An Investigation Into Vaccination Patterns Across Demographics (Report)

AUTHOR1 and AUTHOR2

Research Questions and Results Summary

1. How have vaccination rates for major vaccines for the general US population changed over time (from 1998 to 2016)?
    a. Most have been stagnant but some have been introduced and then rose for a bit but quickly plateaued. Hepatitis A is the notable exception, acting differently than all others.
2. Is there a correlation between vaccination rates and poverty status across the general US population?
    a. Families at or above the poverty level have a higher percentage of their children getting vaccinated than those who are below poverty.
3. Is there a correlation between vaccination rates and location relative to metropolitan area across the general US population?
    a. Families outside the metropolitan statiscal area (MSA) have a lower percentage of their children getting vaccinated than those inside the MSA
4. What is the difference in coverage by a person's race and hispanic origin across the general US population?
    a. There were varying percentages of vaccination of children across all races. A notable discovery was that in some cases Asians have the highest percentage of vaccination in children.
5. What will the vaccination coverage look like in the future in the general US population?
    a. Analysis of the trends over 1998 to 2016 indicates most will continue at a near-constant rate, with the exception of Hepatitis A which might continue to rise.

Motivation and Background

The research that will be done on is to determine how vaccination patterns have changed over time across various demographics. Our reasoning for undergoing this research is to understand which groups of people have a low vaccination coverage in the past and currently. This will allow other researchers to know which groups of people still need vaccinations. This research will also let researchers know which people are still vulnerable to certain diseases that could be addressed. This research could be used as visual data for why certain groups of people do not have high vaccination coverage. With this information the distribution of vaccinations could be improved which would help to ensure everyone is properly protected.

Dataset: https://www.cdc.gov/nchs/data/hus/2017/066.pdf

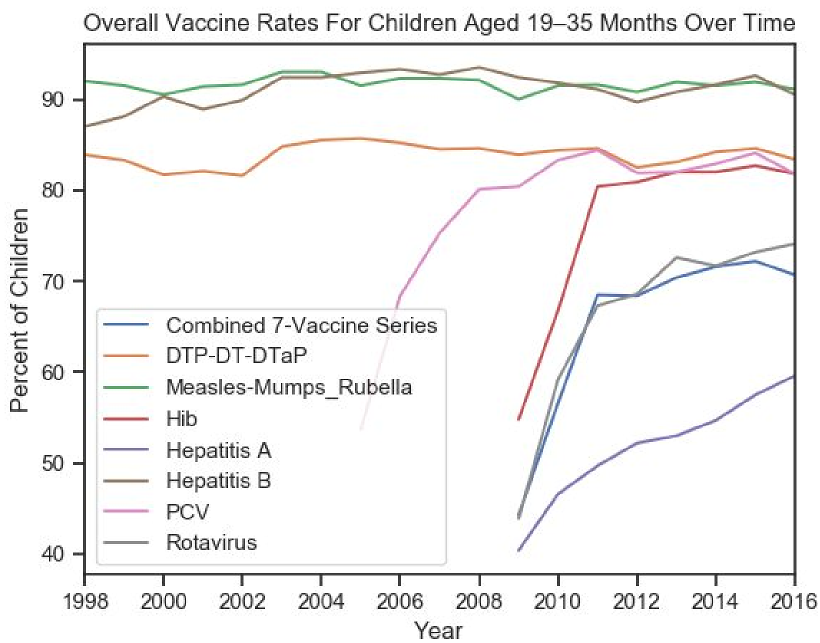
Methodology

For the first and last research question the methodology will be to chart the vaccination rates for major vaccines against time as the constant variable on a scatter plot. From there, we will plot a line of best fit and take note of how well it fits onto the data, as well as the trajectory of the given line. This will be done through polynomial regressions of varying order depending on the data. Through this we will be able to see how the rates have changed thus far as well as extrapolate as to how they will change going forward.

For the remaining three research questions what are effectively bar graphs will be used to take note of the changes between the dependent variable (vaccination rates for major vaccines) as compared to the independent variables (poverty status, location, or race/ethnicity depending on the question being asked). This way, we can analyze and not any changes. The independent variables aren't continuous (i.e. they can't be represented by, say, a number with a wide range of options). Thus something like a scatter plot wouldn't function in the given circumstances.

Results

Research Question 1

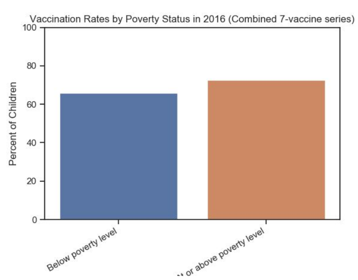


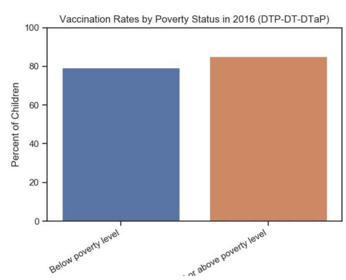
With this question, we plotted all of the vaccinations on the same access. The first thing we noticed was that several were decidedly horizontal in their graphing. Namely, these were Measles-Mumps-Rubella, Hep B, and DTP/DT/DTaP. They all originated in the high 80's / 90's in 1998, when our data begins. The immediate next thing we took note of was that the rest of the vaccines, excluding Hep A, followed a pattern that looks roughly like a quick rise and then a plateau. These ones each didn't start back in 1998 but rather in either 2009 or 2005 (PCV). Finally, the aforementioned Hep A, while also starting in 2009, has a more linearly upward trajectory. The exact trends of these vaccination patterns will be discussed in more depth with research question 4.

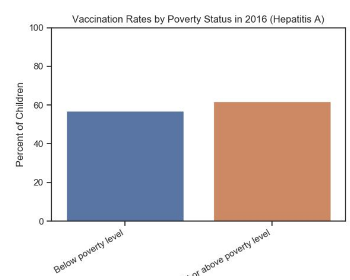
So why does our data look like this? Well the answer can be found with some outside research. Via the Children's Hospital of Philadelphia, we can have distinct information about when each of the non-1998 vaccines came into popularity. For one, PCV (pneumococcal vaccine) was established in 2001, which doesn't entirely explain why it doesn't appear in the data until 2005. Rotavirus, on the other hand, came back into the immunization schedule in 2008, explaining for the most part it's placement in the data (starting at 2009). Others, like Hep A and Hib (both 2000) and the combined 7-series data, don't have good external reasons for being the way they are.

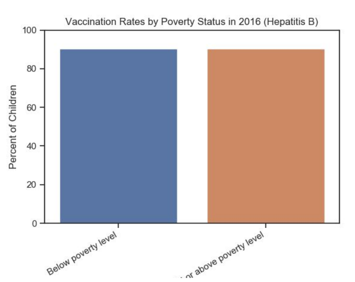
What can this data tell us about the period of time we have access to? I think the biggest conclusion that can be drawn is that once vaccines reach a certain level of usage they tend to stay constant from then on. This is evidenced by the near-level quality of the long-standing vaccines. Additionally, vaccines that were newly introduced took some time to reach a constant level of usage. This would seem to indicate that it takes some time for the public to gain trust and faith in new vaccinations.

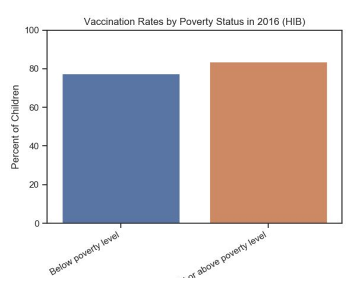
Research Question 2

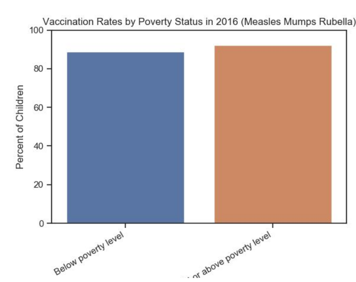


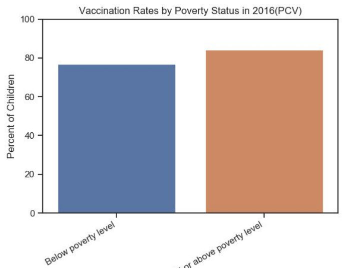


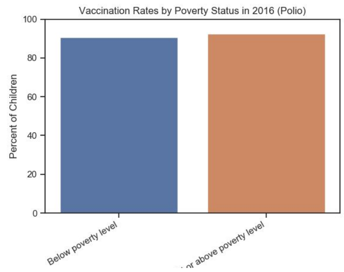


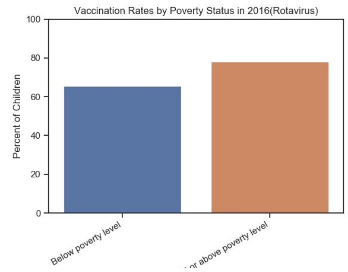


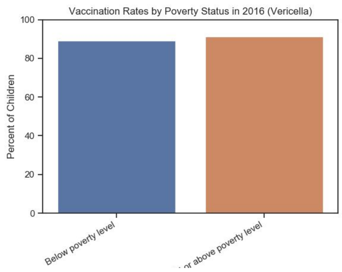






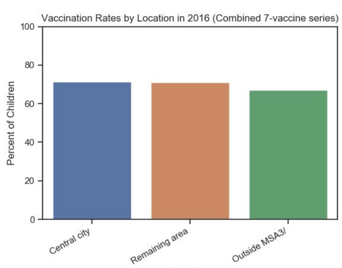


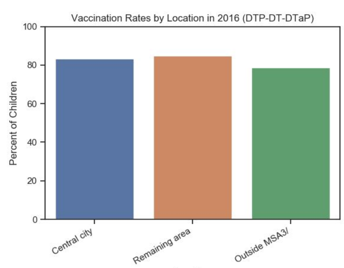


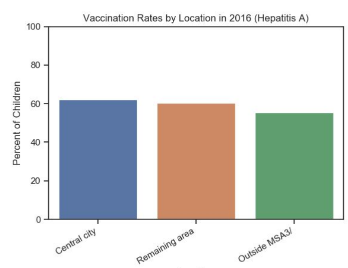


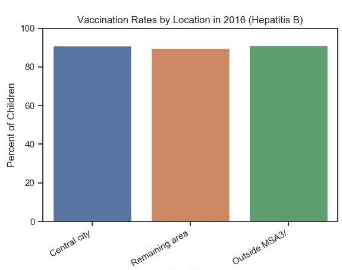
In this research question, we created graphs of the percentage of children getting vaccinated in families below the poverty level and families at or above the poverty level in 2016. The results we found were that families at or above the poverty level have a higher percentage of being vaccinated. This makes sense because families with a higher income are able to afford the costs of getting their children vaccinated. Families in poverty are unable to acquire these vaccines due to lack of transportation, couldn't get off work, didn't have insurance and perhaps didn't know that they could get vaccinations at clinics without insurance.

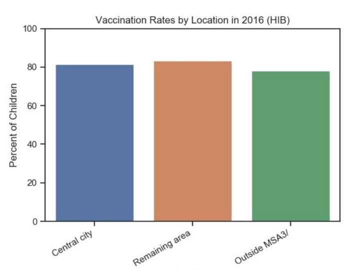
Research Question 3

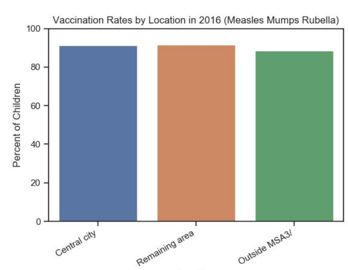


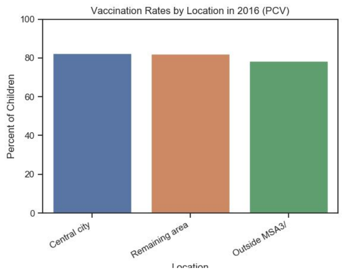


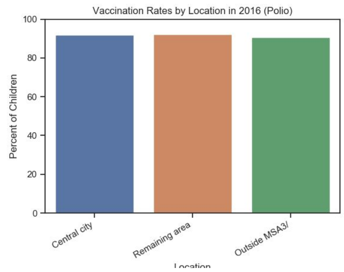


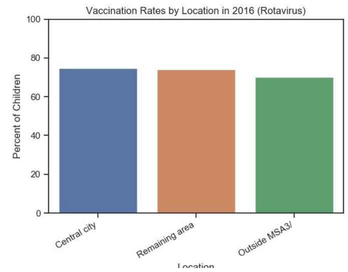


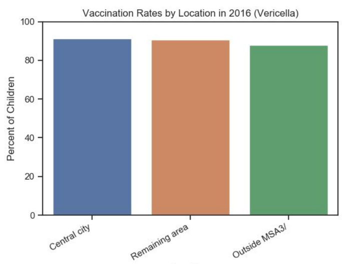






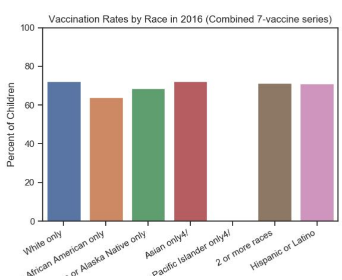


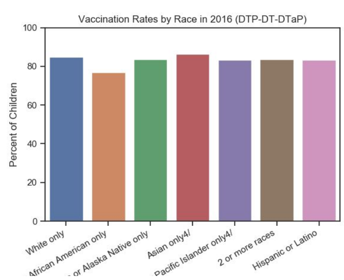


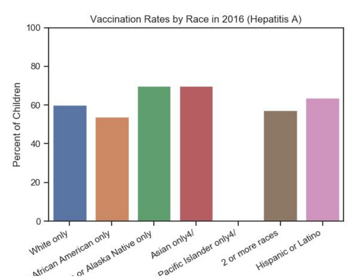


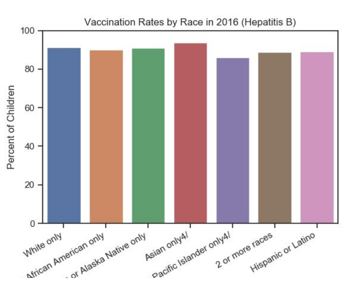
In this research question, we made graphs of the percentage of children getting vaccinated based on their location. These locations are inside the metropolitan statistical area (MSA), specifically in the central city and remaining area, and outside the MSA. We found that families inside the MSA have a higher percentage of children being vaccinated than outside. This is probably due to the fact that families outside the MSA are not in the vicinity of a clinic that offers vaccinations or they are unaware of vaccination programs such as Vaccines for Children, a program where children are able to get vaccinations for free, that are more prevalent inside the MSA region.

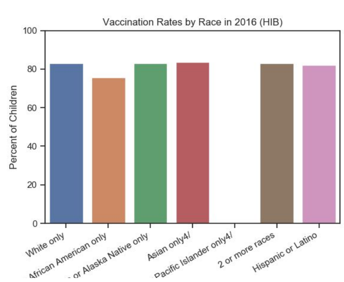
Research Question 4

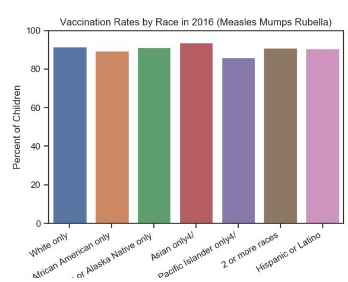


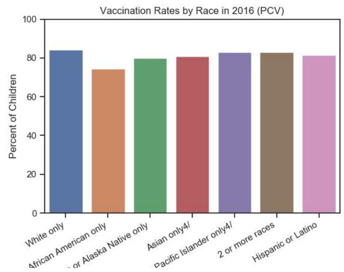


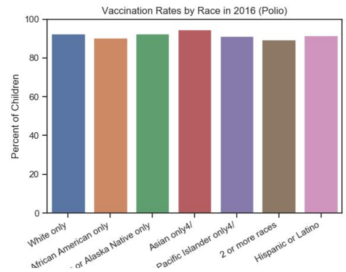


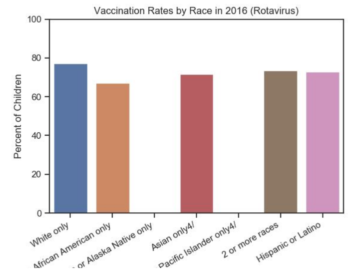


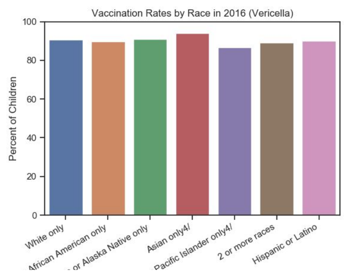






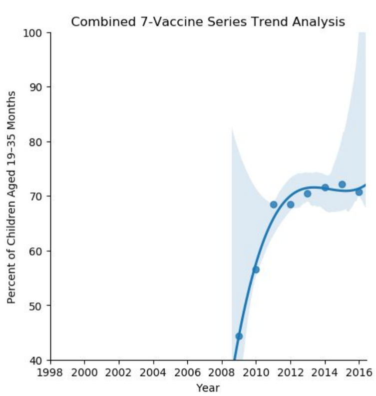


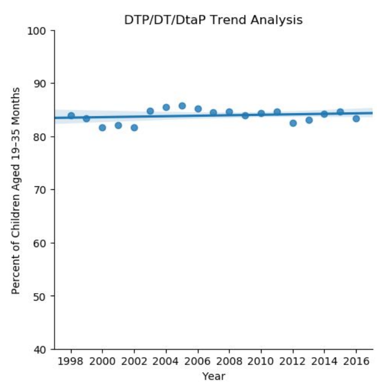


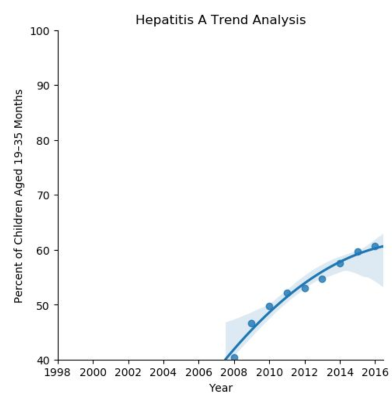


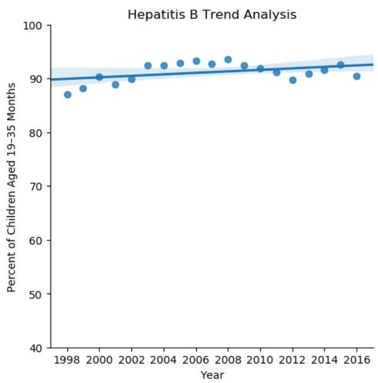
In this research question, we created graphs of the percentage of children getting vaccinated based on their race. Some of the data was missing so data for that race was unable to be graphed. We found that our results produced varying percentages of vaccination coverage based on race. In some cases the coverage is similar to other races in the graph. These results were produced probably due to families having a varied access to care factors such as clinics, vaccination programs or insurance. A surprising result we found was that Asians seem to have a higher percentage of vaccination in some cases and were unable to find the reasoning why this is so.

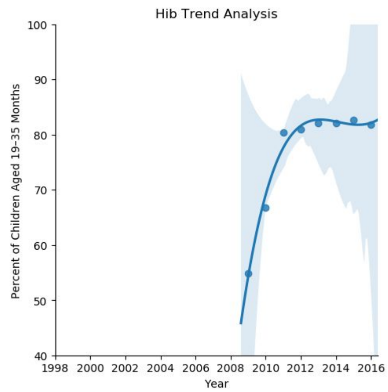
Research Question 5

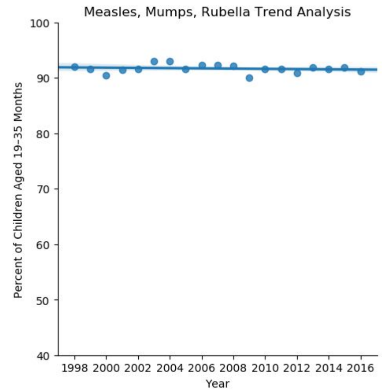


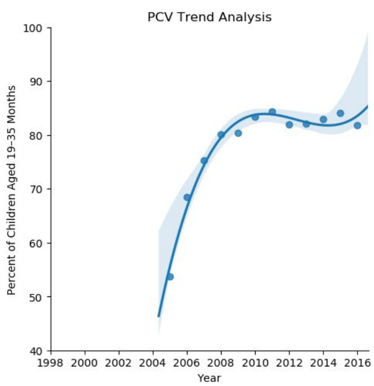


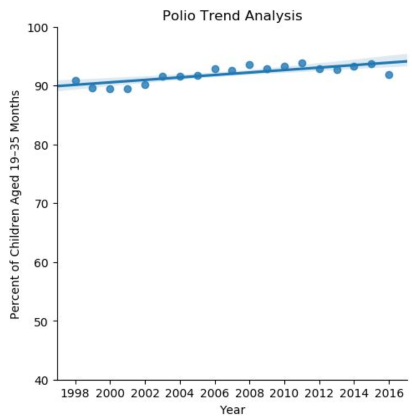


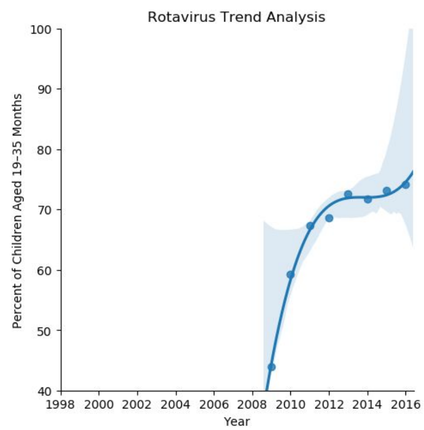


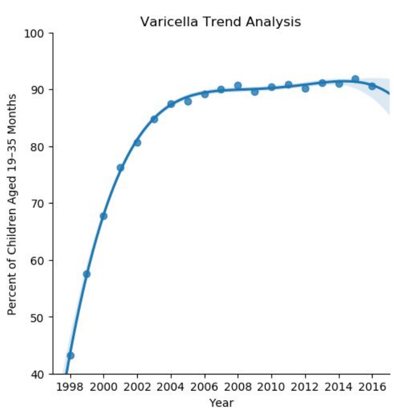












For this research question, the issue revolves around the situation going forward with each of these vaccinations. To get a feel for this, we used seaborn to create polynomial regressions that fit the data. They were of differing order depending on which would fit the data the best. What we found from this was that vaccines such as DTP, Hep. B, Measles, Mumps, Rubella, and Polio that started from 1998 had very flat trendlines. This indicated that though there may be some upward or downward sloping, the data generally indicates that those vaccines will not be changing very much going forward.

On the other hand, we have ones like the 7-vaccine data, Hib, PCV, Rotavirus, and Varicella that seem to rise quickly and then peter out. The regressions done indicate this with must having something of a flat trajectory going to the future. In the specific case of Rotavirus, there appears to be a distinct yet minor upward trend in recent years. The big outlier, as mentioned before, is Hep. A, which still seems to have a more noticeable upward path going forward. However, the order that fit best for it was order 2, which indicated a more parabola like shape. Now, the drop-off associated with a parabola will likely not happen and is more likely just an artifact of the math we used here. However, what this does perhaps show is that Hep. A is on the road to plateauing. If that is the case, it would likely end up around the 60-70% range. This would be much lower than other similarly shaped ones, except perhaps Rotavirus but that's on a little bit of an upward trajectory anyway. What this indicates about Hep. A, we don't know. We couldn't find any reliable outside evidence as to why that would be occurring with this one vaccine specifically.

Reproducing Our Results

To reproduce our results, create a new VSCode directory and then add the main.py file and the original csv file to it ("table066.csv"). Before running the file create three folders at the top level of the directory named "Location", "Race", "Poverty", "CSV Files" and "Trends". Pandas, Seaborn, and MatPlotLib libraries need to be installed. Running the file will create all the necessary csv files and create and save all the figures.

Work Plan Evaluation

1. Setup Git repository to work on code together (Less than an hour)
    a. This ended up taking 0 hours and 0 minutes as we didn't end up going with this method. We realized that since our project can be so distinctly compartmentalized into different methods it would be more straightforward (as neither of us had extensive git knowledge) to just work on individual parts at a time
2. Clean up data (Less than an hour)
    a. Use Python library pandas
    b. This ended up taking more like a full hour to just clean up the data. It ended up being a little more troublesome because we needed to split the original csv file into multiple ones (for each vaccine) and then parse each of those into an individual data frame.
3. Create graphs (Few hours)
    a. Use Python library seaborn and matplotlib
    b. Creating the graphs for the first and last research question was very long in and of itself, taking a few hours themselves.
    c. Creating the graphs for the second, third and fourth research question also took quite some time, taking a few hours to produce.
4. Write report on project (Few hours)

Testing

Truth be told, it was very difficult to do any reliable testing for our code. The reason for this is two-fold. For one, our dataset was already very small so it doesn't really make sense to test on a smaller dataset. It would basically be the same and any aspect of a smaller set that would make it easier to spot flaws or bugs would still be present in our own data. Secondly, there are no other graphs for which we can compare our data. We have no reference for what this *should* look like. As a result, the only way to "test" at all is to just eyeball the graphs based on our data and make sure it looks like what we're expecting. We did just such a thing and found no apparent issues.

Live Presentation or Video?

Live presentation!

Collaboration

The only other people that technically helped us were various users on Stack Overflow.

Works Cited

https://www.chop.edu/centers-programs/vaccine-education-center/vaccine-history/developments-by-year
https://www.npr.org/sections/health-shots/2019/05/20/724468630/the-other-reasons-kids-arent-getting-vaccinations-poverty-and-health-care-access