# Evolution of income (equality) in Canada*

Alicia Yang

26 April 2022

**Abstract**

People have claimed that individuals who are experiencing mental illness are more likely to be involved in violent incidents and crime activities either as victims or as criminals. To investigate more on this issue, this paper aims to explore how the mental health issue interacts with the reported crimes in Canada. By utilizing the data from Statistics Canada, we found that increase in mental health cases does have a positive correlation with the crime rates in Canada. More specifically, teenagers suffering from mental illness are much more likely to be involved in criminal activities. In addition, implications behind this correlation is discussed and possible solutions is proposed.

# Contents



*Code and data are available at: https://github.com/Alicia-y/Telling-stories-with-data-final-paper

# 1 Introduction

In recent decades the world as a whole is developing fast economically, and one evidence of which is the rise in income level. As more countries become more developed in terms of their economy and eventually overcome their issues of poverty, their governments have gradually shift their goal to dealing with the problem of economic inequality. It concerns how the wealth and income of the population is distributed. Although we are generally living higher quality lives and getting higher incomes, another problem is being put forward by people who are claiming that the gap between the rich and poor is getting larger due to technological advancement, globalization, countries' governmental structure and policies, which betters off the people who are already wealthy by making it easy for them to seek for more opportunities, thus, snowballing their assets faster. Income disparity can be a huge issue. It makes the income per capita of a country less reflective of its people's true living quality, because income per capita does not tell the distribution of the income. The richest people may pull up the statistics about the average/median income level, but it doesn't mean the country is performing well economically as a whole. In addition to that, income inequality also have other potential negative effects. It could reduce the stability of the economy and increase the risk of financial crisis. It could also result in corruption, misallocation of resources as the rich is empowered economically, socially and even politically. Canada is one of the country in the world that provides the most benefits and welfare, yet people still claimed that the economic inequality is rising along with the income over the past 2 decades(2). Therefore, I'm interested in whether this is a serious issue in Canada. If so, in which regions and what period of time is it a serious issue? And is it possible to predict how the disparity of income change in future?

The data used in this report is obtained from Open Government Data of Statistics Canada. The dataset is called "Upper income limit, income share and average income by economic family type and income decile", and it provides different kinds information about the income level of the households of Canada between 1976 and 2020 by the Canadian Income Survey. In this dataset, I'm interested in the year the data is from, the geographical location, the income decile so that we could observe the difference between the lowest and highest decile, and the actual value of the income. In section 2, a cleaned dataset is obtained

Associated with

Difficult and challenging to deal with

In section 2, the two datasets are discussed and cleaned to perform more accurate analysis. Visualizations in the forms of tables and graphs are presented to help to explain the possible correlation between mental illness cases and reported crime cases in Canada each year. In section 3, a model is constructed to justify the relationship and to make future predictions. All findings related to this topic is presented in section 4. A discussion is carried out in section 5 on the implications of the findings, possible solutions to this issue, as well as the limitations of this paper.

One of the reason why mental health is such a serious issue is that it's hard to identify and treat. Most patients undergo long-term treatments before fully recovered. To provide possible resolutions, measures are taken in recent years for the patients of mental illness, such as opening different types of community support services to help them cope with this issue. I will also look into the effectiveness of these support services and programs and see if they actually helped the ones suffering from mental health issues.

# 2 Data

## 2.1 Dataset Description and Methodology

The dataset, which is a summary of the income-related statistics in Canada between 1976 and 2020, is obtained from the Open Government Portal of Statistics Canada. The source of the data is from the Canadian Income Survey conducted annually. The survey covers all the possible population in Canada except for the people and households living in remote areas or indigenous settlements. It doesn't really affect the outcome of the survey, because these people only take up less than 2 percent of the population(). The selected respondents were drawn from the Labour Force Survey samples, which is based on a stratified probability sampling(). To reduce some non-sampling errors, telephone interviews ware conducted prior to the main survey to increase

the response rate of the selected participants Once they give consent to completing the survey, they will be able to complete it in an online form. Respondents are protected from the confidentiality rules, so that their privacy won't be disclosed after the survey is made public. One possible bias came with the survey methodology was the over-coverage of the units that are not the target population and the under-coverage of certain sub-population that should be included in the survey, such as certain remote areas.

The original dataset contains 173043 observations and 18 variables with all kinds of information about income level statistics. This report wants to focus on exploring the trend of income inequality and investigating the possible factors that may affect income inequality. Thus, I will be interested in the variables "Year", "Geographical Location", "Income decile", "Income" which tells how the income level of different regions of Canada has evolved throughout these years. R (R Core Team 2020), and R packages "tidyverse" (Wickham et al. 2019), "janitor" (Firke 2021), "knitr" (Xie 2021), "dplyr" (Wickham et al. 2021), and "kableExtra" (Zhu 2021) are utilized to create an extract of the cleaned dataset (Table 1).

Table 1: Extracting the first ten rows from the Income data

| Year | Geographical location | Income decile | Income | Income range |
|---|---|---|---|---|
| 1976 | Canada | Total deciles | 72400 | 184400 |
| 1976 | Canada | Lowest decile | 9200 | NA |
| 1976 | Canada | Second decile | 21200 | NA |
| 1976 | Canada | Third decile | 32400 | NA |
| 1976 | Canada | Fourth decile | 44600 | NA |
| 1976 | Canada | Fifth decile | 57000 | NA |
| 1976 | Canada | Sixth decile | 69100 | NA |
| 1976 | Canada | Seventh decile | 81800 | NA |
| 1976 | Canada | Eighth decile | 97200 | NA |
| 1976 | Canada | Ninth decile | 118300 | NA |

Table 1 shows the first ten rows of the cleaned dataset. It contains 6395 variables and 5 variables in total. The target population of the dataset is a combination of economic families and unattached individuals in Canada. As we known, the information from the dataset is extracted from the Canada Income Survey. Variable "Year" indicates the year of the survey. Variable "Geographical location" indicates where the group of respondents are from. Variable "Income" gives the average value of after-tax income of the respondents in corresponding year and location. Variable "Income decile" gives more information about the variable "Income", by telling us whether each piece of income information represents the average after-tax income of all respondents, the richest 10 percent respondents, the second richest 10 percent respondents, ..., and the poorest 10 percent respondents. This piece of data is extremely helpful because it also gives extra information about the difference regarding the income level between the poorest and richest groups. The last variable "Income range" is not in the original dataset. It indicates the difference in income between the poorest and richest groups, which is obtained from manipulating data from the "Income" column. It's only available at the instances of "Total deciles".

## 2.2 Data Visualization

In order to get further familiarized with the dataset and estimate the possible associations between the nation's average income and other factors, exploratory analysis is carried out by conducting data visualizations to observe whether the patterns between certain factors matches the generally expectations of the trend of income and income inequality in Canada. First of all, we will have a look through the overall trend of income level in Canada.

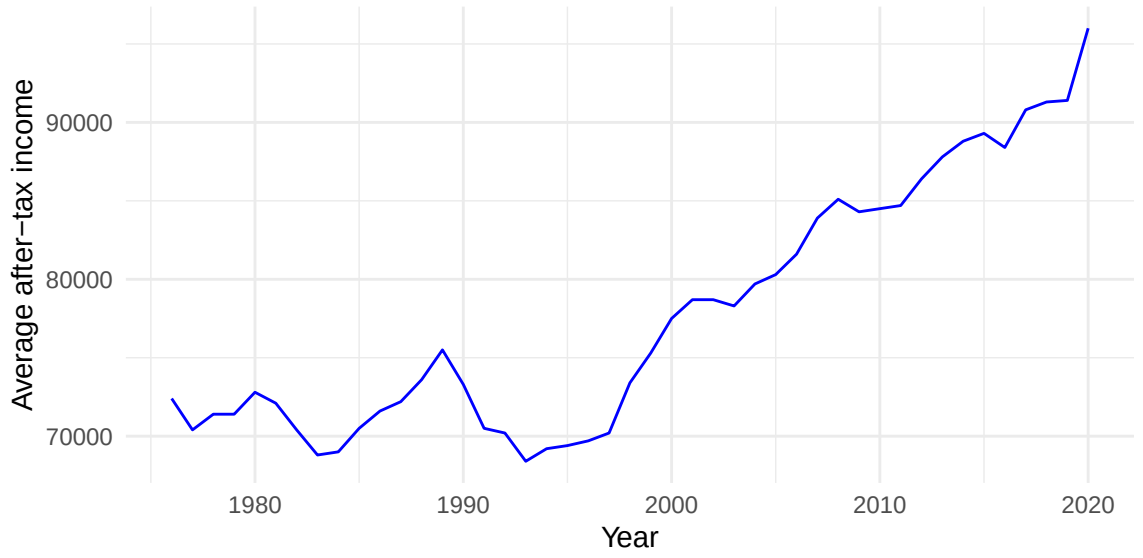


Figure 1: Average after-tax income in Canada per year between 1976 and 2020

As Figure 1 has demonstrated, Canadians have experienced a rapidly growing income level since 1976 to now. Early 1980s and 1990s are basically the only two periods of times that we see an obvious pattern of the decrease in income, and they all lasted for less than five years. This phenomenon is expected as during the time of early 1980s and 1990s, Canada was experiencing recession due to the change in monetary policy and a side effect of the cold war. As a result, the unemployment rate rose and due to an increase in the competitiveness of workforce, people's incomes have dropped. However, during other times, the income level have always demonstrated an increasing trend, indicating that without negative external forces, Canada generally has expanded its economy and enhanced its people's life quality well. The average after-tax income for economic families and unattached individuals in Canada has rose from around $72400 in 1976 to $96000 in 2020.

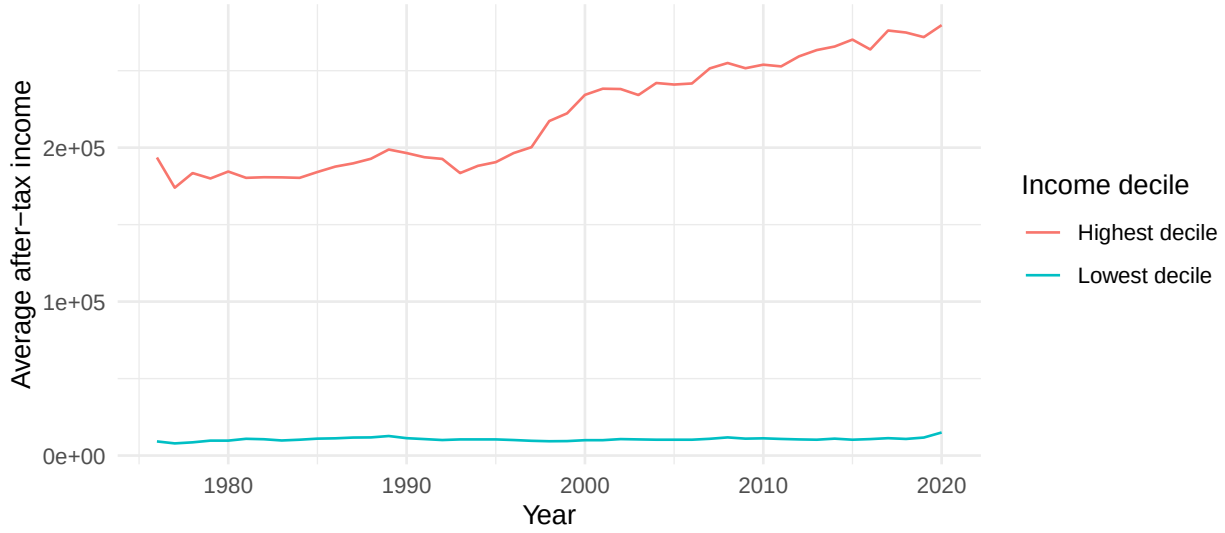


Figure 2: Trend of lowest and highest decile of sverage after-tax income in Canada between 1976 and 2020

Figure 2 is an illustration of how the income level of the richest and poorest groups of economic families and unattached individuals differ between the year of 1976 and 2020. The blue line indicates the trend of the lowest decile of the average after-tax income in Canada, which barely increased. In fact, the lowest decile of average after-tax income was $9200 in 1976, but only $15000 after 44 years in 2020. On the other hand, the red line indicates the trend of the highest decile of the average after-tax income in Canada, which has a very clear increasing pattern, by looking at the data, it increased for $86000 between 1976 and 2020, which is more than 10 times the income level increase of the lowest decile. By comparing the two lines, we can see that the income inequality in Canada as a whole is getting larger as the two lines drift further away from each other as time passed.

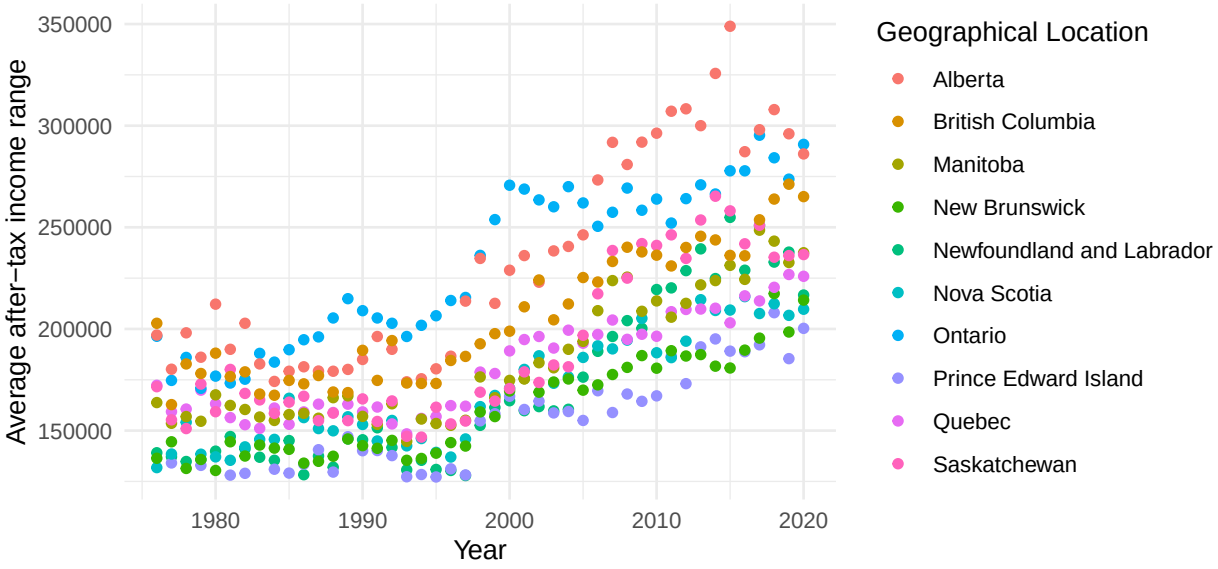


Figure 3: Difference between lowest and highest average after-tax income in the 10 provinces of Canada between 1976 and 2020

Figure 3 has demonstrated the difference between the lowest and highest decile of the average after-tax income of economic families and unattached individuals in the ten provinces of Canada between 1976 and 2020. Overall, there is an increasing pattern in every province, indicating that the gap between the incomes earned by richest and poorest people is getting larger in every region of Canada. By just examining this figure, some particular provinces seem to experience a greater increase in income inequality than the others as time passed. For instance, as shown by the red dots, Alberta clearly has the greatest increase among all the provinces of Canada. On the other hand, Quebec seems to see the smallest increase in income inequality as it has a flatter slope in the figure. The cause the difference in income inequality between these provinces is likely to be difference in their provincial policy and economy.

# 3 Model

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \tag{1}$$

Equation (1) seems useful, eh?

Here's a dumb example of how to use some references: In paper we run our analysis in `R` (R Core Team 2020). We also use the `tidyverse` which was written by (**???**) If we were interested in baseball data then (**???**) could be useful.

We can use maths by including latex between dollar signs, for instance $\theta$.

# 4 Results

# 5 Discussion

## 5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

## 5.2 Second discussion point

## 5.3 Third discussion point

## 5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

# Appendix

## A Additional details

# B References


https://www.nytimes.com/2020/07/01/opinion/economic-inequality-moral-philosophy.html#:~:text=Enough%20economic%20inequality%20can%20transform,society%20ruled%20by%20the%20rich.&text=Large%20ine

(2)https://www.conferenceboard.ca/hcp/hot-topics/canInequality.aspx#:~:text=Income%20inequality%20in%20Canada%20ha

https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5200

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org,%20https://github.com/tidyverse/dplyr.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/,%0Ahttps://github.com/haozhu233/kableExtra.