lab1.2\_estes

andrew.estes

8/21/2021

#load library

library(titanic)

#step 3

summary(titanic\_train)

## PassengerId Survived Pclass Name   
## Min. : 1.0 Min. :0.0000 Min. :1.000 Length:891   
## 1st Qu.:223.5 1st Qu.:0.0000 1st Qu.:2.000 Class :character   
## Median :446.0 Median :0.0000 Median :3.000 Mode :character   
## Mean :446.0 Mean :0.3838 Mean :2.309   
## 3rd Qu.:668.5 3rd Qu.:1.0000 3rd Qu.:3.000   
## Max. :891.0 Max. :1.0000 Max. :3.000   
##   
## Sex Age SibSp Parch   
## Length:891 Min. : 0.42 Min. :0.000 Min. :0.0000   
## Class :character 1st Qu.:20.12 1st Qu.:0.000 1st Qu.:0.0000   
## Mode :character Median :28.00 Median :0.000 Median :0.0000   
## Mean :29.70 Mean :0.523 Mean :0.3816   
## 3rd Qu.:38.00 3rd Qu.:1.000 3rd Qu.:0.0000   
## Max. :80.00 Max. :8.000 Max. :6.0000   
## NA's :177   
## Ticket Fare Cabin Embarked   
## Length:891 Min. : 0.00 Length:891 Length:891   
## Class :character 1st Qu.: 7.91 Class :character Class :character   
## Mode :character Median : 14.45 Mode :character Mode :character   
## Mean : 32.20   
## 3rd Qu.: 31.00   
## Max. :512.33   
##

#step 4

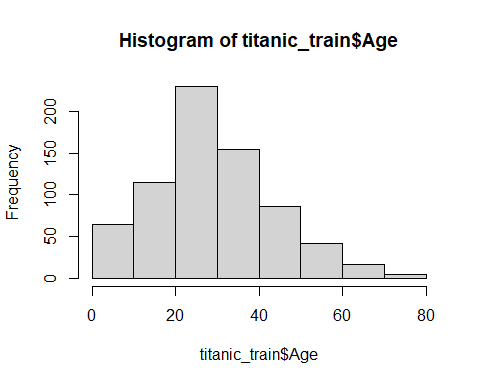
titanic\_table <- table(titanic\_train$Survived, titanic\_train$Pclass)  
titanic\_table

##   
## 1 2 3  
## 0 80 97 372  
## 1 136 87 119

1st class passengers were the only Pclass more likely to survive than not there were less 2 Pclass passengers than either 1Pclass or 3Pclass

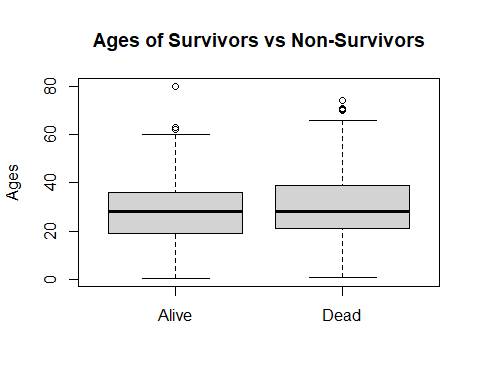
#step 5

hist(titanic\_train$Age)

 Younger passengers outlived older passengers Is that due to age or due to number of passengers within each age group?

#step 6

Alive = titanic\_train$Age[which(titanic\_train$Survived==1)]   
Dead = titanic\_train$Age[which(titanic\_train$Survived==0)]   
my.bp <- boxplot(Alive,Dead,  
 main = "Ages of Survivors vs Non-Survivors",  
 ylab = "Ages",  
 names = c("Alive", "Dead"))

 There appears to be no difference in the age distribution of who survived