Class 17: Vaccination Rate Mini Project

Kaitlyn Powell

Background

"Statewide COVID-19 Vaccines Administered by ZIP Code"

```r
url <- "covid19vaccinesbyzipcode_test.csv"
```

Getting Started

```r
# Import vaccination data
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction          county
1 2021-01-05                    92240                 Riverside       Riverside
2 2021-01-05                    91302               Los Angeles     Los Angeles
3 2021-01-05                    93420           San Luis Obispo San Luis Obispo
4 2021-01-05                    91901                 San Diego       San Diego
5 2021-01-05                    94110             San Francisco   San Francisco
6 2021-01-05                    91902                 San Diego       San Diego
  vaccine_equity_metric_quartile                 vem_source
1                              1 Healthy Places Index Score
2                              4 Healthy Places Index Score
3                              3 Healthy Places Index Score
4                              3 Healthy Places Index Score
5                              4 Healthy Places Index Score
6                              4 Healthy Places Index Score
  age12_plus_population age5_plus_population tot_population
1               29270.5                33093          35278
```

```
2               23163.9                25899          26712
3               26694.9                29253          30740
4               15549.8                16905          18162
5               64350.7                68320          72380
6               16620.7                18026          18896
  persons_fully_vaccinated persons_partially_vaccinated
1                       NA                           NA
2                       15                          614
3                       NA                           NA
4                       NA                           NA
5                       17                         1268
6                       15                          397
  percent_of_population_fully_vaccinated
1                                     NA
2                               0.000562
3                                     NA
4                                     NA
5                               0.000235
6                               0.000794
  percent_of_population_partially_vaccinated
1                                         NA
2                                   0.022986
3                                         NA
4                                         NA
5                                   0.017519
6                                   0.021010
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                               0.023548                  NA
3                                     NA                  NA
4                                     NA                  NA
5                               0.017754                  NA
6                               0.021804                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                        2
2                        NA                       15
3                        NA                        4
4                        NA                        8
5                        NA                       17
6                        NA                       15
                                                               redacted
1 Information redacted in accordance with CA state privacy requirements
2 Information redacted in accordance with CA state privacy requirements
```

```
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
174631 2022-11-22                    94066                 San Mateo
174632 2022-11-22                    92254                 Riverside
174633 2022-11-22                    94065                 San Mateo
174634 2022-11-22                    92280            San Bernardino
174635 2022-11-22                    94929                     Marin
174636 2022-11-22                    92313            San Bernardino
               county vaccine_equity_metric_quartile                 vem_source
174631      San Mateo                              4 Healthy Places Index Score
174632      Riverside                              1 Healthy Places Index Score
174633      San Mateo                              4 Healthy Places Index Score
174634 San Bernardino                             NA            No VEM Assigned
174635          Marin                              4    CDPH-Derived ZCTA Score
174636 San Bernardino                              2 Healthy Places Index Score
       age12_plus_population age5_plus_population tot_population
174631               37730.3                40903          43101
174632                7882.3                 8985           9779
174633               10465.5                11778          12461
174634                   0.0                    0             NA
174635                 174.2                  218            254
174636               10842.9                11847          12547
       persons_fully_vaccinated persons_partially_vaccinated
174631                    38105                         2889
174632                     9456                         1688
174633                    11238                          889
174634                       NA                           NA
174635                       NA                           NA
174636                     7948                          600
       percent_of_population_fully_vaccinated
174631                               0.884086
174632                               0.966970
174633                               0.901854
174634                                     NA
174635                                     NA
```

```
174636                              0.633458
       percent_of_population_partially_vaccinated
174631                                  0.067029
174632                                  0.172615
174633                                  0.071343
174634                                        NA
174635                                        NA
174636                                  0.047820
       percent_of_population_with_1_plus_dose booster_recip_count
174631                              0.951115               27085
174632                              1.000000                3840
174633                              0.973197                8701
174634                                    NA                  NA
174635                                    NA                  NA
174636                              0.681278                4522
       bivalent_dose_recip_count eligible_recipient_count
174631                      9127                    37620
174632                       372                     9430
174633                      3456                    11021
174634                        NA                       14
174635                        NA                      159
174636                      1085                     7921
                                                                redacted
174631                                                                No
174632                                                                No
174633                                                                No
174634 Information redacted in accordance with CA state privacy requirements
174635 Information redacted in accordance with CA state privacy requirements
174636                                                                No
```

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

`persons_fully_vaccinated` details the total number of people fully vaccinated.

What column details the Zip code tabulation area?

`zip_code_tabulation_area` details the 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?

2022-11-22

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

Table 1: Data summary

Name	vax
Number of rows	174636
Number of columns	18
Column type frequency:	
character	5
numeric	13
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	99	0
local_health_jurisdiction	0	1	0	15	495	62	0
county	0	1	0	15	495	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	8613	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.04	18993.88	0	1346.95	13685.10	31756.12	88556.7	
age5_plus_population	0	1.00	20875.24	21105.98	0	1460.50	15364.00	34877.00	101902.0	
tot_population	8514	0.95	23372.77	22628.51	12	2126.00	18714.00	38168.00	111165.0	
persons_fully_vaccinated	14921	0.91	13466.34	14722.46	11	883.00	8024.00	22529.00	87186.0	
persons_partially_vaccinated	14921	0.91	1707.50	1998.80	11	167.00	1194.00	2547.00	39204.0	
percent_of_population_fully_vaccinated	18665	0.89	0.55	0.25	0	0.39	0.59	0.73	1.0	
percent_of_population_partially_vaccinated	18665	0.89	0.08	0.09	0	0.05	0.06	0.08	1.0	
percent_of_population_with_1_plus_dose	19562	0.89	0.61	0.25	0	0.46	0.65	0.79	1.0	
booster_recip_count	70421	0.60	5655.17	6867.49	11	280.00	2575.00	9421.00	58304.0	
bivalent_dose_recip_count	156958	0.10	1646.02	2161.84	11	109.00	719.00	2443.00	18109.0	
eligible_recipient_count	0	1.00	12309.10	14555.83	0	466.00	5810.00	21140.00	86696.0	

> Q5. How many numeric columns are in this dataset?

There are 13 numeric columns.

> Q6. Note that there are "missing values" in the dataset. How many NA values there in the persons_fully_vaccinated column?

There are 15440 NA values in the persons_fully_vaccinated column.

```
sum( is.na(vax$persons_fully_vaccinated) )
```

```
[1] 14921
```

> What percent of persons_fully_vaccinated values are missing (to 2 significant figures)?

89 %

Working with dates

```
library(lubridate)
```

```
Loading required package: timechange


Attaching package: 'lubridate'


The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```
today()
```

```
[1] "2022-11-28"
```

```
# This will give an Error!
#today() - vax$as_of_date
```

```r
# Specify that we are using the year-month-day format
vax$as_of_date <- ymd(vax$as_of_date)
```

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

```
Time difference of 692 days
```

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

```
Time difference of 686 days
```

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

6 days

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

There are 98 unique dates.

Working with ZIP codes

```r
library(zipcodeR)
```

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

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

```r
zip_distance('92037','92109')
```

```
  zipcode_a zipcode_b distance
1     92037     92109     2.33
```

```
reverse_zipcode(c('92037', "92109") )
```

```
# A tibble: 2 x 24
  zipcode zipcode_~1 major~2 post_~3 common_c~4 county state   lat   lng timez~5
  <chr>   <chr>      <chr>   <chr>       <blob> <chr>  <chr> <dbl> <dbl> <chr>
1 92037   Standard   La Jol~ La Jol~ <raw 20 B> San D~ CA     32.8 -117. Pacific
2 92109   Standard   San Di~ San Di~ <raw 21 B> San D~ CA     32.8 -117. Pacific
# ... with 14 more variables: 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>, and abbreviated variable names
#   1: zipcode_type, 2: major_city, 3: post_office_city, ...
```

Pull data for all ZIP codes in the dataset

#zipdata <- reverse_zipcode(vax$zip_code_tabulation_area)

Focus on the San Diego area

```
#vax$county == "San Diego"
```

```
# Subset to San Diego county only areas
sd <- vax[ vax$county == "San Diego" , ]
```

```
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] 10593
```

```
sd.10 <- filter(vax, county == "San Diego" &
                age5_plus_population > 10000)
```

```
which.max(sd$age12_plus_population)
```

```
[1] 53
```

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

There are 107 distinct zip codes listed for San Diego county.

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

92154

```
skimr::skim(sd.10)
```

Table 4: Data summary

Name	sd.10
Number of rows	7524
Number of columns	18
Column type frequency:	
character	4
Date	1
numeric	13

Table 4: Data summary

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
local_health_jurisdiction	0	1	9	9	0	1	0
county	0	1	9	9	0	1	0
vem_source	0	1	23	26	0	2	0
redacted	0	1	2	69	0	2	0

Variable type: Date

skim_variable	n_missing	complete_rate	min	max	median	n_unique
as_of_date	0	1	2021-01-05	2022-11-22	2021-12-14	99

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
zip_code_tabulation_area	0	1.00	92054.57	70.78	91901.0	92017.75	92070.00	92113.25	92173.00	
vaccine_equity_metric_quartile	0	1.00	2.86	0.94	1.0	2.00	3.00	4.00	4.00	
age12_plus_population	0	1.00	36365.97	15210.57	10061.5	25399.93	37240.85	54737.18	76365.20	
age5_plus_population	0	1.00	39922.24	16787.61	10704.0	28218.50	40270.50	49486.75	82971.00	
tot_population	0	1.00	42630.07	17989.65	11417.0	29980.00	43641.00	53267.25	88979.00	
persons_fully_vaccinated	40	0.99	24866.61	15994.53	11.0	13257.25	23486.50	34900.50	87186.00	
persons_partially_vaccinated	40	0.99	3225.32	2704.12	11.0	1716.00	2568.00	3787.00	30455.00	
percent_of_population_fully_vaccinated	40	0.99	0.58	0.25	0.0	0.49	0.64	0.73	1.00	
percent_of_population_partially_vaccinated	40	0.99	0.08	0.06	0.0	0.05	0.06	0.09	0.98	
percent_of_population_with_1_plus_dose	40	0.99	0.64	0.25	0.0	0.55	0.70	0.79	1.00	
booster_recip_count	2526	0.66	10206.24	8011.21	11.0	3891.00	9068.50	14938.75	46665.00	
bivalent_dose_recip_count	6588	0.12	2542.04	2258.18	11.0	701.00	2030.50	3696.75	12081.00	
eligible_recipient_count	0	1.00	24712.17	16035.06	0.0	13153.00	23326.00	34851.00	86696.00	

```r
library(dplyr)

sd.11.15 <- sd %>% filter(as_of_date == "2022-11-15")
nrow(sd.11.15)
```

```
[1] 107
```

```
sd.11.15.vac <- sd.11.15$percent_of_population_fully_vaccinated
```

```
sd.11.15.avg <- mean(sd.11.15.vac, na.rm = TRUE)
sd.11.15.avg
```

```
[1] 0.7369099
```

> Q13. What is the overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2022-1s-15"?

The overall average of "Percent of Population Fully Vaccinated value for all San Diego county as of 2022-11-15 is 0.7369099.

> 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 "2022-11-15"?

```
#ggplot(sd.11.15) +
 # aes(sd.11.15$zip_code_tabulation_area, sd.11.15$percent_of_population_fully_vaccinated)
 #geom_bar()
```

Note: I kept getting an error when trying to create a histogram.

Focus on UCSD/La Jolla

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

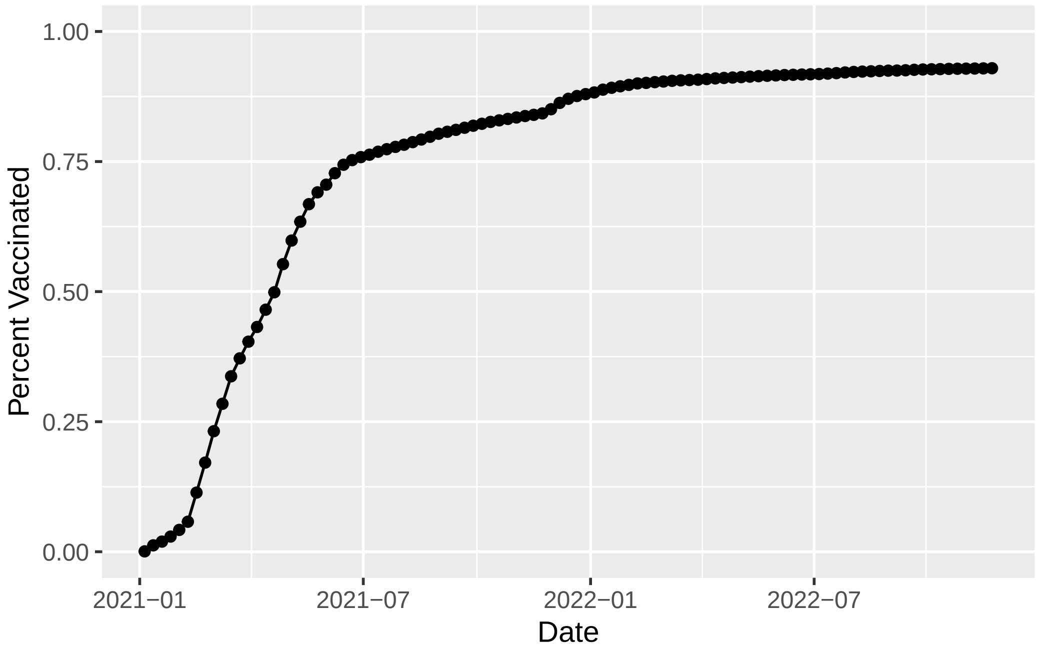
```
ggplot(ucsd) +
  aes(ucsd$as_of_date, ucsd$percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
```

```
Warning: Use of `ucsd$as_of_date` is discouraged.
i Use `as_of_date` instead.

Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
i Use `percent_of_population_fully_vaccinated` instead.

Warning: Use of `ucsd$as_of_date` is discouraged.
i Use `as_of_date` instead.

Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
i Use `percent_of_population_fully_vaccinated` instead.
```



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

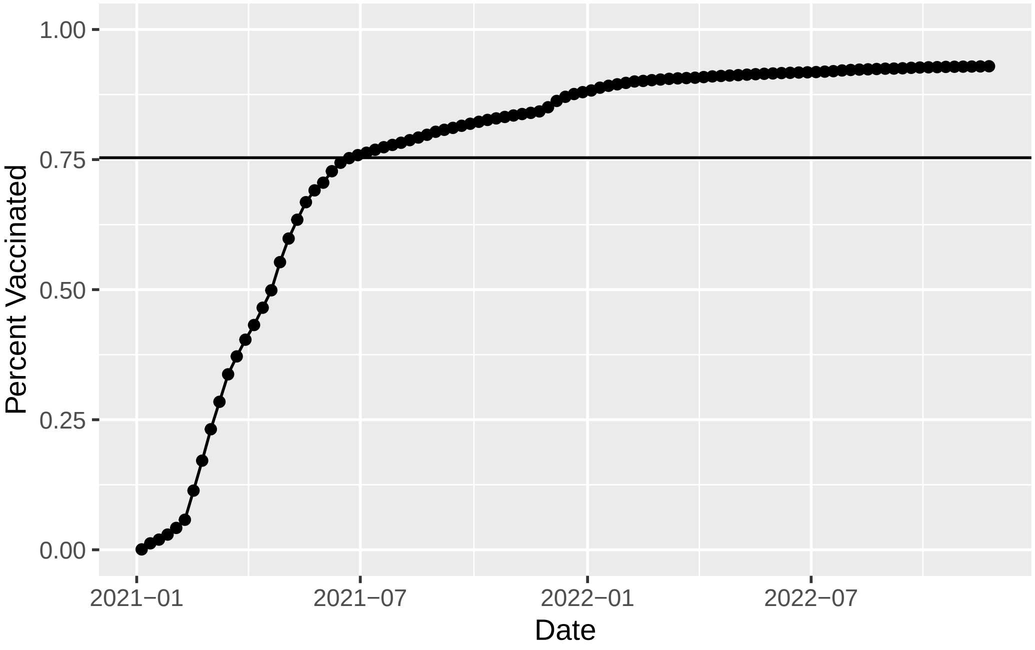
```
p + geom_hline(yintercept = mean(ucsd$percent_of_population_fully_vaccinated))
```

```
Warning: Use of `ucsd$as_of_date` is discouraged.
i Use `as_of_date` instead.

Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
i Use `percent_of_population_fully_vaccinated` instead.

Warning: Use of `ucsd$as_of_date` is discouraged.
i Use `as_of_date` instead.

Warning: Use of `ucsd$percent_of_population_fully_vaccinated` is discouraged.
i Use `percent_of_population_fully_vaccinated` instead.
```



```
m <- mean(ucsd$percent_of_population_fully_vaccinated)
m
```

```
[1] 0.7535428
```

```
# Subset to all CA areas with a population as large as 92037
vax.36 <- filter(vax, age5_plus_population > 36144 &
                  as_of_date == "2022-11-15")

#head(vax.36)
```

> 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 "2022-11-15"?

Min: 0.000760 First Quartile: 0.755672 Median: 0.870625 Third Quartile: 0.915898 Max: 0.929365 Mean: 0.7535428

```
fivenum(ucsd$percent_of_population_fully_vaccinated)
```

```
[1] 0.000760 0.755672 0.870625 0.915898 0.929365
```

> Q18. Using ggplot generate a histogram of this data.

```
#ggplot(ucsd) +
 # aes(ucsd$percent_of_population_fully_vaccinated, ucsd$tot_population) +
  #geom_bar()
```

Note: I kept getting an error when trying to create a histogram. > Q19. Is the 92109 and 92040 ZIP code areas above or below the average value you calculated for all these above?

Based on picture in the lab handout, the average value is slightly above.

```
vax %>% filter(as_of_date == "2022-11-15") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

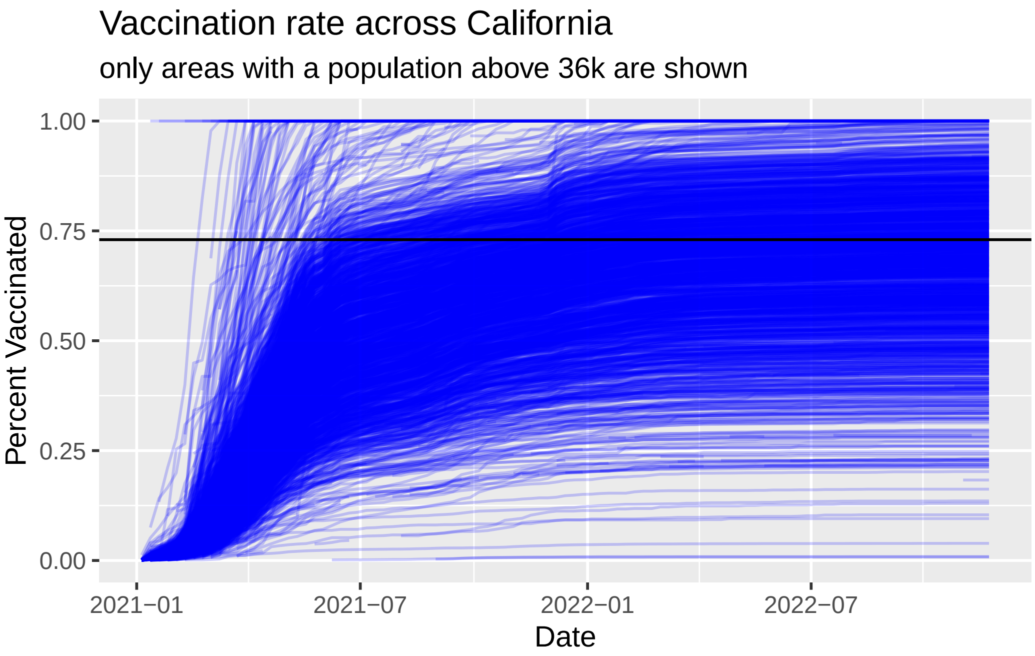
```
  percent_of_population_fully_vaccinated
1                               0.546646
```

> 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, )


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 a population above 36k are shown") +
  geom_hline(yintercept = 0.73)
```

```
Warning: Removed 16568 rows containing missing values (`geom_line()`).
```



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

This data makes me want to be even more cautious when travelling.