# class17_vaccines

Jocelyn Olvera

11/24/2021

#Looking at COVID-19 vaccination rates during the Holiday travel #Data obtained from: https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data-by-zip-code

**Importing Data**

```
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
```

## Q: How many of entries 'nrow(vax)'?

```
nrow(vax)
```

```
## [1] 82908
```

## Using skimr package and the 'skim()' function to get a quick overview of the data.

**using 'skimr::skim' allows you to use the package only once without loading it in with 'library()'.**

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

Table 1: Data summary

| | |
|---|---|
| Name | vax |
| Number of rows | 82908 |
| Number of columns | 14 |
| _____ | |
| Column type frequency: | |
| character | 5 |
| numeric | 9 |
| _____ | |
| 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 | 47 | 0 |
| local_health_jurisdiction | 0 | 1 | 0 | 15 | 235 | 62 | 0 |
| county | 0 | 1 | 0 | 15 | 235 | 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 | 1089 | 0.95 | 2.44 | 1.11 | 1 | 1.00 | 2.00 | 3.00 | 4.0 | |
| age12_plus_population | 0 | 1.00 | 18895.04 | 18993.94 | 0 | 1346.95 | 13685.10 | 31756.12 | 88556.7 | |
| age5_plus_population | 0 | 1.00 | 20875.24 | 21106.04 | 0 | 1460.50 | 15364.00 | 34877.00 | 101902.0 | |
| persons_fully_vaccinated | 8355 | 0.90 | 9585.35 | 11609.12 | 11 | 516.00 | 4210.00 | 16095.00 | 71219.0 | |
| persons_partially_vaccinated | 8355 | 0.90 | 1894.87 | 2105.55 | 11 | 198.00 | 1269.00 | 2880.00 | 20159.0 | |
| percent_of_population_fully_vaccinated | 8355 | 0.90 | 0.43 | 0.27 | 0 | 0.20 | 0.44 | 0.63 | 1.0 | |
| percent_of_population_partially_vaccinated | 8355 | 0.90 | 0.10 | 0.10 | 0 | 0.06 | 0.07 | 0.11 | 1.0 | |
| percent_of_population_with_1_plus_dose | 8355 | 0.90 | 0.51 | 0.26 | 0 | 0.31 | 0.53 | 0.71 | 1.0 | |

## Working with dates

Notice that one of these columns is a date column. Working with time and dates gets annoying quickly. We can use **lubridate** package to make this easy…

```
library(lubridate)
```

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

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

```
today()
```

```
## [1] "2021-11-24"
```

## Q1. What column details the total number of people fully vaccinated?

## Column: "persons_fully_vaccinated"

## Q2. What column details the Zip code tabulation area?

## Column: "zip_code_tabulation_area"

## Q3. What is the earliest date in this dataset?

```
vax$as_of_date[1]
```

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

##This will not work because our data column was read as character…

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

##Instead, try overwritting the first column to match the format and assign to 'd' vector.

```
d <- ymd(vax$as_of_date[1])
```

```
today() - d[1]
```

```
## Time difference of 323 days
```

##I will make the 'as_of_date' column Date format…

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

##Q. When was the dataset last updated? # Q4. What is the latest date in this dataset?

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

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

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

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

```
## Time difference of 1 days
```

##Q. How many days does the dataset span?

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

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

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

##Q. How many different ZIP code areas are in this dataset?

```
length(unique(vax$zip_code_tabulation_area))
```

```
## [1] 1764
```

### Working with ZIP codes

##To work with ZIP codes we use the **zipcodeR**

```
library(zipcodeR)
```

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

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

### Focus on the San Diego area

```
inds <- vax$county == "San Diego"
nrow(vax[inds,])
```

```
## [1] 5029
```

### Sub-setting can get tedious and complicated quickly when you have multiple things we want to subset by.

```
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
```

**We will use the 'filter()' function to do our subsetting from now on.**

**We want to focus on San Diego.**

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

```
## [1] 5029
```

**More subsetting.**

```
sd.20 <- filter(vax, county == "San Diego" &
                age5_plus_population > 20000)
nrow(sd.20)
```

```
## [1] 3055
```

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

```
length(unique(sd))
```

```
## [1] 14
```

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

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

```
sd.now <- filter(vax, county == "San Diego",
                as_of_date == "2021-11-23")

head(sd.now)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-11-23                   92120                San Diego San Diego
## 2 2021-11-23                   91962                San Diego San Diego
## 3 2021-11-23                   92155                San Diego San Diego
## 4 2021-11-23                   92147                San Diego San Diego
## 5 2021-11-23                   91913                San Diego San Diego
## 6 2021-11-23                   92114                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 1                              4 Healthy Places Index Score
## 2                              3 Healthy Places Index Score
## 3                             NA            No VEM Assigned
## 4                             NA            No VEM Assigned
## 5                              3 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               26372.9                28414                   21234
```

```
## 2                1758.7                 2020                       948
## 3                 456.0                  456                        70
## 4                 518.0                  518                        NA
## 5               43514.7                50461                     37974
## 6               59050.7                64945                     43708
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         3198                                0.747308
## 2                          126                                0.469307
## 3                           20                                0.153509
## 4                           NA                                      NA
## 5                         6690                                0.752542
## 6                         6261                                0.673000
##   percent_of_population_partially_vaccinated
## 1                                   0.112550
## 2                                   0.062376
## 3                                   0.043860
## 4                                         NA
## 5                                   0.132578
## 6                                   0.096405
##   percent_of_population_with_1_plus_dose
## 1                               0.859858
## 2                               0.531683
## 3                               0.197369
## 4                                     NA
## 5                               0.885120
## 6                               0.769405
##                                                                  redacted
## 1                                                                      No
## 2                                                                      No
## 3                                                                      No
## 4 Information redacted in accordance with CA state privacy requirements
## 5                                                                      No
## 6                                                                      No
```

```
sd.now$percent_of_population_fully_vaccinated
```

```
##   [1] 0.747308 0.469307 0.153509       NA 0.752542 0.673000 0.171930 0.628913
##   [9] 0.355234 0.686848 0.496899 0.694990 0.725720 0.576161 0.652680 0.806525
##  [17] 0.718495 1.000000 0.633126 0.835713 0.855294 0.657697 0.631422 0.846959
##  [25] 0.769692 1.000000       NA 0.628480 0.844500       NA 0.683163 0.523179
##  [33] 0.082372 0.771474 0.464088 0.592998 0.651956 0.632170 0.571643 0.656561
##  [41] 0.603904 0.626561 0.691278 0.723539 0.813734 0.707481 0.730845 0.617369
##  [49] 0.841184 0.743946 0.759115 1.000000 0.676833 0.944622 0.667700 0.638762
##  [57] 0.766287 1.000000 0.711136 0.743590 0.798508 0.916196 0.694622 0.613783
##  [65] 0.526130 0.641578 0.700739 0.484584 0.370307 0.594036 0.618409 0.682470
##  [73] 0.863395 0.840959 1.000000 0.249635 0.610675 1.000000 0.729044 0.614751
##  [81] 0.586075 0.699525 1.000000 0.769195 0.715999 0.670258 1.000000 0.521976
##  [89] 0.010726 0.732941 0.632636 0.559401 0.010169 0.639952 0.891644 0.713647
##  [97] 0.672947 0.653994 0.569850 0.665486 0.523125 0.673358 0.951807 0.604313
## [105] 0.744649 0.787222 0.894858
```

```
sd.vax <- sd.now$percent_of_population_fully_vaccinated
summary(sd.vax)
```

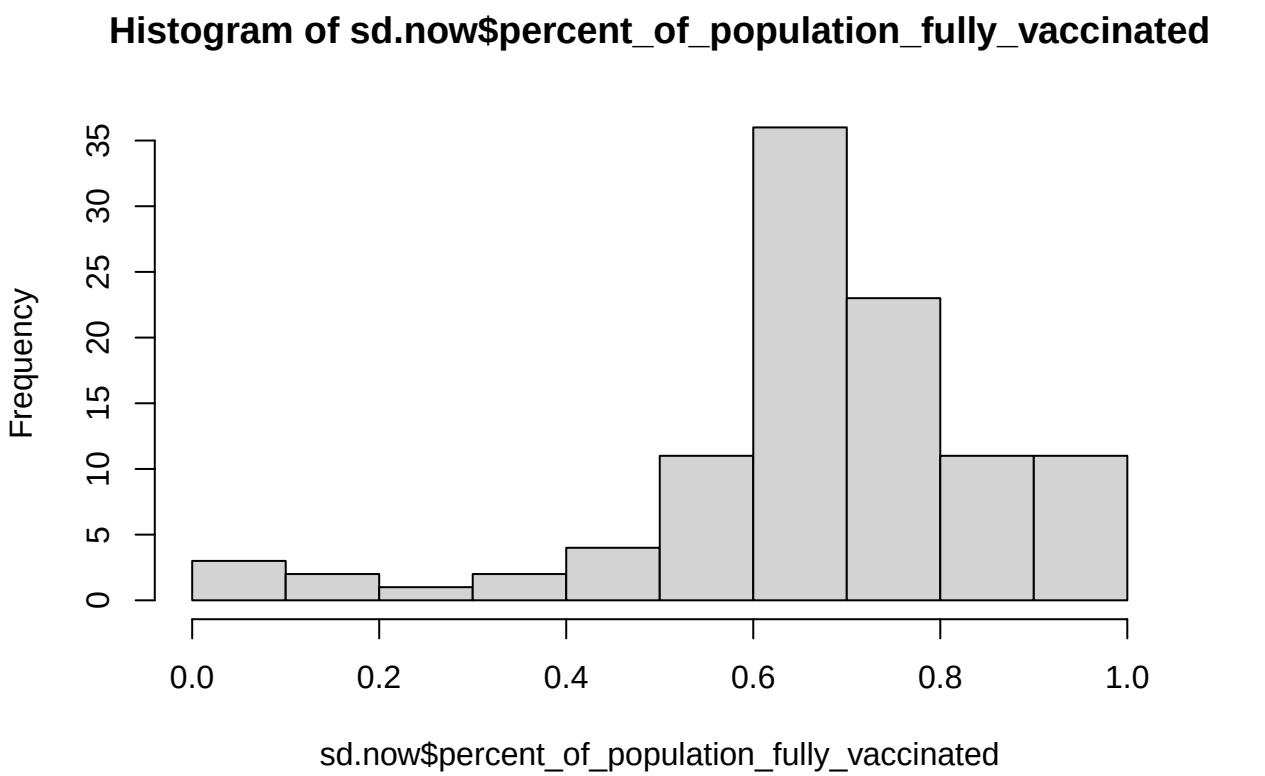
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
```

```
## 0.01017 0.61301 0.67965 0.67400 0.76932 1.00000       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"?**

Base R histogram

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




### This plot above is going to be susceptible to being skewed byt ZIP code areas with small populations. This will have big effects for just a small number of unvaxed folks…

##Q. What is the population population of the 92037 ZIP code area?

```
sd.92037 <- filter(sd.now, zip_code_tabulation_area == "92037")
head(sd.92037)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-11-23                   92037                San Diego San Diego
##   vaccine_equity_metric_quartile               vem_source
## 1                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               33675.6                36144                    33115
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         7660                                0.916196
##   percent_of_population_partially_vaccinated
## 1                                    0.21193
##   percent_of_population_with_1_plus_dose redacted
## 1                                      1       No
```

##Q. What is the average vaccination value for this UCSD/La JOlla ZIP code area?

```
sd.92037$age5_plus_population
```

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

```
sd.92037$percent_of_population_fully_vaccinated
```

```
## [1] 0.916196
```

### Area interested in, Escondido, CA 92025.

```
sd.92025 <- filter(vax, zip_code_tabulation_area == "92025")
sd.92025$age5_plus_population
```

```
##  [1] 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162
## [13] 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162
## [25] 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162
## [37] 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162 49162
```

```
sd.92025$percent_of_population_fully_vaccinated
```

```
##  [1] 0.000590 0.003133 0.006733 0.010476 0.019934 0.030694 0.045645 0.062304
##  [9] 0.080204 0.099487 0.121801 0.136325 0.152740 0.168891 0.197164 0.218400
## [17] 0.252166 0.281111 0.306965 0.336113 0.363268 0.380395 0.402648 0.422460
## [25] 0.435357 0.445324 0.454253 0.462939 0.472052 0.479313 0.485171 0.492250
## [33] 0.500732 0.510415 0.521175 0.531000 0.541292 0.550100 0.557911 0.565335
## [41] 0.571071 0.575892 0.580550 0.584171 0.588056 0.591270 0.594036
```

### Area interested in by classmate 92124.

```
sd.92124 <- filter(sd.now, zip_code_tabulation_area == "92124")
sd.92124$age5_plus_population
```

```
## [1] 29040
```

```
sd.92124$percent_of_population_fully_vaccinated
```

```
## [1] 0.559401
```

```
sd.92103 <- filter(sd.now, zip_code_tabulation_area == "92103")
sd.92103$age5_plus_population
```

```
## [1] 33213
```

```
sd.92103$percent_of_population_fully_vaccinated
```

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

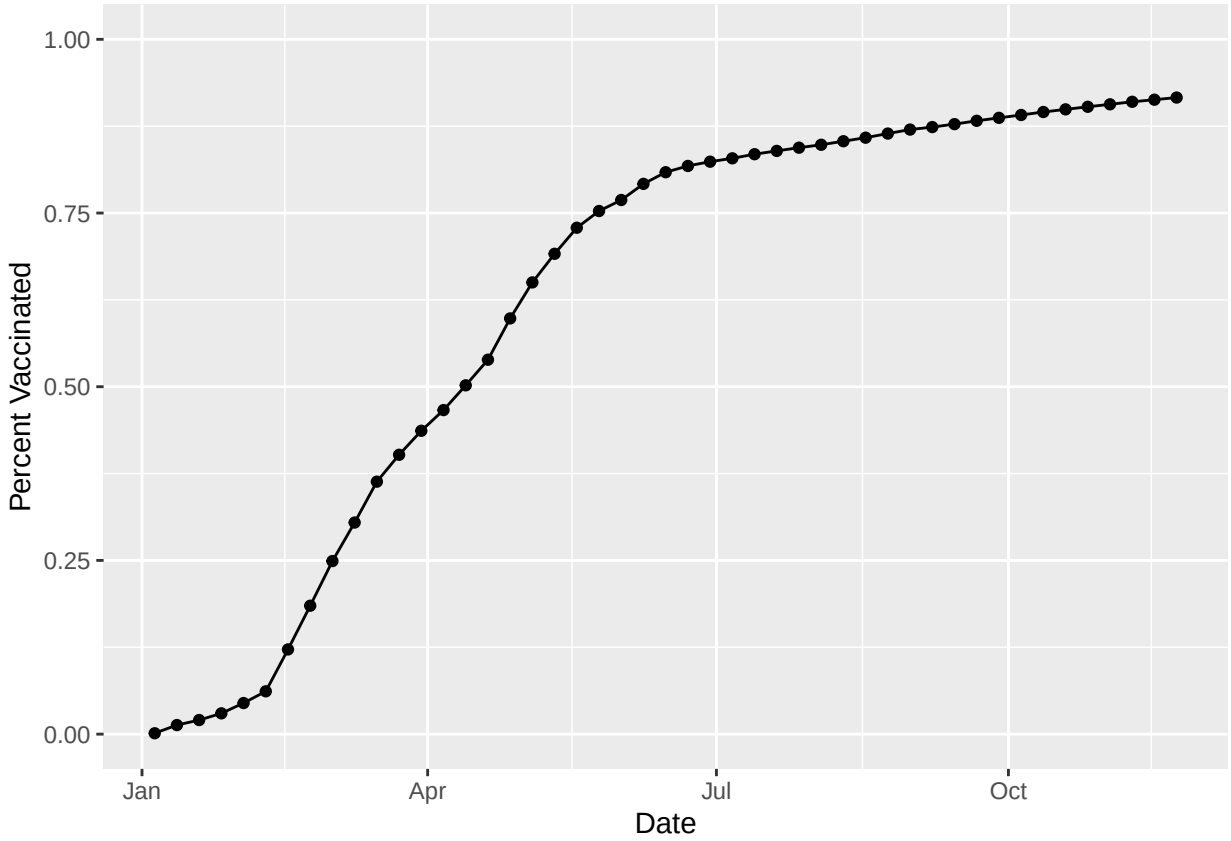
### Focus on UCSD/La Jolla

### Plotting area of interest, begining with La Jolla (92037).

```
sd.92037 <- filter(vax, zip_code_tabulation_area == "92037")
```

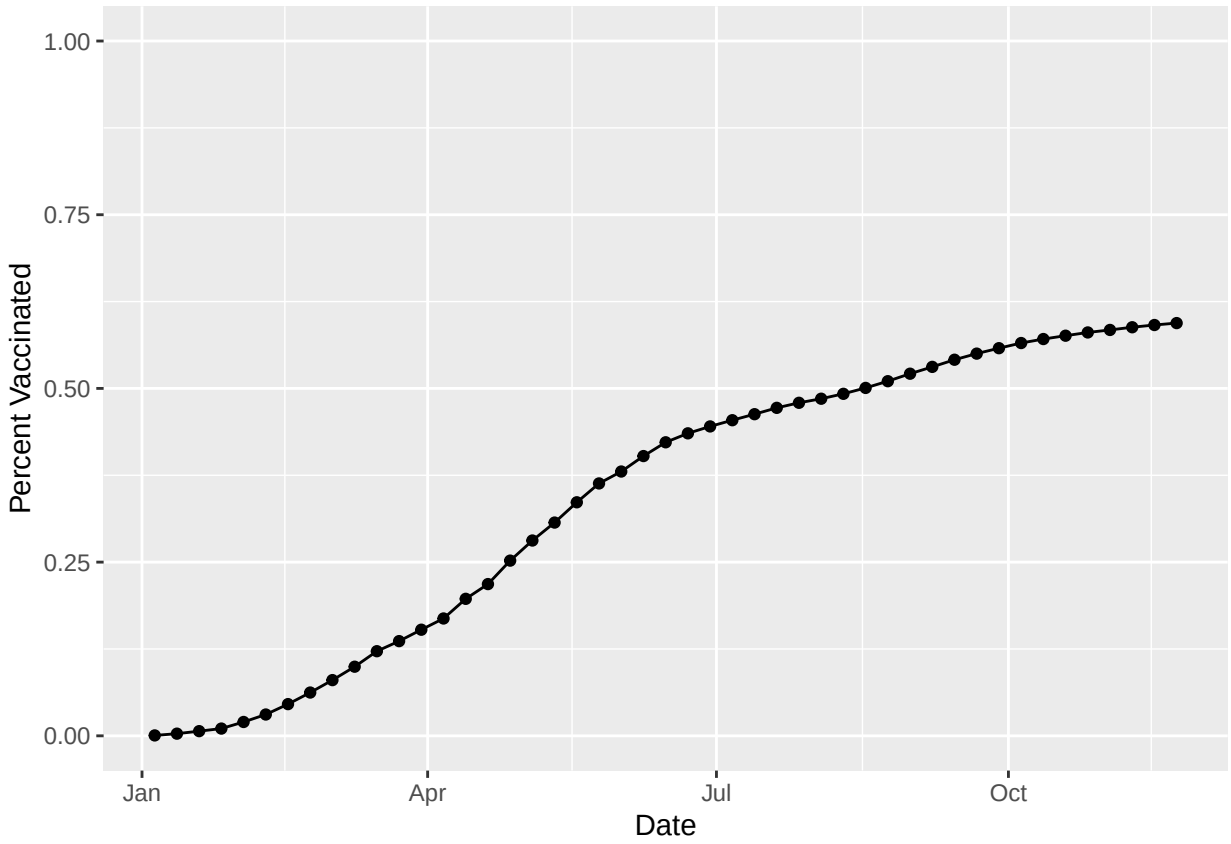
```
library(ggplot2)
ggplot(sd.92037) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated) +
```

```
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x= "Date", y="Percent Vaccinated")
```




**For Escondido, CA 92025.**

```
library(ggplot2)
ggplot(sd.92025) +
  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")
```




**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?**

**Lets make this plot for all San Diego County ZIP code areas that have a population as least as large as 92037.**

```
head(vax$age5_plus_population)
```

```
## [1] 40888  1521 31347 16905  2526  2397
```

```
sd.36 <- filter(vax, county == "San Diego",
                age5_plus_population > 36144)
head(sd.36)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-01-05                   92058                San Diego San Diego
## 2 2021-01-05                   92078                San Diego San Diego
## 3 2021-01-05                   92019                San Diego San Diego
## 4 2021-01-05                   92117                San Diego San Diego
## 5 2021-01-05                   92057                San Diego San Diego
## 6 2021-01-05                   91913                San Diego San Diego
##   vaccine_equity_metric_quartile               vem_source
## 1                              1 Healthy Places Index Score
## 2                              3 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
```

```
## 4                              3 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               34956.0                39695                       NA
## 2               41789.5                47476                       37
## 3               37439.4                40464                       25
## 4               50041.6                53839                       42
## 5               51927.0                56906                       22
## 6               43514.7                50461                       37
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                          688                                0.000779
## 3                          610                                0.000618
## 4                         1143                                0.000780
## 5                          691                                0.000387
## 6                         1993                                0.000733
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                   0.014492
## 3                                   0.015075
## 4                                   0.021230
## 5                                   0.012143
## 6                                   0.039496
##   percent_of_population_with_1_plus_dose
## 1                                     NA
## 2                               0.015271
## 3                               0.015693
## 4                               0.022010
## 5                               0.012530
## 6                               0.040229
##                                                                  redacted
## 1 Information redacted in accordance with CA state privacy requirements
## 2                                                                      No
## 3                                                                      No
## 4                                                                      No
## 5                                                                      No
## 6                                                                      No
```

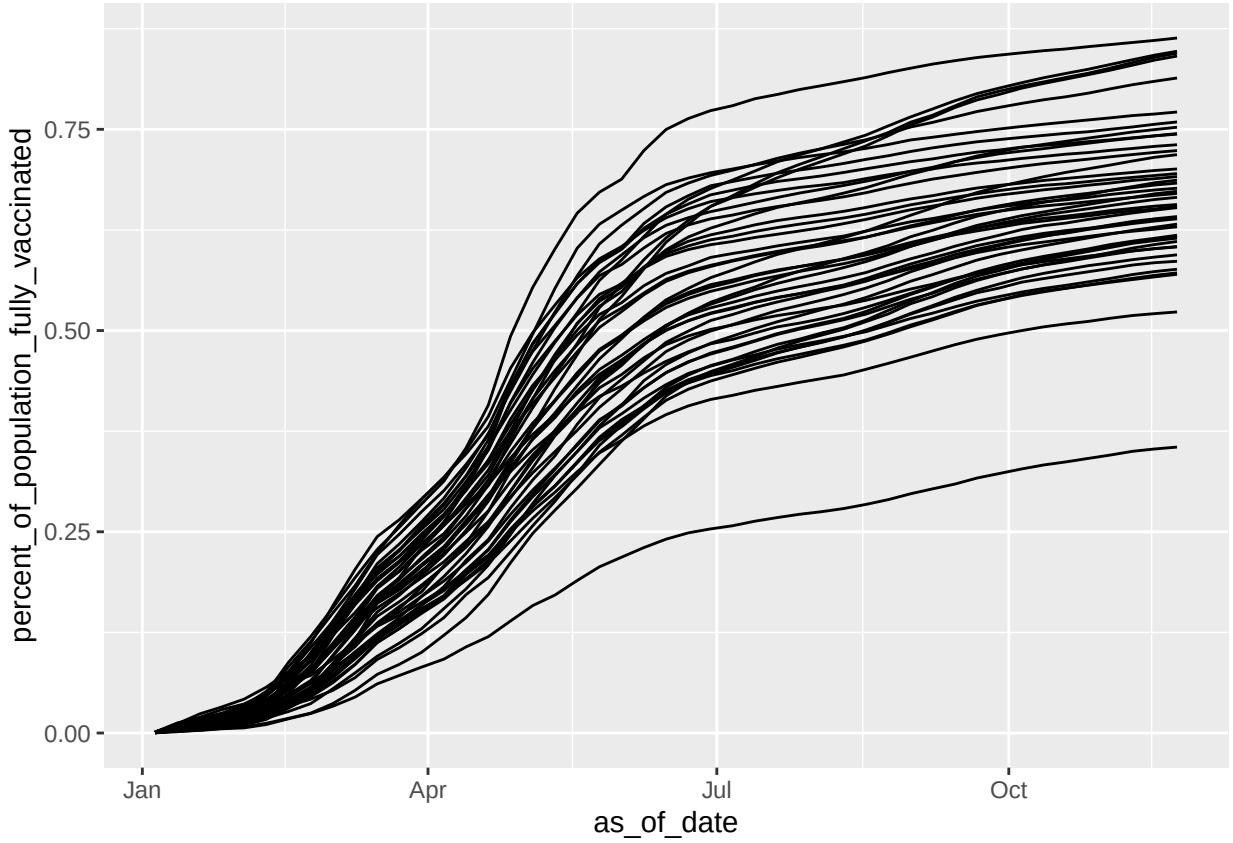
### How many ZIP code areas in San Diego County have a population larger than 92037.

```
length(unique(sd.36$zip_code_tabulation_area))
```

```
## [1] 43
```

```
library(ggplot2)
ggplot(sd.36) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated,
      group = zip_code_tabulation_area) +
  geom_line()
```

```
## Warning: Removed 1 row(s) containing missing values (geom_path).
```




```
  labs(x= "Date", y="Percent Vaccinated")
```

```
## $x
## [1] "Date"
##
## $y
## [1] "Percent Vaccinated"
##
## attr(,"class")
## [1] "labels"
```

### Making a plot for all California with populations as large as La Jolla.

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

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction         county
## 1 2021-01-05                   92395            San Bernardino San Bernardino
## 2 2021-01-05                   92410            San Bernardino San Bernardino
## 3 2021-01-05                   92646                    Orange         Orange
## 4 2021-01-05                   92886                    Orange         Orange
## 5 2021-01-05                   92545                 Riverside      Riverside
## 6 2021-01-05                   92677                    Orange         Orange
##   vaccine_equity_metric_quartile                vem_source
## 1                              1 Healthy Places Index Score
## 2                              1 Healthy Places Index Score
```

```
## 3                              4 Healthy Places Index Score
## 4                              4 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               35012.3                41625                       NA
## 3               49327.5                53307                       18
## 4               43348.1                48075                       34
## 5               35528.1                39692                       NA
## 6               58070.9                63004                       19
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                      NA
## 2                           NA                                      NA
## 3                         1083                                0.000338
## 4                         1057                                0.000707
## 5                           NA                                      NA
## 6                         1059                                0.000302
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                         NA
## 3                                   0.020316
## 4                                   0.021986
## 5                                         NA
## 6                                   0.016808
##   percent_of_population_with_1_plus_dose
## 1                                     NA
## 2                                     NA
## 3                               0.020654
## 4                               0.022693
## 5                                     NA
## 6                               0.017110
##                                                                  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                                                                      No
```

### How many ZIP codes in CA are as large as La Jolla.

```
length(unique(ca.all$zip_code_tabulation_area))
```

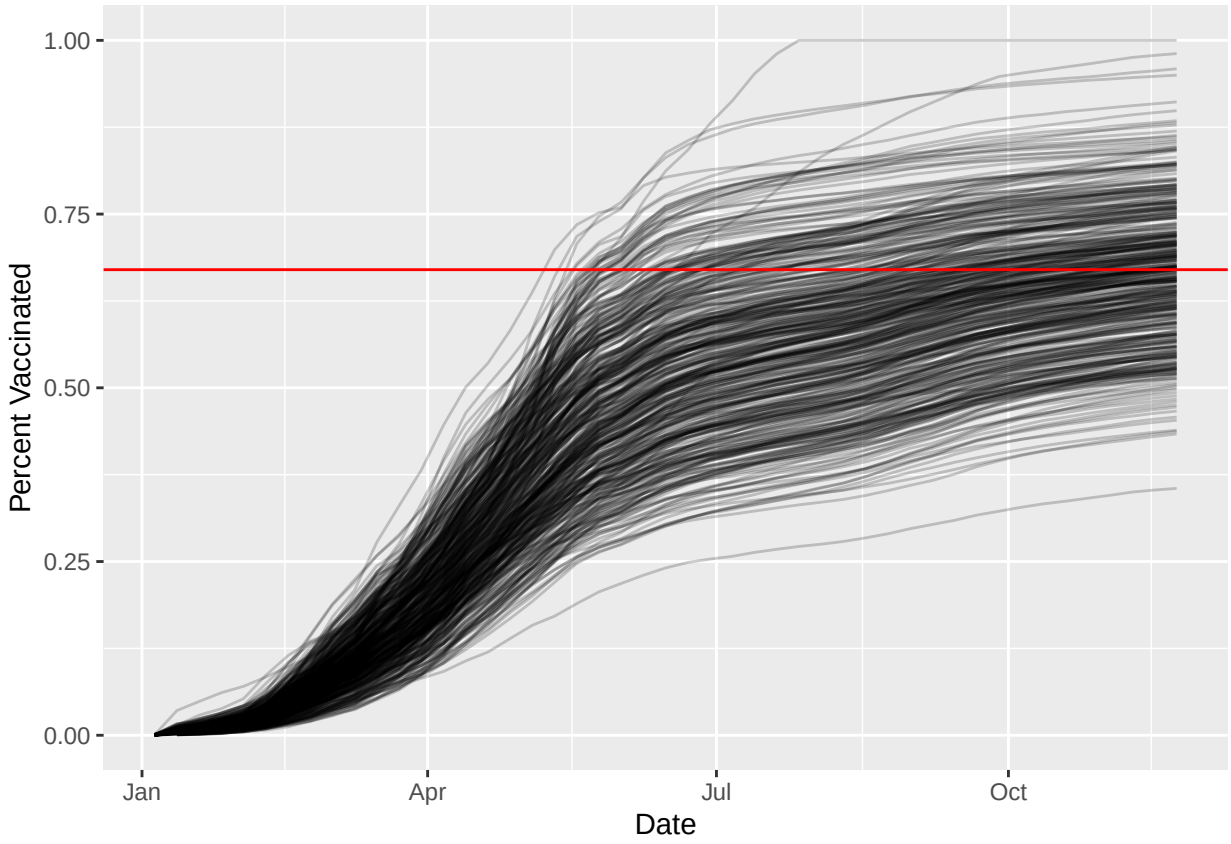
```
## [1] 411
```

### Make the plot for all the ZIP codes as large as La Jolla.

```
library(ggplot2)
ggplot(ca.all) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated,
      group = zip_code_tabulation_area) +
```

```
  geom_line(alpha = 0.2) +
  labs(x= "Date", y="Percent Vaccinated") +
  geom_hline(yintercept=0.67, color = "red")
```

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




**What is the mean across the state for these 36k + population areas.**

```
ca.now <- filter(ca.all,
                 as_of_date == "2021-11-23")
summary(ca.now$percent_of_population_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3552  0.5939  0.6696  0.6672  0.7338  1.0000
```