# What made the Pacific Northwest move during the July 4th weekend?

Eleda Johnson, Sam Stephens, Mahesh Arumugam

08/05/2021

## Contents

# What made the Pacific Northwest move during the July 4th weekend?

## 1 Introduction

The effects of the COVID-19 pandemic have been felt and are still being felt around the world. In particular, the tourism industry suffered extreme economic losses between the many travel restrictions, cessation of communal social activities, and widespread fear of contracting the virus. Specifically, in 2020, there was a 35% reduction in domestic travel, resulting in a loss of over 300 Billion dollars [1]. By contrast, there is an uptick in domestic travel in 2021 [2].

The recent dramatic decline in US COVID-19 cases is primarily attributed to vaccination programs on the local and national levels [3]. As of May 2021, many social restrictions such as interstate travel bans and mask mandates have been lifted, though not consistently across the US.

The rebound of domestic travel this summer is happening with no end in sight for the COVID-19 pandemic. Surges in new cases are still frequent and a 6% US unemployment rate, higher than the near 4% pre-pandemic rate [4], could slow the pace of recovery for the tourism industry. Vaccination rates have been variable at the state and county levels, particularly influenced by those vocal against trusting vaccinations like those in strong support of 2020 Presidential Candidate Donald Trump. People aligning with conservative political ideologies are generally dismissive of the social restrictions and generally fewer restrictions were enacted in areas where those political ideologies were majority in 2020 [5].

For tourism revitalization in Oregon, a key question then is what factors are affecting vacation mobility this summer? In particular we are interested in the impact on travel taken within our state and the Pacific Northwest area for regional context. We would like to understand whether the recent travel trends suggest people are getting more confident to travel since the pandemic started. And perhaps, understanding what makes people move will help to optimize funding allocation on a county level across our state more effectively.

In this study, we seek to establish a causal relationship between these variables[1]. We note that this study does not completely model a formal causal relationship because such models require complete understanding of socio-economic factors that contribute to people going on vacations (or road trips). Such concepts are beyond the scope of this report. Hence, our research question is:

| Research.Question |
|---|
| **What factors contributed to the decision of embarking on road trips during the summer of 2021 by the people in Pacific Northwest (WA, OR, ID)?** |

In particular, we are interested in understanding how vaccination rates are impacting vacation mobility in the Pacific Northwest during summer 2021. We will explore the relationship between number of trips between 100 and 500 miles and COVID-19 vaccinations controlled for key related attributes, including population density and number of COVID-19 cases. In addition, we also consider several additional factors that influence travel, such as employment status, state restrictions, and political ideology. We also note that this study only focuses on a small time period -July 4th weekend.

The rest of the report is organized as follows. First, in Section 2, we formally introduce the variables and data sources and provide an exploratory analysis of the data. Then, in Section 3, we state our models formally and summarize these models. Subsequently, in Section 4, we discuss the Classical Linear Model assumptions and present our conclusions on whether our models and data meet these assumptions. And, in Section 5, we discuss potential causal pathways and how it affect the vacation travel during the pandemic. Finally, in Section 6, we summarize our findings.

[1] Originally, we intended to study the causal theory affecting summer travel. However, in this report, we only focus on causal pathways from COVID-19 related attributes including new cases prior to the travel and vaccinations. This research does not model a complete causal theory for summer travel.

# 2 Data and Exploration

In this section, we identify our variables and data sources, examine the relationship between the outcome variable and explanatory variables, and provide some initial analysis of the data. We gathered data for as many variables we identified as having a potential relationship to the number of trips, specifically those mentioned in Section 1. Included with this report are the cleaning scripts used to prepare interim data files from the primary data sources which were joined into the used dataframe. Smaller datasets, like employment data by county, were cleaned and prepared in Microsoft Excel. Cleaning steps at this point were limited to typecasting a variable name standardization by column.

## 2.1 Data Description

To operationalize our research question, first, we identify the variables and datasets used in our analysis. Table 1 lists the variables and the data sources and more information is provided for each variable. With the exception of population density and the restrictions indicator, all variables are normalized by population and transformed using natural logarithm. Normalization by population helps account for the large variability between county populations. Taking the log transform of these per capita variables allows for a more normal distribution to be used in the model.

The three Pacific Northwest states have a wide range of political views and population densities, which we consider the strongest factors for travel distance not related to the pandemic. We expect these states will have a reasonably diverse sample from which to determine correlations, as seen in the regression table (cf. Table 4).

As discussed in Section 1 we noted that we are interested in the July 4th weekend that typically draws a lot of road trips. Specifically, we focus our study from Wednesday prior to July 4, 2021 to the Wednesday following July 4, 2021. More precisely, our time window is between June 30, 2021 and July 7, 2021. Focusing on this particular weekend poses a number of advantages. As a national holiday it is a four day weekend for many workers and observed by all counties in our study area. This avoids the potential for local events to cause clustering on different time intervals over the summer months. It is also the first major holiday weekend long enough to make a substantial road trip during the pandemic, but also with a significant amount of vaccine distribution and relaxation of state-level social restrictions. Memorial day weekend was also considered, but that is usually only a three day weekend and fewer vaccines had been distributed.

**Dependent Variable (Trips Between 100 and 500 miles)**. The dependent variable of *travel over the July 4th weekend* was operationalized through a rich dataset compiled by the Bureau of Transportation Statistics looking at Trips by Distance [6] on a county level. The data was aggregated at the county-level from the weighted and de-trend GPS of millions of mobile devices. Trips were defined as stays of longer than 10 minutes and encompassed all modes of transportation. These data are reported in distance *buckets* divided at 100 miles, 250 miles, and over 500 miles.

Of interest to the local state tourism would be those traveling for the holiday within the Pacific Northwest. For reference, a Portland, OR to Seattle, WA journey would be around 174 miles and a trip from Portland, OR to Boise, ID would be around 430 miles. So, we focused on trips for a July 4th weekend in the *buckets* between 100 and 500 miles to differentiate longer, holiday travel from daily commute distances and to maintain a focus region within the Pacific Northwest, versus longer car or plane journeys. To calculate this variable we subtracted all trips over 500 miles from all trips over 100 miles. Our assumption was that travelers taking trips over 500 miles would make at least one stop on the way between 100 and 500 miles.

Figure 1 shows distribution of number of trips between 100 and 500 miles during July 4th weekend travel window. By adjusting for population in a county, we get a distribution that shows normality with a natural log transformation. Thus, adjusting for population allows us to effectively control for counties with huge population densities and reduce the variability.

**Population** (used in normalization and to calculate density). In order to be consistent with the output variable and because census data available is relatively old, county population was calculated from the BTS data [6]. The BTS data contains values for *people staying at home* and *people not staying at home*, of which population is the sum.

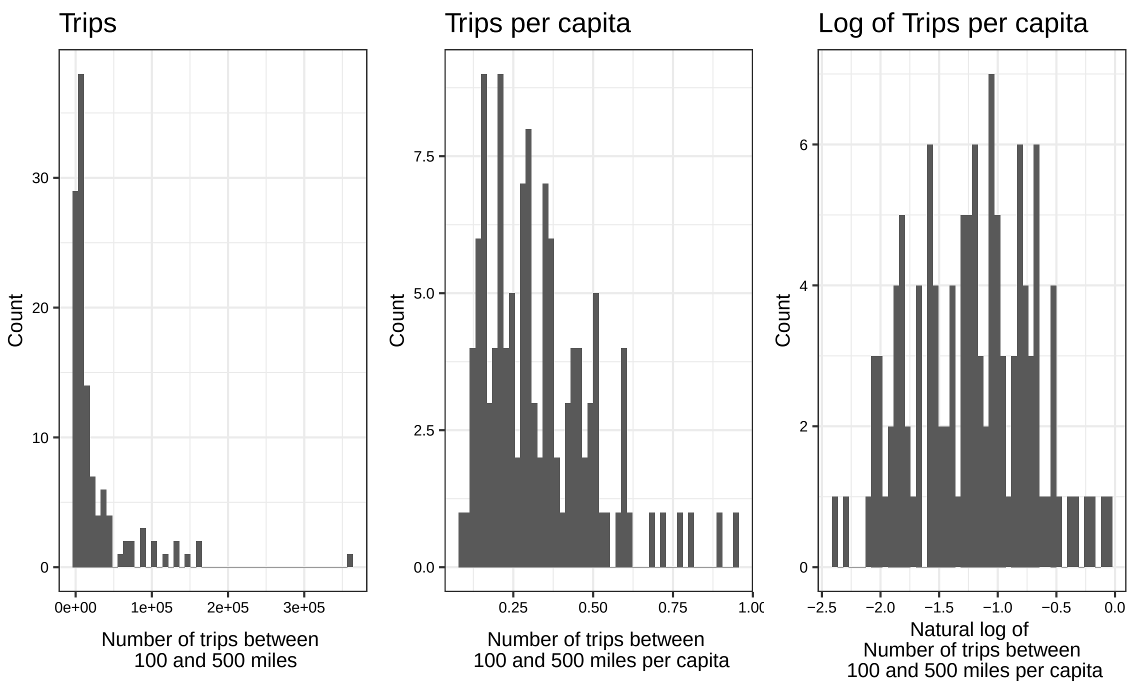


Figure 1: Distribution of the outcome variable

**Vaccination Rate**. We addressed vaccination rates using recently updated data from the Center for Disease Control and Prevention (CDC) [7] looking at vaccination rates from across the US reported at the county-level. The data is compiled from multiple sources across government and private entities and at variable reporting intervals. The CDC data only contained both case and vaccine data for about 6 in 10 US counties, and that entire states were missing. However, all data were available for all counties in our states of interest (Washington, Oregon, and Idaho). We found the number of total cumulative cases in San Juan County in Washington to reduce by 1 in the 10 days preceding the travel window. This is an obvious error in data recording, which we manually corrected by setting San Juan county to zero new cases as to have minimal impact on our analysis. Also, our focus on the recent July 4th weekend gives us fairly high assurance the database is relatively complete for our analysis. We take this variable as the number of total vaccinations in the county at the start of our July 4th time window and normalize it by population using the BTS data sourcing our dependent variable.

**New COVID-19 Cases**. For those traveling over the July 4th weekend we considered the influence of the number of new COVID-19 cases in a vacationers immediate county as having an impact on their travel plans. COVID-19 cases per US county were sourced data from the New York Times dataset [8] based on reports from state and local health officials. To operationalize this, we determined the raw number of new cases per 100,000 people in the 10 day period *prior* to the holiday weekend. The normalization by population was done using the BTS data population calculation. We considered this a look at the intensity of the pandemic in each country and a reasonable window of time to influence travel plans.

**Population Density**. To calculate population density we pulled the size of each county in square miles of land as of 2020 from the US Census Bureau [9]. Distances between destinations are greater in rural areas. We expect fewer people in more dense areas to make trips of our distance of interest. We expect population to be a strong driver of travel distance mostly unrelated to the pandemic, so this variable is included in all our models.

**Conservative Political Ideology**. The impact of political ideology on how people perceive the COVID-19 pandemic has implications for choice to vaccinate, enact local restrictions, and willingness to travel over a major holiday weekend. We chose to use votes for Presidential candidate Donald Trump in the 2020 general election [10] as a proxy for political ideology. The higher a county's voting rate for Donald Trump the more we

Table 1: Variables and Datasets

| Variable | Dataset | Type | Description |
|---|---|---|---|
| **Number of trips between 100 and 500 miles during July 4th Weekend in a county** | Bureau of Transportation (BTS) Trips By Distance dataset | Metric / Number | Trips between 100 and 500 miles to differentiate vacation travel from regular commute and flight trips. |
| **Percentage of people that completed vaccinations in a county** | Center for Disease Control and Prevention (CDC) COVID Vaccination dataset | Metric / Percentage | Percentage of people that are completely vaccinated at the start of the July 4th weekend travel window. |
| **New COVID-19 cases in the county** | New York Time COVID Dataset | Metric / Number | Number of new cases 10 days prior to the July 4th weekend travel window. |
| **Population density of the county** | U.S. Census Bureau Dataset | Metric / Number | Computed using the population of the county and the area. |
| **Number of people voting for Republican Presidential candidate in 2020 General Election** | MIT Election Dataset | Metric / Number | This variable identifies conservative political ideology of the county. |
| **COVID-19 restrictions** | Boston University School of Public Health Dataset for COVID State Policies | Indicator | Identifies restrictions (closure of businesses such as restaurants and bars) that were active during the July 4th weekend. |
| **Employment rate in the county** | Bureau of Labor Dataset for Employment Statistics | Metric / Number | Percentage of people employed in the month of June 2020 in the county |

attribute the county aligning to conservative ideologies (generally dismissive of social restrictions, increased skepticism of vaccinations, more likely to recreationally travel) and visa-versa for more liberal ideologies. Trump votes were also normalized by county population using the BTS population estimate.

**COVID-19 Restrictions In Force**. We used the CUSP [11] dataset, compiled by Boston University School of Public Health, to integrate in control indicator variables for COVID-19 state-level restrictions. The data, last updated on July 27, 2021, includes 272 variables from all 50 states and Washington, D.C. Indicator variables for each statewide restriction that included at least a *start date* was created for active and inactive policies at the start of our model window on June 30, 2021. For the Pacific Northwest region, the main active restrictions over the 2021 July 4th weekend were a general mask restriction and a second round of business closures (restaurants, bars, movie theaters) in both Oregon and Washington.

**Employment Rate**. Given effects of the pandemic on the US economy, employment status likely impacted decisions to travel over the summer months. We looked specifically at the most recent labor statistics for US employment by county [12] provided by the US Bureau of Transportation, available for June 2021.

## 2.2 Exploratory Data Analysis

Next, we generated histograms of our variables of interest for the three states. Variable distributions were initially highly skewed. This kicked off an iterative process of applying variable transformations and re-generating the histograms until we produced reasonably normal distributions for each variable. The result was that we applied the natural log to all our normalized variables. One was added to all variables prior to taking the natural log in order to obtain finite variables. Figure 2 shows the characteristics of the variables.

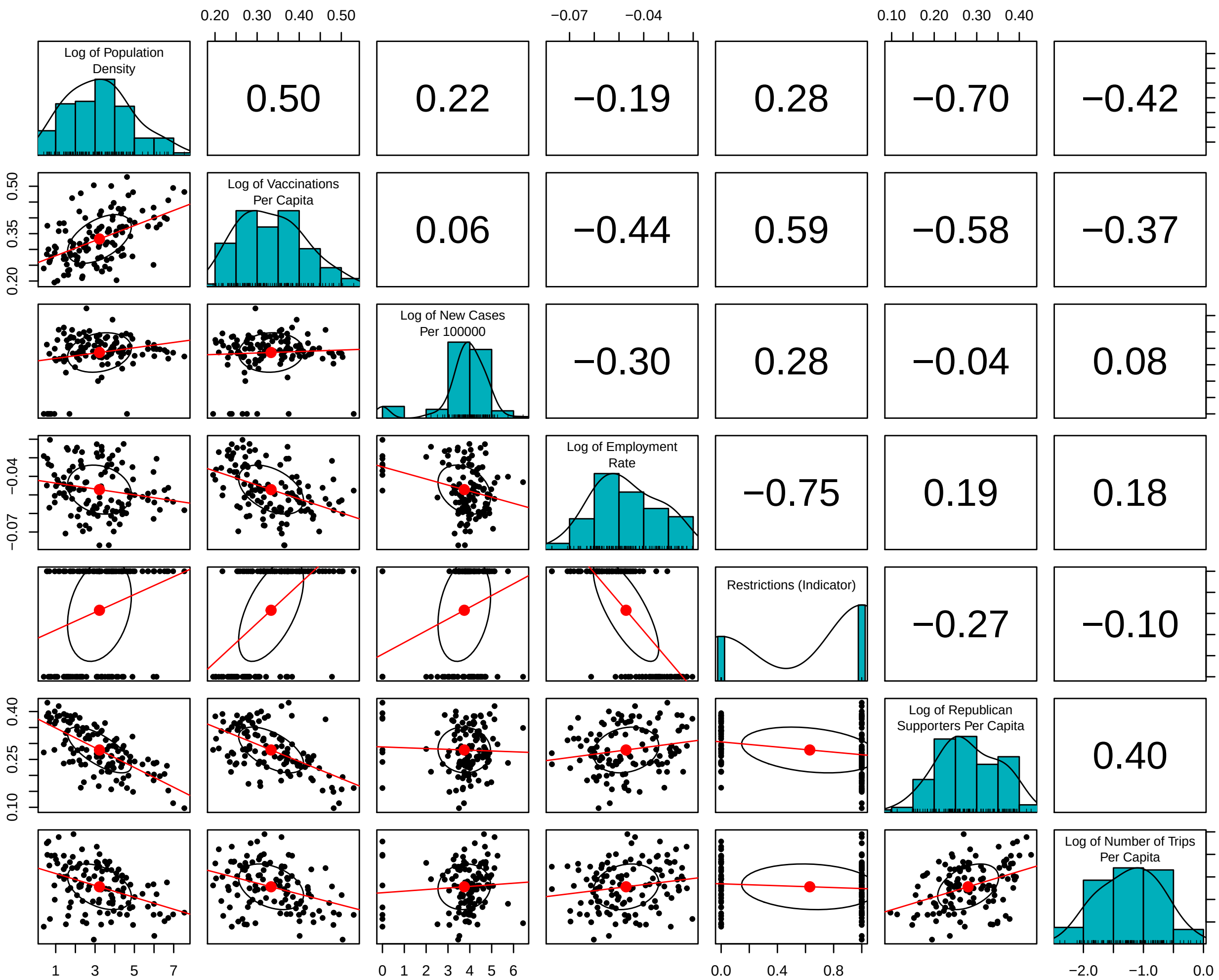


Figure 2: Feature characteristics

The scatterplots in Figure 2 show how the variables are distributed and correlated with other variables (including the outcome variable). We note that population density has the highest correlation with the outcome variable (=-0.4227156). This correlation indicates that as the population density increases there is a decrease in the number of trips. This relation is expected as people in highly dense areas tend to have more social connections and access to recreational places. As a result, they seem not embark on long road trips. On the other hand, people in sparsely populated areas inevitably have to make long road trips for vacations. Population density is also correlated with vaccinations (=0.5030382). Densely populated regions (e.g., urban/cities) tend to be more inclined to get vaccinated compared to people in sparse regions (e.g., rural areas). Furthermore, population density also indicates how people are inclined to vote (correlation with support for Republican Presidential candidate =-0.7006223). Rural regions typically favor conservative policies and vote for Republican candidates in the general election. Whereas, urban areas tend to favor liberal policies. As a result, population density is an important factor in our analysis. And, from our regression table (cf. Table 4), we confirm that population density is statistically (and practically) significant in all our models.

Counter-intuitively, vaccination rate is negatively correlated with the outcome variable (=-0.3688423). In other words, as vaccination rates improve, there is a decrease in the number of trips (holding other covariates constant). And, the number of new COVID-19 cases 10 days prior to the holiday weekend has a small positive correlation with the outcome variable (=0.08353).

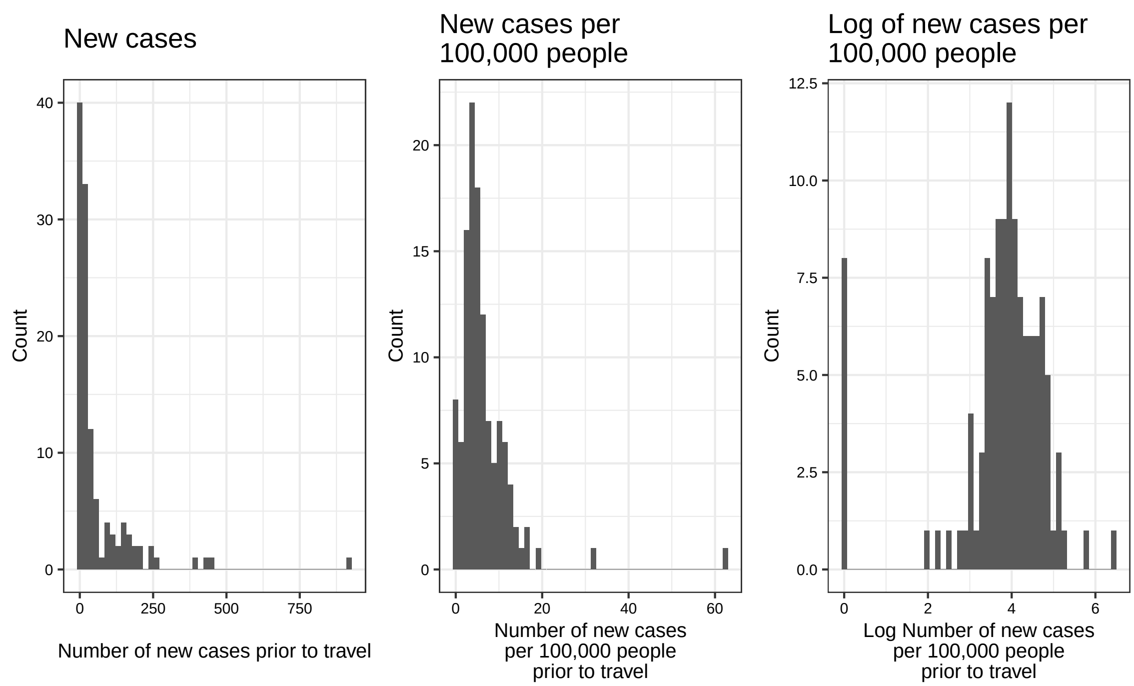


Figure 3: Observations about new cases leading up to the holiday travel

Figure 3 show the number of new COVID-19 cases in the 10 day period prior to the holiday weekend. The raw scatterplot (cf. first plot in Figure 3) shows a right skew that corresponds to King County in WA (where Greater Seattle area belongs to) with a population density of 1066.0298384. To adjust for this skew, we compute new cases per 100,000 residents during the 10 day period prior to the holiday weekend. With this adjustment, we observe a small right skew (cf. second plot in Figure 3) that corresponds to Benewah County in ID. The number of cases in this county is unusally high (= 58) for its population (=9298). Finally, we take a natural logarithm of number of cases per 100,000 residents to make it close to normally distributed (third plot in Figure 3). We note that there a few counties that report 0 cases during the 10 day period leading up to the holiday weekend. Most of these counties are in Idaho and are valid data points. However, one of other counties is San Juan county that had a data-entry issue that showed number of new cases as -1. We fixed this data entry issue such that the number of new cases in San Juan county prior to holiday weekend is 0.

# 3 Models and Regression Table

In this section, we present the regression models. We chose population density and vaccination rate to be our primary variables of interest across all three models. The population density variable was included in all models because it has the largest correlation with trips (=-0.4227156) and because it is a property of each county mostly unrelated to the pandemic. The goals and development process of each model is summarized below.

1. **Model One**. Model One is intended to be the simplest model of pandemic-related travel behavior. The vaccination rate variable was chosen as the other covariate (correlation =-0.3688423). Votes for the

Republican 2020 presidential candidate have a high correlation with trips (=0.08353) but were excluded models One and Two because it is highly inversely correlated with population density (correlation = -0.7006223), almost to the point of being a proxy.

2. **Model Two**. Model Two sought to expand on Model One by adding another variable directly related to the pandemic, the number of new COVID-19 cases per 100,000 people in the 10 days preceding the travel window. Models one and two show statistical significance in their results. However, It should be noted that Model Two achieves significance on vaccines only with the use of classical standard error.
3. **Model Three**. The model three goal was to include as many variables as possible to determine both the stability of vaccines and cases as significant variables, and whether other factors were stronger or more significant drivers of the number of trips.

These models vary in their meaning, precision, and possible applications; all of which will be discussed in conclusion. Model outcome variable is described in Table 2 and the variables are described in Table 3. Model coefficients, errors and adjusted $R^2$ statistics are reported in Table 4.

Table 2: Outcome for the models

| Outcome | Transformation | Model.Outcome |
|---|---|---|
| **Number of trips between 100 and 500 miles during the July 4th weekend travel window by peopel in the Pacific Northwest** | Natural Logarithm | *ln_trips_100_500_per_pop.21* |

Table 3: Covariates for models

| Covariate | Transformation | Model.Variable |
|---|---|---|
| **Population density of a county** | Natural Logarithm | *ln_pop_density* |
| **Vaccinations per capita in the county at the start of the travel window** | Natural Logarithm | *ln_vaccinated_first_per_pop* |
| **Number of COVID-19 cases per 100,000 people in the county during the 10-day period prior to the start of the travel window** | Natural Logarithm | *ln_pcases_per_100kpop.21* |
| **Restrictions on businesses (such as restaurants, bars, movie theatres) that were in effect at the start of the travel window** | None | *rstr_businesses_active* |
| **Number of voters per capita in the county that voted for Republican Presidential candidate during the 2020 General Election** | Natural Logarithm | *ln_trump_votes_per_pop* |
| **Employment rate of the county in the month of June 2021** | Natural Logarithm | *ln_employment_rate* |

Table 4: Regression table

| | *Dependent variable:* | | |
|---|---|---|---|
| | ln_trips_100_500_per_pop.21 | | |
| | (1) | (2) | (3) |
| ln_pop_density | $-0.099^{**}$ | $-0.113^{***}$ | $-0.099^{**}$ |
| | (0.031) | (0.030) | (0.038) |
| ln_vaccinated_first_per_pop | $-1.374^{*}$ | $-1.293^{*}$ | $-1.512$ |
| | (0.641) | (0.619) | (0.820) |
| ln_pcases_per_100kpop.21 | | $0.073^{*}$ | 0.071 |
| | | (0.035) | (0.056) |
| ln_employment_rate | | | $10.985^{*}$ |
| | | | (5.296) |
| rstr_businesses_active | | | $0.331^{*}$ |
| | | | (0.163) |
| ln_trump_votes_per_pop | | | 0.610 |
| | | | (0.928) |
| Constant | $-0.442^{*}$ | $-0.699^{**}$ | $-0.522$ |
| | (0.201) | (0.221) | (0.551) |
| Standard Error Type | Robust | Classical | Robust |
| Observations | 119 | 119 | 119 |
| Adjusted $R^2$ | 0.198 | 0.220 | 0.248 |
| *Note:* | $^{*}$p<0.05; $^{**}$p<0.01; $^{***}$p<0.001 | | |

### 3.1 Interpretation and Significance

Models One and Two show statistical significance for all covariates. Model Three is only significant for business restrictions and employment, but also has the highest adjusted $R^2$, and so explains the most variation. The overall practical significance of the models is mediocre, as the adjusted $R^2$ indicates that only a small amount of the variation in the outcome is explained by any of the models. Population density, vaccinations, and case rate have relatively stable coefficients across all three models. F-tests were not performed between the models because models One and Three were found not to have Homoskedastic errors (cf. Section 4.4) and therefore use Robust instead of Classical standard errors.

One important finding from our models is that the vaccination rate is inversely correlated with road vacation trips in all three models. Especially, in Model Two, a 1% change in vaccination rates results in a decrease of 1.3% in road trips. We suspect this is mainly because people who sought out the vaccine early are also more cautious of traveling long distances during the pandemic. And, it is also important to note that case rate does not seem to influence road trips strongly. A 1% change in case rate equates to only a 0.07% percent change in trips. Perhaps the pandemic itself is not a large driver of whether or not people take road trips.

## 4 Model Assumptions

In this section, we discuss whether our model satisfies the Classical Linear Model assumptions.

## 4.1 IID

We acknowledge a general regional clustering by restricting our samples on the Pacific Northwest states of Washington, Oregon and Idaho. These states share features such as regional weather events and cultural norms. We move the discussion of IID assumptions down to the sub-state level by selecting a representative sampling from this regional cluster grouped by county. At this scale, we have a complete dataset for all our variables of interest at the county level meeting the assumption for identically distributed. There are still potential geographic and demographic clustering issues confounding the assumption of sample independence such as urban vs rural and coastal vs mountain splits. Our sampling process mitigates some of the potential for clustering in variables such as number of new COVID-19 cases over a 10-day window and vaccination rate by normalizing for county total population. However, as we chose to include a major population density as an independent variable, this is likely not enough to ensure strict independence. We can reasonably proceed with the assumptions for IID acknowledging our interpretation of the results likely applies to the local county clusters over suppositions about the entire region.

## 4.2 No Perfect Collinearity

This assumption is satisfied for all models because none of the covariates are perfectly collinear. We use Variance Inflation Factor (VIF) [13] as an indicator of multicollinearity. Specifically, VIF measures how the variance of a regression coefficient inflated due to multicollinearity [2]. The highest VIF of Model Two was for population density at 1.4073067. The highest VIF of any of the models was the binary indicator for restrictions on business in Model three at 2.9081573. Thus, based on VIF scores, we believe that there is no perfect collinearity in our models.

## 4.3 Linear Conditional Expectation

From Figure 4, we observe that linear conditional expectation assumption is satisfied for model two because the smoothing line of the residuals plotted against the fitted values is relatively flat. The very slight waves to the line deviate from zero no more than about a half a standard deviation of the outcome.

## 4.4 Homeskedastic Errors

Model Two satisfies the homoskedastic errors assumption due to failure to reject the homoskedastic null hypothesis of the studentized Breusch-Pagan test [14] (p-value = 0.0596848). Also, while the Scale-Location (Figure 4) admittedly does show a slight rise in standardized residuals with the increasing outcome values, the range of the standardized residuals is within acceptable bounds for the homoskedastic errors assumption. Therefore, we submit Model Two with classical standard errors. If Model Two is used in an application where a higher number of trips per county occur, then robust standard errors should be a consideration with an acknowledgement that statistical significance may decrease.

Models One and Three cannot be considered to have homoskedastic because they may reject the homoskedastic null hypothesis of the Breusch-Pagan test (p-value = 0.0036351 for Model One and p-value = 0.0274211 for Model Three). Models One and Three are therefore presented with robust standard errors.

## 4.5 Normally Distributed Errors

The normal distribution of errors assumption was found to be satisfied for the vast majority of counties. We assessed this assumption by means of the Q-Q plot shown in Figure 5. The model tends to very slightly overpredict trips in counties both with very high and very low rates of travel, but errors are normally distributed across most outcomes.

[2] We adopt the rule of thumb specified at https://www.statisticshowto.com/variance-inflation-factor/ for interpreting VIF. A VIF score of 1 indicates there is no correlation with other regression variables. A VIF score between 1 and 5 is a sign that the variable is moderately correlated with other regression variables. Finally, a VIF score greater than 5 indicates that the variable is higly correlated with other regression variables.

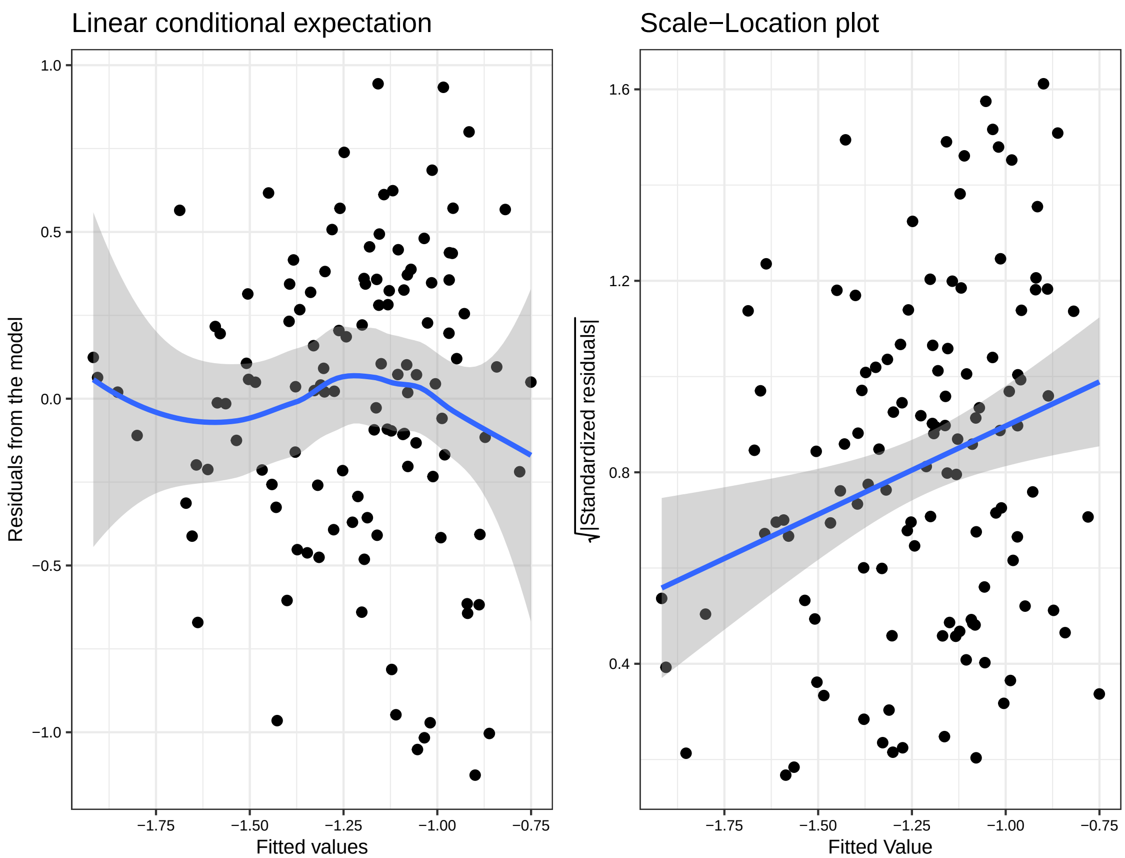


Figure 4: Linear conditional expectation and Homoskedasticity for Model Two

# 5 Potential Causal Pathways

In this section, we seek to reason about potential causal pathways between the number of trips that are between 100 and 500 miles taken during the July 4th weekend in the Pacific Northwest and the vaccinations for COVID-19. First, we focus on the variables related to COVID-19 that showed significance in our models (specifically, Model Two).

## 5.1 Causal Pathways from Simplified Model (Model Two)

Figure 6 shows the casual pathways we identified for Model Two.

**Percentage of people vaccinated in the county**. The percentage of people who are vaccinated tend to be more conservative about their exposure to the pandemic. As a result, vaccinated people are more likely to not take long road trips (greater than 100 miles that are typically inter-state travel in the Pacific Northwest). Hence, high vaccination rates are generally negatively related to long road trips. We distinguish here the causal pathway between vaccination rate and our measure of travel and the reported relationship between vaccination rate and decreasing new cases across much of the country referenced earlier.

**Number of new COVID-19 cases leading up to the July 4th weekend travel window**. On the other hand, if the recent trend in the number of COVID-19 cases increases, since people want to enjoy their much needed vacation after the long pandemic, they would prefer to get vaccinated prior to the travel window. And, rather surprisingly, a high number of new COVID-19 cases leading up to the travel window would encourage people to take the trip in anticipation of going into a lockdown soon. Note that there is no reverse causality from number of trips to number of COVID-19 cases prior to the travel window. Specifically, trips made during the July 4th weekend cannot affect the number of COVID-19 cases in the 10-day period prior to

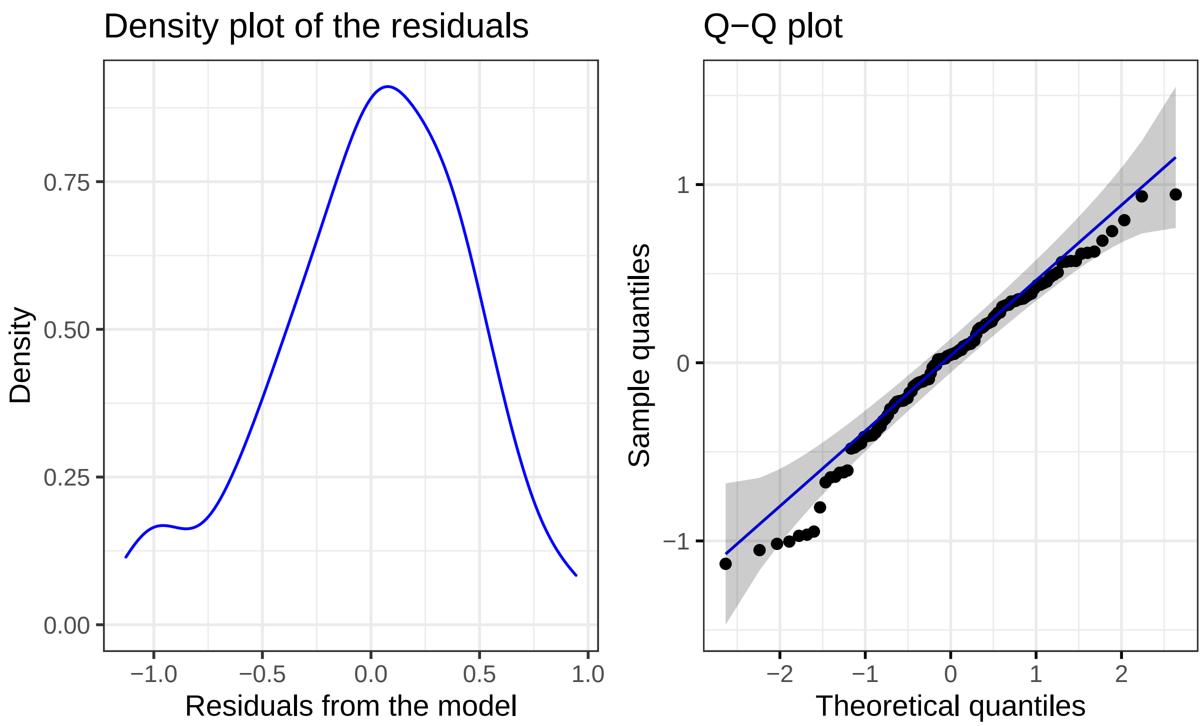


Figure 5: Normally distributed errors

the weekend.

**Population density of the county**. Finally, population density plays an important role in travel. Typically, densely populated counties tend to be urban areas that have an abundance of recreational activities and avenues for enjoying the weekend. On the contrary, sparsely populated areas tend to be rural areas that are remote and require road trips to go to places that have recreational activities. Therefore, population density has a negative impact on the number of trips. However, population density is also a confounding variable for the number of new COVID-19 cases and vaccination rates. Densely populated counties typically have high rates of infection, thus, a high number of new COVID-19 cases leading up to the July 4th weekend travel window. Likewise, densely populated counties generally have a high percentage of people vaccinated.

## 5.2 Causal Pathways from Enhanced Model (Model Three)

For clarity of presentation, we only illustrate the casual pathways from Model Two in Figure 6. However, there are other variables that impact the number of trips made during the July 4th weekend travel window. Specifically, we reason the casual pathways for the variables tht are included in Model Three.

**Restrictions that are active during the travel window**. If businesses are closed during the travel window, there are few options for people who travel to plan for food, recreation and other fun activities. As a result, people in places that have restrictions in place tend to travel to places that have a more relaxed set of restrictions. Thus, restrictions that are in effect during the travel window (especially, closure of restaurants, bars, movie theaters) have a positive impact on the number of trips.

**Conservative Political Ideology**. In addition, as discussed in Section 1, conservative political ideology is generally dismissive of the idea of social restrictions and skeptical about the guidelines proposed by the local state and county officials. As a result, we see an increase in the number of trips by people who are supportive of conservative political ideology.

**Employment rate**. Lastly, people who are employed are generally willing to make vacation trips. This may be an economical indicator suggesting affordability and willingness to take trips is a sign of confidence in

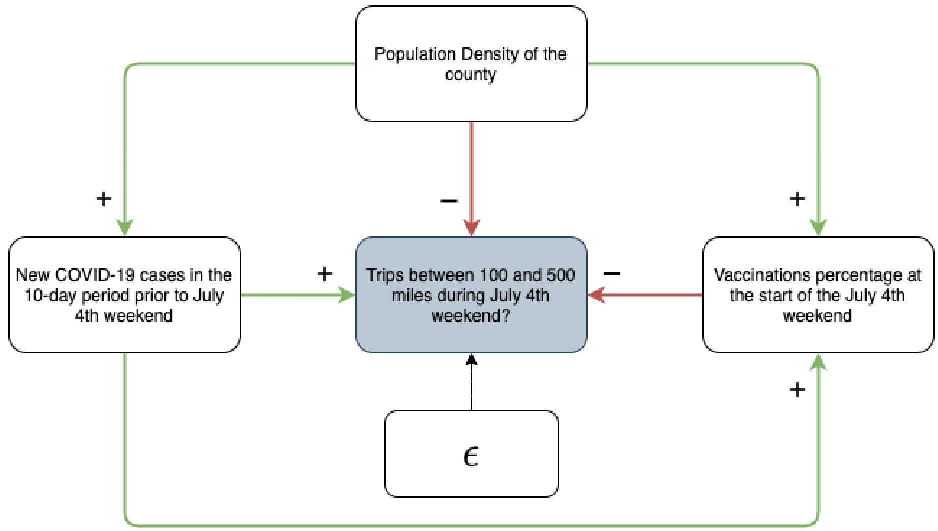


Figure 6: Potential causal pathways for number of trips during the July 4th weekend related to COVID-19 attributes

their ability to navigate the uncertain times surrounding the pandemic.

**Background Variables ($\epsilon$).** The background variables ($\epsilon$) are the variables that we did not explore or measure in our research that could affect the number of trips made during the July 4th weekend travel window. We assume that these variables are independent of the variables discussed above.

## 5.3 Omitted Variables

In exploring the impact of vaccination rates on mid-ranged travel over the July 4th weekend in the Pacific Northwest, we have highlighted several important covariates and the complex relationships among them. Other variables exist which impact people's decisions to travel, and are likely worthy of subsequent studies to establish causal relationships which could be leveraged to promote local tourism. Potentially impactful variables we identified and were unable to operationalize effectively are *COVID-19 fatigue* and *proximity to fun*. We can use our existing control for population density as a reasonable proxy for *proximity to fun* as a simplification. A larger population density is estimated to have more fun places nearby through economy.

Looking at Model Two, we can estimate the bias impact for our covariates and these omitted variables. Impacting all our estimates is a strong positive relation between covid fatigue and mobility; the more restless people are the more they are likely to travel despite the risks.

COVID-19 fatigue is a general term used to describe the individual and group restlessness resulting from extended social restrictions, an indefinite future end for the pandemic, and a drive for the return of pre-COVID-19 normalcy. This is a variable which is too nebulous to operationalize effectively but is potentially a significant omitted variable from our model.

Population density has a negative impact on mobility shown in our model coefficient. Large population centers have had stronger-for-longer restrictions and we estimate a strong positive relationship here with COVID-19 fatigue. We expect the compounding positive effect of COVID-19 fatigue to be away from zero even with an estimated negative coefficient for population density.

Vaccinations also have a negative impact on travel over the July 4th weekend. Estimating the relationship between vaccinations and COVID-19 fatigue is less straightforward. One would suppose as vaccination rate increases that the end of the pandemic becomes more attainable and reduces fatigue. However, more fatigue may drive people to get vaccinated to be done with the pandemic sooner. Optimistically, the hope generated from higher vaccination rates is the stronger driver and the omitted variable bias on vaccinations is towards 0.

The increase in COVID-19 cases leading up to the July 4th weekend have a weak, positive influence on travel in the Northwest. More people ignoring mask policies, gathering socially, and visiting businesses more regularly to relieve COVID-19 fatigue will generally increase the number of COVID-19 cases. Similar to population density, the omitted variable bias effect will be away from 0 on the COVID-19 case variable.

# 6 Conclusion

Perhaps the most important insight from this analysis is that vaccination rate is inversely correlated with road vacation trips in all three models. We suspect this is mainly because people who sought out the vaccine early are also people who are more cautious of traveling long distances during the pandemic. Vaccinated people also tend to be located in more densely populated areas, which could cause some confounding. However, note that case rate was the only additional variable which could be added to Model Two before vaccines became insignificant, so the coefficient on vaccines should be trusted cautiously. An additional limitation of Model Two is the questionable homoskedasticity discussed in Section 4.4.

Another important insight is the high coefficient on employment rate in the model. A 1% increase in employment rate equates to a 11% increase in the number of people taking to the roads. When considered with the high negative correlation between employment and business restrictions, this may be an indicator that the key to increasing road tourism is removing restrictions and increasing employment levels.

Lastly, it is important to note that case rate is not a strong determinant of road travel. A 1% change in case rate equates to only a 0.07% percent change in trips. Perhaps the pandemic itself is not a large driver of whether or not people take road trips.

Overall, the practical significance of these models is relatively low, as the adjusted $R^2$ indicates that only a small amount of the variation in the outcome is explained by any of the models of road travel at the distances we have investigated.

**Limitations and future work**. In our study, we excluded air travel. Typically, trips between 100 and 500 miles during a national holiday tend to be road trips. Trips that are greater than 500 miles are generally inter-state trips and primarily involves flight journeys. Future work may investigate whether vaccinated people are taking to the skies more than the roads. Another finding from our study is that population density plays a (statistically) significant role in determining the number of road trips taken. Future work may investigate if vaccinations and/or other COVID-19 related attributes play any role in vacation travel by using difference panel approach (which is beyond the scope of W203 and this study).

# References


[1] U.S. Travel Association. *Travel: The Hardest-Hit U.S. Industry.* https://www.ustravel.org/system/files/media_root/document/Research_Fact-Sheet_Industry-Table.pdf. 2021.

[2] Alison Sider. *Air Travel is Back, Including All the Things You Hated.* https://www.wsj.com/articles/air-travel-airport-crowds-back-after-covid-11620828493. 2021.

[3] Jeffrey Gettleman, Shalini Venugopal, and Apoorva Mandavilli. *In U.S., Virus Cases Have Recently Dropped in Over Half of States.* https://www.nytimes.com/live/2021/04/28/world/covid-vaccine-coronavirus-cases. 2021.

[4] Trading Economics. *United States Unemployment rate : 1948 - 2021:* https://tradingeconomics.com/united-states/unemployment-rate. 2021.

[5] Anton Gollwitzer et al. *Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic.* https://www.nature.com/articles/s41562-020-00977-7. 2020.

[6] Bureau of Transportation Statistics (BTS). *Trips by Distance.* https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv. 2021.

[7] Centers for Disease Control and Prevention (CDC). *COVID-19 Vaccinations in the United States, County.* https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh. 2021.

[8] New York Times. *New York Times Covid-19 Data by County. Retrieved July 31, 2021*. https://github.com/nytimes/covid-19-data. 2021.
[9] US Census Bureau. *National Counties Gazetter File 2020*. https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html. 2020.
[10] MIT Election Data and Science Lab. *County Presidential Election Returns 2000-2020*. https://doi.org/10.7910/DVN/VOQCHQ. Version V9. 2021.
[11] Boston University School of Public Health. *COVID-19 US State Policy (CUSP) Database*. https://www.tinyurl.com/statepolicies. 2021.
[12] Bureau of Labor Statistics. *US Employment Statistics by County, May 2020 – June 2021*. https://www.bls.gov/lau/. 2021.
[13] J. Fox and G. Monette. "Generalized collinearity diagnostics". In: *JASA* 87 (1992), pp. 178–183.
[14] T. S. Breusch and A. R. Pagan. "A simple test for heteroscedasticity and random coefficient variation". In: *Econometrics* 47 (1979), pp. 1287–1294.