How has government spending on public infrastructure and transportation changed from pre pandemic (2019) to during the pandemic (2020)?

Aaron Lee

Maxwell Liu

Jihee Choo

Shani Lyubomirsky

In this project we explore different factors that affect transportation patterns, namely government spending on public infrastructure and transportation and number of employees. We first focus on the question: “How has government spending on public infrastructure and transportation changed from pre-pandemic (2019) to during the pandemic (2020)?” through means of multiple hypothesis testing to observe significant changes in ridership and spending on specific sectors. Exploring potential trends in government spending would bring insights into how the government reacted to the pandemic and the resulting changes in transportation demands in terms of budget planning and spending. The latter part focuses on answering the question: “How does the number of rail riders affect the number of rail employees?” using a logistic regression model and causal inference methods. We conclude with a discussion of the relationship between these different factors and public and private transportation ridership.

Data Overview

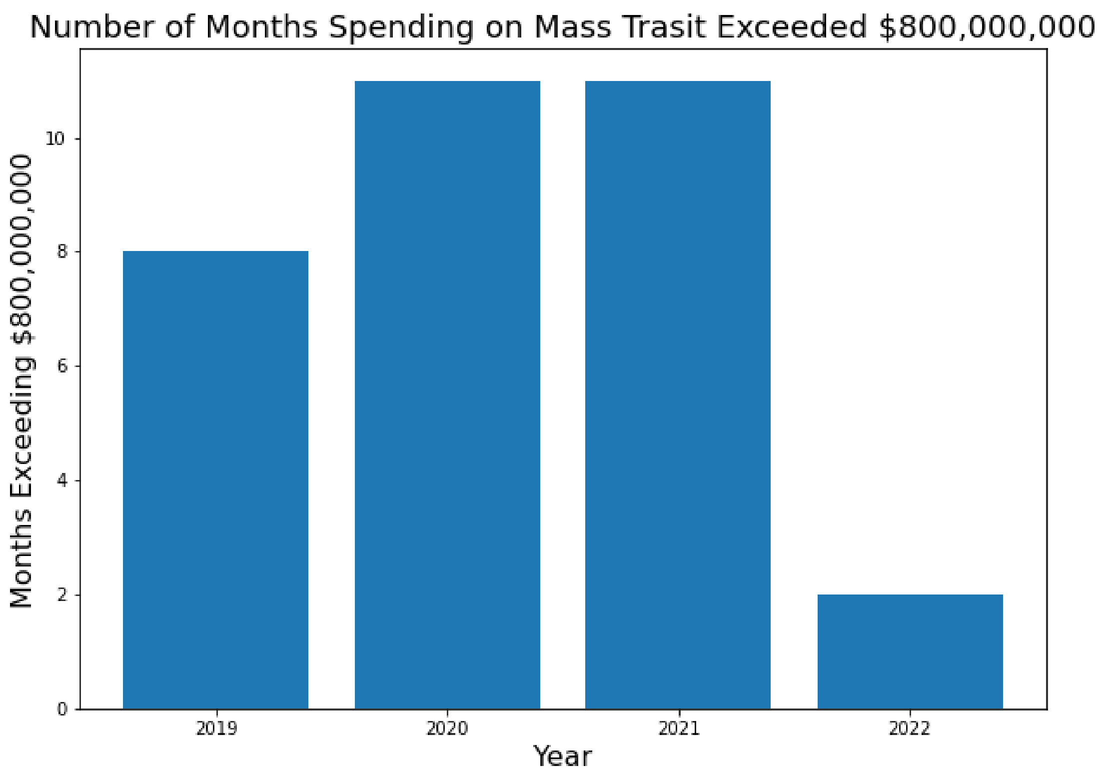
Data Description

Our primary source of data was the Monthly Transportation Statistics dataset from the Bureau of Transportation Statistics. The dataset is a census consisting of transportation statistics including government spending on various transportation sectors, miles traveled by vehicles on highways, and vehicle sales amongst others. We specifically focused on the total miles traveled by highway vehicles, number of rail employees and users, and state and local government construction spending on transportation, mass transit, highway and street, infrastructure, and pavement. Each row represents a monthly or quarterly statistic.

It would have been useful to have data regarding miles traveled by vehicle outside of highways, for example in residential or commercial areas, to compare changes in travel patterns to highways. The data was also aggregated by month, and we think a more thorough analysis can be done if we had access to the statistics by day. Some columns such as rider satisfaction with rail systems and cleanliness score or rail systems might have also been useful as confounders for our causal inference question but they were unavailable. “Participants” of the data could include in our case: riders/employees of rail systems and government agencies, all of which should be aware of the collection of data. The data contains aggregated statistics for the U.S. as a whole, including every state; there may be a possibility of certain counties/states not maintaining their data regularly and being excluded in the dataset. For this reason, we do have concerns over measurement error. In addition because a state would need to work with multiple agencies in rail systems to aggregate data, this further confounds measurement error. Overall, the dataset is not very robust due to the data aggregated by month, but it is very comprehensive going as far back as 1940’s.

EDA

The years of the pandemic were 2020 and 2021 (2022 is only halfway over, and it is still considered a 'Pandemic State' for many parts of the US, so we will consider its data is too convoluted to include in analysis at this time and exclude it from analysis). We observe a trend that there is a significant increase in spending on mass transit in the years of 2020 and 2021 compared to 2019, a 'Pre-Pandemic year' for the US, as shown in the following graph:

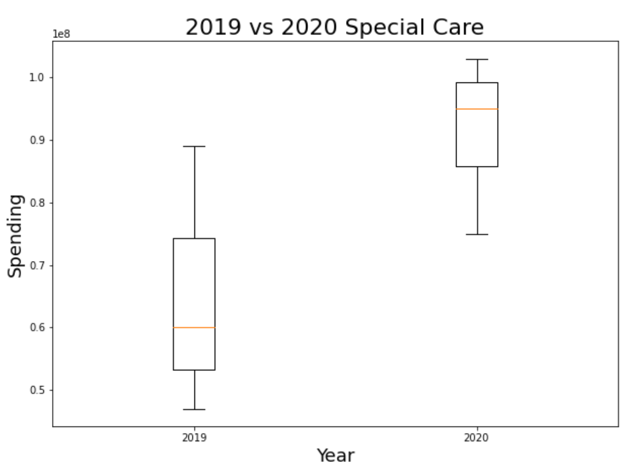


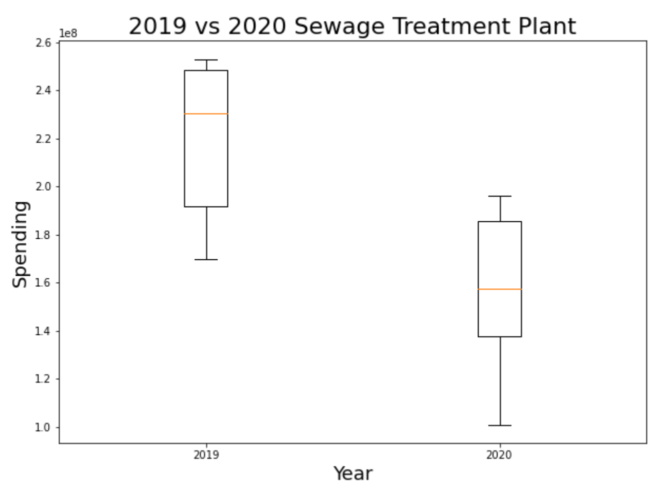
The data set given was very large and included many columns/rows that were unnecessary for our research questions. As such, we dropped the rows which contained irrelevant dates (1970-2018) and only picked the columns regarding government spending on construction that related to transportation.

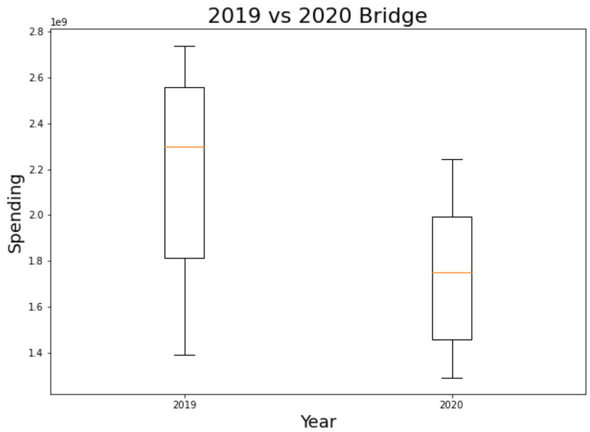
These decisions will impact the models and inferences, because noting the amount of columns on government construction spending alone there were many! causes one to think about if a valid data point to include would be how much the government spent on transit construction in comparison to their many hundreds of other construction projects. We will have to take the government's spending capacity throughout a pandemic state into account as such throughout our models and inferences.

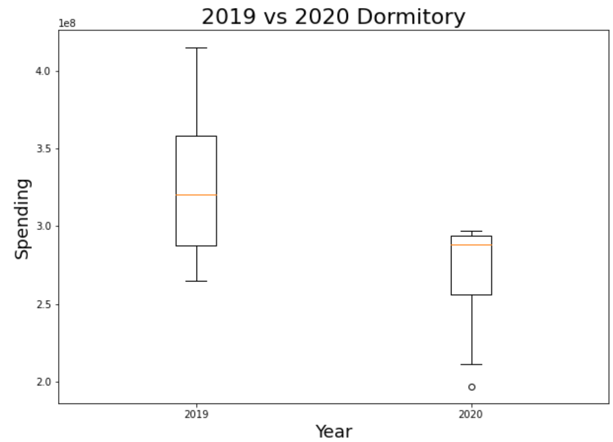
Through cutting all dates before 2019 - the last ‘pre-Pandemic’ year - we lose our ability to gain insight into the usual highs and lows of government transit spending on construction before the pandemic. As such, our baseline of 2019 could be flawed and we must take that into account when performing our models and analysis.

This visualization maps government spending on mass transit construction before and during the pandemic years we define and set in our project. Therefore, seeing that the trend shows an increase of spending during the pandemic, we are further motivated to look into this increase in government spending during the pandemic and gain more insights - look into the potential for confounding variables - throughout our research.







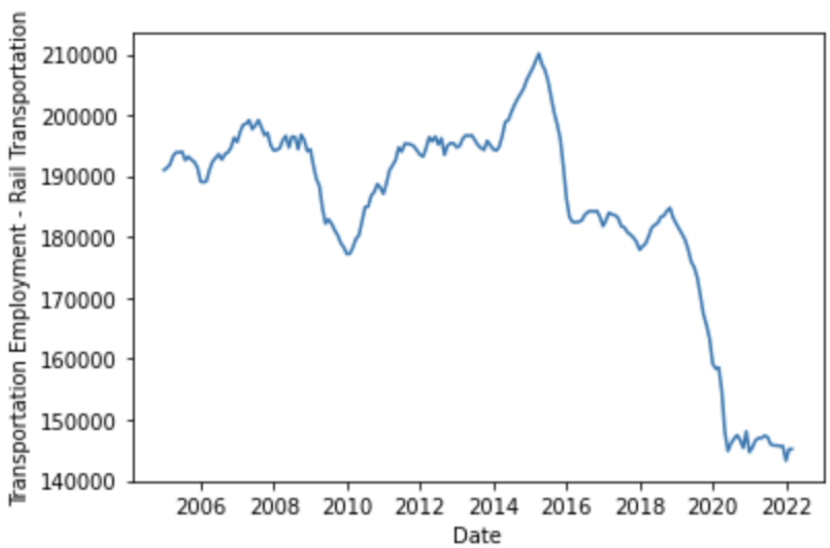


One trend that was fairly consistent was the spread of the different amounts of spending between the two years. Spending in 2019 tends to be more spread out as can be seen with the generally larger box and whiskers on the box plots. In comparison, the spending in 2020 tends to have a smaller spread meaning that the spending tends to be more consistent. This trend could possibly be correlated with the pandemic and may warrant further investigation.

Several things were removed during the data cleaning process. First, all of the columns that were not related to State and Local Government Construction Spending were removed from the dataframe. Subsequently, two dataframes were constructed with all of the data from 2019 and another with all of the data from 2020. Each dataframe ended up having 12 data points, one for each month, with all of the spending for the year.

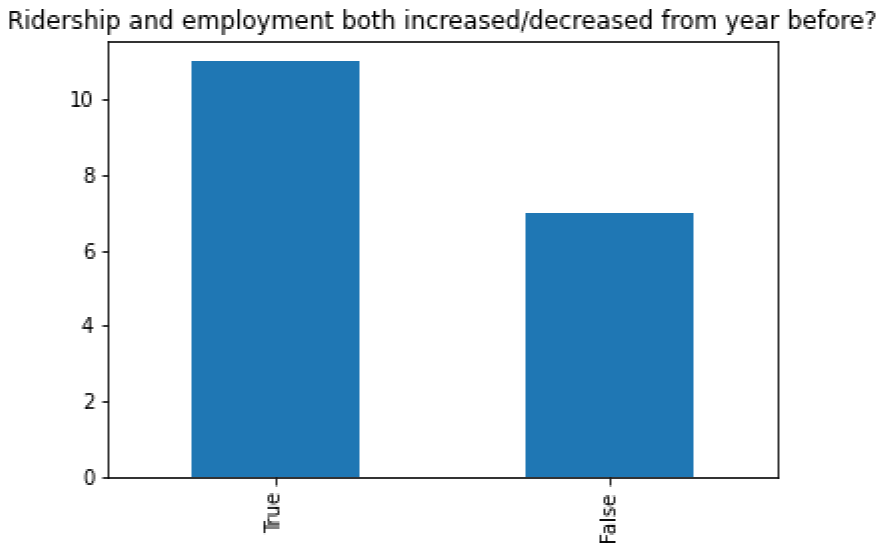
The granularity of the dataset and the relatively short length of the pandemic will limit the amount of data points we will have access to. This will make it harder to show clear patterns in data because of the limited sample size we have available. Moving forward we will likely use more of the available (2018 - 2022) to increase the number of data points we can perform inference on.

These visualizations directly answer our first question "How has government spending on public infrastructure and transportation in highly populated areas changed from pre-pandemic (2019) to during the pandemic (2020)?". We can see the differences in the distributions of spending between 2019 and 2020. From the boxplots we can see that in some cases where the spending in 2019 is greater and others where it is lesser than 2020. This means that spending did not consistently increase or decrease between these two years overall. With this in mind, our future work will likely look at how different spendings were prioritized in different years.



In regards to the second research question, we began by simply looking at trends of employee counts at railways over time. This is seen by the line graph visualization above. We reformatted the ‘Date’ column of the data to be by year rather than by monthly timestamps, allowing for the graph to be more readable.

Looking at the trends of employee counts over time, there are many peaks and valleys worth exploring. This brought us to our next visualization and exploration of potential relationships between railway ridership and number of railway employees.



Interested in exploring the relationship between ridership and employment, we created the visualization above. Working with the same dataset, we created the columns “Ridership increased from year before?” and “Employment increased from year before?” consisting of booleans indicating whether the number of rail employees and riders increased from the year before. Then we created another column, “Ridership and employment both increased/decreased from year before?, which indicates whether the two columns followed the same trend of increasing or decreasing from the year before. Choosing to look at changes in ridership and employment on a yearly basis will impact the results of our analysis, namely as we eventually change to look at 6 month intervals instead, but still provides a general overview of the direction of our research question.

Looking at the bar graph, we see that there are more counts of True than False, meaning there are more years where rail ridership and employment followed the same general pattern of increasing or decreasing. This helps answer our second research question, as it gives general indication that there may be a relationship between number of rail riders and number of rail employees, namely that they have a somewhat positive correlation where an increase in one may indicate an increase in the other variable and vice versa. We will follow up on this observation to look at smaller time intervals and evaluate their significance as well as causal relationship.

Multiple Hypothesis Testing

Overview

Our research question that we explored with hypothesis testing was: How has government spending on public infrastructure and transportation in highly populated areas changed from pre pandemic (2019) to during the pandemic (2020)?

As such, our hypotheses explore government spending in different aspects of transportation. Since there are many different aspects of transportation that make the process come together - roads, highways, streets, pavement - it makes sense to test many hypotheses instead of just one to get a full breadth of understanding.

Additionally, testing multiple hypotheses that are tangentially related to spending - miles traveled and infrastructure - can bolster our argument or bring us in a different direction with our research question based on the results.

We used Bonferroni and B-H methods to create error cutoffs. Bonferroni controls for the Family-Wise Error Rate (FWER) or the probability that any of the tests are a false positive and

Benjamini-Hochberg controls for the False Discovery Rate (FDR) or the expected value of the false discovery proportion. The cutoff p-value for Bonferroni is 0.00833 and the cutoff p-value for Benjamini-Hochberg 0.00240 for our total of 6 hypothesis tests.

Methods/Results

Test 1:

Null Hypothesis 1: There was no change in miles by land traveled in both 2020 and 2019.

Alternative Hypothesis 1: There was a change in miles by land traveled in both 2020 and 2019.

Method: A/B testing.

It is fruitful to use A/B testing for these kinds of hypothesis questions, since we are comparing results from two different years. Scrambling the results between the years is a very usable and applicable method of A/B testing.

P-value: 0.0024

The p-value obtained is significant under both error rates controlled with an alpha value of 0.05. Meaning that the null hypothesis is to be rejected in favor of the alternative hypothesis. This change in miles by land traveled in 2020 vs 2019, we believe is due to the effects of COVID-19 restrictions. Commuting became relatively nonexistent in 2020 which has likely led to there being less miles traveled by land causing the difference in miles traveled by land in 2019 vs 2020. We thought that these changes would also be reflected in the spending, which led to our next hypothesis tests.

The limitation of this result is that we had to assume independence between each month, which may not have been the case, and additionally there was only data for each month, when it would have been ideal to have it for each day.

Test 2:

Null Hypothesis 2: There was no more spending on transportation by land in both 2020 and 2019.

Alternative Hypothesis 2: There was a difference in spending on transportation by land in 2020 vs 2019.

Method: A/B testing

P-value: 0.1276

The p-value obtained is not significant under any threshold, meaning that we fail to reject the null hypothesis. While the p-value is relatively low, it could still be due to random chance. Therefore, spending on transportation by land did not differ between 2020 and 2019 despite what appeared to be significant change in miles traveled by land.

The limitation of this result is that we had to assume independence between each month, which may not have been the case, and additionally there was only data for each month, when it would have been ideal to have it for each day.

Test 3:

Null Hypothesis 3: There was no difference in the spending on mass transit in 2020 and 2019.

Alternative Hypothesis 3: There was a difference in the spending on mass transit in 2020 and 2019.

Method: A/B testing

P-value: 0.076

The P-value obtained from out A/B testing is well above all of the thresholds for the naive, bonferroni, and B-H cutoff values. This means that the differences are not significantly significant and spending on mass transit did not greatly differ between 2019 and 2020.

A possible limitation of this result is that we assumed independence of the data points. In reality, it is very likely that the spending from one month is dependent on the spending of previous and future months. In addition to this, there could be other factors besides the pandemic that affected the spending. For example, budgets for spending on mass transit could have been set before the pandemic and not reconsidered during the pandemic.

Test 4:

Null Hypothesis 4: There was no difference in the spending on highways and streets in 2020 and 2019.

Alternative Hypothesis 4: There was a difference in the spending on highways and streets in 2020 vs 2019.

Methods: A/B Testing

P-value: 0.949

The P-value for the spending on highways and streets is extremely high meaning that it is very likely that there was no difference in the spending on highways and streets between 2020 and 2019. As such, we accept the null hypothesis. This remained the case for all cutoffs before and after the corrections.

A limitation of this test is the assumption of independence between all of the data points. In reality, the spending from month to month is likely dependent on the spending from other months both in the future and in the past. Additionally, there could have been other factors that led to these results. For example, if there was a large project or initiative for highways and streets, similar amounts could have been budgeted for these years leading to very similar spending.

Test 5:

Null Hypothesis 5: There was no difference in the spending on infrastructure in 2020 and 2019

Alternative Hypothesis 5: There was a difference in the spending on infrastructure in 2020 and 2019.

Methods: A/B Testing

P-value: 0.0533

While this P-Value is close to our naive threshold of 0.05, it does not quite go below this cutoff or any other cutoff from the correction techniques we implemented. As such, we accept the null hypothesis meaning that from this test, there was no difference in the spending on infrastructure in 2020 and 2019.

There are limitations to this study that could have affected the results. First, an assumption of independence was made for each of the data points. This is however likely not the case because spending from any given month could affect future months. For example, if a large amount of money was spent to complete a big project, then once that project is completed, there may no longer be a need to spend as much money in future months.

Test 6:

Null Hypothesis 6: There was no difference in the spending on pavement in 2020 and 2019.

Alternative Hypothesis 6: There was a difference in the spending on pavement in 2020 and 2019.

Methods: A/B Testing

P-value: 0.6122

The P-value for this test is very high and above all of the cutoffs before and after the corrections were made so, we accept the null hypothesis. In this case, this means that there was no difference in the spending on pavement from 2020 to 2019.

A limitation of this study is the assumption of independence of the data points. In reality, it is likely that the spending from one month affects the spending of other months. It is also possible that other factors besides the pandemic contributed to either the difference or similarity in the spending. An example of this could be that in previous years many pavements were fixed and spending was no longer as necessary in the following years.

Discussion

After the correction procedures none of the discoveries were changed. This was likely due to the fact that the test for the number of miles traveled by land had the only statistically significant result from using just the naive cutoff. Additionally, with an extremely small P-value of 0.0024, the result would have likely stayed significant under most cutoffs due to how unlikely the null hypothesis was.

The individual tests may guide us to or away from proving our research question, but cannot lead us to prove with reasonable confidence the result of our research question. However, the results in aggregate we believe are comprehensive enough to come to a reasonable conclusion to our research question. This is because our research question covers a very broad topic of government spending on transportation. Transportation is a very multi-faceted topic that cannot be tackled with a singular hypothesis test.

The limitations in our analysis come from our dataset only having data for each month of a given year. This limits the amount of data we have to work with altogether. Additionally to perform the A-B tests the assumption of independence is needed. The pandemic might have not been the only factor leading to these results. We worked to avoid p-hacking through bonferroni and B-H corrections.

We think it would be very useful to have more data on the breakdown of spending to conduct tests on the nuances of transportation. For example, rather than having a column for government spending on mass transportation, there would be one on spending on city roads specifically. A dataset with finer granularity in location would allow us to perform the correction techniques on the p-values for each of the individual cities for the same kind of spending instead of applying the correction techniques to the P-values for the different types of spending.

Causal Inference

Overview

We used causal inference to answer the question of how increases to the number of urban rail riders affects the number of rail employees?

In order to answer this question, we used the Monthly Transportation Statistics dataset put out by the Bureau of Transportation. This dataset includes various data about transportation in the United States on a monthly basis. One variable of interest was the adjusted number of urban rail riders instead of raw counts of the ridership for a given month. The adjustment was done by the Bureau of Transportation to account for the normal fluctuations in ridership between different months out of the year. Very little information was documented about what these adjustments entail, so in our analysis we assume that the data was correctly processed by the Bureau of Transportation. Under this assumption, we used the adjusted ridership to determine our treatment and control groups. A given month was part of the control group if in the month prior, the urban rail ridership was greater than the average ridership from the previous six months. Subsequently, a given month was a part of the control group if the urban rail ridership from the month prior was less than the average ridership from the previous six months.

We chose to look at the difference between the number of rail transportation employees for a given month between these two groups. To do this, we looked at monthly estimates put out by the Bureau of Transportation as a part of the Current Population Survey. In this survey an employed person is defined as “people aged 16 years and older in the civilian noninstitutional population who did any work at all as paid employees”. In our analysis, we assumed that this estimate produced from this survey is an accurate estimate of the number of employees.

We use inverse propensity weighting, using a logistic regression model for the propensity scores. We account for confounders such as government spending. We assume that these are all the confounders that are required for us to draw causal conclusions.

Methods

The variable involved in the treatment is “Transit Ridership - Urban Rail - Adjusted” and we create a new binary variable “Increased Ridership” if there was an increase in the previous month's ridership compared to the average ridership from 6 months before. The outcome variable is the number of employees for rail systems or “Transportation Employment - Rail Transportation”

We identified 3 main confounding variables in our dataset. Each confounding variable was the proportion of the total spending for six months before a given month. We assumed that changing spending amounts to proportions would help mitigate the effects of inflation on the spending. In addition to this, we decided to make the proportions from six months prior because we assumed that construction spending would not have an immediate effect on the treatment or effect.

The first two confounders, construction spending on mass transit and transportation, were chosen because of the same line of reasoning. We held the belief that increased construction on transit in a given month would lead to lower ridership affecting the average used to determine the treatment group. Additionally, we assumed that these spendings would lead to better transit systems which in turn could lead to less of a need for employees to run the rail systems. The final confounder was construction spending on highways and streets. We assumed that this was a confounder because better roads could lead to less of a need to ride the urban rail system and make it easier for rail employees to get to work.

We assume that the unconfoundedness assumption holds. In other words, the treatment is independent of the outcomes given the confounders. This is a reasonable assumption due to the time differential between our treatment “increased ridership” (6 months before) and the outcome number of employees, so accounting for confounders we can isolate the causal effect of number of riders on number of employees. However, as we are unable to account for all of the confounding variables, the unconfoundedness assumption likely does not hold.

We use inverse propensity weighting as our method to adjust for confounding variables. Because our treatment is hand-crafted, this method will help adjust for the bias in the treatment assignment. To find the propensity scores of the data points, we chose to use SciKit-Learn's vanilla logistic regression model where the input was the confounders and the output was whether that data point received the treatment. After creating this model, we measured its performance using an ROC curve that had an AUC of 0.73 showing that the model was an adequate predictor of the treatment. With this in mind, we assumed that the probabilities produced by this model could be used as propensity scores. No adjustment for outliers were necessary because all of the probabilities were within the standard range of [0.1, 0.9].

We do not suspect that there are any colliders in the dataset.

Results

The naive average treatment effect is 2961. We obtained a score of 51,931 using inverse propensity weighting. Our results demonstrate that there is a significant effect of the number of rail riders on the number of rail employees, but due to the lack of access to a diverse set of confounding variables, the results remain inconclusive. Compared to the naive estimate, the inverse propensity score ATE is significantly larger, demonstrating that the treatment likely has a larger effect on the number of employees than what initially might seem.

Discussion

The limitations of this method are that it is necessary for the unconfoundedness assumption to hold by accounting for all confounding variables.

Additional data to add to our confounding variables, such as cleanliness of rail stations, would have been useful for answering our causal question. When exploring the relationship between ridership and number of employees, confounders such as the rail-station conditions can greatly affect both. If the station is too unsanitary to be rideable/workable by either groups, then it would have been important to include in our analysis.

Our confidence that there is a causal relationship between our chosen treatment and outcome is not high. While we did observe a strong relationship in our results, our limited number of confounding variables limit our confidence in the results respectively.

Conclusions

Through our exploration of our research questions, we have gained a better understanding of pre-pandemic government construction spending and the relationship between rail riders and employees. Our first research question: “How has government spending on public infrastructure and transportation changed from pre-pandemic (2019) to during the pandemic (2020)?” was explored through means of multiple hypothesis testing. Our hypothesis tests delved into different aspects of government spending such as differences in spending on pavement and roads as well as differences in miles traveled.

In regards to our multiple hypothesis question, our findings brought us to the conclusion that - aside from the change in miles traveled from pre-pandemic (2019) to pandemic (2020) years - we failed to reject the null hypothesis for any of our other hypothesis tests on government transportation spending. Our study included limitations, such as working under the assumption that each data point was independent when in reality it may not be sound to assume so.

We can reasonably conclude that the hypothesis test that successfully rejected the null hypothesis - the differences in miles traveled between pre-pandemic and pandemic years - was due to the high reduction of commuter traffic during the pandemic year of 2020. As for our other hypothesis tests in which we failed to reject the null, we can speculate as to why there were little to no differences in government spending before and during the pandemic years. One speculation is that road and walkways were preserved during 2020 to ensure essential workers had maintained passages for their work. For example, an ambulance would need maintained highways and an EMT would need even pavement in order to save a life from COVID or anything else. Another speculation is that constant maintenance of roadways is worth the financial investment in the long run. Allowing the roads to deteriorate on the ideology that they can be repaired after the pandemic may end up being more costly and greatly hinder the return to office and school transition.

Our results for the multiple hypothesis testing research question above are not highly generalizable to comparing years outside of the pandemic period. The COVID-19 pandemic has been a very unique period of modern history in which different states, countries, and individuals have reacted differently. Thus, the data gathered during this time period is not highly generalizable without accounting for the variety of limitations.

In regards to our causal inference research question, our findings found that while there is a strong relationship between ridership and number of employees, a conclusive determination drawn from this relationship is not possible due to the limitations of our study. Since our causal inference analysis requires more confounding variables to gain a more accurate conclusion from the results, we refrain from stating with high confidence that the number of employees is caused by ridership of railways.

Our results for the causal inference research question are more generalizable to ridership-employee relationships for railways, as they are not focused on a certain pandemic period.

Overall, this project has been a fruitful way of exploring the differences in government spending on transit before and during the pandemic and the causal relationship between ridership and number of railway employees. While some of our results weren't as conclusive as we had hoped, this process has been a step all data scientists take in their journey through utilizing data to draw ethical conclusions about our world.