Lab 17

Bomin Xie

Getting started

```
# Import vaccination data
vax <- read.csv("https://marcos-diazg.github.io/BIMM143_SP23/class-material/class17/covid1
head(vax)
```

```
  as_of_date zip_code_tabulation_area local_health_jurisdiction      county
1 2021-01-05                   94579                   Alameda     Alameda
2 2021-01-05                   93726                    Fresno      Fresno
3 2021-01-05                   94305               Santa Clara Santa Clara
4 2021-01-05                   93704                    Fresno      Fresno
5 2021-01-05                   94403                 San Mateo   San Mateo
6 2021-01-05                   93668                    Fresno      Fresno
  vaccine_equity_metric_quartile                vem_source
1                              3 Healthy Places Index Score
2                              1 Healthy Places Index Score
3                              4 Healthy Places Index Score
4                              1 Healthy Places Index Score
5                              4 Healthy Places Index Score
6                              1    CDPH-Derived ZCTA Score
  age12_plus_population age5_plus_population tot_population
1               19192.7                20872          21883
2               33707.7                39067          42824
3               15716.9                16015          16397
4               24803.5                27701          29740
5               37967.5                41530          44408
6                1013.4                 1199           1219
  persons_fully_vaccinated persons_partially_vaccinated
1                       NA                           NA
2                       NA                           NA
```

```
3                      NA                            NA
4                      NA                            NA
5                      NA                            NA
6                      NA                            NA
  percent_of_population_fully_vaccinated
1                                     NA
2                                     NA
3                                     NA
4                                     NA
5                                     NA
6                                     NA
  percent_of_population_partially_vaccinated
1                                         NA
2                                         NA
3                                         NA
4                                         NA
5                                         NA
6                                         NA
  percent_of_population_with_1_plus_dose booster_recip_count
1                                     NA                  NA
2                                     NA                  NA
3                                     NA                  NA
4                                     NA                  NA
5                                     NA                  NA
6                                     NA                  NA
  bivalent_dose_recip_count eligible_recipient_count
1                        NA                        4
2                        NA                        2
3                        NA                        8
4                        NA                        5
5                        NA                        7
6                        NA                        0
  eligible_bivalent_recipient_count
1                                 4
2                                 2
3                                 8
4                                 5
5                                 7
6                                 0
                                                                    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: The column details the total number of people fully vaccinated are “persons_fully_vaccinatted”.

Q2: The column details the Zip code tabulation area is “zip_code_tabulation_area”.

Q3: The earliest date in this dataset is “2021-01-05”.

Q4: The latest date in this dataset is “2023-05-23”.

```
# install.packages("skimr")
library(skimr)
skimr::skim_without_charts(vax)
```

Table 1: Data summary

Name	vax
Number of rows	220500
Number of columns	19
_____	
Column type frequency:	
character	5
numeric	14
_____	
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	125	0
local_health_jurisdiction	0	1	0	15	625	62	0
county	0	1	0	15	625	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
zip_code_tabulation_area	0	1.00	93665.11	817.38	90001	92257.75	93658.50	95380.50	97635.0
vaccine_equity_metric_quartile	10875	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0
age12_plus_population	0	1.00	18895.04	18993.87	0	1346.95	13685.10	31756.18	88556.7
age5_plus_population	0	1.00	20875.22	21105.97	0	1460.50	15364.00	34877.00	101902.0
tot_population	10750	0.95	23372.77	22628.50	12	2126.00	18714.00	38168.00	111165.0
persons_fully_vaccinated	17711	0.92	14272.72	15264.17	11	954.00	8990.00	23782.00	87724.0
persons_partially_vaccinated	17711	0.92	1711.05	2071.56	11	164.00	1203.00	2550.00	42259.0
percent_of_population_fully_vaccinated	22579	0.90	0.58	0.25	0	0.44	0.62	0.75	1.0
percent_of_population_partially_vaccinated	22579	0.90	0.08	0.09	0	0.05	0.06	0.08	1.0
percent_of_population_with_1_plus_dose	23732	0.89	0.64	0.24	0	0.50	0.68	0.82	1.0
booster_recip_count	74388	0.66	6373.43	7751.70	11	328.00	3097.00	10274.00	60022.0
bivalent_dose_recip_count	159956	0.27	3407.91	4010.38	11	222.00	1832.00	5482.00	29484.0
eligible_recipient_count	0	1.00	13120.40	15126.17	0	534.00	6663.00	22517.25	87437.0
eligible_bivalent_recipient_count	0	1.00	13016.51	15199.08	0	266.00	6562.00	22513.00	87437.0

Q5: In this dataset, there are 14 numeric columns in this dataset.

Q6: There are 17711 "missing values" in the "persons_fully_vaccinated" column.

Q7: Based on the skimr result, there are 8.04 percent of "persons_fully_vaccinated" values are missing.

Working with dates

```
# install.packages("lubridate")
library(lubridate)
```

```
Attaching package: 'lubridate'

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

    date, intersect, setdiff, union
```

```
today()
```

```
[1] "2023-06-13"
```

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

```
Time difference of 21 days
```

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

```
[1] 125
```

Q9: It has been 21 days passed since the last update.

Q10: There are 125 unique dates in the dataset.

Working with ZIP codes

```
# install.packages("zipcodeR")
library(zipcodeR)
```

```
The legacy packages maptools, rgdal, and rgeos, underpinning this package
will retire shortly. Please refer to R-spatial evolution reports on
https://r-spatial.org/r/2023/05/15/evolution4.html for details.
This package is now running under evolution status 0
```

```
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
```

Focus on San Diego area

```
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] 13375
```

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

```
length(unique(sd$zip_code_tabulation_area))
```

```
[1] 107
```

```
sd$zip_code_tabulation_area[which.max(unique(sd$tot_population))]
```

```
[1] 92154
```

Q11: There are 107 distinct zip codes.

Q12: The zip code area with largest population in this dataset is 92154.

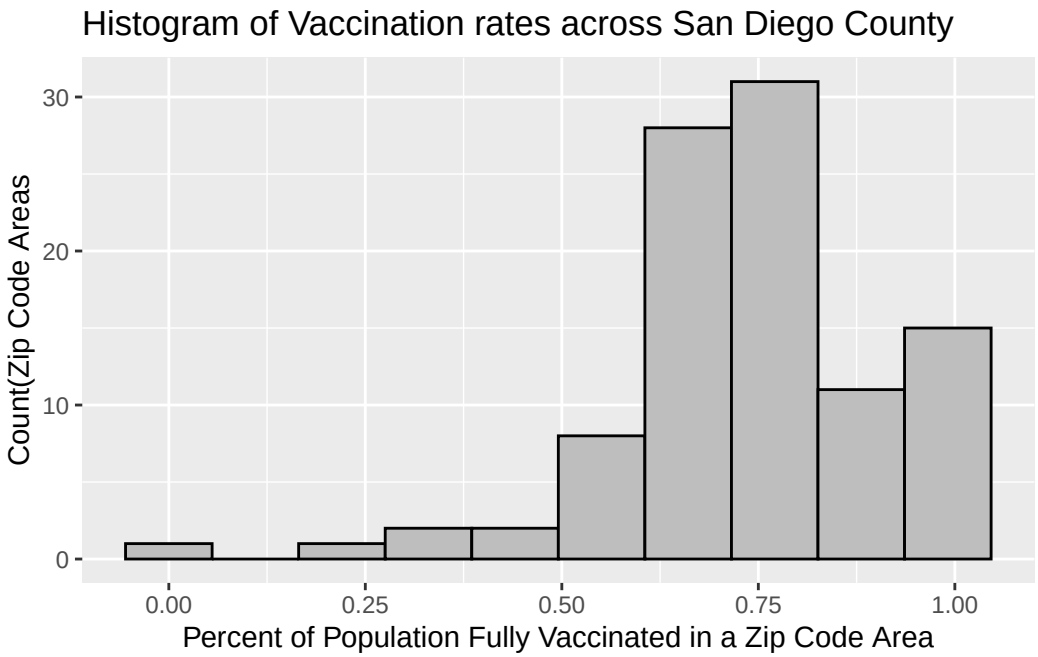
```
avg_percent <- filter(sd, as_of_date == "2023-02-28")
mean(avg_percent$percent_of_population_fully_vaccinated, na.rm = TRUE) * 100
```

```
[1] 74.1269
```

Q13: The overall average "percent of population fully vaccinated value" is 74.13.

Q14:

```
library(ggplot2)
ggplot(avg_percent, aes(percent_of_population_fully_vaccinated)) +
  geom_histogram(bins = 10, na.rm = TRUE, color= "black", fill = "grey") +
  ggtitle("Histogram of Vaccination rates across San Diego County") +
  xlab("Percent of Population Fully Vaccinated in a Zip Code Area") + ylab("Count(Zip Code
```



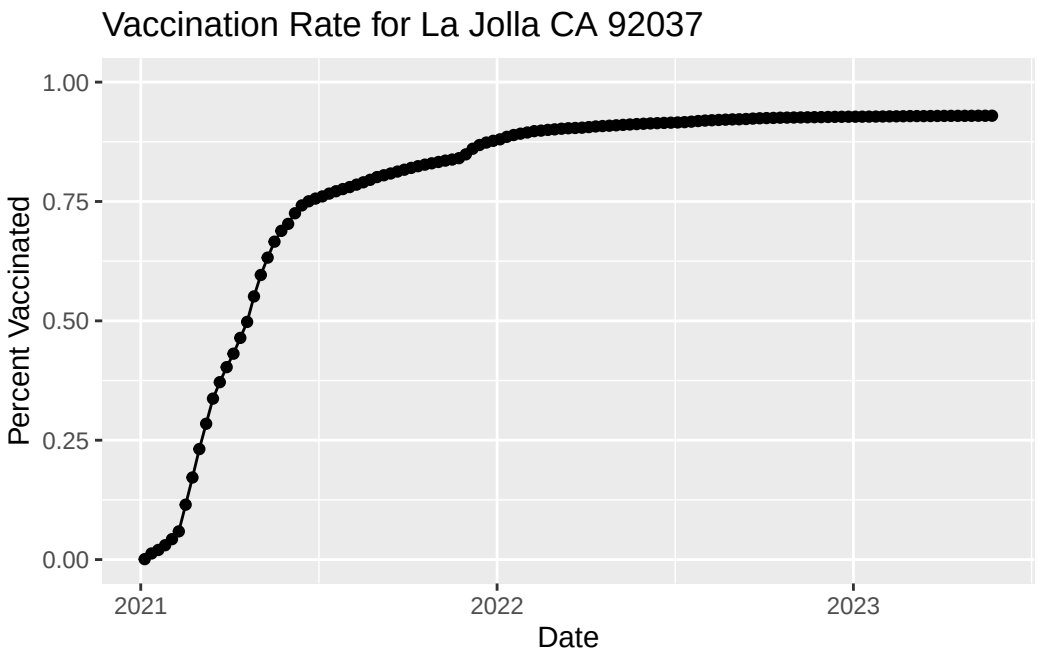
Focus on UCSD/La Jolla

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

```
[1] 36144
```

Q15:

```
vaccination_rate_plot <- ggplot(ucsd) +
  aes(as_of_date, percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(title = "Vaccination Rate for La Jolla CA 92037", x= "Date", y="Percent Vaccinated"
vaccination_rate_plot
```



Comparing to similar sized areas

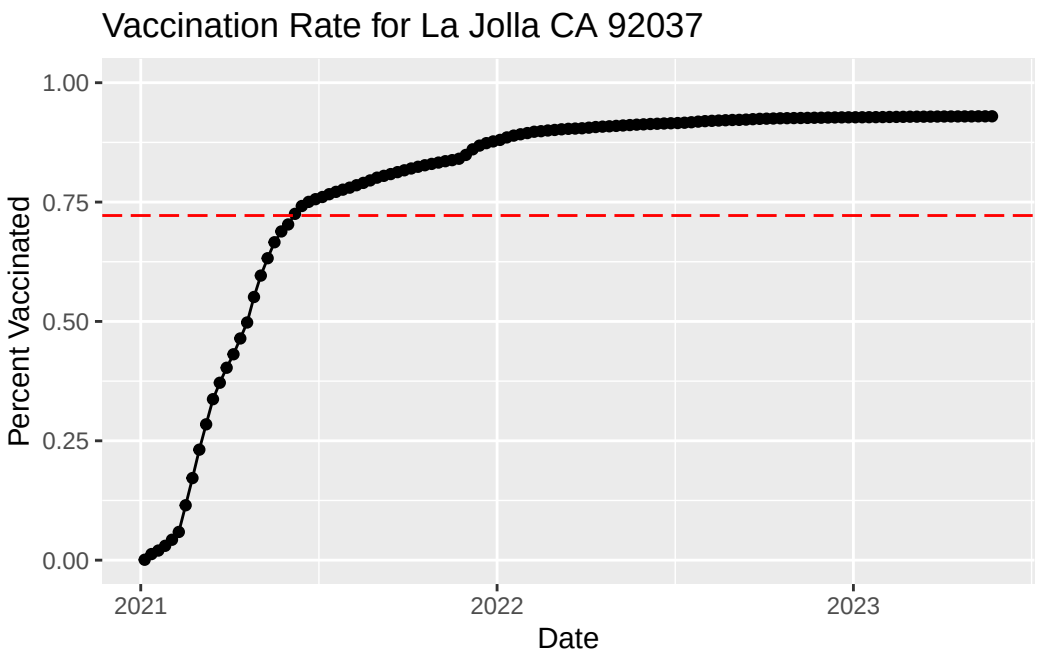
```
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2023-02-28")

#head(vax.36)
```

Q16:

```
mean_92037 <- mean(vax.36$percent_of_population_fully_vaccinated)
mean_92037
```

```
[1] 0.7218591
```

```
vaccination_rate_plot + geom_hline(yintercept = mean_92037, color = "red", linetype = 5)
```



Q17: The 6 number summary is listed below:

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

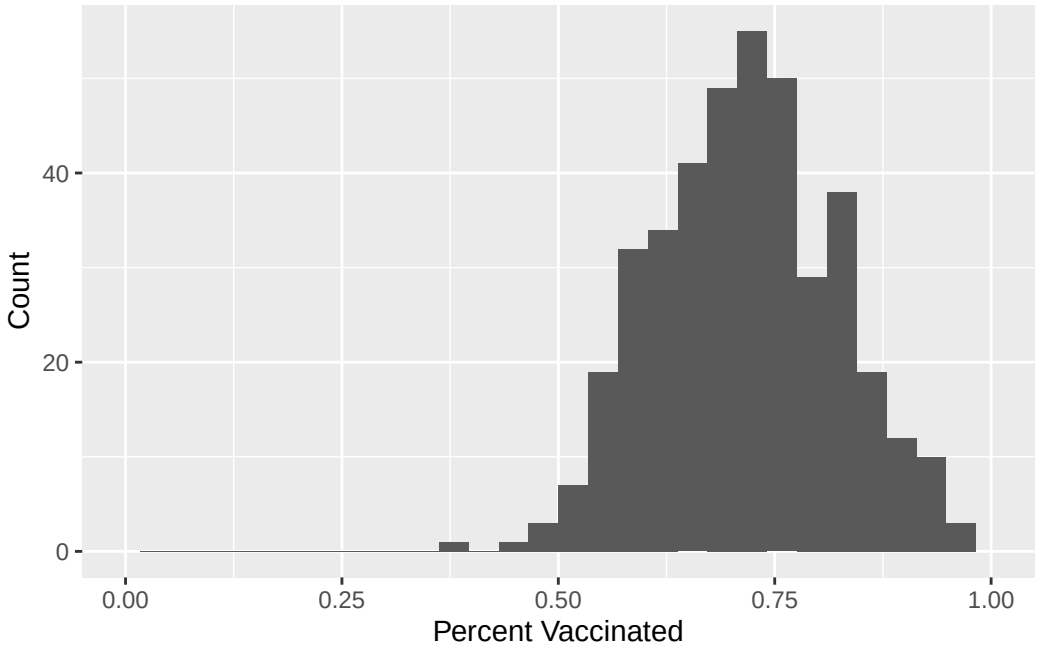
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3809  0.6464  0.7201  0.7219  0.7916  1.0000
```

Q18:

```
ggplot(vax.36, aes(percent_of_population_fully_vaccinated)) +
  geom_histogram(na.rm = TRUE) +
  xlim(0,1) +
```

```
    xlab("Percent Vaccinated") + ylab("Count")
```

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




```
 vax %>% filter(as_of_date == "2023-05-23") %>%
   filter(zip_code_tabulation_area=="92040") %>%
   select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                               0.552434
```

```
 vax %>% filter(as_of_date == "2023-05-23") %>%
   filter(zip_code_tabulation_area=="92109") %>%
   select(percent_of_population_fully_vaccinated)
```

```
  percent_of_population_fully_vaccinated
1                                0.69487
```

Q19: Based on the above result, both the two ZIP code areas are below the average value of 0.7219.

Q20:

```
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", na.rm = TRUE) +
  ylim(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 = mean_92037, linetype=5)
```