Missing data:

china=read.csv("china.csv",header=T)

apply(is.na(china),2,sum)

summary(china)

dim(china)

china=china[-27] #remove V317

india=read.csv("india.csv",header=T)

summary(india)

dim(india)

india=india[-27] #remove V317

pMiss=function(x){

sum(is.na(x))/length(x)\*100

}

ChinaPNA=apply(china,2,pMiss)

IndiaPNA=apply(india,2,pMiss)

apply(rbind(ChinaPNA,IndiaPNA),2,max)

mdChina=md.pattern(china)

mdIndia=md.pattern(india)

head(mdIndia)

aggr\_plot=aggr(china,col=c('navyblue','red'),numbers=TRUE,sortVars=TRUE,labels=names(china),cex.axis=.7,gap=3,ylab=c("Histogram of missing data","Pattern"))

imputed\_china=mice(china,m=5,maxit=50,meth='pmm',seed=500)

imputed\_india=mice(india,m=5,maxit=50,meth='pmm',seed=500)

densityplot(imputed\_china)

densityplot(imputed\_india)  
stripplot(imputed\_china,pch=20,cex=1.2)

stripplot(imputed\_india,pch=20,cex=1.2)

summary(imputed\_china)

summary(imputed\_india)

completeChina=complete(imputed\_china,1)

completeIndia=complete(imputed\_india,1)

write.csv(completeChina,file="CompleteChina.csv")

write.csv(completeIndia,file="CompleteIndia.csv")