

# Problem Set 4

Erica Criollo

2025-03-22

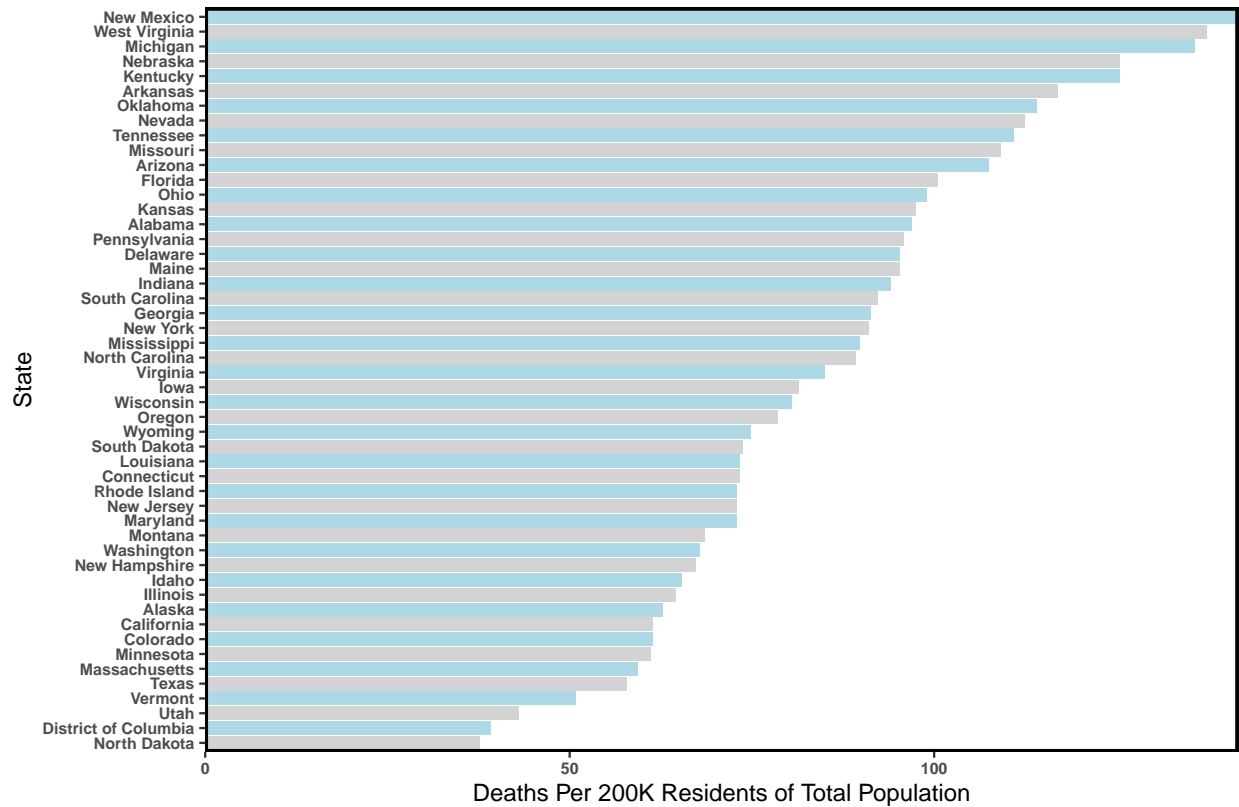
## 1. Summary

This report is on the COVID-19 pandemic, mask use, and vaccinations in counties and states across the United States from 2021 to 2022. The data is sourced from the New York Times who has been collecting and standardizing Covid data from hundreds of state and local sources since the beginning of the pandemic. The results shows that vaccinations are highly associated with fewer COVID-19 deaths while larger population size of a county is linked to more COVID-19 deaths.

## 2. Data

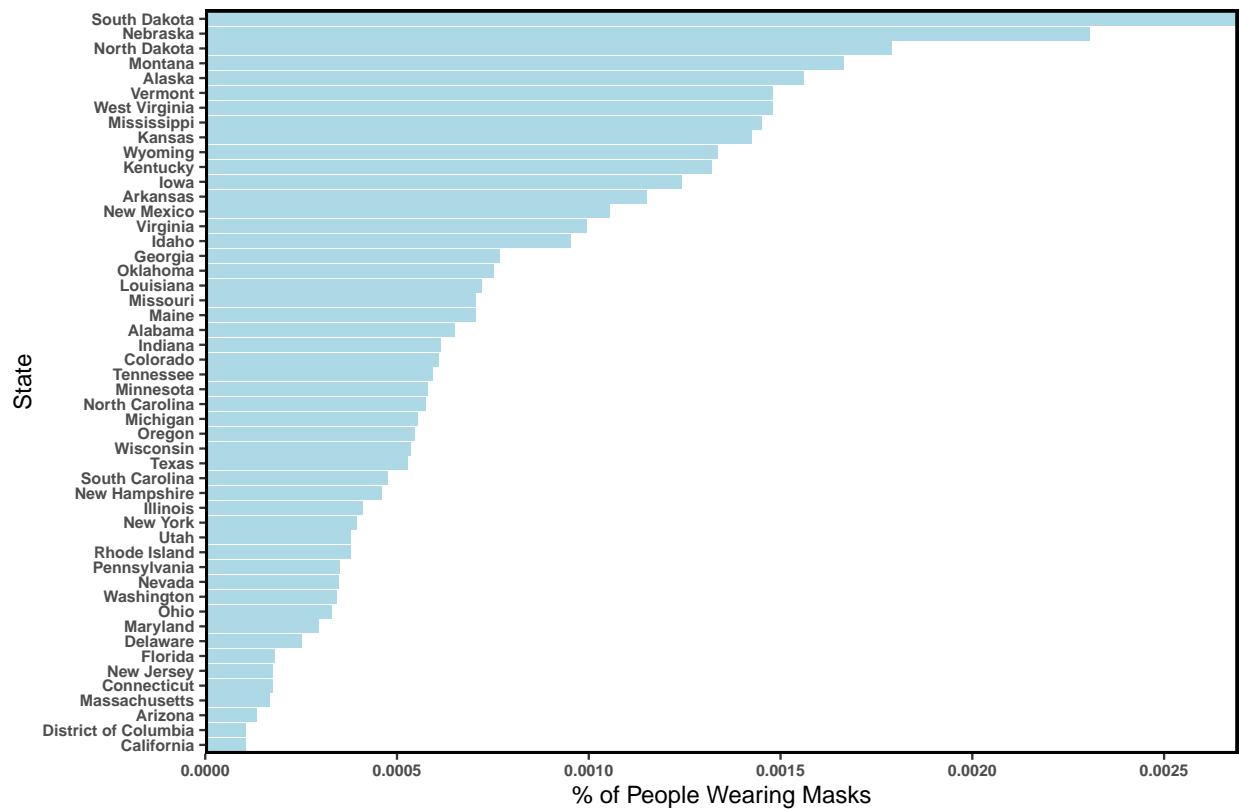
In Figure 1., the graph shows that the rates of new COVID-19 Deaths have been higher in states such as New Mexico, West Virginia, Michigan, and Nebraska. The lowest rates of new deaths include states such as North Dakota, D.C., Utah, and Vermont. We must add that the data was cleaned to remove state counties with unavailable population data which includes Puerto Rico, Hawaii, Northern Mariana Islands, and the Virgin Islands. The lowest rates of new deaths are 38 per 200,000 residents in a state and the highest rates were 142 per 200,000 residents in a state.

**Figure 1. Rates of New COVID-19 Deaths by State (2021–2022)**



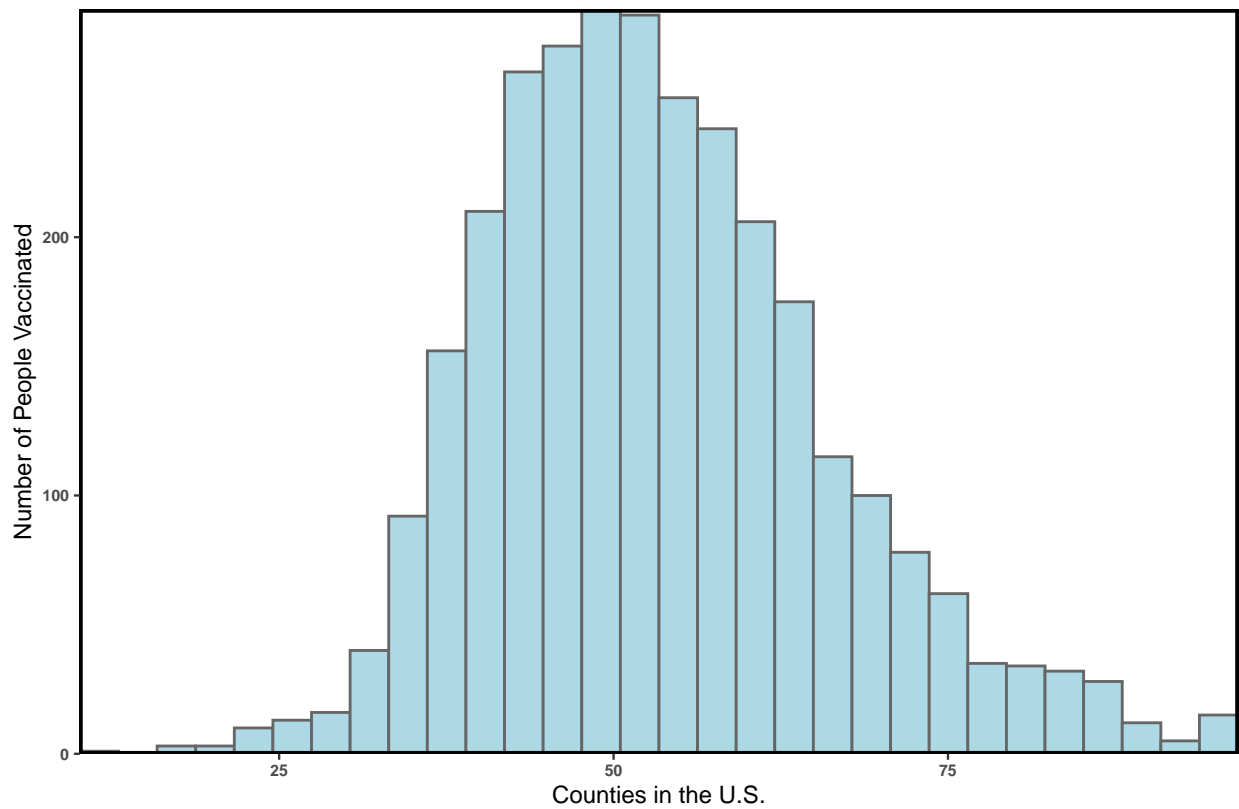
Regarding mask usage, Figure 2. reveals that states with the highest percent of people wearing masks includes South Dakota, Nebraska, North Dakota, and Montana. The lowest percent of states include California, D.C., Arizona, and Massachusetts.

**Figure 2. Mask Usage by State (July 2022)**



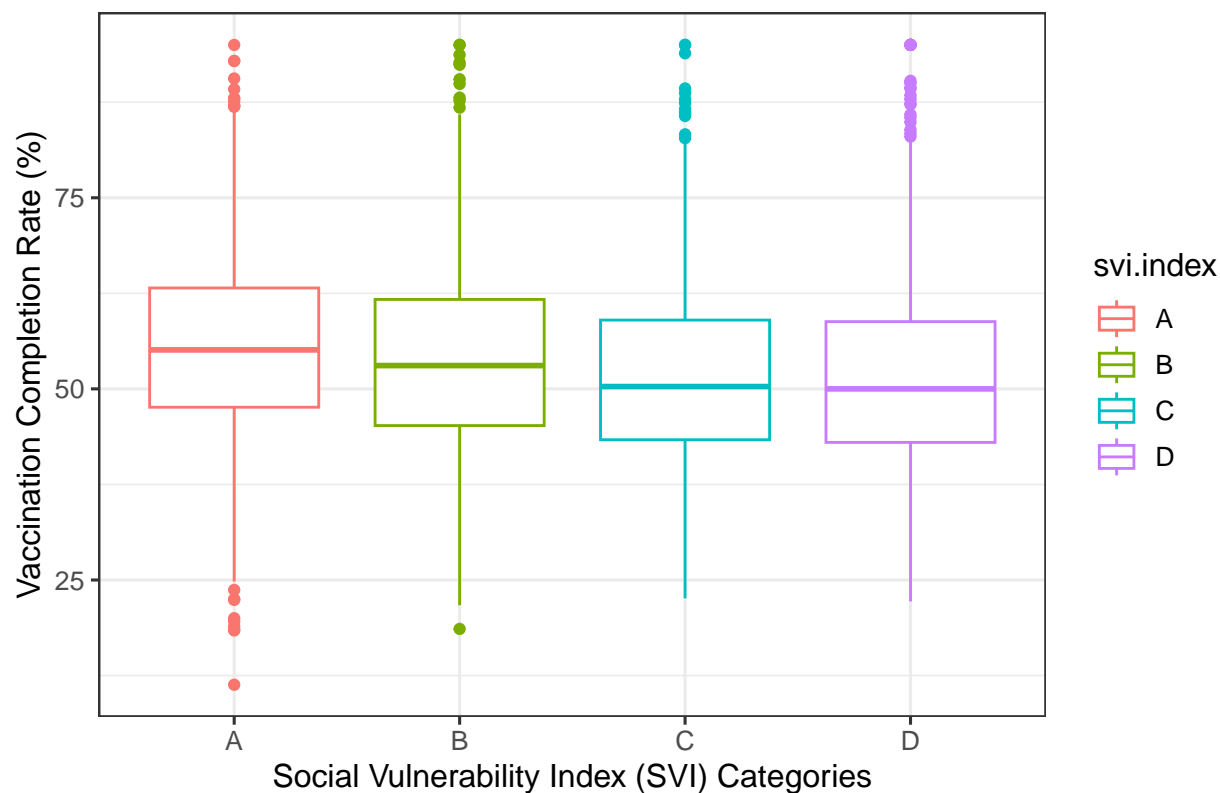
In Figure 3., it shows that the distribution of vaccination rates in the United states is normal.

**Figure 3. Vaccination Rates in the United Staes**



When considering the Social Vulnerability Index (SVI), there was a similar mean of vaccination completion across all levels of social vulnerability. Therefore, vaccinations seem accessible to all groups with slight decrease as the SVI index moves from A to D.

Figure 4. Vaccination Rates Among Different Social Vulnerability Index



```
## List of 6
## $ axis.title      :List of 11
##   ..$ family      : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : num 12
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## $ axis.text       :List of 11
##   ..$ family      : NULL
##   ..$ face         : chr "bold"
##   ..$ colour       : NULL
##   ..$ size         : num 6
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
```

```
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##   $ legend.position: chr "none"
##   $ panel.border    :List of 5
##   ..$ fill          : NULL
##   ..$ colour        : chr "black"
##   ..$ linewidth     : num 1
##   ..$ linetype      : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##   $ panel.grid      : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##   $ plot.title      :List of 11
##   ..$ family        : NULL
##   ..$ face          : chr "bold"
##   ..$ colour        : NULL
##   ..$ size          : num 12
##   ..$ hjust         : num 0.5
##   ..$ vjust         : NULL
##   ..$ angle         : NULL
##   ..$ lineheight    : NULL
##   ..$ margin        : NULL
##   ..$ debug         : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
## - attr(*, "class")= chr [1:2] "theme" "gg"
## - attr(*, "complete")= logi FALSE
## - attr(*, "validate")= logi TRUE
```

State	County	Vaccination Rate
Arizona	Apache	95.0
Arizona	Santa Cruz	95.0
California	Imperial	95.0
Colorado	San Juan	95.0
Georgia	Chattahoochee	95.0
North Dakota	Slope	11.3
Louisiana	Cameron	18.4
Montana	McCone	18.5
Ohio	Holmes	18.6
Texas	Loving	18.9

### 3. Analysis

In the regression, it suggests that higher vaccination rates have a high association with fewer COVID-19 deaths, given that there is a statistically significant negative relationship. Population size is positively associated with COVID-19 deaths which may be due to the spread of the disease through highly populated cities/areas that leads to more deaths.

Table 2: Impact of Mask-Wearing and Vaccination on COVID-19 Deaths (2022)

	m1	m2	m3
Mask-Wearing Rate	-19.411 (14.715)		316.890*** (45.220)
Vaccination Rate		-0.636*** (0.135)	3.239*** (0.412)
Population Size	0.001*** (0.000)	0.001*** (0.000)	
Num.Obs.	3049	3049	3049

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## Appendix: Replication code

```
knitr::opts_chunk$set(
  echo = FALSE,
  eval = TRUE,
  fig.align = 'center',
  message = FALSE,
  warning = FALSE
)

#Loading packages and importing data

library(tidyverse)
library(scales)
library(lfe)
library(modelsummary)
library(gt)
library(data.table)
library(scales)

# create the dataset -----
## 2021-2022 deaths (BIG FILES) *This reports each county's new deaths between 2021-2022*
covid <-
  data.table::fread('https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2022.')
  filter(!is.na(fips),
    !state %in% c('Puerto Rico', 'Hawaii', 'Northern Mariana Islands', 'Virgin Islands')) %>%
  select(fips, county, state, date, deaths) %>%
  group_by(fips, county, state) %>%
  summarise(deaths = max(deaths, na.rm = T) - min(deaths, na.rm = T), .groups = "drop")

## estimated mask usage from July 2020 survey
mask <-
  read_csv('https://raw.githubusercontent.com/nytimes/covid-19-data/master/mask-use/mask-use-by-county')
  mutate(
    fips = as.integer(COUNTYFP),
```

```

    always.mask = ALWAYS, #always masking *Not sure if this is correct*
    .keep = 'none'
  ) # for merging

## prep CDC data from directory
vax <-
  read_csv('cdc vax mar1.csv') %>%
  filter( # drop unknown/incomplete/questionable reports
    FIPS != 'UNK',
    Recip_State != 'VI',
    Completeness_pct > 0,
    !is.na(Administered_Dose1_Recip)
  ) %>%
  mutate(
    fips = as.integer(FIPS),
    population = Census2019,
    vax.complete = Series_Complete_Pop_Pct, # percent vaxd
    svi.index = SVI_CTGY, # social vulnerability index
    .keep = 'none'
  )

## merge
covid <-
  left_join(covid, mask) %>%
  left_join(vax) %>%
  #mutate(deaths.scaled = deaths / population * 100000) %>% *Not doing this anymore*
  ungroup() # scale by population

rm(mask, vax)

summary(covid)
# *The covid dataset now contains:County-level cumulative COVID deaths (adjusted for population size),

#Figure 1

new_covid <- covid %>%
  group_by(state) %>% # Instead I am using states only as county names are shared by several states
  summarise(
    total_deaths = sum(deaths, na.rm = TRUE), # Adding all new deaths in each state
    total_population = sum(population, na.rm = TRUE)
  ) %>%
  mutate(deaths_per_200k = (total_deaths / total_population) * 100000) %>% # Comparison with other s
  arrange(desc(deaths_per_200k)) %>%
  mutate(color_group = factor(row_number() %% 2)) #

new_covid %>%
  ggplot(aes(x = reorder(state, deaths_per_200k), y = deaths_per_200k)) +
  geom_col(aes(fill = color_group)) + #Making the shift in colors
  coord_flip() +
  scale_y_continuous(expand = c(0, 0)) +
  scale_fill_manual(values = c("1" = "lightblue", "0" = "lightgray")) + # alternating colors
  labs(

```



```

    title = 'Figure 1. Rates of New COVID-19 Deaths by State (2021-2022)',
    x = 'State',
    y = 'Deaths Per 200K Residents of Total Population'
) +
theme_bw(base_size = 9) +
theme(
  plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
  plot.caption = element_text(
    family = 'Times New Roman',
    hjust = 0.5,
    size = 15,
    face = "bold"
  ),
  axis.text = element_text(size = 6, face = 'bold'),
  #axis.title = element_text(size = 10, face = 'bold'),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  panel.border = element_rect(colour = "black", size = 1) ,
  legend.position = "none"
)

## stat summary and more
# summary(new_covid$deaths_per_200k)

# find some examples? #

# Mask usage ----- #Make this better
## VIZ: "Always wears a mask"

# Figure 2

covid %>%
  group_by(state) %>%
  summarise(
    mask_wearers = sum(always.mask, na.rm = TRUE), # Group by state and percent of mask wearers in ea
    total_population = sum(population, na.rm = TRUE)
  ) %>%
  mutate(
    mask_percentage = (mask_wearers / total_population) * 100 # percent of mask wearers
  ) %>%
  arrange(desc(mask_percentage)) %>%
  ggplot(aes(x = reorder(state, mask_percentage), y = mask_percentage)) +
  scale_y_continuous(expand = c(0, 0)) +
  geom_col(fill = "lightblue") + #
  coord_flip() +
  labs(
    title = "Figure 2. Mask Usage by State (July 2022)",
    x = "State",
    y = "% of People Wearing Masks"
  ) +
  theme_bw(base_size = 9) +
  theme(
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5),

```

```

    plot.caption = element_text(
      family = 'Times New Roman',
      hjust = 0.5,
      size = 15,
      face = "bold"
    ),
    axis.text = element_text(size = 6, face = 'bold'),
    #axis.title = element_text(size = 10, face = 'bold'),
    panel.grid = element_blank(),
    panel.background = element_blank(),
    panel.border = element_rect(colour = "black", size = 1) ,
    legend.position = "none"
  )
)

tab1<- table(covid$always.mask)
## helpers
# summary(covid$always.mask)
# find hi/lo counties?

# Rates of vaccination -----

## VIZ: overall vax rates
covid %>%
  ggplot(aes(x = vax.complete)) +
  geom_histogram(color = 'gray40', fill = 'lightblue') +
  labs(
    title = "Figure 3. Vaccination Rates in the United Staes",
    x = "Counties in the U.S.",
    y = "Number of People Vaccinated"
  ) +
  theme_bw(base_size = 9) +
  theme(
    plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
    plot.caption = element_text(
      family = 'Times New Roman',
      hjust = 0.5,
      size = 15,
      face = "bold"
    ),
    axis.text = element_text(size = 6, face = 'bold'),
    panel.grid = element_blank(),
    panel.background = element_blank(),
    panel.border = element_rect(colour = "black", size = 1) ,
    legend.position = "none"
  ) +
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0))

## VIZ: vax rates by Social Vulnerability Index category
covid %>%
  filter(!is.na(svi.index)) %>%
  ggplot(aes(y = vax.complete, x = svi.index, color = svi.index)) +
  geom_boxplot()+

```

```

labs(
  title = "Figure 4. Vaccination Rates Among Different Social Vulnerability Indices",
  x = "Social Vulnerability Index (SVI) Categories",
  y = "Vaccination Completion Rate (%)",
  fill = "SVI Category"
) +
theme_bw(base_size = 12)
theme(
  plot.title = element_text(face = "bold", size = 12, hjust = 0.5),
  axis.text = element_text(size = 6, face = "bold"),
  axis.title = element_text(size = 12),
  panel.grid = element_blank(),
  panel.border = element_rect(colour = "black", size = 1),
  legend.position = "none"
)

## find high/low counties
covid %>%
  select(State = state, County = county, Vaccination_Rate = vax.complete) %>%
  arrange(desc(Vaccination_Rate)) %>%
  slice_head(n = 5) %>%
  bind_rows(
    covid %>%
      select(State = state, County = county, Vaccination_Rate = vax.complete) %>%
      arrange(Vaccination_Rate) %>%
      slice_head(n = 5)
  ) %>%
  knitr::kable(digits = 2, col.names = c("State", "County", "Vaccination Rate"))

# Impact on 2022 COVID deaths -----
## regression estimates
mods <-
  list(
    m1 = felm(deaths ~ always.mask + population + svi.index | state, data = covid),
    m2 = felm(deaths ~ vax.complete + population + svi.index | state, data = covid),
    m3 = felm(deaths ~ always.mask + vax.complete + svi.index | state, data = covid)
  )

## regression table

modelsummary(
  mods,
  gof_map = c('nobs'),
  stars = TRUE,
  coef_map = c(
    "always.mask" = "Mask-Wearing Rate",
    "vax.complete" = "Vaccination Rate",
    "population" = "Population Size",
    "svi.index" = "Social Vulnerability Index"
  ),
  output = 'gt',
  title = "Impact of Mask-Wearing and Vaccination on COVID-19 Deaths (2022)"
)

```