# Stats 140SL 'Red versus Blue states'

Natasha Vuong, Kirsten Landsiedel, Yonathan Khalil, Ashlyn Jew, Emily Hou, and Hana Lim

10/10/2018

## Contents



# 1 Statement of the problem

## 1.1 In context

With the onset of the COVID-19 pandemic, the United States' response became political and polarizing. Republican President Donald Trump initially downplayed the severity of the pandemic and delivered messaging that directly conflicted with that of public health experts. Statements were issued by several Republican leaders, and the public's response to COVID safety, as well as official government policies regarding the matter, seemed to vary along party lines. For instance, in July 2020, Georgia governor Brian Kemp, a Republican, banned Georgia cities from requiring people to wear masks in public spaces. A month earlier, California governor Gavin Newsom, a Democrat, issued an executive order requiring all Californians to wear masks while in public. Seeing that the vast majority of students in Stats 140SL are seniors, and that this country will still be very much affected by COVID-19 for the foreseeable future, this issue will be an important one to consider when relocating for jobs or graduate school post-graduation. For these reasons, we want to investigate whether "red" states have conferred vulnerability to death due to COVID-19 when compared to "blue" states (as characterized in the wake of the 2020 election).

# 2 Formal Hypothesis

We define "red" states as those that voted Republican in the 2020 election and "blue" states as those that voted Democrat. Our null hypothesis is that there is no statistically significant difference between the number of deaths in “red” versus “blue” states. Our alternate hypothesis is that there is a difference between the number of deaths in “red” versus “blue” states.

# 3 Analysis

H0: there is no statistically significant difference between the number of deaths in “red” versus “blue” states
Ha: there is a difference between the number of deaths in “red” versus “blue” states.(Death rate)

## 3.1 death rate

```
## # A tibble: 2 x 2
##   redstate  mean
## * <lgl>    <dbl>
## 1 FALSE     47.4
## 2 TRUE      31.9
```

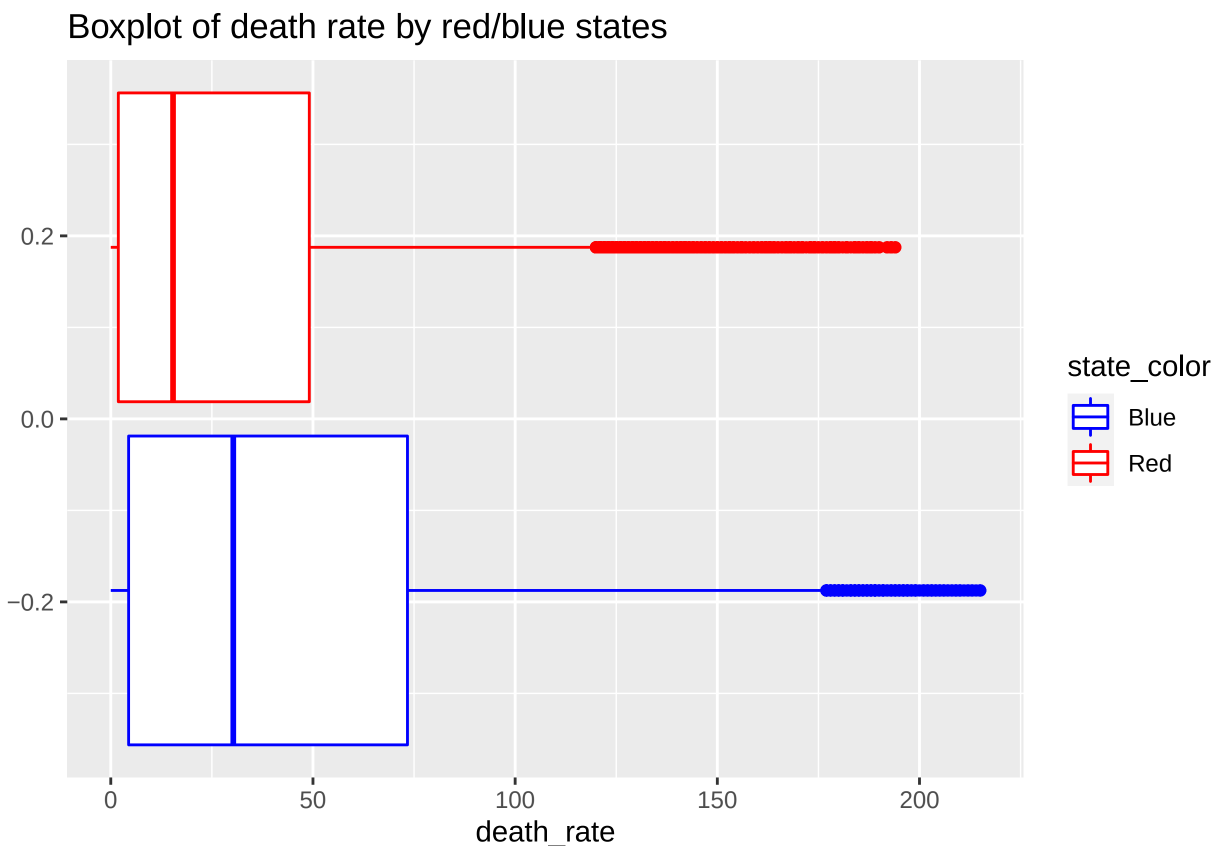
```
##
##  Two Sample t-test
##
## data:  x and y
## t = -23.366, df = 18919, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -16.73273 -14.14268
## sample estimates:
## mean of x mean of y
##  31.93278  47.37048
```

Since p-value is less than the significance value, reject the null and therefore supports the alternative hypothesis.

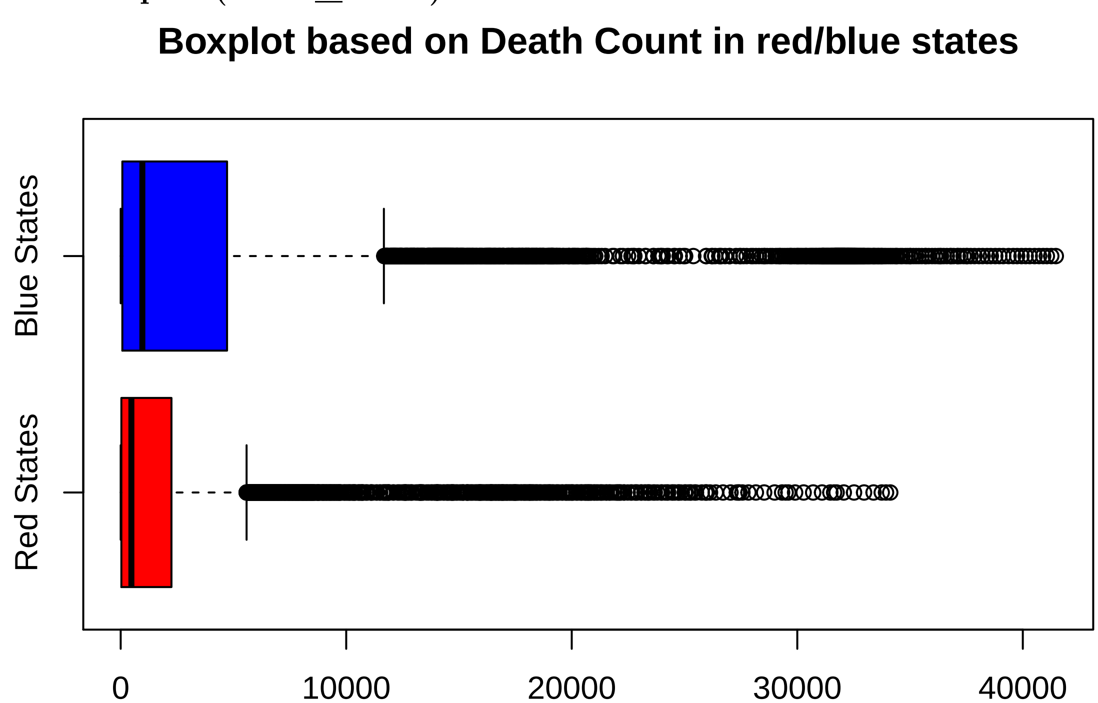
```
##
##  Two Sample t-test
##
## data:  x_2 and y_2
## t = -24.533, df = 18919, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2089.815 -1780.590
## sample estimates:
## mean of x mean of y
##  1923.032  3858.235
```

T-test conducted based on death count also shows that it supports the alternative hypothesis.

### 3.2 Boxplot (death_rate)



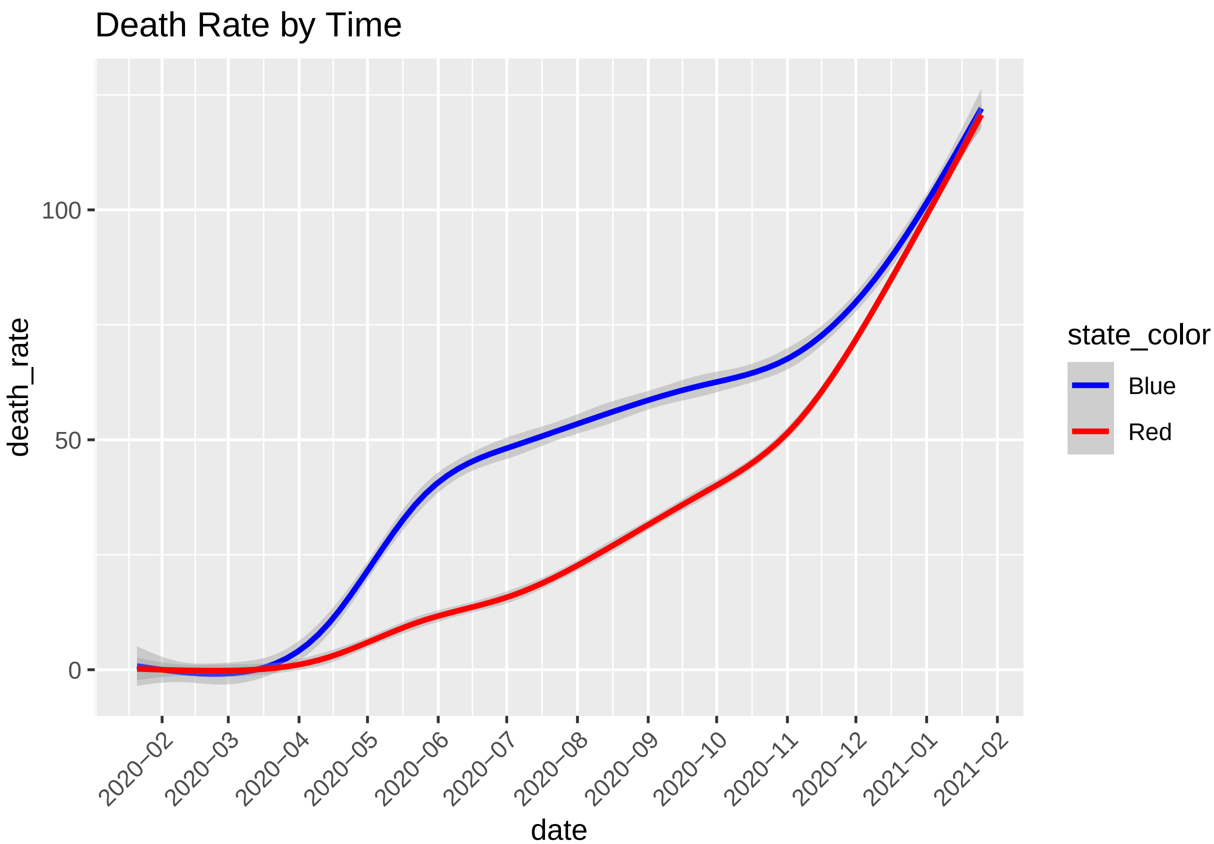

### 3.3 Boxplot (death_count)




Boxplot: The means of both the death rate and death count from the "blue" states are higher than the means of the “red” states. The “blue” states also have wider range than the “red” states.

### 3.4 death rate by time

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




Time based plot: Death rate of the combined blue states and red states over time.

## 4 Additional Analysis

In the hypothesis tests we ran, we rejected the null hypothesis that there is no statistically significant difference between the number of deaths in “red” versus “blue” states. We wanted to explore possible relationships between policy milestones and state partisanship to gain a broader understanding of COVID-19 related deaths. Per the CDC, “The virus that causes COVID-19 most commonly spreads between people who are in close contact with one another (within about 6 feet, or 2 arm lengths)” (https://www.cdc.gov/coronavirus/2019-ncov/faq.html#Spread). We chose to look at policy milestones because they are likely to affect the chances that people are in close contact with each other.

Separating states by red and blue, we looked at the average dates that the states closed schools, closed non-essential businesses, implemented stay-at-home orders, and ended stay-at-home orders. Some states didn’t have dates for closing non-essential businesses, implementing stay-at-home orders, reopening businesses statewide, or ending stay-at-home orders because there wasn’t a state-wide mandate. For those states, we coded them as today’s date.

```
## # A tibble: 2 x 5
##   partisan meanSchoolsClosed meanNonessential~ meanStayAtHomeS~ meanStatewideSt~
## * <chr>    <date>            <date>            <date>           <date>
## 1 blue     2020-03-17        2020-05-22        2020-03-26       2020-06-13
## 2 red      2020-03-17        2020-08-08        2020-07-05       2020-08-01
```

```
## Time difference of 78.49231 days

## Time difference of 100.6538 days

## Time differences in days
## [1] 78.92308 27.72000
```

On average, blue and red states have the same date for schools closing. Red states have an average date for non-essential businesses closing that is 78 days later than blue states. Red states have an average date for implementing stay-at-home orders that is 100 days later than blue states. Blue states have an average mandated stay-at-home period of 79 days, while for red states, it is 28 days.

# 5 Conclusion

Based on the T-test, we can observe that p-value is less than the significance value (alpha=0.05). We therefore reject the null hypothesis and support that there is a difference between the number(rate) of deaths in "red" versus "blue" states.

# 6 Limitations

Our definition of "red" and "blue" states oversimplifies the political demographics of each state. States are not homogeneously "red" or "blue". For example, in the "blue" state of California, the Bay Area and Los Angeles comprises predominantly democratic voters, while Shasta County comprises mostly republican voters. Many states are similar in that large cities are mostly "blue" while rural areas tend to lean "red". Thus, it may be an oversimplification in our analysis to define states by their electoral college votes in the 2020 election. Furthermore, defining states as "blue" does not indicate that those states have more aggressive coronavirus policies compared to "red" states. For instance, Arizona, a "blue" state, does not have a statewide mask mandate, but Texas, a "red" state does have a statewide mask mandate. Also, local citywide policies usually have a more direct impact on coronavirus cases than state policies. And we have many instances where city policies differ from their state policies. For example, Alaska does not have a state mask mandate, but Anchorage, Alaska's largest city, does have a city mask mandate. These are some of the cofounding factors that we have not taken into consideration in our analysis. Another consideration is that COVID cases and deaths may be recorded differently in different cities, counties, and states. We are not certain whether the difference in data collection would affect our results. Overall the binary variable is useful in a model that predicts death rates from Covid-19, however the interpretability is minimal as further analysis is necessary to dissect this simple predictor.

# 7 Division of work

Division of work
Statement of problem: Kirsten, Emily
Formal statistical hypothesis: Kirsten, Emilymm Aim to have done by Saturday
Analysis: Yonatan, Natasha, Hana
Conclusion: Yonatan, Natasha, Hana
Aim to have done by Tuesday
Limitations of the analysis: Ashlyn
Aim to have done by Wednesday
Latex/Rmd: Kirsten