Factors Affecting Canada Employment Rate between 2017 and 2021[*]

An Exploration of Canada Labour Force Survey

Luckyna Laurent   Samita Prabhasavat   Zoie So

27 March 2022

Abstract

Educational degree and sex have strong relevance for employment rate. This report uses the Labour force characteristics by educational degree, month, and sex generated using raw data collected from the Labour Force Survey (LFS) conducted by Statistics Canada. The data are manipulated and explored which results in strong evidences indicating that both educational degree and sex have effects on the employment rate. This paper will explain the importance of educational degree in Canada labour market and the difficulties people might face according to their sex.

1 Introduction

In 1997, Statistics Canada started a Labour Force Survey (LFS) to observe the characteristics of labour market in Canada. The Labour Force Survey provides estimates of employment rates in Canada by potential factors including month, age group, aboriginal group, province, average wage, and educational degree. The LFS estimates are the first of the major monthly economic data series to be released. LFS data are used to produce the well-known unemployment rate as well as other standard labour market indicators such as the employment rate and the participation rate.

The data set used in this report contains Canada labour force characteristics such as employment rate, unemployment rate, and participation rate by month, age group, and educational degree. The time frame interested in this report is between 2017 to 2021 which covers from pre COVID-19 pandemic time to present. The purpose of this report this to explore factors that may affect employment rates in Canada.

Educational degree and sex will be two main factors of this study. Also, the time period between 2017 and 2021 will be taken into consideration to see the employment rate before COVID-19 pandemic and the employment rate as the pandemic is ongoing. The data was downloaded from Statistics Canada website and manipulated in order to make sure that the data is in the right format. The data was then explored and interpreted which leads to the conclusion that both educational degree and sex have strong effect on the employment rate in Canada. However, a supplemental survey was created as more data is needed in order to further explore Canada labour market. Understanding the employment rate can help us understand things that happen in that certain time period including the country's economics performance during COVID-19 pandemic.

This report will first, look at the data from the survey and the methodology used to collect data in Section 2 to provide readers with the context of the data used in this report. Second, it will discuss the insights found from exploring the data in Section 3. Last, it will conclude the insights found from exploring this data set and provide plans for further exploration of this data set in Section 4.

---

[*]Code and data are available at: https://github.com/PSamita/Labour-Force.git

2 Data

2.1 Overview of the data

In this report, data analysis is done using R statistical programming language (R Core Team 2020). The here package is used to access file from another folder in the same R Project (Müller 2020). The janitor package is used to clean the names of the columns (Firke 2021). The tidyverse package is used for data manipulation (Wickham et al. 2019). The knitr package is used to create a table from data frame (Xie 2014). The ggplot2 package is used to create graphs (Wickham 2016). The RColorBrewer is used to assign colours to graphs and figures (Neuwirth 2014). The bookdown package is used to cross reference graphs and figures (Xie 2016).

The data set used in this report is the Labour force characteristics by educational degree, month, and sex generated using raw data collected from the Labour Force Survey (LFS) conducted by Statistics Canada (Canada 2021). The LFS is an active survey carried out monthly by Statistics Canada to estimate the employment rate and unemployment rate as well as other standard labour-related issues such as participation rate and labour demographic characteristics. The survey covers a variety of industries, public and private sectors nationwide, in both provinces and territories.

The LFS is an important indicator of the labour market in Canada. It is rigorous and timely to ensure that the results are of acceptable quality and credible. These data are used by both employees and employers across Canada, as well as different levels of government for evaluation and planning of employment and welfare programs. The data are also used by labour market analysts, economists, consultants, planners, forecasters, and academics in both private and public sectors. The LFS often compares GDP data with labour market trends to evaluate whether they are in line with general economic performance.

Normally, four surveys are used together to provide a more complete picture of current labour market happenings. The Labour Force Survey (LFS, record number 3701), the Survey of Employment, Payrolls, and Hours (SEPH, record number 2612), the Employment Insurance Statistics (EIS, record number 2604), and the Job Vacancy and Wage Survey are the four surveys mentioned (JVWS, record number 5217). The LFS publishes labour market data on a monthly basis, including the unemployment rate and demographic analyses. Later in the report, the SEPH provides more insight on non-farm industry employment and wages, while the JVWS provides preliminary job vacancy indicators. Employment Insurance benefits are broken down by geography, socio-demographics, and previous occupation in the EIS. The JVWS publishes thorough information on job openings by occupation and economic region every quarter.

2.2 Survey Methodology

The subjects of interest for this survey include employment and unemployment rate, hours of work and arrangements, wages, salaries, industries, labour, and unionization and industrial relations.

2.2.1 Target Population

The non-institutionalized population aged 15 and up is the target population in the LFS. The survey is done in all provinces and territories around the country. Persons living on reserves and other Aboriginal settlements in the provinces; full-time members of the Canadian Armed Forces; the institutionalized population; and households in extremely remote locations with very low population density are excluded from the survey's scope. Together, these groups account for fewer than 2% of the Canadian population aged 15 and up being excluded.

The LFS does not include any questions that ask respondents if they are temporary foreign workers. As a result, the LFS is unable to offer counts or employment data for temporary foreign workers. Temporary foreign workers will only be included in the LFS if they identify the designated house as their regular place of residence when contacted for the survey. Furthermore, they cannot be distinguished from a wider group

of responders who are not Canadian citizens or landed immigrants. Because the 'other' group accounted for only 2% of the employed population in 2014, it had a minor impact on overall employment figures. This group also includes Canadian citizens of ancestry who were born elsewhere, as well as international students with a study permit.

Estimates from the National Labour Force Survey are based on provincial LFS results. The results of the territorial LFS are published separately from the national estimates.

2.2.2   Instrument Design

Periodically, the Labour Force Information questionnaire is modified. When items were added to the current Labour Force Information questionnaire in 1997, it was revised. The following were among the additions:

Employees' average weekly and hourly salaries, union membership, employment stability, and establishment size are all factors to consider.

Review committees, focus groups, and pilot tests have all been used to test the Labour Force Survey Information questionnaire.

As previously stated, the questionnaire was significantly expanded in 1997. These revisions included questions concerning employees' average weekly and hourly pay, union membership, job stability, and workplace size.

2.2.3   Sampling

This is a sample survey with a cross-sectional design. The LFS employs a stratified multi-stage design with a probability sample. Each province is divided into several large geographical strata. Within each stratum, the initial stage of sampling consists of picking smaller geographic areas known as clusters. The second round of sampling is choosing houses from within each cluster.

The LFS uses a rotating panel sample design, in which selected dwellings are included in the sample for six months in a row. About one-sixth of the LFS sampled dwellings are in their first month of the survey each month, one-sixth in their second month, and so on. Each of the six rotating groups can be utilized as a representative sample on its own, which is a characteristic of the LFS sample design.

Within selected dwellings, basic demographic information is collected for all household members. All civilian household members aged 15 and up have their labour force information collected. The sample is divided into provinces and strata within provinces in order to meet the demand for accurate estimates at various geographic scales. The monthly LFS sample size has recently been around 56,000 households, resulting in the collection of information on the labour market for around 100,000 people. It should be noted that the LFS sample size may alter from time to time to satisfy data quality or financial needs.

2.2.4   Data Sources

This survey must be completed. Data is gathered directly from survey participants. Responses to survey questions are recorded by the interviewer using a computerized questionnaire on a laptop or desktop computer at the time of the interview. To maintain anonymity, the response data is encrypted and delivered electronically to the relevant Statistics Canada Regional Office. They are then sent to the central office in Ottawa for further processing through a secure line.

Interviewers from a regional office CATI (Computer Assisted Telephone Interview) site or a field interviewer conduct LFS interviews over the phone in English or French. Senior interviewers supervise all LFS interviewers and are responsible for ensuring that their team is conversant with survey topics and processes, as well as checking their interviews on a regular basis.

One competent household member is usually able to provide information on all of the members of the home. This type of 'proxy' reporting accounts for about 65 percent of the data gathered. Respondents are asked

to participate in the survey for six months. It takes about 20 minutes to complete the birth interview. The next five interviews will take between 10 and 12 minutes to complete.

2.2.5 Error Detection

Some editing is done on the spot during the interview. When the information submitted is outside of the anticipated range (too large or too small) or contradictory with earlier entries, the interviewer is prompted to make changes via computer message panels. If the respondent does not know or refuses to answer, interviewers have the option of bypassing the edits or skipping questions. As a result, once the response data reaches head office, it is subjected to additional editing and imputation processes.

The editing and imputation phases of processing entail the detection and modification of logically inconsistent or missing information components. Since the true value of each entry on the questionnaire is unknown, the only way to detect inaccuracies is to look for glaring irregularities (for example, a 15-year-old respondent who is recorded as having last worked in 1940). If a value appears suspicious but is reasonable, it will be included in the monthly statistics. As a result, quality controls and interviewer training must be prioritised in order to ensure that errors are both few and non-systematic.

2.2.6 Imputation

It is possible to notice that all questionnaire items for individuals (persons) in the family are missing during the editing step of processing. Complete (or total) non-response is the term for this. When only a few questionnaire data items are missing, this is known as item non-response. Complete non-response is resolved using imputation and non-response weight adjustment. Imputation is the only way for resolving item non-response.

Carry-forward, deterministic, and donor (hot-deck) imputation are among the imputation methods used in the LFS. When errors or omissions are discovered, the incorrect or missing items are replaced with logically consistent values. Deterministic (or substitution) imputation is the word for this method. These modifications are done either automatically by the edit and imputation system or manually by specialists. These adjustments are based on pre-determined criteria and may include the questionnaire's internal logic, a reference to previous month's data (if available), or the usage of similar records to impute one or more variables.

If available and appropriate, some missing items are handled by carrying forward data from the previous month. Other missing information may necessitate donor (hot-deck) imputation, which entails copying data from another person (i.e., a 'donor') who has similar traits. Editing and imputation changes are recorded in all cases, and this data is utilised to evaluate various aspects of survey performance. These error logs are also used to inform interviewers of previous mistakes in order to avoid making the same mistakes again in the future.

2.2.7 Estimation

The sample data are weighted to allow estimates to be tabulated at the national, province, and sub-provincial levels. The sample design specifies the amount of weighting variables that will be used to calculate individual weights. The fundamental weight, which is the inverse of the likelihood of selection, is the most important component. For example, if 2% of the houses in a certain area are sampled, each household would be given a basic weight of 1/.02=50. The fundamental weight is then changed to account for any sub-sampling that may have happened owing to growth in the area. Non-response and coverage mistake are then factored into the weight.

Imputation procedures such as carry forward, substitution, and donor imputation are used in the LFS to adjust for certain survey non-response. The weights for the responding homes in the same area are adjusted

to account for any remaining non-response. The characteristics of responding households are assumed to be similar to those of non-responding households in this non-response adjustment.

The subweights are the weights generated after the non-response adjustments. To remedy for coverage problems, the final modification to the weight is done. The subweights are modified so that the population estimates from the survey match the control totals. In the LFS tabulations, these final weights are used.

2.2.8 Quality Evaluation

The indicators from the Labour Force Survey that are released each month are the most up-to-date. Any validation of LFS data with data from other sources should be done with caution. Select data from the LFS is compared to analogous data from the Survey of Employment, Payrolls, and Hours (SEPH), Employment Insurance Statistics, and the census on a regular basis. Economists who deal with the LFS frequently compare GDP data with LFS data to examine if labour market patterns are consistent with overall economic performance. Other parallels include:

Manufacturing shipment data and LFS manufacturing employment; Dwelling starts, building permits and construction employment; Retail and wholesale sales and trade employment.

Furthermore, the LFS is subjected to a variety of tests to guarantee that the estimates are of acceptable quality.

2.2.9 Disclosure Control

Statistics Canada is prohibited by law from disclosing any data gathered under the Statistics Act that relates to any identifiable person, business, or organisation without that person, business, or organization's prior knowledge or written authorization. To prohibit the publication or exposure of any material deemed sensitive, certain confidentiality standards are applied to all data that is disclosed or published. Data is suppressed whenever necessary to prevent direct or residual disclosure of personally identifying information.

2.2.10 Revisions and Seasonal Adjustment

Most labour market estimates are subject to seasonal variation, which refers to annual swings caused by weather and regular institutional events such as vacations and holiday seasons. Seasonal adjustment is used to remove seasonal changes from over 3,000 series in the LFS in order to simplify short-term change analysis for significant indicators such as employment and unemployment by age and sex, employment by industry, and employment by worker class (public and private employees or self-employed). At the national and provincial levels, several of these indicators are seasonally adjusted. For census metropolitan areas (CMAs), main labour force status data are also seasonally adjusted and reported as three-month moving averages to smooth out erratic fluctuations caused by limited sample numbers. The technique that is being applied for seasonal adjustment is X-12-ARIMA.

The Labour Force Survey revises its estimates for the previous three years at the start of each year, using the most recent seasonal elements. LFS data is also adjusted every five years after new population estimates from the most recent census become available. All LFS data dating back to the previous census is re-weighted using the new population estimates (because the new population estimates will include the inter-censal period between the two most recent censuses), and all relevant historical LFS estimates are corrected at that time. Generally, this is when the most recent classification systems for industry, vocation, and geography, as well as other revisions, are introduced.

2.3 Data Accuracy

All estimates are subject to sampling and non-sampling errors because the LFS is a sample survey.

Non-sampling mistakes can occur at any point throughout the survey data collection and processing. Coverage errors, non-response errors, response errors, interviewer errors, coding errors, and other processing faults are examples of these.

The average non-response rate to the LFS is around 10% of eligible homes. Interviewers are told to make every effort to schedule LFS interviews with members of eligible households. Each month, a tiny number of non-responding families remain after all attempts to secure interviews have been undertaken. To accommodate for non-responding families, a weight modification is used.

Coefficients of variation for LFS estimates as a function of the standard error and the size of the estimate are used to calculate sampling errors associated with survey estimates.

2.4   Variables

As of 2021, there are 64 variables in the LFS as follow:

- Aboriginal group of person, category
- Aboriginal identity of person, category
- Absence from work of employed person, category
- Absence from work of employed person, duration
- Absences from work of employed person, category
- Absences from work of employed person, ratio
- Activity index of economy, category
- Activity index of economy, index
- Activity prior to unemployment of person, category
- Actual hours worked of employed person, category
- Actual hours worked of employed persons, number
- Age at retirement of person, measure
- Age composition of household, category
- Age of person, category
- Average hourly wage rate of employed persons, value
- Average weekly wage rate of employed persons, value
- Class of worker of employed person, category
- Duration of unemployment of person, category
- Duration of unemployment of person, duration
- Educational attainment of person, category
- Educational attainment of person, proportion
- Establishment size of employed person, category
- Family type of person, category
- Gender of person, category
- Geographic location of economy, name
- Geographic location of person, category
- Geographic location of person, name
- Hourly wage of employed person, category
- Immigrant status of person, category
- Industry of employed person, type
- Job of employed person, type
- Job search method of person, category
- Job tenure of employed person, category
- Job tenure of employed person, duration
- Labour force status of employed person, category
- Labour force status of person, category
- Median hourly wage rate of employed persons, value
- Median weekly wage rate of employed persons, value

- Number of person, count
- Occupation of employed person, type
- Overtime hours of employed person, category
- Overtime hours of employed person, time
- Place of birth of person, category
- Presence of children in the household of person, category
- Reason for absence from work of person, category
- Reason for leaving job by person, category
- Reason for not looking for work by person, category
- Reason for part-time work of employed person, category
- Residence on or off reserve of person, name
- School attendance of person, category
- School attendance of person, proportion
- School attendance of person, status
- Supplementary unemployment rate of active population, category
- Supplementary unemployment rate of active population, rate
- Time lost for absence from work of employed person, category
- Time lost for absence from work of employed person, time
- Type of student of person, category
- Type of work of employed person, category
- Type of work sought of person, category
- Union coverage of employed person, category
- Usual work hours of employed person, category
- Usual work hours of employed person, duration
- Weekly earnings of employee, value
- Weekly wage of employed person, category

2.5   Strengths and Weaknesses

The strengths make LFS become an accredicted indicator of the standard labour market. The rigorous methodology and quality control are the major factors. For instance, the probability sample in both smaller geographic areas and selected dwellings cover each province in the large geographic stratum. The sample data are even weighted to enable tabulations of estimates at geographical levels of aggregation. Other than that, error detection, editing and manipulation are involved to identify any logical inconsistent or missing information items. These measures can correct for coverage errors and make the survey estimates of the population conform to control totals.

In addition, the one-to-one interview can ensure all required information is collected in the right direction. For example, interviewers may misunderstand instructions or respondents may make errors in answering questions. Meanwhile, the survey responses are captured directly by the interviewer who is under supervision and monitoring. These quality controls methods effectively minimize mistakes and adjust missing responses in the survey to affect the result further.

However, there are still some limitations of the survey. The biggest weakness is the options are too binary. Take an example of type of employment, there are only full-time/part-time, employed/unemployed to be chosen. Regarding the latest labour market, freelancing is poised to become the majority workforce in the future. Considering the survey results are nationwide, the persons living in Yukon, Nunavut, and the Northwest Territories, person living on Indian reserves are excluded from the survey. They compromise 2% of the population. Understanding their employment conditions are important for both the overall and specific labour market. Besides, the survey ignores some groups of people including temporary foreign workers and Canadian citizens who were born elsewhere. These people can make huge differences to the results of the survey.

Another weakness found from the LFS is the unstableness of the sample size. As one-to-one interview is both time-consuming and cost-sonsuming, the survey scale is subject to change from time to time in order

to meet data quality or budget requirements. It is no doubt that a larger data set and more accurate result have a positive relation. All in all, the LFS is still a credible and reliable survey as a market indicator in Canada, which can reflect the current labour market to a large extent.

2.6   Focused Aspect

This paper will focus on the Labour force characteristics by educational degree, month, year, and sex data. This data set contains employment rates calculated from the data collected in LFS by year, month, educational degree, and sex.

2.6.1   Variables

As of 2021, there are 5 variables in this data set as follow:

- Year
  - The year when the data was collected which ranges from 2017 to 2021 in this paper.

- Month
  - The month when data was collected which ranges from 1 (January) to 12 (December).

- Educational degree
  - The educational degrees of the participants which are as follow:
    * No degree, certificate or diploma
    * High school graduate
    * High school graduate, some post-secondary
    * Postsecondary certificate or diploma
    * Bachelor's degree
    * Above bachelor's degree

- Sex
  - Sex demographic of the participants which are as follow:
    * Male
    * Female

- Employment rate
  - The employment rates were calculated in percentage.

2.6.2   Data Exploration

Table 1 shows the example of raw data downloaded from Statistics Canada website. The table includes 5 variables which are year, month, educational degree, sex, and employment rate.

Table 1: Canada Employment Rate from 2017 to 2021 by Educational Degree and Sex

Year	Month	Educational Degree	Sex	Employment Rate
2017	01	No degree, certificate or diploma	Males	38.3
2017	02	No degree, certificate or diploma	Males	38.4
2017	03	No degree, certificate or diploma	Males	38.1
2017	04	No degree, certificate or diploma	Males	38.4
2017	05	No degree, certificate or diploma	Males	40.9
2017	06	No degree, certificate or diploma	Males	41.9

The line graphs in Figure 1 represent the data downloaded from Statistics Canada website which is used in this report to further explore the effects of educational degree and sex on employment rate in Canada. The y-axis shows the employment rate in Canada while the x-axis shows the year between 2017 and 2021. Each line graph compares the employment rate between male and female in the same educational level.

Figure 1: Canada Employment Rate from 2017 to 2021 by Educational Degree and Sex

In order to choose the appropriate statistical summary for this data, a box plot was plotted to check the normality of the data. According to a box plot in Figure 2, the data points in this data set lie between the upper thin horizontal line to the bottom thin horizontal line. A thick black horizontal line across the blue

box represents the median employment rate from this data set where 50% of the data is above the median, and another 50% of the data are above the median. Within the blue box itself, 25% of the data is above the median while another 25% of the data is below the median. In this box plot, the thick black horizontal line or the median line is a bit high which indicates a possible skew in the data. In addition, no outlier was detected in this dataset.
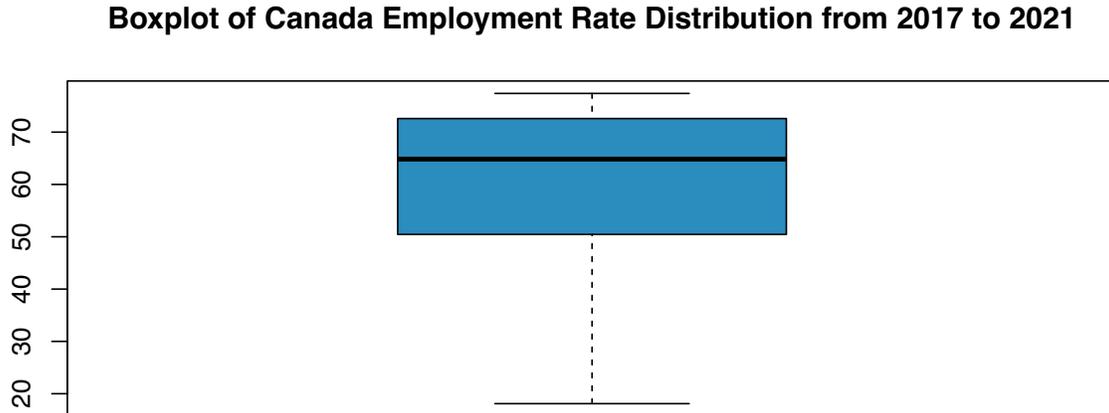
Boxplot of Canada Employment Rate Distribution from 2017 to 2021

Figure 2: Boxplot of Canada Employment Rate Distribution from 2017 to 2021

In order to further confirm the non-normality of the data from looking at the box plot above, a histogram was plotted as in Figure 3. It can be seen that the histogram is left skewed and does not look like a typical bell-shaped diagram which indicates that the data is not normally distributed.

The median was chosen as a measure of central tendency for exploring this dataset as the data is not normally distributed.

3 Results

3.1 Canada Employment Rate by Year

A line graph in Figure 4 shows changes in median employment rate in Canada over a 5-year period from 2017 to 2021. From 2017 to 2018, it can be seen that the median employment rate dropped from around 66 to around 65. In 2019, the median employment rate bounced back to almost the same as in 2017. From 2019 to 2020, the median employment rate in Canada declined dramatically to around 61 which was mainly caused by COVID-19 pandemic. From 2020 to 2021, the median employment rate in Canada started to show a small growth and increased to around 62.

3.2 Canada Employment Rate by Educational Degree

A bar graph in Figure 5 shows the median employment rates in Canada by educational degree between 2017 to 2021. It can be seen that people with educational degree above Bachelor's degree has the highest median employment rate among all educational degree groups at almost 75. The median employment rate for people

Histogram of Canada Employment Rate Distribution from 2017 to 2021
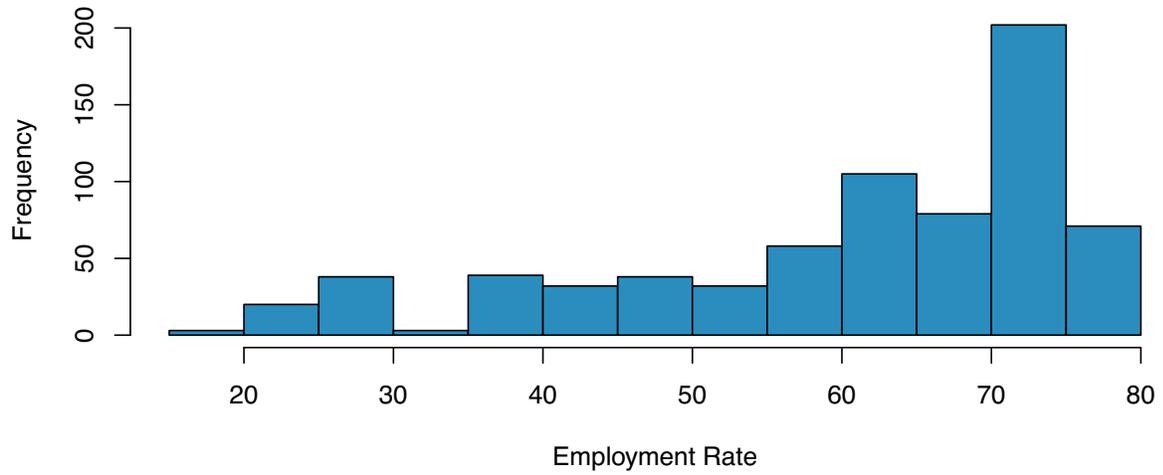

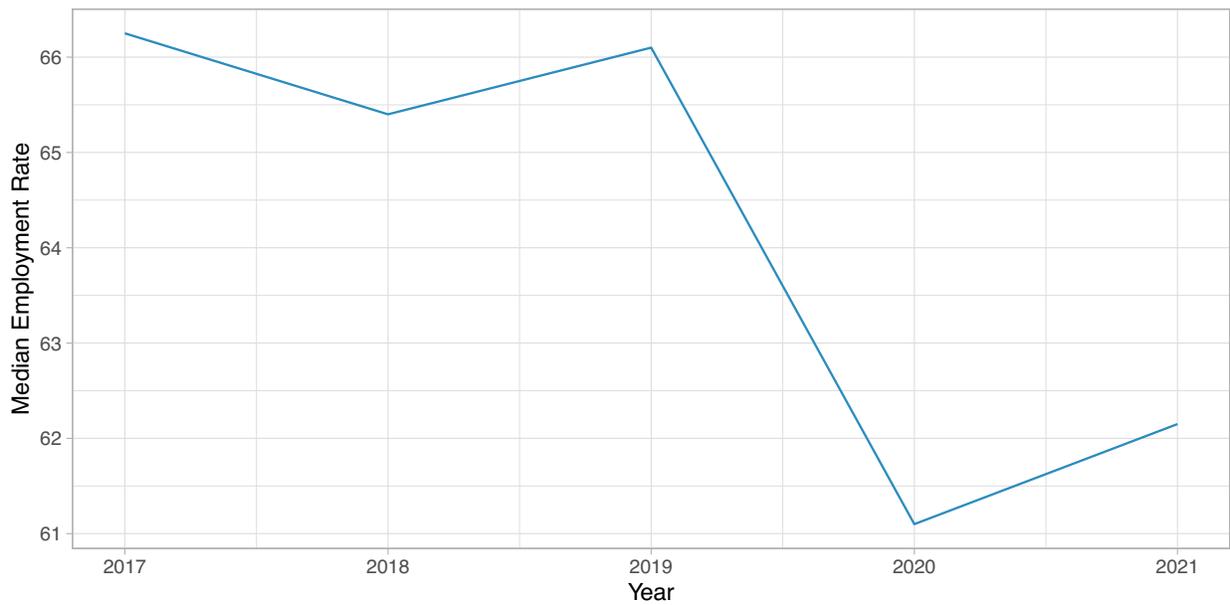Figure 3: Histogram of Canada Employment Rate Distribution from 2017 to 2021

Figure 4: Canada Median Employment Rate from 2017 to 2021

in Canada with Bachelor's degree is a little lower at around 72. People with post-secondary certificate or diploma have the third highest median employment rate in Canada at around 66. It is clear that there is a gap between people who graduated from a college and who are not. The median employment rate between people with high school graduate and some post-secondary study is about 69 which is a little lower than people who completed at least a certificate or diploma at a post-secondary institute. The median employment rate for people who have hogh school graduate alone is around 52. Comparing to other educational degree groups, people who do not have any degree have the lowest employment rate at around 30 which is a huge different from other groups.
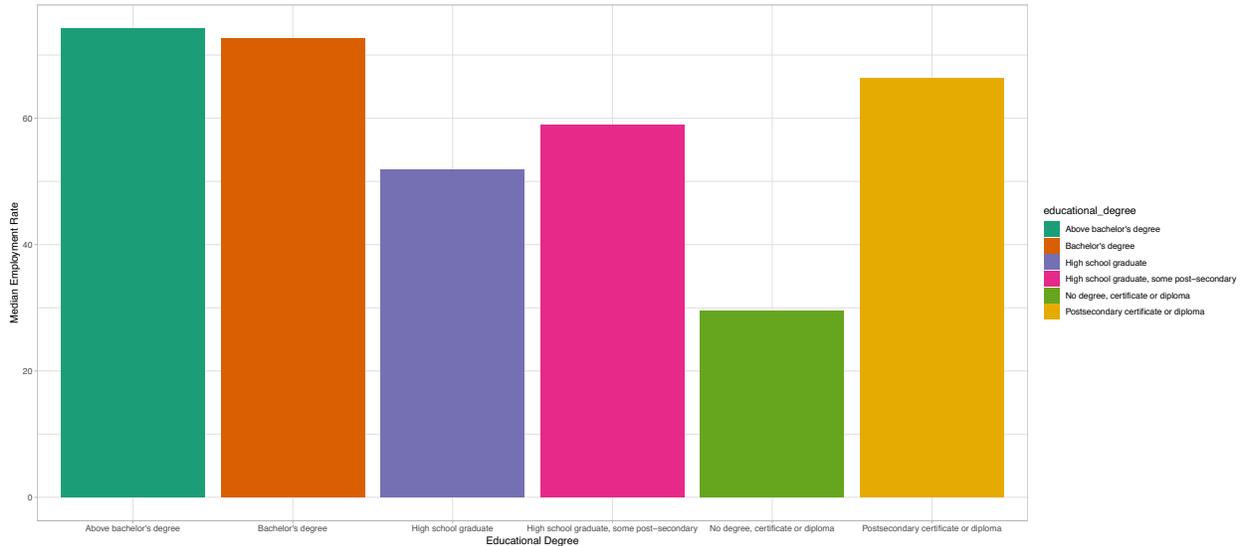

Figure 5: Canada Median Employment Rate from 2017 to 2021 by Educational Degree

3.3 Canada Employment Rate by Sex

A bar graph in Figure 6 shows the median employment rates in Canada by sex between 2017 to 2021. It can be seen that the median employment rate in Canada for male is slightly higher than female. While the median employment rate for male is about 68, the median employment rate for female is around 61.

3.4 Canada Employment Rate by Sex and Educational Degree

A bar graph in Figure 7 shows the median employment rates in Canada by sex and educational degree between 2017 and 2021. It can be seen that the median employment rate in Canada for male is slightly higher than female for every educational level except among people with degrees above Bachelor's degree where the median employment rate between male and female are around the same value. Also, the gap between median employment rate among male and female tends to get larger as the level of education gets lower.

4 Discussion

4.1 Methodology

This paper was created by downloading raw data as csv file into RStudio software. The data was then filtered to keep only variables of interest, filtered out null values, and ensured that all data are consistent and in the
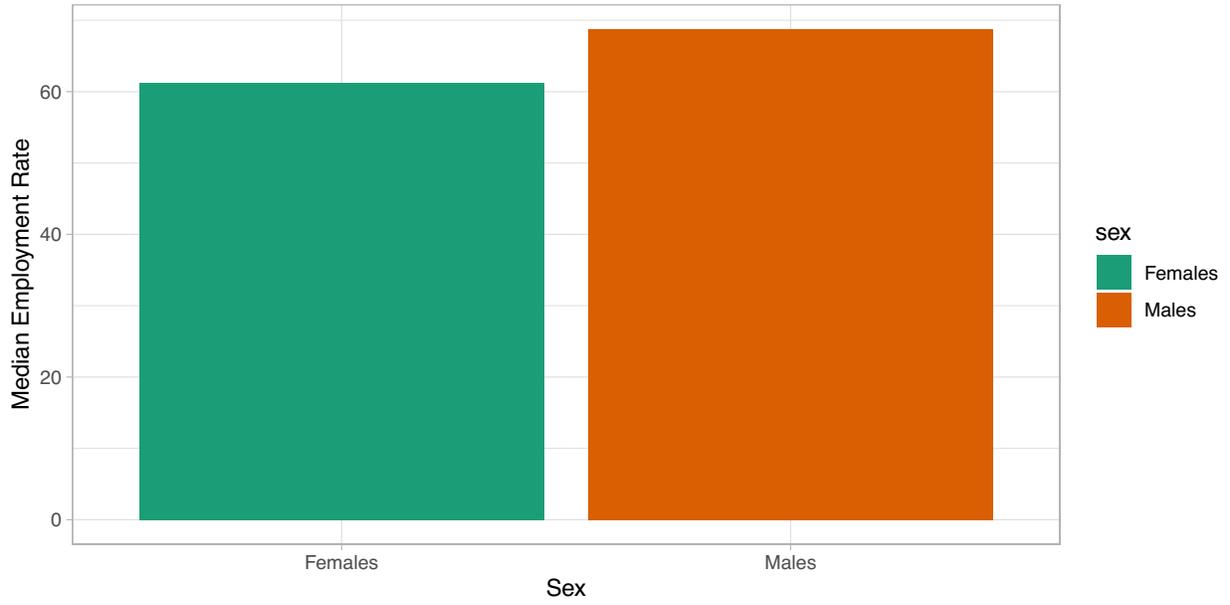
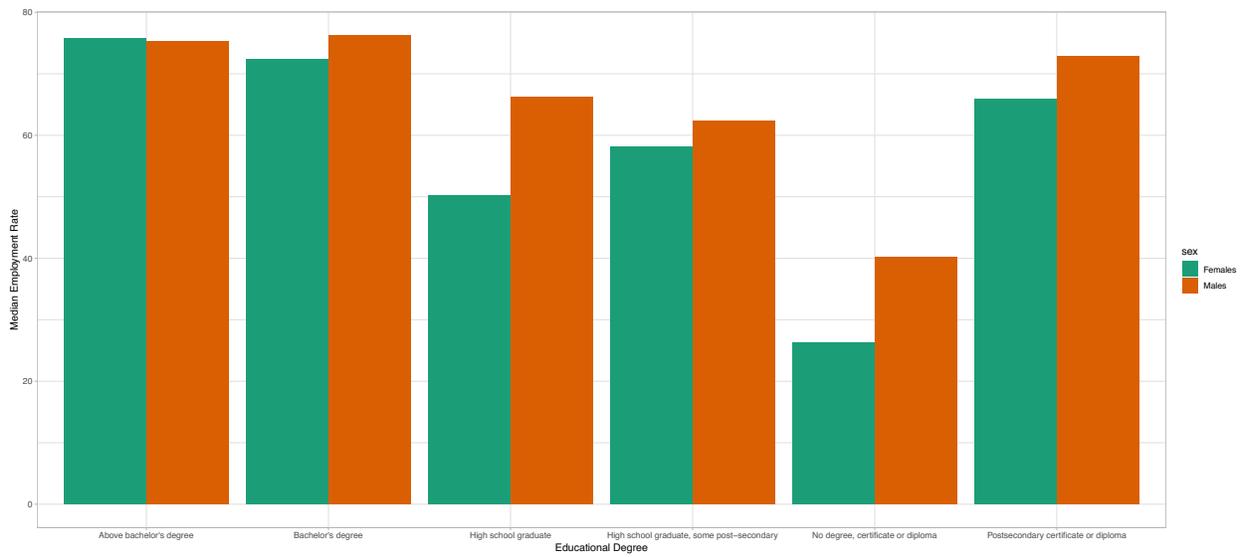Figure 6: Canada Employment Rate from 2017 to 2021 by Sex

Figure 7: Canada Median Employment Rate from 2017 to 2021 by Educational Degree and Sex

right format. For example, the employment rates were calculated in percentage which means all values in employment rate column should be more than or equal to 0 but less than or equal to 100. Moreover, outliers were detected if applicable. In this case, no outlier was detected.

After the data is ready to be used, the next step is data exploration. First, box plot and histogram were used to test the normality of data which led to deciding on which statistical summary will be used in this paper. Second, statistical summary was used to explore the data. Third, graphs are plotted to help finding insights about the data. Last, another survey was designed to gain further information related to this dataset.

4.2 The Impact of Educational Degree on Canada Employment Rate from 2017 to 2021

According to Figure 5, he higher one's education, the more likely one is to find work. People with advanced degrees, such as a master's or PhD, had somewhat higher employment rates than those with a bachelor's degree among university graduates. Having some postsecondary education but not completing a programme leading to a degree, certificate, or diploma, on the other hand, offered minimal advantage above high school completion.

Looking back to Figure 1, the employment rate for people with a degree higher than Bachelor's degree is the highest among all groups regardless the sex. The employment rate for people with Bachelor's degree is slightly lower than people with advanced degrees. These two groups of people can be considered as highly employed compared to other groups with the employment rate falls between 70 to 80. People with post-secondary certificate or diploma and people with some post-secondary study can be seen as middle employed compared to other groups in this study as their employment rates are around the same range which is 60 to 70. However, the employment rate for people with post-secondary certificate or diploma is considerably higher. The employment rate among people with high school graduate falls between 40 to 60 and the employment rate among people with no degree falls between 20 to 40 which is the lowest compared to other groups.

According to a study, people with lower educational degrees like those who do not have degree or those with high school graduate are more likely to be hired in sectors such as construction, extraction, and manufacturing (Chen Ying-Yeh 2017). As a result, employers tend to not look into their educational backgrounds, instead, employers look more into skills relevant to jobs. Even though the employment rate among people with lower educational level is much lower than other groups, people within the group with demanding skills might still be able to find jobs and can be considered in middle employed group.

For people with higher educational levels, the study shows that the amount of knowledge a person can use is most likely to have a positive relation with that person's educational level (Duggan 2019). Nowadays, there are lots of online resources for people to learn and gain knowledge from without attending a post-secondary institution. For some companies especially in tech industry these days, a certification from a certain online institute or a certain company can be used as a substitute to a Bachelor's degree. However, a study shows that most employers still look for a degree from a reliable institution because they believe there is no better indicator to guarantee a person's knowledge (Duggan 2019). For example, a person who wants to become a data analyst must know statistics, data analysis techniques, and programming languages. Most online courses in the market only teach data analysis techniques and programming languages. In fact, statistics is actually the foundation of data analysis that cannot be ignored. That being said, most employers find online courses as supplementary learning but not a substitute for a degree from an educational institution. Moreover, some studies suggest that there are things people cannot get from enrolling in an online course such as class discussion which contribute to knowledge among students and professor in that class (Moja 2021).

In addition, Figure 1 shows that the employment rates among all groups dipped in 2020 when COVID-19 pandemic started in Canada. It can be seen that the dip in graphs for people with higher educational levels is not as huge as the dip in graphs for people with lower educational level. According to Statistics Canada, constructions and manufacturing are among the sectors that were most affected by COVID-19 pandemic when it started. On the other hand, people with higher educational levels especially those with degrees

higher than Bachelor's degree were minimally affected by the pandemic as many of them are already in stable position such as educator, senior employee, or healthcare staff (Canada 2021).

4.3   The Impact of Sex on Canada Employment Rate from 2017 to 2021

Looking back to Figure 1 and Figure 7, the employment rates among male and female with a degree high than Bachelor's degree does not show significant differences. For people with Bachelor's degree or post-secondary study, the employment rate among male is slightly but considerably higher than the employment rate among female. For people with a certificate or diploma from post-secondary institutes, the employment rate among female is significantly lower than male. It is interesting that the differences in employment rate between male and female is larger among people with a certificate or diploma from post-secondary institutes compared to those who have post-secondary study but do not have a degree. For both people with no educational degree and people with high school graduate, the gaps in employment rate between male and female are almost the same amount where males have much higher employment rate than females.

A study suggests that people with lower level of education tend to work more a labours which makes employers consider strength more than knowledge (Chen Ying-Yeh 2017). In this case, employers tend to choose males to work in construction sites as they believe males have more strength than females. Also, as a result of patriarchy from the past, people have the idea of constructions are only for male workers. Even in other sectors, some employers do not want to hire female workers because of problems related to menstruation and pregnancy. For some women, having a menstruation can be very painful which leads to the decrease in the ability to work during the period. However, this does not happen to every woman, and the study suggests that the number of women facing difficulties during menstruation is not significant enough to imply that every woman will have this problem.

For people with higher level of education, the responsibilities male and female take care at work are mostly similar. Therefore, employers tend not to consider sex of the applicants but mostly consider educational degree and knowledge. For example, a male professor and a female professor both have the same job of teaching students. They can both teach in the same department and in the same faculty. In some cases, patriarchy may still have effects on people with higher educational degrees. The study shows that some people still think men are created to be better at Math and programming compared to women. As a result, some employers tend to hire more males as engineers or programmers. Even though many big tech companies are trying to promote diversity in the workplace by hiring more female employees, the idea of diversity in the workplace is still new to most sectors (Howard A. 2021).

4.4   Weaknesses

The weakness of this paper is that the data used does not show all the responses to the survey row by row. Instead, it only shows the employment rates which was calculated from the responses. As a result, there are limitations when exploring and plotting data as the employment rate is the only numerical value that can be used. Also, there are only 5 variables in this dataset which limits the number of aspects that can be explored in this dataset.

4.5   For the future

In order to explore more about this data, a supplemental survey (See Appendix A) was designed and created to collect more data.

The survey was created to be sent out online for people to fill in by themselves. To begin with, since the data downloaded from Statistics Canada website does not show the age group of the participants in the same size interval, the supplemental survey will ask participants to select one age group according to their real age. Next, we are interested in learning more about diversity in the workplace especially for gender and ethnic group. Participants will be asked to identify themselves, but there will also be a choice for participant

to select if they prefer not to answer. Also, we are interested in looking into each sector. Participants will be asked to name to sector or industry they are working in. Last, we are interested in comparing the employment rate between provinces in Canada. Participants will be asked to locate their province of residence and province they are working in.

The data collected from the survey will be used to further explore the factors affecting employment rate in Canada. As we will get raw data from the participants, more data cleaning process will be required. For example, there might be participants who did not finish answering all the questions before submitting which leads to some values missing from the data. We will need to deal with that row of data either by contacting the participant to ask or drop that row of data. After cleaning the data, some summary statistics will be used to further explore the data which will depend on the distribution and the characteristics of the data we collected. Next, data will be used to plot graphs which will help us to easily see the underlying insights behind the data.

We expect to see the comparison of the employment rate as follow:

- The employment rate between age groups
- The employment rate between self-identified genders
- The employment rate between status in Canada
- The employment rate between ethnic groups
- The employment rate between sectors
- The employment rate between provinces in Canada

This data should be able to help us understand more about Canada employment rate as it gives more details and it covers factors that are often mentioned in studies about diversity in the workplace.

Appendix

A   Supplemental Survey

Supplemental Survey to the Labour Force Survey

Please complete this form to participate in the survey

samita.prabhasavat@gmail.com (not shared) Switch accounts

Informed Consent

Overview
Our team consists of Luckyna Laurent, Samita Prabhasavat, and Zoie So. We are students in the Bachelor of Information program at the University of Toronto. Thank you for taking part in our research study, which is about understanding factors affecting Canada labour market such as age group and gender. To participate in the study, you will complete a survey about yourself and your employment information. The survey will take about 7 questions long and will take approximately 10 minutes to complete.

What will my participation involve?
Your participation in this study will involve the completion of this survey.

How will my confidentiality be protected?
The research team will not know who has consented to participate in the study. All identifying information will be removed from the data before it is analyzed and any findings are made public. All data will be stored securely, and access limited to the investigators. To be clear, we intend to publish and make public presentations based on this study. We further intend to release the dataset on which our research is based. However, we will remove any identifying information before the data are used for research and identifying information will never be released.

Inclusion/Exclusion criteria?
You must be over the age of 15 and is currently working in Canada.

Can I withdraw from the study?
You may refuse to participate, withdraw at anytime, and may decline to participate in some part of the study - all without any negative consequences. If you decide to withdraw from the study, you may do so at anytime before the data have been published.

How can I learn more about the study?
You can contact the primary investigator: Samita Prabhasavat, samita.prabhasavat@mail.utoronto.ca .

Do you consent to sharing your data with the primary investigator (Samita Prabhasavat), for the purposes described above?

○ Yes, I authorize the use of the data collected about me through registration and subsequent surveys to be used as part of the research study and wish to register for the study.

○ No, I do not want my data to be included in the study, the data collected about me in this form .

Which category below includes your age?

○ 15 - 19
○ 20 - 24
○ 25 - 29
○ 30 - 34
○ 35 - 39
○ 40 - 44
○ 45 - 49
○ 50 - 54
○ 55 - 59
○ 60 - 64
○ 65 or older

How would you describe your gender identity?

○ Male
○ Female
○ Prefer not to say
○ Other: _____

What is your status in Canada?

○ Citizen
○ Permanent Residence
○ Work Permit
○ Study Permit
○ Other: _____

How would you identify yourself?

○ White
○ Black
○ Asian
○ Indigenous
○ Other: _____

How would you describe your gender identity?

○ Male
○ Female
○ Prefer not to say
○ Other: _____

What is your status in Canada?

○ Citizen
○ Permanent Residence
○ Work Permit
○ Study Permit
○ Other: _____

How would you identify yourself?

○ White
○ Black
○ Asian
○ Indigenous
○ Other: _____

Which province are you working in?

[Choose ▼]

Which province is your province of residence?

[Choose ▼]

Thank you for completing our survey.

References

Canada, Statistics. 2021. "Labour Force Survey." Statistics Canada. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1295903.

Chen Ying-Yeh, Lui Carrie, Chen Mengni. 2017. "Female Labour Force Participation and Suicide Rates in the World." Social science & medicine. https://doi.org/10.1016/j.socscimed.2017.11.014.

Duggan. 2019. "Education Policy, Digital Disruption and the Future of Work : Framing Young People's Futures in the Present." Springer International Publishing. https://doi.org/10.1007/978-3-030-30675-5.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Howard A., Isbell C. 2021. "Diversity in Ai: The Invisible Men and Women." Sloan Management Review. https://librarysearch.library.utoronto.ca/permalink/01UTORONTO_INST/fedca1/cdi_gale_infotracacademiconefile_A657446650.

Moja. 2021. "National and Institutional Responses - Reimagined Operations - Pandemic Disruptions and Academic Continuity for a Global University." Studies in higher education (Dorchester-on-Thames). https://doi.org/10.1080/03075079.2020.1859688.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer.

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. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2016. *Bookdown: Authoring Books and Technical Documents with R Markdown.* Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/bookdown.