

MATH2349 - Preprocessing UN Goods Trading Datasets

Code ▼

Assignment 3

Jigar Mangukiya (S3715807), Vikas Virani (S3715555), Vishwa Gandhi (S3714805)

Required packages

Hide

```
library(dplyr)
library(readr)
library(tidyr)
library(ggplot2)
```

Executive Summary

Two datasets “Imports of Goods” & “Exports of Goods” were taken with country wise Gross values for Imported and Exported goods for different years and different SNA series methods.

We have taken a subset of country column to filter out 7 countries -“United Kingdom”, “United States”, “Japan”, “Germany”, “Australia”, “Canada” and “France” and year column to filter records from 1970 to 2010 & used Join operation to merge these two datasets to combine them in a single dataset. Also, Column names were changed to more meaningful names after joining them.

For better Understanding of data, “Country”, “SNA_Series”, “SNA_System” & “Currency” are converted into unordered Factor and “Year” is converted into ordered Factor.

As merged dataset satisfies the tidy data condition, there was no need to perform any operation to tidy the dataset. We have merged two columns “SNA_Series” & “SNA_Numbers” and mutated one column named “Terms_Of_Trade”.

We have scanned the data to identify potential NAs/missing values and Outliers in the dataset. There were no missing values or NAs found in the combined dataset. Possible outliers are identified but given the context of the Numeric Column on which outliers are identified, there is no significance of considering them as outliers. Detailed Explanation is given in the the relevant “Scan” section of this document.

Since the dataset contains Gross values for different countries in their own currency values and different countries having different currency values, we have transformed the variables “Gross_Imports” & “Gross_Exports” by scaling the values, to express them in terms of Standard Deviations, giving resulting values on a same scale so that they can be compared with other values.

Data

Data for this Data Preprocessing study is acquired from “www.data.un.org”, which is mainly a product of United Nations Statistics Division which brings UN Stats Databases in reach of users through internet. Initially we are referring two datasets, both falling under same category and sub-category, which is “National Accounts Official Country Data”/“table 1.1 Gross domestic product by expenditures at current prices”. Datasets are “Imports of Goods” & “Exports of Goods”. Data descriptions for both of them are follows.

1. Imports of Goods

Source - http://data.un.org/Data.aspx?d=SNA&f=group_code%3a101%3bitem_code%3a21
 (http://data.un.org/Data.aspx?d=SNA&f=group_code%3a101%3bitem_code%3a21)

2. Exports of Goods

Source - http://data.un.org/Data.aspx?d=SNA&f=group_code%3a101%3bitem_code%3a18
 (http://data.un.org/Data.aspx?d=SNA&f=group_code%3a101%3bitem_code%3a18)

Variables - a. Country or Area - Name of territory being observed b. SNA93 Table Code - category table reference number c. Sub Group - Sub Category reference number d. Item - Item that current database represents in the provided category/sub-category e. SNA93 Item Code - "System of National Accounts - 1993" - item number for particular dataset f. Year - Year of observation g. Series - SNA 93 Observation Series h. Currency - Currency for the mentioned value for Import / Export i. Fiscal Year Type - Fiscal Year specification for the observation j. value - Reported gross value of Import/Export for country or Area for given year

However, For this task, we will only consider following columns.

- i. Country or Area
- ii. Year
- iii. Series
- iv. Currency
- v. SNA System
- vi. Value

Both of the datasets, contain identical variable sets. First dataset contains the gross value for Imported Goods and second contains gross value for Exported Goods, for multiple years measured by multiple SNA Series methods. We will consider a subset of countries to carry out the preprocessing task, which includes "Japan", "United Kingdom", "Germany", "United States", "France", "Canada" & "Australia". We will consider data from years 1970 to 2010. So for each country we will have Import & Export Records for 41 years.

We will merge the datasets with an inner join on columns "Country or Area", "Year", "Series", "SNA System", "Currency". This will give us a combined dataset, records of which, describes the Gross Import & Gross Export Values for a country, for a year measured in the same SNA Series.

We will rename the columns to more readable names.

It is important to note that for a given year, Import & Export values can be measured in multiple SNA Systems. For this task we will use only the highest SNA System(Series) in which data is recorded. We will do that by taking a slice which has maximum SNA System year out of all 3 SNA Systems that are used to measure Goods Trading, i.e. "1968", "1993", "2008".

Hide

```
cat("Subsetting Import of Goods Dataset before merging (Below)")
```

```
Subsetting Import of Goods Dataset before merging (Below)
```

Hide

```
ImportGoods <- ImportGoods[0:9506,]
ImportGoods <- ImportGoods %>% select(`Country or Area`,Year,Series,Currency,`SNA System`,Value)
ImportGoods <- ImportGoods %>% subset(ImportGoods$`Country or Area` %in% c("United Kingdom",
"United States","Japan","Germany","Australia","Canada","France") & ImportGoods$Year %in% c(seq(1970,2010)))
cat("Subsetting Export of Goods Dataset before merging (Below)")
```

Subsetting Export of Goods Dataset before merging (Below)

Hide

```
ExportGoods <- ExportGoods[0:9506,]
ExportGoods <- ExportGoods %>% select(`Country or Area`,Year,Series,Currency,`SNA System`,Value)
ExportGoods <- ExportGoods %>% subset(ExportGoods$`Country or Area` %in% c("United Kingdom",
"United States","Japan","Germany","Australia","Canada","France") & ExportGoods$Year %in% c(seq(1970,2010)))
cat("Merging datasets (Below)")
```

Merging datasets (Below)

Hide

```
TradeStatistics <- ImportGoods %>% inner_join(ExportGoods,by = c("Country or Area","Year","Series",
"SNA System","Currency"))
cat("Renaming Columns (Below)")
```

Renaming Columns (Below)

Hide

```
colnames(TradeStatistics) <- c("Country","Year","SNA_Series","Currency","SNA_System","Gross_Imports",
"Gross_Exports")
head(ImportGoods)
```

Country or Area <chr>	Year <int>	Series <int>	Currency <chr>	SNA System <int>	Value <dbl>
Australia	2010	1000	Australian dollar	2008	2.25028e+11
Australia	2010	300	Australian dollar	1993	2.19243e+11
Australia	2009	1000	Australian dollar	2008	2.10755e+11
Australia	2009	300	Australian dollar	1993	2.04995e+11
Australia	2008	1000	Australian dollar	2008	2.24855e+11
Australia	2008	300	Australian dollar	1993	2.20890e+11

6 rows

Hide

head(ExportGoods)

Country or Area <chr>	Year <int>	Series <int>	Currency <chr>	SNA System <int>	Value <dbl>
Australia	2010	1000	Australian dollar	2008	2.47021e+11
Australia	2010	300	Australian dollar	1993	2.46953e+11

Country or Area <chr>	Year <int>	Series <int>	Currency <chr>	SNA System <int>	Value <dbl>
Australia	2009	1000	Australian dollar	2008	2.01809e+11
Australia	2009	300	Australian dollar	1993	2.01751e+11
Australia	2008	1000	Australian dollar	2008	2.31622e+11
Australia	2008	300	Australian dollar	1993	2.31623e+11

6 rows

Hide

```
head(TradeStatistics)
```

Country <chr>	Y... <int>	SNA_Series <int>	Currency <chr>	SNA_Sys... <int>	Gross_Imports <dbl>	Gross_Exports <dbl>
Australia	2010	1000	Australian dollar	2008	2.25028e+11	2.47021e+11
Australia	2010	300	Australian dollar	1993	2.19243e+11	2.46953e+11
Australia	2009	1000	Australian dollar	2008	2.10755e+11	2.01809e+11
Australia	2009	300	Australian dollar	1993	2.04995e+11	2.01751e+11
Australia	2008	1000	Australian dollar	2008	2.24855e+11	2.31622e+11
Australia	2008	300	Australian dollar	1993	2.20890e+11	2.31623e+11

6 rows

Hide

```
TradeStatistics <- TradeStatistics %>% group_by(Country,Year) %>% slice(which.max(SNA_System))
TradeStatistics$Country %>% table()
```

.	Australia	Canada	France	Germany	Japan	United Kingdom	Un
ited States	41	41	41	41	41	41	
41							

Understand

Following is the Summary of variable types. 1. Country - Character 2. Year - int 3. SNA_Series - int 4. Currency - Character 5. SNA_System - int 6. Gross_Imports - double 7. Gross_Exports - double

- “Country”, “SNA_Series” & “SNA_System” are converted into unordered factors.
- Currency is converted into unordered labeled factor variable. We will change the labels to International currency codes for ease of usage.
- Year is converted into a ordered factor variable
- Gross_Imports & Gross_Exports are numeric variables

Hide

```
TradeStatistics$Country <- TradeStatistics$Country %>% as.factor()
attributes(TradeStatistics$Country)
```

```
$levels
[1] "Australia"      "Canada"          "France"           "Germany"          "Japan"            "United Kingdom"
[7] "United States"

$class
[1] "factor"
```

Hide

```
TradeStatistics$SNA_Series <- TradeStatistics$SNA_Series %>% as.factor()
attributes(TradeStatistics$SNA_Series)
```

```
$levels
[1] "20"  "200" "300" "400" "1000"

$class
[1] "factor"
```

Hide

```
TradeStatistics$SNA_System <- TradeStatistics$SNA_System %>% as.factor()
attributes(TradeStatistics$SNA_System)
```

```
$levels
[1] "1968" "1993" "2008"

$class
[1] "factor"
```

Hide

```
TradeStatistics$Year <- TradeStatistics$Year %>% factor(levels = c(seq(1970,2010)),labels = c
(seq(1970,2010)), ordered = TRUE)
attributes(TradeStatistics$Year)
```

```
$levels
[1] "1970" "1971" "1972" "1973" "1974" "1975" "1976" "1977" "1978" "1979" "1980" "1981" "198
2" "1983" "1984" "1985"
[17] "1986" "1987" "1988" "1989" "1990" "1991" "1992" "1993" "1994" "1995" "1996" "1997" "199
8" "1999" "2000" "2001"
[33] "2002" "2003" "2004" "2005" "2006" "2007" "2008" "2009" "2010"

$class
[1] "ordered" "factor"
```

Hide

```
str(TradeStatistics$Year)
```

```
Ord.factor w/ 41 levels "1970"<"1971"<...: 1 2 3 4 5 6 7 8 9 10 ...
```

Hide

```
TradeStatistics$Currency <- TradeStatistics$Currency %>% factor(levels = c("1999 FRF euro / e
uro","1999 DEM euro / euro","Euro","Australian dollar","Canadian dollar","yen","pound sterlin
g","US dollar"),labels = c("EUR","EUR","EUR","AUD","CAD","JPY","GBP","USD"),ordered = FALSE)
```

```
duplicated levels in factors are deprecated
```

Hide

```
attributes(TradeStatistics$Currency)
```

```
$levels
[1] "EUR" "EUR" "EUR" "AUD" "CAD" "JPY" "GBP" "USD"

$class
[1] "factor"
```

Hide

```
TradeStatistics$Gross_Imports %>% class()
```

```
[1] "numeric"
```

Hide

```
TradeStatistics$Gross_Exports %>% class()
```

```
[1] "numeric"
```

Tidy & Manipulate Data I

Following characteristics are observable in the current data.

- Each variable has its own column, as there are variables expressed as values.
- All the values in the records that are unique and have their own row due to inner join operation we performed. only matching rows are copied to the new dataset.
- All the values have their own cell.

That suggests that data is readily in a tidy format.

However, there is a possibility to merge two columns - SNA_Series & SNA_Numbers, as they have contextual binding with each other, and also, we don't need these values separately in further preprocessing tasks. So these two columns are merged together with unite() operation. The resulting column will contain both values joined with a "/". Altered dataset can be seen in the table-output below.

Hide

```
TradeStatistics <- TradeStatistics %>% unite(col = "SNA-Series/System",SNA_Series , SNA_System,sep = "/")
TradeStatistics$`SNA-Series/System` <- as.factor(TradeStatistics$`SNA-Series/System`)
TradeStatistics %>% head()
```

Country <fctr>	Year <ord>	SNA-Series/System <fctr>	Currency <fctr>	Gross_Imports <dbl>	Gross_Exports <dbl>
Australia	1970	1000/2008	AUD	3.806e+09	4.261e+09
Australia	1971	1000/2008	AUD	3.814e+09	4.766e+09
Australia	1972	1000/2008	AUD	3.831e+09	6.110e+09
Australia	1973	1000/2008	AUD	5.795e+09	6.861e+09
Australia	1974	1000/2008	AUD	7.728e+09	8.656e+09
Australia	1975	1000/2008	AUD	7.999e+09	9.628e+09

6 rows

Tidy & Manipulate Data II

We will create a new variable called “Terms_Of_Trade”. Terms of trade is a trading term, which is essentially a ratio of Gross Exports to Gross Imports multiplied by 100. Terms of trade, in fact, describes the amount of export good a country can purchase per unit of import good. So if the Terms of Trade is more than 100%, that indicates a country has an healthy economy with good trade. But that is a blunt and abstract statement to indicate the meaning of ToT.

We will add the new column named “Terms_Of_Trade” in our dataset by using a mutate() operation.

Hide

```
TradeStatistics <- TradeStatistics %>% mutate(Terms_Of_Trade = (Gross_Exports * 100)/ Gross_Imports)
TradeStatistics %>% head()
```

Country <fctr>	Y... <ord>	SNA-Series/System <fctr>	Currency <fctr>	Gross_Imports <dbl>	Gross_Exports <dbl>	Terms_Of_Trade <dbl>
Australia	1970	1000/2008	AUD	3.806e+09	4.261e+09	111.95
Australia	1971	1000/2008	AUD	3.814e+09	4.766e+09	124.96
Australia	1972	1000/2008	AUD	3.831e+09	6.110e+09	159.48
Australia	1973	1000/2008	AUD	5.795e+09	6.861e+09	118.39
Australia	1974	1000/2008	AUD	7.728e+09	8.656e+09	112.00
Australia	1975	1000/2008	AUD	7.999e+09	9.628e+09	120.36

6 rows

Scan I

In this section, We have checked Gross_Imports, Gross_Exports & newly added Terms_Of_Trade variables for null values, Infinite values and NaN values.

As per the results, there are no null/inconsistent values in the numeric variables of dataset.

[Hide](#)

```
na_Count_Gross_Imports <- sum(is.na(TradeStatistics$Gross_Imports) | is.infinite(TradeStatistics$Gross_Imports) | is.nan(TradeStatistics$Gross_Imports))
na_Count_Gross_Exports <- sum(is.na(TradeStatistics$Gross_Exports) | is.infinite(TradeStatistics$Gross_Exports) | is.nan(TradeStatistics$Gross_Exports))
na_Count_TOT <- sum(is.na(TradeStatistics$Terms_Of_Trade) | is.infinite(TradeStatistics$Terms_Of_Trade) | is.nan(TradeStatistics$Terms_Of_Trade))
cat(" Number of Nulls/inconsistencies in Gross_Imports -",na_Count_Gross_Imports,"\n","Number of Nulls/inconsistencies in Gross_Imports -",na_Count_Gross_Exports,"\n","Number of Nulls/inconsistencies in Terms_Of_Trade -",na_Count_TOT,"\n")
```

```
Number of Nulls/inconsistencies in Gross_Imports - 0
Number of Nulls/inconsistencies in Gross_Imports - 0
Number of Nulls/inconsistencies in Terms_Of_Trade - 0
```

Scan II

We will scan all 3 numeric columns for potential outliers in them using box-plots.

It is important to note that we will scan for outliers, considering each country, for variables Gross_Imports & Gross_Exports. We do so to accommodate the effect of different currencies on the value of Imports & Exports, we will need look for outliers in the data of particular groups.

For example, we cannot directly compare, United States' import value in 1970, which is "40900000000" USD, to Japan's import value in 1970, which is "5484100000000" JPY. Simply because a USD and JPY cannot be directly compared.

That's why we can look for outliers in particular country's data, where all observations are made in same currency. Outliers in table 1 & 2 represent, Significantly high or low, Import or Export of Goods by a country in a particular year than it did in other years.

However, for Terms of Trade, we can consider entire variable to scan for outliers, simply because of the fact that ToT is not dependent on any currency, and can be used to compare absolute trade performances of countries.

For example, if United States has a ToT of 110 for any given year & Japan has ToT of 140 for the same or different year, we can say that Japan performed better than united states particularly in "Terms of Trade".

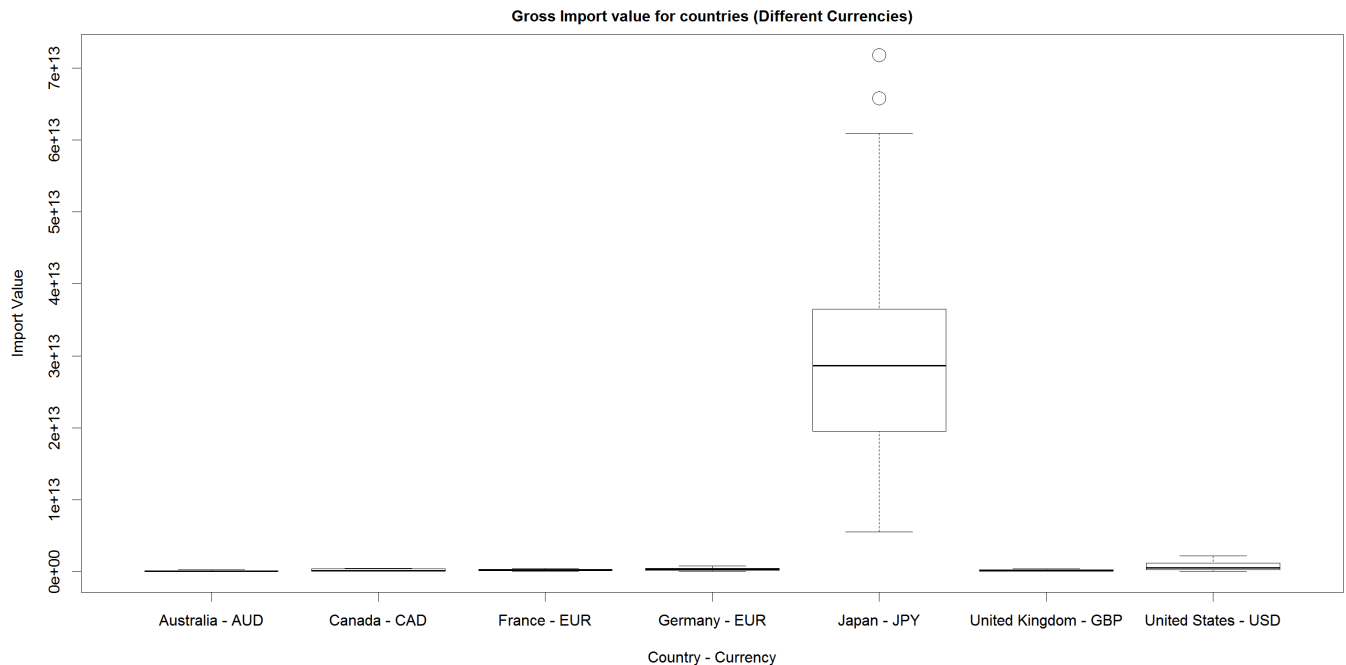
Outliers in table 3 represent, Significantly high or low performance of a country in "Terms of Trade" as compared to all other countries for entire observation period.

These outliers, are not due to any error in measurement, but are extreme occurrences of events in the observation period so we do not need to impute them, as they possess actual statistical values.

NOTE > Box-plots for Gross_Imports & Gross_Exports may seem unrealistic or faulty at first! but the irregular shape for countries other than "Japan", is because of the fact that all the values are represented in different currencies. Value of Japanese being smaller than rest of the countries, the magnitude of numbers representing Import & Export is large. > We need to scale the variables Gross_Imports & Gross_Exports, to suppress the effect of different currencies. Which is done in the next section

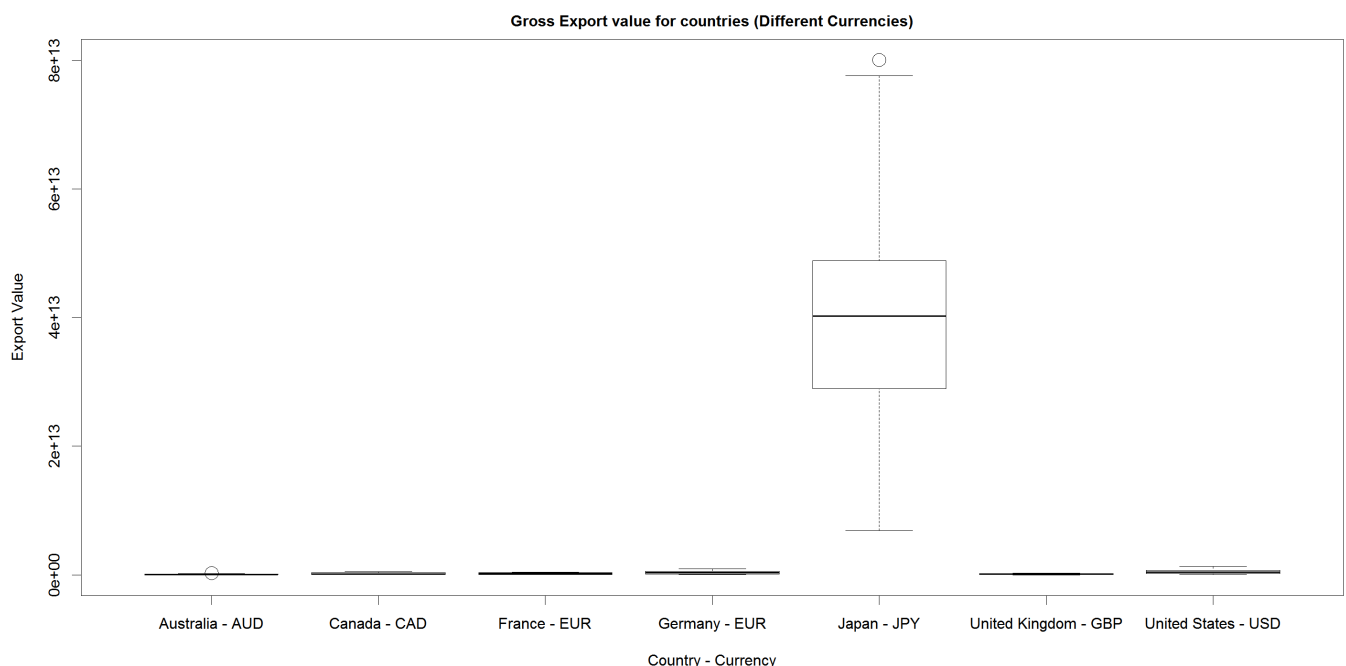
[Hide](#)


```
par(ps = 12, cex = 2, cex.main = 1)
imports_boxplot <- boxplot(Gross_Imports ~ as.factor(x = unite(col = "Country-Currency", Country, Currency, sep = " - ", data = TradeStatistics)$`Country-Currency`),
                           data = TradeStatistics, main="Gross Import value for countries (Different Currencies)",
                           xlab="Country - Currency",
                           ylab="Import Value")
```

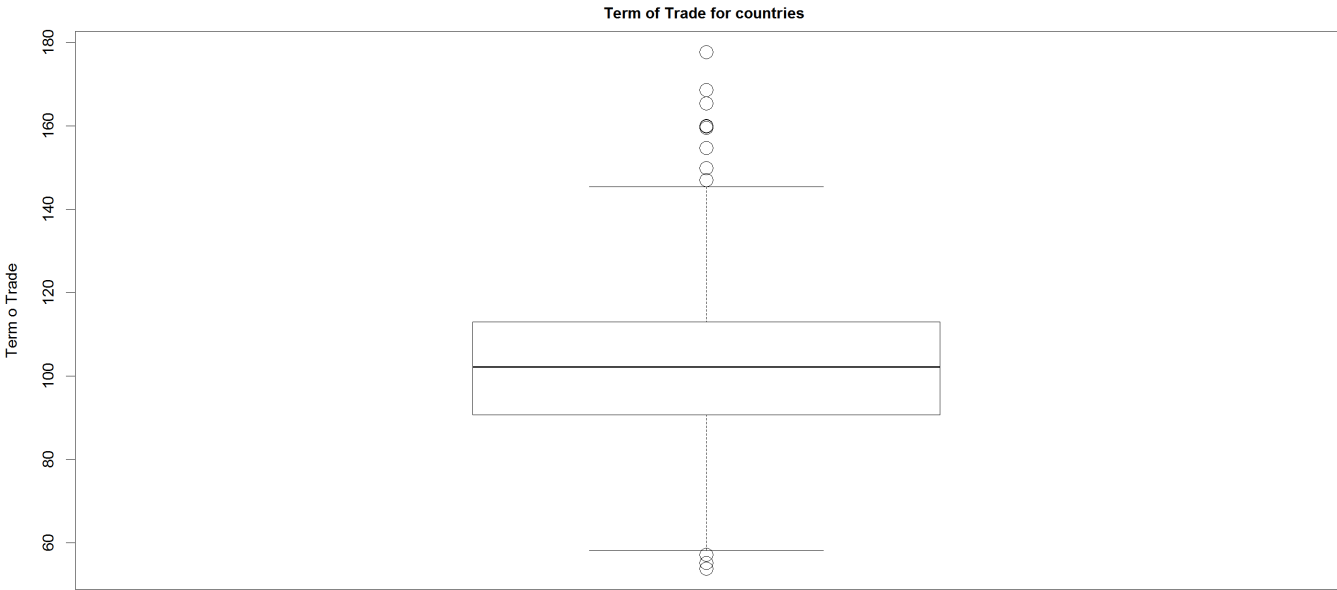


Hide

```
exports_boxplot <- boxplot(Gross_Exports ~ as.factor(x = unite(col = "Country-Currency", Country, Currency, sep = " - ", data = TradeStatistics)$`Country-Currency`),
                           data = TradeStatistics,
                           main="Gross Export value for countries (Different Currencies)",
                           xlab="Country - Currency",
                           ylab="Export Value")
```



```
tot_boxplot <- boxplot(TradeStatistics$Terms_Of_Trade,
  main="Term of Trade for countries ",
  ylab="Term o Trade")
```



```
Import_Outliers <- TradeStatistics %>% subset(Gross_Imports %in% imports_boxplot$out)
Export_Outliers <- TradeStatistics %>% subset(Gross_Exports %in% exports_boxplot$out)
Tot_Outliers <- TradeStatistics %>% subset((Terms_Of_Trade) %in% tot_boxplot$out)
print(Import_Outliers)
```

Coun... <fctr>	Y... <ord>	SNA-Series/System <fctr>	Currency <fctr>	Gross_Imports <dbl>	Gross_Exports <dbl>	Terms_Of_Trade <dbl>
Japan	2007	1000/2008	JPY	6.58363e+13	8.00236e+13	121.5494
Japan	2008	1000/2008	JPY	7.18080e+13	7.76111e+13	108.0814

2 rows

```
print(Export_Outliers)
```

Country <fctr>	Y... <ord>	SNA-Series/System <fctr>	Currency <fctr>	Gross_Imports <dbl>	Gross_Exports <dbl>	Terms_Of_Trade <d
Australia	2010	1000/2008	AUD	2.25028e+11	2.47021e+11	109.77
Japan	2007	1000/2008	JPY	6.58363e+13	8.00236e+13	121.54

2 rows

Hide

```
print(Tot_Outliers)
```

Country <fctr>	Y... SNA-Series/System <ord><fctr>	Currency <fctr>	Gross_Imports <dbl>	Gross_Exports <dbl>	Terms_C
Australia	1972 1000/2008	AUD	3.83100e+09	6.11000e+09	15
Japan	1971 20/1968	JPY	5.62590e+12	8.27300e+12	14
Japan	1986 300/1993	JPY	1.94747e+13	3.45997e+13	17
Japan	1987 300/1993	JPY	1.92915e+13	3.25233e+13	16
Japan	1988 300/1993	JPY	2.16113e+13	3.34258e+13	15
Japan	1992 300/1993	JPY	2.63055e+13	4.20816e+13	15
Japan	1993 300/1993	JPY	2.36823e+13	3.91640e+13	16
Japan	1994 300/1993	JPY	2.46166e+13	3.93485e+13	15
Japan	1998 1000/2008	JPY	3.22116e+13	4.82898e+13	14
United States	2003 1000/2008	USD	1.29619e+12	7.41420e+11	5

1-10 of 13 rows

Previous 1 2 Next

Transform

The difference in values of different currencies creates a major problem while trying to understand nature of variables Gross_Imports & Gross_Exports. As can be seen in the plots on the left hand side in the grid below. Japan's currency JPY has a relatively low value compared to others. So the number representing, even a small trade, in JPY will be a significantly large number. (As on 21 October, 2018, 1 JPY = 0.0089 USD).

This creates a problem when we try to interpret the data which has values expressed in multiple currencies. Smaller currencies tend to overshadow the larger currencies.

To remedy the effects of different currencies, we will scale the values, and express them in terms of Standard Deviations.

Here also, we will group the data by country, and apply scaling to each of the group independently. If we apply the scaling on entire variables, the output will be the same, smaller JPY overshadowing all the other larger currencies as can be seen in left side plots.

Grouped scaling will expose true nature of the values, by abstracting the effect of currency on the data.

As can be observed from the plots on the right hand side, data is more interpretable, as the values are scaled and now expressed in terms of Standard deviations of that group.

By comparing both graphs we can see that Japan's values in scatterplot & shapes in boxplots have not changed. It is because scaling has not changed the values in its absolute, but only represented them in a more suitable and narrow scale, without affecting the true nature of the data.

Hide

```
cat("Plot A1")
```

Plot A1

Hide

```
scatter_plot_before_scale <-
  ggplot(TradeStatistics, aes(x=Gross_Imports, y=Gross_Exports, col
or=Country, shape = Country))+
  scale_shape_manual(values=1:nlevels(TradeStatistics$Country)) +
  geom_point(size=8) +
  theme_grey(base_size = 25) +
  labs(x = "Gross Imports",y = "Gross Exports",title = "Scatter pl
ot - Gross Imports vs. Gross Exports (Before Scaling)",tag = "A1")
cat("Plot A2")
```

Plot A2

Hide

```
boxplot_Import_before_scale <- ggplot(data = TradeStatistics,aes(x = Country, y = Gross_Import
ts, color = Country)) +
  geom_boxplot(size = 1,outlier.color = "black",outlier.size =
3) +
  theme(axis.text.x = element_text(angle = 45, hjust = 2,size =
25)) +
  theme_grey(base_size = 25) +
  labs(x = "Country",y = "Gross Imports",title = "Box plot - Gr
oss Imports By Country (Before Scaling)",tag = "A2")
cat("Plot A3")
```

Plot A3

Hide

```
boxplot_Export_before_scale <- ggplot(data = TradeStatistics,aes(x = Country, y = Gross_Export
ts, color = Country)) +
  geom_boxplot(size = 1,outlier.color = "black",outlier.size =
3) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1,size =
25)) +
  theme_grey(base_size = 25) +
  labs(x = "Country",y = "Gross Exports",title = "Box plot - Gr
oss Exports By Country (Before Scaling)",tag = "A3")
cat("Scaling Gross_Imports")
```

Scaling Gross_Imports

Hide

```
TradeStatistics <- TradeStatistics %>% group_by(Country) %>% mutate(Gross_Imports_Scaled =
(Gross_Imports /(sd(Gross_Imports))))
cat("Scaling Gross_Exports")
```

Scaling Gross_Exports

Hide

```
TradeStatistics <- TradeStatistics %>% group_by(Country) %>% mutate(Gross_Exports_Scaled =  
(Gross_Exports / (sd(Gross_Exports))))  
cat("Plot B1")
```

Plot B1

Hide

```
scatter_plot_after_scale <- ggplot(TradeStatistics, aes(x=Gross_Imports_Scaled, y=Gross_Export  
ts_Scaled, color=Country, shape = Country)) +  
  scale_shape_manual(values=1:nlevels(TradeStatistics$Country)) +  
  geom_point(size = 8) +  
  theme_grey(base_size = 25) +  
  labs(x = "Gross Imports", y = "Gross Exports", title = "Scatter pl  
ot - Gross Imports vs. Gross Exports (After Scaling)", tag = "B1")  
cat("Plot B2")
```

Plot B2

Hide

```
boxplot_Import_after_scale <- ggplot(data = TradeStatistics, aes(x = Country, y = Gross_Import  
s_Scaled, color = Country)) +  
  geom_boxplot(size = 1, outlier.color = "black", outlier.size = 3  
) +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size =  
25)) +  
  theme_grey(base_size = 25) +  
  labs(x = "Country", y = "Gross Imports", title = "Box plot - Gro  
ss Imports By Country (After Scaling)", tag = "B2")  
cat("Plot B3")
```

Plot B3

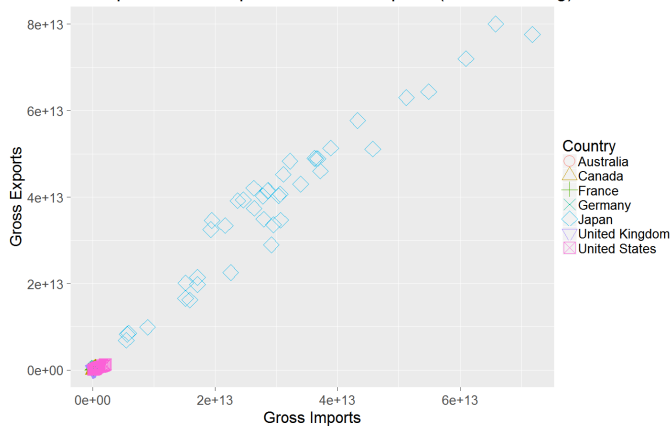
Hide

```

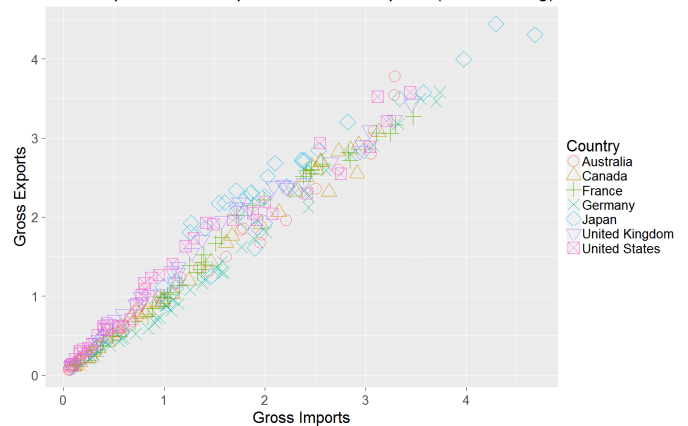
boxplot_Export_after_scale <- ggplot(data = TradeStatistics,aes(x = Country, y = Gross_Export
s_Scaled, color = Country)) +
  geom_boxplot(size = 1,outlier.color = "black",outlier.size = 3
) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1,size =
25)) +
  theme_grey(base_size = 25) +
  labs(x = "Country",y = "Gross Exports",title = "Box plot - Gro
ss Exports By Country (After Scaling)",tag = "B3")
gridExtra::grid.arrange(
  scatter_plot_before_scale,
  scatter_plot_after_scale,
  boxplot_Import_before_scale,
  boxplot_Import_after_scale,
  boxplot_Export_before_scale,
  boxplot_Export_after_scale,
  ncol=2
)

```

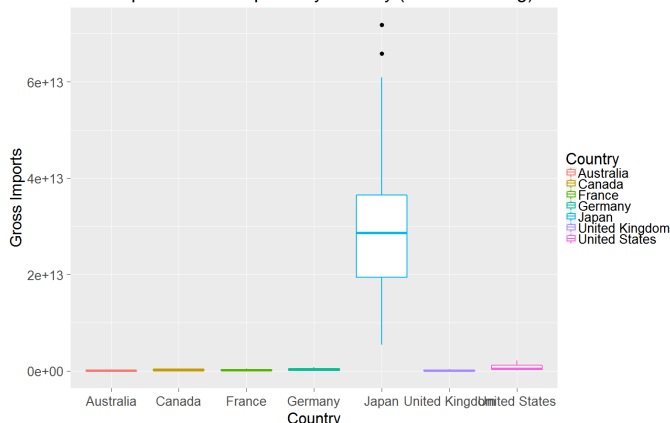
Scatter plot - Gross Imports vs. Gross Exports (Before Scaling)



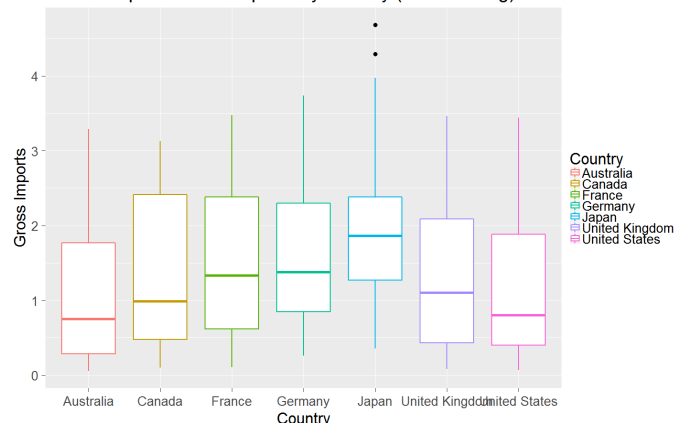
Scatter plot - Gross Imports vs. Gross Exports (After Scaling)



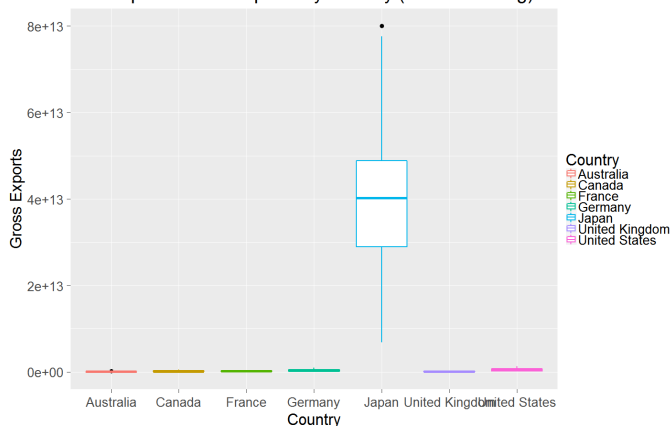
Box plot - Gross Imports By Country (Before Scaling)



Box plot - Gross Imports By Country (After Scaling)



Box plot - Gross Exports By Country (Before Scaling)



Box plot - Gross Exports By Country (After Scaling)

