

# vaccine

May Wu PID:A59010588

11/24/2021

## Mini-Project

### COVID-19 Vaccination Rates

```
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 the data

```
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	93665.1118	17.39	90001	92257.7593	658.5095	380.5097	635.0	
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	18895.0418	993.94	0	1346.95	13685.1031	756.1288	556.7	
age5_plus_population	0	1.00	20875.2421	1106.04	0	1460.50	15364.0034	877.0010	1902.0	
persons_fully_vaccinated	8355	0.90	9585.35	11609.12	11	516.00	4210.00	16095.0071	219.0	
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_8355plus_doses	8355	0.90	0.51	0.26	0	0.31	0.53	0.71	1.0	

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 `install.packages("lubridate")`

```
library(lubridate)
```

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

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

```
today()
```

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

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 characters

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

```
d = ymd(vax$as_of_date)
```

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

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

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?

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

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

Q. how many days since the last update?

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

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

Q. how many different zip code

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

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

To work with zipcode, we can use the **zipcodeR**

```
library(zipcodeR)
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 San Diego county

subset the full CA vax data down to just SD County using `filter()`

```
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
```

More complicated subsetting

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

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

Q. What is the average vaccination rate of SD as of yesterday?

```
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
```

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

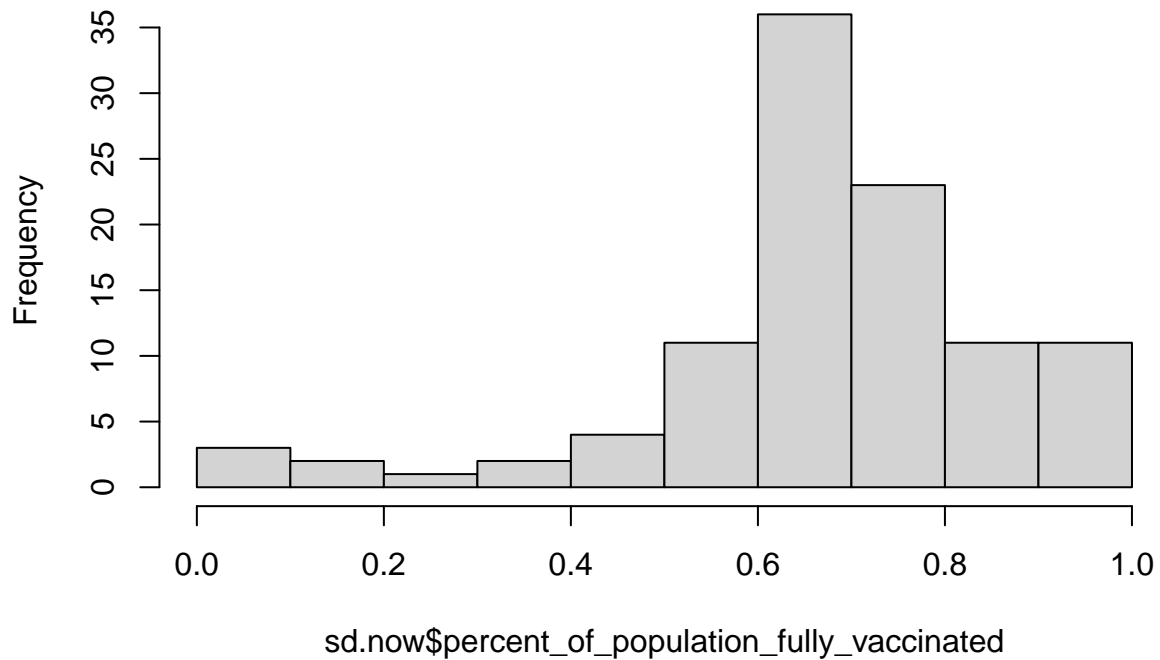
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.01017 0.61301 0.67965 0.67400 0.76932 1.00000      3
```

Q. make a histogram of these value

base R histogram

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

## Histogram of sd.now\$percent\_of\_population\_fully\_vaccinated



Q. What's the average vaccination value for 92037 area?

```
d = filter(sd.now, zip_code_tabulation_area == '92037' )  
d$percent_of_population_fully_vaccinated
```

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

Q. what's the population of the 92037 ZIP code area?

```
d$persons_fully_vaccinated
```

```
## [1] 33115
```

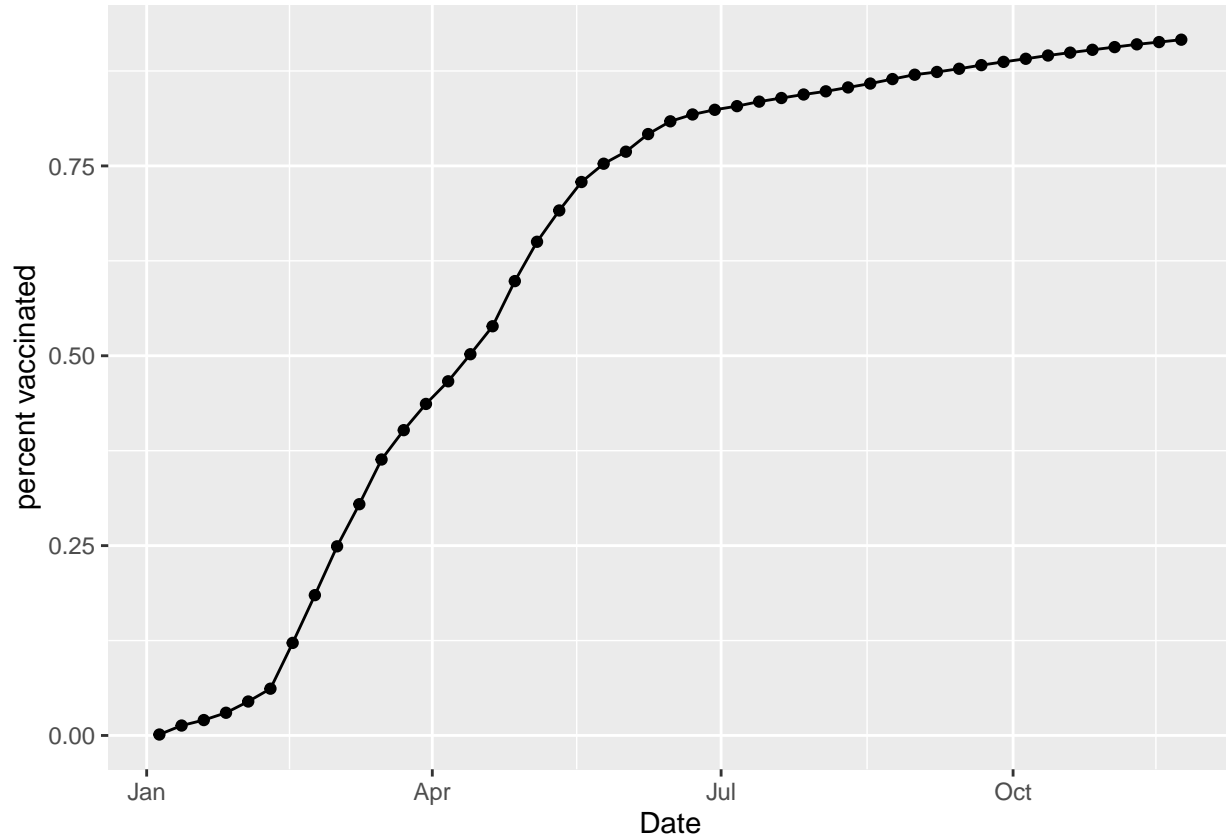
```
d = filter(sd.now, zip_code_tabulation_area == '92124' )  
d$percent_of_population_fully_vaccinated
```

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

time series of vaccination rate for a given ZIP code area.

starting with 92037

```
lj = filter(vax, zip_code_tabulation_area == '92037')
library(ggplot2)
#vax$percent_of_population_fully_vaccinated
ggplot(lj, aes(x=as_of_date, y=percent_of_population_fully_vaccinated)) + geom_line(group = 1) + geom_p
```



Let's make this plot for the whole SD county

```
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
```

Q. how many ZIP code areas in SD county have a population larger than 92037

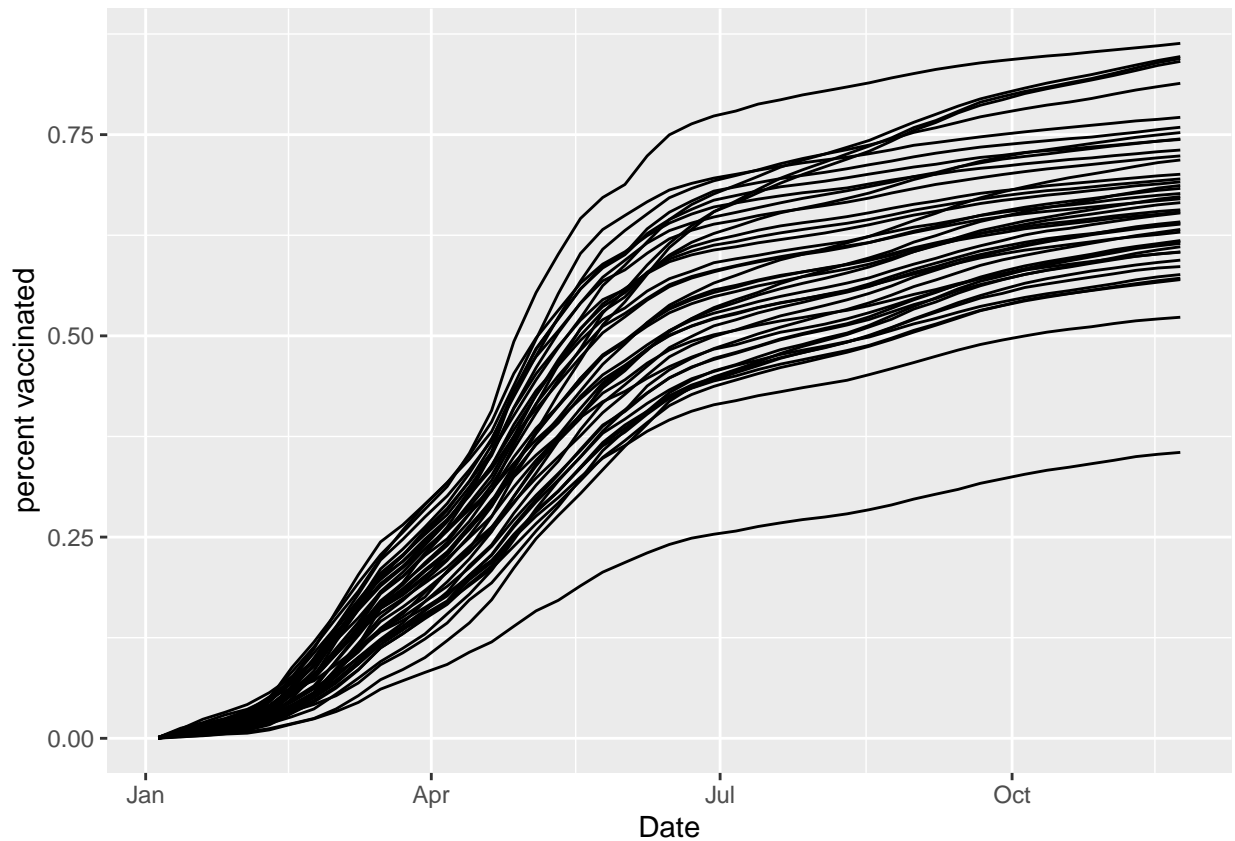
```
length(unique(sd.36$zip_code_tabulation_area))
```

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

lets make the plot

```
ggplot(sd.36, aes(x=as_of_date, y=percent_of_population_fully_vaccinated, group = zip_code_tabulation_a
  geom_line() +
  labs(x='Date', y = 'percent vaccinated')
```

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



Q. Make a plot like this for the all ZIP code areas in the State with a population at least as large as La Jolla

```
ca = filter(vax, age5_plus_population > 36144)
head(ca)
```

```
##   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
```

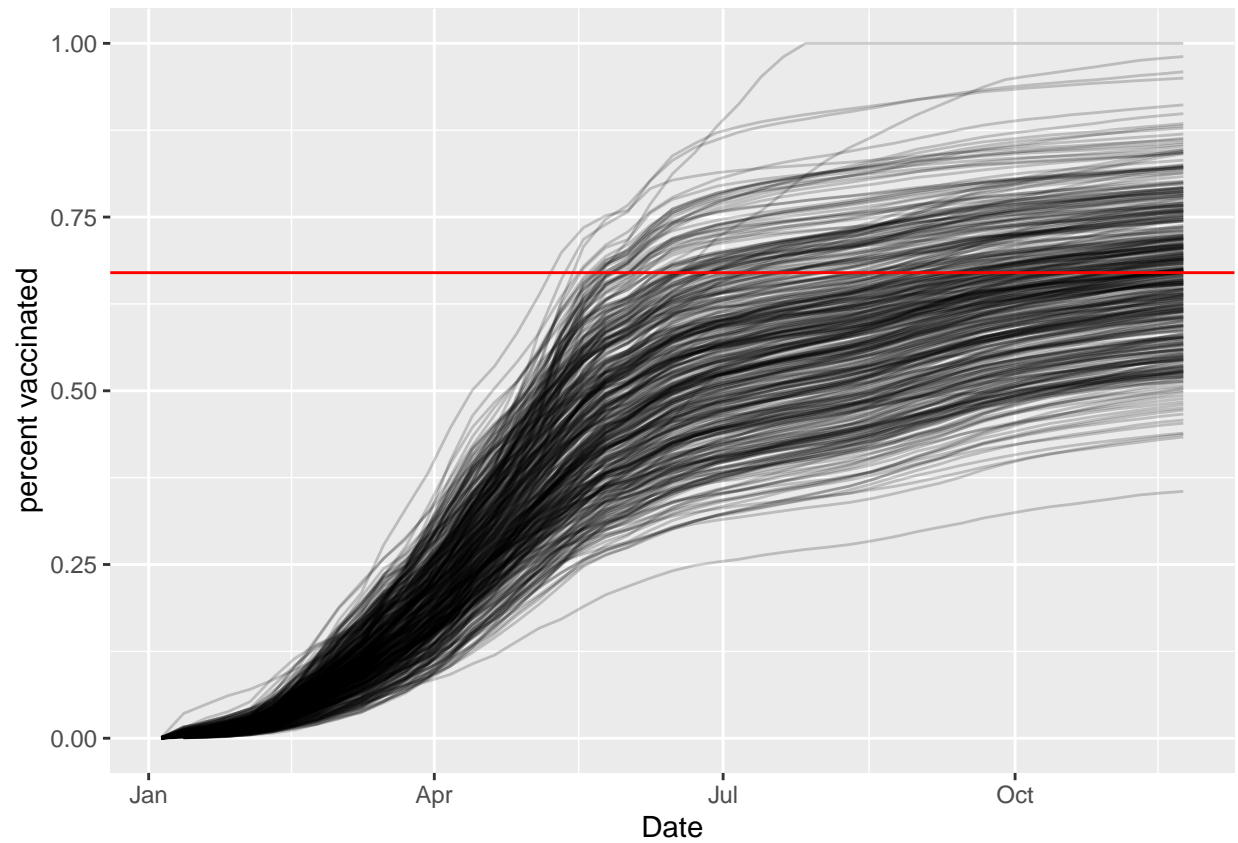
```
## 5          35528.1          39692          NA
## 6          58070.9          63004          19
##  persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1              NA              NA
## 2              NA              NA
## 3             1083             0.000338
## 4             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
```

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

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

```
ggplot(ca, 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).
```



Q. What is the mean across the state for these 36k+ population area?

```
ca.now = filter(ca, 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
```