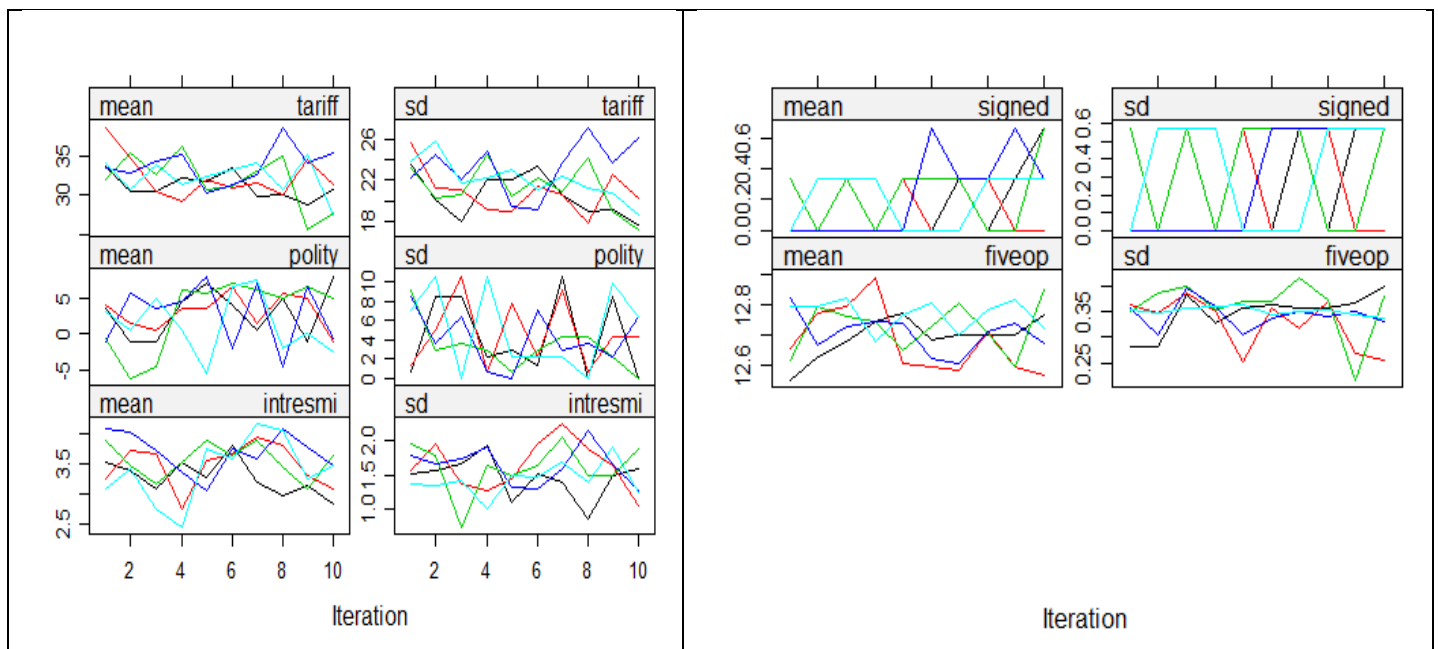


```

#=====
#                               Using R: Missing Data Imputation
#=====
#The freetrade data frame in the Amelia package has economic ...
library(Amelia)
data(freetrade)
summary(freetrade)
str(freetrade)
#We need to change the variable type to be useable
freetrade2 <- freetrade
freetrade2$year <- as.numeric(freetrade$year)
freetrade2$polity <- as.numeric(freetrade$polity)
freetrade2$signed <- as.numeric(freetrade$signed)
freetrade2$country <- as.numeric(as.factor(freetrade$country))
#-----
#(a) Perform the regression using Listwise deletion.
freetrade2.LD <- na.omit(freetrade2)
fit.a <- lm(data=freetrade2, tariff ~ year + country + polity + pop + gdp.pc +
intresmi + signed + fiveop + usheg)
f <- summary(fit.a)
c.a <- f[[4]]
se.a <- f[[6]]
#-----
#(b) Perform the regression using mean imputation.
freetrade2.MI <- freetrade2
freetrade2.MI[is.na(freetrade2$tariff), "tariff"] <- mean(freetrade2$tariff, na.rm=T)
freetrade2.MI[is.na(freetrade2$polity), "polity"] <- mean(freetrade2$polity, na.rm=T)
freetrade2.MI[is.na(freetrade2$intresmi), "intresmi"] <-
mean(freetrade2$intresmi, na.rm=T)
freetrade2.MI[is.na(freetrade2$signed), "signed"] <- mean(freetrade2$signed, na.rm=T)
freetrade2.MI[is.na(freetrade2$fiveop), "fiveop"] <- mean(freetrade2$fiveop, na.rm=T)
fit.b <- lm(data=freetrade2, tariff ~ year + country + polity + pop + gdp.pc +
intresmi + signed + fiveop + usheg)
f <- summary(fit.b)
c.b <- f[[4]]
se.b <- f[[6]]
#-----
#(c) Perform the regression using multiple imputation - in particular, ...
library(mice)
#for this data only these 4 method works. Here we continue using sample method.
#freetrade2.mice <- mice(freetrade2, m=5, maxit=10, method="mean")
#freetrade2.mice <- mice(freetrade2, m=5, maxit=10, method="rf")
freetrade2.mice <- mice(freetrade2, m=5, maxit=10, method="sample")
freetrade2.mice <- mice(freetrade2, m=5, maxit=10, method="cart")
str(freetrade2.mice)
freetrade2.mice$chainMean
freetrade2.mice$chainVar
plot(freetrade2.mice)

```



```
fit <- with(freetrade2.mice, lm(tariff ~ year + country + polity + pop + gdp.pc +
intresmi + signed + fiveop + usheg))
est <- pool(fit)
c.c <- summary(est)
```

#-----
##(d) Compare the coefficients for each of the regression models.

```
coefficient <- data.frame(c.a[,1], c.b[,1], c.c[,1])
coefficient
```

##	c.a...1.	c.b...1.	c.c...1.
## (Intercept)	2.862251e+03	2.862251e+03	1.738884e+03
## year	-1.417200e+00	-1.417200e+00	-8.582695e-01
## country	2.177828e+00	2.177828e+00	8.838798e-01
## polity	3.136722e-01	3.136722e-01	-3.163829e-02
## pop	5.801947e-08	5.801947e-08	3.195999e-08
## gdp.pc	-1.766775e-03	-1.766775e-03	-1.439169e-03
## intresmi	-2.486198e+00	-2.486198e+00	-1.870331e+00
## signed	-1.512382e-03	-1.512382e-03	2.084662e+00
## fiveop	-3.284342e+00	-3.284342e+00	9.863059e-01
## usheg	8.700365e+01	8.700365e+01	-4.587027e+01

##As can be seen by using mean imputation the coefficient would not be changed.
However, the coefficient of mice is highly related to the method we are choosing.