In [1]:
```
.libPaths()
```

'C:/Users/james/OneDrive/Documents/R/win-library/4.0' · 'C:/R-4.0.5/library'

In [2]:
```
# create a variable for the library paths
# autocomplete press tab
lib_path = 'C:/Users/james/OneDrive/Documents/R/win-library/4.0'
```

In [3]:
```
# install libraries
install.packages('ggplot2', lib = lib_path)
install.packages('lubridate', lib = lib_path)
install.packages('GGally', lib = lib_path)
install.packages("ggpubr",lib = lib_path)
install.packages("qpcR",lib = lib_path)
install.packages("matlib",lib = lib_path)
install.packages("car",lib = lib_path)
install.packages("dplyr ",lib = lib_path)
install.packages("ggrepel ",lib = lib_path)
install.packages("latex2exp",lib = lib_path)
install.packages("graphics",lib = lib_path)
install.packages("knitr",lib = lib_path)
install.packages("cowplot",lib = lib_path)
install.packages("gridExtra",lib = lib_path)
install.packages("patchwork",lib = lib_path)
install.packages("corrplot",lib = lib_path)
install.packages("faraway",lib = lib_path)
install.packages("ggthemes",lib = lib_path)
install.packages("lmtest",lib = lib_path)
```

```
package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'lubridate' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'GGally' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'ggpubr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'qpcR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'matlib' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'car' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
Warning message:
"package 'dplyr ' is not available for this version of R

A version of this package for your version of R might be available elsewhere,
see the ideas at
https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages"
Warning message:
"package 'ggrepel ' is not available for this version of R

A version of this package for your version of R might be available elsewhere,
see the ideas at
https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages"
package 'latex2exp' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
Warning message:
"package 'graphics' is in use and will not be installed"
package 'knitr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'cowplot' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'gridExtra' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'patchwork' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages

  There is a binary version available but the source version is later:
         binary source needs_compilation
corrplot   0.84   0.88             FALSE

installing the source package 'corrplot'

package 'faraway' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'ggthemes' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
package 'lmtest' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\james\AppData\Local\Temp\RtmpY5fBBA\downloaded_packages
```

In [4]:
```
# load libraries
library('ggplot2')
library('lubridate')
library('GGally')
library('ggpubr')
library('qpcR')
library('matlib')
library('car')
library('dplyr')
library('ggrepel')
library('latex2exp')
library('graphics')
library('rgl')
library('cowplot')
library('gridExtra')
library('patchwork')
library('corrplot')
library('faraway')
library('ggthemes')
library('lmtest')
```

```
Attaching package: 'lubridate'


The following objects are masked from 'package:base':

    date, intersect, setdiff, union


Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2

Loading required package: MASS

Loading required package: minpack.lm

Loading required package: rgl

Loading required package: robustbase

Loading required package: Matrix


Attaching package: 'matlib'


The following object is masked from 'package:rgl':

    GramSchmidt


Loading required package: carData


Attaching package: 'dplyr'


The following object is masked from 'package:car':

    recode


The following object is masked from 'package:MASS':

    select


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union



Attaching package: 'cowplot'


The following object is masked from 'package:ggpubr':

    get_legend


The following object is masked from 'package:lubridate':

    stamp



Attaching package: 'gridExtra'


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

    combine



Attaching package: 'patchwork'


The following object is masked from 'package:cowplot':

    align_plots


The following object is masked from 'package:MASS':

    area


corrplot 0.88 loaded

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


Attaching package: 'faraway'


The following objects are masked from 'package:car':

    logit, vif


The following object is masked from 'package:robustbase':

    epilepsy


The following object is masked from 'package:GGally':

    happy



Attaching package: 'ggthemes'


The following object is masked from 'package:cowplot':

    theme_map


Loading required package: zoo


Attaching package: 'zoo'


The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

In [5]:
```
# function to find all adjusted r^squared #

### input are strings ###
combination_function = function(dependent_var,independent_vars_vector,data_frame){

    # empty list
    equations = list()

    for(i in 1:length(independent_vars_vector)){

        vector_com = combn(independent_vars_vector,i)
        for (j in 1:ncol(vector_com)){

            model_f = as.formula(paste0(dependent_var,"~",paste0(vector_com[,j],collapse = "+")))
            equations = c(equations, model_f)
    }
}
    equation_output = NULL

    for(k in 1:length(equations)){

        equation = lm( equations[[k]], data= data_frame)
        terms = length(equation$coefficients)
        independent_var = c()

        independent_var[1] = paste0(dependent_var," = (",round(as.numeric(equation$coefficients[1]),4),") + ")

        for(i in 2:terms){

            independent_var[i] = paste0("(",round(as.numeric(equation$coefficients[i]),4),") ",names(equation$c

        }

        eq = paste(independent_var,collapse = "")
        adj_rsquare = summary(equation)$adj.r.squared
        equation_df = data.frame("Equation"=eq, "adjusted r^2"=adj_rsquare)
        equation_output = rbind(equation_output,equation_df)

    }

    equation_output = equation_output[order(-equation_output$adjusted.r.2),]
    return(equation_output)

    }
```

In [ ]:

Dataset #2F – Firm Profits (Source: accessed publicly and merged from a secret database)

#

#

 Description: Profit rates and Market structure of Advertising Intensive firms from (2013 – 2018) are provided.

#

#

 Variables: Firm - name of firm

 PT – (Net Income + Interest Expense)/total assets

 PE – (Net Income/shareholder equity)

 CR – Weighted concentration ratio of firm's product markets

 DRC – Dummy variable for CR > 50

 AS - Weighted average Industry advertising-to-sales ratio of firm's product markets

 CAS – Overall advertising-to-sales ratio of the firm

 GR – Weighted average percent changes in industry sales in the firm's market

 DIV2 - Firm's diversification in the more broad industries

 DIV3 - Firm's diversification in the less broad industries

 SIZ - Firms 2018 total assets (millions)

 LRSIZ – inversely proportional to market shares

#

#

 Research Question*: Find a 'best' multiple linear regression model to predict the average profit per total assets (PT)

In [28]:
```
# upload data

firm_profit_df = read.csv('2F - FirmProfit.csv', header=TRUE)
head(firm_profit_df,5)
```

A data.frame: 5 × 12

	FIRM	PT	PE	CR	DCR	AS	CAS	GR	DIV2	DIV3	SIZ	LRSIZ
	<chr>	<dbl>	<dbl>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1	Alberto-Culver	0.30	0.17	40	0	16.5	28.0	1.67	0.28	0.28	37	0.2769
2	American-Bakeries	0.09	0.06	55	1	0.9	0.4	1.16	0.00	0.00	78	0.2295
3	American-Sugar	0.11	0.07	43	0	0.2	0.5	1.13	0.00	0.00	269	0.1787
4	Anheuser-Busch	0.13	0.10	80	1	3.3	2.6	1.37	0.00	0.13	366	0.1694
5	Armour	0.09	0.06	31	0	0.2	0.4	1.34	0.34	0.35	588	0.1568

In [116…
```
####################################################################
######################## PE boxplot ################################
pe.box = ggplot(data=firm_profit_df, aes(x=PE)) +

geom_boxplot(fill='red') +

#### lable names ####
labs(x='PE',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## PE histogram ################################
pe.his = ggplot(data=firm_profit_df, aes(x=PE))+

#histogram
geom_histogram(bins = 15, fill = 'red', color = 'black')+

# lable names
labs(x='PE', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs PE scatt ###############################
pe.scat = ggplot(data=firm_profit_df, aes(x=PE,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="red") +

#### lable names ####
labs(x='PE', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [117…
```
####################################################################
######################## CR boxplot ################################
cr.box = ggplot(data=firm_profit_df, aes(x=CR)) +

geom_boxplot(fill='orange') +

#### lable names ####
labs(x='CR',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## CR histogram ################################
cr.his = ggplot(data=firm_profit_df, aes(x=CR))+

#histogram
geom_histogram(bins = 15, fill = 'orange', color = 'black')+

# lable names
labs(x='CR', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs CR scatt ###############################
cr.scat = ggplot(data=firm_profit_df, aes(x=CR,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="orange") +

#### lable names ####
labs(x='CR', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [118…
```
####################################################################
######################## AS boxplot ################################
as.box = ggplot(data=firm_profit_df, aes(x=AS)) +

geom_boxplot(fill='yellow') +

#### lable names ####
labs(x='AS',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## AS histogram ################################
as.his = ggplot(data=firm_profit_df, aes(x=AS))+

#histogram
geom_histogram(bins = 15, fill = 'yellow', color = 'black')+

# lable names
labs(x='AS', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs AS scatt ###############################
as.scat = ggplot(data=firm_profit_df, aes(x=AS,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="yellow") +

#### lable names ####
labs(x='AS', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [119…
```
####################################################################
######################## CAS boxplot ################################
cas.box = ggplot(data=firm_profit_df, aes(x=CAS)) +

geom_boxplot(fill='green') +

#### lable names ####
labs(x='CAS',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## CAS histogram ################################
cas.his = ggplot(data=firm_profit_df, aes(x=CAS))+

#histogram
geom_histogram(bins = 15, fill = 'green', color = 'black')+

# lable names
labs(x='CAS', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs CAS scatt ###############################
cas.scat = ggplot(data=firm_profit_df, aes(x=CAS,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="green") +

#### lable names ####
labs(x='CAS', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [120…
```
####################################################################
######################## GR boxplot ################################
gr.box = ggplot(data=firm_profit_df, aes(x=GR)) +

geom_boxplot(fill='blue') +

#### lable names ####
labs(x='GR',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## GR histogram ################################
gr.his = ggplot(data=firm_profit_df, aes(x=GR))+

#histogram
geom_histogram(bins = 15, fill = 'blue', color = 'black')+

# lable names
labs(x='GR', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs GR scatt ###############################
gr.scat = ggplot(data=firm_profit_df, aes(x=GR,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="blue") +

#### lable names ####
labs(x='GR', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [121…
```
####################################################################
######################## DIV2 boxplot ################################
div2.box = ggplot(data=firm_profit_df, aes(x=DIV2)) +

geom_boxplot(fill='#4b0082') +

#### lable names ####
labs(x='DIV2',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## DIV3 histogram ################################
div2.his = ggplot(data=firm_profit_df, aes(x=DIV2))+

#histogram
geom_histogram(bins = 15, fill = '#4b0082', color = 'black')+

# lable names
labs(x='DIV2', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs DIV2 scatt ###############################
div2.scat = ggplot(data=firm_profit_df, aes(x=DIV2,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="#4b0082") +

#### lable names ####
labs(x='DIV2', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [122…
```
####################################################################
######################## DIV3 boxplot ################################
div3.box = ggplot(data=firm_profit_df, aes(x=DIV3)) +

geom_boxplot(fill='violet') +

#### lable names ####
labs(x='DIV3',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## DIV3 histogram ################################
div3.his = ggplot(data=firm_profit_df, aes(x=DIV3))+

#histogram
geom_histogram(bins = 15, fill = 'violet', color = 'black')+

# lable names
labs(x='DIV3', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs DIV3 scatt ###############################
div3.scat = ggplot(data=firm_profit_df, aes(x=DIV3,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="violet") +

#### lable names ####
labs(x='DIV3', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

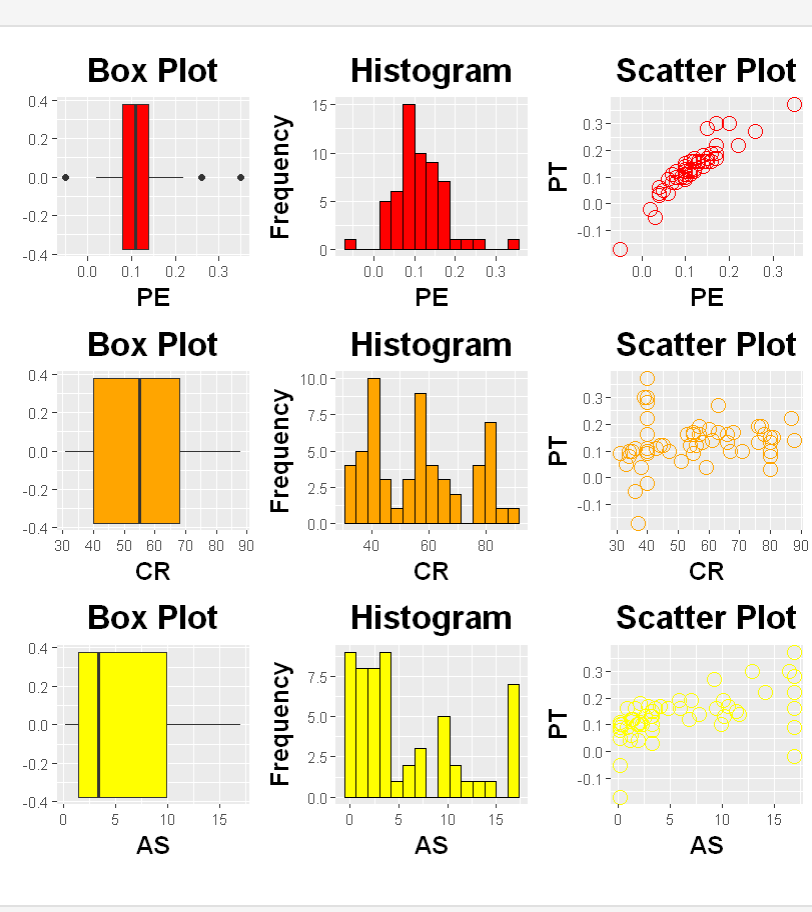
In [123…
```
####################################################################
######################## SIZ boxplot ################################
siz.box = ggplot(data=firm_profit_df, aes(x=SIZ)) +

geom_boxplot(fill='cyan') +

#### lable names ####
labs(x='SIZ',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## SIZ histogram ################################
siz.his = ggplot(data=firm_profit_df, aes(x=SIZ))+

#histogram
geom_histogram(bins = 15, fill = 'cyan', color = 'black')+

# lable names
labs(x='SIZ', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs SIZ scatt ###############################
siz.scat = ggplot(data=firm_profit_df, aes(x=SIZ,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="cyan") +

#### lable names ####
labs(x='SIZ', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [124…
```
####################################################################
######################## LRSIZ boxplot ################################
lrsiz.box = ggplot(data=firm_profit_df, aes(x=LRSIZ)) +

geom_boxplot(fill='brown') +

#### lable names ####
labs(x='LRSIZ',y='', title='Box Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
######################################################################
######################## LRSIZ histogram ################################
lrsiz.his = ggplot(data=firm_profit_df, aes(x=LRSIZ))+

#histogram
geom_histogram(bins = 15, fill = 'brown', color = 'black')+

# lable names
labs(x='LRSIZ', y='Frequency',title='Histogram')+

# title-label
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold'))
#######################################################################
######################## PT vs LRSIZ scatt ###############################
lrsiz.scat = ggplot(data=firm_profit_df, aes(x=LRSIZ,y=PT)) +

#### point settings ###
geom_point(shape=1,size=4,color="brown") +

#### lable names ####
labs(x='LRSIZ', y='PT', title='Scatter Plot') +

#### Title-label ####
theme(plot.title = element_text(hjust=.5, size=20, face='bold')) +

#### x-label ####
theme(axis.title.x = element_text(hjust=.5, size=15, face='bold')) +

#### y-label ####
theme(axis.title.y = element_text(hjust=.5, size=15, face='bold'))
```

In [ ]:

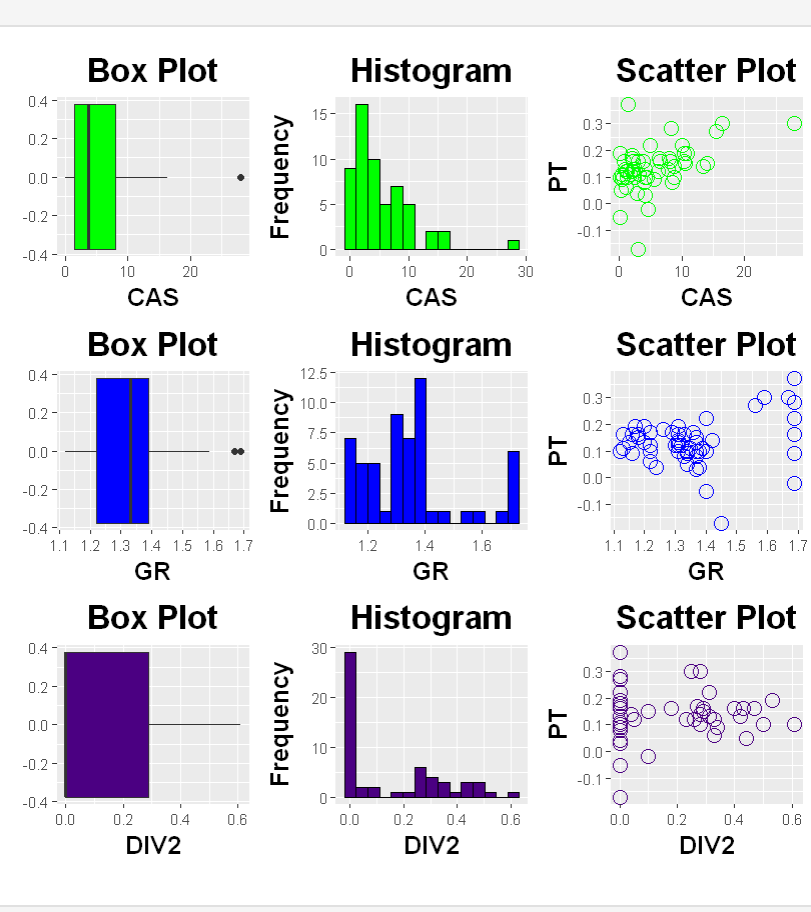
In [125…

```
pe.box+pe.his+pe.scat+
cr.box+cr.his+cr.scat+
as.box+as.his+as.scat
```



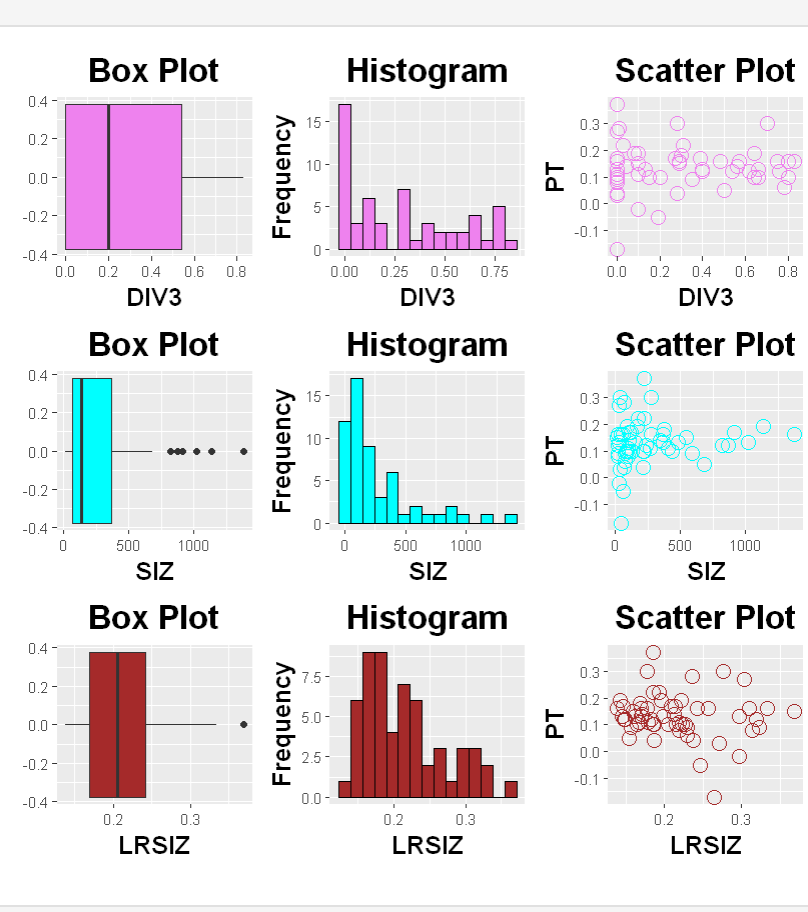
In [126…

```
cas.box+cas.his+cas.scat+
gr.box+gr.his+gr.scat+
div2.box+div2.his+div2.scat
```



In [127…

```
div3.box+div3.his+div3.scat+
siz.box+siz.his+siz.scat+
lrsiz.box+lrsiz.his+lrsiz.scat
```



In [ ]:

In [9]:

```
head(firm_profit_df,3)
```

A data.frame: 3 × 12

	FIRM	PT	PE	CR	DCR	AS	CAS	GR	DIV2	DIV3	SIZ	LRSIZ
	<chr>	<dbl>	<dbl>	<int>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1	Alberto-Culver	0.30	0.17	40	0	16.5	28.0	1.67	0.28	0.28	37	0.2769
2	American-Bakeries	0.09	0.06	55	1	0.9	0.4	1.16	0.00	0.00	78	0.2295
3	American-Sugar	0.11	0.07	43	0	0.2	0.5	1.13	0.00	0.00	269	0.1787

In [ ]:

In [29]:

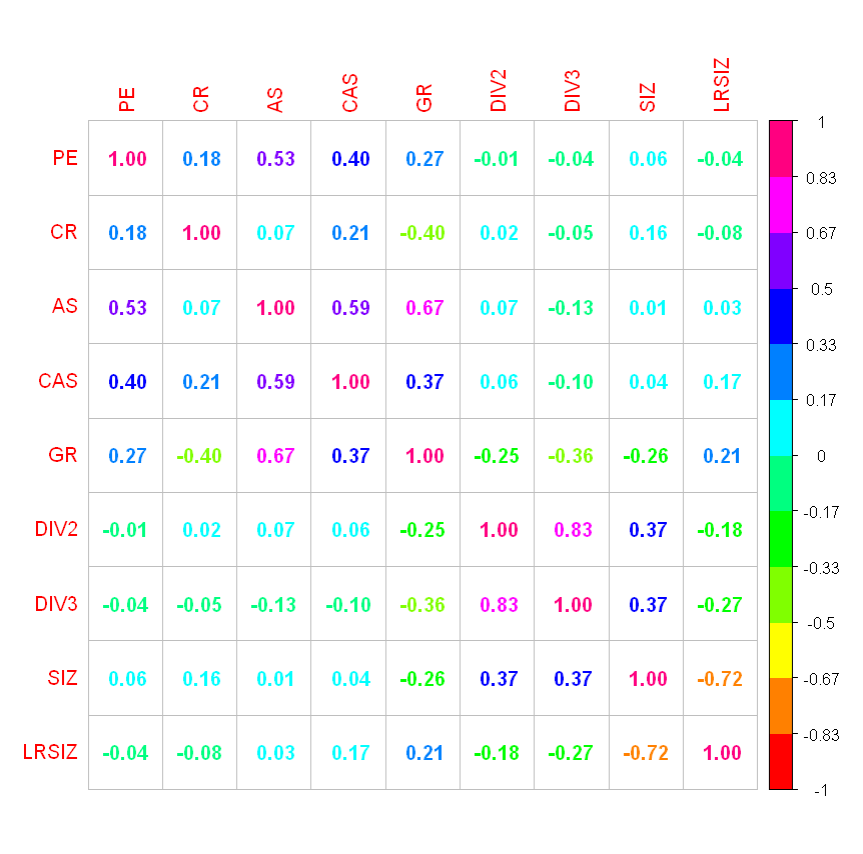
```
###### Full blown model #####
ind_vars = c('PE','CR','AS','CAS','GR','DIV2','DIV3','SIZ','LRSIZ')
model = lm(PT ~PE+CR+AS+CAS+GR+DIV2+DIV3+SIZ+LRSIZ,data=firm_profit_df)

######### eigenn values #########
X = model.matrix(model)[,-1]
eig = eigen(t(X)%*%X)
eig$val
K = sqrt(eig$val[1]/eig$val)
```

9573480.63155159 · 109519.064603014 · 2604.70192550467 · 657.56145197888 · 10.8370390898704 · 4.99823468751705 · 0.413937853275911 · 0.137651174941192 · 0.0793350163191742

In [ ]:

In [30]:

```
### corr maxtrix trail 1###

X = model.matrix(model)[,-1] # remove the intercept part
M = cor(X)
corrplot(M, method="number",col = rainbow(12)) ### AS variable will be removed
```



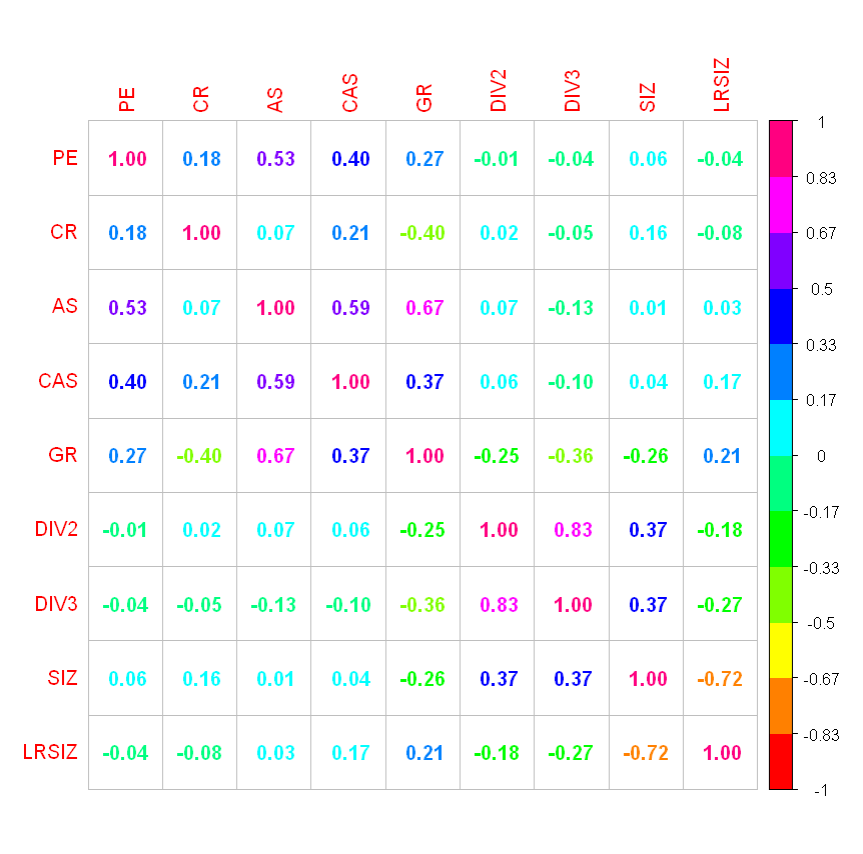
In [132…

```
model = lm(PT ~PE+CR+CAS+GR+DIV2+DIV3+SIZ+LRSIZ,data=firm_profit_df)
round(vif(model),3)
```

PE: 1.323 **CR:** 1.674 **CAS:** 1.713 **GR:** 2.155 **DIV2:** 3.525 **DIV3:** 3.835 **SIZ:** 2.604 **LRSIZ:** 2.397

In [31]:

```
### corr maxtrix trail 2###

X = model.matrix(model)[,-1] # remove the intercept part
M = cor(X)
corrplot(M, method="number",col = rainbow(12)) ### DIV3 variable will be removed
```

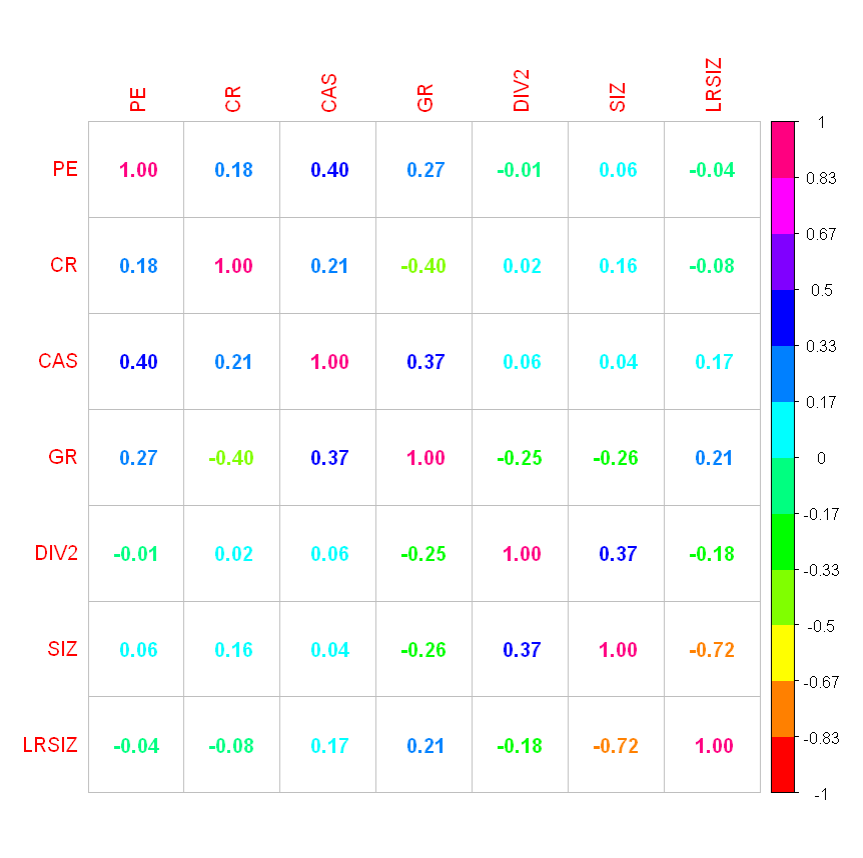


In [32]:

```
### corr maxtrix trail 3###

model = lm(PT ~PE+CR+CAS+GR+DIV2+SIZ+LRSIZ,data=firm_profit_df)
round(vif(model),3)

### corr maxtrix trail 3###

X = model.matrix(model)[,-1] # remove the intercept part
M = cor(X)
corrplot(M, method="number",col = rainbow(12)) ### SIZ variable will be removed

X = model.matrix(model)[,-1]
eig = eigen(t(X)%*%X)
eig$val
```

PE: 1.307 **CR:** 1.575 **CAS:** 1.705 **GR:** 1.984 **DIV2:** 1.281 **SIZ:** 2.603 **LRSIZ:** 2.356

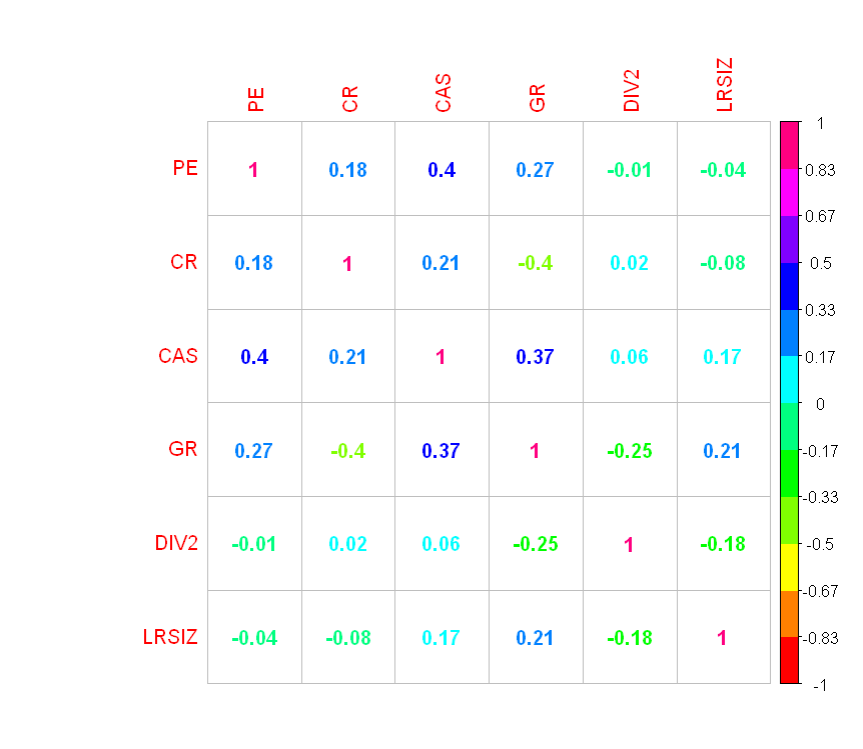
9572627.78965759 · 108528.06170664 · 1451.27764860324 · 12.1258002822566 · 1.71515761884628 · 0.170696562350851 · 0.092362595029722



In [135…

```
model = lm(PT ~PE+CR+CAS+GR+DIV2+LRSIZ,data=firm_profit_df)
round(vif(model),3)

### corr maxtrix trail 4###

X = model.matrix(model)[,-1] # remove the intercept part
M = cor(X)
corrplot(M, method="number",col = rainbow(12)) ### Looks Great!
```

PE: 1.306 **CR:** 1.575 **CAS:** 1.614 **GR:** 1.942 **DIV2:** 1.178 **LRSIZ:** 1.109



In [11]:

```
ind_vars = c('PE','CR','CAS','GR','DIV2','LRSIZ')
results = combination_function('PT',ind_vars,firm_profit_df)
```

In [34]:

```
##### top 16 from the list #####
head(results,16)
```

A data.frame: 16 × 2

	Equation	adjusted.r.2
	<chr>	<dbl>
63	PT = (0.0806) + (1.2197) PE(-4e-04) CR(0.0029) CAS(-0.0385) GR(0.0276) DIV2(-0.1448) LRSIZ	0.8645994
50	PT = (0.0093) + (1.1922) PE(0.0022) CAS(0.0366) DIV2(-0.1431) LRSIZ	0.8644472
59	PT = (0.0231) + (1.1988) PE(-2e-04) CR(0.0024) CAS(0.0363) DIV2(-0.1513) LRSIZ	0.8641758
58	PT = (0.1084) + (1.2222) PE(-5e-04) CR(0.0031) CAS(-0.0521) GR(-0.159) LRSIZ	0.8637812
61	PT = (0.022) + (1.1965) PE(0.0023) CAS(-0.0104) GR(0.0343) DIV2(-0.1396) LRSIZ	0.8621212
28	PT = (0.0202) + (1.1848) PE(0.0024) CAS(-0.1675) LRSIZ	0.8604218
44	PT = (0.0342) + (1.1915) PE(-2e-04) CR(0.0025) CAS(-0.1757) LRSIZ	0.8601874
49	PT = (0.0456) + (1.1948) PE(0.0026) CAS(-0.0221) GR(-0.1569) LRSIZ	0.8593117
27	PT = (-0.023) + (1.211) PE(0.0019) CAS(0.0452) DIV2	0.8579144
57	PT = (0.0535) + (1.2398) PE(-4e-04) CR(0.0025) CAS(-0.0439) GR(0.035) DIV2	0.8577760
43	PT = (-0.0139) + (1.2168) PE(-2e-04) CR(0.002) CAS(0.0453) DIV2	0.8565631
48	PT = (-5e-04) + (1.2174) PE(0.002) CAS(-0.0174) GR(0.041) DIV2	0.8561045
42	PT = (0.0862) + (1.2456) PE(-5e-04) CR(0.0029) CAS(-0.0623) GR	0.8548928
31	PT = (0.0022) + (1.2706) PE(0.0431) DIV2(-0.1018) LRSIZ	0.8511662
26	PT = (0.0249) + (1.2185) PE(0.0023) CAS(-0.0328) GR	0.8509660
8	PT = (-0.0161) + (1.2055) PE(0.002) CAS	0.8502786

In [35]:

```
### all models that are sig at 0.05 ###

model = lm(PT~PE+CAS+LRSIZ,firm_profit_df) # best model
#model = lm(PT~PE+CAS,firm_profit_df)
#model3 =lm(PT~PE+DIV2,firm_profit_df)
#model4 = lm(PT~PE,firm_profit_df)
#model5 = lm(PT~CAS,firm_profit_df)
#model6 = lm(PT~CR+GR,firm_profit_df)
summary(model)
```

```
Call:
lm(formula = PT ~ PE + CAS + LRSIZ, data = firm_profit_df)

Residuals:
      Min        1Q    Median        3Q       Max
-0.094078 -0.017058 -0.000543  0.019753  0.101697

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.020178   0.018659   1.081   0.2844
PE           1.184759   0.076206  15.547   <2e-16 ***
CAS          0.002391   0.000902   2.651   0.0106 *
LRSIZ       -0.167515   0.075489  -2.219   0.0308 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03154 on 53 degrees of freedom
Multiple R-squared:  0.8679,    Adjusted R-squared:  0.8604
F-statistic: 116.1 on 3 and 53 DF,  p-value: < 2.2e-16
```

In [28]:

In [14]:

```
# getting the residuals
residuals = model$residuals

# standardize residuals
stand_residuals = rstandard(model)

# jackknife residuals
jack_residuals = rstudent(model)

#find Cook's distance for each observation
cooks_distance = cooks.distance(model)

#diangoal of the hat matrix
leverage = hatvalues(model)

#make a dataframe of interest

#empty dataframe
df = data.frame()


#add cols of interest
n = length(firm_profit_df$LRSIZ)
df = cbind( c(1:n),firm_profit_df$PE,firm_profit_df$CAS,firm_profit_df$LRSIZ,firm_profit_df$PT,fitted(model),mo


#cols name
col_labels=c('observations',"PE(x2)","CAS(x6)","LRSIZ(x11)","PT(y)",'PT(yhat)',"residuals","stand_residuals","j
colnames(df) = col_labels

#convert matrix to dataframe
df = data.frame(df)
```

In [ ]:

In [37]:

```
head(df,5)
```

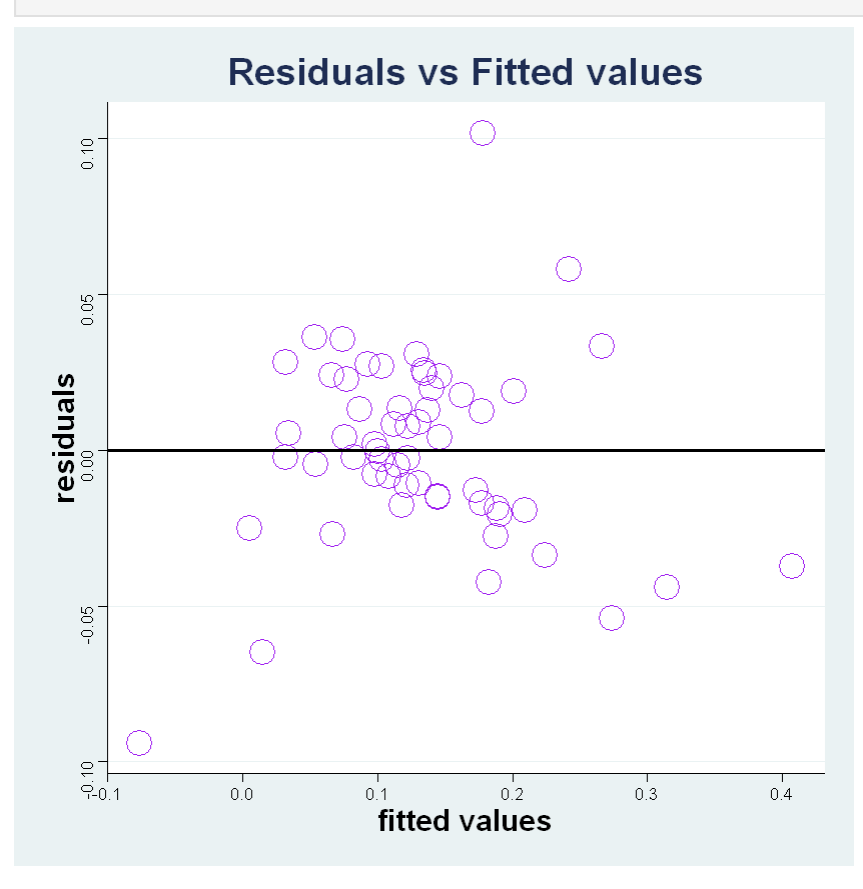
A data.frame: 5 × 11

	observations	PE.x2.	CAS.x6.	LRSIZ.x11.	PT.y.	PT.yhat.	residuals	stand_residuals	jackknife_residuals	cooks_distance	leverage
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	0.17	28.0	0.2769	0.30	0.24214416	0.05785584	2.3179516	2.4220269	0.801446577	0.37369200
2	2	0.06	0.4	0.2295	0.09	0.05377549	0.03622451	1.1726884	1.1769421	0.014596637	0.04072766
3	3	0.07	0.5	0.1787	0.11	0.07437192	0.03562808	1.1531593	1.1568333	0.013981737	0.04035997
4	4	0.10	2.6	0.1694	0.13	0.11649321	0.01350679	0.4352408	0.4318877	0.001557424	0.03183872
5	5	0.06	0.4	0.1568	0.09	0.06595384	0.02404616	0.7842292	0.7813422	0.008920208	0.05483485

In [ ]:

In [142…

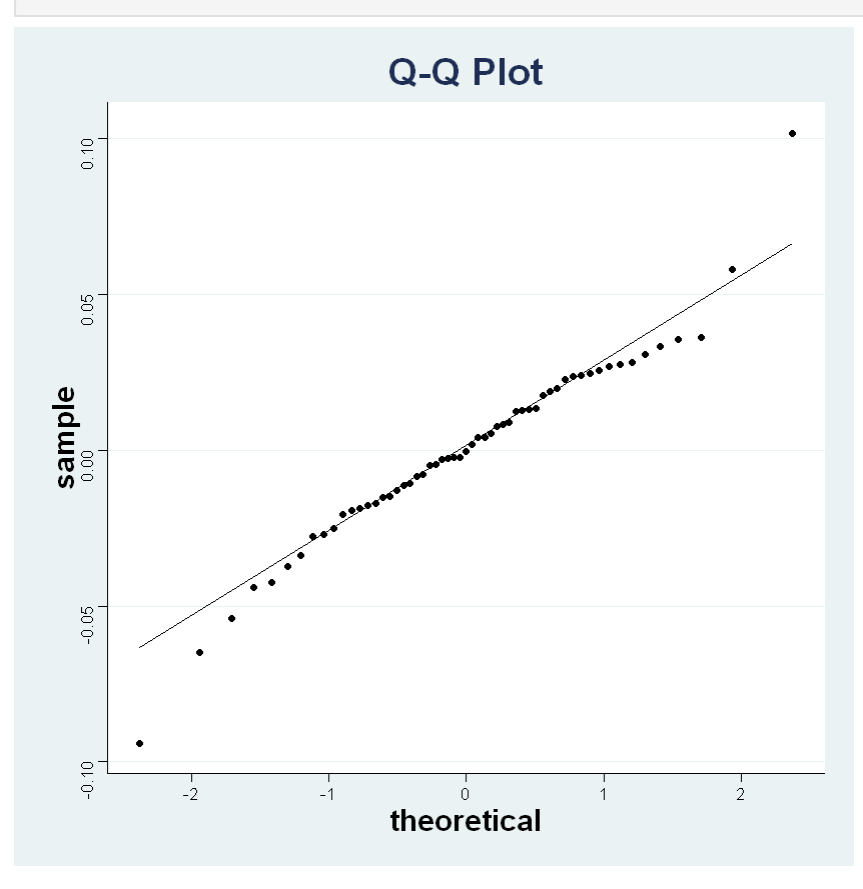
```
### Linearity ###

#upload data
ggplot(data=df, aes(x=PT.yhat., y=residuals)) +

#point settings
geom_point(shape=1,size=7,colour="purple") + # point settings

theme_stata() +


# lable names
labs(x='fitted values', y='residuals', title='Residuals vs Fitted values') +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# add hor line
geom_hline(yintercept=0, col = "black", size=1)
```



In [57]:

In [161…

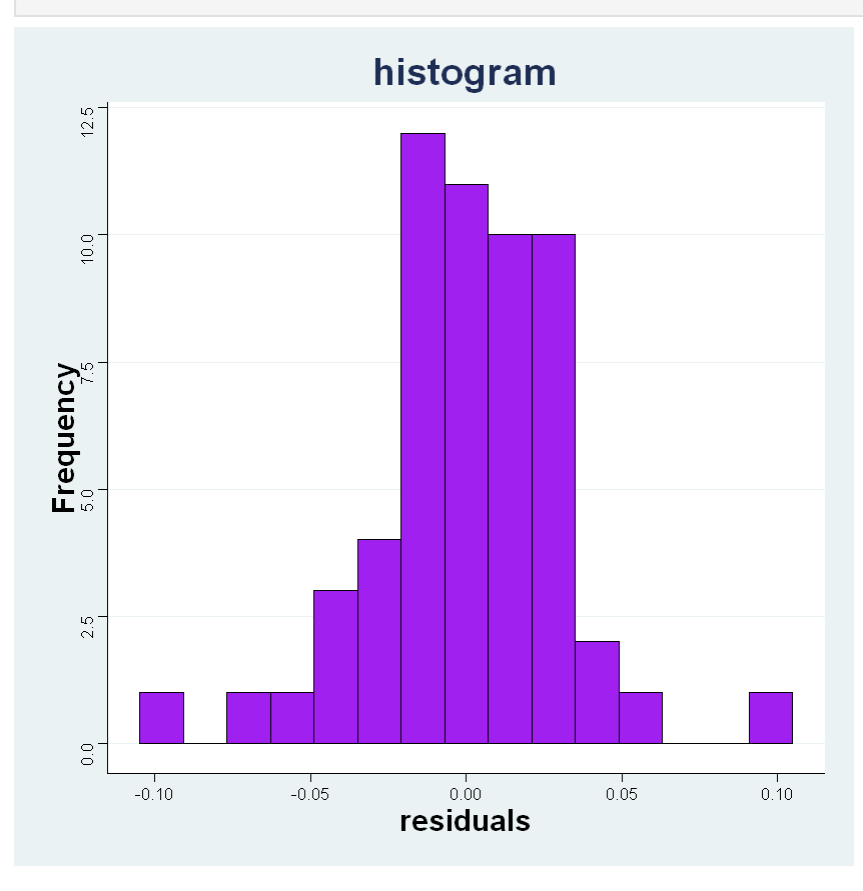
```
### normalirty1 ###

#upload data
ggplot(data=df, aes(sample=residuals)) +

# qq plot
stat_qq() +

# add line
geom_qq_line()+

scale_color_brewer(palette="Dark2")+

theme_stata() +

# lable names
labs( title='Q-Q Plot')+

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold'))  +
# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold'))
```



In [ ]:

In [160…

```
### normalirty2 ###

#upload data
ggplot(data=df, aes(x=residuals))+

#histogram
geom_histogram(bins = 15, fill = 'purple', color = 'black')+

# lable names
labs(x='residuals', y='Frequency',title='histogram')+

theme_stata() +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold'))
```



In [ ]:

In [38]:

```
### normalirty3 ###

shapiro.test(residuals)
```

```
	Shapiro-Wilk normality test

data:  residuals
W = 0.9636, p-value = 0.08402
```

In [ ]:

In [39]:

```
#### independence assumption #####
dwtest(model, alternative=c("two.sided"))
```

```
	Durbin-Watson test

data:  model
DW = 1.6177, p-value = 0.1304
alternative hypothesis: true autocorrelation is not 0
```

In [ ]:

In [156…

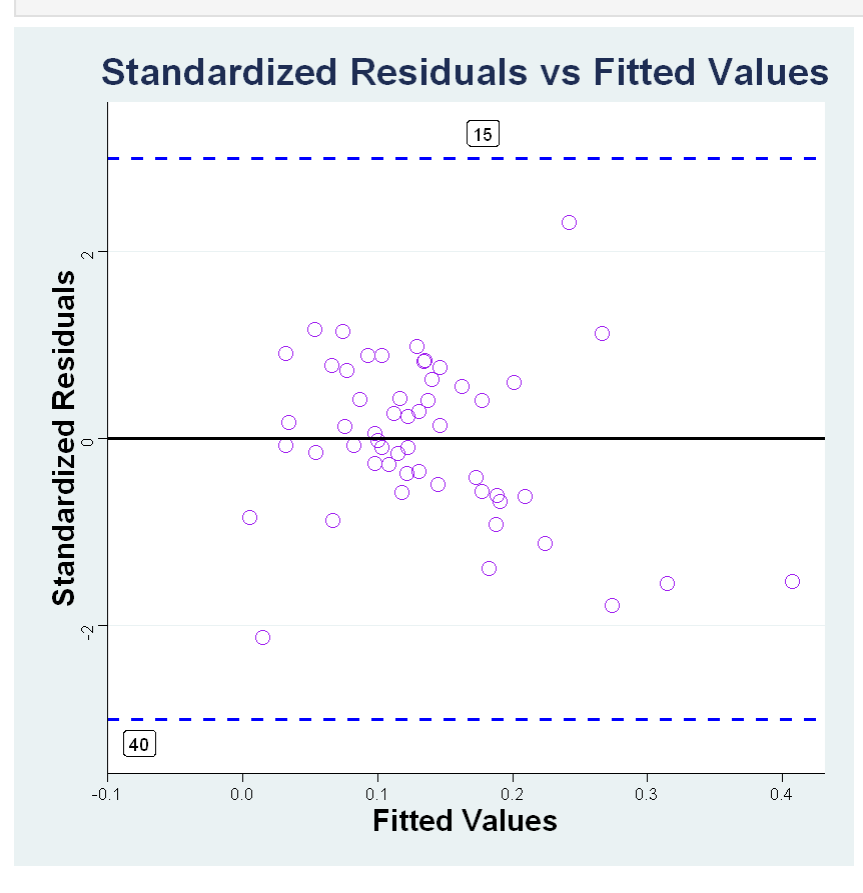
```
head(df,3)
```

A data.frame: 3 × 11

	observations	PE.x2.	CAS.x6.	LRSIZ.x11.	PT.y.	PT.yhat.	residuals	stand_residuals	jackknife_residuals	cooks_distance	leverage
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	0.17	28.0	0.2769	0.30	0.24214416	0.05785584	2.317952	2.422027	0.80144658	0.37369200
2	2	0.06	0.4	0.2295	0.09	0.05377549	0.03622451	1.172688	1.176942	0.01459664	0.04072766
3	3	0.07	0.5	0.1787	0.11	0.07437192	0.03562808	1.153159	1.156833	0.01398174	0.04035997

In [29]:

```
##### standarize residuals #####

# need to be  > 3 std

#upload data
ggplot(data=df, aes(x=PT.yhat., y=stand_residuals)) +

theme_stata() +

#point settings
geom_point(shape=1,size=4,colour="purple") + # point settings

# lable names
labs(x='Fitted Values', y='Standardized Residuals', title='Standardized Residuals vs Fitted Values') +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# add hor line
geom_hline(yintercept=0, col = "black", size=1)+

# add hor line
geom_hline(yintercept=3, linetype='dashed',col = "blue", size=1)+

# add hor line
geom_hline(yintercept=-3,linetype='dashed' ,col = "blue", size=1)+

# add label to potential outliers under conditions
geom_label(data=df %>% filter(abs(stand_residuals)>3),aes(label=observations))


#new_df = df[order(stand_residuals),]
#tail(new_df,6)
```

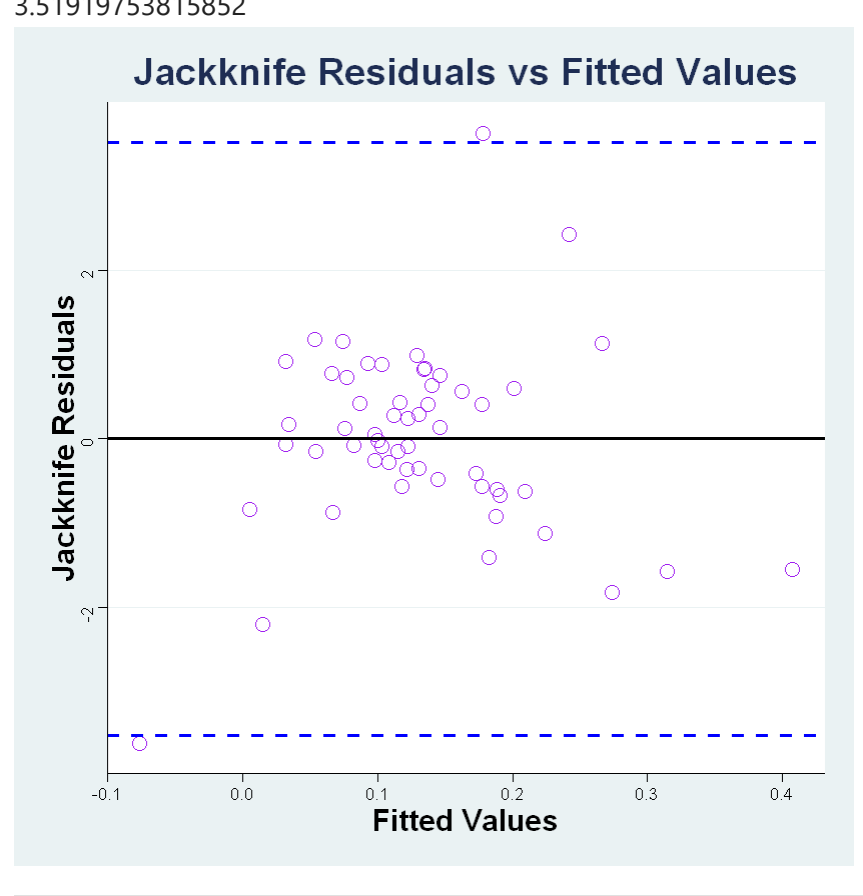


In [ ]:

In [155…

```
##### Jacknfive residual #####

#Bonferroni critical value
bonferroni=qt(.025/n,df=n-2,lower.tail = FALSE)


#upload data
ggplot(data=df, aes(x=PT.yhat., y=jackknife_residuals)) +

theme_stata() +

#point settings
geom_point(shape=1,size=4,colour="purple") + # point settings

# lable names
labs(x='Fitted Values', y='Jackknife Residuals', title='Jackknife Residuals vs Fitted Values') +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# add hor line
geom_hline(yintercept=0, col = "black", size=1)+

# add hor line bonferroni
geom_hline(yintercept=bonferroni, linetype='dashed',col = "blue", size=1)+

# add hor line bonferroni
geom_hline(yintercept=-bonferroni,linetype='dashed' ,col = "blue", size=1)+

# add label to potential outliers under conditions
geom_label( data=df %>% filter(abs(stand_residuals)>bonferroni),aes(label=observations))

bonferroni
```

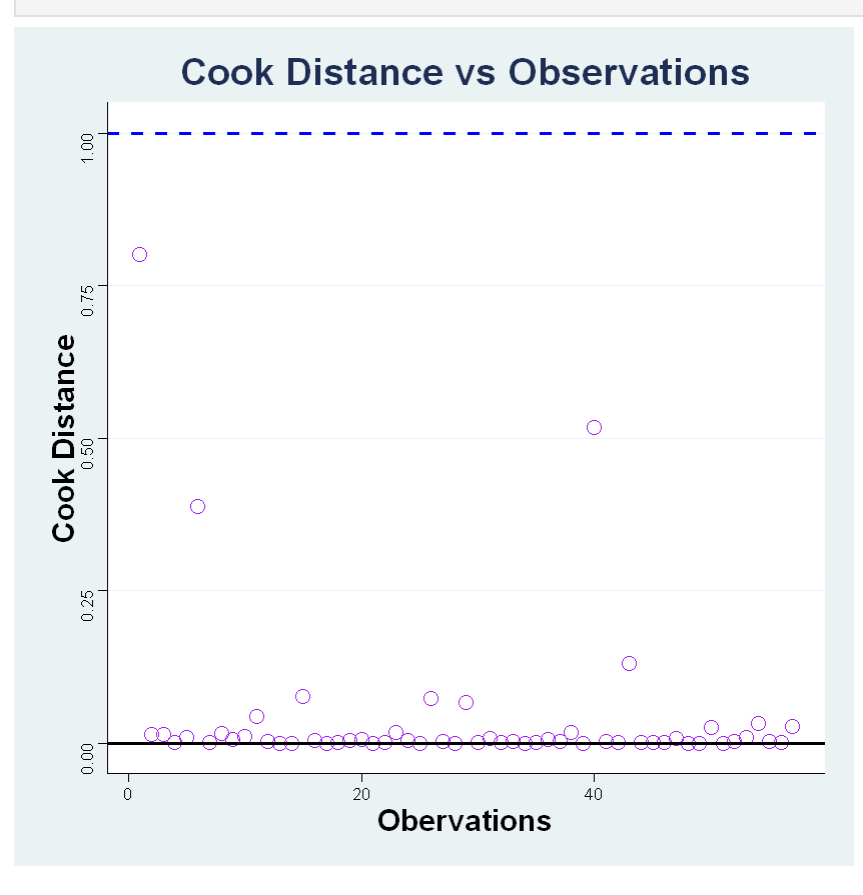
3.51919753815852



In [ ]:

In [46]:

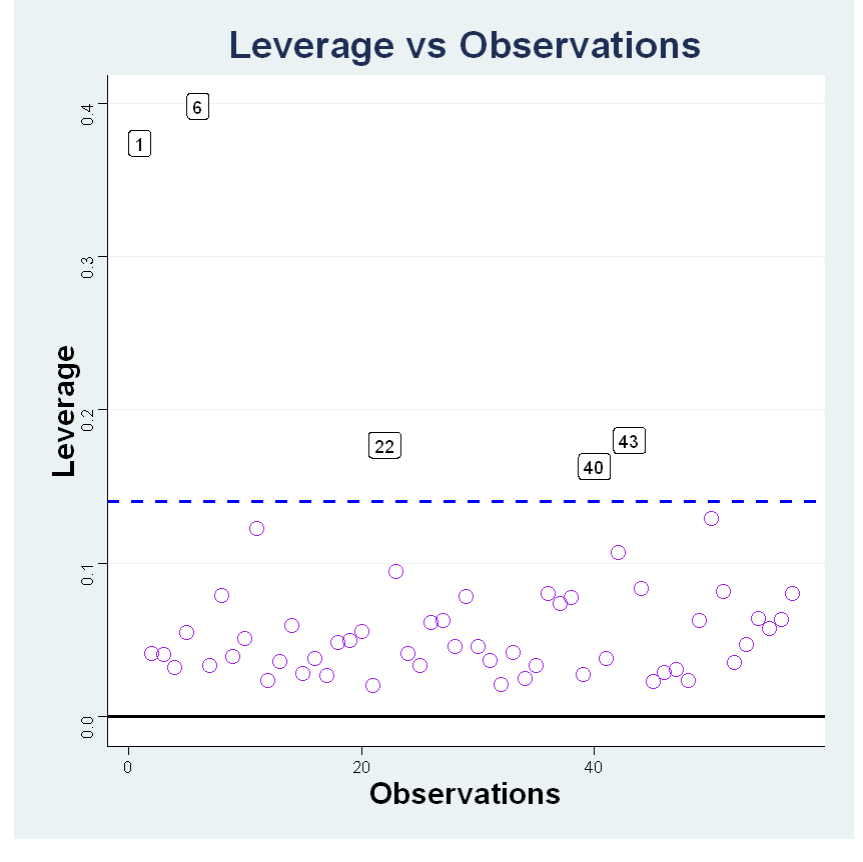
```
### cook distance ###

ggplot(data=df, aes(x=observations, y=cooks_distance)) +

#point settings
geom_point(shape=1,size=4,colour="purple") + # point settings

theme_stata() +

# lable names
labs(x='Obervations', y='Cook Distance', title='Cook Distance vs Observations') +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# add hor line
geom_hline(yintercept=0, col = "black", size=1)+

# add hor cook distance thresh hold
geom_hline(yintercept=1, linetype='dashed',col = "blue", size=1)+


# add label to potential outliers under conditions
geom_label( data=df %>% filter(cooks_distance>1),aes(label=observations))
```



In [ ]:

In [145…

```
### leverage points graph ###

# Outlier = Leverage > 2*p/n, p = rank of X

X = model.matrix(model)
p = as.integer(rankMatrix(X))
c.thes = 2*p/n

#upload data
ggplot(data=df, aes(x=observations, y=leverage)) +

#point settings
geom_point(shape=1,size=4,colour="purple") + # point settings

theme_stata() +

# lable names
labs(x='Observations', y='Leverage', title='Leverage vs Observations') +

# title-label
theme(plot.title = element_text(hjust=.5, size=23, face='bold')) +

# x-label
theme(axis.title.x = element_text(hjust=.5, size=18, face='bold')) +

# y-label
theme(axis.title.y = element_text(hjust=.5, size=18, face='bold')) +

# add hor line
geom_hline(yintercept=0, col = "black", size=1)+

# add hor line for thres hold
geom_hline(yintercept=c.thes, linetype='dashed',col = "blue", size=1)+

# add label to potential outliers under conditions
geom_label( data=df %>% filter(leverage>c.thes),aes(label=observations))
```



In [146…:

```
c.thes
```

0.140350877192982

In [ ]:

In [28]:

```
summary(model)
```

```
Call:
lm(formula = PT ~ PE + CAS + LRSIZ, data = firm_profit_df)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.094078 -0.017058 -0.000543  0.019753  0.101697 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.020178   0.018659   1.081   0.2844    
PE           1.184759   0.076206  15.547   <2e-16 ***
CAS          0.002391   0.000902   2.651   0.0106 *  
LRSIZ       -0.167515   0.075489  -2.219   0.0308 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03154 on 53 degrees of freedom
Multiple R-squared:  0.8679,	Adjusted R-squared:  0.8604 
F-statistic: 116.1 on 3 and 53 DF,  p-value: < 2.2e-16
```

In [7]:

```
Error in eval(expr, envir, enclos): object 'firm_profit_df' not found
Traceback:
```

In [ ]:

In [84]:

```
# before
summary(model)
```

```
Call:
lm(formula = PT ~ PE + CAS + LRSIZ, data = firm_profit_df)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.094078 -0.017058 -0.000543  0.019753  0.101697 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.020178   0.018659   1.081   0.2844    
PE           1.184759   0.076206  15.547   <2e-16 ***
CAS          0.002391   0.000902   2.651   0.0106 *  
LRSIZ       -0.167515   0.075489  -2.219   0.0308 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03154 on 53 degrees of freedom
Multiple R-squared:  0.8679,	Adjusted R-squared:  0.8604 
F-statistic: 116.1 on 3 and 53 DF,  p-value: < 2.2e-16
```

In [ ]:

In [24]:

```
### confindence int for population ###1

# Confidence interval
newdata=data.frame(PE=.20,CAS=4,LRSIZ=.20)

predict(model,newdata,inteval='confidence',interval="confidence",level=.95)

#we are 95% confindent that the expected _____ is between lb and hb when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.2331902	0.2167559	0.2496245

In [40]:

```
### confindence int for population ###2

# Confidence interval
newdata=data.frame(PE=.50,CAS=30,LRSIZ=.50)

predict(model,newdata,inteval='confidence',interval="confidence",level=.95)

#we are 95% confindent that the expected _____ is between lb and hb when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.6005235	0.5292747	0.6717723

In [41]:

```
### confindence int for population ###3

# Confidence interval
newdata=data.frame(PE=.75,CAS=20,LRSIZ=.45)

predict(model,newdata,inteval='confidence',interval="confidence",level=.95)

#we are 95% confindent that the expected _____ is between lb and hb when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.8811811	0.782076	0.9802862

In [ ]:

In [27]:

```
### prediction int for an indivual y-value ###1

# prediction interval
newdata=data.frame(PE=.2,CAS=4,LRSIZ=.20)

predict(model,newdata,inteval='prediction',interval="prediction",level=.95)

#we are 95% confindent that the y _____ is between lb and ub when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.2331902	0.1678309	0.2985495

In [42]:

```
### prediction int for an indivual y-value ###2

# prediction interval
newdata=data.frame(PE=.5,CAS=30,LRSIZ=.50)

predict(model,newdata,inteval='prediction',interval="prediction",level=.95)

#we are 95% confindent that the y _____ is between lb and ub when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.6005235	0.5052442	0.6958027

In [43]:

```
### prediction int for an indivual y-value ###3

# prediction interval
newdata=data.frame(PE=.75,CAS=20,LRSIZ=.45)

predict(model,newdata,inteval='prediction',interval="prediction",level=.95)

#we are 95% confindent that the y _____ is between lb and ub when x2 is __, x6 is ___, and x11 is ____.
```

A matrix: 1 × 3 of type dbl

	fit	lwr	upr
1	0.8811811	0.7636075	0.9987548

In [ ]: