**Assignment-EDA**

# univariate non graphical : Categorical

> Titanic\_Survived\_sum<-summary(Titanic$Survived)

> Titanic\_Survived\_sum

0 1

549 342

> Titanic\_Survived\_table<-as.matrix(Titanic\_Survived\_sum)

> Titanic\_Survived\_table

[,1]

0 549

1 342

> Titanic\_Survived\_table<-as.matrix(Titanic\_Survived\_sum,Titanic\_Survived\_sum/sum(Titanic\_Survived\_sum))

> Titanic\_Survived\_table

[,1]

0 549

1 342

> Titanic\_Survived\_table<-cbind(Titanic\_Survived\_table,Titanic\_Survived\_table/sum(Titanic\_Survived\_table))

> Titanic\_Survived\_table

[,1] [,2]

0 549 0.6161616

1 342 0.3838384

#Pclass

> Titanic\_Pclass\_sum<-summary(Titanic$Pclass)

> Titanic\_Pclass\_table<-as.matrix(Titanic\_Pclass\_sum)

> Titanic\_Pclass\_table

[,1]

1 216

2 184

3 491

> Titanic\_Pclass\_table<-cbind(Titanic\_Pclass\_table,Titanic\_Pclass\_table/sum(Titanic\_Pclass\_table))

> Titanic\_Pclass\_table

[,1] [,2]

1 216 0.2424242

2 184 0.2065095

3 491 0.5510662

#sex

> Titanic\_sex\_table<-table(Titanic$Sex)

> Titanic\_sex\_table

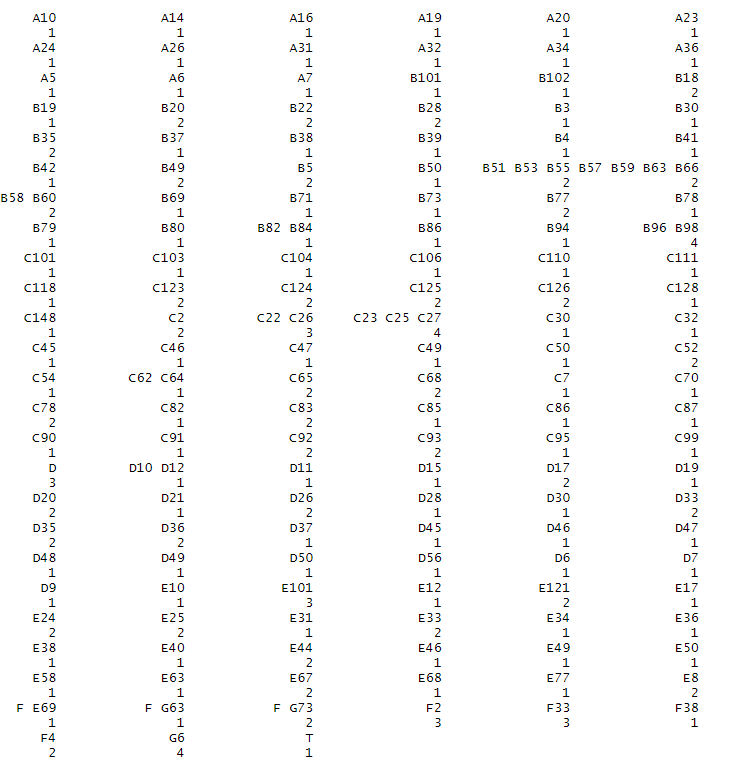
female male

314 577

#Cabin

> Titanic\_Cabin\_table<-table(Titanic$Cabin)

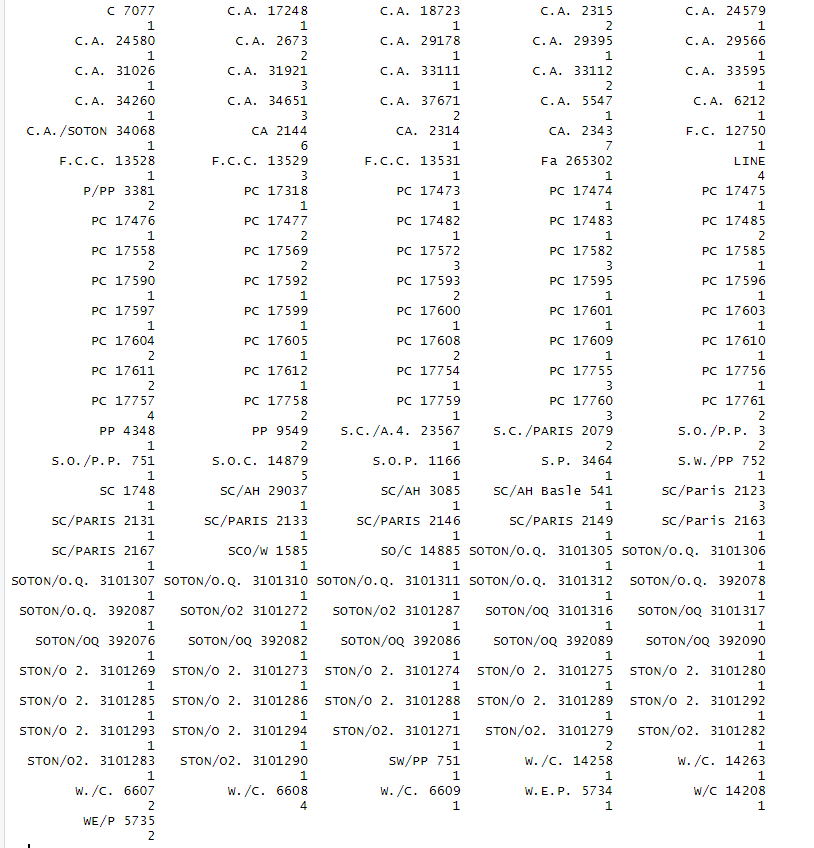
> Titanic\_Cabin\_table



#Ticket

> Titanic\_Cabin\_table<-table(Titanic$Cabin)

> Titanic\_Cabin\_table



-----------------------Quantitative Data------------------

-------------------Age--------------------------

> Titanic\_data\_age\_mean<-mean(Titanic$Age,na.rm = TRUE)

> Titanic\_data\_age\_mean

[1] 29.69912

> Titanic\_data\_age\_median<-median(Titanic$Age,na.rm = TRUE)

> Titanic\_data\_age\_median

[1] 28

> Titanic\_data\_age\_mode<-mode(Titanic$Age)

> Titanic\_data\_age\_mode

[1] "numeric"

>

> Titanic\_data\_age\_central\_tendency<-list(Titanic\_data\_age\_mean,Titanic\_data\_age\_median)

> Titanic\_data\_age\_central\_tendency

[[1]]

[1] 29.69912

[[2]]

[1] 28

> Titanic\_data\_age\_variance<-var(Titanic$Age)

> Titanic\_data\_age\_variance

[1] NA

> Titanic\_data\_age\_variance<-var(Titanic$Age,na.rm = TRUE)

> Titanic\_data\_age\_variance

[1] 211.0191

> Titanic\_data\_age\_standard\_deviation<-sqrt(Titanic\_data\_age\_variance)

> Titanic\_data\_age\_standard\_deviation

[1] 14.5265

--------------Sibsp-----------

> Titanic\_data\_Sibsp\_mean<-mean(Titanic$SibSp)

> Titanic\_data\_Sibsp\_mean

[1] 0.5230079

> Titanic\_data\_Sibsp\_median<-median(Titanic$SibSp)

> Titanic\_data\_Sibsp\_median

[1] 0

> Titanic\_data\_age\_central\_tendency

[[1]]

[1] 29.69912

[[2]]

[1] 28

> Titanic\_data\_Sibsp\_variance<-var(Titanic$SibSp,na.rm = TRUE)

> Titanic\_data\_Sibsp\_variance

[1] 1.216043

> Titanic\_data\_Sibsp\_Standard\_deviation<-sqrt(Titanic\_data\_Sibsp\_variance)

> Titanic\_data\_Sibsp\_Standard\_deviation

[1] 1.102743

------------------Parch----------------------

> Titanic\_data\_Parch\_mean<-mean(Titanic$Parch)

> Titanic\_data\_Parch\_mean

[1] 0.3815937

> Titanic\_data\_Parch\_median<-median(Titanic$Parch)

> Titanic\_data\_Parch\_median

[1] 0

> Titanic\_data\_Parch\_central\_tendency<-list(Titanic\_data\_Parch\_mean,Titanic\_data\_Parch\_median)

> Titanic\_data\_Parch\_central\_tendency

[[1]]

[1] 0.3815937

[[2]]

[1] 0

> Titanic\_data\_Parch\_variance<-var(Titanic$Parch,na.rm = TRUE)

> Titanic\_data\_Parch\_variance

[1] 0.6497282

> Titanic\_data\_Parch\_standard\_deviation<-sqrt(Titanic\_data\_Parch\_variance)

> Titanic\_data\_Parch\_standard\_deviation

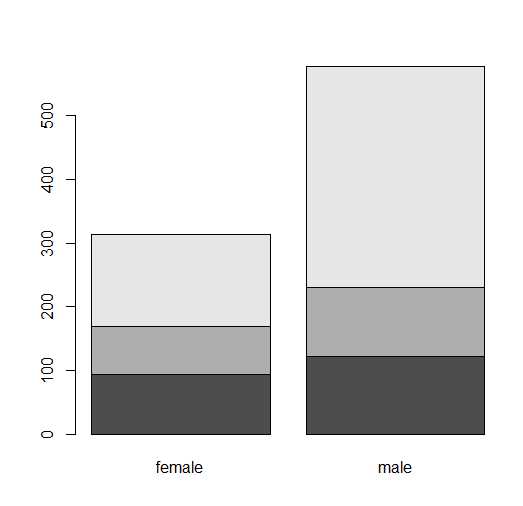
[1] 0.8060572

-----------------------Fare---------------

|  |
| --- |
| > Titanic\_data\_Fare\_mean<-mean(Titanic$Fare)  > Titanic\_data\_Fare\_mean  [1] 32.20421  > Titanic\_data\_Fare\_median<-median(Titanic$Fare)  > Titanic\_data\_Fare\_median  [1] 14.4542  > Titanic\_data\_Fare\_central\_tendency<-list(Titanic\_data\_Fare\_mean,Titanic\_data\_Fare\_median)  > Titanic\_data\_Fare\_central\_tendency  [[1]]  [1] 32.20421  [[2]]  [1] 14.4542 |
|  |
| |  | | --- | | > Titanic\_data\_Fare\_variance<-var(Titanic$Fare,na.rm = TRUE)  > Titanic\_data\_Fare\_variance  [1] 2469.437  > Titanic\_data\_Fare\_standard\_deviation<-sqrt(Titanic\_data\_Fare\_variance)  > Titanic\_data\_Fare\_standard\_deviation  [1] 49.69343  --------------------Univariate graphical method--------  \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Categorical\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  -----------------------Survived--------------------  > t<- table(Titanic$Survived,Titanic$Sex)  > barplot(t) | |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Pclass\*\*\*\*\*\*\*\*\*\*\*\*\*\*

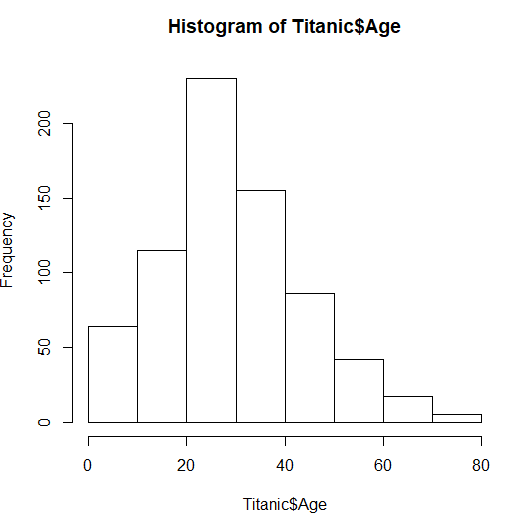
> barplot(table(Titanic$Pclass,Titanic$Sex))



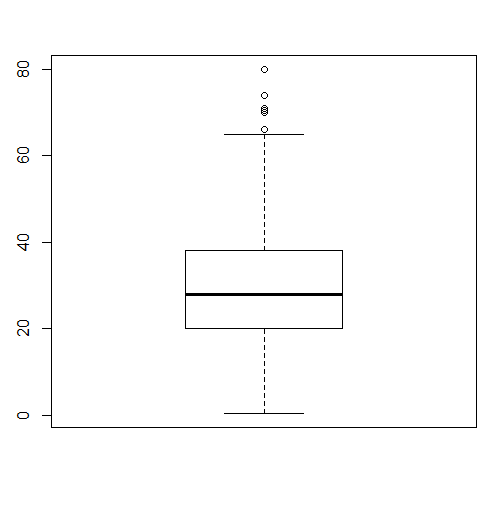
----------------------Quantitative-------------------------------------------------------

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Age\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

hist(Titanic$Age)

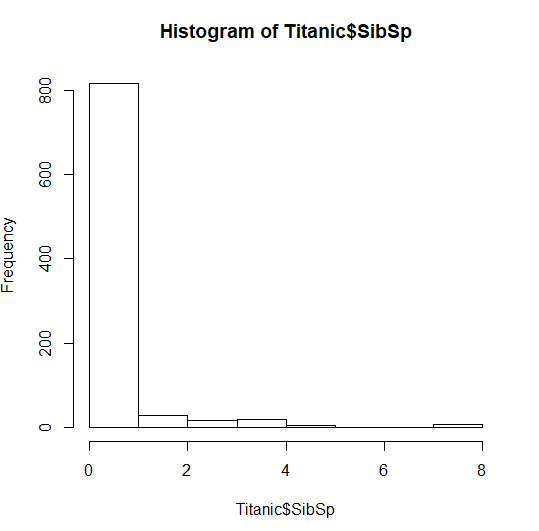


> boxplot(Titanic$Age)

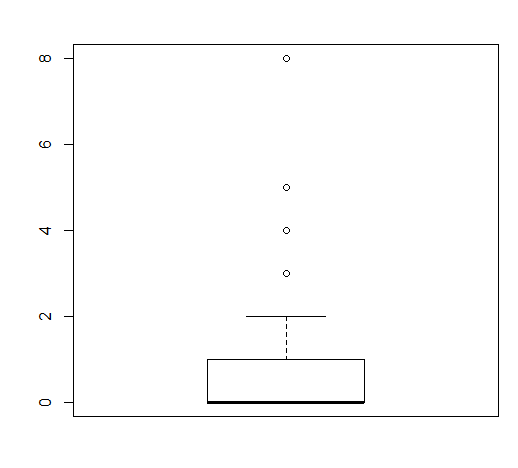


\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Sibsp\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

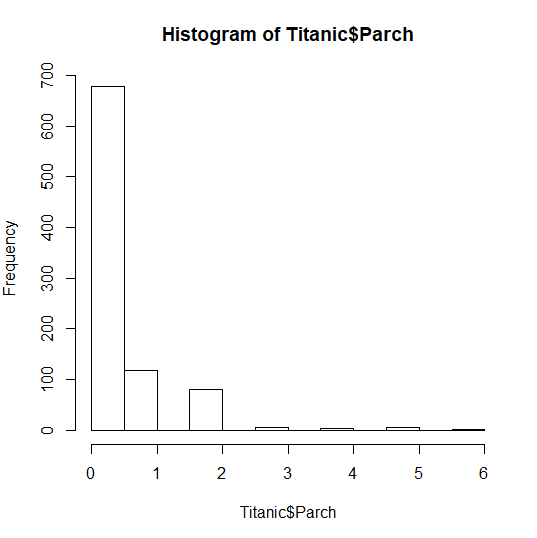
> hist(Titanic$SibSp)



boxplot(Titanic$SibSp)

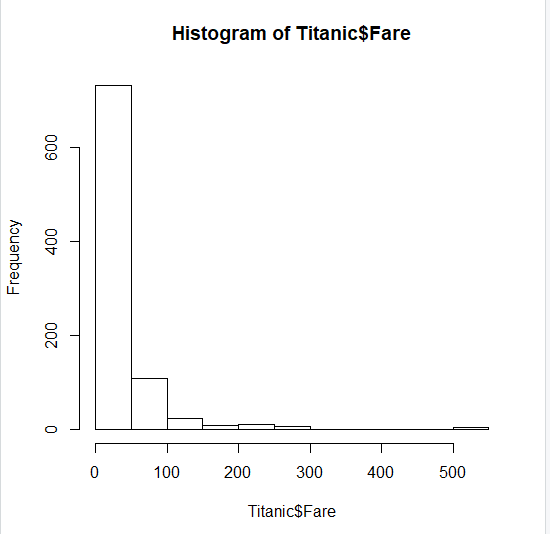


hist(Titanic$Parch)

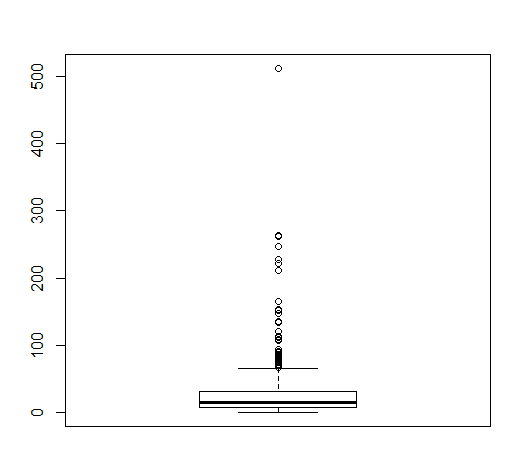


\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Fare\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

> hist(Titanic$Fare)



> boxplot(Titanic$Fare)



**Multivariate**

-----------------------Survived, Pclass-----------------------

> Titanic\_data<- cbind(Titanic$Survived,Titanic$Pclass)

> Titanic\_data<-summary(Titanic\_data)

> Titanic\_data

V1 V2

Min. :1.000 Min. :1.000

1st Qu.:1.000 1st Qu.:2.000

Median :1.000 Median :3.000

Mean :1.384 Mean :2.309

3rd Qu.:2.000 3rd Qu.:3.000

Max. :2.000 Max. :3.000

> Titanic\_chisq\_test<-table(Titanic$Survived,Titanic$Pclass)

> chisq.test(Titanic\_chisq\_test)

Pearson's Chi-squared test

data: Titanic\_chisq\_test

X-squared = 102.89, df = 2, p-value <

2.2e-16

> Titanic\_data=data.frame(Titanic$Survived,Titanic$Pclass)

> colnames(Titanic\_data)=c("Survived","Pclass")

> library(ggplot2)

Find out what's changed in ggplot2 at https://github.com/tidyverse/ggplot2/releases.

> ggplot(data=Titanic\_data,aes(x=Survived,y=Pclass,fill=Survived))+geom\_bar(stat="identity")+theme\_bw()

> ggplot(data=Titanic\_data,aes(x=Survived,y=Pclass,fill=Pclass))+geom\_bar(stat="identity")+theme\_bw()

