Lab 2 part 1

EPI\_Data <- read.csv(file.choose(),header=T)

attach(EPI\_Data)

head(EPI\_Data)

library(dplyr)

####Central Tendancies####

mean(EPI)

mode(EPI)

median(EPI)

mean(DALY)

mode(DALY)

median(DALY)

####Histograms####

hist(EPI)

hist(DALY)

####5 Random Data Points####

sample\_n(data.frame(EPI,DALY),5)

####10% Random Data Points####

sample\_frac(data.frame(EPI,DALY),.1)

###arrange in descending####

help(arrange)

new\_decs\_EPI <- arrange(data.frame(EPI,DALY),desc(EPI))

new\_decs\_DALY <- arrange(data.frame(EPI,DALY),desc(DALY))

new\_decs\_EPI

new\_decs\_DALY

####Mutate####

help(mutate)

mutate(data.frame(EPI), double\_EPI = EPI \* 2)

mutate(data.frame(EPI), double\_DALY = DALY \* 2)

####summarise####

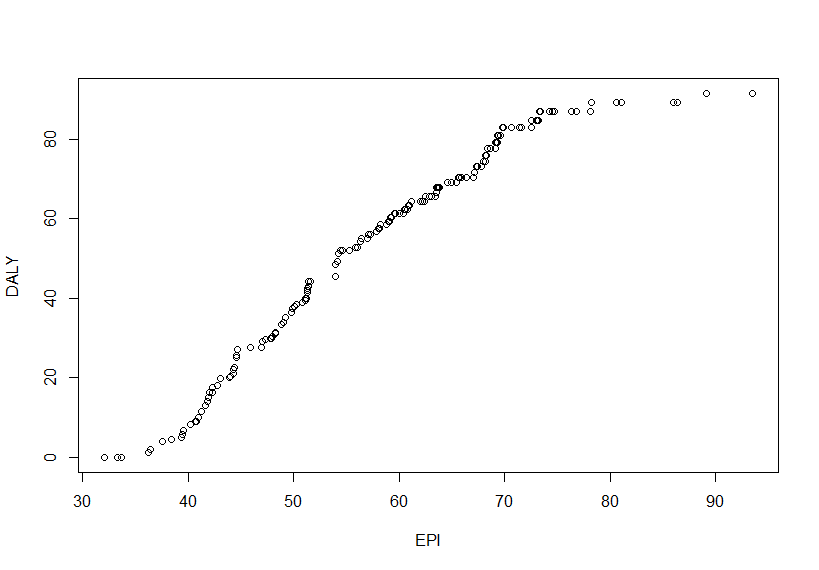
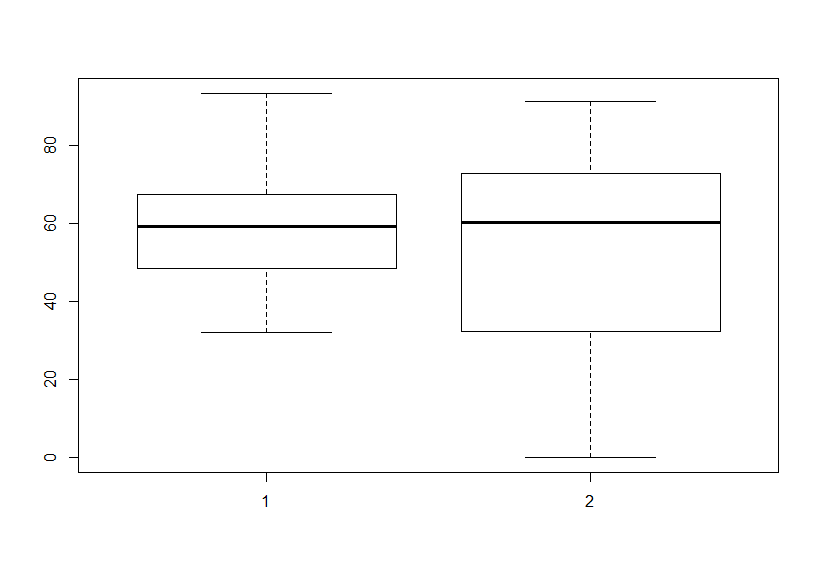
summarise(data.frame(EPI,DALY),mean\_EPI = mean(EPI),mean\_DALY = mean(DALY))

###boxplot###

boxplot(EPI,DALY)

####qq plot###

qqplot(EPI,DALY)



### Regression ###

lmEPI<-lm(EPI~DALY+AIR\_H+WATER\_H)

summary(lmEPI) # AIR\_H is most significant

cEPI<-coef(lmENVH)

DALYNEW<-c(seq(5,95,5))

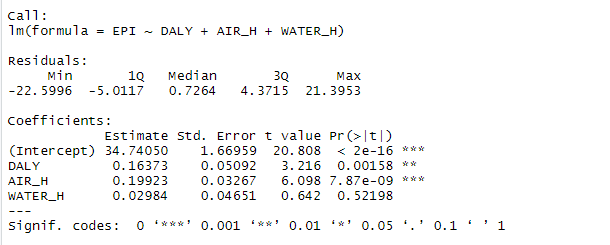
AIR\_HNEW<-c(seq(5,95,5))

WATER\_HNEW<-c(seq(5,95,5))

NEW<-data.frame(DALYNEW,AIR\_HNEW,WATER\_HNEW)

pEPI<-predict(lmEPI,NEW,interval= 'prediction')

cEPI<-predict(lmEPI,NEW,interval= 'confidence')



lab 2 part 2

# Multiple Regression

multiple\_regression <-read.csv(file.choose())

head(multiple\_regression)

attach(multiple\_regression)

mm<-lm(formula = ROLL~UNEM+HGRAD)

help(predict)

NEW <- data.frame(HGRAD = 90000,UNEM = .07)

pmm<-predict(mm,NEW,interval= 'prediction')

cmm<-predict(mm,NEW,interval= 'confidence')

pmm # Roll is about 76000



mm<-lm(formula = ROLL~UNEM+HGRAD+INC)

NEW <- data.frame(INC = 25000)

pmm<-predict(mm,NEW,interval= 'prediction')

cmm<-predict(mm,NEW,interval= 'confidence')

pmm # Roll is about 134000



# Classification

library(class)

abalone <-read.csv(file.choose())

head(abalone)

attach(abalone)

colnames(abalone) <-c("sex", "length", 'diameter', 'height', 'whole\_weight', 'shucked\_wieght', 'viscera\_wieght', 'shell\_weight', 'rings' )

summary(abalone$rings)

str(abalone)

aba <-abalone

aba$sex<-NULL

normalize <-function(x) {

return ((x -min(x)) / (max(x) -min(x)))

}

aba[1:7] <-as.data.frame(lapply(aba[1:7], normalize))

ind<-sample(2, nrow(aba), replace=TRUE, prob=c(0.7, 0.3))

KNNtrain<-aba[ind==1,]

KNNtest<-aba[ind==2,]

KNNpred<-knn(train = KNNtrain[1:7], test = KNNtest[1:7], cl = KNNtrain$rings, k = 55)

KNNpred

table(KNNpred)



# Clustering

dim(iris)

summary(iris)

iris\_data <- select(iris,-c(Species))

set.seed(300)

k.max<-15

wss<-sapply(1:k.max,function(k){kmeans(iris\_data,k,nstart= 20,iter.max = 1000)$tot.withinss})

plot(1:k.max,wss, type= "b", xlab= "Number of clusters(k)", ylab= "Within cluster sum of squares")

icluster<-kmeans(iris\_data,3,nstart = 20)

