**Find how many people survived and died in Titanic disaster:**

table(train$Survived)

0 1

538 338 - We see that in the training set, 338 passengers survived, while 538 died

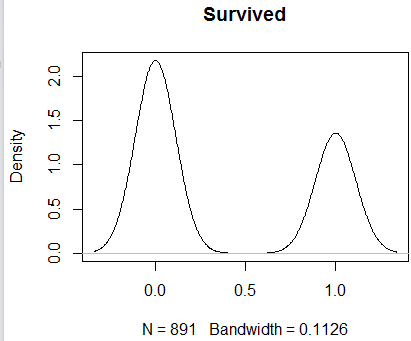
> prop.table(table(train$Survived))

0 1

0.6141553 0.3858447 38% of passengers survived the disaster in the training set and 61% died.

**Histogram and density**

plot(density(train$Survived), main="Survived")



**Finding Correlation Coefficients Pearson:**

> cor(train$Survived, train$SexN, method = "pearson")

[1] -0.5436527

> cor(train$Survived, train$Pclass, method = "pearson")

[1] -0.338481

> cor(train$Survived, train$Age, method = "pearson")

[1] -0.051832

> cor(train$Survived, train$SibSp, method = "pearson")

[1] -0.0353225

Pearson **Correlation Coefficient Calculator**. The Pearson **correlation coefficient** is used to measure the strength of a linear association between two variables, where the value r = 1 means a perfect positive **correlation** and the value r = -1 means a perfect negative **correlation**.

for example, 0.92 or -0.97 would show, respectively, a very strong positive and negative correlation. As with all statistis that demonstrate correlation, this does not prove causation.

> cor(train[,1:7])

Survived Pclass SexN Age SibSp Parch

Survived 1.00000000 -0.33959068 -0.54365274 -0.05183200 -0.03763407 0.08525835

Pclass -0.33959068 1.00000000 0.13320941 -0.38969350 0.09051013 0.02071297

SexN -0.54365274 0.13320941 1.00000000 0.08914444 -0.11547931 -0.24108181

Age -0.05183200 -0.38969350 0.08914444 1.00000000 -0.30108803 -0.20504350

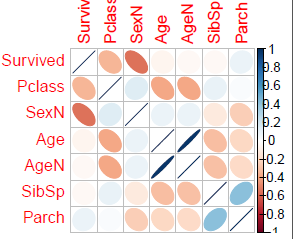
AgeN -0.03919002 -0.38446242 0.08438302 0.97848311 -0.29422982 -0.19641856

SibSp -0.03763407 0.09051013 -0.11547931 -0.30108803 1.00000000 0.41705931

Parch 0.08525835 0.02071297 -0.24108181 -0.20504350 0.41705931 1.00000000

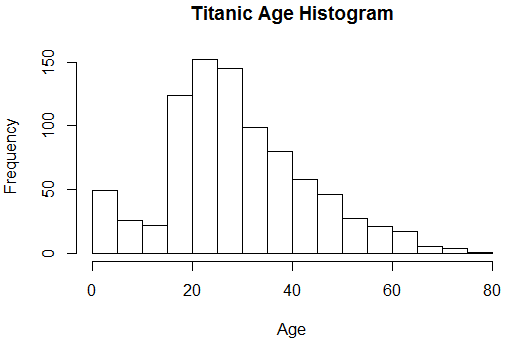
**Visualizing Correlation**

corrplot(cor(train[, c(1,2,3,4,5,6,7)]), method="ellipse")]



In this corrplot negative correlation are showed in brown color and positive correlations are showed in blue color. Darker color and xize of the circles are showing proportional correlation coefficient.

hist(train$Age, xlab= "Age",main="Histogram for Age", breaks=20)

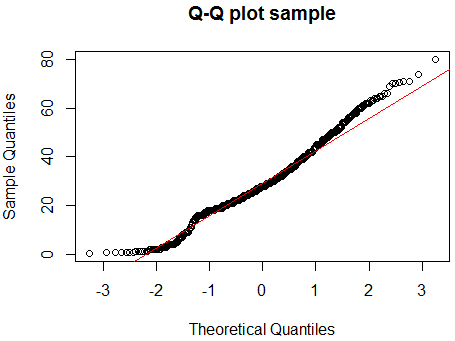


From this histogram we can see that Titanic has more passaged whose ages are between 20 - 30.

Careate a Quantile-Quantile plat to evaluate the fit of my dataset to the normal distribution.

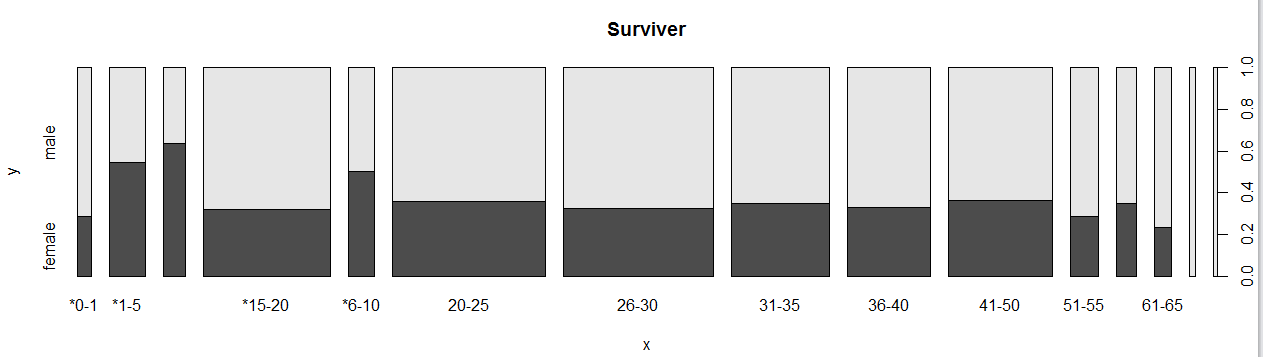
> qqnorm(train$Age, main="Q-Q plot sample")

> qqline(train$Age, col="red")



Based on my QQ Plot the quantiles of Age data represented by a straight line and considered as standard normal distribution.

plot(train$AgeGr,train$Sex, main="Titanic Age Group By Sex", type="l", type="2")



**Titanic Age Groups by Sex**

> prop.table(table(train$AgeGr, train$Sex))

female male

\*0-1 0.004566210 0.011415525

\*1-5 0.021689498 0.018264840

\*11-15 0.015981735 0.009132420

\*15-20 0.045662100 0.097031963

\*6-10 0.014840183 0.014840183

20-25 0.061643836 0.110730594

26-30 0.054794521 0.113013699

31-35 0.038812785 0.071917808

36-40 0.030821918 0.062785388

41-50 0.042237443 0.074200913

51-55 0.009132420 0.022831050

56-60 0.007990868 0.014840183

61-65 0.004566210 0.014840183

66-70 0.000000000 0.006849315

71-80 0.000000000 0.004566210

summary(train$Sex)

female male

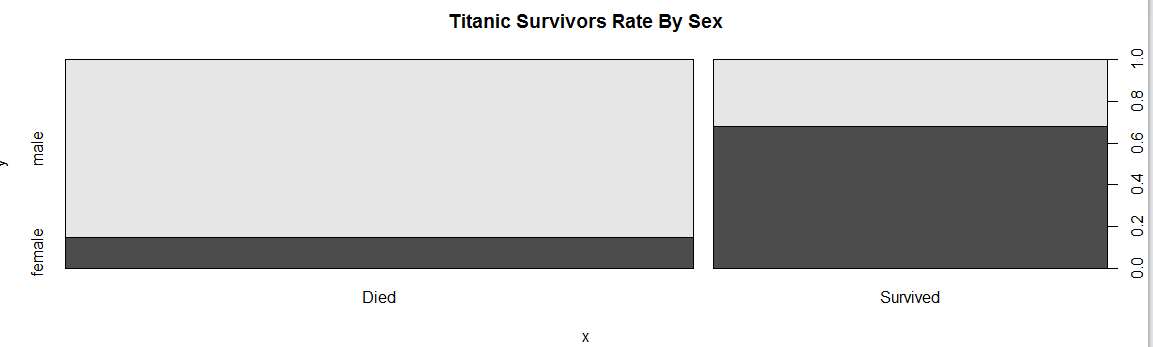
309 567 – Take a look at summary of Sex in the dataset

> prop.table(table(train$Sex))

female male

0.3527397 0.6472603 64% of Titanic passengers are man

> plot(train$Survived,train$Sex, main="Titanic Survivors Rate By Sex", type="l", type="2")



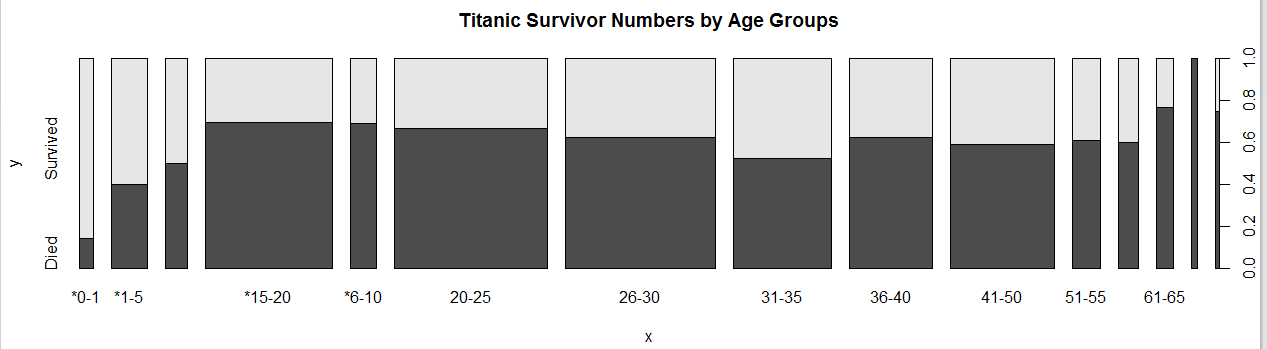
> prop.table(table(train$Sex, train$Survived),1)

0 1

female 0.2556634 0.7443366

male 0.8095238 0.1904762 # Look at Survivor gender patterns

> plot(train$AgeGr,train$SurvivedW, main="Titanic Survivor Numbers by Age Group", type="l", type="2")



> prop.table(table(train$AgeGr, train$Survived))

0 1

\*0-1 0.002283105 0.013698630

\*1-5 0.015981735 0.023972603

\*11-15 0.012557078 0.012557078

\*15-20 0.099315068 0.043378995

\*6-10 0.020547945 0.009132420

20-25 0.115296804 0.057077626

26-30 0.105022831 0.062785388

31-35 0.058219178 0.052511416

36-40 0.058219178 0.035388128

41-50 0.068493151 0.047945205

51-55 0.019406393 0.012557078

56-60 0.013698630 0.009132420

61-65 0.014840183 0.004566210

66-70 0.006849315 0.000000000

71-80 0.003424658 0.001141553

summary(train$Age)

Min. 1st Qu. Median Mean 3rd Qu. Max.

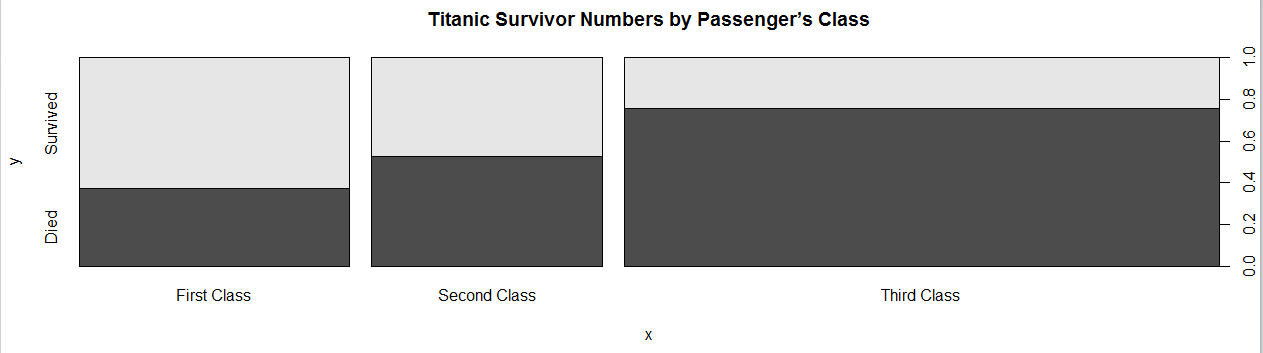
0.42 20.00 28.00 29.51 38.00 80.00 My Median Survivor Age in the dataset is **28** and **Min-0.42** and **Max-80**

> summary(train$AgeGr)

\*0-1 \*1-5 \*11-15 \*15-20 \*6-10 20-25 26-30 31-35 36-40 41-50 51-55 56-60 61-65 66-70 71-80

14 35 22 125 26 151 147 97 82 102 28 20 17 6 4

plot(train$PclassW,train$SurvivedW, main="Titanic Survivor Numbers by Passenger’s Class", type="l", type="2")



> prop.table(table(train$Pclass, train$Survived))

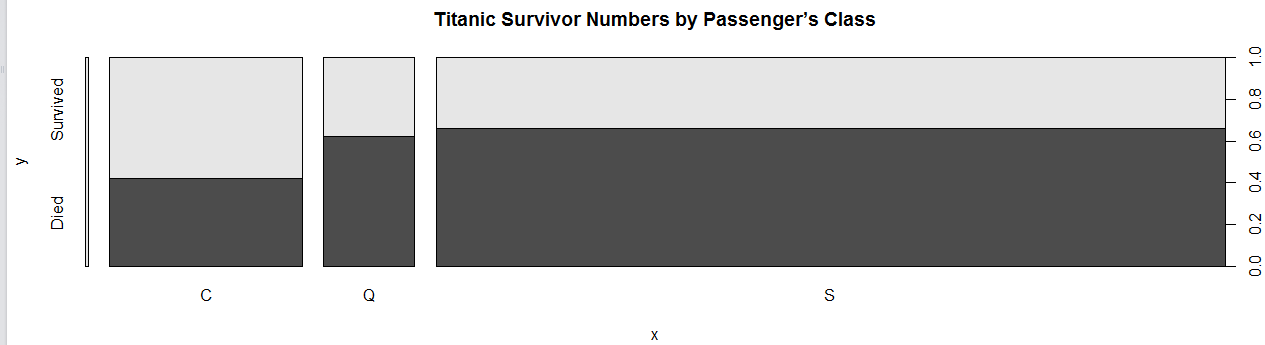
0 1

1 0.09132420 0.15525114

2 0.11073059 0.09931507

3 0.41210046 0.13127854

> plot(train$Embarked,train$SurvivedW, main="Titanic Survivor Numbers by Embarked Area ", type="l", type="2")



**Titanic Survivor Numbers by Embarked Area**

> prop.table(table(train$Embarked, train$Survived))

0 1

0.000000000 0.002283105

C 0.075342466 0.103881279

Q 0.052511416 0.031963470

S 0.486301370 0.247716895

> summary(train$Embarked)

C Q S

2 157 74 643

Southampton passengers have a bigger death rate compare to passengers from Cherbourg and Queenstown.

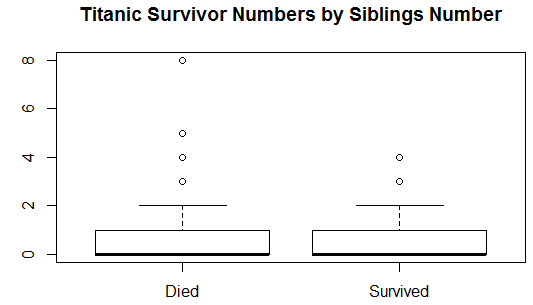
> prop.table(table(train$Embarked))

C Q S

0.002283105 0.179223744 0.084474886 0.734018265 73% of passengers from Southampton died

**Created a Boxplot to see frequencies on Titanic survivor numbers by Sibling number.**

plot(train$SurvivedW,train$SibSp, main="Titanic Survivor Numbers by Siblings Number", type="l", type="2")



> prop.table(table(train$Survived, train$SibSp))

0 1 2 3 4 5

0 0.441780822 0.110730594 0.017123288 0.013698630 0.017123288 0.005707763

1 0.235159817 0.127853881 0.014840183 0.004566210 0.003424658 0.000000000

8

0 0.007990868

1 0.000000000

> summary(train$SibSp)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 0.000 0.000 0.532 1.000 8.000