Lecture 2.8 – Basics of Data Visualization

Learning Objectives:

3.6.5 – Understand the importance of attention and accessibility to communicating ideas.

3.6.6 – Understand the importance of ethical presentation of data through visualization.

Converting Data to Understanding

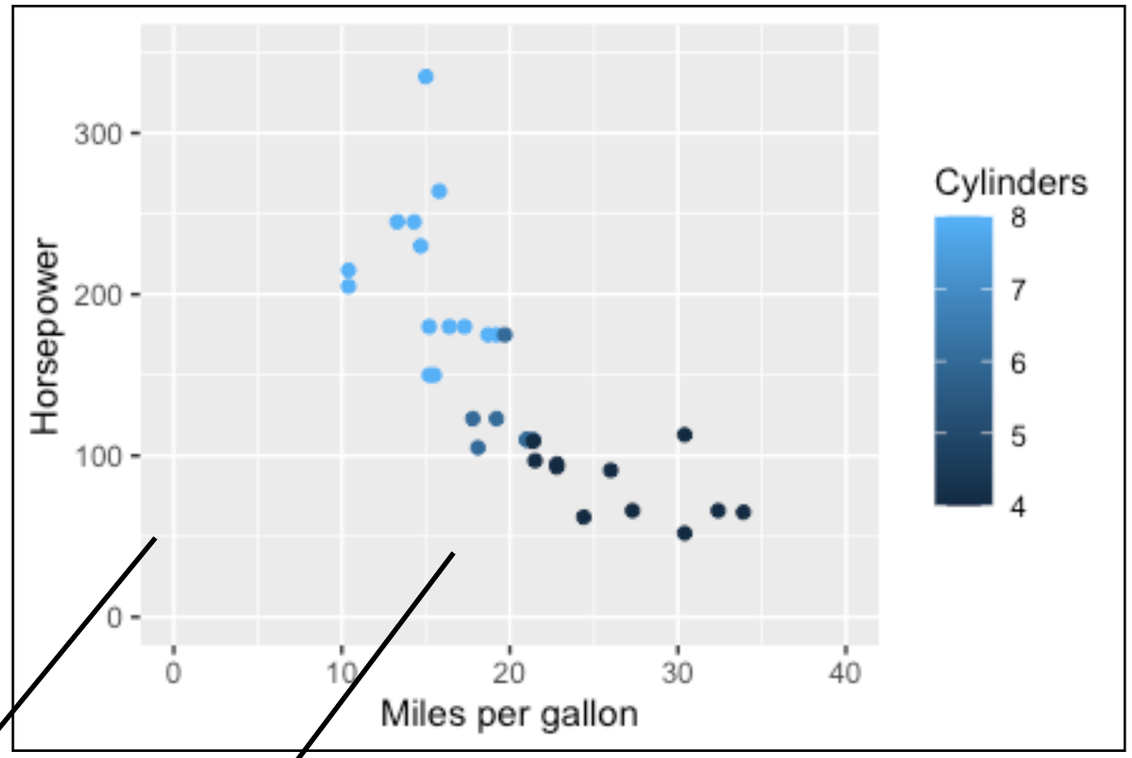
Data | **Mapping** | **Aesthetic attributes** | **Geometric object**

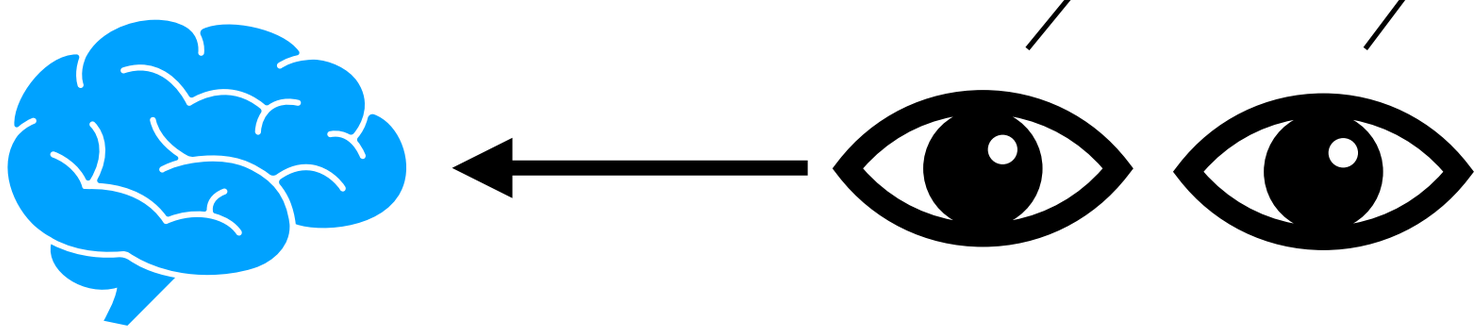
	mpg	cyl	disp	hp
Mazda RX4	21.0	6	160.0	110
Mazda RX4 Wag	21.0	6	160.0	110
Datsun 710	22.8	4	108.0	93
Hornet 4 Drive	21.4	6	258.0	110
Hornet Sportabout	18.7	8	360.0	175
Valiant	18.1	6	225.0	105
Duster 360	14.3	8	360.0	245
Merc 240D	24.4	4	146.7	62
Merc 230	22.8	4	140.8	95

mpg → x

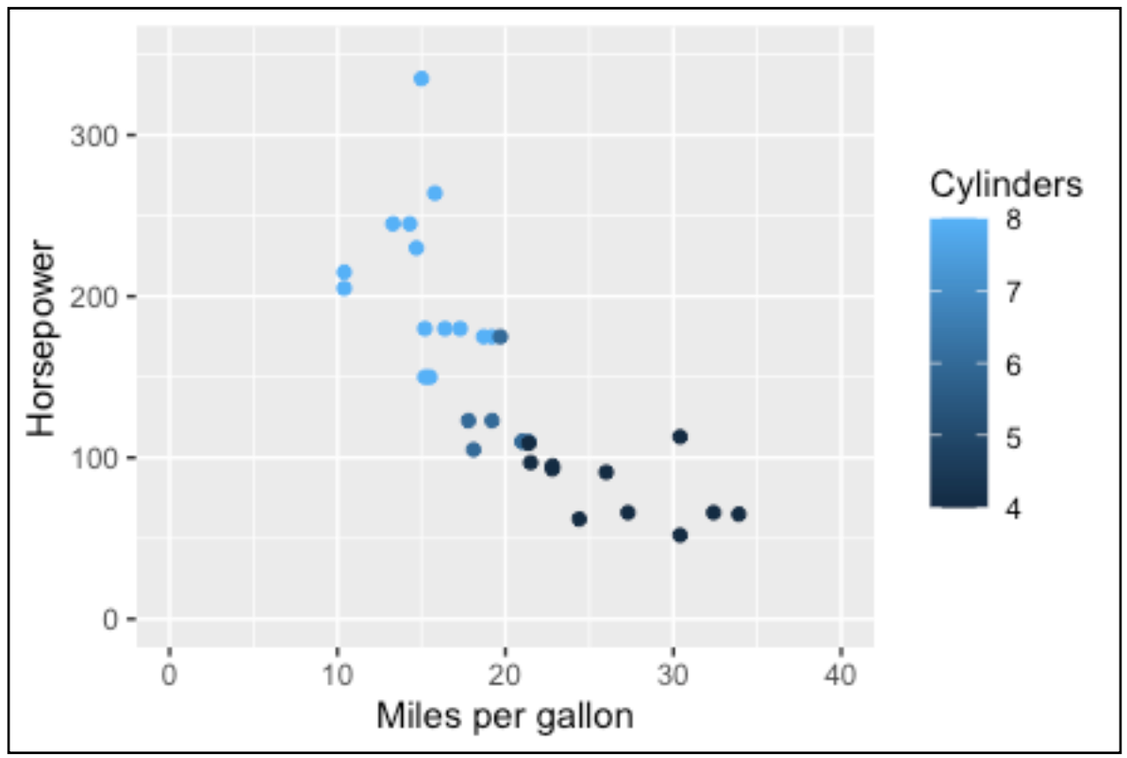
hp → y

cyl → color/fill





Converting Data to Understanding



1. Using guides to orient the viewer to the chart accurately.

2. Representing data faithfully (and not misleading your viewer)

3. Maximizing clarity and understanding using what we know about visual perception and processing

Converting Data to Understanding

Data

	mpg	cyl	disp	hp
Mazda RX4	21.0	6	160.0	110
Mazda RX4 Wag	21.0	6	160.0	110
Datsun 710	22.8	4	108.0	93
Hornet 4 Drive	21.4	6	258.0	110
Hornet Sportabout	18.7	8	360.0	175
Valiant	18.1	6	225.0	105
Duster 360	14.3	8	360.0	245
Merc 240D	24.4	4	146.7	62
Merc 230	22.8	4	140.8	95

Mapping

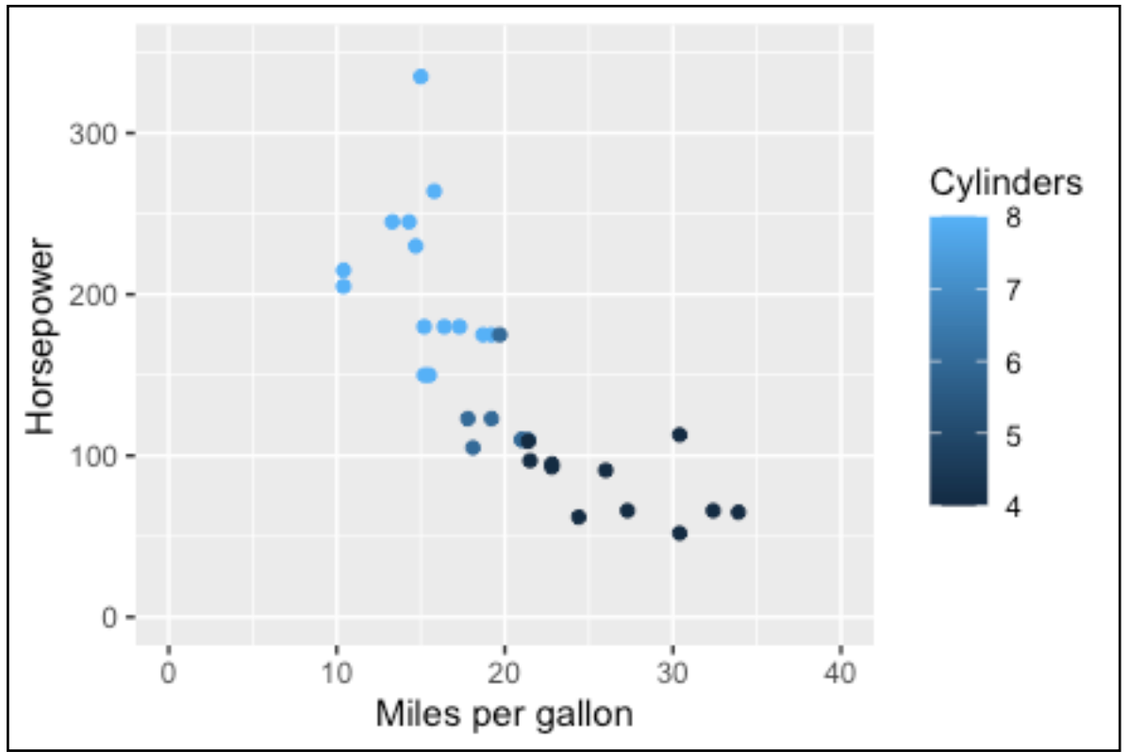
Aesthetic attributes

x

y

color/fill

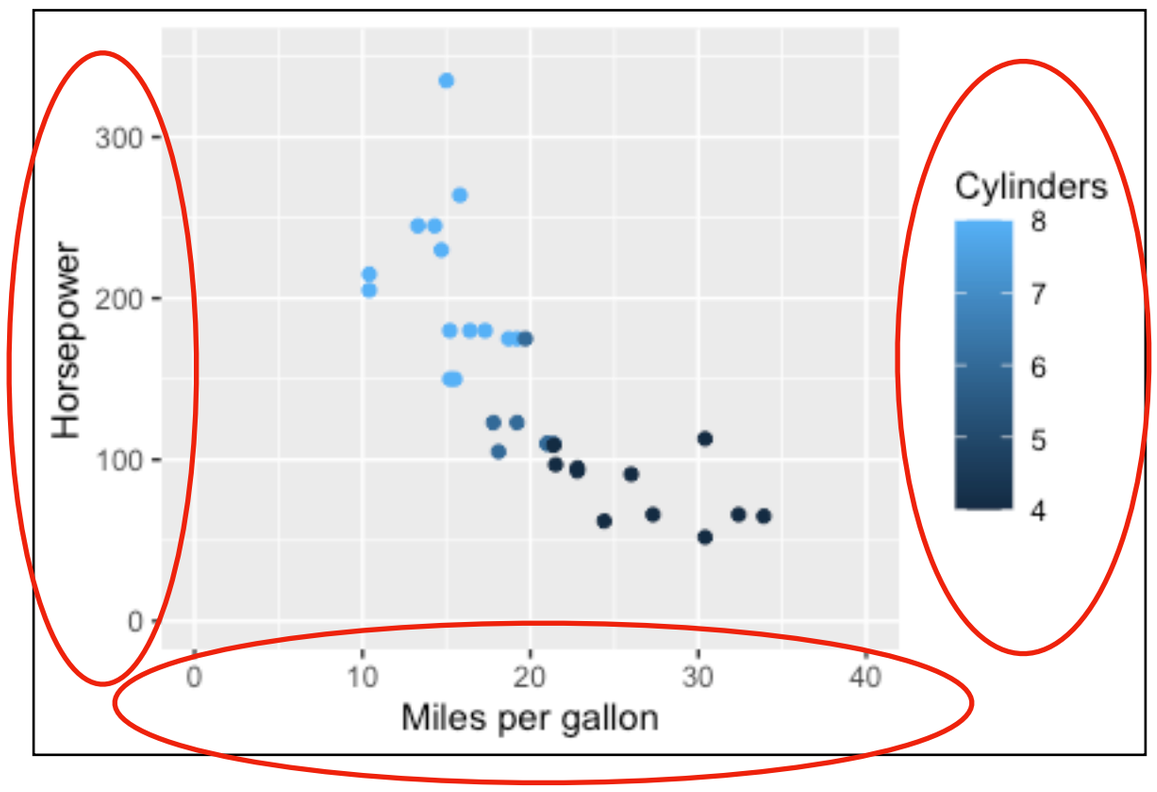
Geometric object



- Get the basics right! Make sure your audience can map elements back to the data using guides!

Visualization Guides

- Guides include any element that helps the viewer map elements back to data.



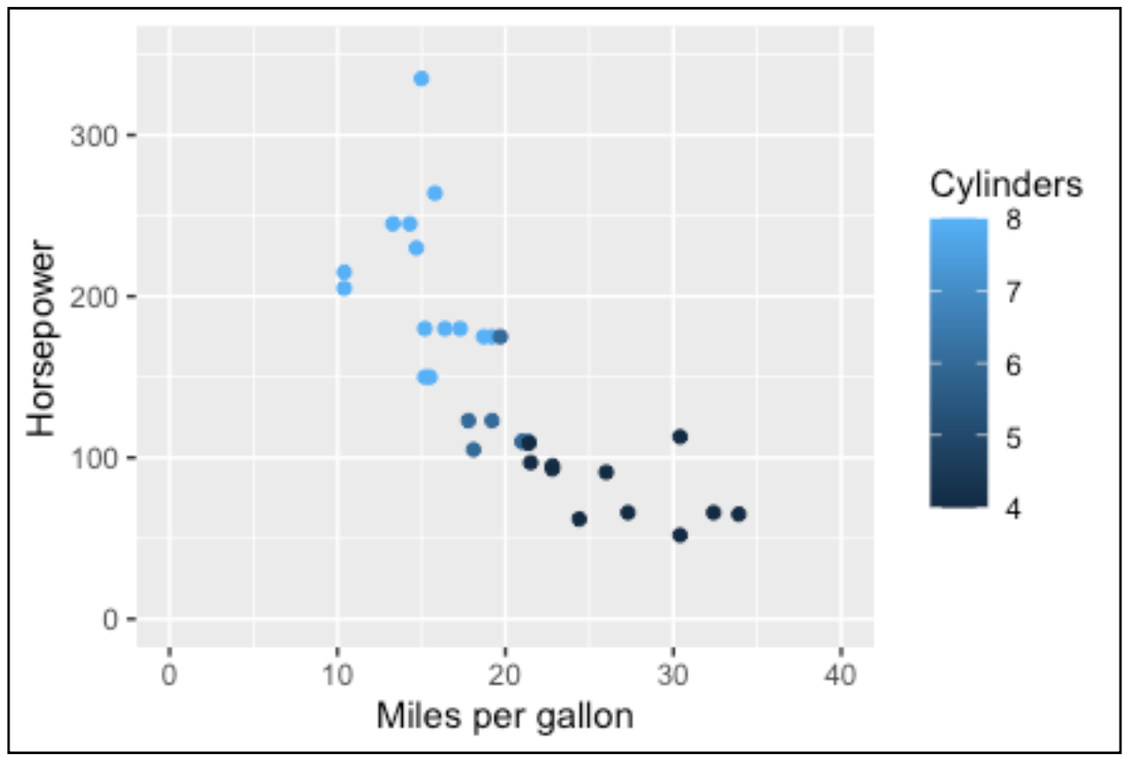
- Axis labels
- Legends
- Scales
- Guide lines
- Figure captions

Make sure they are accessible!

- Guides should be easy to read, clear, and descriptive, containing units where appropriate.

Make sure your guides are descriptive enough that a naïve viewer will understand what the elements represent!

Converting Data to Understanding



The Good, The Bad, The Wrong, and The Ugly

1. Using guides to orient the viewer to the chart accurately.
2. Representing data faithfully (and not misleading your viewer)
3. Maximizing clarity and accessibility using what we know about visual perception and processing

The Good, The Bad, The Wrong, and The Ugly

- **Wrong**: graphs that have problems related to math or scaling, figures that mislead or are deceiving.

    - Figures could be misleading on purpose (deception) or by accident.
    - Unacceptable everywhere and all the time.
    - **First and foremost: you must NOT mislead your viewers!**
    - If your results are so weak you need to resort to deception or near deception, just admit it's not that strong of a conclusion.

- **Bad**: graphs that have problems related to perceptual issues, unclear or confusing, or overly complicated.

    - Unacceptable in this class, you can do better!

- **Ugly**: graphs that are clear and informative but have aesthetic problems (like, it's just ugly).

    - These are not great but acceptable. (I would rather have an honest and ugly graphic than a beautiful but deceiving graphic.)
    - Acceptable but can be improved

WRONG

James Hamblin
@jameshamblin

Fox added a “%” where it doesn’t belong, decreasing fatality by 100x. Based on the CDC data it’s not that .054% of people over 70 don’t survive COVID. It’s 5.4%.

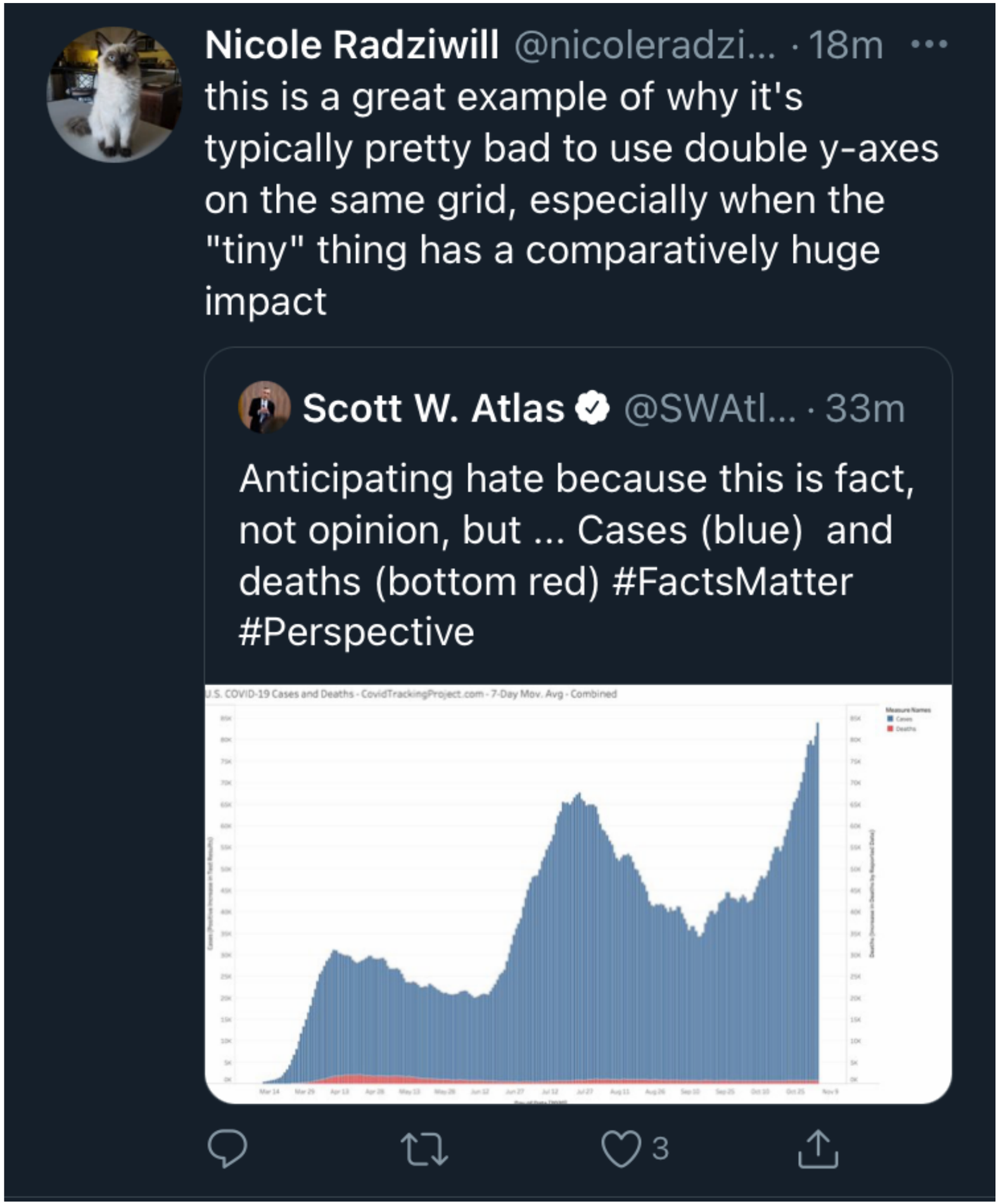
> **Zeducation** @RealZeducation · Sep 25
> Just a reminder that everything in 2020 was over a virus that does this:

6:49 PM · Sep 26, 2020 · Twitter for iPhone

594 Retweets **25** Quote Tweets **1.9K** Likes

- Problems with math.

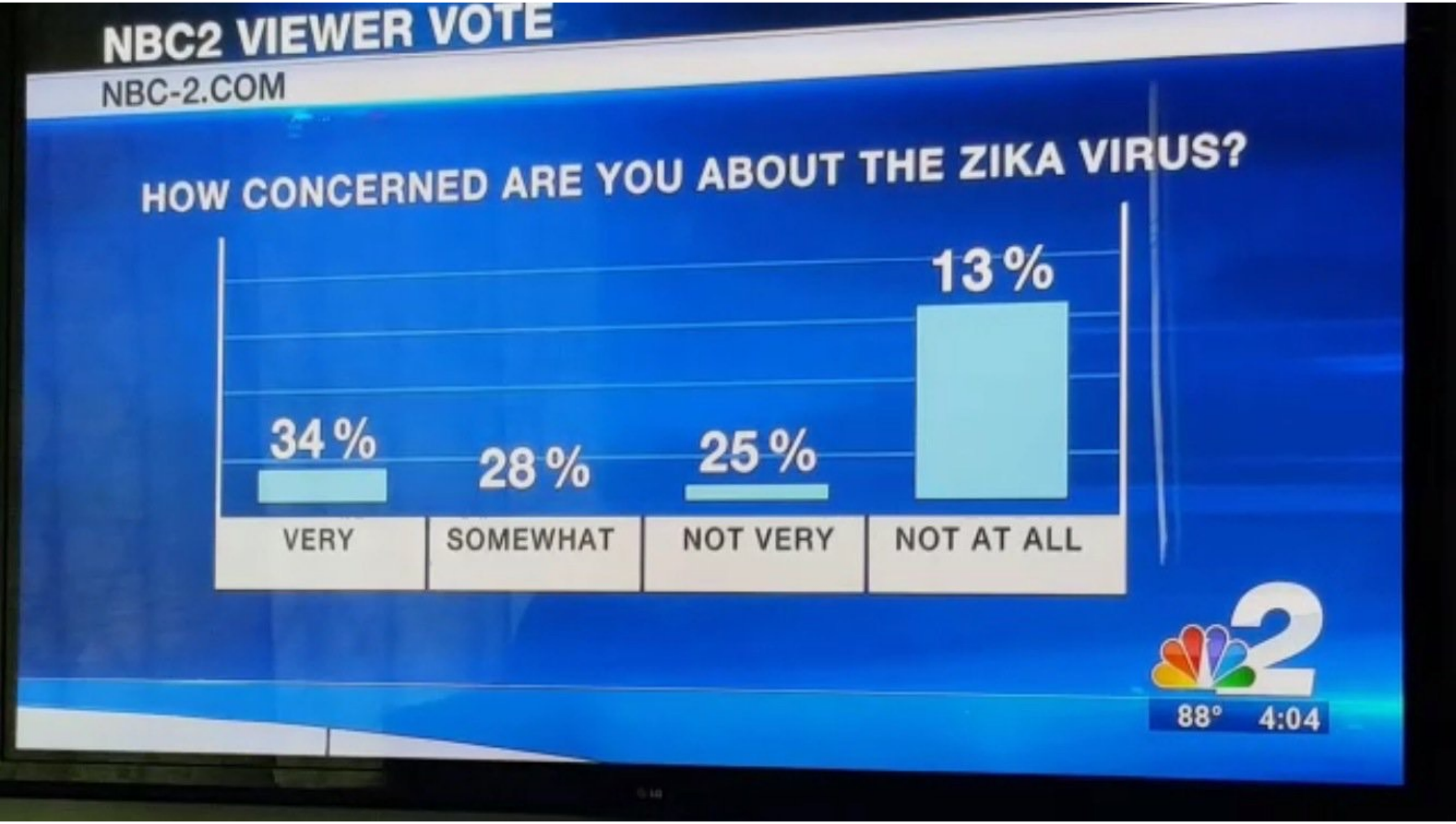
BAD/ WRONG



- Misleading axis scaling.

WRONG

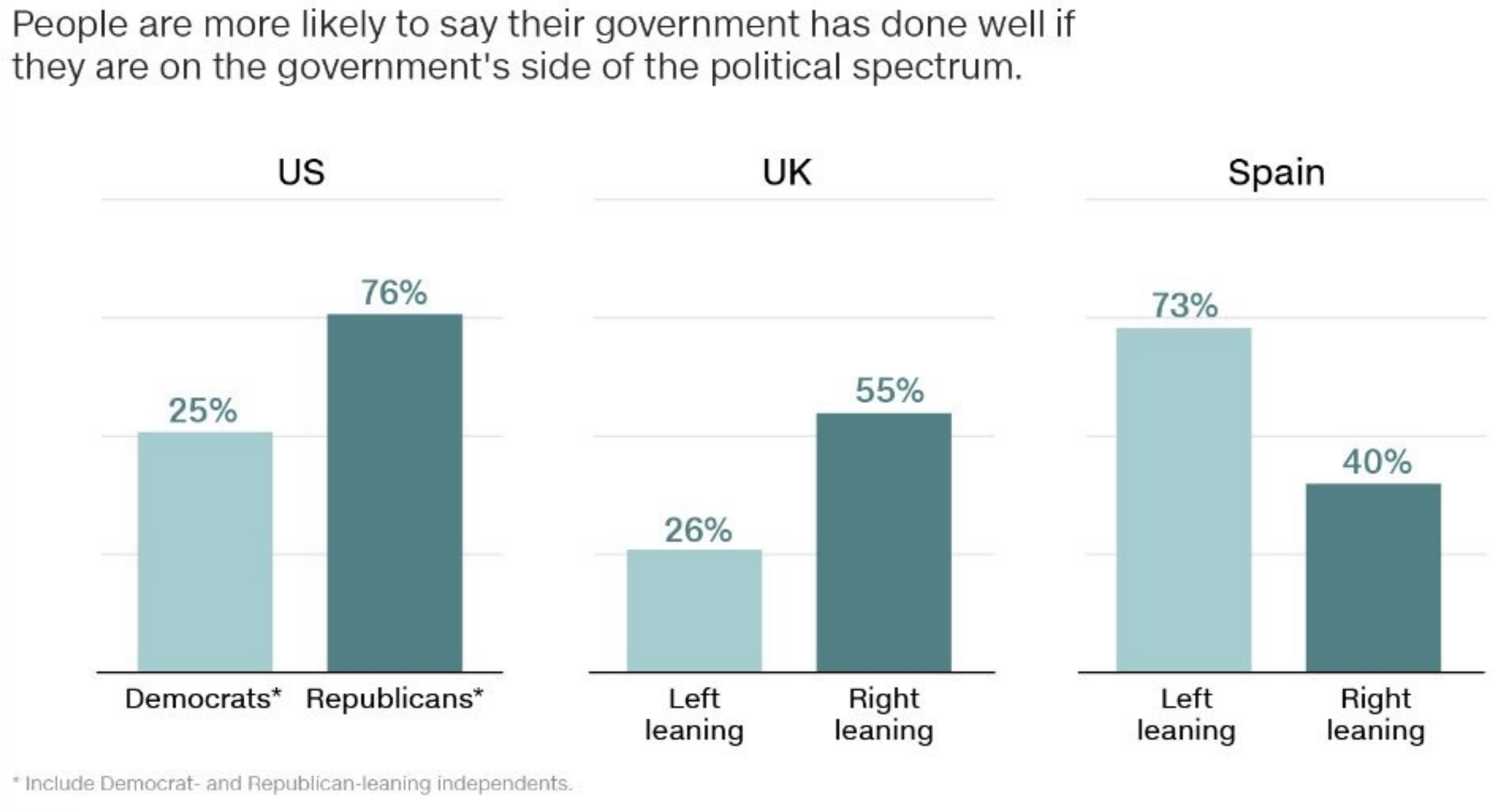
- Numbers don’t match size of bars.



Seriously, wtf?

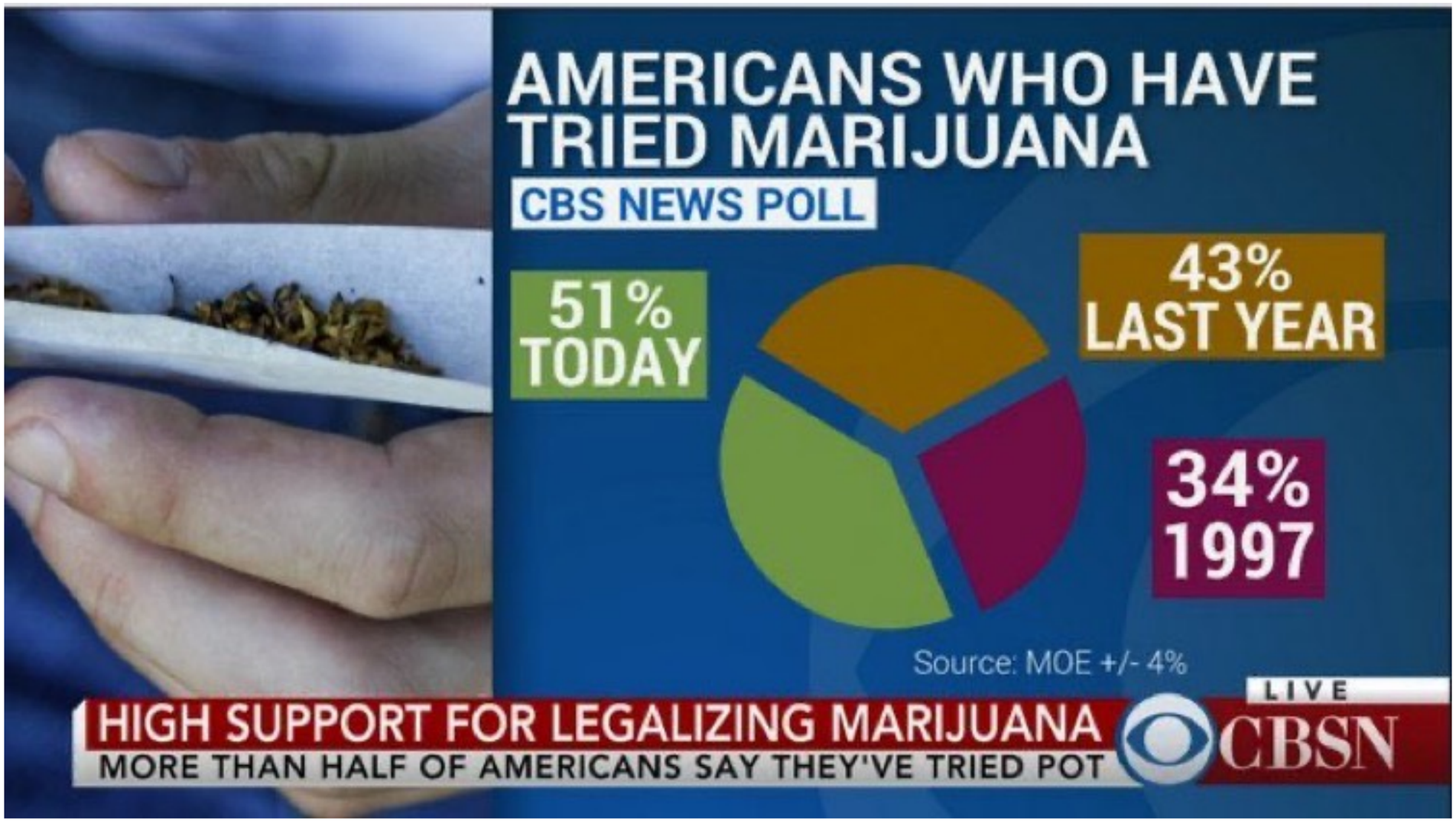
BAD

- Floating, unlabeled y axis, misleading.



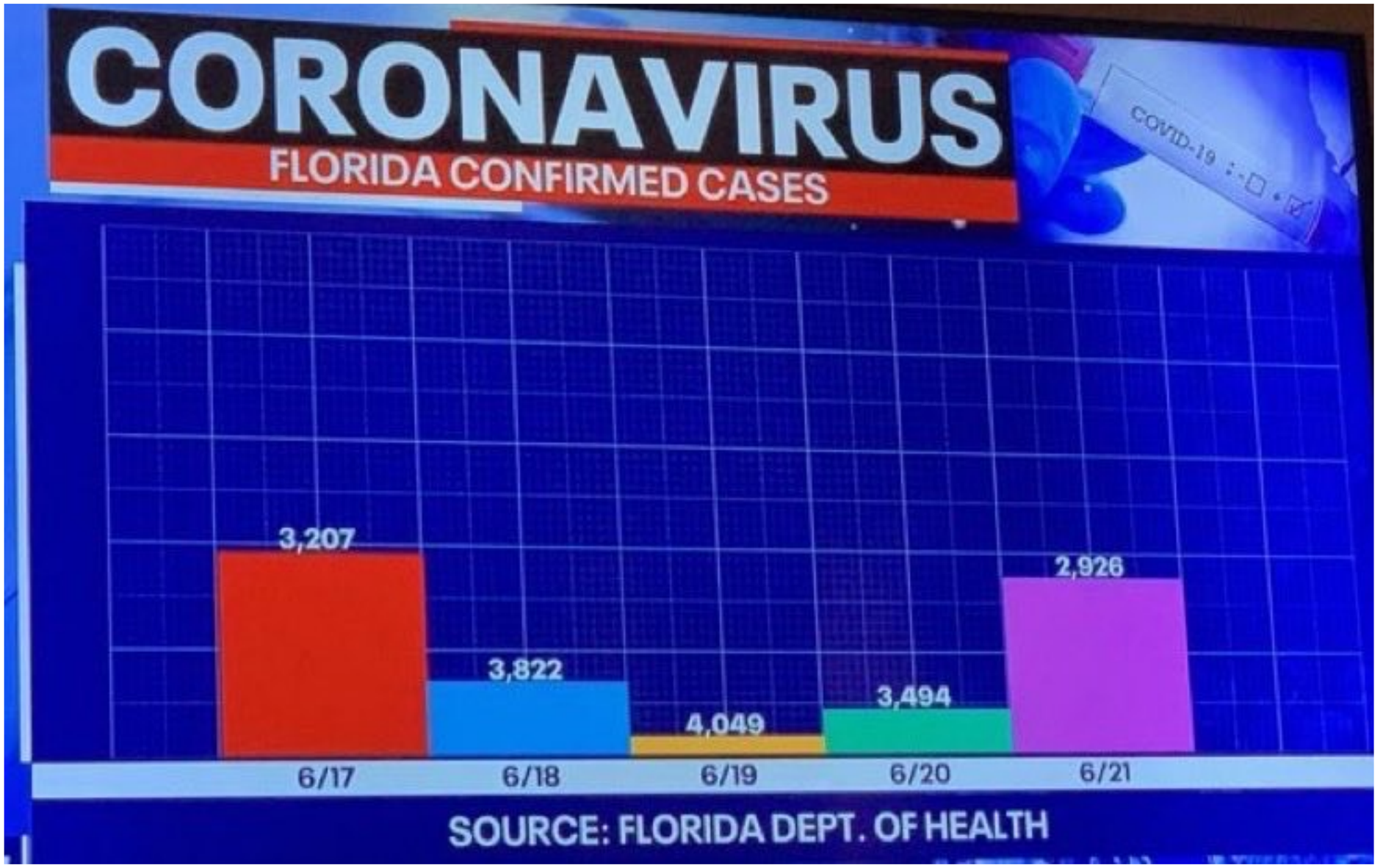
BAD

- Wrong chart type for data. (lolwut)

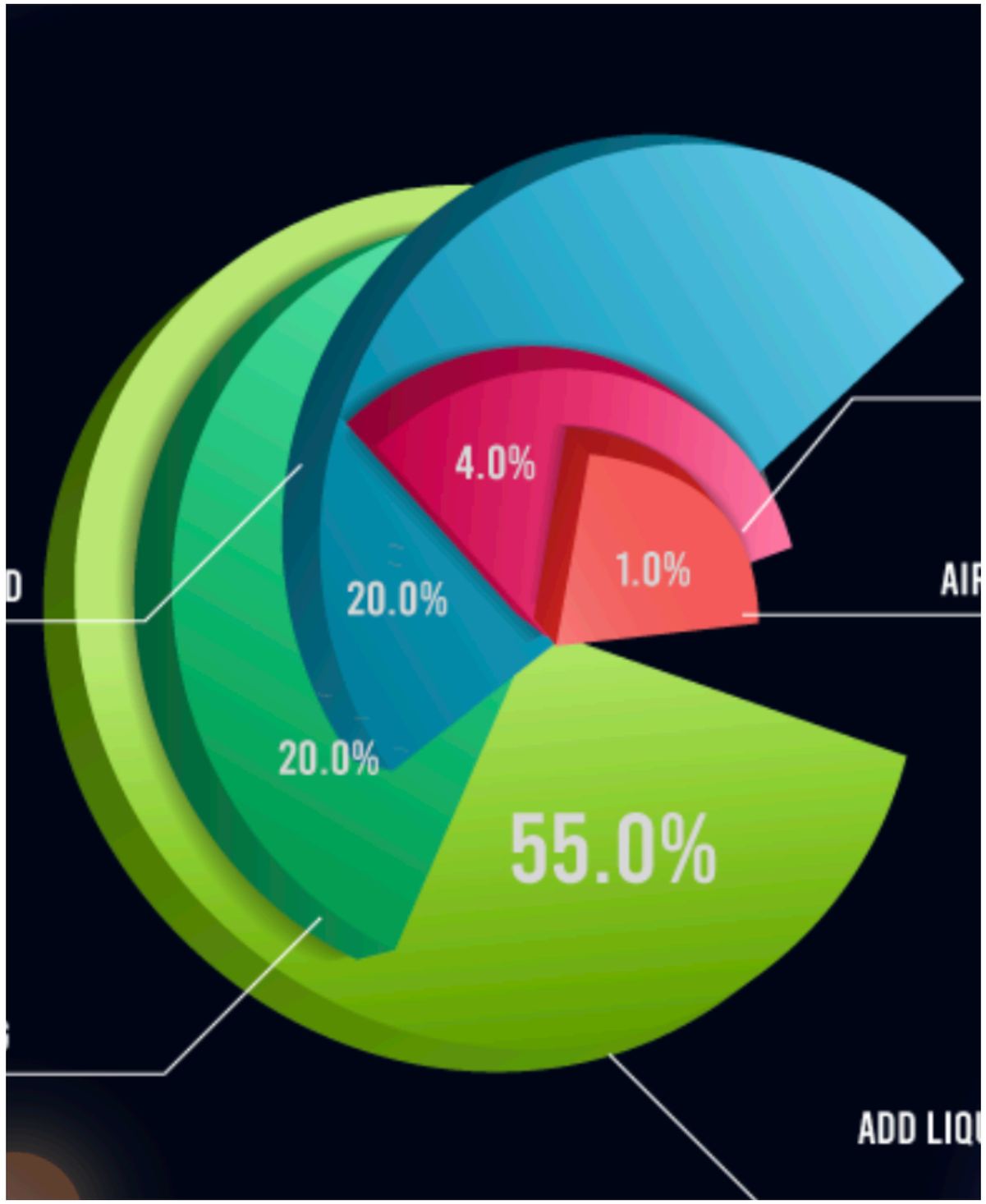


WRONG

😒

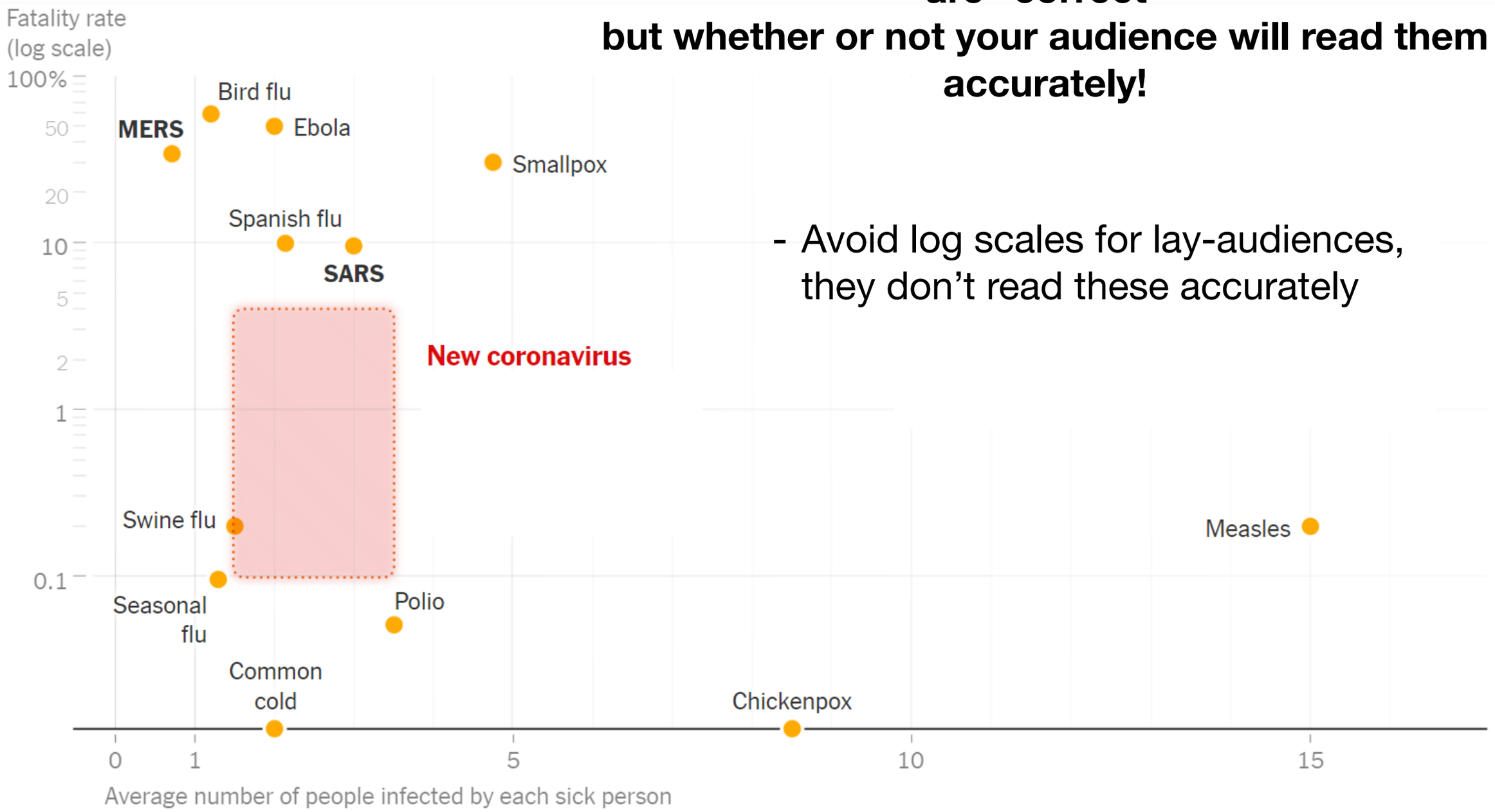

CORONAVIRUS
FLORIDA CONFIRMED CASES
COVID-19
3,207
3,822
4,049
3,494
2,926
6/17
6/18
6/19
6/20
6/21
SOURCE: FLORIDA DEPT. OF HEALTH

PLEASE STOP



BAD

You MUST take into account not only that things are “correct”
but whether or not your audience will read them accurately!



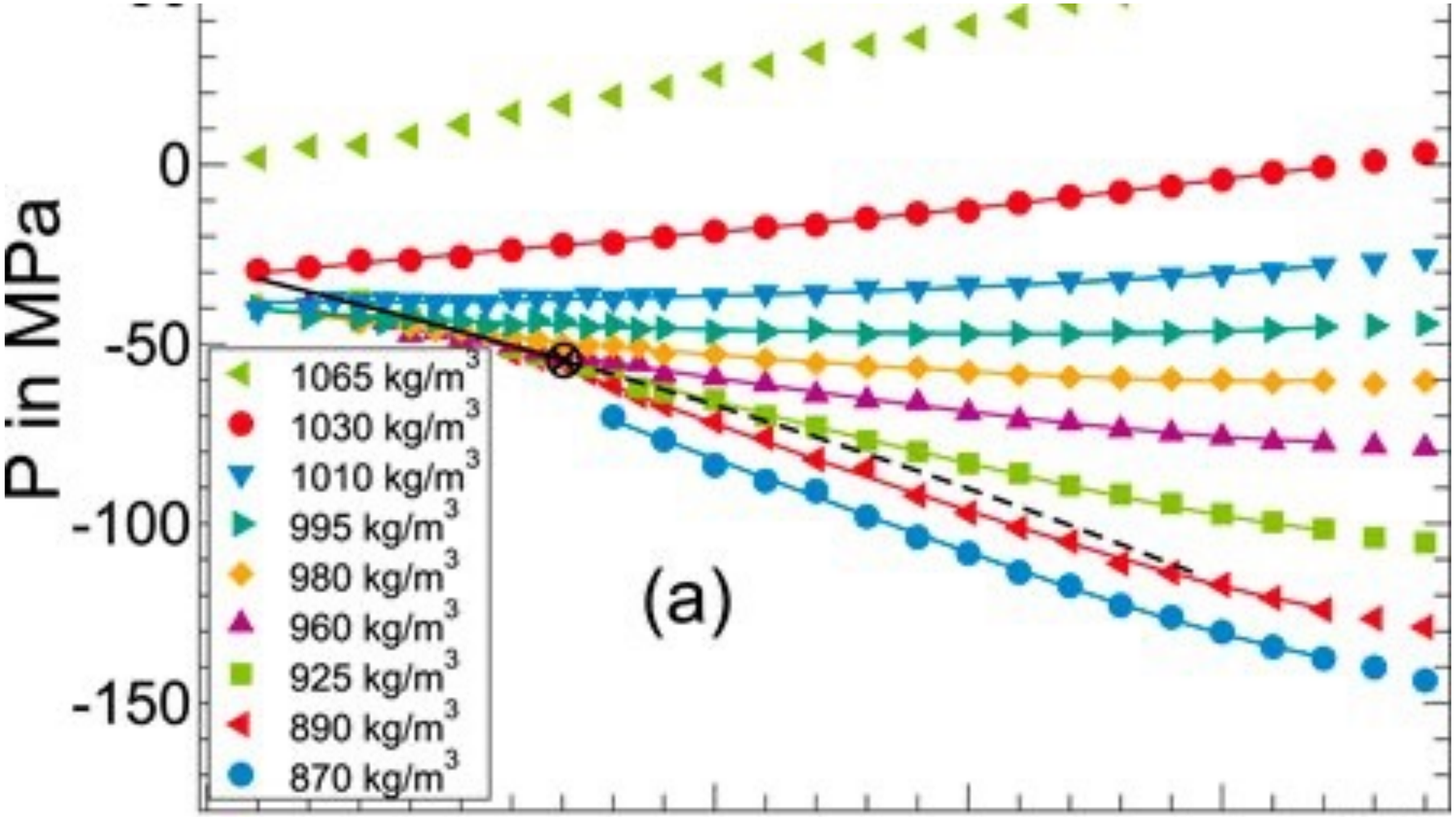
Note: Average case-fatality rates and transmission numbers are shown. Estimates of case-fatality rates can vary. **The preliminary estimates for the new coronavirus are shown in the pink region.**

- Avoid log scales for lay-audiences, they don’t read these accurately

New York Times https://www.nytimes.com/2020/02/18/learning/whats-going-on-in-this-graph-coronavirus-outbreak.html

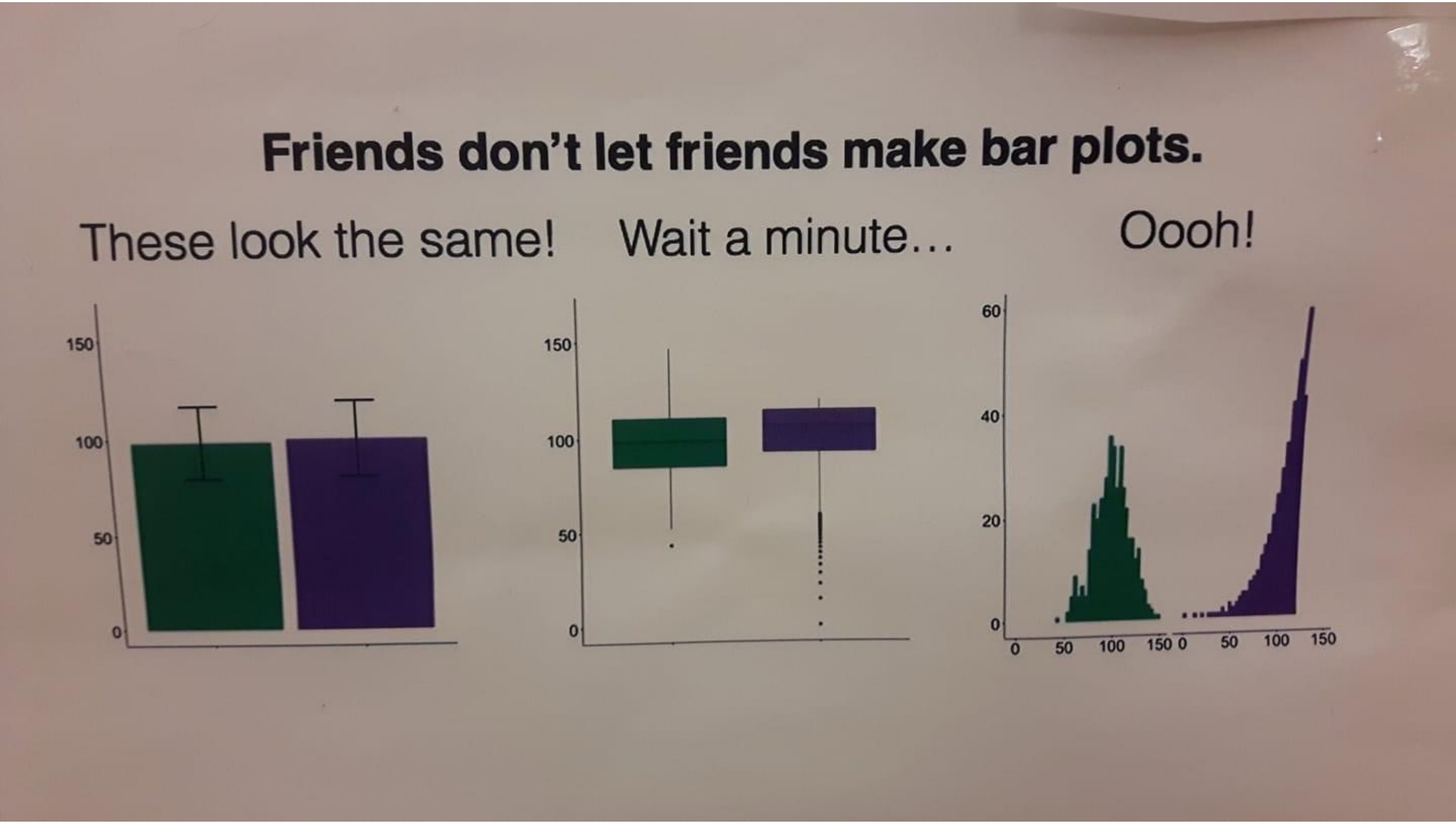
UGLY

- Some perceptual issues, could definitely be clearer.
- Places a large cognitive burden on viewer.



(X axis labels got cut off, my fault)

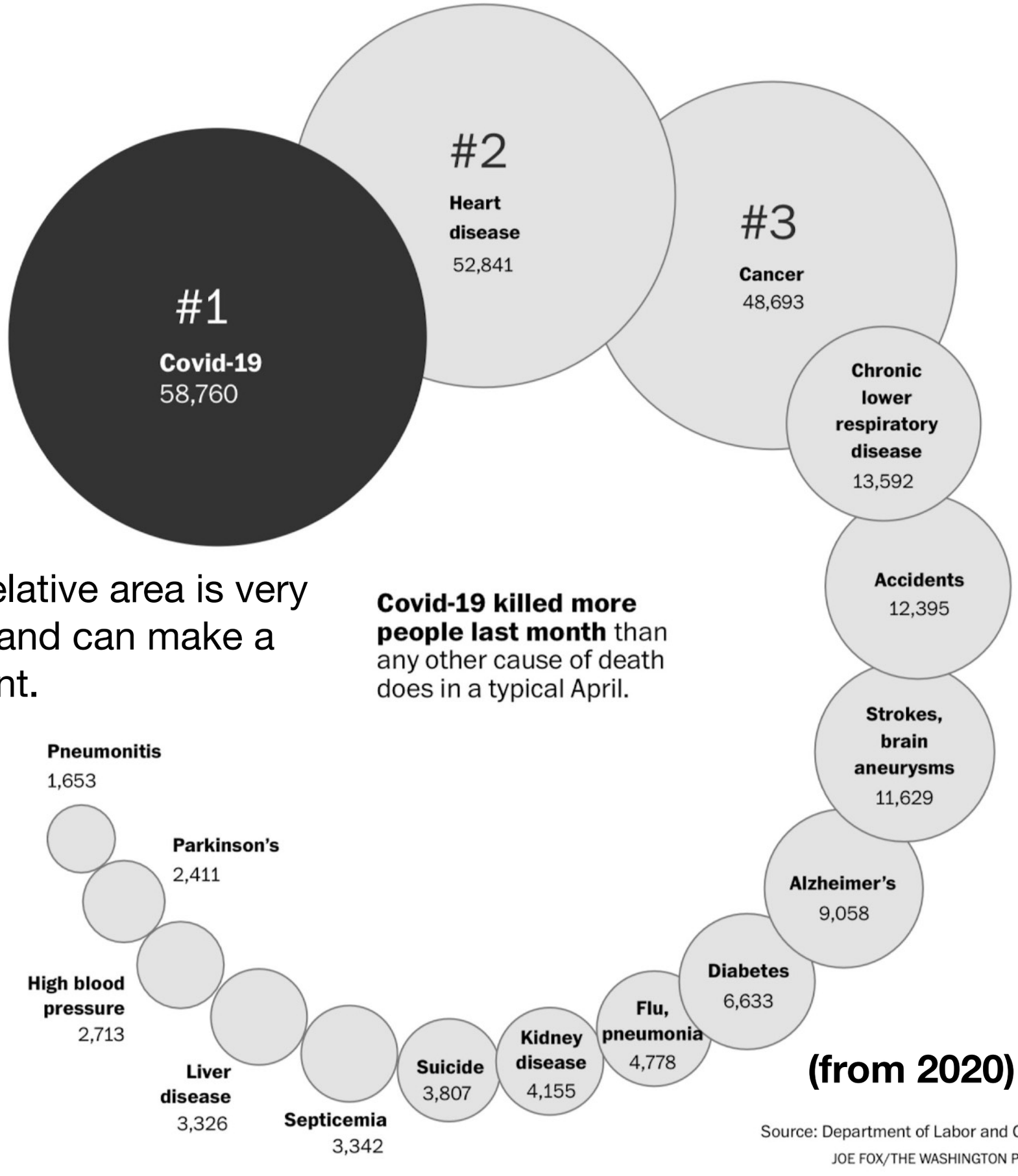
Bad to Good



- Your job is to represent your data accurately, with as much representation to the raw data as possible. Sometimes this means choosing geometric objects carefully!

GOOD

- Remember that relative area is very easily processed and can make a powerful statement.

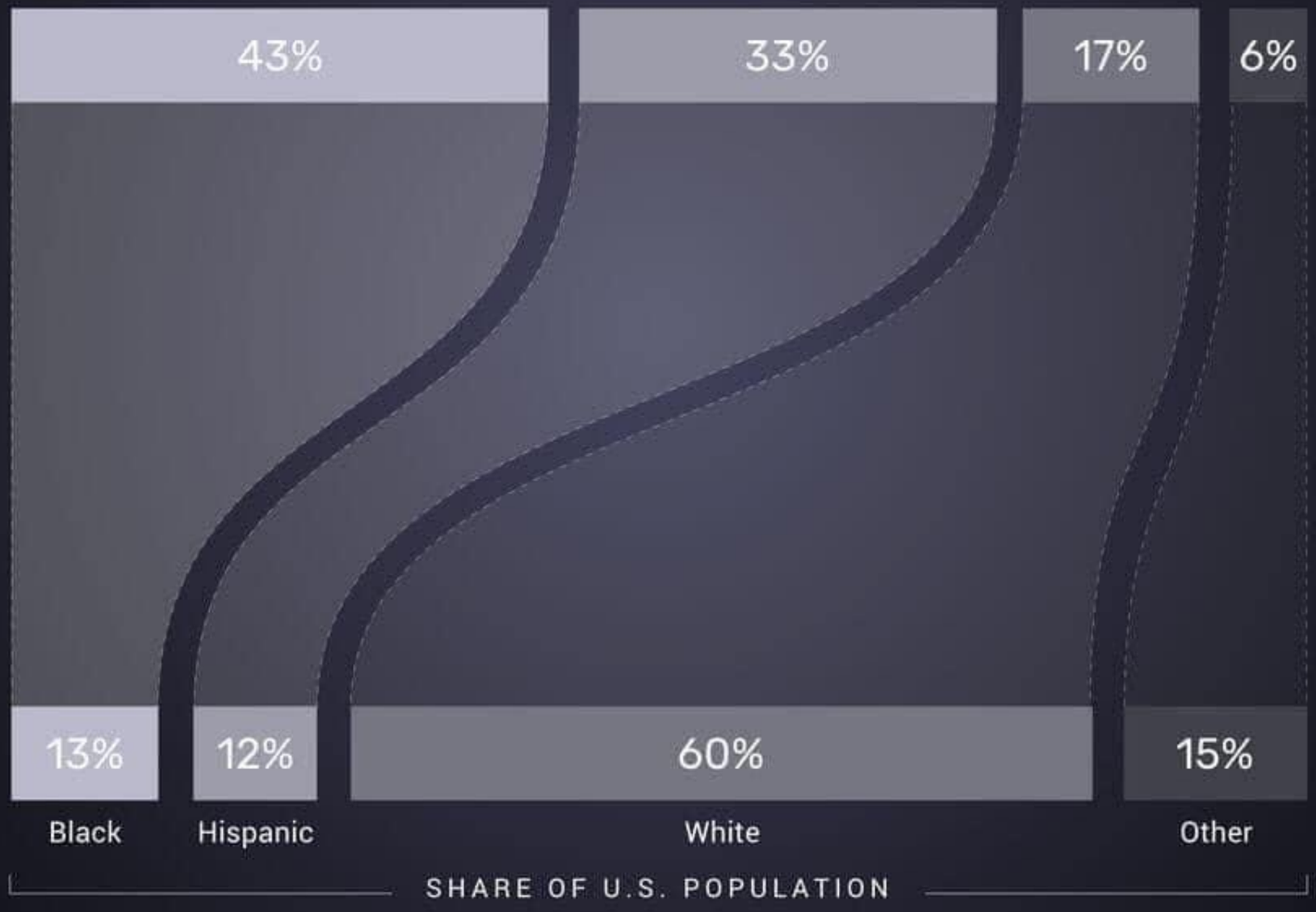


GOOD

U.S. POLICE SHOOTINGS

Fatal police shootings in the U.S. since January 01, 2015

Black Americans are disproportionately affected



 | Source: Washington Post - Updated 2nd June 2020

COVID-19 Visualization Wins and Losses

Brilliant

This is how you do Data Viz, folks.

"All the News That's Fit to Print"

The New York Times

VOL. CLXIX .... No. 58,645 © 2020 The New York Times Company FRIDAY, MARCH 27, 2020

JOB LOSSES SOAR; U.S. VIRUS CASES TOP WORLD

Since the coronavirus descended on Brooklyn Hospital Center three weeks ago, the staff has handled over 800 potential cases.

New Data Shows Staggering Toll of Outbreak

This article is by Ben Casselman, Patricia Cohen and Tiffany Hsu.

More than three million people filed for unemployment benefits last week, sending a collective shudder throughout the economy that is unlike anything Americans have experienced.

The alarming numbers, in a report released by the Labor Department on Thursday, provide some of the first hard data on the economic toll of the coronavirus pandemic, which has shut down whole swaths of American life faster than government statistics can keep track.

Just three weeks ago, barely 200,000 people applied for jobless benefits, a historically low number. In the half-century that the government has tracked applications, the worst week ever, with 695,000 so-called initial claims, had been in 1982.

Thursday's figure of nearly 3.3 million set a grim record. "A large part of the economy just collapsed," said Ben Herzon, executive director of IHS Markit, a business data and analytics firm.

The numbers provided only the first hint of the economic cataclysm in progress. Even comparatively optimistic forecasters expect millions more lost jobs, and with them foreclosures, evictions and bankruptcies. Thousands of businesses have closed in response to the pandemic, and many will never reopen. Some economists say the decline in gross domestic product this year could rival the worst years of the Great Depression.

And there was fresh evidence on Thursday of the relentless course of the virus itself. Cases in the United States now exceed 80,000, the most of any nation, even China and Italy, according to a New York Times database. More than 1,000 deaths across the country have been linked to the virus.

At least 160 million people na-

Continued on Page A17

Labeling Regions by Risk

President Trump told governors he planned to classify counties according to the danger of coronavirus infection. Page A11.

Unfilled Posts Force Scramble By Washington

By JENNIFER STEINHAUER and ZOLAN KANNO-YOUNGS

WASHINGTON — Of the 75 senior positions at the Department of Homeland Security, 20 are either vacant or filled by acting officials, including Chad F. Wolf, the acting secretary who recently was unable to tell a Senate committee how many respirators and protective face masks were available in the United States.

The National Park Service, which like many federal agencies is full of vacancies in key posts, tried this week to fill the job of a director for the national capital region after hordes of visitors flocked to see the cherry blossoms near the National Mall, creating a potential public health hazard as the coronavirus continues to spread.

At the Department of Veterans Affairs, workers are scrambling to order medical supplies on Amazon after its leaders, lacking experience in disaster responses, failed to prepare for the onslaught of patients at its medical centers.

Ever since President Trump came into office, a record high turnover and unfilled jobs has emptied offices across wide swaths of the federal bureaucracy. Now, current and former administration officials and disaster experts say the coronavirus has exposed those failings as never be-

Continued on Page A13

Online Class With No Way To Get There

By NIKITA STEWART

Allia Phillips was excited about picking up an iPad from her school in Harlem last week. She did not want to miss any classes and hoped to land on the fourth-grade honor roll again.

On Monday, the first day that New York City public schools began remote learning, the 10-year-old placed her iPad on a tray she set up over her pillow on a twin bed in a studio that she shares with her mother and grandmother inside a homeless shelter on the Upper West Side.

And then, Allia saw nothing.

"I went downstairs to find out that they don't have any internet," said Kasha Phillips-Lewis, Allia's mother. "You're screwing up my daughter's education. You want to screw me up? Fine. But not my daughter's education."

The Department of Education, which runs the largest school system in the country with more than 1.1 million students, began attempting to teach all students through remote learning this week because schools were closed to slow the spread of the coronavirus.

Shuttering the vast system, which includes 1,800 schools, was a serious challenge for the city, and the large-scale, indefinite school closures are uncharted territory, altering the lives and rou-

Continued on Page A16

Courage at a Brooklyn Hospital, At the Front of an Invisible War

By SHERI FINK

It was not even 9 in the morning and Dr. Sylvie de Souza's green N95 mask, which was supposed to form a seal against her face, was already askew.

In freezing rain on Monday, she trudged in clogs between the emergency department she chairs at the Brooklyn Hospital Center and a tent outside, keeping a sharp eye on the trauma doctors, nurses and other staff members who would screen nearly 100 walk-in patients for the coronavirus that day.

Inside her E.R., more than a dozen people showing signs of infection waited for evaluation in an area used just a few weeks ago for stitches and casts. Another dozen

With Supplies Waning, Workers Soldier On

lay on gurneys arranged one in front of the next, like a New York City car park. One man on a ventilator was waiting for space in the intensive care unit.

Minutes before paramedics wheeled in a heart attack patient, Dr. de Souza pointed to beds reserved for serious emergencies, separated by a newly constructed wall from the suspected virus cases. "This is our safe area," she told a reporter. Then she corrected herself: "This is thought to be safe." There was really no way

Continued on Page A14

No Crowd, but I'll Take You Out to the Ballgame

By DAN BARRY

Professional baseball greeted a new season this afternoon with an Opening Day game for the ages, an extra-inning masterpiece that vividly unfolded on the sun-dappled field of the imagination. The crack of the bat could almost be heard, the blur of white almost seen, the communal joy nearly felt.

A Fan Writes a Fantasy for Opening Day

From the moment the first batter tipped his helmet — and a bird flew out — to the walk-off home run by a slathering peach-hitter, this 11-inning affair redefined what constitutes a perfect game. No one cared about the outcome; the distraction was reward enough.

Don't misunderstand. This game between the New York Gothams and the Cincinnati Greens mattered, but in ineffable ways beyond the columns of wins and losers. It mattered so much that complaints about baseball's slow pace yielded to the universal wish that this game would last forever.

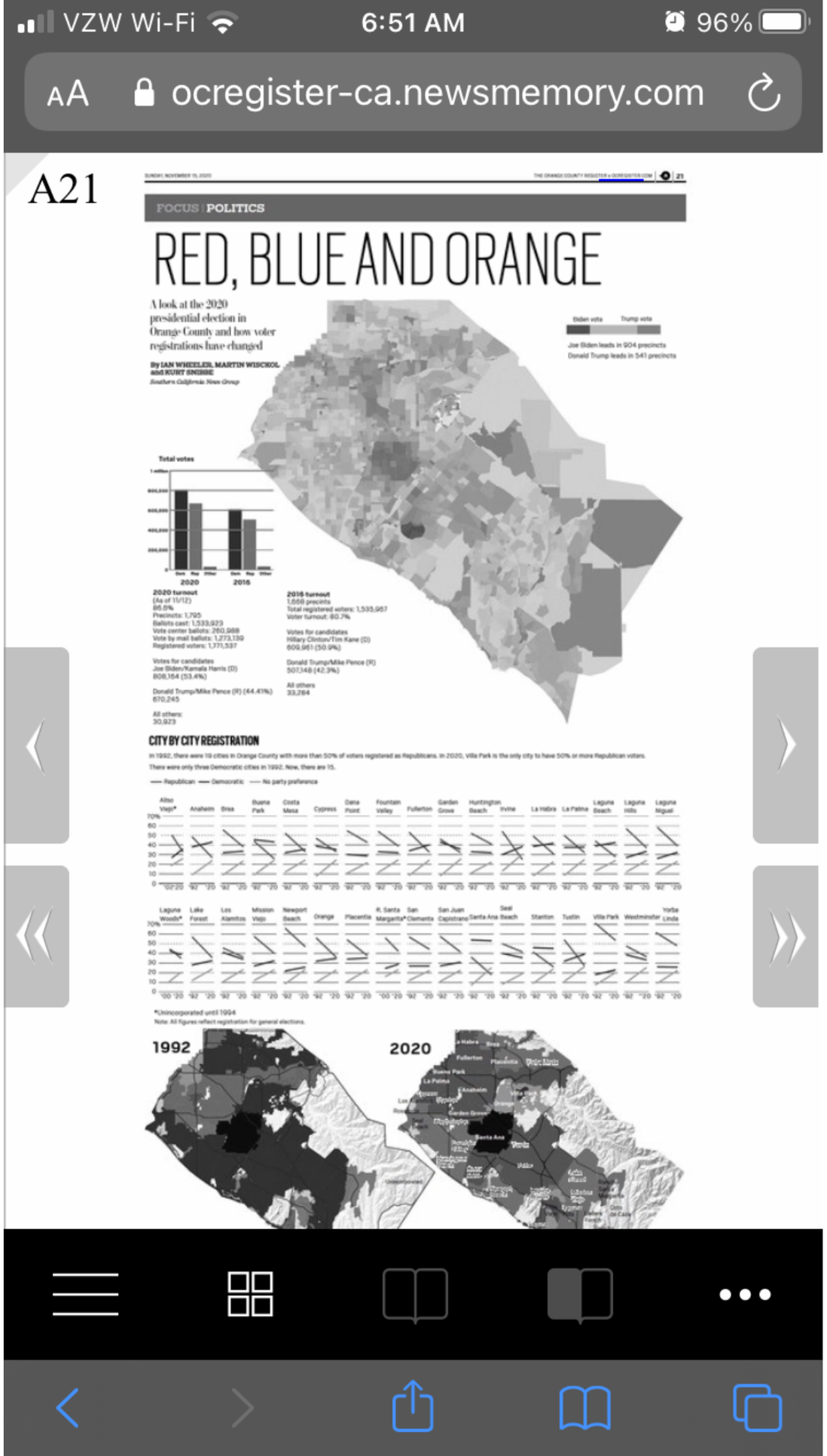
"I could've played into the night," said the redeemed Gothams left fielder Sammy Iona, who missed part of last season after sneezing so hard that he strained a ligament in his back. "I didn't

Continued on Page A13

Nearly 3.3 million unemployment claims were filed last week, a record number.

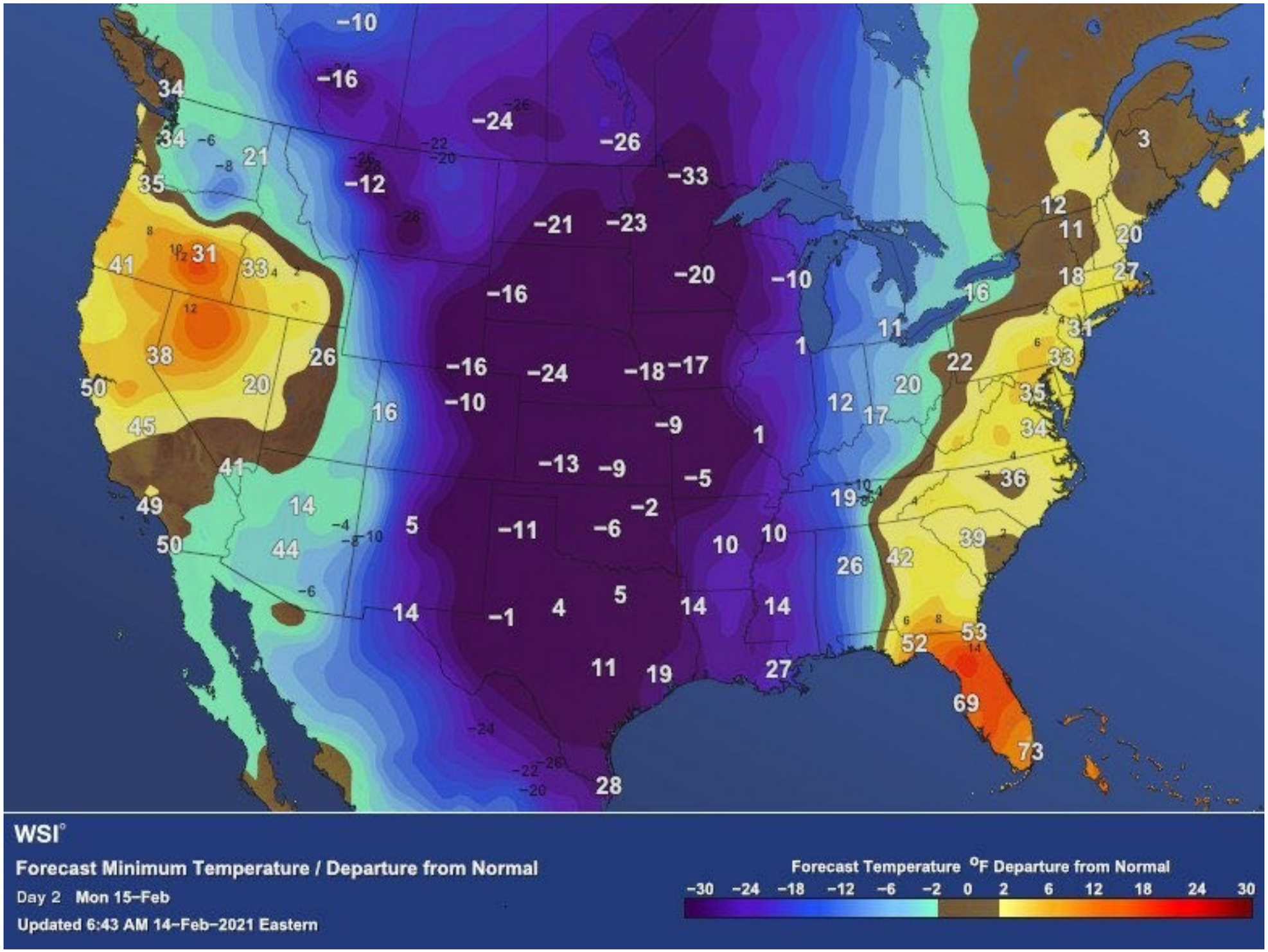
You decide

Orange county voting results in 2020 presidential election (exactly as they appeared on my phone)



Check Your Understanding

Discuss one good thing and one thing that is confusing about this graphic.



How do you move your work from “ugly” to “good?”

- Maybe you have a graph that is ugly or hard to read (but honest), how can you move it into that ***really good***, ***powerful*** category?

  - It has much more to do with storytelling than it does coding!

  - The point of data visualization is to communicate ideas about data to your audience.

  - Figures should be much more than random plots of the data you have.

  - ***Think carefully*** about ***what the point is*** and what idea you want to communicate. **Let these guide your design!**

OCAR Storytelling

- Form figures around OCAR storytelling, basic structure:

  - **Opening**: who are the players? What metrics are on the x and y axes? What values are being shown?

  - **Challenge**: what is the question these data are trying to answer by showing the reader relationships?

  - **Action**: How do the data need to interact in order to show you the answer to the challenge?

  - **Resolution**: What does the relationship mean? Spell it out for the audience either in the caption or by speaking.

What are the OCAR elements to this graphic?

Opening:

Share of deaths, share of population, race

Challenge:

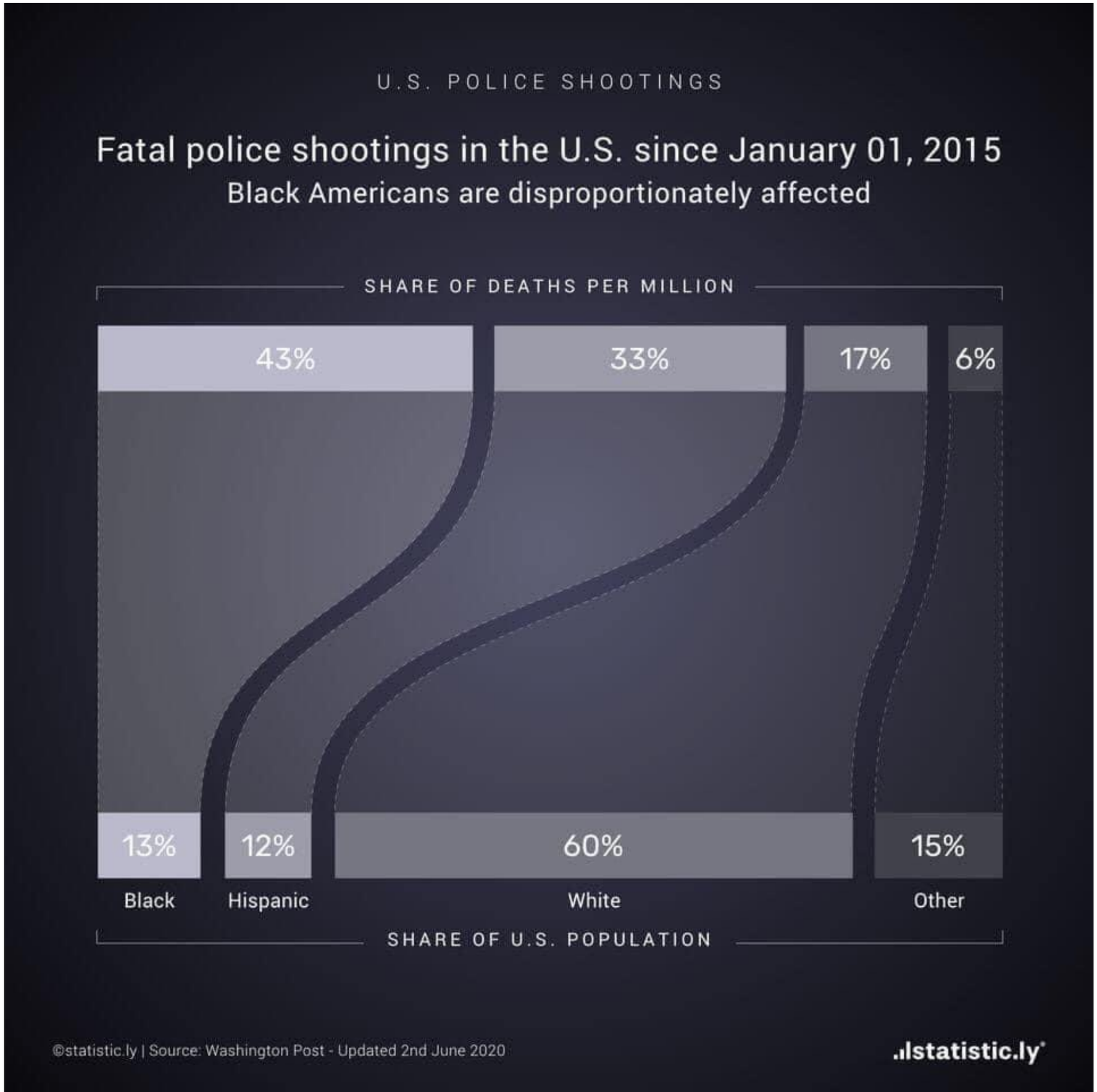
How many black people are killed by police compared to whites?

Action:

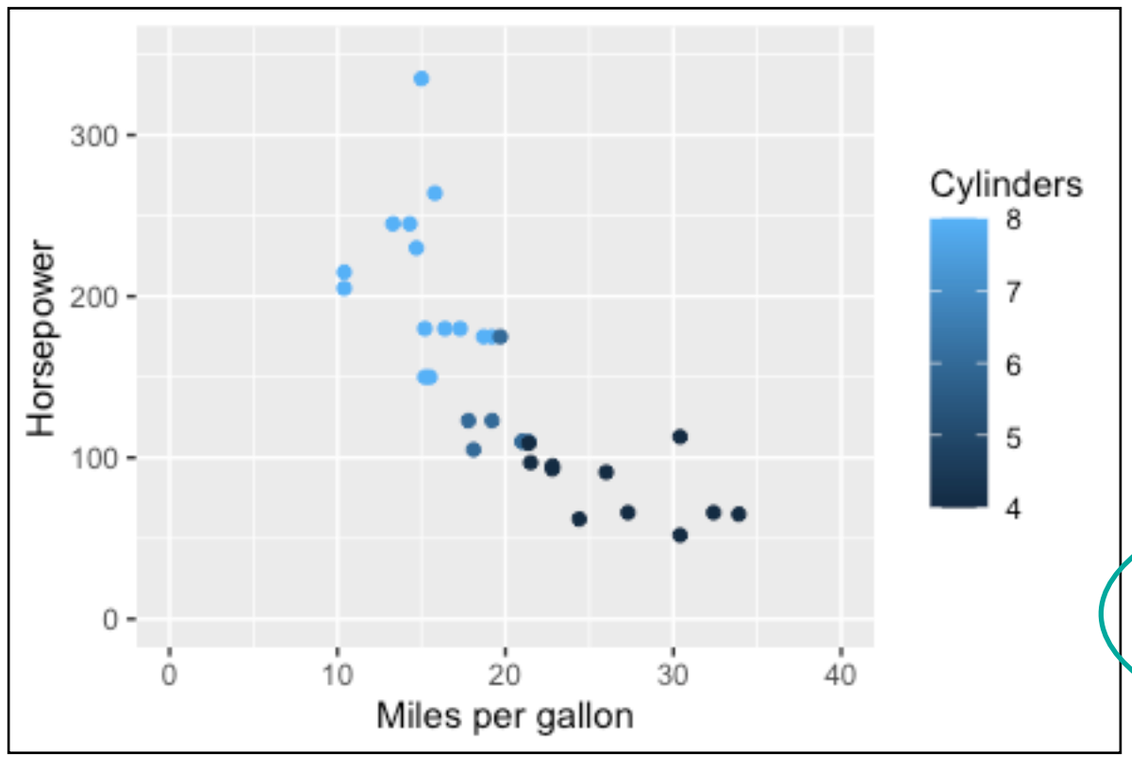
Flow of proportionality between two scales.

Resolution:

Black people are killed disproportionately more than whites.



Converting Data to Understanding



1. Using guides to orient the viewer to the chart accurately.

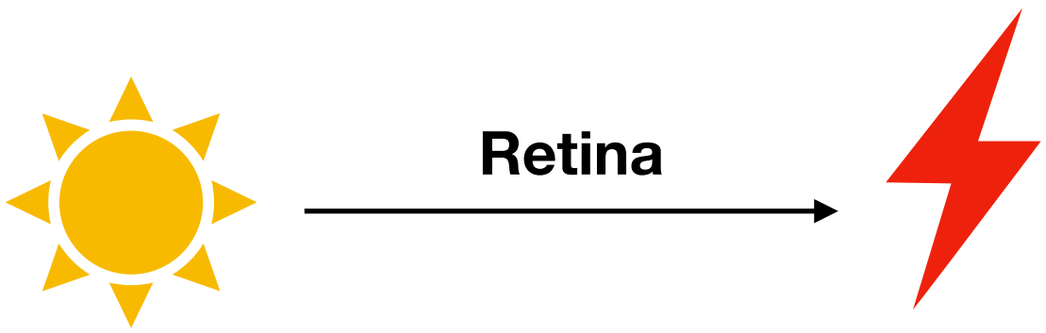
2. Representing data faithfully (and not misleading your viewer)

3. Maximizing clarity and understanding using what we know about visual perception and processing

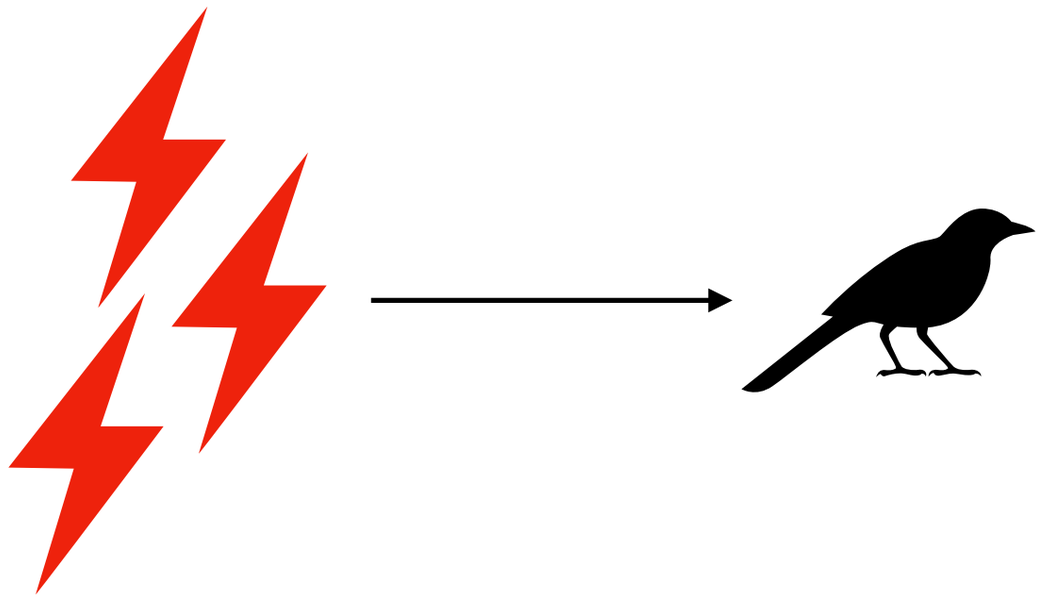
- **attention**
- **working memory**
- **processing**

Visual Cognitive System

1. Encoding



2. Pattern Processing

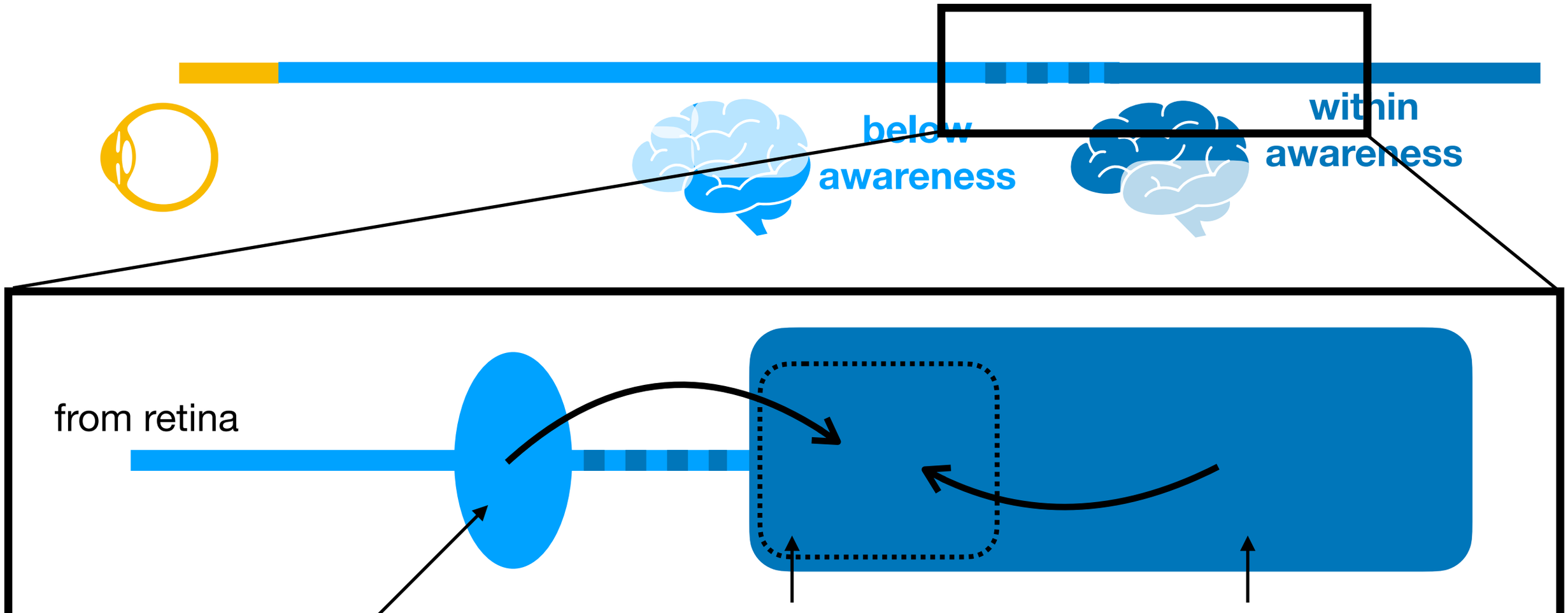


3. Visual Processing

visual working memory

long-term memory

Working memory



Iconic memory buffer

- very short term storage
- holds what is on retina and a few hundred milliseconds later
- lacks semantic content

Visual working memory

- can be drawn from iconic or long-term memory
- combination of external visual info and experiences stored in long-term memory
- context provided by long-term memories

Long-term memory

- information we retain from everyday experiences (for lifetime)
- not really separate from WM

Memory and Attention

- Visual Working Memory: very few available slots

What color was the lamp? pink

- Objects held in iconic memory a very short time (<400 ms).

- Eye movements help reset slots when needed.

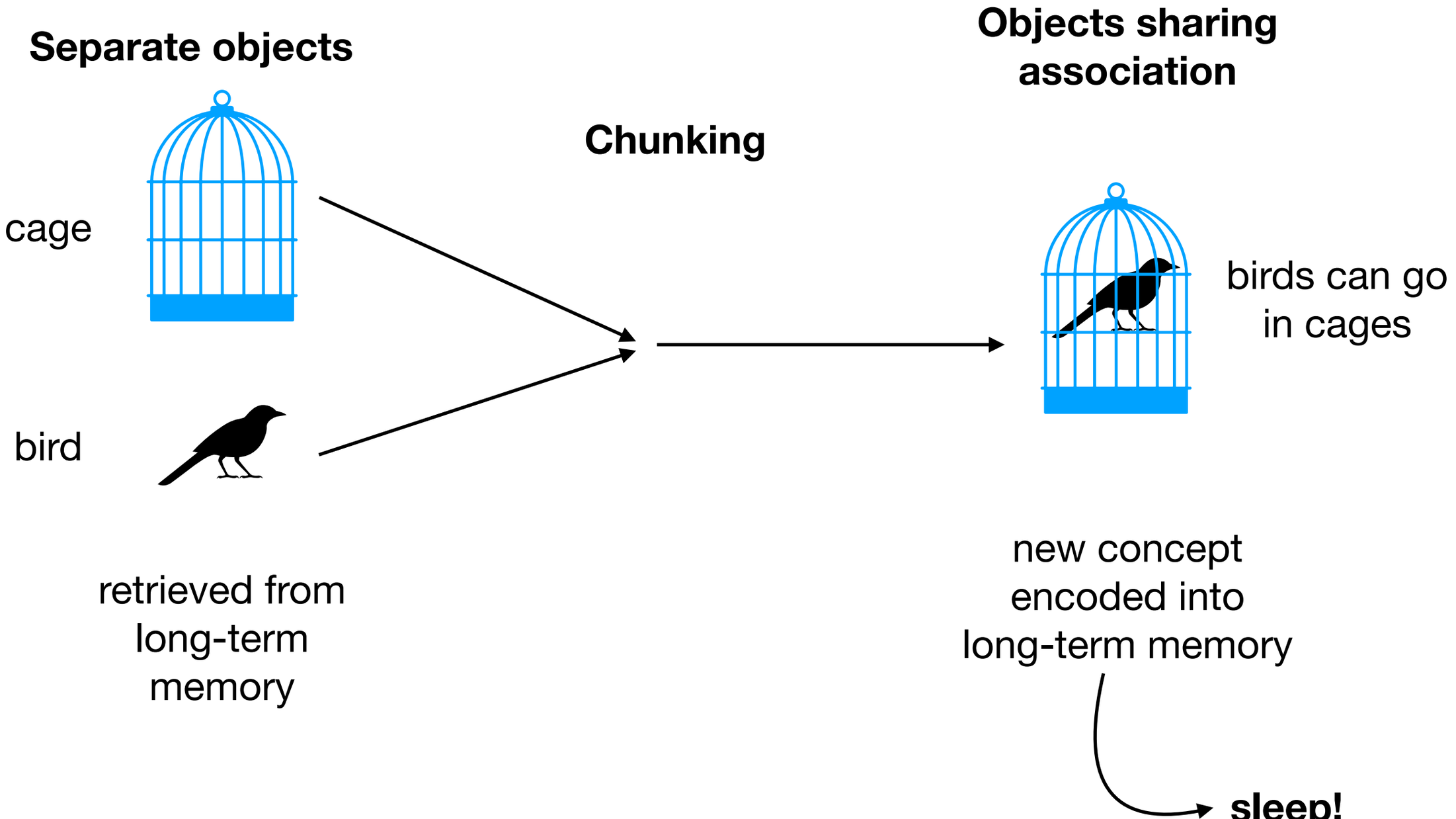
- **Attention helps guide how those slots are filled.**

https://www.youtube.com/watch?v=VkrrVozZR2c

Memory and Attention

- Memory slots are not limited to objects, also concepts and other “chunks.”

  - A **chunk** is just about anything — it is an object, concept, group of objects, plan, etc.



What does this mean for data visualization?

1) Attention is Queen.

- Following OCAR will help you determine how many working memory slots your figure is using, and how to reduce if necessary.

What are the things most important? How do you draw your viewer's attention?

2) Reduce demand on visual working memory by improving processing.

- Limit the number of concepts/relationships you present in each graphic.
- Use elements that make visual processing easier.
- Make guides clear and easy to understand.

How can you revise your graphic to make understanding easier?

Check Your Understanding

Are flawless diamonds on average smaller than those with inclusions?

1) **Attention is Queen.** What is the most important point to draw your viewer's attention to? How can you create a relationship that shows this point?

2) **Reduce demand on working memory.** Is there anything not relevant in your graph that you could exclude? Are your graph elements easy to understand and visually process?

Action Items

1. **Prepare Project 1 for submission.**