

Analysis of Covid data in Canada

David Chae and Parker Johnson

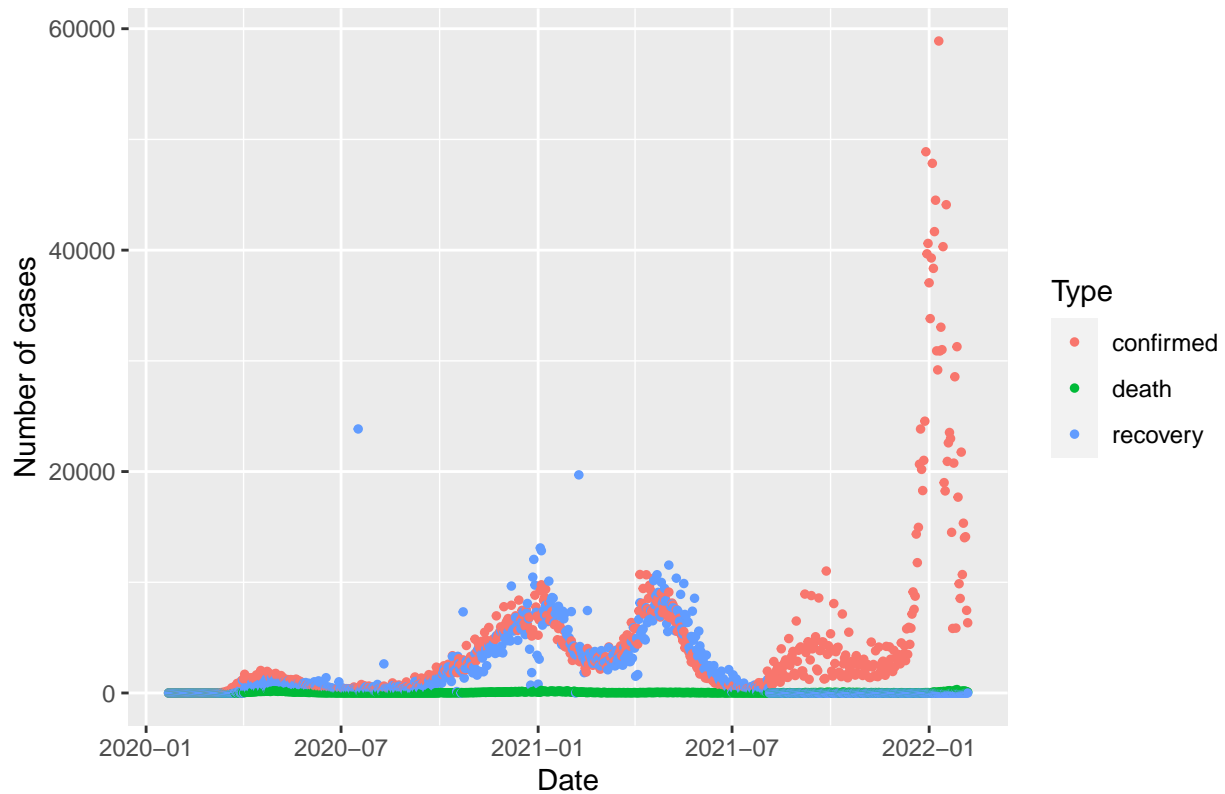
Introduction

With the COVID-19 pandemic still ongoing, there is an abundance of data that can be analyzed. We decided to analyze data regarding Canada's COVID cases and vaccinations, because we were curious to learn about this virus outside of the US and explore provincial data when possible. We used two datasets about COVID cases and vaccinations that were obtained from Rami Krispin on Github, who scraped the data from the Johns Hopkins University Center for Systems Science and Engineering. We used time series analysis to investigate the types of COVID cases and vaccination rates over time, as well as creating maps to investigate total COVID cases per Canadian province.

Results

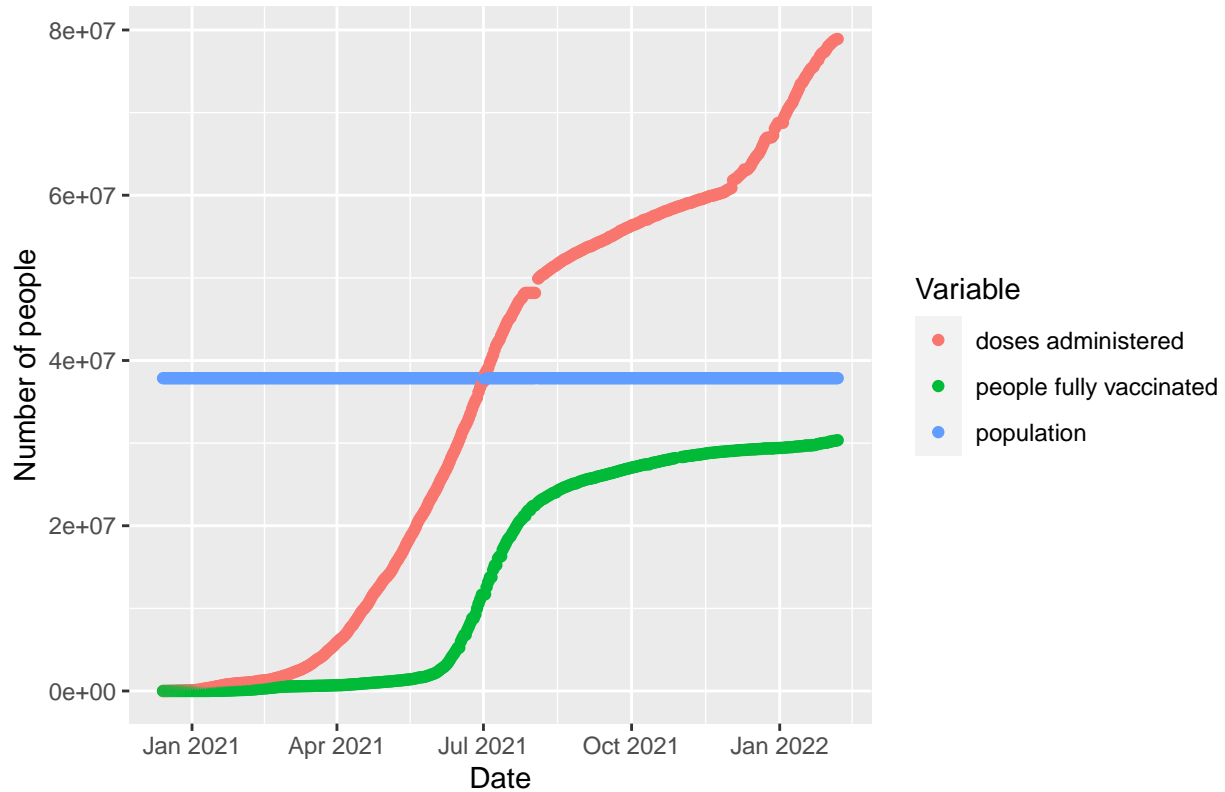
From figure 1, we can observe that from the start of this dataset(January 2020) until about the end of July 2021, the daily confirmed cases and recovery reports tended to stay about the same and that, at least compared to the number of confirmed cases and recovery reports, the number of deaths due to Covid remained negligible throughout the 2 year period. Another observation to note is that starting from the beginning of August 2021, there are no reported cases of recovery. We believed that this was most likely due to the data set not containing any values, as it is extremely unlikely that no one in Canada recovered from Covid between August 2021 and now. We later discovered that the data collection for recovery cases stopped around this time. Finally, from around October 2020 to around December 2021, the confirmed cases seemed to follow a sinusoidal pattern with 3 cycles before suddenly spiking in numbers around late December 2021.

Figure 1: Canada COVID cases by type over time



Observing figure 2, throughout the 2 year dataset, the population of Canada was assumed to be approximately constant. Both the number of doses administered and the number of people fully vaccinated steadily rose throughout the observation period. The rate at which the doses administered increased seemed to shift 3 times throughout the observation period, rapidly increasing around Mid May 2021, decreasing around August 2021 before increasing again in December 2021. The number of people fully vaccinated rose at a slow rate before rising extremely quickly from June 2021 to August 2021 and returning to a slow rise, approaching, but not exceeding, the population of Canada.

Figure 2: Vaccination and doses versus population in Canada



We then looked at cases per province, as shown in the table below. We found the total number of confirmed cases, total deaths, and the total number of fully vaccinated individuals within each province. However, the datasets did not contain vaccination information per province, which will be more thoroughly discussed later. The province with the highest number of confirmed cases is Ontario with 1054061 cases, and the highest number of deaths is Quebec with 13,481 deaths. The lowest number of cases and deaths is Nunavut, and Quebec contains the worst proportion of individuals who die out of those confirmed to have COVID.

province	total vaxxed	confirmed	death	death prop
Alberta	28272392	503434	3652	0.0072542
British Columbia	0	330638	2678	0.0080995
Manitoba	0	122637	1589	0.0129569
New Brunswick	0	30762	259	0.0084195
Newfoundland and Labrador	0	18244	45	0.0024666
Northwest Territories	0	6558	17	0.0025923
Nova Scotia	0	39129	157	0.0040124
Nunavut	0	1935	5	0.0025840
Ontario	0	1054061	11829	0.0112223
Prince Edward Island	0	8766	12	0.0013689
Quebec	0	881055	13481	0.0153010
Saskatchewan	0	123259	1015	0.0082347
Yukon	0	3186	18	0.0056497

From these maps in figures 4 and 5, we can see that Ontario has the highest number of total confirmed COVID cases of the Canadian provinces, while provinces such as the Yukon, Northwest Territories, and Nunavut have a low number of confirmed cases, which is understandable given the populations of each province. The second map shows that Northwest Territories have the highest proportion of confirmed cases per population,

with over 14% of the province's population contracting COVID at some point. Newfoundland and Labrador have the lowest proportion of confirmed COVID cases.

Figure 4: Total Confirmed COVID Cases

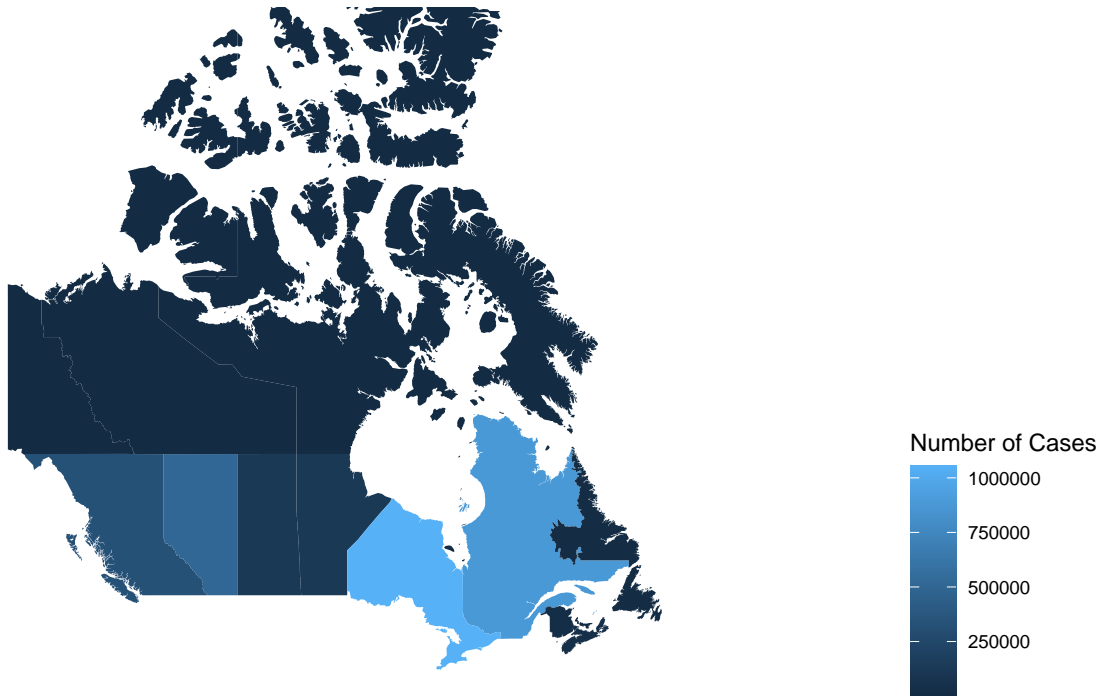
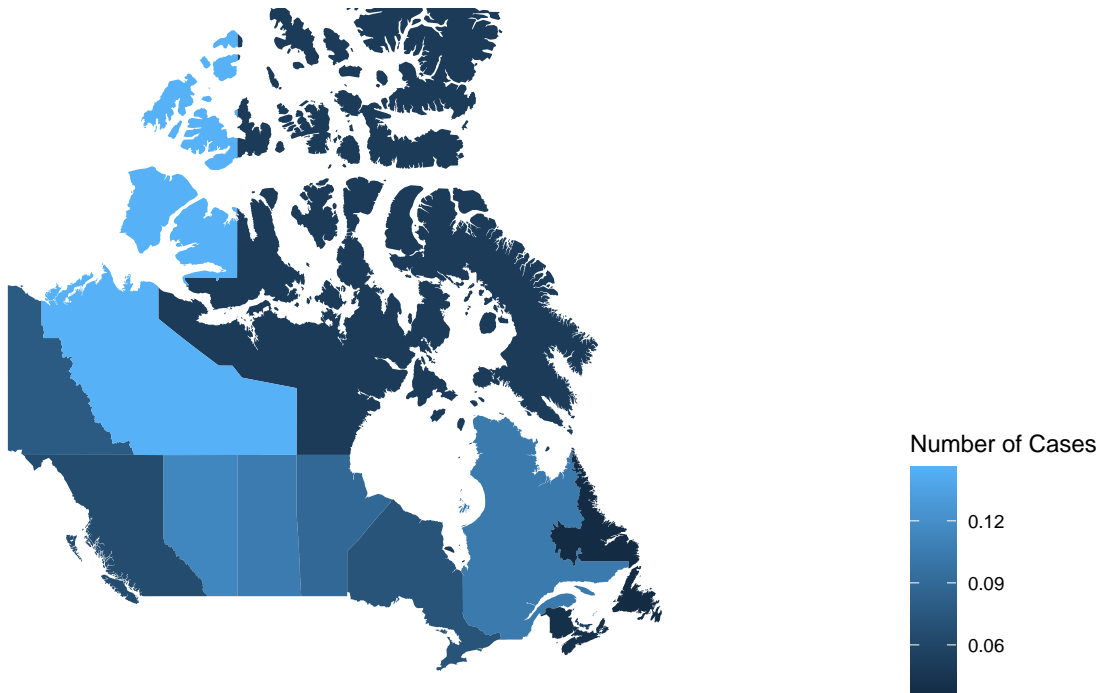


Figure 5: COVID Cases Per Province Population



Discussion

From these results, we were able to achieve a better understanding of the COVID situation in Canada. These results show that, in the time period in which we have recovery data, Canada has done very well as a country in terms of keeping the infection rate and recovery rate relatively equal. Also, as of today, Canada has a large proportion of its population having at least one dose of the vaccine, with the total amount of vaccine doses administered being a little over twice the population and fully vaccinated people approaching the total population. These results also explain the prevalence of the virus within each province of Canada.

One limitation of our analysis is that the original dataset stopped recording values for the number of individuals who recovered from COVID after August 2021. Therefore, our graph of the types of cases is incomplete and does not illustrate the true values for those who recovered from COVID. Another limitation is that the original dataset did not contain vaccination information within each province. While the steps we followed to produce the table were correct, accurate values were not produced for the total number of fully vaccinated individuals because either the information didn't exist, or the dataset had an error and recorded the fully vaccinated information for all of Canada. This was especially the case for October 29th and 31st 2021 where data that would have made sense to have been for Canada as a whole was recorded for Alberta.

Visualization and Data Wrangling

For our cases by type plot, we grouped the data by date and type of case, then summed the cases together before creating a point plot.

For our vaccination plot, after grouping by date, we filtered our data for just data points for Canada as a whole, as we discovered that for some of the dates the data was being recorded for each province as well as Canada as a whole, essentially double counting the data. We then replaced the NA values in the `doses_admin`, `people_partially_vaccinated`, and `people_fully_vaccinated` variables with zero for the sake of analysis.

For our table, we grouped by province and case type, then created a summary table for total cases and people fully vaccinated. We then removed recovery since it contained many NA values, removed 3 strange places listed as a province, and transformed the table to flush the types of cases into columns instead of rows. Finally, we added a column for the proportion of people who died out of confirmed cases.

For our maps, we imported data that contained information for creating a map of Canada. We then created tables corresponding to total cases per province and cases per a province's population, transformed the information into a data frame, then joined this data frame with the data that contained the map of Canada. From here, we could plot the map and fill each province with a color corresponding to the number of cases / cases per province's population.

For our data wrangling steps, first, we decided to look into Canada, so we filtered out the data points containing Canada. We also decided that some variables, such as the country and continent codes, were unnecessary in our observations, so we removed them from our dataset for cleanliness. We noticed that some numerical data points were negative or did not have any value. For the sake of analysis, we set those data points to be zero. Finally, the original data sets contained data points for Canada as a whole and left the province variable for them blank. Since it was useful for us to analyze those data points, we set the province variable to "Canada".