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_mod.csv"
mydata=read.csv(infile)
str(mydata)
```

```
## 'data.frame':    2264 obs. of  25 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" ...
##  $ quoteRange         : num  14.11 1.33 14.11 6.73 2.73 ...
```

```
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  25 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/ 15 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/ 10 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 ...
##  $ quoteRange         : num  14.11 1.33 14.11 6.73 2.73 ...
```

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.   :-25.350   Min.   :  0.000   Min.   :  0.52   Min.   :-128.2000
## 1st Qu.:  9.898   1st Qu.:  9.117   1st Qu.: 11.40   1st Qu.:   0.0000
## Median : 26.420   Median : 24.430   Median : 30.74   Median :   0.0233
## Mean   : 60.347   Mean   : 55.540   Mean   : 69.09   Mean   :  -0.0185
## 3rd Qu.: 72.657   3rd Qu.: 67.442   3rd Qu.: 81.46   3rd Qu.:   0.0452
## Max.   :923.000   Max.   :911.170   Max.   :965.82   Max.   :   1.4638
## 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      : 854   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)  :  41   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    :1265   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)    :  38   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
##
##
##
##
##    quoteRange
## Min.   :-66.93
## 1st Qu.:  2.12
## Median :  5.77
## Mean   : 13.55
## 3rd Qu.: 14.47
## Max.   :422.99
##
```

I'm not sure what the fields "isin" and "wkn" are, but I assume they are related to some kind of designation on the stock market. Name is fairly obvious, it is the name of the fund, with the ticker being the ticker that represents that fund on the stock market, and the fund provider is the name of the company that manages that fund. Quote is the quote price to own a share on that particular day listed in quoteDate, while quote52Low and quote52High represent the lowest and highest quote price in the previous 52 weeks. ytdReturnCUR, threeMonthReturnCUR, monthReturnCUR, sixMonthReturnCUR, yearReturnCUR, weekReturnCUR, yearReturn1CUR represent different time interals for the rate of return on the funds. Fees represent the maintenance/management fee that the fund managers take as part of managing which stocks are part of the fund. fundCurrency shows what currency each fund is traded in, while domicileCountry represents where the fund is managed from. inceptionDate shows when the fund was started, currencyRisk shows whether the fund is hedged against losses or not, UCITSCompliance and SecuritiesLending are boolean values that show if that fund lends out securities or if they comply with UCITS.

Question: What column was added?

The quoteRange column was added, I don't recall it from the earlier version of this file.

Question: Do you see any outliers or obvious errors?

Both quote and quoteRange contain negative numbers. While it's normal for a field like rate of return to contain negative numbers, the quote price for a share of an ETF should never be a negative number; you will never get paid to gain ownership shares of an ETF.

c. Show boxplots of the basic data

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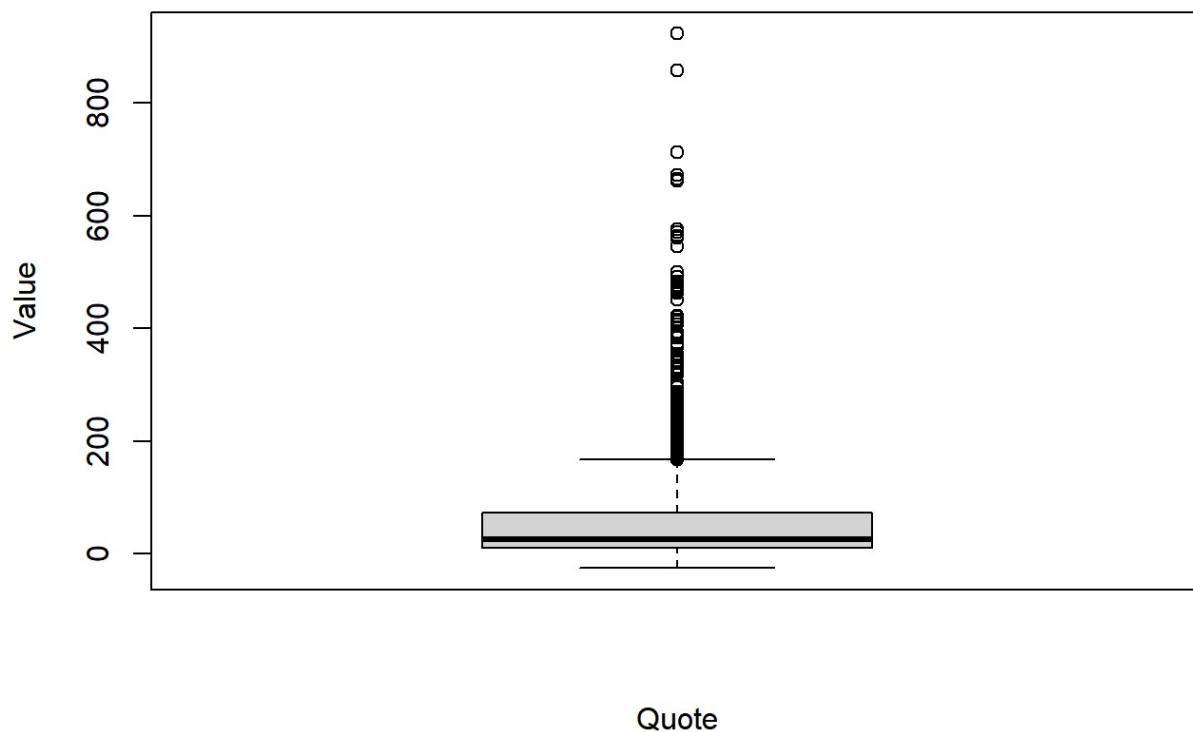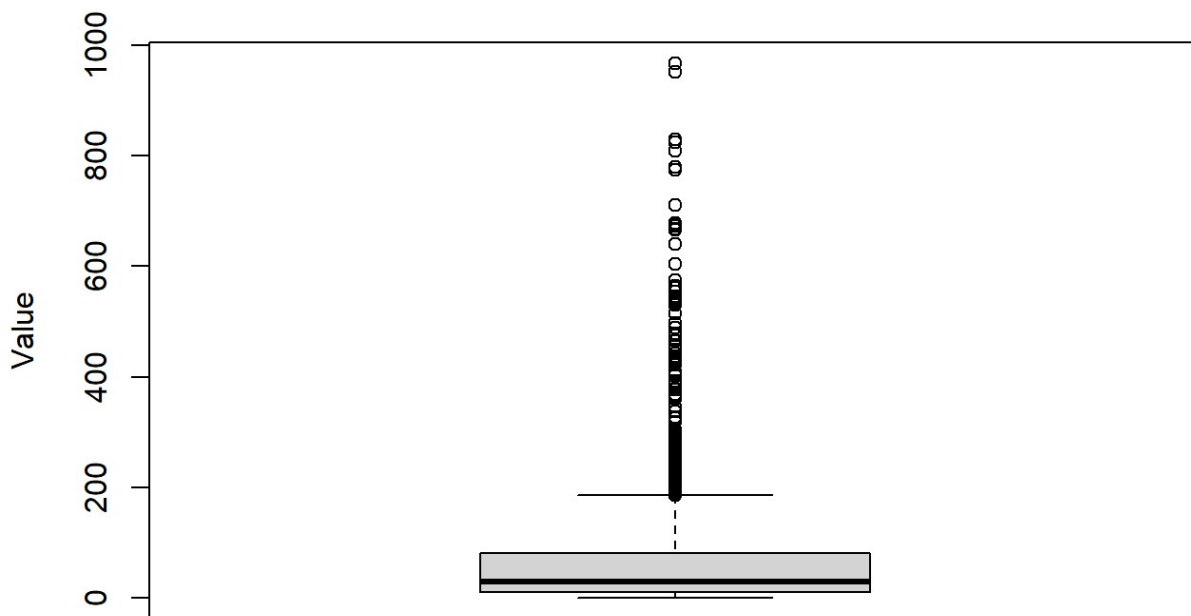


```
boxplot(mydata$quote52Low, xlab="52 Week Low Quote", ylab="Value")
```
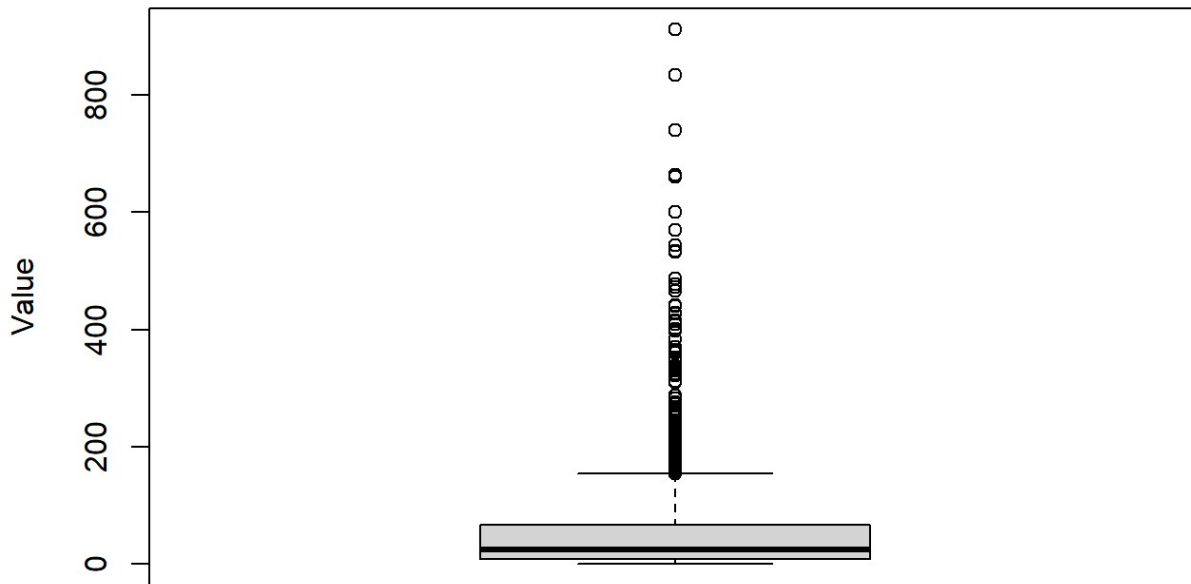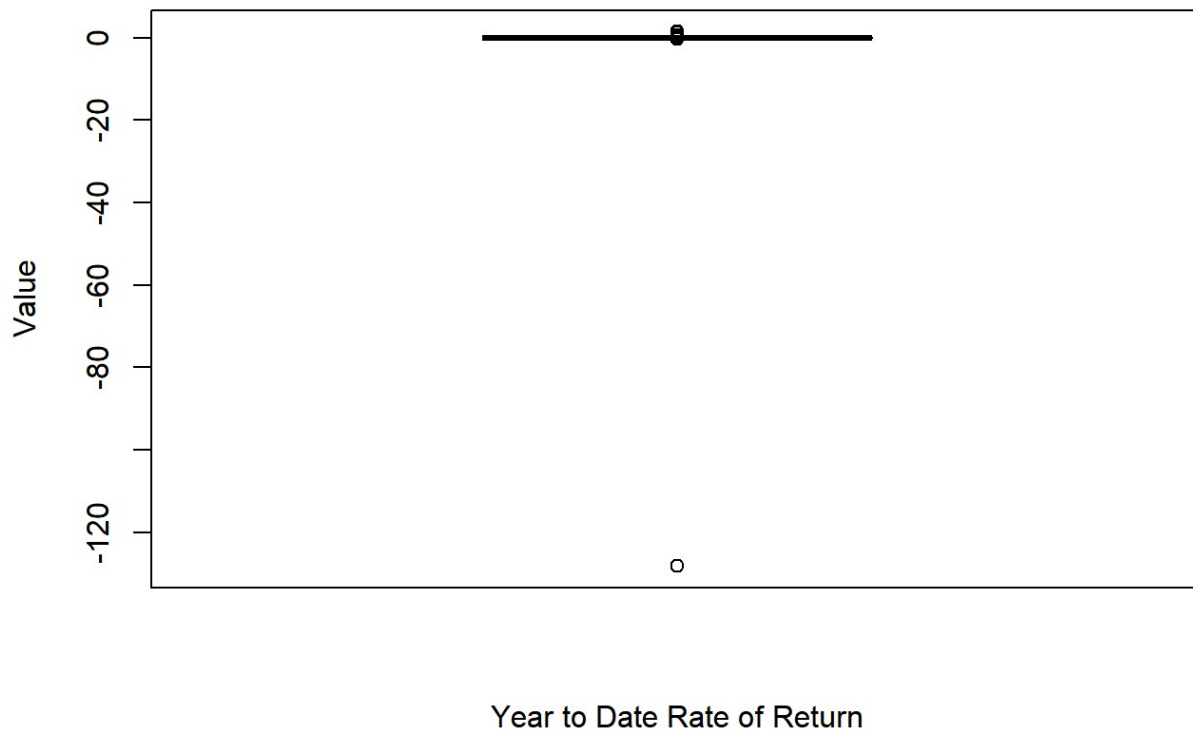
52 Week Low Quote

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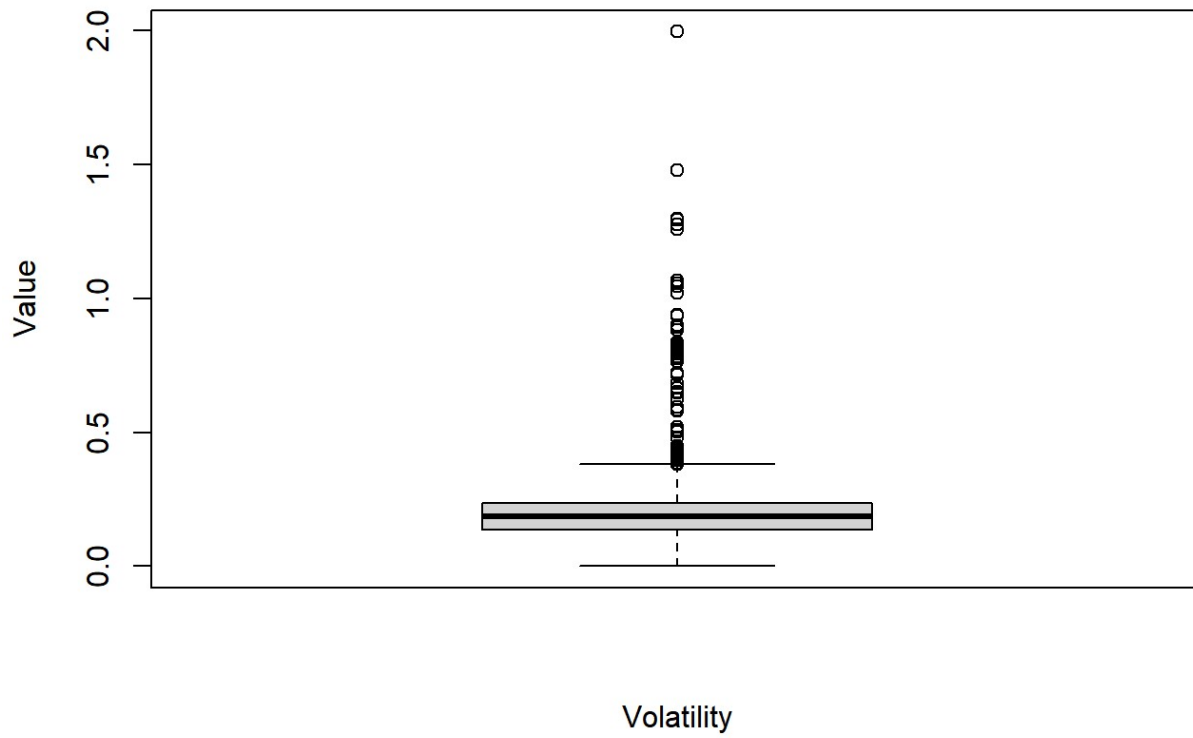

52 Week High Quote

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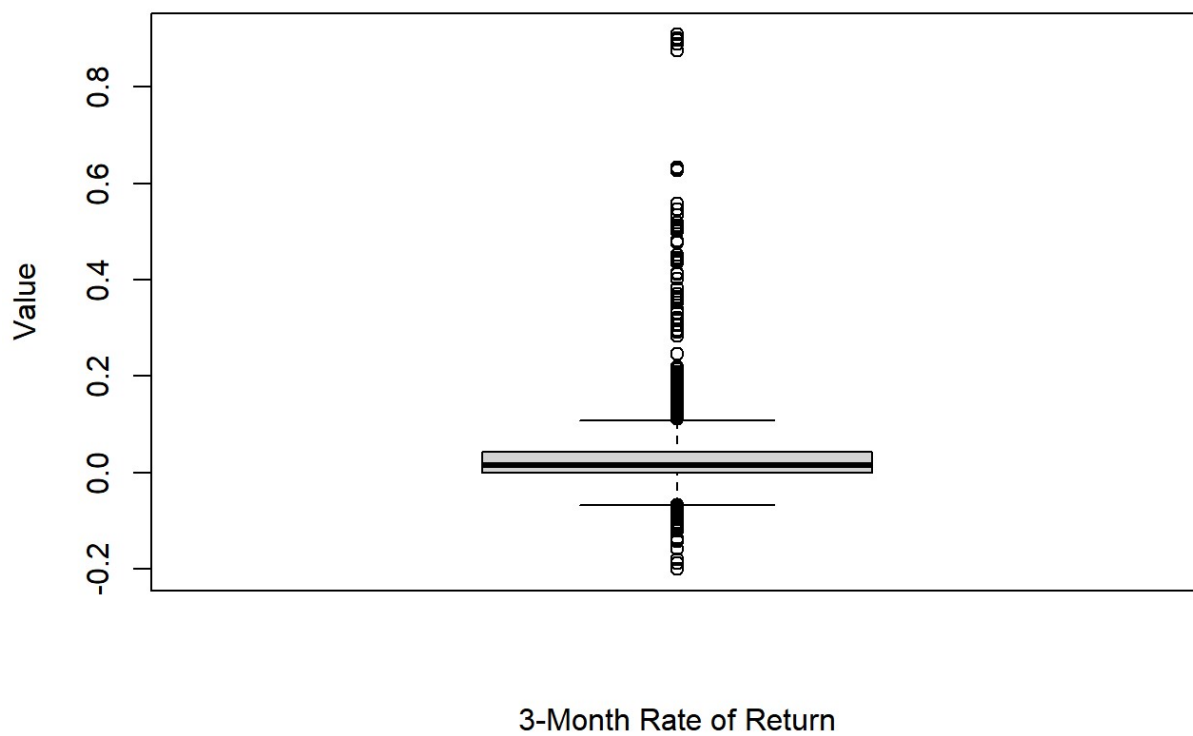
Year to Date Rate of Return

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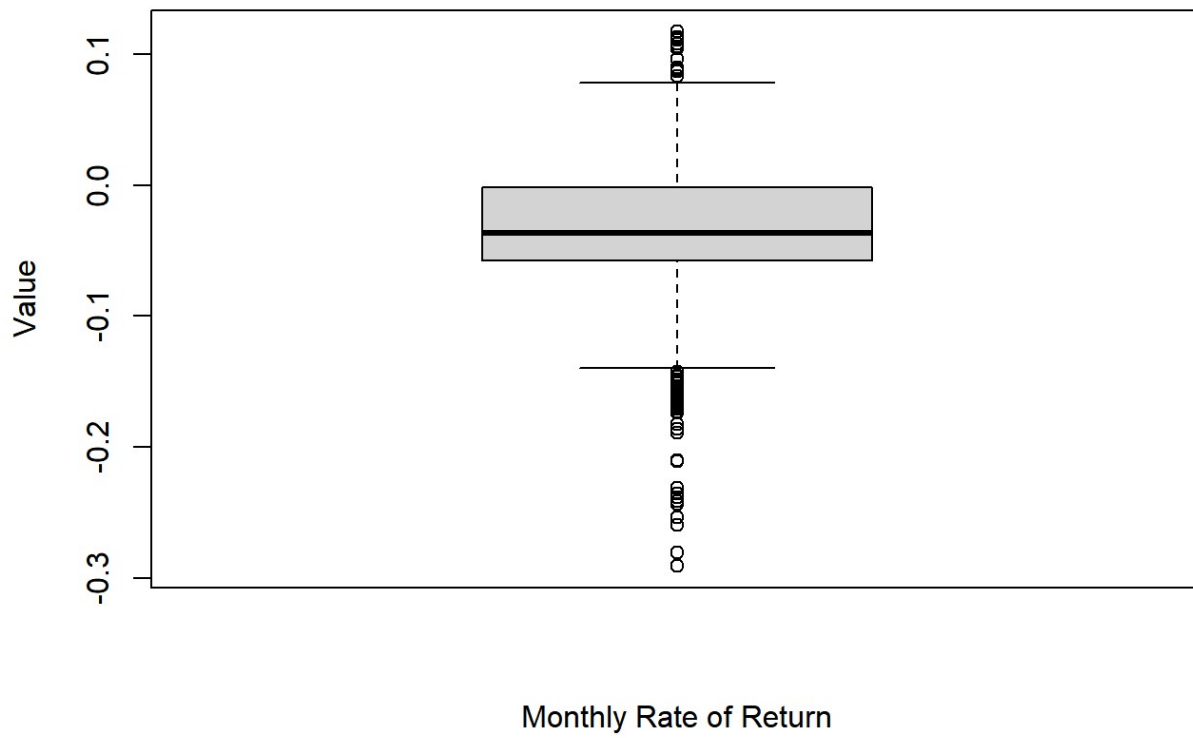

Fee

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

Volatility

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


3-Month Rate of Return

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

Monthly Rate of Return

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


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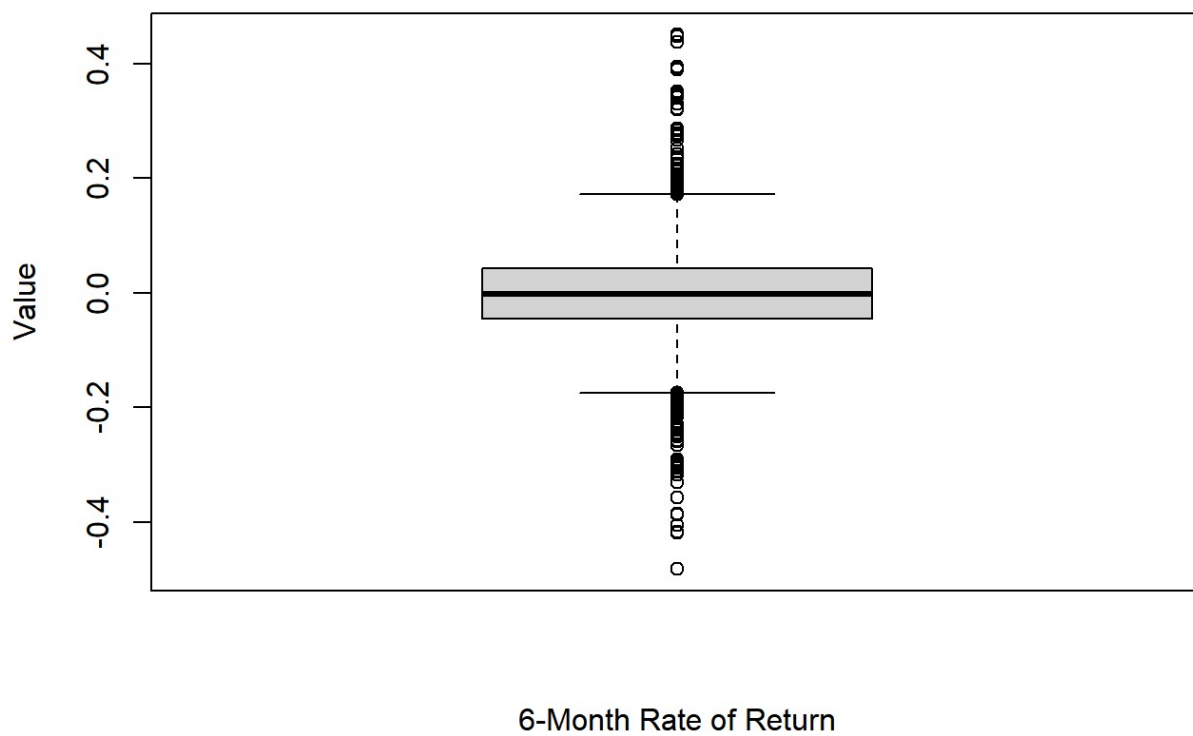
```
boxplot(mydata$yearReturnCUR, xlab="Yearly Rate of Return", ylab="Value")
```

Yearly Rate of Return

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


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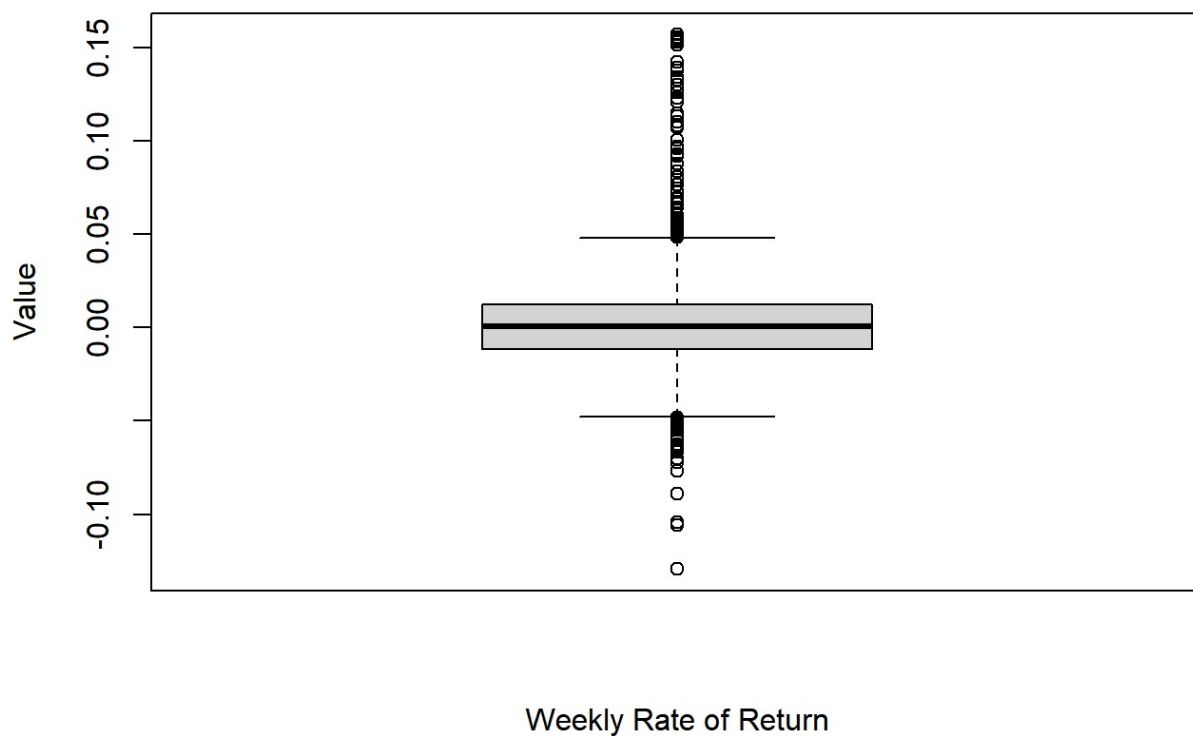
```
boxplot(mydata$yearReturn1CUR, xlab="Yearly Rate of Return", ylab="Value")
```

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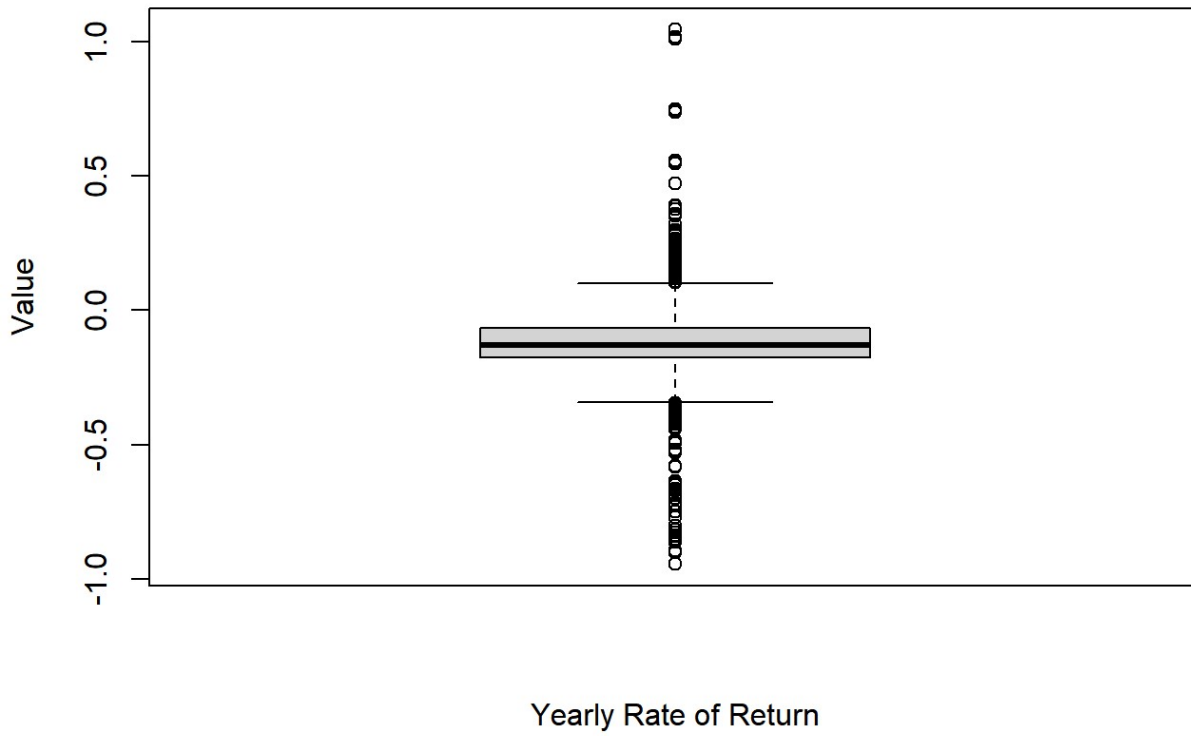
```
boxplot(mydata$quoteRange, xlab="Quote Range", ylab="Value")
```
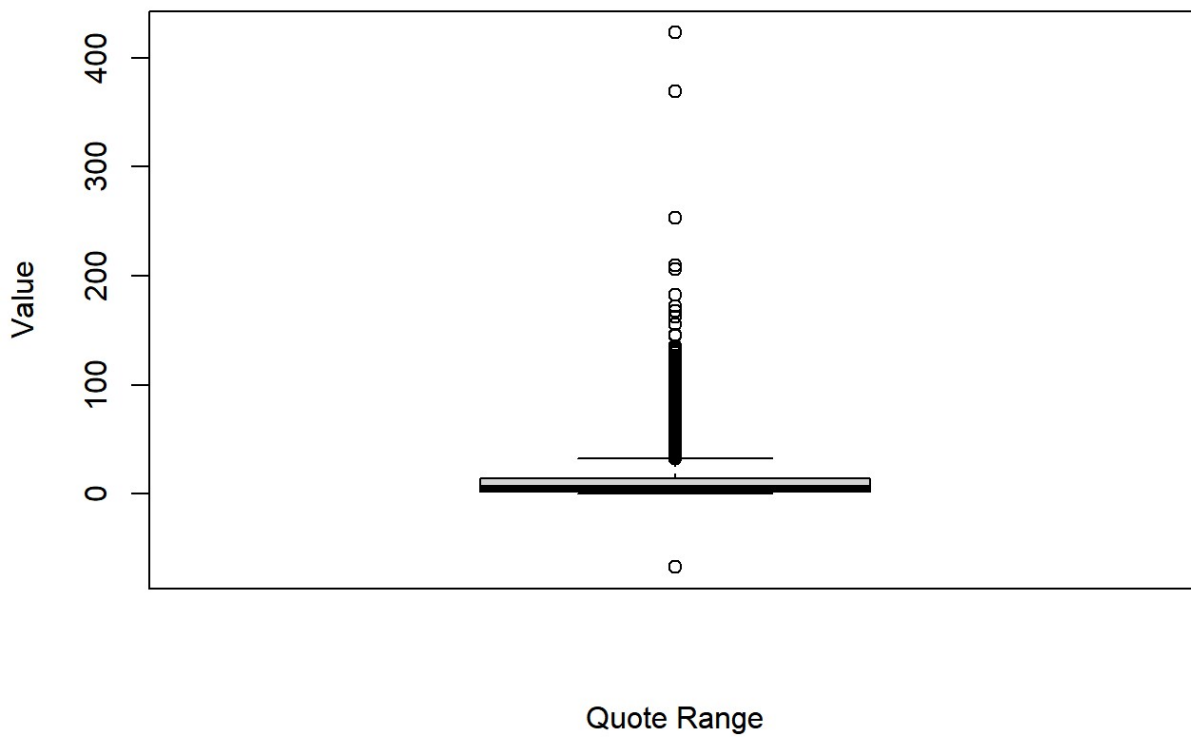

Certain columns

Quote Range

show that some errors are present, such as ytdReturnCUR having a value past -120 on the chart when that would represent a total loss of more than 100%, which is impossible. Additionally, quote and quoteRange showing values in the negative are not

possible. Everything else appears at first glance to be in order.

Question/Action: Show me a boxplot that shows yearReturn grouped by currencyRisk

```
boxplot(mydata$yearReturnCUR ~ mydata$currencyRisk)
```
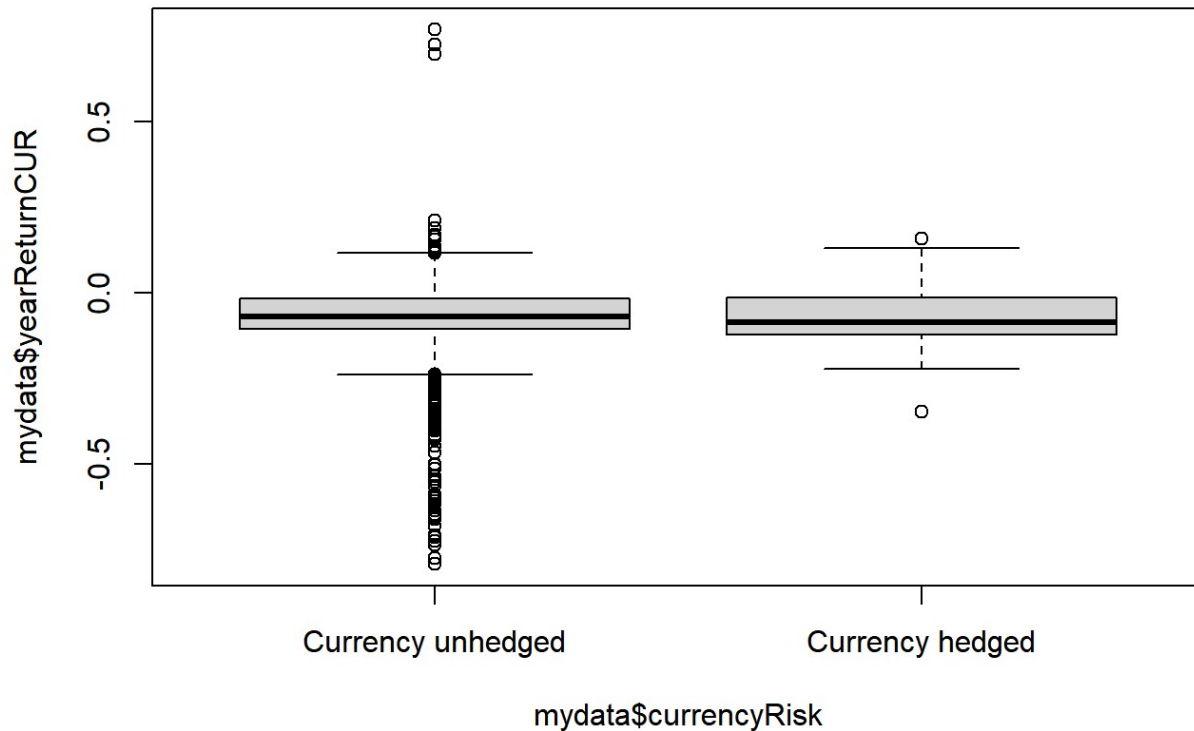

Question/Action: Show me two other interesting box plots with a numeric variable

```
#Graphing the volatility of funds based on the domicile, along with graphing the 6 month rate of return bas
ed on the manager of the fund
par(cex.axis=0.8)
boxplot(mydata$yearVolatilityCUR ~ mydata$domicileCountry, las=2, xlab='')
```
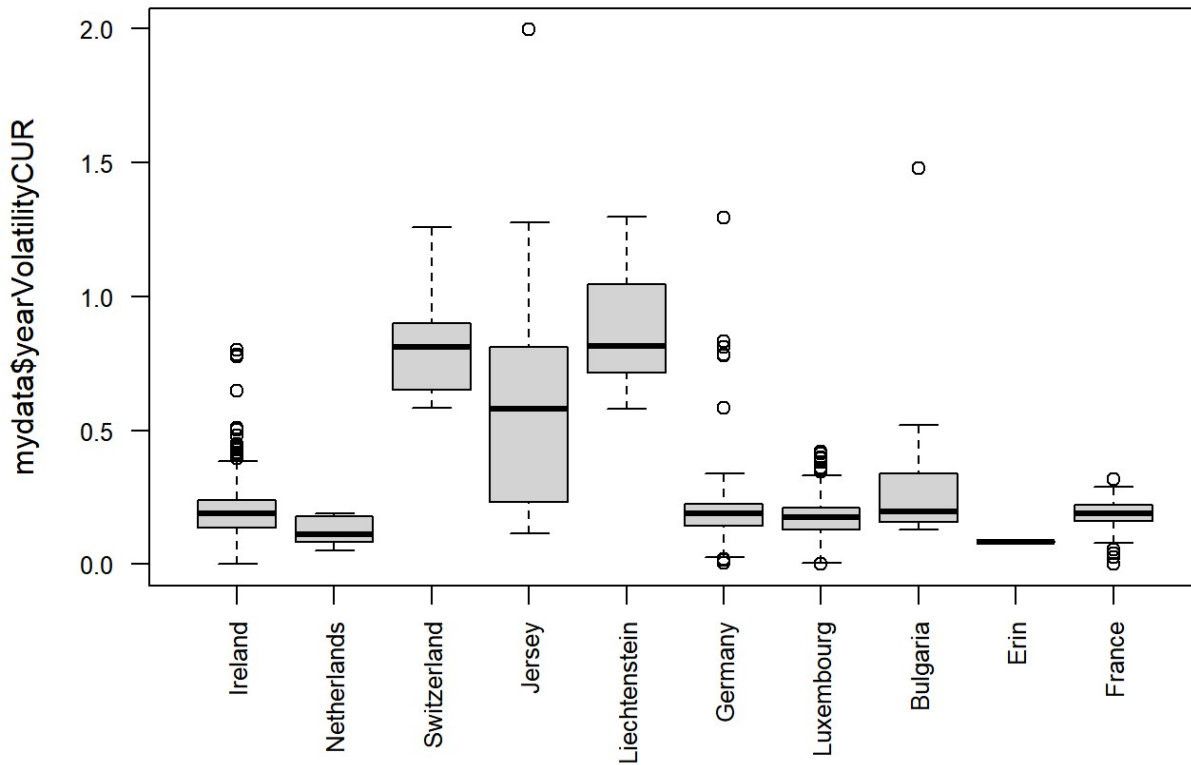
```
boxplot(mydata$sixMonthReturnCUR ~ mydata$fundProvider, las=2, xlab='')
```
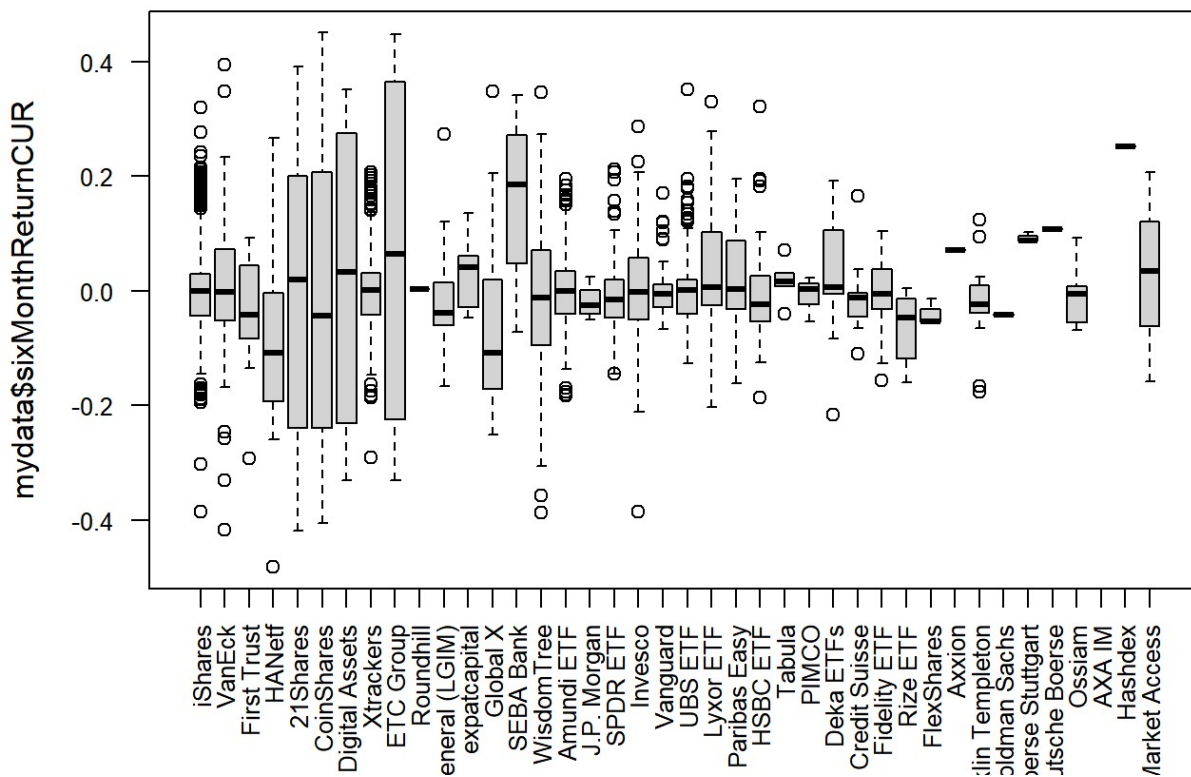

d. Show a histogram of one variable, specifically ytdReturn

```
#I had to avoid the error data in this case to make a readable graph
hist(mydata$ytdReturnCUR[mydata$ytdReturnCUR>-1],xlab="Year-To-Date Rate of Return", xlim=c(-0.5,1))
```

Histogram of mydata$ytdReturnCUR[mydata$ytdReturnCUR > -1]

Frequency (y-axis: 0, 500, 1000, 1500)

Year-To-Date Rate of Return (x-axis: -0.5, 0.0, 0.5, 1.0)

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$defineCalculation(calculationName="total", summariseExpression="n()")
pt$evaluatePivot()
pt$renderPivot()
```
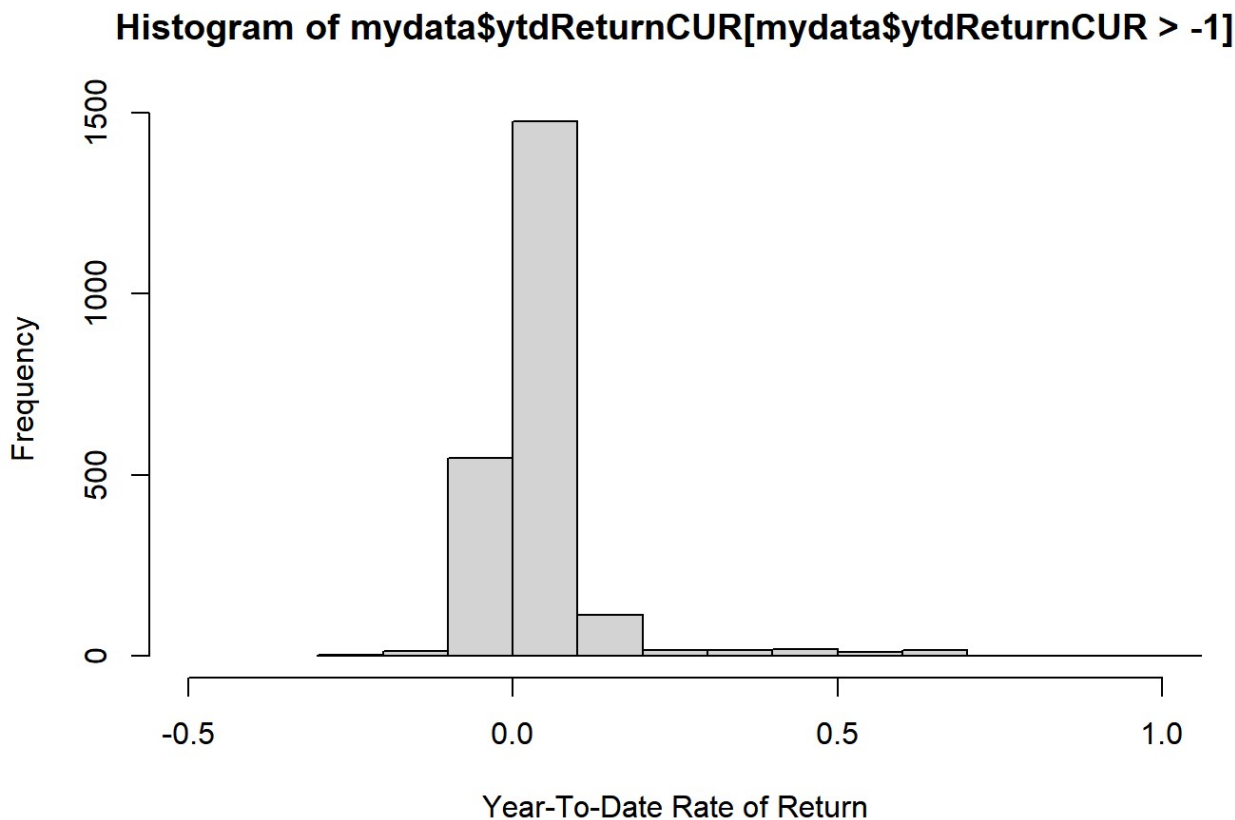
	EUR	USD	USD Hedged	EUR Hedged	GBP Hedged	FAR	GBP	CHF Hedged	CHF	AUD	JPY	HKD	CAD	SEK	BGN	Total
iShares	161	243	11	48	3	1	13	6	1		2				1	490
VanEck	12	31														43
First Trust	2	13					1									16
HANetf	1	23														24
21Shares		22														22
CoinShares		13														13
Deutsche Digital 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	854	1071	30	205	12	1	40	12	9	2	23	2	1	1	1	2264

Question: Create a cross table of counts by the factors currencyRisk and securitiesLending

```
pt2 <- PivotTable$new()
pt2$addData(mydata)
pt2$addColumnDataGroups("securitiesLending")
pt2$addRowDataGroups("currencyRisk")
pt2$defineCalculation(calculationName="total2", summariseExpression="n()")
pt2$evaluatePivot()
pt2$renderPivot()
```

	FALSE	TRUE	Total
Currency unhedged	1415	590	2005
Currency hedged	147	112	259
Total	1562	702	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(5,6,7,8,9,10,12,14)]), center=TRUE, scale=TRUE)
mydata.pca.plot <- autoplot(mydata.pca, data=na.omit(mydata[,c(5,6,7,8,9,10,12,14,18)]), colour='domicileCo
untry')
mydata.pca.plot
```

When you run the
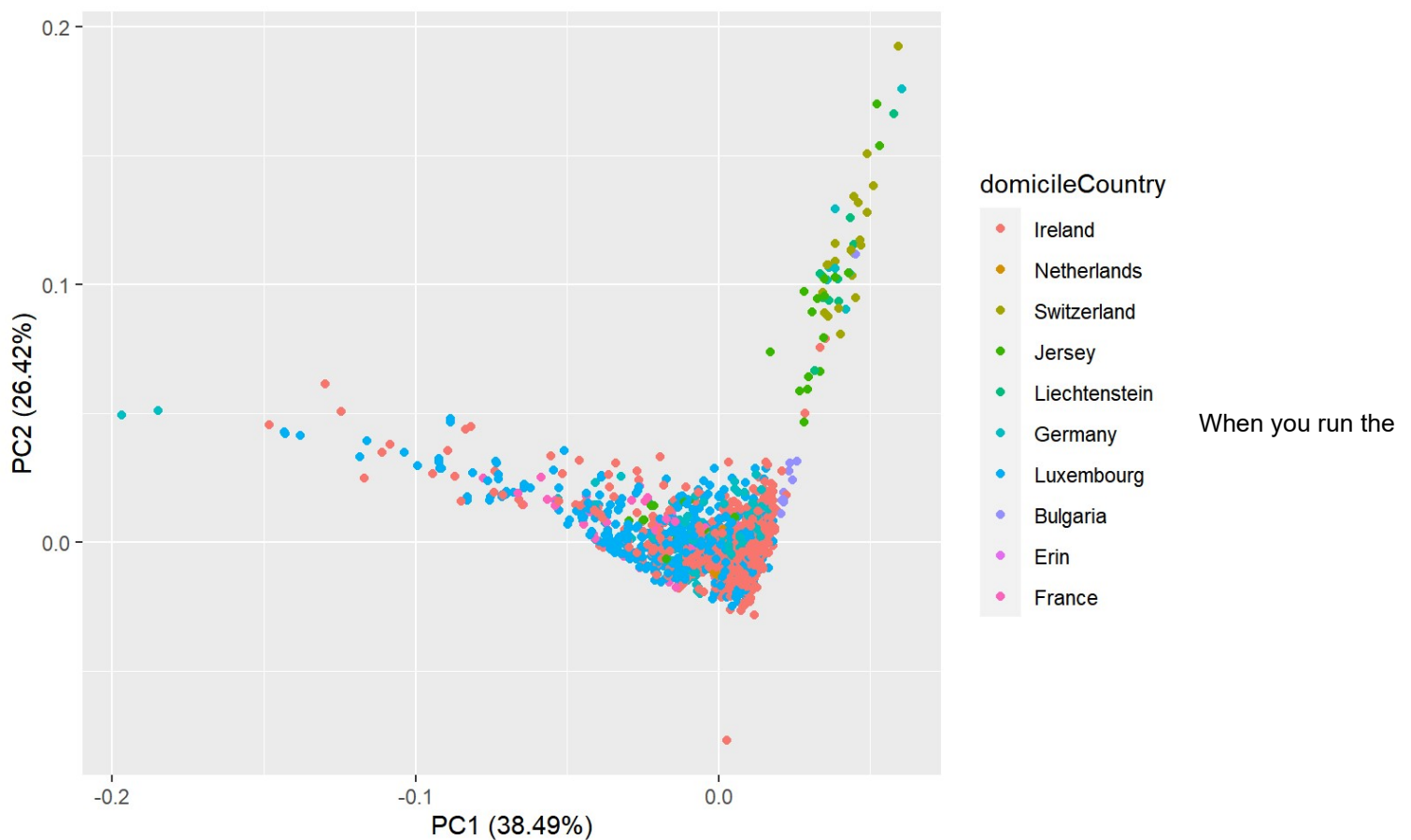
PCA, use the scale.=TRUE option in prcomp or cor=TRUE in princomp, so the data is scaled to a variance of 1. Explain: What impact does this have? Why would I ask you to do this?

By scaling everything to having a variance of 1, you can compare different variables with different magnitudes without it making the graph completely unable to be read. Graphing the relationships between values within a set is easier to compare to other sets without having to worry about issues of scale.

For the PCA, create a second version of the data set that has no rows with Nas. Determine how many rows were removed after you create this

I was unable to create a PCA with values containing NAs, so the first one that I made already has no NAs. However, the rows that were removed is something that I can see:

```
nrow(mydata)
```

```
## [1] 2264
```

```
nrow(na.omit(mydata))
```

```
## [1] 1986
```

Create two different plots of the PCA scores on axis 1 and 2 (ie standard PCA plots), coloring by currencyRisk and by domicileCountry

I already have the first one colored by domicileCountry, so I'll do the next one by currencyRisk

```
mydata.pca <- prcomp(na.omit(mydata[,c(5,6,7,8,9,10,12,14)]), center=TRUE, scale=TRUE)
mydata.pca.plot <- autoplot(mydata.pca, data=na.omit(mydata[,c(5,6,7,8,9,10,12,14,22)]), colour='currencyRisk')
mydata.pca.plot
```
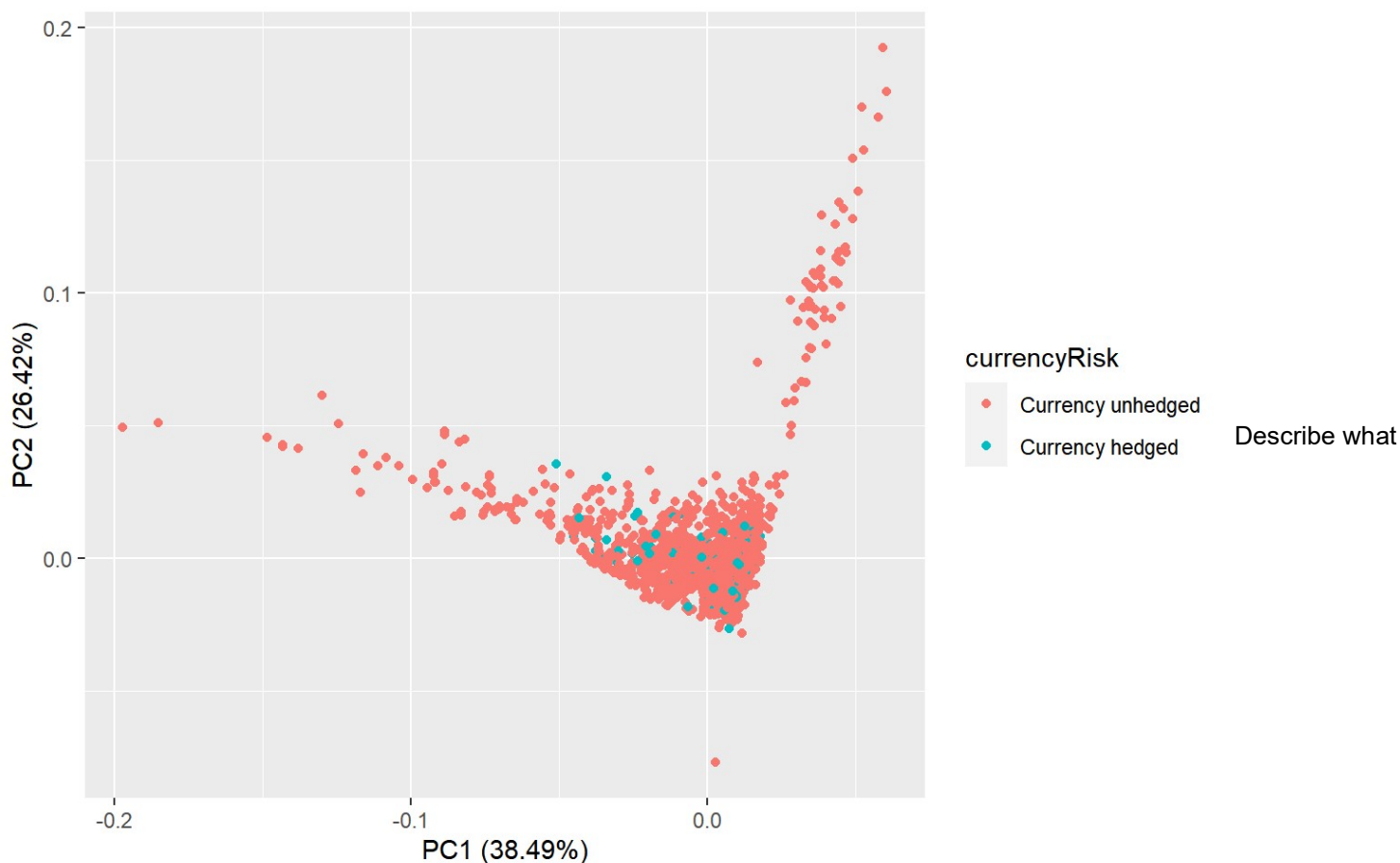
pattern you see.

First off, there's no real discernable pattern when it comes to the domicileCountry. For currencyRisk, it appears that only the unhedged funds show an extreme variation from having extremely low PC1 and high PC2. Additionally, hedged funds are generally speaking in a much tighter cluster of results, likely because of hedging cutting back on profits while also shielding from losses, thus making them have a lower amount of volatility.

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

```
rules <- validator(quote52High-quote52Low==quoteRange,
                    quote52Low < quote52High)
validmydata <- confront(mydata,rules)
head(validmydata)
```

```
## Object of class 'validation'
## Call:
##      head.default(x = validmydata)
##
## Rules confronted: 2
##    With fails   : 2
##    With missings: 0
##    Threw warning: 0
##    Threw error  : 0
```

h. Run an error localization using some basic rules

```
basicRules <- validator(quote52High-quote52Low==quoteRange)
lemydata <- locate_errors(mydata,basicRules)
summary(lemydata)
```

```
## Variable:
##                     name errors missing
## 7           quote52High      1       0
## 1                  isin      0       0
## 2                   wkn      0       0
## 3                  name      0       0
## 4          fundProvider      0       0
## 5                 quote      0       4
## 6            quote52Low      0       0
## 8           ytdReturnCUR      0      31
## 9                  fees      0       0
## 10     yearVolatilityCUR      0     226
## 11           fundCurrency      0       0
## 12 threeMonthReturnCUR      0      35
## 13        monthReturnCUR      0      13
## 14     sixMonthReturnCUR      0      77
## 15         inceptionDate      0       0
## 16                ticker      0       0
## 17          yearReturnCUR      0     228
## 18        domicileCountry      0       0
## 19          weekReturnCUR      0       5
## 20        yearReturn1CUR      0     273
## 21             quoteDate      0       0
## 22           currencyRisk      0       0
## 23      UCITSCompliance      0       0
## 24     securitiesLending      0       0
## 25            quoteRange      0       0
## Errors per record:
##    errors records
## 1       0    2263
## 2       1       1
```

i. Be ready to use the functional relation rule

```
relationrule<-validator(isin~name+wkn)
relationmydata=confront(mydata,relationrule)
summary(relationmydata)
```

```
##   name items passes fails nNA error warning         expression
## 1   V1  2264   2264     0    0 FALSE   FALSE isin ~ name + wkn
```

j. Do error corrections, one being a rule-based correction

```
mymod=modifier(
  if(domicileCountry=='Erin')
  {
    domicileCountry<-'Ireland'
  }
)
mod_data=modify(mydata,mymod)
table(mydata$domicileCountry)
```

```
##
##        Ireland    Netherlands   Switzerland         Jersey Liechtenstein
##           1265             12            28             37             14
##        Germany     Luxembourg      Bulgaria           Erin         France
##            136            705            11              1             55
```

```
table(mod_data$domicileCountry)
```

```
##
##        Ireland    Netherlands   Switzerland         Jersey Liechtenstein
##           1266             12            28             37             14
##        Germany     Luxembourg      Bulgaria           Erin         France
##            136            705            11              0             55
```

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

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

```
##      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.   :-25.35    Min.   :  0.000   Min.   :  0.52    Min.   :-128.2000
## 1st Qu.:  9.90    1st Qu.:  9.117   1st Qu.: 11.40    1st Qu.:   0.0000
## Median : 26.52    Median : 24.430   Median : 30.74    Median :   0.0233
## Mean   : 61.06    Mean   : 55.540   Mean   : 69.09    Mean   :  -0.0185
## 3rd Qu.: 72.89    3rd Qu.: 67.442   3rd Qu.: 81.46    3rd Qu.:   0.0452
## Max.   :923.00    Max.   :911.170   Max.   :965.82    Max.   :   1.4638
##                                                       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      : 854   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)  :  41   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    :1265   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)    :  38   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
##
##
##
##
##    quoteRange
## Min.   :-66.93
## 1st Qu.:  2.12
## Median :  5.77
## Mean   : 13.55
## 3rd Qu.: 14.47
## Max.   :422.99
##
```

Impute all three of these variables using a random forest model

monthReturnCUR, sixMonthReturnCUR, threeMonthReturnCUR

```
mydata.imp2=impute_rf(mydata, monthReturnCUR + sixMonthReturnCUR + threeMonthReturnCUR ~ fees)
summary(mydata.imp2)
```

```
##      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.   :-25.350   Min.   :  0.000   Min.   :  0.52   Min.   :-128.2000
## 1st Qu.:  9.898   1st Qu.:  9.117   1st Qu.: 11.40   1st Qu.:   0.0000
## Median : 26.420   Median : 24.430   Median : 30.74   Median :   0.0233
## Mean   : 60.347   Mean   : 55.540   Mean   : 69.09   Mean   :  -0.0185
## 3rd Qu.: 72.657   3rd Qu.: 67.442   3rd Qu.: 81.46   3rd Qu.:   0.0452
## Max.   :923.000   Max.   :911.170   Max.   :965.82   Max.   :   1.4638
## 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       : 854   1st Qu.:-0.00110
## Median :0.002500   Median :0.1849    EUR Hedged: 205   Median : 0.01630
## Mean   :0.003261   Mean   :0.2025    GBP       :  40   Mean   : 0.03296
## 3rd Qu.:0.004000   3rd Qu.:0.2359    USD Hedged:  30   3rd Qu.: 0.04163
## Max.   :0.025000   Max.   :1.9948    JPY       :  23   Max.   : 0.90850
##                    NA's   :226       (Other)   :  41
## monthReturnCUR     sixMonthReturnCUR   inceptionDate         ticker
## Min.   :-0.29100   Min.   :-0.48250   Length:2264        Length:2264
## 1st Qu.:-0.05750   1st Qu.:-0.04450   Class :character   Class :character
## Median :-0.03630   Median :-0.00120   Mode  :character   Mode  :character
## Mean   :-0.03351   Mean   : 0.00597
## 3rd Qu.:-0.00150   3rd Qu.: 0.04015
## Max.   : 0.11720   Max.   : 0.45040
##
## yearReturnCUR        domicileCountry weekReturnCUR       yearReturn1CUR
## Min.   :-0.79370   Ireland    :1265   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)    :  38   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
##
##
##
##
##    quoteRange
## Min.   :-66.93
## 1st Qu.:  2.12
## Median :  5.77
## Mean   : 13.55
## 3rd Qu.: 14.47
## Max.   :422.99
##
```