Coronavirus EDA Executive Summary

Data Science I (STAT 301-1)

Derex Wangmang

Executive Summary

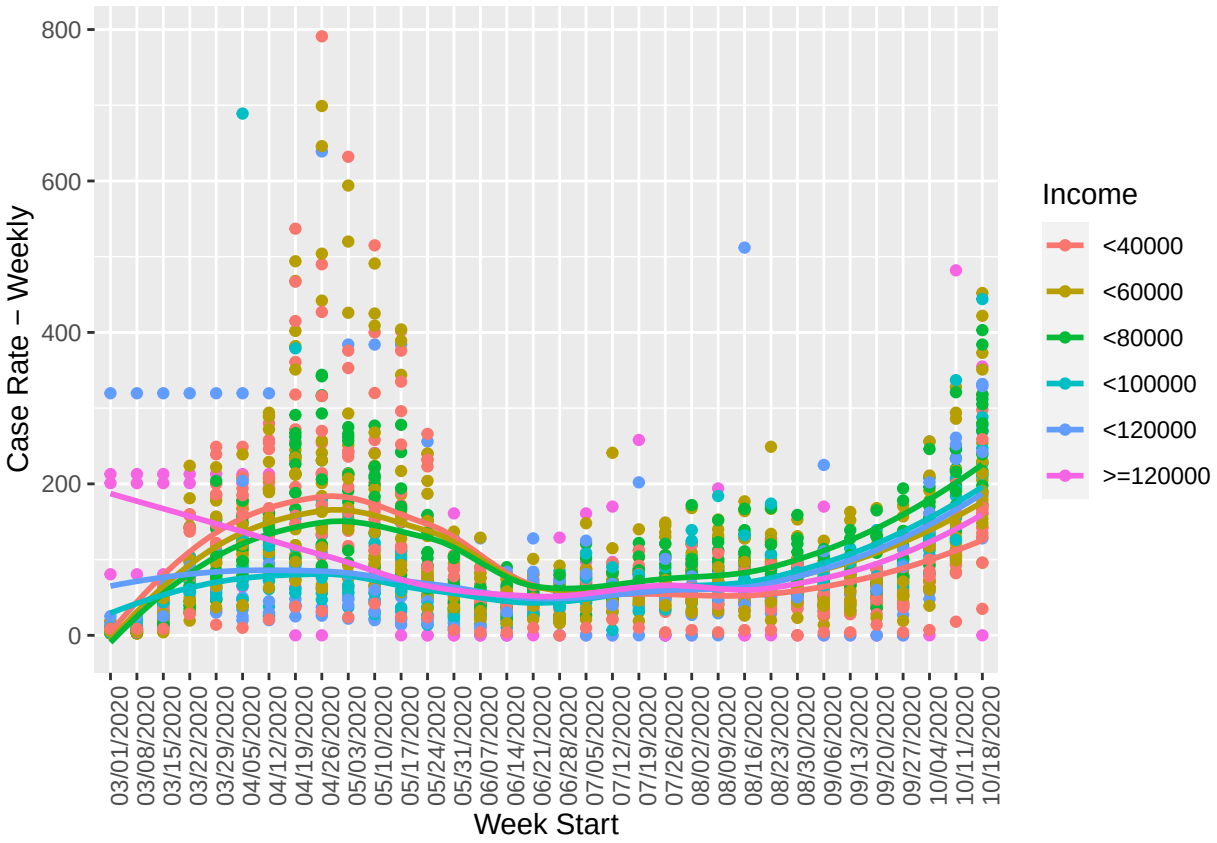
My Exploratory Data Analysis focused on visualizing how the coronavirus exacerbated existing inequalities. As part of my project, I retrieved datasets that included cases and deaths in the Chicagoland area, diabetes rates, asthma rates, and income levels. Because my original dataset about COVID-19 cases was separated by ZIP code, I used ZIP code as the unique key across datasets. I explored the following questions:

- What is the overall trend of coronavirus cases and deaths in the Chicagoland area?
- How does median income relate to coronavirus case and death rates?
- How do medical conditions, specifically diabetes and asthma rates, relate to the rates of coronavirus cases and deaths?

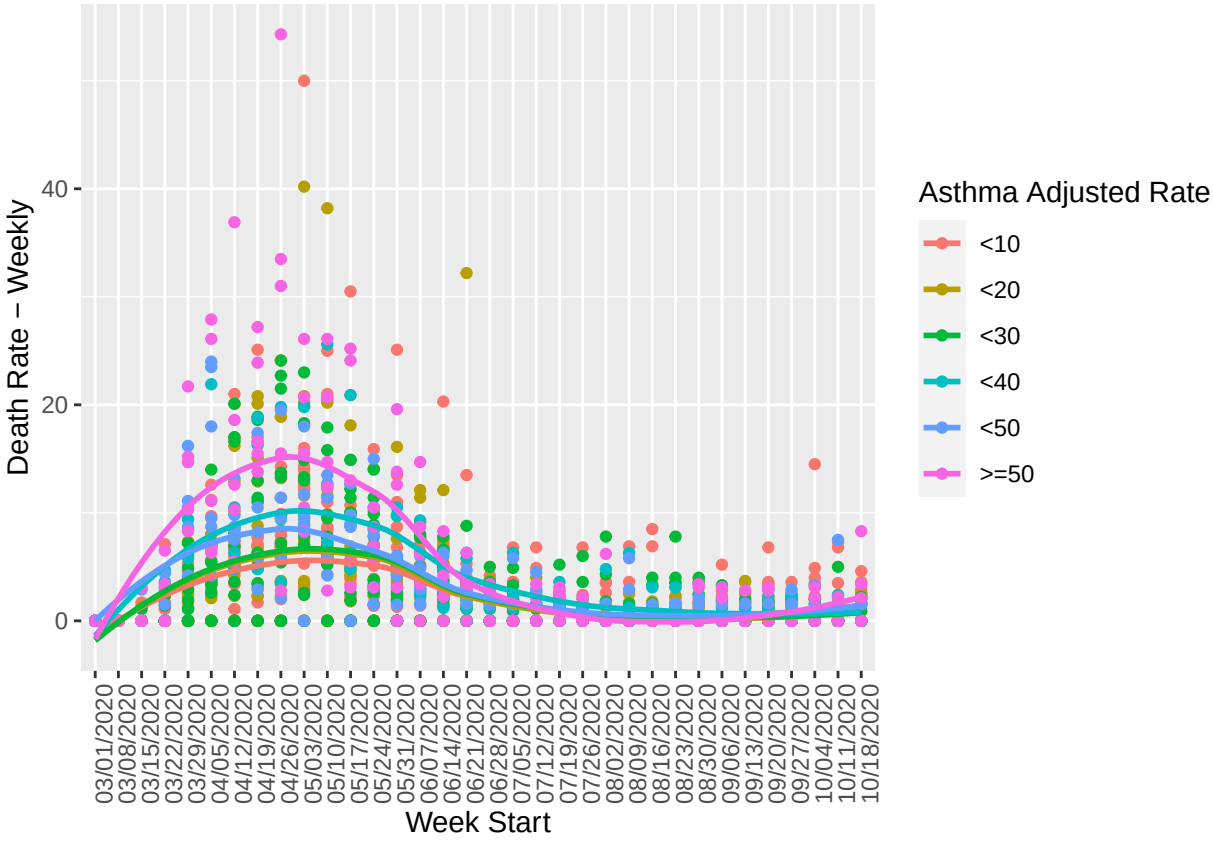
I processed my data by filtering out all observations with ZIP codes outside the ZIP codes from the case data. I then removed rows with missing or incomplete data. Additionally, I disaggregated data that initially represented multiple ZIP codes. I also leveraged `tidycensus` to retrieve data about the median income from the US Census Bureau.

Combing these data with the original case data, I fitted trends whose lines were separated by the previous mentioned factors. With these, there were two significant patterns.

1. The higher the income, the lower the rates. When examining the case and death rates across ZIP codes with varying incomes, I aggregated the data by income levels of 20000. The highest peaks were found in ZIP codes with the lowest incomes, and the lowest peaks in ZIP codes with the highest incomes. This may be a result of white collar work, where workers within those ZIP codes successfully transitioned to a work from environment. Their reduced contact with others may have led to decreased cases. Another potential reason is that those from the lowest income ZIP codes use public transportation to travel, which may expose them to more cases, whereas those from the wealthy population may have their own method of transportation.



2. ZIP codes with higher asthma rates had higher death rates than the death rates for ZIP codes with lower asthma rates. While the trends are the same, with a peak around 04/26 and a constant rate of near 0 afterwards, the death rates were significantly higher than that of the ZIP codes with the lowest asthma rates. Since the coronavirus affects the respiratory system, and asthma is a medical condition that affects the lungs, I expected the death rate to be higher for the ZIP code with higher rates of asthma. This visualization below quantifies my assumption, illustrating how the populations with higher asthma rates have larger peaks that extend over longer periods of time.


Death Rate – Weekly
40
20
0
03/01/2020
03/08/2020
03/15/2020
03/22/2020
03/29/2020
04/05/2020
04/12/2020
04/19/2020
04/26/2020
05/03/2020
05/10/2020
05/17/2020
05/24/2020
05/31/2020
06/07/2020
06/14/2020
06/21/2020
06/28/2020
07/05/2020
07/12/2020
07/19/2020
07/26/2020
08/02/2020
08/09/2020
08/16/2020
08/23/2020
08/30/2020
09/06/2020
09/13/2020
09/20/2020
09/27/2020
10/04/2020
10/11/2020
10/18/2020
Week Start
Asthma Adjusted Rate
<10
<20
<30
<40
<50
>=50