COVID-19 Vaccination Rates

Joshua Cheung

03/01/2022

# Background

We will start by downloading the most recently dated “Statewide COVID-19 Vaccines Administered by ZIP Code” CSV file from: <https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data-by-zip-code>

# Getting started

We move our downloaded CSV file to our project directory and then read/import into an R object called vax. We will use this data to answer the questions below.

# We import the 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 92140 San Diego San Diego  
## 2 2021-01-05 94133 San Francisco San Francisco  
## 3 2021-01-05 94523 Contra Costa Contra Costa  
## 4 2021-01-05 94005 San Mateo San Mateo  
## 5 2021-01-05 94104 San Francisco San Francisco  
## 6 2021-01-05 94549 Contra Costa Contra Costa  
## vaccine\_equity\_metric\_quartile vem\_source  
## 1 NA No VEM Assigned  
## 2 3 Healthy Places Index Score  
## 3 4 Healthy Places Index Score  
## 4 4 Healthy Places Index Score  
## 5 NA No VEM Assigned  
## 6 4 Healthy Places Index Score  
## age12\_plus\_population age5\_plus\_population persons\_fully\_vaccinated  
## 1 3747.7 3737 NA  
## 2 25070.5 25957 NA  
## 3 30457.9 32828 NA  
## 4 3996.1 4364 NA  
## 5 387.8 399 NA  
## 6 25393.8 28468 NA  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA 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  
## 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?**

Inspection of the column names shows that the column “persons\_fully\_vaccinated” details the total number of people fully vaccinated.

**Q2. What column details the Zip code tabulation area?**

Inspection of the column names shows that the column “zip\_code\_tabulation\_area” details the zip code tabulation area.

**Q3. What is the earliest date in this dataset?**

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"

We see that the earliest date in this dataset is 2021-01-05.

**Q4. What is the latest date in this dataset?**

tail(vax$as\_of\_date)

## [1] "2022-02-22" "2022-02-22" "2022-02-22" "2022-02-22" "2022-02-22"  
## [6] "2022-02-22"

We see that the latest date in this dataset is 2022-02-22.

We now call the skim() function from the skimr package to get a quick overview of this dataset:

library(skimr)  
skimr::skim(vax)

Data summary

|  |  |
| --- | --- |
| Name | vax |
| Number of rows | 105840 |
| Number of columns | 15 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| 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 | 60 | 0 |
| local\_health\_jurisdiction | 0 | 1 | 0 | 15 | 300 | 62 | 0 |
| county | 0 | 1 | 0 | 15 | 300 | 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 | 5220 | 0.95 | 2.44 | 1.11 | 1 | 1.00 | 2.00 | 3.00 | 4.0 | ▇▇▁▇▇ |
| age12\_plus\_population | 0 | 1.00 | 18895.04 | 18993.92 | 0 | 1346.95 | 13685.10 | 31756.12 | 88556.7 | ▇▃▂▁▁ |
| age5\_plus\_population | 0 | 1.00 | 20875.24 | 21106.02 | 0 | 1460.50 | 15364.00 | 34877.00 | 101902.0 | ▇▃▂▁▁ |
| persons\_fully\_vaccinated | 18174 | 0.83 | 12064.29 | 12983.91 | 11 | 1059.00 | 7287.50 | 19859.00 | 77213.0 | ▇▃▁▁▁ |
| persons\_partially\_vaccinated | 18174 | 0.83 | 820.71 | 1318.77 | 11 | 76.00 | 370.00 | 1066.00 | 31869.0 | ▇▁▁▁▁ |
| percent\_of\_population\_fully\_vaccinated | 18174 | 0.83 | 0.51 | 0.26 | 0 | 0.33 | 0.54 | 0.70 | 1.0 | ▅▅▇▇▃ |
| percent\_of\_population\_partially\_vaccinated | 18174 | 0.83 | 0.05 | 0.09 | 0 | 0.01 | 0.03 | 0.05 | 1.0 | ▇▁▁▁▁ |
| percent\_of\_population\_with\_1\_plus\_dose | 18174 | 0.83 | 0.54 | 0.27 | 0 | 0.35 | 0.58 | 0.75 | 1.0 | ▅▃▆▇▅ |
| booster\_recip\_count | 64191 | 0.39 | 3923.43 | 5704.10 | 11 | 169.00 | 1072.00 | 5803.00 | 49951.0 | ▇▁▁▁▁ |

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

Using the results from the skim