COVID-19 Vaccination Project

Adrian Cornejo and Victor Guillera

11/30/2021

Problem Statement

The goal of our study is to observe the impact between renter versus homeowner status (housing tenure) on COVID-19 vaccination rates per CA county as an indicator of SES. There are multiple factors behind housing tenure that may affect the overall status of SES and subsequent vaccination rates such as the number of individuals occupying a residency, mortgage status, household income, or generational wealth, and other demographic information. While renter and homeowner rates are not a perfect representation of SES, this comparison will help us elucidate if the financial status is a potential barrier to vaccination and an adequate indicator of SES per our given datasets.

We calculated the renter-to-homeowner ratio of every county in CA and compared the top 5 ratios to the bottom 5 ratios, tracking vaccine progress by percentage of fully vaccinated individuals over the age of 12 from January 2021 through September 2021. Two different datasets were used, one is of COVID-19 vaccination progress throughout CA provided by the CDPH following vaccine dosage by zip code from 1/5/21 to 9/14/21, and the other is demographic information in CA across all counties from the US census data for 2014 to 2019.

Methods

Data Sources - Years/Dates of Data:

COVID-19 Vaccination Status Dataset: This dataset supplied by the CDPH tracked the total number of persons partially and fully vaccinated with a COVID-19 vaccine over the age of 12 by CA county and zip code. The dataset contained weekly totals starting on January 5th, 2021 and ended on September 14th, 2021.

CA County Demographics Dataset: This dataset contained demographic information of every county in CA from 2021 census data. Variables included: population, population density, race/ethnicity, gender, median age, households, families, housing units, average family size, vacant homes, and renter and owner occupancies.

Data Cleaning/Methods Used:

COVID-19 Vaccination Status Dataset: We observed the overall dataset and variable types using the str() function and then select() to keep our variables of interest. Using the names() function we renamed variable columns to names more suitable for our team, from there we converted 'NA' variables to values of 0 with is is.na() function as a string subset. Any duplicates for rows were removed using distinct(). We filtered out values in the county column that were equal to zero by filter(county != "0"). All dates were input as.date, but we wanted to visualize multiple dates in each month as their name (i.e. Jan., Feb., etc.), using mutate() and a case_when() function with the help of str_detect() we converted each month taking rows containing the numeric value of the month to equal their respective month name.

CA County Demographics Dataset: We observed the overall dataset and variable types using the str() function and then select() to keep our variables of interest. Using the names() function we renamed variable columns to names more suitable for our team, from there we converted 'NA' variables to values of 0 with is is.na() function as a string subset. Any duplicates for rows were removed using distinct().

New Variable Calculations:

COVID-19 Vaccination Status Dataset: We summarized monthly vaccination data per month using group_by(county, month) and mutate() with sum() to add the variables county_total_pop_over12, county_fully_vaccinated, county_partially_vaccinated. Lastly, using filter() we kept the last date of each month to use for the cumulative measurement and visuzaliation in our graphs.

CA County Demographics Dataset: Four variables were created using the mutate() function; total_in_use_housing = total_housing_units - vacant_housing, renter_housing_prop = (renters/(total_in_use_housing)), owner_housing_prop = (owners/(total_in_use_housing)), renter_owner_ratio = renters/owners

Decisions Team Made About the Data:

We decided the best way to easily visualize our data to interpret in the scope of our problem statement was to compare the renter-to-owner ratios of the top five counties with the bottom five counties. The two datasets of COVID-19 vaccinations and demographics were combined using right_join() and then using select() we kept our desired variables. Using pivot_wider(), arrange(), and mutate_if(), we arranged our data in an ideal way to view in a table with kable().

Additional Wrangling Done for Visuals:

Originally, we split up the top five and bottom five counties to visualize in separate ggplot point graphs, but made the decision to combine the two to visualize and compare together.

COVID-19 Vaccination Status Dataset: This dataset supplied by the CDPH tracked the total number of persons partially and fully vaccinated with a COVID-19 vaccine over the age of 12 by CA county and zip code. The dataset contained weekly totals starting on January 5th, 2021 and ended on September 14th, 2021.

CA County Demographics Dataset: This dataset contained demographic information of every county in CA from 2021 census data. Variables included: population, population density, race/ethnicity, gender, median age, households, families, housing units, average family size, vacant homes, and renter and owner occupancies.

Data Cleaning/Methods Used:

‘

Results

Table 1: Summary of CA Housing Across All Counties in 2021

	Total in Use Housing	Renter/Housing Ratio	Owner/Housing Ratio	Renter/Owner Ratio
Min	497.00	0.2311765	0.3575537	0.3006887
1st Quartile	19040.75	0.3426035	0.5753113	0.5211578
Median	70284.50	0.3854496	0.6145504	0.6272347
Mean	216853.41	0.3833366	0.6166634	0.6472471
3rd Quartile	207711.50	0.4246887	0.6573965	0.7381916
Max	3241204.00	0.6424463	0.7688235	1.7967828

Table 1 provides a statistical breakdown of the total in-use housing to total housing, renter to in-use housing ratio, owner to in-use housing ratio, and the renter to owner ratio for the counties in California in 2021.

Table 2: Total Number of Renters and Owners in CA in 2021

	Renters	Owners
Total Count	5542127	7035371

Table 2 provides the total renter and owner households in all of California in 2021. There are more owner households than renter households overall.

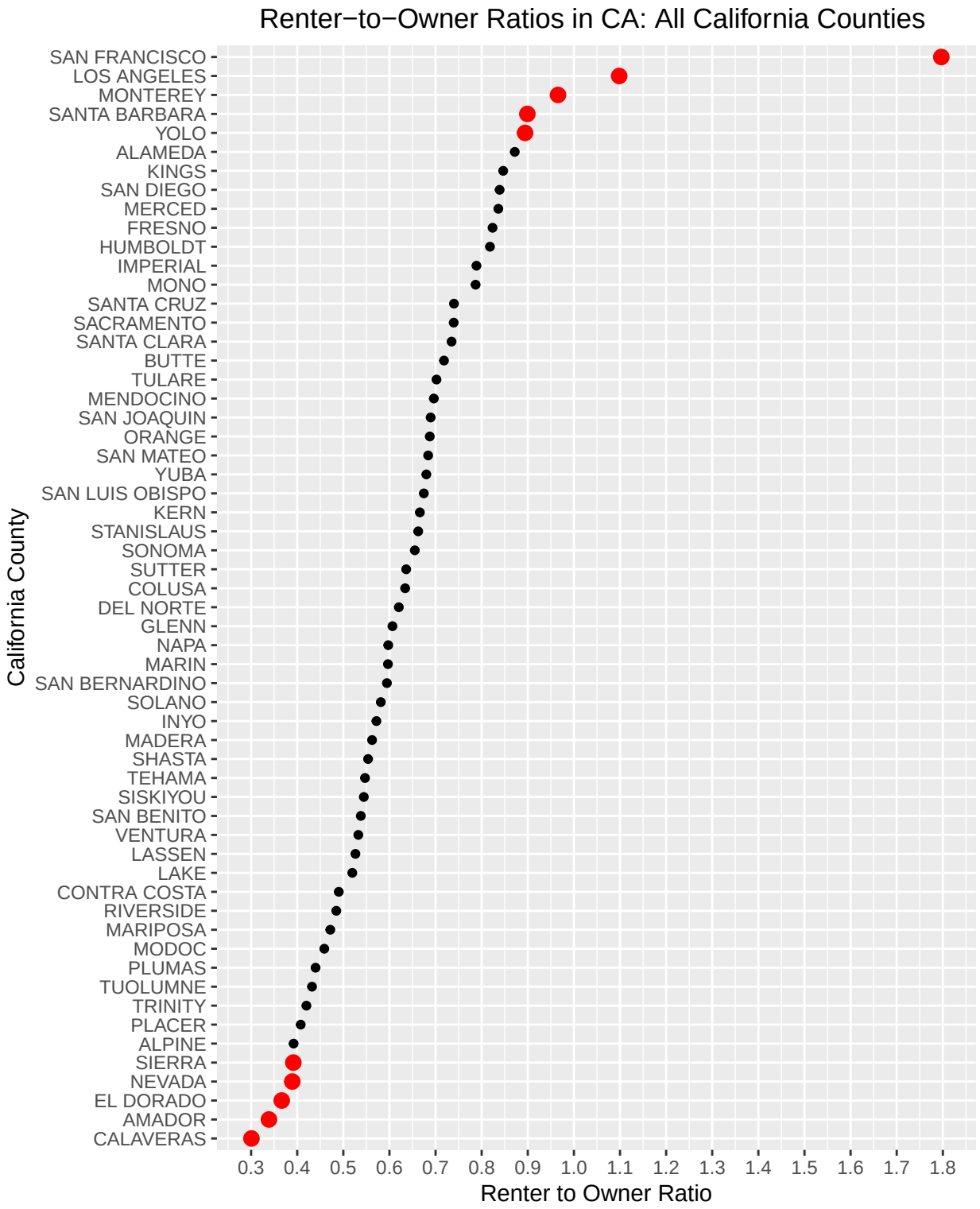


Figure 1: This graph displays the Renter-to-Owner Ratio for every single county in California. Highlighted in red are the top 5 and bottom 5 ratios. We will use their corresponding counties' data for further analysis.

Table 3: COVID-19 Vaccination Rate by Renter/Owner Ratio in Top 5 and Bottom 5 Counties in CA in 2021

County	Renter/Owner Ratio	Jan	Feb	March	April	May	June	July	Aug	Sep
SAN FRANCISCO	1.797	0.009	0.048	0.188	0.419	0.618	0.733	0.768	0.789	0.806
LOS ANGELES	1.098	0.008	0.048	0.152	0.318	0.477	0.576	0.622	0.654	0.682
MONTEREY	0.966	0.006	0.033	0.124	0.309	0.489	0.587	0.633	0.664	0.693
SANTA BARBARA	0.899	0.006	0.044	0.139	0.300	0.483	0.569	0.605	0.631	0.656
YOLO	0.894	0.010	0.051	0.177	0.356	0.497	0.595	0.631	0.654	0.676
SIERRA	0.392	0.006	0.082	0.253	0.365	0.276	0.295	0.272	0.333	0.363
NEVADA	0.389	0.005	0.035	0.163	0.311	0.434	0.501	0.529	0.549	0.569
EL DORADO	0.366	0.007	0.050	0.169	0.316	0.426	0.507	0.540	0.563	0.584
AMADOR	0.339	0.006	0.034	0.140	0.292	0.382	0.429	0.449	0.465	0.484
CALAVERAS	0.301	0.006	0.035	0.148	0.278	0.369	0.414	0.435	0.454	0.474

Note. Pink =Top 5 counties & Blue = Bottom 5 counties.

Table 3: This table shows the vaccination rate at the end of each month for the time period provided in the dataset for the top 5 and bottom 5 renter-to-owner ratios and their counties.

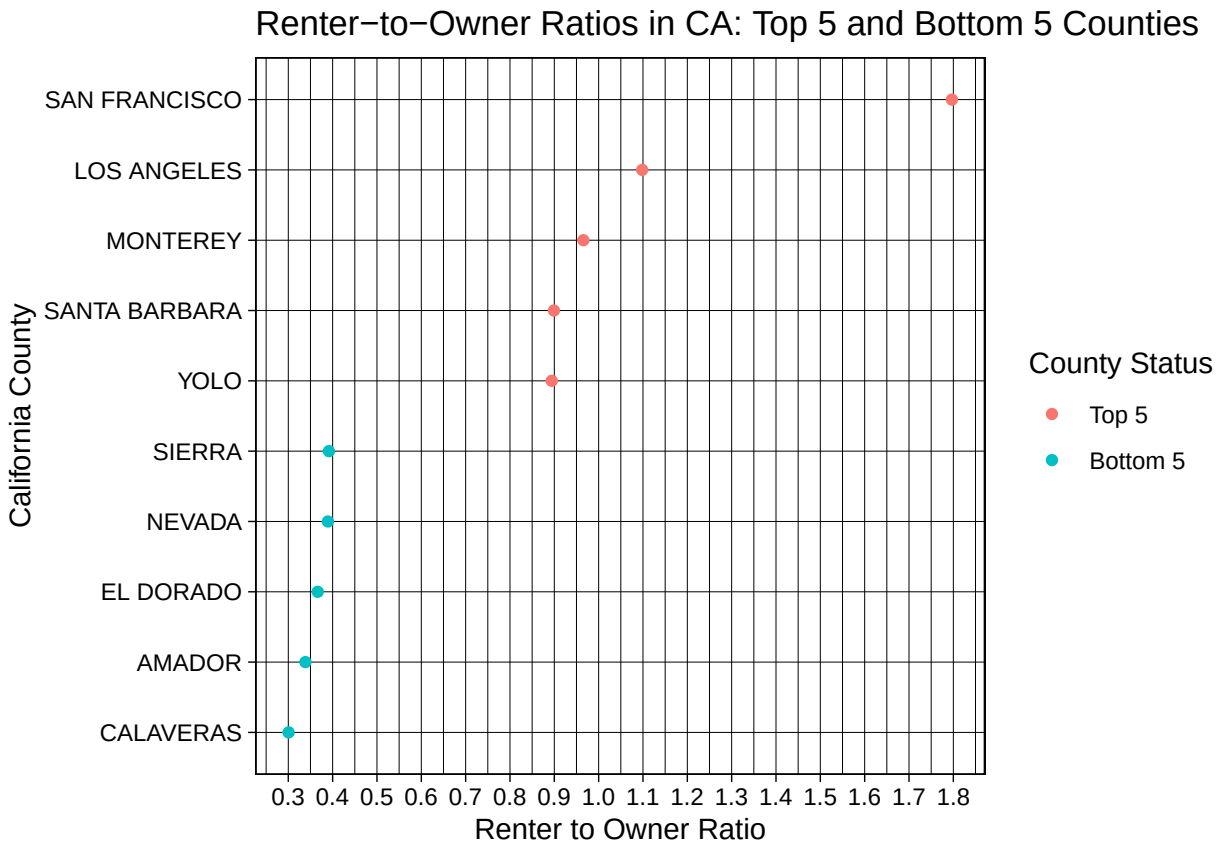


Figure 2: This graph provides a greater visual aid for the disparity of owning a home and the amount of renters for the top 5 and bottom 5 counties.

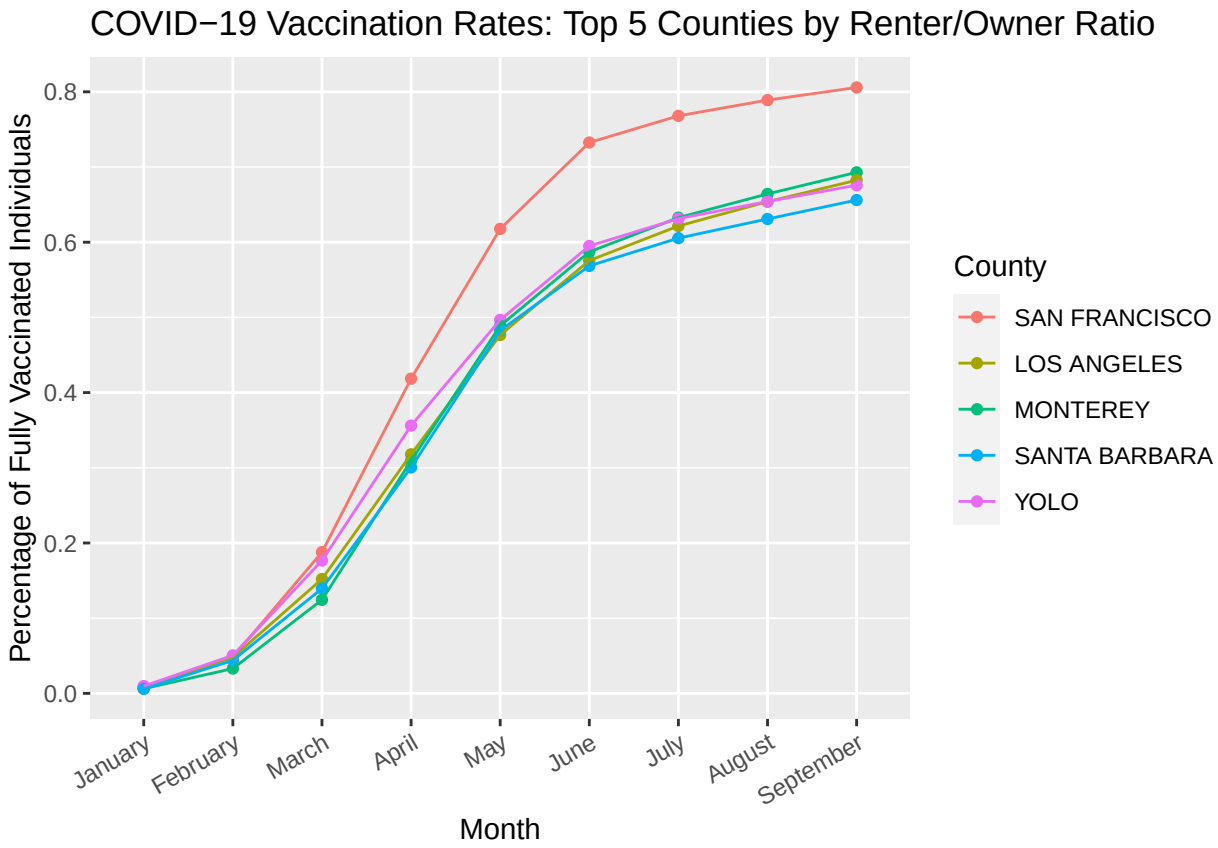


Figure 3: This graph shows the vaccination rates for the top 5 renter-to-owner ratio counties in California over January to September. San Francisco consistently had the highest percentage of fully vaccinated individuals starting in March and reached an astonishing 80% coverage of its "over 12" population. Other counties had a very similar pattern of coverage during this period.

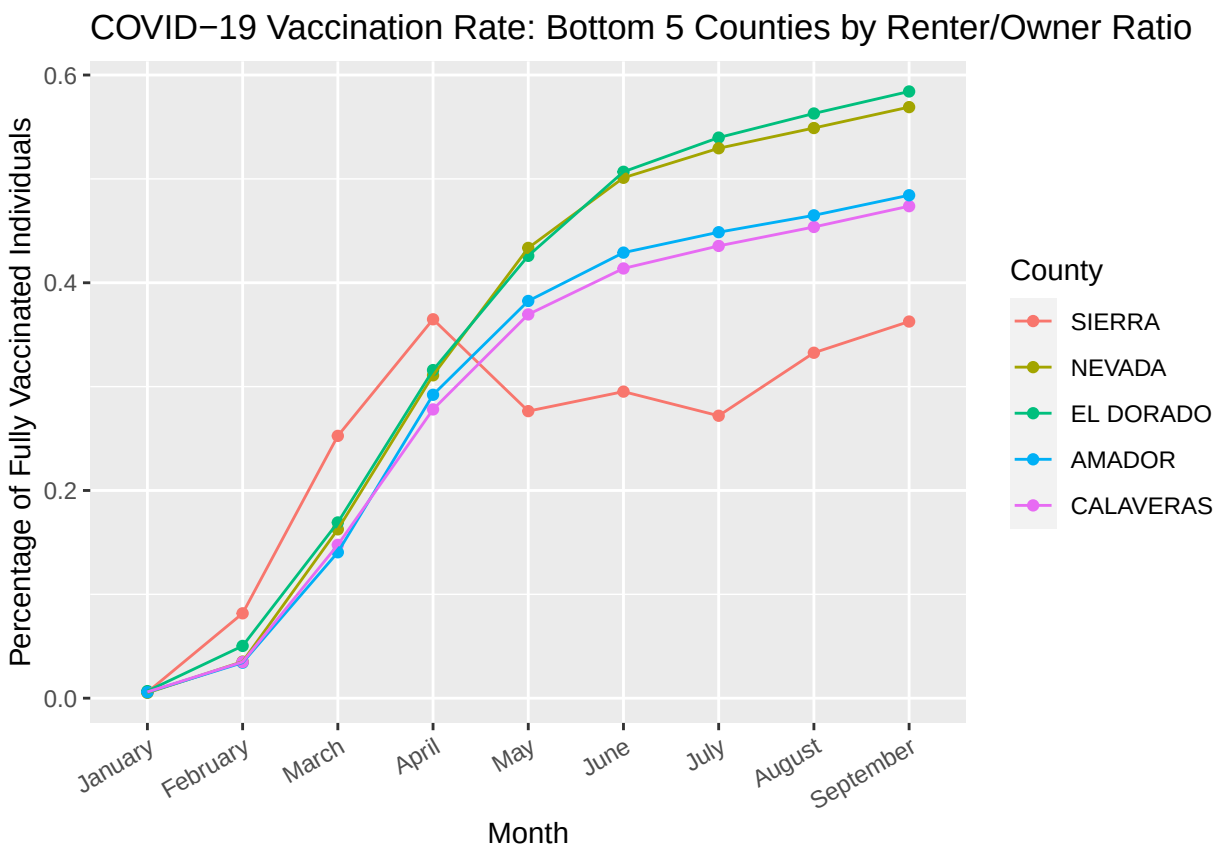


Figure 4: This graph shows the vaccination rates for the bottom 5 renter-to-owner ratio counties over January to September. The Sierra line looks very strange, as we lose coverage. This could potentially be due to low population density and movement out of the county perhaps. We can only speculate with the given data.

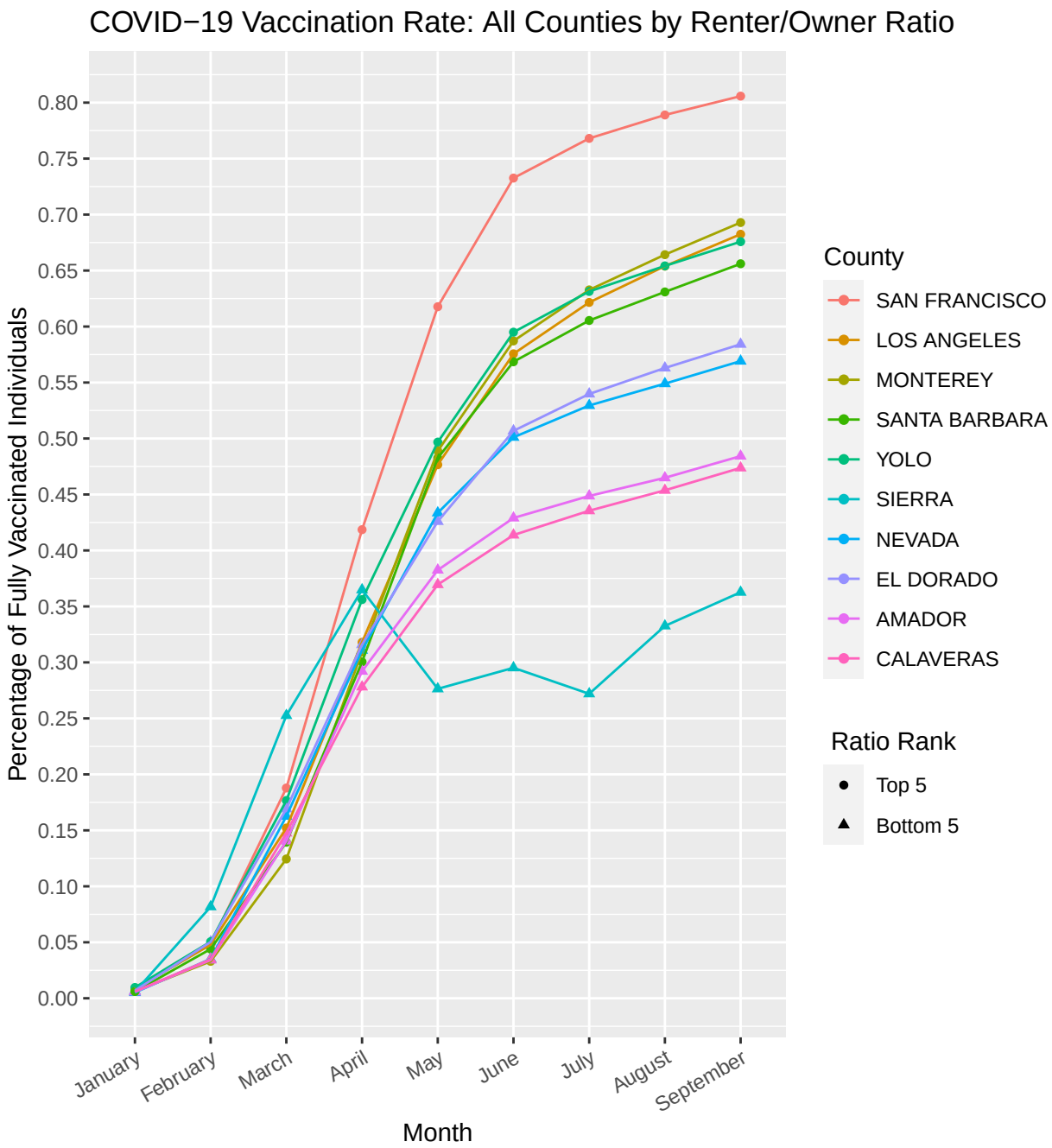


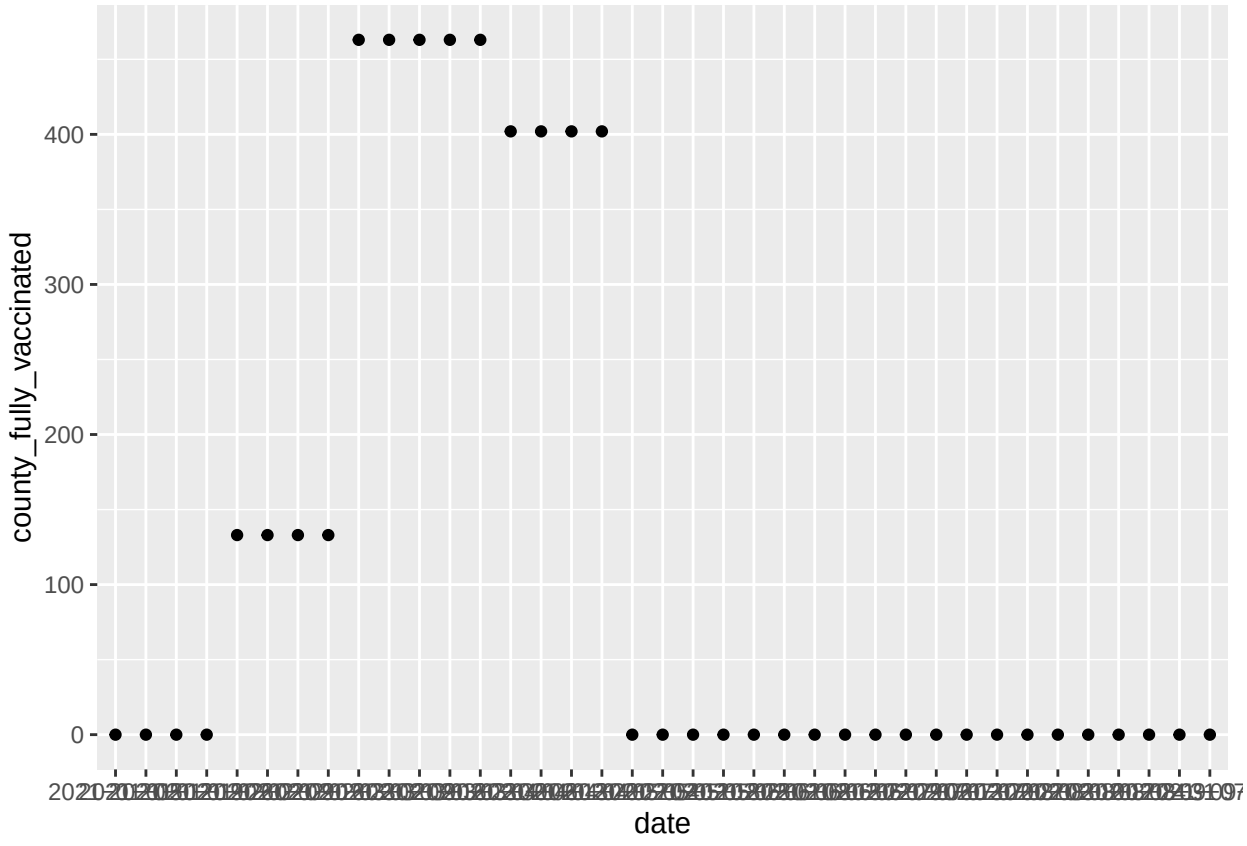
Figure 5: This graph combines the data from Figure 3 and Figure 4 for an easier comparison of vaccination rates over time for each of the top and bottom 5 counties.

Discussion

We see in the figures above, that overall, those counties with the highest renter-to- owner ratios also had the highest vaccination rates for the January-September time period compared to the bottom five renter-to-owner ratio counties. We start to see a divide between the groups after April during the initial vaccine rollout. As the data shows, it appears as if being in a county with more home owners is more associated with being unvaccinated.

This goes against our hypothesis that homeowner/renter rates would be a good predictor for SES and therefore higher vaccination rates. Owning versus renting a home involves many SES factors in and of itself, so we can see how confounding can easily skew our results in this case. Renting in big cities like Los Angeles and San Francisco can be very expensive, and it can be interpreted as showing high SES in some cases for certain counties. It is also important to note the political divide in the country and its impact on vaccinations. Bigger cities like Los Angeles and San Francisco tend to be liberal and more pro vaccine compared to more conservative rural counties (like the bottom 5, which are all in the Sierra Nevada region).

```
##   zip_code
## 1    96118
## 2    95910
## 3    96126
## 4    95936
## 5    96125
## 6    96124
## 7    95944
```




```
## Warning in rm(covidvaxamador): object 'covidvaxamador' not found
```

```
##    zip_code
## 1     95669
## 2     95699
```

```
## 3     95665
## 4     95640
## 5     95666
## 6     95601
## 7     95646
## 8     95689
## 9     95675
## 10    95629
## 11    95642
## 12    95685
```

```
## [1] 185
```