


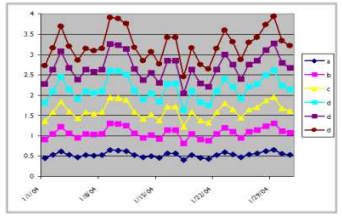


Sonia Kuijpers

Upper: Paul Button
Lower: Frederica Fragapane







Anonymous

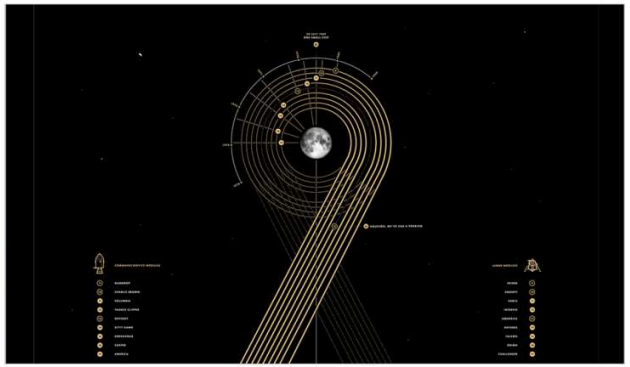
We aim for DataViz that:

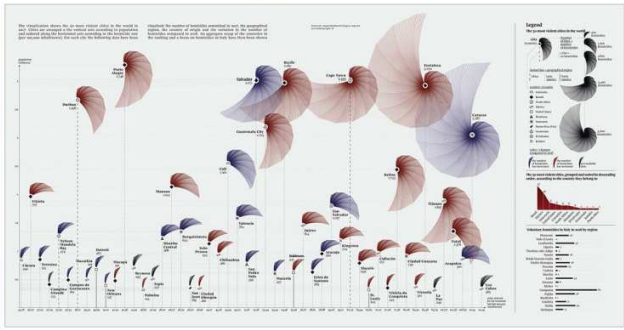
- is informative
- reduces complexity
- is easy to grasp
- is visually appealing
- draws attention

but:

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

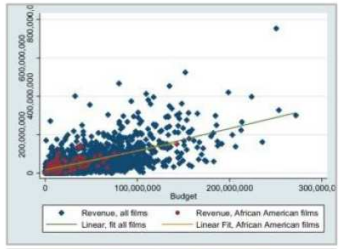
Sonia Kuijpers

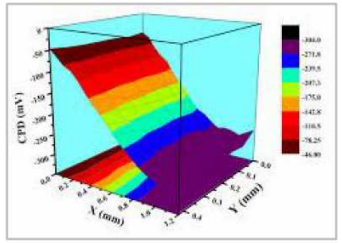


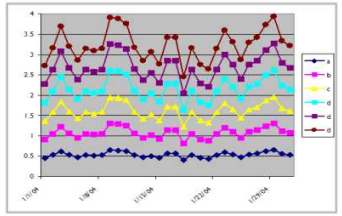


Upper: Paul Button
Lower: Frederica Fragapane

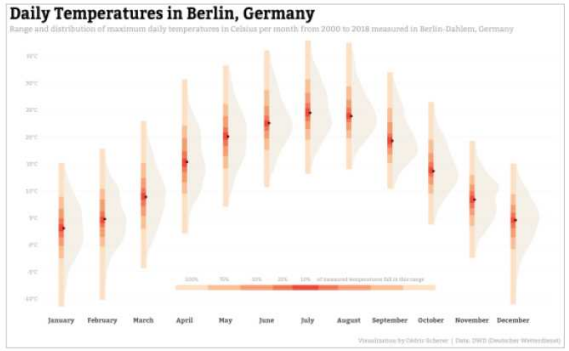
Gradient from poorly designed & uninformative data visualization to data art

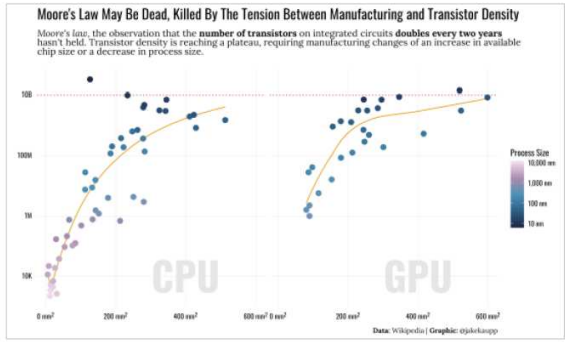






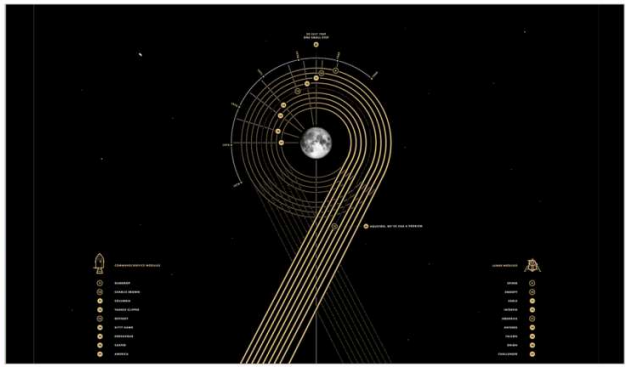
Anonymous

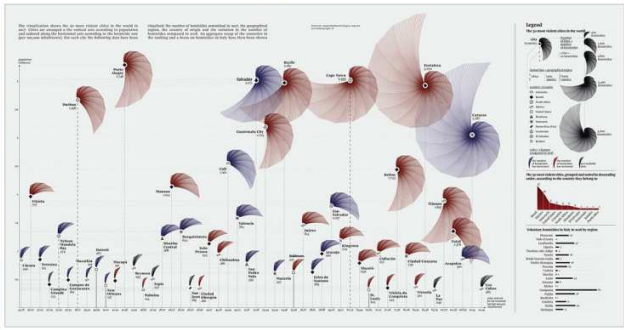




Upper: Cédric Scherer
Lower: Jake Kaupp

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 + continiuous + 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 + continiuous + 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 + continiuous + 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 + continiuous + ordered
  - 1 → quantitative + continuous + ordered
    or: quantitative + discrete + ordered
    or: qualitative + discrete + ordered
    or: qualitative + discrete + unordered

Illustration by Allison Horst (github.com/allisonhorst/stats-illustrations)

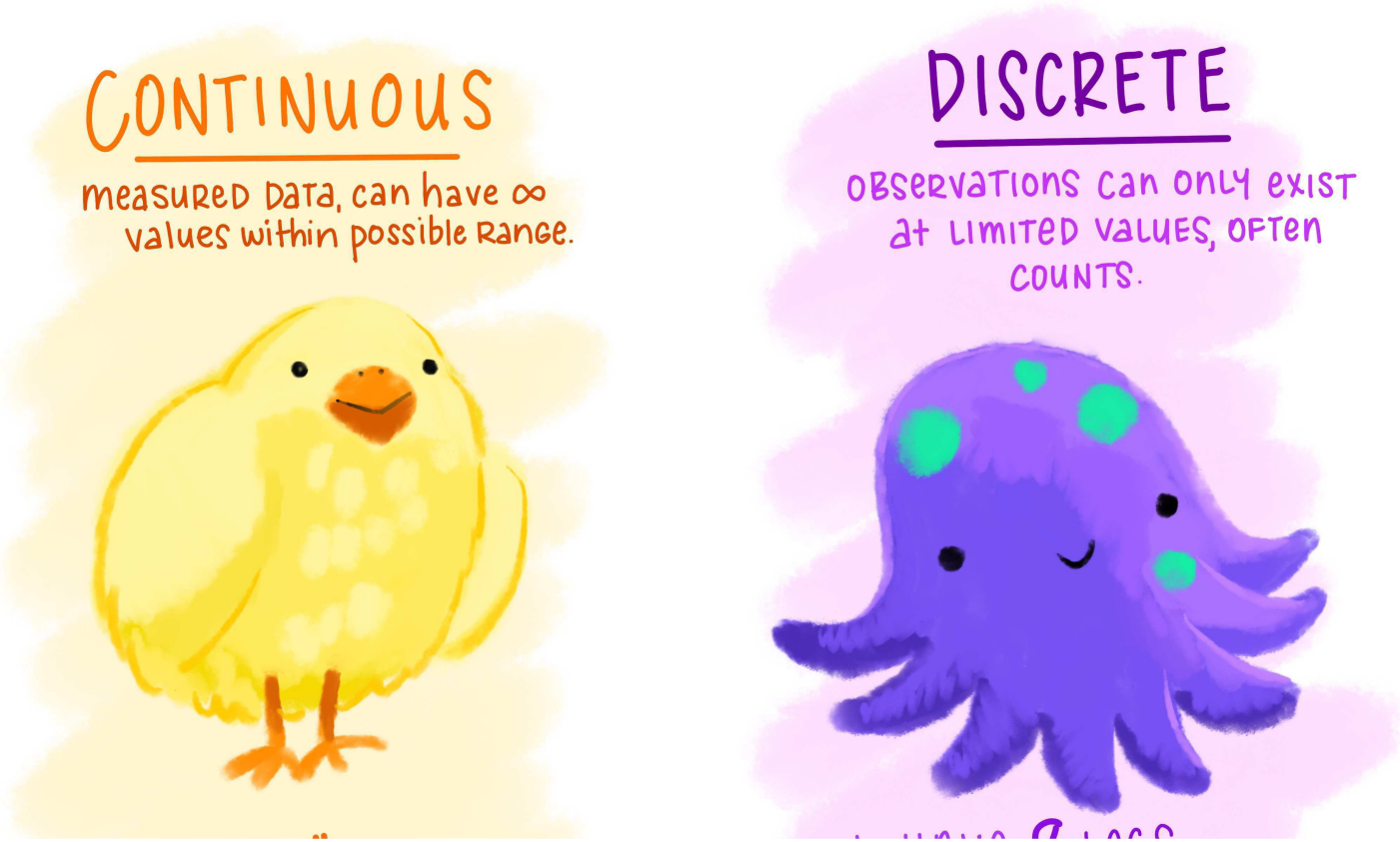
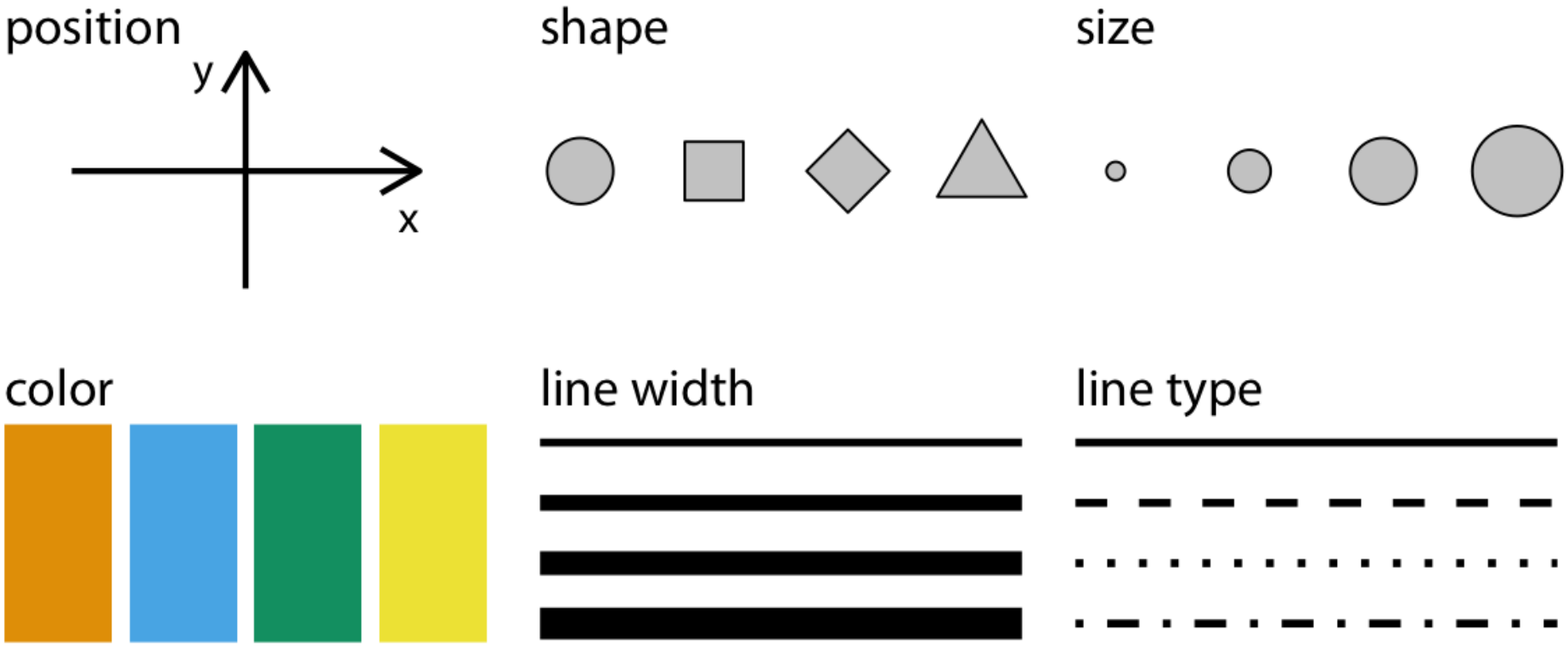


Illustration by Allison Horst (github.com/allisonhorst/stats-illustrations)

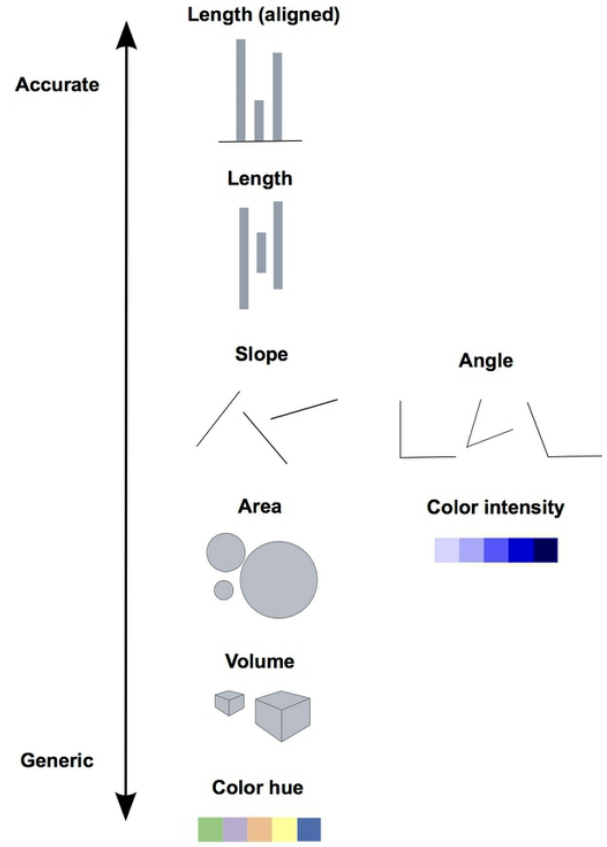
Mapping Data to Aesthetics

Data visualizations map values into quantifiable features (aesthetics)



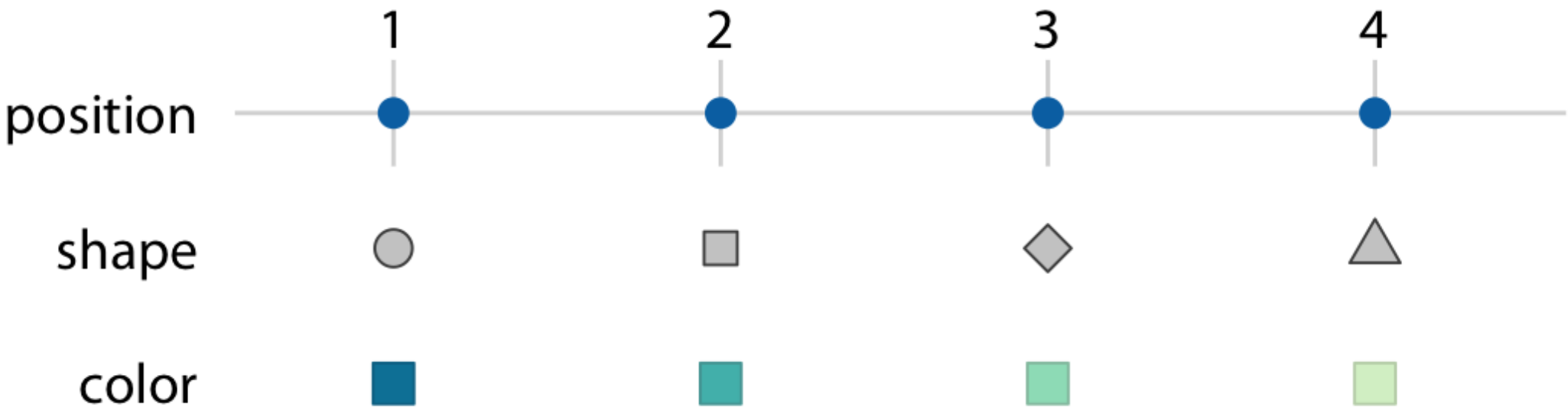
Source: "Fundamentals of Data Visualization" by Claus Wilke

Data visualizations map values into quantifiable features (aesthetics)



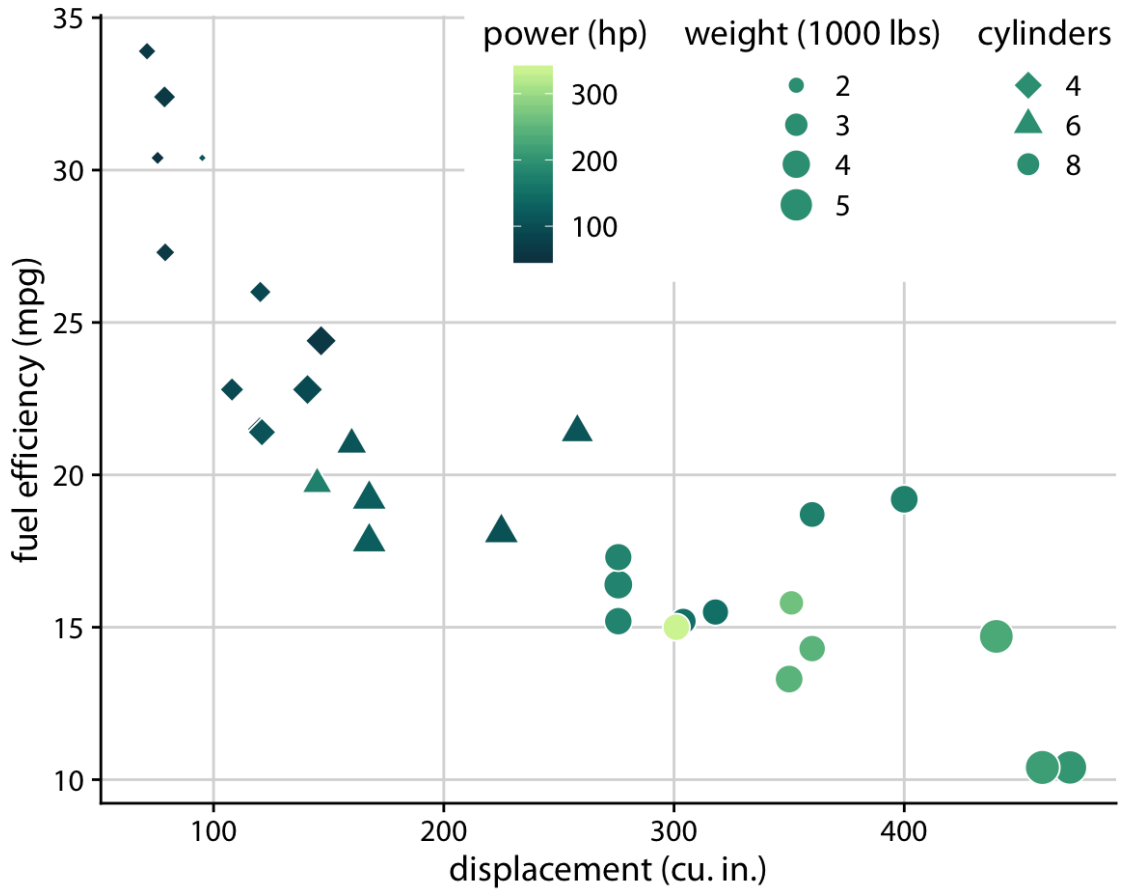
Source: Peter Aldhous based on experiments by William Cleveland and Robert McGill

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

visualcomposer.com/blog/ultimate-guide-to-colors-and-color-palettes

Color Palette Types

Categorical

default_uchicago

Sequential: Single-Hue

cyan_material

Sequential: Multi-Hue

Sunset

Diverging

Geyser

Cyclical

ocean.phase

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

Rainbow Color Map (Still) Considered Harmful

Publisher: IEEE

2 Author(s) David Borland ; Russell M. Taylor Ii **View All Authors**

172	3	9091
Paper Citations	Patent Citations	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
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[a)]

10 March 2017

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

Frédérique Chesterman; Hannah Manssens; Céline Morel; Guillaume Serrell; Bastian Piepers; Tom Kimpe

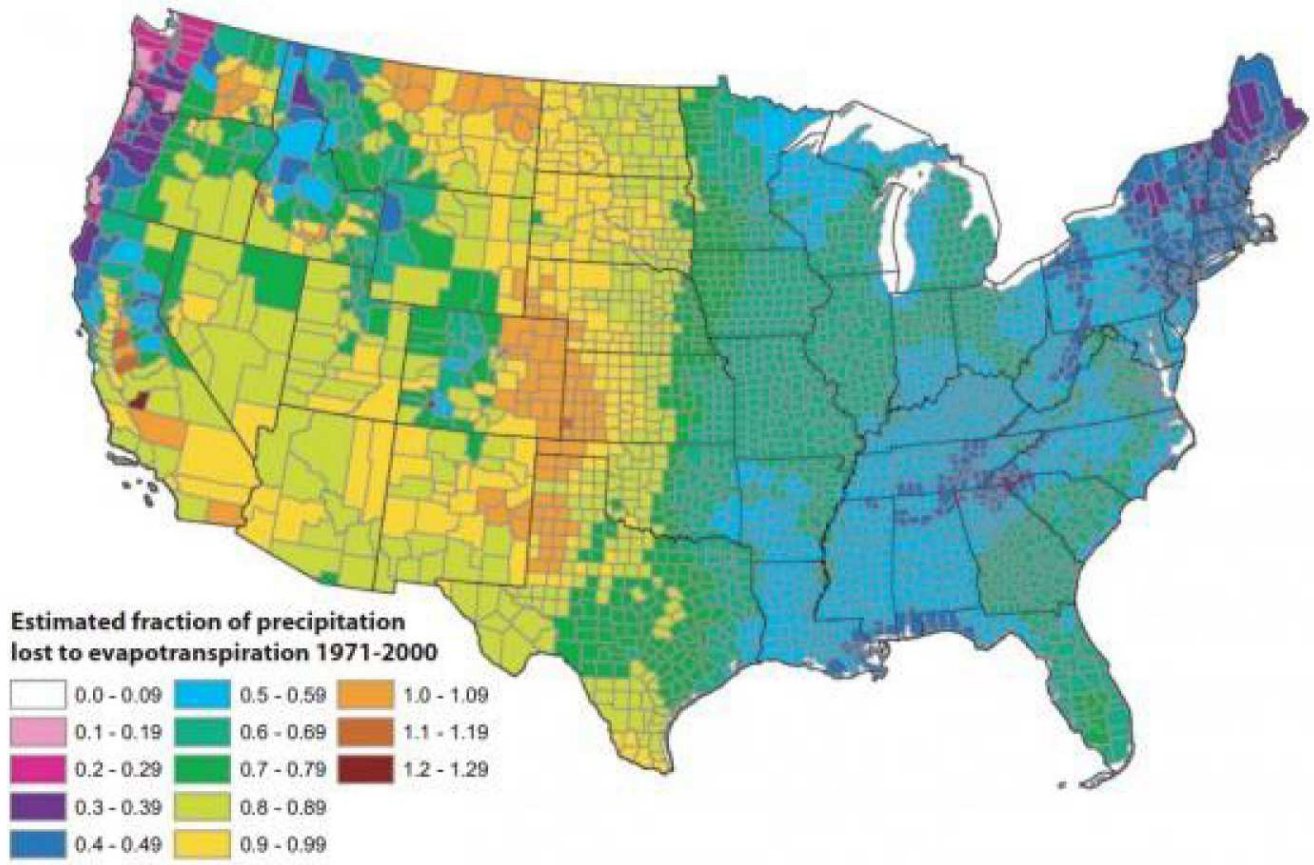


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM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

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

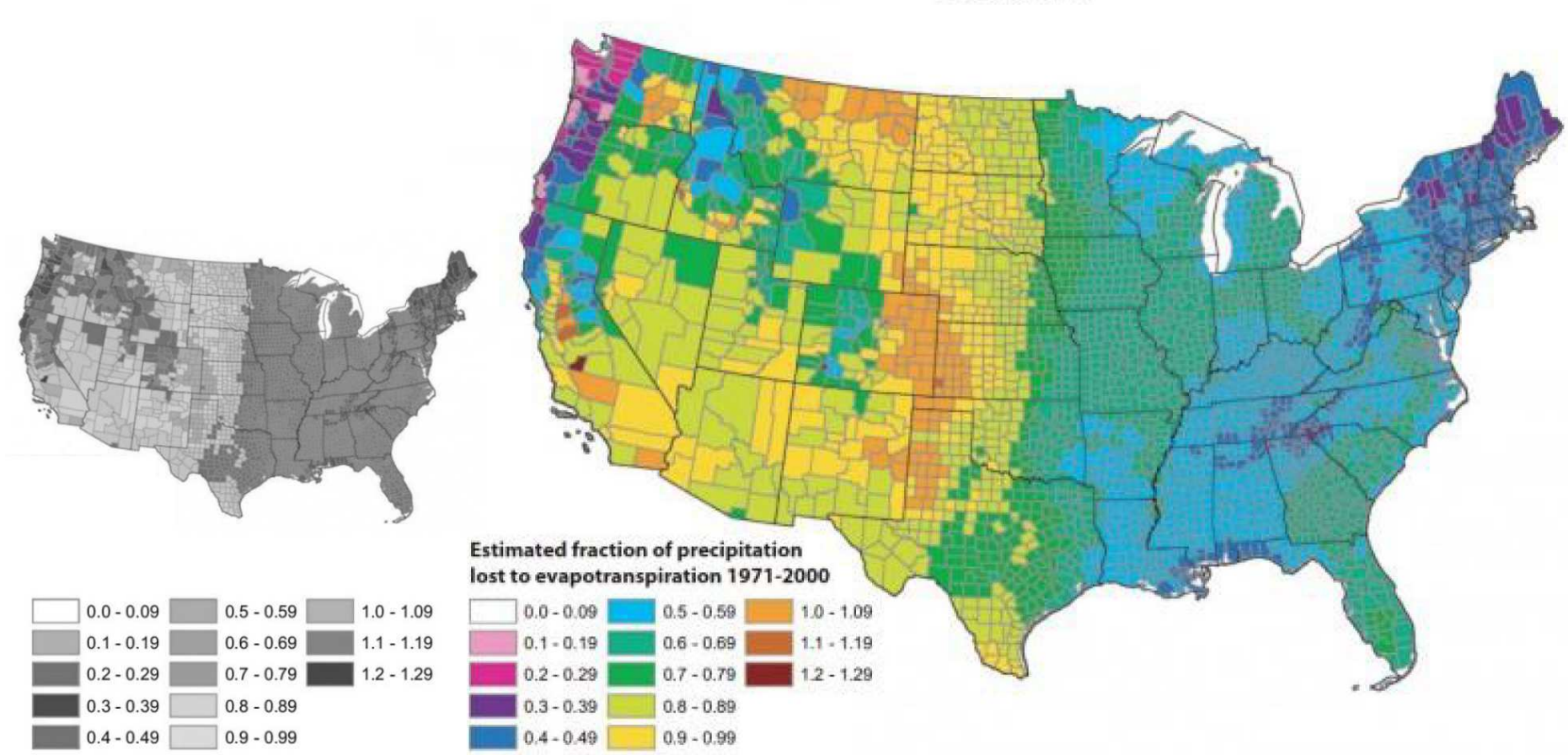
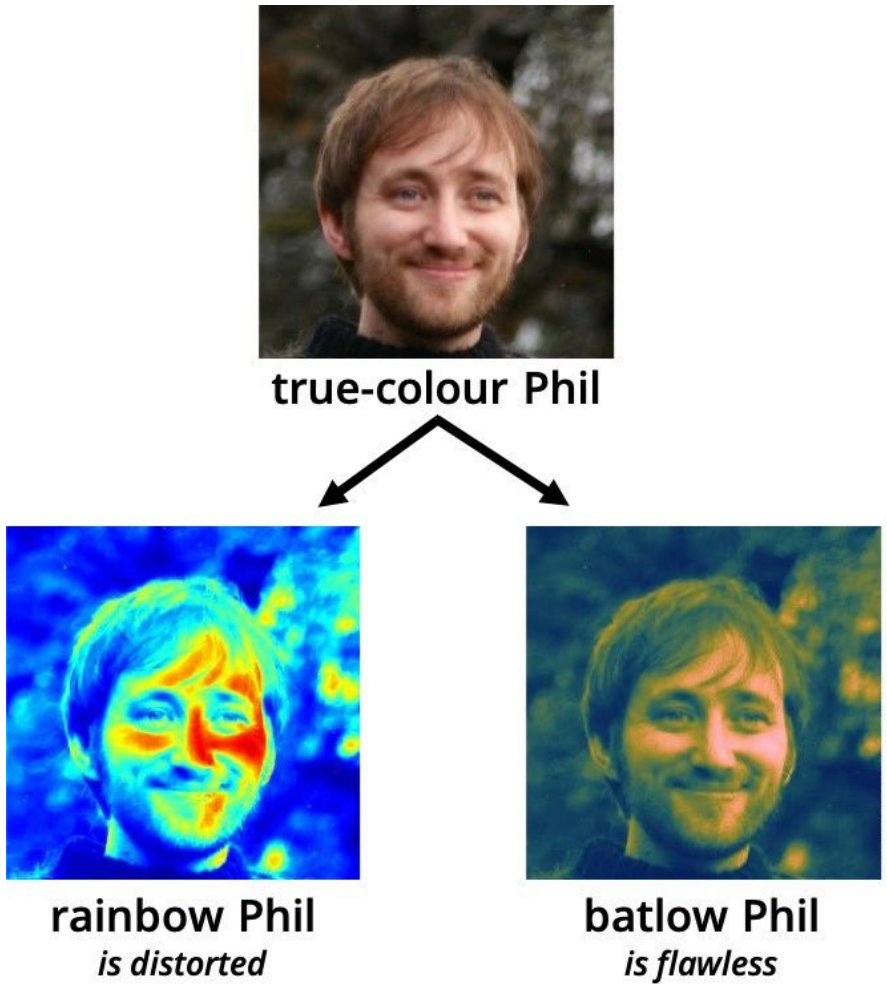


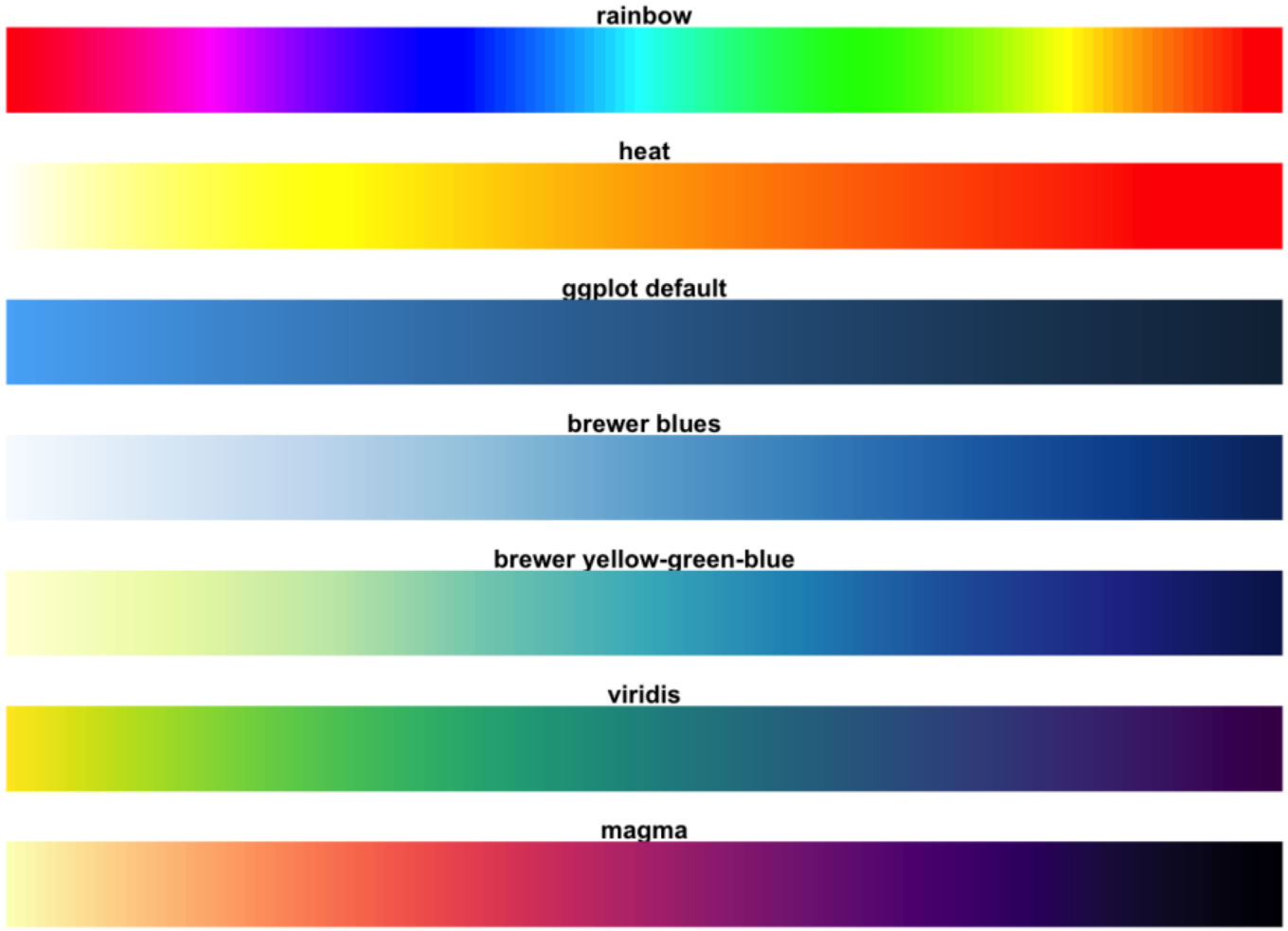
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM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

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



Source: fabiocrameri.ch/batlow

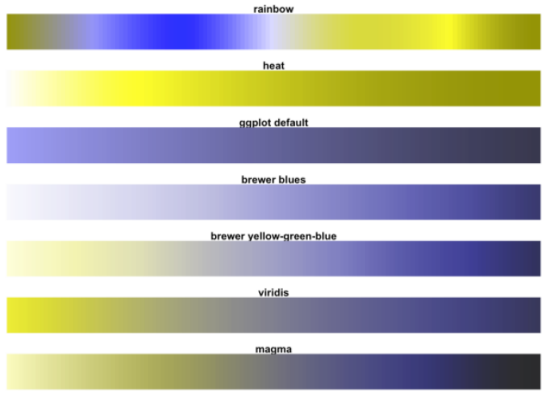
Choice of Color Palettes



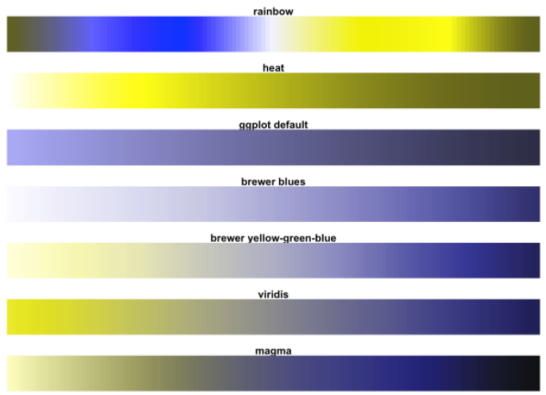
Source: cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html

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

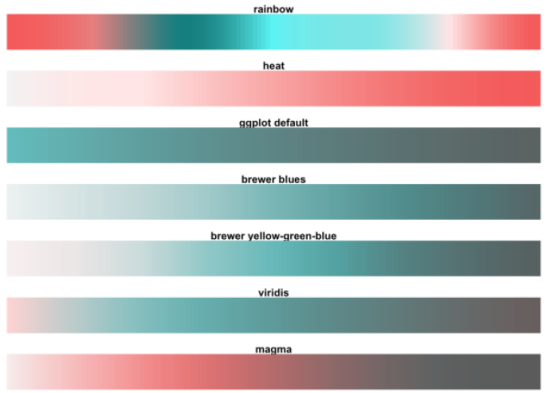
Deutranopia: present in 6% of males



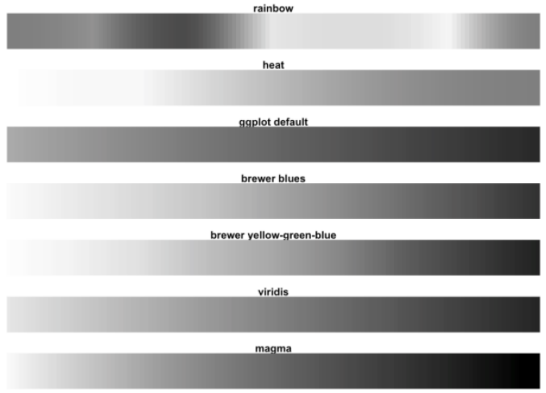
Protanopia: present in 1% of males



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

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!

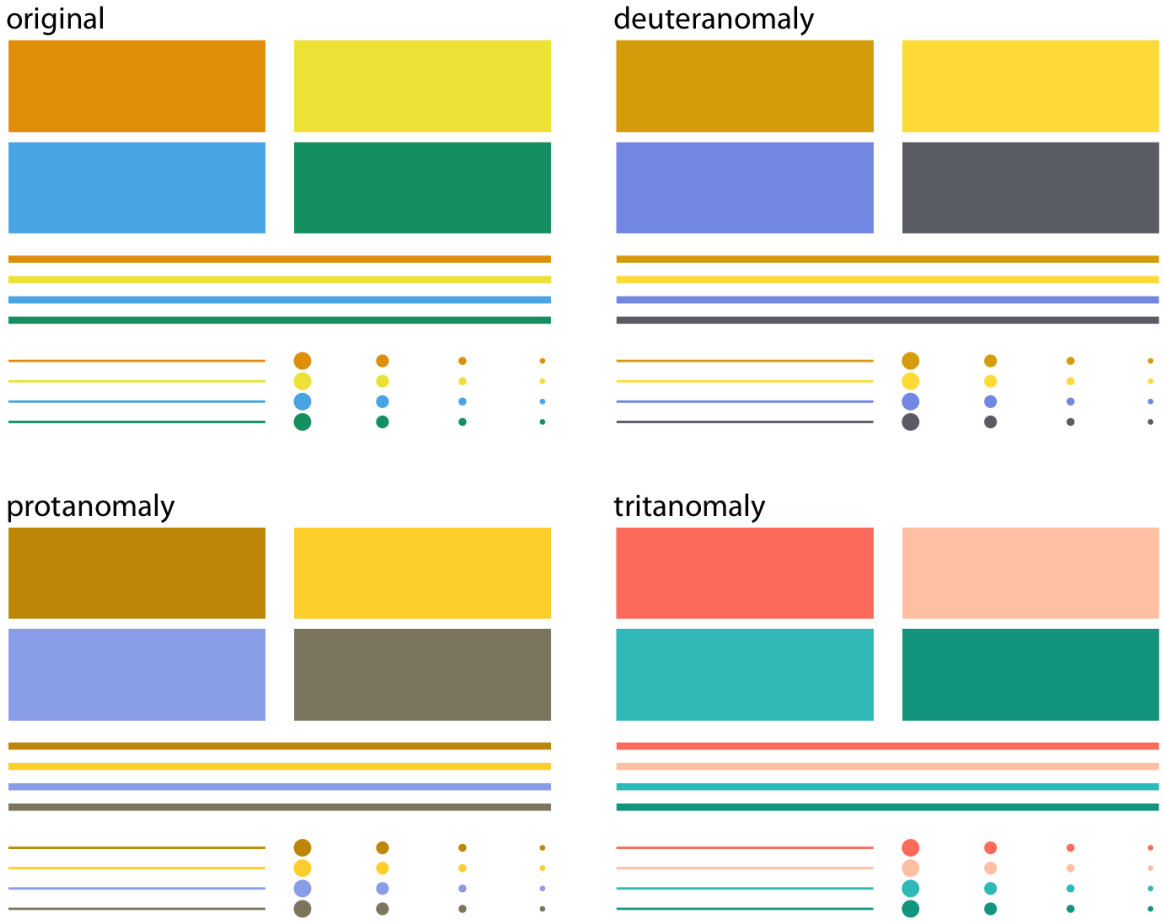
Choice of the Color Palette & Accessibility

Choose color-blind friendly palettes:
projects.susielu.com/viz-palette

Test your final visualization:
color-blindness.com/coblis-color-blindness-simulator

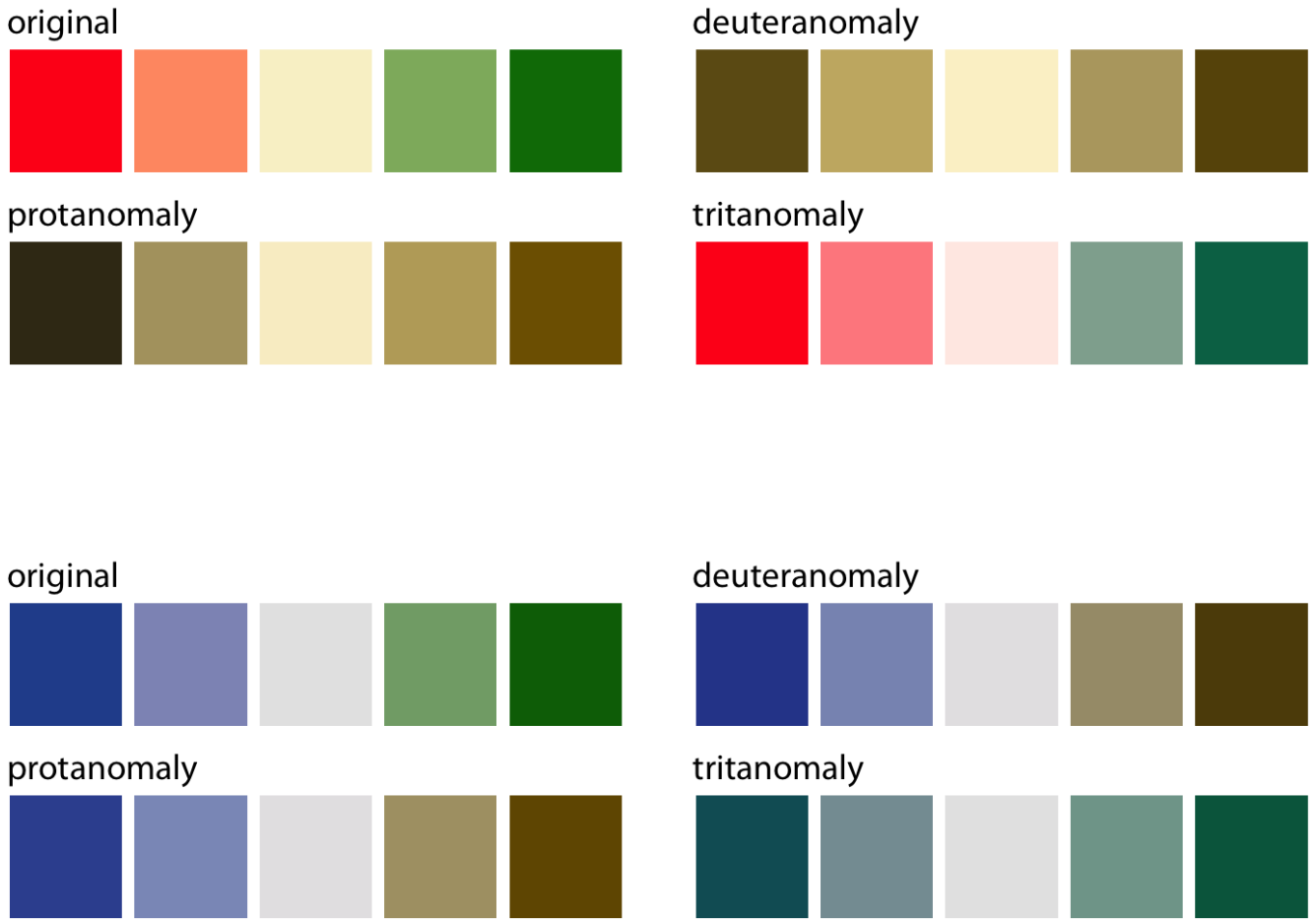
Create a CVD-version of your ggplot in R:
github.com/clauswilke/colorblindr

Choice of the Color Palette & Accessibility



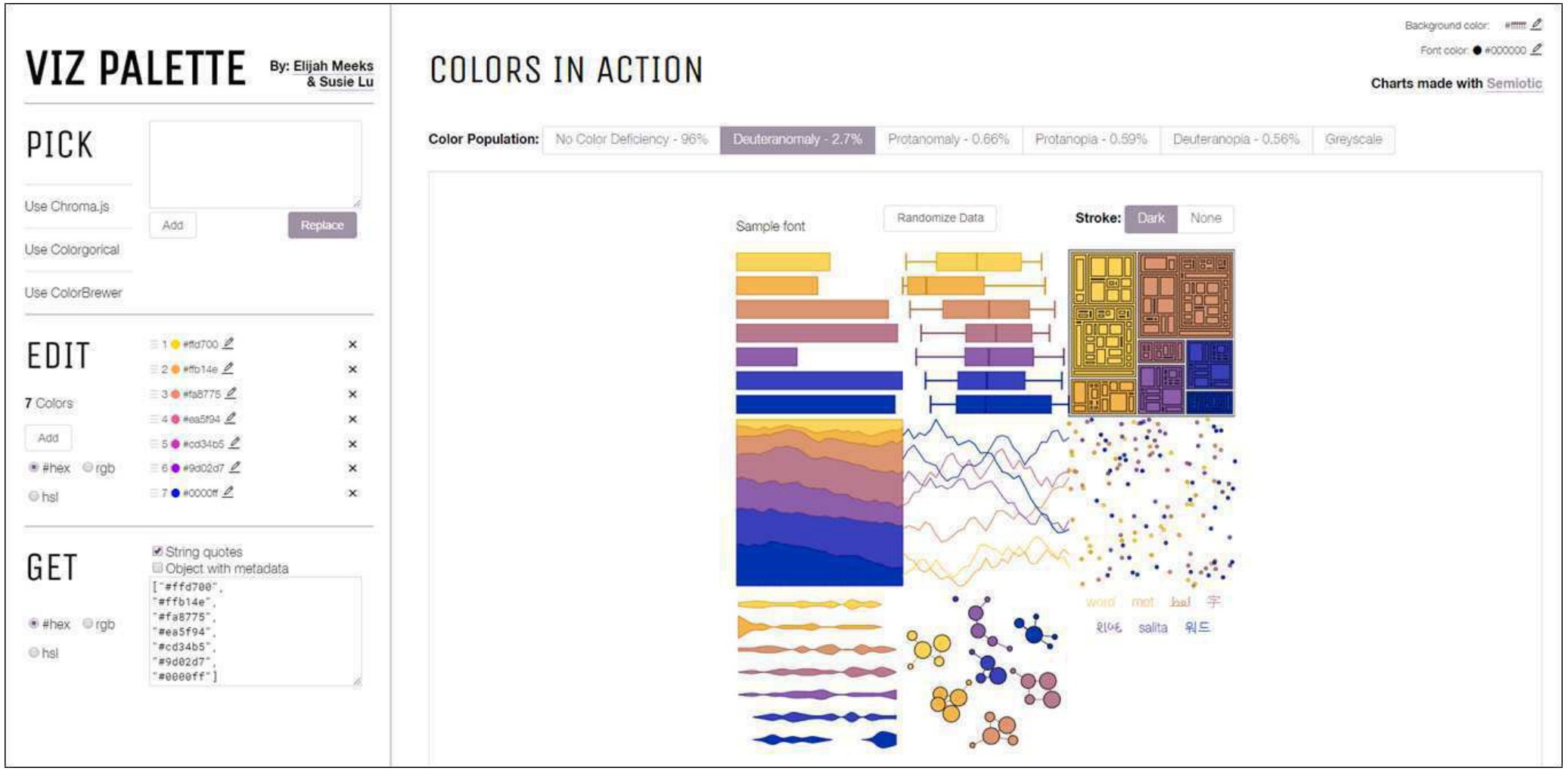
Source: "Fundamentals of Data Visualization" by Claus Wilke

Choice of the Color Palette & Accessibility



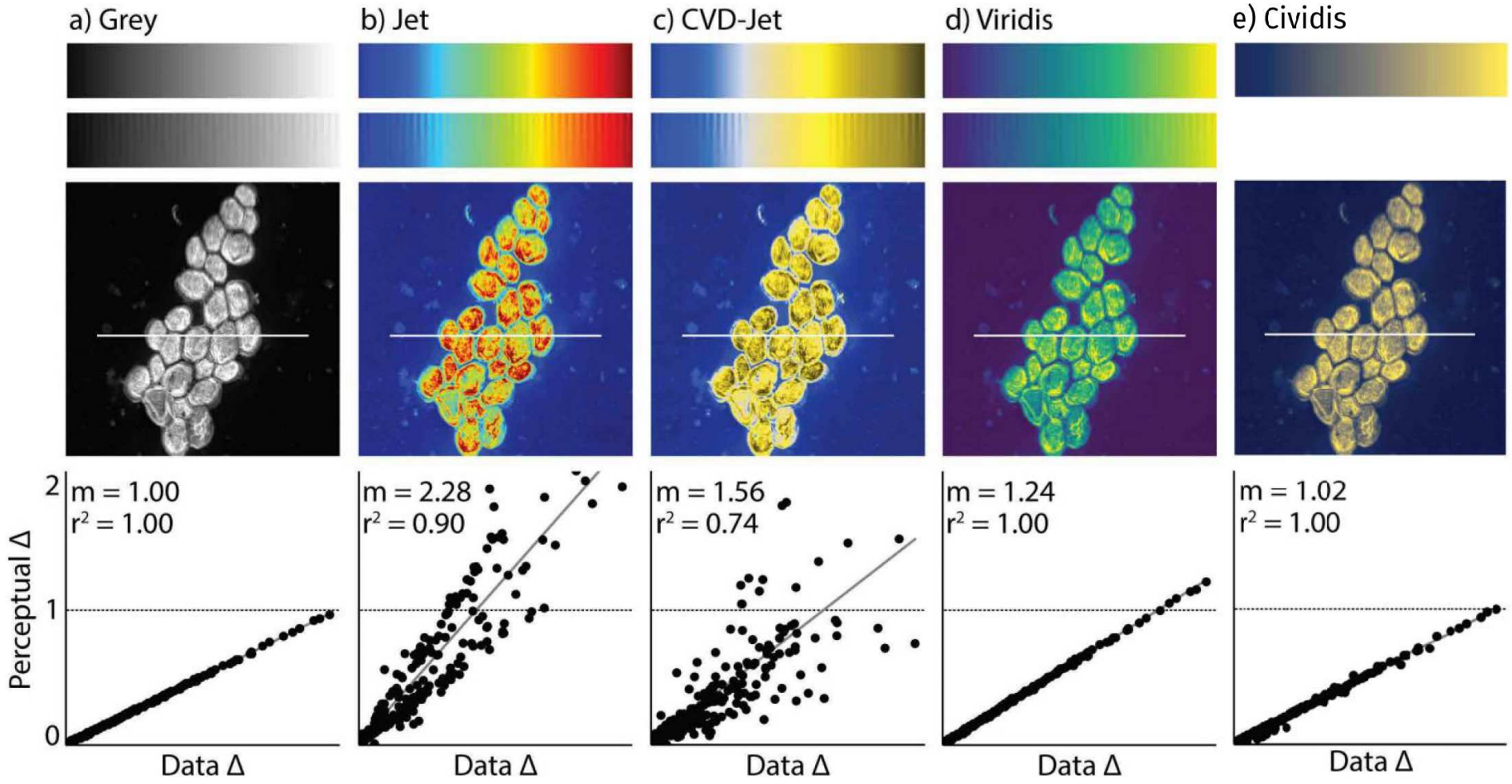
Source: "Fundamentals of Data Visualization" by Claus Wilke

Choice of the Color Palette & Accessibility



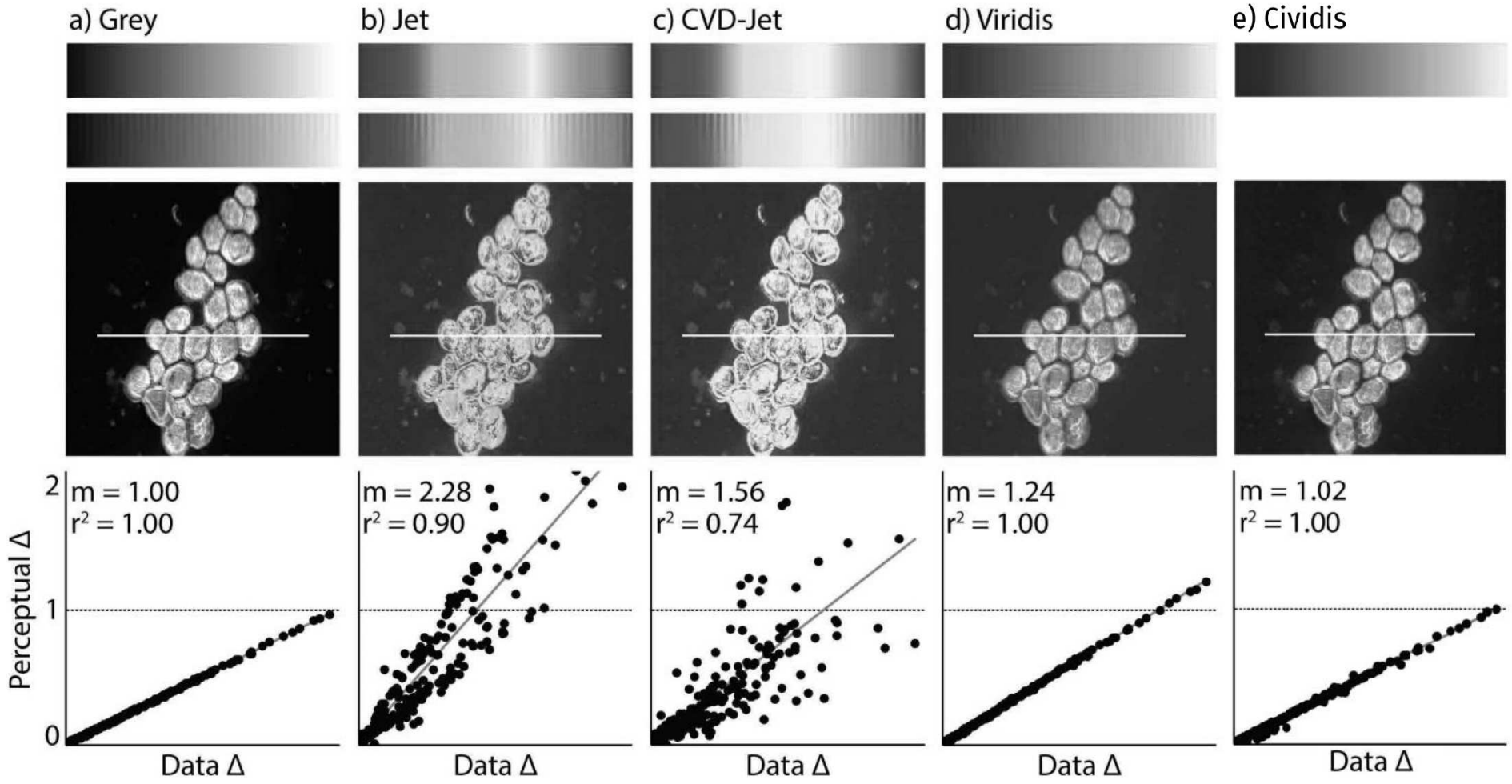
projects.susielu.com/viz-palette

Choice of the Color Palette & Accessibility

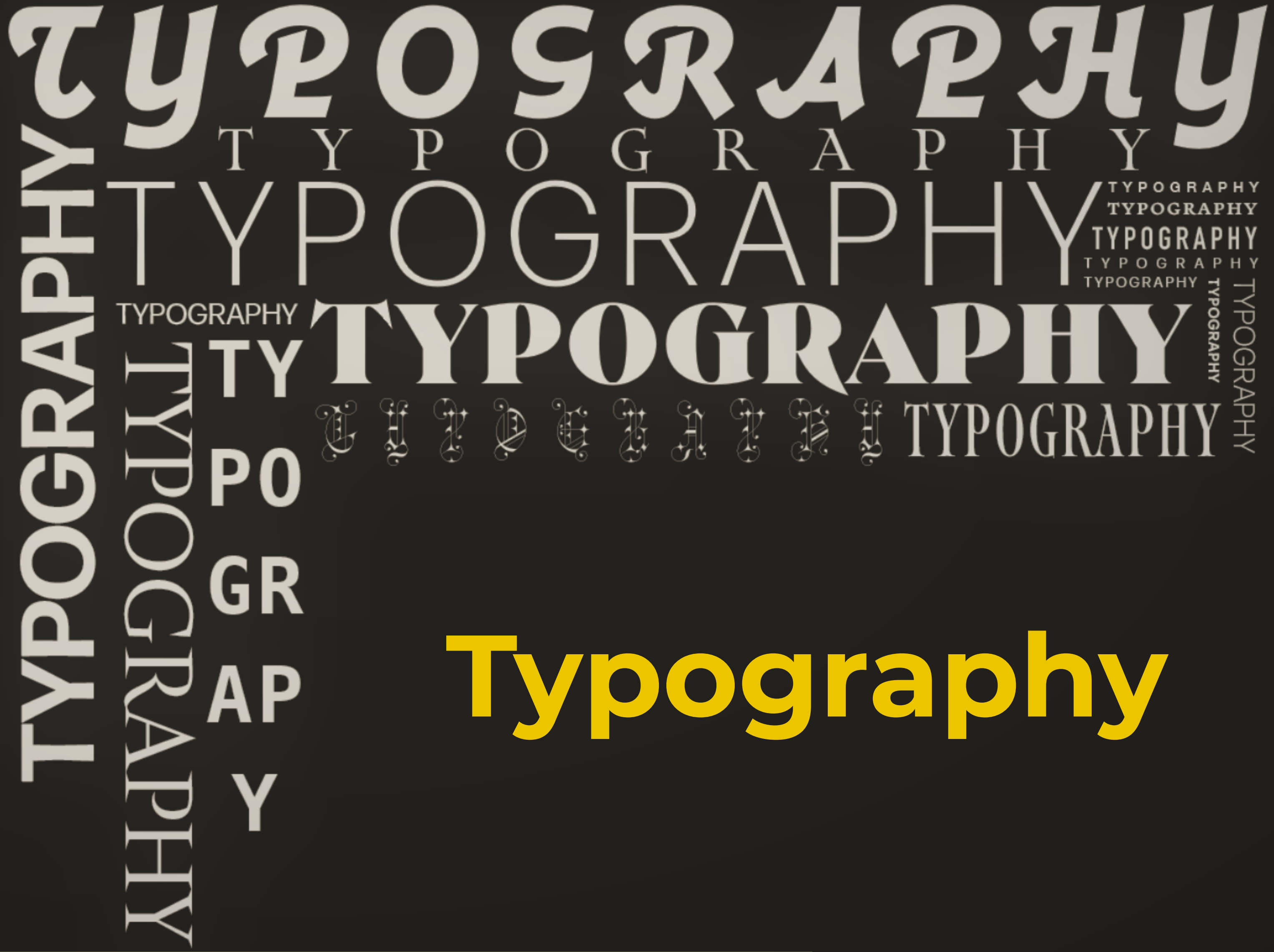


Modified from Nuñez, Anderton & Renslow (2018) PLOSone

Choice of the Color Palette & Accessibility



Modified from Nuñez, Anderton & Renslow (2018) PLOSone


TYPOGRAPHY
T Y P O G R A P H Y
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TY
PO
GR
AP
Y
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY
TYPOGRAPHY

Typography

You'll always be mine!

FONTS MATTER

The Choice of the Font(s)

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

The Choice of the Font(s)

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

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

I am important, too!

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

Yeah, I know I am kinda boring. Sorry.

How to Visualize Hierarchy

I am important!

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

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

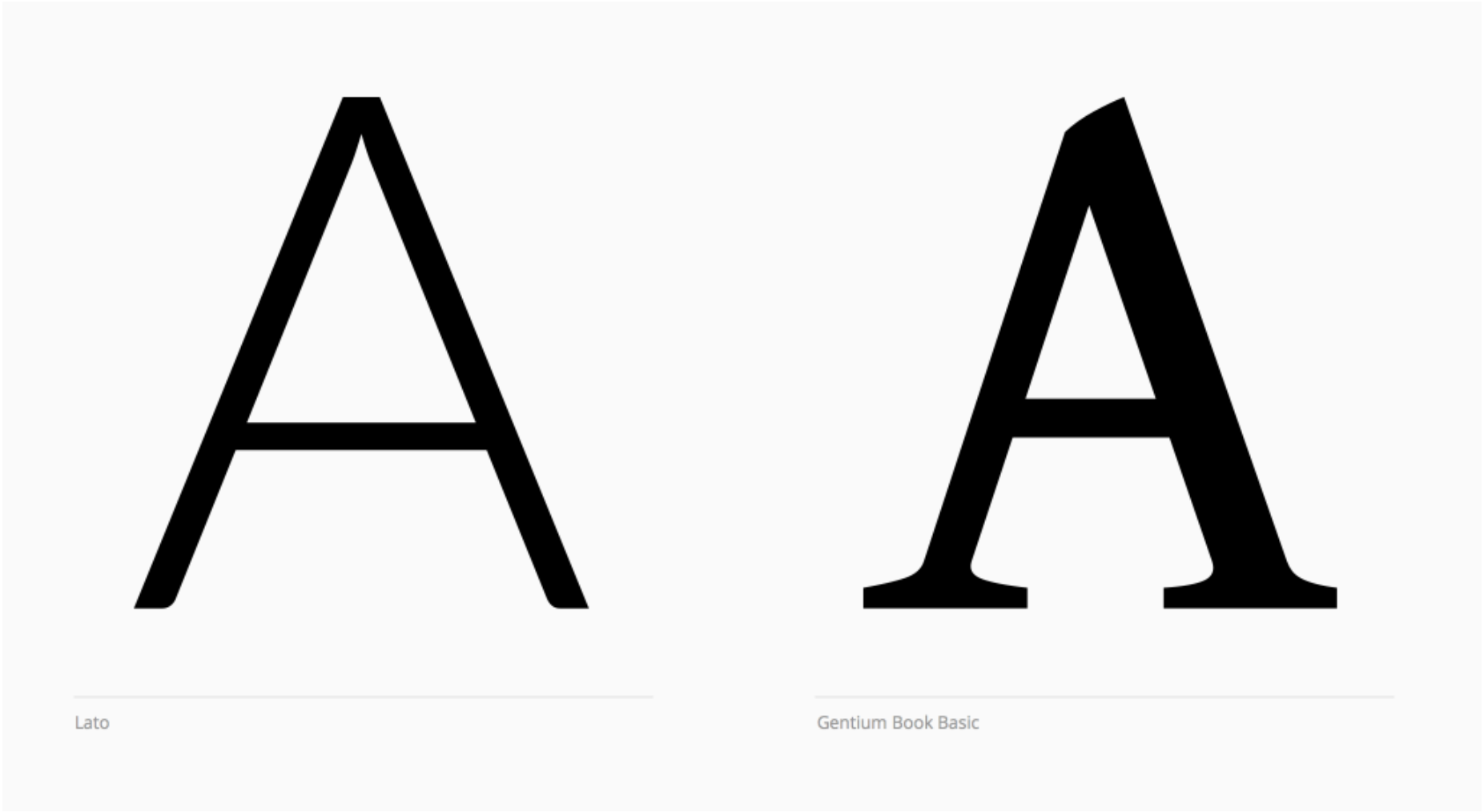
A Tale of Two Cities

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

Source: wordpress.com

Sans-Serif or Serif ?



Source: wordpress.com

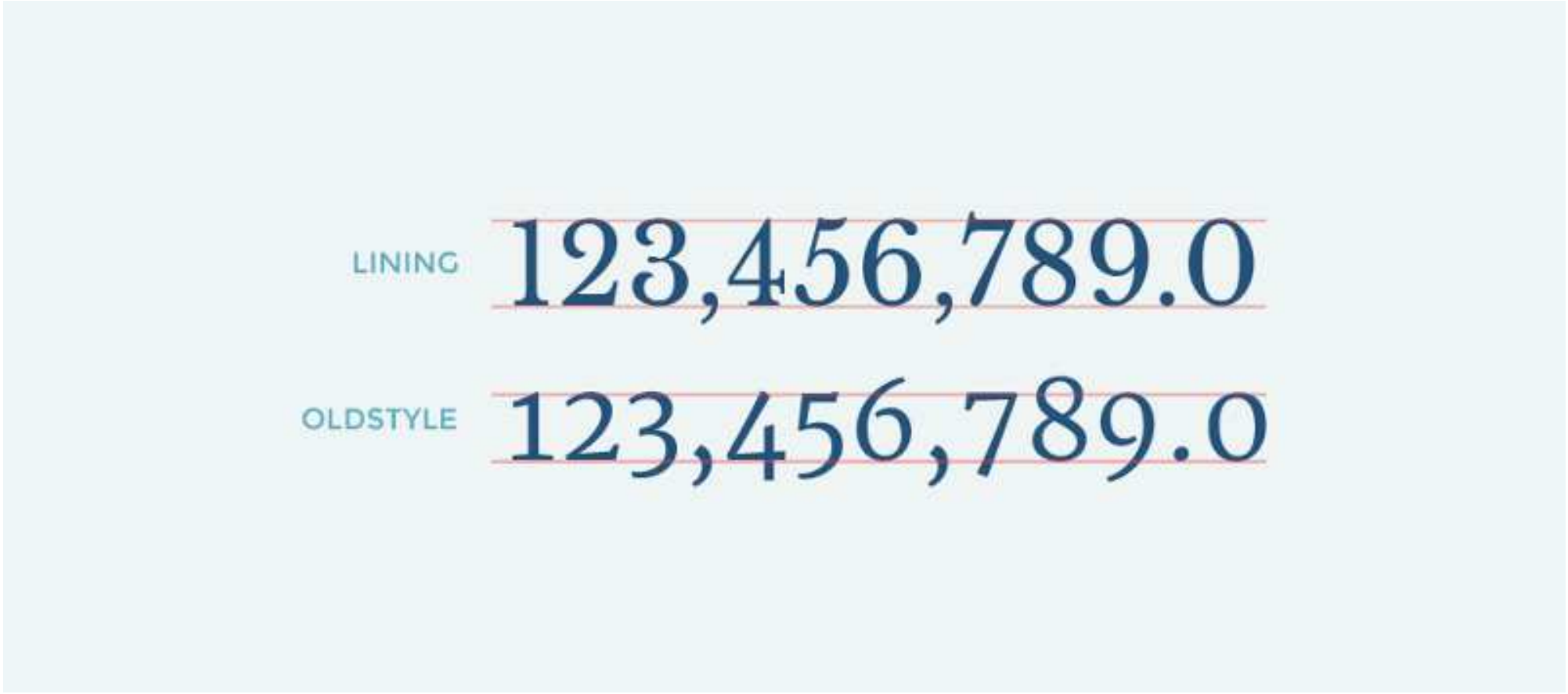
Keep it Simple

Source: canva.com

Tabular (Monospaced) Numbers

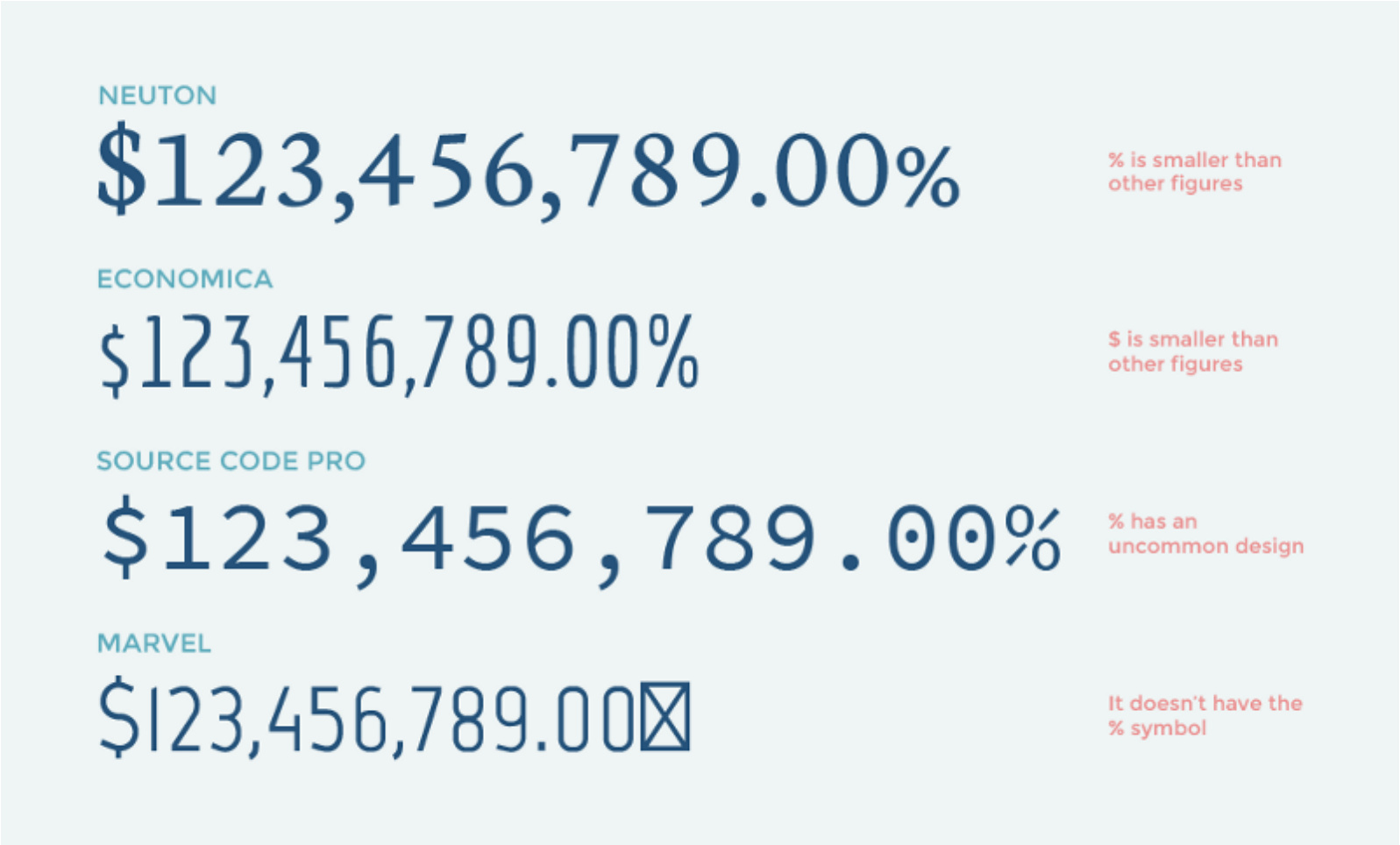
Source: invisionapp.com/inside-design/best-free-fonts-for-numbers

Number Fonts with Lining



Source: invisionapp.com/inside-design/best-free-fonts-for-numbers

Quality of Number Symbols



Source: 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%**

Source: invisionapp.com/inside-design/best-free-fonts-for-numbers

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%**

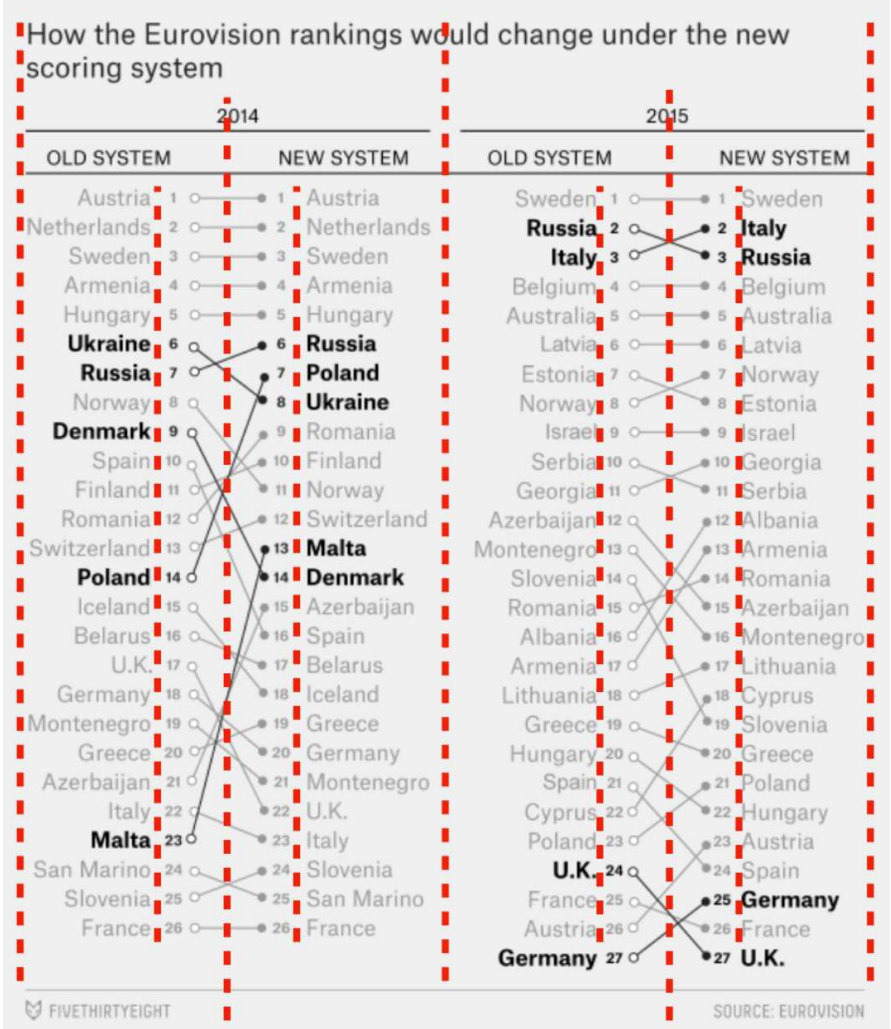
Varela Round

REGULAR $123,456,789.00%

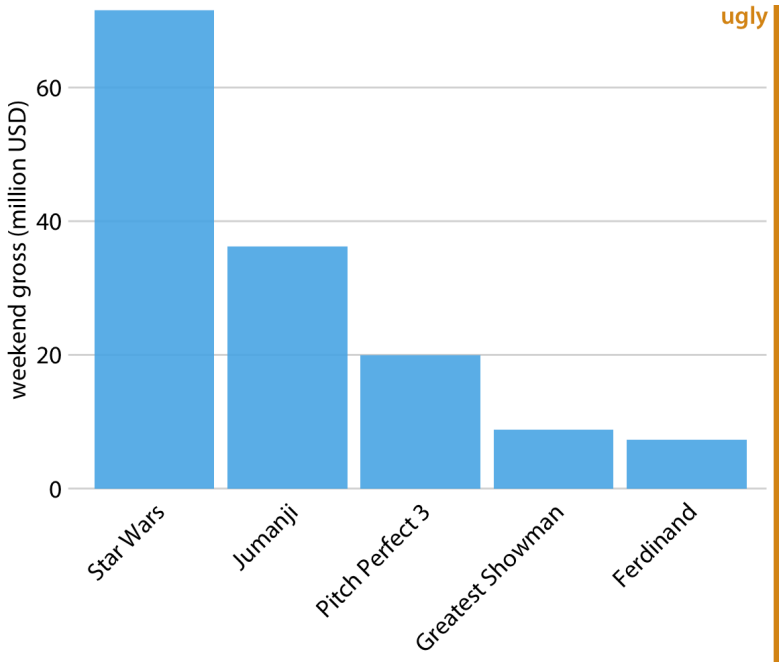
Source: invisionapp.com/inside-design/best-free-fonts-for-numbers

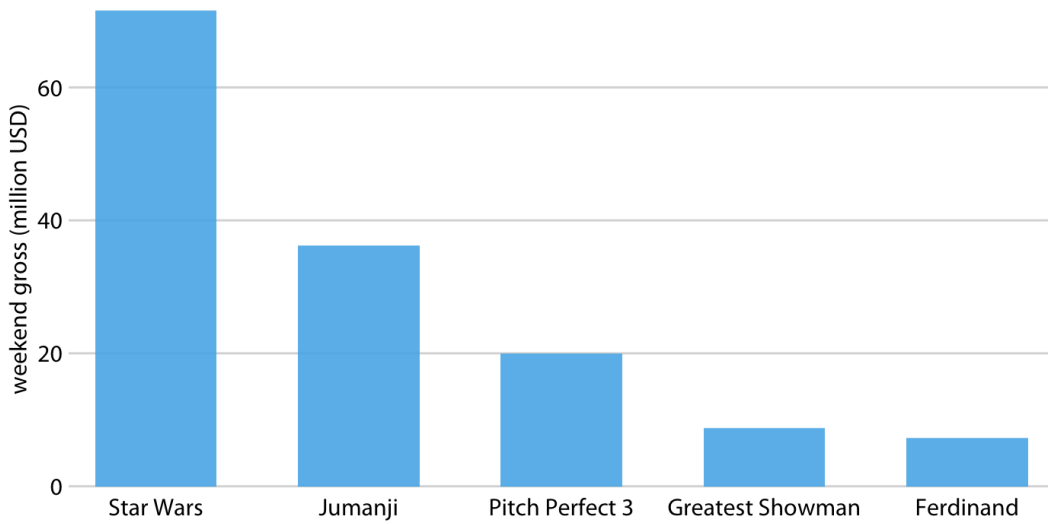
Allign Your Text!

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



(Don't) Rotate Your Text!





Source: "Fundamentals of Data Visualization" by Claus Wilke

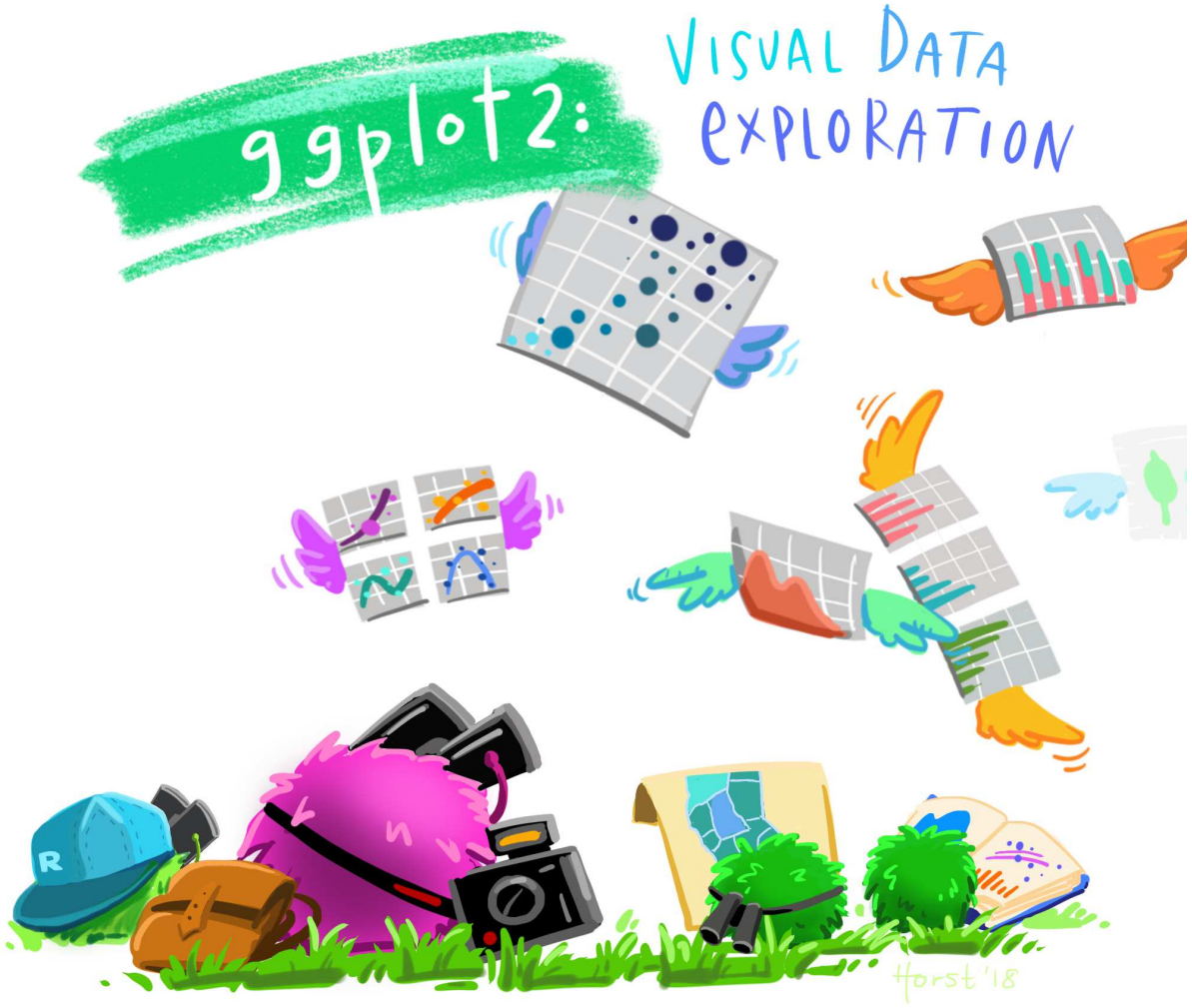


Illustration by Allison Horst (github.com/allisonhorst/stats-illustrations)