Answer the below questions:

a. Visualize the correlation between all variables in a meaningful and clear way of representing. Find out

top 3 reasons for having more crime in a city.

Sol= # I will use the data prepared in the assignment 7.1

crimemodel <- read.csv("E:/Acadgild/Class 7/Assignments/crimemodel.csv")

View(crimemodel)

str(crimemodel)

library(dplyr)

crimemodel1 <- crimemodel

crimemodel1 <- mutate(crimemodel, day = as.numeric(as.factor(day)),

month = as.numeric(as.factor(month)),

season = as.numeric(as.factor(season)))

names(crimemodel1)

str(crimemodel1)

correlation <- cor(crimemodel1[,c("Beat","count","Arrest","day","month","past.crime.1",

"past.crime.7","past.crime.30","past.arrest.30","crime.trend",

"policing","season")])

psych::cor.plot(correlation)

b. What is the difference between co-variance and correlation? Take an example from this dataset and

show the differences if any?

Sol= # Co-Variance is a systematic relationship between a pair of random variables where in a

# change in one variable reciprocated by an equivalent change in another variable.

# Measure of correlation, Lie between -??? and +???, Change in scale affects covariance

# Correlation is statistical measure that indicates how strongly two variables are related.

# Scaled version of covariance, Lie between -1 and +1,

# Change in scale does not affect the correlation. Unit free measure

# Correlation is a special case of covariance which can be obtained when the data is standardized.

cov(crimemodel1$count, crimemodel1$past.crime.30)

cor(crimemodel1$count, crimemodel1$past.crime.30)

# covariance’s is hard to compare as the value ranges from -??? and +???

# we get a different covariance from when we do it in other units

# Then we need to 'normalize' the covariance by means divide the covariance

# by something that represents the diversity and scale in both the covariates,

# and end up with a value that is assured to be between -1 and 1: the correlation