

Titanic_Survival_Analysis

Himanshi

2025-08-08

Project Overview

This project analyzes the `Titanic` dataset from R's built-in `datasets` package. It contains aggregated passenger data by Class, Sex, Age, and Survival status. The goal is to explore survival patterns and understand how class, gender, and age affected outcomes.

Loading and Inspecting the Data

```
data(Titanic)
titanic_df <- as.data.frame(Titanic)

head(titanic_df)
```

```
##   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
```

```
str(titanic_df)
```

```
## '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 ...
```

Survival rate

```
aggregate(Freq ~ Sex + Class + Survived, data=titanic_df, sum)
```

```
##      Sex Class Survived Freq
## 1   Male   1st      No   118
## 2 Female   1st      No    4
## 3   Male   2nd      No   154
## 4 Female   2nd      No    13
## 5   Male   3rd      No   422
## 6 Female   3rd      No   106
## 7   Male  Crew      No   670
## 8 Female  Crew      No    3
## 9   Male   1st     Yes    62
## 10 Female  1st     Yes   141
## 11  Male   2nd     Yes    25
## 12 Female  2nd     Yes    93
## 13  Male   3rd     Yes    88
## 14 Female  3rd     Yes    90
## 15  Male  Crew     Yes   192
## 16 Female  Crew     Yes    20
```

Overall Survival rate

```
total <- sum(titanic_df$Freq)
survived <- sum(titanic_df$Freq[titanic_df$Survived == "Yes"])
survival_rate <- survived / total * 100
survival_rate
```

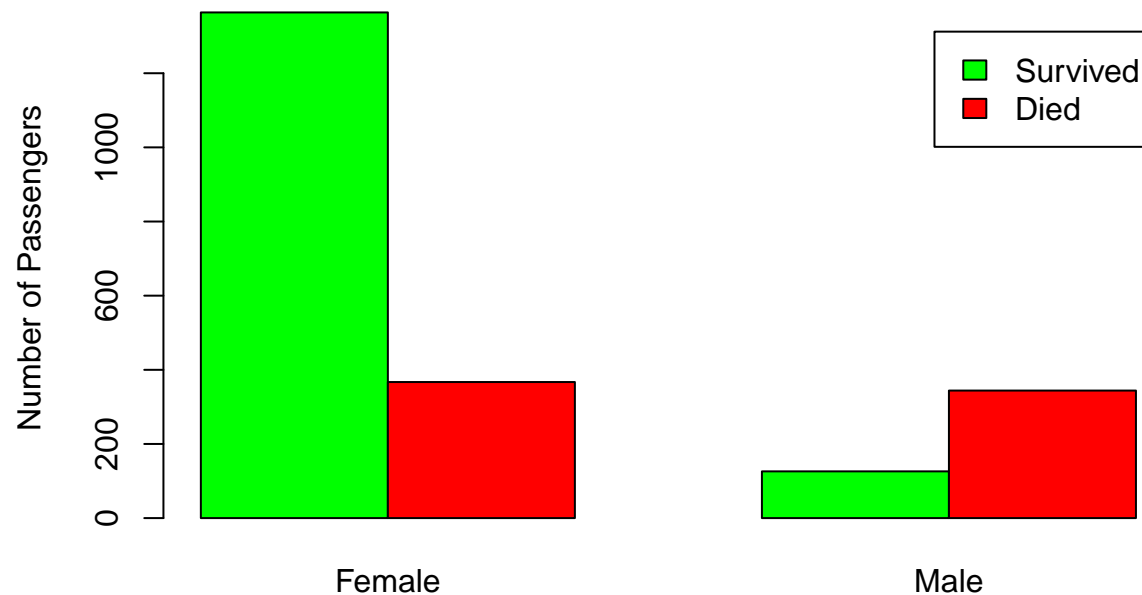
```
## [1] 32.3035
```

Survival By Sex

```
sex_survival <- aggregate(Freq ~ Sex + Survived, data=titanic_df, sum)

barplot(
  height = matrix(sex_survival$Freq, nrow=2, byrow=TRUE),
  beside = TRUE,
  names.arg = c("Female", "Male"),
  col = c("green", "red"),
  legend = c("Survived", "Died"),
  main = "Survival by Sex",
  ylab = "Number of Passengers"
)
```

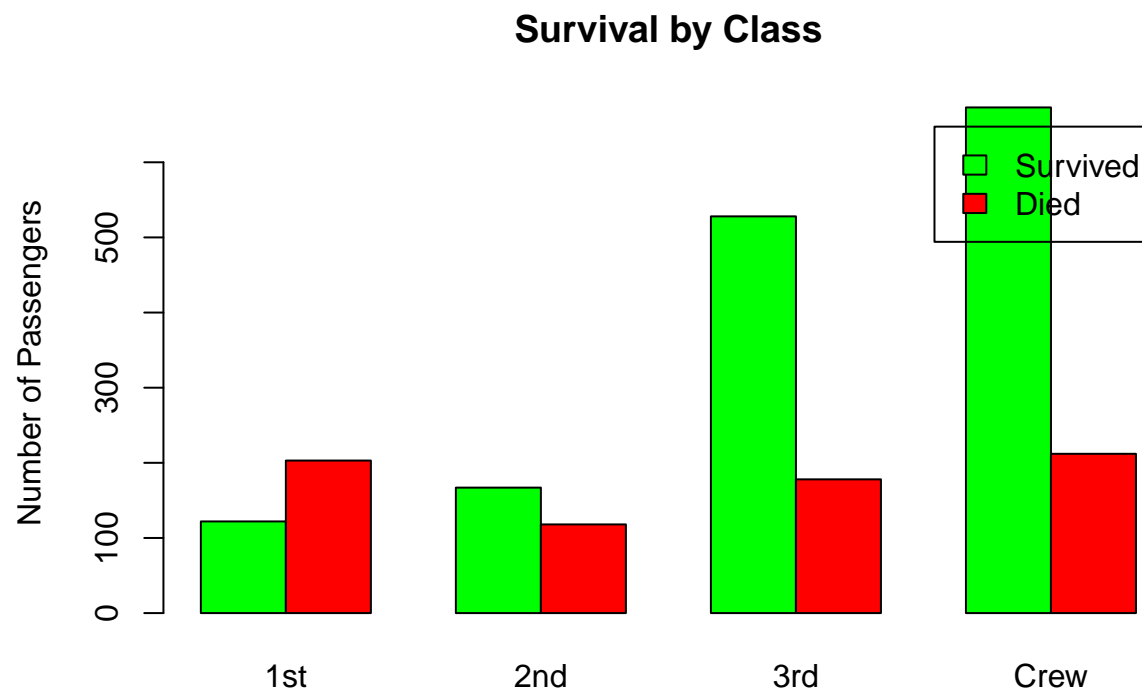
Survival by Sex



Survival By Class

```
class_survival <- aggregate(Freq ~ Class + Survived, data=titanic_df, sum)

barplot(
  height = matrix(class_survival$Freq, nrow=2, byrow=TRUE),
  beside = TRUE,
  names.arg = c("1st", "2nd", "3rd", "Crew"),
  col = c("green", "red"),
  legend = c("Survived", "Died"),
  main = "Survival by Class",
  ylab = "Number of Passengers"
)
```



Heatmap: Class vs Sex vs Survival

```
class_survival <- aggregate(Freq ~ Class + Survived, data=titanic_df, sum)

barplot(
  height = matrix(class_survival$Freq, nrow=2, byrow=TRUE),
  beside = TRUE,
  names.arg = c("1st", "2nd", "3rd", "Crew"),
  col = c("green", "red"),
  legend = c("Survived", "Died"),
  main = "Survival by Class",
  ylab = "Number of Passengers"
)
```

