As COVID 19 has impacted almost everyone in the world directly or non-directly. We wanted to analyze the impact of the virus on the United States. We have seen the impact of the virus here in the US since March of 2019. The volume of fresh and relevant data was overwhelming as we went from website to website to locate the data sets needed. As we found data sets, we had to pivot from our initial proposal to review the impact of restrictions as we found this data would have turned very subjective. For example, one state’s red level could be a near 100% shut down while another state’s red level could have just limited social gatherings and entertainment openings. Given the time constraints for this project we did not feel we could develop a consistent method of mapping state level restrictions.

Our initial collection of data sets included various components needed for analysis. We were able to secure the data need to evaluate relationships between the virus and unemployment, vaccine distribution, contributors to the spread of the virus, and how the virus has impacted altered the health of the nation. This data was collected from the CDC, Bureau of Labor Statistics, US Census, as well as various data sites including Dataworld.com and Keggel.com.

**Analysis of Vaccine Distribution to the US**

Did the Government Distribute the Vaccine as they described?

What possible considerations did the government consider when planning the distribution of the vaccines?

While compiling the data needed to review the distribution of Vaccines to various effects of the virus such as confirmed cases, deaths, age, gender, population, population density. It was discovered that information within the CDC Provisional statistics pertaining to age were inconsistent with gender statistics. Upon reviewing state level case statistics, it was decided to move forward without age grouped information as it did not match state level while gender statistics matched.

We reviewed three potential scenarios to find the methodology applied to the distribution of the vaccine.

1. Total Allocated Dosage vs Population Density

We looked at this relationship due to the ease of spread of the virus. We expected that more densely populated state would exhibit greater spread and be in higher need of the vaccine. This was not the case as there was virtually no relationship (correlation factor of .17) between the distribution and how many people were located per square mile.

1. Total Allocated Dosage vs CDC total confirmed cases per state

While reviewing this relationship we expected there to potentially be a strong relationship between the distribution and total confirmed cases to help slow the spread and work to heard immunity there was a strong relationship between these variables of .97.

1. Total Allocated Dosage vs Population

The near perfect correlation of .99 between these data sets show that the government followed thorough with their roll out of the vaccine based on population. The relationship posed a few additional questions we reviewed from our data.

While developing the answers to why the government distributed the vaccine as they did, we explored a few additional questions that arose. It is heavily reported on the ease at which this virus spreads. We looked to see if there was a correlation between CDC confirmed cases and Population Density. Much to our surprise population density only had a correlation factor of .14 while CDC confirmed cases and Total Population showed a near perfect correlation of .98.

We also reviewed to see if there were any relationships between the Pfizer and Moderna vaccines. We found no clear relationships to the breakdown of distribution of the manufacturers. The only relationship was the total allocations to the total populations.

**Analysis of Passenger Count on the top 10 Airports in the US:**

Does the additional Passenger Count from the local airports going in and out of a state have an impact on the COVID-19 case count within that state?

We developed a theory on what we concluded about the effect that the airports have on the COVID-19 cases:

The correlation between the top 10 US airports with the airport passenger count of the top 10 states with the highest Covid-19 case count should have a great impact on the case counts with that area.

We first discussed our theory that the extra traffic coming and going in any give airport would have a great impact on the COVID-19 case count within that area. To support the theory, we got the top 10 US airports form ffa.gov. We pulled the top 10 from that data to further analyze the effects. Then we add the population counts to the passenger average that passed through those states. We then compared the outcome to the states with the top 10 cases counts. We noticed a pattern between the state’s population and the states with the highest case counts.

Inclusion the data does seem to support that the airports have an effect on the COVID-19 cases to some point. It the study group that was pulled, 70% of the states follow the same pattern as the cases count and population. The other 30% drops off. The data is also showing us that the state’s population plays a bigger role in the case counts.

**Analysis of the Effects Covid-19 Case Count Effects on the Unemployment Within the State:**

Is there a difference in the amount of unemployed in 2019 to 2020? If so, what was the difference?

Does the number of COVID-19 cases have any effect on the amount of people that are unemployed within that state?

After analyzing the data on the amount of people unemployed in 2019 and 2020, it was clear that the amount in 2019 were more than doubled in 2020 for most states.

While looking at the confirmed case count to the amount of people unemployed, it showed that it had a great impact. If you look at the top 5 states and the lowest 5 state effect by confirmed COVID-19 cases, you can see that the amounts are close enough that it indeed had an effect. In regard to the first 3 of the top 5 states, there could be additional factors that played in the lower unemployed numbers verse case count. Were any of the confirmed COVID-19 cases from the same person taking multiple test? How many of the confirmed cases were employed before they contracted COVID-19?

In conclusion, we feel that the data shows that Covid-19 has a huge impact on the amount of unemployment that the is currently plaguing our country.

**Analysis of Covid-19 Deaths vs. Influenza Deaths**

The data utilized was retrieved from the **National Center for Health and Statistics** (**NCHS**).

Utilizing Line Graphs

**Step 1**: Evaluate **Total Deaths** for the past five years.

Findings: There was **325,221** increase in **Total Deaths** from 2019 to 2020. In a normal year, most deaths accumulate during the winter months. Mainly when many people are the least active, given cold weather conditions. This is also coincides with the peak of the flu season.

**Step 2**: Evaluate **Pneumonia Deaths** for the past five years.

Findings: Total pneumonia deaths saw about a **64.6%** increase from 2019 to 2020. Pneumonia deaths normally account for most of the deaths associated with viral respiratory illnesses, given other viral symptoms are mostly manageable with existing medications.

**Step 3**: Evaluate **Covid-19 Deaths** for 2020.

Findings: There were **235,745 Covid-19 Deaths**.

**Step 4a**: Calculate **Covid-19 Pneumonia Deaths** in targeted 2020 timeframe.

**Step 4b**: Calculate **Covid-19 Non-Pneumonia Deaths** in targeted 2020 timeframe.

**Step 4c**: Calculate **Influenza Pneumonia Deaths** in targeted 2020 timeframe.

**Step 4d**: Evaluate **Targeted Pneumonia Deaths** in targeted 2020 timeframe.

Findings: This chart analyzes pneumonia deaths from the first Covid-19 death associated with it to early November, and compares it to previous timeframes in 2016-2019. Noticed that 2020 deaths other than Covid-19 seemed slightly influenced by it.

**Step 4e**: Use correlation analysis to verify **Influenza Pneumonia Deaths** were statistically consistent with the same timeframe in previous years.

Findings: Correlation testing was done comparing 2020 pneumonia deaths other than Covid-19 with the previous years displayed with a minimum factor of **0.82**, which represents a strong correlation.

**Step 4f**: Use ANOVA analysis to verify **Influenza Pneumonia Deaths** were statistically distributed with the same timeframe in previous years.

Findings: ANOVA testing was done on the correlation datasets, generating a **0.33** **pvalue** indicating there is no significant statistical difference in those datasets.

Utilizing Pie Charts

**Step 5**: Compare 2020 **Influenza Pneumonia** and **Non-Pneumonia Deaths**.

Findings: Pneumonia **94.7%** (**168,377**), Non-Pneumonia **5.3%** (**9,361**)

**Step 6**: Compare 2020 **Covid-19 Pneumonia** and **Non-Pneumonia Deaths**.

Findings: Pneumonia **46.0%** (**108,462**), Non-Pneumonia **54.0%** (**127,283**)