**Program and Instructions**

**RRMW\_Data Activity\_CAnD 3 Training**

**Project title: RRWM Data Activity CAnD3**

**Owner: Genan Hamad**

**Date: Oct 2, 2025**

**Tools: R & R studio (2025.05.1+513)**

**Citation of Data:**

Statistics Canada. 2019. Census of Population, 2016 [Canada] Public Use Microdata File (PUMF): Individuals File. Statistics Canada [producer and distributor], accessed September 10, 2021. ID: pumf-98M0001-E-2016-individuals.

**Summary of the main tasks that were performed to produce code for this project:**

1. Create a subset of data from the Education and Income variables
2. Recode the Education variable
3. Generate a descriptive statistics table for the new subset of data
4. Run a simple linear regression

**Detailed tasks that were performed to produce code for this project:**

**Importing Data**

1. Import and Read CSV data file using base R
2. Preview data

**Variable selection**

1. Create a new dataset (data2) with only **Wages (income variable)** and **HDGREE (education variable)** columns
2. View first few rows of data2

**Recoding the HDGREE (education variable)**

1. Recode the HDGREE (education variable), whereas the original values (1, 2, 3, 4, 5, 6, 7) are coded as "low\_edu" and the original values (8, 9, 10, 11, 12, 13) are coded as "high\_edu"
2. Show summary of data2
3. Preview data2

**Descriptive Statistics Table**

Generate a descriptive statistics table for data2, using dfSummary(). data2 contains: "Wages": numeric variable and "HDGREE": categorical variable ("low\_edu" and "high\_edu")

**Regression analysis**

Perform a simple linear regression using data2. The predictor: HDGREE (education variable) is categorical: "low\_edu", "high\_edu", and the outcome variable: Wages (income variable) is numeric. Original documentation and descriptions of these variables are provided in the appendix.

1. Convert HDGREE to a factor
2. Run a simple linear regression analysis using data2
3. View the regression model summary

**Appendix**

**Original Documentation of Variable Used in this Project:**

1. **HDGREE (education variable)**

Refers to the highest certificate, diploma or degree

Original Documentation of **HDGREE**:

A screenshot of a certificate

AI-generated content may be incorrect.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Value** | **Label** | **Records** | **Weighted** | **Percentage (Weighted)** |
| 1 | No certificate, diploma or degree | 141,122 | 5,226,774.0 | 18.2% |
| 2 | Secondary (high) school diploma or equivalency certificate | 204,645 | 7,579,423.6 | 26.5% |
| 3 | Trades certificate or diploma other than Certificate of Apprenticeship | 41,156 | 1,524,386.6 | 5.3% |
| 4 | Certificate of Apprenticeship or Certificate of Qualification | 32,745 | 1,212,759.8 | 4.2% |
| 5 | Program of 3 months to less than 1 year (College, CEGEP, etc.) | 24,062 | 891,167.9 | 3.1% |
| 6 | Program of 1 to 2 years (College, CEGEP and other non-university) | 69,982 | 2,591,915.4 | 9.0% |
| 7 | Program of more than 2 years (College, CEGEP and other non-university) | 53,855 | 1,994,674.2 | 7.0% |
| 8 | University certificate or diploma below bachelor level | 20,853 | 772,341.9 | 2.7% |
| 9 | Bachelor’s degree | 119,686 | 4,432,780.6 | 15.5% |
| 10 | University certificate or diploma above bachelor level | 11,295 | 418,337.8 | 1.5% |
| 11 | Degree in medicine, dentistry, veterinary medicine or optometry | 4,433 | 164,184.5 | 0.6% |
| 12 | Master’s degree | 35,374 | 1,310,156.6 | 4.6% |
| 13 | Earned doctorate | 5,478 | 202,889.3 | 0.7% |
| 88 | Not available | 8,604 | 318,671.0 | 1.1% |
| 99 | Not applicable | 157,131 | 5,819,600.8 | — |

1. **Wages (income variable)**

Refers to gross wages and salaries before deduction.

Original Documentation of **Wages**:

A screenshot of a computer

AI-generated content may be incorrect.