THE CAPITAL ASSET PRICING MODEL

$$R_i = R_f + \beta_i (R_m - R_f)$$

EXAMPLE 3: MULTIPLE LINEAR REGRESSION WITH AN OIL STOCK

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

WE WANT TO FIND A MODEL THAT CAN PREDICT THE RATE OF RETURN ON AN OIL STOCK LIKE EXXON MOBIL

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

$$R_i - R_f = \beta_i (R_m - R_f)$$

RETURNS OF EXXON MOBIL - RISK FREE RATE

THIS IS THE CAPM MODEL THAT WE HAVE ALREADY SEEN

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

$$R_i - R_f = \beta_i(R_m - R_f)$$

RETURNS OF AN INDEX THAT REPRESENTS THE MARKET - RISK FREE RATE

S&P 500

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

$$R_i - R_f = \beta_i (R_m - R_f)$$

THIS MODEL USES ONLY 1 DEPENDENT VARIABLE

THIS KIND OF REGRESSION IS CALLED SIMPLE LINEAR REGRESSION

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

$$R_i - R_f = \beta_i (R_m - R_f) + \beta_j R_j$$

WHAT IF WE ADDED ANOTHER VARIABLE?

RETURNS OF OIL PRICES

BY INCLUDING MORE VARIABLES, WE HOPE TO EXPLAIN MORE OF THE VARIATION IN THE DEPENDENT VARIABLE

EXAMPLE 3: MULTIPLE LINEAR REGRESSION

$$R_i - R_f = \beta_i (R_m - R_f) + \beta_j R_j$$

THIS KIND OF REGRESSION IS CALLED
MULTIPLE LINEAR REGRESSION

LET'S GO THROUGH THE STEPS WE'LL NEED TO DO

STEP 1: BUILD A SIMPLE LINEAR REGRESSION MODEL USING ONLY EXXON MOBIL AND S&P 500

WE'LL DOWNLOAD THE MONTHLY PRICES FOR

EXXON MOBIL (XOM)

^GSPC (S&P 500)

FOR 6 YEARS (JAN 1, 2010 - FEB 1, 2016)

THESE ARE THE FILE PATHS FOR THE DATA DOWNLOADED FROM YAHOO FINANCE

```
xomFile <- '/Users/swethakolalapudi/Desktop/Regression/xom.csv'
snpFile <-'/Users/swethakolalapudi/Desktop/Regression/snp.csv'
```

WE'LL USE A **FUNCTION WE WROTE EARLIER** TO PREPROCESS THE DATA IN THESE FILES

```
names(xom)[2:3] <- c("xom.returns","snp.returns")
```

WE'LL GIVE RETURNS COLUMNS **PROPER NAMES**

```
xom_SM <- lm(xom$xom.returns~xom$snp.returns)
```

PERFORM LINEAR REGRESSION

```r
xom_SM <- lm(xom$xom.returns~xom$snp.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```r
summary(xom_SM)
```

BETA FOR XOM

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.076423 -0.022784 -0.003236  0.017084  0.085066 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.003908   0.003761  -1.039    0.302    
xom$snp.returns  0.868164   0.098480   8.816 5.15e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03186 on 71 degrees of freedom
Multiple R-squared:  0.5226,	Adjusted R-squared:  0.5159 
F-statistic: 77.71 on 1 and 71 DF,  p-value: 5.146e-13
```

```
xom_SM <- lm(xom$xom.returns~xom$snp.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```
summary(xom_SM)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.076423 -0.022784 -0.003236  0.017084  0.085066

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003908   0.003761  -1.039    0.302
xom$snp.returns  0.868164   0.098480   8.816 5.15e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03186 on 71 degrees of freedom
Multiple R-squared:  0.5226,	Adjusted R-squared:  0.5159
F-statistic: 77.71 on 1 and 71 DF,  p-value: 5.146e-13
```

THIS RELATIONSHIP IS STATICALLY SIGNIFICANT

```r
xom_SM <- lm(xom$xom.returns~xom$snp.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```r
summary(xom_SM)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.076423 -0.022784 -0.003236  0.017084  0.085066

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003908   0.003761  -1.039    0.302
xom$snp.returns  0.868164   0.098480   8.816 5.15e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03186 on 71 degrees of freedom
Multiple R-squared:  0.5226,	Adjusted R-squared:  0.5159
F-statistic: 77.71 on 1 and 71 DF,  p-value: 5.146e-13
```

R-SQUARE

= 52%

```
xom_SM <- lm(xom$xom.returns~xom$snp.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```
summary(xom_SM)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.076423 -0.022784 -0.003236  0.017084  0.085066

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003908   0.003761  -1.039    0.302
xom$snp.returns  0.868164   0.098480   8.816 5.15e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03186 on 71 degrees of freedom
Multiple R-squared:  0.5226,    Adjusted R-squared:  0.5159
F-statistic: 77.71 on 1 and 71 DF,  p-value: 5.146e-13
```

ADJUSTED R-SQUARE = 51.5%

R-SQUARE = 52%

SIMPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 51.5%

R-SQUARE = 52%

LET'S SEE WHAT HAPPENS WHEN WE ADD ANOTHER VARIABLE TO THE REGRESSION

STEP 2: ADD ANOTHER VARIABLE TO THE REGRESSION

STEP 2: ADD ANOTHER VARIABLE TO THE REGRESSION

WE'LL ADD THE RETURNS OF OIL PRICES

THESE WILL BE REPRESENTED BY THE RETURNS ON AN OIL ETF (THESE ARE FUNDS WHOSE RETURNS DEPEND HEAVILY ON OIL PRICES)

READ THE PRICES OF USO FROM A FILE AND CONVERT THEM TO RETURNS

```
uso <- read.table('/Users/swethakolalapudi/Desktop/Regression/
uso.csv',header = TRUE, sep =",")[,c("Date","Adj.Close")]
names(uso)[2]<-"uso.returns"
uso[,c("Date")] <- as.Date(uso[,c("Date")])
uso <- uso[order(uso$Date, decreasing = TRUE),]
uso[-nrow(uso),-1] <- uso[-nrow(uso),-1]/uso[-1,-1]-1
```

THIS IS VERY SIMILAR TO HOW WE DID THE PREPROCESSING EARLIER, EXCEPT WE ARE JUST DOING IT FOR ONE SECURITY

MERGE THE RETURNS OF USO AND THE RETURNS WE HAVE FROM THE PREVIOUS STEP INTO ONE DATA FRAME (THIS IS SO THEY ARE ALL ALIGNED BY DATE)

```r
xom <- merge(xom, uso, by = "Date")
```

REGRESS XOM RETURNS ON BOTH S&P AND USO

```r
xom_MLR <- lm(xom$xom.returns~xom$snp.returns + xom$uso.returns)
```

THIS MODEL HAS 2 INDEPENDENT VARIABLES

```
xom_MLR <- lm(xom$xom.returns~xom$snp.returns + xom$uso.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```
summary(xom_MLR)
```

CO-EFFICIENTS FOR S&P AND USO

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.061634 -0.020990 -0.004805  0.017007  0.085419

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003060   0.003741  -0.818   0.4162
xom$snp.returns  0.769195   0.112644   6.829 2.56e-09 ***
xom$uso.returns  0.089186   0.051429   1.734   0.0873 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03142 on 70 degrees of freedom
Multiple R-squared:  0.5422,    Adjusted R-squared:  0.5292
F-statistic: 41.46 on 2 and 70 DF,  p-value: 1.325e-12
```

```
xom_MLR <- lm(xom$xom.returns~xom$snp.returns + xom$uso.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```
summary(xom_MLR)
```

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.061634 -0.020990 -0.004805  0.017007  0.085419 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.003060   0.003741  -0.818   0.4162    
xom$snp.returns  0.769195   0.112644   6.829 2.56e-09 ***
xom$uso.returns  0.089186   0.051429   1.734   0.0873 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03142 on 70 degrees of freedom
Multiple R-squared:  0.5422,	Adjusted R-squared:  0.5292 
F-statistic: 41.46 on 2 and 70 DF,  p-value: 1.325e-12
```

BOTH OF THESE VARIABLES HAVE A STATISTICALLY SIGNIFICANT RELATIONSHIP

```r
xom_MLR <- lm(xom$xom.returns~xom$snp.returns + xom$uso.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```r
summary(xom_MLR)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.061634 -0.020990 -0.004805  0.017007  0.085419

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003060   0.003741  -0.818   0.4162
xom$snp.returns  0.769195   0.112644   6.829 2.56e-09 ***
xom$uso.returns  0.089186   0.051429   1.734   0.0873 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03142 on 70 degrees of freedom
Multiple R-squared:  0.5422,	Adjusted R-squared:  0.5292
F-statistic: 41.46 on 2 and 70 DF,  p-value: 1.325e-12
```

R-SQUARE

= 54%

```
xom_MLR <- lm(xom$xom.returns~xom$snp.returns + xom$uso.returns)
```

LET'S SEE THE RESULTS OF THIS REGRESSION

```
summary(xom_MLR)
```

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.061634 -0.020990 -0.004805  0.017007  0.085419

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.003060   0.003741  -0.818   0.4162
xom$snp.returns  0.769195   0.112644   6.829 2.56e-09 ***
xom$uso.returns  0.089186   0.051429   1.734   0.0873 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03142 on 70 degrees of freedom
Multiple R-squared:  0.5422,    Adjusted R-squared:  0.5292
F-statistic: 41.46 on 2 and 70 DF,  p-value: 1.325e-12
```

ADJUSTED R-SQUARE = 52.9%

R-SQUARE = 54%

LET'S COMPARE THIS WITH THE PREVIOUS RESULTS

SIMPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 51.5%

R-SQUARE = 52%

MULTIPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 52.9%

R-SQUARE = 54%

R-SQUARE IMPROVED WHEN YOU ADDED A NEW VARIABLE

SIMPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 51.5%

R-SQUARE = 52%

MULTIPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 52.9%

R-SQUARE = 54%

ADJUSTED R-SQUARE ALSO IMPROVED

SIMPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 51.5%

R-SQUARE = 52%

MULTIPLE LINEAR REGRESSION

ADJUSTED R-SQUARE = 52.9%

R-SQUARE = 54%

NOTICE HOW THE DIFFERENCE BETWEEN R-SQUARE AND ADJUSTED R-SQUARE INCREASED

THIS HAPPENS AS YOU ADD MORE VARIABLES

VARIABLES THAT TAKE ONE OF A LIMITED SET OF VALUES ARE **CALLED CATEGORICAL VARIABLES**

MONTH   COLOR

QUARTER   GENDER

ARE PERFECT EXAMPLES OF CATEGORICAL VARIABLES

CATEGORICAL VARIABLES

THESE KIND OF VARIABLES USUALLY DON'T HAVE ANY NUMERIC MEANING

HOW DO WE INCLUDE THEM IN A LINEAR REGRESSION?

CATEGORICAL VARIABLES

HOW DO WE INCLUDE THEM IN A LINEAR REGRESSION?

WE CREATE A SET OF "DUMMY VARIABLES" TO REPRESENT THE CATEGORICAL VARIABLE

THE NUMBER OF VARIABLES = NUMBER OF DISTINCT VALUES -1

CATEGORICAL VARIABLES

HOW DO WE INCLUDE THEM IN A LINEAR REGRESSION?

“DUMMY” VARIABLES”

THE NUMBER OF VARIABLES = NUMBER OF DISTINCT VALUES -1

EACH DUMMY VARIABLE IS 1 FOR A PARTICULAR VALUE OF THE CATEGORICAL VARIABLE, ELSE IT IS 0

LET'S TAKE AN EXAMPLE

EXAMPLE 4: USING A CATEGORICAL VARIABLE FOR REGRESSION

EXAMPLE 4: USING A CATEGORICAL VARIABLE FOR REGRESSION

WE HAD EARLIER DONE AN EXERCISE TO REGRESS GOOGLE RETURNS AGAINST NASDAQ

LET'S INCLUDE MONTH, WHICH IS A CATEGORICAL VARIABLE

THERE ARE 12 DISTINCT VALUES FOR MONTH

LET'S INCLUDE MONTH, WHICH IS A CATEGORICAL VARIABLE

THERE ARE 12 DISTINCT VALUES FOR MONTH

THERE WILL BE 11 DUMMY VARIABLES

Var1
Var2
Var3
Var4
Var5
Var6
Var7
Var8
Var9
Var10
Var11

THE VALUE OF EACH OF THESE DEPENDS ON THE VALUE OF MONTH

LET'S INCLUDE MONTH, WHICH IS A CATEGORICAL VARIABLE

M	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec

Var1
Var2
Var3
Var4
Var5
Var6
Var7
Var8
Var9
Var10
Var11

FOR 11 OF THESE MONTHS ONE VARIABLE WILL HAVE THE VALUE 1 , ALL ELSE WILL BE 0

FOR ONE OF THE MONTHS ALL THE VARIABLES WILL BE 0

LET'S INCLUDE MONTH, WHICH IS A CATEGORICAL VARIABLE

M	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec
Var1	1	0	0	0	0	0	0	0	0	0	0	0
Var2	0	1	0	0	0	0	0	0	0	0	0	0
Var3	0	0	1	0	0	0	0	0	0	0	0	0
Var4	0	0	0	1	0	0	0	0	0	0	0	0
Var5	0	0	0	0	1	0	0	0	0	0	0	0
Var6	0	0	0	0	0	1	0	0	0	0	0	0
Var7	0	0	0	0	0	0	1	0	0	0	0	0
Var8	0	0	0	0	0	0	0	1	0	0	0	0
Var9	0	0	0	0	0	0	0	0	1	0	0	0
Var10	0	0	0	0	0	0	0	0	0	1	0	0
Var11	0	0	0	0	0	0	0	0	0	0	1	0

THIS IS CALLED ONE-HOT ENCODING

LET'S SEE HOW TO DO THIS IN R

WE ALREADY HAVE A DATA FRAME CALLED GOOG

LET'S ADD A MONTH COLUMN

```
goog$Month = format(goog$Date,"%m")
```

Date	goog.returns	nasdaq.returns	tbonds.returns	Month
2010-02-01	-0.0287651490	0.019495859	0.02284	02
2010-03-01	0.0509375748	0.045750044	0.02560	03
2010-04-01	-0.0972357416	0.002168238	0.02420	04
2010-05-03	-0.0971921643	-0.103917642	0.02097	05
2010-06-01	-0.1017174844	-0.083434017	0.01795	06
2010-07-01	0.0736930461	0.052983210	0.01598	07
2010-08-02	-0.0852566960	-0.075809642	0.01342	08
2010-09-01	0.1555604124	0.107618795	0.01281	09
2010-10-01	0.1554058941	0.046805211	0.01179	10
2010-11-01	-0.1091324620	-0.018301121	0.01464	11
2010-12-01	0.0486888972	0.041739880	0.02016	12
2011-01-03	-0.0087618121	-0.001724194	0.01952	01
2011-02-01	0.0003502713	0.009069816	0.02137	02
2011-03-01	-0.0656800154	-0.022681285	0.02225	03
2011-04-01	-0.0924544001	0.013499781	0.01975	04

YOU CAN CREATE DUMMY VARIABLES USING THE MODEL.MATRIX FUNCTION

```
dummyVars <- model.matrix(~Month,goog)
```

THE VARIABLE TO BE CONVERTED, THE DATA FRAME THE COLUMN IS IN

Month
02
03
04
05
06
07
08
09
10
11
12
01
02
03

YOU CAN CREATE DUMMY VARIABLES USING THE MODEL.MATRIX FUNCTION

```
dummyVars <- model.matrix(~Month,goog)
```

Month
02
03
04
05
06
07
08
09
10
11
12
01
02
03

YOU CAN CREATE DUMMY VARIABLES USING THE MODEL.MATRIX FUNCTION

```
dummyVars <- model.matrix(~Month,goog)
```

Month
02
03
04
05
06
07
08
09
10
11
12
01
02
03

(Intercept)	Month02	Month03	Month04	Month05	Month06	Month07	Month08	Month09	Month10	Month11	Month12
1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0

YOU CAN CREATE DUMMY VARIABLES USING THE MODEL.MATRIX FUNCTION

```
dummyVars <- model.matrix(~Month,goog)
```

Month
02
03
04
05
06
07
08
09
10
11
12
01
02
03

(Intercept)	Month02	Month03	Month04	Month05	Month06	Month07	Month08	Month09	Month10	Month11	Month12
1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0
1	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0

```r
dummyVars <- model.matrix(~Month,goog)
```

YOU CAN MERGE THIS WITH THE ORIGINAL DATA FRAME AND THEN RUN A LINEAR REGRESSION ON ALL 11 VARIABLES + THE NASDAQ RETURNS

OR

DIRECTLY USE THE CATEGORICAL VARIABLE IN THE LINEAR REGRESSION

```r
goog_MLR <- lm(goog$goog.returns~goog$nasdaq.returns+goog$Month)
```

```r
goog_MLR <- lm(goog$goog.returns~goog$nasdaq.returns+goog$Month)
```

LET'S SEE THE RESULTS OF THE REGRESSION

```r
summary(goog_MLR)
```

```
Residuals:
      Min        1Q    Median        3Q       Max 
-0.167102 -0.027855  0.004201  0.034741  0.121227 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.001029   0.022894  -0.045   0.9643    
goog$nasdaq.returns  0.810490   0.163540   4.956 6.21e-06 ***
goog$Month02         0.001253   0.031531   0.040   0.9684    
goog$Month03        -0.023391   0.032359  -0.723   0.4726    
goog$Month04        -0.048513   0.032311  -1.501   0.1385    
goog$Month05         0.002568   0.032431   0.079   0.9371    
goog$Month06        -0.014763   0.032375  -0.456   0.6500    
goog$Month07         0.074377   0.032477   2.290   0.0255 *  
goog$Month08        -0.011228   0.032519  -0.345   0.7311    
goog$Month09         0.030690   0.032339   0.949   0.3464    
goog$Month10         0.048443   0.033136   1.462   0.1490    
goog$Month11        -0.012551   0.032330  -0.388   0.6992    
goog$Month12         0.025718   0.032315   0.796   0.4293    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05596 on 60 degrees of freedom
Multiple R-squared:  0.5187,	Adjusted R-squared:  0.4225 
F-statistic: 5.389 on 12 and 60 DF,  p-value: 4.221e-06
```

NOTICE HOW LM AUTOMATICALLY CREATED 11 DUMMY VARIABLES FROM MONTH

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -0.001029   0.022894  -0.045   0.9643
goog$nasdaq.returns 0.810490   0.163540   4.956 6.21e-06 ***
goog$Month02        0.001253   0.031531   0.040   0.9684
goog$Month03       -0.023391   0.032359  -0.723   0.4726
goog$Month04       -0.048513   0.032311  -1.501   0.1385
goog$Month05        0.002568   0.032431   0.079   0.9371
goog$Month06       -0.014763   0.032375  -0.456   0.6500
goog$Month07        0.074377   0.032477   2.290   0.0255 *
goog$Month08       -0.011228   0.032519  -0.345   0.7311
goog$Month09        0.030690   0.032339   0.949   0.3464
goog$Month10        0.048443   0.033136   1.462   0.1490
goog$Month11       -0.012551   0.032330  -0.388   0.6992
goog$Month12        0.025718   0.032315   0.796   0.4293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05596 on 60 degrees of freedom
Multiple R-squared:  0.5187,	Adjusted R-squared:  0.4225
F-statistic: 5.389 on 12 and 60 DF,  p-value: 4.221e-06
```

CO-EFFICIENTS FOR EACH OF THE VARIABLES

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -0.001029   0.022894  -0.045   0.9643
goog$nasdaq.returns   0.810490   0.163540   4.956 6.21e-06 ***
goog$Month02          0.001253   0.031531   0.040   0.9684
goog$Month03         -0.023391   0.032359  -0.723   0.4726
goog$Month04         -0.048513   0.032311  -1.501   0.1385
goog$Month05          0.002568   0.032431   0.079   0.9371
goog$Month06         -0.014763   0.032375  -0.456   0.6500
goog$Month07          0.074377   0.032477   2.290   0.0255 *
goog$Month08         -0.011228   0.032519  -0.345   0.7311
goog$Month09          0.030690   0.032339   0.949   0.3464
goog$Month10          0.048443   0.033136   1.462   0.1490
goog$Month11         -0.012551   0.032330  -0.388   0.6992
goog$Month12          0.025718   0.032315   0.796   0.4293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05596 on 60 degrees of freedom
Multiple R-squared:  0.5187,    Adjusted R-squared:  0.4225
F-statistic: 5.389 on 12 and 60 DF,  p-value: 4.221e-06
```

STATISTICAL SIGNIFICANCE

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)         -0.001029   0.022894  -0.045   0.9643    
goog$nasdaq.returns  0.810490   0.163540   4.956 6.21e-06 ***
goog$Month02         0.001253   0.031531   0.040   0.9684    
goog$Month03        -0.023391   0.032359  -0.723   0.4726    
goog$Month04        -0.048513   0.032311  -1.501   0.1385    
goog$Month05         0.002568   0.032431   0.079   0.9371    
goog$Month06        -0.014763   0.032375  -0.456   0.6500    
goog$Month07         0.074377   0.032477   2.290   0.0255 *  
goog$Month08        -0.011228   0.032519  -0.345   0.7311    
goog$Month09         0.030690   0.032339   0.949   0.3464    
goog$Month10         0.048443   0.033136   1.462   0.1490    
goog$Month11        -0.012551   0.032330  -0.388   0.6992    
goog$Month12         0.025718   0.032315   0.796   0.4293    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05596 on 60 degrees of freedom
Multiple R-squared:  0.5187,	Adjusted R-squared:  0.4225 
F-statistic: 5.389 on 12 and 60 DF,  p-value: 4.221e-06
```

ONLY MONTH07 IS STATISTICALLY SIGNIFICANT

AND IT'S CO-EFFICIENT VALUE IS LOW

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.001029   0.022894  -0.045   0.9643
goog$nasdaq.returns  0.810490   0.163540   4.956 6.21e-06 ***
goog$Month02         0.001253   0.031531   0.040   0.9684
goog$Month03        -0.023391   0.032359  -0.723   0.4726
goog$Month04        -0.048513   0.032311  -1.501   0.1385
goog$Month05         0.002568   0.032431   0.079   0.9371
goog$Month06        -0.014763   0.032375  -0.456   0.6500
goog$Month07         0.074377   0.032477   2.290   0.0255 *
goog$Month08        -0.011228   0.032519  -0.345   0.7311
goog$Month09         0.030690   0.032339   0.949   0.3464
goog$Month10         0.048443   0.033136   1.462   0.1490
goog$Month11        -0.012551   0.032330  -0.388   0.6992
goog$Month12         0.025718   0.032315   0.796   0.4293
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05596 on 60 degrees of freedom
Multiple R-squared:  0.5187,     Adjusted R-squared:  0.4225
F-statistic: 5.389 on 12 and 60 DF,  p-value: 4.221e-06
```

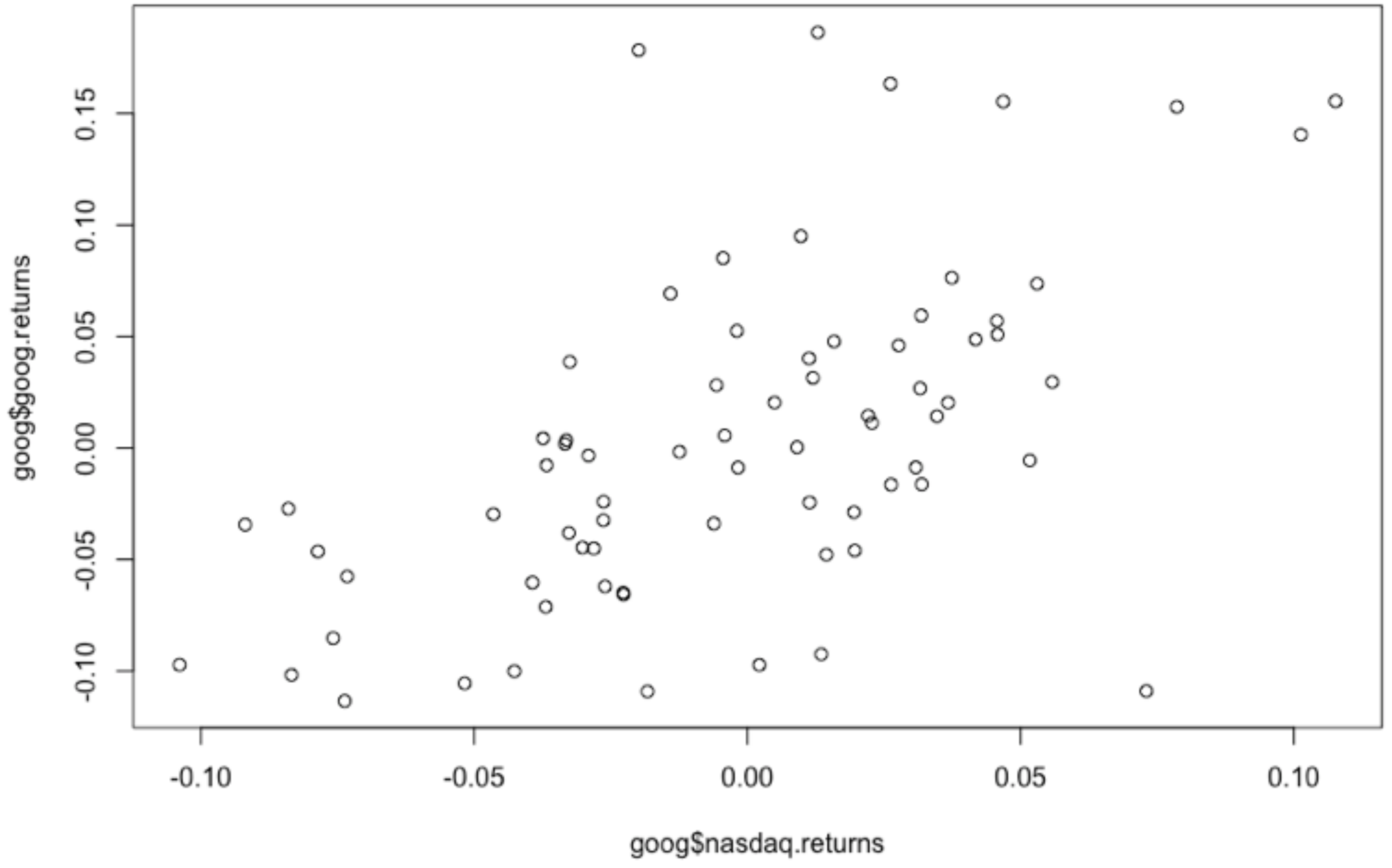
THERE IS A LESSON HERE

ADDING MORE VARIABLES IS NOT ALWAYS A GOOD THING!!

EXAMPLE 5: DEALING WITH OUTLIERS WITH RLM()

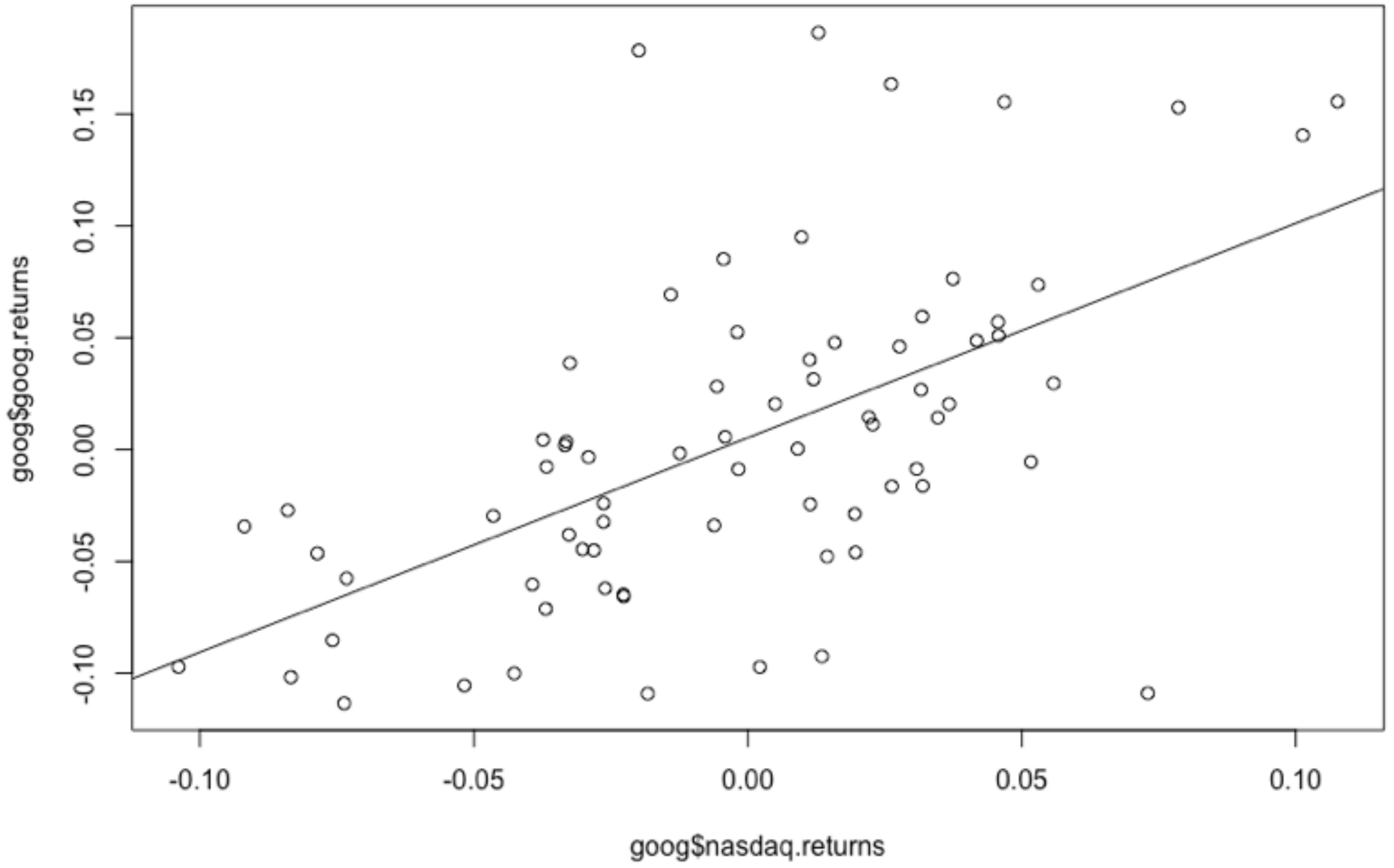
HERE IS THE PLOT FOR GOOGLE VS NASDAQ RETURNS

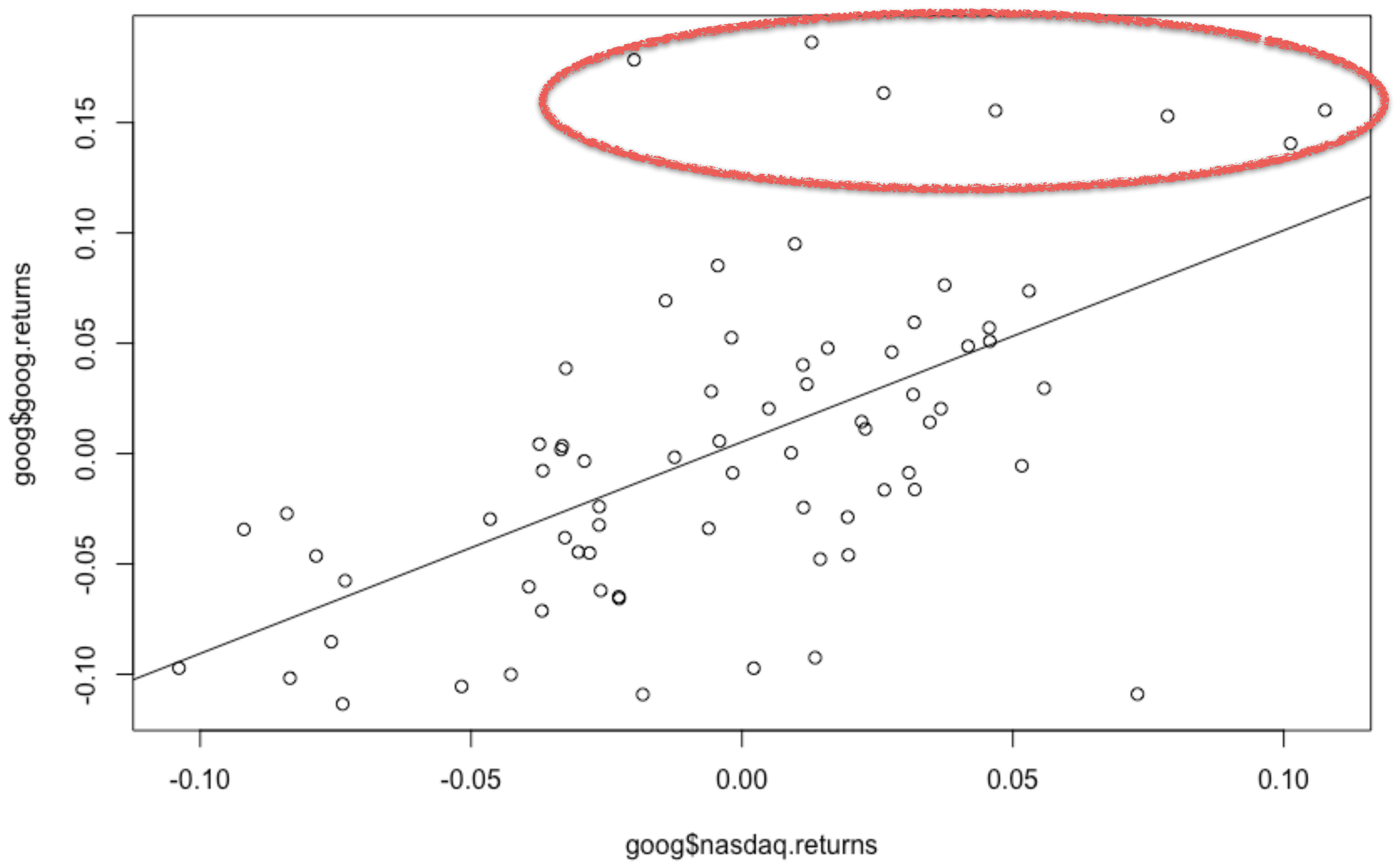
```
plot(goog$nasdaq.returns,goog$goog.returns)
```



HERE IS THE PLOT OF THE FITTED LINE AFTER REGRESSION

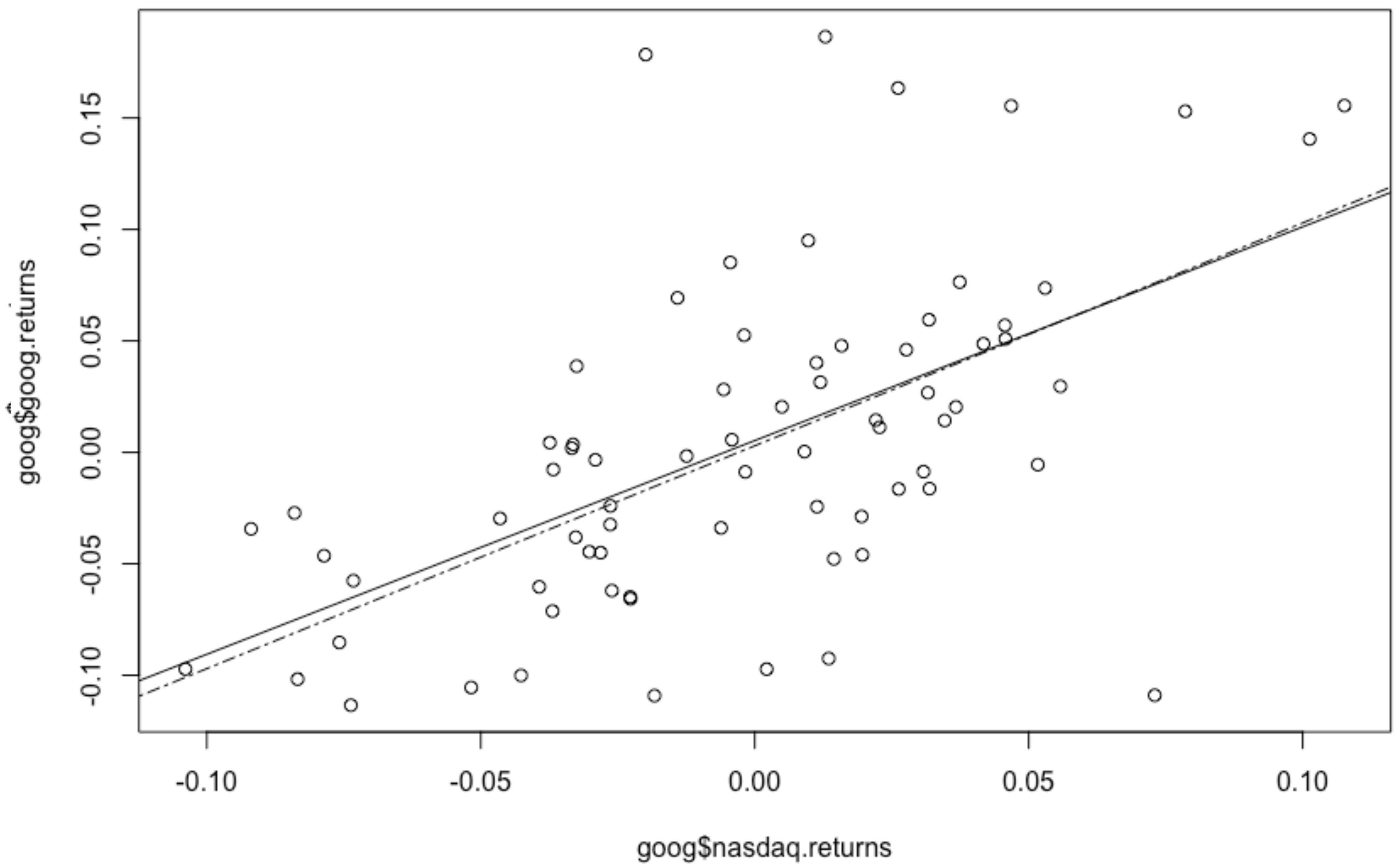
```
abline(googM)
```





YOU CAN SEE THAT THERE ARE SOME OUTLIERS IN THIS PLOT WHICH MIGHT BE DISTORTING THE LINE

RLM() FITS A LINE AFTER TAKING INTO ACCOUNT THE EFFECT OF OUTLIERS

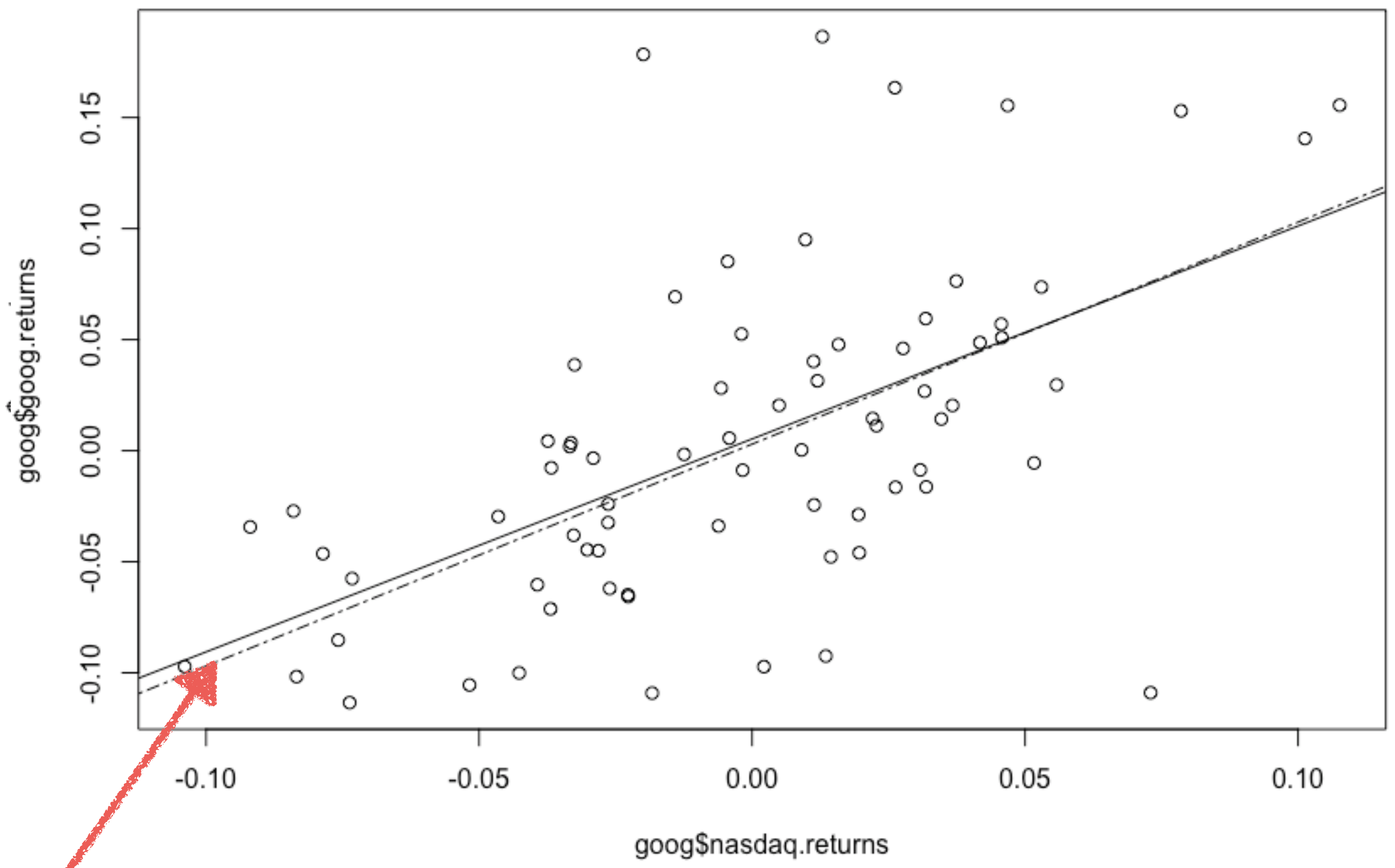


RLM() IS A FUNCTION IN THE MASS PACKAGE

```
require(MASS)
googRLM <- rlm(goog$goog.returns~goog$nasdaq.returns)
```

```
abline(googRLM, lty ='twodash')
```

LET'S SEE THE RESULTS AFTER USING RLM



THE DOTTED LINE IS THE RESULT OF RLM() WHICH HAS BEEN SLIGHTLY ADJUSTED FOR THE EFFECT OF OUTLIERS