Visualization

CS109A Introduction to Data Science

Pavlos Protopapas, Natesh Pillai

As the matplotlib thickens …

ANNOUNCEMENTS

- Homework 4 will be released Oct 20 and is due Oct 27 (Wed) @ 11:59pm

Learning Objectives

- Understand why visualization/plotting is important
- Learn aspects that tend to make visualizes effective and ineffective
- Feel comfortable designing plots that best convey your message
- Gain experience in producing plots with Python

Extra Goal

be more cognizant of broader design choices
(e.g., typography, spacing, colors)

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design

- Communication
  - Motivation
  - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Assume you know a given dataset is credible, complete with the info you want, and has no missing values.

Why do further EDA?

Purposes of EDA:

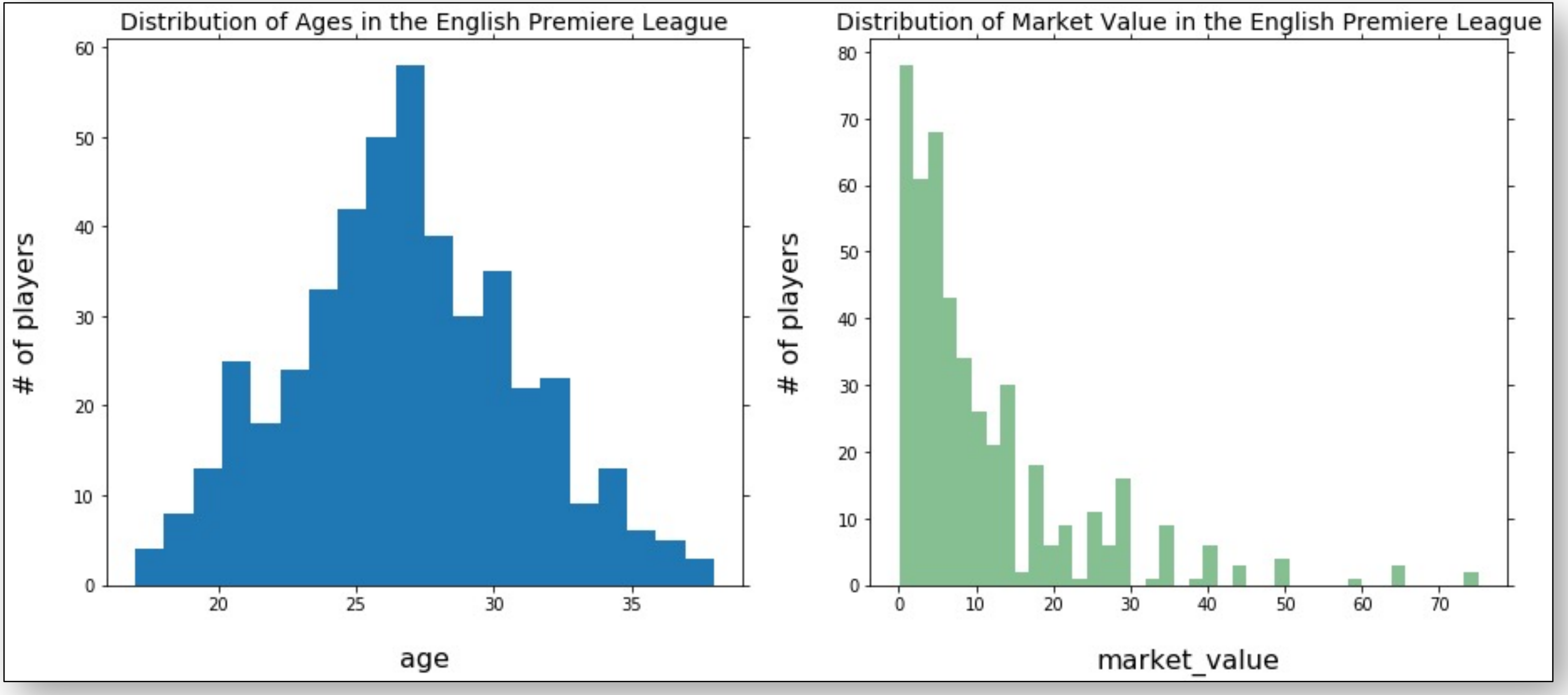
- Maximize insight into a dataset
- Uncover underlying structure
- Detect outliers
- Test underlying assumptions
- Develop parsimonious models

EDA Refresher: English Premier League

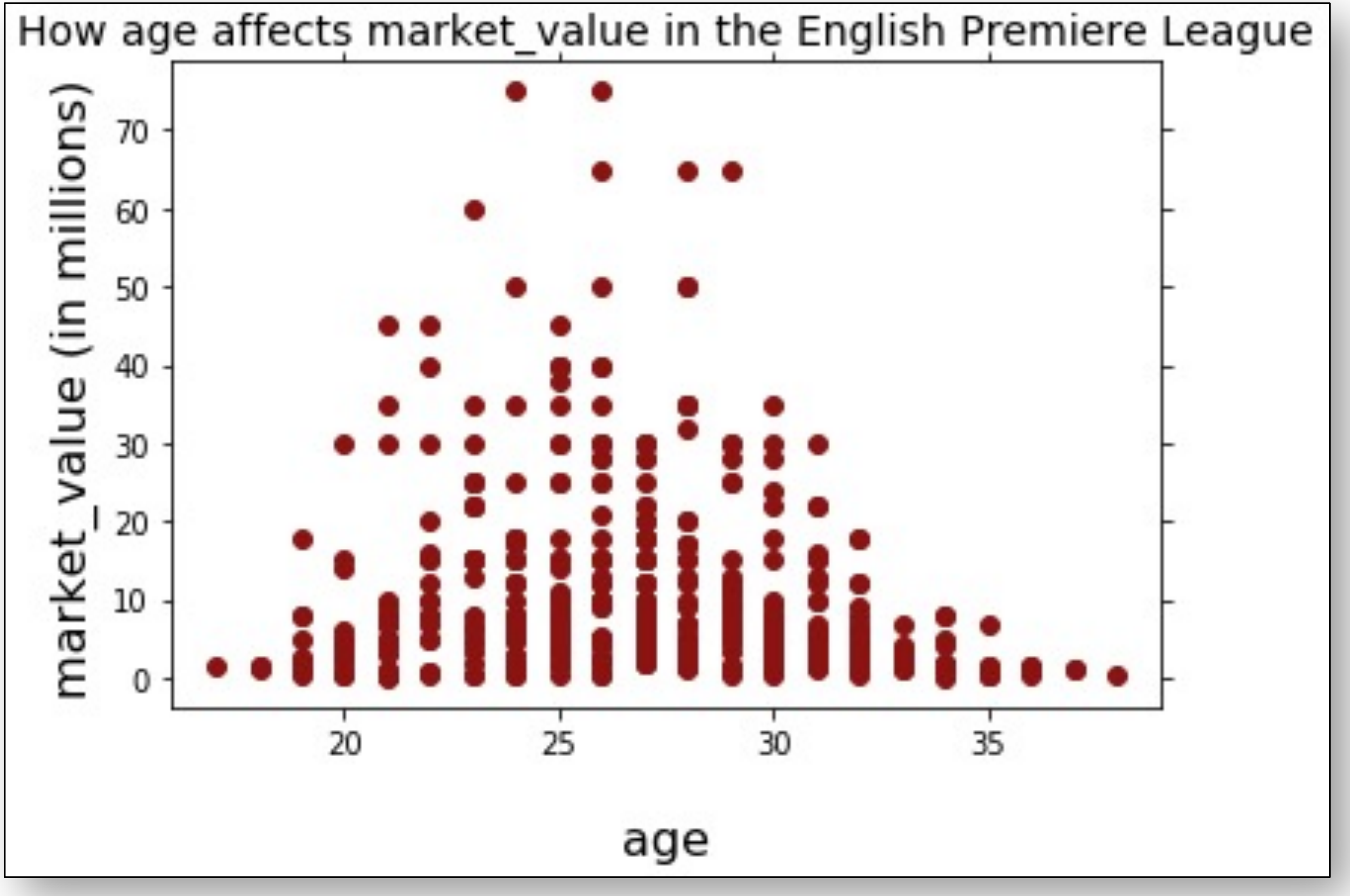
name	club	age	position	market value
Alexis Sanchez	Arsenal	28	LW	65
Mesut Ozil	Arsenal	28	AM	50
Petr Cech	Arsenal	35	GK	7
Theo Walcott	Arsenal	28	RW	20
Laurent Koscielny	Arsenal	31	CB	22

from www.transfermarkt.us

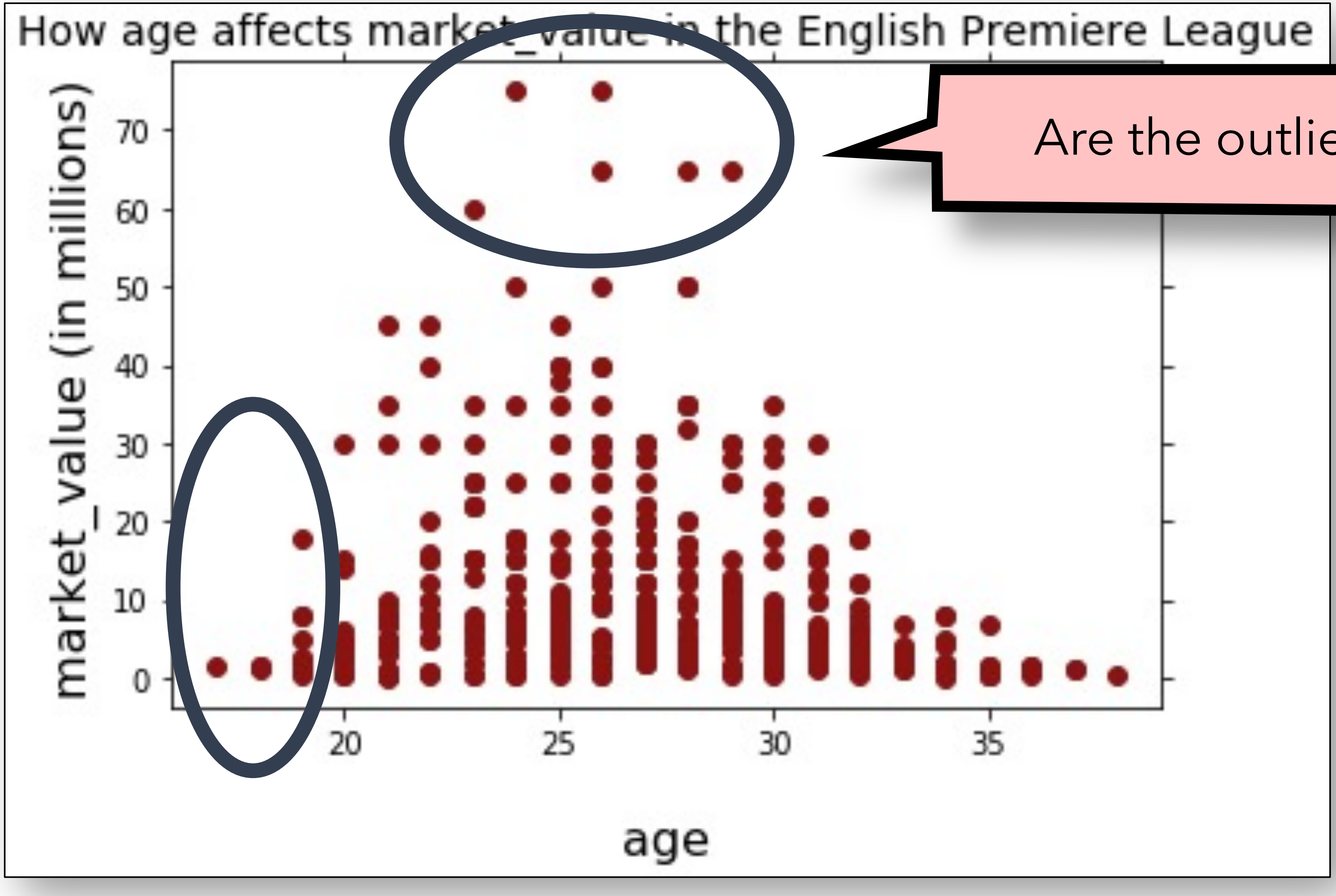
EDA Refresher: English Premier League



EDA Refresher: English Premier League

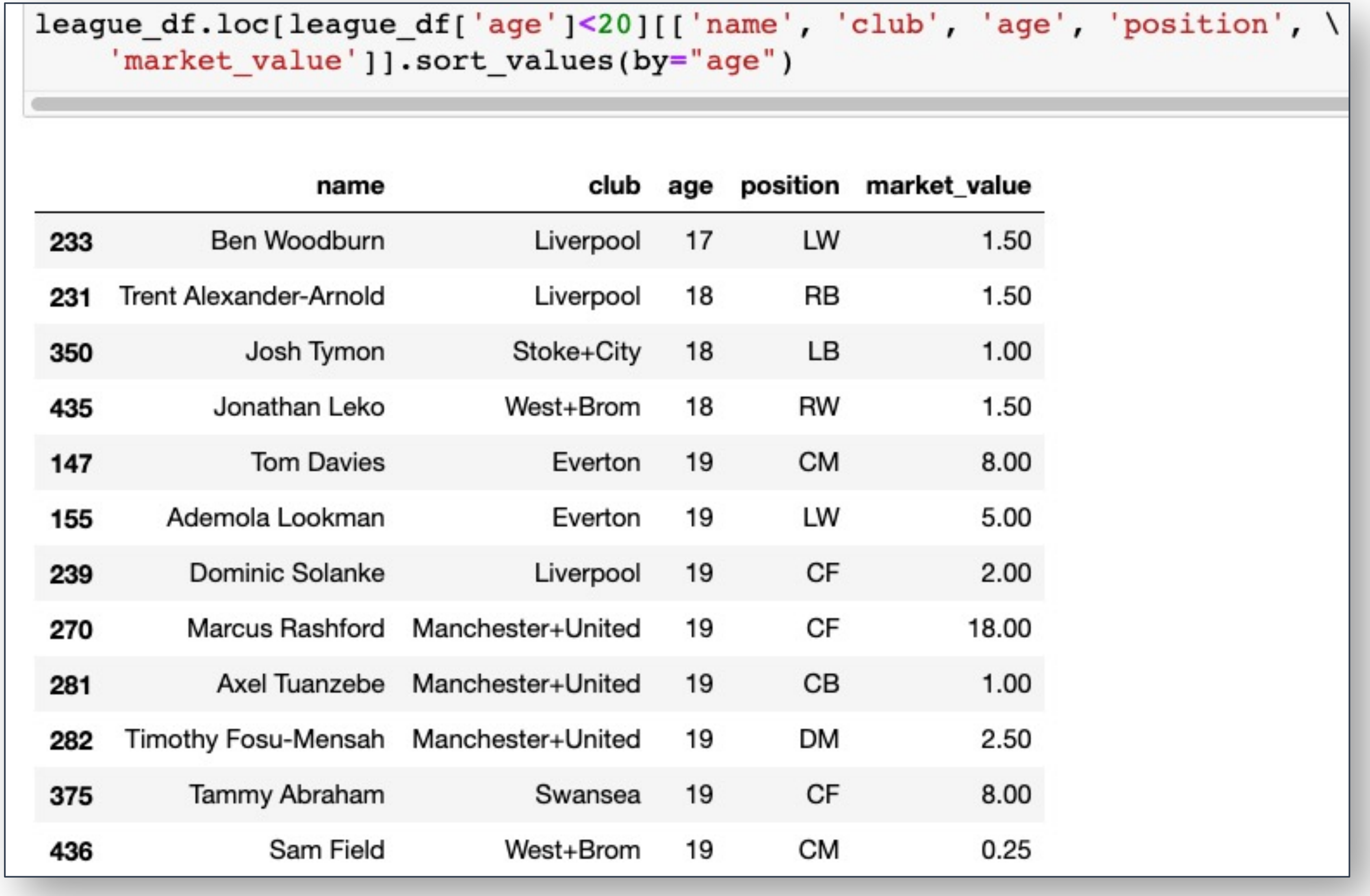


EDA Refresher: English Premier League



Are the outliers legit?

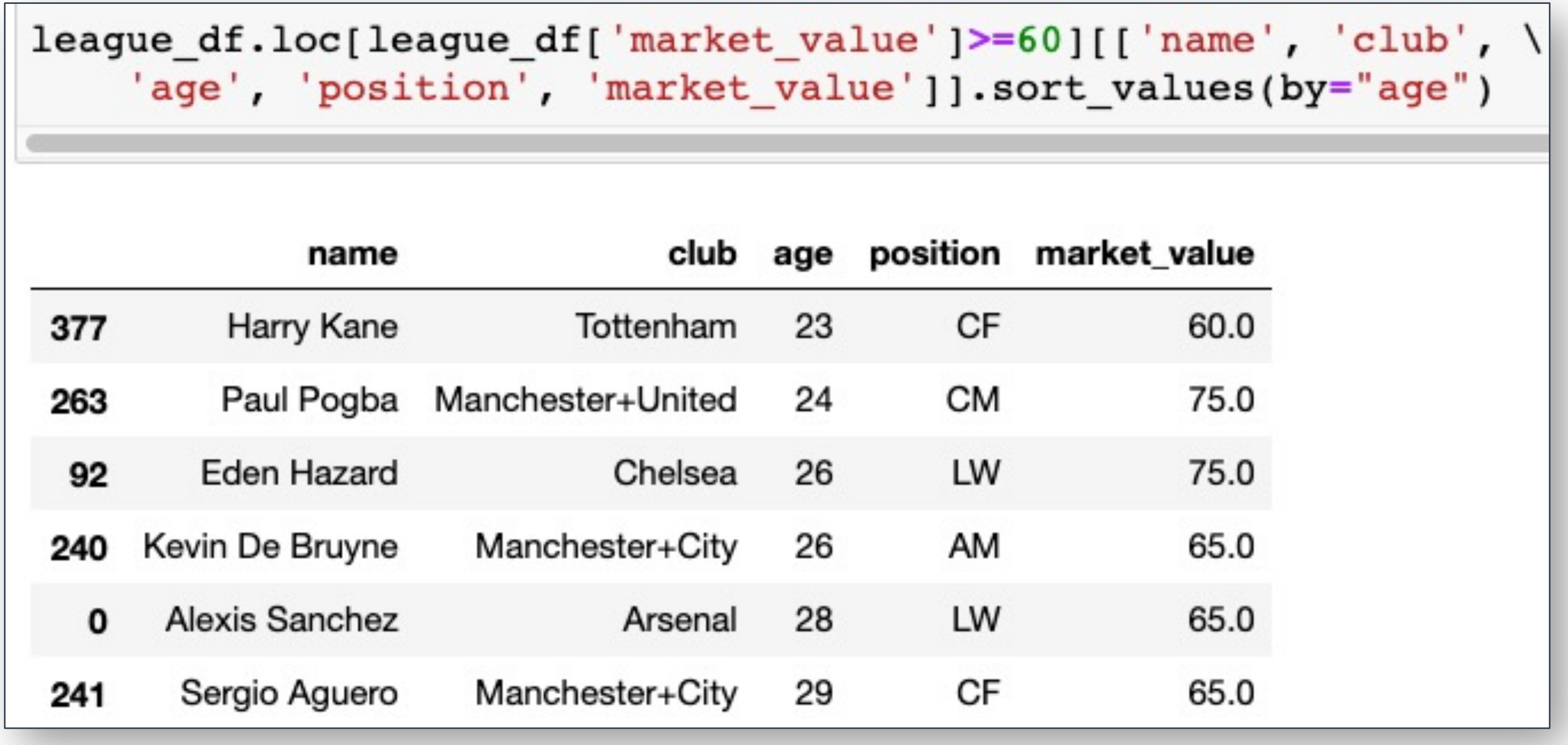
EDA Refresher: English Premier League



```
league_df.loc[league_df['age']<20][['name', 'club', 'age', 'position', \
    'market_value']].sort_values(by="age")
```

	name	club	age	position	market_value
233	Ben Woodburn	Liverpool	17	LW	1.50
231	Trent Alexander-Arnold	Liverpool	18	RB	1.50
350	Josh Tymon	Stoke+City	18	LB	1.00
435	Jonathan Leko	West+Brom	18	RW	1.50
147	Tom Davies	Everton	19	CM	8.00
155	Ademola Lookman	Everton	19	LW	5.00
239	Dominic Solanke	Liverpool	19	CF	2.00
270	Marcus Rashford	Manchester+United	19	CF	18.00
281	Axel Tuanzebe	Manchester+United	19	CB	1.00
282	Timothy Fosu-Mensah	Manchester+United	19	DM	2.50
375	Tammy Abraham	Swansea	19	CF	8.00
436	Sam Field	West+Brom	19	CM	0.25

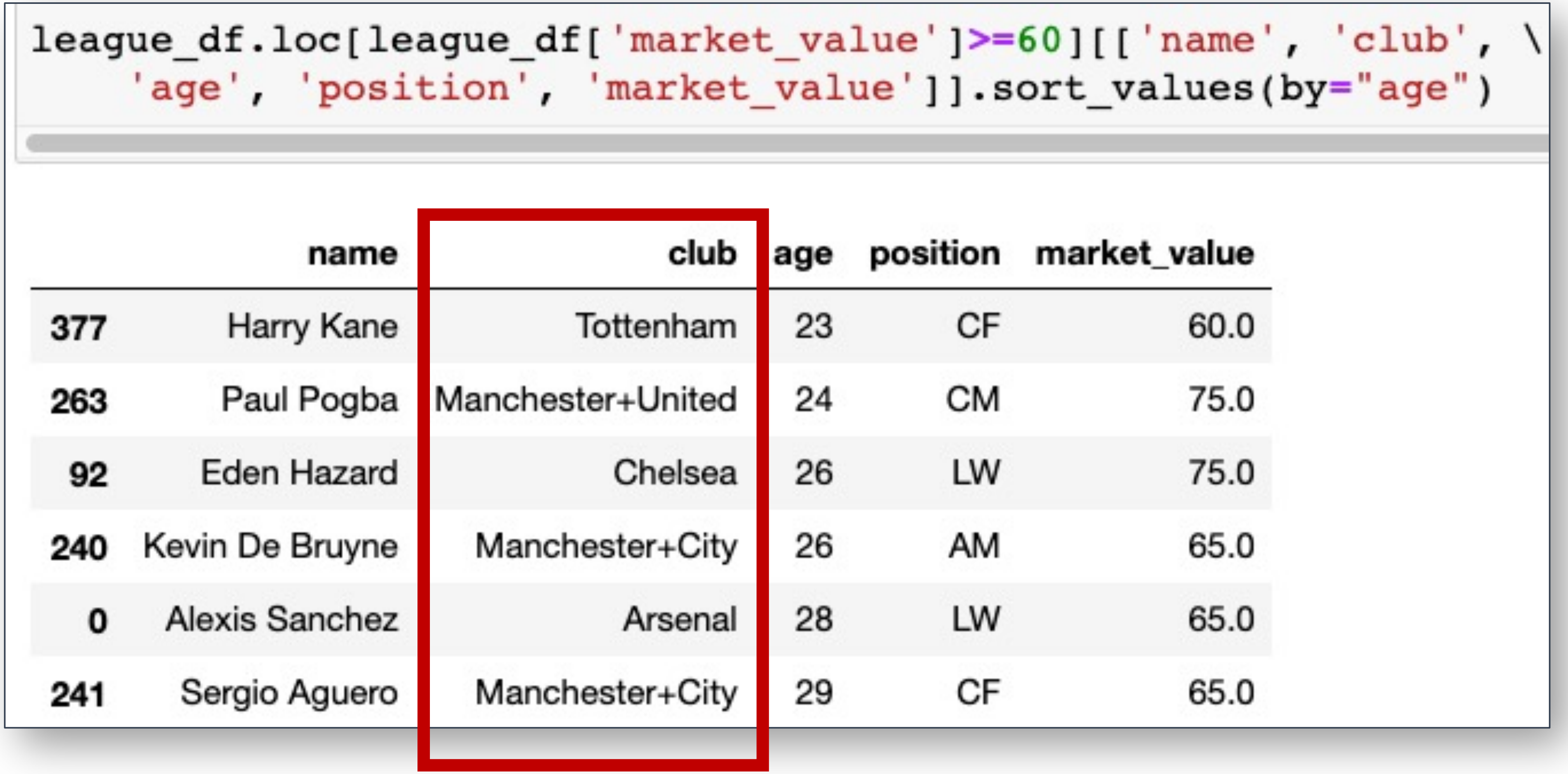
EDA Refresher: English Premier League



```
league_df.loc[league_df['market_value']>=60][['name', 'club', \
    'age', 'position', 'market_value']].sort_values(by="age")
```

	name	club	age	position	market_value
377	Harry Kane	Tottenham	23	CF	60.0
263	Paul Pogba	Manchester+United	24	CM	75.0
92	Eden Hazard	Chelsea	26	LW	75.0
240	Kevin De Bruyne	Manchester+City	26	AM	65.0
0	Alexis Sanchez	Arsenal	28	LW	65.0
241	Sergio Aguero	Manchester+City	29	CF	65.0

EDA Refresher: English Premier League



```python
league_df.loc[league_df['market_value']>=60][['name', 'club', \
    'age', 'position', 'market_value']].sort_values(by="age")
```

	name	club	age	position	market_value
377	Harry Kane	Tottenham	23	CF	60.0
263	Paul Pogba	Manchester+United	24	CM	75.0
92	Eden Hazard	Chelsea	26	LW	75.0
240	Kevin De Bruyne	Manchester+City	26	AM	65.0
0	Alexis Sanchez	Arsenal	28	LW	65.0
241	Sergio Aguero	Manchester+City	29	CF	65.0

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

DISCLAIMER. Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design

- Communication
  - Motivation
  - Key Considerations

DISCLAIMER.

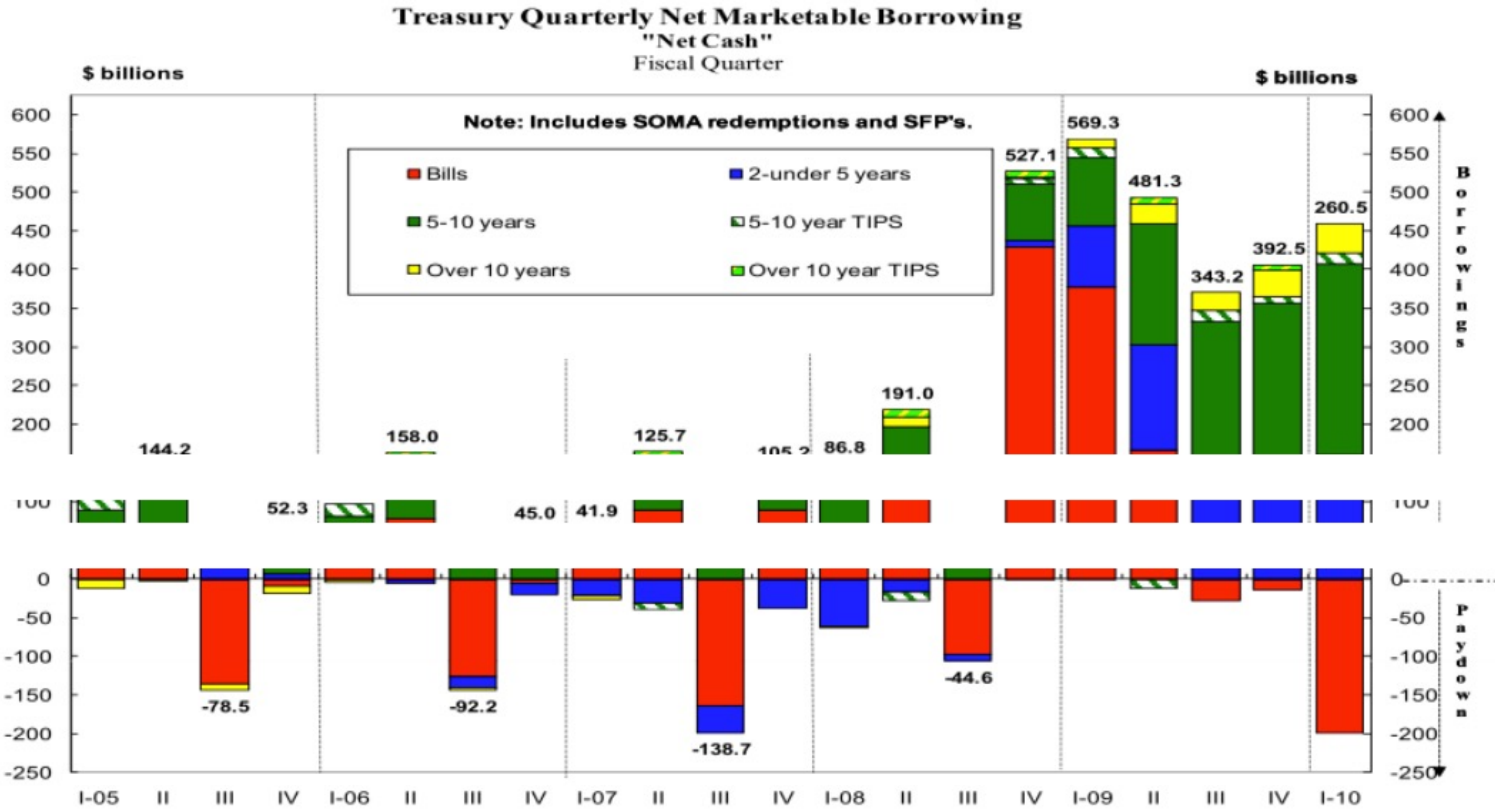
Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.

Agenda

- EDA Refresher
- Effective Visualization
    - Graphical Integrity
    - Scope
    - Displays
    - Sensible Design
- Communication
    - Motivation
    - Key Considerations

DISCLAIMER. Some of these examples involve political data. In no way should this be taken as a signal of my support or endorsement in any beliefs; the point is merely to convey good and bad choices when it comes to effective visualization.

Ineffective Visualization



Ineffective Visualization

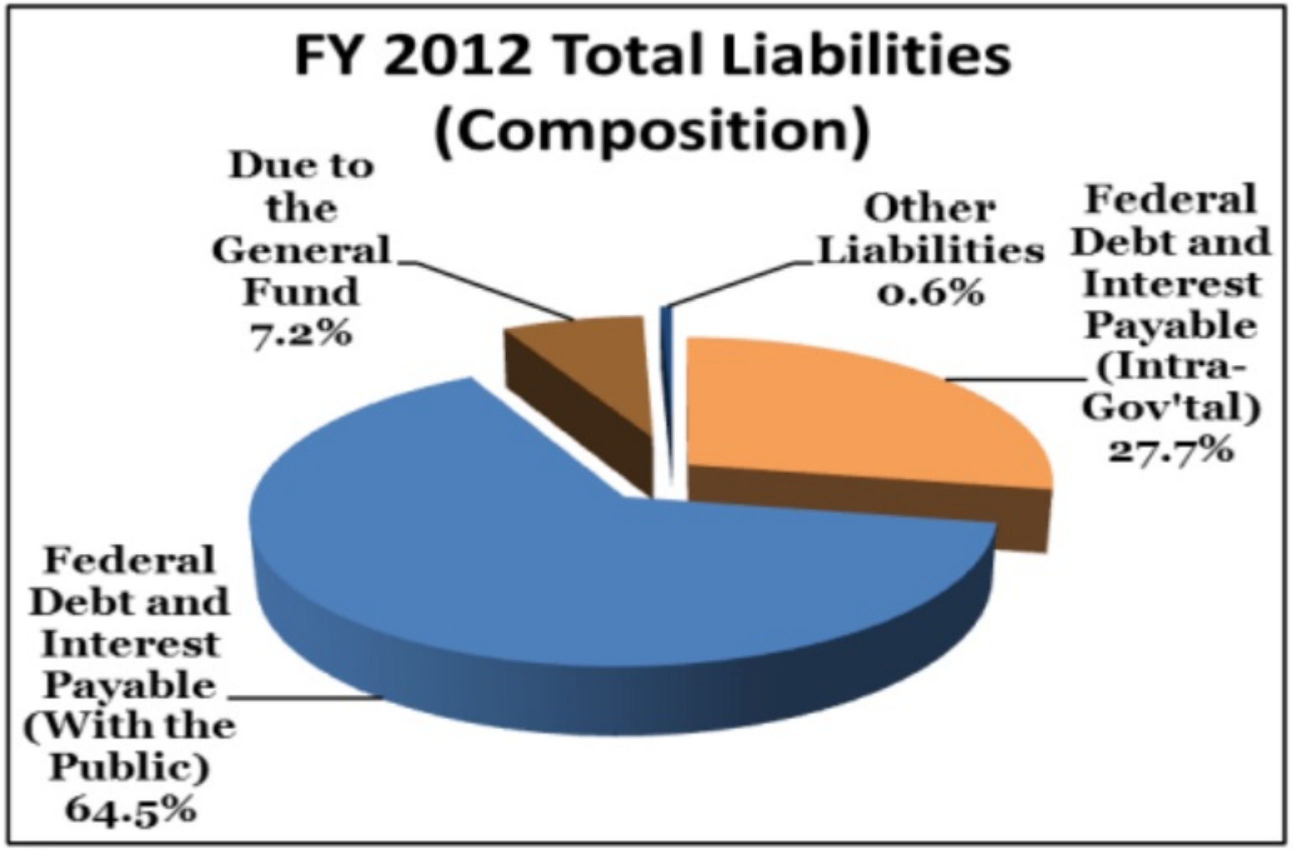
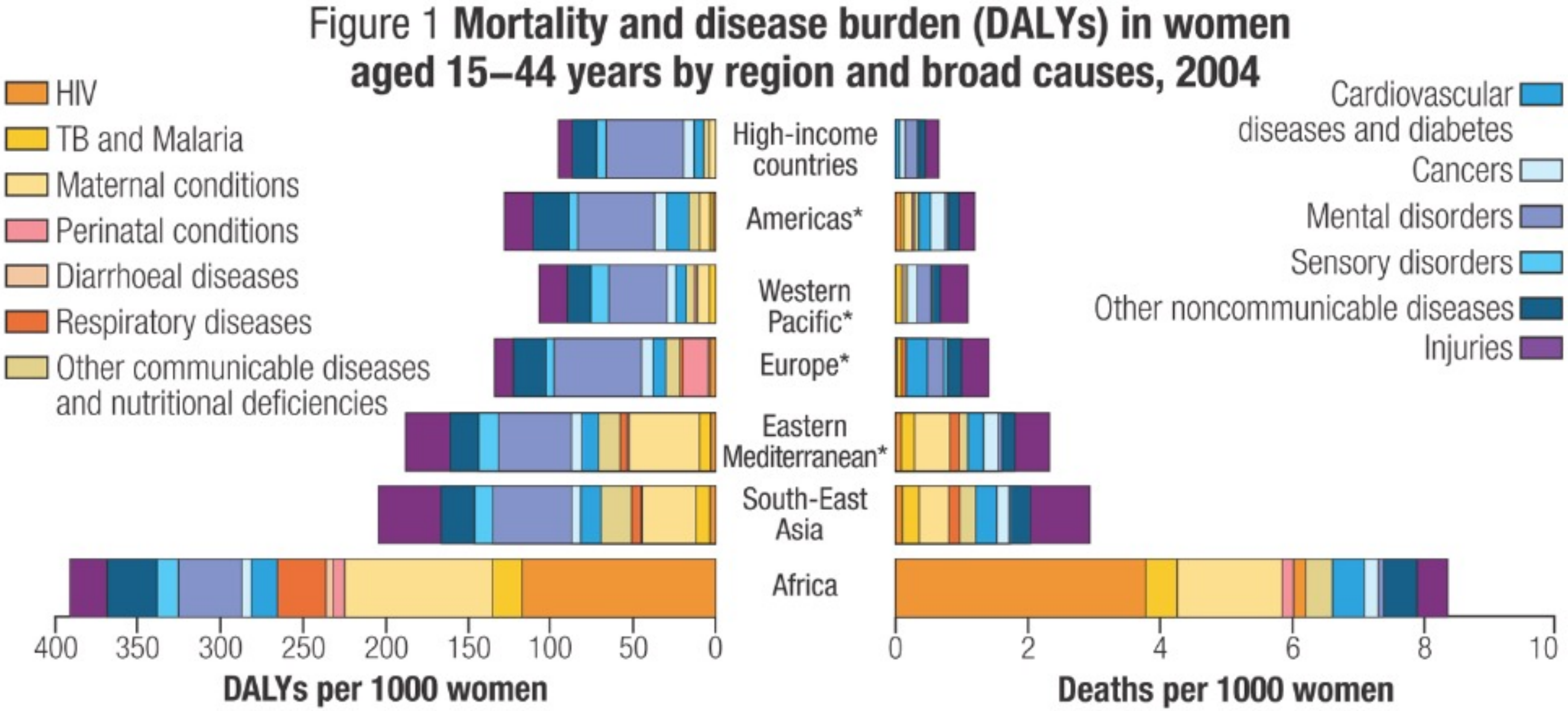


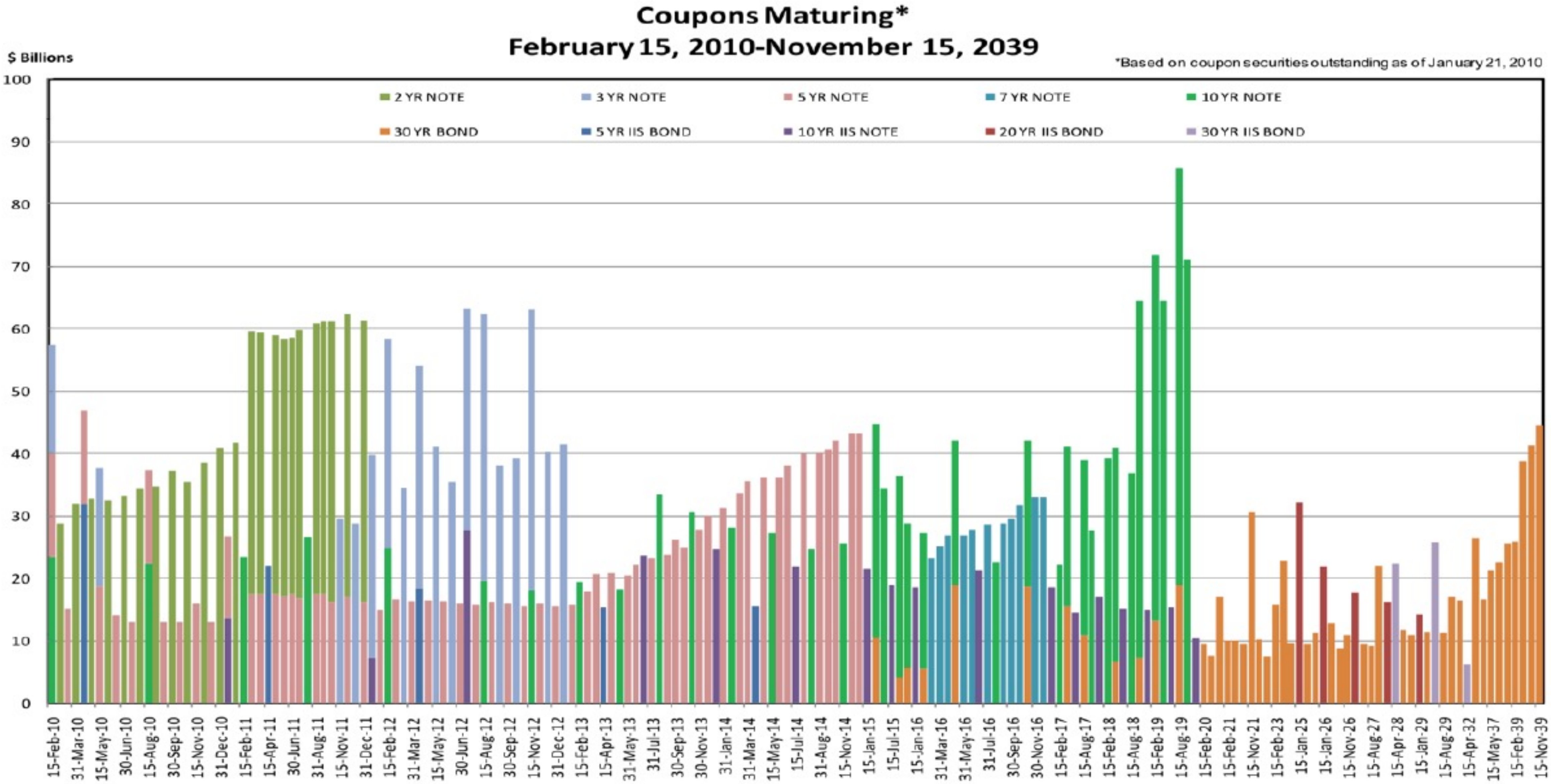
Figure 10

Ineffective Visualization



* High-income countries are excluded from the regional groups.
Source: World Health Organization.[1]

Ineffective Visualization



Ineffective Visualization

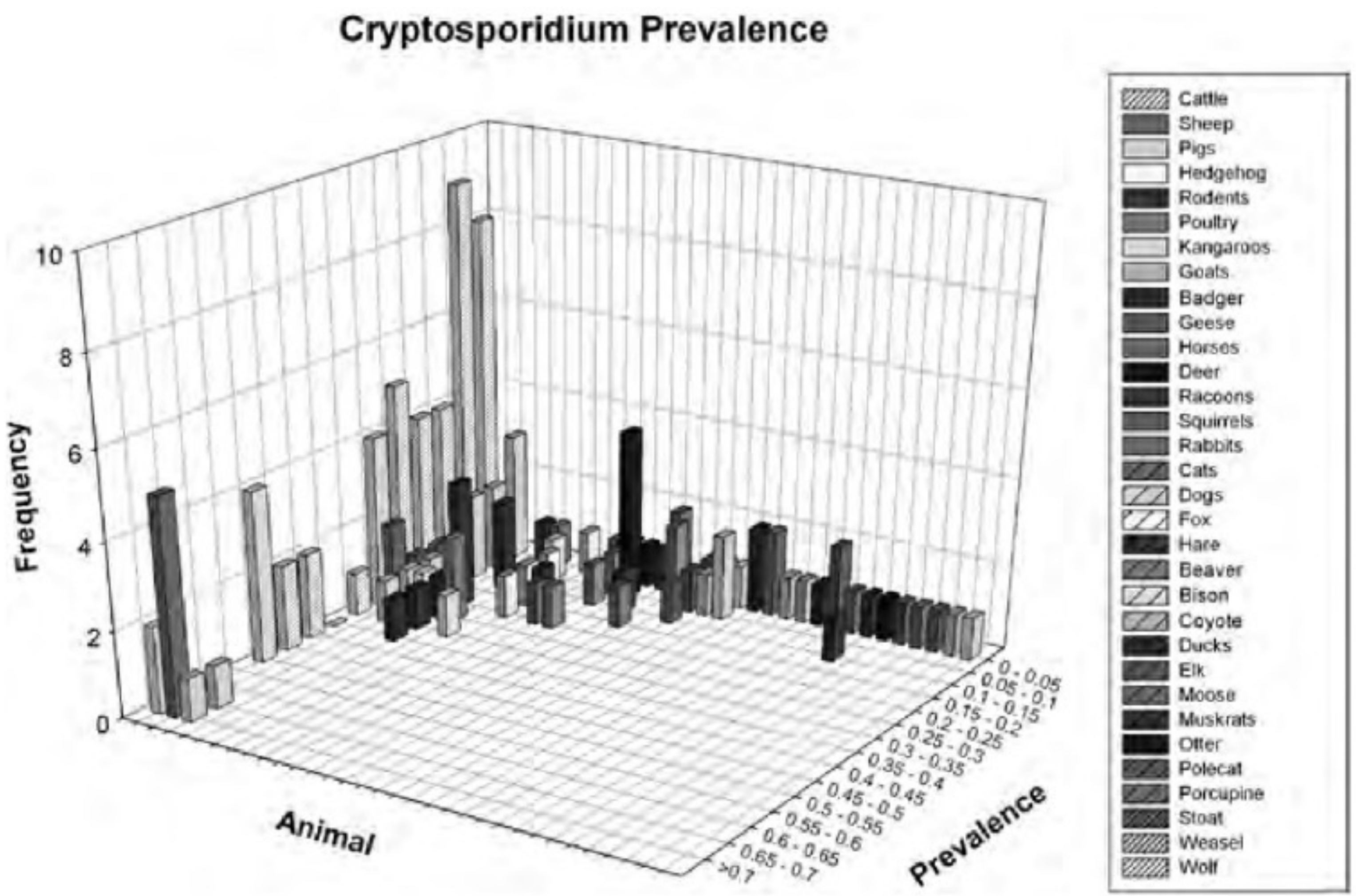


Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

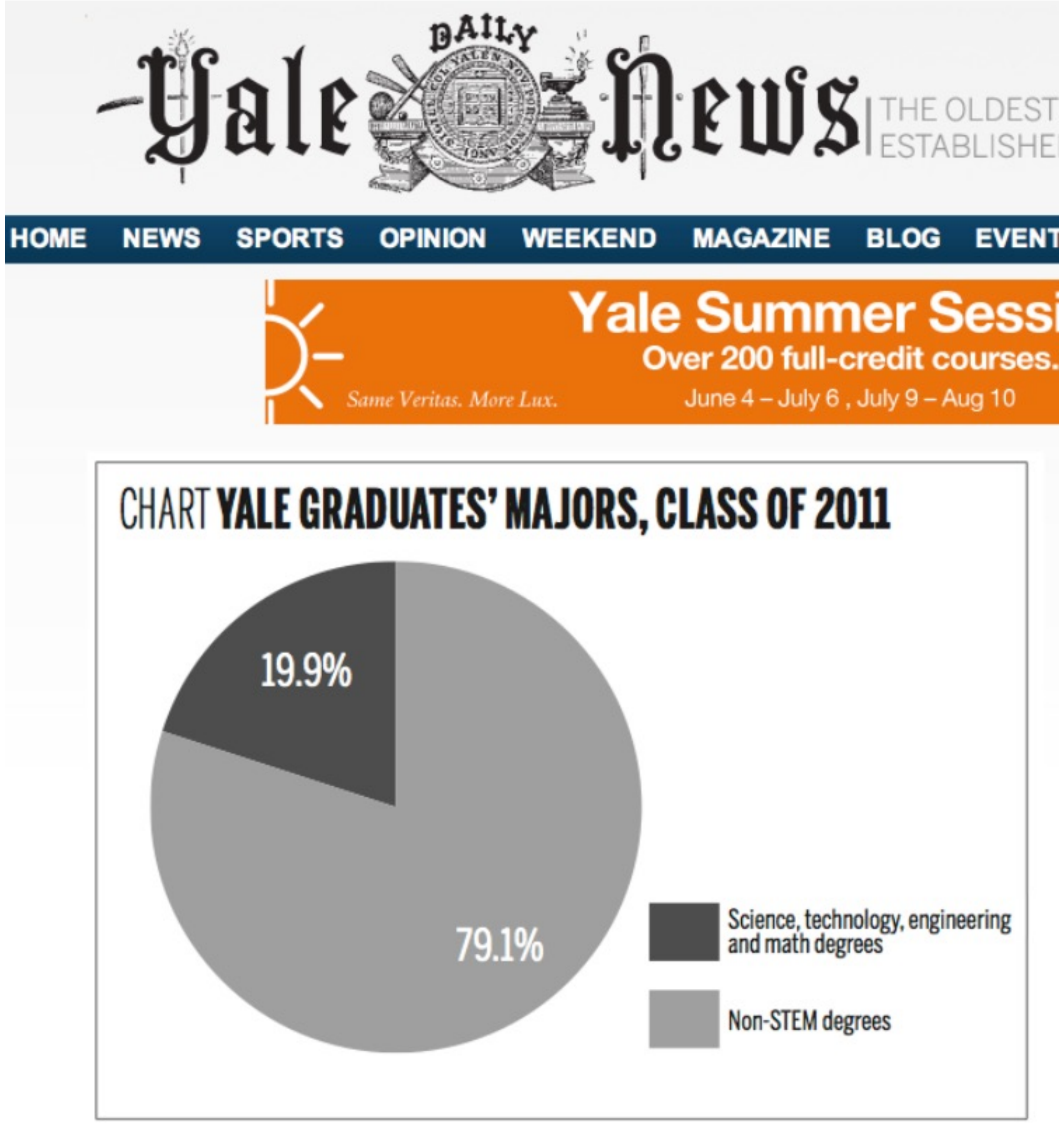
Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Graphical Integrity: What's wrong?

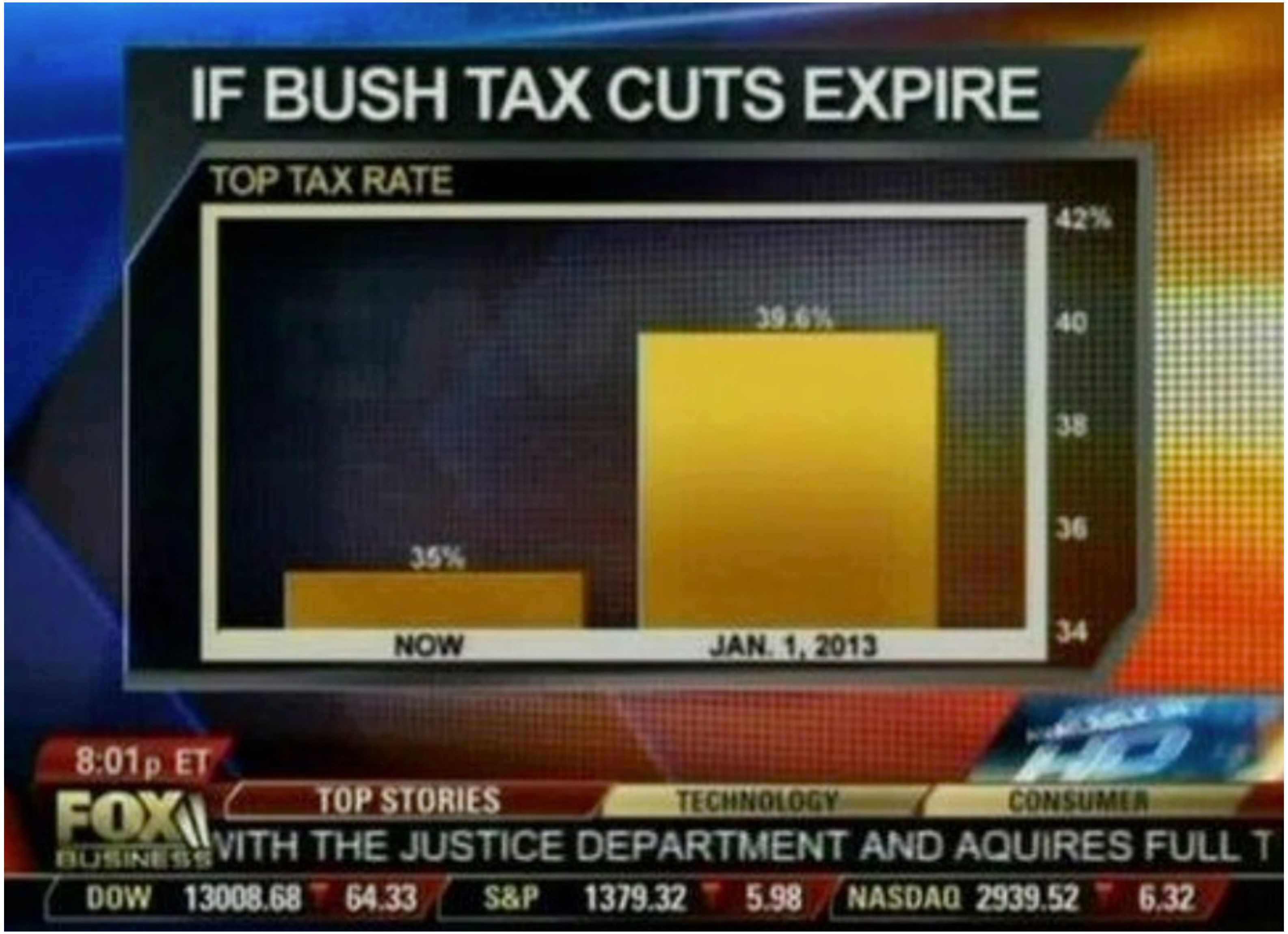


Yale Daily News THE OLDEST ESTABLISHE

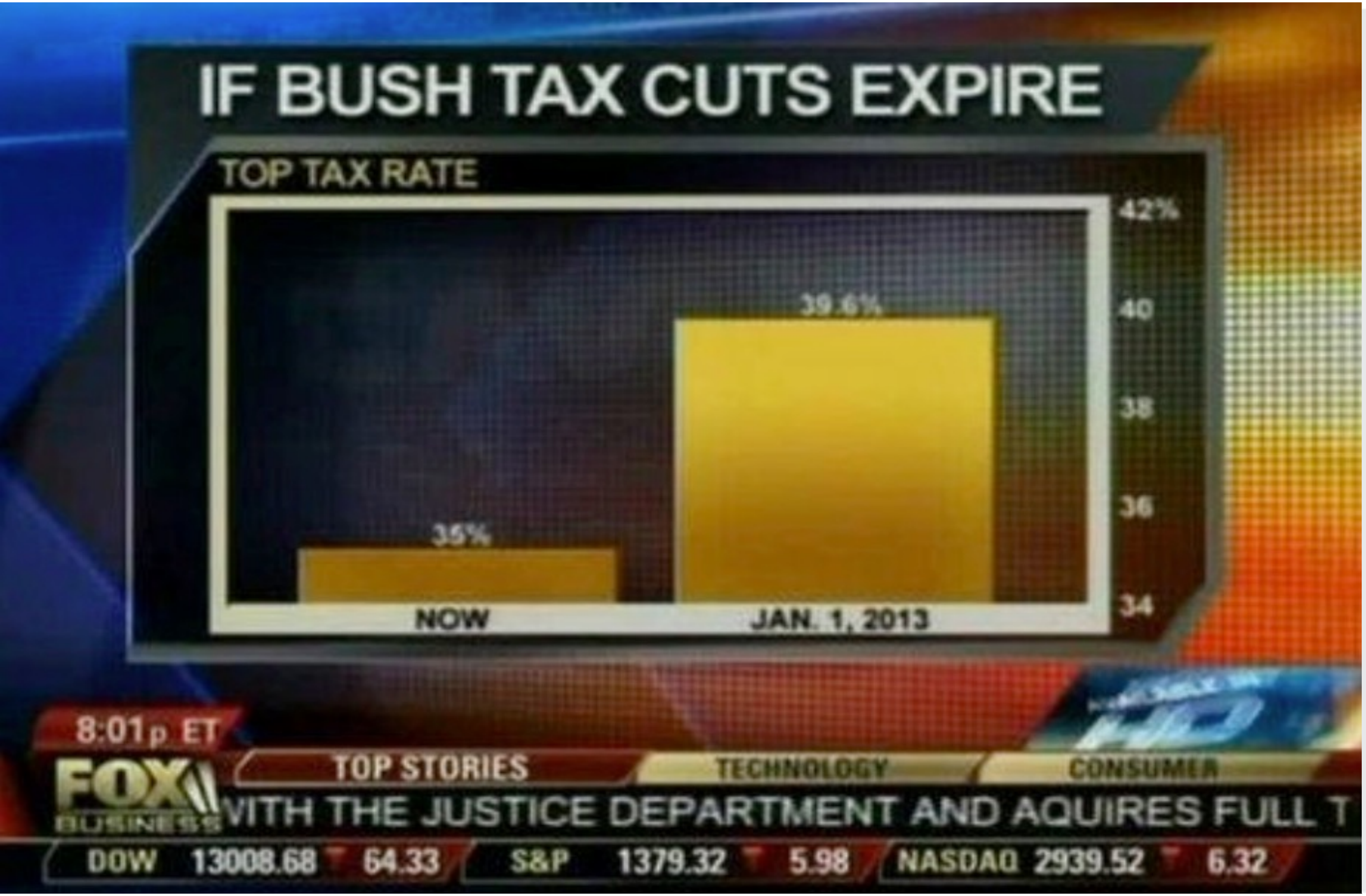
HOME NEWS SPORTS OPINION WEEKEND MAGAZINE BLOG EVENT

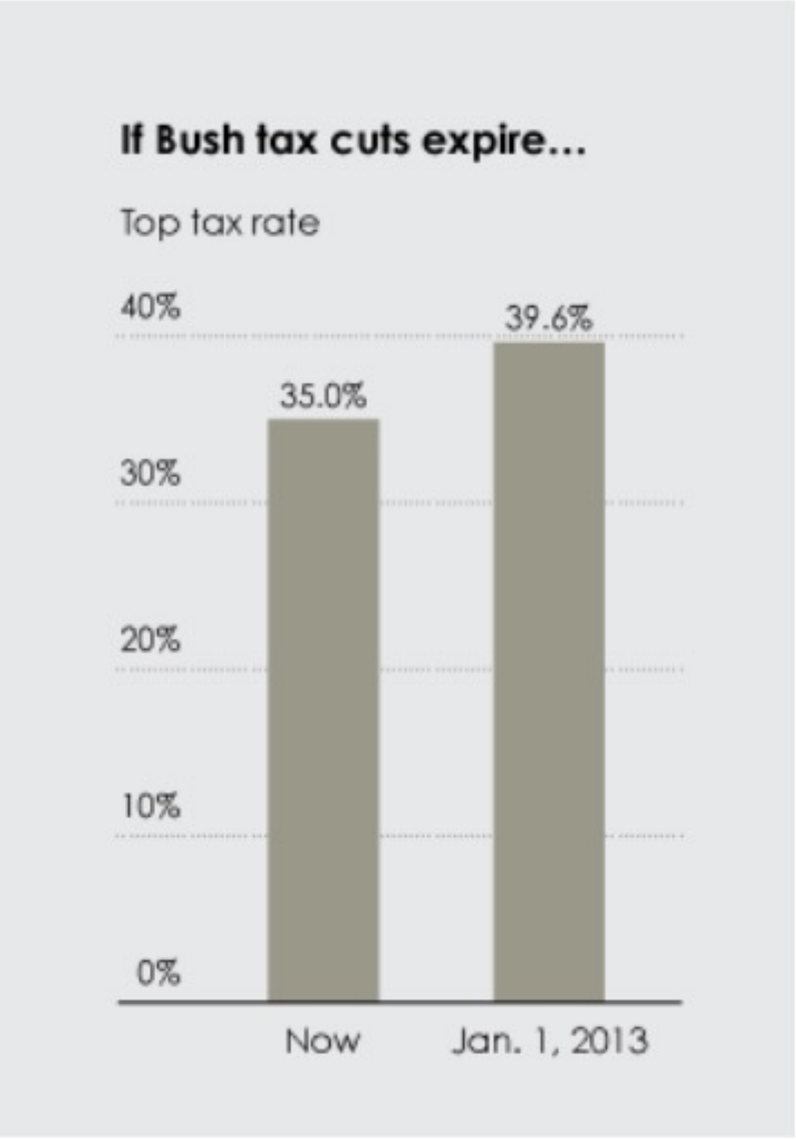
CHART **YALE GRADUATES' MAJORS, CLASS OF 2011**

Graphical Integrity: What’s wrong?

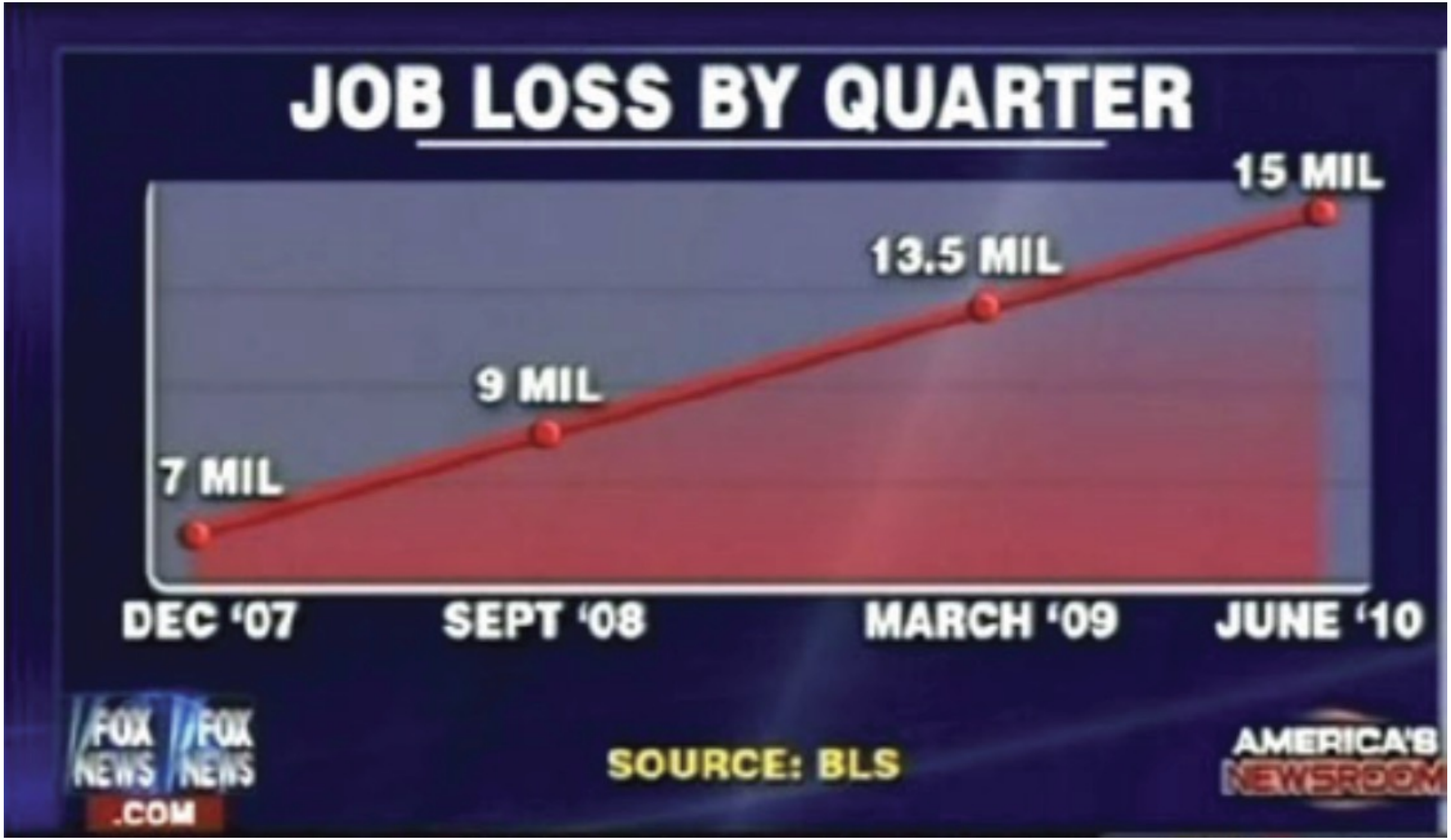


Graphical Integrity: ~~What's wrong?~~ Real Story

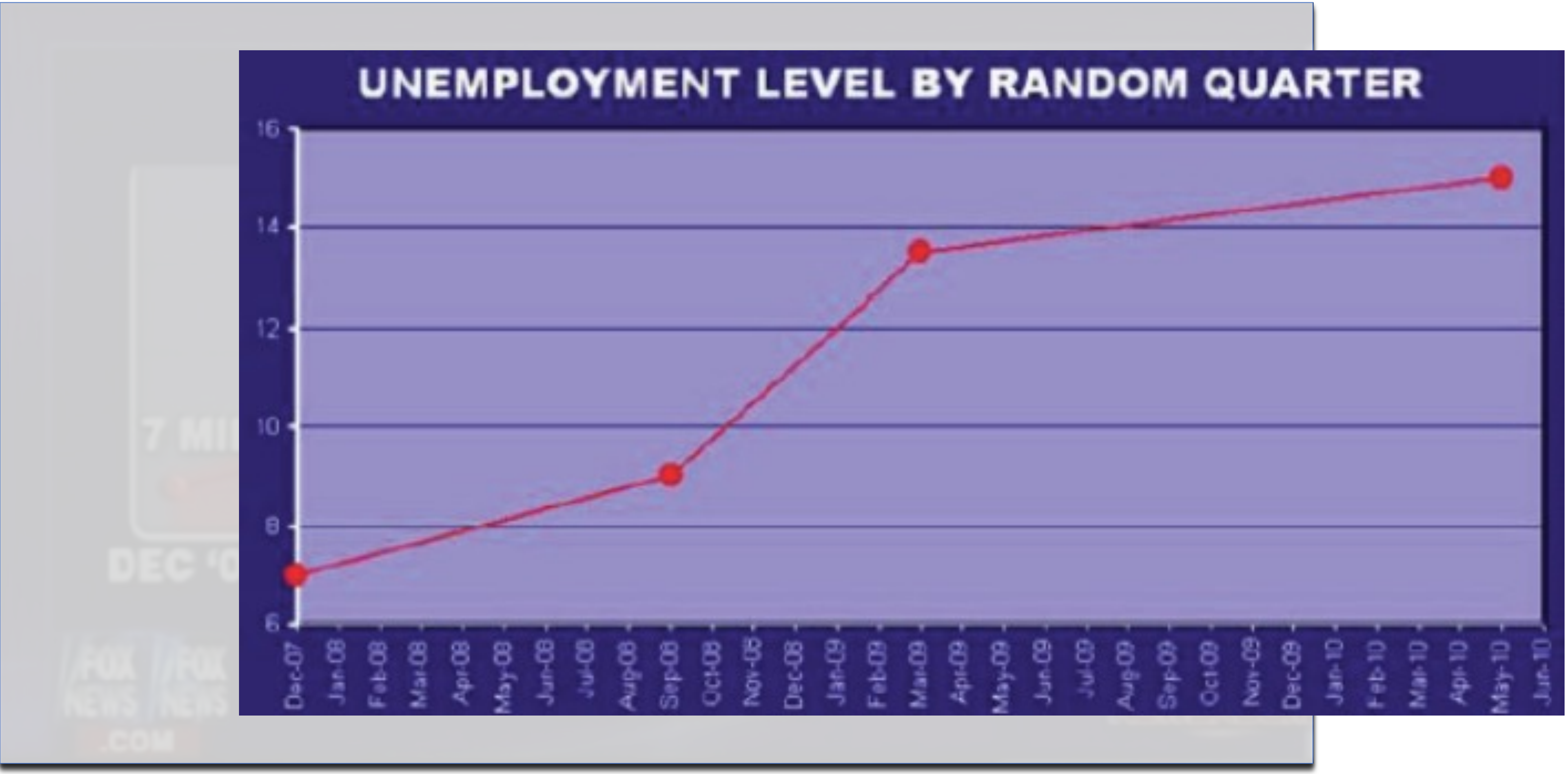




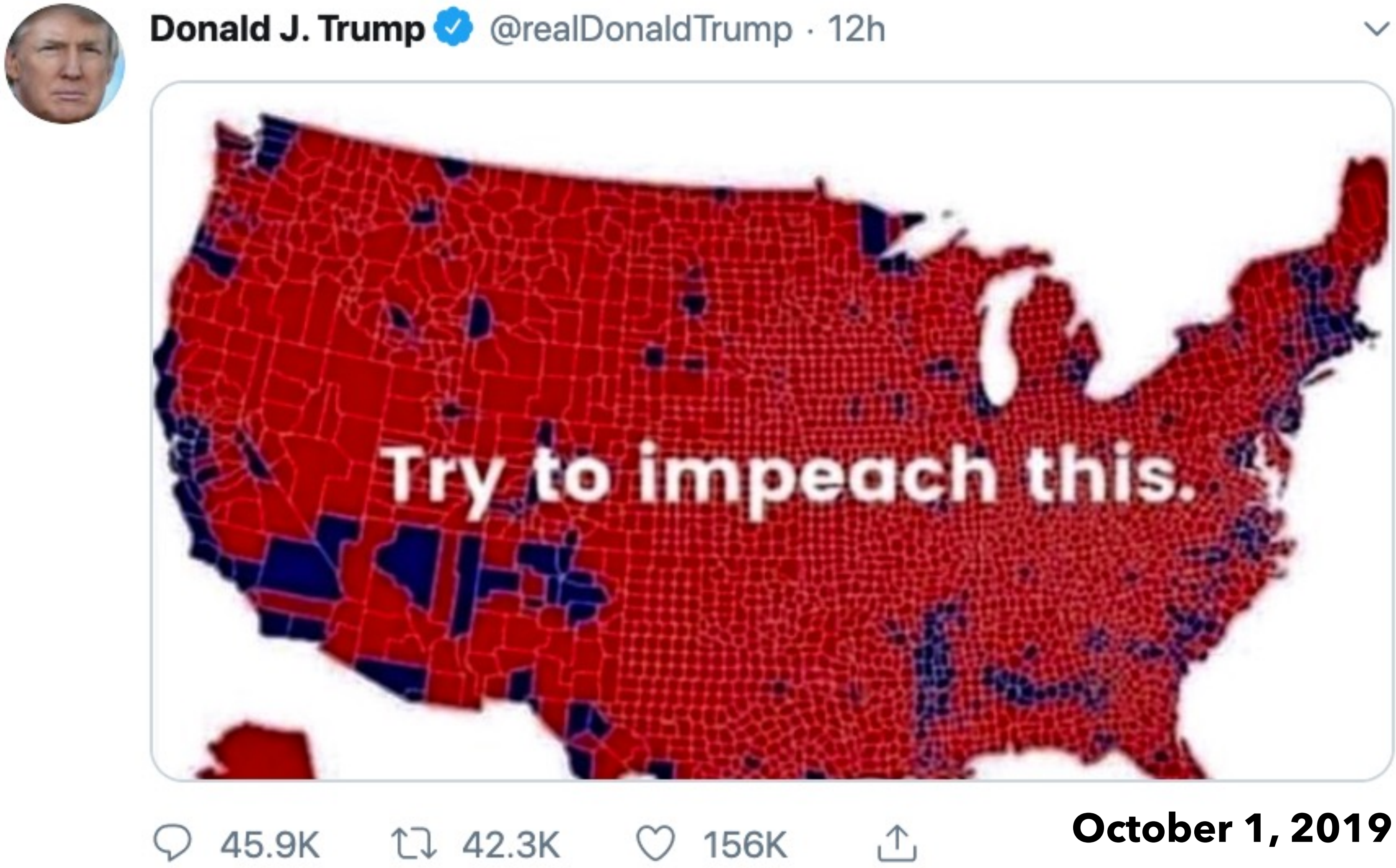
Graphical Integrity: What's wrong?



Graphical Integrity: ~~What's wrong?~~ Real Story



Graphical Integrity: What’s wrong?

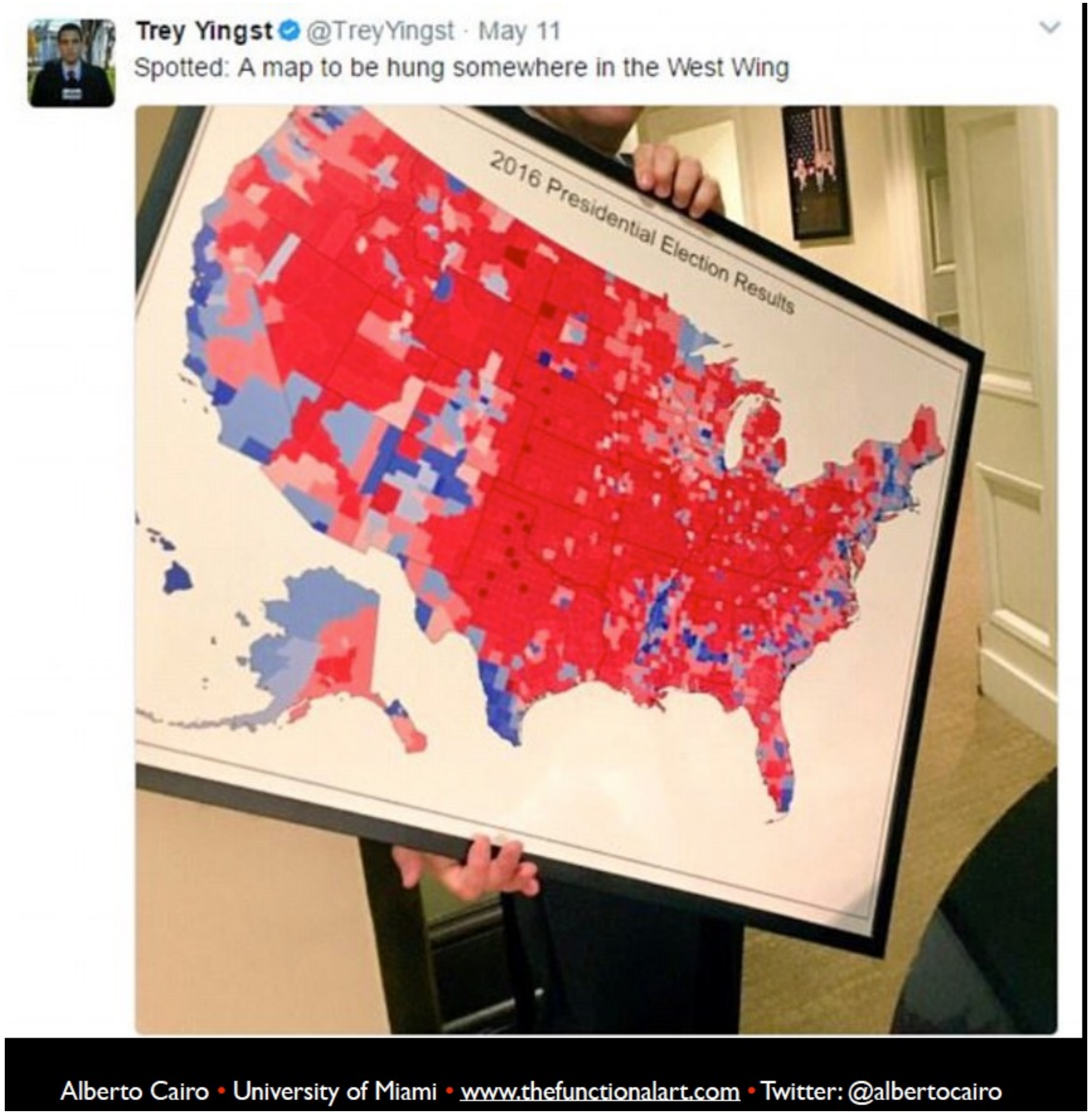


Donald J. Trump @realDonaldTrump · 12h

45.9K 42.3K 156K

October 1, 2019

Graphical Integrity: What's wrong?

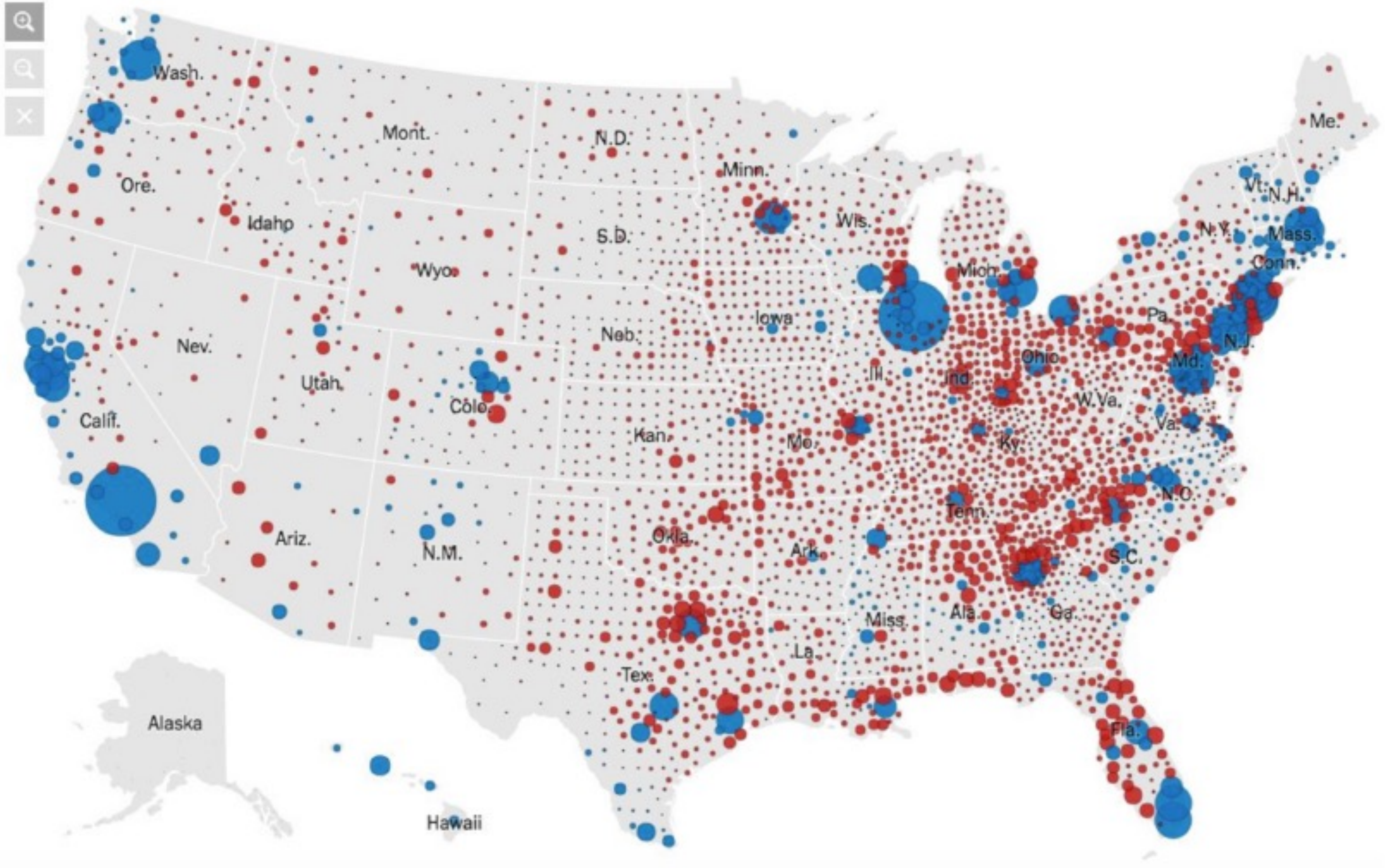


Trey Yingst @TreyYingst · May 11

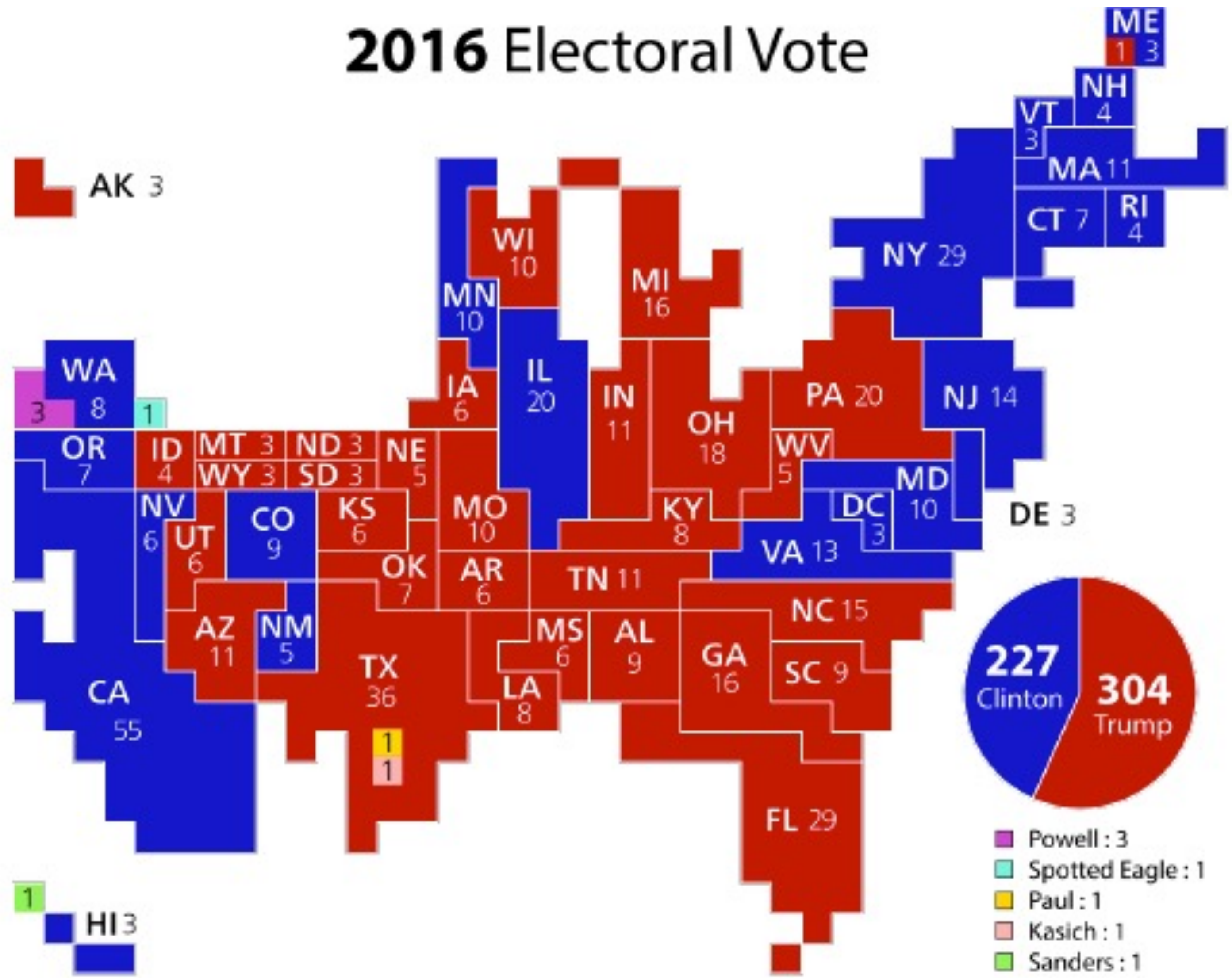
Spotted: A map to be hung somewhere in the West Wing

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

Graphical Integrity: What's wrong?



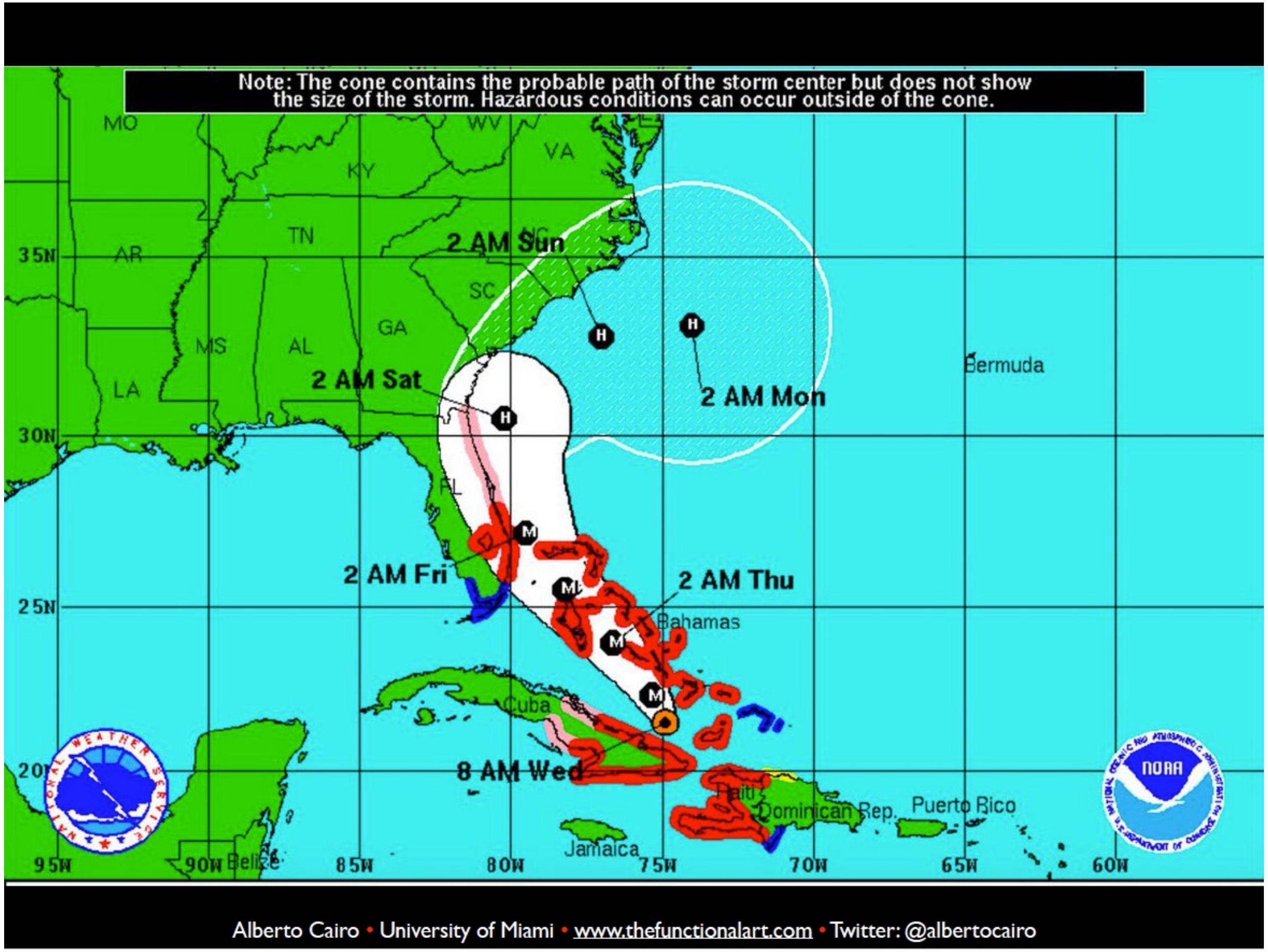
Graphical Integrity: ~~What's wrong?~~ Real Story



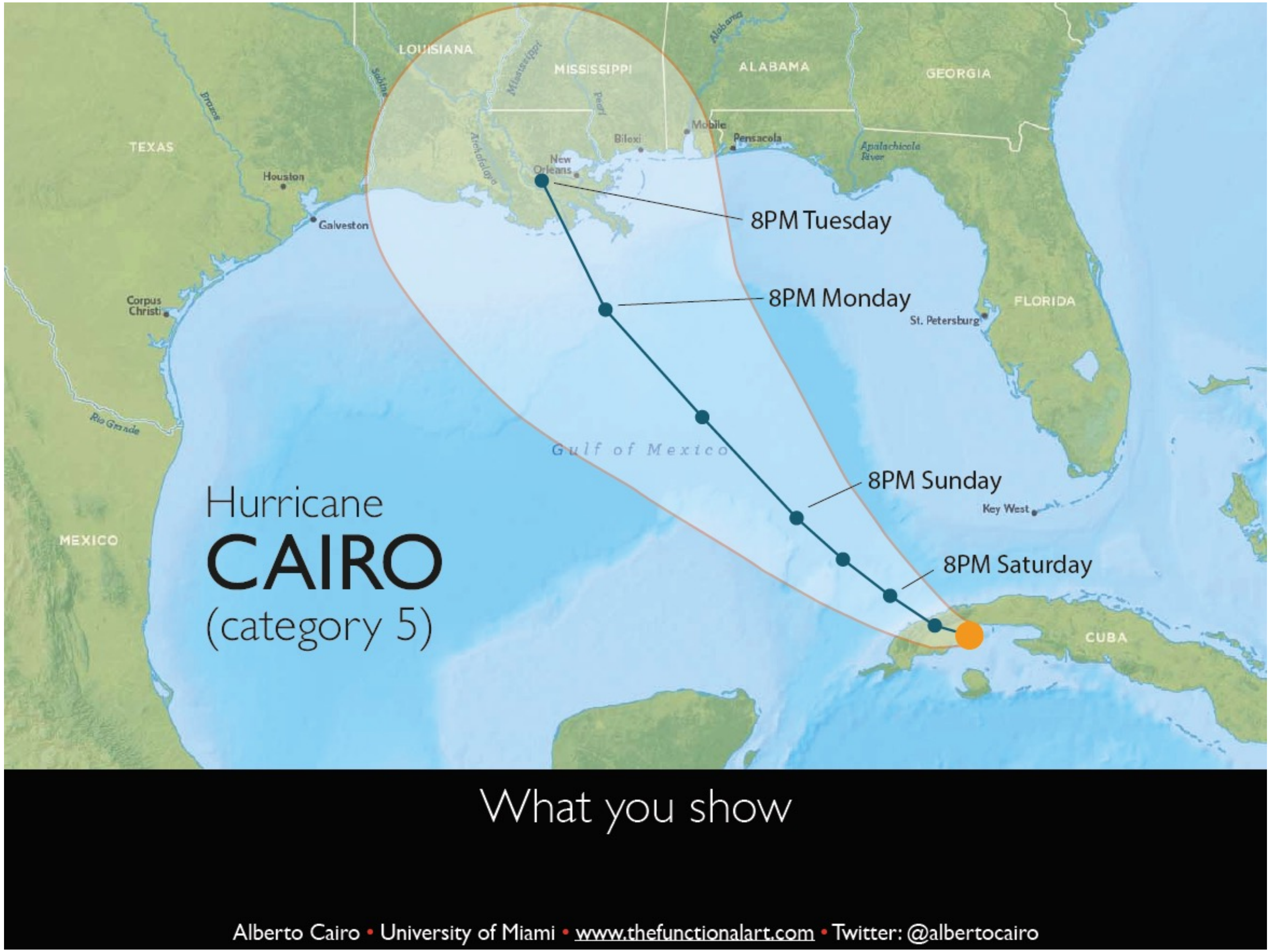
Graphical Integrity: What’s wrong?

Lesson: be proportional

Graphical Integrity: What’s wrong?


Note: The cone contains the probable path of the storm center but does not show the size of the storm. Hazardous conditions can occur outside of the cone.
MO
WV
VA
KY
TN
AR
NC
2 AM Sun
SC
GA
MS
AL
LA
2 AM Sat
2 AM Mon
Bermuda
FL
2 AM Fri
2 AM Thu
Bahamas
Cuba
8 AM Wed
Haiti
Dominican Rep.
Puerto Rico
Jamaica
Belize
35N
30N
25N
20N
95W
90W
85W
80W
75W
70W
65W
60W
H
M
NOAA
NATIONAL WEATHER SERVICE
Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

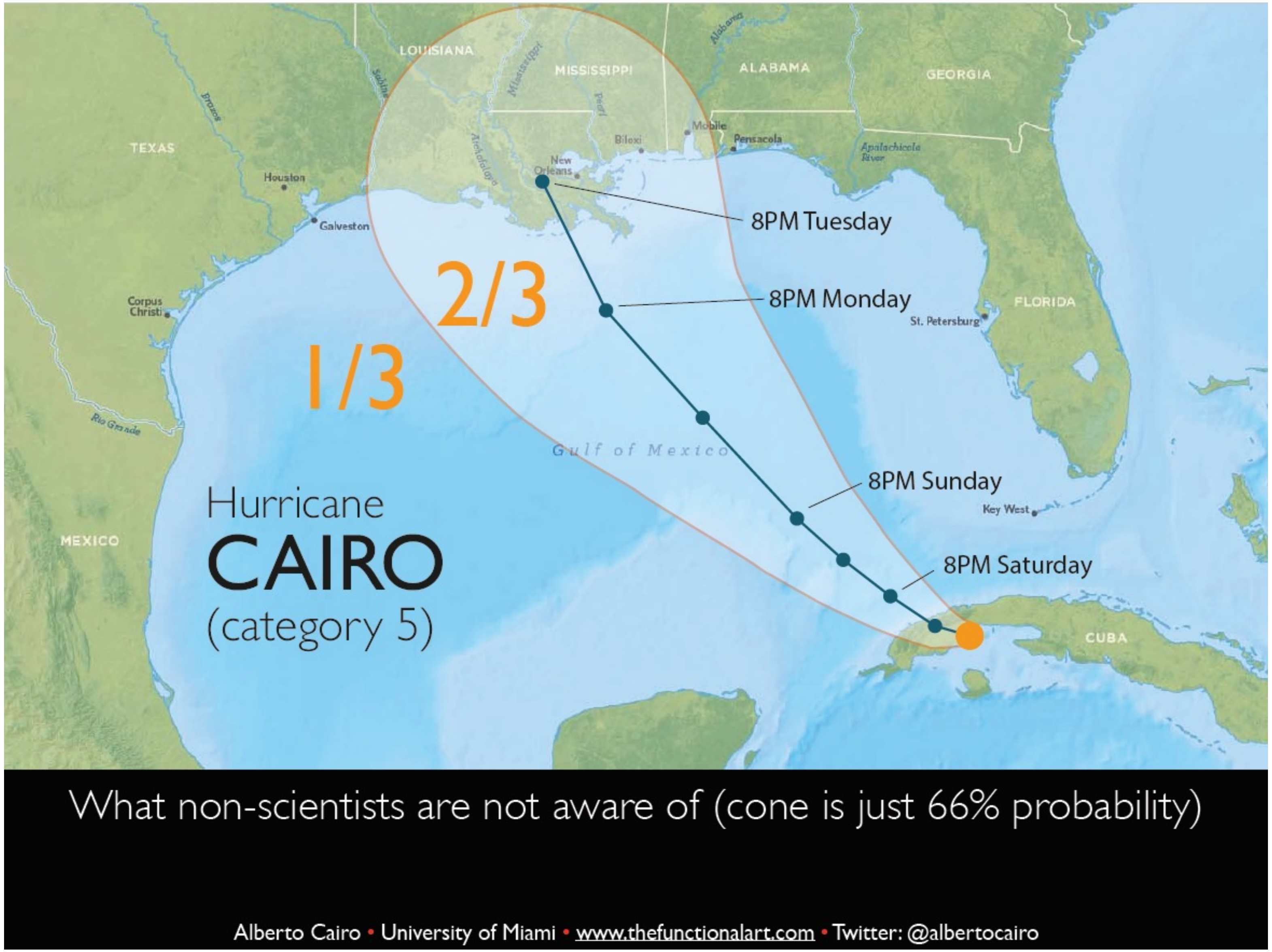
Graphical Integrity: What’s wrong?



What you show

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

Graphical Integrity: What’s wrong?

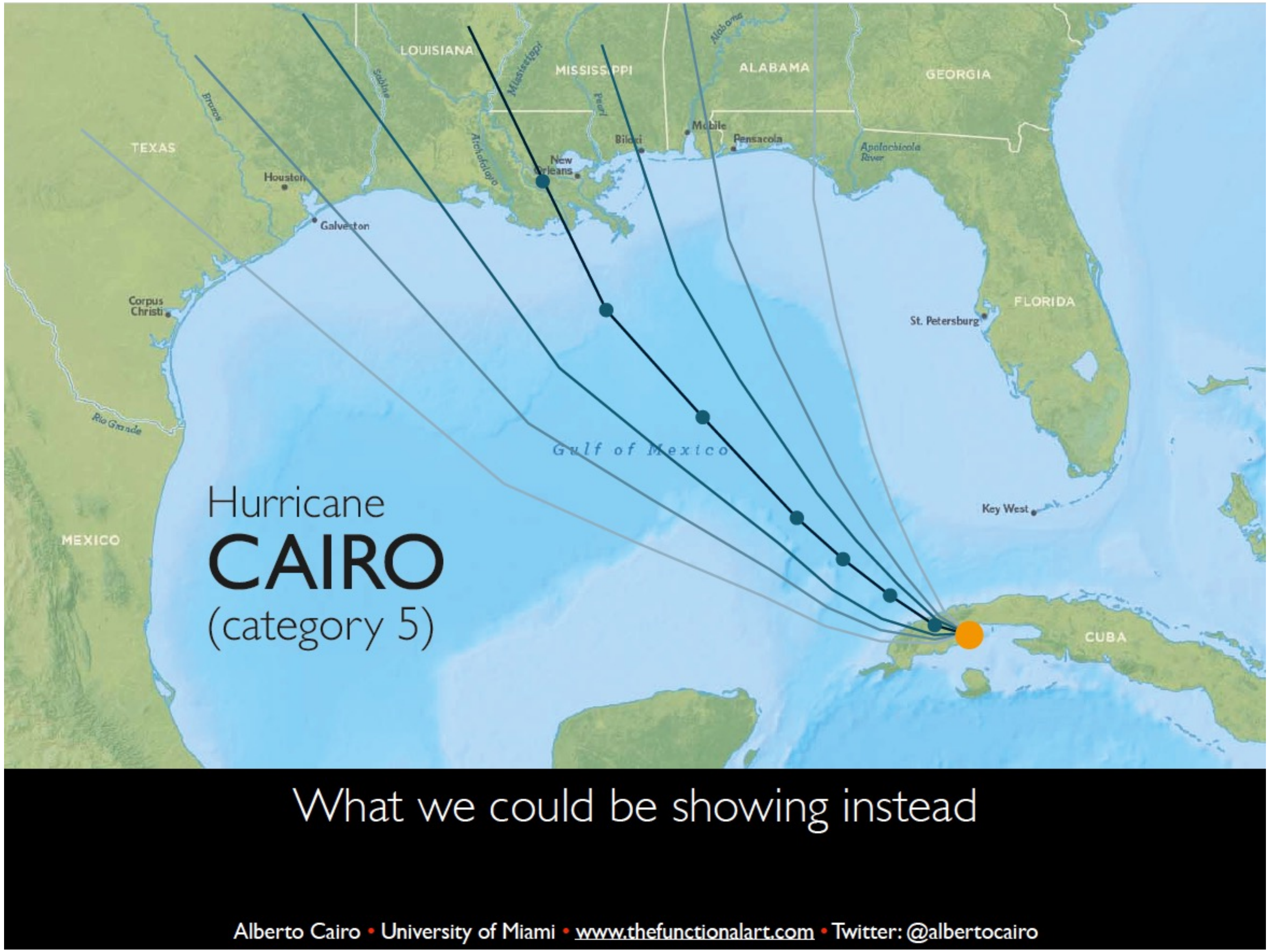


Hurricane
CAIRO
(category 5)

What non-scientists are not aware of (cone is just 66% probability)

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

Graphical Integrity: What's wrong?



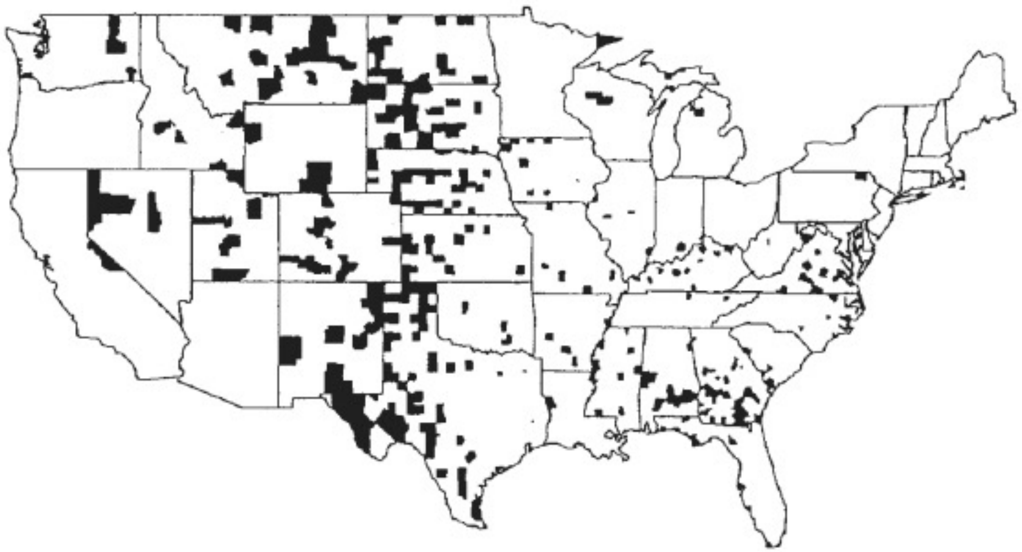
What we could be showing instead

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

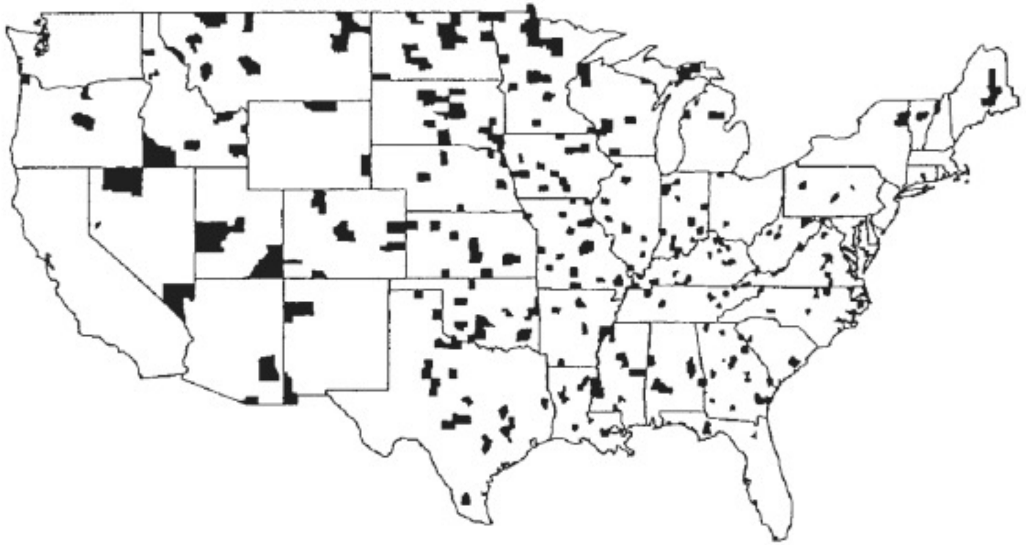
Graphical Integrity: What’s wrong?

Lesson II: include uncertainty when possible

Graphical Integrity: What's wrong?

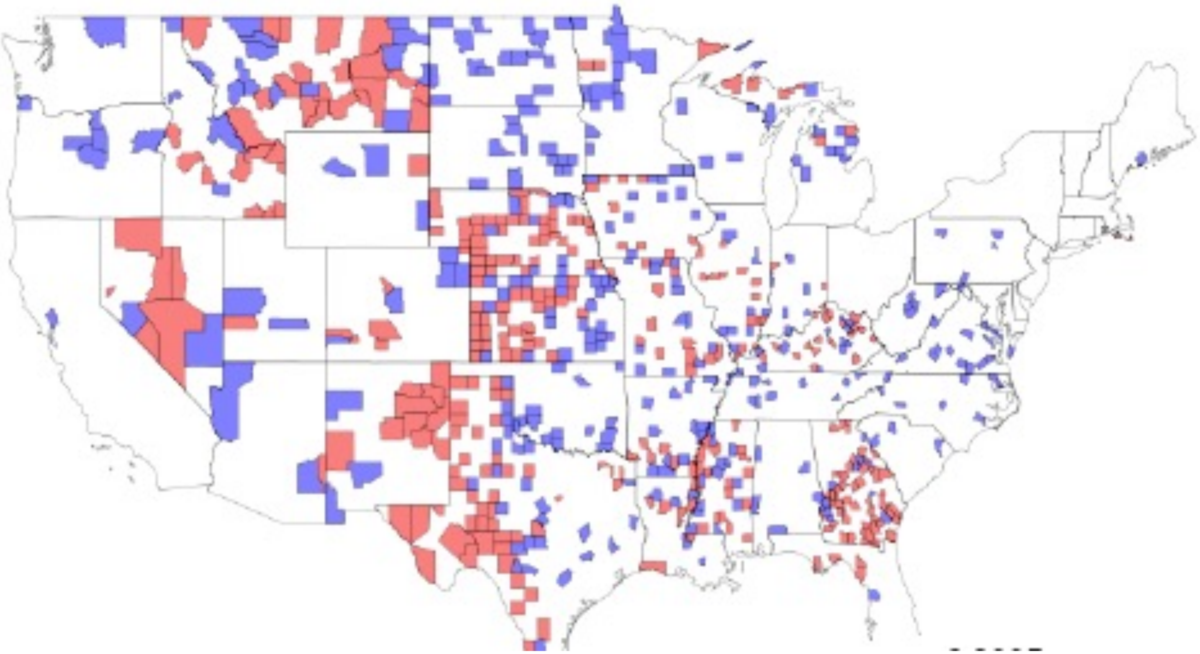


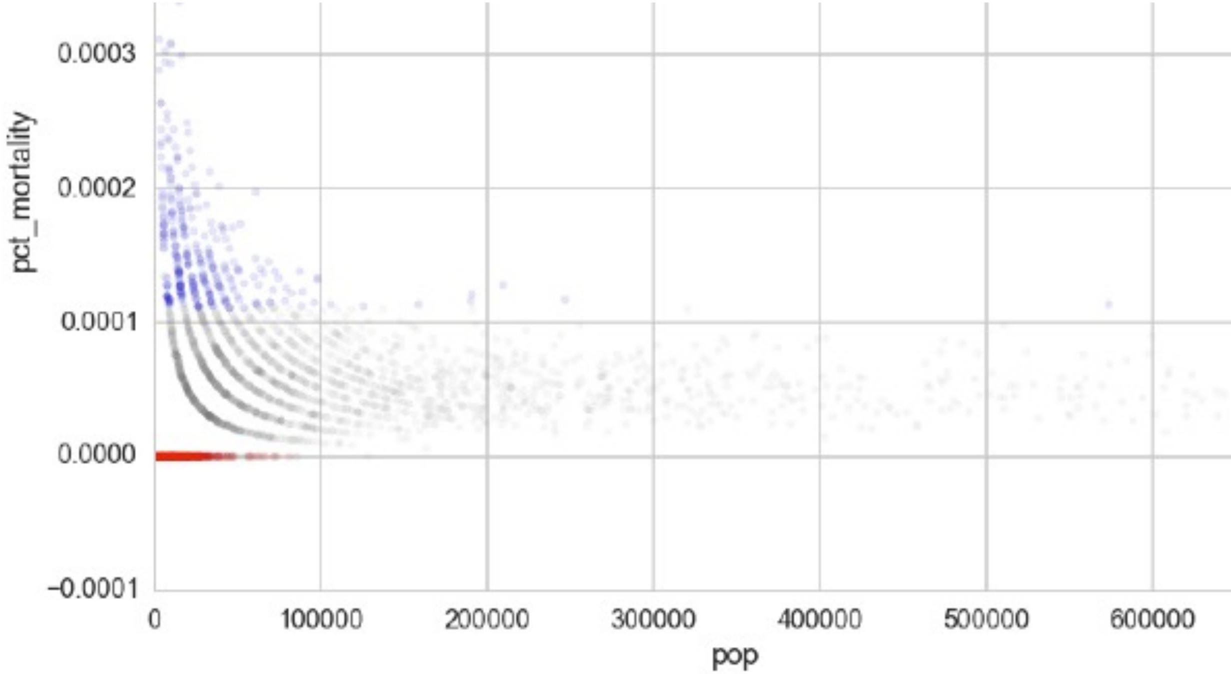
Counties with the LOWEST kidney cancer death rates (1980-1989)



Counties with the HIGHEST kidney cancer death rates (1980-1989)

Graphical Integrity: What’s wrong?





Graphical Integrity: What’s wrong?

Lesson III: plot all the data

Agenda

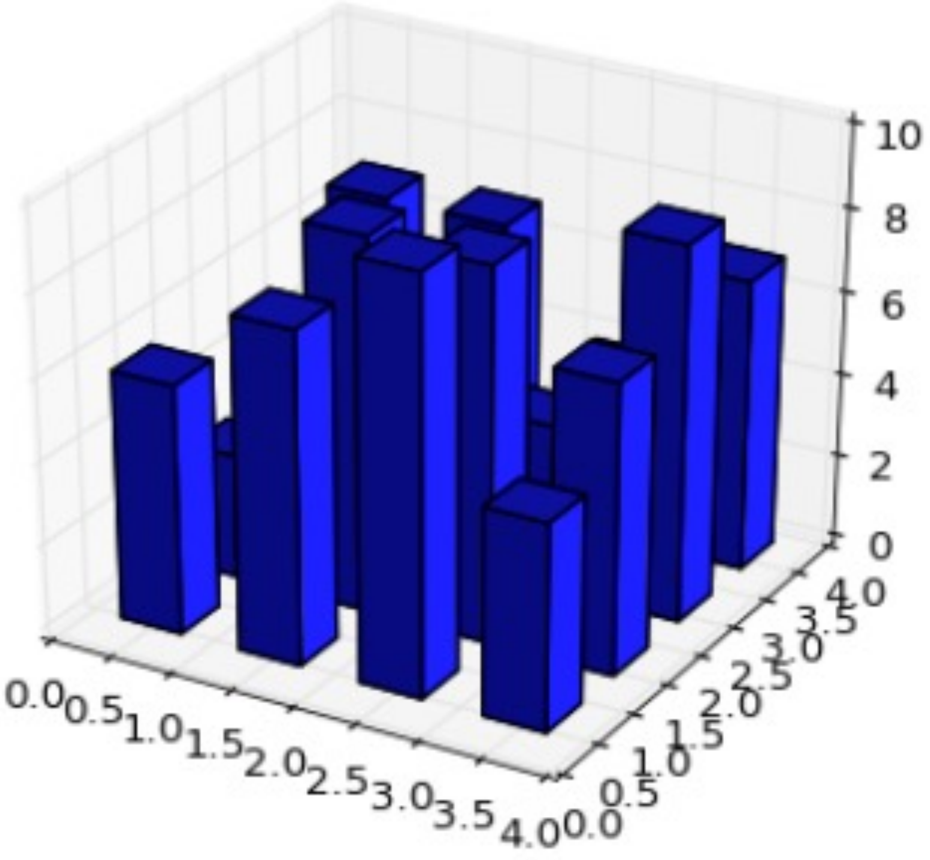
- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

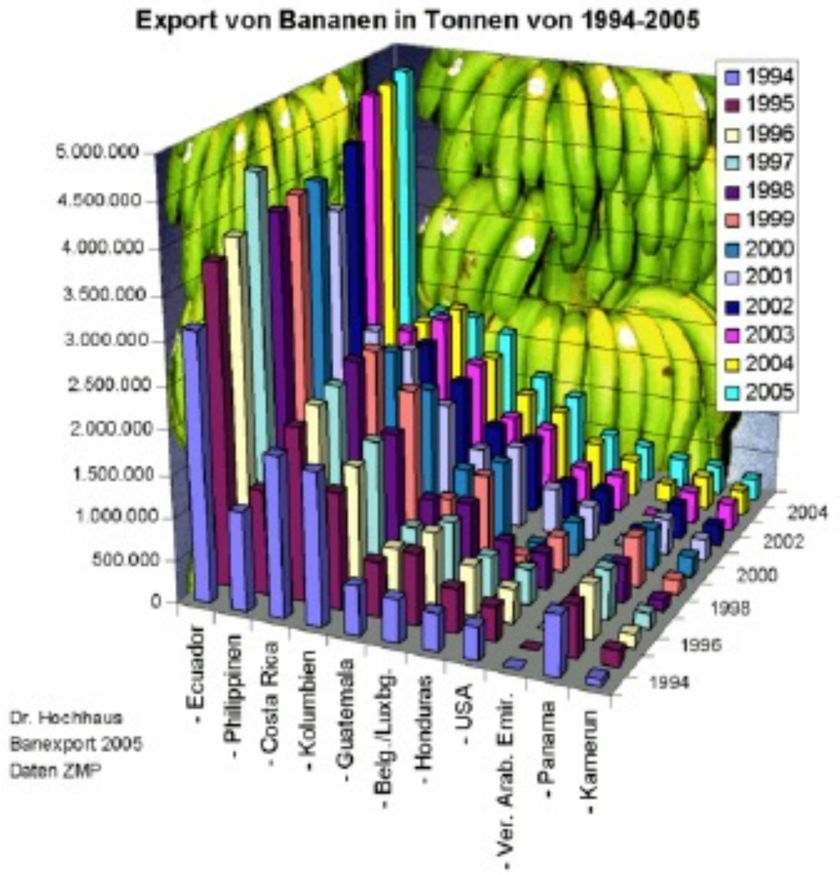
Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design

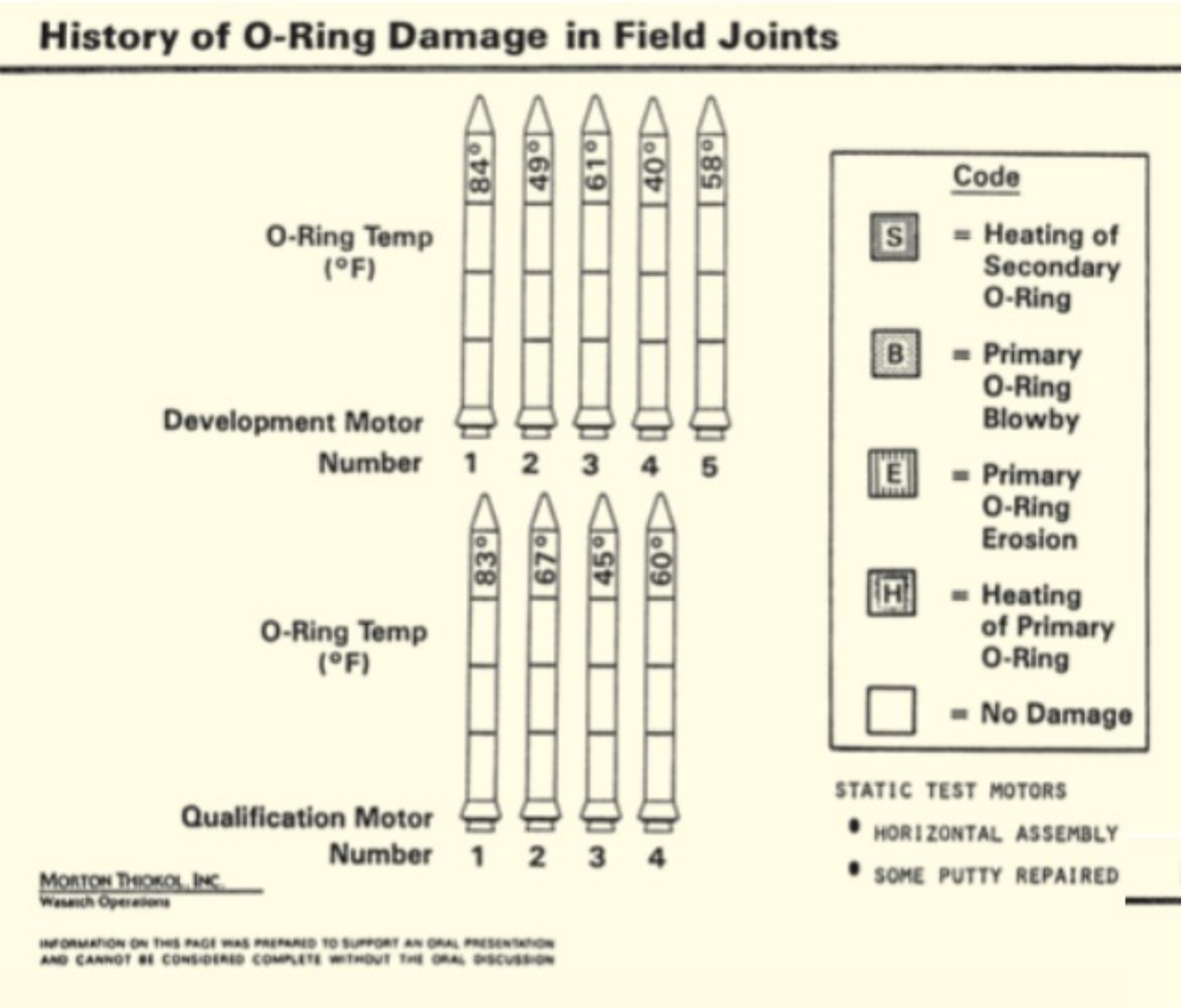
- Communication
  - Motivation
  - Key Considerations

Scope: What's wrong?





Scope: What's wrong?



History of O-Ring Damage in Field Joints

Development Motor Number	1	2	3	4	5
O-Ring Temp (°F)	84°	49°	61°	40°	58°

Qualification Motor Number	1	2	3	4
O-Ring Temp (°F)	83°	67°	45°	60°

Code

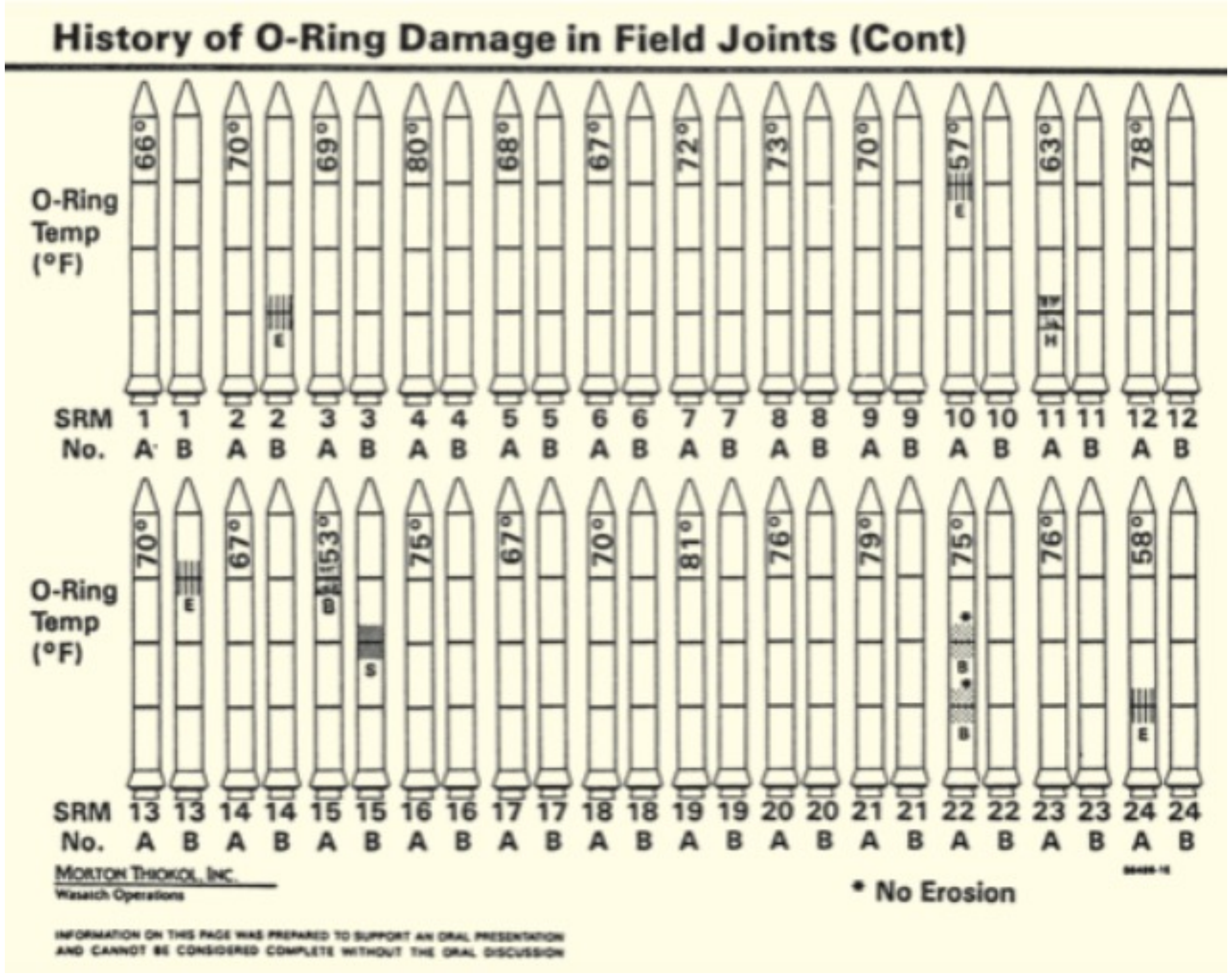
- S = Heating of Secondary O-Ring
- B = Primary O-Ring Blowby
- E = Primary O-Ring Erosion
- H = Heating of Primary O-Ring
- □ = No Damage

STATIC TEST MOTORS

- HORIZONTAL ASSEMBLY
- SOME PUTTY REPAIRED

MORTON THIOKOL, INC.
Wasatch Operations

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION



History of O-Ring Damage in Field Joints (Cont)

SRM No.	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
O-Ring Temp (°F)	66°		70°	E	69°		80°		68°		67°		72°		73°		70°		57° E		63° H		78°	

SRM No.	13	13	14	14	15	15	16	16	17	17	18	18	19	19	20	20	21	21	22	22	23	23	24	24
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
O-Ring Temp (°F)	70°	E	67°		53° B	S	75°		67°		70°		81°		76°		79°		75° • B • B		76°		58° E	

MORTON THIOKOL, INC.
Wasatch Operations

• No Erosion

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Scope

Lesson IV: keep it simple… enough

Scope

"You should have stayed with the soup question. The object of a question is to obtain information that matters only to us"

-- Sean Connery in Finding Forrester (movie)

Scope

- Making plots is effectively providing an answer to an implicit question
- You get to pick the way to express the answer
- Ensure the answer doesn't leave the viewer with uncertainty as to **what it's answering** or the **completeness of the answer**
- A good plot should invoke and inspire new questions

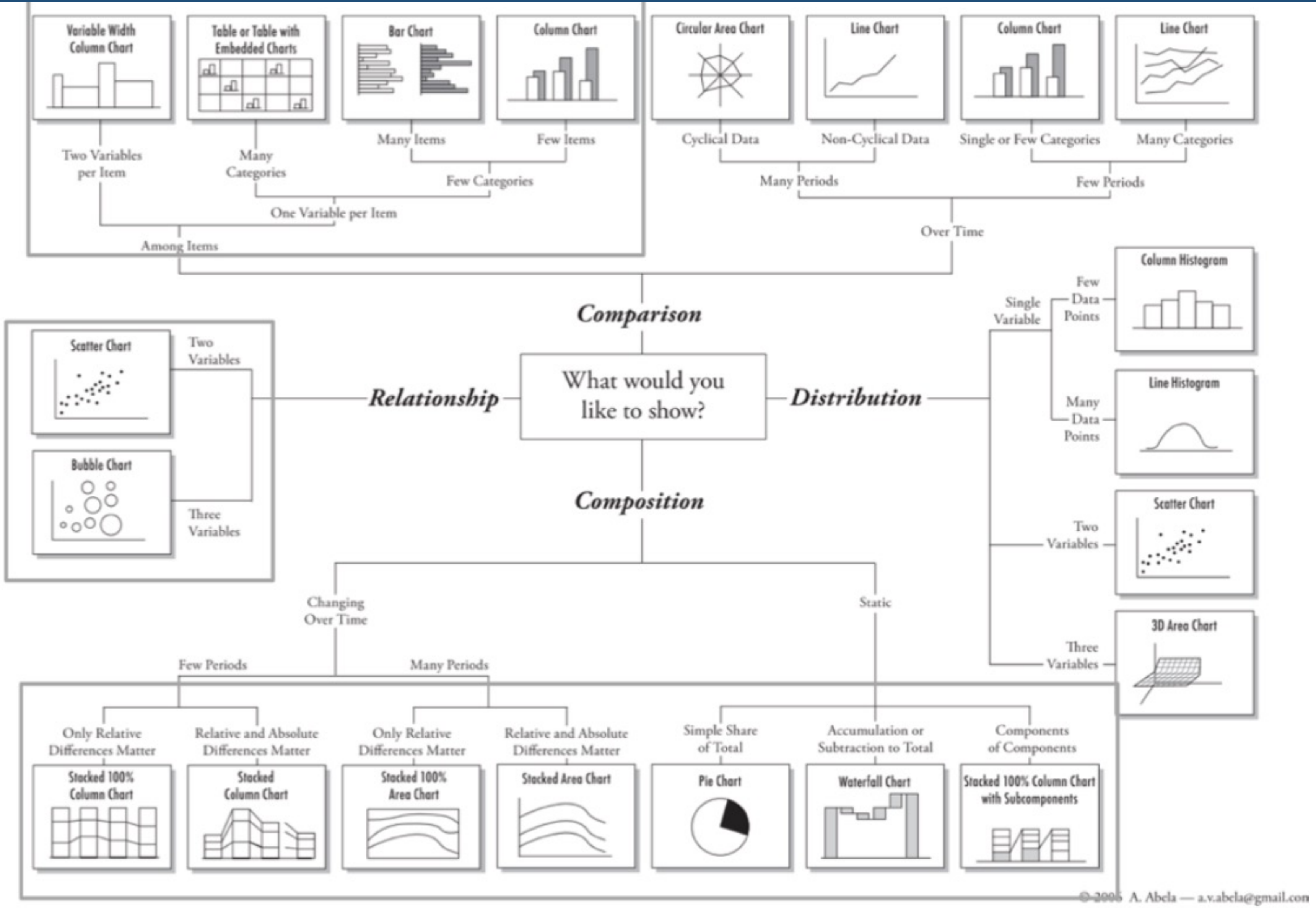
Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

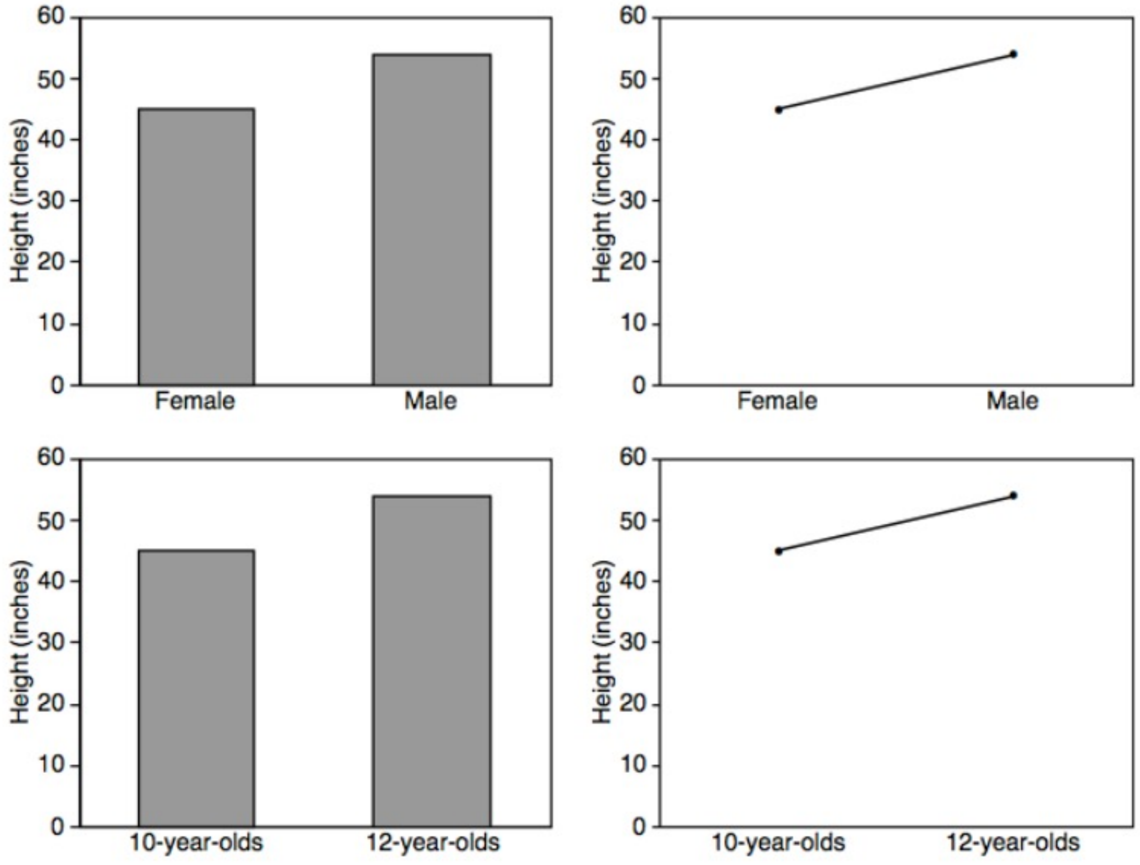
Displays



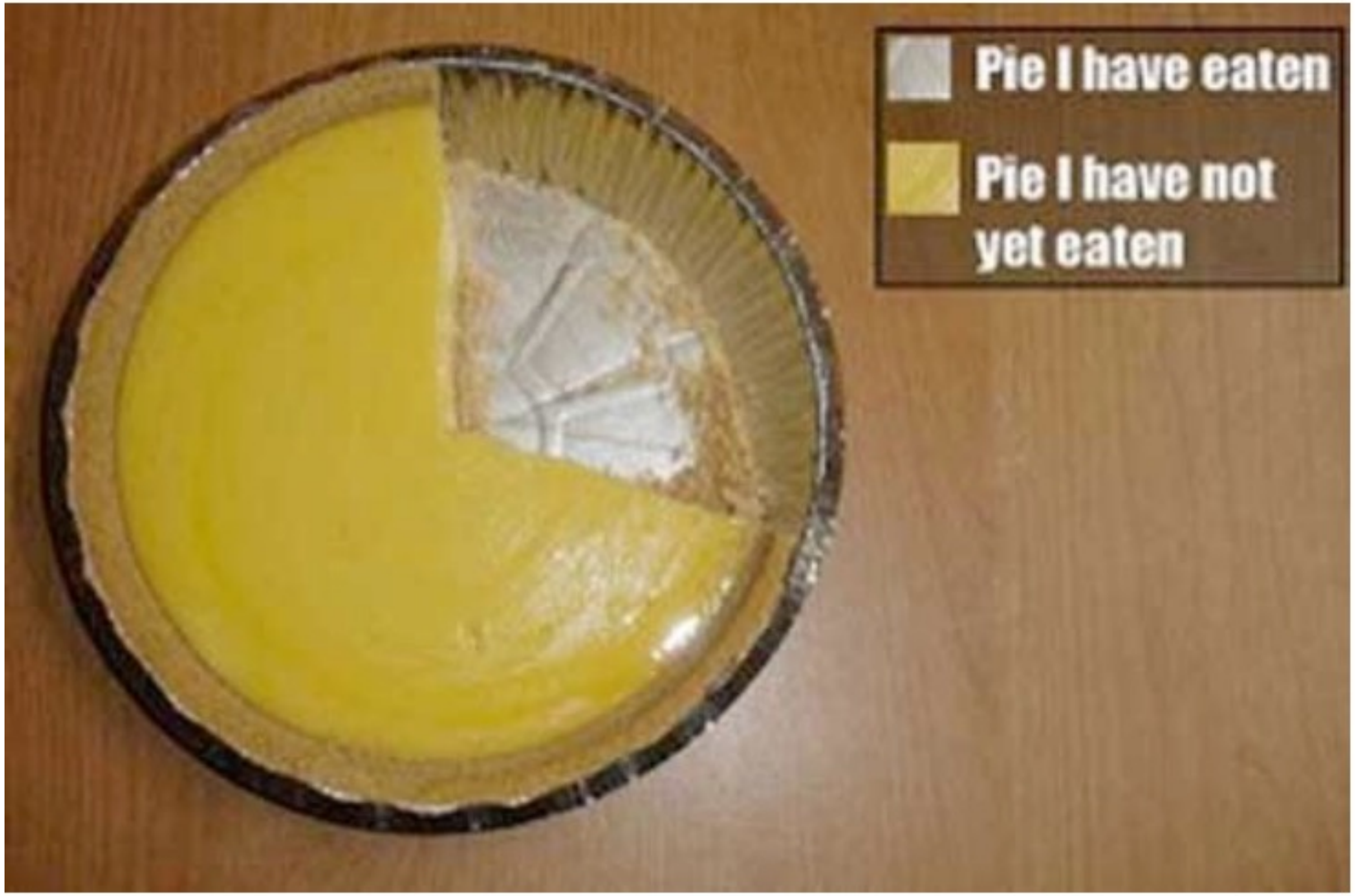
http://extremepresentation.typepad.com/blog/files/choosing_a_good_chart.pdf

Displays: comparisons

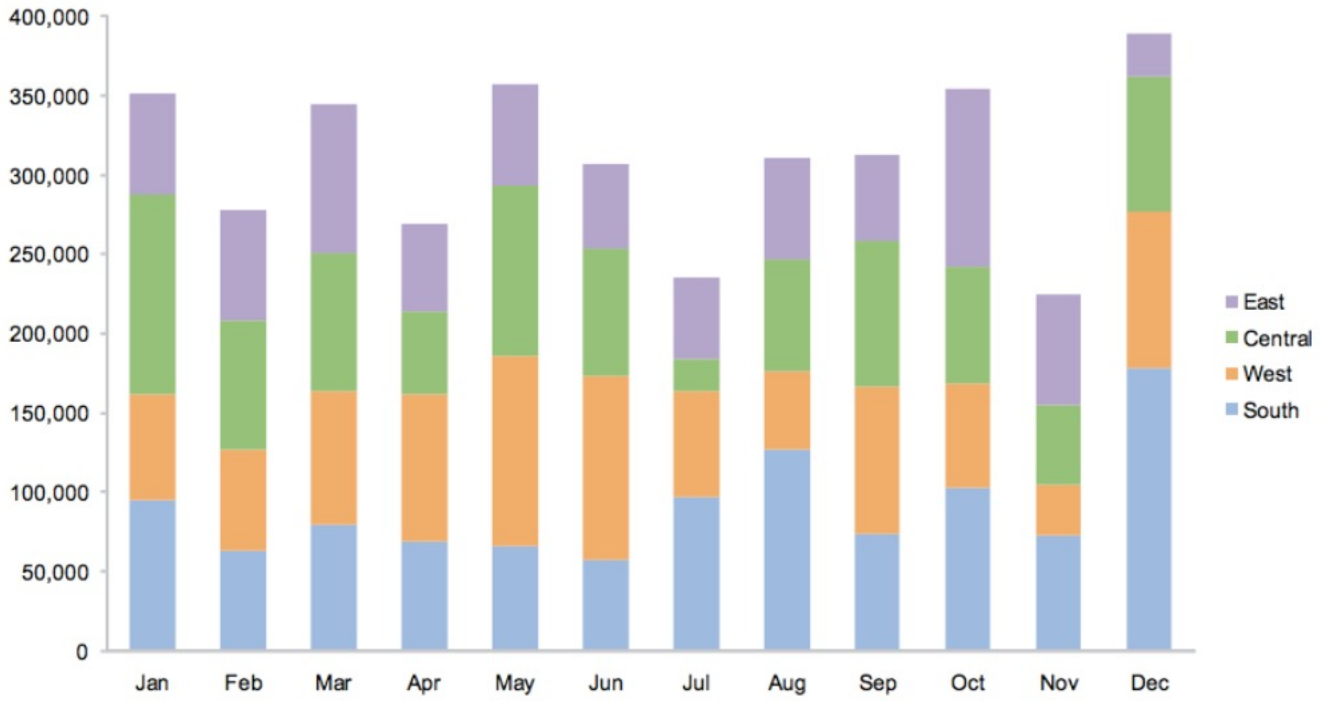
Bars vs. Lines



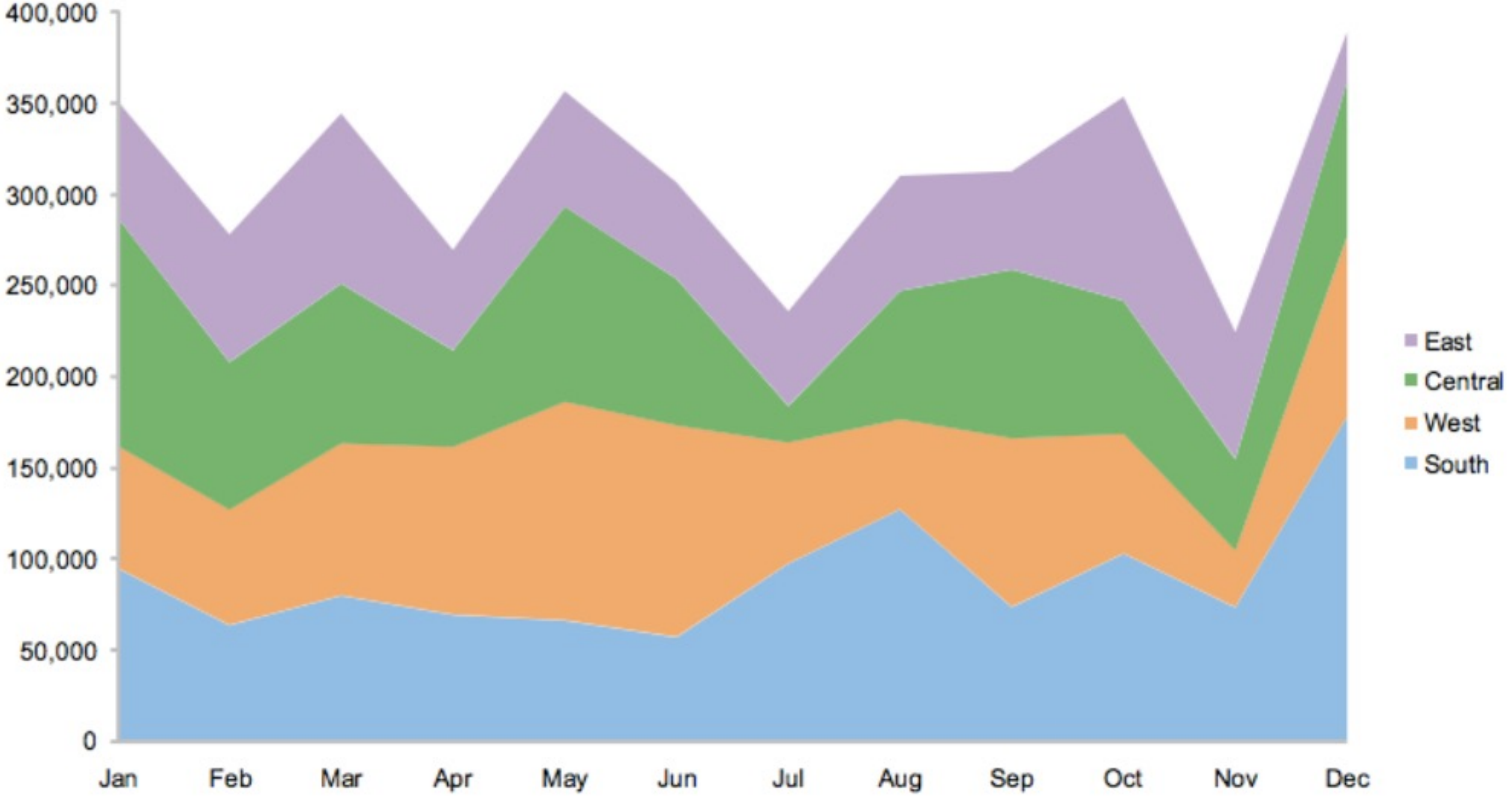
Displays: proportions



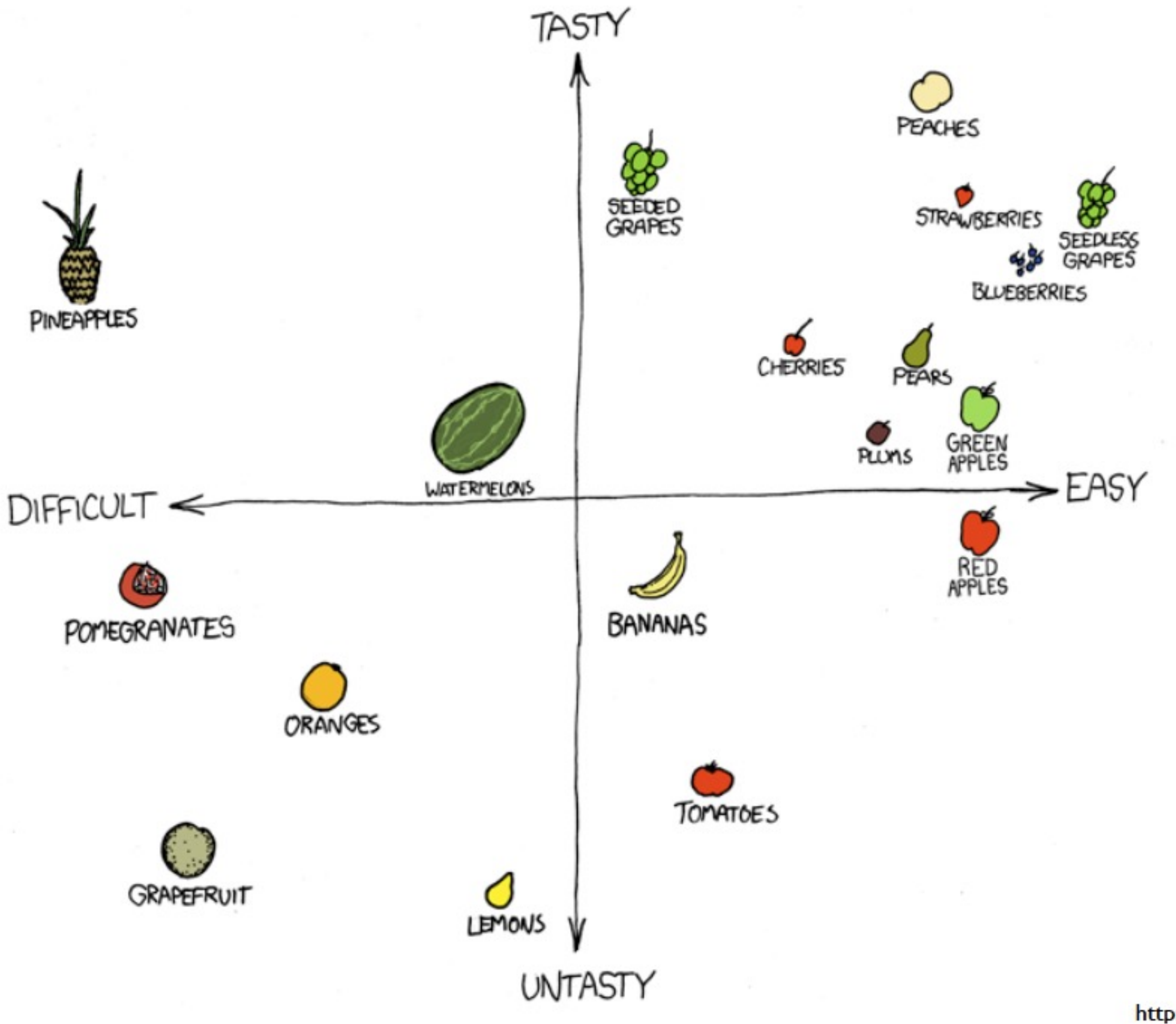
Displays: proportions



Displays: proportions

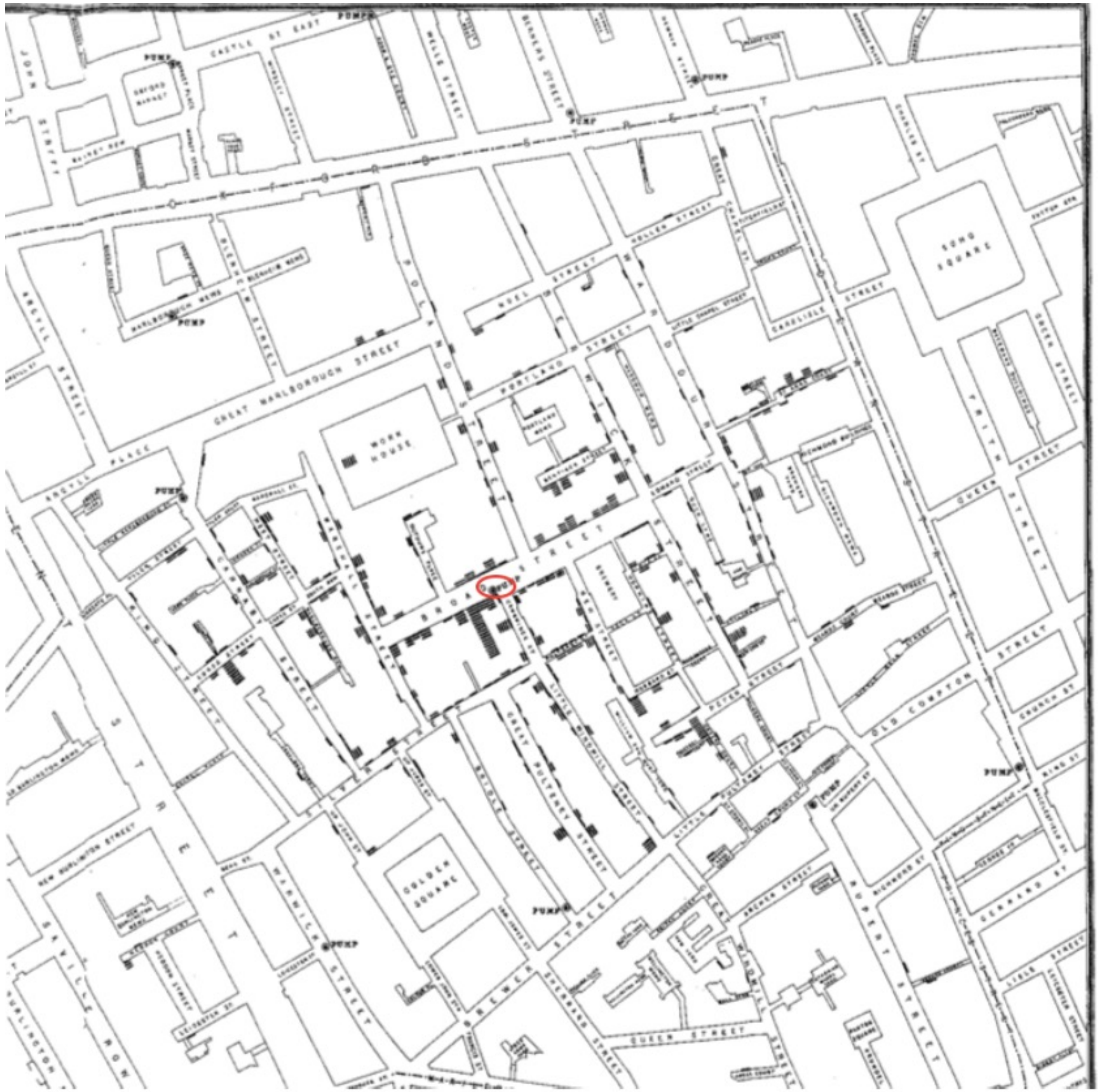


Displays: scatter plot



http://xkcd.com/388/

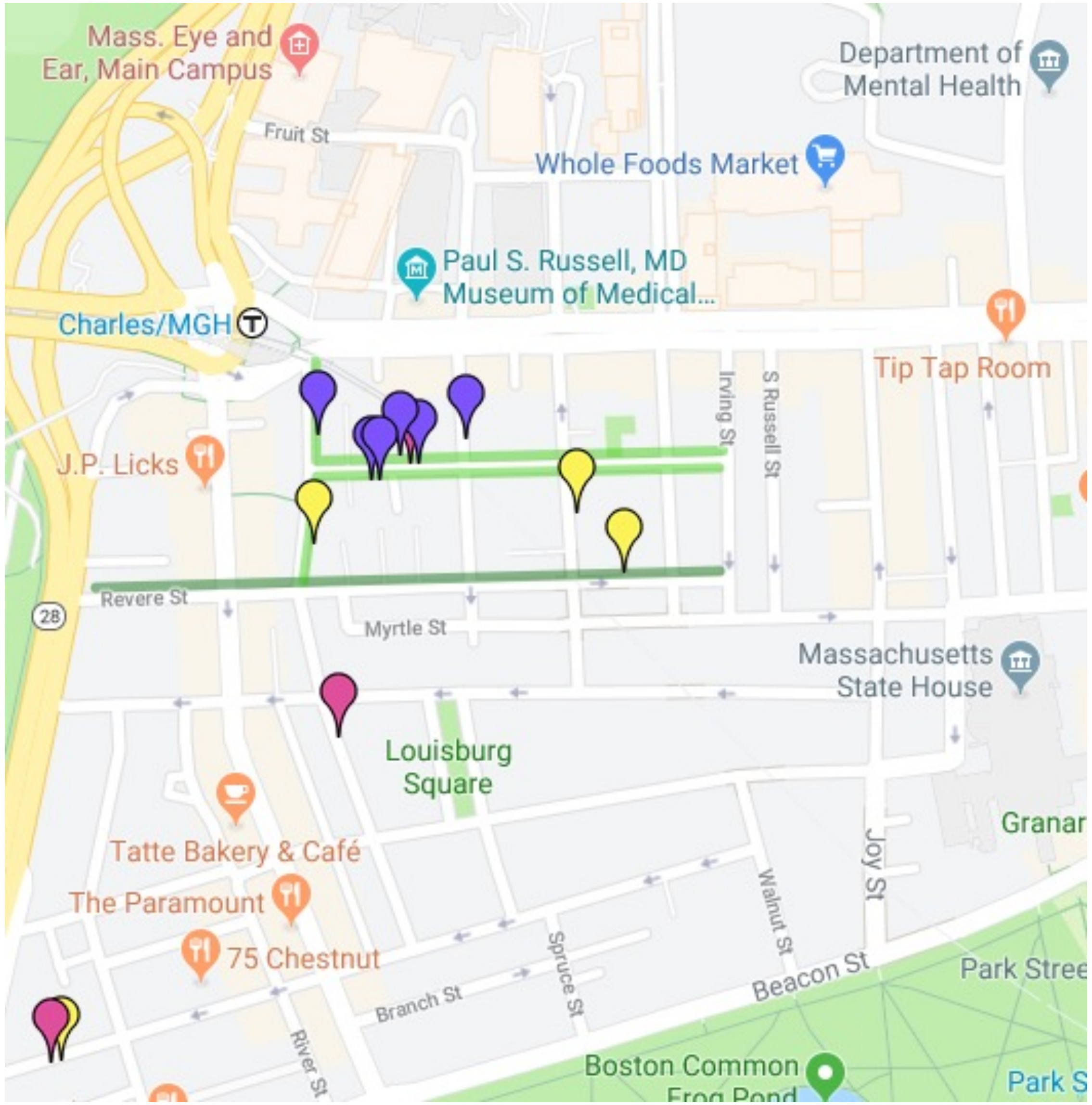
Displays: proportions



London Cholera Epidemic

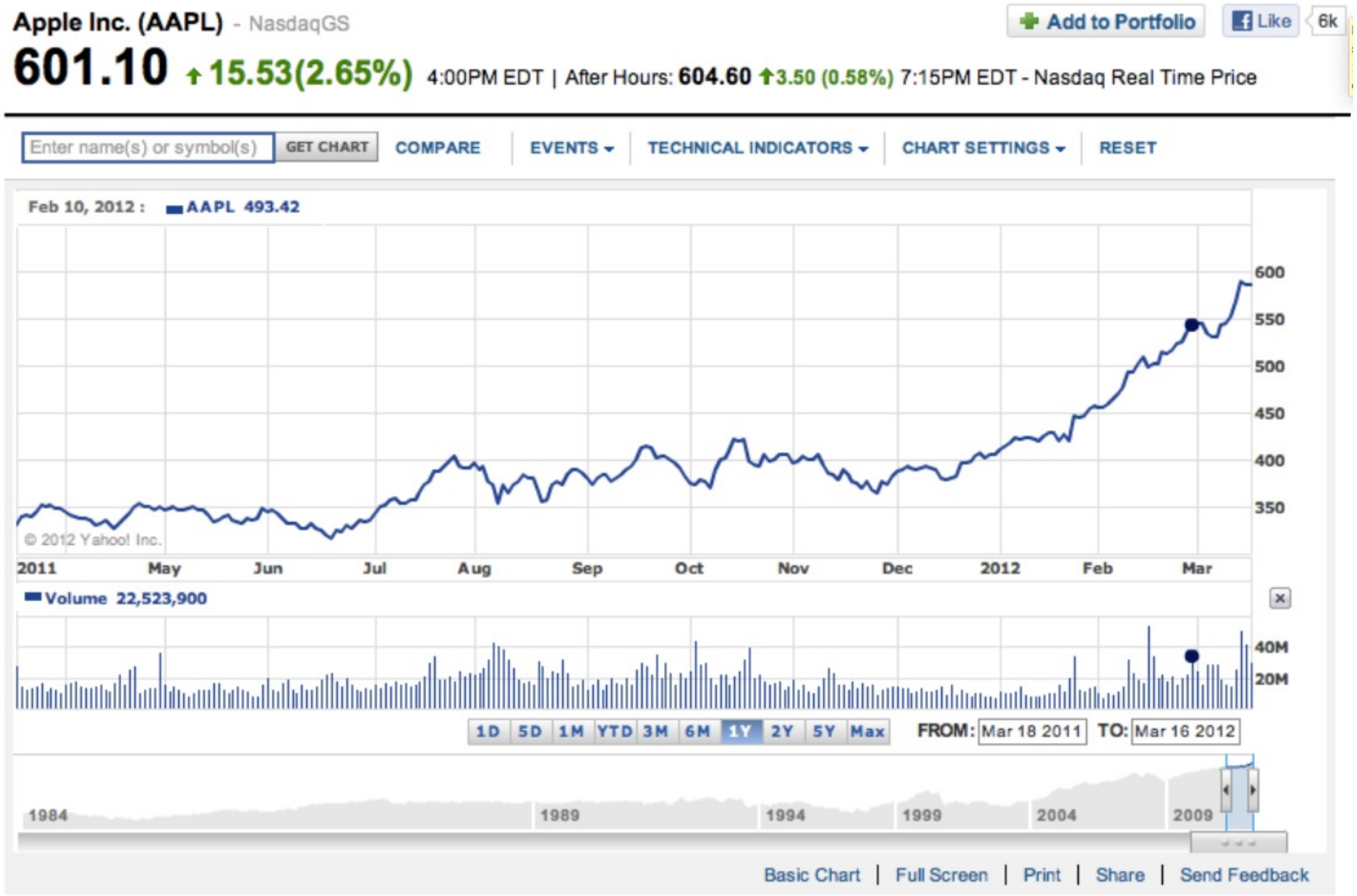
-- Edward Tufte,
Visual and Statistical Thinking

Displays: proportions



- friday @ 10:30am
- sunday @ 5pm
- saturday @ 2pm
- sunday @ 2pm
- friday @ 11:30am
- 1st and 3rd wed
- 2nd and 4th wed
- 2nd and 4th wed
- saturday @ 3pm
- sunday @ 4:30pm
- friday @ 7:30pm
- saturday @ 8pm
- friday @ 8pm
- saturday @ 6pm
- saturday @ 7
- saturday @ 6

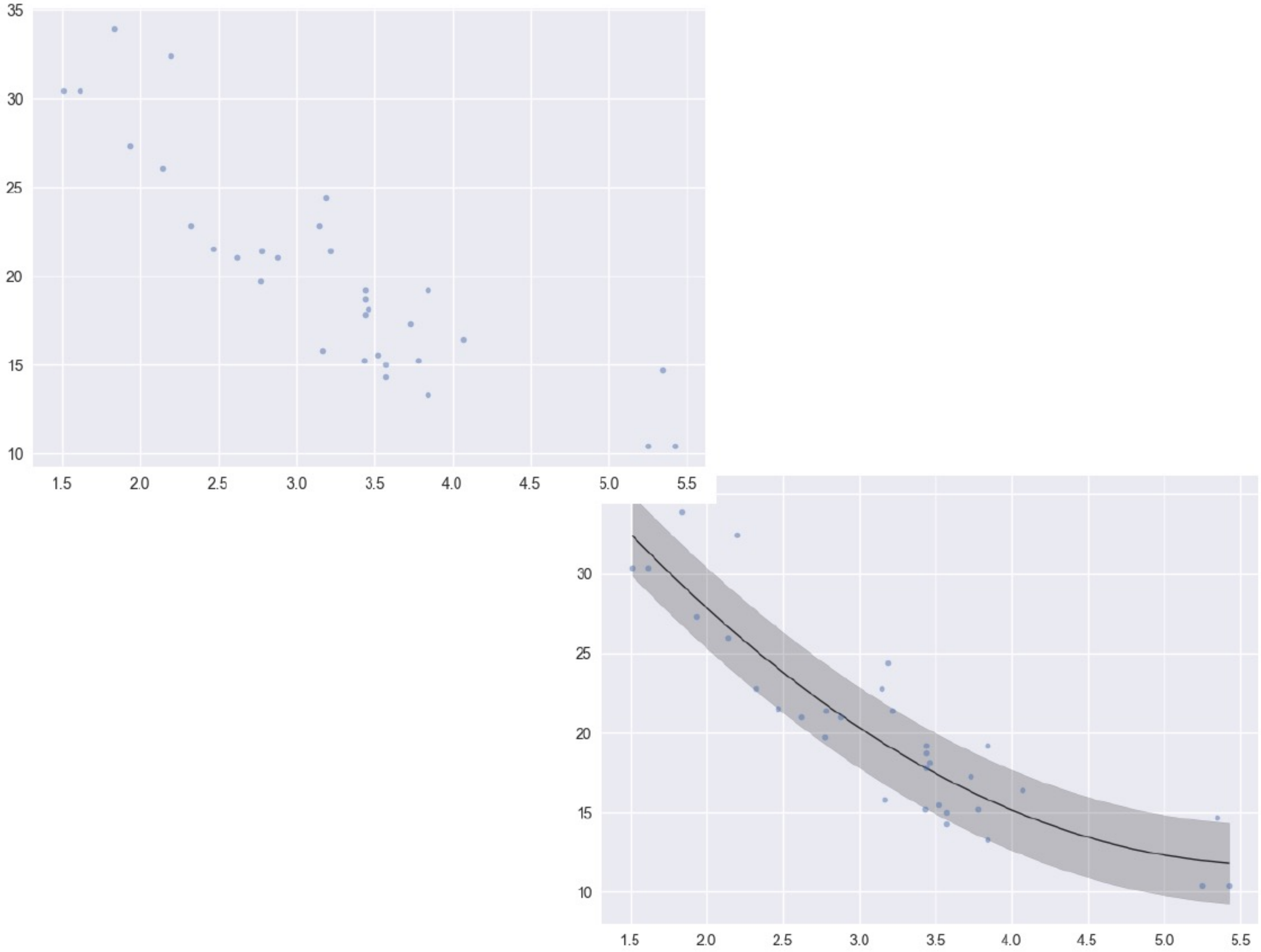
Displays: trends



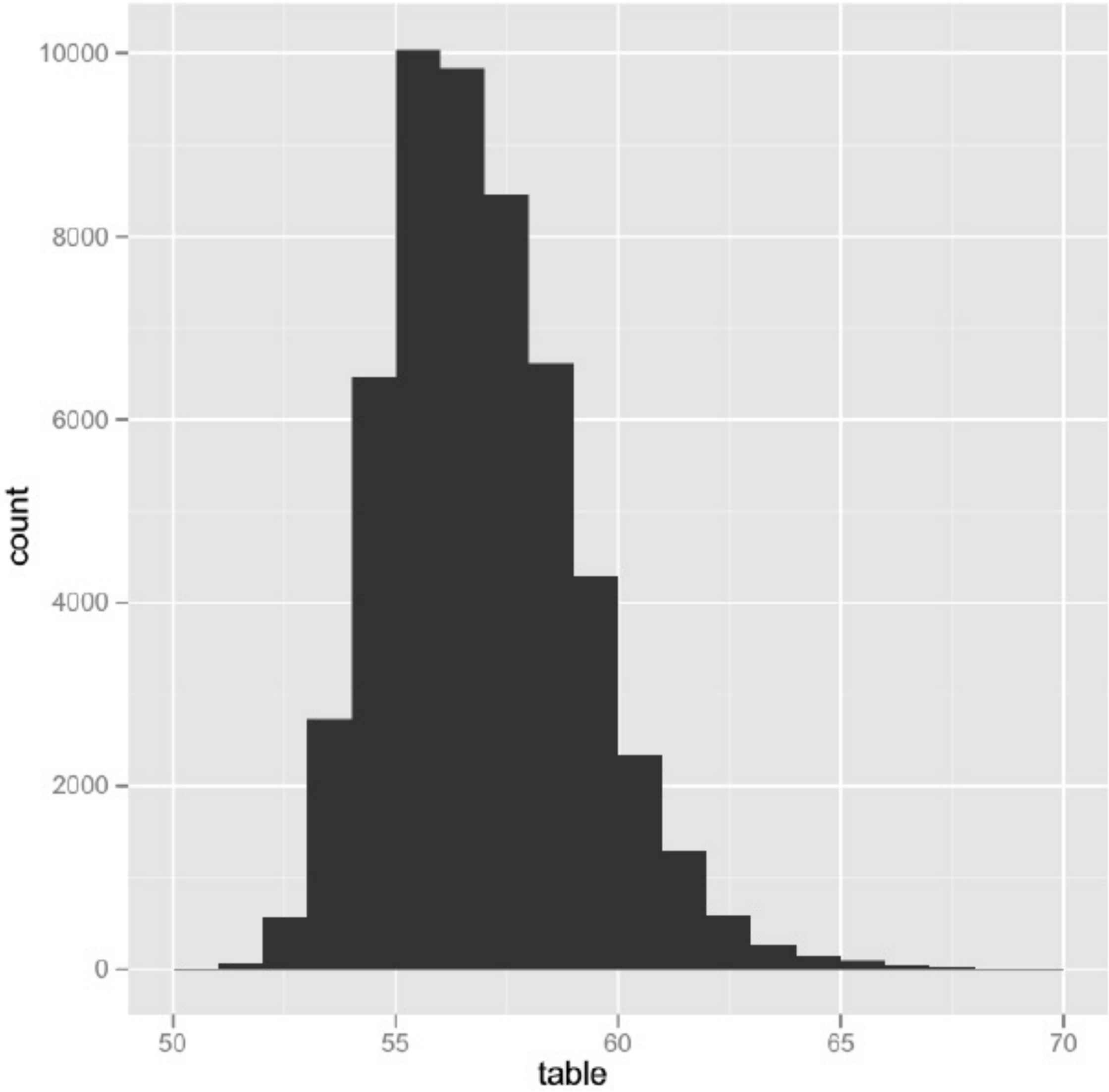
Apple Inc. (AAPL) - NasdaqGS

601.10 ↑15.53(2.65%) 4:00PM EDT | After Hours: 604.60 ↑3.50 (0.58%) 7:15PM EDT - Nasdaq Real Time Price

Displays: trends

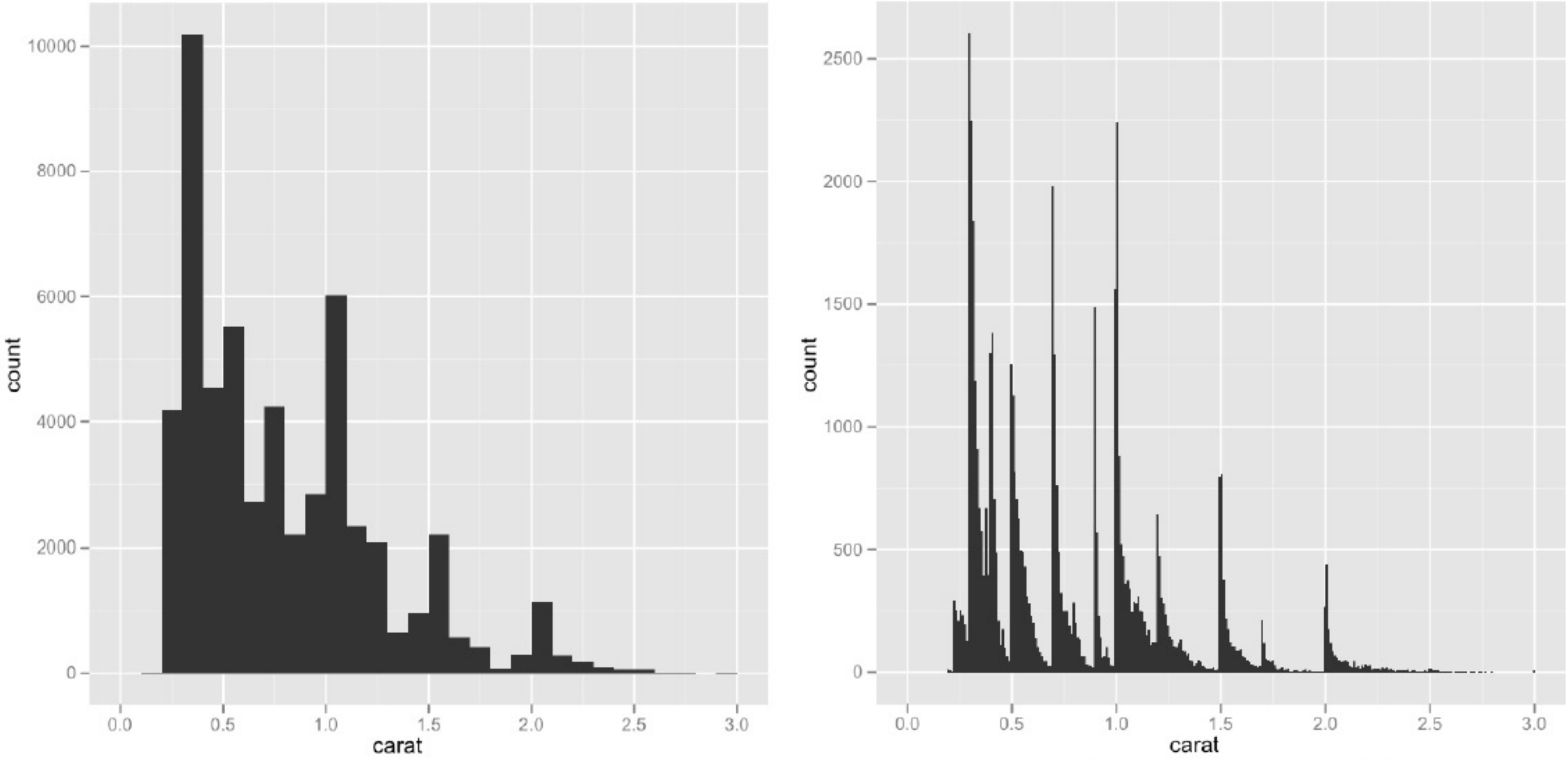


Displays: distributions



Displays: distributions

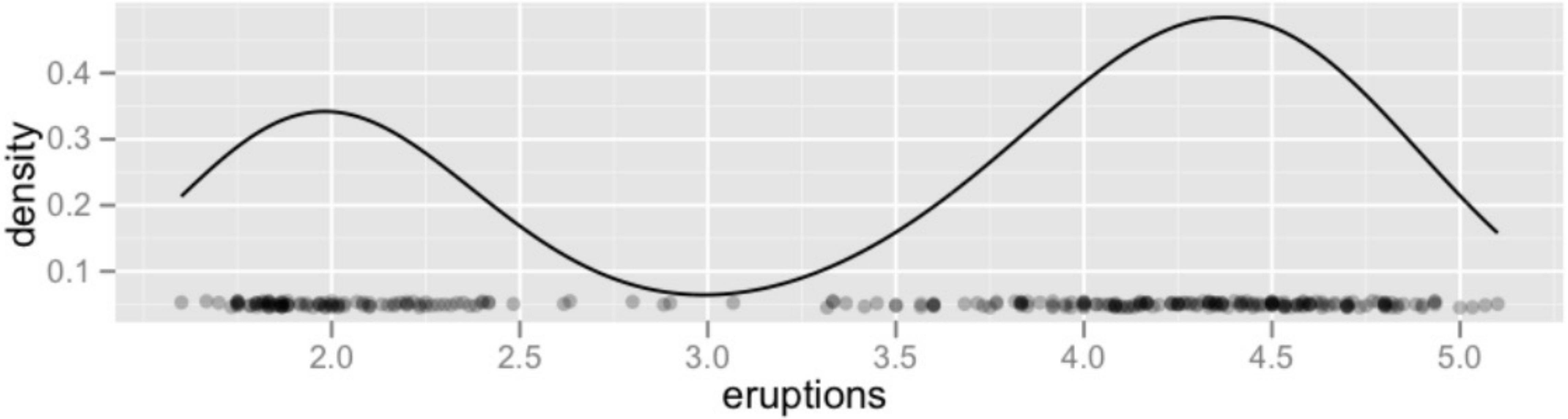
Bin Width



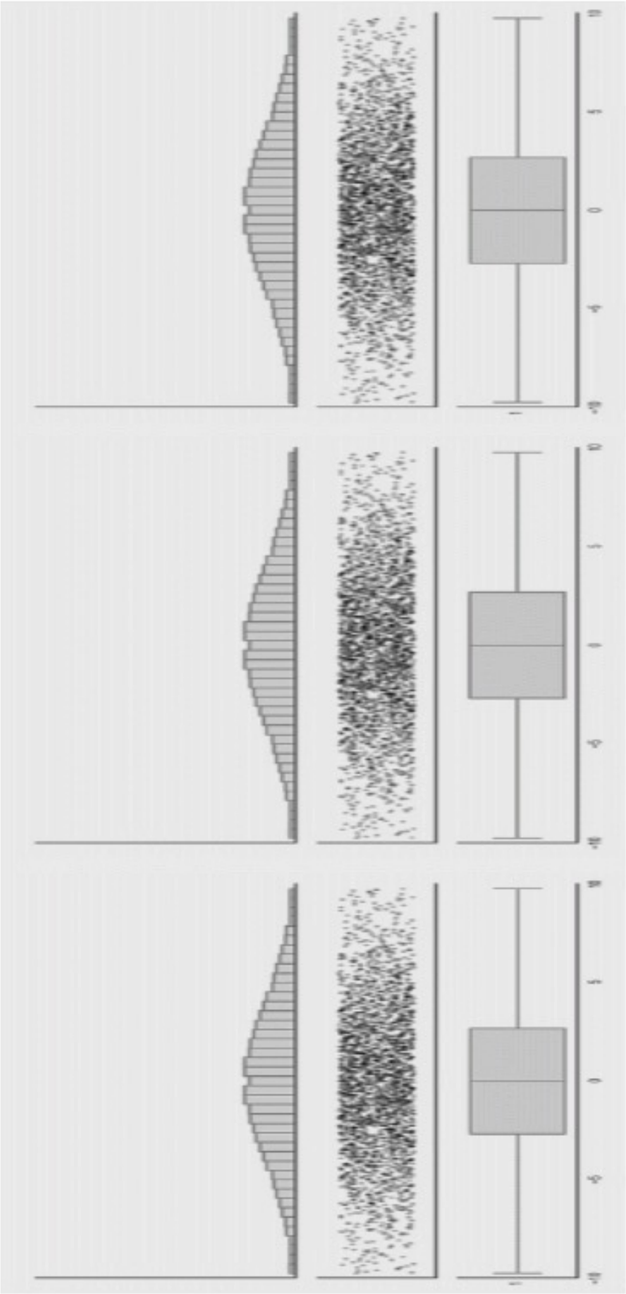
binwidth = 0.1

binwidth = 0.01

Displays: density plots

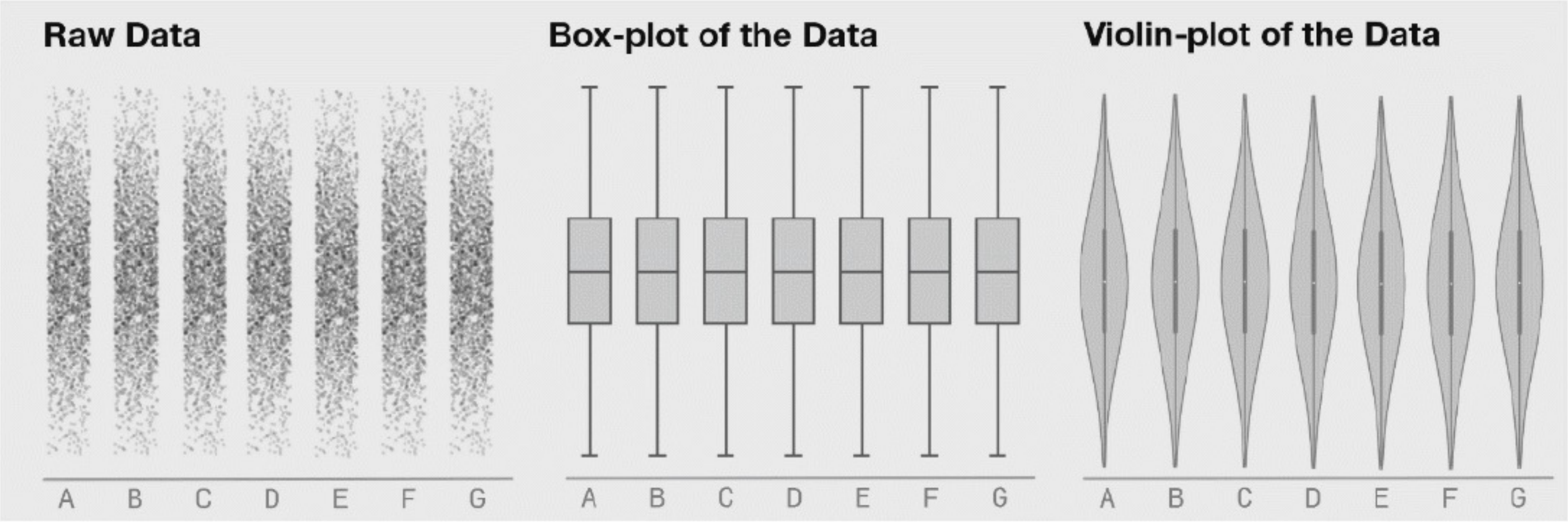


Displays: distributions

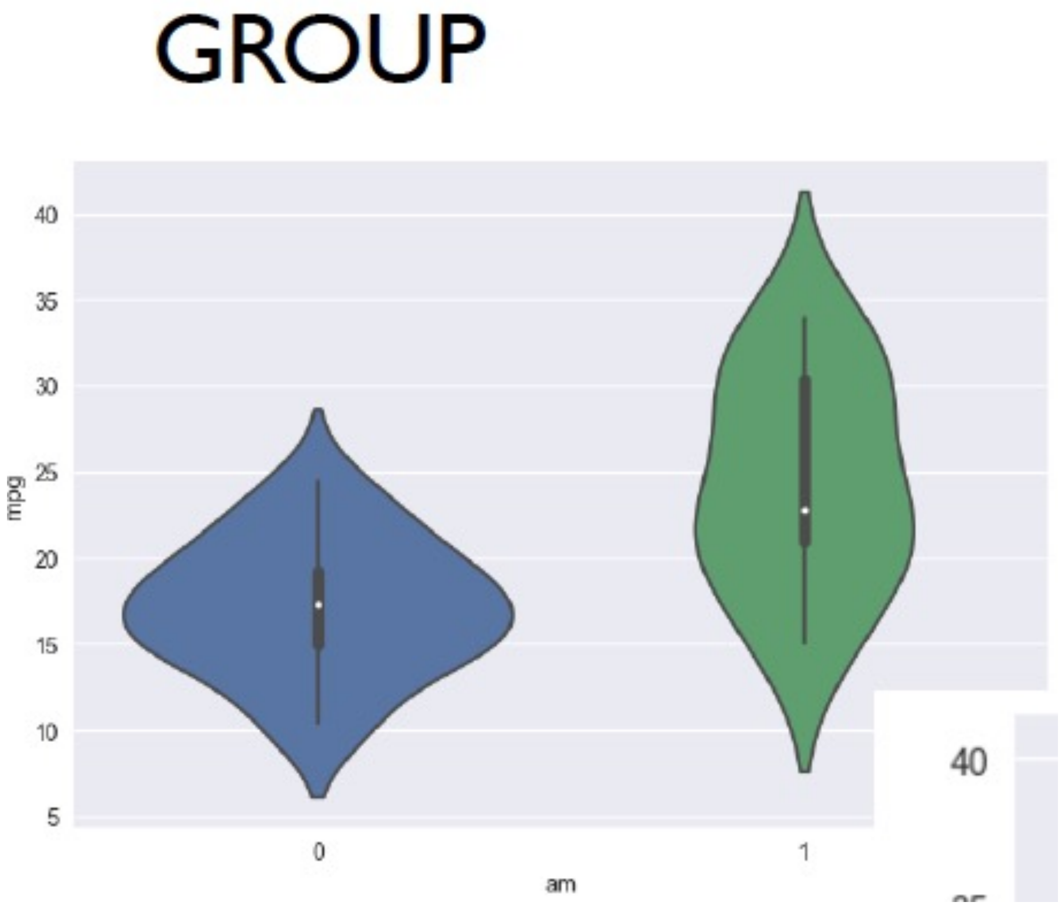


https://www.autodeskresearch.com/publications/samestats

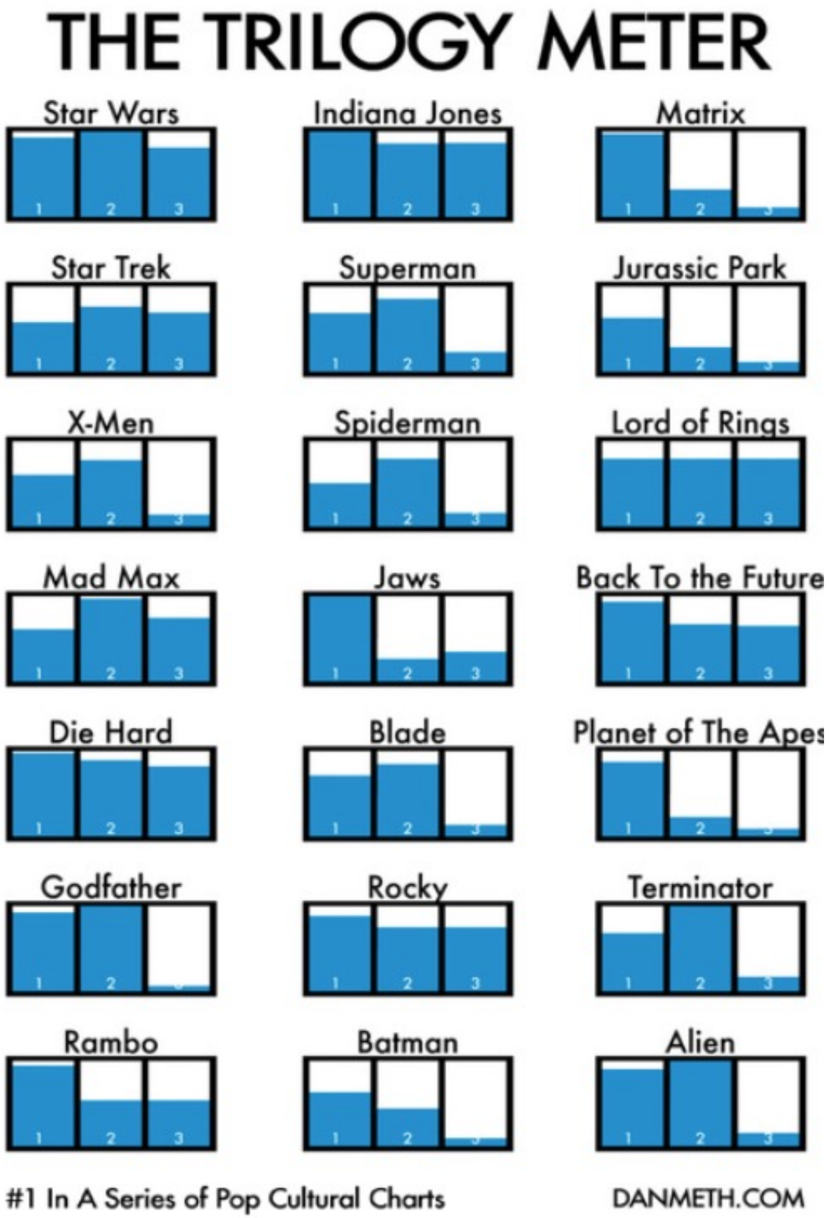
Displays: distributions



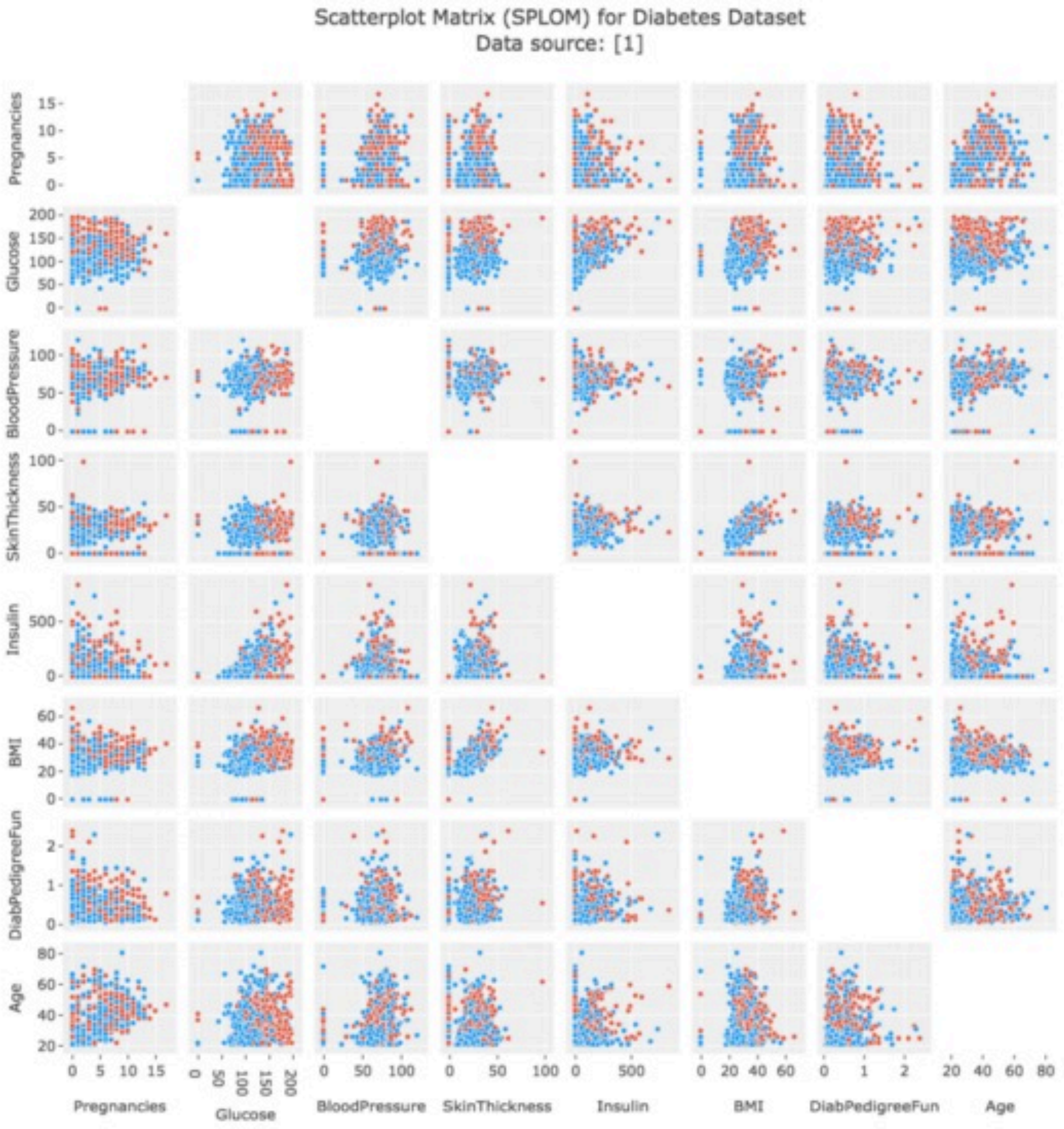
Displays: distributions



Displays: small multiples

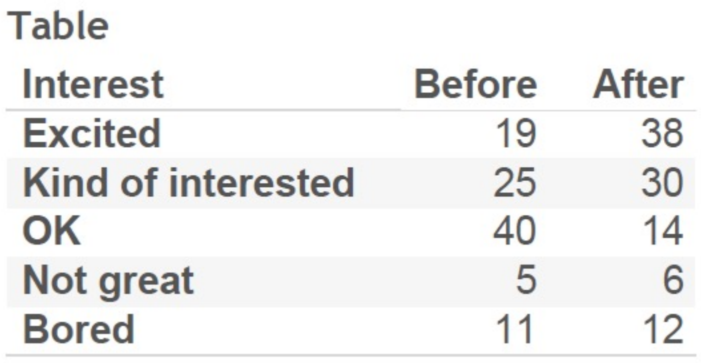


Displays:
scatter plot
matrix



Break-out Discussion!

How do you feel about doing science?

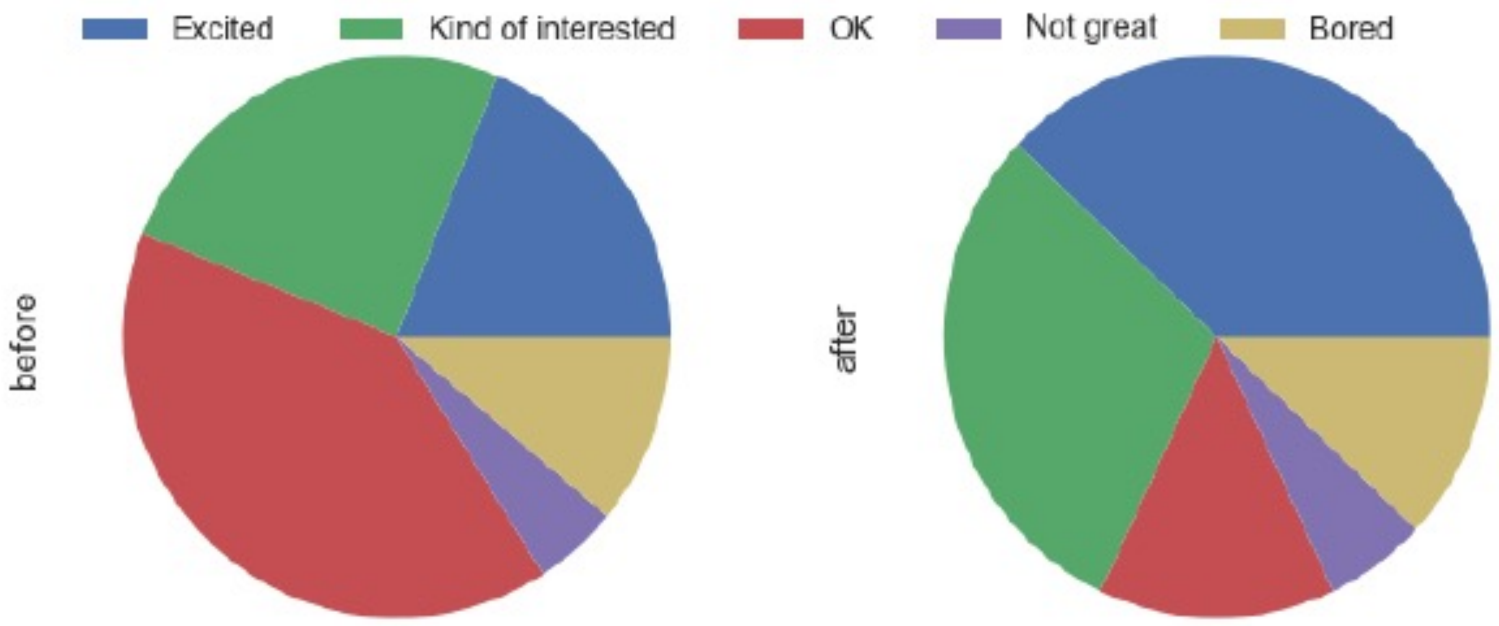


Table

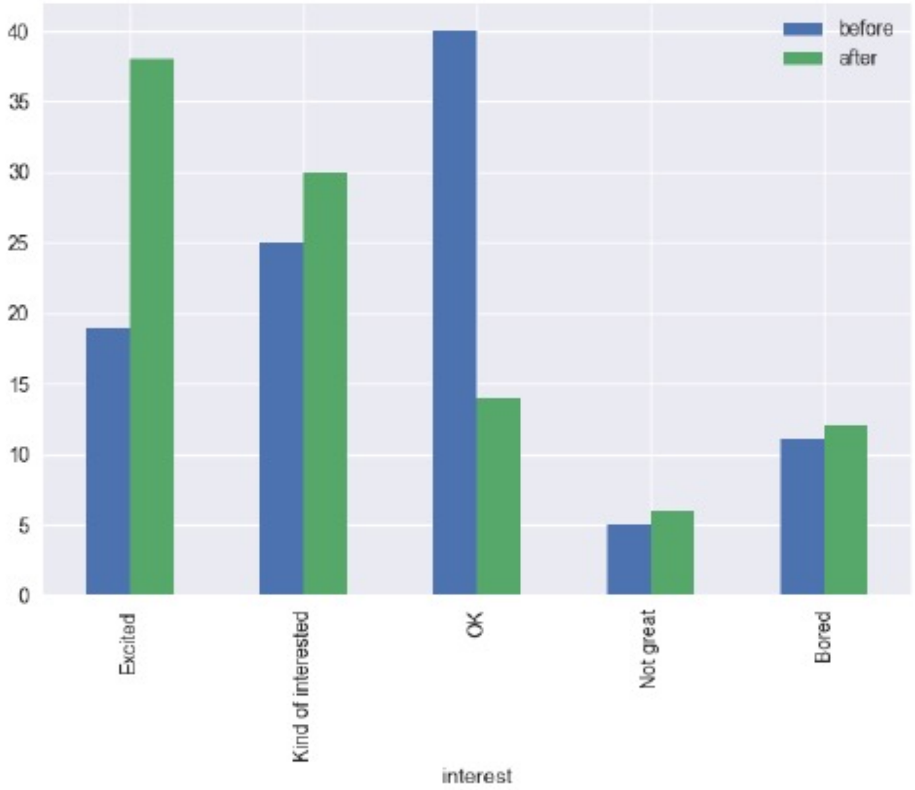
Interest	Before	After
Excited	19	38
Kind of interested	25	30
OK	40	14
Not great	5	6
Bored	11	12

Data courtesy of Cole Nussbaumer

Displays: exercise feedback

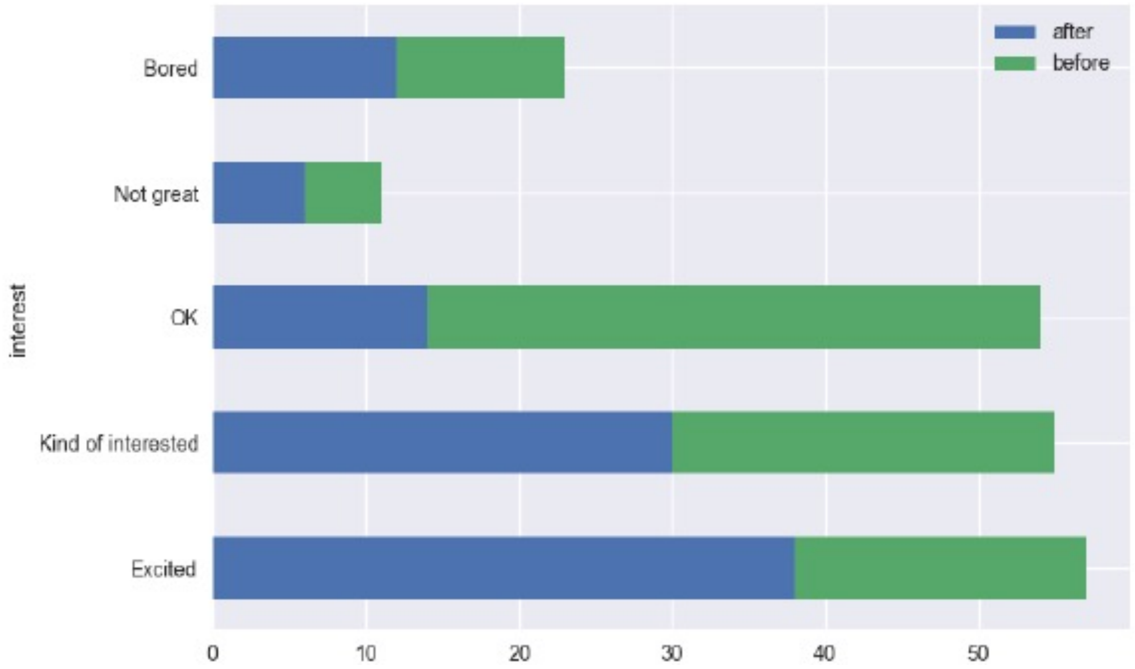


Displays: exercise feedback

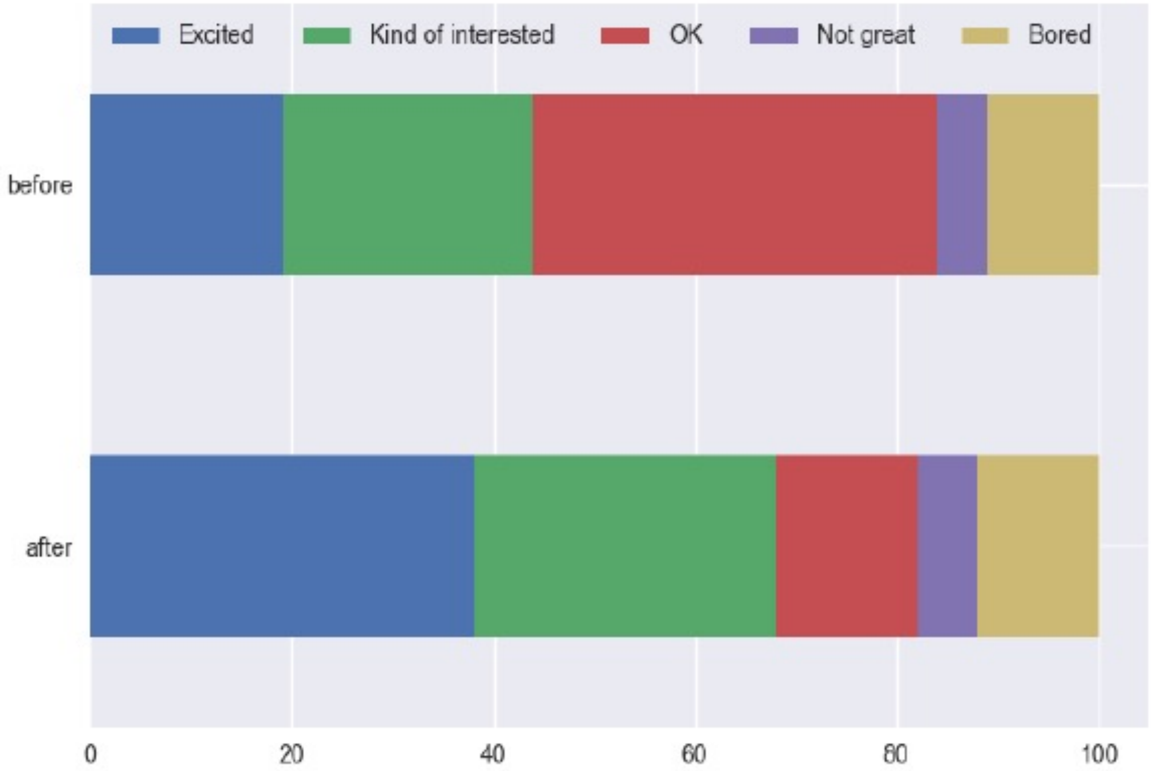


Displays: exercise feedback

Stacked bar, not very useful



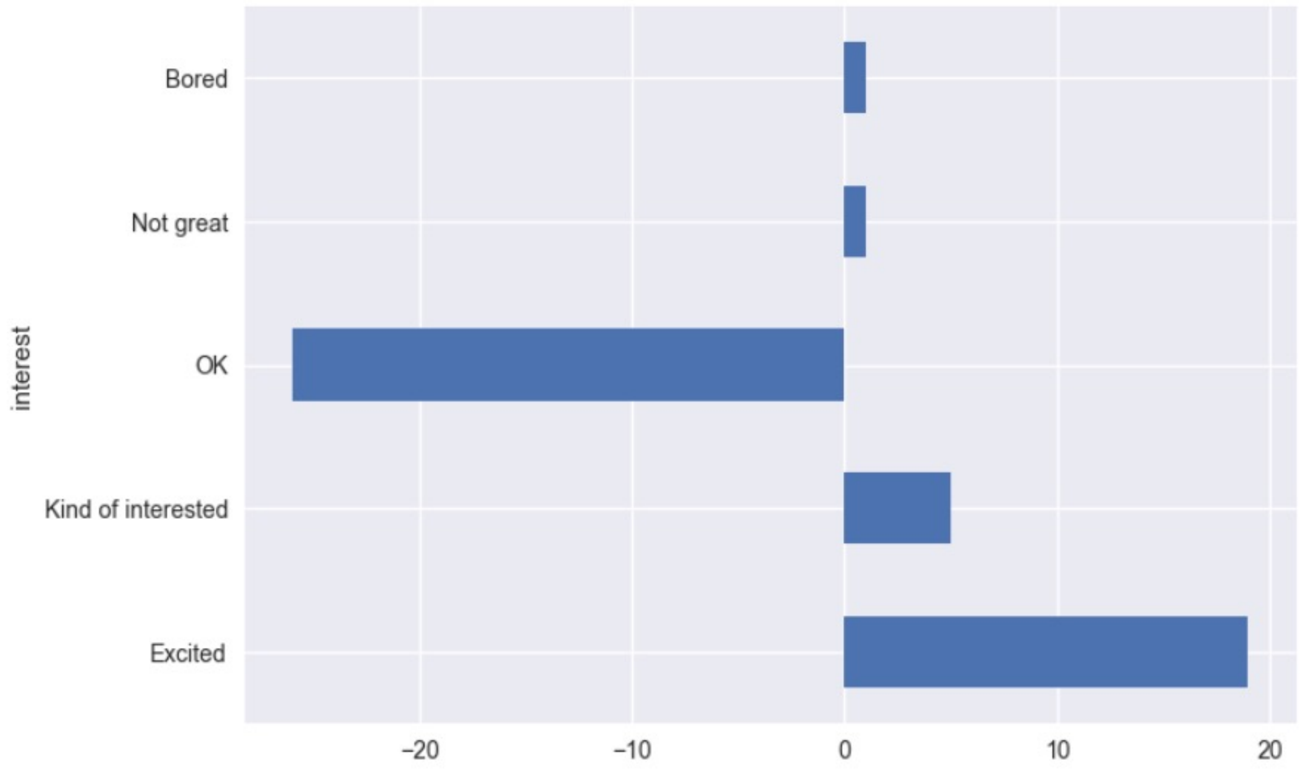
Displays: exercise feedback



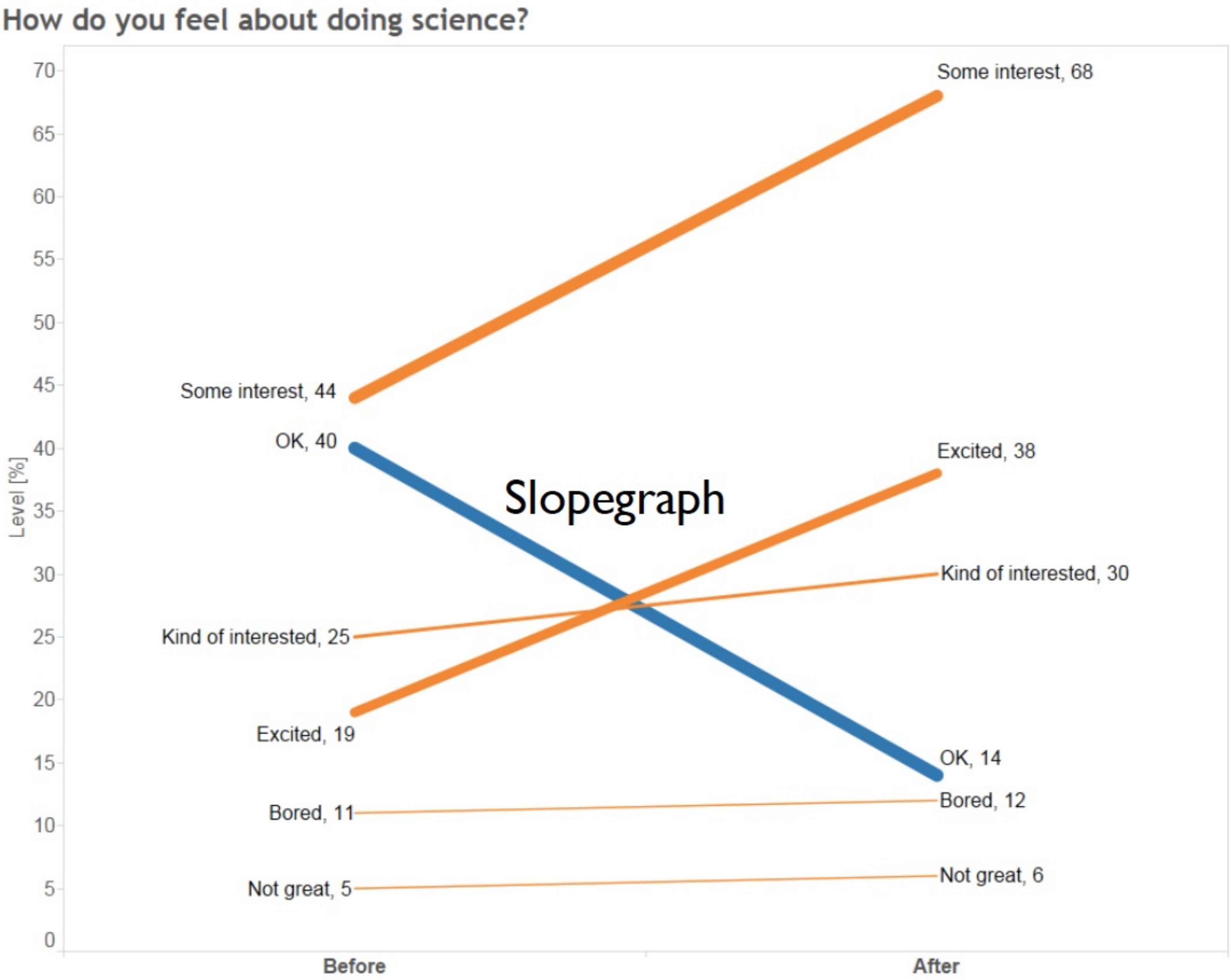
Data Transposed Bar Chart

Displays: exercise feedback

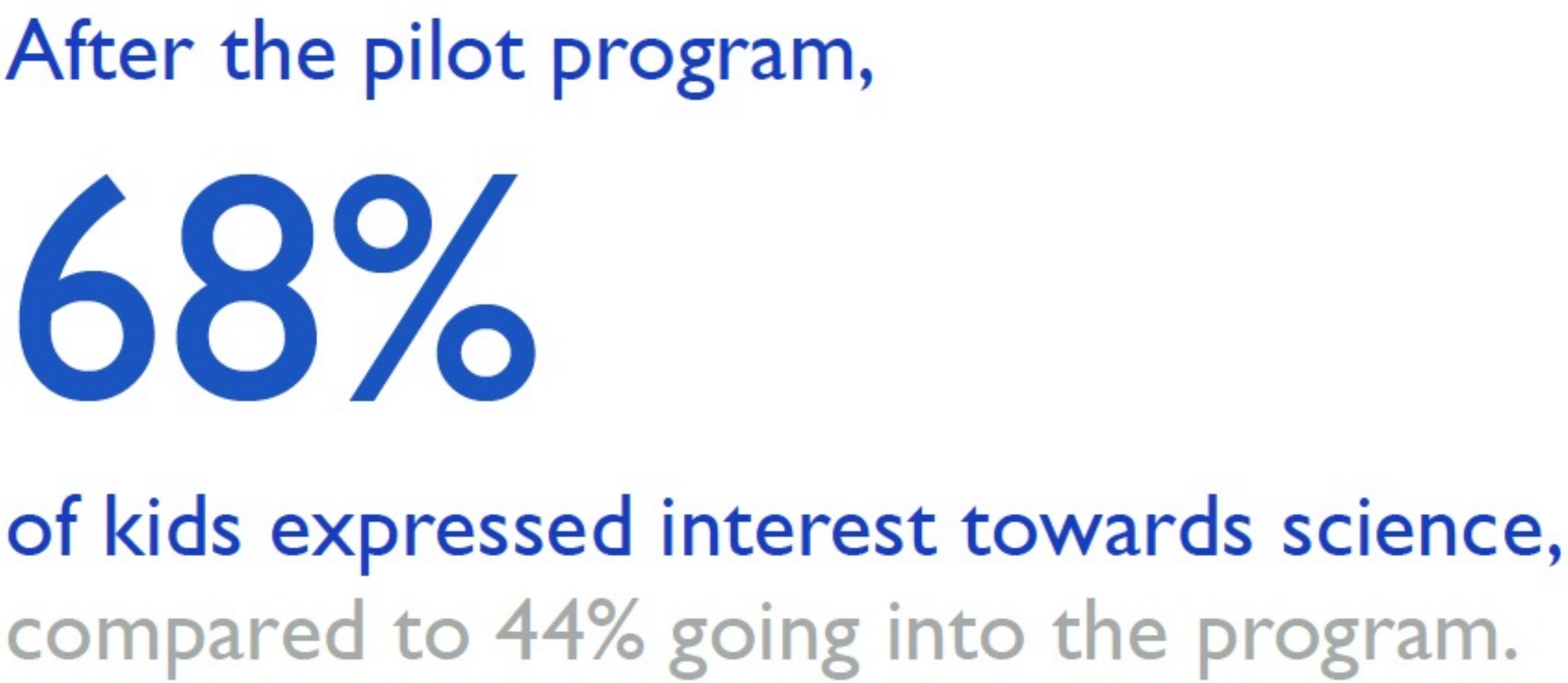
Difference Bar Chart



Displays: exercise feedback



Displays: exercise feedback



After the pilot program,

68%

of kids expressed interest towards science,

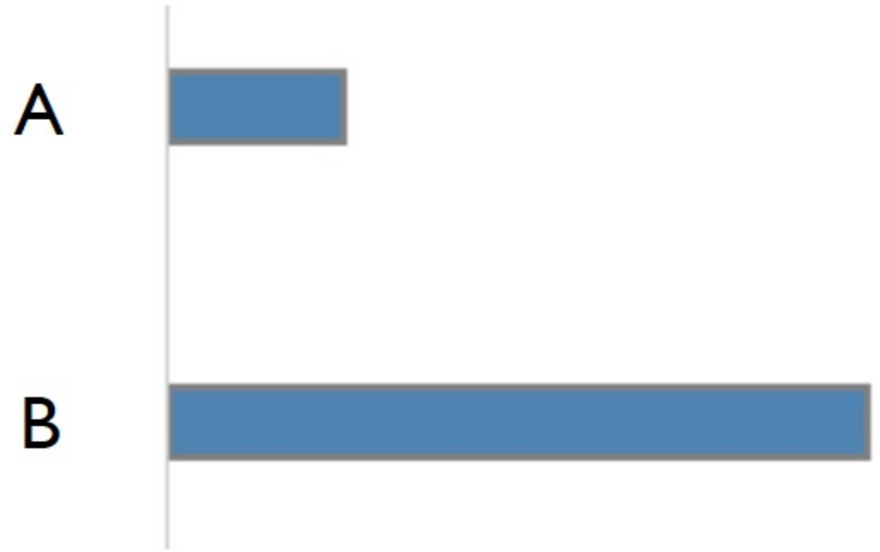
compared to 44% going into the program.

Time for a game

🎉

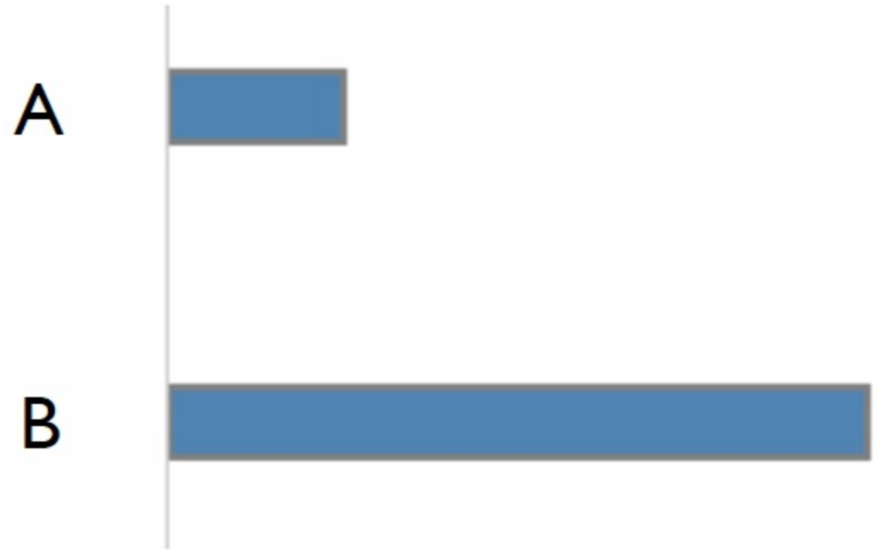
Displays: perceptual effectiveness

How much longer?



Displays: perceptual effectiveness

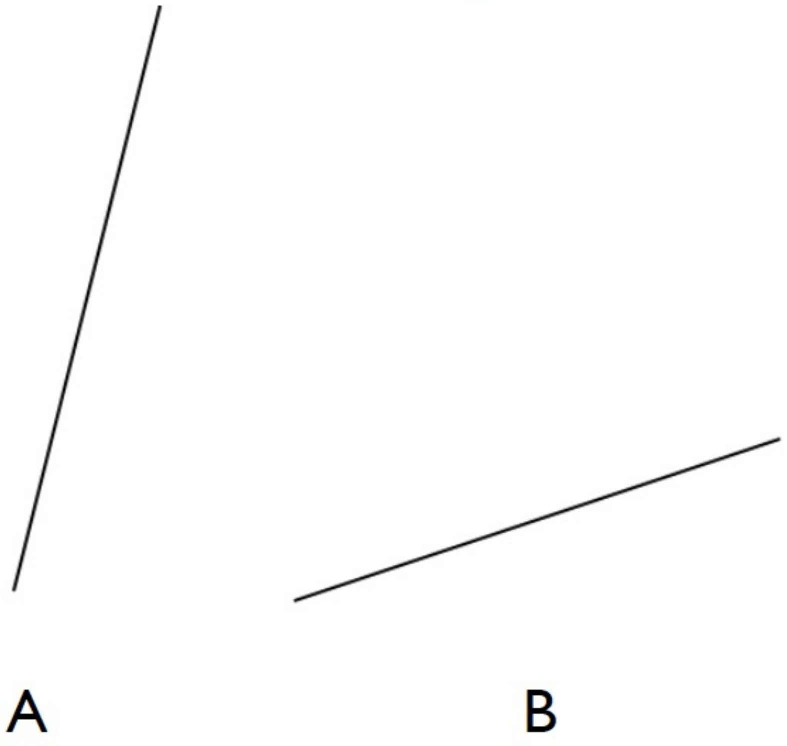
How much longer?



4x

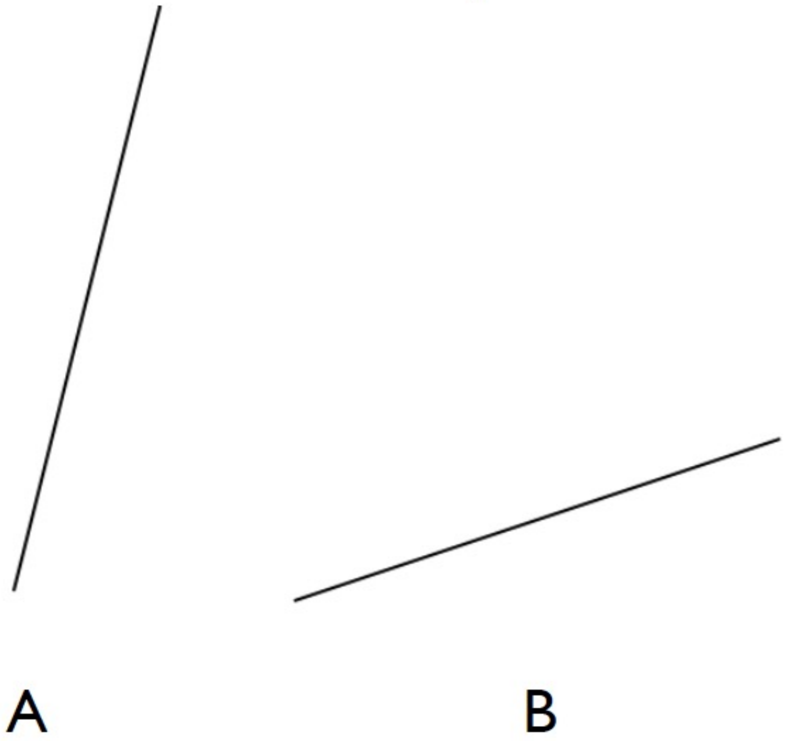
Displays: perceptual effectiveness

How much steeper slope?



Displays: perceptual effectiveness

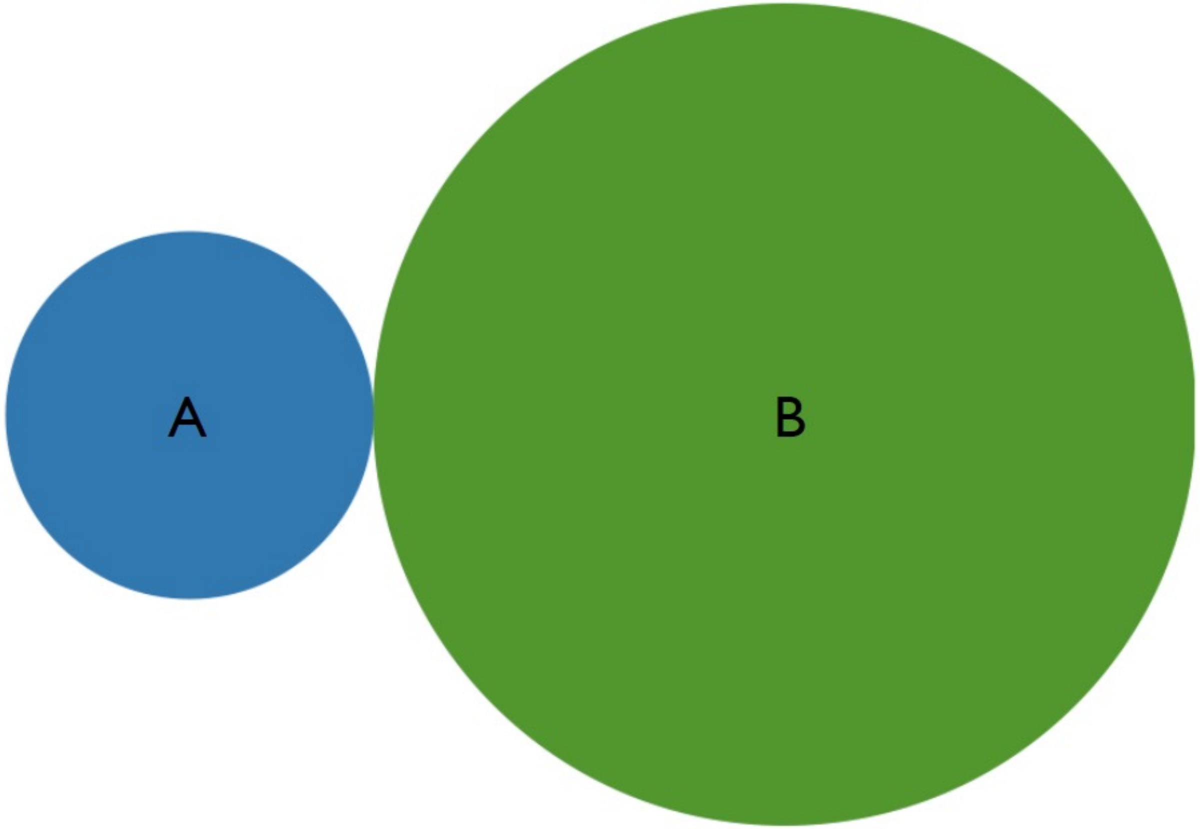
How much steeper slope?



4x

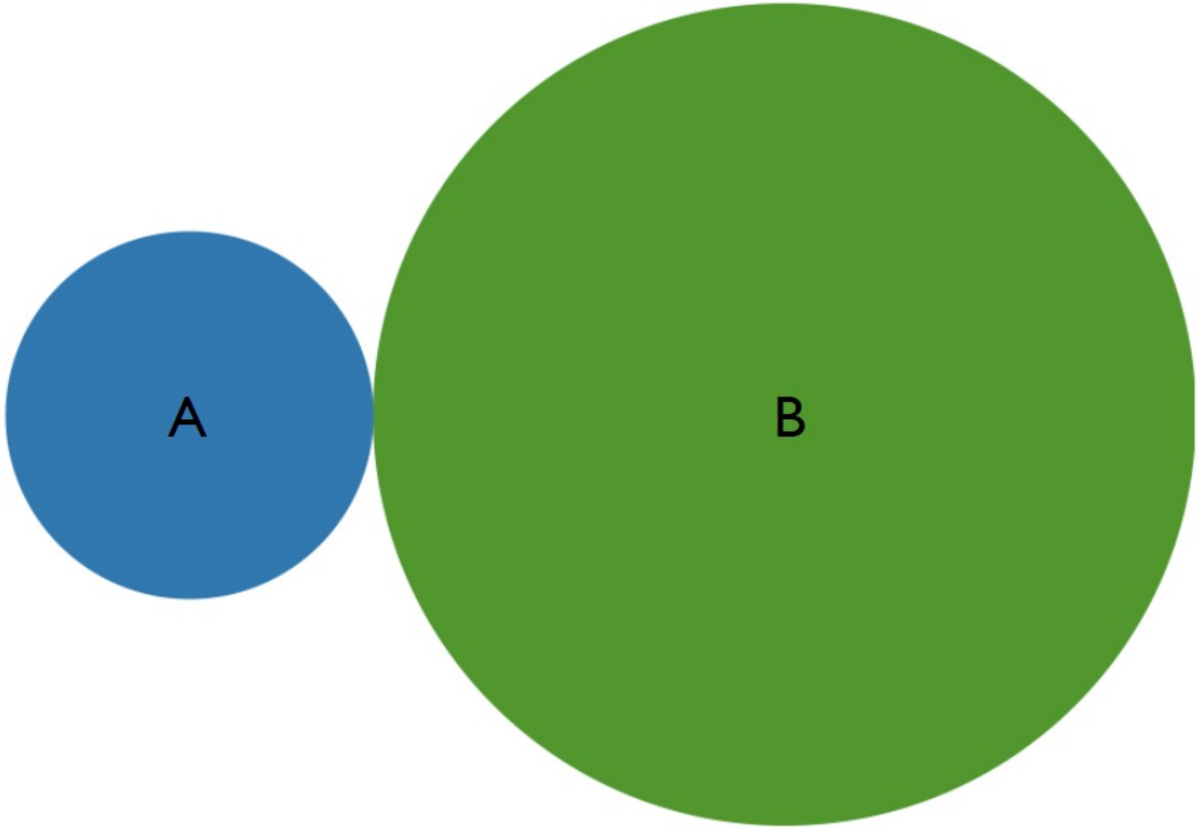
Displays: perceptual effectiveness

How much larger area?



Displays: perceptual effectiveness

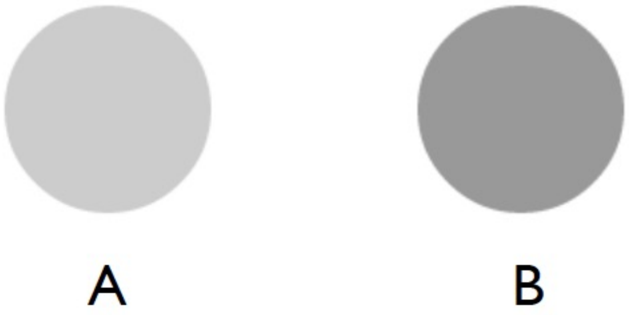
How much larger area?



4.5x

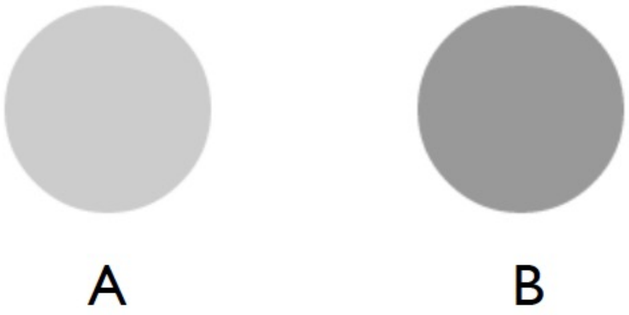
Displays: perceptual effectiveness

How much darker?



Displays: perceptual effectiveness

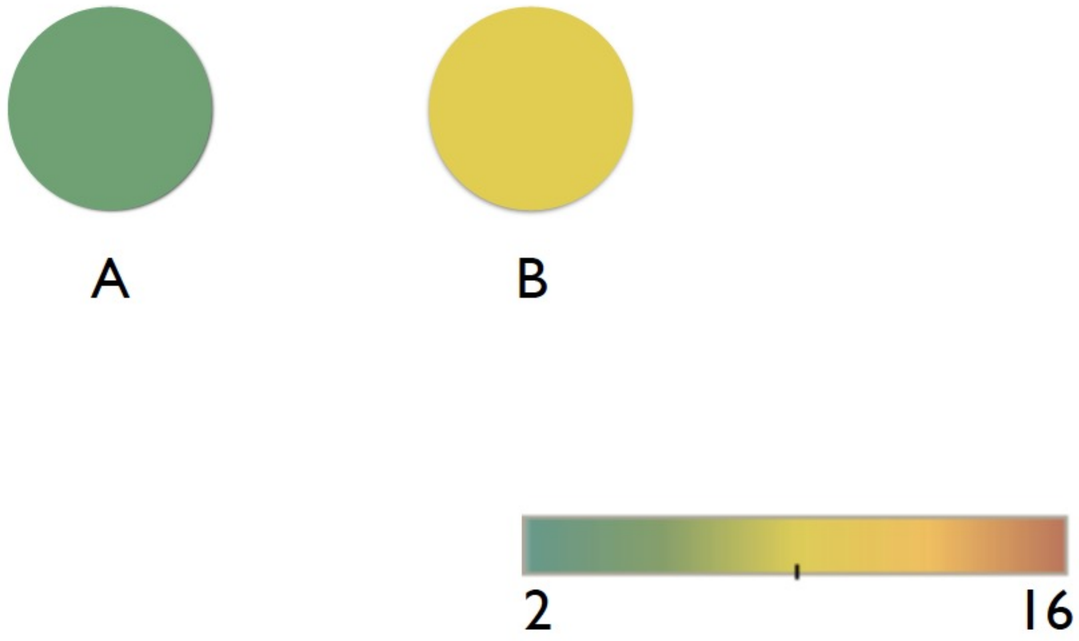
How much darker?



2x

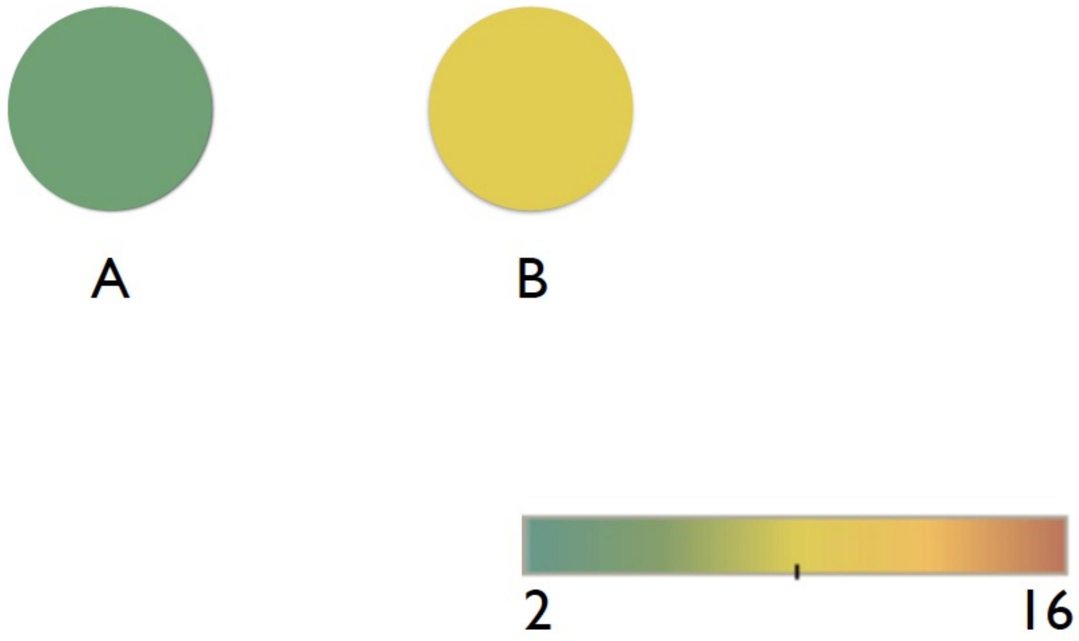
Displays: perceptual effectiveness

How much bigger value?



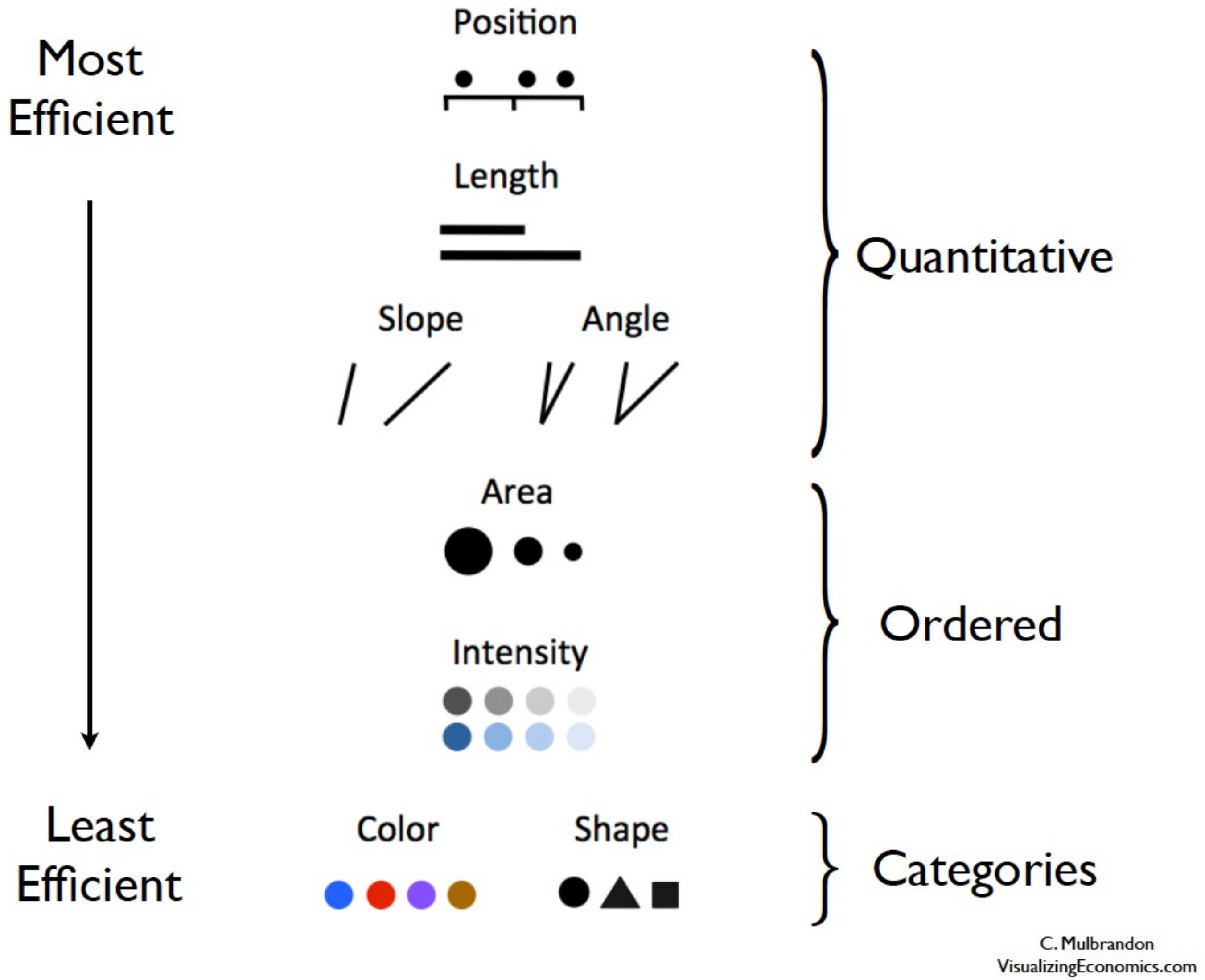
Displays: perceptual effectiveness

How much bigger value?



4x

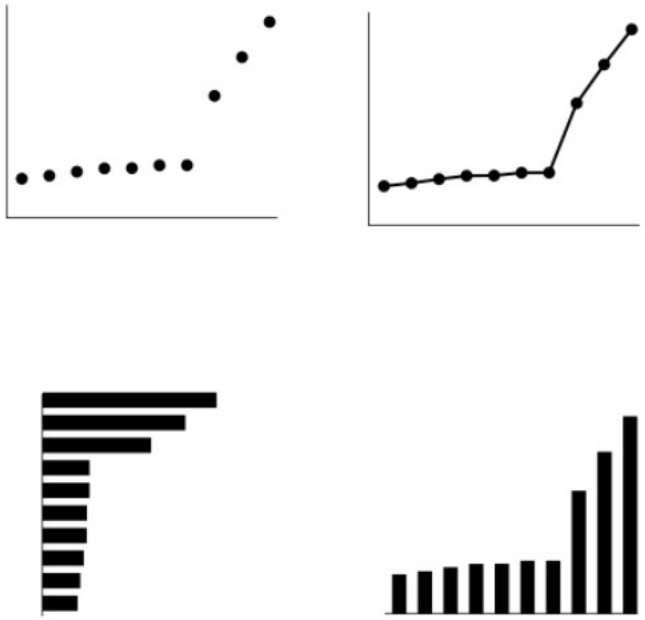
Displays: perceptual effectiveness

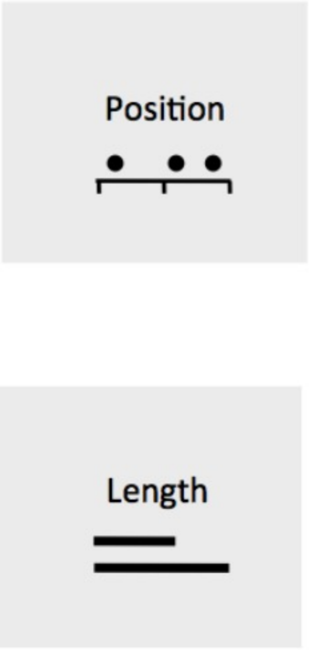


C. Mulbrandon
VisualizingEconomics.com

Displays: perceptual effectiveness

Most Effective

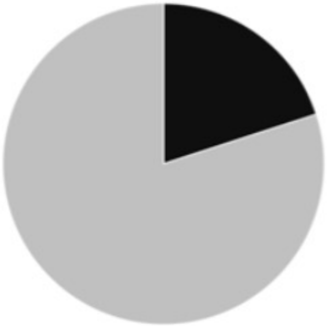


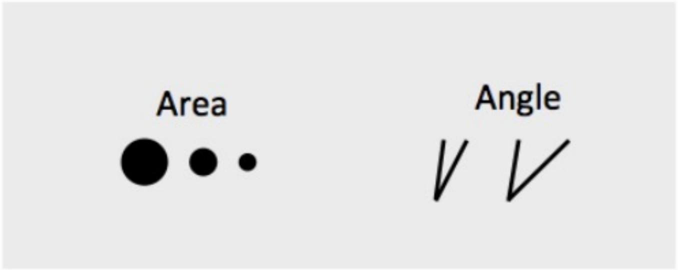


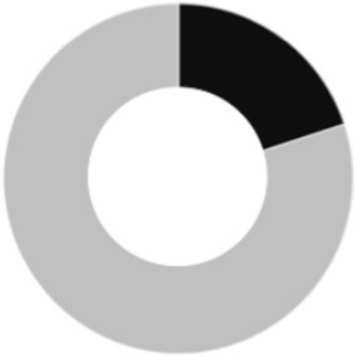
VisualizingEconomics.com

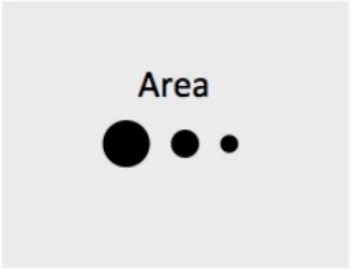
Displays: perceptual effectiveness

Less Effective





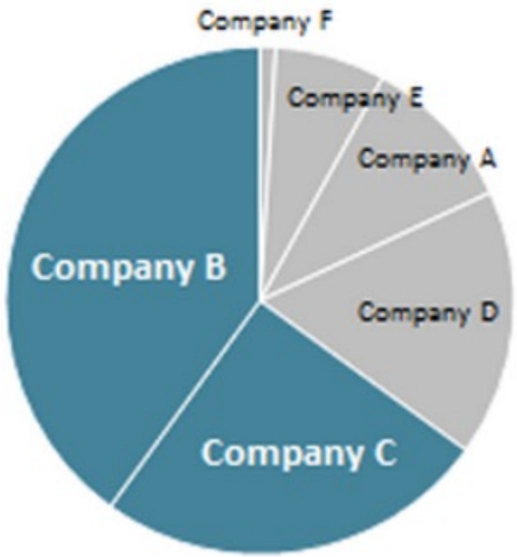


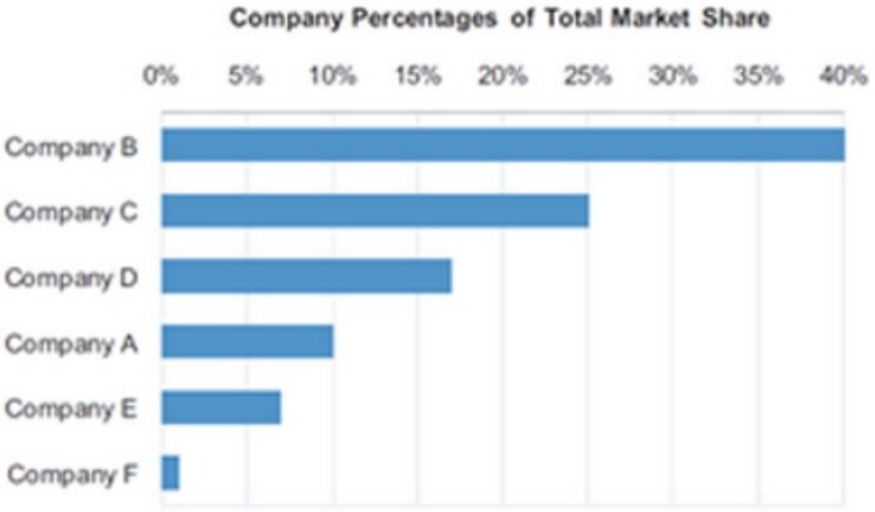


Displays: perceptual effectiveness

Pie vs. Bar Charts

65% of the market is controlled by companies B and C





Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design

- Communication
  - Motivation
  - Key Considerations

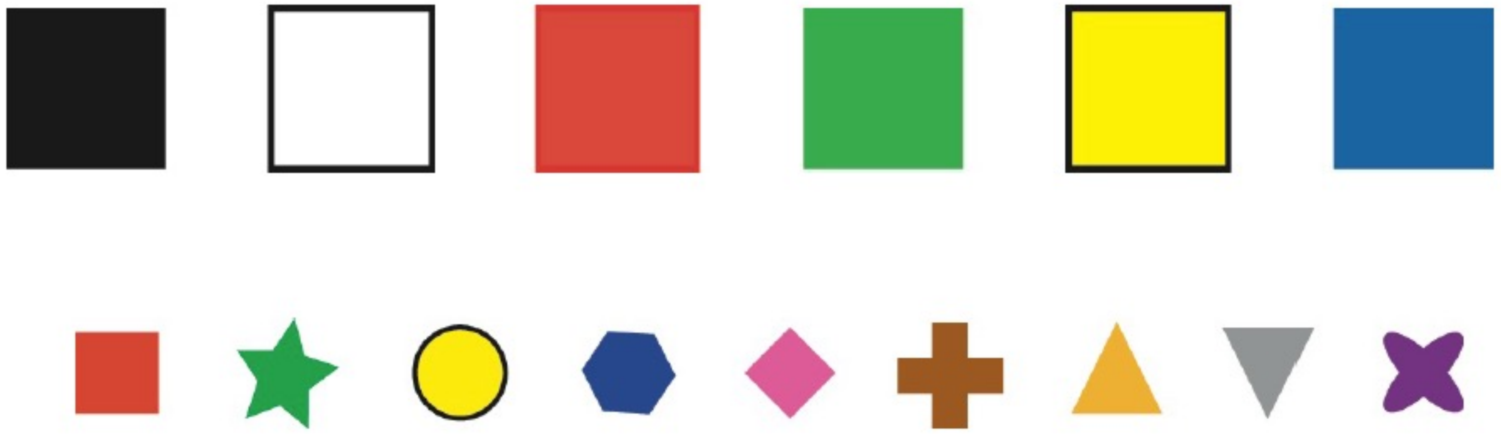
Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

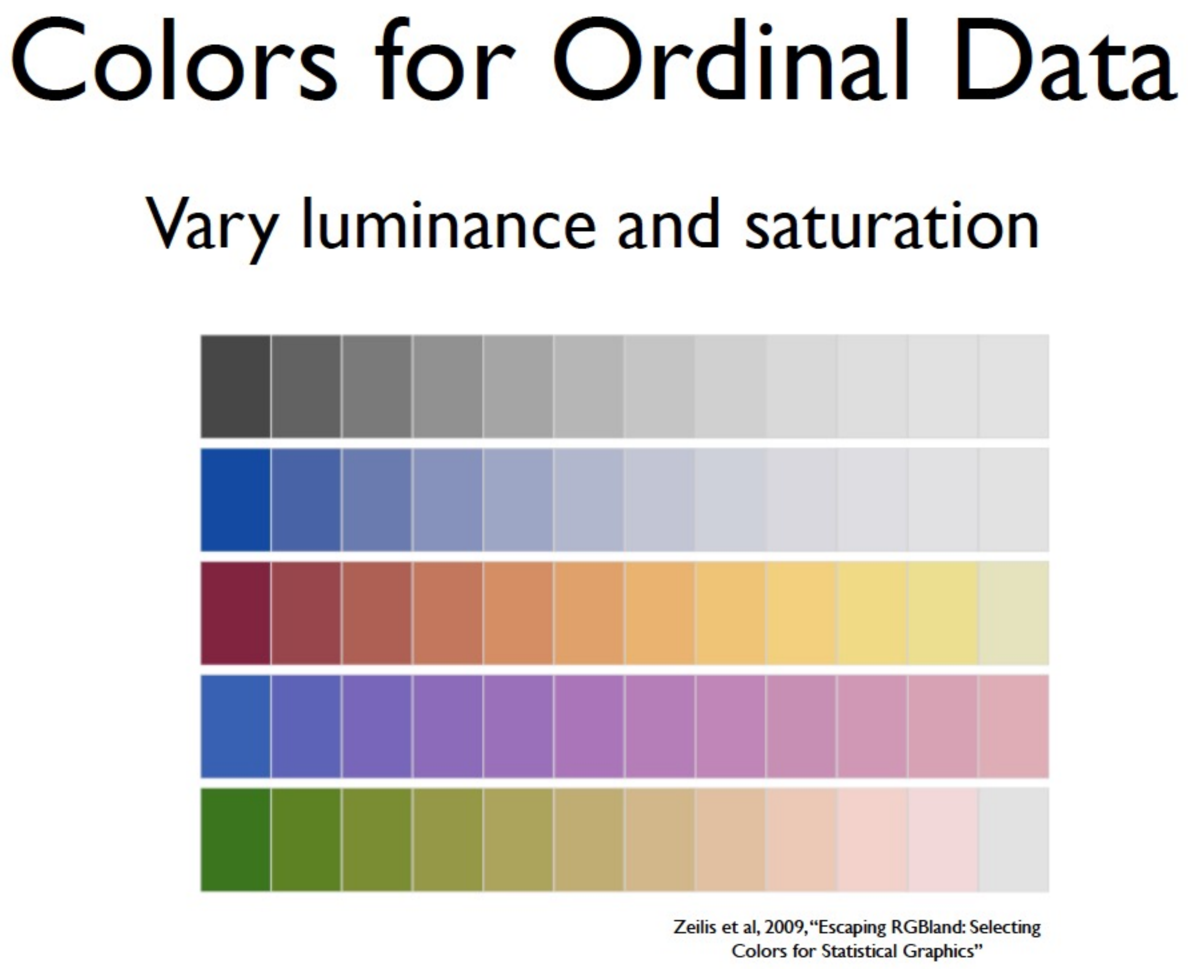
Sensible Design

Colors for Categories

Do not use more than 5-8 colors at once



Sensible Design



Colors for Ordinal Data

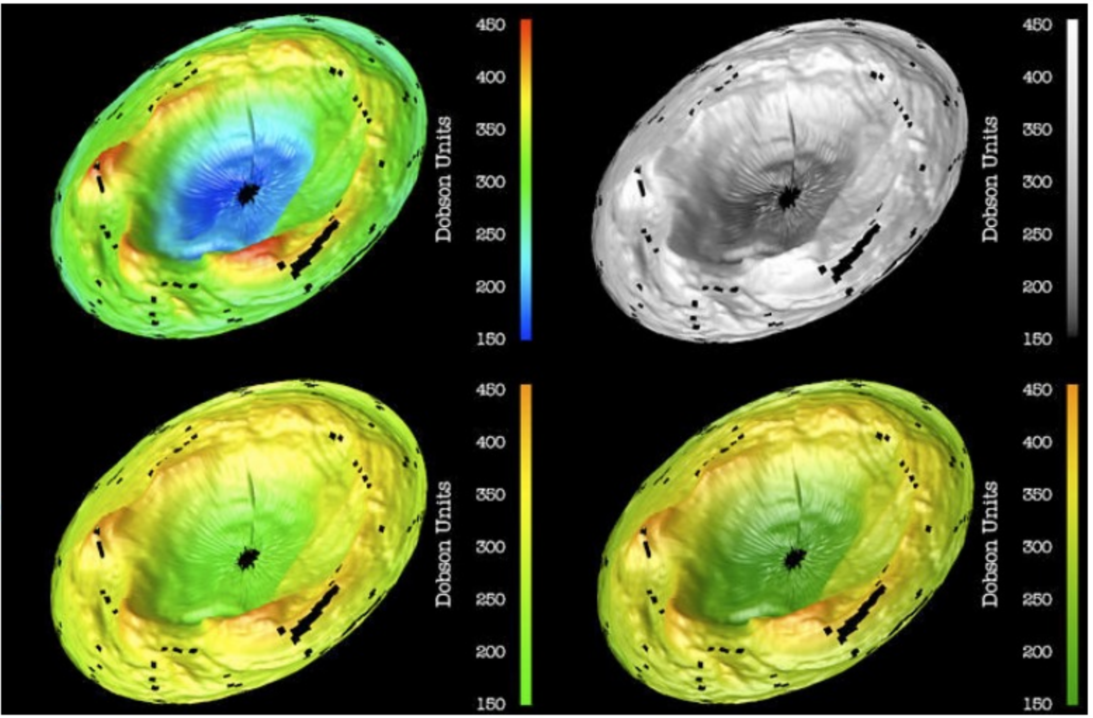
Vary luminance and saturation

Zeilis et al, 2009, "Escaping RGBland: Selecting Colors for Statistical Graphics"

Colors for Quantitative Data

Hue (Rainbow)

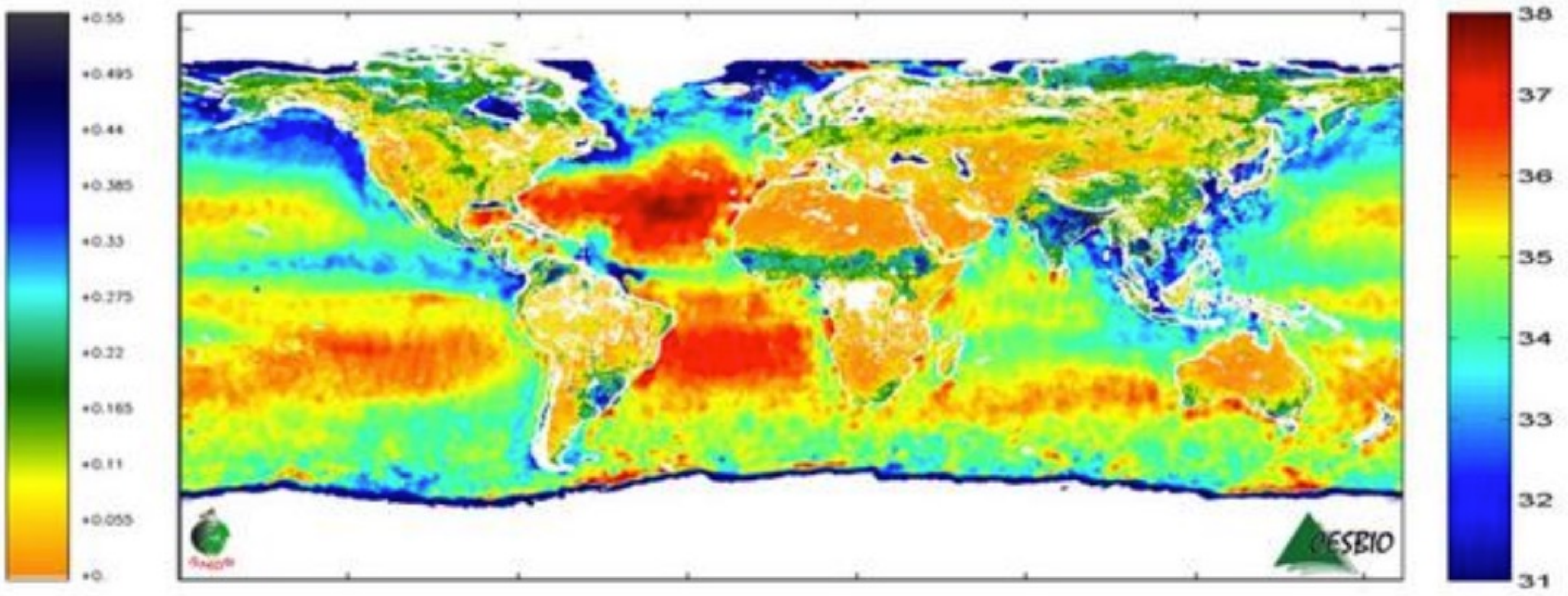
Luminance & Hue



Luminance

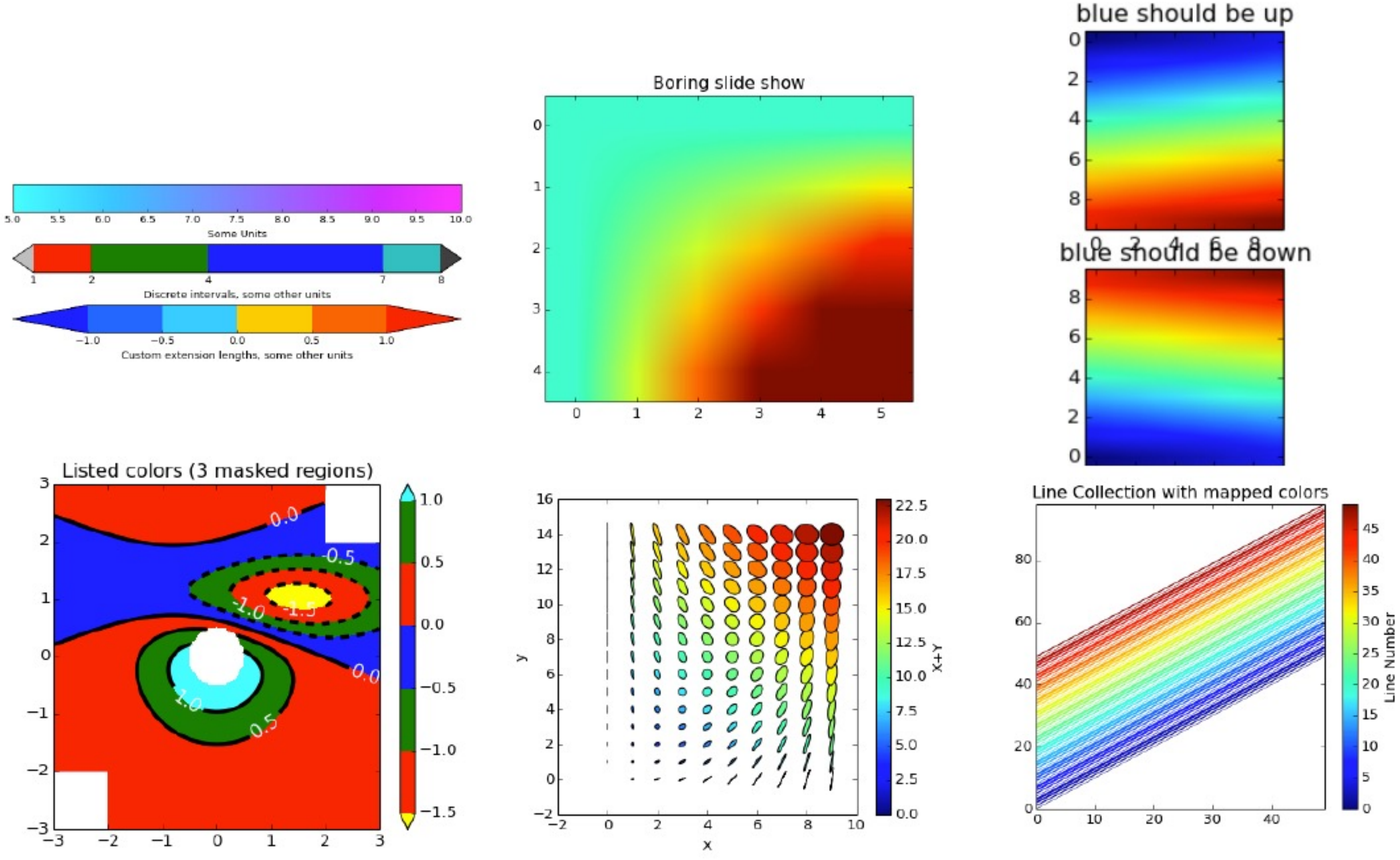
Rogowitz and Treinish, Why should engineers and scientists be worried about color?

Sensible Design: Color Gradients



Sensible Design: Color Gradients

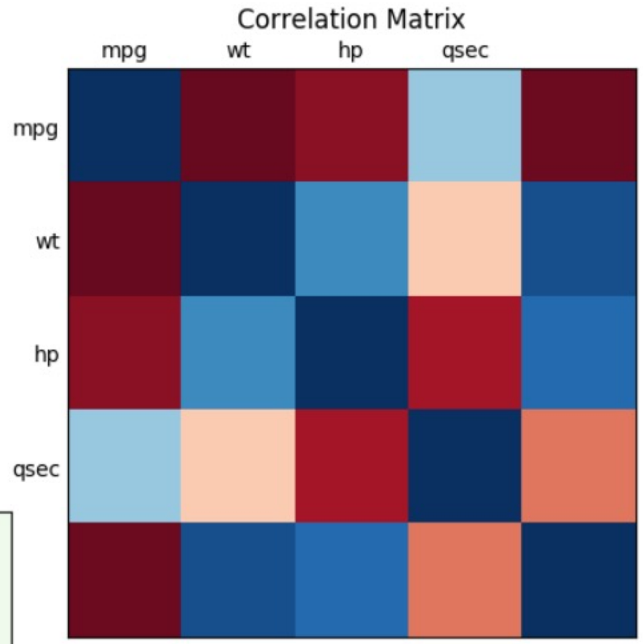
Avoid Rainbow Colors!

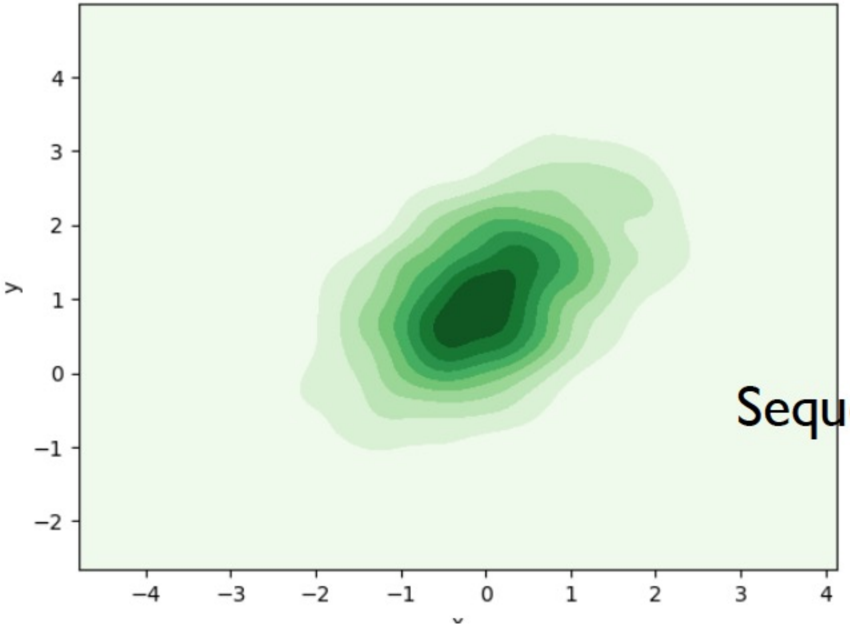


matplotlib gallery

Sensible Design: Color Gradients

Diverging Palette for Quantitative or Ordinal

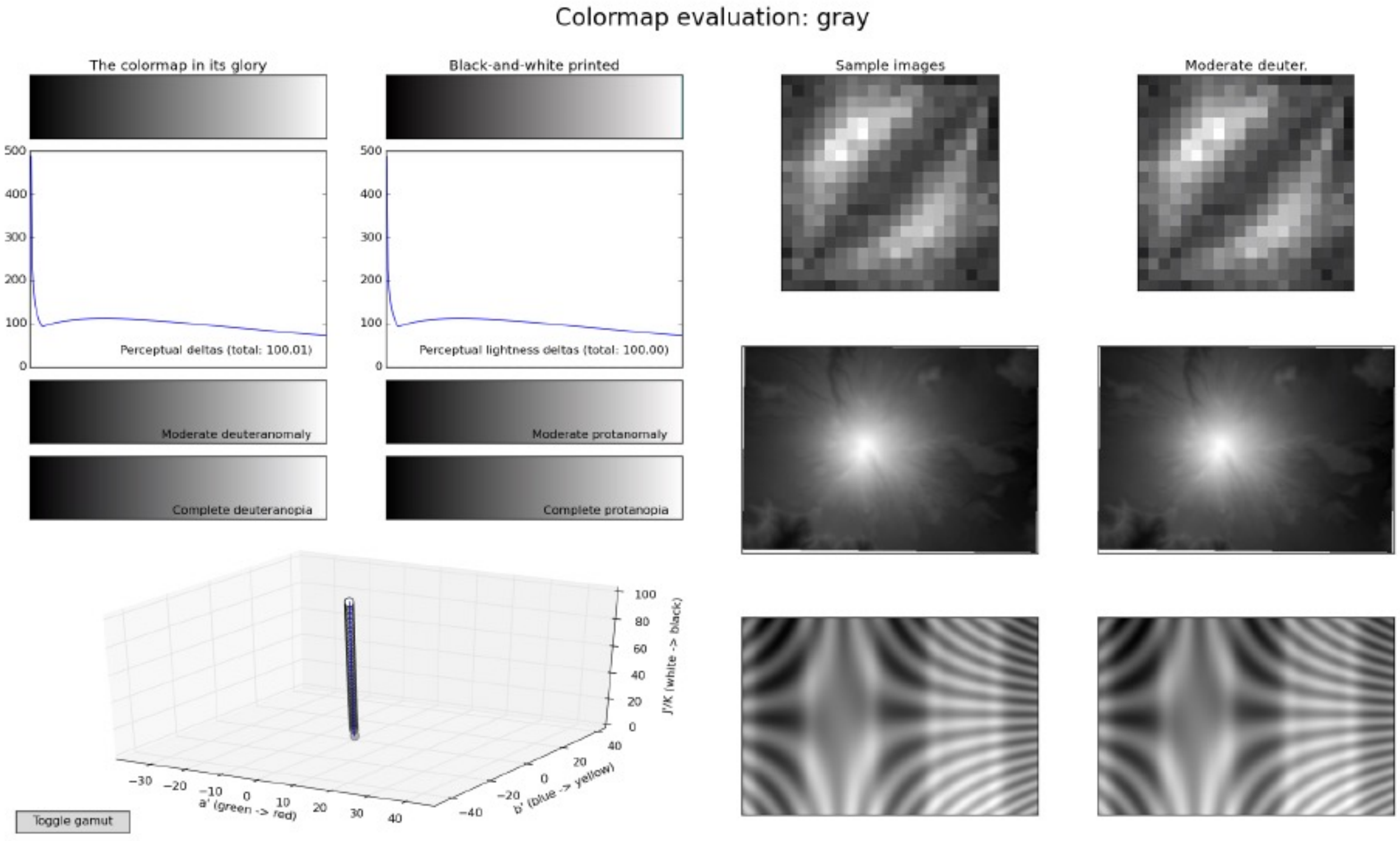




Sequential Palette for Densities

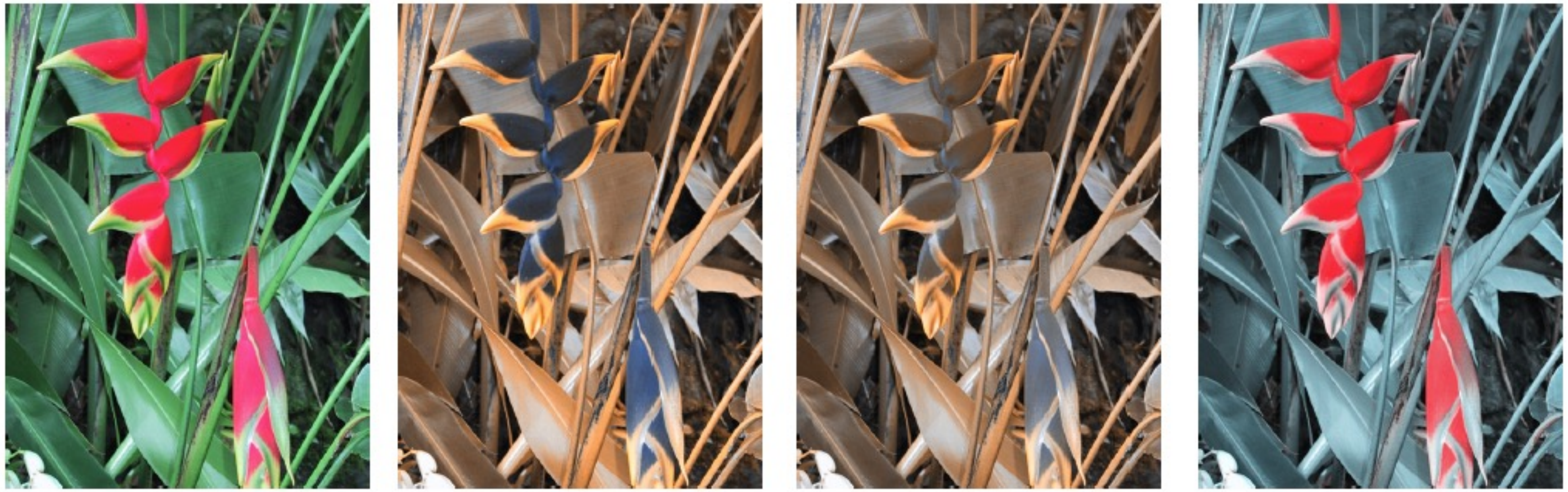
Sensible Design: Color Gradients

Gray



Sensible Design: Color Gradients

Color Blindness



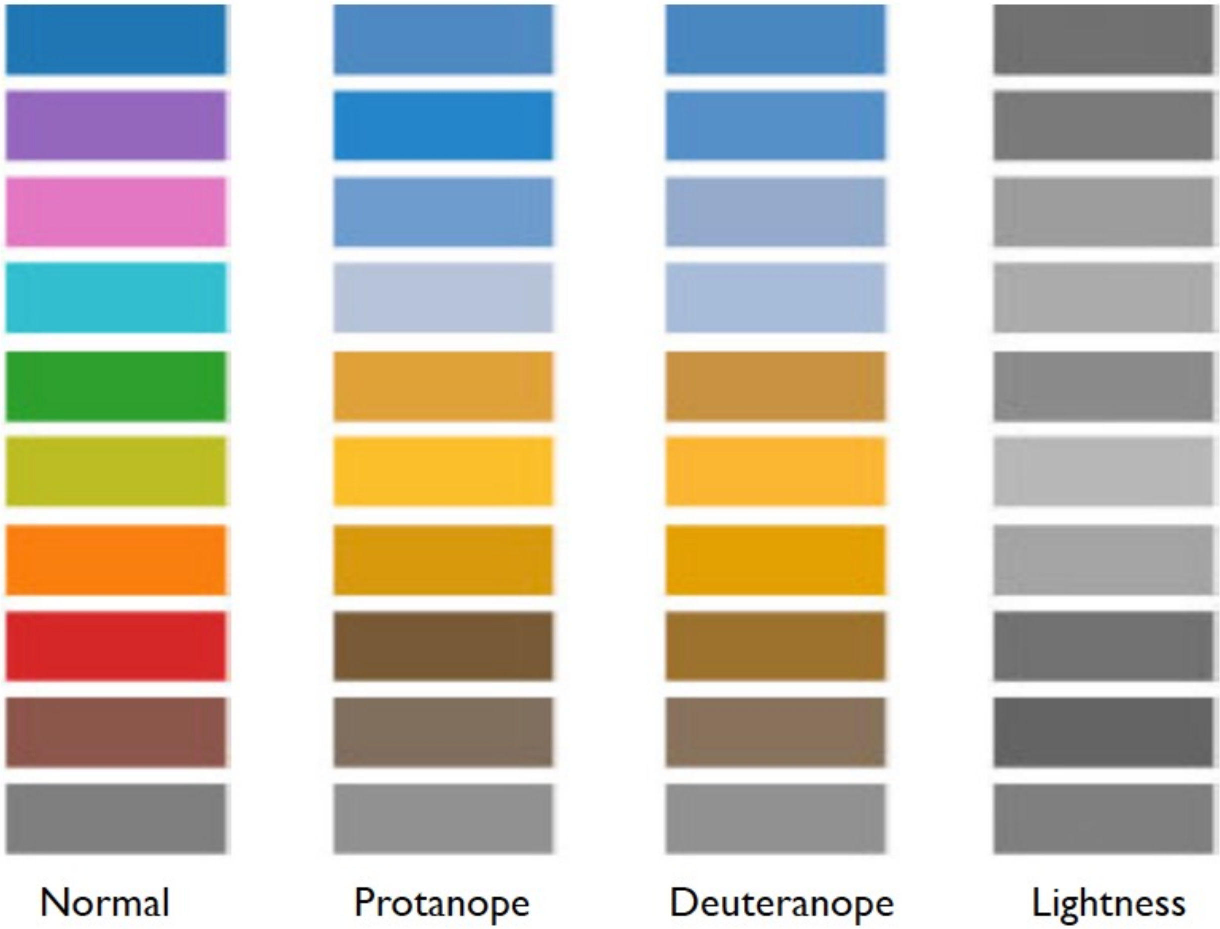
Protanope Deuteranope Tritanope

Red / green deficiencies

Blue / Yellow deficiency

Sensible Design: Color Gradients

Color Blindness

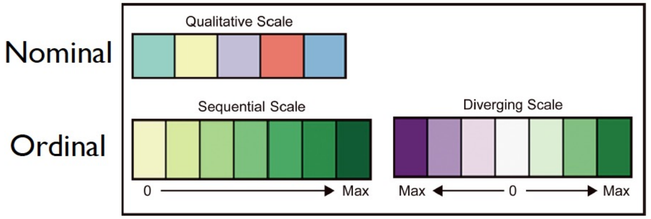


Sensible Design: Color Pickers!

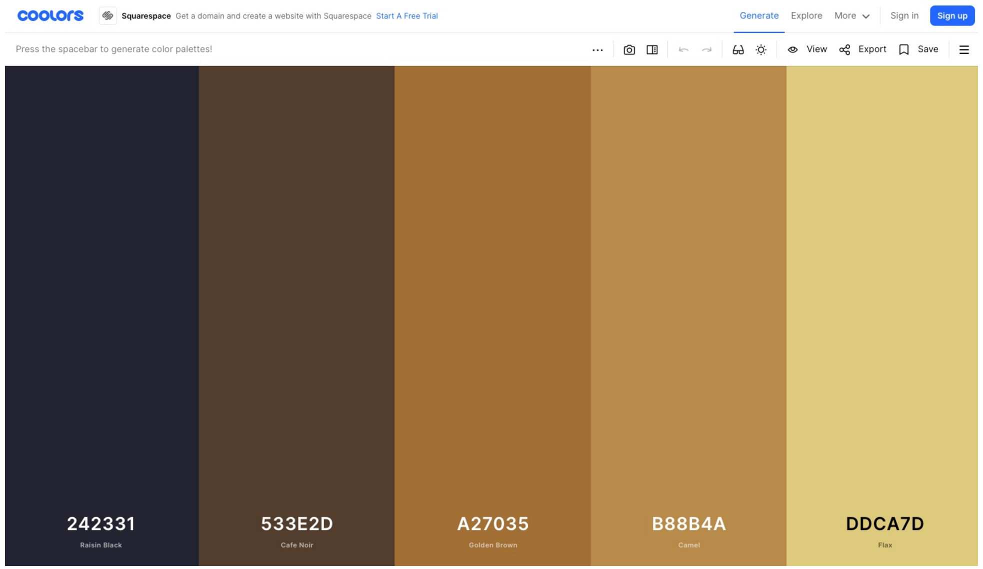
Great sites for selecting color schemes:

- http://colorbrewer2.org
- https://coolors.co/

Color Brewer



Coolors.co


coolors
Squarespace Get a domain and create a website with Squarespace Start A Free Trial
Generate
Explore
More
Sign in
Sign up
View
Export
Save

Press the spacebar to generate color palettes!

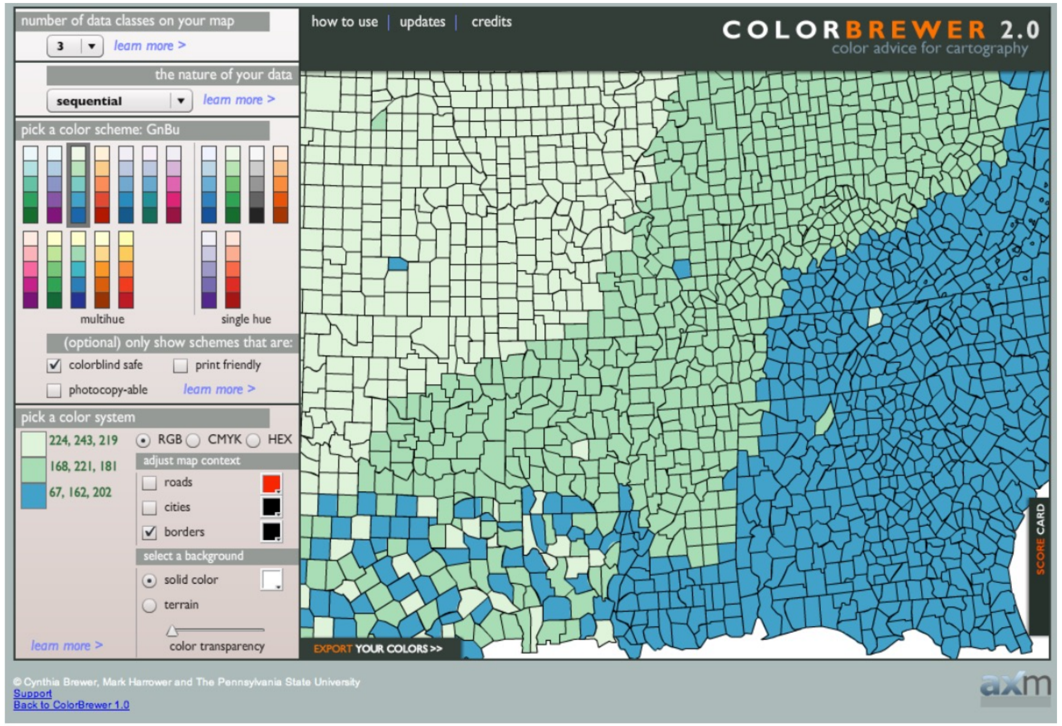
242331 Raisin Black

533E2D Cafe Noir

A27035 Golden Brown

B88B4A Camel

DDCA7D Flax



number of data classes on your map: 3

the nature of your data: sequential

pick a color scheme: GnBu

(optional) only show schemes that are:
☑ colorblind safe □ print friendly
□ photocopy-able

pick a color system
224, 243, 219
168, 221, 181
67, 162, 202
● RGB ○ CMYK ○ HEX

adjust map context
□ roads
□ cities
☑ borders

select a background
● solid color
○ terrain
color transparency

© Cynthia Brewer, Mark Harrower and The Pennsylvania State University
Support
Back to ColorBrewer 1.0

Sensible Design: Typography

How much do you trust this text?

How much do you trust this text?

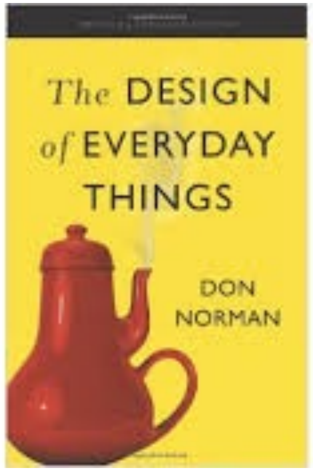
How much do you trust this text?

How much do you trust this text?

Sensible Design: Beyond Data Science

The everyday items that are designed the best are the ones that we never have to think about how to use/interact with it.

Can you think of examples?



Highly recommended

Lesson V: design matters

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

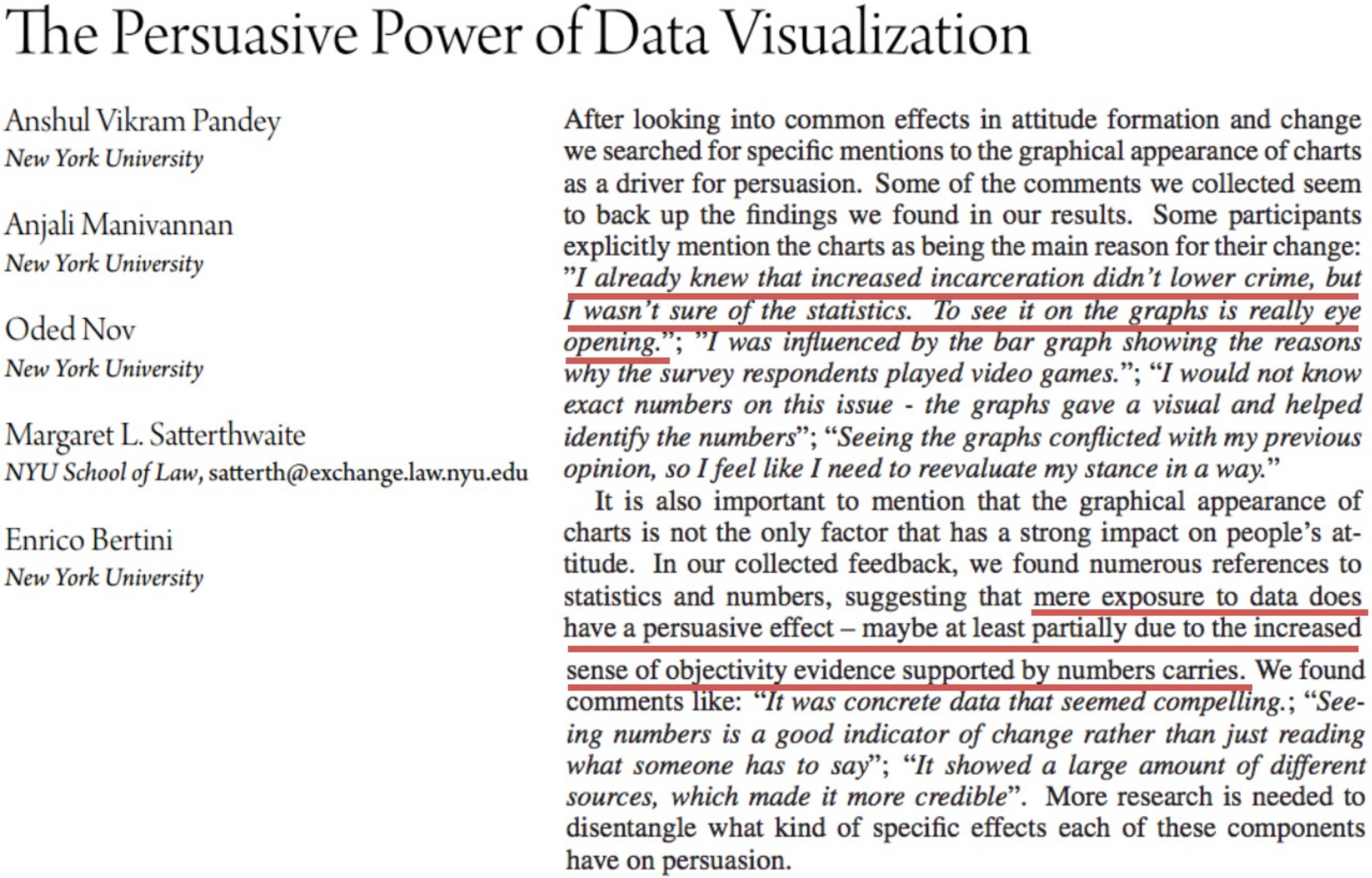
Communication

Analyze (Exploratory)

- Explore the data
- Assess a situation
- Determine how to proceed
- Decide what to do

Communicate (Explanatory)

- Present data and ideas
- Explain and inform
- Provide evidence and support
- Influence and persuade



The Persuasive Power of Data Visualization

Anshul Vikram Pandey
New York University

Anjali Manivannan
New York University

Oded Nov
New York University

Margaret L. Satterthwaite
NYU School of Law, satterth@exchange.law.nyu.edu

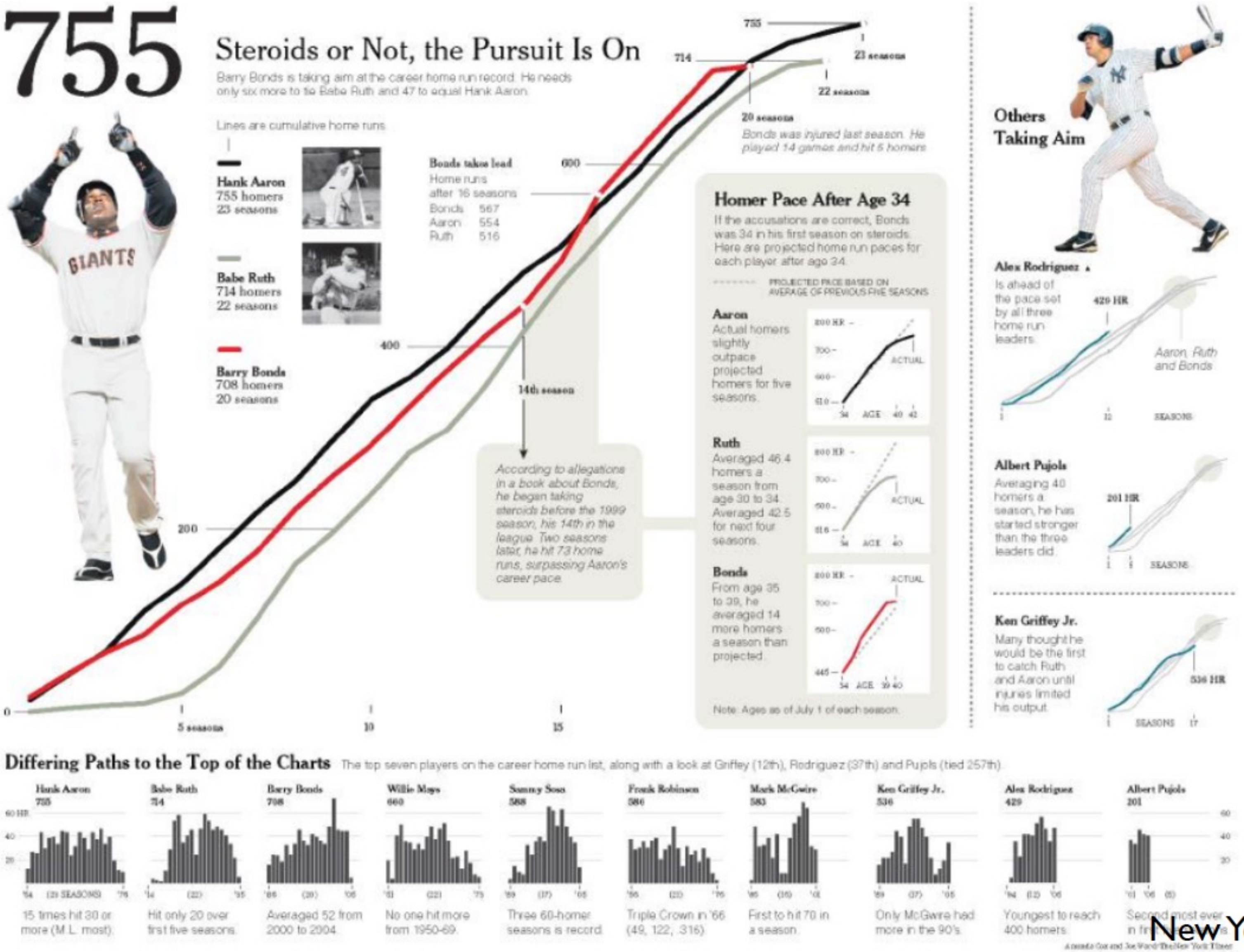
Enrico Bertini
New York University

After looking into common effects in attitude formation and change we searched for specific mentions to the graphical appearance of charts as a driver for persuasion. Some of the comments we collected seem to back up the findings we found in our results. Some participants explicitly mention the charts as being the main reason for their change: *"I already knew that increased incarceration didn't lower crime, but I wasn't sure of the statistics. To see it on the graphs is really eye opening."*; *"I was influenced by the bar graph showing the reasons why the survey respondents played video games."*; *"I would not know exact numbers on this issue - the graphs gave a visual and helped identify the numbers"*; *"Seeing the graphs conflicted with my previous opinion, so I feel like I need to reevaluate my stance in a way."*

It is also important to mention that the graphical appearance of charts is not the only factor that has a strong impact on people's attitude. In our collected feedback, we found numerous references to statistics and numbers, suggesting that mere exposure to data does have a persuasive effect – maybe at least partially due to the increased sense of objectivity evidence supported by numbers carries. We found comments like: *"It was concrete data that seemed compelling.*; *"Seeing numbers is a good indicator of change rather than just reading what someone has to say"*; *"It showed a large amount of different sources, which made it more credible"*. More research is needed to disentangle what kind of specific effects each of these components have on persuasion.

http://lsr.nellco.org/cgi/viewcontent.cgi?article=1476&context=nyu_plltwp

Communication



Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Bonds takes lead

Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

Bonds was injured last season. He played 14 games and hit 5 homers

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

Aaron
Actual homers slightly outpace projected homers for five seasons.

Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

Note: Ages as of July 1 of each season.

Others Taking Aim

Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.

Aaron, Ruth and Bonds

Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.

Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).

Hank Aaron 755 — 15 times hit 30 or more (M.L. most).

Babe Ruth 714 — Hit only 20 over first five seasons.

Barry Bonds 708 — Averaged 52 from 2000 to 2004.

Willie Mays 660 — No one hit more from 1950-69.

Sammy Sosa 588 — Three 60-homer seasons is record.

Frank Robinson 586 — Triple Crown in '66 (49, 122, .316).

Mark McGwire 583 — First to hit 70 in a season.

Ken Griffey Jr. 536 — Only McGwire had more in the 90's.

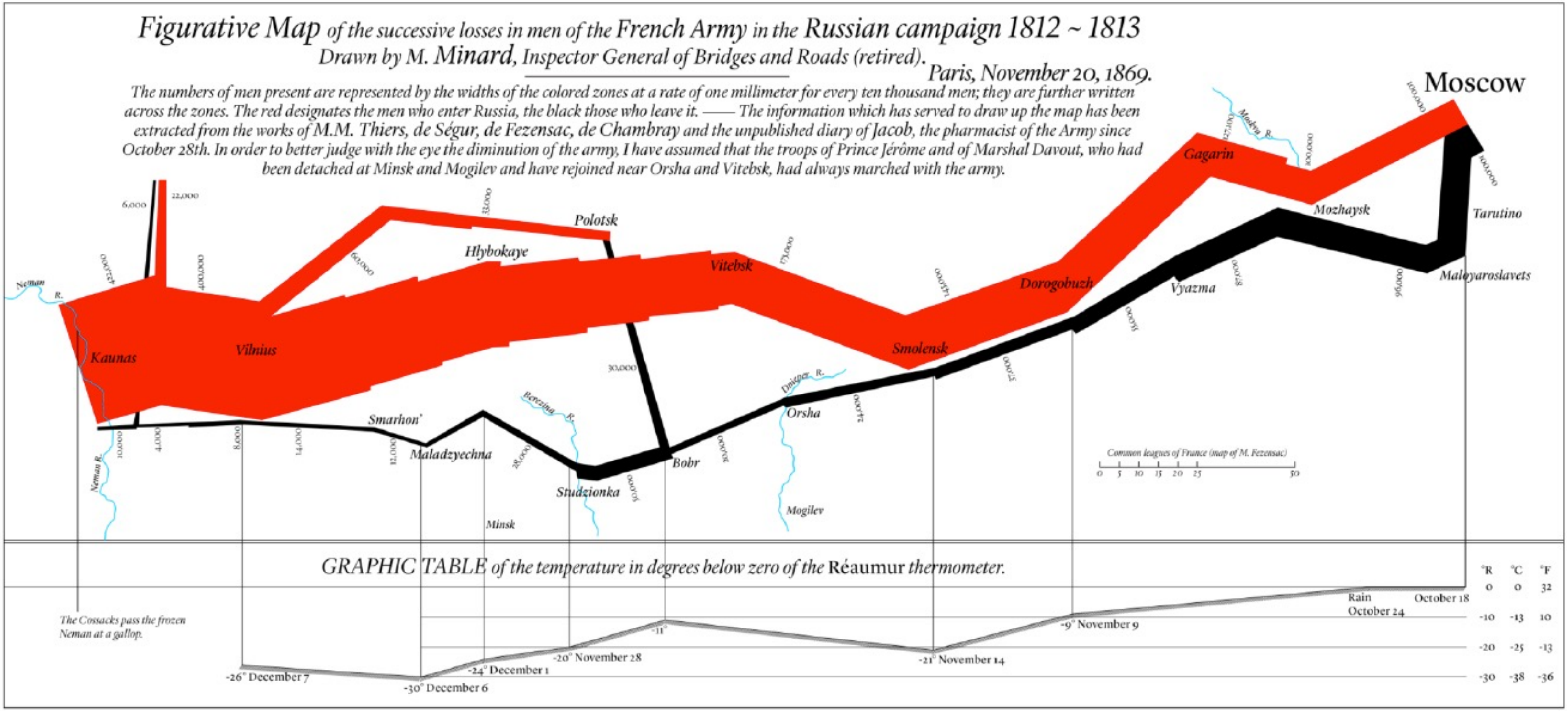
Alex Rodriguez 429 — Youngest to reach 400 homers.

Albert Pujols 201 — Second most ever in fi[illegible]

Amanda Cox and Joe Ward/The New York Times

New York Time:

Minard's Graphic on Napoleon's Russia Campaign



Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812 ~ 1813
Drawn by M. Minard, Inspector General of Bridges and Roads (retired).
Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten thousand men; they are further written across the zones. The red designates the men who enter Russia, the black those who leave it. —— The information which has served to draw up the map has been extracted from the works of M.M. Thiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, the pharmacist of the Army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of Prince Jérôme and of Marshal Davout, who had been detached at Minsk and Mogilev and have rejoined near Orsha and Vitebsk, had always marched with the army.

GRAPHIC TABLE of the temperature in degrees below zero of the Réaumur thermometer.

The Cossacks pass the frozen Neman at a gallop.

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design

- Communication
  - Motivation
  - Key Considerations

Agenda

- EDA Refresher
- Effective Visualization
  - Graphical Integrity
  - Scope
  - Displays
  - Sensible Design
- Communication
  - Motivation
  - Key Considerations

Communication

Key Considerations

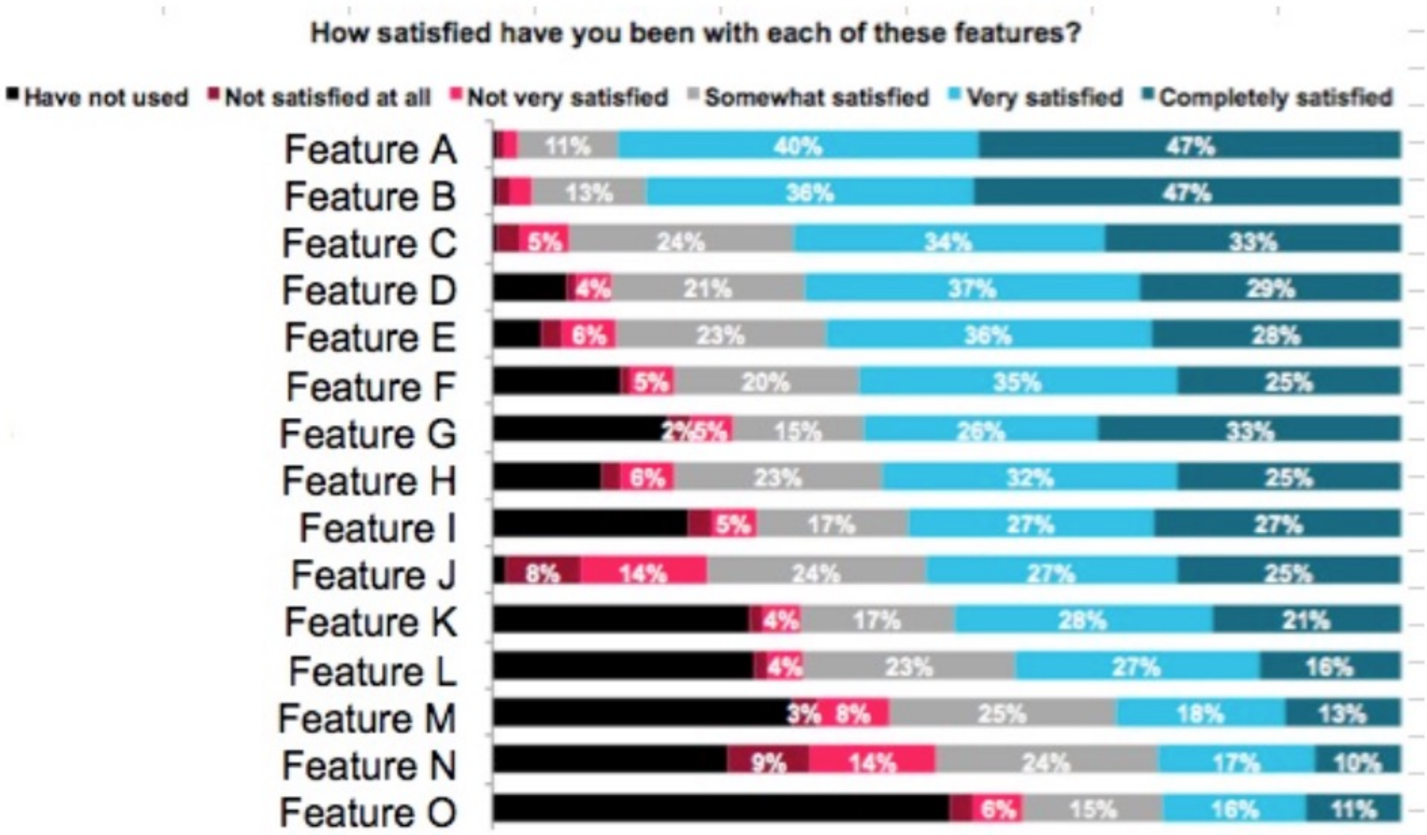
- Who is your **audience**
- What questions are you answering?
- Why should the audience care?
- What are your major insights and surprises?
- What change do you want to affect?

Communication

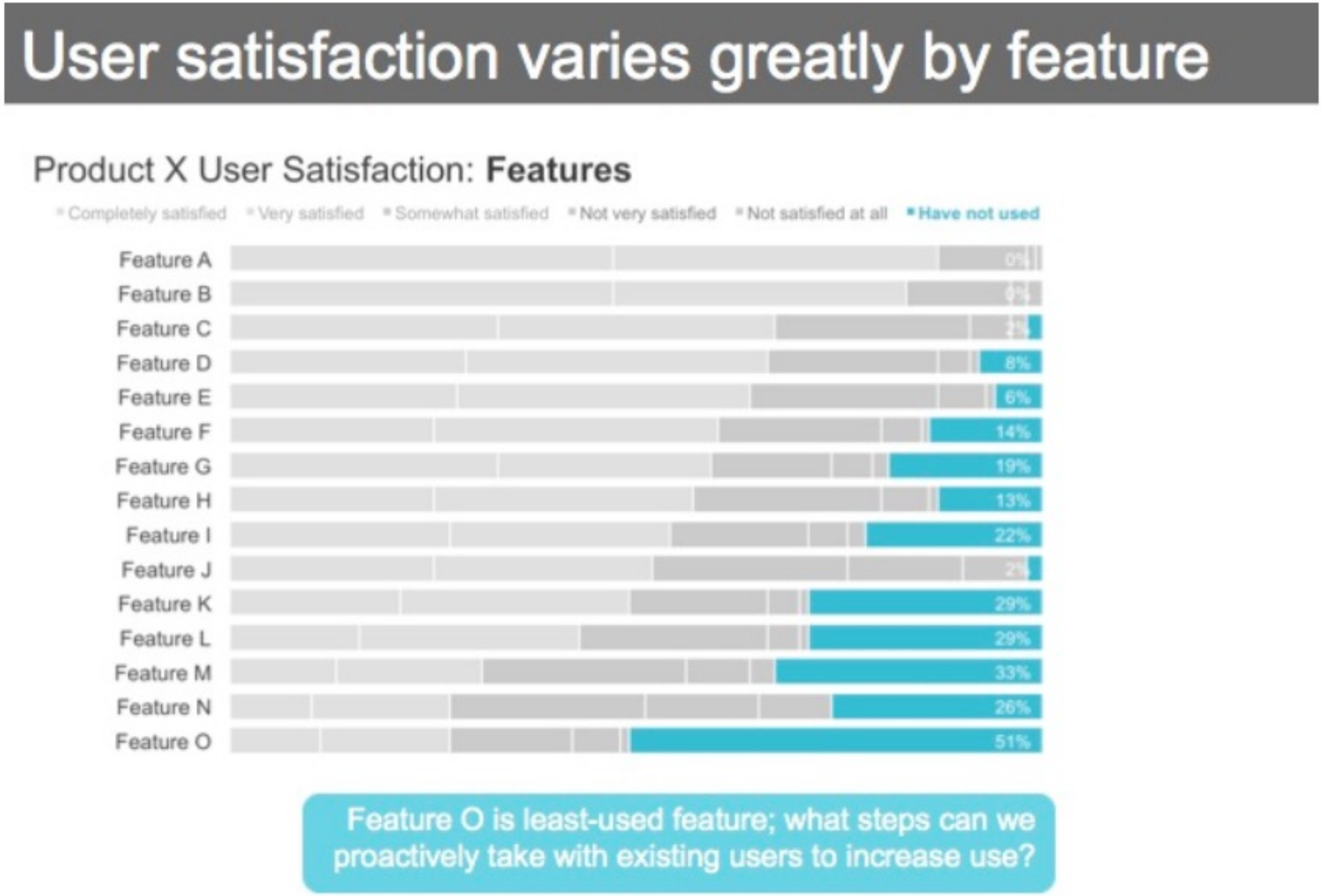
Don’t make them think!

- The audience does not want to spend cognitive energy on dissecting and decoding your intended message.
- Lead them through the **major** steps of your story.
- Point out interesting key facts and insights using **captions** and **annotations**

Don’t Bury the Lead



Communication



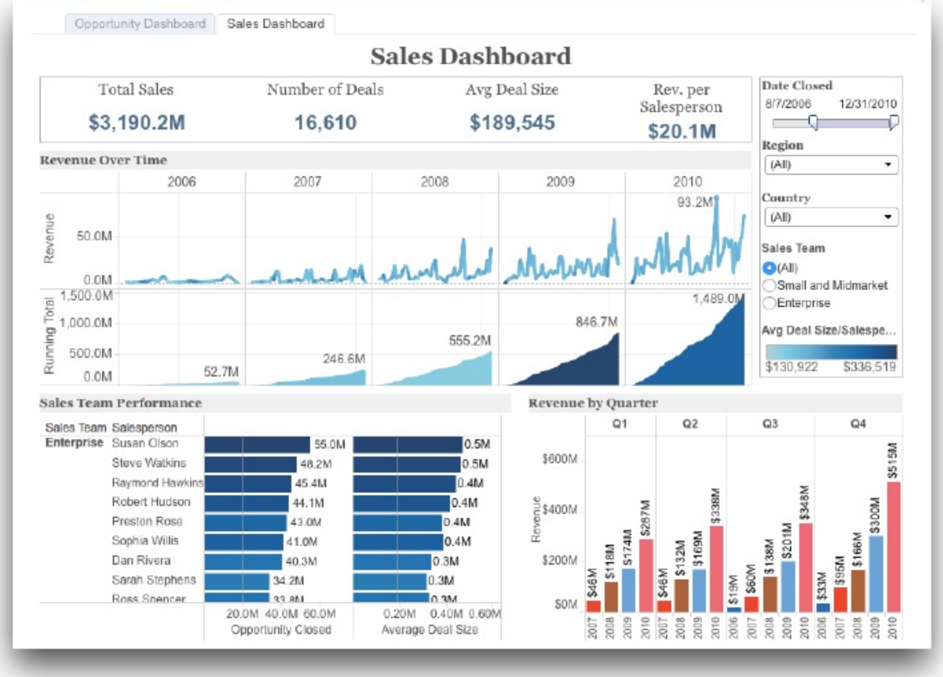
User satisfaction varies greatly by feature

Product X User Satisfaction: **Features**

Feature O is least-used feature; what steps can we proactively take with existing users to increase use?

What is the message?

Exploratory
Neutral



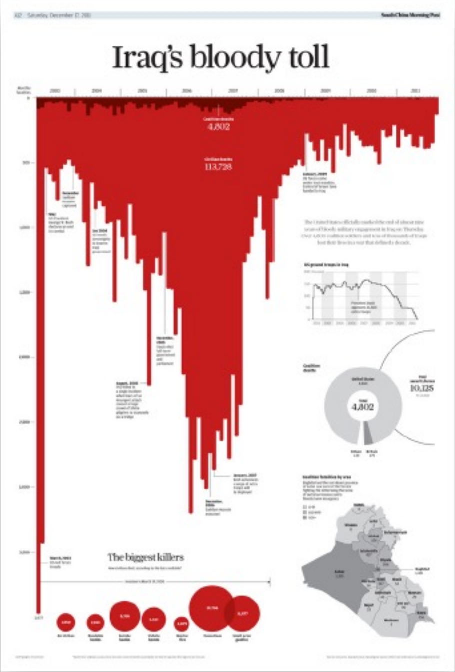
Opportunity Dashboard | Sales Dashboard

Sales Dashboard

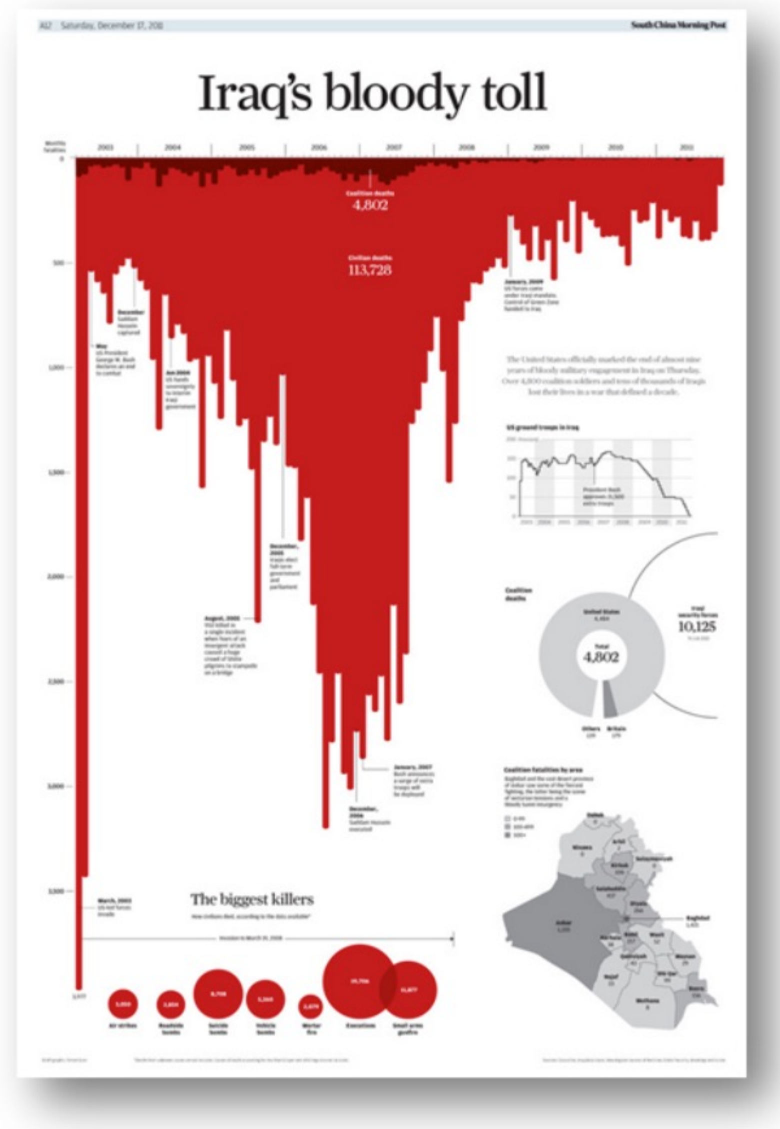
Total Sales	Number of Deals	Avg Deal Size	Rev. per Salesperson
$3,190.2M	16,610	$189,545	$20.1M

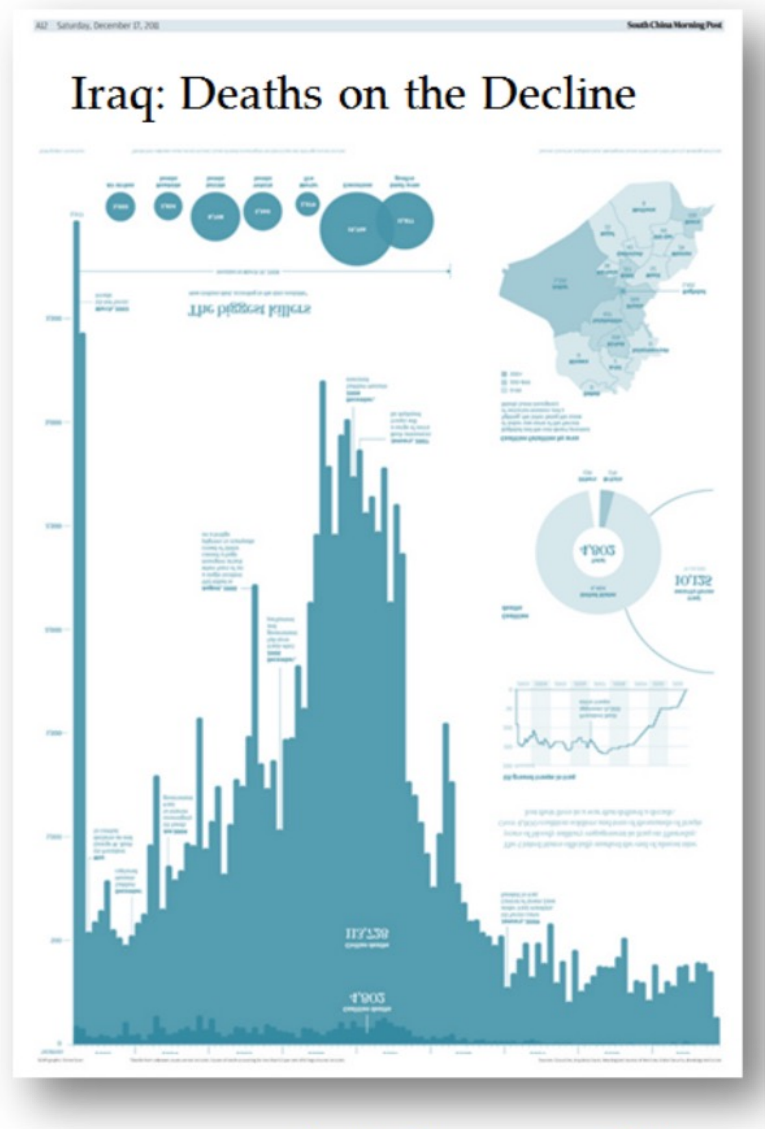
Date Closed: 8/7/2006 – 12/31/2010
Region: (All)
Country: (All)
Sales Team: ● (All) ○ Small and Midmarket ○ Enterprise
Avg Deal Size/Salespe...: $130,922 – $336,519

Explanatory
Opinionated



Communication





Andy Cotgreave, Tableau

Final Takeaways

- How you choose to display your data greatly influences how people interpret the data

- Humans are visual, *emotional* creations; make graphs that don't make others **feel** confused, insulted, etc.

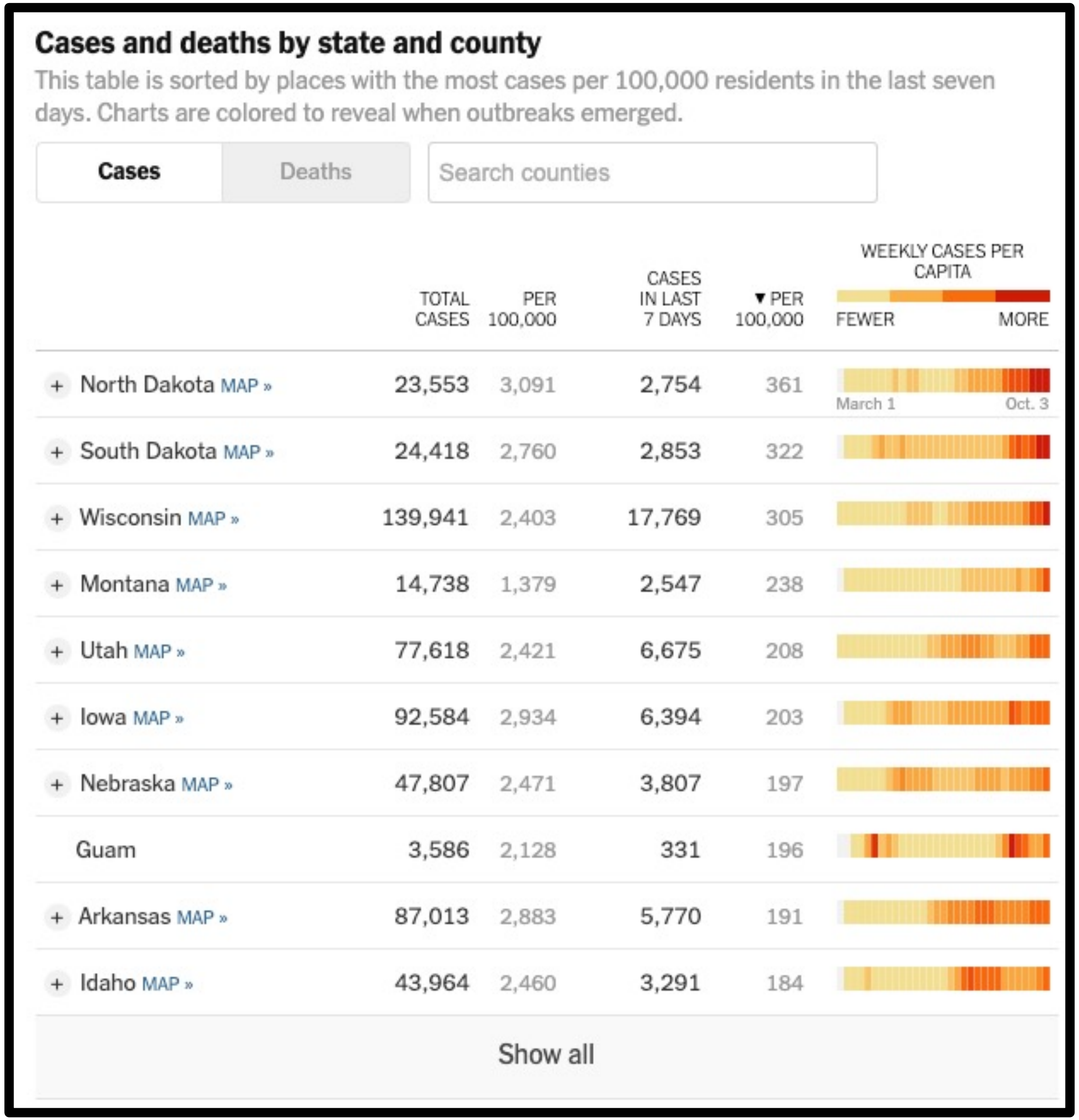
- Your graphs should illicit good feelings and effectively convey your narrative

Suggested Python Packages

- Matplotlib
- Seaborn
- plotly
- ggplot

Further Good Examples

- https://www.nytimes.com/ tends to have incredibly high-quality visualizations that convey information seamlessly
- https://www.reddit.com/r/dataisbeautiful/
- fivethirtyeight.com



Cases and deaths by state and county

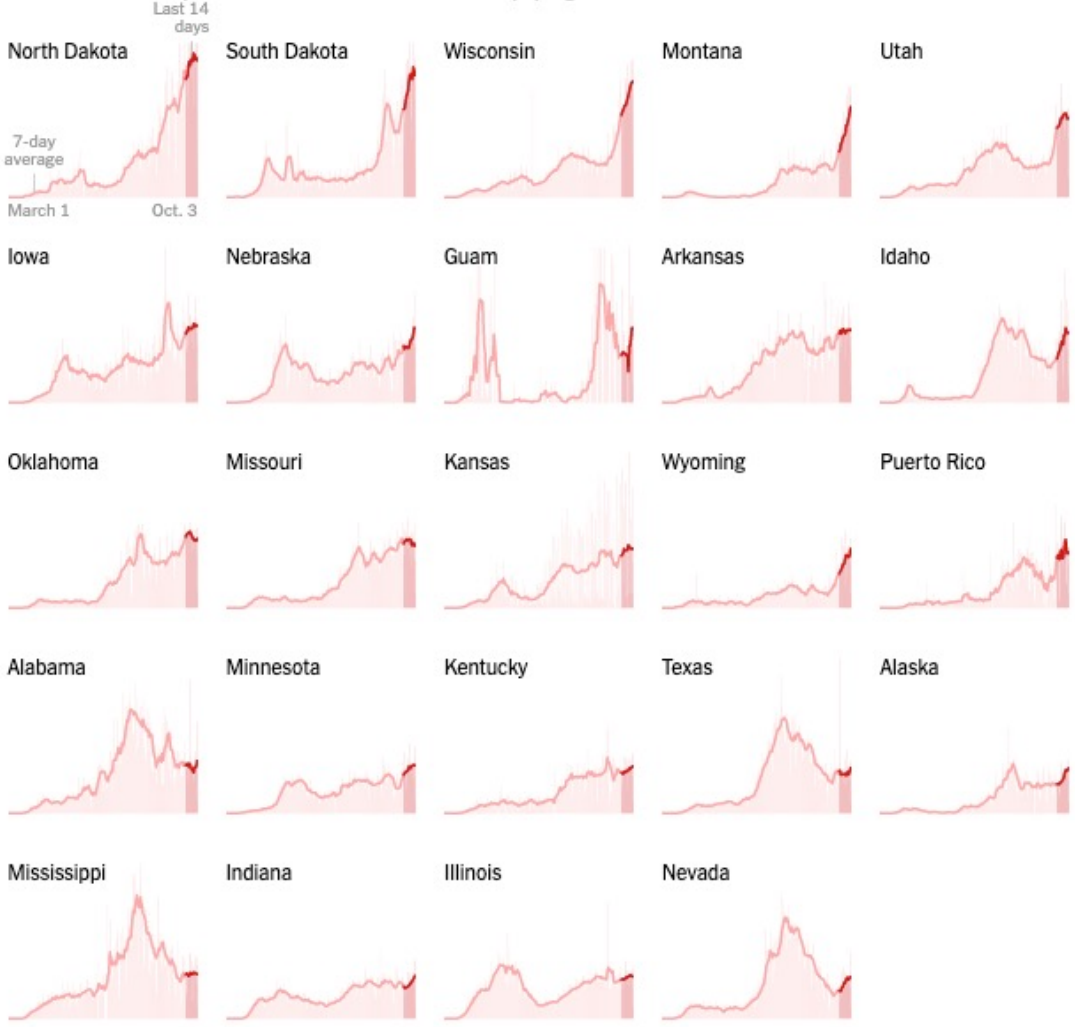
This table is sorted by places with the most cases per 100,000 residents in the last seven days. Charts are colored to reveal when outbreaks emerged.

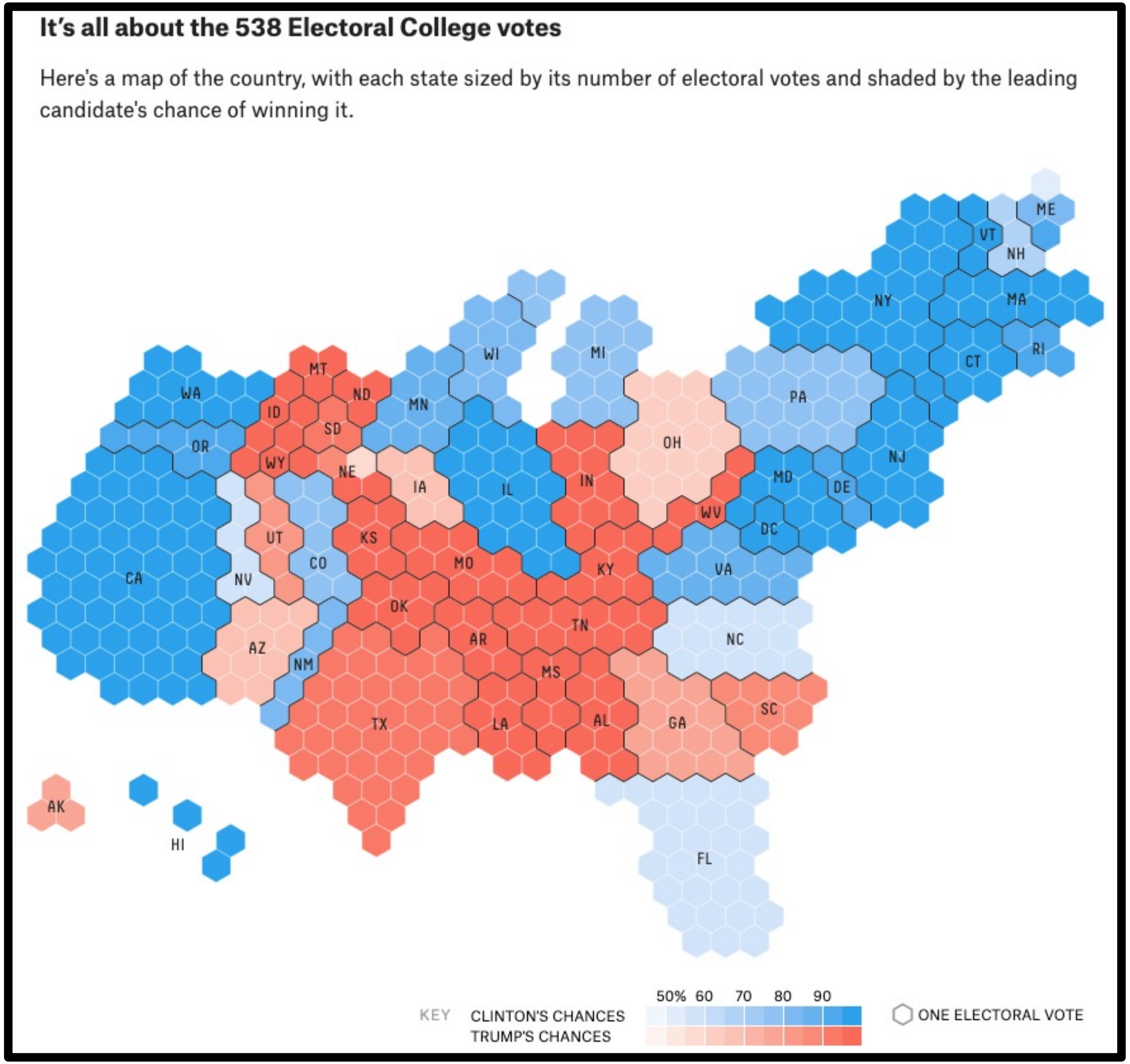
	TOTAL CASES	PER 100,000	CASES IN LAST 7 DAYS	▼ PER 100,000	WEEKLY CASES PER CAPITA
+ North Dakota MAP »	23,553	3,091	2,754	361	
+ South Dakota MAP »	24,418	2,760	2,853	322	
+ Wisconsin MAP »	139,941	2,403	17,769	305	
+ Montana MAP »	14,738	1,379	2,547	238	
+ Utah MAP »	77,618	2,421	6,675	208	
+ Iowa MAP »	92,584	2,934	6,394	203	
+ Nebraska MAP »	47,807	2,471	3,807	197	
Guam	3,586	2,128	331	196	
+ Arkansas MAP »	87,013	2,883	5,770	191	
+ Idaho MAP »	43,964	2,460	3,291	184	

Show all

Where new cases are higher and staying high

States where new cases are higher had a daily average of at least 15 new cases per 100,000 people over the past week. Charts show daily cases per capita and are on the same scale. Tap a state to see detailed map page.



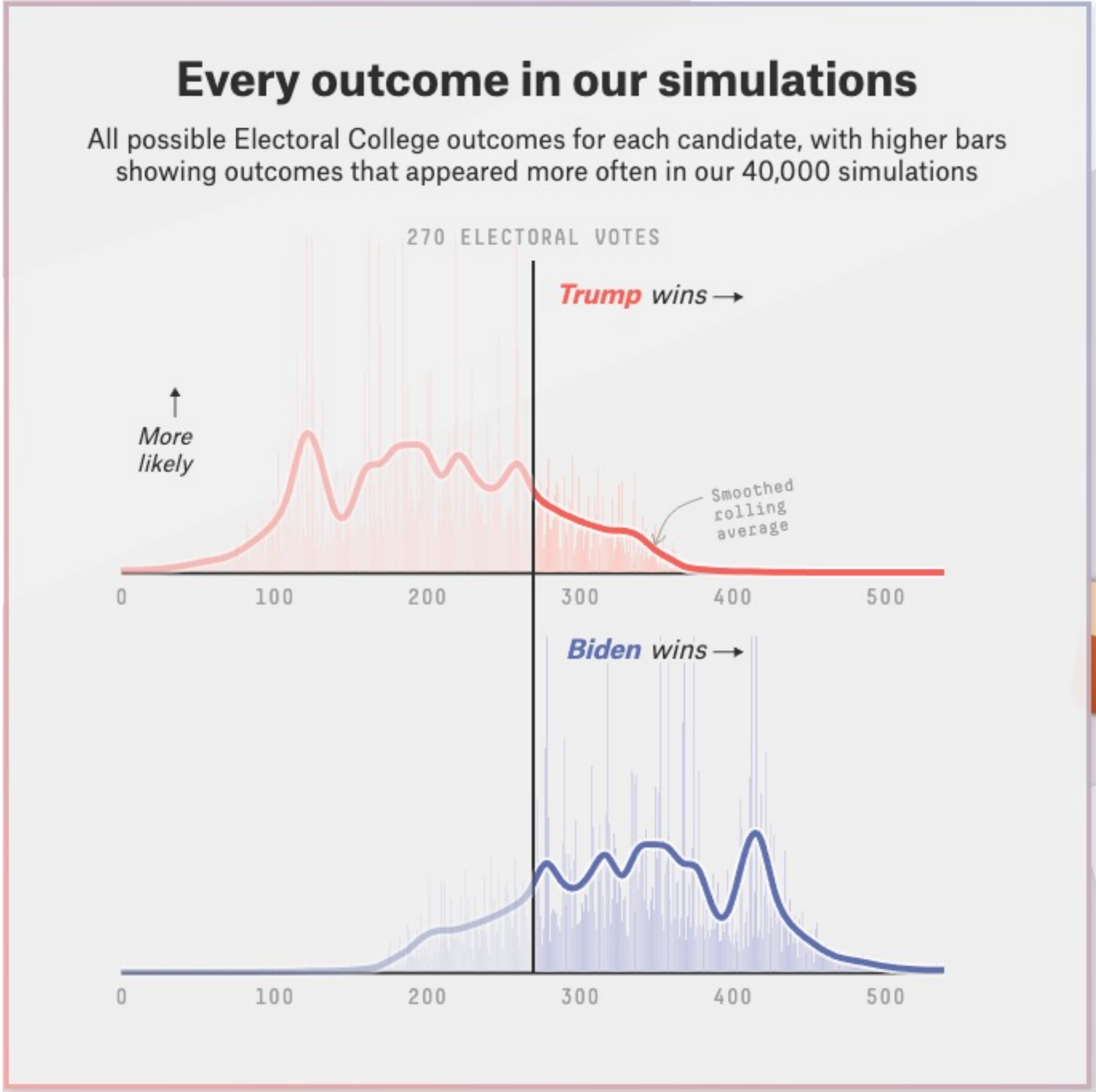


It’s all about the 538 Electoral College votes

Here's a map of the country, with each state sized by its number of electoral votes and shaded by the leading candidate's chance of winning it.

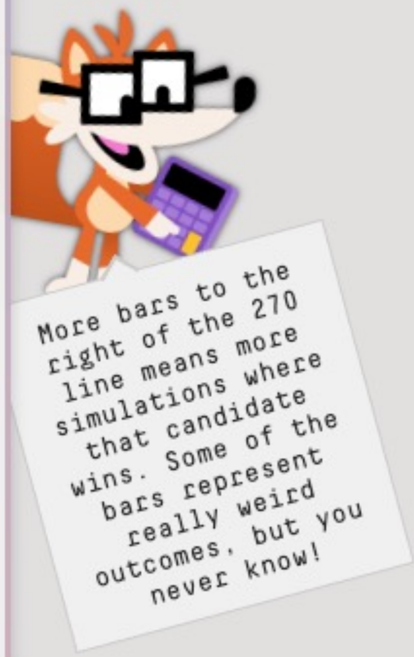
https://projects.fivethirtyeight.com/2016-election-forecast/

https://projects.fivethirtyeight.com/2020-election-forecast/

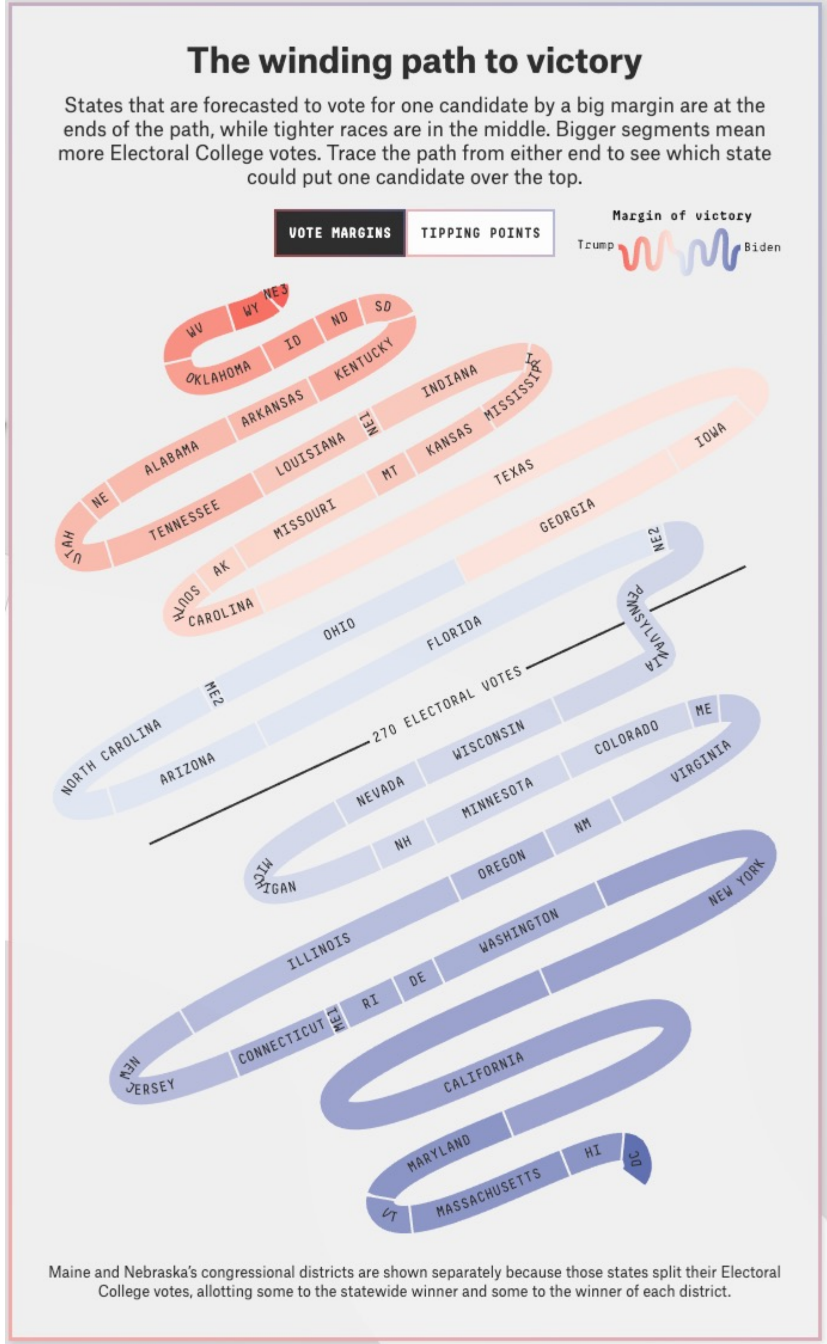


Every outcome in our simulations

All possible Electoral College outcomes for each candidate, with higher bars showing outcomes that appeared more often in our 40,000 simulations



More bars to the right of the 270 line means more simulations where that candidate wins. Some of the bars represent really weird outcomes, but you never know!



The winding path to victory

States that are forecasted to vote for one candidate by a big margin are at the ends of the path, while tighter races are in the middle. Bigger segments mean more Electoral College votes. Trace the path from either end to see which state could put one candidate over the top.

Maine and Nebraska's congressional districts are shown separately because those states split their Electoral College votes, allotting some to the statewide winner and some to the winner of each district.

https://projects.fivethirtyeight.com/2020-election-forecast/

Exercise time!