Determinants of Wage Variation in Canada*

Older Workers, Men, and Higher Education Earn Higher Average Hourly Wages

Yuanyi (Leo) Liu

December 14, 2024

This paper examines wage disparities in Canada by analyzing the effects of education, gender, and age on average hourly earnings. The findings indicate that older workers consistently earn higher wages, men earn more than women across all groups, and higher levels of education are strongly associated with increased wages. These patterns highlight systemic differences in earnings tied to demographic and educational factors. Understanding these disparities provides important context for policymakers aiming to address wage inequality and improve economic outcomes.

Table of contents

*Code and data are available at: Determinants of Wage Variation in Canada.

1 Introduction

Wages are a fundamental component of economic well-being, influencing individual livelihoods, workforce productivity, and broader societal equity. Understanding the factors that shape earnings is essential for addressing wage inequality and improving economic opportunities. In Canada, wage disparities persist across demographic and educational groups, raising important questions about the relationship between age, gender, and education and their roles in determining hourly wages.

This paper analyzes how average hourly wages in Canada vary based on three key factors: age, gender, and education level. Using data spanning two decades, we employ a linear regression model to quantify the relationships between these factors and earnings. The analysis examines whether older workers earn higher wages, how gender influences pay, and the extent to which education impacts earnings. These questions are important in the ongoing concerns about income inequality and the need for policies that promote equitable pay structures. While previous studies have documented broad patterns of wage inequality, this paper offers a focused examination of specific demographic and educational factors, providing a deeper understanding of their combined effects.

In this study, our estimand is the average hourly wage rate in Canada. The object of the estimation is the average hourly wage rate across different demographic and educational groups based on the data. By modeling these relationships, the study quantifies the extent to which these factors contribute to observed wage disparities, providing a detailed understanding of their relative influence within the Canadian labor market.

The findings demonstrate that wages increase with age, particularly for workers over 55, while men consistently earn more than women across all groups. Higher levels of education are strongly associated with higher wages, with university graduates earning substantially more than those with lower educational attainment. These results highlight persistent disparities that reflect both structural and systemic influences on earnings. By examining these relationships, this paper provides a basis for understanding wage inequality in Canada and informs discussions about potential policy interventions.

The structure of the paper is organized as follows: following Section 1, Section 2 presents the data collection and cleaning process, along with an overview of the variables used in the analysis. Section 3 explains the chosen model and why it is appropriate for modeling average hourly wages. Then, Section 4 provides the results, highlighting key trends and predictions. Eventually, Section 5 concludes with a discussion of the findings, addressing wage disparities, potential strategies to improve equity in Canadian earnings, and the limitations of the model used.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to process and analyze data on average hourly wages in Canada. The dataset (Statistics Canada 2020), published on January 11, 2020, covers data from January 1, 1997, to December 31, 2019, but for this analysis, we focus on data from 2000 onwards to ensure relevance to recent labor market trends. The dataset is maintained annually and provides detailed information on wages, demographics, and employment characteristics, making it well-suited for studying wage disparities in Canada. Following methodologies discussed in "Telling Stories with Data" (Alexander 2023), we analyze wage patterns by aggregating data across multiple demographic groups to ensure a balanced and unbiased representation of wage disparities. For key operations, please refer to Appendix B.

The dataset includes variables such as age groups (15+, 25+, 25-34, 25-54, and 25-64), types of work (full-time and part-time), educational levels (e.g., 0-8 years, high school graduate, post-secondary certificate diploma, university degree), and immigration statuses (e.g., very recent immigrants, recent immigrants, established immigrants, non-landed immigrants, and Canadian-born individuals). Wages are provided as both average weekly and average hourly rates. For this analysis, we focus on the age groups "15-24 years", "25-54 years", and "55 years and over", the combined category of full- and part-time workers, and specific education levels, such as high school graduate, post-secondary certificate diploma, and university degree.

This filtering excludes ambiguous or aggregate categories such as "some high school," "some post-secondary," and "total landed immigrants" to ensure a clear and interpretable dataset. This filtering allows us to analyze relationships between wages and specific demographic factors while maintaining focus on education, gender, and age. Alternative datasets, such as those focusing on industry-specific wage trends or detailed time-series data, were not used because they do not provide the necessary demographic granularity. This dataset is uniquely positioned to address our research questions by capturing a broad view of wage disparities in Canada.

2.2 Measurement

Wage data is a measurement of economic outcomes based on information collected through the Labour Force Survey Special Tabulations conducted by Statistics Canada. Respondents' demographic information, including age, gender, education level, and employment type, is reported through standardized surveys administered to a representative sample of the Canadian workforce. Wage data, including both hourly and weekly rates, are derived from employer records and worker-reported earnings, ensuring broad coverage across different industries and occupations. More details on survey methodologies can be found in the Labour Force Survey Special Tabulations.

Each row in the dataset represents an aggregated statistic for a specific combination of age group, education level, gender, and type of work. For example, the average hourly wage rate for "25-54 years" old male workers with a "university degree" is calculated by pooling the wages of individuals who fit this description, smoothing out individual variations to provide a general trend for this subgroup. By categorizing variables such as education levels into distinct groups (e.g., "high school graduate," "post-secondary certificate diploma"), the dataset allows for consistent comparisons across demographic segments.

2.3 Outcome variables

2.3.1 Hourly Wages

The primary outcome variable in this study is the `Average_hourly_wages`, measured in Canadian dollars. This variable reflects the mean earnings of individuals across various demographic and employment categories. The hourly wage data are derived from the Labour Force Survey, which collects self-reported income information and supplements it with employer records where available. These averages are calculated by dividing reported weekly wages by the number of hours worked, standardizing earnings across full-time and part-time workers.

Table 1: Summary Statistics for Average Hourly Wages in Canada (2000–2019), Including Minimum, Maximum, Mean, Median, and Standard Deviation

Minimum Wage	Maximum Wage	Mean Wage	Median Wage	Standard Deviation
7.6	43.39	21.48	20.48	7.4

Table 1 provides an overview of the distribution of average hourly wages in Canada from 2000 to 2019. The wages range from a minimum of \$7.60 to a maximum of \$43.39, with a mean of \$21.48 and a median of \$20.48, indicating that most wages are concentrated around the middle of the distribution, while a smaller proportion of higher earners pushes the maximum upward. The standard deviation of \$7.40 reflects notable variability, emphasizing differences in earning potential influenced by factors such as education, gender, and age.

Using package `ggplot2` (Wickham 2016), Figure 1 presents the trend of average hourly wages in Canada from 2000 to 2019. Each dot represents the average wage for a specific year, and the overall shows a upward trend. The consistent increase in wages suggests improvements in earnings across the labor market, potentially driven by inflation, economic growth, and changes in workforce composition.

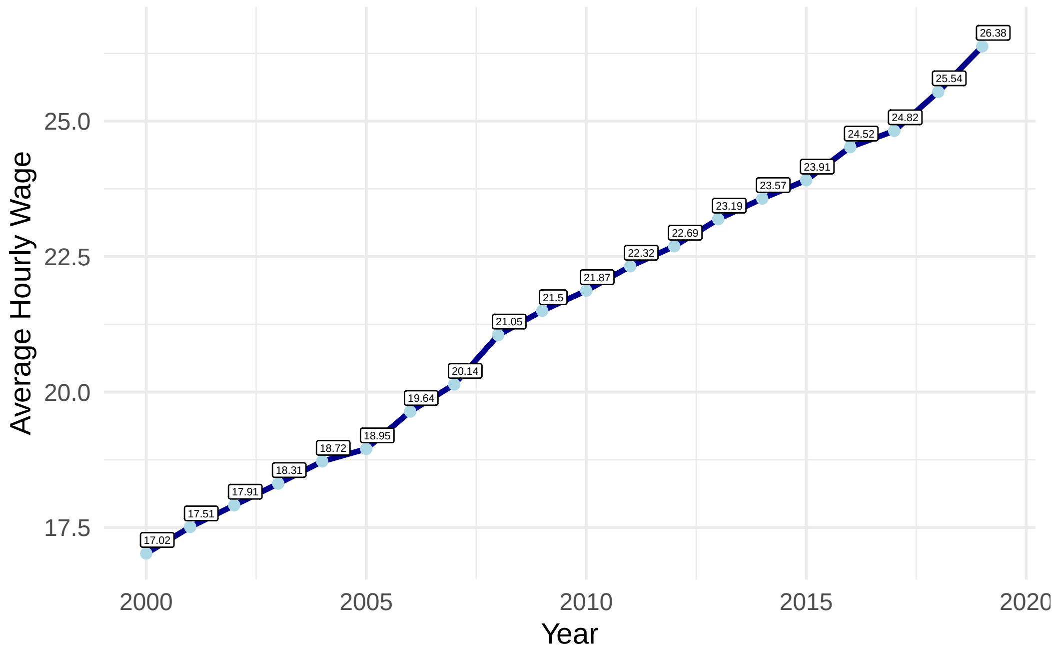

Figure 1: Yearly Trend of Average Hourly Wages in Canada (2000–2019): Steady Growth Reflecting Changes in the Labor Market.

2.4 Predictor variables

2.4.1 Education

The `Education_level` variable represents individuals' highest level of formal education attained and is categorized into levels ranging from "0-8 years" to "Above bachelor's degree." This variable reflects the influence of educational attainment on earning potential. The data for education levels are self-reported in the Labour Force Survey, capturing a wide range of qualifications, including high school completion, trade certifications, and university degrees.

Figure 2 presents the distribution of average hourly wages in Canada from 2000 to 2019 across different education levels. Higher education levels, such as "University degree" and "Above bachelor's degree," are associated with higher median wages, while lower education levels, like "0–8 years," show lower wages. The variability within each education level, shown by the spread of the boxes and whiskers, indicates differences in earning potential within these groups. For instance, individuals with an "Above bachelor's degree" earn higher wages on average and exhibit greater wage variability compared to those with "0–8 years" of education.

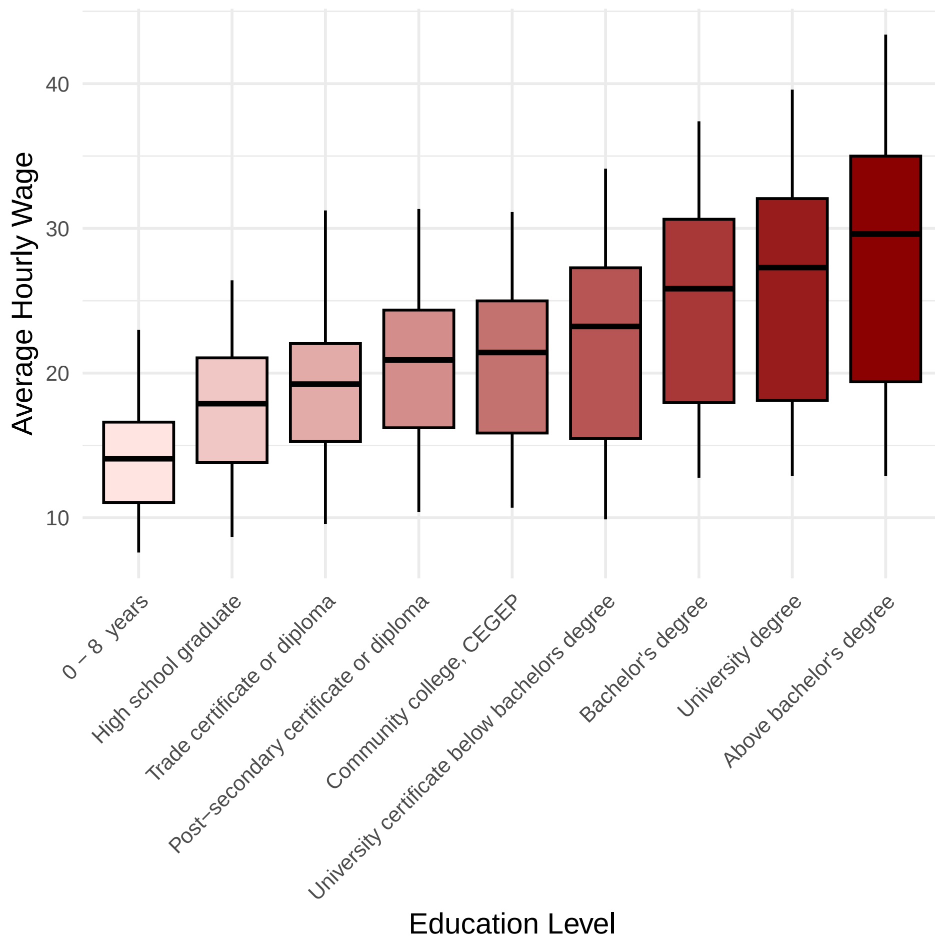

Figure 2: Average hourly wages in Canada by education level from 2000 to 2019. Higher education levels corresponding to higher median wages.

2.4.2 Gender

The `Gender` variable indicates whether an individual identifies as "Male" or "Female". The data are reported as binary classification for simplicity. This variable allows for an investigation of the persistent gender wage gap, capturing differences in average hourly wages for men and women. Including gender in the analysis is essential for highlighting inequalities in the labor market and understanding how these disparities persist across educational levels and age groups.

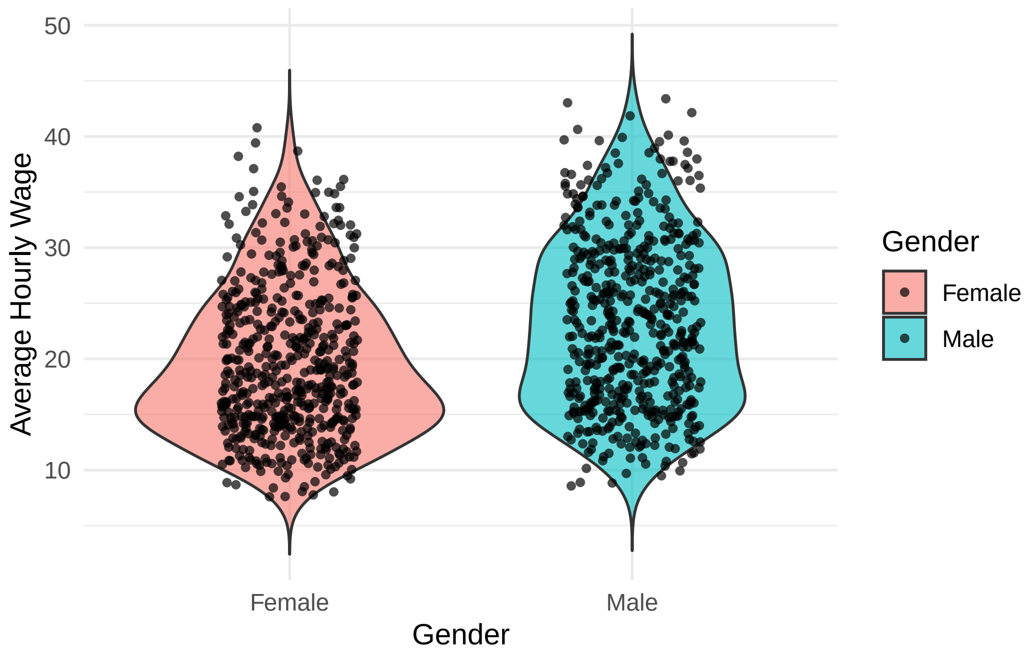

Figure 3: Average hourly wages in Canada by gender from 2000 to 2019. Males showing higher median wages and a broader range of earnings compared to females.

Figure 3 compares the distribution of average hourly wages in Canada from 2000 to 2019 by gender. It displays the spread of wages for each gender, with males generally earning higher wages as shown by the broader and higher distribution. The jittered points represent individual wage observations, further emphasizing the overlap and differences in wage ranges. While both distributions have a similar shape, the median wage for males is higher than that for females, reflecting a persistent gender wage gap over the period analyzed.

2.4.3 Age

The `Age_group` variable divides individuals into three categories: "15-24 years", "25-54 years", and "55 years and over". These groups represent different career stages, from early employment to peak earning years and late career. The Labour Force Survey provides these classifications to standardize analyses of how wages vary with experience and age. Younger workers generally report lower wages due to limited experience, while mid-career and older workers often earn higher wages, reflecting accumulated skills and knowledge over time.

Figure 4: Average hourly wages in Canada by age group from 2000 to 2019. Older age groups showing higher average wages and broader distributions compared to younger groups.

Figure 4 illustrates the distribution of average hourly wages in Canada from 2000 to 2019 across three age groups: "15-24 years," "25-54 years," and "55 years and over." The youngest group, "15-24 years," exhibits the lowest wages, with a narrow distribution concentrated at the lower end. In contrast, the "25-54 years" and "55 years and over" groups show higher wages and broader distributions, reflecting greater variability in earnings as individuals gain experience and qualifications. The older age group generally earns the highest wages, highlighting the strong association between age, experience, and earning potential.

3 Model

Our modeling approach seeks to quantify the relationship between demographic and educational factors and average hourly wages in Canada. To achieve this, we employ a linear regression model to examine how predictors such as education level, gender, and age group influence hourly earnings. The model is implemented using the `stan_lm` function, with a Gaussian distribution to capture the variability in hourly wages.

In this analysis, we focus on predictors that capture key socio-economic and demographic characteristics. Specifically, we include `Education_level`, a categorical variable representing individuals' highest level of education; `Gender`, indicating whether the individual identifies as male or female; and `Age_group`, categorized into "15-24 years," "25-54 years," and "55 years and over." These predictors allow us to explore how differences in education, gender, and age intersect to shape wage outcomes.

The model assumes that the average hourly wage, given these predictors, follows a normal distribution. This Gaussian assumption simplifies parameter estimation and aligns with standard practices in wage modeling. Additionally, we assume moderate priors to avoid overfitting while ensuring interpretability of the coefficients. This balanced approach enables us to identify meaningful relationships between the predictors and wages. Background details and diagnostics are included in Appendix C.

3.1 Alternative model

Initially, the `Education_level` variable was grouped into broader categories, such as "Low," "Medium," and "High" education levels, to simplify the analysis. This approach aimed to reduce the model's complexity while capturing general trends in wage variation. However, this categorization diminished the model's ability to detect differences in wages associated with specific education levels. For instance, combining "Bachelor's degree" and "Above bachelor's degree" into a single category masked the wage attributed to higher education, resulting in a less precise analysis.

An alternative approach was also explored by excluding the `Age_group` variable, based on the hypothesis that its effects might overlap with those of education and gender. However, this exclusion led to a poorer model fit, as it failed to account for wage differences across age groups. Retaining `Age_group` as a distinct predictor improved the model's performance, offering a more accurate depiction of how wages vary across life stages.

3.2 Model set-up

The model predicts the average hourly wage using the following predictor variables:

- Education Level (`Education_level`): A categorical variable representing the highest level of formal education attained by an individual. Levels range from "0–8 years" to "Above bachelor's degree."
- Gender (`Gender`): A binary variable indicating whether an individual identifies as "Male" or "Female."
- Age Group (`Age_group`): A categorical variable representing the age ranges of individuals, categorized as "15–24 years", "25–54 years", and "55 years and over".

The model takes the form:

$$
\begin{aligned}
y_i \mid \mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \beta_0 + \beta_1 \cdot \text{Education level}_i + \beta_2 \cdot \text{Gender}_i \\
&\quad + \beta_3 \cdot \text{Age group}_i + \epsilon_i \\
\epsilon_i &\sim \text{Normal}(0, \sigma^2)
\end{aligned}
$$

Where:

- β_0 is the intercept term.
- $\beta_1, \beta_2, \beta_3$ are the coefficients for each predictor.
- σ^2 is the variance of the error term.

The model is executed in `R` (R Core Team 2023) using the `rstanarm` package (Goodrich et al. 2022). Default priors from `rstanarm` (Goodrich et al. 2022) are used, with the priors set to have a mean of zero and a moderate standard deviation to ensure a reasonable level of regularization.

3.3 Model justification

Existing economic theories and labor market research suggest that education level, age group, and gender notably influence wage outcomes. Higher levels of education are associated with greater specialization and skills, which typically result in higher earnings. Age reflects work experience and career progression, with older individuals often earning more due to accumulated skills and seniority. Gender remains a key determinant, as wage disparities between males and females persist across various labor markets. These predictors collectively capture essential socio-economic dimensions that shape wage structures in Canada.

A Bayesian linear regression model was selected because the dependent variable (average hourly wage) is continuous and approximately normally distributed. This model is well-suited for assessing the contribution of each predictor to wage outcomes while controlling for the effects of others. For instance, the coefficients of education level, age group, and gender directly indicate their respective impacts on average wages.

Further justification for using this model comes from its alignment with the central limit theorem, as the data aggregates wage observations across individuals. Additionally, the predictors align with established labor market theories, giving a solid theoretical underpinning to our model. The inclusion of priors in the Bayesian framework prevents overfitting, balancing interpretability and predictive accuracy.

A key limitation of this model is that it is entirely trained on the analysis dataset, without splitting into training and testing subsets. Each observation in the dataset represents a unique combination of education level, age group, and gender, making it impossible to partition the data without losing combinations. While this approach ensures that all available data contribute to parameter estimation, it restricts the ability to validate predictions on unseen data. To compensate, internal validation methods, such as posterior predictive checks, are employed to evaluate the model's fit and generalizability. Further discussion on this issue is provided in Section 5.

4 Results

Section 4 examines the relationship between education level, gender, and age group with respect to average hourly wages in Canada. Using the analysis dataset, we apply a Bayesian linear regression model to identify the key factors that influence hourly wages. Below, we present the results of our model, visualizations, and key takeaways.

4.1 Model results and interpretation

The linear regression model built on the analysis dataset estimated the effects of education level, age group, and gender on average hourly wages. The results are summarized in Table 2, with a focus on the coefficients and their implications. The intercept, estimated at 5.758, represents the baseline hourly wage for females aged 15-24 years with an education level of "0–8 years."

4.1.1 Education Level

Education level has a strong positive impact on hourly wages. Individuals with "Above bachelor's degree" earn approximately 13.844 CAD more per hour compared to those with "0–8 years" of education. Similarly, holding a bachelor's degree is associated with an increase of 10.726 CAD per hour, and having a university degree contributes an additional 11.740 CAD.

4.1.2 Age Group

Hourly wages vary considerably across age groups. Individuals aged 25–54 years earn 9.630 CAD more per hour than those in the baseline group (15–24 years). Those in the 55+ age group experience an even higher increase of 10.546 CAD, reflecting the wage advantages linked to experience and seniority.

Table 2: Summary of regression model coefficients examining the relationship between education level, age group, gender, and average hourly wages in Canada.

	coefficient
(Intercept)	5.758
Education_levelAbove bachelor's degree	13.844
Education_levelBachelor's degree	10.726
Education_levelCommunity college, CEGEP	6.387
Education_levelHigh school graduate	3.224
Education_levelPost-secondary certificate or diploma	6.132
Education_levelTrade certificate or diploma	5.073
Education_levelUniversity certificate below bachelors degree	7.501
Education_levelUniversity degree	11.740
Age_group25-54 years	9.630
Age_group55 years and over	10.546
GenderMale	3.618
Num.Obs.	1080
R2	0.781
R2 Adj.	0.778
Log.Lik.	−2870.102
ELPD	−2881.5
ELPD s.e.	19.2
LOOIC	5763.1
LOOIC s.e.	38.4
WAIC	5763.1
RMSE	3.45

4.1.3 Gender

The analysis highlights gender disparities in wages, with males earning 3.618 CAD more per hour than females on average. This finding reflects ongoing gender wage gaps, even after accounting for education and age group.

Overall, the model explains a substantial portion of the variance in hourly wages, with an R^2 of 0.781, indicating that 78.1% of the variation is captured by the predictors. The adjusted R^2 of 0.778 supports the accuracy of the model, while the RMSE value of 3.45 suggests a reasonable level of error in predicting wages. These results align with economic theories that emphasize education as a key driver of earning potential, age as a proxy for experience, and persistent gender wage gaps in labor markets. A complete overview of the model and diagnostic details is available in Appendix C.

4.2 Visualization of Results

The relationship between education level, gender, and age group with average hourly wages is illustrated in Figure 5. Each panel represents a distinct age group: "15–24 years," "25–54 years," and "55 years and over." Within each panel, points are color-coded by gender.

Figure 5 presents that wages consistently increase with higher levels of education across all age groups. This trend is evident in the steepness of the linear trendlines, showing that the wage benefits associated with university degrees and qualifications above a bachelor's degree. The gender wage gap is visible in all panels, with males earning higher wages than females at nearly every education level. Notably, the gap widens at higher education levels, suggesting that gender disparities are more significant as qualifications increase.

It also highlights differences in wages across age groups. Older individuals, particularly those aged 55 and over, earn higher wages compared to younger cohorts, reflecting the influence of experience and career progression. Additionally, the intersection of gender, age, and education shows further disparities. For example, males with higher education levels in older age groups consistently exhibit the highest wages, while younger females with lower education levels earn the least.

5 Discussion

5.1 Key Findings

5.1.1 The Impact of Education on Wages

Higher levels of education are strongly associated with increased hourly wages. Individuals with a bachelor's degree or higher earn more than those with lower educational attainment. This

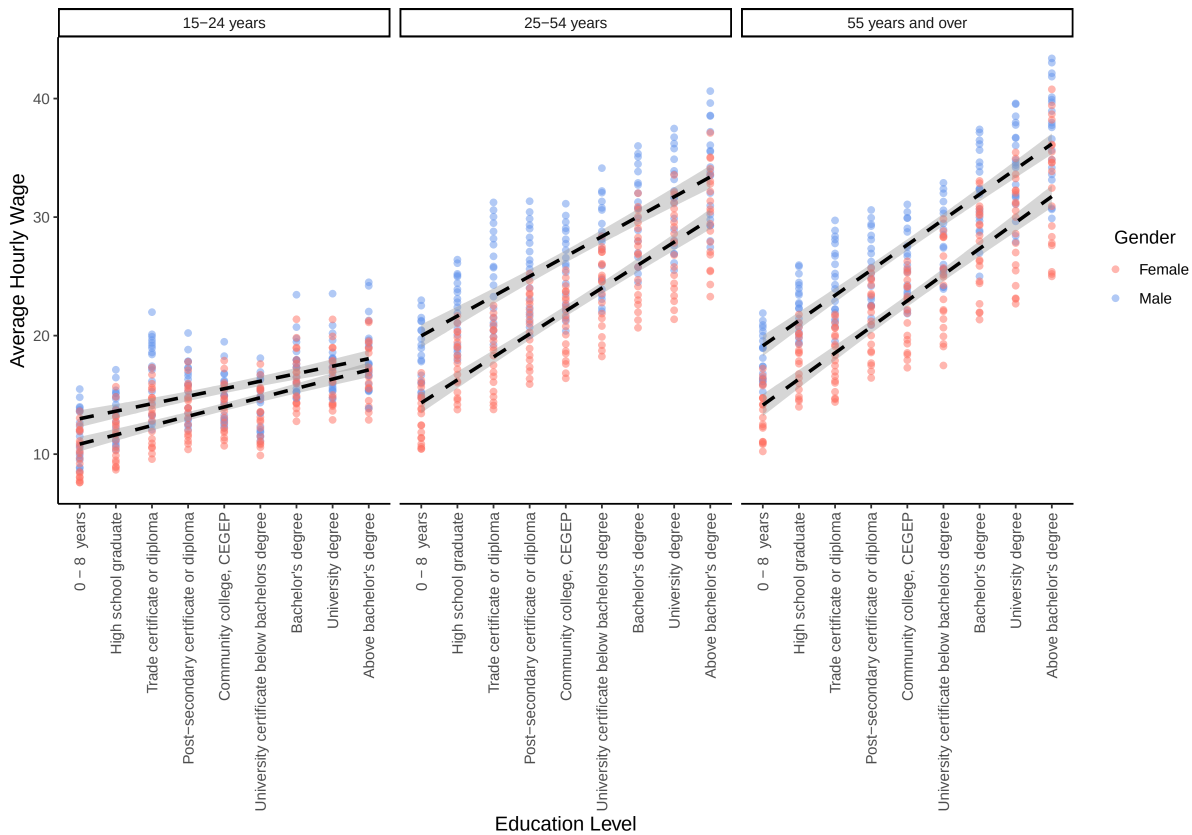

Figure 5: Average hourly wages by education level, gender, and age group in Canada. Wages increase with higher levels of education across all age groups.

trend reflects the demands of a knowledge-based economy, where employers value specialized qualifications and technical expertise. It underscores the role of education in enhancing skills and productivity, which directly translates to higher earnings.

Moreover, the wage premium for higher education shows the importance of equitable access to education. Rising costs of tuition and other barriers could exacerbate socioeconomic inequalities, as those unable to access higher education may remain confined to lower-paying jobs. Policymakers should consider expanding access to affordable education to ensure broader participation in higher-paying sectors.

5.1.2 Age and Experience in Wage Growth

The results reveal a clear wage advantage for middle-aged and older workers compared to younger individuals. Workers aged 25–54 years and 55+ years earn substantially more than their counterparts aged 15–24 years. This trend is likely attributable to experience, seniority, and the accumulation of skills over time, which enhance an individual's productivity and value in the labor market.

In practical terms, this trend reflects a lifecycle wage profile, where earnings typically peak during mid-career and stabilize or slightly decline as workers approach retirement. This understanding is important for retirement planning and addressing income inequality across life stages. Employers could also consider strategies to support younger workers, such as offering training programs and mentorship opportunities, to accelerate their skill development and income growth.

5.1.3 Persistent Gender Wage Disparities

The analysis presents a consistent wage disparity between males and females, with males earning much more across all education levels and age groups. This persistent wage gap suggests systemic inequities in how labor markets reward work. Factors contributing to this disparity may include occupational segregation, differences in work hours, career interruptions, and implicit biases in hiring and promotion processes.

Addressing this gap requires targeted interventions. Initiatives such as wage transparency, mentorship programs for women, and policies promoting work-life balance could help bridge the divide. Beyond fairness, closing the gender wage gap has broader economic implications, as it could boost household incomes and enhance overall productivity in the economy.

5.2 Limitations and Future Research Directions

This study relies on a dataset that represents unique combinations of education level, age group, and gender, limiting the use of traditional validation methods like splitting data into training and testing sets. Future research could explore alternative approaches, such as bootstrapping or cross-validation, to enhance model.

Additionally, the dataset focuses on a specific subset of the population and does not account for variables such as geographic location, race, or immigration status. These factors likely intersect with education, age, and gender to influence wages, and their inclusion could provide a more detailed understanding of wage disparities.

The findings emphasize the economic returns of education, highlighting the need for investment in accessible, high-quality educational systems. The relationship between age and wages underscores the importance of workforce development and retirement planning. Additionally, addressing gender wage disparities requires systemic changes to ensure equitable rewards for labor.

Future studies could investigate the causal mechanisms behind these trends using longitudinal data or natural experiments. Evaluating the effectiveness of policies, such as wage transparency laws or educational subsidies, could provide actionable understanding. Further research into the intersection of race, immigration status, and other factors with wages would deepen our analysis of labor market dynamics.

Appendix

A Idealized methodology and survey

A.1 Overview

This survey is designed to investigate how demographic and educational factors influence average hourly wages in Canada. With a hypothetical budget of $50,000, the survey methodology aims to capture detailed and representative data from working-age individuals across diverse demographics. The survey focuses on three key variables: education level, age group, and gender, and it seeks to ensure accuracy, minimize bias, and provide coverage of the Canadian workforce.

A.2 Sampling approach

The survey employs a stratified random sampling strategy to ensure that the collected data represents the Canadian working population. This approach helps capture wage patterns across various demographic groups.

A.2.1 Stratification variables

- Age Group: 15–24 years, 25–54 years, 55 years and over
- Gender: Male, Female, and Non-binary
- Education Level: 0 – 8 years, High school graduate, Trade certificate or diploma, Post-secondary certificate or diploma, Community college, University certificate below bachelor's degree, University degree, Bachelor's degree, Above bachelor's degree
- Employment Type: Full-time workers, Part-time workers
- Race: White, Black, Indigenous, Asian, Hispanic/Latino, Other
- Hourly Income: <$15, $15-$29, $30-$44, $45-$59, >$60

A.2.2 Sample Size

The survey targets 20,000 respondents across Canada. This sample size ensures a balance between budget constraints and statistical reliability, with an expected margin of error of ±2.5% at a 95% confidence level.

A.3 Recruitment Strategy

To ensure a diverse and representative sample, the survey incorporates two recruitment strategies:

Online survey panels:

- Partner with reputable online survey platforms, such as SurveyMonkey or YouGov, which have access to extensive databases of respondents.
- Validate participant demographics to match the stratification variables.

Social media advertising:

- Use platforms such as Facebook, LinkedIn, and Instagram to reach underrepresented groups, including younger individuals, minority groups, and non-binary respondents.
- Tailor advertisements to target specific demographics and encourage participation.

Incentives: Participants will be entered into a raffle for gift card prizes, ensuring high participation rates.

A.4 Data validation and quality control

To maintain the integrity of the data, we will take the following measures:

- **Attention checks:** At least two ‘attention check’ questions will be included in the survey to identify inattention or robots.
- **Duplicate detection:** Use email or IP-based verification to prevent duplicate responses. Ensure that each respondent provides only one response set to keep the dataset unique.
- **Response time monitoring:** Track the time it takes respondents to complete surveys. Responses with apparently faster-than-average completion times are labelled as potentially low-quality.
- **Data consistency check:** Analyses the consistency of responses, especially between related questions. Inconsistent answers will be labelled for review.
- **Panel partner validation:** For respondents recruited through an online survey panel, rely on the panel provider’s in-built validation systems. These systems use various quality checks, such as identity verification and response consistency, which ensure high-quality data from panel participants.

A.5 Budget allocation

- Online Panel Recruitment: $25,000
- Social Media Advertising: $15,000
- Participant Incentives: $5,000
- Data Cleaning and Validation: $5,000

A.6 Survey implementation

The survey will be available via Google Forms to ensure broad accessibility and easy data collection. The link to the survey will be distributed via email invitations to panellists and targeted social media adverts. To ensure maximum reach and diversity, multiple channels will be used, including Facebook, Instagram, and other social media platforms. The link to the survey is here.

Distribution plan:

- **Panel invitations:** Members of established survey panels (e.g. Lucid, YouGov) will receive email invitations to participate in the survey.
- **Social media activities:** Targeted adverts will be placed on social media platforms to reach under-represented groups, especially young voters and national minority groups.
- **Follow-up reminders:** An automated reminder email will be sent to participants who have not completed the survey to maximize response rates.
- **Survey monitoring:** Response rates will be closely monitored throughout the data collection period and sampling methods will be adjusted if certain population quotas are not reached.
- **Accessibility:** The survey will be mobile-friendly, which will allow respondents to complete the survey on any device.

A.7 Survey structure

Survey introduction:

Welcome!

We are conducting a survey to understand how demographic and educational factors influence hourly wages in Canada. Your participation is important, and your responses will contribute to the research on wage distribution across various groups.

Please note:

- All responses are confidential and will be used for research purposes only.
- This survey will take approximately 10 minutes to complete.
- Your participation is entirely voluntary, and you may withdraw at any time.
- All questions marked with an asterisk are required.
- There are no right or wrong answers, we appreciate your honest opinions.

If you have any questions or concerns, please feel free to contact our research team at liuyuanyi1208@gmail.com.

As a thank-you for your participation, you will be entered into a raffle to win a gift card. We deeply appreciate the time and effort you contribute to this study.

Survey question:

Part 1: Screening for eligibility

1.Are you currently employed?
* Yes / No
2.Are you aged 15 years or older?
* Yes / No
3.Do you currently reside in Canada?
* Yes / No

Part 2: Demographics

4.What is your age group?
* 15–24 years / 25–54 years / 55+
5.What is your gender?
* Male / Female / Non-binary / Prefer not to say
6.Which of the following best describes your race or ethnicity? * White / Black / Indigenous / Asian / Hispanic/Latino / Other
7.What is your highest level of education?
* 0–8 years / High school graduate / Trade certificate or diploma / Post-secondary certificate or diploma / Community college, CEGEP / University certificate below bachelor's degree / Bachelor's degree / University degree / Above bachelor's degree

Part 3: Employment details

8.Are you employed full-time or part-time?
* Full-time / Part-time
9.In which industry or job sector do you work?
* [Open-ended response]
10.How many hours do you typically work per week?
* [Numeric response]

Part 4: Wage

11.What is your average hourly income before taxes?
* [Numeric response]
12.Does your wage include additional benefits or bonuses?
* Yes / No
13.Do you believe your education level has impacted your wages?
* Yes / No / Not Sure

Part 5: Verify and Consent

14.Please select 'Agree' to verify that you are paying attention.
* Agree / Disagree
15.Do you agree to participate in this survey? Your responses will be kept confidential and

used only for research purposes.
* Yes / No

End Part:

Thank you for your participation!

Your responses are important to us in understanding wage trends in Canada. If you have any questions or would like more information about this research, please feel free to contact us at liuyuanyi1208@gmail.com.

B Data Manipulation and Cleaning

During the data cleaning phase, the R packages `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), and `arrow` (Richardson et al. 2024) were used. The raw data was imported using `read_csv` from the `tidyverse` (Wickham et al. 2019) package. Subsequent operations filtered the dataset to focus on Canadian data, both full-time and part-time. Additionally, the focus was placed on records that reported the average hourly wage rate, while entries with ambiguous educational levels such as "Some high school" and aggregate categories like "Total, all education levels" were excluded to maintain data clarity and relevance.

The analysis focused on data from 2000 onward, ensuring the timeframe was appropriate for the study. Age groups were restricted to "15–24 years," "25–54 years," and "55 years and over," representing distinct workforce demographics. Columns were renamed for clarity, including Year, Education_level, and Average_hourly_wages, enabling easier reference in subsequent analyses.

To facilitate gender comparisons, the dataset was reshaped using the `pivot_longer` function, combining the Male and Female wage columns into a single column labeled Gender, with corresponding wage data stored under Average_hourly_wages. This restructuring allowed for straightforward comparisons across gender groups while maintaining a tidy dataset.

Education levels were ordered logically, from "0–8 years" to "Above bachelor's degree," using the `reorder` function. This ensured consistent interpretation in visualizations and modeling, highlighting wage trends across increasing educational attainment levels.

The final step involved saving the cleaned and structured data. This was done using the `write_csv` function to generate a CSV file for broad compatibility and the `write_parquet` function from the `arrow` (Richardson et al. 2024) package for a more compressed and efficient file format, both of which were stored in the `data/02-analysis_data` directory. In all figures and tables, the library `here` (Müller 2020) was used to ensure that the file path should be accessible in all directories.

C Model details

C.1 Posterior predictive check

In Figure 6, we implement a posterior predictive check, which displays the overlap between the observed data (denoted by y) and the replicated data generated from the model (denoted by y_{rep}). In a well-fitting model, the distribution of these replicated data should resemble the distribution of the observed data. Here, the replicated lines mainly follow the shape and central tendency of the observed data's density. This similarity indicates that the model is capturing the overall pattern of the data reasonably well. Deviations between the replicated and observed lines would indicate areas where the model might not be accurately capturing the data structure.

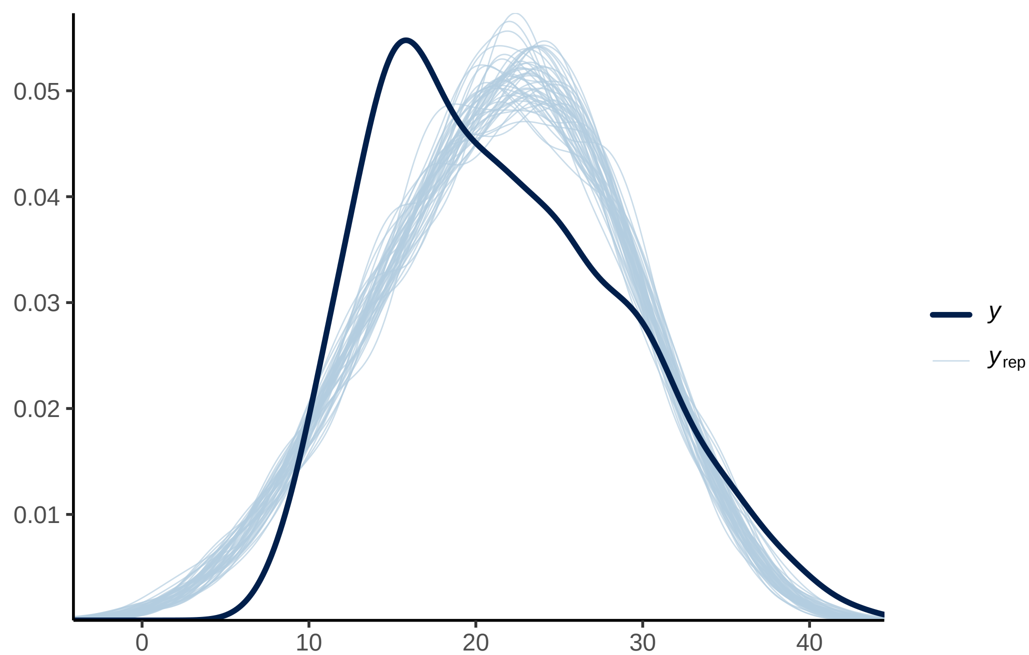

Figure 6: Comparison of Observed Data Density (dark line) and Model-Generated Replications (light lines) shows close alignment, suggesting the model captures the main distributional characteristics of the data.

C.2 Diagnostics

Figure 7 is a visualization of the Gelman-Rubin diagnostic, often denoted as R-hat, which assesses the convergence of a Bayesian model's Markov Chain Monte Carlo (MCMC) samples. It shows the values of for various parameters in the model. An value close to 1 indicates good convergence, meaning the chains for each parameter have mixed well and are sampling from

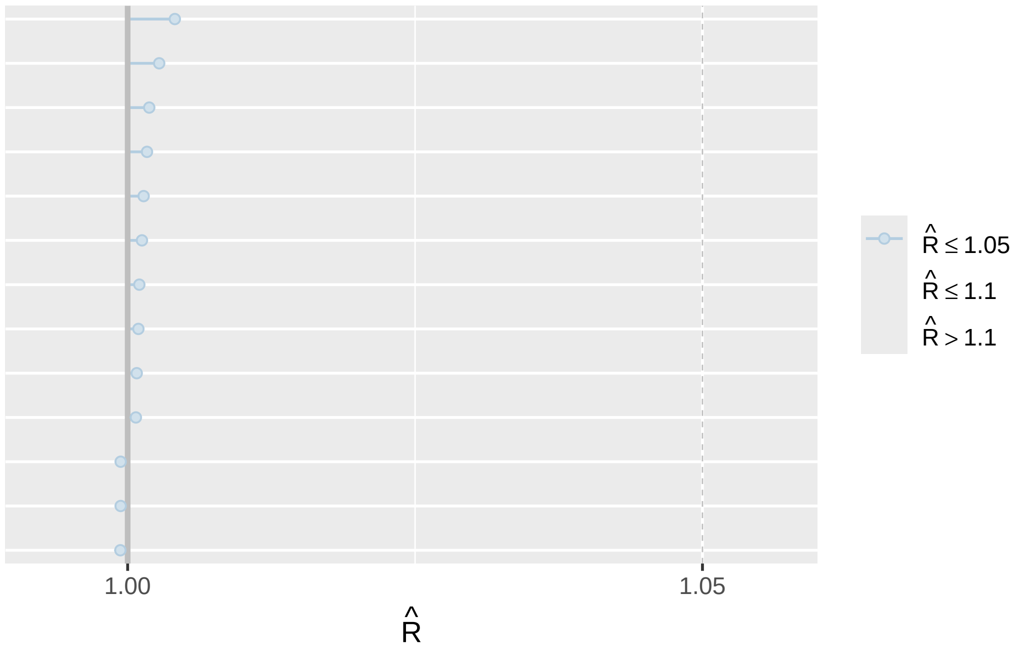

Figure 7: All parameters have R-hat values at or below 1.05, indicating strong convergence of the MCMC chains and reliable parameter estimates.

the same posterior distribution. In Figure 7, all parameters have an value at or below 1.05, suggesting that the model has achieved adequate convergence and the estimates are reliable.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

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

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Statistics Canada. 2020. "Wages by education level." *Open Government Portal*. https://open.canada.ca/data/en/dataset/1f14addd-e4fc-4a07-9982-ad98db07ef86.

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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.