

# Adv R - HW1

Sushmitha - 24209228

## 1. Introduction

The **Titanic dataset** is a widely used dataset in data analysis and machine learning, originally derived from the tragic sinking of the **RMS Titanic in April 1912**. It captures detailed information on the fate of the passengers aboard the ship, summarizing their survival outcomes based on *economic status (passenger class), sex, age group, and survival status*.

In its original format in R, the Titanic dataset is accessed using Titanic (capital T), and is structured as a contingency table object of class table. To facilitate data analysis and visualization, it is common practice to convert this table into a *data frame*, where each row corresponds to a unique combination of the four variables, with an additional column Freq indicating the number of passengers in that group.

Key Variables:

- **Class:** Indicates the socio-economic status of the passenger (1st, 2nd, 3rd, or Crew).
- **Sex:** Gender of the passenger (Male or Female).
- **Age:** Age group of the passenger (Child or Adult).
- **Survived:** Survival outcome (Yes or No).
- **Freq:** Frequency count for each combination of the above variables.

### **i** Note

In this document, the margins have been reduced, headings and sections have been included, color highlighting has been applied, and various formatting aesthetics have been incorporated to enhance formality and presentation.

This analysis presents a **bar chart** to explore the relationship between passenger class, sex, and survival outcome.

## 2. Dataset Manipulation

### 1. Load data and get summary

We use the Titanic dataset built into R. It is originally stored as a **4-dimensional table**, so we convert it into a data frame for plotting. Then basic summary of the data set has been displayed to get some insights from the variables.

Class	Sex	Age	Survived	Freq
1st :8	Male :16	Child:16	No :16	Min. : 0.00
2nd :8	Female:16	Adult:16	Yes:16	1st Qu.: 0.75
3rd :8				Median : 13.50
Crew:8				Mean : 68.78
				3rd Qu.: 77.00
				Max. :670.00

### Analytical Summary:

- **Passenger Class:** The dataset is balanced across all four classes (*1st, 2nd, 3rd, Crew*) with **8 records each**, indicating an equal representation of each category.
- **Sex:** There is an equal split between *males and females* across all classes (**16 records each**).
- **Age:** Both children and adults are represented equally (16 records each), providing a good mix for analysis.
- **Survival:** Each class has an equal number of survival/no survival records initially (16 each), but the frequency (Freq) column captures the actual number of people represented by each record, varying widely from **0 to 670**.
- **Frequency:** The frequency distribution of survival ranges from 0 to 670, with a mean of about **68.78**, showing significant variability in the number of passengers represented by each observation.

### 3. Data plotting and interpretation

A bar chart has been created to illustrate the survival outcomes of passengers in the Titanic dataset. Counts are separated based on *gender, with survival status (Yes or No) indicated by distinct colors within each gender category*. The bars have been grouped according to *passenger class (1st, 2nd, 3rd, and Crew)* to emphasize the influence of class and gender on survival rates.

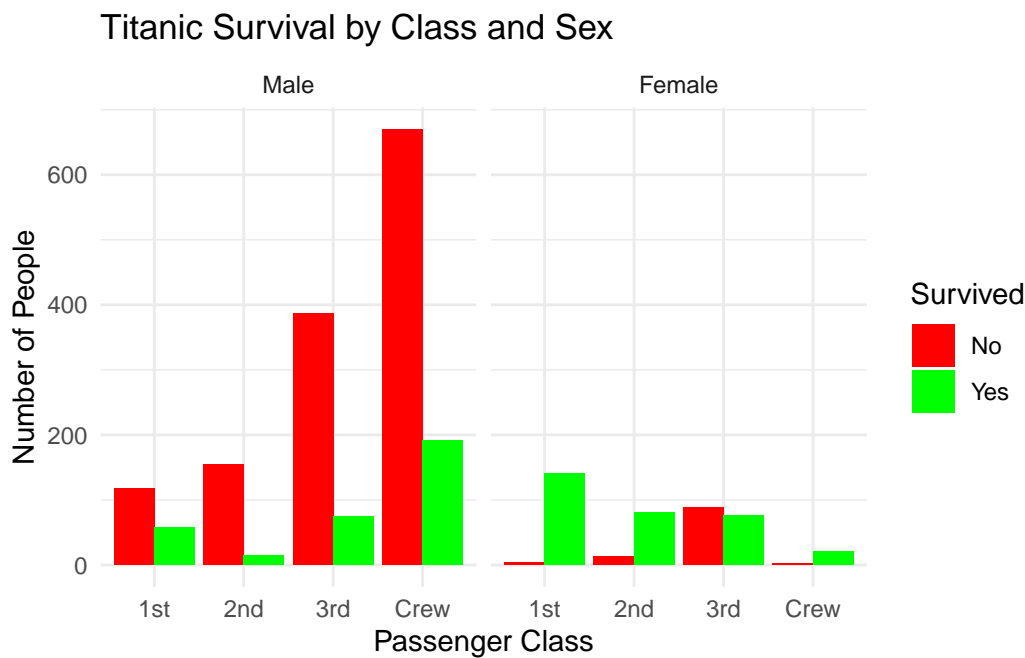


Figure 1: Bar plot representing the survival count by class and sex

### Plot Interpretation:

This plot shows that survival was more likely for passengers in **higher classes, and females** had higher survival rates across all classes. The lowest survival counts were among **third-class males**.

- **First-Class Passengers:** Had higher survival rates, especially females, where survival numbers were significantly higher than fatalities.
- **Second-Class Passengers:** Also show better survival odds, particularly for women.
- **Third-Class Passengers:** Had the lowest survival rates, especially third-class males, who had the highest number of deaths.
- **Crew Members:** Male crew had low survival; female crew had higher survival rates, but the total number was small.

This visualization makes it clear that survival on the Titanic was influenced by both **class and sex**, with *females and upper-class passengers* more likely to survive.

---