Code Chunk

Karen Perez Torres

2024-05-10

```r
# packages
  library(tidyverse)
  library(scales)
  library(lfe)
  library(stargazer)
  library(haven)
  library(ggplot2)
  library(janitor)
  library(knitr)
  library(modelsummary)
  library(gt)
```

```r
# create the data set ------------------
## 2021-2022 deaths (BIG FILES)
    covid = data.table::fread('https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counti
    filter(!is.na(fips), state != 'Puerto Rico') |>
    select(fips, county, state, date, deaths) |>
    group_by(fips, county, state) |>
    summarise(deaths = max(deaths, na.rm = T) - min(deaths, na.rm = T))
```

```r
view(covid)
```

```r
## estimated mask usage from July 2020 survey
  mask = read_csv('https://raw.githubusercontent.com/nytimes/covid-19-data/master/mask-use/mask-use-by-
    mutate(
      fips = as.integer(COUNTYFP),
      always.mask = ALWAYS, #always masking
      .keep = 'none'
    )# for merging
```

```r
view(mask)
```

```r
## prep CDC data from directory
  vax = read_csv('cdc vax mar1.csv') |>
    filter(
      FIPS != 'UNK',
      Recip_State != 'VI',
      Completeness_pct > 0,
      !is.na(Administered_Dose1_Recip)
    ) |> # drop unknown/incomplete/questionable reports
    mutate(
```

```
      fips = as.integer(FIPS),
      population = Census2019,
      vax.complete = Series_Complete_Pop_Pct, # percent vaccinated
      svi.index = SVI_CTGY, # social vulnerability index
      .keep = 'none'
    )
```

```
## merge
  covid =
    left_join(covid, mask) |>
    left_join(vax) |>
    mutate(deaths.scaled = deaths / population * 100000) |>
    ungroup() # scale by population
```

```
 ## add regions to states

  covid <- covid |>
    mutate(region = case_when((state== "Alaska" | state== "Arizona" | state== "Colorado" | state== "Idal
                                  state== "New Mexico" | state== "Montana" | state== "Utah" | state== "Ne
                                  state== "Wyoming" | state== "California" | state== "Hawaii" | state==
                                  state== "Washington") ~ "West",
                               (state== "Connecticut" | state== "Maine" | state== "Massachusetts" | state
                                  state== "Rhode Island" | state== "Vermont" | state== "New Jersey" | sta
                                  state== "Pennsylvania") ~ "Northeast",
                               (state== "Indiana" | state== "Illinois" | state== "Michigan" | state== "Oh
                                  state== "Wisconsin" | state== "Iowa" | state== "Kansas" | state== "Minn
                                  state== "Missouri" | state== "Nebraska" | state== "North Dakota" | stat
                               ~ "Midwest",
                               (state== "Delaware" | state== "District of Columbia" | state== "Florida"
                                  state== "Maryland" | state== "North Carolina" | state== "South Carolina
                                  state== "West Virginia" | state== "Alabama" | state== "Kentucky" | stat
                                  state== "Tennessee" | state== "Arkansas" | state== "Louisiana" | state=
                                  state== "Texas") ~ "South"))
```

```
view(covid)
```

Introduction —

In 2020, COVID-19 swept the world. In response, the United States began an aggressive program of public masking and vaccine development in order to limit deaths, among other programs. By 2022, there is enough evidence by which to evaluate the success of these programs.

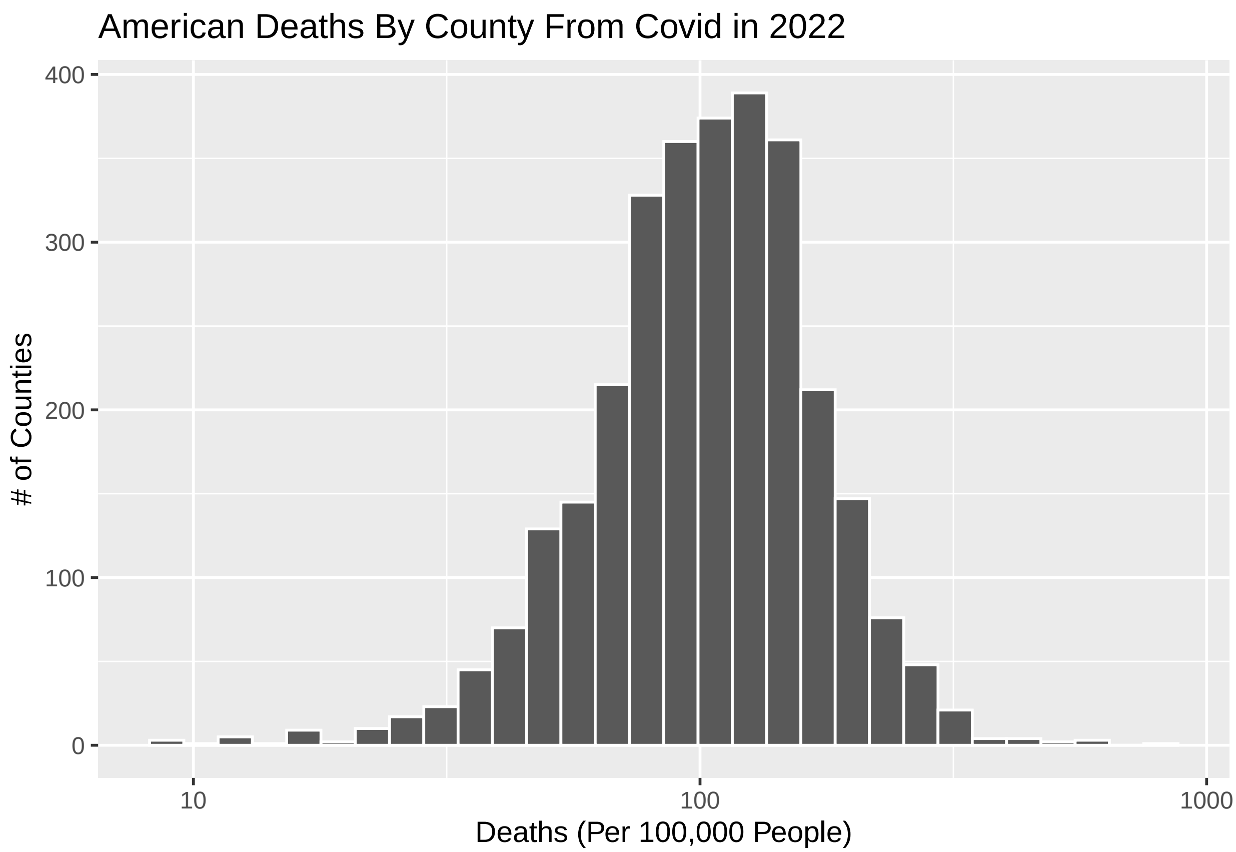
Below you will find examinations of Deaths per 100,000 People, Masking, Vaccinations, and Social Vulnerability at the National and Regional levels. Regression Analysis has also been performed to identify significant relationships between preventative measures or social factors and rates of death.

2022 county-level Death and Masking rates have been sourced from the New York Times. Vaccination and Social Vulnerability Classifications come from the CDC.

Overall Findings: Masks and Vaccines have a significant impact on mitigating deaths from COVID. This held true both nationally and regionally. We can also assess that socially vulnerable populations were more susceptible to the effects of COVID. These results see some variation at the regional level, indicating that geography or population density may play a role.

Covid Deaths in 2022 —

```
## VIZ1 using deaths.scaled to account for population differences
 covid |>
   ggplot(aes(x = deaths.scaled)) +
   geom_histogram(color = 'white') +
   scale_x_log10() + # to mitigate skew
   labs(title='American Deaths By County From Covid in 2022',x="Deaths (Per 100,000 People)",y='# of Co
```

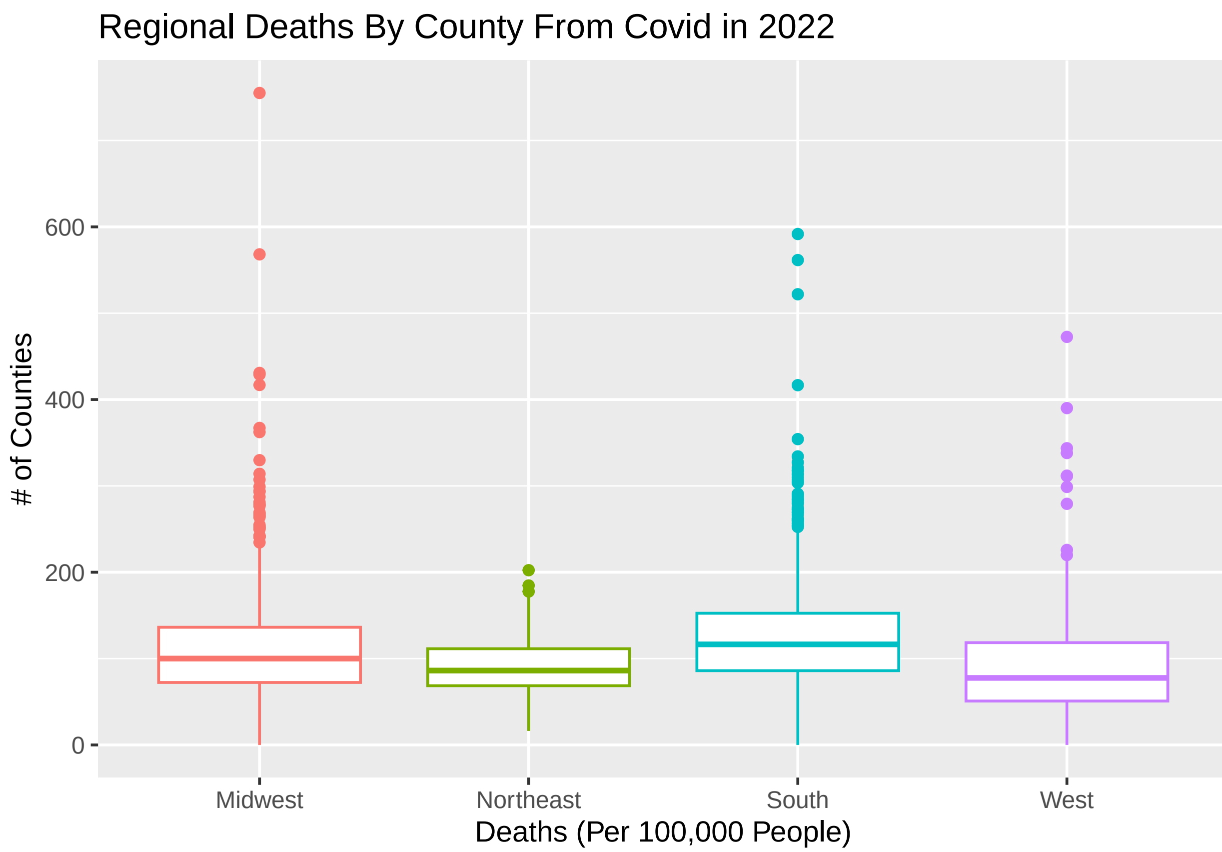


Deaths per 100,000 people was calculated by dividing the total deaths per county by the county's population, then multiplying by 100,000.

Looking at the data nationally for 2022, counties saw an average of 113 deaths per 100,000 due to Covid-19. For comparison, this cause of death is below heart disease, cancer, and accidents, but higher than strokes, Alzheimer's disease and diabetes.

```
# American Deaths by Region
filter(covid,!is.na(region)) |>
    ggplot(aes(y = deaths.scaled, x = region, color = region)) +
    geom_boxplot() +
```

```
  theme(legend.position="none") +
  labs(title='Regional Deaths By County From Covid in 2022',x="Deaths (Per 100,000 People)",y='# of C
```



In our regional analysis, we compared the Midwest, Northeast, South, and West.

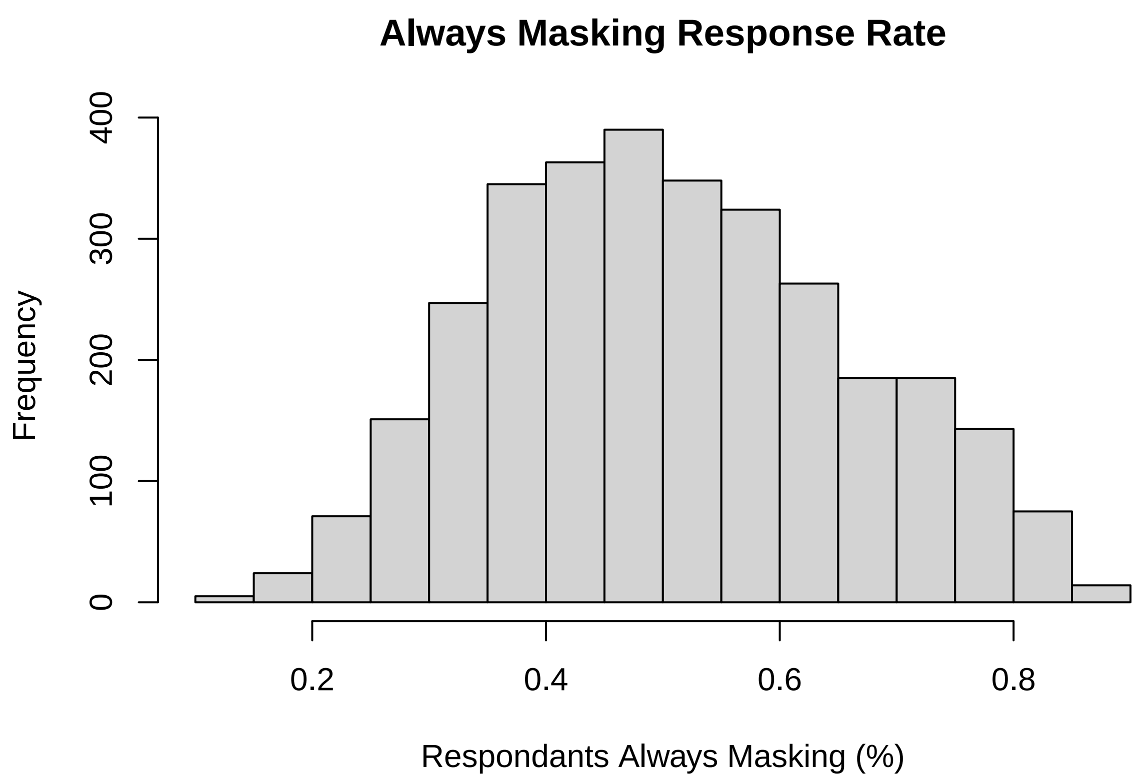
Of these counties, the South saw the highest rates of death, while the West saw the lowest. As we will see later, this did not entirely correspond to rates of masking or vaccination.

The worst rates of death were seen in Kingman County in Kansas, at 755 deaths per 100,000 people. Counties like Harding, New Mexico and Kennedy, Texas did not have any deaths recorded, although they did see hospitalizations.

To note: At the national level, a mean of 113 deaths per 100,000 vs a median of 104 indicates that a small number of counties with high death rates pulled the average rate up.

Masking in America —

```r
hist(covid$always.mask,
   main = "Always Masking Response Rate",
   xlab = "Respondants Always Masking (%)")
```

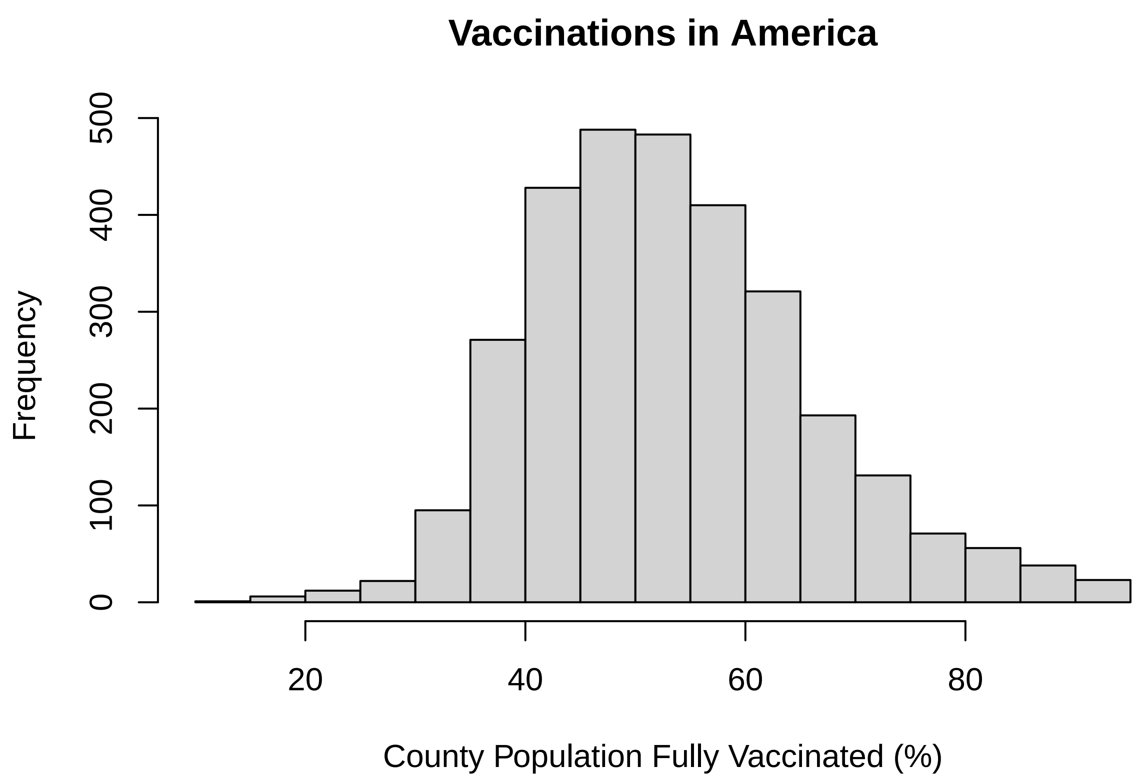


Respondents were asked at the county level if they always used a mask during their activities, versus less stringent mask usage habits.

Across all US counties, 51% of the population reported that they always wore masks during 2022. The lowest percentage of folks who reported that they always masked were in the county of Valley, Montana at 12% and the people who reported always using masks were at a high of 89% in the county of Inyo, California. The middle 50% of counties saw their mask rate fall between 39% and 61%.

Rates of Vaccination —

```r
hist(covid$vax.complete,
     main = "Vaccinations in America",
     xlab = "County Population Fully Vaccinated (%)")
```



Vaccinations are recorded as the percentage of the county population that was fully up to date on their vaccines at the time of data collection.

Across all US counties, 53% of the population reported that they were fully vaccinated during 2022. The lowest percentage of folks who reported that they completed their vaccination was in the county of Slope, North Dakota at 11% and the people who reported complete vaccination was at a high of 95% in the county of Apache, Arizona. The middle 50% of counties saw their full vaccination rate fall between 44% and 61%.

After evaluating each region separately (Appendix A), certain differences were noted against the national trend.

In the South, deaths per 100,000 was far more than the national average; with 125 deaths (an increase of 12 deaths).

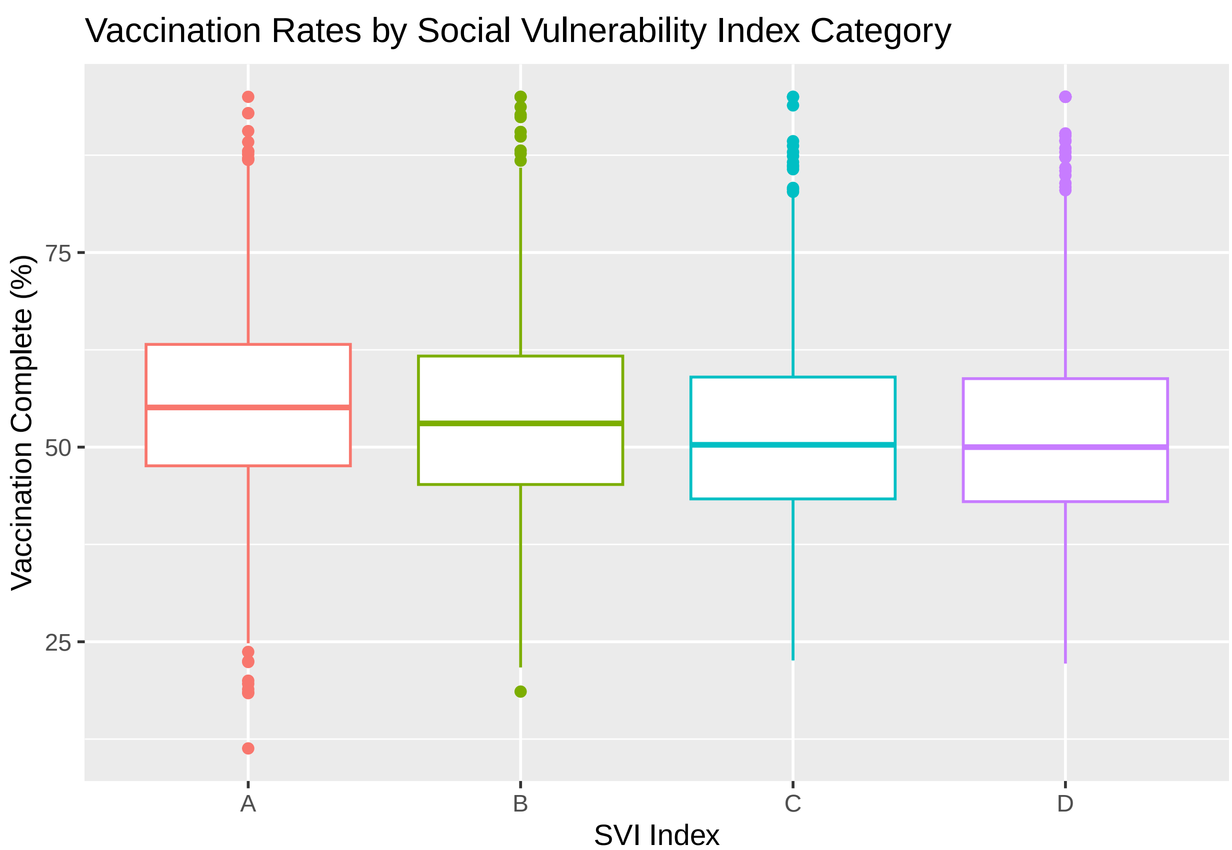
In the Midwest, there were less deaths than the national average with 110 per 100,000 people and was on par with the national average of 53% of the county's population fully vaccinated.

In the Western Region, counties reported 89.8 deaths per 100,000 people (the lowest of all the regions) with the North East region trailing behind with 92 reported deaths. The North East region had the highest reported complete vaccinations (67%) and always use of masks at 71%.

A deeper statistical analysis of counties may reveal additional trends at the national and regional level.

Social Vulnerability —

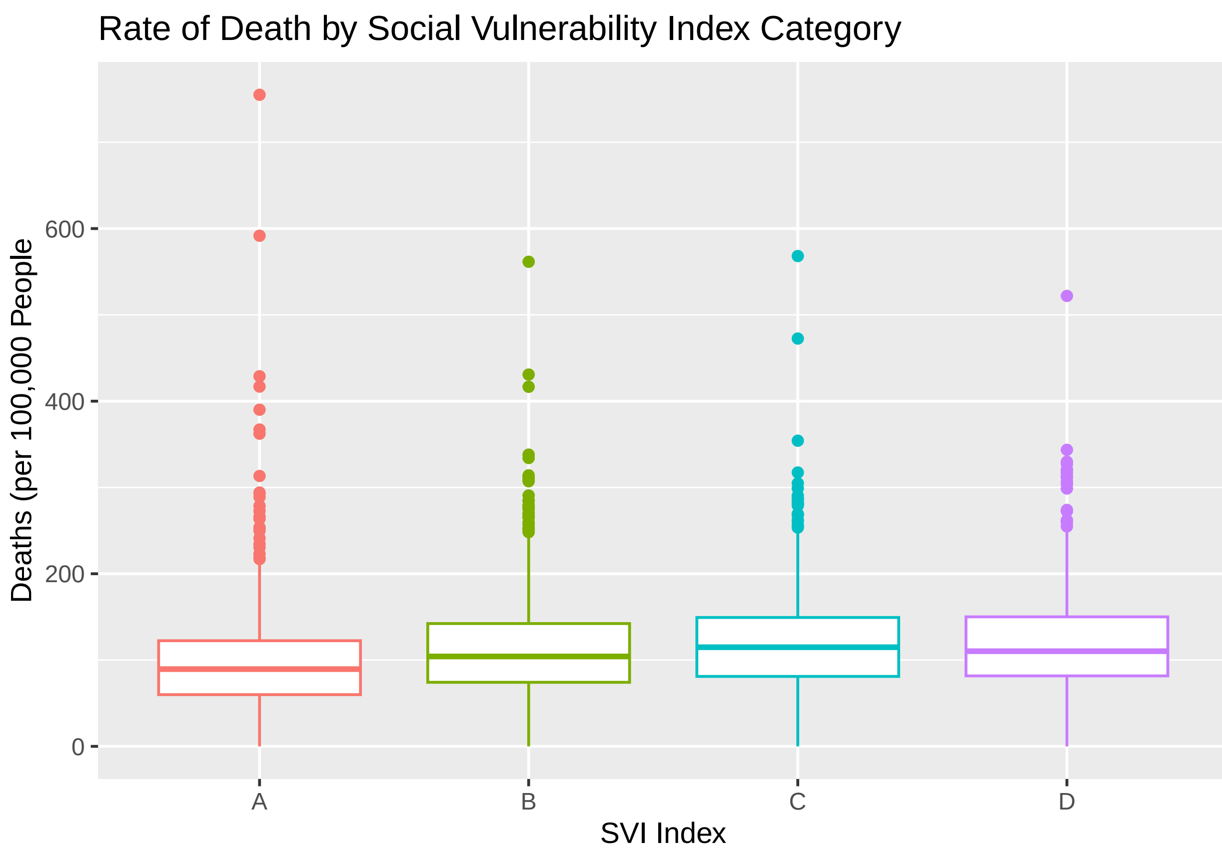
```
filter(covid,!is.na(svi.index)) |>
    ggplot(aes(y = vax.complete, x = svi.index, color = svi.index)) +
    geom_boxplot()+
    theme(legend.position="none") +
    labs(title='Vaccination Rates by Social Vulnerability Index Category',x="SVI Index",y='Vaccination (
```



A final variable of interest is the social vulnerability index. Ranign from Class A to Class D, this is a composite variable assembled by the CDC to evaluate a population's exposure to diseases, natural disasters, and other calamities. It primarily draws from economic and health indicators.

A clear decline in vaccinations can be witnessed from Class A to Class D, indicating that more vulnerable populations had less access or interest in the vaccination.

```
filter(covid,!is.na(svi.index)) |>
    ggplot(aes(y = deaths.scaled, x = svi.index, color = svi.index)) +
    geom_boxplot()+
    theme(legend.position="none") +
    labs(title='Rate of Death by Social Vulnerability Index Category',x="SVI Index",y='Deaths (per 100,(
```



This fall in vaccinations visually corresponds to rates of death, as more vulnerable populations see higher rates of death. This connection is explored further in our regression analysis.

Regression Analysis —

Combined National and Regional Models

```
scovid <- filter(covid, region == 'South')
ncovid <- filter(covid, region == 'Northeast')
mcovid <- filter(covid, region == 'Midwest')
wcovid <- filter(covid, region == 'West')
```

```
combined = list(
    national = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = covid),
    south = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = scovid),
    northeast = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = ncovid),
    west = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = wcovid),
    midwest = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = mcovid)
)
```

```
stargazer::stargazer(
    combined,
    header = FALSE,
      column.labels = c("National", "South", "Northeast", "West", "Midwest"),
    keep.stat = 'n', type = 'text', # change to html in Rmd
    add.lines = list(c('State fixed effects','Yes','Yes','Yes'))
)
```

```
##
## ==========================================================================
##                                      Dependent variable:
##                      -----------------------------------------------------
##                                         deaths.scaled
##                      National    South    Northeast   West     Midwest
##                        (1)        (2)        (3)       (4)       (5)
## --------------------------------------------------------------------------
## always.mask          -83.296*** -98.824*** -60.681**  -21.789  -99.717***
##                       (10.834)   (16.258)  (25.486)  (24.487)  (20.767)
##
## vax.complete         -0.923***  -1.169***  -0.885*** -0.997***  -0.465**
##                       (0.099)    (0.152)    (0.195)   (0.195)   (0.207)
##
## svi.indexB            9.581***   13.098**    5.421   20.747***   7.947*
##                       (2.859)    (6.048)    (4.546)   (7.634)   (4.313)
##
## svi.indexC           13.128***  19.051***    3.772   22.120***   8.364
##                       (3.080)    (5.974)    (5.262)   (7.907)   (5.221)
##
## svi.indexD           17.491***  21.000***    2.780   35.156***   12.362
##                       (3.367)    (5.919)    (8.610)   (8.571)   (8.008)
##
## --------------------------------------------------------------------------
## State fixed effects     Yes        Yes        Yes
## Observations           3,049      1,422       209       431       987
## ==========================================================================
## Note:                                        *p<0.1; **p<0.05; ***p<0.01
```

When masking and vaccinations are considered together, we see that both are still statistically significant at all levels, but are reduced from models in which they are the only treatment evaluated. Nationally, the model predicts that masking would prevent 83 deaths if a county went from 0% to 100% masking, and for every percentage of a population vaccinated, deaths/100K decline by one person.

Evaluating the impact of masking and vaccinations regionally, the South sees greater impacts from both of these activities, both against the national rate and other regions. The impact of masking in this region prevents an additional 15 deaths per 100,000, and each percentage of fully vaccinated persons prevent another .17 deaths, over the national.

Why is this? The South sees lower vaccination rates than the national average, but its masking was actually better. This warrants further investigation into possible factors driving mask and vaccination efficacy.

In evaluating the impact of social vulnerability, we can see that each step up on the vulnerability ladder (from Class A to D) leads to roughly an additional 20 deaths per 100,000. This is true at the National, South, and West regions, but was not statistically significant in the Midwest or Northeast region.

What could cause social vulnerability to not have an impact on deaths in the Midwest or Northeast? Additional research will be required, starting with social safety nets or aggressive campaigns in these regions to reach more vulnerable populations, mitigating their disadvantages.

Not included in this report, we built models that evaluated the impact of the population of counties on the rates of death. Their effect was less significant and/or had only a small impact on the number of deaths per 100,000 people. At the national level, this translated to an additional 10,000 people in a county would see a decline in deaths by .2 persons. Perhaps by investigating the population density of a county, we could draw more specific conclusions in regards to its impact on death rates.

Appendix A —

Summary of Deaths / 100K , Masking, and Vaccinations

National Consolidated —

```
# Deaths
  summary(covid$deaths.scaled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    0.00   74.05  103.60  113.16  142.16  755.03      93
```

```
# Masking
  summary(covid$always.mask)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##  0.1150  0.3930  0.4970  0.5077  0.6130  0.8890       9
```

```
# Vaccinations
  summary(covid$vax.complete)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   11.30   44.30   52.10   53.43   61.00   95.00      93
```

Regional Differences —

South

```
scovid <- filter(covid, region == 'South')
summary(scovid$deaths.scaled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    0.00   86.04  116.51  125.46  152.50  591.72       1
```

```
summary(scovid$always.mask)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##  0.1770  0.4480  0.5270  0.5368  0.6178  0.8800       1
```

```
summary(scovid$vax.complete)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    18.4    42.1    48.6    50.0    56.3    95.0       1
```

Northeast

```r
ncovid <- filter(covid, region == 'Northeast')
summary(ncovid$deaths.scaled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   16.23   68.61   86.17   92.31  111.43  202.38       3
```

```r
summary(ncovid$always.mask)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##  0.4470  0.6390  0.7345  0.7128  0.7880  0.8840
```

```r
summary(ncovid$vax.complete)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   37.70   59.10   66.60   67.03   75.80   92.90       3
```

Midwest

```r
mcovid <- filter(covid, region == 'Midwest')
summary(mcovid$deaths.scaled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    0.00   72.34  100.05  110.07  136.26  755.03      68
```

```r
summary(mcovid$always.mask)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   0.155   0.328   0.403   0.415   0.494   0.788
```

```r
summary(mcovid$vax.complete)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   11.30   46.00   53.40   53.38   60.20   95.00      68
```

West

```r
wcovid <- filter(covid, region == 'West')
summary(wcovid$deaths.scaled)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    0.00   50.88   77.63   89.78  118.47  472.56      15
```

```r
summary(wcovid$always.mask)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##  0.1150  0.3847  0.5635  0.5369  0.6850  0.8890       2
```

```r
summary(wcovid$vax.complete)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   18.50   46.60   56.50   58.27   69.35   95.00      15
```