

# **COVID-19 Cases - Percent Changes**

**Do 'Social-Distancing' guidelines provide a change in the percent increase of the number of COVID-19 cases that a state reports?**

# Important Notes and References

- IMPORTANT: I am not an epidemiologist. This study is being done for educational purposes only and should not be used in any scientific or policy-making way.
- The number of COVID-19 cases for each day/state/county was taken from New York Times' GitHub (<https://github.com/nytimes/covid-19-data>).
- Data for COVID-19 for each day/state was also obtained from <https://covidtracking.com/api/>
- The date at which each state implemented stay-at-home orders was scraped from: <https://www.kff.org/coronavirus-policy-watch/stay-at-home-orders-to-fight-covid19/>

# Hypothesis Testing and Statistical Power

See **COVID-19-USA-Percent-Change.ipynb**

- Null Hypothesis: The social distancing guidelines do not provide a change in the percent increase of COVID-19 cases that a state reports.
- Rejection Threshold: 0.05
- With this null hypothesis, the conditional probability of finding a result equally or more extreme than observed was calculated (the p-value).

# Explore and Clean The Data

- For each day, the percent increase in the number of cases from the day before was calculated. Then, that percent increase from before social distancing and after social distancing was used to calculate the p-value.
  - $\text{Percent Increase} = (\text{That Day's Cases} - \text{Previous Day's Cases}) / (\text{Previous Day's Cases}) * 100$
- A sample of the data that was used is shown below.

[ 34 ] :

	2020-04-02	2020-04-03	2020-04-04	2020-04-05	2020-04-06	2020-04-07	2020-04-08	2020-04-09
state								
Alabama	14.828210	20.866142	6.384365	12.737293	8.962520	9.521436	13.746017	13.565426
Alaska	2.097902	6.849315	8.333333	9.467456	2.162162	11.640212	6.161137	4.017857
American Samoa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Arizona	13.234253	10.562500	14.132278	12.382368	8.638167	4.462475	5.864078	10.711665
Arkansas	9.455128	8.052709	0.677507	14.804845	8.675264	7.551241	8.024072	6.406685

# Statistical Power

## Alternative Hypothesis

- An alternative hypothesis was developed that social social distancing guidelines provide a 30% decrease in percent daily increases.
- Statistical Power: The probability that the test correctly reject the null hypothesis.

$$\text{Mean Percent Daily Increase of COVID-19 Cases (before distancing)} \sim N \left( \text{Mean}(\text{before}), \frac{\sigma}{\sqrt{\text{days}}} \right)$$

$$\text{Mean Percent Daily Increase of COVID-19 Cases (after distancing)} \sim N \left( \text{Mean}(\text{after}), \frac{\sigma}{\sqrt{\text{days}}} \right)$$

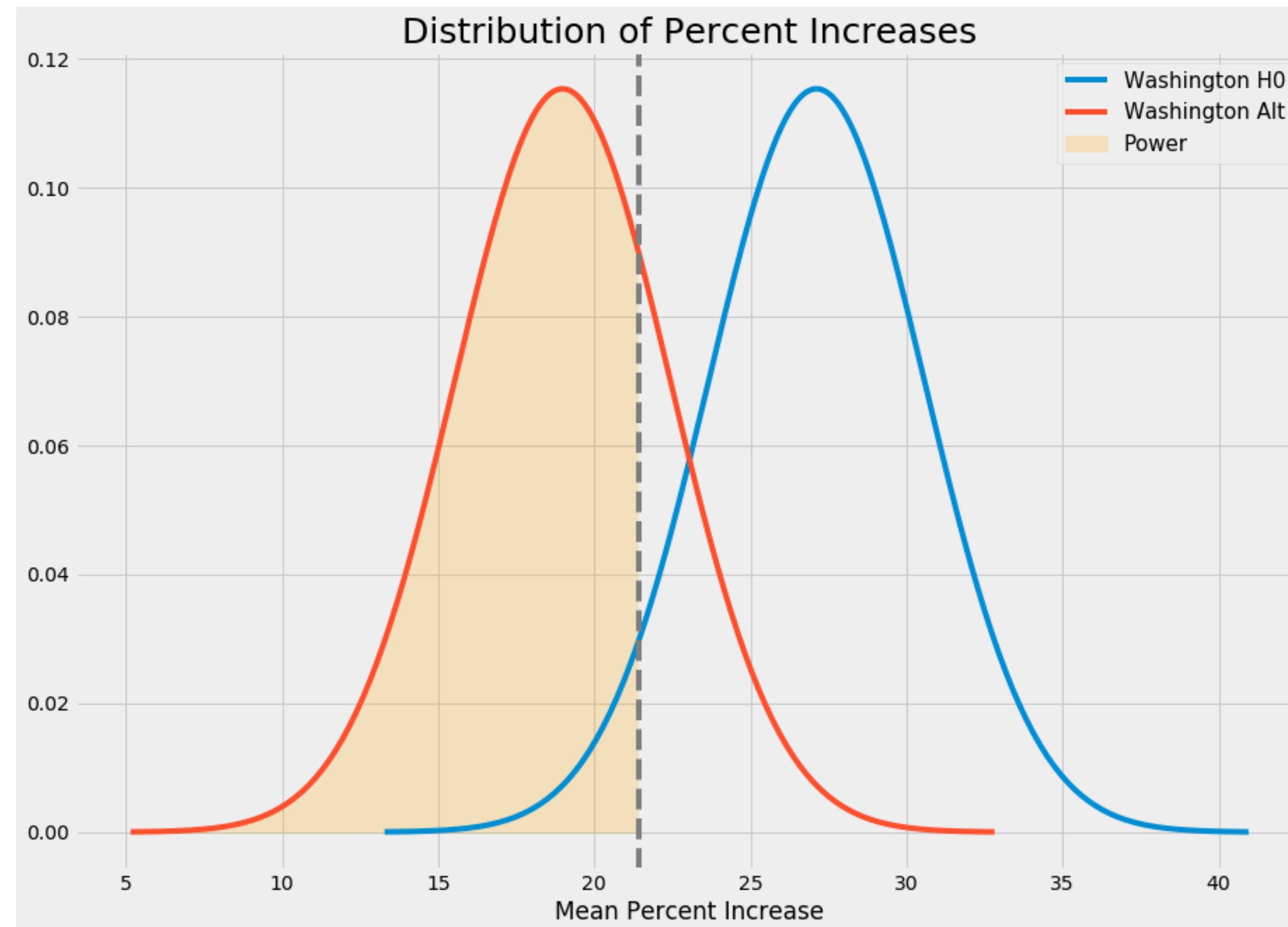
$$\text{Difference in Mean Percent Daily Increase of COVID-19 Cases} | H_0 \sim N \left( 0, \frac{\sigma}{\sqrt{\text{days}}} \right)$$

$$\text{Mean Percent Daily Increase of COVID-19 Cases} | H_A \sim N \left( \text{Mean} * .7, \frac{\sigma}{\sqrt{\text{days}}} \right)$$

# Statistical Power

## Look Only At Washington State

- Using below plot, the power was calculated to be 0.76. That means, there is a 76% chance we will detect a difference of 30%, when such a difference actually exists.



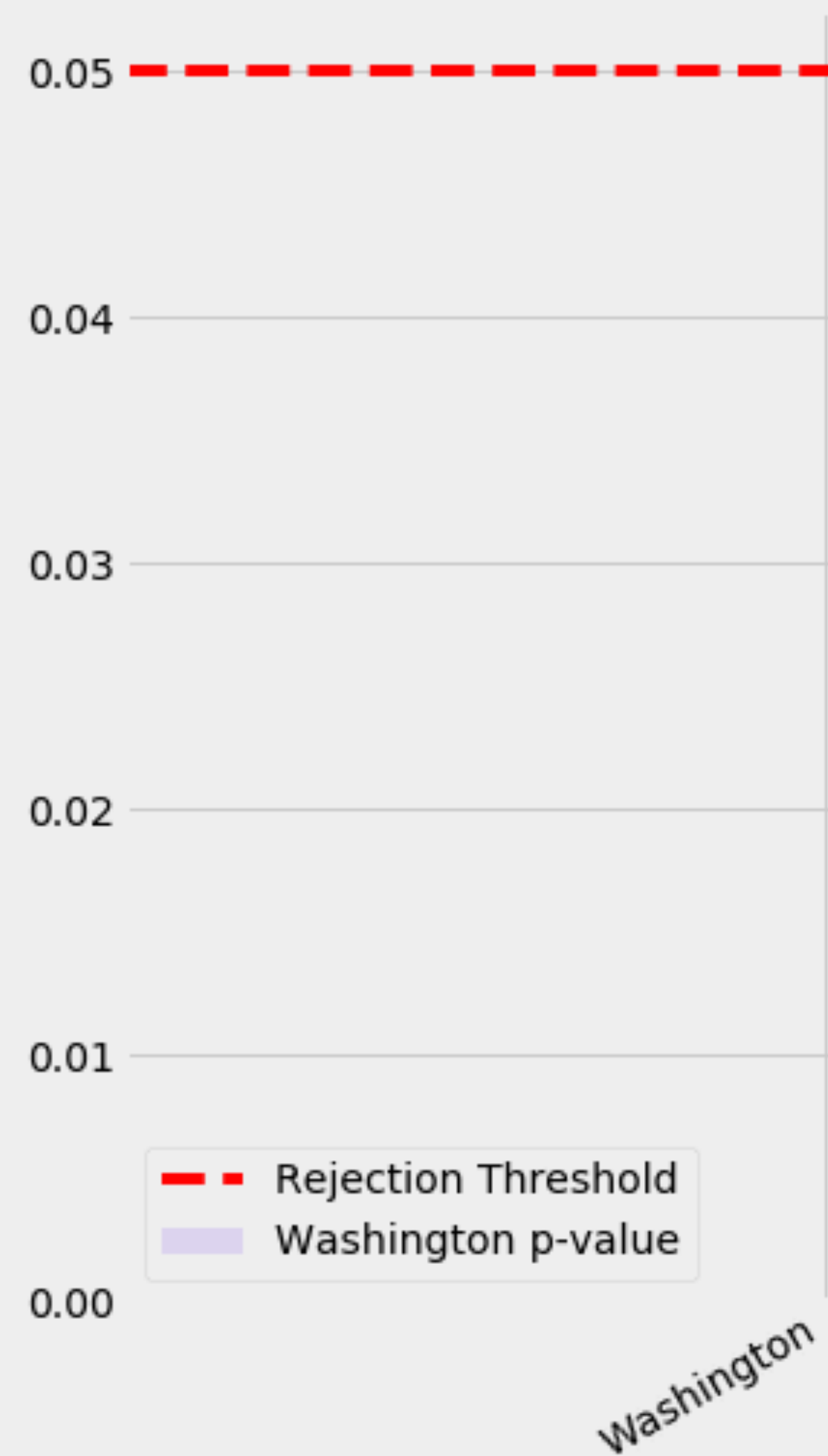
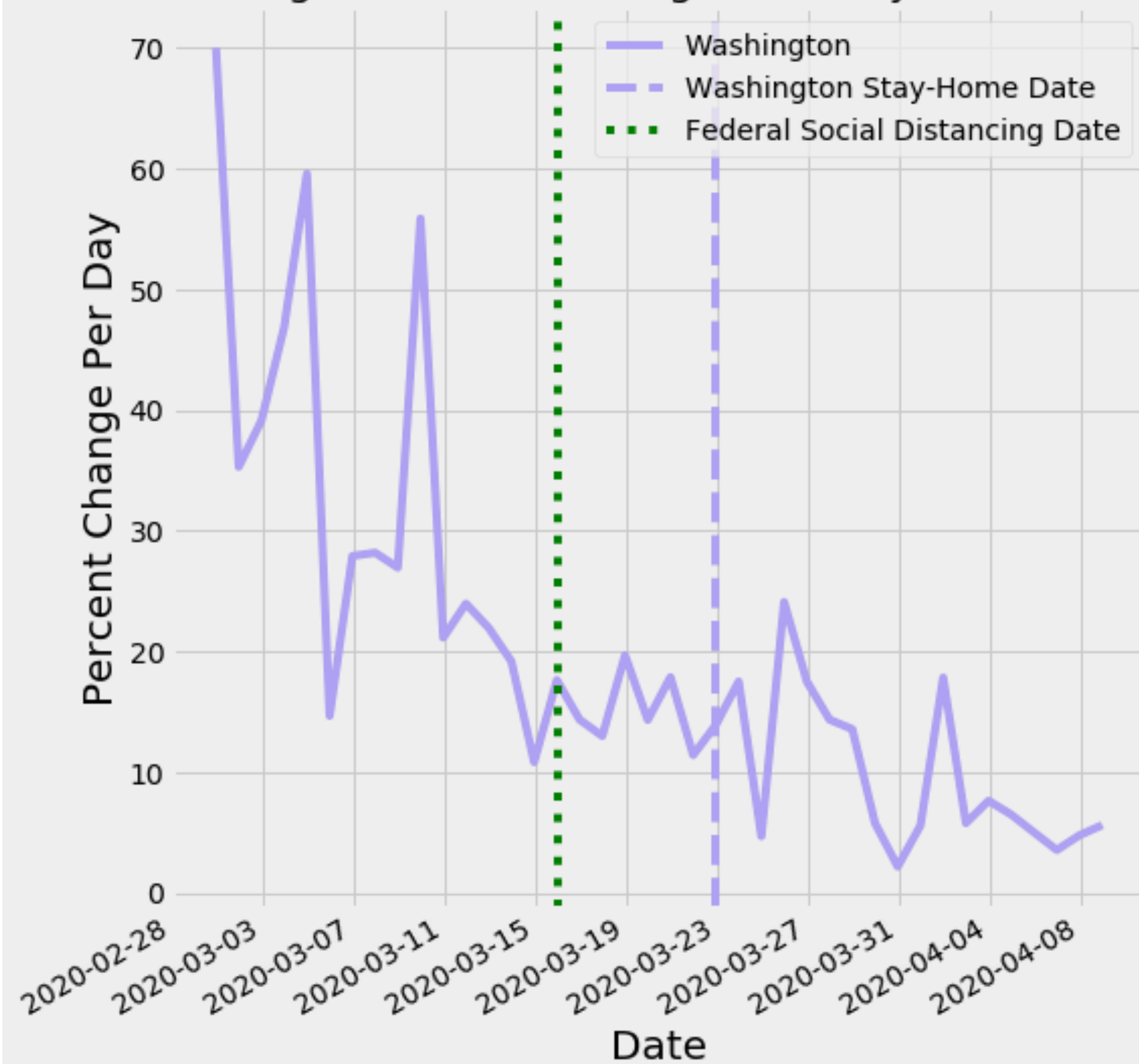
# Hypothesis Testing

**See All-States-COVID-19.ipynb**

- Data from before the Federal Social Distancing date was obtained and compared to data after the Federal Social Distancing. From that data, the p-value was calculated using the Student's t-test.
- A graph of Washington State's percent change over time and a bar chart of the p-value is shown below. The p-value is very low, which means that the null hypothesis can be rejected. This signifies that there MAY be an effect of social distancing on percent increases in reported COVID-19 cases.
- On the next slide, California is shown on a similar plot. The p-value is higher than the rejection threshold, meaning that the null hypothesis cannot be rejected. Additional exploration of the data can be done to look into why.

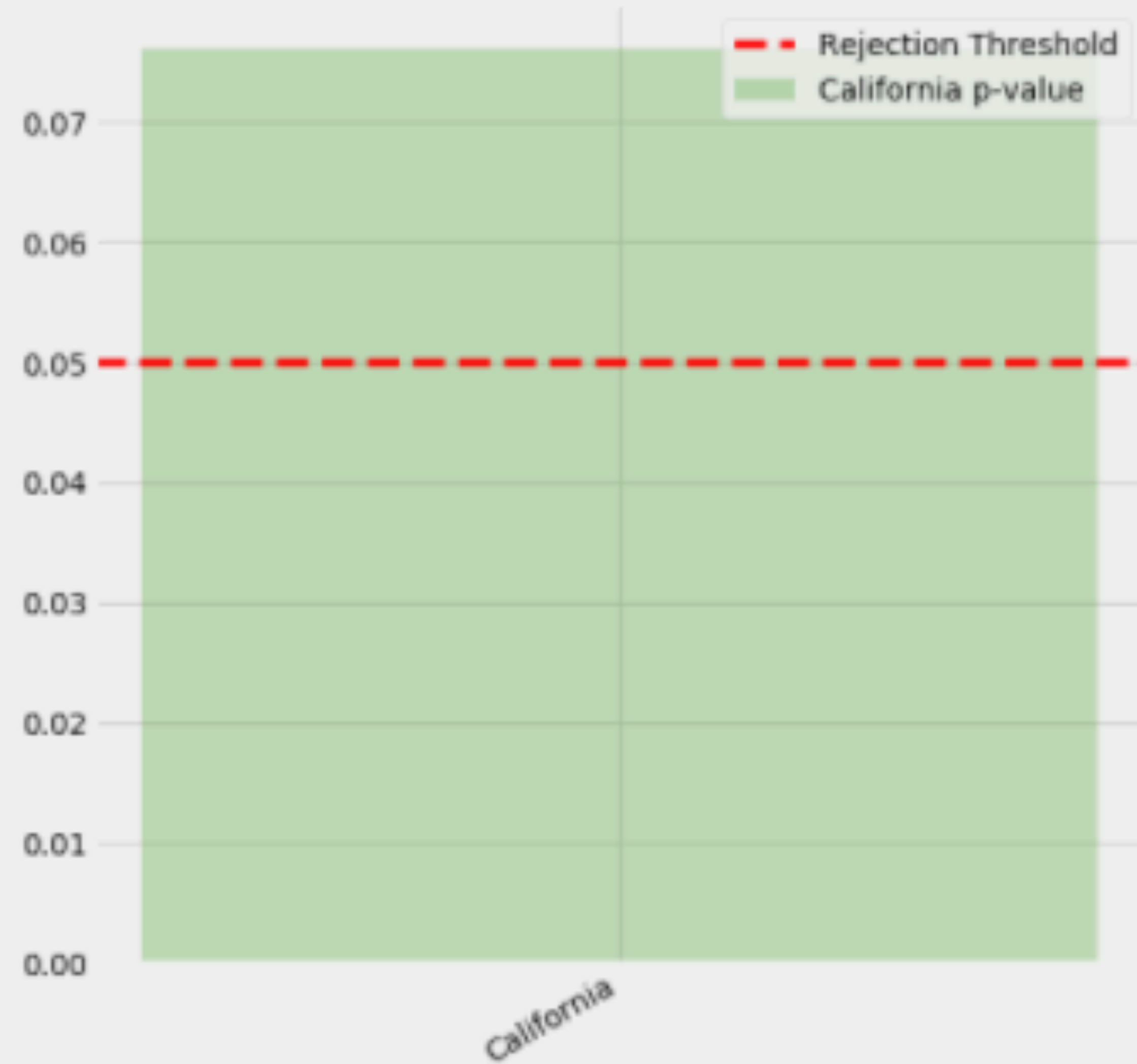
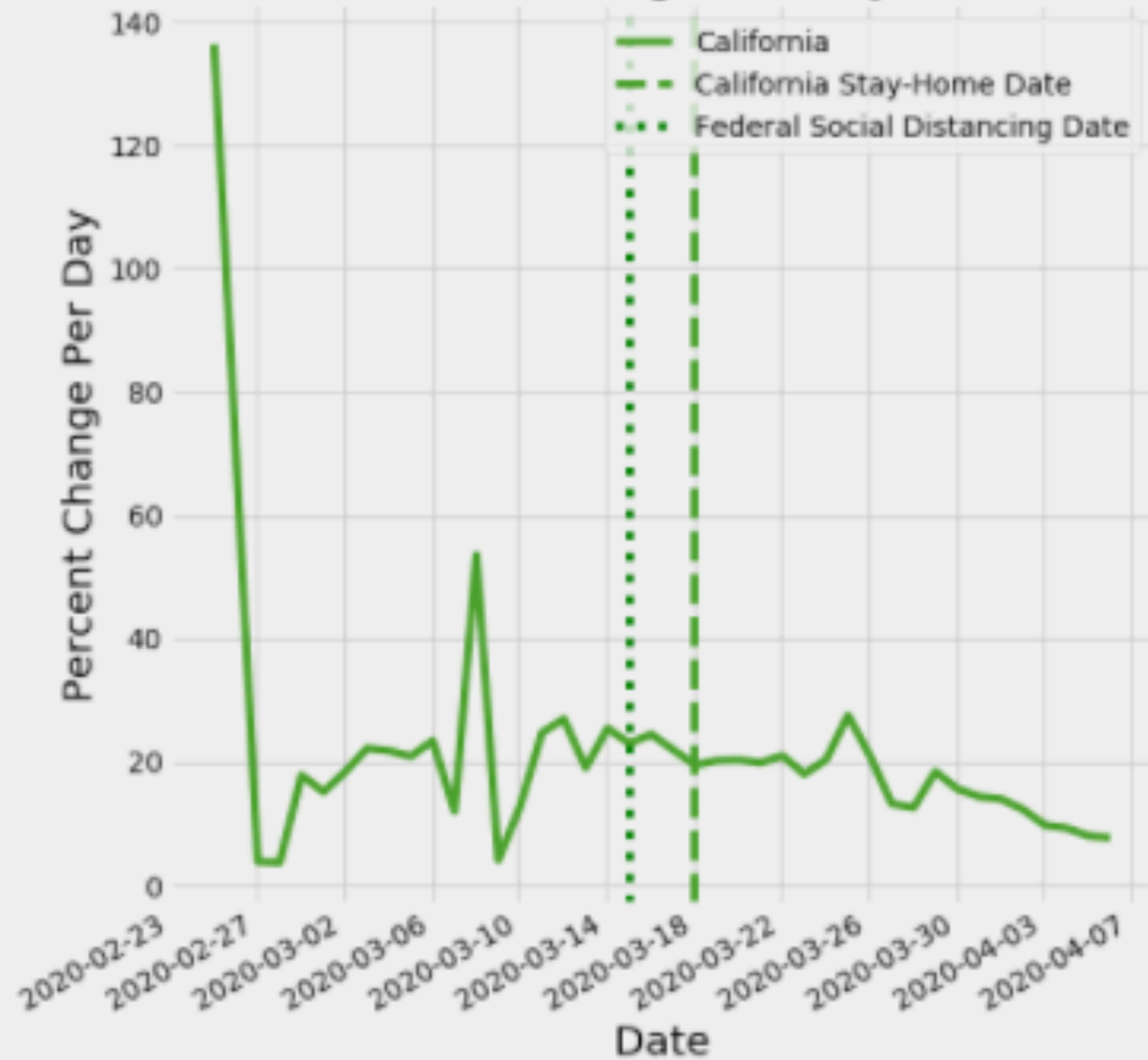


Washington Percent Change and Stay-Home Dates





California Percent Change and Stay-Home Dates



# **Additional Exploratory Data Analysis**

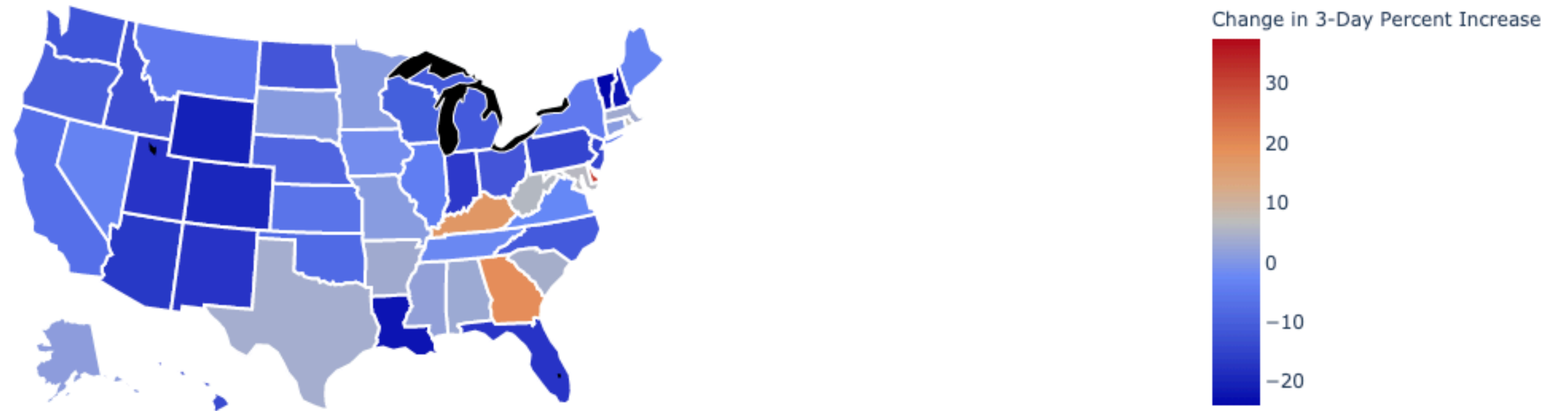
**Choropleth Maps Using Plotly**

# State-Level Choropleth Map

See `choropleth_states.ipynb`

- To further answer the question: "Do social distancing guidelines provide a change in the percent increase of the number of COVID-19 cases that a state reports", an investigation of the 3-day percent changes was performed. To do this, the 3-day percent change was calculated. This is:
  - 3-Day Percent Change =  $(\text{Today's Number of Cases} - \text{Cases From 3 Days Ago}) / (\text{Cases from 3 Days Ago}) * 100$
- The previous 3-day percent was also calculated (i.e., number of cases from 4 days and 7 days ago). The previous 3-day percent change was subtracted from the current 3-day percent change.
  - If the value is positive, that means the state is trending in the wrong direction (i.e., can argue that social distancing is not working). If negative, trending in correct direction (i.e., can argue social distancing is working). See below for the choropleth map. For interactive, see '`choropleth_states.ipynb`'.

## Change in 3-Day Percent Increase



As shown above, most states are in the 'negative', which means that there has been a drop in the 3-day percent change. This can be attributed to social distancing.

# County-Level Choropleth Map

See `choropleth_county.ipynb`

- An investigation of the 3-day percent changes was also performed at the county level.
- Many counties data was unavailable or no cases reported.



Again, as shown above, most counties are in the ‘negative’, which means that there has been a drop in the 3-day percent change. This can be attributed to social distancing.



# Lessons Learned

- I found this dataset to be quite manageable in terms of size, which allowed me to easily clean the data.
- Stick to the question that I am trying to answer, rather than getting side-tracked. The GitHub Project tab was a great way to stay focused.
- Plotly is great for visualizations and I would like to use this more.



# Questions?