Project

Group 1

2020-01-29

# Dataset:

mydata <- read.csv("C:/Users/Lea/Documents/Harrisburg University/Analytical Methods 1/dataproj1.csv", header = TRUE)  
View(mydata)  
  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

library(knitr)  
library(ggExtra)  
library(corrplot)

## corrplot 0.84 loaded

# Summary

dim(mydata)

## [1] 477 11

summary(mydata)

## Date Adj\_Close Personal\_Consumption\_Expenditure  
## 1/1/1980: 1 Min. : 102.1 Min. : 1695   
## 1/1/1981: 1 1st Qu.: 330.2 1st Qu.: 3684   
## 1/1/1982: 1 Median : 980.3 Median : 6438   
## 1/1/1983: 1 Mean : 989.4 Mean : 6993   
## 1/1/1984: 1 3rd Qu.:1366.4 3rd Qu.:10044   
## 1/1/1985: 1 Max. :2980.4 Max. :14696   
## (Other) :471   
## Medicare Personal\_Interst\_Income Personal\_Dividend\_Income  
## Min. : 33.1 Min. : 309.9 Min. : 61.3   
## 1st Qu.:103.0 1st Qu.: 799.5 1st Qu.: 165.6   
## Median :211.0 Median : 998.5 Median : 358.4   
## Mean :297.7 Mean :1009.9 Mean : 462.9   
## 3rd Qu.:500.3 3rd Qu.:1273.9 3rd Qu.: 747.0   
## Max. :814.9 Max. :1760.9 Max. :1329.4   
##   
## Real\_Disposable\_Personal\_Income Personal\_Savings\_Rate Personal\_Interest\_PMT  
## Min. :21288 Min. : 2.200 Min. : 41.3   
## 1st Qu.:27024 1st Qu.: 5.900 1st Qu.:113.7   
## Median :32646 Median : 7.200 Median :189.3   
## Mean :32602 Mean : 7.342 Mean :180.1   
## 3rd Qu.:38166 3rd Qu.: 8.600 3rd Qu.:240.4   
## Max. :45884 Max. :13.200 Max. :364.6   
##   
## Personal\_Current\_Taxes Social\_Security   
## Min. : 281.5 Min. : 108.8   
## 1st Qu.: 579.9 1st Qu.: 231.1   
## Median :1029.5 Median : 383.2   
## Mean :1045.7 Mean : 460.5   
## 3rd Qu.:1455.4 3rd Qu.: 673.0   
## Max. :2223.1 Max. :1040.3   
##

# Missing data

#Missing data  
apply(mydata,2,function(x) sum(is.na(x)))

## Date Adj\_Close   
## 0 0   
## Personal\_Consumption\_Expenditure Medicare   
## 0 0   
## Personal\_Interst\_Income Personal\_Dividend\_Income   
## 0 0   
## Real\_Disposable\_Personal\_Income Personal\_Savings\_Rate   
## 0 0   
## Personal\_Interest\_PMT Personal\_Current\_Taxes   
## 0 0   
## Social\_Security   
## 0

#Function to Calculate Percent Missing by Row/Column  
percentmiss = function(x) { sum(is.na(x))/length(x)\*100 }  
  
#Percent of Missing Values for each Row Summed into a Frequency Table  
missing\_all=apply(mydata, 1, percentmiss)  
table(missing\_all)

## missing\_all  
## 0   
## 477

There is no missing data.

# Outliers

mahal = mahalanobis(mydata[ , -1],   
 colMeans(mydata[ , -1], na.rm = TRUE),  
 cov(mydata[ , -1], use = "pairwise.complete.obs"))  
cutoff = qchisq(1-.001, ncol(mydata[, -1]))  
cutoff

## [1] 29.5883

ncol(mydata[,-1])

## [1] 10

summary(mahal < cutoff)

## Mode FALSE TRUE   
## logical 4 473

# Distribution

head(mydata)

## Date Adj\_Close Personal\_Consumption\_Expenditure Medicare  
## 1 1/1/1980 114.16 1697.3 33.1  
## 2 2/1/1980 113.66 1701.4 33.6  
## 3 3/1/1980 102.09 1708.2 34.2  
## 4 4/1/1980 106.29 1695.2 34.7  
## 5 5/1/1980 111.24 1700.1 35.3  
## 6 6/1/1980 114.24 1718.8 35.9  
## Personal\_Interst\_Income Personal\_Dividend\_Income  
## 1 309.9 61.3  
## 2 316.2 61.3  
## 3 321.4 62.4  
## 4 323.6 63.4  
## 5 327.2 64.2  
## 6 330.3 64.9  
## Real\_Disposable\_Personal\_Income Personal\_Savings\_Rate Personal\_Interest\_PMT  
## 1 21720 9.9 41.6  
## 2 21585 10.1 42.7  
## 3 21430 10.2 43.2  
## 4 21388 11.3 42.4  
## 5 21288 11.4 42.3  
## 6 21312 11.2 42.0  
## Personal\_Current\_Taxes Social\_Security  
## 1 281.5 108.8  
## 2 284.6 109.7  
## 3 288.4 109.5  
## 4 289.4 109.4  
## 5 292.0 110.1  
## 6 295.1 110.4

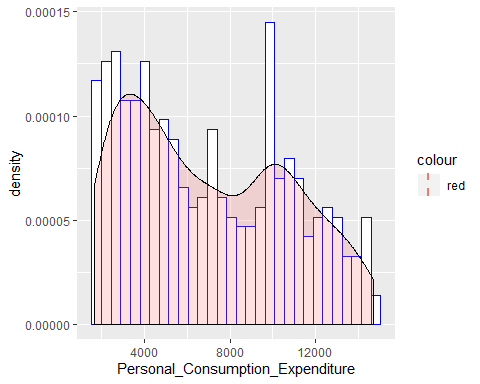
ggplot(mydata, aes(x=Personal\_Consumption\_Expenditure, Y=Adj\_Close)) +   
 geom\_histogram(aes(y=..density..), bindwidth =0.5, colour = "blue", fill = "white") +   
 geom\_density( alpha =0.2, fill = "#FF6666")+  
 geom\_vline( aes(xintercept=mean(mpg), color="red"), linetype ="dashed", size = 1)

## Warning: Ignoring unknown parameters: bindwidth

## Warning in mean.default(mpg): argument is not numeric or logical: returning NA

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 477 rows containing missing values (geom\_vline).

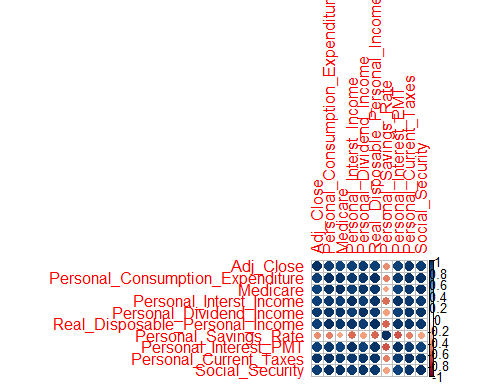


# Additivity: correlation

#Correlation Matrix  
correlation\_matrix <- cor(mydata[ , -1])  
round(correlation\_matrix,2)

## Adj\_Close Personal\_Consumption\_Expenditure  
## Adj\_Close 1.00 0.95  
## Personal\_Consumption\_Expenditure 0.95 1.00  
## Medicare 0.93 0.99  
## Personal\_Interst\_Income 0.93 0.97  
## Personal\_Dividend\_Income 0.95 0.97  
## Real\_Disposable\_Personal\_Income 0.94 0.99  
## Personal\_Savings\_Rate -0.47 -0.51  
## Personal\_Interest\_PMT 0.93 0.96  
## Personal\_Current\_Taxes 0.98 0.98  
## Social\_Security 0.94 0.99  
## Medicare Personal\_Interst\_Income  
## Adj\_Close 0.93 0.93  
## Personal\_Consumption\_Expenditure 0.99 0.97  
## Medicare 1.00 0.95  
## Personal\_Interst\_Income 0.95 1.00  
## Personal\_Dividend\_Income 0.97 0.94  
## Real\_Disposable\_Personal\_Income 0.96 0.98  
## Personal\_Savings\_Rate -0.39 -0.57  
## Personal\_Interest\_PMT 0.93 0.98  
## Personal\_Current\_Taxes 0.96 0.97  
## Social\_Security 1.00 0.95  
## Personal\_Dividend\_Income  
## Adj\_Close 0.95  
## Personal\_Consumption\_Expenditure 0.97  
## Medicare 0.97  
## Personal\_Interst\_Income 0.94  
## Personal\_Dividend\_Income 1.00  
## Real\_Disposable\_Personal\_Income 0.95  
## Personal\_Savings\_Rate -0.42  
## Personal\_Interest\_PMT 0.93  
## Personal\_Current\_Taxes 0.98  
## Social\_Security 0.97  
## Real\_Disposable\_Personal\_Income  
## Adj\_Close 0.94  
## Personal\_Consumption\_Expenditure 0.99  
## Medicare 0.96  
## Personal\_Interst\_Income 0.98  
## Personal\_Dividend\_Income 0.95  
## Real\_Disposable\_Personal\_Income 1.00  
## Personal\_Savings\_Rate -0.58  
## Personal\_Interest\_PMT 0.98  
## Personal\_Current\_Taxes 0.97  
## Social\_Security 0.97  
## Personal\_Savings\_Rate Personal\_Interest\_PMT  
## Adj\_Close -0.47 0.93  
## Personal\_Consumption\_Expenditure -0.51 0.96  
## Medicare -0.39 0.93  
## Personal\_Interst\_Income -0.57 0.98  
## Personal\_Dividend\_Income -0.42 0.93  
## Real\_Disposable\_Personal\_Income -0.58 0.98  
## Personal\_Savings\_Rate 1.00 -0.64  
## Personal\_Interest\_PMT -0.64 1.00  
## Personal\_Current\_Taxes -0.51 0.96  
## Social\_Security -0.41 0.93  
## Personal\_Current\_Taxes Social\_Security  
## Adj\_Close 0.98 0.94  
## Personal\_Consumption\_Expenditure 0.98 0.99  
## Medicare 0.96 1.00  
## Personal\_Interst\_Income 0.97 0.95  
## Personal\_Dividend\_Income 0.98 0.97  
## Real\_Disposable\_Personal\_Income 0.97 0.97  
## Personal\_Savings\_Rate -0.51 -0.41  
## Personal\_Interest\_PMT 0.96 0.93  
## Personal\_Current\_Taxes 1.00 0.97  
## Social\_Security 0.97 1.00

#Plot of Correlation Matrix  
corrplot(correlation\_matrix)



#Symnum BiVariate Correlation Table  
symnum(correlation\_matrix)

## A P\_C\_E M P\_I\_I P\_D R P\_S P\_I\_P P\_C\_T S  
## Adj\_Close 1   
## Personal\_Consumption\_Expenditure \* 1   
## Medicare \* B 1   
## Personal\_Interst\_Income \* B \* 1   
## Personal\_Dividend\_Income B B B \* 1   
## Real\_Disposable\_Personal\_Income \* B B B B 1   
## Personal\_Savings\_Rate . . . . . . 1   
## Personal\_Interest\_PMT \* B \* B \* B , 1   
## Personal\_Current\_Taxes B B B B B B . B 1   
## Social\_Security \* B B B B B . \* B 1  
## attr(,"legend")  
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '\*' 0.95 'B' 1

# We do not meet the assumption for additivity. Some variables are colinear or multi colinear. Most variables except Personal\_Savings\_Rate have a strong correlation > 0.9.  
  
#Correlation Output for Pearson - all variables included  
cor(mydata[ , -1], use="pairwise.complete.obs", method = "pearson")

## Adj\_Close Personal\_Consumption\_Expenditure  
## Adj\_Close 1.0000000 0.9451312  
## Personal\_Consumption\_Expenditure 0.9451312 1.0000000  
## Medicare 0.9277856 0.9878936  
## Personal\_Interst\_Income 0.9315099 0.9685456  
## Personal\_Dividend\_Income 0.9540759 0.9741080  
## Real\_Disposable\_Personal\_Income 0.9364394 0.9906543  
## Personal\_Savings\_Rate -0.4678895 -0.5070630  
## Personal\_Interest\_PMT 0.9332979 0.9625234  
## Personal\_Current\_Taxes 0.9769375 0.9800688  
## Social\_Security 0.9405055 0.9909287  
## Medicare Personal\_Interst\_Income  
## Adj\_Close 0.9277856 0.9315099  
## Personal\_Consumption\_Expenditure 0.9878936 0.9685456  
## Medicare 1.0000000 0.9459383  
## Personal\_Interst\_Income 0.9459383 1.0000000  
## Personal\_Dividend\_Income 0.9739015 0.9428803  
## Real\_Disposable\_Personal\_Income 0.9603753 0.9752834  
## Personal\_Savings\_Rate -0.3920397 -0.5655979  
## Personal\_Interest\_PMT 0.9267078 0.9816611  
## Personal\_Current\_Taxes 0.9638697 0.9667459  
## Social\_Security 0.9967807 0.9506044  
## Personal\_Dividend\_Income  
## Adj\_Close 0.9540759  
## Personal\_Consumption\_Expenditure 0.9741080  
## Medicare 0.9739015  
## Personal\_Interst\_Income 0.9428803  
## Personal\_Dividend\_Income 1.0000000  
## Real\_Disposable\_Personal\_Income 0.9536107  
## Personal\_Savings\_Rate -0.4169426  
## Personal\_Interest\_PMT 0.9331148  
## Personal\_Current\_Taxes 0.9759912  
## Social\_Security 0.9724848  
## Real\_Disposable\_Personal\_Income  
## Adj\_Close 0.9364394  
## Personal\_Consumption\_Expenditure 0.9906543  
## Medicare 0.9603753  
## Personal\_Interst\_Income 0.9752834  
## Personal\_Dividend\_Income 0.9536107  
## Real\_Disposable\_Personal\_Income 1.0000000  
## Personal\_Savings\_Rate -0.5812812  
## Personal\_Interest\_PMT 0.9774687  
## Personal\_Current\_Taxes 0.9712868  
## Social\_Security 0.9678597  
## Personal\_Savings\_Rate Personal\_Interest\_PMT  
## Adj\_Close -0.4678895 0.9332979  
## Personal\_Consumption\_Expenditure -0.5070630 0.9625234  
## Medicare -0.3920397 0.9267078  
## Personal\_Interst\_Income -0.5655979 0.9816611  
## Personal\_Dividend\_Income -0.4169426 0.9331148  
## Real\_Disposable\_Personal\_Income -0.5812812 0.9774687  
## Personal\_Savings\_Rate 1.0000000 -0.6372868  
## Personal\_Interest\_PMT -0.6372868 1.0000000  
## Personal\_Current\_Taxes -0.5095444 0.9576291  
## Social\_Security -0.4069762 0.9289456  
## Personal\_Current\_Taxes Social\_Security  
## Adj\_Close 0.9769375 0.9405055  
## Personal\_Consumption\_Expenditure 0.9800688 0.9909287  
## Medicare 0.9638697 0.9967807  
## Personal\_Interst\_Income 0.9667459 0.9506044  
## Personal\_Dividend\_Income 0.9759912 0.9724848  
## Real\_Disposable\_Personal\_Income 0.9712868 0.9678597  
## Personal\_Savings\_Rate -0.5095444 -0.4069762  
## Personal\_Interest\_PMT 0.9576291 0.9289456  
## Personal\_Current\_Taxes 1.0000000 0.9717039  
## Social\_Security 0.9717039 1.0000000

#Correlation Output for Spearman  
cor(mydata[ , -1], use="pairwise.complete.obs", method = "spearman")

## Adj\_Close Personal\_Consumption\_Expenditure  
## Adj\_Close 1.0000000 0.9529989  
## Personal\_Consumption\_Expenditure 0.9529989 1.0000000  
## Medicare 0.9496997 0.9994735  
## Personal\_Interst\_Income 0.9525667 0.9750768  
## Personal\_Dividend\_Income 0.9607475 0.9899380  
## Real\_Disposable\_Personal\_Income 0.9550611 0.9977592  
## Personal\_Savings\_Rate -0.5942996 -0.5479286  
## Personal\_Interest\_PMT 0.9392945 0.9659843  
## Personal\_Current\_Taxes 0.9798700 0.9813790  
## Social\_Security 0.9506965 0.9996217  
## Medicare Personal\_Interst\_Income  
## Adj\_Close 0.9496997 0.9525667  
## Personal\_Consumption\_Expenditure 0.9994735 0.9750768  
## Medicare 1.0000000 0.9735181  
## Personal\_Interst\_Income 0.9735181 1.0000000  
## Personal\_Dividend\_Income 0.9882956 0.9765709  
## Real\_Disposable\_Personal\_Income 0.9971828 0.9763186  
## Personal\_Savings\_Rate -0.5458614 -0.5641767  
## Personal\_Interest\_PMT 0.9649663 0.9849560  
## Personal\_Current\_Taxes 0.9788818 0.9826688  
## Social\_Security 0.9996640 0.9740756  
## Personal\_Dividend\_Income  
## Adj\_Close 0.9607475  
## Personal\_Consumption\_Expenditure 0.9899380  
## Medicare 0.9882956  
## Personal\_Interst\_Income 0.9765709  
## Personal\_Dividend\_Income 1.0000000  
## Real\_Disposable\_Personal\_Income 0.9922059  
## Personal\_Savings\_Rate -0.5718060  
## Personal\_Interest\_PMT 0.9737700  
## Personal\_Current\_Taxes 0.9837875  
## Social\_Security 0.9884044  
## Real\_Disposable\_Personal\_Income  
## Adj\_Close 0.9550611  
## Personal\_Consumption\_Expenditure 0.9977592  
## Medicare 0.9971828  
## Personal\_Interst\_Income 0.9763186  
## Personal\_Dividend\_Income 0.9922059  
## Real\_Disposable\_Personal\_Income 1.0000000  
## Personal\_Savings\_Rate -0.5395296  
## Personal\_Interest\_PMT 0.9671093  
## Personal\_Current\_Taxes 0.9820136  
## Social\_Security 0.9974088  
## Personal\_Savings\_Rate Personal\_Interest\_PMT  
## Adj\_Close -0.5942996 0.9392945  
## Personal\_Consumption\_Expenditure -0.5479286 0.9659843  
## Medicare -0.5458614 0.9649663  
## Personal\_Interst\_Income -0.5641767 0.9849560  
## Personal\_Dividend\_Income -0.5718060 0.9737700  
## Real\_Disposable\_Personal\_Income -0.5395296 0.9671093  
## Personal\_Savings\_Rate 1.0000000 -0.6193007  
## Personal\_Interest\_PMT -0.6193007 1.0000000  
## Personal\_Current\_Taxes -0.5761230 0.9689638  
## Social\_Security -0.5453532 0.9650782  
## Personal\_Current\_Taxes Social\_Security  
## Adj\_Close 0.9798700 0.9506965  
## Personal\_Consumption\_Expenditure 0.9813790 0.9996217  
## Medicare 0.9788818 0.9996640  
## Personal\_Interst\_Income 0.9826688 0.9740756  
## Personal\_Dividend\_Income 0.9837875 0.9884044  
## Real\_Disposable\_Personal\_Income 0.9820136 0.9974088  
## Personal\_Savings\_Rate -0.5761230 -0.5453532  
## Personal\_Interest\_PMT 0.9689638 0.9650782  
## Personal\_Current\_Taxes 1.0000000 0.9794486  
## Social\_Security 0.9794486 1.0000000

#Correlation Output for Kendall  
cor(mydata[ , -1], use="pairwise.complete.obs", method = "kendall")

## Adj\_Close Personal\_Consumption\_Expenditure  
## Adj\_Close 1.0000000 0.8571454  
## Personal\_Consumption\_Expenditure 0.8571454 1.0000000  
## Medicare 0.8510496 0.9910546  
## Personal\_Interst\_Income 0.8307500 0.8776483  
## Personal\_Dividend\_Income 0.8700949 0.9464847  
## Real\_Disposable\_Personal\_Income 0.8521846 0.9700277  
## Personal\_Savings\_Rate -0.4400990 -0.3903632  
## Personal\_Interest\_PMT 0.8136113 0.8642722  
## Personal\_Current\_Taxes 0.8982862 0.9145973  
## Social\_Security 0.8527457 0.9897283  
## Medicare Personal\_Interst\_Income  
## Adj\_Close 0.8510496 0.8307500  
## Personal\_Consumption\_Expenditure 0.9910546 0.8776483  
## Medicare 1.0000000 0.8709361  
## Personal\_Interst\_Income 0.8709361 1.0000000  
## Personal\_Dividend\_Income 0.9417657 0.8817725  
## Real\_Disposable\_Personal\_Income 0.9649747 0.8792952  
## Personal\_Savings\_Rate -0.3866944 -0.4040362  
## Personal\_Interest\_PMT 0.8581046 0.9093593  
## Personal\_Current\_Taxes 0.9078160 0.8972267  
## Social\_Security 0.9905077 0.8716643  
## Personal\_Dividend\_Income  
## Adj\_Close 0.8700949  
## Personal\_Consumption\_Expenditure 0.9464847  
## Medicare 0.9417657  
## Personal\_Interst\_Income 0.8817725  
## Personal\_Dividend\_Income 1.0000000  
## Real\_Disposable\_Personal\_Income 0.9427078  
## Personal\_Savings\_Rate -0.4127126  
## Personal\_Interest\_PMT 0.8774546  
## Personal\_Current\_Taxes 0.9150942  
## Social\_Security 0.9399431  
## Real\_Disposable\_Personal\_Income  
## Adj\_Close 0.8521846  
## Personal\_Consumption\_Expenditure 0.9700277  
## Medicare 0.9649747  
## Personal\_Interst\_Income 0.8792952  
## Personal\_Dividend\_Income 0.9427078  
## Real\_Disposable\_Personal\_Income 1.0000000  
## Personal\_Savings\_Rate -0.3805512  
## Personal\_Interest\_PMT 0.8648087  
## Personal\_Current\_Taxes 0.9058331  
## Social\_Security 0.9668007  
## Personal\_Savings\_Rate Personal\_Interest\_PMT  
## Adj\_Close -0.4400990 0.8136113  
## Personal\_Consumption\_Expenditure -0.3903632 0.8642722  
## Medicare -0.3866944 0.8581046  
## Personal\_Interst\_Income -0.4040362 0.9093593  
## Personal\_Dividend\_Income -0.4127126 0.8774546  
## Real\_Disposable\_Personal\_Income -0.3805512 0.8648087  
## Personal\_Savings\_Rate 1.0000000 -0.4573959  
## Personal\_Interest\_PMT -0.4573959 1.0000000  
## Personal\_Current\_Taxes -0.4217347 0.8660725  
## Social\_Security -0.3858384 0.8584357  
## Personal\_Current\_Taxes Social\_Security  
## Adj\_Close 0.8982862 0.8527457  
## Personal\_Consumption\_Expenditure 0.9145973 0.9897283  
## Medicare 0.9078160 0.9905077  
## Personal\_Interst\_Income 0.8972267 0.8716643  
## Personal\_Dividend\_Income 0.9150942 0.9399431  
## Real\_Disposable\_Personal\_Income 0.9058331 0.9668007  
## Personal\_Savings\_Rate -0.4217347 -0.3858384  
## Personal\_Interest\_PMT 0.8660725 0.8584357  
## Personal\_Current\_Taxes 1.0000000 0.9082203  
## Social\_Security 0.9082203 1.0000000

# Regression model

library(MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

Multiple\_lin1 <- lm(Adj\_Close~Personal\_Consumption\_Expenditure +Medicare + Personal\_Interst\_Income + Personal\_Dividend\_Income + Real\_Disposable\_Personal\_Income + Personal\_Savings\_Rate + Personal\_Interest\_PMT + Personal\_Current\_Taxes + Social\_Security, data = mydata)  
summary(Multiple\_lin1)

##   
## Call:  
## lm(formula = Adj\_Close ~ Personal\_Consumption\_Expenditure + Medicare +   
## Personal\_Interst\_Income + Personal\_Dividend\_Income + Real\_Disposable\_Personal\_Income +   
## Personal\_Savings\_Rate + Personal\_Interest\_PMT + Personal\_Current\_Taxes +   
## Social\_Security, data = mydata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -411.94 -59.94 0.46 52.07 362.88   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 324.08132 218.06296 1.486 0.138   
## Personal\_Consumption\_Expenditure -0.03545 0.05188 -0.683 0.495   
## Medicare -6.57362 0.51397 -12.790 < 2e-16 \*\*\*  
## Personal\_Interst\_Income -1.15544 0.10310 -11.207 < 2e-16 \*\*\*  
## Personal\_Dividend\_Income 0.55034 0.09008 6.109 2.11e-09 \*\*\*  
## Real\_Disposable\_Personal\_Income -0.06636 0.01414 -4.693 3.55e-06 \*\*\*  
## Personal\_Savings\_Rate 30.53300 5.72892 5.330 1.53e-07 \*\*\*  
## Personal\_Interest\_PMT 9.39676 0.50384 18.650 < 2e-16 \*\*\*  
## Personal\_Current\_Taxes 1.05707 0.07232 14.618 < 2e-16 \*\*\*  
## Social\_Security 6.34717 0.40760 15.572 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 105.2 on 467 degrees of freedom  
## Multiple R-squared: 0.9796, Adjusted R-squared: 0.9792   
## F-statistic: 2496 on 9 and 467 DF, p-value: < 2.2e-16

confint(Multiple\_lin1)

## 2.5 % 97.5 %  
## (Intercept) -104.42478107 752.58741575  
## Personal\_Consumption\_Expenditure -0.13739433 0.06650155  
## Medicare -7.58360319 -5.56364243  
## Personal\_Interst\_Income -1.35802495 -0.95284978  
## Personal\_Dividend\_Income 0.37333038 0.72735512  
## Real\_Disposable\_Personal\_Income -0.09414983 -0.03857289  
## Personal\_Savings\_Rate 19.27535069 41.79064795  
## Personal\_Interest\_PMT 8.40667412 10.38683636  
## Personal\_Current\_Taxes 0.91496526 1.19917159  
## Social\_Security 5.54621325 7.14813595

Multiple\_lin2 <- lm(Adj\_Close~Medicare + Personal\_Interst\_Income + Personal\_Dividend\_Income + Real\_Disposable\_Personal\_Income + Personal\_Savings\_Rate + Personal\_Interest\_PMT + Personal\_Current\_Taxes + Social\_Security, data = mydata)  
summary(Multiple\_lin2)

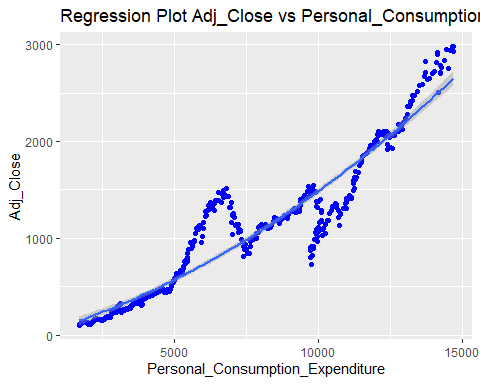
##   
## Call:  
## lm(formula = Adj\_Close ~ Medicare + Personal\_Interst\_Income +   
## Personal\_Dividend\_Income + Real\_Disposable\_Personal\_Income +   
## Personal\_Savings\_Rate + Personal\_Interest\_PMT + Personal\_Current\_Taxes +   
## Social\_Security, data = mydata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -408.46 -62.50 -0.40 55.76 357.97   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 430.801425 152.075724 2.833 0.00481 \*\*   
## Medicare -6.809802 0.380140 -17.914 < 2e-16 \*\*\*  
## Personal\_Interst\_Income -1.123695 0.091981 -12.217 < 2e-16 \*\*\*  
## Personal\_Dividend\_Income 0.536805 0.087824 6.112 2.07e-09 \*\*\*  
## Real\_Disposable\_Personal\_Income -0.074856 0.006733 -11.118 < 2e-16 \*\*\*  
## Personal\_Savings\_Rate 32.847667 4.617237 7.114 4.25e-12 \*\*\*  
## Personal\_Interest\_PMT 9.377462 0.502765 18.652 < 2e-16 \*\*\*  
## Personal\_Current\_Taxes 1.044211 0.069784 14.963 < 2e-16 \*\*\*  
## Social\_Security 6.275119 0.393498 15.947 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 105.1 on 468 degrees of freedom  
## Multiple R-squared: 0.9796, Adjusted R-squared: 0.9793   
## F-statistic: 2811 on 8 and 468 DF, p-value: < 2.2e-16

confint(Multiple\_lin2)

## 2.5 % 97.5 %  
## (Intercept) 131.96565400 729.63719673  
## Medicare -7.55679403 -6.06281079  
## Personal\_Interst\_Income -1.30444182 -0.94294826  
## Personal\_Dividend\_Income 0.36422716 0.70938190  
## Real\_Disposable\_Personal\_Income -0.08808691 -0.06162577  
## Personal\_Savings\_Rate 23.77458552 41.92074913  
## Personal\_Interest\_PMT 8.38950544 10.36541761  
## Personal\_Current\_Taxes 0.90708277 1.18133992  
## Social\_Security 5.50187790 7.04836089

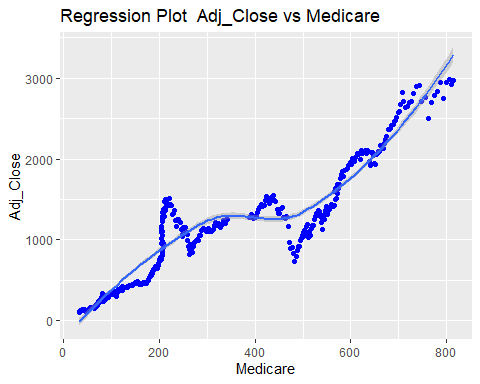
# Regression plots

library(ggplot2)  
  
w <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Consumption\_Expenditure)) + geom\_point(colour="blue")  
w<- w + stat\_smooth(method="lm", formula = y~poly(x,2))+ ggtitle("Regression Plot Adj\_Close vs Personal\_Consumption\_Expenditure")  
w

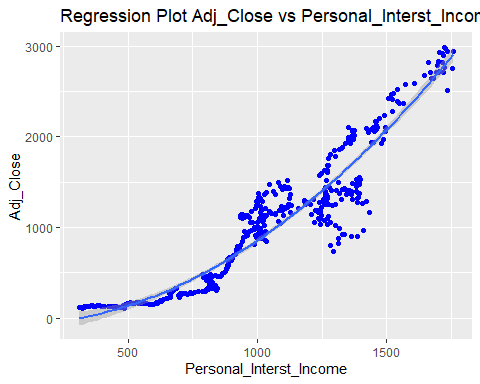


wa <- ggplot(mydata, aes(y=Adj\_Close, x=Medicare)) + geom\_point(colour="blue")  
wa<- wa + stat\_smooth()+ ggtitle("Regression Plot Adj\_Close vs Medicare")  
wa

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

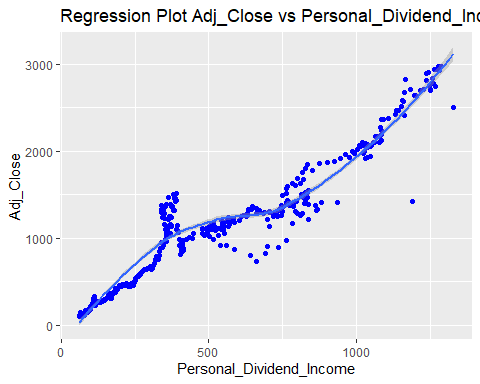


g <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Interst\_Income)) + geom\_point(colour="blue")  
g<- g + stat\_smooth(method="lm", formula = y~poly(x,2))+ ggtitle("Regression Plot Adj\_Close vs Personal\_Interst\_Income")  
g

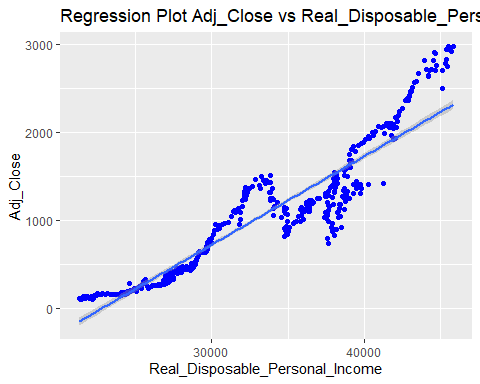


gl <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Dividend\_Income)) + geom\_point(colour="blue")  
gl<- gl + stat\_smooth()+ ggtitle("Regression Plot Adj\_Close vs Personal\_Dividend\_Income")  
gl

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

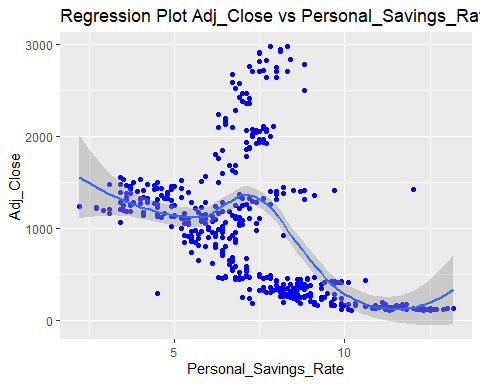


am <- ggplot(mydata, aes(y=Adj\_Close, x=Real\_Disposable\_Personal\_Income)) + geom\_point(colour="blue")  
am<- am + stat\_smooth(method="lm", formula = y~x)+ ggtitle("Regression Plot Adj\_Close vs Real\_Disposable\_Personal\_Income")  
am

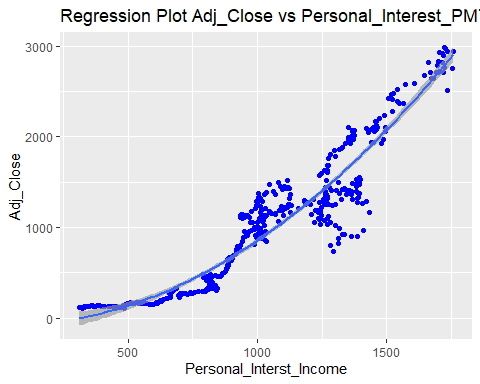


ama <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Savings\_Rate)) + geom\_point(colour="blue")  
ama<- ama + stat\_smooth()+ ggtitle("Regression Plot Adj\_Close vs Personal\_Savings\_Rate")  
ama

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

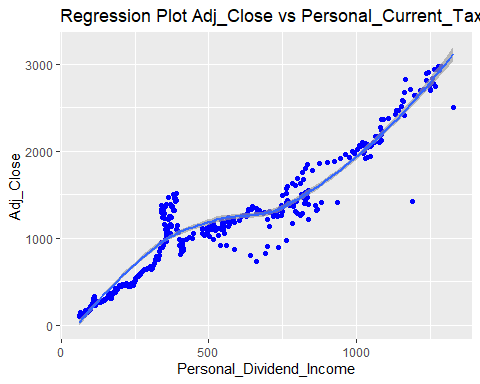


nm <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Interest\_PMT)) + geom\_point(colour="blue")  
nm<- g + stat\_smooth(method="lm", formula = y~poly(x,2))+ ggtitle("Regression Plot Adj\_Close vs Personal\_Interest\_PMT")  
nm

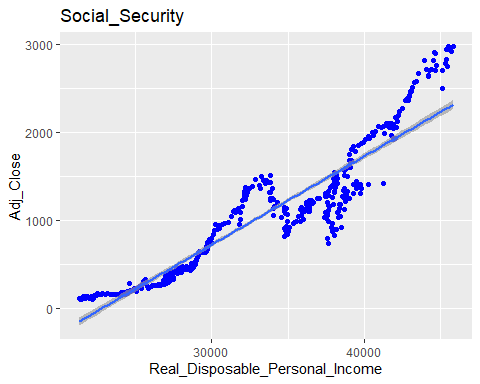


glb <- ggplot(mydata, aes(y=Adj\_Close, x=Personal\_Current\_Taxes)) + geom\_point(colour="blue")  
glb<- gl + stat\_smooth()+ ggtitle("Regression Plot Adj\_Close vs Personal\_Current\_Taxes")  
glb

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'  
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



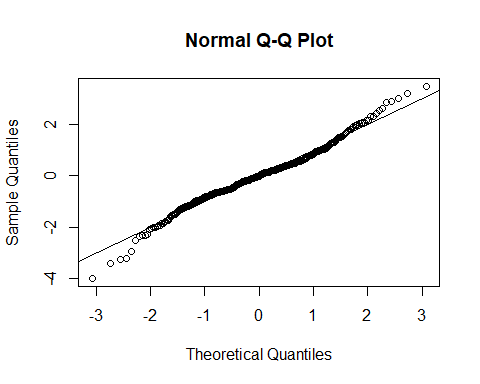
m <- ggplot(mydata, aes(y=Adj\_Close, x=Social\_Security)) + geom\_point(colour="blue")  
m<- am + stat\_smooth(method="lm", formula = y~x)+ ggtitle("Social\_Security")  
m



# Verify assumptions for the model Multiple\_lin2:

## Linearity:

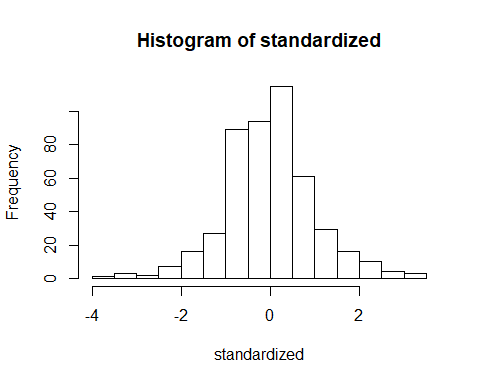
final = mydata  
random = rchisq(nrow(final), 7)  
fake = lm(random ~ ., data = final)  
standardized = rstudent(Multiple\_lin2)  
fitted = scale(Multiple\_lin2$fitted.values)  
{qqnorm(standardized)  
abline(0,1)}



The assumption for linearity is met since the Normal Q-Q plot is nearly linear.

## Normality:

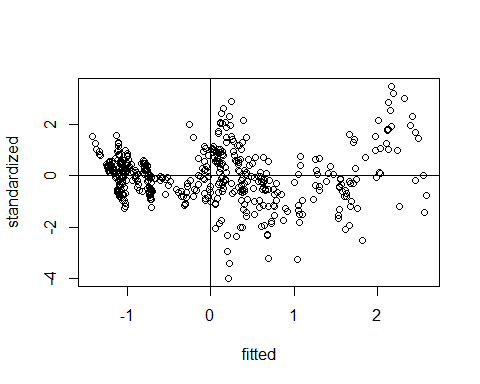
hist(standardized, breaks = 15)



The assumption for normality is met is the histogram presents a normal distribution.

## Homogeneity and Homoscedasticity:

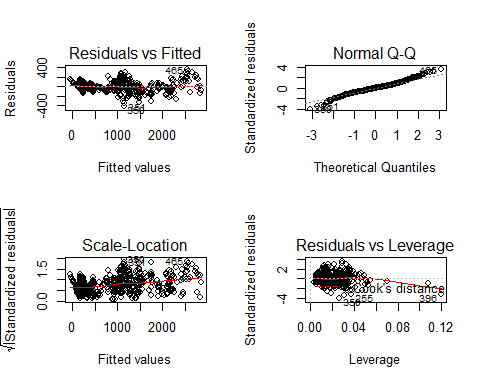
{plot(fitted, standardized)  
abline(0,0)  
abline(v = 0)}



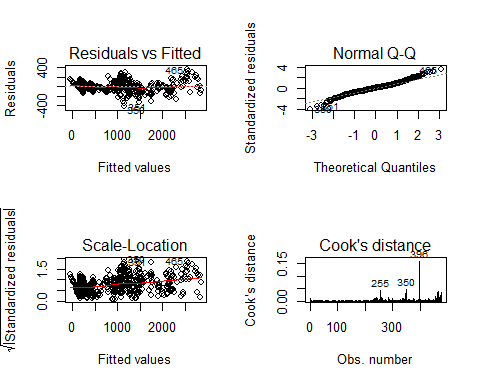
The assumption for homogeneity is met because the plot shows that the spread is nearly consistent across the ranges of values. The assumption for homoscedasticity is met since the variance around the regression line is similar for all values of the predictor variable.

# Assumptions for model Multiple\_lin2

# Diagnostics plots to check for linearity, normality, homoscedasticity, homogeneity, influential values, outliers and high leverage points  
par(mfrow = c(2, 2))  
plot(Multiple\_lin2)



plot(Multiple\_lin2, 1) # Residuals vs Fitted  
plot(Multiple\_lin2, 2) # Normal Q-Q  
plot(Multiple\_lin2, 3) # Scale-Location  
plot(Multiple\_lin2, 4) # Cook's Distance



plot(Multiple\_lin2, 5) # Residuals vs Leverage  
  
# Normality  
standardized = rstudent(Multiple\_lin2)  
hist(standardized, breaks=15)

