Class 17: Vaccination Rate Mini Project

Elena

Table of contents

Exploring Data

```
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction          county
1 2021-01-05                   92240                 Riverside       Riverside
2 2021-01-05                   91302               Los Angeles     Los Angeles
3 2021-01-05                   93420           San Luis Obispo San Luis Obispo
4 2021-01-05                   91901                 San Diego       San Diego
5 2021-01-05                   94110             San Francisco   San Francisco
6 2021-01-05                   91902                 San Diego       San Diego
  vaccine_equity_metric_quartile                vem_source
1                              1 Healthy Places Index Score
2                              4 Healthy Places Index Score
3                              3 Healthy Places Index Score
4                              3 Healthy Places Index Score
5                              4 Healthy Places Index Score
6                              4 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
1               29270.5                33093          35278
2               23163.9                25899          26712
3               26694.9                29253          30740
4               15549.8                16905          18162
5               64350.7                68320          72380
```

```
6              16620.7                18026          18896
  persons_fully_vaccinated persons_partially_vaccinated
1                       NA                           NA
2                       15                          614
3                       NA                           NA
4                       NA                           NA
5                       17                         1268
6                       15                          397
  percent_of_population_fully_vaccinated
1                                     NA
2                               0.000562
3                                     NA
4                                     NA
5                               0.000235
6                               0.000794
  percent_of_population_partially_vaccinated
1                                         NA
2                                   0.022986
3                                         NA
4                                         NA
5                                   0.017519
6                                   0.021010
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                               0.023548                  NA
3                                     NA                  NA
4                                     NA                  NA
5                               0.017754                  NA
6                               0.021804                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                        2
2                        NA                       15
3                        NA                        4
4                        NA                        8
5                        NA                       17
6                        NA                       15
                                                               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?

2021-01-05

Q4. What is the latest date in this dataset?

2022-11-22 (seems that there are new entries since the latest one noted on 2022-11-15 for the online worksheet)

Data Overview

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

Table 1: Data summary

Name	vax
Number of rows	174636
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	99	0
local_health_jurisdiction	0	1	0	15	495	62	0
county	0	1	0	15	495	59	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
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	8613	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	20875.22	21105.98	0	1460.50	15364.00	34877.00	101902.0	
tot_population	8514	0.95	23372.72	22628.51	12	2126.00	18714.00	38168.00	111165.0	
persons_fully_vaccinated	14921	0.91	13466.31	14722.46	11	883.00	8024.00	22529.00	87186.0	
persons_partially_vaccinated	14921	0.91	1707.50	1998.80	11	167.00	1194.00	2547.00	39204.0	
percent_of_population_fully_vaccinated	18665	0.89	0.55	0.25	0	0.39	0.59	0.73	1.0	
percent_of_population_partially_vaccinated	18665	0.89	0.08	0.09	0	0.05	0.06	0.08	1.0	
percent_of_population_with_1_plus_dose	19762	0.89	0.61	0.25	0	0.46	0.65	0.79	1.0	
booster_recip_count	70421	0.60	5655.17	6867.49	11	280.00	2575.00	9421.00	58304.0	
bivalent_dose_recip_count	156958	0.10	1646.02	2161.84	11	109.00	719.00	2443.00	18109.0	
eligible_recipient_count	0	1.00	12309.11	14555.83	0	466.00	5810.00	21140.00	86696.0	

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?

14921

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

```
[1] 14921
```

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

8.54%

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

```
[1] 0.08544057
```

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

Could be unreported or the data may be difficult to acquire.

Dates

```
library(lubridate)
```

```
today()
```

```
[1] "2022-11-27"
```

```
vax$as_of_date <- ymd(vax$as_of_date)
today() - vax$as_of_date[1]
```

```
Time difference of 691 days
```

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

```
Time difference of 686 days
```

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

5 days

```
today() - max(vax$as_of_date)
```

```
Time difference of 5 days
```

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

99 (98 if including up to 2022-11-15 for the online worksheet)

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

```
[1] 99
```

Working with ZIP codes

```
library(zipcodeR)
library(dplyr)
library(tidyverse)
library(ggplot2)
```

```
geocode_zip('92037')
```

```
# A tibble: 1 x 3
  zipcode   lat   lng
  <chr>   <dbl> <dbl>
1 92037    32.8 -117.
```

```
zip_distance('92037','92109')
```

```
  zipcode_a zipcode_b distance
1     92037     92109     2.33
```

```
reverse_zipcode(c('92037', "92109"))
```

```
# A tibble: 2 x 24
  zipcode zipcode_~1 major~2 post_~3 common_c~4 county state   lat   lng timez~5
  <chr>   <chr>      <chr>   <chr>       <blob> <chr>  <chr> <dbl> <dbl> <chr>
1 92037   Standard   La Jol~ La Jol~ <raw 20 B> San D~ CA     32.8 -117. Pacific
2 92109   Standard   San Di~ San Di~ <raw 21 B> San D~ CA     32.8 -117. Pacific
# ... with 14 more variables: radius_in_miles <dbl>, area_code_list <blob>,
#   population <int>, population_density <dbl>, land_area_in_sqmi <dbl>,
#   water_area_in_sqmi <dbl>, housing_units <int>,
#   occupied_housing_units <int>, median_home_value <int>,
#   median_household_income <int>, bounds_west <dbl>, bounds_east <dbl>,
#   bounds_north <dbl>, bounds_south <dbl>, and abbreviated variable names
#   1: zipcode_type, 2: major_city, 3: post_office_city, ...
```

Focus on SD

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

```
[1] 10593
```

```
sd.10 <- filter(vax, county == "San Diego" &
                age5_plus_population > 10000)
```

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

107

```
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?

92154

```
which.max(sd$age12_plus_population)
```

```
[1] 53
```

```
sd$zip_code_tabulation_area[53]
```

```
[1] 92154
```

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

0.74

```
#Filter for 2022-11-15 data
vax.sd.1115 <- sd %>% filter(as_of_date == "2022-11-15")
```

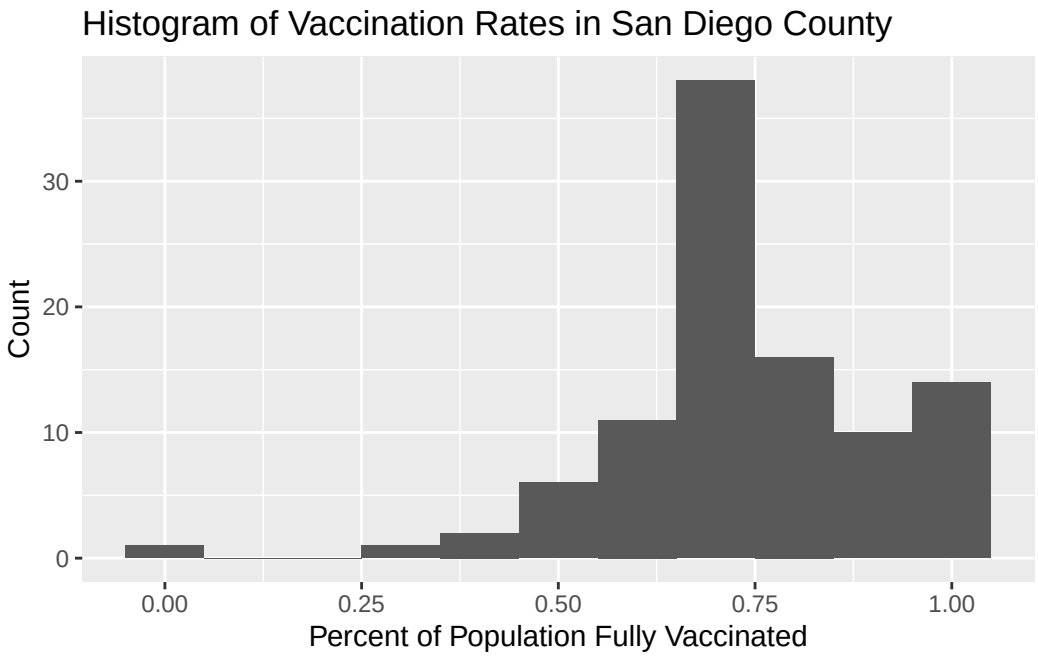
```
mean(vax.sd.1115$percent_of_population_fully_vaccinated, na.rm=T)
```

```
[1] 0.7369099
```

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"?

```
ggplot(vax.sd.1115) +
  aes(percent_of_population_fully_vaccinated) +
  geom_histogram(binwidth=0.1) +
  labs(title="Histogram of Vaccination Rates in San Diego County", x="Percent of Populatio
```

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



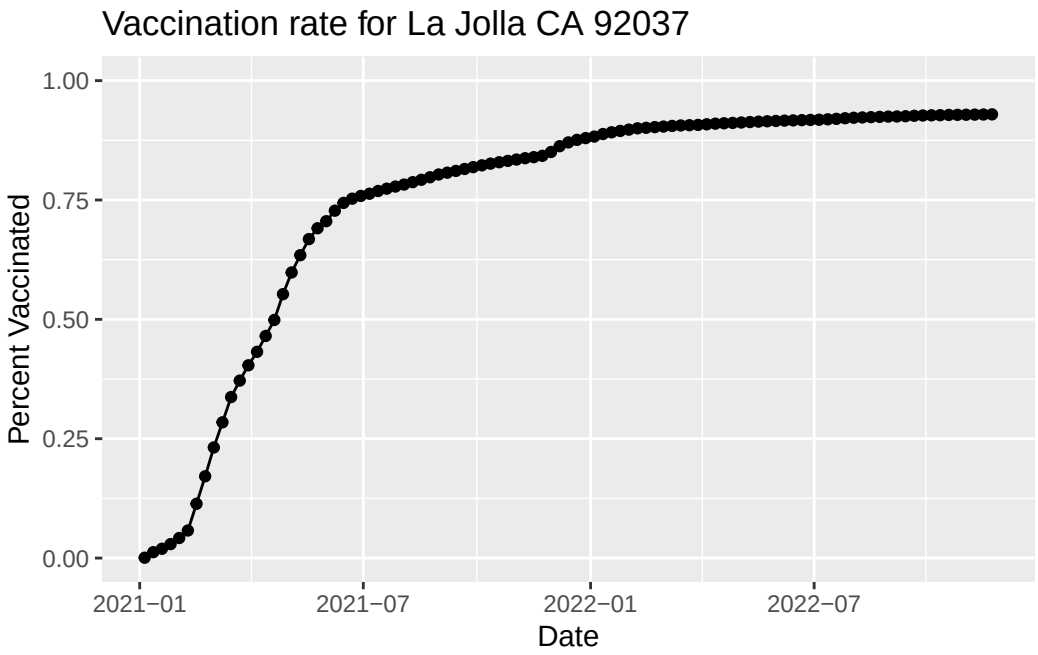
Focus on UCSD/La Jolla

```
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 92037", x="Date", y="Percent Vaccinated")
```



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?

```
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2022-11-15")
head(vax.36)
```

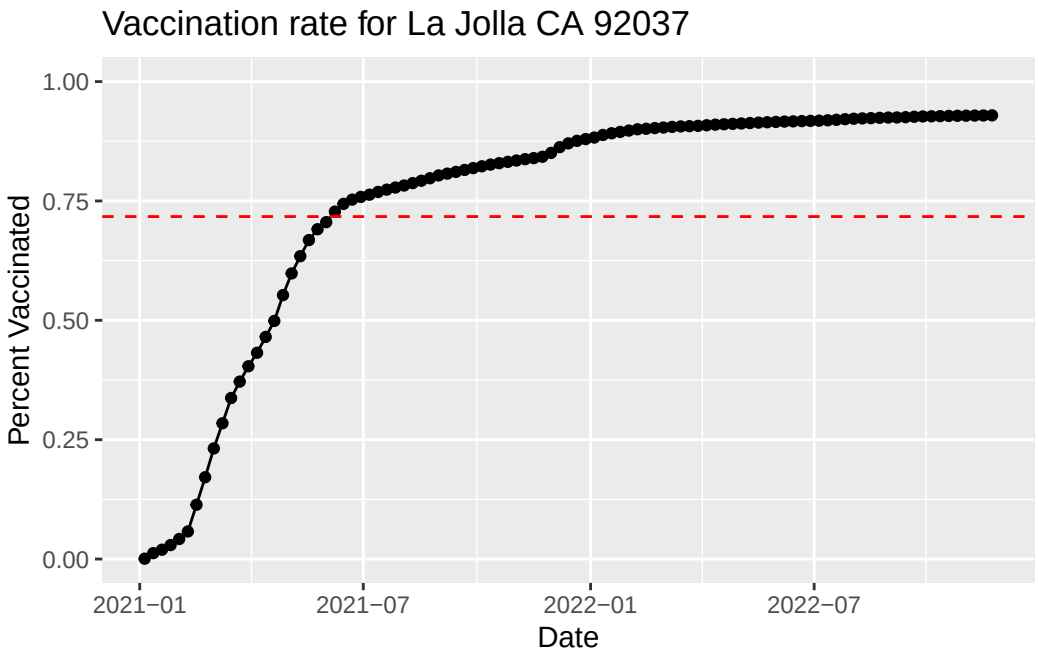
```
  as_of_date zip_code_tabulation_area local_health_jurisdiction        county
1 2022-11-15                   92236                 Riverside     Riverside
2 2022-11-15                   92130                 San Diego     San Diego
3 2022-11-15                   94121           San Francisco San Francisco
4 2022-11-15                   94551                 Alameda       Alameda
5 2022-11-15                   94112           San Francisco San Francisco
6 2022-11-15                   94303             Santa Clara   Santa Clara
  vaccine_equity_metric_quartile                vem_source
1                              1 Healthy Places Index Score
2                              4 Healthy Places Index Score
3                              4 Healthy Places Index Score
4                              4 Healthy Places Index Score
5                              3 Healthy Places Index Score
6                              3 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
1               38505.3                42923          45477
2               46300.3                53102          56134
3               39105.0                41363          43616
4               38947.9                43399          47227
5               75681.8                81107          84707
6               40033.3                44989          48244
  persons_fully_vaccinated persons_partially_vaccinated
1                    30465                         3858
2                    52380                         5751
3                    36566                         2373
4                    32557                         2333
5                    78358                         4646
6                    41275                         4175
  percent_of_population_fully_vaccinated
1                               0.669899
2                               0.933124
3                               0.838362
```

```
4                                0.689373
5                                0.925048
6                                0.855547
  percent_of_population_partially_vaccinated
1                                   0.084834
2                                   0.102451
3                                   0.054407
4                                   0.049400
5                                   0.054848
6                                   0.086539
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.754733               12943
2                               1.000000               34821
3                               0.892769               28345
4                               0.738773               20223
5                               0.979896               56744
6                               0.942086               26288
  bivalent_dose_recip_count eligible_recipient_count redacted
1                      1395                    30375       No
2                     11203                    51780       No
3                     10994                    36013       No
4                      5568                    32234       No
5                     16019                    77580       No
6                      8573                    40853       No
```

```
mean.vax.36 <- mean(vax.36$percent_of_population_fully_vaccinated)
mean.vax.36
```

```
[1] 0.7172851
```

```
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 92037", x="Date", y="Percent Vaccinated") +
  geom_hline(yintercept=mean.vax.36, col="red", lty="dashed")
```



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"?

- Min 0.3785
- First Q 0.6396
- Median 0.7155
- Mean 0.7173
- Third Q 0.7880
- Max 1.0000

```
summary(vax.36$percent_of_population_fully_vaccinated)
```

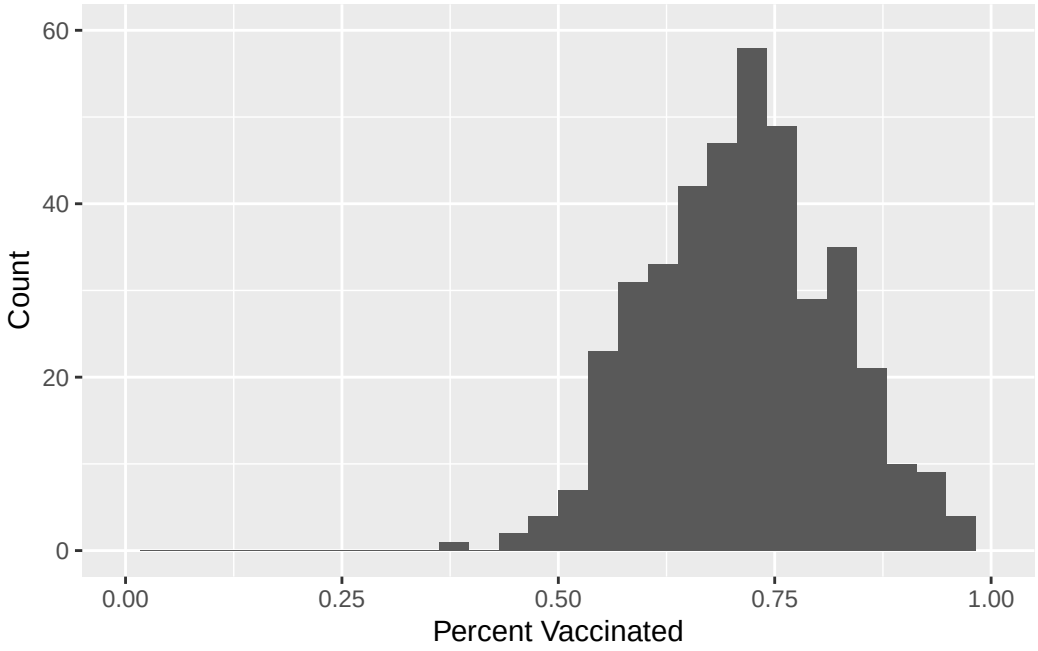
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
 0.3785  0.6396  0.7155  0.7173  0.7880  1.0000 
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36) +
  aes(percent_of_population_fully_vaccinated) +
  geom_histogram() +
  labs(x="Percent Vaccinated", y="Count") +
  xlim(0,1) + ylim(0,60)
```

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

```
Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



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

Both 92109 (0.693299) and 92040 (0.546646) are below the average value (0.7172851)

```
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                               0.693299
```

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

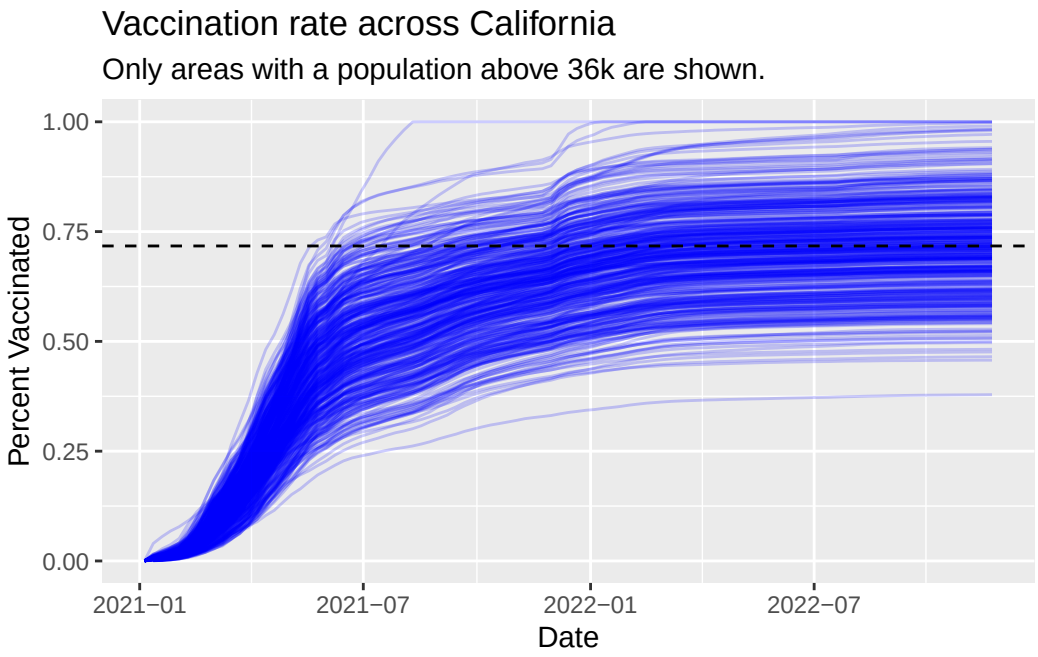
```
  percent_of_population_fully_vaccinated
1                               0.546646
```

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)
```

```
ggplot(vax.36.all) +
  aes(x=as_of_date,
      y=percent_of_population_fully_vaccinated,
      group=zip_code_tabulation_area) +
  geom_line(alpha=0.2, color="blue") +
  ylim(0,1) +
  labs(x="Date", y="Percent Vaccinated",
       title="Vaccination rate across California",
       subtitle="Only areas with a population above 36k are shown.") +
  geom_hline(yintercept = mean.vax.36, linetype="dashed")
```

```
Warning: Removed 184 rows containing missing values (`geom_line()`).
```



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

Not so great! Seems like a lot of the vaccination rates in certain counties are below the average and have plateaued since July 2021. Folks could also travel to other parts of the country/world, where vaccination rates could vary. Finally, vaccination isn't (and shouldn't be) the sole protection against COVID-19 - other methods, such as wearing masks and ventilation, also play an important role in preventing the spread of the virus!