The first step of stage II involved finding the center values of the weekly data (deaths and cases) for the state Oregon. Here it was clear that the mode was not a useful measure of center for this dataset. For normalized cases, Oregon was similar to Washington in mean weekly cases. California, North Carolina, Virginia, and New York were higher, but similar to each other. For normalized deaths, all the previous states had the same mean weekly deaths. The top 5 counties with the highest case rates were Malheur, Umatilla, Morrow, Jefferson, and Marion Counties. The top 5 counties with the highest death rates were Malheur, Hood River, Jefferson, Morrow, and Umatilla Counties. The daily trends for Oregon and these counties were all distinct.

The next step of stage II involved fitting a distribution to the new case data from the state of Oregon and comparing it to the distributions of 5 other states. For the normalized data, all states were skewed to the right, with the highest peak at the lowest range of new cases and a long tail out to the higher ranges of new cases. Oregon's data had a skew value of 3.5 and a kurtosis value of 17.1, showing it was skewed and had a high peakedness at the lower ranges. The Washington data had a higher skew and a much higher kurtosis value (5.8 and 44.5 respectively). The California data had a lower skew value, and a lower kurtosis value (2.2 and 4.2 respectively). The North Carolina data had a lower skew value and a higher kurtosis value (4.0 and 29.4 respectively). The Virginia data had a lower skew and a lower kurtosis value (1.8 and 3.4 respectively). Finally, the New York data had a lower skew value and a very low kurtosis value (1.2 and 0.6 respectively).

After this, the next task involved modeling a Poisson distribution for Oregon and 5 other states and for counties in North Carolina for both normalized new cases and new deaths. All the new case distributions had a similar shape to each other and all the new death distributions had a similar shape to each other. The differences came with at what number of new cases or new deaths had the highest probability of occurring on any given day.

In terms of correlation of the enrichment data with COVID-19 new cases, the only data variable available from the presidential election data is the number of votes each candidate of a party received. Performing correlation on the normalized number of new cases for a given county and the number of people who voted for either the Republican, Democratic, or other parties resulted in all negative correlation values. Meaning when one goes up the other goes down. The highest correlation was found between the number of people voting for other parties and the mean number of new cases for a county with a value of -0.52, where the value for the democratic voting was -0.41 and for the republican voting was -0.21. Regardless, none of the correlation values were strong.

The next step was to develop hypotheses relating the COVID-19 case data to the enrichment dataset, the presidential election data in this case. The three hypotheses are:

1.) Does a higher number of voters (normalized) for a certain party lead to a higher initial increase in COVID-19 cases?
2.) Does a higher number of voters (normalized) for a certain party lead to a higher number of COVID-19 cases overall?
3.) Do counties with a certain party winner have a higher number of COVID-19 cases overall (normalized)?