```r
#getwd()
#setwd("D:/Desktop/final_analysis")
#install.packages("ggpubr")
#install.packages("moments")
#install.packages("faraway")
library(readxl)
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.0.5
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```r
library(moments)
library(MASS)
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 4.0.5
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car
```

```r
M <- read_excel("MANOVA Data.xlsx")
```

```
## New names:
## * `` -> ...11
```

```r
M=M[,c(4,6,7,8,9,10,12,13,14,15)]
str(M)# structure of data we have
```

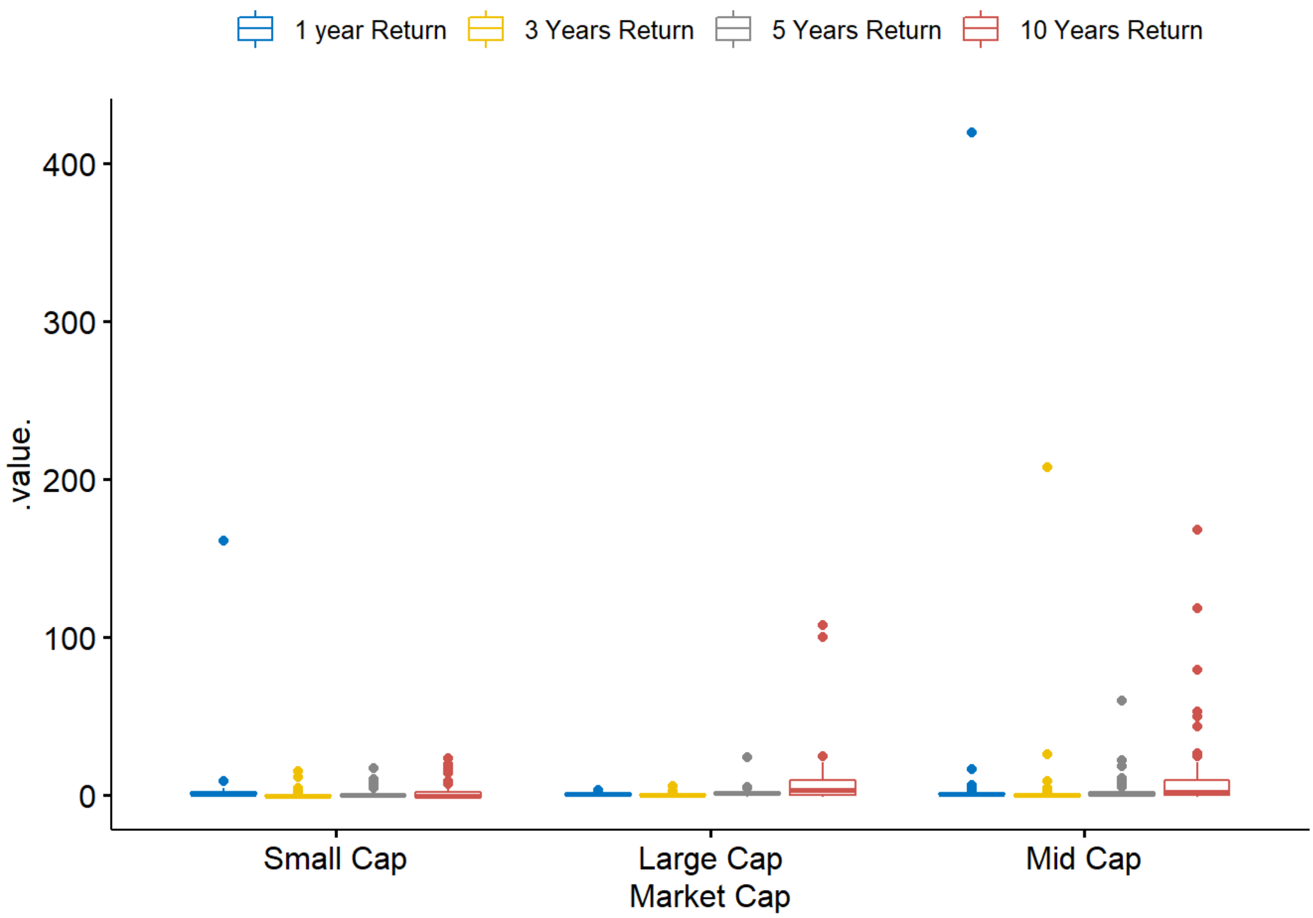
```
## tibble [180 x 10] (S3: tbl_df/tbl/data.frame)
##  $ Market Cap                      : chr [1:180] "Small Cap" "Large Cap" "Small C
ap" "Large Cap" ...
##  $ Price to Book Ratio (High/Low): chr [1:180] "High" "High" "High" "High" ...
##  $ 1 year Return                   : num [1:180] 4.456 0.117 0.497 0.858 1.115
...
##  $ 3 Years Return                  : num [1:180] 11.258 1.218 -0.275 0.407 1.566
...
##  $ 5 Years Return                  : num [1:180] 17.2 2.39 -0.15 1.86 3.87 ...
##  $ 10 Years Return                 : num [1:180] 14.17597 4.27072 0.00304 19.9766
1 10.12747 ...
##  $ 1 Year Risk (SD)                : num [1:180] 0.0318 0.0175 0.0283 0.0227 0.02
4 ...
##  $ 3 Years Risk (SD)               : num [1:180] 0.0376 0.0172 0.0241 0.0234 0.02
```

```
45 ...
## $ 5 Years Risk (SD)              : num [1:180] 0.036 0.0157 0.0219 0.022 0.024
...
## $ 10 Years Risk (SD)             : num [1:180] 0.037 0.0152 0.021 0.0212 0.0247
...
```

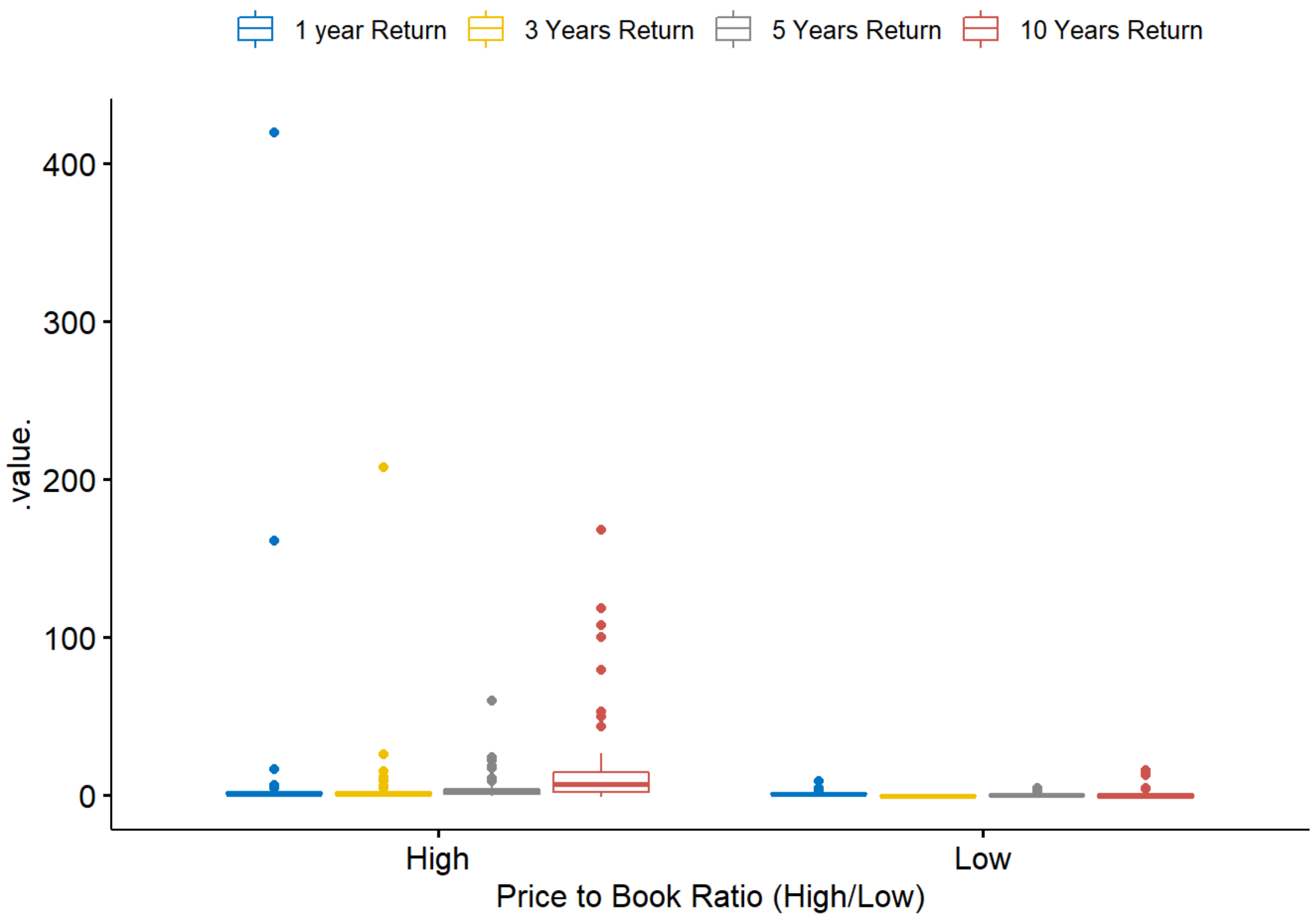
```
summary(M)
```

```
##   Market Cap         Price to Book Ratio (High/Low) 1 year Return
##  Length:180         Length:180                     Min.   : -0.3454
##  Class :character   Class :character               1st Qu.:  0.5167
##  Mode  :character   Mode  :character               Median :  0.9215
##                                                    Mean   :  4.5866
##                                                    3rd Qu.:  1.6837
##                                                    Max.   :419.7248
##  3 Years Return     5 Years Return    10 Years Return    1 Year Risk (SD)
##  Min.   : -0.9557   Min.   :-0.9397   Min.   : -0.9940   Min.   :0.003612
##  1st Qu.: -0.3514   1st Qu.:-0.1565   1st Qu.: -0.5595   1st Qu.:0.023714
##  Median :  0.1078   Median : 0.7093   Median :  1.4177   Median :0.029172
##  Mean   :  1.7926   Mean   : 1.9558   Mean   :  7.8997   Mean   :0.030477
##  3rd Qu.:  0.6192   3rd Qu.: 1.8909   3rd Qu.:  7.1673   3rd Qu.:0.035476
##  Max.   :207.4500   Max.   :59.7404   Max.   :167.8394   Max.   :0.221383
##  3 Years Risk (SD) 5 Years Risk (SD) 10 Years Risk (SD)
##  Min.   :0.01209   Min.   :0.01152   Min.   :0.01379
##  1st Qu.:0.02309   1st Qu.:0.02119   1st Qu.:0.02098
##  Median :0.02865   Median :0.02676   Median :0.02615
##  Mean   :0.03168   Mean   :0.02938   Mean   :0.02860
##  3rd Qu.:0.03468   3rd Qu.:0.03293   3rd Qu.:0.03283
##  Max.   :0.33804   Max.   :0.31068   Max.   :0.25595
```

```
# from summary we can see the mean and median of Returns column are not same a ide
ntifier that these are not normal
# max values of Returns are very far from the mean which indicate that there are o
utliers as well for Returns
#will check quantitatively the normality and the outliers


#Box plot for outliers if any:
# wrt to Market Caps factor 1
ggboxplot(
  M, x = c("Market Cap"), y = c("1 year Return", "3 Years Return","5 Years Return"
,"10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
# from box plot you can see alot of outliers which are extreme for all the Market
 caps, we have to remove it

#wrt to Price to Book Ratio (High/Low)
ggboxplot(
  M, x = c("Price to Book Ratio (High/Low)"), y = c("1 year Return", "3 Years Retu
rn","5 Years Return","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
# from the box plot you can see most of the outliers are for high PB ratio (just a
observation) we have to remove all this

# Removing outliers using mahalanobis function:
cutoff=qchisq(1-0.05,4)
cutoff
```

```
## [1] 9.487729
```

```
mahal=mahalanobis(M[,-c(1,2)],colMeans(M[,-c(1,2)]),cov(M[,-c(1,2)]))
summary(mahal<cutoff)
```

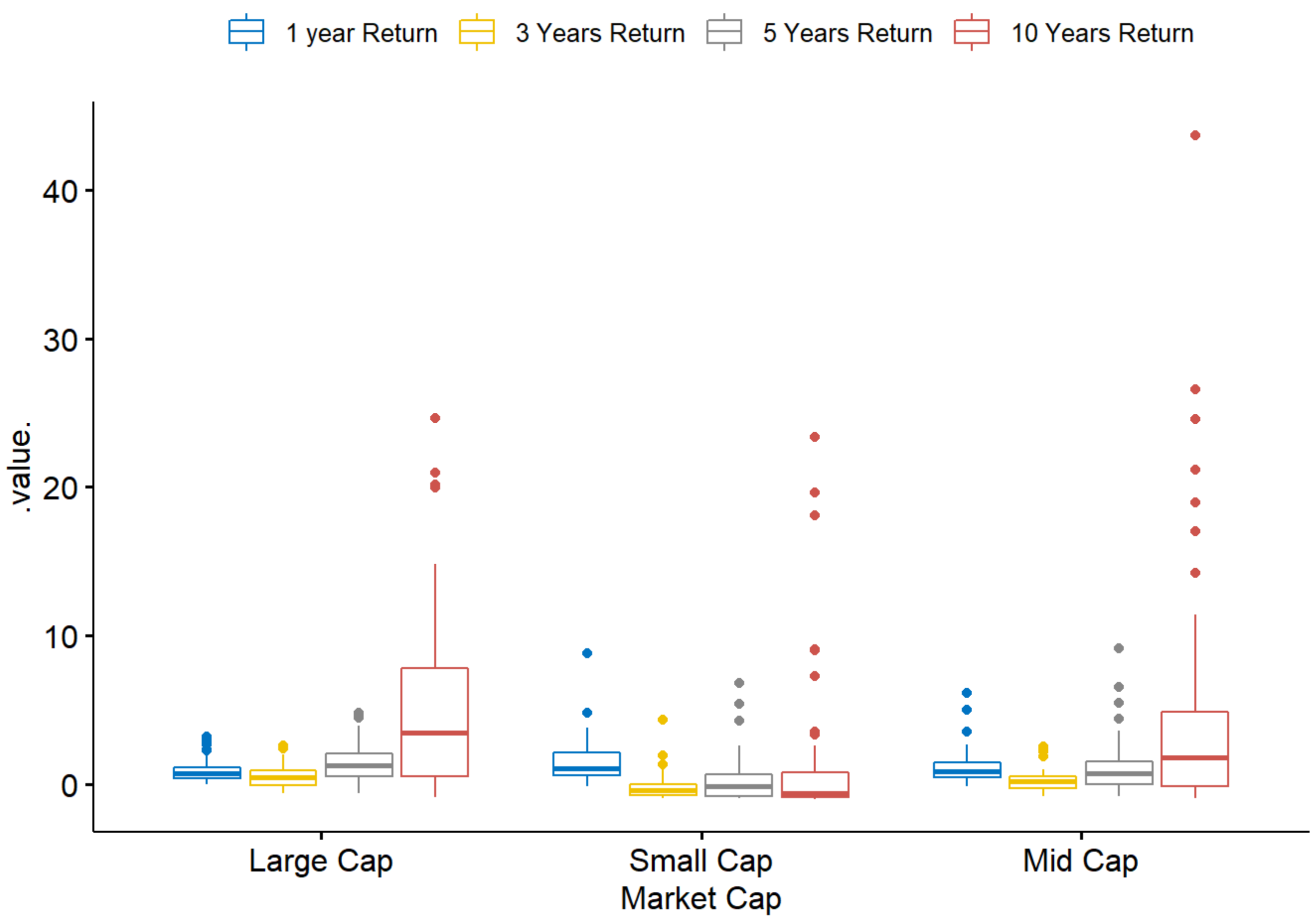
```
##     Mode   FALSE    TRUE
## logical      21     159
```

```
# 21 outliers
noout=subset(M,mahal<cutoff)
Mo=noout
Mo
```

```
## # A tibble: 159 x 10
##    `Market Cap` `Price to Book ~ `1 year Return` `3 Years Return`
##    <chr>        <chr>                      <dbl>            <dbl>
##  1 Large Cap    High                       0.117            1.22
##  2 Small Cap    High                       0.497           -0.275
```

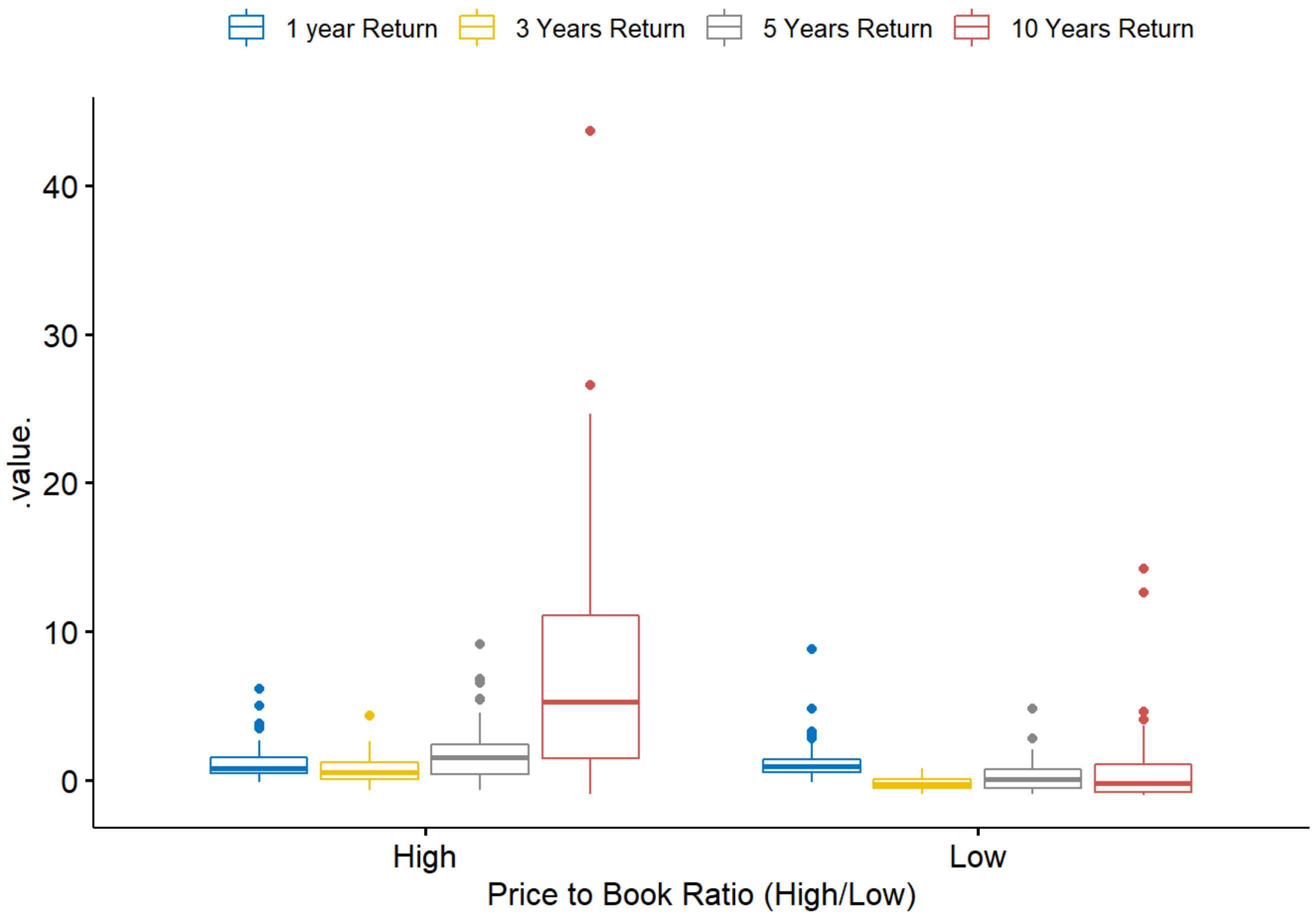
```
##   3 Large Cap    High                        0.858            0.407
##   4 Large Cap    High                        1.12             1.57
##   5 Mid Cap      High                        1.56             0.509
##   6 Large Cap    High                        0.510            0.530
##   7 Large Cap    High                        0.249            0.345
##   8 Large Cap    High                        0.323            0.579
##   9 Large Cap    High                        0.589            2.02
##  10 Small Cap    High                        1.73             0.560
## # ... with 149 more rows, and 6 more variables: `5 Years Return` <dbl>, `10
## #   Years Return` <dbl>, `1 Year Risk (SD)` <dbl>, `3 Years Risk (SD)` <dbl>,
## #   `5 Years Risk (SD)` <dbl>, `10 Years Risk (SD)` <dbl>
```

```
#analysis for Returns:
Mrto=Mo[,c(1,2,3,4,5,6)]#return table seperate
Mrso=Mo[,c(1,2,7,8,9,10)]# risk table separate


#analysis for Returns:
# box plot to verify if the outliers are removed

ggboxplot(
  Mrto, x = c("Market Cap"), y = c("1 year Return", "3 Years Return","5 Years Retu
rn","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
ggboxplot(
  Mrto, x = c("Price to Book Ratio (High/Low)"), y = c("1 year Return", "3 Years R
```

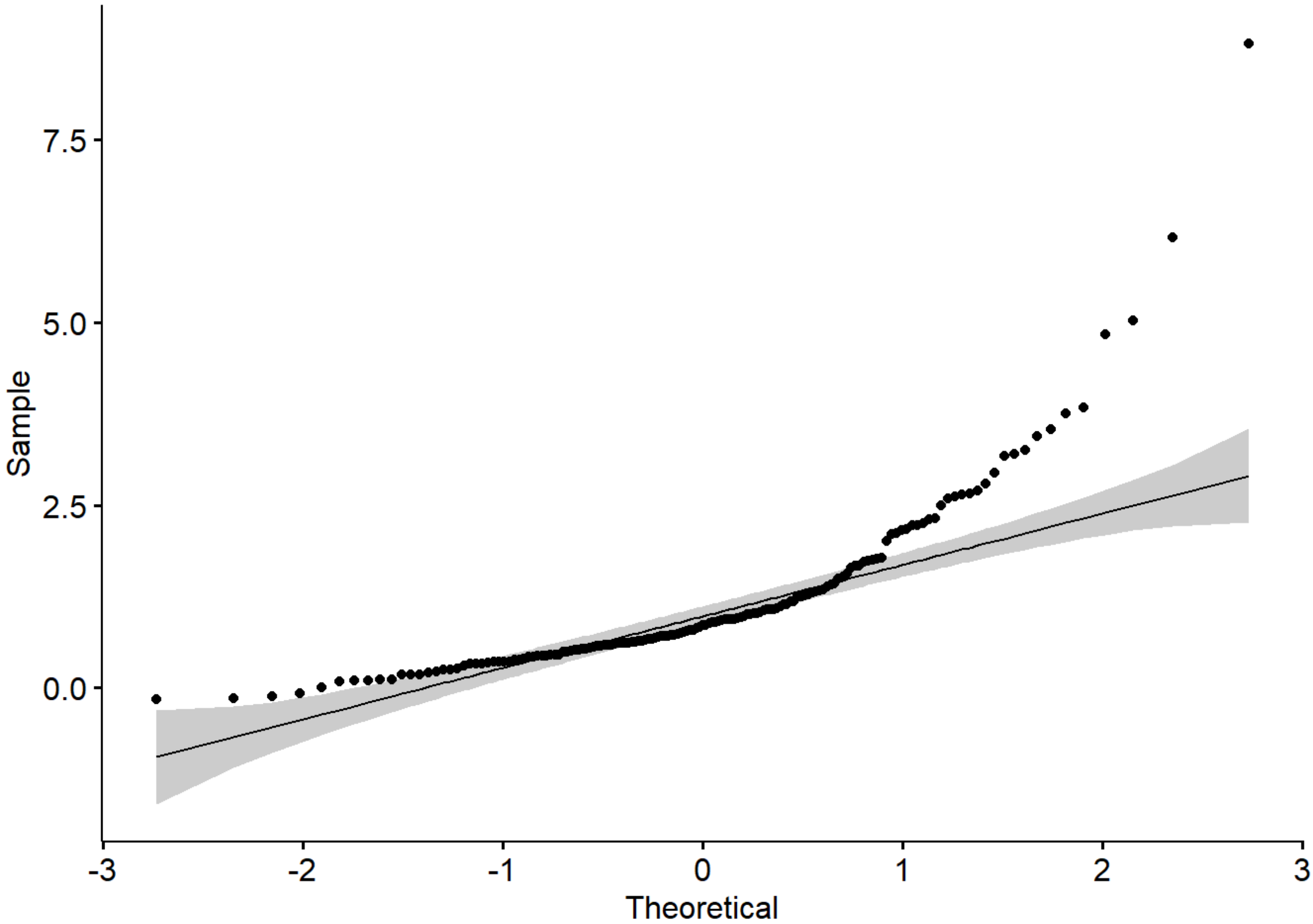
```
eturn","5 Years Return","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
#now the data looks quite good as compared to last box plot
## since the data is not yet normal we will normalize the data and plot the box pl
ot again
# Now doing the normality checkup
# Normality
R=Mrto[,-c(1,2)]
R=as.matrix(R)

shapiro.test(R[,1])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R[, 1]
## W = 0.75103, p-value = 3.974e-15
```

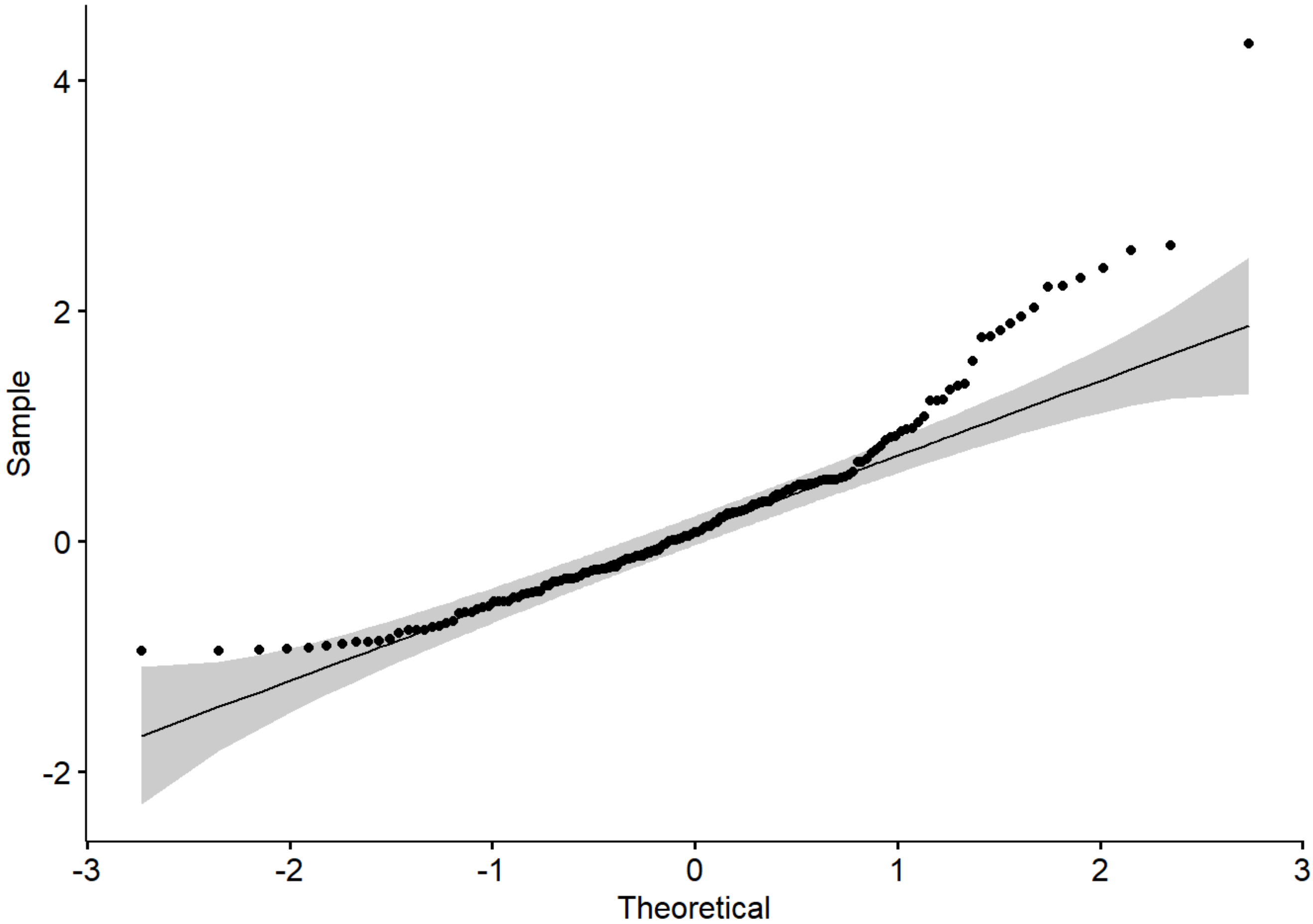
```
ggqqplot(R[,1])
```




```
shapiro.test(R[,2])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R[, 2]
## W = 0.9046, p-value = 1.155e-08
```

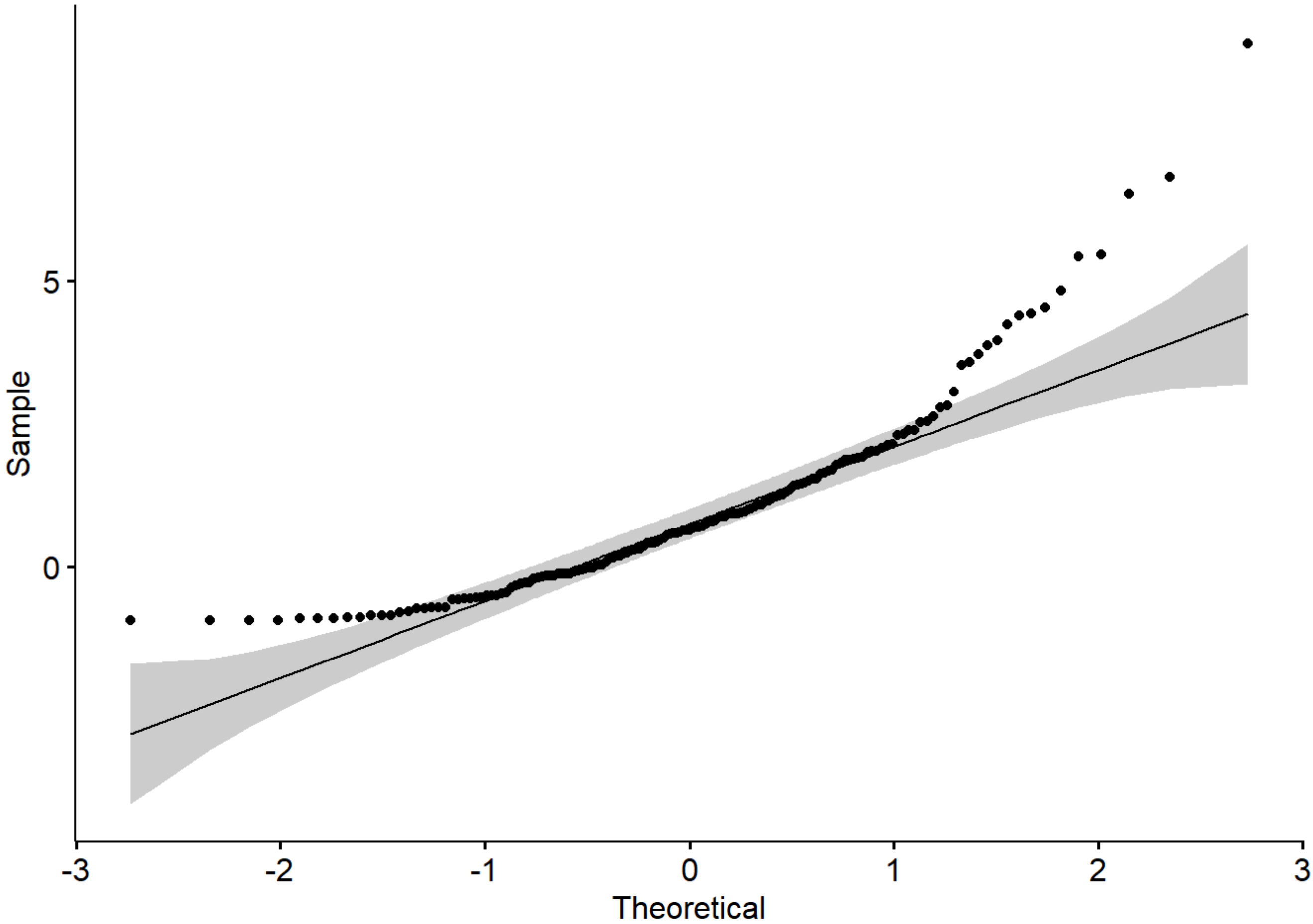
```
ggqqplot(R[,2])
```




```
shapiro.test(R[,3])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R[, 3]
## W = 0.8595, p-value = 4.987e-11
```

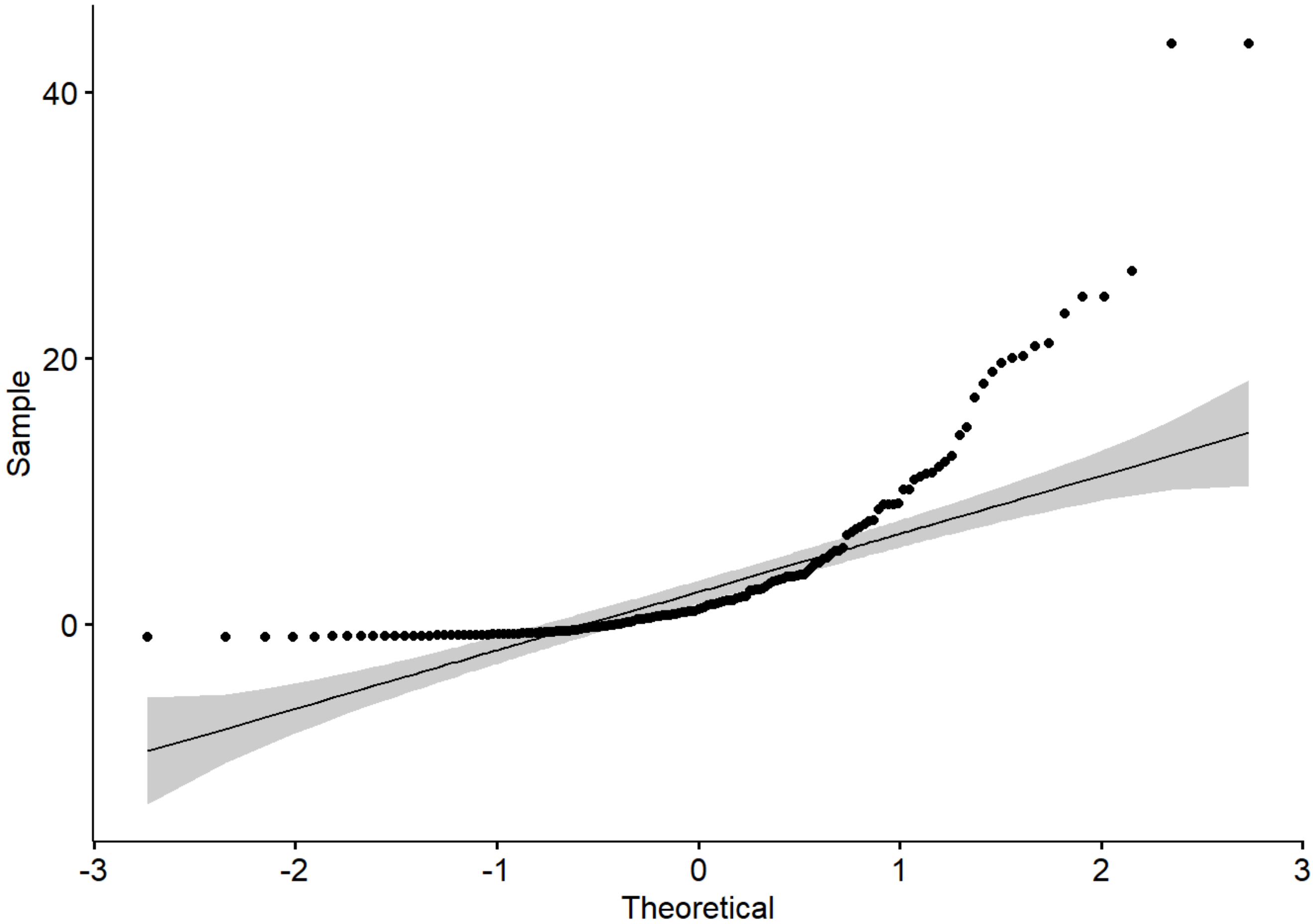
```
ggqqplot(R[,3])
```




```
shapiro.test(R[,4])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R[, 4]
## W = 0.67818, p-value < 2.2e-16
```

```
ggqqplot(R[,4])
```



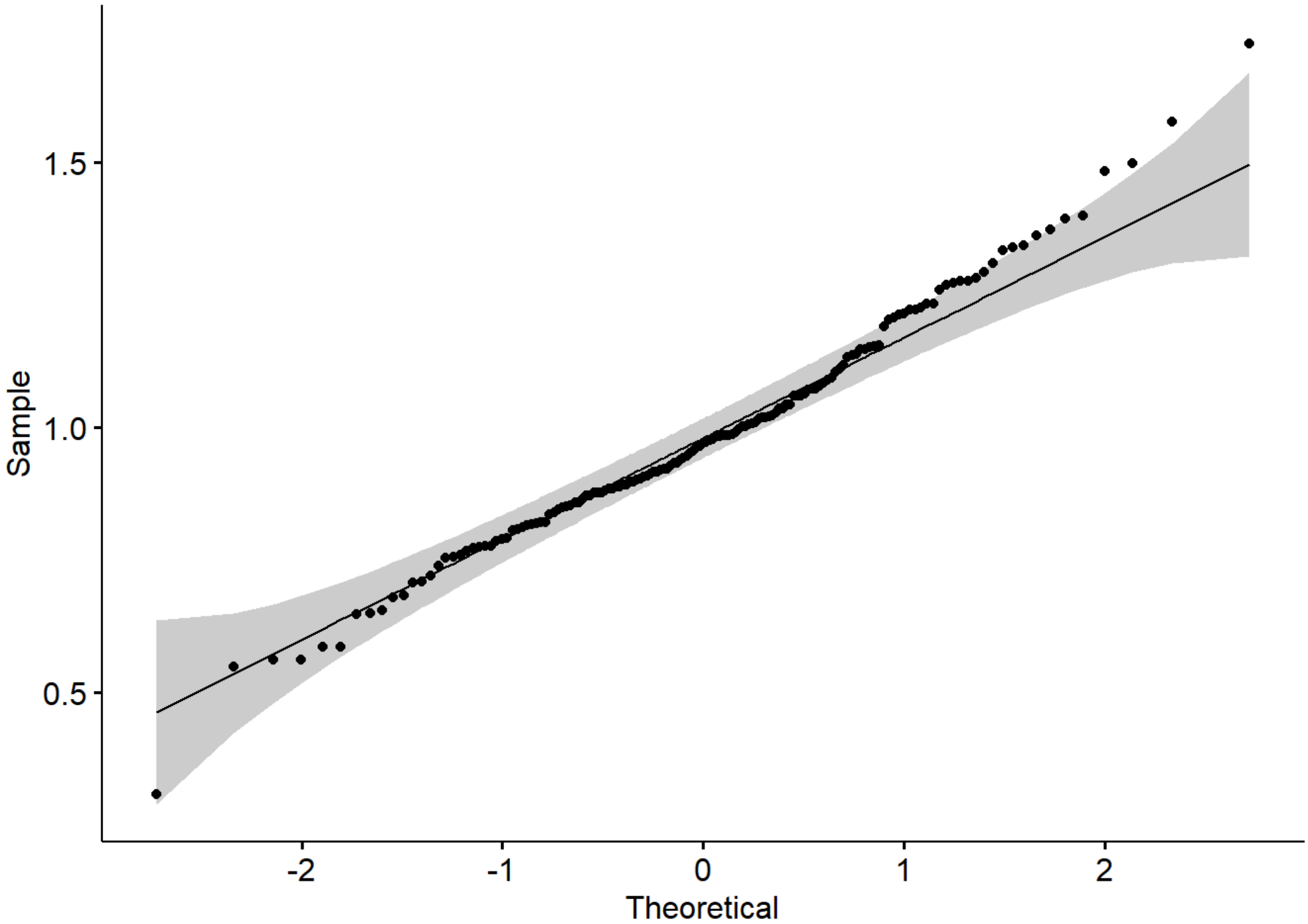

```
# you can see that the data is not normal
# we have to perform some transfromation to make it Normal which is one of the key assumptions for doing Manova

#power Transformation:
# and replacing the coulmn values with particular transformation value
R1=R[,1]
R1n=(R1)^(1/4)
shapiro.test(R1n)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R1n
## W = 0.9852, p-value = 0.09716
```

```
ggqqplot(R1n)
```

```
## Warning: Removed 4 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 4 rows containing non-finite values (stat_qq_line).

## Warning: Removed 4 rows containing non-finite values (stat_qq_line).
```



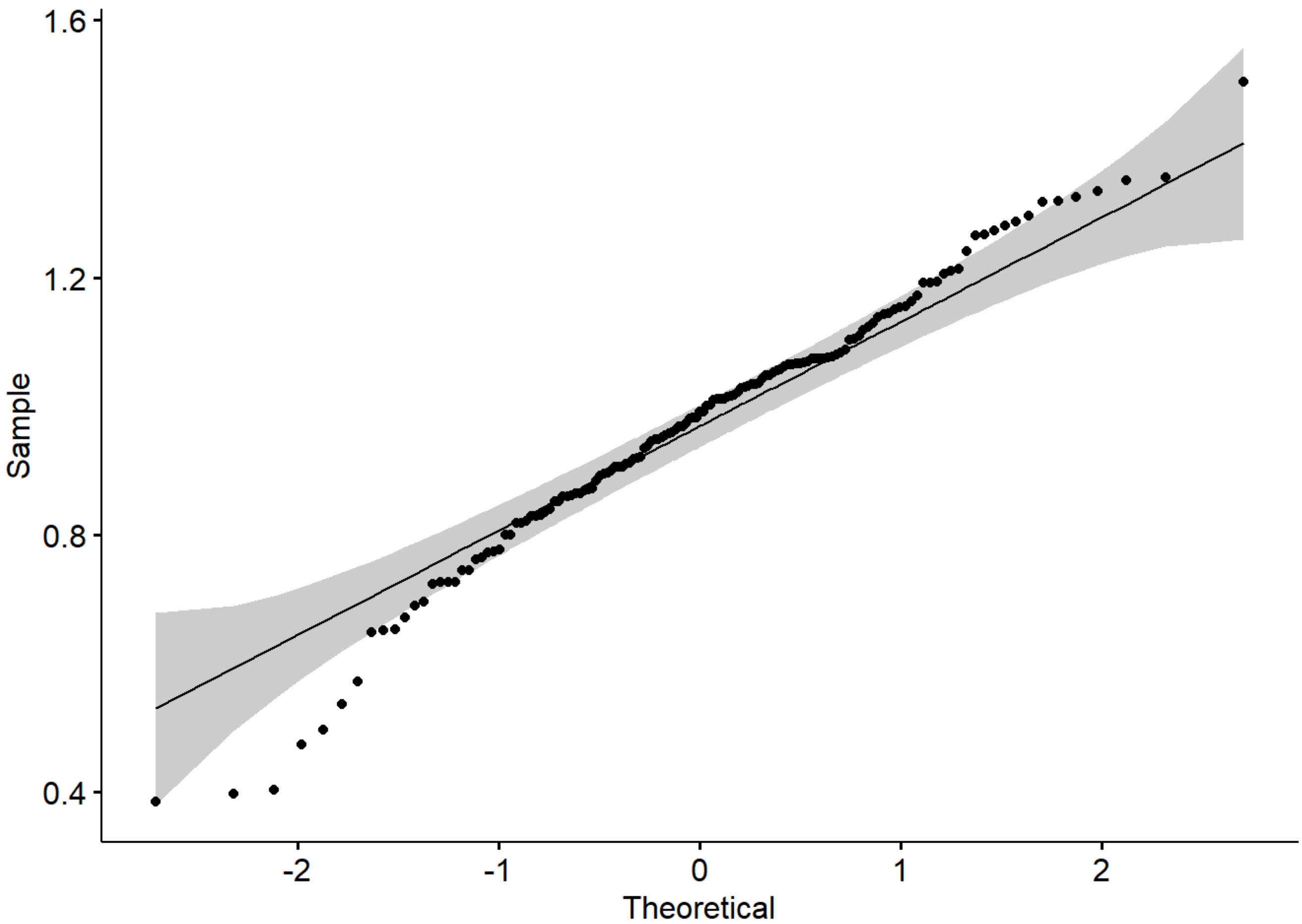

```
#replacing values
Mrto[,3]=(Mrto[,3])^(1/4)

#similarly for others
R2=R[,2]
R2n=(R2+0.8)^(1/4)
shapiro.test(R2n)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R2n
## W = 0.98189, p-value = 0.04956
```

```
ggqqplot(R2n)
```

```
## Warning: Removed 12 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 12 rows containing non-finite values (stat_qq_line).

## Warning: Removed 12 rows containing non-finite values (stat_qq_line).
```



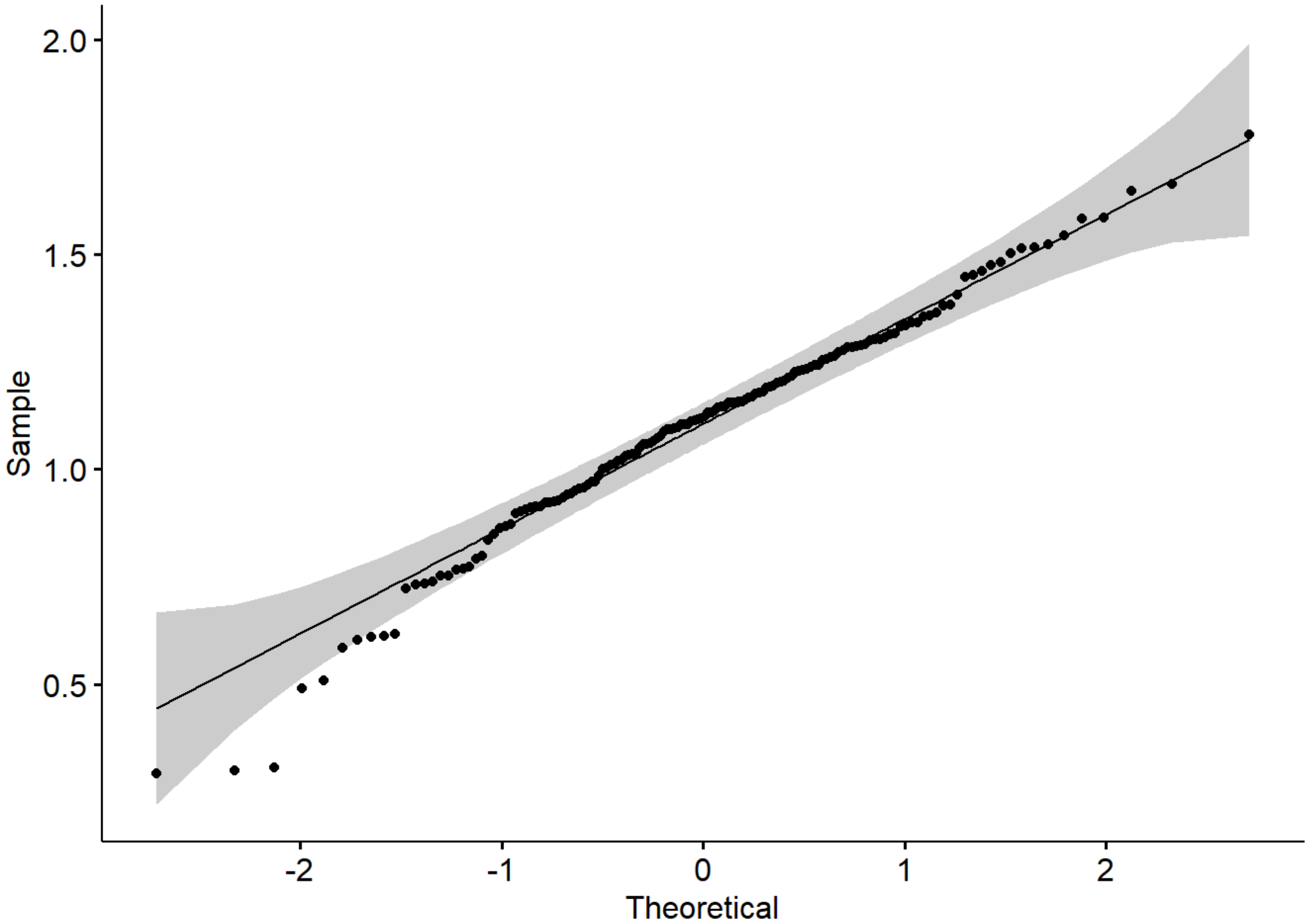

```
Mrto[,4]=(Mrto[,4]+0.8)^(1/4)

R3=R[,3]
R3n=(R3+0.85)^(1/4)
shapiro.test(R3n)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R3n
## W = 0.98054, p-value = 0.03194
```

```
ggqqplot(R3n)
```

```
## Warning: Removed 9 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 9 rows containing non-finite values (stat_qq_line).

## Warning: Removed 9 rows containing non-finite values (stat_qq_line).
```



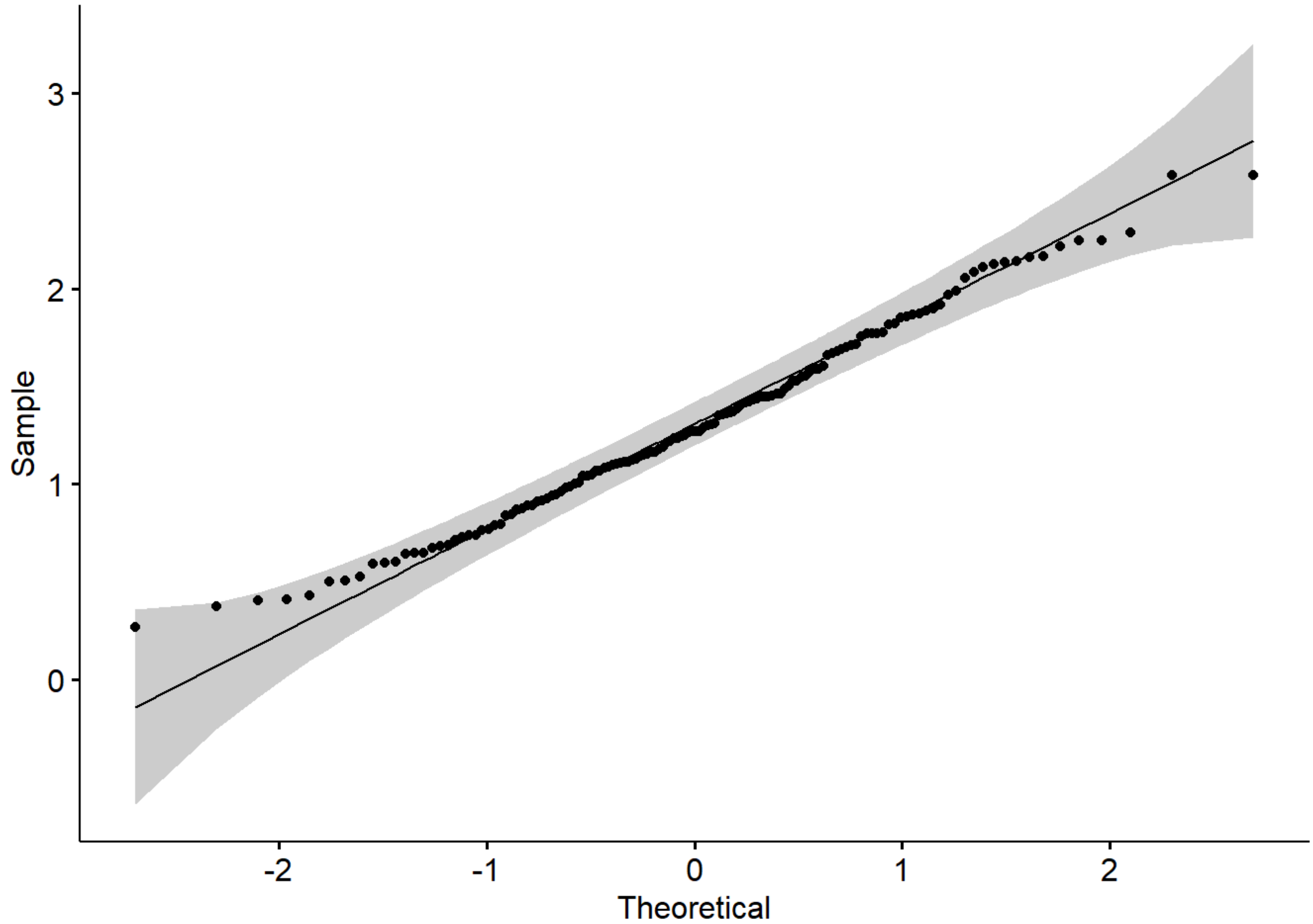

```
Mrto[,5]=(Mrto[,5]+0.85)^(1/4)

R4=R[,4]
R4n=(R4+0.85)^(1/4)
shapiro.test(R4n)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  R4n
## W = 0.98758, p-value = 0.2426
```

```
ggqqplot(R4n)
```

```
## Warning: Removed 19 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 19 rows containing non-finite values (stat_qq_line).

## Warning: Removed 19 rows containing non-finite values (stat_qq_line).
```




```
Mrto[,6]=(Mrto[,6]+0.85)^(1/4)
Mrto
```

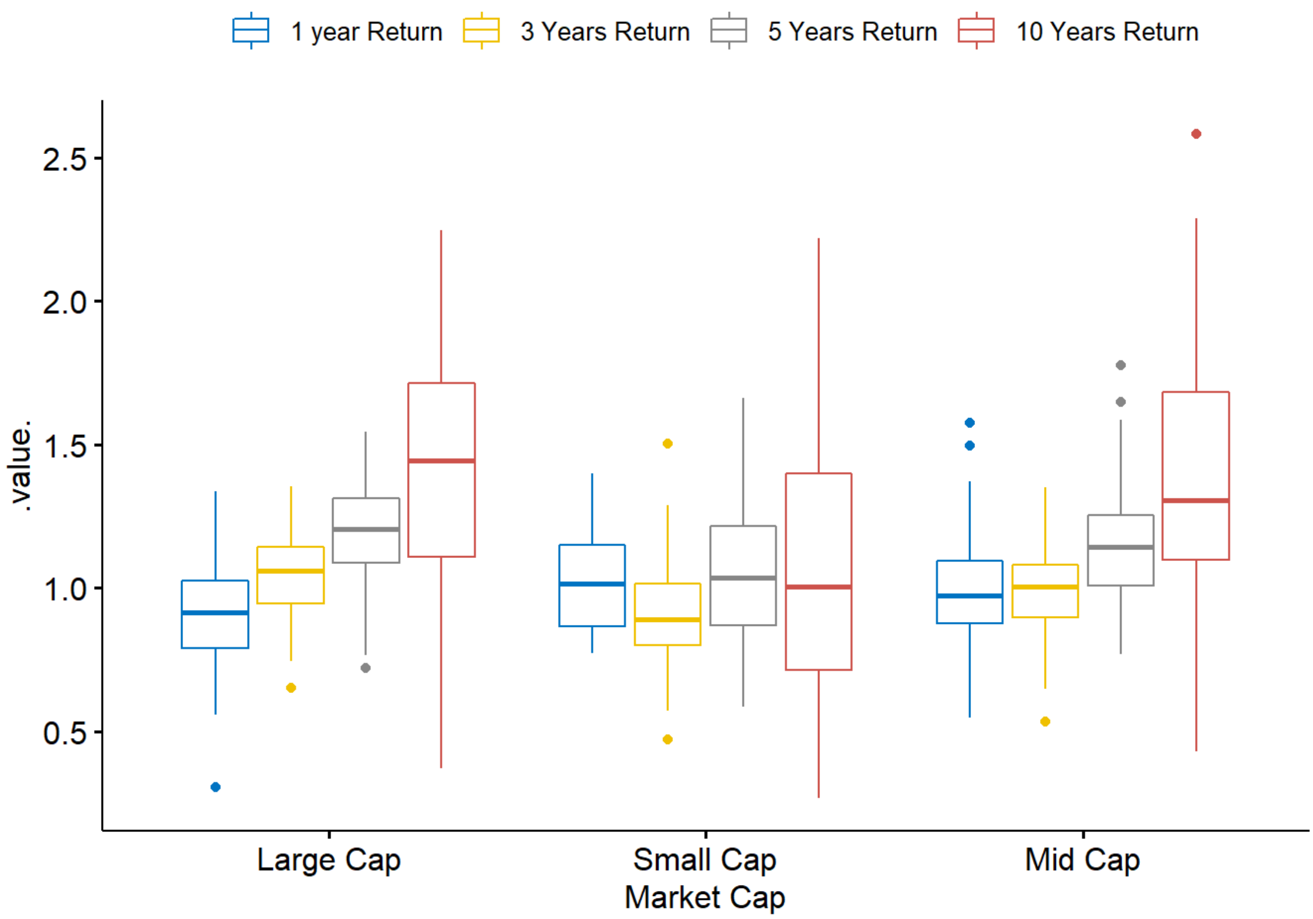
```
## # A tibble: 159 x 6
##    `Market Cap` `Price to Book ~ `1 year Return` `3 Years Return`
##    <chr>        <chr>                      <dbl>            <dbl>
##  1 Large Cap    High                       0.585            1.19
##  2 Small Cap    High                       0.840            0.851
##  3 Large Cap    High                       0.963            1.05
##  4 Large Cap    High                       1.03             1.24
##  5 Mid Cap      High                       1.12             1.07
##  6 Large Cap    High                       0.845            1.07
##  7 Large Cap    High                       0.706            1.03
##  8 Large Cap    High                       0.754            1.08
##  9 Large Cap    High                       0.876            1.30
## 10 Small Cap    High                       1.15             1.08
## # ... with 149 more rows, and 2 more variables: `5 Years Return` <dbl>, `10
## #   Years Return` <dbl>
```

```
Mrton=na.omit(Mrto)  # omit rows containing NA
Mrton
```

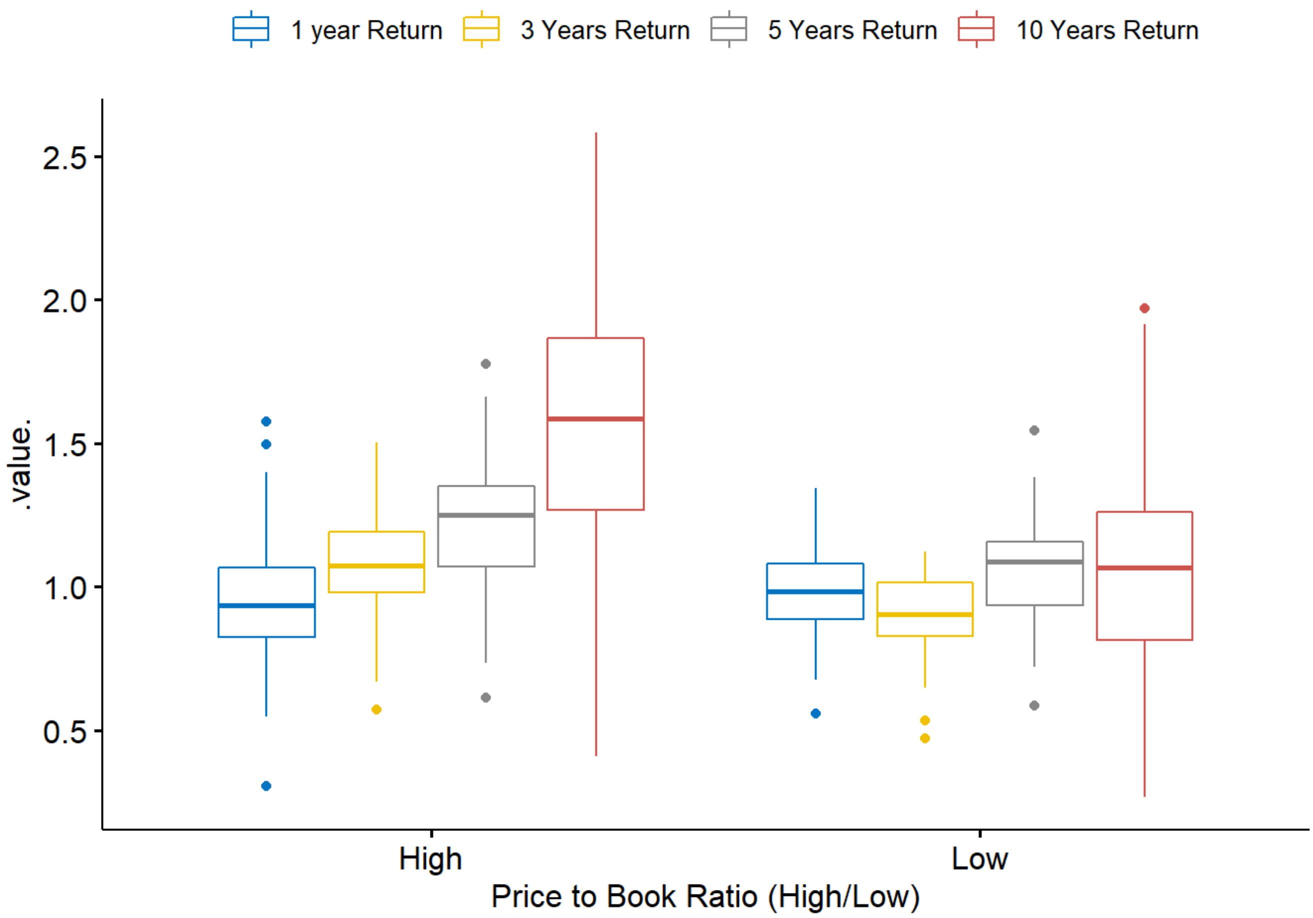
```
## # A tibble: 133 x 6
##    `Market Cap` `Price to Book ~ `1 year Return` `3 Years Return`
##    <chr>        <chr>                      <dbl>            <dbl>
##  1 Large Cap    High                       0.585            1.19
##  2 Small Cap    High                       0.840            0.851
```

```
## 3 Large Cap    High                        0.963            1.05
## 4 Large Cap    High                        1.03             1.24
## 5 Mid Cap      High                        1.12             1.07
## 6 Large Cap    High                        0.845            1.07
## 7 Large Cap    High                        0.706            1.03
## 8 Large Cap    High                        0.754            1.08
## 9 Large Cap    High                        0.876            1.30
## 10 Small Cap   High                        1.15             1.08
## # ... with 123 more rows, and 2 more variables: `5 Years Return` <dbl>, `10
## #   Years Return` <dbl>
```

```
# do the box plot again to check the normality and outliers if any
# now we have removed the outliers and made it normal

ggboxplot(
  Mrton, x = c("Market Cap"), y = c("1 year Return", "3 Years Return","5 Years Ret
urn","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
ggboxplot(
  Mrton, x = c("Price to Book Ratio (High/Low)"), y = c("1 year Return", "3 Years
 Return","5 Years Return","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```



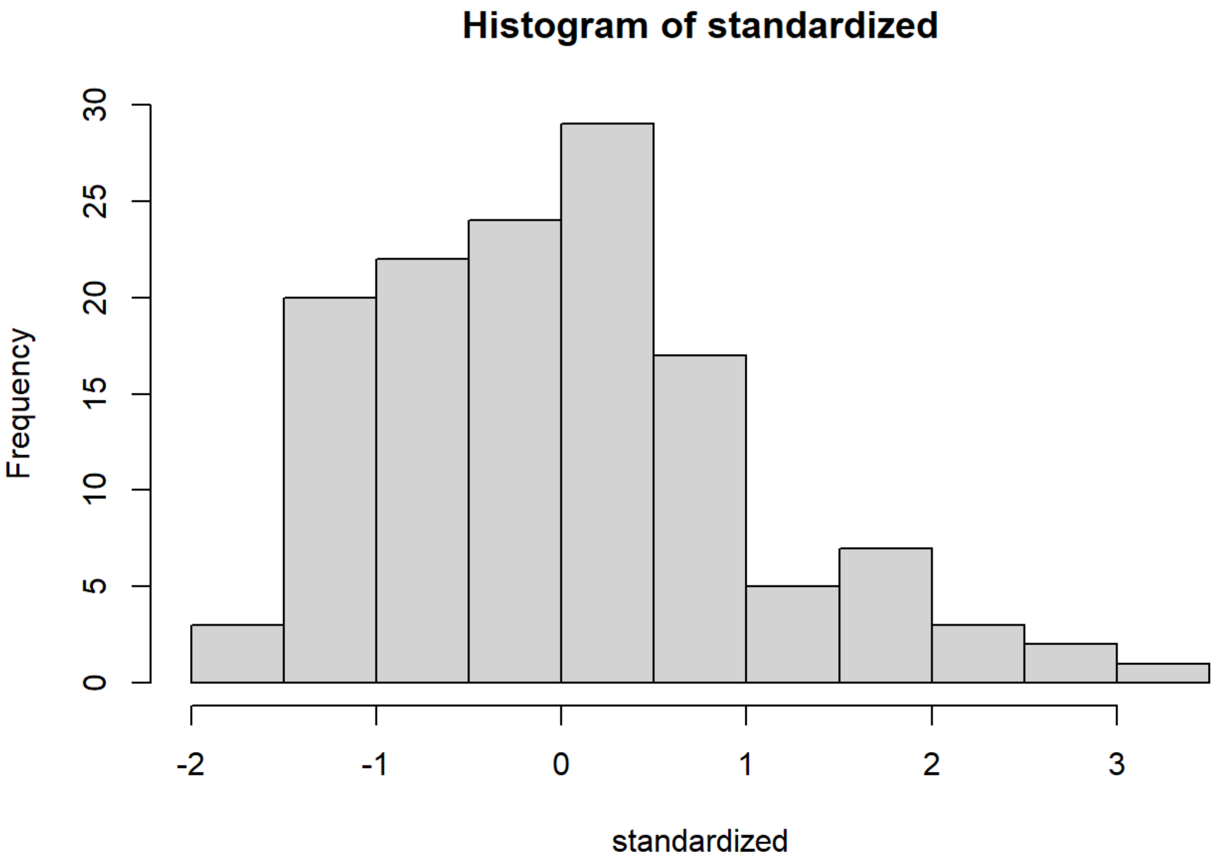

```
# so now you know, getting a good data is so much important it took lot of effort
 to reach till here..


##additivity
# checking correlation should have some correlation but coefficents should be less
than 0.99
# correlation close to 1 makes Manova unstable

correl=cor(Mrton[,-c(1,2)],use="pairwise.complete.obs")
symnum(correl)
```

```
##                   1yR 3YR 5YR 1YR
## 1 year Return     1
## 3 Years Return        1
## 5 Years Return        ,   1
## 10 Years Return       ,   ,   1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1
```

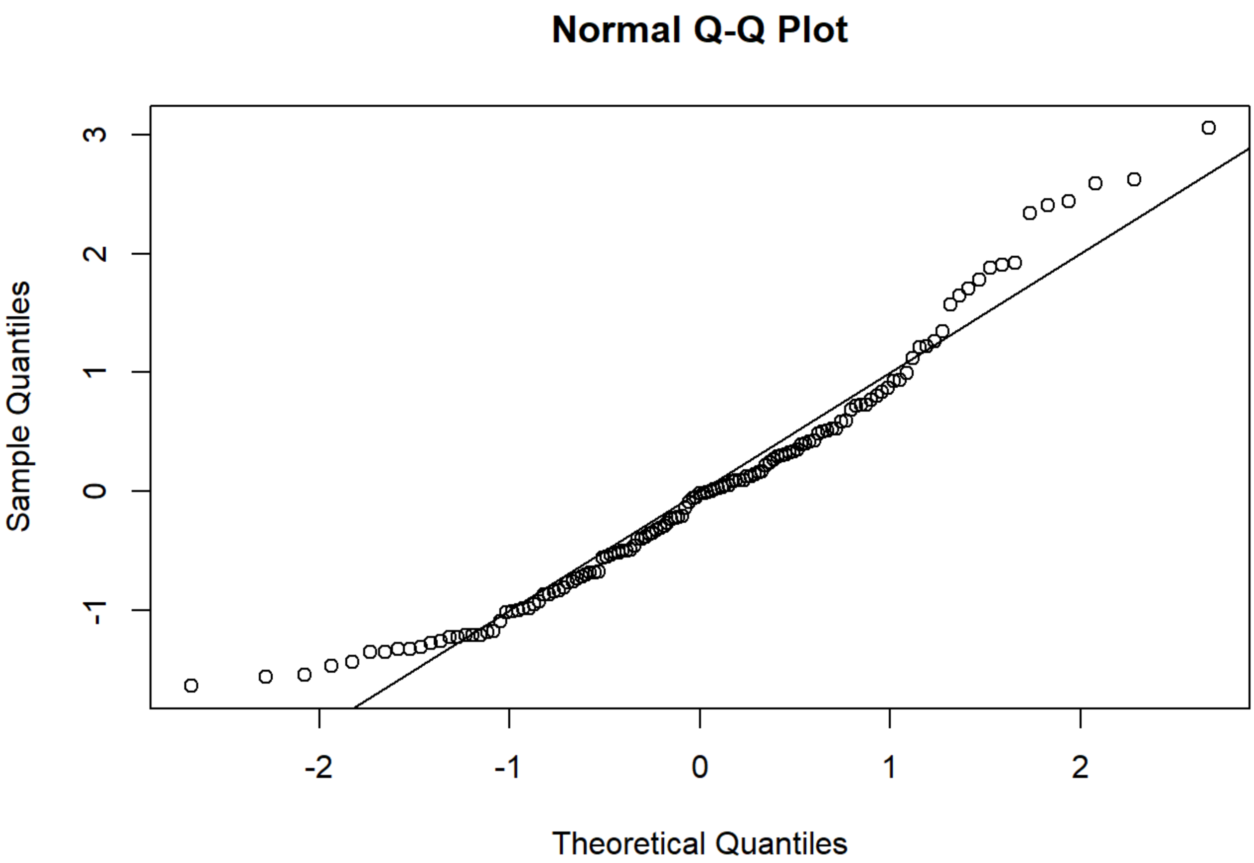
```
correl
```

```
##                   1 year Return 3 Years Return 5 Years Return 10 Years Return
## 1 year Return       1.000000000      0.1983257      0.1882161      0.002850151
## 3 Years Return      0.198325655      1.0000000      0.7846231      0.655625983
## 5 Years Return      0.188216091      0.7846231      1.0000000      0.796460521
## 10 Years Return     0.002850151      0.6556260      0.7964605      1.000000000
```

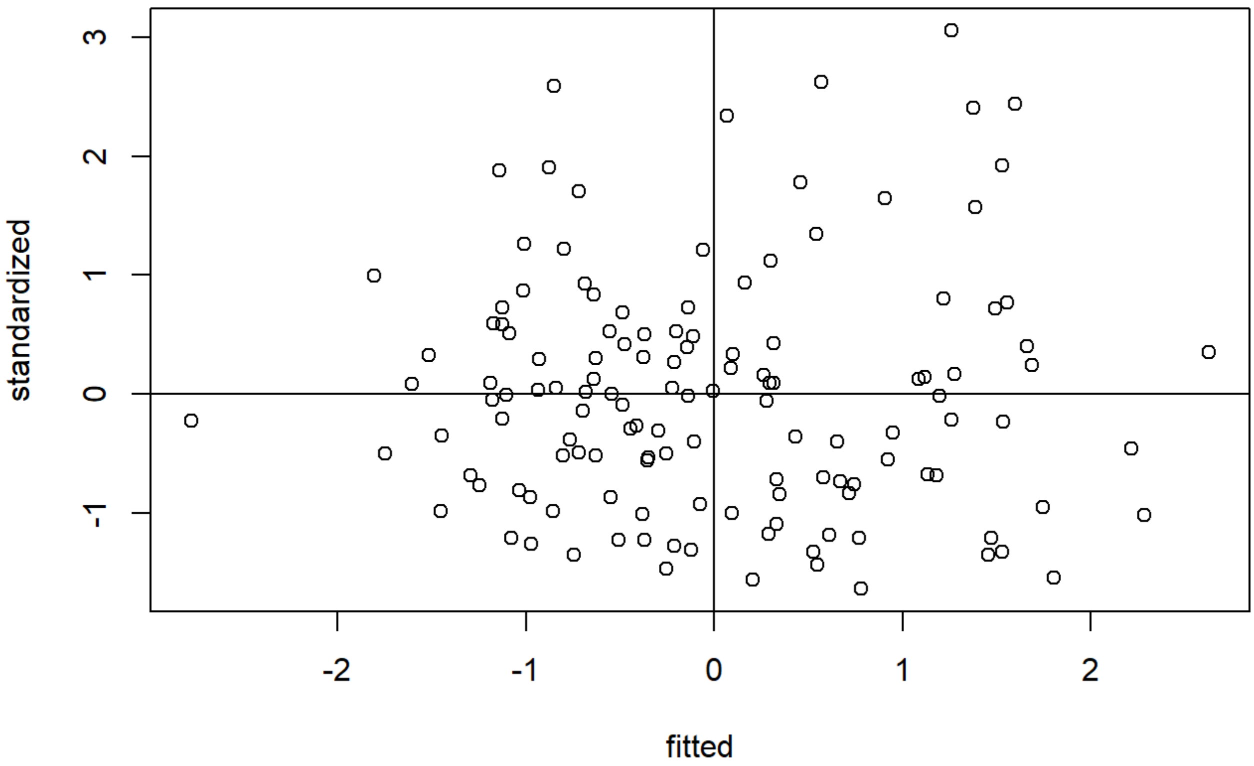
```
# all good

##assumption set up
random=rchisq(nrow(Mrton),7) # any no. let say 7 more than 3
fake=lm(random~.,data=Mrton[,])
standardized=rstudent(fake)
fitted=scale(fake$fitted.values)

##normality
hist(standardized)
```




```
#linearity
qqnorm(standardized)
abline(0,1)
```




```
##homogenity
plot(fitted,standardized)
abline(0,0)
abline(v=0)
```




```
#install.packages("energy")
library(energy)
```

```
## Warning: package 'energy' was built under R version 4.0.5
```

```
##leven's test

library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following objects are masked from 'package:faraway':
##
##     logit, vif
```

```
#install.packages("car")
leveneTest(Mrton$`1 year Return` ~ Mrton$`Market Cap`*Mrton$`Price to Book Ratio
 (High/Low)`,
           data= Mrton, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value  Pr(>F)
## group   5  1.9464 0.09115 .
##       127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Levene's Test of Equality of Variance: Used to examine whether or not the varianc
e between
#independent variable groups are equal; also known as homogeneity of variance. ass
umption for Manova
# since the p value is not significant or greater than 0.05 we cannot reject the h
ypothesis of
# homogeneity in variance

leveneTest(Mrton$`3 Years Return` ~ Mrton$`Market Cap`*Mrton$`Price to Book Ratio
 (High/Low)`,
           data=Mrton, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   5  1.7941 0.1186
##       127
```

```
leveneTest(Mrton$`5 Years Return` ~ Mrton$`Market Cap`*Mrton$`Price to Book Ratio
(High/Low)`,
           data= Mrton, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value  Pr(>F)
## group   5  3.0313 0.01281 *
##       127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
leveneTest(Mrton$`10 Years Return` ~ Mrton$`Market Cap`*Mrton$`Price to Book Ratio
(High/Low)`,
           data= Mrton, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value  Pr(>F)
## group   5  2.9263 0.01555 *
##       127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# for 5 year and 10 years return Levins test has a significant p value to reject t
he hypothesis
# It voilates the assumption of Manova for this specific data
```

```
# even then lets consider all the dependent variable
# we are checking for two dimension multi normality test
# multinorm

DV=cbind(Mrton$`1 year Return`,Mrton$`3 Years Return`,Mrton$`5 Years Return`,Mrton
$`10 Years Return`)
Dv1=cbind(Mrton$`1 year Return`,Mrton$`3 Years Return`)
mvnorm.etest(Dv1, R=200)
```

```
##
##  Energy test of multivariate normality: estimated parameters
##
## data:  x, sample size 133, dimension 2, replicates 200
## E-statistic = 0.65053, p-value = 0.515
```

```
Dv2=cbind(Mrton$`1 year Return`,Mrton$`5 Years Return`)
mvnorm.etest(Dv2, R=200)
```

```
##
##  Energy test of multivariate normality: estimated parameters
##
## data:  x, sample size 133, dimension 2, replicates 200
## E-statistic = 0.78316, p-value = 0.22
```

```
Dv3=cbind(Mrton$`1 year Return`,Mrton$`10 Years Return`)
mvnorm.etest(Dv3, R=200)
```

```
##
##  Energy test of multivariate normality: estimated parameters
##
## data:  x, sample size 133, dimension 2, replicates 200
## E-statistic = 0.91579, p-value = 0.085
```

```
Dv4=cbind(Mrton$`5 Years Return`,Mrton$`10 Years Return`)
mvnorm.etest(Dv4, R=200)
```

```
##
##  Energy test of multivariate normality: estimated parameters
##
## data:  x, sample size 133, dimension 2, replicates 200
## E-statistic = 0.6805, p-value = 0.37
```

```
#p values are high enough to not to reject the hypothesis they are indeed multi va
riate normal
output=lm(DV~ Mrton$`Market Cap`*Mrton$`Price to Book Ratio (High/Low)`, data=Mrto
n)
#contrasts=list(Manova$`Market Cap`= contr.Sum , Manova$`Price to Book Ratio (Hig
h/Low)`=contr.sum))
manova_out=Manova(output, type= "III")
summary(manova_out, multivariate=T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##          [,1]     [,2]     [,3]      [,4]
## [1,] 5.068734 1.617314 1.775321  1.710411
## [2,] 1.617314 2.817429 2.636821  3.332710
## [3,] 1.775321 2.636821 5.019336  7.104009
## [4,] 1.710411 3.332710 7.104009 19.294001
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##          [,1]     [,2]     [,3]     [,4]
## [1,] 20.51402 27.39896 30.75192 41.02403
## [2,] 27.39896 36.59464 41.07293 54.79258
## [3,] 30.75192 41.07293 46.09924 61.49784
## [4,] 41.02403 54.79258 61.49784 82.04005
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1  0.932025 425.0513      4    124 < 2.22e-16 ***
## Wilks             1  0.067975 425.0513      4    124 < 2.22e-16 ***
## Hotelling-Lawley  1 13.711332 425.0513      4    124 < 2.22e-16 ***
## Roy               1 13.711332 425.0513      4    124 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrton$`Market Cap`
##
## Sum of squares and products for the hypothesis:
##            [,1]       [,2]       [,3]       [,4]
## [1,]  0.4445232 -0.3797087 -0.3495510 -0.9167075
## [2,] -0.3797087  0.4479101  0.4534134  1.2348344
## [3,] -0.3495510  0.4534134  0.4688729  1.2869514
## [4,] -0.9167075  1.2348344  1.2869514  3.5423217
##
## Multivariate Tests: Mrton$`Market Cap`
##                  Df test stat  approx F num Df den Df     Pr(>F)
## Pillai            2 0.3499297  6.627175      8    250 7.4253e-08 ***
## Wilks             2 0.6603257  7.148952      8    248 1.6297e-08 ***
## Hotelling-Lawley  2 0.4988735  7.670180      8    246 3.6322e-09 ***
## Roy               2 0.4655106 14.547207      4    125 8.8075e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrton$`Price to Book Ratio (High/Low)`
##
## Sum of squares and products for the hypothesis:
##            [,1]       [,2]       [,3]      [,4]
```

```
## [1,]  0.2108984 -0.3521655 -0.3206741 -1.065547
## [2,] -0.3521655  0.5880580  0.5354726  1.779287
## [3,] -0.3206741  0.5354726  0.4875895  1.620179
## [4,] -1.0655470  1.7792872  1.6201795  5.383589
##
## Multivariate Tests: Mrton$`Price to Book Ratio (High/Low)`
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.3728771  18.4321      4    124 6.5798e-12 ***
## Wilks             1 0.6271229  18.4321      4    124 6.5798e-12 ***
## Hotelling-Lawley  1 0.5945837  18.4321      4    124 6.5798e-12 ***
## Roy               1 0.5945837  18.4321      4    124 6.5798e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrton$`Market Cap`:Mrton$`Price to Book Ratio (High/Low)`
##
## Sum of squares and products for the hypothesis:
##             [,1]          [,2]          [,3]         [,4]
## [1,]  0.12616702 -0.0137396725  0.0258350548 -0.058198693
## [2,] -0.01373967  0.0078946487  0.0008341787  0.025126342
## [3,]  0.02583505  0.0008341787  0.0073696758 -0.001206225
## [4,] -0.05819869  0.0251263419 -0.0012062249  0.082017205
##
## Multivariate Tests: Mrton$`Market Cap`:Mrton$`Price to Book Ratio (High/Low)`
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            2 0.0483577 0.7743115      8    250 0.62573
## Wilks             2 0.9518073 0.7751185      8    248 0.62502
## Hotelling-Lawley  2 0.0504595 0.7758141      8    246 0.62441
## Roy               2 0.0467508 1.4609635      4    125 0.21807
```

```
# from the summary of Manova we can see that for our factor 1  which is Market Cap
# the wilks coeeficient section the p value is less than 0.05 we can easily reject the null hypothesis that the
# Mean of return vectors for all the Market Caps are equal, similarly for factor 2 which Price to book ratio.
# since for interaction we have high P value we cannot reject the null hypothesis that there is zero interaction
# which means since there is no interaction we have to perform analysis for factor 1 and factor 2 seperately to
# from where the difference is coming from

# consider for factor1
# Now we have to find out from where the difference is coming from which level of Market Cap

R1.lm <- lm(Mrton$`1 year Return` ~ Mrton$`Market Cap`, data = Mrton)
R1.av <- aov(R1.lm)
summary(R1.av)
```

```
##                       Df Sum Sq Mean Sq F value Pr(>F)
## Mrton$`Market Cap`    2  0.299 0.14935   3.677  0.028 *
## Residuals           130  5.280 0.04062
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(R1.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = R1.lm)
##
## $`Mrton$`Market Cap``
##                             diff         lwr       upr     p adj
## Mid Cap-Large Cap   0.06656198 -0.02822151 0.1613455 0.2226307
## Small Cap-Large Cap 0.11980221  0.01213540 0.2274690 0.0251904
## Small Cap-Mid Cap   0.05324023 -0.05685340 0.1633339 0.4873576
```

```
# for small cap - Large cap interval is always positive

R2.lm <- lm(Mrton$`3 Years Return` ~ Mrton$`Market Cap`, data = Mrton)
R2.av <- aov(R2.lm)
summary(R2.av)
```

```
##                      Df Sum Sq Mean Sq F value Pr(>F)
## Mrton$`Market Cap`   2  0.306 0.15314   4.874 0.0091 **
## Residuals          130  4.085 0.03142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(R2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = R2.lm)
##
## $`Mrton$`Market Cap``
##                              diff        lwr         upr     p adj
## Mid Cap-Large Cap   -0.04618104 -0.1295470  0.03718496 0.3903355
## Small Cap-Large Cap -0.12470495 -0.2194024 -0.03000755 0.0062264
## Small Cap-Mid Cap   -0.07852392 -0.1753558  0.01830798 0.1364019
```

```
# for small cap - Large cap interval is always negative

R3.lm <- lm(Mrton$`5 Years Return` ~ Mrton$`Market Cap`, data = Mrton)
```

```
R3.av <- aov(R3.lm)
summary(R3.av)
```

```
##                        Df Sum Sq Mean Sq F value Pr(>F)
## Mrton$`Market Cap`    2  0.375 0.18744   3.841 0.0239 *
## Residuals           130  6.343 0.04879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(R3.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = R3.lm)
##
## $`Mrton$`Market Cap``
##                              diff        lwr         upr     p adj
## Mid Cap-Large Cap   -0.03709747 -0.1409865  0.06679158 0.6747805
## Small Cap-Large Cap -0.13702762 -0.2550377 -0.01901759 0.0184054
## Small Cap-Mid Cap   -0.09993016 -0.2206001  0.02073983 0.1254815
```

```
# for small cap - Large cap interval is always negative

R4.lm <- lm(Mrton$`10 Years Return` ~ Mrton$`Market Cap`, data = Mrton)
R4.av <- aov(R4.lm)
summary(R4.av)
```

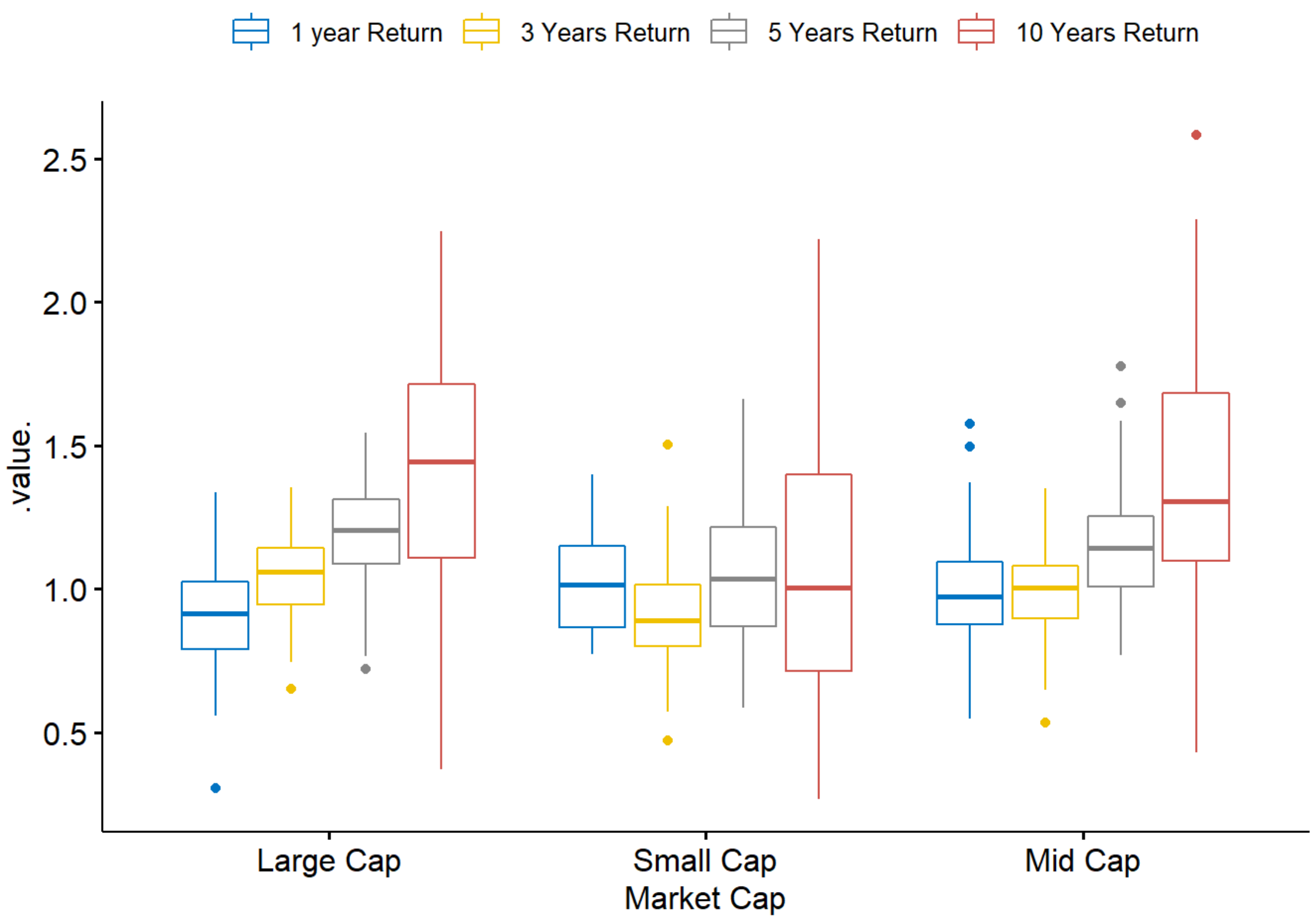
```
##                        Df Sum Sq Mean Sq F value Pr(>F)
## Mrton$`Market Cap`    2  2.005  1.0023   4.266 0.0161 *
## Residuals           130 30.543  0.2349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(R4.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = R4.lm)
##
## $`Mrton$`Market Cap``
##                              diff        lwr         upr     p adj
## Mid Cap-Large Cap   -0.01745386 -0.2454210  0.21051325 0.9820017
```

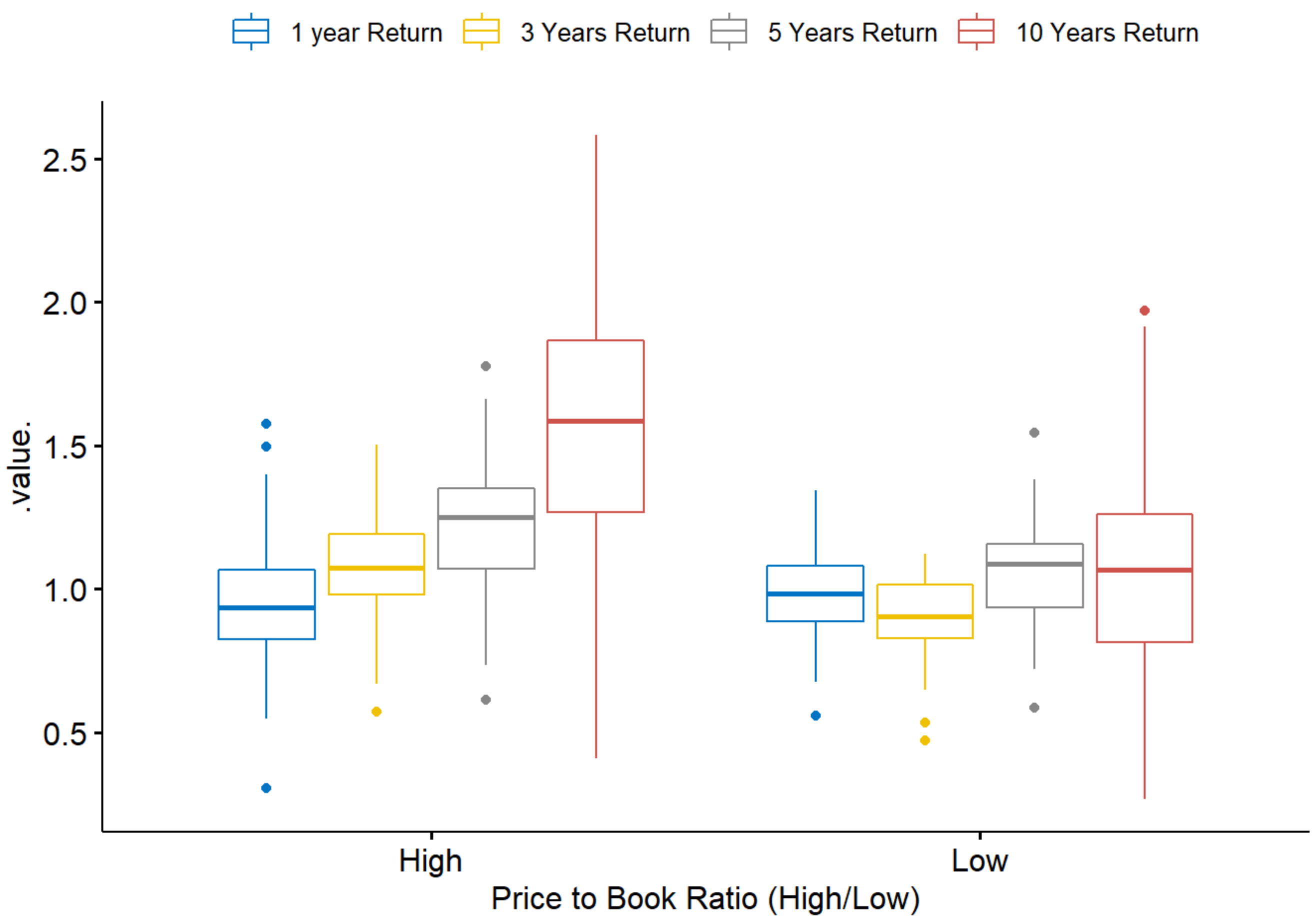
```
## Small Cap-Large Cap -0.29802644 -0.5569797 -0.03907322 0.0196893
## Small Cap-Mid Cap   -0.28057259 -0.5453626 -0.01578253 0.0350709
```

```
# for small cap - Large cap interval is always negative
# and for small cap - Mid cap interval is always negative as well


ggboxplot(
  Mrton, x = c("Market Cap"), y = c("1 year Return", "3 Years Return","5 Years Ret
urn","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
# final result from this entire analysis:
#Theres is no significant difference between the returns of Mid cap and Large Cap
 companies irrespective of no. of years of return choosen
#for 10 years of return theres a significant difference between samll cap and mid
 cap companies returns was higher for Mid cap compared to small cap companies
#except for 10 years of returns small and mid cap companies dont have significantl
y different return
# small cap companies gave higher  1 years of returns as compared to large cap com
panies which is little unobvious
# as obvious except for 1 years of return Large cap companies gave higher returns
 as compared to Small cap companies

#similarly we have to do analysis for factor 2 seperately..

# for visualization..
```

```
ggboxplot(
  Mrton, x = c("Price to Book Ratio (High/Low)"), y = c("1 year Return", "3 Years
 Return","5 Years Return","10 Years Return"),
  merge = TRUE, palette = "jco"
)
```




```
R1_f2.lm <- lm(Mrton$`1 year Return` ~ Mrton$`Price to Book Ratio (High/Low)`, dat
a = Mrton)
R1_f2.av <- aov(R1_f2.lm)
summary(R1_f2.av)
```

```
##                                         Df Sum Sq Mean Sq F value Pr(>F)
## Mrton$`Price to Book Ratio (High/Low)`   1  0.043 0.04328   1.024  0.313
## Residuals                              131  5.535 0.04225
```

```
#p value is greater than 0.5 we cannot reject the null hypothesis
# theres no significant difference between 1 year returns of company with high pb
 ratio as compared to company with low pb ratio
# no need to look at bornferroni interval as only two levels and theres no differe
nce seen

R2_f2.lm <- lm(Mrton$`3 Years Return` ~ Mrton$`Price to Book Ratio (High/Low)`, da
ta = Mrton)
R2_f2.av <- aov(R2_f2.lm)
summary(R2_f2.av)
```

```
##                                            Df Sum Sq Mean Sq F value   Pr(>F)    
## Mrton$`Price to Book Ratio (High/Low)`   1  1.004  1.0042   38.84 5.82e-09 ***
## Residuals                              131  3.387  0.0259                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# p values is very less we are rejecting the null hypothesis
# theres indeed a signifacant difference between the 3 years return for company wi
th high pb ratio and company with low pb ratio
tukey.test <- TukeyHSD(R2_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = R2_f2.lm)
## 
## $`Mrton$`Price to Book Ratio (High/Low)``
##                diff        lwr        upr p adj
## Low-High -0.1749032 -0.2304181 -0.1193882     0
```

```
# Low -High is always negative
# the company with high Pb ratio had higher 3 years return as compared to company
 with low pb ratio



R3_f2.lm <- lm(Mrton$`5 Years Return` ~ Mrton$`Price to Book Ratio (High/Low)`, da
ta = Mrton)
R3_f2.av <- aov(R3_f2.lm)
summary(R3_f2.av)
```

```
##                                            Df Sum Sq Mean Sq F value   Pr(>F)    
## Mrton$`Price to Book Ratio (High/Low)`   1  1.017  1.0171   23.37 3.68e-06 ***
## Residuals                              131  5.701  0.0435                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# p values is very less we are rejecting the null hypothesis
# theres indeed a signifacant difference between the 5 years return for company wi
th high pb ratio and company with low pb ratio
tukey.test <- TukeyHSD(R3_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = R3_f2.lm)
## 
## $`Mrton$`Price to Book Ratio (High/Low)``
##                diff        lwr        upr   p adj
## Low-High -0.1760225 -0.2480502 -0.1039948 3.7e-06
```

```
# Low -High is always negative
# the company with high Pb ratio had higher 5 years return as compared to company
 with low pb ratio


R4_f2.lm <- lm(Mrton$`10 Years Return` ~ Mrton$`Price to Book Ratio (High/Low)`, d
ata = Mrton)
R4_f2.av <- aov(R4_f2.lm)
summary(R4_f2.av)
```

```
##                                           Df Sum Sq Mean Sq F value   Pr(>F)
## Mrton$`Price to Book Ratio (High/Low)`   1  8.881   8.881   49.16 1.13e-10 ***
## Residuals                              131 23.666   0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# p values is very less we are rejecting the null hypothesis
# theres indeed a signifacant difference between the 10 years return for company w
ith high pb ratio and company with low pb ratio
tukey.test <- TukeyHSD(R4_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = R4_f2.lm)
##
## $`Mrton$`Price to Book Ratio (High/Low)``
##                diff       lwr       upr p adj
## Low-High -0.5201415 -0.666896 -0.373387     0
```

```
# Low -High is always negative
# the company with high Pb ratio had higher 10 years return as compared to company
with low pb ratio

#final analysis:
# except for the 1 years return all other had significant difference between the c
ompany with high and low pb ratio
# as predicted from box plot the company with high pb ratio had higher 3 , 5 and 1
0 years returns
# for 1 years return there was no difference between them

# here we came at the end of our Analysis for return

### Analysis for Risk..


str(Mrso)
```

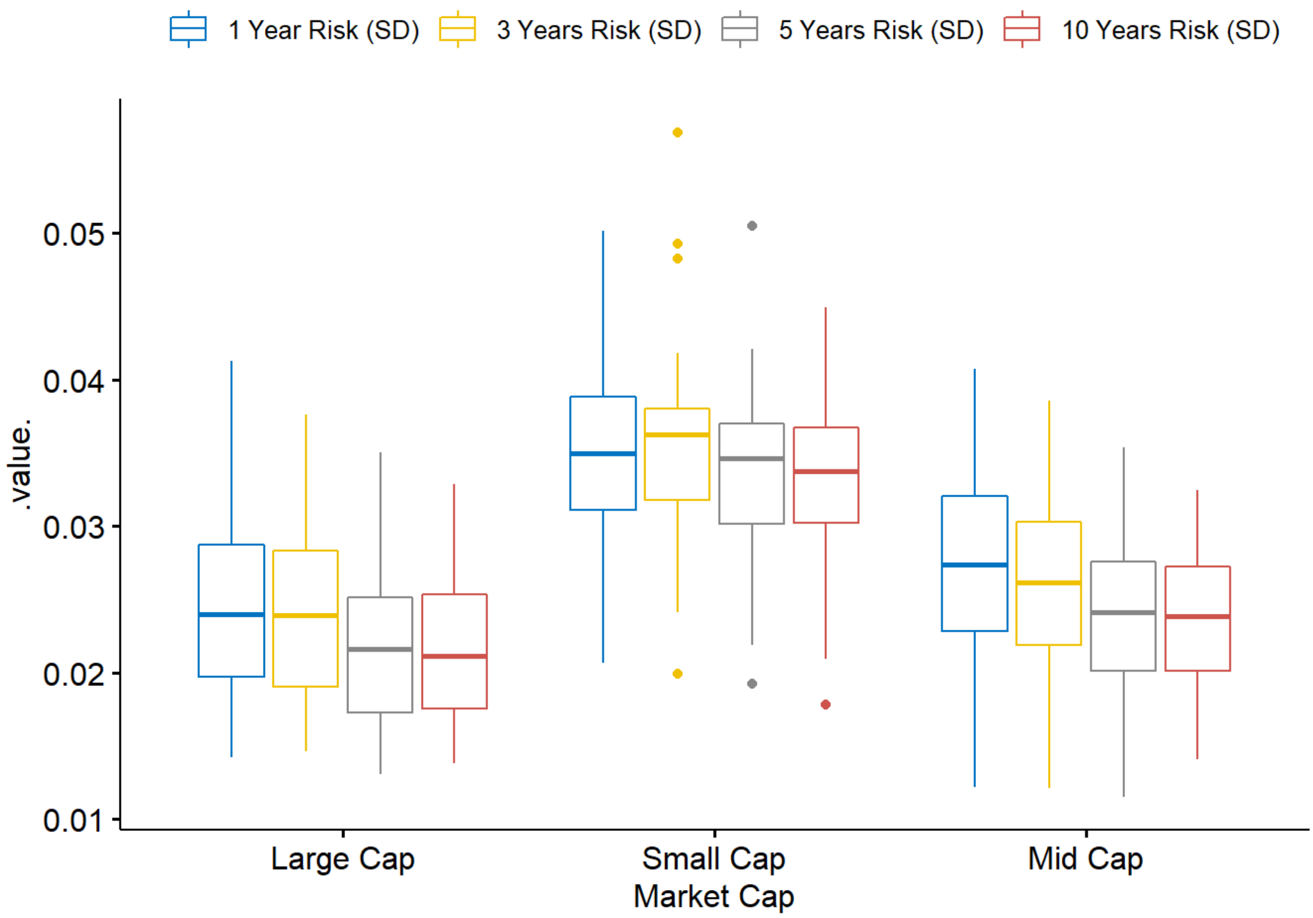
```
## tibble [159 x 6] (S3: tbl_df/tbl/data.frame)
##  $ Market Cap                        : chr [1:159] "Large Cap" "Small Cap" "Large C
ap" "Large Cap" ...
```

```
##  $ Price to Book Ratio (High/Low): chr [1:159] "High" "High" "High" "High" ...
##  $ 1 Year Risk (SD)              : num [1:159] 0.0175 0.0283 0.0227 0.024 0.027
1 ...
##  $ 3 Years Risk (SD)             : num [1:159] 0.0172 0.0241 0.0234 0.0245 0.02
32 ...
##  $ 5 Years Risk (SD)             : num [1:159] 0.0157 0.0219 0.022 0.024 0.02
...
##  $ 10 Years Risk (SD)            : num [1:159] 0.0152 0.021 0.0212 0.0247 0.020
9 ...
```

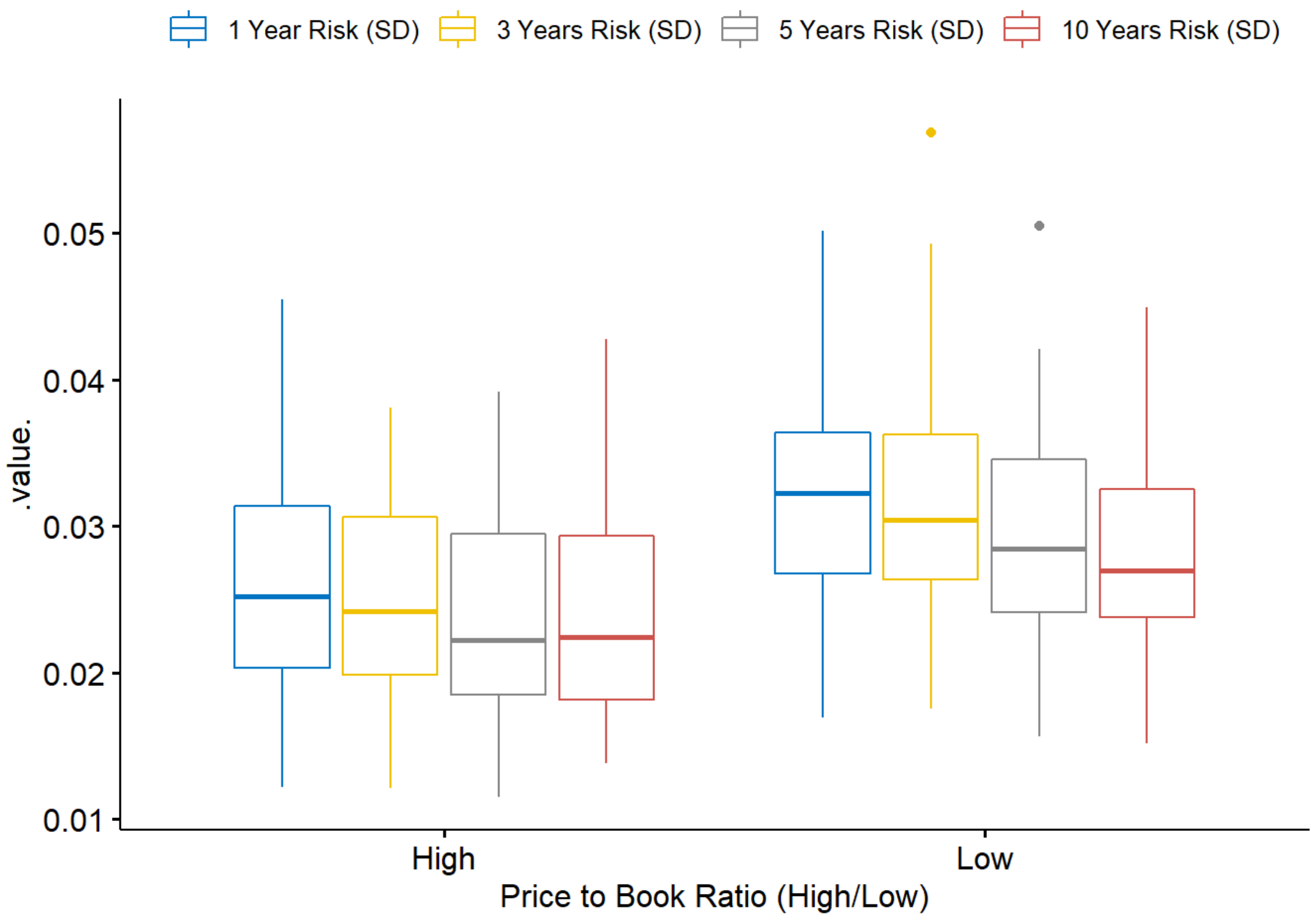
```
Mrso
```

```
## # A tibble: 159 x 6
##    `Market Cap` `Price to Book ~ `1 Year Risk (S~ `3 Years Risk (~
##    <chr>        <chr>                       <dbl>            <dbl>
##  1 Large Cap    High                       0.0175           0.0172
##  2 Small Cap    High                       0.0283           0.0241
##  3 Large Cap    High                       0.0227           0.0234
##  4 Large Cap    High                       0.0240           0.0245
##  5 Mid Cap      High                       0.0271           0.0232
##  6 Large Cap    High                       0.0171           0.0179
##  7 Large Cap    High                       0.0142           0.0146
##  8 Large Cap    High                       0.0163           0.0156
##  9 Large Cap    High                       0.0187           0.0196
## 10 Small Cap    High                       0.0365           0.0376
## # ... with 149 more rows, and 2 more variables: `5 Years Risk (SD)` <dbl>, `10
## #   Years Risk (SD)` <dbl>
```

```
# visualization
# wrt to Market Caps:
ggboxplot(
  Mrso, x = c("Market Cap"), y = c("1 Year Risk (SD)", "3 Years Risk (SD)","5 Year
s Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```




```
# As we have already removed the outliers the box plot looks good and data is perf
ect
#wrt to Price to Book Ratio (High/Low):
ggboxplot(
  Mrso, x = c("Price to Book Ratio (High/Low)"), y = c("1 Year Risk (SD)", "3 Year
s Risk (SD)","5 Years Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```




```
# lets try to remove few more outliers
cutoff=qchisq(1-0.05,4)
cutoff
```

```
## [1] 9.487729
```

```
mahal=mahalanobis(Mrso[,-c(1,2)],colMeans(Mrso[,-c(1,2)]),cov(Mrso[,-c(1,2)]))
summary(mahal<cutoff)
```

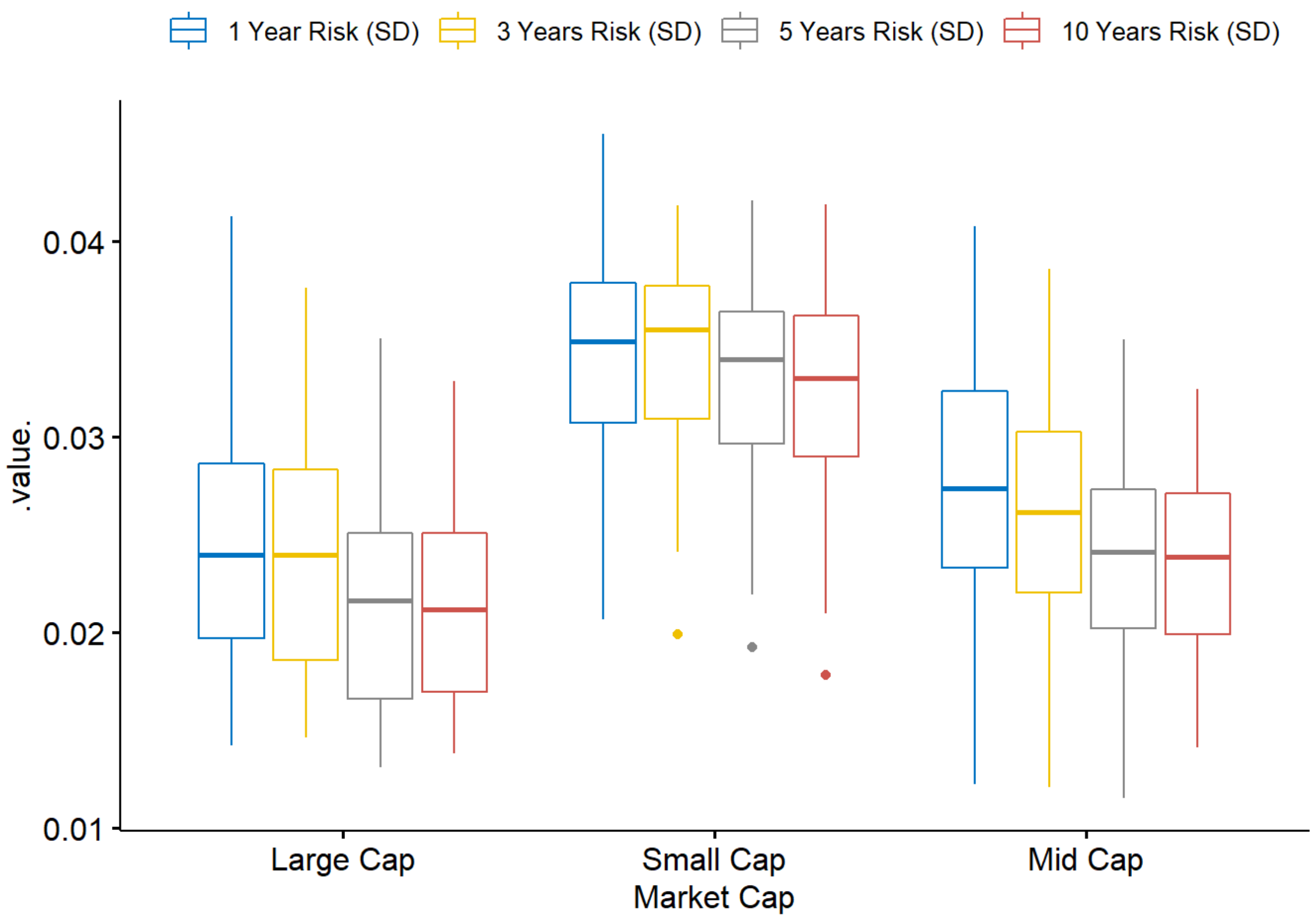
```
##    Mode   FALSE    TRUE
## logical      15     144
```

```
# 15 more outliers
noout=subset(Mrso,mahal<cutoff)
Mrsoo=noout
Mrsoo
```

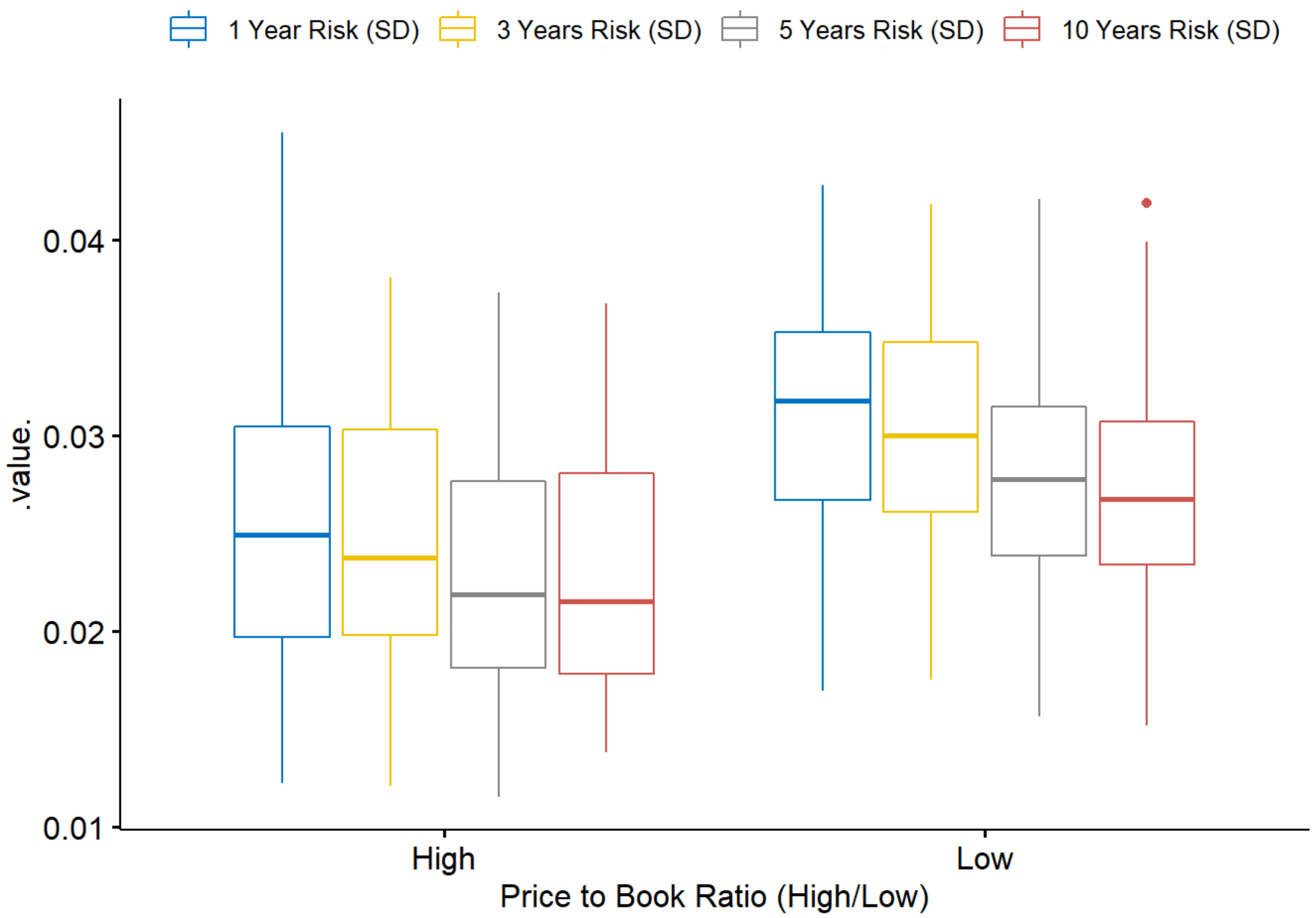
```
## # A tibble: 144 x 6
##    `Market Cap` `Price to Book ~ `1 Year Risk (S~ `3 Years Risk (~
##    <chr>        <chr>                       <dbl>            <dbl>
##  1 Large Cap    High                       0.0175           0.0172
##  2 Small Cap    High                       0.0283           0.0241
##  3 Large Cap    High                       0.0227           0.0234
##  4 Large Cap    High                       0.0240           0.0245
##  5 Mid Cap      High                       0.0271           0.0232
```

```
##   6 Large Cap    High                        0.0171          0.0179
##   7 Large Cap    High                        0.0142          0.0146
##   8 Large Cap    High                        0.0163          0.0156
##   9 Large Cap    High                        0.0187          0.0196
##  10 Small Cap    High                        0.0365          0.0376
## # ... with 134 more rows, and 2 more variables: `5 Years Risk (SD)` <dbl>, `10
## #    Years Risk (SD)` <dbl>
```

```
# see the box plot again
ggboxplot(
  Mrsoo, x = c("Market Cap"), y = c("1 Year Risk (SD)", "3 Years Risk (SD)","5 Yea
rs Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```



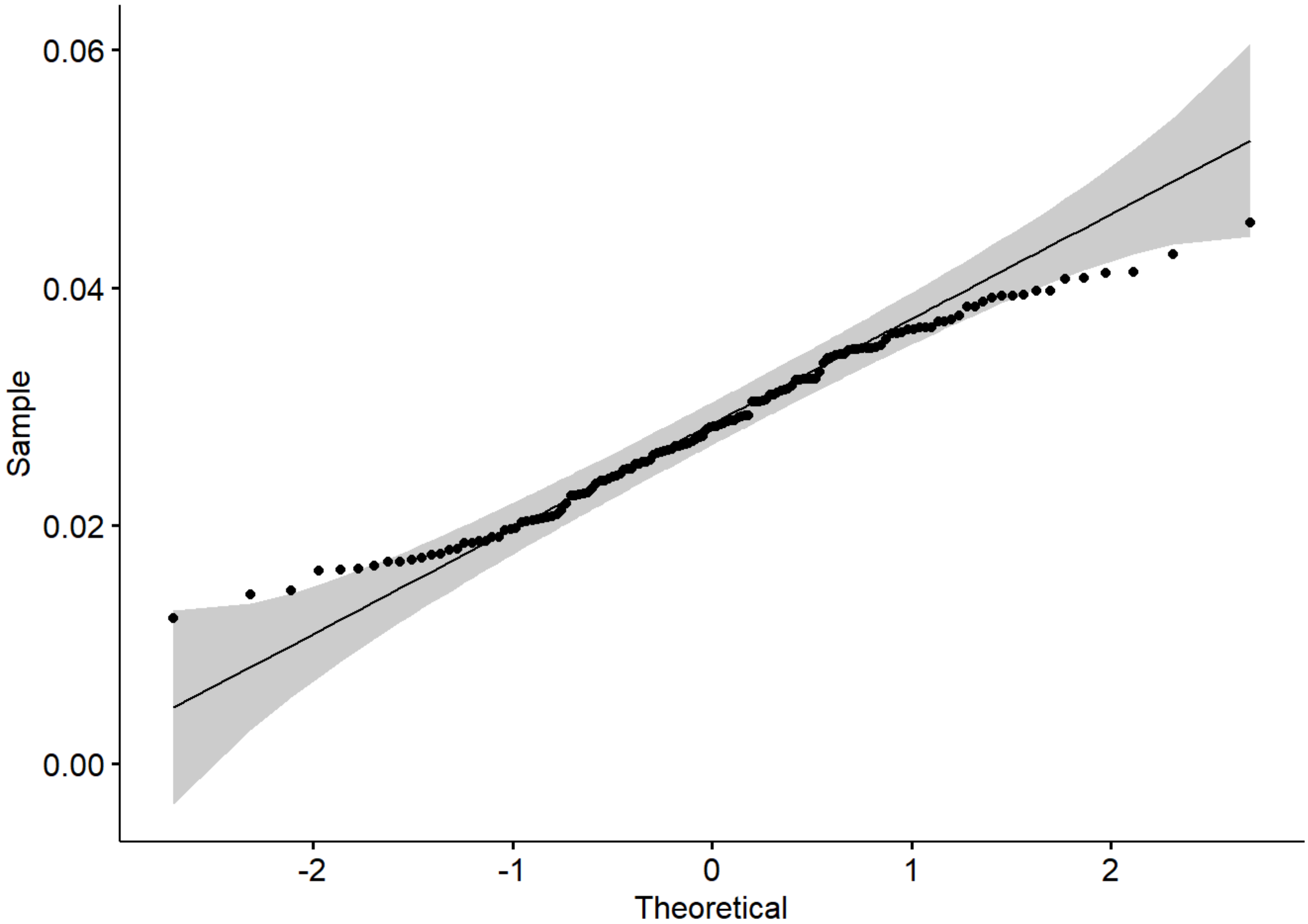

```
ggboxplot(
  Mrsoo, x = c("Price to Book Ratio (High/Low)"), y = c("1 Year Risk (SD)", "3 Yea
rs Risk (SD)","5 Years Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```




```
# now it look more good

# lets check for normality

Rs=Mrsoo[,-c(1,2)]
Rs=as.matrix(Rs)

shapiro.test(Rs[,1])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Rs[, 1]
## W = 0.98052, p-value = 0.03816
```

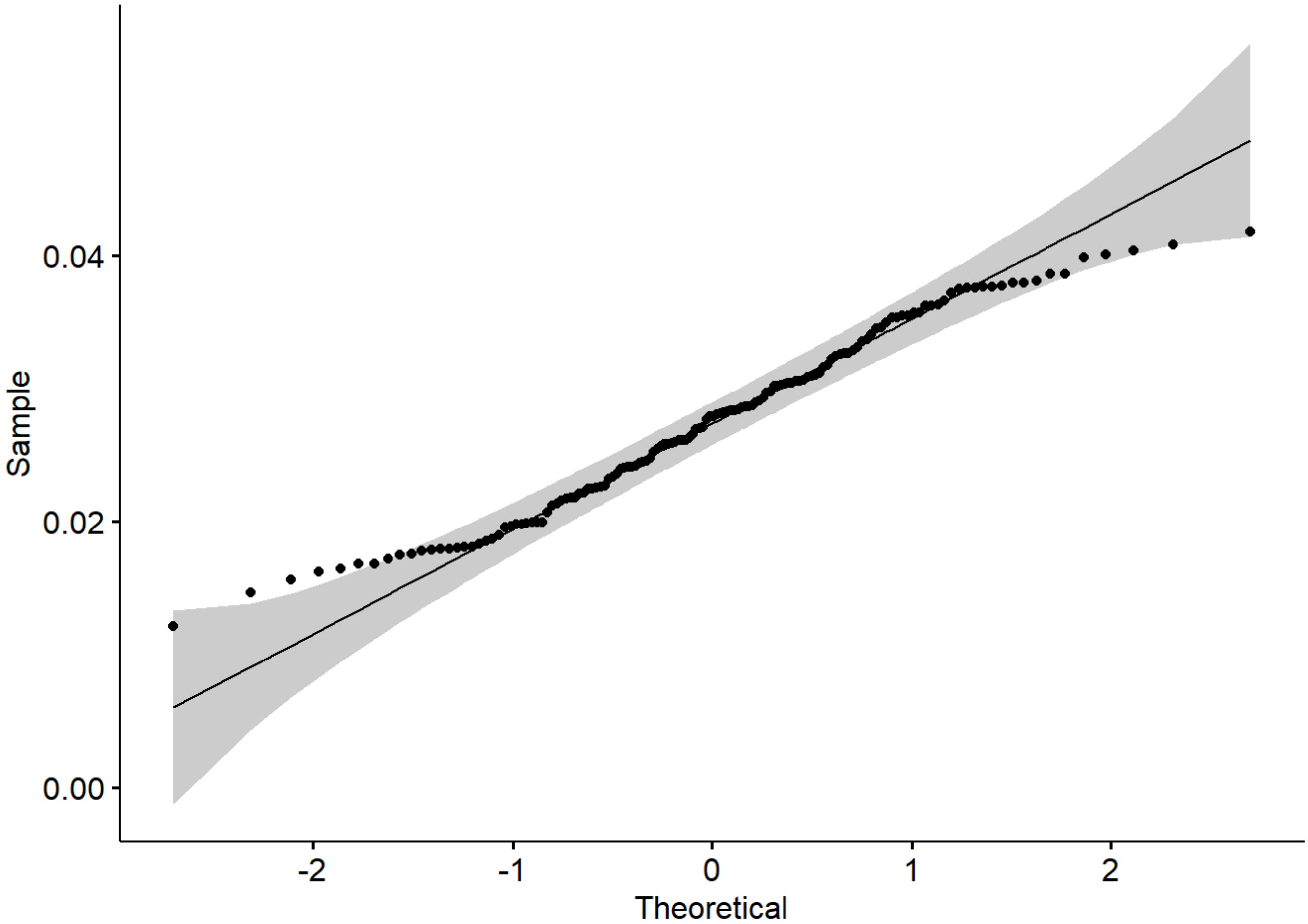
```
ggqqplot(Rs[,1])
```




```
shapiro.test(Rs[,2])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Rs[, 2]
## W = 0.97664, p-value = 0.01453
```

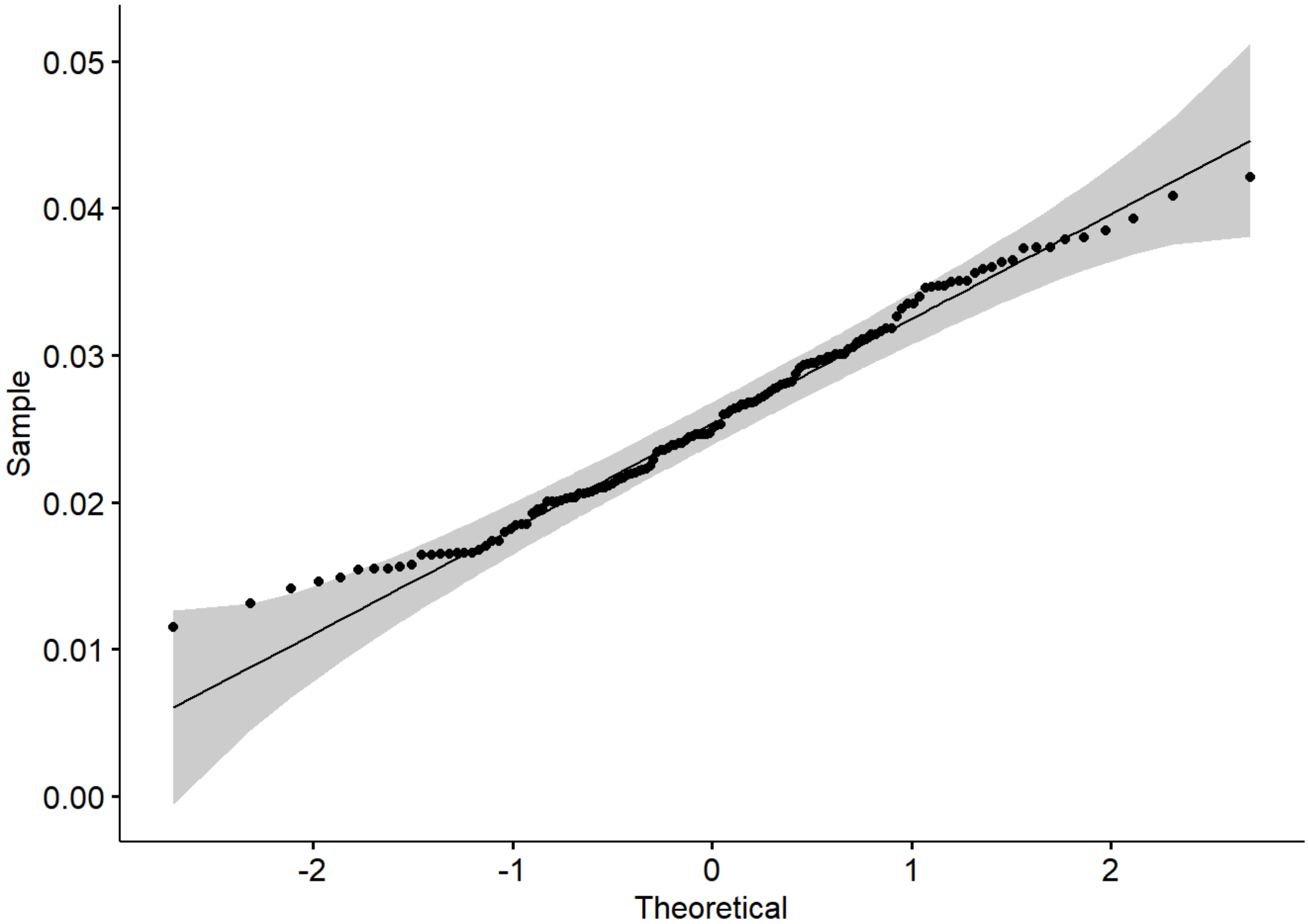
```
ggqqplot(Rs[,2])
```




```r
shapiro.test(Rs[,3])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Rs[, 3]
## W = 0.98109, p-value = 0.04414
```

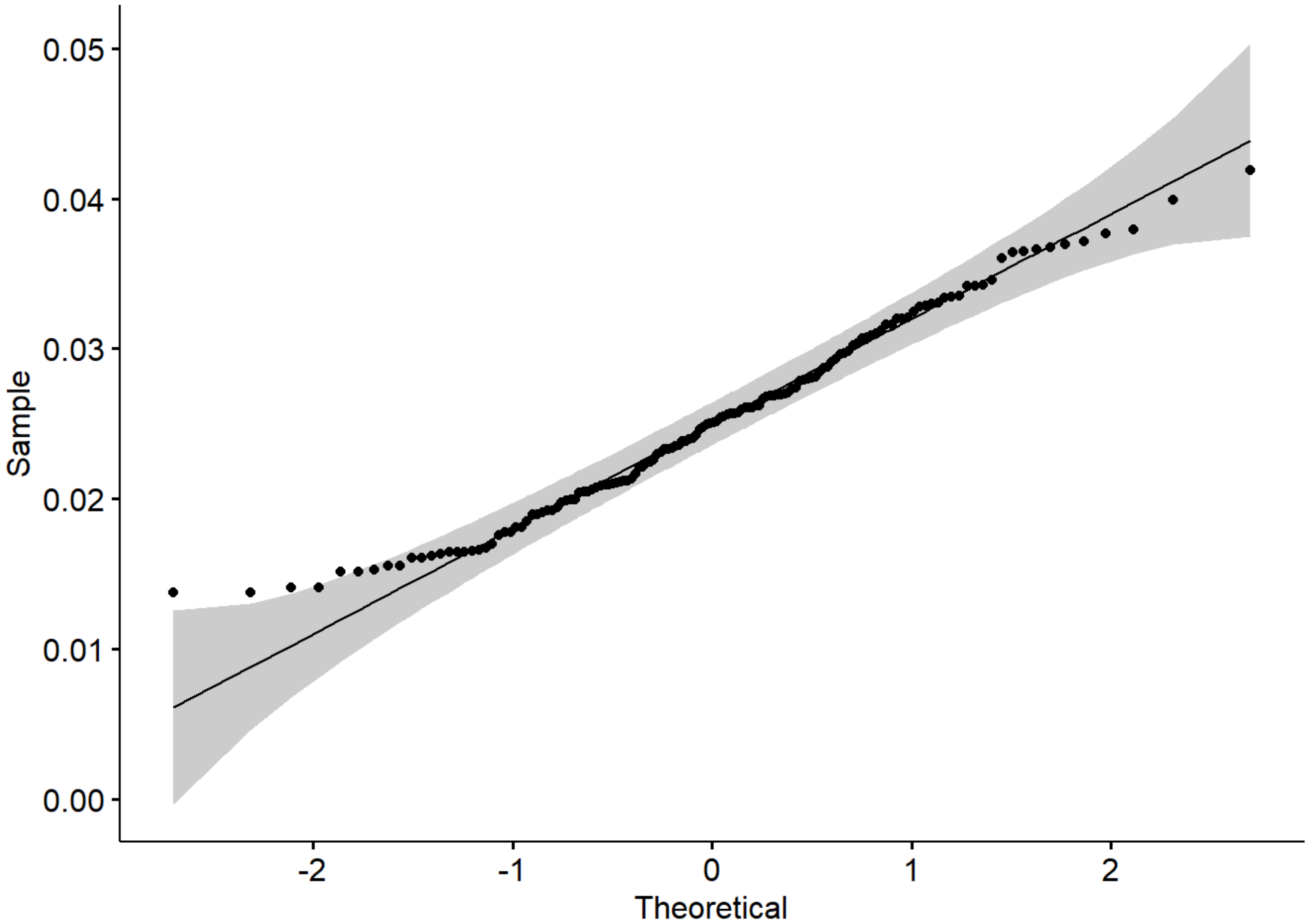
```r
ggqqplot(Rs[,3])
```




```
shapiro.test(Rs[,4])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Rs[, 4]
## W = 0.97816, p-value = 0.02113
```

```
ggqqplot(Rs[,4])
```

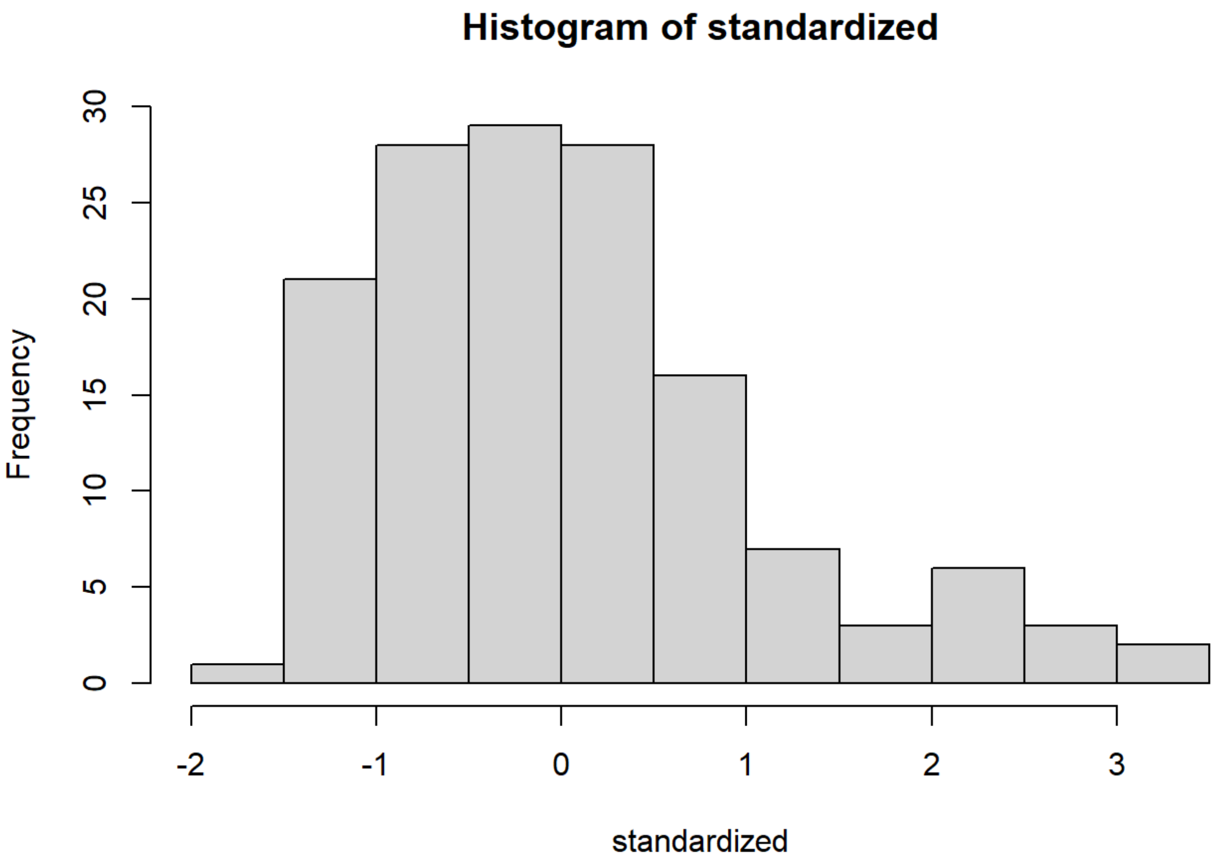



```
#Data looks quite normal p value is not very less, no need to do the transformatio
n

correl_rs=cor(Mrso[,-c(1,2)],use="pairwise.complete.obs")
correl_rs
```

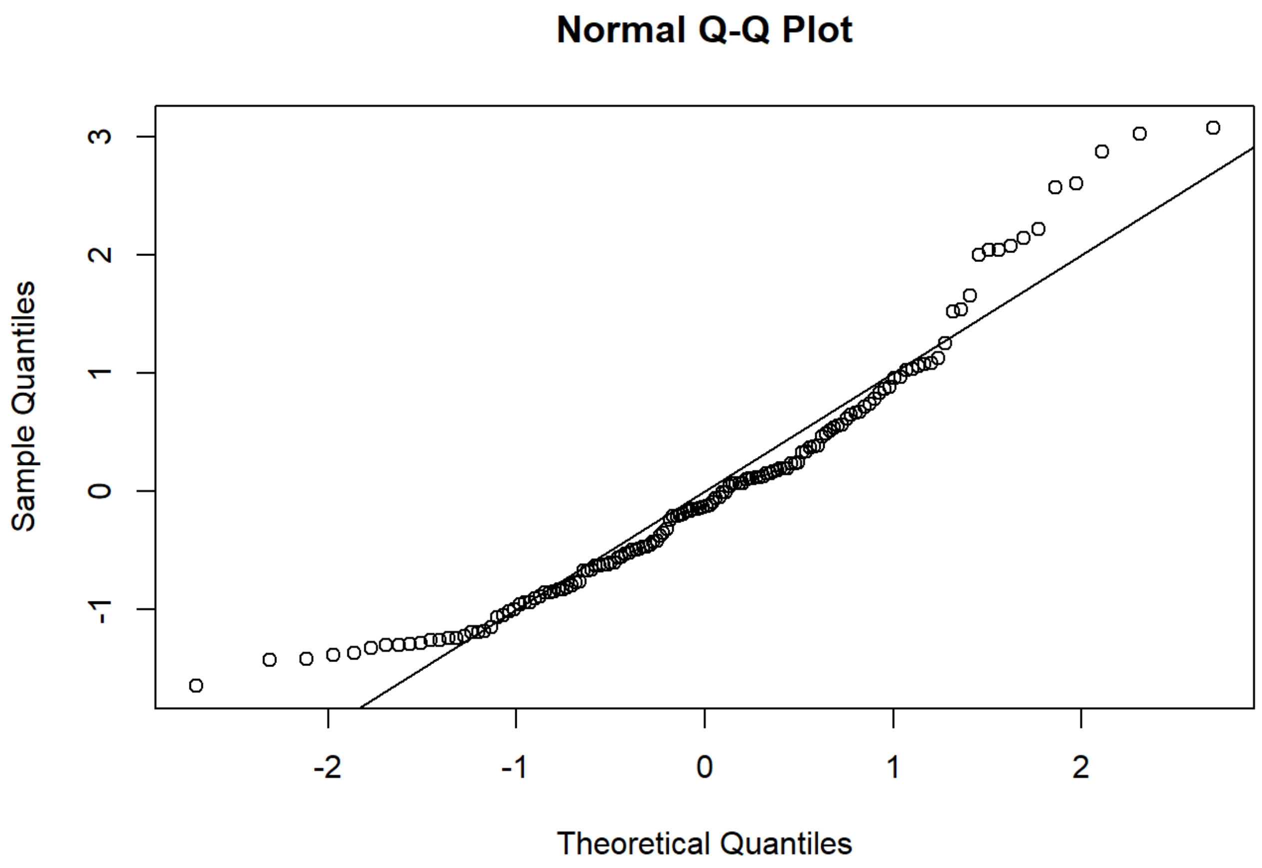
```
##                    1 Year Risk (SD) 3 Years Risk (SD) 5 Years Risk (SD)
## 1 Year Risk (SD)          1.0000000         0.9158521         0.8902871
## 3 Years Risk (SD)         0.9158521         1.0000000         0.9774418
## 5 Years Risk (SD)         0.8902871         0.9774418         1.0000000
## 10 Years Risk (SD)        0.8293213         0.9100177         0.9511580
##                    10 Years Risk (SD)
## 1 Year Risk (SD)            0.8293213
## 3 Years Risk (SD)           0.9100177
## 5 Years Risk (SD)           0.9511580
## 10 Years Risk (SD)          1.0000000
```

```
# therse a quite good correlation between the dependent variable since the correla
tion values should be less than
# 0.99 we are still allowed to apply Manova if the correlation is 1 very close to
 1 it makes Manova unstable
# we are good to go..

##assumption set up

random=rchisq(nrow(Mrsoo),7) # any no. let say 7 more than 3
fake=lm(random~.,data=Mrsoo[,])
```

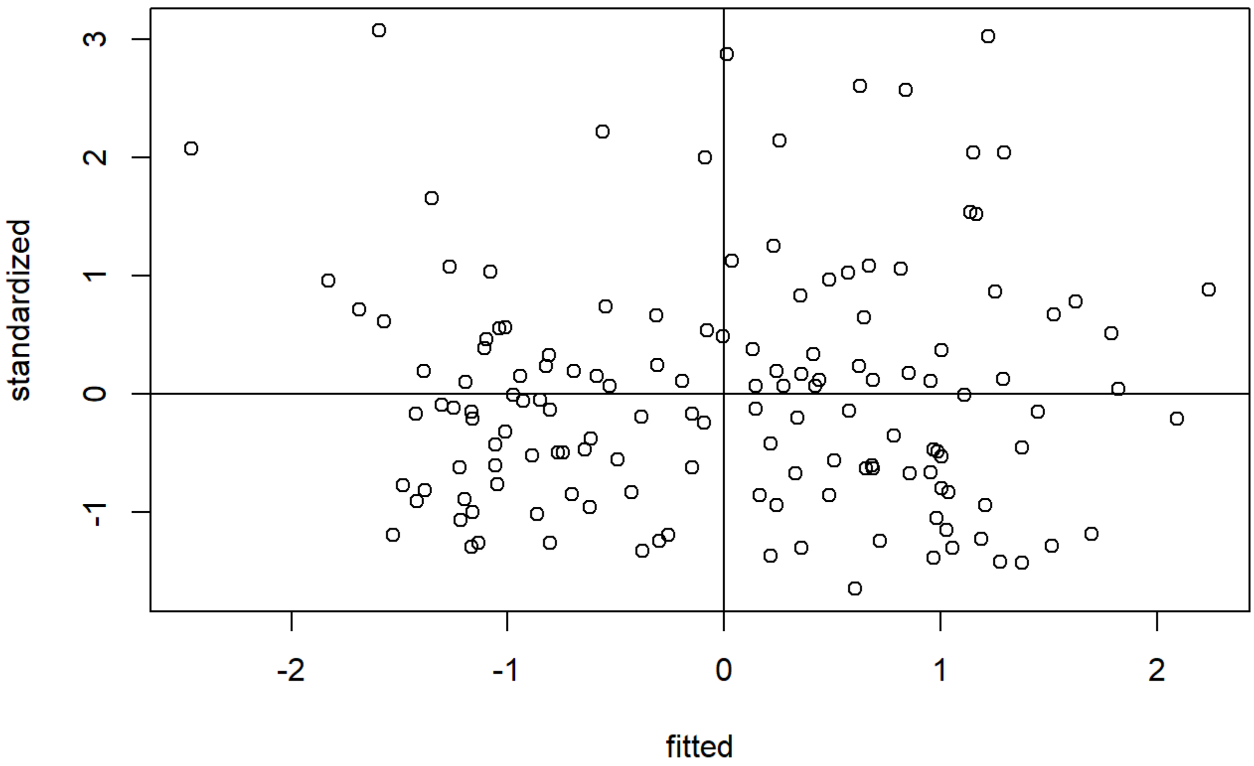
```
standardized=rstudent(fake)
fitted=scale(fake$fitted.values)

##normality
hist(standardized)
```




```
#linearity
qqnorm(standardized)
abline(0,1)
```




```
##homogenity
plot(fitted,standardized)
abline(0,0)
abline(v=0)
```




```
#Levins test

library(car)
#install.packages("car")
leveneTest(Mrsoo$`1 Year Risk (SD)` ~ Mrsoo$`Market Cap`*Mrsoo$`Price to Book Rati
o (High/Low)`,
           data= Mrsoo, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   5  1.4222   0.22
##       138
```

```
# since the p value is greater than 0.05 we cannot reject the null hypothesis
# 1 years Risk has homogeneous variance across all the levels of both the factor..
leveneTest(Mrsoo$`3 Years Risk (SD)` ~ Mrsoo$`Market Cap`*Mrsoo$`Price to Book Rat
io (High/Low)`,
           data= Mrsoo, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   5  1.4285 0.2178
##       138
```

```
#since the p value is greater than 0.05 we cannot reject the null hypothesis
# 3 years Risk has homogeneous variance across all the levels of both the factor..
```

```
leveneTest(Mrsoo$`5 Years Risk (SD)` ~ Mrsoo$`Market Cap`*Mrsoo$`Price to Book Rat
io (High/Low)`,
           data= Mrsoo, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   5  0.6679 0.6484
##       138
```

```
#since the p value is greater than 0.05 we cannot reject the null hypothesis
# 5 years Risk has homogeneous variance across all the levels of both the factor..
leveneTest(Mrsoo$`10 Years Risk (SD)` ~ Mrsoo$`Market Cap`*Mrsoo$`Price to Book Ra
tio (High/Low)`,
           data= Mrsoo, center= mean)
```

```
## Levene's Test for Homogeneity of Variance (center = mean)
##        Df F value Pr(>F)
## group   5  0.5173 0.7629
##       138
```

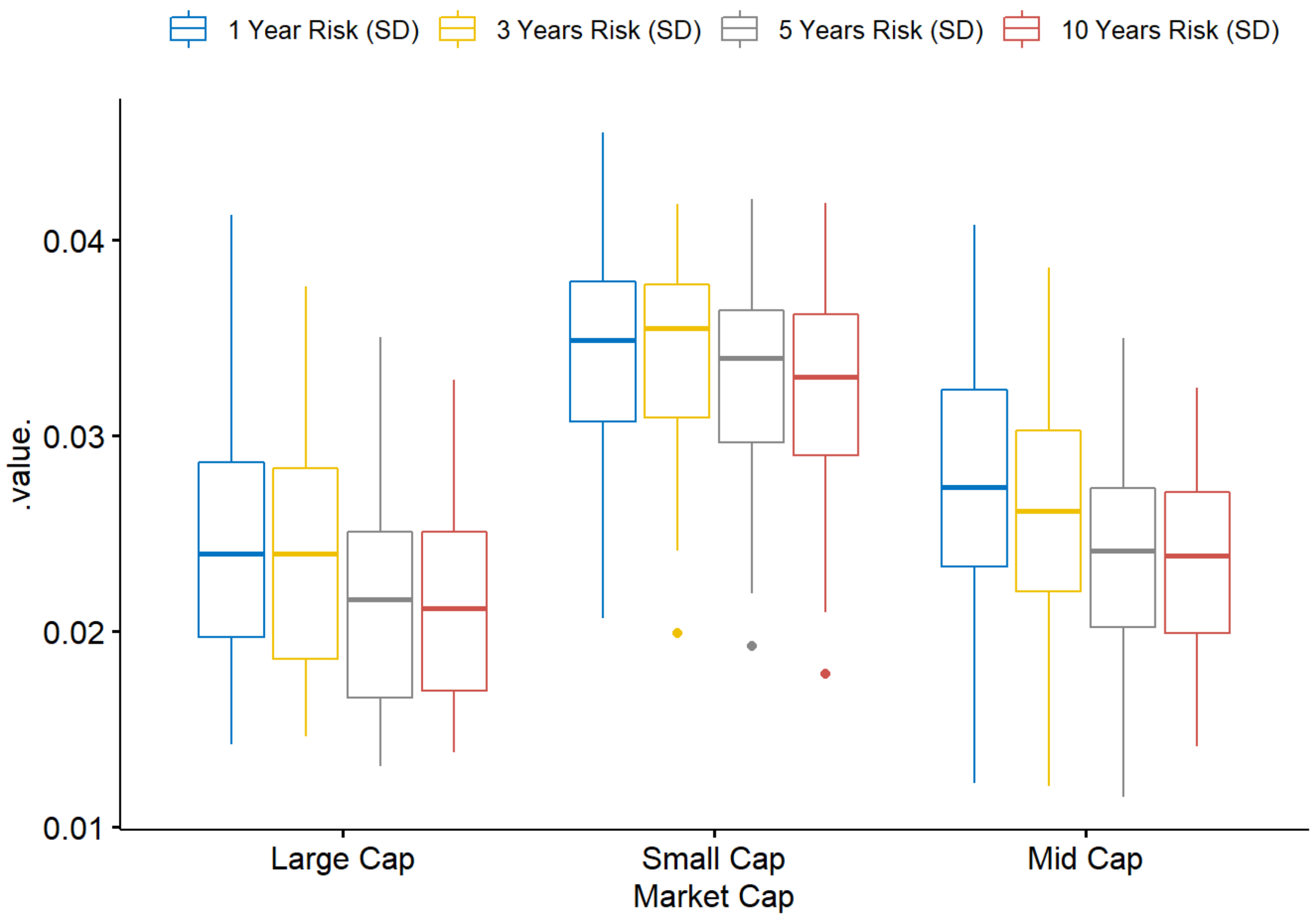
```
#since the p value is greater than 0.05 we cannot reject the null hypothesis
# 10 years Risk has homogeneous variance across all the levels of both the facto
r..

# our assumptions for doing Manova stands true..


# Manova
DV_rs=cbind(Mrsoo$`1 Year Risk (SD)`,Mrsoo$`3 Years Risk (SD)`,Mrsoo$`5 Years Risk
(SD)`,Mrsoo$`10 Years Risk (SD)`)
output=lm(DV_rs~ Mrsoo$`Market Cap`*Mrsoo$`Price to Book Ratio (High/Low)`, data=M
rsoo)
#contrasts=list(Manova$`Market Cap`= contr.Sum , Manova$`Price to Book Ratio (Hig
h/Low)`=contr.sum))
manova_out=Manova(output, type= "III")
summary(manova_out, multivariate=T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##             [,1]        [,2]        [,3]        [,4]
## [1,] 0.004505561 0.003415446 0.003051699 0.002634155
## [2,] 0.003415446 0.003209226 0.002936781 0.002587409
## [3,] 0.003051699 0.002936781 0.002892653 0.002605438
## [4,] 0.002634155 0.002587409 0.002605438 0.002713305
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##            [,1]       [,2]        [,3]        [,4]
```

```
## [1,] 0.01159528 0.01151254 0.010619619 0.010602556
## [2,] 0.01151254 0.01143039 0.010543840 0.010526899
## [3,] 0.01061962 0.01054384 0.009726051 0.009710424
## [4,] 0.01060256 0.01052690 0.009710424 0.009694822
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1  0.798295 133.5734      4    135 < 2.22e-16 ***
## Wilks             1  0.201705 133.5734      4    135 < 2.22e-16 ***
## Hotelling-Lawley  1  3.957730 133.5734      4    135 < 2.22e-16 ***
## Roy               1  3.957730 133.5734      4    135 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrsoo$`Market Cap`
##
## Sum of squares and products for the hypothesis:
##             [,1]        [,2]        [,3]        [,4]
## [1,] 0.001488066 0.001508260 0.001522729 0.001478312
## [2,] 0.001508260 0.001553469 0.001575402 0.001524906
## [3,] 0.001522729 0.001575402 0.001599611 0.001547075
## [4,] 0.001478312 0.001524906 0.001547075 0.001497076
##
## Multivariate Tests: Mrsoo$`Market Cap`
##                  Df test stat  approx F num Df den Df     Pr(>F)
## Pillai            2 0.4019990  8.553164      8    272 2.1665e-10 ***
## Wilks             2 0.6106664  9.438862      8    270 1.6868e-11 ***
## Hotelling-Lawley  2 0.6168150 10.331652      8    268 1.3417e-12 ***
## Roy               2 0.5811252 19.758258      4    136 7.6762e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrsoo$`Price to Book Ratio (High/Low)`
##
## Sum of squares and products for the hypothesis:
##              [,1]         [,2]         [,3]         [,4]
## [1,] 0.0008211147 0.0007195262 0.0006313284 0.0005233582
## [2,] 0.0007195262 0.0006305062 0.0005532203 0.0004586082
## [3,] 0.0006313284 0.0005532203 0.0004854079 0.0004023931
## [4,] 0.0005233582 0.0004586082 0.0004023931 0.0003335755
##
## Multivariate Tests: Mrsoo$`Price to Book Ratio (High/Low)`
##                  Df test stat approx F num Df den Df    Pr(>F)
## Pillai            1 0.1706408 6.944067      4    135 4.099e-05 ***
## Wilks             1 0.8293592 6.944067      4    135 4.099e-05 ***
## Hotelling-Lawley  1 0.2057501 6.944067      4    135 4.099e-05 ***
## Roy               1 0.2057501 6.944067      4    135 4.099e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Mrsoo$`Market Cap`:Mrsoo$`Price to Book Ratio (High/Low)`
```

```
##
## Sum of squares and products for the hypothesis:
##              [,1]         [,2]         [,3]         [,4]
## [1,] 9.958855e-05 7.343190e-05 5.312079e-05 4.577719e-05
## [2,] 7.343190e-05 5.929881e-05 4.584233e-05 3.992766e-05
## [3,] 5.312079e-05 4.584233e-05 3.697661e-05 3.241224e-05
## [4,] 4.577719e-05 3.992766e-05 3.241224e-05 2.843787e-05
##
## Multivariate Tests: Mrsoo$`Market Cap`:Mrsoo$`Price to Book Ratio (High/Low)`
##                  Df test stat  approx F num Df den Df  Pr(>F)
## Pillai            2 0.0396708 0.6880507      8    272 0.70201
## Wilks             2 0.9606115 0.6849836      8    270 0.70470
## Hotelling-Lawley  2 0.0407097 0.6818872      8    268 0.70741
## Roy               2 0.0313303 1.0652301      4    136 0.37626
```

```
# look at wilks coefficient row p value
# we can clearly see from the interaction part that p value is large we, cannot re
ject the null hypothesis of
# zero interaction
# p value for Market cap and PB ratio is very less we are rejecting the null hypot
hesis that there is no difference
# between the mean Risk vector for both the factors for there respective levels

#Now since there is no interaction between the two factor we have to consider both
the factor seperately
#Basically now we have to do Analysis to check from where the difference is actual
ly coming from
# lets start with factor 1 Market Cap (Analysis):
# for visualization see box plot
ggboxplot(
  Mrsoo, x = c("Market Cap"), y = c("1 Year Risk (SD)", "3 Years Risk (SD)","5 Yea
rs Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```




```
Rs1.lm <- lm(Mrsoo$`1 Year Risk (SD)` ~ Mrsoo$`Market Cap`, data = Mrsoo)
Rs1.av <- aov(Rs1.lm)
summary(Rs1.av)
```

```
##                     Df   Sum Sq   Mean Sq F value   Pr(>F)
## Mrsoo$`Market Cap`   2 0.002070 0.0010348   25.27 4.17e-10 ***
## Residuals          141 0.005773 0.0000409
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(Rs1.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs1.lm)
##
## $`Mrsoo$`Market Cap``
##                            diff          lwr         upr     p adj
## Mid Cap-Large Cap   0.003001510 4.305101e-05 0.005959968 0.0459217
## Small Cap-Large Cap 0.009515113 6.317439e-03 0.012712786 0.0000000
## Small Cap-Mid Cap   0.006513603 3.302922e-03 0.009724284 0.0000116
```

```
# for Mid Cap - Large Cap, Small Cap - Large Cap, and Small Cap-Mid cap  the diffe
rence of means are always positive

Rs2.lm <- lm(Mrsoo$`3 Years Risk (SD)` ~ Mrsoo$`Market Cap`, data = Mrsoo)
Rs2.av <- aov(Rs2.lm)
summary(Rs2.av)
```

```
##                      Df   Sum Sq   Mean Sq F value   Pr(>F)    
## Mrsoo$`Market Cap`   2 0.002568 0.0012840   42.15 4.48e-15 ***
## Residuals          141 0.004296 0.0000305                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(Rs2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = Rs2.lm)
## 
## $`Mrsoo$`Market Cap``
##                            diff           lwr         upr    p adj
## Mid Cap-Large Cap   0.002084736 -0.0004672041 0.004636676 0.132671
## Small Cap-Large Cap 0.010321811  0.0075635265 0.013080096 0.000000
## Small Cap-Mid Cap   0.008237075  0.0054675703 0.011006580 0.000000
```

```
# for Small Cap - Large Cap, and Small Cap-Mid cap  the difference of means are al
ways positive

Rs3.lm <- lm(Mrsoo$`5 Years Risk (SD)` ~ Mrsoo$`Market Cap`, data = Mrsoo)
Rs3.av <- aov(Rs3.lm)
summary(Rs3.av)
```

```
##                      Df   Sum Sq  Mean Sq F value Pr(>F)    
## Mrsoo$`Market Cap`   2 0.002846 0.001423   52.62 <2e-16 ***
## Residuals          141 0.003813 0.000027                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(Rs3.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = Rs3.lm)
```

```
##
## $`Mrsoo$`Market Cap``
##                              diff           lwr         upr     p adj
## Mid Cap-Large Cap   0.002061682 -0.0003425588 0.004465922 0.1085125
## Small Cap-Large Cap 0.010826542  0.0082278997 0.013425185 0.0000000
## Small Cap-Mid Cap   0.008764860  0.0061556472 0.011374074 0.0000000
```

```
# for Small Cap - Large Cap, and Small Cap-Mid cap  the difference of means are al
ways positive

Rs4.lm <- lm(Mrsoo$`10 Years Risk (SD)` ~ Mrsoo$`Market Cap`, data = Mrsoo)
Rs4.av <- aov(Rs4.lm)
summary(Rs4.av)
```

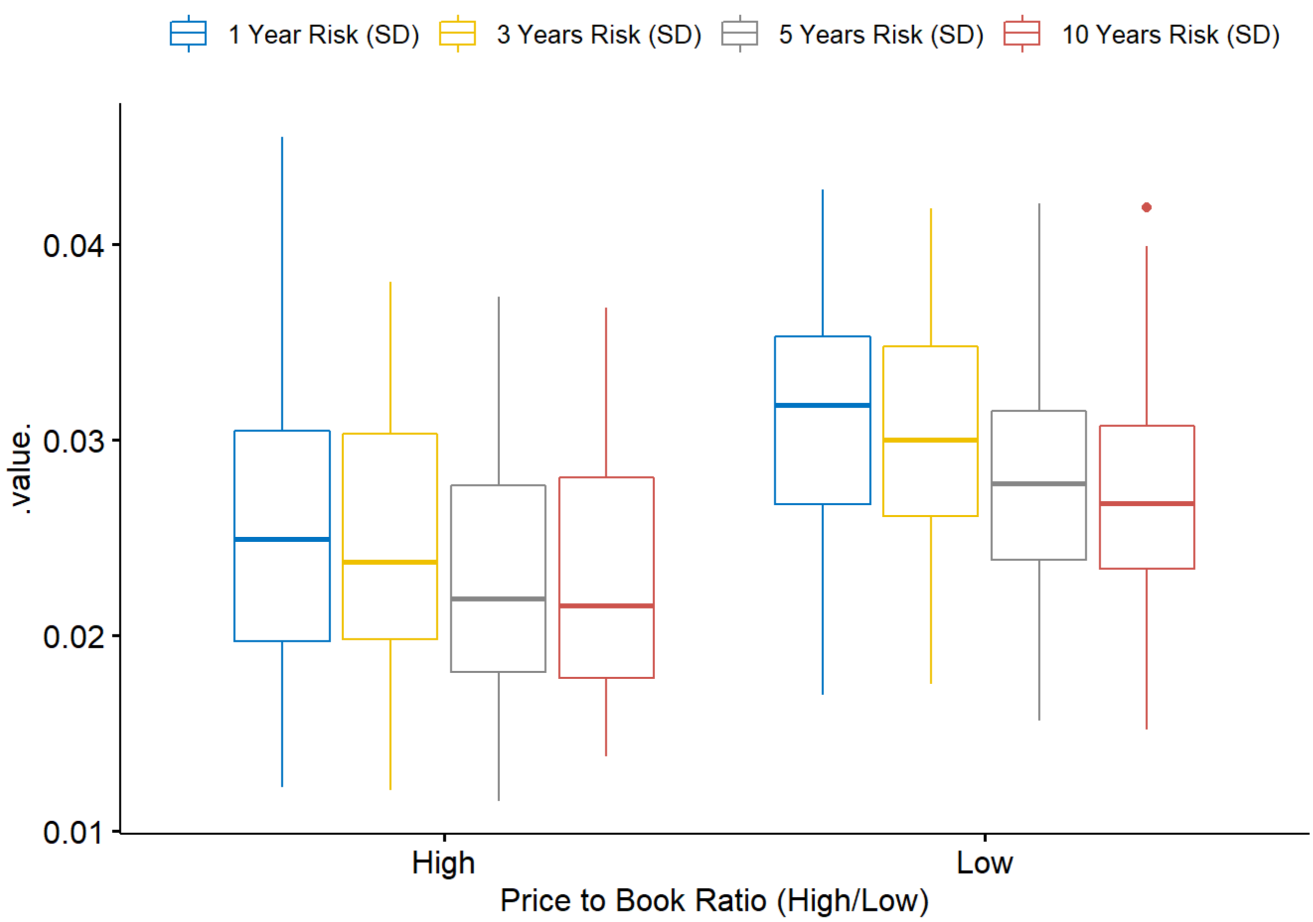
```
##                      Df   Sum Sq   Mean Sq F value Pr(>F)
## Mrsoo$`Market Cap`   2 0.002689 0.0013446   56.77 <2e-16 ***
## Residuals          141 0.003339 0.0000237
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis now we have to see from bornfer
onii intervals to know from where the difference is actually coming from
tukey.test <- TukeyHSD(Rs4.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs4.lm)
##
## $`Mrsoo$`Market Cap``
##                              diff          lwr         upr     p adj
## Mid Cap-Large Cap   0.002316560 6.653313e-05 0.004566586 0.0420135
## Small Cap-Large Cap 0.010613473 8.181514e-03 0.013045432 0.0000000
## Small Cap-Mid Cap   0.008296913 5.855061e-03 0.010738765 0.0000000
```

```
# for Mid Cap - Large Cap, Small Cap - Large Cap, and Small Cap-Mid cap  the diffe
rence of means are always positive

# final result :

# theres a significant difference for 1 year return risk between all the Market ca
ps, of which Small cap companies
# have higher 1 year return risk than Mid cap companies and for the Large cap comp
anies we have the least 1 year return risk

# Theres no significant difference between the 3 years return risk for Mid cap and
Large Cap companies
# while theres a significant difference for 3 years return risk between Small and
 Large cap comp and between Small and Mid cap comp
# again for small cap companies we have higher 3 year return risk among all, thoug
h Mid cap and large cap have no difference
```

```
# Theres no significant difference between the 5 years return risk for Mid cap and Large Cap companies
# while theres a significant difference for 5 years return risk between Small and  Large cap comp and between Small and Mid cap comp
# again, for small cap companies we have higher 5 year return risk among all, though Mid cap and large cap have no difference

# Theres a significant difference for 10 year return risk between all the Market caps, of which Small cap companies
# have higher 10 year return risk than Mid cap companies and for the Large cap companies we have the least 10 year return risk

### lets start with factor 2 PB ratio (Analysis):
## graph for visualization:
ggboxplot(
  Mrsoo, x = c("Price to Book Ratio (High/Low)"), y = c("1 Year Risk (SD)", "3 Years Risk (SD)","5 Years Risk (SD)","10 Years Risk (SD)"),
  merge = TRUE, palette = "jco"
)
```




```
Rs1_f2.lm <- lm(Mrsoo$`1 Year Risk (SD)` ~ Mrsoo$`Price to Book Ratio (High/Low)`, data = Mrsoo)
Rs1_f2.av <- aov(Rs1_f2.lm)
summary(Rs1_f2.av)
```

```
##                                      Df   Sum Sq   Mean Sq F value    Pr(>F)
## Mrsoo$`Price to Book Ratio (High/Low)`   1 0.001071 0.0010713   22.47 5.14e-06
## Residuals                           142 0.006771 0.0000477
```

```
##
## Mrsoo$`Price to Book Ratio (High/Low)` ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis,
tukey.test <- TukeyHSD(Rs1_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs1_f2.lm)
##
## $`Mrsoo$`Price to Book Ratio (High/Low)``
##                 diff         lwr        upr   p adj
## Low-High 0.005455209 0.003180078 0.00773034 5.1e-06
```

```
#low - high mean is always positive
#there is a significant difference for 1 year return risk between the companies wi
th high and low pb ratio
# companies with Low pb ratio have higher 1 year return risk as compared to other
 one

Rs2_f2.lm <- lm(Mrsoo$`3 Years Risk (SD)` ~ Mrsoo$`Price to Book Ratio (High/Low)
`, data = Mrsoo)
Rs2_f2.av <- aov(Rs2_f2.lm)
summary(Rs2_f2.av)
```

```
##                                          Df   Sum Sq  Mean Sq F value   Pr(>F)
## Mrsoo$`Price to Book Ratio (High/Low)`   1 0.000906 0.000906   21.59 7.61e-06
## Residuals                               142 0.005958 0.000042
##
## Mrsoo$`Price to Book Ratio (High/Low)` ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis,
tukey.test <- TukeyHSD(Rs2_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs2_f2.lm)
##
## $`Mrsoo$`Price to Book Ratio (High/Low)``
##                 diff         lwr         upr   p adj
## Low-High 0.005016666 0.002882616 0.007150716 7.6e-06
```

```
#low - high mean is always positive
#there is a significant difference for 3 year return risk between the companies wi
th high and low pb ratio
# companies with Low pb ratio have higher 1 year return risk as compared to other
 one

Rs3_f2.lm <- lm(Mrsoo$`5 Years Risk (SD)` ~ Mrsoo$`Price to Book Ratio (High/Low)
`, data = Mrsoo)
Rs3_f2.av <- aov(Rs3_f2.lm)
summary(Rs3_f2.av)
```

```
##                                          Df   Sum Sq  Mean Sq F value   Pr(>F)
## Mrsoo$`Price to Book Ratio (High/Low)`   1 0.000764 7.64e-04    18.4 3.29e-05
## Residuals                               142 0.005895 4.15e-05
##
## Mrsoo$`Price to Book Ratio (High/Low)` ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#p value less than 0.05 reject the null hypothesis,
tukey.test <- TukeyHSD(Rs3_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs3_f2.lm)
##
## $`Mrsoo$`Price to Book Ratio (High/Low)``
##                 diff         lwr         upr    p adj
## Low-High 0.004606674 0.002483944 0.006729403 3.29e-05
```

```
#low - high mean is always positive
#there is a significant difference for 5 year return risk between the companies wi
th high and low pb ratio
# companies with Low pb ratio have higher 1 year return risk as compared to other
 one

Rs4_f2.lm <- lm(Mrsoo$`10 Years Risk (SD)` ~ Mrsoo$`Price to Book Ratio (High/Low)
`, data = Mrsoo)
Rs4_f2.av <- aov(Rs4_f2.lm)
summary(Rs4_f2.av)
```

```
##                                          Df   Sum Sq  Mean Sq F value   Pr(>F)
## Mrsoo$`Price to Book Ratio (High/Low)`   1 0.000507 5.07e-04   13.04 0.000422
## Residuals                               142 0.005521 3.89e-05
##
## Mrsoo$`Price to Book Ratio (High/Low)` ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#p value less than 0.05 reject the null hypothesis,
tukey.test <- TukeyHSD(Rs4_f2.av)
tukey.test
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Rs4_f2.lm)
##
## $`Mrsoo$`Price to Book Ratio (High/Low)``
##                 diff         lwr         upr     p adj
## Low-High 0.003752745 0.001698295 0.005807196 0.0004223
```

```r
#low - high mean is always positive
#there is a significant difference for 5 year return risk between the companies wi
th high and low pb ratio
# companies with Low pb ratio have higher 1 year return risk as compared to other
 one
# finally one can say that irrespective of no. of years of return risk the compani
es with low PB ratio will have
# higher risk than the company with high PB ratio..

# inferences and final conclusion are on slides
```