Assignment_1

Tirumal Achina

2023-09-07

#Importing the data. #Source: chatGpt and github #Downloaded the dataset and read it from my desktop.

```
titanic <- read.csv("C:/Users/tirum/Desktop/titanic/titanic.csv")
#titanic
#the above line would give the whole dataset.
```

#Descriptive analysis: This shows the summary of whole dataset that we have given. It includes the Distribution, Measures of Central Tendency(Mean,Median,Mode),Measures of Variability(Standard Deviation,Range,Variance,Interquartile Range).

```
head(titanic) #This function shows the frst 6 rows and columns of the dataset
```

```
##   PassengerId Survived Pclass     Lname
## 1           1        0      3    Braund
## 2           2        1      1   Cumings
## 3           3        1      3 Heikkinen
## 4           4        1      1  Futrelle
## 5           5        0      3     Allen
## 6           6        0      3     Moran
##                                                 Name    Sex Age SibSp Parch
## 1                             Mr. Owen Harris   male  22     1     0
## 2 Mrs. John Bradley (Florence Briggs Thayer) female  38     1     0
## 3                                 Miss. Laina female  26     0     0
## 4         Mrs. Jacques Heath (Lily May Peel) female  35     1     0
## 5                           Mr. William Henry   male  35     0     0
## 6                                 Mr. James   male  NA     0     0
##             Ticket    Fare Cabin Embarked
## 1        A/5 21171  7.2500              S
## 2         PC 17599 71.2833   C85        C
## 3 STON/O2. 3101282  7.9250              S
## 4           113803 53.1000  C123        S
## 5           373450  8.0500              S
## 6           330877  8.4583              Q
```

```
tail(titanic) #This function shows the last 6 rows and columns of dataset.
```

```
##     PassengerId Survived Pclass        Lname                         Name    Sex
## 151         151        0      2      Bateman           Rev. Robert James   male
## 152         152        1      1        Pears  Mrs. Thomas (Edith Wearne) female
## 153         153        0      3          Meo                  Mr. Alfonzo   male
```

```
## 154         154        0      3 van Billiard           Mr. Austin Blyler   male
## 155         155        0      3        Olsen              Mr. Ole Martin   male
## 156         156        0      1     Williams           Mr. Charles Duane   male
##      Age SibSp Parch      Ticket    Fare Cabin Embarked
## 151 51.0     0     0 S.O.P. 1166 12.5250              S
## 152 22.0     1     0      113776 66.6000    C2        S
## 153 55.5     0     0  A.5. 11206  8.0500              S
## 154 40.5     0     2    A/5. 851 14.5000              S
## 155   NA     0     0   Fa 265302  7.3125              S
## 156 51.0     0     1    PC 17597 61.3792              C
```

```
summary(titanic) #This shows the whole summary that includes mean, median,mode and so on.
```

```
##   PassengerId        Survived          Pclass         Lname
##  Min.   :  1.00   Min.   :0.0000   Min.   :1.000   Length:156
##  1st Qu.: 39.75   1st Qu.:0.0000   1st Qu.:2.000   Class :character
##  Median : 78.50   Median :0.0000   Median :3.000   Mode  :character
##  Mean   : 78.50   Mean   :0.3462   Mean   :2.423
##  3rd Qu.:117.25   3rd Qu.:1.0000   3rd Qu.:3.000
##  Max.   :156.00   Max.   :1.0000   Max.   :3.000
##
##      Name               Sex                 Age            SibSp
##  Length:156         Length:156         Min.   : 0.83   Min.   :0.0000
##  Class :character   Class :character   1st Qu.:19.00   1st Qu.:0.0000
##  Mode  :character   Mode  :character   Median :26.00   Median :0.0000
##                                        Mean   :28.14   Mean   :0.6154
##                                        3rd Qu.:35.00   3rd Qu.:1.0000
##                                        Max.   :71.00   Max.   :5.0000
##                                        NA's   :30
##      Parch           Ticket              Fare            Cabin
##  Min.   :0.0000   Length:156         Min.   :  6.750   Length:156
##  1st Qu.:0.0000   Class :character   1st Qu.:  8.003   Class :character
##  Median :0.0000   Mode  :character   Median : 14.454   Mode  :character
##  Mean   :0.3974                      Mean   : 28.110
##  3rd Qu.:0.0000                      3rd Qu.: 30.372
##  Max.   :5.0000                      Max.   :263.000
##
##    Embarked
##  Length:156
##  Class :character
##  Mode  :character
##
##
##
##
```

#Transforming variables: This can be done in variable ways such as log,exponential,normalization and so on...For this dataset I choosed Normalization Transformation. #Normalization Transformation: This changes the scale of variable where the mean becomes 0 and the standard deviation becomes 1. It can be done by min-max or z-score normalization.I have Normalized Fare column in the titanic table.

```r
min_max_normalize <- function(x) {
  return((x - min(x)) / (max(x) - min(x)))
}
#Here the Age got normalized and we can find that values in the Age_Norm column
titanic$Age_norm <- min_max_normalize(titanic$Age)
head(titanic)
```

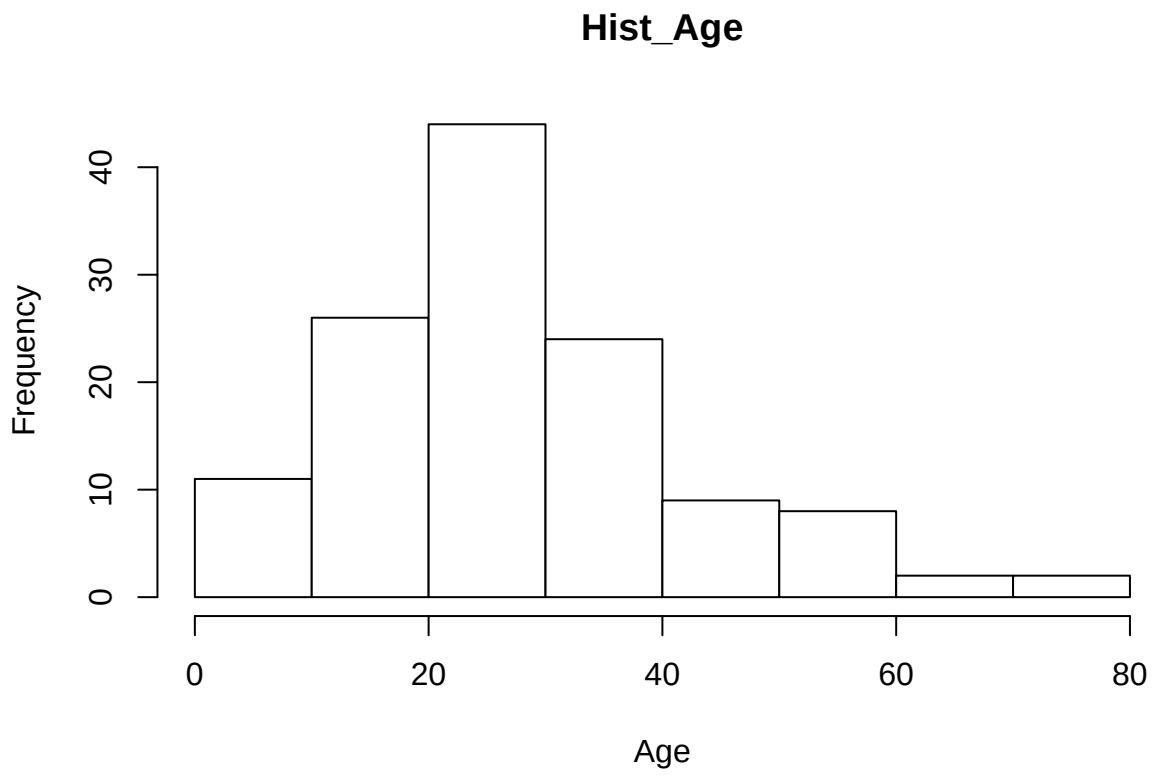
```
##   PassengerId Survived Pclass     Lname
## 1           1        0      3    Braund
## 2           2        1      1   Cumings
## 3           3        1      3 Heikkinen
## 4           4        1      1  Futrelle
## 5           5        0      3     Allen
## 6           6        0      3     Moran
##                                                 Name    Sex Age SibSp Parch
## 1                             Mr. Owen Harris   male  22     1     0
## 2 Mrs. John Bradley (Florence Briggs Thayer) female  38     1     0
## 3                                 Miss. Laina female  26     0     0
## 4         Mrs. Jacques Heath (Lily May Peel) female  35     1     0
## 5                           Mr. William Henry   male  35     0     0
## 6                                  Mr. James   male  NA     0     0
##             Ticket    Fare Cabin Embarked Age_norm
## 1        A/5 21171  7.2500             S       NA
## 2         PC 17599 71.2833   C85        C       NA
## 3 STON/O2. 3101282  7.9250             S       NA
## 4           113803 53.1000  C123        S       NA
## 5           373450  8.0500             S       NA
## 6           330877  8.4583             Q       NA
```

```r
#Log Transformation to Fare Column.
titanic$Fare_log = log(titanic$Fare)
titanic$Fare_log
```

```
##   [1] 1.981001 4.266662 2.070022 3.972177 2.085672 2.135148 3.948596 3.048088
##   [9] 2.409941 3.403555 2.815409 3.279030 2.085672 3.442819 2.061048 2.772589
##  [17] 3.371597 2.564949 2.890372 1.977547 3.258097 2.564949 2.083085 3.569533
##  [25] 3.048088 3.446410 1.977547 5.572154 2.064226 2.066331 3.322183 4.987167
##  [33] 2.047693 2.351375 4.408800 3.951244 1.978128 2.085672 2.890372 2.419630
##  [41] 2.248657 3.044522 2.066331 3.727600 2.064226 2.085672 2.740840 2.047693
##  [49] 3.076353 2.879198 3.681036 2.054124 4.340282 3.258097 4.126799 3.569533
##  [57] 2.351375 1.978128 3.323236 3.848018 1.978128 4.382027 4.424547 3.328627
##  [65] 3.322183 2.724304 2.351375 2.099036 2.070022 2.159003 2.351375 3.848018
##  [73] 4.297285 2.670985 4.034166 2.034706 2.066331 2.085672 3.367296 2.523727
##  [81] 2.197225 2.251292 2.052520 3.852273 2.351375 2.763170 3.537330 2.085672
##  [89] 5.572154 2.085672 2.085672 2.061048 4.113739 3.024077 1.981001 2.085672
##  [97] 3.545419 4.148806 3.135494 3.258097 2.066331 2.066331 4.347532 2.158045
## [105] 2.070022 2.066331 2.034706 2.050913 2.066331 3.184284 3.951244 2.670985
## [113] 2.085672 2.284930 2.671269 2.070022 2.047693 3.044522 5.511495 3.442819
## [121] 4.297285 2.085672 3.403555 2.564949 4.347532 2.419630 2.047693 1.965951
## [129] 3.107198 1.942332 2.066331 1.953028 2.674149 3.258097 2.564949 2.711099
## [137] 3.268934 3.972177 2.221017 4.371976 2.724304 2.047693 2.763170 1.909543
## [145] 2.442347 3.604138 2.053585 3.537330 3.258097 2.564949 2.527727 4.198705
## [153] 2.085672 2.674149 1.989585 4.117071
```

#Hist plots

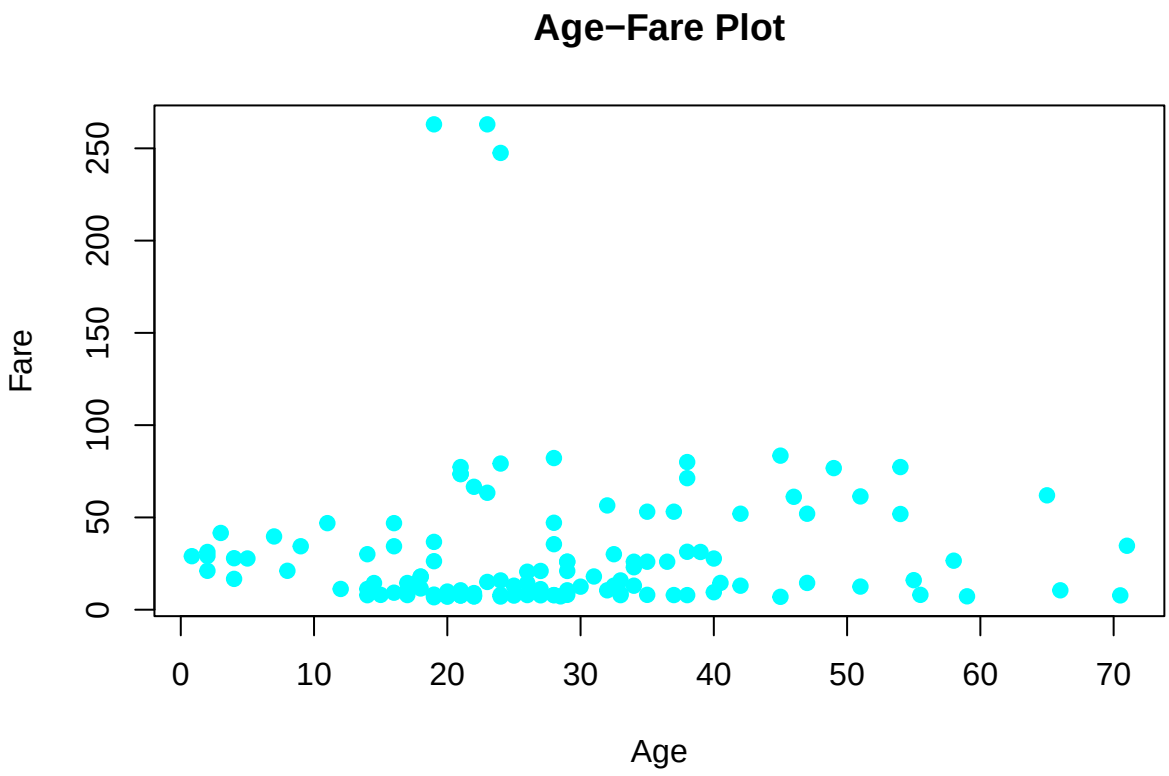
```
hist(titanic$Age, xlab = "Age", ylab = "Frequency",main = "Hist_Age",col = "white")
```




```
#plot(titanic$Age, xlab = "Age",ylab = "Frequency",main = "Hist_Age)
```

#Marking the Scatter plot for Age and Fare

```
x = titanic$Age #Age on the x-axis
y = titanic$Fare #Fare on the y-axis
plot(x,y, xlab ="Age",ylab = "Fare",main = "Age-Fare Plot",pch = 19,col =  "cyan")
```




```
x = titanic$Age
y = titanic$Pclass
plot(x,y, xlab ="Age",ylab = "Pclass",main = "Age-Pclass Plot",pch = 19,col =  "cyan")
```


Age−Pclass Plot
Pclass
1.0
1.5
2.0
2.5
3.0
0
10
20
30
40
50
60
70
Age