PHW251 Team Project: Milestone #4

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-7474-4488-87
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",
            "age12_plus_population",
            11:14))
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)


#get latest date and convert into character for use in plot captions
date <- vax %>%
  mutate(date = as.Date.character(as_of_date)) %>%
  select(- as_of_date)

current_date <- as.Date(max(date$date))
current_date_chr <- as.character(current_date)

caption <- paste("Source: California Department of Health, current to", current_date_chr)
```

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 5+ (at county level)
- Percent eligible population fully vaccination = # of persons fully vaccinated / population 5+ (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(age5_plus_population, na.rm =T),
         county_fully =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age5_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),
                  age5_plus_population*county_partial,
                  persons_partially_vaccinated),
         persons_fully_vaccinated_2 =
           ifelse(is.na(persons_fully_vaccinated),
                  age5_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(age5_plus_population, na.rm =T),
         county_fully_2 =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age5_plus_population, na.rm = T),
         county_eligible_pop =
           sum(age5_plus_population, na.rm = T)) %>%
  ungroup()
```

```r
#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:

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

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

4. Data visualization of data elements

```r
# 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 of descriptive statistics

```
# 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) %>%
      kable_styling(latex_options = c("striped", "scale_down"))
```

The table describes the partial vaccinated rate, fully vaccinated rate, total eligible population, and median age of each county in California. The summarized average rates, total eligible population, and median age for the entire state of California is calculated and presented at the top of the table.

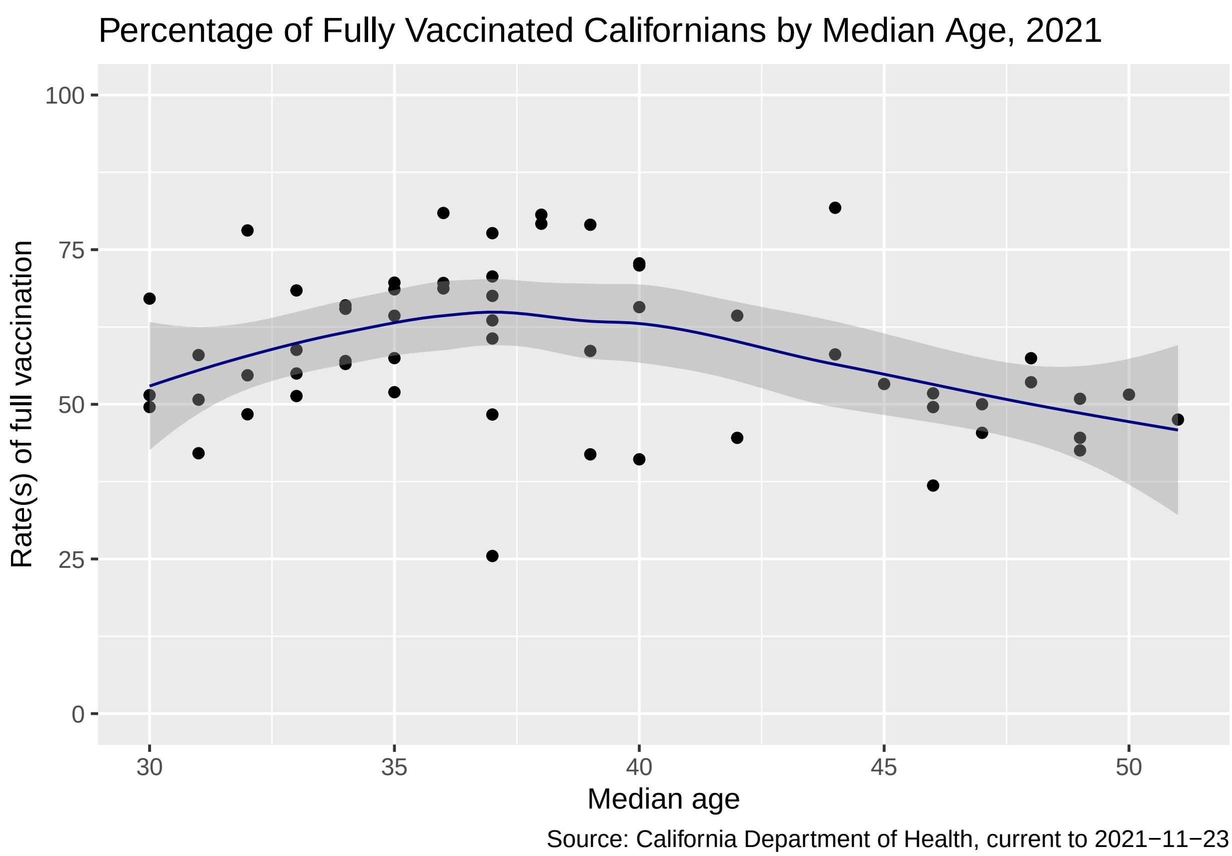
Table 1: Vaccination Rates for California Counties

California	**6.97**	**62.8**	**33,330,578**	**38**
County Name	Partially Vaccinated Rate	Fully Vaccinated Rate	Total Eligible Population	Median Age
Alameda	7.67	77.67	1,565,553	37
Alpine	6.07	49.55	890	46
Amador	9.53	53.57	37,067	48
Butte	5.08	48.35	213,817	37
Calaveras	7.71	50.89	43,656	49
Colusa	6.52	56.52	20,153	34
Contra Costa	6.67	79.19	1,071,086	38
Del Norte	5.85	41.91	25,963	39
El Dorado	6.12	58.07	179,821	44
Fresno	7.60	57.96	911,080	31
Glenn	4.93	51.96	26,204	35
Humboldt	7.09	60.64	128,806	37
Imperial	18.50	78.11	161,453	32
Inyo	6.95	51.77	18,268	46
Kern	6.78	50.75	818,823	31
Kings	6.32	42.09	137,655	31
Lake	6.41	53.28	60,336	45
Lassen	3.06	25.48	26,919	37
Los Angeles	8.57	68.58	9,463,365	35
Madera	6.92	51.34	143,316	33
Marin	9.11	81.77	246,959	44
Mariposa	20.18	44.59	15,319	49
Mendocino	8.42	64.33	81,751	42
Merced	11.18	49.56	248,786	30
Modoc	3.36	36.87	9,384	46
Mono	7.68	67.53	12,259	37
Monterey	8.38	68.41	387,591	33
Napa	9.25	72.79	133,221	40
Nevada	7.43	57.45	92,519	48
Orange	6.94	69.62	2,986,910	36
Placer	6.24	65.71	367,860	40
Plumas	5.25	51.57	19,548	50
Riverside	6.85	57.00	2,255,664	34
Sacramento	6.85	64.32	1,427,122	35
San Benito	8.23	66.00	55,001	34
San Bernardino	6.14	54.69	1,991,511	32
San Diego	11.86	69.67	3,101,086	35
San Francisco	7.48	80.65	835,425	38
San Joaquin	9.36	58.80	688,728	33
San Luis Obispo	6.71	58.63	268,922	39
San Mateo	8.05	79.03	696,222	39
Santa Barbara	8.02	65.44	417,143	34
Santa Clara	7.22	80.93	1,833,854	36
Santa Cruz	6.57	70.67	278,046	37
Shasta	6.23	44.58	161,537	42
Sierra	3.15	47.53	2,632	51
Siskiyou	6.24	45.39	40,245	47
Solano	9.68	63.57	414,116	37
Sonoma	7.18	72.45	475,030	40
Stanislaus	8.95	54.98	505,820	33
Sutter	6.46	57.47	90,595	35
Tehama	4.67	41.11	66,959	40
Trinity	6.01	42.55	12,260	49
Tulare	7.06	51.48	420,906	30
Tuolumne	6.96	50.01	52,564	47

Scatterplot of Median Age vs Fully Vaccinated Rate

```
ggplot(data = clean_data, aes(x= Median_Age,y= Fully_Vax_Rate)) +
  geom_point() +
  geom_smooth(col = "navy blue", size = 0.5) +
  scale_y_continuous(limits=c(0, 100)) +
  labs(x= "Median age",
       y= "Rate(s) of full vaccination",
       title= "Percentage of Fully Vaccinated Californians by Median Age, 2021",
       caption = caption )
```

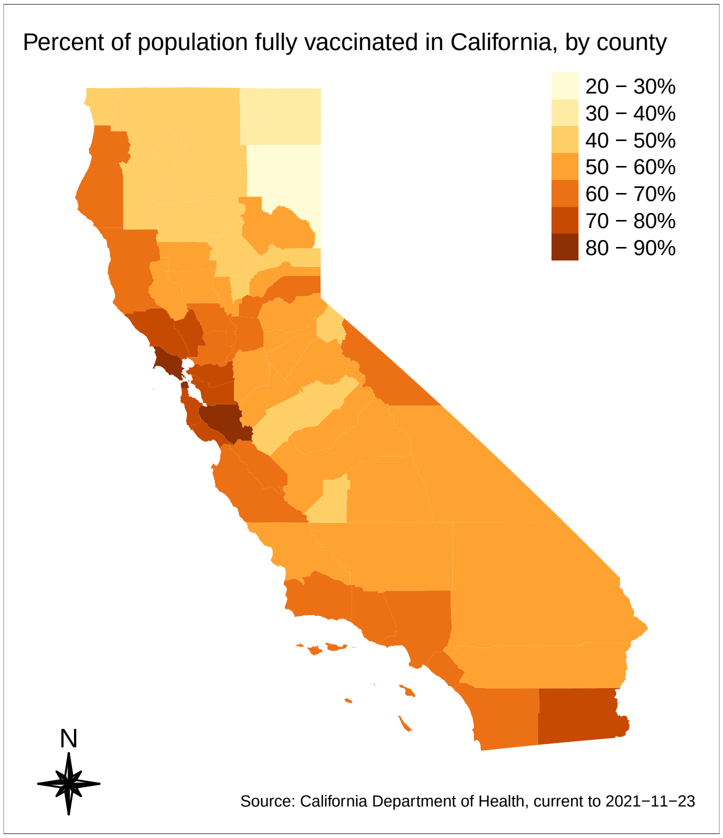
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



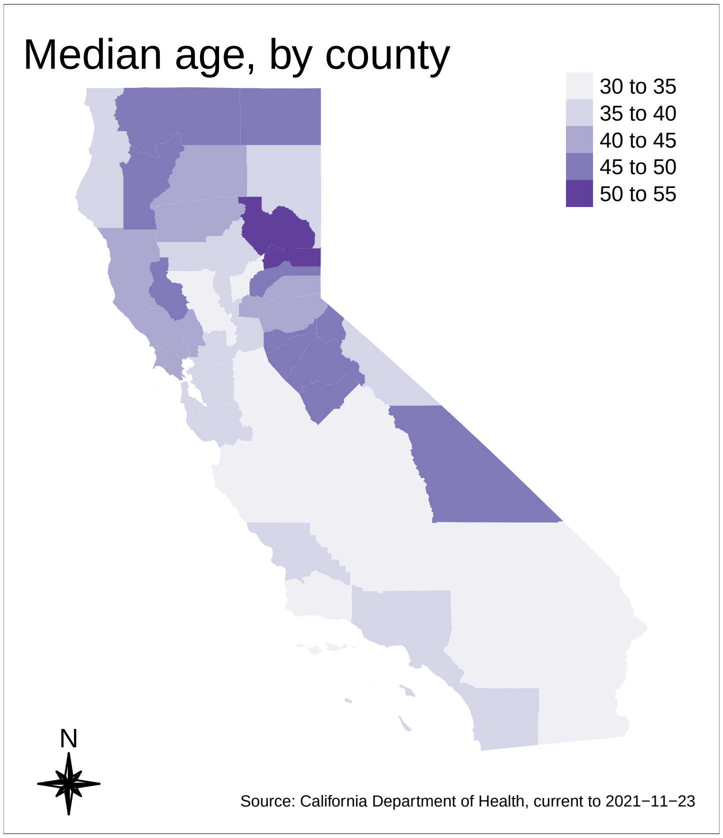
The scatterplot above shows the rate of fully vaccinated individuals stratified by median age of California residents, as measured from California Department of Public Health's latest online data repository for COVID-19. A trend line was fitted to follow the points on the plot, and a general trend of increasing fully vaccinated rates can be seen near the median age of 36-38.

Map of vaccination rates and median age, by County

```r
#get CA boundary with high definition
#ca <- USAboundaries::us_states(resolution = "high", states = "CA")

#CA county boundary
ca_co <- USAboundaries::us_counties(resolution = "high", states = "CA") %>%
  select(- state_name)

#Join dataset with sf polygon
clean_data_rename <- clean_data %>%
  rename(name = County)

ca_covid <- dplyr::left_join(clean_data_rename, ca_co, by = "name")

#Make joined dataset into sf object
ca_covid <- st_as_sf(ca_covid, sf_column_name = "geometry")

#Create plot
legend_title = " "
tm_shape(ca_covid) +
  tm_fill(col = "Fully_Vax_Rate",
          title = legend_title,
          labels = c('20 - 30%','30 - 40%','40 - 50%','50 - 60%', '60 - 70%', '70 - 80%', '80 - 90%')) +
  tm_compass(type = "8star",
             position = c("left", "bottom"),
             size = 2) +
  tm_layout(scale = 1,
            title = "Percent of population fully vaccinated in California, by county",
            frame.lwd = 0,
            inner.margins = 0.1,
            legend.position = c("right", "top")) +
  tm_credits(caption, size = 0.5)
```




```
tm_shape(ca_covid) +
  tm_fill(col = "Median_Age",
          title = legend_title,
          palette = "Purples") +
  tm_compass(type = "8star",
             position = c("left", "bottom"),
             size = 2) +
  tm_layout(scale = 1,
            title = "Median age, by county",
            frame.lwd = 0,
            inner.margins = 0.1,
            legend.position = c("right", "top")) +
  tm_credits(caption, size = 0.5)
```



The map above shows the rates of full vaccination, by county, symbolized by a color ramp (from low to high rates of vaccination going from yellow to dark red). A color ramp was made with classes of 10 percent each. A second map shows the median age of each country, symbolized by a purple color ramp (younger to older median age going from light purple to darker purple). The five-step color ramp ranges from 30 to 55, with each step covering five years of age.

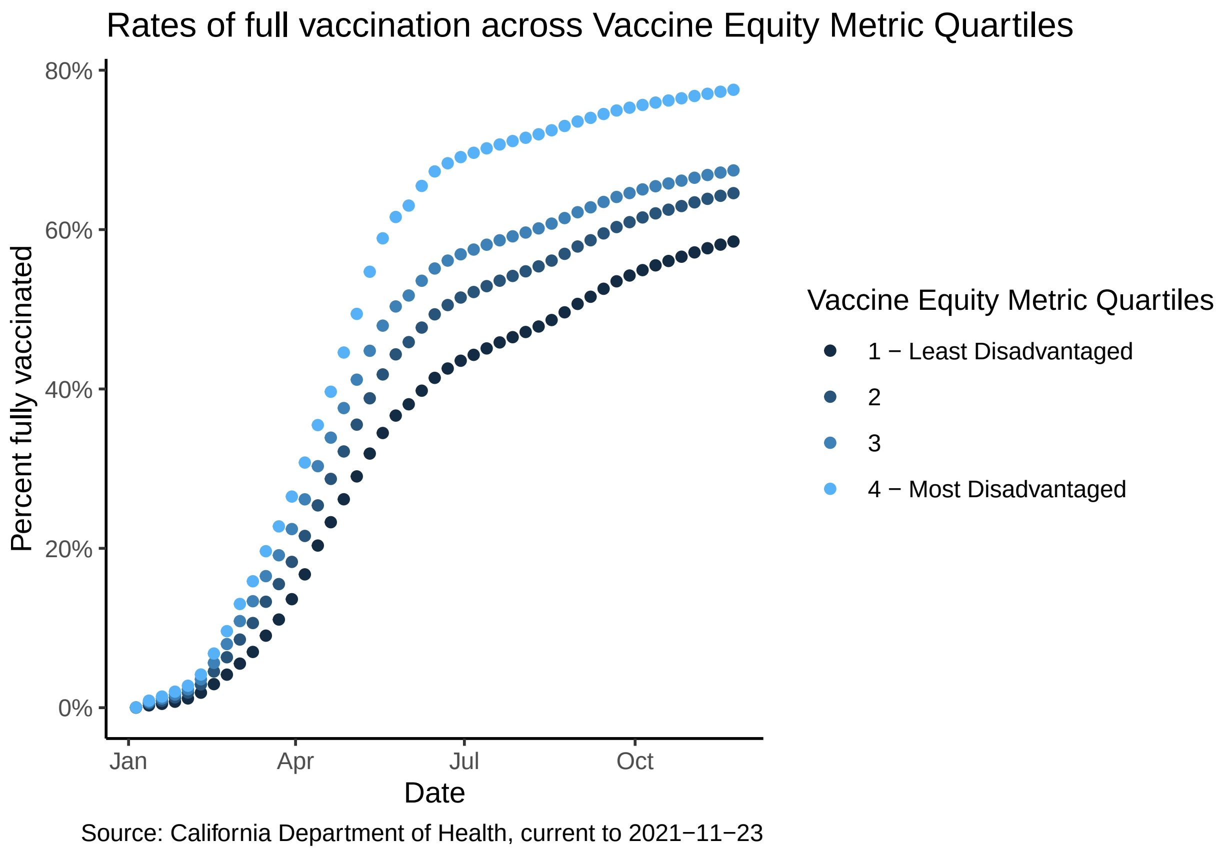
Vaccination rates over time (2021), by vaccine equity metric quartiles

```
#data per day at ZIP code level -> want state-wide totals per day
vax_longitudinal <- read.csv(file_path_vax) %>%
  group_by(county) %>%
  mutate(county_partial =
           sum(persons_partially_vaccinated, na.rm = T)/
           sum(age5_plus_population, na.rm =T),
         county_fully =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age5_plus_population, na.rm = T)) %>%
  ungroup() %>%
  mutate(persons_partially_vaccinated_2 =
           ifelse(is.na(persons_partially_vaccinated),
                  age5_plus_population*county_partial,
                  persons_partially_vaccinated),
         persons_fully_vaccinated_2 =
           ifelse(is.na(persons_fully_vaccinated),
                  age5_plus_population*county_fully,
                  persons_fully_vaccinated)) %>%
   group_by(county) %>%
  mutate(county_partial_2 =
           sum(persons_partially_vaccinated, na.rm = T)/
           sum(age5_plus_population, na.rm =T),
         county_fully_2 =
           sum(persons_fully_vaccinated, na.rm = T)/
           sum(age5_plus_population, na.rm = T),
         county_eligible_pop =
           sum(age5_plus_population, na.rm = T)) %>%
  ungroup()

vax_long_agg <- vax_longitudinal %>%
  group_by(as_of_date, vaccine_equity_metric_quartile) %>%
  summarize(percent_full_vem =
              sum(persons_fully_vaccinated, na.rm = T)/
              sum(age5_plus_population, na.rm = T)) %>%
  ungroup() %>%
  filter(is.na(vaccine_equity_metric_quartile) == FALSE) %>%
  mutate(date = as.Date.character(as_of_date)) %>%
  select(- as_of_date)
```

```
## `summarise()` has grouped output by 'as_of_date'. You can override using the `.groups` argument.
```

```
ggplot(data = vax_long_agg, aes(x = date, y = percent_full_vem)) +
  geom_point(stat = "identity", aes(col = vaccine_equity_metric_quartile)) +
  theme_classic() +
  labs(x = "Date",
       y = "Percent fully vaccinated",
       title = "Rates of full vaccination across Vaccine Equity Metric Quartiles", caption = caption) +
  guides(color=guide_legend(title="Vaccine Equity Metric Quartiles")) +
```

```
  scale_color_continuous(labels =
                           c("1 - Least Disadvantaged", "2", "3", "4 - Most Disadvantaged")) +
  scale_y_continuous(labels = scales::percent)
```



The visualization above is points plotted by month and percent of full vaccination in aggregate ZIP codes falling under four categories of the Vaccine Equity Metric (VEM) quartiles, which is based on demographic indicators of socioeconomic disadvantage and health vulnerability (where 1 is least disadvantaged/vulnerable and 4 and most disadvantaged/vulnerable). It can be seen that rates of vaccination follow, from highest to lowest, the order of vulnerability, from least to most – suggesting areas of improvement for equitable vaccine delivery.