ReadMe File

Research question:

Is there any significant association between total income and sex, age at immigration, and level of education of Canadian Immigrants from 2000 to 2015 who immigrated at the age of 20 years and older?

Dataset: Census 2016.csv

Variables:

TotInc => Total Income

Sex => Sex

AGEIMM => Age at immigration

HDGREE => Highest educational qualification

YRIMM => Year of immigration

Software R studio:

1. Install necessary packages and load the libraries
2. Clean up data by:
   1. Loading the CSV file
   2. Select only TotInc, Sex, AGEIMM, HDGREE, YRIMM (variables) and discard others
   3. Filter AGEIMM and select range 5 to 13 [meaning 20 years and above]
   4. Remove cases with TotInc less than 1000 and those labelled 88888888 and 99999999
   5. Remove cases with HDGREE = 88 or 99
   6. Convert all variables in the dataset to numeric
   7. Re-code Sex as Female =1, Male = 2
   8. Creating HDGREE.R variable such that 1 to 8 = below bachelor, 9 = bachelor, and 10 to 13 = above bachelor
3. Run descriptive statistics to create frequency and percentage summary table for Sex and Education levels.
4. Run a multivariate regression analysis with TotInc as the dependent variable and Sex, age at immigration, and education qualification (HDGREE.R) as independent variables.
5. Write the 2 or 3 tabular output in any file format.

#################################### End ##################