

Advanced Programming with R - Homework 1

Pranav Agwan, Student No - 24219261

Titanic Disaster

In this section we load the **Titanic** dataset which contains the **survival data** of passengers from the **Titanic disaster**. It **summarizes** the counts of **passengers** based on **class**, **age**, **sex** and **survival status**. Lets see the **preview** of the data.

	Class	Sex	Age	Survived	Freq
1	1st	Male	Child	No	0
2	2nd	Male	Child	No	0
3	3rd	Male	Child	No	35
4	Crew	Male	Child	No	0
5	1st	Female	Child	No	0
6	2nd	Female	Child	No	0

We can see **top 6 values** of our data with **5 variables**. Now we will see the structure.

```
'data.frame': 32 obs. of 5 variables:
 $ Class : Factor w/ 4 levels "1st","2nd","3rd",...: 1 2 3 4 1 2 3 4 1 2 ...
 $ Sex : Factor w/ 2 levels "Male","Female": 1 1 1 1 2 2 2 2 1 1 ...
 $ Age : Factor w/ 2 levels "Child","Adult": 1 1 1 1 1 1 1 1 2 2 ...
 $ Survived: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Freq : num 0 0 35 0 0 0 17 0 118 154 ...
```

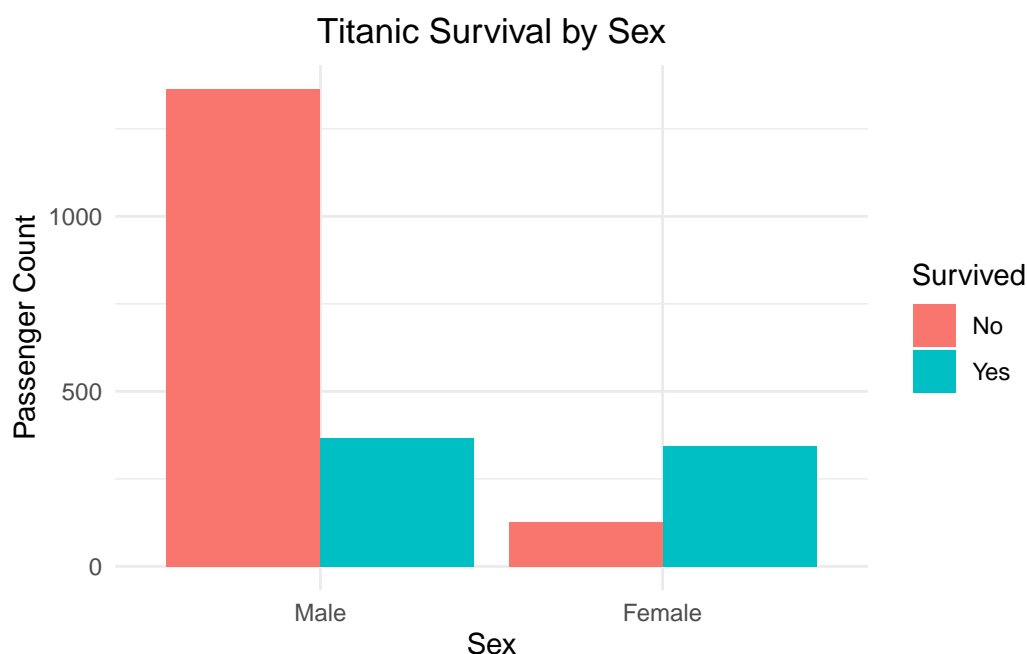
There are total **32 observations** of **5 variables** in the **Titanic dataset**. The variables are as follows:

- ***Class*** - It shows the **passenger class**. It is a **factor variable** with 4 levels: “1st”, “2nd”, “3rd” and “Crew”.
- ***Sex*** - It shows **passenger’s sex**. It is also a **factor variable** with 2 levels: “Male” and “Female”.

- **Age** - It shows the **age group** of the **passenger**. It is also a **factor variable** with 2 levels: “**Child**” and “**Adult**”.
- **Survived** - It shows the **survival status** of the **passenger**. It is also a **factor variable** with 2 levels: “**Yes**” and “**No**”.
- **Freq** - It shows the **count** of the **passengers** that fall into the specific groups mentioned above. It is a **numeric variable**.

Graphical Analysis of Survival of Passengers

We will now show a simple **bar plot** which displays the **number** of **male** and **female** passengers who **survived** or **did not survive** the incident. We will use the **Sex**, **Survived** and **Freq** column for the plot.



The **red bar** shows the passenger which **did not survive** the incident and the **blue bar** shows the **passenger** which **survived**. There are **2 bars** both for “**Male**” and “**Female**”. From the above graph we can **interpret** the following:

- There are approx **1350-1400 males** which **did not survive** the incident.
- There are approx **350-400 males** which **survived** the incident.
- There are approx **100-150 females** which **did not survive** the incident.
- There are approx **350-400 females** which **survived** the incident.