

PM566_Midterm

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-National-and-Jurisdictional>. 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 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.

Totally 8 summary tables and 8 summary figures (boxplots) were planned 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”), 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 the age and vaccination status, 8 grouped scatterplots were planned to create, by the same approach mentioned in the part of summary tables and figures.

Results

1. Summary tables

From the summary table of the percent of people with at least one dose grouped by age (Table 1), we could notice that 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 Percent of people with at least one dose would be larger 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%. This trend were consistent with the results form all the other 2 summary tables of Percent of people with at least one dose grouped by age stratidied by sex (Table 2 & 3).

For the summary table of the percent of people completed a primary series grouped by age (Table 5), we could find that there was still a trend similar to the one above between the vaccination status and age level, but what different was that the statistics were lower than the one of people with at least one dose. For the major part of observed objects, aged from 12-17 years old to 75+ years old, the final vaccination rate will went up to 70%-90%, with group of 65-74 years age would up to 95%. But for the low age-level(<5 years) objects, the vaccination rate were all under or around 5%, and for 5-11 years objects the vaccination rate stayed in the middle, which is around 40%. This trend were consistent with the results form all the other 2 summary tables of Percent of people with at least one dose grouped by age stratidied by sex (Table 6 & 7).

From the summary tables for the percentage of 2 vaccination status grouped by sex (Table 4 & 8), we could find that the statistics of female were always larger than males; and 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+%.

2. Summary figures

All the summary figures (Figure 1-3 & 5-7) showed the similar trend mentioned above relatively. Excepting these information we gained and mentioned, we still could find that for the object age level from 5-11 years old to 65-74 years old, with the age level went up, the major part of the statistics would with higher values, which means it might take shorter time to have a relatively high vaccination rate for those people with higher age level. And the two figures (Figure 4 & 8) for the percent group by sex were still show that the females would have higher vaccination rate, making it to be reasonable that we’d better use the stratified data for analyze the association between age and vaccination rate.

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

cat	min	q1	median	q3	max	days_recorded
Ages_<2yrs	0	3.725	5.20	6.200	7.1	122
Ages_<5yrs	0	4.725	7.00	8.200	9.1	122
Ages_02-04_yrs	0	5.350	8.15	9.475	10.4	122
Ages_05-11_yrs	0	1.500	16.10	42.700	46.2	667
Ages_12-17_yrs	0	35.700	72.80	82.300	84.0	676
Ages_18-24_yrs	0	54.100	77.65	88.325	90.7	676
Ages_25-39_yrs	0	58.300	78.05	86.400	88.2	676
Ages_25-49_yrs	0	62.300	81.05	89.000	90.6	676
Ages_40-49_yrs	0	69.375	86.35	93.400	94.8	676
Ages_50-64_yrs	0	78.200	91.35	95.000	95.0	676
Ages_65-74_yrs	0	90.900	95.00	95.000	95.0	676
Ages_75+_yrs	0	85.700	93.35	95.000	95.0	676

Table 2: Table 2.Summary of Percent of people with at least one dose in females grouped by age

cat	min	q1	median	q3	max	days_recorded
Female_Ages_<2yrs	0	3.725	5.30	6.275	7.1	122
Female_Ages_<5yrs	0	4.750	7.10	8.275	9.1	122
Female_Ages_02-04_yrs	0	5.425	8.20	9.575	10.5	122
Female_Ages_05-11_yrs	0	1.700	18.20	43.500	46.9	661
Female_Ages_12-17_yrs	0	37.400	75.15	84.800	86.6	676
Female_Ages_18-24_yrs	0	57.600	80.35	91.300	93.7	676
Female_Ages_25-39_yrs	0	60.600	80.25	88.700	90.5	676
Female_Ages_25-49_yrs	0	64.600	83.05	90.900	92.5	676
Female_Ages_40-49_yrs	0	71.500	87.90	94.700	95.0	676
Female_Ages_50-64_yrs	0	79.000	91.45	95.000	95.0	676
Female_Ages_65-74_yrs	0	90.100	95.00	95.000	95.0	676
Female_Ages_75+_yrs	0	83.800	91.25	95.000	95.0	676

Table 3: Table 3.Summary of Percent of people with at least one dose in males grouped by age

cat	min	q1	median	q3	max	days_recorded
Male_Ages_<2yrs	0	3.700	5.20	6.200	7.0	122
Male_Ages_<5yrs	0	4.650	6.90	8.100	9.0	122
Male_Ages_02-04_yrs	0	5.250	8.05	9.375	10.2	122
Male_Ages_05-11_yrs	0	1.675	17.90	42.000	45.4	660
Male_Ages_12-17_yrs	0	33.975	70.35	79.600	81.3	676
Male_Ages_18-24_yrs	0	50.375	74.55	84.900	87.2	676
Male_Ages_25-39_yrs	0	55.475	75.15	83.300	85.0	676
Male_Ages_25-49_yrs	0	59.300	78.25	86.000	87.6	676
Male_Ages_40-49_yrs	0	66.300	83.75	91.000	92.3	676
Male_Ages_50-64_yrs	0	76.575	90.15	95.000	95.0	676
Male_Ages_65-74_yrs	0	91.100	95.00	95.000	95.0	676
Male_Ages_75+_yrs	0	87.700	95.00	95.000	95.0	676

Table 4: Table 4.Summary of Percent of people with at least one dose grouped by sex

cat	min	q1	median	q3	max	days_recorded
Sex_Female	0	59.1	75.0	84.2	86.6	676
Sex_Male	0	54.7	71.3	80.6	83.0	676

Table 5: Table 5.Summary of Percent of people completed a primary series grouped by age

cat	min	q1	median	q3	max	days_recorded
Ages_<2yrs	0	0.700	1.40	2.100	2.8	99
Ages_<5yrs	0	0.400	2.00	3.200	4.4	116
Ages_02-04_yrs	0	0.725	2.50	3.975	5.4	114
Ages_05-11_yrs	0	1.400	11.25	36.100	39.2	630
Ages_12-17_yrs	0	20.950	64.20	73.200	74.7	667
Ages_18-24_yrs	0	43.200	67.85	74.900	76.6	676
Ages_25-39_yrs	0	48.575	69.05	74.700	75.9	676
Ages_25-49_yrs	0	52.275	72.15	77.500	78.6	676
Ages_40-49_yrs	0	58.775	77.55	82.400	83.4	676
Ages_50-64_yrs	0	67.900	82.45	87.000	88.6	676
Ages_65-74_yrs	0	80.200	89.85	94.100	95.0	676
Ages_75+_yrs	0	75.500	83.05	86.725	88.8	676

Table 6: Table 6.Summary of Percent of people completed a primary series in females grouped by age

cat	min	q1	median	q3	max	days_recorded
Female_Ages_<2yrs	0	0.700	1.40	2.100	2.8	97
Female_Ages_<5yrs	0	0.900	2.20	3.350	4.5	107
Female_Ages_02-04_yrs	0	1.600	2.90	4.400	5.5	100
Female_Ages_05-11_yrs	0	1.500	11.80	36.700	39.9	629
Female_Ages_12-17_yrs	0	34.525	67.05	75.875	77.3	650
Female_Ages_18-24_yrs	0	46.675	70.75	77.900	79.6	676
Female_Ages_25-39_yrs	0	50.800	71.30	77.100	78.4	676
Female_Ages_25-49_yrs	0	54.575	74.25	79.700	80.9	676
Female_Ages_40-49_yrs	0	60.900	79.25	84.100	85.1	676
Female_Ages_50-64_yrs	0	68.800	82.75	87.300	88.9	676
Female_Ages_65-74_yrs	0	79.475	89.05	93.225	95.0	676
Female_Ages_75+_yrs	0	73.600	81.35	85.000	86.9	676

Table 7: Table 7.Summary of Percent of people completed a primary series in males grouped by age

cat	min	q1	median	q3	max	days_recorded
Male_Ages_<2yrs	0	0.700	1.40	2.100	2.8	98
Male_Ages_<5yrs	0	0.525	2.00	3.175	4.3	114
Male_Ages_02-04_yrs	0	1.275	2.70	4.200	5.4	104
Male_Ages_05-11_yrs	0	1.400	15.10	35.500	38.5	615
Male_Ages_12-17_yrs	0	27.600	62.00	70.500	71.9	657
Male_Ages_18-24_yrs	0	39.575	64.65	71.600	73.2	676
Male_Ages_25-39_yrs	0	45.900	66.30	71.825	72.9	676
Male_Ages_25-49_yrs	0	49.500	69.45	74.700	75.8	676
Male_Ages_40-49_yrs	0	55.975	75.05	80.000	81.0	676
Male_Ages_50-64_yrs	0	66.300	81.40	86.000	87.6	676
Male_Ages_65-74_yrs	0	80.600	90.25	94.525	95.0	676
Male_Ages_75+_yrs	0	77.675	84.95	88.800	91.1	676

Table 8: Table 8.Summary of Percent of people completed a primary series grouped by sex

cat	min	q1	median	q3	max	days_recorded
Sex_Female	0	49.3	66.20	73.925	75.8	676
Sex_Male	0	45.1	62.55	70.300	72.1	676

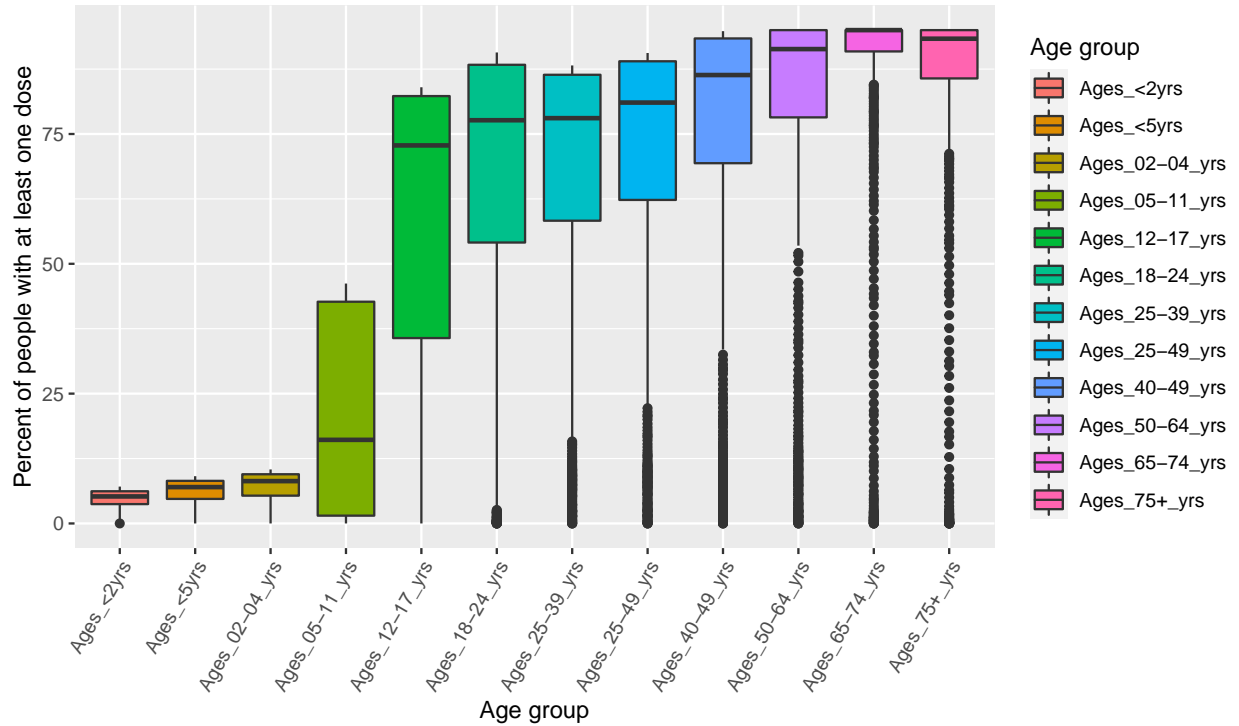


Figure 1.Percent of people with at least one dose grouped by age

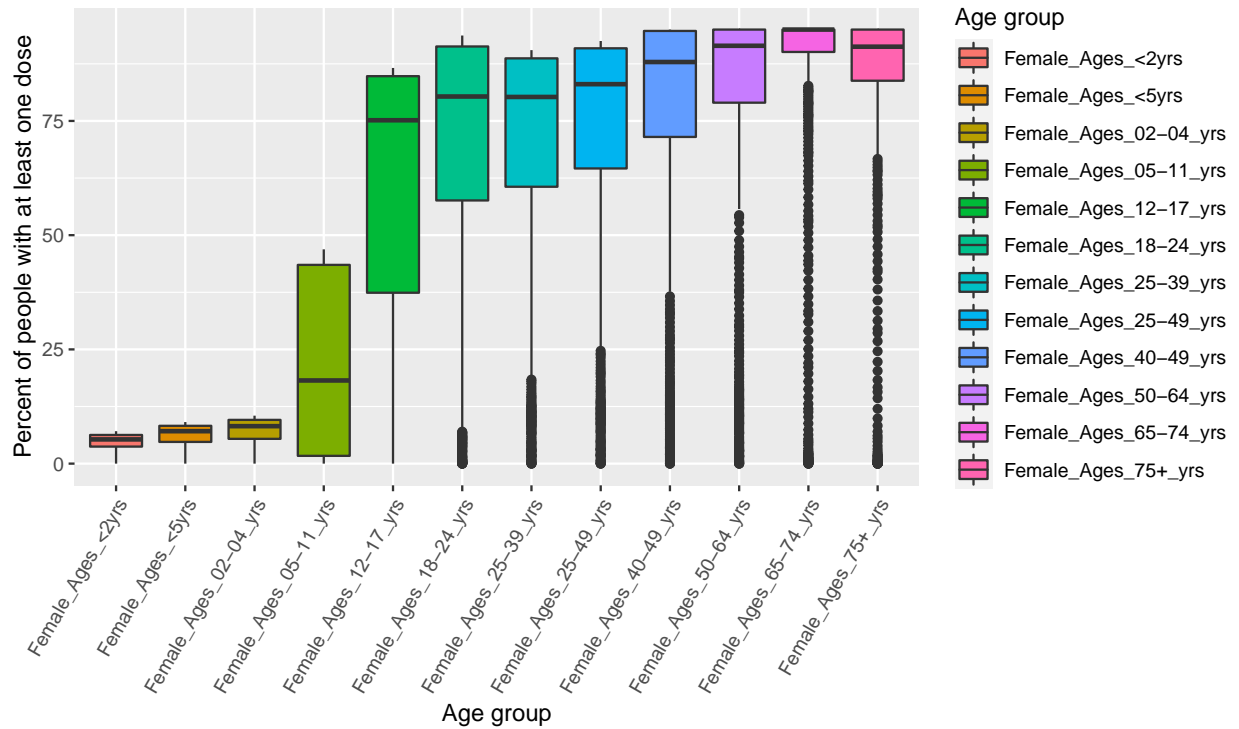


Figure 2.Percent of people with at least one dose of females grouped by age

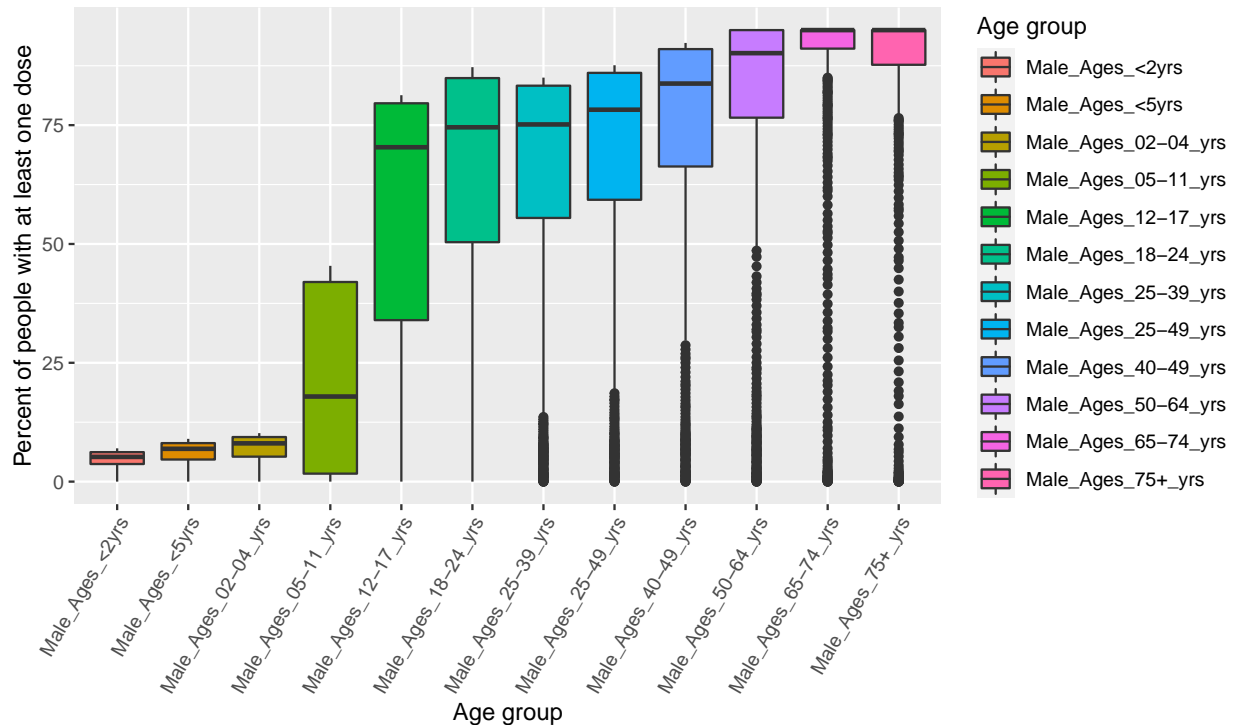


Figure 3.Percent of people with at least one dose of males grouped by age

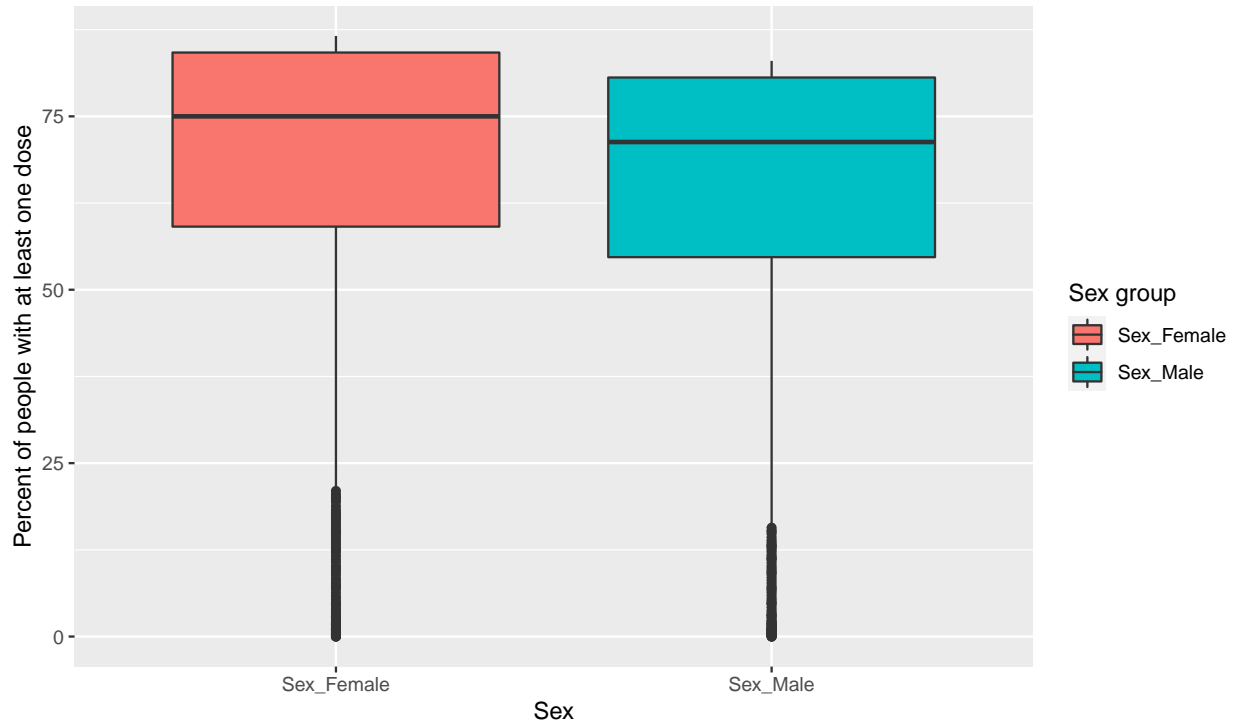


Figure 4.Percent of people with at least one dose grouped by sex

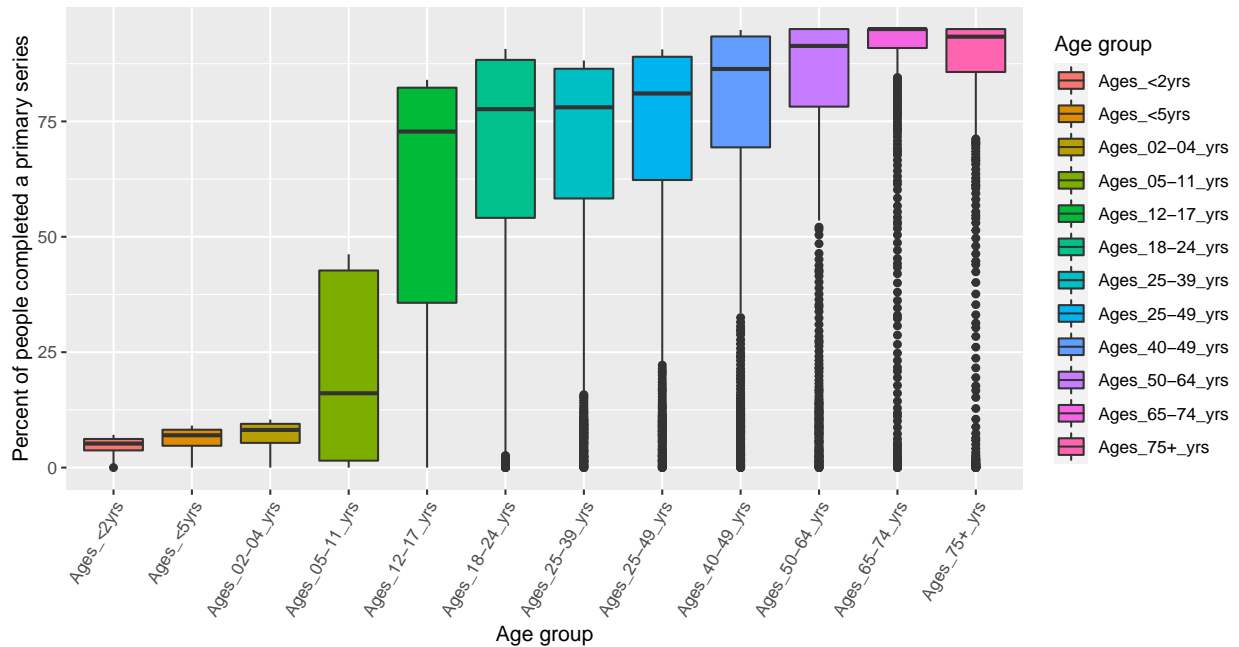


Figure 5.Percent of people completed a primary series grouped by age

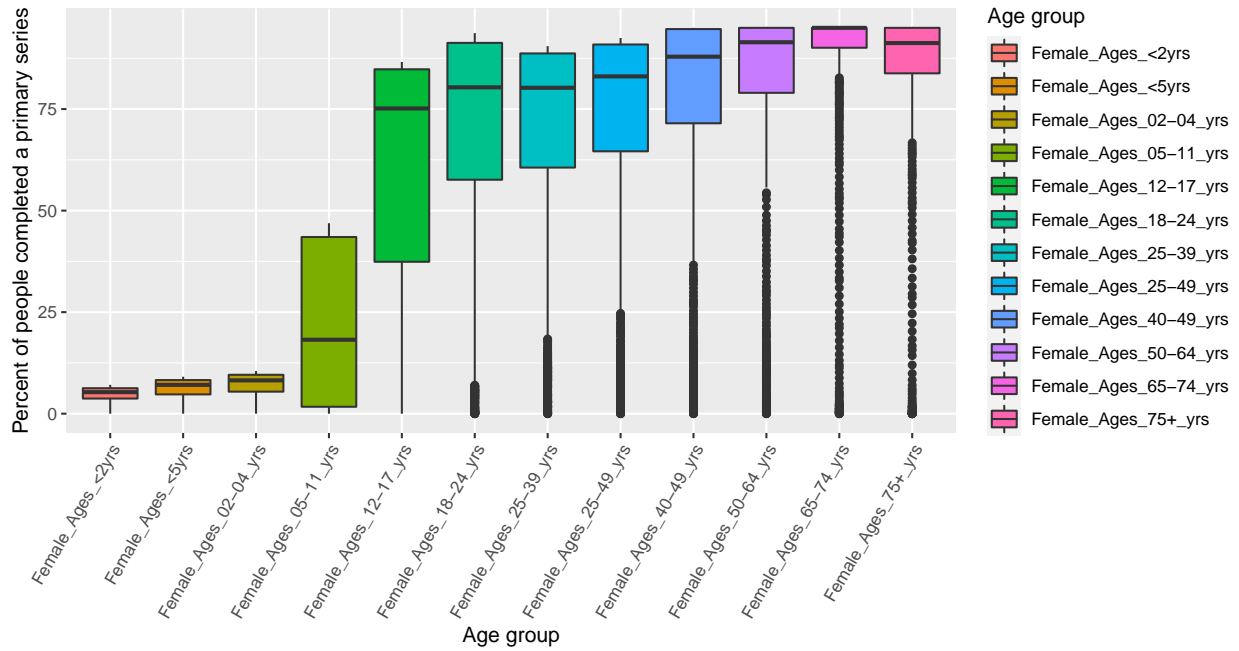


Figure 6.Percent of people completed a primary series of females grouped by age

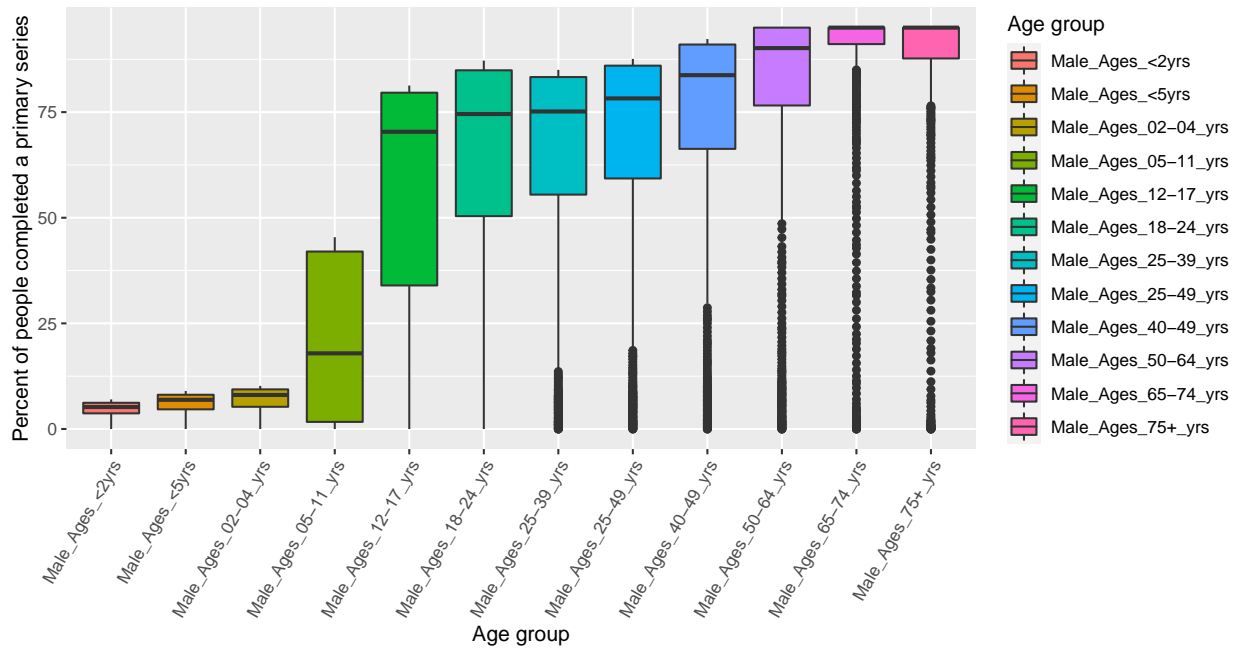


Figure 7.Percent of people completed a primary series of males grouped by age

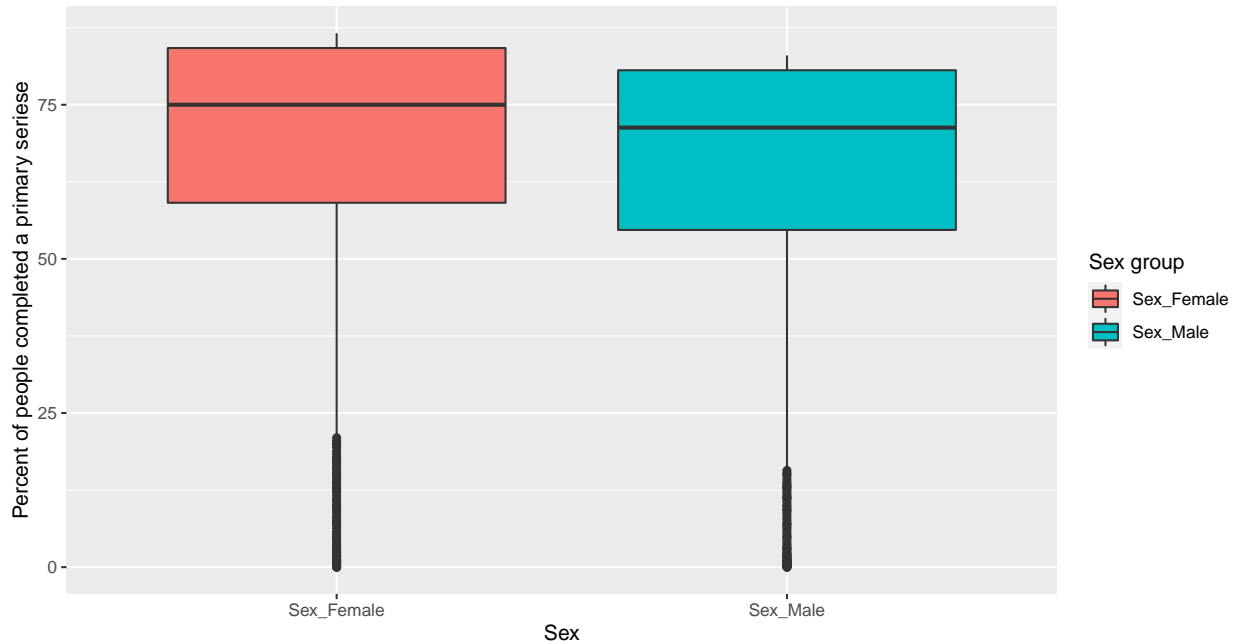


Figure 8.Percent of people completed a primary series grouped by sex

3. Visualization of the association

These 4 figures (Figure 9-12) all verified the trend that 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 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.

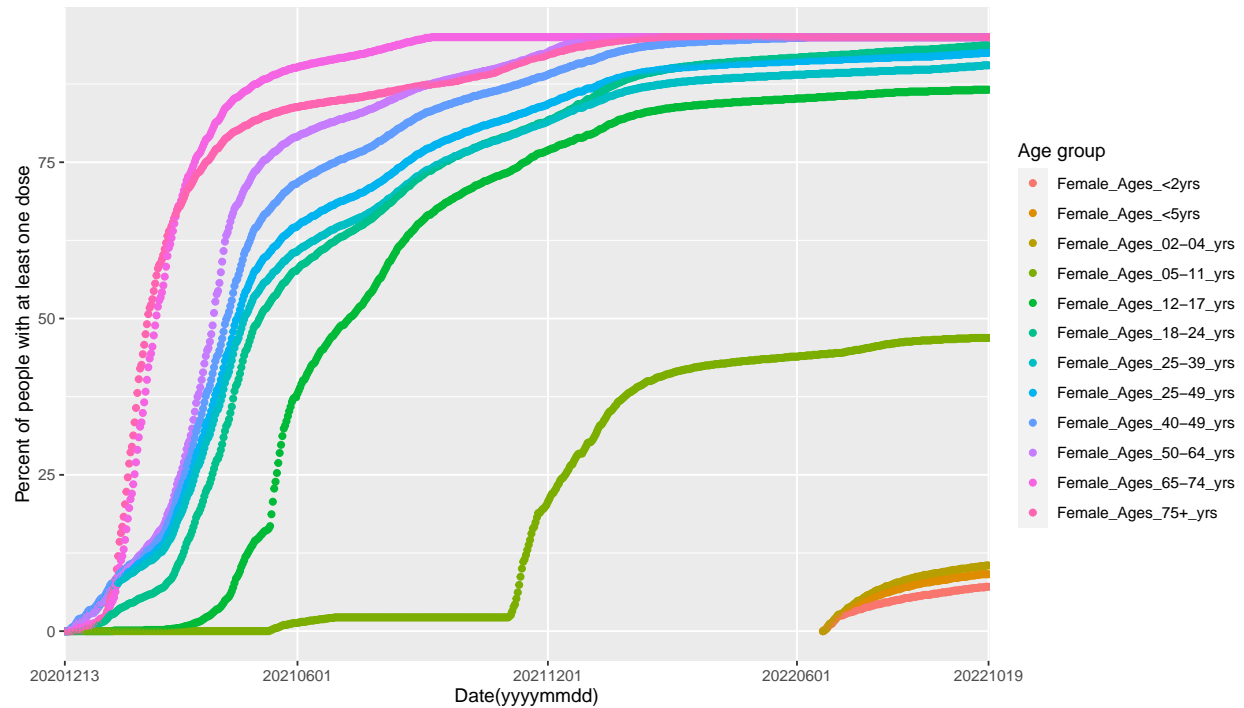


Figure 9.2020–2022 Percent of people with at least one dose of females grouped by age

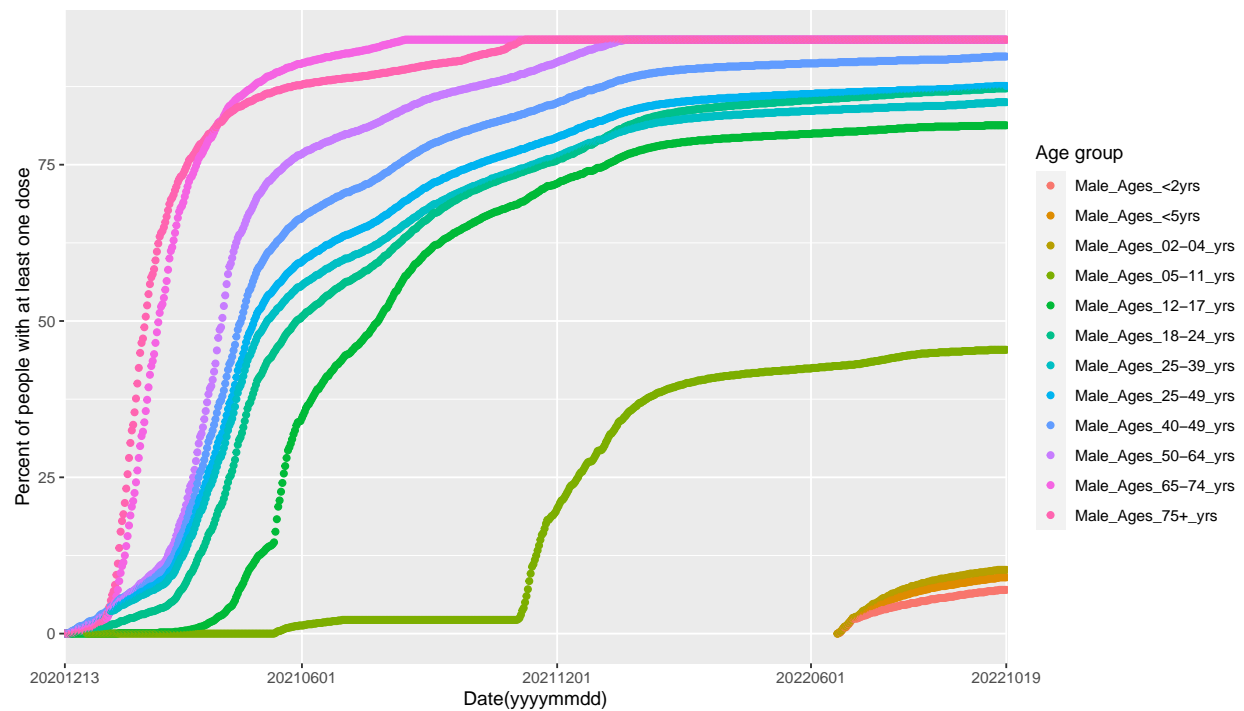


Figure 10.2020–2022 Percent of people with at least one dose of males grouped by age

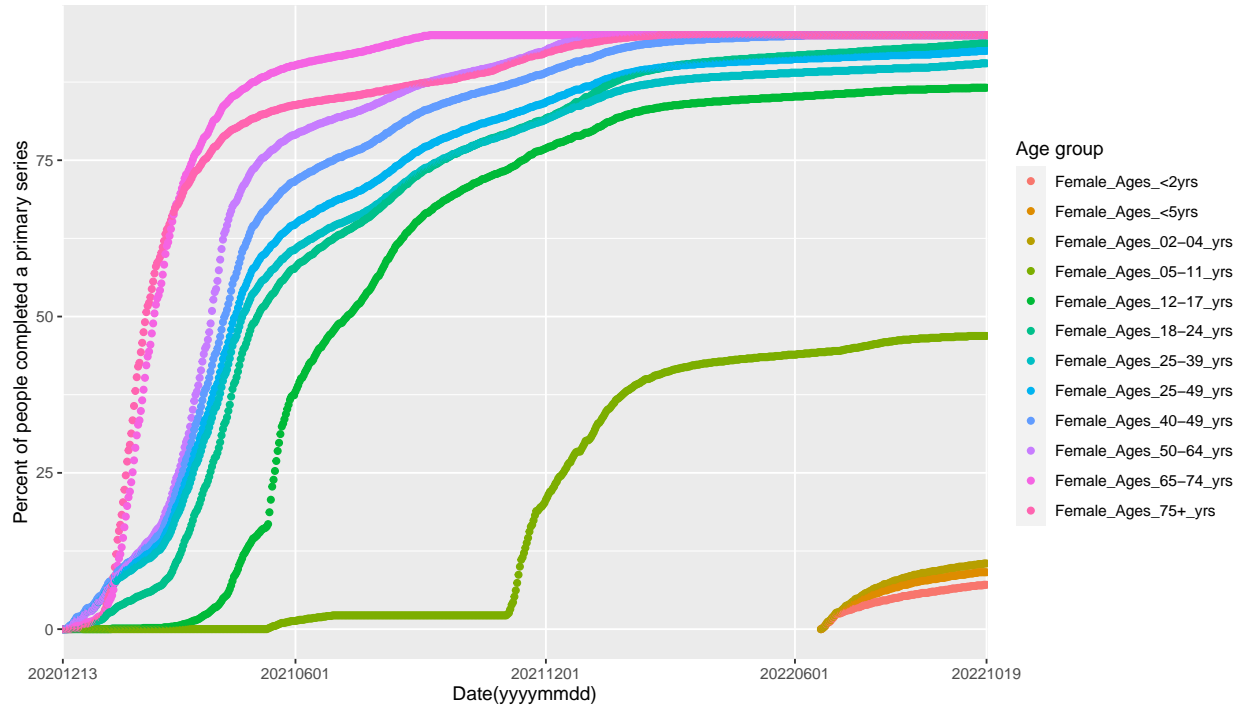


Figure 11.2020–2022 Percent of people completed a primary series of females grouped by age

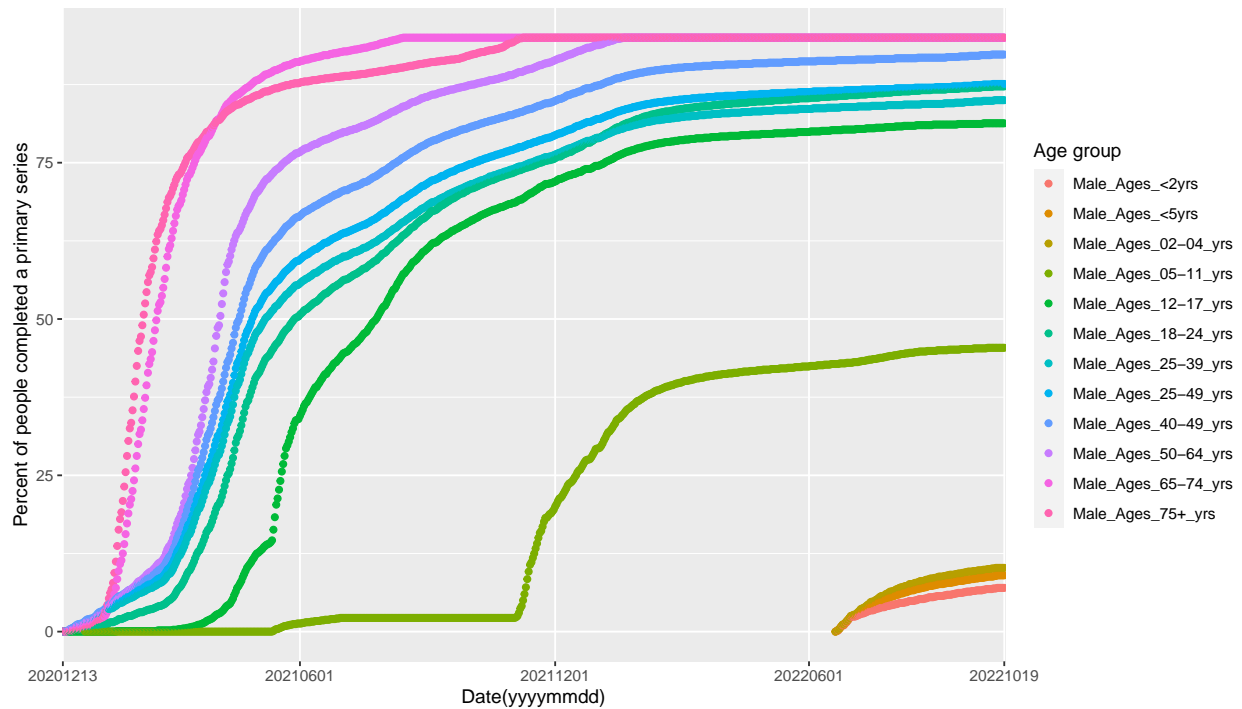


Figure 12.2020–2022 Percent of people completed a primary series of males grouped by age

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.