Lab 19: Vaccination Rates

Nicholas Do (PID: 15053002)

11/28/2021

submit notes PLEASE READ -I started this assignment late, and I believe some of the data I downloaded from the website has been updated, which may have affected some of my answers below. -The skimr function was outputting a unicode character that was not mapped in LaTeX which was giving me an error when trying to knit so I just commented out where I used the function; the answers I got are from the data given when I used the function. ————

Read the CSV

```
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction         county
## 1 2021-01-05                   92395            San Bernardino San Bernardino
## 2 2021-01-05                   93206                      Kern           Kern
## 3 2021-01-05                   91006                 Los Angeles    Los Angeles
## 4 2021-01-05                   91901                   San Diego      San Diego
## 5 2021-01-05                   92230                   Riverside      Riverside
## 6 2021-01-05                   92662                      Orange         Orange
##   vaccine_equity_metric_quartile               vem_source
## 1                              1 Healthy Places Index Score
## 2                              1 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              3 Healthy Places Index Score
## 5                              1 Healthy Places Index Score
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               35915.3                40888                      NA
## 2                1237.5                 1521                      NA
## 3               28742.7                31347                      19
## 4               15549.8                16905                      12
## 5                2320.2                 2526                      NA
## 6                2349.5                 2397                      NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                           NA                                     NA
## 3                          873                                0.000606
## 4                          271                                0.000710
## 5                           NA                                     NA
## 6                           NA                                     NA
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                         NA
```

```
## 3                                    0.027850
## 4                                    0.016031
## 5                                          NA
## 6                                          NA
##   percent_of_population_with_1_plus_dose
## 1                                     NA
## 2                                     NA
## 3                                0.028456
## 4                                0.016741
## 5                                     NA
## 6                                     NA
##                                                                    redacted
## 1 Information redacted in accordance with CA state privacy requirements
## 2 Information redacted in accordance with CA state privacy requirements
## 3                                                                       No
## 4                                                                       No
## 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?

```
head(vax[,9])
```

```
## [1] NA NA 19 12 NA NA
```

```
head(vax$persons_fully_vaccinated)
```

```
## [1] NA NA 19 12 NA NA
```

Column 9, or vax$persons_fully_vaccinated represents the column for persons fully vaccinated.

Q2. What column details the Zip code tabulation area?

```
head(vax$zip_code_tabulation_area)
```

```
## [1] 92395 93206 91006 91901 92230 92662
```

Q3. What is the earliest date in this dataset?

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

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

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

```
## [1] "2021-11-23"
```

The earliest date is 01/05/2021

Q4. What is the latest date in this dataset?

The latest date is 11/23/2021

```
#check for skimr package
library(skimr)
```

```
#skimr::skim(vax)
```

Q5. How many numeric columns are in this dataset?

There are 9 numeric columns in the dataset.

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

There are 8355 missing values in the persons_fully_vaccinated column

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

```
p <- 8355 / 82908
p
```

```
## [1] 0.1007744
```

10.08% missing

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

This data is probably missing because those individuals are not fully vaccinated yet.

#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] "2021-11-28"
```

```
#Convert to lubridate
#Specify that we are using the Year-mont-day format
vax$as_of_date <- ymd(vax$as_of_date)
```

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

```
## Time difference of 327 days
```

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

```
## Time difference of 322 days
```

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

```
today() - max(vax$as_of_date)
```

```
## Time difference of 5 days
```

It has been 5 days since the last update of the dataset.

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

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

```
##  [1] "2021-01-05" "2021-01-12" "2021-01-19" "2021-01-26" "2021-02-02"
##  [6] "2021-02-09" "2021-02-16" "2021-02-23" "2021-03-02" "2021-03-09"
## [11] "2021-03-16" "2021-03-23" "2021-03-30" "2021-04-06" "2021-04-13"
## [16] "2021-04-20" "2021-04-27" "2021-05-04" "2021-05-11" "2021-05-18"
## [21] "2021-05-25" "2021-06-01" "2021-06-08" "2021-06-15" "2021-06-22"
## [26] "2021-06-29" "2021-07-06" "2021-07-13" "2021-07-20" "2021-07-27"
## [31] "2021-08-03" "2021-08-10" "2021-08-17" "2021-08-24" "2021-08-31"
## [36] "2021-09-07" "2021-09-14" "2021-09-21" "2021-09-28" "2021-10-05"
## [41] "2021-10-12" "2021-10-19" "2021-10-26" "2021-11-02" "2021-11-09"
## [46] "2021-11-16" "2021-11-23"
```

There are 47 unique dates in the dataset.

#Working with ZIP codes

```
library(zipcodeR)
```

```
## Warning: package 'zipcodeR' was built under R version 4.1.2
```

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

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

```
#Distance between zipcodes

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
## # ... 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 SD 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] 5029
```

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

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

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

```
## [1] 107
```

There are 107 unique zip codes in San Diego County

> Q12. What San Diego County Zip code area has the largest 12 + Population in this dataset?

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

```
## [1] 60
```

```
sd$zip_code_tabulation_area[60]
```

```
## [1] 92154
```

92154 is the zip code that has the largest 12+ population in this dataset.

```
anothersd <- filter(vax, county == "San Diego", as_of_date == "2021-11-09")
nrow(anothersd)
```

```
## [1] 107
```

```
head(anothersd[,1])
```

```
## [1] "2021-11-09" "2021-11-09" "2021-11-09" "2021-11-09" "2021-11-09"
## [6] "2021-11-09"
```

```
#skimr::skim(anothersd)
```

> Q13. What is the overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2021-11-09"?

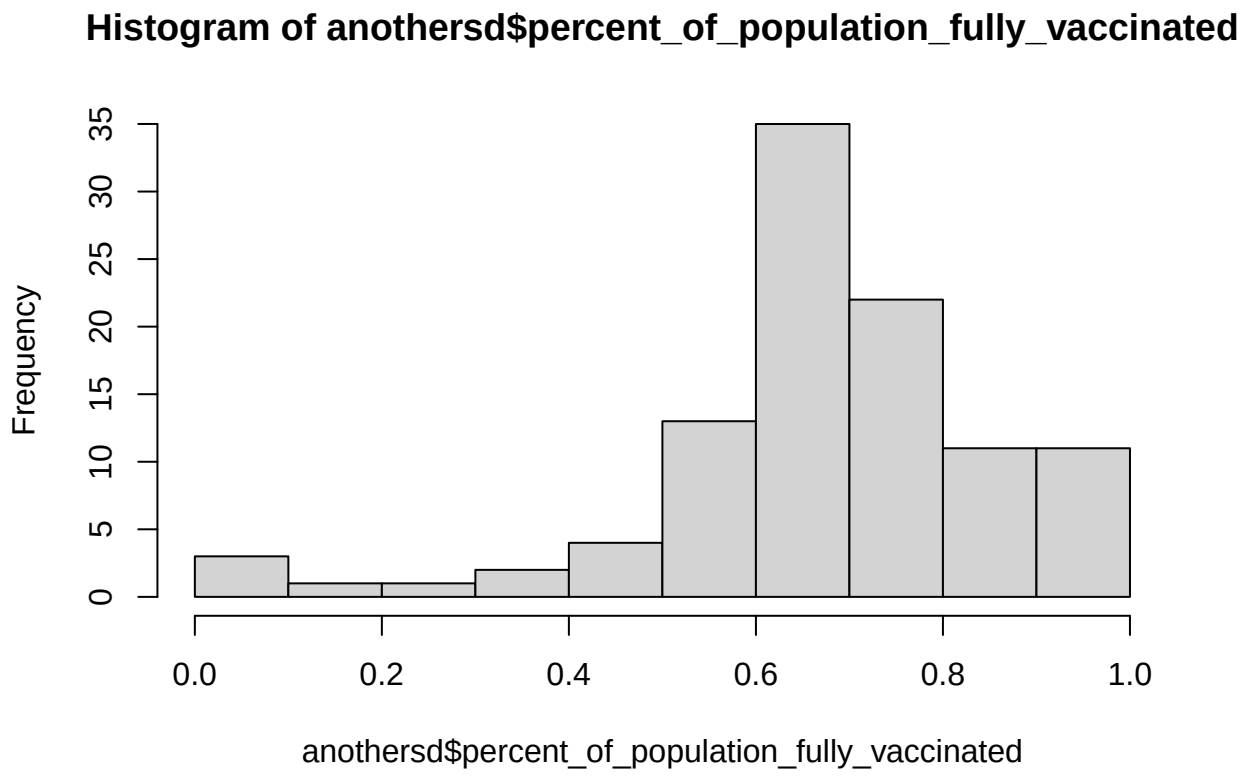
```
anothersd$percent_of_population_fully_vaccinated
```

```
##   [1] 0.627280 0.349969 0.726784 0.010169 0.555131       NA       NA 0.010554
##   [9] 0.081823 0.693335 1.000000 0.634490 0.709656 0.648791 0.762300 0.668860
##  [17] 1.000000 0.690254 0.848730 0.517660 0.625352 0.646069 0.762842 1.000000
##  [25] 0.492764 0.767034 0.464356 0.720677 0.623351       NA 0.746636 0.569588
##  [33] 0.836430 0.517243 0.151316       NA 1.000000 0.830541 0.678779 0.801810
##  [41] 0.710623 0.652142 0.627087 0.581408 0.666651 0.738929 0.890253 1.000000
##  [49] 0.933735 0.781638 0.608357 0.707833 0.665280 1.000000 0.598772 0.724055
##  [57] 0.602453 0.249635 0.829675 0.883379 0.665486 0.741958 0.677454 0.832869
##  [65] 0.460573 0.622995 0.368601 0.857858 0.671967 0.661618 0.632871 0.933972
##  [73] 0.611054 0.805400 0.834455 0.609508 0.564870 0.521700 0.736378 0.598903
##  [81] 0.586433 0.686428 0.718609 0.726210 0.760556 0.789963 0.688959 0.622141
##  [89] 0.701772 0.677109 0.910082 0.635522 0.705998 0.996125 0.588056 0.624514
##  [97] 0.647376 0.739499 0.752886 0.651786 0.607745 1.000000 0.563331 0.659194
## [105] 0.479223 0.696381 0.518689
```

0.963 or 96.3%

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 "2021-11-09"

```
hist(anothersd$percent_of_population_fully_vaccinated)
```



#Focus on UCSD/La Jolla

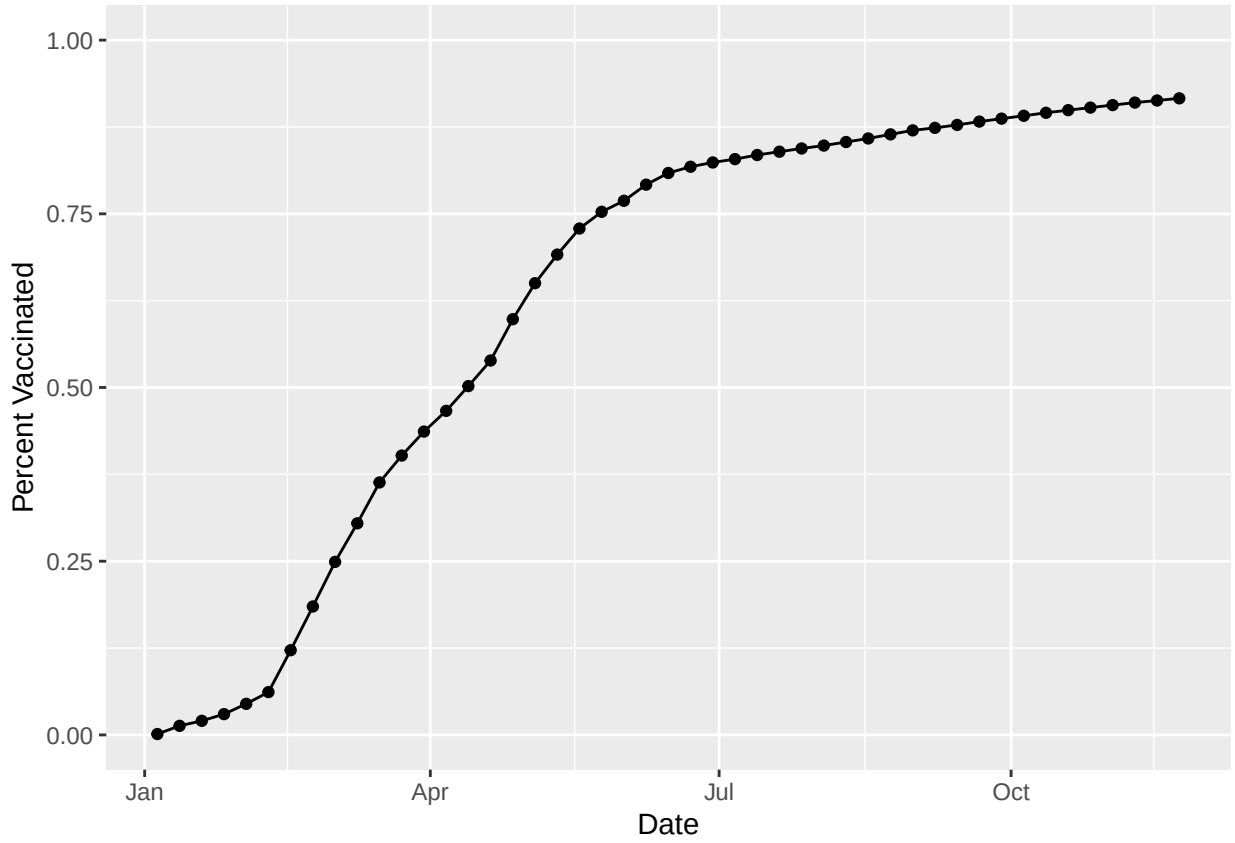
```
ucsd <- filter(sd, zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

```
## [1] 36144
```

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

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

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

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

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

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



#Compare to other similar areas

```
# Subset to all CA areas with a population as large as 92037
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2021-11-16")

head(vax.36)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction         county
## 1 2021-11-16                   92020                San Diego      San Diego
## 2 2021-11-16                   92563                Riverside      Riverside
## 3 2021-11-16                   92806                   Orange         Orange
## 4 2021-11-16                   93291                   Tulare         Tulare
## 5 2021-11-16                   92335           San Bernardino San Bernardino
## 6 2021-11-16                   92618                   Orange         Orange
```

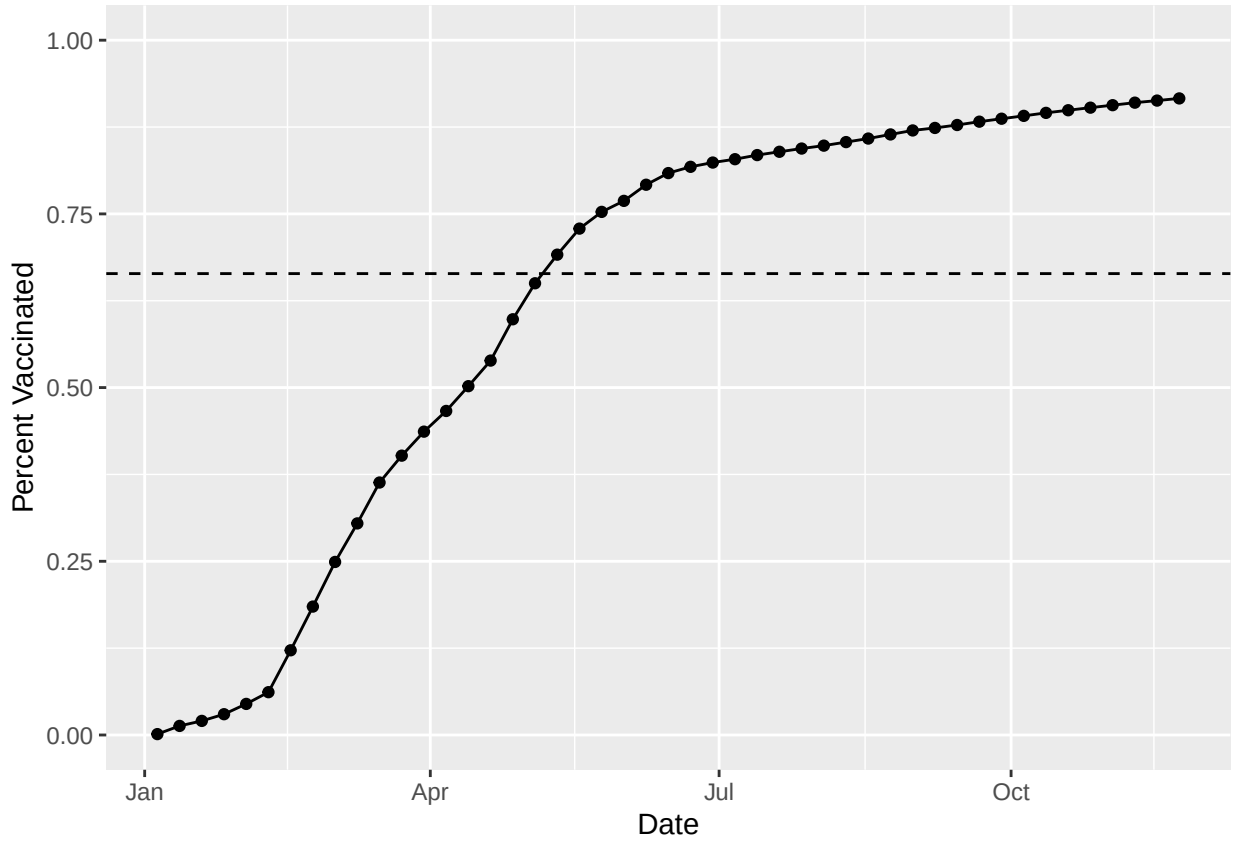
```
##   vaccine_equity_metric_quartile                vem_source
## 1                              2 Healthy Places Index Score
## 2                              3 Healthy Places Index Score
## 3                              2 Healthy Places Index Score
## 4                              1 Healthy Places Index Score
## 5                              1 Healthy Places Index Score
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               49284.5                54991                     35128
## 2               55897.8                63794                     36051
## 3               33050.9                36739                     24810
## 4               46879.7                54254                     27936
## 5               79670.3                91867                     49820
## 6               40348.0                44304                     39695
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         5161                                0.638795
## 2                         4224                                0.565116
## 3                         2355                                0.675304
## 4                         4012                                0.514911
## 5                         5970                                0.542306
## 6                         3936                                0.895969
##   percent_of_population_partially_vaccinated
## 1                                   0.093852
## 2                                   0.066213
## 3                                   0.064101
## 4                                   0.073948
## 5                                   0.064985
## 6                                   0.088841
##   percent_of_population_with_1_plus_dose redacted
## 1                               0.732647       No
## 2                               0.631329       No
## 3                               0.739405       No
## 4                               0.588859       No
## 5                               0.607291       No
## 6                               0.984810       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 "2021-11-16". Add this as a straight horizontal line to your plot from above with the geom_hline() function?

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

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

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

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

## Warning: Use of 'ucsd$as_of_date' is discouraged. Use 'as_of_date' instead.

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



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 "2021-11-16"?

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

```
##    as_of_date         zip_code_tabulation_area local_health_jurisdiction
##  Min.   :2021-11-16   Min.   :90001            Length:411
##  1st Qu.:2021-11-16   1st Qu.:91762            Class :character
##  Median :2021-11-16   Median :92646            Mode  :character
##  Mean   :2021-11-16   Mean   :92862
##  3rd Qu.:2021-11-16   3rd Qu.:94517
##  Max.   :2021-11-16   Max.   :96003
##     county          vaccine_equity_metric_quartile  vem_source
##  Length:411         Min.   :1.000                  Length:411
##  Class :character   1st Qu.:1.000                  Class :character
```

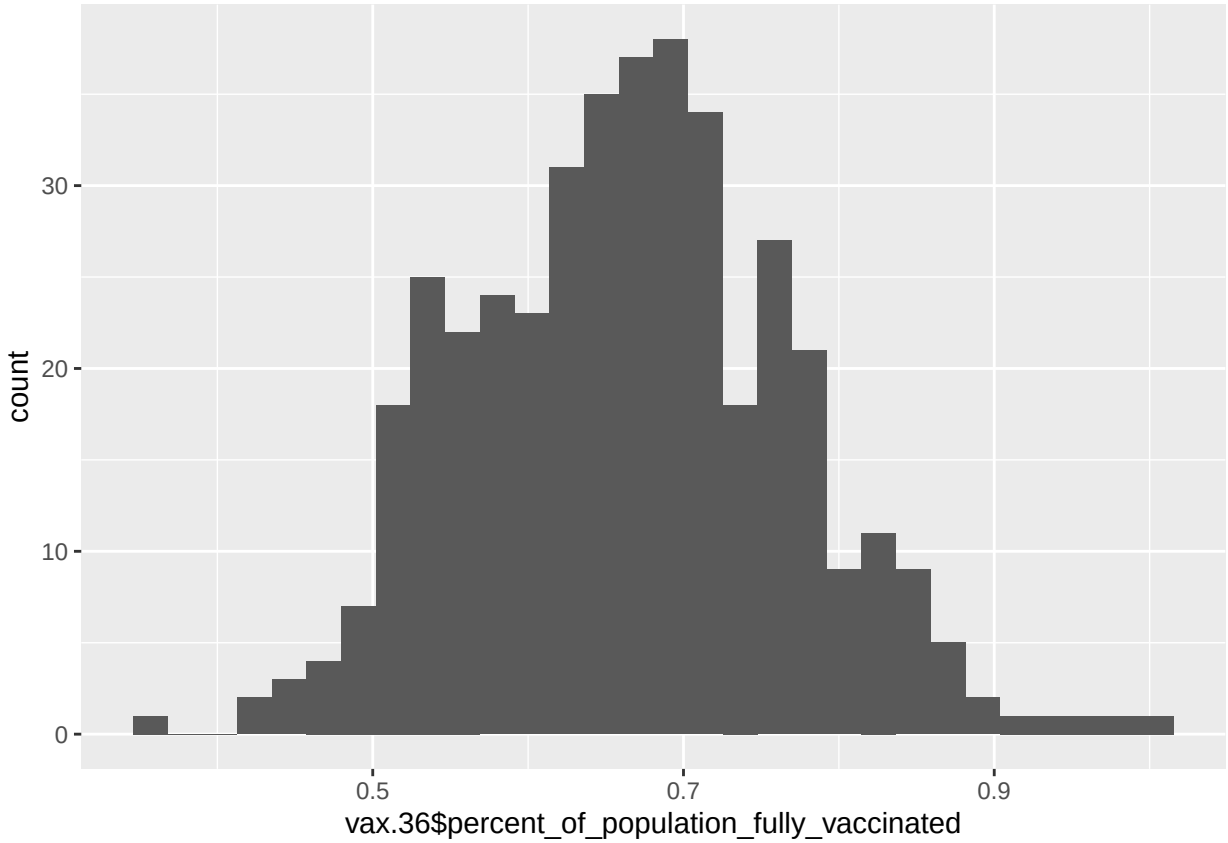
```
##  Mode  :character   Median :2.000                  Mode  :character
##                     Mean   :2.353
##                     3rd Qu.:3.000
##                     Max.   :4.000
##  age12_plus_population age5_plus_population persons_fully_vaccinated
##  Min.   :31651         Min.   : 36181       Min.   :14008
##  1st Qu.:37694         1st Qu.: 41613       1st Qu.:27522
##  Median :43985         Median : 48573       Median :32367
##  Mean   :46847         Mean   : 52012       Mean   :34420
##  3rd Qu.:53932         3rd Qu.: 59168       3rd Qu.:39186
##  Max.   :88557         Max.   :101902       Max.   :71044
##  persons_partially_vaccinated percent_of_population_fully_vaccinated
##  Min.   : 1855                Min.   :0.3529
##  1st Qu.: 2857                1st Qu.:0.5905
##  Median : 3556                Median :0.6662
##  Mean   : 3929                Mean   :0.6640
##  3rd Qu.: 4544                3rd Qu.:0.7298
##  Max.   :14916                Max.   :1.0000
##  percent_of_population_partially_vaccinated
##  Min.   :0.04695
##  1st Qu.:0.06123
##  Median :0.06957
##  Mean   :0.07557
##  3rd Qu.:0.08320
##  Max.   :0.33759
##  percent_of_population_with_1_plus_dose   redacted
##  Min.   :0.4180                         Length:411
##  1st Qu.:0.6689                         Class :character
##  Median :0.7394                         Mode  :character
##  Mean   :0.7384
##  3rd Qu.:0.8075
##  Max.   :1.0000
```

Q18. Using ggplot generate a histogram of this data.

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

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

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



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 == "2021-11-16") %>%
#  filter(zip_code_tabulation_area=="92040") %>%
#  select(percent_of_population_fully_vaccinated)

#vax %>% filter(as_of_date == "2021-11-16") %>%
#  filter(zip_code_tabulation_area=="92109") %>%
#  select(percent_of_population_fully_vaccinated)
```

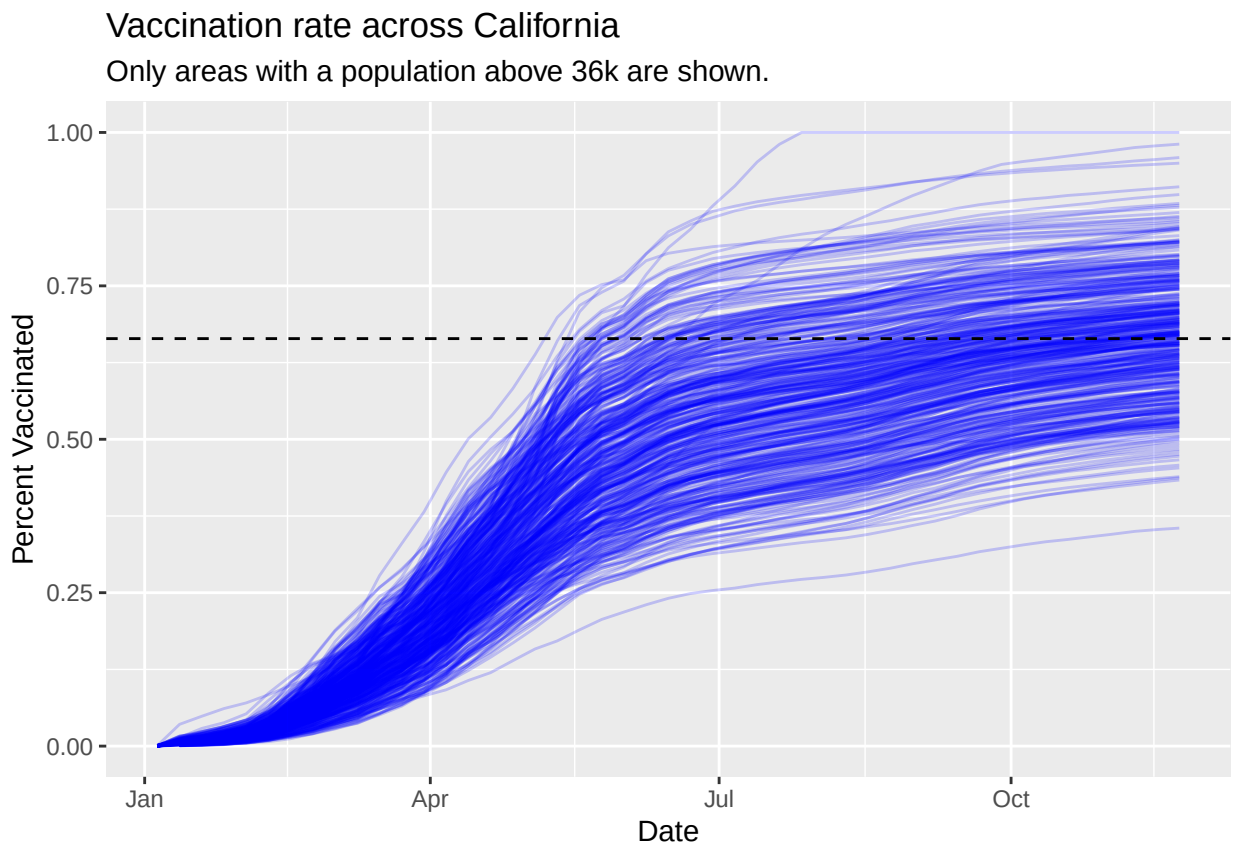
92040 is under, 92109 is above.

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="blue") +
```

```
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated",
       title="Vaccination rate across California",
       subtitle="Only areas with a population above 36k are shown.") +
  geom_hline(yintercept = vaxmean, linetype="dashed")
```

```
## Warning: Removed 176 row(s) containing missing values (geom_path).
```



Q21. How do you feel about traveling for Thanksgiving and meeting for in-person class next Week?

I feel comfortable with it due to the high vaccination rate across California, according to this dataset.