# Lecture 09 – The Good, The Bad, The Ugly

## Today's Learning Objectives:

1. Describe the OCAR system of story telling.

2. Describe what aspects of graphs distinguish them as good, bad, wrong, and ugly.

3. Practice implementing OCAR on graphs.

# The Bad, The Wrong, and The Ugly

- **Wrong**: graphs that have problems related to math or scaling, figures that mislead.
    - Figures could be misleading on purpose or by accident.
    - Unacceptable in publications and this class.
    - **First and foremost: you must NOT mislead your readers!**

- **Bad**: graphs that have problems related to perceptual issues, unclear or confusing, overly complicated or deceiving.
    - Unacceptable in publications and this class.
    - 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.

- **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.
    - We can improve these in class!

**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%.

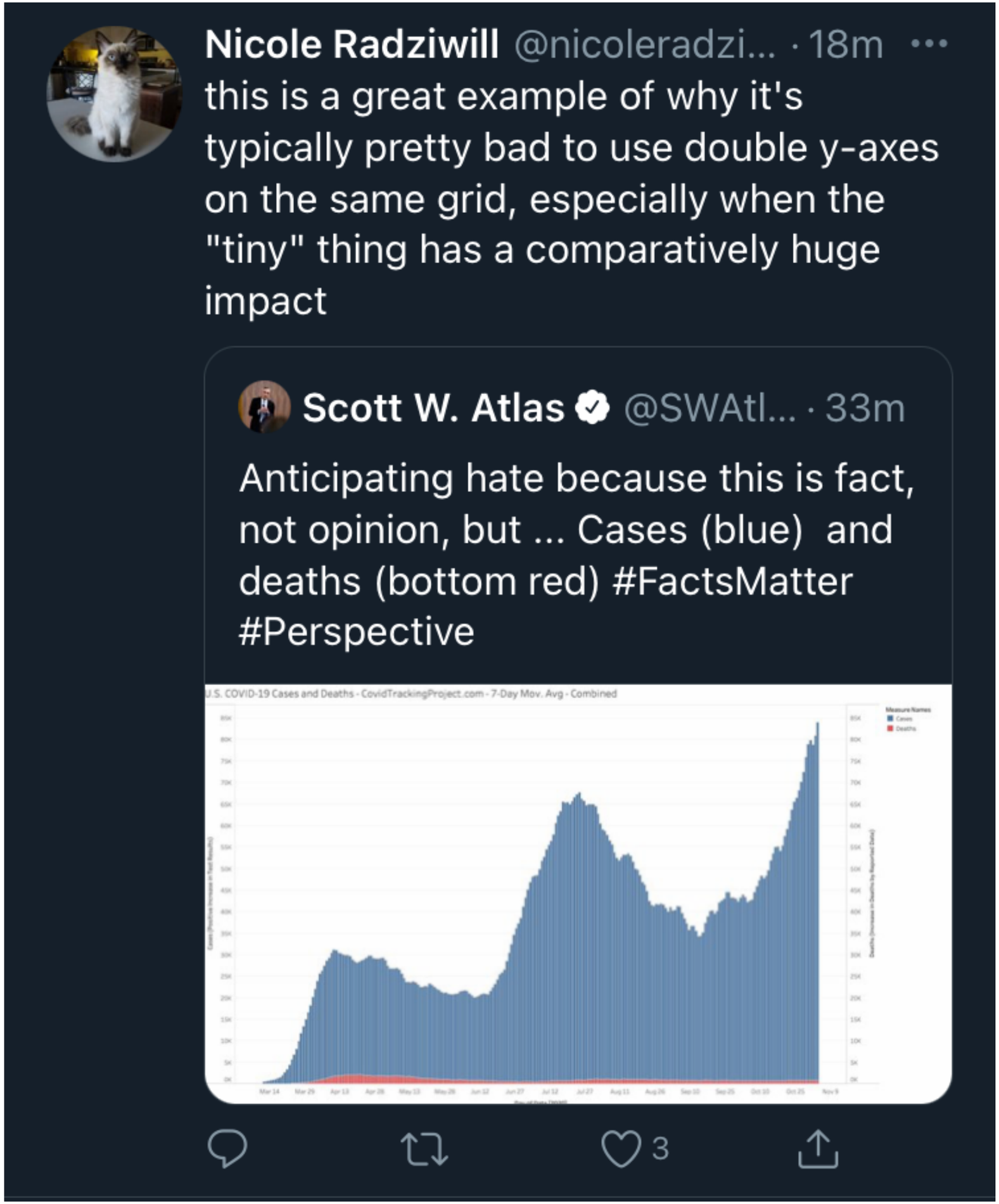
>  **Zeducation** @RealZeducation · Sep 25
>
> Just a reminder that everything in 2020 was over a virus that does this:


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

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

- Problems with math.

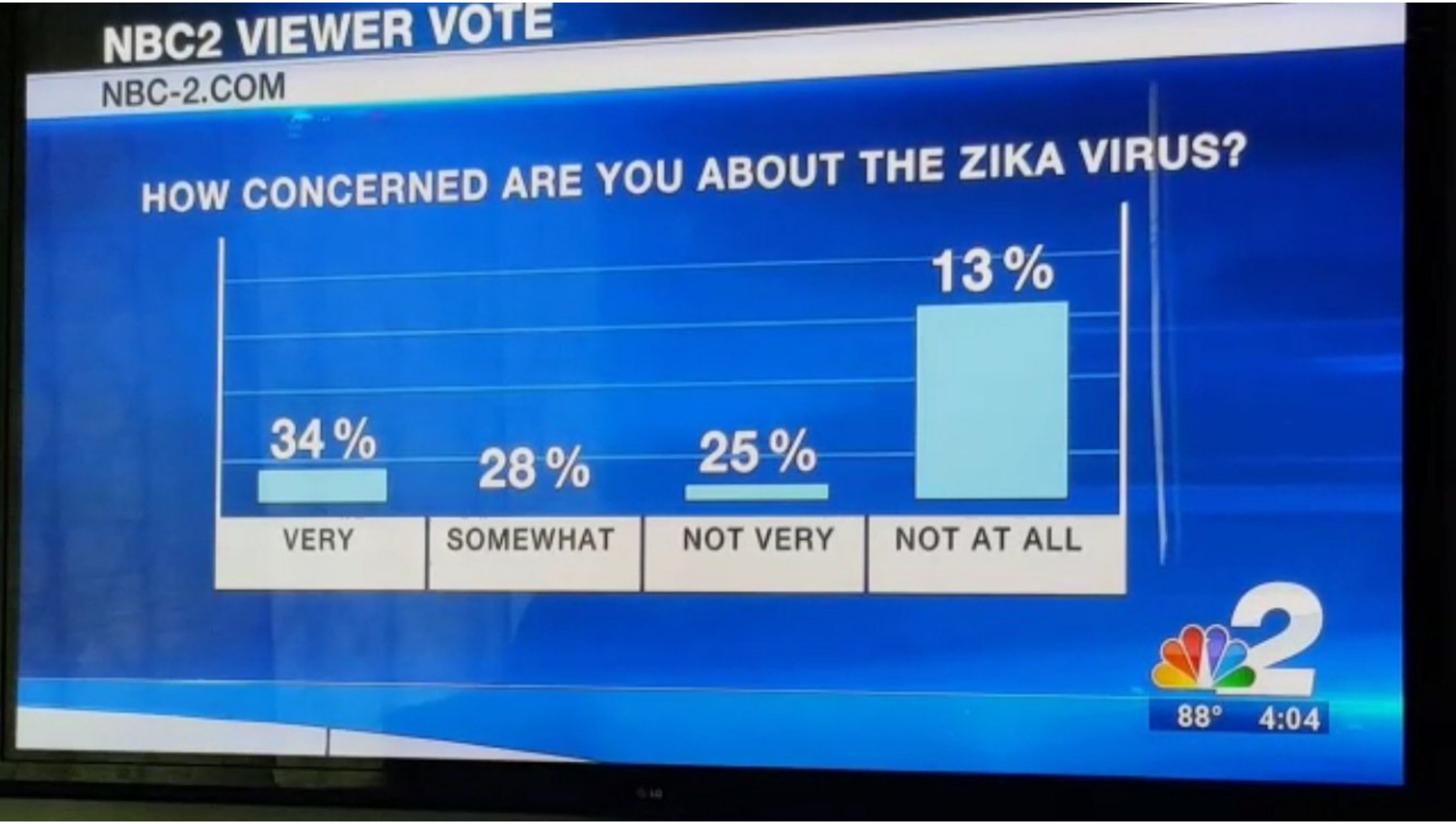
**BAD/ WRONG**




- Misleading axis scaling.

# WRONG

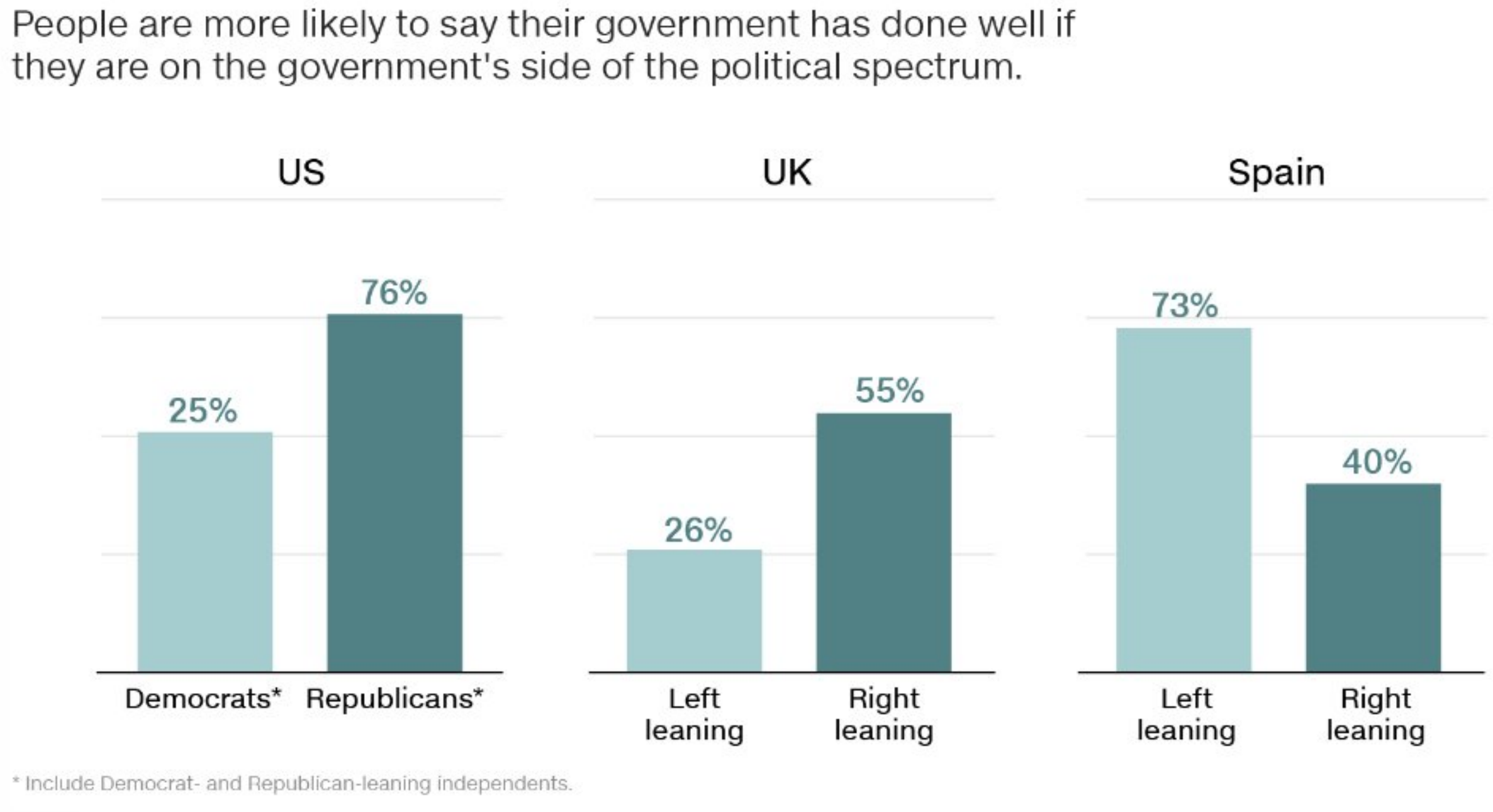
- Numbers don’t match size of bars.




**Seriously, wtf?**

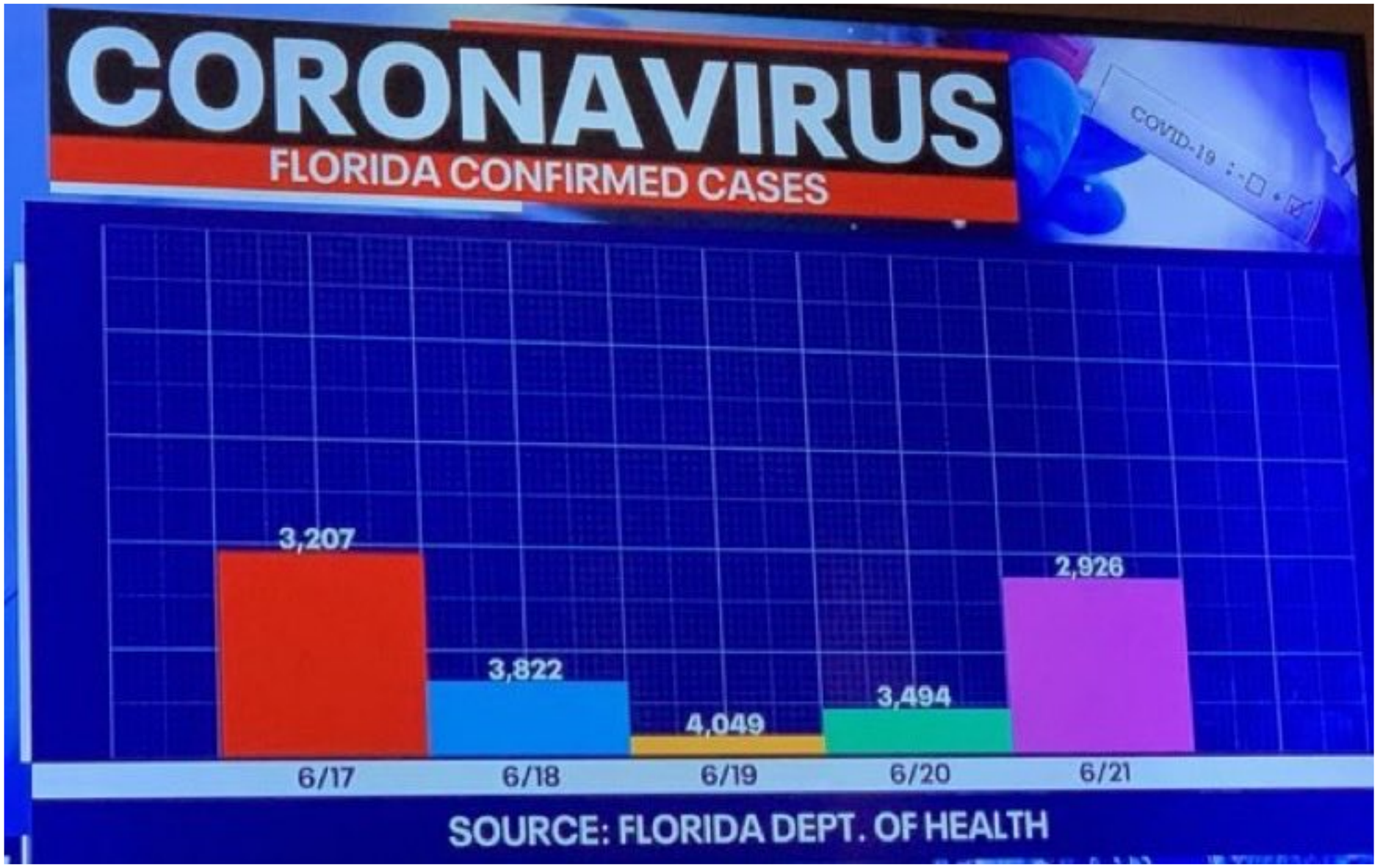
# BAD

- Floating, unlabeled y axis, misleading.



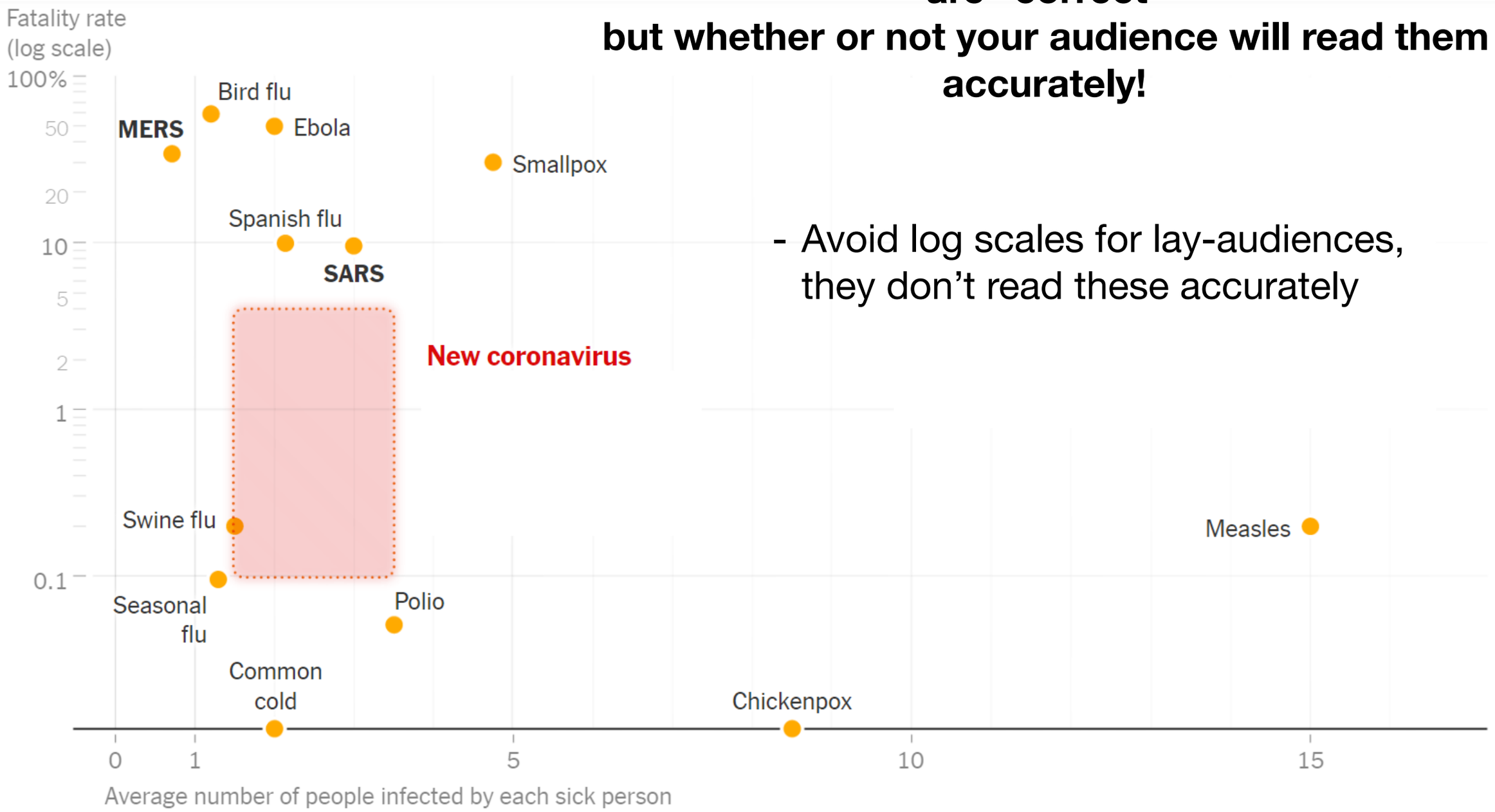
# WRONG

😒


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

# BAD

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



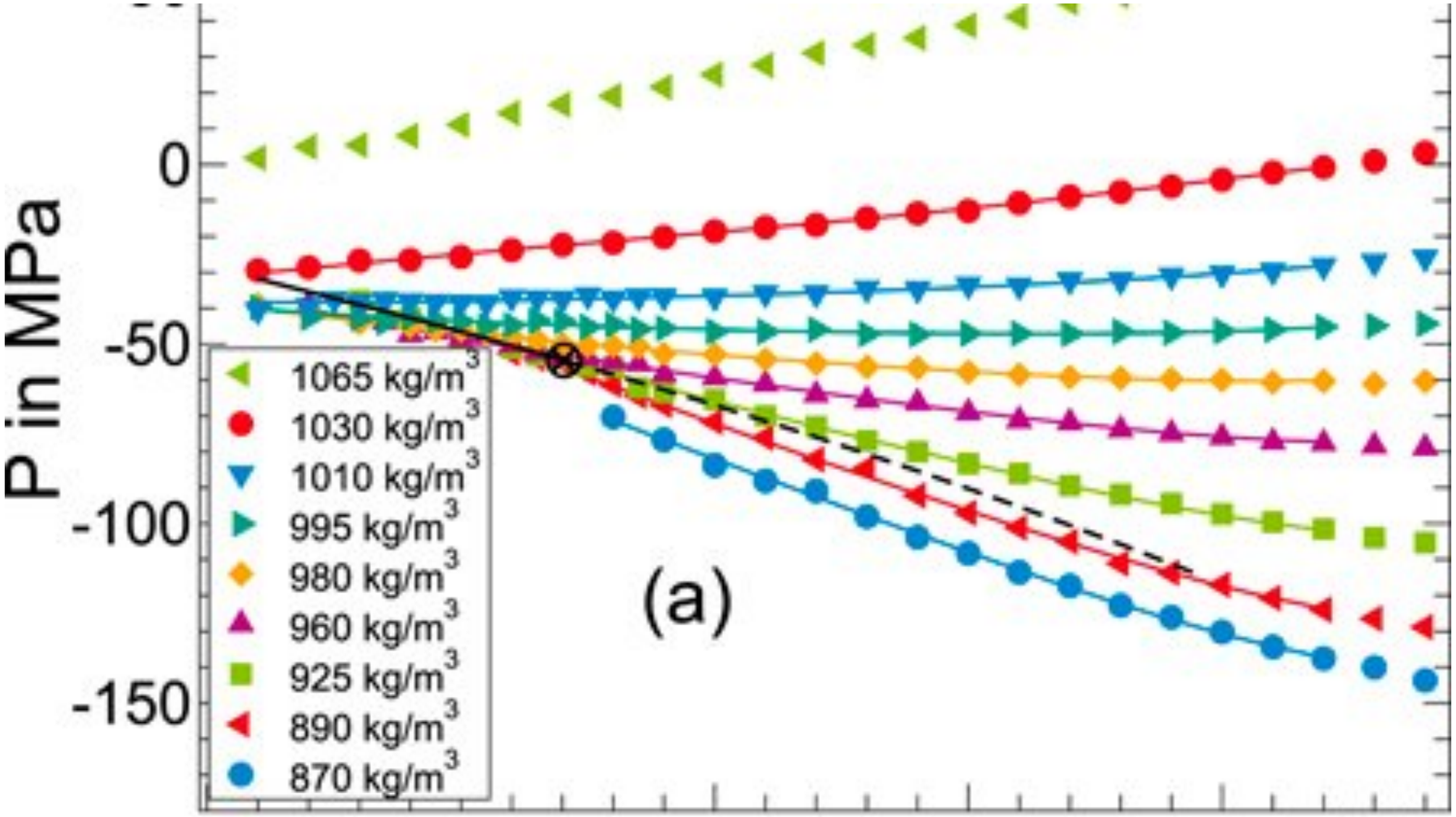

Note: Average case-fatality rates and transmission numbers are shown. Estimates of case-fatality rates can vary. **The preliminary estimates for the new coronavirus are shown in the pink region.**

- 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**

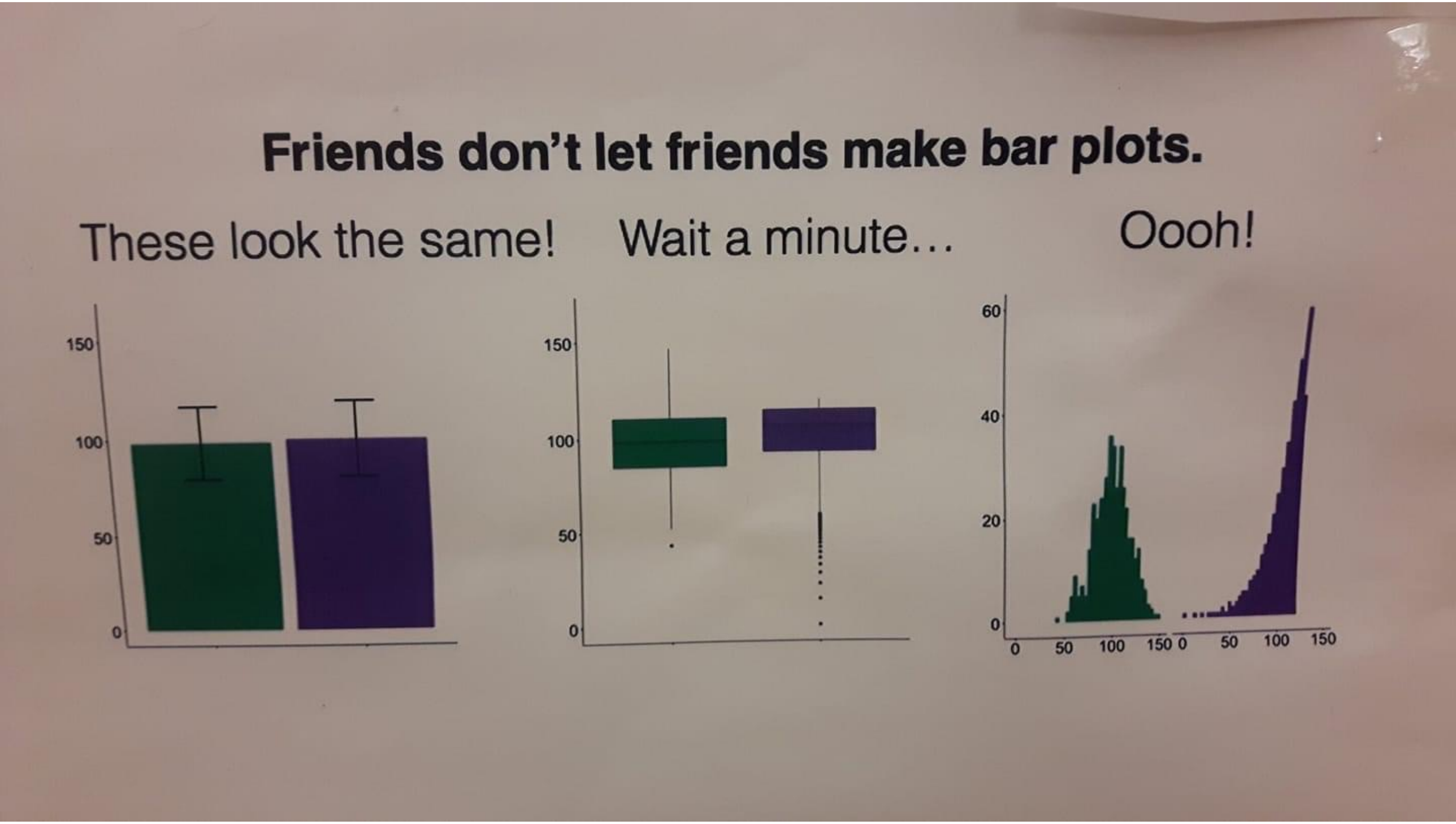
# UGLY

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



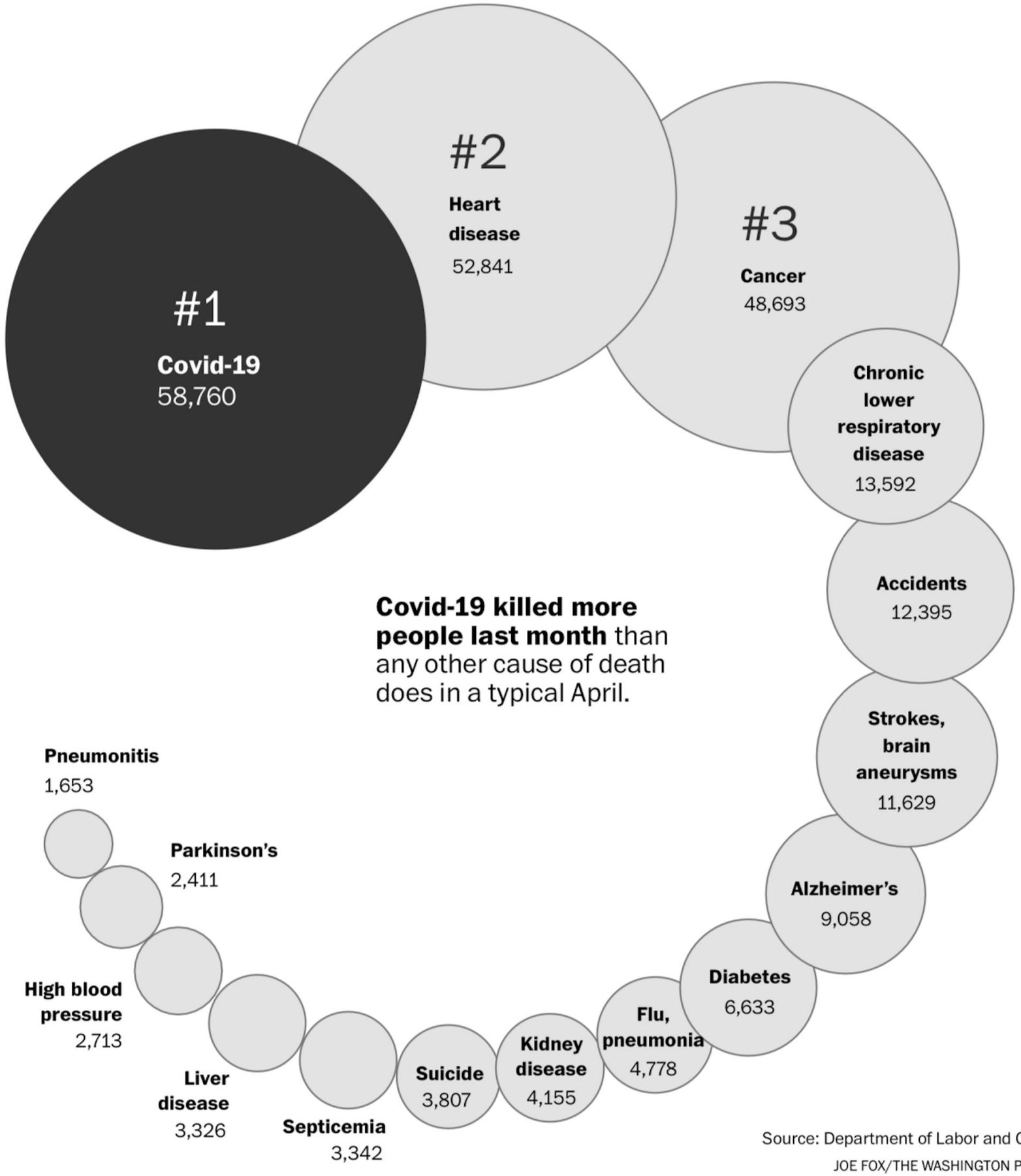

**(X axis labels got cut off, my fault)**

# Bad to Good

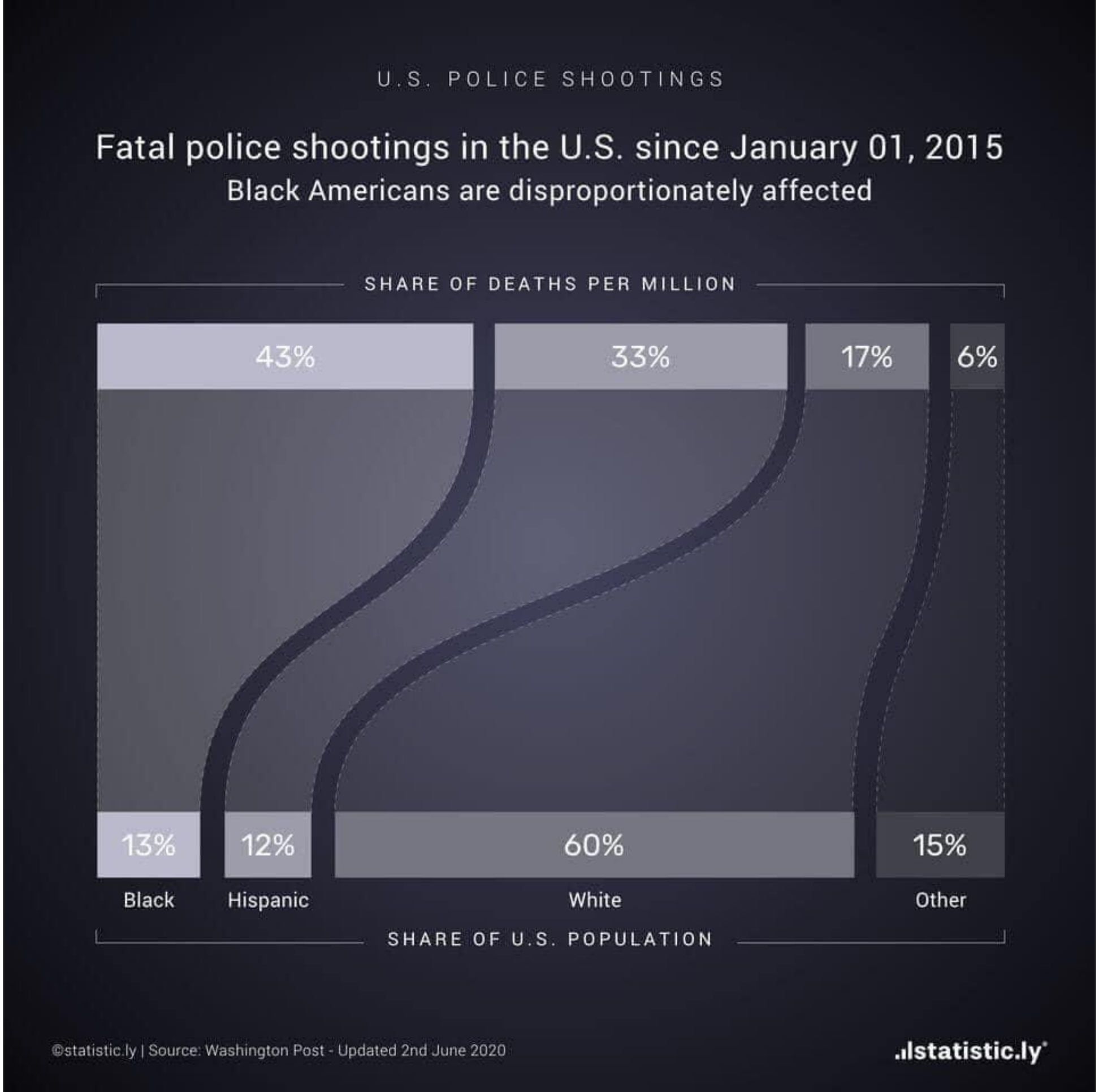



- Your job is to represent your data accurately, with as much representation to the raw data as possible.

# GOOD

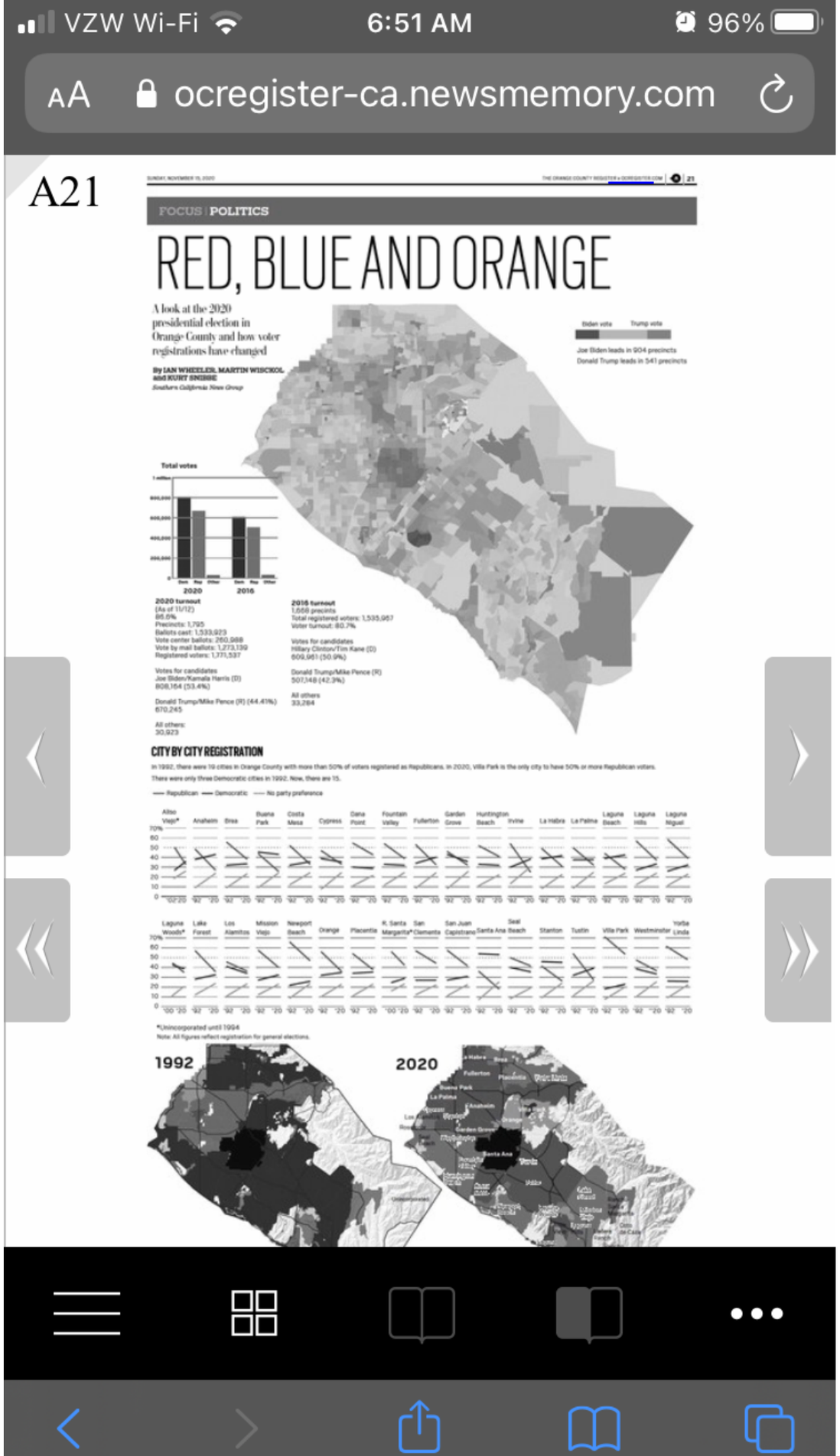

#1
Covid-19
58,760
#2
Heart disease
52,841
#3
Cancer
48,693
Chronic lower respiratory disease
13,592
Accidents
12,395
Strokes, brain aneurysms
11,629
Alzheimer's
9,058
Diabetes
6,633
Flu, pneumonia
4,778
Kidney disease
4,155
Suicide
3,807
Septicemia
3,342
Liver disease
3,326
High blood pressure
2,713
Parkinson's
2,411
Pneumonitis
1,653
Covid-19 killed more people last month than any other cause of death does in a typical April.
Source: Department of Labor and CDC
JOE FOX/THE WASHINGTON POST

GOOD

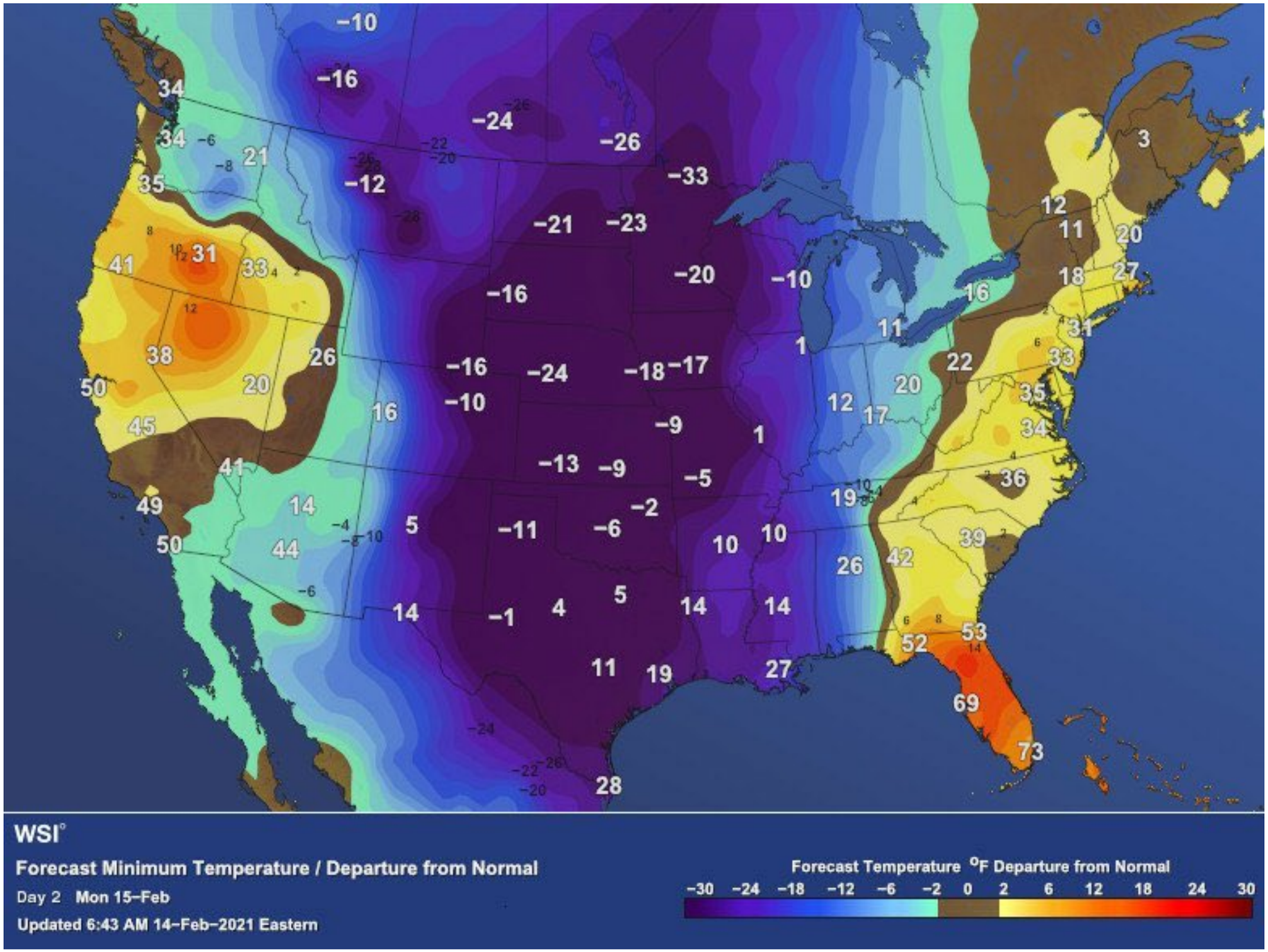

U.S. POLICE SHOOTINGS
Fatal police shootings in the U.S. since January 01, 2015
Black Americans are disproportionately affected
SHARE OF DEATHS PER MILLION
43%
33%
17%
6%
13%
12%
60%
15%
Black
Hispanic
White
Other
SHARE OF U.S. POPULATION
©statistic.ly | Source: Washington Post - Updated 2nd June 2020
statistic.ly

# You decide

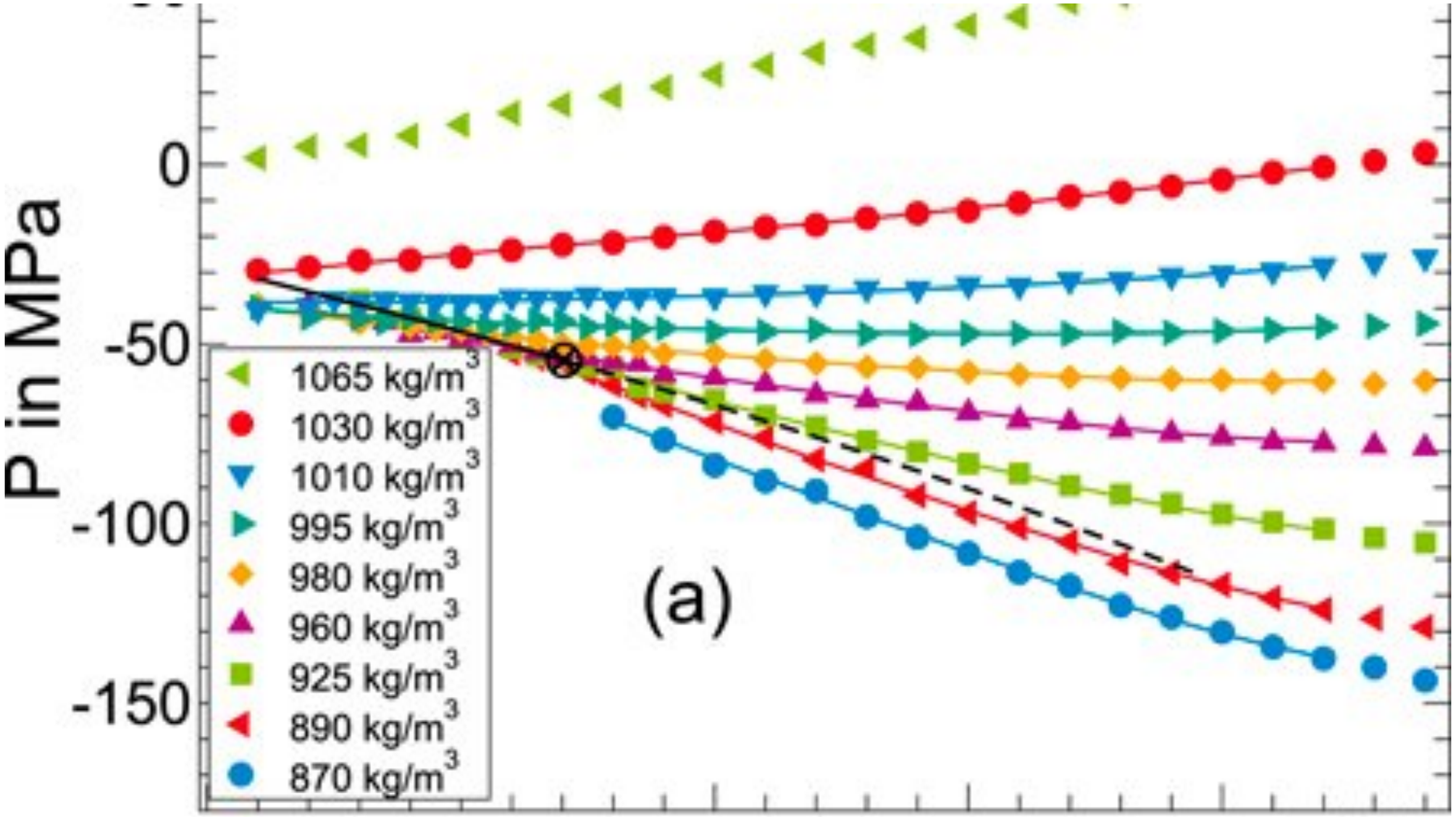
**Orange county voting results in 2020 presidential election**



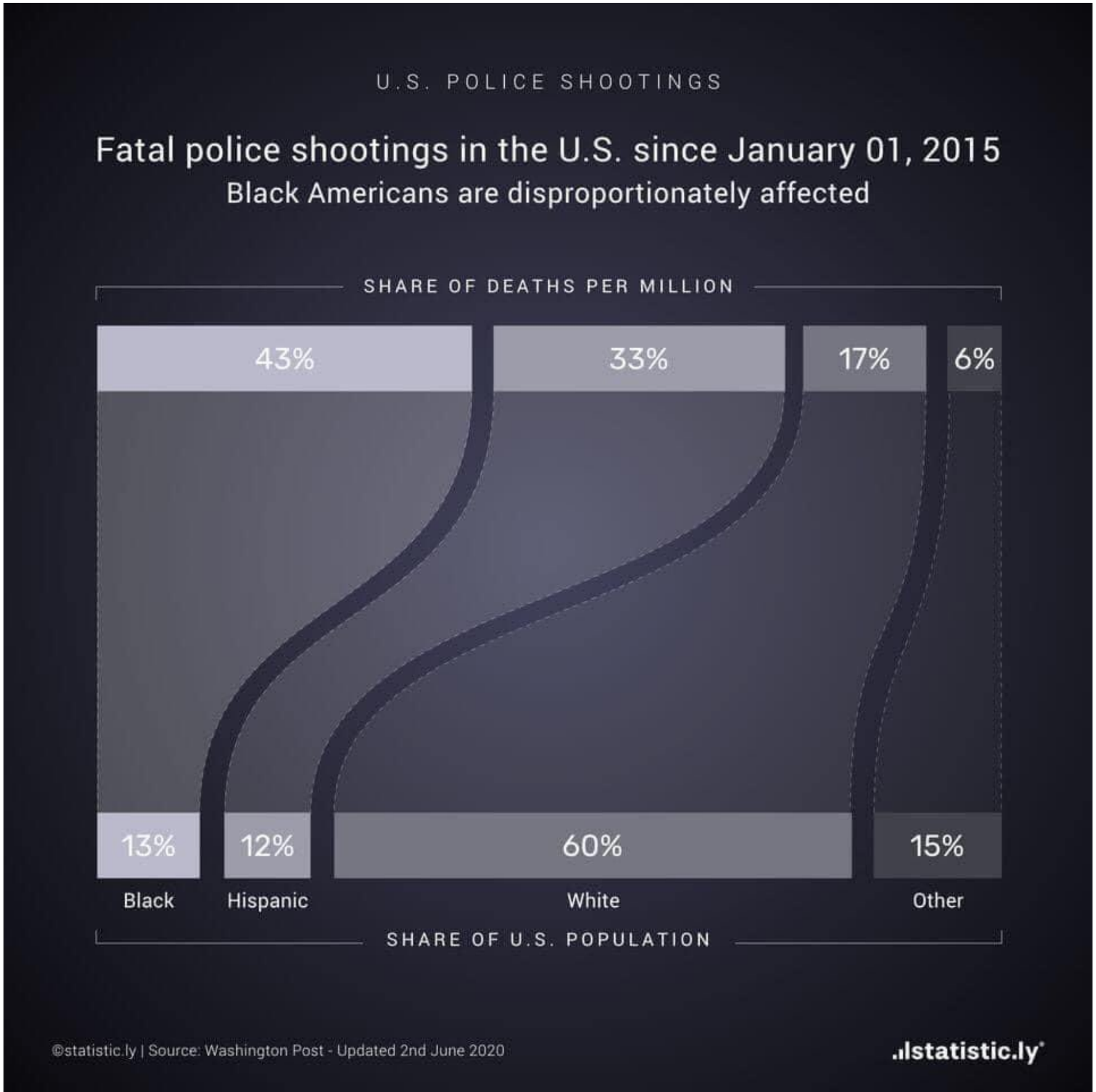
# You decide


WSI
Forecast Minimum Temperature / Departure from Normal
Day 2 Mon 15-Feb
Updated 6:43 AM 14-Feb-2021 Eastern
Forecast Temperature °F Departure from Normal
-30 -24 -18 -12 -6 -2 0 2 6 12 18 24 30

# How would you improve this graph?



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



# 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:

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

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

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

  - **R**esolution: 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:**

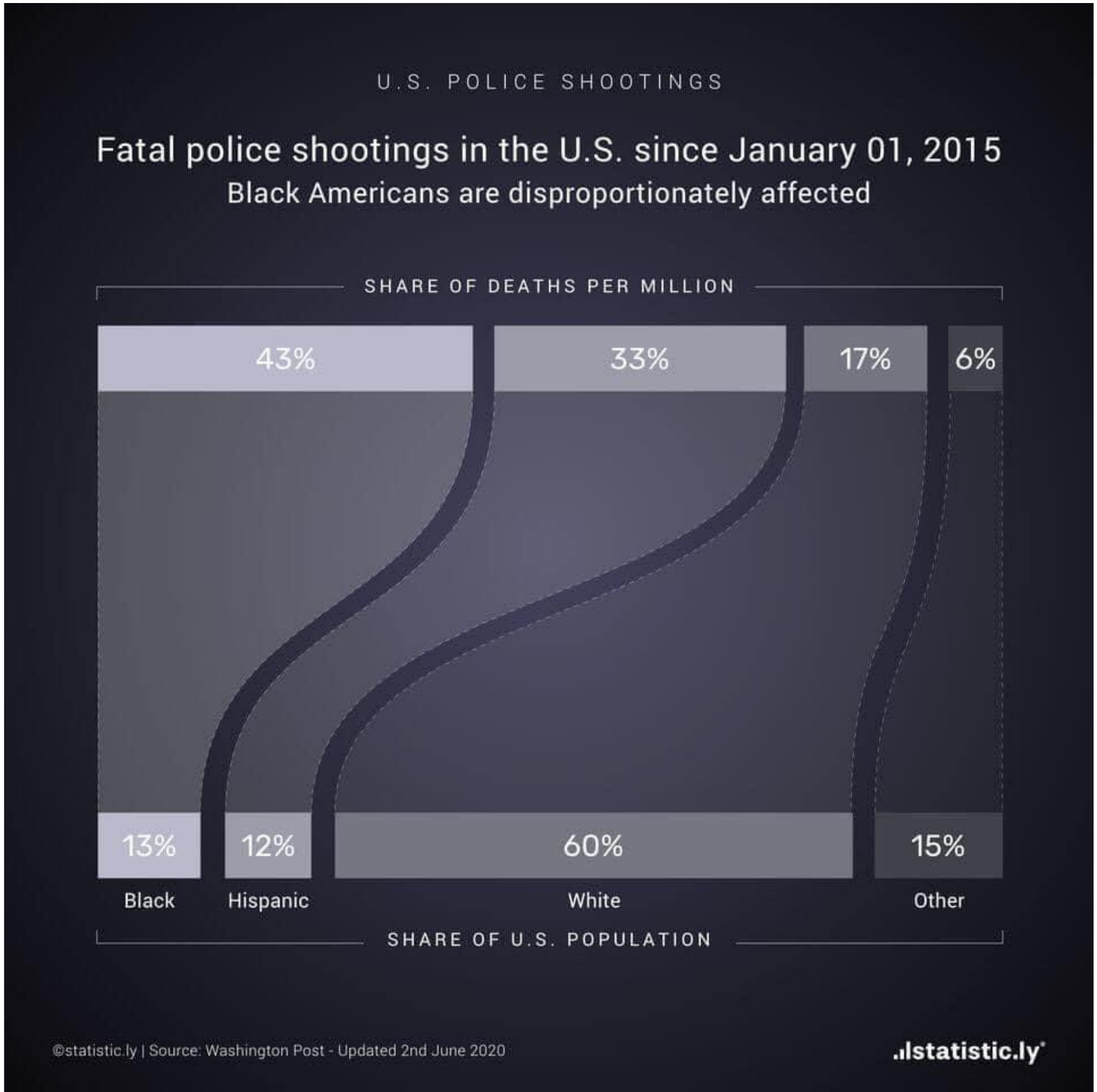
How many black people are killed by police compared to whites?

**Action:**

Flow of proportionality between two scales.

**Resolution:**

Black people are killed disproportionately more than whites.



## Group work:

Use the diamonds data set to answer the following questions:

1. Are flawless diamonds on average smaller than those with inclusions? (Ch 6)
2. What is the distribution of price per carat of the diamonds in this set? (Ch 7)
3. What proportion of diamonds in the set are of each cut? Does this change with clarity? (Ch 10)
4. What determines a diamond's cost? (Ch 12,14)

Use the beavers data set to answer the following questions:

5. Does the body temperature of a beaver vary with time? (Ch 13)
6. How much variation is associated with beaver temperature measurements? (Ch 16)

https://www.kaggle.com/shivam2503/diamonds

**Bring several graphs with you for Wed 3/3 to present in class. Be prepared to present:**

a) Examples of your graphics answering each question.
b) Explain the graphing choices you made for each (how does it follow OCAR?).
c) Provide code to the rest of the class reproducing your graphs.