Project 1 Proposal

Arthur, Jon, and Sonny

The project will study how COVID-19 impacted airport traffic based on information in 2020 as compared to the baseline of pre-Covid traffic capacity. We will first look at data from the TSA showing the passenger throughput in the USA between March 1st, 2020 and Dec 1st, 2020. This data will then be compared to the same timepoints in 2019. This data was sourced from <https://www.tsa.gov/coronavirus/passenger-throughput?page=1>

After looking at how Covid affected travel numbers we will attempt to find relationships between travel numbers and total number of cases using data from the New York Times from March 1st to July 28th. <https://www.kaggle.com/roche-data-science-coalition/uncover>

We will find the correlations between this data and the travel numbers and also explore the deviations in case numbers between the two datasets.

We also will look through a few of the US’s most frequently traveled airports and look at individual trip numbers through a proportion of number of trips compared to the baseline of March 15th, 2020. We will obtain this data from <https://www.kaggle.com/terenceshin/covid19s-impact-on-airport-traffic>

We will look for the relationship between this data and the individual state data case numbers from the NYT dataset. We will examine the Atlanta airport and Georgia, Boston Logan and Massachusetts, San Francisco and California, LAX and California, Dallas-Fort Worth and Texas.

Hypotheses:

Traveler numbers have decreased in 2020 compared to 2019 by a statistically significant margin.

Traveler numbers decrease as Covid case numbers increase in a strong negative relationship.

Research Questions:

1. How have traveler numbers been affected by Covid in the US as a whole? (Sonny)

Using matplotlib, show in a line graph tsa[“total traveler throughput 2020”] on y axis and tsa[“date”] on x axis

1. How does this compare to more normal years? (Sonny)

Using matplotlib, show in one chart the 2020 line using tsa[“total traveler throughput 2020”] and the 2019 line using tsa[“total traveler throughput 2019”] with tsa[“date”] on x axis. Make sure to include a legend and different colors for the lines. Display line slope and analyze the differences between both years’ numbers.

1. How have case numbers changed throughout the year? (Arthur)

Using nyt\_nationwide create 2 line graphs. One with nyt\_nationwide["Total Cases"] on the y axis and nyt\_nationwide[“Date”] on the x and the other with nyt\_nationwide["Case Rate of Change"] on the y axis and nyt\_nationwide[“Date”] on the x. Talk about the differences between what these two graphs are measuring.

1. How do case numbers correlate with traveler numbers? (Jon)

Using combined\_total\_data, make 2 graphs: one with both 2020 travelers numbers line and nyt total cases line, the other with 2020 travelers numbers and nyt rate of change line. Then find the correlation between these lines for graphs. Talk about what this correlation means. Also explore how there could be a delayed reaction between case numbers affecting traveler numbers and vice versa. Utilizing a for loop, test earlier and later dates for case numbers and rate of change numbers against the original date numbers of travelers. Calculate the correlations for all of these and chart them in a bar graph. Find strongest correlation among all calculations and analyze all of this data.

1. What do case numbers look like in particular states? (Jon)

Using our 4 sample states of Georgia, California, Massachusetts, and Texas use the dataframes with the respective names to plot cases numbers and number rate of change against the date. Use similar coding structure as found in question 3. Analyze the data and talk about the differences between these data points. Create a graph with all 4 lines and observe some similarities and differences between states.

1. What do traveler numbers look like in specific airports? (Arthur)

Use our 5 sample airport dataframes for LAX, SFO, ATL, BOS, and DFW to plot traveler baseline percent vs date. Talk about what the percent baseline number means and why it is still a useful comparison tool even if the number by itself is hard to utilize. Create a graph with all 5 lines and observe some similarities and differences between airports.

1. What is the relationship between an airport and its state? (Together Saturday)

Use the merged dataframes of airport and state to make graphs of airport numbers vs date and state numbers/rate of change vs date. Find the correlation and analyze trends in the data.