

# CSC110 Project: **Effect of COVID-19 on Crime in Canada**

Wangzheng Jiang  
Khushil Nimesh Nagda  
Nicholas Harold Poon  
Raghav Rengarajan Srinivasan

December 14, 2021

## 1 Problem Description and Research Question

During the beginning of the COVID-19 pandemic, a newfound fear of catching the virus caused many people to become increasingly conservative regarding their usual activities. Going to public places only increased the risk of acquiring COVID-19. As a result, the general population was advised to stay indoors as much as possible and not to go to public gatherings and events where the virus may spread. In addition, during this period of time, public health measures by the Canadian government caused many businesses and organizations to begin implementing stricter social distancing and general operating procedures to manage the spread of COVID-19 across the population. For instance, restaurants began limiting the number of people able to eat indoors, and face masks have become a necessity to enter almost any public building. The provincial government even went as far as shutting down non-essential businesses and putting the whole province into lockdown to ensure that the virus was contained as much as possible.

Similarly, measures such as the requirement for proof of vaccinations for many indoor facilities and event spaces became prevalent strictly as a result of the COVID-19 pandemic. For this project, our group aims to investigate how the COVID-19 pandemic and the public health measures initiated by the government have impacted crime and calls for service (calling the police, EMS, etc.) in Canada. We are particularly interested in exploring the effect of the pandemic on crime because this topic is one that does not seem to be super prevalent in the media. Typically, we tend to only hear about radical cases of crime from the news (terrorism, extreme violence, abuse, etc), but this does not really give us a sense of the trend in the amount of crime in general, which we believe is important to analyze. Therefore, our research question is “**How has COVID-19 and the Canadian governments response to the pandemic changed the amount of (reported) crime in Canada?**” Our group hypothesizes that as a result of this trend of decreasing activity outdoors and in public spaces, organized-crime activities during the COVID-19 pandemic will have decreased, which may be reflected by the computations that we plan to perform on our data in order to analyze this trend.

## 2 Dataset Description

Both of our datasets on crime and COVID are from **Statistics Canada** and are in **CSV** format. In the dataset on selected police-reported crime and calls for service during the COVID-19 pandemic[1], we are provided with information about the number of 911 calls for each type of service required, for each month from March 2019 to August 2020. This includes calls for enforcing physical crimes and requests for help with mental health and suicide prevention. For this project, we will only be focusing on the 911 calls that relate to crimes. Moreover, we only used the `REF_DATE`, `GEO`, `Violations and calls for service`, and `Value` columns. The `REF_DATE` column provided the date of the call, the `GEO` column told us the location of the call down to a region of Canada, the `Violations and calls for service` column gave us the type of incident that provoked the call, and the `Value` column gave us the number of calls for the same emergency type in the month of the call. This data was stored for every observation listed in the dataset as `EmergencyCall` instances, which contained attributes corresponding to these values. There was sufficient data in the dataset before the pandemic to compare the change in crime before and during the pandemic, which we were able to analyze through statistical models.

The second data set[2] that we will use provides us with a daily update about the number of COVID-19 cases in each province and in Canada as a whole. This daily update starts on January 31, 2021 and is still continuously being updated. It is important to note that in the beginning of the dataset, there is no update on the number of COVID-19 cases everyday, probably because there were not new cases every day at the start of the pandemic. For this dataset, we only used the `prname`, `date`, `num_conf`, and `num_deaths` columns. The `prname` column gave us the province or territory for which the data corresponds to, the `date` column gave us the date for which the data was collected, the `num_conf` column gave us the total number of active cases in the particular province, and the `num_deaths` column told us the number of new deaths recorded in the particular province on the given day. As with the emergency call data, this information was stored in a list of instances of the `CovidData` class, with attributes corresponding to each value.

Filtering the list of all `EmergencyCases` instances to get just the calls induced by crimes, we were able to analyze this data in conjunction with the list of `CovidData` instances using statistical models to answer our research question.

## 3 Computational Description

The data in `police_data.csv` is first read and put into a list of `EmergencyCall` dataset instances in the `read.py` file. We have neglected the `Royal Canadian Mounted Police [99C01]` rows because this information does not provide us with the exact location of the crime. In the `analyze.py` file, we have many helper functions that filter this data. For instance, we filter the calls by just crime-related emergencies, the type of crime reported, and the location in which the crime was reported. We distinguish between "physical" and "non-physical" crimes as well as "public and "private" crimes, and use keyword filtering to sort the `GEO` column into a province or territory, or Canada as a whole if it's a total. We also have a helper function that rewrites the `REF_DATE` column to assume that each crime occurs on the last day of each month. This is to ensure that the date formats are parallel across the 2 datasets and so they can be converted into `datetime.date` objects. The function `emergency_call_to_dict` returns a dictionary mapping relevant attribute names of `EmergencyCall` to the list of instance attributes for all the `EmergencyCall` instances in the input list. This dictionary is then translated into a `pandas` data frame. The purpose of `get_police_data_totals()` is to

add up the number of crimes in a certain category per month for a year. There are 2 keys; the first key is the date and maps to a list of 12 `datetime.date` objects that refer to the last day of each month for a certain year. The second key is the total crimes in a certain category for a certain location for the year. This key maps to a list of integers. The integer at index `i` represents the total number of crimes, in the category that is passed in as a parameter, for the `datetime.date.month` object at the same index in the previous list.

Similarly, we organized the data from the `covid_data.csv` dataset into a list of `CovidData` instances. A helper function is used to convert the `date` column into a `datetime.date` object. The function, `get_monthly_cases`, returns a list of `CovidData` with only the data for the last day of each month for a particular year for a particular location. A helper function, `check_if_monthly_case`, is called in the `get_monthly_cases` function. This helper function checks to see whether the `CovidData` instance is data for the last day of the month. Lastly, `covid_data_to_dict` returns a dictionary mapping relevant attribute names of `CovidData` to the list of instance attributes for all the `CovidData` instances in the input list to be processed with `pandas`. One change to be noted is that instead of the total number of cases per month, we recorded the number of new cases per month in the dictionary instead to analyze the change in cases per month as related to the crime. Finally, although the `_num_deaths` attribute is defined in the `CovidData` class, it is unused because we decided to focus on cases rather than deaths. We still decided to read and save this column of data for potential future analysis of the pandemic.

Since the relationship between COVID cases and crime rate in each province/territory is different, we separated out data for each province/territory for both crime and COVID data sets as described above, and put them into separate Pandas Data Frames (in the `pandas` library). We also compiled a result from the dataset for Canada as a whole as described above as the main set of data that will be extrapolating results from. We also categorized the specific crimes into different categories (*physical*, *non-physical*, *public*, *private*) with keyword filtering, and processed each type of crime separately in order to provide more detailed conclusions and insight. After trimming our data, we plotted our data onto a line graph with `matplotlib`, to show the amount of a certain type of crime committed in Canada over time, combined with current active covid cases over time. This is an overview of our data which we extracted from our data sets.

Then in order to find correlations (or lack thereof) between different types of crime and COVID cases, we used `numpy.polyfit` to find and `numpy.poly1d` to plot the line of best fit onto the scatter plots that have the number of COVID cases as the explanatory variable and the amount of a certain type of crime as the dependent variable. We are also using `numpy.corrcoef` to find the correlation coefficient (r-value) of the scatter plot to determine the relationship. And in order to show our different attempts and models visually, we used `matplotlib.scatter` and `matplotlib.plot`.

Finally, in order to present our findings visually in a GUI, we used `matplotlib.pyplot` and `Tkinter`, which allows us to present our graphs both visually and interactively. In total, we have 5 graphs as part of our final product. The first graph contains COVID data per month from early 2020 till now as well as crime data for the four different types of crime we analyzed: physical, non-physical, public, and private crime, which is an overview of all the data that we have. The next four graphs contain scatterplots that display the correlation between COVID and each of the four types of crime, and using statistical analysis, we found the correlation for each, the results of which will be presented in the Discussion section.

## 4 Instructions

In order to download the two necessary datasets used in this project, navigate to the first two links listed in the references below. The first link should yield the crimes dataset, where the `csv` file can be downloaded. The second link should yield the COVID-19 dataset at the very bottom left of the page. The file name should be `covid19-download.csv` for the second dataset. The crime dataset file should be renamed to `police_data.csv` and the COVID dataset file should be renamed to `covid_data.csv`. If this doesn't work, navigate to this Github repository to download the files:  
<https://github.com/rsrinivasan1/Datasets-for-CSC110-Project>

After downloading both files, they should be placed in a directory titled `data_sets` so that they can be accessed by our project. You can run our project by running the `main.py` file.

In the pop-up windows that shows up, there should be five buttons, each of which correspond to a graph of our final data results as described above.

## 5 Changes since Proposal

For the most part, we stuck to the project plan outlined in the proposal for the final submission. However, we did encounter some issues with graphing the data once we aggregated it into `pandas` DataFrames. We wanted the graphs to be more interactive and as a result had to use `Tkinter` in combination with `matplotlib` to show our visual representations and compare the crime data with the COVID data. Throughout the process, in fact, we realized that some of the libraries that we initially planned to use did not have all of the functionality necessary, and as a result had to continuously update our plan with new libraries such as `Tkinter` and `numpy`. We also realized that what we really wanted to compare between the two datasets was the change in new COVID cases per month as related to the crime per month. As a result, we instead decided to look at the differences in the number of cases per month and compare those to the crime data.

## 6 Discussion

Based on the scatter plots of the amount of different types of crime in relation to the number of covid cases, we successfully drew the following conclusions on the relationship between the COVID and crime situations in Canada:

(1) The total physical crime in Canada has a medium-strength negative correlation with the amount of covid cases. This means as covid cases increase, the amount of physical crime in Canada decreases by a fair amount. This is likely due to people having less chances of physical contact while distancing restrictions are being enforced. This result is within our expectations.

(2) The total non-physical crime in Canada has almost no correlation with the amount of COVID cases. This means as COVID cases increase, the amount of non-physical crime in Canada is not affected. This is likely due to distancing restrictions having little to no effect on people's willingness and capability of committing crimes that require no physical contact. This result is within our expectations.

(3) The total public crime (i.e. crimes committed in public, or to strangers) in Canada has a

low-strength negative correlation with the amount of COVID cases. This means as COVID cases increase, the amount of public crime in Canada decreases somewhat. This is likely due to people having less chances of interacting with strangers while distancing restrictions are being enforced. This result is within our expectations.

(4) The total private crime (i.e. crimes committed in private, or to victims that know the perpetrator) in Canada actually has a medium-strength negative correlation with the amount of COVID cases. This means as COVID cases increase, the amount of private crime in Canada decreases by a fair amount. This result falls out of our expectations and needs to be further examined.

During this project, we've encountered numerous obstacles, especially when filtering and cleaning the data sets that we have and transforming it into data types that we can use for analysis. One of the greatest challenges was that we split our team into two groups, one working on the data wrangling and the other working on the data displaying. Working on this project meant that both groups had to constantly coordinate with each other. Also, the fact that the two data sets that we used recorded their data so differently meant that we had to keep track of many details to guarantee that the final Data Frames contain the correct data. For example, keeping track of the many helper functions and what they do respectively was a great challenge, especially since the people working on the data analysis were not the same people working on data displaying. In addition, when it comes to displaying our data visually, we had no experience with doing so previously, so we had to learn by experience and experimentation, trial and error, which took a substantial amount of time.

The next step for our project is to figure out why we found the correlations that we did, such as why the amount of private crime in Canada has a negative correlation with the amount of COVID cases. This may require some additional research as we likely do not have all of the data necessary to completely understand why these trends occur. In the future, we would also want to separate each province/territory out and analyze them separately, to see if the specific circumstances in each province/territory have impacts on relations between COVID and crime there specifically. For this project, although we did split the data into the different provinces of Canada, we only found it relevant to examine the data for Canada as a whole to analyze the relationship between COVID rates and crime.

## References

- [1] Selected Police-Reported Crime and Calls for Service during the COVID-19 Pandemic, Government of Canada, Statistics Canada, 22 Oct. 2021,  
<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510016901&pickMembers%5B0%5D=1.20&cubeTimeFrame.startMonth=01&cubeTimeFrame.startYear=2019&cubeTimeFrame.endMonth=10&cubeTimeFrame.endYear=2021&referencePeriods=20190101%2C20211001>
- [2] Public Health Agency of Canada. "Covid-19 Daily Epidemiology Update." Canada.ca, 28 May 2021,  
<https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html#tiles>
- [3] Global Affairs Canada. "Covid-19 Travel: Checklists for Requirements and Exemptions." Travel Restrictions in Canada, Travel.gc.ca, 25 Oct. 2021,  
<https://travel.gc.ca/travel-covid/travel-restrictions/exemptions>
- [4] Public Health Agency of Canada. "Government of Canada." Canada.ca, / Gouvernement Du Canada, 5 Nov. 2021,

<https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/epidemiological-economic-research-data/mathematical-modelling.html>

- [5] “Public Health Infobase - Data on COVID-19 in Canada.” Open Government Portal, Public Health Agency of Canada,  
<https://open.canada.ca/data/en/dataset/261c32ab-4cfd-4f81-9dea-7b64065690dc>
- [6] “Covid-19 Intervention Timeline in Canada.” CIHI, Canadian Institute for Health Information,  
<https://www.cihi.ca/en/covid-19-intervention-timeline-in-canada>