1. Use the given link Data Set

Answer the below questions:

Data Preprocessing:

setwd("C:/Users/Ashwin/Desktop/AcadGild assignments/assignment\_13.1")

unzip("crime.zip")

crime\_data <- read.table("Crimes\_-\_2001\_to\_present.csv",header = F,sep=",")

names(crime\_data) <- c("Case", "Number", "Date", "Block", "IUCR", "Primary Type", "Description",

"Location Desc", "Arrest", "Domestic", "Beat", "District", "Ward", "Community Area",

"FBI Code", "X Coordinate", "Y Coordinate", "Year", "Updated On",

"Latitude", "Longitude", "Location")

#checking missing values

summary(crime\_data[,1:22])

#imputing missing values

crime\_data <- crime\_data[1:291267,]

crime\_data$Ward[171220] <- round(mean(crime\_data$Ward,na.rm = T))

crime\_data$Ward[235381] <- round(mean(crime\_data$Ward,na.rm = T))

for(i in 1:291267){

if(is.na(crime\_data$`X Coordinate`[i])){

crime\_data$`X Coordinate`[i] <- round(mean(crime\_data$`X Coordinate`,na.rm = T))

crime\_data$`Y Coordinate`[i] <- round(mean(crime\_data$`Y Coordinate`,na.rm = T))

crime\_data$Latitude[i] <- mean(crime\_data$Latitude,na.rm = T)

crime\_data$Longitude[i] <- mean(crime\_data$Longitude,na.rm = T)

}

}

plot\_data <- crime\_data[,c(11,12,13,14,16,17,20,21)]

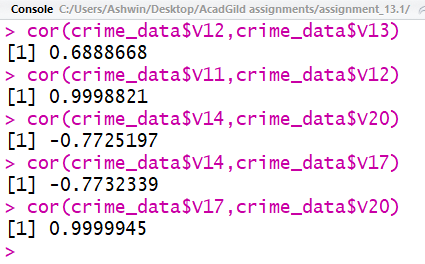
any(grepl("corrplot",installed.packages()))

install.packages("corrplot")

library(corrplot)

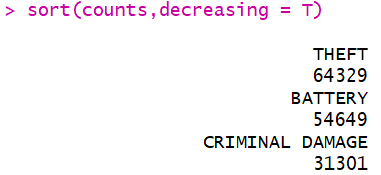
a. Find out top 5 attributes having highest correlation (select only Numeric features).

ANS.



b. Find out top 3 reasons for having more crime in a city.

ANS.



c. Which all attributes have correlation with crime rate?

ANS.

Crimes <- read.csv("C:/Users/potdarjs/Downloads/Crimes\_-\_2001\_to\_present.csv", header=FALSE)

View(Crimes)

names(Crimes) <- c("Case", "Number", "Date", "Block", "IUCR", "Primary Type", "Description",

                   "Location Desc", "Arrest", "Domestic", "Beat", "District", "Ward", "Community Area",

                   "FBI Code", "X Coordinate", "Y Coordinate", "Year", "Updated On",

                   "Latitude", "Longitude", "Location")

head(Crimes)

str(Crimes)

Crimes <- na.omit(Crimes)

names(Crimes)

c <- **cor(Crimes[c(11,12,13,14,18,20,21)])**

c

library(reshape2)

m <- melt(c)

library(dplyr)

m

top <- m%>%select(X1, X2, value)%>%filter(value != 1)

top[order(top$value, decreasing = T)[1:10],]

x <- as.data.frame(table(Crimes$Description))

x[order(x$Freq, decreasing = T)[1:3],]

crime <- Crimes

head(crime)

table([is.na](http://is.na/" \t "_blank)(crime))

crime$Date <- as.POSIXlt(crime$Date, format= "%m/%d/%Y %H:%M:%S")

crime$`Updated On` <- as.POSIXlt(crime$`Updated On`, format= "%m/%d/%Y %H:%M:%S")

library(chron)

crime$Time <- times(format(crime$Date,"%H:%M:%S"))

crime$Date <- as.POSIXct(crime$Date)

crime$`Updated On` <- as.POSIXct(crime$`Updated On`)

time.tag <- chron(times=c("00:00:00", "06:00:00", "12:00:00", "18:00:00","23:59:00"))

time.tag

crime$time.tag <- cut(crime$Time, breaks= time.tag,

                      labels= c("00-06","06-12", "12-18", "18-00"), include.lowest =TRUE)

table(crime$time.tag)

crime$date <- as.POSIXlt(strptime(crime$Date, format = "%Y-%m-%d"))

crime$date <- as.POSIXct(crime$date)

crime$day <- as.factor(weekdays(crime$Date, abbreviate = TRUE))

crime$month <- as.factor(months(crime$Date, abbreviate = TRUE))

str(crime$day)

str(crime$month)

crime$Arrest <- ifelse(as.character(crime$Arrest) == "true",1,0)

crime$crime <- as.character(crime$`Primary Type`)

crime$crime <- ifelse(crime$crime %in% c("CRIM SEXUAL ASSAULT","PROSTITUTION", "SEX OFFENSE","HUMAN TRAFFICKING"), 'SEX', crime$crime)

crime$crime <- ifelse(crime$crime %in% c("MOTOR VEHICLE THEFT"), "MVT", crime$crime)

crime$crime <- ifelse(crime$crime %in% c("GAMBLING", "INTERFEREWITH PUBLIC OFFICER", "INTERFERENCE WITH PUBLIC OFFICER", "INTIMIDATION",

                                         "LIQUOR LAW VIOLATION", "OBSCENITY", "NON-CRIMINAL", "PUBLIC PEACE VIOLATION",

                                         "PUBLIC INDECENCY", "STALKING", "NON-CRIMINAL (SUBJECT SPECIFIED)","NON - CRIMINAL"),

                      "NONVIO", crime$crime)

crime$crime <- ifelse(crime$crime == "CRIMINAL DAMAGE", "DAMAGE",crime$crime)

crime$crime <- ifelse(crime$crime == "CRIMINAL TRESPASS","TRESPASS", crime$crime)

crime$crime <- ifelse(crime$crime %in% c("NARCOTICS", "OTHER NARCOTIC VIOLATION", "OTHER NARCOTIC VIOLATION"), "DRUG", crime$crime)

crime$crime <- ifelse(crime$crime == "DECEPTIVE PRACTICE","FRAUD", crime$crime)

crime$crime <- ifelse(crime$crime %in% c("OTHER OFFENSE", "OTHEROFFENSE"), "OTHER", crime$crime)

crime$crime <- ifelse(crime$crime %in% c("KIDNAPPING", "WEAPONS VIOLATION", "CONCEALED CARRY LICENSE VIOLATION","OFFENSE INVOLVING CHILDREN"), "VIO", crime$crime)

table(crime$crime)