SESSION 12: Generalized Linear Models--Assignment 1

```
#Problem 1

#Answer the below questions:

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

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

#c. Which all attributes have correlation with crime rate?

library(readr)

library(data.table)

getwd()

p<-"D:/ACADGILD/Assignments from Dashboard"

setwd(p)

COBRA.YTD2017 <- read.csv("D:/ACADGILD/Assignments from Dashboard/COBRA-YTD2017.csv")

View(COBRA.YTD2017)

str(COBRA.YTD2017)

summary(COBRA.YTD2017)

sum(is.na(COBRA.YTD2017))

COBRA_YTD[,is.na(COBRA.YTD2017$loc_type)]<- mean(COBRA.YTD2017, na.rm=TRUE)

library(car)

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

fit<-lm(beat~MinOfucr+MaxOfnum_victims+loc_type+neighborhood+x+y,data =COBRA.YTD2017, na.action = na.omit)

fit

summary(fit)

vif(fit)

fit1<-lm(formula=MinOfucr~beat+MaxOfnum_victims+loc_type+neighborhood+x+y,data =COBRA.YTD2017)
```

```
fit1

summary(fit1)

vif(fit1)

vif(fit)>5

vif(fit1)>5

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


library(ggplot2)

COBRA.YTD2017$hour <- sub(":.*", "", COBRA.YTD2017$occur_time)

COBRA.YTD2017$hour <- as.numeric(COBRA.YTD2017$hour)

ggplot(aes(x = hour), data = COBRA.YTD2017) + geom_histogram(bins = 24, color='white',
fill='red') +

ggtitle('Histogram of Crime Time')


z<-table(COBRA.YTD2017$UC2.Literal)

hist(z)

#c. Which all attributes have correlation with crime rate?

library(ggplot2)

pairs(COBRA.YTD2017)

install.packages("corrplot")

library(corrplot)

rank1<-sample(COBRA.YTD2017[1:100,22:23], 20, replace=T)

rank2<-sample(COBRA.YTD2017[1:100,22:23], 20, replace=T)

cbind(rank1,rank2)

plot(rank1, rank2)

cor(rank1,rank2, method="spearman")

cor(rank1,rank2, method="pearson")
```