Project 3

Lorraine Schemenauer

03/16/2023

#Panel Data Models ## Assessing the Data Net exports are an important variable when calculating GDP. By observing panel data for net exports, we can determine which component places a heavier emphasis on GDP. In theory, we should expect that higher exports means higher GDP since countries like the United States are able to gather revenue for their economies. Meanwhile, it is hard to calculate imports since consumption is also a part the GDP, yet it still has an influence on the GDP. If we run this data through the 3 models, we will be able to assess how imports and exports affect GDP.

```
data1 <- read.csv("~/Downloads/top_six_economies.csv", header=FALSE, skip=1)
View(data1)
#Problem 1
#Convert to Panel Data
library(plm)
library(AER)
```

```
## Loading required package: car

## Loading required package: carData

## Loading required package: lmtest

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

## Loading required package: sandwich

## Loading required package: survival
```

```
colnames(data1)<-c("ID", "Country", "Year", "GDP", "GDP.PPP", "GDP.Per.capita", "GDP.growth", "Imports"
#get specific data
data1<-data.frame(data1$Year, data1$Country, data1$ID, data1$GDP, data1$Imports, data1$Exports)
View(data1)
head(data1)
```

```
##   data1.Year data1.Country data1.ID    data1.GDP data1.Imports data1.Exports
## 1       1991 United States       33 6.158129e+12      10.12554      9.660905
## 2       1992 United States       34 6.520327e+12      10.24168      9.708915
## 3       1993 United States       35 6.858559e+12      10.49744      9.547180
## 4       1994 United States       36 7.287236e+12      11.16231      9.893147
## 5       1995 United States       37 7.639749e+12      11.81416     10.639224
## 6       1996 United States       38 8.073122e+12      11.94044     10.746636
```

```r
#change to panel data
paneldata<-pdata.frame(data1, index=c("data1.Country", "data1.Year"))
View(paneldata)
#convert imports/exports to gdp calc
paneldata$data1.Imports<-paneldata$data1.Imports*paneldata$data1.GDP/100
paneldata$data1.Exports<-paneldata$data1.Exports*paneldata$data1.GDP/100
View(paneldata)

#preparing panel data frame
for (i in 1: nrow(paneldata))
{
  if(paneldata$data1.Country [i] == "United States")
  {
    paneldata$data1.ID[i] = 1
  }
  if(paneldata$data1.Country [i] == "China")
 {
      paneldata$data1.ID[i] = 2
  }
    if(paneldata$data1.Country [i] == "Japan")
    {
      paneldata$data1.ID[i] = 3
    }
    if(paneldata$data1.Country [i] == "Germany")
    {
paneldata$data1.ID[i] = 4
}
 if(paneldata$data1.Country [i] == "United Kingdom")
{
paneldata$data1.ID[i] = 5
}
if(paneldata$data1.Country [i] == "India")
{
paneldata$data1.ID[i] = 6
}
}

View(paneldata)
```

Problem 2

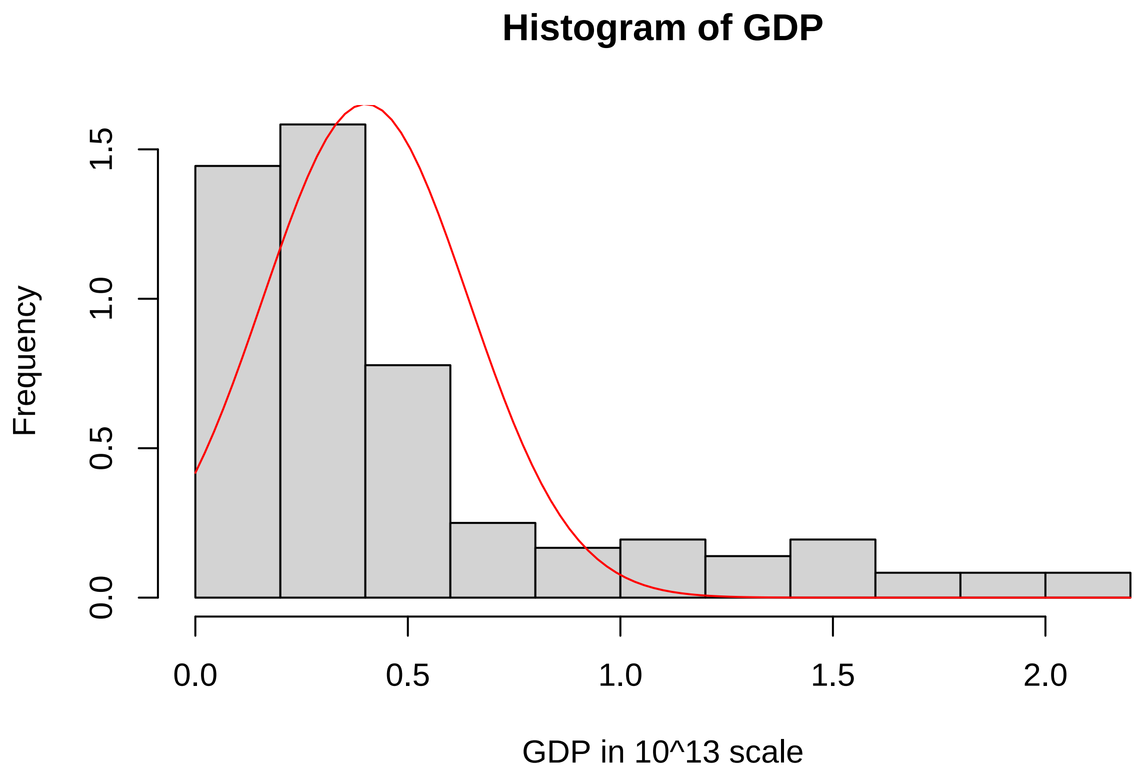
Histograms

For our histograms, we had to apply a transfromation of 10ˆ12 to each variable to create an appropriate scale. Our results showed us that each histogram is skewed right. For the histogram of GDP, all countries are distributed around the GDP mean of 10ˆ12. Based on this histogram, we can conclude that countries that are rich grow at a slower rate. For the histogram of imports, the countries are distributed around the GDP mean of about 10ˆ13. This makes sense because countries trade to maximize GDP. Lastly, in the histogram for exports, we can see that the countries are distributed around the GDP mean of 10ˆ12. This means that countries export to generate high revenue contributing to their GDP.

```
library(MASS)
hist(paneldata$data1.GDP/10^13, prob = TRUE, xlab = "GDP in 10^13 scale", ylab = "Frequency", main = "Hi
fit1<-fitdistr(paneldata$data1.GDP/10^13, densfun="logistic")
```

```
## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced
```

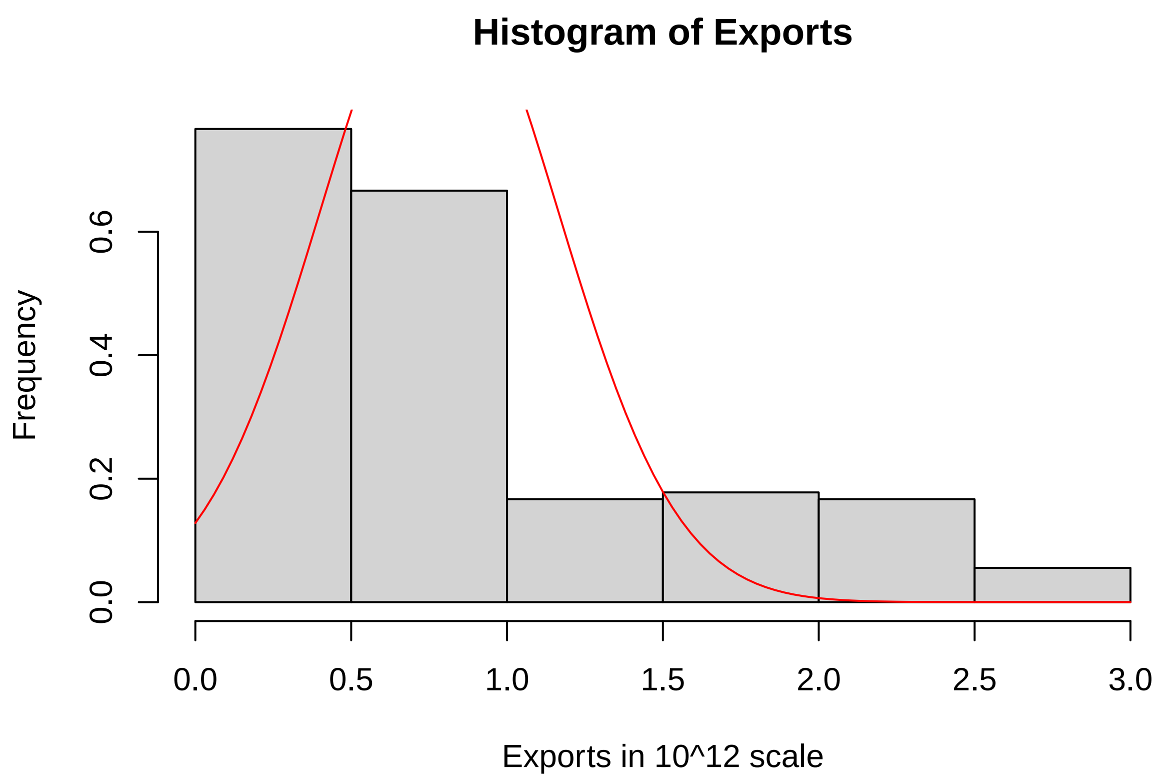
```
curve(dnorm(x,fit1$estimate[1], fit1$estimate[2]), col="red", add=T)
```




```
hist(paneldata$data1.Exports/10^12, prob = TRUE, xlab = "Exports in 10^12 scale", ylab = "Frequency", ma
fit1<-fitdistr(paneldata$data1.Exports/10^12, densfun="logistic")
```

```
## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced
```

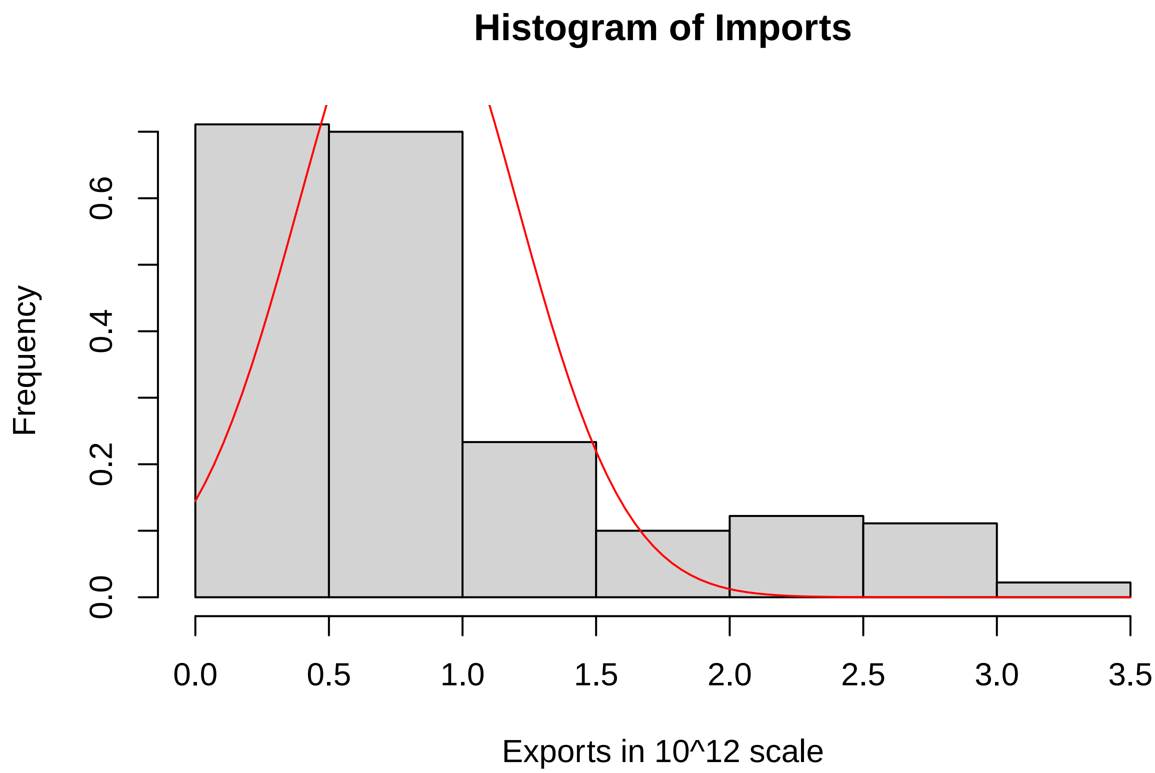
```
curve(dnorm(x,fit1$estimate[1], fit1$estimate[2]), col="red", add=T)
```




```
hist(paneldata$data1.Imports/10^12, prob = TRUE, xlab = "Exports in 10^12 scale", ylab = "Frequency", ma
fit1<-fitdistr(paneldata$data1.Imports/10^12, densfun="logistic")
```

```
## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced

## Warning in densfun(x, parm[1], parm[2], ...): NaNs produced
```

```
curve(dnorm(x,fit1$estimate[1], fit1$estimate[2]), col="red", add=T)
```



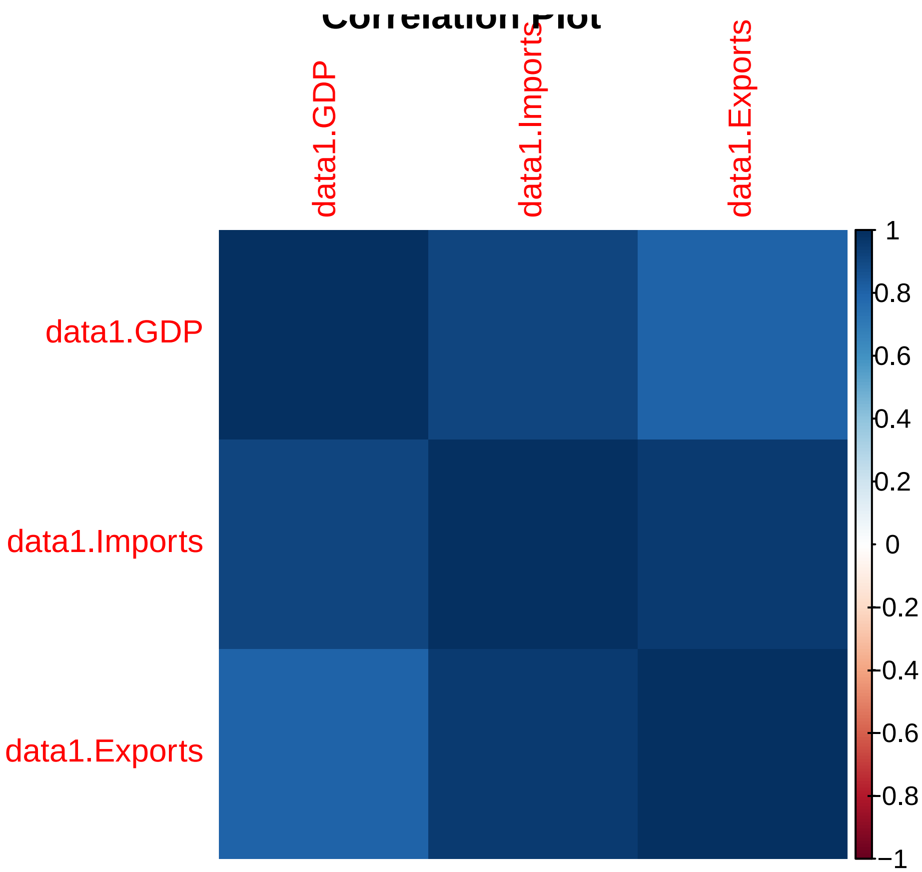
Correlation Plots

From the correlation plot, we can conclude that there is strong positive correlation given that our coefficient values are close to one. In this data, it shows that imports are more strongly correlated to GDP than exports because an economy wants to minimize the cost in trading. It also shows that with an increase in imports, there will be a increase in GDP. However, based on the actual formula for GDP, net exports is exports minus imports. There should be a negative correlation between imports and GDP.

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

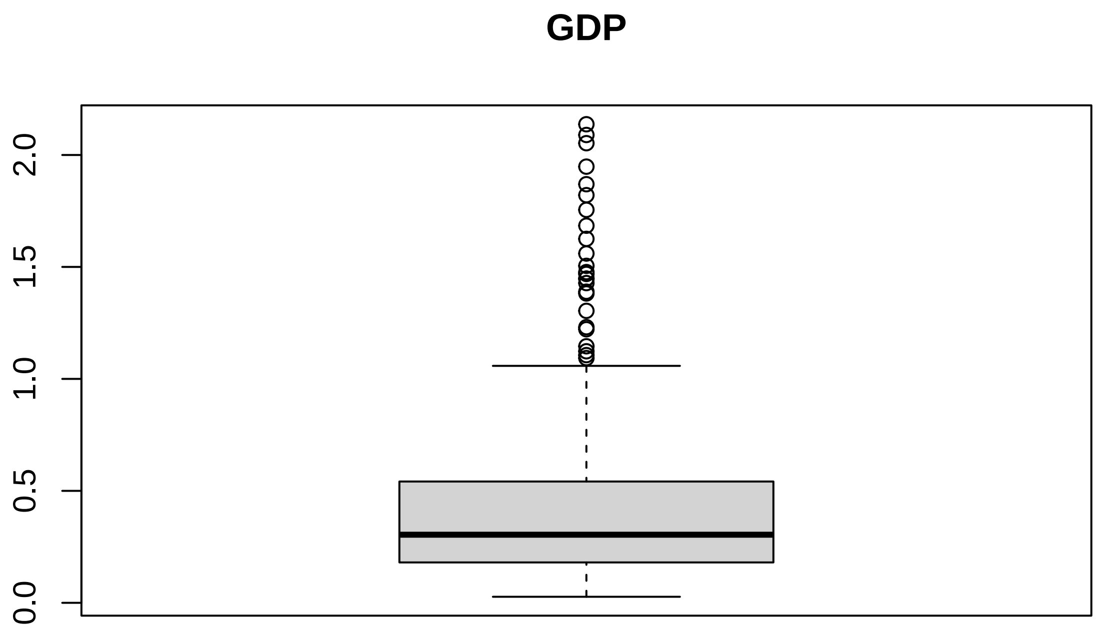
```
M1 = cor(paneldata[,4:6])
corrplot(M1, method = 'shade', main = "Correlation Plot")
```



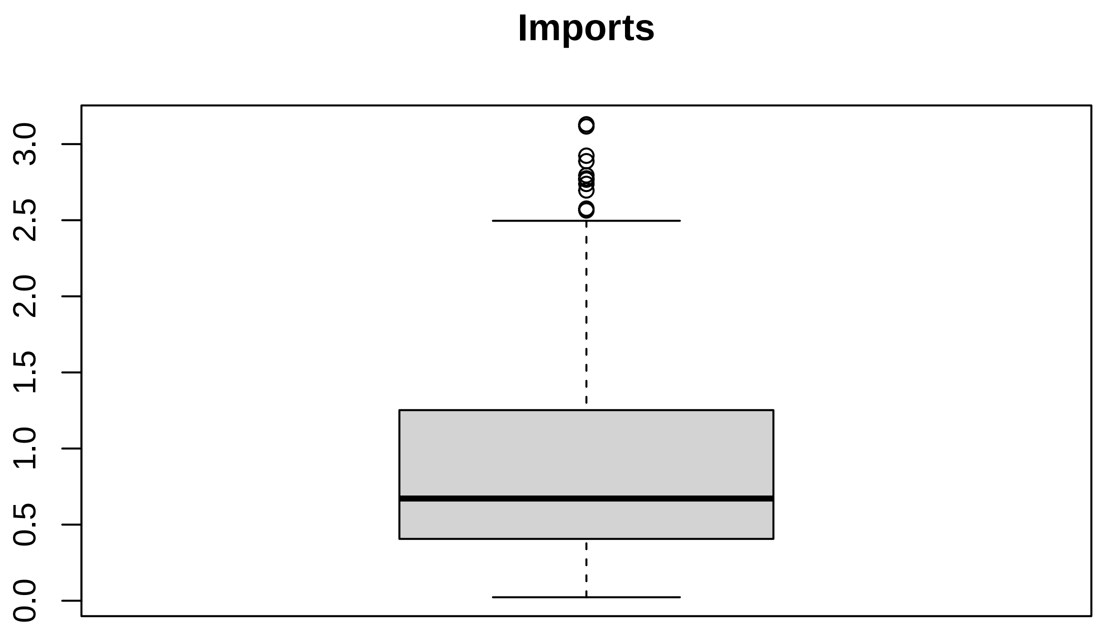
Box Plots

For the boxplot of GDP, the median is around 10^12. As well, over time there are more outliers because countries develop at different rates. This creates a gap between the most and least developed out of the top 6 countries. For the boxplot of exports, the median is around 10^11 GDP which shows that the top 6 countries can get most of its GDP mostly from exports. It also shows that over time, these countries become more reliant on exports for their GDP. For the boxplot of imports, the median is lower than exports at about 10^10 GDP. Given our dataset is the top 6 countries, it makes sense that the median for exports is higher than imports. Again, higher exports means higher GDP.

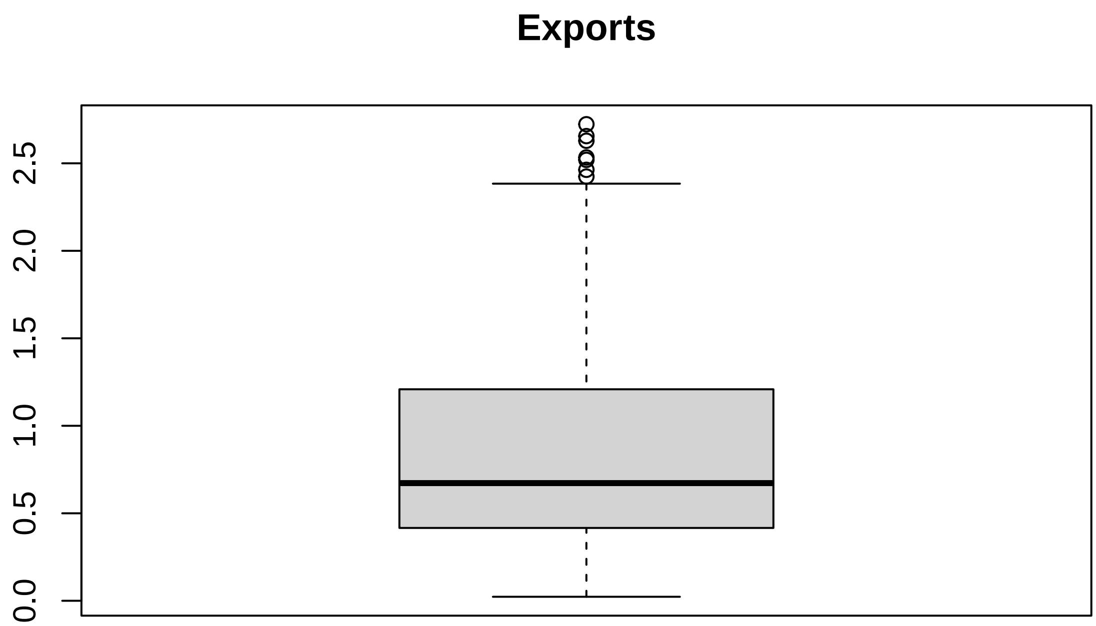
```
boxplot(paneldata$data1.GDP/10^13, main = "GDP")
```




```
boxplot(paneldata$data1.Imports/10^12, main = "Imports")
```



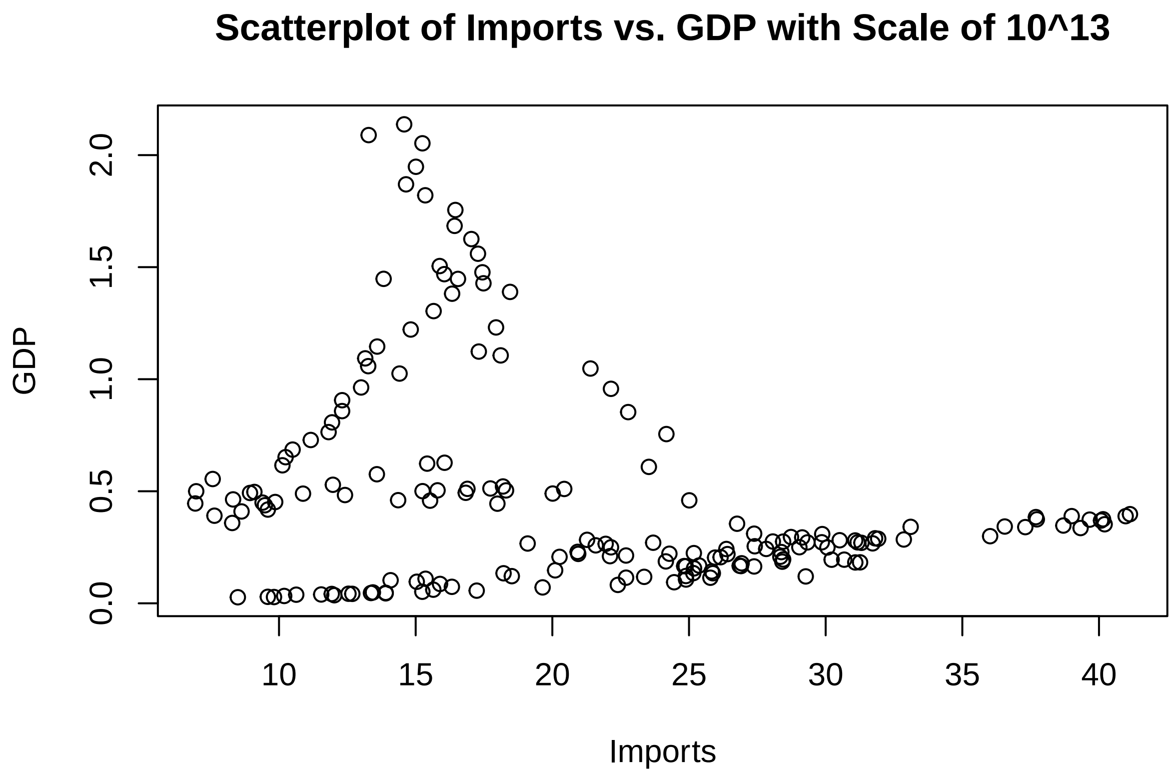

```
boxplot(paneldata$data1.Exports/10^12, main = "Exports")
```



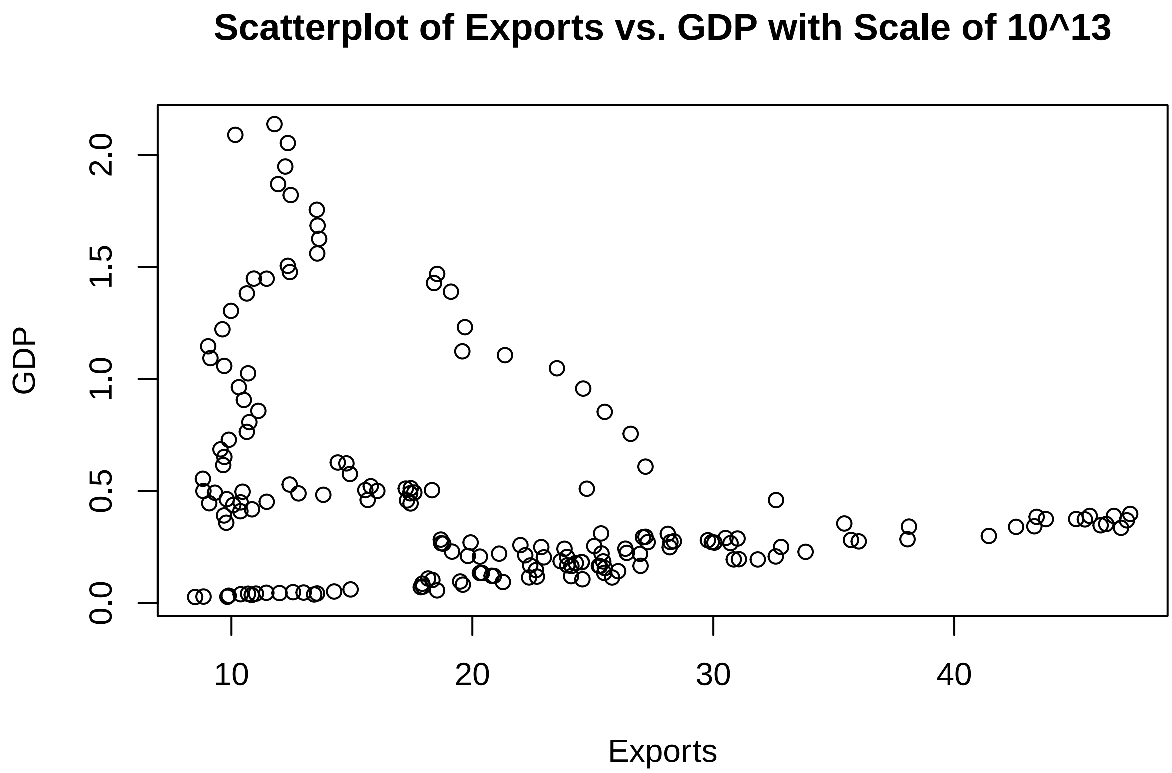
Scatter Plots

The scatter plots show the relationship between our variables imports and exports against GDP. This comparison aligns with our overall question of what the magnitude imports and exports have on GDP. Imports tend to be around between 25-30 while exports tend to be around 20-30.

```
plot(data1$data1.Imports, data1$data1.GDP/10^13, xlab="Imports", ylab="GDP", main="Scatterplot of Impor
```




```
plot(data1$data1.Exports, data1$data1.GDP/10^13, xlab="Exports", ylab="GDP", main="Scatterplot of Expor
```



Statistical Summary

The statistical summary gives us statistics of our data that can better help us understand our panel data. For GDP, the minimum is 2.701x10^11, the maximum is 2.137x10^13, and the mean is 4.915x10^12. This makes sense given we are observing the top 6 countries. For imports, the minimum is 2.294x10^10, the maximum is 3.130x10^12 and the mean is 9.106x10^11. For the exports, the minimum is 2.294x10^10, the maximum is 2.723x10^12, and the mean is 6.725x10^11. Imports has a higher mean than exports which is shocking given we are taking the top 6 GDPs.

```
#statistical summary
summary(paneldata)
```

```
##   data1.Year         data1.Country    data1.ID      data1.GDP
##  1991   :  6   China          :30    Min.   :1.0   Min.   :2.701e+11
##  1992   :  6   Germany        :30    1st Qu.:2.0   1st Qu.:1.813e+12
##  1993   :  6   India          :30    Median :3.5   Median :3.041e+12
##  1994   :  6   Japan          :30    Mean   :3.5   Mean   :4.915e+12
##  1995   :  6   United Kingdom:30    3rd Qu.:5.0   3rd Qu.:5.354e+12
##  1996   :  6   United States :30    Max.   :6.0   Max.   :2.137e+13
##  (Other):144
##  data1.Imports       data1.Exports
##  Min.   :2.294e+10   Min.   :2.294e+10
##  1st Qu.:4.088e+11   1st Qu.:4.165e+11
##  Median :6.714e+11   Median :6.725e+11
##  Mean   :9.106e+11   Mean   :8.775e+11
```

```
##   3rd Qu.:1.252e+12   3rd Qu.:1.192e+12  
##   Max.   :3.130e+12   Max.   :2.723e+12  
##
```

Problem 3

Pooled Model

Pooled model is simply our OLS model.

Coef Test

We performed the Cluster-Robust Standard errors to account for endogeneity of the time component. While the significance changed, we can note that the interpretation for the betas remains the same.

```
library(plm)
library(AER)
library(car)
library(gplots)
```

```
## 
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
## 
##     lowess
```

```
paneldata<-pdata.frame(data1, index=c("data1.Country", "data1.Year"))
poolEffect<- plm(data1.GDP~data1.Exports+data1.Imports, model="pooling", data=paneldata)
summary(poolEffect)
```

```
## Pooling Model
## 
## Call:
## plm(formula = data1.GDP ~ data1.Exports + data1.Imports, data = paneldata, 
##     model = "pooling")
## 
## Balanced Panel: n = 6, T = 30, N = 180
## 
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
## -6.30e+12 -3.01e+12 -1.25e+12  0.00e+00  2.38e+12  1.48e+13 
## 
## Coefficients:
##                  Estimate  Std. Error t-value  Pr(>|t|)    
## (Intercept)    7.5802e+12  9.4607e+11  8.0123 1.459e-13 ***
## data1.Exports -2.8294e+11  1.1532e+11 -2.4535   0.01511 *  
## data1.Imports  1.5746e+11  1.3559e+11  1.1613   0.24710    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Total Sum of Squares:    4.2109e+27
## Residual Sum of Squares: 3.7413e+27
## R-Squared:      0.11152
## Adj. R-Squared: 0.10148
## F-statistic: 11.1081 on 2 and 177 DF, p-value: 2.8535e-05
```

```
coeftest(poolEffect, vcov=vcovHC(poolEffect, type="HC0", cluster="group"))
```

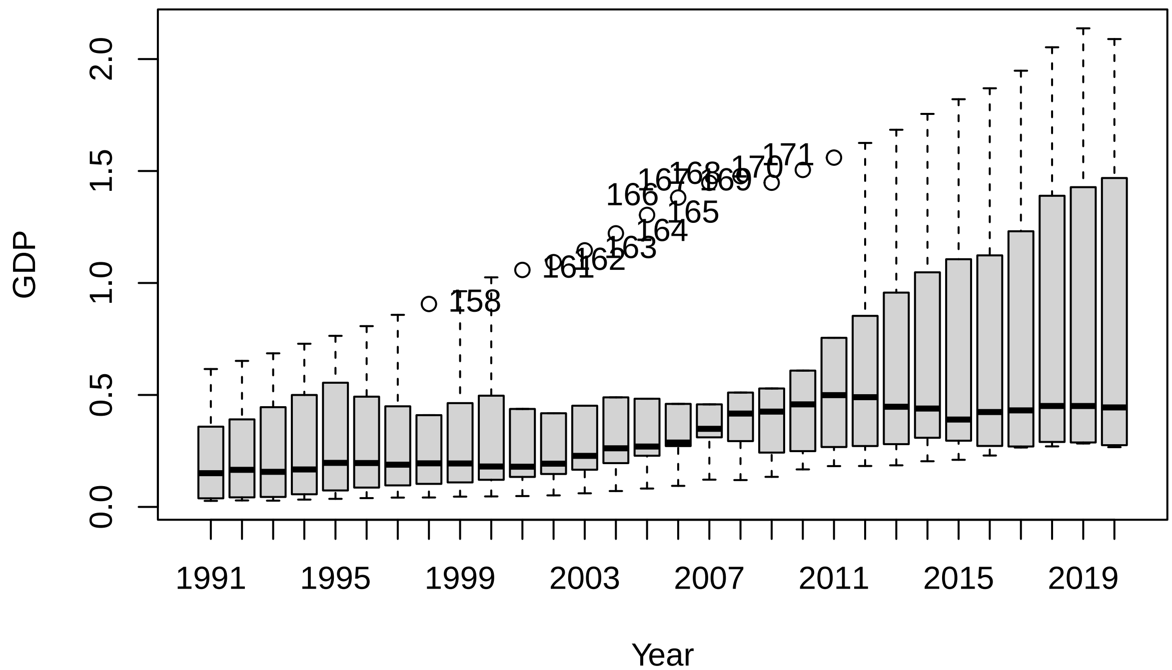
```
##
## t test of coefficients:
##
##                  Estimate  Std. Error t value Pr(>|t|)
## (Intercept)    7.5802e+12  2.5354e+12  2.9898 0.003189 **
## data1.Exports -2.8294e+11  4.0990e+11 -0.6903 0.490937
## data1.Imports  1.5746e+11  3.7651e+11  0.4182 0.676311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pooled Scatter Plot and Plot Means

Overtime, as an economy grows larger, it also becomes harder for it to track down each component in GDP. In the scatterplot, it represents this because the confidence interval get larger and larger overtime. From 1991 to early 2000s, there is not much overlap, meaning there is no significant difference. If it had not been for the financial crisis, the data would have followed the same pattern. By 2015, the median returns. The medians is increasing with time. Yet, due to the confidence interval increasing with time, the prediction of the median is less accurate.

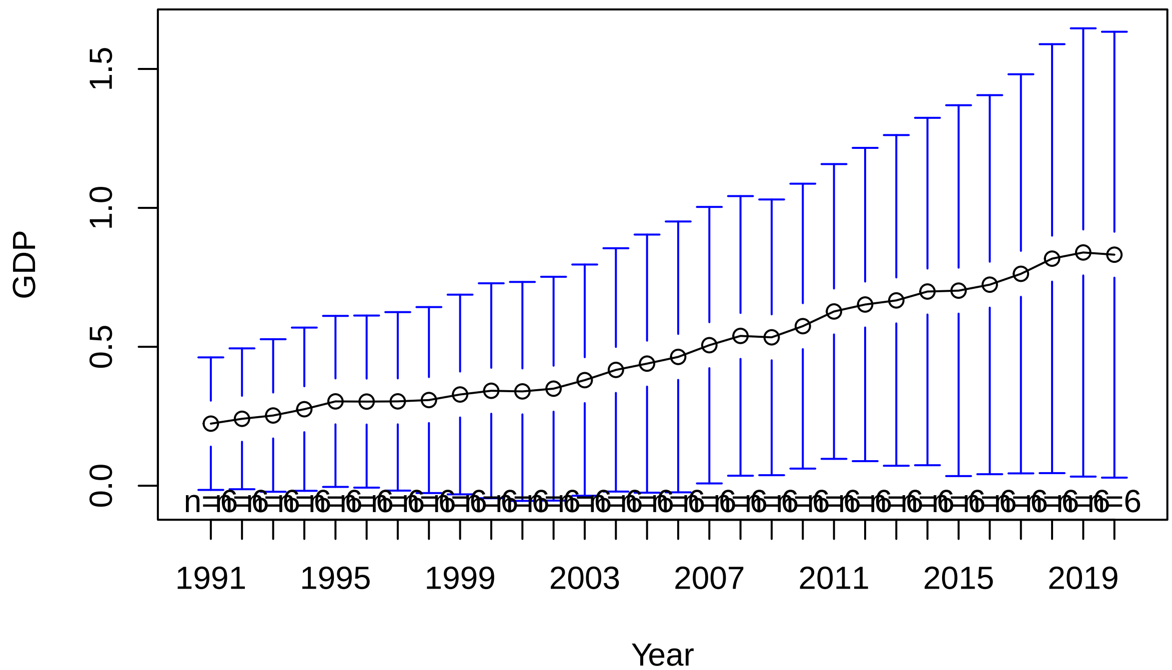
Because we do not have a lot of outliers in the scatterplot, we are able to use the plot means to better visualize if there is a significant difference across GDPs. First, plotting GDP versus the individual, we can conclude there is a large gap between China compared to the other countries. This is a significiant difference. Next, plotting GDP versus the year, there is also a significant difference. As stated above, the error bands grow as GDP grows overtime.

```
scatterplot(data1.GDP/10^13~data1.Year|data1.ID, data=paneldata, xlab="Year", ylab="GDP")
```

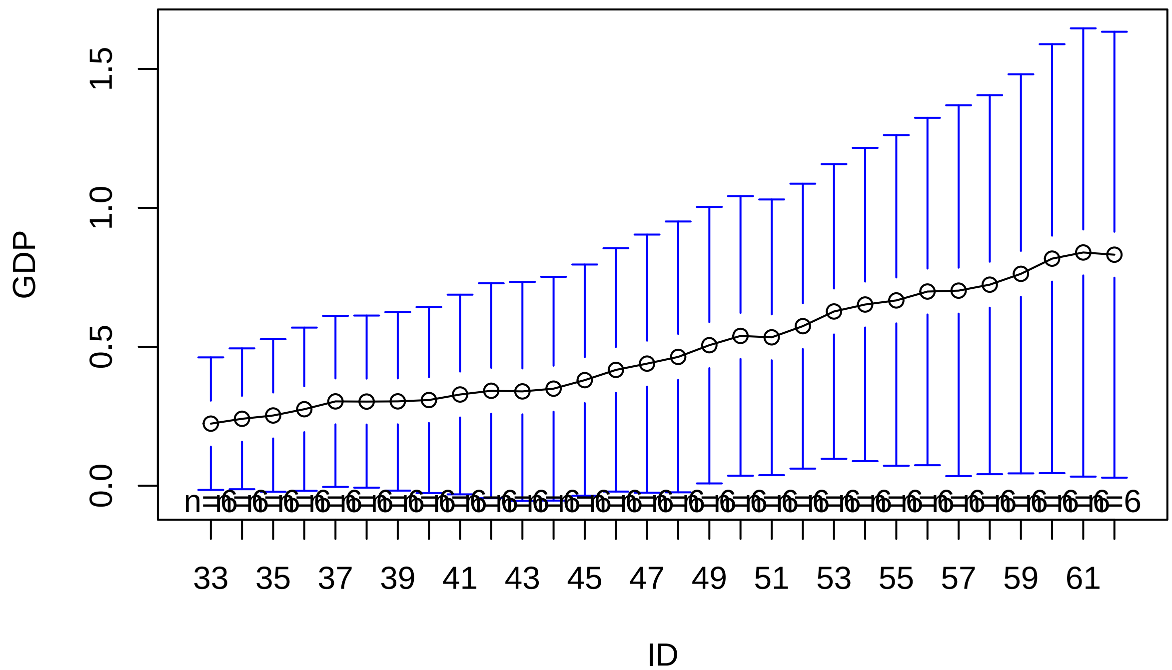



```
##  [1] "158" "161" "162" "163" "164" "165" "166" "167" "168" "169" "170" "171"
```

```
plotmeans(paneldata$data1.GDP/10^13~paneldata$data1.Year, data=paneldata, xlab="Year", ylab="GDP")
```




```
plotmeans(paneldata$data1.GDP/10^13~paneldata$data1.ID, data=paneldata, xlab="ID", ylab="GDP")
```



Comparing Fixed Effects, Pool Effects, and Random Effects

We run the fixed effects model and compare it to the pooled model. To identify the preferred model we ran the pooled model, the fixed effects model, and the random effect model. To compare the pool effects to the fixed effects model, we used the Pftest function. Testing both firm and time effects versus the pooled model, we obtained a low p value meaning that both firm and time effects were better than the pool model. Testing only the time effects against the pooled model, we obtained a p value equal to one. This meant that the time effect was not significantly different. Lastly, testing the firm effects only against the pooled model resulted in a low p value, so the firm fixed effects is the preferred model.

From these three tests, we can conclude that the fixed effects model including the firm is preferred over the pooled model. To verify these findings, we plotted all of these effects onto a coefficient plot. This plot showed us that none of our betas crossed zero. Next, we compared this to the random effects model. This resulted in a large p-value, so we failed to reject the null hypothesis and concluded that the random effects model was preferred. Given the random effects model using GLS, this is the best model to use out of the three. Plotting the coefficients of the random effects model, the beta values did not cross the insignificance line validating that it is the best model.

```
#fixed effects
fixedEffect.full<- plm(data1.GDP~data1.Exports+data1.Imports, model="within", data=paneldata, effect="tw
fixedEffect.time<- plm(data1.GDP~data1.Exports+data1.Imports, model="within", data=paneldata, effect="ti
fixedEffect.firm<- plm(data1.GDP~data1.Exports+data1.Imports, model="within", data=paneldata, effect="in
pFtest(fixedEffect.full, poolEffect)
```

```
##
```

```
##  F test for twoways effects
##
## data:  data1.GDP ~ data1.Exports + data1.Imports
## F = 30.565, df1 = 34, df2 = 143, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

```
#fixed effect full preferred

pFtest(fixedEffect.time, poolEffect)
```

```
##
##  F test for time effects
##
## data:  data1.GDP ~ data1.Exports + data1.Imports
## F = 2.7421, df1 = 29, df2 = 148, p-value = 3.961e-05
## alternative hypothesis: significant effects
```

```
#Including time effect does not help

pFtest(fixedEffect.firm, poolEffect)
```

```
##
##  F test for individual effects
##
## data:  data1.GDP ~ data1.Exports + data1.Imports
## F = 65.877, df1 = 5, df2 = 172, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

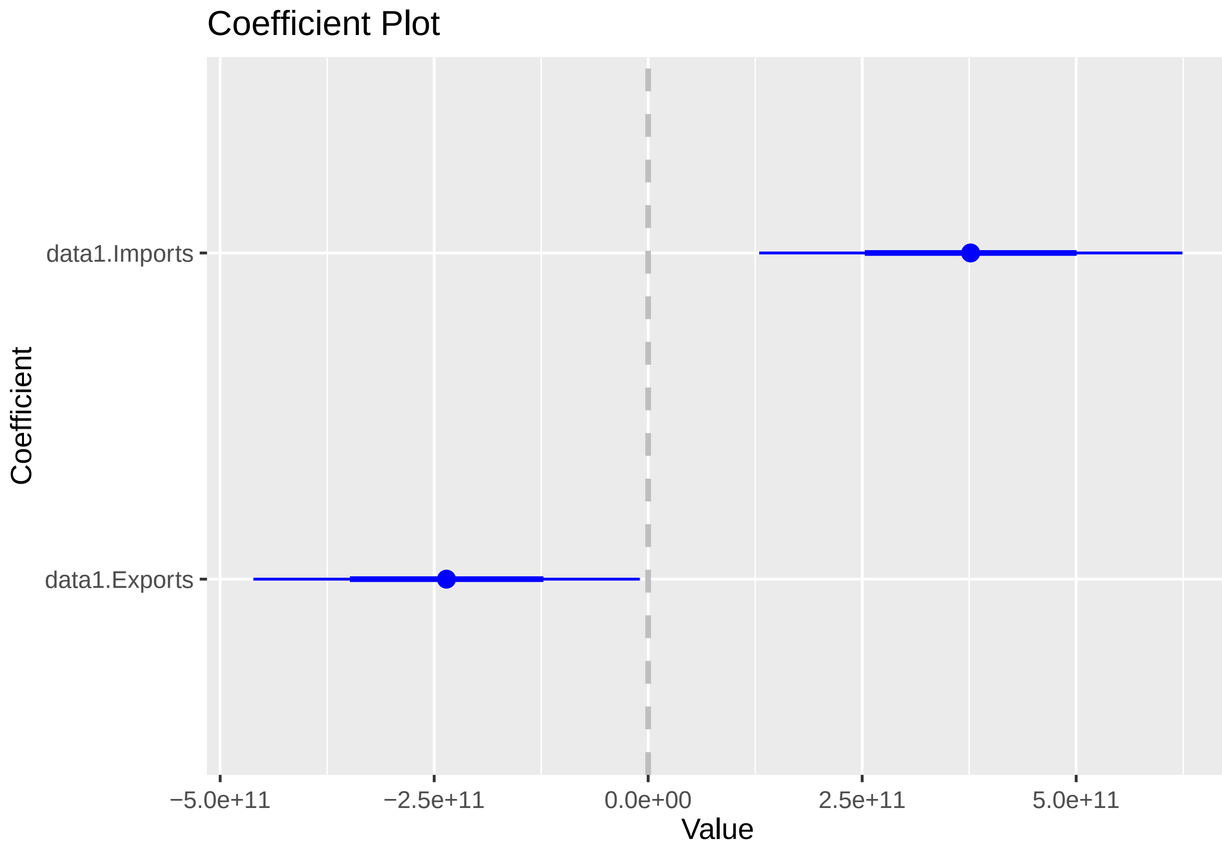
```
#Firm affects the significance, we want to include firm effects
```

Coefficient Plot for Fixed Effects firm model

```
library(coefplot)
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
coefplot(fixedEffect.firm)
```



Random Effects Model

```
#random effect model
randomeffect<-plm(data1.GDP~data1.Exports+data1.Imports, data=paneldata, model="random")

#random effect compared to fixed effect firm
phtest(fixedEffect.firm, randomeffect)
```

```
##
##  Hausman Test
##
## data:  data1.GDP ~ data1.Exports + data1.Imports
## chisq = 2.3699, df = 2, p-value = 0.3058
## alternative hypothesis: one model is inconsistent
```

```
#Fail to reject, Use Random effects
```

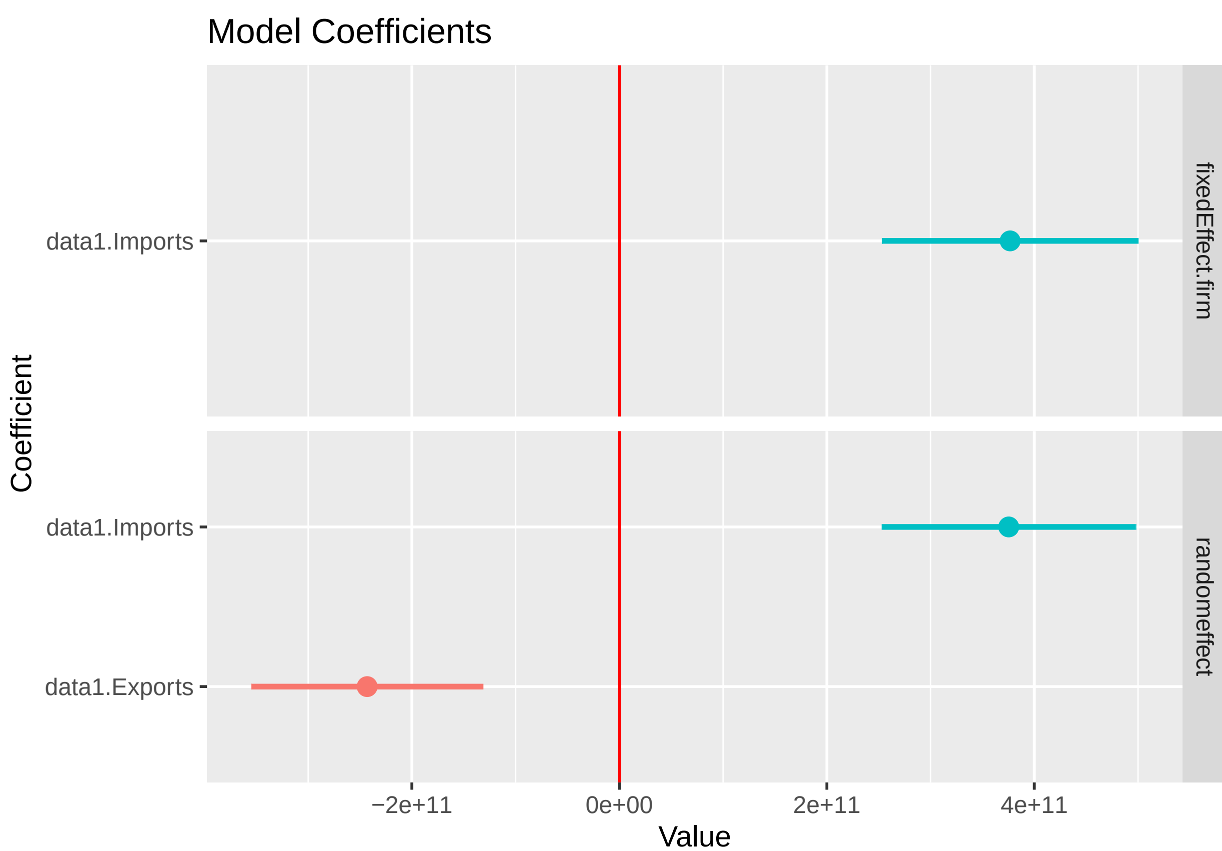
Random Effects Plot

```
#random effects plot
ce <- function(model.obj) {
```

```
  summ.model <- summary(get(model.obj))$coefficients
  extract <- summ.model[2:nrow(summ.model),drop=FALSE, 1:2]
  return(data.frame(extract, vars = row.names(extract), model = model.obj))
}
coefs <- do.call(rbind, sapply(paste0(list(
  "fixedEffect.firm", "randomeffect"
)), ce, simplify= FALSE))
names(coefs)[2] <- "se"
gg_coef <- ggplot(coefs, aes(vars, Estimate)) +
  geom_hline(yintercept = 0, lty = 1, lwd = 0.5, colour = "red") +
  geom_errorbar(aes(ymin = Estimate - se, ymax = Estimate + se, colour = vars),
                lwd = 1, width = 0
  )+
  geom_point(size = 3, aes(colour = vars)) + facet_grid(model ~ ., scales="free") + coord_flip() +
  guides(colour = FALSE) +
  labs(x = "Coefficient", y = "Value") + ggtitle("Model Coefficients")
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.
```

```
gg_coef
```




```
#This verifies that the random effects is the best model because each beta value does not cross the ins
```