Class 17

Izabelle Querubin

COVID-19 Vaccination Rates

Getting Started

```
# Import vaccination data
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction      county
1 2021-01-05                   96065                     Shasta      Shasta
2 2021-01-05                   90044                Los Angeles Los Angeles
3 2021-01-05                   96035                     Tehama      Tehama
4 2021-01-05                   93230                      Kings       Kings
5 2021-01-05                   96068                     Lassen      Lassen
6 2021-01-05                   96038                   Siskiyou    Siskiyou
  vaccine_equity_metric_quartile                vem_source
1                              2    CDPH-Derived ZCTA Score
2                              1 Healthy Places Index Score
3                              2 Healthy Places Index Score
4                              2 Healthy Places Index Score
5                              1    CDPH-Derived ZCTA Score
6                              2    CDPH-Derived ZCTA Score
  age12_plus_population age5_plus_population tot_population
1                 358.9                  385            403
2               79804.5                91088          99443
3                3118.1                 3357           3629
4               54911.7                62296          67605
5                 170.3                  174            204
6                 580.5                  633            732
  persons_fully_vaccinated persons_partially_vaccinated
```

```
1                        NA                             NA
2                        17                            526
3                        NA                             NA
4                        NA                             NA
5                        NA                             NA
6                        NA                             NA
  percent_of_population_fully_vaccinated
1                                     NA
2                                0.000171
3                                     NA
4                                     NA
5                                     NA
6                                     NA
  percent_of_population_partially_vaccinated
1                                         NA
2                                    0.005289
3                                         NA
4                                         NA
5                                         NA
6                                         NA
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                                0.00546                  NA
3                                     NA                  NA
4                                     NA                  NA
5                                     NA                  NA
6                                     NA                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                        0
2                        NA                       17
3                        NA                        0
4                        NA                        2
5                        NA                        0
6                        NA                        0
  eligible_bivalent_recipient_count
1                                 0
2                                17
3                                 0
4                                 2
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?

"persons_fully_vaccinated"

Q2. What column details the Zip code tabulation area?

"zip_code_tabulation_area"

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

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

Q3. What is the earliest date in this dataset?

2021-01-05

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

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

Q4. What is the latest date in this dataset?

2023-06-13

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

Table 1: Data summary

Name	vax
Number of rows	225792
Number of columns	19

Table 1: Data summary

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	128	0
local_health_jurisdiction	0	1	0	15	640	62	0
county	0	1	0	15	640	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	11136	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	11008	0.95	23372.77	22628.50	12	2126.00	18714.00	38168.00	111165.0
persons_fully_vaccinated	18076	0.92	14346.77	15312.25	11	960.00	9099.50	23900.00	87743.0
persons_partially_vaccinated	18076	0.92	1713.81	2082.46	11	164.00	1206.00	2555.00	44974.0
percent_of_population_fully_vaccinated	23037	0.91	0.58	0.25	0	0.44	0.62	0.76	1.0
percent_of_population_partially_vaccinated	23037	0.91	0.08	0.09	0	0.05	0.06	0.08	1.0
percent_of_population_with_1_plus_dose	24214	0.89	0.65	0.24	0	0.51	0.68	0.82	1.0
booster_recip_count	74804	0.67	6498.71	7875.86	11	337.00	3196.00	10469.00	60109.0
bivalent_dose_recip_count	160370	0.29	3496.05	4079.95	11	230.00	1924.00	5621.00	29816.0
eligible_recipient_count	0	1.00	13191.71	15175.51	0	539.00	6755.00	22629.00	87473.0
eligible_bivalent_recipient_count	0	1.00	13079.35	15253.05	0	255.00	6626.00	22621.25	87473.0

Q5. How many numeric columns are in this dataset?

14

```
num_na <- sum(is.na(vax$persons_fully_vaccinated))
```

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

18076

ncol(vax$persons_fully_vaccinated, na.rm = TRUE)

```
percent_missing <- round((num_na / nrow(vax)) * 100, 2)
percent_missing
```

```
[1] 8.01
```

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

8.01%

```
library(lubridate)
```

```
Attaching package: 'lubridate'

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

    date, intersect, setdiff, union
```

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

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

```
Time difference of 894 days
```

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

```
Time difference of 889 days
```

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

3 days have passed

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

```
num_unique_dates <- length(unique(vax$as_of_date))
num_unique_dates
```

```
[1] 128
```

128 unique dates

Working with Zip Codes

```
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.
```

```
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_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>
```

```
# Pull data for all ZIP codes in the dataset
zipdata <- reverse_zipcode( vax$zip_code_tabulation_area )
```

Focus 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] 13696
```

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

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction    county
1 2021-01-05                   92173                San Diego San Diego
2 2021-01-05                   92139                San Diego San Diego
3 2021-01-05                   92078                San Diego San Diego
4 2021-01-05                   92117                San Diego San Diego
5 2021-01-05                   92123                San Diego San Diego
6 2021-01-05                   92118                San Diego San Diego
  vaccine_equity_metric_quartile                vem_source
1                              1 Healthy Places Index Score
2                              2 Healthy Places Index Score
3                              3 Healthy Places Index Score
4                              3 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               25332.5                28487          31000
2               30679.9                33923          36105
3               41789.5                47476          50510
4               50041.6                53839          56983
5               28353.3                30426          32473
6               19835.0                21470          22548
  persons_fully_vaccinated persons_partially_vaccinated
1                       NA                           NA
2                       15                          838
3                       29                          728
4                       32                         1157
5                      360                         3139
6                       13                          496
  percent_of_population_fully_vaccinated
1                                     NA
2                                0.000415
3                                0.000574
4                                0.000562
5                                0.011086
6                                0.000577
  percent_of_population_partially_vaccinated
```

```
1                                       NA
2                                 0.023210
3                                 0.014413
4                                 0.020304
5                                 0.096665
6                                 0.021998
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                                0.023625                 NA
3                                0.014987                 NA
4                                0.020866                 NA
5                                0.107751                 NA
6                                0.022575                 NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                       6
2                        NA                      15
3                        NA                      29
4                        NA                      32
5                        NA                     360
6                        NA                      13
  eligible_bivalent_recipient_count
1                                 6
2                                15
3                                29
4                                32
5                               360
6                                13
                                                                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
distinct_codes <- unique(sd$zip)
distinct_zip_codes <- length(distinct_codes)
distinct_zip_codes
```

```
[1] 107
```

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

107

```
largest_population <- sd$zip[which.max(sd$tot_population)]
largest_population
```

```
[1] 92154
```

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

92154

```
library(dplyr)

sd_filtered <- filter(vax, county == "San Diego" & as_of_date == "2023-05-23")
```

```
sd_filtered <- mutate(sd_filtered, percent_of_population_fully_vaccinated = as.numeric(per
```

```
average_percent_vaccinated <- mean(sd_filtered$percent_of_population_fully_vaccinated, na.
average_percent_vaccinated <- round(average_percent_vaccinated, 4)
average_percent_vaccinated
```

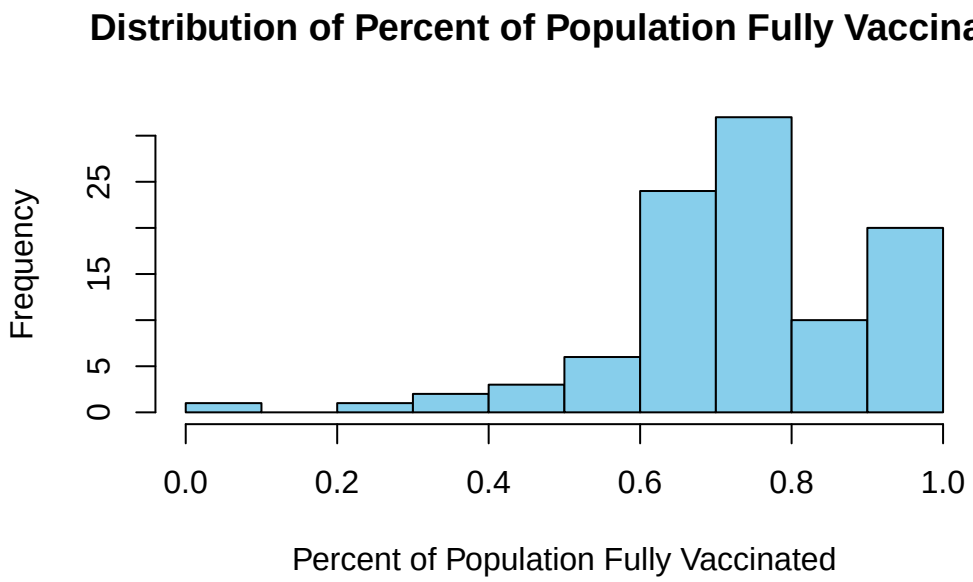
```
[1] 0.7421
```

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"?

74.21%

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"?

```
hist(sd_filtered$percent_of_population_fully_vaccinated, breaks = 10, col = "skyblue", bor
     xlab = "Percent of Population Fully Vaccinated", ylab = "Frequency",
     main = "Distribution of Percent of Population Fully Vaccinated")
```



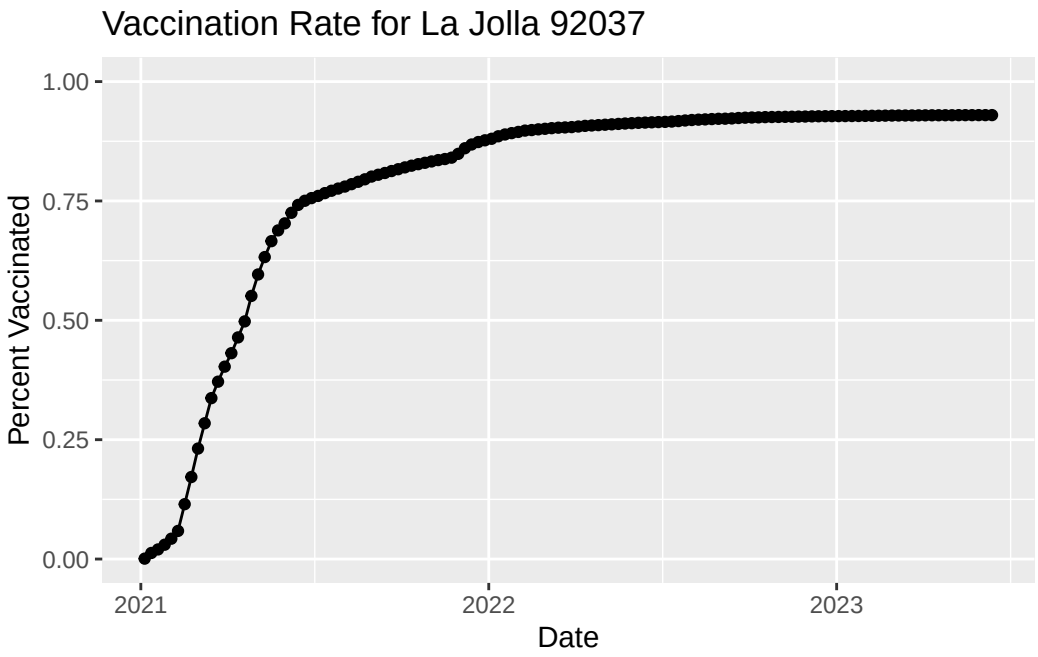

```
ucsd <- filter(sd, zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

```
[1] 36144
```

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

```
library(ggplot2)

ggplot(ucsd) +
  aes(x = as.Date(as_of_date), y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group = 1) +
  ylim(c(0, 1)) +
  labs(x = "Date", y = "Percent Vaccinated", title = "Vaccination Rate for La Jolla 92037"
```




```
# Subset to all CA areas with a population as large as 92037
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2023-05-23")

head(vax.36)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction     county
1 2023-05-23                   92113                 San Diego  San Diego
2 2023-05-23                   95355               Stanislaus Stanislaus
3 2023-05-23                   92084                 San Diego  San Diego
4 2023-05-23                   92104                 San Diego  San Diego
5 2023-05-23                   92083                 San Diego  San Diego
6 2023-05-23                   95382               Stanislaus Stanislaus
  vaccine_equity_metric_quartile                vem_source
1                              1 Healthy Places Index Score
2                              2 Healthy Places Index Score
3                              2 Healthy Places Index Score
4                              3 Healthy Places Index Score
5                              2 Healthy Places Index Score
6                              2 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
1               47799.7                53883          58408
```

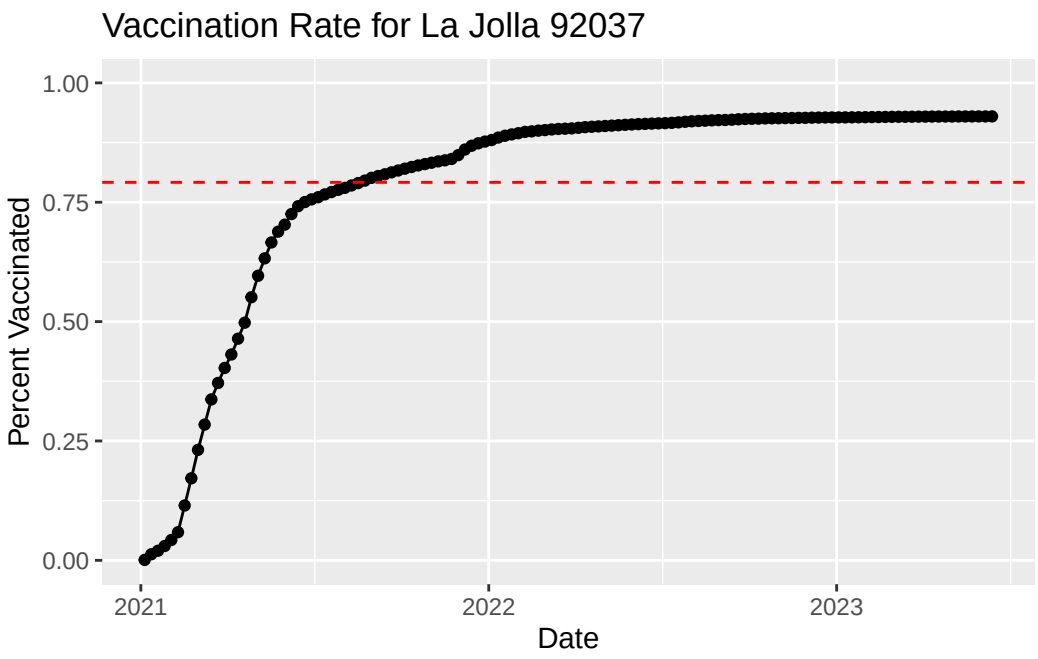
```
2               50941.6                56248          59621
3               42677.7                47784          51619
4               40343.9                42839          45435
5               32246.5                36283          39509
6               32843.7                36425          38700
  persons_fully_vaccinated persons_partially_vaccinated
1                    39245                         5049
2                    39604                         3206
3                    32976                         3047
4                    34881                         4005
5                    26259                         2572
6                    24250                         2104
  percent_of_population_fully_vaccinated
1                               0.671911
2                               0.664263
3                               0.638835
4                               0.767712
5                               0.664633
6                               0.626615
  percent_of_population_partially_vaccinated
1                                   0.086444
2                                   0.053773
3                                   0.059029
4                                   0.088148
5                                   0.065099
6                                   0.054367
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.758355               19775
2                               0.718036               22847
3                               0.697864               18102
4                               0.855860               23130
5                               0.729732               13542
6                               0.680982               13492
  bivalent_dose_recip_count eligible_recipient_count
1                      5992                    39173
2                      8292                    39578
3                      6822                    32896
4                     10702                    34777
5                      4496                    26192
6                      4257                    24240
  eligible_bivalent_recipient_count redacted
1                             39173       No
2                             39578       No
```

3	32896	No
4	34777	No
5	26192	No
6	24240	No

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_percent_vaccinated <- mean(ucsd$percent_of_population_fully_vaccinated, na.rm = TRUE)

ggplot(ucsd) +
  aes(x = as.Date(as_of_date), y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group = 1) +
  ylim(c(0, 1)) +
  labs(x = "Date", y = "Percent Vaccinated", title = "Vaccination Rate for La Jolla 92037") +
  geom_hline(yintercept = mean_percent_vaccinated, linetype = "dashed", color = "red")
```



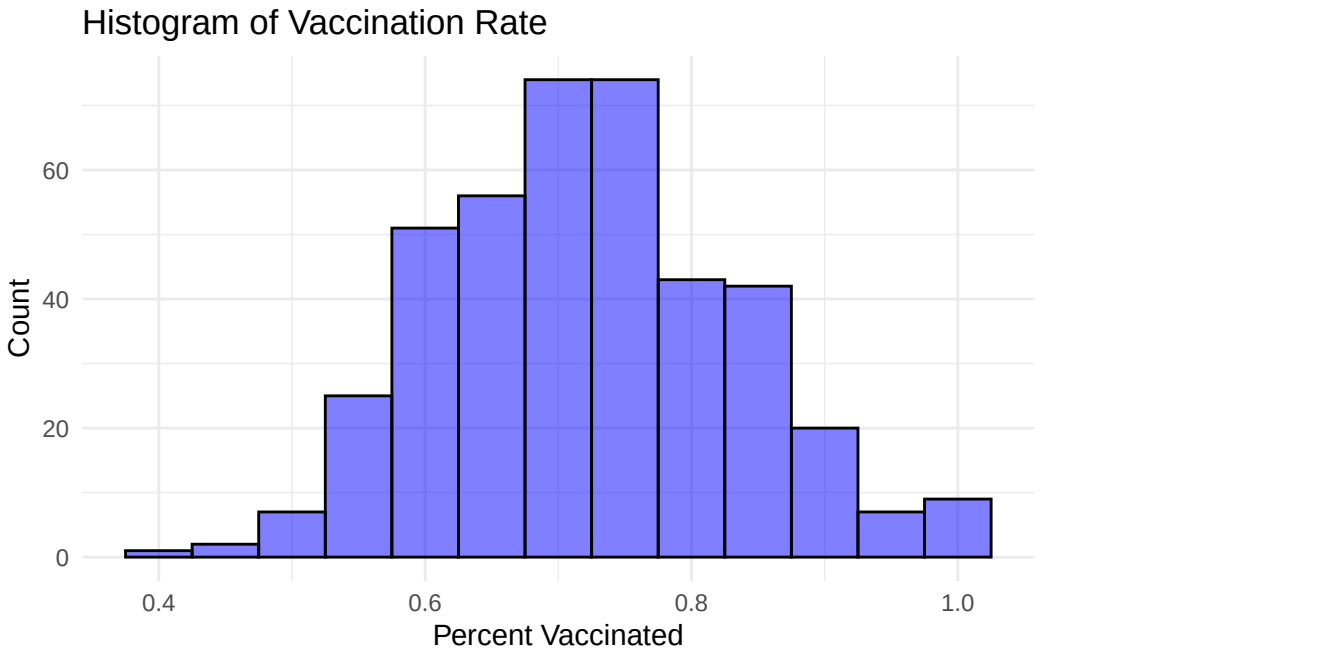
Q17. What is the 6 number summary (Min, 1st Qu., Median, Mean, 3rd Qu., and Max) of the *"Percent of Population Fully Vaccinated"* values for ZIP code areas with a population as large as 92037 (La Jolla) *as_of_date* "2023-05-23"?

```
summary_stats <- summary(ucsd$percent_of_population_fully_vaccinated)
summary_stats
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.000865 0.788900 0.904358 0.791645 0.926064 0.929863
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36, aes(x = percent_of_population_fully_vaccinated)) +
  geom_histogram(binwidth = 0.05, color = "black", fill = "blue", alpha = 0.5) +
  labs(x = "Percent Vaccinated", y = "Count", title = "Histogram of Vaccination Rate") +
  theme_minimal()
```



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                               0.552887
```

```
vax %>% filter(as_of_date == "2023-05-23") %>%
  filter(zip_code_tabulation_area=="92109") %>%
  select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                               0.694063
```

The 92109 and 92040 zip code areas are below the average value.

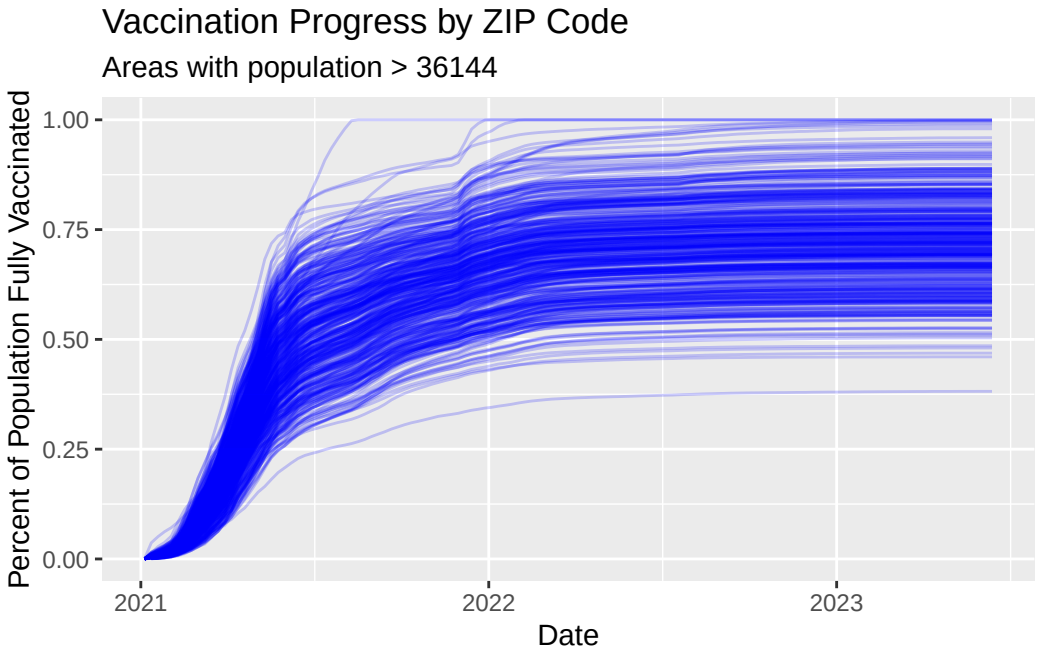
Q20. Finally make a time course plot of vaccination progress for all areas in the full dataset with a `age5_plus_population > 36144`.

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

ggplot(vax.36.all) +
  aes(x = as.Date(as_of_date), y = percent_of_population_fully_vaccinated, group = zip_cod
  geom_line(alpha = 0.2, color = "blue") +
  ylim(0, 1) +
  labs(x = "Date", y = "Percent of Population Fully Vaccinated",
       title = "Vaccination Progress by ZIP Code",
       subtitle = "Areas with population > 36144") +
  geom_hline(yintercept = mean(vax.36.all$percent_of_population_fully_vaccinated), linetyp
```

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

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


Vaccination Progress by ZIP Code
Areas with population > 36144
Percent of Population Fully Vaccinated
1.00
0.75
0.50
0.25
0.00
2021
2022
2023
Date