The Impact of Air Travel on the Spread of COVID-19

PandAIRmic: jloh4, jneronha, tseet, ysun59

Hypothesis

It is an intuitive idea that air travel can act as a vector for COVID-19 spread. We sought to investigate the relationship between the amount of commercial air travel and COVID-19 infection rates. Specifically, we hypothesize that the number of domestic air passenger arrivals in a US state is positively correlated with the number of new COVID-19 cases in that state.

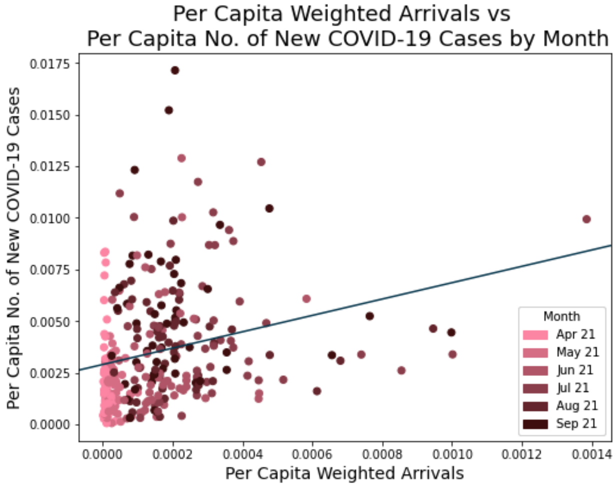
Data

We have 3 main sets of data that we used. Our COVID-19 data is obtained from The Atlantic's The COVID Tracking Project. It gives us the daily number of positive cases for each of the 50 states, 5 territories and D.C., and spans January 2020 to February 2021. Our air passenger arrival data is compiled from two datasets from the USDOT. The DB1B gives a 10% sample of all flight itineraries by quarter, and covers the first three quarters of 2020. The T100 breaks down the number of passengers flying between two airports (including layovers) by month, and covers Jan to Nov 2020. We combined the two datasets to estimate the number of passengers travelling between every pair of states per month, excluding layovers. Our policy data comes from the COVID-19 US State Policy Database developed by the Boston University School of Public Health. It tracks the dates when each US state implemented COVID-19-related policies.

Findings

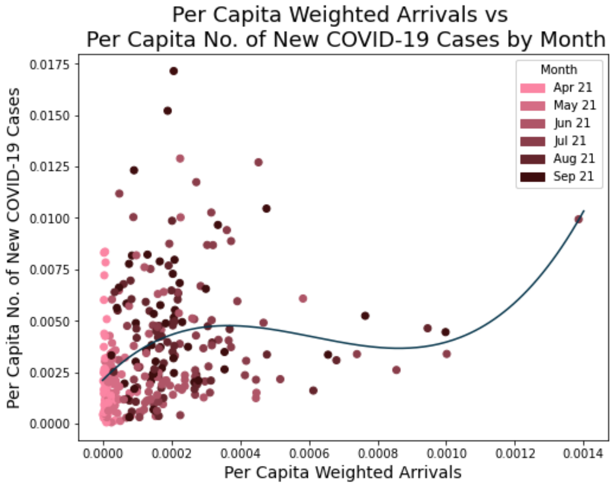
Claim #1: There is a positive correlation between the number of domestic air passenger arrivals in a state and the state's number of new COVID-19 cases.

Support for Claim #1: We first performed simple regression as a baseline analysis, and tested different metrics for the X and Y variables. We found that our best performing model is new COVID-19 cases per capita against weighted passenger arrivals, where the number of flights for each origin state was multiplied with the origin state's per capita COVID-19 cases. We did this weighing because having more flights from states with more COVID-19 cases will correlate with more COVID-19 cases in the destination state. The p-value was <0.05, and R^2 value was 0.065. The below figure shows the regression plot.



Claim #2: The relationship between the number of domestic air passenger arrivals in a state and the state's number of new COVID-19 cases is non-linear.

Support for Claim #2: We performed a polynomial regression with our X variable being the weighted passenger arrivals per state from the simple regression and our Y variable being COVID-19 cases per capita. We fitted examples to a polynomial curve of degree k = {1,...,9} and calculated R^2 values, finding that degree 3 provides the optimal R^2 value. With degree 3, the p-value was <0.05 for every X term, and R^2 value was 0.118. The below figure shows the polynomial regression plot.



Claim #3: A state's policies towards COVID-19 is also correlated to the number of COVID-19 cases.

Support for Claim #3: We performed a multiple regression that includes additional variables such as whether a state had a mask mandate or stay-at-home order together with the weighted passenger arrivals per state. The p-values of the weighted per-capita arrivals,stay-at-home order and quarantine order were < 0.05, and the R^2 value was 0.211. The below table shows our full regression results.

X-Variable	**P-Value**	**Coefficient**
Weighted passenger arrivals per capita	0.0331	1.9856
Stay at Home Notice	0.0318	-0.0010
Business Closure	0.0673	-0.0009
Facemask Mandate	0.0885	-0.0005
Quarantine Mandate	2.44 x 10^-6	-0.0019

Claim #4: There are no significant factors that vary across states but not over time, which are correlated with the number of COVID-19 cases.

Support for Claim #4: We performed a fixed-effects regression model to control for omitted variables in our panel data that vary across entities (states) but not over time (eg. Healthcare infrastructure, social/political norms). The p-value was > 0.05, and the R^2 value was 0.2743. Adding time fixed effects to the model does not significantly change the results. Thus our null hypothesis that there are no significant factors is not rejected.