

class14

Vince (PID: A15422556)

3/3/2022

Getting Started

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

Q2. What column details the Zip code tabulation area? `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? 2022-03-01

```
# 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                92549                Riverside    Riverside
## 2 2021-01-05                92130                San Diego      San Diego
## 3 2021-01-05                92397            San Bernardino San Bernardino
## 4 2021-01-05                94563            Contra Costa    Contra Costa
## 5 2021-01-05                94519            Contra Costa    Contra Costa
## 6 2021-01-05                91042            Los Angeles    Los Angeles
##   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                        4 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                2348.4                2461                NA
## 2                46300.3                53102                61
## 3                3695.6                4225                NA
## 4                17216.1                18896                NA
## 5                16861.2                18678                NA
## 6                23962.2                25741                NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                        NA                NA
## 2                        27                0.001149
## 3                        NA                NA
## 4                        NA                NA
```

```
## 5 NA NA
## 6 NA NA
## percent_of_population_partially_vaccinated
## 1 NA
## 2 0.000508
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## percent_of_population_with_1_plus_dose booster_recip_count
## 1 NA NA
## 2 0.001657 NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
## 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
```

```
head(vax$as_of_date)
```

```
## [1] "2021-01-05" "2021-01-05" "2021-01-05" "2021-01-05" "2021-01-05"
## [6] "2021-01-05"
```

```
tail(vax$as_of_date)
```

```
## [1] "2022-03-01" "2022-03-01" "2022-03-01" "2022-03-01" "2022-03-01"
## [6] "2022-03-01"
```

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

```
## [1] "2022-03-01"
```

Skim Package

```
library(skimr)
skimr::skim(vax)
```

Table 1: Data summary

Name	vax
Number of rows	107604
Number of columns	15

Column type frequency:

Table 1: Data summary

character	5
numeric	10
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	61	0
local_health_jurisdiction	0	1	0	15	305	62	0
county	0	1	0	15	305	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_area0	1.00	93665.11	17.39	0.0001	92257.73	3658.55	380.57	635.0		
vaccine_equity_metric	5407	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.04	993.91	0	1346.95	3685.31	756.13	556.7	
age5_plus_population	0	1.00	20875.21	106.02	0	1460.50	5364.04	877.00	1902.0	
persons_fully_vaccinated	18338	0.83	12155.61	6063.81	1	1066.25	374.50	20005.07	7744.0	
persons_partially_vaccinated	18338	0.83	831.74	1348.68	11	76.00	372.00	1076.00	34219.0	
percent_of_population_fully_vaccinated	18338	0.83	0.51	0.26	0	0.33	0.54	0.70	1.0	
percent_of_population_partially_vaccinated	18338	0.83	0.05	0.09	0	0.01	0.03	0.05	1.0	
percent_of_population_1_plus_doses	18338	1.00	0.54	0.28	0	0.36	0.58	0.75	1.0	
booster_recip_count	64317	0.40	4100.55	900.21	11	176.00	1136.00	154.50	60602.0	

Q5. How many numeric columns are in this dataset? 9 columns

Q6. Note that there are “missing values” in the dataset. How many NA values there in the persons_fully_vaccinated column? 18338

Q7. What percent of persons_fully_vaccinated values are missing (to 2 significant figures)? 17.04%

```
(sum(is.na(vax$persons_fully_vaccinated)) / 107604) * 100
```

```
## [1] 17.04212
```

Q8. [Optional]: Why might this data be missing? It’s possible the patients didn’t follow up for their second dose so they aren’t considered “fully vaccinated.”

Working with Dates

Q9. How many days have passed since the last update of the dataset? 2 days

Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)? 61 unique entries

```
library(lubridate)
```

```
##
```

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

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

```
##
```

```
##     date, intersect, setdiff, union
```

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

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

```
## Time difference of 2 days
```

```
length(unique(vax$as_of_date))
```

```
## [1] 61
```

Working with ZIP codes

```
library(zipcodeR)
```

```
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(ggplot2)
```

Focus on the San Diego area.

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

```
dim(sd)
```

```
## [1] 6527  15
```

Q11. How many distinct zip codes are listed for San Diego County? 107

```
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?
92154

```
sd[which.max(sd$age12_plus_population), ]
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction   county
## 91 2021-01-05                92154                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 91                2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 91                76365.2                82971                18
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 91                22                0.000217
##   percent_of_population_partially_vaccinated
## 91                0.000265
##   percent_of_population_with_1_plus_dose booster_recip_count
## 91                0.000482                NA
##                                     redacted
## 91 Information redacted in accordance with CA state privacy requirements
```

Q13. What is the overall average “Percent of Population Fully Vaccinated” value for all San Diego “County” as of “2022-03-01”? 70.5%

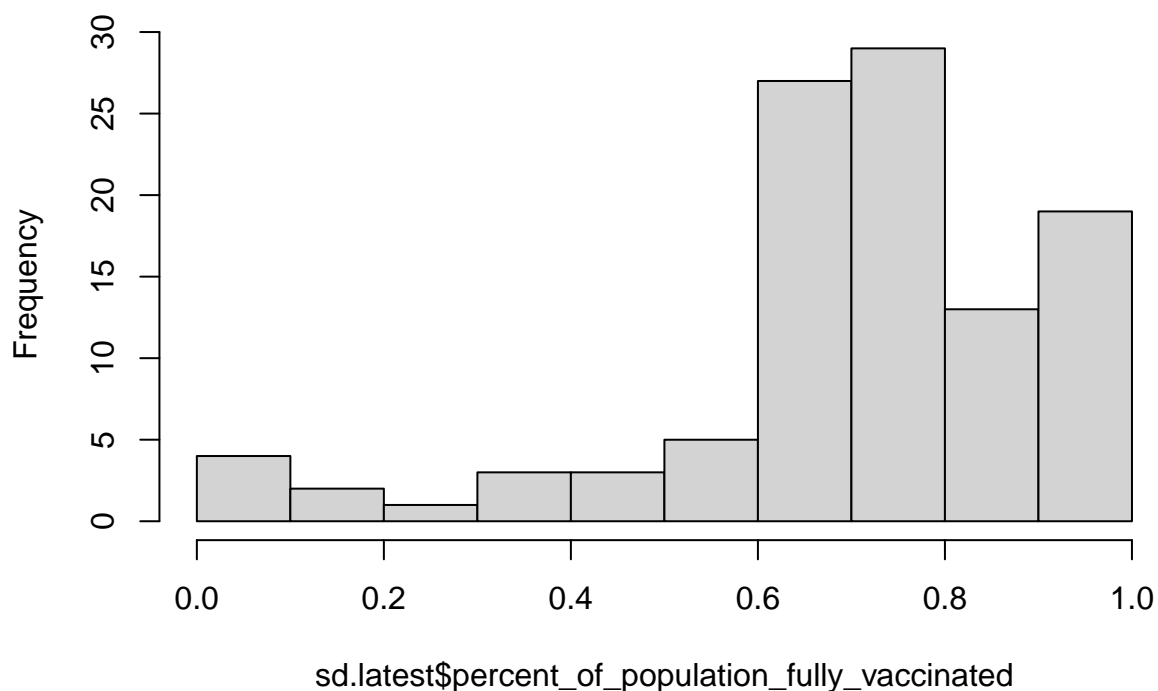
```
#Filter to the day
sd.latest <- filter(sd, as_of_date=="2022-03-01")
mean(sd.latest$percent_of_population_fully_vaccinated, na.rm=TRUE)
```

```
## [1] 0.7052904
```

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 “2022-03-01”?

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

Histogram of sd.latest\$percent_of_population_fully_vaccinated



Focus on UCSD/La Jolla

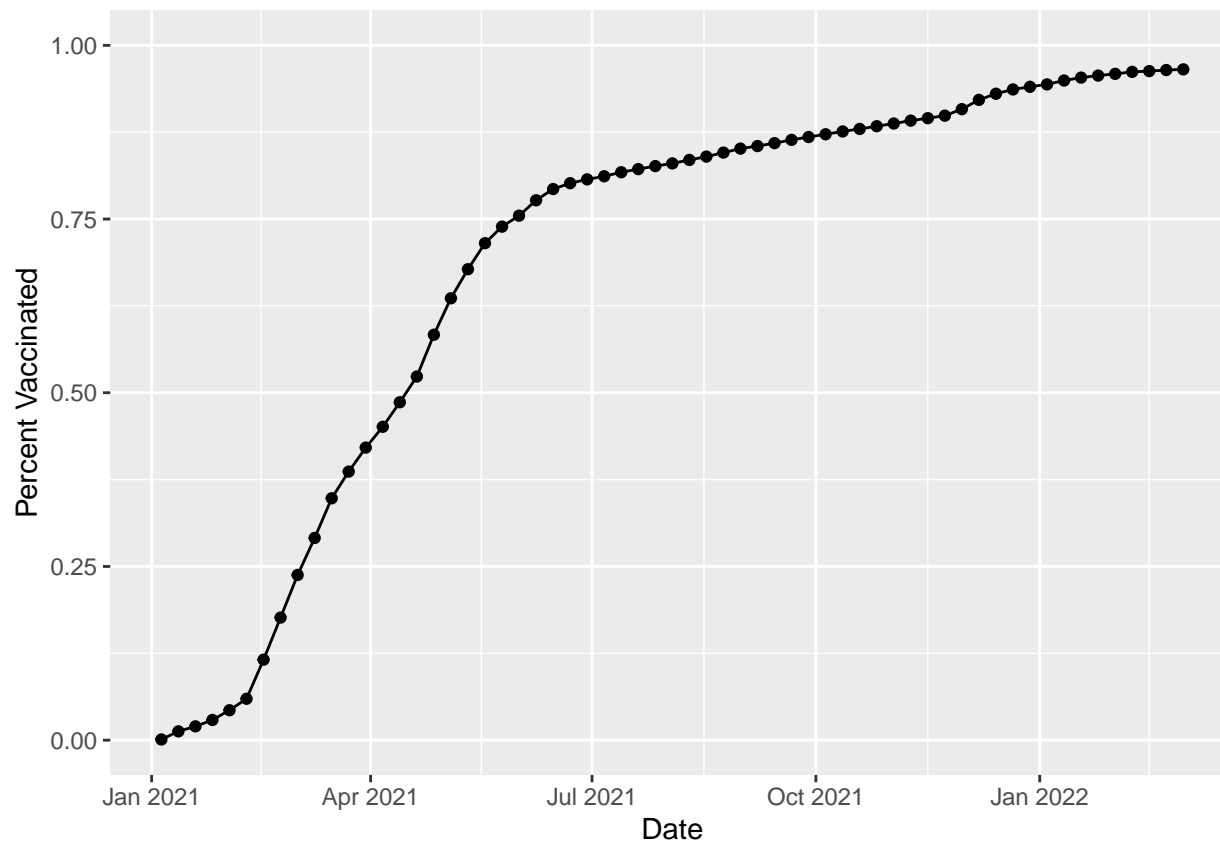
```
ucsd <- filter(sd, zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

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

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

```
baseplot <- ggplot(ucsd) +
  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")
```

```
baseplot
```



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 “2022-03-01”. Add this as a straight horizontal line to your plot from above with the `geom_hline()` function?

```
vax.36 <- filter(vax, age5_plus_population > 36144 &
  as_of_date == "2022-03-01")
```

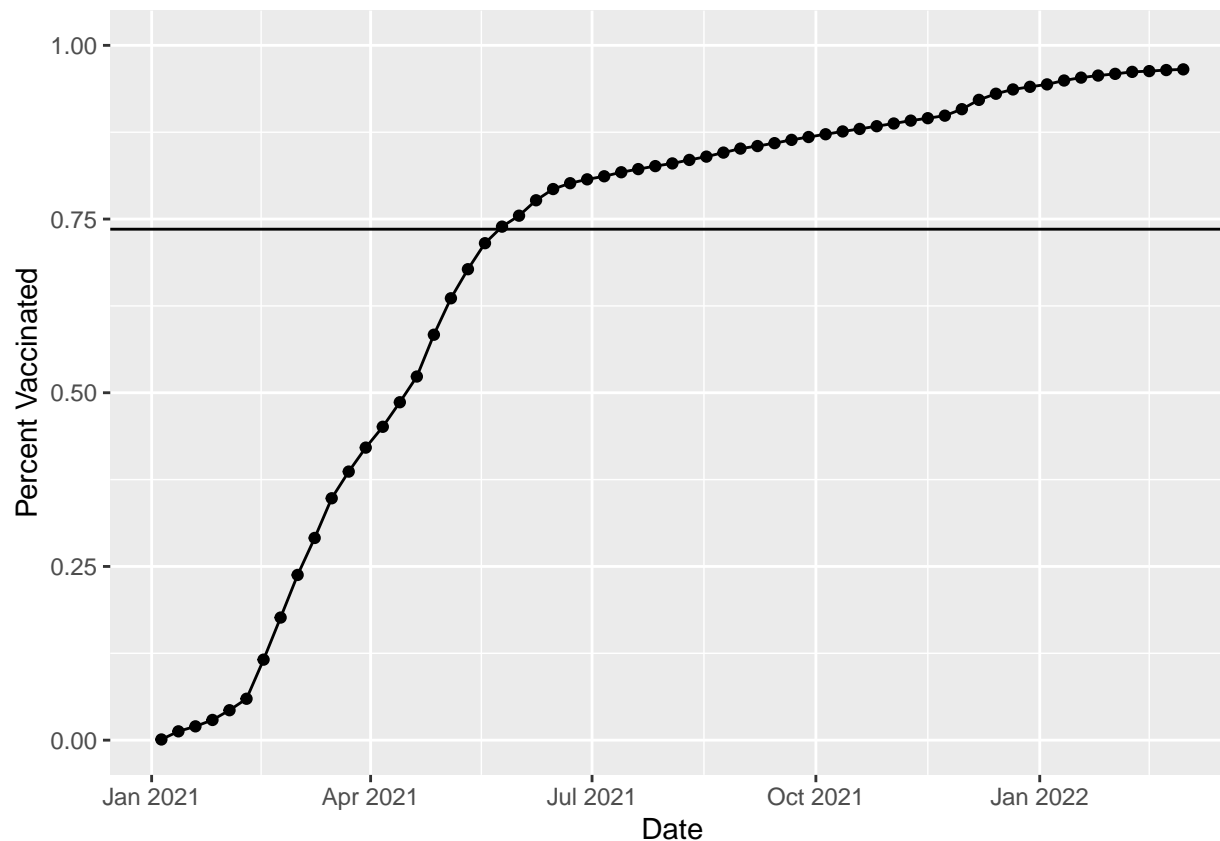
```
head(vax.36)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2022-03-01          95628      Sacramento Sacramento
## 2 2022-03-01          90808      Long Beach Los Angeles
## 3 2022-03-01          92507      Riverside Riverside
## 4 2022-03-01          92626      Orange Orange
## 5 2022-03-01          93257      Tulare Tulare
## 6 2022-03-01          90011      Los Angeles Los Angeles
##   vaccine_equity_metric_quartile      vem_source
## 1                3 Healthy Places Index Score
## 2                4 Healthy Places Index Score
## 3                1 Healthy Places Index Score
## 4                3 Healthy Places Index Score
## 5                1 Healthy Places Index Score
## 6                1 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1                35579.0                38694                28842
```

## 2	33952.3	37179	29383
## 3	51432.5	55253	34455
## 4	44238.8	47883	33767
## 5	61519.8	70784	42919
## 6	87902.8	101902	65342
##	persons_partially_vaccinated	percent_of_population_fully_vaccinated	
## 1	1990		0.745387
## 2	2112		0.790312
## 3	3947		0.623586
## 4	2937		0.705198
## 5	5868		0.606338
## 6	15255		0.641224
##	percent_of_population_partially_vaccinated		
## 1		0.051429	
## 2		0.056806	
## 3		0.071435	
## 4		0.061337	
## 5		0.082900	
## 6		0.149703	
##	percent_of_population_with_1_plus_dose	booster_recip_count	redacted
## 1	0.796816	16913	No
## 2	0.847118	17253	No
## 3	0.695021	15073	No
## 4	0.766535	17595	No
## 5	0.689238	17740	No
## 6	0.790927	19928	No

```
ave.36 <- mean(vax.36$percent_of_population_fully_vaccinated)
```

```
baseplot +  
  geom_hline(yintercept = ave.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 “2022-03-01”?

```
summary(vax.36)
```

```
##      as_of_date      zip_code_tabulation_area local_health_jurisdiction
## Min.   :2022-03-01  Min.   :90001                Length:411
## 1st Qu.:2022-03-01  1st Qu.:91762                Class :character
## Median :2022-03-01  Median :92646                Mode  :character
## Mean   :2022-03-01  Mean   :92862
## 3rd Qu.:2022-03-01  3rd Qu.:94517
## Max.   :2022-03-01  Max.   :96003
##      county      vaccine_equity_metric_quartile vem_source
## Length:411      Min.   :1.000                Length:411
## Class :character 1st Qu.:1.000                Class :character
## Mode  :character Median :2.000                Mode  :character
##                      Mean   :2.353
##                      3rd Qu.:3.000
##                      Max.   :4.000
## age12_plus_population age5_plus_population persons_fully_vaccinated
## Min.   :31651          Min.   : 36181          Min.   :15443
## 1st Qu.:37694          1st Qu.: 41613          1st Qu.:30658
## Median :43985          Median : 48573          Median :35351
## Mean   :46847          Mean   : 52012          Mean   :38227
```

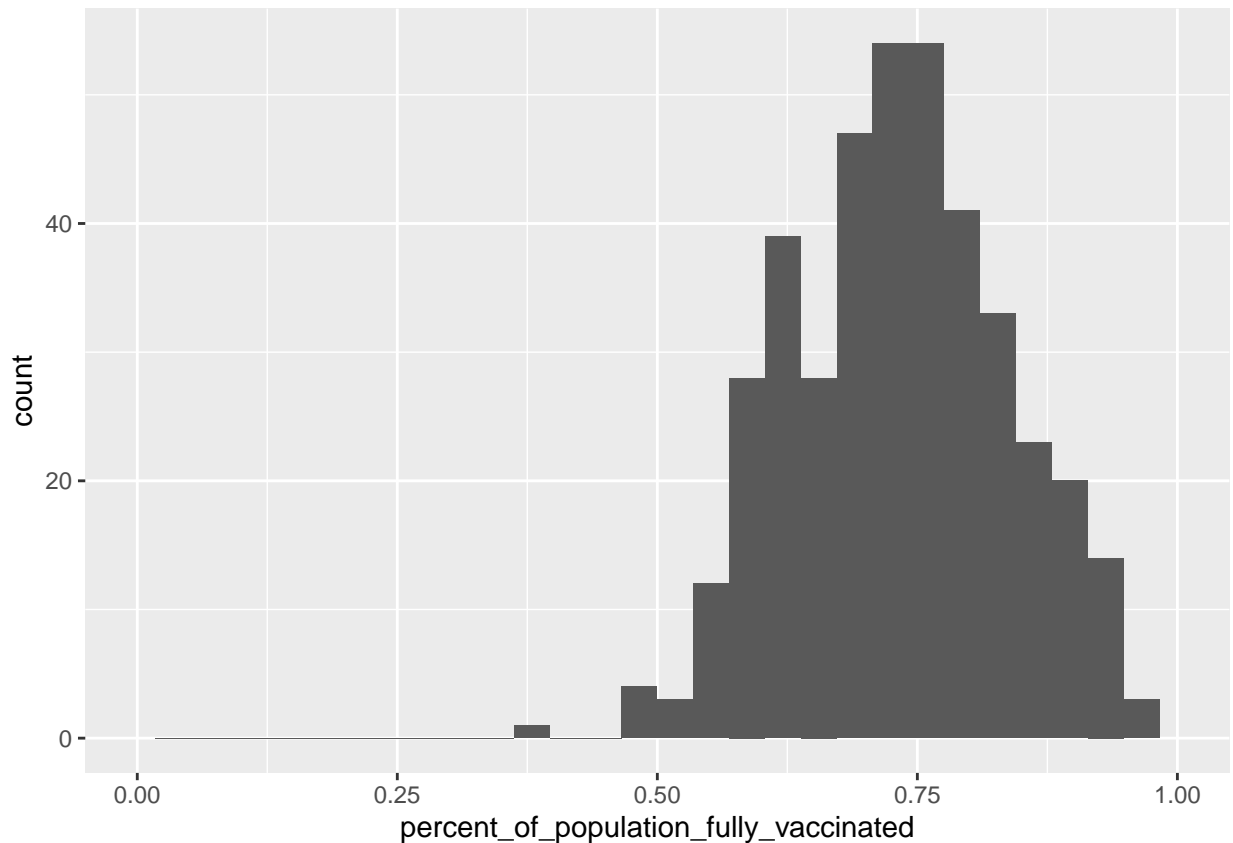
```
## 3rd Qu.:53932      3rd Qu.: 59168      3rd Qu.:43494
## Max. :88557      Max. :101902      Max. :77744
## persons_partially_vaccinated percent_of_population_fully_vaccinated
## Min. : 1753      Min. :0.3890
## 1st Qu.: 2846      1st Qu.:0.6554
## Median : 3671      Median :0.7350
## Mean : 4564      Mean :0.7354
## 3rd Qu.: 5186      3rd Qu.:0.8044
## Max. :34219      Max. :1.0000
## percent_of_population_partially_vaccinated
## Min. :0.04001
## 1st Qu.:0.06016
## Median :0.07024
## Mean :0.08775
## 3rd Qu.:0.08947
## Max. :0.92817
## percent_of_population_with_1_plus_dose booster_recip_count redacted
## Min. :0.5006      Min. : 5062      Length:411
## 1st Qu.:0.7414      1st Qu.:13600      Class :character
## Median :0.8194      Median :17595      Mode :character
## Mean :0.8155      Mean :18800
## 3rd Qu.:0.8899      3rd Qu.:23151
## Max. :1.0000      Max. :50602
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36) +
  aes(percent_of_population_fully_vaccinated) +
  geom_histogram() +
  xlim(c(0,1))
```

```
## `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? The 92109 and 92040 zip code areas are both below the calculated average value.

```
vax %>% filter(as_of_date == "2022-03-01") %>%
  filter(zip_code_tabulation_area=="92109") %>%
  select(percent_of_population_fully_vaccinated)
```

```
## percent_of_population_fully_vaccinated
## 1 0.723778
```

```
vax %>% filter(as_of_date == "2022-03-01") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

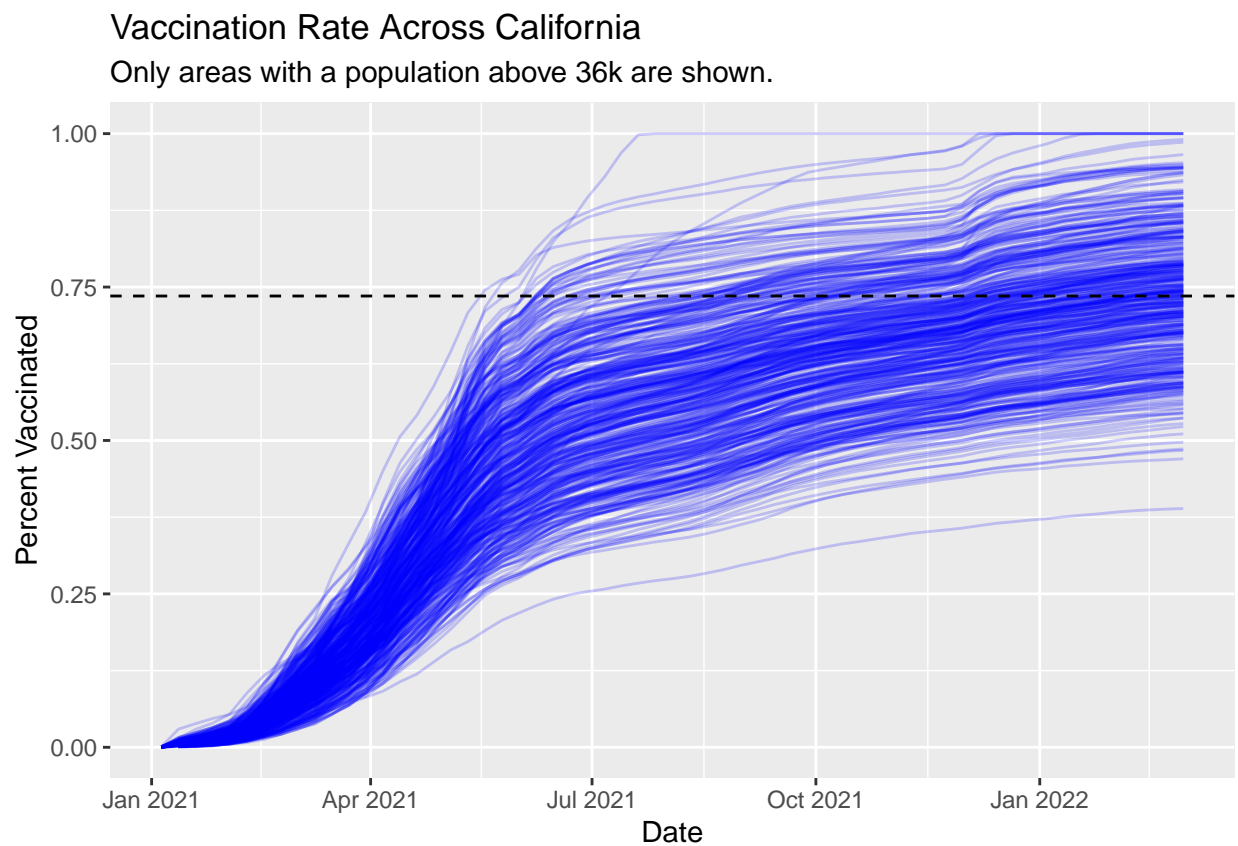
```
## percent_of_population_fully_vaccinated
## 1 0.551981
```

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)
```

```
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 = ave.36, linetype=2)
```

Warning: Removed 311 row(s) containing missing values (geom_path).



Q21. How do you feel about traveling for Spring Break and meeting for in-person class afterwards?
It might be okay, but it is still safer to stay at home.