Boston University
Department of Mathematics and Statistics

MA213 - Fall 2025

Basic Statistics and Probability
Project 1

Project 1 : Exploration of the Data

Your group will select a dataset of your interest and conduct exploratory data analysis. This should include :

- One numerical exploratory data anlysis
- One categorical exploratory data analysis

You will present your analyses through either:

- An in-class presentation
- A pre-reorded video presentation

Suggested Outline

- Introduction
    1. Introduce the dataset and its source
    2. Tell the story of your reasoning and possible hypothesis
    3. Define your variables of interest
- Data Analysis
    1. Numerical Exploratory Data Analysis Part
        (a) Summary statistics
        (b) The relationship between two numerical variables
        (c) Distribution of variable(s) of interest
        (d) Discuss shape, central tendency, spreadity and outliers.
    2. Categorical Exploratory Data Analysis Part
        (a) Summarize categorical variables

(b) Contingency Table (two categorical varaibles)
(c) Visualizations: bar plots, pie charts, etc.

- Conclusion
  1. Summarize key insights
  2. Suggest future analysis or potential applications

Project requirements

- Source of data must be cited
- R code should be included (Rscript or Rmd files)
- References should be provided at the end

The following is the deliverables and deadlines.

Item	**Description**	**Due**
Initial Submission	Brief introduction of the analysis and roles of group members	Week 5 (after Lab 3)
R Script or RMarkdown File	Include data preparation, analysis, and visualizations	Week 7
Slide File	4–5 slides (excluding Title and Reference slides); 5-minute presentation	Week 7

Table 1: Deliverables and Deadlines

Data Sources to Consider

- OpenIntro Datasets
- fivethirtyeight R Package
- datsets R package
- Kaggle datasets
- US government's open data