The Role of Transportation in *Driving* Early COVID-19 Spread

Group Members: Nina Rice, Sunaina Martin, Holly Kular, Richard Vagnino, & Hannah Lloyd

On March 19th, 2020, Governor Gavin Newsom declared a *Stay at Home Order* for the state of California. As the first reported COVID-19 cases in California were in the San Francisco Bay Area, we took an interest in investigating how people's compliance with California's *Stay at Home Order* was related to early COVID-19 spread in the SF Bay Area. Studies conducted early in the pandemic have suggested that stay at home orders are associated with slowing the virus's spread (Medline et al., 2020). Latency between COVID-19's appearance and the resulting institution of preventative policies is positively related with the harm that the virus causes — in other words, illness, hospitalization, and death. This elicits the question on whether compliance with the stay at home order, as measured by movement(i.e., transportation), influences COVID-19 spread.

For the purposes of this project, we defined the Bay Area as: San Francisco County, Alameda County, San Mateo County, and Santa Clara County. These are four of the five most populous counties in the San Francisco Bay and together make up a significant portion of the region's geographic area. It is likely that they are indicative of the Bay Area's behavior as a whole. Transportation during the *Stay at Home Order* offers an approximation of compliance (or lack thereof) since movement from transportation will include people spreading COVID-19 through contact with non-household members. This analysis endeavors to reveal which counties in particular contributed to the early spread of COVID-19. Furthermore, these counties often instituted unique regulations from one another, for example taking different approaches to penalizing disobeying the *Stay at Home Order*.

Since we were only able to get ridership from public transportation agencies at the monthly level, we primarily focused on private vehicle transportation. Our measure of travel by car comes from the CalTrans Performance Measurement System (PeMS), provided by California's Department of Transportation. This dataset includes measurements of freeway traffic through vehicle detection stations, weight-in-motion sensors, and traffic census stations. We compared daily vehicle travel reports as measured by vehicle miles traveled on the freeways, binned by county (Fig. 1). To understand the effect of alternative transportation mediums, we examined cell phone GPS data provided by the Bureau of Transportation Statistics, which uses mobile data to track how many people are staying at home and how far they are travelling (Fig. 2-4). Trips are defined as movements that include a stay of longer than 10 minutes at an anonymized location away from home. Trips capture travel by all modes of transportation including driving, walking, public transit and air travel.

The spread of COVID-19 was measured via confirmed cases (Fig. 5) and hospitalization (Fig. 6) rates by county. Cases are defined in this context as positive PCR tests. Case data is available by day beginning in February 2020, updated weekly. Hospitalization data is also available by day. Case count and hospitalization data came from California's Open Data Portal.

We found that consistent data input within our four counties started late in the month of March. Thus, we decided on examining the range of dates between March 30, 2020 and September 30, 2020.

Cleaning and transformation:

Data were condensed by taking the mean for each week in order to standardize the time variable for all the datasets. In order to account for an effect of seasonality, we normalized the movement data to the average movement in January and February of 2020. The hospitalization and case numbers were log transformed, as were the vehicle travel data.

Analysis:

ANCOVAs (Fig. 7 & 11) were used to determine the relationships between movement and COVID spread. We used time as the covariate and compared the COVID-19 cases and hospitalizations as a function of vehicle movement, short trips, medium trips, long trips and county.

Assumption Checks. Due to the seasonality of the data, the assumption of linearity is violated. Further ANCOVA model assumptions were checked. A correlation matrix was computed for each model. In the model predicting COVID-19 cases by week, county, VMT, and trips the correlation matrix (Fig. 8) shows that some but not all covariates satisfy the assumption of independent variables. For the same model predicting cases, the distribution of residuals (Fig. 9) is normally distributed, so the assumption of normality in the response variable is not violated. Checking for equality of variance, the residuals were plotted against the model predicted COVID-19 values (Fig. 10), this plot shows no clear trend, so the assumption is not violated. In the model predicting COVID-19 hospitalizations by week, county, VMT, and trips the correlation matrix (Fig. 12) some but not all covariates satisfy the assumption of independent variables. The correlation matrix (Fig. 12) shows that some but not all covariates satisfy the assumption of independent variables. For the same model predicting cases, the distribution of residuals (Fig. 13) is slightly skewed, so the assumption of normality in the response variable may be violated. Checking for equality of variance, the residuals were plotted against the model predicted COVID-19 values (Fig. 14), this plot shows a clear trend, so the assumption of homoscedasticity is violated.

Results:

COVID-19 cases as a function of movement

There is a main effect of time ($F(1,99) = 167$, $p < 0.001$), region ($F(3,99) = 31.3$, $p < 0.001$), vehicle movement ($F(1,99) = 59.2$, $p < 0.001$), short trips ($F(1,99) = 56.6$, $p < 0.001$) and long trips ($F(1,99) = 8.84$, $p = 0.03$) on the number of COVID-19 cases (Fig. 7). The main effect of time indicates that as we progress through the pandemic, the cases increase. This finding doesn't address any of our research questions, but demonstrates a general trend in COVID-19 spread. We believe this finding demonstrates that when restrictions were eased with time, individuals could interact more within the community and thus, increase COVID-19 spread. The main effect of a region could be due to the population, size(in terms of land mass), or density within a region. This could also be a result of COVID-19 based restrictions within a county. The main effect of vehicle movement provides a partial answer to our research question such that as a community moves more (and potentially violates the lockdown restrictions) COVID-19 cases increase. Finally, there were main effects of two types of trip size: short and long trips. Short trips are within the community. Given our county size, long indicates between communities. We suspect that medium trips have no significant main effect since medium trips are likely trips for work/professional purposes. If an individual is going into their office or work, they are seeing the same people and might not likely be contributing to community spread.

There is an interaction between the county and the movement ($F(3,32) = 18.7$, $p < 0.01$), specifically, it appears that there is an interaction between county type and the number of trips taken that were under 10 miles ($F(3,43) = 7.85$, $p = 0.0003$) in predicting the number of COVID-19 cases. The necessity of cars is different across counties and the use of private transportation could interact with COVID-19 spread. The possibility that the number of car owners between counties could be due to the available public transportation options. Additionally, the type of jobs people in these counties hold, could account for the interaction between county and vehicle movement. The availability of proximate resources, (e.g., such as grocery stores), could account for the differences in the numbers of short trips between counties.

COVID-19 hospitalizations as a function of movement

There is a main effect of time ($F(1,99) = 9.37$, $p = 0.003$), region ($F(3,99) = 54.5$, $p < 0.001$), vehicle movement ($F(1,99) = 16.2$, $p <0.001$), short trips ($F(1,99) = 58.5$, $p < 0.001$) and long trips ($F(1,99) = 6.66$, $p = 0.0113$) on the number of hospitalizations (Fig. 11) . The main effect of time demonstrates that as the pandemic progresses, there is an increase in the number of hospitalizations related to COVID-19. With fewer restrictions to movement towards the end of the period of interest, greater movement of individuals resulted in greater COVID-19 spread, as seen above and therefore, greater number of hospitalizations related to COVID-19. The main effect of the region accounts for the fact that the four counties in question have different population densities. However, given that many people moved out of more densely populated cities in the Bay Area to the suburbs, the trends observed could also have been influenced by this exodus. The main effect of vehicle movement indicates that as more people drive, there is an increase in the number of hospitalizations, providing an answer to our initial question of whether non-compliance with stay at home orders can predict COVID-19 spread. We see the same main effects for the short and long trip types, with short trips representing community spread and long trips potentially accounting for between community spread.

There is an interaction between the county and the movement ($F(3,32) = 13.6$, $p < 0.01$), specifically, it appears that there is an interaction between county type and the number of trips taken under 10 miles ($F(3,43) = 6.93$, $p< 0.001$) in predicting the number of COVID-19 cases. The possibility that the number of car owners between counties could be different based on the available public transportation and the type of jobs people in these counties hold, could account for the interaction between county and vehicle movement. Additionally, the availability of proximate resources (e.g., grocery stores),could account for the differences in the numbers of short trips between counties.

Discussion:

COVID-19 related hospitalization rates were integrated as a response variable to establish a validity check on comparing our models with COVID-19 cases. Future directions would integrate these two measurements simultaneously as co-response variables. The nature of the pandemic imposed a limitation in the COVID-19 case measurement since there was no access to COVID-19 tests for the majority of the population. Since it took some time for COVID-19 tests to have widespread availability, we built models for both hospitalization rates (a seemingly valid measurement for early COVID-19) and COVID-19 cases (a seemingly valid measurement of COVID-19 over time within county communities). We found similar results in comparing both of these measurements, demonstrating a strong validity of our response variables.

There's been a debate on whether a lockdown was effective. These analyses show that movement, as measured by transportation, does contribute to the spread of the disease. Future directions would contribute to this question by comparing different geographic regions with differing lockdown restrictions to see if those restrictions had an influence on COVID-19 spread.

In summary, the results from these analyses provide new insight into how COVID-19 spreads within a region by looking at transportation and county. As the trips dataset overlaps with the VMT dataset, and the results with both datasets are consistent, we focused on the trips and case datasets for figures 15 and 16. Since there is an effect of short trips, there is more of an effect of within community movement on the spread of COVID-19 . Finally, there was an effect of long trips, such that COVID-19 spread was also potentially facilitated by community spread for long trips, which may indicate noncompliance with lockdown restrictions (Fig. 15&16) Transportation is an indicator of behavior, and our results signify that movement predicts COVID-19 spread.

There are important caveats to consider with the aforementioned analyses. With data that spans over time, there is likely a seasonality effect which cannot be captured well with a linear regression. Additionally, the COVID-19 cases and hospitalization data is non-linear. This data would be better fit with a model that is beyond this course. A limitation for the vehicle miles traveled transportation data is that San Mateo County had a marked decline in the % of lane points observed meaning data was imputed due to missing samples. This could be due to sensor failure or notably the August Fire Complex which impacted a large portion of San Mateo County's freeway system.

References

Medline, A., Hayes, L., Valdez, K., Hayashi, A., Vahedi, F., Capell, W., Sonnenberg, J., Glick, Z., & Klausner, J. D. (2020). Evaluating the impact of stay-at-home orders on the time to reach the peak burden of covid-19 cases and deaths: Does timing matter? *BMC Public*

Figures & Tables

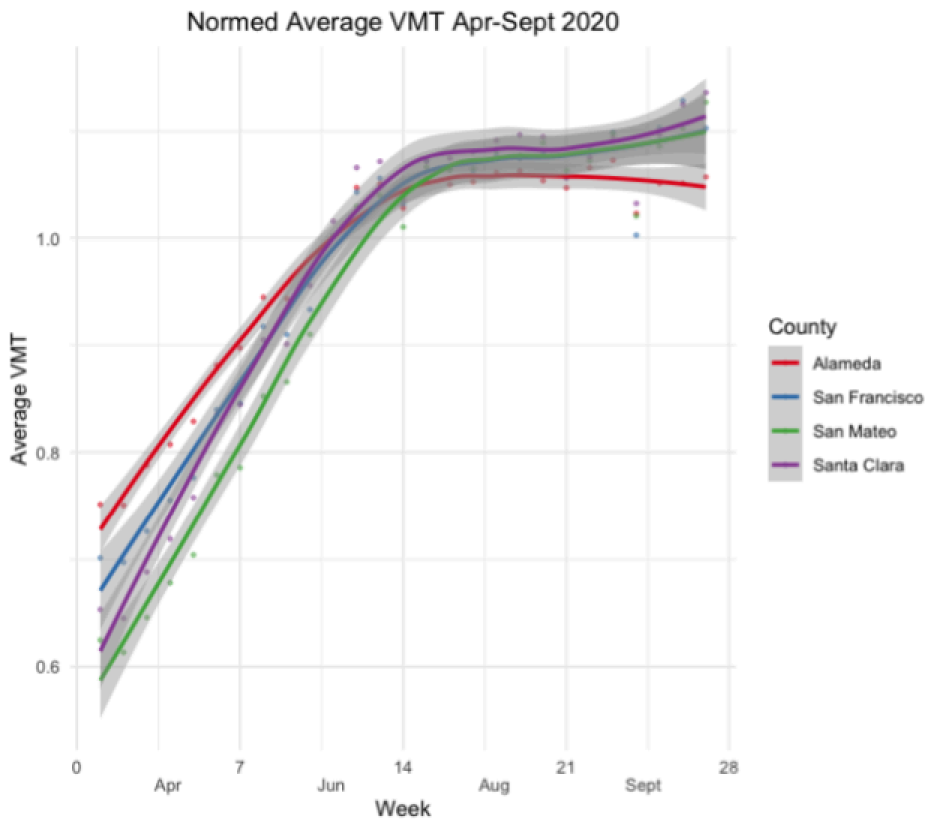

Figure 1. Weekly averages of vehicle miles traveled (VMT) in each county for April-September 2020, normed to January-February 2020 VMT. Fitted with loess method.

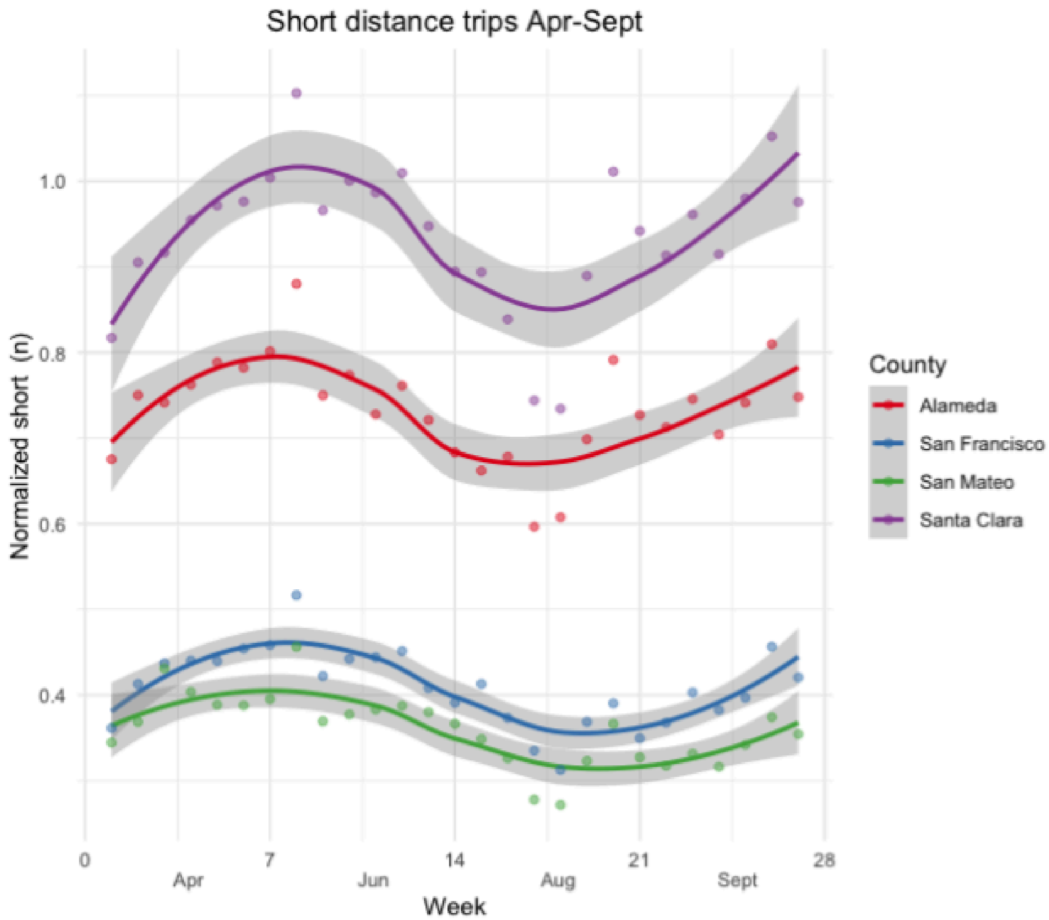

Figure 2. Weekly averages of short distance (0-10 mi) trip counts in each county for April-September 2020, normed to January-February 2020 short distance trip counts. Fitted with loess method.

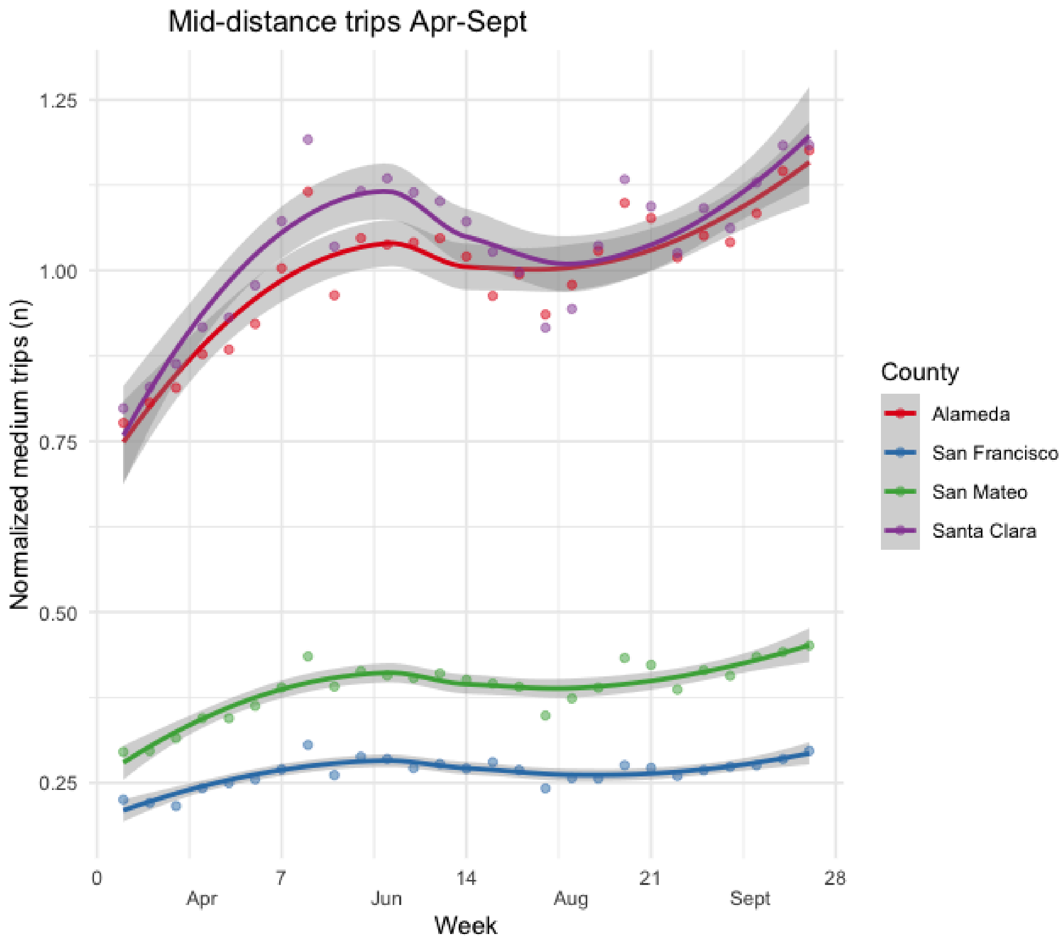

Figure 3. Weekly averages of medium distance (10-50 mi) trip counts in each county for April-September 2020, normed to January-February 2020 medium distance trip counts. Fitted with loess method.

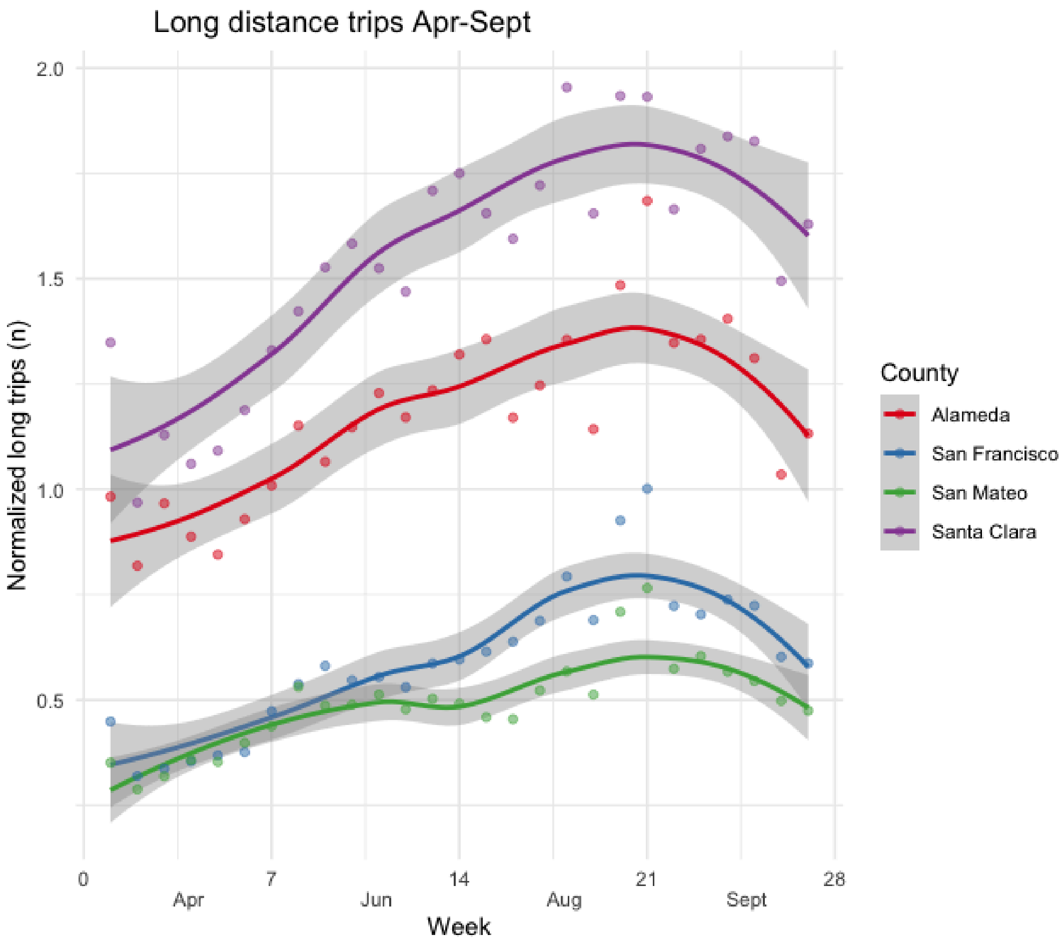

Figure 4. Weekly averages of long distance (>50 mi) trip counts in each county for April-September 2020, normed to January-February 2020 long distance trip counts. Fitted with loess method.

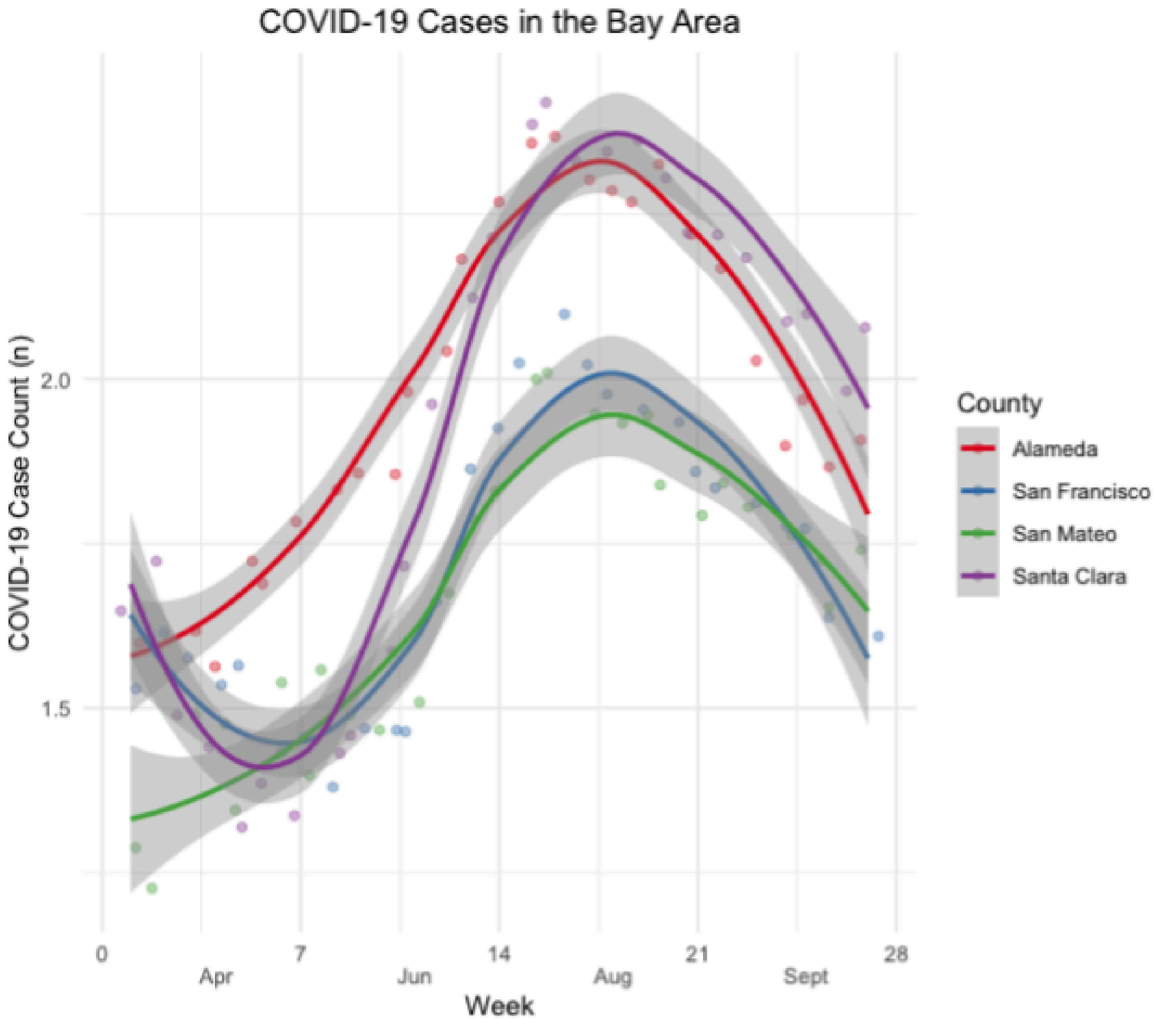

Figure 5. Weekly averages of COVID-19 case counts in each county for April-September 2020. Fitted with loess method.

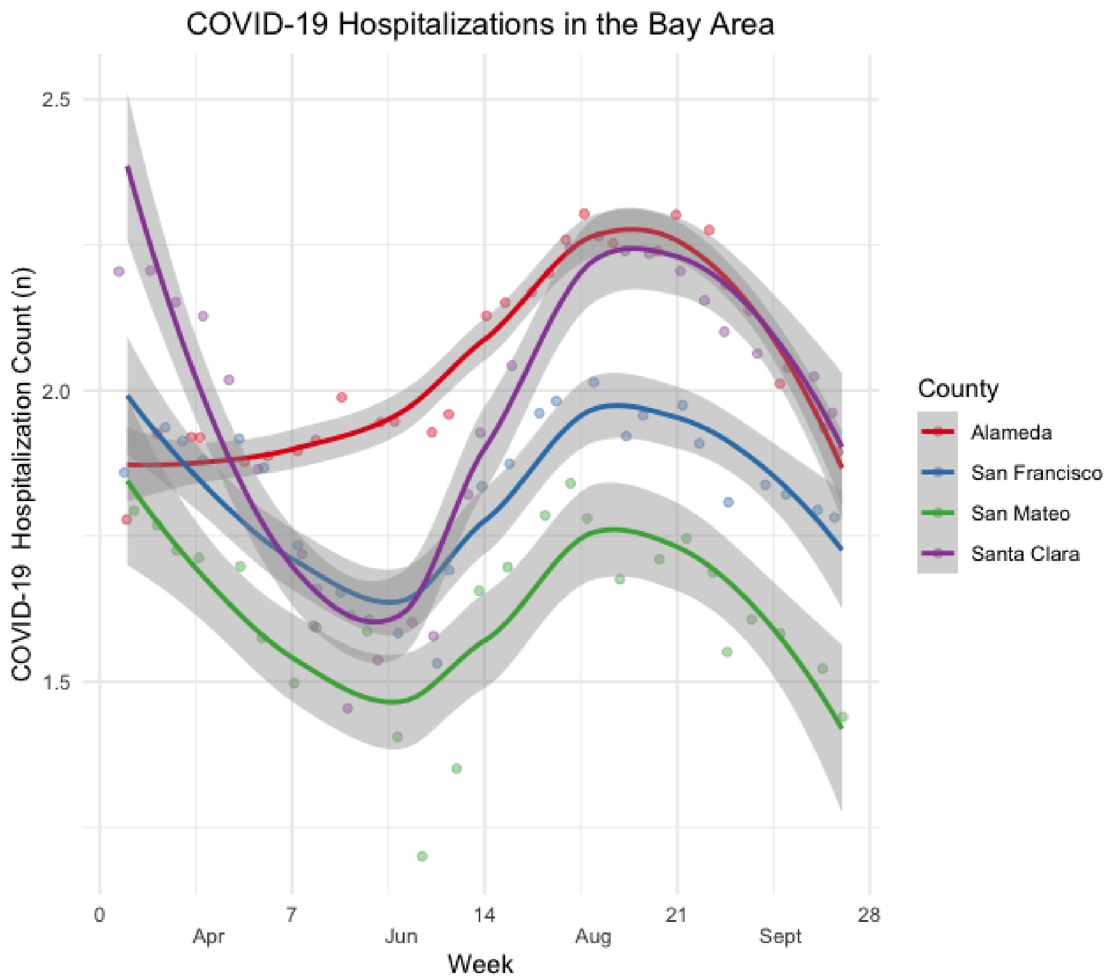

Figure 6. Weekly averages of COVID-19 case counts in each county for April-September 2020. Fitted with loess method.

Figure 7. Summary of ANCOVA model predicting Covid-19 cases by week, county, VMT(mi_norm), and trips (normal_short = short trips, normal_medium = medium trips, normal_long = long trips.).

```
Analysis of Variance Table

Response: log10(m_cases)
              Df Sum Sq Mean Sq  F value    Pr(>F)    
week           1 3.4010  3.4010 166.6428 < 2.2e-16 ***
County         3 1.9155  0.6385  31.2849 2.607e-14 ***
mi_norm        1 1.2088  1.2088  59.2306 1.068e-11 ***
normal_short   1 1.1556  1.1556  56.6223 2.464e-11 ***
normal_medium  1 0.0776  0.0776   3.8003  0.054073 .  
normal_long    1 0.1805  0.1805   8.8437  0.003694 ** 
Residuals     99 2.0205  0.0204                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

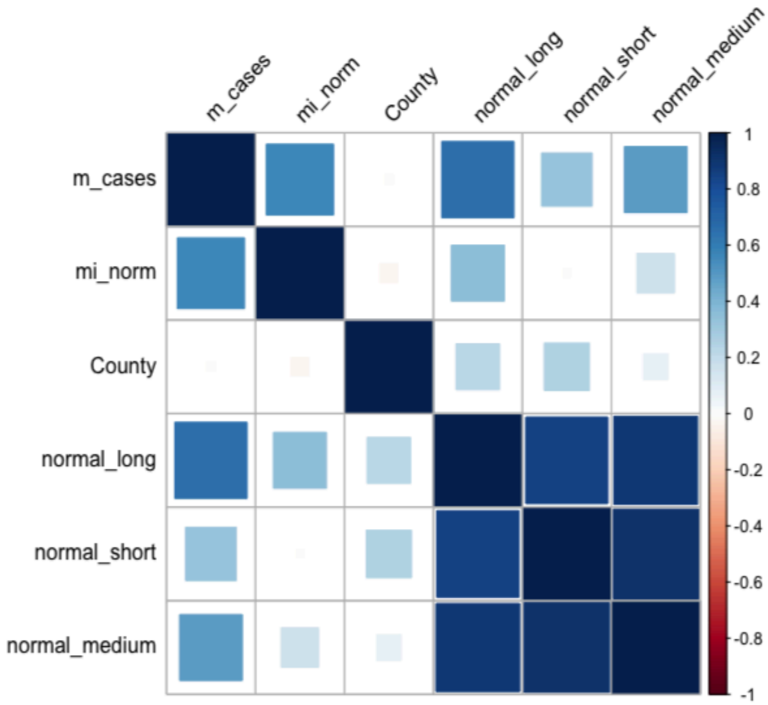

Figure 8. Correlation matrix for the covariates in ANCOVA model of Covid-19 cases predicted by week, county, VMT(mi_norm), and trips (normal_short = short trips, normal_medium = medium trips, normal_long = long trips.).

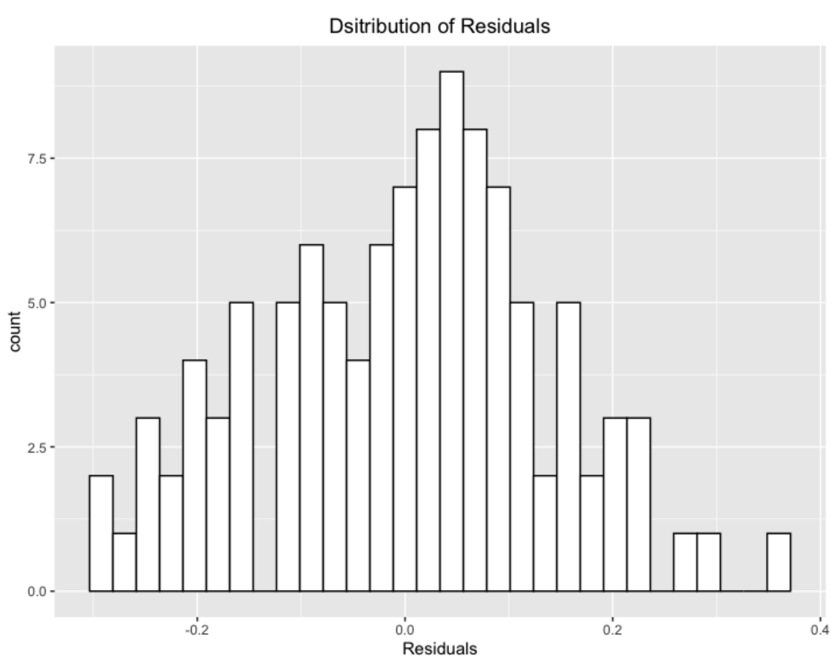

Figure 9. Distribution of residuals of ANCOVA model predicting Covid-19 cases by week, county, VMT, and trips.

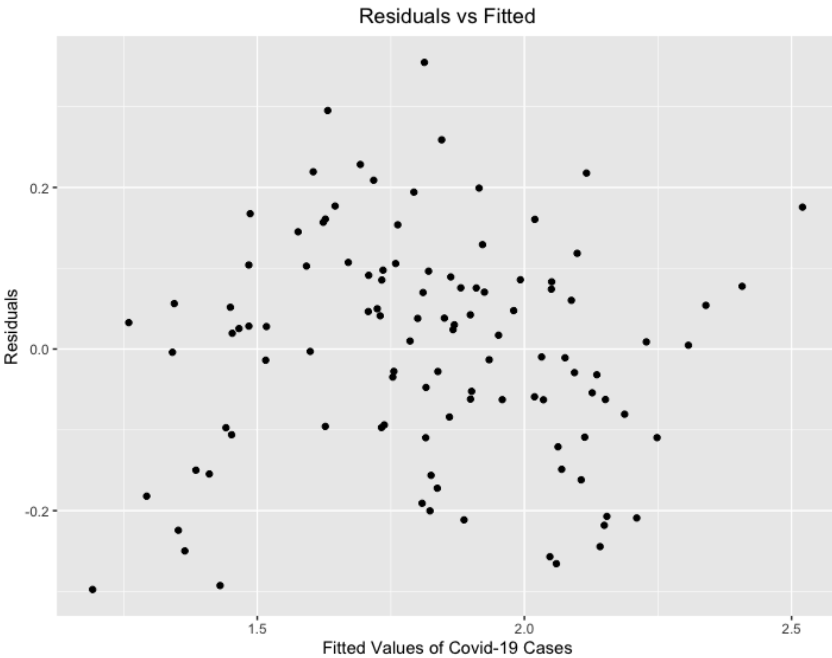

Figure 10. Check for homoscedasticity (equality of variance) in ANCOVA model predicting Covid-19 cases by week, county, VMT, and trips.

Figure 11. Summary of ANCOVA model predicting Covid-19 hospitalization by week, county, VMT(mi_norm), and trips (normal_short = short trips, normal_medium = medium trips, normal_long = long trips.).

```
Analysis of Variance Table

Response: log10(m_hosp)
              Df  Sum Sq Mean Sq F value    Pr(>F)    
week           1 0.16396 0.16396  9.3692 0.0028404 ** 
County         3 2.86273 0.95424 54.5276 < 2.2e-16 ***
mi_norm        1 0.28378 0.28378 16.2158 0.0001108 ***
normal_short   1 1.02386 1.02386 58.5057 1.345e-11 ***
normal_medium  1 0.00091 0.00091  0.0521 0.8198719    
normal_long    1 0.11662 0.11662  6.6640 0.0113033 *  
Residuals     99 1.73252 0.01750                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

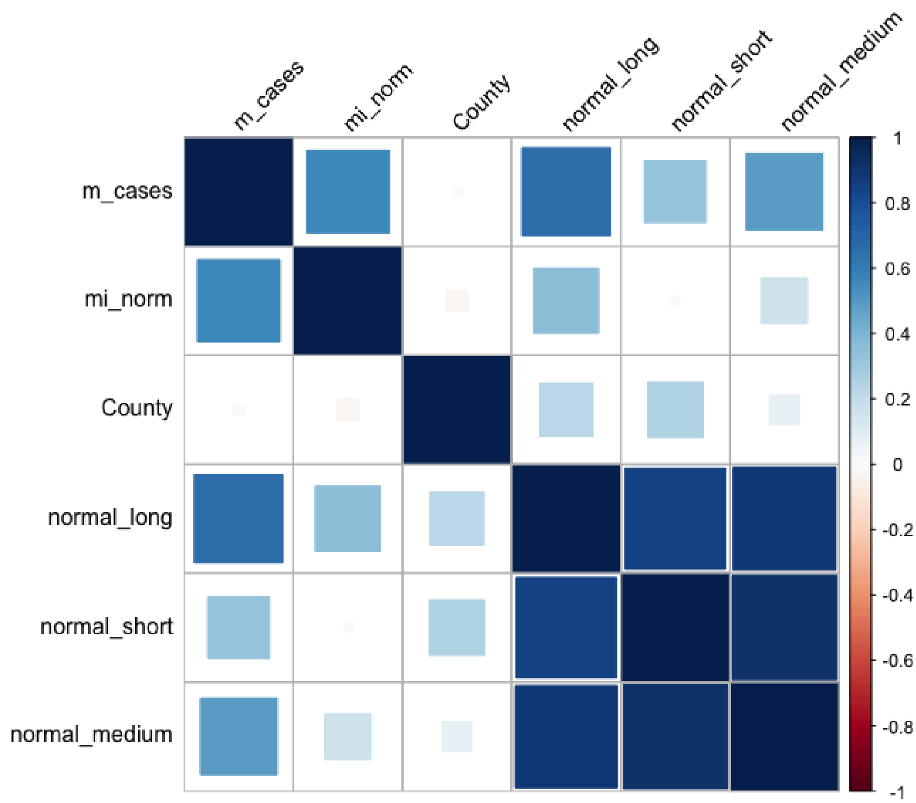

Figure 12. Correlation matrix for the covariates in ANCOVA model of Covid-19 hospitalizations predicted by week, county, VMT(mi_norm), and trips (normal_short = short trips, normal_medium = medium trips, normal_long = long trips.).

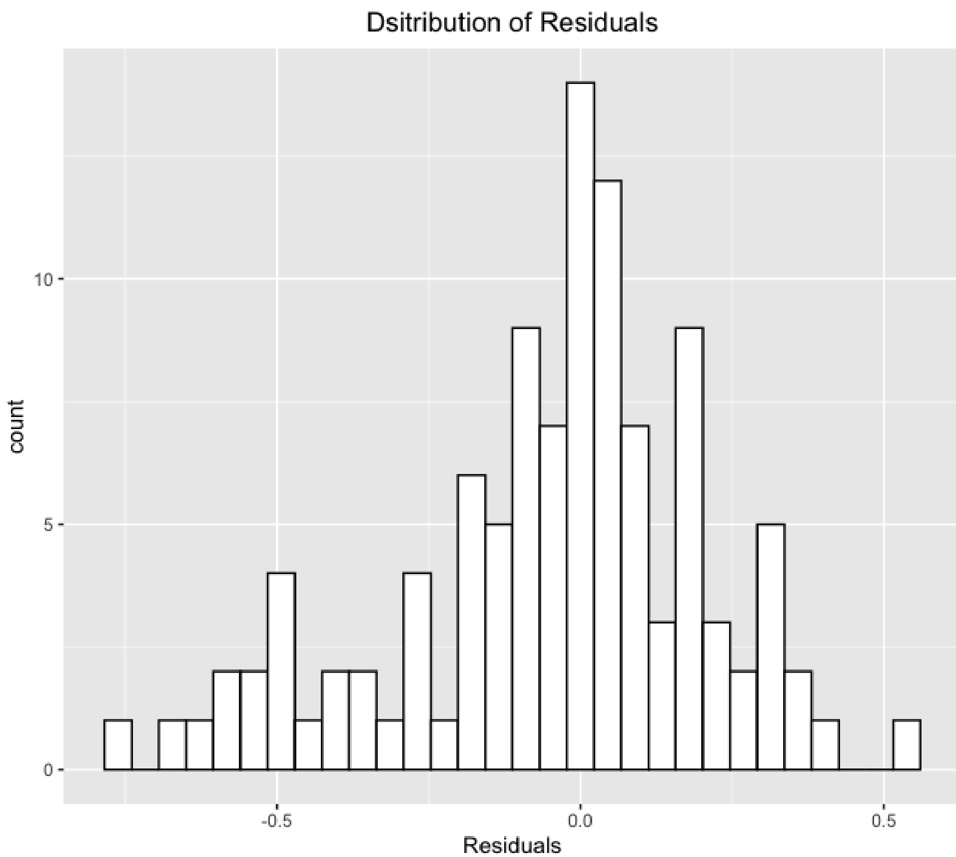

Figure 13. Distribution of residuals of ANCOVA model predicting Covid-19 hospitalizations by week, county, VMT, and trips.

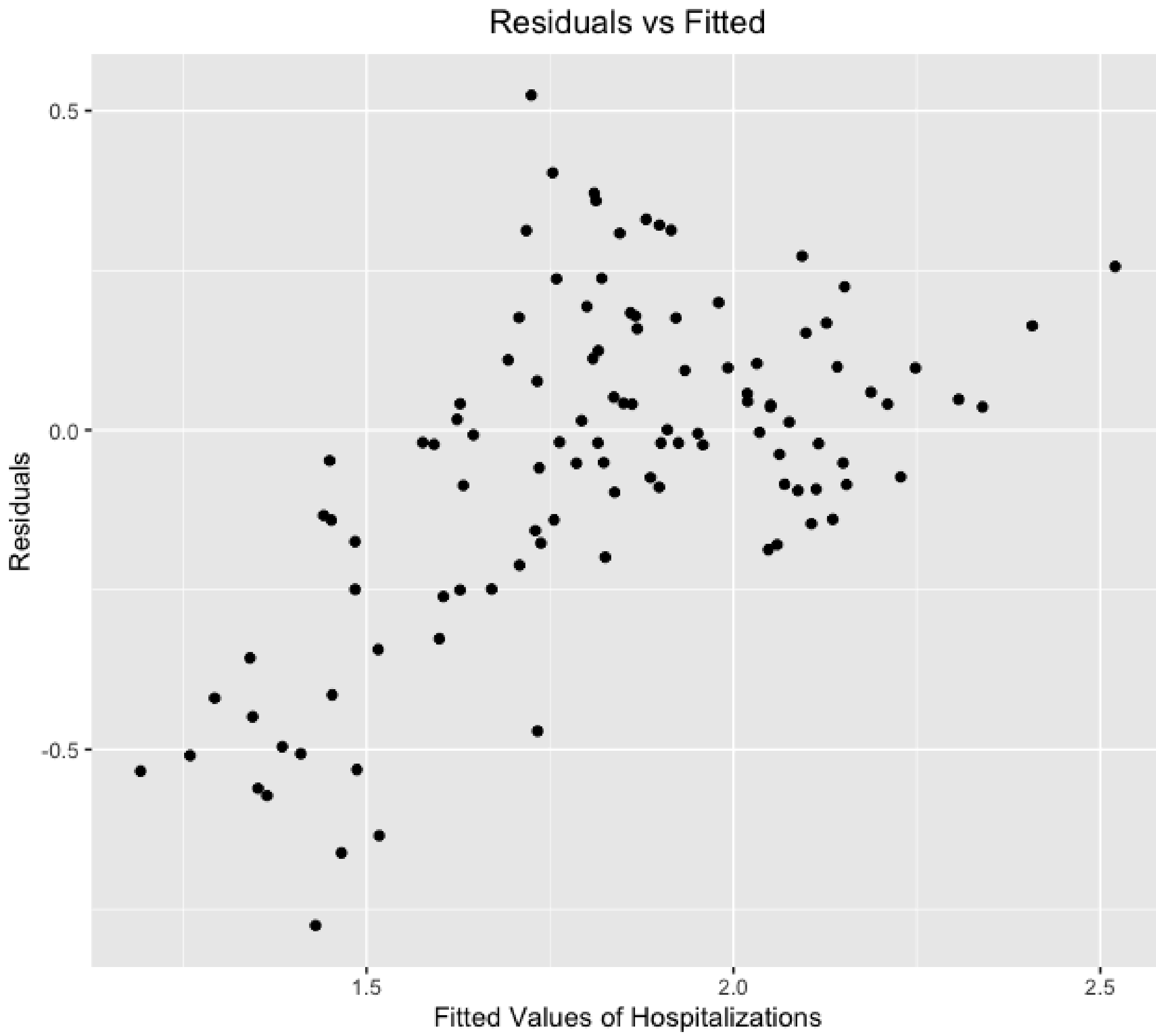

Figure 14. Check for homoscedasticity (equality of variance) in ANCOVA model predicting Covid-19 hospitalizations by week, county, VMT, and trips.

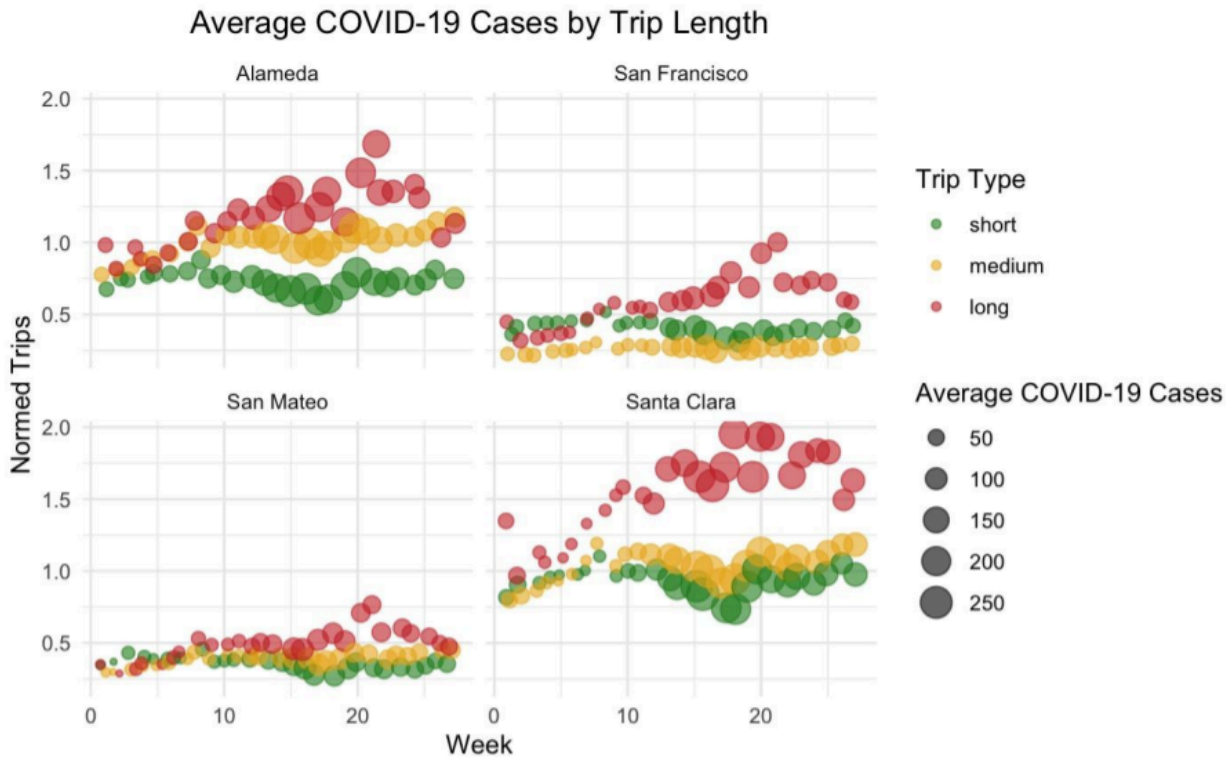

Figure 15. Average COVID-19 Cases by Trip Length in each county for April-September 2020. Trips are measured as weekly averages of (long distance (>50 mi), medium distance (10-50 mi, short distance(0-10) normed to January-February 2020 long distance trip counts.

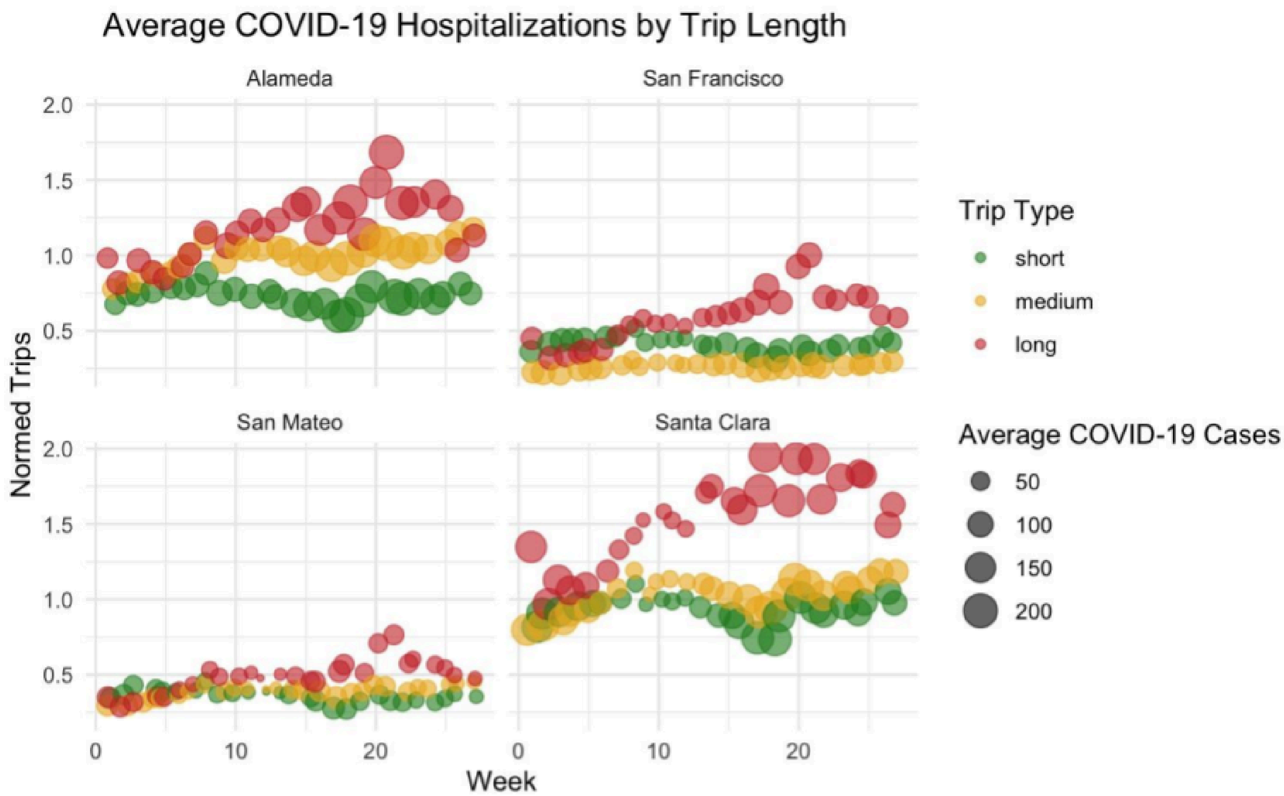

Figure 16. Average COVID-19 Hospitalizations by Trip Length in each county for April-September 2020. Trips are measured as weekly averages of (long distance (>50 mi), medium distance (10-50 mi, short distance(<10) normed to January-February 2020 long distance trip counts.