class17

Thisha Thiagarajan A15474979

12/3/2021

## Understanding the Data

vax <- read.csv("covid19vaccinesbyzipcode\_test.csv")  
head(vax)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-01-05 92091 San Diego San Diego  
## 2 2021-01-05 92116 San Diego San Diego  
## 3 2021-01-05 95360 Stanislaus Stanislaus  
## 4 2021-01-05 94564 Contra Costa Contra Costa  
## 5 2021-01-05 95501 Humboldt Humboldt  
## 6 2021-01-05 95492 Sonoma Sonoma  
## vaccine\_equity\_metric\_quartile vem\_source  
## 1 4 CDPH-Derived ZCTA Score  
## 2 3 Healthy Places Index Score  
## 3 1 Healthy Places Index Score  
## 4 4 Healthy Places Index Score  
## 5 2 Healthy Places Index Score  
## 6 4 Healthy Places Index Score  
## age12\_plus\_population age5\_plus\_population persons\_fully\_vaccinated  
## 1 1238.3 1303 NA  
## 2 30255.7 31673 45  
## 3 10478.5 12301 NA  
## 4 17033.0 18381 NA  
## 5 20566.6 22061 NA  
## 6 25076.9 28024 NA  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 NA NA  
## 2 898 0.001421  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA  
## percent\_of\_population\_partially\_vaccinated  
## 1 NA  
## 2 0.028352  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA  
## percent\_of\_population\_with\_1\_plus\_dose  
## 1 NA  
## 2 0.029773  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA  
## redacted  
## 1 Information redacted in accordance with CA state privacy requirements  
## 2 No  
## 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

tail(vax)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 84667 2021-11-30 95971 Plumas Plumas  
## 84668 2021-11-30 95747 Placer Placer  
## 84669 2021-11-30 93927 Monterey Monterey  
## 84670 2021-11-30 90004 Los Angeles Los Angeles  
## 84671 2021-11-30 90005 Los Angeles Los Angeles  
## 84672 2021-11-30 90640 Los Angeles Los Angeles  
## vaccine\_equity\_metric\_quartile vem\_source  
## 84667 2 Healthy Places Index Score  
## 84668 4 Healthy Places Index Score  
## 84669 1 Healthy Places Index Score  
## 84670 1 Healthy Places Index Score  
## 84671 1 Healthy Places Index Score  
## 84672 1 Healthy Places Index Score  
## age12\_plus\_population age5\_plus\_population persons\_fully\_vaccinated  
## 84667 5364.3 5710 3126  
## 84668 56213.3 63125 48518  
## 84669 13829.2 16740 11694  
## 84670 52412.5 57024 41305  
## 84671 34648.2 37529 25204  
## 84672 53600.6 58943 41337  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 84667 339 0.547461  
## 84668 4589 0.768602  
## 84669 1637 0.698566  
## 84670 5612 0.724344  
## 84671 4001 0.671587  
## 84672 4896 0.701305  
## percent\_of\_population\_partially\_vaccinated  
## 84667 0.059370  
## 84668 0.072697  
## 84669 0.097790  
## 84670 0.098415  
## 84671 0.106611  
## 84672 0.083063  
## percent\_of\_population\_with\_1\_plus\_dose redacted  
## 84667 0.606831 No  
## 84668 0.841299 No  
## 84669 0.796356 No  
## 84670 0.822759 No  
## 84671 0.778198 No  
## 84672 0.784368 No

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

vax$persons\_fully\_vaccinated

Q2. What column details the Zip code tabulation area?

vax$zip\_code\_tabulation\_area

Q3. What is the earliest date in this dataset?

2021-01-05

Q4. What is the latest date in this dataset?

2021-11-30

skimr::skim(vax)

Data summary

|  |  |
| --- | --- |
| Name | vax |
| Number of rows | 84672 |
| 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 | 48 | 0 |
| local\_health\_jurisdiction | 0 | 1 | 0 | 15 | 240 | 62 | 0 |
| county | 0 | 1 | 0 | 15 | 240 | 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 | 4176 | 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 | 8472 | 0.90 | 9709.47 | 11714.06 | 11 | 526.00 | 4309.50 | 16316.00 | 71552.0 | ▇▂▁▁▁ |
| persons\_partially\_vaccinated | 8472 | 0.90 | 1891.41 | 2100.88 | 11 | 197.00 | 1268.50 | 2874.00 | 20158.0 | ▇▁▁▁▁ |
| percent\_of\_population\_fully\_vaccinated | 8472 | 0.90 | 0.43 | 0.27 | 0 | 0.21 | 0.45 | 0.63 | 1.0 | ▇▆▇▇▂ |
| percent\_of\_population\_partially\_vaccinated | 8472 | 0.90 | 0.10 | 0.10 | 0 | 0.06 | 0.07 | 0.11 | 1.0 | ▇▁▁▁▁ |
| percent\_of\_population\_with\_1\_plus\_dose | 8472 | 0.90 | 0.51 | 0.26 | 0 | 0.31 | 0.54 | 0.71 | 1.0 | ▅▅▇▇▅ |

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?

sum( is.na(vax$persons\_fully\_vaccinated) )

## [1] 8472

8472

Q7. What percent of persons\_fully\_vaccinated values are missing?

10.0%

8472/84672

## [1] 0.1000567

1-0.900

## [1] 0.1

Q8. [Optional]: Why might this data be missing?

This data may be missing because this surverying tools used to gather this data did not get enough responses. Or the researchers may not have access to that data.

## Working with Dates

library(lubridate)

##   
## Attaching package: 'lubridate'

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

today()

## [1] "2021-12-03"

vax$as\_of\_date <- ymd(vax$as\_of\_date)

today() - vax$as\_of\_date[1]

## Time difference of 332 days

vax$as\_of\_date[nrow(vax)] - vax$as\_of\_date[1]

## Time difference of 329 days

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

3 days

today() - vax$as\_of\_date[84672]

## Time difference of 3 days

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

There are 48.

#unique\_count = n\_distinct(vax$as\_of\_date)  
#unique\_count  
unique\_count = length(unique(vax$as\_of\_date))  
unique\_count

## [1] 48

## Working with ZIP codes

#install.packages("zipcodeR")  
library(zipcodeR)  
geocode\_zip('92037')

## # A tibble: 1 × 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 × 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>

#To Pull data for all ZIP codes in the dataset  
#zipdata <- reverse\_zipcode( vax$zip\_code\_tabulation\_area )

## Focus on the San Diego area

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

## [1] 5136

Can also do this with dpylr

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] 5136

head(sd)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-01-05 92091 San Diego San Diego  
## 2 2021-01-05 92116 San Diego San Diego  
## 3 2021-01-05 92118 San Diego San Diego  
## 4 2021-01-05 91977 San Diego San Diego  
## 5 2021-01-05 92060 San Diego San Diego  
## 6 2021-01-05 92083 San Diego San Diego  
## vaccine\_equity\_metric\_quartile vem\_source  
## 1 4 CDPH-Derived ZCTA Score  
## 2 3 Healthy Places Index Score  
## 3 3 Healthy Places Index Score  
## 4 2 Healthy Places Index Score  
## 5 3 CDPH-Derived ZCTA Score  
## 6 2 Healthy Places Index Score  
## age12\_plus\_population age5\_plus\_population persons\_fully\_vaccinated  
## 1 1238.3 1303 NA  
## 2 30255.7 31673 45  
## 3 19835.0 21470 18  
## 4 53851.0 59911 18  
## 5 166.0 166 NA  
## 6 32246.5 36283 16  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 NA NA  
## 2 898 0.001421  
## 3 469 0.000838  
## 4 945 0.000300  
## 5 NA NA  
## 6 442 0.000441  
## percent\_of\_population\_partially\_vaccinated  
## 1 NA  
## 2 0.028352  
## 3 0.021844  
## 4 0.015773  
## 5 NA  
## 6 0.012182  
## percent\_of\_population\_with\_1\_plus\_dose  
## 1 NA  
## 2 0.029773  
## 3 0.022682  
## 4 0.016073  
## 5 NA  
## 6 0.012623  
## redacted  
## 1 Information redacted in accordance with CA state privacy requirements  
## 2 No  
## 3 No  
## 4 No  
## 5 Information redacted in accordance with CA state privacy requirements  
## 6 No

sd.10 <- filter(vax, county == "San Diego" &  
 age5\_plus\_population > 10000)

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

There are 107 distinct zip codes.

#unique\_zipcode = n\_distinct(sd$zip\_code\_tabulation\_area)  
#unique\_zipcode  
  
#can also do this with unique and length  
test\_unique = length(unique(sd$zip\_code\_tabulation\_area))  
test\_unique

## [1] 107

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

92154

row\_largest12 <- sd[which.max(sd$age12\_plus\_population),]  
row\_largest12$zip\_code\_tabulation\_area

## [1] 92154

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

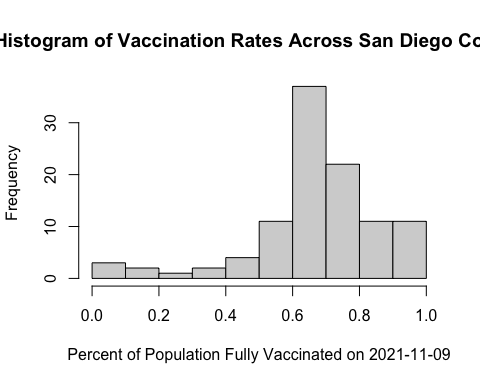
The overall average is 0.6722183. 67.2%

sd.date <- filter(vax, county == "San Diego" &  
 as\_of\_date == "2021-11-16")  
mean(sd.date$percent\_of\_population\_fully\_vaccinated, na.rm = TRUE)

## [1] 0.6722183

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.date$percent\_of\_population\_fully\_vaccinated, main = "Histogram of Vaccination Rates Across San Diego County", xlab = "Percent of Population Fully Vaccinated on 2021-11-09", ylab = "Frequency")



## Focus on UCSD/La Jolla

ucsd <- filter(sd, zip\_code\_tabulation\_area=="92037")  
head(ucsd)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-01-05 92037 San Diego San Diego  
## 2 2021-01-12 92037 San Diego San Diego  
## 3 2021-01-19 92037 San Diego San Diego  
## 4 2021-01-26 92037 San Diego San Diego  
## 5 2021-02-02 92037 San Diego San Diego  
## 6 2021-02-09 92037 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 33675.6 36144 46  
## 2 33675.6 36144 473  
## 3 33675.6 36144 734  
## 4 33675.6 36144 1083  
## 5 33675.6 36144 1620  
## 6 33675.6 36144 2232  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 1270 0.001273  
## 2 1572 0.013087  
## 3 3518 0.020308  
## 4 6220 0.029963  
## 5 8416 0.044821  
## 6 9663 0.061753  
## percent\_of\_population\_partially\_vaccinated  
## 1 0.035137  
## 2 0.043493  
## 3 0.097333  
## 4 0.172089  
## 5 0.232846  
## 6 0.267347  
## percent\_of\_population\_with\_1\_plus\_dose redacted  
## 1 0.036410 No  
## 2 0.056580 No  
## 3 0.117641 No  
## 4 0.202052 No  
## 5 0.277667 No  
## 6 0.329100 No

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:

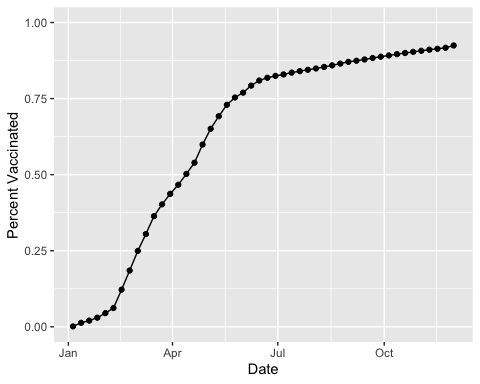
library(ggplot2)  
  
ggplot(ucsd) +  
 aes(x = ucsd$as\_of\_date,  
 y = ucsd$percent\_of\_population\_fully\_vaccinated) +  
 geom\_point() +  
 geom\_line(group=1) +  
 ylim(c(0,1)) +  
 labs(x = "Date", y="Percent Vaccinated")

## Warning: Use of `ucsd$as\_of\_date` is discouraged. Use `as\_of\_date` instead.

## Warning: Use of `ucsd$percent\_of\_population\_fully\_vaccinated` is discouraged.  
## Use `percent\_of\_population\_fully\_vaccinated` instead.

## Warning: Use of `ucsd$as\_of\_date` is discouraged. Use `as\_of\_date` instead.

## Warning: Use of `ucsd$percent\_of\_population\_fully\_vaccinated` is discouraged.  
## Use `percent\_of\_population\_fully\_vaccinated` instead.



## Comparing 92037 to other similar sized areas?

# Subset to all CA areas with a population as large as 92037  
vax.36 <- filter(vax, age5\_plus\_population > 36144 &  
 as\_of\_date == "2021-11-16")  
  
head(vax.36)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-11-16 92345 San Bernardino San Bernardino  
## 2 2021-11-16 92553 Riverside Riverside  
## 3 2021-11-16 92058 San Diego San Diego  
## 4 2021-11-16 91786 San Bernardino San Bernardino  
## 5 2021-11-16 92507 Riverside Riverside  
## 6 2021-11-16 93021 Ventura Ventura  
## vaccine\_equity\_metric\_quartile vem\_source  
## 1 1 Healthy Places Index Score  
## 2 1 Healthy Places Index Score  
## 3 1 Healthy Places Index Score  
## 4 2 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 66047.5 75539 35432  
## 2 61770.8 70472 37411  
## 3 34956.0 39695 14023  
## 4 45602.3 50410 30834  
## 5 51432.5 55253 31939  
## 6 32753.7 36197 24918  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 4389 0.469056  
## 2 4846 0.530863  
## 3 2589 0.353269  
## 4 3132 0.611664  
## 5 3427 0.578050  
## 6 2012 0.688400  
## percent\_of\_population\_partially\_vaccinated  
## 1 0.058102  
## 2 0.068765  
## 3 0.065222  
## 4 0.062131  
## 5 0.062024  
## 6 0.055585  
## percent\_of\_population\_with\_1\_plus\_dose redacted  
## 1 0.527158 No  
## 2 0.599628 No  
## 3 0.418491 No  
## 4 0.673795 No  
## 5 0.640074 No  
## 6 0.743985 No

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?

The calculated mean is 0.6645132.

mean\_vax36 <- mean(vax.36$percent\_of\_population\_fully\_vaccinated)  
mean\_vax36

## [1] 0.6645132

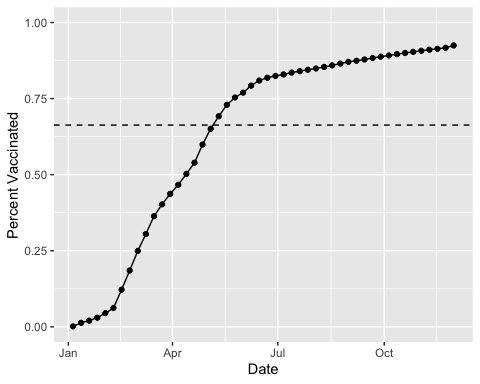
ggplot(ucsd) +  
 aes(x = ucsd$as\_of\_date,  
 y = ucsd$percent\_of\_population\_fully\_vaccinated) +  
 geom\_point() +  
 geom\_line(group=1) +  
 geom\_hline(yintercept = 0.6629812, linetype= "dashed") +  
 ylim(c(0,1)) +  
 labs(x = "Date", y="Percent Vaccinated")

## Warning: Use of `ucsd$as\_of\_date` is discouraged. Use `as\_of\_date` instead.

## Warning: Use of `ucsd$percent\_of\_population\_fully\_vaccinated` is discouraged.  
## Use `percent\_of\_population\_fully\_vaccinated` instead.

## Warning: Use of `ucsd$as\_of\_date` is discouraged. Use `as\_of\_date` instead.

## Warning: Use of `ucsd$percent\_of\_population\_fully\_vaccinated` is discouraged.  
## Use `percent\_of\_population\_fully\_vaccinated` instead.



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 “2021-11-16”?

Min: 0.353269 1st Quartile: 0.591029 Median: 0.666919 3rd Quartile: 0.731112 Max: 1.000000 Mean: 0.6645132

fivenum(vax.36$percent\_of\_population\_fully\_vaccinated)

## [1] 0.353269 0.591029 0.666919 0.731112 1.000000

mean(vax.36$percent\_of\_population\_fully\_vaccinated)

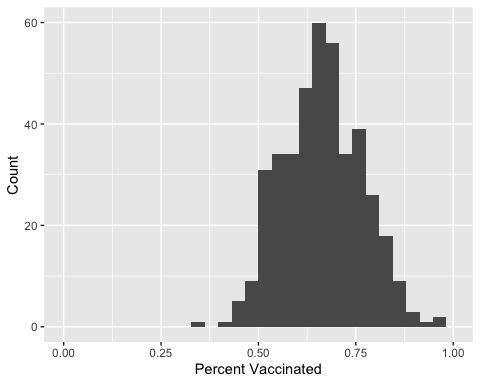
## [1] 0.6645132

Q18. Using ggplot generate a histogram of this data.

ggplot(vax.36) +  
 aes(x = percent\_of\_population\_fully\_vaccinated) +  
 geom\_histogram() +  
 xlim(c(0,1)) +  
 labs(x = "Percent Vaccinated", y="Count")

## `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 == "2021-11-16") %>%   
 filter(zip\_code\_tabulation\_area=="92040") %>%  
 select(percent\_of\_population\_fully\_vaccinated)

## percent\_of\_population\_fully\_vaccinated  
## 1 0.52142

vax %>% filter(as\_of\_date == "2021-11-16") %>%   
 filter(zip\_code\_tabulation\_area=="92109") %>%  
 select(percent\_of\_population\_fully\_vaccinated)

## percent\_of\_population\_fully\_vaccinated  
## 1 0.68912

92040 ZIP code area is below and 92109 ZIP code area is above.

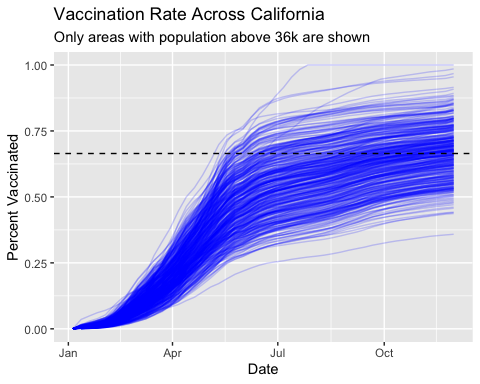
Q20. Finally make a time course plot of vaccination progress for all areas in the full dataset with a age5\_plus\_population > 36144.

vax.36.all <- filter(vax, age5\_plus\_population > 36144)  
head(vax.36.all)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-01-05 91789 Los Angeles Los Angeles  
## 2 2021-01-05 91320 Ventura Ventura  
## 3 2021-01-05 91311 Los Angeles Los Angeles  
## 4 2021-01-05 92705 Orange Orange  
## 5 2021-01-05 92508 Riverside Riverside  
## 6 2021-01-05 92802 Orange Orange  
## 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 3 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 39345.3 42376 23  
## 2 38216.8 42334 11  
## 3 36345.5 38912 16  
## 4 40093.0 44215 16  
## 5 32415.3 36303 NA  
## 6 35113.6 39393 13  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 1121 0.000543  
## 2 771 0.000260  
## 3 903 0.000411  
## 4 768 0.000362  
## 5 NA NA  
## 6 512 0.000330  
## percent\_of\_population\_partially\_vaccinated  
## 1 0.026454  
## 2 0.018212  
## 3 0.023206  
## 4 0.017370  
## 5 NA  
## 6 0.012997  
## percent\_of\_population\_with\_1\_plus\_dose  
## 1 0.026997  
## 2 0.018472  
## 3 0.023617  
## 4 0.017732  
## 5 NA  
## 6 0.013327  
## redacted  
## 1 No  
## 2 No  
## 3 No  
## 4 No  
## 5 Information redacted in accordance with CA state privacy requirements  
## 6 No

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 Vaccinated",  
 title="Vaccination Rate Across California",  
 subtitle="Only areas with population above 36k are shown") +  
 geom\_hline(yintercept = 0.6645132, linetype= "dashed")

## Warning: Removed 177 row(s) containing missing values (geom\_path).



Q21. How do you feel about traveling for Thanksgiving and meeting for in-person class next Week?

Would prefer to have time to get tested before meeting.