

Weekly Status Submission Template

Week Number: 1-3 Weeks

Date: September 18, 2025

Student Name: Abdullah Siddiqui

Original Plan for This Week: (5 points)

According to my project timeline, the first three weeks were planned for:

- **Week 1:** Access and load the 2023 BRFSS dataset and identify relevant variables for analysis.
- **Week 2:** Clean and recode the selected variables, removing invalid responses and preparing the dataset for analysis.
- **Week 3:** Conduct descriptive statistics and produce initial exploratory visualizations such as histograms, bar charts, boxplots.

Tasks Accomplished This Week: (10 points)

Over the first three weeks, I:

- Successfully loaded the 2023 BRFSS dataset (1.1 GB .XPT file) using R.
- Created a subset of the data containing only the relevant variables: mental health days, exercise, smoking, alcohol, sex, education, income, and age group.
- Cleaned the dataset by recoding invalid codes (e.g., 77 = “Don’t know”, 99 = “Refused”, 888 = “No drinks”) and saving the cleaned dataset as BRFSS2023_subset_clean.csv.
- Computed descriptive statistics, including:
 - Average poor mental health days: 4.36
 - Exercise: 75% reported exercising, 25% reported no exercise
 - Smoking: 20% daily, 9% some days, 71% not at all
 - Alcohol: 47% non-drinkers
 - Education: 70%+ had at least some college education
 - Income: distributed across categories, with more than half reporting \$35,000 or higher
- Produced three exploratory plots and saved them as PNGs:
 - Histogram of poor mental health days (hist_mentalhealth.png)
 - Bar chart of exercise (bar_exercise.png)
 - Boxplot of mental health days by sex (boxplot_mentalhealth_sex.png)

Comparison of Planned vs Actual: (10 points)

Week	Planned	Actual
Week 1	Load BRFSS dataset, identify key variables	Successfully loaded dataset, selected subset of 8 variables
Week 2	Clean and recode dataset	Cleaned all variables, handled missing/invalid codes, saved cleaned CSV
Week 3	Run descriptive statistics and create initial plots	Generated descriptive stats and 3 visualizations (histogram, bar chart, boxplot)

Self-Rating of Progress: (5 points)

- **Done** Met the Planned Tasks
- [] Did Not Meet the Planned Tasks

(If you did not meet the planned tasks, please explain why and how you plan to get back on schedule)

Plan for Next Week: (5 points)

For Week 4, my plan is to move beyond descriptive statistics and exploratory plots by:

- Investigating relationships between variables such as correlation between education, income, and poor mental health days.
- Creating cross-tabulations such as exercise vs. smoking, alcohol vs. mental health.
- Preparing initial drafts of more advanced analysis regression models for later weeks.

Evidence of Tasks Completed: (15 points)

(Please provide details of datasets, data analysis, graphs, code etc. as evidence of tasks completed. You can include links to your GitHub repository or other platforms like Kaggle where your work is stored – make sure that the links are set to public access so I can view them)

- **Dataset:** Subset of the 2023 BRFSS (BRFSS2023_subset_clean.csv).

- **Code:** Full R script (BRFSS2023_analysis.R) showing data loading, cleaning, and analysis.
- **Descriptive Statistics:** Average poor mental health days = 4.36, 75% exercised, 21% smoked, 47% non-drinkers.
- **Plots:**
 - Histogram of poor mental health days (hist_mentalhealth.png)
 - Bar chart of exercise (bar_exercise.png)
 - Boxplot of poor mental health days by sex (boxplot_mentalhealth_sex.png)

All files (code, dataset, plots) are saved locally and can be shared via GitHub.

<https://github.com/abdullahs1357/SeniorThesis.git>