```
# Alex Gibbons and Lauren Kawecki
# CMPT 363 - Data Mining
# Dr. Ankur Agrawal
# Due: February 19, 2024

# Assignment #2
# Titanic Dataset charts
library(graphics)
library(ggplot2)
# 1. Bar Chart of Class Survival counts by age and gender
        # This bar chart displays the survival counts based on age and gender within each of the 4 classes
        # Each group is labeled below the x axis and is visiually grouped by a horizontal gray line
barplot(as.matrix(Titanic[,,,"Yes"]), beside=TRUE, legend=rownames(Titanic),
     main="Survival by Age and Class", xlab="Age", ylab="Count",
     col=rainbow(length(rownames(Titanic))),
     args.legend=list(x="topright"))
text(c(3, 7, 11, 15), -0.5, c("Male Child", "Female Child", "Male Adult", "Female Adult"), pos=1, xpd=TRUE)
segments(x0 = c(1, 5, 9, 13) + 0.4, x1 = c(5, 9, 13, 17) - 0.4, y0 = 0, y1 = 0, col = "lightgray", lwd = 5)
```
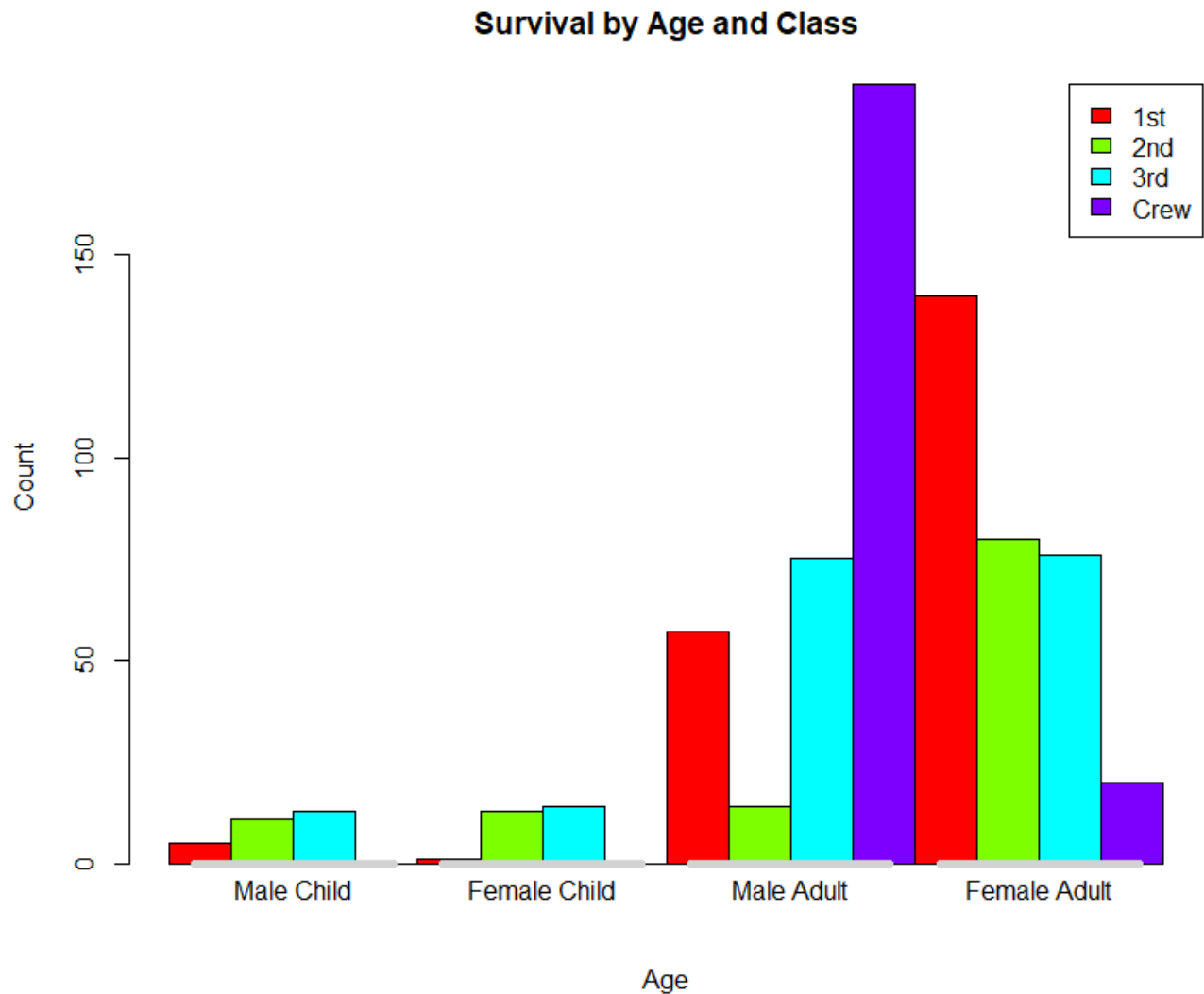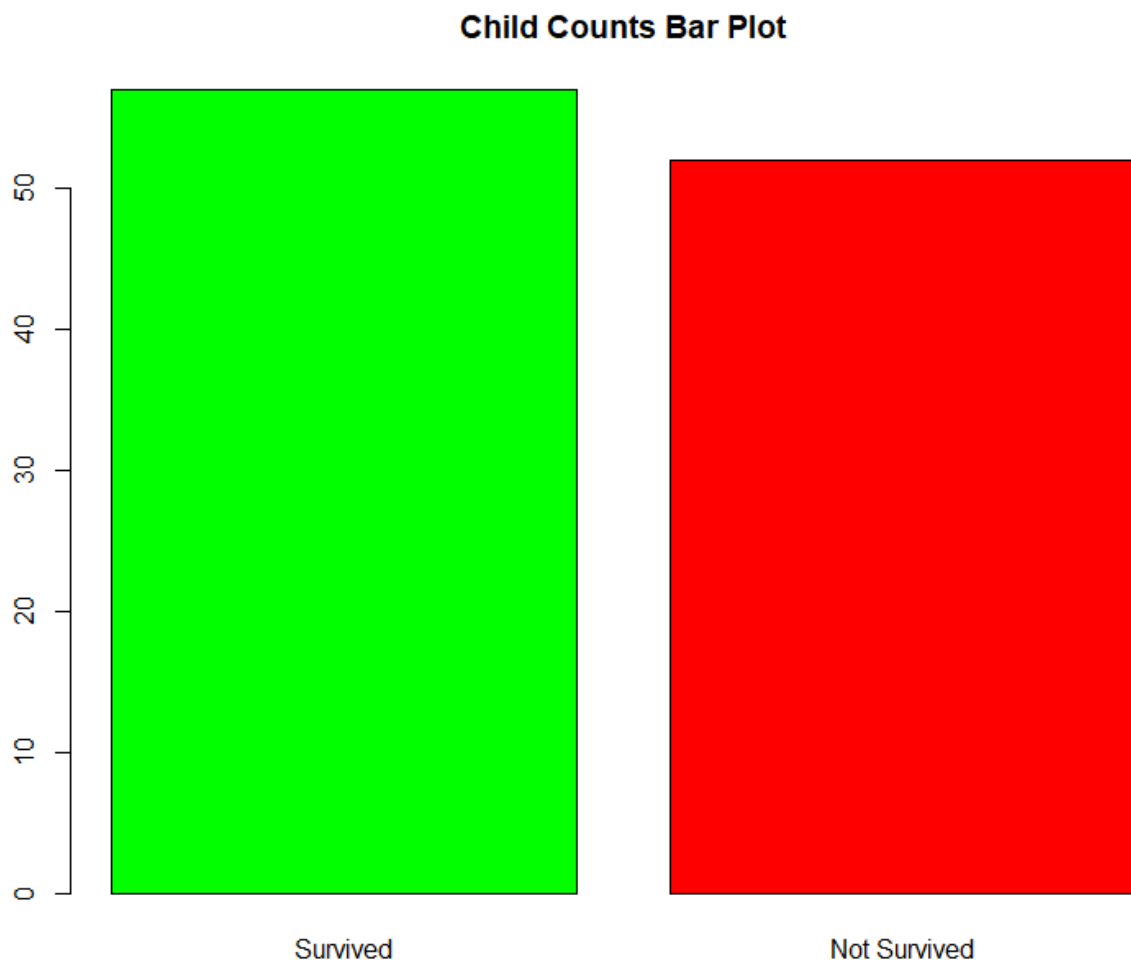

Survival by Age and Class

2. Bar chart of Children Survival Counts
        # This bar chart displays the counts of children that survived vs those that did not
barplot(c(sum(Titanic[,,1,2]),sum(Titanic[,,1,1])),
        main="Child Counts Bar Plot",
        col=c("green","red"),
        names.arg = c("Survived","Not Survived"))

Child Counts Bar Plot

```
# 3. Bar chart of Adult Survival Counts
        # This bar chart displays the counts of adults that survived vs those that did not
barplot(c(sum(Titanic[,,2,2]),sum(Titanic[,,2,1])),
    main="Adult Counts Bar Plot",
    col=c("green","red"),
    names.arg = c("Survived","Not Survived"))
```

Adult Counts Bar Plot
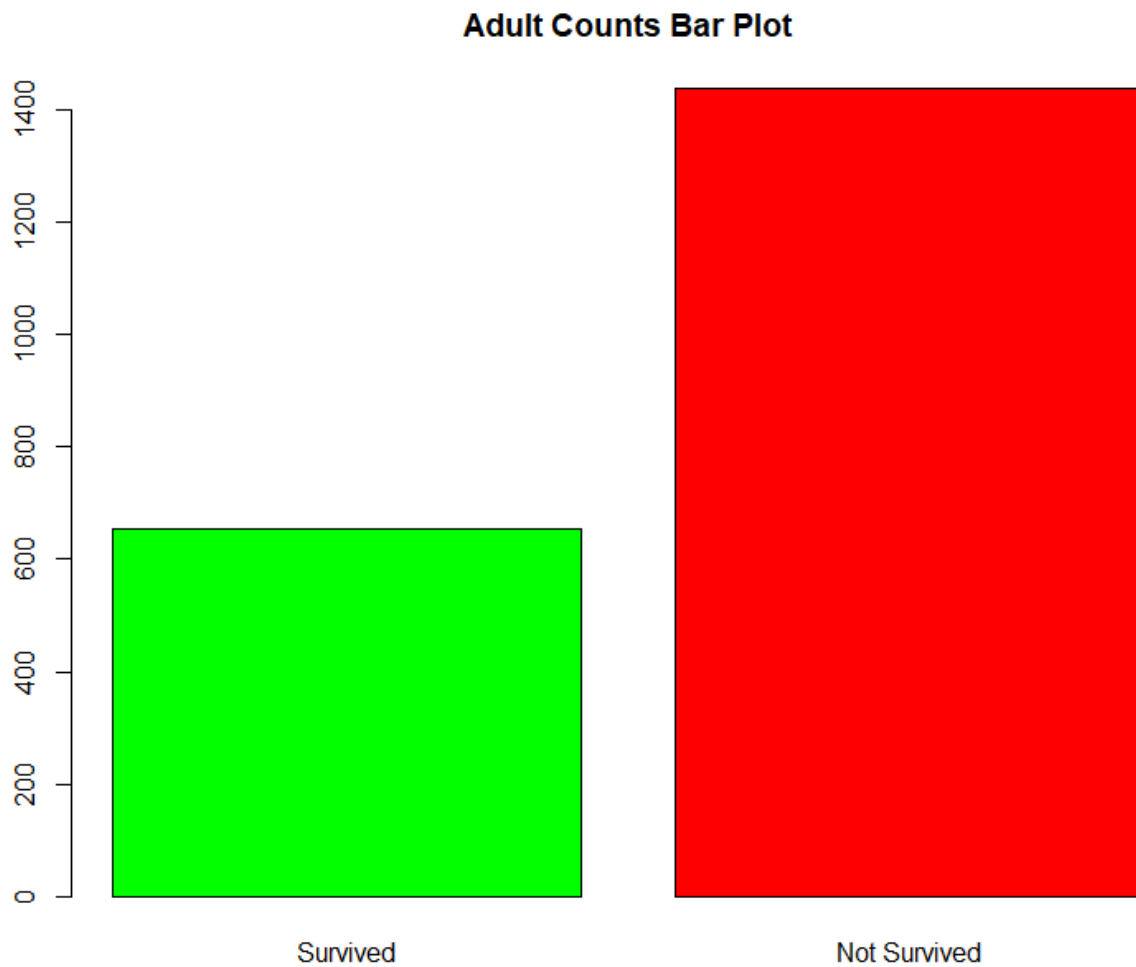
4. Pie Chart Survival Rate by Class
        # This pie chart illustrates the survival rate of passengers from each Class
        # Each slice represents a class
class_survive_slices<-c(sum(Titanic[1,,,"Yes"]),sum(Titanic[2,,,"Yes"]),sum(Titanic[3,,,"Yes"]),sum(Titanic[4,,,"Yes"]))
class_survive_pct<-round(class_survive_slices/sum(class_survive_slices)*100)
class_survive_pie_lbls<-paste(rownames(Titanic), " ", class_survive_pct,"%",sept="")
pie(class_survive_slices,labels=class_survive_pie_lbls,col=rainbow(length(class_survive_pie_lbls)),main="Pie Chart Surviving Classes")
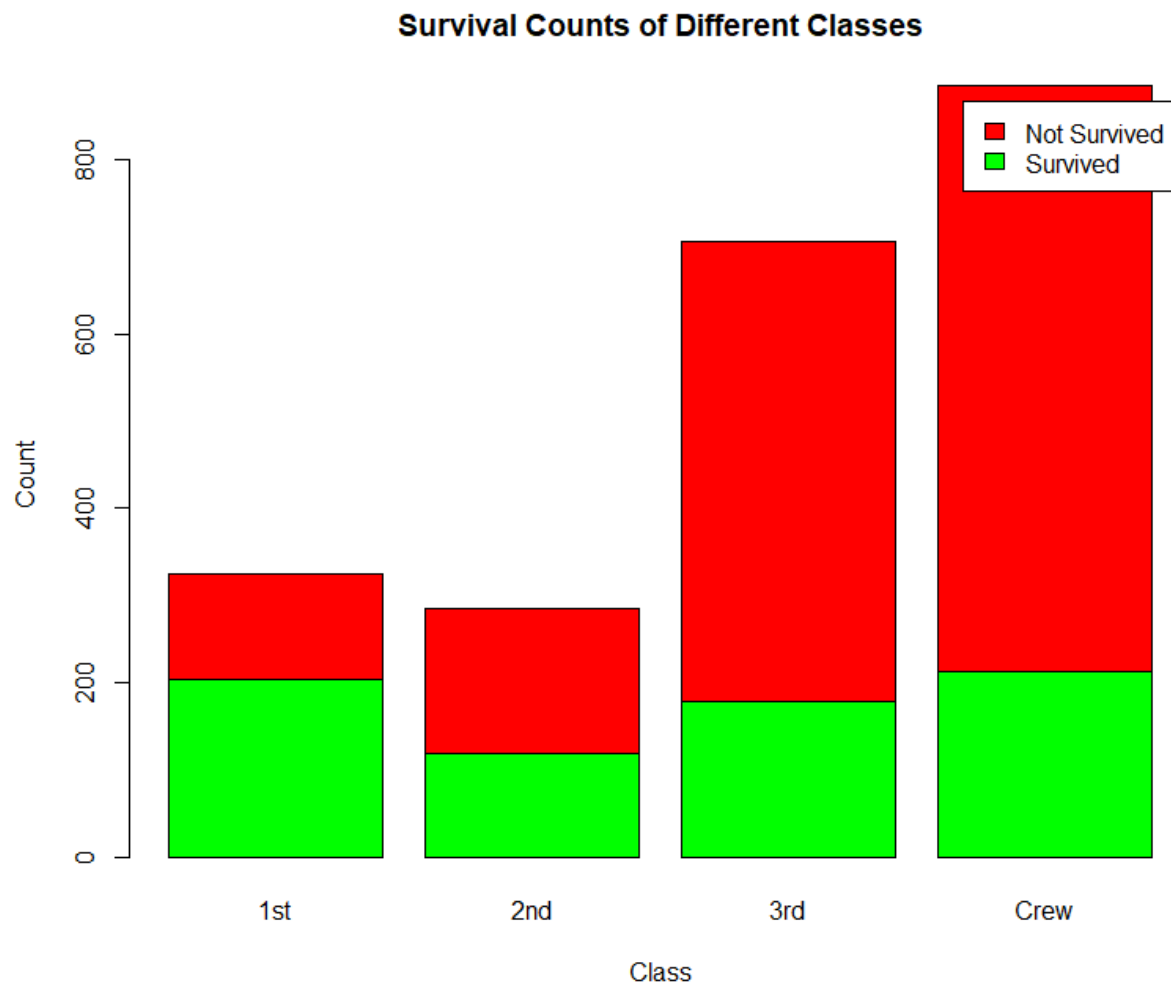
Pie Chart Surviving Classes
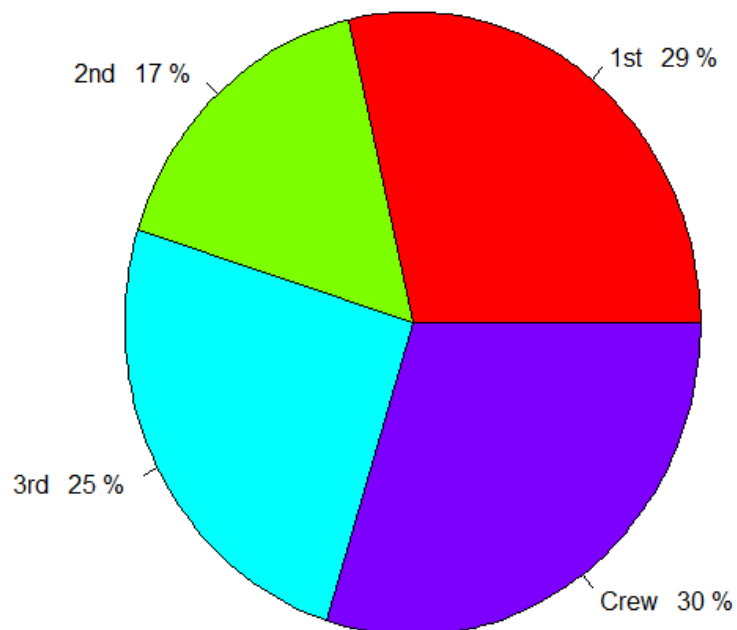
```
# 5. Stacked Bar Chart of Class Survival Counts
        # This stacked bar chart compares the survival counts between classes
        # Each bar is broken in two segments representing survivors and non-survivors
class_survived <- c(sum(Titanic[1,,,"Yes"]),sum(Titanic[2,,,"Yes"]),sum(Titanic[3,,,"Yes"]),sum(Titanic[4,,,"Yes"]))
class_not_survived <- c(sum(Titanic[1,,,"No"]),sum(Titanic[2,,,"No"]),sum(Titanic[3,,,"No"]),sum(Titanic[4,,,"No"]))
class_bar_data <- rbind(class_survived,class_not_survived)
barplot(class_bar_data, col = c("green","red"),
      names.arg = rownames(Titanic),
      legend.text = c("Survived","Not Survived"),
      xlab = "Class", ylab="Count",
      main = "Survival Counts of Different Classes")
```

```
# 6. Stacked Bar Chart of Survival Counts Adult vs Child
        # This stacked bar chart compares the survival counts between ages
        # Each bar represents total count of those that survived and those that did not
        # Each bar is broken in two colored segments representing child and adults
adult_survival <- c(sum(Titanic[,,2,"Yes"]), sum(Titanic[,,2,"No"]))
child_survival <- c(sum(Titanic[,,1,"Yes"]), sum(Titanic[,,1,"No"]))
age_bar_data <- rbind(adult_survival, child_survival)
barplot(age_bar_data, col = c("green", "red"),
    names.arg = c("Survived", "Not Survived"),
    legend.text = c("Adult", "Child"),
    xlab = "Survival", ylab = "Count",
    main = "Survival Counts of Adults and Children")
```
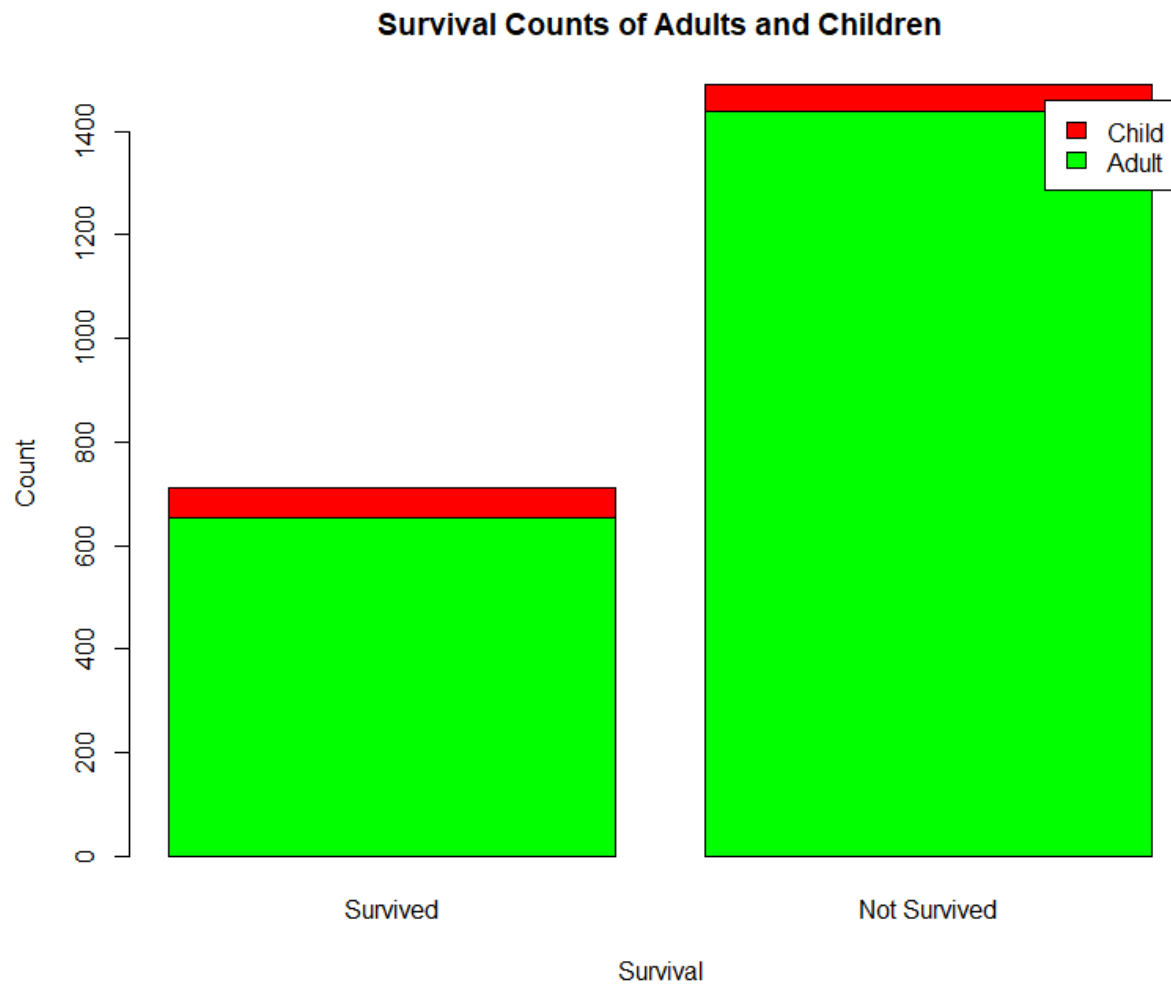
Survival Counts of Adults and Children

7. Stacked Bar Chart male vs female survive not survive
        # This stacked bar chart compares survival counts between males and females
        # Each bar represents total count of survivors and non-survivors and is broken up
        #   by color to indicate male or female
male_survival<-c(sum(Titanic[,1,,2]),sum(Titanic[,1,,1]))
female_survival<-c(sum(Titanic[,2,,2]),sum(Titanic[,2,,1]))
gender_data<-rbind(male_survival,female_survival)
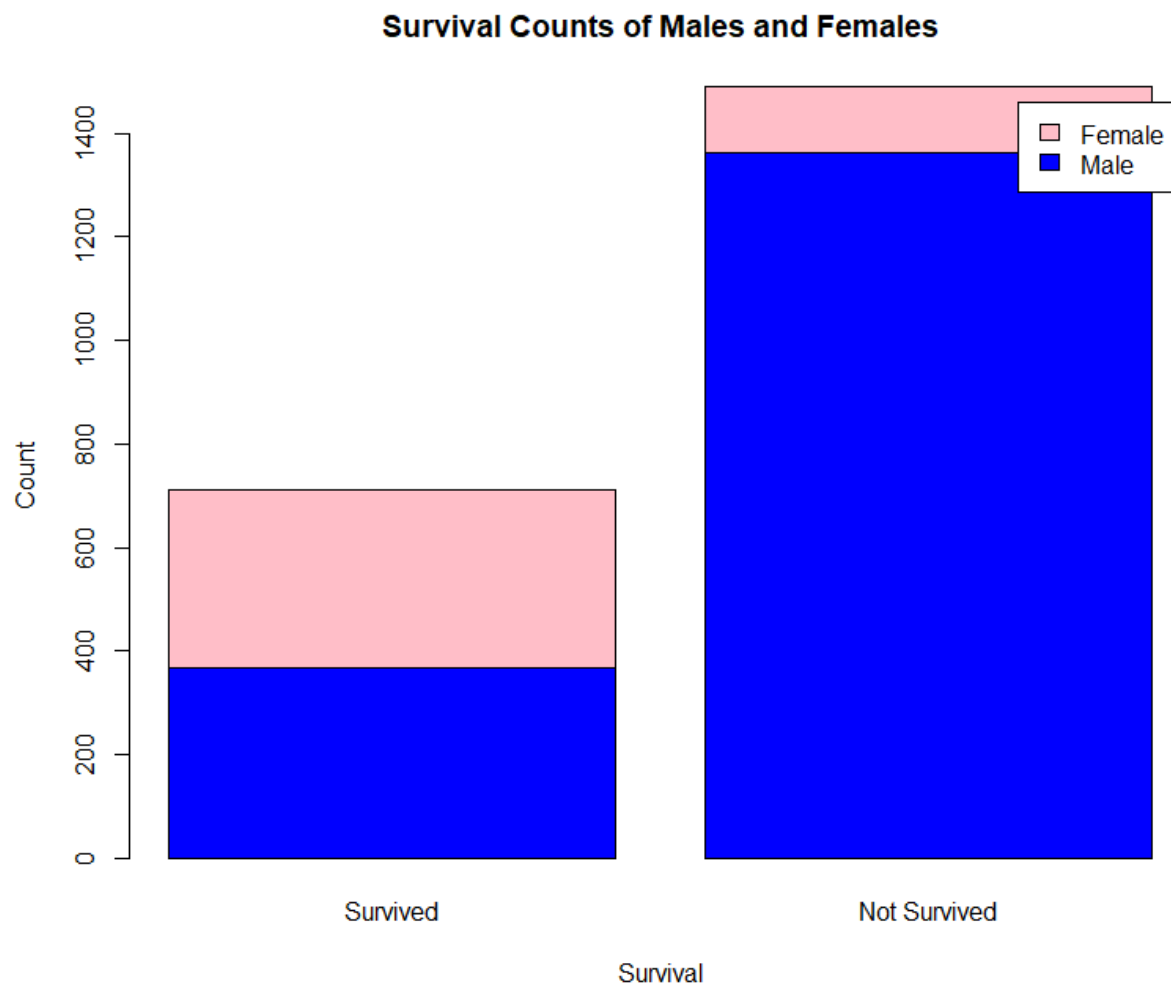barplot(gender_data, col=c("blue","pink"),
    names.arg = c("Survived","Not Survived"),
    legend.text = c("Male","Female"),
    xlab = "Survival", ylab = "Count",
    main = "Survival Counts of Males and Females")

**Survival Counts of Males and Females**

8. Mosaic Plot of all data

      # This mosaic plot provides and overview of survival data on the Titanic

      # It displays the proportions of survivors and non-survivors across different categories(class, sex, age).

      # The colors indicate whether they survived (green) or did not survive (green)

```
mosaicplot(Titanic, main = "Survival on the Titanic", col=c("red","green"))
```


Survival on the Titanic
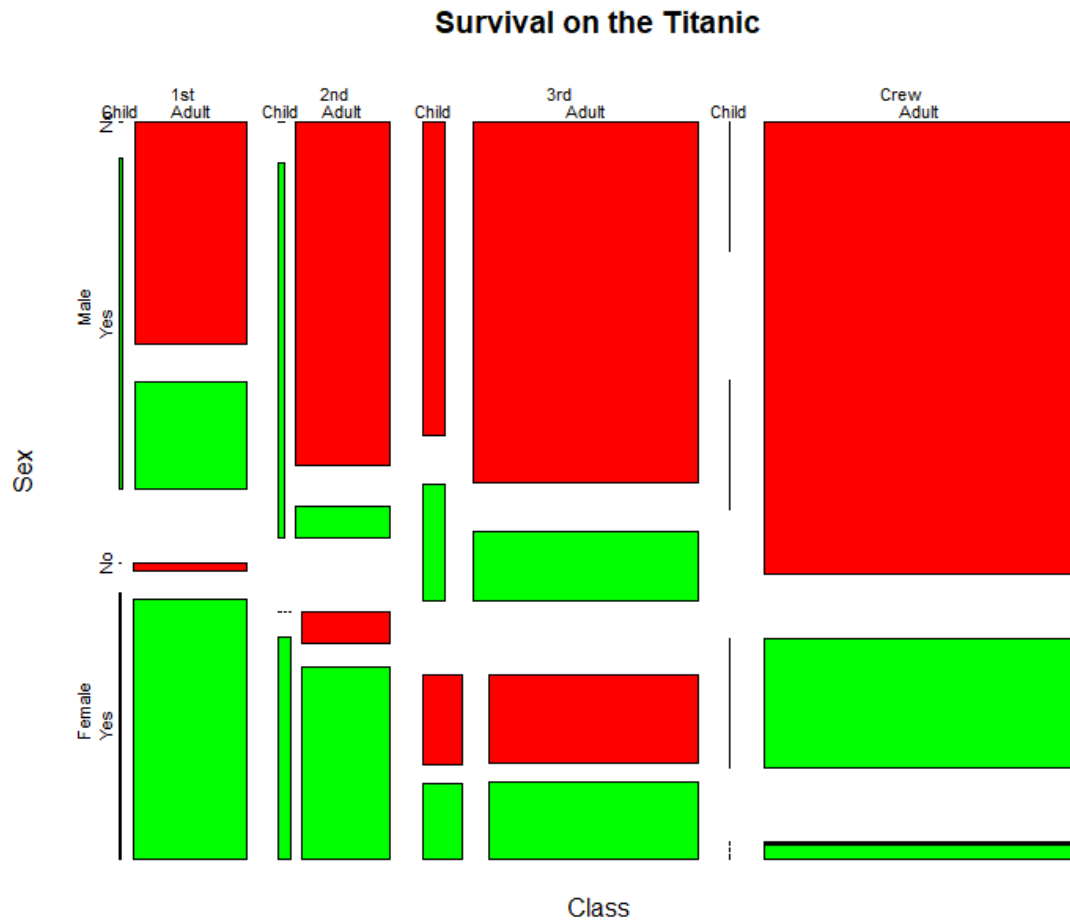
```
# 9. Scatter Plot of Passenger Survival Rate by Class
        # This scatter plot visualizes the survival rate of passengers across different classes.
        # The survival rate is calculated by dividing the count of survivors divided by the total
        #          number of passengers in that class.
        # Each point on the plot represents a class with its corresponding survival rate.
class_survival <- c(sum(Titanic[1,,,"Yes"]), sum(Titanic[1,,,"No"]),
            sum(Titanic[2,,,"Yes"]), sum(Titanic[2,,,"No"]),
            sum(Titanic[3,,,"Yes"]), sum(Titanic[3,,,"No"]),
            sum(Titanic[4,,,"Yes"]), sum(Titanic[4,,,"No"]))
class_survival_rate <- class_survival[seq(1, length(class_survival), by = 2)] /
            (class_survival[seq(1, length(class_survival), by = 2)] +
            class_survival[seq(2, length(class_survival), by = 2)])
plot(x = c(1,2,3,4), y = class_survival_rate,
    main = "Survival Rate by Passenger Class",
    xlab = "Passenger Class", ylab = "Survival Rate",
    col = rainbow(length(rownames(Titanic))), pch = 16,
        xlim = c(0.5,4.5))
legend("topright", legend = rownames(Titanic), col = rainbow(length(rownames(Titanic))), pch = 16, title = "Passenger
Class")
```
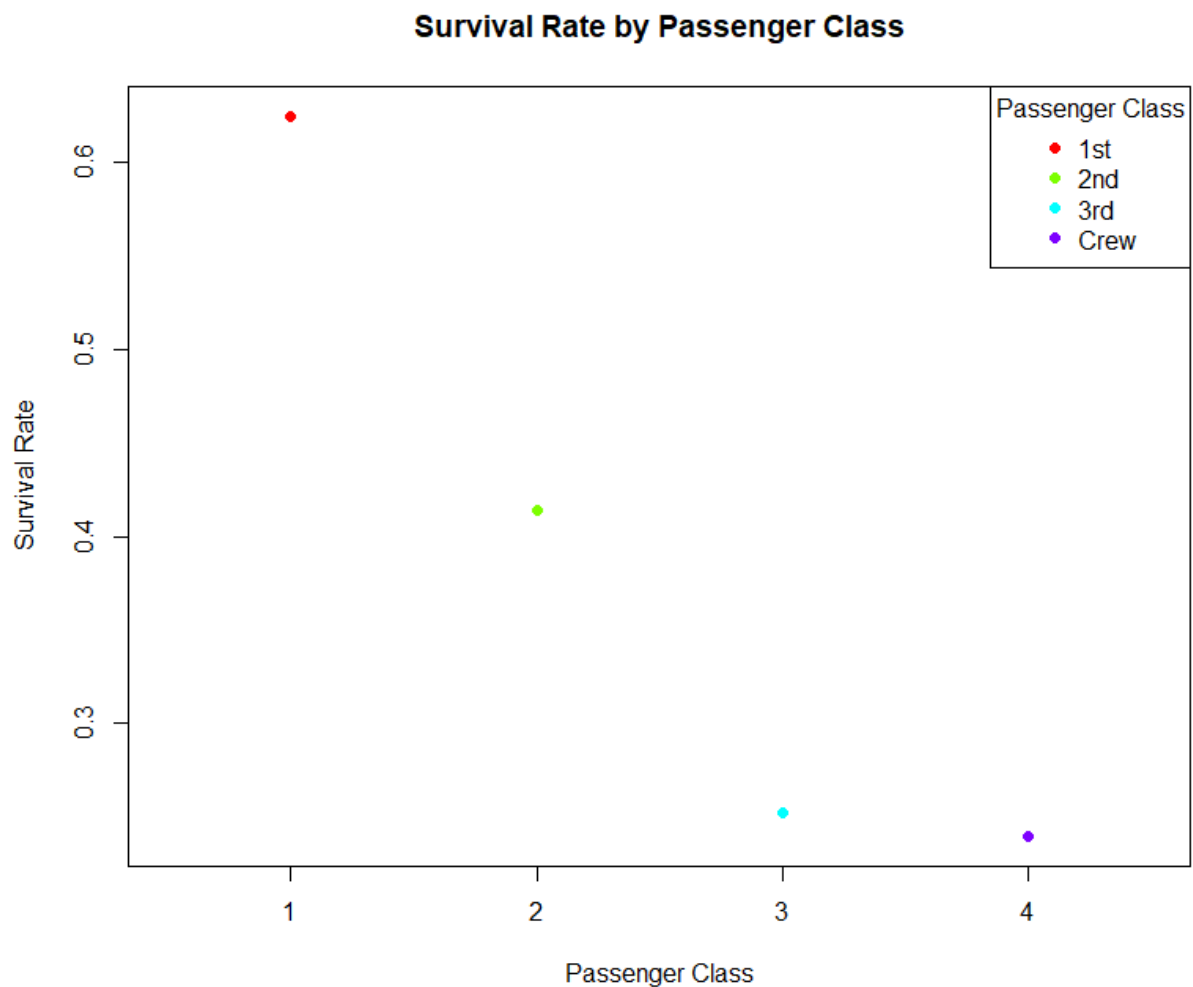

Survival Rate by Passenger Class

```
# 10. Heatmap of Survival Rates by class and gender
          # This heatmap displays the Survival Rates of classes and genders/
          # It is color-coded such that red indicates lower survival rates and green indicates higher rates
survival_rates <- apply(Titanic[, , , "Yes"], c(1, 2), sum) /
          apply(Titanic, c(1, 2), sum)
survival_df <- as.data.frame(survival_rates)
survival_df_long <- reshape(survival_df, direction = "long", idvar = "Class", varying = list(c("Male", "Female")), timevar
= "Gender", v.names = "Survival Rate")
survival_df_long$Class <- c("1st","2nd","3rd","Crew","1st","2nd","3rd","Crew")
survival_df_long$Gender<-c("Male","Male","Male","Male","Female","Female","Female","Female")
heatmap<-ggplot(survival_df_long, aes(x = Gender, y = Class, fill = `Survival Rate`)) +
 geom_tile() +
 scale_fill_gradient(low = "red", high = "green") +
 labs(title = "Survival Rates by Class and Gender",
      x = "Gender",
      y = "Class",
      fill = "Survival Rate") +
 theme_minimal() +
 theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(heatmap)
```


Survival Rates by Class and Gender