

---

---

# Analyzing COVID-19 Data by States in the US using SAS & R

— By Robby Born —

---

---

# Background Info



- In the US, response was tepid and varied with each state implementing different requirements as each wave hit
- Covid-19 data was compiled by the New York Times covid data tracker and the state population count was obtained from the Department of Agriculture using 2020 Census data

# Covid-19 Response by State

- During the first wave of the pandemic, all but 7 states implemented stay-at-home orders
- Not all states had a mask mandate, states such Texas and Florida, which only had recommendations, as opposed to a mandate in California and New York
- By the end of May 2020, almost all states had some type of reopening
- While all states saw cases rise after May 2020, states that reopened saw a bigger influx of cases and deaths during that time period

# Cleaning the Data



State_Name	Deaths	Cases	State	Population	Case Proportion	Deaths Per Million
Washington	12627	1469439	53	7705281	0.190705	1638.746205
Illinois	37777	3090590	17	12812508	0.241217	2948.44694
California	89484	9134794	6	39538223	0.231037	2263.227662
Arizona	29681	2014020	4	7151502	0.281622	4150.316954
Massachusetts	23751	1714167	25	7029917	0.243839	3378.560515
Wisconsin	14349	1591806	55	5893718	0.270085	2434.626156
Texas	87801	6707045	48	29145505	0.230123	3012.505702
Nebraska	4175	478183	31	1961504	0.243784	2128.468767
Utah	4730	928847	49	3271616	0.283911	1445.768697
Oregon	7315	707512	41	4237256	0.166974	1726.353093
Florida	73583	5867926	12	21538187	0.272443	3416.397118
New York	67413	5027587	36	20201249	0.248875	3337.070891
Rhode Island	3527	364184	44	1097379	0.331867	3214.021774
Georgia	35943	2434702	13	10711908	0.227289	3355.424636
New Hampshire	2459	304365	33	1377529	0.220950	1785.080387
North Carolina	23268	2635871	37	10439388	0.252493	2228.86629
New Jersey	33308	2216804	34	9288994	0.238648	3585.748898
Colorado	12206	1364255	8	5773714	0.236287	2114.063842
Maryland	14377	1017235	24	6177224	0.164675	2327.420861
Nevada	10164	714652	32	3104614	0.230190	3273.836941
Tennessee	25757	1974299	47	6910840	0.285681	3727.043312
Hawaii	1379	235797	15	1455271	0.162030	947.5898303

```
library(tidyverse) library(dplyr)
covid_states <- read_excel("~/covid_states.xlsx")
Covid_Real <- read_excel("~/Covid_Real.xlsx")
covid = covid_states
covid_st = covid %>%
  group_by(fips) %>%
  filter(state != "District of Columbia") %>%
  filter(state != "Guam") %>%
  filter(state != "Virgin Islands") %>%
  filter(state != "Puerto Rico") %>%
  filter(state != "Northern Mariana Islands") %>%
  filter(state != "American Samoa")
Covid_set = Covid_st %>%
  rename(State_Name = state)

covid_R = select(Covid_Real, -8, -9, -10, -11)
covid_R = covid_R %>%
  arrange(State_Name)

Covid_P = covid_R %>%
  left_join(Covid_set, by = "State_Name")

Covid_P <- Covid_P[, c(1,9,8,4,5,6,7,2,3)]
Covid_P = Covid_P %>%
  select(-Deaths, -Cases)

Covid_P = Covid_P %>%
  rename(Total_Deaths = Deaths) %>%
  rename(Total_Cases = Cases)

Covid_P = Covid_P %>%
  rename(fips = state) %>%
  rename(state = State_Name)
Covid_P = Covid_P %>%
  rename(Case_Proportion = "Case Proportion") %>%
  rename(Deaths_Per_Million = "Deaths Per Million")

Covid_P[2, 5] = 733391
Covid_P[2, 6] = 0.33689805 #Alaska value was incorrect, so
Covid_P[2, 7] = 1625.3267 #I did this to input the correct value
```

# Codes

## SAS:

```

proc import out = covid_p
  datafile = "C:\Users\robby\Documents\Covid_Real.xlsx"
  dbms = xlsx replace;
run;

proc sql;
  select *
  from covid_p;
quit;

title "Covid-19 Cases US";
proc gmap data=covid_p map= maps.us;
  id state;
  choro cases;
run;
quit;

title "Covid-19 Deaths US";
proc gmap data=covid_p map= maps.us;
  id state;
  choro deaths;
run;
quit;

title "Covid-19 Case Rate by State";
proc gmap data=covid_p map= maps.us;
  id state;
  choro case_proportion;
run;
quit;

title "Covid-19 Deaths per Million";
proc gmap data=covid_p map= maps.us;
  id state;
  choro Deaths_Per_Million;
run;
quit;

```

## R:

```

covid_Cases = plot_usmap(data = Covid_P, values = "Total_Cases", color =
"blue") +
scale_fill_continuous(low = "white", high = "blue", name = "Total Cases",
label = scales::comma) +
labs(title = "Covid-19 Cases by State", subtitle = "Choropleth Map") +
theme(legend.position = "right")
covid_Cases

```

```

covid_Deaths = plot_usmap(data = Covid_P, values = "Total_Deaths", color =
"blue") +
scale_fill_continuous(low = "white", high = "blue", name = "Total Deaths",
label = scales::comma) +
labs(title = "Covid-19 Deaths by State", subtitle = "Choropleth Map") +
theme(legend.position = "right")
covid_Deaths

```

```

covid_Case_Rate = plot_usmap(data = Covid_P, values = "Case_Proportion", color =
"blue") +
scale_fill_continuous(low = "white", high = "blue", name = "Cases",
label = scales::comma) +
labs(title = "Covid-19 Case Rate by State", subtitle = "Choropleth Map") +
theme(legend.position = "right")
covid_Case_Rate

```

```

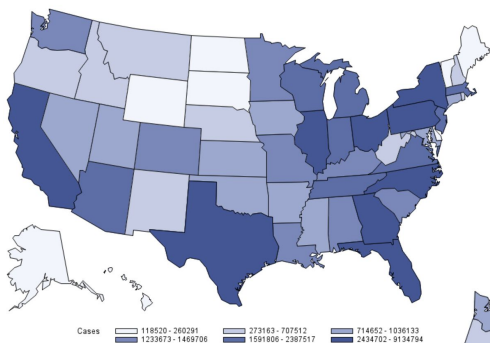
covid_Death_Rate = plot_usmap(data = Covid_P, values = "Deaths_Per_Million",
color = "blue") +
scale_fill_continuous(low = "white", high = "blue", name = "Deaths",
label = scales::comma) +
labs(title = "Covid-19 Deaths per Million", subtitle = "Choropleth Map") +
theme(legend.position = "right")
covid_Death_Rate

```

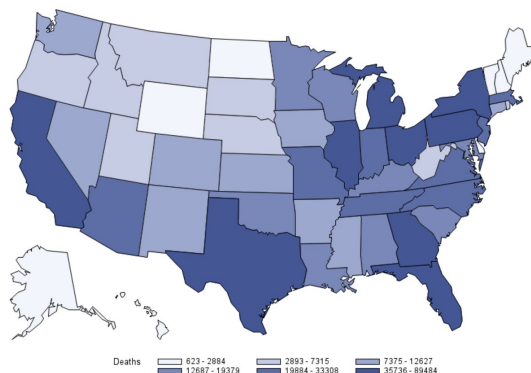
# Data Visualization and Analysis

## Total Case and Death rate by State - SAS

Covid-19 Cases US

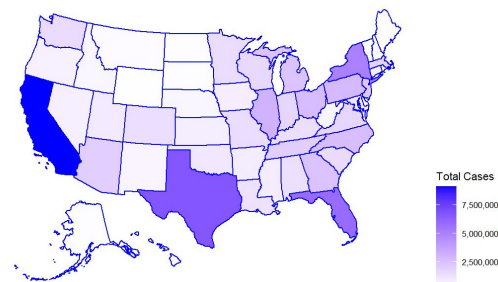


Covid-19 Deaths US

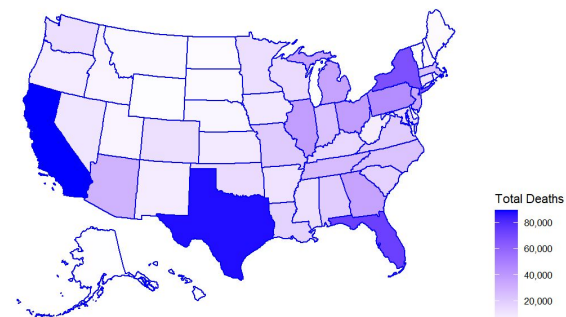


## Total Case and Death rate by State - R

Covid-19 Cases by State  
Choropleth Map



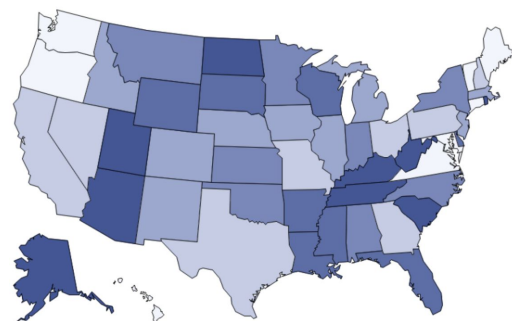
Covid-19 Deaths by State  
Choropleth Map



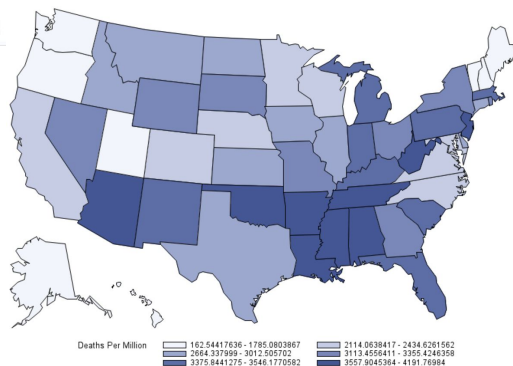
# Data Visualization and Analysis

## Case and Death by State per Capita - SAS

Covid-19 Case Rate by State

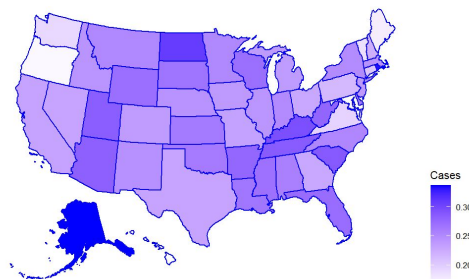


Covid-19 Deaths per Million

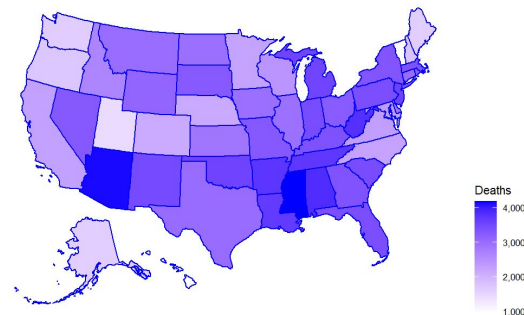


## Case and Death by State per Capita - R

Covid-19 Case Rate by State  
Choropleth Map



Covid-19 Deaths per Million  
Choropleth Map



# What Do We See?

- When we compare case and death rates, states such as Arizona, Mississippi, Alabama, and West Virginia had high case and death rates
- States such as Utah, Rhode Island, and Alaska had very high case rates, but relatively low death rates
- The most prominent correlation with high death rates are race/ethnicity, comorbidities, poverty level, and health disparity and access to medical care
- States and regions, such as Arizona and states in the South, with an older, rural population, more people of color, higher poverty rates saw the highest death rates, and low access to medical care
- Diligent masking, greater access to healthcare, more education, and reduced poverty help create a more equitable solution to reduce the mortality rate



# Thank you!

I would like to thank

- My professor, Dr Olga
- All of my STAT495 class
- The New York Times for providing open access to their Covid-19 state and local data sets
- And the US Department of Agriculture and US Census Bureau for providing the 2020 census population numbers to the public