Validation and Replication Report:

Life Satisfaction Determinants Analysis

1. Summary
This replication reproduces survey-weighted regression modeling to examine whether household income, education, religious attendance predicts self-rated life satisfaction (0–10 scale) using the 2017 General Social Survey (Cycle 31-Family).
All data were accessible via the CAnD3 course site, and the code executed without any error. The process of replication reproduced the same outputs with the authors. Results indicate all three predictor variables significantly influence life satisfaction scores.

2. Data Description
Data Sources: General Social Survey (GSS) dataset
Analysis Data Files
Replication package included five files:

1. Program Output
2. RRWM - Project Program.docx
3. RRWM Assignment.R
4. Readme.md
5. gss-12M0025-E-2017-c-31.pdf (codebook of gss2017)

No dataset provided but can be retrieved and downloaded from CAnD3 course portal.

3. Data checks
File format and readability
- dataset gss-12M0025-E-2017-c-31_F1.csv was read in R using read.csv().
- CSV format is an archive-ready, open and non-proprietary format suitable for preservation.
- The script shows the csv.file could be accessed using a relative path.
- No encoding or delimiter problems, all variables can access and read consistently.
Data structure and content
- The dataset corresponds to the 2017 General Social Survey (GSS), Cycle 31, Public Use Microdata File.

- Ruth's master script selected three variables for analysis:
  SLM_01-Overall life satisfaction
  FAMINCG2-Family income (before tax)
  EHG3_01B-Education - Highest certificate, diploma or degree
  REE_03-Engaging in religious or spiritual activities on own - Past 12 months
  And adding a weighting factor:
  WGHT_PER- for analysis at the person level4. Code Description

Data cleaning and variable recoding
1.Variable Type Conversions

SLM_01: Converted from original format to numeric type for quantitative analysis
Categorical Variables: Three key demographic variables are converted from numeric codes to meaningful factor variables with descriptive labels
2. Factor Variable Recoding
FAMINCG2: Income categories recoded from numeric codes (1-6) to descriptive income ranges with dollar amounts
EHG3_01B: Education levels transformed from codes (1-7) to detailed educational attainment labels
REE_03: Religious participation frequency converted from codes (1-6) to explicit frequency descriptions
3. Data Quality Filtering
Applied complete-case analysis by removing observations with missing values across all four key variables
Ensures dataset integrity for subsequent statistical modeling
Survey design definition (bootstrap replicate weights)
Weighted linear regression analysis
Results visualization and coefficient estimation
Three tables in Project program, and description tables in script

4. Stated Requirements
All clear in Ruth's README file, including RQ, dataset, software and packages, workflow, and data source information.

5. Missing Requirements
Original data file paths and access methods not specified.

6. Computing Environment of the Replicator
Original computing environment stated as RStudio: 2025.09.0 Build 387, "Cucumberleaf Sunflower" Release (af5fc22a, 2025-09-11) for Windows; R: version 4.5.1 (2025-06-13 ucrt), and Quarto: 1.7.32.
Operating System: macOS Sequoia (version 15.6.1)
Hardware:
      Processor: Apple M4 chip
      Installed RAM: 16 GB
      System Type: macOS Sequoia
Software:

      R Studio: 2025.09.0 Build 387 ("Cucumberleaf Sunflower" Release)

      R version: 2025.05.1+513

      R packages loaded as noted in R script: haven, dplyr, skimr, gtsummary, emmeans,

      forcats, ggplot2, broom . All packages were installed from CRAN using

      install.packages() before. Those commands were not included in the script.

7. Replication Steps

Load GSS data

L7-9: accessing the gss2017.csv downloaded before because I cannot access directly it by my side currently.

Clean and recode variables

L12-35: Converted SLM_01 from original format to numeric type; Converted FAMINCG2, EHG3_01B, and REE_03 from numeric codes to meaningful factor variables with descriptive labels.

L37-43: Clear all missing value.

Description

L47-48: Describing SLM_01, FAMINCG2, EHG3_01B, REE_03

Run weighted regression model

L53-63:

Bivariate regression: Modeled life satisfaction as a function of household income.

Multivariate regression: Modeled life satisfaction as a function of household income, education, and religious attendance.

L66: Perform Wald test

Tidy coefficients and create visualization

L71-106: Drop intercept, clear labels, and create plot figure.

8. Findings

Tables1: Weighted Descriptive Statistics for household income, Education, Religious attendance, and Gender

Table 2: Survey-weighted Linear Regression Prediction Life Satisfaction from Household Income

Table3: Survey-Weighted Linear Regression Predicting Life Satisfaction from Houselhole Income, Education, Religious Attendance, and Gender.

Description table (in Program Output)

Coefficients table (in Program Output)