

# COVID-19 Vaccination Rates

Samantha Ruiz

3/3/2022

#Import Vaccine Data

```
vax <- read.csv( "~/Downloads/covid19vaccinesbyzipcode_test (1).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?

person\_fully\_vaccinated

Q2. What column details the Zip code tabulation area?

zip\_code\_tabulation\_area

Q3. What is the earliest date in this dataset?

"2022-03-01"

Q4. What is the latest date in this dataset?

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

```
## [1] "2022-03-01"
```

```
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	93665.111817.39	90001	92257.7593658.5095380.5097635.0					
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	18895.0418993.91	0	1346.95	13685.1031756.1288556.7				
age5_plus_population	0	1.00	20875.2421106.02	0	1460.50	15364.0034877.00101902.0				
persons_fully_vaccinated	18338	0.83	12155.6113063.88	11	1066.25	7374.50	20005.0077744.0			
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_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?

18338

Q7. What percent of persons\_fully\_vaccinated values are missing (to 2 significant figures)?

```
round(18338/107604,2)*100
```

```
## [1] 17
```

```
library(lubridate)
```

```
##
```

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

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

```
##
```

```
## date, intersect, setdiff, union
```

```
today()
```

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

```
age<- today() -ymd("2000-04-13")
age
```

```
## Time difference of 7994 days
```

```
time_length(age, "year")
```

```
## [1] 21.88638
```

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

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

```
## Time difference of 422 days
```

```
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?

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

```
## Time difference of 2 days
```

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

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

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

```
library(zipcodeR)
```

```
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
## # ... 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

## Subset to San Diego county only areas

```
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")
dim(sd)
```

```
## [1] 6527  15
```

```
nrow(sd)
```

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

```
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
```

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

```
sd[which.max(sd$age12_plus_population),]
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction   county
## 91 2021-01-05                92154             San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 91                        2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 91                76365.2                82971                18
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 91                        22                                0.000217
##   percent_of_population_partially_vaccinated
## 91                                0.000265
##   percent_of_population_with_1_plus_dose booster_recip_count
## 91                        0.000482                NA
##
##                                     redacted
## 91 Information redacted in accordance with CA state privacy requirements
```

```
sd$as_of_date[nrow(sd)]
```

```
## [1] "2022-03-01"
```

let's do it with the most recent data set >Q13. What is the overall average “Percent of Population Fully Vaccinated” value for all San Diego “County” as of “2022-02-22”?

```
sd.latest<- filter(sd, as_of_date=="2022-03-01")
mean(sd.latest$percent_of_population_fully_vaccinated, na.rm=TRUE)
```

```
## [1] 0.7052904
```

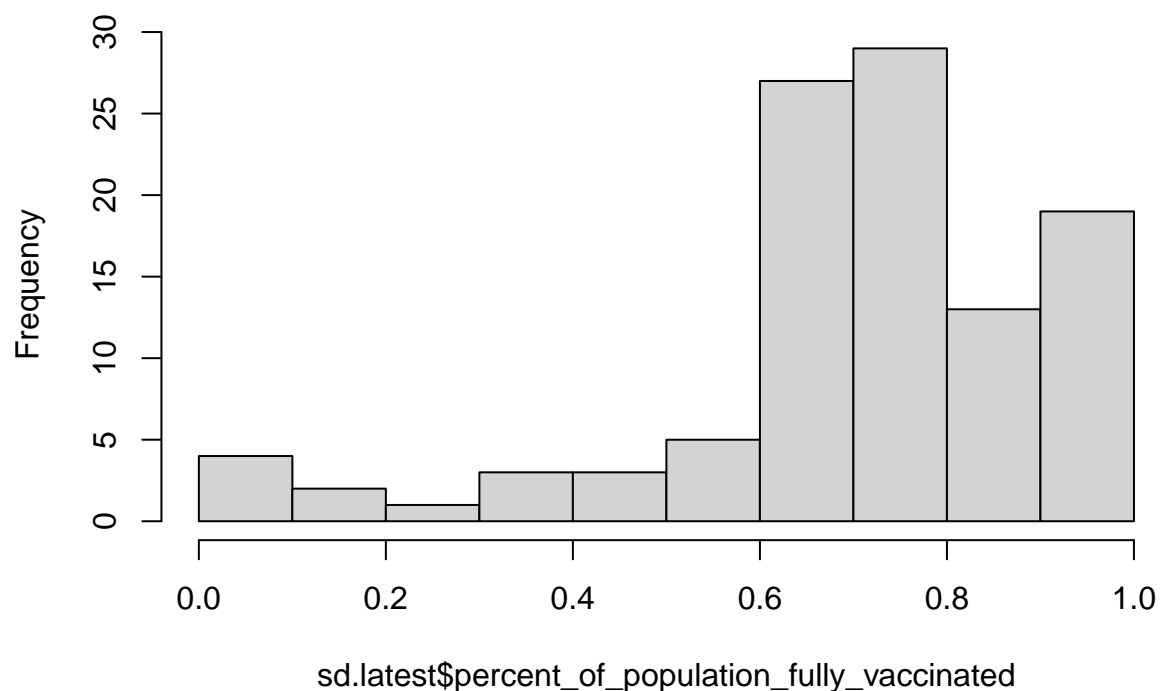
```
summary(sd.latest$percent_of_population_fully_vaccinated, na.rm=T)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.01017 0.65132 0.72452 0.70529 0.82567 1.00000     1
```

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”?

```
hist(sd.latest$percent_of_population_fully_vaccinated)
```

**Histogram of sd.latest\$percent\_of\_population\_fully\_vaccinated**

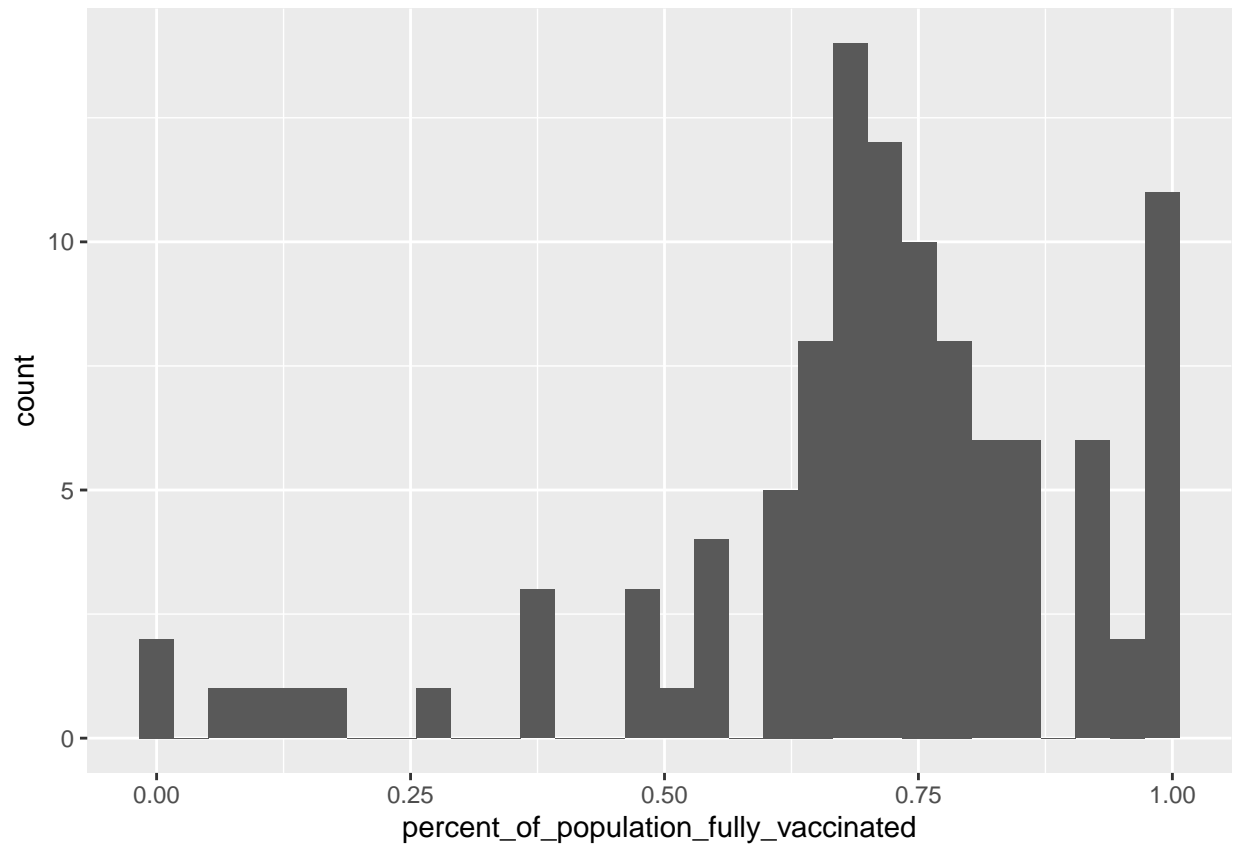


```
library(ggplot2)

ggplot(sd.latest) +
  aes(percent_of_population_fully_vaccinated)+
  geom_histogram()

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

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



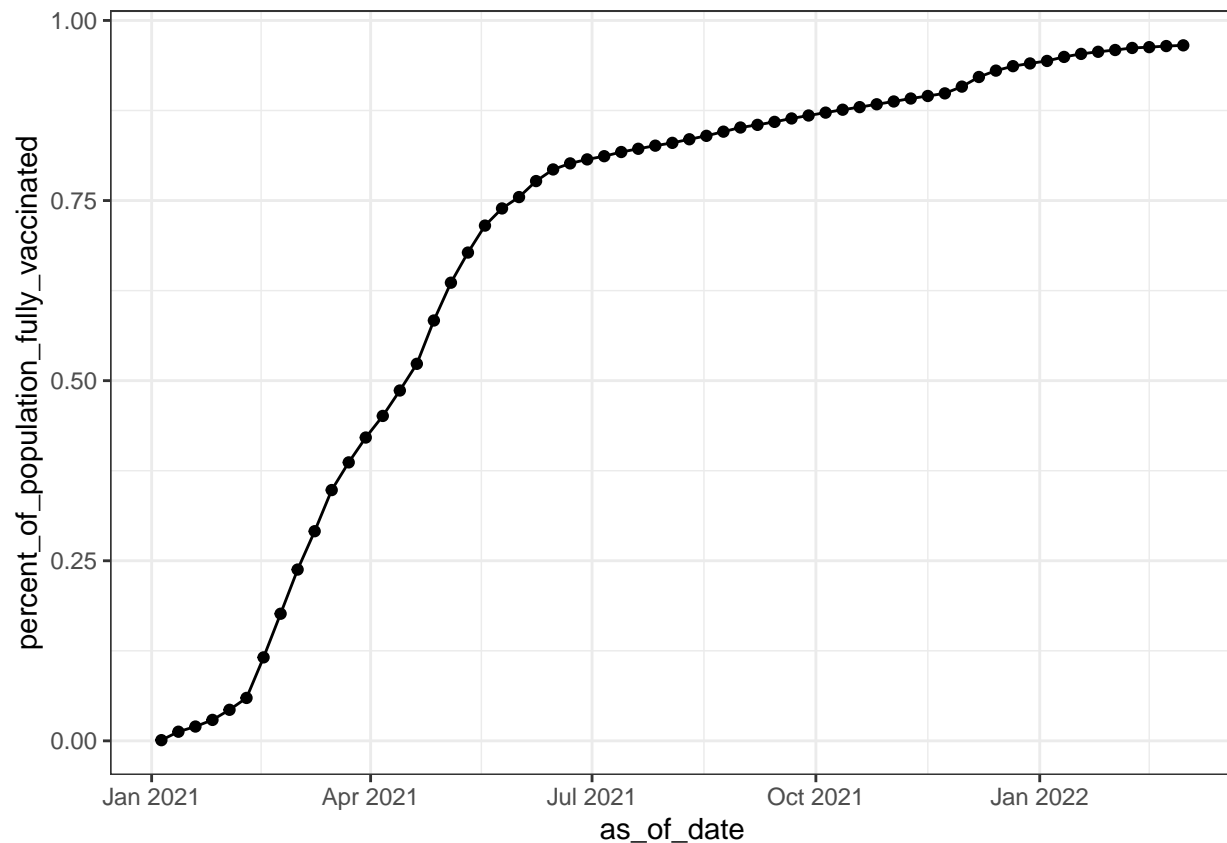
```
ucsd <- filter(sd, zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

```
## [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()+
  theme_bw()
```



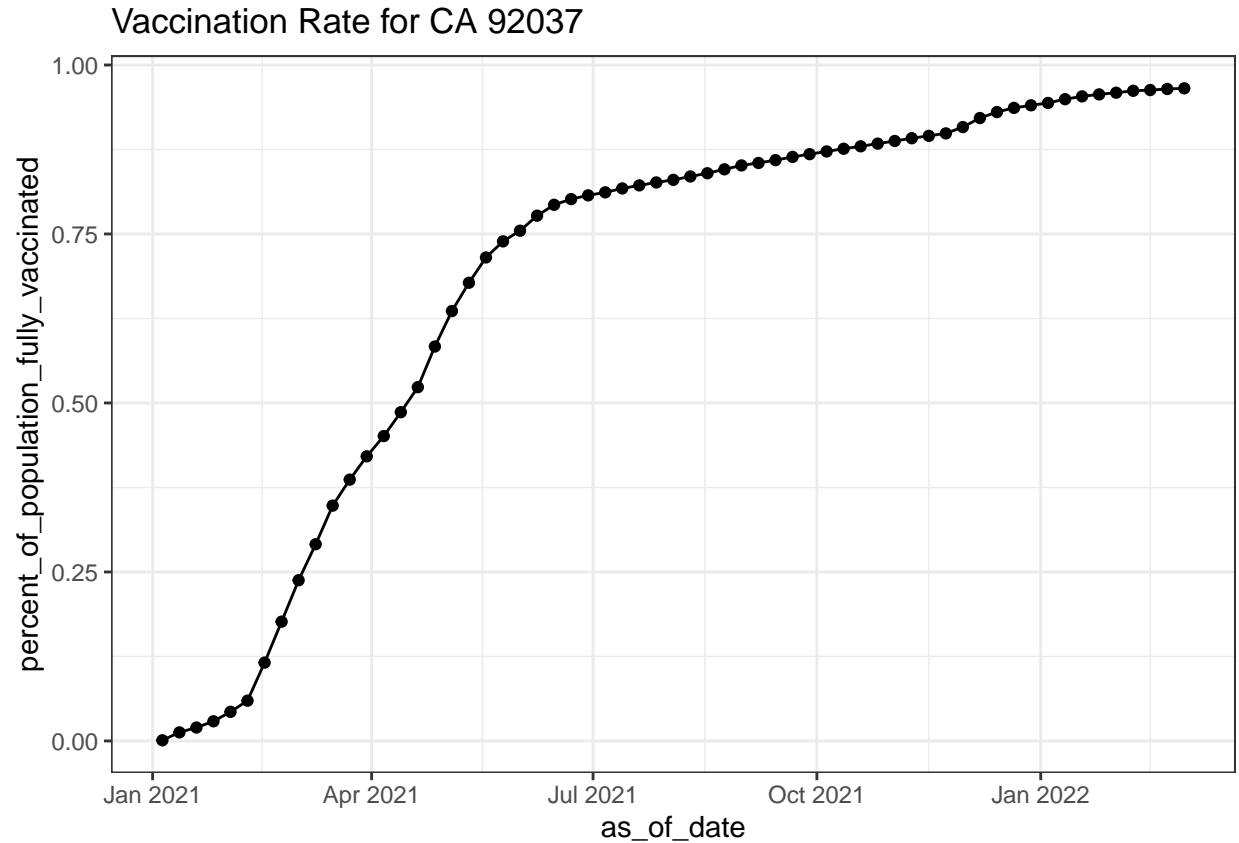


```
vax.36 <- filter(vax, age5_plus_population > 36144 &
  as_of_date == "2022-03-01")
```

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?

```
baseplot<- ggplot(ucsd) +
  aes(as_of_date, percent_of_population_fully_vaccinated)+
  geom_point()+
  geom_line()+
  theme_bw()
```

```
baseplot+
  labs(title = "Vaccination Rate for CA 92037")
```



```
vax.36 <- filter(vax, age5_plus_population>36144 & as_of_date == "2022-03-01")
head(vax.36)
```

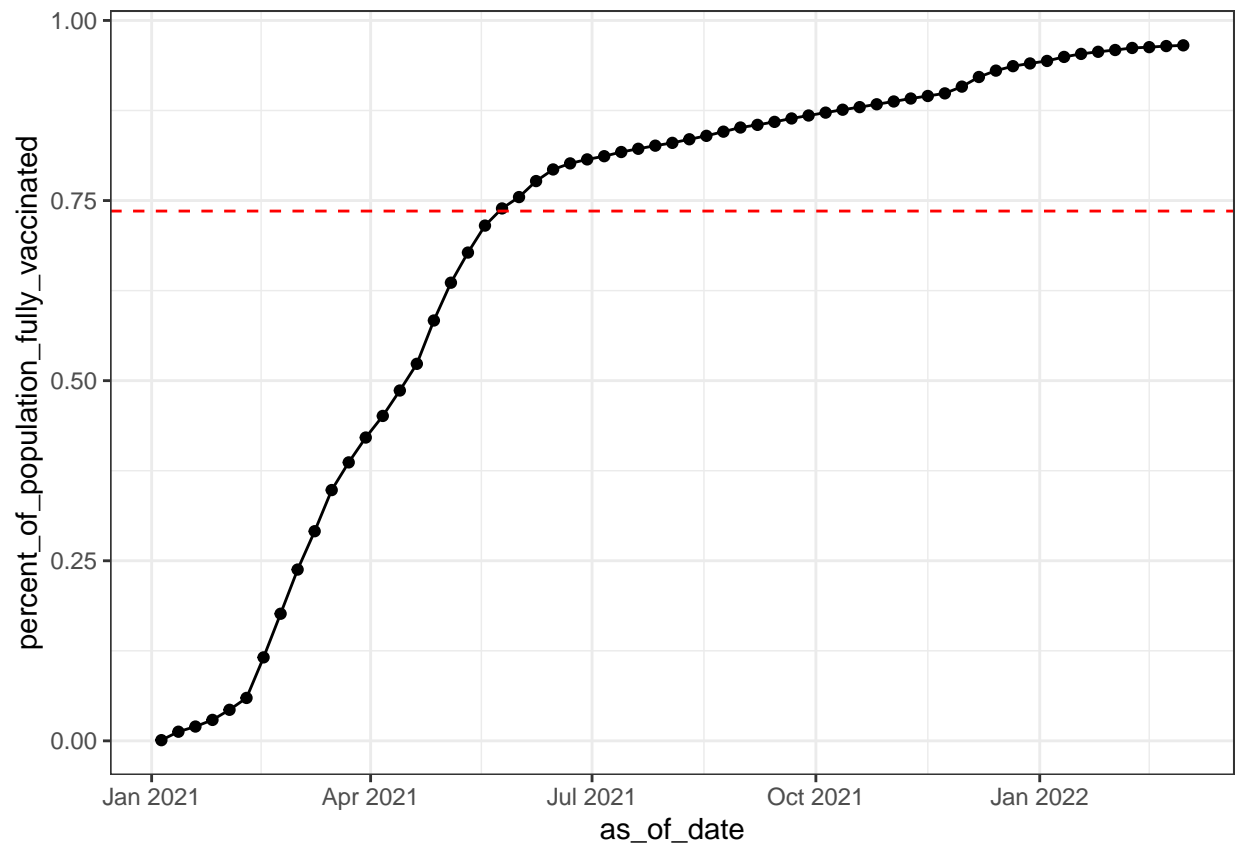
```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2022-03-01           95628             Sacramento Sacramento
## 2 2022-03-01           90808             Long Beach Los Angeles
## 3 2022-03-01           92507             Riverside Riverside
## 4 2022-03-01           92626              Orange    Orange
## 5 2022-03-01           93257              Tulare    Tulare
## 6 2022-03-01           90011             Los Angeles Los Angeles
##   vaccine_equity_metric_quartile      vem_source
## 1                             3 Healthy Places Index Score
## 2                             4 Healthy Places Index Score
## 3                             1 Healthy Places Index Score
## 4                             3 Healthy Places Index Score
## 5                             1 Healthy Places Index Score
## 6                             1 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1                35579.0                38694                28842
## 2                33952.3                37179                29383
## 3                51432.5                55253                34455
## 4                44238.8                47883                33767
## 5                61519.8                70784                42919
## 6                87902.8                101902                65342
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                        1990                        0.745387
```

```
## 2                2112                0.790312
## 3                3947                0.623586
## 4                2937                0.705198
## 5                5868                0.606338
## 6               15255                0.641224
## percent_of_population_partially_vaccinated
## 1                0.051429
## 2                0.056806
## 3                0.071435
## 4                0.061337
## 5                0.082900
## 6                0.149703
## percent_of_population_with_1_plus_dose booster_recip_count redacted
## 1                0.796816            16913        No
## 2                0.847118            17253        No
## 3                0.695021            15073        No
## 4                0.766535            17595        No
## 5                0.689238            17740        No
## 6                0.790927            19928        No
```

```
ave.36 <- mean(vax.36$percent_of_population_fully_vaccinated, na.rm=T)
ave.36
```

```
## [1] 0.7353974
```

```
baseplot+geom_hline(yintercept= ave.36, linetype=2, 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 “2022-02-22”?

```
summary(vax.36$percent_of_population_fully_vaccinated, na.rm=T)
```

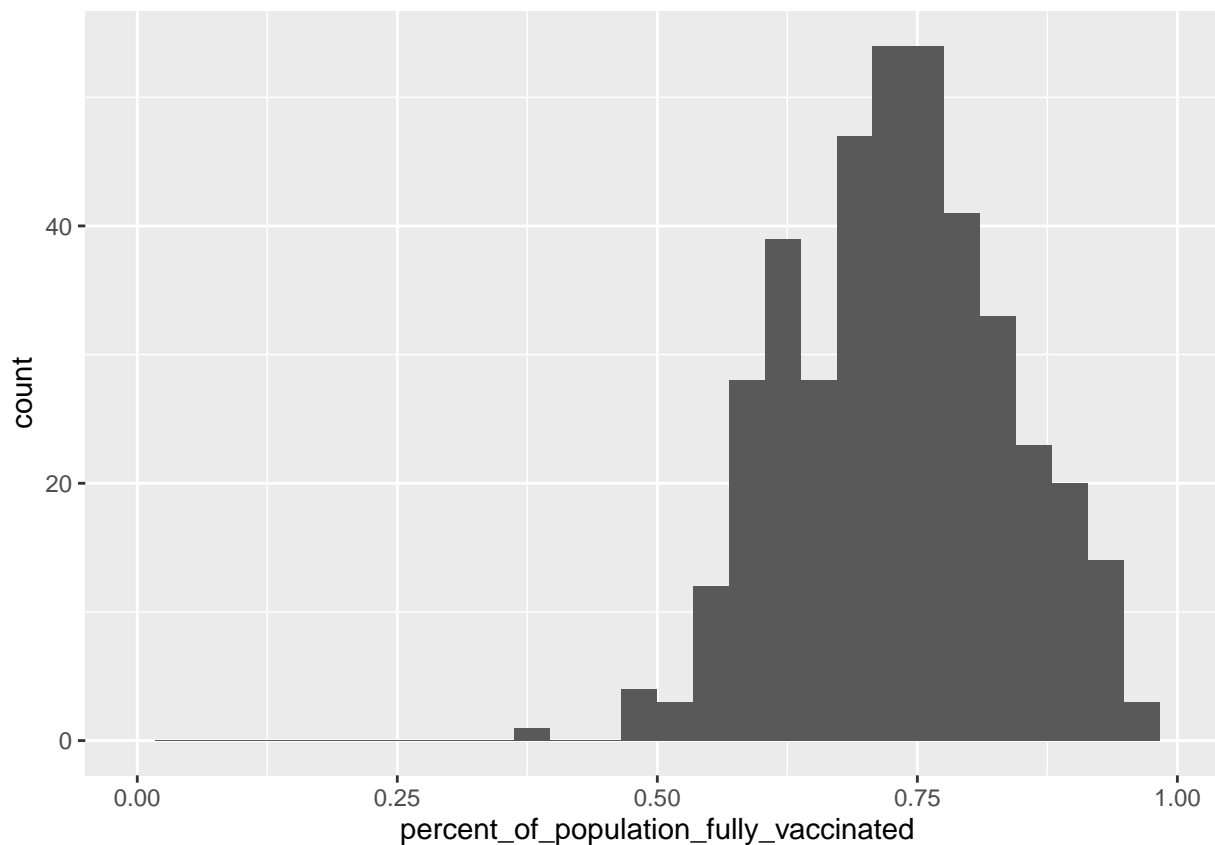
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3890  0.6554  0.7350  0.7354  0.8044  1.0000
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36)+
  aes(percent_of_population_fully_vaccinated)+
  geom_histogram()+
  xlim(c(0,1))
```

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

```
## Warning: Removed 2 rows containing missing values (geom_bar).
```



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 == "2022-03-01") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
## percent_of_population_fully_vaccinated
## 1 0.551981
```

```
vax %>% filter(as_of_date == "2022-03-01") %>%
  filter(zip_code_tabulation_area=="92037") %>%
  select(percent_of_population_fully_vaccinated)
```

```
## percent_of_population_fully_vaccinated
## 1 0.965444
```

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 vaccinate",
       title="Vaccination RatesCross California",
       subtitle= "only areas with a population abouve 36k shown") +
  geom_hline(yintercept = ave.36, linetype=2)
```

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

Vaccination RatesCross California  
only areas with a population abouve 36k shown

