

Data Visualization

Aviral (Avi) Chawla | achawla1@uvm.edu | Vermont Complex Systems Institute



Visualization Pipeline

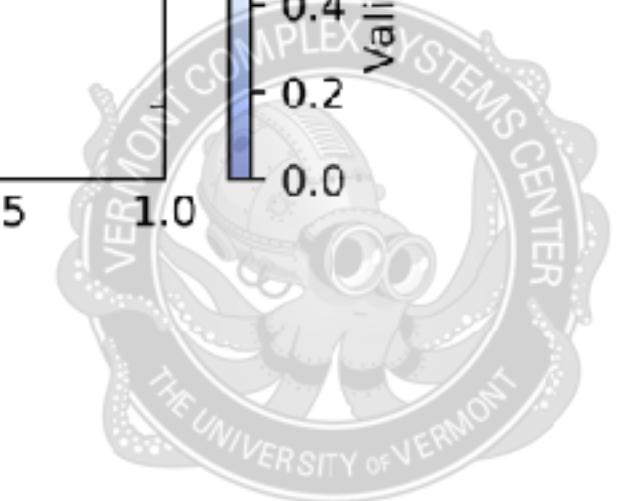
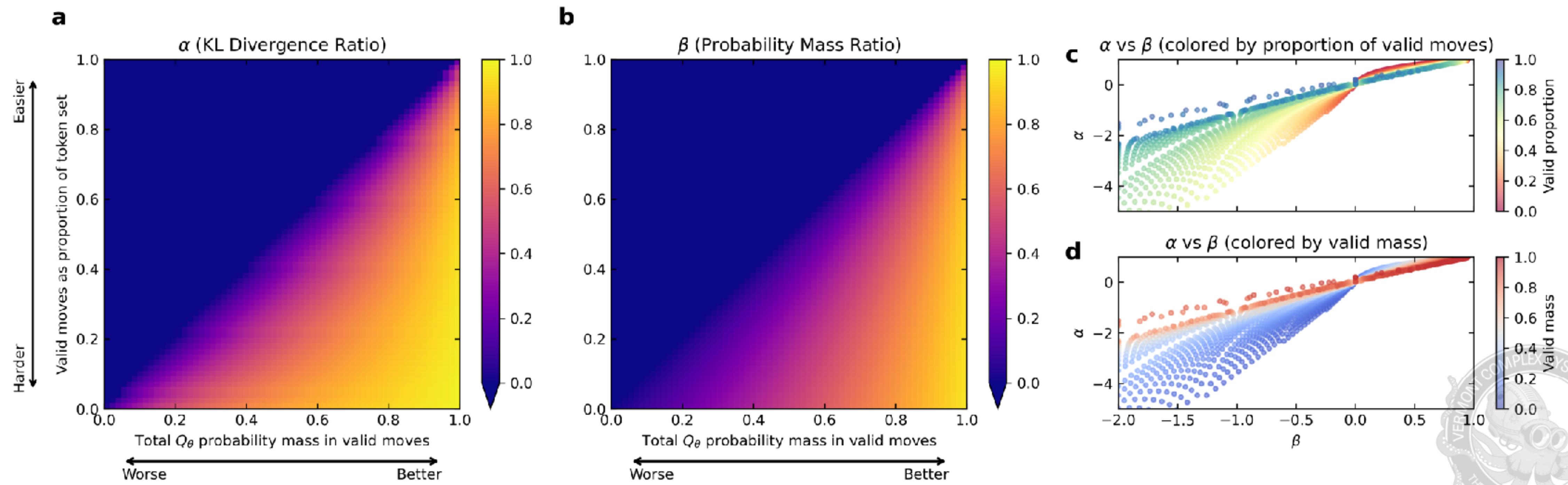
- 1. Determine goals**
- 2. Choose visualization type**
- 3. Build a prototype**
- 4. Feedback & Iterate**

Visualization Pipeline

1. Determine goals (Why make visualizations?)
2. Choose visualization type
3. Build a prototype
4. Feedback & Iterate

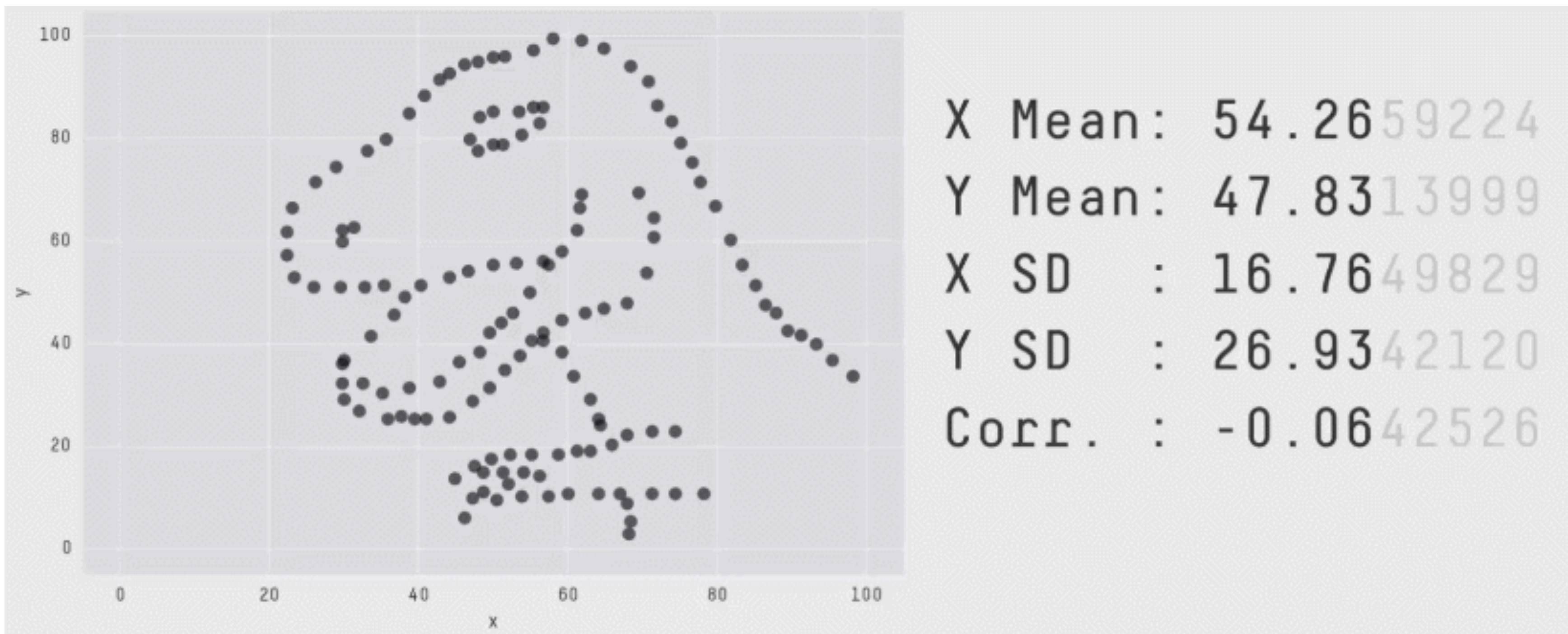
Why make visualizations?

We make visualizations to **a) explore**, **b) analyse**, and **c) communicate**.

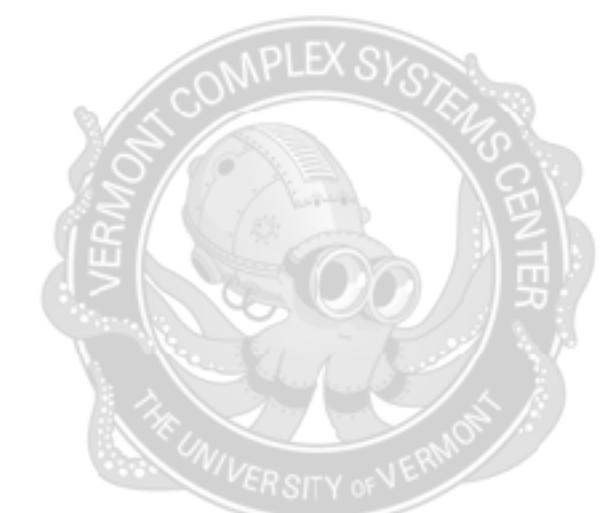


a) Explore

Understand patterns & Outliers

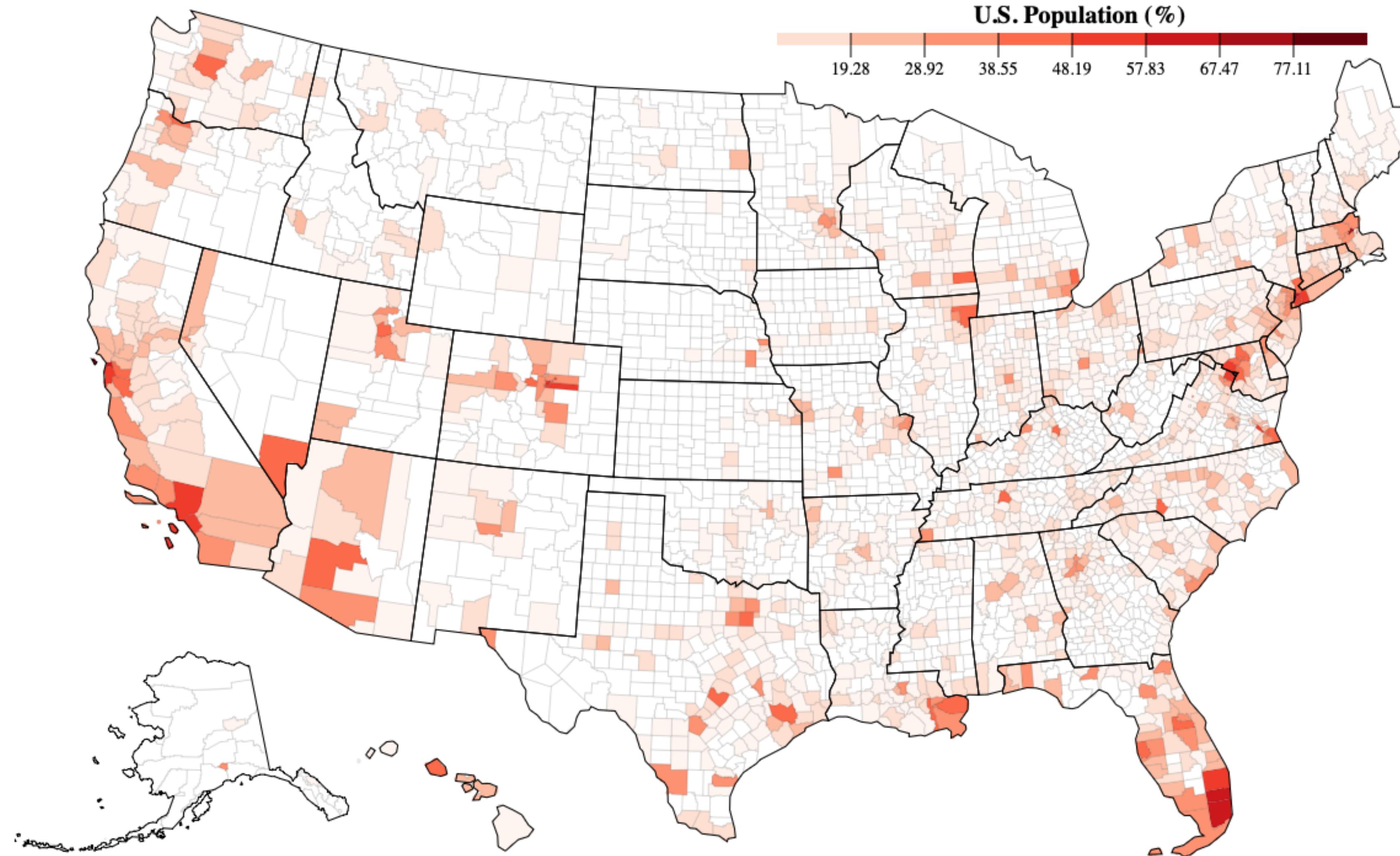


Anscombe's Quartet (1973)



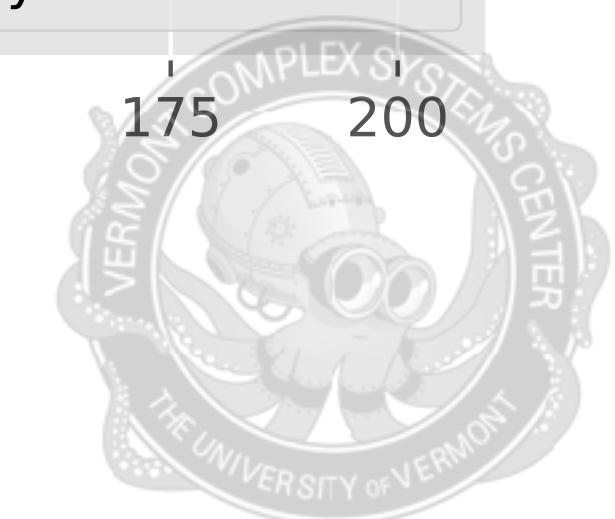
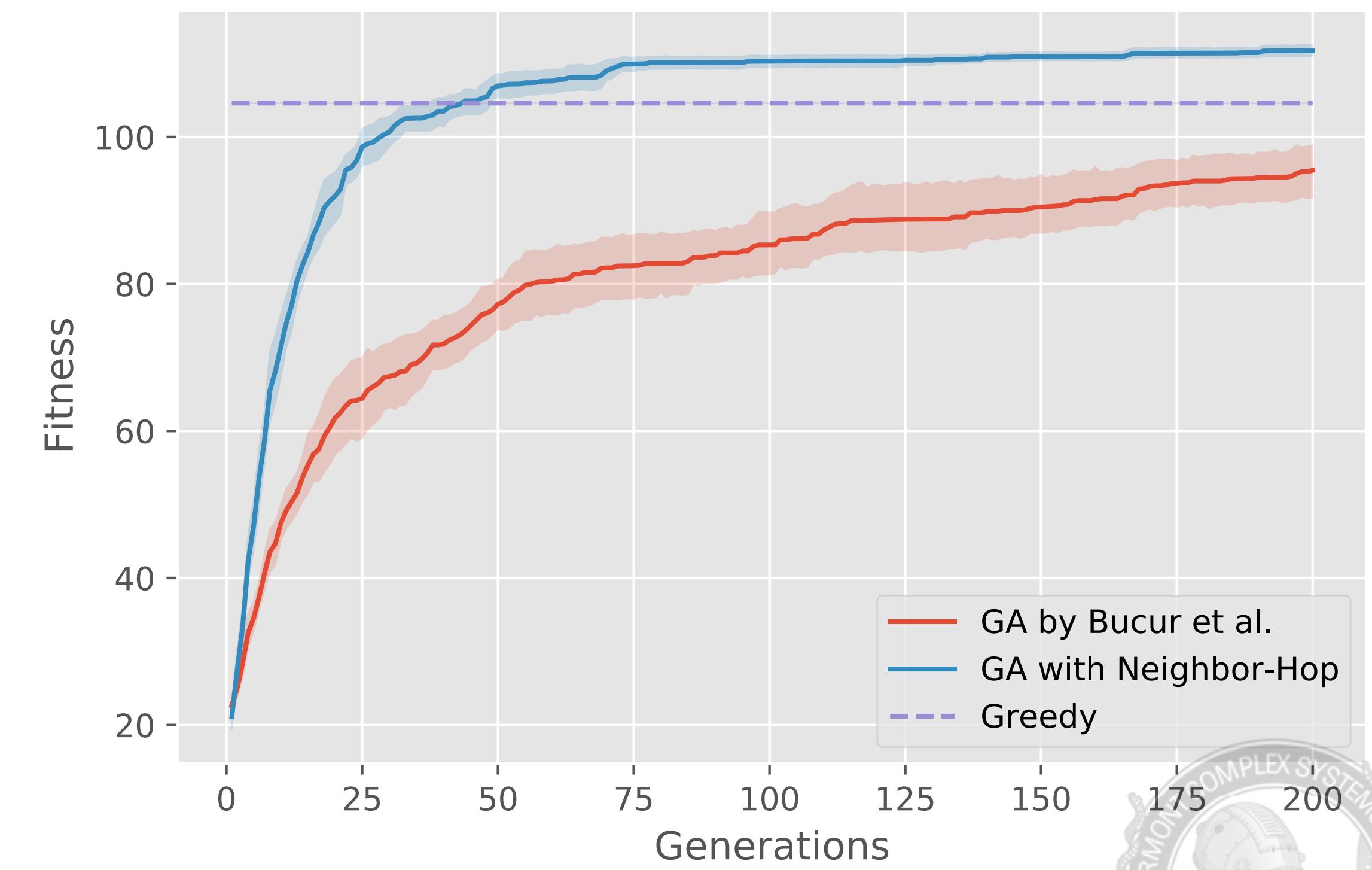
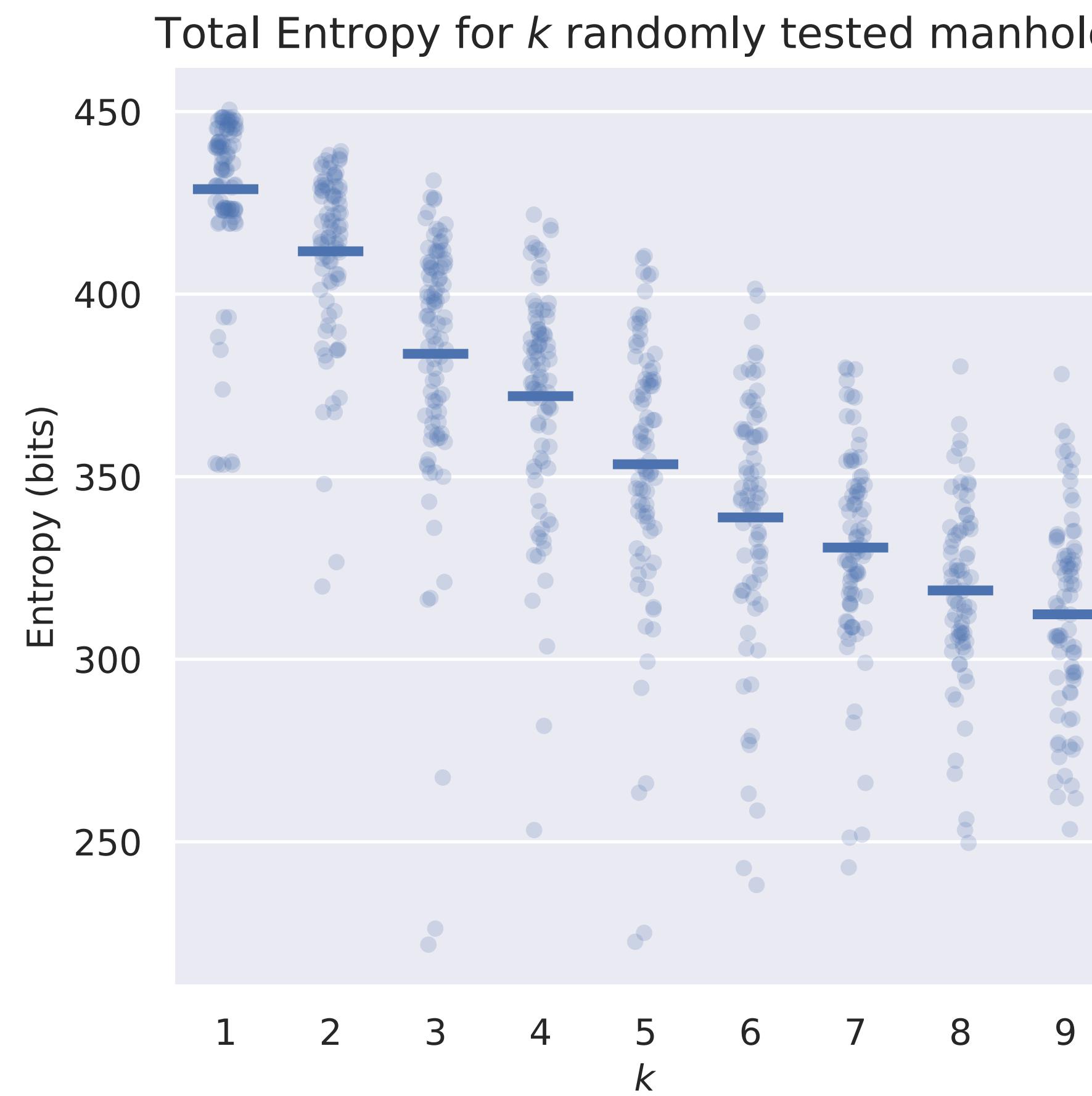
a) Explore

Understand patterns & Outliers



b) Analyse

Answer well-defined questions

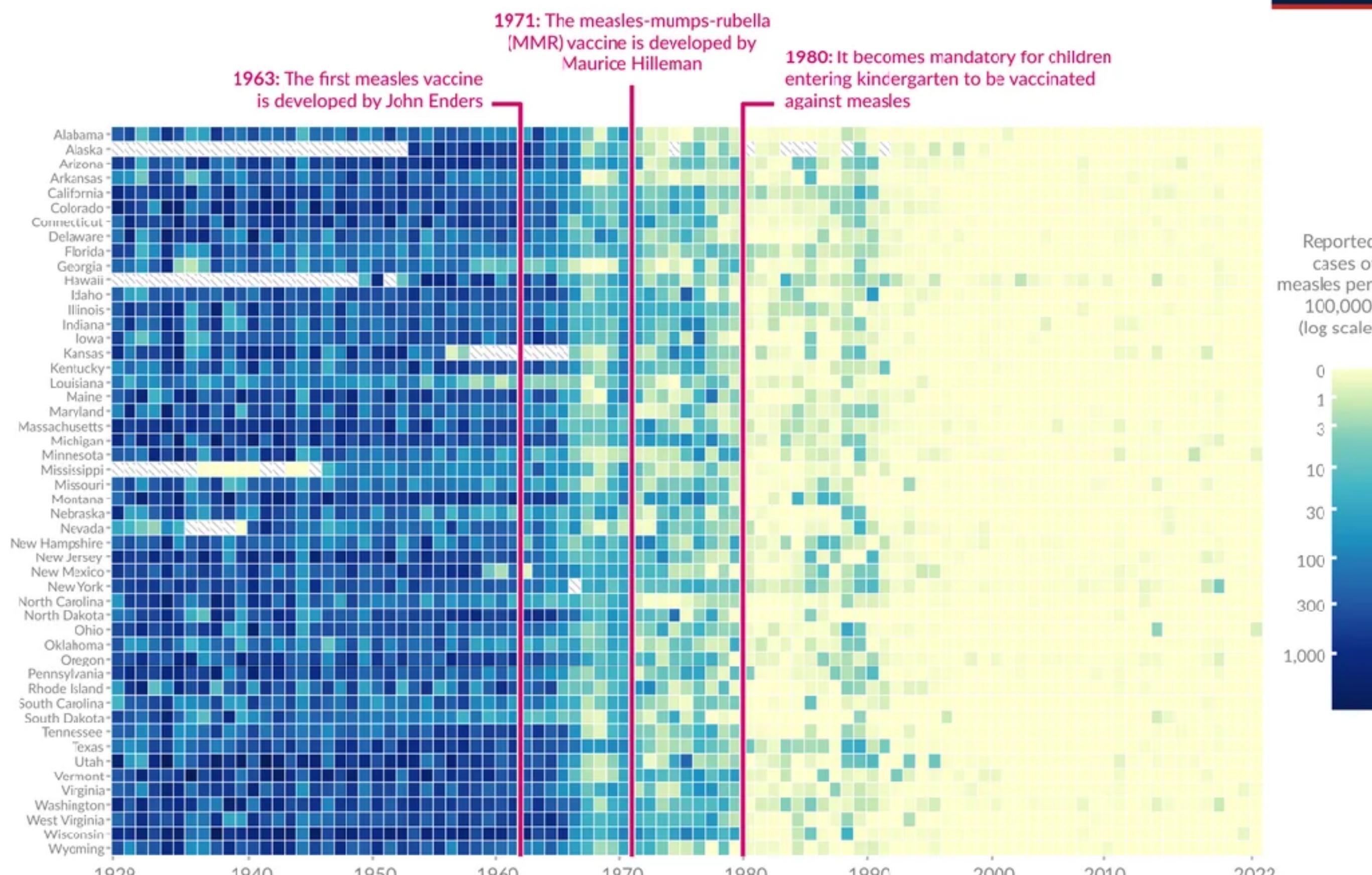


c) Communicate

Inform or share a story

Vaccines reduced measles cases across US states

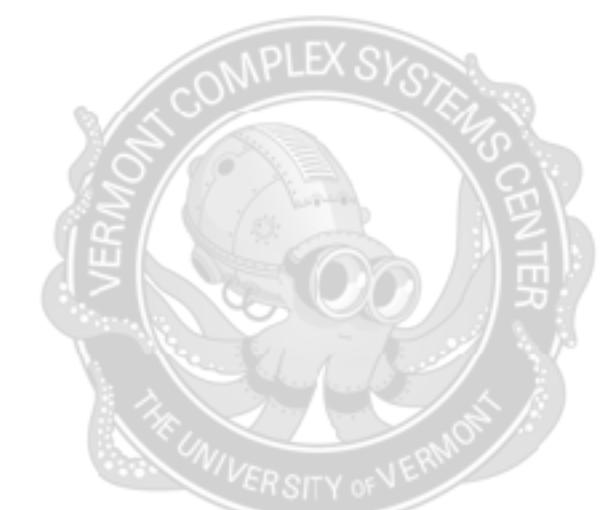
Our World
in Data



Data source: Project Tycho (2018); Centers for Disease Control and Prevention (1959–2022)

OurWorldInData.org — Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Fiona Spooner



You don't visualize if

- You have a well defined question on a well-defined dataset with a single answer
 - What is the current unemployment rate?
- No human intervention necessary / possible
 - Is this bottle broken?
 - What shows to recommend?
 - High frequency trading



Visualization Pipeline

1. Determine goals
2. Choose visualization type (Learning Fundamentals)
3. Build a prototype
4. Feedback & Iterate

Scaffoldings of a Visualization

Marks

Basic geometric elements that depicts items and links.

Marks as Items/Nodes

Points



Lines



Areas



Marks as Links

Containment



Connection



Channels

Control the appearance (properties) of marks. They either encode Categorical data or Quantitative data.

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

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: Categorical Attributes

Spatial region



Color hue



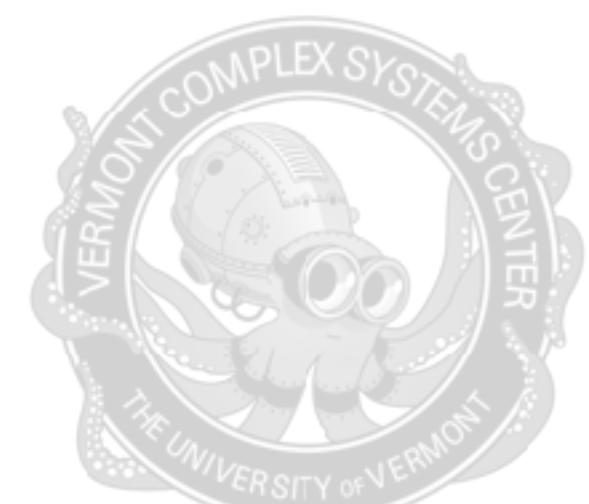
Motion



Shape

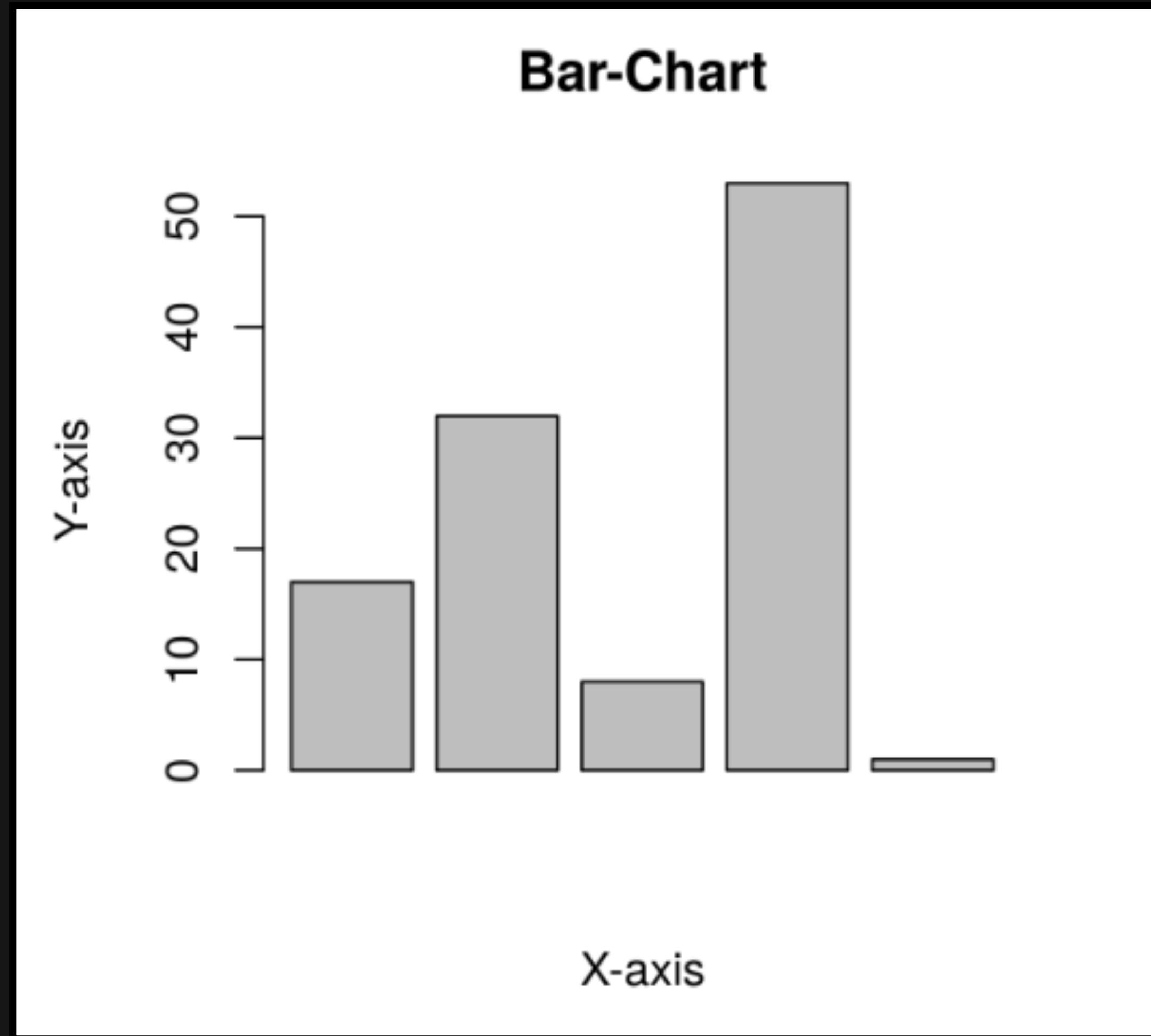


Effectiveness ↑ ↓ Least

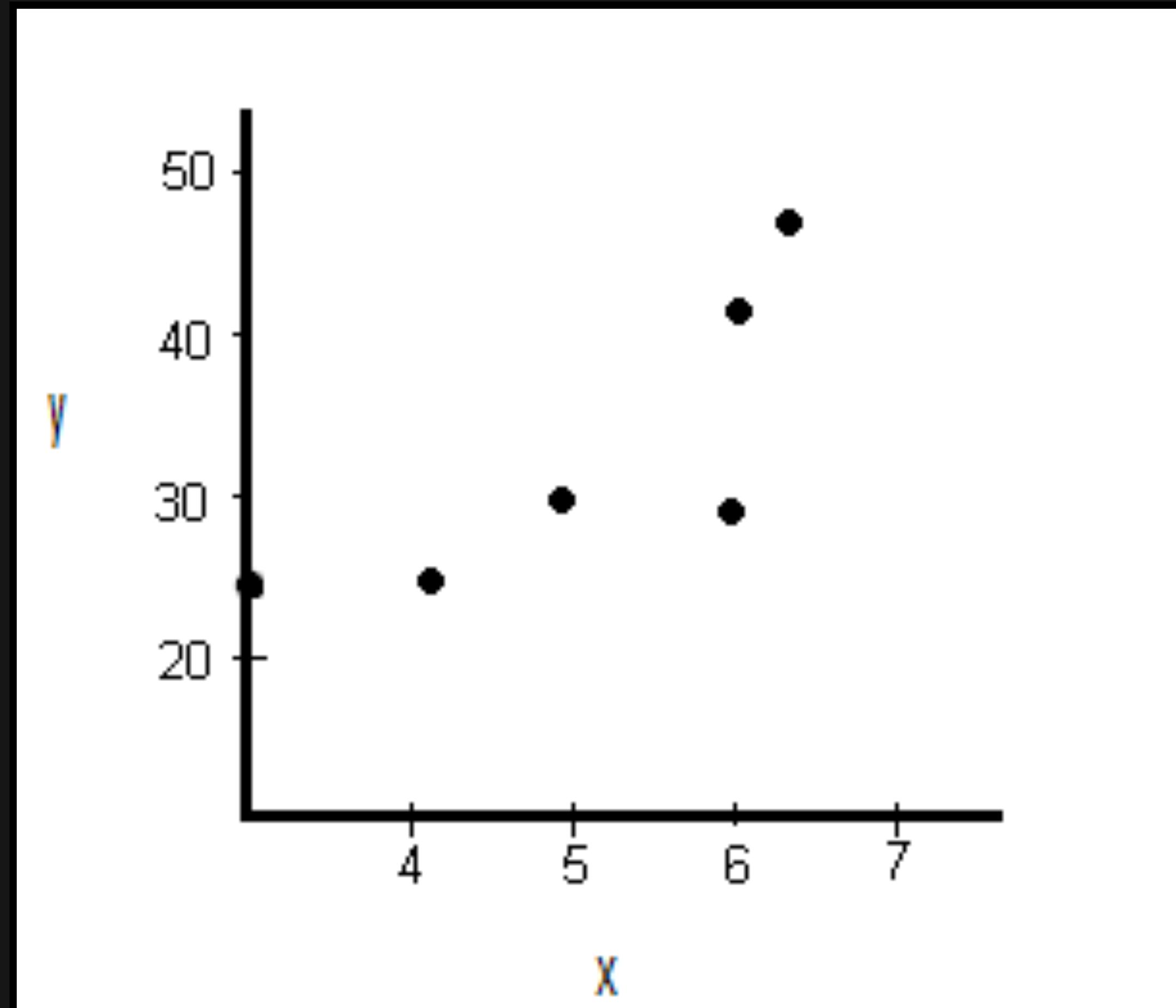


Quiz Time

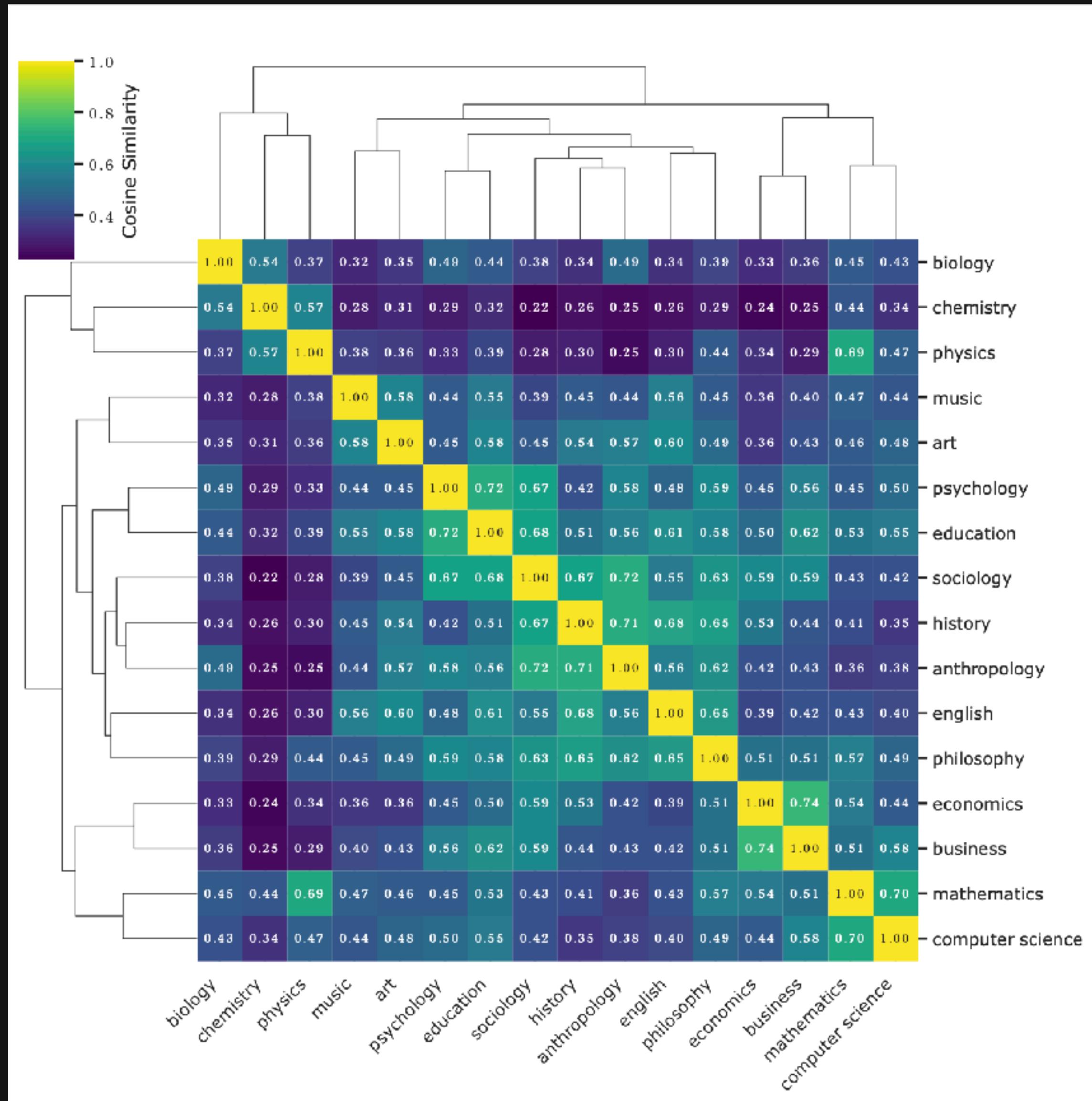
You have to identify appropriate marks and channels for the given visualization.



You have to identify appropriate marks and channels for the given visualization.



You have to identify appropriate marks and channels for the given visualization.



You have to identify appropriate marks and channels for the given visualization.

Scaffoldings of a Visualization

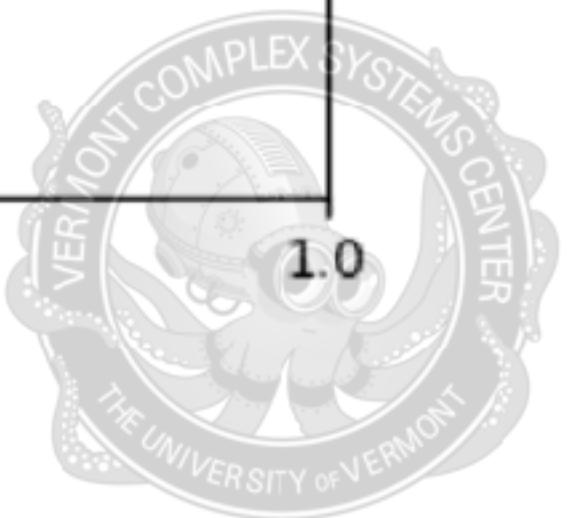
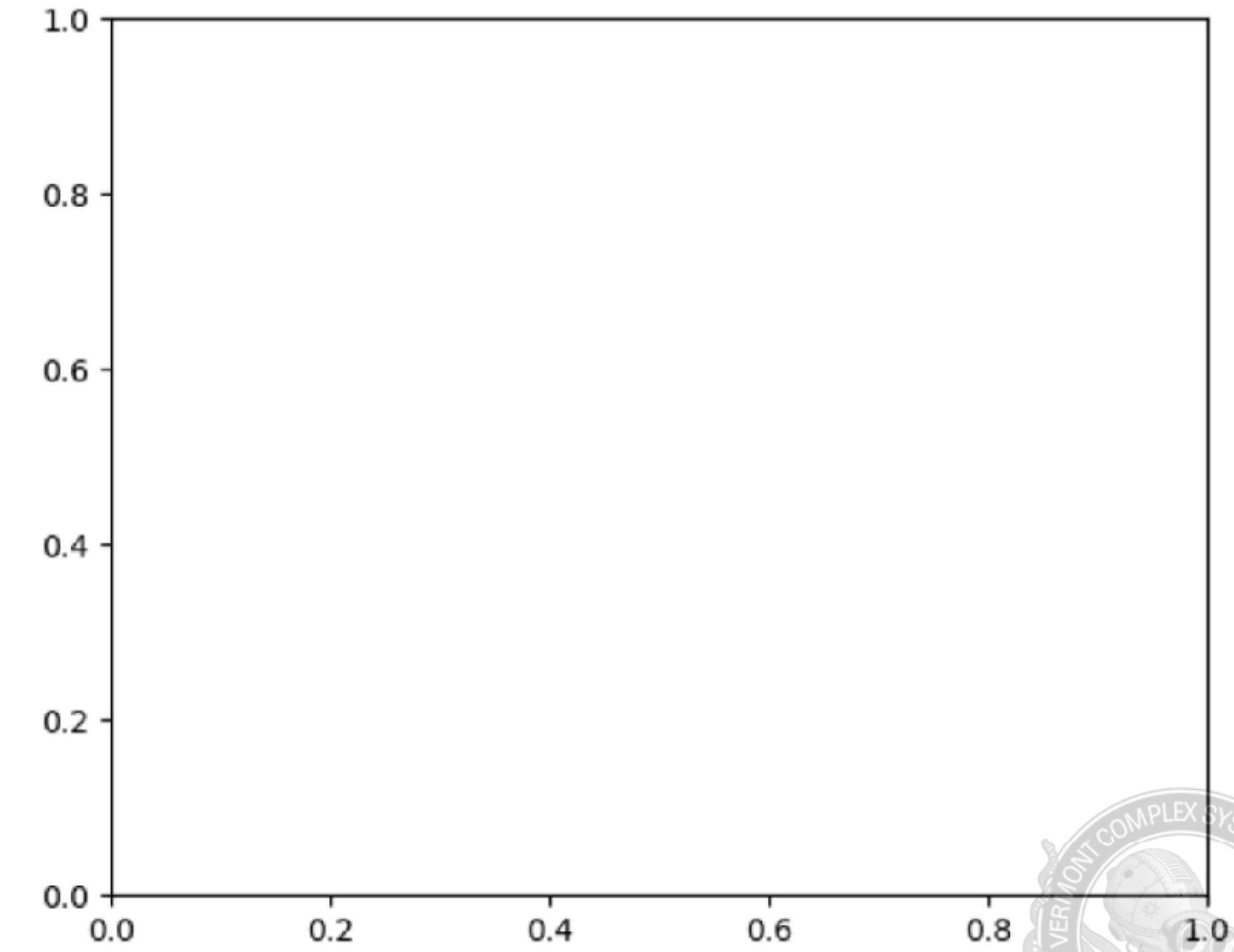
Using Axis & Scale

Axis

Reference lines(s) that provides a framework for plotting the data

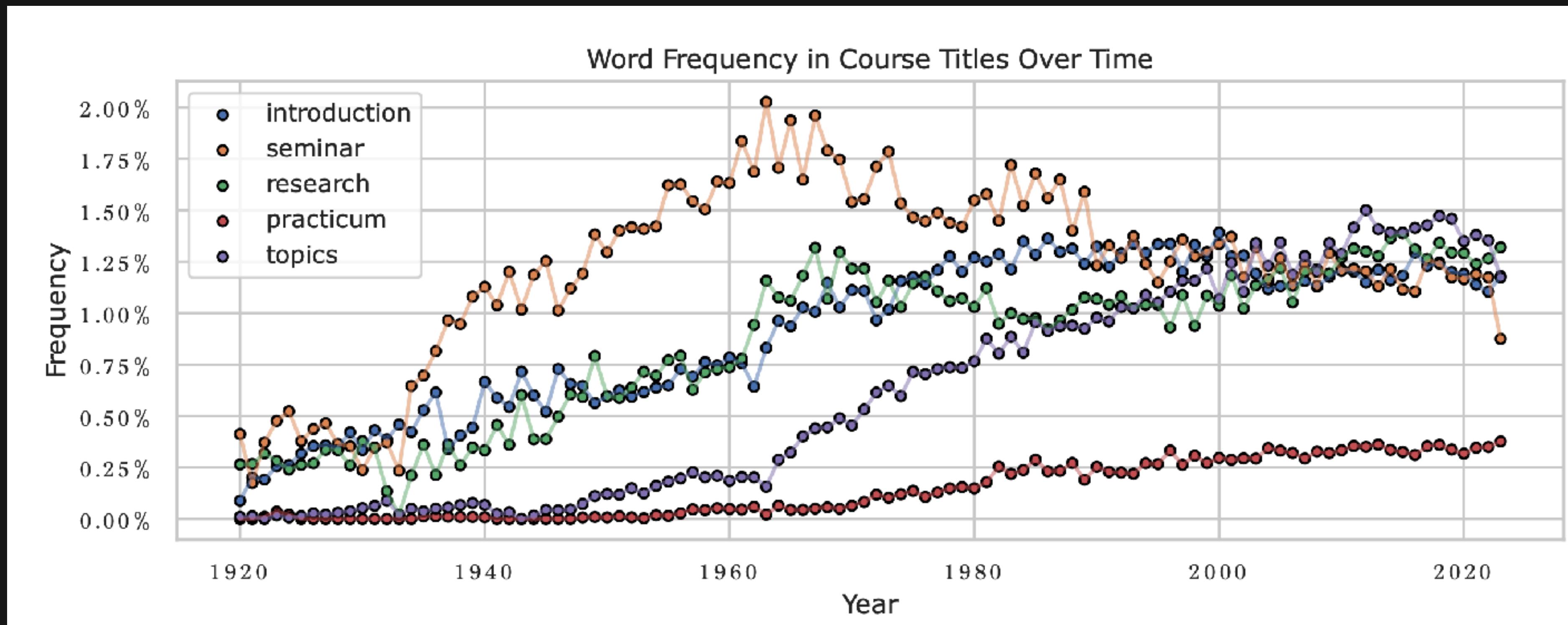
Scale

Represents the specific units or intervals by which the data is measured along



Quiz Time

What are the axis and scale used for the given visualization?



What are the axis and scale used for the given visualization?

Scaffoldings of a Visualization

Using Colors

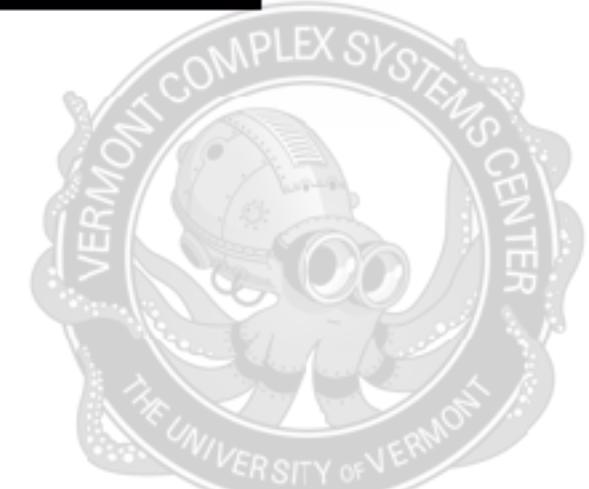
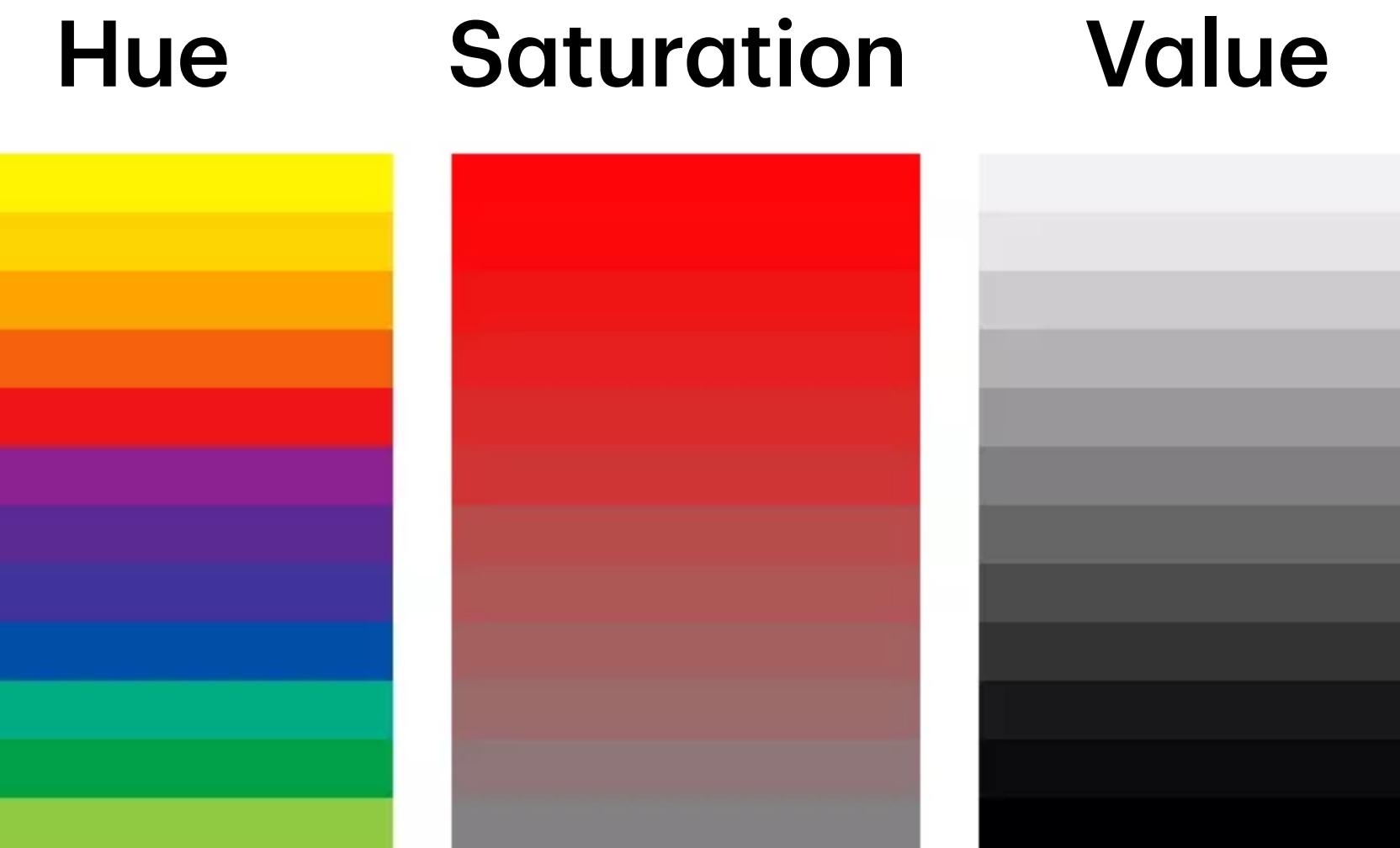
Colors

Modify the marks to encode data, highlight information, and set a tone.

They encode-

- Categorical vs Ordered
- Sequential vs Diverging
- Segments vs Continuous
- Univariate vs Bivariate

Modify colors along



Quiz Time

Identify how color is used to encode the data



Identify how color is used to encode the data

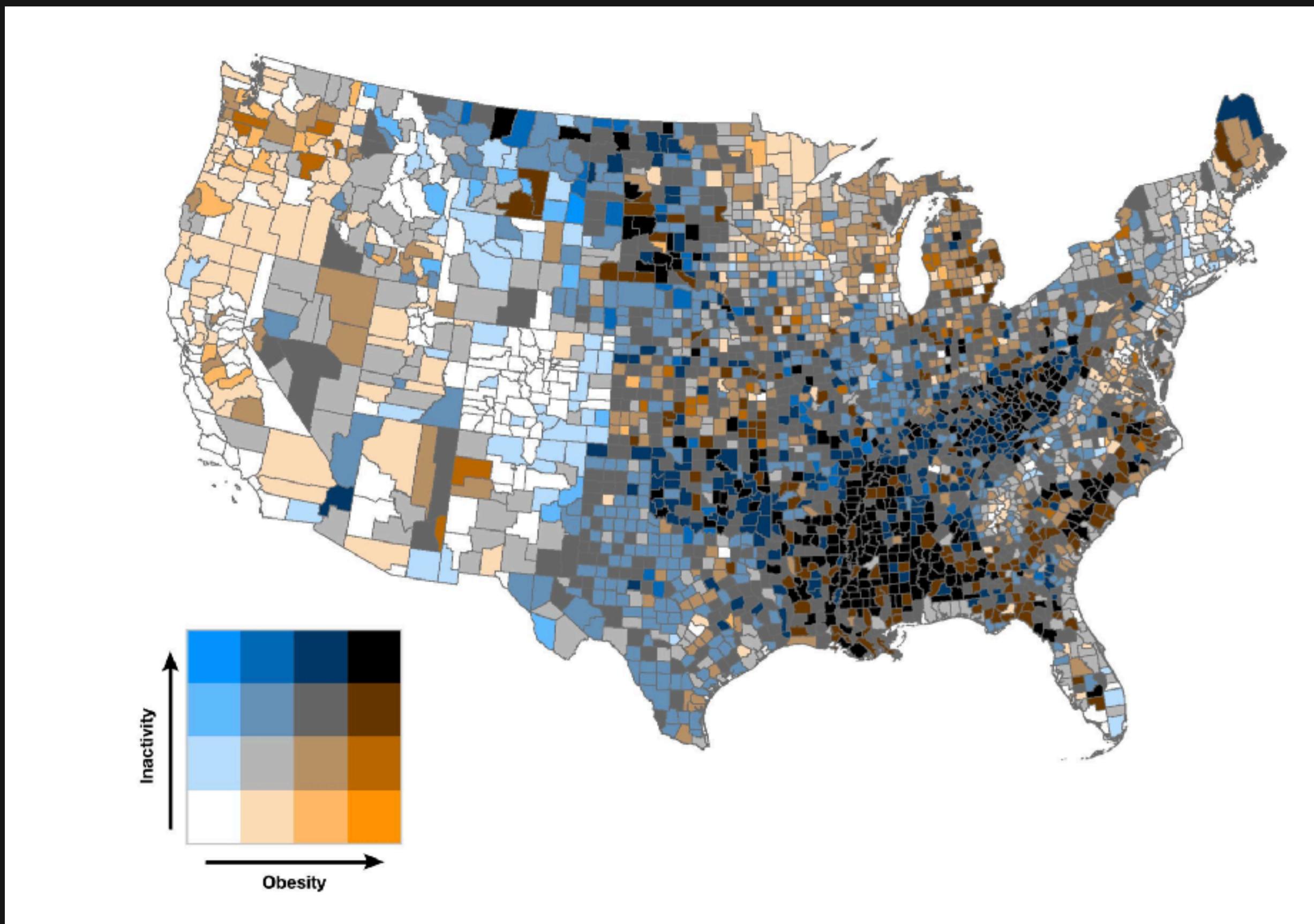
- Categorical vs Ordered
- Sequential vs Diverging
- Segments vs Continuous
- Univariate vs Bivariate



Identify how color is used to encode the data

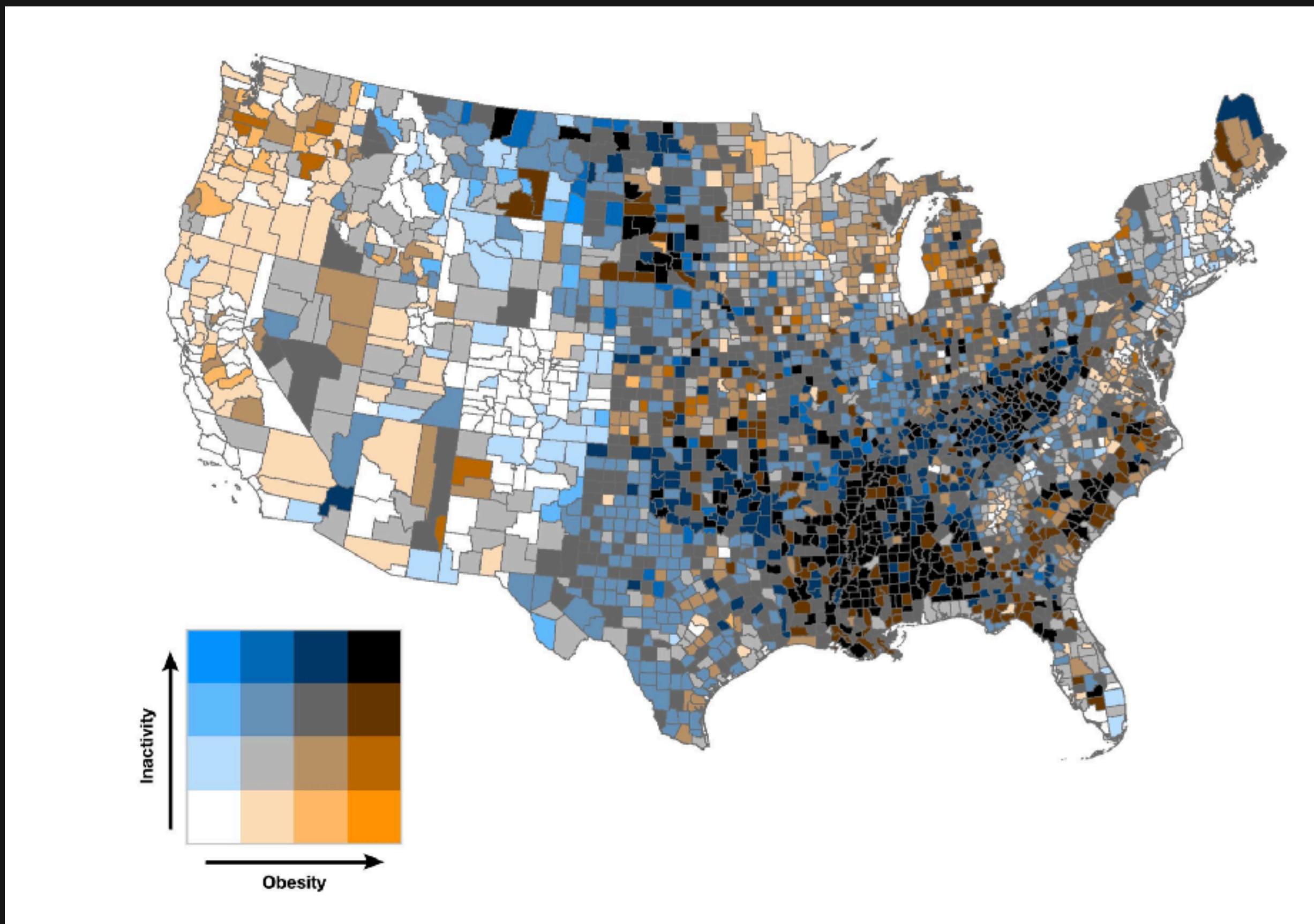
- **Categorical** vs Ordered
- Sequential vs **Diverging**
- **Segments** vs Continuous
- **Univariate** vs Bivariate

Identify how color is used to encode the data



- Categorical vs Ordered
- Sequential vs Diverging
- Segments vs Continuous
- Univariate vs Bivariate

Identify how color is used to encode the data



- Categorical vs **Ordered**
- **Sequential** vs Diverging
- **Segments** vs Continuous
- Univariate vs **Bivariate**

Scaffoldings of a Visualization

Using Colors

Qualitative Color

Used for encoding nominal information

Quantitative Color

Used for encoding magnitude

Okabe Ito



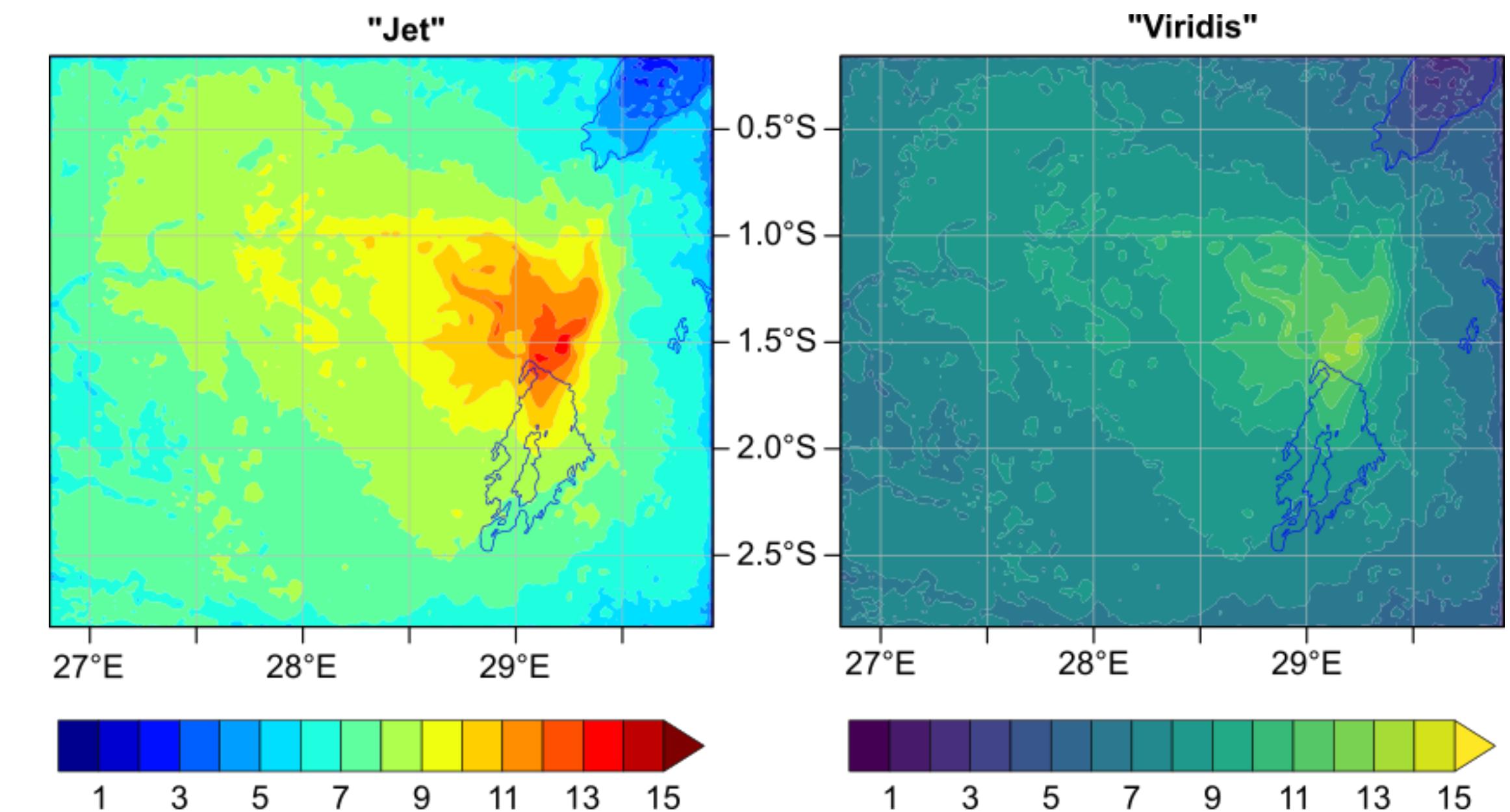
ColorBrewer Dark2



ggplot2 hue

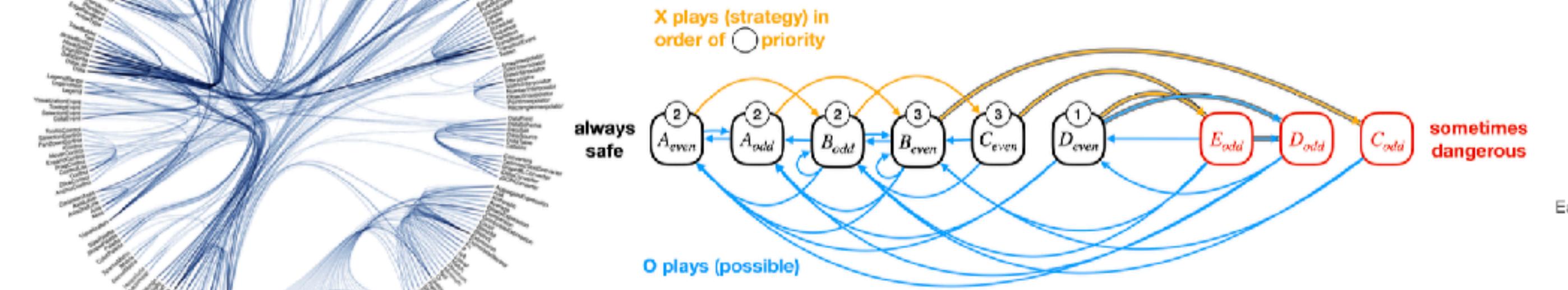
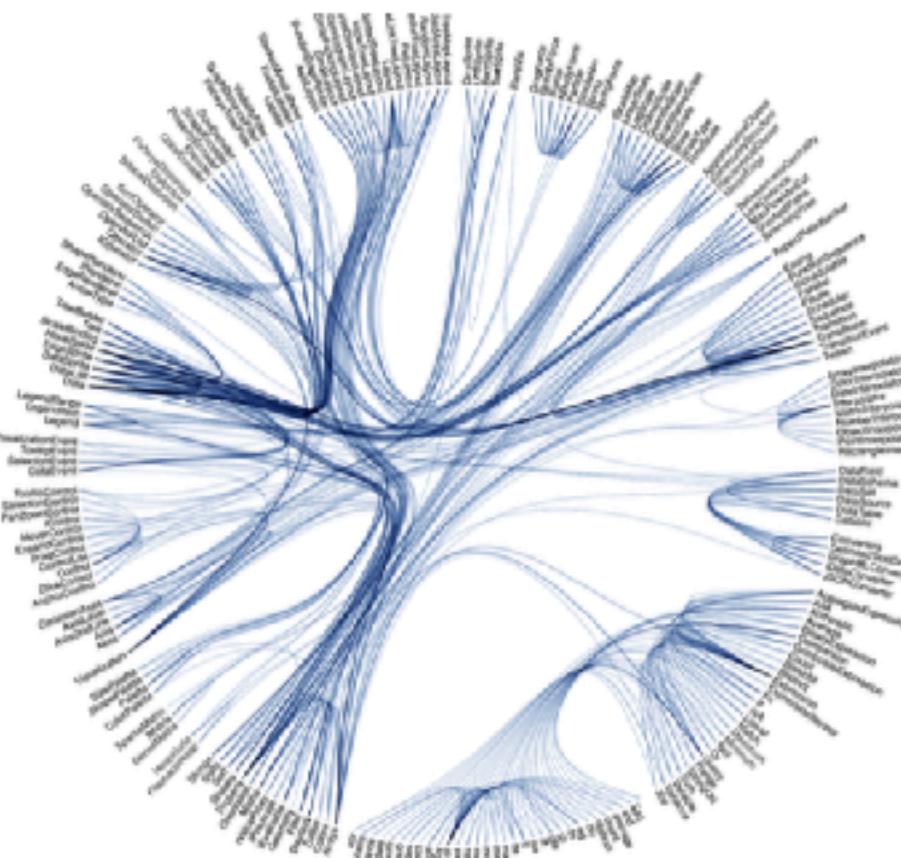
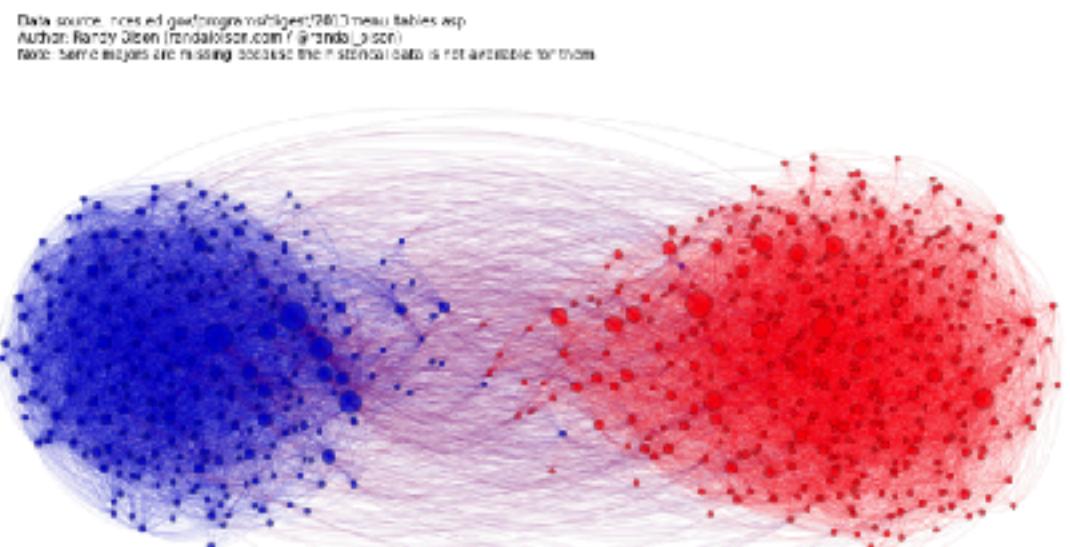
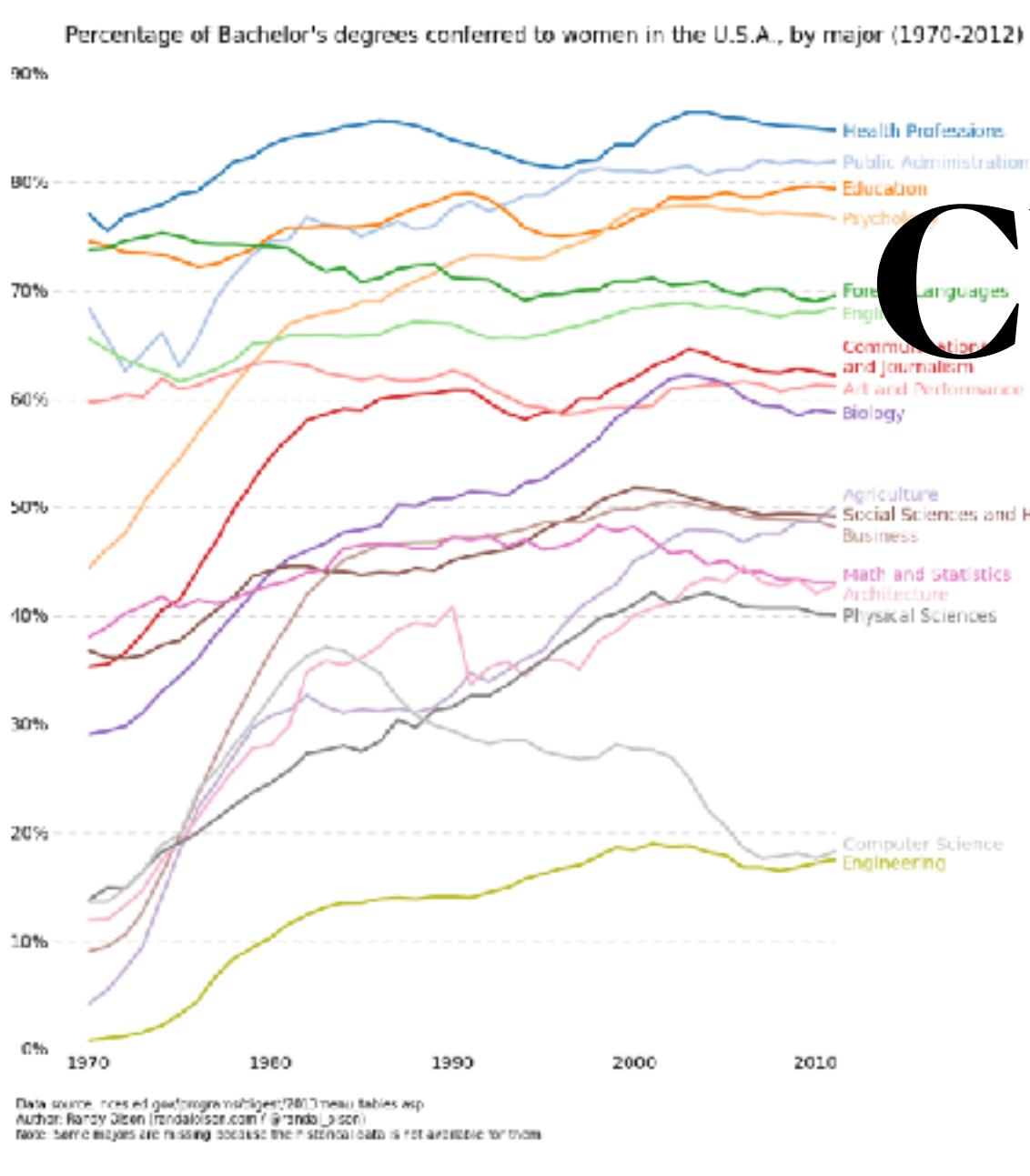


Qualitative Color Map



Quantitative Color Map

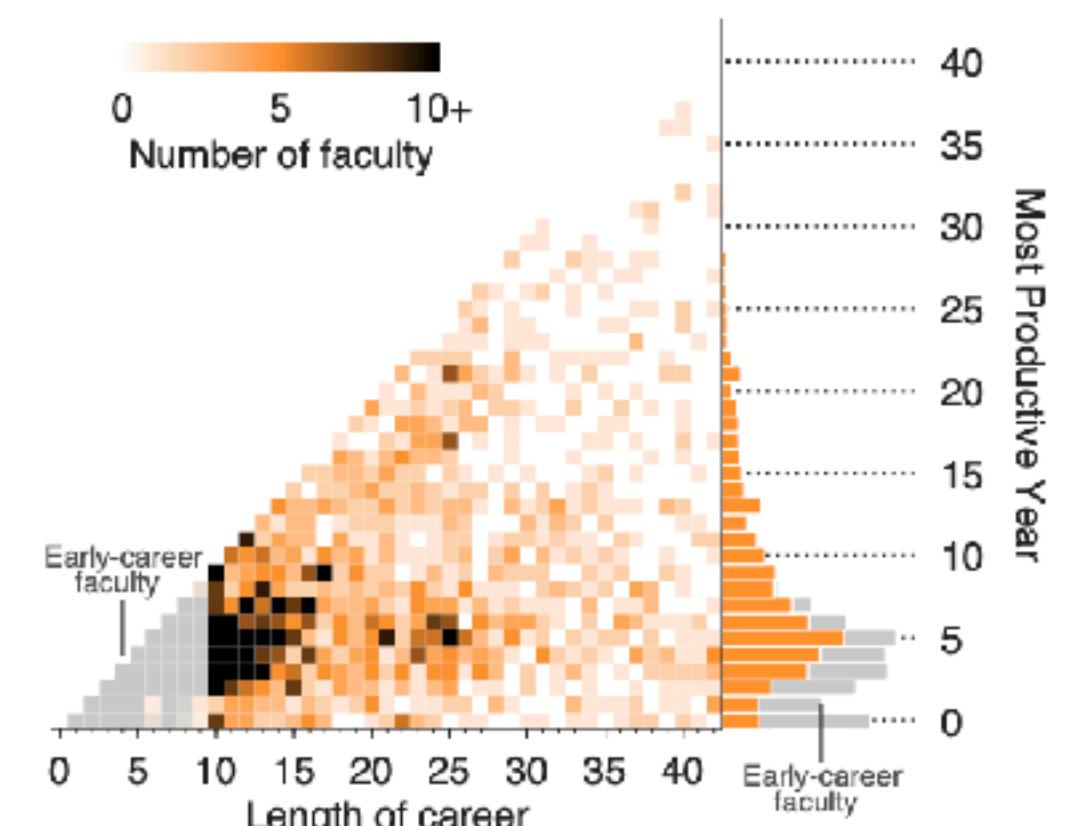
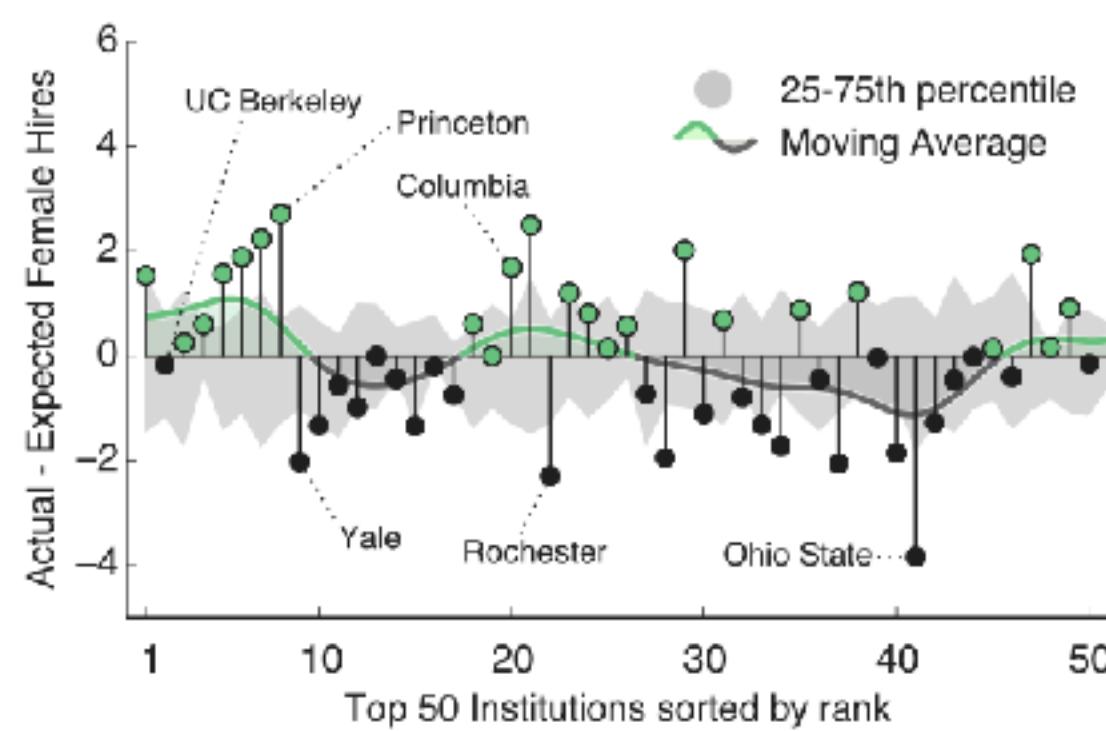
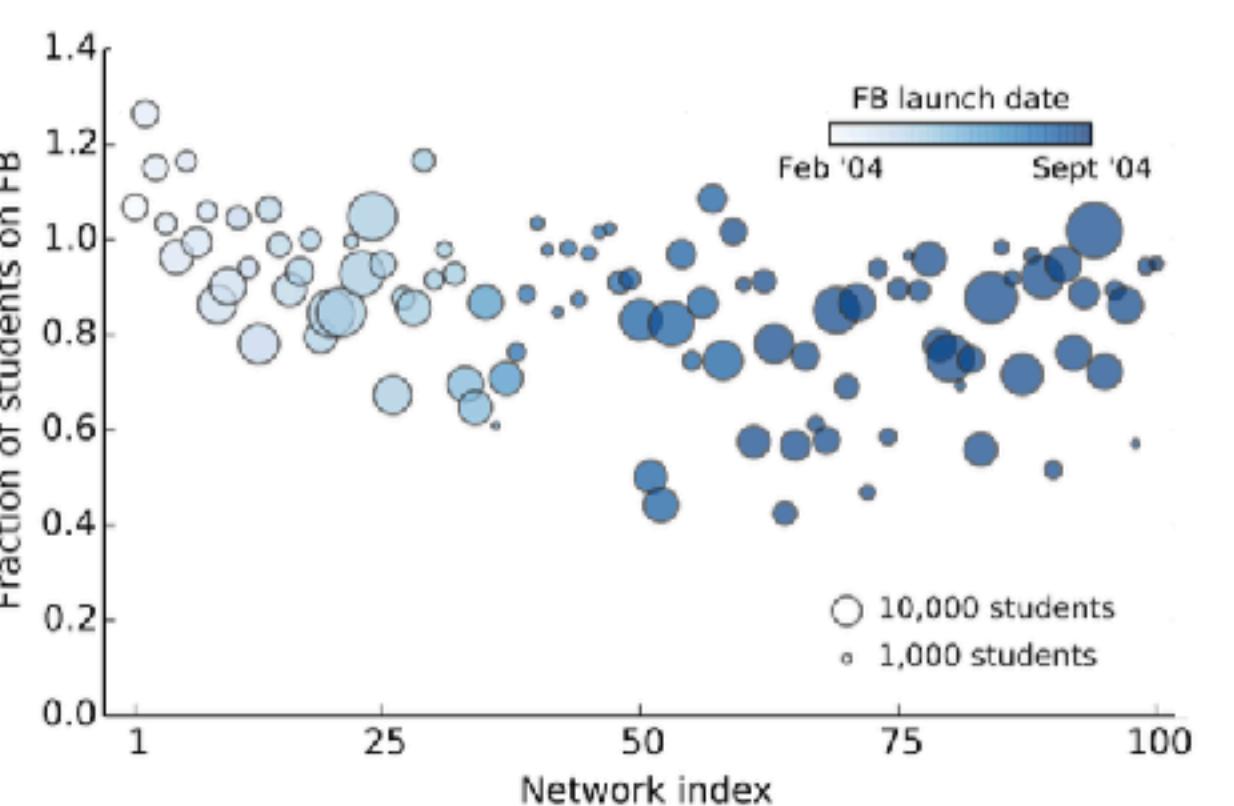
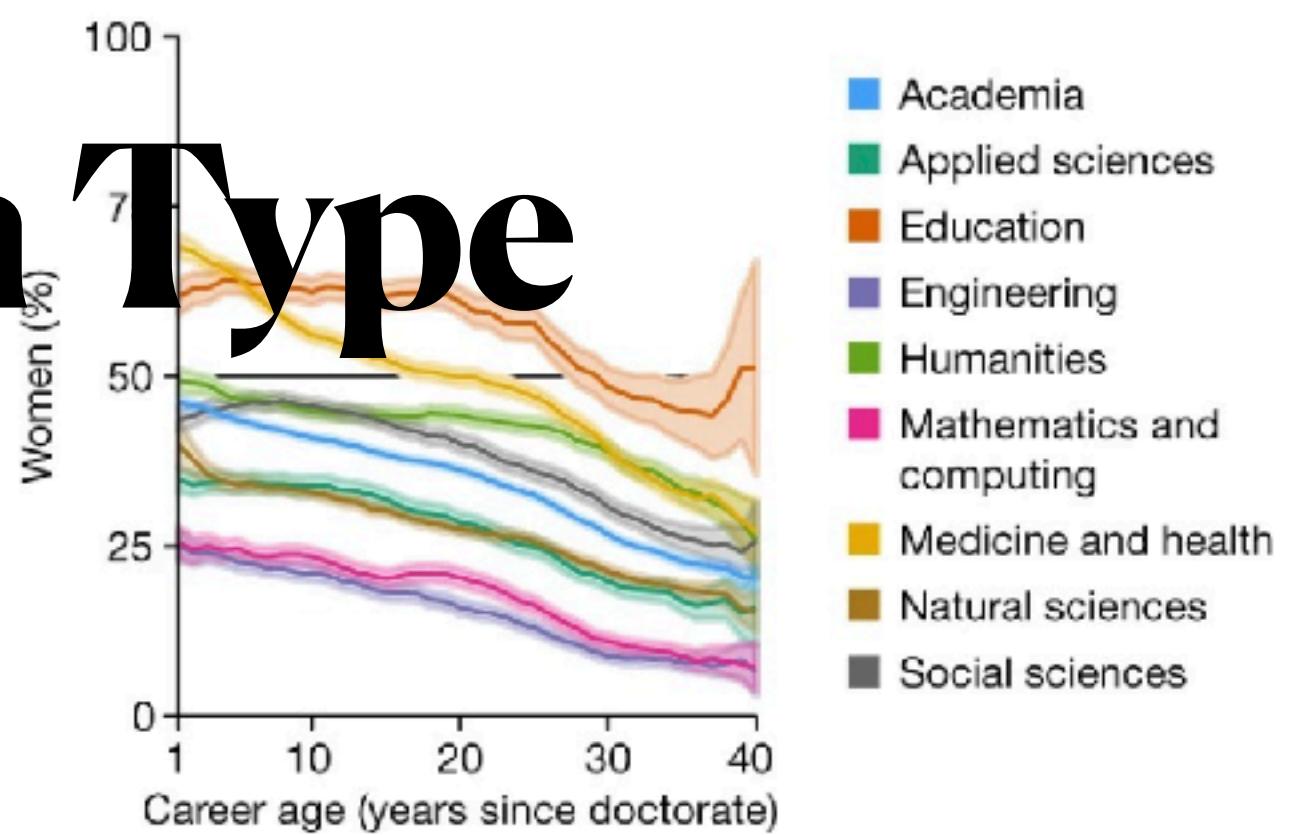
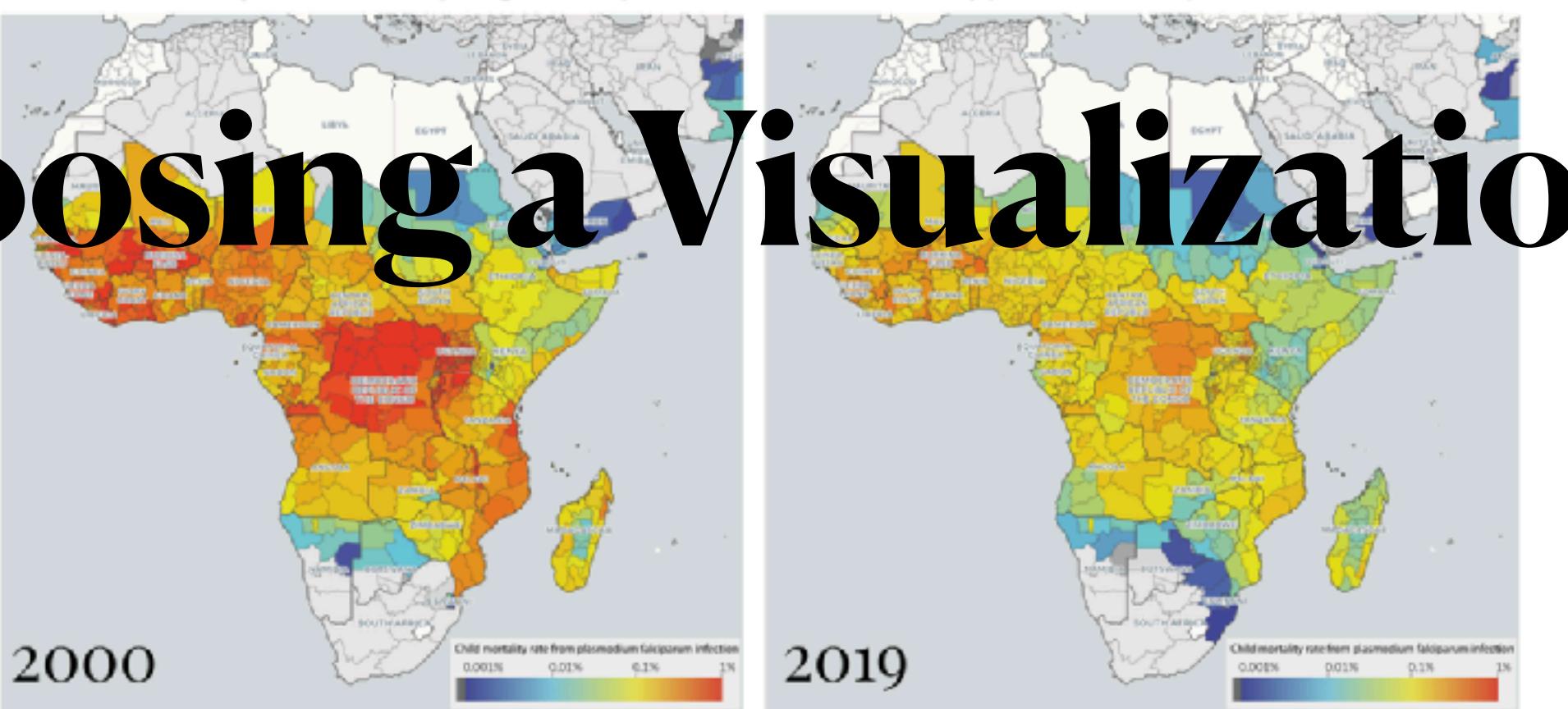




Malaria mortality rate of children in 2000 and 2019

Shown is the mortality rate of children younger than 5 years due to malaria (caused by plasmodium falciparum)

Our World in Data



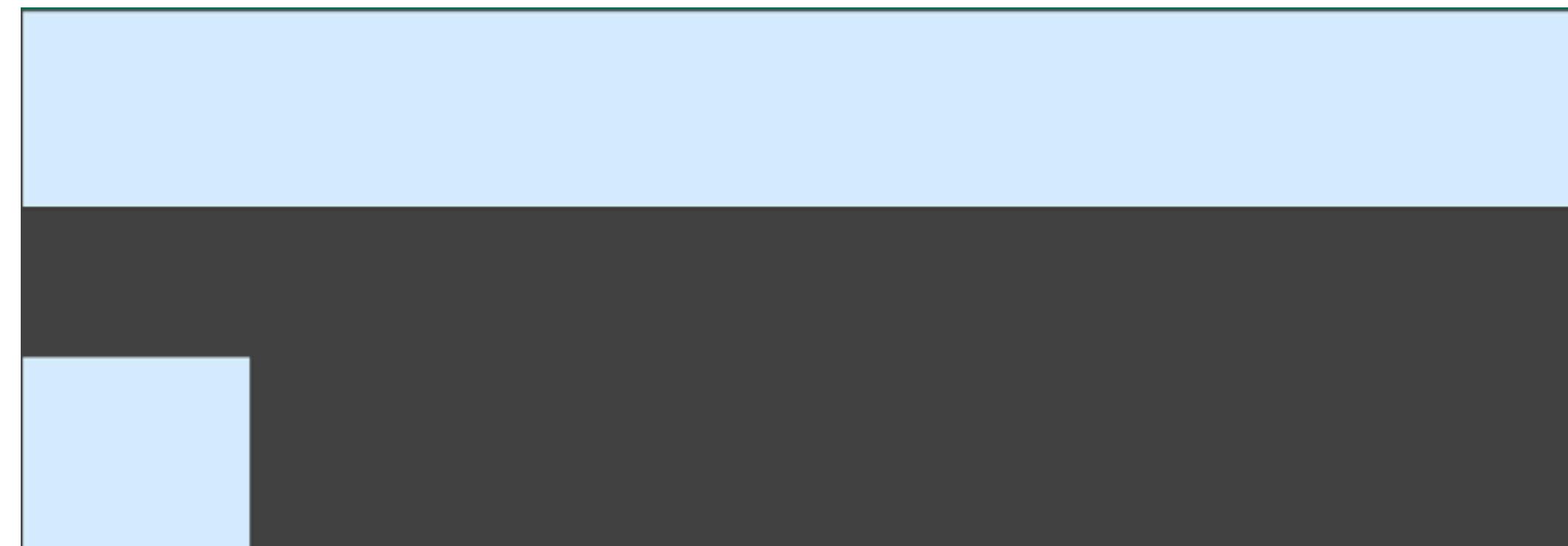
- █ Academia
- █ Applied sciences
- █ Education
- █ Engineering
- █ Humanities
- █ Mathematics and computing
- █ Medicine and health
- █ Natural sciences
- █ Social sciences

Word or phrase

- Always
- Certainly
- Slam dunk
- Almost certainly
- Almost always
- With high probability
- Usually
- Likely
- Frequently
- Probably
- Often
- Serious possibility
- More often than not
- Real possibility
- With moderate probability
- Maybe
- Possibly
- Might happen
- Not often
- Unlikely
- With low probability
- Rarely
- Never

Rules of Thumb

Oughts and Nots

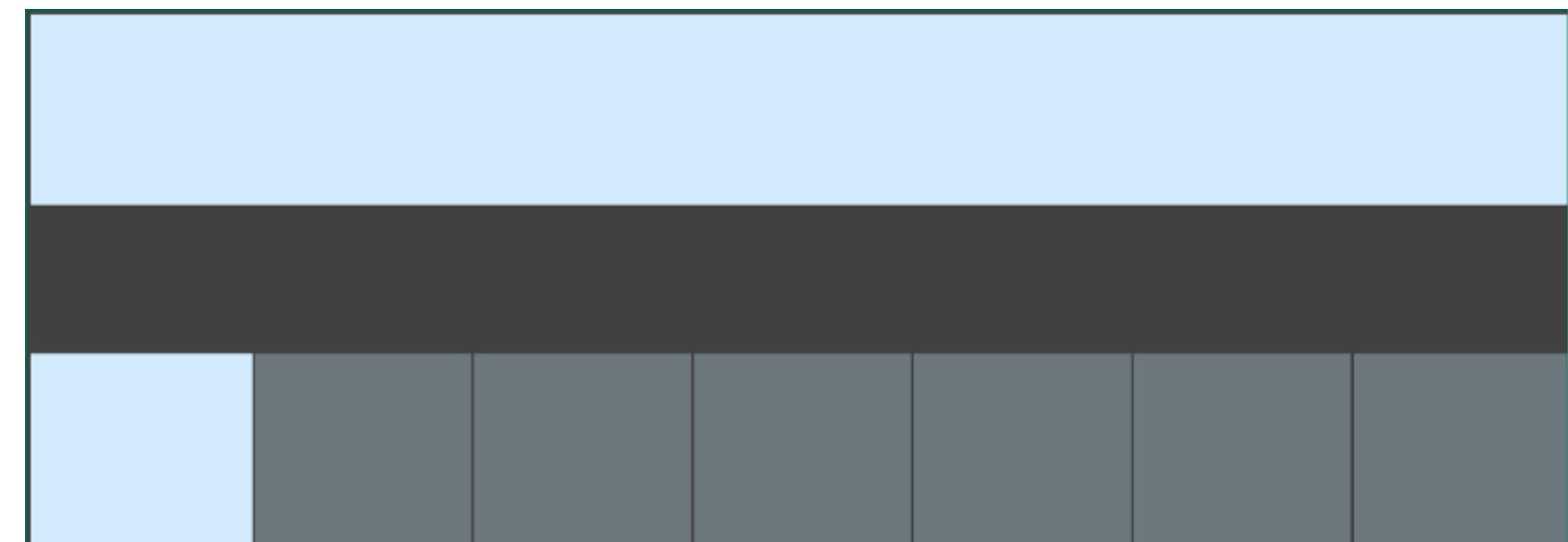


How much longer is the top bar?



Rules of Thumb

Oughts and Nots

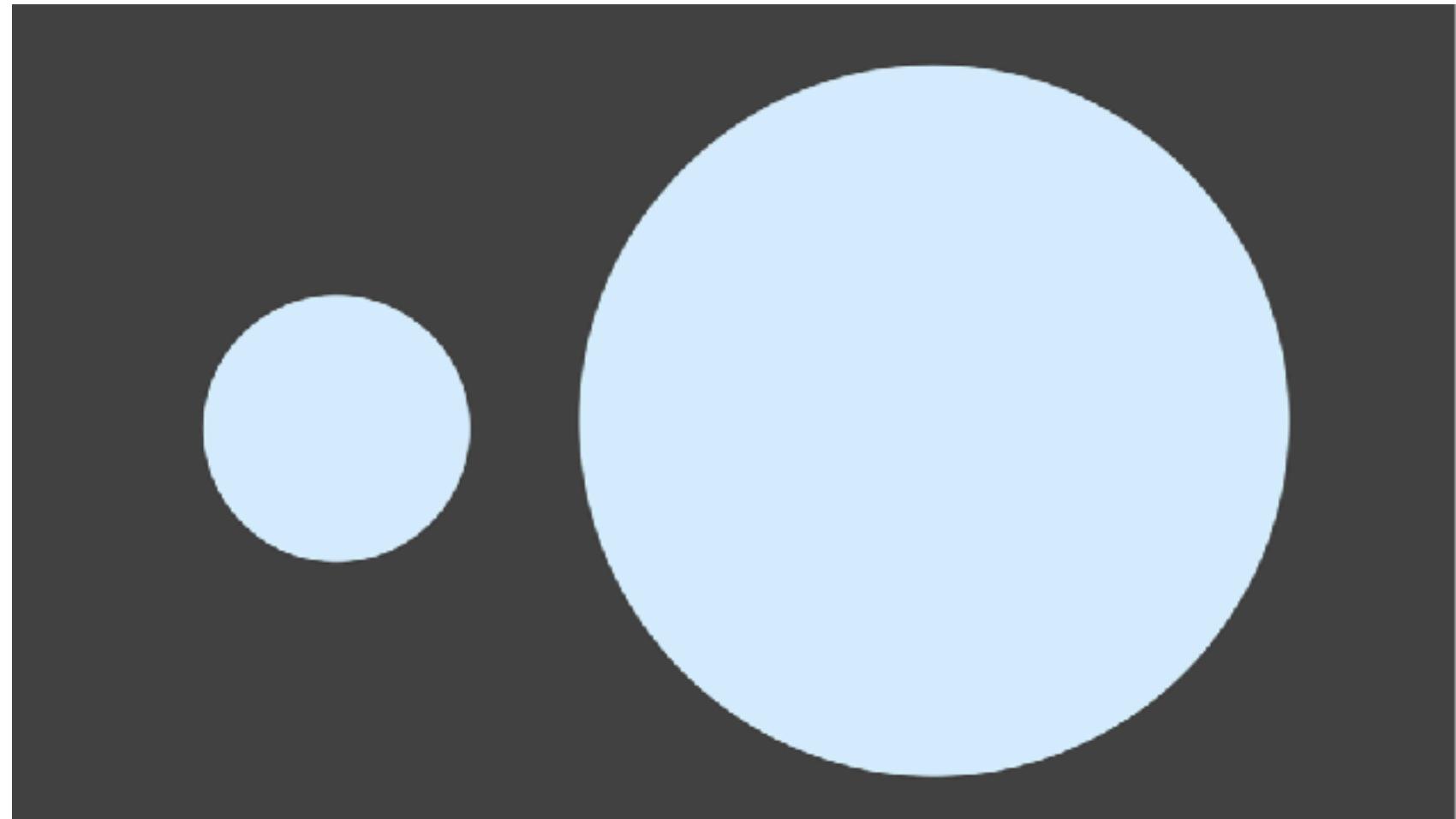


7 times!



Rules of Thumb

Oughts and Nots

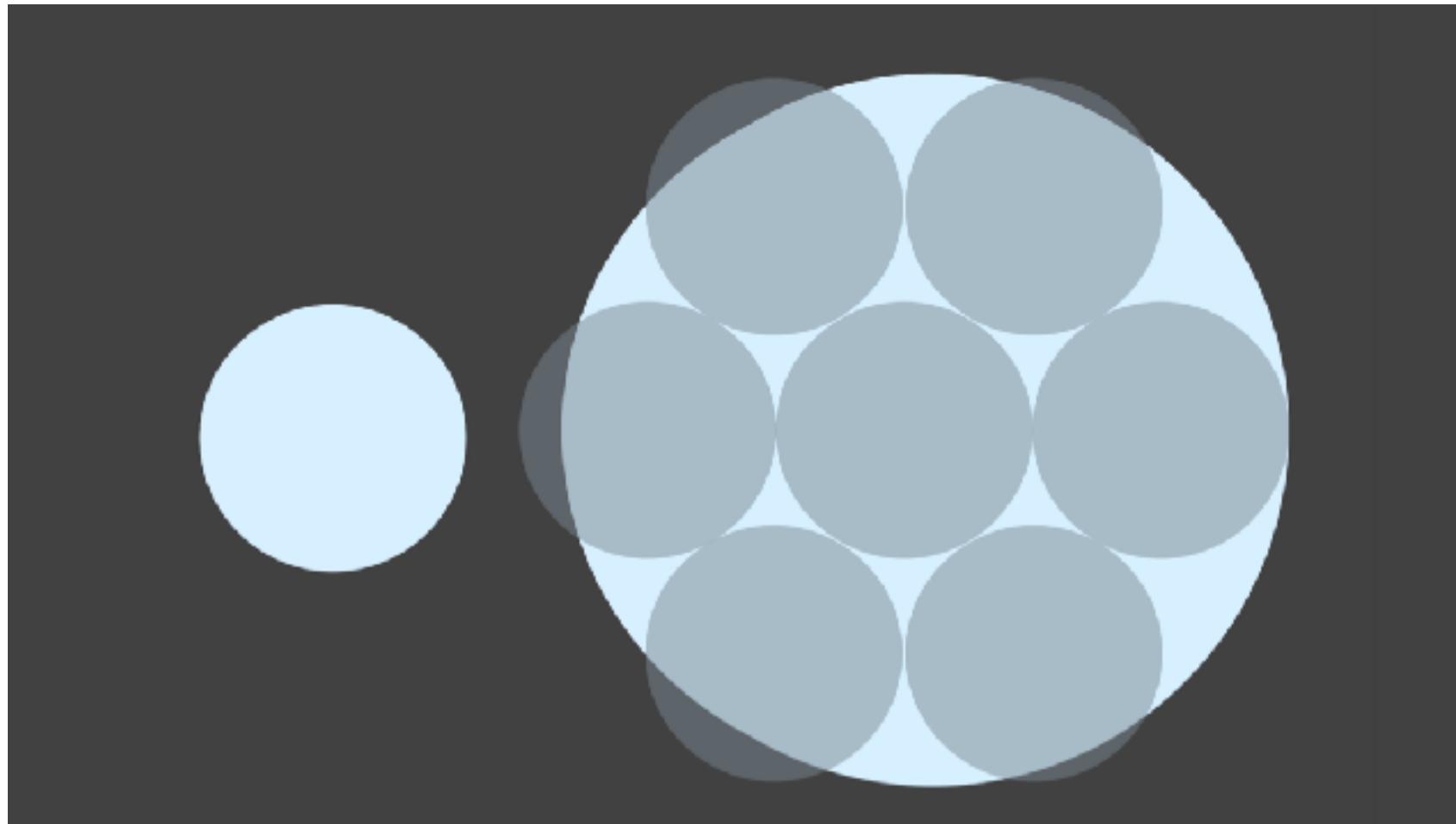


How much bigger is the circle on the right?



Rules of Thumb

Oughts and Nots



Also, 7 times



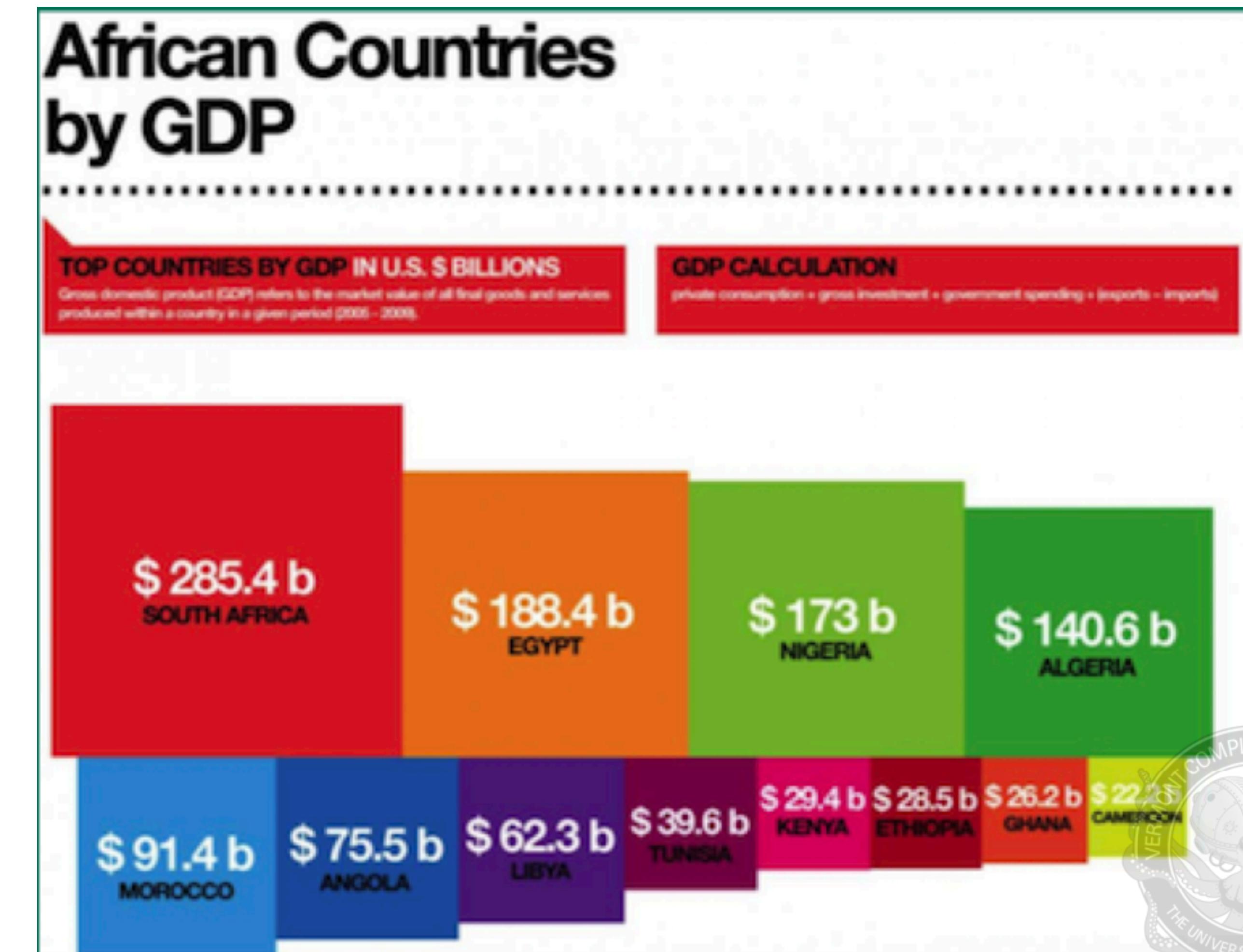
Rules of Thumb Oughts and Nots

Understanding Markers

Lengths are easy to compare

Angles are hard (Don't use pie charts!)

Areas are also hard



Rules of Thumb

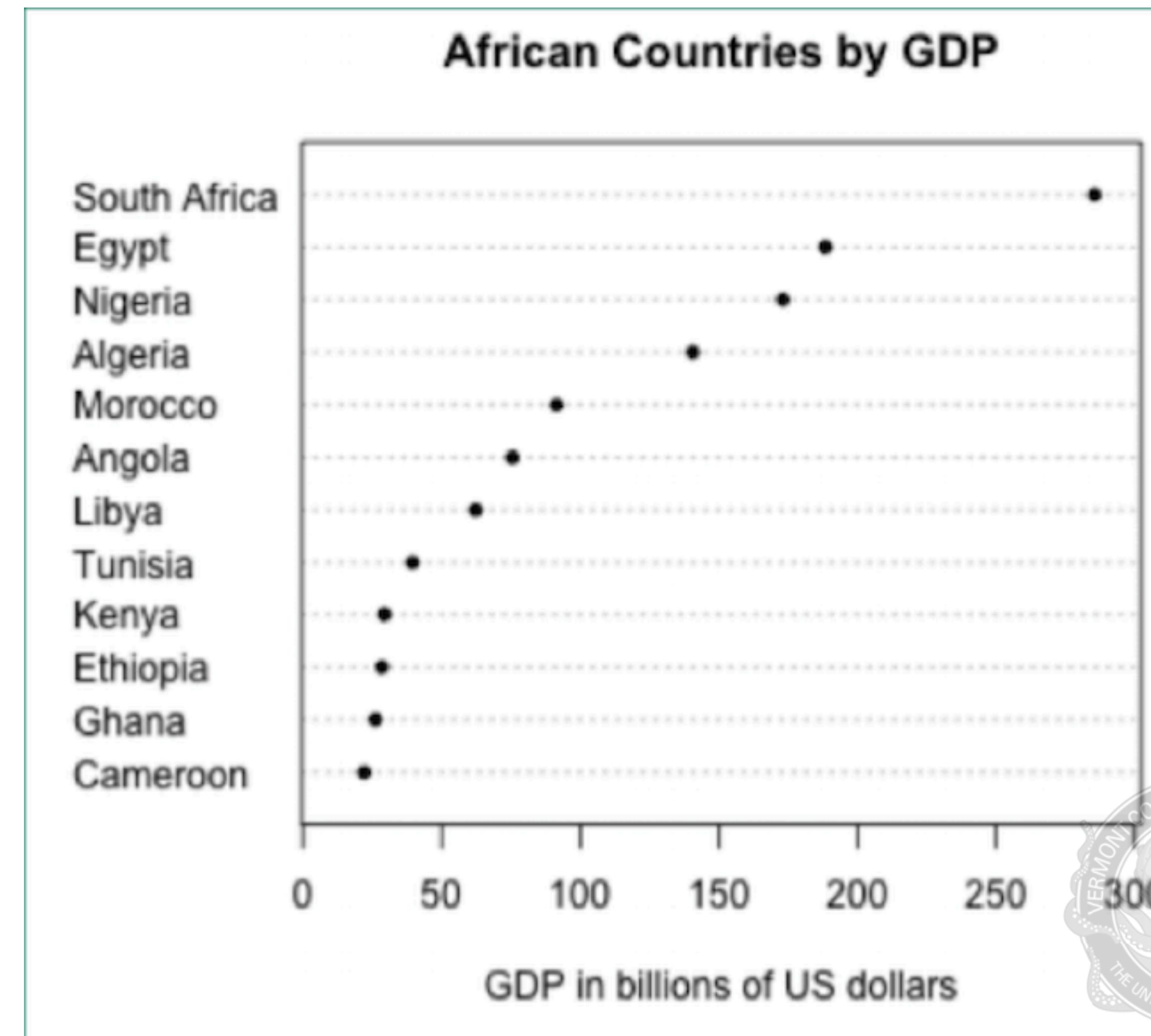
Oughts and Nots

Understanding Markers

Lengths are easy to compare

Angles are hard (Don't use pie charts!)

Areas are also hard



Rules of Thumb

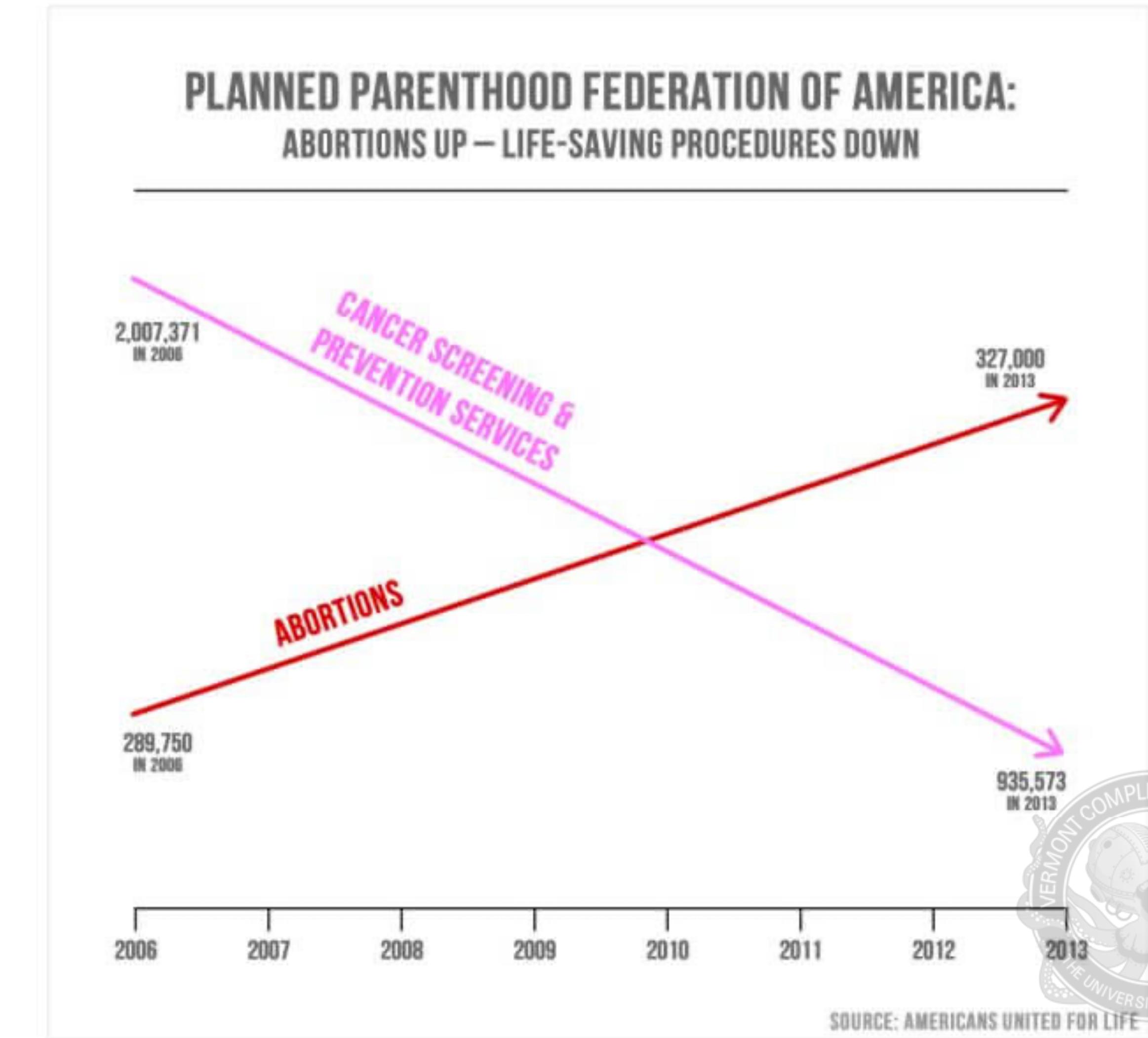
Oughts and Nots

Respecting the Scales

Keep your double scales consistent

Don't truncate your y-axis to mislead

Reveal the data



Rules of Thumb

Oughts and Nots

Respecting the Scales

Keep your double scales consistent

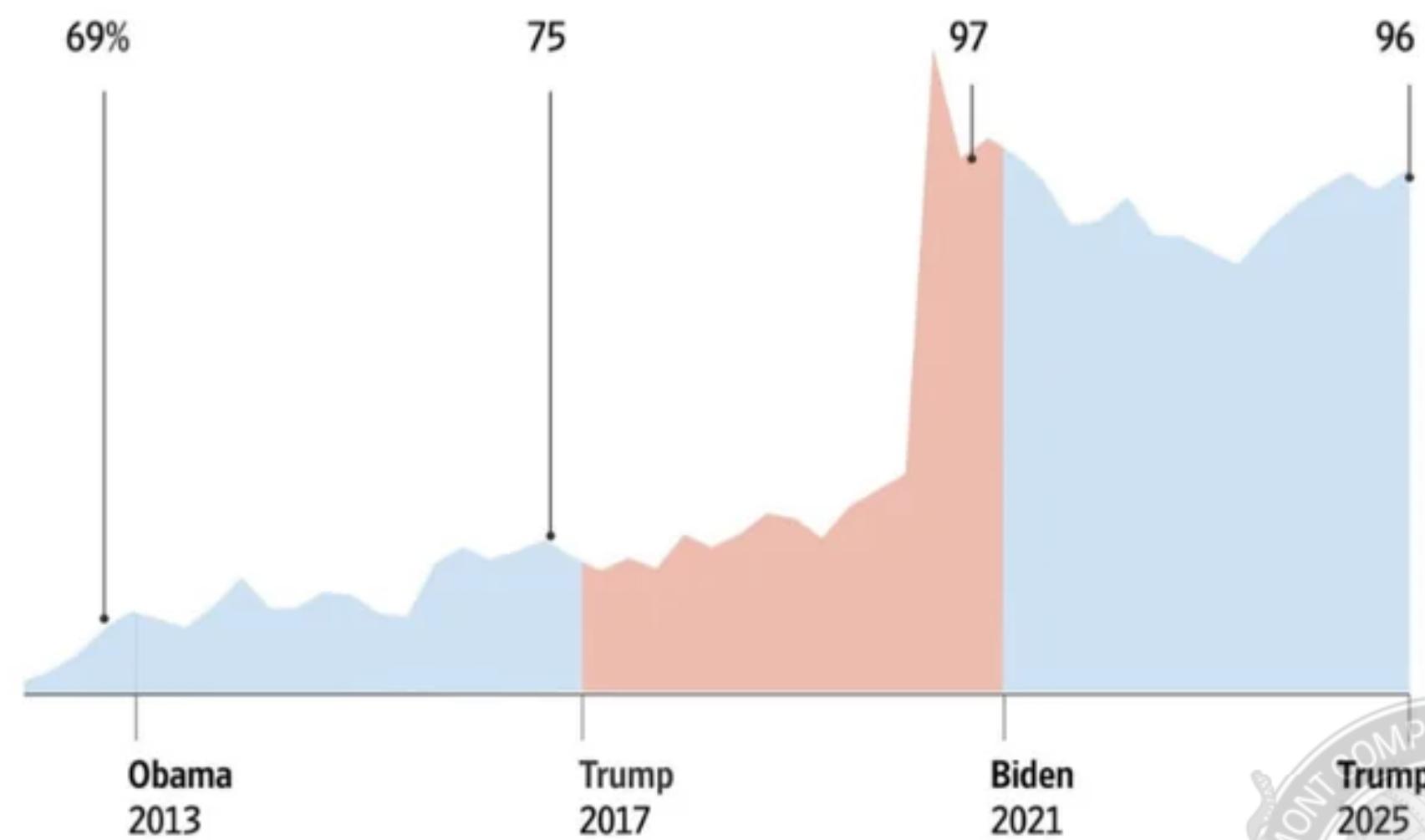
Don't truncate your y-axis to mislead

Reveal the data

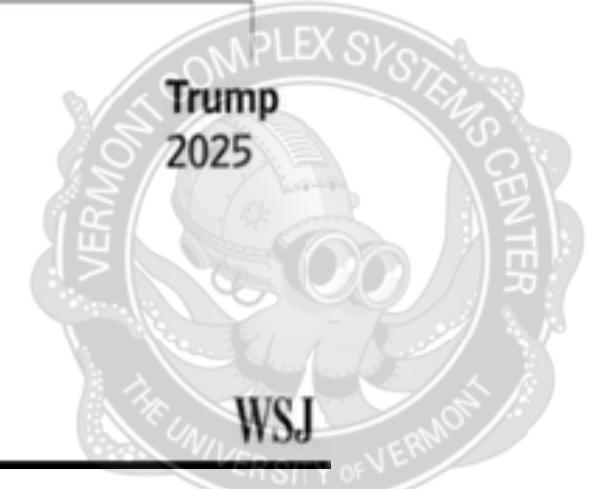
Trump inherited a federal debt of about \$36.2 trillion on Inauguration Day—more than \$16 trillion higher than when he last entered the White House.

Federal debt

Debt held by the public as a percentage of gross domestic product, Q3 before inauguration



Sources: Treasury Department via St. Louis Fed



WSJ

Rules of Thumb

Oughts and Nots

Art of Color

Do not use more than 6-10 colors at a time

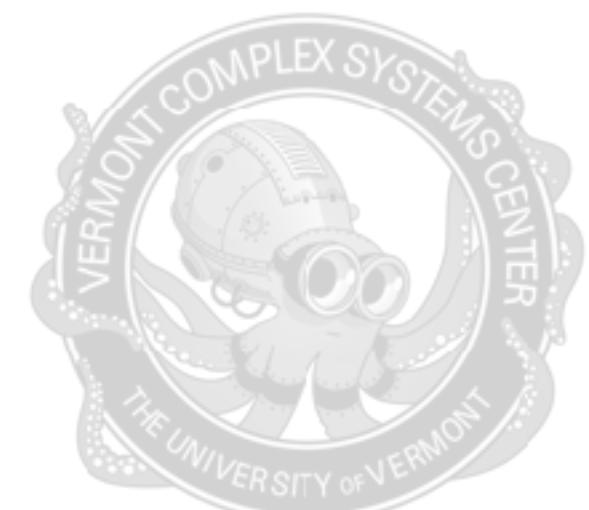
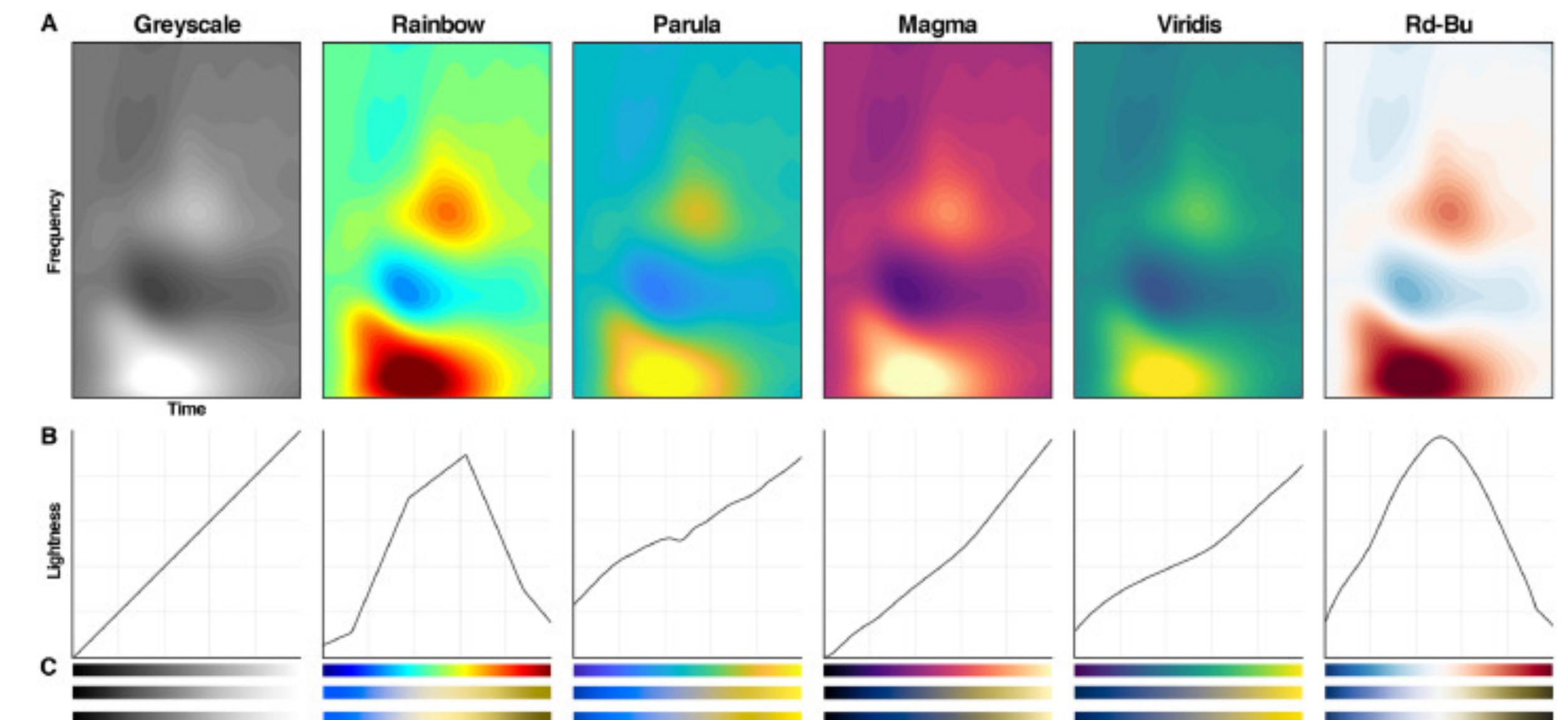
Use Value/Brightness

Saturation works, but not as well

Do not use hue

Do not use rainbow color map

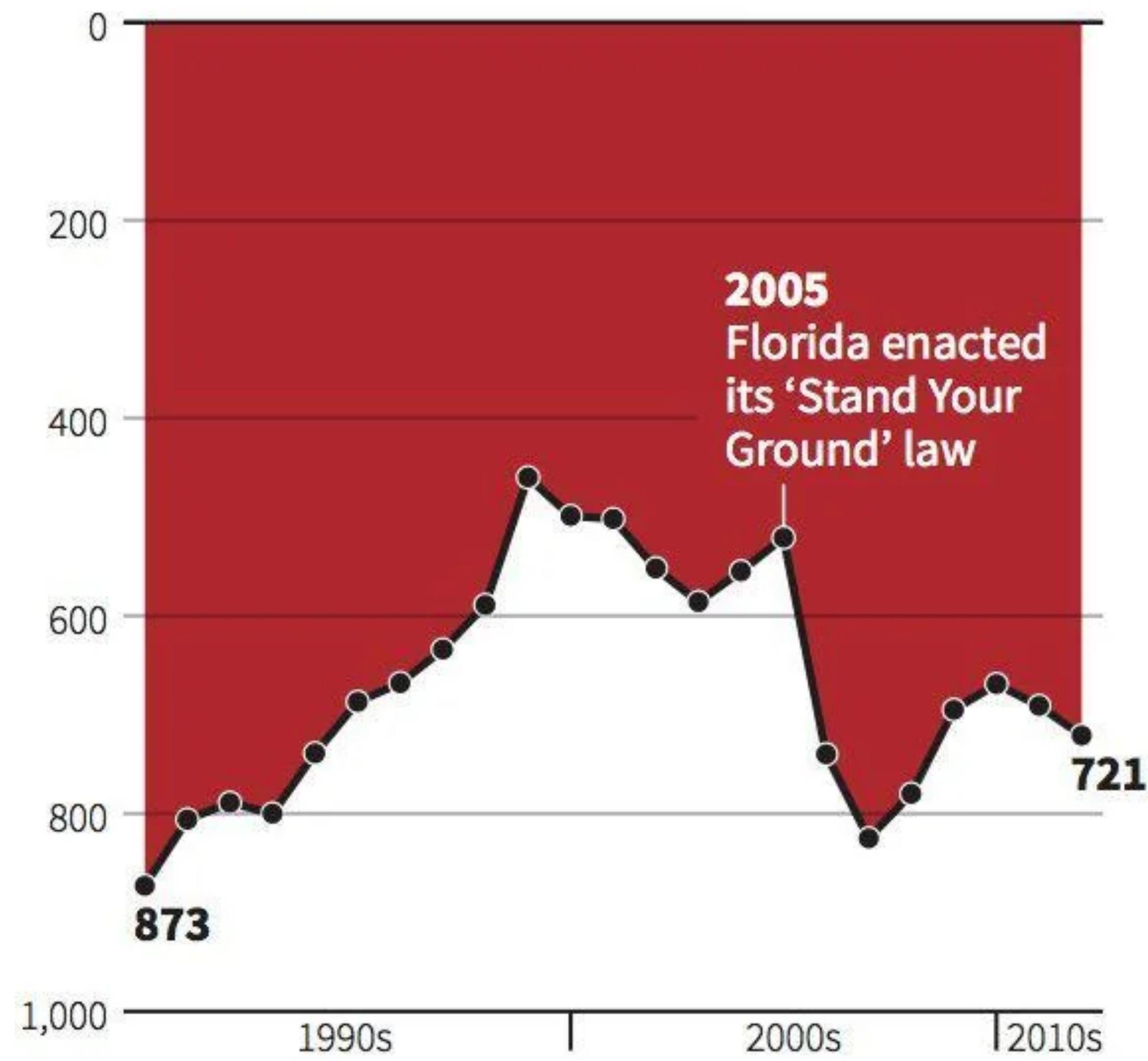
Be color blind friendly



Critiquing Visualizations

Gun deaths in Florida

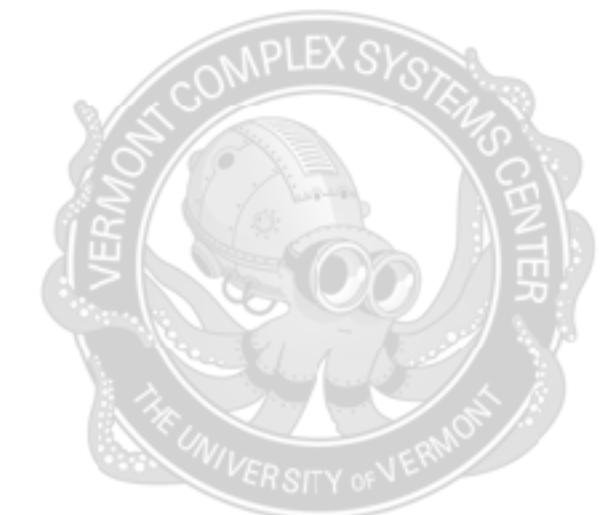
Number of murders committed using firearms

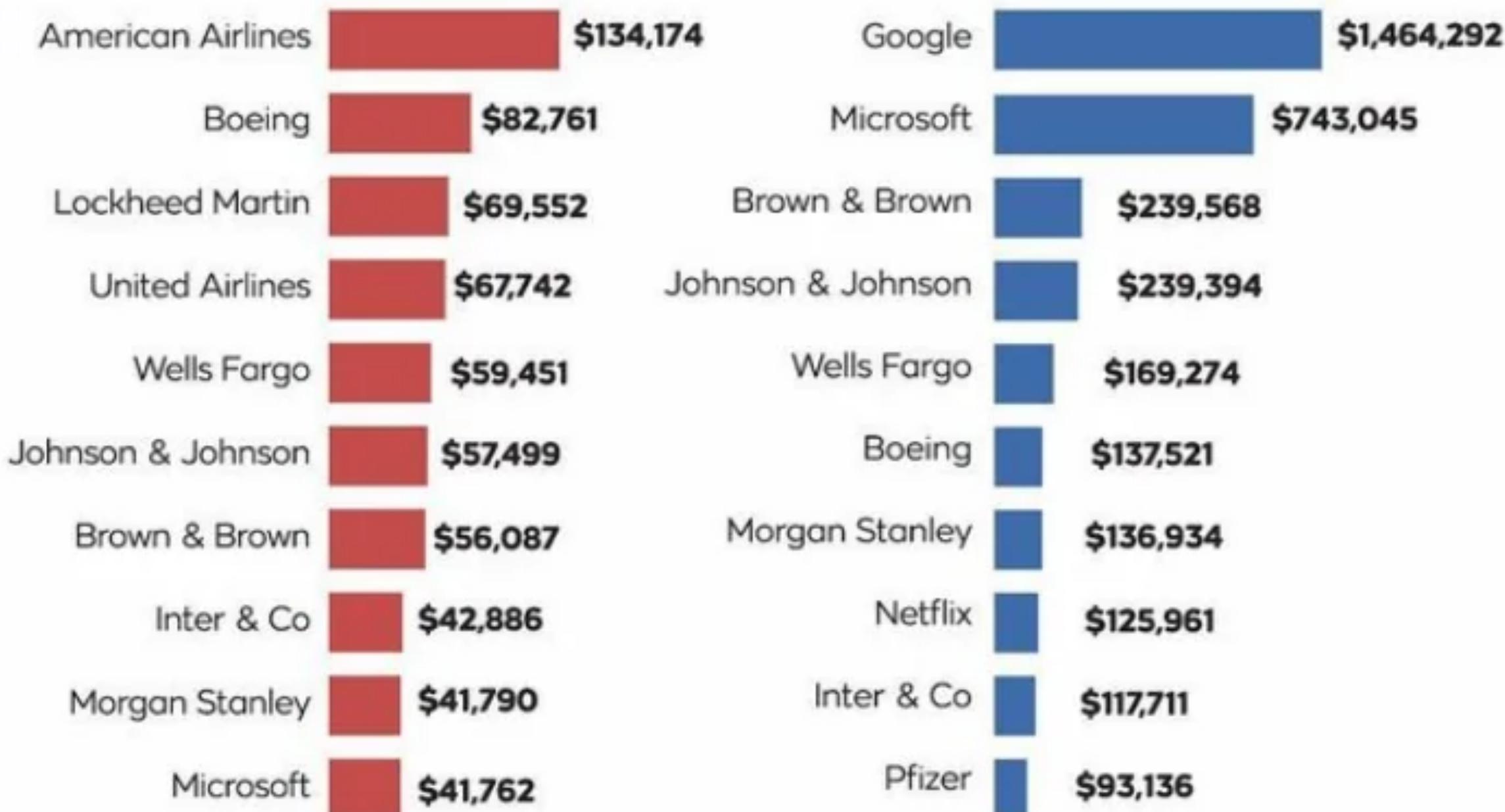


Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

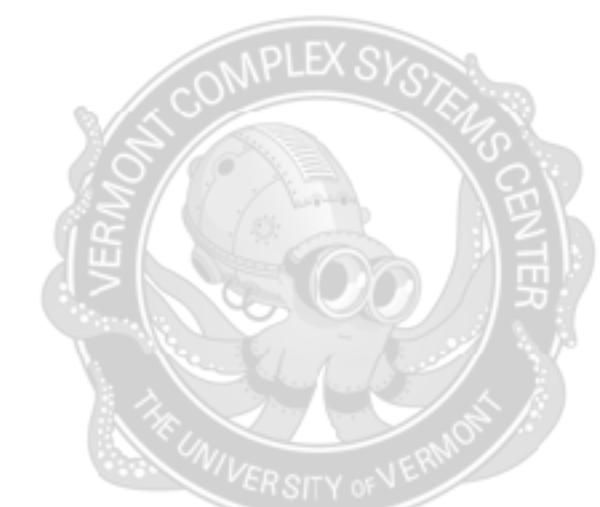
REUTERS





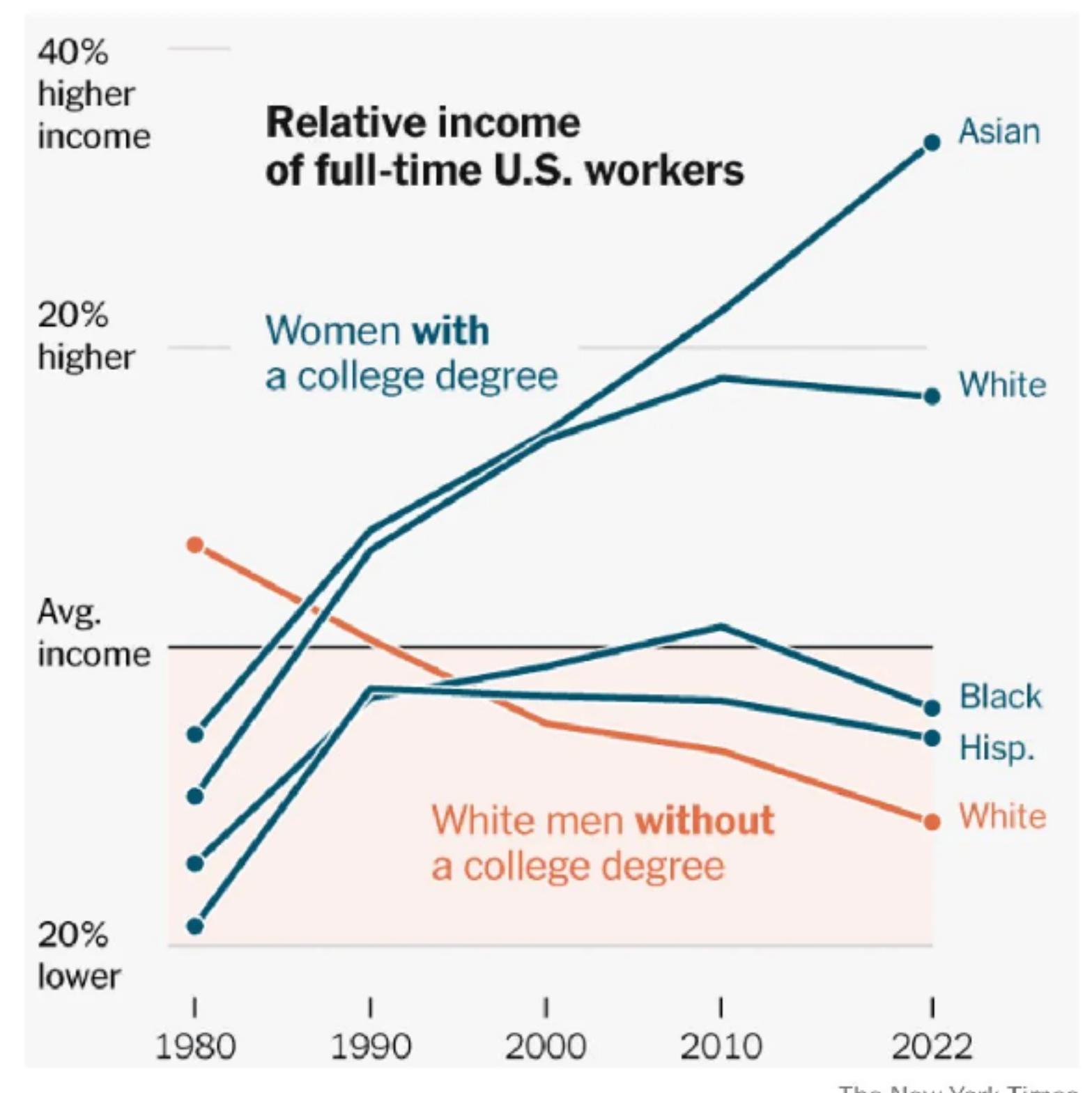
Note: Data excludes donations to affiliated PACs, leaving out many large contributions.

Source: Quiver Quantitative



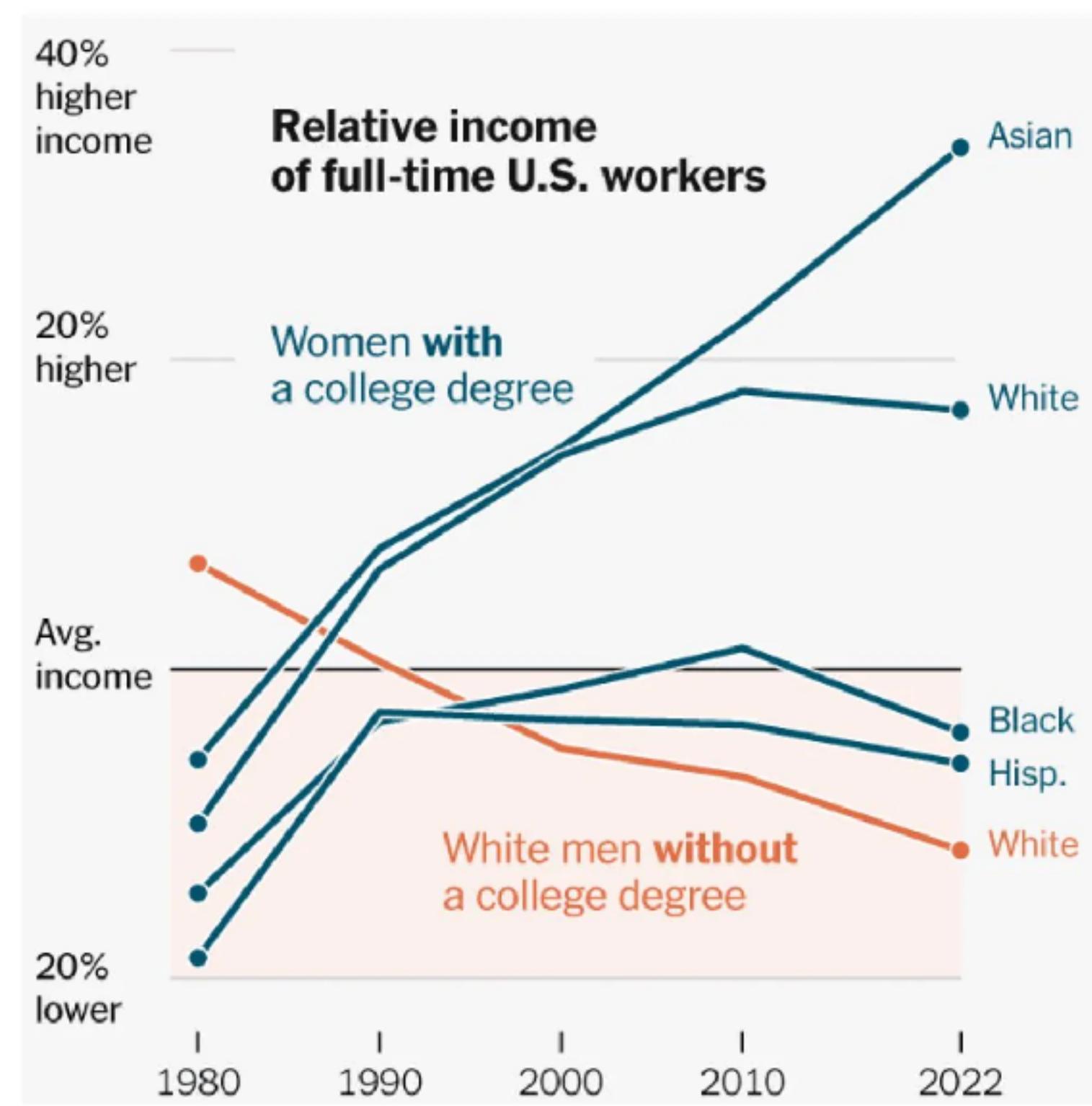
They Used to Be Ahead in the American Economy. Now They've Fallen Behind.

One economic question that really matters in an election year: Are others doing better than you?



They Used to Be Ahead in the American Economy. Now They've Fallen Behind.

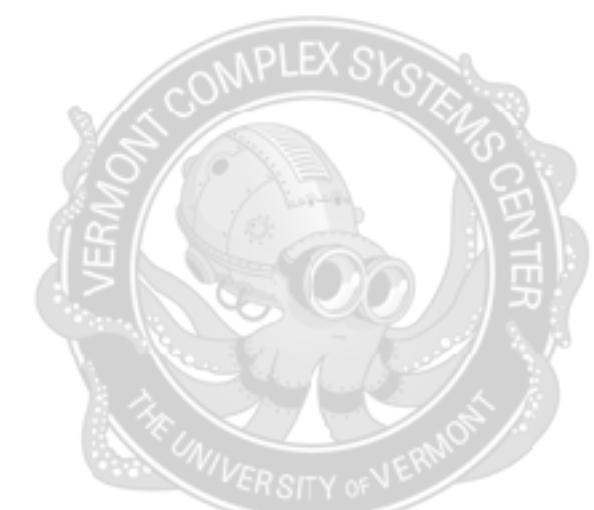
One economic question that really matters in an election year: Are others doing better than you?



"The zero axis is average income instead of median, so more than anything the chart is distorted by increasing income disparity before it even begins to dissect information, so even for those who can figure out what the vertical axis is trying to show the result will be badly distorted by such confounding factors."

“‘Men are falling behind’ is a wild way to summarize ‘the gender wage gap is no longer so huge that even women getting a college education does not make up for it.’”

“Including the question ‘Are others doing better than you?’ REALLY throws off the chart. It (incorrectly) looks like they asked a bunch of white guys ‘Are Asians doing better than you? Are Black people doing better than you? How about women with degrees?’”



Inspired By

- Ethan Kerzner (Staff Engineer, Relevant)
- Alice Patania (Asst. Prof., UVM)
- Nick Cheney (Assoc. Prof., UVM)
- Jonathan St-Onge (Research Scientist, UVM)
- Dan Larremore (Assoc. Prof., CU Boulder)



Visualization Pipeline

1. Determine goals
 2. Choose visualization type
 3. Build a prototype
 4. Feedback & Iterate
- # Coding Time!