Analysis of Covid-19 Vaccination Rates in California One Month After Open Access

Callum Weinberg: 220C Final Project

June 10th 2021

Abstract

California Covid-19 vaccination count data was studied for cumulative counts through May 18th, 2021. May 18th was chosen as it is roughly a month after open eligibility in California. Data was accessed at the Zip Code Tabulation Area (ZCTAs). A quasi-poison regression model was fit to determine if there was a statistically significant relationship between vaccination rates up to that date and demographic characteristics of the ZCTAs. More educated and older ZCTAs had higher vaccination rates. ZCTAs with higher make-ups of Asian, Hispanic, and Female as percentages of the population also had higher vaccination rates. More work is needed to determine if there is an advantage in using the ZCTA level versus county level for analyzing the factors relating to vaccination uptake.

Introduction

On December 14th, 2021, individuals in California began to receive one of the three Covid-19 vaccines available approved for use in the United States.[10] As of April 15th 2021, all individuals in California over the age of 16 were reportedly able to sign up for their first Covid-19 vaccine.[13] This news was broadcast and published weeks in advance of the date.[13] Given the saliency of the Covid-19 pandemic, it seems reasonable to assume most adults in the state were aware they could sign up for their first vaccination as of that date. 12-15 year old individuals were able to sign up for their first vaccination on May 13th.[15]

Estimates vary, but according to experts, "about 85% of Americans will need to be vaccinated to bring the COVID-19 pandemic under control."[14] In order to improve vaccine uptake, it is important to understand which factors relate to an individuals decision to choose to vaccinate themselves or not.[1] Geographic factors have been considered. Demographic factors such as race, age, and gender are also important for understanding how vaccination uptake relates to different groups of Californians.

Chernyavskiy, Richardson, and Ratcliffe from the University of Virginia have very recently (May 31st) released a preprint paper which analyzes vaccine uptake by U.S. counties. Their analysis covers February through May and focused on individuals with one or more vaccination shots. They found that vaccination rates were higher in counties that were older, more educated, and had a higher percentage of racial minorities. They used the 2015 American Community survey as a data source for demographic variables. Their paper also includes a time-analysis component which indicated slowing uptake rates over the course of their time frame. Their analysis also analyzes how a geographic regions uptake rate compares to the expectation given its demgraphics (i.e. isolating geography from the other factors).[5]

Beleche et al. also studied vaccination uptake hesitancy, for January through March 2021.[16] They found declines in vaccine uptake hesitancy for young individuals (below the age of 24) and black individuals over the period of study.[16] Their vaccine data came from a national survey, rather than

[1] Note that the word "relate" is chosen intentionally here. This report makes no claims that any of these factors are causal in any way. Any claims of causality, at a minimum, would likely require an experimental design with a control group. This report focuses on data that effectively comes from a study.

state information. They did use the American Community Survey for demographic information. The focus of their study was regarding changes in hesitancy over the period. However, their results also indicate absolute levels of uncertainly. Table 1 in their report indicates Asian Americans and Americans over the age of 65 had lower levels of hesitancy by March. Their analysis made use of a logistic regression model to predict percentages of vaccine hesitancy. Their report also includes some analysis of county level vaccination uptake rates.[16] Their work (see Figure 2) suggests that as of March 17-19th, hesitancy was likely lower in California than parts of the South, Midwest, and Upper Mountain West. Additionally, there was possibly lower hesitancy in the Bay Area in California and higher hesitancy in Northern California and rural counties in Southern California.[16]

Purpose

The purpose of this report is to better understand what demographic factors relate to the choice of individuals to receive a Covid-19 vaccine in the state of California. Specifically, a regression model is developed and analyzed in order to determine which factors relate to cumulative vaccine uptake as of May 18th, 2021 at the Zip Code Tabulation Area geographic level. The justification for using the May 18th data is this is roughly one month after open eligibility in California, and seems like a reasonable amount of time for most individuals who wanted to be vaccinated to get at least their first shot.

This report is written in June, 2021 and only applies to Covid-19 vaccination uptake prior to May 18th, 2021 in the state of California. There has been some research regarding vaccination uptake at the county level.[5][16] This analysis focuses on the relationship between demographics and vaccine uptake at the the Zip-Code Tabulation area level. The primary benefit of Zip Code level analysis is that Zip Codes generally represent a smaller geographic areas than counties, and therefore should be more representative of each individual within a region than the county level. While there has been some data analysis at the Zip Code level in the news,[9][14] to the best of my knowledge there has been no published regression analysis on California vaccine uptake counts at the Zip Code level using data from the end of May or beginning of June.[2] Given the importance of understanding the Covid-19 pandemic, many researchers are studying similar topics, asking similar questions, and performing similar (but slightly different) analyses.

Data, Assumptions, and Methods

Data

Zip Code Tabulation Area

An integral part of this report is the choice of performing the analysis at the Zip Code Tabulation Area (ZCTAs). ZCTAs are used by the U.S. Census Bureau, while Zip Codes are used by the U.S. Postal Service. The U.S. Census Bureau created ZCTAs to be similar to Zip Codes but used for census data, although they differ occasionally. The U.S. Census Bureau notes "In most instances the ZCTA code is the same as the ZIP Code for an area."[3] Both the vaccine uptake data and the demographic data was retrieved at the ZCTA geographic level.

ZCTA's were chosen because they provide a more granular geographic level than county, in terms of population. Figure 1 compares the distribution of populations by ZCTA versus County for California. Table 1 includes population summary statistics of by ZCTA and County for California.

[2]There have been some Zip Code level analyses in other locations and time frame. Woody et al. performed a spatial analysis of vaccinations in Austin, Texas using data from Mid April [17] for example.

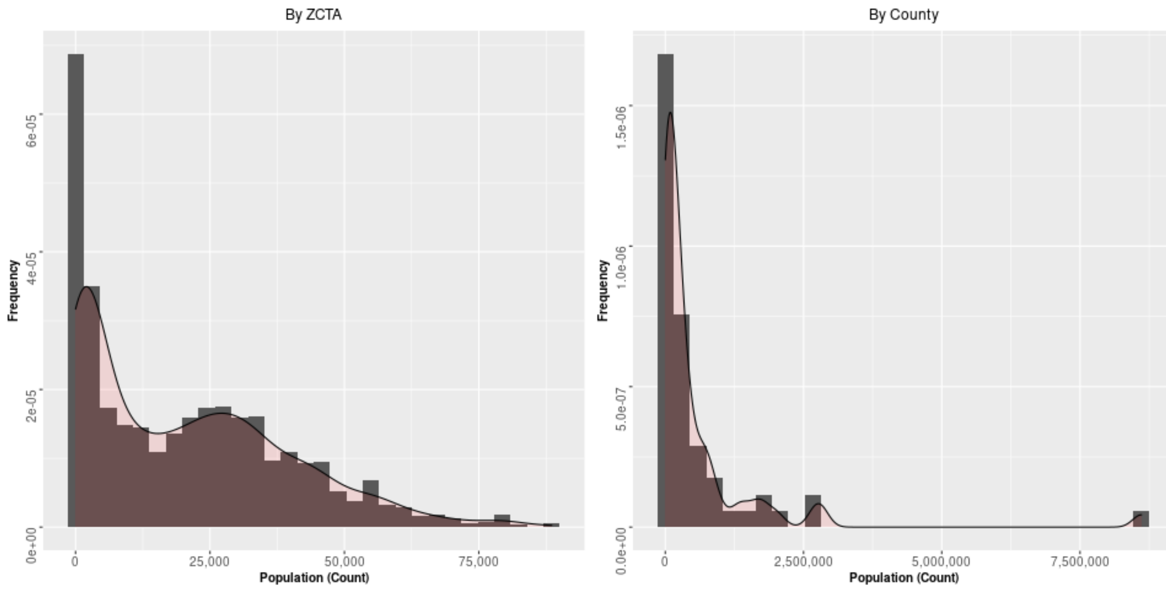

Figure 1: California Populations by ZCTA and by County. Population is the 12 Year-Old Plus Population, as reported with the Vaccination Uptake data from the California Open Data Portal.

Geographic Area	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
ZCTA	9	1,845	16,321	19,842	32,449	88,557
County	812	42,294	155,470	573,698	566,455	8,611,498

Table 1: Summary Statistics for California 12 Year-Old Plus Population, by ZCTA and County

ZCTA's have smaller populations, as illustrated by Figure 1 and summarized in Table 1. The median ZCTA 12 year-old and over population is 16,321 for ZCTA's but is 155,470 for Counties. The benefit of using ZCTA's for analysis is that each geographic unit represents fewer individuals. Therefore, any relationship between demographic factors and vaccine uptake rates is more likely due to the demographic accurately representing the demographic in the given geographic area and less likely due to an averaged result of the demographic factor. For example, say County 1 (not a real ZCTA in California) is made up of rich people in one half of the county and poor people in the other half. The average wealth will average out to moderate wealth. The extent of such averaging problems will be less influential at the ZCTA level given the counties are now broken up into smaller geographic units (although it still will happen). Analyzing data at smaller geographic levels is especially and improvement over the largest geographic areas, such as Los Angeles County.

ZCTA is the smallest geographic area for which vaccination uptake data was found. Note that ZCTA is not guaranteed by definition to be smaller than County,[12] it just happens to be the case for California, based on the above evidence. The ideal geographic level would likely be census block, given it is the smallest geographic unit for which the U.S. Census compiles data.[12] ZCTAs do not necessarily fall within county or even state lines[3], while census blocks do.[12]

California Vaccine Data

The vaccine uptake data comes from the California Open Data Portal as part of the CA.gov network of websites.[11] The portal appears to state that vaccine uptake data comes from the California Immunization Registry. [11] The data dictionary (available at same source) provides information on the vaccine count data. "Fully Vaccinated" and "Partially Vaccinated" counts are provided. Similar to the

[3]ZCTAs can be mapped to Counties using a relationship file available from the U.S. Census, but it was not necessary for this analysis.[2]

analysis by Chernyavskiy et al.,[5] this study will make use of the "Population with 1+ Dose" variable to answer the research question. This can be calculated by adding "Fully Vaccinated" and "Partially Vaccinated" counts.

The Open Data Portal data also includes the Vaccine Equity Metric, which is a metric derived from the California Healthy Places Index (HPI) produced Public Health Alliance of Southern California. The Vaccine Equity Metric places ZCTAs into four quartiles, with one indicating the lowest overall community health levels and four indicating the highest community health levels.[11][6]

The Portal also cites the American Community Survey's 2015-2019 5-Year data as a source, and the data dictionary clarifies that the 12 year-old plus population comes from this source. The following note is included for populations: "Please note that since the 12+ eligible population group involves taking portions from ACS age ranges (i.e., 4/5ths of the 10-14 age range and 1/3 of the 15-17 age range), the population for some zip codes will not be a whole number."[11] 19 ZCTAs were listed in the data but did not include population or vaccine count data. The missing observations include the following note: "Information redacted in accordance with CA state privacy requirements.

American Community Survey Demographic Data

The demographic data for this report comes from the U.S. Census Bureau's American Community Survey (ACS). The ACS is a survey that the U.S. Census conducts annually, and "collects and produces information on social, economic, housing, and demographic characteristics about our nation's population every year."[1] As of 2017, 3.5 million households were surveyed each year.[1] The ACS has become the substitute for the "long" form census, and it asks more detailed questions of households than the 10 year census.[1] For ZCTAs, the data from the ACS is only available as a five-year estimate.[1] The data used in this report is from the 2015 through 2019 estimates.[4] Since the ACS is a survey (and not a census), statistics from the sample are reported with a margin of error.[4]

The following variables were retrieved from the ACS survey 2015 through 2019 estimates, which are available online.[4] Note that they are five year estimates, although they are listed in the 2019 ACS data sets.

The following data came from table DP05 from the 2019 ACS.[4] Age was retrieved as a percent of the population per ZCTA, available in five-year increments up to age 84 (85 and over is its own category). Median age is also considered. Percent of population that is male or female is retrieved. Percent of the population belonging to a certain race is retrieved (White, Black, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander). Additionally, percentage of the population that is Hispanic or Latino is retrieved.

The following data came from table S1501 (Educational Attainment) from the 2019 ACS.[4] For population over 25 years of age, "Less than 9th Grade", "9th to 12th grade, no diploma", "High school graduate (includes equivalency)", "Some college, no degree", "Associate's degree", "Bachelor's degree", and "Graduate or professional degree."

For percent of population living below the poverty threshold, the following variable was chosen: "Percent below poverty level, from Population for whom poverty status is determined." This is variable $S1701\_C03\_001E$ from table S1701. Despite the table being called "Poverty Status in the Past 12 Months", the website indicates that this is also a 5 year estimate (2015-2019). Finally, median income by ZCTA over the past 12 months, adjust for inflation in 2019 dollars, was also considered. This data came from S1901. Similar to poverty, it is assumed to be a 5-Year Estimate. This variable had a non-trivial number of missing Zip Codes that were present in the vaccination data (60).

Assumptions

Assuming that individuals have had the opportunity to get at least a first vaccine is not integral to the analysis portion of this report, but rather is driving the choice of using data from May, 18th. No claims of causality are made in this report, especially regarding this point. Clearly not everyone who wants a vaccine has been able to get one. There is reporting on non-trivial amounts of individuals who are not-vaccine hesitant but have not been vaccinated. There is some evidence some individuals are too busy to make time for vaccination, even if they are open to getting one. Convenience and transportation appear to be issues as well.[8]

Similar to [5], the vaccination uptake data used for analysis will be the count or percentage of individuals who have had at least one shot. This includes individuals who are fully or partially vaccinated. Open registration for vaccination was roughly 5 weeks before the data was collected, which is likely not enough time for many people who were vaccinated after open registration to be fully vaccinated (most vaccinations have been Moderna or Pfizer, which have either four or three weeks in between shots). Therefore, including partially vaccinated individuals makes sense for approaching the issue of vaccine interest a month after open access.

The population per zip code data from the California Open Access data[11] includes 12-15 year olds. It is not likely many 12 to 15 year olds were vaccinated as of May 18th. It is assumed this does not cause any issues in comparing across ZCTAs.

The ACS data chosen has Hispanic as a separate group, not part of the racial subgroups. Some studies and data sets separate "White Hispanic" and "White Non-Hispanic" (and possibly other Hispanic-Racial demographic subsets as well). For this analysis the simpler metric of "Hispanic" or "Not Hispanic" is chosen, with the clarification that this is an ethnic, not racial, categorization.

Methods

Data Processing

For the vaccination data, the 19 ZCTAs with missing vaccination information were removed from the data set.

The age variables were combined into to the following five categories: Under 15, 15 to 24, 25 to 44, 45 to 64, and 65 and Over. Having 65 and Over is something over researches have done[5]. The Under 15 group represents individuals were mostly unable to get vaccinated for this period. 15 to 24 was chosen to represent young adults, and the remaining 25 to 64 was split in half.

For educational attainment, the following four categories were chosen: Less Than High School Diploma ("Less than 9th Grade" and "9th to 12th grade, no diploma"), High School Graduate, Some College or Associates Degree ("Some college, no degree" and "Associate's degree"), Bachelor's Degree, Graduate or Professional Degree. High School Graduate includes "Equivalency", as reported by the ACS. Over 25 population and over 25 education values were chosen for simplicity, as the ACS reports over 25 and under 25 educational attainment. The percentages add to 100 percent for the California ZCTA data.

The ACS and Vaccination data were combined in R. After merging and excluding ZCTAs without vaccination counts, most of the ACS data was fully present. The exception is median income, which is missing about 60 ZCTAs. Four other ZTCAs had missing information.

The population data presents an issue, as 64 of the ZCTAs have a smaller population (which comes indirectly from the ACS) than the number of individuals either fully or partially vaccinated. 11 of these ZCTAs have over twice as many individuals vaccinated as population. For analysis using percentages, any percentage over 1 is set to one. The regression model tries to work around the issue. The issue is discussed further in the analysis section.

Statistical Code

RStudio was used to clean, visualize, and map the data. It was also used to fit regression models. R code to create ZCTA level maps was based on example code provided by Stack Overflow user diegheman, provided in this stack overflow post.[7]

Analysis

Vaccination Rates and Demographic Data

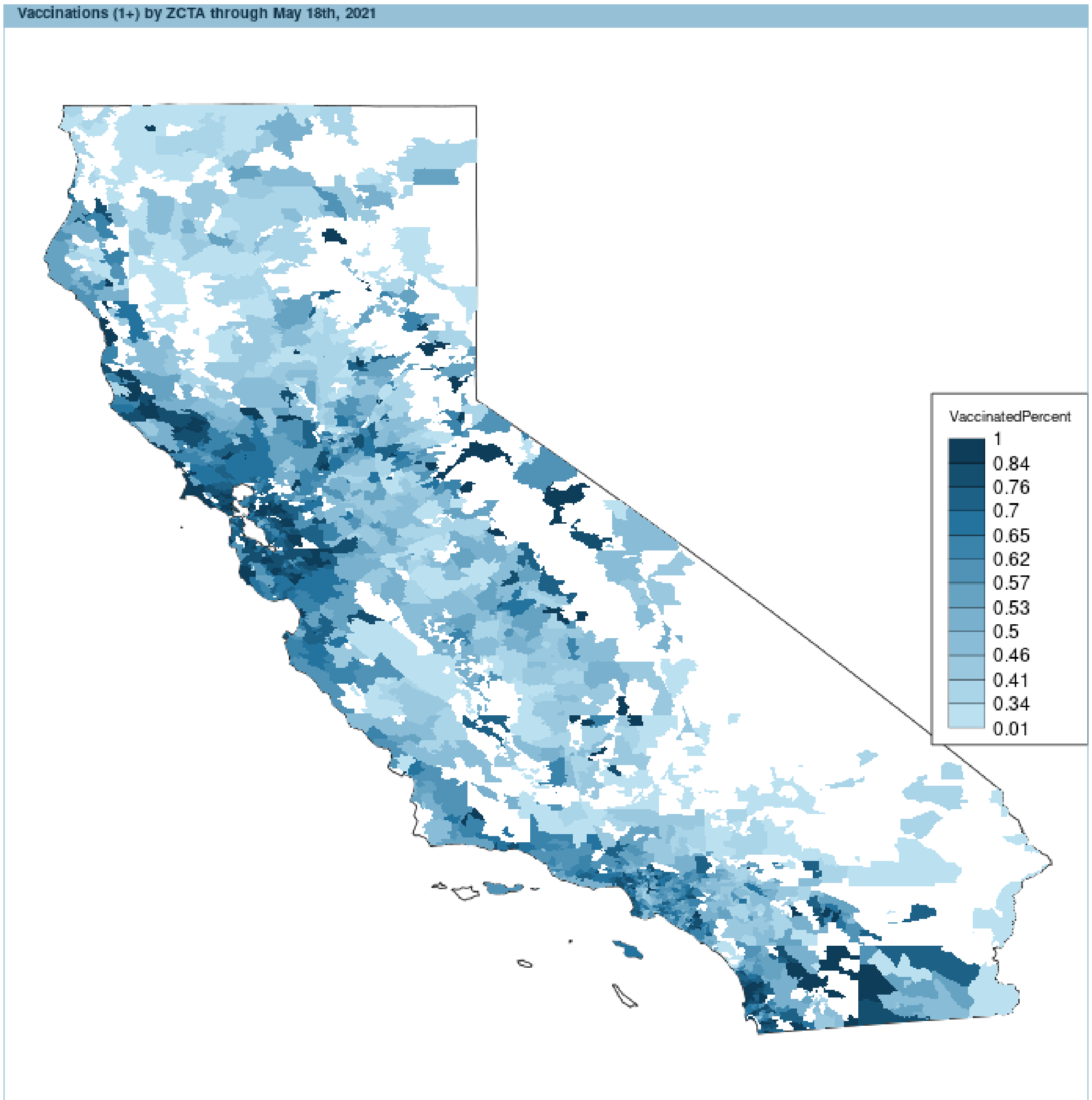

Figure 2: Vaccinations rates (Partially and Fully Vaccinated) for Covid-19 vaccine through May 18th, 2021. ZCTAs with percentages over 1 are set to 1 for this map.

The cumulative vaccination rate (partially and fully vaccinated Californians, as of May 18th 2021) is shown in Figure 2. White areas indicates no ZCTAs (unlike counties, not all geographic areas in the U.S. are covered by ZCTAs). Typically, these are low population areas - In California's case, often deserts, National Forests, and National Parks. Northern California, parts of the Central Valley, and areas north and east of Los Angeles tend to have lower vaccination rates. The Bay Area, parts of Los Angeles, San Diego, Imperial County (southern border), parts of the northern coast, and a few ZCTAs along the Sierra Nevada tend to hand higher vaccination rates. The map makes clear the advantage of analyzing the data as compared to a county level analysis (See [14], for example). A county level analysis averages all of Los Angeles together, and as the LA Times interactive shows, the result is a county that is somewhere in the middle with regards to vaccination rates.[14] There is clearly variation within the county, with central ZCTAs having lower vaccination rates and the eastern and north ZCTAs of the county having higher rates.

Variable	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Vaccination (1+) %	0.01	0.46	0.57	0.60	0.70	6.33*
Vaccination (1+) Count	27	1,056	9,984	12,092	19,813	60,792

Table 2: Summary Statistics for Vaccination Rate and Vaccination County (1+), by ZCTA. As noted in the data section, 64 ZCTAs have vaccination counts greater than the population, 11 of which are more than twice as large.

Table 2 and Figure 3 illustrate the distribution of vaccination rates. The ZCTA level-distribution of vaccination rates is almost symmetric, and looks like it could almost be normally distributed. The one issue is the over representation of ZCTAs with a vaccination rates with value 1[4]. The data otherwise meets rough expectations for May 18th, with the median value being .57 and the mean value being .6 (as a reminder, this is the 12+ population).

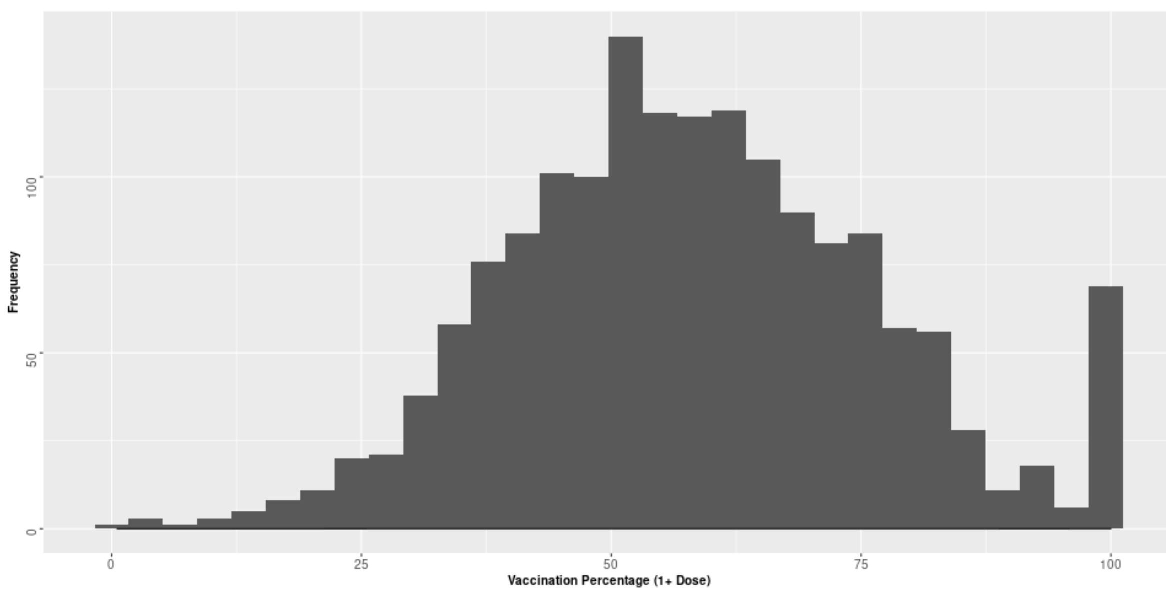

Figure 3: Vaccinations rates Histogram. Note that percentages over 1 have been set to 1 for this visual.

The covariates are mostly percentages (Race, Ethnicity, Age Group, Female, Percent in Poverty). Median Income is measured in 2019 inflation-adjusted dollars. However, Median Income and Percent in Poverty are strongly correlated (-.66 correlation). Median Income was excluded from the model given potential collinearity, and that median income had 60 missing ZCTAs. Due to over specification, at least one race and one education percentage needed to initially be excluded from the model. Preliminary analysis (see "Exploratory Data Analysis.Rmd") suggested that of the race categories that made up a significant percentage of some ZCTAs, percentage white was the least informative. Less than high-school education was the least informative of the education percentages.

However, education as a whole appears highly informative based on preliminary analysis. See Figure 4.

[4]This is an issue from a statistical modeling problem. Assuming it accurately represents the state of vaccination in California, it is subjectively great for public health that there are counties with 100% vaccination rates.

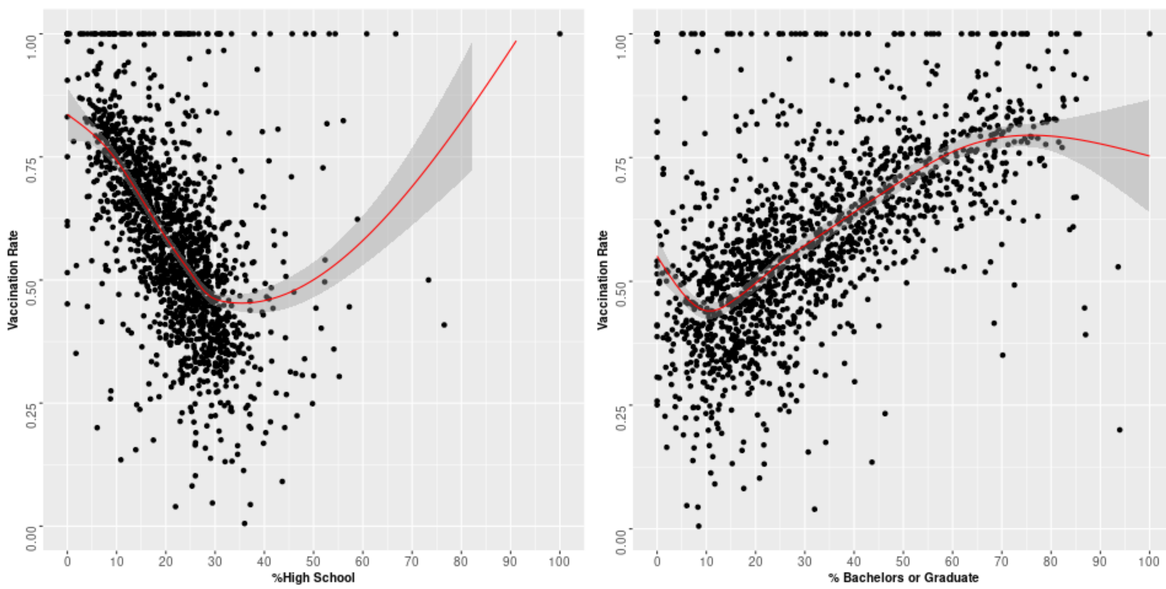

Figure 4: Left: scatterplot of % of individuals in ZCTA only having completed high school vs. vaccination rate. Right: scatterplot of % of individuals in ZCTA having received bachelors, graduate, or professional degree vs. vaccination rate.

Based on Figure 4, there is a clear relationship between education attainment level and vaccination rate, for a given ZCTA. The model in the next section will seek to determine if this relationship is evident when controlling for other variables.

Poisson Regression Model

Given the response variable is either a count variable or a percentage (depending on how it is recorded), a generalized linear regression model is appropriate. Specifically, either the Poisson class of models or the binomial class of models are possible for the given response data. In this report, a Poisson approach is chosen. There are a few justifications for this decision. First, the vaccination data appears to be actual counts, not estimates (see the Data section for a more in depth discussion). The population data for each ZCTA comes (indirectly) from the ACS, which is a survey. A binomial regression model would assume the number of trials (population of each ZCTA) is known, which is a strong assumption. There is clear evidence it would be a problematic assumption, given a few ZCTAs have a higher vaccination count than population (discussed in Conclusion). Therefore a Poisson model will be used.

For the Poisson regression, it is still necessary to account for ZCTA population size. As Figure 1 shows, there is significant spread in the ZCTA population sizes, although they are much smaller than counties for the most part. Therefore the regression model will include an offset. Model 1 below indicates what the offset is mathematically (effectively it accounts for ZCTA population size). It is especially useful in this case, given the ZCTA's which have higher vaccination counts than populations can still be modeled.

$$\text{Model 1: } ln(\frac{\lambda_i}{N_i}) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

$$\text{Model 1 with offset: } ln(\lambda_i) = ln(N_i) + \beta_0 + \sum_{j=1}^{p} \beta_j x_j$$

Note that both models are mathematically equivalent, and this is a statistical method that is possible the log link is chosen for a Poisson regression model. The above model can be interpreted as follows: For vaccination count (1+) y_i and population estimate N_i in ZCTA $i \in \{1, ...1, 625\}$, the random variable is modeled by $Y \sim Poisson(\lambda_i)$. Then the above model can be fit for covariates $x_1, ..., x_p$. The covariates included in the final model are: % Female, Median Age, % Asian, %Hispanic, % Age 25 to 45, % Age 65 and Over, % High School Education Attainment Only, % Bachelors Educational Attainment, Graduate, or Professional Educational Attainment, % Percentage in Poverty, and Vaccination Equity Metric. Vaccination Equity Metric was treated as a categorical variable. Level 1 (bottom quartile) was excluded to avoid over specification.

However, the above model is adjusted slightly for the regression analysis. The data was highly over-dispersed. This is common for geographic count data. In order to account for the overdispersion, the quasi-Poisson regression model was used instead. It adds a parameters that estimates the over-dispersion. In practice, this will affect the coefficients' standard deviations and p-values, although not the estimates themselves. Generally quasi-Poisson is used over other options (for example, negative binomial) when the source of the over-dispersion is not of interest. Future work could consider different methods for modeling the over-dispersion.

Table 3 shows the coefficients, reported as their exponentiated values, as well as the 95% confidence intervals and the p-values for the coefficients in the model.

Covariates	Coefficient ($\exp(\beta)$)	95% LB	95% UB	P-value
Intercept	0.171	0.134	0.217	~ 0
% Female	1.008	1.005	1.011	~ 0
Median Age (Years)	1.011	1.008	1.014	~ 0
% Asian	1.004	1.004	1.005	~ 0
% Hispanic	1.005	1.005	1.005	~ 0
% Age 25 to 44	1.005	1.004	1.006	~ 0
% Age 65 and Over	1.002	1.000	1.005	0.0453
% High School Only	0.990	0.987	0.992	~ 0
% Bachelors +	1.003	1.002	1.005	~ 0
% Under Poverty Threshold	0.997	0.996	0.999	0.0022
VEM Q2	1.098	1.072	1.126	~ 0
VEM Q3	1.115	1.078	1.152	~ 0
VEM Q4	1.171	1.121	1.224	~ 0

Table 3: Results of model 1, fitted for available data (1,625 ZCTAs for Vaccination (1+ dose) counts from May 18th, 2021. LB refers to "Lower Bound" and UB refers to "Upper Bound."

The coefficients are reported as exponentiated (See Model 1: the mean count value, λ_i, is linked by the natural logarithm function) so that the results can be interpreted as change in vaccination rate. The percentage coefficients can be interpreted as follows: a unit change multiplied by the coefficient in the covariate results in a percentage change in the vaccination rate for a given ZCTA. So for example, each additional percentage a ZCTA that is Hispanic relates to a 0.5% increase in the vaccination rate, for the ZCTA. The median-age variable can be interpreted as follows: each additional year for the median age of a ZCTA relates to a 1.1% increase in the vaccination rate for the ZCTA.

The strongest negative relationships is High School Education. Each additional percentage increase in high school only educational attainment for a given ZCTA relates to a one percentage decrease in the expected vaccination rate of the ZCTA. VEM Q2 can be interpreted as follows: a ZCTA categorized as Q2 has a vaccination rate approximately 1.098% that of one with VEM Q1.

All the coefficients included in the model are highly statistically significant, with the exception of "Age 65 and Over." This may be due to the inclusion of Median Age as another covariate. [5] Assuming this model accurately represents the nature of vaccination rates and demographics, then there is roughly a five percentage chance these result for Age 65 and Over is due to chance in the observed sample.

[5] % Under the Poverty Threshold is at a level such that it would generally be considered statistically significant, most settings.

Evaluating the Model

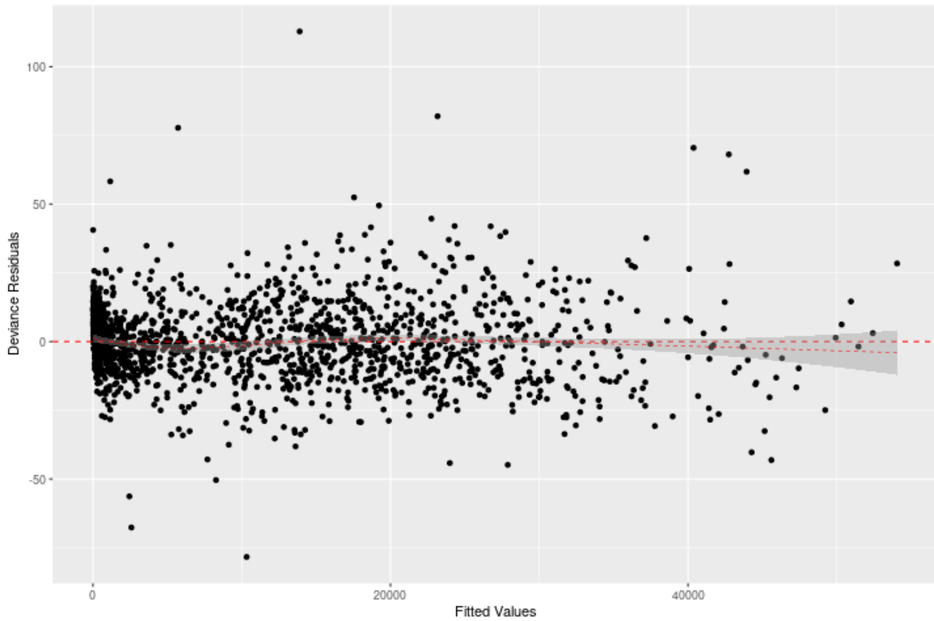

Figure 5: Plot of residuals versus fitted values from the quassipoisson regression model.

Residuals versus covariate plots and partial residual plots show no concerns with regards to the model specification. See "Models.Rmd" for partial residual plots for a complete list of these plots. The residuals versus fitted plot in Figure 5 shows no issues or irregular patterns, and suggests the model is at least a reasonable fit for the data.

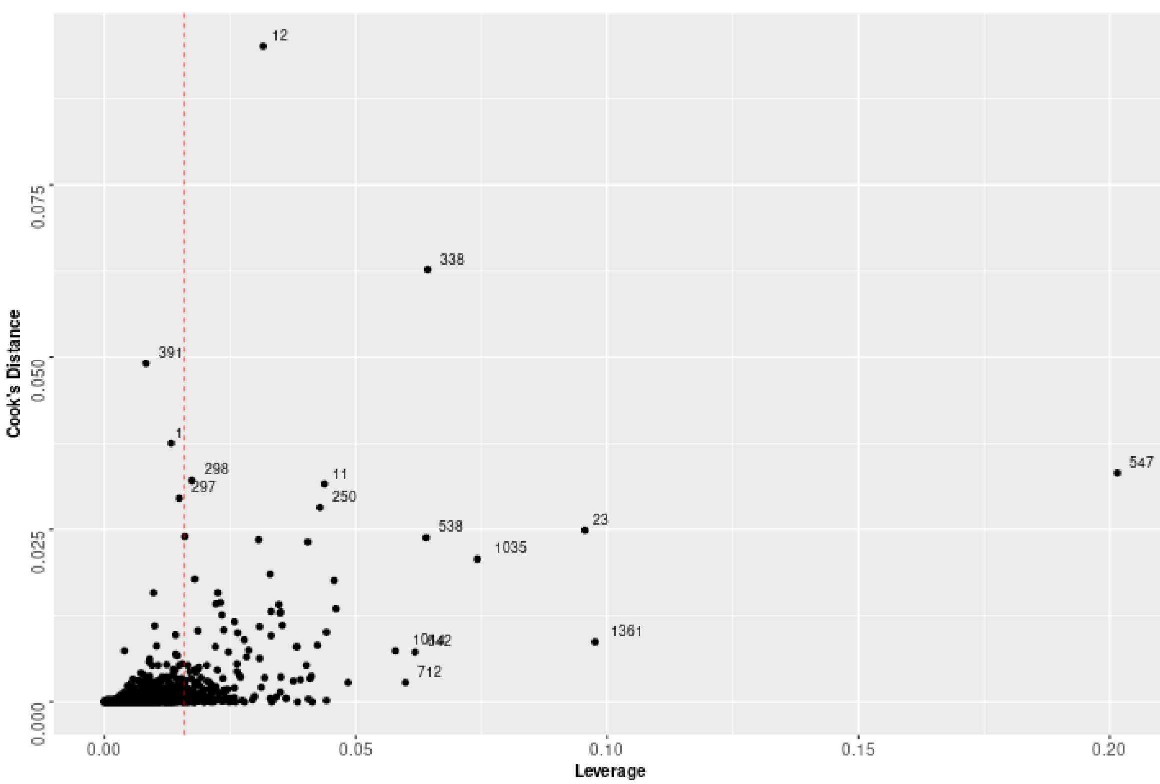

Figure 6: Plot of leverage versus Cook's distance for the quassipoisson regression model.

Figure 6 shows leverage versus Cook's distance for Model 1. Note none of the observations have a high Cook's distance. For this size data set and with this number of covariates, high Cook's distance would be something above .9 or so. These points are all significantly lower.[6] A number of points have high leverage; all of those to the right of the red-dotted line. Observation 547 stands out a little. It is a ZCTA in Orange county, with an estimated population of 52,094. Roughly 67 percent of its residents have been vaccinated with at least one dose. None of its characteristics are particularly extreme. High leverage is in itself not a major concern, and since there are no highly influential points, it can be said the model does not suffer from any serious outliers. Overall this model does not suffer from any identifiable issues.

Results

The fitted version of Model 1, shown in table 3, reveals the following relationships. ZCTAs with higher percentages female, Asian, and Hispanic have higher vaccination rates, all else equal. Older ZCTAs have higher vaccination rates as well. The Age 65 and Over variable was not highly significant by itself, but the result was also capture by median age. Additionally, even ZCTAs with higher representation of the 25 to 44 year old demographic showed higher vaccination rates. In some ways age is not surprising: no individuals under 12 had been vaccinated as of May 18th, and very few under 16. This model may not be the best for capturing the relationship, and future models would be better suited to re-scale the age percentages for ZCTAs based only on the ages that can be vaccinated.

Education appears to be an important factor, according to Model 1. Each additional percentage of having only a high school degree for a ZCTA related to a one percent decrease in percentage vaccinated. In other words, less educated ZCTAs have lower vaccination rates. It is the clearest result from this model. However the relationship was not as strong as maybe would have been expected based on Figure 4. This suggests some of the other covariates included in the model may have a confounding relationship with education.

Finally, ZCTAs with VEM ratings of 2, 3, or 4 all had significantly higher vaccination rates as compared to ZCTAs with a VEM rating of 1. This means that the least healthy ZCTAs are getting vaccinated the least; likely a public health concern.

Conclusion

Comparison with Other Work

The results of this report are in line with national and county level analyses. This report find older ZCTAs and more educated ZCTAs had higher vaccination rates, which is what Chernyavskiy et al. found at the county level.[5] Beleche et al. found that at the national level, Asian Americans and older Americans had higher vaccination rates,[16] which this report also found. In general, older, female, more educated, Asian, and Hispanic Californians were more likely to be vaccinated as of May 18th.

Future Work

The vaccine count data from the California Open Data Portal is available in this form between April 21st and June 1st, updated at a weekly frequency.[11] Studying changes in uptake by week could prove an interesting and useful extension of this analysis.

The ACS has many more tables available which estimate ZCTA level demographics. Future work could include studying more of these characteristics in the context of a regression analysis. Additionally, the Vaccine Equity Metric was not originally considered for this work. Future work should focus on understanding exactly how it is produced, and performing a more in depth analysis into its relationship with vaccinations.

64 of the ZCTAs had a total vaccination count (either fully or partially vaccinated) greater than the population. Most of these ZCTAs were small populations, with the 3rd quartile being a population of 847 individuals. In this researcher's opinion, there are two likely explanations for this discrepancy.

[6] Using the rule of thumb $F_{.5}(p, n-p)$

The simplest explanation is that the ACS 5 year estimates are based on a survey, so there is some uncertainty. Further, populations change over time. Small ZCTAs can see a large change in percentage population growth quickly with just the addition of a few hundred individuals. Therefore, it is possible the discrepancy is due to the ACS 5 year estimates being uncertain and out of date. The other significant possibility is that some vaccination counts are inaccurately recorded, and perhaps record the ZCTA that an individual was vaccinated in, rather than the ZCTA the individual lives in. The data dictionary for the vaccination data states:

> Zip Code Tabulation Area from 2010 Census designations. ZCTA vaccination metrics are based on the stated residence of individuals; however, there may be instances where providers erroneously enter administration ZIP code as the residence ZIP code. Vaccination records with missing ZIP codes are not included in this data set.[11]

Which suggests manual error could be part of the issue. Future work may look to find a more reliable source for ZCTA population. The 2020 census may be available in the near future.

References

[1] "U.S. Census Bureau". American community survey information guide. Online, 10 2017. Available at $https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS\_Information\_Guide.pdf$.

[2] "U.S. Census Bureau". 2010 zip code tabulation area (zcta) relationship file record layouts. Online, 2 2019. Available at https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/2010-zcta-record-layout.html.

[3] "U.S. Census Bureau". Zip code tabulation areas (zctas). Online, 8 2020. Available at https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html.

[4] "U.S. Census Bureau". American community survey data. Online, 4 2021. Available at https://www.census.gov/programs-surveys/acs/data.html. Note the actual data is retrieved from the U.S. Census Bureau's "Explore Census Data" form, by filtering based on ZTCA. The different demographic data is stored in a number of different tables.

[5] Pavel Chernyavskiy, Jeanita W. Richardson, and Sarah J. Ratcliffe. Covid-19 vaccine uptake in united states counties: geospatial vaccination patterns and trajectories towards herd immunity. *medRxiv*, 2021. Preprint.

[6] "COVID19.CA.GOV". Vaccination progress data. Online, 6 2021. Available at https://covid19.ca.gov/vaccination-progress-data/.

[7] "dieghernan". Mapping my data to a zip code area map in r. Online, 2 2020. Available at https://stackoverflow.com/questions/60419762/mapping-my-data-to-a-zip-code-area-map-in-r. The mapping code in my report was based on the example diegheman provided in a comment. The original forum post was created by user XinXin.

[8] Amy Harmon and Josh Holder. They haven't gotten a covid vaccine yet. but they aren't 'hesitant' either. Online, 5 2021. Accessed June 9th, 2021. Available at https://www.nytimes.com/2021/05/12/us/covid-vaccines-vulnerable.html.

[9] Grace Manthey. Map: See how many people are vaccinated in your zip code. Online, 4 2021. Available at https://abc7news.com/california-vaccination-percentage-covid-rates-ca-vaccine-rate-how-many-people-have-gotten-the-covid-19/10560392/.

[10] "Office of Governor Newsom". Governor newsom launches "vaccinate all 58" campaign based on safety and equity as first vaccines arrive to california. Online, 12 2020. Available at https://www.gov.ca.gov/2020/12/14/governor-newsom-launches-vaccinate-all-58-campaign-based-on-safety-and-equity-as-first-vaccines-arrive-to-california/.

[11] "California Department of Public Health". Covid-19 vaccine progress dashboard data by zip code. Online, 5 2021. May 18th Data used. Available at https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data-by-zip-code. Data Dictionary included on website.

[12] Katy Rossiter. Understanding geographic relationships: Counties, places, tracts and more. Online, 7 2014. Available at https://www.census.gov/newsroom/blogs/random-samplings/2014/07/understanding-geographic-relationships-counties-places-tracts-and-more.html.

[13] Nouran Salahieh, Chip Yost, and Carlos Saucedo. Starting april 15, all californians 16 and older will be eligible for covid-19 vaccines; those 50 and older can sign up april 1. Online, 3 2021. Available at https://ktla.com/news/california/starting-april-15-all-californians-16-and-older-will-be-eligible-for-covid-19-vaccines-those-50-and-older-become-eligible-april-1/.

[14] "Los Angeles Times Staff". Tracking coronavirus vaccinations in california. Online, 6 2021. Accessed June 9th, 2021. Available at https://www.latimes.com/projects/california-coronavirus-cases-tracking-outbreak/covid-19-vaccines-distribution/.

[15] Tom Tapp. California prepares to vaccinate 12 to 15 year olds starting thursday; officials address parental consent, school vaccination issues. Online, 5 2021. Available at https://deadline.com/2021/05/california-los-angeles-plans-to-vaccinate-12-15-year-olds-thursday-1234754440/.

[16] TrinidadBeleche, Joel Ruhter, Allison Kolbe, Jessica Marus, Laina Bush, and Benjamin Sommers. May 2021issue brief1covid-19 vaccine hesitancy: Demographic factors, geographic patterns, and changes over time, 5 2021.

[17] Spencer Woody, Emily Javan, Kaitlyn Johnson, Remy Pasco, Maureen Johnson-Leon, Michael Lachmann, Spencer J.Fox, and Lauren Ancel Meyer. Spatial distribution of covid-19 infections and vaccinations in austin, texas. *The University of Texas at Austin COVID-19 Modeling Consortium*, 4 2021.