library(readxl)

library(tidyverse)

library(dunn.test)

#set working directory

getwd()

setwd()

#attach data

brookhaven\_census\_data <- read\_excel("Brookhaven\_ACS2017\_data.xlsx")

brookhaven\_census\_data

names(brookhaven\_census\_data)

#add new variable:Children school age percentage

brookhaven\_census\_data$Children\_School\_Age\_Pct <- brookhaven\_census\_data$Age\_Under\_18\_Pct - brookhaven\_census\_data$Age\_Under\_5\_Pct

brookhaven\_census\_data

#reorder Districts: North, Middle, South

brookhaven\_census\_data$Districts <- factor(brookhaven\_census\_data$Districts, levels=c("North", "Middle", "South"))

#Visualization

#boxplot

#400 x 250

#word: width 3.5

#household size

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Mean\_HH\_Size)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, shape="x", size=3) +

labs(y="Mean", title="Household Size") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#highschool graduate

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Highschool\_Grad\_Pct)) +

geom\_boxplot() +

geom\_jitter(width = 0.2, alpha = 0.7, shape="x", size=3) +

labs(title="High School Graduate", y="Percent")+

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#bachelor's and higher graduate

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Bachelor\_Grad\_Up\_Pct)) +

geom\_boxplot() +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Bachelor's Degree or Higher", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#Spanish spoken at home

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Spanish\_Lang\_Home\_Pct)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Spanish Spoken at Home", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#Median income

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Median\_HH\_Income\_USD)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Household Income (Median)", y="USD") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5)) +

scale\_y\_continuous(labels = scales::comma)

#house value

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Median\_House\_Value\_USD)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="House Value (Median)", y="USD") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5)) +

scale\_y\_continuous(labels = scales::comma)

#Median age

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Median\_Age\_Years)) +

geom\_boxplot(outlier.shape=NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Age (Median)", y="Years") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#Age 65 and over

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Age\_65\_Over\_Pct)) +

geom\_boxplot(outlier.shape=NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Age 65 and Over", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#race white

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Race\_White\_Pct)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Race: White", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#hispanic or latino

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=HispLatino\_Pct)) +

geom\_boxplot(outlier.shape = NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Hispanic or Latino\n(of any race)", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#White not hispanic or latino

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=White\_NoHispanicLatino\_Pct)) +

geom\_boxplot(outlier.shape=NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="White\n(not Hispanic or Latino)", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#age 5-18

brookhaven\_census\_data %>%

ggplot(aes(x=Districts, y=Children\_School\_Age\_Pct)) +

geom\_boxplot(outlier.shape=NA) +

geom\_jitter(width = 0.2, alpha = 0.7, color = "black", shape="x", size=3) +

labs(title="Age 5-18", y="Percent") +

theme\_classic() +

theme(plot.title = element\_text(hjust = 0.5))

#

#Anova

anova\_HH\_size <- aov(Mean\_HH\_Size ~ Districts, data=brookhaven\_census\_data)

summary(anova\_HH\_size)

anova\_Highschool\_Grad\_Pct <- aov(Highschool\_Grad\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Highschool\_Grad\_Pct)

TukeyHSD(anova\_Highschool\_Grad\_Pct) #match

anova\_Bachelor\_Grad\_Up\_Pct <- aov(Bachelor\_Grad\_Up\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Bachelor\_Grad\_Up\_Pct)

TukeyHSD(anova\_Bachelor\_Grad\_Up\_Pct)

anova\_Spanish\_Lang\_Home\_Pct <- aov(Spanish\_Lang\_Home\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Spanish\_Lang\_Home\_Pct)

TukeyHSD(anova\_Spanish\_Lang\_Home\_Pct)

anova\_Median\_HH\_Income\_USD <- aov(Median\_HH\_Income\_USD ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Median\_HH\_Income\_USD)

TukeyHSD(anova\_Median\_HH\_Income\_USD) #match

anova\_Median\_House\_Value\_USD <- aov(Median\_House\_Value\_USD ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Median\_House\_Value\_USD)

TukeyHSD(anova\_Median\_House\_Value\_USD) #match

anova\_Median\_Age\_Years <- aov(Median\_Age\_Years ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Median\_Age\_Years)

anova\_Age\_65\_Over\_Pct <- aov(Age\_65\_Over\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Age\_65\_Over\_Pct)

anova\_Race\_White\_Pct <- aov(Race\_White\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Race\_White\_Pct)

anova\_HispLatino\_Pct <- aov(HispLatino\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_HispLatino\_Pct)

TukeyHSD(anova\_HispLatino\_Pct)

anova\_White\_NoHispanicLatino\_Pct <- aov(White\_NoHispanicLatino\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_White\_NoHispanicLatino\_Pct)

anova\_Children\_School\_Age\_Pct <- aov(Children\_School\_Age\_Pct ~ Districts, data=brookhaven\_census\_data)

summary(anova\_Children\_School\_Age\_Pct )

#Kruskal-Wallis & Dunn Test

#match

dunn.test(brookhaven\_census\_data$Highschool\_Grad\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni") #match

dunn.test(brookhaven\_census\_data$Median\_HH\_Income\_USD, g=brookhaven\_census\_data$Districts, method="bonferroni") #match

dunn.test(brookhaven\_census\_data$Median\_House\_Value\_USD, g=brookhaven\_census\_data$Districts, method="bonferroni") #match

dunn.test(brookhaven\_census\_data$HispLatino\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni") #match

dunn.test(brookhaven\_census\_data$White\_NoHispanicLatino\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni") #match

#significant but not match

dunn.test(brookhaven\_census\_data$Bachelor\_Grad\_Up\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni")

dunn.test(brookhaven\_census\_data$Spanish\_Lang\_Home\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni")

#not significant

dunn.test(brookhaven\_census\_data$Mean\_HH\_Size, g=brookhaven\_census\_data$Districts, method="bonferroni")

dunn.test(brookhaven\_census\_data$Median\_Age\_Years, g=brookhaven\_census\_data$Districts, method="bonferroni")

dunn.test(brookhaven\_census\_data$Age\_65\_Over\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni")

dunn.test(brookhaven\_census\_data$Race\_White\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni")

dunn.test(brookhaven\_census\_data$Children\_School\_Age\_Pct, g=brookhaven\_census\_data$Districts, method="bonferroni")