

# The Future of Park Visitation in the U.S.

W203: Lab 2

Jonathan Whiteley, Max Hoff, Connor McCormick

## Contents

<b>Introduction</b>	<b>2</b>
<b>Research Question and Theoretical Model</b>	<b>2</b>
<b>Data and Variables of Interest</b>	<b>2</b>
Datasets . . . . .	3
Variables . . . . .	4
EDA . . . . .	4
<b>Model Building Process</b>	<b>7</b>
<b>Model Limitations</b>	<b>10</b>
<b>Omitted Variables</b>	<b>13</b>
<b>Results</b>	<b>13</b>
<b>Conclusion</b>	<b>14</b>
<b>References</b>	<b>14</b>

## Introduction

The Covid-19 pandemic has upended lifestyles to a degree that most people could not have imagined a decade ago. The local community impacts of the pandemic have ranged from uncertain financial situations and limited recreational options, all the way to widespread anxiety and loss of loved ones. One of the more interesting changes to lifestyles has been shifts in how many citizens visit public parks, as local travel and indoor recreational activities have been curbed by stay-at-home orders.

Given they are free to use while providing a gathering place for a wide range of diverse groups, parks serve a valuable function to society. The importance of parks on a community's physical and mental wellbeing is well documented (Rice 2020, Samuelson 2020), but became especially relevant to consider as the pandemic spread. As developed world vaccinations increase, this raises an interesting question around the future of public spaces and more specifically public parks. Which lifestyle changes will become a permanent fixture of American life? Will increased public park usage go the way of cloud computing and remote work remaining here to stay? The answers to these questions have important implications for city planners (Honey-Roses 2020), local government officials, and Department of Tourism officials across the country.

## Research Question and Theoretical Model

Our main goal for our research project is to explore whether a causal relationship exists between visits to public parks and Covid-19 vaccination rates in the United States. Specifically, how do Covid-19 vaccination rates within a county impact visits to parks within that county?

Finding the answer to this question will bring important insight into how the Department of Tourism should allocate resources towards parks, especially in counties with limited park space. Additionally, it will allow county officials to better understand park use as a passing trend or a mainstay of society after vaccinated citizens have the opportunity to return to indoor activities. If our evidence shows that there is not a meaningful relationship between park visits and Covid-19 vaccination rates within a county, our default action/recommendation is one of inaction.

If our evidence does indeed indicate a meaningful relationship between park visits and Covid-19 vaccination rates within a county, the Department of Tourism and policymakers in states with lower vaccination rates should reallocate resources to account for this. We predict that counties with higher vaccination rates will show lower park usage.

Through our explanatory modeling process, we plan to leverage the data sources below to operationalize and test parameters of the following causal theory: counties with high vaccination rates are returning to pre-pandemic activities faster and will not continue use of public parks as seen during the pandemic.

The relationships explained through our analysis are related by the structural equation model (Figure 1).

## Data and Variables of Interest

To begin our modeling process, we needed to define the exact variable that we wanted to measure. Since we were researching how the rates of vaccinations affect park visits within a county, it made sense that our outcome variable was change in park visits from a baseline date (pre-Covid-19) to current date (more details to follow).

Our next step was to think of all the possible variables that could influence park visits and then explore data that was readily available in a comparable format to our output variable. From a conceptual level, there were all sorts of variables that we thought of that could influence park visits, but as we learned, not much data was available. Most of our ideation around influential variables were related to quantifying a propensity to visit a park. For example, the amount of parks available in a given region, the quality of these parks, the amount of lifestyle alternatives compared to parks (e.g., dining availability, amusement parks, etc.). We would then be able to control for these variables, leaving us with the true effect that vaccinations had on

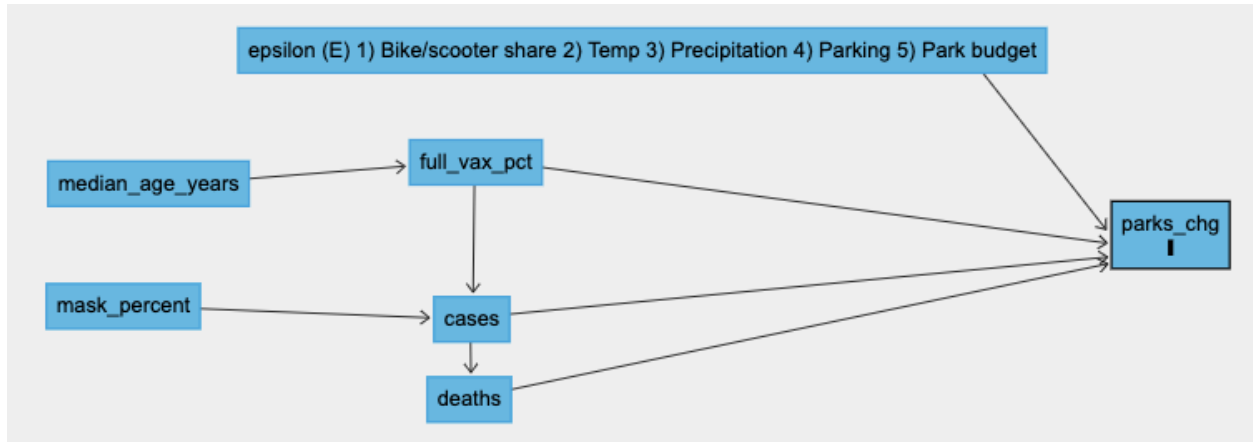


Figure 1: Structural Equation Model

park visitation. Ideation for the amount of variables to control for seemed endless, but unfortunately, we were able to find very little data that was actually of use.

After much searching, we were able to obtain a couple datasets that we hypothesized to have an influence on park visitation. The first dataset was obtaining age data by county from the U.S. census. Our hypothesis was that age might be a useful predictor in determining propensity to visit a park. Without being able to find other measures that indicate park visitation propensity, we thought age might be a useful substitute for at least a portion of the variables that make up propensity. Our reasoning behind this was that the average age of a population might be influential in park visitation, as younger people might be more active, and thus more likely to visit a park.

Another distinct influential factor that we thought might influence visitation was attitudes towards Covid-19. Across the U.S., attitudes and policies have varied widely that most likely had a significant influence on mobility of a population. We searched for several datasets that might quantify these attitudes in some way, and we eventually stumbled upon a New York Times dataset that measured attitudes towards wearing masks. We believed that this might be a good proxy for general attitudes towards Covid-19. We thought attitudes towards Covid-19 might be one of the most influential factors in park visitation, and being able to control for this in some way would have a significant positive effect in being able to interpret our results.

As for predictor and target variables, the datasets we used were a mobility dataset from Google, which gave us a percentage change in visitation to parks by U.S. county, and a vaccination dataset from the CDC, which gave us vaccination rates for all U.S. counties.

We also stumbled upon an additional dataset from the New York Times which gave us case counts and death counts by U.S. county, which we thought could potentially be useful.

Additional detail on each of these datasets is below:

## Datasets

- **COVID-19 Community Mobility Report:** This data that was collected by Google, describes the percent increase/decrease of visits to specific categories of locations such as retail centers, transit stations, grocery store, etc. In our case, we will be looking at the percent trends over time of visits to parks. The data coverage begins from a baseline measurement taken on January 1st, 2021 and is updated routinely. Since we will not be performing a time-series analysis, we decided to select a slice of the data from June 30th, 2021. The variables that we are concerned with in this data set are the FIPS code (used for merging data by county) and the park visits change from baseline.
- **CDC Data on Vaccinations:** This data set was provided by the CDC and contained much information on vaccination rates at a county level. It contains both hard vaccination rates and percentages.

Additionally, it provides these vaccination numbers drilled down to age subsets as well as at a cumulative level. Lastly, this data contains a FIPS code for data merging. The variables of concern for this data set are the FIPS code, the hard cumulative vaccination rate, and the percentage vaccination rate. We decided not to use any of the age subset rates since we hypothesize that all age demographics are reasonably likely to visit a park.

- **New York Times Covid-19 Data:** This data set is similar to the vaccination data provided by the CDC but instead reports on actual Covid-19 cases at a county level. In addition to case data, this data set also contains reports on deaths from Covid-19 as well. This data set is structured in a time series format, similar to the Covid-19 Community Mobility Report. So in order to choose a proper cross-section for analysis, we selected cases from 2-weeks before our selected date for the park visits percent change (June 30th, 2021) up until 2-weeks after, based on the CDC's recommendation of a two week life-cycle for virus contraction to full recovery. The variables that we are concerned with in this data set are the FIPS code, date (for cross-section selection), county, case counts, and death counts.
- **Mask-Wearing Survey Data:** In the same repository as the NYT Covid-19 data set, we found another data set that came from a large number of interviews conducted online by a survey firm named Dynata at the request of The New York Times. They surveyed people at a county level about their mask use between July 2nd and July 14th. Since this survey collection period is so close to the covid case date cross section we selected and the date selected for park visit percent change, we're assuming that the respondent's mask wearing habits are not going to change that much so we can use this as one of our explanatory variables. The variables of concern for this data set are FIPS code and mask use.
- **Census Demographic Data:** Lastly, we utilized the United States Census Bureau's Population Estimates Program data releases to retrieve a data set that describes the population size and median age of an area at the county level. The variables of concern in this data set are the FIPS code, the population size, and the median age.

## Variables

To begin our modeling process, we needed to define the exact variables that we wanted to measure. The table below explicitly states the ones we'll reference throughout this report:

Covariate	Variable Name
Number of citizens vaccinated in a particular county	full_vax_count
Percent of citizens vaccinated in a particular county	full_vax_pct
Log transformed count of cases in a particular county	log_10_cases
Log transformed count of deaths in a particular county	log_10_deaths
Percentage of the population that reported wearing masks frequently	mask_percent
Median age of the population of a particular county	median_age_years

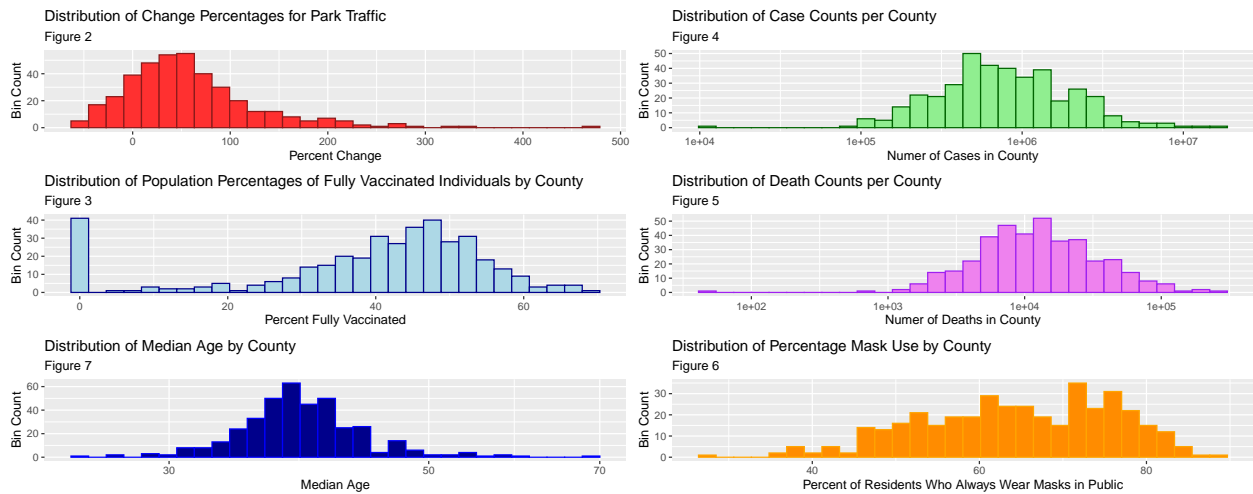
## EDA

Now that we had our data collected and our hypothesis and research goal laid out, it was time to explore the data. A good first step in any Exploratory Data Analysis (EDA) is conducting data quality checks.

### Data Quality Checks

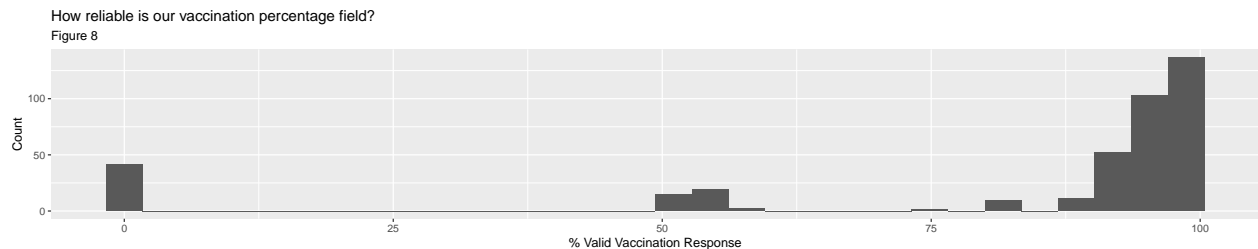
We examined each of the variables to dig up any potential errors or problems with the dataset, and determine whether any adjustments needed to be made. We started our investigation by looking at the distributions for each of the variables:

## Distributions



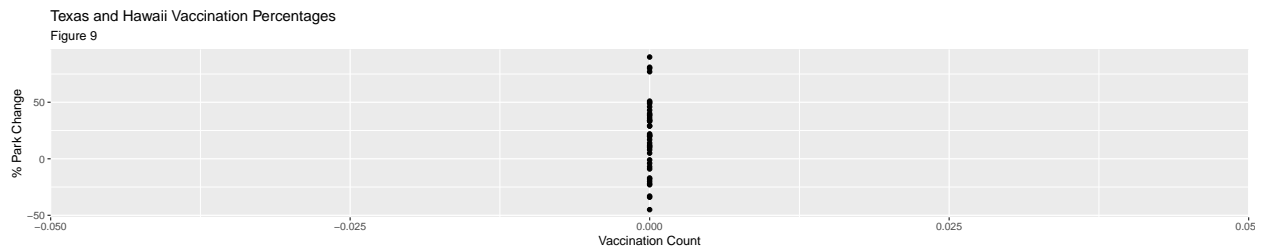
At first glance, in Figures 2 through 7, it seems like we may need to conduct some log transformations to several of the variables, given the long tailed distributions.

**CDC Dataset Investigation** While examining the CDC vaccination dataset we stumbled upon the `Completeness_pct` variable, which states the proportion of vaccination people that listed a valid FIPS code (county code).



We were hoping that almost all counties would have a value of close to 100 percent, but as you can see in Figure 8, a material proportion of the counties have values that might impact our results.

Let's dig further into this. It looks like the counties with 100 percent invalid FIPS codes are coming from two states; Hawaii and Texas. Let's see how these counties are distributed in terms of vaccination percentages as a whole:

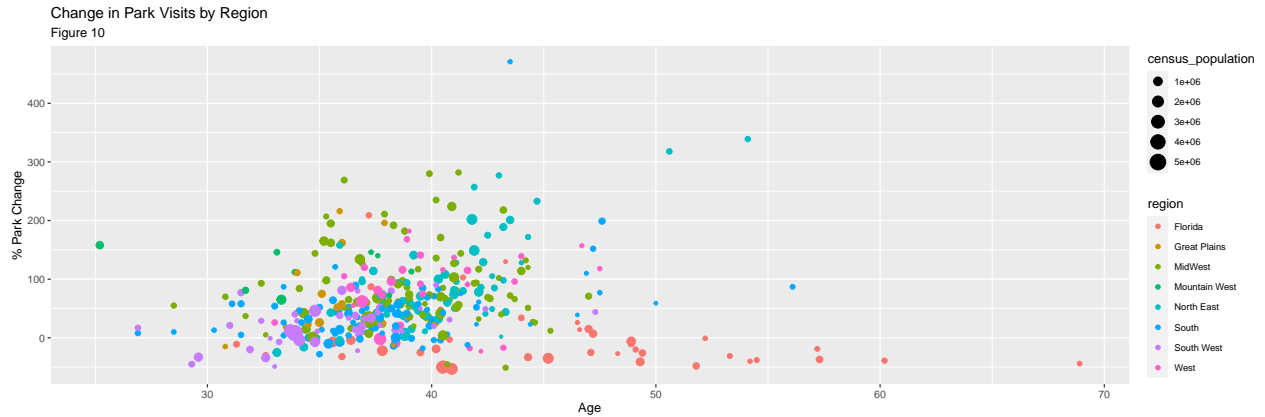


As you can see in Figure 9, all counties within these two states have invalid results, thus we should exclude these in our analyses going forward.

**Region Analysis** Our next set of investigations was related to how each of these variables change by region, since we thought some of the variances for each of our variables might be clustered well by region. This could potentially invalidate our findings, since counties might be not well distributed. If certain regions

contain more counties than others, then our data might not meet the Independent and Identically Distributed (IID) assumption of the Classic Linear Model (CLM).

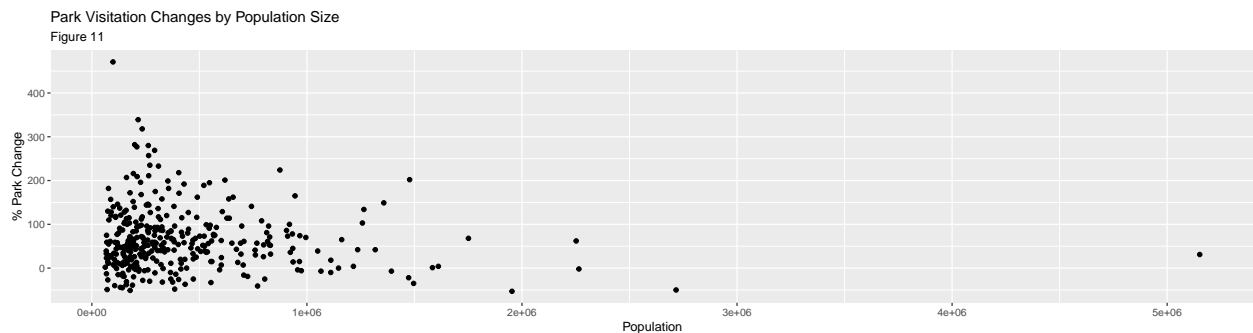
Figure 10 below shows how park visits relates to the median age of a county, broken down by region (shown by color). We would not have to worry about IID if there is no discernible difference between the different regions. Also, we were hoping to find not much difference in county size, as shown by the size of each of the points on the scatter.



As you can see in Figure 10, unfortunately there is a discernible difference between the regions of the country, as we suspected in our original hypothesis. We'll have to revisit this when we build our model. As for county size, it looks like county size is fairly well distributed, so this shouldn't be a problem.

**County size investigation** Another topic that we were interested in was whether county size could have an impact on changes in park visits. Theoretically, smaller counties could have larger swings in park visitation, since they will have smaller bases to work off of. For example, if a county has only two residents, if one person goes to a park who never did before, then park visitation will have a 50% increase. Compare that to a county with 10 residents, if 1 person goes, then then will only see a 10% increase.

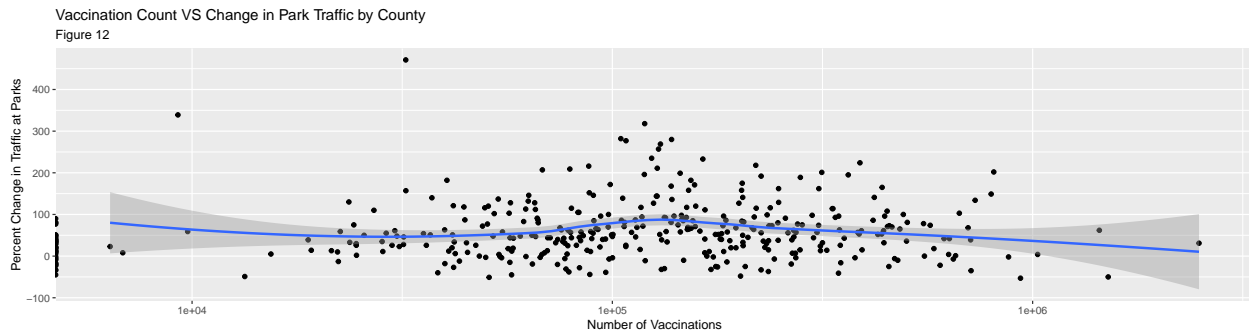
Figure 11 plots how park visitation is affected by population size. We were hoping park visitation would not be affected by population size.



As you can see from Figure 11, unfortunately it does look like smaller counties can have larger swings in park visitation. This is one more item we'll have to consider when building our model and interpreting results.

## Variable Correlations

Now that we've pressure tested our dataset, let's take an initial look at how strong the correlations are:



At first glance, Figure 12 shows that the correlation does not look very strong. We have not applied any transformations yet though, so it's possible that the data just needs some cleaning.

Diving more deeply into the specific variables, Figure 13 below the correlation between each of them:

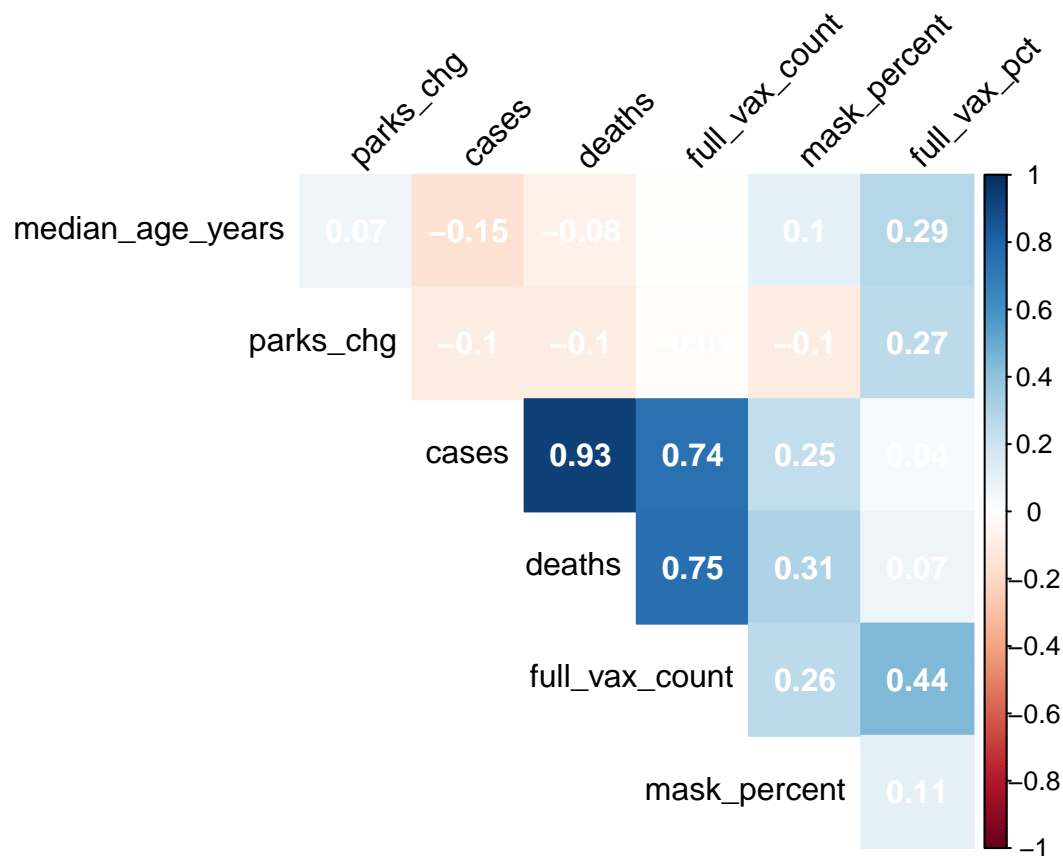


Figure 13

Unsurprisingly, there are very strong correlations between cases, deaths, and vaccination rates. The correlation between deaths and cases might even be strong enough to bring about problems of perfect collinearity (or near perfect collinearity), which we'll discuss in more detail later. As for the other variables, correlations were high enough to make us believe that a strong model might be built.

## Model Building Process

To begin our modeling process, we needed to define the exact variable that we wanted to measure. Since we are researching how the rates of vaccinations effect park visits within a county, it made sense that our outcome variable would be the percent change in park visits from the Covid-19 Community Mobility Report

data set from Google. Note that the parks category includes regions such as national parks, public beaches, marinas, dog parks, plazas, and public gardens. Given the flexibility in the definition of a park for this analysis, we want to make clear the range of the outcome variable that we are analyzing.

To measure the percent change in park visits, we hypothesized that the main explanatory variable would be the rate of vaccinations within a county. In a way, the vaccination rate of a particular region somewhat symbolizes a transition state out of pandemic operations for that region. We are unsure how this will affect the number of visitations to a particular park or set of parks however. On the one hand we have parks becoming a replacement for other locations that were closed during the pandemic, so this may indicate that park traffic will decrease as vaccination rates increase and people become more comfortable being in close quarters again. On the other hand, we have the possibility that people were nervous about going to parks during the pandemic, so in this case, we may expect to see an increase in park traffic as regions transition out of pandemic tendencies. In addition to the vaccination rate of a region, we also have identified the number of cases in a region as an explanatory variable. One thing we've noted for this additional explanatory variable, however, is that there is likely a strong level of collinearity between the number of cases and vaccination rate/count. We will discuss this further in the model limitations section. In addition to these two main explanatory variables, we have identified a few other covariates we plan on including in the larger versions of the model; we will discuss these further below. The final conclusion we plan on reaching is one that explains the causal nature of vaccination rates, and case rates for that matter, on the percent change in visits to parks in an area.

Because of the nature of our variables and the causal effect that we are searching for, we have determined our model to be an explanatory model. We decided to start off with a base model or *limited model* that just regresses the change in park traffic on vaccination rate. Since this is the key variable which we are trying to determine a causal relationship for, we decided it best to use this as the only explanatory variable in our first model.

In our second model however, we decided to include the case count variable as a covariate. As mentioned before, we identified some collinearity between cases and vaccination rates in our EDA which may give rise to a little uncertainty in the results. That being said, we feel that including this covariate will still influence the outcome variable so it is important to include, as long as we are aware of the potential effects. Another alteration we made to the second model was adding a log transformation to the cases covariate. While performing the EDA, we noticed that the distribution of the cases variable had a pretty strong right skew, so we added a log transformation to bring the variable closer to a normal distribution.

Finally, for our third model, we included all relevant covariates that were available and that we believed would have an influence on the change in park visits from baseline. So in addition to vaccination percentage and case count variables, we also included vaccination count, mask-wearing percentages, and the median age in years per county. Furthermore, we also decided to include vaccination count in addition to vaccination percentage since we figured that vaccination count would carry some additional information because of the raw size of the vaccine count numbers speaking to the size of the county in question. Again, we needed to perform a log transformation on the covid cases variable in order to achieve linearity. It is important to point out that more so than model two, there will be some collinearity between some of the explanatory variables in the third model iteration. For starters, we still have the collinearity between vaccination percentage and case counts, but now we also have collinearity between vaccination count, mask percentage, and median age. Addressing the final variable of median age, it is possible that there is some collinearity between this variable and the other explanatory variables, however according to our correlation plot in the EDA section, it seems like there is only a small correlation between median age and vaccine count. At a high level, the variables we've included in the third model combine to more or less explain the severity of the pandemic in a particular region in addition to the response. And to wrap it all up, this explanation is the key that we are using to measure our outcome variable of change from baseline in visits to parks.

## F-Test

We run an F-test to get an indication if our additional features are improving performance of the model. The null hypothesis is that the short model is as good at explaining the variance in `parks_chg` as the two long



models. The rejection criteria used is  $p\text{-value} < 0.05$  for the F-Test. We conclude that we are able to reject the null hypothesis for Model 3 but not Model 2, as our test returned a p-value much lower than 0.05 for Model 3. As a result, we can conclude that the additional features in Model 3 are improving the performance of our model and that our model improved more than if all those indicators had been set to zero.

```
## Analysis of Variance Table
##
## Model 1: parks_chg ~ full_vax_pct
## Model 2: parks_chg ~ full_vax_pct + log10(cases)
## Model 3: parks_chg ~ full_vax_pct + full_vax_count + log10(cases) + mask_percent +
##           median_age_years
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      388 1720195
## 2      387 1711902   1      8293 1.9208 0.16657
## 3      384 1657841   3     54061 4.1740 0.00631 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 13

### Regression Table

```
##
## Table 1: The relationship between vaccination rates and park traffic
## =====
##                               Dependent variable:
## -----
##                               parks_chg
##                               (1)      (2)      (3)
## -----
## full_vax_pct                1.095***    1.138***    1.439***
##                               (0.175)    (0.175)    (0.274)
##
## full_vax_count                                -0.0001
##                                              (0.00003)
##
## log10(cases)                                -11.516
##                                              (7.742)
##
## mask_percent                                -0.647
##                                              (0.335)
##
## median_age_years                                -0.136
##                                              (1.224)
##
## Constant                17.176*           83.340           1.591
##                               (6.970)    (46.291)    (86.494)
## -----
## Observations                390           390           390
## R2                0.073           0.077           0.106
## Adjusted R2        0.070           0.072           0.095
## Residual Std. Error   66.584 (df = 388)   66.510 (df = 387)   65.706 (df = 384)
## F Statistic          30.459*** (df = 1; 388) 16.201*** (df = 2; 387) 9.144*** (df = 5; 384)
## =====
```

## Note:

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

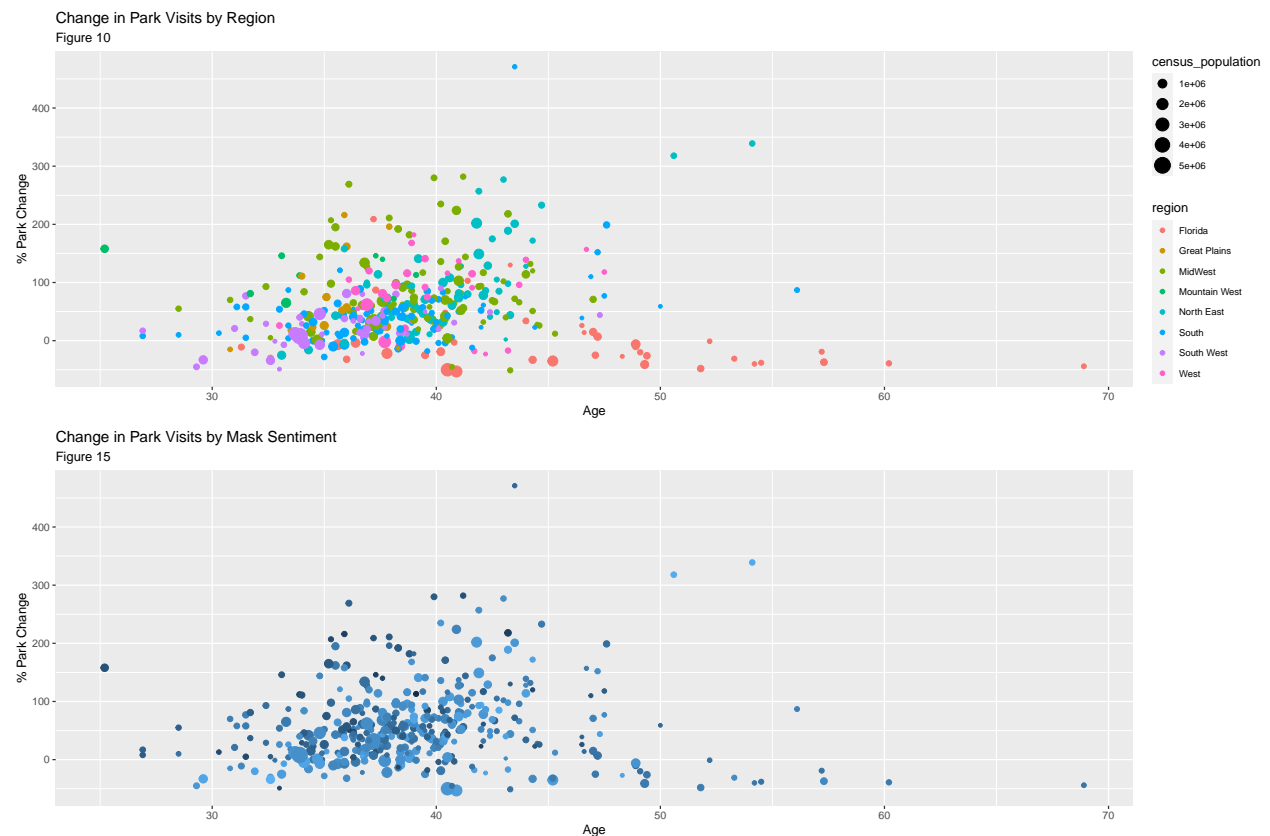
Figure 14

## Model Limitations

Our limitations around finding data set some tight constraints in how we built our model. Those constraints led us to make trade-offs which added limitations to our model. Upon Assessing all five assumptions of the CLM, we found evidence that our model does not fit all the necessary assumptions for the CLM framework - specifically the assumptions for linear conditional expectation, and homoskedastic errors. There was also debatability for whether we met the IID assumption, which we'll discuss below. That being said, our sample size is rather large at 390 counties, so with the CLT holding due to our large sample size, we are able to have some flexibility with the CLM assumptions.

### IID Sampling

As discussed in the EDA section, there are some clusters when you organize the data by region, which might lead one to believe that IID is violated. This on it's own might be enough to invalidate our model, but we believe the addition of the mask wearing variable is enough to counteract this problem.

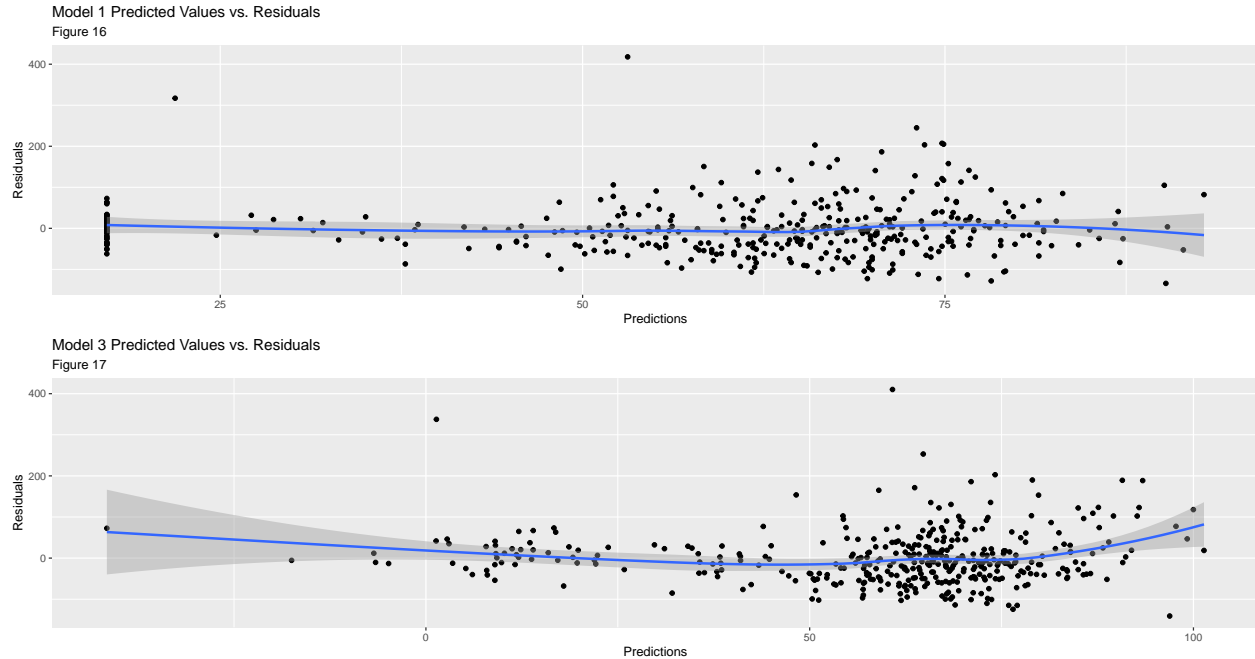


Shown in Figures 10 is the same distribution of park visits by age as shown in the EDA section, but shown in Figure 15 is the chart replicated but broken out by mask sentiment. As you can see, the mask sentiment shows a rough similarity in clustering compared to region. Although there is some subjectivity in our conclusion, we believe that the mask sentiment variable will control enough for this clustering to satisfy most of the IID assumption.

There are other potential problems with the data gathering process that could lead to IID issues, for example using only data from Google users which could exclude some older populations, but we believe these problems to be minimal, and most likely will not influence our results much.

## Linear Conditional Expectation

To assess whether there is a linear conditional expectation, we've looked at the predicted vs. residuals of both Model 1 and Model 3. The plot of both, shown below in Figures 16 and 17, indicates a pretty clear non-linear relationship. This is concerning and does not give us confidence that there is a linear conditional expectation here.



## No Perfect Collinearity

When looking for perfect collinearity, or the absence thereof. It is useful to look at the correlation plot above. The results of the plot show a very strong level of collinearity between cases and deaths which makes sense since more cases generally speaking will lead to more people dying. We comment more about this in the results section of the report, but in short, we decided to take out the deaths variable from our model because we believed this was an instance of near perfect collinearity. There is also the collinearity between cases and the vaccine count variables to consider. The correlation reported in the plot between these two variables is only 0.74 which is not as bad as the 0.93 between cases and deaths, but is still pretty high. We decided to keep these variables both in the model because we believe that there is still some explanatory power in the cases variable that does not exist in the vaccine count variable and vice versa, but we wanted to point out that there could be some resulting uncertainty from leaving these variables both in the model.

## Homoskedastic Errors

To assess whether the distribution of the errors is homoskedastic or heteroskedastic, we can reference Figures 16 and 17 again. We see that the band increases in thickness as it goes from left to right for `full_vax_count`, getting quite wide on the upper side of the predicted values, which is slightly concerning. For `cases` we see a wide band on the left that tapers off and gets wide out to the right.

We can also leverage a scale-location plot, shown below in Figure 18. Homoskedasticity would show up on this plot as a flat smoothing curve, which does not quite look like it is the case here, perhaps suggesting problems with homoskedasticity.

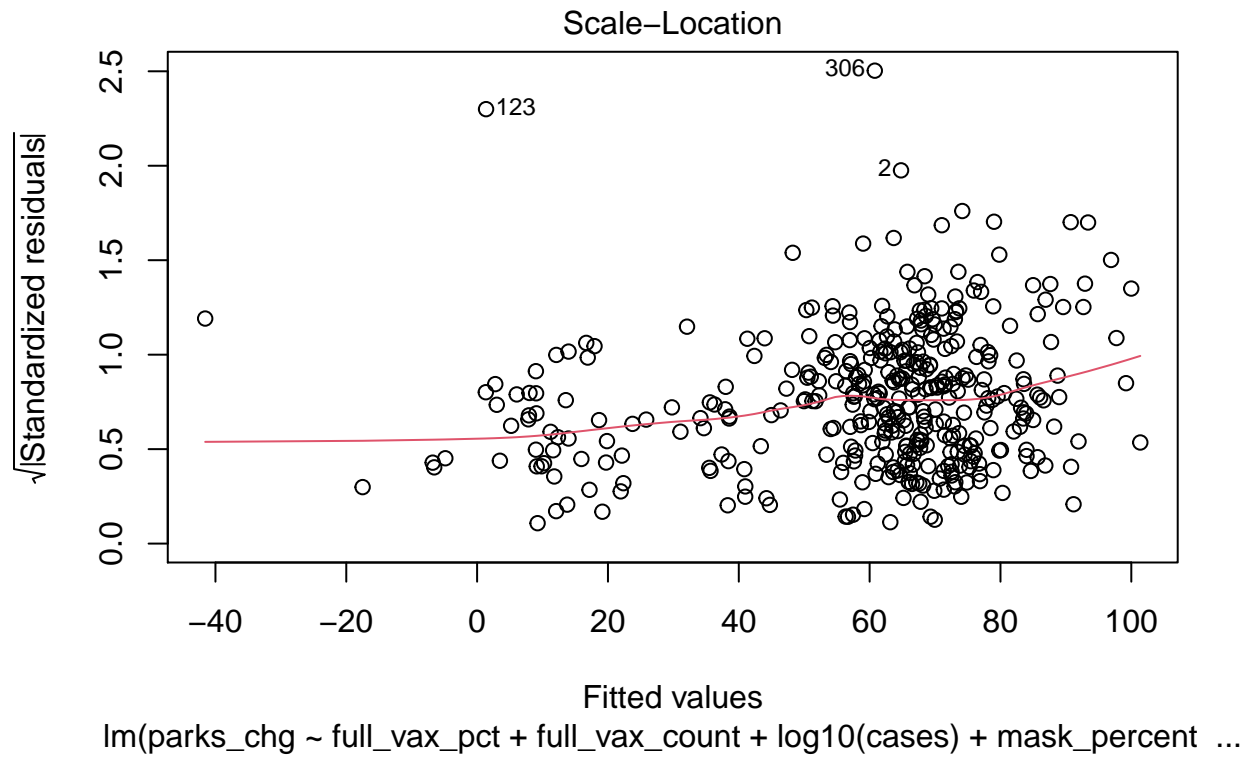
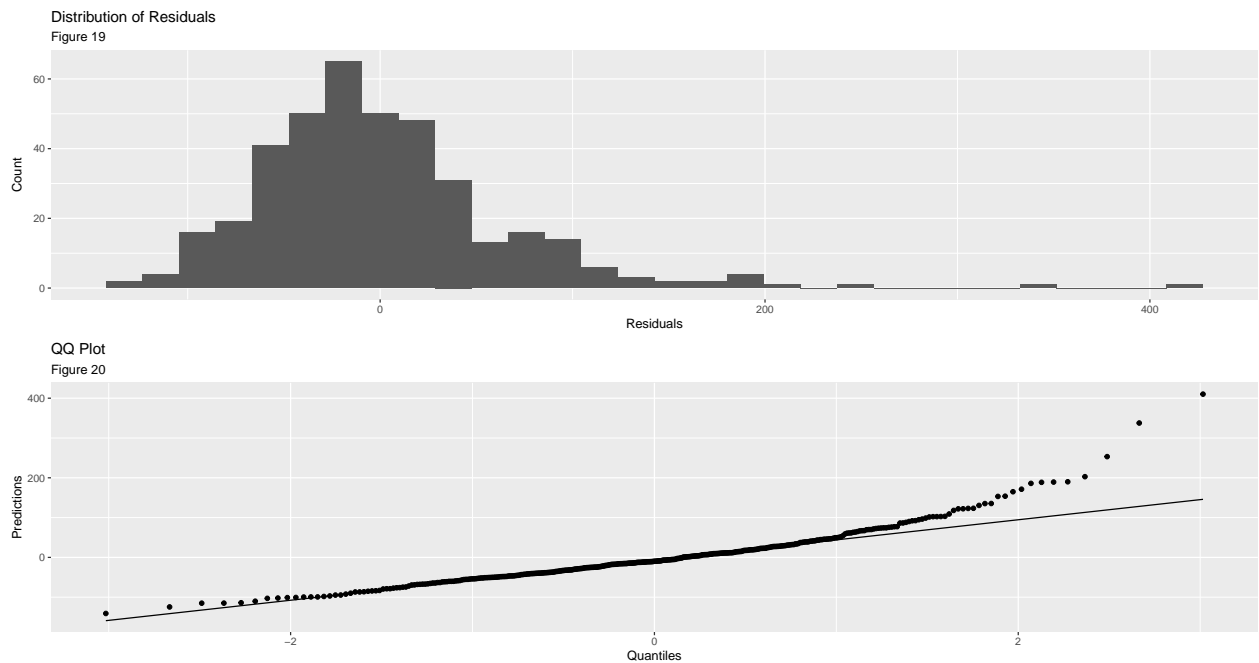


Figure 18

### Normally Distributed Errors

We are looking for evidence of normally distributed errors by examining the histogram of residuals and a qqplot, shown below in Figures 19 and 20. The normal shape of the histogram, yet slightly skewed right and the qqplot distribution shows the points from the quantiles forming a line that is roughly straight except off to the right side. This give us some confidence in stating that the errors are close to normally distributed.



## Omitted Variables

We identified several omitted variables which could cause omitted variable bias in our model. Park traffic is tied to the transportation option proximity and availability. As such, we aimed to operationalize this concept through two variables: changes in proximity to bike/scooter sharing stations and vehicle parking spot availability. Related to bike sharing, findings suggest moderate-frequency riders in San Antonio county increased bike share usage 12% during the pandemic (Jobe, Jeffrey, and Griffin 2021). This gives us reason to believe that counties expanding bike & scooter sharing station footprints during the pandemic would see increased use. These variables were omitted due to a lack of comprehensive county-level data set from which could operationalize the concepts.

Another variable driving park traffic is park maintenance/cleanliness. This concept would be difficult to measure and a proxy for it could be a county level public park budget allocated to maintenance. Total investment in parks in dollars, square acres dedicated to parks, and percent of residents within a 10 minute walk of a park would also have been relevant variables, however the dataset containing these variables (The Trust for Public Land) was at the city level and only included the 100 largest US cities. The expected relationship between the above variables with our output variable, `parks_chg`, and primary input variable, `full_vax_pct` are both positive. Given these observations, the direction of the bias will be positive. Since, the coefficient associated with the included variable `full_vax_pct` is positive, adding in the omitted variable would push the coefficient in the negative direction, towards zero in this case.

Weather was also an important topic we discussed regarding omitted variables which may impact park usage in a given county. The difference between baseline and the average temperature and average precipitation are two additional variables which may impact the change in park traffic. If the baseline weather and precipitation in a given county was especially high or low given the average for the season, this could give an unrepresentative change in the park traffic from baseline to June 30, 2021. Having the average temperature and precipitation for each county would have allowed additional explanation within the model, but the data itself was difficult to collect and standardize for our purposes. We also discussed a potential reverse causal pathway between park traffic and vaccination rate, however determined that it was not a clear two way relationship. We proceeded with our model construction leveraging the causal theory outlined initially.

## Results

When considering the results of our modeling efforts, we notice that across all three models, the percentage of fully vaccinated individuals in a particular county was a variable that had a highly statistically significant effect in all three specifications of the model with  $p\text{-value} < 0.001$ . Because of these results for this variable, we concluded that this variable had a very strong causal effect that was shown in all three models. When looking into the practical significance for model 1, we were able to conclude that for every 1% increase in percent of the county population fully vaccinated, there was a corresponding 1.095% increase in park visits. For models two and three, there was a 1.138% and 1.439% increase respectively.

Next, we will discuss the results of the number of cases per county under the log transformation. At first, we were including both  $\log_{10}(\text{cases})$  and  $\log_{10}(\text{deaths})$  in our third model, and it was interesting to see that the results of our coefficient were very divergent in model three with a coefficient of about 44.0 for  $\log_{10}(\text{cases})$  and about 49.0 for  $\log_{10}(\text{deaths})$ . It was unclear why these variables had such different effects. That's when we realized that we had an instance of near perfect collinearity between the  $\log_{10}(\text{cases})$  and  $\log_{10}(\text{deaths})$  variables. When looking at the practical significance for the third model, we had about a 44% increase for an increase of 1 in  $\log_{10}(\text{cases})$  and about a 49% decrease for an increase of 1 in  $\log_{10}(\text{deaths})$ . This was much larger than any other coefficient which was odd, but additionally, this would translate to an unreal practical significance - if there were 10 cases in a county ( $\log_{10}(10) = 1$ ) then there would be about a 44% increase in park traffic. This was frankly unbelievable. So in response to this realization, we decided to remove  $\log_{10}(\text{deaths})$  from our third model. This left us with a normal looking  $\log_{10}(\text{cases})$  coefficient with a believable practical significance. When assessing this practical significance, we determined that any increase of 1 on the log 10 scale could be interpreted as an 11.516% decrease in park visits in model two and a 10% increase in the third.

Coming to our last couple covariates, we will assess the percent of diligent mask wearers and median age of a county together. Both of these covariates were added solely in the last model. The coefficient for the `mask_percent` variable did not have a statistically significant effect in the regression. The `median_age_years` variable also did not have a statistically significant effect. When looking at the practicality of these variables, we can interpret them as follows: A 1% increase in mask wearers that report always wearing a mask would translate to a 0.647% decrease in park visits. A 1 year increase in the median age of a county could be interpreted as a 0.136% decrease in park visits.

Lastly, we noted that with the addition of more covariates with each model iteration, our R2 value was increasing which indicates more variance being explained by our model, however it is still quite low. We also noted that the intercept for the first model turned out to have a statistically significant effect in the regression, though none of the other models had statistically significant intercepts.

## Conclusion

Based on the results of our coefficient tests, we are able to conclude that the percentage of vaccinated individuals has a positive causal effect on the percentage change from baseline in park visits. Additionally, our F-Test gave us additional confidence that our third model was pretty effective at explaining the change in park traffic in response to our explanatory variables.

Although it is tempting to discard our initial hypothesis, and conclude that park visitation is here to stay, since vaccination rates had significant, positive coefficients across all three models, we can't be certain of this. For one, we don't know if vaccinations in the U.S. will continue to increase, or what the future of the pandemic will be like. It is very possible to have additional waves of the virus, and the pandemic become a persistent part of life. Assuming that the pandemic will eventually end and normal life will resume though, we still believe our results should be taken with a grain of salt. Although our sample size was large enough to overcome some of the deficiencies in our CLM assumptions, we believe our modeling may come short in other ways. Most importantly, we recognize that our analysis was a snapshot into the causal relationship between the aforementioned factors. More sampling over time and time series analysis into behavior changes would allow a more comprehensive understanding. Any effects that we observed may only apply to this particular moment in time. We don't know for sure if the relationship is linear, and if interests will change further once the pandemic has fully ended worldwide. Interests in parks may just be a temporary fad, since pandemic life is still a recent memory. As time passes and these memories fade, so may fondness towards parks.

As for additional research on the topic, it would be interesting to reconduct this study once the pandemic is fully over. Also, it would be interesting to examine the impact of rural vs urban populations and how each's access to public spaces was impacted. The ability to collect more detailed data might provide rich new ways to examine these affects, and the release of public datasets (similar to Google's mobility report) should continue to be encouraged.

Overall, the future of parks in the U.S. looks brighter than it did before the pandemic, and society as a whole will benefit.

## References

- "2010 Eval Estimates." Census.gov, [www2.census.gov/programs-surveys/popest/datasets/2010/2010-eval-estimates/](http://www2.census.gov/programs-surveys/popest/datasets/2010/2010-eval-estimates/).
- "COVID-19 Community Mobility Reports." Google, [www.google.com/covid19/mobility/](http://www.google.com/covid19/mobility/).
- "COVID-19 Vaccinations in the United States, County." Cdc.gov, Centers for Disease Control and Prevention, [data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh](https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh).
- "Guidance for Administrators in Parks and Recreational Facilities." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, [www.cdc.gov/coronavirus/2019-ncov/community/parks-rec/park-administrators.html](http://www.cdc.gov/coronavirus/2019-ncov/community/parks-rec/park-administrators.html).
- Honey-Rosés, Jordi, et al. "The Impact of Covid-19 on Public Space: An Early Review of the Emerging Questions – Design, Perceptions and Inequities." *Cities & Health*, 2020, pp. 1–17., doi:

10.1080/23748834.2020.1780074.

- Jobe, Jeffrey, and Greg P. Griffin. “Bike Share Responses to Covid-19.” *Transportation Research Interdisciplinary Perspectives*, vol. 10, 2021, p. 100353., doi:10.1016/j.trip.2021.100353.
- Korevaar, Hannah M., et al. “Quantifying the Impact of US STATE NON-PHARMACEUTICAL Interventions On COVID-19 Transmission.” *MedRxiv*, Cold Spring Harbor Laboratory Press, 1 Jan. 2020, [www.medrxiv.org/content/10.1101/2020.06.30.20142877v1](http://www.medrxiv.org/content/10.1101/2020.06.30.20142877v1).
- “Nytimes Covid-19-Data/Mask-Use.” [GitHub.com/Nytimes/Covid-19-Data](https://github.com/Nytimes/Covid-19-Data), [github.com/nytimes/covid-19-data/tree/master/mask-use](https://github.com/nytimes/covid-19-data/tree/master/mask-use).
- Rice, William L, et al. “Changes in Recreational Behaviors of Outdoor ENTHUSIASTS during the Covid-19 Pandemic: Analysis across Urban and Rural Communities.” *Journal of Urban Ecology*, vol. 6, no. 1, 2020, doi:10.1093/jue/juaa020.
- Samuelsson, Karl, et al. “Urban Nature as a Source of Resilience during Social Distancing amidst the Coronavirus Pandemic.” 2020, doi:10.31219/osf.io/3wx5a.
- Volenec, Zoe M., et al. “Public Parks and The Pandemic: HOW Park Usage Has Been Affected BY COVID-19 Policies.” *Public Library of Science ONE*, Public Library of Science, [journals.plos.org/plosone/article?id=10.1371/journal.pone.0251799](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0251799).