R_Assignment_5

Maria Paz Paredes Lagos

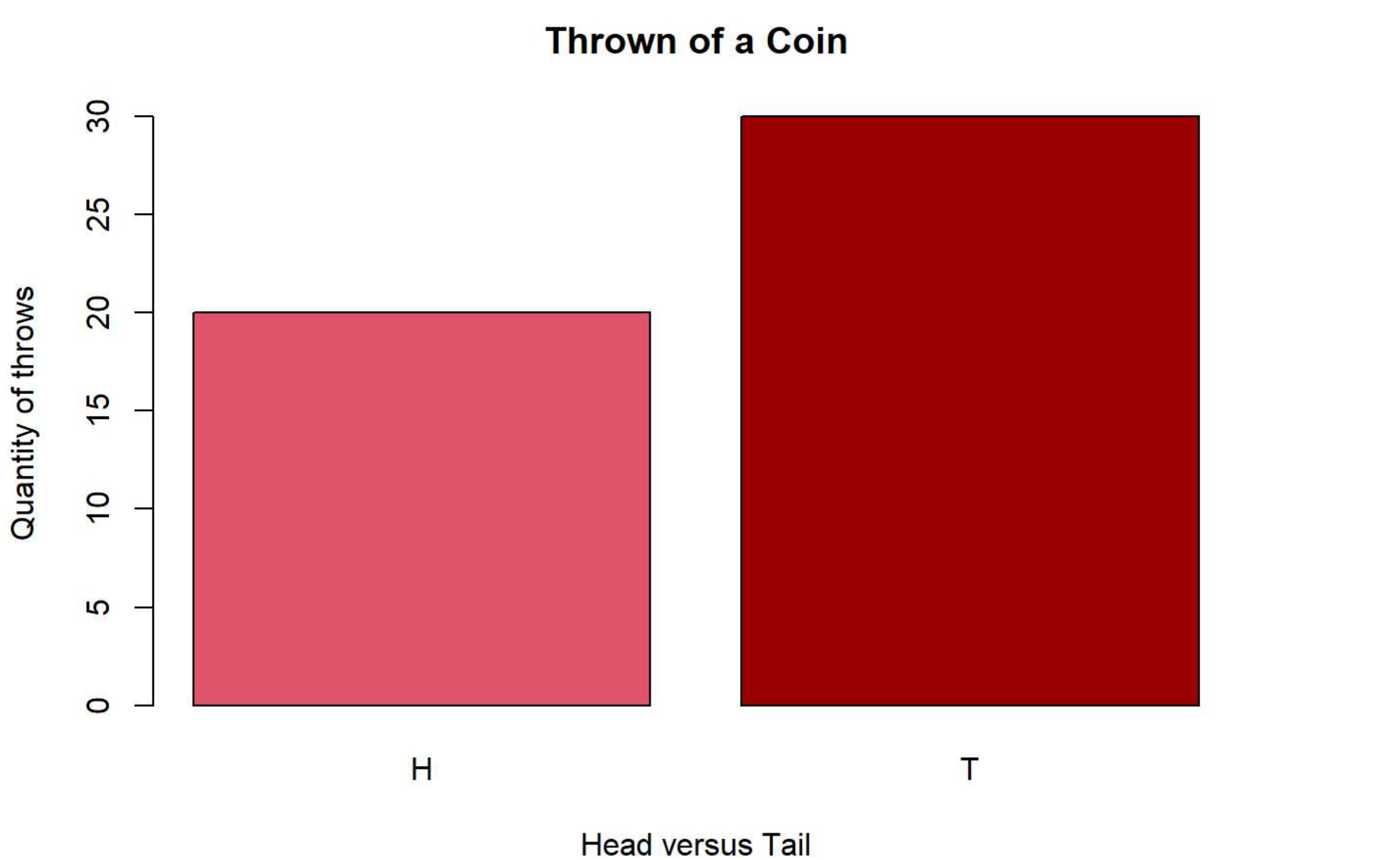
2023-12-01

Question 1: This code takes a sample of probabilities of a coin (Head versus Tail) randomly 50 times, with replacement in every run. We called the function “Coin.Throw” and “Tab” to assigned the table of this one. Barplot will help us to visualize the probability of the coin and the colors will differentiate between tail and head.

```
Coin.Throw<-sample(c('H','T'),50 ,replace=TRUE)
Coin.Throw
```

```
##  [1] "T" "H" "T" "T" "T" "T" "H" "H" "T" "H" "H" "T" "H" "H" "H" "H" "T" "T" "T"
## [20] "T" "H" "T" "T" "H" "T" "H" "T" "H" "T" "H" "H" "T" "T" "T" "T" "T" "H" "H"
## [39] "H" "H" "T" "T" "T" "T" "T" "T" "T" "T" "H" "T"
```

```
tab<-table(Coin.Throw)
barplot(tab,
        xlab= "Head versus Tail",
        ylab= "Quantity of throws",
        main= "Thrown of a Coin",
        col=c("#DF536B","#990000"))
```



Question 2: Similar to the problem 1, this code will show the probability of a coin with a sample size of 100. The relative frequency will measure the likelihood of the event occurring compared to the total of trails. We can calculated by dividing the number of favorable outcomes (cumulative sum of elements) by the total number of flips.

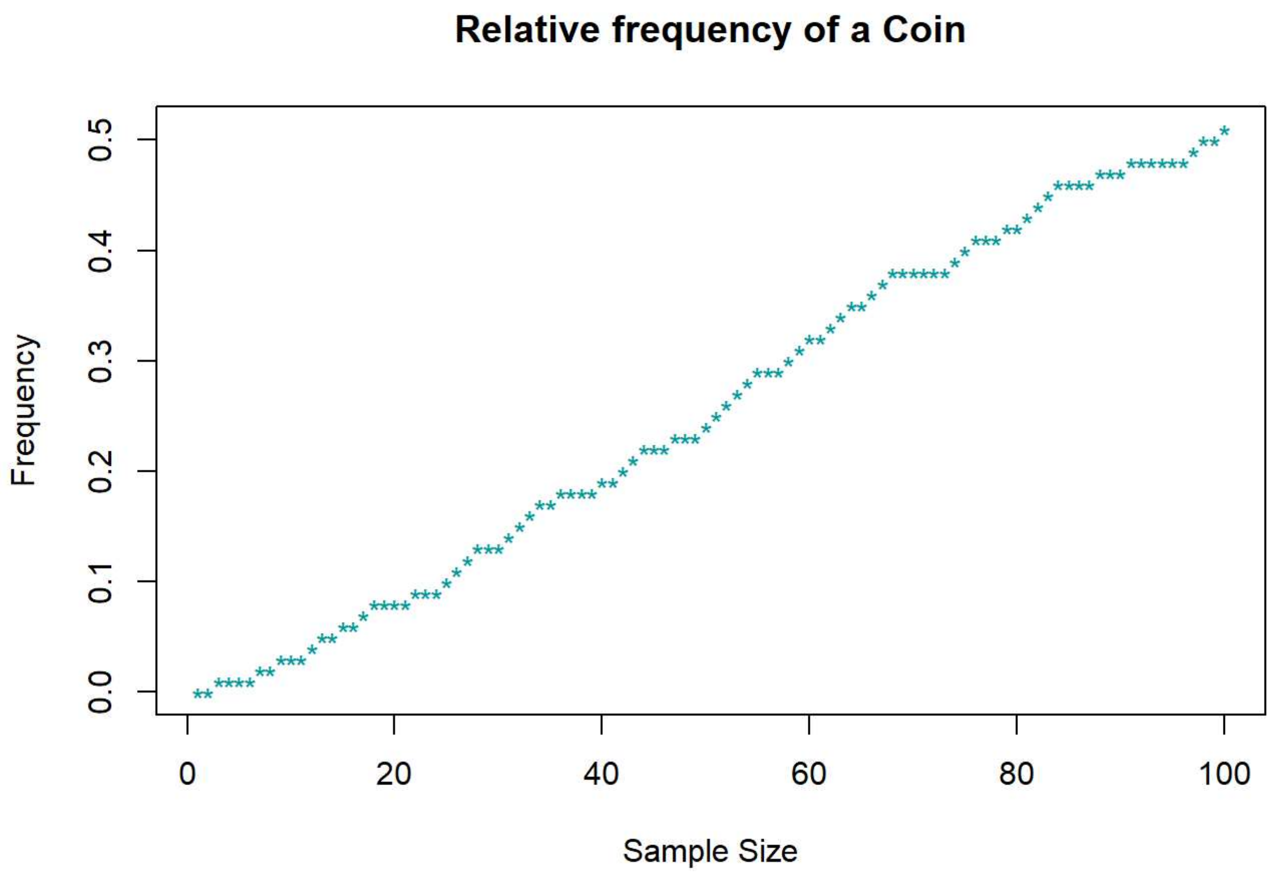
```
CoinThrow2<-sample(c(0, 1), 100, replace=TRUE)
table(CoinThrow2)
```

```
## CoinThrow2
##  0  1
## 49 51
```

```
cumsum(CoinThrow2)
```

```
##   [1]  0  0  1  1  1  1  2  2  3  3  3  4  5  5  6  6  7  8  8  8  8  9  9  9 10
##  [26] 11 12 13 13 13 14 15 16 17 17 18 18 18 18 19 19 20 21 22 22 22 23 23 23 24
##  [51] 25 26 27 28 29 29 29 30 31 32 32 33 34 35 35 36 37 38 38 38 38 38 38 39 40
##  [76] 41 41 41 42 42 43 44 45 46 46 46 46 47 47 47 48 48 48 48 48 48 49 50 50 51
```

```
Relfreq<-cumsum(CoinThrow2)/length(CoinThrow2)
num<-1:100
plot(num, Relfreq,
     pch ="*",
     ylab = "Frequency",
     xlab = "Sample Size",
     main = "Relative frequency of a Coin",
     col = "#009999")
```



Question 3: Two players gambling on a coin game, Amy wins 1 dollar if head comes up or loss 1 dollar otherwise. This parameter will help us to write our code, our sample will between (1, -1) from winning or losing, with a sample space of 100 and replacement in every run. As well as in the last question, we need to know the relative frequency and plot it.

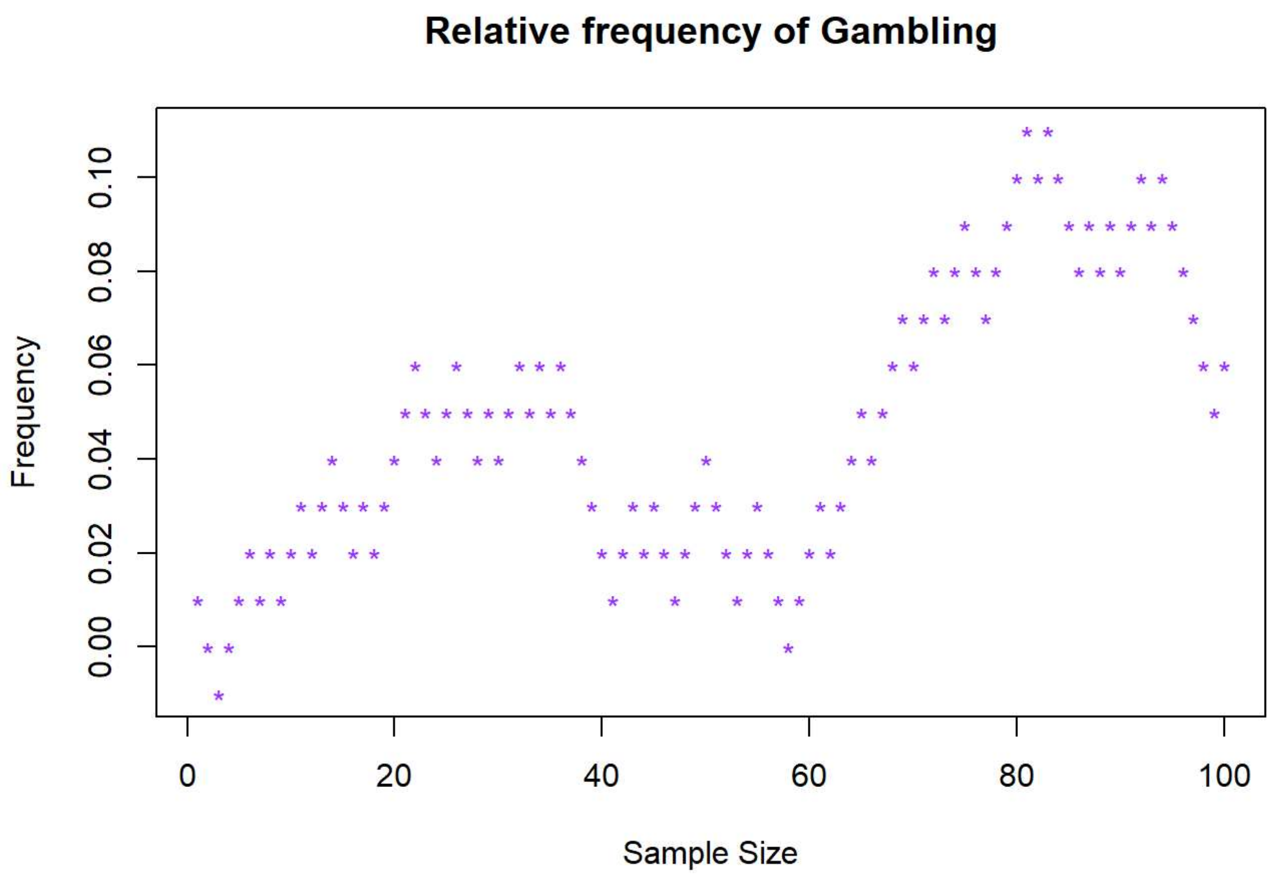
```
Gambling <-sample(c(1, -1),100, replace = TRUE)
table(Gambling)
```

```
## Gambling
## -1  1
## 47 53
```

```
cumsum(Gambling)
```

```
##   [1]  1  0 -1  0  1  2  1  2  1  2  3  2  3  4  3  2  3  2  3  4  5  6  5  4  5
##  [26]  6  5  4  5  4  5  6  5  6  5  6  5  4  3  2  1  2  3  2  3  2  1  2  3  4
##  [51]  3  2  1  2  3  2  1  0  1  2  3  2  3  4  5  4  5  6  7  6  7  8  7  8  9
##  [76]  8  7  8  9 10 11 10 11 10  9  8  9  8  9  8  9 10  9 10  9  8  7  6  5  6
```

```
Relfreq <- cumsum(Gambling)/length(Gambling)
num<-1:100
plot(num, Relfreq,
     pch = "*",
     ylab = "Frequency",
     xlab = "Sample Size",
     main = "Relative frequency of Gambling",
     col = "#9933FF")
```



Question 4: We were giving data from the Titanic and our first step is to load our data into the program, we did this by storing the exact location of the data in the variable “mydata1” with the mode of reading The head function will give us part of the data, and the function names will give us the names of the variables for the two-way tables.

```
mydata1 <-read.csv("C:\\Users\\maria\\Downloads\\TitanicDataA5.csv", header = TRUE)

head(mydata1)
```

```
##                                             Name    Sex   Age    Pclass
## 1                            Mr. Owen Harris Braund   Male 18-30 3rd Class
## 2 Mrs. John Bradley (Florence Briggs Thayer) Cumings Female 31-60 1st Class
## 3                             Miss. Laina Heikkinen Female 18-30 3rd Class
## 4        Mrs. Jacques Heath (Lily May Peel) Futrelle Female 31-60 1st Class
## 5                           Mr. William Henry Allen   Male 31-60 3rd Class
## 6                                   Mr. James Moran   Male 18-30 3rd Class
##       Fate
## 1     Died
## 2 Survived
## 3 Survived
## 4 Survived
## 5     Died
## 6     Died
```

```
names(mydata1)
```

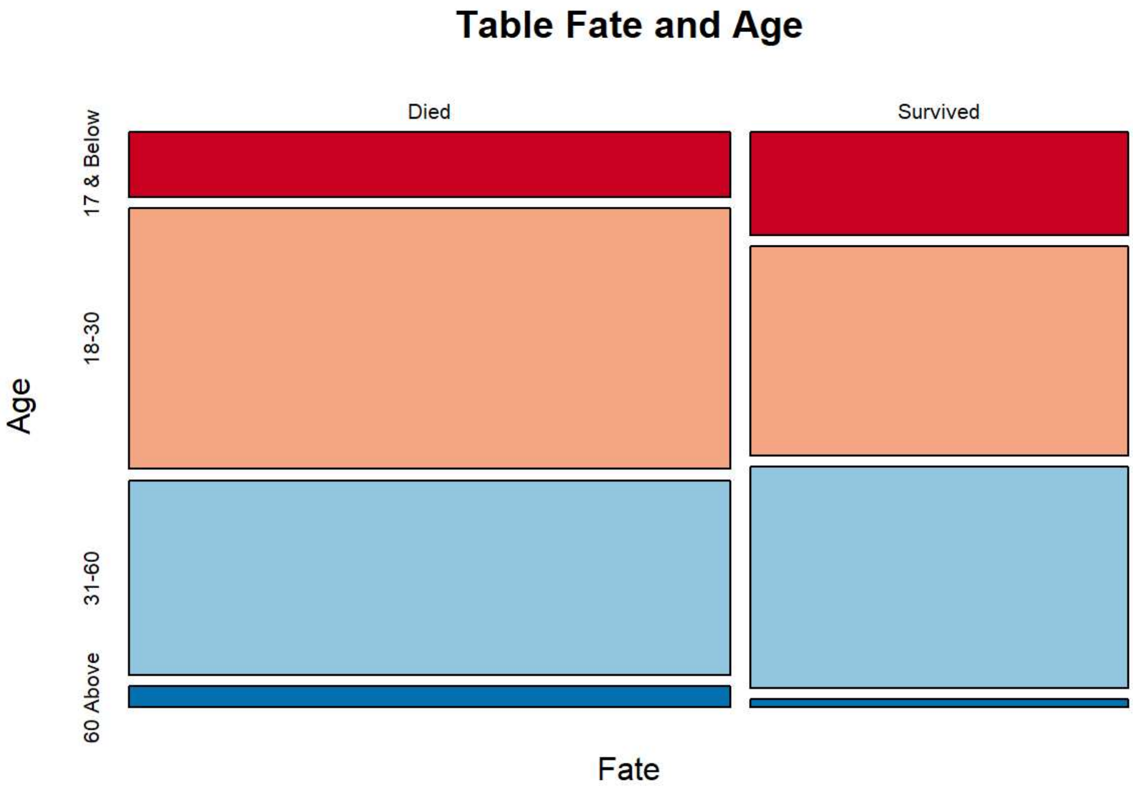
```
## [1] "Name"   "Sex"    "Age"    "Pclass" "Fate"
```

C. We will store every variable with their respective names from the giving data. This will help us to create easy and clear table with the data. The tables will help us with the probabilities of the variables.

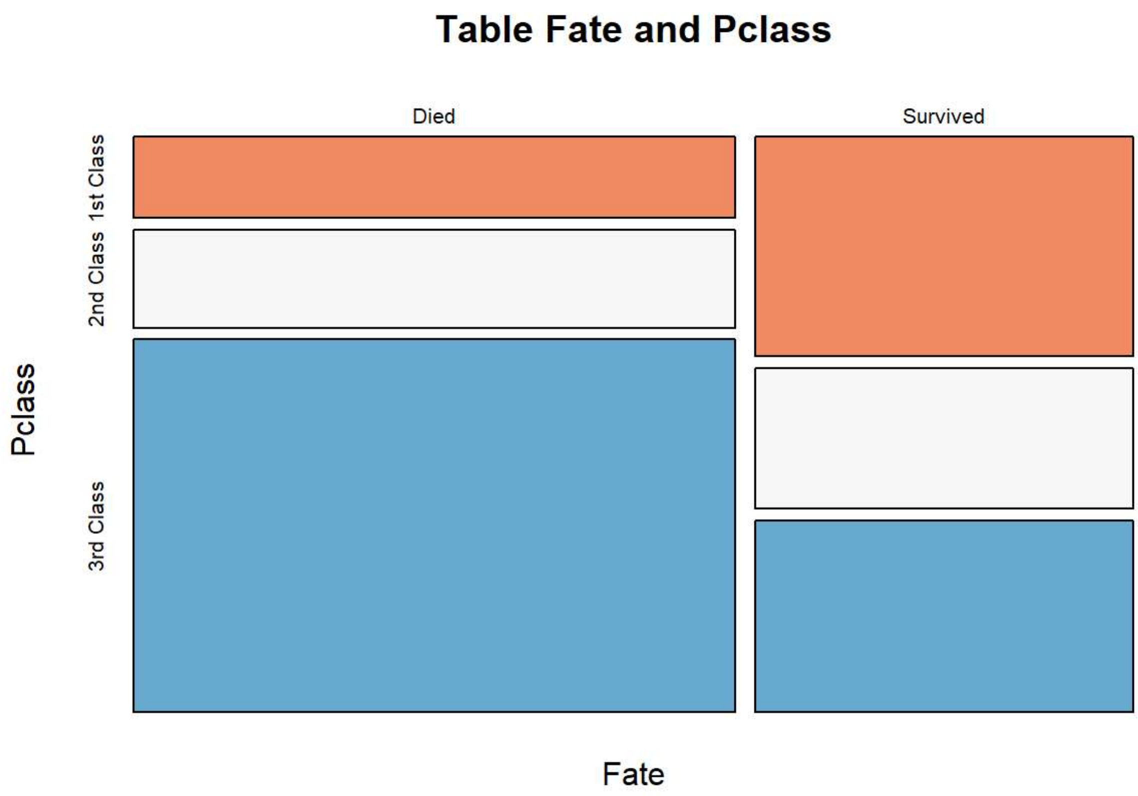
```
Fate<-mydata1$Fate
Age<- mydata1$Age
Sex<-mydata1$Sex
Pclass<-mydata1$Pclass

TableFA<-table(Fate, Age)
TableFS<-table(Fate, Sex)
TableFP<-table(Fate, Pclass)
```

D. Installing a colors palette to plot graphs from the tables that we already did. The colors will help us to differentiate between the outputs. Each graph have their respective title and axis names.

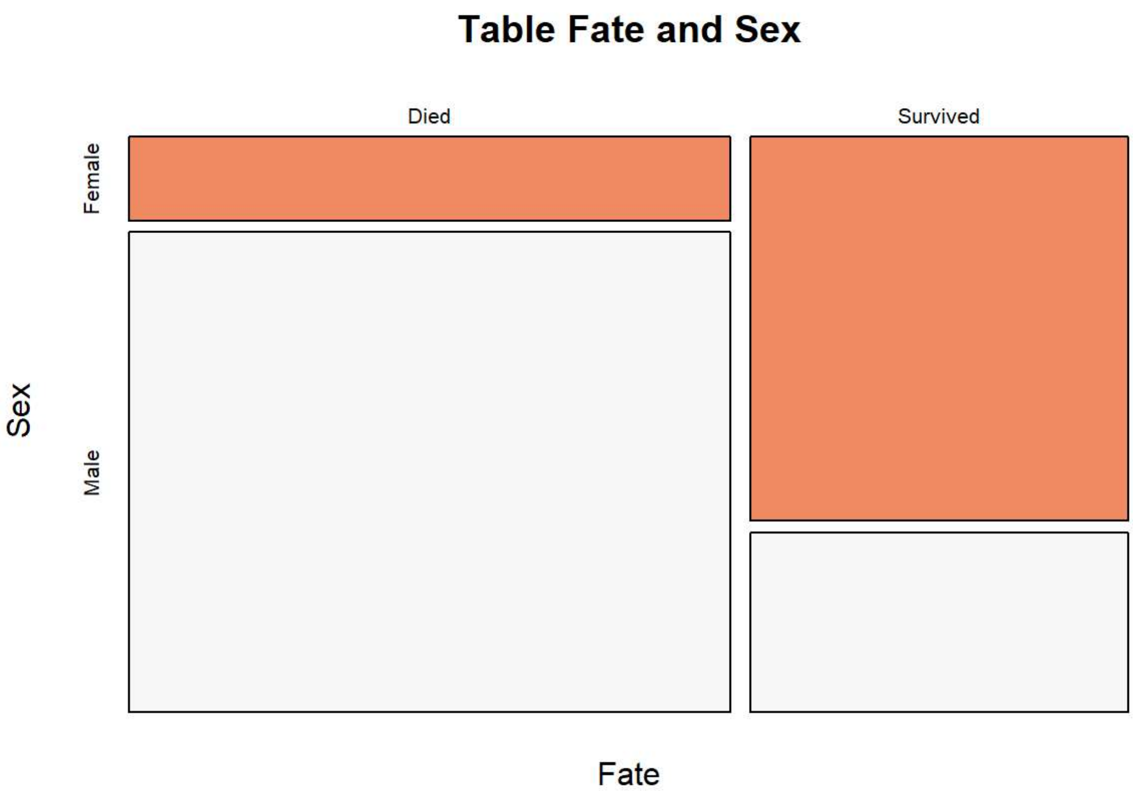
```
library("RColorBrewer")
plot(TableFA,
     main = "Table Fate and Age",
     col = brewer.pal(n= 4, name = "RdBu"))
```




```
plot(TableFP,
     main = "Table Fate and Pclass",
     col = brewer.pal(n= 3, name = "RdBu"))
```




```
plot(TableFS,
     main = "Table Fate and Sex",
     col = brewer.pal(n= 3, name = "RdBu"))
```



(E)

To calculate relevant probabilities, I worked with conditional probabilities based on the variables. We will use margin in our probabilities. The “margin” parameter determines whether we want to calculate probabilities along rows (margin == 1) or columns (margin ==2)

```
prob_survival_given_class<-prop.table(TableFP, margin = 1)

TableSA<-table(mydata1$Sex, mydata1$Age)
prob_sex_given_age<-prop.table(TableSA, margin = 2)

Prob_fate_given_class<-prop.table(TableFP, margin = 2)

cat("Probability of survival given passenger class:")
```

```
## Probability of survival given passenger class:
```

```
prob_survival_given_class
```

```
##           Pclass
## Fate        1st Class 2nd Class 3rd Class
##   Died      0.1467890 0.1779817 0.6752294
##   Survived  0.3976608 0.2543860 0.3479532
```

```
cat("Probability of being male or female given age group:")
```

```
## Probability of being male or female given age group:
```

```
prob_sex_given_age
```

```
##
##          17 & Below     18-30     31-60  60 Above
##   Female  0.4846154 0.3265823 0.3511905 0.1538462
##   Male    0.5153846 0.6734177 0.6488095 0.8461538
```

```
cat("Probability of different fates given passanger class:")
```

```
## Probability of different fates given passanger class:
```

```
Prob_fate_given_class
```

```
##           Pclass
## Fate       1st Class 2nd Class 3rd Class
##   Died     0.3703704 0.5271739 0.7556468
##   Survived 0.6296296 0.4728261 0.2443532
```

F. This code uses the “itself” function to create a new variable “Survived” with values 1 for “Survived” and 0 for “Died”. The Overall survival rate is calculated by dividing the number of passenger who survived by the total of passengers

```
mydata1$Survived<-ifelse(mydata1$Fate == "Survived", 1, 0)

total_passengers<-nrow(mydata1)
total_passengers
```

```
## [1] 887
```

```
total_survived<-sum(mydata1$Survived == 1)
total_survived
```

```
## [1] 342
```

```
overall_survival_rate<-total_survived / total_passengers

print("Overall Survival Rate:")
```

```
## [1] "Overall Survival Rate:"
```

```
round(overall_survival_rate, 2)
```

```
## [1] 0.39
```

After finding the overall survival rate, we will do the same with every category (females, males, kids, young, adults, 1st class, 2nd class, 3rd class) To find each category we will sum the survival data with each category data.

```
survived_females <-sum(mydata1$Survived == 1 & mydata1$Sex == "Female")
female_survival_rate <-survived_females/total_survived
print("Overall Female Survival Rate:")
```

```
## [1] "Overall Female Survival Rate:"
```

```
round(female_survival_rate, 2)
```

```
## [1] 0.68
```

```
survived_males <-sum(mydata1$Survived == 1 & mydata1$Sex == "Male")
male_survival_rate<-survived_males/total_survived
print("Overall Male Survival Rate:")
```

```
## [1] "Overall Male Survival Rate:"
```

```
round(male_survival_rate, 2)
```

```
## [1] 0.32
```

```
kids_survival<-sum(mydata1$Survived == 1 & mydata1$Age == "17 & Below")
kids_rate<-kids_survival/total_survived
print("Overall Kids Survival Rate:")
```

```
## [1] "Overall Kids Survival Rate:"
```

```
round(kids_rate, 2)
```

```
## [1] 0.19
```

```
young_survival<-sum(mydata1$Survived == 1 & mydata1$Age == "18-30")
young_rate<-young_survival/total_survived
print("Overall Young Survival Rate:")
```

```
## [1] "Overall Young Survival Rate:"
```

```
round(young_rate, 2)
```

```
## [1] 0.39
```

```
adults_survival<-sum(mydata1$Survived == 1 & mydata1$Age == "31-60")
adults_rate<-adults_survival/total_survived
print("Overall Adults Survival Rate:")
```

```
## [1] "Overall Adults Survival Rate:"
```

```
round(adults_rate, 2)
```

```
## [1] 0.41
```

```
elderly_survival<-sum(mydata1$Survived == 1 & mydata1$Age == "61 Above")
elderly_rate<-elderly_survival/total_survived
print("Overall Elderly Survival Rate:")
```

```
## [1] "Overall Elderly Survival Rate:"
```

```
round(elderly_rate, 2)
```

```
## [1] 0
```

```
Fclass_survival<-sum(mydata1$Survived == 1 & mydata1$Pclass == "1st Class")
Fclass_rate<-Fclass_survival/total_survived
print("Overall First Class Survival Rate:")
```

```
## [1] "Overall First Class Survival Rate:"
```

```
round(Fclass_rate, 2)
```

```
## [1] 0.4
```

```
Sclass_survival<-sum(mydata1$Survived == 1 & mydata1$Pclass == "2nd Class")
Sclass_rate<-Sclass_survival/total_survived
print("Overall Second Class Survival Rate:")
```

```
## [1] "Overall Second Class Survival Rate:"
```

```
round(Sclass_rate, 2)
```

```
## [1] 0.25
```

```
Tclass_survival<-sum(mydata1$Survived == 1 & mydata1$Pclass == "3rd Class")
Tclass_rate<-Tclass_survival/total_survived
print("Overall Third Class Survival Rate:")
```

```
## [1] "Overall Third Class Survival Rate:"
```

```
round(Tclass_rate, 2)
```

```
## [1] 0.35
```

After we got every survival rate, we can determinate that females, adults, and first class have a significantly higher likelihood of survival.

Question 5: (A) To solve this question we will have a similar approach from the last question to load our data into the program. As well as, determinate each variable to create two ways tables.

```
mydata2 <-read.csv("C:\\Users\\maria\\Downloads\\ChildDataB52.csv", header = TRUE)
attach(mydata2)
```

```
## The following objects are masked _by_ .GlobalEnv:
##
##     Age, Sex
```

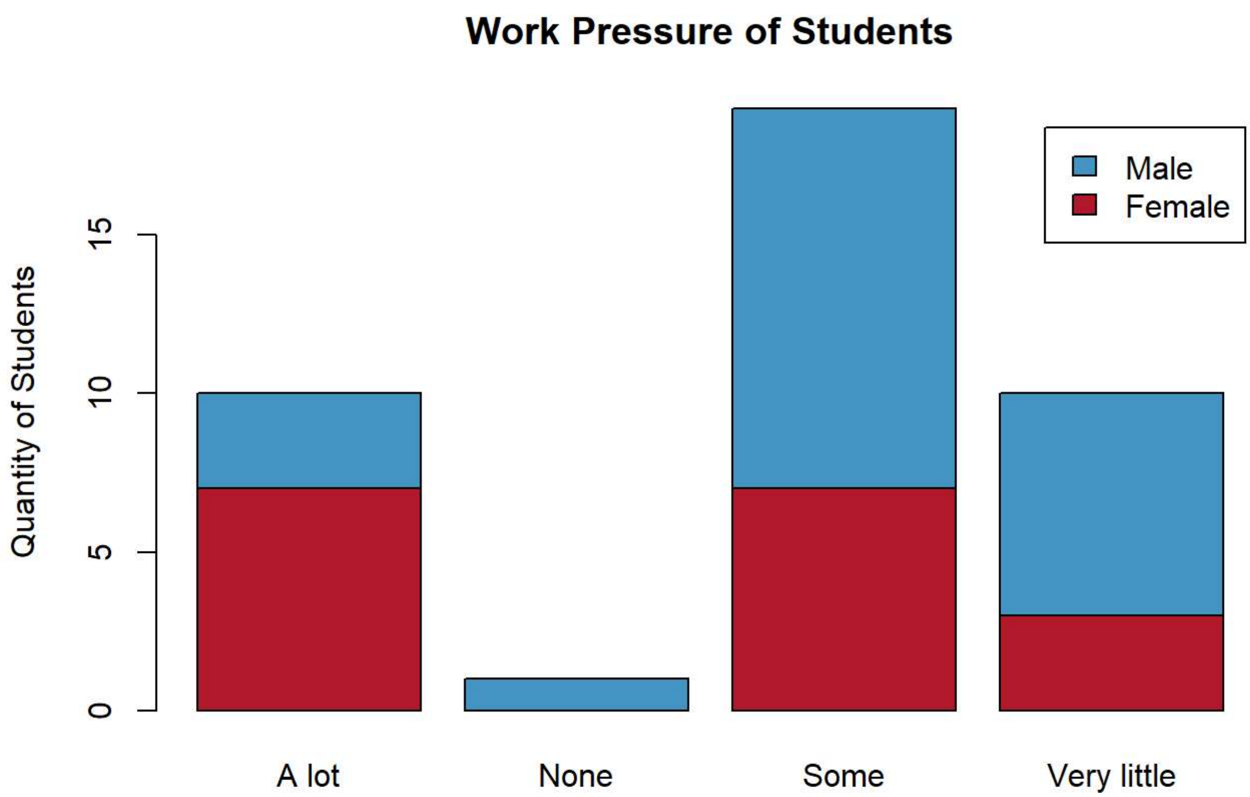
```
names(mydata2)
```

```
##  [1] "Grade"           "Height"          "Footlength"      "Age"
##  [5] "Sex"             "Domhand"         "FavSchSubject"   "HomeOccupants"
##  [9] "SchWorkPressure" "PlannedQual"     "Preferred2be"
```

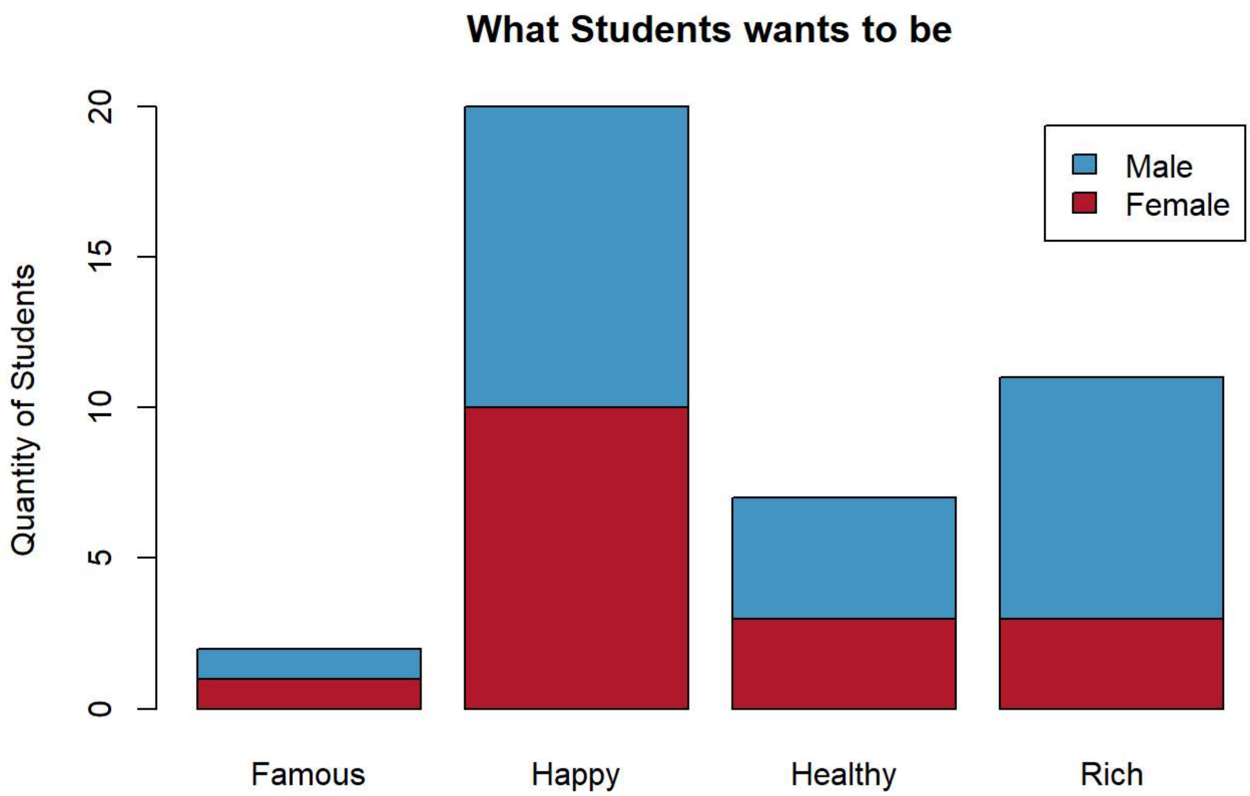
```
Sex<-mydata2$Sex
Domhand<-mydata2$Domhand
SchWorkPressure<-mydata2$SchWorkPressure
Preferred2be<-mydata2$Preferred2be

table_SD<-table(Sex, Domhand)
table_SW<-table(Sex, SchWorkPressure)
table_SP<-table(Sex, Preferred2be)
```

D. After we have every table, we will have a clean code to plot each graph. Every graph will have their respective names and we will use the function “legend.text” to differentiate the stack data.

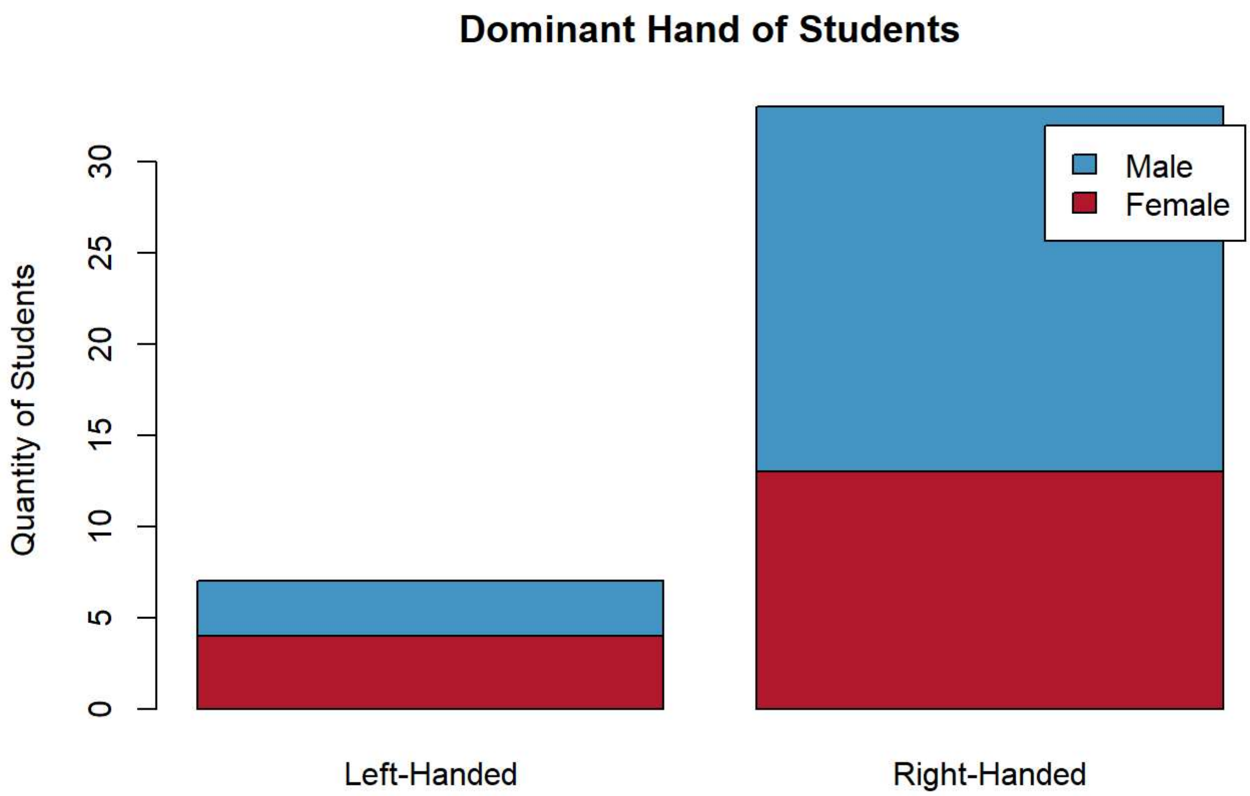
```
barplot(table_SW,
        ylab = "Quantity of Students",
        main = "Work Pressure of Students",
        col = c("#B2182B", "#4393C3"),
        legend.text = rownames(table_SD),
        beside = FALSE)
```




```
barplot(table_SP,
        ylab = "Quantity of Students",
        main = "What Students wants to be",
        col = c("#B2182B", "#4393C3"),
        legend.text = rownames(table_SD),
        beside = FALSE)
```




```
barplot(table_SD,
        ylab = "Quantity of Students",
        main = "Dominant Hand of Students",
        col = c("#B2182B", "#4393C3"),
        legend.text = rownames(table_SD),
        beside = FALSE)
```



E. After installing the package “psych” we will load it, the function “describe” is used to obtain descriptive statistics that includes columns 2 to 4.

```
library("psych")
describe(mydata2[,2:4])
```

```
##            vars  n   mean    sd median trimmed   mad min max range  skew
## Height        1 40 164.85 25.47    170  169.19 11.12  63 200   137 -2.81
## Footlength    2 40  24.61  3.26     25   24.80  2.97  12  30    18 -1.20
## Age           3 40  16.30  1.44     17   16.50  1.48  12  18     6 -1.28
##            kurtosis   se
## Height         8.59 4.03
## Footlength     3.51 0.52
## Age            1.06 0.23
```