# R + Git Assignment 1  
# Author: Harsh Patel  
  
# Load dataset  
data <- as.data.frame(Titanic)  
  
# Inspect data  
head(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

# Descriptive stats  
summary(data)

## 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

table(data$Class)

##   
## 1st 2nd 3rd Crew   
## 8 8 8 8

table(data$Sex)

##   
## Male Female   
## 16 16

# New variable: Survival Rate = Survived count / Total count per Class  
library(dplyr)

##   
## Attaching package: 'dplyr'

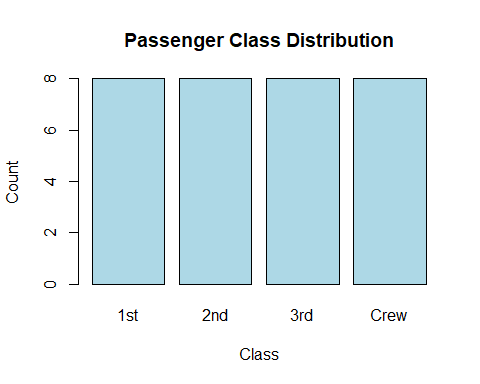
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

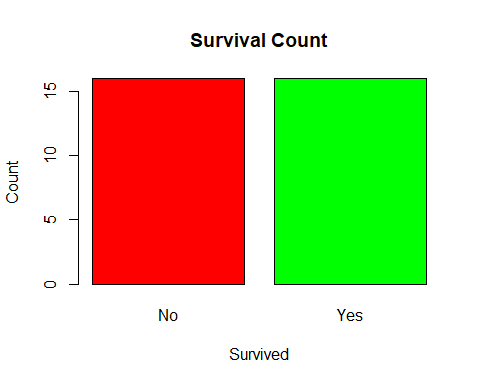
data <- data %>%  
 group\_by(Class, Survived) %>%  
 summarise(Total = sum(Freq), .groups = "drop") %>%  
 group\_by(Class) %>%  
 mutate(SurvivalRate = Total / sum(Total))  
  
head(data)

## # A tibble: 6 × 4  
## # Groups: Class [3]  
## Class Survived Total SurvivalRate  
## <fct> <fct> <dbl> <dbl>  
## 1 1st No 122 0.375  
## 2 1st Yes 203 0.625  
## 3 2nd No 167 0.586  
## 4 2nd Yes 118 0.414  
## 5 3rd No 528 0.748  
## 6 3rd Yes 178 0.252

# Plots  
barplot(table(as.data.frame(Titanic)$Class),  
 main="Passenger Class Distribution",  
 xlab="Class", ylab="Count", col="lightblue")



barplot(table(as.data.frame(Titanic)$Survived),  
 main="Survival Count",  
 xlab="Survived", ylab="Count", col=c("red","green"))



plot(jitter(as.numeric(data$Class)), data$SurvivalRate,  
 pch=19, col=c("red","green")[as.numeric(data$Survived)],  
 xaxt="n", xlab="Passenger Class", ylab="Survival Rate",  
 main="Survival Rate by Class")  
axis(1, at=1:4, labels=levels(data$Class))  
legend("topright", legend=levels(data$Survived),  
 col=c("red","green"), pch=19)

