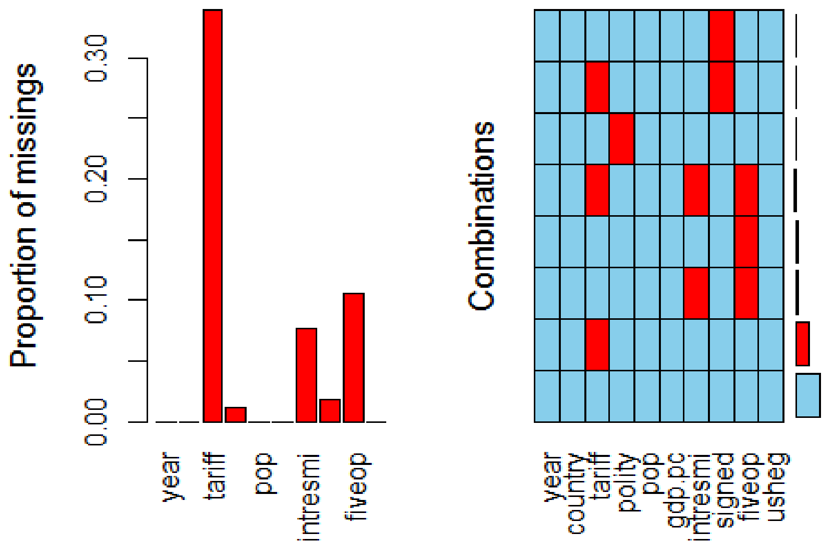
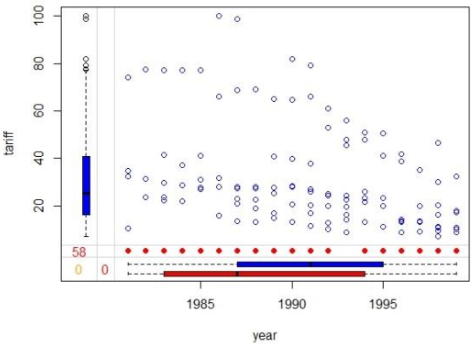
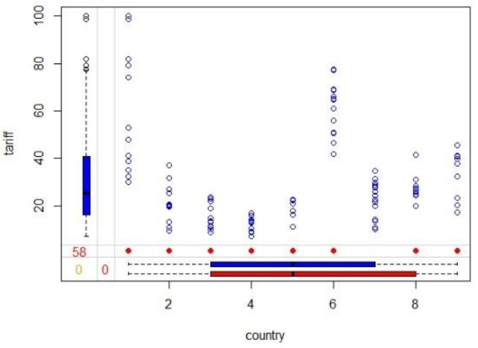
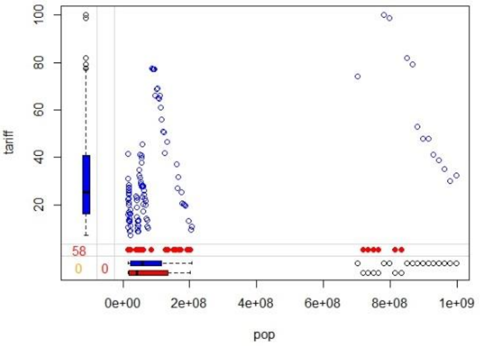
```
#==========================================================================================
#                          Using R: Missing Data Exploration
#==========================================================================================

#-----------------------------------------------------------------------------------------
#(a) Explore the "missingness" in the freetrade using your ...
library(Amelia)
data(freetrade)
summary(freetrade)
str(freetrade)
#We need to change the variable type to be useable
freetrade$year <- as.numeric(freetrade$year)
freetrade$polity <- as.numeric(freetrade$polity)
freetrade$signed <- as.numeric(freetrade$signed)
freetrade$country <- as.factor(freetrade$country)
#Exploring Missing Data
aggregate(freetrade, by=list(freetrade$country), function(x) mean(is.na(x)))
mean(is.na(freetrade$tariff))
## [1] 0.3391813
mean(is.na(freetrade$polity))
## [1] 0.01169591
mean(is.na(freetrade$intresmi))
## [1] 0.07602339
mean(is.na(freetrade$signed))
## [1] 0.01754386
mean(is.na(freetrade$fiveop))
## [1] 0.1052632
#Pattern of Missing Data
library(mice)
md.pattern(freetrade)
md.pairs(freetrade)
library(VIM)
summary(aggr(freetrade))
```



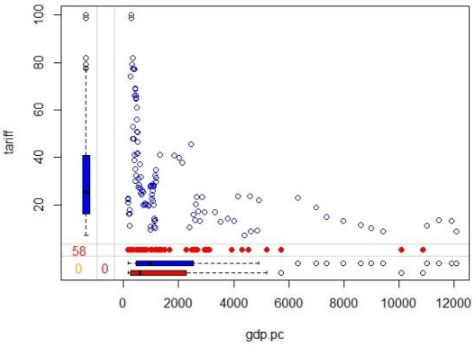
```
# Using Margin Plot and Scatter plot with Missing Data can help to see the
relationship of missing data and other variables.
for(i in c("year","country","pop","gdp.pc")){
  marginplot(freetrade[c(i, "tariff")], col = c("blue", "red", "orange"))
}
```

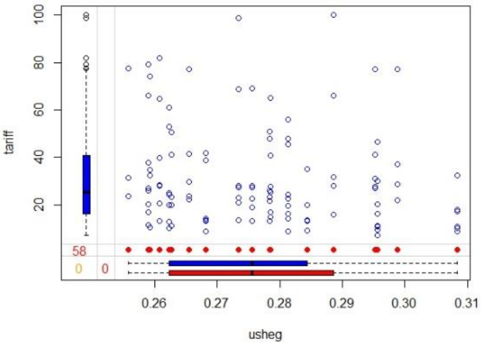


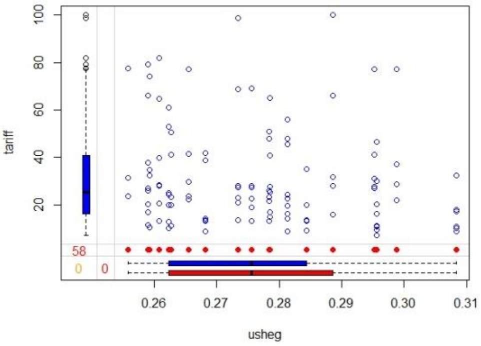




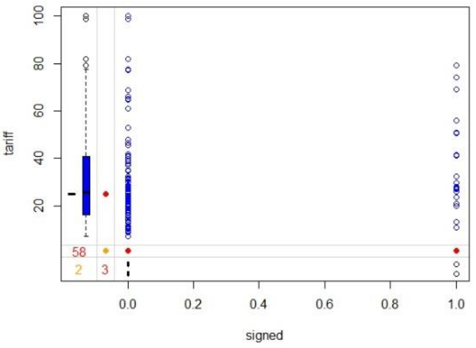
```
for(i in c("usheg","polity","usheg","polity")){
  marginplot(freetrade[c(i, "tariff")], col = c("blue", "red", "orange"))
}
```

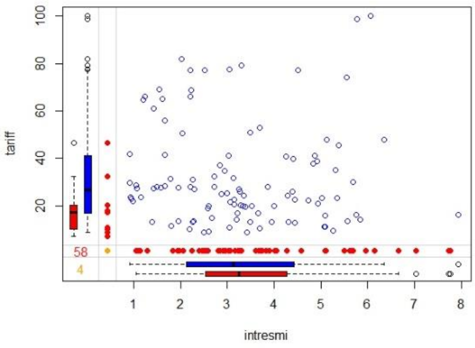


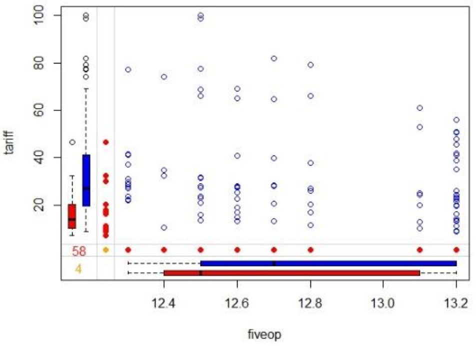




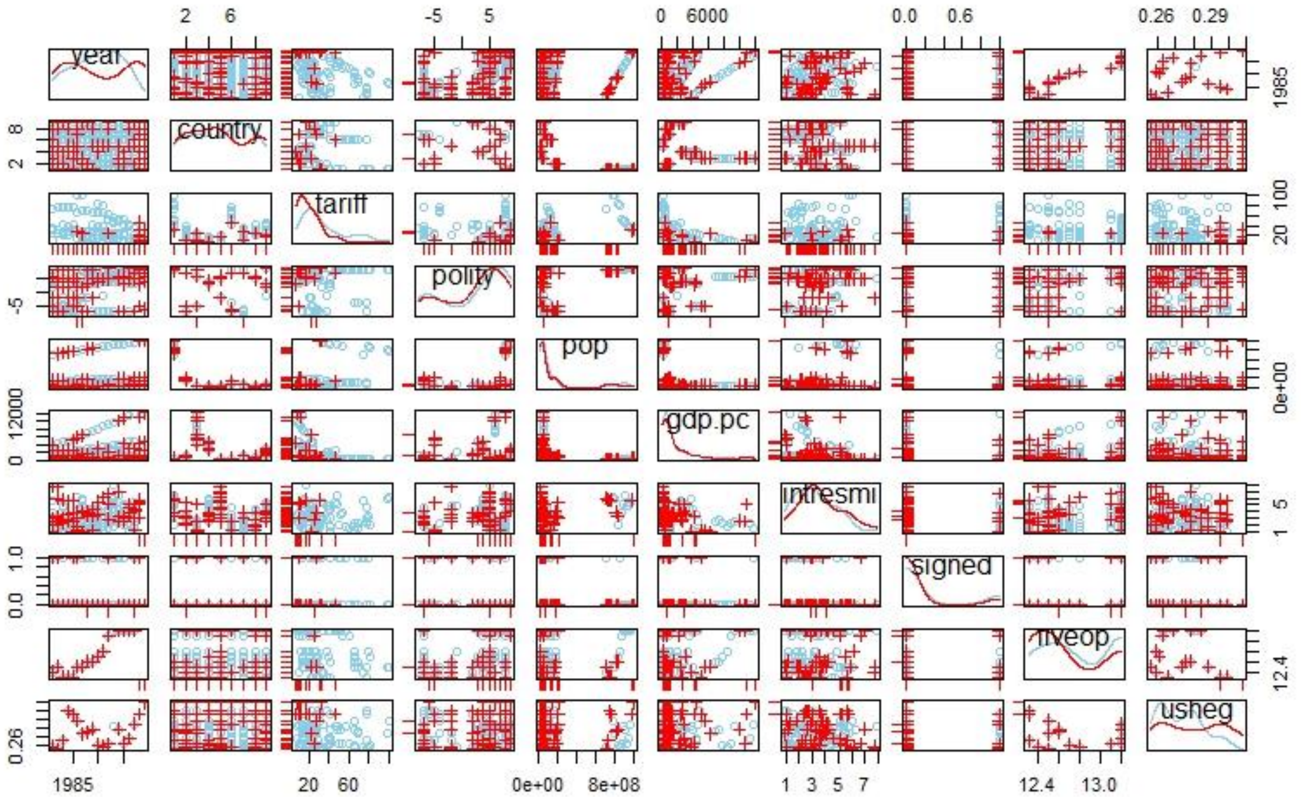
```
for(i in c("signed","intresmi","fiveop")){
  marginplot(freetrade[c(i, "tariff")], col = c("blue", "red", "orange"))
}
```







```r
scattmatrixMiss(freetrade)
```



```r
#There is missing data in tariff=58, polity=2, intresmi=13, signed=3 and fiveop=18
#The number of missing data in polity and signed are so small and can be neglected.
#The number of missing data in intresmi is less than 10 percent. but it is better to
be imputed. The number of missing data in fiveop and more specifically tariff are so
many and cannot be ignored. Most of the missing is for Nepal, Thailand, SriLanka and
Indonesia. It shows that missing is not MCAR. It can be the MAR. There is not an easy
way to say it is MNAR. Based on the margin plots and scatter plot the missing range of
Tariff is in the range of observed data and the densities of observed and missing data
```

```
is not have much deviation. There is chance that the missing data does not change the
overall trend.

#-------------------------------------------------------------------------------------
#(b) Implement your own statistical test (e.g. ANOVA, ...
aov.freetrade <- aov(tariff~country, data=freetrade)
summary(aov.freetrade)
##               Df Sum Sq Mean Sq F value Pr(>F)
## country        8  37349    4669   37.07 <2e-16 ***
## Residuals    104  13098     126

chisq.test(freetrade$country, freetrade$tariff)
## data:  freetrade$country and freetrade$tariff
## X-squared = 831.96, df = 736, p-value = 0.007819
#Effect of removal of Nepal
No.Nep.freetrade <- freetrade[freetrade$country != "Nepal",]
aov.freetrade <- aov(tariff~country, data=No.Nep.freetrade)
summary(aov.freetrade)
##              Df Sum Sq Mean Sq F value Pr(>F)
## country       7  35981    5140   38.76 <2e-16 ***
## Residuals    98  12995     133

chisq.test(No.Nep.freetrade$country, No.Nep.freetrade$tariff)
##  Pearson's Chi-squared test
##
## data:  No.Nep.freetrade$country and No.Nep.freetrade$tariff
## X-squared = 684.79, df = 602, p-value = 0.01063
#Effect of removal of Philippines
No.Phi.freetrade <- freetrade[freetrade$country != "Philippines",]
aov.freetrade <- aov(tariff~country, data=No.Phi.freetrade)
summary(aov.freetrade)
##              Df Sum Sq Mean Sq F value Pr(>F)
## country       7  35975    5139   36.27 <2e-16 ***
## Residuals    86  12188     142

chisq.test(No.Nep.freetrade$country, No.Phi.freetrade$tariff)
##  Pearson's Chi-squared test
##
## data:  No.Nep.freetrade$country and No.Phi.freetrade$tariff
## X-squared = 639.33, df = 574, p-value = 0.03012
#Both tests reject the hypothesis of being independence.
#Deleting the Nepal record increase the chance of being independent and the Deletion
#of Philippines increase this hypothesis more. However, they are still beyond the
critical point and we can still assume that they are dependent variables. Also, by
deletion of variables our sample is smaller and our results are less conclusive.
```