Covid-19 Vaccine Project

Devanshi

Table of contents

#Importing vax data

```
url <- "https://data.chhs.ca.gov/dataset/ead44d40-fd63-4f9f-950a-3b0111074de8/resource/ec3
vax <- read.csv(url)
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction         county
1 2021-01-05                   91606               Los Angeles    Los Angeles
2 2021-01-05                   95312                    Merced         Merced
3 2021-01-05                   91350               Los Angeles    Los Angeles
4 2021-01-05                   91708            San Bernardino San Bernardino
5 2021-01-05                   95305                  Tuolumne       Tuolumne
6 2021-01-05                   91351               Los Angeles    Los Angeles
  vaccine_equity_metric_quartile                vem_source
1                              1 Healthy Places Index Score
2                              1    CDPH-Derived ZCTA Score
3                              4 Healthy Places Index Score
4                             NA            No VEM Assigned
5                             NA            No VEM Assigned
6                              3 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
```

```
1               38210.0                41964          44295
2                 187.4                  236            276
3               29940.2                33775          36173
4                3517.3                 3794             NA
5                   0.0                    0             NA
6               27874.9                30641          32711
  persons_fully_vaccinated persons_partially_vaccinated
1                       14                          482
2                       NA                           NA
3                       65                         1225
4                       NA                           NA
5                       NA                           NA
6                       31                          644
  percent_of_population_fully_vaccinated
1                               0.000316
2                                     NA
3                               0.001797
4                                     NA
5                                     NA
6                               0.000948
  percent_of_population_partially_vaccinated
1                                   0.010882
2                                         NA
3                                   0.033865
4                                         NA
5                                         NA
6                                   0.019688
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.011198                  NA
2                                     NA                  NA
3                               0.035662                  NA
4                                     NA                  NA
5                                     NA                  NA
6                               0.020636                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                       14
2                        NA                        0
3                        NA                       65
4                        NA                        6
5                        NA                        0
6                        NA                       31
                                                                redacted
1 Information redacted in accordance with CA state privacy requirements
```

```
2 Information redacted in accordance with CA state privacy requirements
3 Information redacted in accordance with CA state privacy requirements
4 Information redacted in accordance with CA state privacy requirements
5 Information redacted in accordance with CA state privacy requirements
6 Information redacted in accordance with CA state privacy requirements
```

Q1. What column details the total number of people fully vaccinated?

persons_fully_vaccinated

Q2. What column details the Zip code tabulation area?

zip_code_tabulation_area

Q3. What is the earliest date in this dataset?

```
min(vax$as_of_date)
```

```
[1] "2021-01-05"
```

Q4. What is the latest date in this dataset?

```
max(vax$as_of_date)
```

```
[1] "2022-11-29"
```

```
skimr::skim(vax)
```

Table 1: Data summary

Name	vax
Number of rows	176400
Number of columns	18
Column type frequency:	
character	5
numeric	13
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
as_of_date	0	1	10	10	0	100	0
local_health_jurisdiction	0	1	0	15	500	62	0
county	0	1	0	15	500	59	0
vem_source	0	1	15	26	0	3	0
redacted	0	1	2	69	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
zip_code_tabulation_area	0	1.00	93665.11	1817.39	90001	92257.75	93658.50	95380.50	97635.0	
vaccine_equity_metric_quartile	8700	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.01	18993.88	0	1346.95	13685.13	31756.18	88556.7	
age5_plus_population	0	1.00	20875.24	21105.98	0	1460.50	15364.00	34877.00	101902.0	
tot_population	8600	0.95	23372.72	22628.51	12	2126.00	18714.00	38168.00	111165.0	
persons_fully_vaccinated	15048	0.91	13504.91	14748.88	11	887.00	8076.00	22588.00	87207.0	
persons_partially_vaccinated	15048	0.91	1707.77	2001.11	11	167.00	1195.00	2547.00	39228.0	
percent_of_population_fully_vaccinated	18834	0.89	0.55	0.25	0	0.40	0.59	0.73	1.0	
percent_of_population_partially_vaccinated	18834	0.89	0.08	0.09	0	0.05	0.06	0.08	1.0	
percent_of_population_with_1_plus_dose	19739	0.89	0.62	0.25	0	0.46	0.65	0.79	1.0	
booster_recip_count	70611	0.60	5643.35	6858.00	11	281.00	2585.00	9377.00	58376.0	
bivalent_dose_recip_count	157094	0.11	1770.66	2315.50	11	117.00	778.00	2643.75	18815.0	
eligible_recipient_count	0	1.00	12345.61	14582.42	0	468.00	5851.00	21198.25	86706.0	

Q5. How many numeric columns are in this dataset?

13

Q6. Note that there are "missing values" in the dataset. How many NA values there in the persons_fully_vaccinated column?

```
sum( is.na(vax$persons_fully_vaccinated) )
```

```
[1] 15048
```

Q7. What percent of persons_fully_vaccinated values are missing (to 2 significant figures)?

```
(sum( is.na(vax$persons_fully_vaccinated) ) / length(vax$persons_fully_vaccinated)) * 100
```

```
[1] 8.530612
```

Q8. [Optional]: Why might this data be missing?

Perhaps some people have taken just one dose of the vaccine and therefore are not fully vaccinated.

Working with dates

```
library(lubridate)
```

```
Loading required package: timechange
```

```
Attaching package: 'lubridate'
```

```
The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
vax$as_of_date <- ymd(vax$as_of_date)
```

```
vax$as_of_date[nrow(vax)] - vax$as_of_date[1]
```

```
Time difference of 693 days
```

Q9. How many days have passed since the last update of the dataset?

```
today() - vax$as_of_date[length(vax$as_of_date)]
```

```
Time difference of 2 days
```

Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)?

```
length(unique(vax$as_of_date))
```

```
[1] 100
```

Working with Zip codes

Install the `zipcodeR` package.

```
library(zipcodeR)

zip_distance('90024','92122')
```

```
  zipcode_a zipcode_b distance
1     90024     92122   109.55
```

Pulling data for all zip codes in this data set

```
zipdata <- reverse_zipcode(vax$zip_code_tabulation_area)
```

Focussing on the San Diego Area

```
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
sd <- filter(vax, county == "San Diego")

nrow(sd)
```

```
[1] 10700
```

Q11. How many distinct zip codes are listed for San Diego County?

```
length(unique(sd$zip_code_tabulation_area))
```

```
[1] 107
```

Q12. What San Diego County Zip code area has the largest 12 + Population in this dataset?

```
sd$zip_code_tabulation_area[match(max(sd$age12_plus_population), sd$age12_plus_population)
```

```
[1] 92154
```

Filtering data as of 2022-11-15.

Q13. What is the overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2022-11-15"?

```
latest_sd <- filter(sd, as_of_date == "2022-11-15")
mean(latest_sd$percent_of_population_fully_vaccinated, na.rm = T)
```

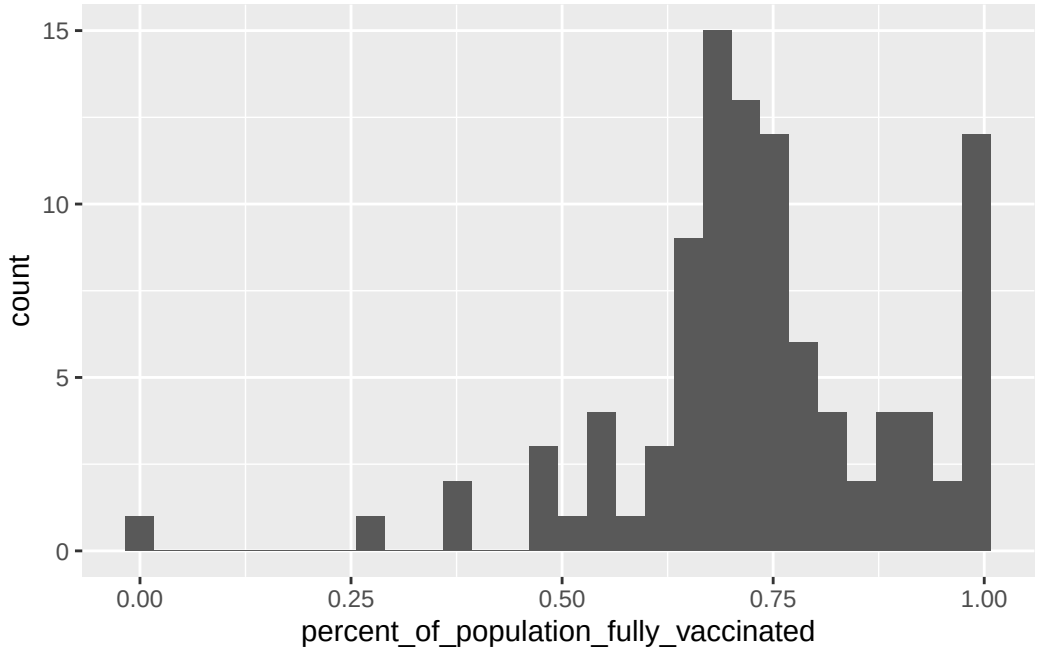
```
[1] 0.7370352
```

Q14. Using either ggplot or base R graphics make a summary figure that shows the distribution of Percent of Population Fully Vaccinated values as of "2022-11-15"?

```
library(ggplot2)
ggplot(latest_sd) + aes(percent_of_population_fully_vaccinated) + geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 8 rows containing non-finite values (`stat_bin()`).
```



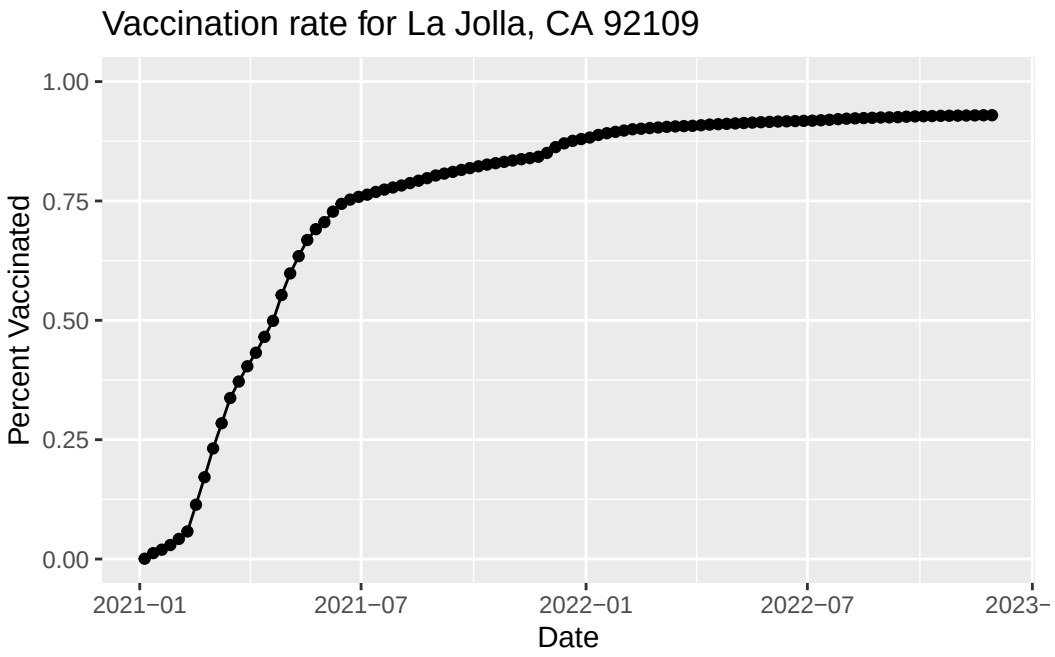
Focusing on UCSD/La Jolla

```
ucsd <- filter(sd, zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

```
[1] 36144
```

Q15. Using ggplot make a graph of the vaccination rate time course for the 92037 ZIP code area:

```
ggplot(ucsd) +
  aes(x = as_of_date, y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(title = "Vaccination rate for La Jolla, CA 92109", x = "Date",
       y="Percent Vaccinated")
```



Comparing to similar sized areas

```
vax_36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2022-11-15")
head(vax_36)
```

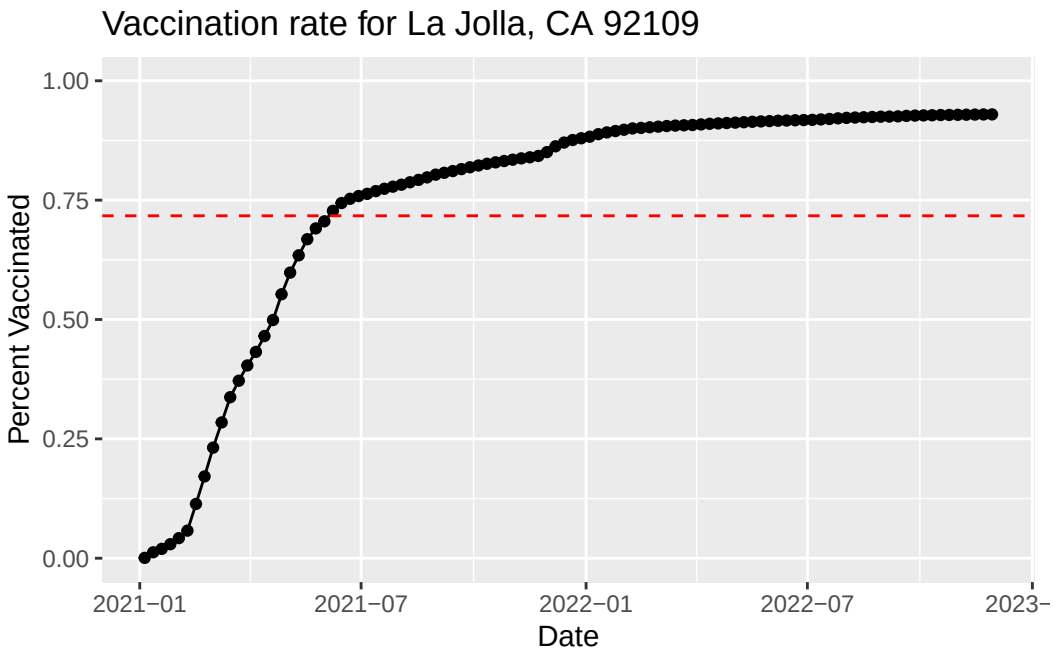
```
  as_of_date zip_code_tabulation_area local_health_jurisdiction        county
1 2022-11-15                   95762                 El Dorado     El Dorado
2 2022-11-15                   94509              Contra Costa  Contra Costa
3 2022-11-15                   94117             San Francisco San Francisco
4 2022-11-15                   94134             San Francisco San Francisco
5 2022-11-15                   93292                    Tulare        Tulare
6 2022-11-15                   93277                    Tulare        Tulare
  vaccine_equity_metric_quartile                vem_source
1                              4 Healthy Places Index Score
2                              2 Healthy Places Index Score
3                              4 Healthy Places Index Score
4                              3 Healthy Places Index Score
5                              1 Healthy Places Index Score
6                              2 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
```

```
1               36212.0                40775          43052
2               57011.2                63454          68166
3               41018.7                42594          44650
4               37371.7                40281          42418
5               32860.1                38048          42031
6               42465.0                47872          52271
  persons_fully_vaccinated persons_partially_vaccinated
1                    36381                         2568
2                    49212                         3647
3                    31064                         3191
4                    40529                         2477
5                    23440                         2456
6                    28825                         2980
  percent_of_population_fully_vaccinated
1                               0.845048
2                               0.721943
3                               0.695722
4                               0.955467
5                               0.557684
6                               0.551453
  percent_of_population_partially_vaccinated
1                                   0.059649
2                                   0.053502
3                                   0.071467
4                                   0.058395
5                                   0.058433
6                                   0.057011
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.904697               23336
2                               0.775445               26763
3                               0.767189               23835
4                               1.000000               28615
5                               0.616117               11359
6                               0.608464               14773
  bivalent_dose_recip_count eligible_recipient_count redacted
1                      7867                    36218       No
2                      6256                    49001       No
3                     10371                    30601       No
4                      7520                    40219       No
5                      2253                    23373       No
6                      3244                    28750       No
```

```
vax_36_mean <- mean(vax_36$percent_of_population_fully_vaccinated,
                    na.rm = T)
```

Q16. Calculate the mean "Percent of Population Fully Vaccinated" for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date "2022-11-15". Add this as a straight horizontal line to your plot from above with the geom_hline() function?

```
ggplot(ucsd) +
  aes(x = as_of_date, y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) + geom_hline(yintercept = vax_36_mean, linetype = "dashed", col = "re
  labs(title = "Vaccination rate for La Jolla, CA 92109",
       x = "Date",
       y="Percent Vaccinated")
```



Q17. What is the 6 number summary (Min, 1st Qu., Median, Mean, 3rd Qu., and Max) of the "Percent of Population Fully Vaccinated" values for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date "2022-11-15"?

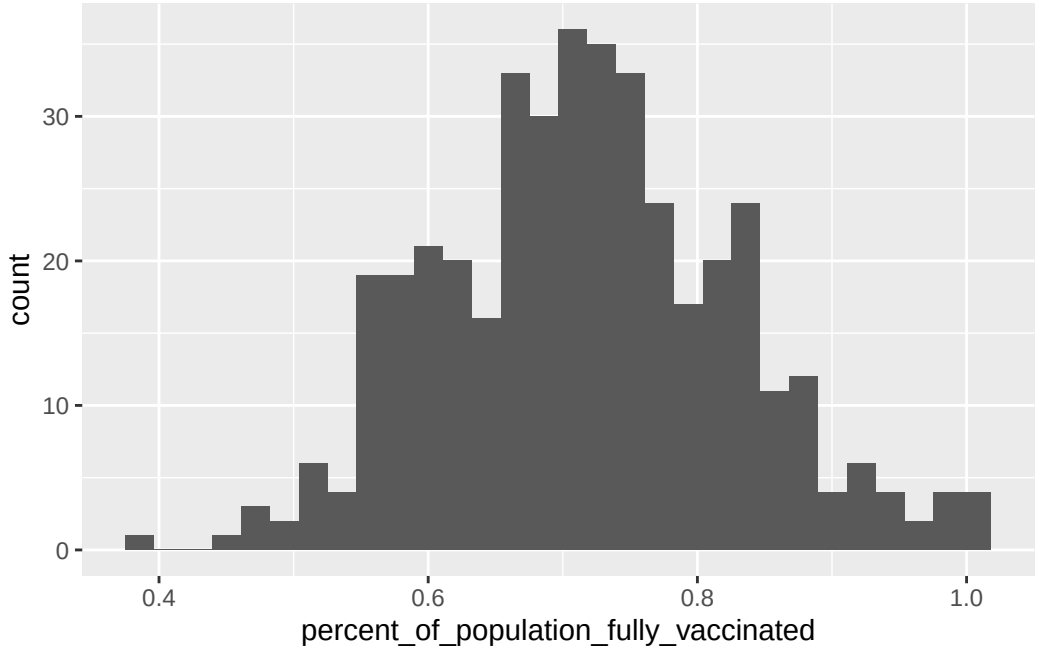
```
summary(vax_36$percent_of_population_fully_vaccinated)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3783  0.6396  0.7157  0.7173  0.7881  1.0000
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax_36) + aes(percent_of_population_fully_vaccinated) +
  geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Q19. Is the 92109 and 92040 ZIP code areas above or below the average value you calculated for all these above?

```
vax %>% filter(as_of_date == "2022-11-15") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                               0.547121
```

```
vax %>% filter(as_of_date == "2022-11-15") %>%
  filter(zip_code_tabulation_area=="92109") %>%
  select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                               0.693129
```

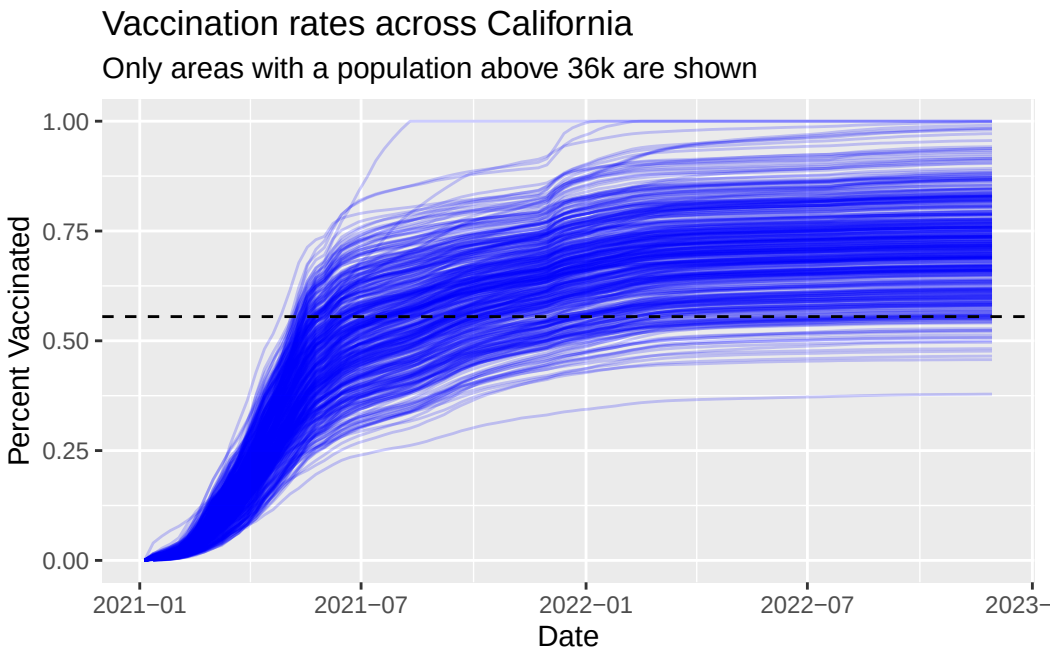
Both of these zipcodes' means are lower than the calculated mean for the La Jolla zipcode.

> Q20. Finally make a time course plot of vaccination progress for all areas in the full dataset with a age5_plus_population > 36144.

```
vax_36_all <- filter(vax, age5_plus_population > 36144)
mean_vax_36_all <- mean(vax_36_all$percent_of_population_fully_vaccinated,
                        na.rm = T)

ggplot(vax_36_all) + aes(x = as_of_date, y = percent_of_population_fully_vaccinated, group
```

```
Warning: Removed 182 rows containing missing values (`geom_line()`).
```



Q21. How do you feel about traveling for Thanksgiving Break and meeting for in-person class afterwards?

Not great. We need to be more careful.