```r
library(readr)
Titanic <- read_csv("Titanic.csv")
```

```
## Rows: 891 Columns: 8
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (2): Sex, Embarked
## dbl (6): Survived, Pclass, Age, SibSp, Parch, Fare
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
summary(Titanic)
```

```
##     Survived          Pclass          Sex                 Age
##  Min.   :0.0000   Min.   :1.000   Length:891         Min.   : 0.42
##  1st Qu.:0.0000   1st Qu.:2.000   Class :character   1st Qu.:20.12
##  Median :0.0000   Median :3.000   Mode  :character   Median :28.00
##  Mean   :0.3838   Mean   :2.309                      Mean   :29.70
##  3rd Qu.:1.0000   3rd Qu.:3.000                      3rd Qu.:38.00
##  Max.   :1.0000   Max.   :3.000                      Max.   :80.00
##                                                      NA's   :177
##      SibSp           Parch            Fare            Embarked
##  Min.   :0.000   Min.   :0.0000   Min.   :  0.00   Length:891
##  1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:  7.91   Class :character
##  Median :0.000   Median :0.0000   Median : 14.45   Mode  :character
##  Mean   :0.523   Mean   :0.3816   Mean   : 32.20
##  3rd Qu.:1.000   3rd Qu.:0.0000   3rd Qu.: 31.00
##  Max.   :8.000   Max.   :6.0000   Max.   :512.33
##
```
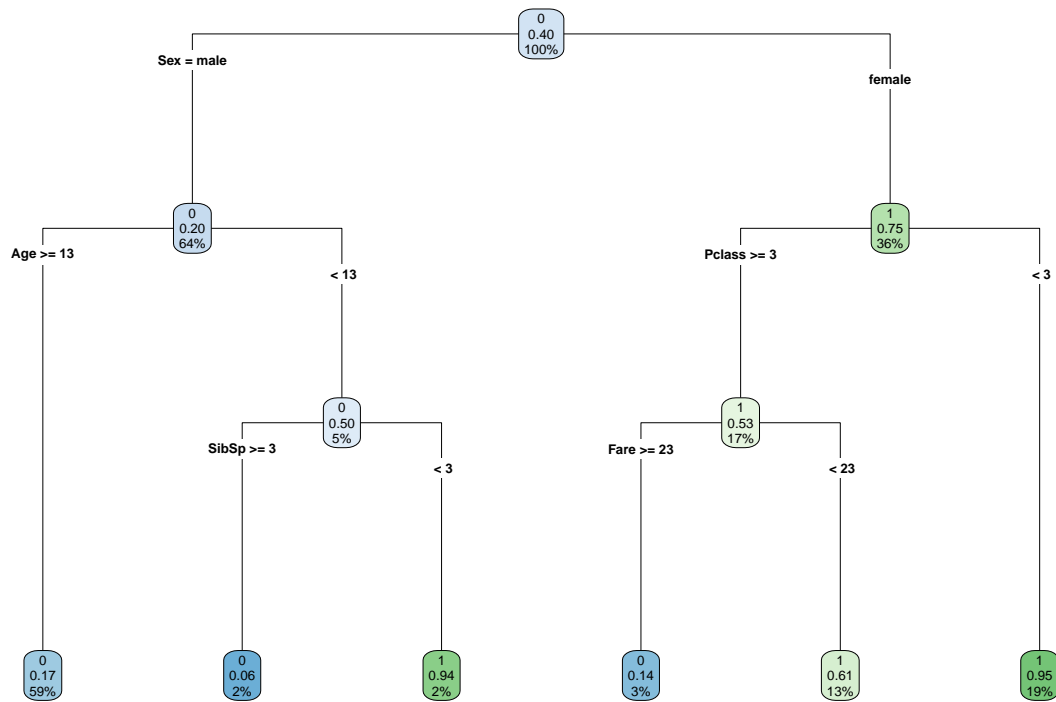
```r
library('ggplot2')
library('lattice')
library('caret')
library('rpart')
library('rpart.plot')
set.seed(80)
```

```r
?createDataPartition
trainIndex <- createDataPartition(Titanic$Survived, p=0.8, list=FALSE, time=1)
```

```r
trainset <- Titanic[trainIndex, ]
testset <- Titanic[-trainIndex,]
```

```r
tree<-rpart(Survived ~ ., trainset, method="class")
```

```r
rpart.plot(tree, cex=0.4, type=4)
```

#validate

```
predict_tree <- predict(tree, testset, type="class")
table(testset$Survived, predict_tree)
```

```
##     predict_tree
##       0   1
##   0 106  13
##   1  16  43
```

#TN = 106 FP = 13 FN = 16 TP = 43 #Sensitivity= TP/(TP+FN) #specificity= TN/(TN+FP) #Accuracy= (TP+TN+FP+FN)/(TP+TN)

#sen = TF/actual positive = 43/(43+16)= 0.729= 72.9% #specificity= TN/ actual Negative 106/(106+13)= 0.891 = 89.1% #accuracy = (43+106)/(106+13+16+43)= 0.837= 83.7%

#3. Will the following passengers survive? (You can answer this question by looking at the tree) Pclass Sex Age SibSp Parch Fare Embarked #3male 22 1 0 7.25 S #according to the descion tree, a male greater than 13 does not survive