Mini-Project: COVID-19 Vaccination Rates

Vivian Cai

3/5/2022

Overview of the data

```
# 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                    92549                 Riverside      Riverside
## 2 2021-01-05                    92130                 San Diego      San Diego
## 3 2021-01-05                    92397            San Bernardino San Bernardino
## 4 2021-01-05                    94563              Contra Costa   Contra Costa
## 5 2021-01-05                    94519              Contra Costa   Contra Costa
## 6 2021-01-05                    91042               Los Angeles    Los Angeles
##   vaccine_equity_metric_quartile                 vem_source
## 1                              3 Healthy Places Index Score
## 2                              4 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                              3 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1                2348.4                 2461                       NA
## 2               46300.3                53102                       61
## 3                3695.6                 4225                       NA
## 4               17216.1                18896                       NA
## 5               16861.2                18678                       NA
## 6               23962.2                25741                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                           27                               0.001149
## 3                           NA                                     NA
## 4                           NA                                     NA
## 5                           NA                                     NA
## 6                           NA                                     NA
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                   0.000508
## 3                                         NA
## 4                                         NA
## 5                                         NA
## 6                                         NA
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                                     NA                  NA
## 2                               0.001657                  NA
## 3                                     NA                  NA
## 4                                     NA                  NA
## 5                                     NA                  NA
## 6                                     NA                  NA
##                                                                 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 **persons_fully_vaccinated** column.

Q2. What column details the Zip code tabulation area?

The **zip_code_tabulation_area** column.

Q3. What is the earliest date in this dataset?

2021-01-05

Q4. What is the latest date in this dataset?

2022-03-01 I got red mark on the course website for this question but I think the data might have updated to include more recent dates.

```
tail(vax)
```

```
##        as_of_date zip_code_tabulation_area local_health_jurisdiction
## 107599 2022-03-01                    91945                 San Diego
## 107600 2022-03-01                    91741               Los Angeles
## 107601 2022-03-01                    91768               Los Angeles
## 107602 2022-03-01                    91345               Los Angeles
## 107603 2022-03-01                    91356               Los Angeles
## 107604 2022-03-01                    94402                 San Mateo
##             county vaccine_equity_metric_quartile                 vem_source
## 107599   San Diego                              2 Healthy Places Index Score
## 107600 Los Angeles                              3 Healthy Places Index Score
## 107601 Los Angeles                              1 Healthy Places Index Score
## 107602 Los Angeles                              2 Healthy Places Index Score
## 107603 Los Angeles                              3 Healthy Places Index Score
## 107604   San Mateo                              4 Healthy Places Index Score
##        age12_plus_population age5_plus_population persons_fully_vaccinated
## 107599               22820.5                25486                    18164
## 107600               22895.7                25243                    19051
## 107601               29837.1                32658                    20587
## 107602               16767.4                18029                    14872
## 107603               26392.1                28379                    22863
## 107604               21862.1                24150                    23094
##        persons_partially_vaccinated percent_of_population_fully_vaccinated
## 107599                         4032                               0.712705
## 107600                         1438                               0.754704
## 107601                         2467                               0.630382
## 107602                         1371                               0.824893
## 107603                         2114                               0.805631
## 107604                         1697                               0.956273
##        percent_of_population_partially_vaccinated
## 107599                                   0.158205
## 107600                                   0.056966
## 107601                                   0.075540
## 107602                                   0.076044
## 107603                                   0.074492
## 107604                                   0.070269
##        percent_of_population_with_1_plus_dose booster_recip_count redacted
## 107599                               0.870910                6542       No
## 107600                               0.811670               10331       No
## 107601                               0.705922                8694       No
## 107602                               0.900937                6715       No
## 107603                               0.880123               12372       No
## 107604                               1.000000               16049       No
```

```r
# install.packages("skimr")
library(skimr)
skimr::skim(vax)
```

Table 1: Data summary

Name	vax
Number of rows	107604
Number of columns	15
Column type frequency:	
character	5
numeric	10
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	61	0
local_health_jurisdiction	0	1	0	15	305	62	0
county	0	1	0	15	305	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	5307	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.04	18993.91	0	1346.95	13685.10	31756.12	88556.7	
age5_plus_population	0	1.00	20875.24	21106.02	0	1460.50	15364.00	34877.00	101902.0	
persons_fully_vaccinated	18338	0.83	12155.61	13063.88	11	1066.25	7374.50	20005.00	77744.0	
persons_partially_vaccinated	18338	0.83	831.74	1348.68	11	76.00	372.00	1076.00	34219.0	
percent_of_population_fully_vaccinated	18338	0.83	0.51	0.26	0	0.33	0.54	0.70	1.0	
percent_of_population_partially_vaccinated	18338	0.83	0.05	0.09	0	0.01	0.03	0.05	1.0	
percent_of_population_with_1_plus_dose	18338	0.83	0.54	0.28	0	0.36	0.58	0.75	1.0	
booster_recip_count	64317	0.40	4100.55	5900.21	11	176.00	1136.00	6154.50	50602.0	

Q5. How many numeric columns are in this dataset?

9

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

18174

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

17.17%

```r
n = 18174/105840
n
```

```
## [1] 0.171712
```

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

Some zip codes might not have residents or they failed to provide the data.

Working with dates

```r
# install.packages("lubridate")
library(lubridate)
```

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

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
today()
```

```
## [1] "2022-03-05"
```

```
# converting data into a lubridate format
# Specify that we are using the year-month-day format
vax$as_of_date <- ymd(vax$as_of_date)
```

```
# How many days have passed since the first vaccination reported in this dataset?
today() - vax$as_of_date[1]
```

```
## Time difference of 424 days
```

```
# how many days the dataset span?
vax$as_of_date[nrow(vax)] - vax$as_of_date[1]
```

```
## Time difference of 420 days
```

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

6

Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)? 61 as shown by the code chunk below

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

```
## [1] 61
```

Working with ZIP codes

```
# install.packages("zipcodeR")
library(zipcodeR)
```

```
# find the centroid of the La Jolla 92037
geocode_zip('92037')
```

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

```
# Calculate the distance between the centroids of any two ZIP codes in miles
zip_distance('92037','92109')
```

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

```
# pull census data about ZIP code areas
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
## # ... with 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

```
# using base R
sd <- vax[ which(vax$county == "San Diego") , ]
nrow(sd)
```

```
## [1] 6527
```

```
# Using dplyr
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] 6527
```

```
# Using dplyr is often more convenient when we are subsetting across multiple criteria
# for example all San Diego county areas with a population of over 10,000.
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                    92130                 San Diego San Diego
## 2 2021-01-05                    91945                 San Diego San Diego
## 3 2021-01-05                    92103                 San Diego San Diego
## 4 2021-01-05                    92075                 San Diego San Diego
## 5 2021-01-05                    92084                 San Diego San Diego
## 6 2021-01-05                    92116                 San Diego San Diego
##   vaccine_equity_metric_quartile                 vem_source
## 1                              4 Healthy Places Index Score
## 2                              2 Healthy Places Index Score
## 3                              4 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                              2 Healthy Places Index Score
## 6                              3 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               46300.3                53102                       61
## 2               22820.5                25486                       NA
## 3               32146.4                33213                       45
## 4               11136.3                12177                       NA
## 5               42677.7                47784                       12
## 6               30255.7                31673                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           27                               0.001149
## 2                           NA                                     NA
## 3                           30                               0.001355
## 4                           NA                                     NA
## 5                           17                               0.000251
## 6                           NA                                     NA
##   percent_of_population_partially_vaccinated
## 1                                   0.000508
## 2                                         NA
## 3                                   0.000903
## 4                                         NA
## 5                                   0.000356
## 6                                         NA
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                               0.001657                  NA
## 2                                    NA                  NA
## 3                              0.002258                  NA
## 4                                    NA                  NA
## 5                              0.000607                  NA
## 6                                    NA                  NA
##                                                                 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
```

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

```
ardsd <- sd %>%
  arrange(desc(age12_plus_population))
head(ardsd)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-01-05                    92154                 San Diego San Diego
## 2 2021-01-12                    92154                 San Diego San Diego
## 3 2021-01-19                    92154                 San Diego San Diego
## 4 2021-01-26                    92154                 San Diego San Diego
## 5 2021-02-02                    92154                 San Diego San Diego
## 6 2021-02-09                    92154                 San Diego San Diego
##   vaccine_equity_metric_quartile                 vem_source
## 1                              2 Healthy Places Index Score
## 2                              2 Healthy Places Index Score
## 3                              2 Healthy Places Index Score
## 4                              2 Healthy Places Index Score
## 5                              2 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               76365.2                82971                       18
## 2               76365.2                82971                      282
## 3               76365.2                82971                      671
## 4               76365.2                82971                      986
## 5               76365.2                82971                     1381
## 6               76365.2                82971                     2136
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           22                               0.000217
## 2                           37                               0.003399
## 3                           93                                0.008087
## 4                          216                                0.011884
## 5                          432                                0.016644
## 6                          761                                0.025744
##   percent_of_population_partially_vaccinated
## 1                                   0.000265
## 2                                   0.000446
## 3                                   0.001121
## 4                                   0.002603
## 5                                   0.005207
## 6                                   0.009172
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                               0.000482                  NA
## 2                               0.003845                  NA
## 3                               0.009208                  NA
## 4                               0.014487                  NA
## 5                               0.021851                  NA
## 6                               0.034916                  NA
##                                                                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
```

```
# Using dplyr select all San Diego "county" entries on "as_of_date" "2022-02-22"
sd0222 <- sd %>%
  filter(as_of_date == "2022-02-22")

head(sd0222)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2022-02-22                    92064                 San Diego San Diego
## 2 2022-02-22                    92103                 San Diego San Diego
## 3 2022-02-22                    92118                 San Diego San Diego
## 4 2022-02-22                    92083                 San Diego San Diego
## 5 2022-02-22                    92056                 San Diego San Diego
## 6 2022-02-22                    92069                 San Diego San Diego
##   vaccine_equity_metric_quartile                 vem_source
## 1                              4 Healthy Places Index Score
## 2                              4 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              2 Healthy Places Index Score
## 5                              3 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               42177.1                46855                    34266
## 2               32146.4                33213                    46456
## 3               19835.0                21470                    14954
## 4               32246.5                36283                    24146
## 5               45552.2                49110                    34782
## 6               41447.3                46850                    32505
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         6861                               0.731320
## 2                         8434                               1.000000
## 3                         7405                               0.696507
## 4                         5924                               0.665491
## 5                         7362                               0.708247
## 6                         7043                               0.693810
##   percent_of_population_partially_vaccinated
## 1                                   0.146430
## 2                                   0.253937
## 3                                   0.344900
## 4                                   0.163272
## 5                                   0.149908
## 6                                   0.150331
##   percent_of_population_with_1_plus_dose booster_recip_count redacted
## 1                               0.877750               15499       No
## 2                               1.000000               14627       No
## 3                               1.000000                5721       No
## 4                               0.828763                7322       No
## 5                               0.858155               15441       No
## 6                               0.844141               12168       No
```

Q13. What is the overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2022-02-22"? 0.7041551

```
mean(sd0222$percent_of_population_fully_vaccinated, na.rm = TRUE)
```

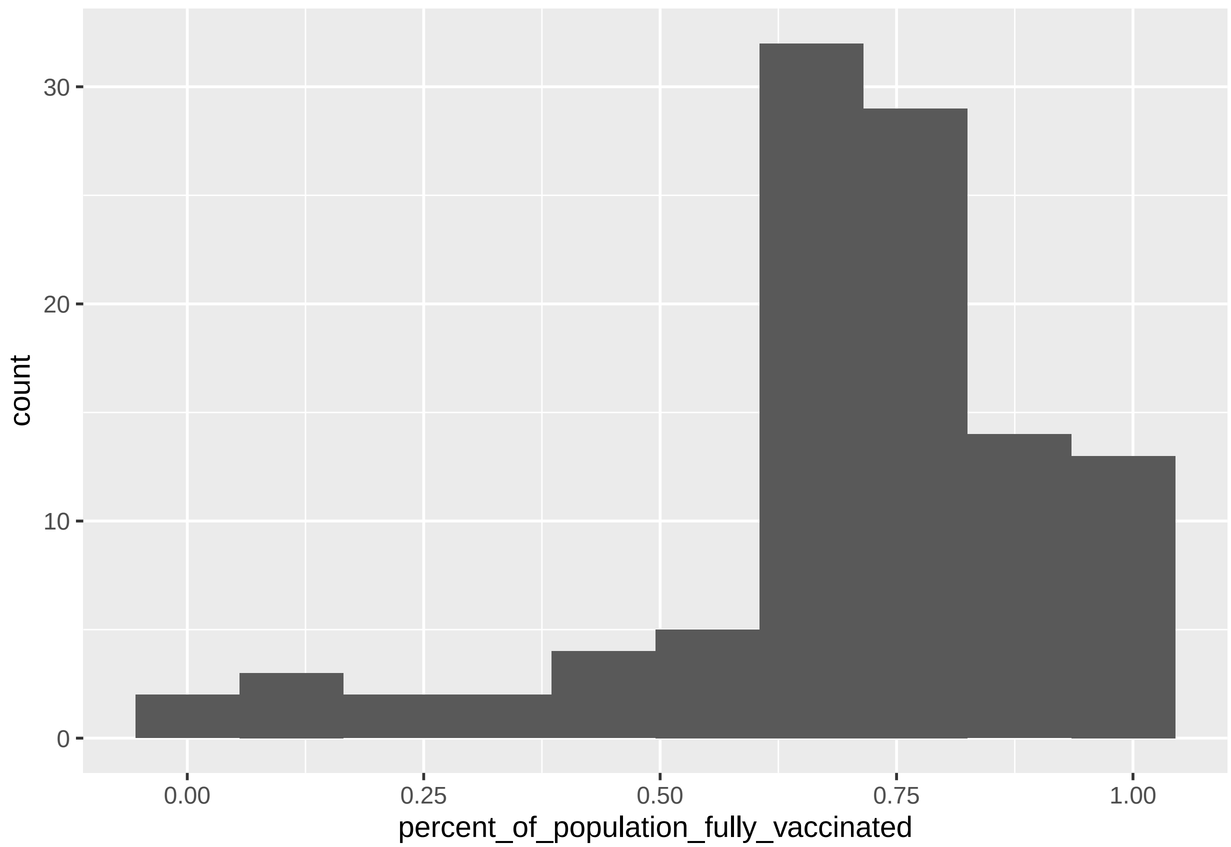
```
## [1] 0.7041551
```

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 "2022-02-22"?

I couldn't get the same graph as shown on the instruction page. Perhaps the data was updated by the time I downloaded it.

```
library(ggplot2)
ggplot(sd0222, aes(x = percent_of_population_fully_vaccinated)) +
  geom_histogram(bins = 10)
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



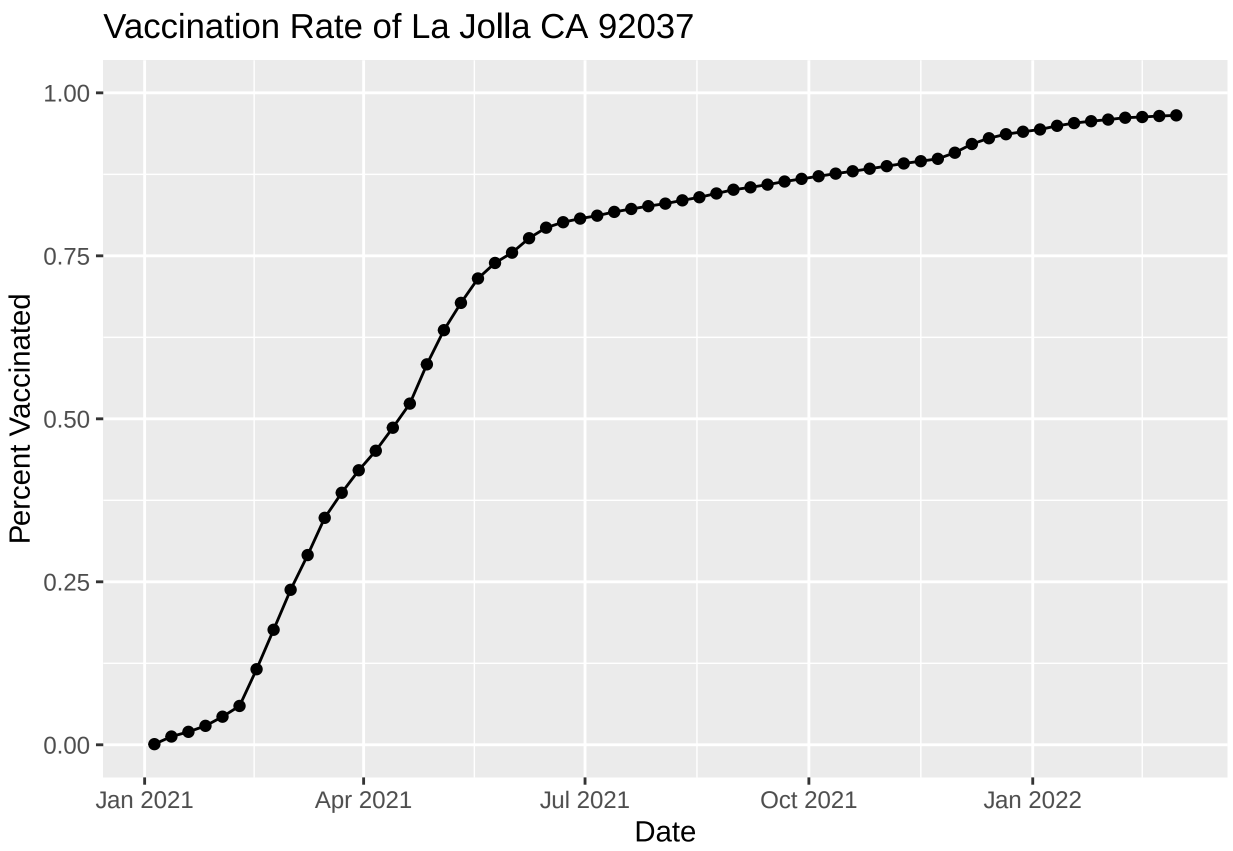
Focus on UCSD/La Jolla

```r
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:

```r
plt <- ggplot(ucsd) +
  aes(x = 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 of La Jolla CA 92037")
plt
```



Comparing to similar sized areas

```
# Subset to all CA areas with a population as large as 92037
vax.36 <- filter(vax, age5_plus_population > 36144 &
                as_of_date == "2022-02-22")

head(vax.36)
```

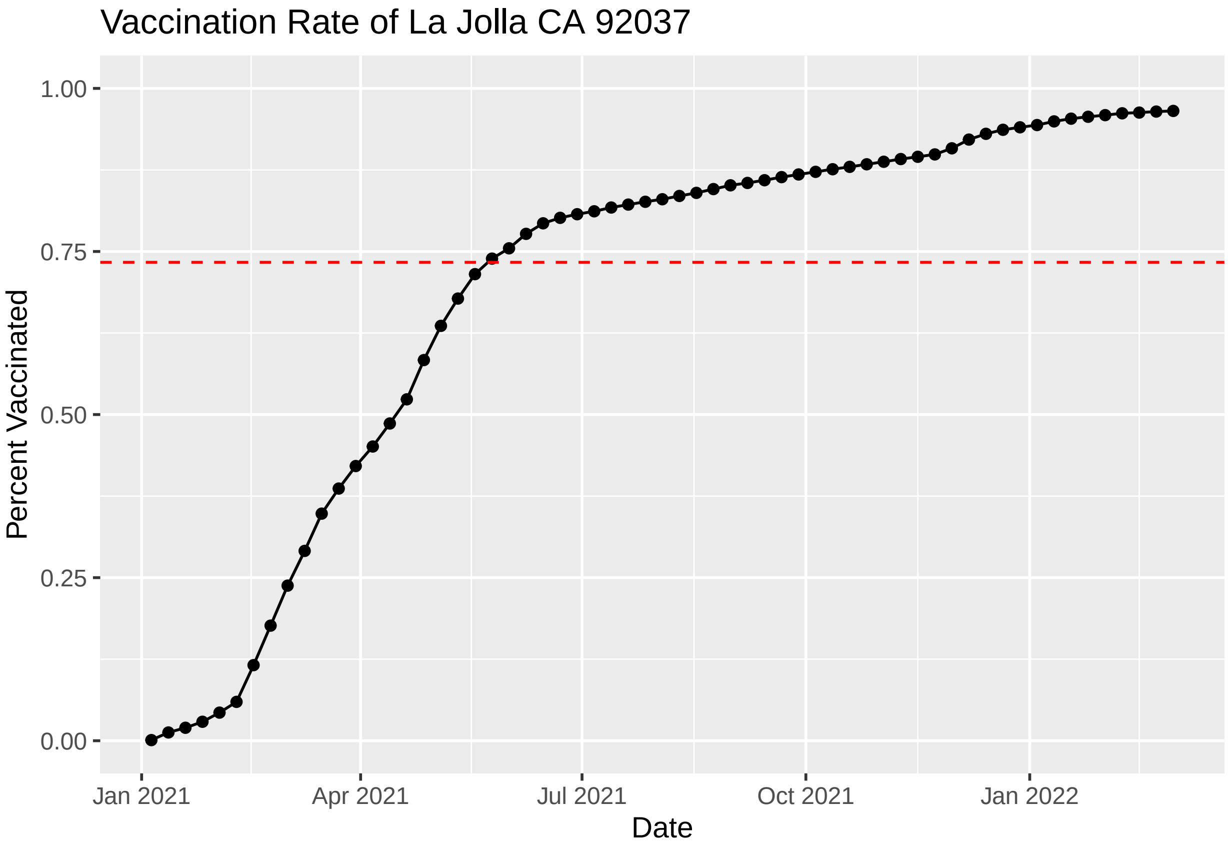
```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction      county
## 1 2022-02-22                    92840                    Orange      Orange
## 2 2022-02-22                    92064                 San Diego   San Diego
## 3 2022-02-22                    92508                 Riverside   Riverside
## 4 2022-02-22                    95403                    Sonoma      Sonoma
## 5 2022-02-22                    90001               Los Angeles Los Angeles
## 6 2022-02-22                    92802                    Orange      Orange
##   vaccine_equity_metric_quartile                 vem_source
## 1                              2 Healthy Places Index Score
## 2                              4 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              3 Healthy Places Index Score
## 5                              1 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1              47302.5                 51902                     40725
## 2              42177.1                 46855                     34266
## 3              32415.3                 36303                     21925
## 4              38545.9                 42294                     33158
## 5              47175.7                 54805                     43075
## 6              35113.6                 39393                     29268
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         4324                               0.784652
## 2                         6861                               0.731320
## 3                         1714                               0.603945
## 4                         2833                               0.783988
## 5                        13917                               0.785968
## 6                         6138                               0.742975
##   percent_of_population_partially_vaccinated
## 1                                   0.083311
## 2                                   0.146430
## 3                                   0.047214
## 4                                   0.066983
## 5                                   0.253937
## 6                                   0.155814
##   percent_of_population_with_1_plus_dose booster_recip_count redacted
## 1                               0.867963               20654       No
## 2                               0.877750               15499       No
## 3                               0.651159               10753       No
## 4                               0.850971               18659       No
## 5                               1.000000               13408       No
## 6                               0.898789               12816       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 "2022-02-22". Add this as a straight horizontal line to your plot from above with the geom_hline() function?

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

```
## [1] 0.733385
```

```
plt + geom_hline(yintercept = mean.36, color = "red", linetype = "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 "2022-02-22"?

As shown by the output of the chunk below.

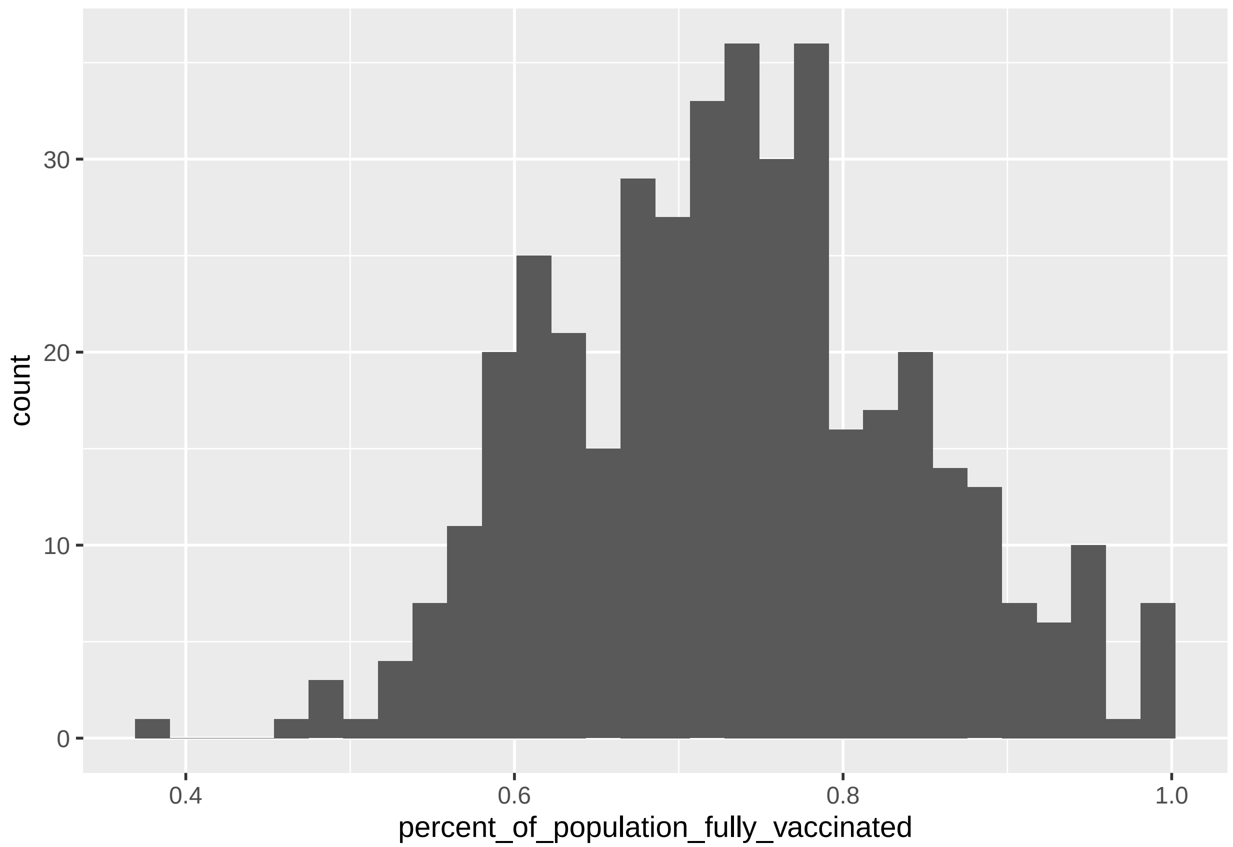
```
summary(vax.36$percent_of_population_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3881  0.6539  0.7333  0.7334  0.8027  1.0000
```

Q18. Using ggplot generate a histogram of this data.

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

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



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

The mean calculated before was 0.7334, as shown below, the 92109 and 92040 ZIP code areas are both **below** 0.7334.

```
vax %>% filter(as_of_date == "2022-02-22") %>%
  filter(zip_code_tabulation_area %in% c("92040","92109")) %>%
  select(percent_of_population_fully_vaccinated)
```

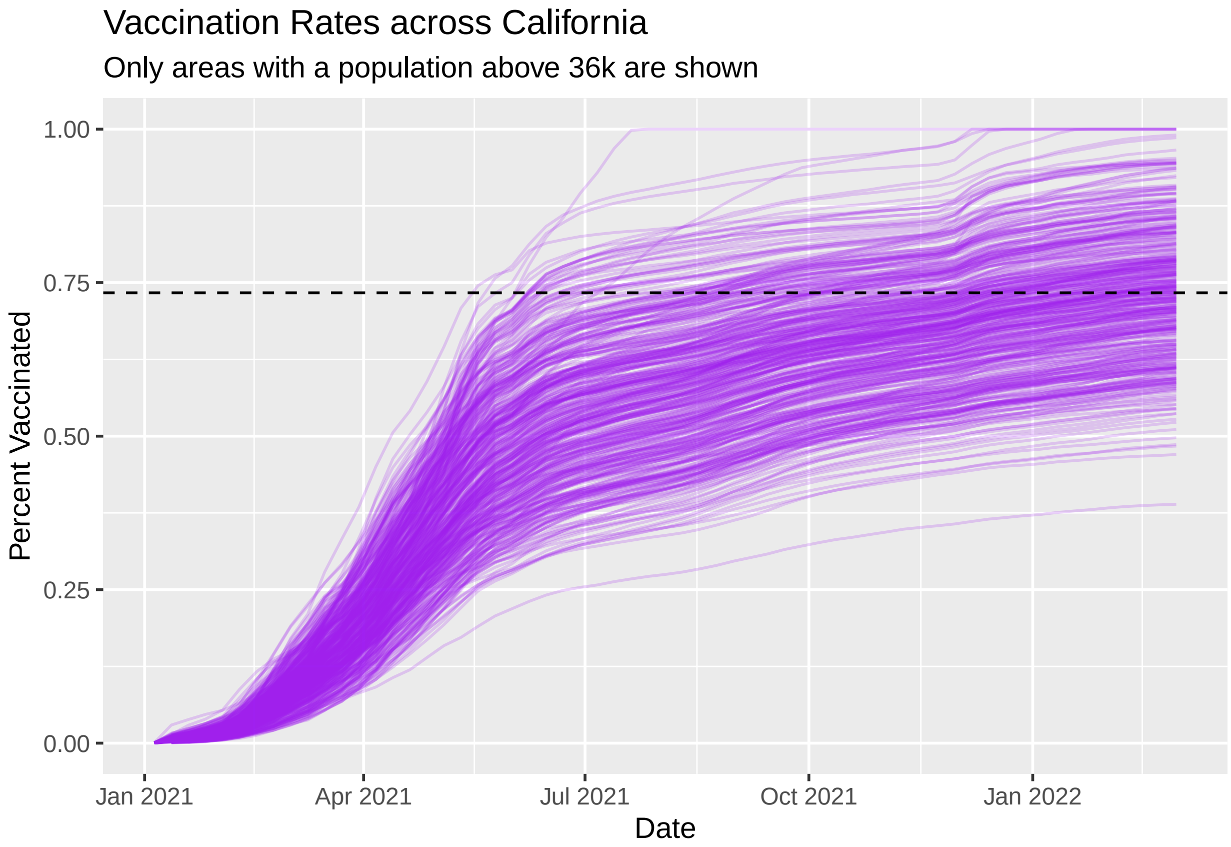
```
##   percent_of_population_fully_vaccinated
## 1                               0.551304
## 2                               0.723044
```

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

```
## Warning: Removed 311 row(s) containing missing values (geom_path).
```



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

Based on the figures, the majority of CA has a vaccination rate above 70%. I feel relatively safe about going back in person with the vaccination status of the state.