Vital Events Canada and the Impact of a Public Health Crisis

Katie Schilling

501130072

Supervisor: Tamer Abdou, PhD

Table of Contents

[Abstract - Overview of Data 3](#_Toc99959599)

[Glimpse of Cleaned Data Sets 3](#_Toc99959600)

[Vital Events 3](#_Toc99959601)

[Covid 3](#_Toc99959602)

[Combined data sets for Analysis 3](#_Toc99959603)

[Literature Review 4](#_Toc99959604)

[Research Questions 7](#_Toc99959605)

[Methodology 7](#_Toc99959606)

[Visualizing the Data 10](#_Toc99959607)

[Variables overview analysis 10](#_Toc99959608)

[Exploratory Data Analysis 12](#_Toc99959609)

[Vital Events EDA 12](#_Toc99959610)

[Covid-19 EDA 13](#_Toc99959611)

[Combined Vital Events & Covid EDA 14](#_Toc99959612)

[Prediction Modelling 15](#_Toc99959613)

[Initial Predictions 15](#_Toc99959614)

[Prediction Modelling Analyzed against introduction of COVID-19 16](#_Toc99959615)

[Causality Analysis 17](#_Toc99959616)

[Usefulness of Data Analysis: Real-world application 19](#_Toc99959617)

[Github repository Link 20](#_Toc99959618)

[References 20](#_Toc99959619)

# Abstract - Overview of Data

Every year the registrar general collects and releases information on vital events in Canada. This information includes marriages, births, stillbirths, and deaths. Based on my research, the vital events data is collected to monitor, and predict future population growth, decline, or stability.

Much of this information is used to predict and allocate resources in anticipation of what the trend analysis is showing. It allows sufficient resources to areas such as health care, housing, and financial support.

In addition, this information is used to monitor for events such as public health crises. In March 2020, this was an event we experienced with Covid-19. While exact numbers, across any data collection is nearly impossible, and predicting this public health crises may not have been possible, with the introduction of the daily number of Covid-19 positive cases, we may be able to re-analyze the future trends and the impact Covid-19 had on vital events

## Glimpse of Cleaned Data Sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Births | Marriages | Deaths | Stillbirths |
| 1994-01-1 | 11631 | 2078 | 8094 | 75 |
| 1944-02-1 | 11254 | 2650 | 6428 | 62 |
| 1944-03-1 | 13003 | 2557 | 6503 | 73 |
| 1944-04-1 | 12576 | 3967 | 6224 | 74 |
| 1994-05-1 | 13240 | 6493 | 6483 | 67 |
| 1994-06-1 | 13072 | 7754 | 6187 | 66 |

### Vital Events

### Covid

|  |  |
| --- | --- |
| Date | Covid |
| 2020-02-1 | 11 |
| 2020-03-1 | 8533 |
| 2020-04-1 | 44688 |
| 2020-05-1 | 37711 |
| 2020-06-1 | 13257 |
| 2020-07-1 | 12108 |

### Combined data sets for Analysis

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | Births | Marriages | Deaths | Stillbirths | Covid |
| 1994-01-1 | 11631 | 2078 | 8094 | 75 | 0 |
| 1944-02-1 | 11254 | 2650 | 6428 | 62 | 0 |
| 1944-03-1 | 13003 | 2557 | 6503 | 73 | 0 |
| 1944-04-1 | 12576 | 3967 | 6224 | 74 | 0 |
| 1994-05-1 | 13240 | 6493 | 6483 | 67 | 0 |
| 1994-06-1 | 13072 | 7754 | 6187 | 66 | 0 |

# Literature Review

Vital events are a key piece of information used when making decisions when it comes to infrastructure, resources, housing, health care etc. Understanding, and being able to determine trends, future population growth and declines, can allow officials properly, or reasonable allocate the resources needed to handle whatever the situation might be. Where we may see an increase in deaths, births or still births, we may find a need for added funds, and resources allocated to hospital and health care sectors overall, whereas if the trend is the opposite, the resources can be allocated elsewhere. Vital event statistics for Births, Stillbirths, deaths, and marriages are the responsibility of each province and territory. Working in collaboration with Stats Canada, alongside each province and territory, as well as vital statistics registrars, vital events data has been complied, analyzed, and published since 1921. Each statistic sheds light on its ability to trend, forecast and use the data to understand how its impacting public health, and resources. Birth statistics are used to monitor the number of new births, family sizes, as well as for target prevention programs for women who are at a high risk of having a baby with low birth weight, as well as gestation period.

Death statistics can provide an overall outlook on population trends, but it can also be a good statistic to monitor trends in public health, such as a health crisis, infectious diseases, unintentional injuries, and suicides.

Common public data regarding my second dataset, which is covid-19 positive cases, is popular, easy to find, and multi-faceted. On a single search you can find articles that compare vaccinated against unvaccinated, deaths, as well as deaths by vaccination status. Hospitalizations based on vaccination status, ICU admissions, again over all as well as by vaccination status. Age groups for all of these variables. The current case counts as well as predicted best-case and worst-case scenarios. Projected outcomes based on public health measures. Visualizations of data for each year, month and day, as well as even more knowledge regarding the different variants, rate of transmission, positivity rates, and even positivity rates by regions/postal code. There is no shortage of information when it comes to both topics. Almost too much information, making it hard to digest, compare, contrast, and fully understand the broader picture. Currently, it does seem as we are through the worst of Covid-19. While the trends show the case counts, and waves getting larger each time, the mortality in each wave, which also seems to coincide with a new variant, seems to be decreasing. It would seem at this point we are in more of a stable position than we were a year or 2 ago.

In my opinion, the largest goal of understanding vital events data and trends is to always ensure we are prepared. As we have seen over the past 2 years, with each wave of Covid-19, our hospitals, and other health care settings have been on the brink of collapse. Each wave created a need for increased help, and then each subsequent wave, the current resources were inefficient to keep the system from on the brink of collapse. Had sufficient funding and resources been allocated, and hospital infrastructure increased prior to the arrival of Covid-19, both our hospitals, and our healthcare workers would have been in a better position to manage the public health crisis. Additionally, with increased births, we also see an increased number of resources, primarily funds, being sent to middle- and lower-income families to help them support the financial burden of supporting children. The proper resources allocated for Child Tax Benefit payments, and funding for the subsidized daycare program for all provinces also helps to lower the burden in supporting children. Middle to low-income families often spend a long time on a waiting list for subsidized daycare funding to become available. This means many pay out of pocket or are unable to work due to the cost of childcare, the result is then the income level for those families shrinking, and the burden on the government greater as they rely on funding such as welfare to manage day to day expenses. Recognizing, and addressing an increased need for childcare, or the high rate in which the cost of living is increasing, could have created an opportunity to allocate necessary funds and resources to attempt to offset this issue, prior to it becoming a much larger issue. When looking at the structure and allocation of funds to each province for health care, there is a lot of ups and downs. There are increases, and the reverse course situations, where the total funding going to each province is adjusted and reduced. When looking at the cash component of the CHT (Canada Health Transfer), from 2013-14 to 2016-17 years, there was an annual growth of 6%, however from 2017-18 to date, the annual increase had fluctuated between 3% and 4.6%. Looking at these numbers, and the health crisis and pressure we have put on our healthcare systems and wondering if the trend in vital statistics coincides with the decrease in CHT to provinces for their healthcare infrastructure, or if we were in a position to be able to better prepare our healthcare system by allowing them to increase resources while not in the middle of a health care crisis.

Looking at the analyses I have found, I have been unable to find an analysis that has been done with the same information. There are many analyses done on the Vital Events data, primarily based on the provinces specific data rather than data for Canada as a whole, and many analyses done on Covid-19 in reference to many different faucets of vitals events, along with many other statistics, there does not seem to be one that encompasses both. A lot of the data used is more enriched in features than the broad information. Many of the features analyzed are breakdowns, or subgroups of specific types of data such as birth weights, marriages, divorces, death by cause, ages, sex, province, city and many more variables. Finding a single article or analysis that looks at the broader picture of the data, prior to breaking it down into smaller sub-groups and analysis was not something I was able to accomplish, nor was I able to find one that took into account the Covid-19 positive cases, or the vital events as a whole alongside any covid-19 data.

In my research I have found many related analyses. One of which is comparing the Covid-19 pandemic with preterm and stillbirth rates using historical data for stillbirths at term and preterm gestation. Methodology used for the analysis utilized Laney control P` charts, and interrupted time-series analyses for the most recent 4 years. The overall results were that from January 2020- December 2020, the preterm birth rate in Ontario was 7.87% with no special cause variation, resulting in the interpretation that overall, there was no unusual change, overall or by subgroup, during the first 12 months of the pandemic vs the previous 17.5 years. I was also able to find a government produced population projection for the province of Ontario. Within this population projection, it analyzes what the actual growth was, vs. the projected growth and then looks at future growth, all the way up to 2046. Currently, based on the impact of Covid-19, they are projecting the growth to drop from 1.3% for 2019-2020 to 0.6% for 2020-2021, before increasing again to roughly 2.1% for the 2021-2022 year. From that point on, it is projected that the population increase will have a steady growth, reaching 1.0% by 2046. From what I gather based on this information and the historical trend line, it does appear that it will take approximately 25 years for the province of Ontario to get back close to the 1.3% increase in 2019-2022. I’m not entirely sure what economic or housing impact that growth, or lack thereof will have, but with the large proportion of Ontario’s population being within the 15-64 age bracket, we have a primarily working class, younger population, which may lead to a higher increase in population, and considerations may need to be taken into account on the impact of this, with the vital events data for all of Canada, and the allocation of funds and resources to support a potentially fast rate of growth.

From my research, I am seeing a lot of analysis on Covid-19 itself. Analysis on the number of positive cases, number of deaths, hospitalizations, both in ICU and regular hospitalizations. I am also seeing analysis of the number of vaccinated, and unvaccinated. There is a lot of data analysis done around Covid-19 itself. While I think it is important to look at the impact Covid has had on our country, I also think it’s important to take those numbers into account, as well as the other Vital statistics we track every year, to see how the overall image has changed with the introduction of covid-19, vs where we anticipated seeing these trend lines go over time. While the data may seem daunting to look at individually, I do feel that the larger picture, of comparing covid cases against our other vital stats, it may paint a picture that does not paint as harsh of a picture as just looking at the covid numbers. We’ve put a lot of emphasis on vital statistics, and on pandemic statistics, however, combining the 2, and comparing for causality, trends, forecasting and time series analysis, we can see the actual impact on these stats.

While researching to see if anyone has completed an analysis the same as mine, or similar, I have found that there are not a lot of articles, papers, and analysis to use in comparison. There are a lot of time series datasets, and analysis. There are many options for each province or territory. Many considers more robust variables such as province/territory, age, sex, weight, cause of death etc. These variables are highly valuable if you want to find trends within each area. Such as deaths. When analyzing deaths by location, age, and cause, it allows you to see trends in certain causes of death, which age groups are being most affected, and whether it is affecting men or women more, or both equally. When looking at a specific area of the Vital events dataset, this could be a great asset, however I was unable to find a trend analysis, or time series forecast of the vital events data provided by Stats Canada. It may be that the data itself is too broad, and those who are researching are looking for more refined results, however I do believe looking at the analysis of the broader variables, before breaking them down into subgroups, can help to identify which variable to look at first. If there is a trend that’s alarming, rapidly changing, or even not changing at all like historical data, it may be a clue that there is something going on there, and it warrants a deeper look into what that might be. I feel the analysis I am doing, is beneficial, as in the larger scheme of things it’s important to not only look at the finer details, but to also be able to step back and see the big picture. When looking for research materials, there were a lot of articles that came up with information regarding Covid-19 and deaths. While I think it’s an important statistic, I also think its important to step back, and look at deaths as a whole. While we may have seen a large increase in deaths due to Covid-19, deaths attributed to other factors and health conditions may have decrease, essentially continuing the trend from historical data, rather than seeing a large jump in deaths. From reading articles regarding Covid-19 and deaths, you get this impression that we have our usual annual deaths, and a large spike in deaths from covid, which leads to the assumption that, we have had a drastic increase in deaths. In my analysis, I will be able to see deaths, compared to the number of covid positive cases, and see if there is a direct correlation between the 2. Determine that if we know the value for one, we can use it to better forecast the other.

# Research Questions

Are there any alarming trends, or changes in variables that show an impact on either the growth or decline of the vital events, that are apparent when visualizing the data?

When analyzing the vital events data, using forecast for time series, what is the trend, prior to adding in the occurrence of COVID-19?

When looking at the overall data, and plotting it as a time series, is there a trend of steady decline in stillbirths, as our knowledge and medical advancements are made?

Based on history, what numbers did we expect to see in 2020 and 2021? How did those numbers differ with the introduction of COVID-19?

Is there any causality between the variables, when compared both individually to one another, and when comparing one variable to the remaining?

What does this analysis tell us about the data, and its ability to predict future trends, when all the data is used in conjunction to one another?Was there any indication that there was an impending public health crisis emerging?

# Methodology

When selecting my topic for this project, I wanted to use data that either has been not well analyzed before or is new data. With the two datasets selected, my logic is to determine if our vital events data as reported is sufficient for predicting future trends, and allowing the proper allocation of resources, while additionally introducing the existence of a public health crisis.

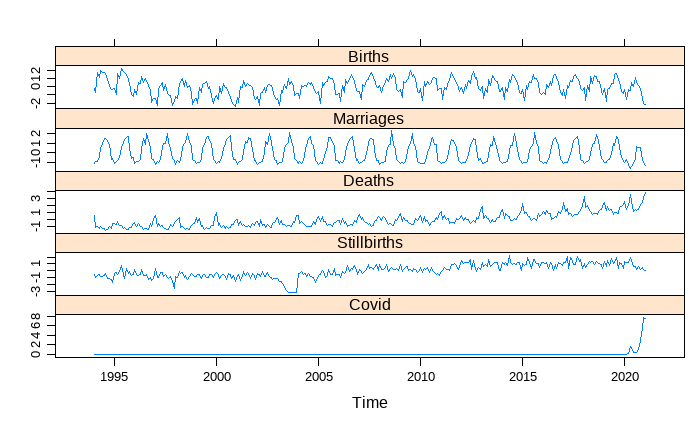
The population of this dataset is all of Canada, it is a very broad population. However, it also provides a lot of information. With a larger quantity of information, it can help to show a larger picture. A smaller population or group may provide a more targeted analysis, I wanted to understand the data on a broader scope. As well as the impact of Covid-19 on a broader scope. With the level of attention, publicization and emphasis placed on Covid-19 and its impact on our daily lives, most specifically deaths, I wanted to see if when analyzed, if the data correlated with the information publicly available. There is no selection or sampling method for this data.

Data Collection for both datasets used are the same. The information in the Vital Events dataset is obtained from the official records of live births, stillbirths, deaths and marriages. Additionally, the number of divorces is also collected though not included in the vital events dataset. The official recording of these events is the responsibility of each province and territory. Due to being official recordings and considering the possibility of an error on input, it is fair to assume the data itself is unbiased. It is not based on any information other than what is provided and is not intended to skew the analysis of the data one way or the other. Once recorded, Statistics Canada, in collaboration with the provinces and territories registrars, compile, analyze and publish the data for the public. The collections, compilation and analyzation of this data has been done since 1912.

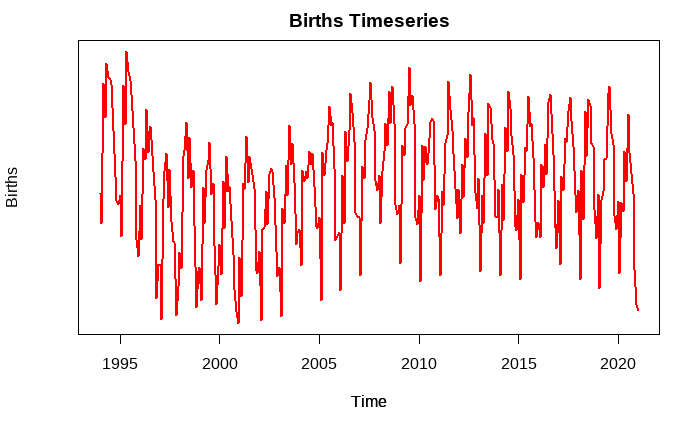
For my analysis, I will be using Granger Causality test in 2 forms. One to analyze a single variable against all other variables, and one to analyze each variable against a single other variable to see if there is a more targeted causality. Additionally, I will be looking at the data as a timeseries. Understanding the pattern and seasonality of the time series, visualizing the time series, and inferring speculation based on the visualization of the timeseries.

I will also be running a prediction model on the time series, prediction from 2020-2022. The purpose of this is to predict what the values should have been for the vital events without the introduction of Covid-19, and then comparing the prediction values to the actual values to see the level of variance and what possible role covid-19 played in the variance.

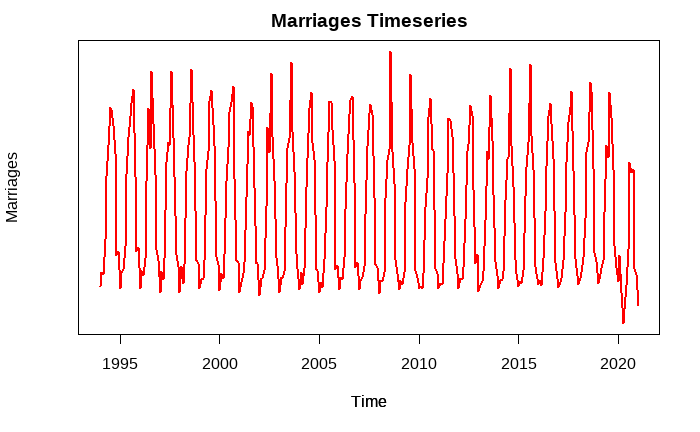
# Visualizing the Data

The visualization of the time series data shows a consistent trend for all variables, except for Covid-19 and stillbirths. Stillbirths are, in general, unpredictable, but stable in their movement, whereas Covid-19 was unexpected, so the results show a lot of volatility. When looking at births, deaths and marriages, there is a level of consistency, however a trending increase in deaths over the past 10 years can be seen. In this visualization, you can start to see the variances in 2020-2021, where the introduction of Covid-19 began. In addition, in this visualization, we can see that not only are we not seeing a decrease in stillbirths over time, and as medical advancements are being made, but rather we have been at a higher rate since approx. 2005 and are seeing a slow and steady increase. This may be since births are seeing a slow increase as well. As more births occurs, the likelihood of a stillbirth occurring, and occurrences of stillbirths increase with it. The risk of it happening appears to remain the same, the upward trend is accounted for by the upward trend in births. My conclusion would be that we are not seeing a decrease in stillbirths with medical advancements.

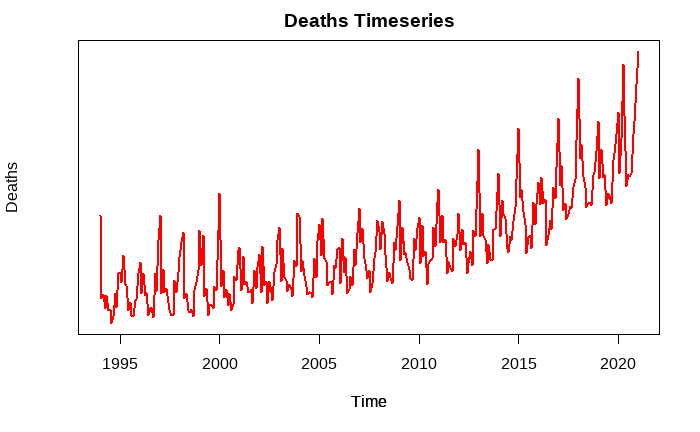
# Variables overview analysis

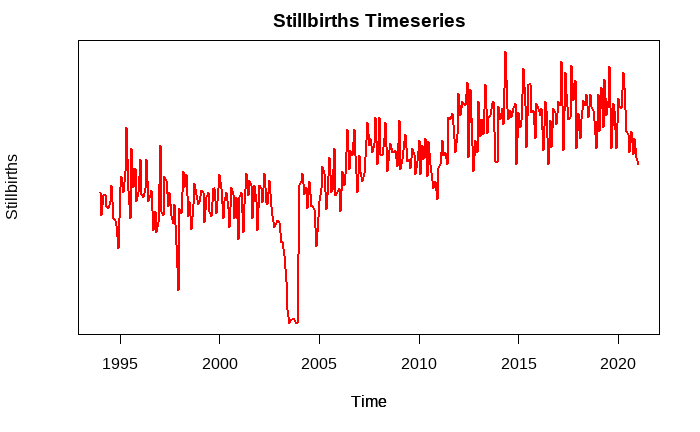
Understanding the patterns of these variables on their own is important being able to see changes or factors that influence the growth or decline of these variables. Primarily I am looking for any trends that show significant change or may have been an indication that there was an underlying, unknown cause for change, or the ability to possibly predict the emergence of a public health crisis, prior to its discovery.

Starting with births, we can see that the trend is stable. There are regular upward trends, followed by a downward trend, the increase and decrease of these levels do vary slightly. We can see from approx. 1988-2005 the upward trend was lower, creating an even lower downward trend then previously, and then from 2005-2020, the trend was stable. We can visually see a change in the pattern in 2020 with a rather sharp decline.



Marriages, like births is stable, there seems to be no indication of any changes in the pattern, or what is expected over time, except for 2020. Again, here we can see the upward trend was cut short, followed by a sharp decline. This would lead to the conclusion there was an outside factor that caused a drastic change in the pattern. Without any knowledge of any existing outside factors, this should lead to further investigation as to why this occurred.

When looking at deaths on its own, we can clearly see a slow and gradual increase of deaths. This can reasonably be explained by the growing population. As there are more people in existence, the number of deaths would increase accordingly. The pattern itself, and steady incline look stable, and one could reasonably predict the following years numbers based on historical data. In 2020, like the other variables we see a sharp change. In this case a sharp increase. However, if you look at the end of 2019, the pattern does not stay the same. There is a sudden reverse. Based on research, in Canada, we did not see the presentation or peak of Covid-19 until March 2020, however, there appears to be a increase in deaths at the end of 2019, beginning of 2020, that if analyzed, may have indicated an underlying factor that is causing the disruption of the pattern. While the public health crisis, Covid-19, may not have been able to be predicted, paying close attention to vital events numbers, primarily deaths, on a month-to-month basis, may have allowed for a quicker response, or an indication of a public health crisis emerging.

When looking at Stillbirths independently, we can see that overall, this is an extremely hard variable to predict. While it appears fairly involatile in its movement, we can see where there are periods of extreme volatility. This may be since there are so many underlying issues that can cause Stillbirths, along with no reason at all. While logically I feel this is something that we should be able to analyze, and decrease over time as we make medical advancements, the reality is this may be something we will never be able to fully understand and prevent in the future. From approx. 2012-2019 the pattern seemed relatively stable, but we do see a sharp decline in 2020. When compared to births, which we also saw a sharp decline, this makes sense. Less births leads to less still births.

# Exploratory Data Analysis

## Vital Events EDA

When looking at the vital events dataset, all variables are balanced, with only some slight variations to the balance for marriages and deaths. There does appear to be either an outlier, or a larger than average amount of marriages on the left side of the graph, but it does not seem to affect the normality of the data itself.

Chart, histogram

Description automatically generatedChart, bar chart

Description automatically generatedWhen looking at the correlation plot, we can see there is a strong relation between deaths and years, meaning that as the years pass by, the number of deaths seems to grow with it. The next strongest correlations are stillbirths when compared to years, and stillbirths when compared to deaths. On the opposite side, deaths and marriages seem to have a negative correlation

Graphical user interface, application

Description automatically generated

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

## Covid-19 EDA

Covid-19 was an unexpected health crisis, therefore the data it provides would not have a normal shape. There were no reported cases prior to 2020, therefore it took a sharp upward swing. However, when looking at the origin Q-Q plot, it does seem to have some level of normally to it, but overall, it is skewed to the right which was not a surprise in this situation. There also is no correlation plot for this dataset, as there is only the one value.

Chart, bar chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Graphical user interface, application

Description automatically generated

## Combined Vital Events & Covid EDA

Chart

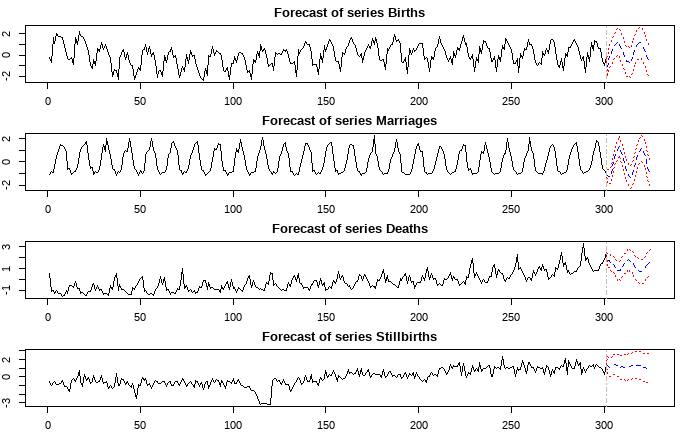
Description automatically generatedA picture containing timeline

Description automatically generatedGraphical user interface, text, application

Description automatically generated with medium confidenceWhen combing the two data sets together, and running a further EDA, the results we get are not unexpected. What I was really looking for in the analysis, was the correlation matrix to further understand how Covid-19 plays a role when compared to the other variables. We can see by the correlation matrix below that there is a negative correlation with births and deaths, almost no impact or correlation with stillbirths, and a positive correlation with deaths.

# Prediction Modelling

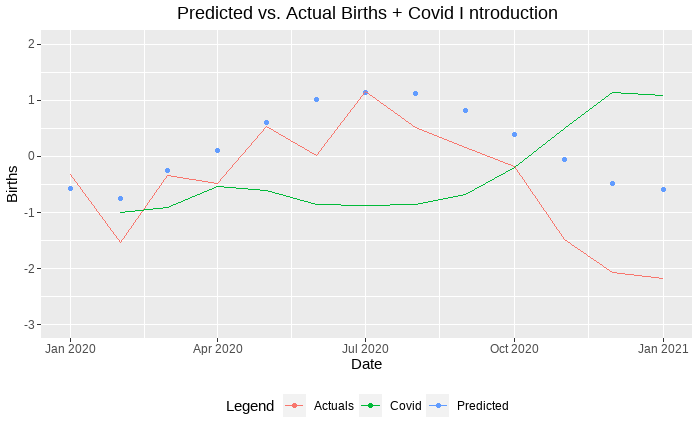
## Initial Predictions

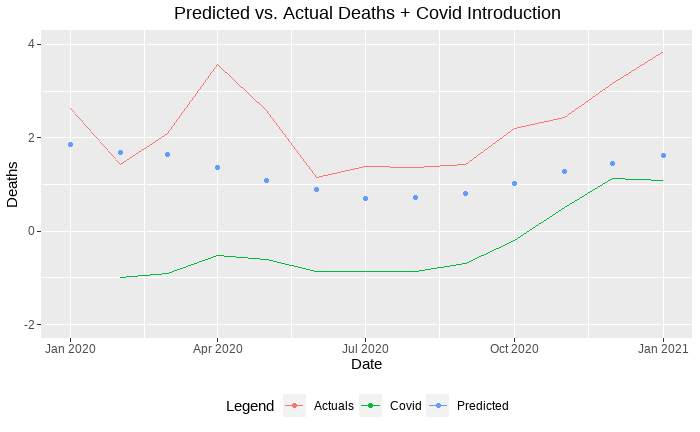
When first looking at the vital events data, I wanted to understand what the data showed, without any outside factors. Whether there was a consistency to the trends, and what the future predictions would be based on historical data. My initial forecast which included the predicted, and then the upper and lower ranges for that prediction, and the confidence intervals produced the following graph.

When looking at just the graph, we can see that Births and marriages look very similar. Because the historical trend is consistent, it shows a consistent forward trend, however, with any prediction, the further you look into the future the more unstable the prediction becomes. The initial confidence interval for births started at 0.93, and slowly increased with each prediction, the final one being a confidence interval of 1.5. This is consistent with all variables, although when looking at both deaths and stillbirths, the variance in the confidence interval is more predominant as they have less consistent historical data, making it harder to more accurately predict future values. Based on this, I would assume that only the prediction for the first 12 months would be reasonably reliable, and that predicting 24 months in advance would result in data with high variance. This may pose a challenge in terms of using this data to predict future needs, as being able to predict further into the future, allows more time to analyze, and allocate resources as needed, ensuring the stability of various factors such as heath care needs, housing needs, and financial assistance.

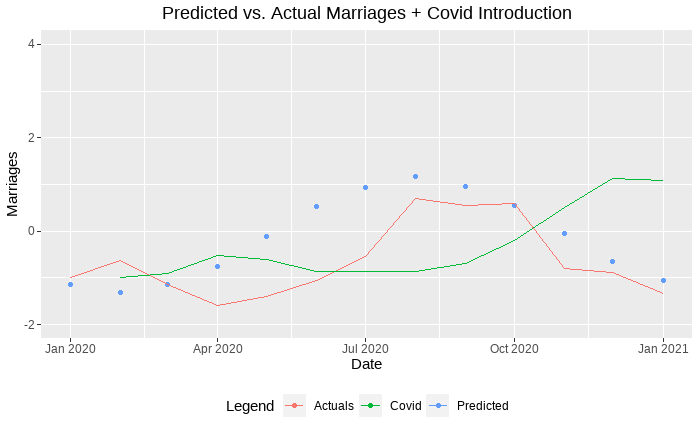
## Prediction Modelling Analyzed against introduction of COVID-19

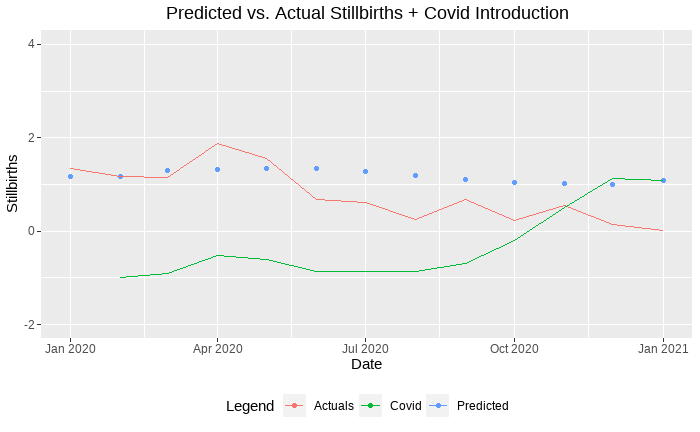
Next, I analyzed each variable, the actual values, the predicted values, and the values of COVID-19 for the same time frames. Analyzing these values together, shows a clearer picture of the predicted values vs. the actual values, as well as understand how COVID-19 may have created a greater variance between the 2. These visuals not only helped me understand the difference between the predicted and actual values, but see the significance the introduction of COVID-19 had on the results

Starting with births, we can see that the predicted values vs. actual values for 2020-21 start out relatively stable, however as we see the number of COVID-19 cases increase, we also see the variance between predicted and actual grow. This would indicate that there is some causality between births and covid, as on the surface it appears COVID-19 has some impact on the number of births.



When looking at deaths, predicted vs. actual values we can see that the variance starts off relatively stable, and then there is a large variance when there is the first slight increase in COVID-19. When COVID-19 stabilizes, so does the variance between actual and predicted values. The further we look, as COVID-19 numbers increase so does the variance between predicted and actuals. Again, this would assume some causality between deaths and COVID-19.

Next, when looking at Marriages, the variance between actuals and predicted values seems to be stable. Even with the introduction of COVID-19. While the initial introduction of COVID-19 does seem to play some part in the variance, as you look towards the end of the chart where COVID-19 numbers are the highest, there is vary little variance between the predicted and actual values. This would indicate that initially there may have been some causality between marriages and covid-19, or indication that one might affect the other, it stabilized and no longer appears to be a factor.



Finally, when looking at Stillbirths actual values vs. predicted values, this is the most stable variable. The variance between predicted and actual is minimal in comparison and shows no indication that there is causality between stillbirths, and COVID-19.

# Causality Analysis

When looking at causality, I took a multipronged approach. I wanted to understand the greater picture, as well as a more targeted picture. First, I looked at overall causality. One variable, compared to all others. The result of this was:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Granger Test | Deaths | Births | Marriages | Stillbirths |
| F-Test | 5.7512 | 9.293 | 14.545 | 4.1624 |
| Df1 | 18 | 18 | 18 | 18 |
| Df2 | 1084 | 1084 | 1024 | 1084 |
| p-value | 2.627e-13 | 2.2e-16 | 2.2e-16 | 1.407e-08 |

The results shows that there are no granger causalities when comparing one variable to all other variables, meaning that one does not influence or indicate the results of all other variables. Additionally, this test also showed that there was no instantaneous causality between the selected variable when compared to the other variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Granger Test **Deaths** | F | Pr(>F) | Significance |
| Deaths ~ Births | 7.4423 | 5.739e-12 | \*\*\* |
| Deaths ~Marriages | 6.7127 | 1.16e-10 | \*\*\* |
| Deaths ~ Stillbirths | 2.1829 | 0.01257 | \* |
| Deaths ~ Covid | 5.9415 | 2.869e-09 | \*\*\* |

|  |  |  |  |
| --- | --- | --- | --- |
| Granger Test **Marriages** | F | Pr(>F) | Significance |
| Marriages ~ Births | 4.2563 | 3.291e-06 | \*\*\* |
| Marriages~ Deaths | 9.9338 | 2.624e-16 | \*\*\* |
| Marriages ~ Stillbirths | 2.9507 | 0.000677 | \*\*\* |
| Marriages ~ Covid | 22.364 | 2.2e-16 | \*\*\* |

Next, I wanted to look at each variable, compared to all other variables one by one, to see if when compared one on one, if any of the variables has a causality with another. These results showed a different picture than the above results. When looking at the causality analysis for each variable, against one of the other variables, the number I am paying most attention to is the P-Value. In this scenario a P-value above 0.05 indicates we can reject the null hypothesis that X does not cause Y, and a P-Value of 0.05 or lower indicates a stronger significance, and we reject the null hypothesis. Based on this information, we can see more targeted levels of causality in variables. Most notably looking at Marriages, all variable when evaluated against it had significance, that knowing marriages, can help predict the values of the other variables. All variables had some level of significance, but what I found most intriguing is the lack of significance between Covid-19 and the other variables. Overall, there was little to no significance. When looking at other variables and their ability to predict Covid, deaths and marriages had the most significance, followed by births. There appears to be no major impact on stillbirths.

|  |  |  |  |
| --- | --- | --- | --- |
| Granger Test **Births** | F | Pr(>F) | Significance |
| Births ~ Deaths | 4.8443 | 2.82e-07 | \*\*\* |
| Births ~Marriages | 9.9338 | 2.624e-16 | \*\*\* |
| Births ~ Stillbirths | 2.7585 | 0.001439 | \*\* |
| Births ~ Covid | 2.7104 | 0.001734 | \*\* |

|  |  |  |  |
| --- | --- | --- | --- |
| Granger Test **Covid** | F | Pr(>F) | Significance |
| Covid ~ Deaths | 1.6482 | 0.7796 | . |
| Covid ~Marriages | 1.803 | 0.04714 | \* |
| Covid ~ Stillbirths | 1.6179 | 0.08576 | . |
| Covid ~ Births | 2.7104 | 0.7079 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Granger Test **Stillbirths** | F | Pr(>F) | Significance |
| Stillbirths ~ Births | 4.3946 | 1.85e-06 | \*\*\* |
| Stillbirths ~ Marriages | 4.5966 | 7.962e-07 | \*\*\* |
| Stillbirths ~ Deaths | 4.63 | 6.926e-07 | \*\*\* |
| Stillbirths ~ Covid | 1.4109 | 0.1597 |  |

When looking at Covid, and its causality to the other variables, they are all borderline. This may be due to insufficient data for the analysis as there is more historical data for the other variables. I would not entirely reject the null hypothesis in this case, rather look more closely, or re-analyze the information once there is more data available for Covid.

# Usefulness of Data Analysis: Real-world application

When analyzing the vital events data, in a real-world application, it is easy to confidently predict the following years patterns when it is just the vital events. Whether there will be an increase or decrease in the numbers, and what resources will be needed such as housing, financial support, medical resources etc. Based on review of the CHT program, which is the Canada Health Transfer, the allocated funds to provinces and territories declined in the years between 2019-209, which one would assume means there was less of a burden on the health care system resulting in requiring less funds. From 2019-2020 there is a sharp incline which one could equate to the introduction of Covid-19, and the increased burden on the health care systems across Canada.

When analyzing Ontario’s population projection, we can also see that when last updated in 2020, prior to the major occurrence of Covid-19, there was a projected decrease in population growth as well as net migration. Looking at the chart, without the raw data to analyze with it, it is hard to tell if this chart had indicators that included Covid-19, or if it could have been a precursor to realizing that the numbers were dropping and that there may have been an outside factor at play.

During my analysis, I did not find any indication, other than an increase in deaths in late 2019, to indicate to presence or impending public health crisis. While the data itself did not seem to provide us with the tools to predict Covid-19, the data, combined with Covid-19 cases, in the future may allow us to see the pattern that exists prior, and predict future public health measures. While there does seem to be some level of direct causality between one variable when compared to another, there does not appear to be an overall causality. When looking at the broader picture, all variables run their own course, and many correlate with one another, such as, when marriages increase, births increase as well. Deaths seem to follow a similar pattern alongside births and deaths, however, there seems to be a greater number of deaths occurring, and at a larger rate as time progresses.

I think future research and analysis would be needed in this area, to better understand how we can predict a public health crisis or a change in trend. I know there are global initiatives and databases that track certain information, but I think on a local or national bases rather than global, it may allow for a more targeted approach, and once detected, can assist in providing knowledge on a more global level. However, with how the data is submitted, and the fact that the vital events data is monthly totals rather than daily, it can be a challenge to predict it sooner. If examined, and daily numbers uploaded to a database, which then can detect a change in trend, or a sudden increase, and allow for closer examination. I think future research is important. For the past 2 years we had to endure hardships no one expected. From the closure of non-essential business, school closures, staffing shortages, and exhausted healthcare workers. A large portion of small business were permanently closed, and it appears that the development of social skills in youth has been delayed, not to mention the mental and physical toll covid has taken on us, as well as the grief, and loss from deaths. I think the true impact of Covid-19 on us as a society is largely unknown because it may take many years for us to feel the full impact. Whether we are in pandemic, or endemic, it is important we do not drop the ball on this one. We need to fully understand all aspects of harm caused by covid, so that we can fully understand how to better prepare for a future pandemic or public health crisis, and ensure we are continuously working towards undoing the damage that has been done to us as a whole over the past 2 years.

# Github repository Link

https://github.com/Kschilling13/CIND820-Capstone-Project/

# References

Elias, B. (2014, October 30). *Moving beyond the historical quagmire of measuring infant mortality for the first nations population in Canada*. Social Science & Medicine. Retrieved April 4, 2022, from https://www.sciencedirect.com/science/article/pii/S0277953614007096

Government of Canada, S. C. (2020, April 8). *Frequently asked questions on vital statistics*. Government of Canada, Statistics Canada. Retrieved April 4, 2022, from https://www.statcan.gc.ca/en/about/relevant/vscc/faq

*What now?: Canada health transfer: Background and future: Canada west foundation*. Canada West Foundation |. (2021, September 8). Retrieved April 4, 2022, from https://cwf.ca/research/publications/what-now-canada-health-transfer-background-and-future/

Shah, P. S., Ye, X. Y., Yang, J., & Campitelli, M. A. (2021, August 3). *Preterm birth and stillbirth rates during the COVID-19 pandemic: A population-based Cohort Study*. CMAJ. Retrieved April 4, 2022, from https://www.cmaj.ca/content/193/30/E1164

*Ontario population projections*. ontario.ca. (n.d.). Retrieved April 4, 2022, from https://www.ontario.ca/page/ontario-population-projections

Canada, P. H. A. of. (2021, May 28). *Covid-19 Daily Epidemiology update*. Canada.ca. Retrieved April 4, 2022, from https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

Government of Canada, Statistics Canada. (2021, December 22). *This summary provides an update on the work the agency has and is undertaking using existing and new data sources to provide critical insights on the social and economic impacts of covid-19 on Canadians. it includes data from fall 2021.* Government of Canada, Statistics Canada. Retrieved April 4, 2022, from https://www150.statcan.gc.ca/n1/pub/11-631-x/11-631-x2021003-eng.htm

Government of Canada, S. C. (2021, October 4). *Survey on covid-19 and mental health, February to May 2021*. The Daily - . Retrieved April 4, 2022, from https://www150.statcan.gc.ca/n1/daily-quotidien/210927/dq210927a-eng.htm

Wikimedia Foundation. (2022, April 2). *Timeline of the covid-19 pandemic in Canada*. Wikipedia. Retrieved April 4, 2022, from https://en.wikipedia.org/wiki/Timeline\_of\_the\_COVID-19\_pandemic\_in\_Canada