PS4

Nicholas Rangel

2025-03-20

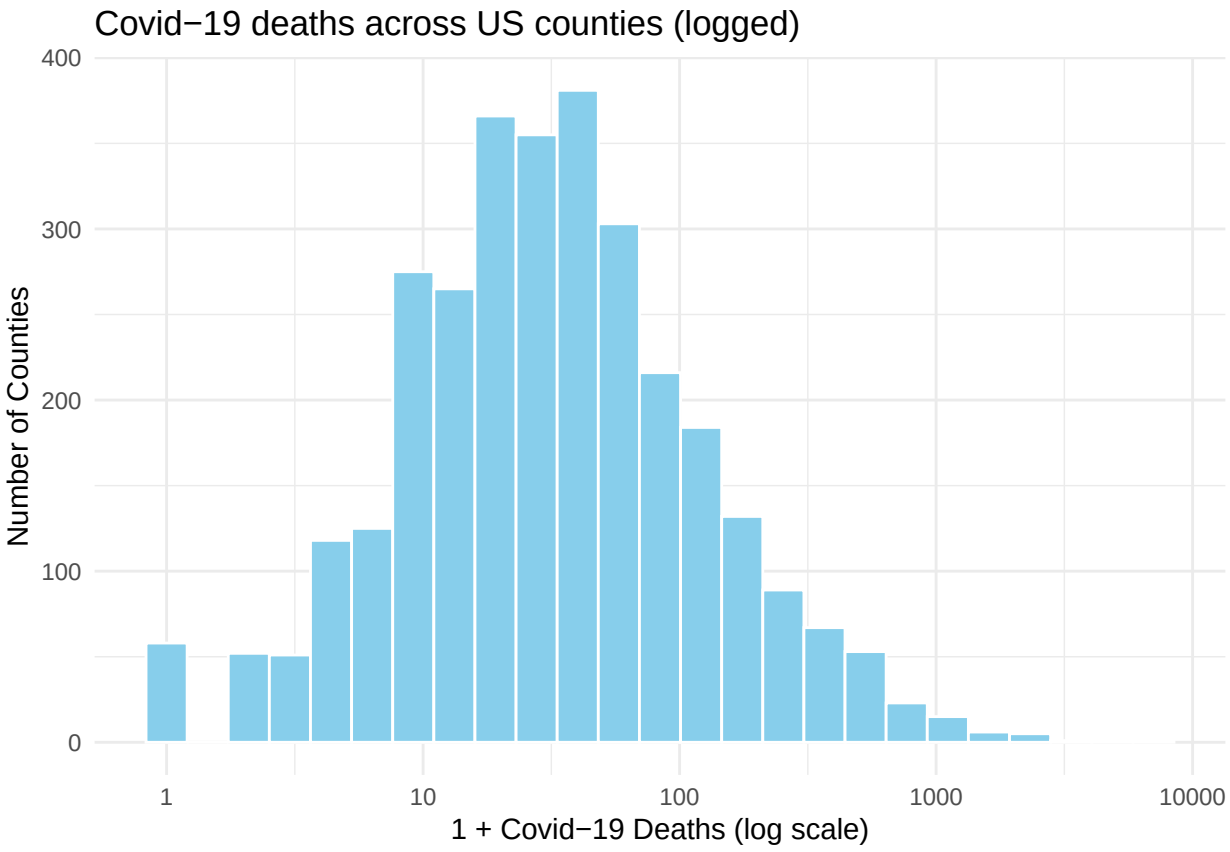
My Responses

The following report analyzes data on COVID-19 deaths on a county level, along with mask usage, and vaccination rates across the United States. The goal is to determine how mask-wearing and vaccination impact COVID-19 death rates in 2022 by using regression models and visualizations. Key findings include:

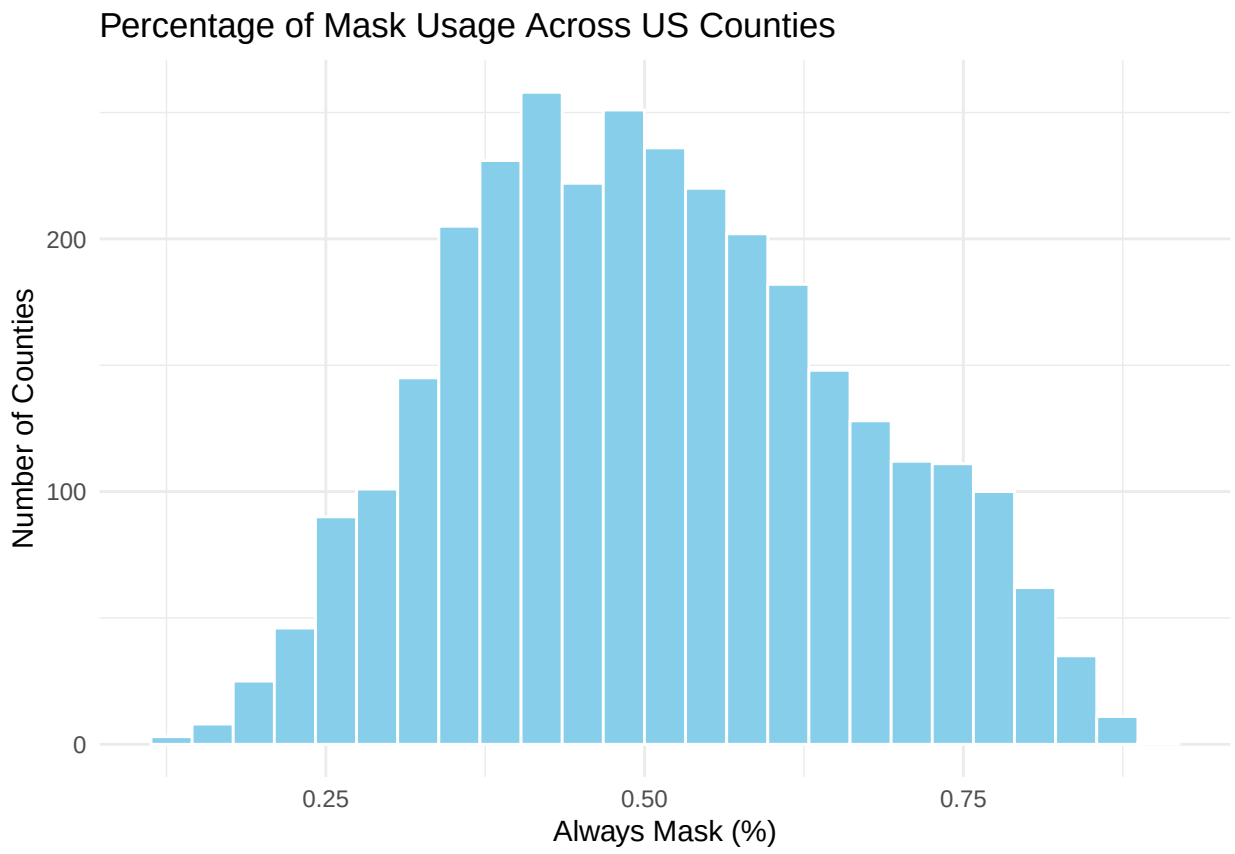
- A wide variation in Covid-19 deaths across counties, with the distribution heavily skewed.
- Differences in the percentage of people who always wear a mask.
- Significant variation in vaccination rates that, along with mask use, help explain differences in death rates.

The analysis uses three primary data sources:

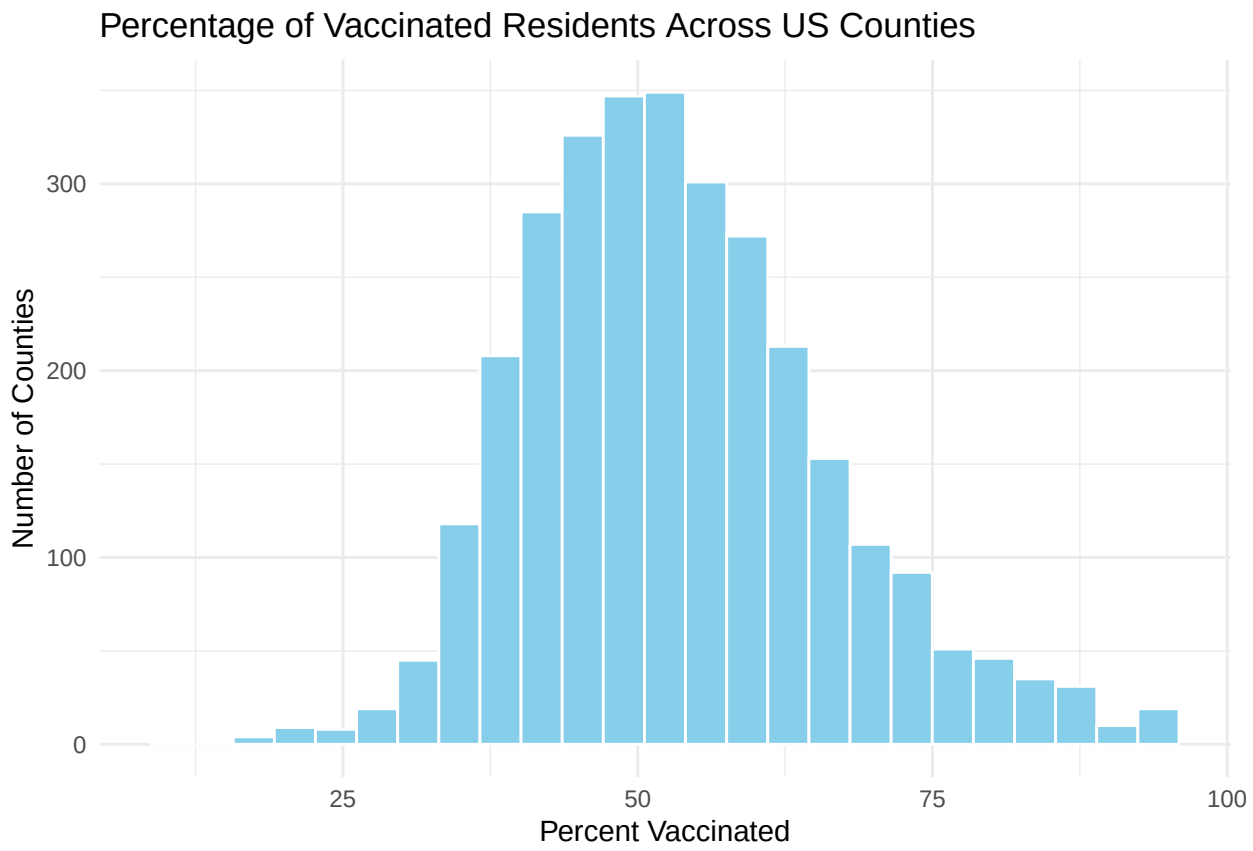
- COVID-19 Deaths: Derived from the New York Times dataset on county-level Covid-19 data for 2021-2022. We can see that the outliers are very far from the mean deaths, which is 84, and the median is only 29. The highest outlier represents Los Angeles County at 7034 deaths, the second highest is almost half of that with 4602 deaths in Maricopa County, representing Phoenix, Arizona.
- Mask Usage: Based on a July 2020 survey (also from the New York Times) that reports the percentage of respondents who always wear a mask. The mean and median for respondents who always wear masks is 50%, while the maximum is 88.9% in Inyo county, California, and the minimum five counties come from Montana. This could infer that population density impacts mask usage.
- Vaccination Rates: Sourced from the CDC vaccination data file ("cdc vax mar1.csv"). This file provides county-level vaccination data. The boxpot showing the vaccination rates by social vulnerability index takes the first histogram (Vaccination rates by county) created to another level. It shows that counties with a SVI of A and B have noticeably higher percentage of fully vaccinated residents than counties with SVI of C and D. The table created provides a list of the counties with highest fully vaccinated rates, with over 10 counties having a 95% completion rate, and Slope County in North Dakota being the lowest with just 11.3% fully vaccinated.



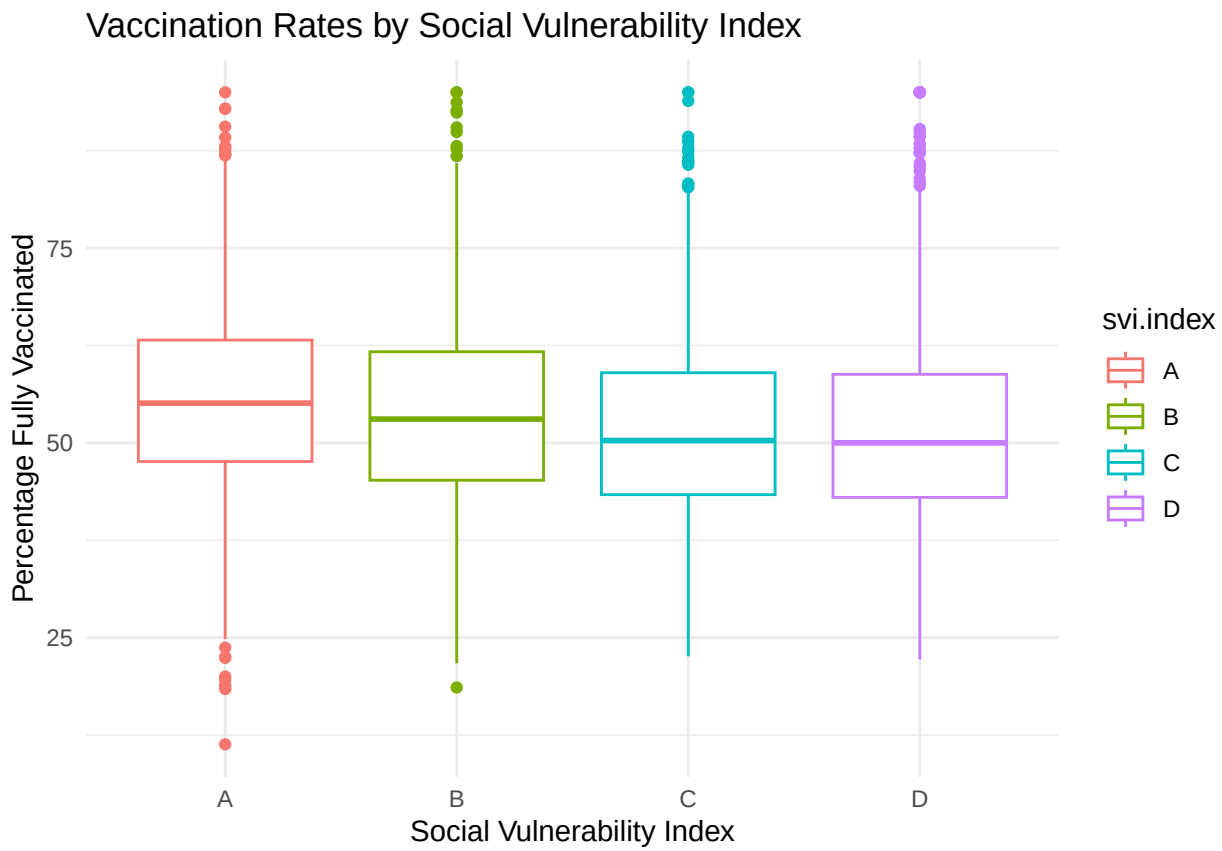

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   12.00   29.00   84.14   70.00 7034.00
```




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




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




```
## # A tibble: 15 x 3
##    vax.complete state        county
##           <dbl> <chr>        <chr>
##  1         95   Arizona      Apache
##  2         95   Arizona      Santa Cruz
##  3         95   California   Imperial
##  4         95   Colorado     San Juan
##  5         95   Georgia      Chattahoochee
##  6         95   Kansas       Geary
##  7         95   New Mexico   McKinley
##  8         11.3 North Dakota Slope
##  9         95   Texas        Brooks
## 10         95   Texas        Irion
## 11         95   Texas        Maverick
## 12         95   Texas        Presidio
## 13         95   Texas        Starr
## 14         95   Texas        Webb
## 15         95   Wyoming      Teton
```

Analysis: The regression table above indicates that higher vaccination rates are significantly associated with lower COVID-19 deaths, and mask usage also plays a role once controlling for other factors. The tripple asterisks next to the values on the table indicate that P-Values are very low, leading to statistically significant results.

	m1	m2	m3
always.mask	-112.887***		-83.296***
	(10.212)		(10.834)
population	-0.000***	-0.000***	
	(0.000)	(0.000)	
svi.indexB	11.415***	9.968***	9.581***
	(2.884)	(2.878)	(2.859)
svi.indexC	15.142***	14.399***	13.128***
	(3.111)	(3.103)	(3.080)
svi.indexD	19.811***	18.353***	17.491***
	(3.401)	(3.393)	(3.367)
vax.complete		-1.123***	-0.923***
		(0.093)	(0.099)
Num.Obs.	3049	3049	3049

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

```
knitr::opts_chunk$set(
  echo = FALSE,
  eval = TRUE,
  fig.align = 'center',
  message = FALSE,
  warning = FALSE
  )
library(tinytex)
library(tidyverse)
library(knitr)
library(lfe)
library(haven)
library(modelsummary)
library(gt)
library(data.table)
  #Covid-19 Deaths:

  covid <-
    data.table::fread('https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2022.
    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))
  ## Here I have cleaned the data to include only the variables I want to investigate.

  #Mask Usage:

  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,
```

```
    .keep = 'none'
  )
## Here I have cleaned the data to show data for individuals who always wear a mask.

#Vaccination Rates:
vax <-
  read_csv('cdc vax mar1.csv') %>%
  filter(
    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,
    svi.index = SVI_CTGY,
    .keep = 'none'
  )
##Here I have extracted data from CDC reports to get percentages on fully vaccinated residents.

covid <-
  left_join(covid, mask) %>%
  left_join(vax) %>%
  mutate(deaths.scaled = deaths / population * 100000) %>%
  ungroup() # scale by population

covid %>%
  ggplot(aes(x = (1 + deaths))) +
  geom_histogram(color = 'white', fill = 'skyblue', bins = 25) +
  scale_x_log10() +
  ggtitle("Covid-19 deaths across US counties (logged)") +
  labs(x = "1 + Covid-19 Deaths (log scale)", y = "Number of Counties") +
  theme_minimal()
## Here I have created a histogram to better visualize Covid-19 deaths across US counties.

summary(covid$deaths)
## While the mean deaths is 84, there are huge outliers, as the max is 7034, being Los Angeles County

covid %>%
  ggplot(aes(x = always.mask)) +
  geom_histogram(color = 'white', fill = 'skyblue', bins = 25) +
  ggtitle("Percentage of Mask Usage Across US Counties") +
  labs(x = "Always Mask (%)", y = "Number of Counties") +
  theme_minimal()
## Here I have used merged data to create a histogram visualizing the distribution of individuals who

summary(covid$always.mask)

covid %>%
  ggplot(aes(x = vax.complete)) +
  geom_histogram(color = 'white', fill = 'skyblue', bins = 25) +
```

```
    ggtitle("Percentage of Vaccinated Residents Across US Counties") +
    labs(x = "Percent Vaccinated", y = "Number of Counties") +
    theme_minimal()
  ## Here I have created a histogram showing the percentage of vaccinated residents across US counties.

summary(covid$vax.complete)

  covid %>%
    filter(!is.na(svi.index), !is.na(vax.complete)) %>%
    ggplot(aes(x = svi.index, y = vax.complete, color = svi.index)) +
    geom_boxplot() +
    labs(
      title = "Vaccination Rates by Social Vulnerability Index",
      x = "Social Vulnerability Index",
      y = "Percentage Fully Vaccinated"
    ) +
    theme_minimal()
  ## This boxplot shows vaccination rates by social vulnerability index.

  covid %>%
    select(vax.complete, state, county) %>%
    filter(vax.complete %in% c(min(vax.complete, na.rm = T),
                               max(vax.complete, na.rm = T)))
  ## This provides a list of the counties with highest fully vaccinated rates, with over 10 counties ha
 mods <-
    list(
      m1 = felm(deaths.scaled ~ always.mask + population + svi.index | state, data = covid),
      m2 = felm(deaths.scaled ~ vax.complete + population + svi.index | state, data = covid),
      m3 = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, data = covid)
    )
## Here I have created regression estimates for each of the variables individually, and then combined.

 modelsummary(
    mods,
    gof_map = c('nobs'),
    stars = TRUE,
    output = 'gt'
  )

 ## Here I have created a regression table that shows the relationship between deaths and mask usage, d
```