

# PM566 Final Project: An initial Research on the Association between Covid-19 Vaccination Status and Age

Yating Zeng

2022-12-07

## Introduction

COVID-19 has been here for around 3 years, with vaccine widely used. It would be likely that some of the people tend to not take the vaccine than the others. Thus, in this project, the question of my interest is: What is the association between age and two vaccination status (at least one dose & completed a primary series) in California state? For this project, I'll use the dataset on Covid-19 vaccination from the Centers for Disease Control and Prevention (CDC) website, which provided data for select demographic characteristics (age, sex, and age by sex) of people receiving COVID-19 vaccinations in the United States at the national and jurisdictional levels, fitting my analysis interest well. All the data were cumulative data, which were counted since the date it started observing.

## Methods

### 1.Dataset

In this project, the dataset used was a public resource from CDC website, named "COVID-19 Vaccination Age and Sex Trends in the United States, National and Jurisdictional". The link of the dataset is shown below: <https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Age-and-Sex-Trends-in-the-United-States>. The CSV file of the data was then downloaded and read into R studio for further analysis in this project.

### 2.Data cleaning, wrangling and EDA

After checking the summary of the content of the dataset, the dimensions and the original properties for each variable were known. I filtered the data to create a new dataset to keep only the information of California. For simplifying the typing in analysis, 7 columns were renamed to be shorter. Then the proportion of missing values of each column and column "Demographic\_Category" were checked. Considering that age and vaccination status of primary dose series were the main factors towards this analysis, the information about "Booster", "Age\_unknown" and all the "Age>65" levels of the "Demographic\_Category" column, and the missing values of "dose1"(count of people take at least one dose) and "census" (census statistics used for calculating the percentage of vaccination) were removed.

Because the information this dataset was about the information strongly rely on time series, and all the statistics were cumulative data, a new variable "date" was created for further reorder the data by the time recorded. Based on the category from "Demographic\_Category" variable (now named "cat"), the original dataset was split into 4 subset for better analysis, which are 1. objects from both sex categorized only by

Table 1: Table 1.Summary of Percent of people with at least one dose grouped by age

Group	Min	Q1	Median	Q3	Max	Records
Ages_<2yrs	0	3.73	5.20	6.20	7.1	122
Ages_<5yrs	0	4.73	7.00	8.20	9.1	122
Ages_02-04_yrs	0	5.35	8.15	9.47	10.4	122
Ages_05-11_yrs	0	1.50	16.10	42.70	46.2	667
Ages_12-17_yrs	0	35.70	72.80	82.30	84.0	676
Ages_18-24_yrs	0	54.10	77.65	88.32	90.7	676
Ages_25-39_yrs	0	58.30	78.05	86.40	88.2	676
Ages_25-49_yrs	0	62.30	81.05	89.00	90.6	676
Ages_40-49_yrs	0	69.38	86.35	93.40	94.8	676
Ages_50-64_yrs	0	78.20	91.35	95.00	95.0	676
Ages_65-74_yrs	0	90.90	95.00	95.00	95.0	676
Ages_75+_yrs	0	85.70	93.35	95.00	95.0	676

age level; 2. objects were all females categorized by age level; 3. objects were all males categorized by age level; and 4. objects from both sex categorized only by sex level.

Summary tables and boxplots were created by 2 vaccination status (“at least one dose” and “completed a primary series”) and 4 categorical groups (“age”; “female\_age”; “male\_age”; and “sex”), showing the minimum, 1st quantile, median, 3rd quantile, maximum, and the number of recorded objects of “the percentage of people” with the 2 kinds of vaccination status grouped by age, sex or age groups stratified by sex. The reason to use data stratified by sex was to remove the possible confounding effect from sex on the association between vaccination status and age level. Then to find out the association between sex and vaccination and between age and vaccination status, grouped scatter plots were planed to create, by the same approach mentioned in the part of summary tables and figures. In this report, totally 2 tables, 2 box plots and 3 scatter plots created were shown here, more information could be found in the supplements website [Link].

## Results

### 1. Vaccination status trends of “At least one dose” and “A primary series” with little difference

From the summary tables and figures shown bellow, we could notice that both the percent of people with at least one dose grouped by age (Table 1 & Figure 1) and the percent of people completed a primary series grouped by age (Table1 & Figure 2) are showing consistent trends and similar data structure. Excepting a little part of the data, most part the data were showing a trend that the statistics (the minimum, 1st quantile, median, 3rd quantile, maximum) of vaccination rate would increase, when the age level was higher. And for the major part of observed objects, who were aged from 12-17 years old to 75+ years old, the final vaccination rate will went up to around 90%, while for the low age-level(<5 years) objects, the vaccination rate were all under or around 10%, and for 5-11 years objects the vaccination rate stayed in the middle, which is around 50%. Beside, the sample size are also nearly the same for these two subsets among each age group, with more than 600 records for each age group which is no less than “5-11 years” level and with around 100 records for age groups “<2 years”, “<5 years” and “2-4 years” (Table 1 & 2). Though in the one for the percent of people with at least one dose, the final vaccination would go up to 80+%, while for the one of the percent of people completed a primary series, the rate could up to 70+%, this make sense for the reason that vaccination rate would always higher if contained the sample who only take one dose. Thus, we would focus on the the results of the sample with at least one does for the next in this report.

Table 2: Table 2.Summary of Percent of people completed a primary series grouped by age

Group	Min	Q1	Median	Q3	Max	Records
Ages_<2yrs	0	0.70	1.40	2.10	2.8	99
Ages_<5yrs	0	0.40	2.00	3.20	4.4	116
Ages_02-04_yrs	0	0.72	2.50	3.98	5.4	114
Ages_05-11_yrs	0	1.40	11.25	36.10	39.2	630
Ages_12-17_yrs	0	20.95	64.20	73.20	74.7	667
Ages_18-24_yrs	0	43.20	67.85	74.90	76.6	676
Ages_25-39_yrs	0	48.58	69.05	74.70	75.9	676
Ages_25-49_yrs	0	52.27	72.15	77.50	78.6	676
Ages_40-49_yrs	0	58.77	77.55	82.40	83.4	676
Ages_50-64_yrs	0	67.90	82.45	87.00	88.6	676
Ages_65-74_yrs	0	80.20	89.85	94.10	95.0	676
Ages_75+_yrs	0	75.50	83.05	86.73	88.8	676

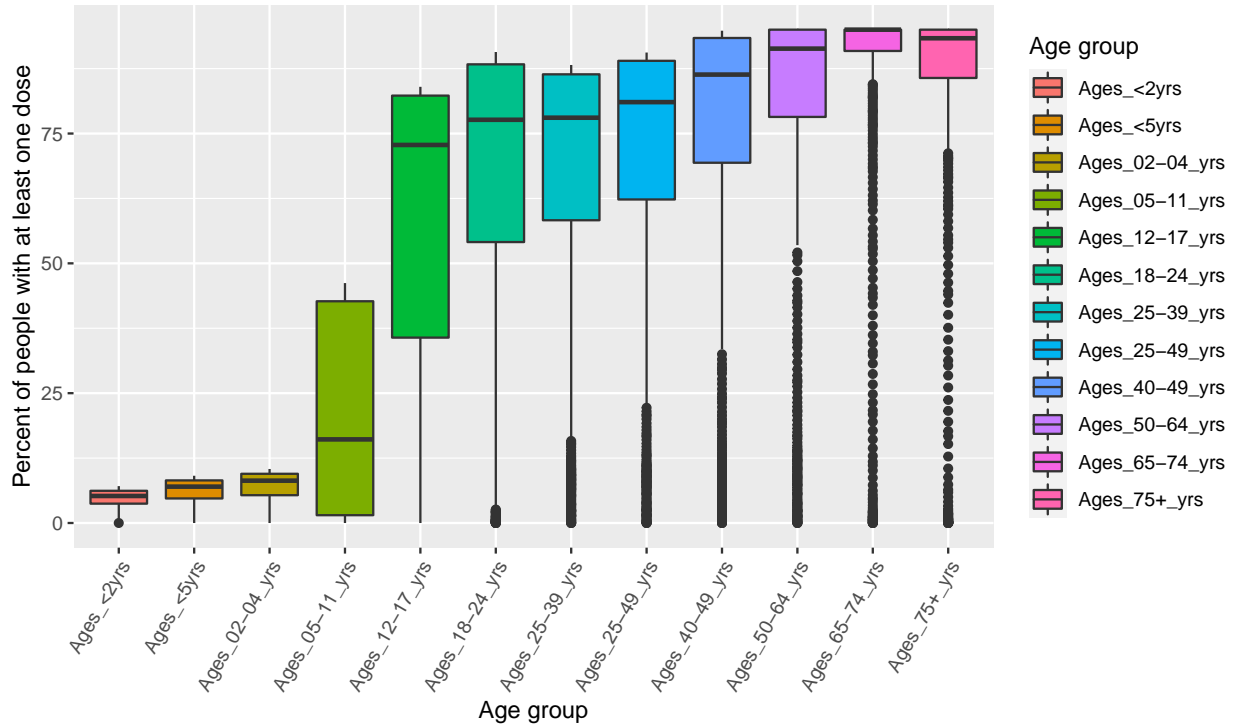


Figure 1.Percent of people with at least one dose grouped by age

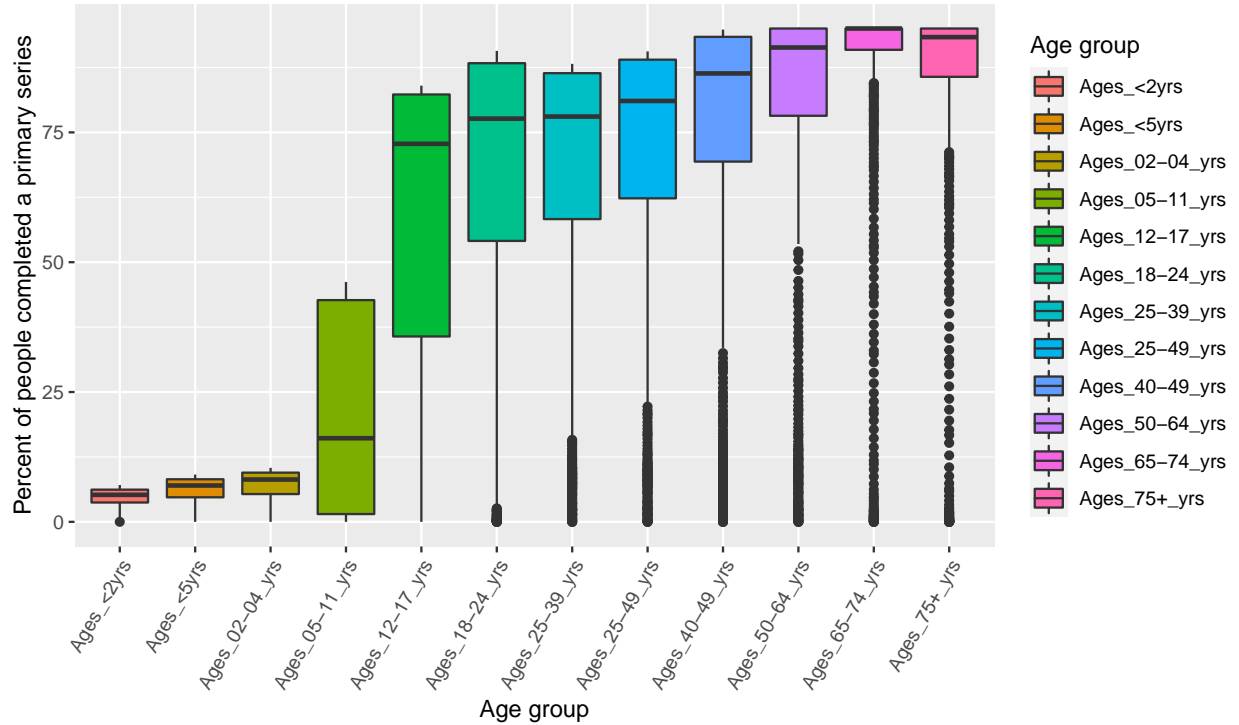


Figure 2. Percent of people completed a primary series grouped by age

## 2. Females have higher vaccination rate than males

The results for the percent group by sex shown in Figure 1. suggests that compared to the male, females would have higher vaccination rate across all the time recorded, and went up to a higher level in the end. Considering over this issue, we would better to analyze the data stratified by gender to prevent possible confounding effect. This difference might also give light on the direction for further study, which could help government to take different measure towards males and females on publicity, which would not be discussed more in this research.

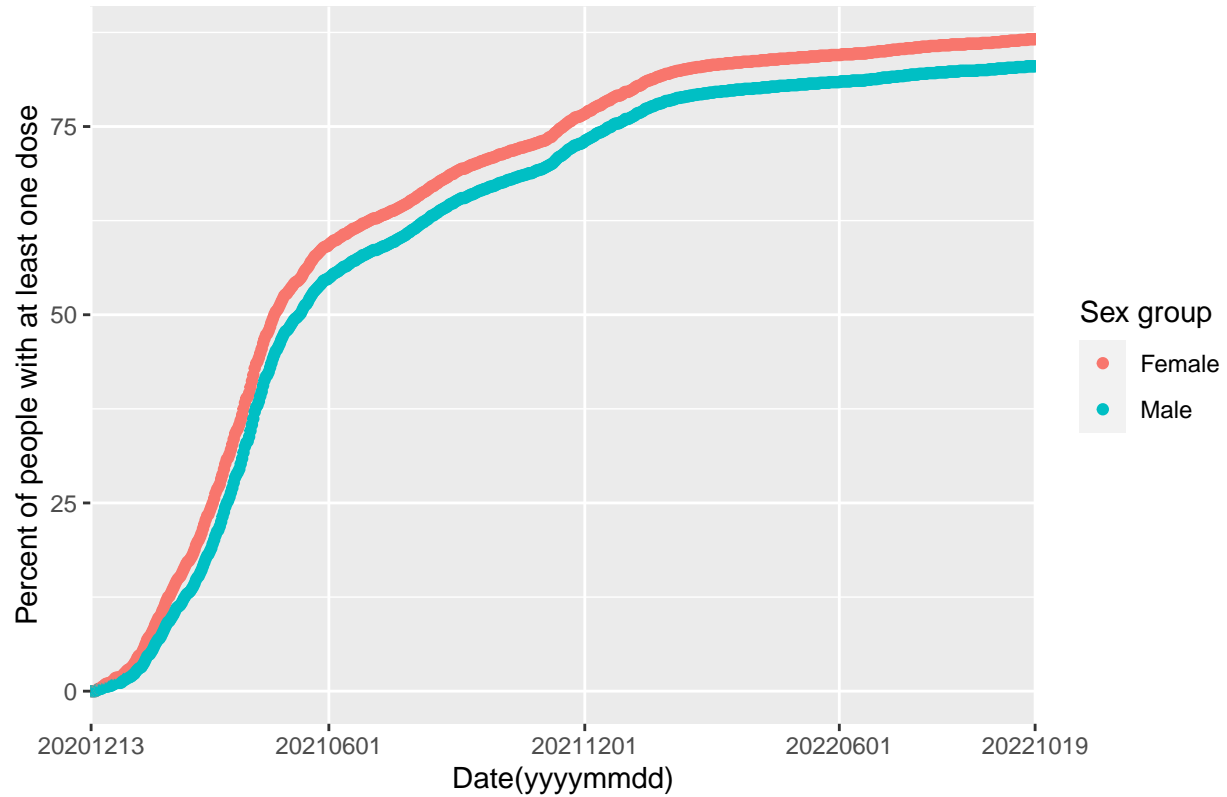


Figure 8. Percent of people with at least one dose grouped by sex

### 3. The vaccination rate would be higher with the age level increased

Based on the figures bellow (Figure 5 & 6), we could find that in both gender, the vaccination rate were higher with the age level being higher for the same time point, and the objects with higher age might take shorter time to have a relatively high vaccination rate. Which needs to be mentioned is that, this trend was also observed in the low age-level group(<5 years), but with obviously lower vaccination rate than the major part of the sample objects. During the very beginning of the record period, vaccination rate of people age from 18 to 64 increased within relatively shorter time, which might be due to the reason that for those people, they would be more easier to accept latest information and be more willing to try new things. After that, the vaccination rate of people aged over 65 increased to a high level with much shorter time, which might be due to the more support, publicity and education from the government, considering that Covid-19 is deadlier for those people. The trend for people aged from 12 to 64, would be more useful for extending to the other vaccination rate expectation in the future.

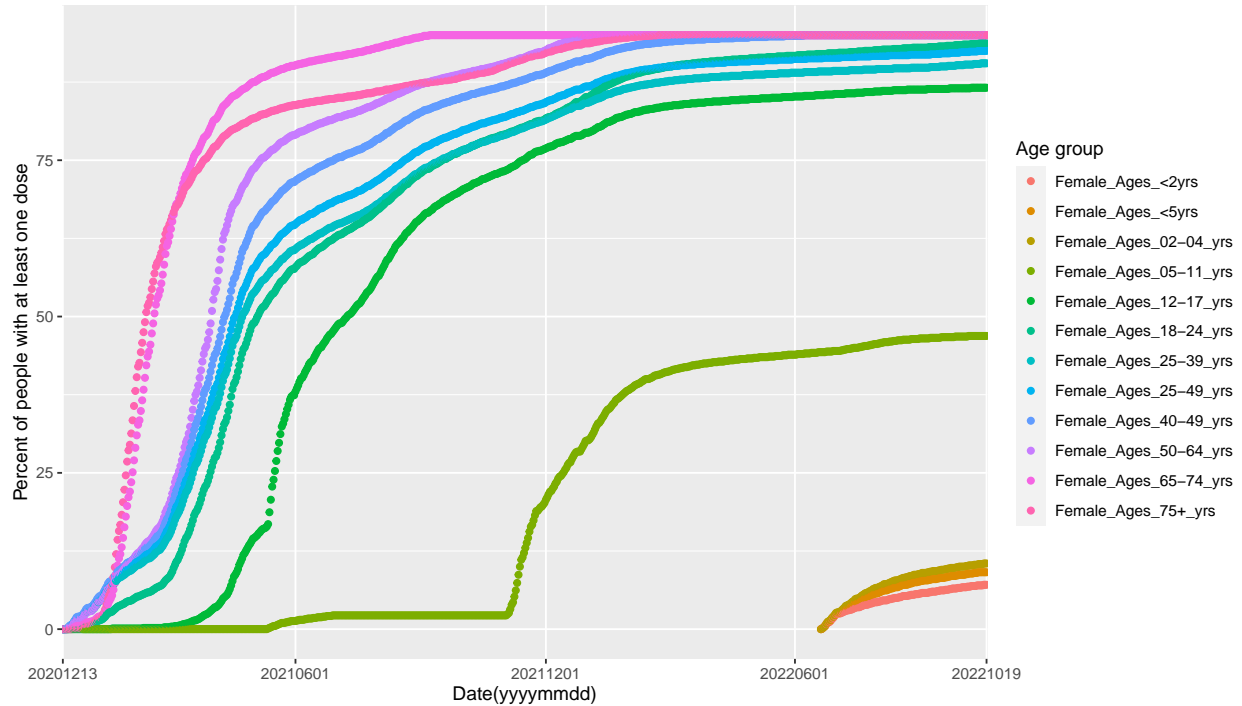


Figure 9.2020–2022 Percent of people with at least one dose of females

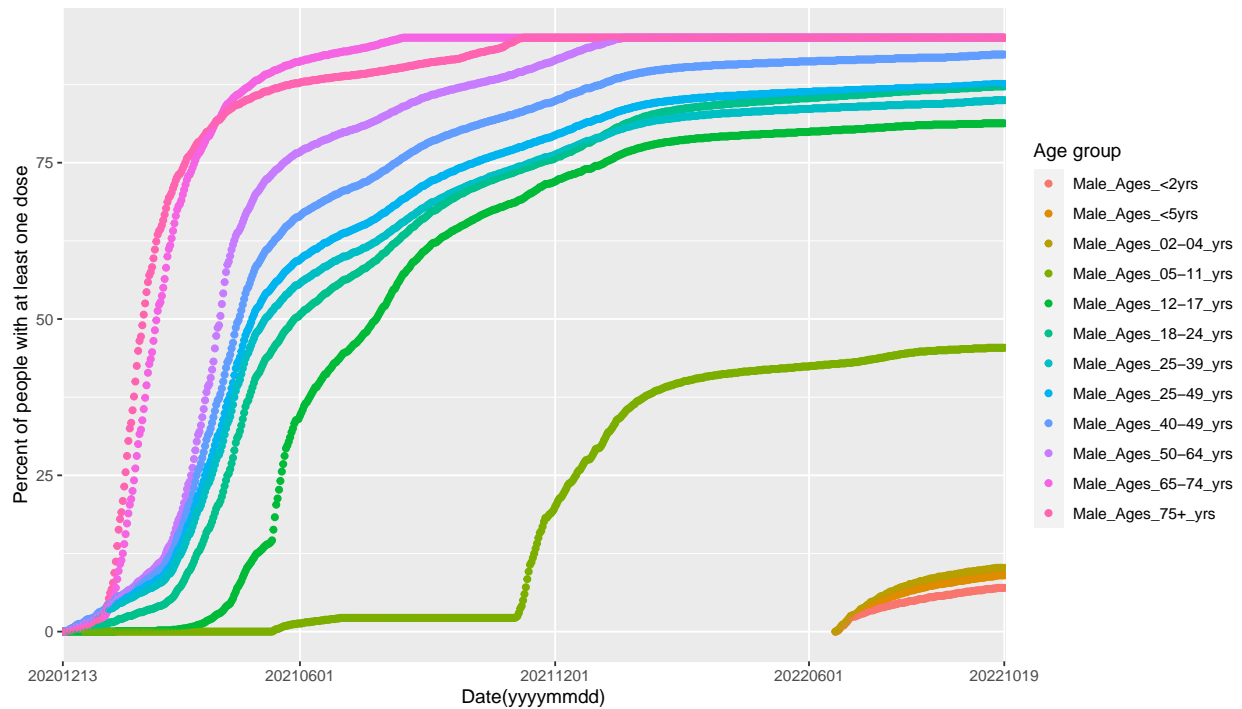


Figure 10.2020–2022 Percent of people with at least one dose of males

## Conclusion

We could believe that there could be an association between age and the two vaccination status (at least one dose & completed a primary series) in California state. For both 2 kinds of vaccination status (take at least

one dose & with completed series) and both sex, the vaccination rate would be higher with the age level being higher for the same time point, and the objects with higher age might also take shorter time to have a relatively high vaccination rate. And the final vaccination rate would be higher with the age level being higher, but the rate for people with age less than 5 years old would keep in a low level, even though they follow the same trend mentioned above.