Class 17

Arturo Agramont

Covid-19 Vaccine data

We will start by loading the data

```
vax = read.csv("covid19.csv")
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction         county
1 2021-01-05                   92840                    Orange         Orange
2 2021-01-05                   93662                    Fresno         Fresno
3 2021-01-05                   92310            San Bernardino San Bernardino
4 2021-01-05                   91911                 San Diego      San Diego
5 2021-01-05                   95329                  Tuolumne       Tuolumne
6 2021-01-05                   93668                    Fresno         Fresno
  vaccine_equity_metric_quartile                vem_source
1                              2 Healthy Places Index Score
2                              1 Healthy Places Index Score
3                              1 Healthy Places Index Score
4                              2 Healthy Places Index Score
5                              2 Healthy Places Index Score
6                              1    CDPH-Derived ZCTA Score
  age12_plus_population age5_plus_population tot_population
1               47302.5                51902          54735
2               24501.3                28311          30725
3                6804.4                 8254           9872
4               71642.8                79225          84026
5                2252.1                 2399           2570
6                1013.4                 1199           1219
  persons_fully_vaccinated persons_partially_vaccinated
1                       NA                           NA
2                       NA                           NA
```

```
3                    NA                          NA
4                    29                        1429
5                    NA                          NA
6                    NA                          NA
  percent_of_population_fully_vaccinated
1                                     NA
2                                     NA
3                                     NA
4                               0.000345
5                                     NA
6                                     NA
  percent_of_population_partially_vaccinated
1                                         NA
2                                         NA
3                                         NA
4                                   0.017007
5                                         NA
6                                         NA
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                                     NA                  NA
3                                     NA                  NA
4                               0.017352                  NA
5                                     NA                  NA
6                                     NA                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                        9
2                        NA                        1
3                        NA                        1
4                        NA                       29
5                        NA                        0
6                        NA                        0
  eligible_bivalent_recipient_count
1                                 9
2                                 1
3                                 0
4                                29
5                                 0
6                                 0
                                                             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?

The column that details the number of people fully vaccinated is: persons_fully_vaccinated

- **Q2.** What column details the Zip code tabulation area?

the column that details this information is zip_code_tabulation_area

- **Q3.** What is the earliest date in this dataset?

The earliest date in this data set is January 5th, 2021.

- **Q4.** What is the latest date in this dataset?

```
tail(vax)
```

```
       as_of_date zip_code_tabulation_area local_health_jurisdiction
224023 2023-06-06                   95124               Santa Clara
224024 2023-06-06                   95304               San Joaquin
224025 2023-06-06                   94608                   Alameda
224026 2023-06-06                   95111               Santa Clara
224027 2023-06-06                   92543                 Riverside
224028 2023-06-06                   95110               Santa Clara
            county vaccine_equity_metric_quartile                vem_source
224023 Santa Clara                               4 Healthy Places Index Score
224024 San Joaquin                               3 Healthy Places Index Score
224025     Alameda                               3 Healthy Places Index Score
224026 Santa Clara                               2 Healthy Places Index Score
224027   Riverside                               1 Healthy Places Index Score
224028 Santa Clara                               2 Healthy Places Index Score
       age12_plus_population age5_plus_population tot_population
224023               42989.9                48326          51455
224024               12620.5                13832          14282
224025               28371.8                29534          31013
224026               51458.6                57367          61830
224027               30702.6                35154          38314
224028               16943.3                18597          19928
       persons_fully_vaccinated persons_partially_vaccinated
224023                    43953                         3072
224024                     9589                          891
```

```
       25413                          2494
       51423                          3915
       20016                          3313
       17277                          1842
       percent_of_population_fully_vaccinated
                                    0.854203
                                    0.671405
                                    0.819431
                                    0.831684
                                    0.522420
                                    0.866971
       percent_of_population_partially_vaccinated
                                        0.059703
                                        0.062386
                                        0.080418
                                        0.063319
                                        0.086470
                                        0.092433
       percent_of_population_with_1_plus_dose booster_recip_count
                                    0.913906               32790
                                    0.733791                5482
                                    0.899849               18568
                                    0.895003               32999
                                    0.608890                9564
                                    0.959404               11016
       bivalent_dose_recip_count eligible_recipient_count
                           16667                    43653
                            2068                     9570
                           10348                    25294
                           10899                    51255
                            3278                    20008
                            4455                    17193
       eligible_bivalent_recipient_count redacted
                                   43653       No
                                    9570       No
                                   25294       No
                                   51255       No
                                       0       No
                                   17193       No
```

The latest data in this data set is June 6th, 2023

We will now call skimr for an overview of the data

```
skimr::skim_without_charts(vax)
```

Table 1: Data summary

Name	vax
Number of rows	224028
Number of columns	19
_____	
Column type frequency:	
character	5
numeric	14
_____	
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	127	0
local_health_jurisdiction	0	1	0	15	635	62	0
county	0	1	0	15	635	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
zip_code_tabulation_area	0	1.00	93665.11	817.38	90001	92257.75	93658.50	95380.50	97635.0
vaccine_equity_metric_quartile	11040	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0
age12_plus_population	0	1.00	18895.04	18993.87	0	1346.95	13685.10	31756.12	88556.7
age5_plus_population	0	1.00	20875.22	21105.96	0	1460.50	15364.00	34877.00	101902.0
tot_population	10922	0.95	23372.77	22628.50	12	2126.00	18714.00	38168.00	111165.0
persons_fully_vaccinated	17966	0.92	14323.68	15297.82	11	958.25	9069.00	23853.75	77720.0
persons_partially_vaccinated	17966	0.92	1712.96	2078.76	11	164.00	1205.00	2553.00	44088.0
percent_of_population_fully_vaccinated	22883	0.90	0.58	0.25	0	0.44	0.62	0.75	1.0
percent_of_population_partially_vaccinated	22883	0.90	0.08	0.09	0	0.05	0.06	0.08	1.0
percent_of_population_with_1_plus_dose	24051	0.89	0.65	0.24	0	0.50	0.68	0.82	1.0
booster_recip_count	74675	0.67	6458.34	7836.24	11	334.00	3167.00	10406.00	60080.0
bivalent_dose_recip_count	160239	0.28	3468.29	4058.09	11	228.00	1897.00	5576.00	29720.0
eligible_recipient_count	0	1.00	13168.70	15160.56	0	537.00	6715.50	22595.25	87451.0

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
eligible_bivalent_recipient_count	0	1.00	13056.95	15236.74	0	258.00	6603.00	22585.00	87451.0

Q5. How many numeric columns are in this dataset?

There are 14 numeric columns in this data set.

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] 17966
```

There are 17966 NA values

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

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

```
[1] 8.019533
```

8.0% of persons_fully_vaccinated values are missing

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

This could be because not all of the patients or people in the study are fully vaccinated.

Working with dates

Using lubridate

```
library(lubridate)
```

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

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

    date, intersect, setdiff, union
```

```
today()
```

```
[1] "2023-06-13"
```

Converting vax column of dates into year month day format in order to plot

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

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

```
Time difference of 889 days
```

Then we can determine the days between the data set

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

```
Time difference of 882 days
```

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

```
table(vax$as_of_date)
```

```
2021-01-05 2021-01-12 2021-01-19 2021-01-26 2021-02-02 2021-02-09 2021-02-16
      1764       1764       1764       1764       1764       1764       1764
2021-02-23 2021-03-02 2021-03-09 2021-03-16 2021-03-23 2021-03-30 2021-04-06
      1764       1764       1764       1764       1764       1764       1764
2021-04-13 2021-04-20 2021-04-27 2021-05-04 2021-05-11 2021-05-18 2021-05-25
      1764       1764       1764       1764       1764       1764       1764
2021-06-01 2021-06-08 2021-06-15 2021-06-22 2021-06-29 2021-07-06 2021-07-13
      1764       1764       1764       1764       1764       1764       1764
2021-07-20 2021-07-27 2021-08-03 2021-08-10 2021-08-17 2021-08-24 2021-08-31
      1764       1764       1764       1764       1764       1764       1764
```

```
2021-09-07 2021-09-14 2021-09-21 2021-09-28 2021-10-05 2021-10-12 2021-10-19
      1764       1764       1764       1764       1764       1764       1764
2021-10-26 2021-11-02 2021-11-09 2021-11-16 2021-11-23 2021-11-30 2021-12-07
      1764       1764       1764       1764       1764       1764       1764
2021-12-14 2021-12-21 2021-12-28 2022-01-04 2022-01-11 2022-01-18 2022-01-25
      1764       1764       1764       1764       1764       1764       1764
2022-02-01 2022-02-08 2022-02-15 2022-02-22 2022-03-01 2022-03-08 2022-03-15
      1764       1764       1764       1764       1764       1764       1764
2022-03-22 2022-03-29 2022-04-05 2022-04-12 2022-04-19 2022-04-26 2022-05-03
      1764       1764       1764       1764       1764       1764       1764
2022-05-10 2022-05-17 2022-05-24 2022-05-31 2022-06-07 2022-06-14 2022-06-21
      1764       1764       1764       1764       1764       1764       1764
2022-06-28 2022-07-05 2022-07-12 2022-07-19 2022-07-26 2022-08-02 2022-08-09
      1764       1764       1764       1764       1764       1764       1764
2022-08-16 2022-08-23 2022-08-30 2022-09-06 2022-09-13 2022-09-20 2022-09-27
      1764       1764       1764       1764       1764       1764       1764
2022-10-04 2022-10-11 2022-10-18 2022-10-25 2022-11-01 2022-11-08 2022-11-15
      1764       1764       1764       1764       1764       1764       1764
2022-11-22 2022-11-29 2022-12-06 2022-12-13 2022-12-20 2022-12-27 2023-01-03
      1764       1764       1764       1764       1764       1764       1764
2023-01-10 2023-01-17 2023-01-24 2023-01-31 2023-02-07 2023-02-14 2023-02-21
      1764       1764       1764       1764       1764       1764       1764
2023-02-28 2023-03-07 2023-03-14 2023-03-21 2023-03-28 2023-04-04 2023-04-11
      1764       1764       1764       1764       1764       1764       1764
2023-04-18 2023-04-25 2023-05-02 2023-05-09 2023-05-16 2023-05-23 2023-05-30
      1764       1764       1764       1764       1764       1764       1764
2023-06-06
      1764
```

7 days have passed between updates.

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

```
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
```

```
n_distinct(vax$as_of_date)
```

```
[1] 127
```

There are 127 unique dates in this data set

Working with zip codes

We are going to use zipcodeR package

```
library(zipcodeR)
```

```
The legacy packages maptools, rgdal, and rgeos, underpinning this package
will retire shortly. Please refer to R-spatial evolution reports on
https://r-spatial.org/r/2023/05/15/evolution4.html for details.
This package is now running under evolution status 0
```

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

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

calculating distance between zip codes

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

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

Collecting census data of the zip codes:

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

```
# A tibble: 2 x 24
  zipcode zipcode_type major_city post_office_city common_city_list county state
  <chr>   <chr>        <chr>      <chr>                     <blob> <chr>  <chr>
1 92037   Standard     La Jolla   La Jolla, CA           <raw 20 B> San D~ CA
2 92109   Standard     San Diego  San Diego, CA          <raw 21 B> San D~ CA
# i 17 more variables: lat <dbl>, lng <dbl>, timezone <chr>,
#   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>
```

Focusing on San Diego

```
sd = vax[vax$county == "San Diego",]
head(sd)
```

```
   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
4  2021-01-05                   91911                San Diego San Diego
20 2021-01-05                   91941                San Diego San Diego
23 2021-01-05                   91963                San Diego San Diego
27 2021-01-05                   92054                San Diego San Diego
29 2021-01-05                   91980                San Diego San Diego
33 2021-01-05                   92056                San Diego San Diego
   vaccine_equity_metric_quartile                vem_source
4                               2 Healthy Places Index Score
20                              3 Healthy Places Index Score
23                              2    CDPH-Derived ZCTA Score
27                              2 Healthy Places Index Score
29                             NA            No VEM Assigned
33                              3 Healthy Places Index Score
   age12_plus_population age5_plus_population tot_population
4                71642.8                79225          84026
20               27354.6                29757          31918
23                1010.3                 1089           1182
27               35176.1                39270          41807
29                   0.0                    0             NA
33               45552.2                49110          52337
```

	persons_fully_vaccinated	persons_partially_vaccinated
4	29	1429
20	22	803
23	NA	NA
27	14	421
29	NA	NA
33	15	811

	percent_of_population_fully_vaccinated
4	0.000345
20	0.000689
23	NA
27	0.000335
29	NA
33	0.000287

	percent_of_population_partially_vaccinated
4	0.017007
20	0.025158
23	NA
27	0.010070
29	NA
33	0.015496

	percent_of_population_with_1_plus_dose	booster_recip_count
4	0.017352	NA
20	0.025847	NA
23	NA	NA
27	0.010405	NA
29	NA	NA
33	0.015783	NA

	bivalent_dose_recip_count	eligible_recipient_count
4	NA	29
20	NA	22
23	NA	0
27	NA	14
29	NA	0
33	NA	15

	eligible_bivalent_recipient_count
4	29
20	22
23	0
27	14
29	0
33	15

redacted

```
4  Information redacted in accordance with CA state privacy requirements
20 Information redacted in accordance with CA state privacy requirements
23 Information redacted in accordance with CA state privacy requirements
27 Information redacted in accordance with CA state privacy requirements
29 Information redacted in accordance with CA state privacy requirements
33 Information redacted in accordance with CA state privacy requirements
```

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

```
n_distinct(sd$zip_code_tabulation_area)
```

```
[1] 107
```

There are 107 distinct zip code areas

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

```
sd.80 <- filter(vax, county == "San Diego" &
                tot_population > 80000)
head(sd.80)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction    county
1 2021-01-05                   91911                San Diego San Diego
2 2021-01-05                   92154                San Diego San Diego
3 2021-01-05                   92126                San Diego San Diego
4 2021-01-12                   92154                San Diego San Diego
5 2021-01-12                   91911                San Diego San Diego
6 2021-01-12                   92126                San Diego San Diego
  vaccine_equity_metric_quartile                vem_source
1                              2 Healthy Places Index Score
2                              2 Healthy Places Index Score
3                              4 Healthy Places Index Score
4                              2 Healthy Places Index Score
5                              2 Healthy Places Index Score
6                              4 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
1               71642.8                79225          84026
2               76365.2                82971          88979
3               71820.2                77775          82658
4               76365.2                82971          88979
5               71642.8                79225          84026
```

```
6               71820.2                77775          82658
  persons_fully_vaccinated persons_partially_vaccinated
1                       29                        1429
2                       18                        1404
3                       34                        1866
4                      275                        1867
5                      311                        1893
6                      423                        2178
  percent_of_population_fully_vaccinated
1                               0.000345
2                               0.000202
3                               0.000411
4                               0.003091
5                               0.003701
6                               0.005117
  percent_of_population_partially_vaccinated
1                                   0.017007
2                                   0.015779
3                                   0.022575
4                                   0.020982
5                                   0.022529
6                                   0.026350
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.017352                  NA
2                               0.015981                  NA
3                               0.022986                  NA
4                               0.024073                  NA
5                               0.026230                  NA
6                               0.031467                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                       29
2                        NA                       18
3                        NA                       34
4                        NA                      275
5                        NA                      311
6                        NA                      423
  eligible_bivalent_recipient_count
1                                29
2                                18
3                                34
4                               275
5                               311
6                               423
```

```
                                                                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
```

Zip code 92154 has the highest total population

Q13. What is the overall average (with 2 decimal numbers) "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2023-05-23"?

```
sd2 <- filter(vax, county == "San Diego" & as_of_date == "2023-05-23")
mean(na.omit(sd2$percent_of_population_fully_vaccinated))*100
```

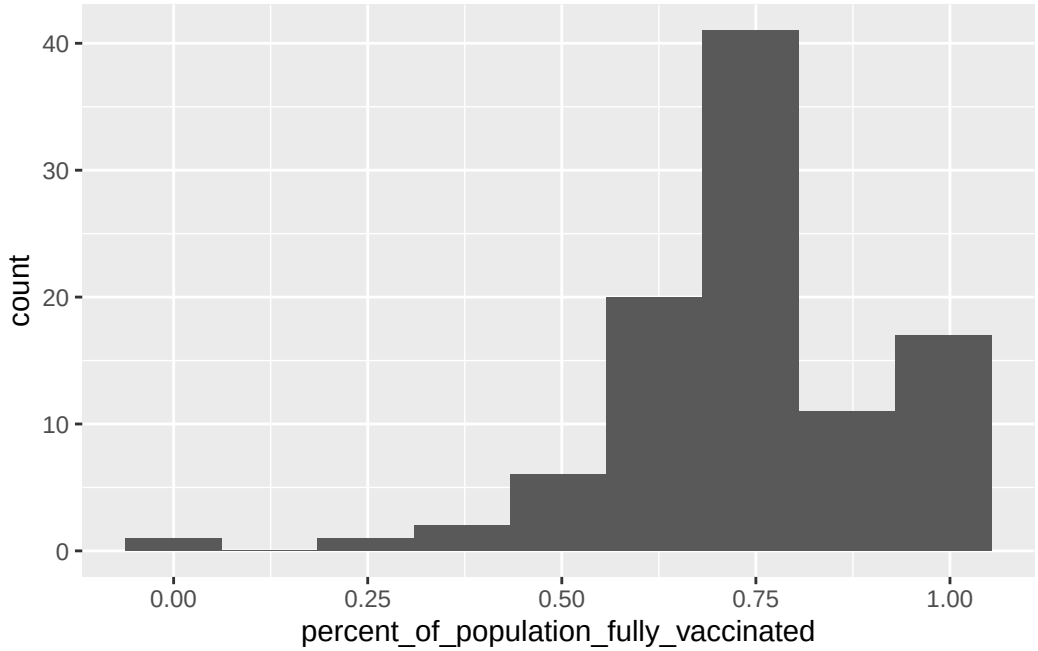
```
[1] 74.20143
```

74.20% is the average of percent of population fully vaccinated

- **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 "2023-05-23"?

```
library(ggplot2)
ggplot(sd2)  + geom_histogram(aes(percent_of_population_fully_vaccinated), bins = 9)
```

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



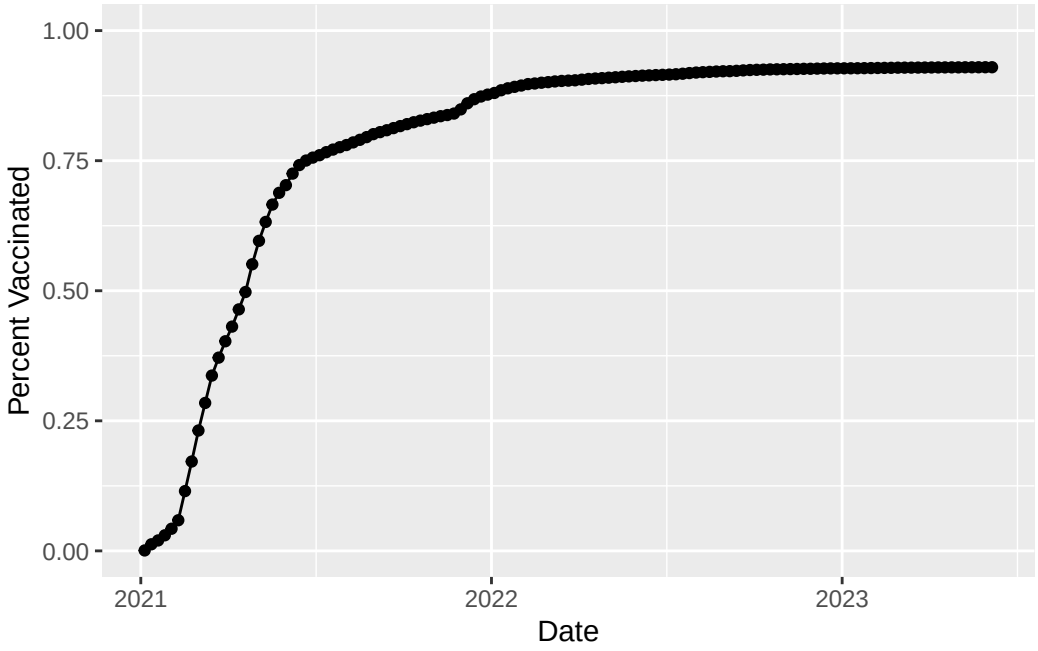
UCSD/ La Jolla

```
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(as_of_date,
      percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
```



Looking at data of larger sized populations

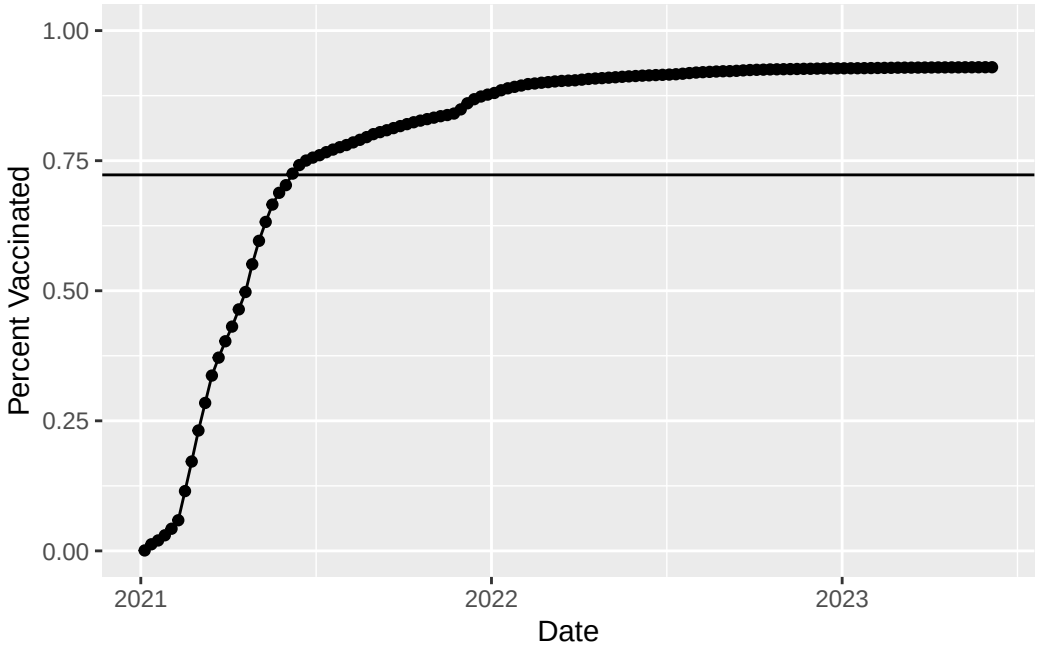
```
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2023-05-23")
```

- **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* "2023-05-23". Add this as a straight horizontal line to your plot from above with the `geom_hline()` function?

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

```
[1] 0.7227246
```

```
ggplot(ucsd) +
  aes(as_of_date,
      percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated") +
  geom_hline(yintercept = mean(vax.36$percent_of_population_fully_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* "2023-05-23"?

```
summary(vax.36)
```

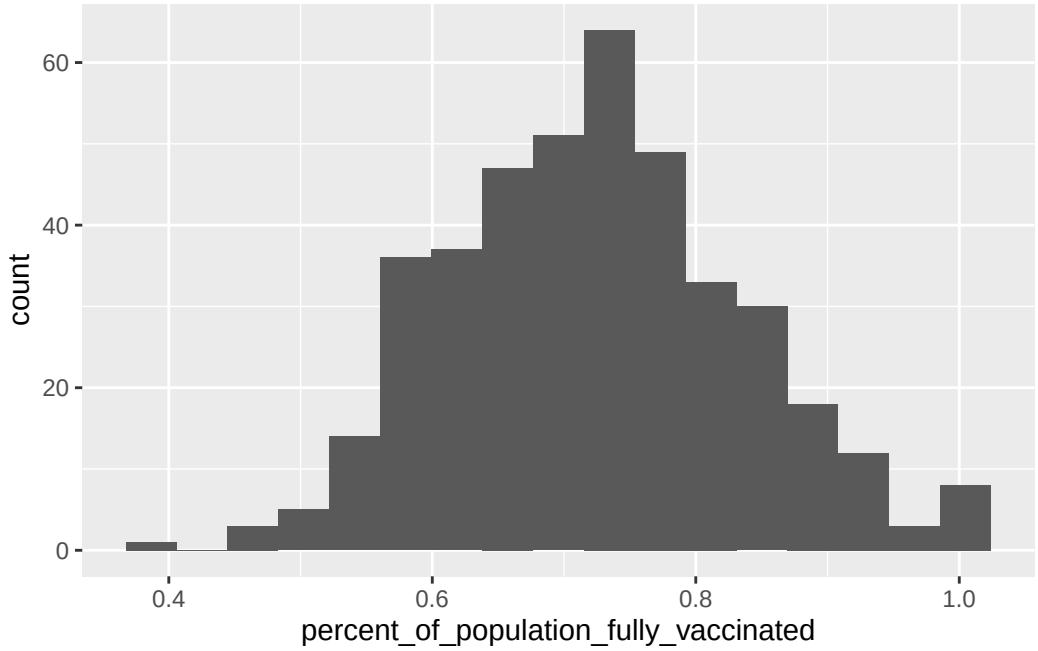
```
  as_of_date         zip_code_tabulation_area local_health_jurisdiction
Min.   :2023-05-23   Min.   :90001            Length:411
1st Qu.:2023-05-23   1st Qu.:91762            Class :character
Median :2023-05-23   Median :92646            Mode  :character
Mean   :2023-05-23   Mean   :92862
3rd Qu.:2023-05-23   3rd Qu.:94517
Max.   :2023-05-23   Max.   :96003
   county          vaccine_equity_metric_quartile  vem_source
Length:411         Min.   :1.000                   Length:411
Class :character   1st Qu.:1.000                   Class :character
Mode  :character   Median :2.000                   Mode  :character
                   Mean   :2.353
                   3rd Qu.:3.000
                   Max.   :4.000
age12_plus_population age5_plus_population tot_population
Min.   :31651         Min.   : 36181       Min.   : 38007
```

```
1st Qu.:37694         1st Qu.: 41612       1st Qu.: 44393
Median :43985         Median : 48573       Median : 52212
Mean   :46847         Mean   : 52012       Mean   : 55641
3rd Qu.:53932         3rd Qu.: 59168       3rd Qu.: 62910
Max.   :88557         Max.   :101902       Max.   :111165
persons_fully_vaccinated persons_partially_vaccinated
Min.   :17566            Min.   : 1813
1st Qu.:32280            1st Qu.: 2956
Median :37329            Median : 3654
Mean   :40136            Mean   : 4270
3rd Qu.:45422            3rd Qu.: 4848
Max.   :87720            Max.   :40934
percent_of_population_fully_vaccinated
Min.   :0.3816
1st Qu.:0.6471
Median :0.7207
Mean   :0.7227
3rd Qu.:0.7924
Max.   :1.0000
percent_of_population_partially_vaccinated
Min.   :0.04471
1st Qu.:0.05937
Median :0.06761
Mean   :0.07665
3rd Qu.:0.07974
Max.   :1.00000
percent_of_population_with_1_plus_dose booster_recip_count
Min.   :0.4474                         Min.   : 9135
1st Qu.:0.7148                         1st Qu.:18074
Median :0.7910                         Median :22777
Mean   :0.7928                         Mean   :24091
3rd Qu.:0.8707                         3rd Qu.:28800
Max.   :1.0000                         Max.   :59845
bivalent_dose_recip_count eligible_recipient_count
Min.   : 2948             Min.   :17499
1st Qu.: 6234             1st Qu.:32184
Median : 8697             Median :37247
Mean   : 9645             Mean   :40025
3rd Qu.:12212             3rd Qu.:45310
Max.   :29485             Max.   :87451
eligible_bivalent_recipient_count   redacted
Min.   :    0                     Length:411
1st Qu.:32094                     Class :character
```

```
Median :37178                     Mode   :character
Mean   :39926
3rd Qu.:45310
Max.   :87451
```

- **Q18.** Using ggplot generate a histogram of this data.

```
ggplot(vax.36)+ geom_histogram(aes(percent_of_population_fully_vaccinated), bins = 17)
```



- **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 == "2023-05-23") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                                0.55265
```

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

```r
vax.36.all <- filter(vax, age5_plus_population > 36144)


ggplot(vax.36.all) +
  aes(as_of_date,
      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 above 36k population") +
  geom_hline(yintercept = mean(vax.36.all$percent_of_population_fully_vaccinated), linetyp
```

```
Warning: Removed 185 rows containing missing values (`geom_line()`).
```

```
Warning: Removed 1 rows containing missing values (`geom_hline()`).
```