Class17_VaccinationMiniProject

Caitriona Brennan

#Background

In this Thanksgiving class when many of our classmates are traveling let's have a look at COVID-19 vaccination rates around the States.

We get vaccination rate data from CA.GOV.

#Import Data

```r
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 entries do we have?

```
nrow(vax)
```

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

We can use the **skimr** package and the ‘skim()’ function to get a quick overview of structure of this dataset. IF we only want to use Skimr once we can call it like this rather than library(skimr)

```
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	4089	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	

#Notice that one of these is a date column. Working with time and dates get's annoying quickly.

```
library(lubridate)
```

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

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

```
today()
```

```
## [1] "2021-11-27"
```

> Q. How many days since the first entry in the dataset?

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

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

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

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

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

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

```
## Time difference of 326 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? What is the last date in this dataset?How many days since the last update?

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

> Q. How many days does this dataset span?

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

> Q. How many different ZIP codes are recorded in this dataset?

```
zipcodes <- unique(vax[,2])
length(zipcodes)
```

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

```
library(zipcodeR)
```

##Focus in on San Diego County

We want to subset the full CA vax data down to just San Diego.

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

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

Subsetting 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 in on san diego county

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

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

To do more complicated subsetting...

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

nrow(sd.20)
```

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

```r
head(sd.20)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-01-05                    92011                 San Diego San Diego
## 2 2021-01-05                    92081                 San Diego San Diego
## 3 2021-01-05                    92124                 San Diego San Diego
## 4 2021-01-05                    92058                 San Diego San Diego
## 5 2021-01-05                    92078                 San Diego San Diego
## 6 2021-01-05                    92123                 San Diego San Diego
##   vaccine_equity_metric_quartile                 vem_source
## 1                              4 Healthy Places Index Score
## 2                              2 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              1 Healthy Places Index Score
## 5                              3 Healthy Places Index Score
## 6                              3 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               20503.6                23247                       NA
## 2               25558.0                27632                       14
## 3               25422.4                29040                       29
## 4               34956.0                39695                       NA
## 5               41789.5                47476                       37
## 6               28353.3                30426                       48
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                          346                               0.000507
## 3                          575                               0.000999
## 4                           NA                                     NA
## 5                          688                               0.000779
## 6                          994                               0.001578
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                   0.012522
## 3                                   0.019800
## 4                                         NA
## 5                                   0.014492
## 6                                   0.032669
##   percent_of_population_with_1_plus_dose
## 1                                      NA
## 2                                0.013029
## 3                                0.020799
## 4                                      NA
## 5                                0.015271
## 6                                0.034247
##                                                                 redacted
## 1 Information redacted in accordance with CA state privacy requirements
## 2                                                                     No
## 3                                                                     No
## 4 Information redacted in accordance with CA state privacy requirements
## 5                                                                     No
## 6                                                                     No
```

```
length(unique(sd.20[,2]))
```

```
## [1] 65
```

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

nrow(sd.now)
```

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

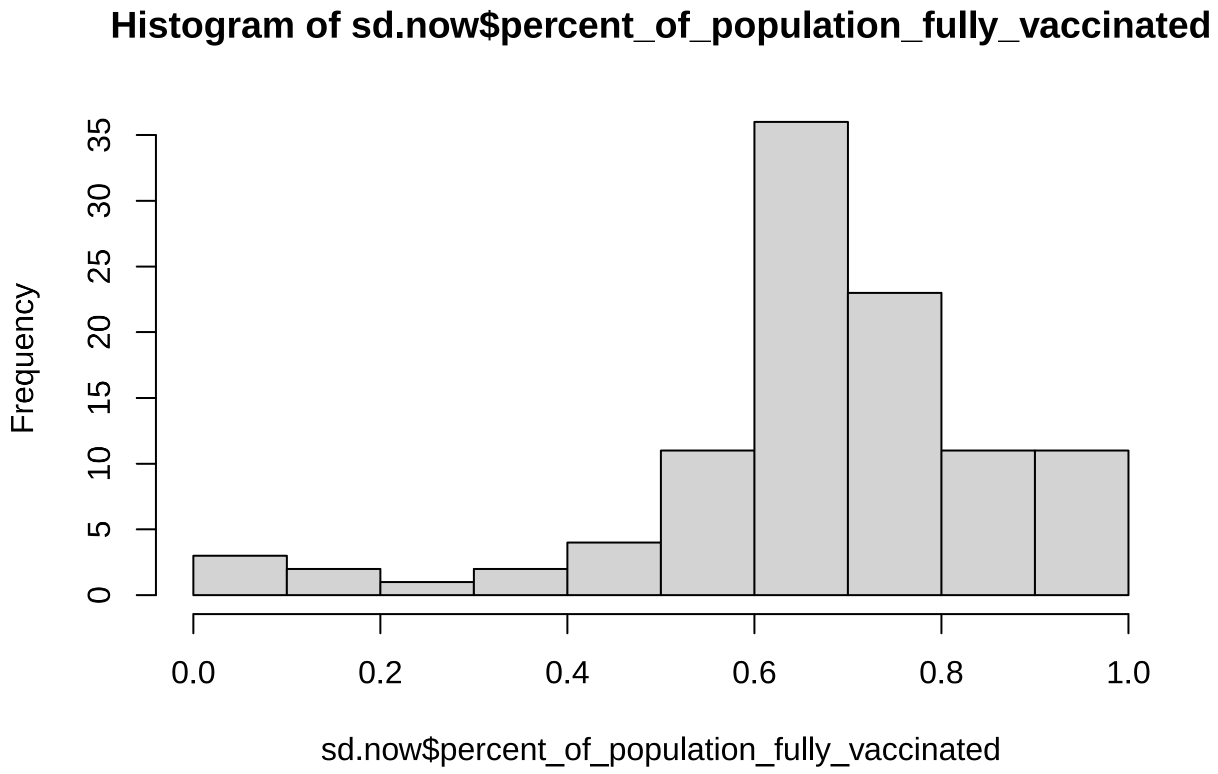
```
summary(sd.now$percent_of_population_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
## 0.01017 0.61301 0.67965 0.67400 0.76932 1.00000       3
```

Q. Make a histogram of these values:

R based histogram

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



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

> Q. what is the population of the 92037 ZIP code area?

> Q. what is the average vaccination value for this UCSD/La Jolla ZIP code area?

```
lj <- filter(sd.now, sd.now$zip_code_tabulation_area==92037)
lj$age5_plus_population
```

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

```
lj$percent_of_population_fully_vaccinated
```

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

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

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

```
Hillcrest
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-11-23                    92103                 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               32146.4                33213                    44547
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         5214                                      1
##   percent_of_population_partially_vaccinated
## 1                                   0.156987
##   percent_of_population_with_1_plus_dose redacted
## 1                                      1       No
```

```
Hillcrest$percent_of_population_fully_vaccinated
```

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

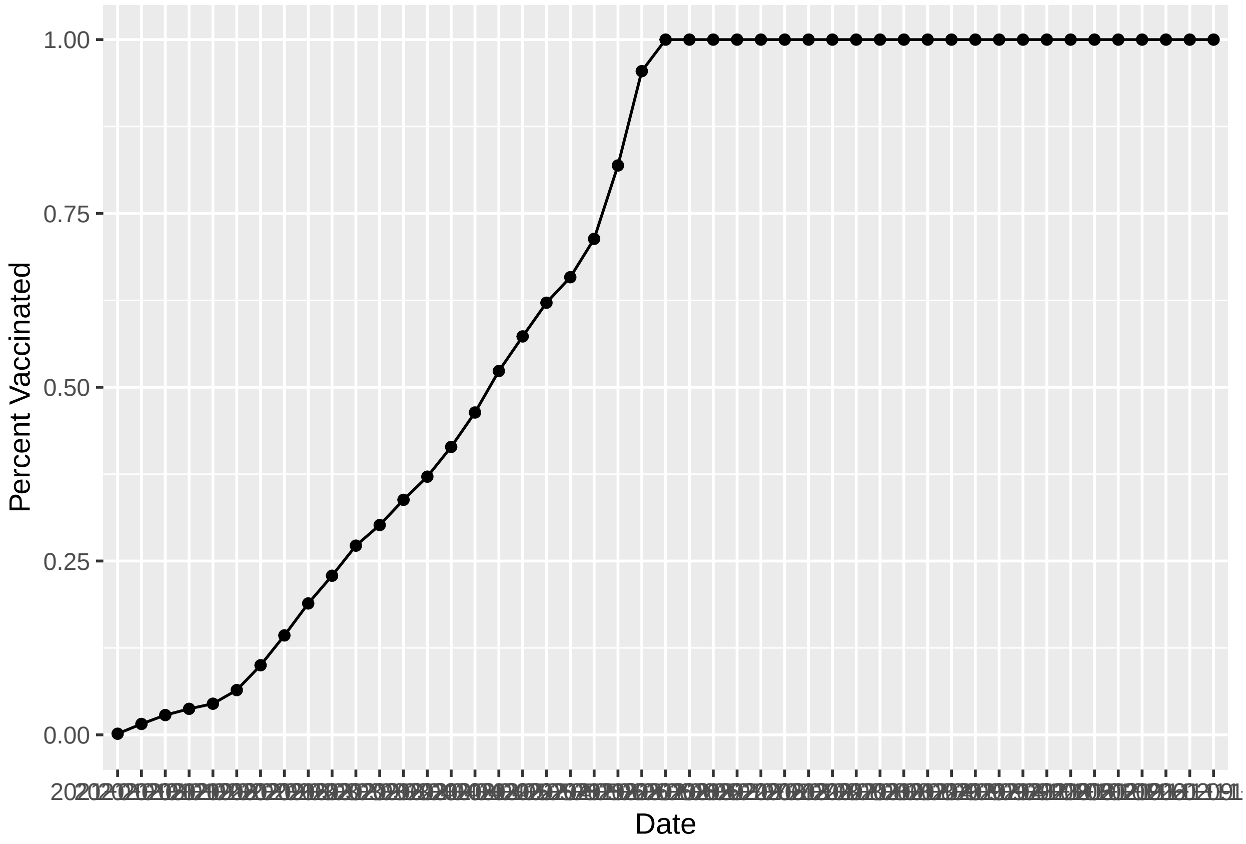
Time series of vaccination rate for a given ZIP code area. Start with 92037

```
hillcrest <- filter(vax, vax$zip_code_tabulation_area==92103)
head(hillcrest)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-01-05                    92103                 San Diego San Diego
## 2 2021-01-12                    92103                 San Diego San Diego
## 3 2021-01-19                    92103                 San Diego San Diego
## 4 2021-01-26                    92103                 San Diego San Diego
## 5 2021-02-02                    92103                 San Diego San Diego
## 6 2021-02-09                    92103                 San Diego San Diego
##   vaccine_equity_metric_quartile                 vem_source
## 1                              4 Healthy Places Index Score
## 2                              4 Healthy Places Index Score
## 3                              4 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                              4 Healthy Places Index Score
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               32146.4                33213                       53
## 2               32146.4                33213                      520
## 3               32146.4                33213                      944
## 4               32146.4                33213                     1242
## 5               32146.4                33213                     1487
## 6               32146.4                33213                     2137
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         1383                               0.001596
## 2                         1361                               0.015657
## 3                         2434                               0.028423
## 4                         3629                               0.037395
## 5                         5438                               0.044772
## 6                         6151                               0.064342
##   percent_of_population_partially_vaccinated
## 1                                   0.041640
## 2                                   0.040978
## 3                                   0.073285
## 4                                   0.109264
## 5                                   0.163731
## 6                                   0.185199
##   percent_of_population_with_1_plus_dose redacted
## 1                               0.043236       No
## 2                               0.056635       No
## 3                               0.101708       No
## 4                               0.146659       No
## 5                               0.208503       No
## 6                               0.249541       No
```

```
library(ggplot2)
```

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



Let's make this plot for all San Diego county ZIP code areas that have a population as least as large as 92037.

```
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?

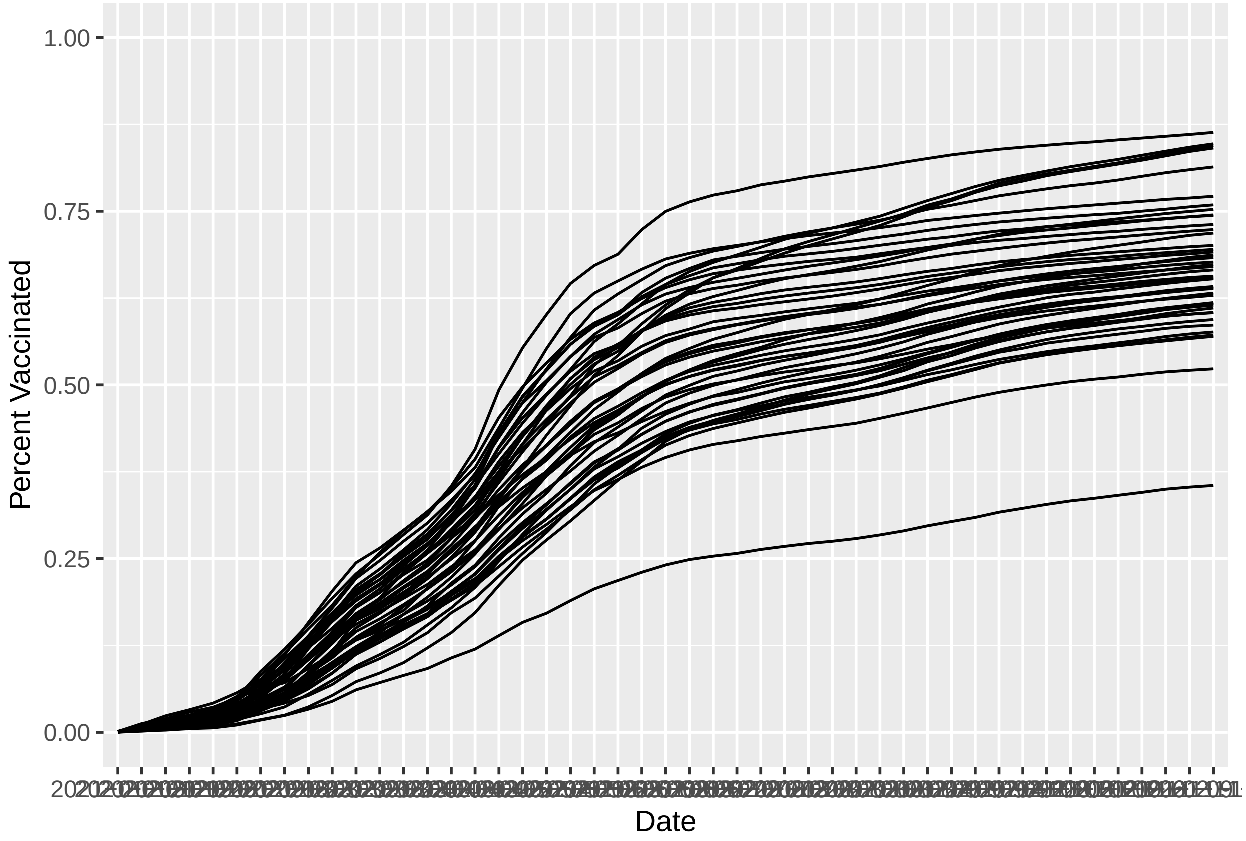
```r
length(unique(sd.36$zip_code_tabulation_area))
```

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

Let's make the plot

```r
ggplot(sd.36) +
  aes(x=as_of_date,
      y=percent_of_population_fully_vaccinated,
      group=zip_code_tabulation_area) +
  geom_line() +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
```

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



Q. Make a plot like this for all ZIP codes in CA that have a population at least as large as La Jolla (>31644)

```
ca.36 <- filter(vax,
                age5_plus_population > 36144)
head(ca.36)
```

```
##   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
```

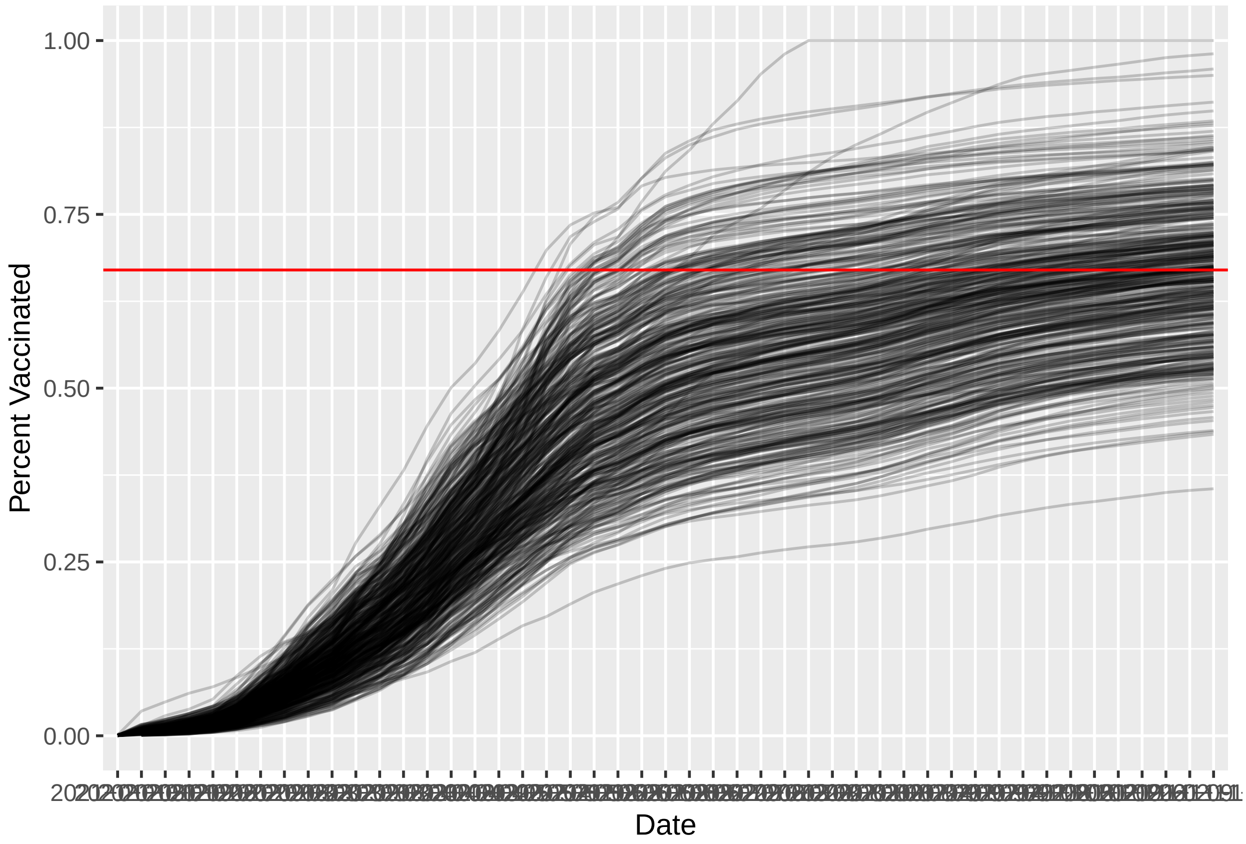
How many zipcode areas?

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

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

```
ggplot(ca.36) +
  aes(x=as_of_date,
      y=percent_of_population_fully_vaccinated,
      group=zip_code_tabulation_area) +
  geom_line(alpha=0.2) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated") +
  geom_hline(yintercept = 0.67, color="red")
```

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



Q. What is the mean accross the state for these 36K + population areas?

```
ca.now <- filter(ca.36, 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
```