

United States COVID-19 Cases by Red & Blue States

By: Laura Faulds, Amjad Abdullah, Rafia Tazeen

Datasets Used and Their Limitations

— — —

- Used 3 different datasets:
 - “US Covid-19 Cases & Deaths by State Over Time” from data.gov
 - “State Codes” from Kaggle
 - “State Population” from Kaggle
- Cleaned, formatted & merged them to make two base dataframes
- Had some limitations in the dataset:
 - Dropped rows with missing values
 - Dropped columns that were irrelevant
 - Did not have data for all 50 states
 - Had to eliminate data for territories
 - Only have data for 2021

Null and Alternative Hypotheses

Null Hypothesis: The type of state government (red/Republican or blue/Democrat) has no impact on the number of COVID-19 cases.

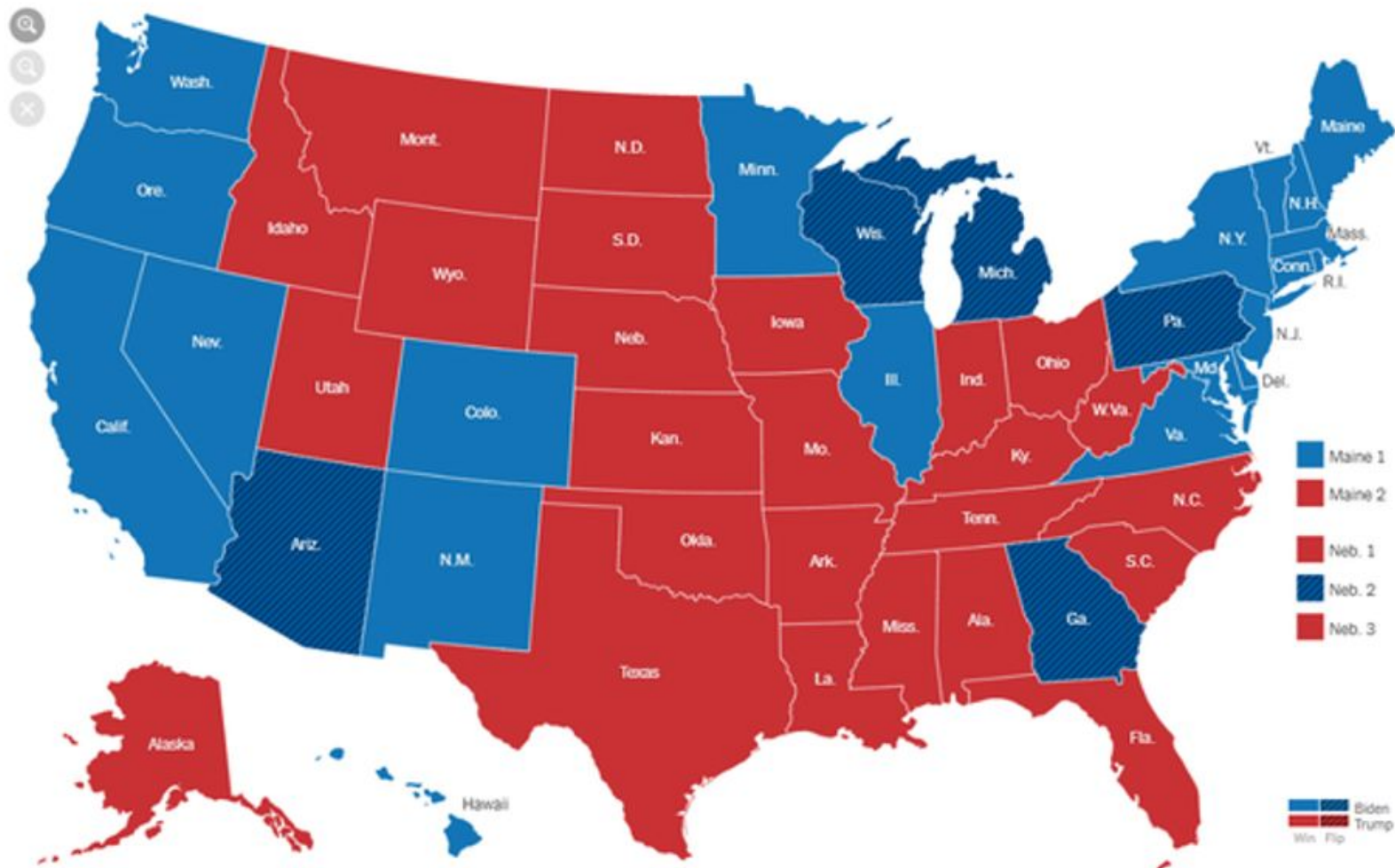
Alternative Hypothesis: The type of state government (red/Republican or blue/Democrat) has an impact on the number of COVID-19 cases.

Analyzing the Data

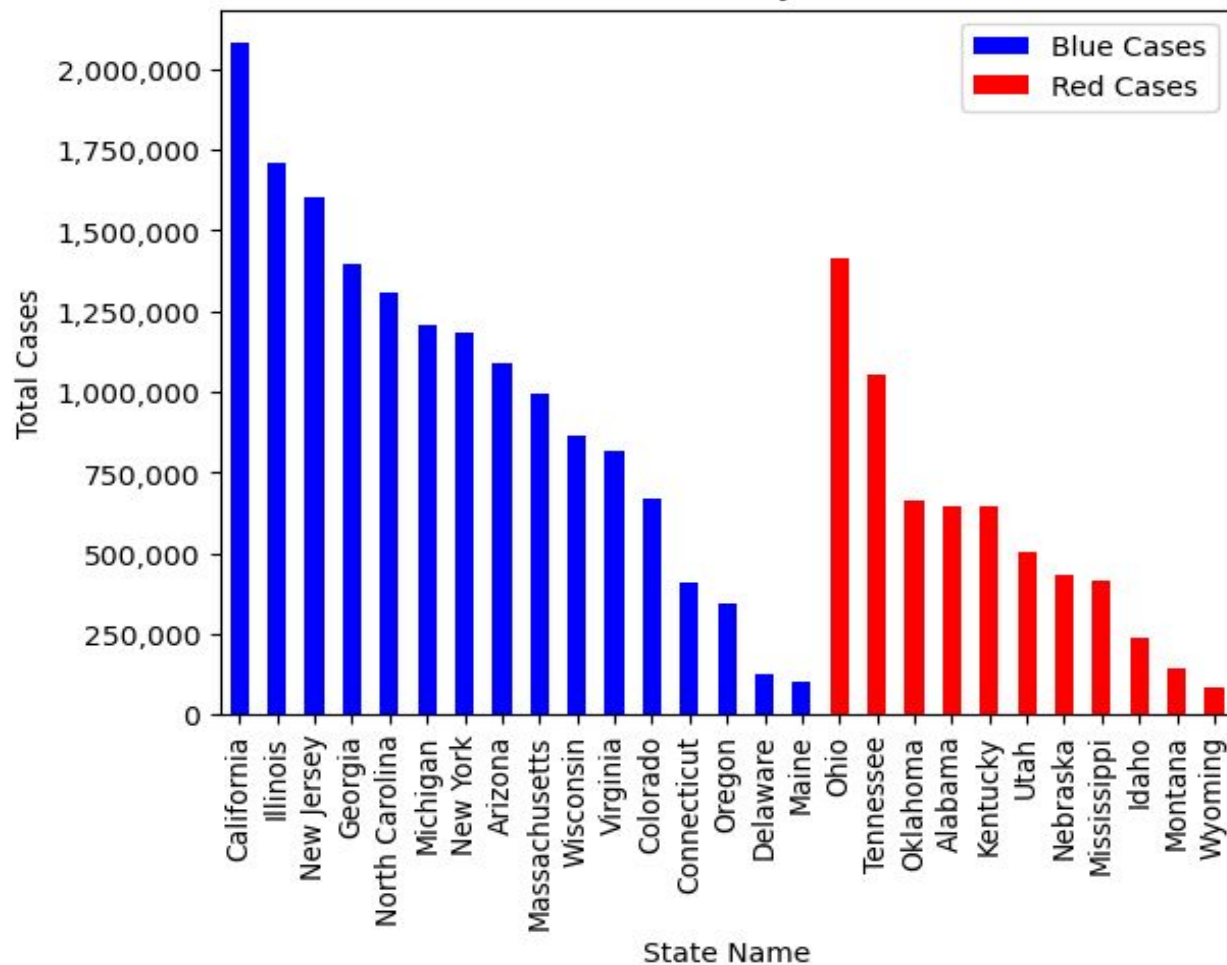
— — —

- Incorporating the data for the governing party for each state (Democrat or Republican)
- Determining state total averages
- Charting to see the distribution of cases by state and their respective governing party
- After analyzing the data according to Total Cases, we decided to re-examine according to Case Prevalence

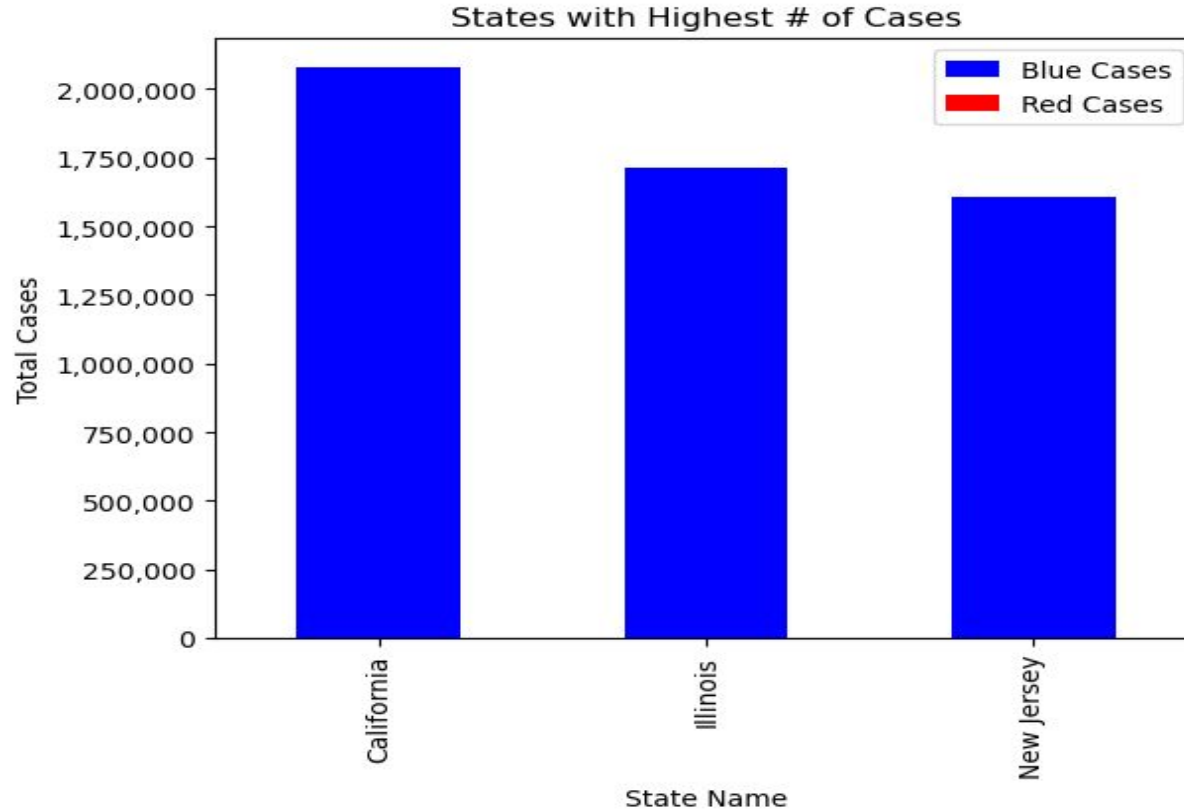
State Name	State Color	Blue Cases	Red Cases	State Population	Case/Population Percentage
Wisconsin	Blue	522523	0	5880101	8.886293
Oregon	Blue	113909	0	4256301	2.676244
Illinois	Blue	963389	0	12686469	7.593831
Colorado	Blue	350126	0	5811297	6.024920
Michigan	Blue	591739	0	10037504	5.895280
...
New York	Blue	1562082	0	19857492	7.866462
Massachusetts	Blue	1140614	0	6989690	16.318521
New Jersey	Blue	1564253	0	9267961	16.878071
Colorado	Blue	957696	0	5811297	16.479901
Illinois	Blue	2181009	0	12686469	17.191616



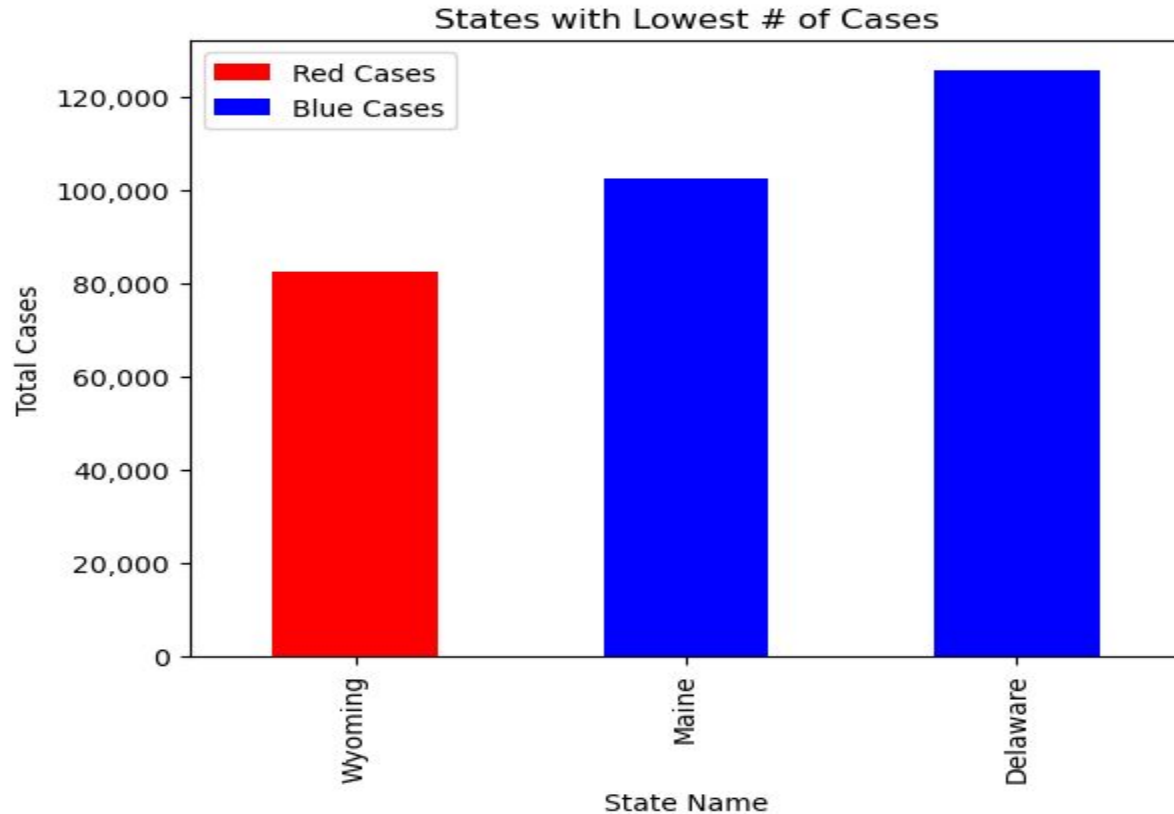
Total Cases by State



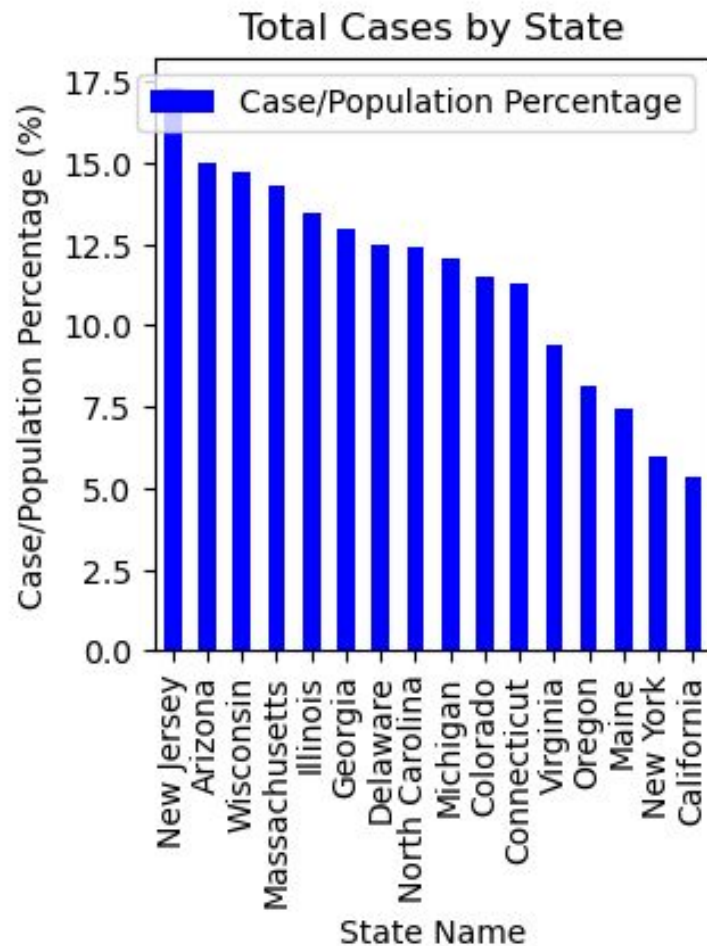
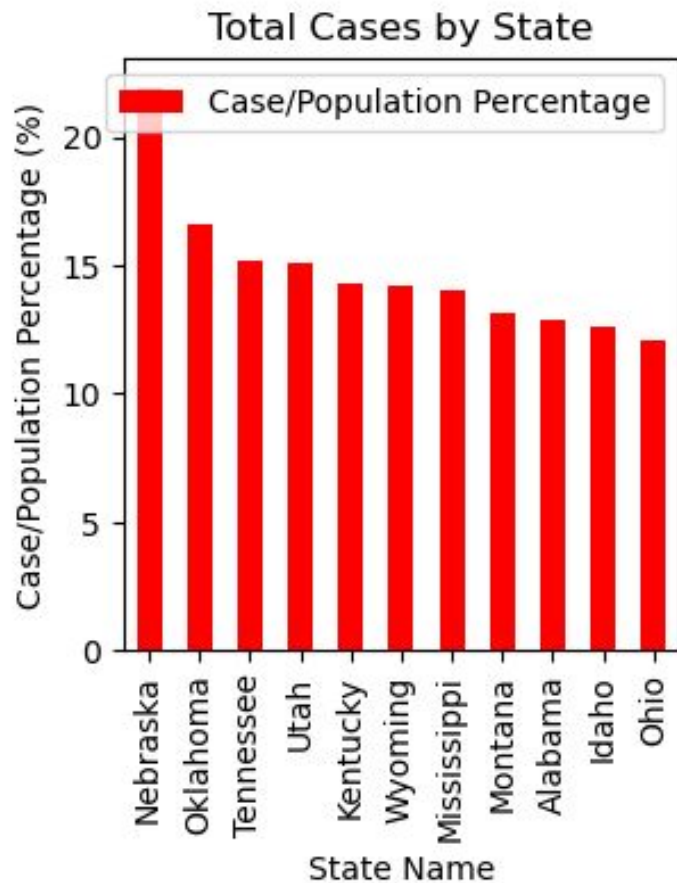
The three highest number of cases by state

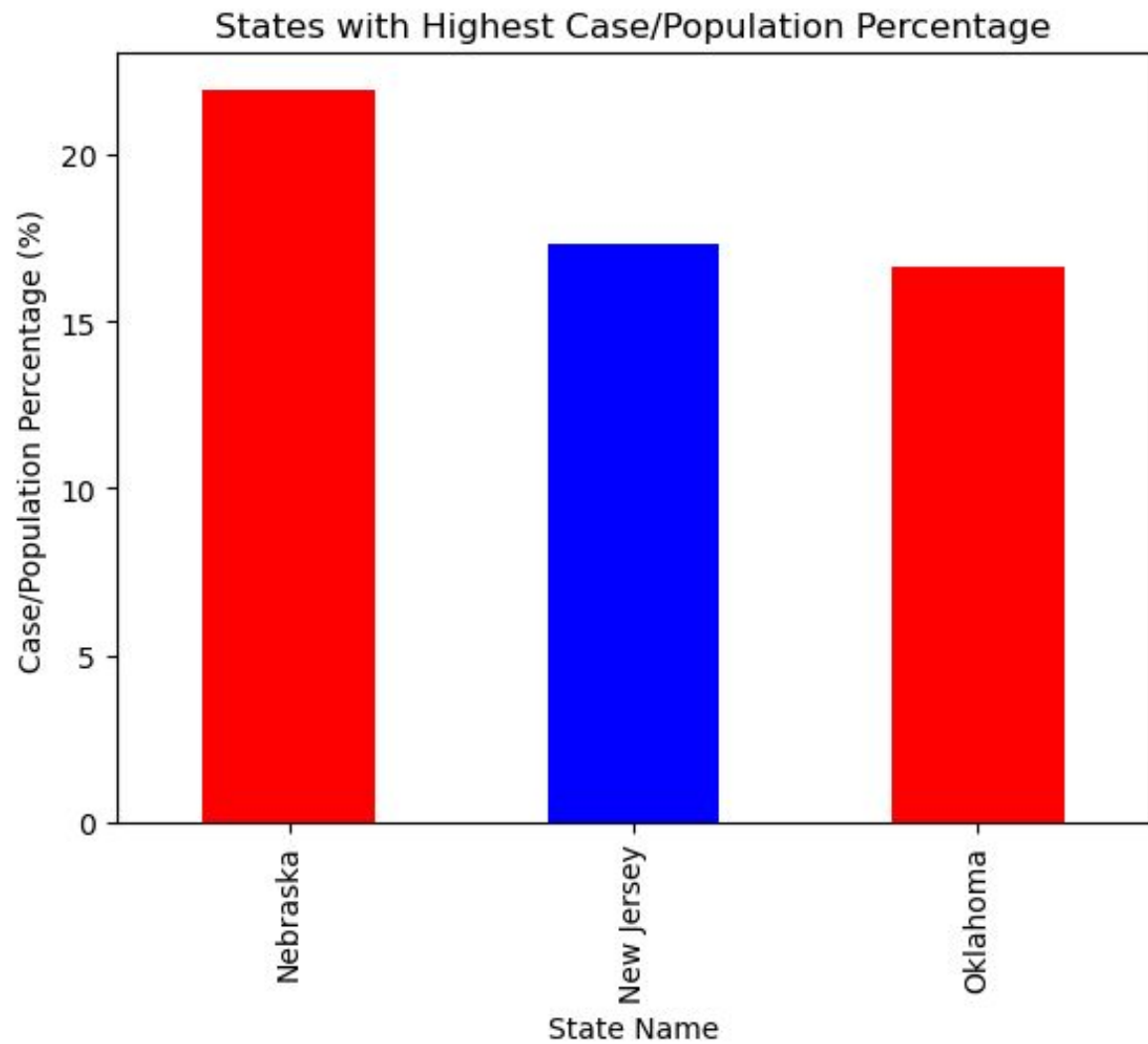


The three lowest number of cases by state

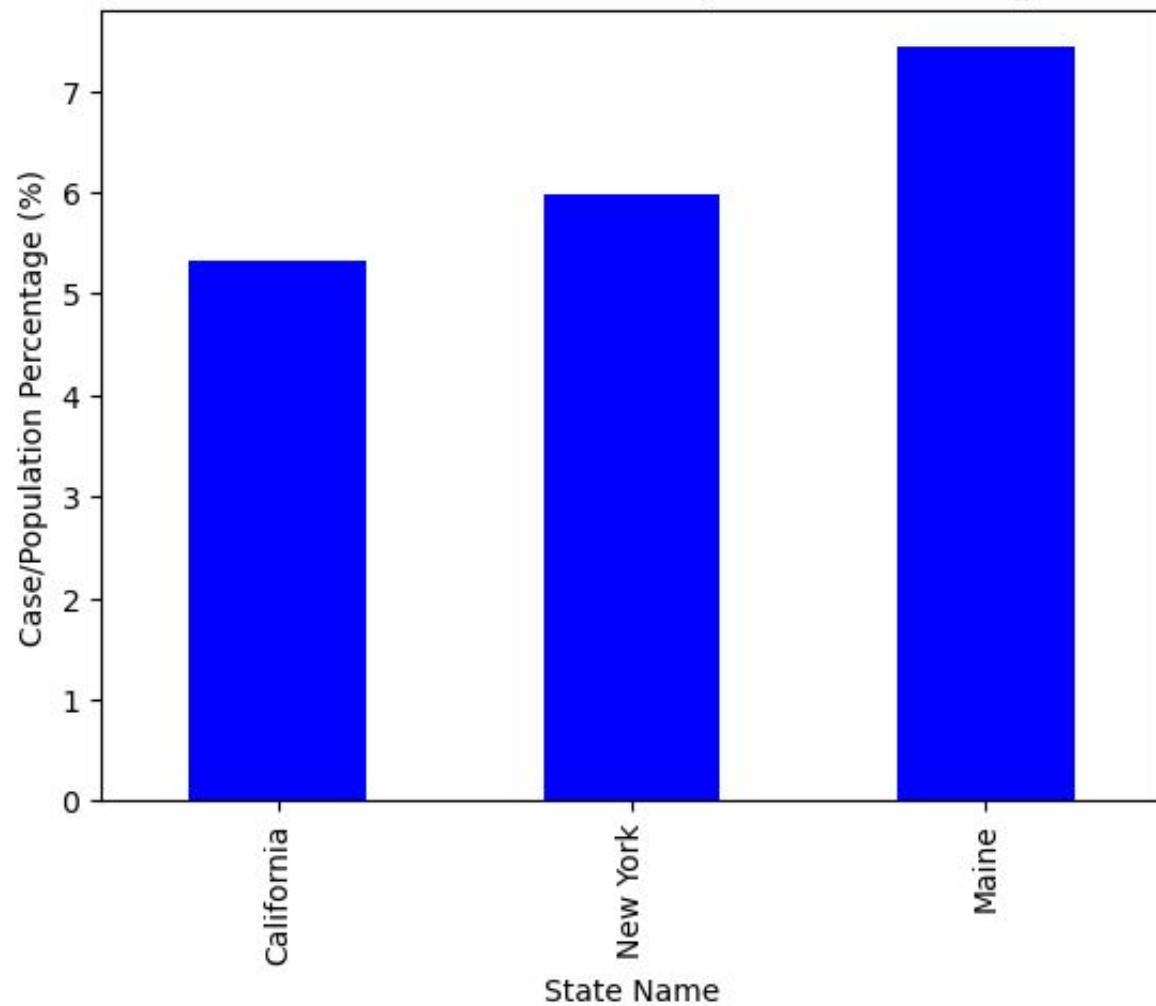


Case/Population Percentage by Red/Blue States



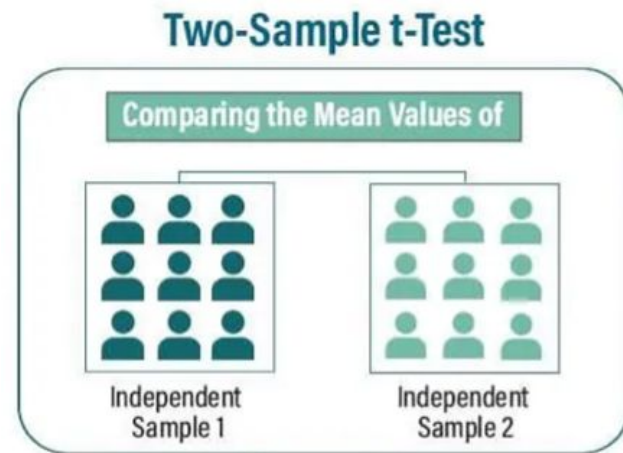


States with Lowest Case/Population Percentage



Statistical Test: t-test

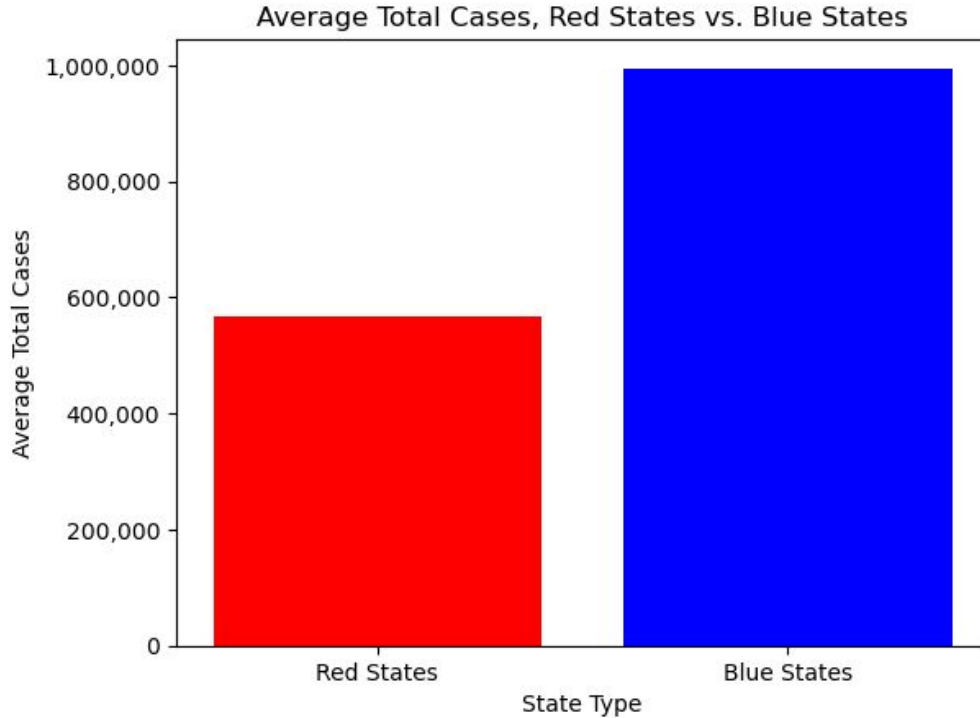
- Based on the data, independent t-test was most appropriate to test the hypotheses
- Did 2 t-tests:
 - Using total number of cases per state by color
 - Using total number of cases by population percentage per state by color
- Results:
 - Total Cases:
 - t-statistic: 33.592271324857755
 - p-value: 8.142141668872524e-242
 - Case Prevalence
 - t-statistic: -22.420479534465183
 - p-value: 3.330387755169245e-110



P-value

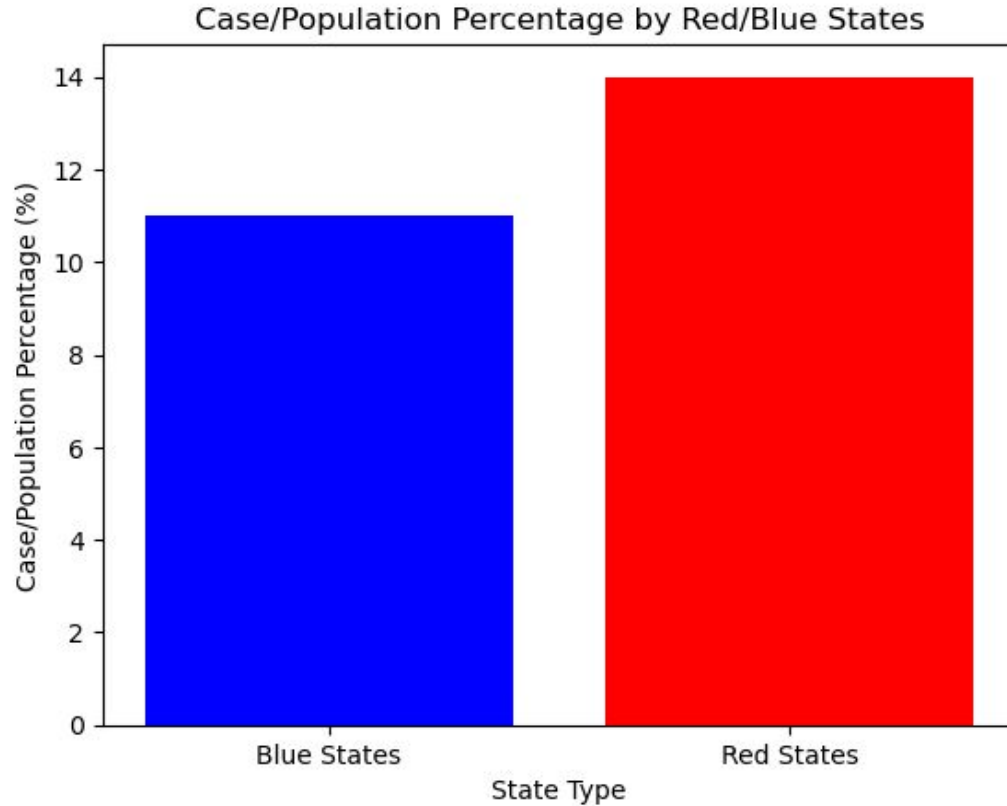
— — —

- P-value is less than 0.05 in both tests
- Can reject the null hypothesis
- Can accept the alternate hypothesis that the type of state government (**red/Republican** or **blue/Democrat**) has an impact on the number of COVID-19 cases.



Blue States:

- Larger populations, tend to be clustered around either coastline
- Larger cities/business hubs (i.e. LA, NYC, Chicago, Atlanta)
- Greater access to large airports
- Higher populations = higher number of cases



Red States:

- Smaller populations, tend to be located in the middle of the country (more isolated)
- Republican legislation more relaxed about vaccination and masking
- Although less unique individuals were diagnosed with COVID-19, a greater majority of the population got sick