Class 17 COVID-19 Vaccination Rates Mini-Project

Bangyan Hu (PID: A15540189)

11/23/2021

#Getting Started

```
# 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                   92804                    Orange    Orange
## 2 2021-01-05                   92626                    Orange    Orange
## 3 2021-01-05                   92250                  Imperial  Imperial
## 4 2021-01-05                   92637                    Orange    Orange
## 5 2021-01-05                   92155                 San Diego San Diego
## 6 2021-01-05                   92259                  Imperial  Imperial
##   vaccine_equity_metric_quartile                vem_source
## 1                              2 Healthy Places Index Score
## 2                              3 Healthy Places Index Score
## 3                              1 Healthy Places Index Score
## 4                              3 Healthy Places Index Score
## 5                             NA            No VEM Assigned
## 6                              1    CDPH-Derived ZCTA Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               76455.9                84200                       19
## 2               44238.8                47883                       NA
## 3                7098.5                 8026                       NA
## 4               16027.4                16053                       NA
## 5                 456.0                  456                       NA
## 6                 119.0                  121                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         1282                                0.000226
## 2                           NA                                      NA
## 3                           NA                                      NA
## 4                           NA                                      NA
## 5                           NA                                      NA
## 6                           NA                                      NA
##   percent_of_population_partially_vaccinated
## 1                                   0.015226
## 2                                         NA
## 3                                         NA
## 4                                         NA
## 5                                         NA
## 6                                         NA
```

```
##   percent_of_population_with_1_plus_dose
## 1                                0.015452
## 2                                      NA
## 3                                      NA
## 4                                      NA
## 5                                      NA
## 6                                      NA
##                                                                    redacted
## 1                                                                        No
## 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
```

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

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

persons fully vaccinated (persons_fully_vaccinated).

Q2. What column details the Zip code tabulation area?

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?

2021-11-16

Q5. How many numeric columns are in this dataset?

There are 9 numeric columns are in this dataset.

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] 8256
```

There are 8256 "missing values" for persons_fully_vaccinated in the dataset.

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

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

```
## [1] 0.101745
```

10.% are missing.

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

The data is missing might be that people are still not fully vaccinated yet.

#Working with dates

```
library(lubridate)
```

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

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

```
today()
```

```
## [1] "2021-11-29"
```

```
# This will give an Error!
#today() - vax$as_of_date[1]
```

```
# Speciffy that we are using the Year-mont-day format
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 315 days
```

Q. How many days since the last entry?

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

```
## Time difference of 13 days
```

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

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

```
## Time difference of 13 days
```

7 days have passed since the last update of the dataset (on Nov. 23). 13 days have passed since the last update of the dataset (on Nov. 29).

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

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

```
## [1] 46
```

There are 46 unique dates in the dataset.

#Working with ZIP codes

```
library(zipcodeR)
```

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

```
## # A tibble: 1 x 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 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>
```

```
# Pull data for all ZIP codes in the dataset
#zipdata <- reverse_zipcode( vax$zip_code_tabulation_area )
```

#Focus on the San Diego area

```
# 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] 4922
```

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

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?

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

```
##    as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 23 2021-01-05                   92154                San Diego San Diego
##    vaccine_equity_metric_quartile                vem_source
## 23                              2 Healthy Places Index Score
##    age12_plus_population age5_plus_population persons_fully_vaccinated
## 23               76365.2                82971                        32
##    persons_partially_vaccinated percent_of_population_fully_vaccinated
## 23                         1336                                0.000386
##    percent_of_population_partially_vaccinated
## 23                                   0.016102
##    percent_of_population_with_1_plus_dose redacted
## 23                                0.016488       No
```

The zip code area 92154 in San Diego has the largest 12+ population in this dataset

What is the population in the 92037 ZIP code area?

```
filter(sd, zip_code_tabulation_area == "92037") [1,]
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-01-05                   92037                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 1                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               33675.6                36144                        44
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         1265                                0.001217
##   percent_of_population_partially_vaccinated
## 1                                   0.034999
##   percent_of_population_with_1_plus_dose redacted
## 1                                0.036216       No
```

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

```
sd.now <- filter(sd, as_of_date == "2021-11-09")

head(sd.now)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2021-11-09                   92075                San Diego San Diego
## 2 2021-11-09                   92130                San Diego San Diego
## 3 2021-11-09                   92060                San Diego San Diego
## 4 2021-11-09                   92091                San Diego San Diego
## 5 2021-11-09                   92020                San Diego San Diego
## 6 2021-11-09                   92004                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 1                              4 Healthy Places Index Score
## 2                              4 Healthy Places Index Score
## 3                              3    CDPH-Derived ZCTA Score
## 4                              4    CDPH-Derived ZCTA Score
## 5                              2 Healthy Places Index Score
## 6                              2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               11136.3                12177                     9504
## 2               46300.3                53102                    45517
## 3                 166.0                  166                      153
## 4                1238.3                 1303                     1159
## 5               49284.5                54991                    34904
## 6                2151.8                 2186                     2582
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                         1623                                0.780488
## 2                         6642                                0.857162
## 3                           34                                0.921687
## 4                          221                                0.889486
## 5                         4688                                0.634722
## 6                          514                                1.000000
##   percent_of_population_partially_vaccinated
## 1                                   0.133284
## 2                                   0.125080
## 3                                   0.204819
## 4                                   0.169609
## 5                                   0.085250
## 6                                   0.235133
##   percent_of_population_with_1_plus_dose redacted
## 1                               0.913772       No
## 2                               0.982242       No
## 3                               1.000000       No
## 4                               1.000000       No
## 5                               0.719972       No
## 6                               1.000000       No
```

```
mean(sd.now$percent_of_population_fully_vaccinated, na.rm=TRUE)
```

```
## [1] 0.6727567
```

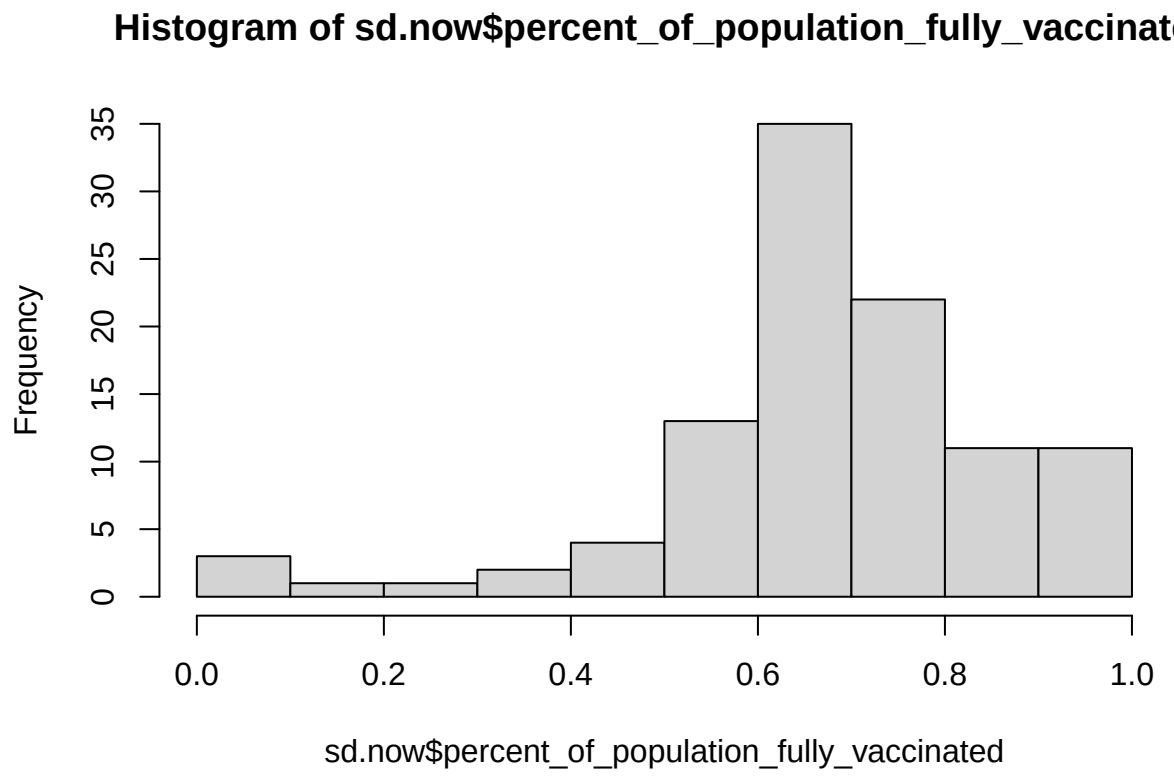
```
summary( sd.now$percent_of_population_fully_vaccinated )
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.01017 0.60776 0.67700 0.67276 0.76164 1.00000       4
```

The overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2021-11-09" is 67.27567% (0.6727567).

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"?

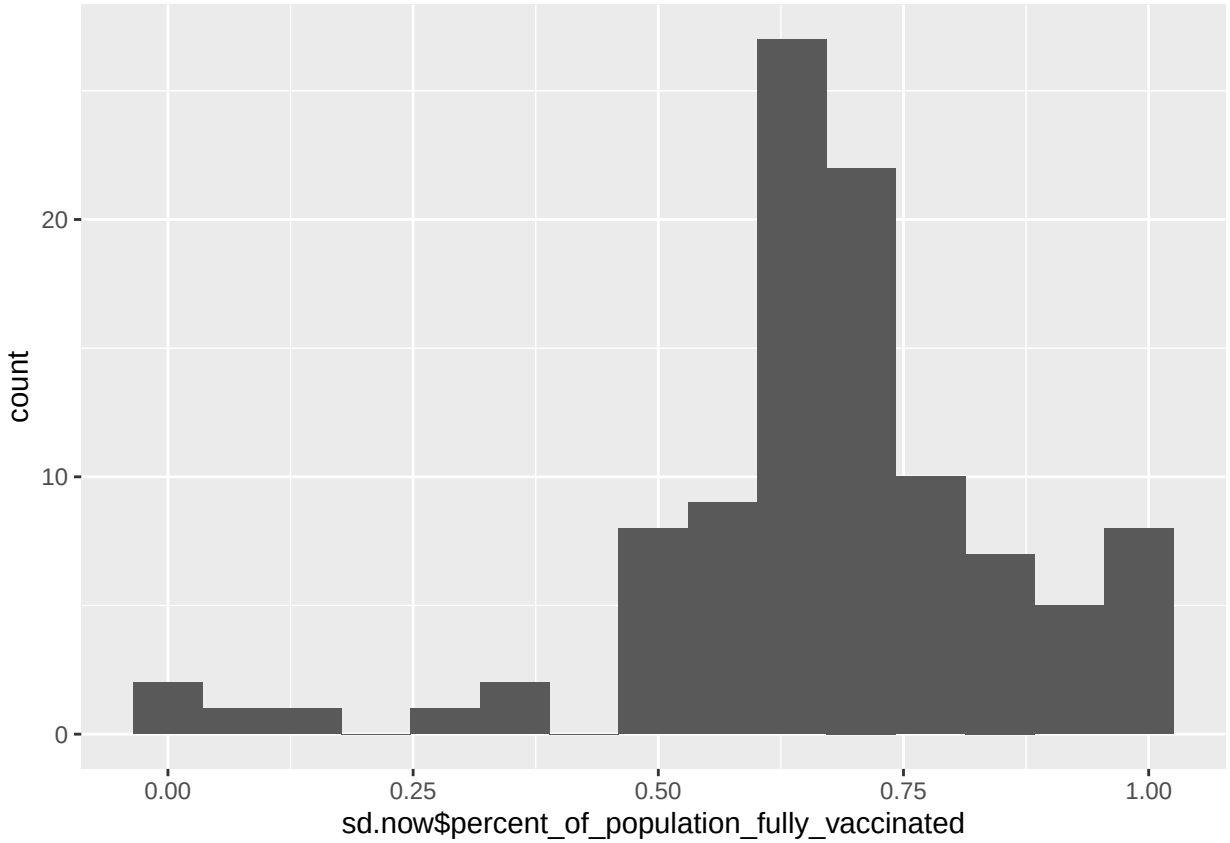
```
hist(sd.now$percent_of_population_fully_vaccinated)
```




```
library(ggplot2)
ggplot(sd.now) + aes(sd.now$percent_of_population_fully_vaccinated) + geom_histogram(bins=15)
```

```
## Warning: Use of `sd.now$percent_of_population_fully_vaccinated` is discouraged.
## Use `percent_of_population_fully_vaccinated` instead.
```

```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```



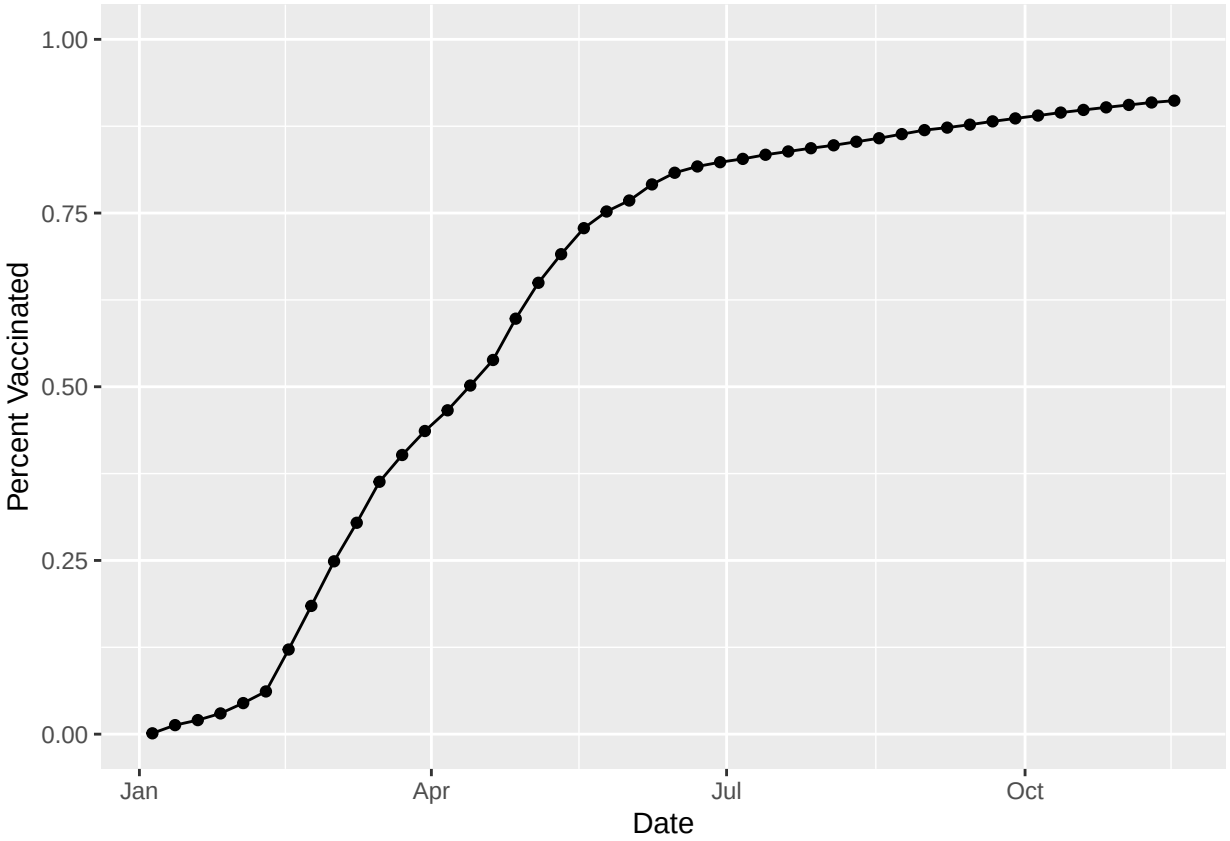
#Focus on UCSD/La Jolla

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

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

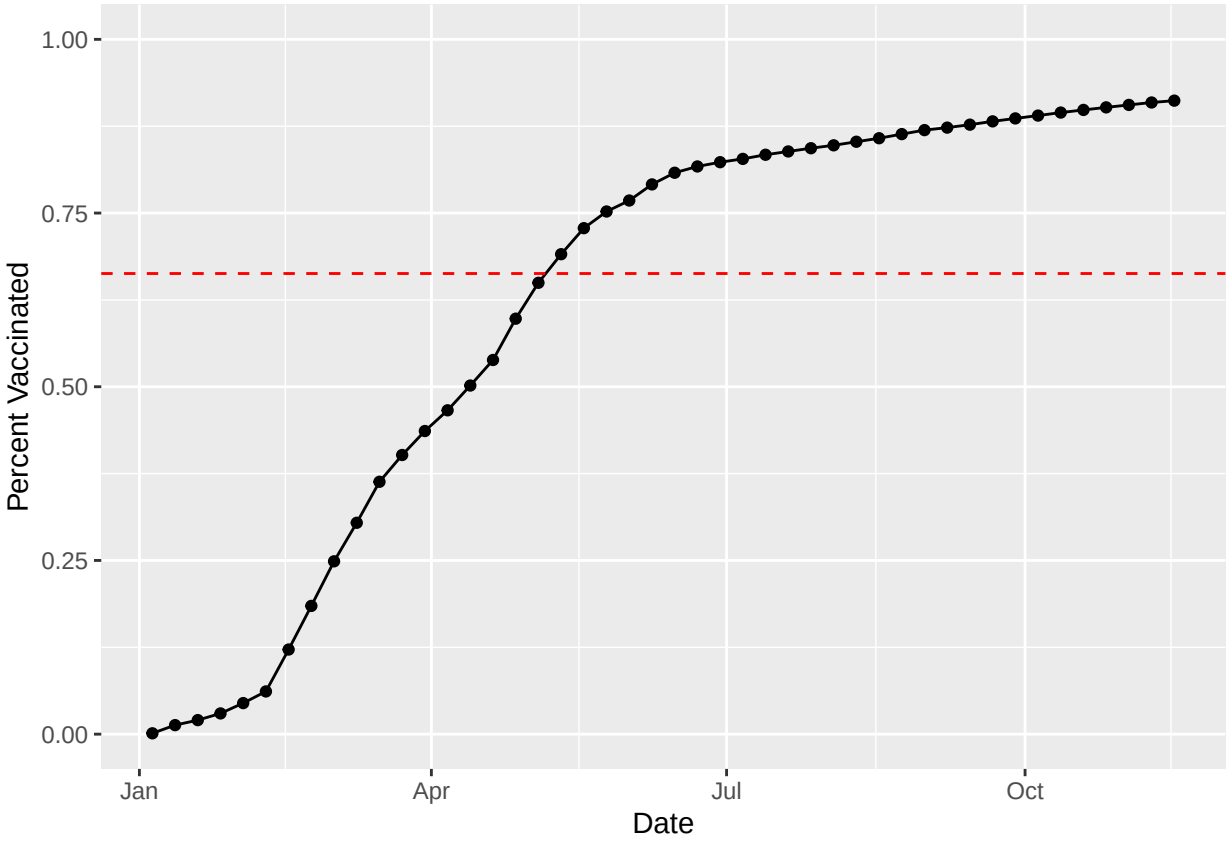
Q15. Using ggplot make a graph of the vaccination rate time course for the 92037 ZIP code area:

```
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")
```



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?

```
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") +
  geom_hline(yintercept = 0.6629812, linetype = "dashed", color = "red")
```



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"?

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

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

```
# 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")

summary(vax.36$percent_of_population_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3519  0.5891  0.6649  0.6630  0.7286  1.0000
```

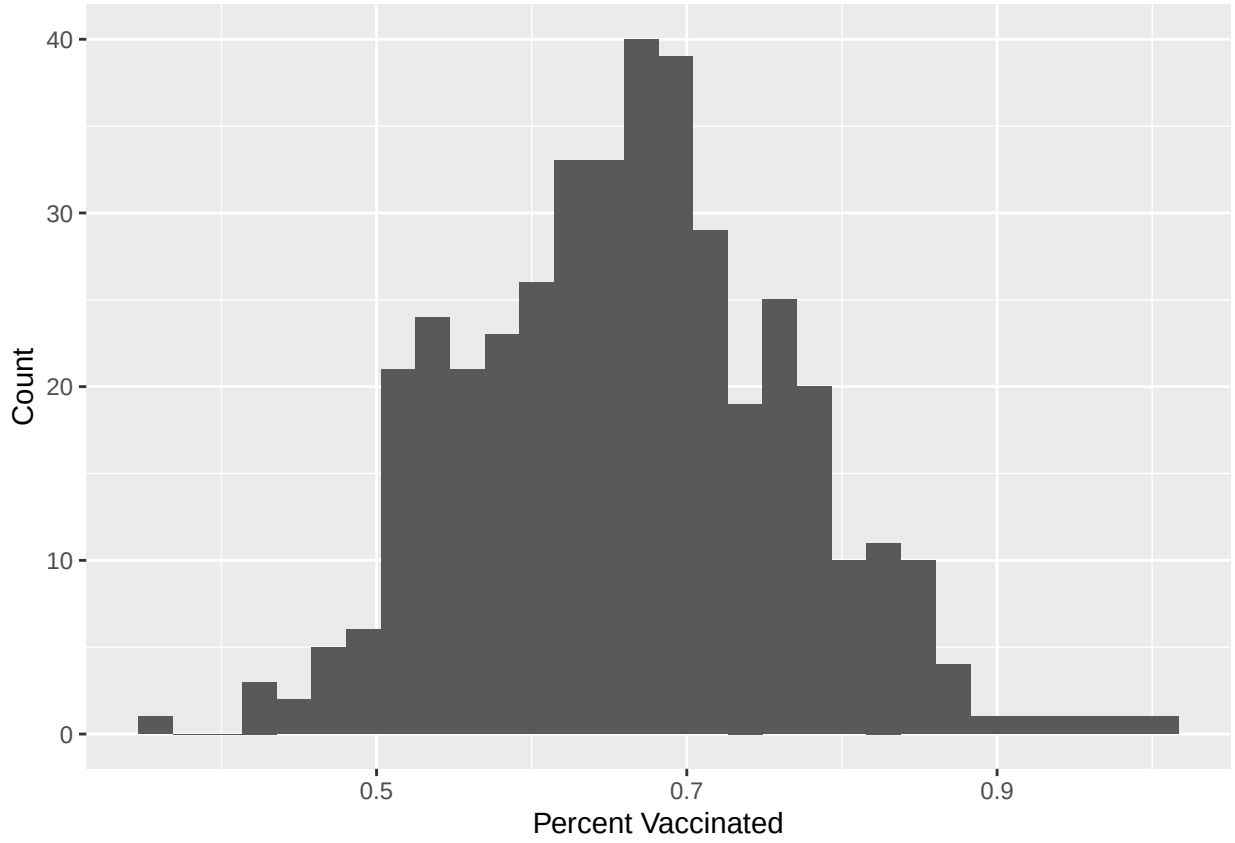
Min. 1st Qu. Median Mean 3rd Qu. Max. 0.3519 0.5891 0.6649 0.6630 0.7286 1.0000

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36) + aes(vax.36$percent_of_population_fully_vaccinated) +
  geom_histogram() + labs(x = "Percent Vaccinated", y="Count")
```

```
## Warning: Use of ‘vax.36$percent_of_population_fully_vaccinated‘ is discouraged.
## Use ‘percent_of_population_fully_vaccinated‘ instead.
```

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



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                               0.520463
```

```
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                               0.687763
```

```
mean(vax.36$percent_of_population_fully_vaccinated)
```

```
## [1] 0.6629812
```

Thus, the 92040 ZIP code area is below the average value you calculated for all these above, and the the 92109 ZIP code area is above the average value you calculated for all these above

Q20. Finally make a time course plot of vaccination progress for all areas in the full dataset with a age5_plus_population > 36144.

First, we need to subset the full "vax" dataset to include only ZIP code areas with a population as large as 92037.

```
vax.36.all <- filter(vax, age5_plus_population > 36144)
nrow(vax.36.all)
```

```
## [1] 18906
```

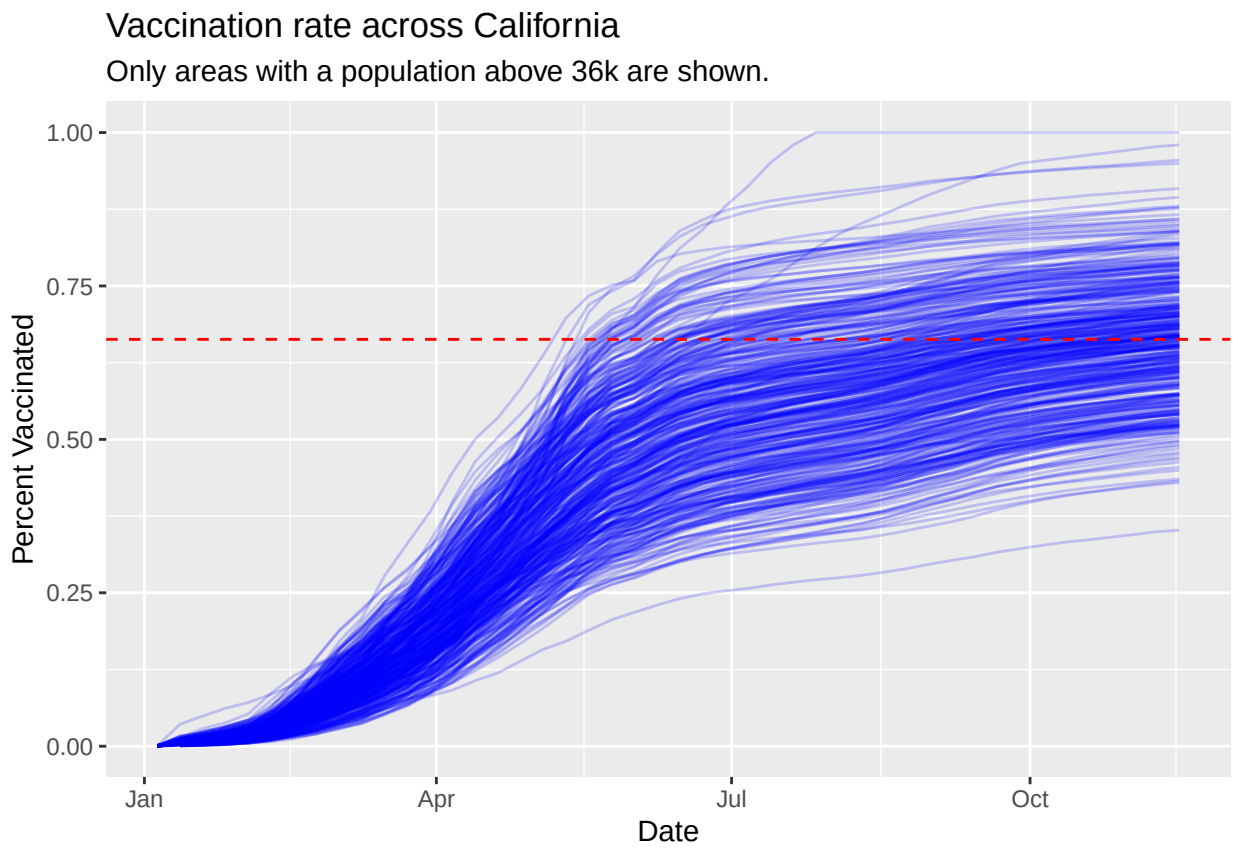
```
length(unique(vax.36.all$zip_code_tabulation_area))
```

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

Thus, let's make a final figure that shows all these ZIP areas

```
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 = 0.6629812, linetype = "dashed", color = "red")
```

```
## Warning: Removed 180 row(s) containing missing values (geom_path).
```



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

Great.

#About this document

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.3.5   dplyr_1.0.7     zipcodeR_0.3.3  lubridate_1.8.0
## [5] skimr_2.1.3
##
## loaded via a namespace (and not attached):
```

```
##  [1] httr_1.4.2         tidyr_1.1.4        bit64_4.0.5        jsonlite_1.7.2
##  [5] assertthat_0.2.1   sp_1.4-6           highr_0.9          blob_1.2.2
##  [9] yaml_2.2.1         tidycensus_1.1     pillar_1.6.4       RSQLite_2.2.8
## [13] lattice_0.20-45    glue_1.5.0         uuid_1.0-3         digest_0.6.28
## [17] rvest_1.0.2        colorspace_2.0-2   htmltools_0.5.2    pkgconfig_2.0.3
## [21] raster_3.5-2       purrr_0.3.4        scales_1.1.1       terra_1.4-22
## [25] tzdb_0.2.0         tigris_1.5         tibble_3.1.6       proxy_0.4-26
## [29] farver_2.1.0       generics_0.1.1     ellipsis_0.3.2     withr_2.4.2
## [33] cachem_1.0.6       repr_1.1.3         cli_3.1.0          magrittr_2.0.1
## [37] crayon_1.4.2       memoise_2.0.1      maptools_1.1-2     evaluate_0.14
## [41] fansi_0.5.0        xml2_1.3.2         foreign_0.8-81     class_7.3-19
## [45] tools_4.1.2        hms_1.1.1          lifecycle_1.0.1    stringr_1.4.0
## [49] munsell_0.5.0      compiler_4.1.2     e1071_1.7-9        rlang_0.4.12
## [53] classInt_0.4-3     units_0.7-2        grid_4.1.2         rappdirs_0.3.3
## [57] labeling_0.4.2     base64enc_0.1-3    rmarkdown_2.11     gtable_0.3.0
## [61] codetools_0.2-18   DBI_1.1.1          curl_4.3.2         R6_2.5.1
## [65] knitr_1.36         rgdal_1.5-27       fastmap_1.1.0      bit_4.0.4
## [69] utf8_1.2.2         KernSmooth_2.23-20 readr_2.1.0        stringi_1.7.5
## [73] Rcpp_1.0.7         vctrs_0.3.8        sf_1.0-4           tidyselect_1.1.1
## [77] xfun_0.28
```