***Analysis Design Plan: GSS 2017 (Self-Rated Mental Health)***

Dataset: **General Social Survey 2017 (Cycle 31 – Family, Social Support and Well-Being)**  
Weight variable: **WGHT\_PER (Person Weight – General)**

**1. Data Cleaning**

* Loaded the 2017 GSS microdata file (gss-12M0025-E-2017-c-31\_F1.csv).
* Created analytic variables by recoding Statistics Canada numeric codes into labelled categories.
  + **Dependent variable:**
    - *SRH\_115 – Self-rated mental health*
      * 1 = Excellent, 2 = Very good, 3 = Good, 4 = Fair, 5 = Poor.
      * Values 96–99 (Valid skip, Refusal, Don’t know, Not stated) set to missing.
  + **Predictors:**
    - *AGEGR10 – Age group of respondent*
      * 15–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75 and over.
    - *SEX – Sex of respondent*
      * 1 = Male, 2 = Female.
    - *MARSTAT – Marital status*
      * 1 = Married, 2 = Living common-law, 3 = Widowed, 4 = Separated, 5 = Divorced, 6 = Single (never married).
    - *FAMINCG2 – Total family income before tax*
      * 1 = Less than $25 000, 2 = $25 000–$49 999, 3 = $50 000–$74 999, 4 = $75 000–$99 999, 5 = $100 000–$124 999, 6 = $125 000 and more.
* Converted *Self-rated mental health* into a numeric scale
* Checked for missing values and removed invalid or non-response categories.

**2. Data Merging**

* Not applicable.  
  The analysis uses a single dataset (no merge operations required).

**3. Definition of Analytic Sample**

* Included respondents with valid responses on:
  + SRH\_115 (Self-rated mental health)
  + AGEGR10 (Age group)
  + SEX (Sex)
  + MARSTAT (Marital status)
  + FAMINCG2 (Family income).
* Excluded respondents with missing or invalid values on any of these variables.
* Applied survey weights using **WGHT\_PER** to obtain population-representative estimates for Canada.

**4. Table of Summary / Descriptive Statistics**

* Produced a weighted frequency table showing the distribution of:
  + **Self-rated mental health** (Excellent → Poor).
  + **Sex** (Male, Female).
  + **Age group** (Seven 10-year intervals from 15 to 75+).
  + **Marital status** (Six categories).
  + **Family income** (Six income bands).
* Percentages are weighted using WGHT\_PER.
* Table caption: *“Table 1. Weighted Descriptive Statistics for Key Variables (GSS 2017)”*.

**5. Table of Multivariate Analyses**

* Estimated a **survey-weighted linear regression model** with:
  + **Dependent variable:** Self-rated mental health (numeric 1–5)
  + **Independent variables:** Sex, Age group, Marital status, and Family income.
* Model type: **Weighted least squares (svyglm)** accounting for complex sampling design.
* Interpretation:
  + Positive coefficients → lower self-rated mental health (i.e., worse outcomes).
  + Negative coefficients → better self-rated mental health (closer to “Excellent”).

**6. Data Accessibility Statement**

This project uses the **2017 General Social Survey: Families** from Statistics Canada.  
Access and reuse are governed by the *Statistics Canada Open Licence Agreement*.  
All analyses are descriptive and reproducible using open-source R packages