

Vaccination Rate Mini Project

Jack Reddan (PID: A59010543)

Getting Started

```
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
```

```
library(lubridate)
```

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

```
#library(zipcodeR)
```

```
library(ggplot2)
```

```
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
```

Q01: What column details the total number of people fully vaccinated?

persons_fully_vaccinated

Q02: What column details the Zip code tabulation area?

zip_code_tabulation_area

Q03: What is the earliest date in this dataset?

```
vax %>%
  arrange(as_of_date) %>%
  head(1)[1]
```

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

Q04: What is the latest date in this dataset?

```
vax %>%
  arrange(desc(as_of_date)) %>%
  head(1)[1]
```

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

```
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_quarter	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	1756.12	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_8355plus_dose	8355	0.90	0.51	0.26	0	0.31	0.53	0.71	1.0	

Q05: How many numeric columns are in this dataset?

9

Q06: Note that there are “missing values” in the dataset. How many NA values there in the `persons_fully_vaccinated` column?

8355

Q07: What percent of `persons_fully_vaccinated` values are missing (to 2 significant figures)?

```
round((1 - 0.899)*100, 2)
```

```
## [1] 10.1
```

Q08: Why might this data be missing?

Some area codes, such as military bases, may have certain restrictions for data release and transparency, resulting in a lack of data in the report.

Working with Dates

```
today()

## [1] "2021-11-29"
vax$as_of_date <- ymd(vax$as_of_date)

today() - vax$as_of_date[1]

## Time difference of 328 days
vax$as_of_date[nrow(vax)] - vax$as_of_date[1]

## Time difference of 322 days
Q09: How many days have passed since the last update of the dataset?
last_update <- vax %>%
  arrange(desc(as_of_date)) %>%
  head(1)[1]

today() - last_update

## Time difference of 6 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] 47
```

Working with ZIP Codes

Focus on the San Diego Area

```
sd <- vax[vax$county == "San Diego",]

With dplyr:
sd <- vax %>%
  filter(county == "San Diego")

sd.10 <- vax %>%
  filter(county == "San Diego") %>%
  filter(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 %>%
  arrange(desc(age12_plus_population)) %>%
  head(1)[2]
```

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

Q13: What is the overall average “Percent of Population Fully Vaccinated” value for all San Diego “County” as of “2021-11-09”?

```
sd.yest <- sd %>%
  filter(as_of_date == "2021-11-23") %>%
  filter(!is.na(percent_of_population_fully_vaccinated))

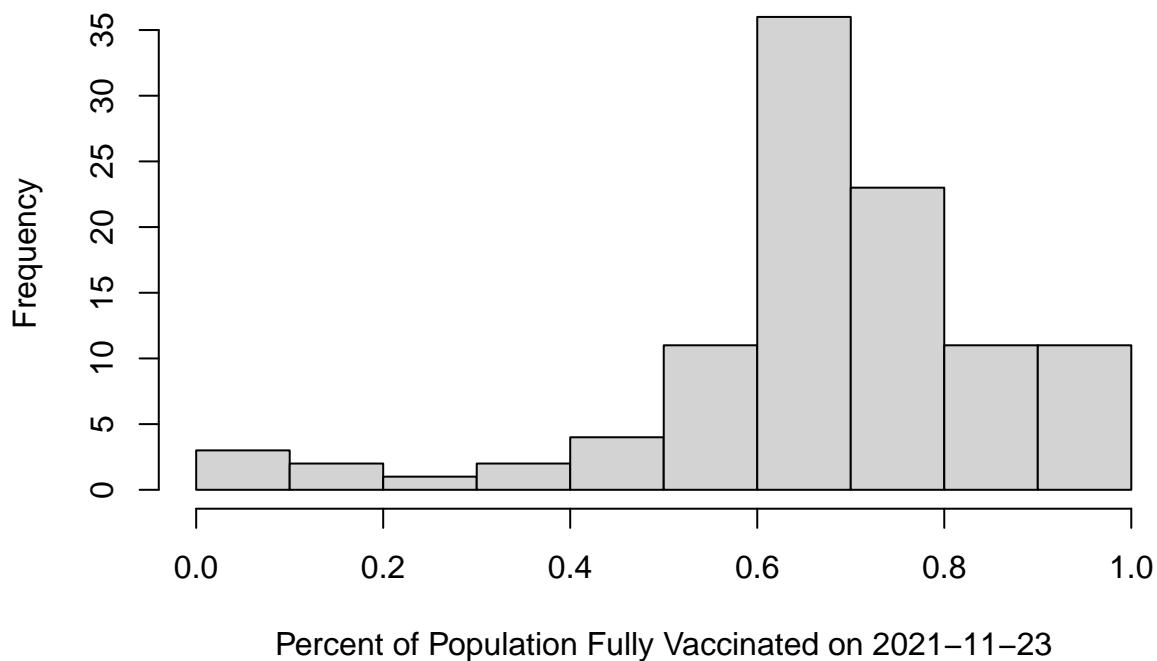
paste(round(mean(sd.yest$percent_of_population_fully_vaccinated)*100, 2), "%", sep = "")
```

```
## [1] "67.4%"
```

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(sd.yest$percent_of_population_fully_vaccinated,
     xlab = "Percent of Population Fully Vaccinated on 2021-11-23",
     ylab = "Frequency",
     main = "Histogram of Vaccination Rates Across San Diego County")
```

Histogram of Vaccination Rates Across San Diego County



Focus on UCSD/La Jolla

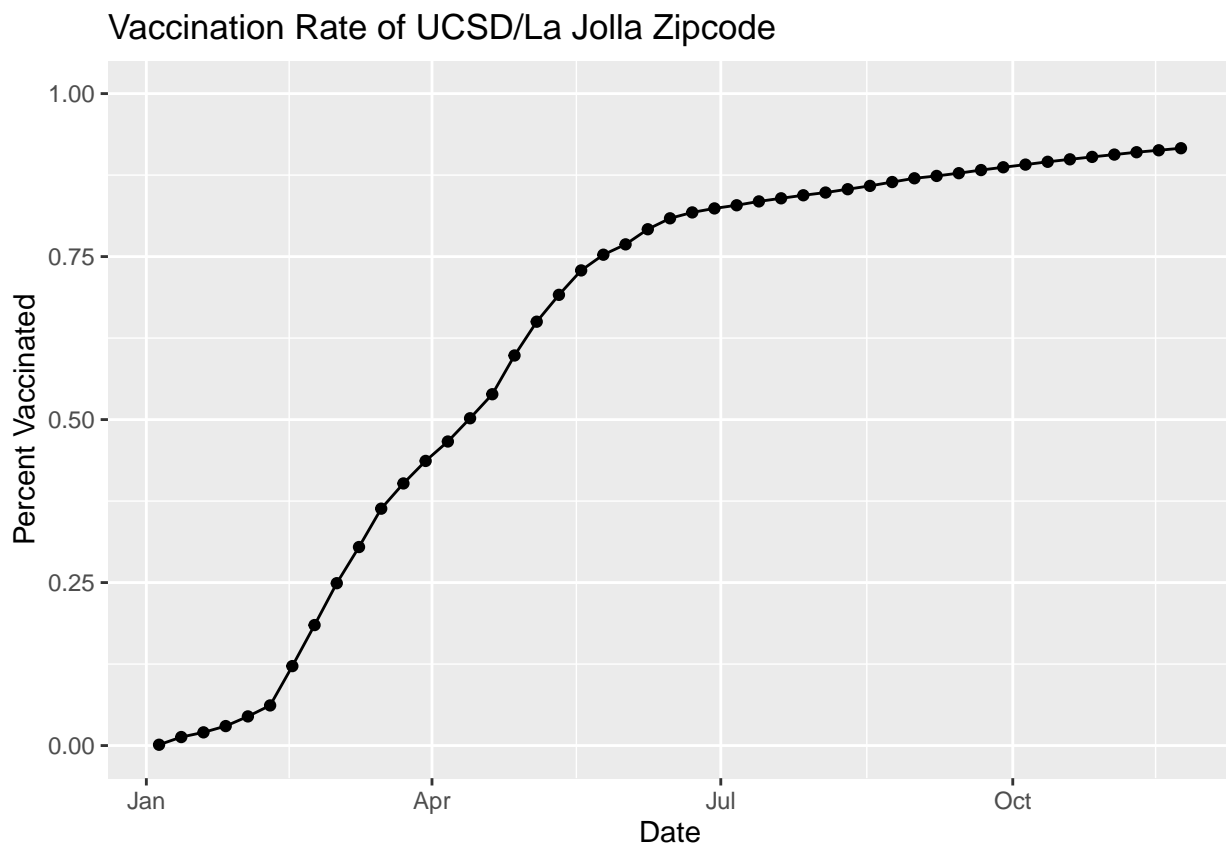
```
ucsd <- sd %>%
  filter(zip_code_tabulation_area == "92037")
```

```
ucsd$age5_plus_population[1]
```

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

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

```
ggplot(data = ucsd) +  
  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",  
       title = "Vaccination Rate of UCSD/La Jolla Zipcode")
```



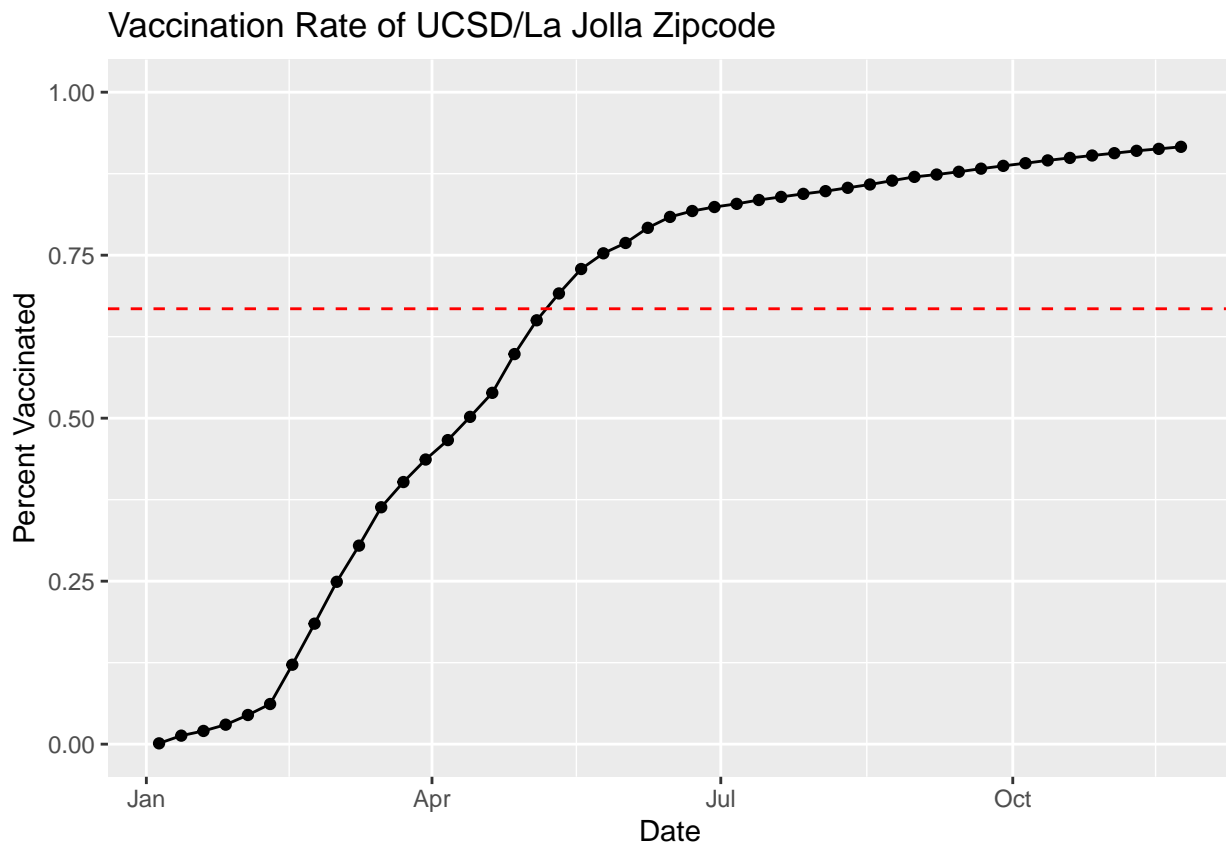
Compare to Similar Sized Areas

```
vax.lj_pop <- vax %>%  
  filter(age5_plus_population >= ucsd$age5_plus_population) %>%  
  filter(as_of_date == "2021-11-23")
```

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 “2021-11-23”. Add this as a straight horizontal line to your plot from above with the `geom_hline()` function?

The mean is 66.78%.

```
ggplot(data = ucsd) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group = 1) +
  geom_hline(yintercept = mean(vax.lj_pop$percent_of_population_fully_vaccinated),
            linetype = 2,
            col = "red") +
  ylim(c(0,1)) +
  labs(x = "Date",
       y = "Percent Vaccinated",
       title = "Vaccination Rate of UCSD/La Jolla Zipcode")
```



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 “2021-11-23”?

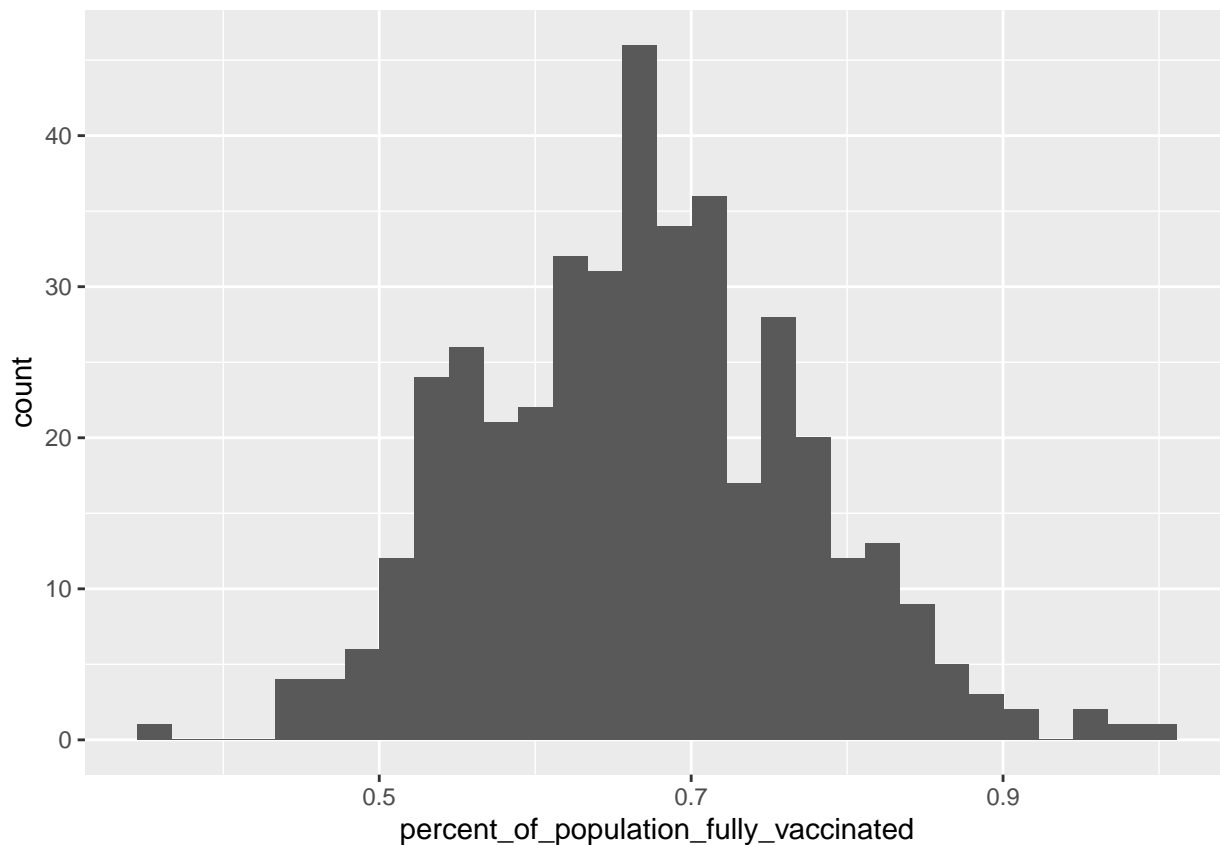
```
summary(vax.lj_pop$percent_of_population_fully_vaccinated)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3552  0.5939  0.6698  0.6678  0.7350  1.0000
```

Q18: Using ggplot generate a histogram of this data.

```
ggplot(data = vax.lj_pop) +
  aes(x = percent_of_population_fully_vaccinated) +
  geom_histogram()
```

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



Q20: Is the 92109 and 92040 ZIP code areas above or below the average value you calculated for all these above?

```
avg_vax_rate <- mean(vax.lj_pop$percent_of_population_fully_vaccinated)
```

```
zc_92109 <- vax %>%
  filter(as_of_date == "2021-11-23") %>%
  filter(zip_code_tabulation_area == 92109)
```

```
zc_92040 <- vax %>%
  filter(as_of_date == "2021-11-23") %>%
  filter(zip_code_tabulation_area == 92040)
```

```
zc_92109$percent_of_population_fully_vaccinated > avg_vax_rate
```

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

```
zc_92040$percent_of_population_fully_vaccinated > avg_vax_rate
```

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

```
sd.lj_pop <- sd %>%
  filter(age5_plus_population >= ucsd$age5_plus_population)
```

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

```
## [1] 44
```

```
ggplot(data = sd.lj_pop) +
  aes(x = as_of_date,
```

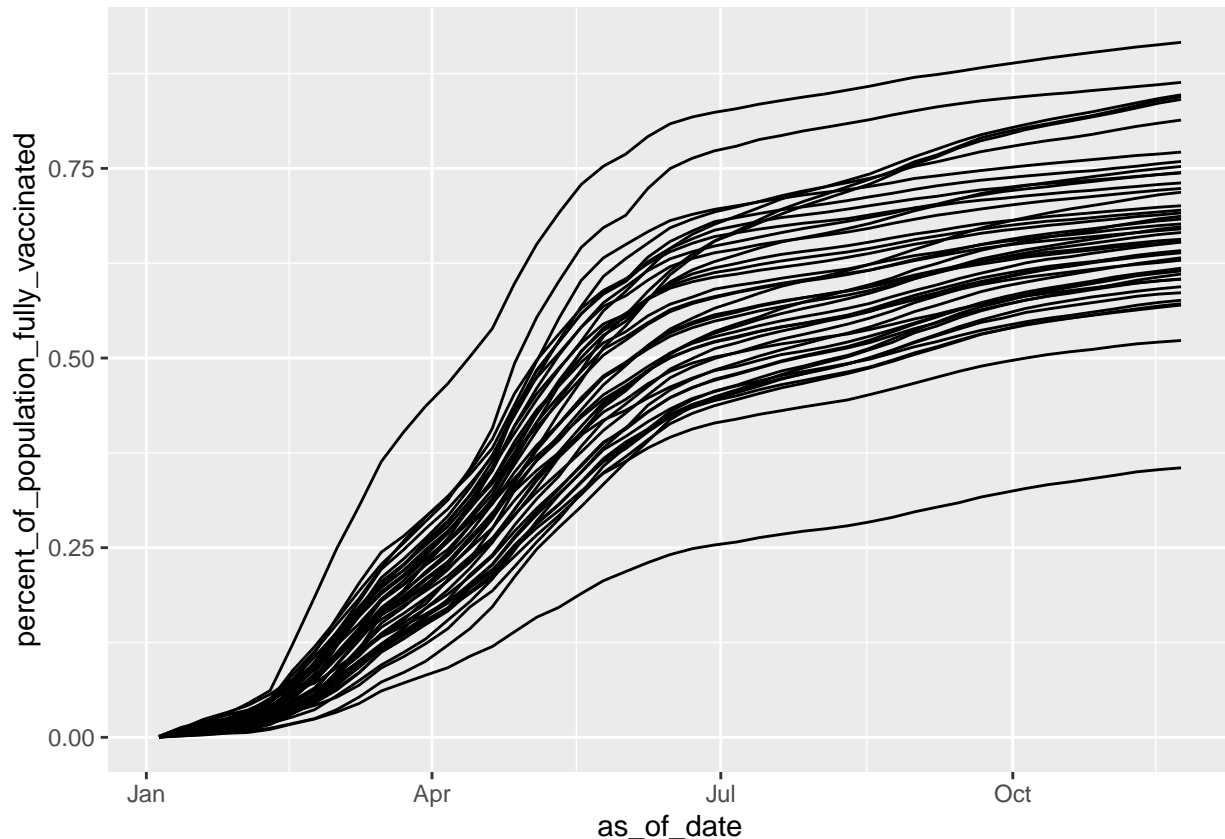


```

y = percent_of_population_fully_vaccinated,
group = zip_code_tabulation_area) +
geom_line()

```

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



```

vax.lj_pop_all <- vax %>%
  filter(age5_plus_population >= ucsd$age5_plus_population)

length(unique(vax.lj_pop_all$zip_code_tabulation_area))

```

```
## [1] 412
```

```
mean.lj_pop <- mean(vax.lj_pop$percent_of_population_fully_vaccinated)
```

```

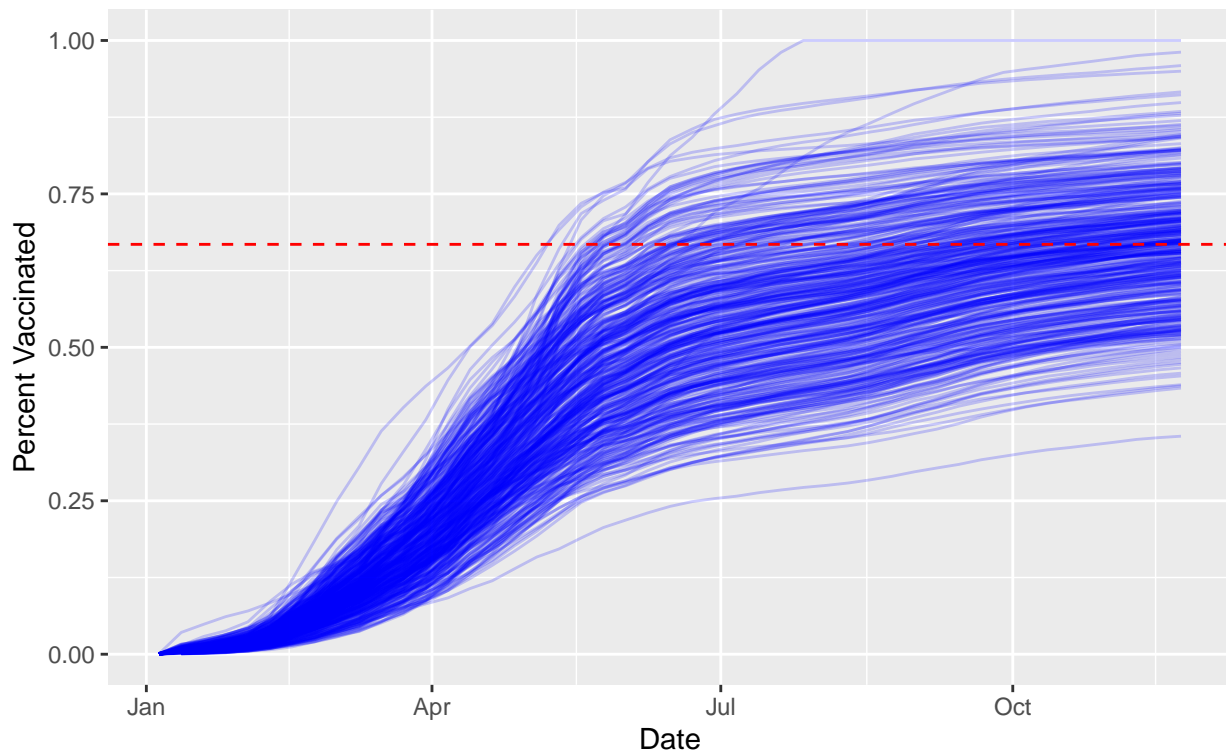
ggplot(data = vax.lj_pop_all) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated,
      group = zip_code_tabulation_area) +
  geom_line(alpha = 0.2,
            color = "blue") +
  geom_hline(yintercept = mean.lj_pop,
             color = "red",
             linetype = 2) +
  labs(x = "Date",
       y = "Percent Vaccinated",
       title = "Vaccination Rates Across California",
       subtitle = "Only areas with a population above or equalt othat of La Jolla")

```

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

Vaccination Rates Across California

Only areas with a population above or equal to that of La Jolla



Session Information

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Arch Linux
##
## Matrix products: default
## BLAS:   /usr/lib/libblas.so.3.10.0
## LAPACK: /usr/lib/liblapack.so.3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
```

```
## [1] ggplot2_3.3.5    lubridate_1.8.0 dplyr_1.0.7
##
## loaded via a namespace (and not attached):
## [1] highr_0.9          pillar_1.6.4      compiler_4.1.2    base64enc_0.1-3
## [5] tools_4.1.2        digest_0.6.28     jsonlite_1.7.2    evaluate_0.14
## [9] lifecycle_1.0.1    tibble_3.1.6      gtable_0.3.0      pkgconfig_2.0.3
## [13] rlang_0.4.12       DBI_1.1.1         yaml_2.2.1        xfun_0.28
## [17] fastmap_1.1.0      repr_1.1.3        withr_2.4.2       stringr_1.4.0
## [21] knitr_1.36         generics_0.1.1    vctrs_0.3.8       grid_4.1.2
## [25] tidyselect_1.1.1   glue_1.5.0        R6_2.5.1          fansi_0.5.0
## [29] rmarkdown_2.11     farver_2.1.0      tidyr_1.1.4       purrr_0.3.4
## [33] skimr_2.1.3        magrittr_2.0.1    scales_1.1.1      ellipsis_0.3.2
## [37] htmltools_0.5.2    assertthat_0.2.1 colorspace_2.0-2  labeling_0.4.2
## [41] utf8_1.2.2         stringi_1.7.6     munsell_0.5.0     crayon_1.4.2
```