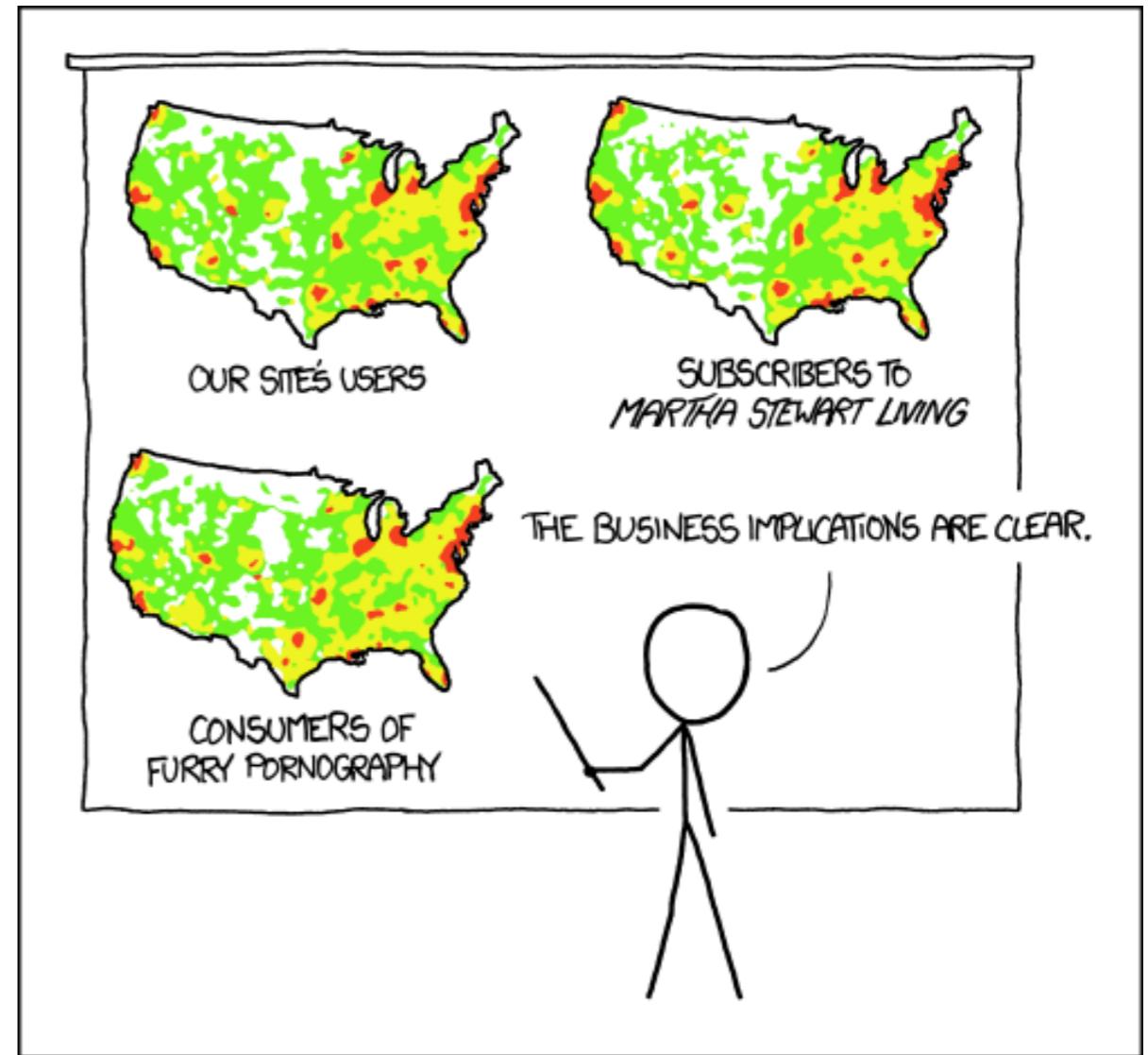


# Data visualization

Why care & what to do about it



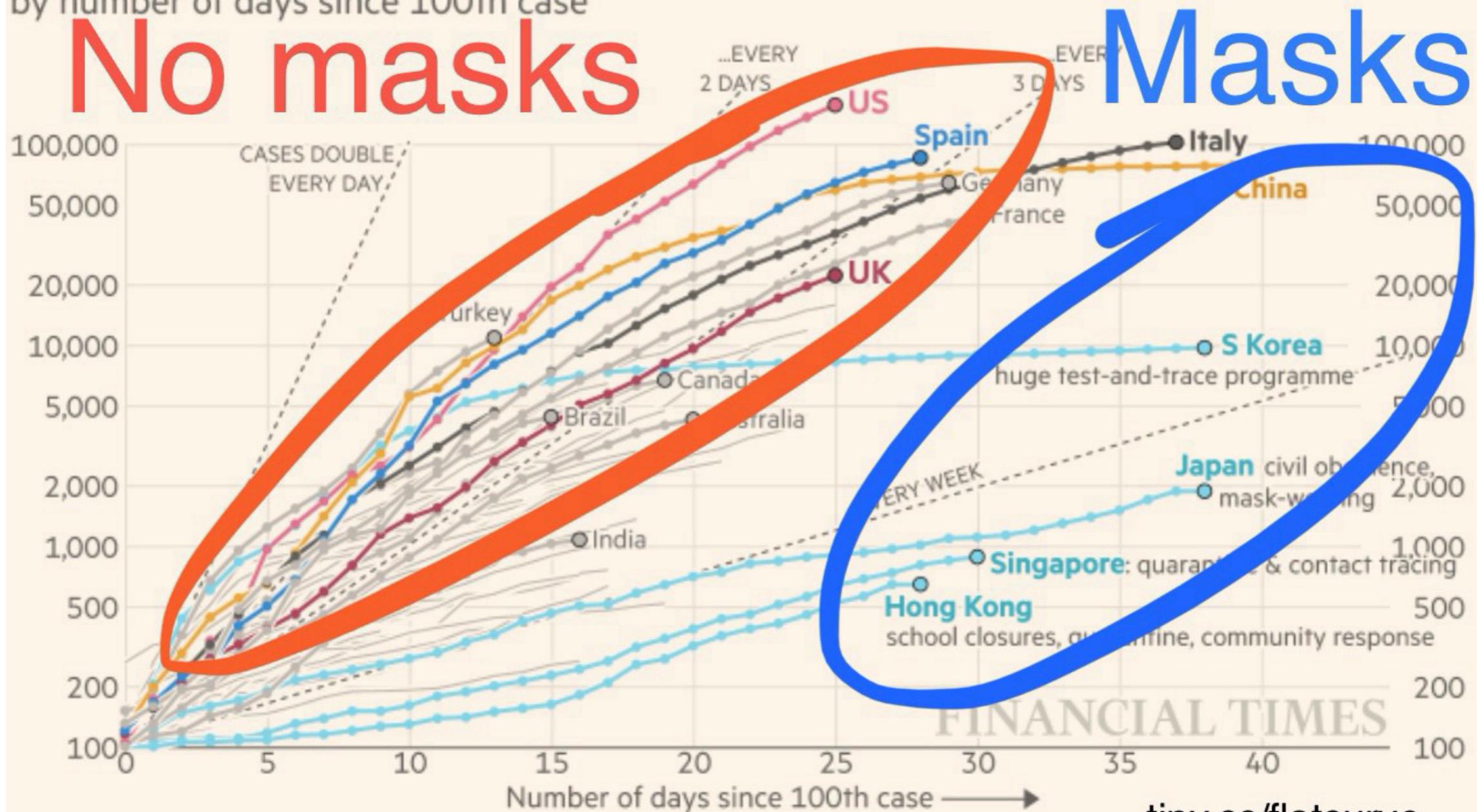
# What are your impressions?

What would your initial impression be if this showed up on your newsfeed?

What do you think the author is trying to communicate? What questions do you still have about it?

## Country by country: how coronavirus case trajectories compare

Cumulative number of confirmed cases,  
by number of days since 100th case



FT graphic: John Burn-Murdoch / @jburnmurdoch

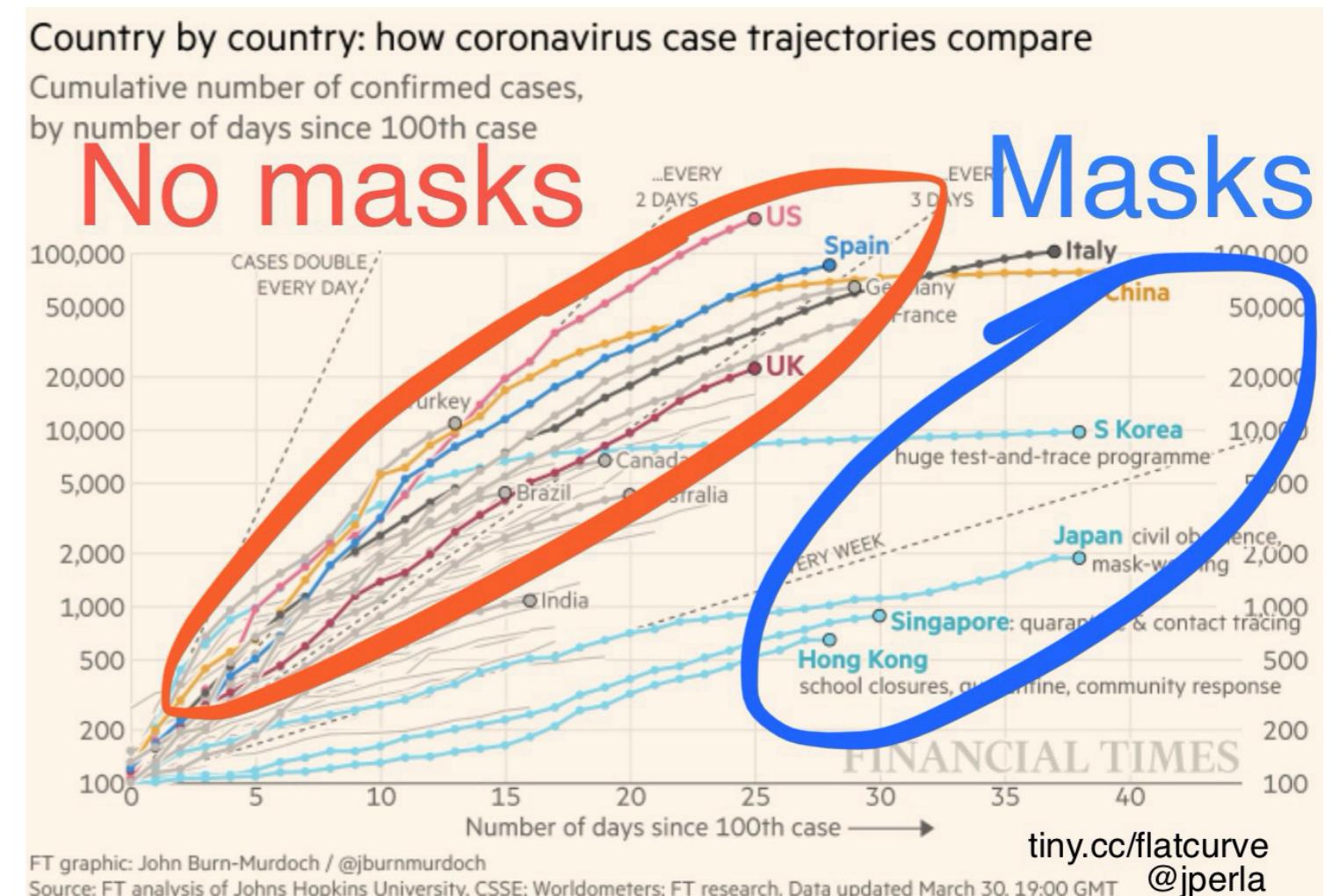
Source: FT analysis of Johns Hopkins University, CSSE; Worldometers; FT research. Data updated March 30, 19:00 GMT

# What are your impressions of this visualization?

Impressions here:

*Plots like this were widely circulated on social media, especially Twitter*

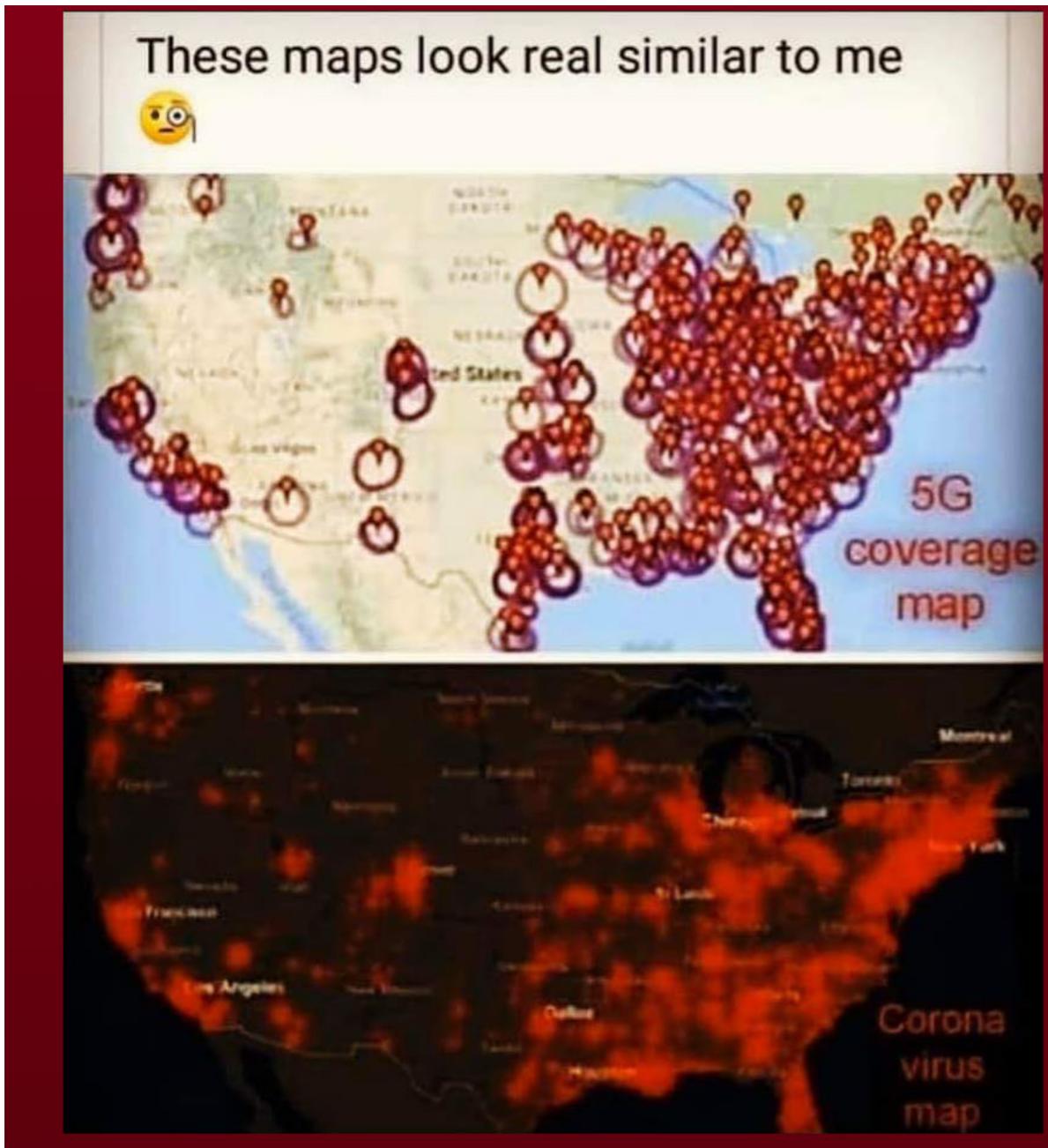
1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_
4. \_\_\_\_\_



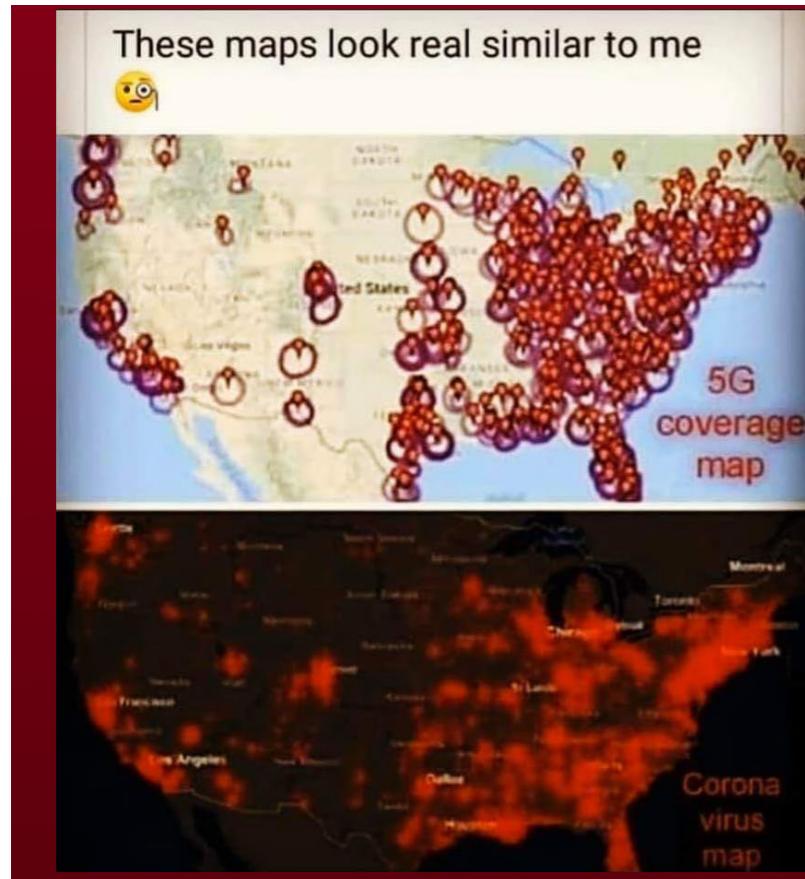
# What are your impressions of this visualization?

What would your initial impression be if this showed up on your newsfeed?

What do you think the author is trying to communicate? What questions do you still have about it?



# What are your impressions of this visualization?



Subway sandwich shops:



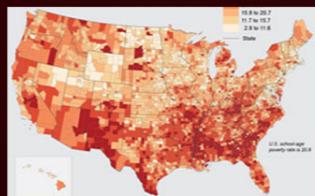
Hair restoration centers:



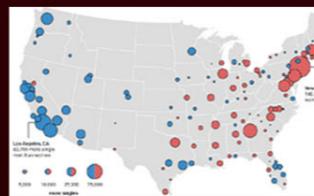
McDonald's Locations:



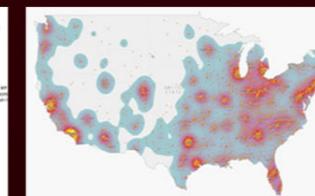
Children in Poverty:



Single People:



Taco Bell:



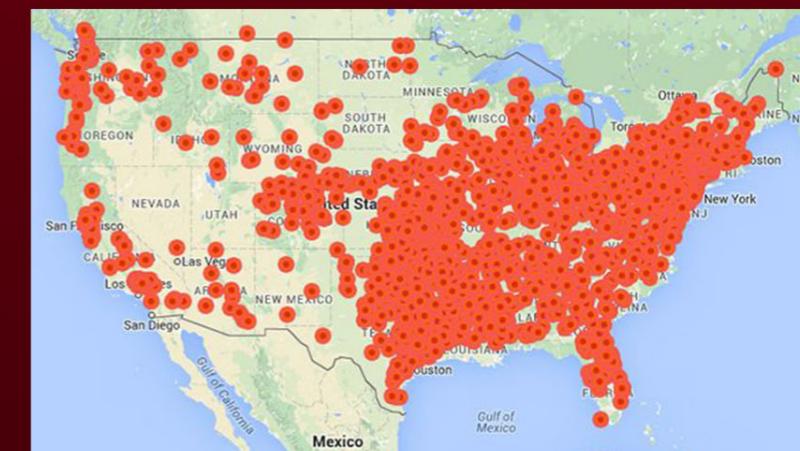
The Elderly:



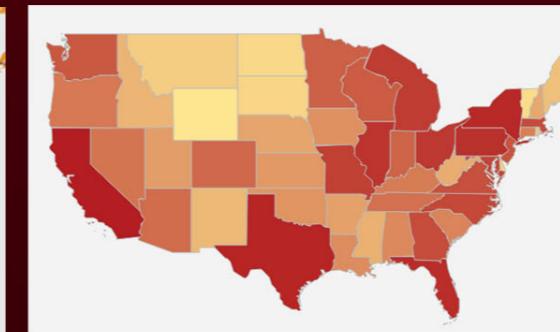
Juggalos:



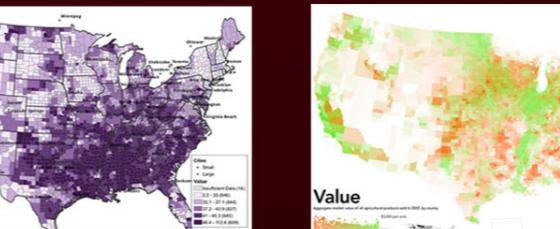
The KKK:



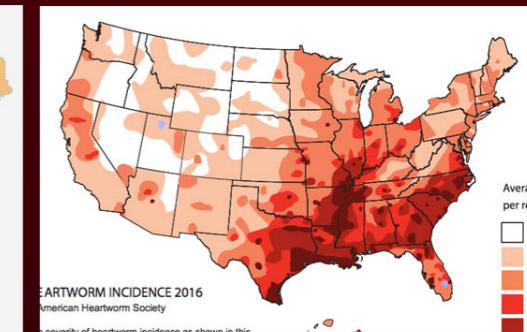
"Karens":



Agriculture:

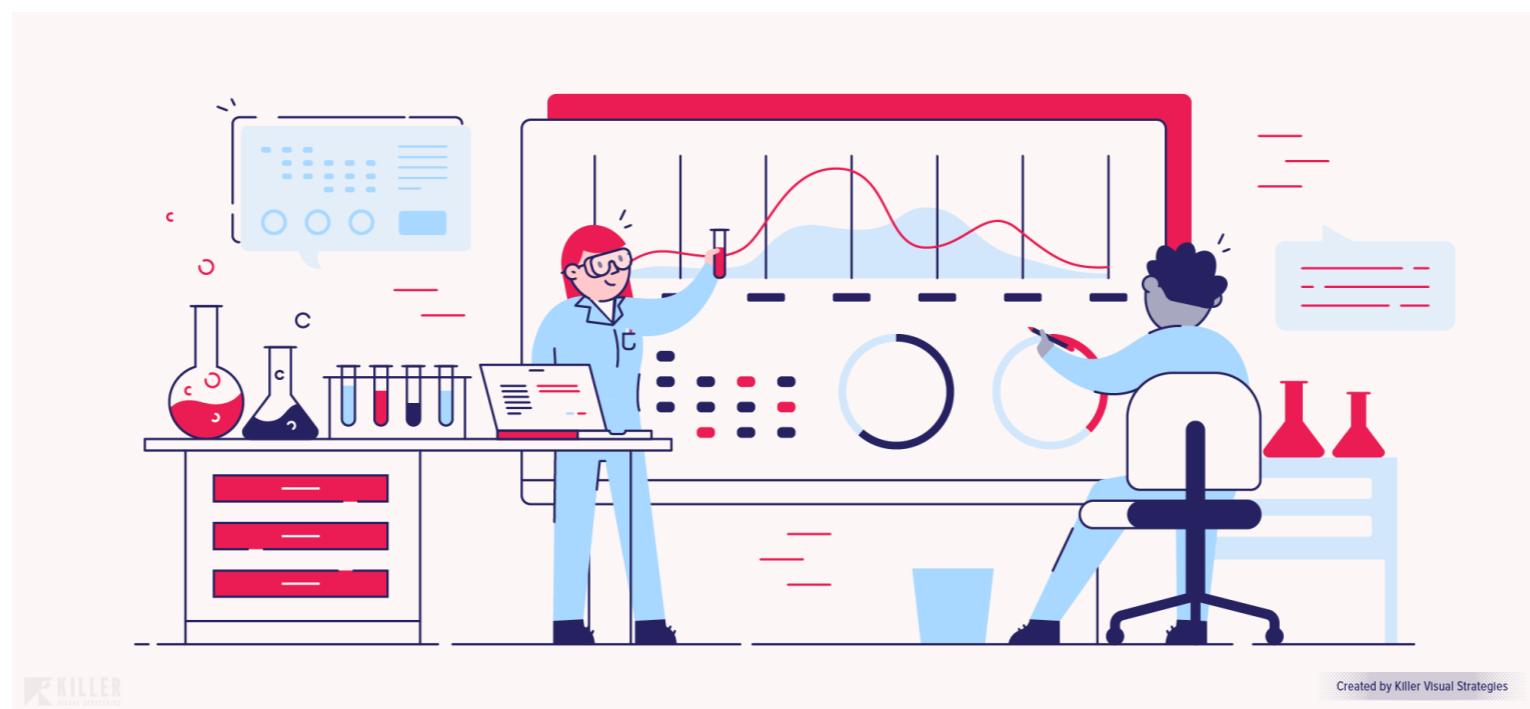


Dogs with Heartworm:



# Why care about data visualization?

1. Data visualization is a powerful way to tell a story or communicate a message
2. It's all around us, on the news, media, research articles
3. As scientists, we'll run into the need to create or judge data visualizations
4. Now that we know R, there are many tools in R that we can use to make effective data visualizations



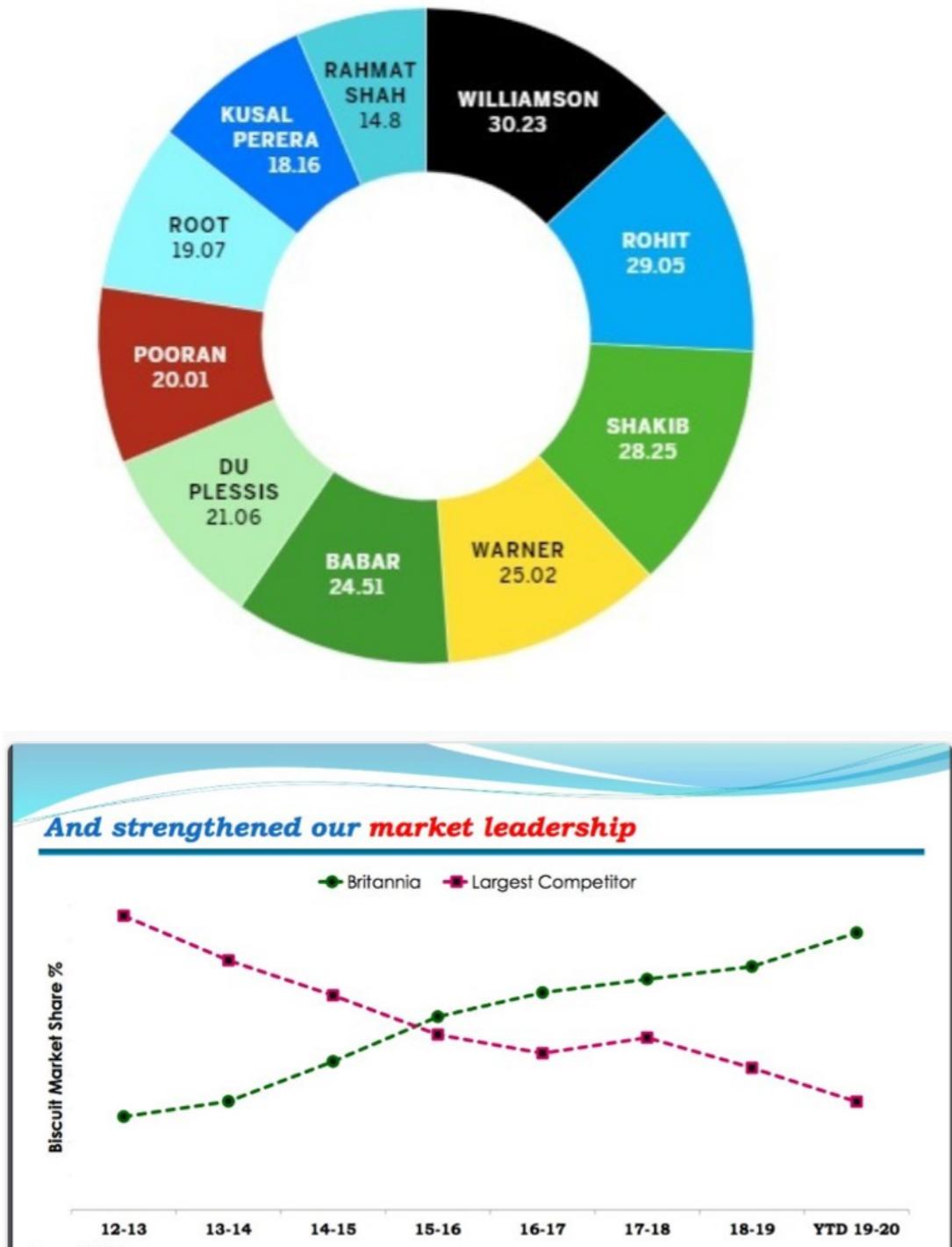
# What's our role in data visualization?

1. Be critical of data visualizations
2. Create effective and meaningful visualizations in your own work (by using R!)



Created by Killer Visual Strategies

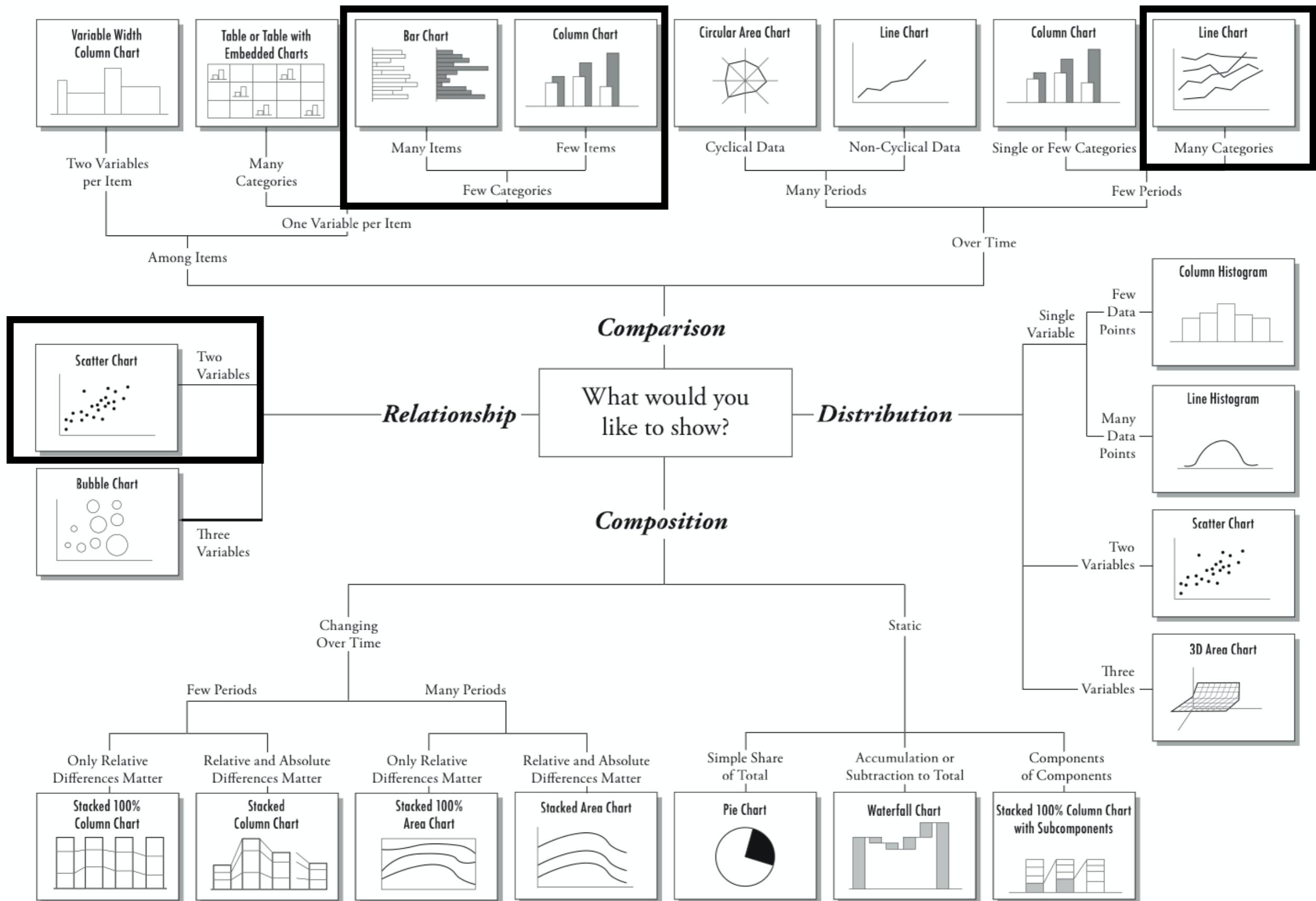
# Choose the right kind of plot for your data



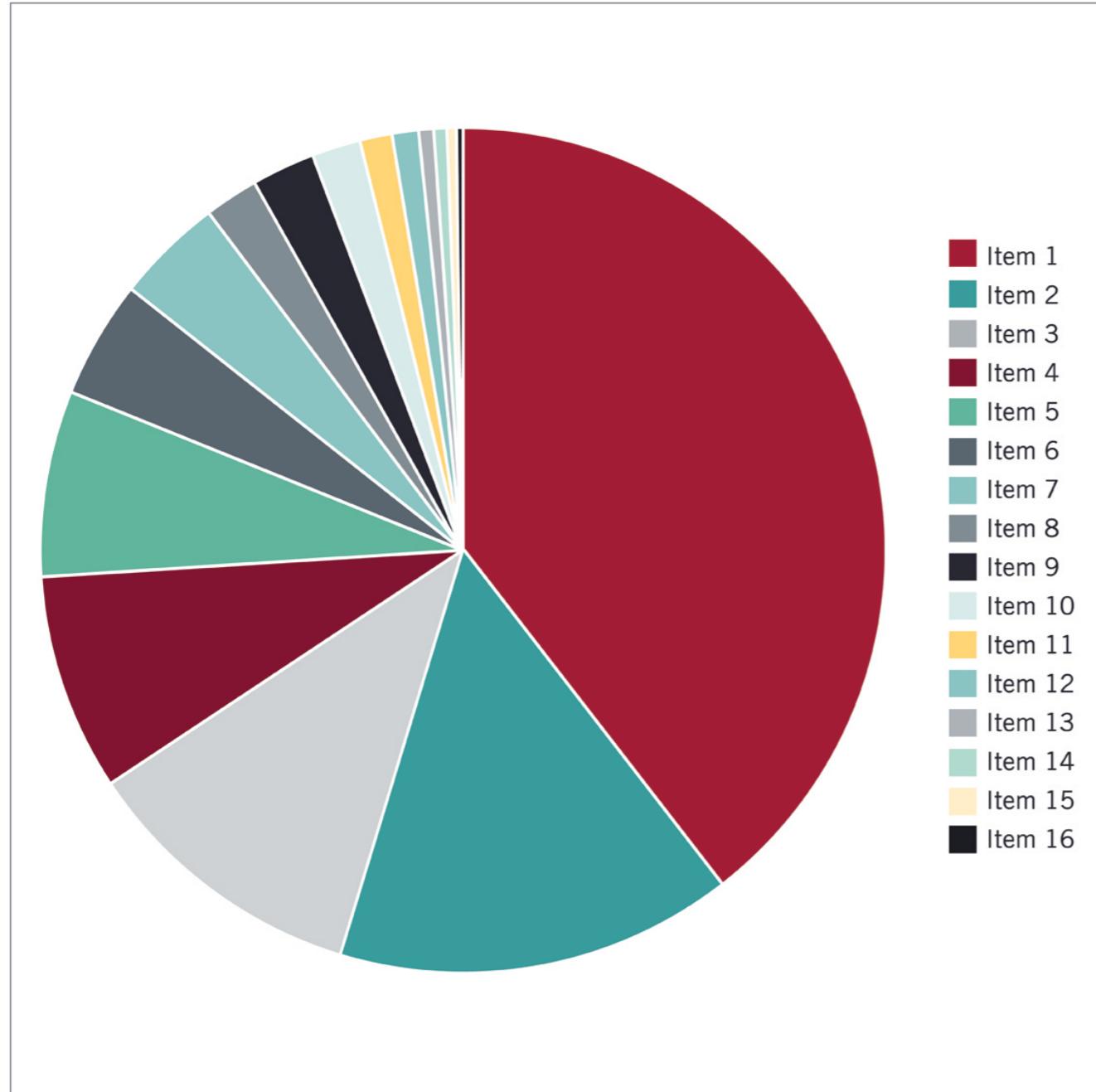
Before plotting, ask:

1. How many variables do I have?
2. What kinds of variables are they? (Quantitative or qualitative?)
3. Do the variables change over time?

# Chart Suggestions—A Thought-Starter



# Keep it simple for the reader



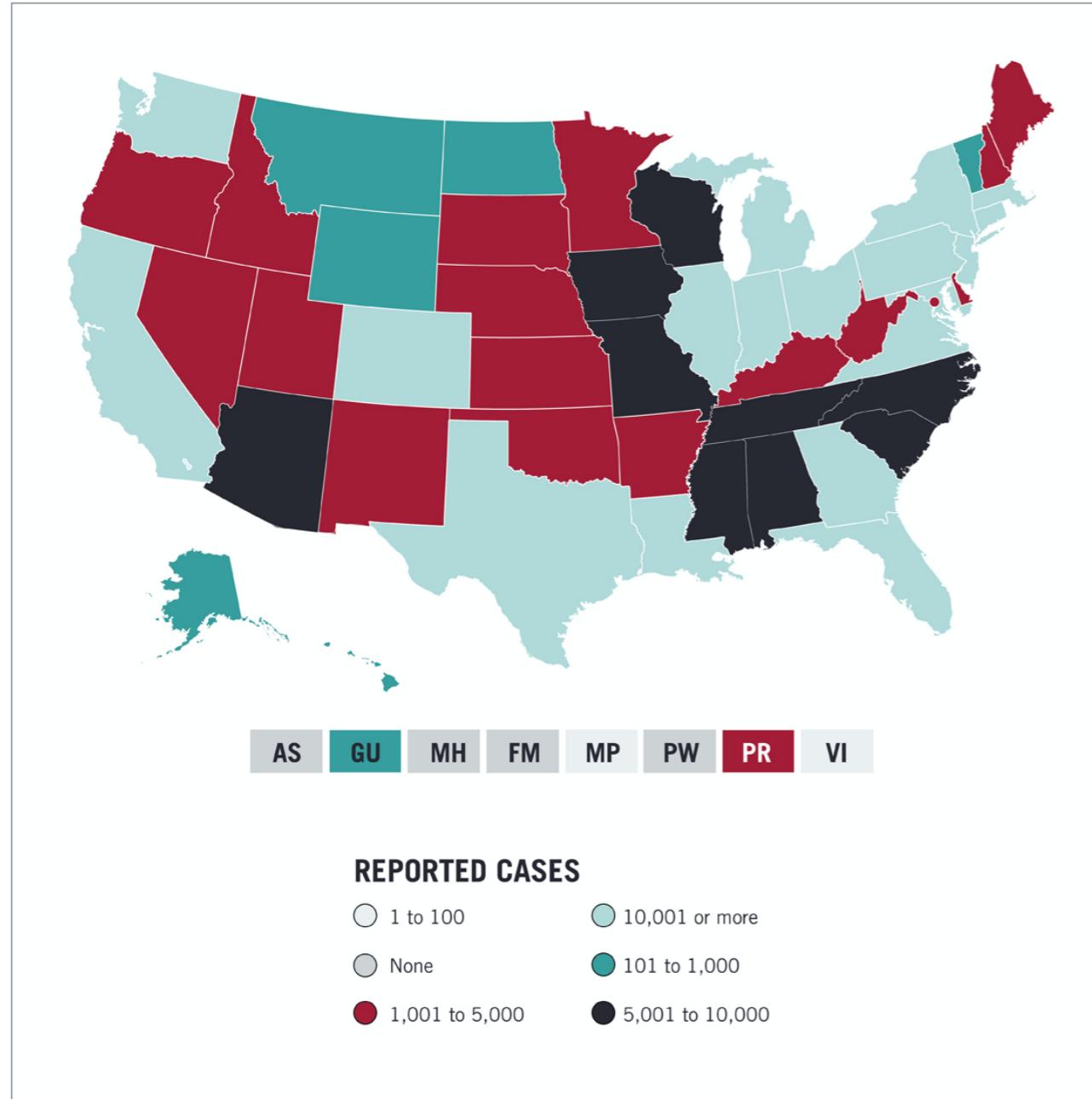
Too many variables?

1. Grouping data
2. Filter to include the most important ones

Too much shifting between plot and legend?

1. Directly label the plot where possible
2. Reduce the number of variables (above)

# Choose your colors wisely



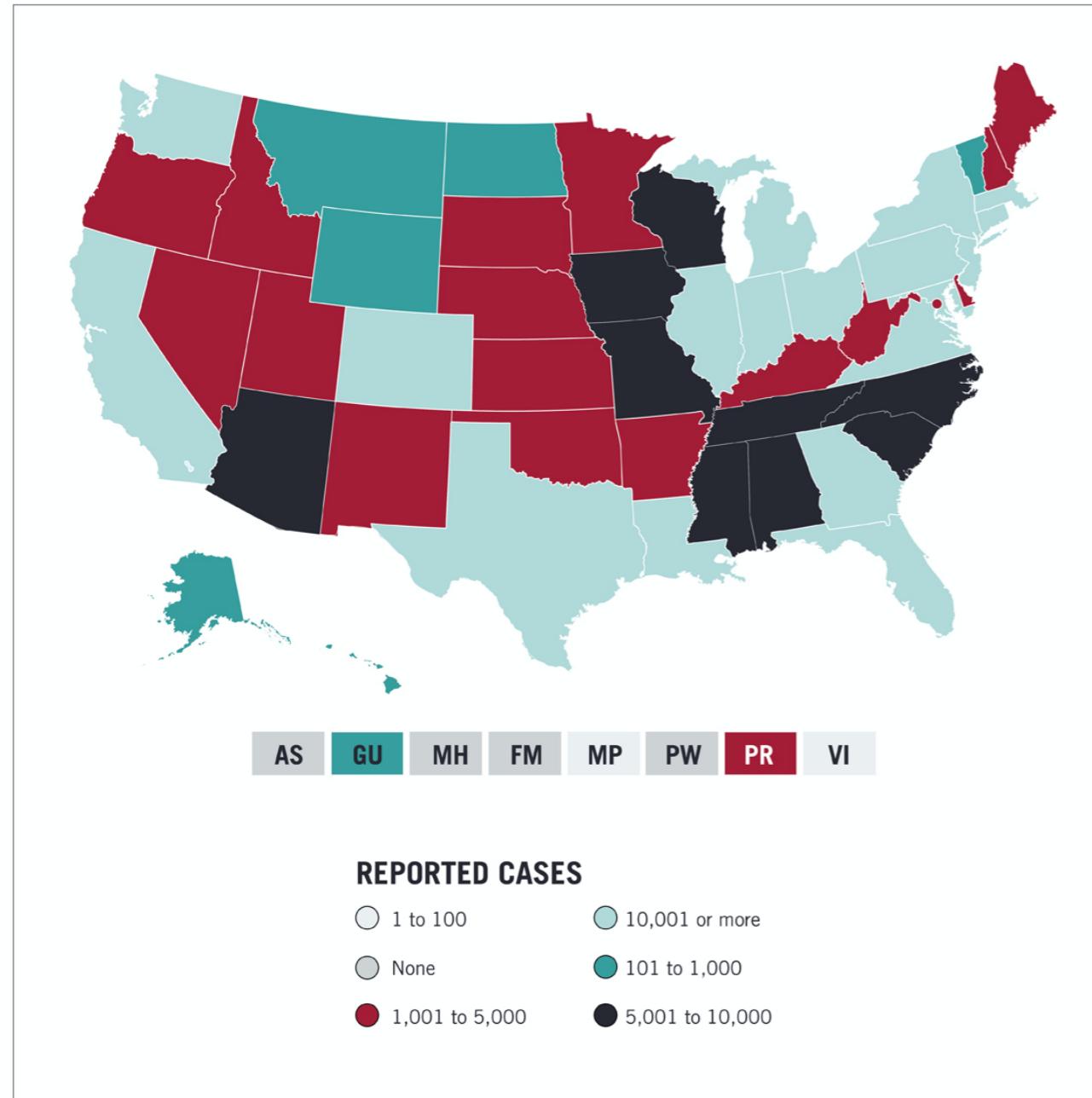
There's a hierarchical ordering in the **reported cases** variable

None < “1 to 100” < “101 to 1,000” < “1,001 to 5,000” < ...

Do the colors reflect that ordering?

Does the ordering in the legend reflect that hierarchy?

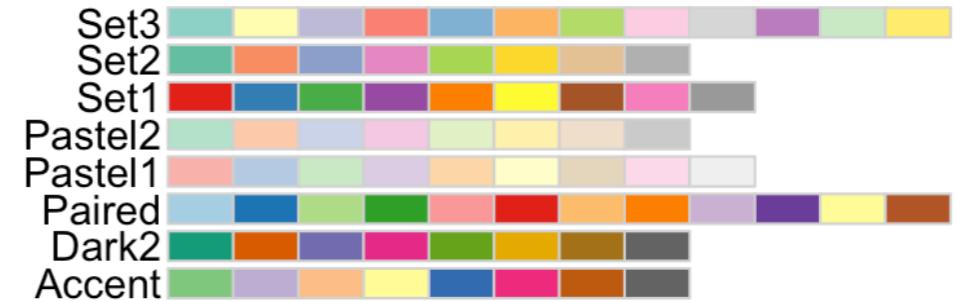
# Color palette should match the type of variable



Numerical/Continuous variable -> Use continuous color palette



Discrete/categorical variable -> Use discrete color palette



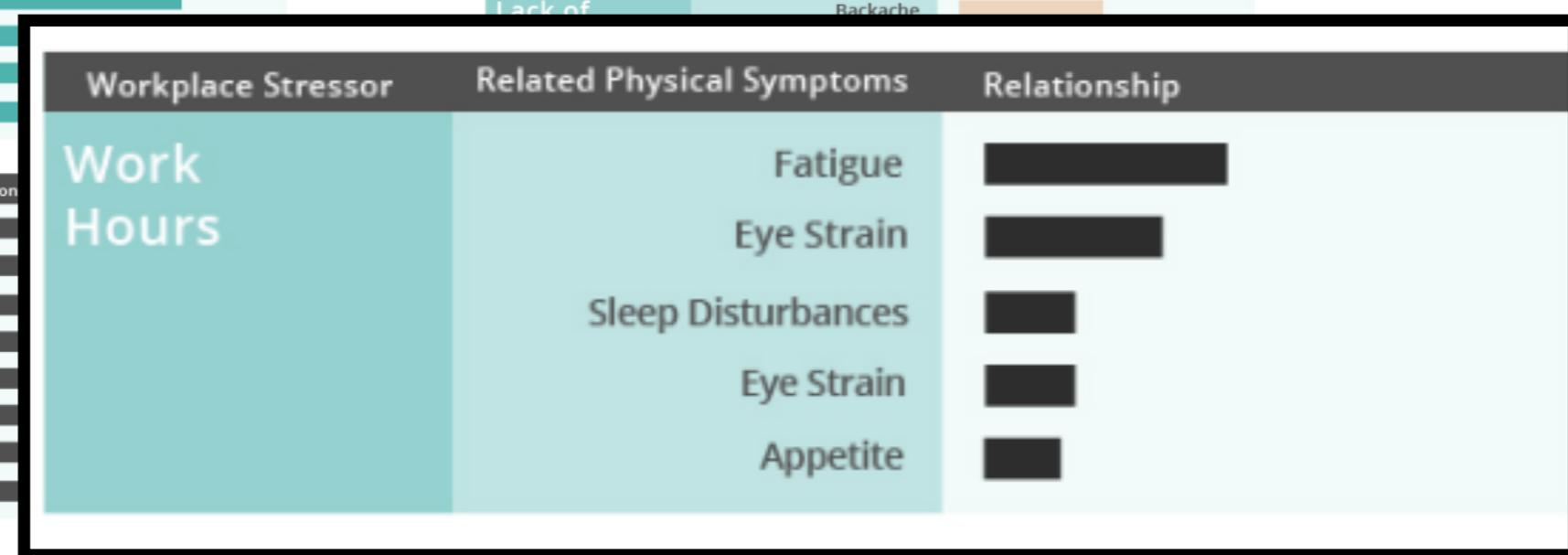
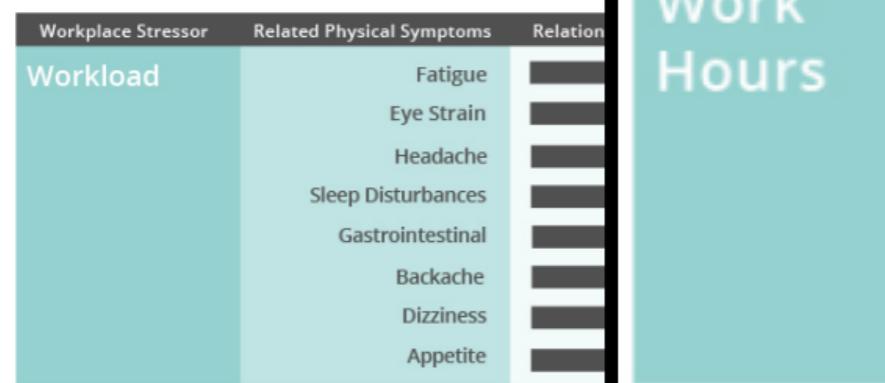
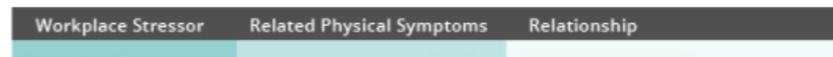
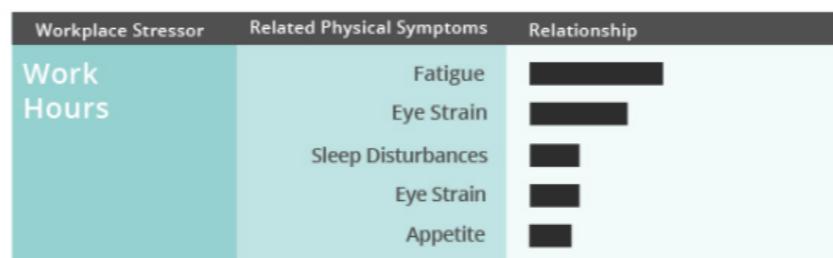
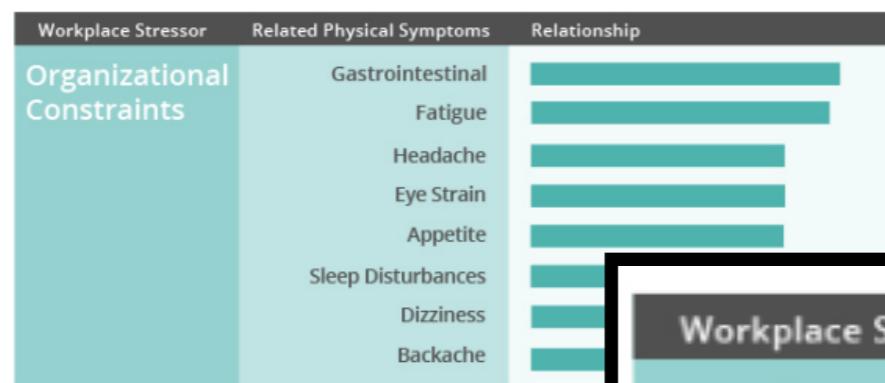
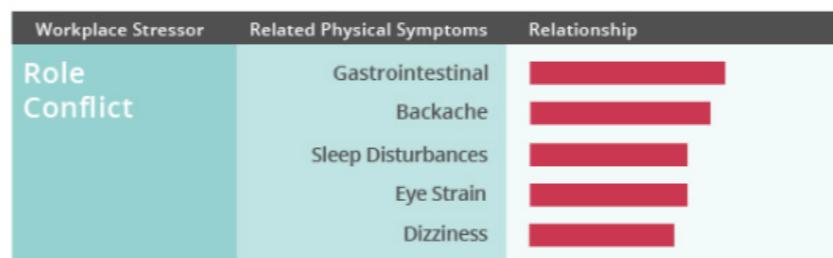
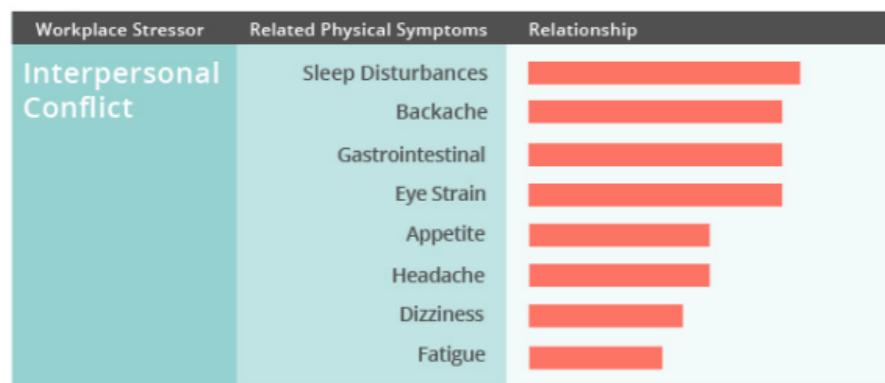
**1 in 20 people are color blind in some way**

1. Use color blind friendly discrete & continuous palettes
2. View plots with a color blind simulator

# Labels should be clear & easy to read, but not overly simplified

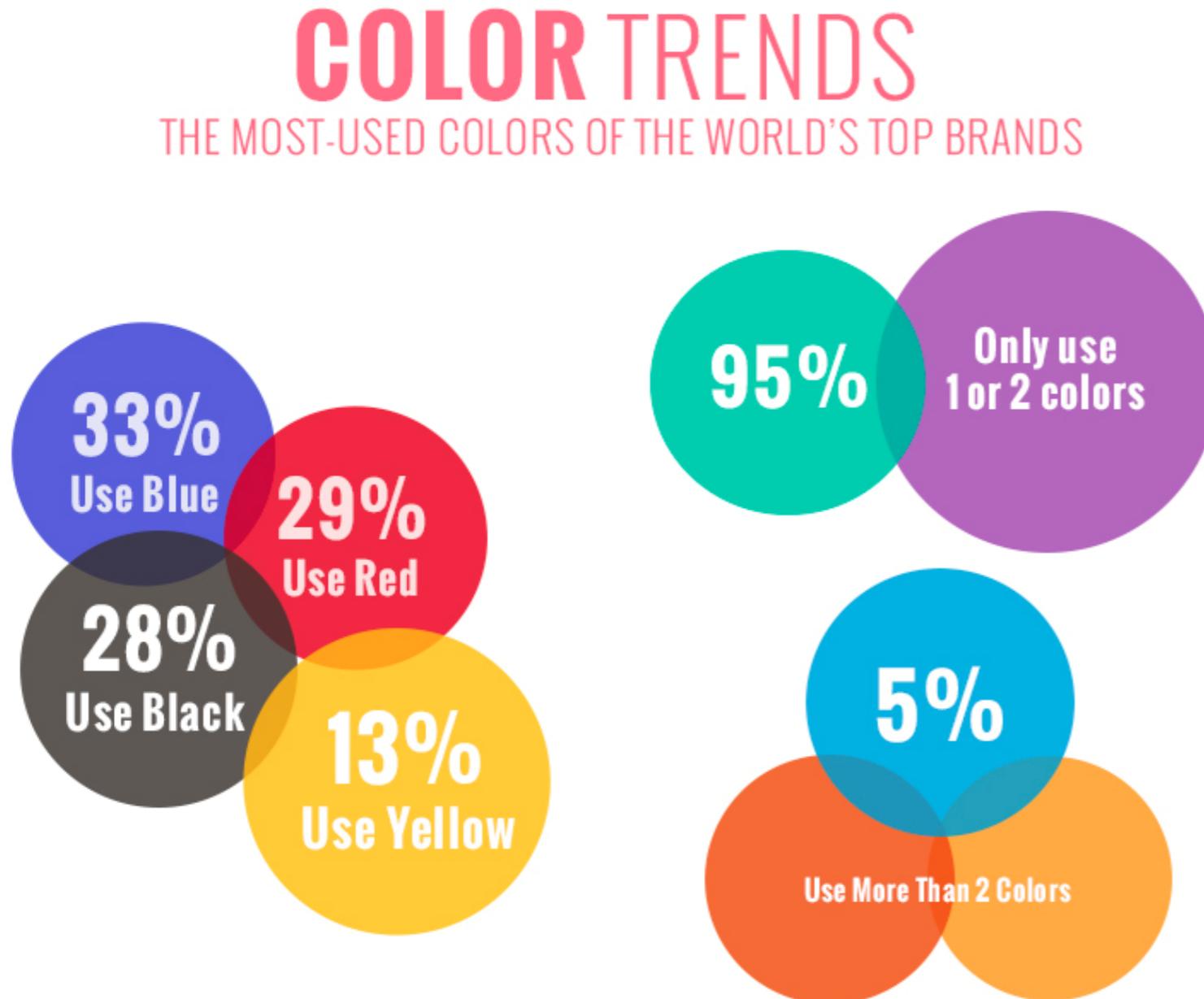
## WORK PLACE STRESSORS AND THEIR RELATIONSHIP TO PHYSICAL SYMPTOMS:

A comparison of workplaces stressors and what health symptoms they cause.



Source: Excerpt from [full infographic](#)

# Leave meaningless details out



**Meaningless details distract the reader from the main message**

What does the size of the bubbles have to do with the percentages inside the bubbles?

What do the colors mean?

Are these Venn diagrams?  
What does the overlap between circles mean?

Common visualization aesthetics that could be used in a meaningless way: **color, size, shape, annotations**

# How to proof read plots

1. **Plot type.** Is the type of plot appropriate for the type of data shown?
2. **Clutter.** Is this plot overwhelming? Are the number of variables shown manageable?
3. **Legend.** Is the legend manageable? Can the legend be replaced with direct labels on the plot?
4. **Color.** Do the colors make sense for the type of variable they represent? Does there even need to be color?
5. **Labels.** Are the labels clearly shown and accurate?

*There are more things to think about, but this should cover the most essential!*