Class 17

Zavier Annis

```
# Import vaccination data
vax <- read.csv("data.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
```

```r
tail(vax)
```

```
       as_of_date zip_code_tabulation_area local_health_jurisdiction
174631 2022-11-22                   94066                San Mateo
```

```
174632 2022-11-22                   92254                Riverside
174633 2022-11-22                   94065                San Mateo
174634 2022-11-22                   92280           San Bernardino
174635 2022-11-22                   94929                    Marin
174636 2022-11-22                   92313           San Bernardino
              county vaccine_equity_metric_quartile                vem_source
174631      San Mateo                              4 Healthy Places Index Score
174632      Riverside                              1 Healthy Places Index Score
174633      San Mateo                              4 Healthy Places Index Score
174634 San Bernardino                             NA            No VEM Assigned
174635          Marin                              4    CDPH-Derived ZCTA Score
174636 San Bernardino                              2 Healthy Places Index Score
       age12_plus_population age5_plus_population tot_population
174631               37730.3                40903          43101
174632                7882.3                 8985           9779
174633               10465.5                11778          12461
174634                   0.0                    0             NA
174635                 174.2                  218            254
174636               10842.9                11847          12547
       persons_fully_vaccinated persons_partially_vaccinated
174631                    38105                         2889
174632                     9456                         1688
174633                    11238                          889
174634                       NA                           NA
174635                       NA                           NA
174636                     7948                          600
       percent_of_population_fully_vaccinated
174631                               0.884086
174632                               0.966970
174633                               0.901854
174634                                     NA
174635                                     NA
174636                               0.633458
       percent_of_population_partially_vaccinated
174631                                   0.067029
174632                                   0.172615
174633                                   0.071343
174634                                         NA
174635                                         NA
174636                                   0.047820
       percent_of_population_with_1_plus_dose booster_recip_count
174631                               0.951115               27085
174632                               1.000000                3840
```

```
174633                                0.973197                8701
174634                                      NA                  NA
174635                                      NA                  NA
174636                                0.681278                4522
       bivalent_dose_recip_count eligible_recipient_count
174631                      9127                    37620
174632                       372                     9430
174633                      3456                    11021
174634                        NA                       14
174635                        NA                      159
174636                      1085                     7921
                                                                redacted
174631                                                                No
174632                                                                No
174633                                                                No
174634 Information redacted in accordance with CA state privacy requirements
174635 Information redacted in accordance with CA state privacy requirements
174636                                                                No
```

Q1. The "persons_fully_vaccinated" column details the total number of people fully vaccinated.

Q2. The zip_code_tabulation_area details the Zip code tabulation area.

Q3. The earliest data in this dataset is 2021-01-05.

Q4. The latest date in this dataset AS OF November 21st is 2022-11-15. However, the dataset has been updated and the latest data is now 2022-11-22.

```
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
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	31756.12	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	67186.0	
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	19362	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. There are 13 numeric columns in this dataset.

Q6. As of 11/21, there are 15,440 NA values in the "persons_fully_vaccinated" column.

Q7. As of 11/21, 8.93% of the "persons_fully_vaccinated" values are missing.

Q8. This data is likely missing because the associated county either lacks the public health data to report or is behind / not currently reporting it to the federal or state authorities.

```
library(lubridate)
```

```
Warning: package 'lubridate' was built under R version 4.2.2
```

```
Loading required package: timechange
```

```
Warning: package 'timechange' was built under R version 4.2.2
```

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

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

    date, intersect, setdiff, union
```

```
today()
```

```
[1] "2022-11-28"
```

```
# Specify that we are using the year-month-day format
vax$as_of_date <- ymd(vax$as_of_date)
today() - vax$as_of_date[1]
```

```
Time difference of 692 days
```

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

```
Time difference of 686 days
```

```
today() - vax$as_of_date[174636]
```

```
Time difference of 6 days
```

Q9. As of today, Monday 11/28, 6 days have passed since the last update on 11/22.

Q10. As of 11/22, there are 98 unique dates in the dataset (taken from n_unique value in skimr data). As of 11/28, there are 99 unique dates.

```
library(zipcodeR)
```

```
Warning: package 'zipcodeR' was built under R version 4.2.2
```

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

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

```
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, ...
```

```
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] 10593
```

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

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

```
[1] 107
```

Q11. There are 107 distinct zip codes listed for San Diego county.

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

```
[1] 53
```

```
sd[53,]
```

```
   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
53 2021-01-05                   92154                San Diego San Diego
   vaccine_equity_metric_quartile                vem_source
53                              2 Healthy Places Index Score
   age12_plus_population age5_plus_population tot_population
53               76365.2                82971          88979
   persons_fully_vaccinated persons_partially_vaccinated
53                       17                         1379
   percent_of_population_fully_vaccinated
53                               0.000191
   percent_of_population_partially_vaccinated
53                                   0.015498
   percent_of_population_with_1_plus_dose booster_recip_count
53                               0.015689                  NA
   bivalent_dose_recip_count eligible_recipient_count
53                        NA                       17
                                                                  redacted
53 Information redacted in accordance with CA state privacy requirements
```

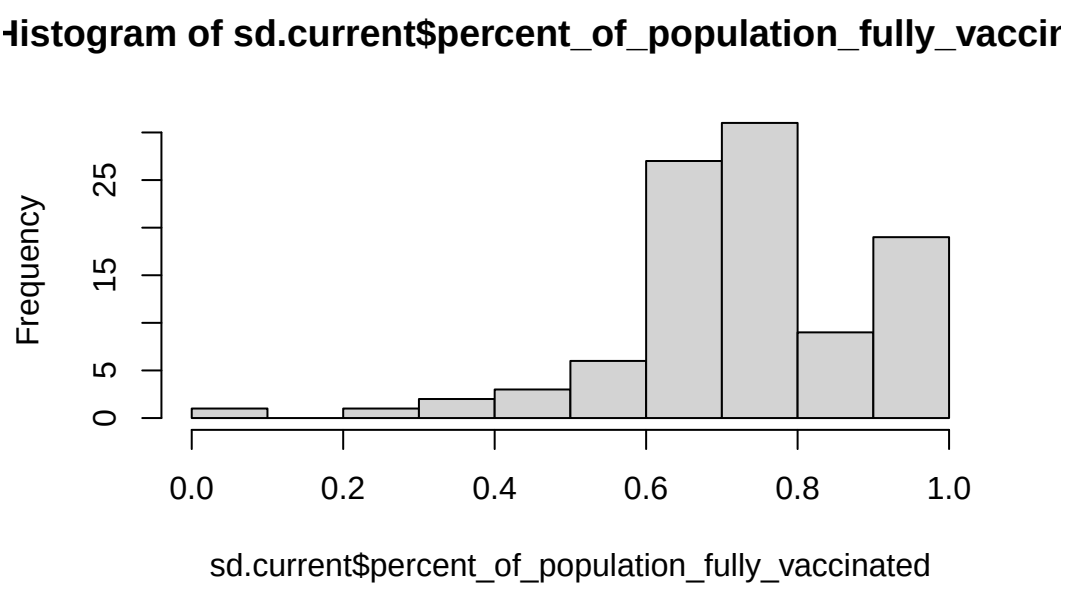
Q12. The 92154 zip code has the largst 12+ population for San Diego county in this dataset.

```
sd.current <- filter(vax, county == "San Diego" &
                 as_of_date == "2022-11-15")
mean(sd.current$percent_of_population_fully_vaccinated, na.rm = TRUE)
```

```
[1] 0.7369099
```

Q13. The overall average value for "Percent of Population Fully Vaccinated" is 73.69% as of 2022-11-15. NOTE THAT VALUE MAY BE SLIGHTLY DIFFERENT BECAUSE "as_of_date" DATA CAN BE RETROACTIVELY UPDATED FOR 2022-11-15.

```
hist(sd.current$percent_of_population_fully_vaccinated)
```




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

```
[1] 36144
```

Q15.

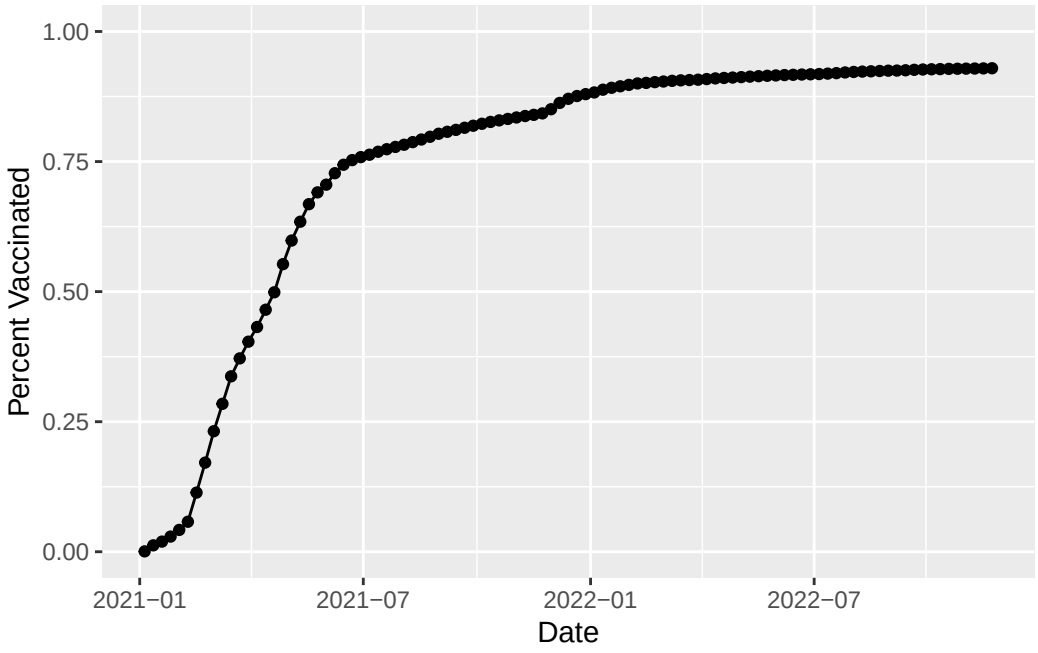
```
library(ggplot2)
ggplot(ucsd) +
  aes(x=ucsd$as_of_date,
      y=ucsd$percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
```

```
Warning: Use of `ucsd$as_of_date` is discouraged. Use `as_of_date` instead.
```

```
Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
Use `percent_of_population_fully_vaccinated` instead.
```

```
Warning: Use of `ucsd$as_of_date` is discouraged. Use `as_of_date` instead.
```

```
Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
Use `percent_of_population_fully_vaccinated` instead.
```



```r
# Subset to all CA areas with a population as large as 92037
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
```

Q16.

```
mean(vax.36$percent_of_population_fully_vaccinated, na.rm = TRUE)
```

```
[1] 0.7172851
```

NOTE THAT VALUE ABOVE MAY BE SLIGHTLY DIFFERENT DUE TO RETROACTIVE UPDATES OF THE 2022-11-15 DATE.

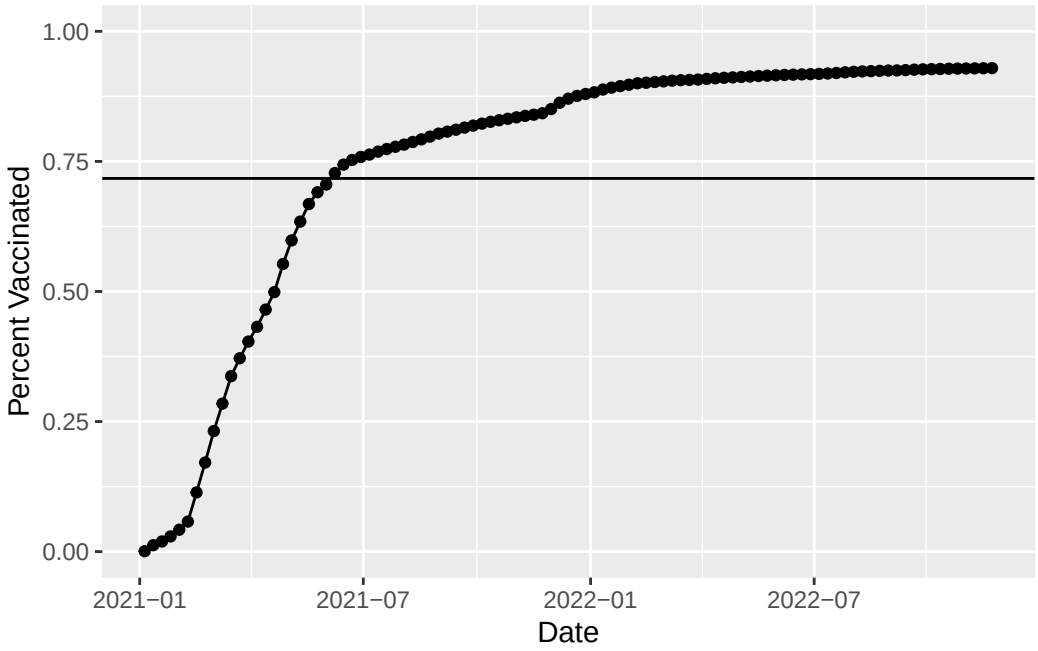
```
ggplot(ucsd) +
  aes(x=ucsd$as_of_date,
      y=ucsd$percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  geom_hline(yintercept = 0.7172851) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
```

```
Warning: Use of `ucsd$as_of_date` is discouraged. Use `as_of_date` instead.
```

```
Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
Use `percent_of_population_fully_vaccinated` instead.

Warning: Use of `ucsd$as_of_date` is discouraged. Use `as_of_date` instead.

Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
Use `percent_of_population_fully_vaccinated` instead.
```




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

```
[1] 0.3785010 0.6396185 0.7155240 0.7879820 1.0000000
```

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

```
[1] 0.7172851
```

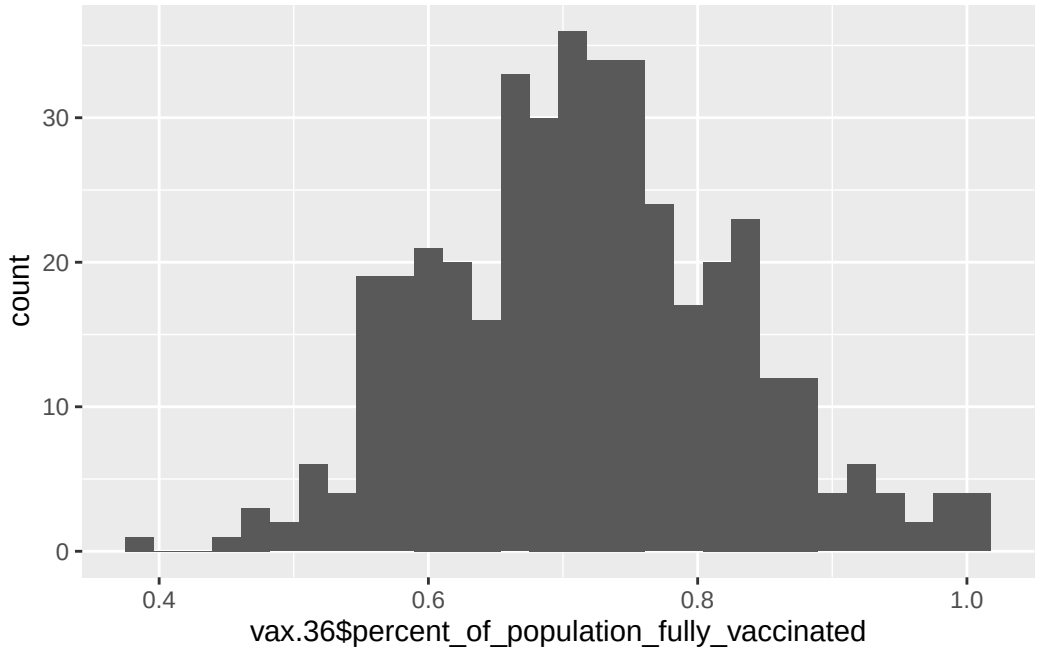
Q17. The six-number summary is displayed in the above code chunk in the following order: minimum, first quartile, median, third quartile, max, and mean on the second line.

Q18.

```
ggplot(vax.36) +
  aes(x=vax.36$percent_of_population_fully_vaccinated) +
  geom_histogram()
```

```
Warning: Use of `vax.36$percent_of_population_fully_vaccinated` is discouraged.
Use `percent_of_population_fully_vaccinated` instead.
```

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




```
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
```

```
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
```

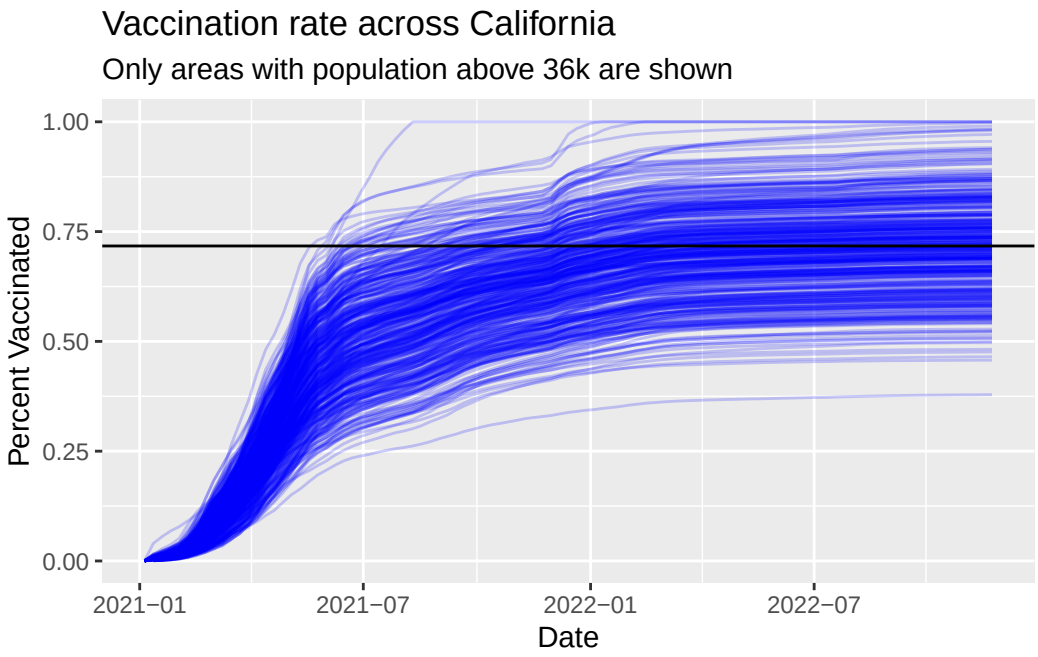
Q19. Both the 92109 and 92040 ZIP code areas are below the average value calculated for all these above.

Q20.

```
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 population above 36k are shown") +
  geom_hline(yintercept = 0.7172851)
```

```
Warning: Removed 184 row(s) containing missing values (geom_path).
```



Q21. Yes, I am very comfortable returning to class in-person after break!