# Data Visualization for Effective Science Communication with Diverse Audiences

**Dr. Ali Sarvghad**

**Dr. Narges Mahyar**

Manning College of Information and Computer Sciences

University of Massachusetts Amherst

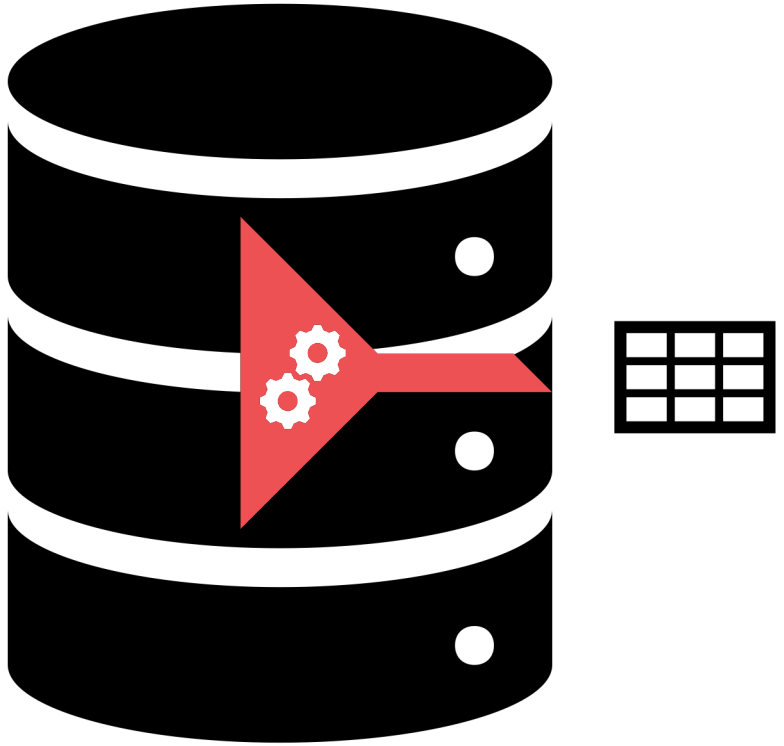
May 8th 2023

# Agenda

1. What is data visualization and why is it important?
2. Data visualization for communication
3. Know your audience
4. Is your visualization well-designed?
5. Building blocks of data visualization
6. How to choose effective mapping?
7. Effective use of color
8. Avoiding misleading visualization design
9. Storytelling with data
10. Wrap up and recommendation

# What is data visualization and why is it important?

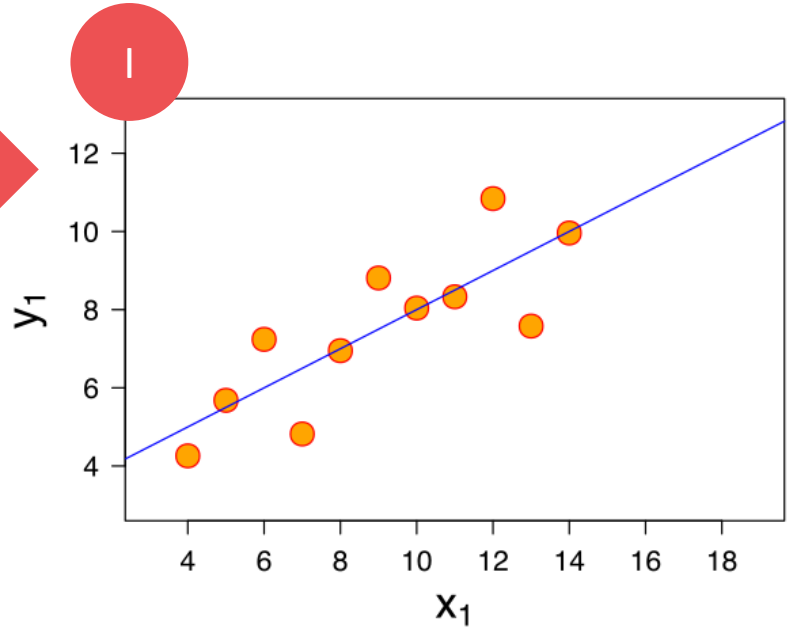
# Data Challenge

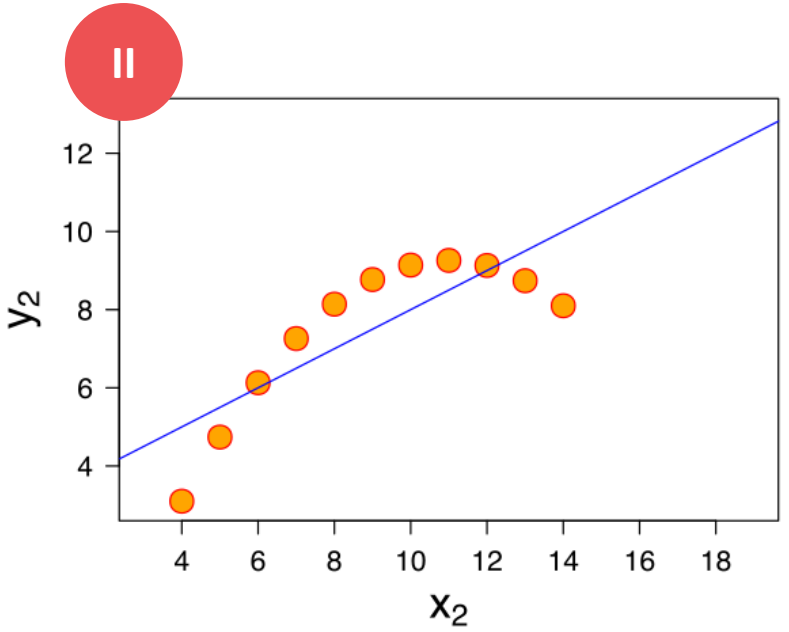


2.5 quintillion bytes of data **each day**!

| | I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|---|
| | x | y | x | y | x | y | x | y |
| | 10 | 8,04 | 10 | 9,14 | 10 | 7,46 | 8 | 6,58 |
| | 8 | 6,95 | 8 | 8,14 | 8 | 6,77 | 8 | 5,76 |
| | 13 | 7,58 | 13 | 8,74 | 13 | 12,74 | 8 | 7,71 |
| | 9 | 8,81 | 9 | 8,77 | 9 | 7,11 | 8 | 8,84 |
| | 11 | 8,33 | 11 | 9,26 | 11 | 7,81 | 8 | 8,47 |
| | 14 | 9,96 | 14 | 8,1 | 14 | 8,84 | 8 | 7,04 |
| | 6 | 7,24 | 6 | 6,13 | 6 | 6,08 | 8 | 5,25 |
| | 4 | 4,26 | 4 | 3,1 | 4 | 5,39 | 19 | 12,5 |
| | 12 | 10,84 | 12 | 9,13 | 12 | 8,15 | 8 | 5,56 |
| | 7 | 4,82 | 7 | 7,26 | 7 | 6,42 | 8 | 7,91 |
| | 5 | 5,68 | 5 | 4,74 | 5 | 5,73 | 8 | 6,89 |
| SUM | 99,00 | 82,51 | 99,00 | 82,51 | 99,00 | 82,50 | 99,00 | 82,51 |
| AVG | 9,00 | 7,50 | 9,00 | 7,50 | 9,00 | 7,50 | 9,00 | 7,50 |
| STDEV | 3,32 | 2,03 | 3,32 | 2,03 | 3,32 | 2,03 | 3,32 | 2,03 |

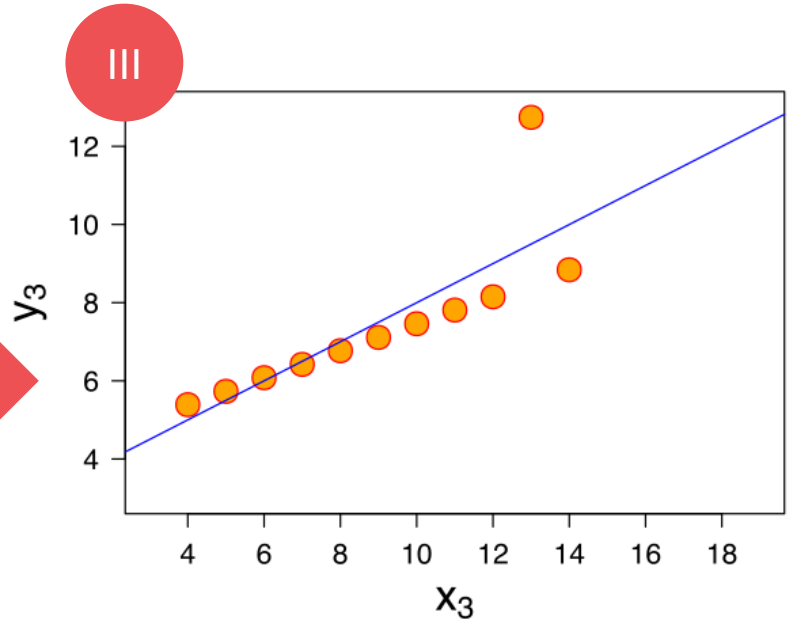
Weak correlation

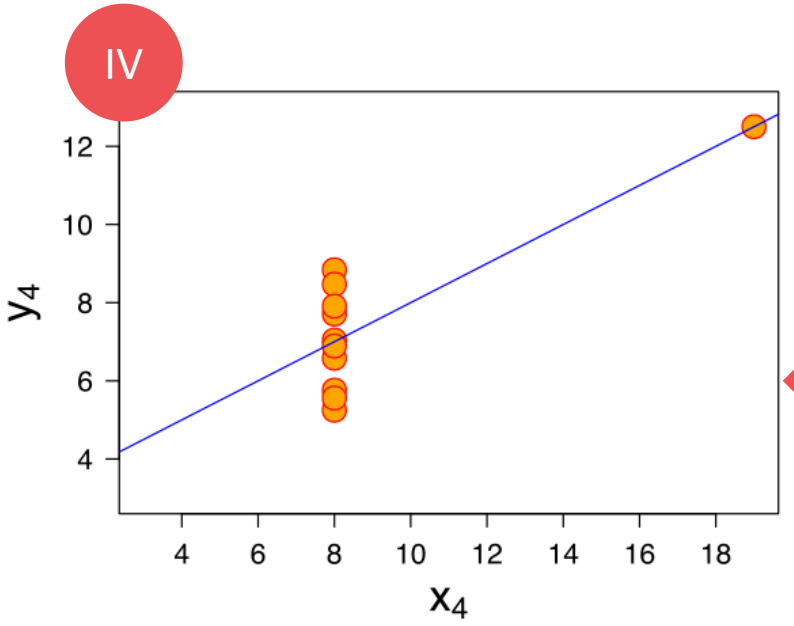


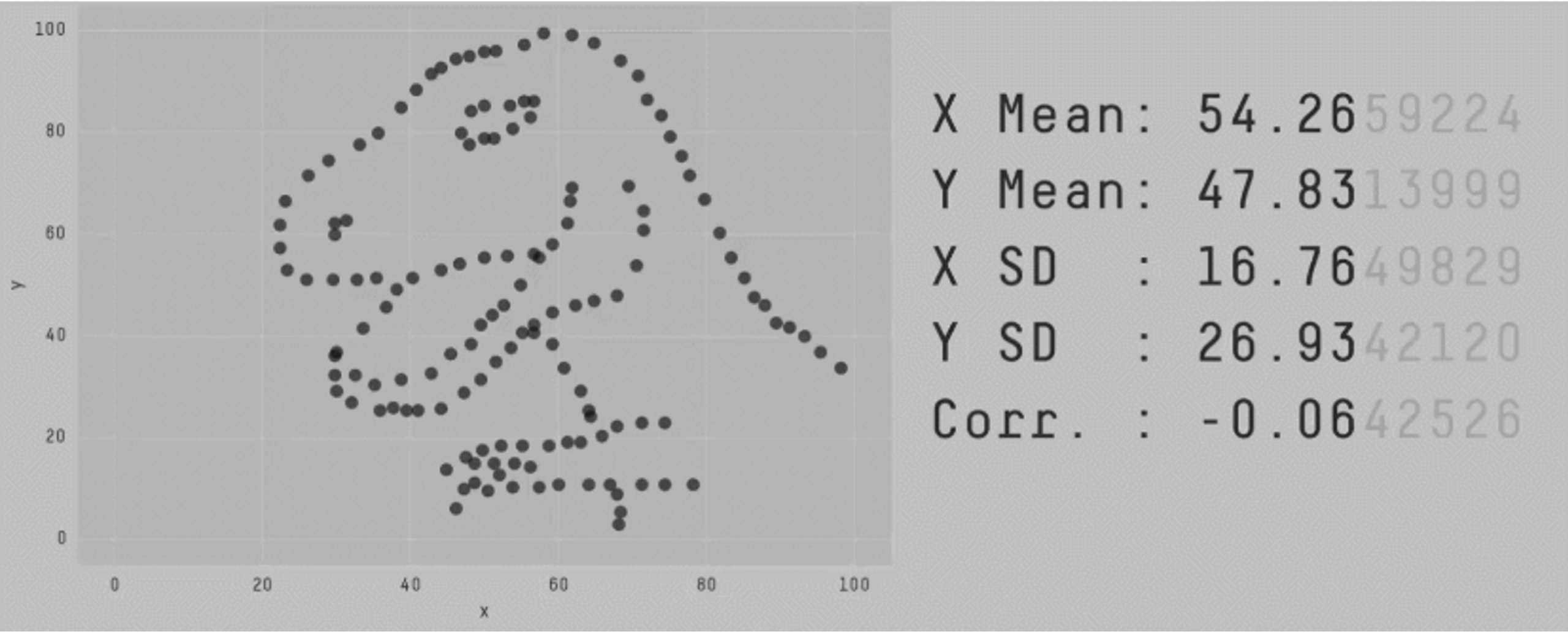

Non-linear relationship

outlier




outlier




*Matejka, Justin, and George Fitzmaurice. "Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing." CHI2017, ACM, 2017.*

# What is data visualization?

The engineering definition:

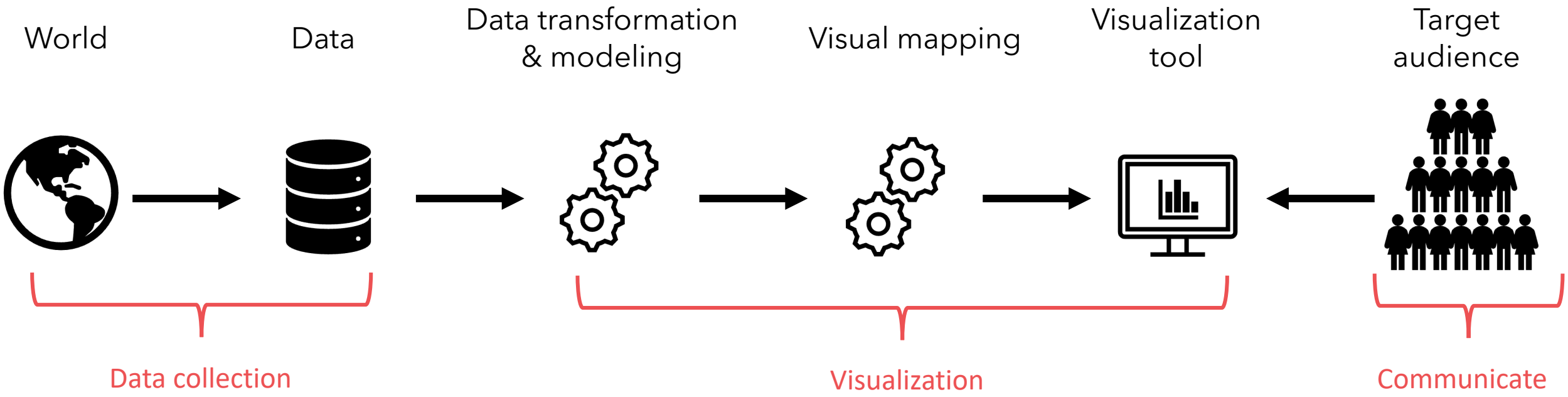
- Binding data values and structures to graphical elements on a display

Other definitions:

- Use of computer-supported, **interactive** visual representations of data to **amplify cognition** (Card, Mackinlay, Schneiderman' 98)

- The **depiction of information using spatial or graphical representations**, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system (Hearst' 03)

# Data visualization & communication pipeline

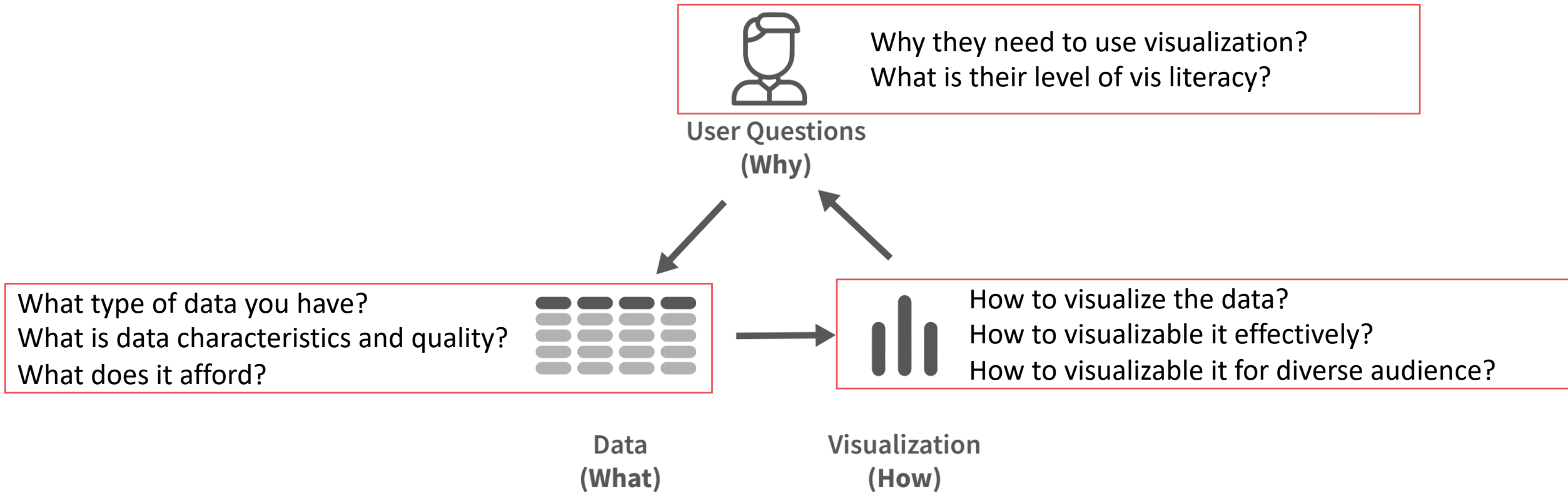


# Data visualization for communication

# Data visualization for communication

A primary goal of data visualization is to
**communicate** information **clearly** and **efficiently**
via statistical graphics, plots and information graphics.

# Data Visualization for Communication






Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021

# Know your audience

# Who is your audience?

- General public
- Domain expert/s
- Analyst/s

# Who is your audience?

What is their level of familiarity with visualization?

- VLAT: Visualization Literacy Assessment Test
- CALVI: Critical Thinking Assessment for Literacy in Visualizations.

What is their cognitive characteristics?

- People with cognitive impairments
- Children

Ge, Lily W., Yuan Cui, and Matthew Kay. "CALVI: Critical Thinking Assessment for Literacy in Visualizations." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.

# What type of device they will use?

- Desktop
- Tablet
- Phone
- Watch

# Ways to learn about your audience

Techniques:

- Questionnaire
- Interview
- Contextual inquiry
- Focus group

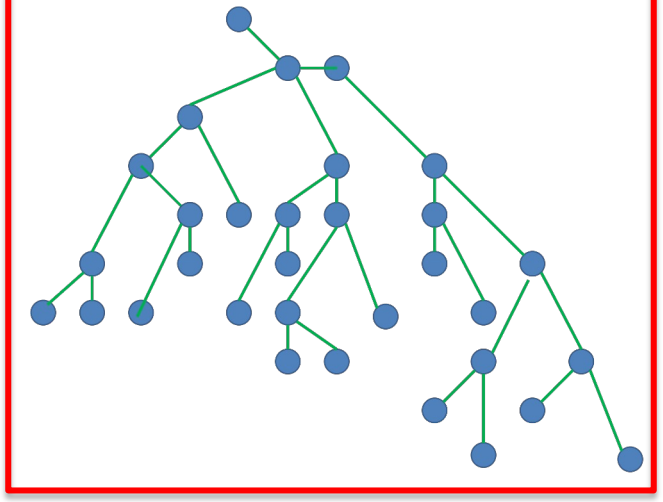
# What is their objective of using visualization?

- Discover insights
- Make decisions
- Explore and communicate

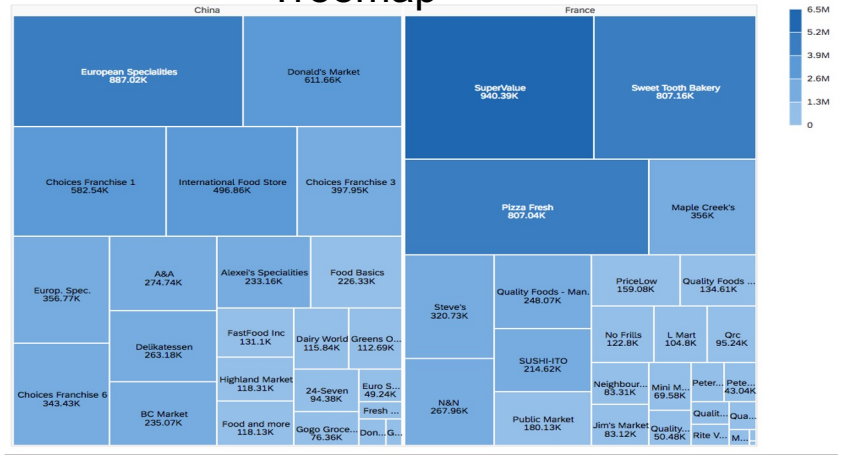
What : Hierarchical data

Why : Understand topology of hierarchy

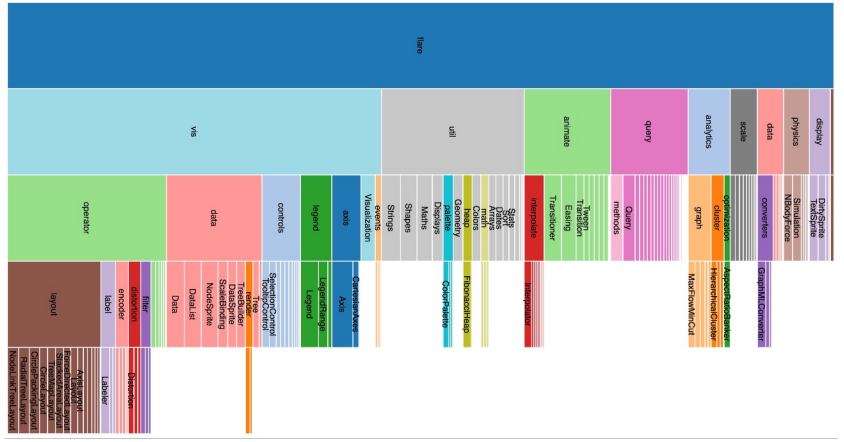
Tree



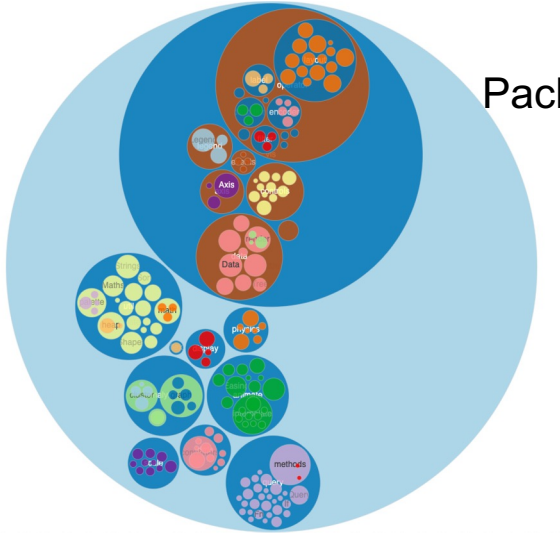
Treemap



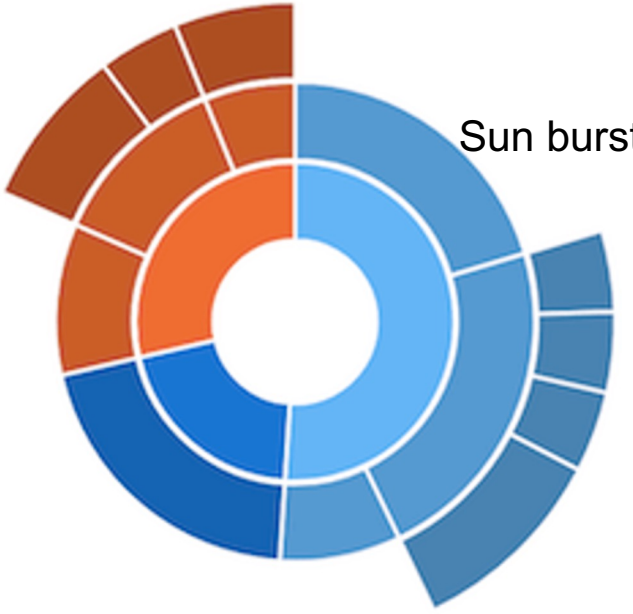

Icicle plot



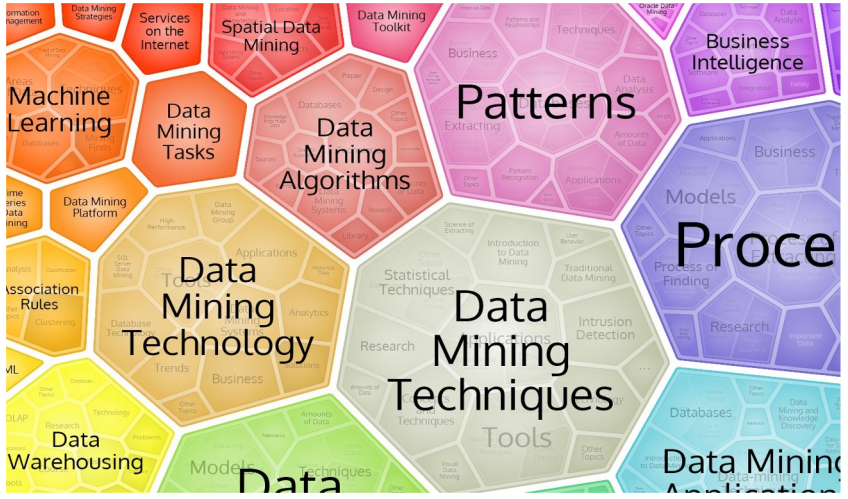

Packed circles




Sun burst



Voronoi treemap



# Who Do We Mean When We Talk About Visualization Novices?

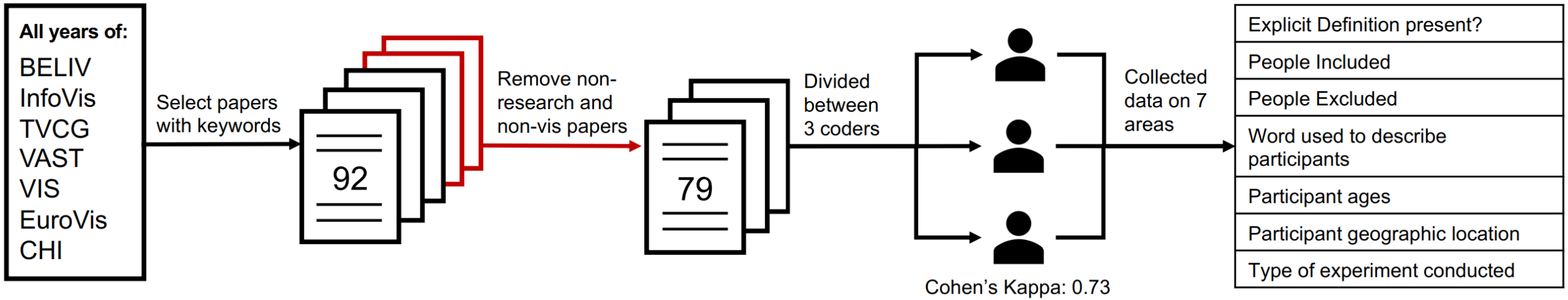
Alyxander Burns
Mount Holyoke College, University of Massachusetts Amherst
Massachusetts, USA
ajburns@mtholyoke.edu

Christiana Lee*
Ria Chawla*
christianale@umass.edu
rchawla@umass.edu
University of Massachusetts Amherst
Massachusetts, USA

Evan Peck
Bucknell University
Pennsylvania, USA
evan.peck@bucknell.edu

Narges Mahyar
University of Massachusetts Amherst
Massachusetts, USA
nmahyar@cs.umass.edu

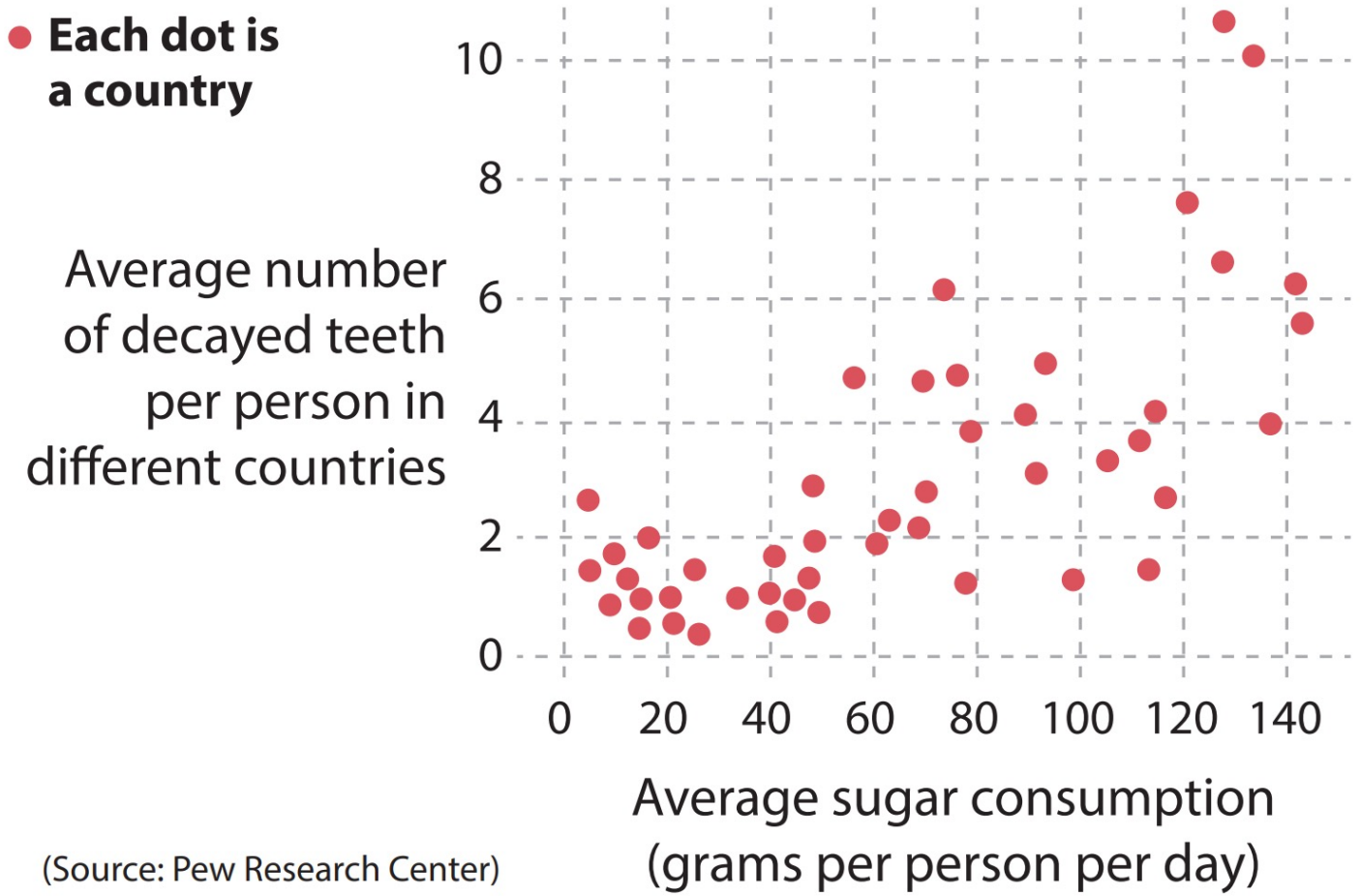
Burns, A., Lee, C., Chawla, R., Peck, E. and Mahyar, N., 2023, April. Who Do We Mean When We Talk About Visualization Novices?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).




80% of papers rely on implicit definitions for novices (often ambiguous)

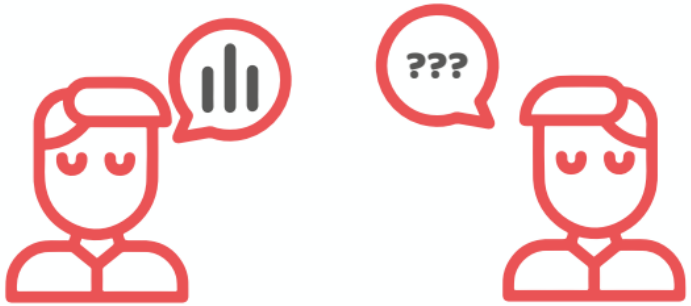
A novice is most likely to be young
A novice is most likely to be a university student
A novice is most likely to be a US resident
A novice is most likely to be lack traditional STEM experience

# Relationship between sugar consumption per person and average number of decayed teeth




Source: Adopted from How charts Lie, by Alberto Cairo

# Data Visualization and Communication




The old adage “a picture is worth 1000 words” should often be “a picture could be worth 1000 words if the picture was created properly”.

Visualization is a powerful tool, but only if used in the right way. **It’s not a silver bullet.**

Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021

# Ineffective & misleading visual design

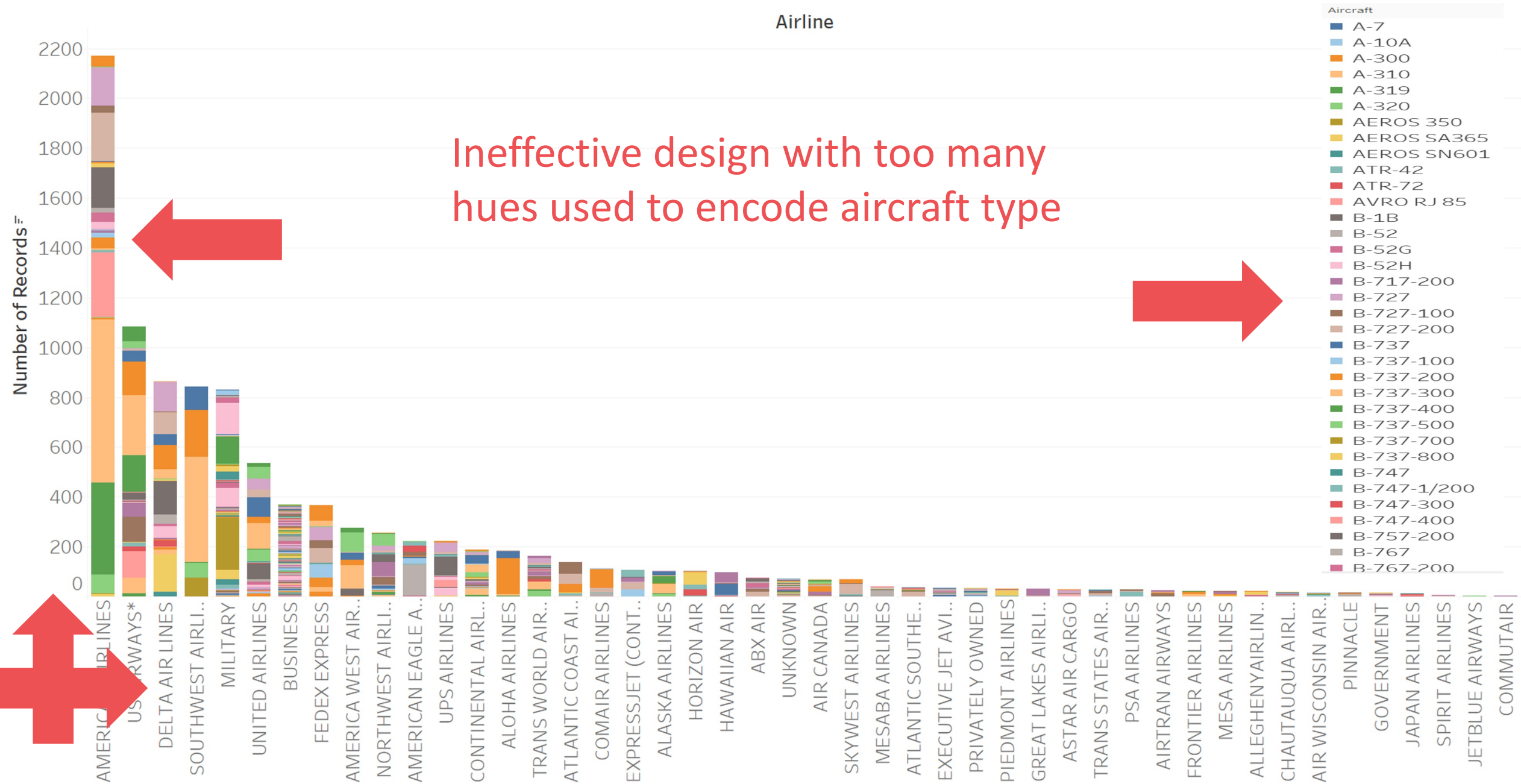
Ineffective visualization design

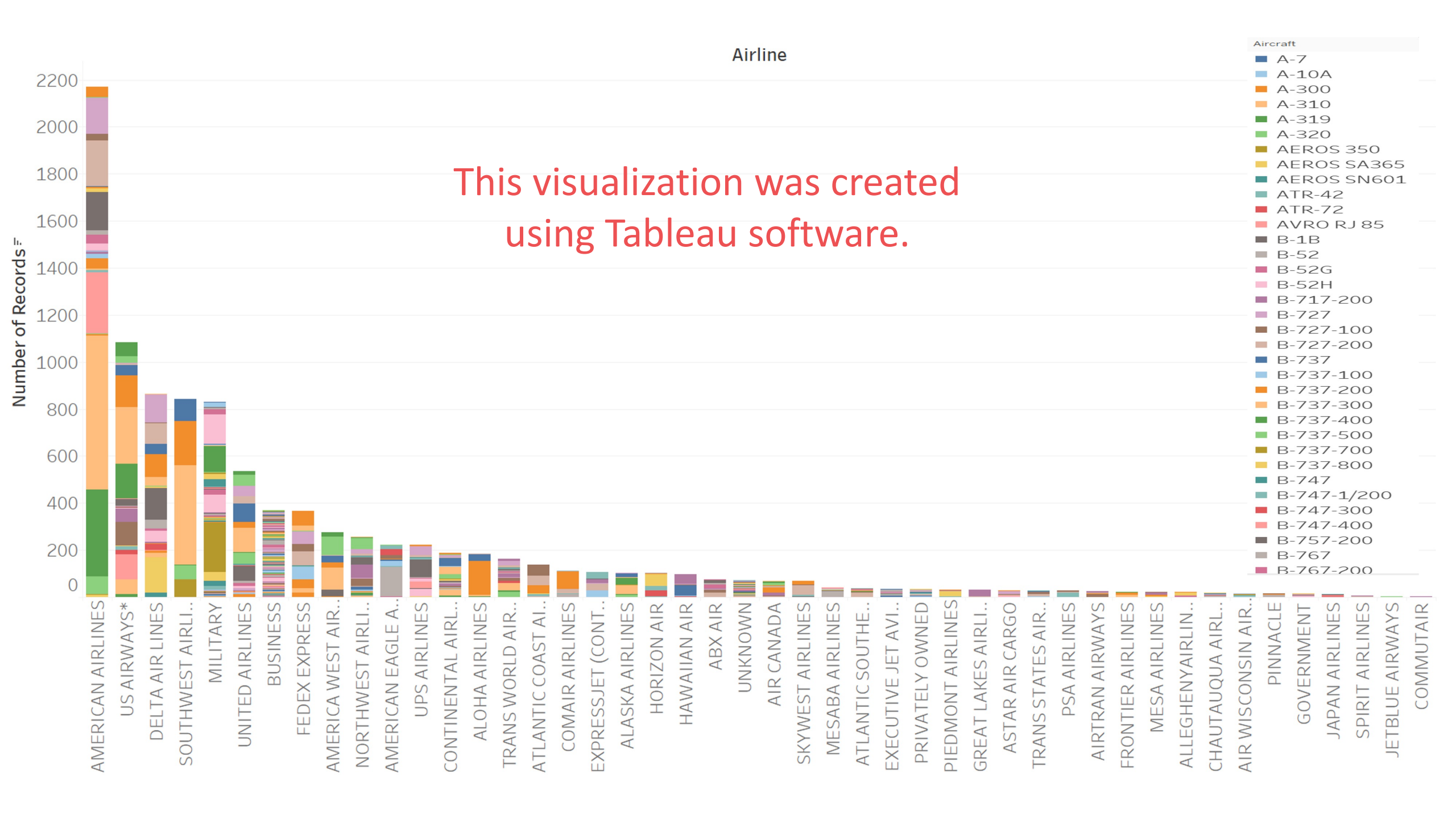
- Visualizations that do not adhere to empirically stablished guidelines of visualization design (e.g., improper use of color)

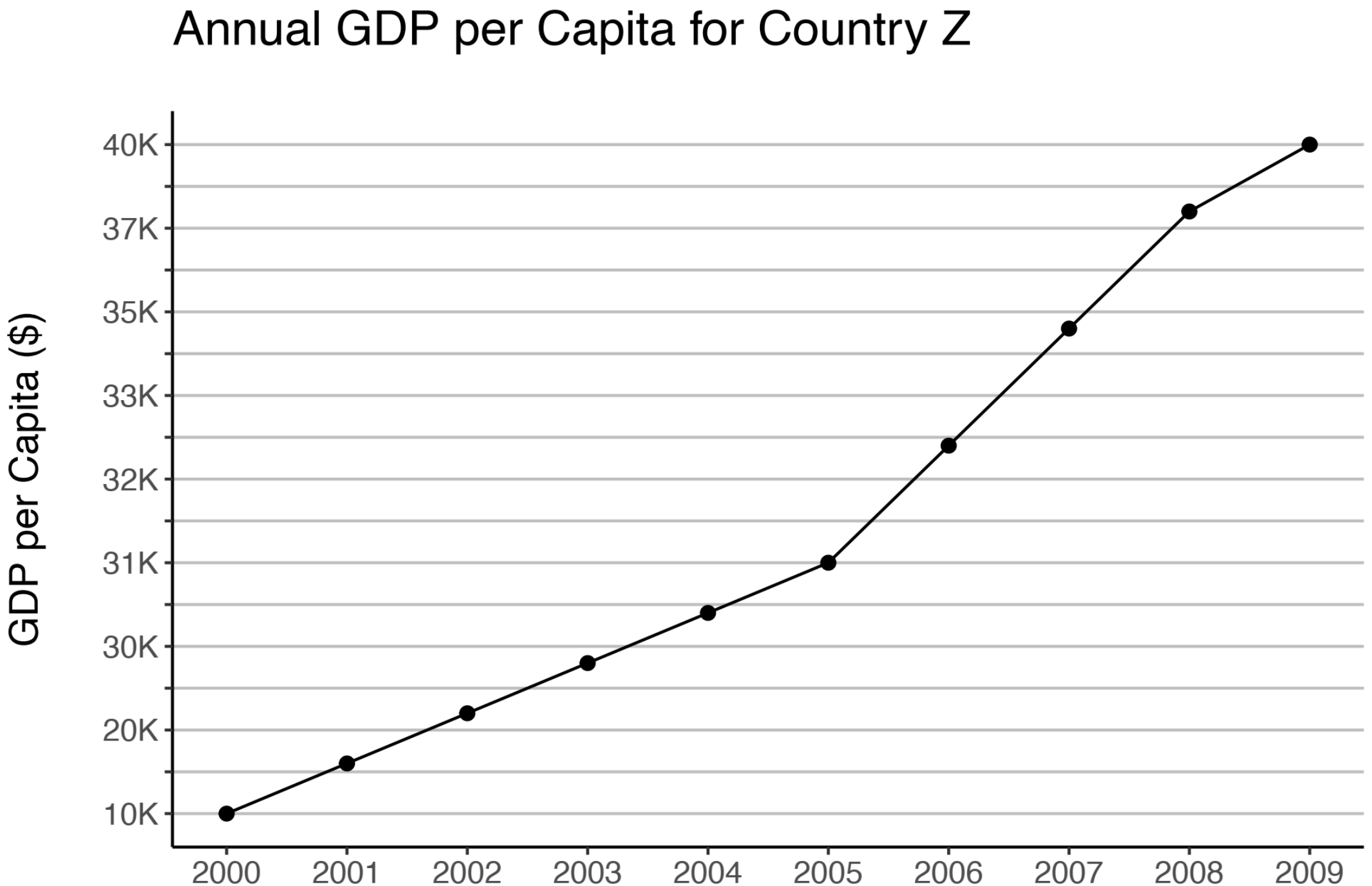
Misleading visualization design

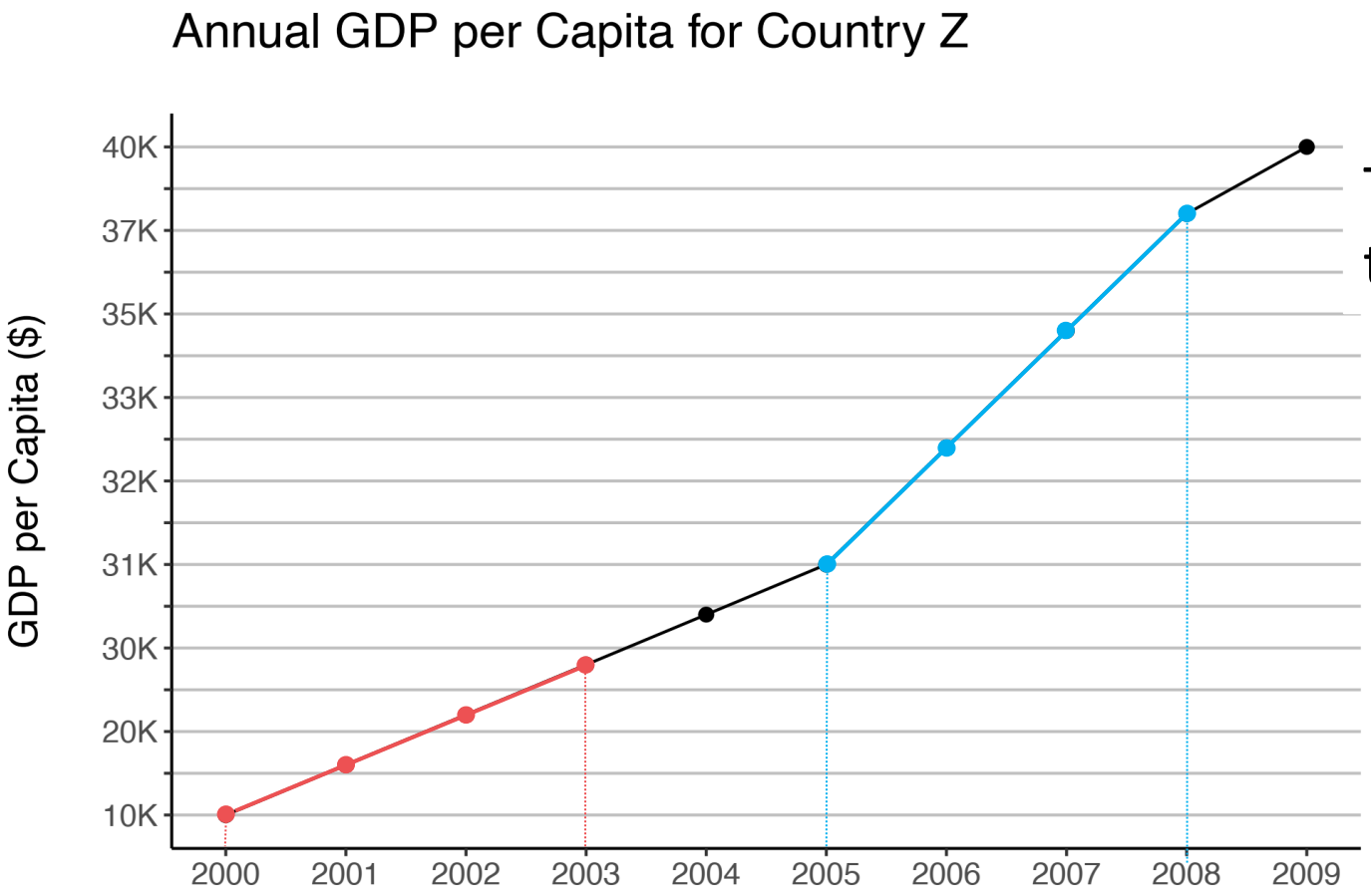
- Visualizations that are effectively designed, but can still cause in misleading conclusions (e.g., truncated axis)

# Is your visualization well-designed?


Airline
Number of Records
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Ineffective design with too many hues used to encode aircraft type
Aircraft
A-7
A-10A
A-300
A-310
A-319
A-320
AEROS 350
AEROS SA365
AEROS SN601
ATR-42
ATR-72
AVRO RJ 85
B-1B
B-52
B-52G
B-52H
B-717-200
B-727
B-727-100
B-727-200
B-737
B-737-100
B-737-200
B-737-300
B-737-400
B-737-500
B-737-700
B-737-800
B-747
B-747-1/200
B-747-300
B-747-400
B-757-200
B-767
B-767-200
DELTA AIR LINES
SOUTHWEST AIRLI..
MILITARY
UNITED AIRLINES
BUSINESS
FEDEX EXPRESS
AMERICA WEST AIR..
NORTHWEST AIRLI..
AMERICAN EAGLE A..
UPS AIRLINES
CONTINENTAL AIRL..
ALOHA AIRLINES
TRANS WORLD AIR..
ATLANTIC COAST AI..
COMAIR AIRLINES
EXPRESSJET (CONT..
ALASKA AIRLINES
HORIZON AIR
HAWAIIAN AIR
ABX AIR
UNKNOWN
AIR CANADA
SKYWEST AIRLINES
MESABA AIRLINES
ATLANTIC SOUTHE..
EXECUTIVE JET AVI..
PRIVATELY OWNED
PIEDMONT AIRLINES
GREAT LAKES AIRLI..
ASTAR AIR CARGO
TRANS STATES AIR..
PSA AIRLINES
AIRTRAN AIRWAYS
FRONTIER AIRLINES
MESA AIRLINES
ALLEGHENYAIRLIN..
CHAUTAUQUA AIRL..
AIR WISCONSIN AIR..
PINNACLE
GOVERNMENT
JAPAN AIRLINES
SPIRIT AIRLINES
JETBLUE AIRWAYS
COMMUTAIR


Airline
Number of Records
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
This visualization was created using Tableau software.
Aircraft
A-7
A-10A
A-300
A-310
A-319
A-320
AEROS 350
AEROS SA365
AEROS SN601
ATR-42
ATR-72
AVRO RJ 85
B-1B
B-52
B-52G
B-52H
B-717-200
B-727
B-727-100
B-727-200
B-737
B-737-100
B-737-200
B-737-300
B-737-400
B-737-500
B-737-700
B-737-800
B-747
B-747-1/200
B-747-300
B-747-400
B-757-200
B-767
B-767-200
AMERICAN AIRLINES
US AIRWAYS*
DELTA AIR LINES
SOUTHWEST AIRLI..
MILITARY
UNITED AIRLINES
BUSINESS
FEDEX EXPRESS
AMERICA WEST AIR..
NORTHWEST AIRLI..
AMERICAN EAGLE A..
UPS AIRLINES
CONTINENTAL AIRL..
ALOHA AIRLINES
TRANS WORLD AIR..
ATLANTIC COAST AI..
COMAIR AIRLINES
EXPRESSJET (CONT..
ALASKA AIRLINES
HORIZON AIR
HAWAIIAN AIR
ABX AIR
UNKNOWN
AIR CANADA
SKYWEST AIRLINES
MESABA AIRLINES
ATLANTIC SOUTHE..
EXECUTIVE JET AVI..
PRIVATELY OWNED
PIEDMONT AIRLINES
GREAT LAKES AIRLI..
ASTAR AIR CARGO
TRANS STATES AIR..
PSA AIRLINES
AIRTRAN AIRWAYS
FRONTIER AIRLINES
MESA AIRLINES
ALLEGHENYAIRLIN..
CHAUTAUQUA AIRL..
AIR WISCONSIN AIR..
PINNACLE
GOVERNMENT
JAPAN AIRLINES
SPIRIT AIRLINES
JETBLUE AIRWAYS
COMMUTAIR


Annual GDP per Capita for Country Z
GDP per Capita ($)
40K
37K
35K
33K
32K
31K
30K
20K
10K
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009



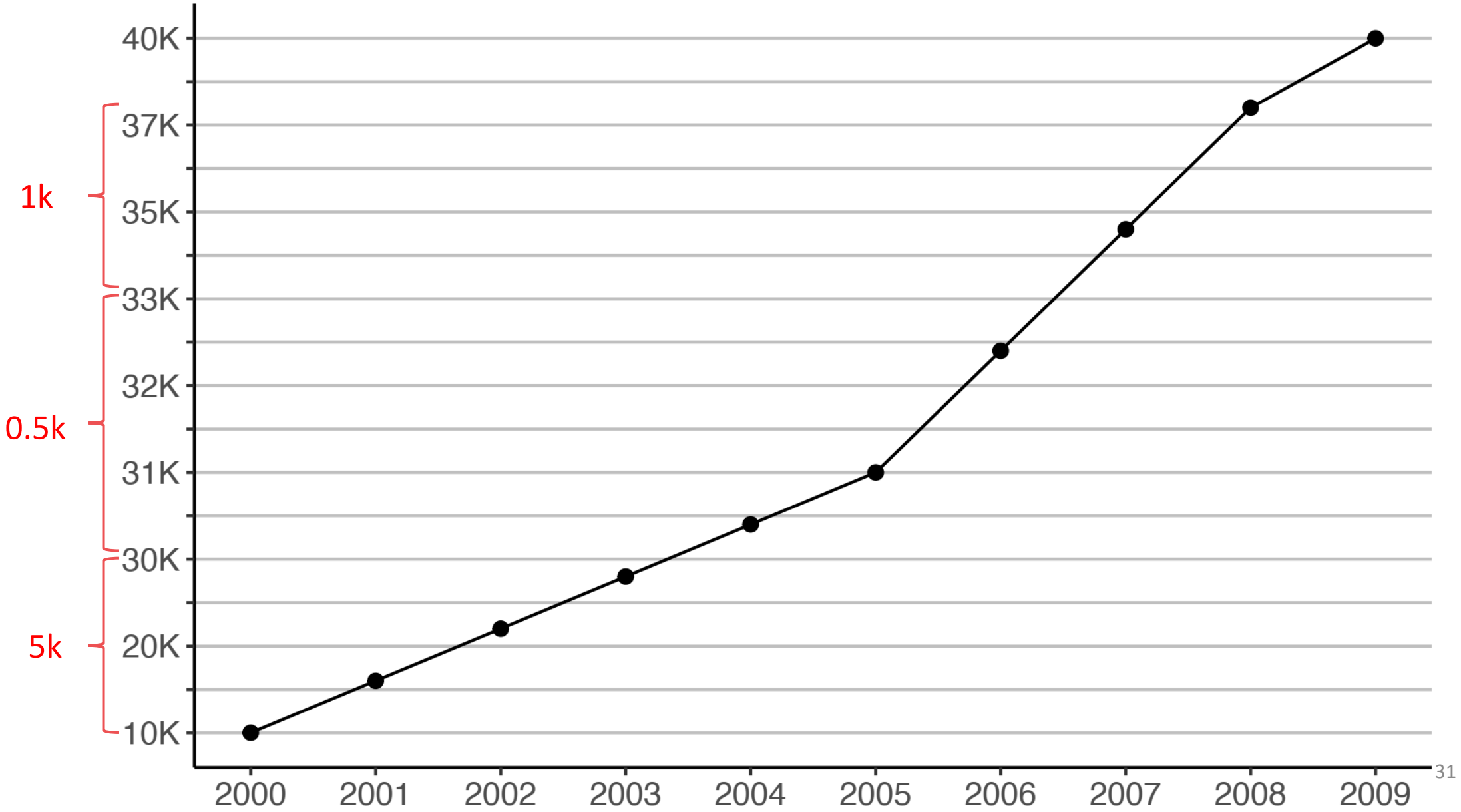

The GDP per capita grew faster from 2000 to 2003 than from 2005 to 2008.

- ○ True
- ○ False
- ○ Cannot be inferred / Inadequate Information

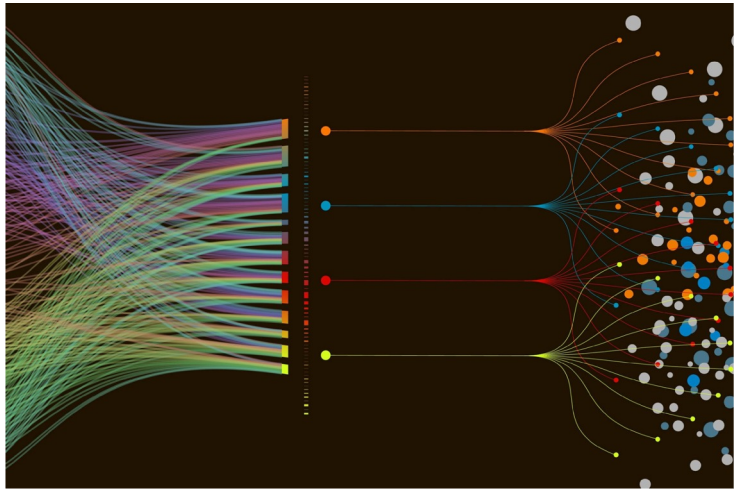
**GDP per capita for country Z grew faster from 2000 to 2033 than from 2005 to 2008.**

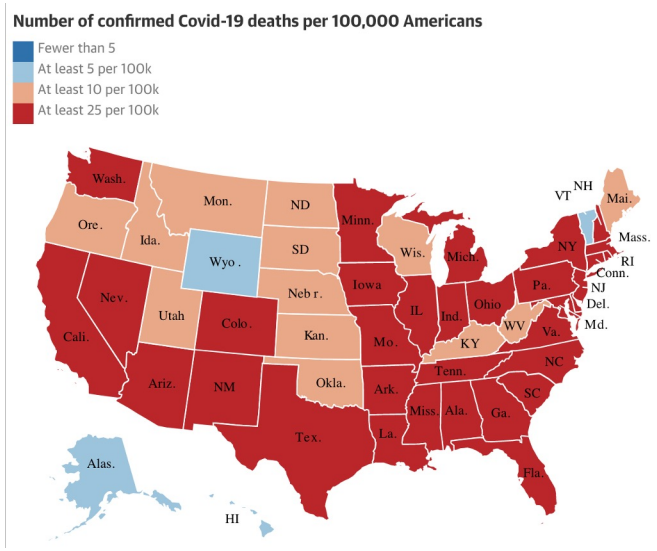
ⓘ Start presenting to display the poll results on this slide.

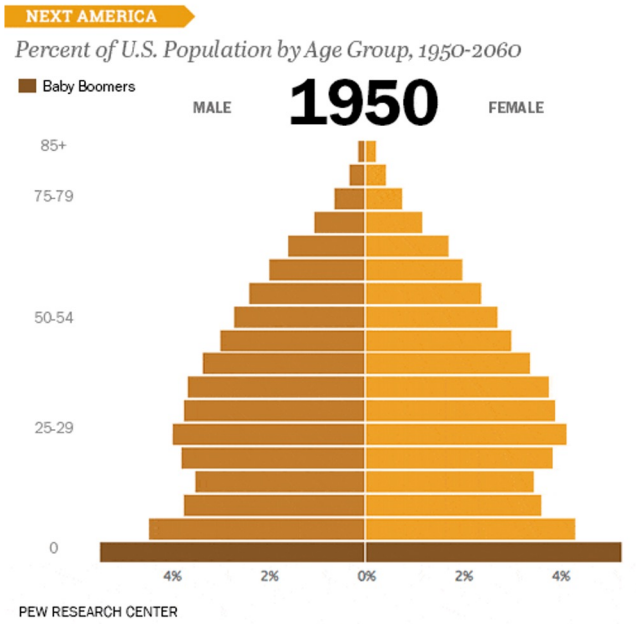
# None!

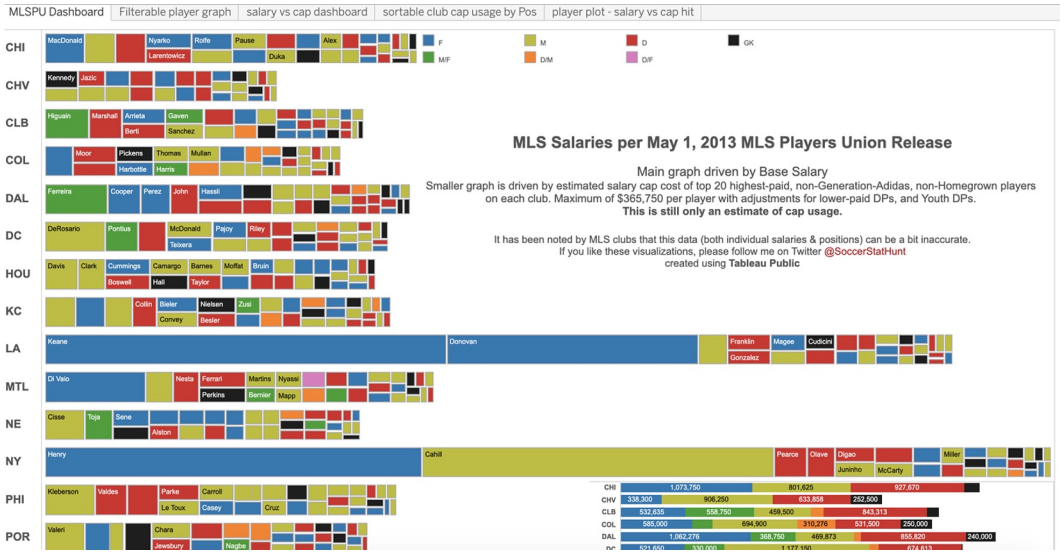

40K
37K
35K
33K
32K
31K
30K
20K
10K
1k
0.5k
5k
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009

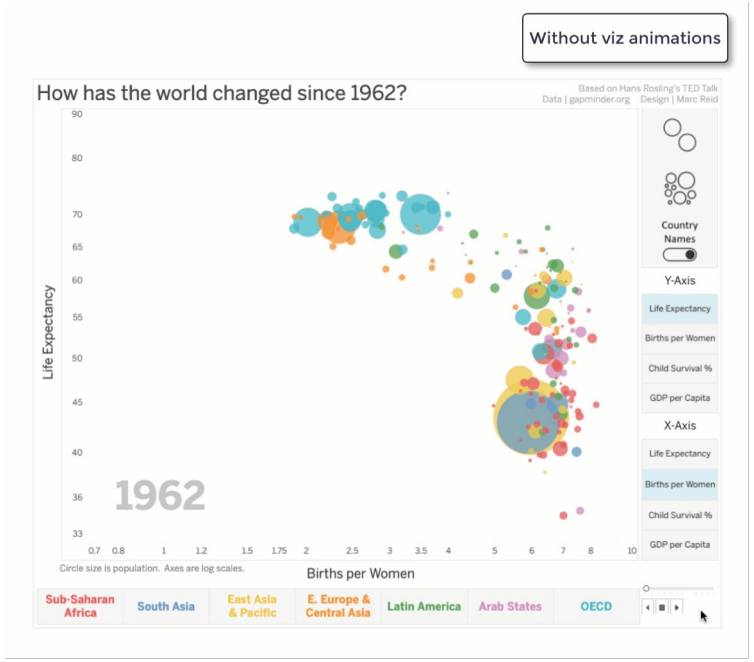
# Building blocks of data visualization

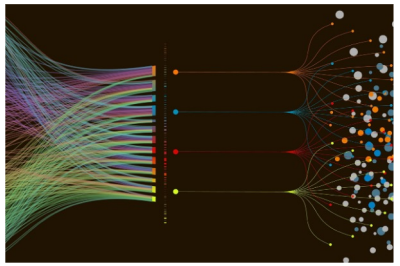


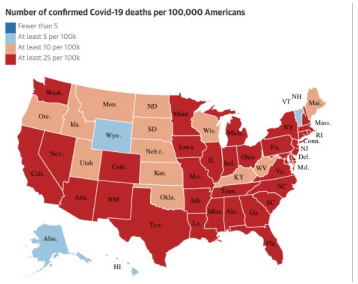

Number of confirmed Covid-19 deaths per 100,000 Americans
Fewer than 5
At least 5 per 100k
At least 10 per 100k
At least 25 per 100k
Wash.
Ore.
Mon.
Ida.
ND
Wyo.
SD
Minn
Wis.
Mich.
Nev.
Utah
Colo.
Neb r.
Iowa
IL
Ind.
Ohio
Cali.
Kan.
Mo.
WV
Va.
KY
Ariz.
NM
Okla.
Ark.
Tenn.
NC
SC
Miss.
Ala.
Ga.
Tex.
La.
Fla.
Alas.
HI
VT
NH
Mai.
NY
Mass.
RI
Conn.
Pa.
NJ
Del.
Md.

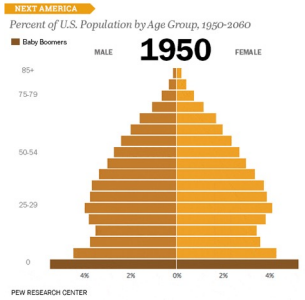

NEXT AMERICA
Percent of U.S. Population by Age Group, 1950-2060
Baby Boomers
MALE
1950
FEMALE
85+
75-79
50-54
25-29
0
4%
2%
0%
2%
4%
PEW RESEARCH CENTER

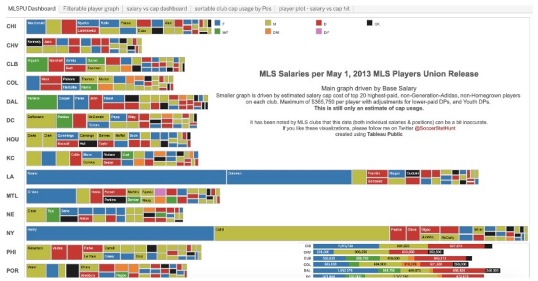

MLSPU Dashboard
Filterable player graph
salary vs cap dashboard
sortable club cap usage by Pos
player plot - salary vs cap hit
MLS Salaries per May 1, 2013 MLS Players Union Release
Main graph driven by Base Salary
Smaller graph is driven by estimated salary cap cost of top 20 highest-paid, non-Generation-Adidas, non-Homegrown players on each club. Maximum of $365,750 per player with adjustments for lower-paid DPs, and Youth DPs.
This is still only an estimate of cap usage.
It has been noted by MLS clubs that this data (both individual salaries & positions) can be a bit inaccurate.
If you like these visualizations, please follow me on Twitter @SoccerStatHunt
created using Tableau Public
CHI
CHV
CLB
COL
DAL
DC
HOU
KC
LA
MTL
NE
NY
PHI
POR


Without viz animations
How has the world changed since 1962?
Based on Hans Rosling's TED Talk
Data | gapminder.org
Design | Marc Reid
Life Expectancy
1962
Country Names
Y-Axis
Life Expectancy
Births per Women
Child Survival %
GDP per Capita
X-Axis
Life Expectancy
Births per Women
Child Survival %
GDP per Capita
Circle size is population. Axes are log scales.
Births per Women
Sub-Saharan Africa
South Asia
East Asia & Pacific
E. Europe & Central Asia
Latin America
Arab States
OECD

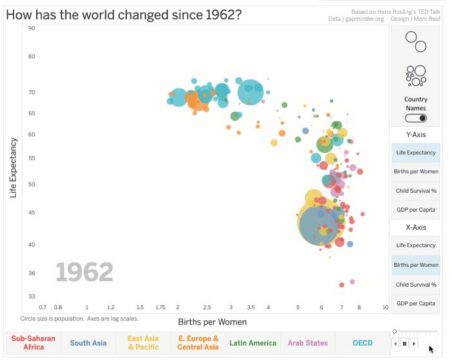




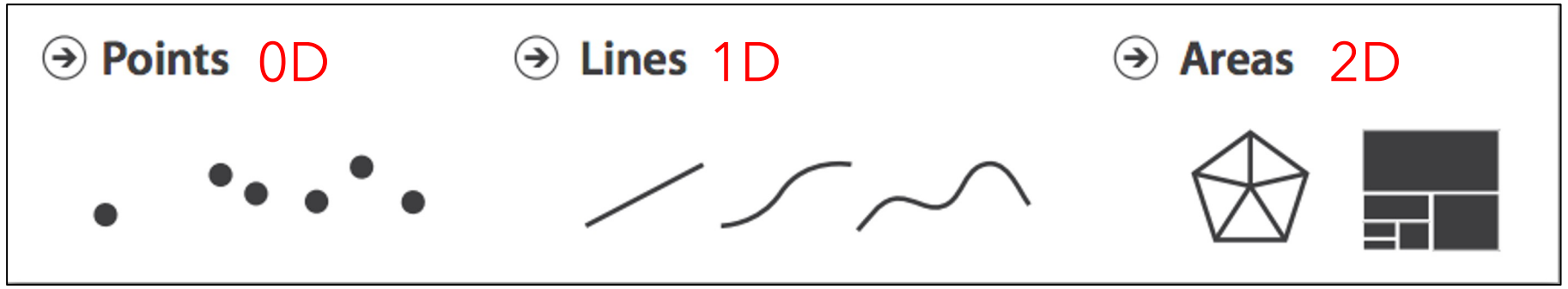

Without viz animations



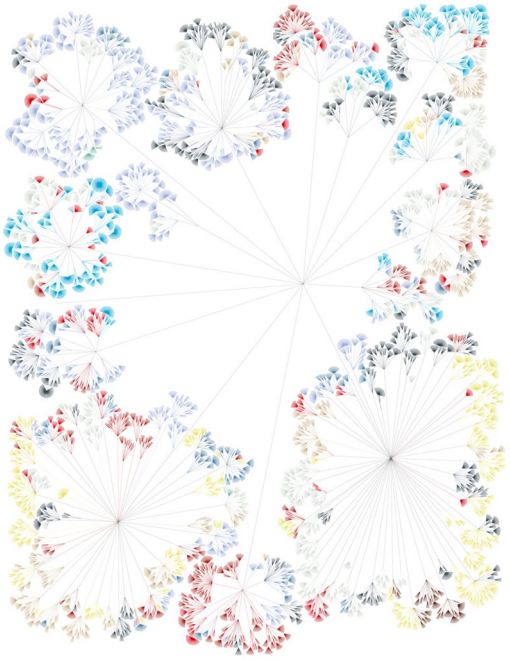

All these visualizations, despite their apparent differences, are built using a few primitive geometric objects that we call:

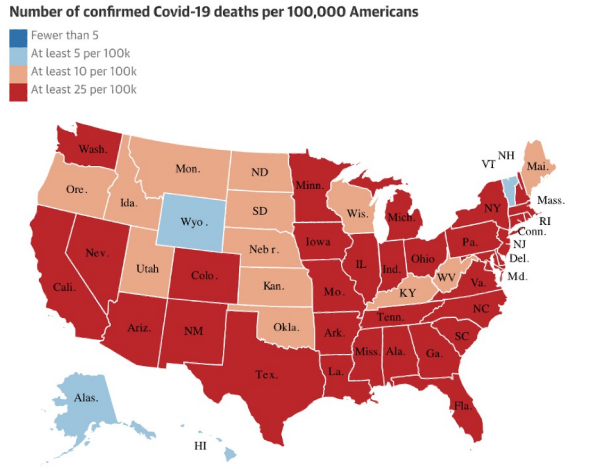
# Marks

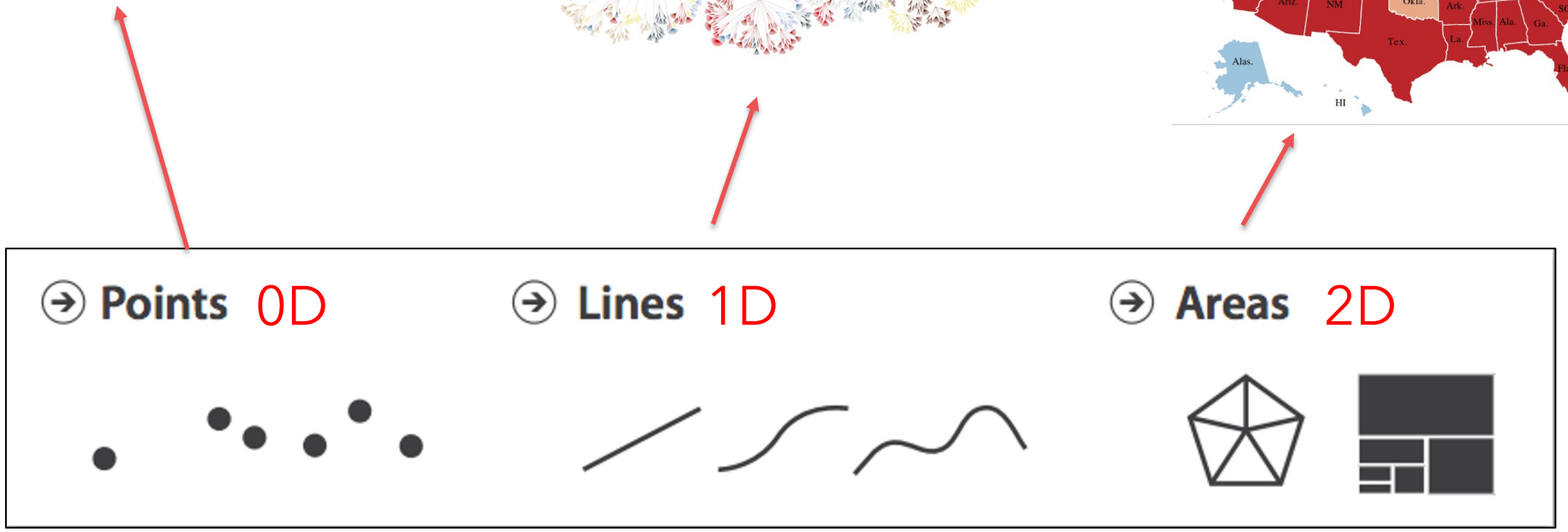
- **Marks** ae the building blocks of any type of visualization
- They are **geometric primitive objects** which are classified according to the number of their spatial dimensions



Ground living area partially explains sale price of apartments
Sale price (k$)
600
400
200
0
1000
2000
3000
4000
5000
Ground living area




Number of confirmed Covid-19 deaths per 100,000 Americans
Fewer than 5
At least 5 per 100k
At least 10 per 100k
At least 25 per 100k
Wash.
Ore.
Cali.
Nev.
Ida.
Mon.
Wyo.
Utah
Ariz.
Colo.
NM
ND
SD
Nebr.
Kan.
Okla.
Tex.
Minn.
Iowa
Mo.
Ark.
La.
Wis.
Ill.
Mich.
Ind.
Ohio
KY
Tenn.
Miss.
Ala.
Ga.
Fla.
SC
NC
Va.
WV
Pa.
NY
VT
NH
Mai.
Mass.
RI
Conn.
NJ
Del.
Md.
Alas.
HI


Points 0D
Lines 1D
Areas 2D

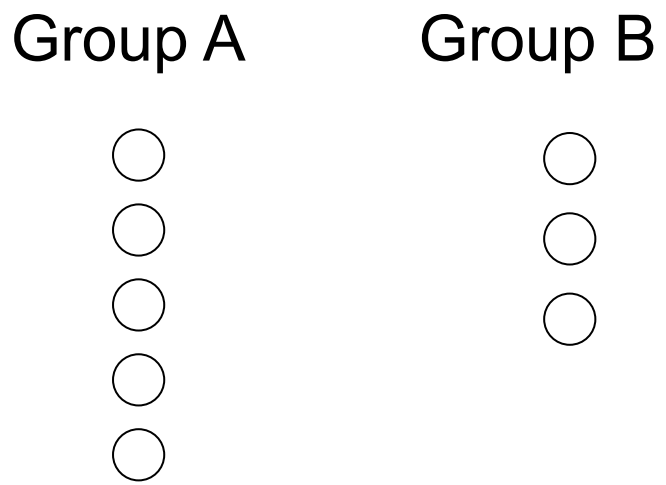
# Visual Channels

We use some visual properties of marks, that we call visual channels, to encode data.

- Position
- Color
- Shape
- Size
- Angle
- Motion
- Text
- …

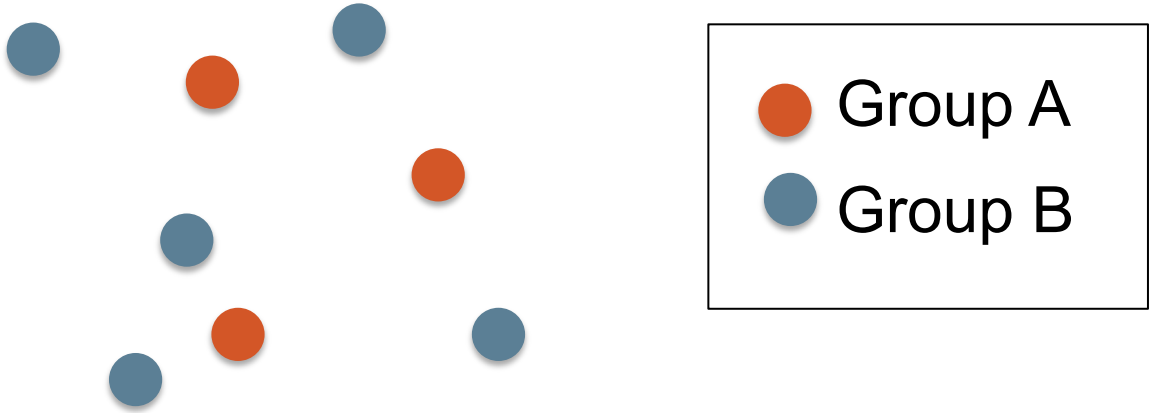
# Visual Channels

- **Position**
- Color
- Shape
- Size
- Angle
- Motion
- Text
- …



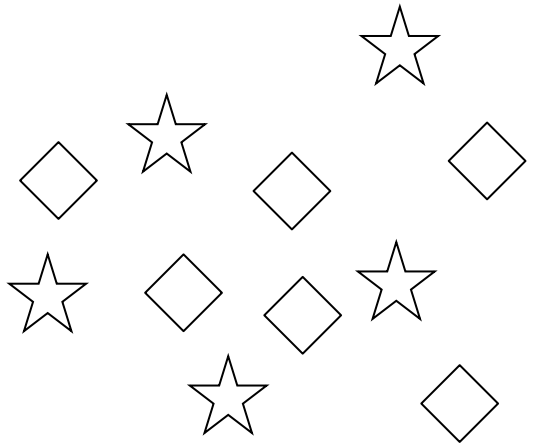
# Visual Channels

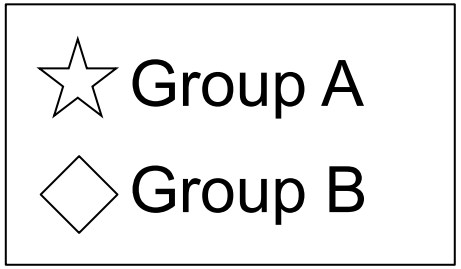
- Position
- **Color hue**
- Shape
- Size
- Angle
- Motion
- Text
- …



# Visual Channels

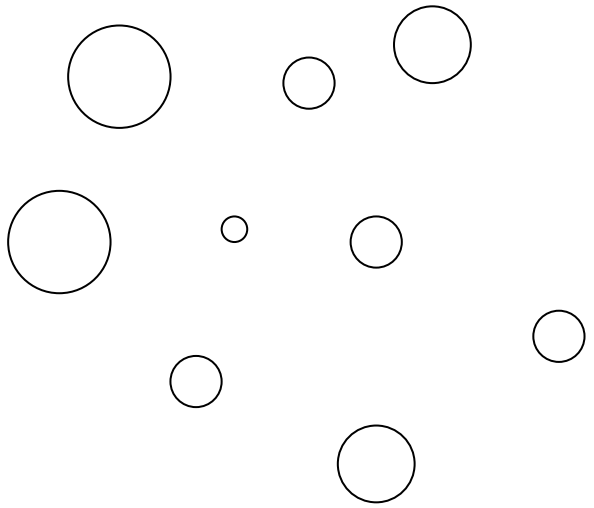
- Position
- Color
- **Shape**
- Size
- Angle
- Motion
- Text
- …



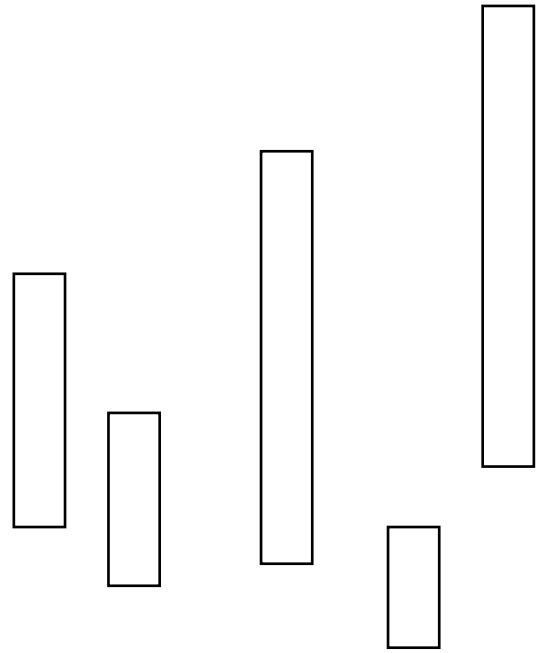


# Visual Channels

- Position
- Color
- Shape
- **Size**
- Angle
- Motion
- Text
- …



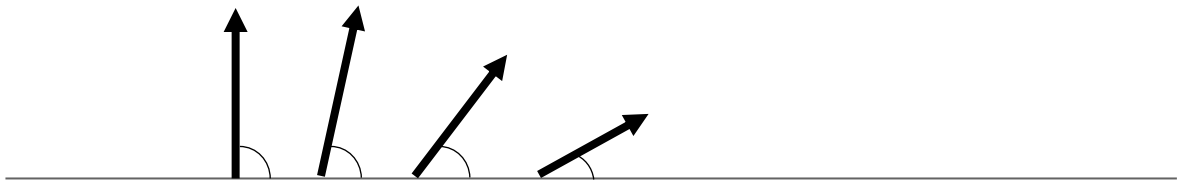
Area



Height

# Visual Channels

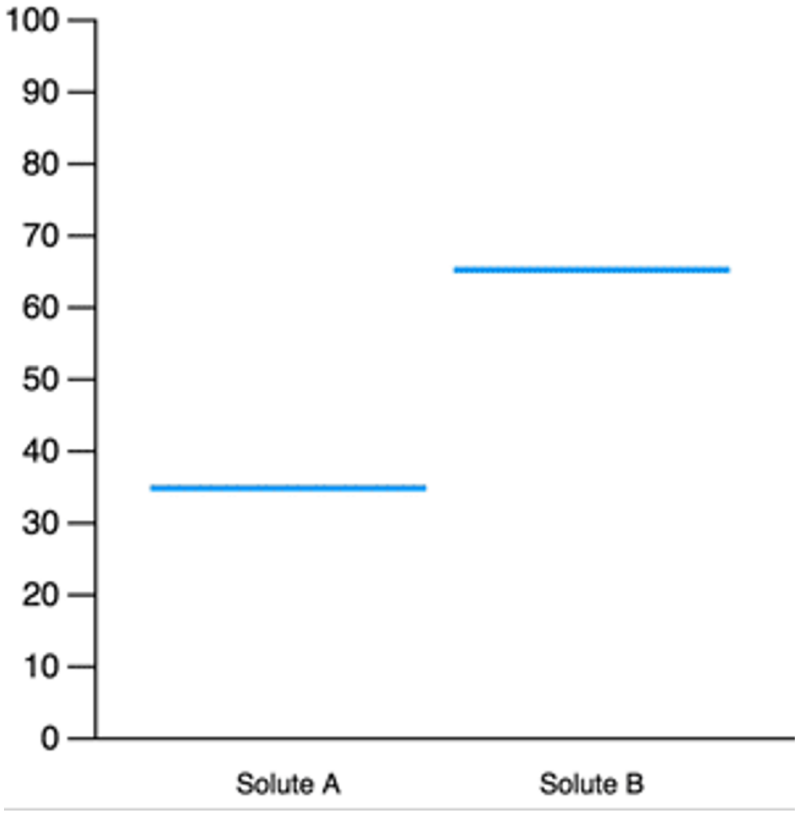
- Position
- Color
- Shape
- Size
- **Angle**
- Motion
- Text
- …



Wind direction

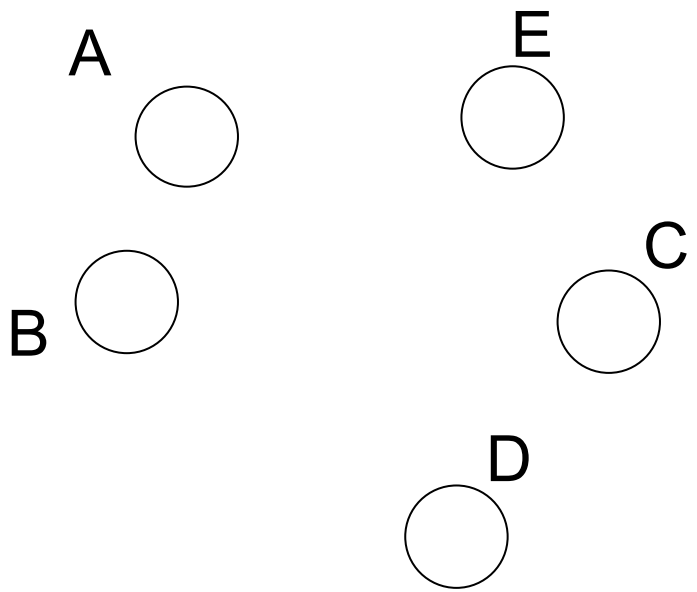
# Visual Channels

- Position
- Color
- Shape
- Size
- Angle
- **Motion**
- Text…
- …



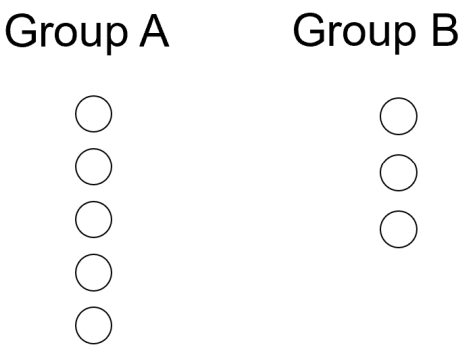
# Visual Channels

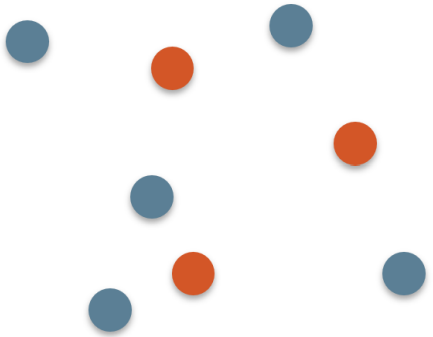
- Position
- Color
- Shape
- Size
- Angle
- Motion
- **Text**
- **…**

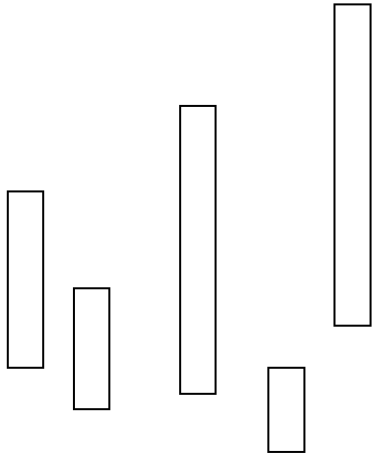


# Visual Encoding

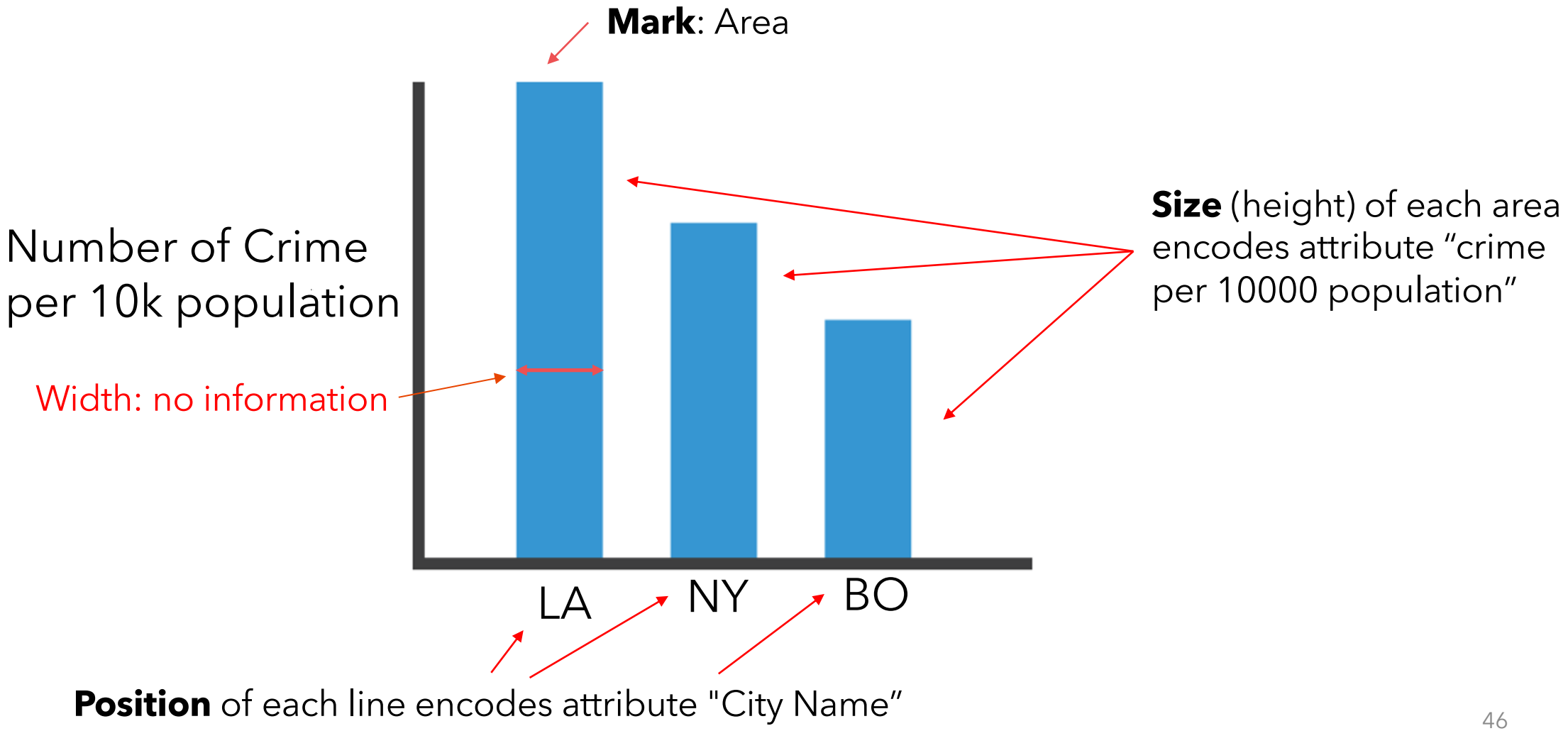
The process of mapping a mark's certain property to a data value is known as visual encoding



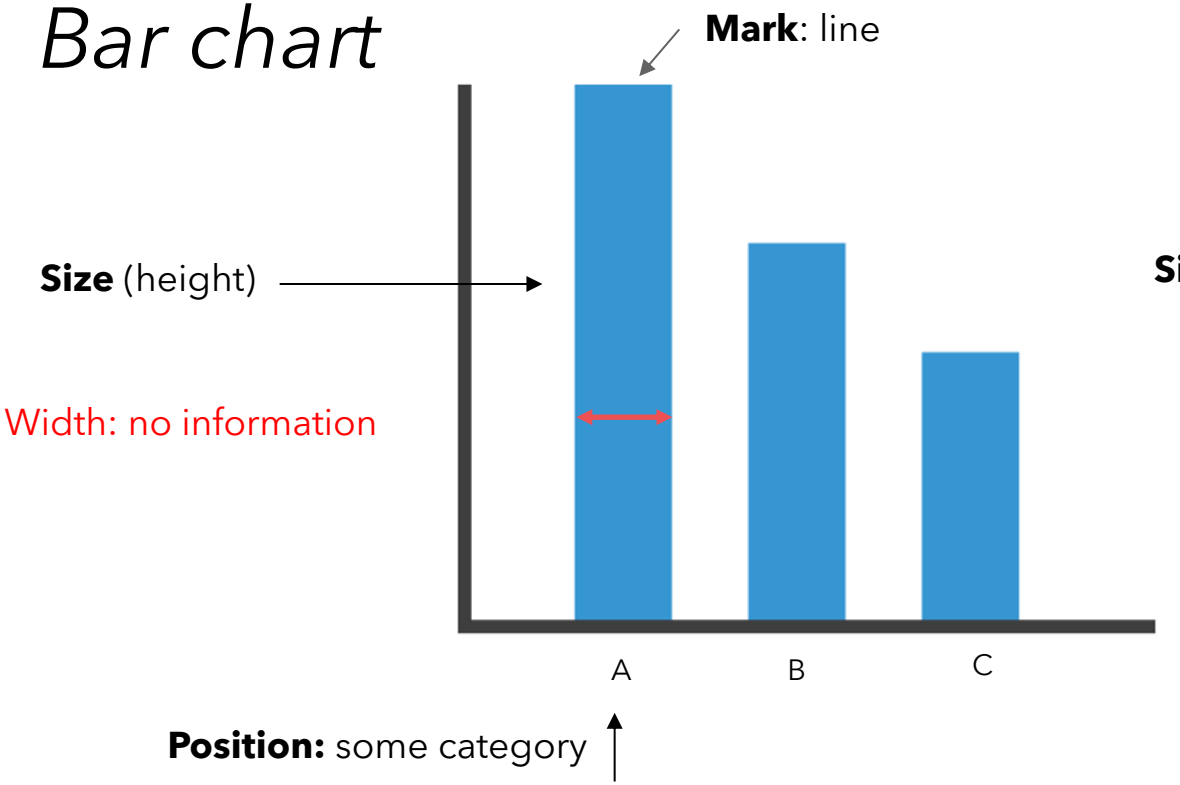


# Visual Encoding

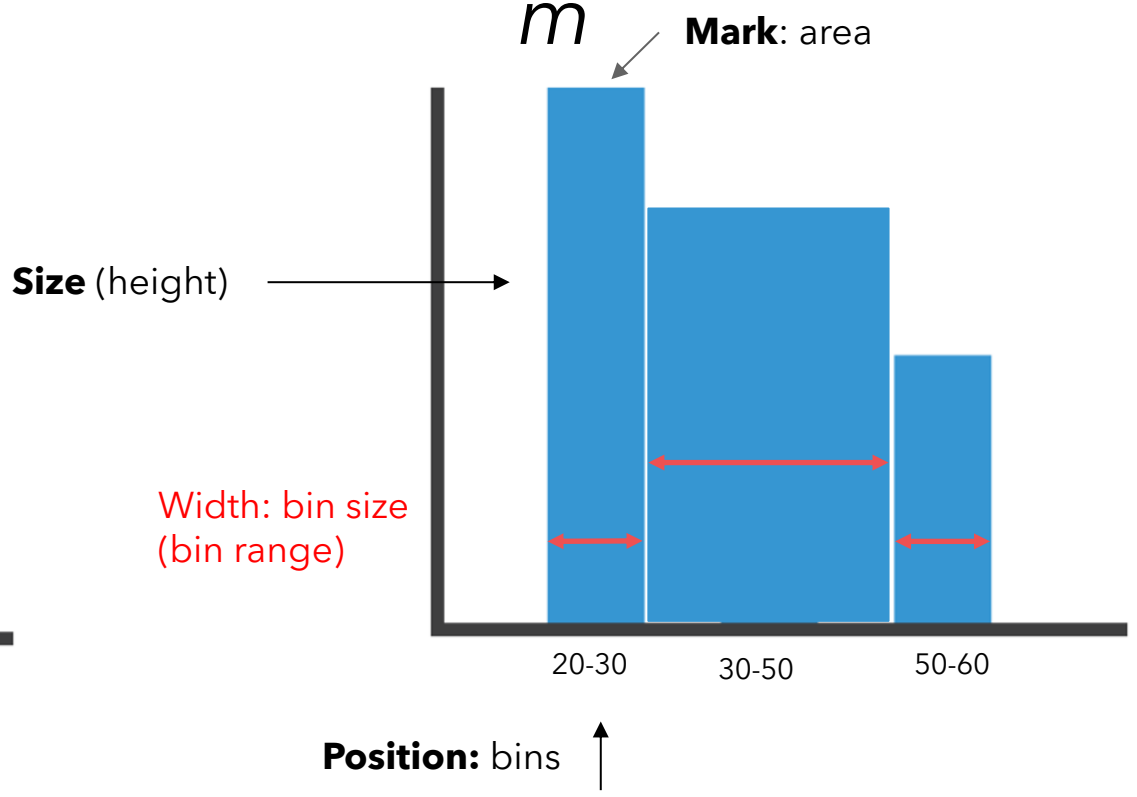


# Visual Encoding

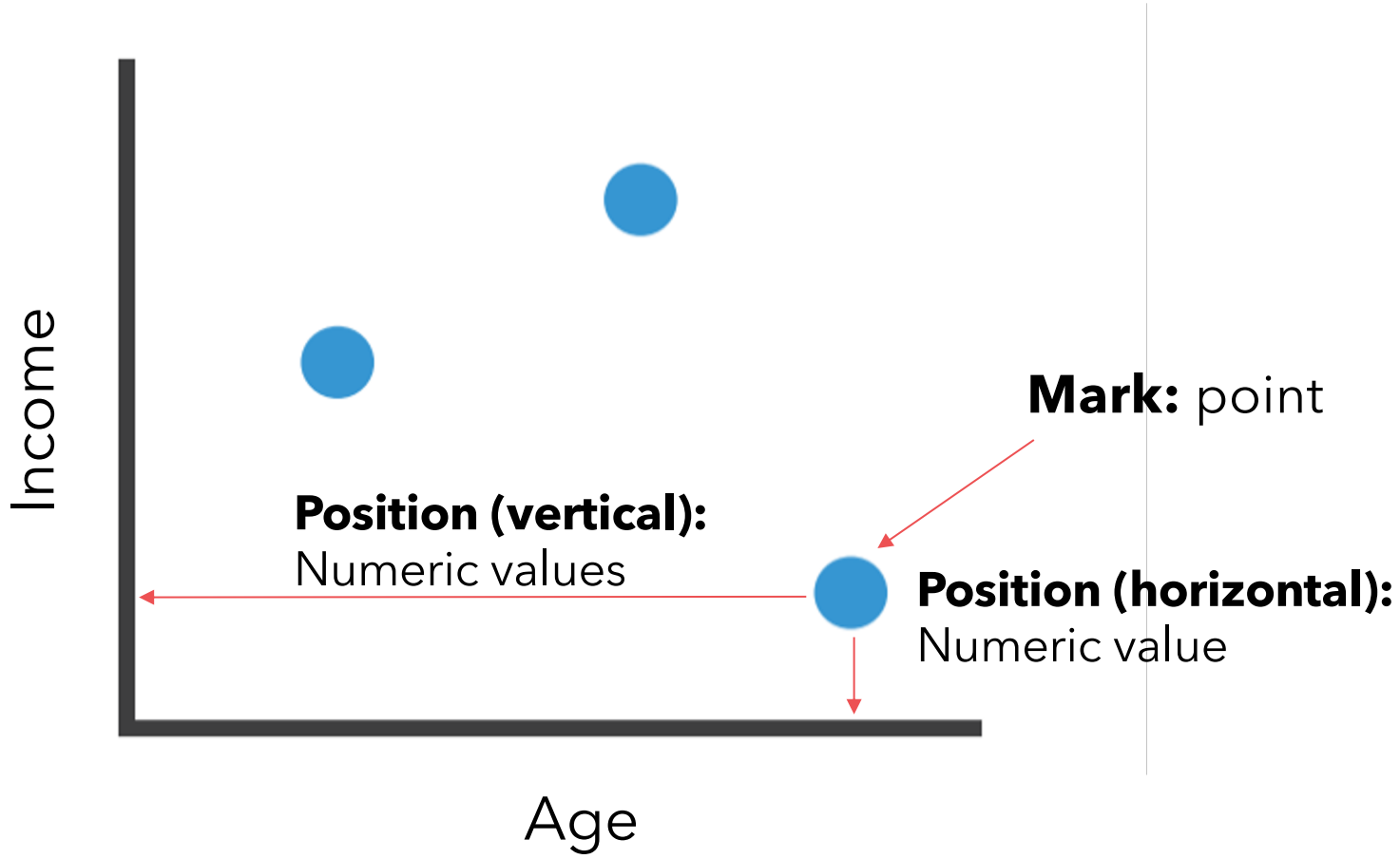
*Bar chart*



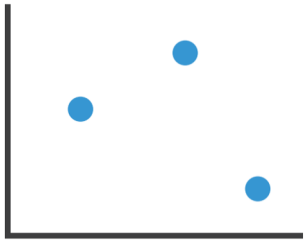

*Histogram*



# Visual Encoding

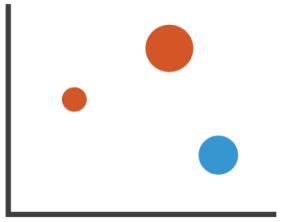


# Visual Encoding



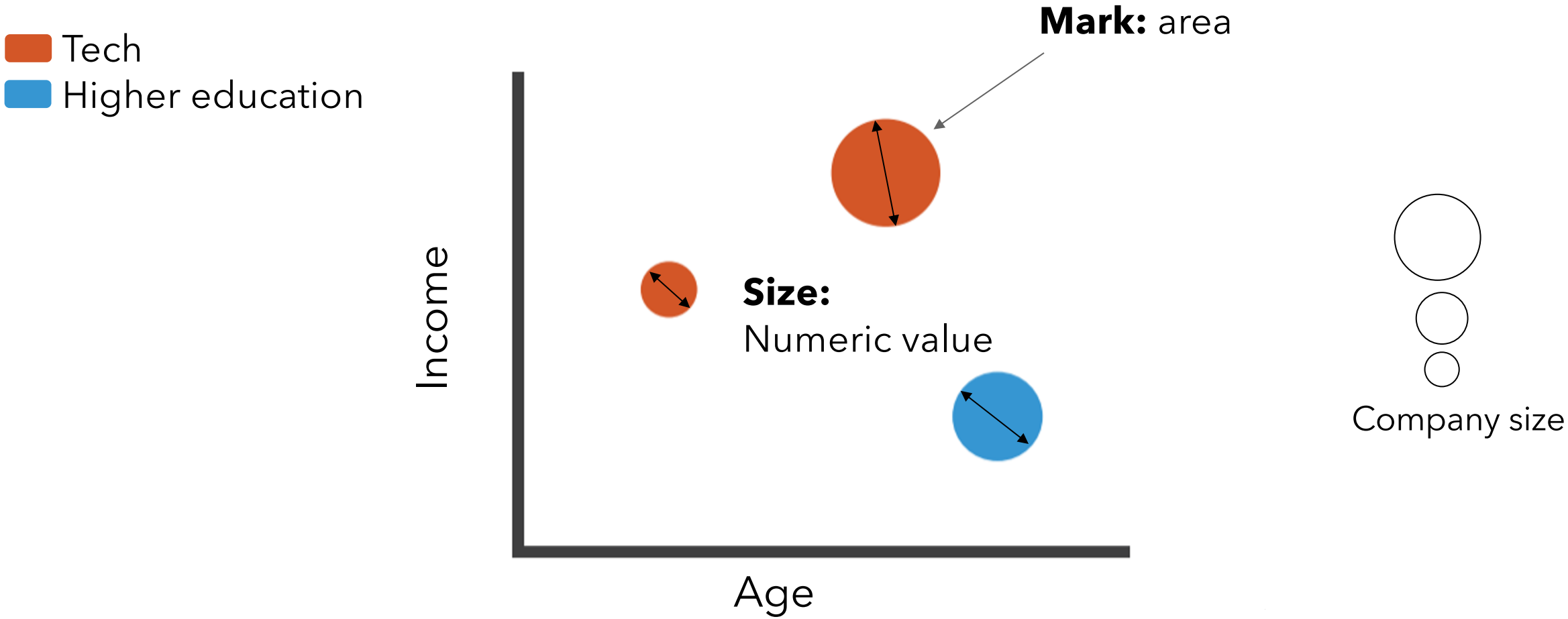
Position (X & Y) — 2 channels

Position (X & Y) + Color — 3 channels



Position (X & Y) + Color + Size — 4 channels

# Visual Encoding

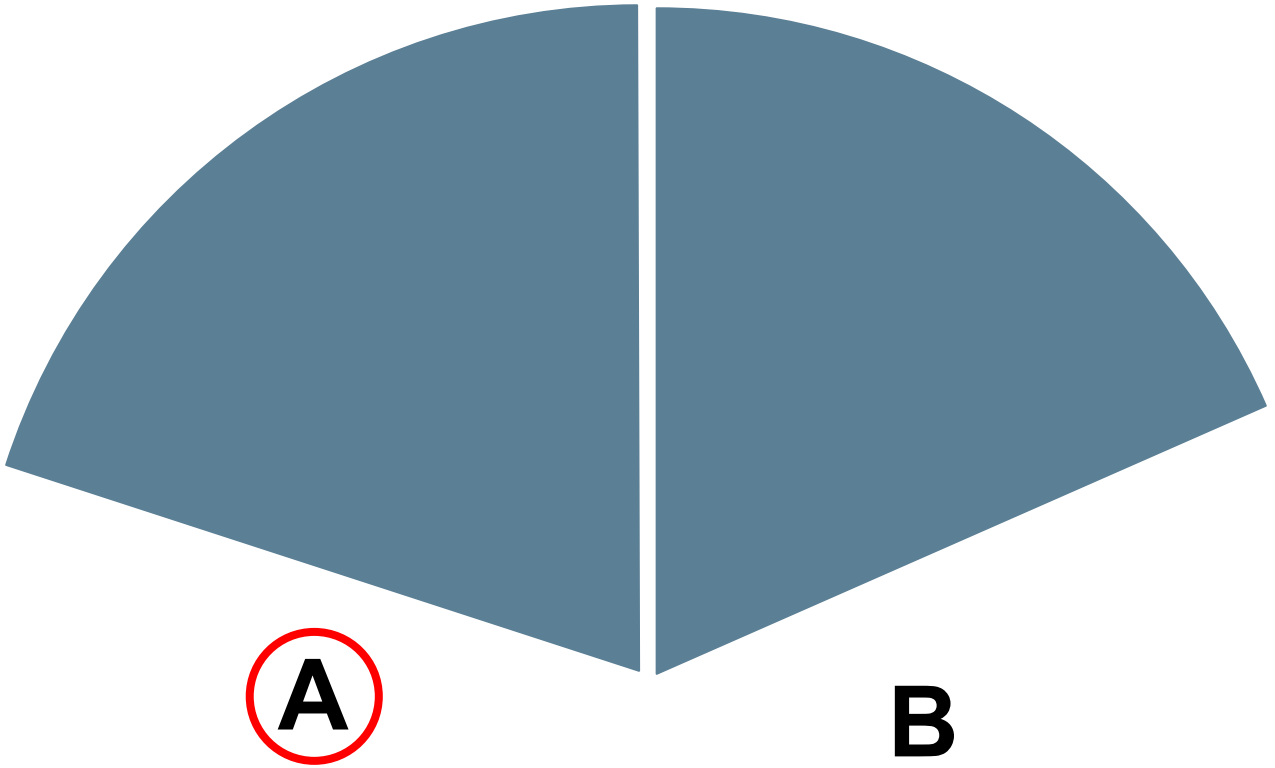


# Visual Encoding

Visual encoding channels vary in terms of:

- Accuracy
- Discriminability
- Separability
- Visual Popout
- Support for perceptual grouping
- …

A
B


A
B

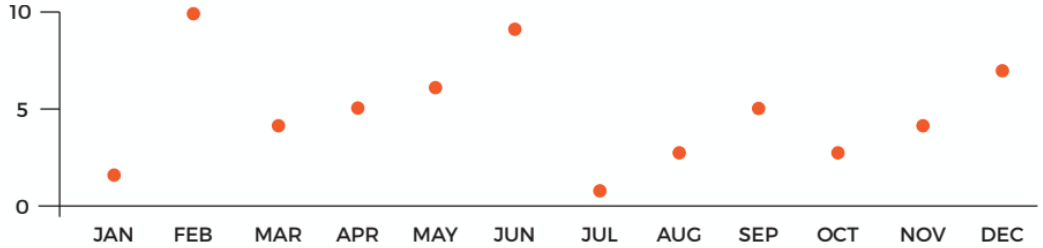
A
B

# How to choose effective mapping?

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

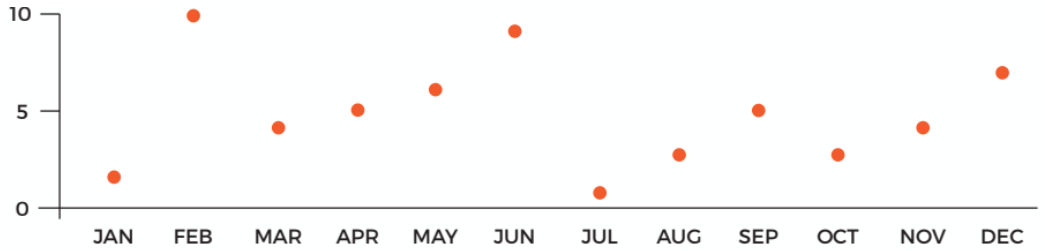
| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

Dot plot



| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

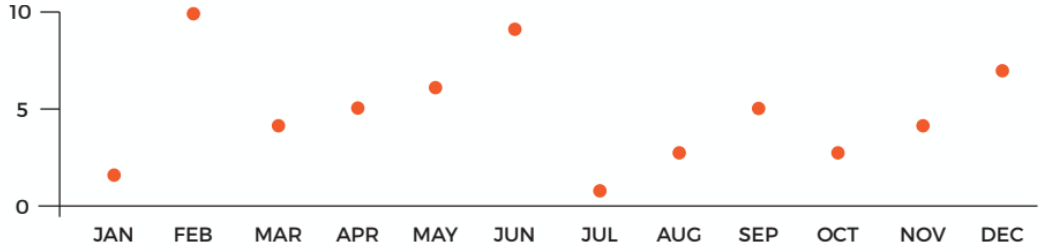
Dot plot




Line

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

## Dot plot

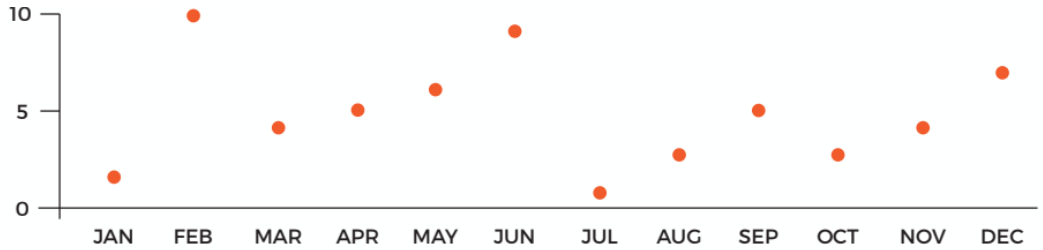



## Line

## Aligned Bars

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

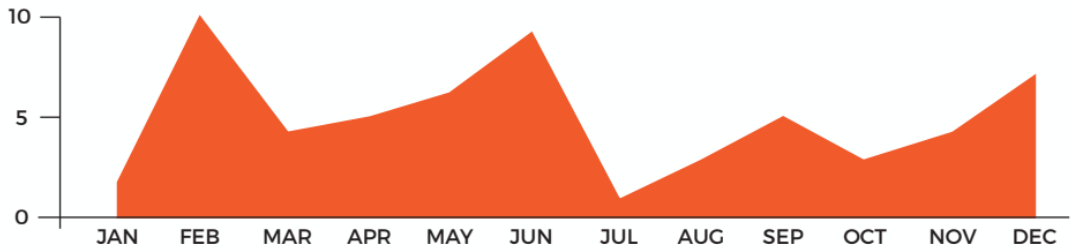
## Dot plot




## Line

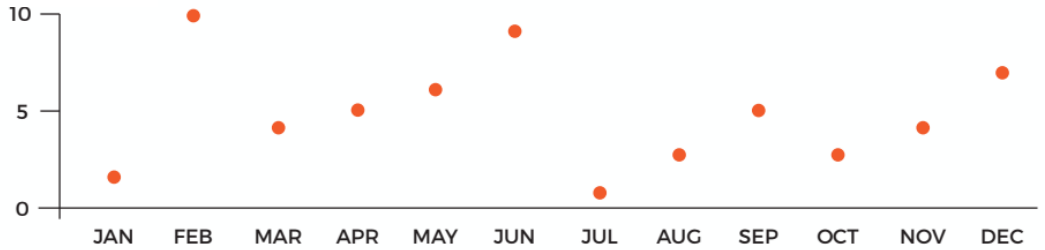
## Aligned Bars

## Area



| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

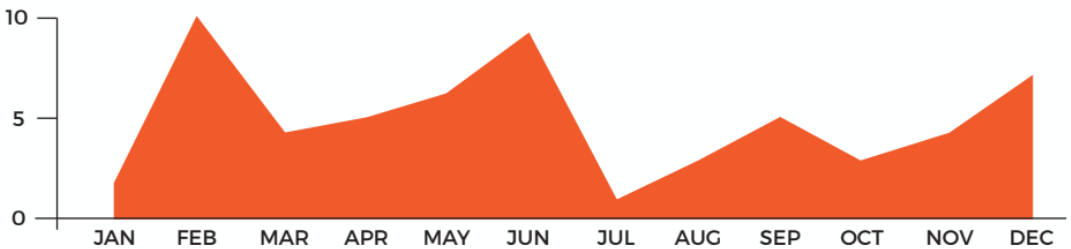
## Dot plot




## Line

## Aligned Bars

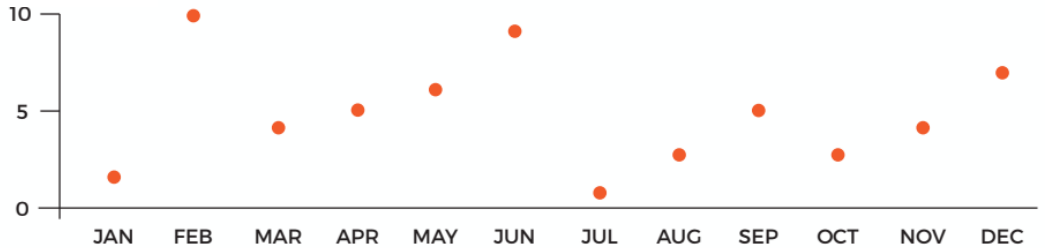
## Area




## Size

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

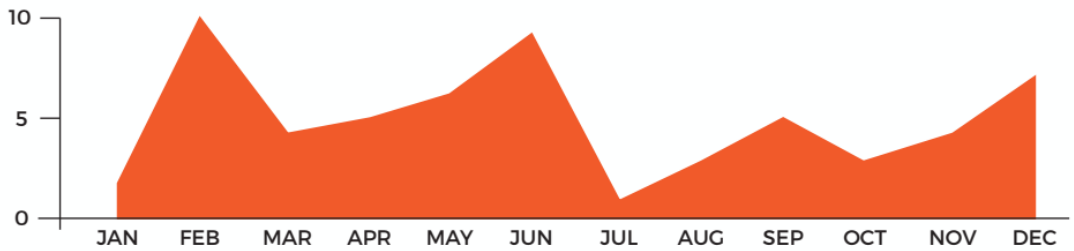
## Dot plot




## Line

## Aligned Bars

## Area

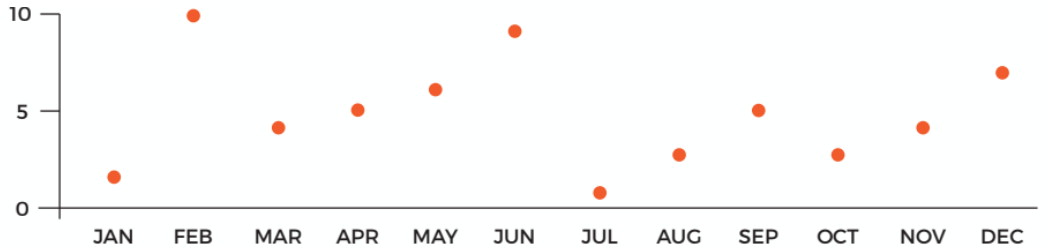



## Size

## Saturation

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

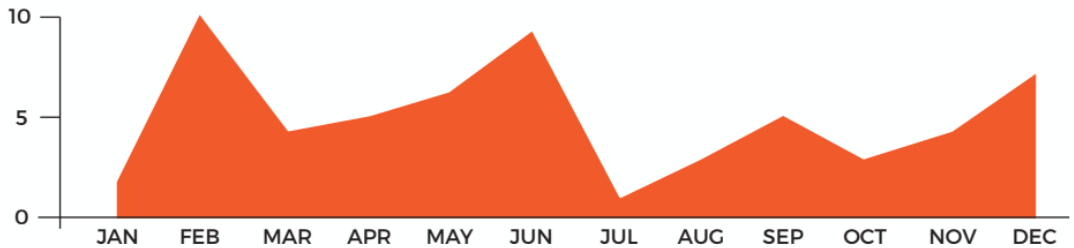
## Dot plot




## Line

## Aligned Bars

## Area



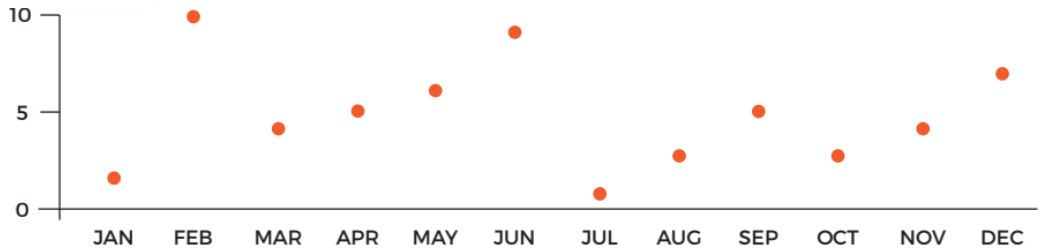

## Size

## Saturation

## Size & Saturation

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

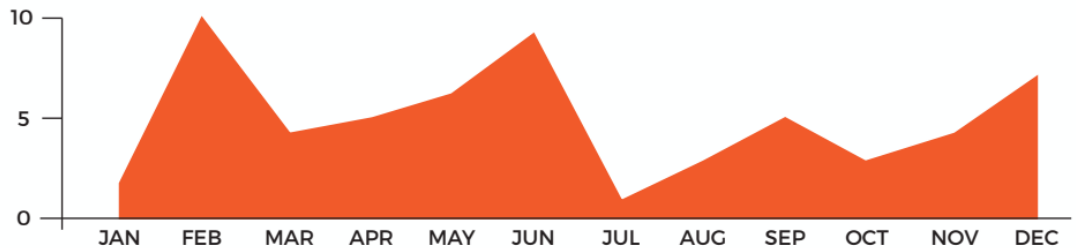
## Dot plot




## Line

## Aligned Bars

## Area



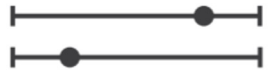

## Size

## Saturation

## Size & Saturation

## Size, Saturation, & Position

Position on common scale

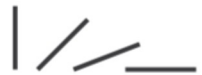


Position on unaligned scale

Length (1D size)



Tilt/angle



Area (2D size)

Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)

Spatial region

Color hue

Motion

Shape

## Channels: Expressiveness Types and Effectiveness Ranks

### Magnitude Channels

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Most

Effectiveness

Least

Same

Same

### Identity Channels

- Spatial region
- Color hue
- Motion
- Shape

**Channels:** Expressiveness Types and Effectiveness Ranks

## Magnitude Channels

Position on common scale

Position on unaligned scale

Length (1D size)

Tilt/angle

Area (2D size)

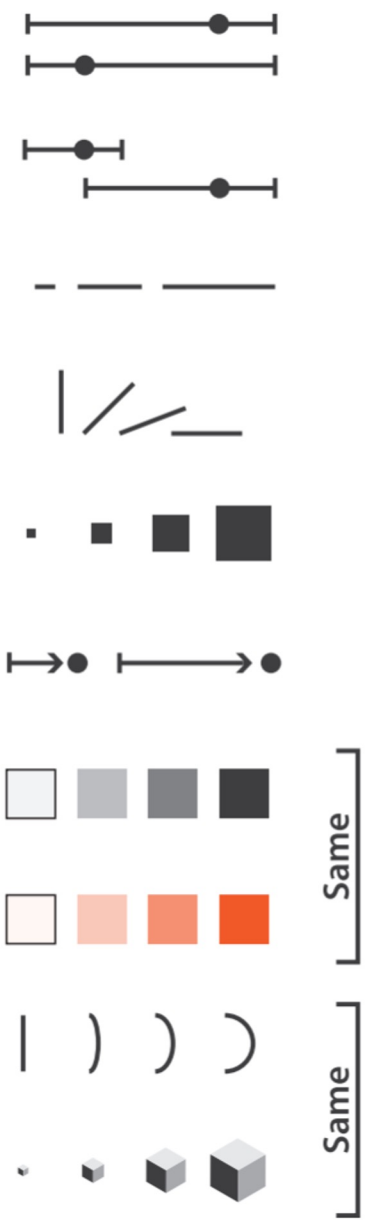
Depth (3D position)

Color luminance

Color saturation

Curvature

Volume (3D size)




## Identity Channels

Spatial region

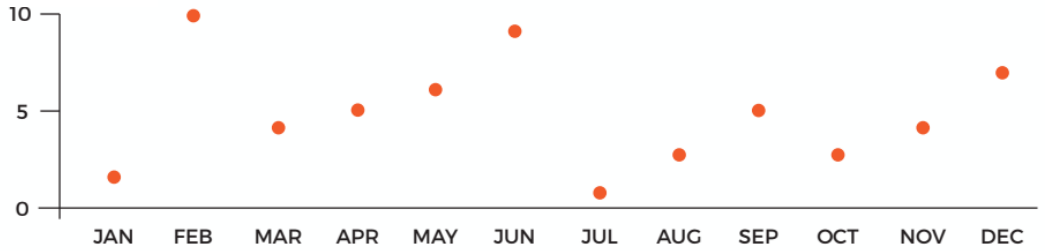
Color hue

Motion

Shape

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

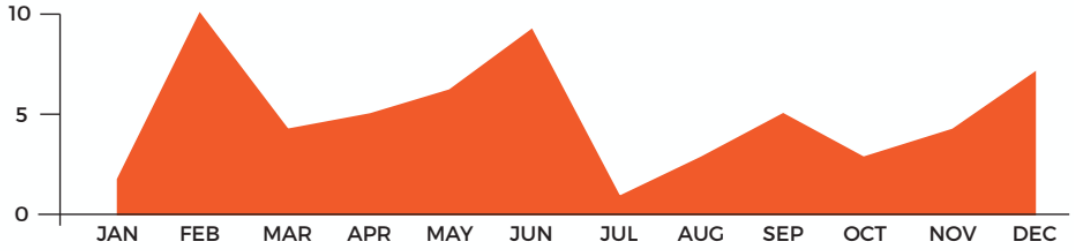
## Dot plot




## Line

## Aligned Bars

## Area




## Size

## Saturation

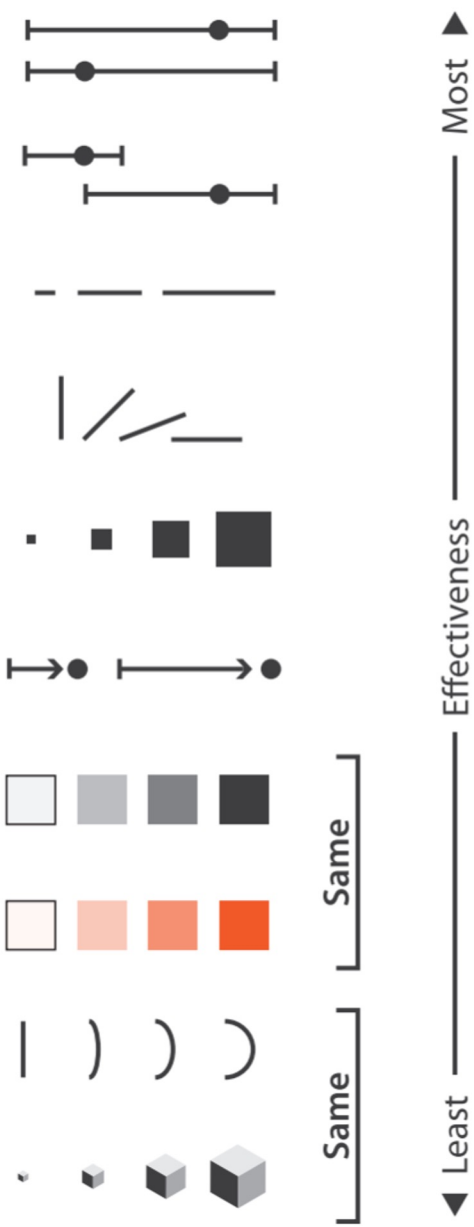
## Size & Saturation

## Size, Saturation, & Position

# **Channels:** Expressiveness Types and Effectiveness Ranks

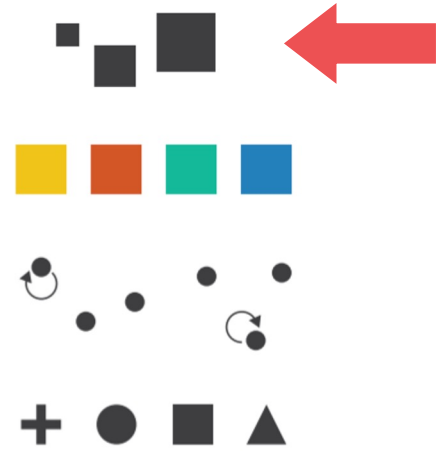
## Magnitude Channels

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



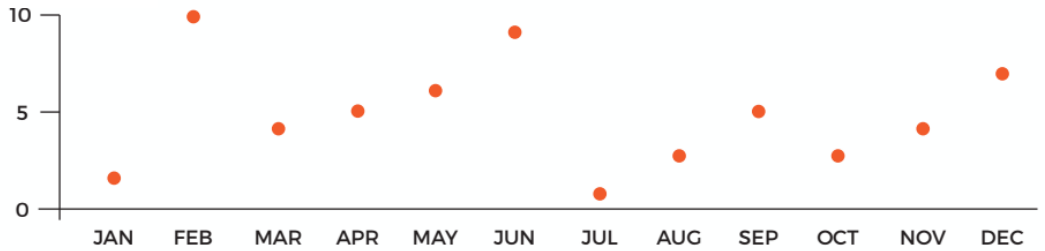

## Identity Channels

- Spatial region
- Color hue
- Motion
- Shape



| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

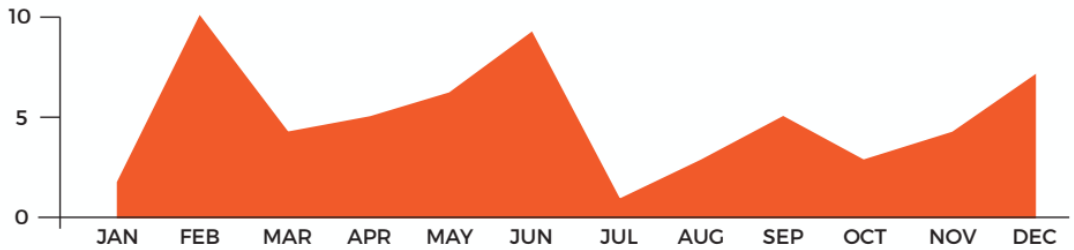
## Dot plot




## Line

## Aligned Bars

## Area




## Size

## Saturation

## Size & Saturation

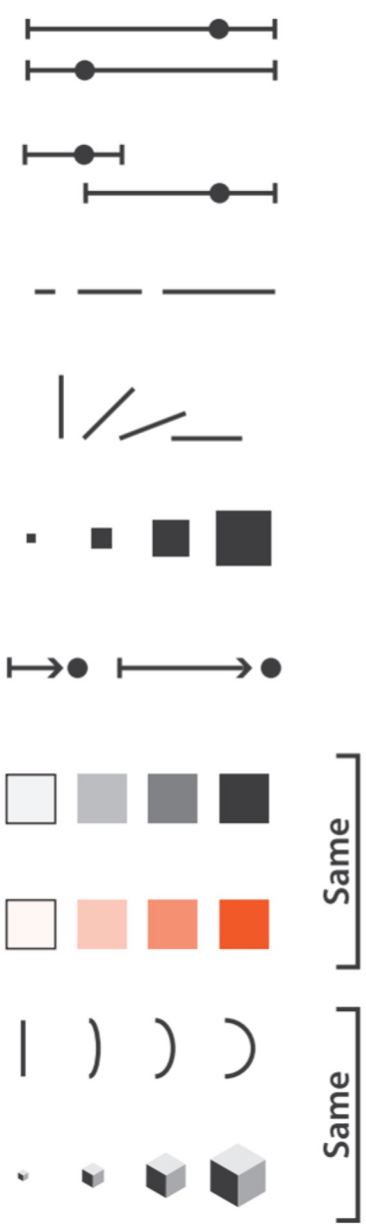
## Size, Saturation, & Position

Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021

## **Channels:** Expressiveness Types and Effectiveness Ranks

### Magnitude Channels

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



Most

Effectiveness

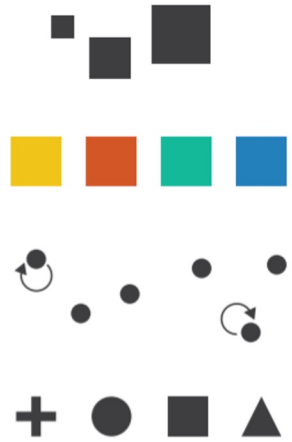
Least

Same

Same

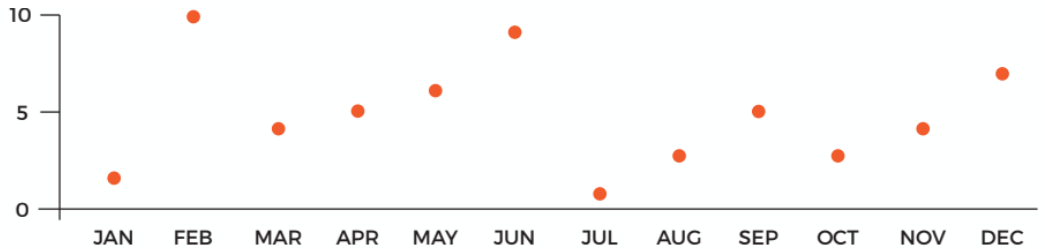
### Identity Channels

- Spatial region
- Color hue
- Motion
- Shape



| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

## Dot plot

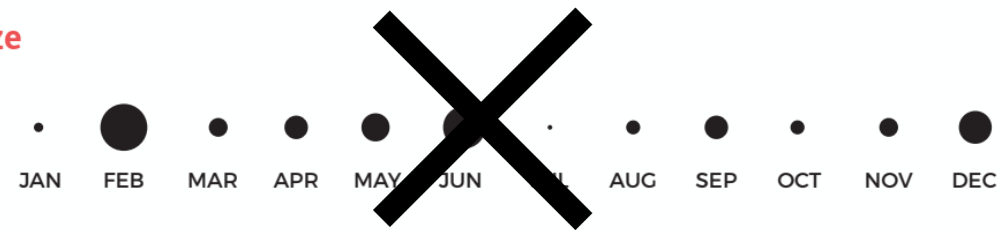



## Line

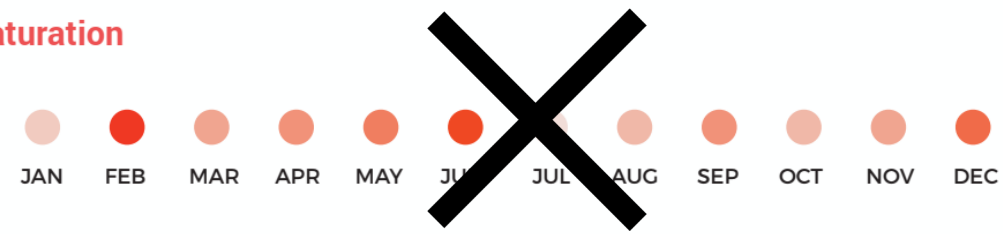
## Aligned Bars

## Area

## Size




## Saturation




## Size & Saturation

## Size, Saturation, & Position

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

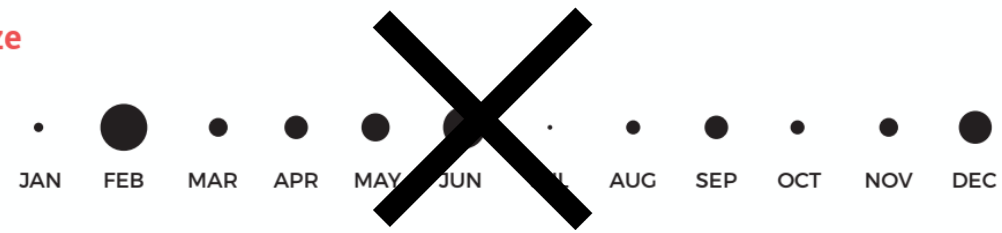
## Dot plot

## Line

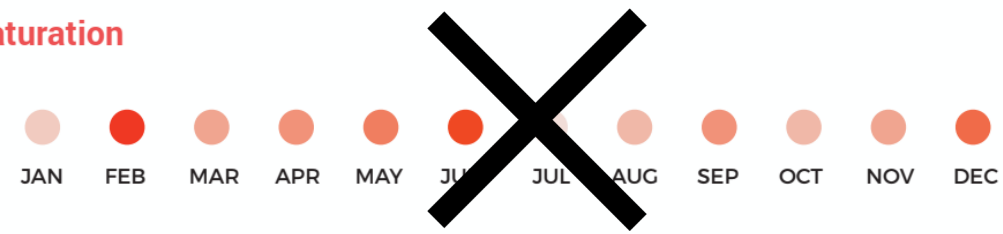
## Aligned Bars

## Area

## Size




## Saturation




## Size & Saturation

## Size, Saturation, & Position

| JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 10 | 4 | 5 | 6 | 9 | 1 | 3 | 5 | 3 | 4 | 7 |

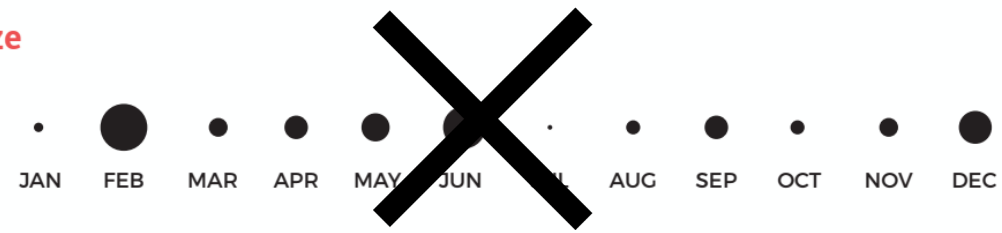
## Dot plot

## Line

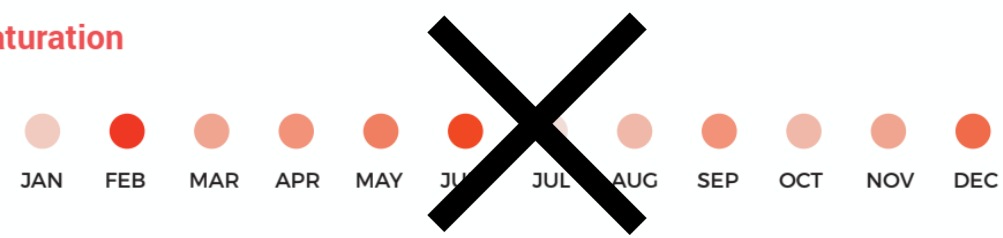
## Aligned Bars

## Area

## Size



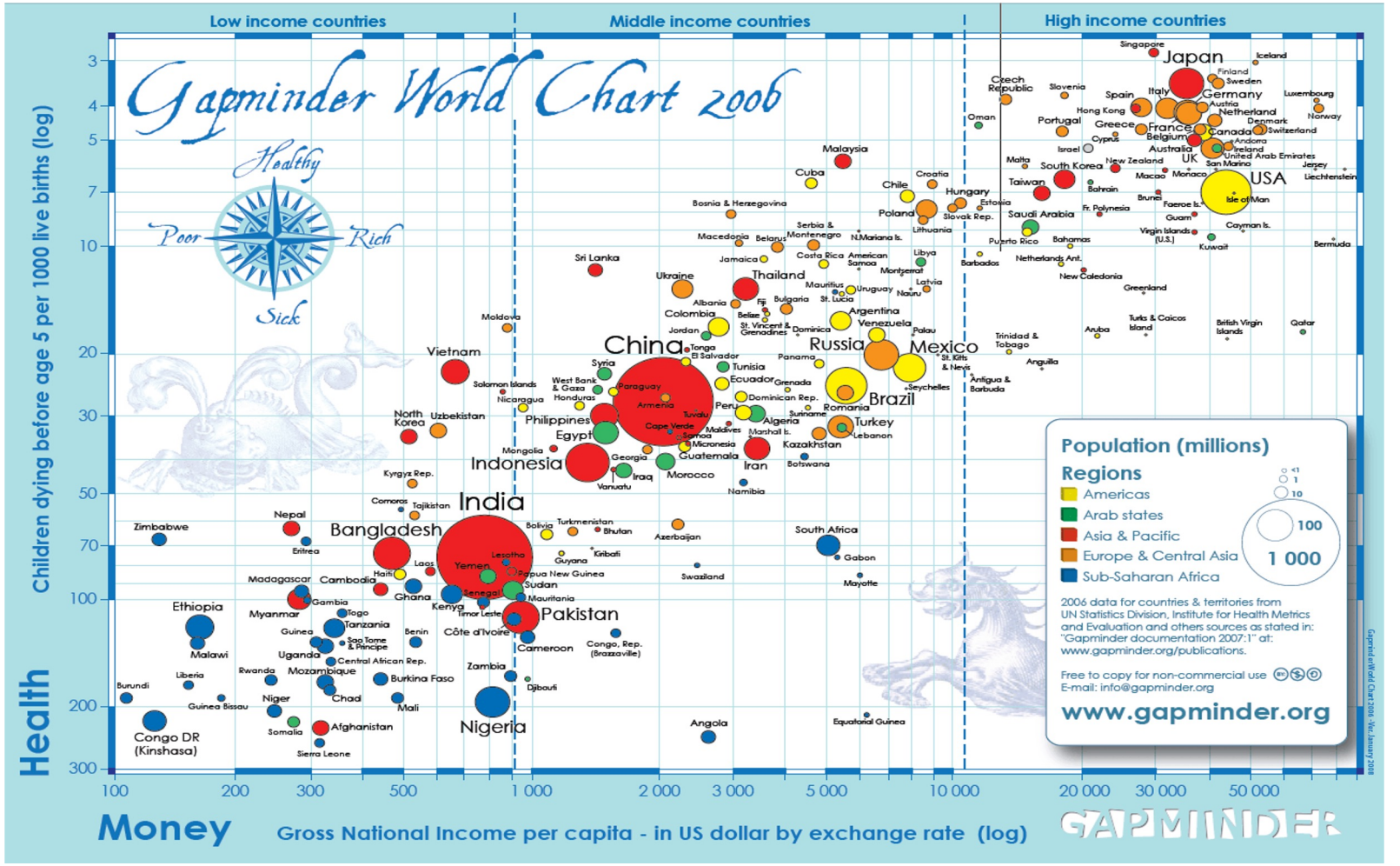

## Saturation




## Size & Saturation

## Size, Saturation, & Position

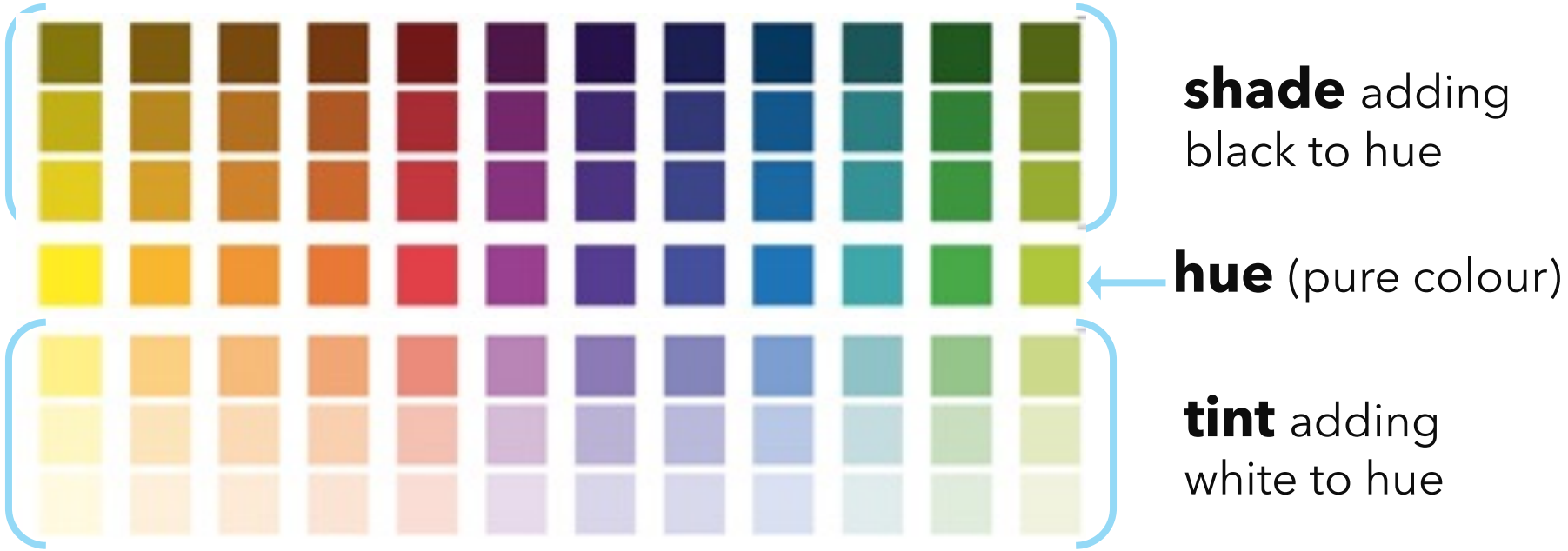
# Activity




https://tinyurl.com/y9s9z22a

# Effective use of color

# Very basic color theory




even though this shows discrete colors
. . . it's really a continuous gradient (infinite colors from white to black)

image from Williams (2008). The non-designer's design book

# Color relationships

- Complementary colors: are a good place to start
  - Work best in as a combination of main color/accent
  - But are hard to read when used as text on background

# Combinations of colors

- Most color combinations can look good together
  - The key is to vary the contrast to enhance readability
  - To increase contrast, you can adjust shade/tint of colors that bump into one another

This is pretty hard to see

This has some contrast

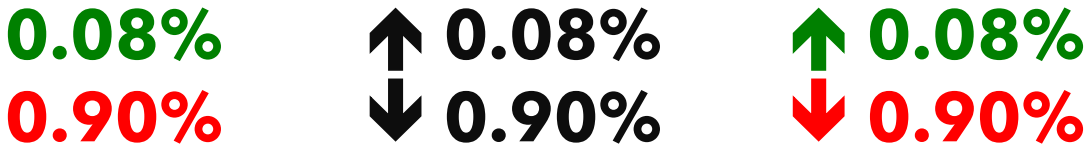
# Guidelines for choosing colors

- Limit your palette – use what you need;
  - Think about what you intend to communicate with color! Make purposeful decisions to:
  - Communicate information or hierarchy
  - Evoke certain emotions or feelings
    - e.g., calm and warmth
    - e.g., edginess and tension
    - e.g., energy and excitement

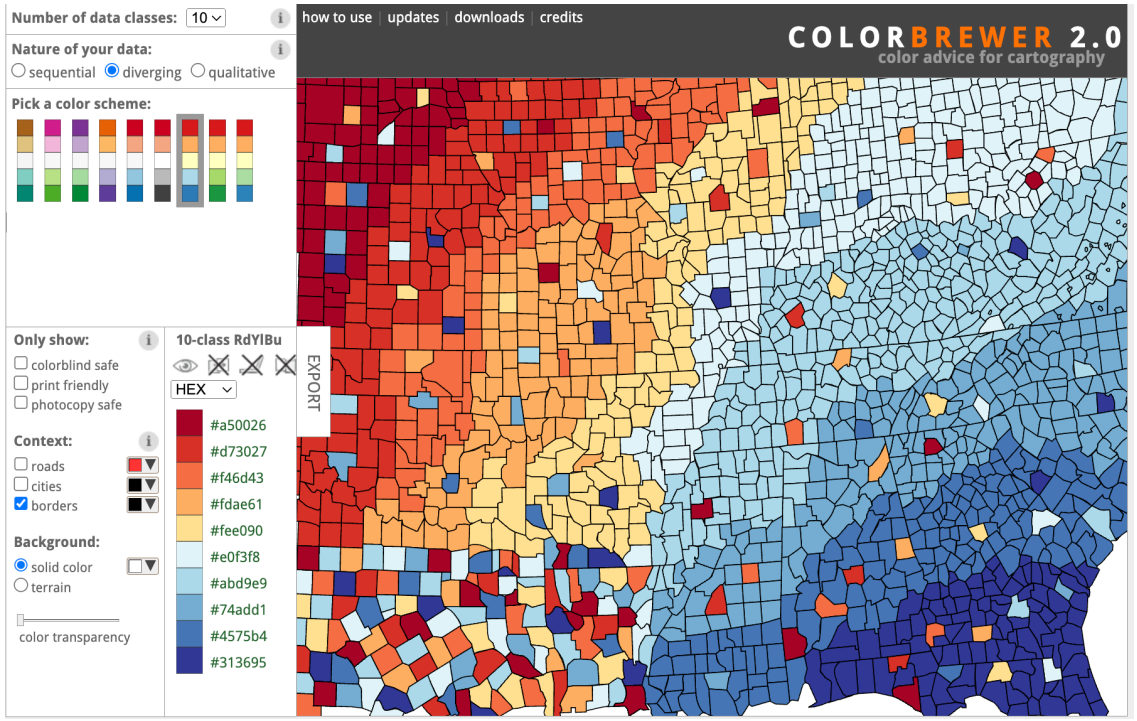
# Guidelines for choosing colors

- Use color-blind safe colors
  - Always for critical information
    - http://colorbrewer2.org
  - When color is the only way to distinguish the content

- Less important to avoid common color-blind combinations if multiple visual cues distinguish same content
  - e.g., showing changes in stock prices



# Choosing colors resources




- Color index books
- Color advice for maps and data: http://colorbrewer2.org/
- Color blind simulators: http://www.vischeck.com/
- Tips on Designing Colorblind-Friendly Visualizations: https://www.tableau.com/blog/examining-data-viz-rules-dont-use-red-green-together

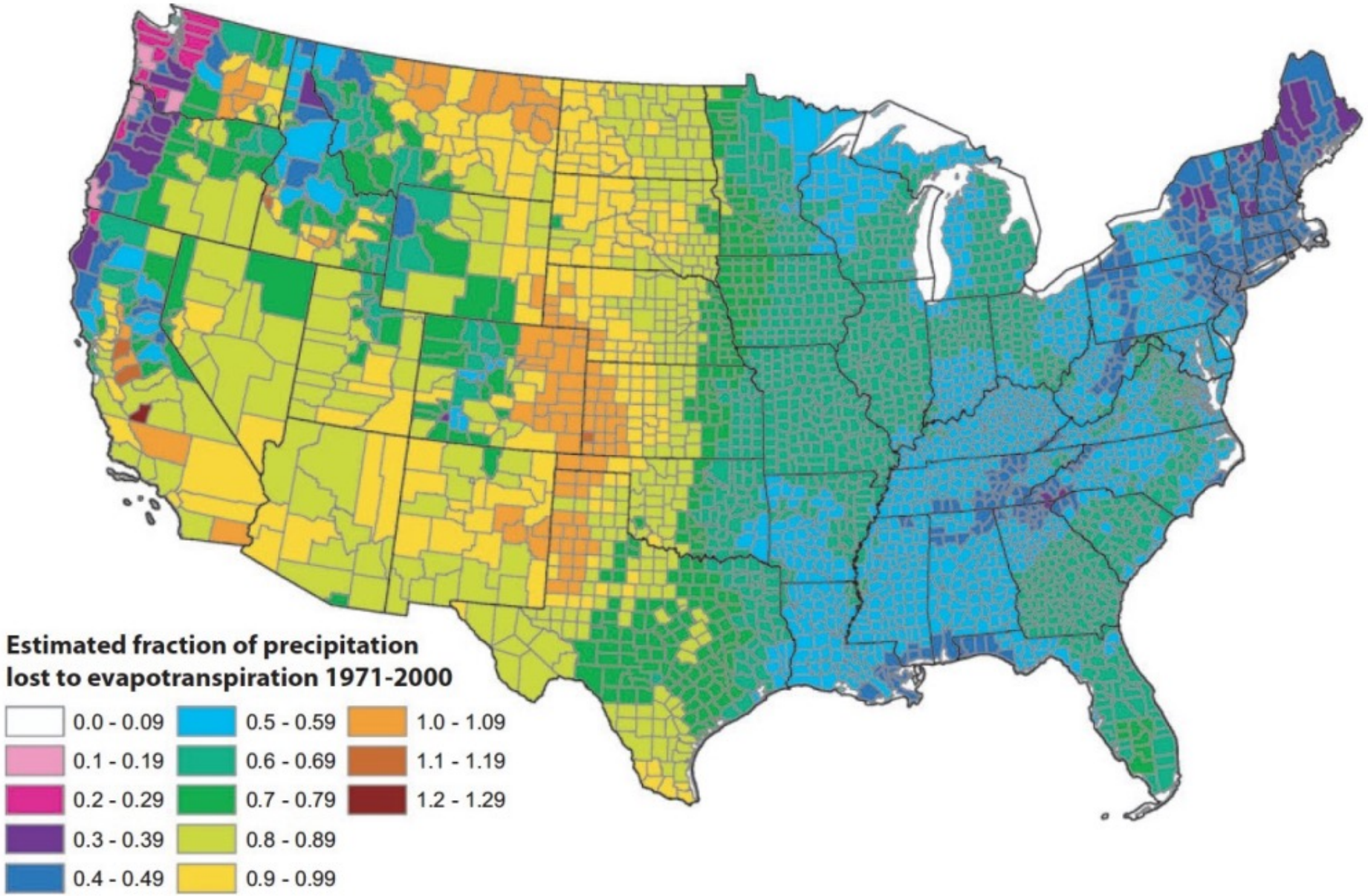



FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ($P$) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of $ET/P$ were made first at the 800-m resolution of the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

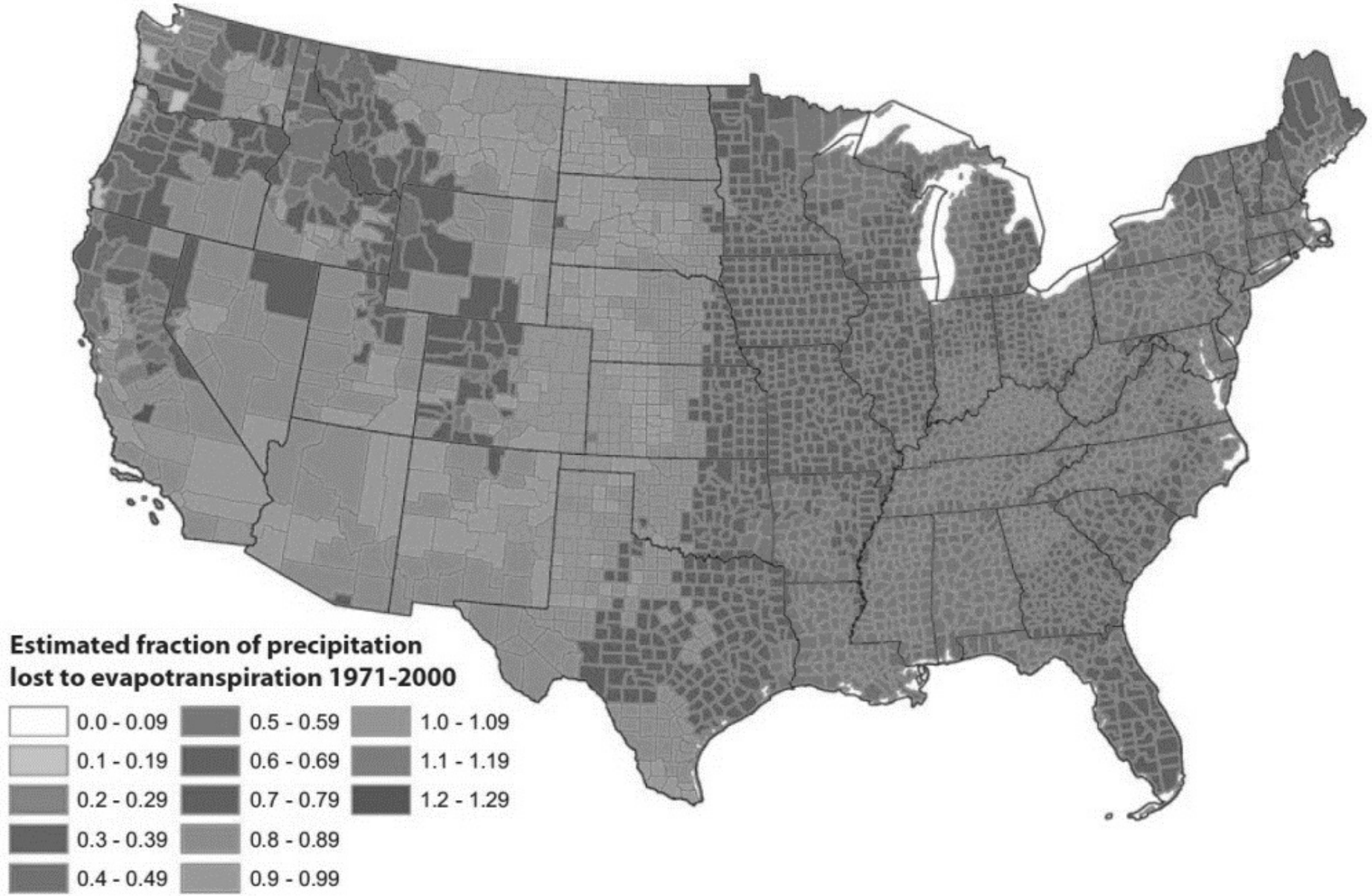
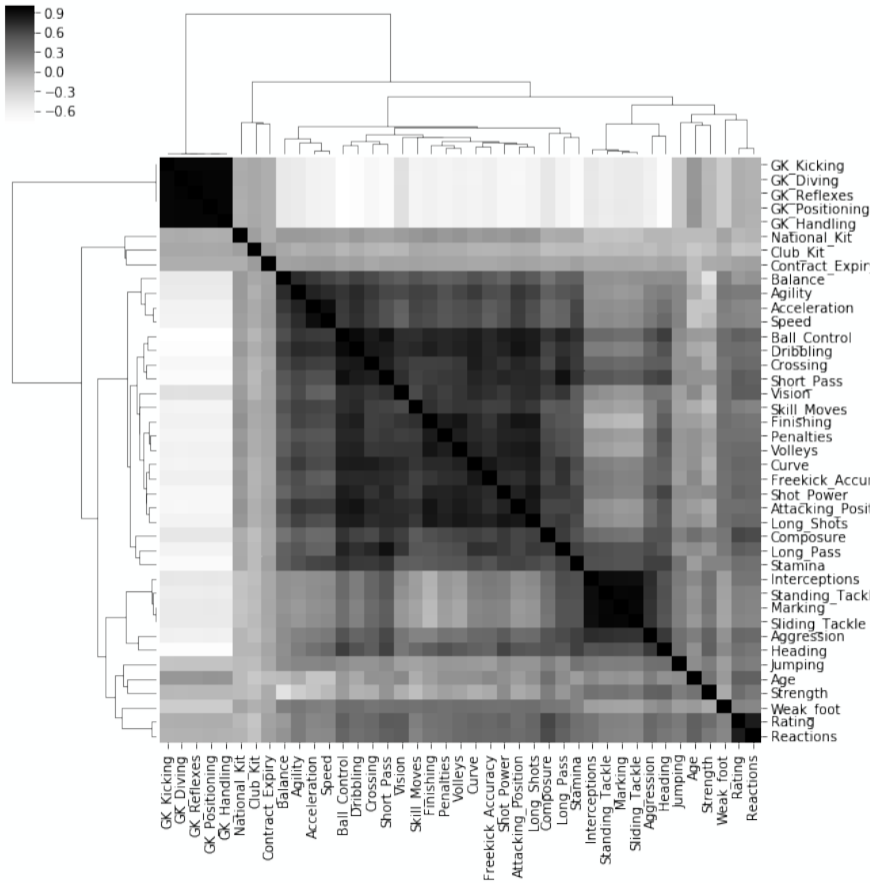
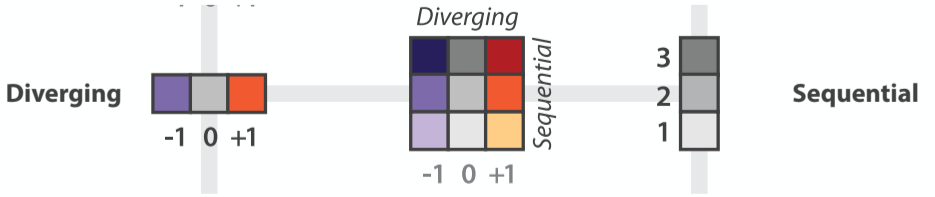
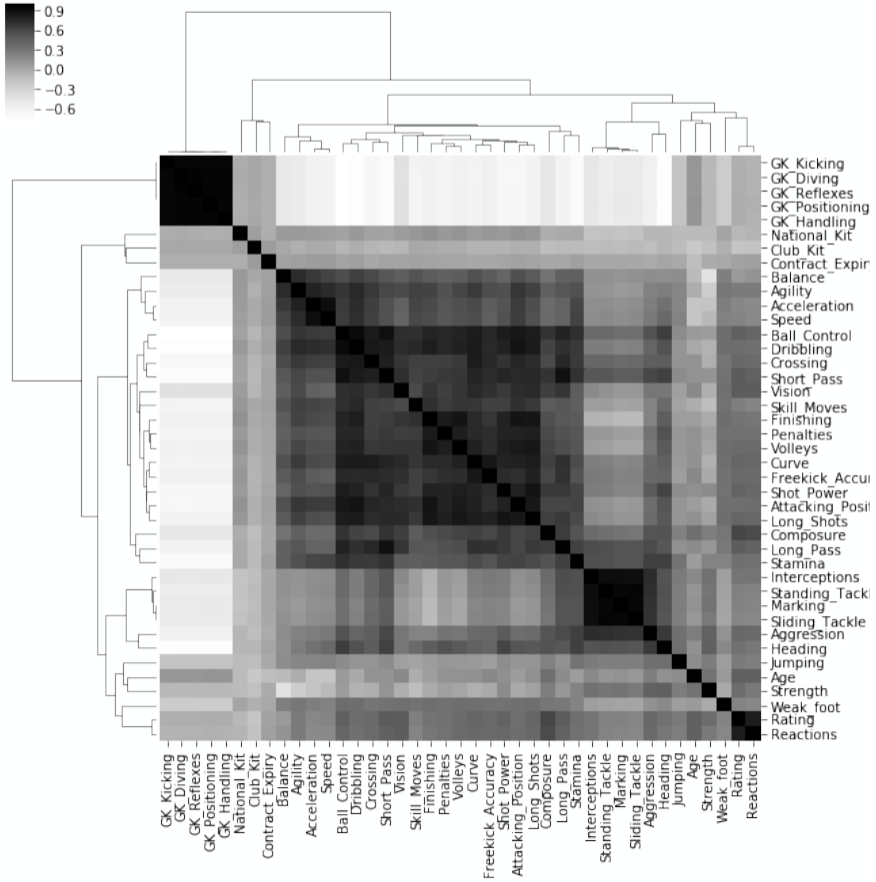



FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ($P$) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of $ET/P$ were made first at the 800-m resolution of the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

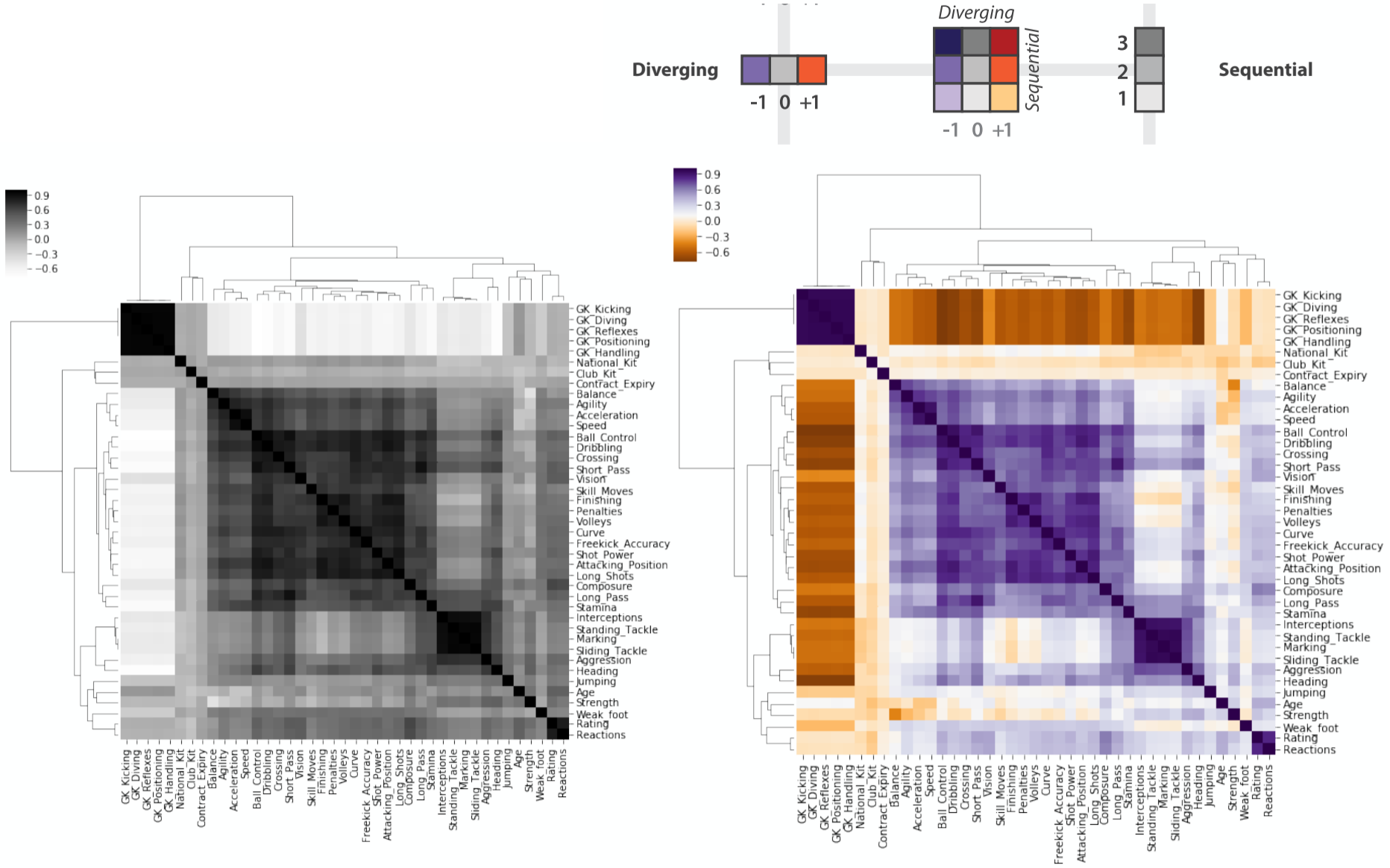



Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021



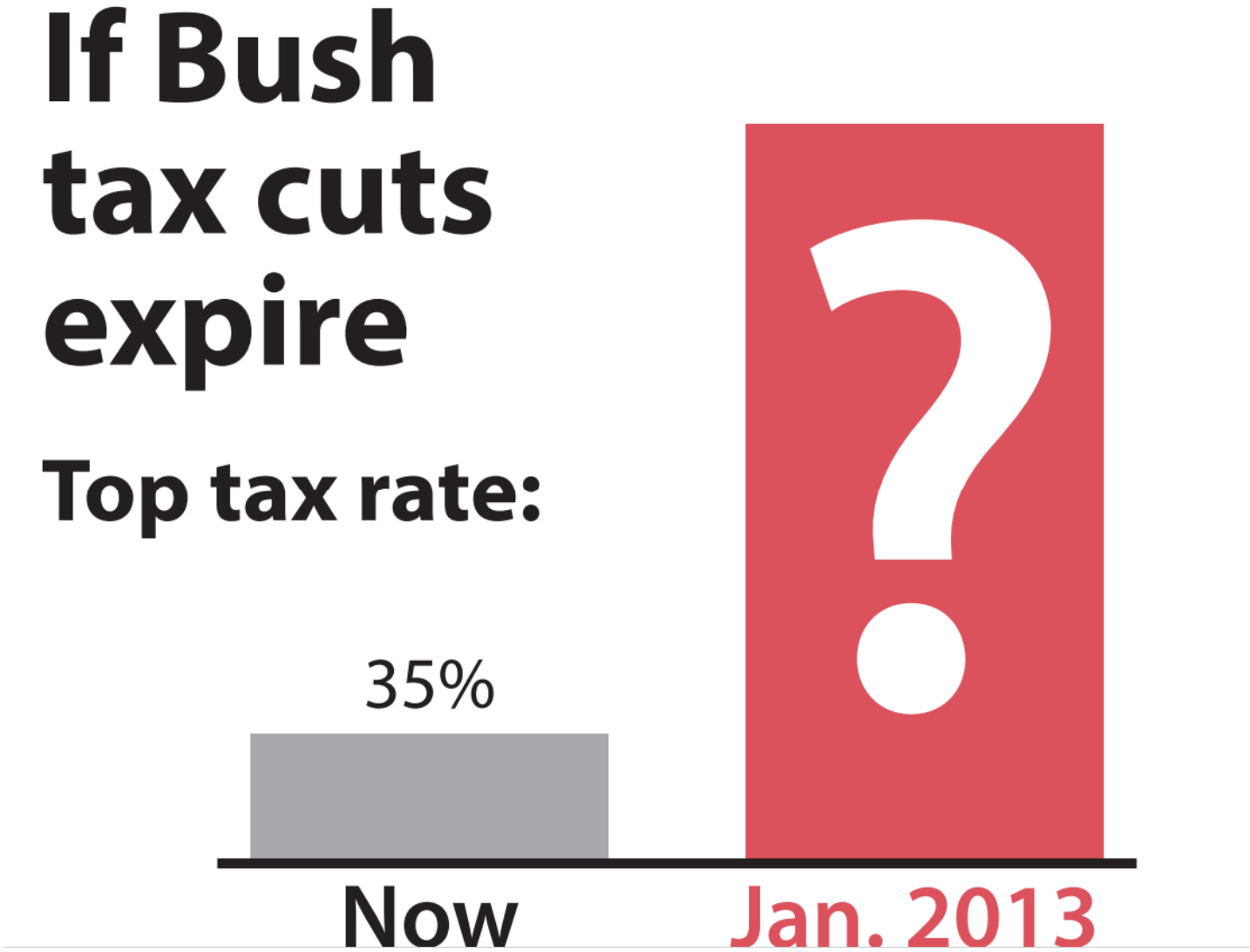




Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021

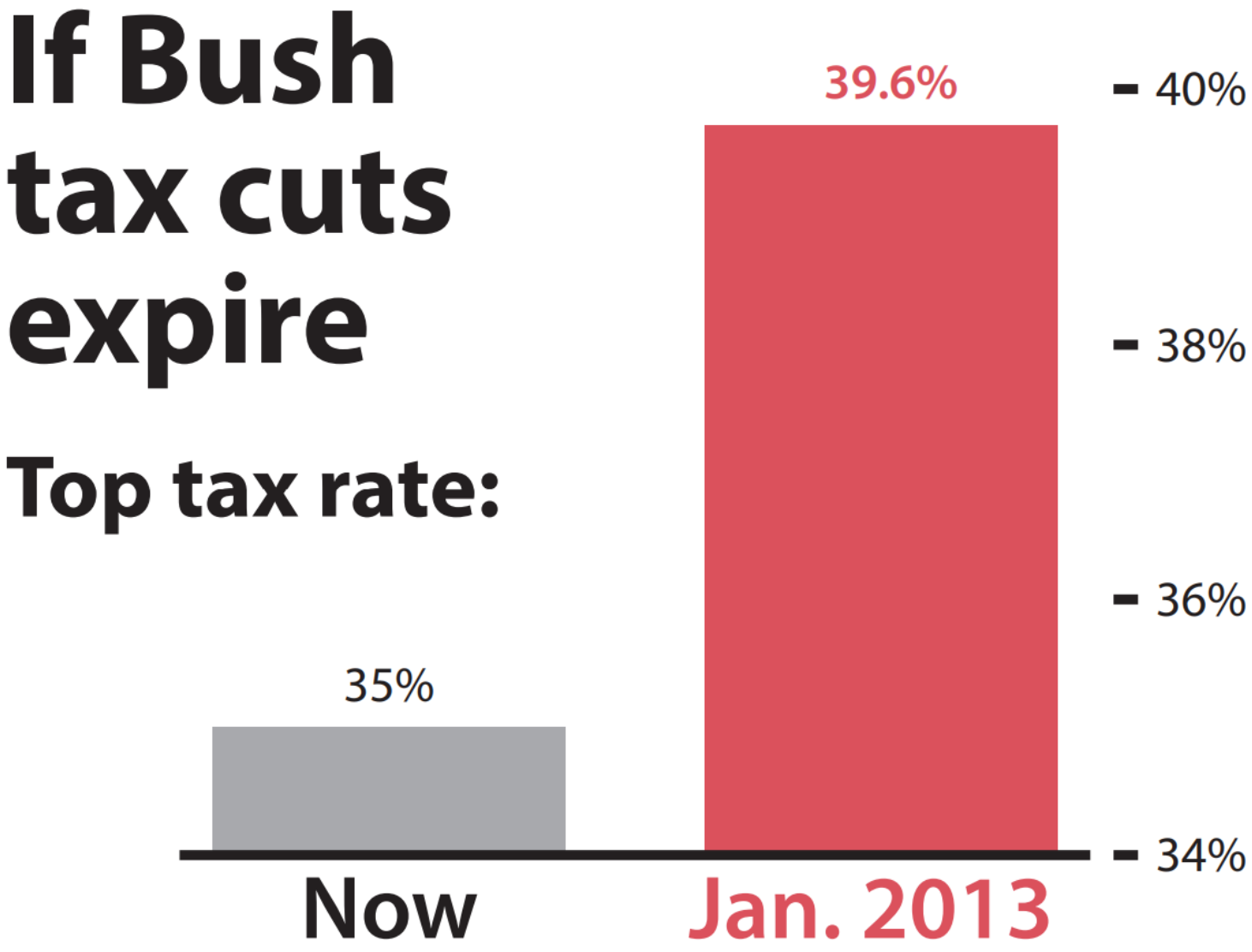



Source: Communicating Effectively with Data Visualization, Eamonn Maguire, CERN, 2021

# Avoiding misleading visualization design



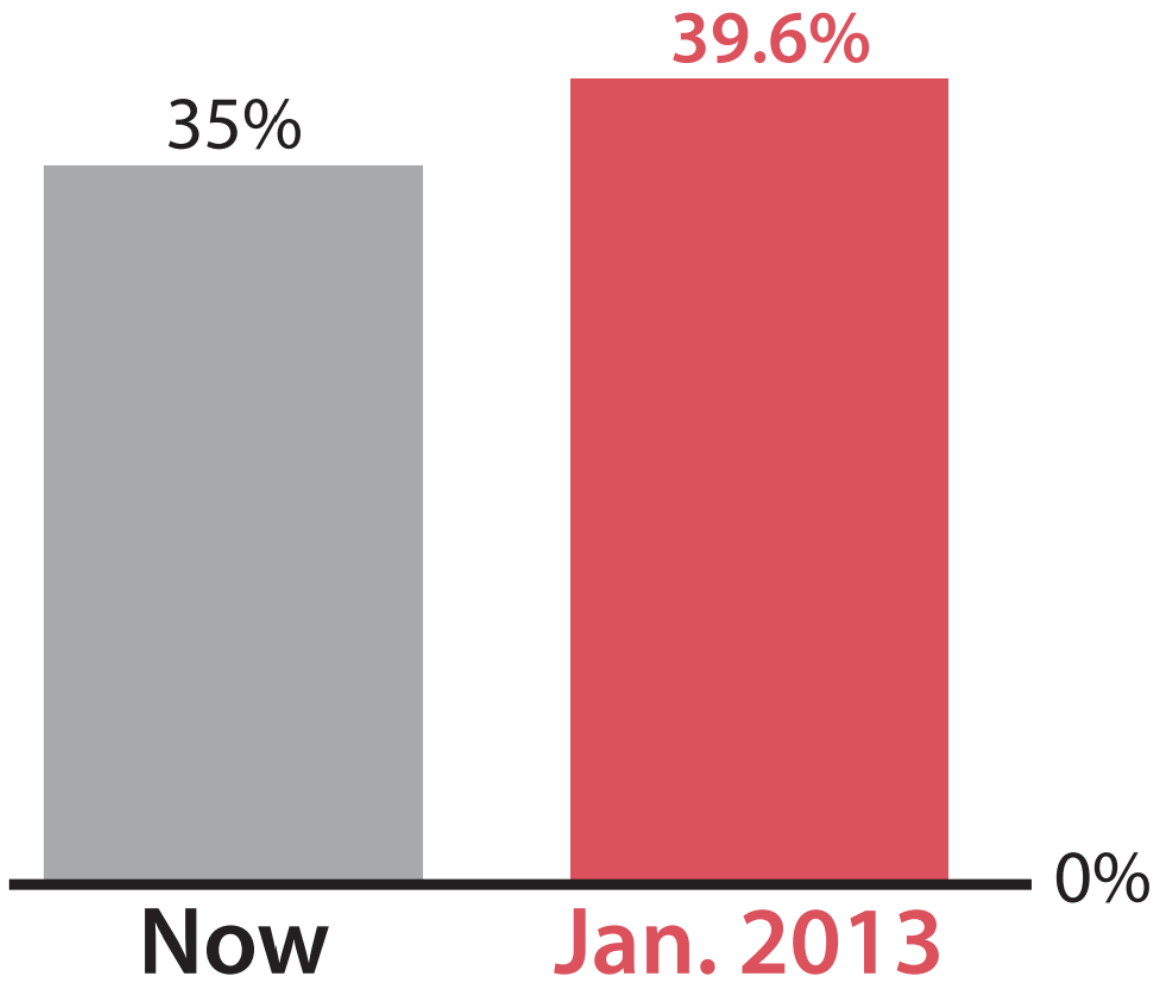

Source: How charts Lie, by Alberto Cairo




Source: How charts Lie, by Alberto Cairo

# If Bush tax cuts expire

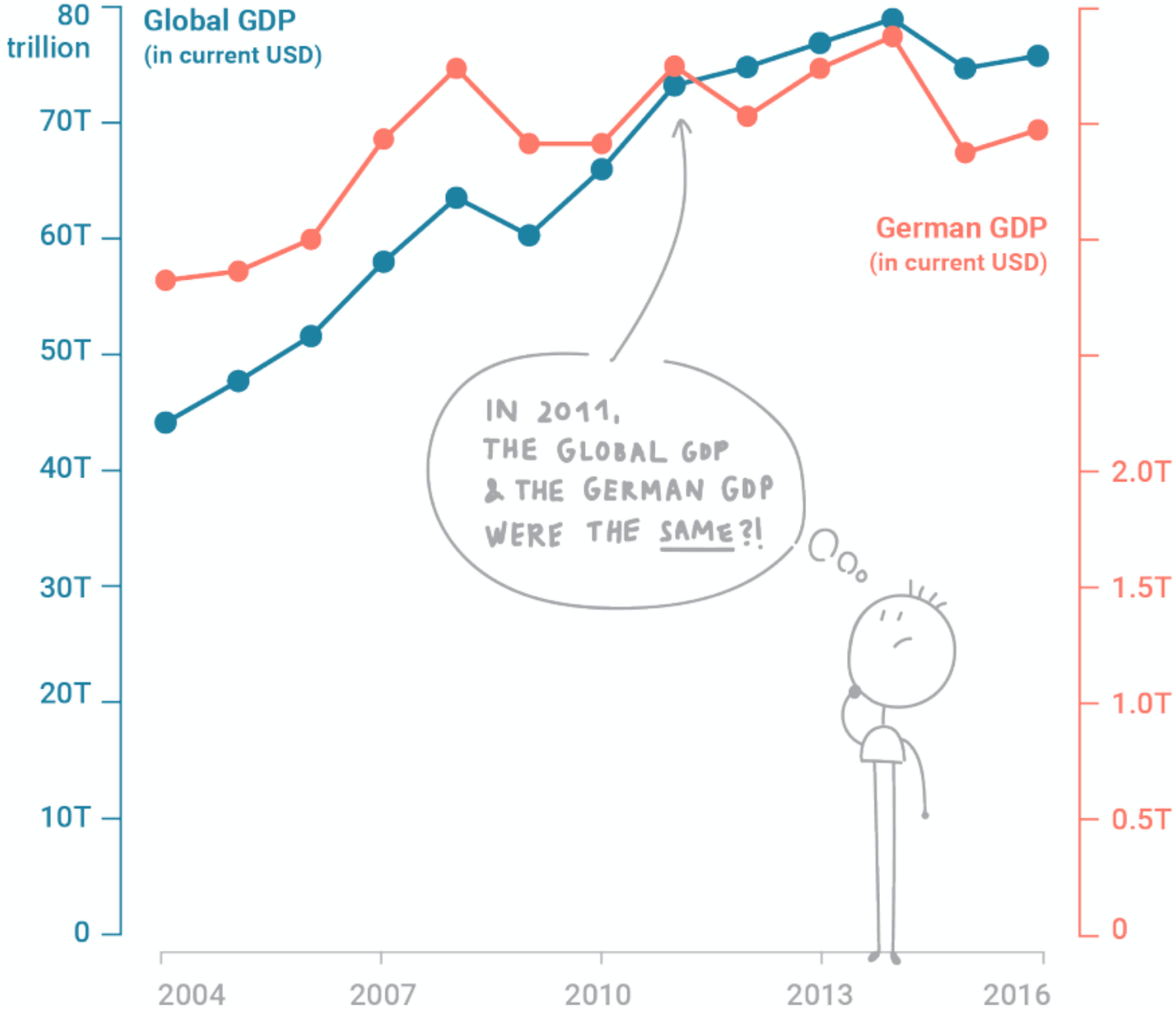
## Top tax rate:

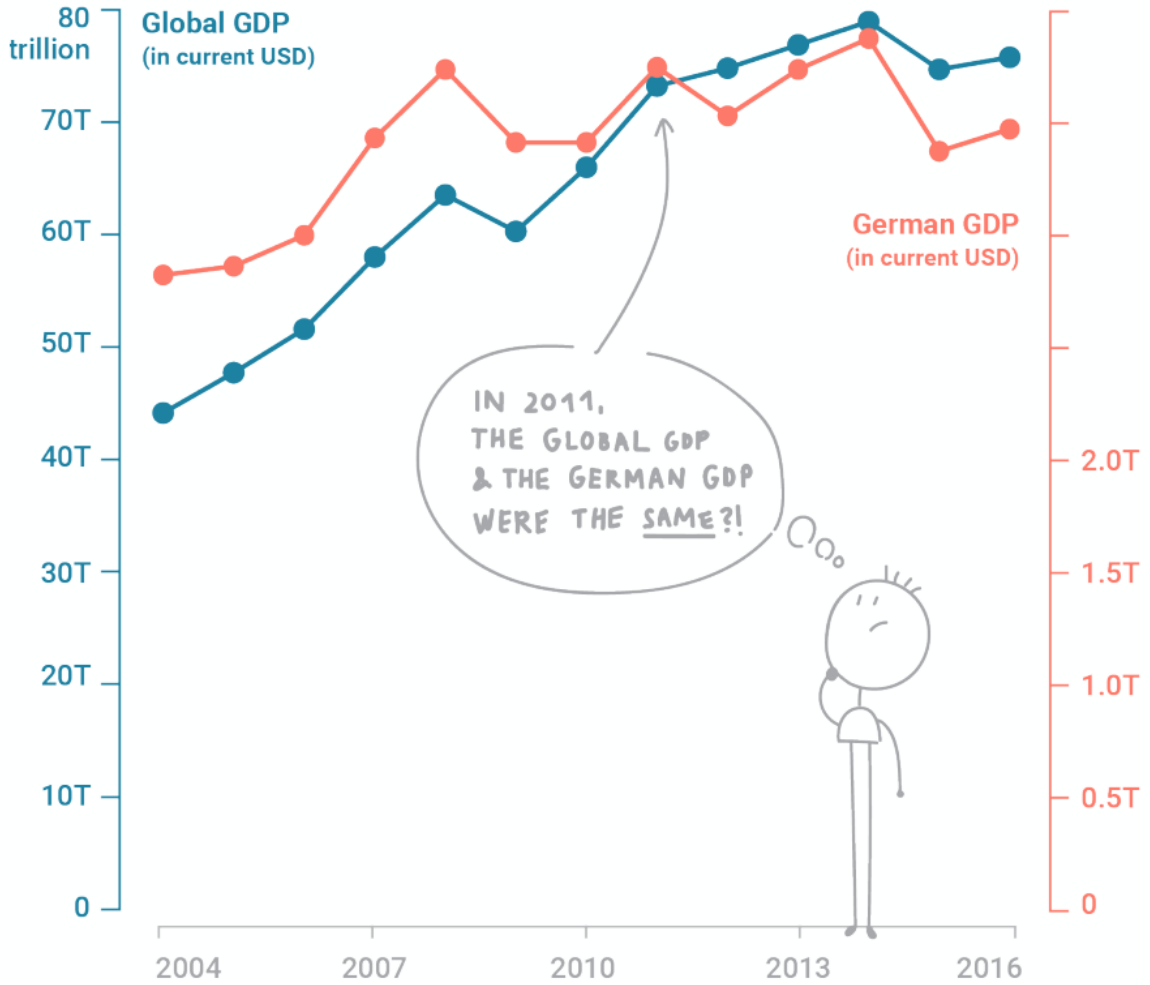



Source: How charts Lie, by Alberto Cairo

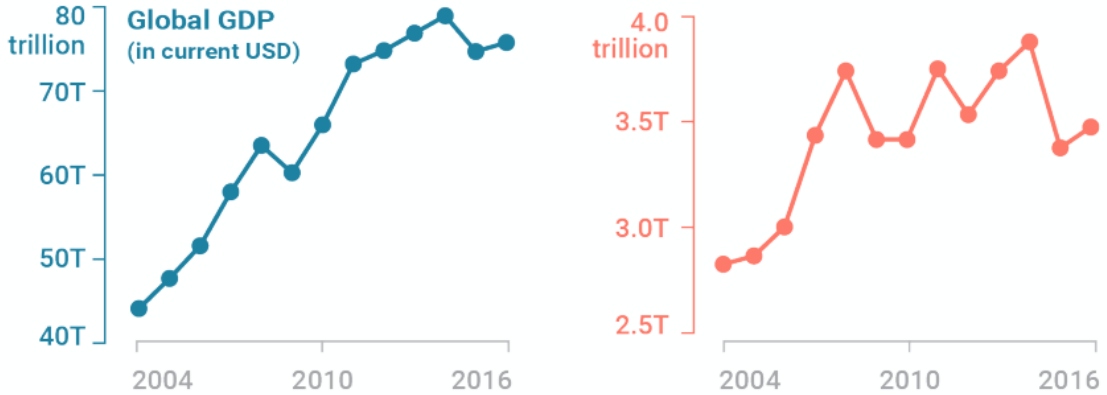
# Misleading visualization designs

- Truncated axis
- Dual Axis
- Inverted axis
- Values as areas/volume
- Uneven binning
- Unclear encoding
- Cherry-picking data
- Cherry-picking time frame
- Setting arbitrary threshold

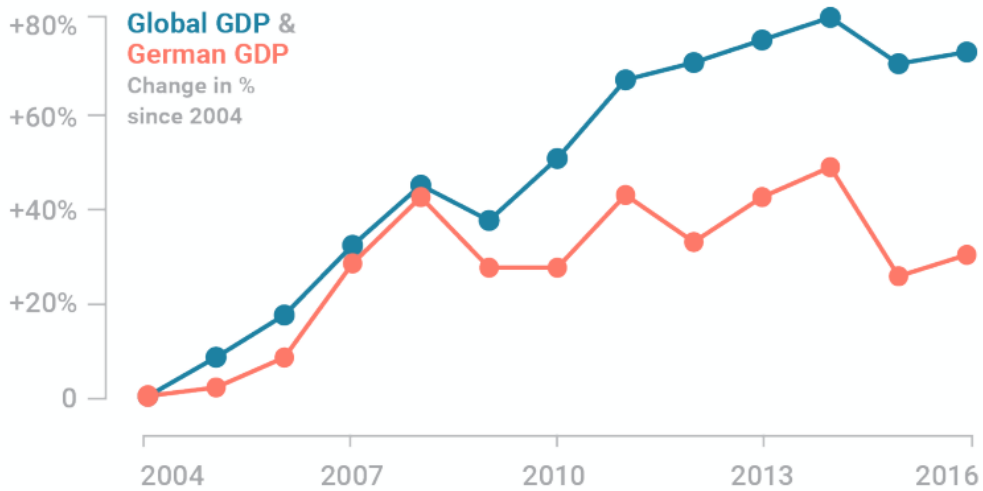

80 trillion
70T
60T
50T
40T
30T
20T
10T
0
Global GDP
(in current USD)
German GDP
(in current USD)
2.0T
1.5T
1.0T
0.5T
0
IN 2011,
THE GLOBAL GDP
& THE GERMAN GDP
WERE THE SAME?!
2004
2007
2010
2013
2016



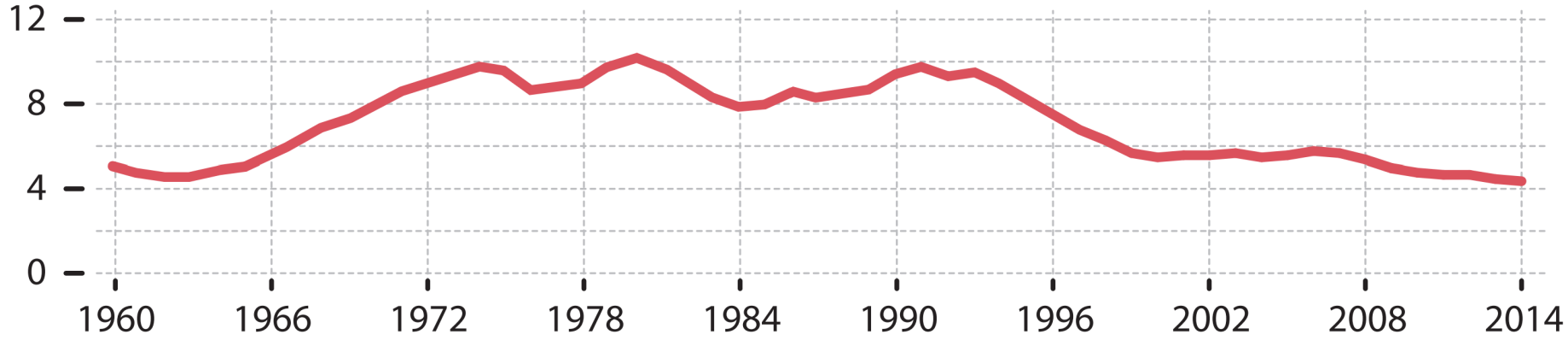

## Solution 1 - Juxtapose




## Solution 2 - Index

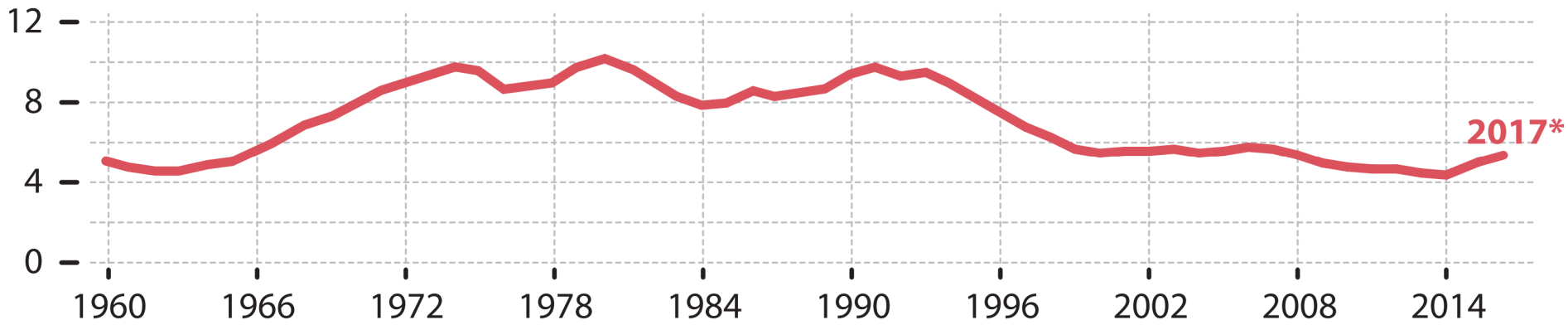


## U.S. murder rate (yearly murders per 100,000 people)




(Source: Bureau of Crime Statistics)

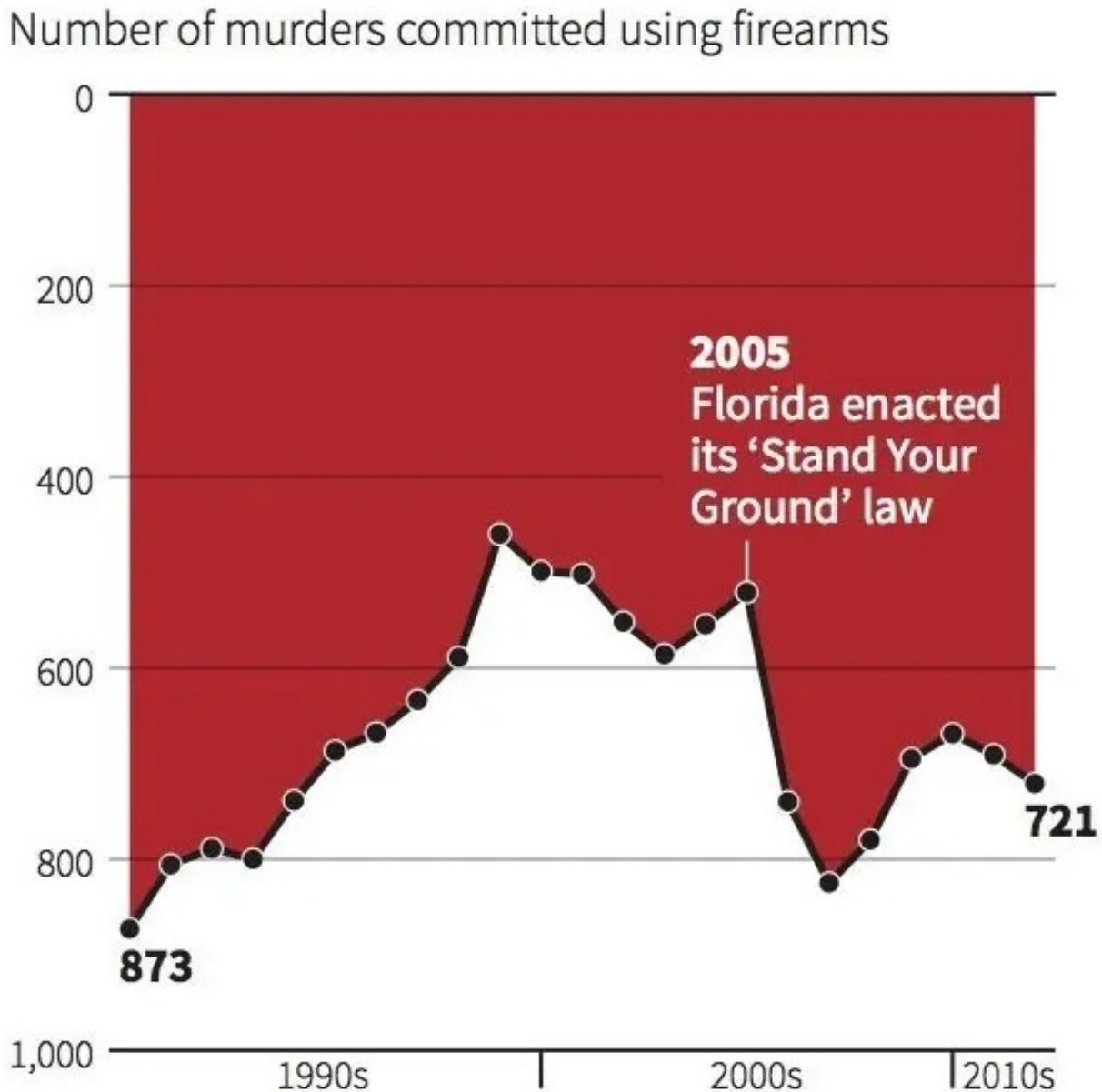
## U.S. murder rate (yearly murders per 100,000 people)



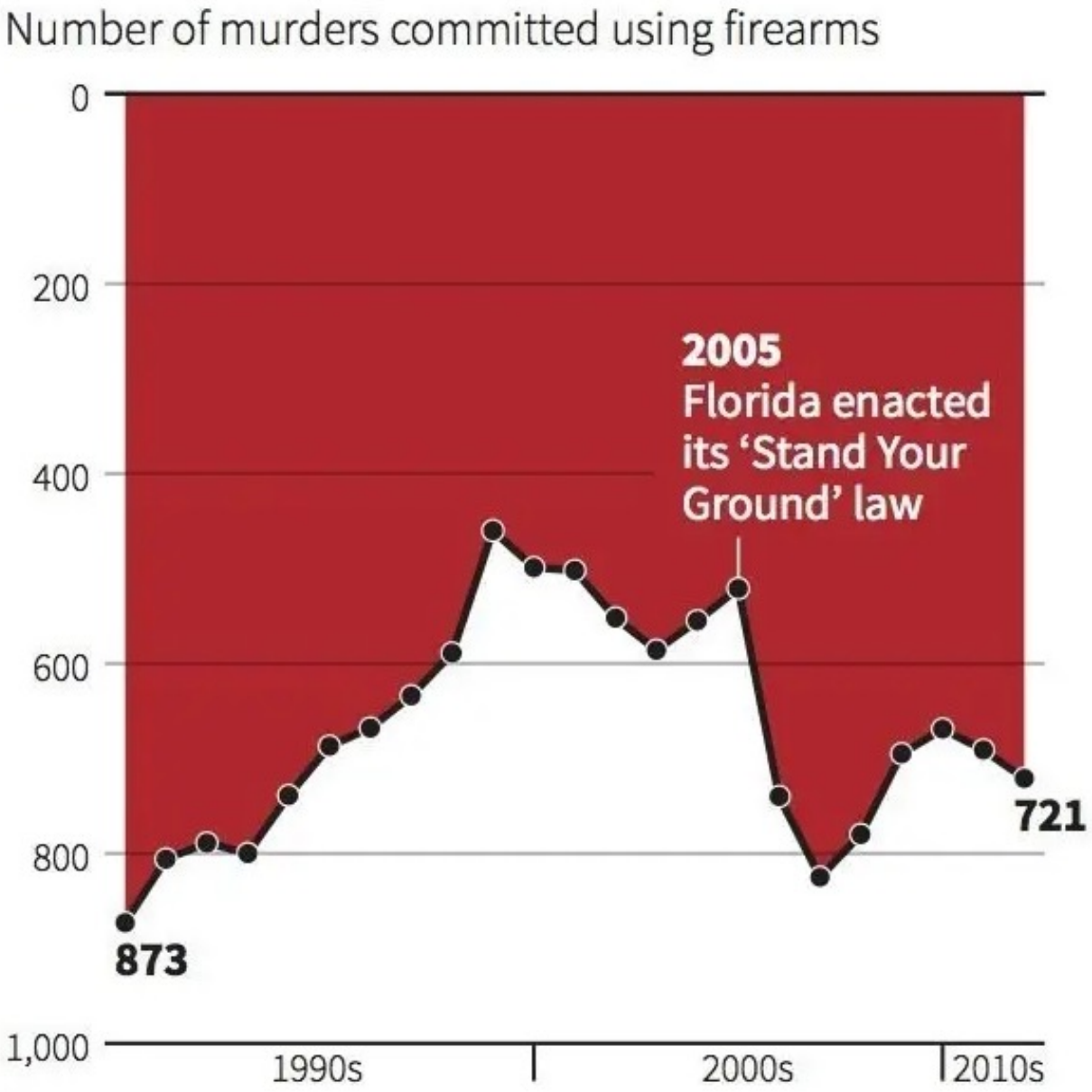

*Preliminary 2017 estimate (obtained on January 31, 2018)

Source: How charts Lie, by Alberto Cairo

# Gun deaths in Florida



# Gun deaths in Florida




The 2005 "Stand Your Ground" law resulted in a decrease in gun death in Florida.

- ◯ True
- ◯ False

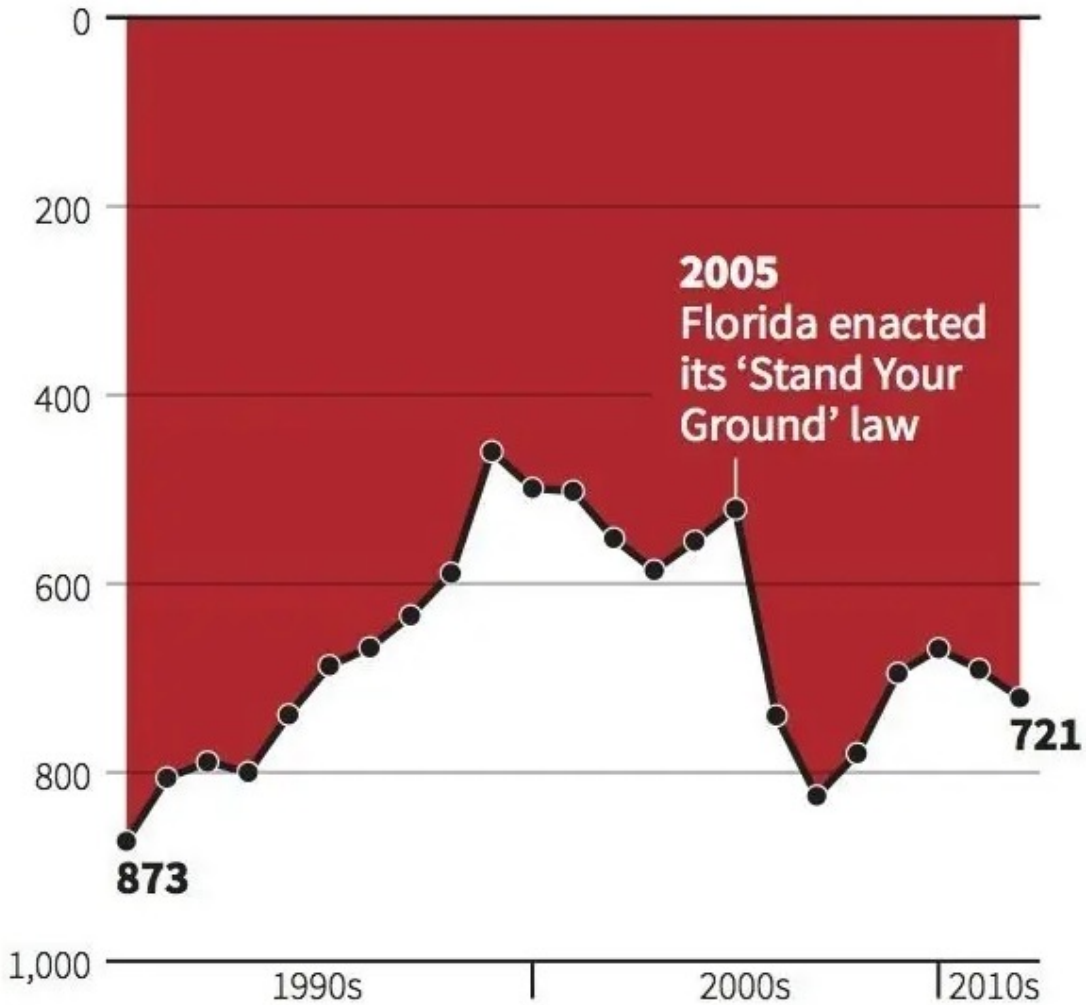
slido

# The 2005 "Stand Your Ground" law resulted in a decrease in gun death in Florida.

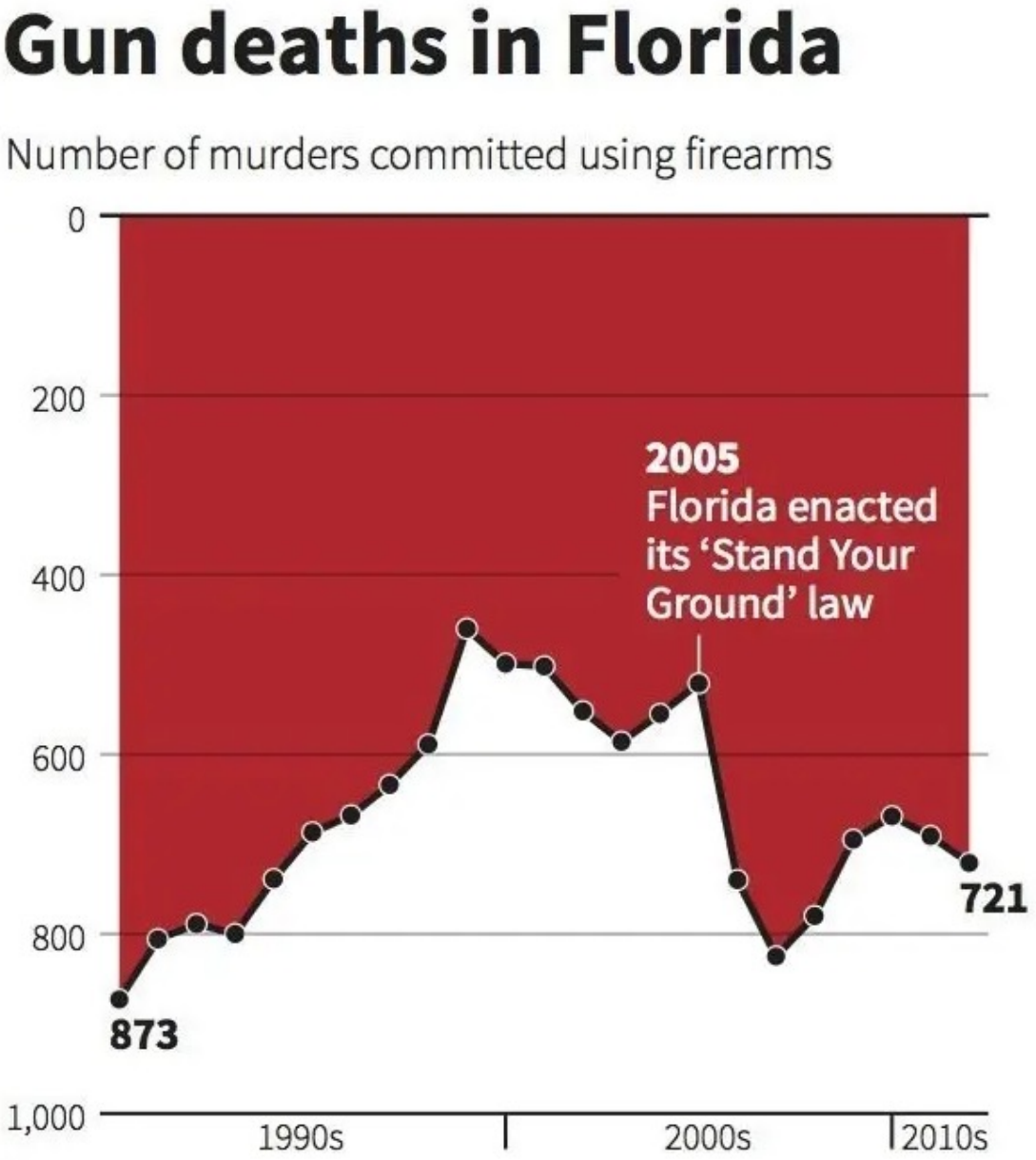
ⓘ Start presenting to display the poll results on this slide.

# Gun deaths in Florida

Number of murders committed using firearms



Inverted Y axis



In the book, Cairo explores 5 main ways that charts lie:

- by being designed poorly
- by displaying dubious data
- by displaying insufficient data
- by concealing or confusing uncertainty
- by suggesting misleading patterns

# Misleading Beyond Visual Tricks: How People *Actually* Lie with Charts

Maxim Lisnic
maxim.lisnic@utah.edu
University of Utah
Salt Lake City, Utah, USA

Cole Polychronis
cole.polychronis@utah.edu
University of Utah
Salt Lake City, Utah, USA

Alexander Lex
alex@sci.utah.edu
University of Utah
Salt Lake City, Utah, USA

Marina Kogan
kogan@cs.utah.edu
University of Utah
Salt Lake City, Utah, USA

Lisnic, M., Polychronis, C., Lex, A. and Kogan, M., 2023, April. Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-21).

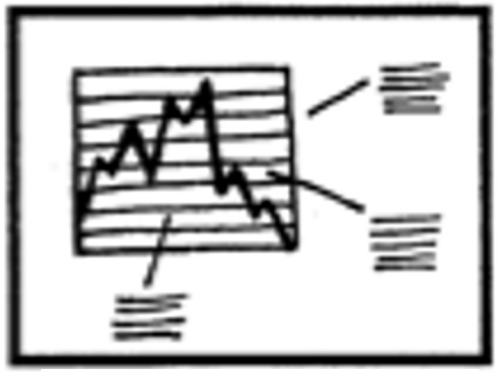
# Storytelling with data

# Storytelling with data

- Tell a **narrative** to your audience using data, visualizations, and story structure
- Narrative is the choice of **which events** to relate and in **what order** to relate them
- A narrative is a representation or specific manifestation of the story, rather than the story itself

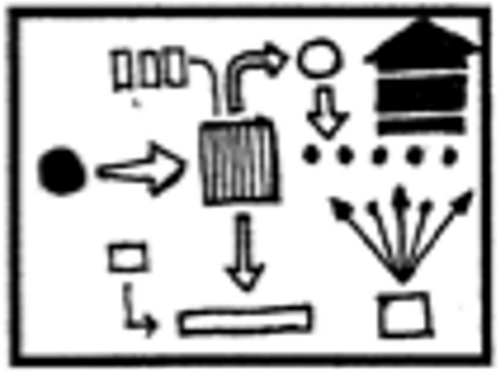
# Visual Storytelling Genres

Magazine Style

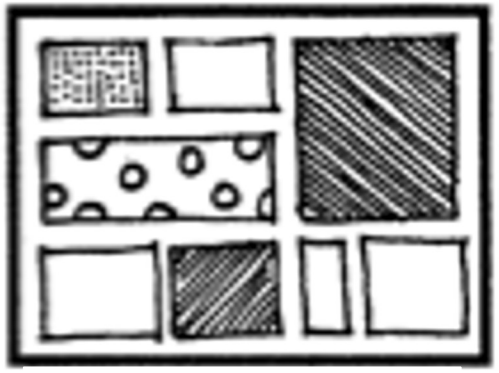


Annotated Chart

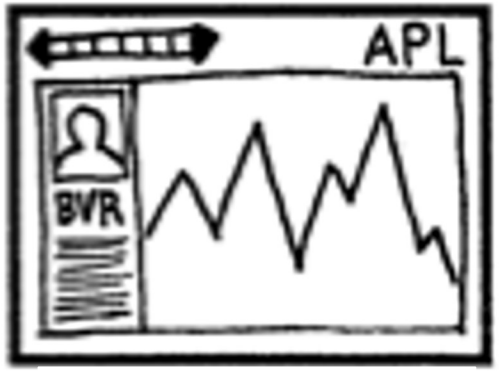
Partitioned Poster



Flow Chart



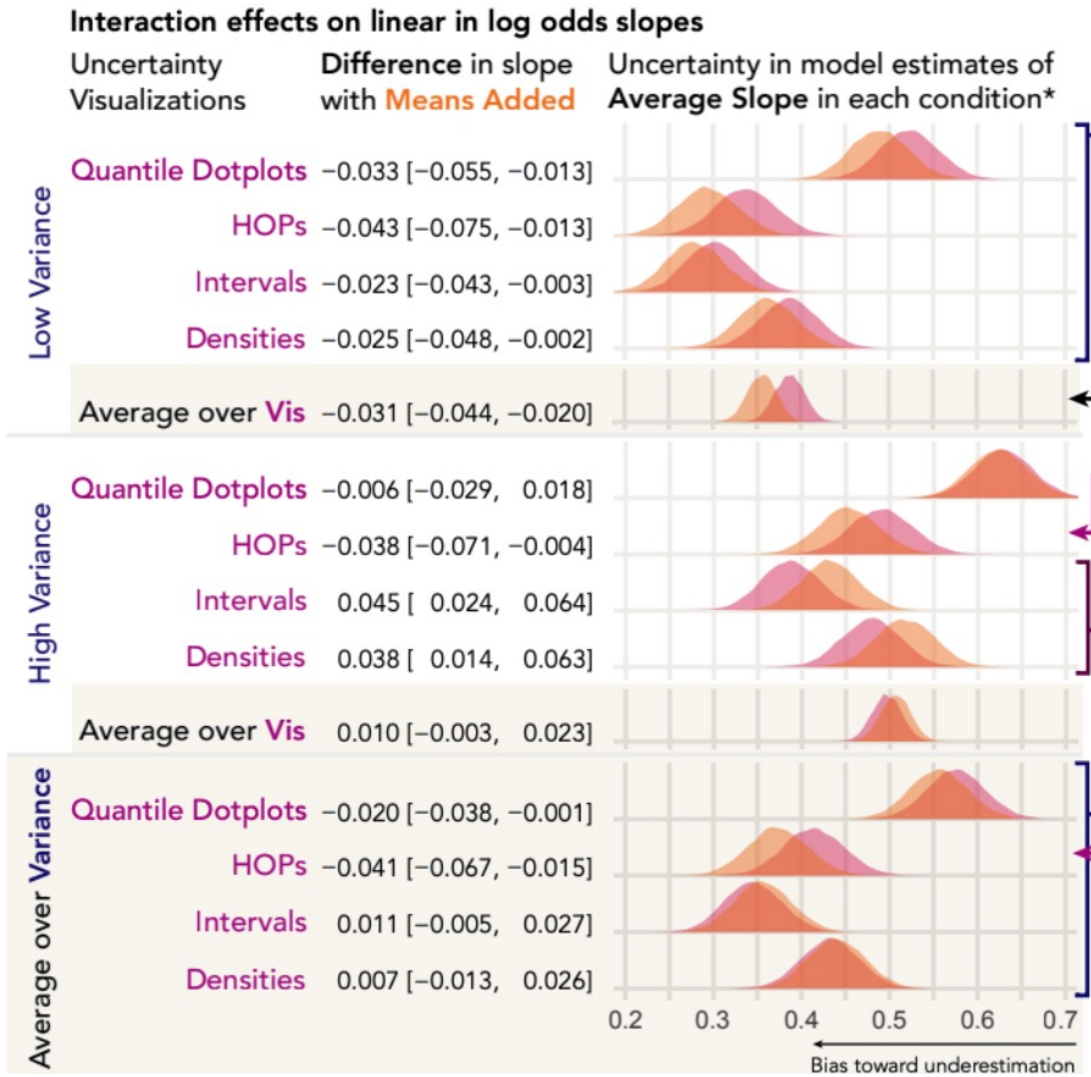
Comic Strip




Slide Show

Video/Animation/Film

Segel, E. and Heer, J., 2010. Narrative visualization: Telling stories with data. TVCG, 16(6), pp.1139-1148.




# 4 Results

## 4.1 Probability of Superiority Judgments

For each uncertainty visualization, **adding means** at **low variance** decreases LLO slopes. Recall that a slope of one corresponds to no bias, and a slope less than one indicates underestimation. When we **average over uncertainty visualizations**, **adding means** at **low variance** reduces LLO slopes for the average user, indicating a very small 0.8 percentage points increase in probability estimation error.

At **high variance**, the effect of **adding means** changes directions for different uncertainty visualizations. **Adding means** decreases LLO slopes for **HOPs**, whereas **adding means** increases LLO slopes for **intervals and densities**. Because differences in LLO slopes represent changes in the exponent of a power law relationship, these slope differences of similar magnitude indicate a very small increase in probability of superiority estimation error of 0.3 percentage points for HOPs and small reductions in error of about 1.5 and 1.0 percentage points for intervals and densities, respectively.

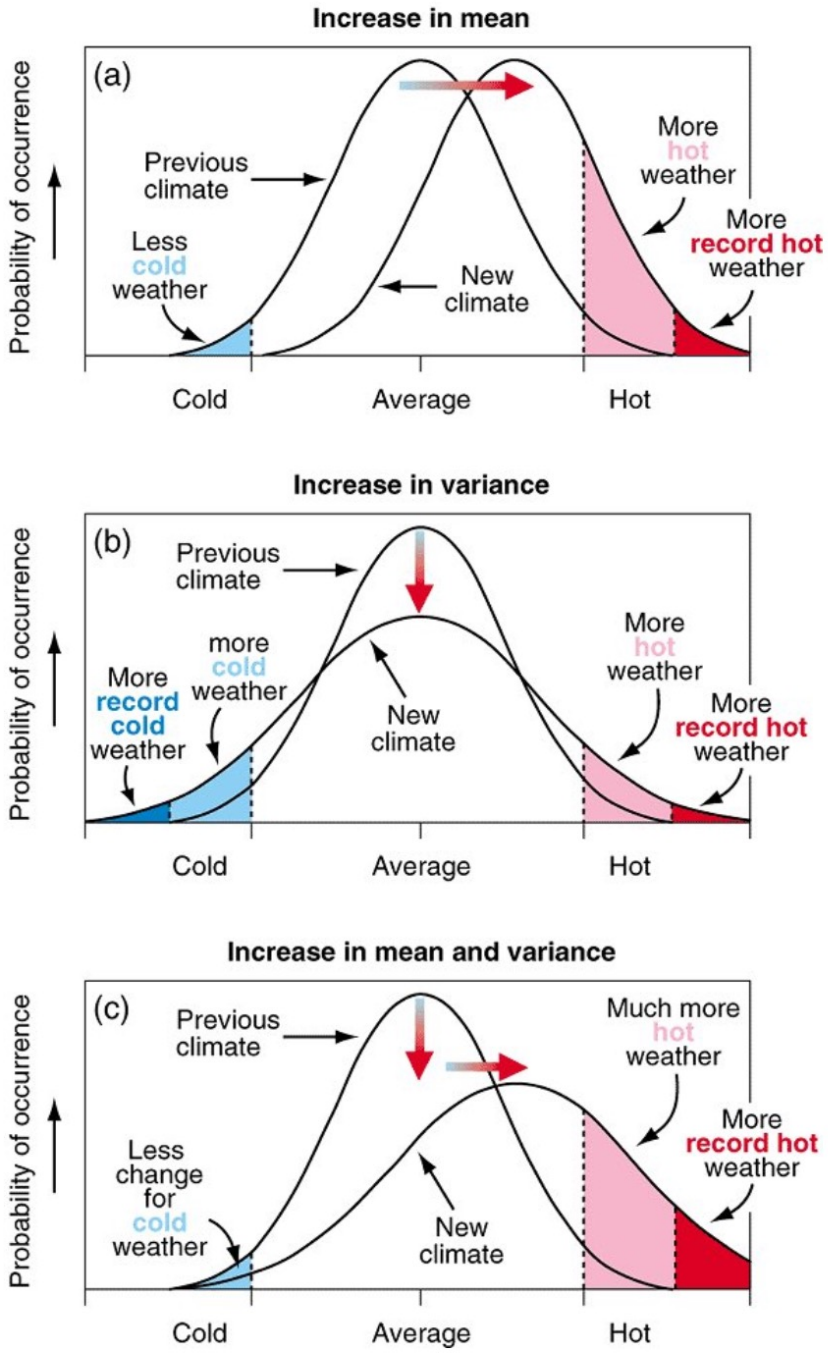
Users of all uncertainty visualizations underestimate effect size. When we **average over variance**, users show an average estimation error of 8.6, 14.0, 14.8, and 12.4 percentage points in probability of superiority units for quantile dotplots, HOPs, intervals, and densities, respectively, each **without means**. In this marginalization, **adding means** only has a reliable impact on LLO slopes for **HOPs**, but the difference is practically negligible.

A composition of charts can help tell a story.

We don't need to read a caption or have an explanation to interpret this (at first) rather complex visualization.

Here, **annotations** are used inline to communicate what is happening to the climate when we have not just an increase in the mean, but also variance.

Colour effectively communicates warm vs cold.



# What does a story add to visualizations?

- Explanations and semantics of data
- Drawing attention to important features
- Contextual information that will augment understanding
- Thought provoking and action inspiring

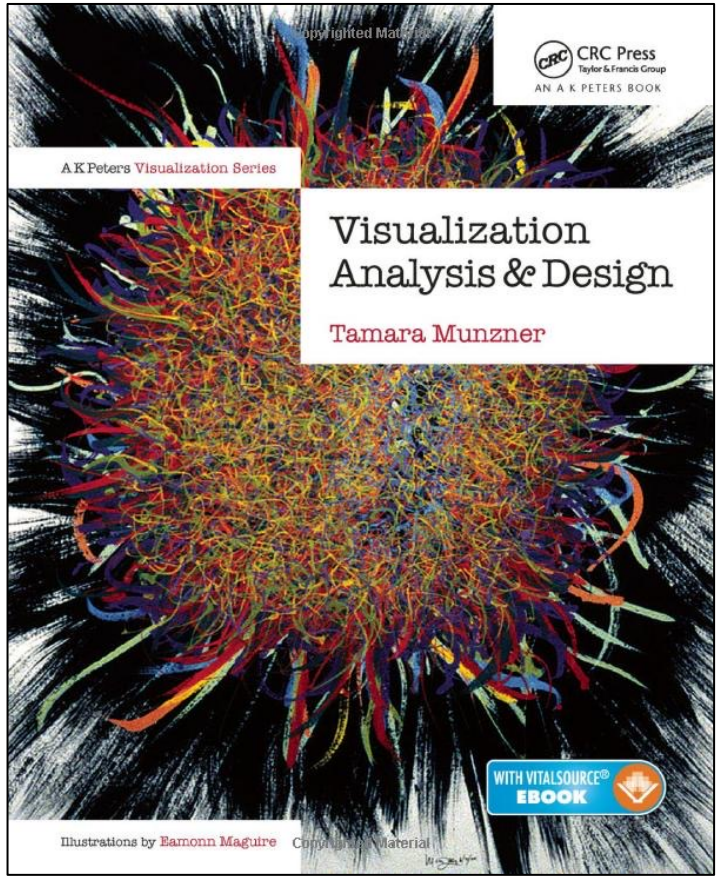
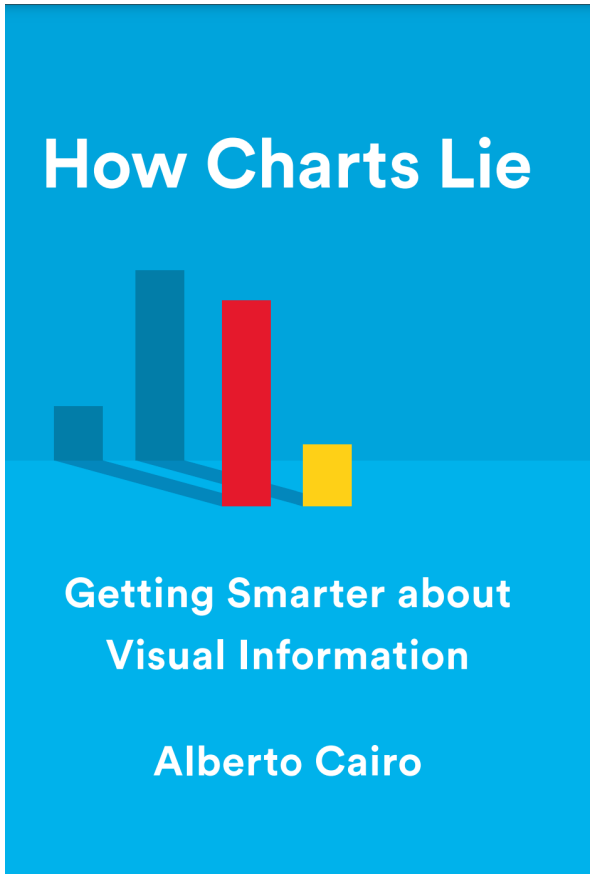
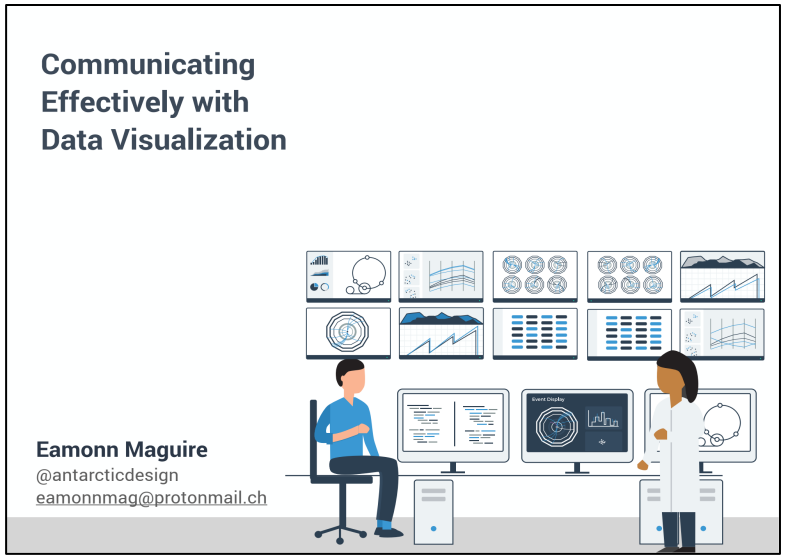
# Tips to design effective data stories

- Use Data-Driven Story to Present Statistics
- Address Your Audience's Main Concerns
- Ensure Your Data Story Does Not Lose Focus
- The Data Must Support Your Story
- Highlight the Human in Your Storytelling
- Add the "Aha!" Moments
- Use Proper Data Visualization Methods

# Use the right chart type

- This information should give you a sense of which chart type will be most appropriate:
- **Pie charts** are best for comparisons between a relatively low number of categories;
- **Bar charts** are best for very precise comparisons between categories, and for when you want to show negative and positive values in the dataset;
- **Scatter charts** are great to show correlation and clustering, especially if you have a lot of data to show;
- **Line charts** emphasize trends over time;
- **Bubble charts** showcase distribution or relationships in large data sets;
- **Area charts** let you compare volumes of data easily.
-

# Wrap up and recommendation

- Know your audience
- Determine the best visual & effective mapping
- Use colors wisely
- Highlight the most important information
- Use storytelling and annotation
- Avoid clutter
- Consider an Infographic
- Incorporate interactivity
- Communicate uncertainty

# Workshop Heavily Influenced By…