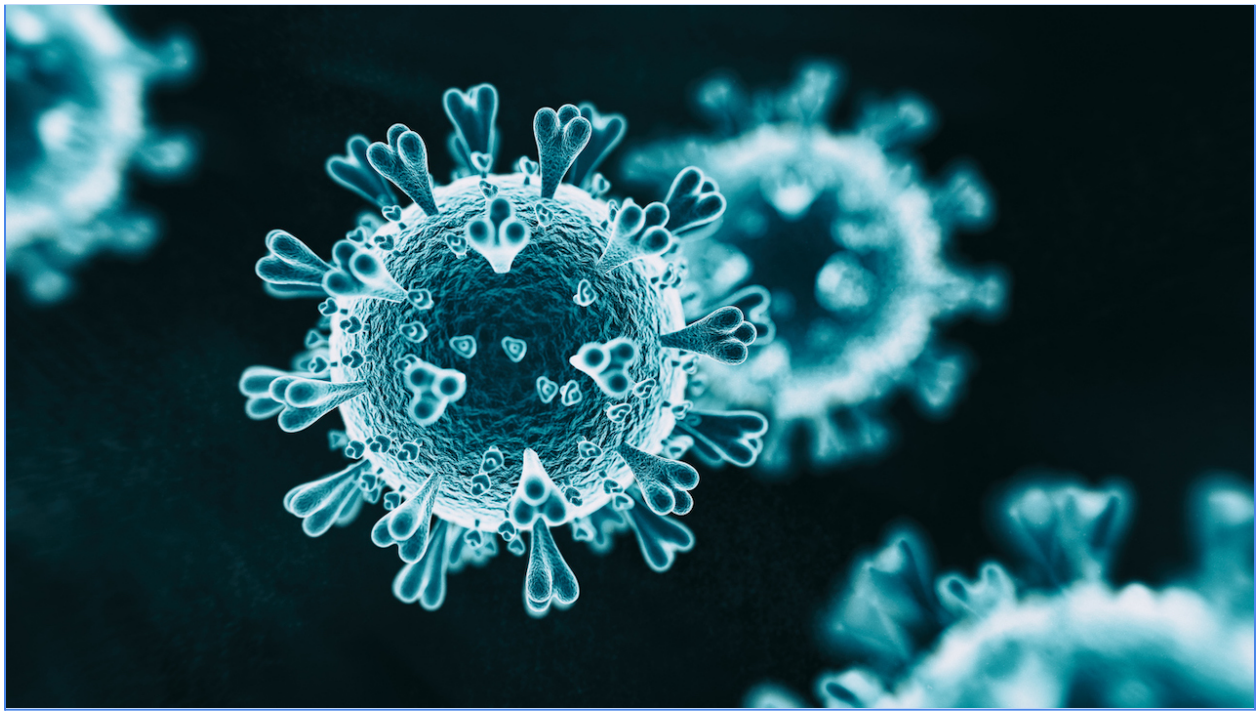
Canada v.s. The USA COVID-19 Analysis

A data science project bringing forth insight on the fight against the pandemic



Veer Sandhu

Data Science Intern, Science For All Audiences

August 8th, 2021

Code Repository
Interactive Project Link

INTRODUCTION

The COVID-19 pandemic has impacted every nation on earth and been dealt with through countless methods and strategies. Investigating where countries succeeded and where they failed can provide valuable information on resource management, prioritization, planning, and organization. Canada and the U.S may seem similar at first glance, however, they differ economically, socially, and politically. This project compares the two countries regarding how they dealt with the pandemic and where certain damages were dealt. Note that the documentation and report are written in plain language to provide easy understanding and comprehension.

The areas regarding measurements include the number of cases, deaths, tests performed, vaccines distributed, financial budget, and the COVID-19 action plans. These comparisons can bring many issues and problems to the spotlight. For example, one can tell if a nation misused its resources, lacked the capacity to fight the virus, had an inconsistent recovery, and even tell if a nation prioritized certain groups of people.

This report answers the question: How do Canada and the U.S. compare regarding their fight against the COVID-19 pandemic based on features such as cases, deaths, tests performed, vaccines distributed, and similar measurements?

I used the Data Science process to bring forth insight and information on these areas. Its steps are as follows: Ask an interesting question, get useful data, analyze/explore the data, model the data, and communicate the results/findings. This system allows for a structured way to provide information and recommendations regarding the topic and its future direction.

TOOLS AND TECHNOLOGIES

Data Science requires a set of tools just like any other piece of development. Specifically, the core requirement is a programming language that can perform statistical analysis. A programming language is a way to give the computer instructions so it can perform certain actions and calculations. Additionally, a way to visualize data and make graphs is also necessary. Other tools may include software that can download and process data (e.g. Excel/Google Sheets).

I used the Python programming language for this project which is a freely available and open-source platform. Python has excellent support for Data Science and is excellent for analyzing numerical information. This language is also simple and easy to use, meaning less time is spent on fixing errors and more resources can be focused on the data aspect. Furthermore, there are many libraries (or add-ons) that can be installed with Python giving it new features and abilities.

Regarding such add-ons, I included Pandas, Plotly, and Scikit-Learn throughout my developments. Pandas is a library used by many Data Scientists to read spreadsheet files, process data, and organize information in the form of tables. Plotly allows for data visualization in the form of interactive and appealing graphs. Finally, the Scikit-Learn library allows for data modeling and the prediction of new information.

I also used Excel in this project, however, not as an analysis tool but as a method to view the data after it was downloaded online. This allowed for an efficient way to scan the data and verify its condition to be processed.

DATA

I collected data for this project came through various sources which include the Canadian/American Government websites, Datalab, Kaggle, and Our World in Data. These datasets included information regarding the total COVID-19 cases, deaths, tests performed, and vaccines distributed. I also gathered information on the Government's COVID-19 action plans and financial reports through verified Government websites.

I downloaded these datasets in the form of spreadsheets and combined them into a singular Excel file. I then cleaned the data and removed missing values with Python and Pandas so it could be ready for investigation.

However, there is a crucial factor that must be examined before analysis begins. Recall that Canada and the U.S. are the two countries being compared and their population sizes drastically differ. Canada has a population of 37 million while the U.S. has a population of 328 million. Based on this information, it would be unwise to simply compare the total cases or deaths between the two countries. To fix this issue, new measurements were added that were proportionate to groups of the population such as cases per 100 people, deaths per 100 cases, and tests per case. These values will allow analysis to be well-grounded and fair, without including the factor of population size.

Table 1: Main Data Spreadsheet

	Country	Date	Cases	Deaths	Tests	Cases/Population	Diff Cases/Population	Deaths/Cases	Tests/Cases
0	United States	2020-01-22	1.0	0.0	0.0	3.046923e-07	0.000000	0.000000	0.000000
1	United States	2020-01-23	1.0	0.0	0.0	3.046923e-07	0.000000	0.000000	0.000000
2	United States	2020-01-24	2.0	0.0	0.0	6.093845e-07	0.000030	0.000000	0.000000
3	United States	2020-01-25	2.0	0.0	0.0	6.093845e-07	0.000000	0.000000	0.000000
4	United States	2020-01-26	5.0	0.0	0.0	1.523461e-06	0.000091	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...
537	Canada	2021-07-16	1430127.0	26450.0	37747547.0	3.804541e+00	0.101091	1.849486	26.394542
538	Canada	2021-07-17	1430437.0	26452.0	37794221.0	3.805366e+00	0.082469	1.849225	26.421451
539	Canada	2021-07-18	1430740.0	26457.0	37798244.0	3.806172e+00	0.080607	1.849183	26.418667
540	Canada	2021-07-19	1431378.0	26466.0	37907486.0	3.807869e+00	0.169726	1.848987	26.483211
541	Canada	2021-07-20	1431691.0	26468.0	0.0	3.808702e+00	0.083267	1.848723	0.000000

The table above shows the combined dataset that will be used for analysis. The first column label is “Country” which states whether that row comes from Canada or the U.S. The next column is “Date” which reads when that row of data was collected. Other columns include measurements such as cases, deaths, tests, or other equivalent measurements. Note that the screenshot does not include every single row of data to save space. The final size of this table was 1084 rows with 9 columns.

ANALYSIS

Using the numerical values collected so far, I built various graphs to bring forth insight into different areas. The axis labels are in green text and the graph title is in black text. The horizontal axis (left to right) mostly represents the date when the information was collected. The vertical axis (up and down) is a measurement from the data (deaths, cases, etc). Consider that these charts are mostly double line graphs meaning there are two lines represented by the legend on the right. Data from Canada is represented by the red line while data from the U.S. is represented by the blue line.

Graph 1

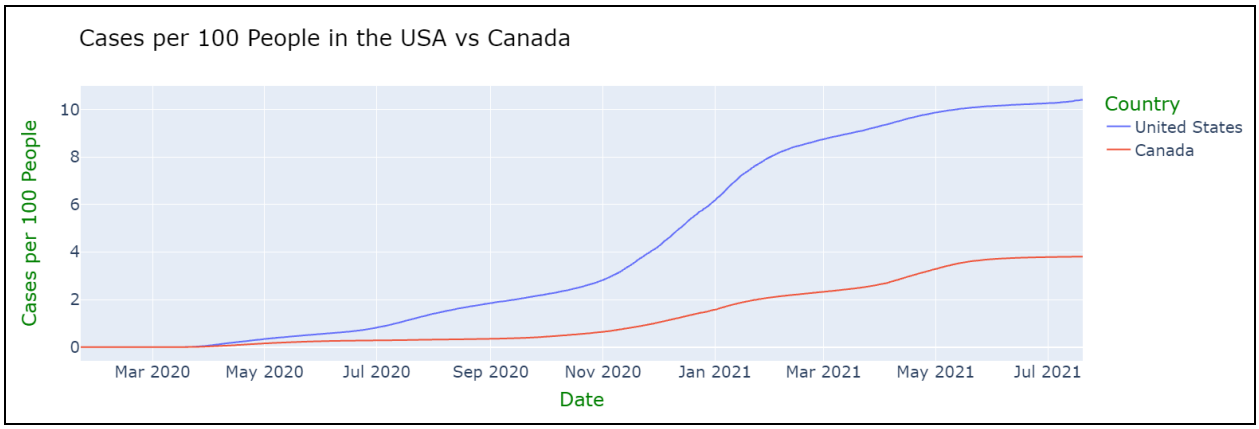


Figure 1: Comparison of COVID-19 cases per 100 people in the U.S. and Canada. The horizontal axis represents when the information was collected and the vertical axis represents the case value. Data: Open Source Platform (Our World in Data, 2021).

Is it clear that the U.S. has a higher peak of over 10 cases per 100 people, meaning 10 in 100 Americans have had COVID-19 at some point in their lives in July 2021. Canada also had its peak at the same time but had a lower value of around 4 cases per 100 people. Considering that the U.S. values were always higher than Canada's, it shows that their social distancing and spread prevention measures are weak for their population size. Another insight is that both nations are getting closer to "flattening the curve" as the latest dates show the lines being flat rather than on a steep incline (going up). Therefore, it is predictable that the cases will not drastically increase any time soon and a large spread of the virus will not occur

Graph 2

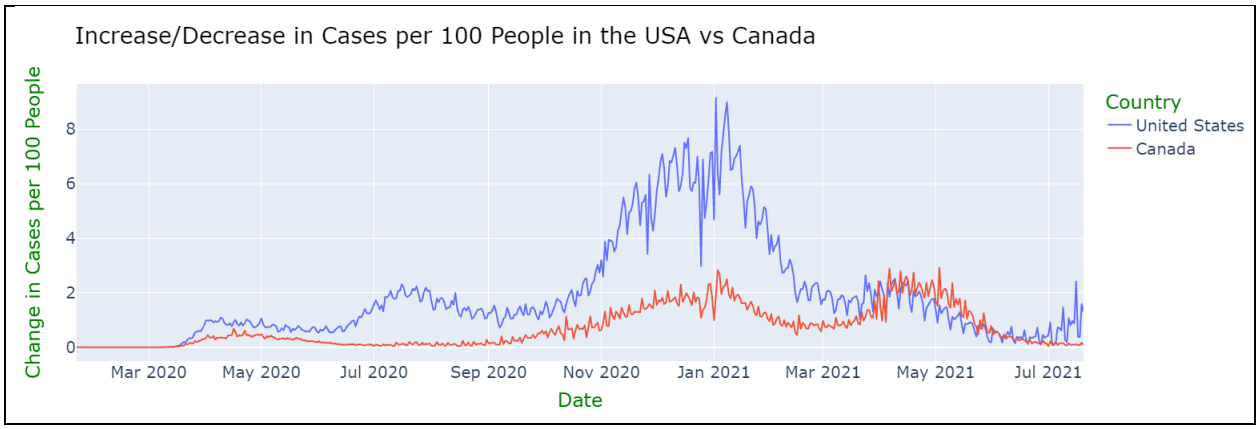


Figure 2: Increase/Decrease in cases per 100 people (daily) in the U.S. and Canada. The horizontal axis represents when the information was collected and the vertical axis represents the change in case value. Data: Open Source Platform (Our World in Data, 2021).

Using the measurements of the increase/decrease in cases, one can find the date on which the pandemic produced the highest increase in cases (spike/peak). In January 2021, both nations had a large increase in cases per 100 people. The U.S. reached an increase of +8 cases per 100 people while Canada reached +2. This event occurred due to many people visiting each other during the holidays (Christmas and New Year) which caused a large increase in the spread. Canada reached an increase of +2 cases per 100 people once again in April 2021 due to the government lifting lockdown measures earlier than required (resulting in another spread of the virus). Furthermore, Canada is currently on a decline of cases (line slopping down) since May 2021, meaning their prevention measures seem to be working. The increase of cases measurement during July 2021 is the lowest it has been since the start of the pandemic. However, the U.S. seems to be on an increase (line slopping up) as of the latest date, July 2021, which means the nation's containment of the virus needs to be improved. As such, it might result in

another spread of COVID-19 which will be detrimental to not only public health but economic growth as well (job losses, quarantining workers, etc).

Graph 3

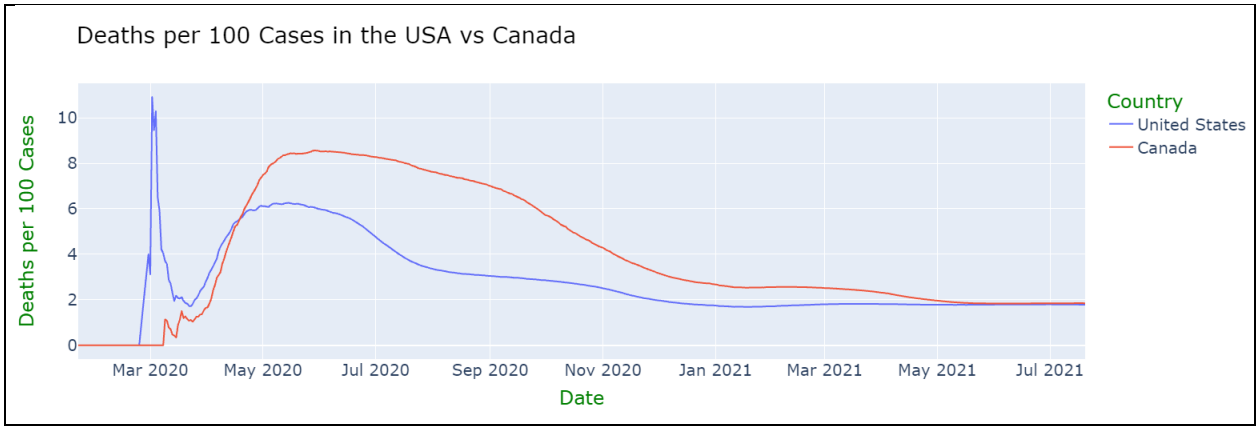


Figure 3: Comparison of COVID-19 deaths per 100 cases in the U.S. and Canada. The horizontal axis represents when the information was collected and the vertical axis represents the death value. Data: International Public Health (World Health Organization, 2021).

It is clear that the U.S.had its highest death to case ratio in March 2020. This high measurement occurred due to overcrowded hospitals, poor resource management, and a lack of medical equipment. Cities such as New York felt this first hand making it one of the worst medical emergencies in the history of the U.S. Furthermore, both countries reach a high death count during May 2020 before they start declining. This occurred due to cases reaching an all-time high since the start of the pandemic and public health struggled to treat the virus due to a lack of experience. Another key point in this graph is that Canada's death count relative to the population ranging from May 2020 to May 2021 was higher than the U.S. by 2 deaths per 100 cases. This can be justified by Canada's smaller medical workforce, larger vulnerable population (elders), and lack of experience at the time. However, both countries have lowered their death

counts to 2 deaths per 100 cases as of July 2021 which suggests that their treatment of the virus has improved.

Graph 4

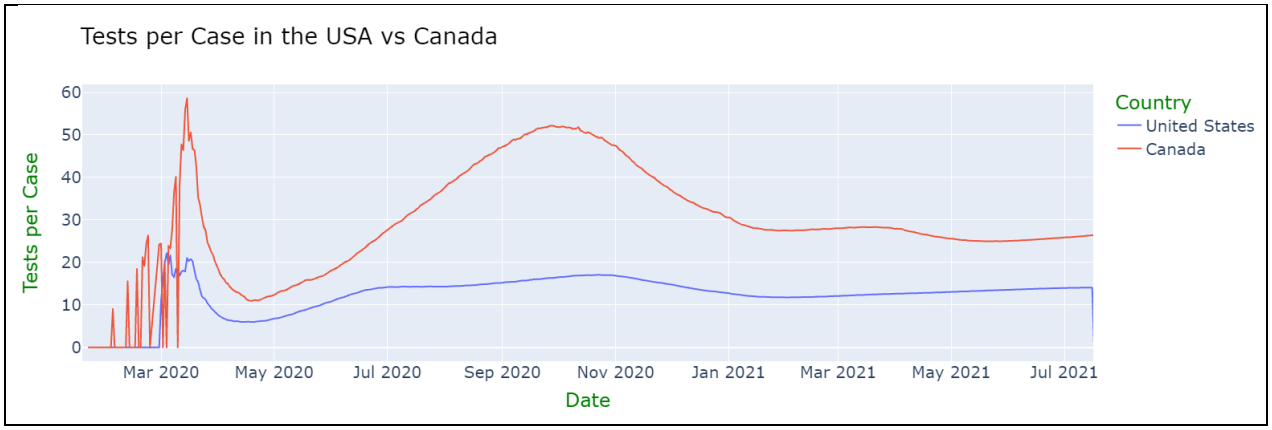


Figure 4: Comparison of COVID-19 tests performed for every case in the U.S. and Canada. The horizontal axis represents when the information was collected and the vertical axis represents the test to case ratio. Data: Government websites (Government of Canada, 2021) and catalogs (United States Government, 2021).

The measurement of tests per case allows one to determine if a nation's testing systems were adequate for their case count and population. It is clear that no matter which date is focused on, Canada has performed significantly more tests for every case than the U.S. This piece of information is crucial as the U.S. has a larger economy and more resources (Datalab, 2021). However, Canada still managed to provide nationwide testing services to all groups of people including those less fortunate. Performing a large number of tests can help a nation find the individual who has the virus as more people are getting checked for it. On the other hand, if a small amount tests are being performed, there is a lower chance the infected person will be found. This specific scenario occurred in the U.S. where there was an extremely high case count but few tests being performed, as such, those who had the virus would not have knowledge of

their infection until clear symptoms appeared. If the population size of a country is not matched by the tests performed, the case count will drastically increase due to the unawareness of positive cases (those with the virus) and high-risk areas.

Graph 5

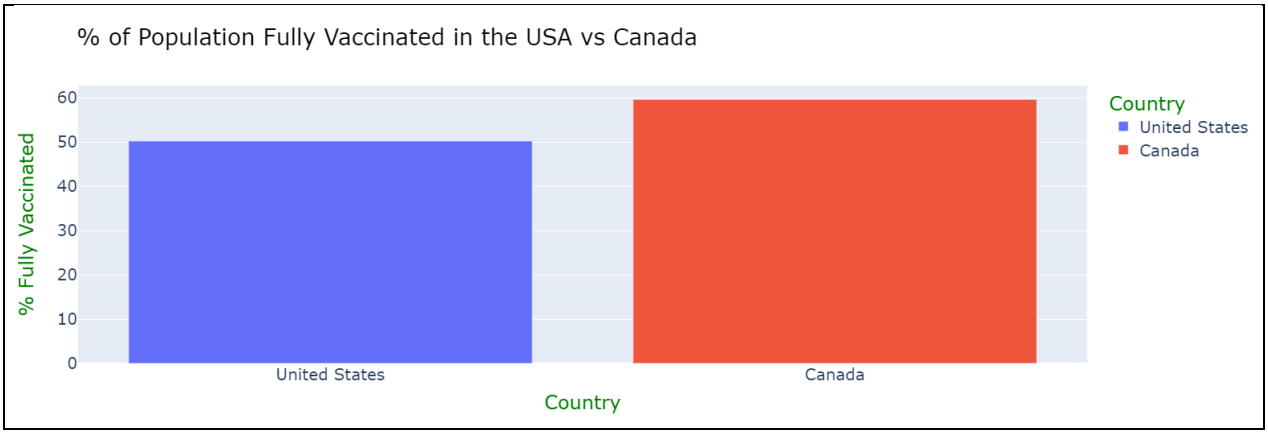


Figure 5: Comparison of the percentage of the population fully vaccinated in the U.S. and Canada. The horizontal axis represents the country and the vertical axis represents the percent vaccinated. Data: International Public Health (World Health Organization, 2021).

As stated in the previous graph analysis, the U.S. has more resources and a larger economy than Canada when regarding factors such as the healthcare budget and GDP (Department of Finance Canada, 2021). However, Canada has vaccinated 10% more of its population than the U.S. (measuring 60% to 50%). The vaccine distribution plan of Canada was far superior to that of the U.S. due to its organization and planning (Lewis, 2021). The public was vaccinated based on prioritization which was done by the age group starting with the elderly, then the middle-aged, and ending with youth. The vaccine distribution in Canada was done fairly and effectively for its population which will provide a smooth recovery from the virus.

DATA MODELING

After the data has been analyzed, it can be modeled through mathematical means which allows one to predict future information and simulate certain scenarios. For example, one can model the date and the number of cases to forecast how many people will get the virus based on previous trends. Data can be modeled through various means, however, this project used Linear Regression (the line of best fit) which allows for easier understanding and prediction of data numerical.

MODEL 1

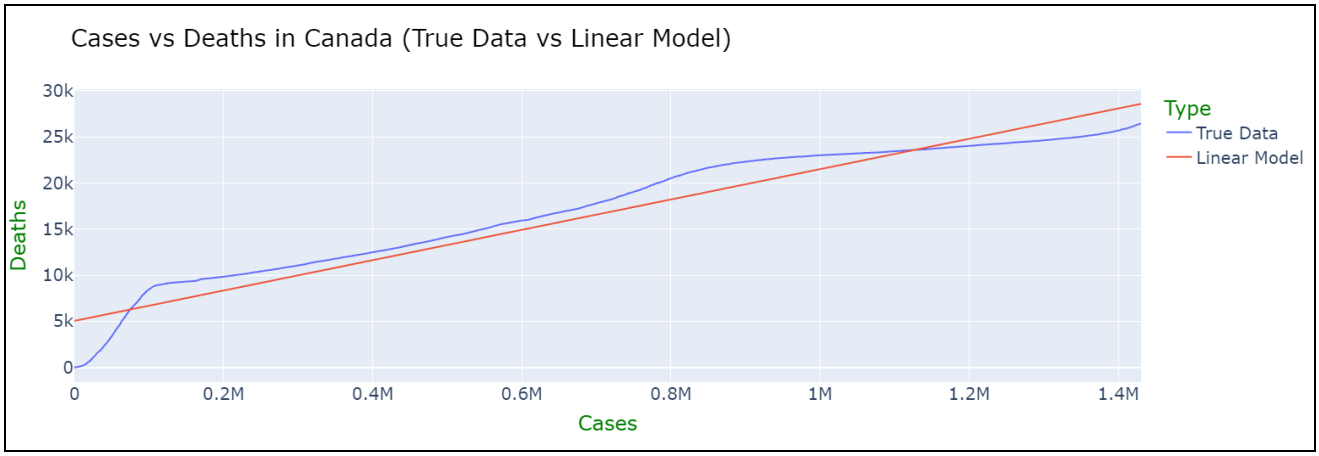


Figure 6: Model of the total cases to the total deaths in Canada through Linear Regression (90% accuracy)

This model proves to be useful is due to the fact that we can forecast the death count given the number of total cases during that time period. For example, if there was a scenario where Canada reached double its current case count, the model can predict the amount deaths it would cause. The scenario that was simulated in this project was Canada reaching a total of 2 million cases which is much higher than its current total of 1.4 million.

Even though the 2 million value is not on the horizontal axis for the total case count, it can still be used for prediction as this model extrapolates information in the sense that the line continues after the visible grid. Therefore, it can be used to predict any death count if given a viable input value (case count). The model predicts that if Canada reached 2 million cases, the total death count will reach around 38,000. If this is compared to the U.S, it displays that when they reached 2 million cases, the death count was 116,000 (over 3 times the value of Canada). This proves that at the same case count, Canada would have far fewer deaths than the U.S. according to the trained model. It must be known that health care in the U.S. is not as accessible as in Canada due to the fact that the patients must directly pay for their care in hospitals. As such, those who cannot afford those facilities will have a hard time receiving the necessary treatment.

MODEL 2

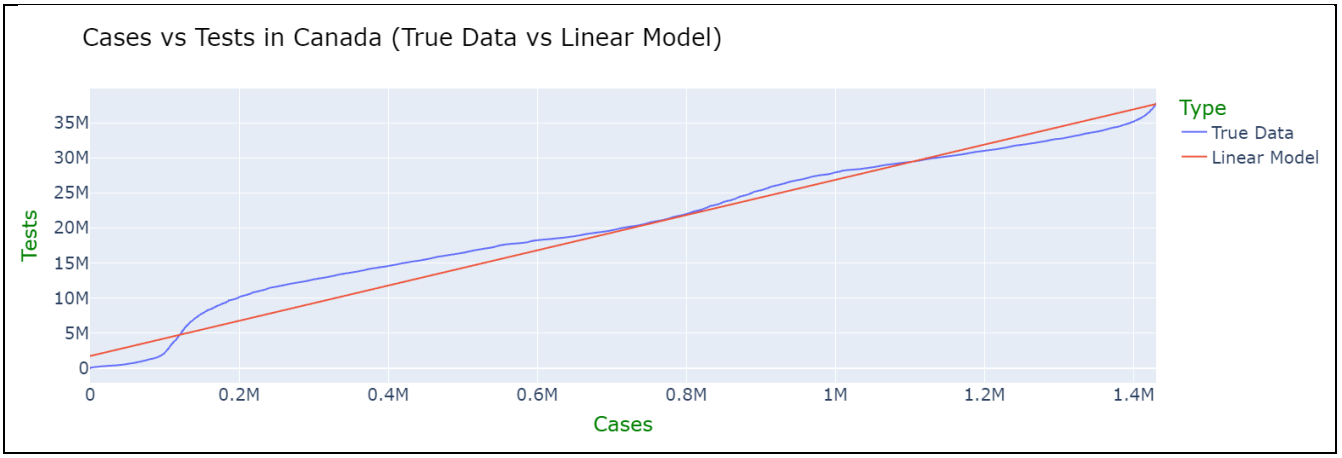


Figure 7: Model of the total cases to the tests performed in Canada through Linear Regression (96% accuracy)

The scenario being simulated in this project was Canada reaching 2 million cases as stated in the previous description. This model predicts that if Canada reached a total of 2 million cases at some point, the COVID-19 tests will reach a total of 52 million (this counts tests from

the start of the pandemic until now). Canada has currently performed a total of 39 million tests and this number has proven to rise with the case count based on the linear model. This is far different from the U.S. due to the fact that when 2 million cases were reached in their nation, there were only 25 million total tests. This shows how Canada's testing facilities are far more efficient and applicable for its population size compared to the U.S. As justified in the analysis, if a nation does not perform enough tests, the case count will drastically rise due to the lack of information for those who have the virus.

CORE FINDINGS AND FINAL INSIGHTS

There were several conclusions and insights gained from this project comparing Canada and the U.S. fighting COVID-19. The numerical differences are as follows: the U.S. had a case count (relative to the population) triple than that of Canada, they also had 4 times the maximum increase (most cases in a day), Canada performed twice as many tests relative to the case count, Canada vaccinated 10% more of their population, and the U.S. had a peak death count over 5 times than that of Canada. The similarities are as follows: both countries currently have 2 deaths for every 100 cases and their cases are roughly increasing by 2 for every 100 people per day which has improved greatly since the start of the pandemic.

Furthermore, based on external research, there were many differences in the U.S. and Canada's action plans regarding COVID-19. Canada has spent $150 billion on fighting the pandemic while the U.S. has spent $3 trillion (United States Government, 2020). However, it must be known that the U.S. has a far larger economy than Canada. Nevertheless, Canada has provided superior services to the public such as financial aid for families, services on indigenous reserves, a high amount of testing facilities, and an excellent vaccine distribution plan. On the

other hand, the COVID-19 action plan of the U.S. had a late response, flawed isolation systems, overcrowding in health care facilities, and slow testing procedures (Lewis, 2021).

This project proved that Canada has fought the pandemic better than the U.S. while suffering fewer damages and employing successful systems (vaccine availability, testing services, and health care control). This analysis shows the importance of resource management, organization, and planning. Those with more tools at their disposal will not find any success if they do not find ways to use them to their greatest ability.

REFERENCES

Datalab. (2021). The federal response to Covid-19: U.S. Treasury Data Lab. Federal Response to COVID-19 | U.S. Treasury Data Lab. https://datalab.usaspending.gov/federal-covid-funding/.

Department of Finance Canada. (2021, July 31). Canada's COVID-19 Economic Response Plan. Canada.ca. https://www.canada.ca/en/department-finance/economic-response-plan.html.

Government of Canada. (2021, April 21). COVID-19 Planned Expenditures. Canada.ca. https://www.canada.ca/en/treasury-board-secretariat/services/planned-government-spending/supplementary-estimates/supplementary-estimates-c-2020-21/covid-19-planned-expenditures.html.

Government of Canada. (2020, April 22). Interactive data visualizations of COVID-19. Interactive data visualization of COVID-19 in the world - Public Health Infobase. https://health-infobase.canada.ca/covid-19/international/.

Lewis, T. (2021, March 11). How the U.S. Pandemic Response Went Wrong and What Went Right. Scientific American. https://www.scientificamerican.com/article/how-the-u-s-pandemic-response-went-wrong-and-what-went-right-during-a-year-of-covid/.

Our World in Data. (2021, March 22). Coronavirus Pandemic (Covid-19) - Statistics and Research. Coronavirus Pandemic (COVID-19). https://ourworldindata.org/coronavirus.

The White House. (2021, January 21). The Biden-Harris plan to beat COVID-19. The White House. https://www.whitehouse.gov/priorities/covid-19/.

The World Bank. (2020). GDP - Canada, United States. The World Bank Data. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=CA-US.

United States Government. (2021, September 2). Covid-19 Diagnostic Laboratory Testing (PCR testing) time series. Data.gov. https://catalog.data.gov/dataset/covid-19-diagnostic-laboratory-testing-pcr-testing-time-series-80872.

United States Government. (2020). Financing the Government. U.S. Department of the Treasury. https://home.treasury.gov/policy-issues/financing-the-government.

United States Government. (2021). Government Response to Coronavirus, COVID-19. USAGov. https://www.usa.gov/federal-covid-response.

World Health Organization. (2021, July). WHO COVID-19 Explorer. COVID-19 Explorer. https://worldhealthorg.shinyapps.io/covid/.