Lab 17: Vaccination Mini-Project (Quarto Doc)

Forrest Wang A15911047

*Quarto was halting my rendering and the "head" command generated too many pages so I just used ">" to show the code I was writing

#Getting Started: Importing Data >vax <- read.csv("covid19vaccinesbyzipcode_test.csv") >head(vax)

Q1. What column details the total number of people fully vaccinated? >head(vax) >vax["persons_fully_vaccinated"]

A1. "persons_fully_vaccinated", aka the 11th row down from the "Data Dictionary" details total number of fully vaccinated individuals under the requirements that they have "2 Pfizer doses >= 17 days apart, 2 Moderna doses >= 24 days apart, and 1 dose of J&J, a combination of Pfizer and Moderna doses >=17 days apart"

Q2. What column details the Zip code tabulation area?

    head(vax) vax["zip_code_tabulation_area"]

A2: "zip_code_tabulation_area", aka the 4th row down from the "Data Dictionary" details the Zip code tabulation area.

Q3. What is the earliest date in this dataset?

    head(vax)

A3: The earliest date in the dataset is from 2021-01-05, Jan 5 of 2021

Q4. What is the latest date in this dataset?

    tail(vax)

A4: The latest dart in the dataset is from 2022-11-29, Nov 29 of 2022

Skimming…

    skimr::skim(vax)

Q5. How many numeric columns are in this dataset? A5. There are 18 columnes in the 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))

A6. There are 15048 NA values in the column "persons_fully_vaccinated"

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

    15048/2179043333

A7: 6.9E-6 or 6.9E-06% of all persons fully vaccinated values are missing

Q8. [Optional]: Why might this data be missing? A8. This data might be missing because certain vaccine administering organizations like the VA do not disclose their information.

#Working with dates *Quarto was halting my rendering so I just used ">" to show the code I was writing

What is today's date?

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

    today()

Converting data to lubridate format: $>$vax$as_of_date <- ymd(vax$as_of_date)

Q9. How many days have passed since the last update of the dataset? $>$today() - vax$as_of_date[1]

A9. 697 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)?

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

A10. There has been 693 unique days that are detailed.