library(lubridate)

ftse<-read.csv("C:/Users/pc/Desktop/akano-documents/ehi r assignment/ftse.csv")

spx<-read.csv("C:/Users/pc/Desktop/akano-documents/ehi r assignment/spx.csv")

ftse$new\_date<-dmy(ftse$Date)

spx$new\_date<-dmy(spx$Date)

install.packages("tidyverse")

library(tidyverse)

#DROP THE COLUMN Date

spx<-spx %>% select(-Date)

ftse<-ftse %>% select(-Date)

#arrange by new\_date

ftse<-ftse %>% arrange(new\_date)

spx<-spx %>% arrange(new\_date)

head(ftse)

head(spx)

#merge two data sets

total<-merge(ftse, spx, by="new\_date")

head(total)

nrow(total)

tail(total)

#Generate time series plot

library(ggplot2)

ts\_spx<- ggplot(spx, aes(x=new\_date, y=spx)) + geom\_line()

ts\_spx

#GENERATE SCATTER PLOT

scatter.smooth(total$spx ,total$ftse, main="UK VS US")

boxplot(total$spx , main="BOXPLOT FOR spx")

boxplot(total$ftse , main="BOXPLOT FOR ftse")

#GENERATE HISTOGRAM

hist(total$ftse , main="HISTOGRAM FOR FTSE")

hist(total$spx , main="HISTOGRAM FOR SPX")

#VIEW SUMMARY OF DATA

summary(total$ftse)

summary(total$spx)

#FIT LINEAR REGRESSION MODEL

model <- lm(total$ftse~total$spx)

summary(model)

#RENAME COLUMN new\_date to date

total<-total %>% rename(date=new\_date)

#generate year variable

total$year<-year(total$date)

#ARITHMETRIC MEAN

ftse\_mean <- aggregate(ftse ~ year, total, FUN=mean)

ftse\_mean

spx\_mean <- aggregate(spx ~ year, total, FUN=mean)

#GEOMETRIC MEAN

library('psych')

ftse\_geo\_mean <- aggregate(ftse ~ year, total, FUN=geometric.mean)

ftse\_geo\_mean

spx\_geo\_mean <- aggregate(spx ~ year, total, FUN=geometric.mean)

#GENERATE COLUMN FOR RUWAR AND COVID

total$RUWAR <- ifelse((total$date < as.Date('2022-02-24')), 0, 1)

head(total)

total$COVID <- ifelse((total$date < as.Date('2020-01-29')), 0, 1)

head(total)

tail(total)

#FIT MULTIPLE REGRESSION MODEL

model2 <- lm(total$ftse ~ total$spx+ total$RUWAR + total$COVID)

summary(model2)

#REGRESSION EQUATION

# ftse=5087+0.7881+4967RUWAR-1429COVID