PHW251 Team Project: Milestone #2

Scenario Two: COVID Vaccination Progress

Saira Mayet, Jessica Pak, Andrew Nguyen

9/28/2021

Description of dataset

- What is the data source?

  There are two datasets of interest: one describing COVID-19 vaccine administration across the state of california, sourced from the California Open Data Portal ("cov_vax_admin.csv"). Fields include date, ZIP code, county, and raw counts of vaccination status, amongst other population information. The dataset spans January 5th, 2021, to September 21st, 2021.

  The other describes demographics (e.g. population, race/ethnicity, age, household size, etc.) for each California county, updated to 2012. This dataset was rehosted on Avery Richards' GitHub, and is sourced from Census Data.

- How does the dataset relate to the group statement and question?

  Problem statement: We are monitoring the state level COVID-19 vaccination rates among counties in California and in relation to age.
  Question: Is there any correlation between median age and vaccinated person prevalence on the county level?

  The group statement and question relates to exploring, analyzing, and visualizing vaccination rates at county level and to explore if there is a correlation between age and vaccination rate. These two described datasets have necessary fields to support these analyses by including vaccination information at the ZIP level and county demographic data.

Load libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
## Warning: package 'stringr' was built under R version 4.1.1
```

```
## Warning: package 'forcats' was built under R version 4.1.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

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

Import Statement

```
file_path_vax <- "https://data.chhs.ca.gov/dataset/ead44d40-fd63-4f9f-950a-3b0111074de8/resource/ec32ee
file_path_county <- "https://raw.githubusercontent.com/Averysaurus/reproducable_examples-/main/ca_count

vax_temp <- read.csv(file_path_vax)
county_temp <- read.csv(file_path_county)

str(vax_temp)
```

```
## 'data.frame':    74088 obs. of  13 variables:
##  $ as_of_date                                : chr  "2021-01-05" "2021-01-05" "2021-01-05" "2021-01-
##  $ zip_code_tabulation_area                  : int  93428 95327 95668 95826 95240 93631 93555 93544
##  $ local_health_jurisdiction                 : chr  "San Luis Obispo" "Tuolumne" "Sutter" "Sacrament
##  $ county                                    : chr  "San Luis Obispo" "Tuolumne" "Sutter" "Sacrament
##  $ vaccine_equity_metric_quartile            : num  3 2 2 3 1 2 2 1 2 3 ...
##  $ vem_source                                : chr  "Healthy Places Index Score" "Healthy Places Ind
##  $ age12_plus_population                     : num  5532 7384 562 33966 39229 ...
##  $ persons_fully_vaccinated                  : num  NA NA NA NA 41 NA NA NA NA NA ...
##  $ persons_partially_vaccinated              : num  NA NA NA NA 399 NA NA NA NA NA ...
##  $ percent_of_population_fully_vaccinated    : num  NA NA NA NA 0.00104 ...
##  $ percent_of_population_partially_vaccinated: num  NA NA NA NA 0.0102 ...
##  $ percent_of_population_with_1_plus_dose    : num  NA NA NA NA 0.0112 ...
##  $ redacted                                  : chr  "Information redacted in accordance with CA stat
```

```
vax <- vax_temp %>% select(-c("local_health_jurisdiction", "vem_source", 10:13))
county <- county_temp %>% select(c("name", "pop2012", "med_age"))

head(vax)
```

```
##   as_of_date zip_code_tabulation_area          county
## 1 2021-01-05                    93428 San Luis Obispo
## 2 2021-01-05                    95327        Tuolumne
## 3 2021-01-05                    95668          Sutter
## 4 2021-01-05                    95826      Sacramento
## 5 2021-01-05                    95240     San Joaquin
## 6 2021-01-05                    93631          Fresno
##   vaccine_equity_metric_quartile age12_plus_population persons_fully_vaccinated
## 1                              3                5532.1                       NA
## 2                              2                7383.5                       NA
## 3                              2                 562.0                       NA
## 4                              3               33965.9                       NA
## 5                              1               39228.8                       41
## 6                              2               13395.1                       NA
##   persons_partially_vaccinated
## 1                           NA
## 2                           NA
## 3                           NA
## 4                           NA
## 5                          399
## 6                           NA
```

```
head(county)
```

```
##          name pop2012 med_age
## 1        Kern  851089    30.7
## 2       Kings  155039    31.1
## 3        Lake   65253    45.0
## 4      Lassen   35039    37.0
## 5 Los Angeles 9904341    34.8
## 6      Madera  153025    33.1
```

Determine data types

```
print("These are the data types for the vaccination dataset:")
```

```
## [1] "These are the data types for the vaccination dataset:"
```

```
sapply(vax, class)
```

```
##                     as_of_date       zip_code_tabulation_area
##                    "character"                      "integer"
##                         county vaccine_equity_metric_quartile
##                    "character"                      "numeric"
##          age12_plus_population       persons_fully_vaccinated
##                      "numeric"                      "numeric"
##   persons_partially_vaccinated
##                      "numeric"
```

```
print("These are the datatypes for the county demographics dataset:")
```

```
## [1] "These are the datatypes for the county demographics dataset:"
```

```
sapply(county, class)
```

```
##        name     pop2012     med_age
## "character"   "integer"   "numeric"
```

Identifying desired type/format for each data

as_of_date: character -> date vaccine_equity_metric_quartile: integer -> factor

```
vax$as_of_date <- as_date(vax$as_of_date)
class(vax$as_of_date)
```

```
## [1] "Date"
```

```
vax$vaccine_equity_metric_quartile <- as.factor(vax$vaccine_equity_metric_quartile)
class(vax$vaccine_equity_metric_quartile)
```

```
## [1] "factor"
```

Basic descriptives of data elements

```
print("Here are the simple frequencies for the county and vaccine equity metric (by quartile) variables
```

```
## [1] "Here are the simple frequencies for the county and vaccine equity metric (by quartile) variables
```

```
table(vax$county) #how many ZIP code time entries exist in each county
```

```
##
##                         Alameda          Alpine          Amador           Butte
##             210            2058              42             504             756
##       Calaveras          Colusa    Contra Costa       Del Norte       El Dorado
##             756             294            1806             168             924
##          Fresno           Glenn        Humboldt        Imperial            Inyo
##            2310             252            1470             630             420
##            Kern           Kings            Lake          Lassen     Los Angeles
##            2058             294             588             546           12180
##          Madera           Marin        Mariposa       Mendocino          Merced
##             504            1176             336            1092             798
##           Modoc            Mono        Monterey            Napa          Nevada
##             462             294            1176             420             504
##          Orange          Placer          Plumas       Riverside      Sacramento
##            3696            1218             672            2940            2268
##      San Benito  San Bernardino       San Diego   San Francisco     San Joaquin
##             168            3738            4494            1134            1344
## San Luis Obispo       San Mateo   Santa Barbara     Santa Clara      Santa Cruz
##             924            1218             966            2436             714
##          Shasta          Sierra        Siskiyou          Solano          Sonoma
##            1092             294             882             630            1512
##      Stanislaus          Sutter          Tehama         Trinity          Tulare
##            1008             378             546             546            1386
##        Tuolumne         Ventura            Yolo            Yuba
##             546            1134             714             462
```

```
table(vax$vaccine_equity_metric_quartile) # ZIP code time entries categorized by vaccine equity metric
```

```
##
##     1     2     3     4
## 18690 18606 16884 16254
```

```
table(county$name) #how many counties exist in the ca_county_demographic dataset
```

```
##
##         Alameda          Alpine          Amador           Butte       Calaveras
##               1               1               1               1               1
##          Colusa    Contra Costa       Del Norte       El Dorado          Fresno
##               1               1               1               1               1
##           Glenn        Humboldt        Imperial            Inyo            Kern
##               1               1               1               1               1
##           Kings            Lake          Lassen     Los Angeles          Madera
##               1               1               1               1               1
##           Marin        Mariposa       Mendocino          Merced           Modoc
##               1               1               1               1               1
##            Mono        Monterey            Napa          Nevada          Orange
##               1               1               1               1               1
##          Placer          Plumas       Riverside      Sacramento      San Benito
##               1               1               1               1               1
##  San Bernardino       San Diego   San Francisco     San Joaquin San Luis Obispo
##               1               1               1               1               1
##       San Mateo   Santa Barbara     Santa Clara      Santa Cruz          Shasta
##               1               1               1               1               1
##          Sierra        Siskiyou          Solano          Sonoma      Stanislaus
##               1               1               1               1               1
##          Sutter          Tehama         Trinity          Tulare        Tuolumne
##               1               1               1               1               1
##         Ventura            Yolo            Yuba
##               1               1               1
```

```
print("Here are summary statistics for numeric variables of interest.")
```

```
## [1] "Here are summary statistics for numeric variables of interest."
```

```
summary(vax$age12_plus_population)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    1347   13685   18895   31756   88557
```

```
summary(vax$persons_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA’s
##      11     462    3674    8895   14812   70322    7742
```

```
summary(vax$persons_partially_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA’s
##      11     203    1295    1946    2973   20273    7742
```

```
summary(county$pop2012)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1148   48492  180662  650129  645995 9904341
```

```
summary(county$med_age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.60   33.70   37.05   38.49   43.08   51.00
```