**Homework Assignment 1**

**STA 141A A02**

**Submitted By:**

hwdata1<-readRDS("college\_scorecard\_2013.rds")

#1 How many observations

dim(hwdata1)

nrow(hwdata1)

sum(hwdata1$main)

#2

ncol(hwdata1)

#features = 51

length(is.factor(hwdata1))

class(data3)

var\_class<-sapply(hwdata1,class)

var\_class

table(var\_class)

#names of all integer variables

names(var\_class=="integer")

names(var\_class=="factor")

length(var\_class=="factor")

length(var\_class)

#names of all factor variables

names(var\_class[var\_class=="factor"])

#3

#no of NA in dataset

is.na(hwdata1)

#max. number of NA

which.max(is.na(hwdata1))

#analysing data with more NA

sort(colSums(is.na(hwdata1)))

#which feature has most missing values.

#4

props = prop.table(table(hwdata1$ownership))

degree\_dat<-hwdata1[c("highest\_degree","ownership")]

head(degree\_dat)

degree\_dat<-hwdata1[c("primary\_degree","highest\_degree","ownership")]

table(degree\_dat)

degree\_dat\_tab<-table(degree\_dat)

dimnames(degree\_dat\_tab)

prop.table(degree\_dat\_tab,margin=1)

prop.table(degree\_dat\_tab,margin=2)

mosaicplot(prop.table(degree\_dat\_tab,margin=1),color=TRUE, shade=TRUE)

mosaicplot(degree\_dat\_tab,color=TRUE, shade=TRUE,cex.axis =.45,las=2)

#5

summary(hwdata1$undergrad\_pop)

# what will hapen if ....

decile(hwdata1$undergrad\_pop)

quantile(hwdata1$undergrad\_pop[which(!is.na(hwdata1$undergrad\_pop))], probs = seq(0, 1, length=11))

boxplot(hwdata1$undergrad\_pop,quantile(hwdata1$undergrad\_pop[which(!is.na(hwdata1$undergrad\_pop))], probs = seq(0, 1, length=11)))

abline(h = min(hwdata1$undergrad\_pop), col = "Blue")

abline(h = max(hwdata1$undergrad\_pop), col = "Yellow")

abline(h = median(hwdata1$undergrad\_pop), col = "Green")

d <- density(hwdata1$undergrad\_pop[which(!is.na(hwdata1$undergrad\_pop))]) # returns the density data

plot(d)

#6

#analysis of tuition fee of 5 populous states

states=hwdata1[hwdata1$state %in% c("CA","TX","FL","IL","NY"),]

boxplot(states$tuition~states$state)

states1=droplevels(states)

boxplot(states1$tuition~states1$state,xlab="States",ylab="tuition fees",main="States Tution fees Distribution")

#7

c<- split(hwdata1,hwdata1$ownership)

names(c)

c<-hwdata1$ownership[which(hwdata1$ownership=="Public")]

plot(hwdata1$spend\_per\_student ~ hwdata1$avg\_10yr\_salary,c)

abline(a=0,b=1)

title("Public Distribution")

c1<-hwdata1$ownership[which(hwdata1$ownership=="For Profit")]

plot(hwdata1$spend\_per\_student ~ hwdata1$avg\_10yr\_salary,c1)

abline(a=0,b=1)

names(hw)

title("For Profit")

c2<-hwdata1$ownership[which(hwdata1$ownership=="Nonprofit")]

plot(hwdata1$spend\_per\_student ~ hwdata1$avg\_10yr\_salary,c2)

abline(a=0,b=1)

title("Nonprofit Distribution")

#8.

hwdata1$new<- hwdata1$avg\_10yr\_salary/hwdata1$cost

b<-data.frame(hwdata1$name,hwdata1$new)

w<-b[!(rowSums(is.na(b))),]

w$hwdata1.new

?data.frame

#hwdata1$rank<-rank(w$hwdata1.new)

ord <- order(w$hwdata1.new, decreasing=T)

tsu <- w[ord, ]

#9.

b<-data.frame(hwdata1[c("name","race\_white","race\_black","race\_asian","race\_native","race\_pacific","race\_other","race\_hispanic")])

#table(hwdata1$name,b)

w<-b[!(rowSums(is.na(b))),]

wmean<-(w$race\_white+w$race\_black+w$race\_asian+w$race\_hispanic+w$race\_other+w$race\_native+w$race\_pacific)/7

po<-data.frame(w$name,wmean)

ord <- order(po$wmean)

tsu <- po[ord, ]

head(tsu)

which.max(wmean)

po[29,]

summary(b)

#10

grep("University of California-Davis",hwdata1$name,fixed = TRUE)

summary(hwdata1[123,])

s<- hwdata1[123,]

q<- mean(hwdata1$tuition[which(!is.na(hwdata1$tuition))])

p<- mean(hwdata1$avg\_10yr\_salary[which(!is.na(hwdata1$avg\_10yr\_salary))])

scatter.smooth(q,s$tuition)

plot(q,s$tuition,xlab = "Other colleges mean tution",ylab= "UC-Davis tution",main ="tuition")

abline(h=13800,col="blue")

abline(v=17600,col="yellow")

plot(p,s$tuition,xlab = "Other colleges mean avg sal",ylab= "UC-Davis avg-sal",main ="avg-sal")

r<- mean(hwdata1$undergrad\_pop[which(!is.na(hwdata1$undergrad\_pop))])

plot(r,s$tuition,xlab = "Other colleges mean undergrad pop",ylab= "UC-Davis undergrad pop",main ="undergrad pop")