Final Project Study:
COVID19 Daily Rate of Infection, Vaccination, and Unemployment Rates in Fairfield County, Connecticut

Hasnah Said

1. Introduction

The latest threat to global health is the ongoing outbreak of respiratory disease named Coronavirus Disease 19 (COVID-19). COVID-19 was first detected in late 2019 in Wuhan, China, and its highly contagious nature made spread rapidly worldwide. In order to combat overburdening the health system with new cases, countries around the world implemented strict lockdown measures and restricted movements to slow down the spread and maintain the rate of infection at a manageable level. The lockdowns measurements helped considerably in controlling COVID-19 transmission, but it also affected different aspects of daily life; many people lost their jobs during the pandemic, schools had to adopt a remote model, and the supply chain was slowed down leading to global shortages. In December 2020, the first vaccine to combat COVID-19 was developed by Pfizer. The increase in vaccine administration to the public had a significant effect in slowing down the spread as well as opening up the world and slowly returning to what life used to be pre-COVID19.

In this paper, I would like to analyze the correlation of vaccines on COVID-19 rate of infection, hospitalization, and deaths related to COVID19 as well as the impact of COVID-19 on unemployment rates in Fairfield County, Connecticut.

2. Background/Related Work

2.1: Current COVID-19 Research

The safety and efficacy of COVID-19 Vaccines have been studied and researched thoroughly. The New England Journal of Medicine published a paper "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine" written by doctors supported by BioNTech and Pfizer. This paper shows that a two-dose regimen of BNT162b2 conferred 95% protection against COVID-19 in people 16 years and older. The CDC published a study that demonstrated the comparative effectiveness of the three vaccines used in the United States (Moderna, Pfizer-BioNTecc, and Johnson&Johnson) in preventing hospitalizations. This article found that Moderna had the highest effect in preventing hospitalization at 93%, it is followed by Pfizer at 88%, and finally Johnson&Johnson at 71%.

2.2: Assumptions Research Questions

The assumption I had at the beginning of the analysis were:
- Vaccination impacted COVID-19 cases, hospitalizations, deaths by lowering them
- COVID-19 increased the unemployment rate in Fairfield County, CT

This paper will aim at investigating my assumptions statistically and show whether they hold or not. The research questions I aim to answer in the subsequent sections are:
1. What is the impact of vaccination on:
   a. Confirmed COVID-19 cases?
   b. Hospitalization due to COVID-19?
   c. Deaths due to COVID-19?
2. Is there a correlation between confirmed COVID-19 cases and the rate of unemployment in Fairfield County, CT?

The period I will be conducting my analysis is from February 1, 2020, through October 15, 2021 in Fairfield County, CT.

3. Methodology

3.1: Dataset
- COVID-19 Confirmed Cases
  - This .csv file contains the number of daily confirmed cases by county on daily basis and it is updated daily by John Hopkins University on their Github Repository. For this analysis, I used the CONVENIENT_us_confirmed_cases.csv instead of the raw version because it is easier to use. The columns I used from this CSV file:

Column name	Description
Province_State	This column contains dates
Connecticut.1	This column contains the number of confirmed cases

- COVID-19 Deaths
  - This .csv file is downloaded from Kaggle and it is updated daily by John Hopkins University on their Github Repository, the dataset has the number of deaths related to COVID-19 and it's given at the county level on daily basis. For this analysis, I used the CONVENIENT_us_deaths.csv instead of the raw version because it is easier to use. The columns I used from this CSV file:

Column name	Description

Province_State	This column contains dates
Connecticut.1	This column contains the number of deaths

- COVID-19 Hospitalization
  - This dataset is obtained from the Connecticut Open Data that is provided by the Connecticut State Agencies. The data is provided at the county level. The columns I used for this analysis are:

Column name	**Description**
Date updated	Date dataset was last updated
County Code	FIPS of my assigned county and I used to filter my data
Hospitalized cases	Number of patients currently hospitalized with COVID-19

- Vaccination
  - The vaccination data were obtained from the CDC and it is organized at the level county. The data represents all vaccination partners. The columns I used for this analysis are:

Column name	**Description**
Date	Date data are reported on CDC COVID Data Tracker
FIPS	Federal Information Processing Standard State Code. I used this column to filter out my assigned county
Series_Complete_Yes	Total number of people who are fully vaccinated (have the second dose of a two-dose vaccine or one dose of a single-dose vaccine) based on the jurisdiction and county where the recipient lives

  - The Vaccination data are from December 07, 2020, to October 15, 2021.

- Unemployment Rate in Fairfield County

○ I obtained the dataset from FRED Economic Data. The data is obtained from the US Bureau of Labor Statistics. The data is conveniently organized and I used the filters on the website to get only the information needed for this analysis:

Column name	Description
Date	Date of unemployment rate reported on monthly bases
CTFAIR1URN	Rate of unemployment in Fairfield County

○ I used two versions of the unemployment rate dataset, the rates are shown monthly and the last rate available is for October 2021:
  ■ Version 1: generated over the period between February 1, 2020, to October 1, 2021
  ■ Version 2: generated over the period between January 1, 2018, to October 1, 2021. I used this version to see how the unemployment rate changed pre-COVID and post-COVID

● Mask Mandates
  ○ This dataset is obtained from the CDC and it provides information on public mask mandates by county on daily basis. The columns I used for this are:

Column name	Description
Date	Date of unemployment rate reported on monthly bases
FIPS_County	U.S. county FIPS codes. I used this column to filter out my assigned county
date	Daily date in the dataset
Face_Masks_Required_in_Public	A requirement for individuals operating in a personal capacity to wear masks 1) anywhere outside their homes or 2) both in retail businesses and in restaurants/food establishments.

● Mask Use
  ○ This dataset is published by the New York Times and it estimates the prevalence of mask-wearing in counties in the US. I used all the columns in this dataset:

Column name	Description
COUNTYFP	The county FIPS code

NEVER	The estimated share of people in this county who would say never in response to the question "How often do you wear a mask in public when you expect to be within six feet of another person?"
RARELY	The estimated share of people in this county who would say rarely
SOMETIMES	The estimated share of people in this county who would say sometimes
FREQUENTLY	The estimated share of people in this county who would say frequently
ALWAYS	The estimated share of people in this county who would say always

3.2: Analysis

3.2.1: Tools, methods, and libraries

I used visualizations to see COVID-19 trends and Pearson correlation coefficient to understand the relationship of vaccination, rate of infection, and unemployment in Fairfield County, Connecticut.

The following are the tools and libraries I used to conduct the analysis for this project:
- Jupyter-notebook with Python. I used the following libraries to analyze the datasets:
  - Pandas
  - NumPy
  - Matplotlib
  - Seaborn
- RStudio: I used R to calculate Pearson's correlation coefficient and graph it. The library I used here:
  - ggpubr

3.2.2: Data Preprocessing

To begin my analysis, I started by cleaning my data. The county I was assigned for this project is Fairfield County, CT (FIPS=9001), I used Pandas to read the CSVs into dataframes, drop unneeded columns, convert date columns to datetime using pd.to_datetime(<column_name>), rename columns, and filter data.

After the data was cleaned, I merged the cleaned columns and calculated the daily rate of infection and the change in the daily rate of infection using the following formulas:

Population at risk = Fairfield population - rolling average of confirmed cases
Daily rate of infection = Rolling average of confirmed cases / population at risk
Change in the daily rate of infection = (difference in the daily rate of infection) * 100

All the data was added to a final dataframe that I used to perform some data exploration and find the correlation between the columns

3.2.2: Exploratory Data Analysis

I did an initial analysis and created visualizations using matplotlib to detect COVID-19 trends from 2020-02-01 to 2021-10-15. Below is what I found during

- **Estimates of mask use, voluntary response to the question "How often do you wear a mask in public when you expect to be within six feet of another person?" in Fairfield County, CT:**
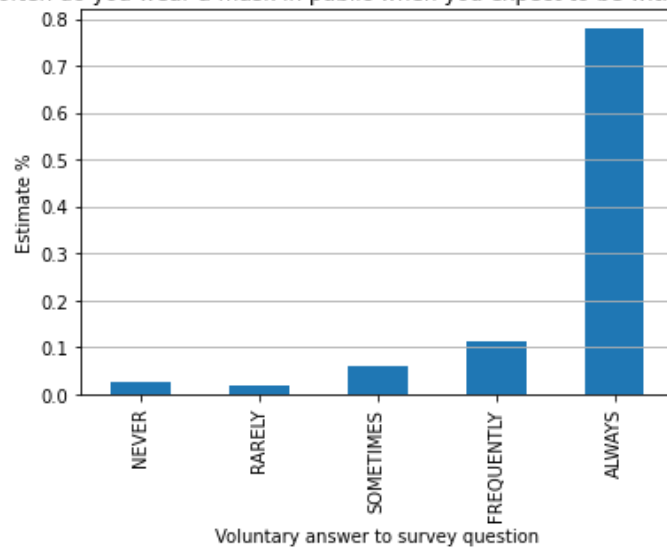
Figure 1: mask use in Fairfield County

The bar chart shows that 80% of people who responded to the survey question always wore their masks in public areas.

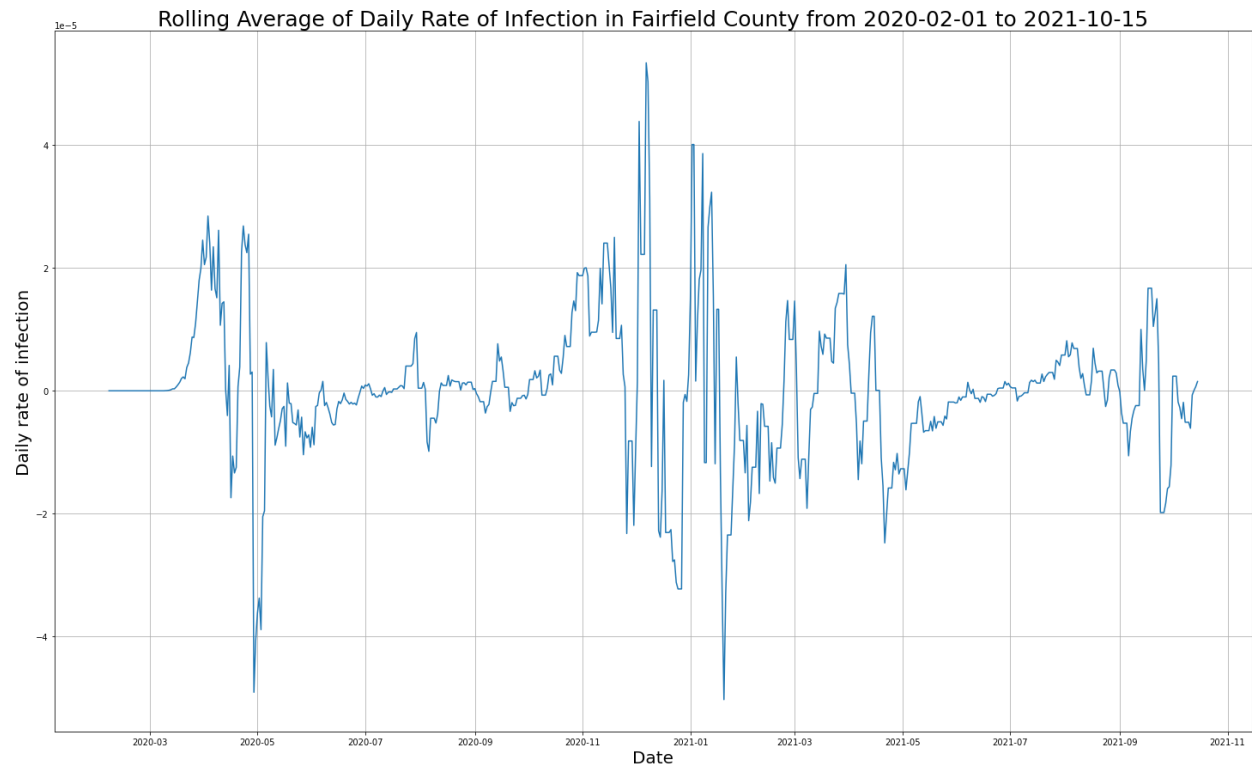- **The trend in daily rate of COVID-19 infections in Fairfield County:**

Figure 2: daily rate of infection

The graph shows a spike in COVID-19 cases during the beginning of the pandemic and another spike in cases during the holiday season in 2020, which is consistent with the first wave and second wave of the COVID-19 cases.

● **Trend in confirmed cases as mask mandate policies change in Fairfield County:**
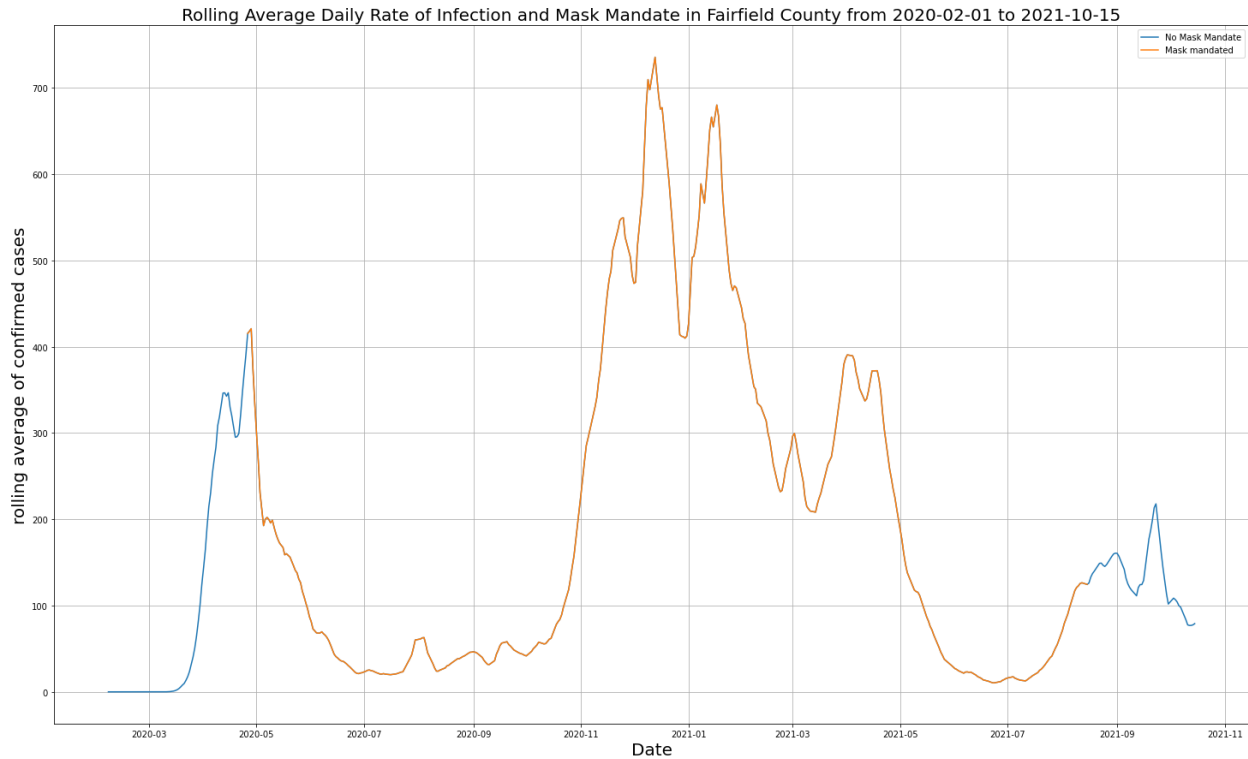
Figure 3: daily rate of infection as mask mandate policies change

The graph shows a drop in COVID cases after masks were mandated around May 2020 and a slight rise in cases after the mandate was lifted. This shows the effectiveness of masks in combating COVID-19.
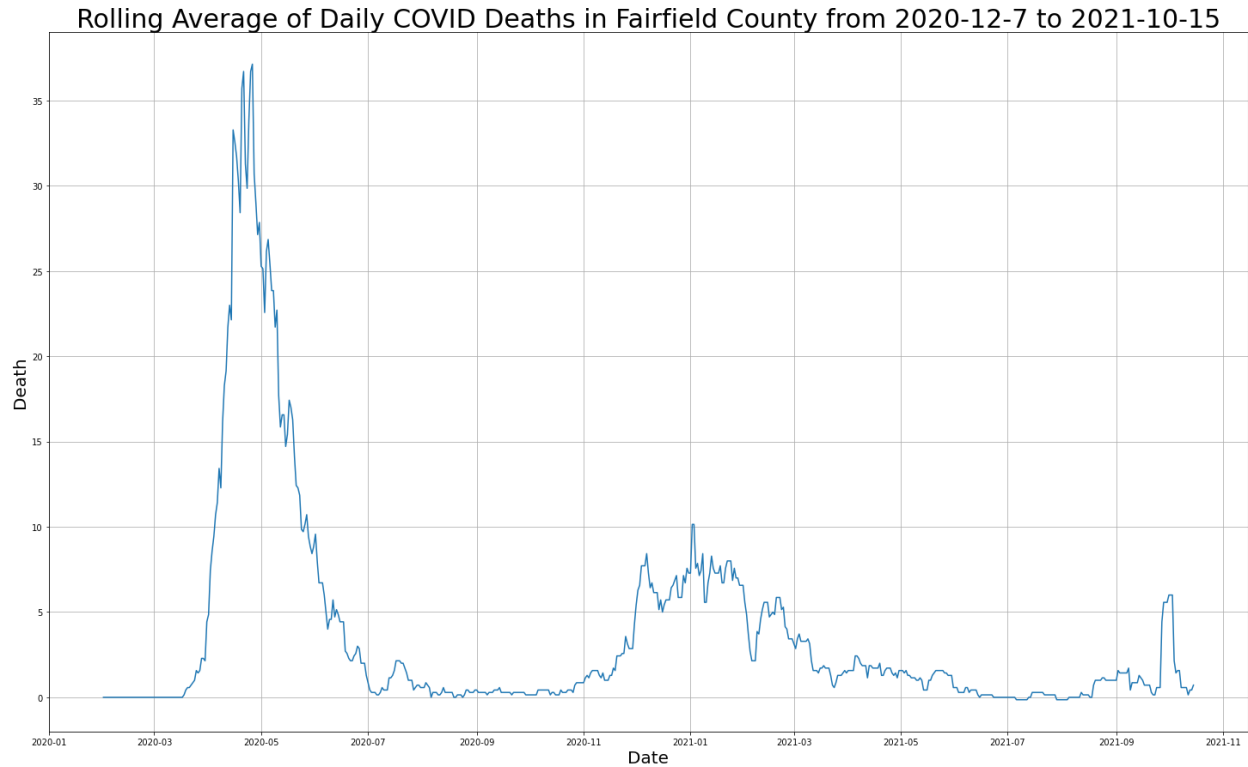
- **The trend in COVID-19 deaths in Fairfield County:**

Figure 4: Covid-19 related deaths

The figure a significant spike in COVID-19 related deaths at the beginning of the pandemic. The number of death dropped but rose again during the 2020 holiday season and again in late 2021, which coincides with the rise Delta variant of COVID-19.

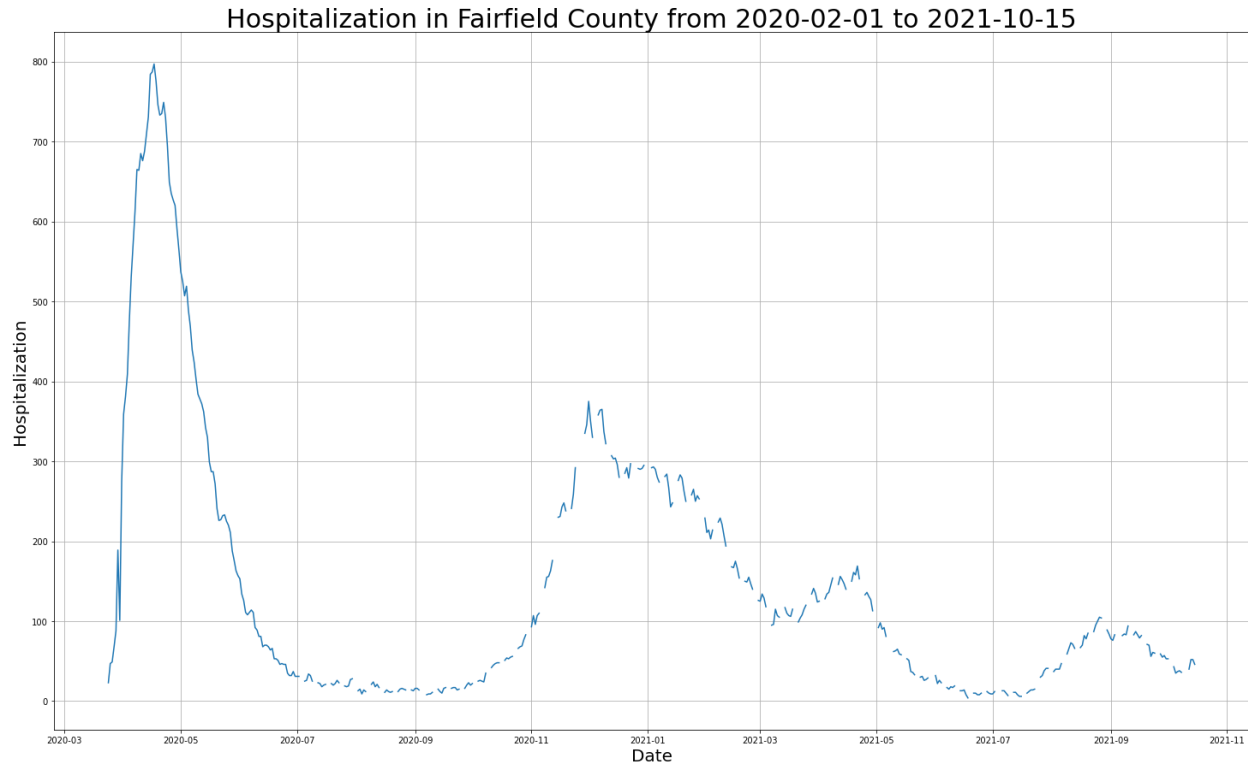- **The trend in COVID-19 hospitalization in Fairfield County:**

Figure 5: Covid-19 related hospitalizations

The graph shows the hospitalizations rose at the beginning of the pandemic and again during the 2020 holiday season, the significant drop in COVID-19 related hospitalization could be due to mask mandates.

● **The trend in being fully vaccinated by either receiving two does of Pfizer and Moderna vaccines or one does of Johnson&Johnson vaccine:**
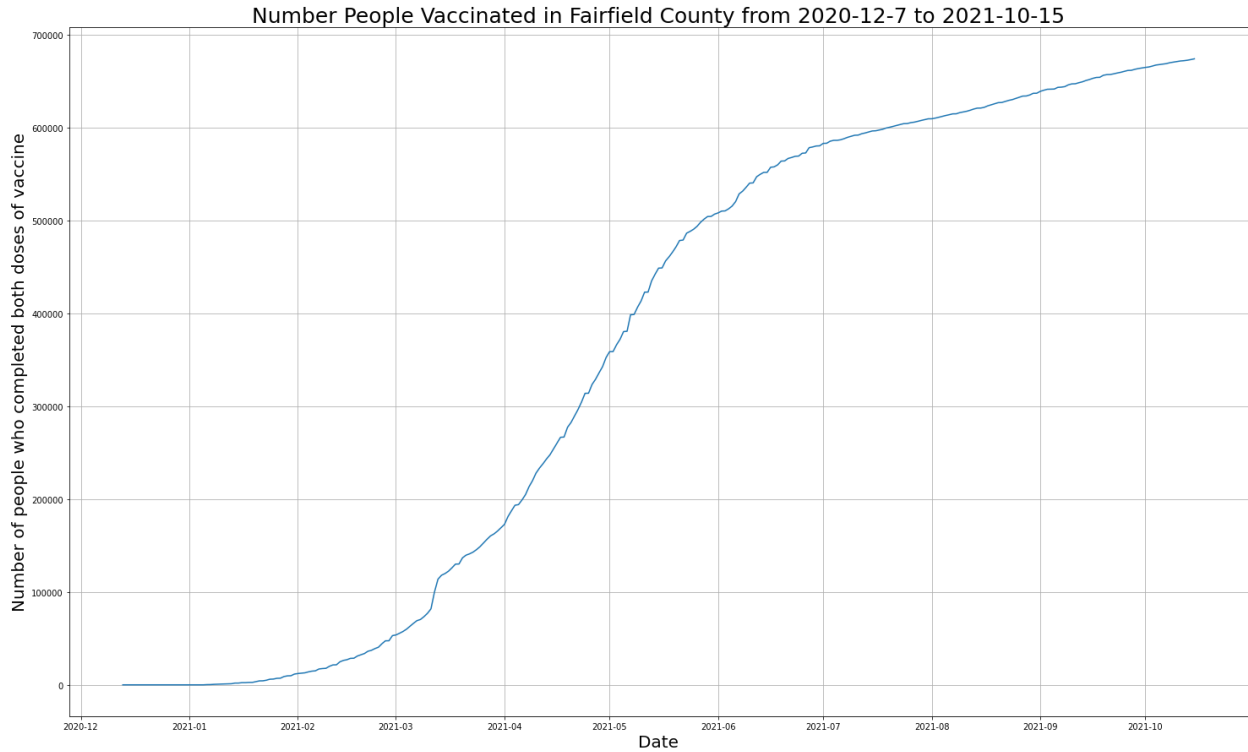
Figure 6: Vaccination trend

The graph shows a steady increase in people who are fully vaccinated in Fairfield County. As of October 15, 2021, 674134.0 people are fully vaccinated which is about 70.4% of Fairfield County residents.

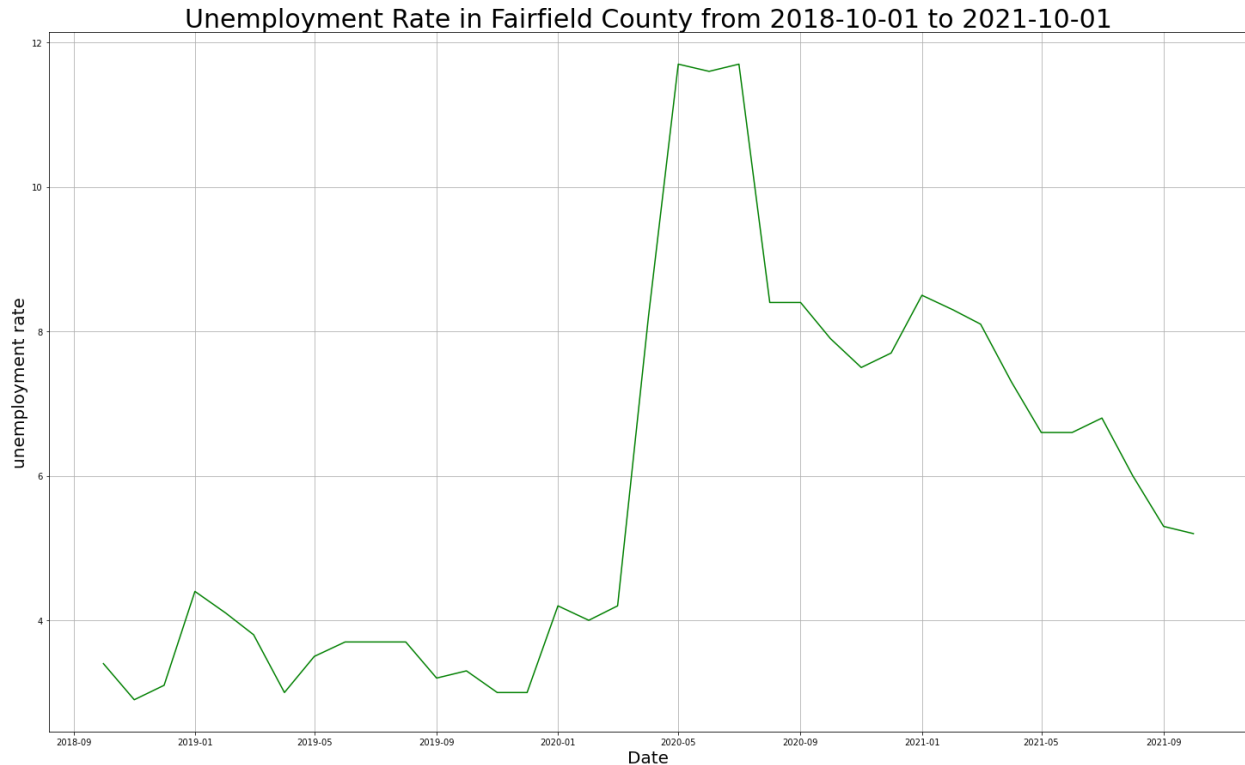- **The Unemployment Rate in Fairfield County from 2018-01-01 to 2021-10-01**

Figure 7: Unemployment Trend

For this graph, I used unemployment rates from 2018 to investigate whether the trend changed when the pandemic happened. The graph shows a clear spike in the unemployment rate between February 2020 and May 2020. The unemployment rate seems to drop at a steady rate since the beginning of 2021.

4. Findings

After completing the initial analysis shown in section 3.2.2, I performed the rest of my research in RStudio using the library 'ggpubr' to show statistically if there are correlations in my datasets. To find the correlations, I used Pearson's Correlation Coefficient that measures if there is a linear correlation between two sets of data.

Pearson's coefficient value and sign indicate the strength and direction of the correlation:
- Strength: Pearson coefficient can range in value from -1 to +1. The larger the coefficient, the stronger the relationship between the two sets of data. 1 is a perfect linear correlation and 0 means there's no correlation between the two sets of data.
- Direction: Pearson's coefficient sign indicates how the two seats are related. A positive coefficient indicates that both sets of data increase and decrease together. A negative coefficient indicates that if one set of data increases, the other decreases, and vice versa.

To show the correlation in my dataset, I created a scatter plot using 'ggscatter' function and drew a correlation line to show the relationship between the two variables. The findings are shown in the following subsections:

4.1: Correlation between vaccination and confirmed COVID-19 cases:
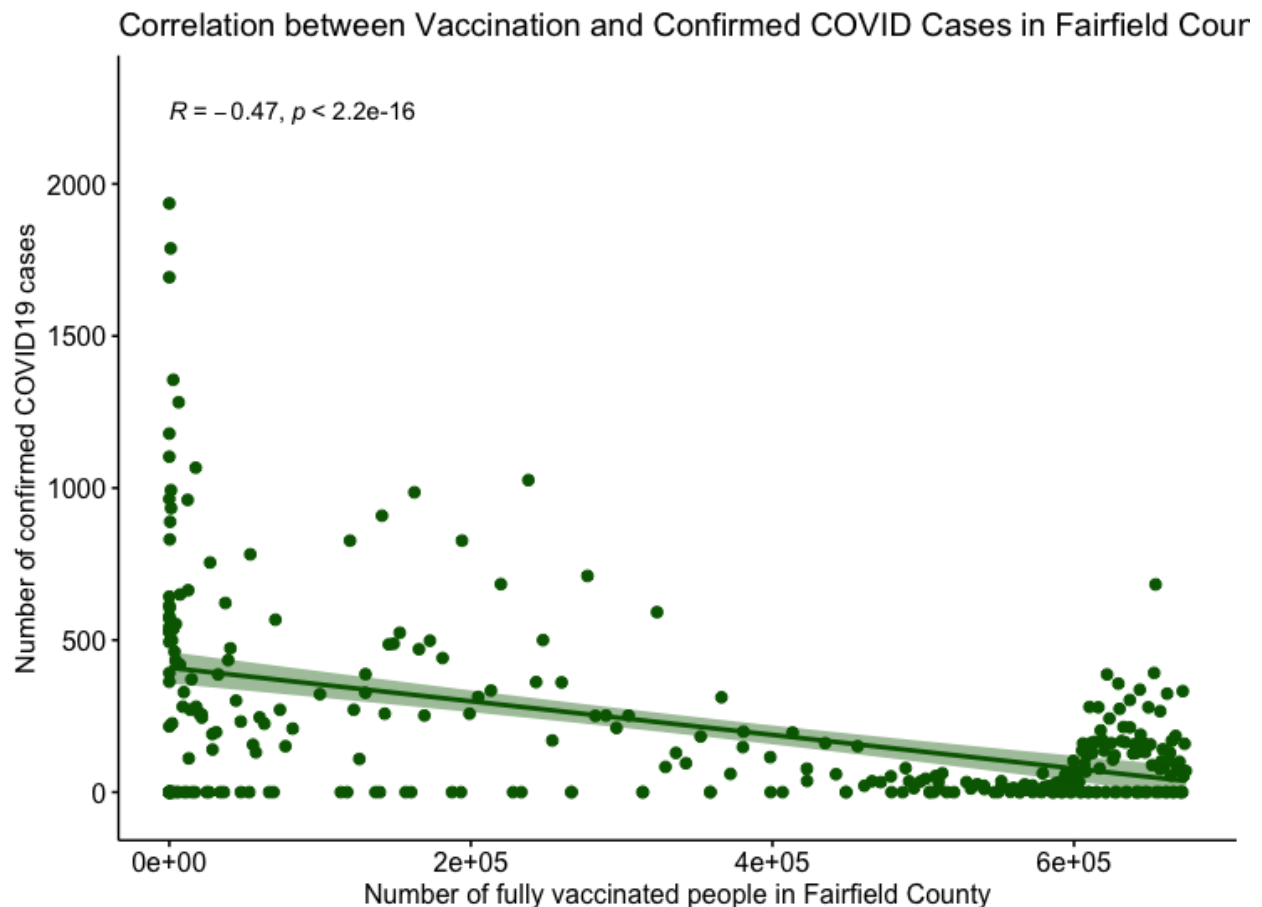

Figure 8: correlation between vaccination and number of covid cases

The figure above shows a negative correlation number of people who are fully vaccinated and the rate of infection. The correlation coefficient is -0.47 which means there is a correlation between the COVID-19 cases and vaccination. The negative sign indicates being fully vaccinated contributes to the drop of COVID-19 cases.

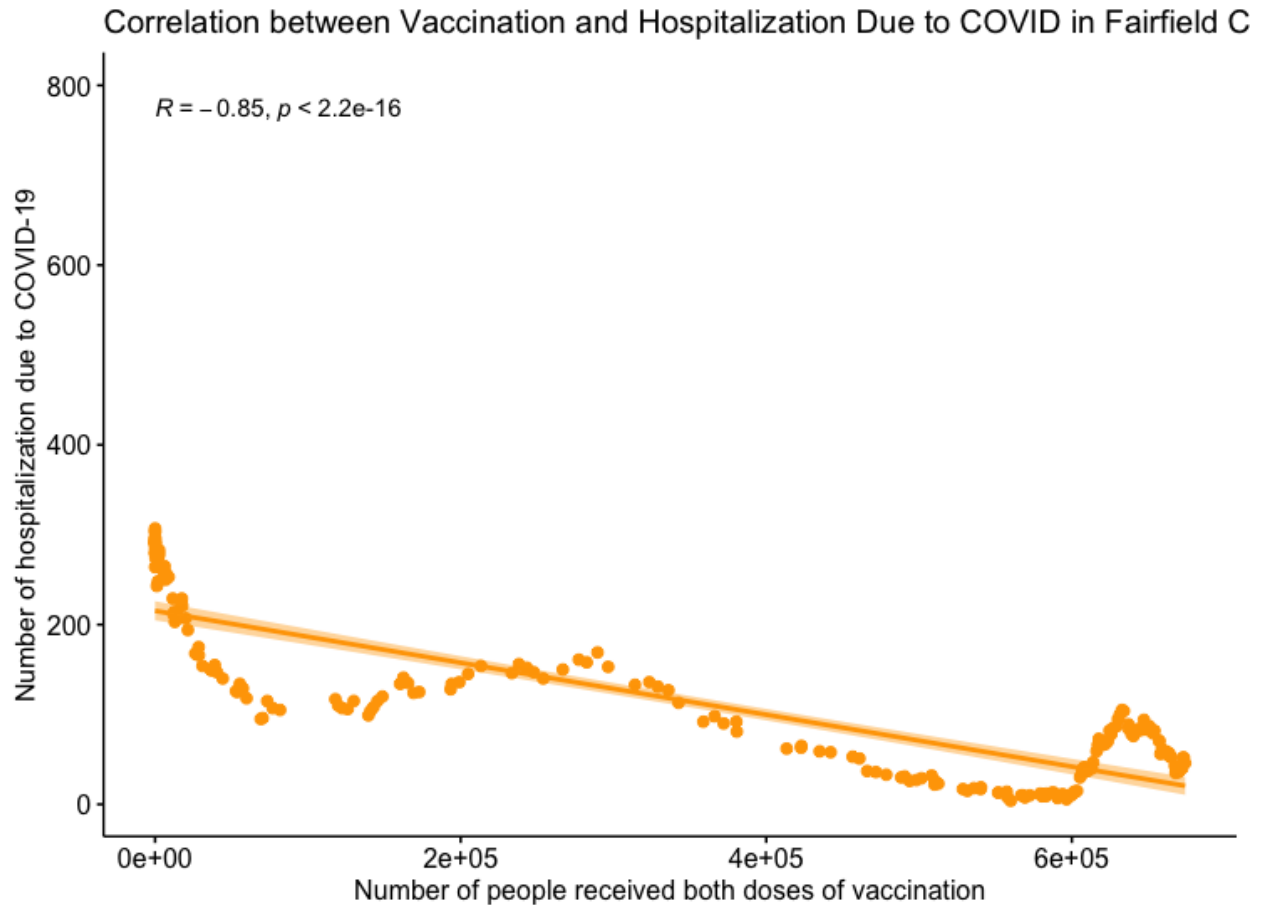## 4.2: Correlation between vaccination and hospitalization

Figure 9: correlation between vaccination and hospitalizations

The graph shows a strong correlation between vaccination and hospitalization due to COVID-19. The correlation coefficient is -0.85, the negative sign indicates that as the vaccination rate goes up, the number of hospitalizations goes down.

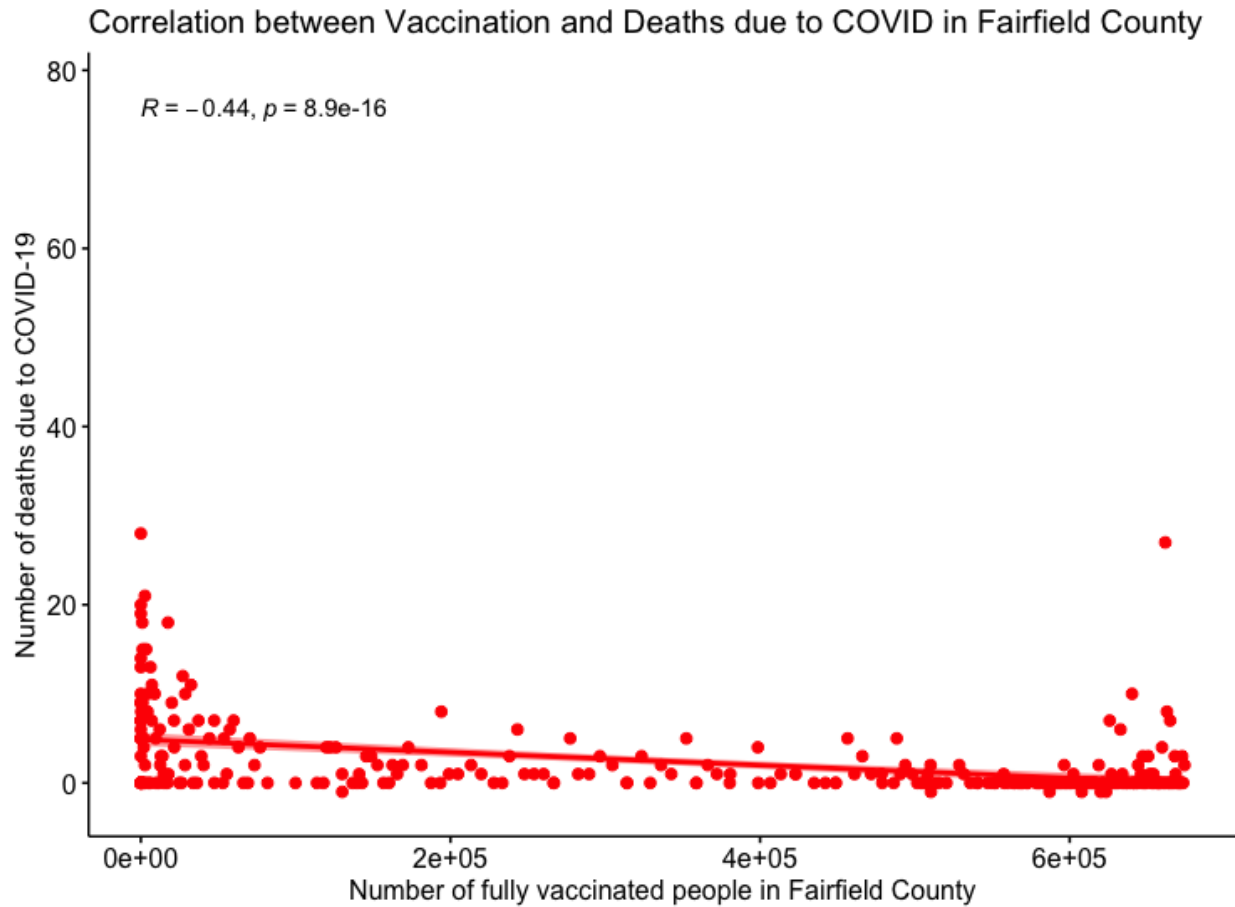## 4.3: Correlation between vaccination and death

Figure 10: correlation between vaccination and deaths

The graph shows that there is a correlation between vaccination and death due to COVID-19. The negative sign indicates that as the number of fully vaccinated people increases, the number of deaths related to COVID-19 decreases.

4.4: Correlation between COVID-19 cases and unemployment

Correlation between COVID Cases and Unemployment Rate in Fairfield County
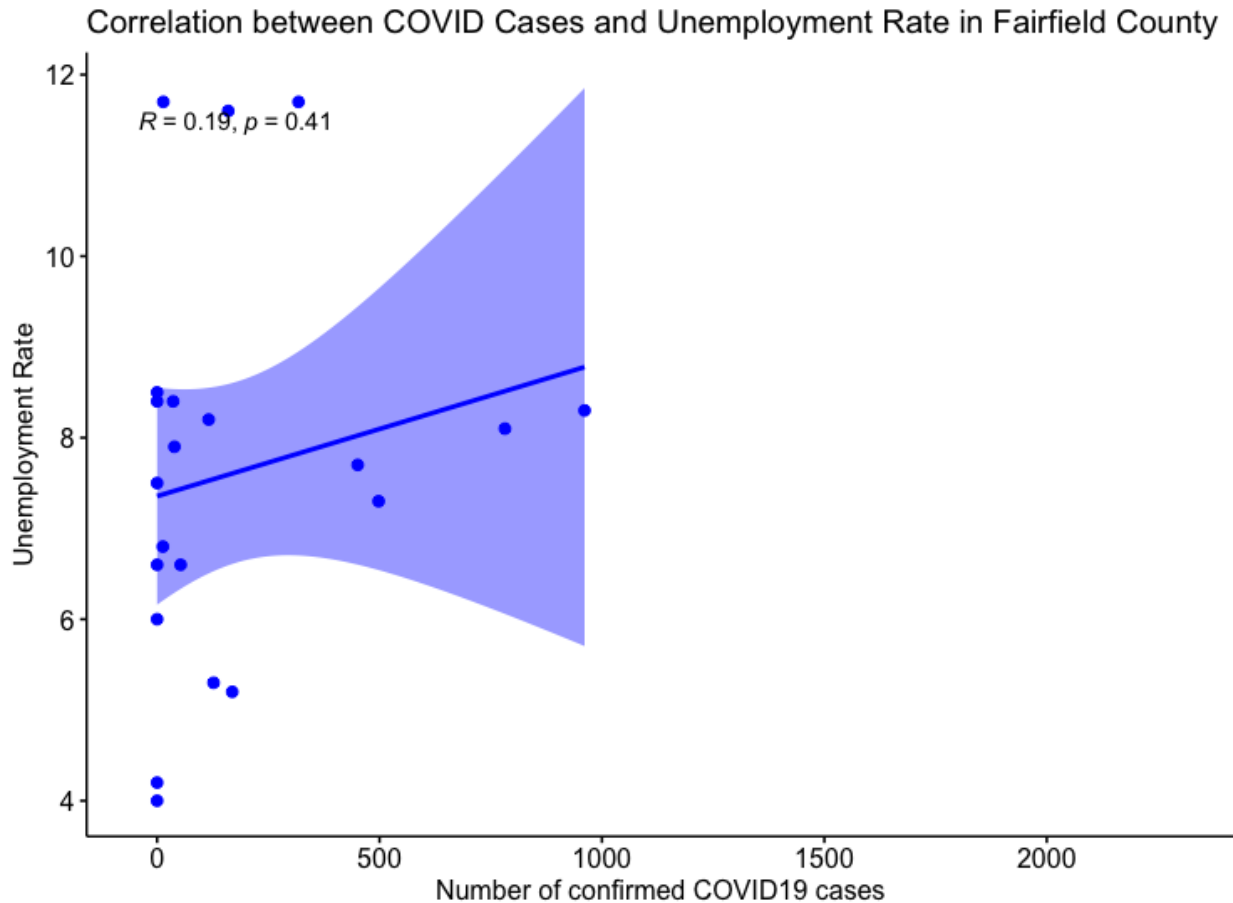
$R = 0.19, p = 0.41$

Figure 11: correlation between covid cases and unemployment

The final graph shows that there is a correlation between the increase in COVID-19 cases and unemployment rates in Fairfield County. <u>The positive correlation coefficient indicates that unemployment rates increased with the rise of COVID-19 cases.</u>

5. Discussion/Implications

Originally, I wanted to investigate the impact COVID-19 had on Community Mobility for this project but unfortunately, I had some difficulty coming up with results that made sense to me, so I pivoted to analyzing other aspects of COVID-19 in my assigned county.

The results from my analysis were not surprising and matched the assumptions I had when I started working on the final project. The visualizations and correlation plots show that being fully vaccinated and wearing masks in public are definitely crucial in combating the COVID-19 pandemic and slowing down the spread.

In section 3, the visualizations showed the same trend in hospitalizations and deaths related to COVID-19; as the number of confirmed cases increases, the number of hospitalizations and

deaths increases as shown in Figure 12. This correlation supports the trends observed in Figure 4 and Figure 5.
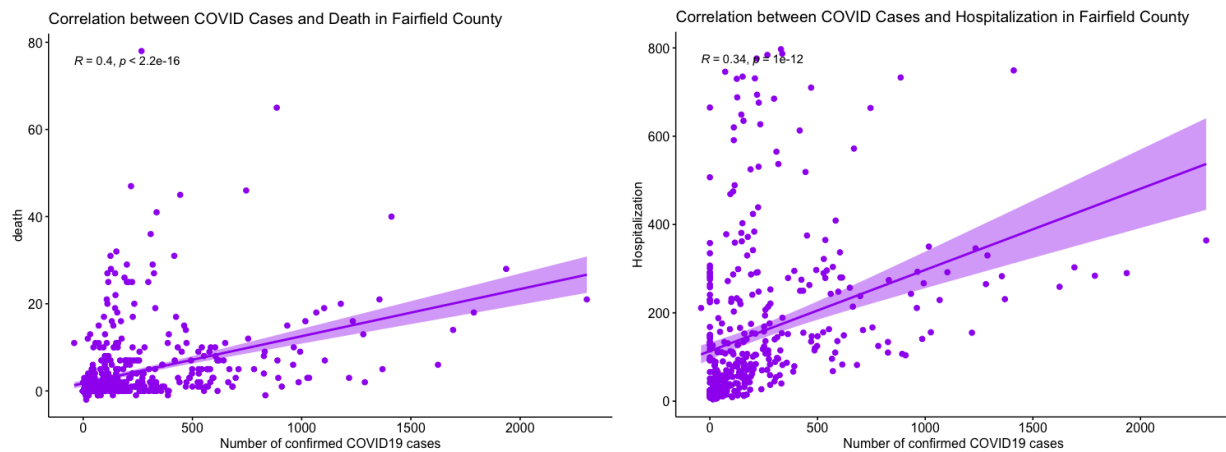

Figure 12: correlation between covid cases and death (left) and correlation between covid cases and hospitalization (right)

From a human-centered perspective, the main concern with the datasets I used in this analysis was the privacy of the participants. The identity of the participants is preserved and it is very difficult to try to identify any of the participants. This analysis also shows the importance of participating in combating the COVID-19 pandemic by getting vaccinated, wearing masks in public, and adhering to any movement restrictions that are put in place to slow down the spread of COVID-19.

The human-centered principles I made an effort to follow while conducting this study are:
- Reproducibility: I documented my analysis process in the Jupyter notebook and R markdown by adding comments and using variable names that would make it easy to follow what I did. I will also list the libraries, tools, and methods I used in the README file on the final project Github repository
- Licensing my Github repository: I will use MIT license for my Github repository
- Including any licensing, terms of use, copyright information of the datasets I used: please check sections 8 and 9 for reference and data sources
- Checking for any fairness, bias, and privacy issues in the datasets I used: Privacy was the most applicable issue that I might have faced in my analysis. The datasets I used didn't have any identifying personal information that could harm the participants

6. Limitation
Below I show the list of limitations I had while conducting this analysis and future work might be needed to verify my findings:
- Missing data points and blanks in the datasets
- Cannot check if the data is representative of the entire population in Fairfield County, CT
- Data was unavailable for the month of October in the unemployment rate dataset

- Mask use survey was collected by voluntary participation which might not be representative of the entire county
- Vaccination booster data is not available
- COVID-19 variant data not available
- Data might not have been prepared thoroughly by checking assumptions for Pearson's Coefficient Correlation test
- The rolling average was not applied to all the dataset
- The scales of the datasets might vary and need to be normalized for more in-depth analysis

7. Conclusion

For the final project, I was assigned to Fairfield County in Connecticut. The research questions I investigated in this analysis as well as my findings/answers to them are listed below:

1. **Is there a correlation between getting fully vaccinated and the number of daily confirmed COVID-19 cases?**
   Yes, there was a correlation between vaccination and COVID-19 cases. The increase in vaccination leads to the decrease of COVID-19 cases and lowers the rate of infection.
2. **Is there a correlation between vaccination and COVID-19 related hospitalizations?**
   I found that there was a linear correlation between the two; as the number of people who are fully vaccinated increases, the number of COVID-19 related hospitalizations decreases.
3. **Is there a correlation between vaccination and COVID-19 related deaths?**
   Yes, my analysis showed that increasing vaccination decreases COVID-19 deaths
4. **Is there a correlation between COVID-19 cases and the unemployment rate in Fairfield County?**
   Yes, there is a correlation between unemployment and COVID-19 cases. The unemployment rate spiked significantly at the beginning of the pandemic and stayed at higher rates than it was in 2018 and 2019

My study showed the impact of COVID-19 vaccines had on decreasing the rate of COVID-19 infections, deaths and hospitalization. It also showed how COVID-19 contributed to big spikes in unemployment in Fairfield County, Connecticut.

8. Reference

- ❖ Self WH, Tenforde MW, Rhoads JP, et al. Comparative Effectiveness of Moderna, Pfizer-BioNTech, and Janssen (Johnson & Johnson) Vaccines in Preventing COVID-19 Hospitalizations Among Adults Without Immunocompromising Conditions — United States, March–August 2021. MMWR Morb Mortal Wkly Rep 2021;70:1337–1343. DOI: http://dx.doi.org/10.15585/mmwr.mm7038e1external icon.

- ❖ Polack, Fernando P., et al. "Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine." New England Journal of Medicine (2020).
- ❖ Pearson's Correlation Coefficient
- ❖ Unemployment in these 25 CT towns doubled — or worse — during the coronavirus pandemic
- ❖ Fairfield COVID-19 Page

9. Data Source

The following is a list of the datasets that are used in this analysis and their citation, licenses, or terms of service:

- COVID-19 Confirmed Cases
  - License: Public Domain U.S. Government
- COVID-19 Deaths
  - License: Attribution 4.0 International (CC BY 4.0)
- COVID-19 Hospitalization
  - Terms of Service
  - Data provided by the Department of Public Health
- COVID-19 Vaccination
  - License: Public Domain U.S. Government
- Unemployment Rate in Fairfield County, Connecticut
  - Fred Economic Data: Terms of Service
  - US Bureau of Labor Statistics: Terms of Service
- Mask use
  - License
- Mask Mandate
  - Citation: CDC, COVID-19 Community Intervention & Critical Populations Task Force, Monitoring & Evaluation Team, Mitigation Policy Analysis Unit, the CDC, Center for State, Tribal, Local, and Territorial Support, Public Health Law Program, and Max Gakh, Assistant Professor, School of Public Health, University of Nevada, Las Vegas, "U.S. State and Territorial Orders Requiring Masks in Public," (August 15, 2021).