

Covid Data Between State With The Highest Vaccine Rate and State With The Lowest Vaccine Rate

B. Han

2025-02-17

I will be comparing the Covid deaths per thousand vs Cases per thousand between the state with the highest vaccine rate Rhode Island and the state with the lowest vaccine rate Wyoming. The hope is to find a similar correlation in the beginning and some alterations between the two states as the the vaccine is rolled out later on in the data. For most of the clean up I followed the videos in the course, but decided to focus my analysis on Wyoming and Rhode Island

```
knitr::opts_chunk$set(echo = TRUE)
library(stringr)
library(readr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(tidyr)
library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
## 
##     date, intersect, setdiff, union

library(ggplot2)
```

R Markdown

Importing and Transforming

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv", "time_series_covid19_confirmed_global.csv", "time_series_covid19_deaths_US.csv"

file_names <-
c("time_series_covid19_confirmed_US.csv", "time_series_covid19_confirmed_global.csv", "time_series_covid19_deaths_US.csv")

urls <- str_c(url_in, file_names)

global_cases <- read_csv(urls[2])

## Rows: 289 Columns: 1147
## -- Column specification -----
## Delimiter: ","
## chr (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

global_deaths <- read_csv(urls[4])

## Rows: 289 Columns: 1147
## -- Column specification -----
## Delimiter: ","
## chr (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

US_cases <- read_csv(urls[1])

## Rows: 3342 Columns: 1154
## -- Column specification -----
## Delimiter: ","
## chr (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
## dbl (1148): UID, code3, FIPS, Lat, Long_, 1/22/20, 1/23/20, 1/24/20, 1/25/20...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

US_deaths <- read_csv(urls[3])

## Rows: 3342 Columns: 1155
## -- Column specification -----
## Delimiter: ","
## chr (6): iso2, iso3, Admin2, Province_State, Country_Region, Combined_Key
```

```

## dbl (1149): UID, code3, FIPS, Lat, Long_, Population, 1/22/20, 1/23/20, 1/24...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

```

Cleaning and joining cases and deaths

```

global_cases <- global_cases %>%
  pivot_longer(cols =
    -c('Province/State',
      'Country/Region', Lat, Long),
    names_to = "date",
    values_to = "cases") %>%
  select(-c(Lat, Long))

global_deaths <- global_deaths %>%
  pivot_longer(cols = -c('Province/State',
    'Country/Region', Lat, Long),
    names_to = "date",
    values_to = "deaths") %>%
  select(-c(Lat, Long))

global <- global_cases %>%
  full_join(global_deaths) %>%
  rename(Country_Region = 'Country/Region',
    Province_State = 'Province/State') %>%
  mutate(date = mdy(date))

```

```

## Joining with 'by = join_by('Province/State', 'Country/Region', date)'

```

```
global <- global %>% filter(cases > 0)
```

```
global %>% filter(cases > 28000000)
```

```

US_cases <- US_cases %>%
  pivot_longer(cols = -(UID:Combined_Key),
    names_to = "date",
    values_to = "cases") %>%
  select(Admin2:cases) %>%
  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))

```

```

US_deaths <- US_deaths %>%
  pivot_longer(cols = -(UID:Population),
    names_to = "date",
    values_to = "deaths") %>%
  select(Admin2:deaths) %>%

```

```

  mutate(date = mdy(date)) %>%
  select(-c(Lat, Long_))

US <- US_cases %>%
  full_join(US_deaths)

## Joining with `by = join_by(Admin2, Province_State, Country_Region,
## Combined_Key, date)`

global <- global %>%
  unite("Combined_Key",
    c(Province_State, Country_Region),
    sep = ", ",
    na.rm = TRUE,
    remove = FALSE)

uid_lookup_url <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/UIDs/UIDs.csv"

uid <- read_csv(uid_lookup_url) %>%
  select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

## Rows: 4321 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (7): iso2, iso3, FIPS, Admin2, Province_State, Country_Region, Combined_Key
## dbl (5): UID, code3, Lat, Long_, Population
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

global <- global %>%
  left_join(uid, by = c("Province_State", "Country_Region")) %>%
  select(-c(UID, FIPS)) %>%
  select(Province_State, Country_Region, date, cases, deaths, Population, Combined_Key)

```

US data by State

```

US_by_state <- US %>%
  group_by(Province_State, Country_Region, date) %>%
  summarize(cases = sum(cases), deaths = sum(deaths),
            Population = sum(Population)) %>%
  mutate(deaths_per_mill = deaths * 1000000 / Population) %>%
  select(Province_State, Country_Region, date, cases,
         deaths, deaths_per_mill, Population) %>%
  ungroup()

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

```

```

US_totals <- US_by_state %>%
  group_by(Country_Region, date) %>%
  summarize(cases = sum(cases), deaths = sum(deaths),
            Population = sum(Population)) %>%
  mutate(deaths_per_mill = deaths * 1000000 / Population) %>%
  select(Country_Region, date, cases,
         deaths, deaths_per_mill, Population) %>%
  ungroup()

```

```

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

```

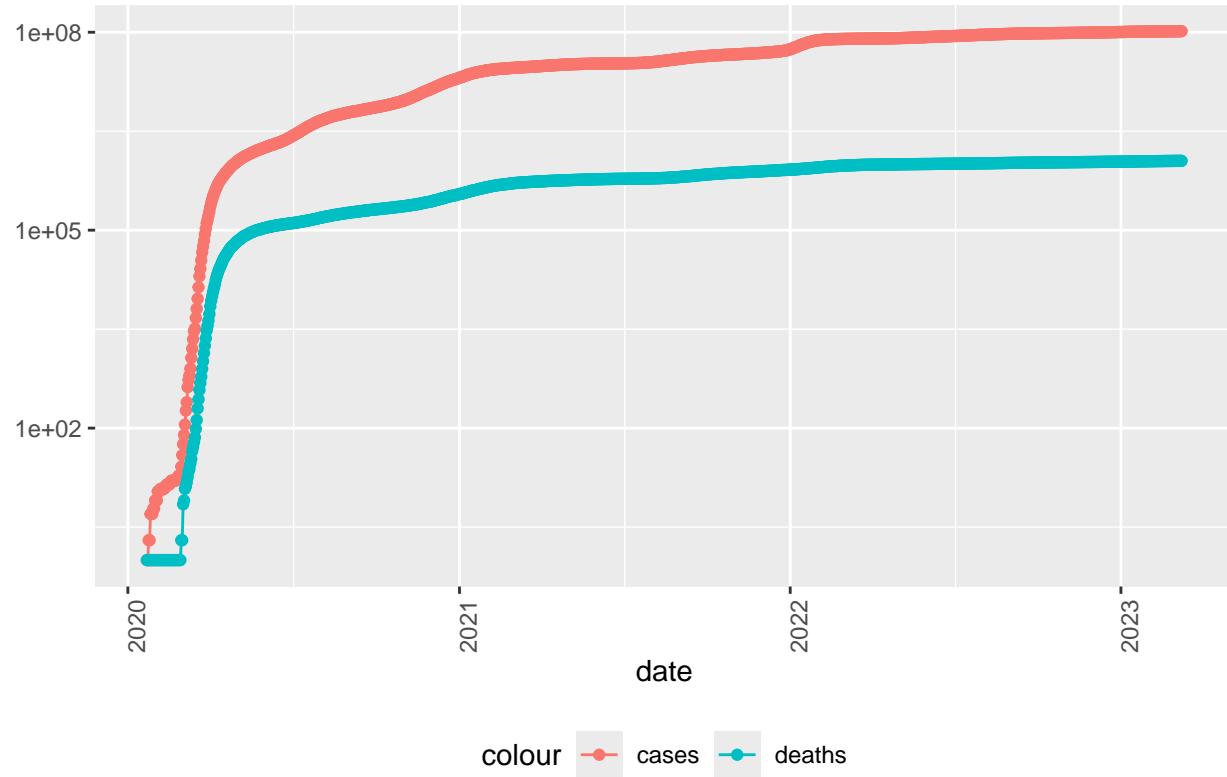
US Totals Plot

```

US_totals %>%
  filter(cases > 0) %>%
  ggplot(aes(x = date, y = cases)) +
  geom_line(aes(color = "cases")) +
  geom_point(aes(color = "cases")) +
  geom_line(aes(y = deaths, color = "deaths")) +
  geom_point(aes(y = deaths, color = "deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = "COVID19 in US", y = NULL)

```

COVID19 in US

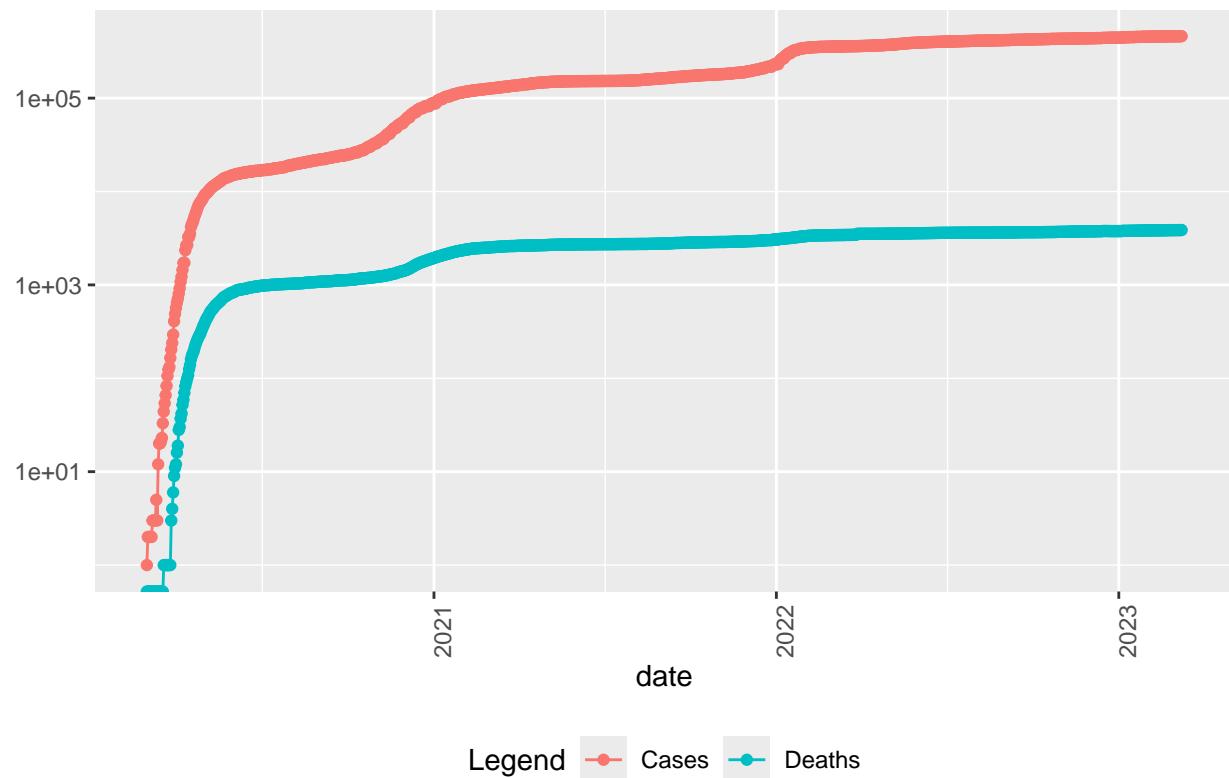


Rhode Island and Wyoming State Level Analysis

```
state <- "Rhode Island"
US_by_state %>%
  filter(Province_State == state) %>%
  filter(cases > 0) %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = cases, color = "Cases")) +
  geom_point(aes(y = cases, color = "Cases")) +
  geom_line(aes(y = deaths, color = "Deaths")) +
  geom_point(aes(y = deaths, color = "Deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = str_c("COVID-19 in ", state), y = NULL, color = "Legend")
```

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
## log-10 transformation introduced infinite values.
```

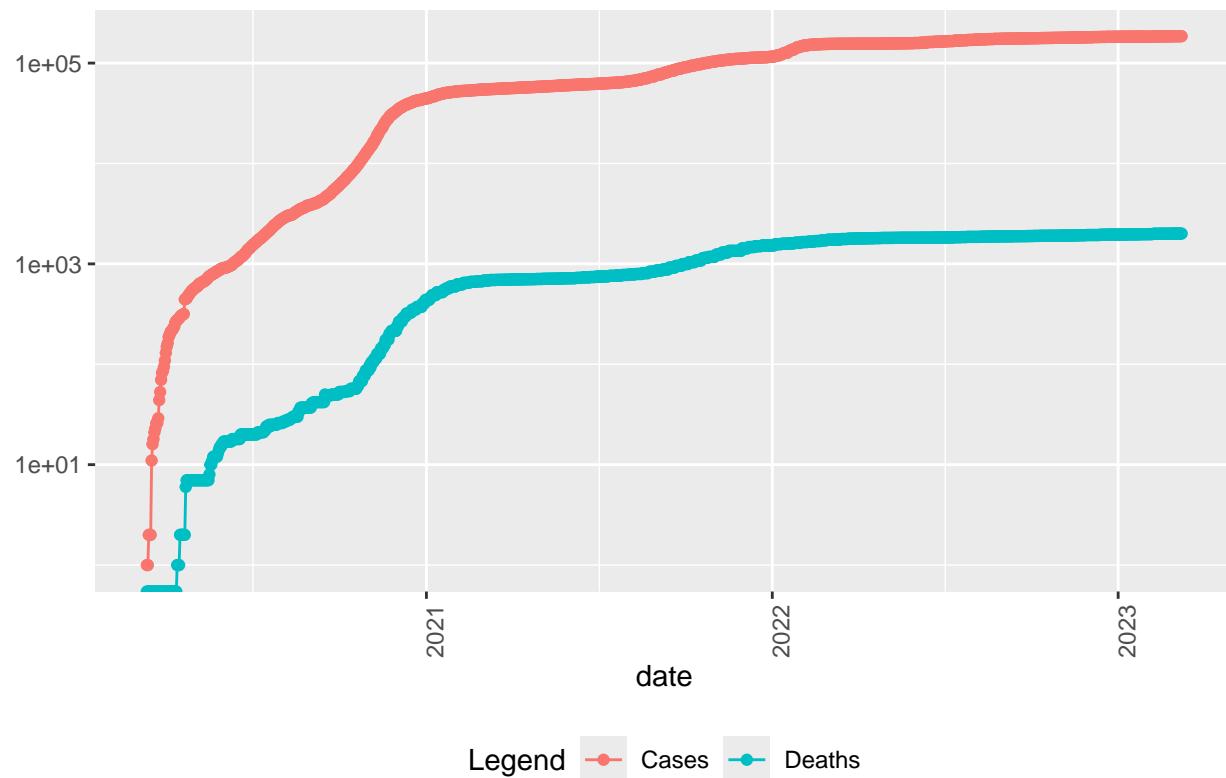
COVID-19 in Rhode Island



```
state <- "Wyoming"
US_by_state %>%
  filter(Province_State == state) %>%
  filter(cases > 0) %>%
  ggplot(aes(x = date)) +
  geom_line(aes(y = cases, color = "Cases")) +
  geom_point(aes(y = cases, color = "Cases")) +
  geom_line(aes(y = deaths, color = "Deaths")) +
  geom_point(aes(y = deaths, color = "Deaths")) +
  scale_y_log10() +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 90)) +
  labs(title = str_c("COVID-19 in ", state), y = NULL, color = "Legend")

## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
## log-10 transformation introduced infinite values.
```

COVID-19 in Wyoming



```
max(US_totals$date)
```

```
## [1] "2023-03-09"
```

Chose the highest vaccine rate state and lowest rate state

```
highest_lowest_vaccine_states <- c("Wyoming", "Rhode Island")

States_selected <- US_by_state %>%
  filter(Province_State %in% highest_lowest_vaccine_states)

States_selected <- States_selected %>%
  mutate(cases_per_thou = 1000 * cases / Population,
        deaths_per_thou = 1000 * deaths / Population)
```

Rhode Island Linear Model

```
RI_Model <- lm(deaths_per_thou ~ cases_per_thou, data = filter(States_selected, Province_State == "Rhode Island"))
summary(RI_Model)
```

```
##
```

```

## Call:
## lm(formula = deaths_per_thou ~ cases_per_thou, data = filter(States_selected,
##   Province_State == "Rhode Island"))
##
## Residuals:
##       Min     1Q Median     3Q    Max
## -1.00020 -0.25792 -0.07057  0.54704  0.60606
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.000e+00 2.183e-02 45.80 <2e-16 ***
## cases_per_thou 6.768e-03 8.736e-05 77.47 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4566 on 1141 degrees of freedom
## Multiple R-squared:  0.8403, Adjusted R-squared:  0.8401
## F-statistic:  6002 on 1 and 1141 DF,  p-value: < 2.2e-16

```

Wyoming Linear Model

```

WY_Model <- lm(deaths_per_thou ~ cases_per_thou, data = filter(States_selected, Province_State == "Wyoming"))
summary(WY_Model)

```

```

##
## Call:
## lm(formula = deaths_per_thou ~ cases_per_thou, data = filter(States_selected,
##   Province_State == "Wyoming"))
##
## Residuals:
##       Min     1Q Median     3Q    Max
## -0.28928 -0.09340 -0.02561  0.07511  0.45225
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.578e-02 5.977e-03 4.312 1.75e-05 ***
## cases_per_thou 1.103e-02 3.085e-05 357.643 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1234 on 1141 degrees of freedom
## Multiple R-squared:  0.9912, Adjusted R-squared:  0.9912
## F-statistic: 1.279e+05 on 1 and 1141 DF,  p-value: < 2.2e-16

```

Combine the Rhode Island and Wyoming plots with the LM model

```

States_selected %>%
  ggplot(aes(x = cases_per_thou, y= deaths_per_thou, color = Province_State)) +
  geom_point() +
  geom_smooth(method = "lm", aes(color = Province_State)) +

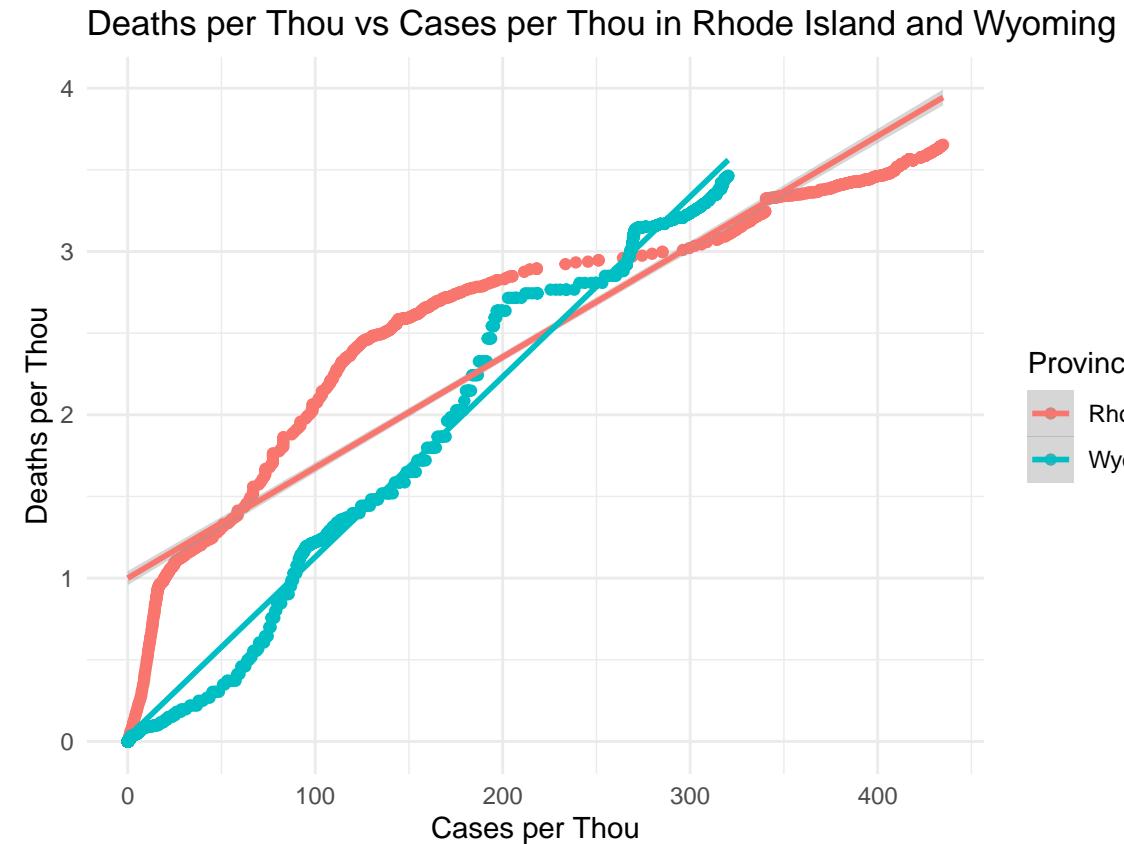
```

```

  labs(title = "Deaths per Thou vs Cases per Thou in Rhode Island and Wyoming",
       x = "Cases per Thou",
       y = "Deaths per Thou") + theme_minimal()

```

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



We can see that although Rhode Island starts with a very sharp slope, the slope tends to taper off, while in Wyoming the slope of is steep and remains steep to the point where the deaths per thousand crosses Rhode Island's at some point. We see two very long tapering off sections of deaths for Rhode Island, which probably correlates to the vaccine roll out, and while we see two smaller tapering sections in Wyoming the deaths per thousand continued to rise sharply. My conclusion would be that the Covid vaccine did provide a significant decrease in mortality rate in patients and did in fact help prevent deaths.

Some bias that have risen during this analysis is that Rhode island is about 1 million in population size and Wyoming is about half of that in population and this could skew results as there is 2 times the amount of people in Rhode Island. Another bias is that Wyoming is a very big state than Rhode Island, making proximity to other residents much bigger than Rhode Island. Also, Rhode Island's proximity to New York which was one of the origin sites for Covid is much closer than to Wyoming, making it likely that Covid had infected Rhode Island much earlier than Wyoming.