

Degrees of Difference: Analyzing the Impact of Education on Earnings in Canada*

Yuanyi (Leo) Liu

April 18, 2024

In this paper, we explore the connection between education and hourly wages in Canada, focusing on data from the year 2000. Our findings indicate a clear trend: higher educational levels correlate with increased average hourly wages for individuals aged 25 to 54. This research highlights the importance of education in determining earning potential and suggests that investment in education could have long-term economic benefits. The study provides evidence for policy implications regarding educational incentives and workforce development strategies in Canada.

Table of contents

1	Introduction	2
2	Data	3
2.1	Source and Methodology	3
2.2	Variables	3
2.3	Measurements	5
3	Model	5
3.1	Model set-up	6
3.2	Model justification	6
4	Results	6
5	Discussion	8
5.1	Findings	8
5.2	Insights on Education and Wages	9

*Code and data are available at: <https://github.com/leoyliu/Analyzing-the-Impact-of-Education-on-Earnings-in-Canada>

5.3	Limitations	9
5.4	Future Research Directions	9
	Appendix	10
	A Data Manipulation and Cleaning	10
	References	12

1 Introduction

The relationship between education and earnings is a well-established topic of interest within the field of labor economics, providing insights into the broader socio-economic fabric of a country. As Canada’s economy continues to diversify and specialize in various sectors, the value of education in this landscape remains a critical question for policymakers and the public. While previous studies have explored this link, there is an ongoing need to update and deepen our understanding of how this dynamic plays out in the modern economy.

This paper narrows the focus to the Canadian labor market, where the interplay between educational achievement and wages within the core working age group, those aged 25-54, offers a mirror to the efficacy and value of educational advancements in the workplace. This demographic is selected for its relevance to the labor force and policy implications, excluding younger individuals still likely in the education system and older individuals, who may present outlier educational experiences that do not align with the central workforce.

This investigation seeks to address a gap in the current literature by providing a targeted analysis of the 25-54 age group in Canada since the year 2000, a period marked by rapid technological change and economic evolution. By applying a linear model to wage data categorized by education level, this study uncovers a positive correlation between educational attainment and hourly wages, suggesting that higher education can be linked to improved wage outcomes in this key demographic.

In this study, our primary focus is to estimate the effect of education on hourly wages in Canada. Our estimand is the incremental average hourly wage rate for individuals within the specified education levels, compared to the next lower education level. By quantifying this effect, we aim to capture the economic value of educational attainment and how it translates into wage premiums across various stages of a typical Canadian worker’s career span.

The structure of the paper is organized as follows: Following Section 1, Section 2 presents the data, detailing the data sources, analytical techniques, and the rationale behind the chosen methods. Section 3 then delves into the specifics of the linear model analysis, laying out the statistical underpinnings that support our investigation. After that, Section 4 discusses the results, elaborating on the observed trends and patterns in wage rate data. Section 5 interprets these findings in light of the current economic and educational context in Canada,

exploring potential factors influencing these trends, drawing connections to broader socio-economic issues, and providing suggestions for future research in this area.

2 Data

This section aims to offer an insightful understanding of the dataset utilized in our analysis, which serves as the foundation for our examination of the relationship between education and average hourly wage rates in Canada, specifically within the 25-54 age demographic.

2.1 Source and Methodology

This study utilizes a dataset sourced from the Open Government Portal of Canada [], specifically designed to track the correlation between educational attainment and hourly wages across various demographic segments of the Canadian workforce. The dataset covers the period from the year 2000 onwards, providing a longitudinal view of wage trends in relation to educational background.

Alternative datasets, such as those from Statistics Canada’s Labour Force Survey, were considered but were not selected due to their less detailed categorization of education levels and their broader focus on employment without specific wage breakdowns by education.

The data was processed and cleaned using R(R Core Team 2020), a powerful statistical programming language. The selection of this dataset was motivated by its direct relevance to the research question and its regular updates, ensuring that the data remains relevant for observing current trends. Initial data processing involved filtering the dataset to focus solely on entries categorized under ‘Canada’ in geography, eliminating data irrelevant to the national focus of this research. This step was crucial to maintaining the clarity and relevance of the analysis. For key operations, please refer to the Section [A](#).

2.2 Variables

To better understand data, key variables extracted for this study include the year of data collection, education level, age group, and the average hourly wage rate. Notably, we have streamlined the ‘Education level’ from a broad range of categories into a numeric variable that aligns with ascending educational attainment, facilitating a quantitative analysis of its impact on wages. The age group has been limited to the 25-54 years range to focus on the most economically active segment of the population, avoiding potential outliers from younger individuals with less work experience and older individuals whose advanced education, such as doctoral degrees, could distort the analysis. Wage rates are presented in Canadian dollars per hour, reflecting pre-tax earnings, inclusive of tips, commissions, and bonuses.

Table 1: First Ten Rows of Wage Rates Spanning from 2000 to 2019 For Cleaned Data

Year	Education Level	Age	Hourly Wage
2000	0 - 8 years	25-54 years	13.1
2000	High school graduate	25-54 years	15.8
2000	Post-secondary certificate or diploma	25-54 years	18.0
2000	Trade certificate or diploma	25-54 years	17.4
2000	Community college, CEGEP	25-54 years	18.1
2000	University certificate below bachelors degree	25-54 years	20.2
2000	University degree	25-54 years	23.4
2000	Bachelor's degree	25-54 years	22.5
2000	Above bachelor's degree	25-54 years	25.6
2001	0 - 8 years	25-54 years	13.3

Table 1, created with `kableExtra` (Zhu 2021), outlines the first ten records of our dataset, featuring average hourly wages for Canadians aged 25-54 across varying levels of education from the years 2000 to 2019. This snapshot reveals the variable structure of our analysis, which includes the `Year` of wage data, the defined `Education level` ranging from “0 - 8 years” to “Above bachelor’s degree,” and the consistent `Age group` focus. It provides an early indication of the ascending trend in wages with higher educational attainment, setting the stage for our deeper investigation into the economic value of education within Canada’s workforce.

To transition from a tabular overview to a more graphical representation, we will visualize the data, which will allow us to spot patterns and trends that are not immediately apparent in numerical form.

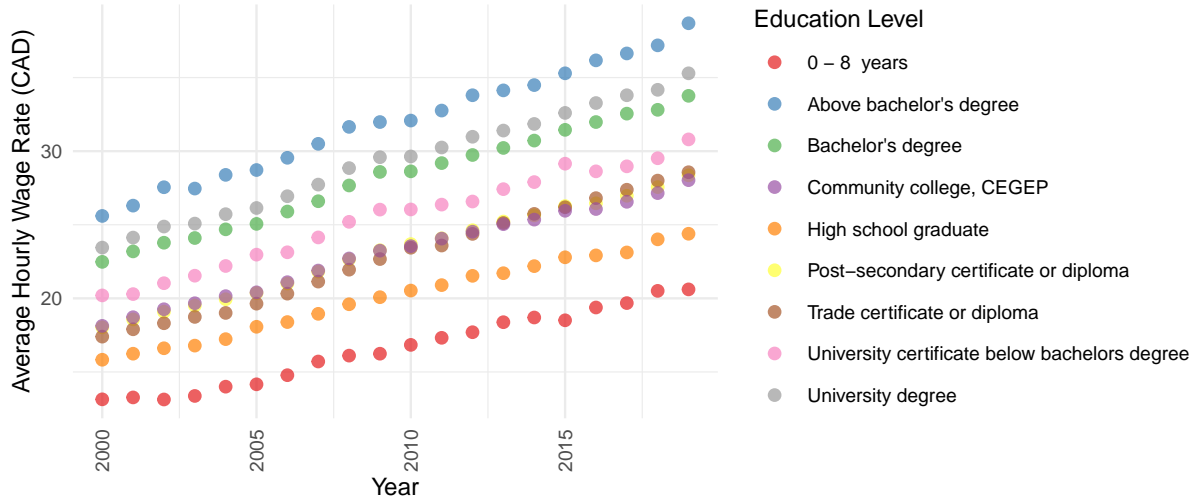


Figure 1: Visualization of Wage Rate from 2000 to 2019

Figure 1 illustrates the distribution of average hourly wage rates across different educational levels from 2000 to 2019 for Canadian workers aged 25-54. Each dot represents the wage rate for a specific education level in a given year, with color coding to distinguish between educational categories, ranging from “0 – 8 years” to “University degree.” The spread and upward trend of the points suggest that individuals with higher educational qualifications tend to have higher average hourly wages, and this trend persists over the 20-year period.

2.3 Measurements

The measurement of variables within this dataset was handled to ensure accuracy and relevance. The ‘Average Hourly Wage Rate’ is calculated based on actual wages reported by employers, providing an objective measure of income. This data is then vetted and standardized by statistical agencies to ensure it reflects accurate and fair representations of wages across different demographics.

For educational attainment, data entry occurs as individuals enter the workforce or update their qualifications, with categorizations reflecting the highest level of formal education completed. This data is collected through surveys and employment records, often as part of broader demographic data collection efforts by governmental agencies.

For an in-depth technical breakdown of the statistical analyses and additional visualizations, readers are directed to the appendices, where these aspects are discussed in detail. This ensures that the main text remains accessible while still providing resources for more technically inclined readers.

Correlations between the variables will be examined at Section 3, providing insights into how education levels are associated with hourly wages. By understanding these relationships, we can better infer the potential impact of education on earnings within the Canadian economy.

3 Model

The objective of our modeling approach is to quantify the relationship between educational attainment and hourly wages. Our analysis employs a Bayesian framework to assess how changes in the level of education correlate with variations in wage rates among Canadian workers aged 25-54.

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [?@sec-model-details](#).

3.1 Model set-up

Define y_i represent the average hourly wage rate for the i^{th} individual. The predictor variable, x_i , corresponds to the numeric value assigned to each education level. The model can be described by the following equations:

$$\begin{aligned}y_i | \mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) \\ \mu_i &= \beta_0 + \beta_1 x_i \\ \beta_0 &\sim \text{Normal}(0, 2.5) \\ \beta_1 &\sim \text{Normal}(0, 2.5) \\ \sigma &\sim \text{Exponential}(1)\end{aligned}$$

We run the model in R (R Core Team 2020) using the `rstanarm` package of (`rstanarm?`). We use the default priors from `rstanarm`.

The model is executed in R (R Core Team 2020) using the `rstanarm` package (`rstanarm?`). Default priors from `rstanarm` are utilized, set with a mean of zero and a conservative standard deviation, reflecting no strong prior beliefs about the magnitude of the education coefficients.

3.2 Model justification

Given the economic theory and previous empirical findings, we anticipate that higher educational qualifications would correspond to higher hourly wages. This is encapsulated in the assumption that β_i should have a positive effect on y_i . Through Bayesian analysis, we aim to capture the degree to which educational attainment can predict wage rates, thus providing empirical evidence to support or refine this hypothesis.

Mathematical notation is incorporated using LaTeX syntax, enabling precise formulation of our statistical model. For example, the average effect of education on hourly wage rate is denoted by β .

4 Results

Section 4 explores the relationship between education and average hourly wages in Canada. Utilizing a dataset that captures 20 years of wage data across varying educational levels, we apply a linear modeling approach to uncover patterns and draw conclusions about the economic value of educational advancement. Below, we present the results of our model and its implications.

The results obtained from our linear model, which predicts the average hourly wage rate based on the numeric representation of education levels, show a clear and positive relationship between educational attainment and wage rate. Table 2 indicates that the intercept is estimated

Table 2: Model Summary

	Linear Model
(Intercept)	16.143
Education_numeric	1.672
Num.Obs.	180
R2	0.610
R2 Adj.	0.607
Log.Lik.	-478.097
ELPD	-480.2
ELPD s.e.	7.1
LOOIC	960.5
LOOIC s.e.	14.2
WAIC	960.4
RMSE	3.43

at 16.143, suggesting that when the education level is at the baseline (0 - not accounted for in the numeric scale), the average hourly wage rate would be approximately 16.14. Importantly, the coefficient for Education_numeric is positive (1.672), supporting the hypothesis that higher educational levels correlate with higher wage rates. The model used 180 observations for the analysis, with an R-squared value of 0.610, which means that approximately 61% of the variance in the average hourly wage rate is explained by the model. The adjusted R-squared, which accounts for the number of predictors in the model, is very close to the R-squared, indicating a good fit without overfitting. Notably, the RMSE (Root Mean Square Error) of 3.43 reflects the standard deviation of the prediction errors, which measures how spread out these errors are around the true regression line.

Figure 2 illustrates the fitted linear regression line across the spread of average hourly wage rates for different education levels over the years 2000 to 2019. Each point in the scatter plot corresponds to the average hourly wage rate for a given education level, differentiated by color to represent each year. The dashed line across the plot signifies the best-fit linear regression line, which inclines upward as the education level increases. The diversity of colors along the regression line also reveals the variation in wage rates across different years, although the increasing trend is consistent throughout.

In conclusion, the results presented align with the initial hypothesis, showing that as education level increases, there is a corresponding increase in the average hourly wage rate. This validates the underlying assumption that more advanced education tends to lead to better-paying job opportunities. The analysis effectively captures the essence of the data, allowing for a meaningful interpretation that leads us into the discussion. In the next Section 5, we will delve into the implications of these findings, examining their relevance to current labor market theories and the potential for informing education and economic policy decisions.

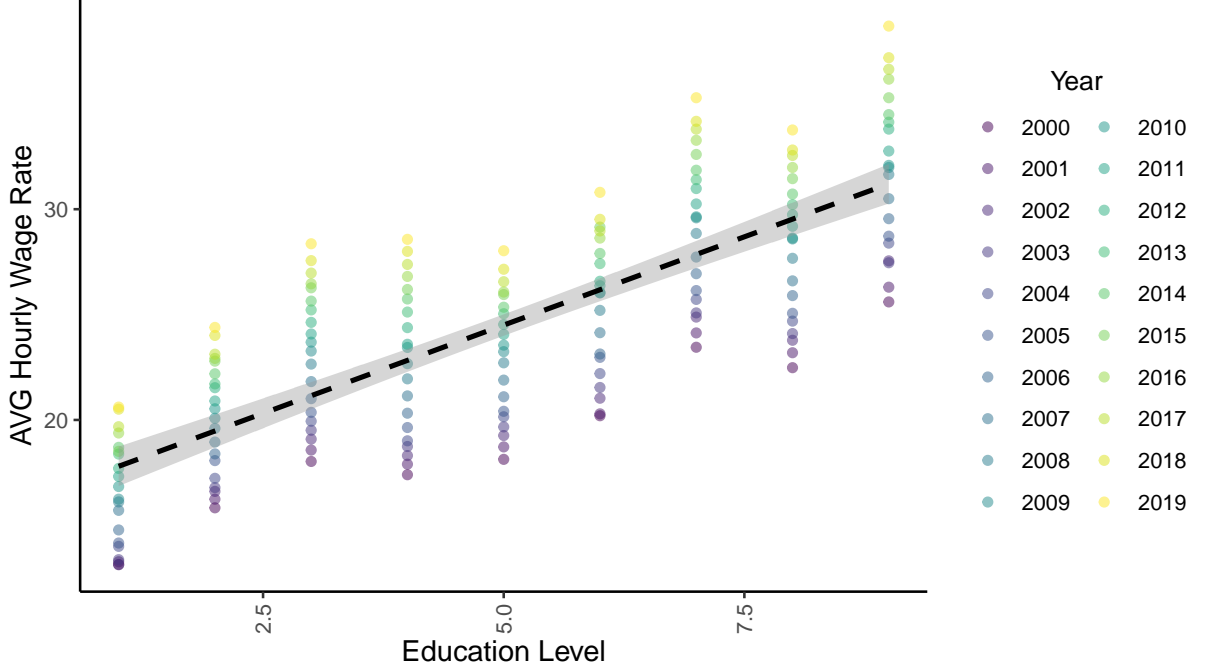


Figure 2: Model Visualization

5 Discussion

In this paper, we investigated the relationship between educational attainment and average hourly wage rates from 2000 to 2019. By deploying a linear Bayesian model, we have been able to highlight the correlation between higher educational attainment and increased wage rates, aligning with our original hypothesis.

5.1 Findings

Our findings reinforce the perception of education as an investment yielding tangible financial returns. The analysis showcases a clear, positive trajectory in wage rates corresponding with ascending educational levels. This outcome underlines the vital role of education in driving economic growth, suggesting that educational enhancements could be pivotal in strategies aimed at combating poverty and fostering economic resilience.

Further examination reveals the potential of education as a tool for social mobility. The data indicates that as individuals climb the educational ladder, they unlock higher earnings, which could lead to a more equitable distribution of income. These insights could serve as a valuable asset for policymakers striving for educational equity and highlight the importance of making higher education accessible to all strata of society.

5.2 Insights on Education and Wages

The positive correlation observed in this study goes beyond confirming a long-suspected link between education and wages; it also points to the complex interplay of various wage determinants. Despite the clear trend, the dispersion of wage data around the regression line suggests multiple underlying factors at play. This finding suggests a need for a comprehensive approach to wage enhancement, integrating elements such as vocational training, relocation incentives, and specialized educational programs.

Expanding on the above, it is evident that education serves as one of many pillars supporting an individual's earning capacity. While education remains a significant factor, the disparity between the predicted and observed wages implies other variables are also influential. Recognizing this complexity is crucial for a complete understanding of how education impacts wage rates and the formulation of holistic policies that address the multifaceted nature of the job market.

5.3 Limitations

The primary limitation of this study is its relatively small sample size, with only 180 observations. This constraint may impact the robustness and generalizability of the findings. Additionally, the focus on the age group of 25-54 years might not encapsulate the full spectrum of the workforce, especially considering the varied career stages and life circumstances of younger and older workers. Such a limited scope might overlook the nuanced effects of education on wages across different life stages.

5.4 Future Research Directions

Future research should aim to disaggregate the age groups to understand the differential impact of education on wages across a more granular age spectrum. Distinguishing between early-career individuals, mid-career professionals, and late-career individuals could yield more nuanced insights into the education-wage dynamic at different life stages. Additionally, expanding the sample size and including more varied age groups would enhance the representativeness and depth of the analysis.

Incorporating a more comprehensive range of variables and exploring more complex, perhaps non-linear models, would allow for a better understanding of the multifarious factors influencing wages. Such an enriched analysis could serve as the foundation for more effective, tailored economic policies that address both the direct and indirect ways education can influence economic outcomes.

Appendix

A Data Manipulation and Cleaning

Most of the data in our dataset was previously cleaned for the project we are replicating. Thus, `?@fig-trends-in-birth-rate` directly used data from `outputs/data/fig_1.csv` without any further cleaning necessary, as it simply displays the birth rate from 1980 to 2020. Thus, the cleaned data is directly stored to `outputs/data/fig_1.csv` from `inputs/data/fig_1.csv`.

In `?@fig-trends-in-birth-rate-in-young-people`, the dataset was imported from `inputs/data/fig_2.csv` and used `dplyr` (Wickham et al. 2023) to select the birth rates of the six age groups (from 15 to 44). Furthermore, the selected data was renamed using `tidyverse` (Wickham et al. 2019) and changed into names that represent the content of the data better. Moreover, these data is then pivoted using `tidyverse` (Wickham et al. 2019) in order to be graphed properly using `ggplot2` (Wickham 2016).

In `?@fig-birth-rate-comparison-old-and-young`, the first dataset was imported from `inputs/data/fig_3.csv` and the dataset containing American states information was from `mapdata` (Richard A. Becker and Ray Brownrigg. 2022). The first dataset was first sliced to remove birth rate information regarding the states of Alaska and Hawaii, as these two states can't be properly shown in the dataset from `mapdata` (Richard A. Becker and Ray Brownrigg. 2022). Then, a temporary data frame is created with tibble from `tidyverse` (Wickham et al. 2019) with the abbreviated state names from `inputs/data/fig_3.csv` and the state full names. Moreover, the abbreviated state name in the first dataset is switched with the state full name using `left_join` from `dplyr` (Wickham et al. 2023), which is then selected and renamed using `tidyverse` (Wickham et al. 2019). After this, the adjusted dataset with the state full names and birth rate is merged with the second dataset that contains the American states information for further graphing with `haven` (Wickham, Miller, and Smith 2023).

In all three figures, the library `here` (Müller 2020) was used to ensure that the file path should be accessible in all directories.

The transformation from raw to analysis-ready data involved substantial cleaning, including filtering by geography (Canada), type of work (both full- and part-time), and specific wage type (average hourly rate). Moreover, we excluded broad or ambiguous categories like 'Some high school' and 'Total, all education levels' to ensure clarity in our analysis. The resulting 'Education_numeric' variable was constructed using a sorted factorization process, mapping each education level to a corresponding numeric value.

In presenting our findings, we will incorporate graphs that visualize the relationship between education and wage rates, providing an intuitive understanding of the data's characteristics. Tables summarizing the observations, along with discussions on the graphs, will be included to elucidate key trends and observations. Summary statistics and any notable relationships between the variables will also be featured, offering a detailed overview of the dataset. Should

the volume of this information warrant it additional detail, these elements may be relegated to appendices to maintain the accessibility of the data section.

References

- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richard A. Becker, Original S code by, and Allan R. Wilks. R version by Ray Brownrigg. 2022. *Mapdata: Extra Map Databases*. <https://CRAN.R-project.org/package=mapdata>.
- 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 Grolmund, 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, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. <https://haven.tidyverse.org>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <http://haozhu233.github.io/kableExtra/>, <https://github.com/haozhu233/kableExtra>.