**Motivation**

As COVID-19 swept through America in 2020, much research has been conducted on its effects over peoples’ health and lives. The initial wave of the pandemic quickly leads to a rise in the unemployment rate in many industries, yet the economy seems to slowly recover towards the end of 2020 despite the continuing high number of confirmed cases per day. For this assignment, I’m planning to study how the number of confirmed COVID-19 cases impact the unemployment rate in Denver, Colorado.

Understanding the relationship between infectious diseases and the economy is necessary for the public as well as policy makers. As public health measures were taken to slow the spread of coronavirus, economic activities were stalled due to slow flow of goods and people. Thus, it is important to quantify how peoples’ lives have been impacted by COVID-19 through economic indicators, such as unemployment rate, to better investigate changes in human activities during the pandemic.

**Research Question**

My research question is how the unemployment rate was impacted by the pandemic in Denver, Colorado? It is unsurprising to find out there is a huge spike in unemployment rate in the initial couple months of the outbreak, so I’m more interested in learning the trend of unemployment rate after months of the outbreak to see if the economy has recovered, and if so when it started to recover.

**Data Used**

I will use data provided by the U.S. Bureau of Labor Statistics. Specifically, I will use the dataset about unemployment rate in Denver, Colorado as the additional data for my project. The dataset provides monthly unemployment rates ranging from 1990 to 2021, but I will most likely only use data in recent years, such as 2019-2021, to fit in the context of the research. The link of the data is provided below.

<https://beta.bls.gov/dataViewer/view/timeseries/LAUCN080310000000003>

According to the copyright information listed on the website of the U.S Bureau of Labor Statistics, I am free to use anything they publish on the public domain without specific permission(Linking and copyright information, 2016). However, I do need to cite the Bureau of Labor Statistics as the source once using the data.

There is no ethical consideration to use the dataset because the data presented doesn’t reveal any personal information. According to the U.S Bureau of Labor Statistics, the surveys designed strictly protect the confidentiality of the participants to ensure no information, either about individuals or businesses, will be misused(Looking for information about a specific survey?, n.d.). In addition, the response provided will only be used to generate statistics about groups with specific characteristics rather than specific individuals.

**Unknowns and Dependencies**

There are potential external factors that may affect the relationship between unemployment rate and confirmed COVID cases. For example, there are different types of orders issued by the government that ask people to stay at home or require restaurants to control the percentage of dine-in customers. Those policies weren’t necessarily issued with the mask policy at the same time, nor are they for sure related to the trend of confirmed cases. However, it is possible that they may affect the unemployment rate in a way, so they should be taken into consideration when analyzing data. Furthermore, the final visualization could even mark the critical time about when different policies went effective to provide additional information and give a more complete context.

**Methodology**

To answer my research question, I need to test to see if there is any correlation between the two time series – unemployment rate and COVID-19 cases. Data about the unemployment rate can be directly downloaded as a csv file from the link given in the previous section, with no change needed to be made. As for the COVID-19 cases, I will continue to use the data from A4. However, to improve the visualization, I decided to apply a smoothing technique by averaging confirmed cases in a rolling window of a week to avoid showing too many spikes.

To measure the correlation between the two time series, I choose to use both the Pearson correlation coefficient and the cross correlation function(Wright, 1921; Podobnik & Stanley, 2008). I will start with calculating the Pearson correlation, which ranges from -1 to 1, to measure the global synchrony of the two time series. In addition, it’s also possible to look at local synchrony from a moment-to-moment perspective by using Pearson r(Jin Hyun Cheong, 2021). However, one assumption held by using Pearson correlation coefficient is the homoscedasticity of the data, which might not be true given the fluctuations in both two time series. As a result, though Pearson correlation coefficient provides an overall insight, I need to further my investigation with the Time Lagged Cross Correlation.

The Time Lagged Cross Correlation can provide insights about dynamic signals by repetitively calculating correlation between two time series, with one time series constantly shifting. By looking at the correlation at different offset, I am able to detect how the two time series synchronized over time. In addition, the Cross Correlation method assumes the stationarity of time series, which makes sense in our case since neither unemployment rate nor COVID-19 are heavily affected by seasons.

As for the final output, I’m planning to use different visualizations to show the trend of unemployment rate and COVID-19 cases and the two types of correlations. As discussed in the previous section, critical time nodes may also be shown on the graphs if important policies other than masking are considered to potentially affect unemployment rate so that a more comprehensive background is given for viewers to fully understand the relationship.

**Timeline To Completion**

I will finish data collection and model building by Nov. 15th to have a sense of how the time series of unemployment rate and COVID-19 cases relate to each other. Then, I will spend around a week to analyze what I find from the model and hope to finish it by Nov. 23rd. Around that time, my data analysis should contain reasonings by potentially using other resources online to explain what I observe from the results of the model. By Dec 1st, I should complete the visualization about my findings and start to work on the presentation. I plan to finish the presentation slides by Dec. 6th to turn it in on time and will work on my documentation afterwards. I will finish my report by Dec. 10th, leaving two days for unexpected difficulties in case things don’t go as planned. A brief outline about my timeline is shown as below:

Nov. 15th: Data collection & Model building

Nov. 23rd: Data analysis

Dec 1st: Visualization

Dec 6th: Presentation preparation

Dec 10th: Documentation

Reference

Jin Hyun Cheong, P. D. (2021, March 13). *Four ways to quantify synchrony between time series*

*data*. Medium. Retrieved November 10, 2021, from

https://towardsdatascience.com/four-ways-to-quantify-synchrony

-between-time-series-data-b99136c4a9c9.

Podobnik, B., & Stanley, H. E. (2008). Detrended cross-correlation analysis: a new method for

analyzing two nonstationary time series. *Physical review letters*, *100*(8), 084102.

U.S. Bureau of Labor Statistics. (2016, August 5). *Linking and copyright information*. U.S.

Bureau of Labor Statistics. Retrieved November 10, 2021, from

https://www.bls.gov/bls/linksite.htm.

U.S. Bureau of Labor Statistics. (n.d.). *Looking for information about a specific survey?*

U.S. Bureau of Labor Statistics. Retrieved November 10, 2021, from

https://www.bls.gov/respondents/.

Wright, S., 1921. Correlation and causation. *Journal of agricultural research,* 20(7), pp. 557–585