**Homework Assignment 2**

**STA 141A A02**

**Submitted By:**

#Libraries to import

airfare<- read.csv("airfare/airfare.csv")

library("readxl")

b<- read\_xlsx("airfare/cpi\_1996\_2017.xlsx",range="A12:O34")

#1

table1a<-airfare[airfare$table%in%"1a"]

table1a<-split(airfare,airfare$table=="1a")

table6<-split(airfare,airfare$table=="6")

temp<- split(airfare,airfare$table)

table1a<- temp[[1]]

table6<-temp[[2]]

#2

#range of values

range(table1a$year)

range(table6$year)

range(table1a$quarter)

range(table6$quarter)

#2b

# no data for table 1a

table1a[which(is.na(table1a))]

#how many na present in table

summary(table1a)

table(table1a$year)

table(table6$year)

#null values for data in quarter

w<-table(table1a$year, table1a$quarter)

q<-table(table6$year, table6$quarter)

#data.frame(table1a$year,table6$year)

table(table1a$year)

View(table6)

#3

d1<- data.frame(table6[table6$year==2017,],table6[table6$year==2017,])

d2<-data.frame("City1"=d1$city1,"City2"=d1$city2)

d3<-split(d2,d2$City1)

#plot(density(d2))

z<- sapply(d3,nrow)

w<-data.frame("CityName"=levels(unique(d2$City1)),"City1 Connections"=z)

ord<-order(w$City1.Connections,decreasing="True")

tsu<-w[ord,]

head(tsu)

hist(z,main="City 1 connections",xlab = "City plot for year 2017",ylab="Connections in year 2017")

d4<-split(d2,d2$City2)

z1<-sapply(d4,nrow)

z1[names="Wichita, KS"]

z1\_1<-table(d2$City2)

sort(z1\_1)

q<-as.factor(z1\_1)

city2\_level<-levels(q)

z1\_2<-table(d2$City1)

sort(z1\_2)

q1<-as.factor(z1\_2)

city1\_level<-levels(q1)

city\_levels<-union(city1\_level,city2\_level)

z4<- as.factor(q,levels=city\_levels)

levels(q1)<-city\_levels

levels(q)<-city\_levels

w<- c(q,q1)

qw<-sort(w,decreasing = T)

head(qw)

qw1<-sort(w)

head(qw1)

#Similarly for year 2007 and 1997.

#4

par(mfrow=c(1,1))

table6$dates<-c("02/15","05/15","08/15","11/15")[table6$quarter]

#table6$dates=="May-15"

table(table6$dates)

wer<-paste(table6$dates,sep="/",table6$year)

table(wer)

q<-table6[table6$year==1999,]

mean(q$fare)

q1<-table6[table6$year==2001,]

mean(q1$fare)

q2<-table6[table6$year==2002,]

mean(q2$fare)

Dates<-as.Date(wer,"%m/%d/%Y")

listwanted<-data.frame("Avg Passengers"=table6$passengers/4,"Dates of quarter"=Dates)

ListSplitted<-split(listwanted,listwanted$Dates.of.quarter)

qwe<-tapply(listwanted$Avg.Passengers,listwanted$Dates.of.quarter,sum)

#plot(qwe,ylab = "Dates per Quarter",xlab="Avg Passengers")

table(listwanted)

table(qwe)

head(qwe)

View(qwe)

barplot(qwe,ylab = "Total Passengers",xlab="Mid-Dates per quarter",main = "Total Passenger Distribution per Quarter")

#5

library("readxl")

b<- read\_xlsx("airfare/cpi\_1996\_2017.xlsx",range="A12:O34")

View(b)

colnames(b)

cpi<-b[,c("Feb", "May", "Aug","Nov")]

head(cpi)

dim(cpi)

z<-data.frame("Year Number"=b$Year,"CPI Val"=b$HALF1)

x<- z[z$Year.Number==2017,]

x$CPI.Val

w<-data.frame("Year Number table6"=a$year,"fare"=a$fare)

ww<-w[w$Year.Number.table6==2017,]

qw<-b[b%in%c("Feb","May","Aug"),]

table6$real17\_fare=table6$fare\*(x$CPI.Val)/(b$Feb)

View(table6)

##merging two datasets in R

newTable<-merge(table6,b,by.x="year",by.y="Year")

er<-c("Feb","May","Aug","Nov")

newTable$real17\_fare=newTable$fare\*(x$CPI.Val)/(newTable$Feb)

nrow(newTable)

head(newTable)

#6

boxplot(newTable$real17\_fare~newTable$year,ylab="Airfare",xlab="Time",main="Airfare Variation With Time")

#7

par(mfrow=c(1,1))

t1<-table1a[table1a$year==2015,]

smoothScatter(t1$fare,t1$miles)

scatter.smooth(t1$fare,t1$miles,main="Table1a for 2015")

abline( 4.025,324.949,col="red")

par(mfrow=c(1,1))

t1<-table6[table6$year==2015,]

smoothScatter(t1$fare,t1$miles)

scatter.smooth(t1$fare,t1$miles,main="Table6 for 2015")

par(mfrow=c(2,4))

model<-lm(miles~fare,table1a[table1a$year==2015,])

plot(model,which=1,main="Table 1a miles and fare")

summary(model)

model1<-lm(miles~fare,table6[table6$year==2015,])

plot(model,which=1,main="Table 6 miles and fare")

summary(model1)

model3<-lm(miles~fare,table1a[table1a$year==2015,])

plot(model3,which=2,main="Table 1a miles and fare")

model4<-lm(miles~fare,table6[table6$year==2015,])

plot(model4,which=2,main="Table 6 miles and fare")

model5<-lm(miles~fare,table1a[table1a$year==2015,])

plot(model5,which=3,main="Table 1a miles and fare")

model6<-lm(miles~fare,table6[table6$year==2015,])

plot(model6,which=3,main="Table 6 miles and fare")

model7<-lm(miles~fare,table1a[table1a$year==2015,])

plot(model7,which=5,main="Table 1a miles and fare")

model8<-lm(miles~fare,table6[table6$year==2015,])

plot(model8,which=5,main="Table 6 miles and fare")

plot(lm(miles~fare,table1a), las = 1)

par(mfrow=c(2,3))

q1<-table6[table6$year==2015,]

m<-smoothScatter(q1$miles,q1$fare,main="Table 6")

m2<-scatter.smooth(q1$miles,q1$fare,main="Table 6")

plot(lm(miles~fare,q1),which=1)

plot(lm(miles~fare,q1),which=2)

plot(lm(miles~fare,q1),which=3)

plot(lm(miles~fare,q1),which=5)

par(mfrow=c(2,2))

q2<-table1a[table1a$year==2015,]

plot(lm(miles~fare,q2),which=1)

plot(lm(miles~fare,q2),which=2)

plot(lm(miles~fare,q2),which=3)

plot(lm(miles~fare,q2),which=5)

#8

par(mfrow=c(2,4))

q1<-table6[table6$year==2015,]

plot(lm(fare~passengers+miles,q1),which=1,main = "Fare with miles and passenger Table6")

plot(lm(fare~passengers+miles,q1),which=2 ,main = "Fare with miles and passenger Table6")

plot(lm(fare~passengers+miles,q1),which=3,main = "Fare with miles and passenger Table6")

plot(lm(fare~passengers+miles,q1),which=5,main = "Fare with miles and passenger Table6")

q2<-table1a[table1a$year==2015,]

plot(lm(fare~miles+passengers,q2),which=1,main = "Fare with miles and passenger Table1a")

plot(lm(fare~miles+passengers,q2),which=2,main = "Fare with miles and passenger Table1a")

plot(lm(fare~miles+passengers,q2),which=3,main = "Fare with miles and passenger Table1a")

plot(lm(fare~miles+passengers,q2),which=5,main = "Fare with miles and passenger Table1a")

#9

qw<-table6[table6$year==2015,]

qw1<-qw[qw$fare>qw$lg\_fare,]

#few of the city pairs

aw<-data.frame("City1"=qw1$city1,"City2"=qw1$city2)

head(aw)

#plot graphics.

plot(qw1$lg\_fare,qw1$fare)

#graphics and model to find a fit

par(mfrow=c(2,2))

plot(lm(fare~lg\_fare,qw1),which=1)

summary(lm(fare~lg\_fare,qw1))

plot(lm(fare~lg\_fare,qw1),which=2)

plot(lm(fare~lg\_fare,qw1),which=3)

plot(lm(fare~lg\_fare,qw1),which=5)

#10

par(mfrow=c(2,2))

states=table1a[table1a$airport1 %in% c("SMF","OAK","SFO","SJC"),]

qw<-data.frame("Airport Names"=states$airport1,"Avg fare"=states$fare)

qwe<-split(qw,qw$Airport.Names)

table1a$airport1

o<-droplevels(qw)

table(table1a$city2)

#View(table1a$city1)

boxplot(o$Avg.fare~o$Airport.Names,main="Airport1",,xlab="Airportnames",ylab="Fare")

#barplot(o$Airport.Names,o$Avg.fare,main="Airport1",,xlab="Airportnames",ylab="Fare")

#Airport2

states=table1a[table1a$airport2 %in% c("SMF","OAK","SFO","SJC"),]

qw2<-data.frame("Airport Names"=states$airport2,"Avg fare"=states$fare)

qwe<-split(qw2,qw2$Airport.Names)

o<-droplevels(qw2)

boxplot(o$Avg.fare~o$Airport.Names,main="Airport2",,xlab="Airportnames",ylab="Fare")

#long distance connections

#airport1

states=table1a[table1a$airport1 %in% c("SMF","OAK","SFO","SJC"),]

qw<-data.frame("Airport Names"=states$airport1,"distance"=states$miles)

qwe<-split(qw,qw$Airport.Names)

table1a$airport1

o<-droplevels(qw)

boxplot(o$distance~o$Airport.Names,main="Airport1",xlab="Airportnames",ylab="Miles")

#Airport2

states=table1a[table1a$airport2 %in% c("SMF","OAK","SFO","SJC"),]

qw2<-data.frame("Airport Names"=states$airport2,"distance"=states$miles)

qwe<-split(qw2,qw2$Airport.Names)

o<-droplevels(qw2)

boxplot(o$distance~o$Airport.Names,main="Airport2",xlab="Airportnames",ylab="Miles")

## Do this differ by year

par(mfrow=c(2,2))

table2<-table1a[table1a$year==2017,]

states=table2[table2$airport2 %in% c("SMF","OAK","SFO","SJC"),]

qw2<-data.frame("Airport Names"=states$airport2,"distance"=states$miles)

qwe<-split(qw2,qw2$Airport.Names)

o<-droplevels(qw2)

boxplot(o$distance~o$Airport.Names,main="Airport2 per 2017",xlab="Airportnames",ylab="Miles")

table2<-table1a[table1a$year==2016,]

states=table2[table2$airport2 %in% c("SMF","OAK","SFO","SJC"),]

qw2<-data.frame("Airport Names"=states$airport2,"distance"=states$miles)

qwe<-split(qw2,qw2$Airport.Names)

o<-droplevels(qw2)

boxplot(o$distance~o$Airport.Names,main="Airport2 in 2016",xlab="Airportnames",ylab="Miles")

table2<-table1a[table1a$year==2015,]

states=table2[table2$airport2 %in% c("SMF","OAK","SFO","SJC"),]

qw2<-data.frame("Airport Names"=states$airport2,"distance"=states$miles)

qwe<-split(qw2,qw2$Airport.Names)

o<-droplevels(qw2)

boxplot(o$distance~o$Airport.Names,main="Airport2 in 2015",xlab="Airportnames",ylab="Miles")

Similarly for year 2014.