The US government wants to understand the net inflow and outflow of cash from the country every year. The US trade in goods and services is summarised in terms of Balance of Payments. The Balance of Payments is the difference between Exports and Imports and signifies if there was a Net outflow of cash from the US or a net inflow. This data has been provided from the US Census Bureau, Economic Indicator Division.

* This data is available for Goods and Services from the year 1960 to 2015.
* All Values are in Millions of Dollars

Questions:-

1. Is the Balance of Payments of Goods and Services correlated?
2. Is the correlation real or coincidence ?

# D- understand the correlation between Balance of Payments of Goods and Services

# Understadn if Balance of Payments for Exports and Imports can be said to belong to the same population

# Import data

BOP <- read.csv("F:/springer book/Case study/CaseStudy6/BOP.csv", stringsAsFactors=FALSE)

# Check Format

str(BOP)

# Conclusion :- all are integers

# Check Descriptive statistics

install.packages("pastecs")

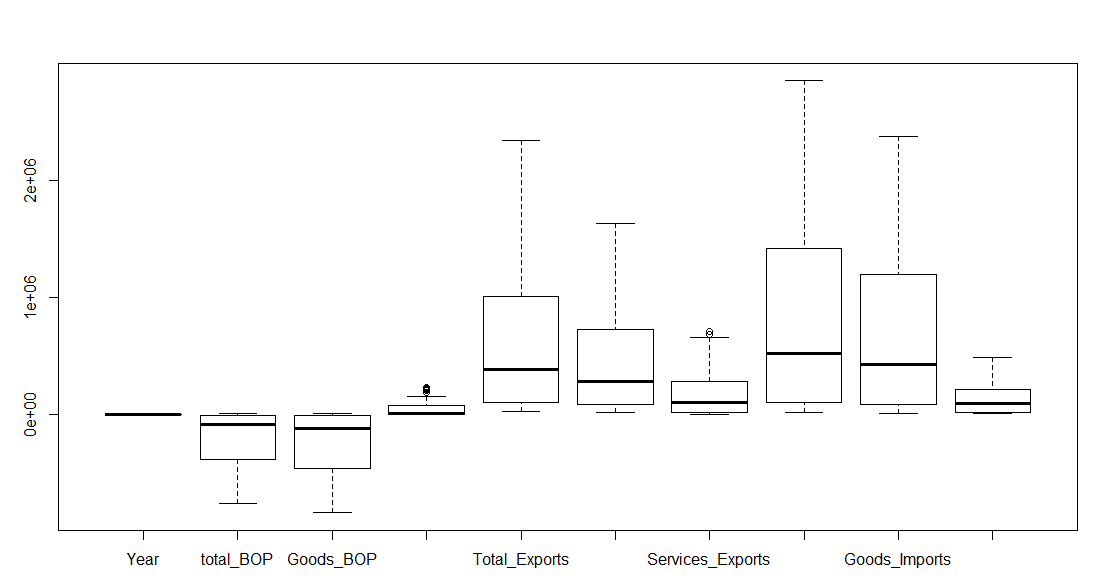
library(pastecs)

stat.desc(BOP)

# Conclusion :- No null or na values

# Draw boxplot

boxplot(BOP)



# There are outliers in most of the variables

# Remove outliers for total\_BOP, Goods\_BOP, Services\_BOP

# Remove outliers using The boxplot.stats function; is a ancillary

#function that produces statistics for drawing boxplots. It returns among

#other information a vector stats with five elements: the extreme of the

#lower whisker, the lower ‘hinge’, the median, the upper ‘hinge’ and the

#extreme of the upper whisker, the extreme of the whiskers are the adjacent

#values (last non-missing value, i.e. every value beyond is an outlier.

#Outliers are then all values outside the interval id1$stats[1] < and id1$stats[5]

id1 <- boxplot.stats(BOP$total\_BOP)

id2 <- boxplot.stats(BOP$total\_BOP, coef=2)

id1$stats

> id1$stats

[1] -761716.0 -378145.5 -87002.5 -2797.5 12404.0

id1$stats[1]

> id1$stats[1]

[1] -761716

id1$stats[5]

> id1$stats[5]

[1] 12404

# Conclusion - Outliers for BOP$total\_BOP is Below -761716 and above 12404

summary(BOP$total\_BOP)

> summary(BOP$total\_BOP)

Min. 1st Qu. Median Mean 3rd Qu. Max.

-761700 -375300 -87000 -189100 -3545 12400

# Conclusion - no outliers for total\_BOP

id1 <- boxplot.stats(BOP$Goods\_BOP)

id2 <- boxplot.stats(BOP$Goods\_BOP, coef=2)

id1$stats

id1$stats[1]

id1$stats[5]

summary(BOP$Goods\_BOP)

# Conclusion - There are values below the lower limit of -837289 which should be dropped

id1 <- boxplot.stats(BOP$Services\_BOP)

id2 <- boxplot.stats(BOP$Services\_BOP, coef=2)

id1$stats

id1$stats[1]

id1$stats[5]

summary(BOP$Services\_BOP)

# Conclusion - there are values above 154020 which should be dropped

# subset to remove outliers from 2 variables

NEWBOP<- subset(BOP, Goods\_BOP>-837289)

NEWBOP<-subset(NEWBOP, Services\_BOP<154020)

dim(NEWBOP)

> dim(NEWBOP)

[1] 49 10

# A- Analyse the data by running correlation between Services and Good BOP

cor(NEWBOP$Goods\_BOP, NEWBOP$Services\_BOP)

cor.test(NEWBOP$Goods\_BOP, NEWBOP$Services\_BOP)$p.value

> cor(NEWBOP$Goods\_BOP, NEWBOP$Services\_BOP)

[1] -0.7944912

> cor.test(NEWBOP$Goods\_BOP, NEWBOP$Services\_BOP)$p.value

[1] 9.497043e-12

# Conclusion - The correlation coefficient is -.79 and the p value is <=.05. Thus we can conclude that the correlation is real and should be accepted

# You can aslo use

#install.packages("Hmisc")

#library(Hmisc)

# rcorr(NEWBOP, type="pearson")

**Task for you: - Please run the correlation for**

1. **Goods\_Exports vs Services\_Exports**
2. **Goods\_Imports vs Services\_Imports**

**What is your conclusion?**