PHW251 Team Project: Milestone #3

Scenario Two: COVID Vaccination Progress

Saira Mayet, Jessica Pak, Andrew Nguyen

10/28/2021

Import Statement

```
# from Milestone #2
file_path_vax <-
  "https://data.chhs.ca.gov/dataset/ead44d40-fd63-4f9f-950a-3b0111074de8/resource/ec32eece-7474-4488-871
file_path_county <-
  "https://raw.githubusercontent.com/Averysaurus/reproducable_examples-/main/ca_county_demographics.csv

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

# subset data to only include variables of interest
vax <- vax_temp %>%
  select(-c("vaccine_equity_metric_quartile",
            "local_health_jurisdiction",
            "vem_source",
            10:13))
county <- county_temp %>%
  select(c("name",
           "med_age"))
```

1. Subset data, as needed:

Partial subsetting of data was already performed in Milestone #2 (see import statement above).

We are keeping the following variables in these datasets:

- County: name, median age
- Vax: date, ZIP , county, vaccine equity quartile, population 12+, number person fully vaccinated, number partially vaccinated

Now we will subset for the latest date of data, as the CDPH's "COVID-19 Vaccines Administered by Zip Code" excel sheet is a continually-updated dataset.

```
vax_latest <- vax %>%
  mutate(date = as_date(as_of_date)) %>%
  filter(date == max(date)) %>%
  select( - as_of_date)
```

2. Create new variables needed for analysis

Clean variables:

- Mean imputation: county-level means of fully and partially vaccinated that will be used to replace NA values in dataset
- Percent eligible population partially vaccinated = # of persons partially vaccinated / population 12+ (at county level)
- Percent eligible population fully vaccination = # of persons fully vaccinated / population 12+ (at county level)
- Merging relational data: county demographic dataset with vaccine administration dataset using key variable "county".

Since our final clean dataset needs to undergo mean imputation before our new variables are created, we will perform that first:

```
#aggregate county-level counts + %s for vaccination

vax_temp <- vax_latest %>%
  group_by(county) %>%
  mutate(county_partial =
           sum(persons_partially_vaccinated, na.rm = T)/
           sum(age12_plus_population, na.rm =T),
         county_fully =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age12_plus_population, na.rm = T)) %>%
  ungroup()

#mean imputation: for ZIPs that have eligible population counts but lack vaccination numbers

vax_temp <- vax_temp %>%
  mutate(persons_partially_vaccinated_2 =
           ifelse(is.na(persons_partially_vaccinated),
                  age12_plus_population*county_partial,
                  persons_partially_vaccinated),
         persons_fully_vaccinated_2 =
           ifelse(is.na(persons_fully_vaccinated),
                  age12_plus_population*county_fully,
                  persons_fully_vaccinated))

#make second county-level aggregate with imputed data + original data

vax_temp <- vax_temp %>%
  group_by(county) %>%
  mutate(county_partial_2 =
           sum(persons_partially_vaccinated, na.rm = T)/
           sum(age12_plus_population, na.rm =T),
         county_fully_2 =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age12_plus_population, na.rm = T),
         county_eligible_pop =
           sum(age12_plus_population, na.rm = T)) %>%
  ungroup()
```

```
#create new dataset with only one entry per county

vax_aggregate <- vax_temp %>%
  distinct(county, .keep_all = T) %>%
  select(county,
         county_partial_2,
         county_fully_2,
         county_eligible_pop) %>%
  filter(county_partial_2 != 0)
```

Merging both county and vax_latest dataset:

```
county <- county %>%
  rename(county = name)

data <- left_join(vax_aggregate, county, by="county")
```

4. Data dictionary based on clean dataset

Variable Name	Definition
County Name	Names of the 58 counties
Median Age	Median age of population per county as recorded by United States American Community Survey (ACS) data.
Percentage fully vaccinated	number of eligible people fully vaccinated (2+ doses for mRNA vaccine, 1 dose for adenovirus vaccine) divided by number of persons 12+
Percent partially vaccinated	number of eligible people partially vaccination (1 dose for mRNA vaccine) divide by number of persons 12+
Eligible Population	total number of people eligible to receive vaccinations (age 12+)

5. Table with descriptive statistics with 4 data elements

```
# final clean data
clean_data <- data %>%
  rename(County = county,
         Partially_Vax_Rate = county_partial_2,
         Fully_Vax_Rate = county_fully_2,
         Eligible_Population = county_eligible_pop,
         Median_Age = med_age) %>%
  mutate(Partially_Vax_Rate = round((Partially_Vax_Rate*100), digits =2),
         Fully_Vax_Rate = round((Fully_Vax_Rate*100), digits = 2),
         Eligible_Population = round(Eligible_Population, digits = 0),
         Median_Age = round(Median_Age, digits = 0)) %>%
  arrange(County)

# table with descriptive statistics (averages calculated for CA as a whole) and 4 data elements
kable(clean_data,
      booktabs = T,
      col.names = c("County Name",
                    "Partially Vaccinated Rate",
                    "Fully Vaccinated Rate",
                    "Total Eligible Population",
                    "Median Age"),
      align = "c",
      caption = "Vaccination Rates for California Counties",
      format.args=list(big.mark=",")) %>%
      add_header_above(c("California"= 1,
      "6.97"=1,
      "62.8"=1,
      "33,330,578"=1,
      "38"=1),
      bold = T)
```

Table 2: Vaccination Rates for California Counties

California	**6.97**	**62.8**	**33,330,578**	**38**
County Name	Partially Vaccinated Rate	Fully Vaccinated Rate	Total Eligible Population	Median Age
Alameda	5.35	83.83	1,430,500	37
Alpine	5.29	53.80	812	46
Amador	9.24	55.75	34,928	48
Butte	5.15	51.75	195,513	37
Calaveras	7.06	53.64	40,548	49
Colusa	6.89	62.61	17,791	34
Contra Costa	4.84	86.33	968,349	38
Del Norte	5.79	44.59	23,641	39
El Dorado	5.74	62.23	164,929	44
Fresno	7.84	64.30	799,799	31
Glenn	4.89	56.52	23,495	35
Humboldt	6.16	64.75	118,282	37
Imperial	17.05	86.38	142,270	32
Inyo	4.92	54.36	16,760	46
Kern	7.20	56.16	716,569	31
Kings	6.75	46.05	121,666	31
Lake	6.92	57.20	55,179	45
Lassen	3.50	25.66	25,302	37
Los Angeles	8.01	73.49	8,620,134	35
Madera	6.93	56.52	126,670	33
Marin	5.66	88.60	225,657	44
Mariposa	18.41	44.50	14,187	49
Mendocino	8.25	68.66	74,612	42
Merced	11.97	54.93	216,985	30
Modoc	3.45	39.02	8,555	46
Mono	6.00	73.21	11,139	37
Monterey	7.89	75.22	344,680	33
Napa	8.50	78.18	122,228	40
Nevada	6.68	60.97	85,795	48
Orange	6.25	75.13	2,719,660	36
Placer	5.40	71.41	332,902	40
Plumas	4.76	52.83	18,316	50
Riverside	7.00	62.44	2,016,228	34
Sacramento	6.12	70.18	1,284,073	35
San Benito	8.34	71.83	49,034	34
San Bernardino	6.39	60.07	1,769,305	32
San Diego	9.82	75.03	2,825,624	35
San Francisco	5.29	84.12	792,610	38
San Joaquin	9.70	65.02	607,696	33
San Luis Obispo	6.22	61.92	249,268	39
San Mateo	5.99	85.34	637,296	39
Santa Barbara	7.47	70.59	378,790	34
Santa Clara	4.97	88.10	1,664,302	36
Santa Cruz	5.58	75.81	255,646	37
Shasta	6.90	47.39	147,137	42
Sierra	1.82	27.95	2,411	51
Siskiyou	6.59	48.59	36,737	47
Solano	9.95	68.50	375,537	37
Sonoma	6.12	77.77	435,690	40
Stanislaus	9.49	60.60	447,618	33
Sutter	7.18	63.17	80,375	35