# Math for DS Project
# Ryan L'Abbate & Antonio Singh

**Country of Choice: Canada**

Our World In Data:

https://ourworldindata.org/coronavirus-data-explorer

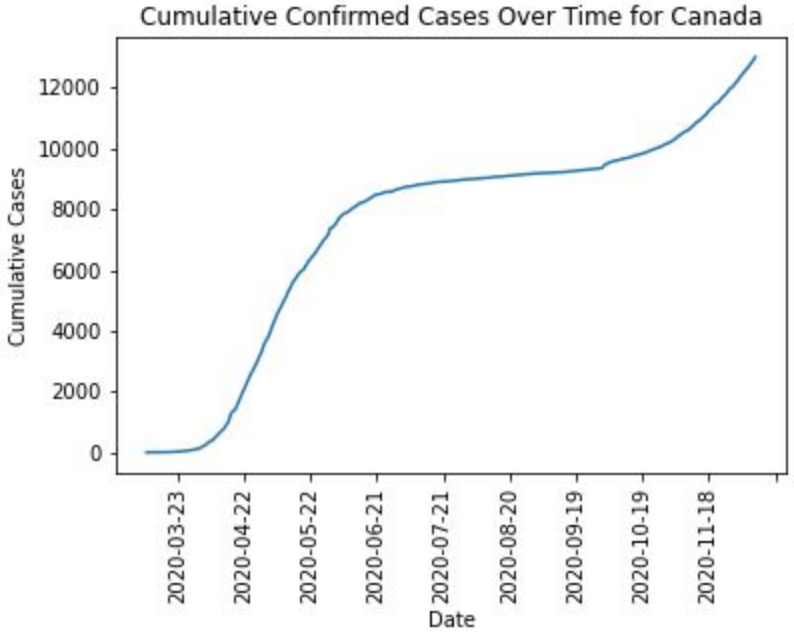
Mobility Data:

https://www.kaggle.com/therealcyberlord/coronavirus-covid-19-visualization-prediction

UNDP HDI Data:

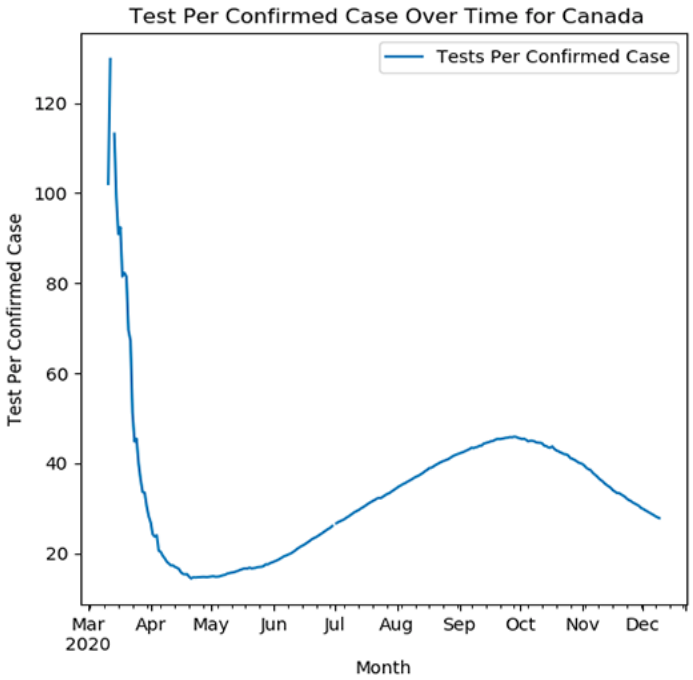
http://hdr.undp.org/en/data

**1. Cumulative Confirmed Cases Over Time for Country X**




Interpretation: Covid had two major rises in Canada. Covid spread rapidly in Canada beginning in early April 2020 and continued to rise rapidly before tapering off around the end of June 2020. Cases stayed relatively stable until October where they began to rapidly increase again. The second rise looks more exponential while the first rise looks more linear.

**2. Test per Confirmed Case Over Time for Country X**

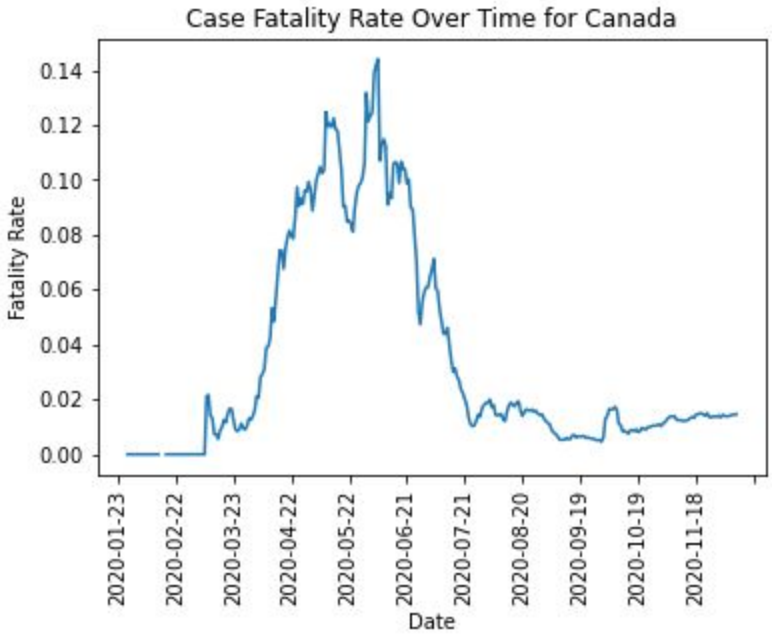



-To obtain the Test Per Confirmed Case Over Time for Canada, the Our World In Data csv was read in and stored. Only the rows with the condition of column location equaling Canada were then obtained as a data frame since we only needed the data for Canada. Test Per Confirmed Case Over Time was then calculated by dividing the number of tests by the number of cases. The resulting values were then visualized. Since we were looking at the change over time, the dates were converted to datetime objects so we can just view them as overall months starting from March to December.

-Based on the visualization, we see a large number of tests were done but the confirmed cases were on a decline as the spread of coronavirus began from March to June but a steady increase started happening from July. We found it strange that with testing being so high that it resulted in such smaller positivity rates initially but we figured that it could potentially be due to either test inaccuracies during the initial run of testing starting such as false positives or not much knowledge of the virus at the time when it came to testing compared to the present testing methods. With the testing, since the numbers of tests were very low in the beginning of the Spring, only the most severe cases of people were being tested so we figured the numbers would be much higher but obviously this was not the case initially. As the steady increase began, it hit its peak in the fall especially in September to November with the highest peak being in October. We can conclude that as testing began to happen more regularly from the Summer to Fall, the coronavirus cases were beginning to be confirmed on a more regular basis in Canada. We can

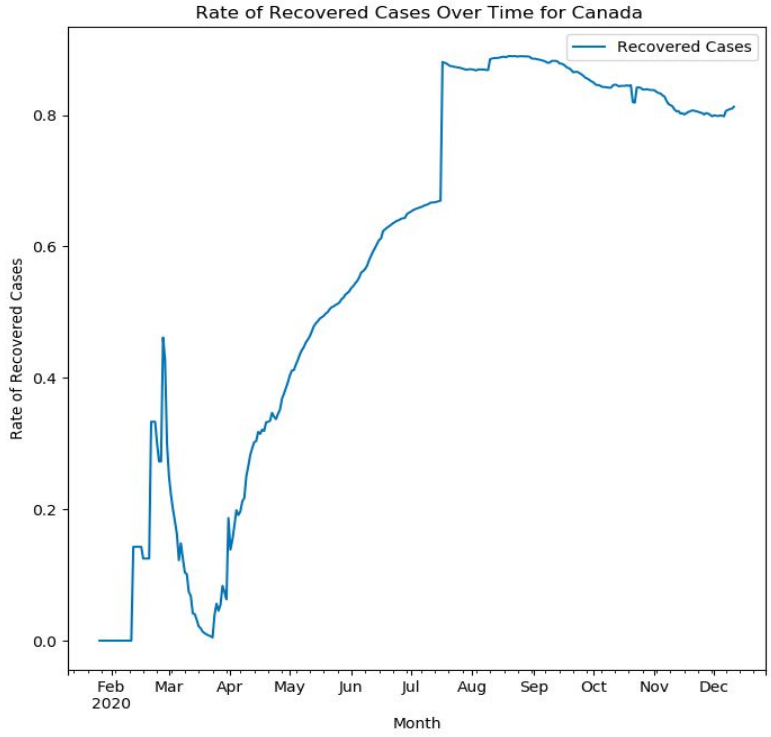
see that even in the present month of December with coronavirus spiking in areas again that a steady decrease is happening in December.

### 3. Case Fatality Rate Over Time for Country X




Interpretation: Fatality rates increase rapidly in early April, which is also when the first major rise in Covid occurred (see graph 1). The fatality rate decreases in late June, which is when the first rise ended. Notably, no significant rise occurred later in the year with the second rise. This may be due to people taking more precautions as the pandemic went on.

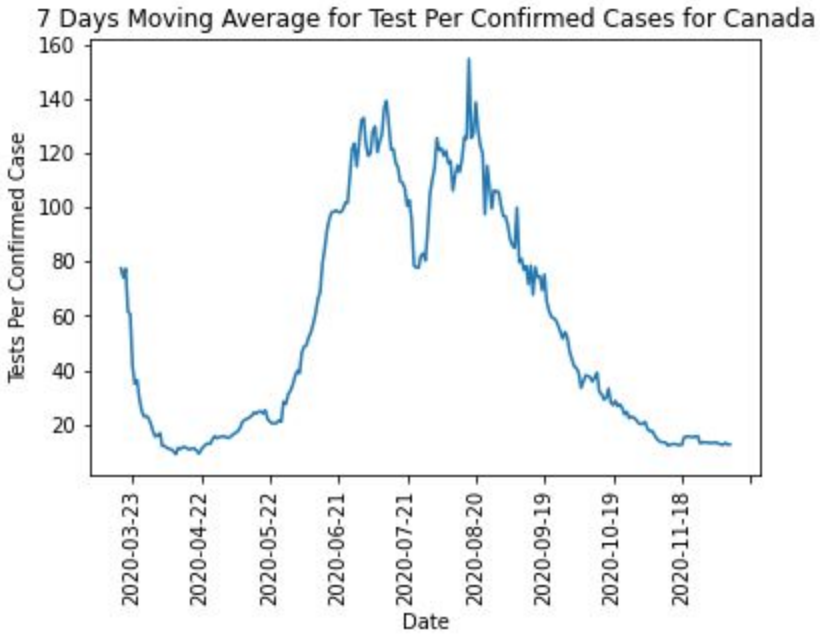
## 4. Rate of Recovered Cases Over Time for Country X

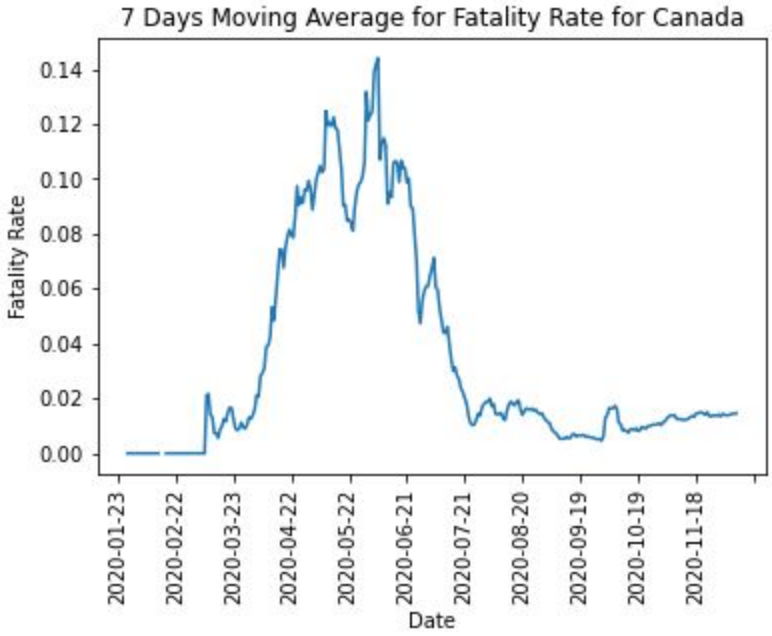


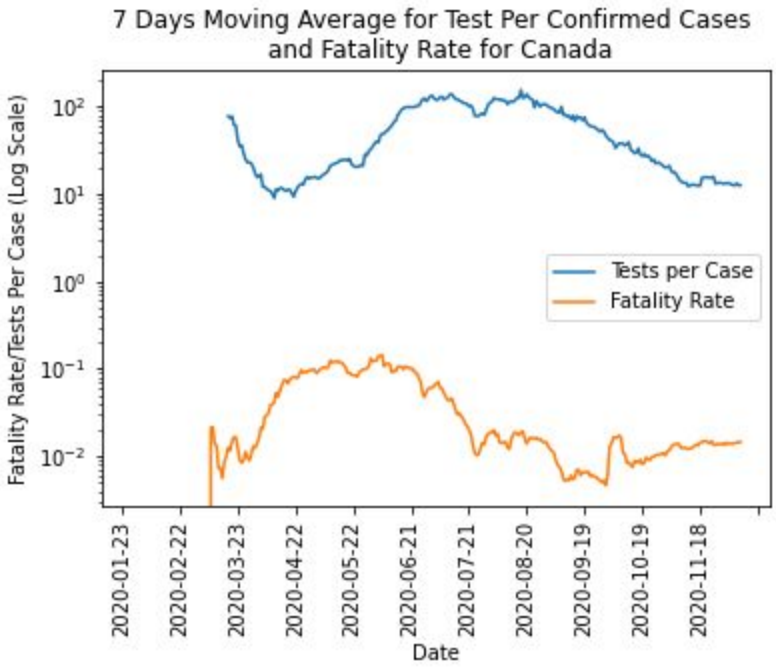

-To obtain the Rate of Recovered Cases Over Time for Canada, the John Hopkins University coronavirus case data csv was used with a big focus on the confirmed and recovered cases. Only the rows with the conditional of column location equaling Canada were then obtained as a data frame since we only needed the data for Canada. Rate of Recovered Cases Over Time was then calculated by dividing the number of recovered cases by the number of confirmed cases.The resulting values were then visualized. Since we were looking at the change over time, the dates were converted to datetime objects so we can just view them as overall months starting from March to December.

-Based on the visualization, we can see overall Canada has had a very high recovery rate. There was a small drop in April which was the beginning of the coronavirus spread, this makes sense as to why there was a drop in the recovery but surprisingly within a month the recovery rate began to increase again from May to August. There's a steep increase in August which could probably be the result of either new policies in regard to the coronavirus or increased testing which led to people identifying positivity and quarantining faster to recover. There had been a decrease in the fall, however, beginning in the middle of October. While the decline has continued from October to the present, we can conclude that Canada has had a very high recovery rate since April which shows they've dealt with recovery from the coronavirus pretty well.

## 5. 7 Days Moving Average for Test Per Confirmed Cases and Fatality Rate for Country X

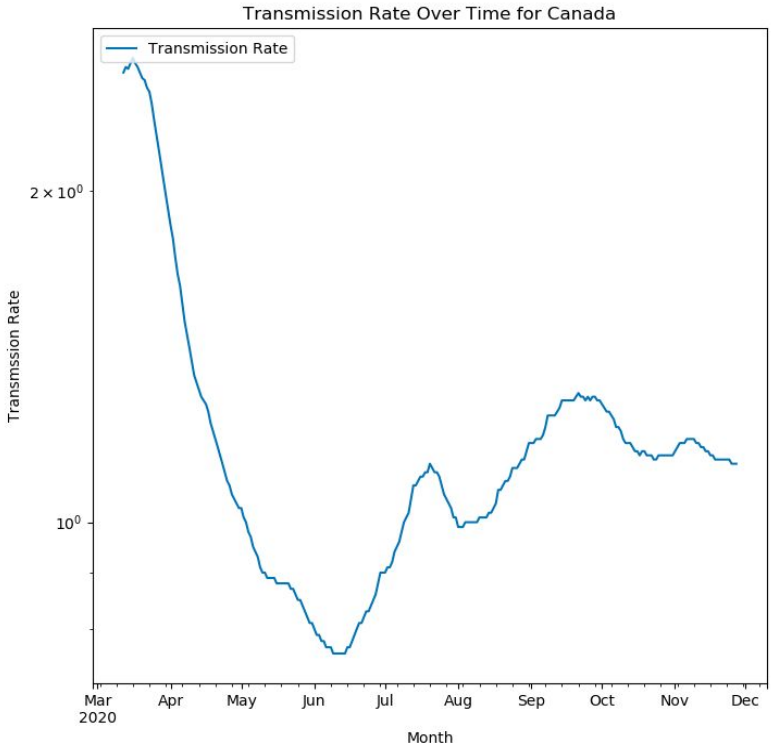


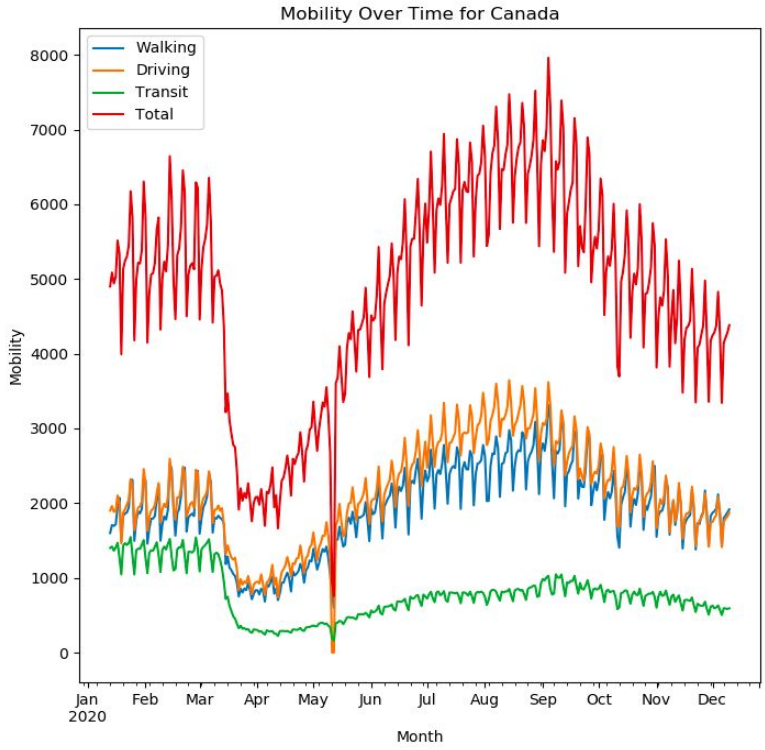



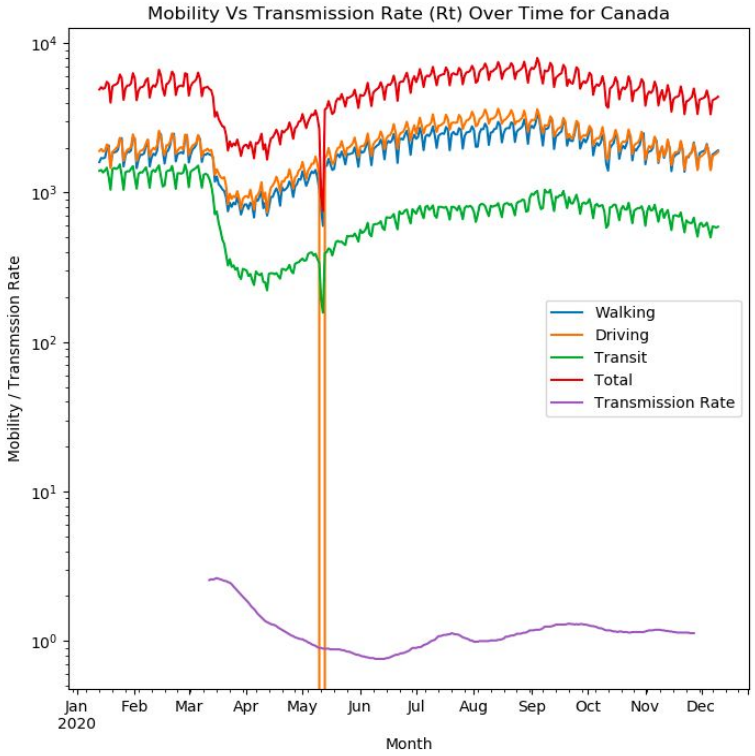



Interpretation: From April to June, tests per case was low and fatality rate was high. However, in July and August, tests per case was high and fatality rate was low. Finally, tests per case declined and fatality rate increased in a steady manner up to present day. However, despite the Tests per case being near the same value in November and April, the fatality rate was much higher in April. This may suggest that only severe cases were discovered due to a lack of tests early in the pandemic. More mild cases may have been documented later on as testing became more available. These mild cases are less likely to be fatal, resulting in a lower measured fatality rate by the data. From July to August, both tests per case and the fatality rate decline. There was a higher positivity rate from testing, but a lower fatality rate. This further suggests that a lot more mild cases were being discovered, with the cases becoming a bit more severe around October and November.

## 6. Mobility Vs Transmission Rate (Rt) Over Time for Country X



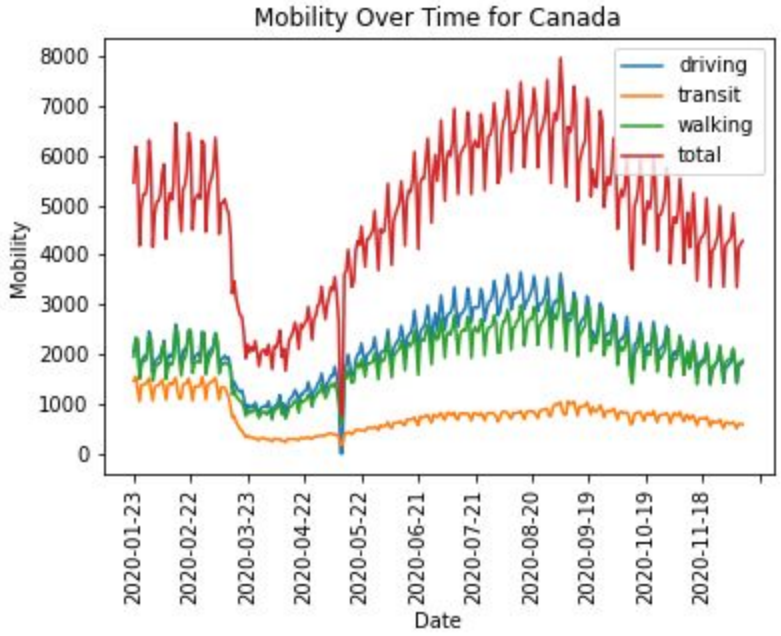



-To obtain the Mobility Vs Transmission Rate (Rt) Over Time for Country X for Canada, the mobility data csv provided by Apple was used to find the mobility data.The Our World In Data csv was used to get the transmission rate. Only the rows with the conditional of column country equaling Canada were then obtained as a data frame since we only needed the data for Canada. Since we had three different types of transportation in the mobility data, the data was split into 4 different kinds of data which were walking, transit, driving and total to represent both the individual transportation types and the overall transportation types data. The transmission rate data was then obtained by selecting the column reproduction_rate from all columns with Canada as their location value from the OWID data. The resulting values were then visualized both individually and then together on one graph.

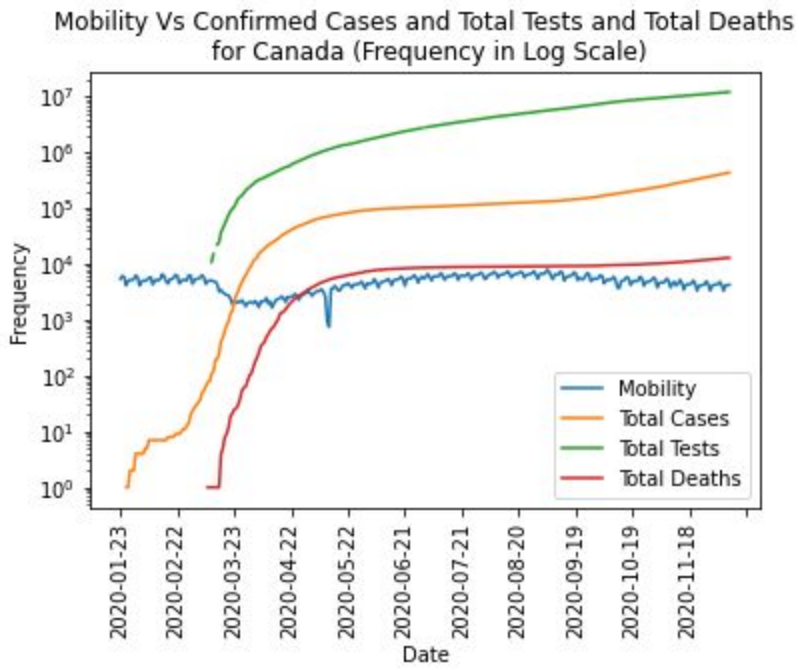
-Based on the individual visualizations, we can see the transmission rate was very high in the initial spread of the coronavirus for Canada but by May there was a steep decline for the transmission rate dropping to almost non-existent levels. Since fall began, we can see there has been a steady increase for the transmission rate to the point that it only saw a very small decline in November. For the individual visualization of Mobility, we can see mobility in the form of transit has declined completely since April which makes sense since that was the beginning of the spread of the virus so many people in Canada would probably choose to not use public transit at the time. All three types of mobility experienced this drop in that same time even though transit had the biggest decline. Surprisingly as time has passed, mobility increased with driving at the highest since many residents of Canada probably wanted to travel alone due to covid concerns. When both visualizations are plotted together, we can see that even though the

transmission rate has been somewhat consistent from March to the present, mobility has still been very high. After the short term decline of mobility in May, however, we can see a small decrease in transmission from May to July which means the lack of mobility helped to decrease the transmission rate. Similarly, once the mobility data started to increase again especially in July so did the transmission rate. We can conclude mobility has a direct effect on transmission rates but in smaller increments.

**7. Mobility Vs Confirmed Cases and Total Tests and  Total Deaths (Frequency in Log Scale) Over Time for Country X**

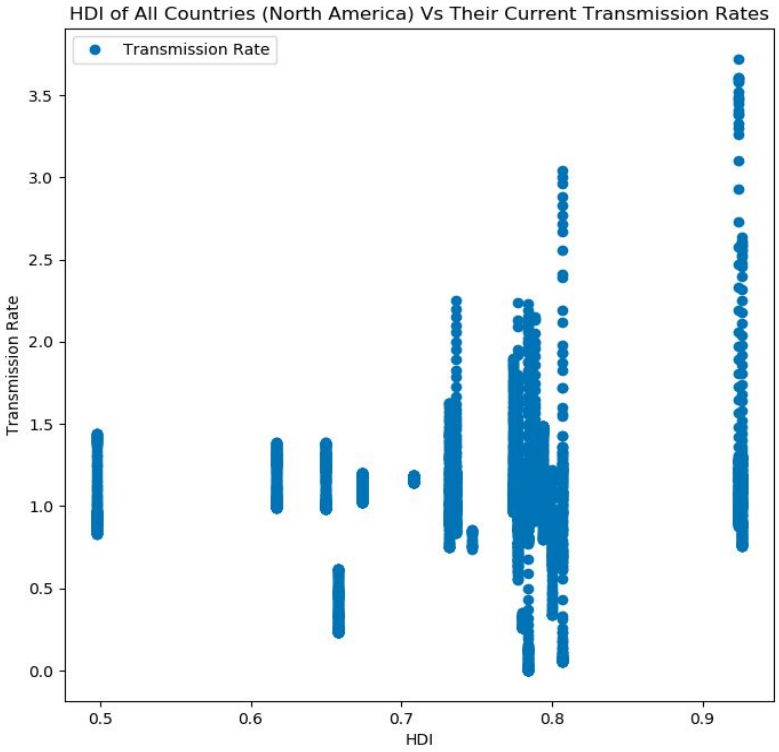



All of the mobility types trend similarly, so no need to look at them separately.

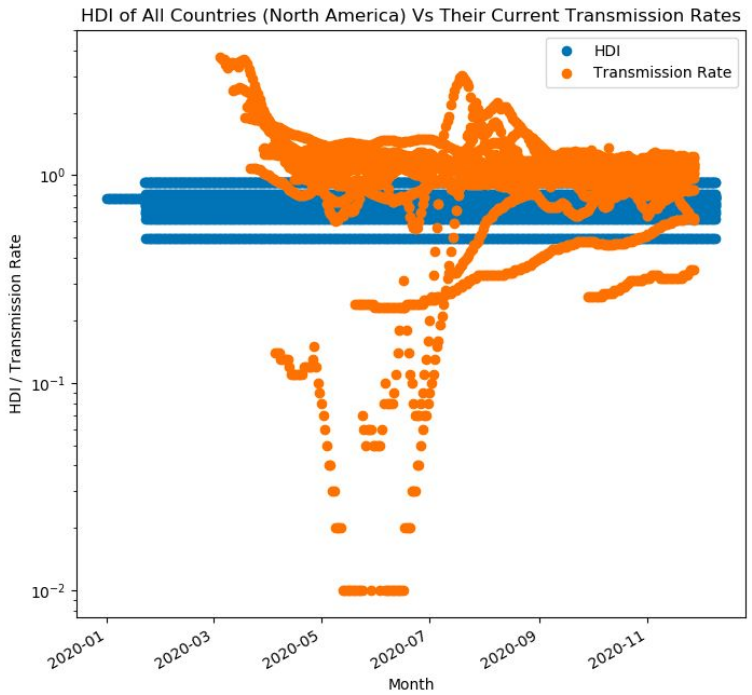



Interpretation: The rapid rise in cases in February and March may have caused people to travel less, as evidenced by the noticeable decrease in mobility around mid-March. Right after this decrease, tests, cases, and deaths all begin to flatten out. This may suggest that the decreased mobility of the people prevented the further spread of Covid. After the mobility increases slightly in August, we see the cases and deaths begin to increase a while after. This further demonstrates the correlation between mobility and covid transmission.

## 8. HDI of All Countries Vs Their Current Transmission Rates



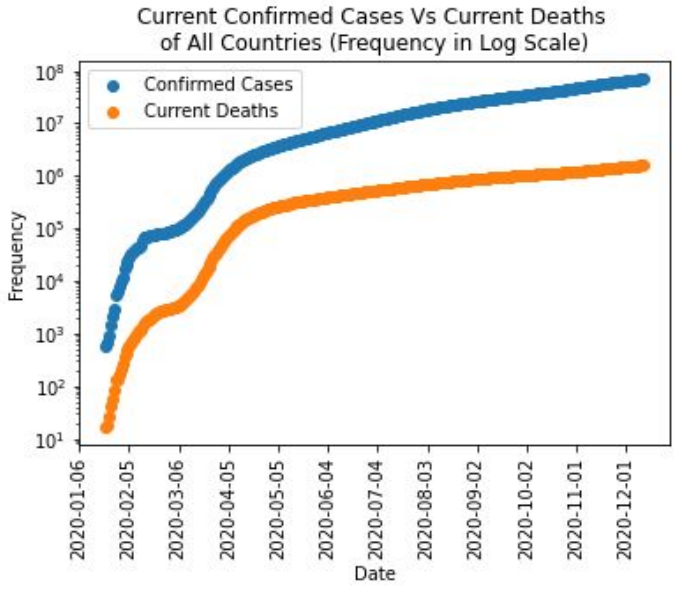

Alternate Method that was tried but didn't yield as great or interesting results:



-Since we couldn't visualize all the countries, the continent of North America was used in place of all countries. To obtain the HDI of All Countries vs Their Current Transmission Rates, the Our World In Data csv was read in and stored. Only the rows with the conditional of column continent equaling North America were then obtained as a data frame since we only needed the data for North America. The human development index column was then selected along with the reproduction rate column as well and both columns of data were then visualized. Since the line graph visualized was a bit messy, a scatter plot was used instead to visualize this data.

-Based on the visualization, we can see for North America as the HDI increases there's a small gap but gradual increase in their transmission rate with the highest HDI yielding the highest transmission rate. With our initial visualization (see above), we didn't really see much interpretation outside of the fact that the transmission rate was consistent and becoming higher than the HDI so as a result we tried to visualize the data using HDI as a label for X rather than data. From researching HDI's, we saw that Canada and the United States actually have two of the highest HDI in the 0.9 (0.92 for both of them as of 2020) range which makes sense as to why the transmission rate would be so high since the United States being included for the North America date probably skewed the results. We can conclude that based on this, in North America as HDI increases especially in terms of life expectancy and standard of living, there is a higher chance of transmission rate increasing though this may not be a final conclusion since the United States being part of this data could have skewed the results exponentially.

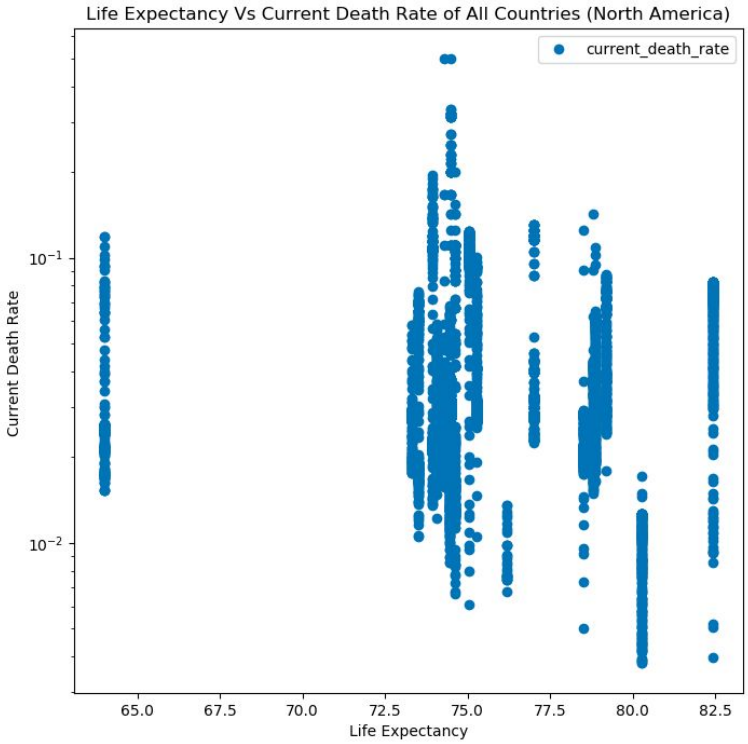
**9. Current Confirmed Cases Vs Current Deaths of All Countries**



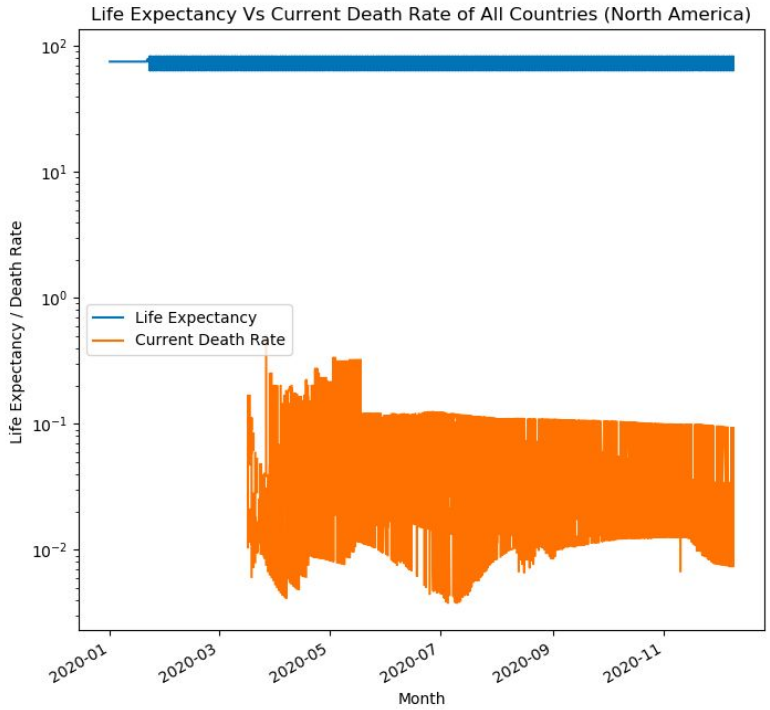

Interpretation: The trends of confirmed cases and current deaths follow almost the exact same trend, with confirmed cases being roughly two orders of magnitude greater than confirmed

deaths. This suggests the fatality rate did not vary significantly worldwide during the initial rise of Covid since the deaths and cases always trended similarly. However, the confirmed cases appear to increase more rapidly than the deaths after April, suggesting the fatality rate gradually decreased later on as covid knowledge, awareness, and treatments improved.

**10. Life Expectancy Vs Current Death Rate of All Countries**




Alternate Method that was tried but didn't yield as great or interesting results:




-To obtain the Life Expectancy Vs Current Death Rate of All Countries, the Our World In Data csv was read in and stored. Only the rows with the conditional of column continent equaling North America were then obtained as a data frame since we only needed the data for North America. The life expectancy column of data was then selected to be used for visualization. For the death rate, the columns of data for total deaths and total cases were stored and total deaths were divided by total cases to find the current death rate. Both the life expectancy and current death rate data were then visualized.

-Based on our visualization, we noticed that in regard to the death rate and life expectancy's relationship when visualized it seems that more healthy and younger people are less affected by the death rate which is to be expected. We tried the method of just plotting both graphs side by side to one another but it yielded less than stellar results as shown above. By plotting the life expectancy as the x label and using the death rate as the y label, we're able to further understand and conclude that a deciding factor for the death rate in North America is life expectancy in regard to who's healthier and younger compared to who's more unhealthy and older.