**Assignment #4 Code**

dat<-read.csv(file = "https://raw.githubusercontent.com/mattdemography/EDU\_7043/master/Data/GSS\_2016\_AA.csv", stringsAsFactors=F)

names(dat)

str(dat)

#subset and removing 'non-answers'

bob<-dat[c(1:2867), c(38,39,16,32)]

bob$AGE<- ifelse(bob$AGE==99| bob$AGE==98, NA, bob$AGE)

bob$NATCHLD<- ifelse(bob$NATCHLD==9| bob$NATCHLD==8, NA, bob$NATCHLD)

#subsets

wh<-subset(bob, bob$RACE==1)

b<-subset(bob, bob$RACE==2)

o<-subset(bob, bob$RACE==3)

nw<-subset(bob, bob$RACE==2 | bob$RACE==3)

m<-subset(bob, bob$SEX==1)

f<-subset(bob, bob$SEX==2)

#mean,median and range

summary(bob$NATCHLD)

summary(wh$NATCHLD)

summary(b$NATCHLD)

summary(o$NATCHLD)

summary(nw$NATCHLD)

summary(m$NATCHLD)

summary(f$NATCHLD)

#standard deviation

sd(bob$NATCHLD,na.rm=T)

sd(wh$NATCHLD,na.rm=T)

sd(b$NATCHLD,na.rm=T)

sd(o$NATCHLD,na.rm=T)

sd(nw$NATCHLD,na.rm=T)

sd(m$NATCHLD,na.rm=T)

sd(f$NATCHLD,na.rm=T)

#variance

var(bob$NATCHLD,na.rm = T)

var(wh$NATCHLD,na.rm=T)

var(b$NATCHLD,na.rm=T)

var(o$NATCHLD,na.rm=T)

var(nw$NATCHLD,na.rm=T)

var(m$NATCHLD,na.rm=T)

var(f$NATCHLD,na.rm=T)

#Comparing male and female views of national spending on children

t.test(m$NATCHLD, f$NATCHLD)

wilcox.test(m$NATCHLD, f$NATCHLD)

#comparing different racial groups views of national spending on children

racenatchild<-aov(bob$NATCHLD ~ as.factor(bob$RACE))

summary(racenatchild)

#Results of ANOVA

TukeyHSD(racenatchild)

#normality

hist(bob$NATCHLD)

boxplot(wh$NATCHLD, b$NATCHLD, o$NATCHLD)

boxplot(m$NATCHLD,f$NATCHLD)