Final Exam

Mike Kozlowski

2023-05-10

a. Load the data set, make sure all data types are correctly set

```
infile="C:\\Users\\Mike\\Documents\\DAT511\\Final\\justetf_alt.csv"
mydata=read.csv(infile)
str(mydata)
```

```
## 'data.frame':    2264 obs. of  24 variables:
##  $ isin               : chr  "IE00B0M62Y33" "IE00BMTX2B82" "NL0009272749" "IE000RN036E0" ...
##  $ wkn                : chr  "A0HGWF" "A2P9XA" "A1JN2C" "A3DGK2" ...
##  $ name               : chr  "iShares AEX UCITS ETF" "iShares AEX UCITS ETF EUR (Acc)" "VanEck AEX UCITS
ETF" "First Trust Alerian Disruptive Technology Real Estate UCITS ETF Acc" ...
##  $ fundProvider       : chr  "iShares" "iShares" "VanEck" "First Trust" ...
##  $ quote              : chr  "73.98" "6.68" "72.34" "17.44" ...
##  $ quote52Low         : num  62.8 5.8 63.2 17.1 10.4 ...
##  $ quote52High        : num  76.92 7.13 77.27 23.8 13.12 ...
##  $ ytdReturnCUR       : num  0.0541 0.0503 0.051 0.0029 -0.0599 ...
##  $ fees               : num  0.003 0.003 0.003 0.006 0.004 0.025 0 0.015 0.0035 0.0149 ...
##  $ yearVolatilityCUR  : num  0.188 0.191 0.19 NA 0.247 ...
##  $ fundCurrency       : chr  "EUR" "EUR" "EUR" "USD" ...
##  $ threeMonthReturnCUR: num  0.0415 0.0341 0.0345 0.0115 -0.0445 ...
##  $ monthReturnCUR     : num  -0.0488 -0.0552 -0.0544 -0.0708 -0.0908 ...
##  $ sixMonthReturnCUR  : num  0.1006 0.0915 0.092 -0.1012 -0.1214 ...
##  $ inceptionDate      : chr  "11/18/2005" "7/29/2020" "12/14/2009" "3/31/2022" ...
##  $ ticker             : chr  "IUSJ" "AYE7" "2TCA" "FTGT" ...
##  $ yearReturnCUR      : num  0.0336 0.0421 0.0453 NA 0.0258 ...
##  $ domicileCountry    : chr  "Ireland" "Ireland" "Netherlands" "Ireland" ...
##  $ weekReturnCUR      : num  -0.0037 -0.0277 -0.0277 0.0034 -0.0331 0.0388 0.0201 0.0364 -0.058 -0.0723
...
##  $ yearReturn1CUR     : num  -0.118 -0.118 -0.116 NA 0.263 ...
##  $ quoteDate          : chr  "3/21/2023" "3/17/2023" "3/17/2023" "3/21/2023" ...
##  $ currencyRisk       : chr  "Currency unhedged" "Currency unhedged" "Currency unhedged" "Currency unhed
ged" ...
##  $ UCITSCompliance    : chr  "Yes" "Yes" "Yes" "Yes" ...
##  $ securitiesLending  : chr  "Yes" "Yes" "No" "No" ...
```

```
mydata$fundProvider<-unlist(lapply(mydata$fundProvider,as.factor))
mydata$quote<-unlist(lapply(mydata$quote,as.numeric))
```

```
## Warning in lapply(mydata$quote, as.numeric): NAs introduced by coercion

## Warning in lapply(mydata$quote, as.numeric): NAs introduced by coercion

## Warning in lapply(mydata$quote, as.numeric): NAs introduced by coercion

## Warning in lapply(mydata$quote, as.numeric): NAs introduced by coercion
```

```
mydata$fundCurrency<-unlist(lapply(mydata$fundCurrency,as.factor))
mydata$domicileCountry<-unlist(lapply(mydata$domicileCountry,as.factor))
mydata$currencyRisk<-unlist(lapply(mydata$currencyRisk,as.factor))
mydata$UCITSCompliance<-ifelse(mydata$UCITSCompliance=="Yes",1,0)
mydata$UCITSCompliance<-unlist(lapply(mydata$UCITSCompliance,as.logical))
mydata$securitiesLending<-ifelse(mydata$securitiesLending=="Yes",1,0)
mydata$securitiesLending<-unlist(lapply(mydata$securitiesLending,as.logical))
str(mydata)
```

```
## 'data.frame':    2264 obs. of  24 variables:
##  $ isin               : chr  "IE00B0M62Y33" "IE00BMTX2B82" "NL0009272749" "IE000RN036E0" ...
##  $ wkn                : chr  "A0HGWF" "A2P9XA" "A1JN2C" "A3DGK2" ...
##  $ name               : chr  "iShares AEX UCITS ETF" "iShares AEX UCITS ETF EUR (Acc)" "VanEck AEX UCITS
ETF" "First Trust Alerian Disruptive Technology Real Estate UCITS ETF Acc" ...
##  $ fundProvider       : Factor w/ 40 levels "iShares","VanEck",..: 1 1 2 3 4 5 6 2 2 7 ...
##  $ quote              : num  73.98 6.68 72.34 17.44 10.7 ...
##  $ quote52Low         : num  62.8 5.8 63.2 17.1 10.4 ...
##  $ quote52High        : num  76.92 7.13 77.27 23.8 13.12 ...
##  $ ytdReturnCUR       : num  0.0541 0.0503 0.051 0.0029 -0.0599 ...
##  $ fees               : num  0.003 0.003 0.003 0.006 0.004 0.025 0 0.015 0.0035 0.0149 ...
##  $ yearVolatilityCUR  : num  0.188 0.191 0.19 NA 0.247 ...
##  $ fundCurrency       : Factor w/ 14 levels "EUR","USD","USD Hedged",..: 1 1 1 2 2 2 2 2 1 2 ...
##  $ threeMonthReturnCUR: num  0.0415 0.0341 0.0345 0.0115 -0.0445 ...
##  $ monthReturnCUR     : num  -0.0488 -0.0552 -0.0544 -0.0708 -0.0908 ...
##  $ sixMonthReturnCUR  : num  0.1006 0.0915 0.092 -0.1012 -0.1214 ...
##  $ inceptionDate      : chr  "11/18/2005" "7/29/2020" "12/14/2009" "3/31/2022" ...
##  $ ticker             : chr  "IUSJ" "AYE7" "2TCA" "FTGT" ...
##  $ yearReturnCUR      : num  0.0336 0.0421 0.0453 NA 0.0258 ...
##  $ domicileCountry    : Factor w/ 9 levels "Ireland","Netherlands",..: 1 1 2 1 1 3 4 5 2 5 ...
##  $ weekReturnCUR      : num  -0.0037 -0.0277 -0.0277 0.0034 -0.0331 0.0388 0.0201 0.0364 -0.058 -0.0723
...
##  $ yearReturn1CUR     : num  -0.118 -0.118 -0.116 NA 0.263 ...
##  $ quoteDate          : chr  "3/21/2023" "3/17/2023" "3/17/2023" "3/21/2023" ...
##  $ currencyRisk       : Factor w/ 2 levels "Currency unhedged",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ UCITSCompliance    : logi  TRUE TRUE TRUE TRUE TRUE FALSE ...
##  $ securitiesLending  : logi  TRUE TRUE FALSE FALSE FALSE FALSE ...
```

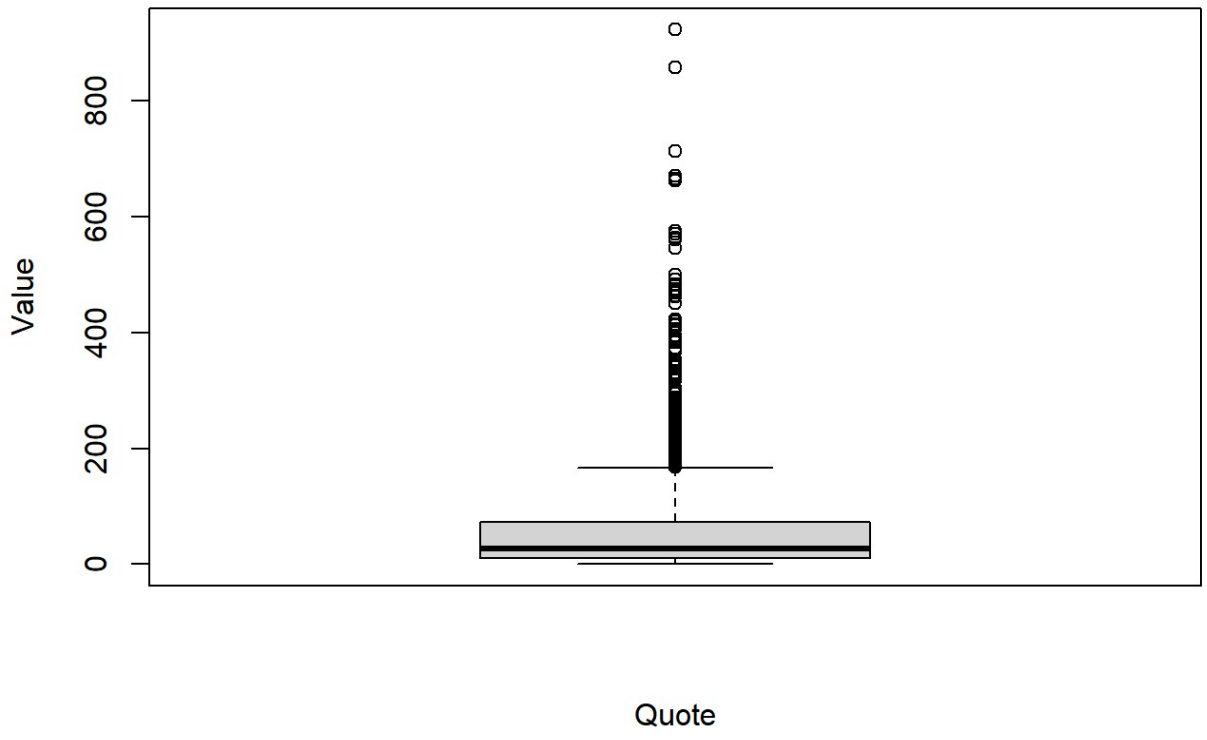
b. Do a basic data summary

```
summary(mydata)
```

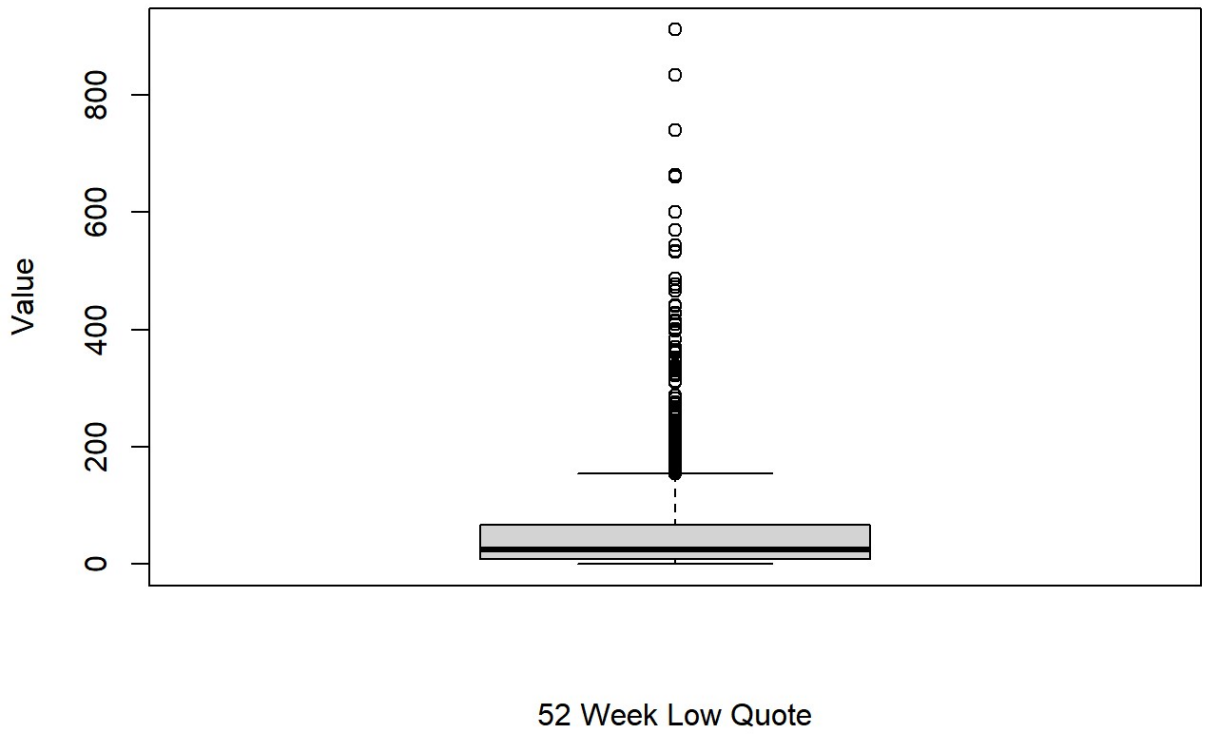
```
##      isin               wkn                name               fundProvider
##  Length:2264        Length:2264        Length:2264        iShares   :490
##  Class :character   Class :character   Class :character   Xtrackers :283
##  Mode  :character   Mode  :character   Mode  :character   Amundi ETF:243
##                                                           Lyxor ETF :229
##                                                           Invesco   :133
##                                                           UBS ETF   :132
##                                                           (Other)   :754
##      quote          quote52Low       quote52High      ytdReturnCUR
##  Min.   :  0.34   Min.   :  0.000   Min.   :  0.52   Min.   :-0.22020
##  1st Qu.:  9.90   1st Qu.:  9.117   1st Qu.: 11.40   1st Qu.: 0.00000
##  Median : 26.42   Median : 24.430   Median : 30.74   Median : 0.02330
##  Mean   : 60.37   Mean   : 55.540   Mean   : 69.09   Mean   : 0.03886
##  3rd Qu.: 72.66   3rd Qu.: 67.442   3rd Qu.: 81.46   3rd Qu.: 0.04520
##  Max.   :923.00   Max.   :911.170   Max.   :965.82   Max.   : 1.46380
##  NA's   :4                                           NA's   :31
##       fees          yearVolatilityCUR     fundCurrency  threeMonthReturnCUR
##  Min.   :0.000000   Min.   :0.0011    USD       :1071   Min.   :-0.20040
##  1st Qu.:0.001500   1st Qu.:0.1383    EUR       : 855   1st Qu.:-0.00150
##  Median :0.002500   Median :0.1849    EUR Hedged: 205   Median : 0.01600
##  Mean   :0.003261   Mean   :0.2025    GBP       :  40   Mean   : 0.03249
##  3rd Qu.:0.004000   3rd Qu.:0.2359    USD Hedged:  30   3rd Qu.: 0.04260
##  Max.   :0.025000   Max.   :1.9948    JPY       :  23   Max.   : 0.90850
##                     NA's   :226       (Other)   :  40   NA's   :35
##  monthReturnCUR     sixMonthReturnCUR  inceptionDate         ticker
##  Min.   :-0.29100   Min.   :-0.48250   Length:2264        Length:2264
##  1st Qu.:-0.05750   1st Qu.:-0.04485   Class :character   Class :character
##  Median :-0.03620   Median :-0.00150   Mode  :character   Mode  :character
##  Mean   :-0.03337   Mean   : 0.00599
##  3rd Qu.:-0.00140   3rd Qu.: 0.04210
##  Max.   : 0.11720   Max.   : 0.45040
##  NA's   :13         NA's   :77
##  yearReturnCUR         domicileCountry weekReturnCUR       yearReturn1CUR
##  Min.   :-0.79370   Ireland    :1266   Min.   :-0.129300   Min.   :-0.94380
##  1st Qu.:-0.10960   Luxembourg : 705   1st Qu.:-0.011800   1st Qu.:-0.17520
##  Median :-0.07190   Germany    : 136   Median : 0.000800   Median :-0.12990
##  Mean   :-0.07585   France     :  55   Mean   : 0.001726   Mean   :-0.12590
##  3rd Qu.:-0.01720   Jersey     :  37   3rd Qu.: 0.012300   3rd Qu.:-0.06405
##  Max.   : 0.76820   Switzerland:  28   Max.   : 0.156700   Max.   : 1.04420
##  NA's   :228        (Other)    :  37   NA's   :5           NA's   :273
##   quoteDate                   currencyRisk  UCITSCompliance securitiesLending
##  Length:2264        Currency unhedged:2005   Mode :logical   Mode :logical
##  Class :character   Currency hedged  : 259   FALSE:113       FALSE:1562
##  Mode  :character                            TRUE :2151      TRUE :702
##
##
##
##
```

c. Show boxplots of the basic data

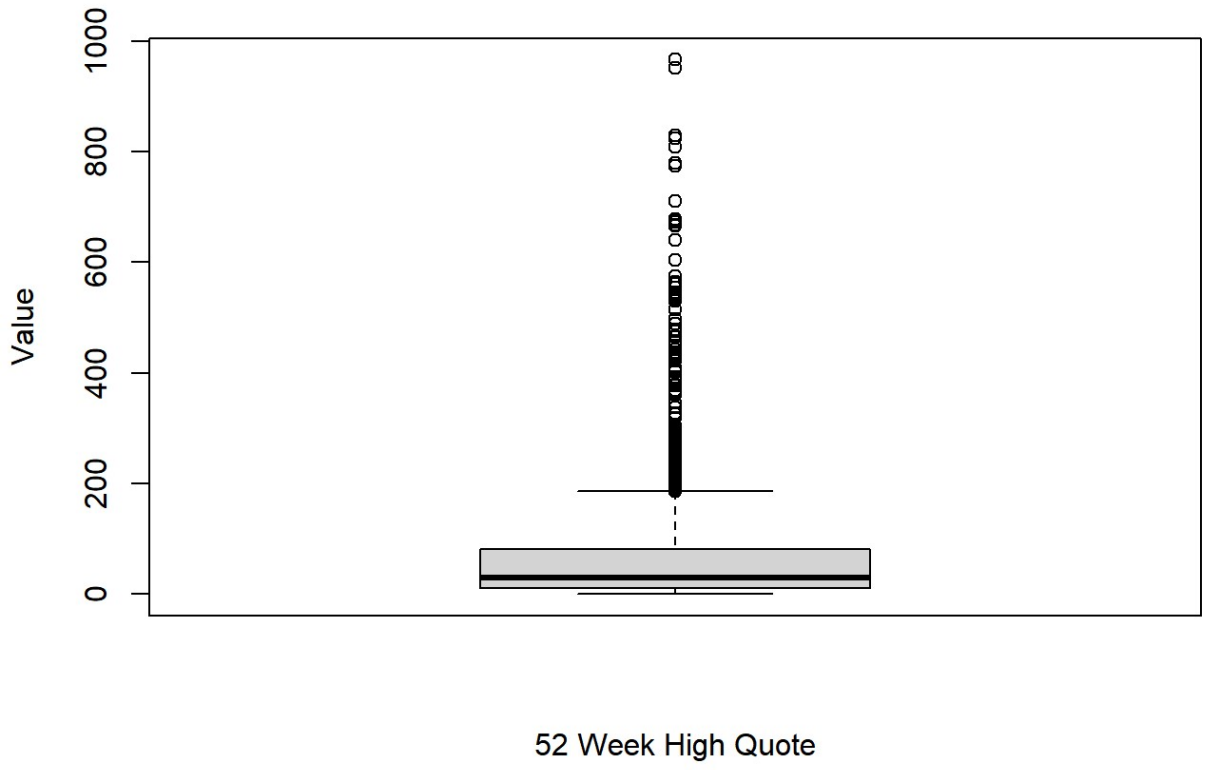
```
boxplot(mydata$quote, xlab="Quote", ylab="Value")
```



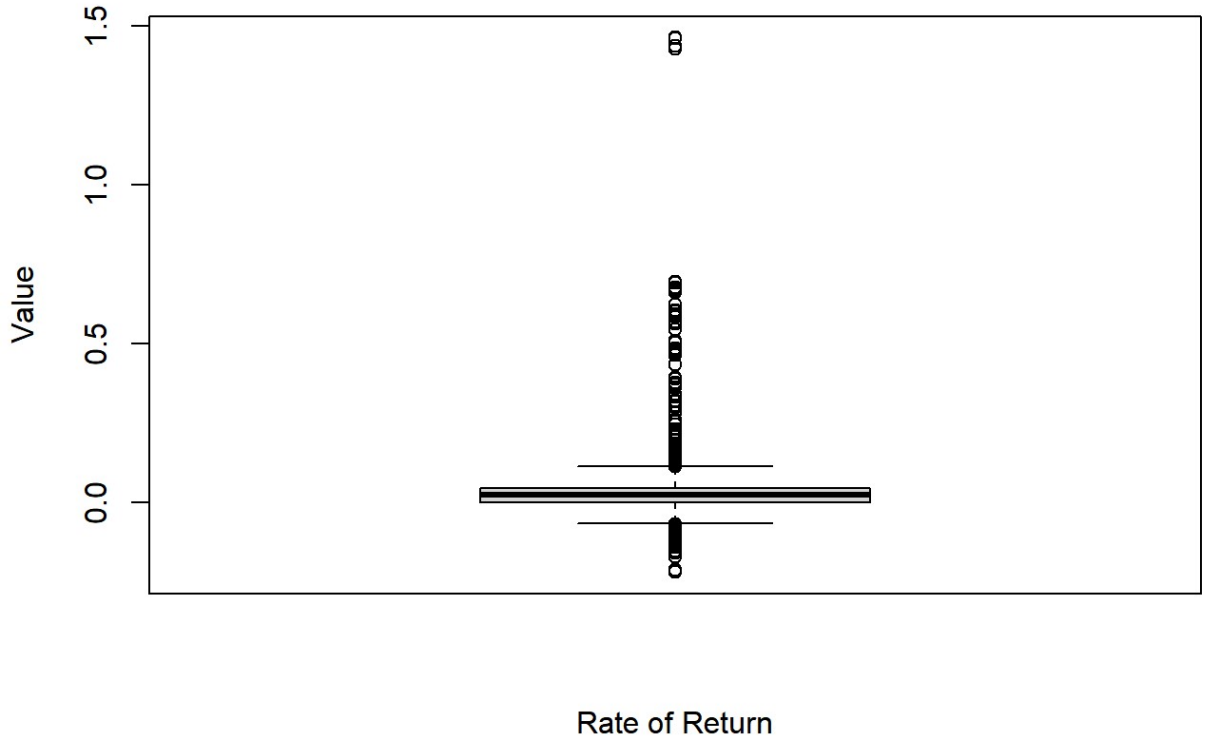

```
boxplot(mydata$quote52Low, xlab="52 Week Low Quote", ylab="Value")
```



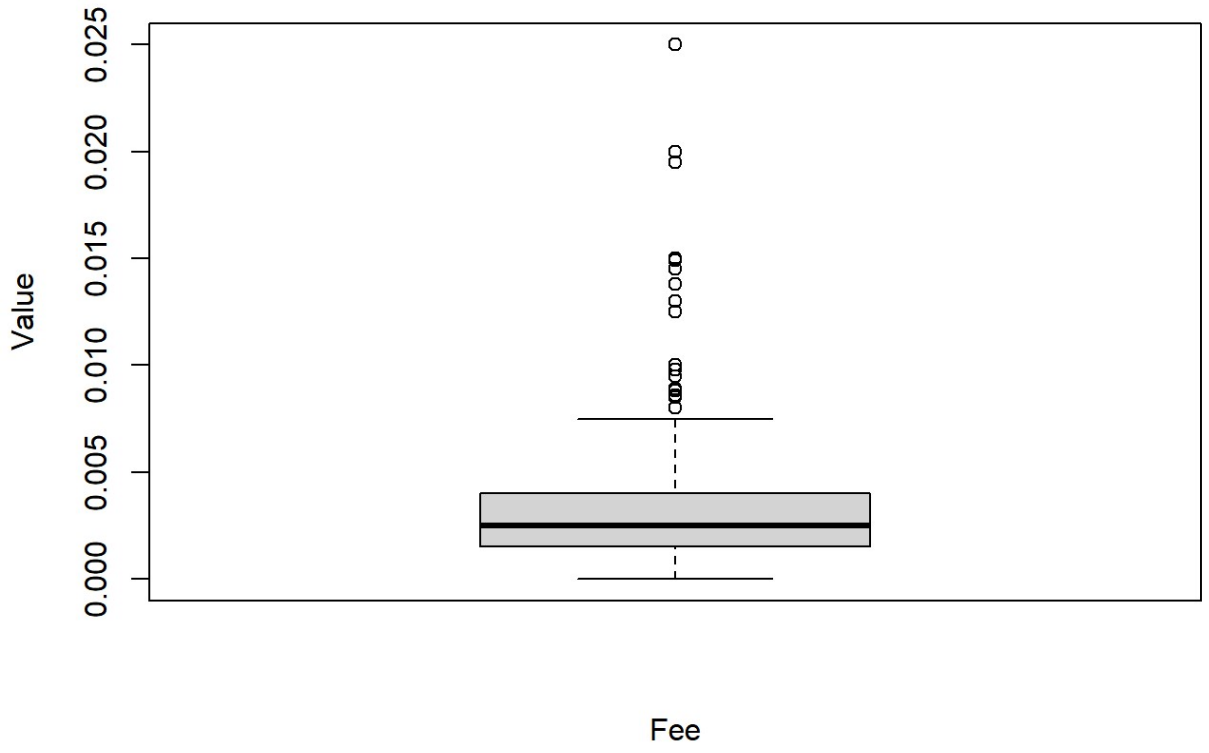

```
boxplot(mydata$quote52High, xlab="52 Week High Quote", ylab="Value")
```



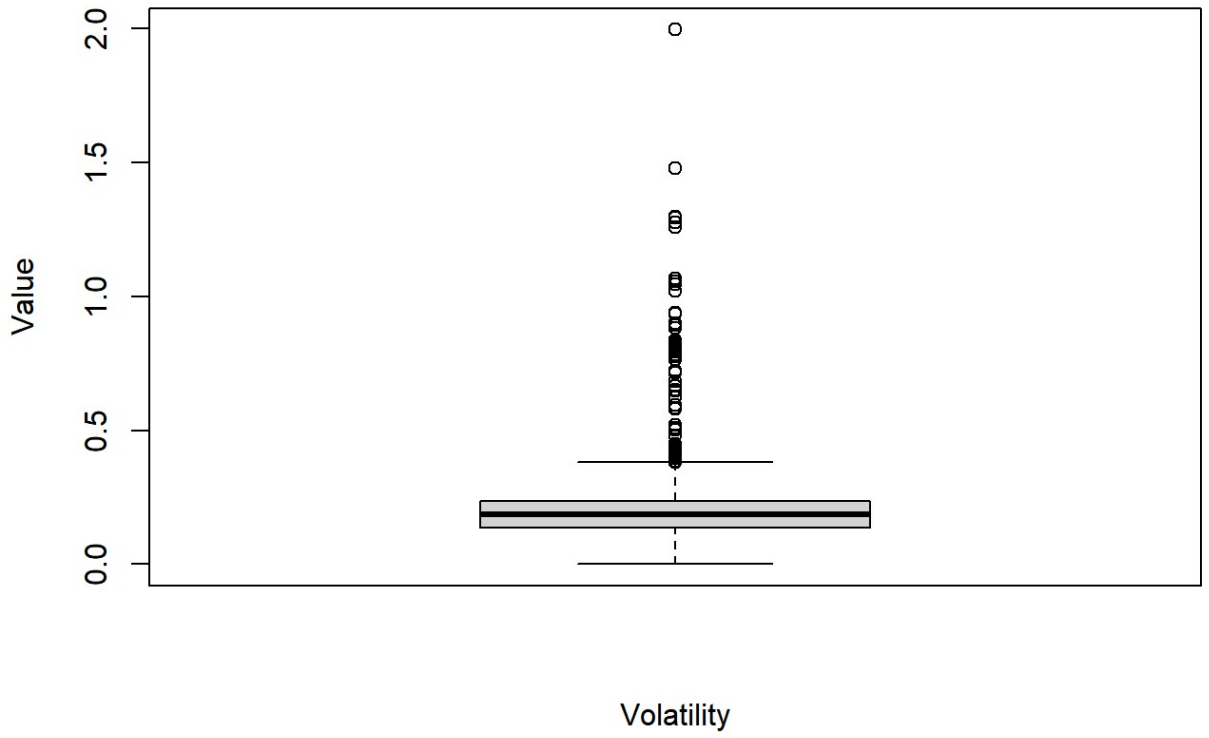

```
boxplot(mydata$ytdReturnCUR, xlab="Rate of Return", ylab="Value")
```



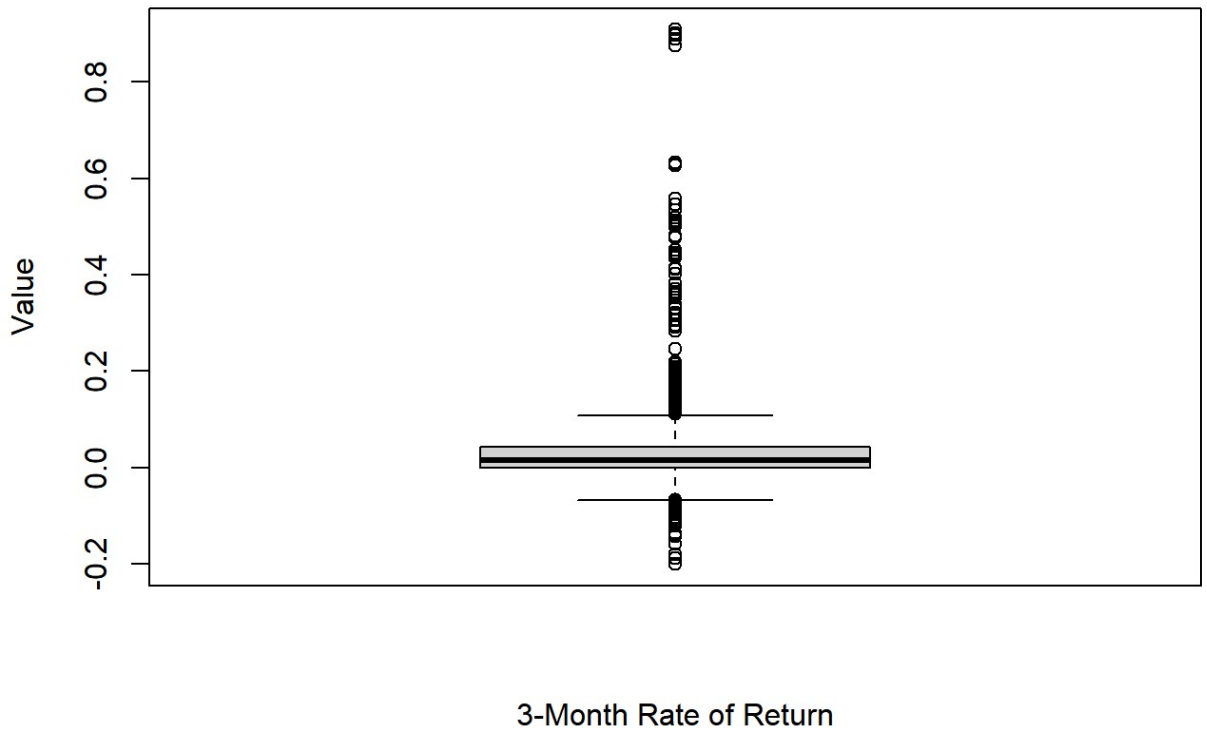

```
boxplot(mydata$fees, xlab="Fee", ylab="Value")
```



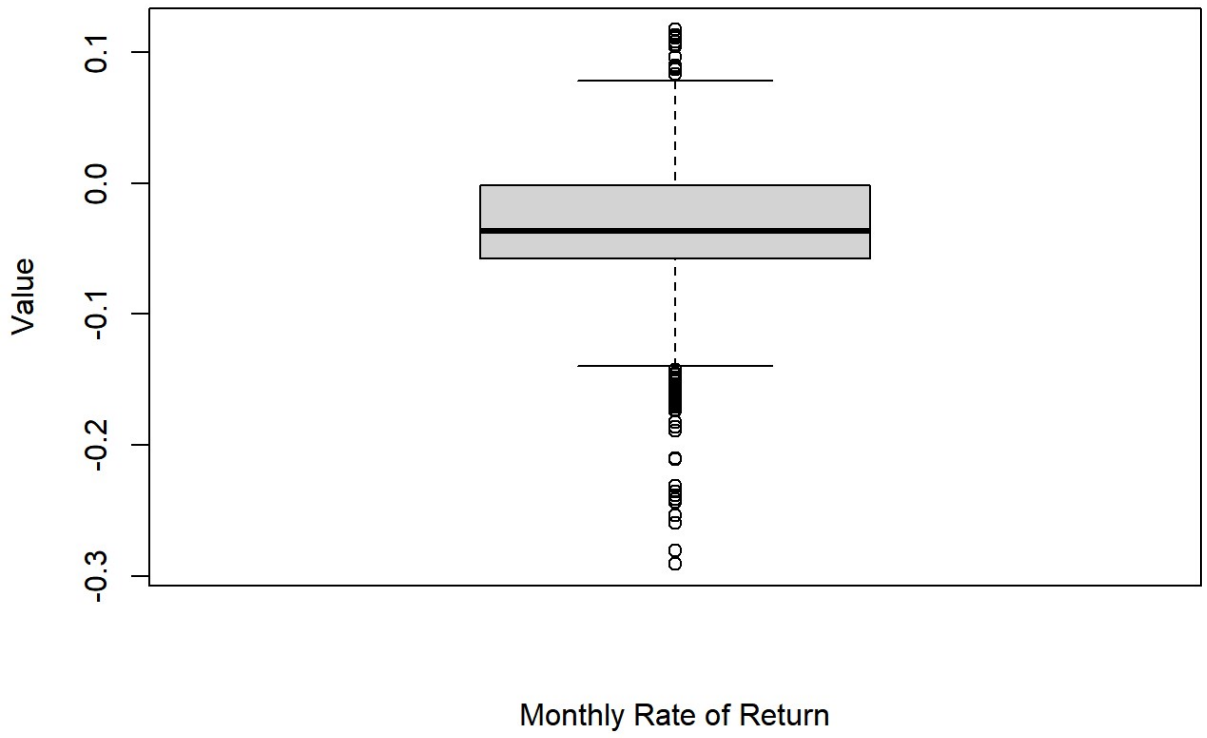

```
boxplot(mydata$yearVolatilityCUR, xlab="Volatility", ylab="Value")
```



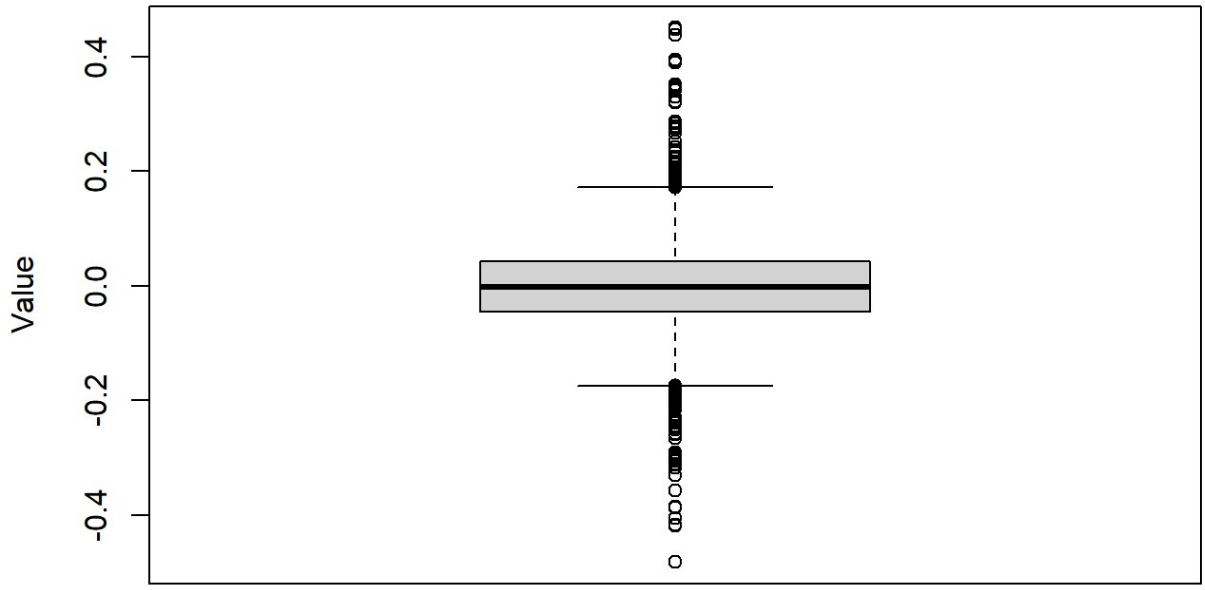

```
boxplot(mydata$threeMonthReturnCUR, xlab="3-Month Rate of Return", ylab="Value")
```




```
boxplot(mydata$monthReturnCUR, xlab="Monthly Rate of Return", ylab="Value")
```

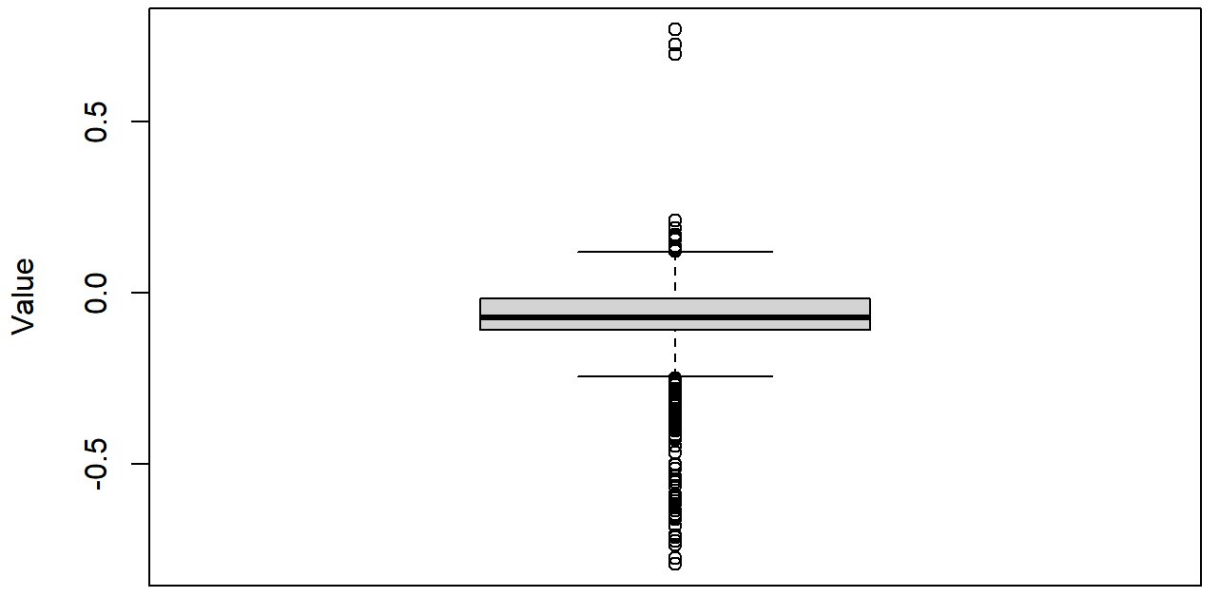



```
boxplot(mydata$sixMonthReturnCUR, xlab="6-Month Rate of Return", ylab="Value")
```



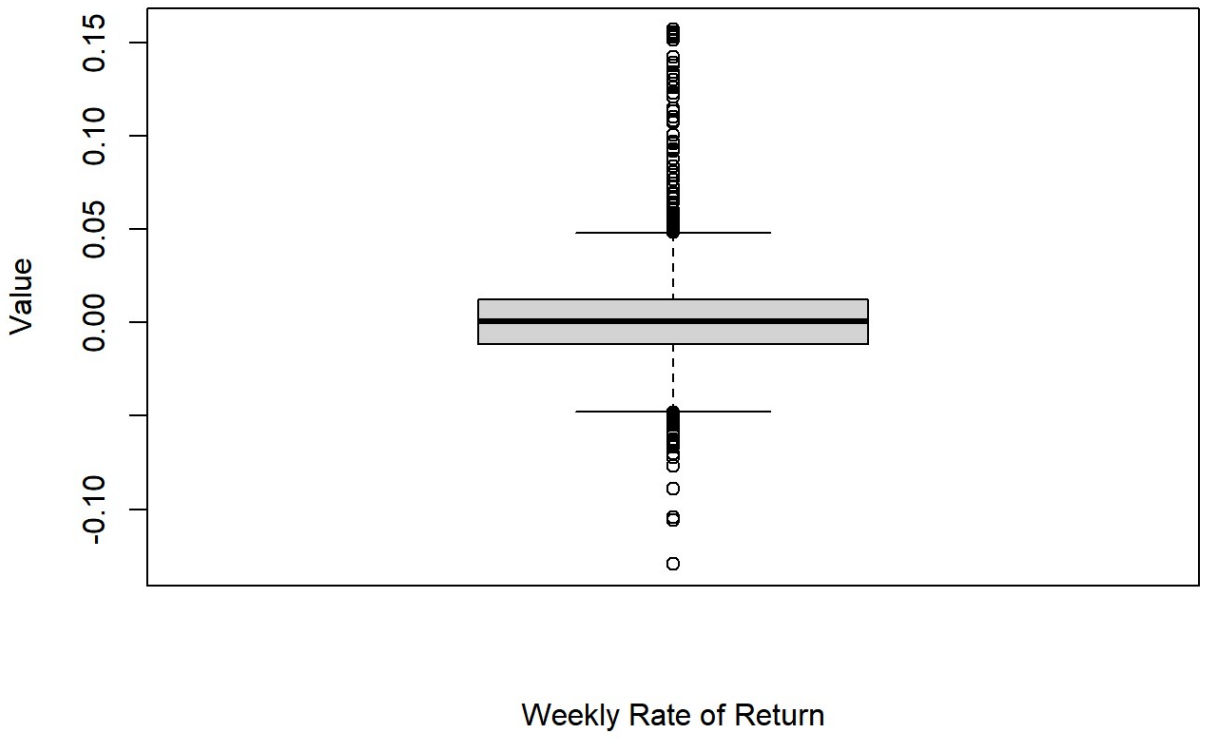
6-Month Rate of Return

```
boxplot(mydata$yearReturnCUR, xlab="Yearly Rate of Return", ylab="Value")
```

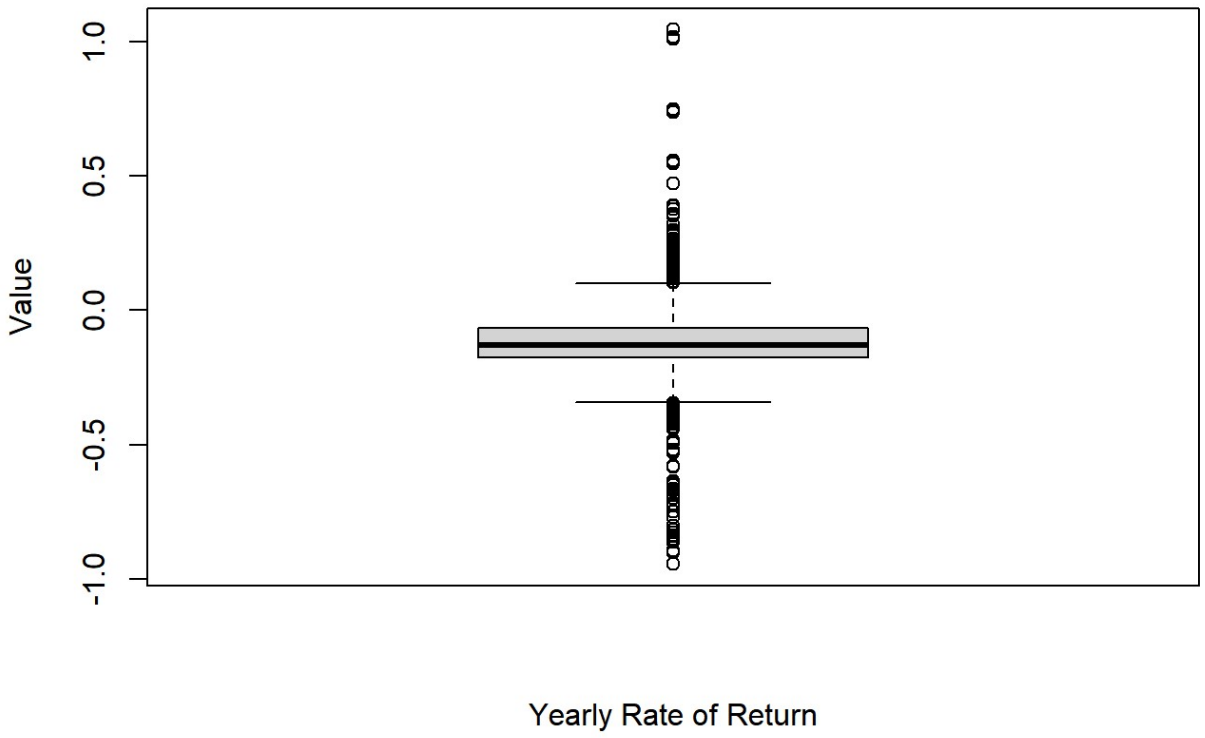


Yearly Rate of Return

```
boxplot(mydata$weekReturnCUR, xlab="Weekly Rate of Return", ylab="Value")
```

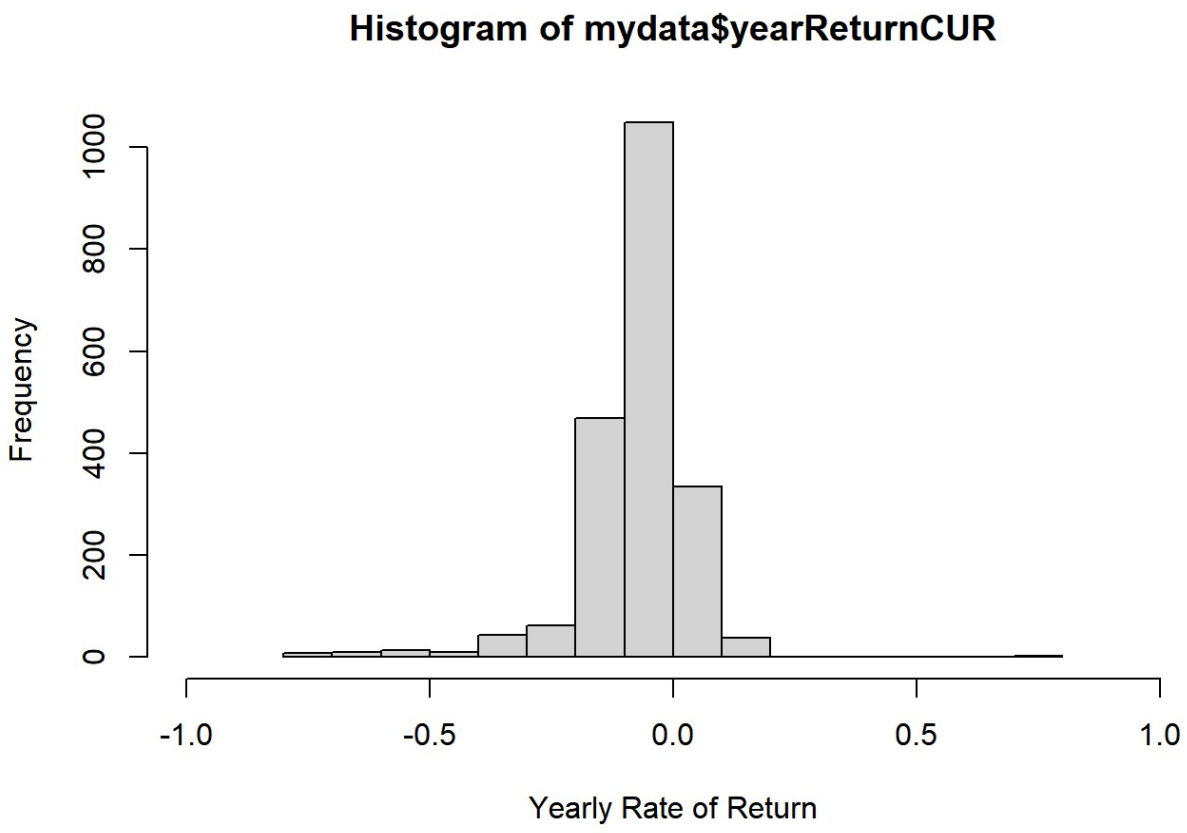



```
boxplot(mydata$yearReturn1CUR, xlab="Yearly Rate of Return", ylab="Value")
```



d. Show a histogram of one variable

```
hist(mydata$yearReturnCUR,xlim=c(-1,1),xlab="Yearly Rate of Return")
```



e. Create a cross-table (aka Pivot Table) showing the number of funds per several other factor variables (fund provider, domicile, fund currency, etc)

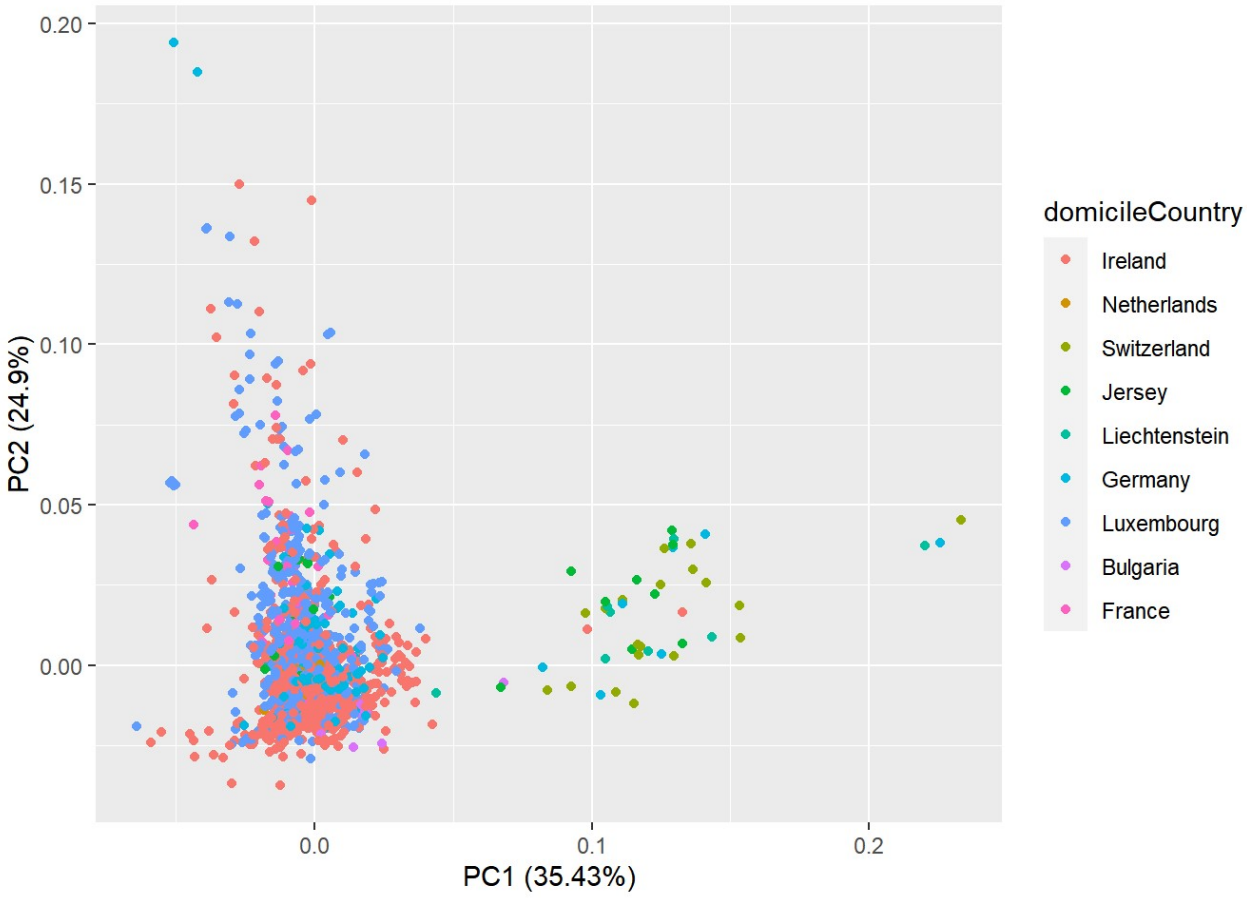
```
pt <- PivotTable$new()
pt$addData(mydata)
pt$addColumnDataGroups("fundCurrency")
pt$addRowDataGroups("fundProvider")
pt$defineCalculation(calculationName="TotalFunds", summariseExpression="n()")
pt$evaluatePivot()
pt$renderPivot()
```

	EUR	USD	USD Hedged	EUR Hedged	GBP Hedged	GBP	CHF Hedged	CHF	AUD	JPY	HKD	CAD	SEK	BGN	Total
iShares	162	243	11	48	3	13	6	1		2			1		490
VanEck	12	31													43
First Trust	2	13				1									16
HANetf	1	23													24
21Shares		22													22
CoinShares		13													13
Deutsche Digital Assets		4													4
Xtrackers	85	132	7	37	3	6	4	3	2	4					283
ETC Group		10													10
Roundhill		1													1
Legal & General (LGIM)	4	36		1											41
expatcapital	10													1	11
Global X		33													33
SEBA Bank		4													4
WisdomTree	10	41	2	6											59

Amundi ETF	151	59	4	20	1	1		1		6					243
J.P. Morgan	2	12		4											18
SPDR ETF	31	58	4	10	1	8				2					114
Invesco	45	71		15	1					1					133
Vanguard	21	31		9		3									64
UBS ETF	26	62		31	2	2	1	3		4		1			132
Lyxor ETF	151	54	2	12	1	5	1	1		2					229
BNP Paribas Easy	63	14		5											82
HSBC ETF	7	42		1		1					1				52
Tabula	3			2											5
PIMCO	3	4		1											8
Deka ETFs	46	5								2	1				54
Credit Suisse	1	7		1											9
Fidelity ETF	2	10		2											14
Rize ETF		8													8
FlexShares		5													5
Axxion	1														1
Franklin Templeton	3	15													18
Goldman Sachs		4													4
Boerse Stuttgart	3														3
Deutsche Boerse	1														1
Ossiam	6	1													7
AXA IM		1													1
Hashdex		2													2
Market Access	3														3
Total	855	1071	30	205	12	40	12	9	2	23	2	1	1	1	2264

f. Show PCA plot, using all the real number variables. Be ready to color code the PCA plot using one of the factors

```
mydata.pca <- prcomp(na.omit(mydata[,c(5,6,7,8,9,10,12,13,14,17,19,20)]), center=TRUE, scale=TRUE)
mydata.pca.plot <- autoplot(mydata.pca, data=na.omit(mydata), colour='domicileCountry')
mydata.pca.plot
```



g. Use a validator structure to carry out a rule-based validation

```
#rules <- validator(<will be specified during exam>)
```

h. Run an error localization using some basic rules

```
#will be specified during exam
```

i. Be ready to use the functional relation rule

```
#will be specified during exam
```

j. Do two error corrections, one using a typographic correction, and the other being a rule-based correction

```
#v<-validator(<will be specified during exam>)
#correct_typos(mydata,v)
```

k. Do an imputation on one variable using a regression model or a random forest model

```
mydata.imp=impute_lm(mydata, quote ~ quote52Low + quote52High)
head(mydata.imp)
```

```
##          isin    wkn
## 1 IE00B0M62Y33 A0HGWF
## 2 IE00BMTX2B82 A2P9XA
## 3 NL0009272749 A1JN2C
## 4 IE000RN036E0 A3DGK2
## 5 IE00BKPTXQ89 A2P4PH
## 6 CH1146882316 A3GVVU
##                                                                  name
## 1                                              iShares AEX UCITS ETF
## 2                                   iShares AEX UCITS ETF EUR (Acc)
## 3                                               VanEck AEX UCITS ETF
## 4 First Trust Alerian Disruptive Technology Real Estate UCITS ETF Acc
## 5                  HANetf Alerian Midstream Energy Dividend UCITS ETF
## 6                                              21Shares Algorand ETP
##   fundProvider quote quote52Low quote52High ytdReturnCUR  fees
## 1      iShares 73.98      62.81       76.92       0.0541 0.003
## 2      iShares  6.68       5.80        7.13       0.0503 0.003
## 3       VanEck 72.34      63.16       77.27       0.0510 0.003
## 4  First Trust 17.44      17.07       23.80       0.0029 0.006
## 5       HANetf 10.70      10.39       13.12      -0.0599 0.004
## 6     21Shares  2.14       1.64        9.68       0.3049 0.025
##   yearVolatilityCUR fundCurrency threeMonthReturnCUR monthReturnCUR
## 1            0.1879          EUR              0.0415        -0.0488
## 2            0.1906          EUR              0.0341        -0.0552
## 3            0.1898          EUR              0.0345        -0.0544
## 4                NA          USD              0.0115        -0.0708
## 5            0.2473          USD             -0.0445        -0.0908
## 6            0.8803          USD              0.1889        -0.2438
##   sixMonthReturnCUR inceptionDate ticker yearReturnCUR domicileCountry
## 1            0.1006    11/18/2005   IUSJ        0.0336         Ireland
## 2            0.0915     7/29/2020   AYE7        0.0421         Ireland
## 3            0.0920    12/14/2009   2TCA        0.0453     Netherlands
## 4           -0.1012     3/31/2022   FTGT            NA         Ireland
## 5           -0.1214     7/27/2020   JMLP        0.0258         Ireland
## 6           -0.4185    11/22/2021   ALC0       -0.7128     Switzerland
##   weekReturnCUR yearReturn1CUR quoteDate      currencyRisk UCITSCompliance
## 1       -0.0037        -0.1178 3/21/2023 Currency unhedged            TRUE
## 2       -0.0277        -0.1179 3/17/2023 Currency unhedged            TRUE
## 3       -0.0277        -0.1162 3/17/2023 Currency unhedged            TRUE
## 4        0.0034             NA 3/21/2023 Currency unhedged            TRUE
## 5       -0.0331         0.2626 3/21/2023 Currency unhedged            TRUE
## 6        0.0388        -0.9014 3/21/2023 Currency unhedged           FALSE
##   securitiesLending
## 1              TRUE
## 2              TRUE
## 3             FALSE
## 4             FALSE
## 5             FALSE
## 6             FALSE
```