

2020 Covid-19 and Unemployment Impacts by U.S. Regions

Python for Data Science

Introduction

The COVID-19 (Covid) pandemic has enormously impacted various aspects of society throughout the world. This is clearly a topic that will be researched through numerous lenses for years to come and will benefit greatly from the data collection efforts undertaken by agencies across the globe, nationally, and locally.

Our goal for this project was to examine a small slice of these impacts with data currently publicly available. In reviewing Covid case and death figures provided by the U.S. Centers for Disease Control (CDC), we are interested to learn how strongly correlated these numbers are with monthly unemployment figures posted by the U.S. Bureau of Labor Statistics (BLS).

We hypothesized that the U.S. Census Bureau-defined Midwest region would show a weaker correlation than other regions. We believed that employment in the leisure and hospitality industry had been strongly impacted by the pandemic, and as the Northeast, South, and West regions are more popular travel destinations, we suspected that the unemployment rates would be higher in those regions. Specifically, we explored our data sets to shed light on the following:

1. We suspect that the Midwest as an inland region was the slowest to be impacted by Covid from a case count perspective as the virus was first imported from abroad.
2. We suspect that the region with the highest proportion of leisure and hospitality industry employment will show the highest unemployment rate increase due to Covid, and vice versa.
3. We suspect that Covid case counts are correlated to the unemployment rate, and the correlation will be stronger on the Coast and less strong in the Midwest.

The data sets chosen to inform our analysis are the most complete, accurate and reputable available. They include:

- U.S. CDC COVID-19 data, (daily), inclusive of the entire U.S. population of around 330 million as a whole
- U.S. unemployment rate data and U.S. labor force by region data, (monthly), sample size is about 60,000 eligible households of approximately 110,000 individuals, representing the underlying population of over 160.5 million in the U.S. civilian labor force
- U.S. Leisure and Hospitality (L&H) industry employment data, (monthly), sample size is about 69,400 business establishments covering about 2.8 million employees that represent an estimated 15.7 million employees in the L&H industry
- This paper presents the findings of our data analysis exploring the relationships between Covid cases and deaths, unemployment, and the hospitality and leisure industry through a regional lens. We are interested in the topic because there has been an enormous economic impact to families and individuals and will be for years to come for society as a whole in terms of relief package funding. As there will be future epidemics/pandemics, insights drawn from the analysis of these data sets will allow for better response planning at all levels.

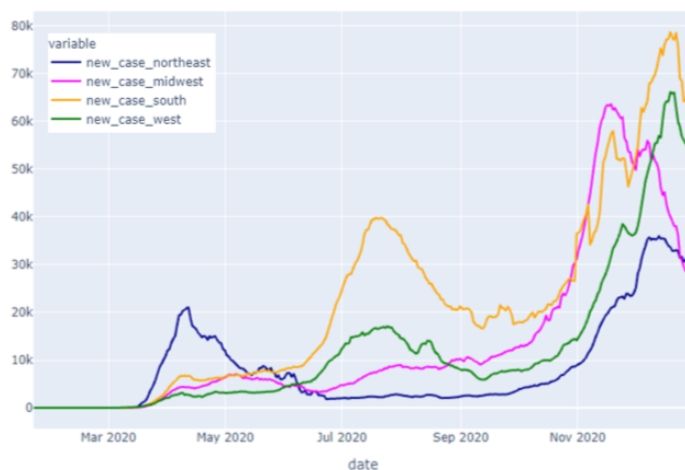
Data Preparation and Analysis

In reviewing the data sets, we identified variables and data types; reviewed completeness, basic metrics, and outliers; rolled values up to the regional level, and harmonized date fields across data sets. Note that (Un)Employment source data is captured at the monthly level, while Covid figures are recorded daily. For inter-data set analysis, we rolled Covid data up to the monthly to facilitate comparison. For the full data dictionary, see the EDA workbook, and for the key items, see the table below. Fields calculated to facilitate analysis are noted.

Field	Source	Category	Description
Month	Harmonized across data sets	int64	Month in 2020
Region	Harmonized across data sets	object	Region name, as defined by U.S. Census BLS data
New cases	U.S. CDC Covid-19 data	int32	Daily confirmed Covid cases in reporting area
New deaths	U.S. CDC Covid-19 data	int32	Daily confirmed Covid deaths in reporting area
Cumulative cases	Calculated from CDC New cases	int32	Summation of new Covid cases
Cumulative deaths	Calculated from CDC New deaths	int32	Summation of new Covid deaths
Cumulative case rate	Calculated from Covid cumulative cases	float64	Cumulative Covid cases per regional capita
Cumulative death rate	Calculated from Covid cumulative deaths	float64	Cumulative covid deaths per regional capita
unemp_rate	Monthly U.S. BLS unemployment rate data	float64	Percent unemployed by area based on unemployment claims
unemp_rate_percent_change	Calculated from U.S. unemployment rate	float64	Month-to-month month percent change of unemployment
LH_employment	Monthly U.S. BLS L&H industry employment data	int64	Size of H&L workforce
LH_employment_percent_change	Calculated from U.S. L&H industry employment data	float64	Month-to-month percent change of L&H employment
labor force	Monthly U.S. BLS labor force by region data	int64	Size of U.S. workforce in 2020
LH_employment rate	Calculated BLS L&H employment and labor force	float64	Percent of labor force employed in L&H
LH_emprate_percent_change	Calculated from U.S. L&H industry employment data	float64	Month-to-month percent change of L&H employment rate

A look at the Covid-19 data

Daily new covid counts (7 day moving average)



Daily new covid deaths (7 day moving average)

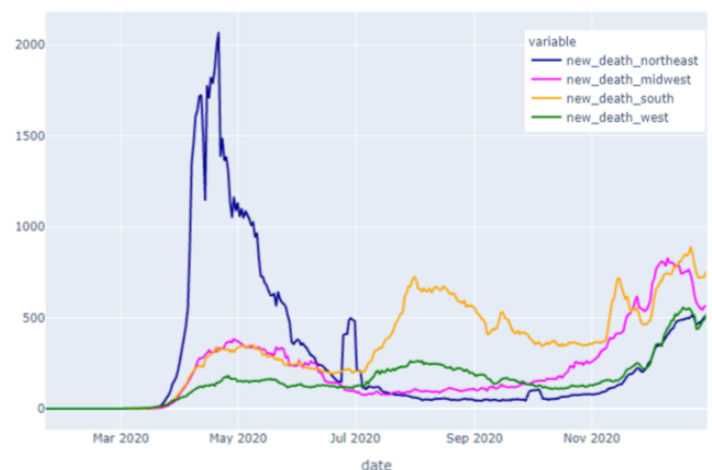


Figure 1: 7 Day Moving Average of New Cases and Deaths and Correlation Heat Map

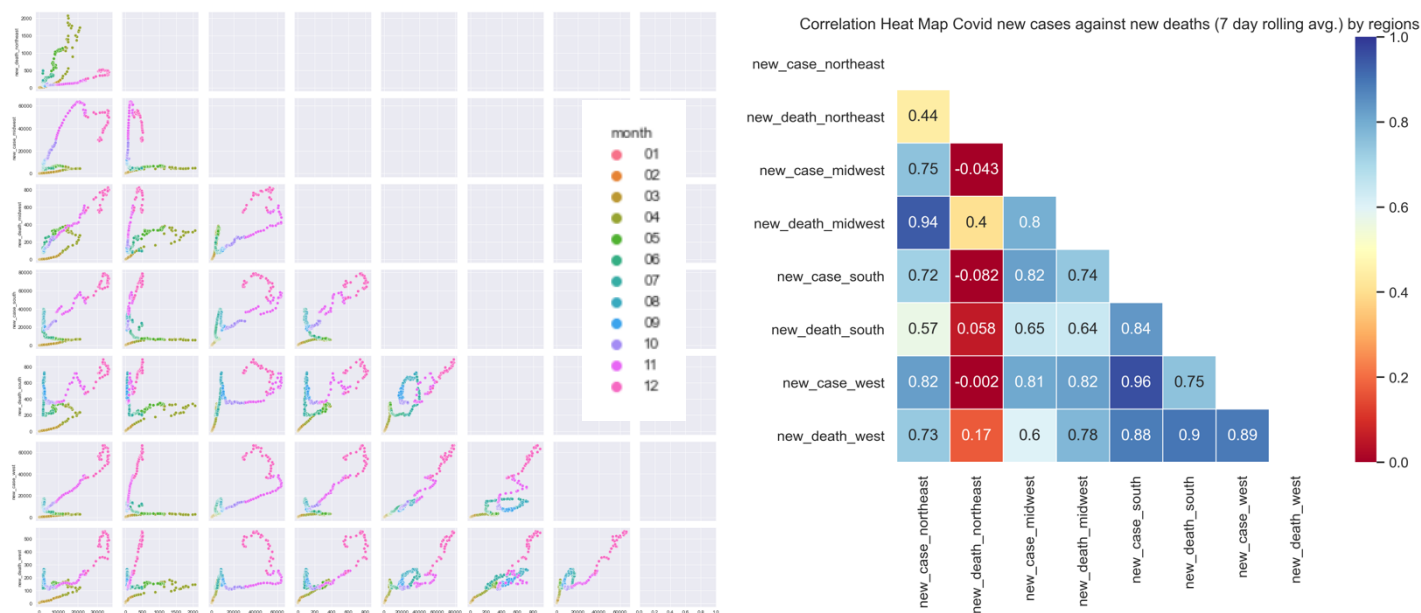


Figure 2a (left): Pair-plots. 2b (right): Corresponding correlation heat map between new Covid cases and deaths (7-day rolling averages)

In reviewing Covid cases and deaths by region, Figure 1, we can see that all four regions have a general trend of growth for new Covid cases. New cases in the Northeast rise first in March, and spike in April. April deaths in the Northeast are remarkably high in April, and interestingly, new case numbers are low following, perhaps reflecting the significant restrictions put in place in the at the regional level in the Northeast.

One might have predicted that the number of cases and deaths would be similarly correlated across regions, but as we see in Figure 2(b), the Northeast data is an aberration, as the correlation coefficient is markedly lower than that of other regions. This may be due to the timing of new cases, hospital capacity, improvements in treatment protocols over time. The contrast in numbers from the first and second halves of 2020 is stark, explaining the low correlation coefficient. We see the situation much improved in the second half of 2020, as evidenced in the top left tile in pairplot Figure 2(a).

We also noted that the new cases in the West and South are very highly correlated, perhaps reflecting weather seasonality, hospitality and leisure employment rates, or state and local policy similarities.

Which regions were impacted first by Covid cases and deaths?

As seen in Figure 1, in terms of new cases and new deaths, the Midwest is not the slowest to be impacted, but does see a significant spike in new case numbers later than the other regions.

When we look at the cumulative cases normalized by regional population size, it is confirmed that the Midwest was on par with the South and the West to be impacted (Figure 3).

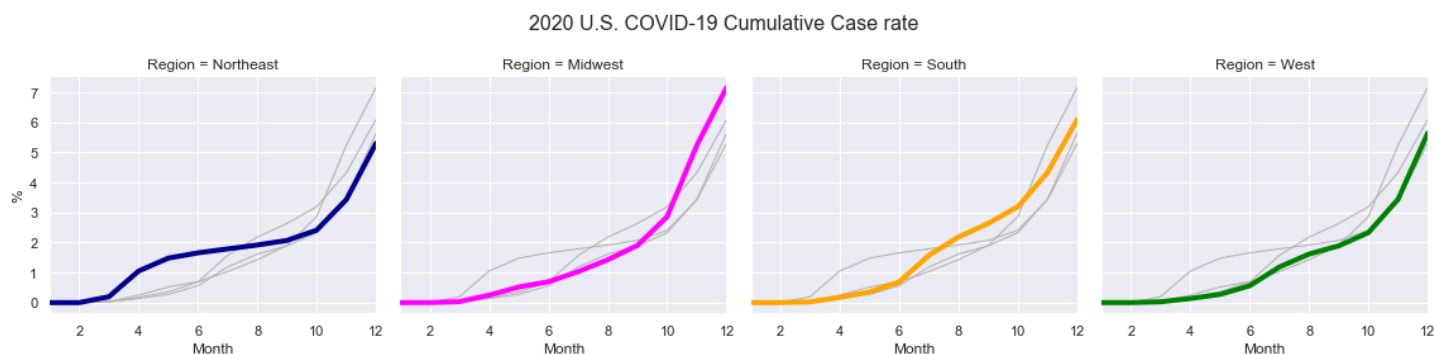


Figure 3: Cumulative Covid Case Rate by Region

A look at the (Un)Employment data

2020 U.S. Unemployment Rate across all industries

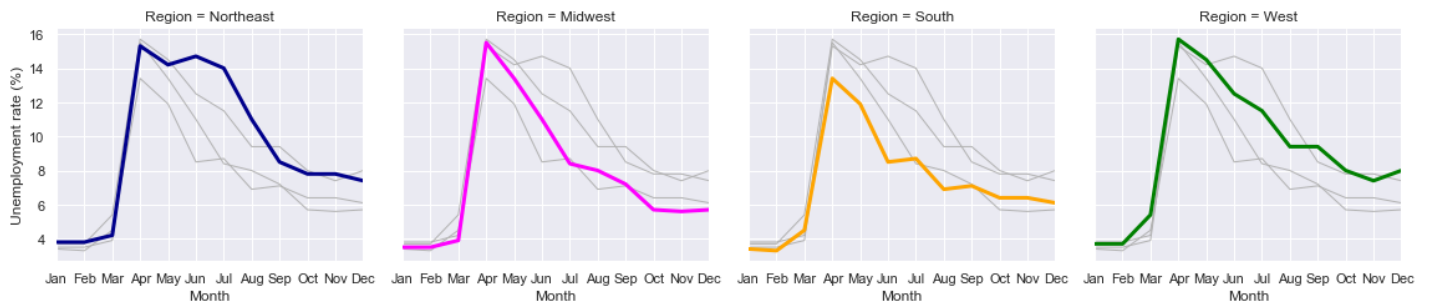


Figure 4: Unemployment Rates by Region, All Industries

As seen in Figure 4, all regions experienced a catastrophic increase in unemployment, although the recovery curves are much different. The Northeast, Midwest, South and West, started out 2020 with less than 4% unemployment, and all but the South peaked in April with over 15%. *From a month-to-month difference perspective, the South and West had an early unemployment rate uptick in March, but the Midwest had the largest percent increase in April of all regions.*

The Northeast took the longest to get under 14% unemployment, and as of December, the West has the highest unemployment. To the lay observer, this could be due to stronger Covid restrictions in the Northeast, a more significant fear in the population due to the onset date or death rates, and/or a higher number of L&H employees in the West. Further analysis will be required.

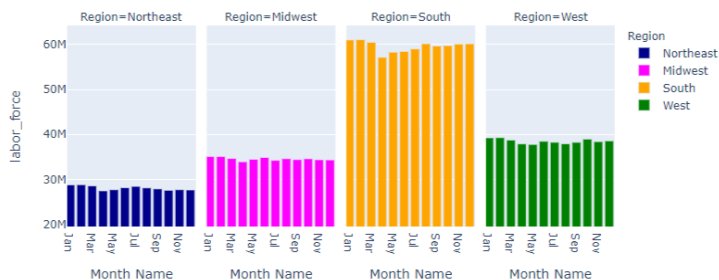


Figure 5: U.S. Labor Force Population by Region

As seen in Figure 5, the South represents 37% of the U.S. labor force in January, which was larger than anticipated, and the Northeast the smallest at 18%, which is consistent with the overall population figures.

Which regions have the highest percentage of H&L jobs? Are unemployment figures higher?

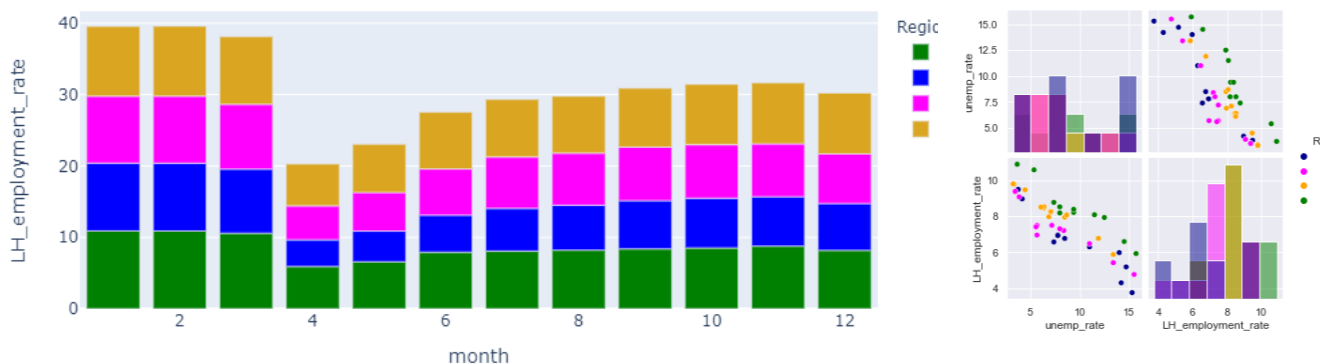


Figure 6a (left): 2020 U.S. L&H Employment Rate by Month. Figure 6b (right): Unemployment Rate and L&H Employment Rate Pair-plot

As suspected, Figure 6a shows the Midwest started out the year with the smallest percentage of hospitality and leisure workers as a percentage of the labor force at 9.4%, and the West was highest at 10.9%. We can see that the number of people employed in the L&H industry is still significantly less than early pandemic numbers. As a percentage of the labor

force, the South has recovered the most, while the Northeast the least, with a delta of 1.2% and 2.9%, respectively. (Note that the Midwest has a difference of 1.4%.)

Referring back to figure 4, there are a few observations worth noting. The West was the first to reach a 5% unemployment rate and peaked at the highest rate of all regions, which is aligned to our initial thought that the region with the highest percentage of H&L workers would experience the most unemployment, though the South, the second highest of L&H employment, had the lowest peak unemployment. This could be due to the “snowbird season” or reflect a lighter hand in restrictions put into place to curb the transmission rate of the virus.

Are there relationships between Covid numbers and unemployment figures?



Figure 7: New Covid Cases by Region

Again, referring to figure 4, in comparison to the above, we observe that new case numbers do not appear to be correlated to unemployment, as unemployment across regions hits an apex in April, and new cases do not peak until much later in the year.

However, the monthly percent change in cumulative new cases and percent change in unemployment, are very highly correlated with correlation coefficients between .94 and .99, as shown in the below Figures 7, 8a, and 8b.

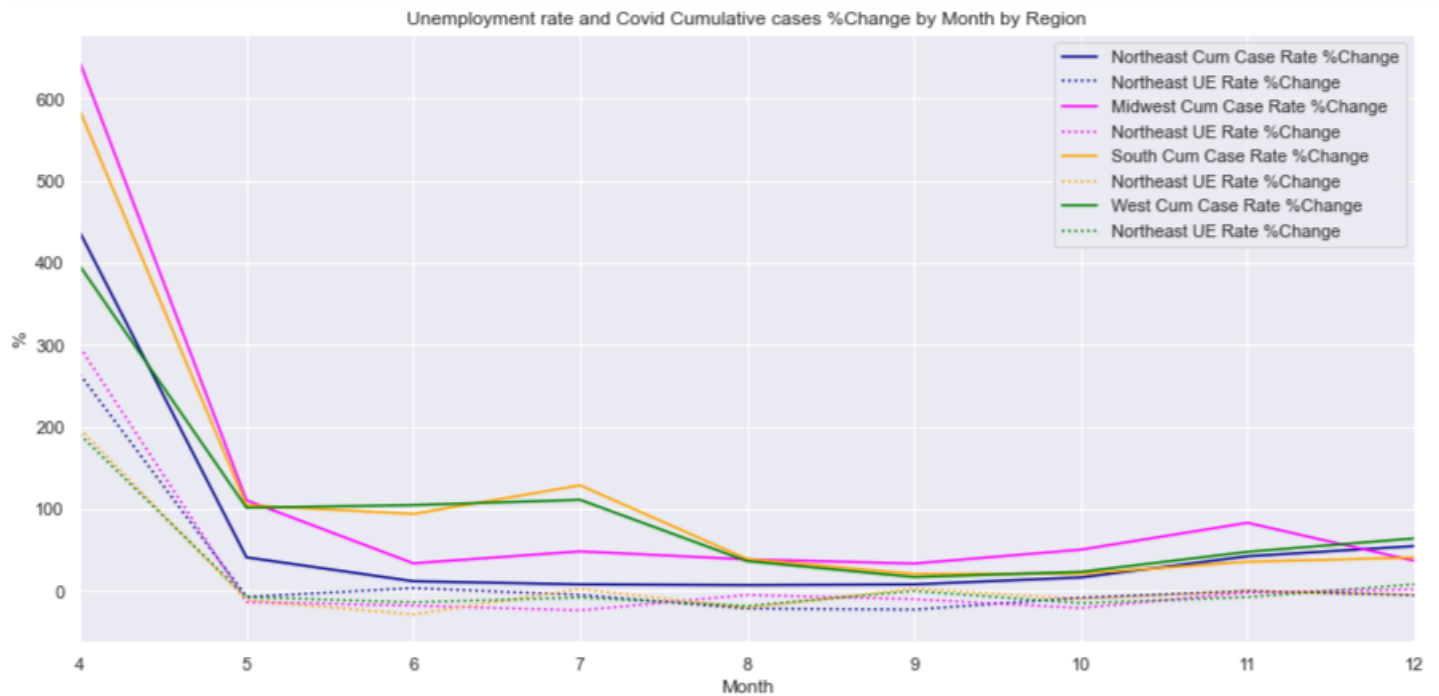
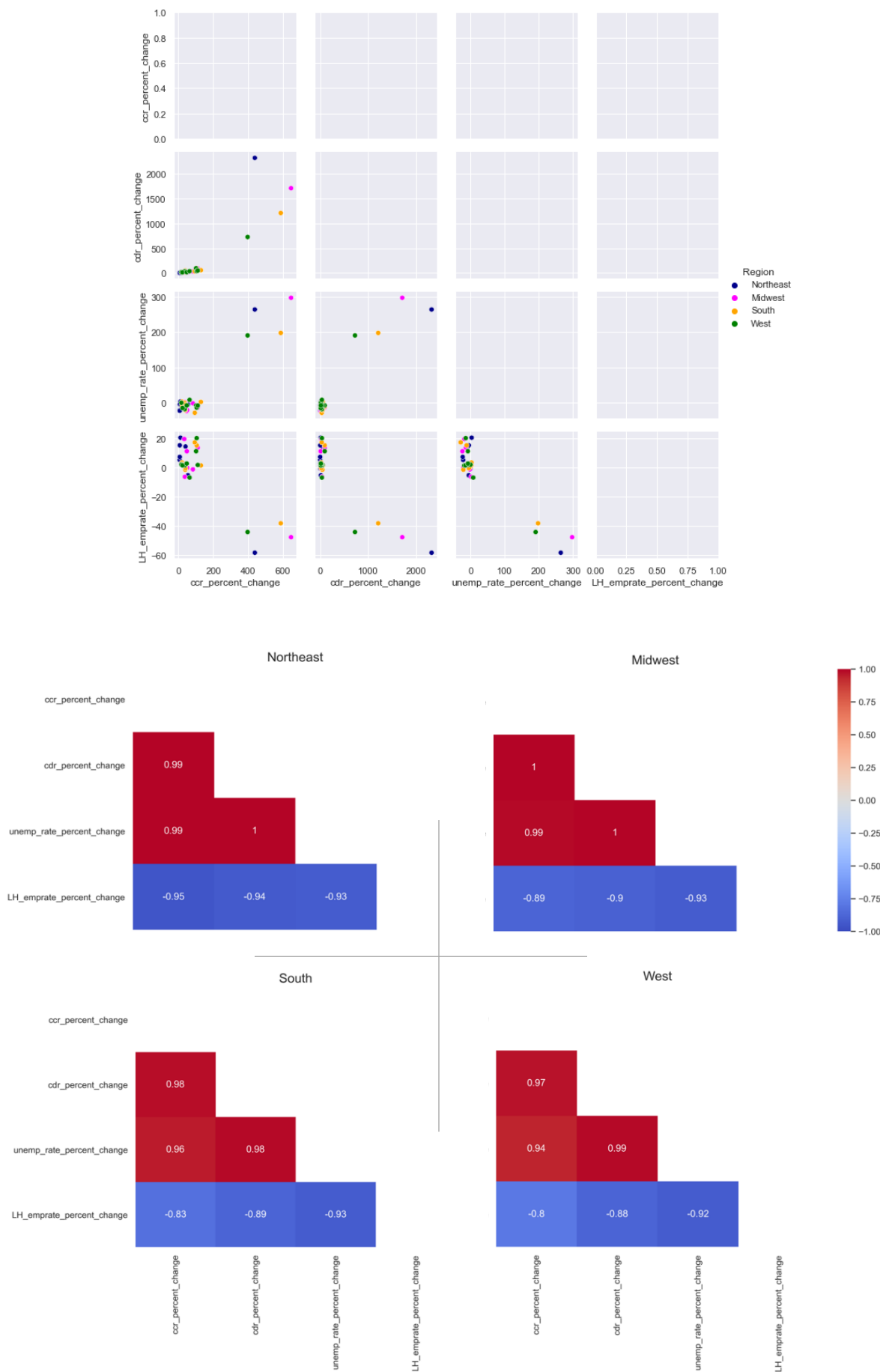


Figure 7: Covid cumulative case rate monthly % change vs unemployment rate monthly % change

Figure 8a (top): Pair-plots. 8b (bottom): Corresponding correlation heat map between monthly Covid cumulative case rate (ccr) %change, Covid cumulative death rate (cdr) %change, unemployment rate %change, and L&H employment rate %change



Conclusion

Through our data analysis, we have validated that the unemployment is correlated to Covid-19 rates, and while general trends across regions in the analyzed data is similar through a wide-angle lens, as just this small slice of data shows, regional impacts across various measures were not uniform.

As discussed above, and relating back to our original hypotheses:

- The Midwest was not the slowest to be impacted by Covid cases. Perhaps the bigger lesson here is that given the amount people travel, viruses move quickly. The Northeast was the most significantly impacted early on especially in terms of Covid deaths, likely as a result of high population density or travel hubs in New York City.
- The West had the highest percentage of H&L workers, and showed the highest peak unemployment figure, though the South had the next highest percent of H&L workers, but the lowest peak unemployment rate.
- The rate of change in Covid case counts and deaths is strongly correlated to the rate of change in unemployment. The one difficulty we encountered with the data is that in the first three months of 2020, there were no cases, then a few, and then many, so the rate of change, based on the previous month, did not allow for a meaningful analysis, thus we had to start with April. The data disclosed that this relationship was strongest in the Midwest and Northeast, but only slightly so. The visualizations clearly illustrate that with the introduction of the Covid virus into the U.S., the impact to employment/unemployment was swift and vast.

As we know, the actual Covid-19 virus did not cause unemployment directly, but our collective reaction to it at government, societal, and individual levels did. Fascinating and socially impactful analysis that will influence future policy we expect to see is unemployment, labor force, and GDP figures as they relate to different policies enacted across states and regions. Cross-sector analysis will be similarly important though economic and social/humanitarian perspectives. We additionally anticipate comparator analysis with other countries to help inform better ways of managing the response to future epidemics/pandemics on all levels.