COVID-19 Vaccination Rates Mini-Project

Vera Sophia Beliaev

3/3/2022

Goal: view and compare vaccination rates around San Diego, CA

Getting Started

```
#import vax 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_vacinated column details the total number of people fully vaccinated.

Q2. What column details the Zip code tabulation area? The zip_code_tabulation_area column details the Zip code tabulation area.

Q3. What is the earliest date in this dataset? 2021-01-05 is the earliest date in this dataset.

Q4. What is the latest date in this dataset? 2022-03-01 is the latest date in this dataset.

```
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
library(skimr)
```

```r
#get an overview of the data
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
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	5807	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	13685.31	31756.13	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? 15 numeric columns are in the dataset.

Q6. Note that there are "missing values" in the dataset. How many N.A values there in the persons_fully_vaccinated column? There are 18338 NA values in the persons_fully_vaccinated column

Q7. What percent of persons_fully_vaccinated values are missing (to 2 significant figures)? 17% of persons_fully_vaccinated values are missing.

```
18338/107604 * 100
```

```
## [1] 17.04212
```

Q8. [Optional]: Why might this data be missing? The data may be missing for certain groups of people, such as military.

Working with Dates

```
library(lubridate)
```

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

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

```
#today's date
today()
```

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

We want to convert the data in the as_of_date col into a lubridate format

```
#Specify use of year-month-day
vax$as_of_date <- ymd(vax$as_of_date)
#Now we can do calculations with the dates


#Days passed since the first vax in the datasheet
today() - vax$as_of_date[1]
```

Time difference of 422 days

> Q9. How many days have passed since the last update of the dataset? 2 days have passed since the last update of the dataset.

```
today() - vax$as_of_date[107604]
```

Time difference of 2 days

> Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)? There are 61 unique dates in the dataset.

```
#List of all unique dates in the dataset
unique_dates <- unique(vax$as_of_date)


#number of unique dates
length(unique_dates)
```

[1] 61

Working with Zip Codes

```
library(zipcodeR)


#give centroid (latitude & longitude) of zipcode 92037
geocode_zip('92037')
```

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

```
#dist in miles b/w the centroids of 2 zipcodes
zip_distance('92037','92109')
```

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

```
#get census dataabout zipcodes
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(unique(vax$zip_code_tabulation_area))
```

```
head(zipdata)
```

```
## # A tibble: 6 x 24
##   zipcode zipcode_type major_city post_office_city common_city_list county state
##   <chr>   <chr>        <chr>      <chr>                      <blob> <chr>  <chr>
## 1 90001   Standard     Los Angel~ Los Angeles, CA        <raw 44 B> Los A~ CA
## 2 90002   Standard     Los Angel~ Los Angeles, CA        <raw 47 B> Los A~ CA
## 3 90003   Standard     Los Angel~ Los Angeles, CA        <raw 23 B> Los A~ CA
## 4 90004   Standard     Los Angel~ Los Angeles, CA        <raw 34 B> Los A~ CA
## 5 90005   Standard     Los Angel~ Los Angeles, CA        <raw 34 B> Los A~ CA
## 6 90006   Standard     Los Angel~ Los Angeles, CA        <raw 23 B> Los A~ 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>
```

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

#Subset of San Diego county only areas

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

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

```
#Further subset: SD county + population over 10,000
sd.10 <- filter(vax, county == "San Diego" & age5_plus_population >10000)
```

Q11. How many distinct zip codes are listed for San Diego County? 107 distinct zip codes are listed for San Diego County.

```
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? The zip code 92154 has the largest 12+ population in this dataset.

```
largest_12_pop <- sd %>%
  slice_max(age12_plus_population, n=1)
head(largest_12_pop)
```

```
##    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
```

```r
#select all sd county and as of date 2022-02-22 using dplyr
sd0222 <- filter(sd, 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"? The overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2022-02-22" is ~70.4%.

```
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"?

```
library(ggplot2)
```

```
ggplot(sd0222, aes(percent_of_population_fully_vaccinated)) + geom_histogram() + labs(title= "Histogram
```

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

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

Histogram of Vaccination Rates Across San Diego Cuunty
As of 2022−02−22

Focus on UCSD/La Jolla

UCSD is in 92037 zip code with a age 5+ population size of 36,133

```
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=
```

Vaccination rate for La Jolla CA 92109
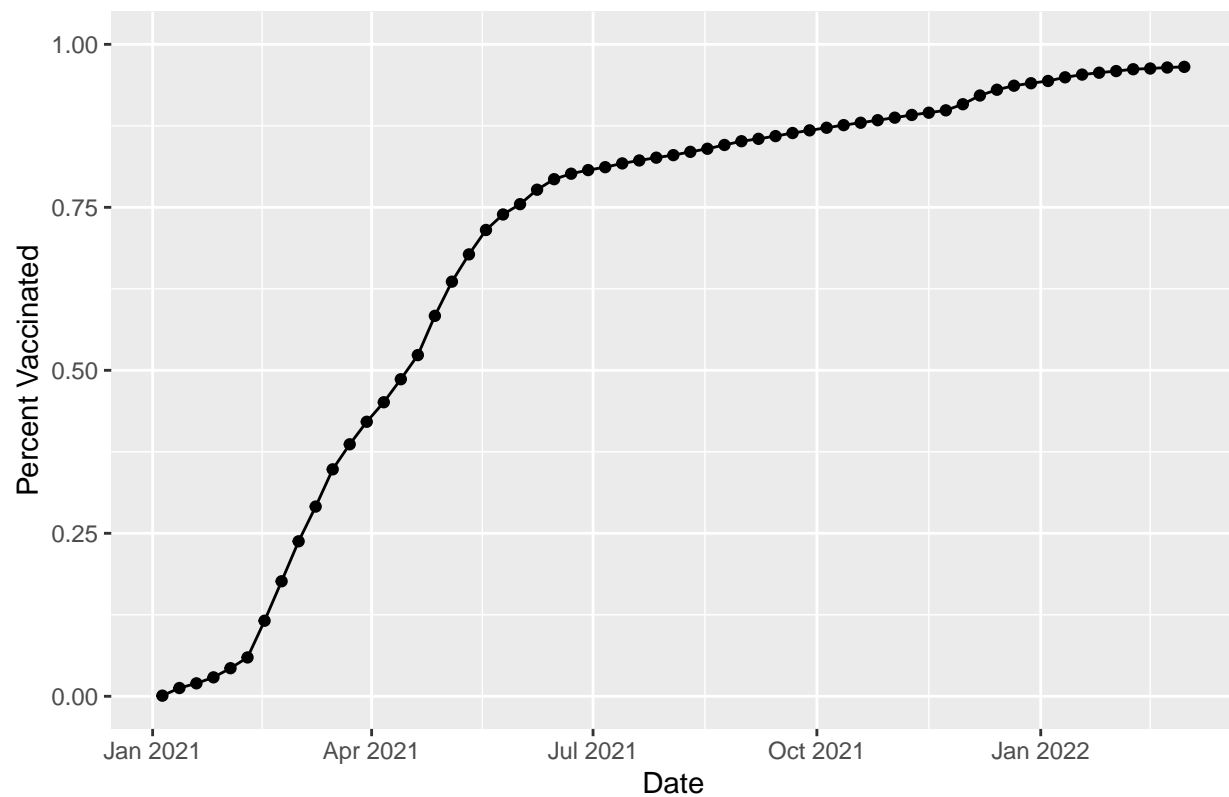
Comparing to Similar Sized Areas

Let's look at all zip code areas in original dataset that have a population such as zip code 92037 or larger

```
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
```
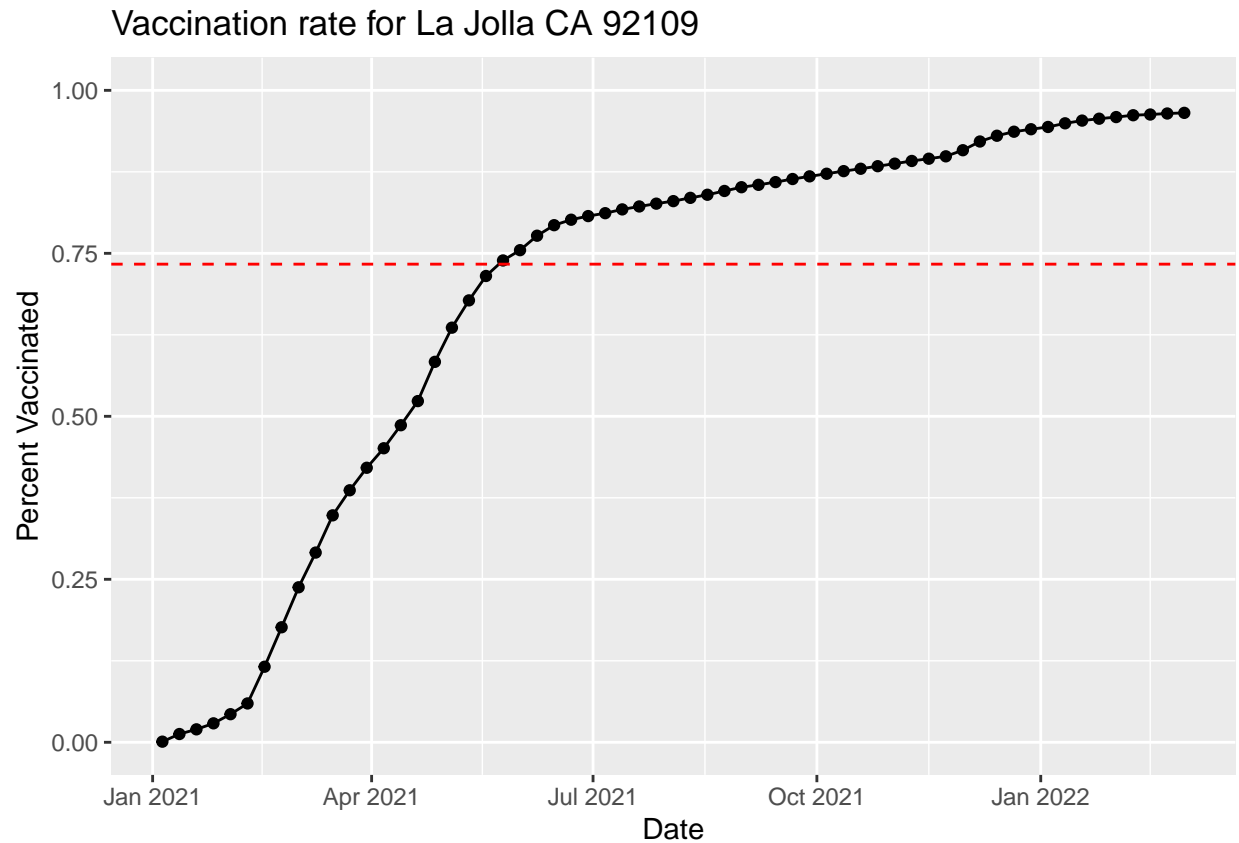
Q16. Calculate the mean "Percent of Population Fully Vaccinated" for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date "2022-02-22". Add this as a straight horizontal line to your plot from above with the geom_hline() function? The mean "Percent of Population Fully Vaccinated" for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date "2022-02-22" is ~73.3%.

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

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

```
#add this percentage to the plot
ggplot(ucsd) + aes(as_of_date, percent_of_population_fully_vaccinated) +geom_point() + geom_line(group=
```

Vaccination rate for La Jolla CA 92109

> 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"?
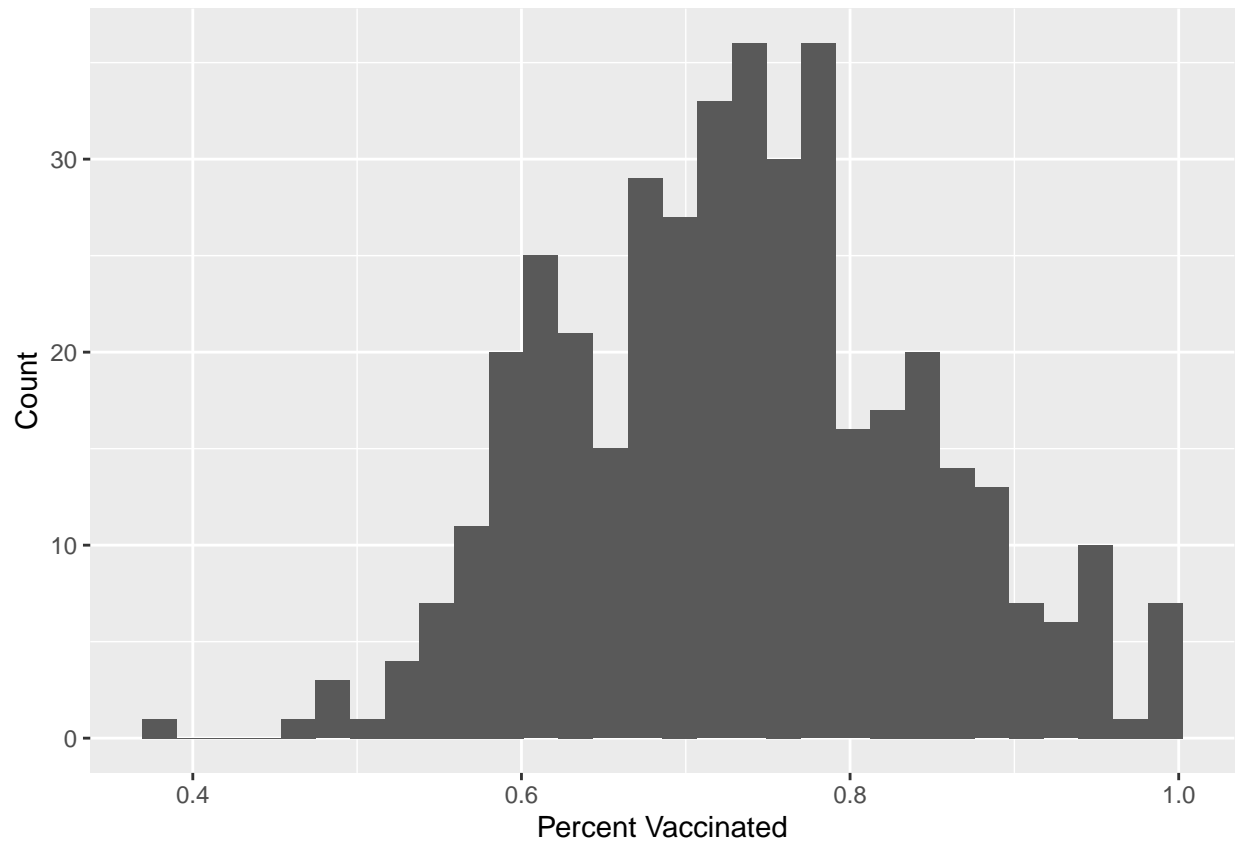
```
summary(vax.36$percent_of_population_fully_vaccinated)
```

```
##    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(percent_of_population_fully_vaccinated)) + geom_histogram() +labs(y= "Count", x= "Pe
```

```
## `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? Both zipcodes 92109 and 92040 are below the average value of 73.3%.

```
#92040
vax %>%
  filter(as_of_date == "2022-02-22") %>%
  filter(zip_code_tabulation_area == "92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
##   percent_of_population_fully_vaccinated
## 1                               0.551304
```

```
#92109
vax %>%
  filter(as_of_date == "2022-02-22") %>%
  filter(zip_code_tabulation_area == "92109") %>%
  select(percent_of_population_fully_vaccinated)
```

```
##   percent_of_population_fully_vaccinated
## 1                               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.

14

```
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_
```

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