

# PM566 Final Project

Yating Zeng

2022-12-07

## Contents

<b>Introduction</b>	<b>1</b>
<b>Methods</b>	<b>1</b>
1.Dataset . . . . .	1
2.Data cleaning, wrangling and EDA . . . . .	2
<b>Results</b>	<b>2</b>
1. Vaccination status trends of “At least one dose” and “A primary series” with little difference . . . . .	2
2. Females have higher vaccination rate than males . . . . .	4
3. The vaccination rate would be higher with the age level increased . . . . .	5
<b>Conclusion</b>	<b>7</b>
Link to my report PDF: <b>Report</b> [download]	
Link to my supplements PDF: <b>Supplements</b> [download]	

## 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 the COVID-19 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-Uni/5i5k-6cmh>. 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. Totally *25 tables and figures* were created, with 5 figures would shown below and more information could be found in the supplementary. All the tables and figure were planed to create by 2 vaccination status (“at least one dose” and “completed a primary series”) and 4 categorical groups (“age”; “female\_age”; “male\_age”; and “sex”). The summary tables and figures would show 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 association between the age and vaccination status would be shown with scatterplots.

## Results

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

From the summary figures shown bellow, we could notice that both the percent of people with at least one dose grouped by age (Figure 1) and the percent of people completed a primary series grouped by age (Figure 2) are showing consistent trends and similar data structure. Excepting a little part of the data, most of 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. Thus, we would focus on the the results of the sample with at least one does for the next in this report.

Figure 1. Percent of people with at least one do

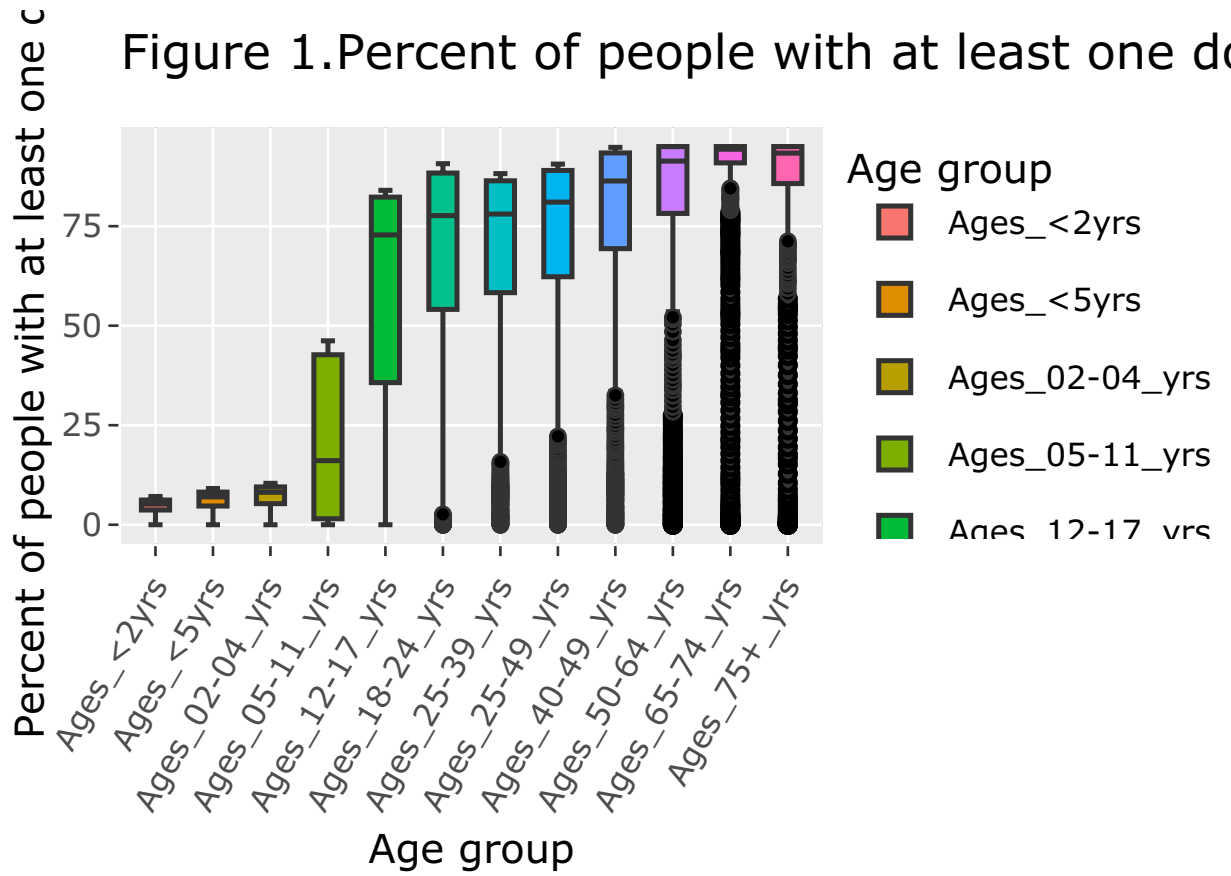


Figure 1.

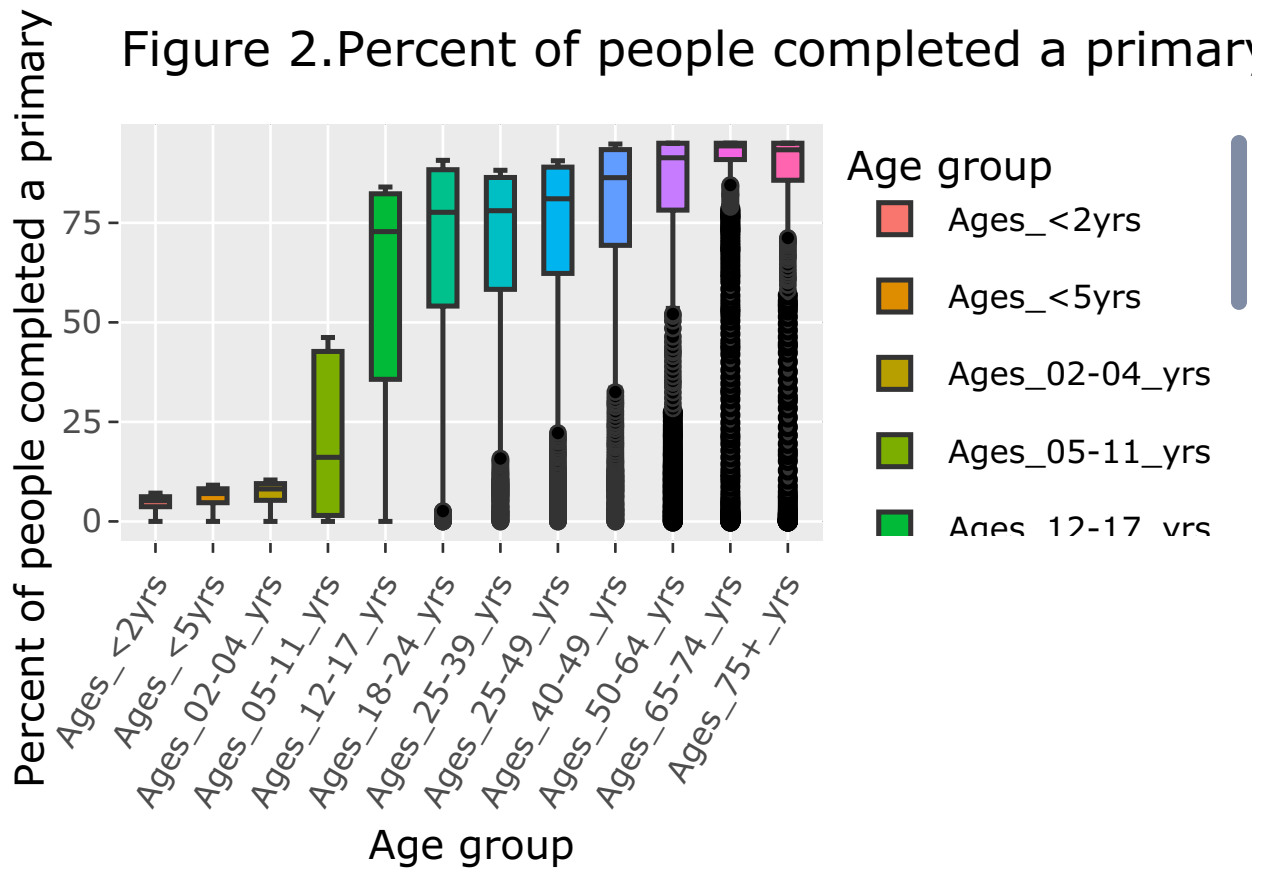
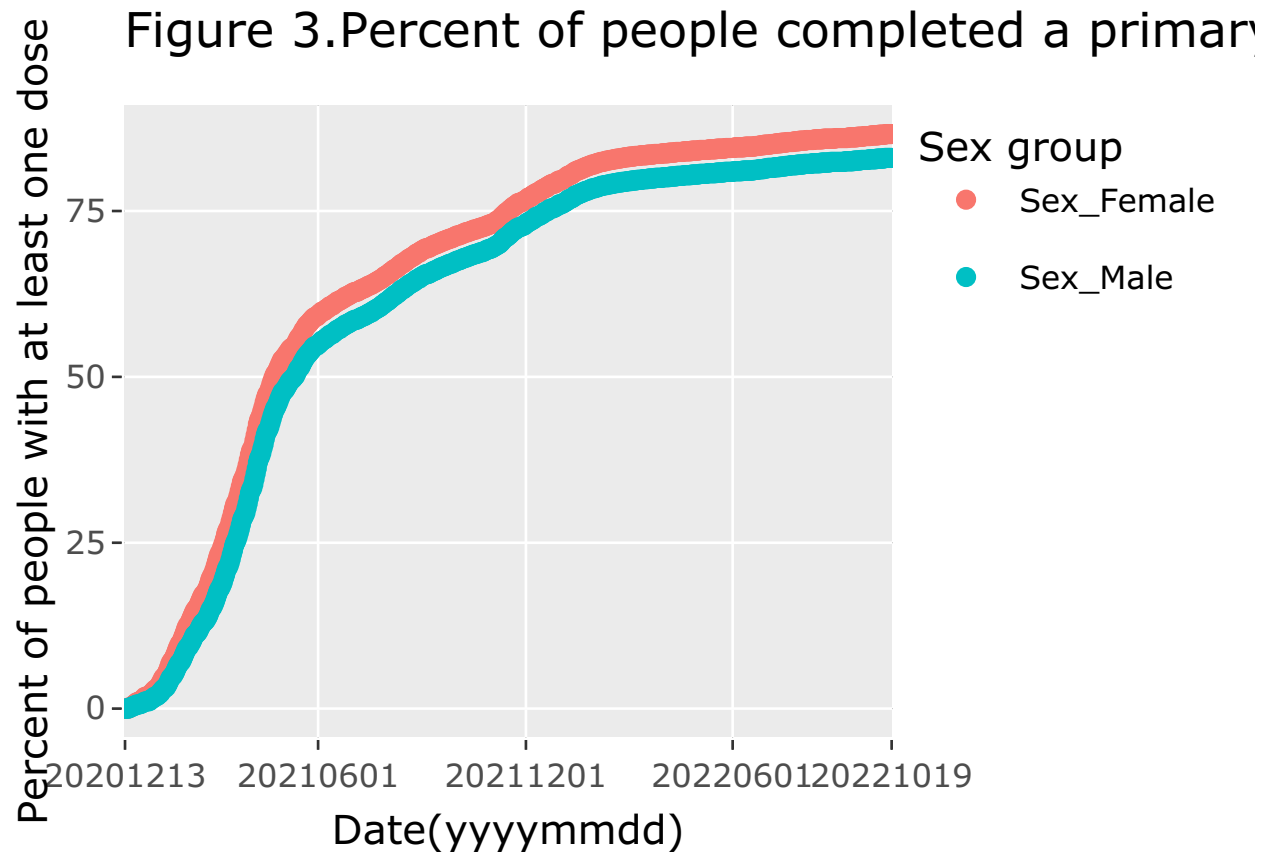


Figure 2.

## 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.



### 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.

Figure 4. Percent of people with at least one dose

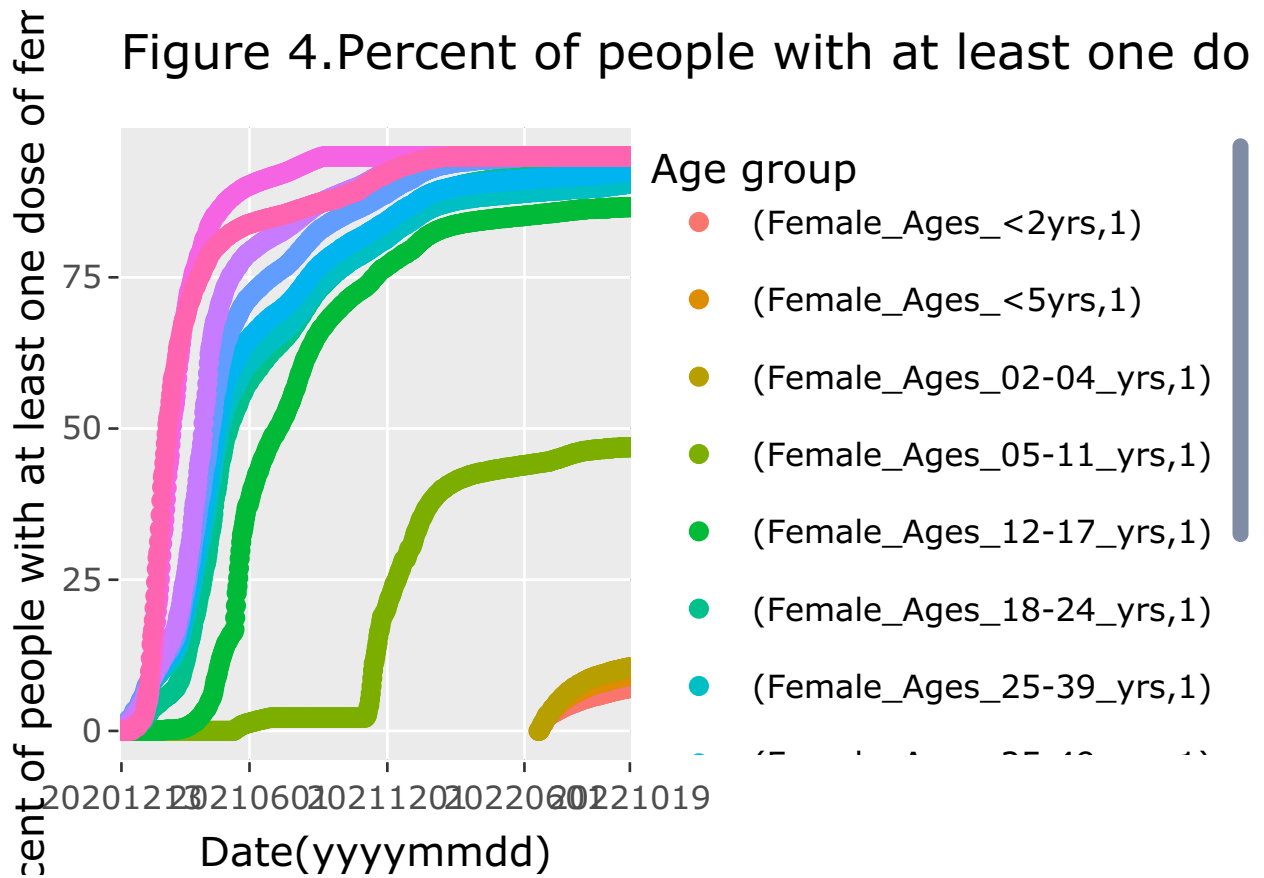


Figure 4.

Figure 5. Percent of people with at least one dose

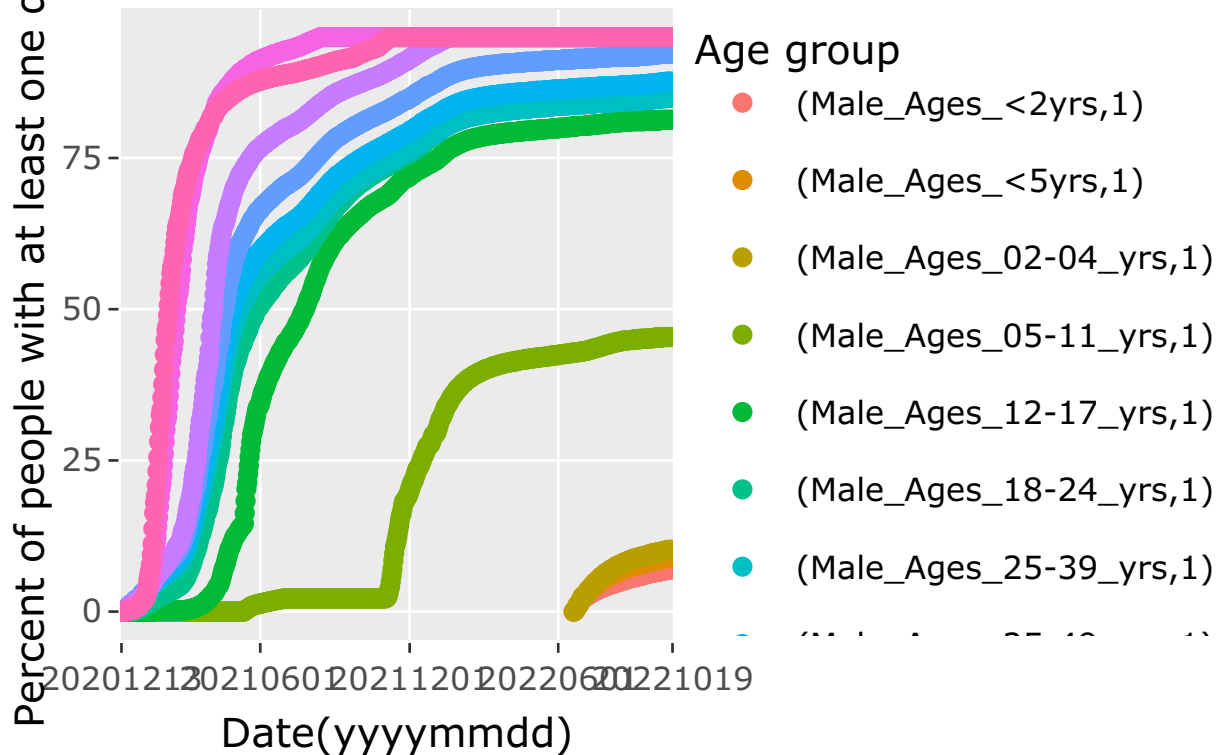


Figure 5.

## 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 two 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.