Class_17.Rmd

Kaito Tanaka

11/29/2021

Getting Started

```
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction         county
## 1 2021-01-05                   92395            San Bernardino San Bernardino
## 2 2021-01-05                   93206                      Kern           Kern
## 3 2021-01-05                   91006                 Los Angeles    Los Angeles
## 4 2021-01-05                   91901                   San Diego      San Diego
## 5 2021-01-05                   92230                   Riverside      Riverside
## 6 2021-01-05                   92662                      Orange         Orange
##   vaccine_equity_metric_quartile                vem_source
## 1                              1 Healthy Places Index Score
## 2                              1 Healthy Places Index Score
## 3                              3 Healthy Places Index Score
## 4                              3 Healthy Places Index Score
## 5                              1 Healthy Places Index Score
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               35915.3                40888                       NA
## 2                1237.5                 1521                       NA
## 3               28742.7                31347                       19
## 4               15549.8                16905                       12
## 5                2320.2                 2526                       NA
## 6                2349.5                 2397                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                           NA                                     NA
## 3                          873                                0.000606
## 4                          271                                0.000710
## 5                           NA                                     NA
## 6                           NA                                     NA
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                         NA
## 3                                   0.027850
## 4                                   0.016031
## 5                                         NA
## 6                                         NA
```

```
##   percent_of_population_with_1_plus_dose
## 1                                     NA
## 2                                     NA
## 3                                0.028456
## 4                                0.016741
## 5                                     NA
## 6                                     NA
##                                                                    redacted
## 1 Information redacted in accordance with CA state privacy requirements
## 2 Information redacted in accordance with CA state privacy requirements
## 3                                                                       No
## 4                                                                       No
## 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] "2021-11-23" "2021-11-23" "2021-11-23" "2021-11-23" "2021-11-23"
## [6] "2021-11-23"
```

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

The column "persons_fully_vaccinated" details the total number of people fully vaccinated.

Q2. What column details the Zip code tabulation area?

The column "zip_code_tabulation_area" details the zip code tabulation area.

Q3. What is the earliest date in this dataset?

The earliest date is 2021-01-05.

Q4. What is the latest date in this dataset?

The latest date is 2021-11-23.

Get a quick overview of the dataset

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

Q5. How many numeric columns are in this dataset?

```
ncol(vax)
```

```
## [1] 14
```

9 numeric columns.

> 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] 8355
```

8256 NA values are present in "persons_fully_vaccinated".

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

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

```
## [1] 10.07744
```

10.17% of the data is missing here

##Working with Dates

We will use the **lubridate** package to make life a lot easier when dealing with dates and times:

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

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

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

what is today's date?

```
today()
```

```
## [1] "2021-11-30"
```

```
#Specify our format
vax$as_of_date <- ymd(vax$as_of_date)
```

```
#Example of what we can now do easily
today() - vax$as_of_date[1]
```

```
## Time difference of 329 days
```

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

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

```
## Time difference of 322 days
```

7 days have passed since the last update of the dataset.

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

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

```
## [1] 47
```

There are 47 unique dates in the dataset.

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