Commonly Used Epidemiology Statistics in the Time of COVID-19 And What They Do and Do Not Mean

To be updated periodically: Updated 08/28/2020

Collin Catalfamo, MPH [1], Jordan Arias, MPH [1] Mark Nichter[2]

[1] Mel and Enid Zuckerman College of Public Health, University of Arizona

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

The Natural Progression of an Infectious Disease

Latent Period

- The time interval between exposure to the pathogen of interest (i.e. a virus or bacterium) and being capable of transmitting the disease.

Incubation Period

- The time interval between exposure to the pathogen of interest and the appearance of the first symptom.

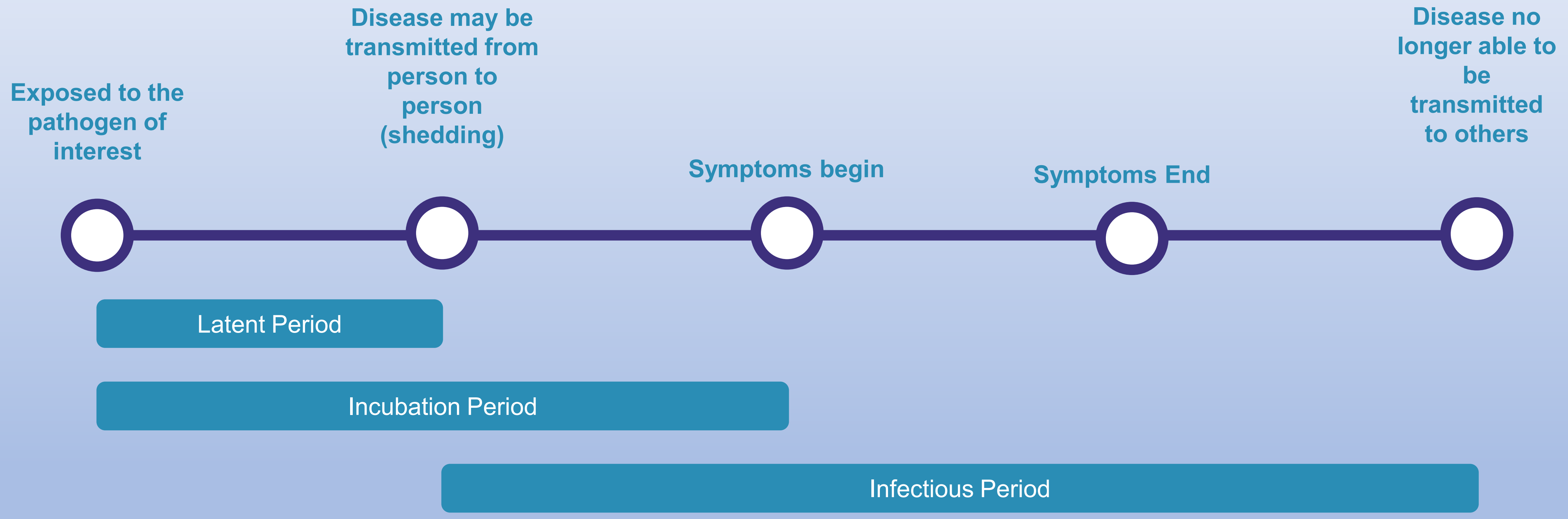
Infectious Period

- The time interval in which transmission between individuals is possible.

****When the incubation period is longer than the latent period, you can infect others before you know you are sick!**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

The Natural Progression of an Infectious Disease



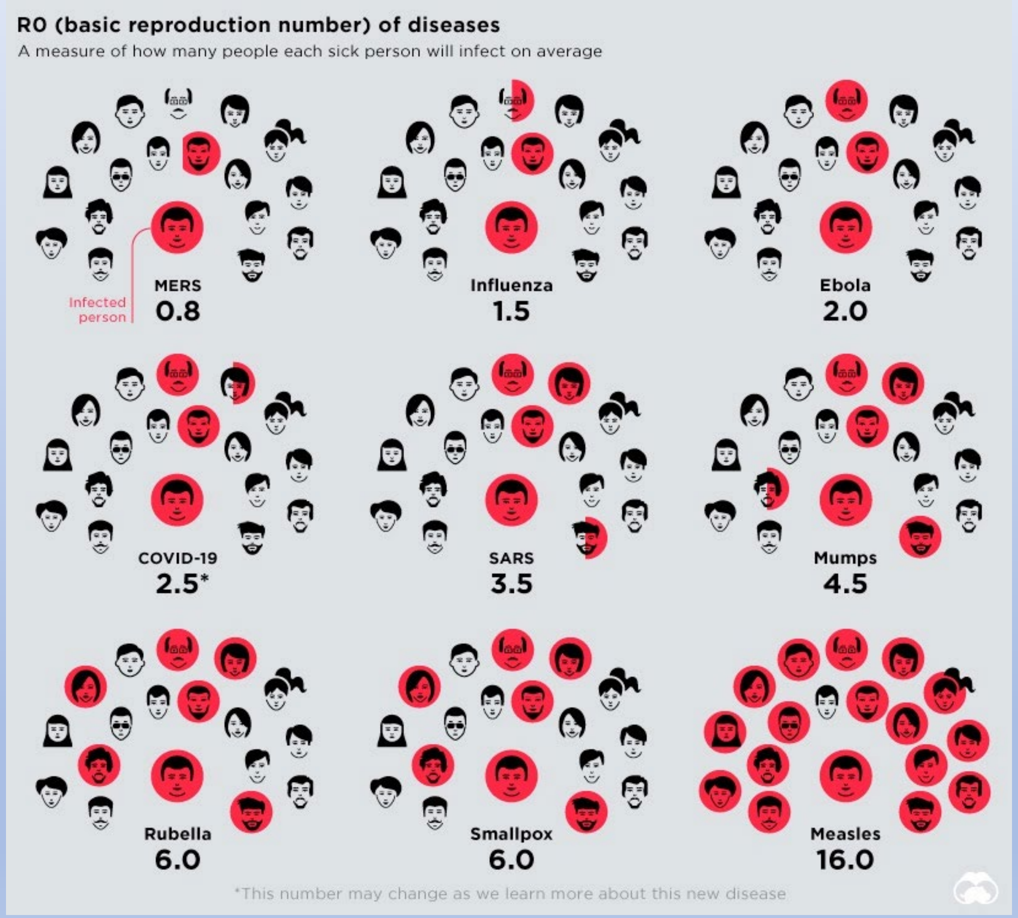
HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Basic Reproduction Number (R_0) (How contagious is a disease)

Pronounced 'R naught' (R_0)

This number represents the average number of people a sick person will infect.

R_0 may differ for each disease!



HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Incidence vs. Prevalence

Incidence

- The number of *new* cases of a disease divided by the average population at risk during a specific time period.

  An example of **incidence** is when individuals talk about "daily cases", i.e. *the number of new COVID-19 cases over a one-day period*

Prevalence

- The number of *existing* cases of a disease in a population at a given time period.

  An example of **prevalence** is when individuals talk about the total number of cases we have had.

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Mortality Rate

The **mortality rate** is the number of deaths that occur in a population during a given time period.

$$\frac{Number\ of\ deaths\ in\ a\ given\ time\ period}{Total\ population\ at\ midpoint\ of\ time\ period} \times 100{,}000$$

This tells us how many people in a population are dying.

This rate can be broken down by specific causes, age groups, sexes, and racial/ethnic groups.

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Case Fatality Rate

The **case fatality rate** is the number of deaths caused by a disease among those who have the disease.

$$\frac{Number\ of\ deaths\ due\ to\ disease\ 'X'}{Number\ of\ cases\ of\ disease\ 'X'} \times 100$$

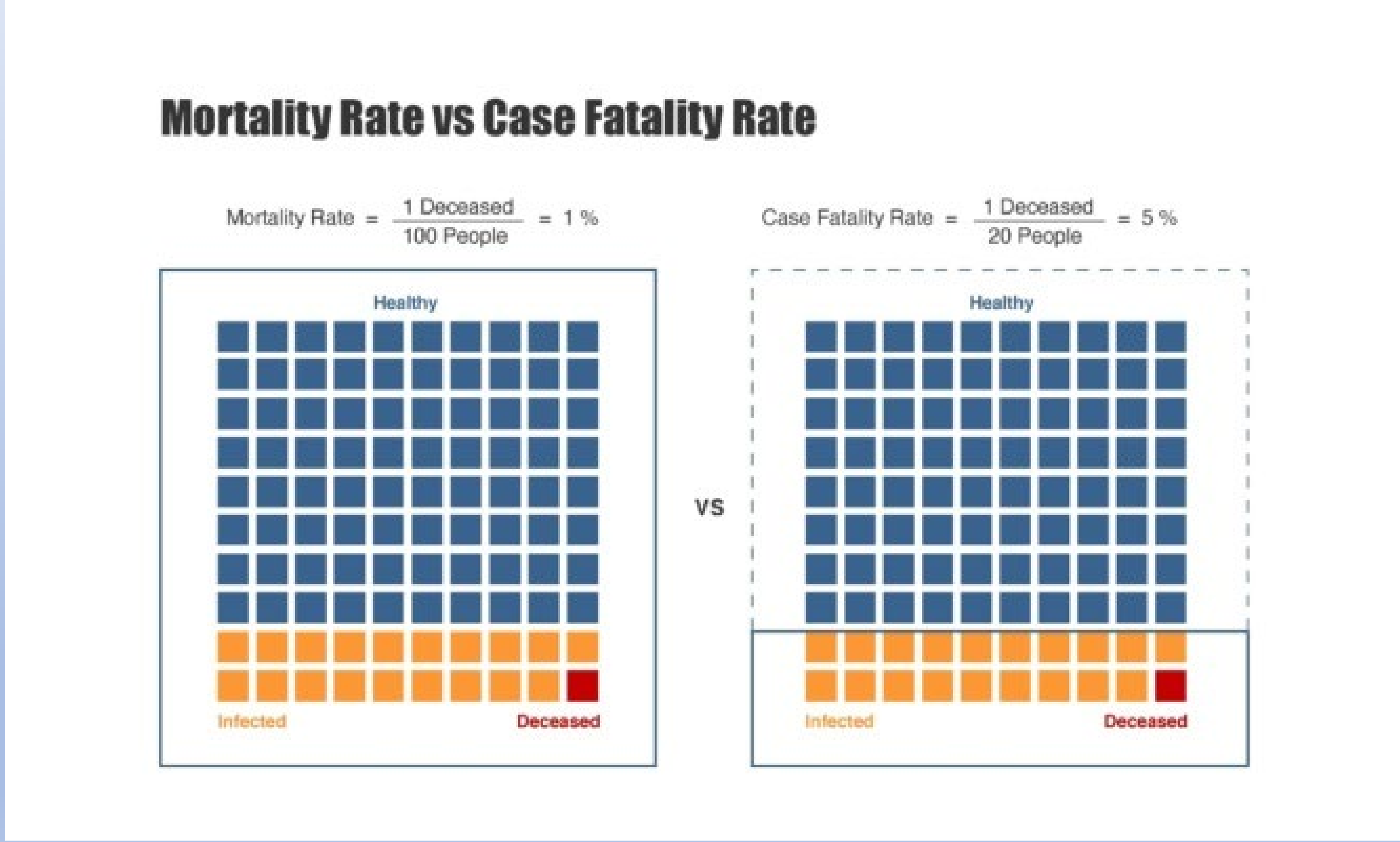
This tells us how deadly the disease is.

Mortality Rate vs. Case Fatality Rate

It is important to differentiate between the case fatality rate and the mortality rate.

The (cause-specific) mortality rate is the proportion of the total population that died from the disease over time.

The case fatality rate is the percentage of those infected with the disease that died due to the disease.



HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Which statistic is more appropriate?

Mortality Rate vs Case fatality Rate

- It is likely you have seen both of these rates used to either support precautions in place or to downplay the severity of the COVID-19 outbreak.
- **Look for a Case Fatality Rate (CFR) when discussing the severity of the disease for those who acquire the infection.**
  - The CFR tells you the proportion of people that die after acquiring the infection.
  - I.e. **How many people died out of the total number of people that contracted COVID-19**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Which statistic is more appropriate?

- A **Mortality Rate** (sometimes referred to as a death rate by non-epidemiologists/public health researchers**) is NOT an appropriate statistic when evaluating *how severe COVID-19 is if you contract the illness*.**
  - **This is because the mortality rate tells you the proportion of people that die from the infection out of the total number of BOTH healthy persons who never contracted COVID-19 and those who did contract COVID-19.**
- **If a report is discussing the severity of COVID-19 and uses a Mortality Rate (or death rate) instead of a Case Fatality Rate (CFR),**
  - **It is likely someone is intentionally trying to mislead you into thinking the outbreak is nothing to worry about.**

Which statistic is more appropriate?

- **Consider the following example**:

  Suppose there was a car crash on the freeway and investigators wanted to determine how severe the incident was so that they can decide whether or not to make swift changes to prevent similar accidents from happening in the future.

  > In this accident, there were 4 people in the two cars involved and 3 of them died.

Which statistic is more appropriate?

- The first investigator decides to calculate the severity of the accident by finding the proportion of people that died in the accident and comes up with a percentage of 75%, concluding it is important to prevent these types of accidents in the future.
- The other investigator decides to calculate the severity of the accident by finding the proportion of people that died in that accident out of everyone that drives (whether they were in the accident or not) and comes up with a percentage of 1%, concluding that this type of accident was NOT severe so no changes need to be made.
  - I.e. 3 people died from this accident out of the 300 people that drive in this area (including the drivers and passengers in the accident as well as those who were not).

Which statistic is more appropriate?

- **Which investigator's method is correct?**

- **The first investigator was correct.**
  - **If you were concerned with how severe that accident was, you would <u>NOT</u> include anyone who was not involved in the accident in your calculation.**

- The same can be said when talking about the severity of COVID-19.
  - **If you want to figure out how severe the infection is to people who acquire it, you wouldn't want to include people who have not acquired the infection in your calculation.**
  - Doing so underestimates how lethal the infection is which can cause people to garner a false sense of safety.

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

When are mortality rates useful?

- **Mortality Rates** can't tell you how severe COVID-19 is once you acquire the infection, BUT they can be used to tell us **how severe our outbreak is or how widespread they are.**
- One commonly used mortality rate for COVID-19 is the **"death rate per capita"** (or a standardized mortality rate). These look at how many deaths have occurred in a country but are adjusted to a "per person" scale.
- How does this actually work?

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Standardized Mortality Rates: “death per capita”

- Let’s walk through an example:
- Below are the daily death rates (or mortality rates) of India and the United States from COVID-19

USA	**INDIA**
909-910 deaths/day	965 deaths/day

- So India’s outbreaks must be much more widespread and larger, right? **Not quite.**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Standardized Mortality Rates: “death per capita”

- We need to consider the number of people that live in the US and in India to make an accurate comparison. This is because we are wanting to look at the *scale of the outbreak* and not how deadly/severe it is.

- The US has a population of around 330,000,000 people while India has a population of 1,380,000,000 people. Because these are two VASTLY different sizes, we make a *per capita* rate. I.e. “how bad is it per person”

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Standardized Mortality Rates: “death per capita”

- The US’s death per capita rate would be:
  909 deaths per day/330,000,000 people in the US = 0.00000275

- India’s death per capita rate would be:
  965 deaths per day/1,380,000,000 people in the US = 0.000000699

- These are very small numbers, so instead of “per person” we’ll adjust this to be “per 1,000,000 people”.

Standardized Mortality Rates: “death per capita”

- Once we increase our scale, the US’s rate per 1,000,000 persons comes out to:
  2.75 deaths per 1,000,000 people

- And India’s death rate per 1,000,000 persons comes out to:
  0.7 deaths per 1,000,000 people

- What does this mean? Between 2-3 people out of every million people die each day in the US from COVID-19. **That is more than 3 times the rate of deaths from COVID-19 in India.**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Mortality and Case Fatality Rates

- Both rates can be misleading if not used to answer the types of questions they are not meant to be used with.
  - **"per capita" mortality rates** are really effective at answering "how widespread/severe is the outbreak?"
  - **Case fatality rates (CFR)** are effective at answering "how severe is the infection if I acquire the infection?"
- If a person uses either of these statistics to answer different questions (I.e. A mortality rate to describe how severe the infection is, or a CFR to describe how widespread the infection is), **that person is likely trying to mislead you**

Testing statistics and rates

- There are **four types of tests** used to determine if someone is currently infected with the virus that causes COVID-19 and if they have ever been infected with the virus.
  - **PCR tests (a swab or spit test) and antigen tests (a rapid test) can determine whether a person is ACTIVELY infected. These are DIAGNOSTIC tests.**

    These kinds of tests are used primarily to help determine INCIDENCE

  - **A serology/antibody test (a blood sample) determines whether a person has EVER HAD the infection.**

    These kinds of tests are used primarily to determine PREVALENCE

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Testing statistics and rates

- **For each of the tests, there are two statistics to pay attention to**

➢ **One is the percent positive, and the other is the testing rate.**

  - **Percent positive** is the percentage of tests that are positive out of the total number of tests completed.

$$Percent\ Positive = \frac{Number\ of\ positive\ tests}{Total\ number\ of\ tests\ performed}$$

  - **Testing rates** are the number of tests completed divided by the population, and then multiplied by 1,000,000 (this makes the rates of different population sizes comparable)
    - Testing rates are discussed in more detail in the one of our other presentations

$$Testing\ rate = \left(\frac{total\ number\ of\ tests\ performed}{total\ number\ of\ people\ in\ the\ population}\right) * 1{,}000{,}000$$

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

What do these rates/statistics mean for each of the tests used currently?

DIAGNOSTIC Tests

PCR Tests (the swab or spit) & Antigen (rapid tests)

- For **DIAGNOSTIC** tests, the **percent positive** gives us the percent of people who were tested via a swab, spit, or finger prick who have **ACTIVE** infections.
  - **I.e. Out of everyone who was tested, __% currently have COVID-19.**
- **Testing rates** for **DIAGNOSTIC** tests tell us how many people in the population are being tested for **ACTIVE** infections.
  - I.e. **How many people in the population are being tested to see if they currently have COVID-19**

Serology Tests (the blood sample)

- For serology tests, the **percent positive** gives us the percent of people who tested positive for the presence of antibodies to COVID-19.
  - **I.e. Out of everyone that was tested, __% have had a PAST infection with COVID-19.**
- **Testing rates** for serology tests tell us how many people in the population are being tested to determine if they had a **PAST** infection with COVID-19.
  - **I.e. How many people in the population are being tested to see if they have ever had COVID-19**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Testing statistics and rates

- The **percent positive** is dependent on both the testing rate and how many people are actually sick or have had the infection.
  - For **DIAGNOSTIC tests (the swab, spit, or rapid tests)**, a community with a very high **testing rate** will likely have a lower **percent positive** unless they are experiencing high rates of community transmission of the virus. This is because they are completing a high number of tests (which increases the denominator for the calculation)

  - The **percent positive** for **serology tests (the blood sample)**, while still being influenced by the **testing rate**, is largely dependent on **how many people contracted the infection**.
    - If very few people contracted COVID-19 in a community, then their **percent positive** will be LOW, whereas a community in which a large proportion acquired the infection will have a HIGHER **percent positive.**

I've heard both the percent positive and the testing rate used to criticize or praise our testing program. Which statistic is appropriate to use?

- As impressive as a **testing rate** may be, it tells us *very little* about how successful we have been at containing an outbreak other than we've tested a lot of people.
    - Testing programs need to be scaled to the size of the pandemic, not the size of the population. **The reason the U.S. has needed a huge testing program is because we have still record massive numbers of new cases of COVID-19 daily.**
- The **percent positive** is the statistic to follow when assessing the success of a testing program and measuring the extent of community spread of the virus. The percent positive tells us what proportion of the people we tested have a positive test result (and, therefore, have the infection).
    - If we have a testing program that is scaled up enough to the size of our outbreak and less than 5% of those tests come back positive over a period of two weeks, then we know our outbreaks are under control.
    - **A percent positive greater than 5% is a sign that restrictions should NOT be relaxed because we are still experiencing community transmission.**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Testing statistics and rates

- **DIAGNOSTIC tests (the swab, spit, or rapid tests)** are helpful in determining where we are currently at in the outbreak because they are important in calculating **incidence rates**, which **tell us how many new cases of COVID-19 we are experiencing**.
- **Serology tests (the blood sample)** are helpful in determining **prevalence rates**, or how many people have ever had COVID-19. **These rates are crucial to determining how close we are to achieving *herd immunity*.**

- **Because these tests are measuring different aspects of the outbreak, they should not be combined into one statistic.**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Limitations of tests

- **Testing only gives us information about a particular snapshot in time**.
  - **This is why it is important to continually retest members of the population to fully understand the outbreak.**

  - E.g. Just because someone tests negative to a DIAGNOSTIC test (**the swab, spit, or rapid tests**) or serology test (**the blood sample**), this does not mean that they will never catch COVID-19.
  - This also <u>DOES NOT</u> mean that if you test negative once and then test positive a second time that the tests are faulty, inaccurate, or misleading. More likely than not, this just means that you managed to acquire the infection at some point in time after the first time you were tested.

Rates and statistics are likely to change as outbreaks progress and as new science emerges.

- ***Science is an iterative process***. When new information and data becomes available, estimates of these different statistics and rates will be updated to reflect the new information.

  - This does not mean that scientists and researchers were "wrong" and/or were trying to mislead you.

  - Rather, they made estimates based on the best information that was available at that point in time.

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Misinformation cloaked in statistics about COVID-19 is rampant: Look at the data presented with a critical eye

- **A more common (and worrying) trend that you should look out for in and statistics is the combination of both PCR (the swab) and serology (the blood sample) test statistics into one category.**
- Pennsylvania, Texas, Georgia, and Vermont have been combining the testing results from PCR (**the swab or spit**) tests and serology (**the blood sample**) tests to form one metric.
- **This is done for only one reason:** to make testing rates appear higher and to make the outbreaks appear smaller.
    - Remember, both of these tests present ***different*** information that cannot and should not be combined. PCR (**the swab or spit**) tests tell us *how many new (current) infections* there are. Serology (**the blood sample**) tests tell us *who has had the infection in the past*.
    - Because of these distinctions, PCR (**the swab or spit**) tests generally have HIGHER **percent positive** rates whereas serology (**the blood sample**) tests have MUCH LOWER **percent positive** rates.
    - **Combining the two tests is *extremely inappropriate* and removes any ability to evaluate how well your community is performing in testing efforts and the progression of the outbreak.**

HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

How to effectively read graphs and figures

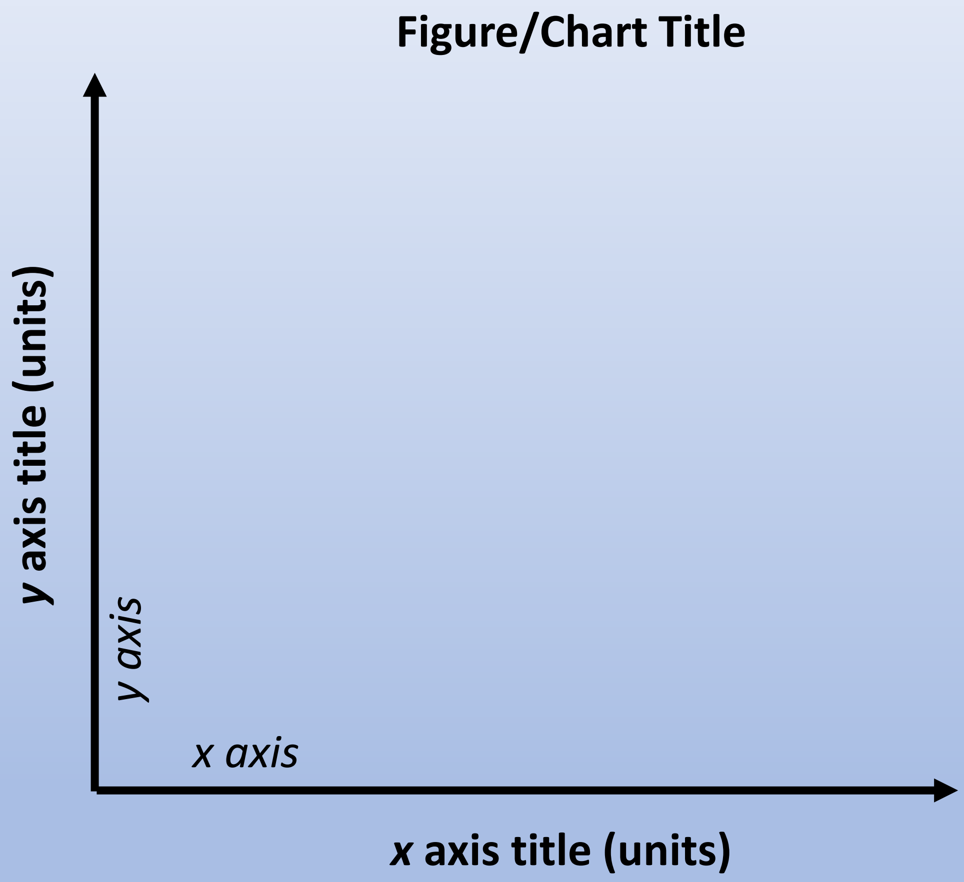
Overview

- Often, scientists and other individuals reporting on epidemiologic data will opt for the use of a figure or a graph to visually portray patterns observed as opposed to using a table. When done correctly, these are much easier for individuals who are not accustomed to working with these types of data to view and understand patterns compared to tables.
- This next section will discuss what to look out for when reading a graph or figure so that you are able look at them with a critical lens. Ultimately, this will allow you to evaluate the data presented in graphs and figures and help you determine if the authors of the graphic are intentionally manipulating the presentation of the data to tell a story contrary to the one the data actually tells.

Titles and Captions

- The first components to look for with any figure or graph are the Title, axes, and any axes titles that are present. The purpose of these are to provide context for the data that is presented.

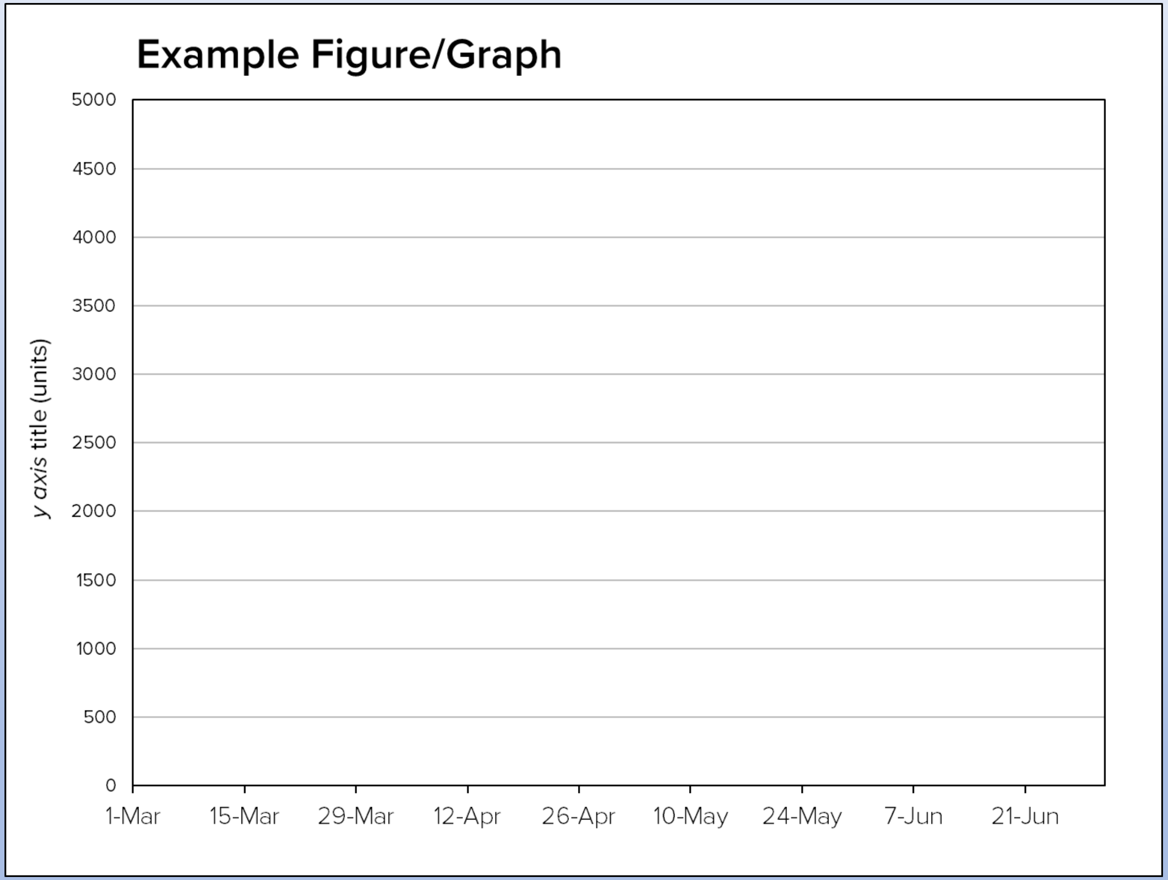
- These should be informative enough that it you are able to determine what the types of data are that are present, the units that are used, and give you a hint at the relationship between the two.



HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

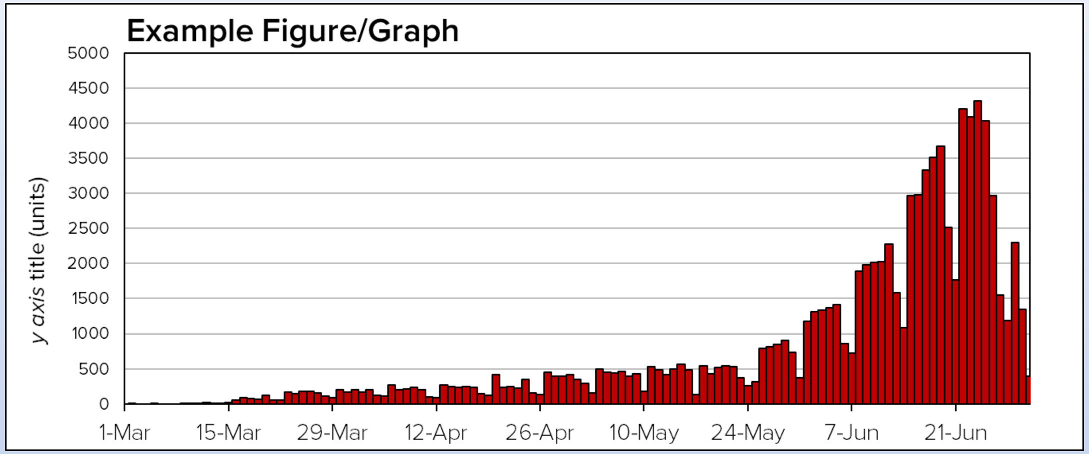
Titles and Captions

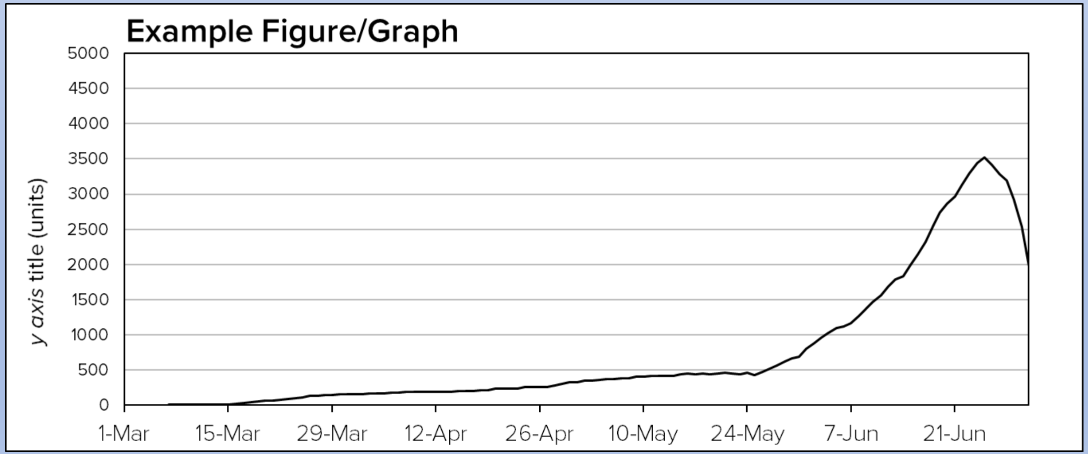
- In outbreaks and the time of COVID-19, the *x axis* is commonly representing time so will display dates.
  - This is to help portray how the outbreak has evolved as it progresses through time.
  - If using a Date format, these will not have a *x axis title* as it should be self-explanatory
- **Time should ALWAYS progress from the past to present in sequential order** when using time as the *x-axis.*
  - NOTE: The example starts March 1st and moves to June 21st in the correct order



Common Types of Graphs using Time

- There are two types of plots that use time as an *x axis* commonly used: a bar chart and a line graph.
  - Bar charts (shown in the top chart) are used to display **counts**
  - Line graphs (shown in the bottom chart) are used to better display **trends.**
- **These two types can be plotted together because they are closely related.**
  - I.e. it can be helpful to display a trendline on top of daily cases so that we can see how many cases we see each day, but also the overall trend in cases day to day.

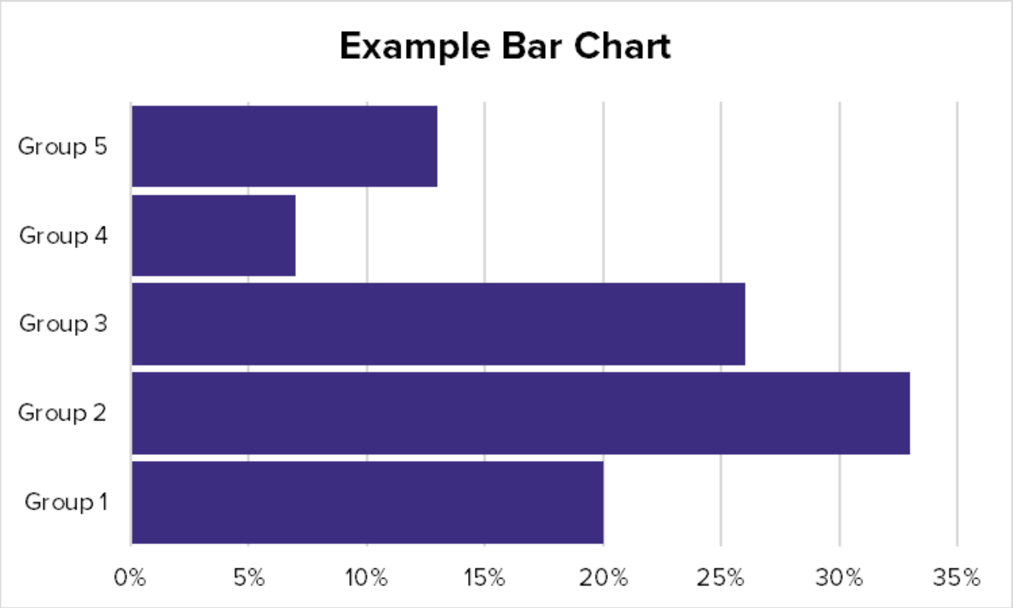


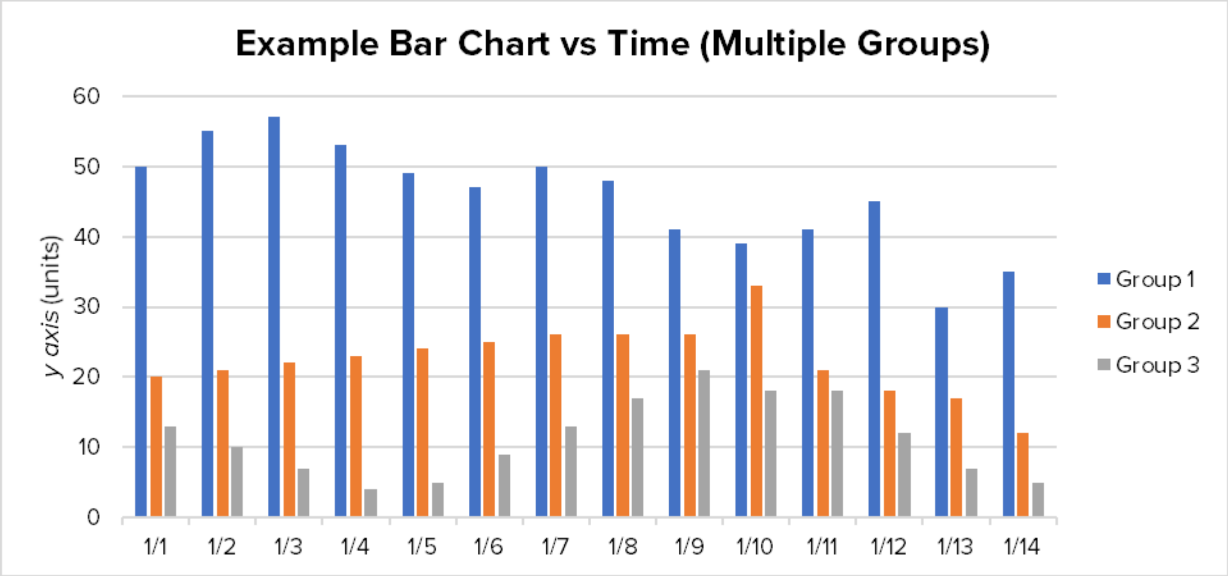


HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families

Bar Charts Hints

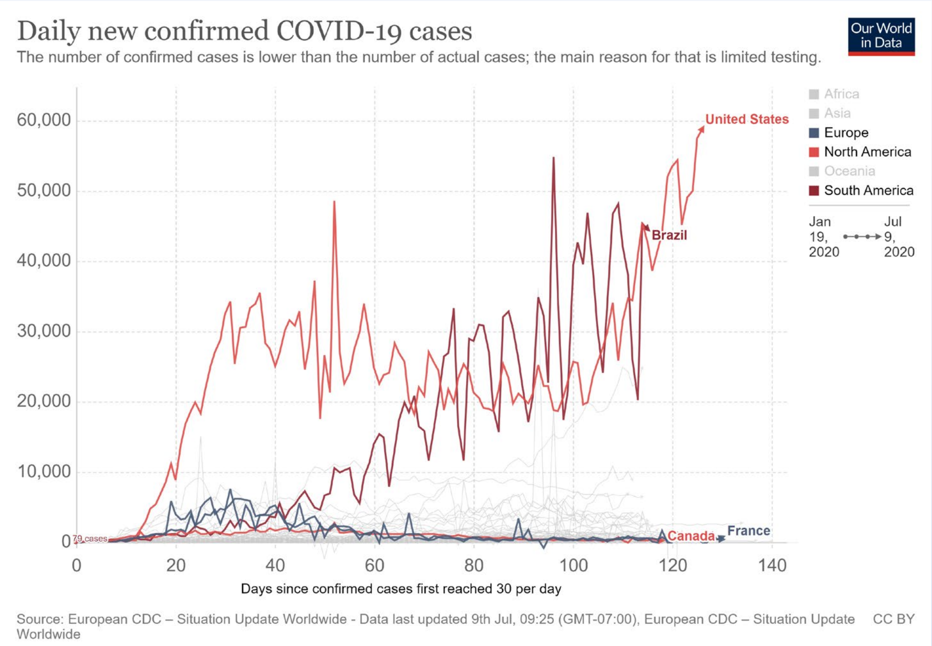
- Bar charts can also be used to compare different groups to each other without time as an axis.
  - NOTE: The example to the top right is an example of a horizontal bar chart. If a bar chart is comparing groups and not using time, bars can be horizontal or vertical.
- The example to the top right displays data to **compare** a **count** (in this case a percent) across different **groups**.
- If multiple groups across time (like the bottom right chart) are being compared, the ordering of groups should be consistent at each time point.
  - A change in organization at different time points could be a sign that the author of the figure is trying to mislead you

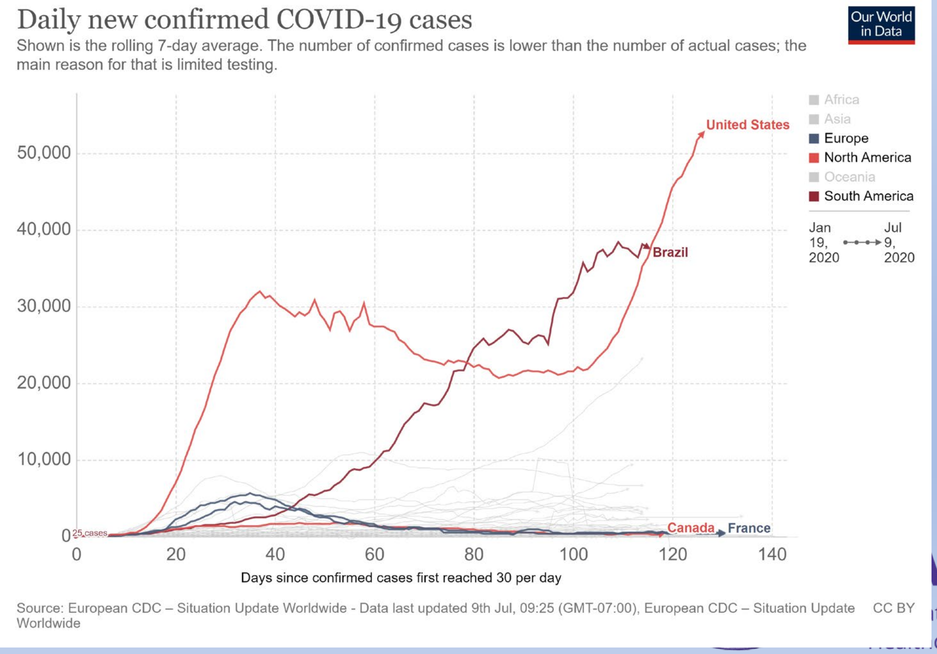




Line Graph Hints

- Because line graphs are used to display trends over time, sometimes the lines can vary in appearance and be difficult to read.
  - For example, a graph showing the percent positive over time may be very smooth because as time goes on there aren't drastic changes day to day.
  - A graph showing the daily number of cases can be hard to read because of the drastic changes that can occur (shown in the top right).
- In line graphs showing cases over time, **a common way to better show the trend in cases is to plot a rolling 7-day average** instead of the daily number (shown in the bottom right)
  - This is done by averaging the past 7 days at each point. **Not only does this smooth out the line, but it better captures the trend in increase or decrease of cases.**





Axis Units

- It is always crucially important to take note of the units used for the data that is being presented. If different territories, counties, cities, states, or countries are being compared **you want to make sure that the data is comparable**.
- One way that epidemiologists do this is by creating a **per capita rate**. A **per capita rate** is just the number of, for example, new cases identified or tests completed per person in that population. **This gives us a rate of increase or decrease in cases or tests that has been adjusted for the population, allowing us to compare between populations of different sizes.**
  - **Most commonly you will see these depicted as "per 1,000,000 people" or "per 100,000 people".** This just means that instead of showing you the rate per 1 person, they increased it to make the rate more interpretable. Sometimes per capita rates can be small, so increasing the rate to a larger group of people makes them more easy to understand.
  - **A graph comparing absolute numbers from populations of different sizes can be misleading**. For example, countries that are larger often will have more cases than countries that are smaller because they have more people. A more accurate and correct comparison would the number of cases per person, or per capita.

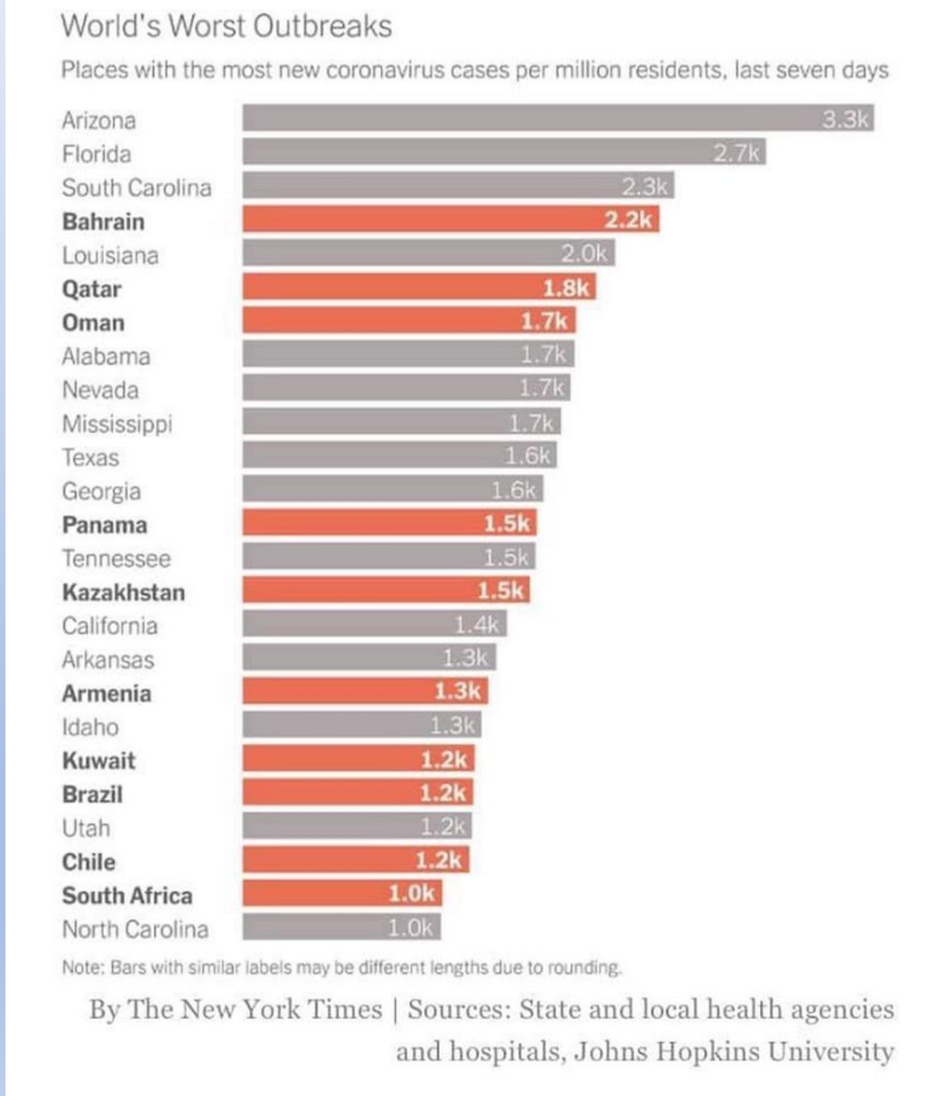
Examples

- The next set of slides will show you real world examples of graphs that have been published including a critical evaluation of each.

- These have been selected to show you strengths and limitations, as well as to reinforce the importance of evaluating these with a critical eye so that you do not fall victim to those wanting to mislead you.

Pay attention to the titles and captions!

- The chart to the right shows **a per capita rate** ("per million residents") of COVID-19 cases for each of the countries/States listed.
- After reading the figure caption we can see that the authors opted to use a 7-day total rather than the more commonly reported 7-day rolling average of daily cases.
  - Without carefully reading the titles and captions, at first glance, this would make the outbreaks appear many times larger than they are.



STED
Support for
Families

Why is a 7-day average preferred over daily totals?

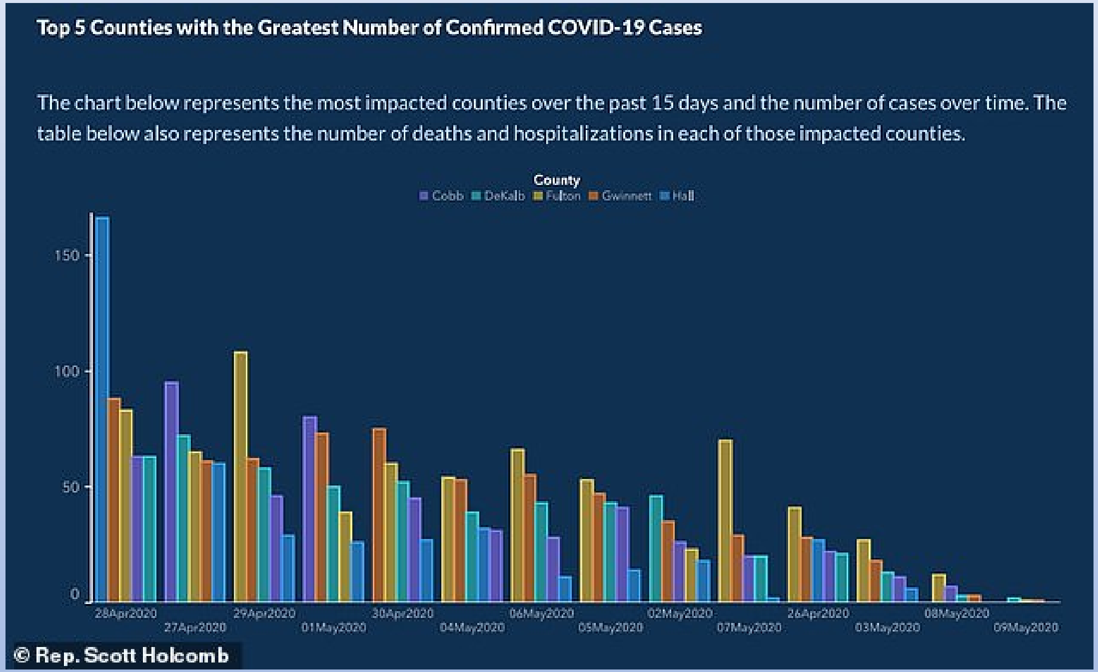
- **There is often a high variability in the number of cases reported day by day.**
  - This is even further complicated by the fact that if often takes 1 week or longer for complete data to be reported for each day. There is also a lag between receiving test results after getting tested (results can come back on average between 2-14 days).
- Because of this, **its recommended that a 7-day rolling average to be used** for COVID-19 cases and other statistics. This allows researchers and public health professionals to characterize the trend in these rates and adjust for significant differences between each day.
  - 3-day rolling averages can also be used if data is reliably and more completely reported in a shorter time frame.

Misinformation through figure manipulation

- **As the outbreaks continue there will be individuals that will try to mislead you to further their own agendas**.
- The image to the right is a clear cut example of this. Here are ways the author of this chart was trying to mislead you:

1. The dates are not in the correct order
2. The order of the bars (groups) changes at each date as to make the outbreak appear as though it were consistently declining.

- Both of these were done to make the outbreak appear as though it were consistently declining



HCW HOSTED
Coordinating Community Support for Healthcare Workers and Families