

# Data Visualization

# Outline

---

0:10                  What this workshop is about

---

0:15                  Basics: form or content?

---

0:20                  Types of charts

---

0:35                  Instruments

---

0:50                  Design principles

---

0:55                  Color use

---

1:10                  Visualization software

---

# What problems do you have with visualizations?

1. I do not know what I want to visualize and I am not sure where to start
2. I know what I want to show, but not sure how to do it
3. My graphs look bad, no matter what I do 😣
4. I have no problems! I am very good at visuals

# What software do you use for visualization?

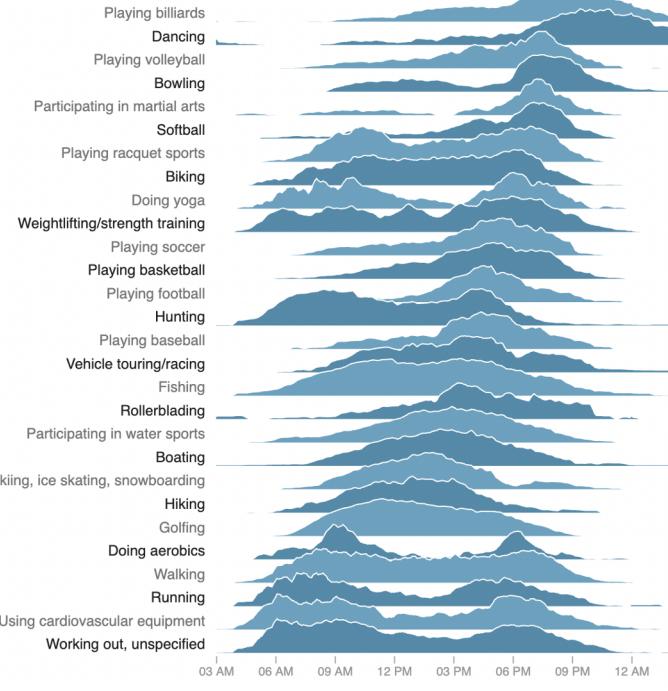
1. Python or R
2. Tableau
3. Excel
4. Other (please, share in the chat!)

# Learning objectives

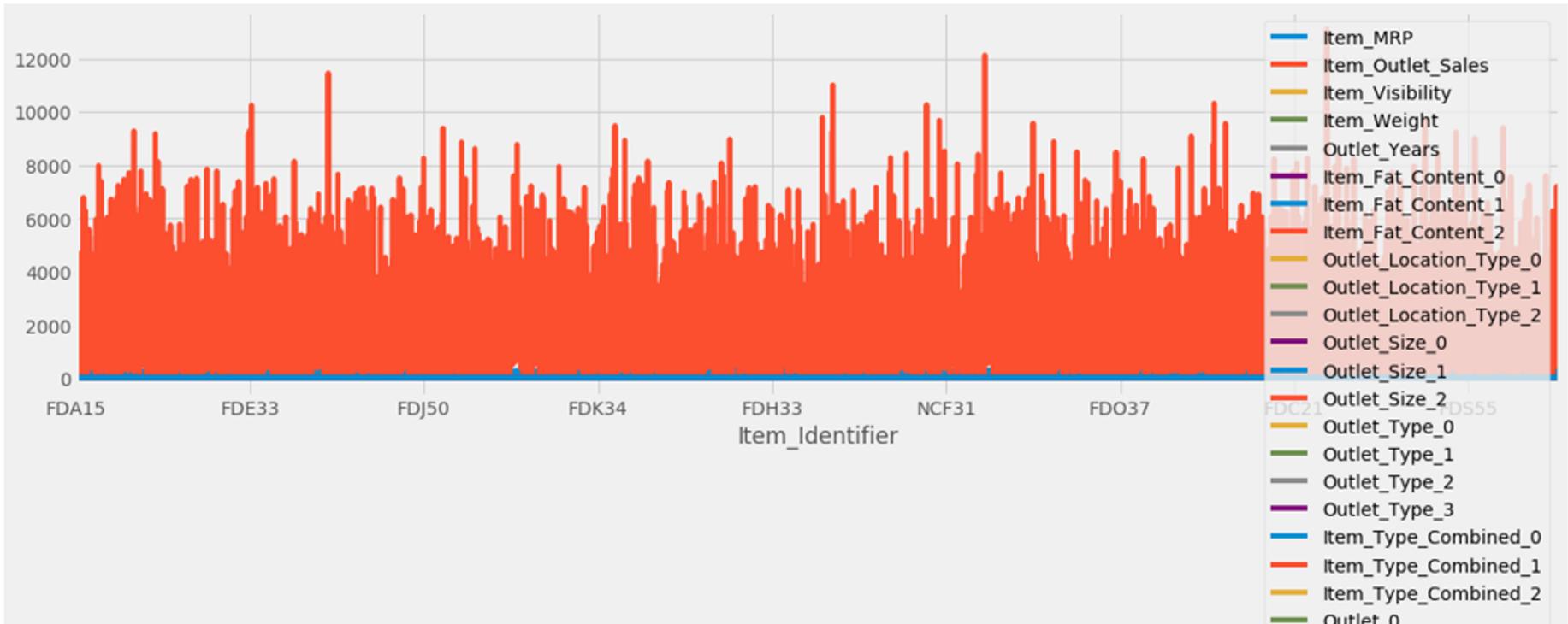
1. Understand basic principles behind effective data visualization
2. Learn about graph types and visualization instruments
3. Be familiar with aesthetics features that make graphs clearer and *nicer!*

# Why do we need visuals?

"20030100013280", 1, 68, -1, -1, 1, -1, -1, "04:00:00", "05:00:00", 130124, 13, 1301, 97, 60, 60, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 2, 30, -1, -1, -1, -1, -1, "05:00:00", "05:30:00", 10201, 1, 102, 0, 90, 90, 30, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 3, 600, -1, -1, -1, -1, -1, "05:30:00", "15:30:00", 10101, 1, 101, 0, 690, 690, 600, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 4, 150, -1, -1, -1, -1, -1, "15:30:00", "18:00:00", 120303, 0, 840, 840, 150, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 5, 5, -1, -1, -1, -1, -1, "18:00:00", "18:05:00", 110101, 11, 1101, 0, 845, 845, 5, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 6, 175, -1, -1, -1, -1, -1, "18:05:00", "21:00:00", 120303, 12, 1203, 0, 1020, 1020, 175, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 7, 270, -1, -1, -1, -1, -1, "21:00:00", "01:30:00", 10101, 1, 101, 0, 1290, 1290, 270, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 8, 10, -1, -1, -1, -1, -1, "01:30:00", "01:45:00", 10201, 1, 100, 0, 1300, 10, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013280", 9, 140, -1, -1, -1, -1, -1, "01:45:00", "04:30:00", 130124, 13, 1301, 1470, 1440, 170, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 1, 180, -1, -1, -1, -1, -1, "04:00:00", "07:00:00", 10101, 11, 1101, 0, 240, 240, 60, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 2, 60, -1, -1, -1, -1, -1, "07:00:00", "08:00:00", 110101, 11, 1101, 0, 240, 240, 60, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 3, 100, -1, -1, -1, -1, -1, "07:00:00", "08:30:00", 10101, 11, 1101, 0, 240, 240, 60, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 4, 60, -1, -1, -1, -1, -1, "09:00:00", "09:30:00", 10101, 11, 1101, 0, 240, 240, 60, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 5, 10, -1, -1, -1, -1, -1, "10:00:00", "10:10:00", 151201, 18, 1512, 0, 370, 370, 10, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 6, 530, -1, -1, -1, -1, -1, "10:10:00", "19:00:00", 120101, 12, 1201, 0, 900, 900, 530, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 7, 18, -1, -1, -1, -1, -1, "19:00:00", "19:10:00", 151201, 18, 1512, 0, 910, 910, 18, 1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 8, 30, -1, -1, -1, -1, -1, "19:10:00", "19:40:00", 110101, 11, 1101, 0, 940, 940, 30, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 9, 60, -1, -1, -1, -1, -1, "19:40:00", "20:40:00", 120303, 12, 1203, 0, 1000, 1000, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013344", 10, 140, -1, -1, -1, -1, -1, "20:40:00", "01:00:00", 10101, 1, 101, 0, 1620, 1440, 620, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 1, 300, -1, -1, -1, -1, -1, "04:00:00", "09:00:00", 10101, 1, 101, 97, 300, 300, 300, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 2, 60, -1, -1, -1, -1, -1, "09:00:00", "09:30:00", 10101, 1, 101, 97, 300, 300, 300, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 3, 30, -1, -1, -1, -1, -1, "10:00:00", "10:30:00", 110101, 11, 1101, 0, 390, 390, 30, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 4, 5, -1, -1, -1, -1, -1, "10:30:00", "10:35:00", 180701, 18, 1807, 0, 395, 395, 5, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 5, 60, -1, -1, -1, -1, -1, "10:35:00", "11:35:00", 70101, 7, 701, 0, 455, 455, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 6, 15, -1, -1, -1, -1, -1, "11:35:00", 180701, 18, 1807, 0, 460, 460, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 7, 15, -1, -1, -1, -1, -1, "11:40:00", "11:55:00", 20101, 2, 201, 0, 475, 475, 15, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 8, 30, -1, -1, -1, -1, -1, "11:55:00", "12:00:00", 110101, 11, 1101, 0, 490, 490, 15, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 9, 60, -1, -1, -1, -1, -1, "12:00:00", "12:15:00", 110101, 11, 1101, 0, 520, 520, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 10, 180, -1, -1, -1, -1, -1, "12:15:00", "12:30:00", 110101, 11, 1101, 0, 550, 550, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 11, 40, -1, -1, -1, -1, -1, "12:30:00", "17:55:00", 20201, 2, 202, 0, 635, 635, 180, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 12, 45, -1, -1, -1, -1, -1, "17:55:00", "18:35:00", 120101, 12, 1201, 0, 875, 875, 40, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 13, 45, -1, -1, -1, -1, -1, "18:35:00", "19:20:00", 110101, 11, 1101, 0, 920, 920, 45, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 13, 18, -1, -1, -1, -1, -1, "19:20:00", "22:20:00", 120101, 12, 1201, 0, 1100, 1100, 180, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 14, 60, -1, -1, -1, -1, -1, "22:20:00", "23:20:00", 20203, 2, 202, 0, 1160, 1160, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 15, 20, -1, -1, -1, -1, -1, "23:20:00", "23:40:00", 10201, 1, 102, 0, 1180, 1180, 20, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013352", 16, 26, -1, -1, -1, -1, -1, "23:40:00", "09:45:00", 10101, 1, 101, 0, 1785, 1440, 605, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 1, 360, -1, -1, -1, -1, -1, "04:00:00", "10:00:00", 30101, 1, 301, 0, 380, 380, 201, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 2, 20, -1, -1, -1, -1, -1, "10:00:00", "20:20:00", 20201, 2, 202, 0, 385, 385, 180, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 3, 15, -1, -1, -1, -1, -1, "10:20:00", "10:25:00", 30101, 3, 301, 0, 385, 385, 5, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 4, 15, -1, -1, -1, -1, -1, "10:25:00", "10:40:00", 20203, 2, 202, 0, 400, 400, 15, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 5, 20, -1, -1, -1, -1, -1, "10:40:00", "11:00:00", 10201, 1, 102, 0, 420, 420, 20, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 6, 10, -1, -1, -1, -1, -1, "11:00:00", "11:10:00", 20681, 2, 206, 0, 430, 430, 10, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 7, 120, -1, -1, -1, -1, -1, "11:10:00", "11:20:00", 10481, 1, 500, 0, 550, 550, 120, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 8, 30, -1, -1, -1, -1, -1, "11:20:00", "13:00:00", 10481, 1, 500, 0, 550, 550, 170, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 9, 75, -1, -1, -1, -1, -1, "13:00:00", "15:00:00", 10101, 1, 101, 0, 730, 730, 75, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 10, 75, -1, -1, -1, -1, -1, "15:00:00", "17:15:00", 20201, 2, 202, 0, 750, 750, 75, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 11, 15, -1, -1, -1, -1, -1, "17:15:00", "17:30:00", 10101, 1, 101, 0, 810, 810, 810, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 12, 25, -1, -1, -1, -1, -1, "17:30:00", "17:45:00", 110101, 11, 1101, 0, 850, 850, 25, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 13, 25, -1, -1, -1, -1, -1, "17:45:00", "18:10:00", 110101, 11, 1101, 0, 875, 875, 25, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 14, 20, -1, -1, -1, -1, -1, "18:10:00", "18:35:00", 20203, 2, 202, 0, 875, 875, 25, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 15, -15, -1, -1, -1, -1, -1, "18:35:00", "23:00:00", 120303, 12, 1203, 0, 1140, 1140, 265, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100013348", 16, 285, -1, -1, -1, -1, -1, "23:00:00", "23:15:00", 10201, 1, 102, 0, 1155, 1155, 15, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100014165", 1, 60, -1, -1, -1, -1, -1, "04:00:00", "05:00:00", 10101, 1, 101, 97, 60, 60, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100014165", 2, 30, -1, -1, -1, -1, -1, "05:00:00", "05:30:00", 10201, 1, 102, 0, 90, 90, 30, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100014165", 3, 15, -1, -1, -1, -1, -1, "05:30:00", "05:45:00", 1080501, 18, 10805, 0, 105, 105, 15, -1, -1, -1, -1, -1, -1, -1, -1, -1  
 "20030100014165", 4, 435, -1, -1, -1, -1, -1, "05:45:00", "13:00:00", 50101, 5, 501, 0, 540, 540, 435, -1, -1, -1, -1, -1, -1, -1, -1, -1



# Do we need to learn how to visualize?



# Graph should make a point

## 1. Hypothesis

*What do you want to show?*

## 2. Proof

*How to emphasize that?*

## 3. Explaining

*What does the chart show?*

# BASICS

# Form

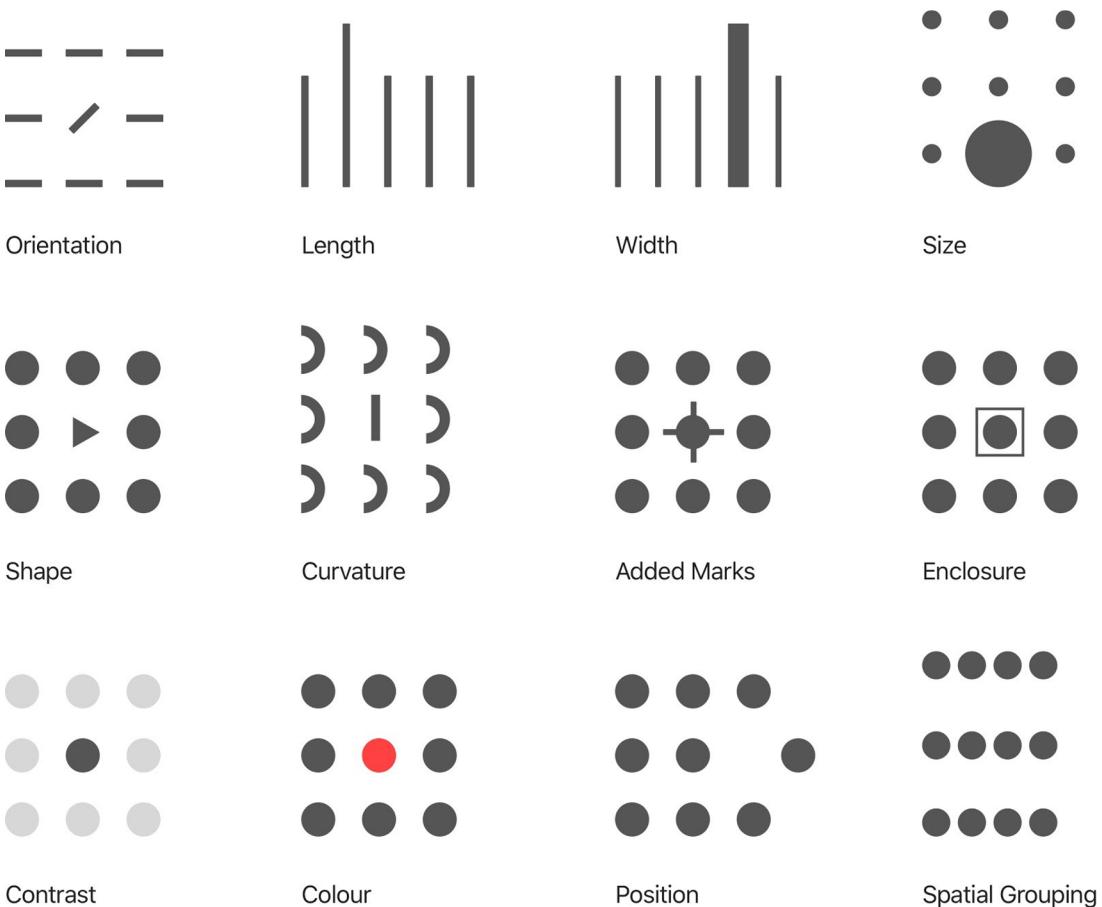
Graph type  
Features  
Design

# Content

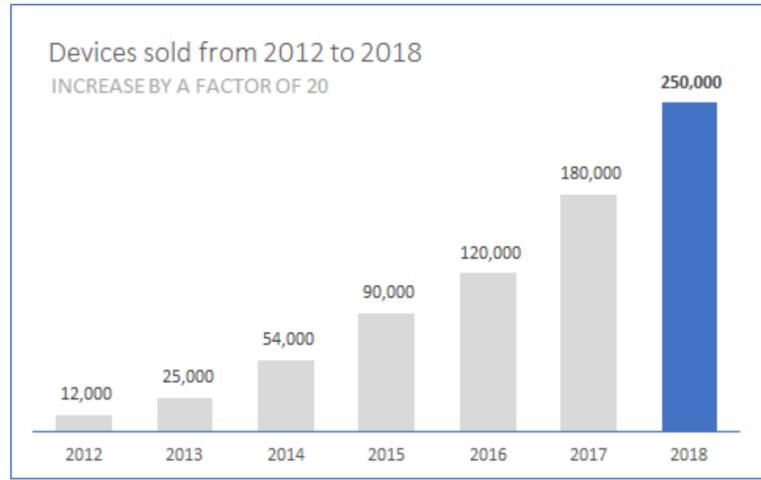
Data format  
Transformation  
Context

**Both are important!**

# Preattentive Features



# Preattentive features establish the hierarchy of information



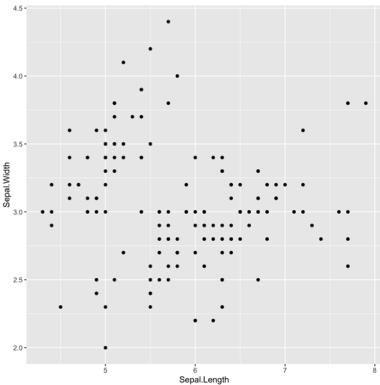
- Remove 3D
- Use direct labeling
- Remove grid (if appropriate)
- Use color meaningfully

If possible, declutter!

# Choosing a Graph Type

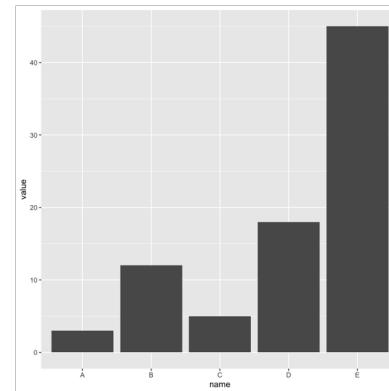
Consider:

- Purpose
- Variable types
- Dimensions

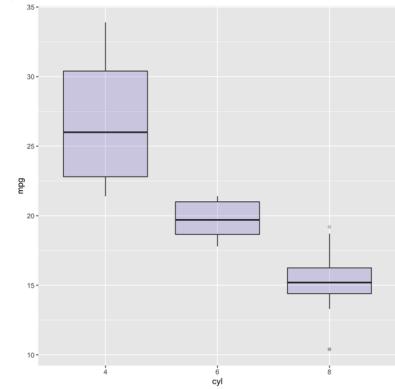


scatterplot

## Hall of Fame

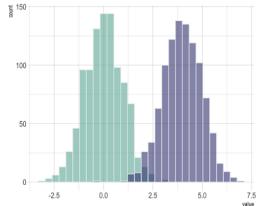


barplot

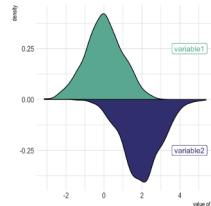


boxplot

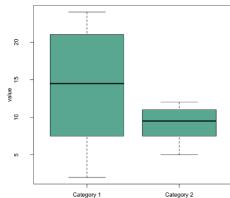
# Distribution



histogram

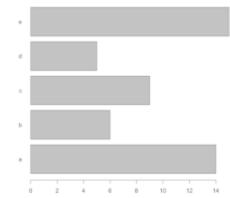


density plot



boxplot

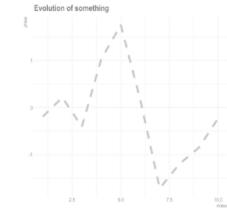
# Comparison



barplot

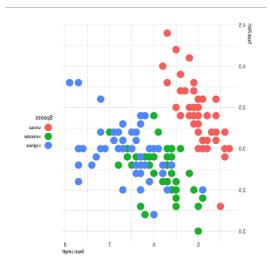


spider plot

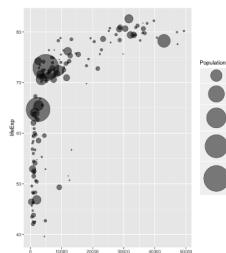


line

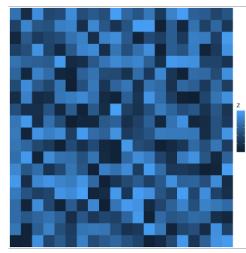
# Relationship



scatterplot

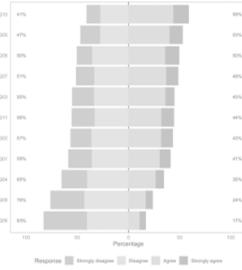


bubble plot

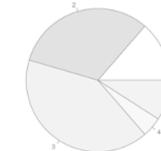


heatmap

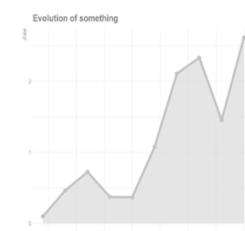
# Part of the whole



barplot

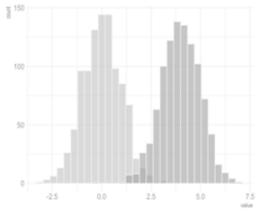


pie chart



area plot

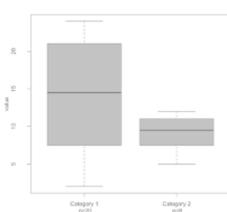
# Distribution



histogram

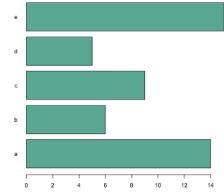


density plot

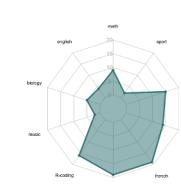


boxplot

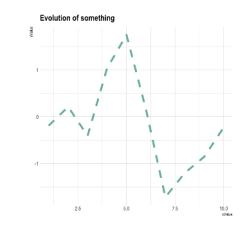
# Comparison



barplot

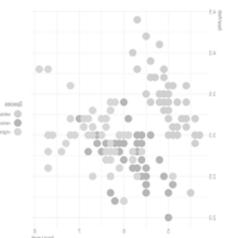


spider plot

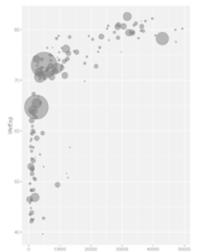


line

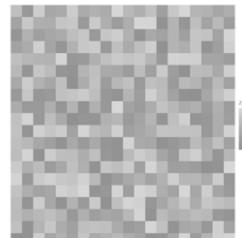
# Relationship



scatterplot

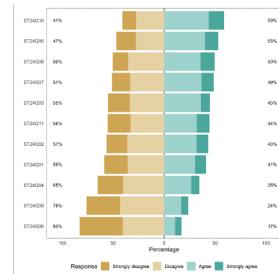


bubble plot

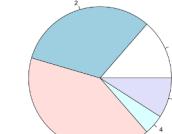


heatmap

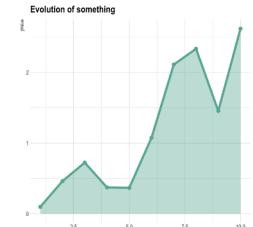
# Part of the whole



barplot



pie chart



area plot

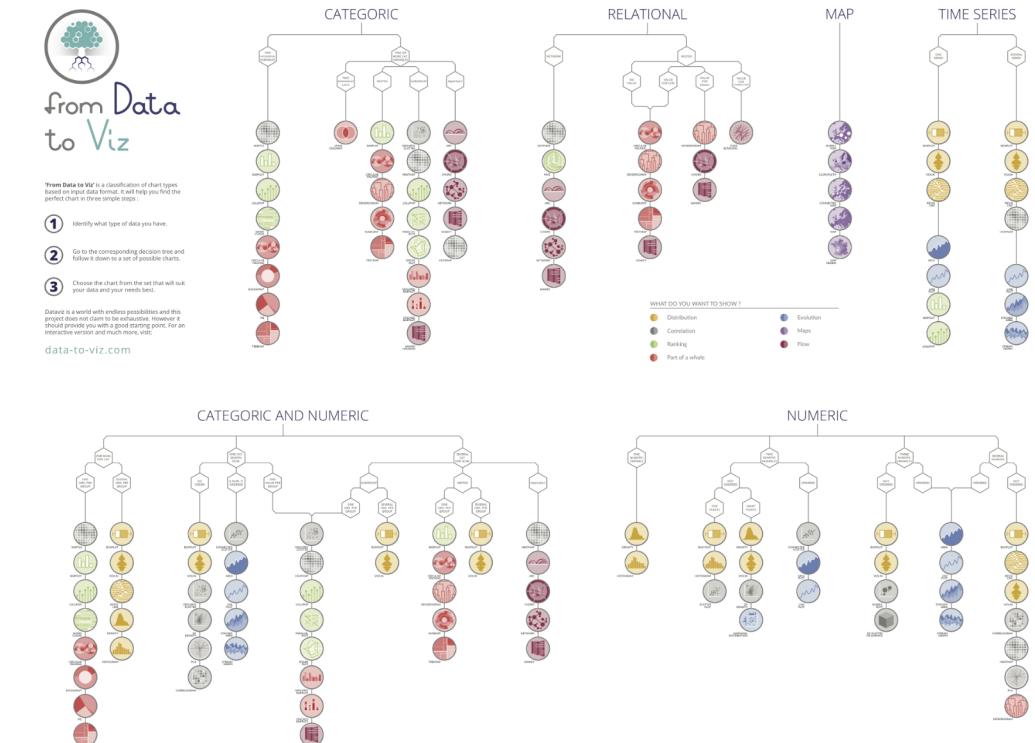
# Graph Types and Data Types

Numeric/Categorical

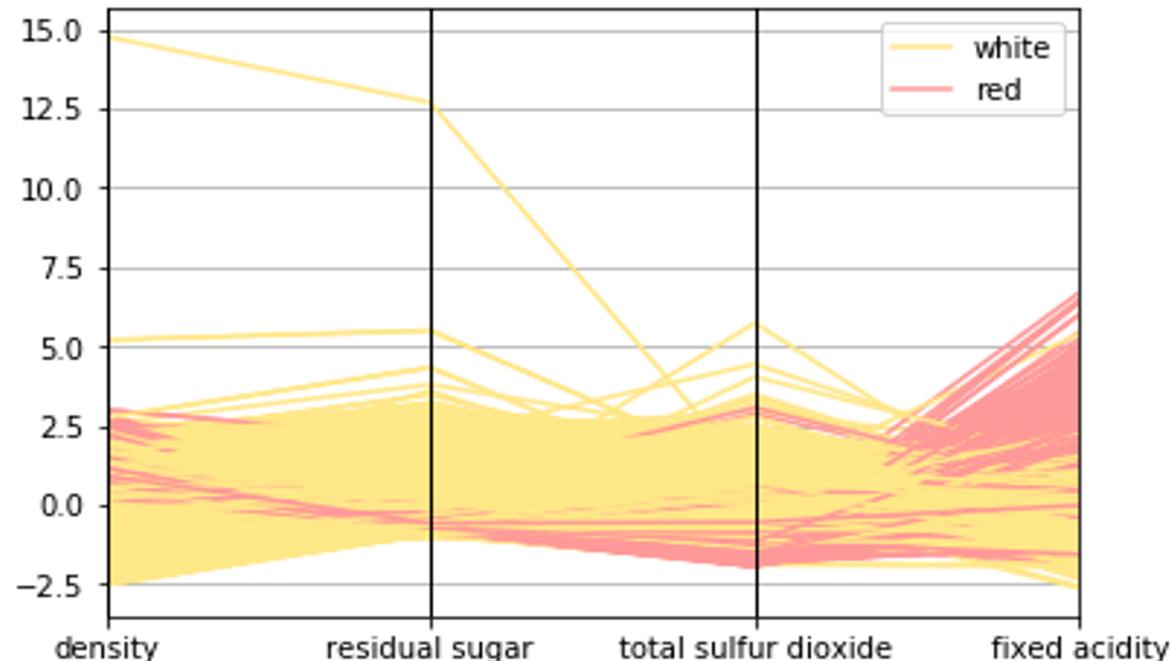
Is it ordered?

How many observations  
per group?

<https://www.data-to-viz.com/>

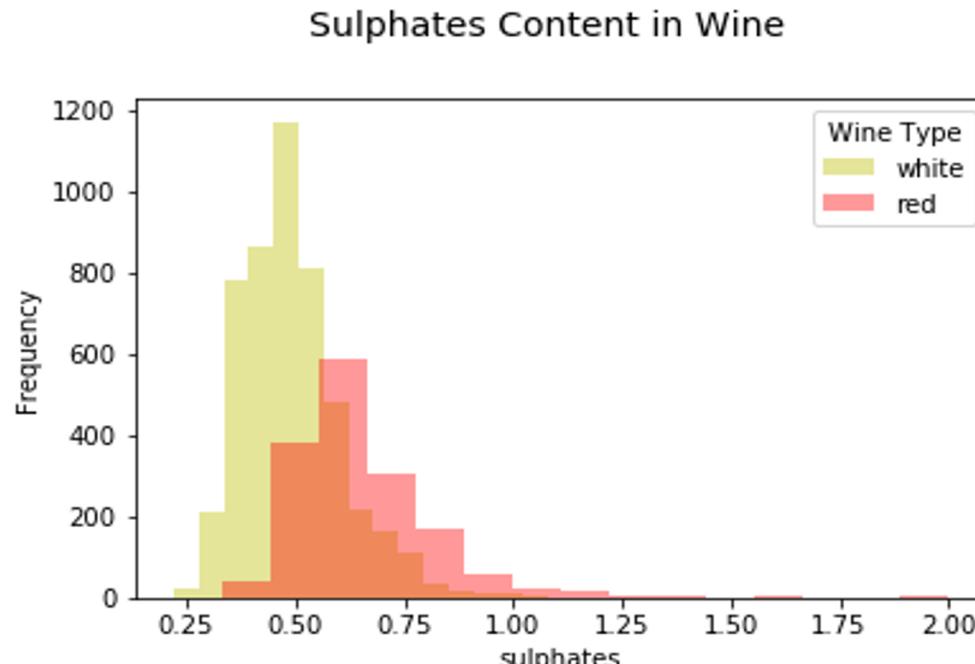


# Multiple dimensions I



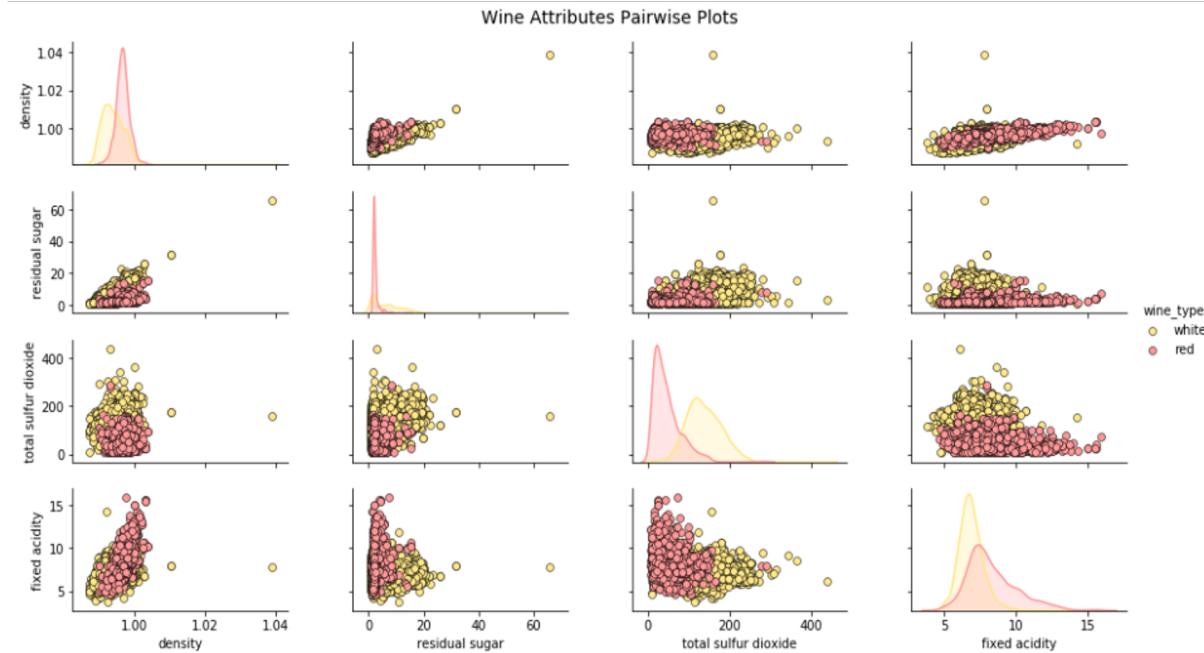
graph types

# Multiple dimensions II



colours

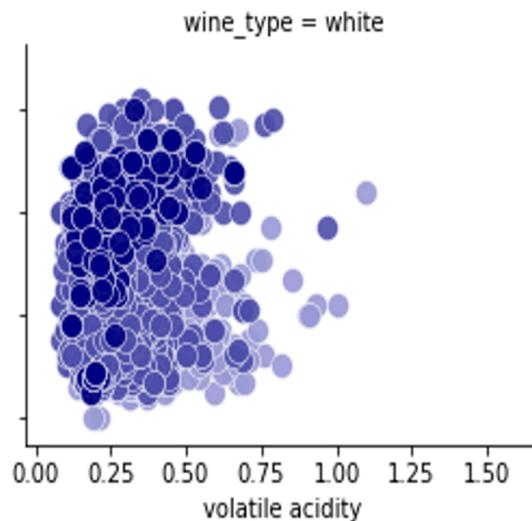
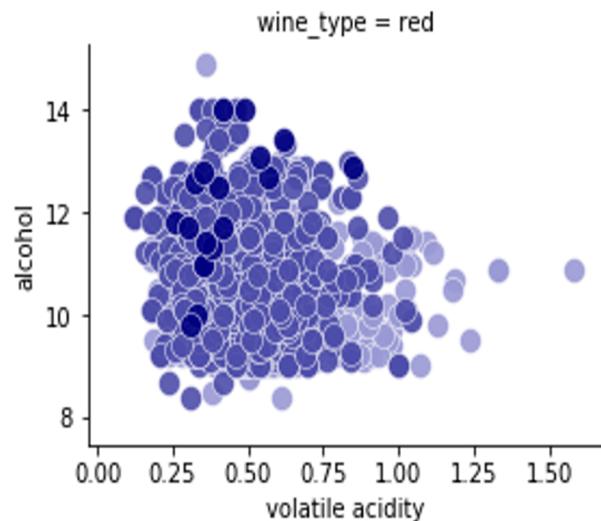
# Multiple dimensions III



pairwise correlations

# Multiple dimensions IV

Wine Type - Alcohol - Quality - Acidity



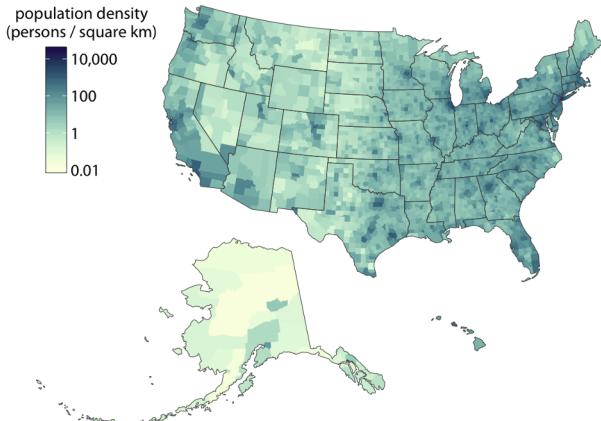
Wine Quality Class

- low
- medium
- high

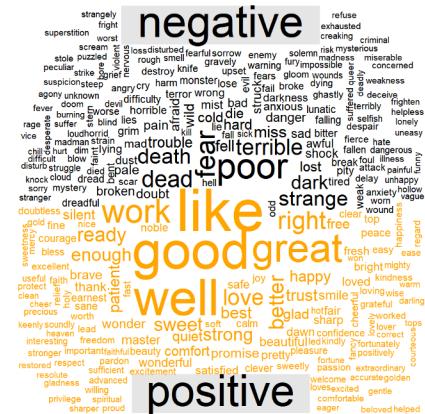
facets

# Other Graph Types

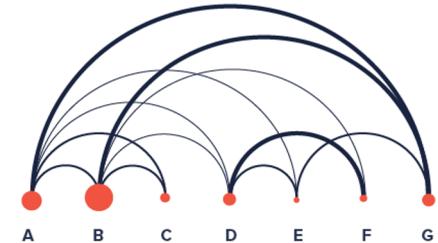
geospatial data



wordcloud



connections



# Having good data is more important than choosing a right graph type

Ideally, data should be:

- ✓ High volume
- ✓ Historical
- ✓ Consistent
- ✓ Clean
- ✓ Clear
- ✓ Richly segmented

# Long vs Wide formats

“Long” format

country	year	metric
x	1960	10
x	1970	13
x	2010	15
y	1960	20
y	1970	23
y	2010	25
z	1960	30
z	1970	33
z	2010	35

“Wide” format

country	yr1960	yr1970	yr2010
x	10	13	15
y	20	23	25
z	30	33	35

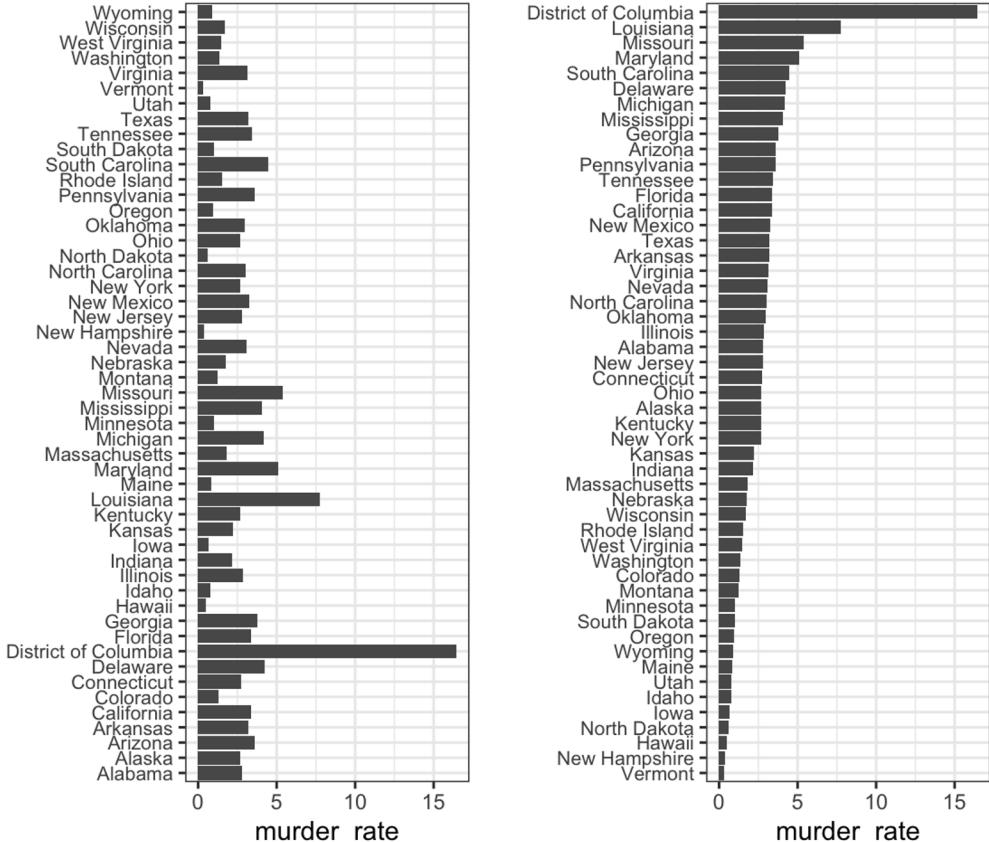
# INSTRUMENTS

Most of the data visualization software allows you to:

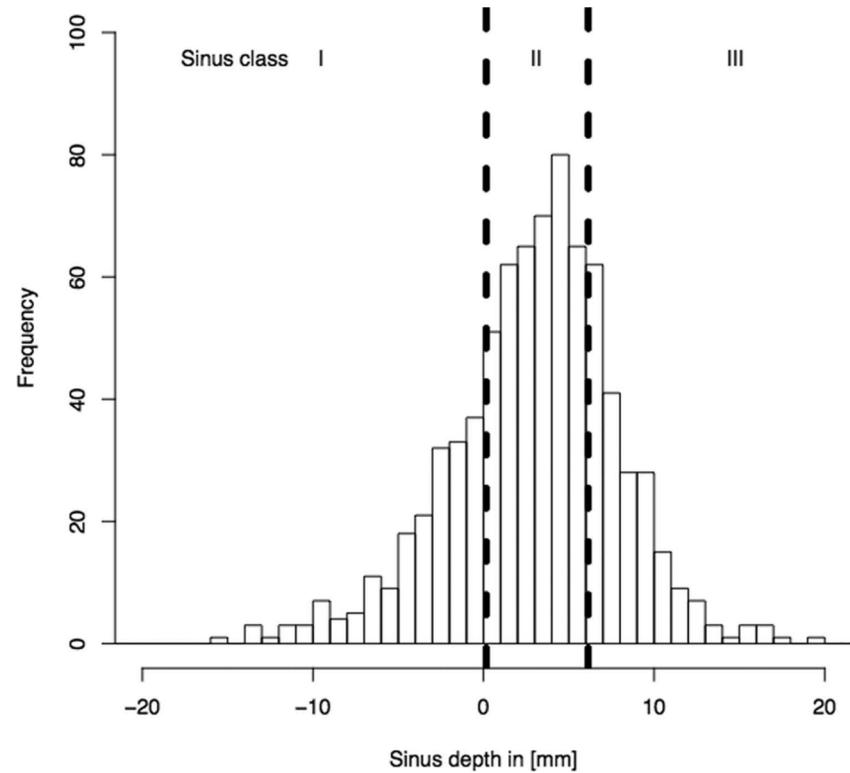
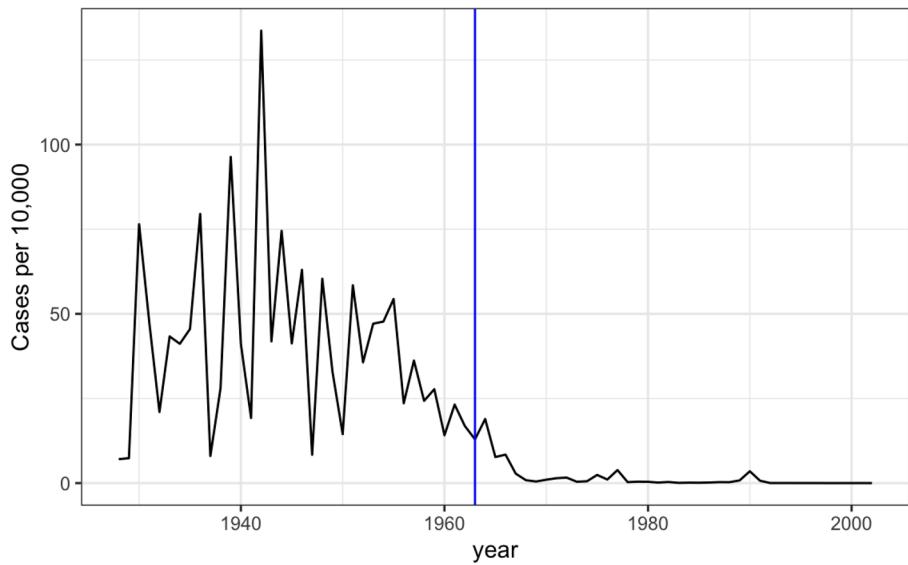
- Aggregate
- Filter
- Compare
- Annotate

and much more!

# Data ordering

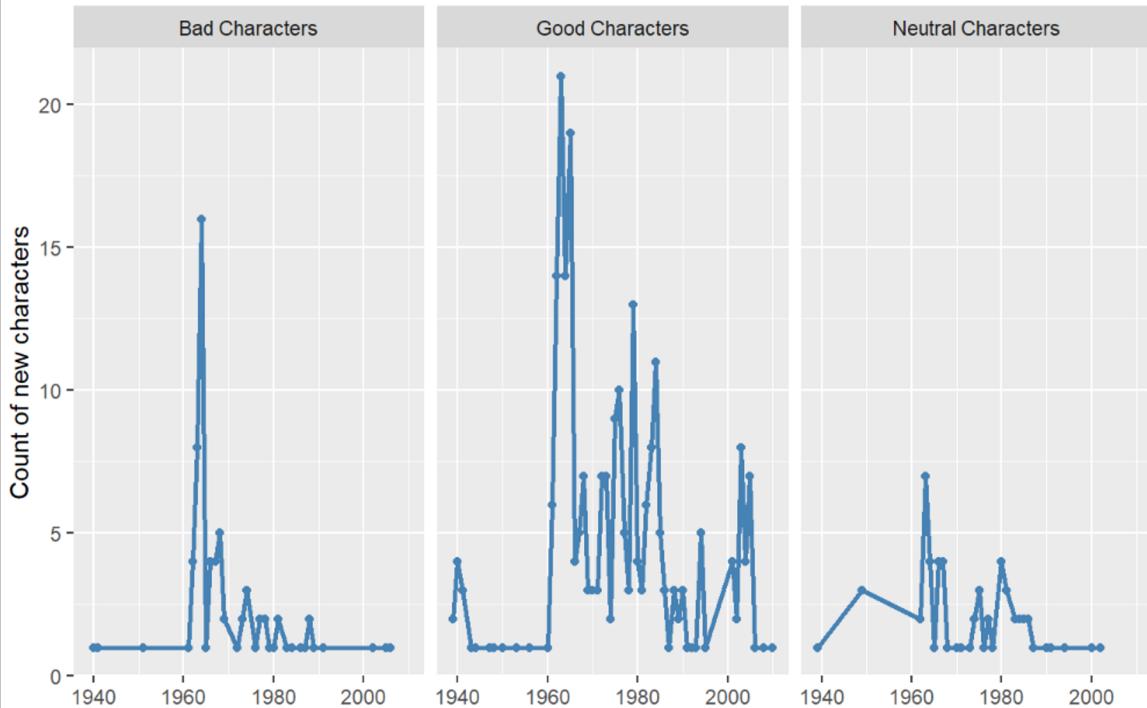


# Reference Lines and Regions

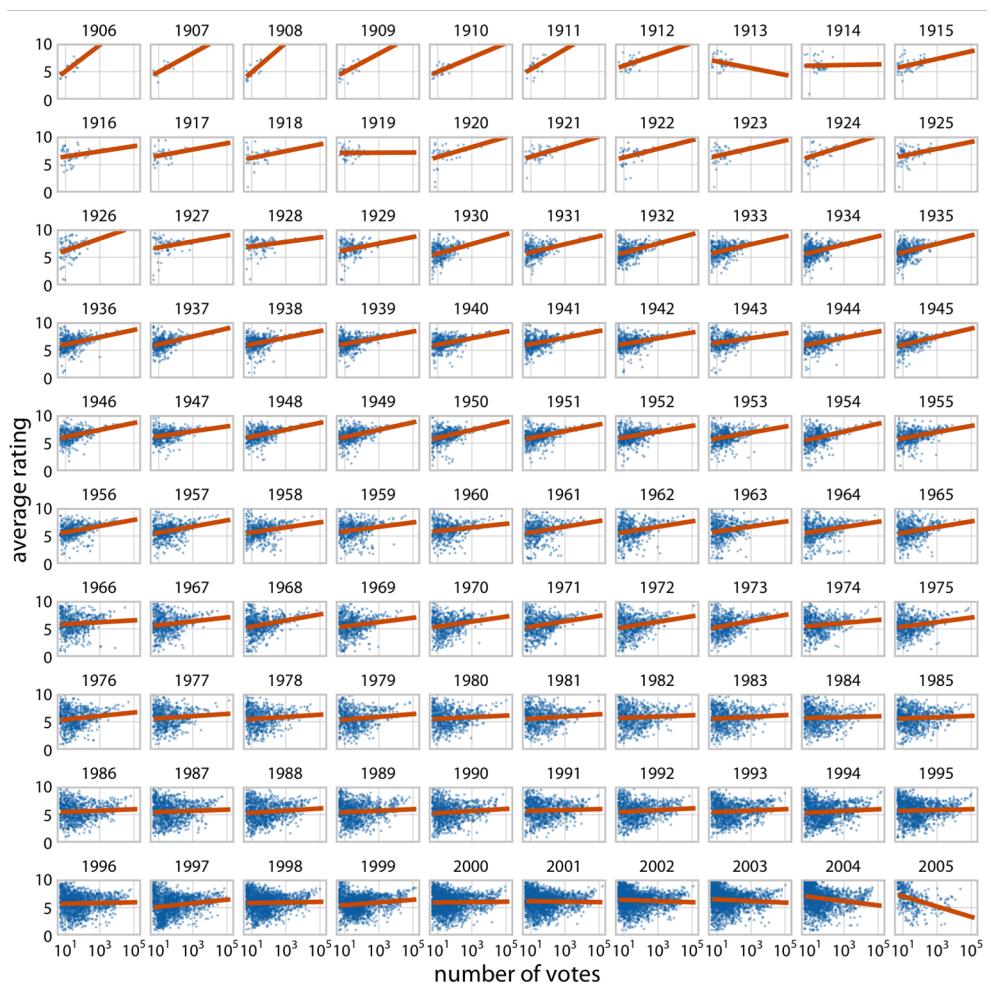


# Facets

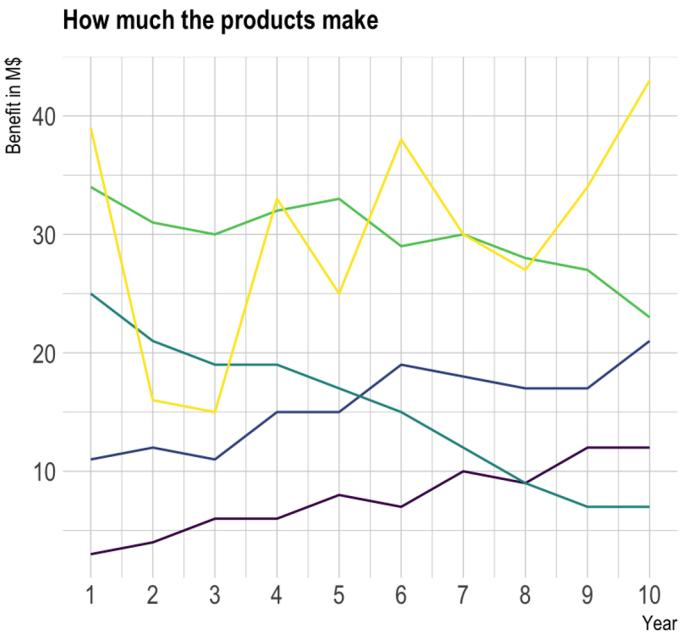
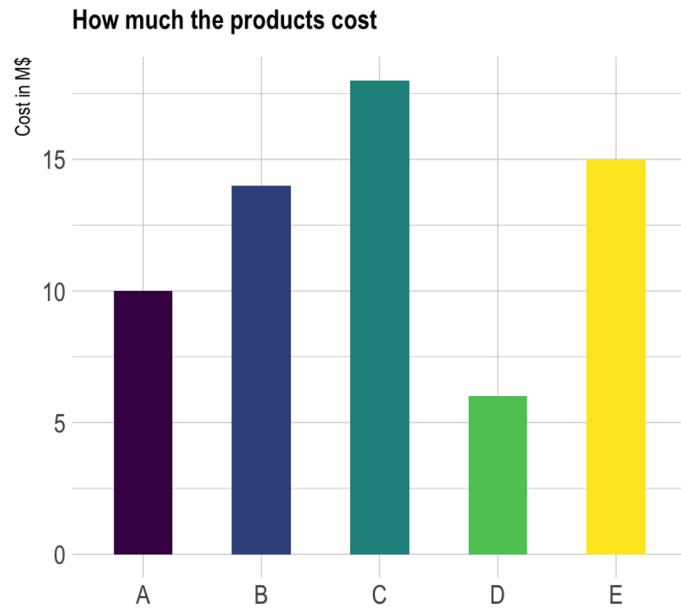
New Marvel characters by alignment  
(limited to characters with more than 100 appearances)



Useful for  
multidimensional  
or large scale data



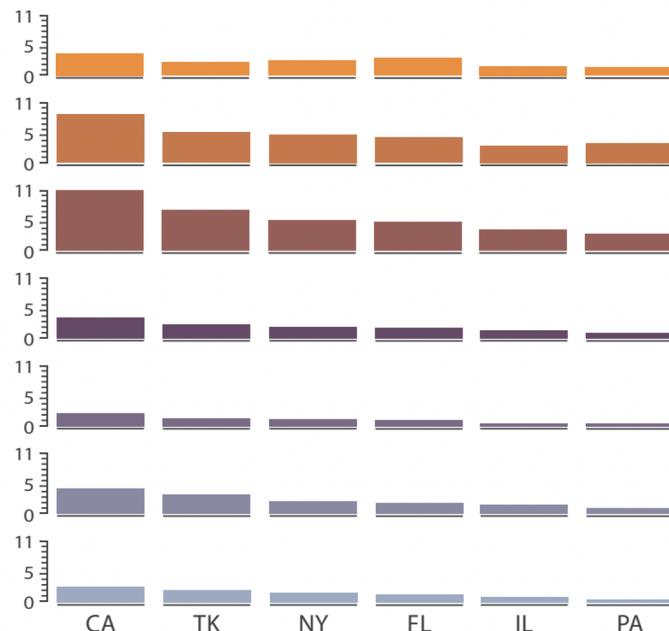
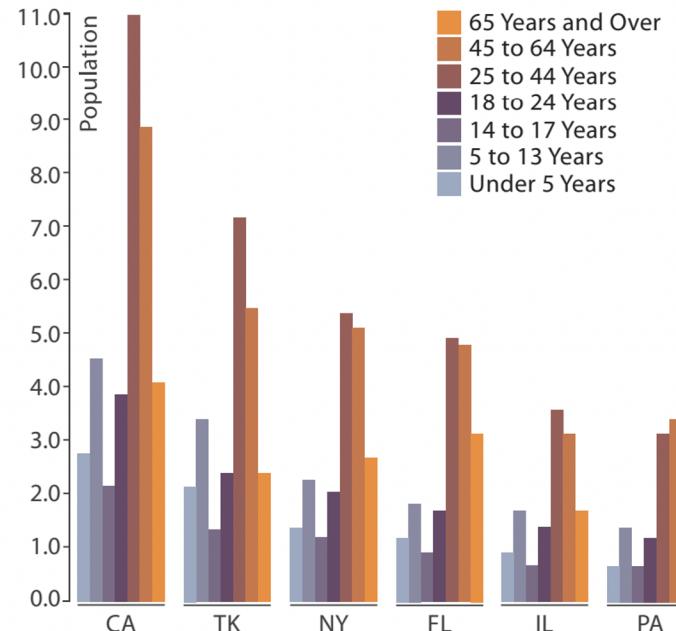
# Multiple Views



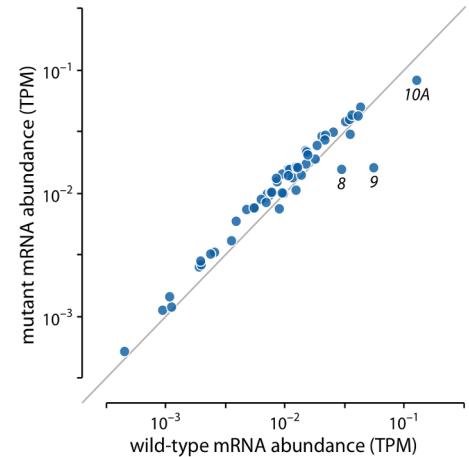
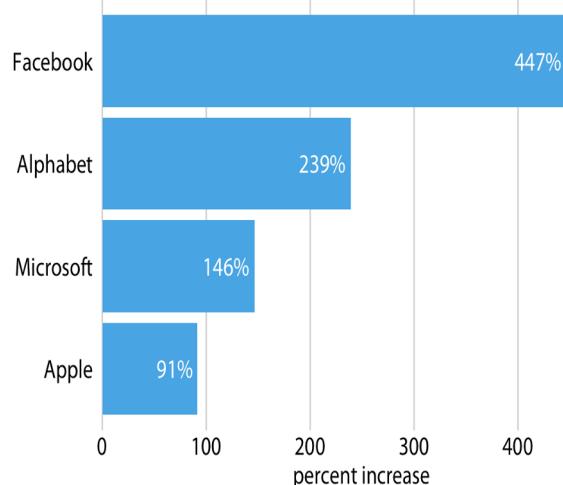
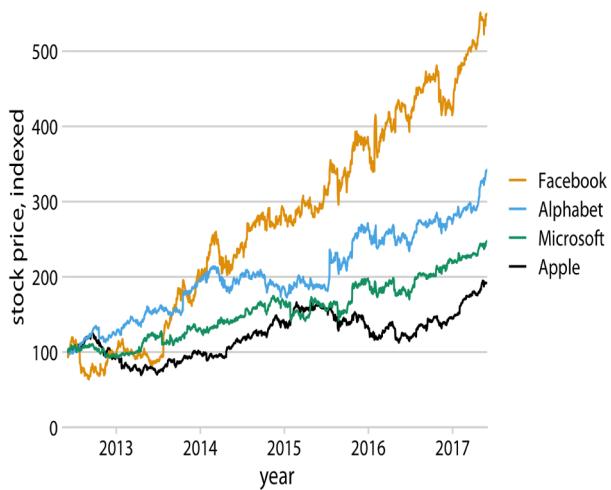
product

- A
- B
- C
- D
- E

# Multiple Views



# Context



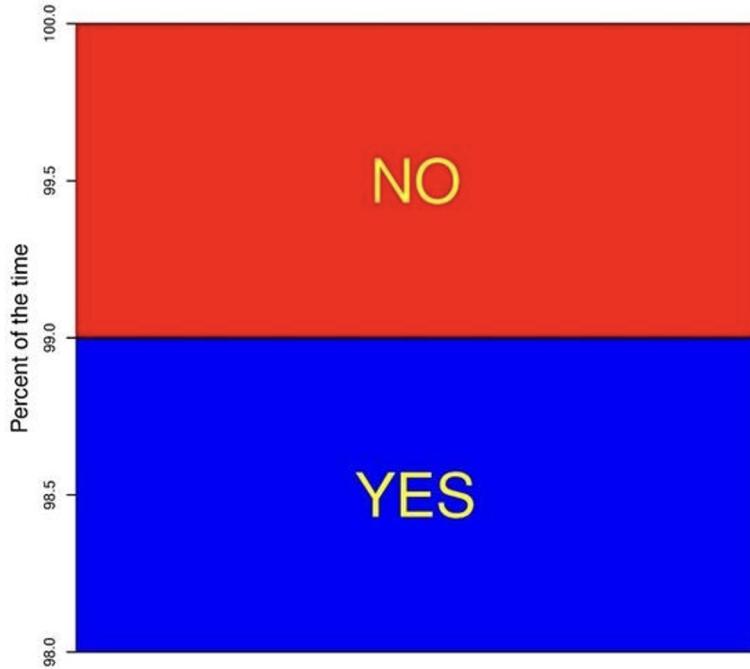
Perpendicular grid lines are  
the most useful

For paired data,  
use a diagonal line

# DESIGN PRINCIPLES

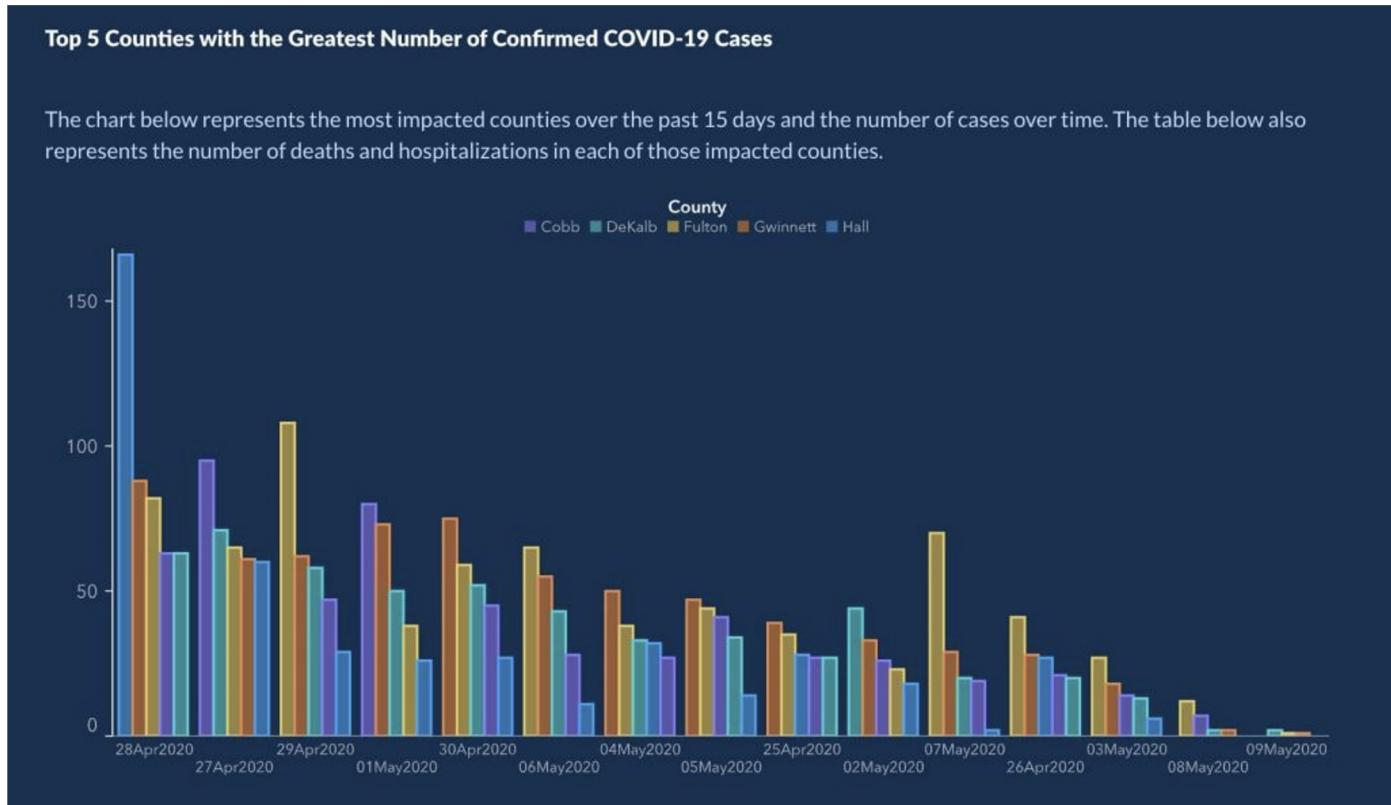
# Cutting Y-scales

Is truncating the Y-axis misleading?



If cutting Y-axis (for example, when using Likert scale data), use **points** instead of bars

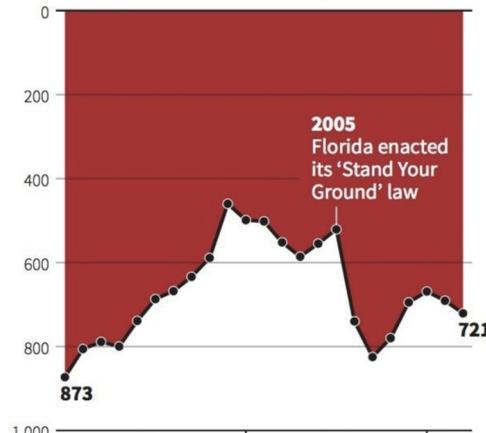
# Use reasonable X-axis



# flipped

## Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

# dual



# Avoid misleading axes!

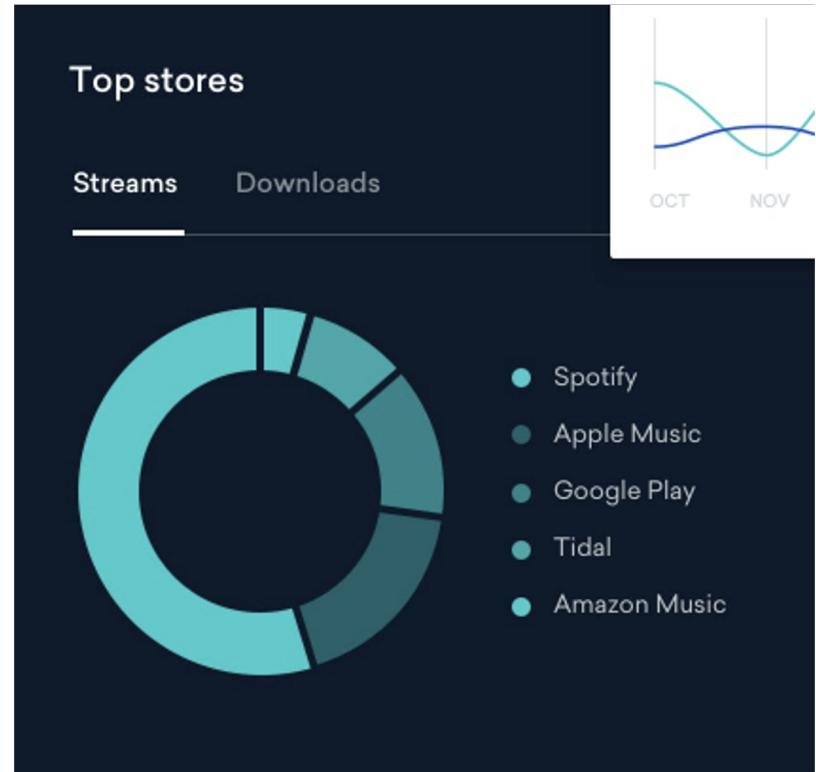
# Colour Use

Gradient and diverging palettes

Colorblind-friendly

Do not add more than 8 colours

Use gray for context



# Creating Palettes

SANFORD AND SELNICK

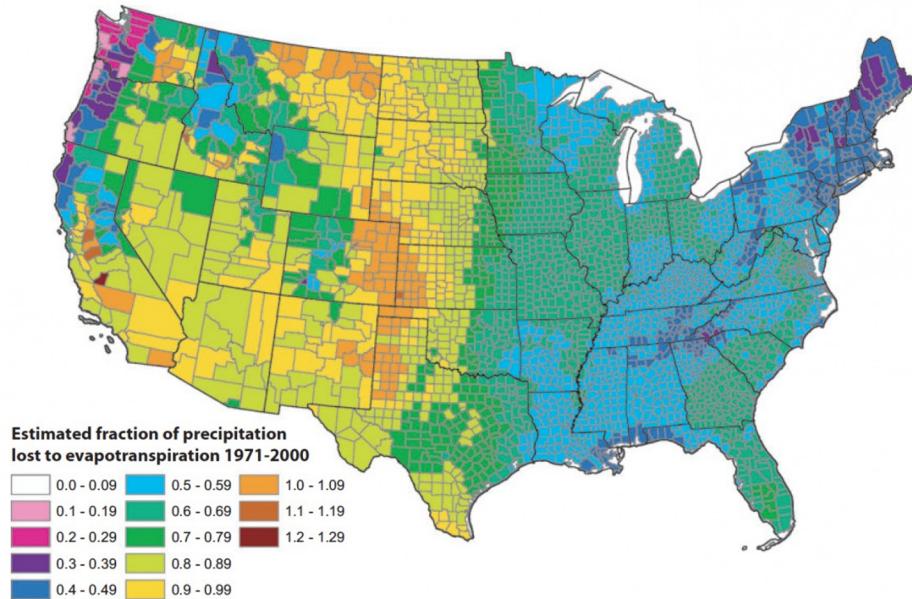


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ( $P$ ) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of  $ET/P$  were made first at the 800-m resolution of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

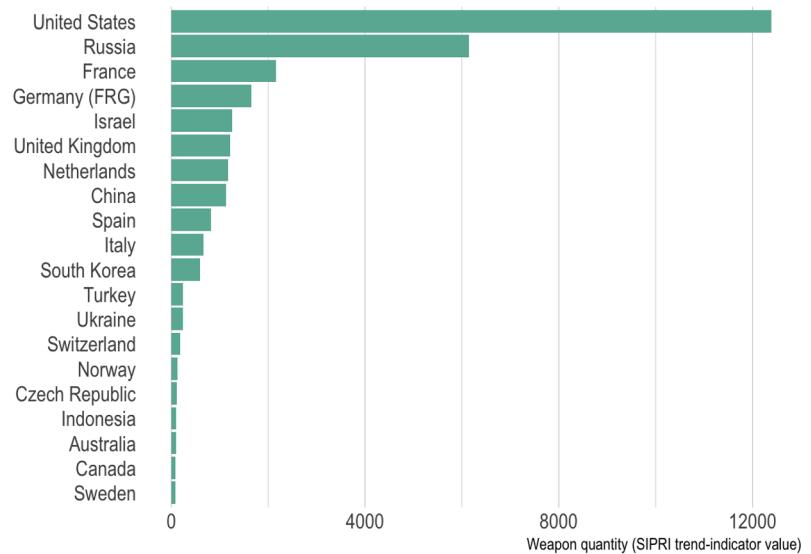
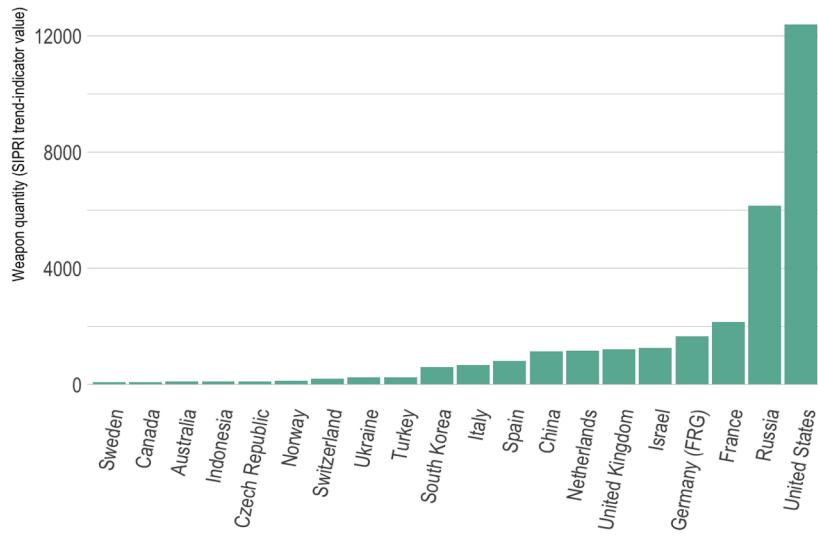
**Use** meaningful colors

**Use** color generators

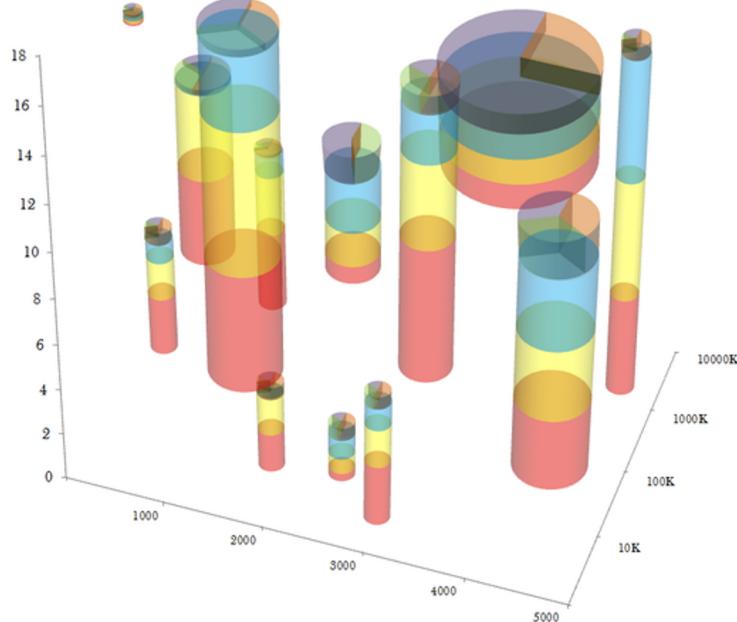
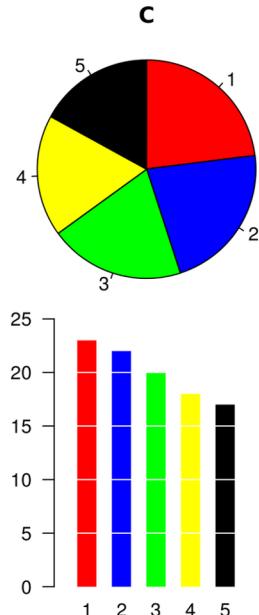
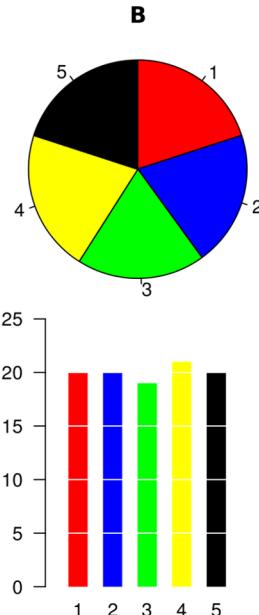
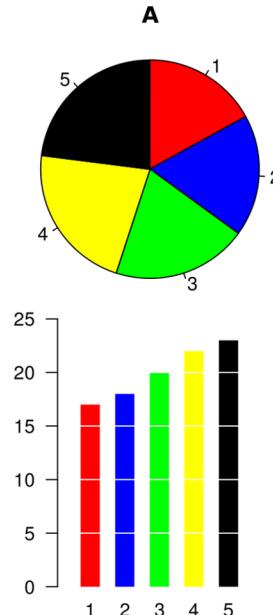
**Avoid** pure and bright  
colours

**Avoid** rainbow palette

# Use horizontal bar chart



# Avoid pie charts and 3D



# Tables

1. Remove style and borders
2. **Do not use vertical lines**, remove all grids if possible
3. **Left align your text, BUT right align your numbers**
4. Align headers correspondingly
5. Group and separate

Company	Last Trade	Trade Time	Change	Prev Close
GOOG Google Inc.	597.74	12:12PM	14.81 (2.54%)	582.93
AAPL Apple Inc.	378.94	12:22PM	5.74 (1.54%)	373.20
AMZN Amazon.com Inc.	191.55	12:23PM	3.16 (1.68%)	188.39
ORCL Oracle Corporation	31.15	12:44PM	1.41 (4.72%)	29.74
MSFT Microsoft Corporation	25.50	12:27PM	0.66 (2.67%)	24.84
CSCO Cisco Systems, Inc.	18.65	12:45PM	0.97 (5.49%)	17.68
YHOO Yahoo! Inc.	15.81	12:25PM	0.11 (0.67%)	15.70

# **SOFTWARE**

# Data Visualization Tools



Many features  
Easy to learn & use  
No data cleaning  
Data security



Data cleaning  
Popular  
Confusing  
Lacks customization



Free  
Data exploration  
Requires learning  
Debugging