

# miniproject14\_Sasse.R

Noah

2022-03-07

```
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
```

```
#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
```

```
#Q5: How many numeric columns are in this dataset?
#9 columns
```

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

```
## [1] 18338
```

```
#Q6: Note that there are "missing values" in the dataset.
#How many NA values there in the persons_fully_vaccinated column?
#18338 NAs
```

```
#Q7: What percent of persons_fully_vaccinated values are missing (to 2 significant figures)?
#17 percent
```

```
#Q9: How many days have passed since the last update of the dataset?
#5 days
```

```
#Q10: How many unique dates are in the dataset (i.e. how many different dates are detailed)?
#420
```

```
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>
```

```
zipdata <- reverse_zipcode( vax$zip_code_tabulation_area )
```

```
sd <- vax[vax$county == "San Diego", ]
sd.10 <- filter(vax, vax$county == "San Diego" &
                vax$page5_plus_population > 10000)
```

```
#Q11: How many distinct zip codes are listed for San Diego County?
#107 counties
```

```
#Q12: What San Diego County Zip code area has the largest 12 + Population in this dataset?
#92154
```

```
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
```

```
Q13 <- filter(vax, as_of_date == "2022-02-22")
mean(Q13$percent_of_population_fully_vaccinated, na.rm = TRUE)
```

```
## [1] 0.6972217
```

```
#Q13: What is the overall average "Percent of Population Fully Vaccinated"
#value for all San Diego "County" as of "2022-02-22"?
#69.7%
```

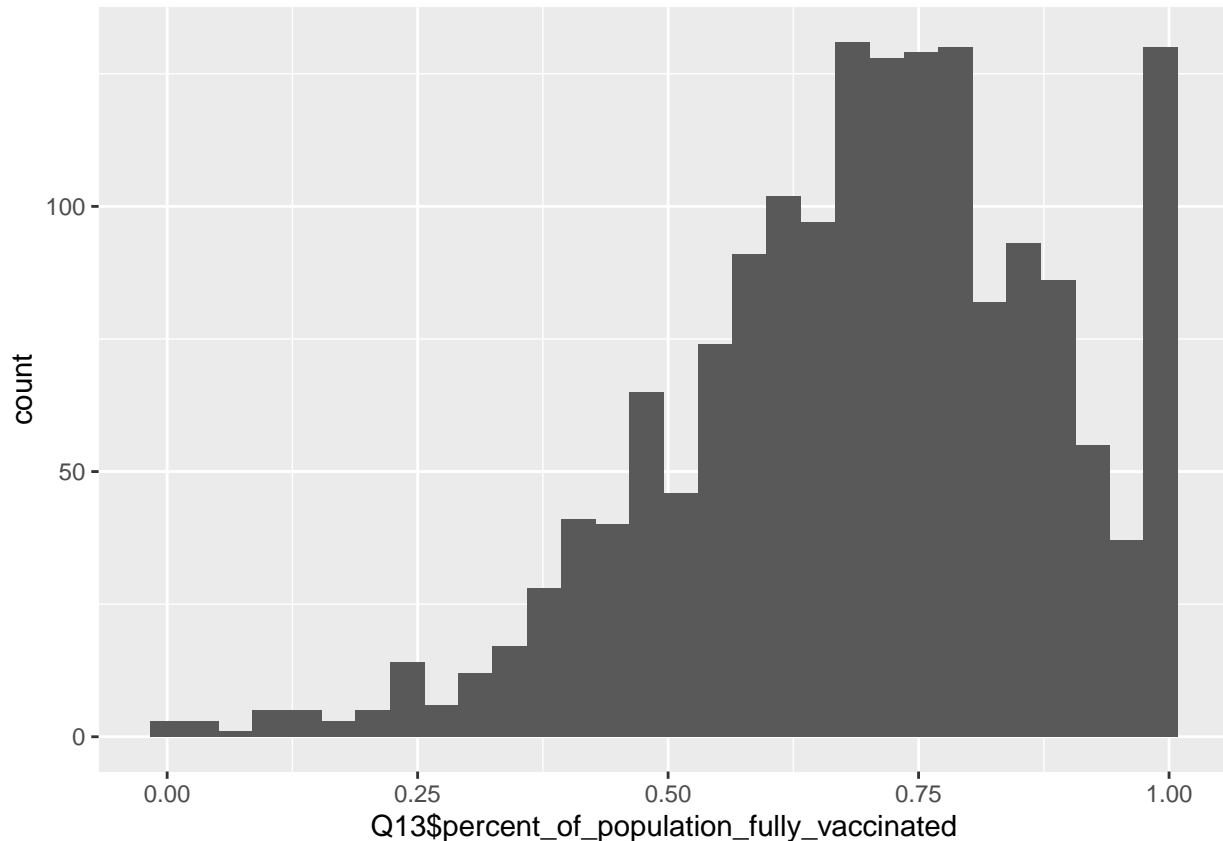
```
#Q14: Using either ggplot or base R graphics make a summary figure that shows the distribution
#of Percent of Population Fully Vaccinated values as of "2022-02-22"?
```

```
library(ggplot2)
ggplot(Q13, aes(Q13$percent_of_population_fully_vaccinated)) +
  geom_histogram()
```

```
## Warning: Use of 'Q13$percent_of_population_fully_vaccinated' is discouraged. Use
## 'percent_of_population_fully_vaccinated' instead.
```

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

```
## Warning: Removed 105 rows containing non-finite values (stat_bin).
```



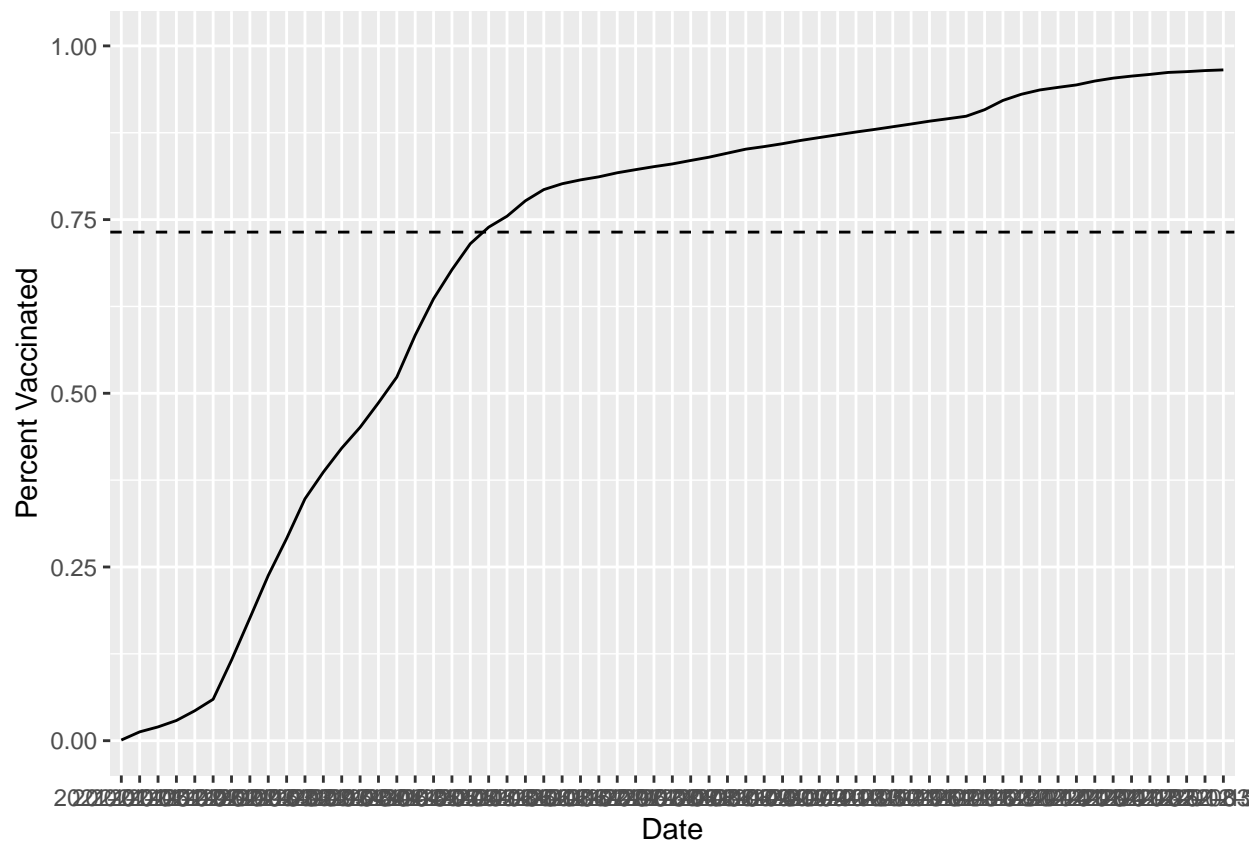
```
#Q15 and 16:
```

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

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

```
ggplot(ucsd) +
  aes(as_of_date, percent_of_population_fully_vaccinated) +
  geom_line() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x = "Date", y="Percent Vaccinated") +
  geom_hline(yintercept = 0.732, linetype= 'dashed')
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



#16-18:

```
vax.36 <- filter(vax, age5_plus_population > 36144 &
  as_of_date == "2022-02-22")
head(vax.36)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction   county
## 1 2022-02-22          92840                Orange      Orange
## 2 2022-02-22          92064              San Diego    San Diego
## 3 2022-02-22          92508              Riverside    Riverside
## 4 2022-02-22          95403                Sonoma      Sonoma
## 5 2022-02-22          90001            Los Angeles    Los Angeles
## 6 2022-02-22          92802                Orange      Orange
```

```
## vaccine_equity_metric_quartile vem_source
## 1 2 Healthy Places Index Score
## 2 4 Healthy Places Index Score
## 3 3 Healthy Places Index Score
## 4 3 Healthy Places Index Score
## 5 1 Healthy Places Index Score
## 6 2 Healthy Places Index Score
## age12_plus_population age5_plus_population persons_fully_vaccinated
## 1 47302.5 51902 40725
## 2 42177.1 46855 34266
## 3 32415.3 36303 21925
## 4 38545.9 42294 33158
## 5 47175.7 54805 43075
## 6 35113.6 39393 29268
## persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1 4324 0.784652
## 2 6861 0.731320
## 3 1714 0.603945
## 4 2833 0.783988
## 5 13917 0.785968
## 6 6138 0.742975
## percent_of_population_partially_vaccinated
## 1 0.083311
## 2 0.146430
## 3 0.047214
## 4 0.066983
## 5 0.253937
## 6 0.155814
## percent_of_population_with_1_plus_dose booster_recip_count redacted
## 1 0.867963 20654 No
## 2 0.877750 15499 No
## 3 0.651159 10753 No
## 4 0.850971 18659 No
## 5 1.000000 13408 No
## 6 0.898789 12816 No
```

*#Q19*

```
vax %>% filter(as_of_date == "2022-02-22") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
## percent_of_population_fully_vaccinated
## 1 0.551304
```

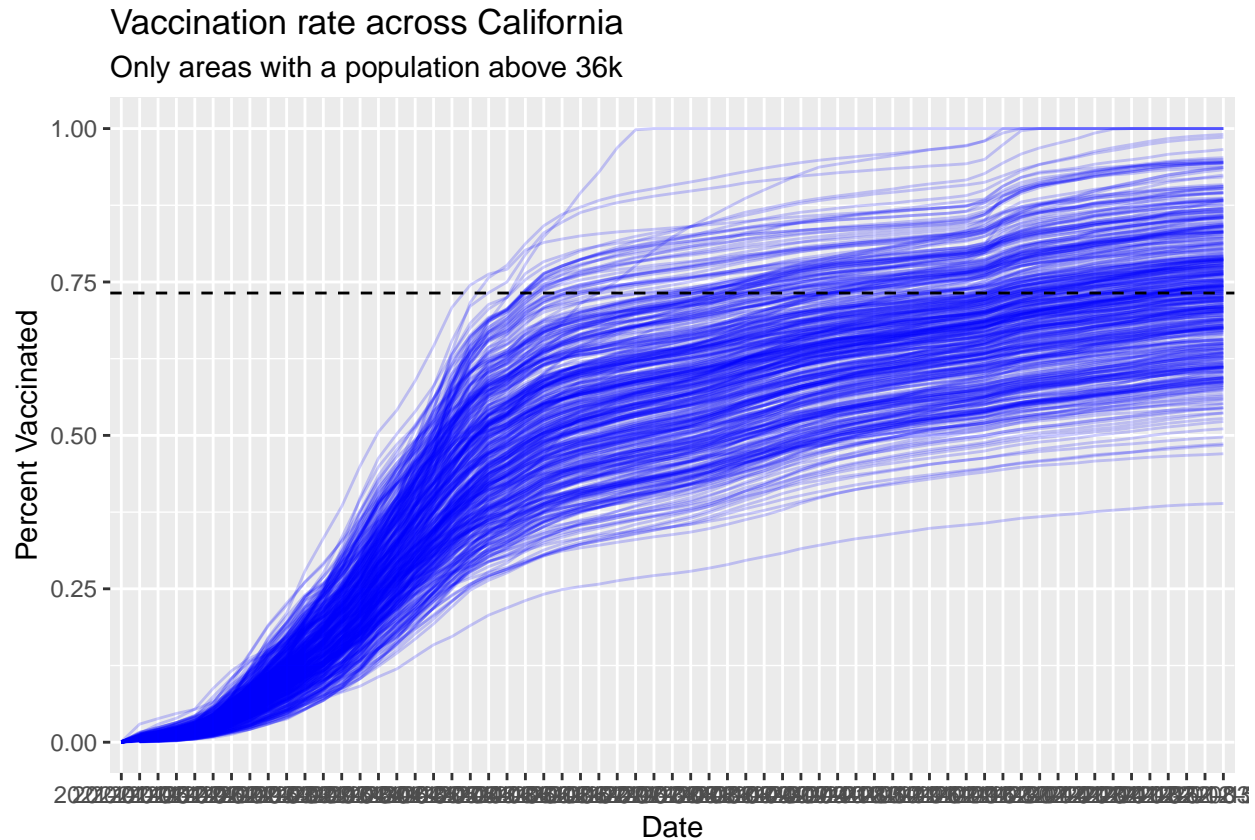
```
vax.36.all <- filter(vax, vax$age5_plus_population>36144)
```

*#Q20:*

```
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(0,1) +
```

```
labs(x="Date", y="Percent Vaccinated",
      title="Vaccination rate across California",
      subtitle="Only areas with a population above 36k") +
geom_hline(yintercept = 0.732, linetype= 'dashed')
```

```
## 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?*  
*#Comfortable*

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
```

```
##
## attached base packages:
## [1] stats      graphics  grDevices utils      datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.3.5  dplyr_1.0.8    zipcodeR_0.3.3
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.8      lattice_0.20-45  tidyr_1.2.0      class_7.3-20
## [5] digest_0.6.29   utf8_1.2.2       R6_2.5.1         RSQLite_2.2.10
## [9] evaluate_0.15   e1071_1.7-9      httr_1.4.2       highr_0.9
## [13] pillar_1.7.0    rlang_1.0.2      curl_4.3.2       uuid_1.0-3
## [17] rstudioapi_0.13 raster_3.5-15     blob_1.2.2       rmarkdown_2.12
## [21] labeling_0.4.2  rgdal_1.5-28     readr_2.1.2      stringr_1.4.0
## [25] foreign_0.8-82  munsell_0.5.0    bit_4.0.4        proxy_0.4-26
## [29] compiler_4.1.2  xfun_0.30        pkgconfig_2.0.3  tigris_1.6
## [33] htmltools_0.5.2 tidyselect_1.1.2 tibble_3.1.6     codetools_0.2-18
## [37] fansi_1.0.2     withr_2.5.0      crayon_1.5.0     tzdb_0.2.0
## [41] sf_1.0-6        tidycensus_1.1   rappdirs_0.3.3   grid_4.1.2
## [45] gtable_0.3.0    jsonlite_1.8.0   lifecycle_1.0.1  DBI_1.1.2
## [49] magrittr_2.0.2  scales_1.1.1     units_0.8-0      KernSmooth_2.23-20
## [53] cli_3.1.1       stringi_1.7.6    cachem_1.0.6     farver_2.1.0
## [57] sp_1.4-6        xml2_1.3.3       ellipsis_0.3.2   generics_0.1.2
## [61] vctrs_0.3.8     tools_4.1.2      bit64_4.0.5      glue_1.6.2
## [65] purrr_0.3.4     hms_1.1.1        fastmap_1.1.0    yaml_2.3.5
## [69] colorspace_2.0-3 terra_1.5-21     maptools_1.1-2   classInt_0.4-3
## [73] rvest_1.0.2     memoise_2.0.1    knitr_1.37
```