# R Codes for Data Visualization

#Read data

data1 <- read.csv("C:/MA 299/R/Boston\_housing.csv")

#Columns’ names

colnames(data1)

#Descriptive statistics

summary(data1)

#Histogram

hist(data1$medv,main="Distribution of medv",xlab="medv")

#Bar chart

counts <- table(data1$chas)  
barplot(counts, main="Bar chart of chas",   
   xlab="Chas")

#Pie chart

mytable <- table(data1$chas)  
lbls <- paste(names(mytable), "\n", mytable, sep="")  
pie(mytable, labels = lbls,   
   main="Pie Chart of chas")

#Scatter plot

plot(data1$lstat,data1$medv,

main="Relationship Between lstat and medv",

xlab="lstat",

ylab="medv")

#Scatter plot (Matrix)

pairs(~medv+lstat+dis,data=data1,   
   main="Simple Scatterplot Matrix")

#Box plot

boxplot(medv~chas,data=data1, main="Housing Data",   
   xlab="chas", ylab="medv")

#Histogram (with ggplot 2)

install.packages("ggplot2")

library(ggplot2)

qplot(data = data1, x = medv)

qplot(data = data1, x = medv, binwidth = 5)

#Scatter plot (with ggplot 2)

qplot(data = data1, x = lstat, y = medv, color = chas)

qplot(data = data1, x = lstat, y = medv, size = chas)

#Box plot (with ggplot 2)

qplot(x = chas, y = medv, data = data1, geom = "boxplot")

#Stacked Plot

BankL <- read.csv("C:/MA 299/R/UniversalBankLogisticglm.csv")

attach(BankL)

library(ggplot2)

qplot(x = PersonalLoanf, y = Income, data = BankL, geom = "boxplot")

counts <- table(CDf, PersonalLoanf)

counts

mytable<-prop.table(counts, 2)

mytable

barplot(mytable, main="Stacked plot between PersonalLoanf and CDf", xlab="PersonalLoanf", col=c("darkblue","red"))

#Stacked Plot (an alternative)

BankL <- read.csv("C:/MA 299/R/UniversalBankLogisticglm.csv")

attach(BankL)

library(ggplot2)

library(ggthemes)

library(extrafont)

library(plyr)

library(scales)

library(reshape2)

counts <- table(CDf, PersonalLoanf)

mytable<-prop.table(counts, 2)

mydata<-as.data.frame(mytable)

plots <- ggplot() + geom\_bar(aes(y = Freq, x = PersonalLoanf, fill = CDf),

data = mydata, stat="identity")

#Heatmap with values

library(gplots)  
heatmap.2(cor(housing.df), Rowv = FALSE, Colv = FALSE, dendrogram = "none",   
 cellnote = round(cor(housing.df),2),   
 notecol = "black", key = FALSE, trace = 'none', margins = c(10,10))

#Pairwise scatter plots  
plot(housing.df[, c(1, 3, 12, 13)])

#Alternative plots  
library(GGally)  
ggpairs(housing.df[, c(1, 3, 12, 13)])

#Parallel coordinates plots  
library(MASS)  
parcoord(housing.df[housing.df$CAT..MEDV == 0, -14], main = "CAT.MEDV = 0")  
parcoord(housing.df[housing.df$CAT..MEDV == 1, -14], main = "CAT.MEDV = 1")

#Scatter plot with labelled points  
utilities.df <- read.csv("C:/MA 299/R/Utilities.csv")  
plot(utilities.df$Fuel.Cost ~ utilities.df$Sales,   
 xlab = "Sales", ylab = "Fuel Cost", xlim = c(2000, 20000))  
text(x = utilities.df$Sales, y = utilities.df$Fuel.Cost,   
 labels = utilities.df$Company, pos = 4, cex = 0.8, srt = 20, offset = 0.2)