Presenting Your Research

Rachel Silverman, PhD, ScM

Research Scientist, Epidemiologist

COLLEGE OF SCIENCE
VIRGINIA TECH

CENTER FOR BIOSTATISTICS
AND HEALTH DATA SCIENCE

Basics to consider for data presentation

- Who is your audience?
  - Lay audience?
  - Providers?
  - Researchers?
- What is the medium?
  - Report?
  - PowerPoint?
  - Poster?
  - Pamphlet?
- What type of data?
  - Nominal?
  - Ordinal?
  - Continuous?
  - Temporal?
  - Geographic?

Your answers to these will inform what information you include and how you present the data

COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Basic guidelines for tables and figures

Figures and tables should be:

- Easy to interpret
- Self-explanatory
- Aesthetically pleasing

Keep them simple and to the point

Minimize and define abbreviations

Summarize the key points: What should the audience look for and take away?

Guidelines are helpful but depends on how you want to communicate and what looks best

Presenting Tables

Publication

Title

Column titles are descriptive

Data points are merged when appropriate

Footnotes
(includes abbreviations, other details to fully understand table)

Table 2

HIV incidence following sexually transmitted infections by infection type and associated number needed to treat with pre-exposure prophylaxis to prevent one HIV infection

Panel A: Primary analysis – Incidence of HIV diagnosis among MSM with STIs in Washington State							
	Total number of STIs	No. of STIs based on order*	Subsequent HIV diagnoses	Incidence of HIV diagnosis per 100 person-yrs (95% CI)	Number needed to treat with PrEP by estimate of efficacy^		
					44%	73%	92%
Overall	10080	7064	280	1.61 (1.43–1.81)	141	85	68
Rectal GC	985	871	79	4.13 (3.31–5.15)	55	33	26
Early syphilis	771	675	51	2.83 (2.15–3.72)	80	48	38
Rectal CT	1480	939	29	1.59 (1.10–2.29)	143	86	68
Urethral GC	2087	1537	66	1.63 (1.28–2.07)	140	84	67
Late syphilis	188	155	4	1.04 (0.39–2.78)	218	131	104
Pharyngeal GC	1205	528	11	1.08 (0.60–1.95)	210	127	100
Urethral CT	3124	2283	39	0.62 (0.45–0.84)	369	222	176
Pharyngeal CT	240	76	1	1.08 (0.15–7.65)	211	127	101

STI = sexually transmitted infection. MSM = men who have sex with men. GC = gonorrhea. CT = chlamydial infection. CI = confidence interval. PrEP = HIV pre-exposure prophylaxis.

*STI types are presented in hierarchical order based on prior analyses identifying descending incidence of HIV diagnosis. Men diagnosed with concurrent infections or infections at >1 site were assigned the higher order STI for that time point. Repeat infections of the same type were not included in this count.

^This represents an estimate of the number of HIV-negative MSM with each infection that would need to be treated with PrEP for one year in order to prevent one HIV infection among PrEP recipients based on three estimates of PrEP effectiveness from iPrex[17]: overall (44%), with at least 90% adherence (73%), and with detectable blood levels of PrEP drug (92%).

†The date of HIV infection was estimated as the midpoint between the last negative and first positive HIV test from HIV surveillance, HIV partner services, or STI partner services data.

Katz, et al. (2016)

Tables

What is useful for a website/report/publication may be difficult to look at in a PowerPoint

From CDC Website

Demographic characteristics of COVID-19 cases in the U.S.

	No. of cases (% of total)						
Age group (years)	< 18	18-44	45-64	65-74	75+	Unknown	Total
Totals	18,722	300,445	293,466	91,403	104,879	54,439	863,354
Race missing/unspecified	12,431 (66.4%)	186,923 (62.2%)	170,758 (58.2%)	44,159 (48.3%)	42,718 (40.7%)	18,065 (33.2%)	475,054 (55.0%)
Race specified	6,291 (33.6%)	113,522 (37.8%)	122,708 (41.8%)	47,244 (51.7%)	62,161 (59.3%)	36,374 (66.8%)	388,300 (45.0%)
Among those with race specified							
American Indian or Alaska Native	140 (2.2%)	1,200 (1.1%)	1,003 (0.8%)	269 (0.6%)	184 (0.3%)	140 (0.4%)	2,936 (0.8%)
Asian	291 (4.6%)	5,923 (5.2%)	6,437 (5.2%)	2,143 (4.5%)	2,284 (3.7%)	1,737 (4.8%)	18,815 (4.8%)
Black or African American	1,458 (23.2%)	31,169 (27.5%)	37,939 (30.9%)	14,974 (31.7%)	13,343 (21.5%)	11,730 (32.2%)	110,613 (28.5%)
Native Hawaiian or other Pacific Islander	40 (0.6%)	465 (0.4%)	349 (0.3%)	108 (0.2%)	69 (0.1%)	78 (0.2%)	1,109 (0.3%)
White	3,124 (49.7%)	56,399 (49.7%)	60,415 (49.2%)	25,112 (53.2%)	41,592 (66.9%)	15,943 (43.8%)	202,585 (52.2%)
Multiple/other	1,238 (19.7%)	18,366 (16.2%)	16,565 (13.5%)	4,638 (9.8%)	4,689 (7.5%)	6,746 (18.5%)	52,242 (13.5%)
Ethnicity missing/unspecified	12,260 (65.5%)	194,525 (64.7%)	182,413 (62.2%)	50,709 (55.5%)	53,879 (51.4%)	21,528 (39.5%)	515,314 (59.7%)
Ethnicity specified	6,462 (34.5%)	105,920 (35.3%)	111,053 (37.8%)	40,694 (44.5%)	51,000 (48.6%)	32,911 (60.5%)	348,040 (40.3%)

https://www.cdc.gov/covid-data-tracker/index.html

Last updated on May 3, 2020

Tables

Some of the data

Age group (years)	< 18	18-44	45-64	65-74	75+	Unknown	Total
Totals	18,722	300,445	293,466	91,403	104,879	54,439	863,354
Race missing/unspecified	12,431 (66.4%)	186,923 (62.2%)	170,758 (58.2%)	44,159 (48.3%)	42,718 (40.7%)	18,065 (33.2%)	475,054 (55.0%)
Race specified	6,291 (33.6%)	113,522 (37.8%)	122,708 (41.8%)	47,244 (51.7%)	62,161 (59.3%)	36,374 (66.8%)	388,300 (45.0%)

Less information, a little easier to digest

Tables

Number of Covid-19 Cases in the US by Age and Race Information Availability

Title

Units

Age group (years)	Total	Race missing/ unspecified	Race specified
< 18	18,722	12,431 (66.4%)	6,291 (33.6%)
18-44	300,445	186,923 (62.2%)	113,522 (37.8%)
45-64	293,466	170,758 (58.2%)	122,708 (41.8%)
65-74	91,403	44,159 (48.3%)	47,244 (51.7%)
75+	104,879	42,718 (40.7%)	62,161 (59.3%)
Unknown	54,439	18,065 (33.2%)	36,374 (66.8%)
Total	863,354	475,054 (55.0%)	388,300 (45.0%)

Source information

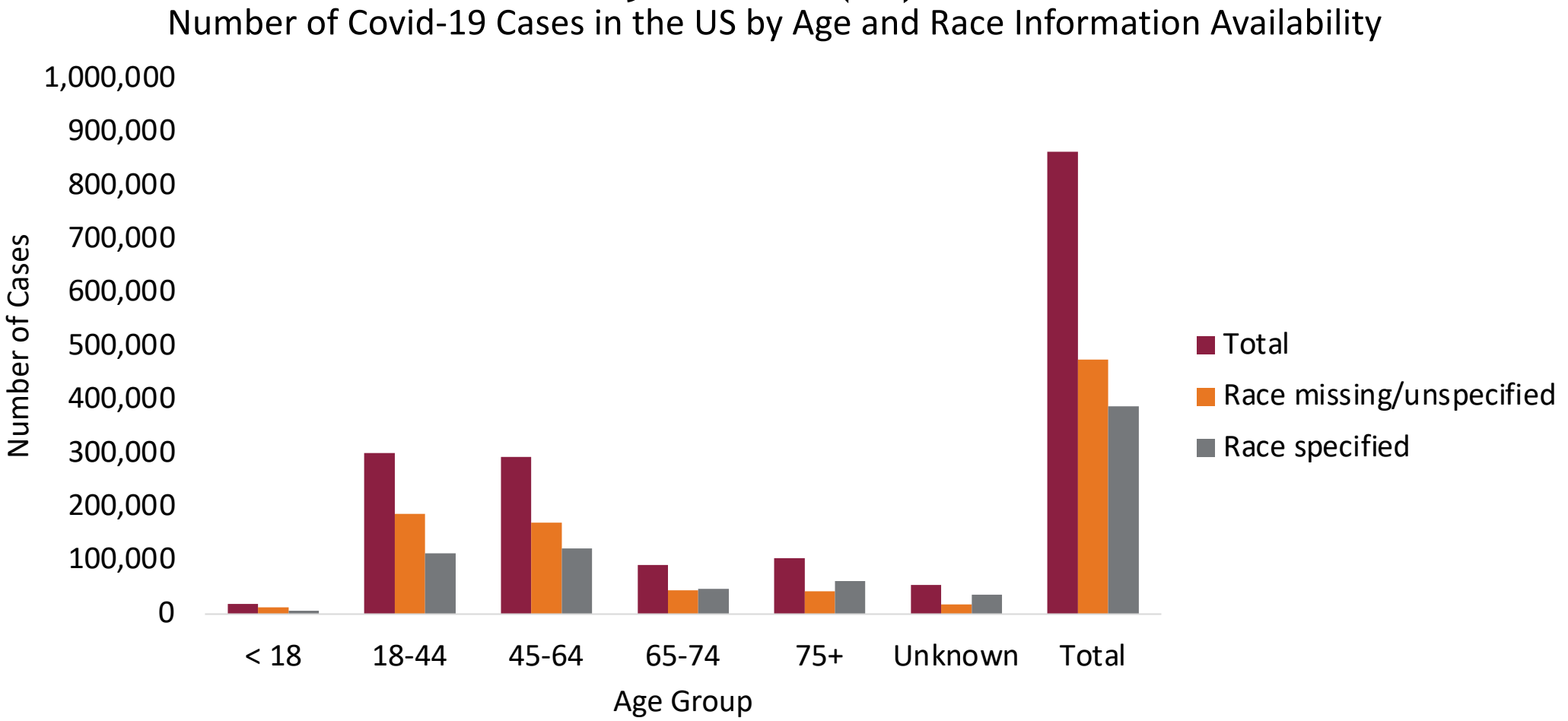
Last updated on May 3, 2020

https://www.cdc.gov/covid-data-tracker/index.html

Same information, easier to digest

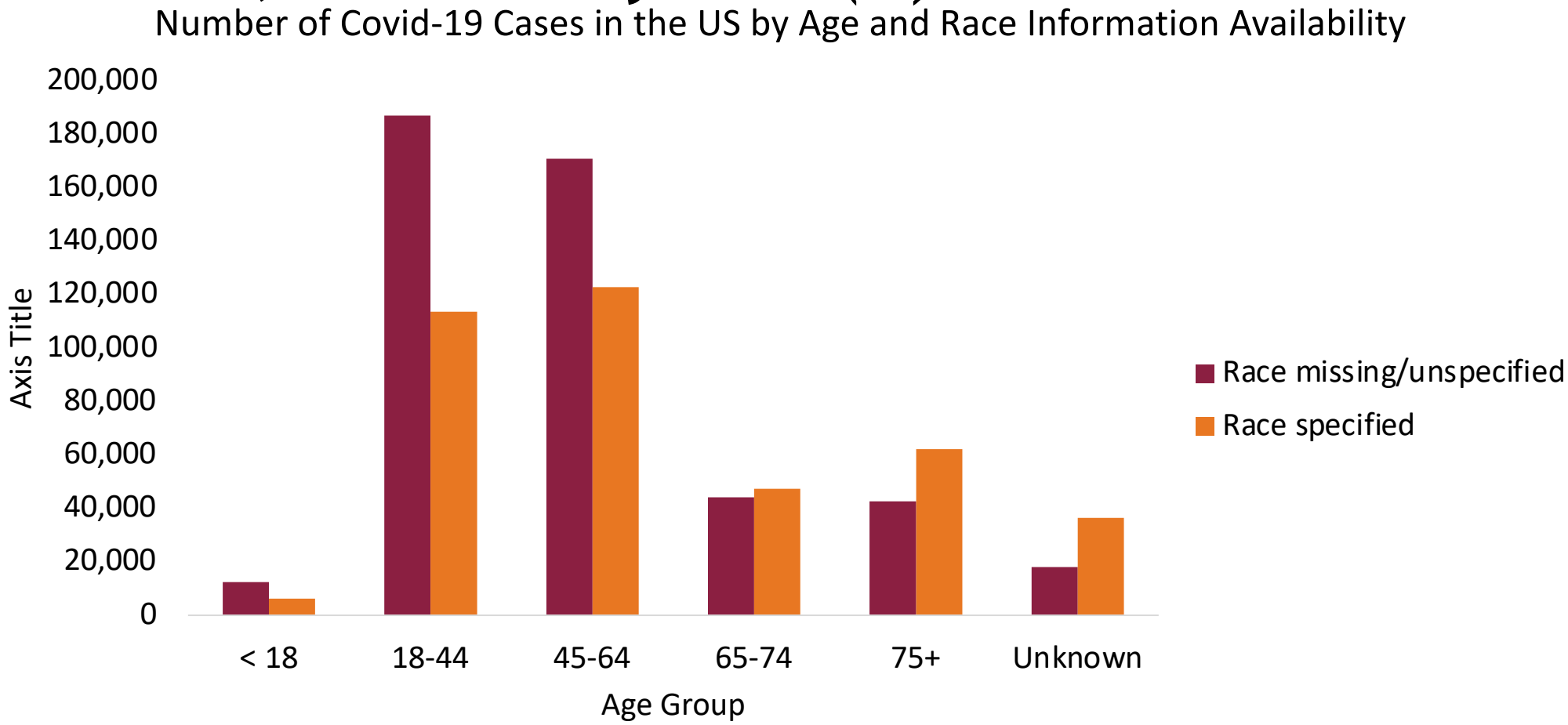
COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Same data, in chart form (1)



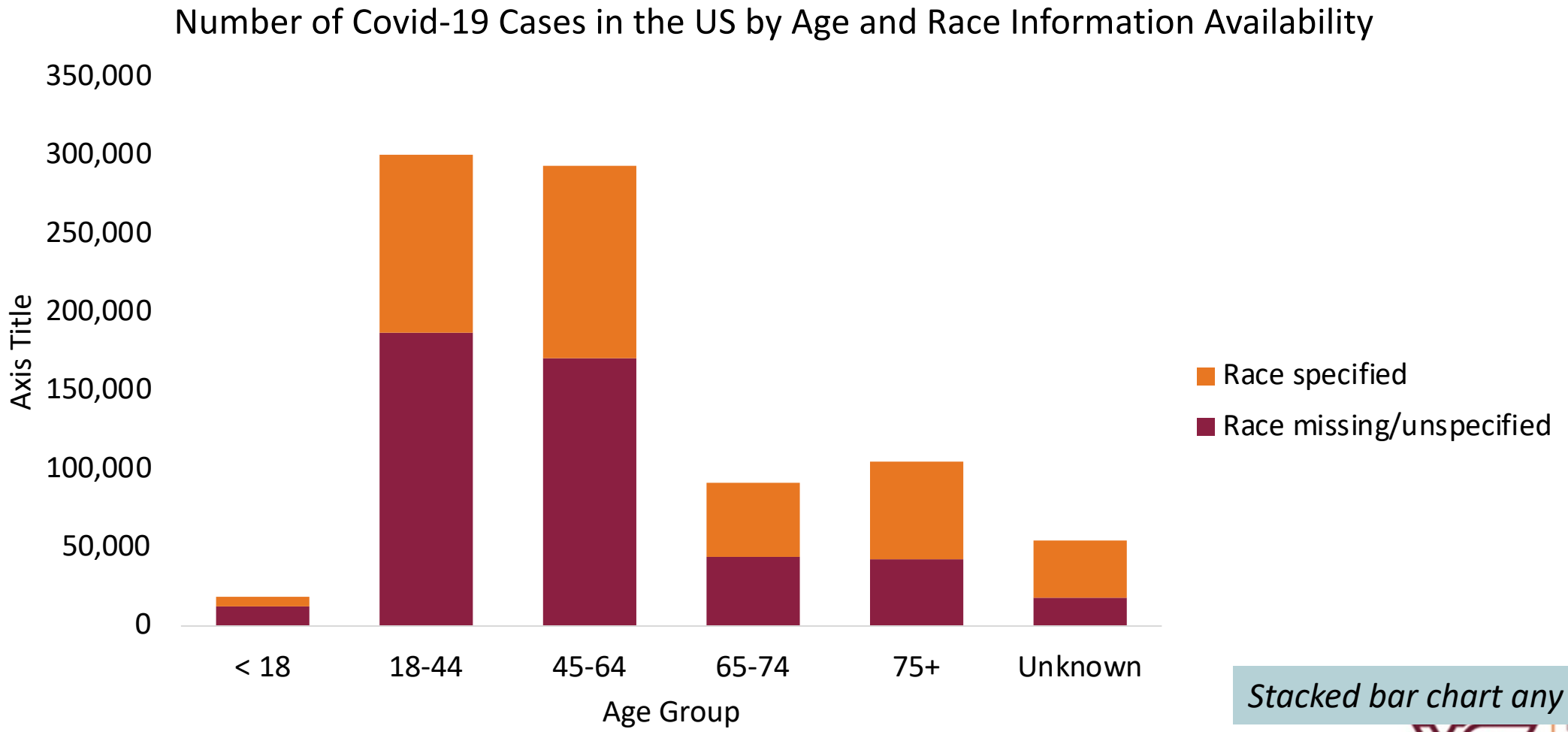
Scale is very large since total counts are included, difficult to compare subgroups (y-axis scale is 0-1,000,000)

Same data, in chart form (2)



Better, but still not super helpful to see relationship since number of people within a group are so different

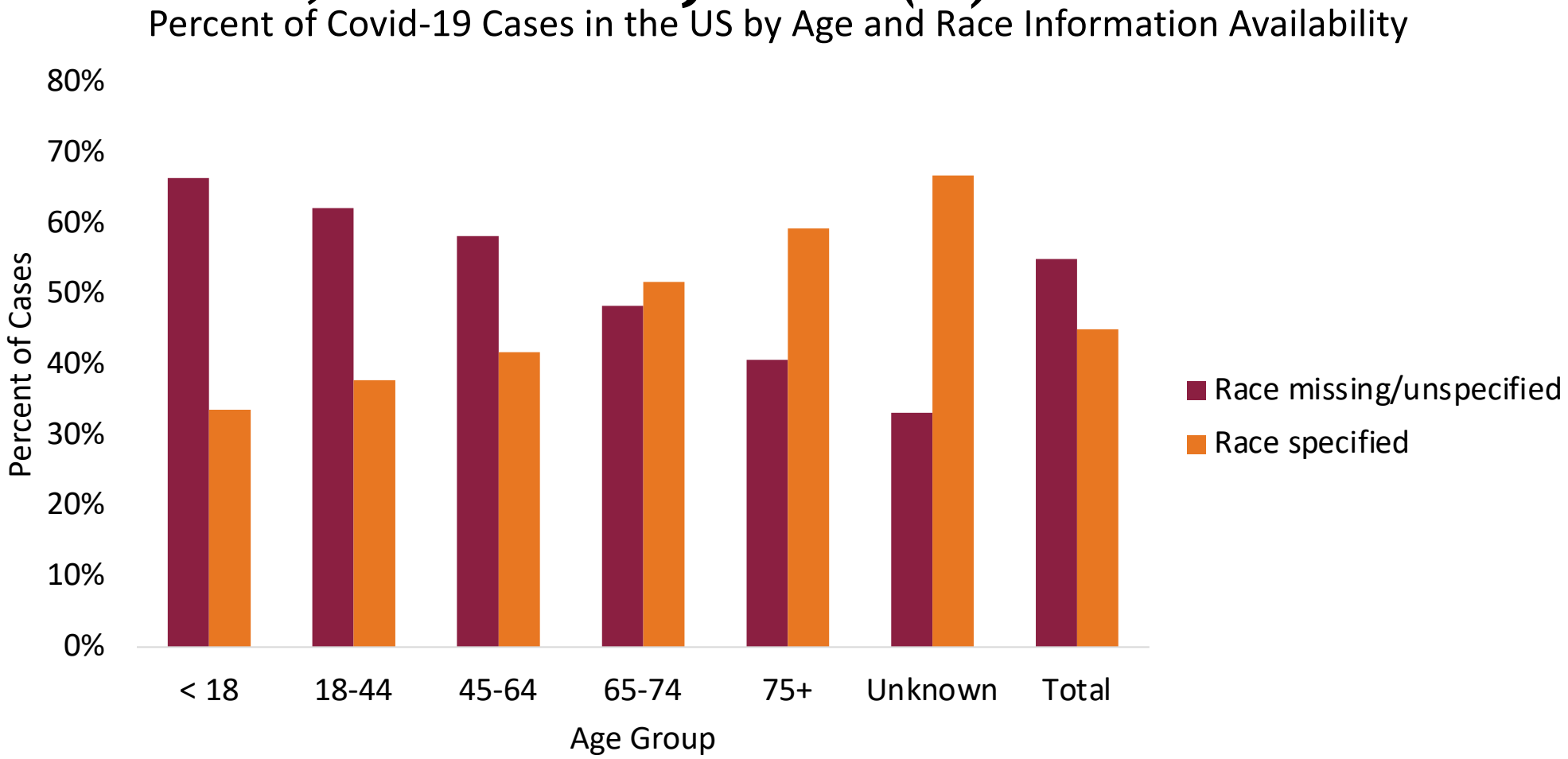
Same data, in chart form (3)



Stacked bar chart any better?

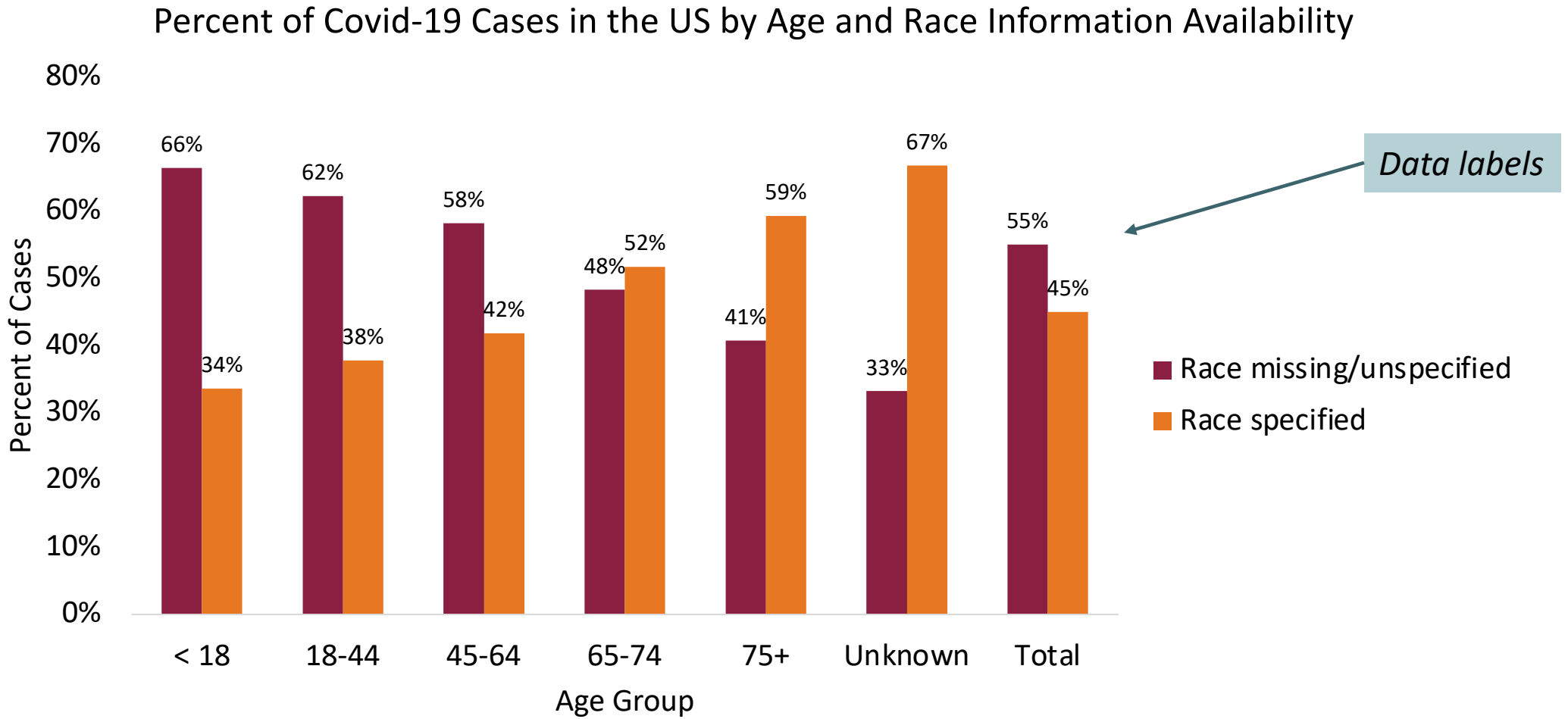
COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Same data, in chart form (4)



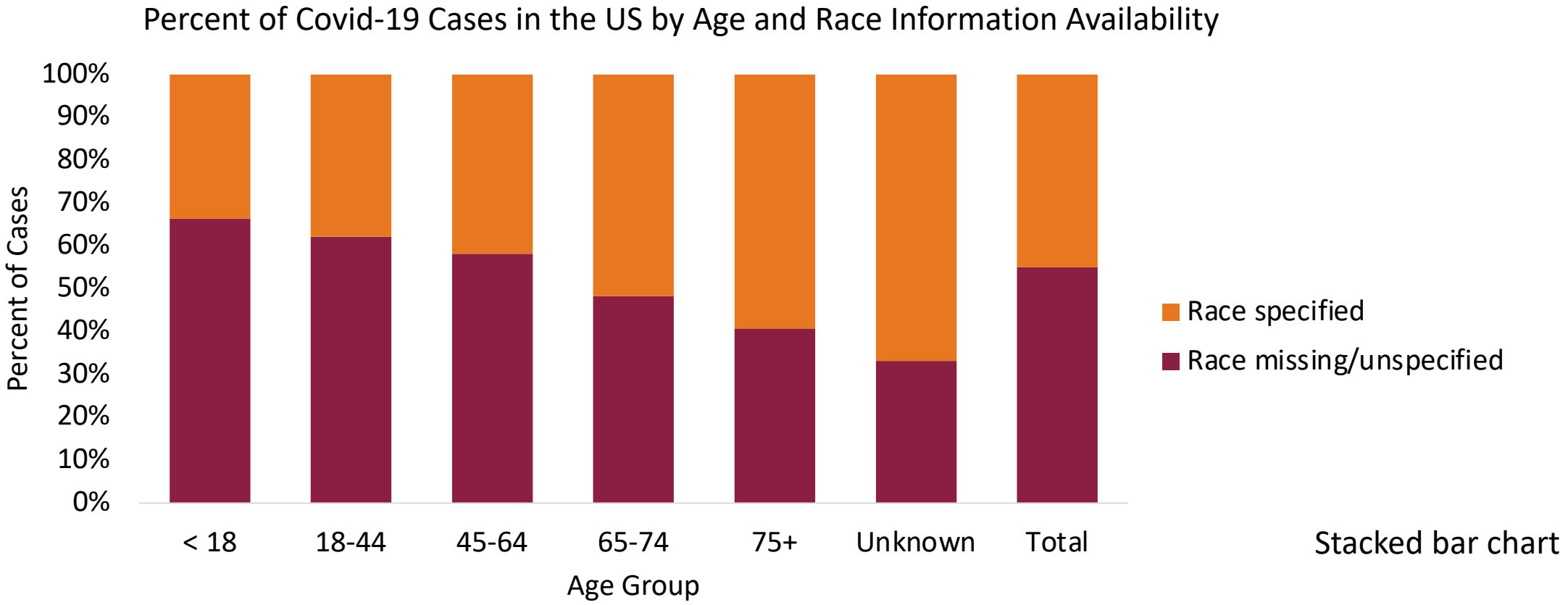
Using percentages, it is easier to see trends, and can include the total on the same scale for comparison (y-axis scale is 0-80%)

Same data, in chart form (4)



Using percentages, it is easier to see trends, and can include the total on the same scale for comparison (y-axis scale is 0-80%)

Same data, in chart form (5)



Stacked bar chart

Can clearly see the decrease in proportion missing race information with increasing age group

COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Figures/Graphs/Charts - what to make?

- Use the appropriate chart type for your data
  - *A pie chart is probably **never** the appropriate type for your data*
- Complex Charts: Allow audience to explore data
  - Can be interactive
- Simple, classic Charts: Communicate insights quickly and clearly

Depends on your audience and need

COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Chart Components

Required:

- Title
- Axis labels & scale values

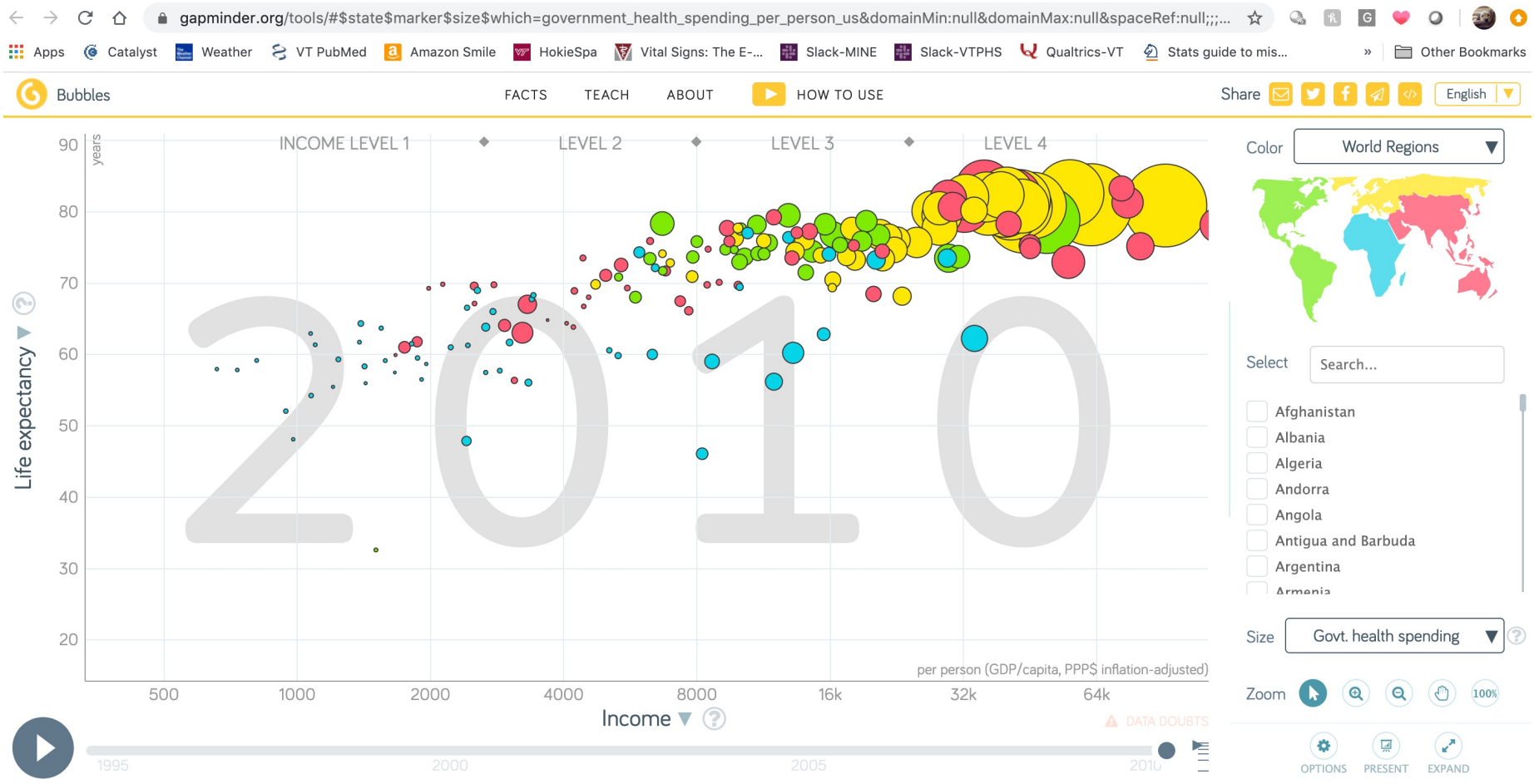
As needed

- Data source
- Legend
- Value labels

- Easy to read font
  - Size
  - Style
  - Color
- Clear visuals
  - Colors
  - Shapes
  - Labels

Charts should be easy to understand and self explanatory to your audience

Complex chart example



https://www.gapminder.org/tools/#$state$marker$size$which=government_health_spending_per_person_us&domainMin:null&domainMax:null&spaceRef:null;;;&chart-type=bubbles

Complex chart example

Virginia Department of Health COVID-19 Surveillance Data

Using the most recently available data from VDH. Raw data is available here.
Authors are not affiliated with Virginia Department of Health. See important notes below about this data.

Select date range to view

2020-05-08 — 2021-04-26

2020-05-08 2020-07-19 2020-09-29 2020-12-10 2021-02-20 2021-04-26

Plot age groups and health districts individually or sum totals for selected groups:

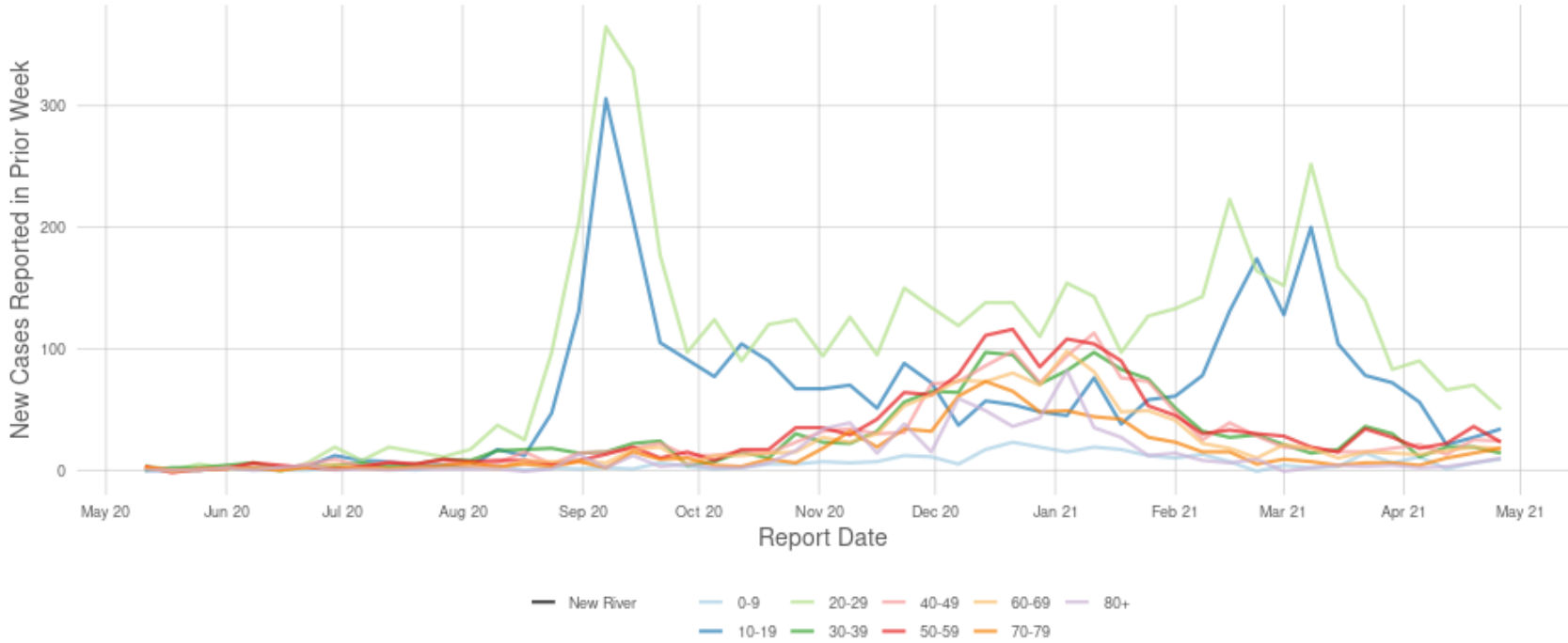
(•) Plot individually ( ) Sum total

Age Group

0-9 10-19 20-29 30-39 40-49 50-59 60-69 70-79 80+

Missing

New River



Download data | Download plot pdf

https://rsilverman.shinyapps.io/VDH-COVID-data/

COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Simple charts

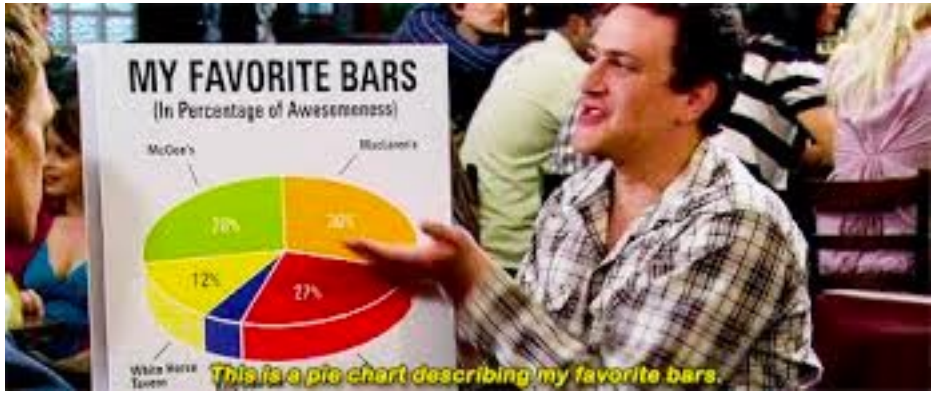
- Bar charts/histograms
  - Display the number, frequency, mean, or other data points, to compare categories or groups (nominal or ordinal).
- Line graphs
  - Displays how variables change over a continuous variable.
- Scatter plots
  - Displays the relationship between two of quantitative measurements for the same individuals.

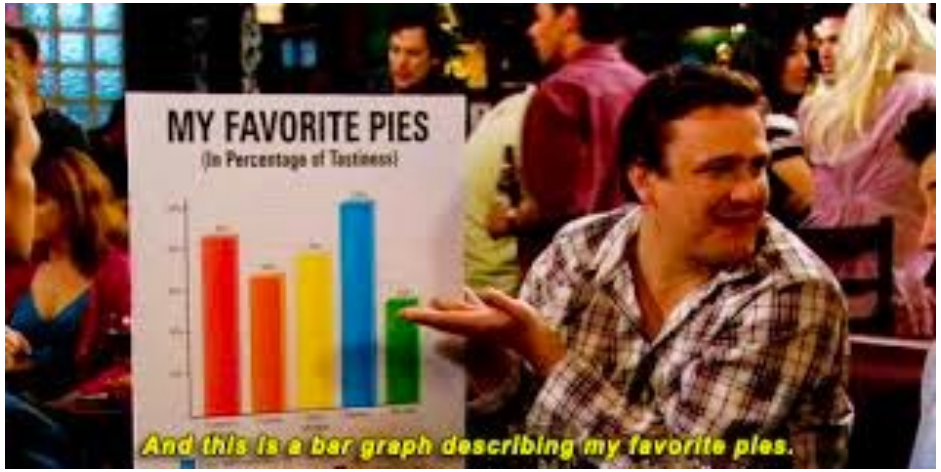
All can be grouped (use different color or symbol) to visualize comparisons within and across categories in the same chart

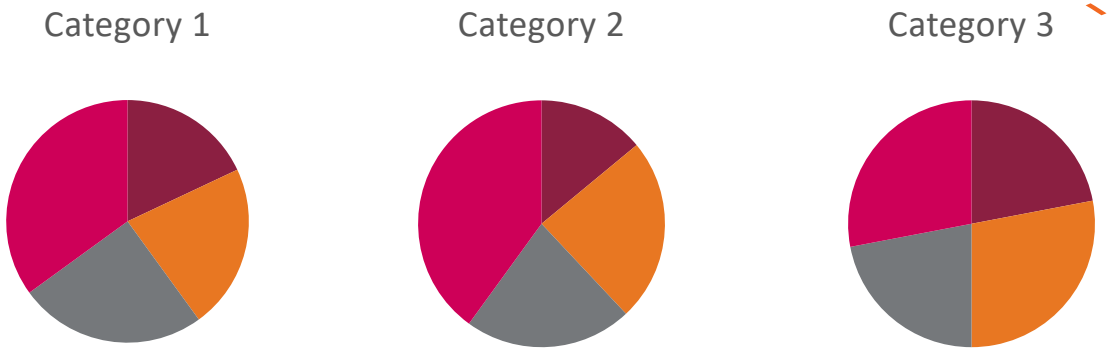
Bar charts

- Categorized data
  - Ordinal data: arrange in category order (e.g. age group, Likert scale)
  - Nominal data: often arrange from largest to smallest bar for easier interpretation
- Can be vertical or horizontal
  - Horizontal usually preferred when categories have long titles
- Grouped bar charts
  - Compare within and across categories
- Stacked bar chart
  - Better alternative to pie chart; can condense a figure and save space

Bar vs. Pie



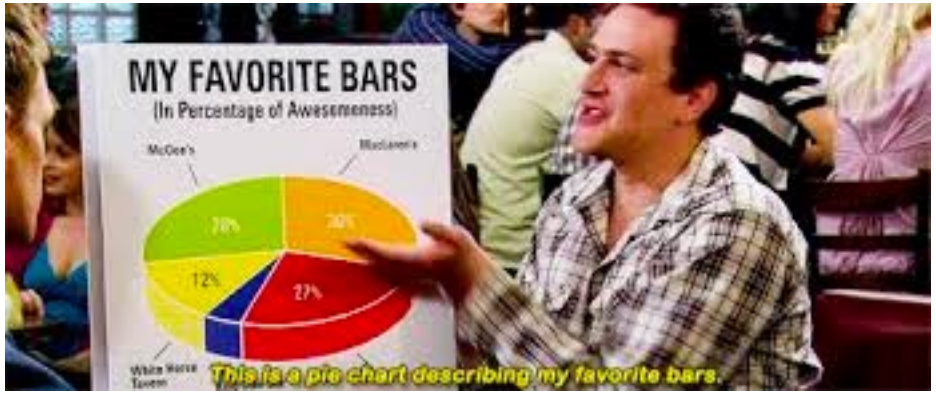


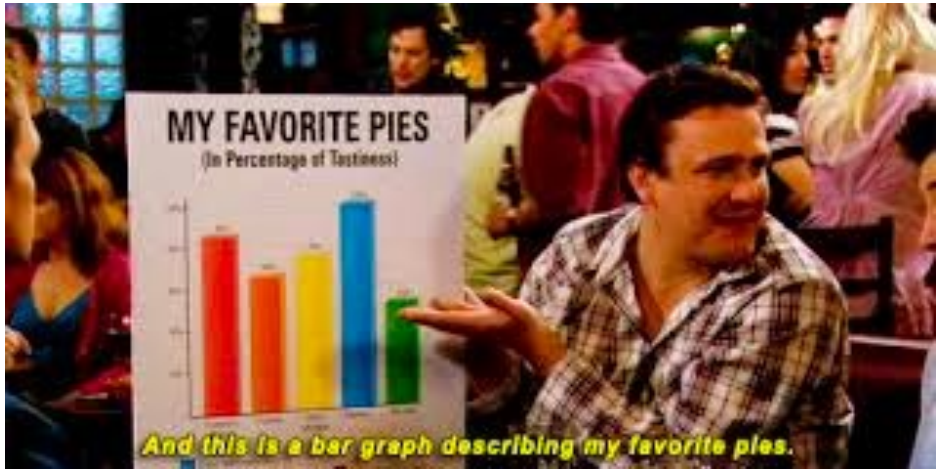


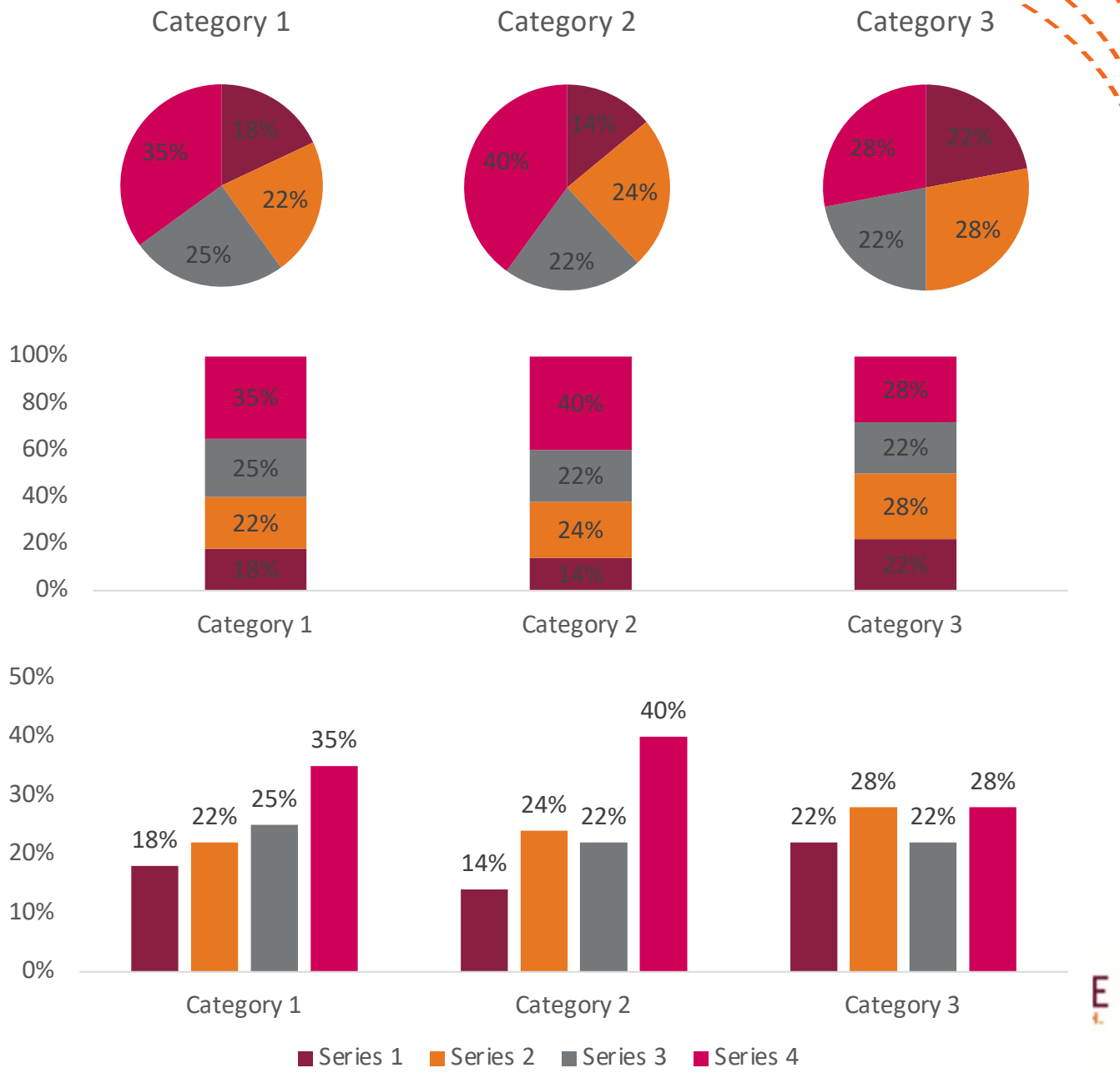
COLLEGE OF SCIENCE
VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Bar vs. Pie


MY FAVORITE BARS
(In Percentage of Awesomeness)
This is a pie chart describing my favorite bars.


MY FAVORITE PIES
(In Percentage of Tastiness)
And this is a bar graph describing my favorite pies.


Category 1
Category 2
Category 3
35%
18%
22%
25%
40%
14%
24%
22%
28%
22%
28%
22%
100%
80%
60%
40%
20%
0%
35%
25%
22%
18%
40%
22%
24%
14%
28%
22%
28%
22%
Category 1
Category 2
Category 3
50%
40%
30%
20%
10%
0%
18%
22%
25%
35%
14%
24%
22%
40%
22%
28%
22%
28%
Category 1
Category 2
Category 3
Series 1
Series 2
Series 3
Series 4

Among MSM, STD diagnosis increases risk of HIV

Slide title has main point

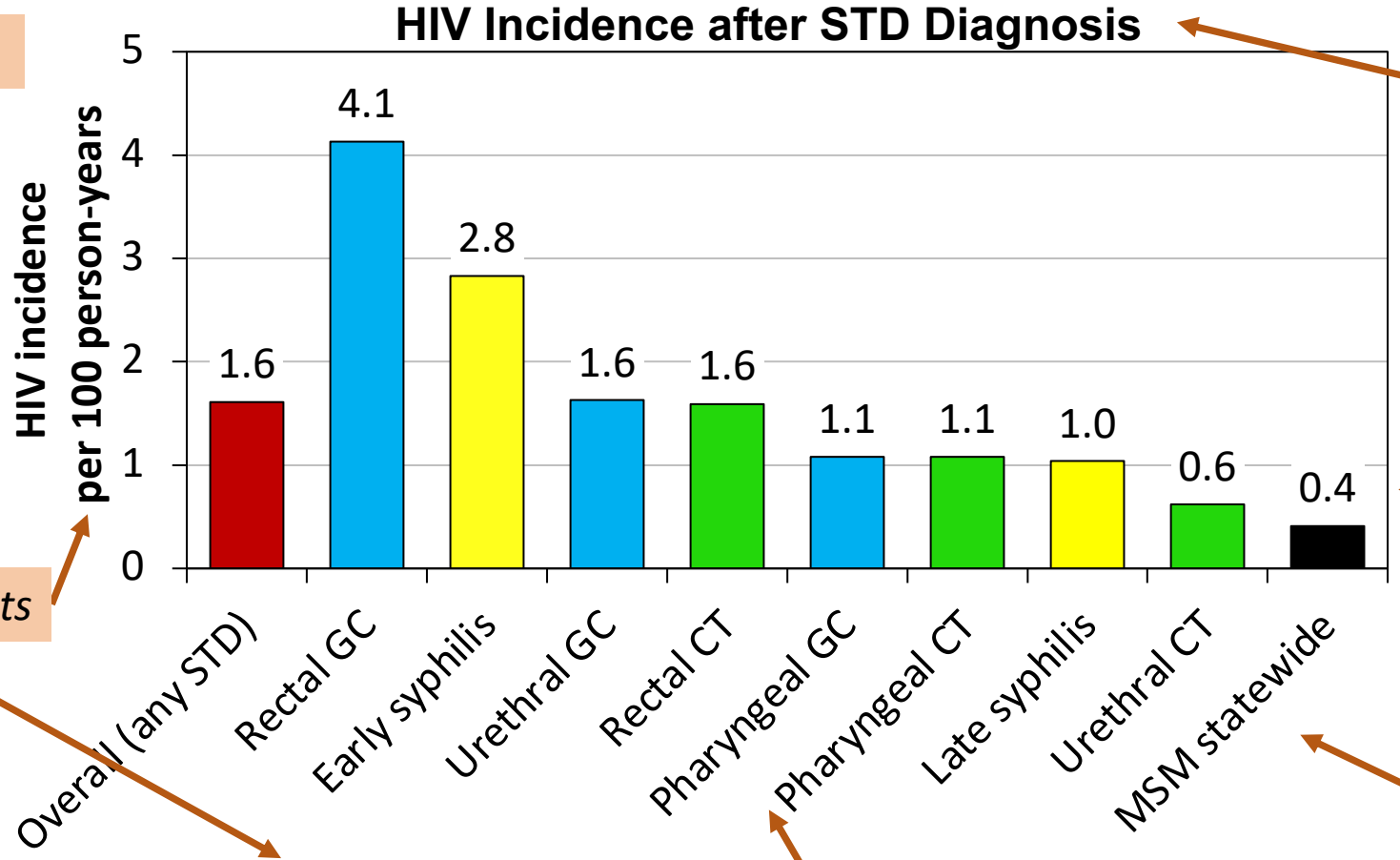


Chart title

Color coded by STD (could use a legend)

Data labels

Titles for axes

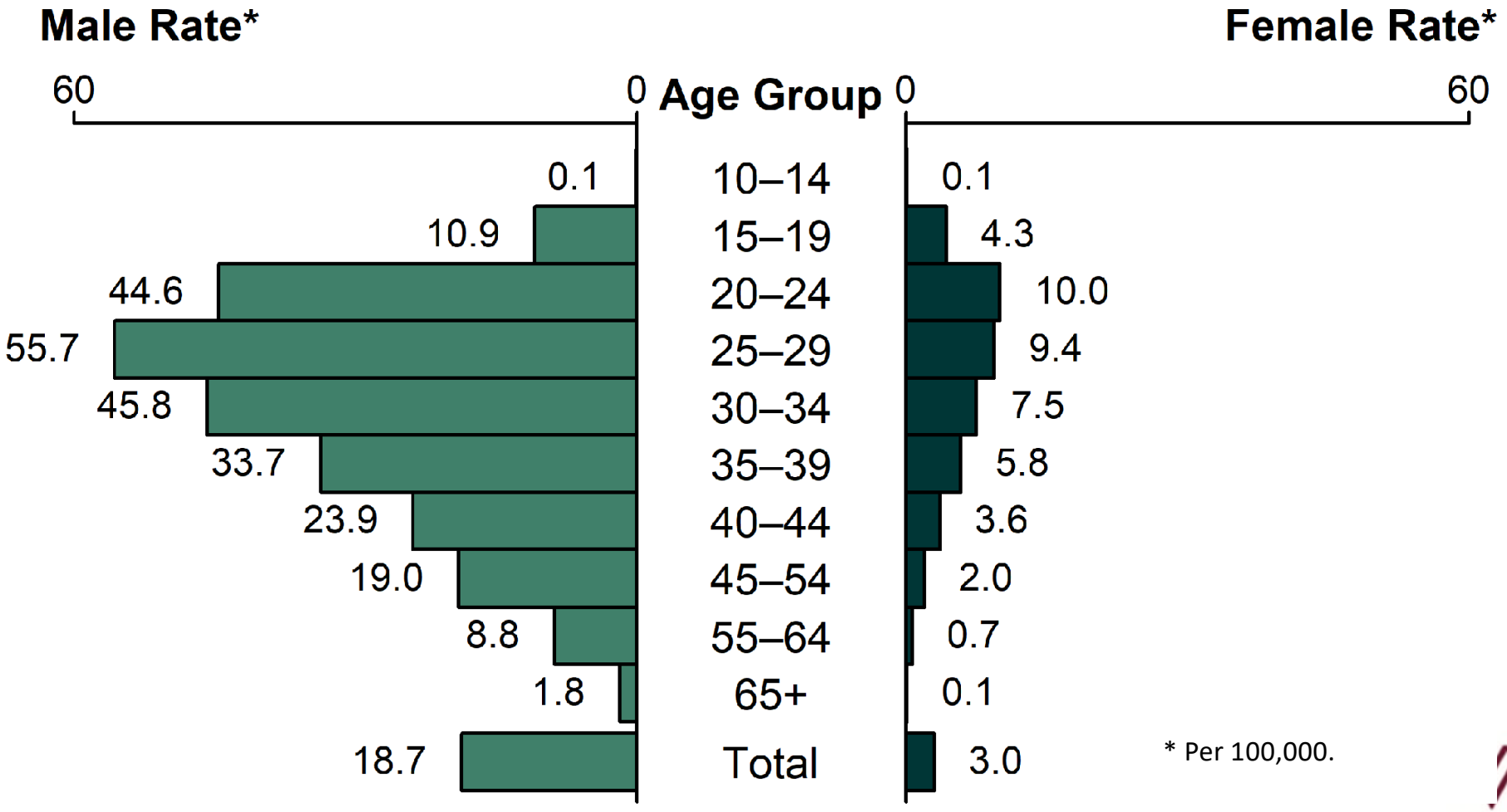
units

Nominal data ordered in descending value order (other than the overall)

Adapted from Katz, et al (2016)

May need to define abbreviations, but these were clear to this audience

Primary and Secondary Syphilis — Rates of Reported Cases by Age Group and Sex, United States, 2018

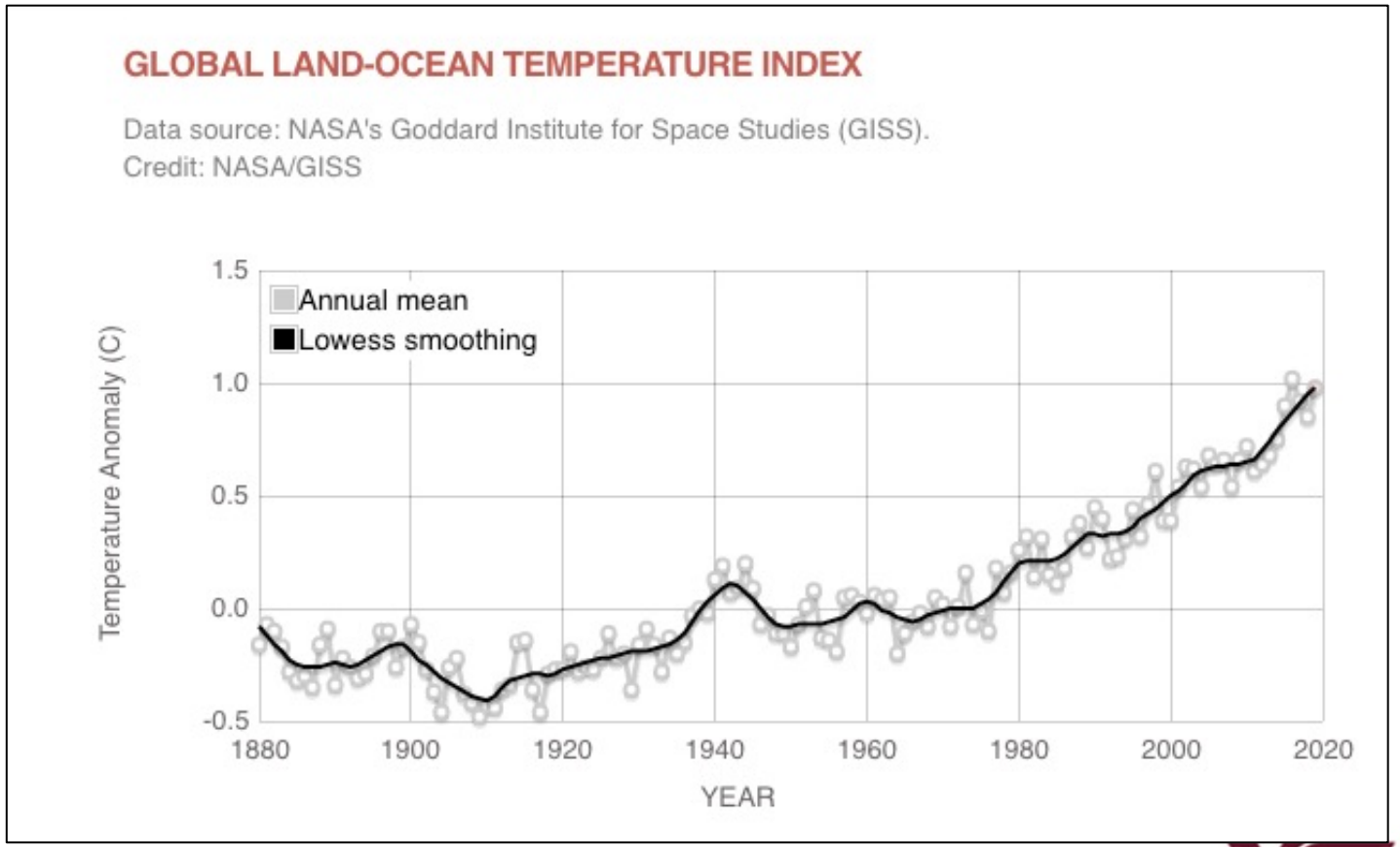


* Per 100,000.

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Line graphs

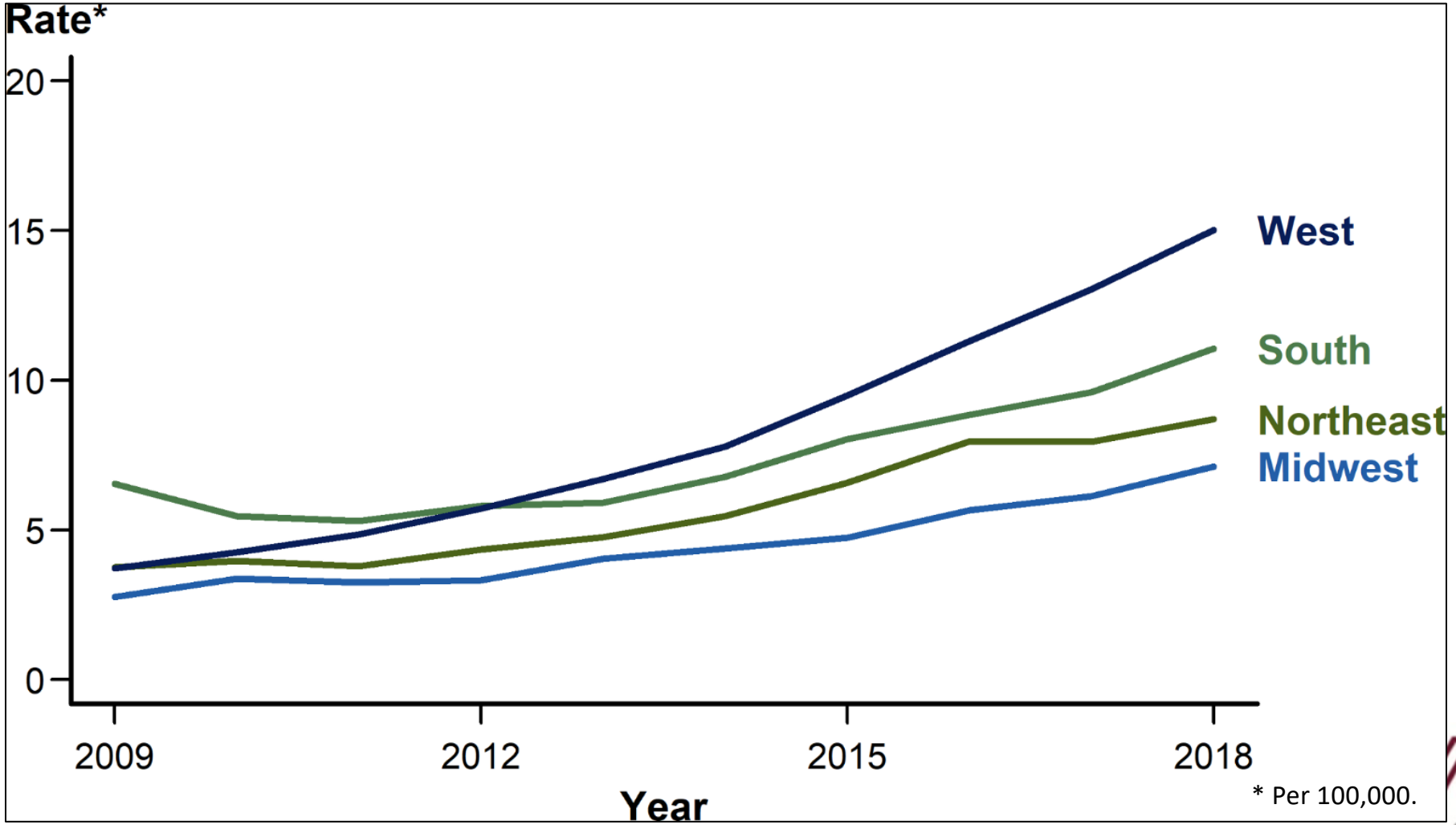
- Displays how variables change over a continuous variable
  - Time series
  - Distance
  - Others?



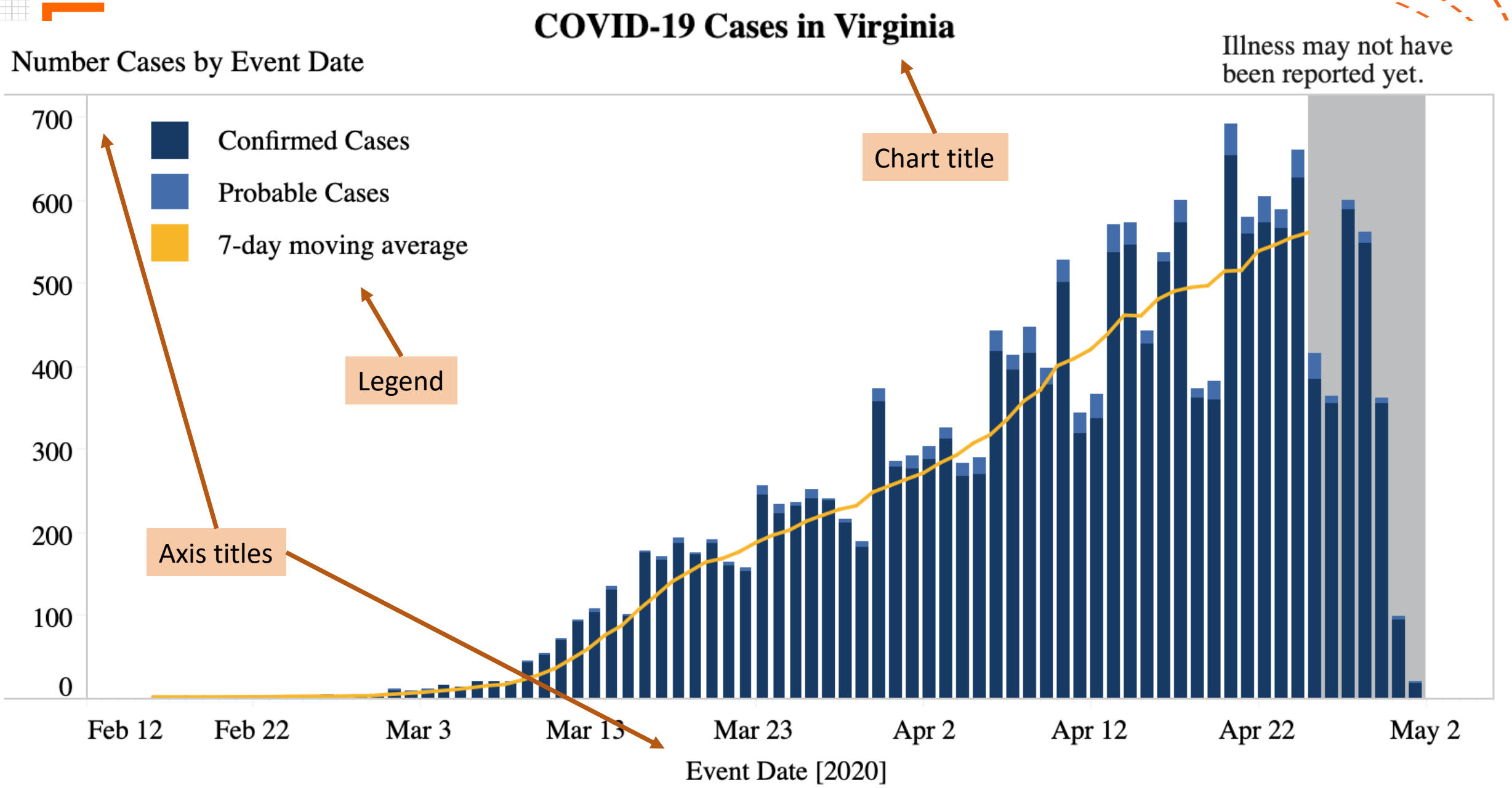
https://climate.nasa.gov/vital-signs/global-temperature/

Can easily interpret and compare trends

Primary and Secondary Syphilis — Rates of Reported Cases by Region, United States, 2009–2018



CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE



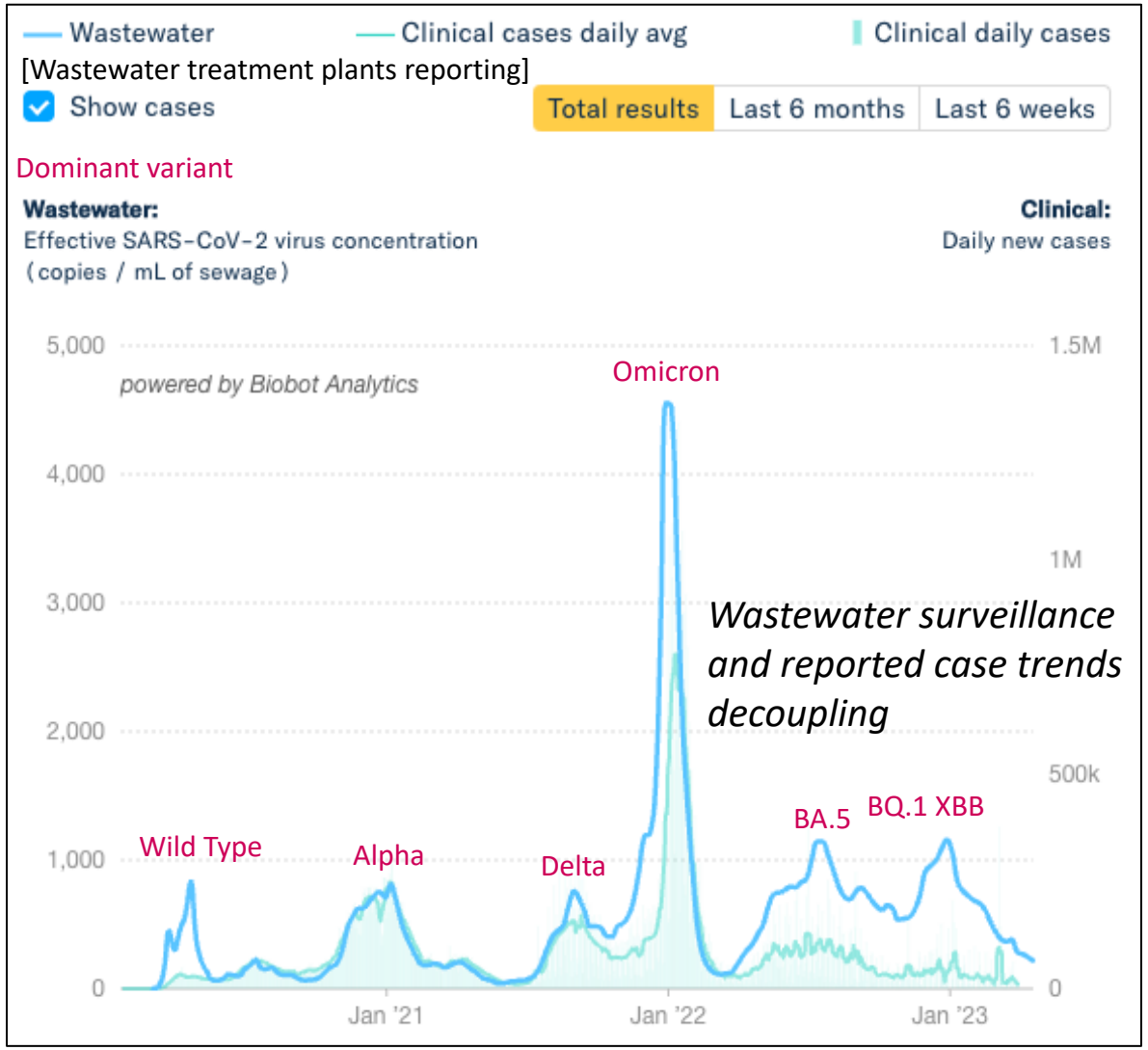
https://www.vdh.virginia.gov/coronavirus/

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

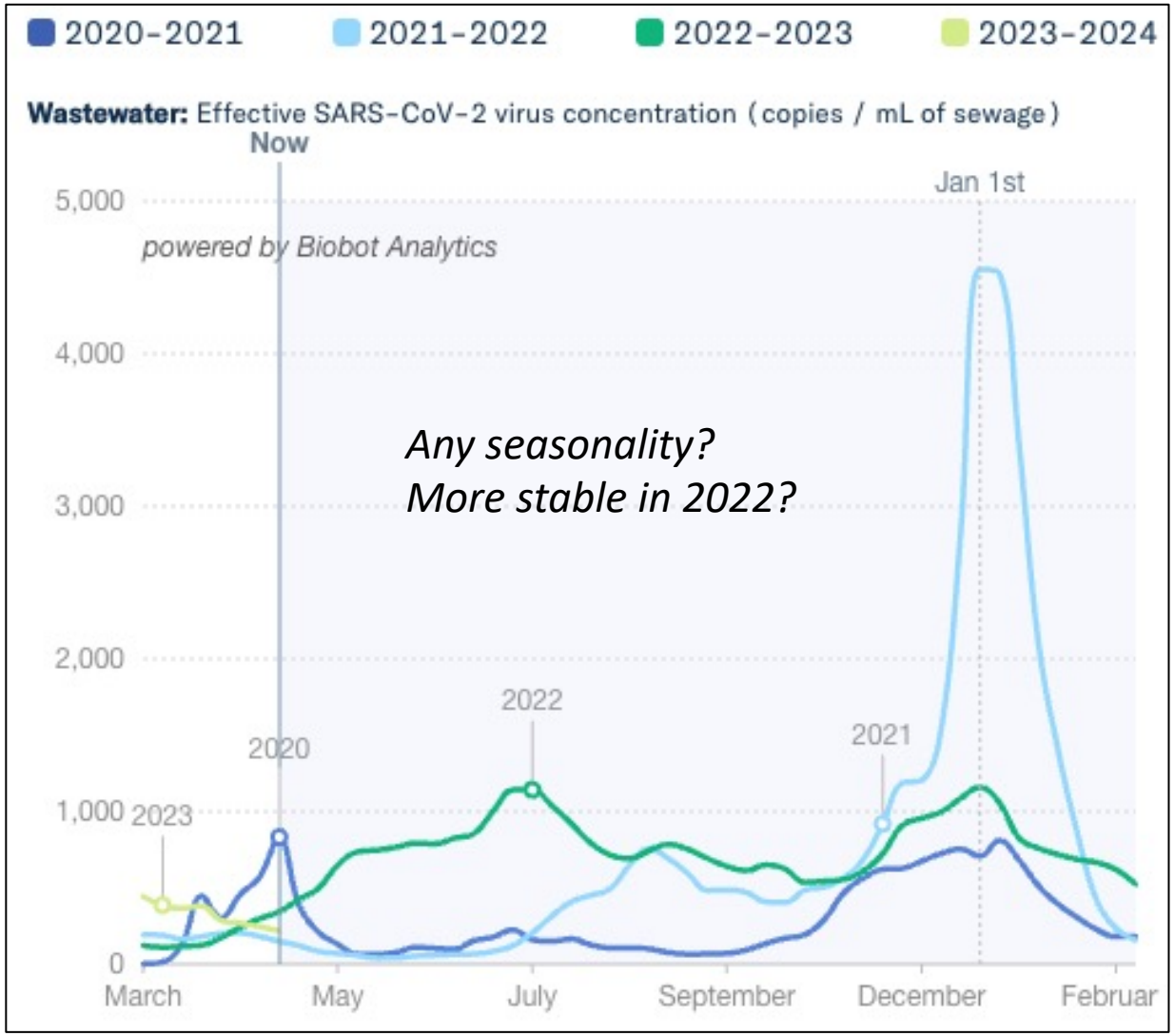
Wastewater Surveillance

USA Nationwide

Waste-water Concentrations vs. Reported Clinical Cases



Wastewater concentrations year over year

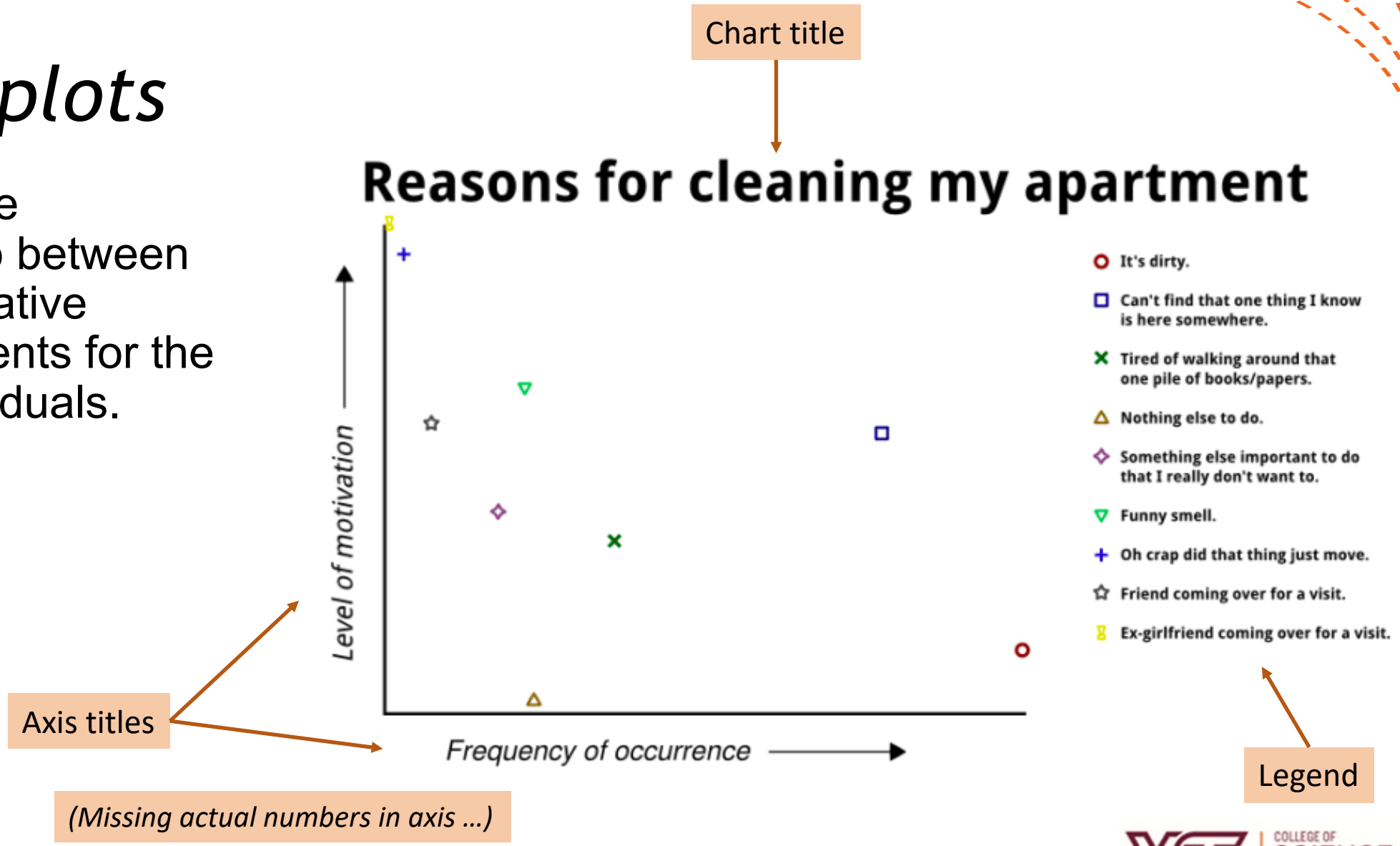


Source: https://biobot.io/data/

Data last updated April 24, 2023 from samples collected during the week of April 17, 2023.

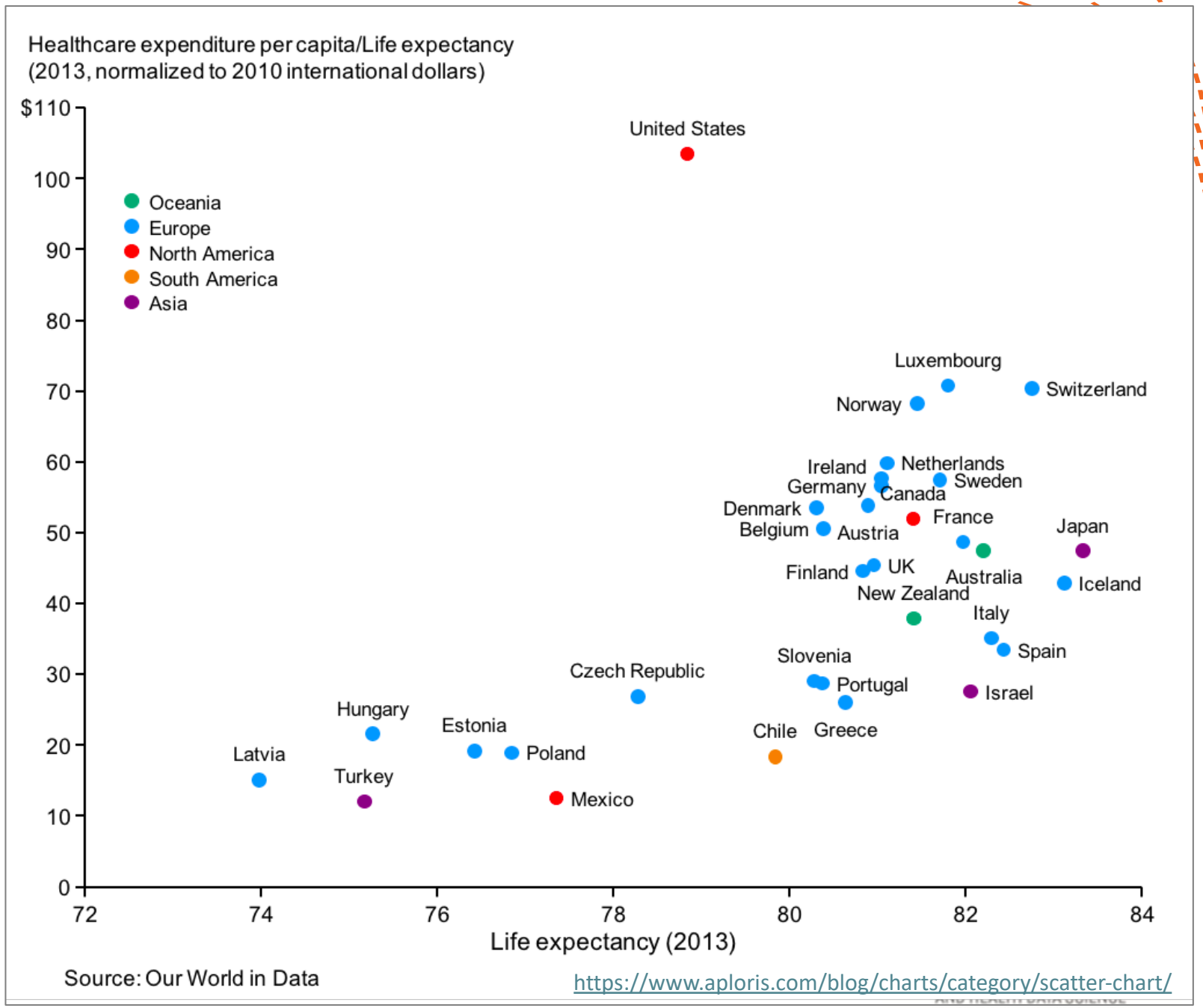
Scatter plots

- Displays the relationship between two quantitative measurements for the same individuals.



https://www.mentalfloss.com/article/64631/9-scatterplots-good-laugh

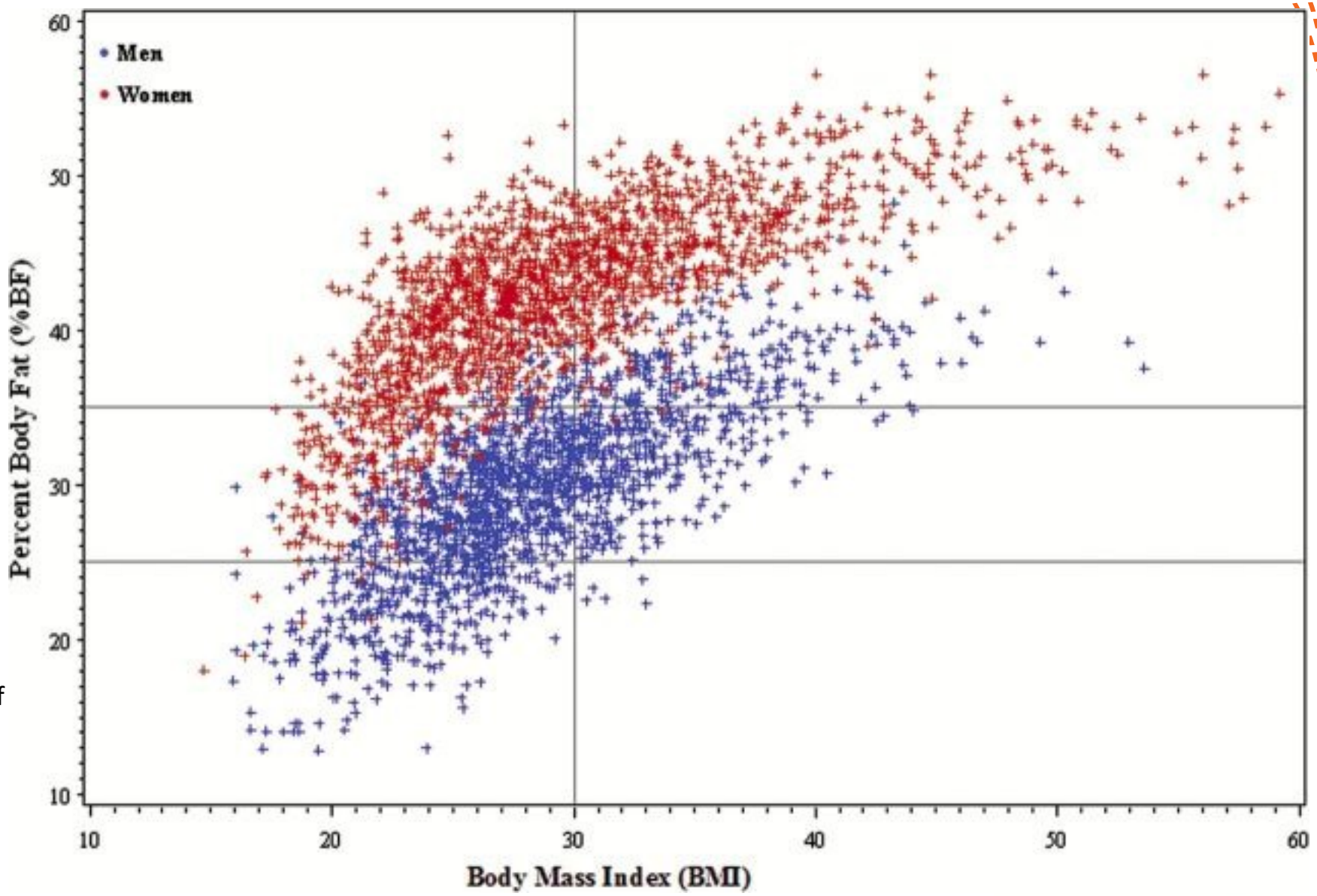
Scatter plots



https://www.aploris.com/blog/charts/category/scatter-chart/

Scatter plots

Peterson et al. (2015)

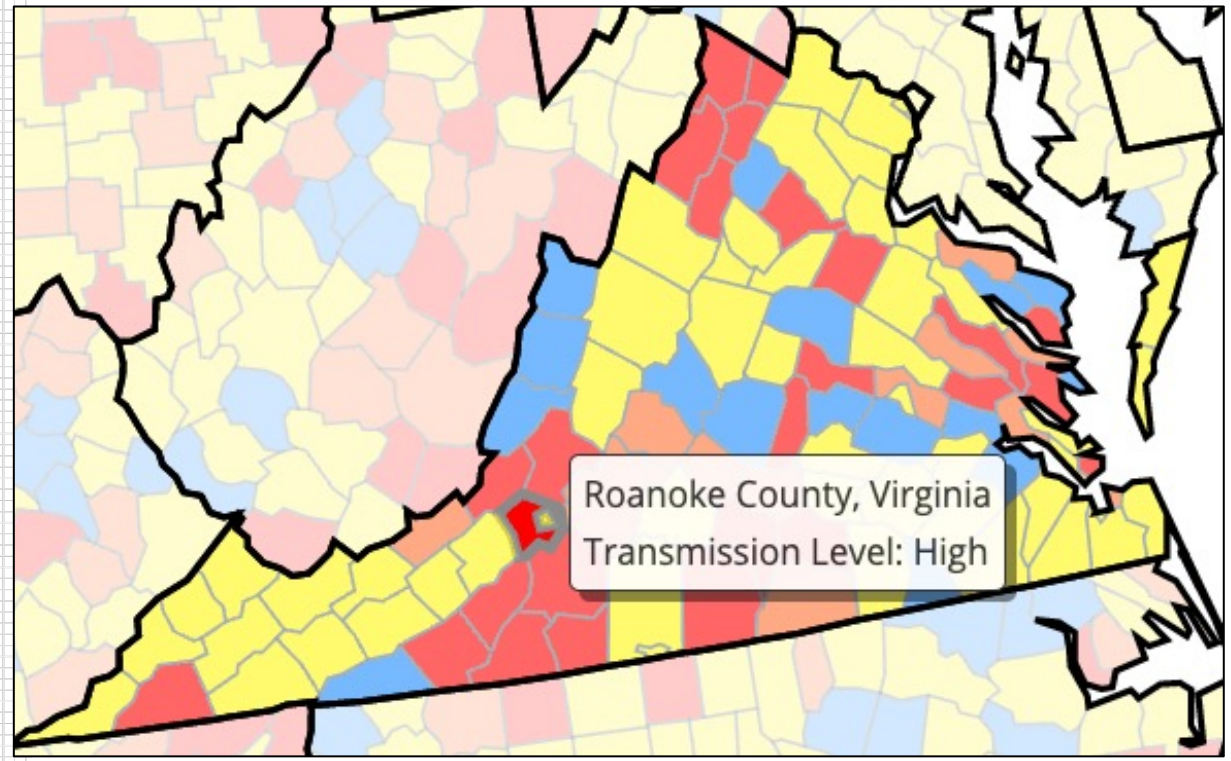


Scatter plot depicting the correlation between body mass index (BMI) and percent body fat (%BF) for the whole sample, by sex. The vertical line represents the standard BMI cutoff for obesity (BMI ≥ 30). The horizontal lines represent the sex-specific %BF cutoffs for obesity in men (≥ 25%) and women (≥ 35%).

Geographic/Geospatial data

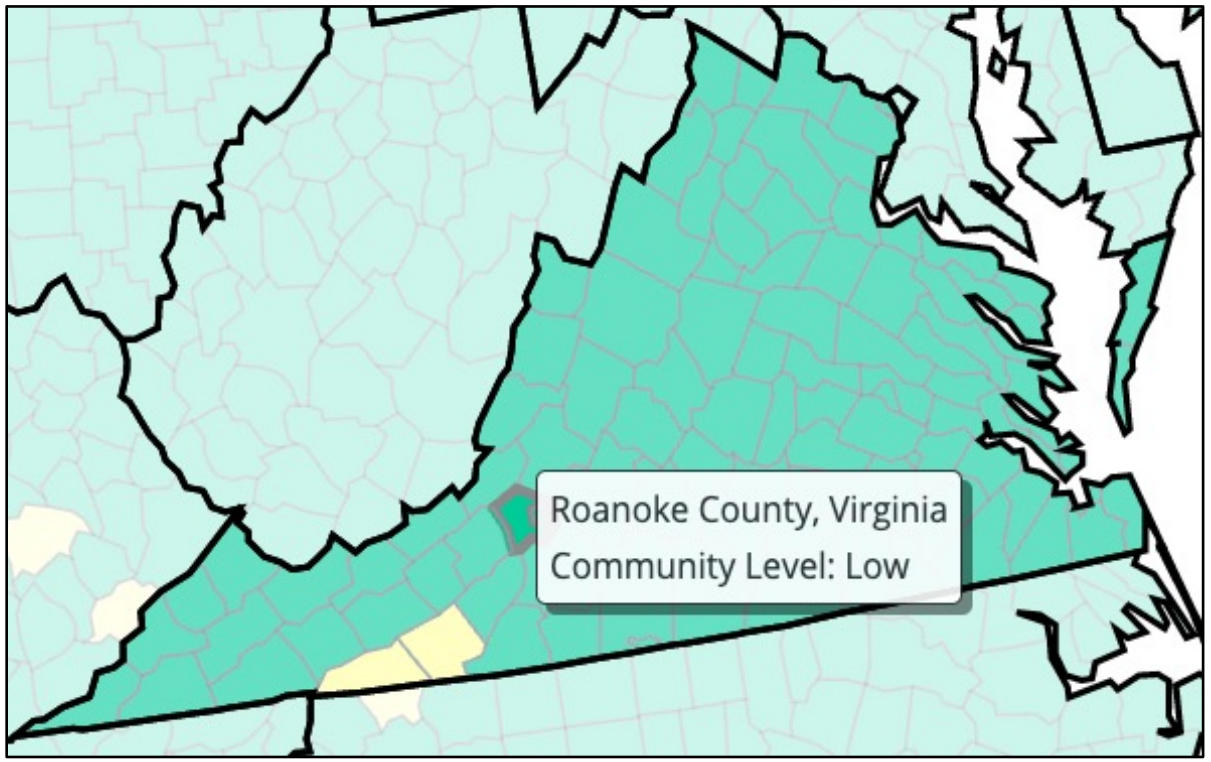
CDC COVID Data Tracker

Transmission Levels
Based on incidence and percent tests that are positive



Less of a Lag than hospital reports, but more lag than other sources (wastewater)
Underreported, depends on testing patterns

Community Levels
Based on incidence and hospitalization

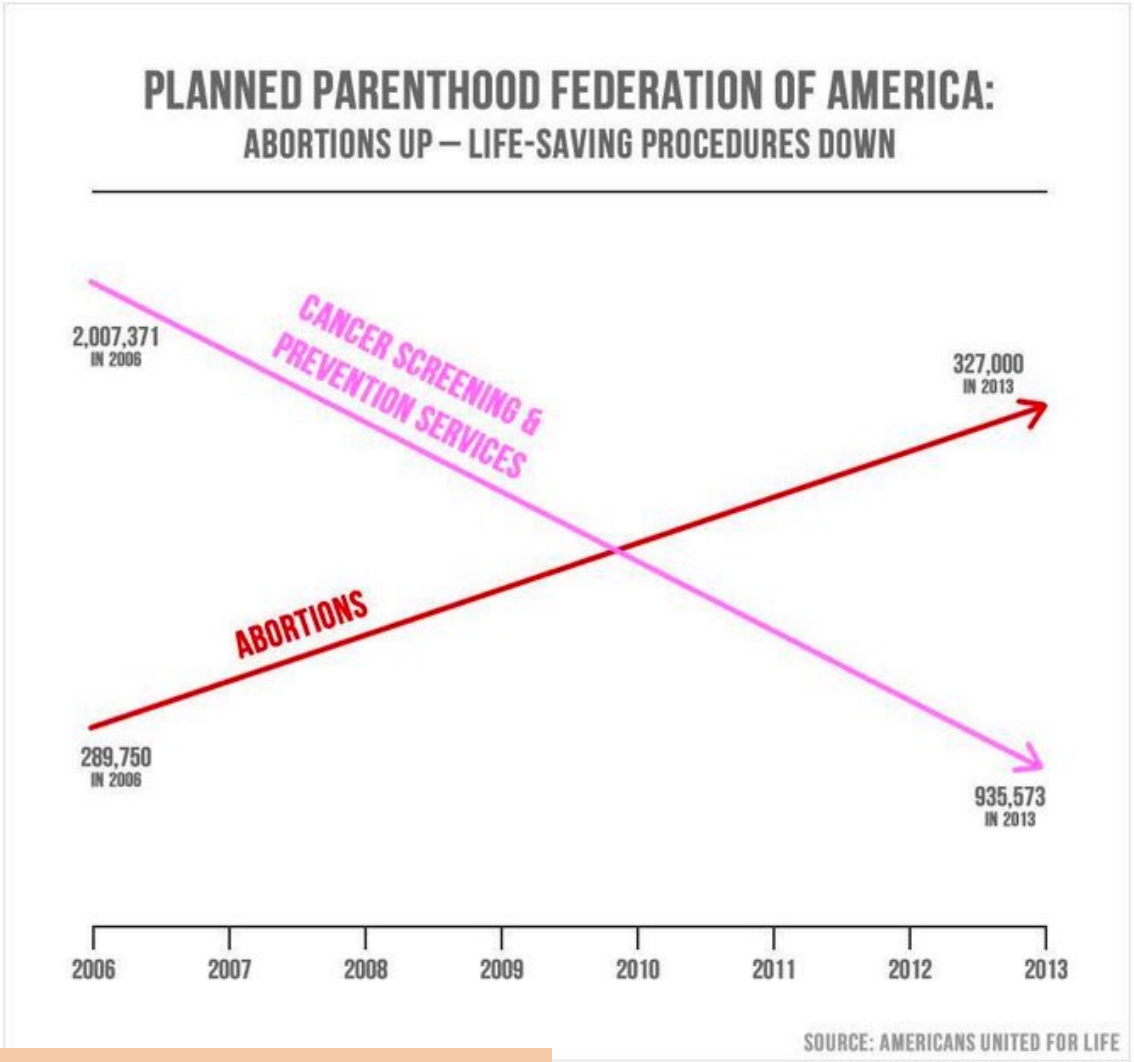


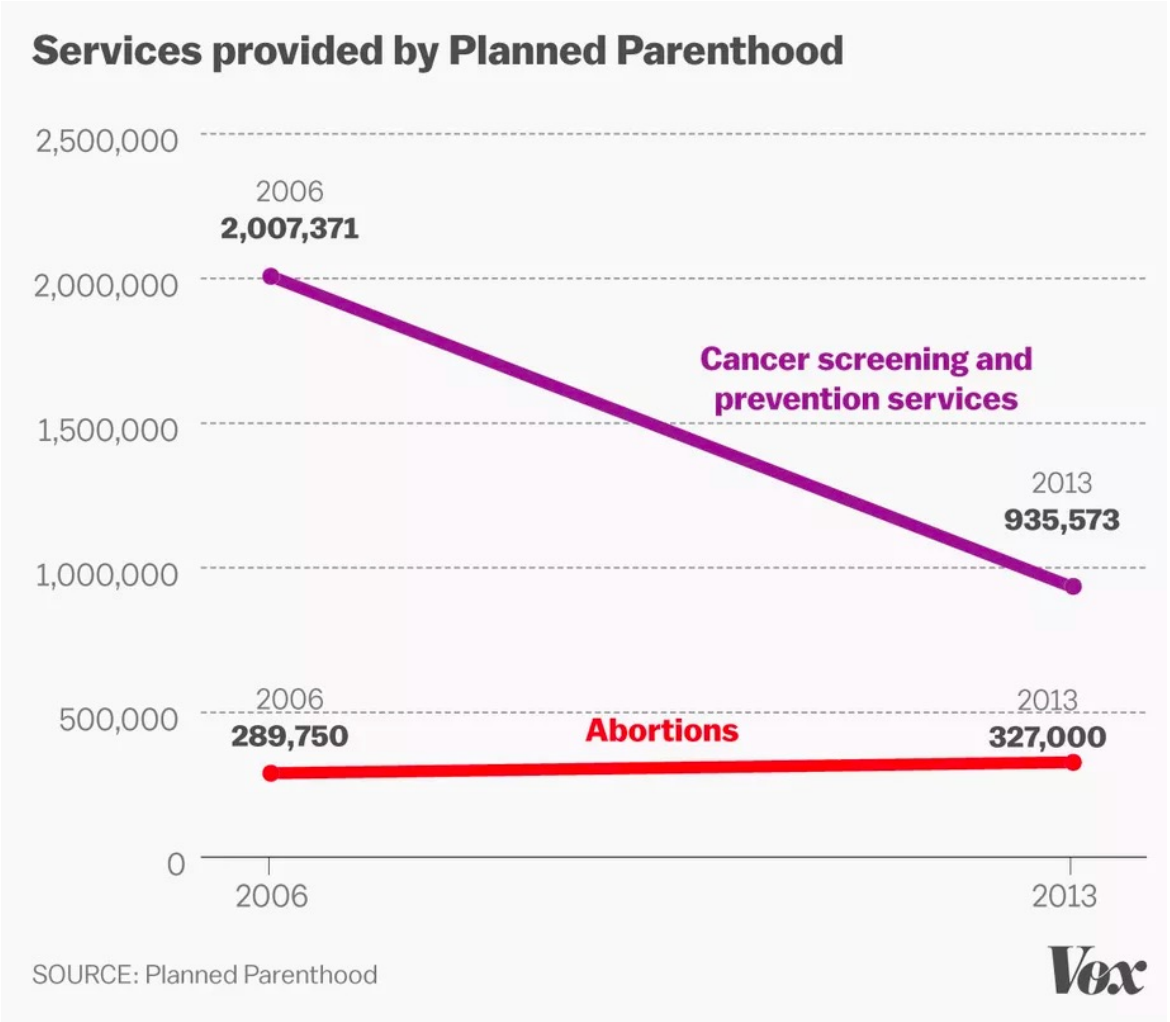
Hospitalizations is a lagging metric

As of 4/26/2022

Distorting data

Charts can communicate and miscommunicate data





No axis labels
Different scales for same units

https://www.vox.com/2015/9/29/9417845/planned-parenthood-terrible-chart

COLLEGE OF SCIENCE VIRGINIA TECH
CENTER FOR BIOSTATISTICS & DATA SCIENCE

Distorting data

Charts can communicate and miscommunicate data

And on April 23, when the state announced it was extending stay-at-home, DHHS charted hospitalizations with a chart with a Y-axis ranging from 0 to 500. Charting hospitalizations that way showed a steady increase, and the state said "there has **not** been a downward trajectory over the past 14 days."

The chart NCDHHS used last week to show that hospitalizations have been increasing.

April 23 – Extending Stay-at-home

But during a news conference on Thursday – and with a formal decision on reopening looming -- the state used a different graphic, with a Y-axis that goes from 0 to 1,000.

The chart NCDHHS used Thursday to show that hospitalizations have been "leveling" -- with a different Y-axis.

Cohen said this slide shows that hospitalizations are leveling – even though there are more COVID-19 patients in hospitals than there were when stay-at-home was extended on April 23.

April 30 – Ending Stay-at-home

https://www.wfae.org/post/analysis-columbia-epidemiologist-nc-s-data-i-can-only-share-your-befuddlement?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+wfae%2FZyar+%28Coronavirus%29#stream/0

Resources

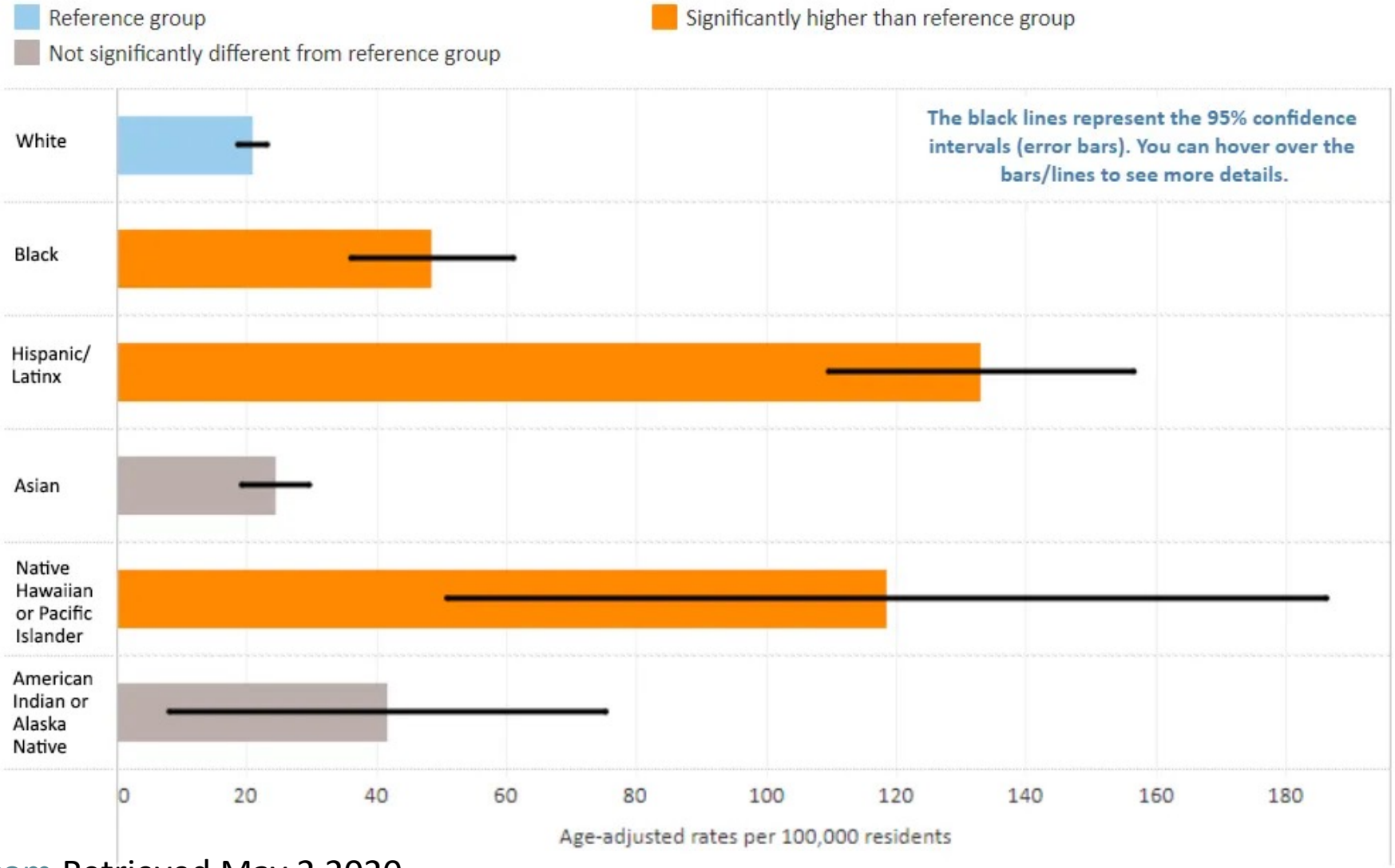
- https://serialmentor.com/dataviz/
- https://multimedia.journalism.berkeley.edu/tutorials/visualizing-data-a-guide-to-chart-types/

Summarizing data in text

- Providing a clear description and interpretation of data visualizations.
  - Include number(s) of people.
  - Include percentages, if applicable.
  - Interpretation should reflect the data.
  - Discuss limitations and how these may impact results.
- Level of detail may depend on the audience

Hospitalized cases per 100,000 residents (Age-adjusted)

Excluding skilled nursing and assisted living facilities



Seattle/King County

publichealthinsider.com Retrieved May 2 2020

COLLEGE OF SCIENCE VIRGINIA TECH

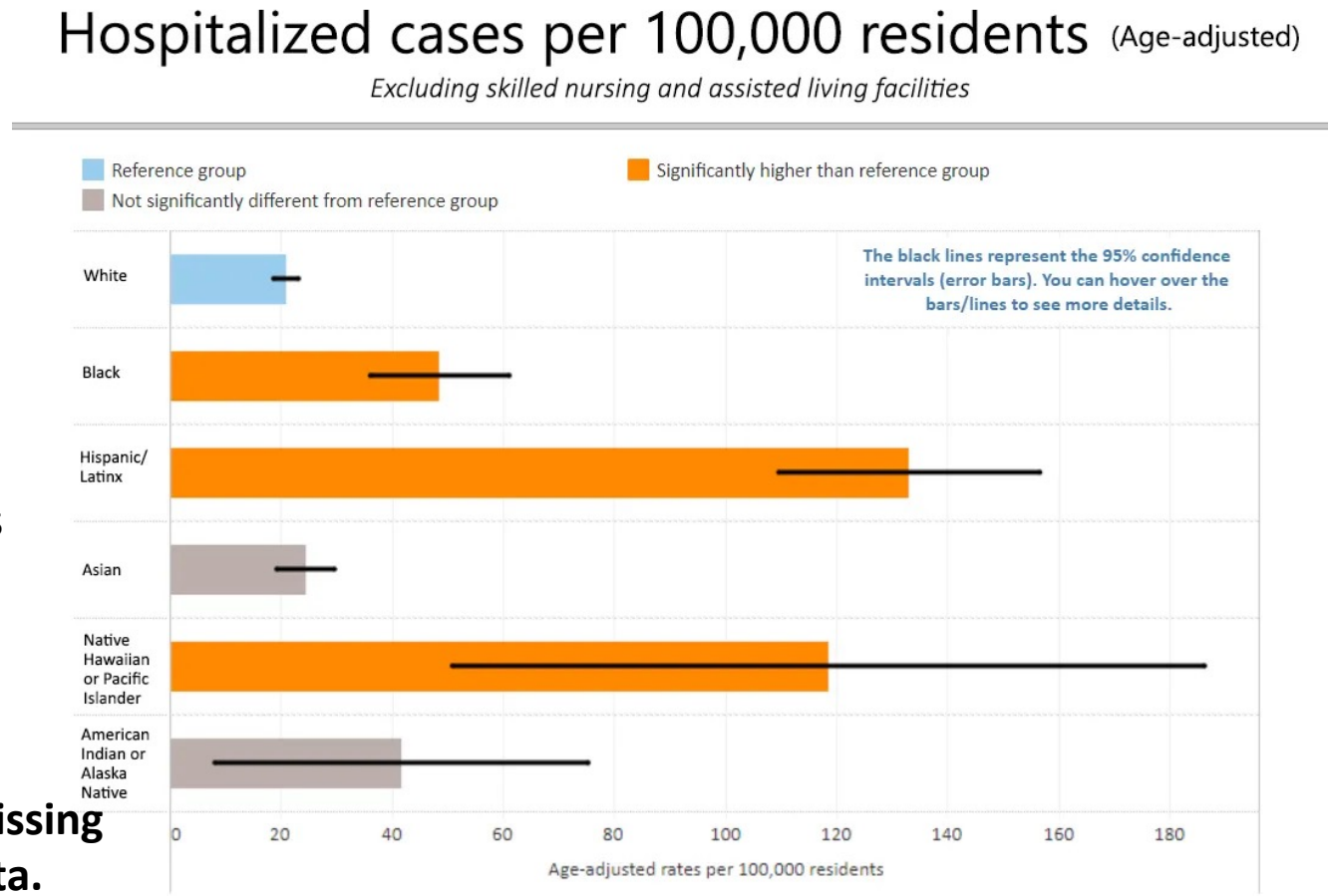
BIOSTATISTICS & DATA SCIENCE

Summarizing data in text

Executive Summary Seattle/King County

Overall, through May 1, 2020, there have been 6,348 confirmed cases of COVID-19 in King County, which translates into an overall age-adjusted rate of 285.0 cases per 100,000 residents. The rate of confirmed cases is highest among Native Hawaiian/Pacific Islanders (666.0 per 100,000), Hispanics (627.6 per 100,000), and Blacks (327.6 per 100,000). These rates are higher, with statistical significance, than Whites (148.6 per 100,000). Rates are higher for American Indians/Alaska Natives, but the rates are not statistically significant due to small case numbers.

Small numbers, limited availability of testing, and missing data should be considered when interpreting the data.



Seattle/King County

publichealthinsider.com Retrieved May 2 2020

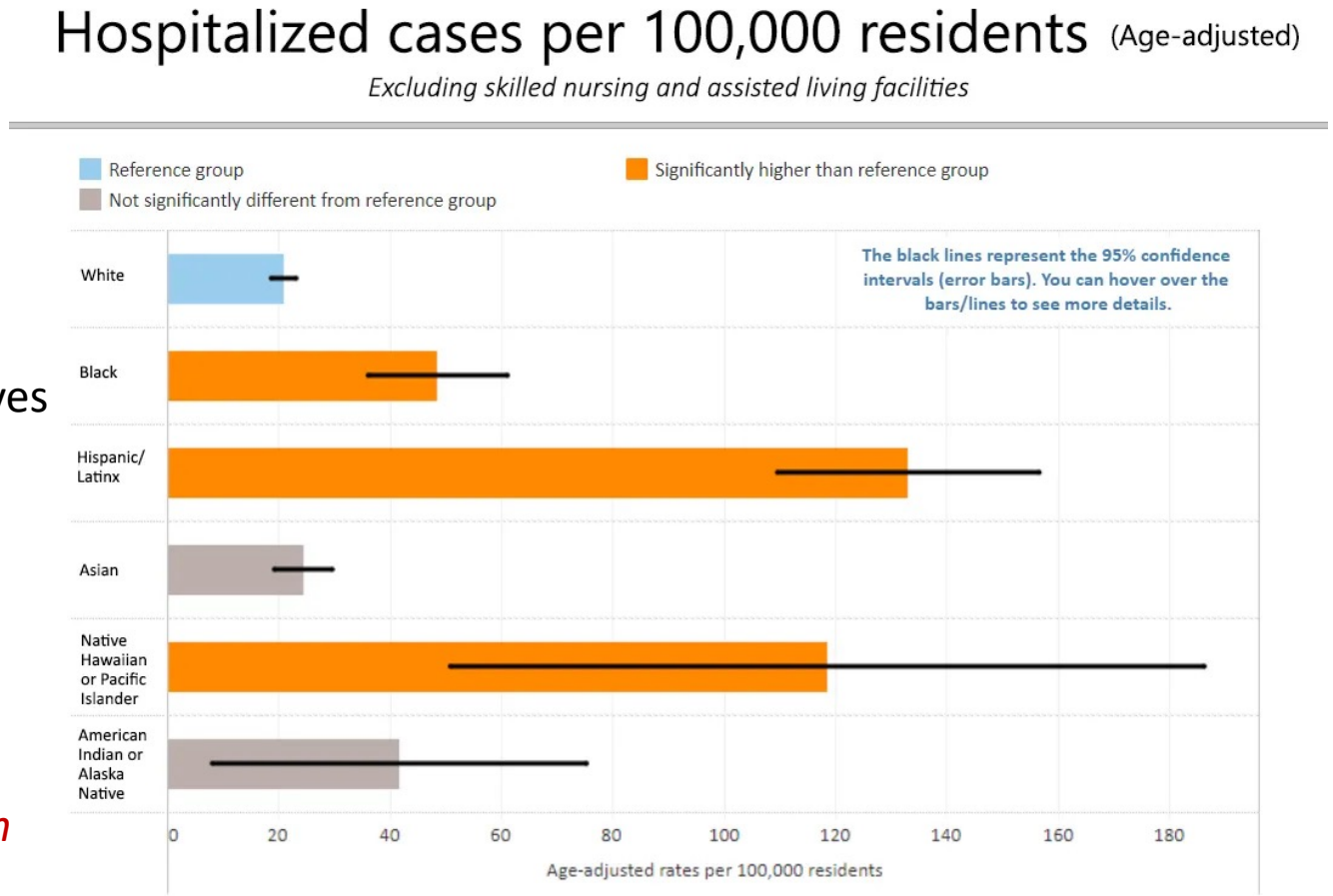
Summarizing data in text

Blog post/Article Seattle/King County

Hispanic/Latinx, Native Hawaiian/Pacific Islanders and Blacks had significantly higher rates of COVID-19 cases and hospitalizations as compared to Whites. The rates were also higher among American Indian/Alaskan Natives (though not statistically significant due to small population numbers) and slightly lower among Asian populations as compared to Whites.

Note: it does not say "there was no difference"

Statistical significance is just one piece of information



publichealthinsider.com Retrieved May 2 2020

Final tips

- Look at how other's have presented and discussed similar data.
  - What looks good? What doesn't?
  - What components or data points are typically included and how are these presented?
- Get feedback
  - Ask colleagues/friends/etc. for their thoughts.
  - Is it clear to them? Hard to look at?
- Keep your audience in mind
  - Understanding their background and level of understanding informs what you will present and how to effectively communicate data

Presentation is an art.

- *What works best for your situation?*
- *What you want to communicate?*
- *Who will receive this information?*

Questions?

Contact me at: rsilverman@vt.edu
CBHDS website https://biostat.centers.vt.edu/
(virtual walk-in hours & requests for collaboration)

COLLEGE OF SCIENCE
VIRGINIA TECH
CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE