Vaccine Impacts on Travel

Mai La, Matt Lyons, Matt Pribadi

Contents

1 Introduction

The 2019 novel coronavirus (COVID-19) pandemic had a massive impact on many aspects of life around the world. Although many industries were affected, the tourism industry, in particular, was devastated. Public policy, the unknown effectiveness of vaccines and treatment, and a tumultuous job market resulted in varying disruptions of travel and vacationing during the pandemic. Between federal policy, the provisioning of government workers, and differing approaches of state policy, there is quite a bit of data to be examined in terms of each of these individual impacts on the effect of travel and vacation destinations within the United States. As part of the tourism bureau, our team is interested in understanding how to restore our tourism industry to pre-pandemic levels and help city, county, and state economies to recover.

Our team proposes examining a specific research question to model the disruptions of travel from baseline based on the COVID-19 pandemic and vaccination rates. The critical question is: **How have vaccination rates affected travel to the most popular tourist county within each of the 50 U.S. States?**

There are many aspects that could indicate increases in tourism, such as accommodation booking and occupancy rate, nearby airport traffic, and increased travel to restaurants, retail and recreation areas. We are focusing primarily in the month of June 2021, on retail and recreational mobility data changes as a metric indicator for quantifying travel. This variable is further explained in the variables section below, but includes a number of significant indicators of travel such as traffic to museums, restaurants, and shopping centers. We believe that the retail and recreation data will best capture tourist activities across the varying destinations: from beautiful natural attractions, to small towns, to the busiest cities. This time period was chosen for many reasons: length of time from the start of the pandemic, the simultaneous start of summer vacation trips,

the lifting of several public mandates among the states, and the push by the U.S. government to increase vaccination rates. Note that while we are examining several variables, there are additional factors that may be correlated with these variables that are omitted from the model; these will be discussed in detail later.

This paper is structured as follows: Section 2 will include a more thorough definition of our research question and the causal model of interest, Section 3 will describe our data, variables, and the three main statistical models we have generated, Section 4 will further summarize the analysis of the regressions and coefficients that we obtained from these models, and Section 5 will include the conclusions and limitations of our study.

2 Research Question and Design

2.1 Research Question

Our goal of this report is to examine whether there is a statistically significant relationship between destination vaccination rates and changes in travel to popular tourist counties within each of the 50 U.S. states. These changes are measured by Google's Community Mobility Report as percent changes in "Retail and Recreation"-categorized trips taken compared to pre-pandemic levels.

Additionally, we are interested in understanding the relationships, not only between destination vaccination rates and mobility, but also what influence from political leaning, average income, education level, and population density may wield at these destinations. These factors could have a major impact on both travel and people's willingness to receive the vaccine.

The answers to this question will help clarify which factors are most important in rehabilitating the tourism industry nationwide. This will help the industry understand where and how to market tourism and in what other areas they should consider applying resources.

2.2 Causal Model

We hypothesize that a tourist destination is more attractive to travelers when it has a higher vaccination rate, because people feel it is safer to travel to these places while the COVID-19 pandemic is still active. Therefore, our causal model focuses primarily on understanding the relationship between mobility and vaccination rates.

Figure 1 summarizes the possible pathways that may exist between mobility, vaccination rates, and related attributes of tourism destinations. Our simplified model takes a look at the relationship between the solid green and orange lines. With increasing covariates in the model, we then examine the relationships to the simplified model with the solid blue lines. The dotted blue lines represent potential relationships which may exist between the covariates. For example, the average income and cost of living in an area are likely to affect its population density and desirability to tourists.

- **Vaccination Rate:** While it is intuitive to think improving vaccination rate could increase travel to a destination, there may also exist a reverse pathway from mobility back to vaccinations. One can argue that as more people come to the destination, a perception that the area is crowded could motivate community members to receive the vaccine. This reverse pathway could inflate the effect of vaccinations in a regression model.

- **Control Variables:** We categorized the variables possibly affecting both travel and vaccinations into four groups: conception factors, economic factors, demographics, and area qualities/characteristics. We believe the measures within each group will be somewhat related to one another. To study their effects, we will include at least one from each group as control variables:

  - **Political leaning:** When it comes to what factors affect a person's decision to receive (or not to receive) the COVID-19 vaccine, we believe politics plays an important role. With the development of social networks and ease in sharing information or spreading different theories about the pandemic and vaccinations, people with the same political views are highly influenced by others in the same group, and groups are often defined by their opposition to the opinions of the "other" group. Political leaning at a destination could also affect its travel inflow and attractiveness to
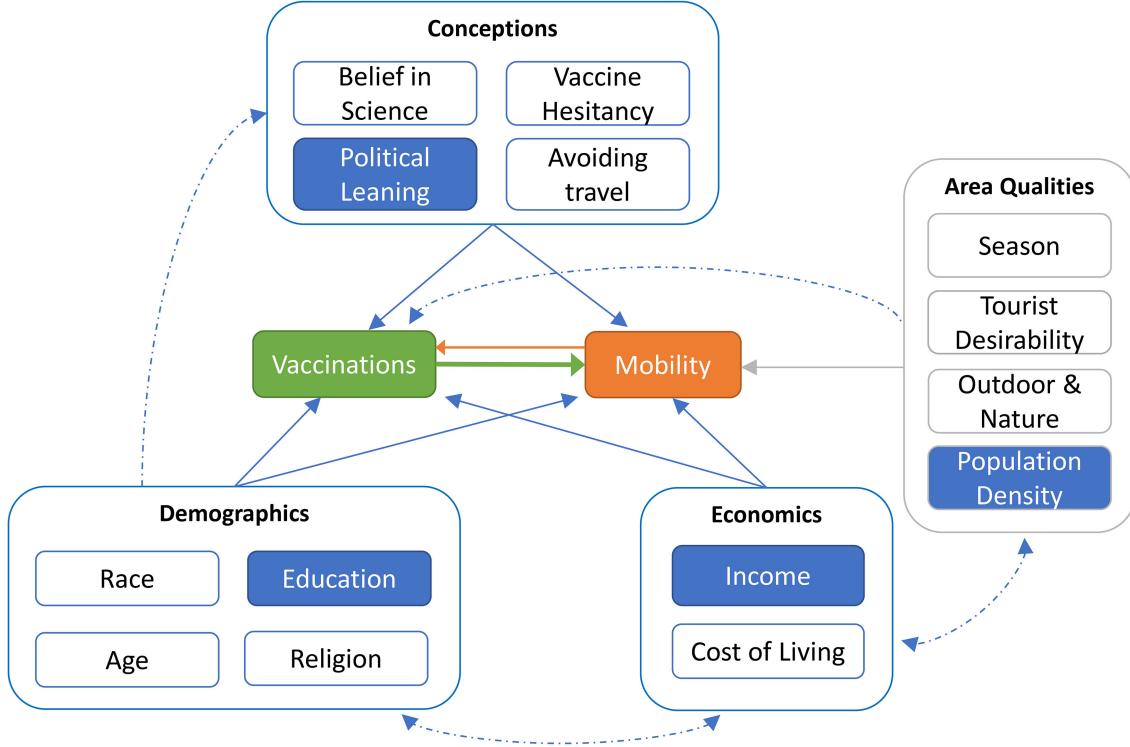
Figure 1: Variables of the causal relationship between vaccination rate and mobility

people from different states or counties. It could also correlate with people's willingness to forgo travel during the pandemic in order to protect themselves and others.

– **Income:** The average annual income at a tourist destination reflects the region's economics and, via taxation and/or charitable giving, investment in infrastructures and recreational facilities and services. A region with higher income may have a higher vaccination rate, perhaps due to the area's ability to purchase greater quantities of the vaccine or via improved infrastructure that speeds up distribution. On the other hand, a region with higher income is also likely to have higher costs and may for this reason be less attractive as a travel destination.

– **Education:** Education is likely to be positively correlated with vaccination rates. More educated individuals are more likely to believe in the science and effectiveness of vaccinations. Education may also have a positive impact on travel, as regions with better-educated populations may be more attractive to tourists.

– **Population** Density: We believe it is important to examine the effect of population density, as it is a measure of how close to one another people live, being of crucial importance in a pandemic. We believe that people living in high-density areas will be vaccinated at higher rates and that travelers will be less likely to choose to travel to densely-populated areas (tourist sites or not).

- **Omitted variables:** These are discussed in section 5.2 below.

3 Variables and Exploratory Data Analysis

3.1 Variable Selection and Data

The section below will examine our group's primary variables and exploratory data analysis. Each variable has been chosen with a specific intention and will be further explored with the model analysis in the following section. **Figure 2** shows a summary of the variables used in the model and how they are defined in terms of

this report.

Variable Name	Description
Attraction	The major attraction located in the county of interest
FIPS	United States FIPS (Federal Information Processing System) code
county	County name
state	State Name
mobility	June 2021 Google mobility data (divided by 100) for Retail and Recreation
vaccination_rate	Number of people completely vaccinatinated normalized by population in the county
annual_income	Average annual income within the county
pct_dem	Percentage of the population in the county that voted Democrat in the 2020 presidential election
some_college	Percentage of the population in the county that reported having at least some college education
income_index	Average county annual income normalized on $64,013 (average U.S. household income)
pop_density	Population density per square mile

Figure 2: Covariates used for model analysis

3.1.1 Travel Destination Mobility Data

As our main outcome variable, we examined mobility data as provided by **Google**. This dataset is looking at COVID-19 Community Mobility, intended for understanding the impacts of COVID-19 on travel across various sectors and locations within the United States, over the course of the pandemic. We primarily focused on the Retail and Recreation travel data from this dataset, which gives information on mobility trends for places such as restaurants, cafes, shopping centers, theme parks, and museums, among others. The data was generated anonymously, based on how the number of visits and the length of stay at these places changed compared to a baseline. This baseline is calculated based on the median value corresponding to the day of the week, during a 5-week period between January 3 and February 6, 2020. This baseline is significant because the percent changes in mobility are based on this initial 5-week period at the beginning of the year. Based on seasonality, mobility could be different depending on different time-periods of the year.

This dataset was chosen because it offered granularity in terms of dates of interest (data from June 2021), county selection, and affected industries within scope. The data were first filtered to include only the counties containing selected popular tourist destinations within each of the 50 U.S. states, as described below in section 3.1.2. Data was then generated by pulling June 2021 mobility data from the Retail and Recreation column, aggregated by the counties of interest (and the states and FIPS) to then give a mean of the mobility score in June 2021 for that county. The number is then divided by 100, to return the values to percentages. This mean mobility designation is then used in the models as an outcome variable.

3.1.2 Counties of Interest

In order to best filter out specific information from mobility data, counties of interest for each state were chosen. We cross-referenced **several popular tourist guides** suggesting top attractions in each state, picking the most commonly-referenced attractions unless they were closed during the pandemic (e.g. Disneyland and Universal Studios). In a handful of cases, alternate counties had to be chosen due to unavailability in data.

These counties were chosen solely due to the attractions located in them, in order to maximize the chance that the retail and recreational spend was due to tourism rather than local travel.

3.1.3   Vaccination Rate

One of our primary covariates for our model is the vaccination rate of the county of interest. We chose the destination of the mobility data because that is where the attractions are physically located and may be a major factor of travel to and from that county. In order to determine the vaccination rate for fully-vaccinated individuals (of all ages) in each county, we used public data from the **Centers for Disease Control and Prevention (CDC)**, which included data for the total number of people fully vaccinated and normalized by the reported population of the county. Two counties' data were unavailable in this dataset: Honolulu County, Hawaii and Bexar County, Texas. However, each of those data elements were separately available from official sources (Bexar County's **website** and the **Hawai'i Department of Health**).

To utilize vaccination rates for our model, we took the average percent of the population who were fully vaccinated during the month of June 2021 in order to ensure consistency within the time period.

3.1.4   Destination County Political Leaning

There can be little doubt that this pandemic has been politicized. Rumors about poisonous ingredients or microchips, general mistrust of science, and skepticism over the results of the 2020 election have led some to refuse the vaccine, believing that the U.S. government does not have its citizens' best interests at heart. Others have continued traveling just as before, while refusing to wear face masks. However, we believe that these political beliefs will explain only a portion of attitudes toward vaccination and travel. We created a new variable, `per_dem`, that calculates the percent of the county who voted for the Democratic party in the 2020 Presidential Election. By controlling for political leanings, we believe we will gain a clearer view of what is driving travel behaviors.

To measure the political leaning of a destination, we used a **dataset** scraped from several news sources' coverage of the 2020 presidential election, including the New York Times and Politico. The data source listed the number and percentage of votes each candidate received.

3.1.5   Destination County Income

As a control variable, including the average annual income of the county of interest may be an important variable to include due to its direct impact on travel rates for the reasons mentioned in the introduction. Both vaccination rates and travel are factors of privilege and therefore may skew the outcome variables if this variable was not included in the model building exercise.

To create this dataset, we started with extracting the annual average income for 2020 by county from the **U.S. Bureau of Labor Statistics report**. We chose the aggregated data for annual average pay for all industries at county level. We then created an `income_index` by standardizing the county's average annual income by dividing it by the U.S average household income for 2020, which was reported as \$64,013 in the dataset.

3.1.6   Destination County Education Level

Our goal in adding the county's education level as a control variable was to isolate vaccination rate from both income level and education level, as we believe they will be somewhat correlated. If the true cause of the effect is vaccination rate, the statistically significant effect will remain present after controlling for income and education level.

This dataset was obtained from the **Department of Agriculture's Economic Research Service**. For the model building exercise, we took a subset of data for the percentage of adults who have reported at least some college education in each county from the 2015–2019 series.

3.1.7 Destination County Population Density

In addition to the income index and education levels in the county, population density within the county seems to be an important aspect of travel as well. There are specific attractions across the United States that are in larger cities and in those cities, the infrastructure that the city provides may be a valuable point of interest. For example, San Francisco is listed as one of the attractions of choice, where tourists can see the Golden Gate Bridge, Alcatraz, and the naturally sloping elevation changes the city provides. Population density is also another indicator of urban versus rural, that could reflect the differences in the behavior of vaccinations and travel between different states and counties.

The dataset was generated from population density figures available from **the US Census Bureau** at the county level. Because of varying degrees of population density across the counties, we expect this to be a variable that would need transformation for our models below.

3.1.8 Exploratory Data Analysis

While examining the data, we generated several histograms to understand the distribution of the data as shown in **Figure 3**. Based on the distribution of each variable, we chose to transform only one of our variables, population density, due to its high variability. For this variable we chose to do a log transform, which generated more normalized data. Although several of the variables are correlated with one another, examining the Variance Inflation Factors, all of which are well under 5, suggests that they are moderately, though not problematically, correlated with one another. A further analysis of the limitations of these variables and models are described in Section 5.

The following plots of distribution show some skewness for most variables. For retail and recreation mobility data, the low density count on the right could potentially inflate the fitted model's error on this side due to less data points. There is also one outlier at about a 200% increase in the mobility rate, which belongs to Wyoming for traveling activities near the Yellowstone National Park. The region has the lowest population density per square mile in the dataset and substantially less than the other regions, which implies being highly inflated by tourist inflow. Other factors not included in this study like season and having beautiful natural scenery may also explain this result.

The distribution of education looks slightly different than the rest of variables: while most show one mode, education data show three modes. Due to small sample size (50 observations), the distribution of these variables might not be well representative of the population, but could also be an artifact of the bin size we created.
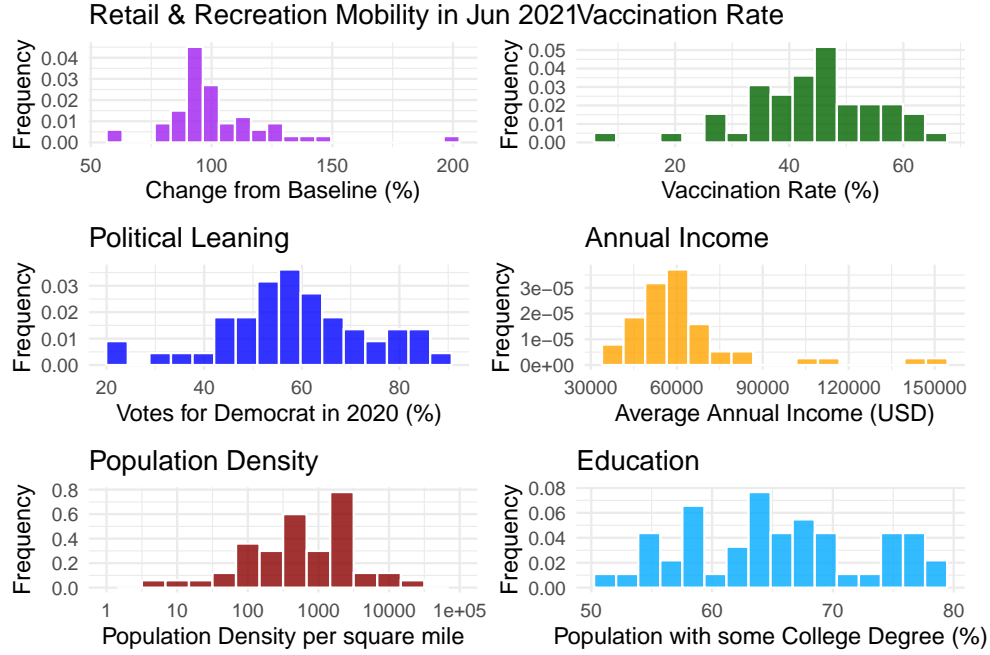
Figure 3: Histograms of Variables of Interest

Figure 4 below shows scatter plots between mobility rate and its predictors, which describe linearity trends between these variables. Vaccinations effect on mobility has a slight reduction trend but not obvious. Political leaning, income and population density show a clear decreasing tendency. This could reflect the high correlation between these variables. Whereas, education seems to have a neutral effect on mobility.

3.2 Model Building Process

The section below will summarize the models that will help to answer our research question and explore the relationship between vaccination rates and changes in travel to the counties of interest. We will discuss three models that vary from a simplified model to models that include control variables in order to provide a more comprehensive understanding of the causal relationships behind our research question. The goal of our modeling is to enable the research team to build a data-driven plan in order to resuscitate the tourism industry, and in turn, the economy.

Model 1

We begin first with our simplified model, which looks primarily at our output variable and the main covariate of our research question: vaccination rate. Here we develop a linear model with mobility data as a function of vaccination rate as shown in the simplified model below. Here, our primary coefficient of interest is β_1 that measures how vaccination rates of the county of interest affect mobility rates.

Model 2

Adding on to the covariate in Model 1, we add on political leaning in the form of how much of the county voted for the Democratic party in the 2020 election. This linear model is shown below under Simplified Model with Political Leaning, with β_1 and β_2 being the coefficients of interest in order to understand the effects of the county's political leaning and vaccination rates on mobility and travel within this county.

Model 3

Adding on to the covariate in Model 2, we add on control variables in the form of a calculated income index, average college education in the county, and the log of the population density within the county. This
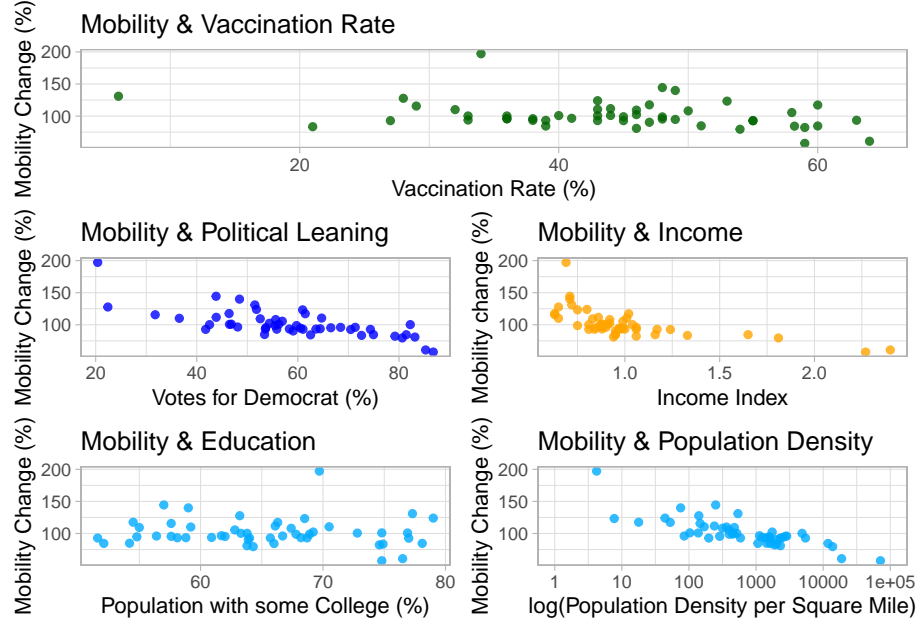
Figure 4: Exploration of Variables of Interest

linear model is shown below under Full Model with Control Variables. With the addition of β_3, β_4, and β_5 coefficients, our team is exploring the effects of the county's annual income and education on mobility and travel within this county.

$$\textbf{Simplified Model:} \tag{1}$$
$$mobility = \beta_1 \cdot vaccination\_rate + \epsilon \tag{2}$$
$$\tag{3}$$
$$\textbf{Simplified Model with Political Leaning:} \tag{4}$$
$$mobility = \beta_1 \cdot vaccination\_rate + \beta_2 \cdot pct\_dem + \epsilon \tag{5}$$
$$\tag{6}$$
$$\textbf{Full Model with Control Variables:} \tag{7}$$
$$mobility = \beta_1 \cdot vaccination\_rate + \beta_2 \cdot pct\_dem + \beta_3 \cdot income\_index + \tag{8}$$
$$\beta_4 \cdot some\_college + \beta_5 \cdot log(pop\_density) + \epsilon \tag{9}$$

4 Results

4.1 Regression Table

Table 1 below is a summary of the coefficients, constants, and significance of our linear regression models using robust standard error. A summary of the significant variables are described below.

- For the simplified model (model 1) only looking at the effects of vaccination rates on changes in mobility, the linear model produces a β_1 of -0.629 with a p-value less than 0.05 (at 0.028 with the coefficient test) and an epsilon of 1.295.

- For the model including political leaning and vaccination rates (model 2), the linear model produces a β_2 of -1.010 with a p-value less than 0.01 and an epsilon of 1.644 with a p-value less than 0.01. We no

longer have a statistically significant coefficient for vaccination rate.

- For the full model including the control variables (model 3), the linear model produces a β_5 of -0.204 with a p-value less than 0.05 and an epsilon of 1.325 with a p-value less than 0.01. We do not have statistically significant coefficients for political leaning, vaccination rates, and college education.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Mon, Aug 02, 2021 - 03:43:33 PM

Table 1: Mobility and Covariate Model Regression

	Dependent variable:		
	Mobility in June 2021		
	(1)	(2)	(3)
Vaccination Rate	−0.629**	−0.083	−0.156
	(0.278)	(0.245)	(0.262)
Percent Democratic Vote		−1.010***	−0.427
		(0.269)	(0.266)
Income Index			−0.008
			(0.137)
College Education			0.046
			(0.362)
log(Population Density)			−0.147***
			(0.056)
Constant	1.295***	1.644***	1.724***
	(0.133)	(0.177)	(0.308)
Observations	50	50	50
R^2	0.107	0.518	0.684
Adjusted R^2	0.089	0.497	0.648
Residual Std. Error	0.209 (df = 48)	0.155 (df = 47)	0.130 (df = 44)
F Statistic	5.762** (df = 1; 48)	25.233*** (df = 2; 47)	19.019*** (df = 5; 44)

Note: *p<0.1; **p<0.05; ***p<0.01

4.2 Analysis of Results

The results are interesting and somewhat surprising. In Model 1, as expected, we found that increased mobility was correlated with vaccination rate ($p < 0.05$). The direction was not what we had anticipated, however: as vaccination rate in a destination county increases, travel to that county tends to increase less strongly. In particular, with 1% increase in vaccination rate, we see a 0.6 percent decrease in mobility rate. From a causal standpoint, this does not seem logical—would travelers not want to minimize their exposure to risk by traveling to counties with a greater share of vaccinated individuals? As there are many omitted variables in this model and the effect of them vary from biased towards zero or biased away from zero, the coefficient of vaccination in this model mostly is not representative and the practical significance of this coefficient is not very valuable.

In Model 2, we see that the effect initially explained by vaccination rate is instead better explained by political leaning, with counties that voted in greater proportion for Joe Biden in 2020 seeing relatively fewer visitors. We lost the statistical significance for vaccination as in model 1, and see statistical significance for political leaning instead ($p < 0.01$). Specifically, with a 1% increase of votes for Democrats, we have a 1% reduction in mobility rate (holding vaccinations constant). This, too, is surprising. We expected to find that vaccination rate was somewhat correlated with political leaning, but that the original effect would persist. Instead, we find that a significantly greater amount of the mobility effect is better explained by the Democratic share of the county's overall vote in 2020. Why might this be?

Model 3, which we expected to clarify and confirm our results, does bring clarity, but with a different explanation than we had anticipated. After introducing controls for income, education level, and population density, the most sensible explanation emerges: all else being equal, travelers are choosing to travel to destinations with lower population density. Only population density has statistical significance, with each tenfold increase to population density per square mile corresponding to a 15% lower increase in mobility rate, all else remaining the same ($p < 0.01$). In a way, this confirms what we initially believed: travelers appear to be attempting to mitigate their risk of exposure to COVID-19. However, the operationalization was not what we expected, with this behavior operating via population density rather than through vaccination rates.

We attribute the initially confusing results from Models 1 and 2 as consequences of having selected variables which are correlated with one another. As we discovered when examining the correlation plot, virtually all of our variables had some degree of positive correlation. Well-educated individuals tend to (1) earn more, (2) choose to receive the vaccine at a higher level than their less-educated counterparts, and (3) choose to live in cities, which by definition have higher population densities (and also tend to (4) vote blue). We believe, then, that the results from Models 1 and 2, though statistically significant, were capturing only part of the true explanation by way of their correlation. This did give us pause as to the variables we had chosen, but after examining their VIF (as mentioned earlier), we do believe they are capturing distinct attributes of a tourist destination and why travelers may choose one rather than another. The F-test result is also in favor of Model 3 with very highly significant results ($p < 0.001$) compared to Models 1 and 2. Adjusted R-squared value is also improved in Model 3.

4.3 Practical Significance

As we did not find consistent results for the coefficient of vaccination and lost its statistical significance once adding the control variables into the model, one could argue that there is no relationship between vaccination and travel. However, due to the limitations of the linear regression model and particularly the effects of omitted variables as discussed in the following section, we would argue that the result of this study does not imply that vaccination is not a meaningful predictor of mobility during the COVID-19 pandemic. In order to better answer this question, a more complicated model, or at least one incorporating more data, would be required.

Regarding the magnitude of the effect discovered, a 15% reduction in relative mobility increases per tenfold density increase (*ceteris paribus*) is not insignificant: consider a county with 10 people per square mile and one with 10,000. The second county would have likely seen roughly half the increase of the less densely-populated one. We believe the effect does have practical significance and reflects the caution people are displaying when choosing travel destinations (everything else held equal).

5 Conclusions and Limitations

5.1 CLM Assumptions and Limitations of the Models

With small sample size (50 observations in the dataset) we use the Classical Linear Model (CLM) to explain the relationship between mobility and its explained variables. In the following sections we will validate our model against these assumptions.

5.1.1 Assumption 1: Independent and Identically Distributed Data

We select 50 different counties randomly from the same distribution (i.e. top tourism destinations), from rural areas with beautiful scenery to urban and modern environments with entertaining retail activities. The samples are identically distributed.

All the attractions are from different states, so they could be considered as independent. However, they may not satisfy the IID condition fully because the counties in nearby states could be considered as clustered data. For example, there may be regional clustering behavior within the South, the Pacific Northwest, or the Midwest. However, these regions are not monoliths, with each state in the region having their own distinct characteristics and tourist attractions, so we do not believe these effects invalidate the model. Besides, by including the control variables with the demographic, geographic and economic information for these regions, the clustering effect could be mitigated.

5.1.2 Assumption 2: Linear Conditional Expectation

Figure 5 shows the residual versus fitted plot for model 3. Although we have an outlier among our residuals (Wyoming's portion of Yellowstone National Park, which has seen particularly large increases in visitors), the shape of the residuals plot suggests than a linear conditional expectation appears reasonable—most of the data points show the same general trend and there is no obvious other shape they appear to be approximating. The deviation of the residuals away from 0 on the further right is mostly because of not enough data points on this side, rather than being representative for the population. We believe if more samples were selected and the effect of large sample size kicked in, this deviation could be eased out.

5.1.3 Assumption 3: Homoskedasticity

Examining **Figure 5** and the Scale-Location plot, we do not appear to have homoskedastic error terms. Further, the studentized Breusch-Pagan test results in a p-value of 0.017, rejecting the null hypothesis that the error terms are homoskedastic.

Because heteroskedasticity is present, we used heteroskedastic-consistent standard errors. The models and tables above have already been modified to use robust standard errors for the coefficient t-test.
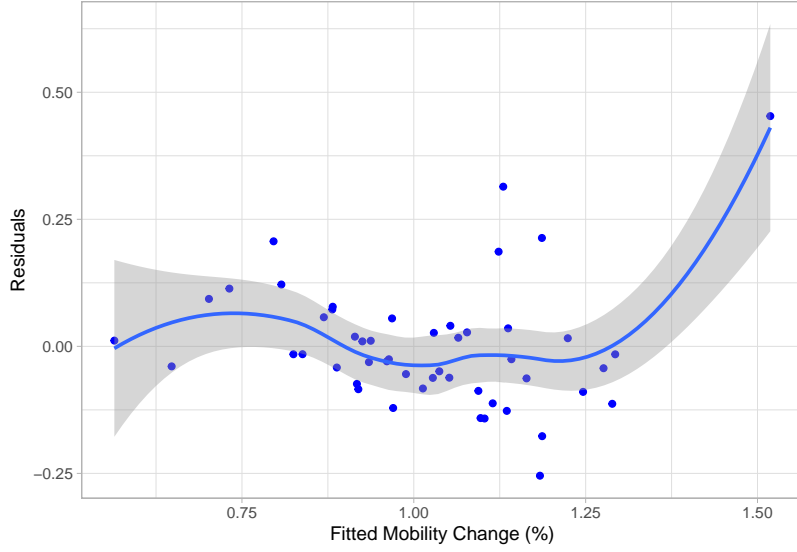
Figure 5: Model 3 residual plot versus fitted values plot

5.1.4 Assumption 4: No Perfect Collinearity

No covariates were dropped when we performed linear regression, and the correlation matrix in **Figure 6** also confirms no perfect collinearity exists. There is a high correlation between political leaning, income and population density, which may represent what happens in most major cities (e.g. New York, San Francisco) versus rural areas (e.g. Wyoming, Maine). However, by including the three highly correlated control variables in our full model, one of the variables may be effectively redundant. The variance inflation factors for each variable suggest dropping income if we chose to drop a variable, but without correction, it would be inappropriate to run further tests.
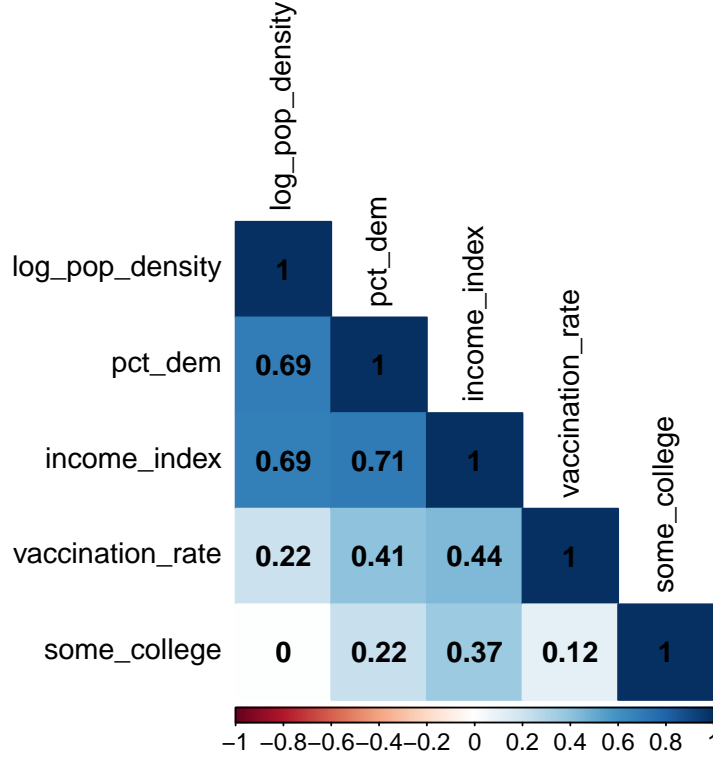
Figure 6: Correlation of the covariates for our model

5.1.5 Assumption 5: Normally Distributed Errors

Giving the small sample size, the residual distribution looks quite normal as shown in **Figure 7**. There's a slight heavy tail on the right which is caused by the outliers and lack of data points on this side. In order to help alleviate this we would need to redesign our model to include more than just the 50 attractions and counties we have generated. In addition, we applied t-test for testing the statistical significance of coefficients using the robust standard error.

5.2 Omitted Variables and Bias

The following summarizes the possible impacts and direction of bias for our omitted variables. Note: we conceptually believe that vaccination rates, when all the correct controls have been applied, will have a weak positive correlation with mobility. Although we do not have all of the correct controls in place or the statistical resolving power to make this effect and its direction clear,*these biases will be discussed under the assumption that the "true" direction of this relationship is positive.*

- **Tourist Desirability:** Although using a baseline specific to each area will help minimize this effect, the percent increase from baseline may not at all be proportional to other factors we are measuring. If vaccination rates have a positive impact on mobility to a destination, we believe similarly that desirability will have a positive effect on mobility. Additionally, because we believe that many desirable areas will tend to be higher-income and more liberal, we believe that more desirable locations will tend to have higher vaccination rates. Therefore, the direction of this bias will likely be away from zero.

- **Race:** This will vary tremendously by region and by group and resists simple explanations. However, a majority of the population is white. From that perspective, "whiteness" **will tend to be correlated positively with vaccination rates**. We believe a desire to travel to "familiar" areas, particularly in uncertain times, will cause this bias to be narrowly away from zero.
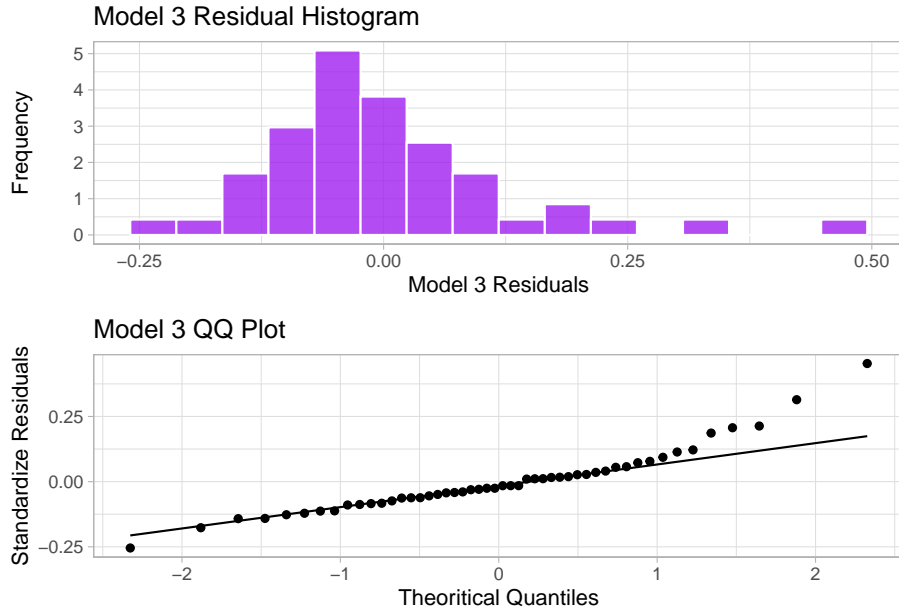
Figure 7: Error Distribution for Model 3

- **Age:** If only because of the late rollout, the youngest populations are less likely to be vaccinated. Additionally, older populations, being more at risk, are more likely to be vaccinated. However, we believe that travelers will find destinations with younger populations more appealing, so the effect of this bias will likely be toward zero.

- **Gender:** Although not noted in the diagram above, we would expect a small gender effect in the negative direction. This is because, given the **rare blood-clotting side effects** of some vaccines on women in certain age groups, we expect women to vaccinate at a slightly lower level than men. Although we do not anticipate a causal reason for destinations with more women to be more traveled-to (besides perhaps that **women tend to do more of the hospitality and tourism-related work**), knowing that **the US population skews slightly female** is enough to predict that the bias would be toward zero.

- **Religion:** If we were to designate a generic "religiosity" variable, it seems likely that high religiosity would be negatively correlated with vaccination rates For example, Catholics and Jehovah's Witnesses have each displayed some amount of hesitancy toward the vaccine, though for different reasons (Catholics oppose the use of fetal cell lines in the development of certain vaccines, and Jehovah's Witnesses had a long-standing ban on receiving vaccines which has since been reversed but likely has had a long-standing chilling effect nonetheless). Under ordinary circumstances, it may be the case that travelers enjoy visiting cities, but in the current climate, population density is negatively correlated with mobility, so we would expect a positive correlation between mobility and religiosity. Therefore, the direction of this bias would likely be toward zero.

- **Cost of living:** Areas with high cost of living will likely tend to have higher vaccination rates. However, we suspect travelers may prefer to avoid traveling to the most expensive areas in favor of more affordable vacations. Therefore, we expect this bias to be slightly toward zero.

- **Belief in science:** This is likely to be positively correlated both with vaccination. The closest variable we were able to measure was education level, which turned out to be weakly positively correlated with mobility. Therefore, we expect this bias to be slightly away from zero.

- **Vaccine hesitancy:** This will be negatively correlated with vaccination rates. As for its relationship

14

to mobility, there are almost two camps: COVID deniers and minimizers will likely tend to be traveling more frequently than others, and the hyper-cautious will tend to be traveling less. Because that first camp is likely to be larger, we expect this bias to be toward zero.

- **Avoiding travel:** This is a different sort of caution measure. In this instance, we would expect those who avoid travel to generally be more cautious and have higher rates of vaccination, as well as lower rates of travel. Therefore, the bias would likely be toward zero.

There is no such thing as a perfect model. Even if we had access to data measuring each of the variables listed above, or more variables still, there is no guarantee the results would be significantly more meaningful or accurate. However, it is our responsibility to reason about the variables we could not measure or chose to omit. Of the variables we omitted, we predict that three would be away from zero and six would be toward zero. Though we cannot be sure of the relative magnitudes of their effects, it does not appear that they would tend to inflate effects or create effects where none otherwise existed. Therefore, despite having omitted these variables, we feel confident that the effect discovered by our model is real.

5.3   Conclusions

Our initial question sought to examine the effect of vaccination rates on travel to popular tourist destinations in each state. We expected that a push for higher vaccination rates would, in turn, increase tourist spending. In the end, this did not turn out to be the case. Instead, we found that the driving factor behind people's chosen destinations is a different operationalization of risk: population density. Although the mechanism was not what we expected, this behavior makes perfect sense in the framework of our causal model.

There are various approaches the bureau could take when examining how to further encourage tourism in the United States while it recovers from COVID-19. The first approach would be to amplify our "strengths"—that is, to promote travel to rural destinations like national parks and outdoor vistas. We expect travelers to have less hesitance in returning to these ever-popular destinations. On the other hand, we also know that national parks are already seeing record numbers of tourists and they may not effectively have additional capacity, given the limited staff and resources available to maintain the parks.

The other main approach would be to shore up our current "weaknesses"—that is, to encourage travelers to return to cities and other, largely man-made, attractions. Although this has the most potential in terms of revenue, given the recent guidance from the CDC relating to the spread of the Delta variant, we feel such a recommendation may be ill-advised, even irresponsible.

Instead, we recommend a third, middle path: promote travel to small towns and rural areas which are still beautiful but which will be less crowded than cities. Nearly every state has hundreds of miles of national forests, grasslands, and other underutilized natural splendor. As the Delta variant winds down, vaccination rates continue to increase, and tourists feel safer, we can slowly adjust our sights upward in terms of promoting travel to larger population centers like New York City, Orlando, and San Francisco.