df = read.table(file = "../aneurysm\_data.csv", sep = ',', header = TRUE)

# Q1.1

plot(df$Age, df$BloodPressure, main="Relationship Between Blood Pressure And Age", xlab = "Age", ylab = "Blood Pressure")

lines(lowess(df$BloodPressure~df$Age, f = 2/3), col = 'blue', lwd = 2)

legend("topright", legend=c("Blood Pressure", "Smooth Line"),

col=c("black", "blue"), pch = c(1,NA), lty= c(NA,1), cex=1)

# Q1.2

boxplot(df$Aneurisms\_q1 ~ floor(df$Age), main="Box Plot of Aneurisms\_q1", xlab = "Age", ylab = "Aneurisms\_q1")

boxplot(df$Aneurisms\_q2 ~ floor(df$Age), main="Box Plot of The Aneurisms\_q2", xlab = "Age", ylab = "Aneurisms\_q2")

boxplot(df$Aneurisms\_q1, df$Aneurisms\_q2, main="Box Plot For The Aneurisms", names=c("Aneurisms\_q1","Aneurisms\_q2"))

?floor

# Q1.3

hist(df$Aneurisms\_q4, col = rgb(0,1,1,0.25), xlim = c(50,400), ylim = c(0,25), xlab = "Aneurisms",

main = "Histogram of Aneurisms\_q3 and Aneurisms\_q4")

hist(df$Aneurisms\_q3, add = T, col =rgb(1,0,0,0.25))

legend("topright", legend=c("Aneurisms\_q4", "Aneurisms\_q3"),

fill=c(rgb(0,1,1,0.25), rgb(1,0,0,0.25)))

# Q2.1

library("ggplot2")

p <- ggplot(df, aes(y=BloodPressure, x=Age))

p + geom\_point() + geom\_smooth(method = 'loess') + ggtitle("Relationship Between Blood Pressure And Age") + xlab("Age") + ylab("Blood Pressure")

# Q2.2

quart\_1 <- subset(df, df$Aneurisms\_q1 <= quantile(df$Aneurisms\_q1,0.25))

quart\_2 <- subset(df, df$Aneurisms\_q1 > quantile(df$Aneurisms\_q1,0.25) & df$Aneurisms\_q1 <= quantile(df$Aneurisms\_q1, 0.5))

quart\_3 <- subset(df, df$Aneurisms\_q1 > quantile(df$Aneurisms\_q1,0.5) & df$Aneurisms\_q1 <= quantile(df$Aneurisms\_q1, 0.75))

quart\_4 <- subset(df, df$Aneurisms\_q1 > quantile(df$Aneurisms\_q1,0.75))

quart\_1$Quartile = "1st quartile"

quart\_2$Quartile = "2nd quartile"

quart\_3$Quartile = "3rd quartile"

quart\_4$Quartile = "4th quartile"

quart = rbind(quart\_1, quart\_2)

quart = rbind(quart, quart\_3)

quart = rbind(quart, quart\_4)

ggplot(data = quart, aes(x=Age))+geom\_density(aes(fill =Quartile, alpha=Quartile)) +

scale\_fill\_manual(values = c("#4E84C4", "#D16103", "#52854C", "#FFDB6D")) +

scale\_alpha\_manual(values = c(0.5, 0.5,0.5,0.5)) + xlim(c(7, 25)) +

ggtitle(" Density of age for quartiles of Aneurysms\_q1") + xlab("Age")

# Q2.3

library(reshape2)

library(knitr)

newdf = melt(df,id.vars=c("ID","Gender","Group","BloodPressure","Age", "Aneurisms\_q1", "Aneurisms\_q2"),

variable.name="q",value.name="Aneurisms")

ggplot(newdf,aes(x= Aneurisms) )+

geom\_histogram(data = subset(newdf, q == "Aneurisms\_q3"), aes(fill = q),bins = 14, alpha = 0.4) +

geom\_histogram(data = subset(newdf, q == "Aneurisms\_q4"), aes(fill = q),bins = 14, alpha = 0.4) +

scale\_fill\_manual(name="Aneurisms", values=c("orangered2","dodgerblue2"),labels=c("q3","q4")) +

ggtitle("Histogram of Aneurisms\_q3 and Aneurisms\_q4")

ggplot(df,aes(df$Aneurisms\_q3)) + geom\_histogram(bins =15, fill = 'red', alpha = 0.4) +

geom\_histogram(data =df, aes(df$Aneurisms\_q4), bins =15, fill = 'blue', alpha = 0.4)

# Q3.1

library("plotly")

Sys.setenv("plotly\_username"="WeiH")

Sys.setenv("plotly\_api\_key"="lGL9DmeMxywlqPPfsFvx")

newdf2 = df

newdf2$Age = floor(df$Age)

p <- plot\_ly(newdf2, x = ~Age, y=~Aneurisms\_q1, split = ~Age,

type = "violin", box = list(visible = T),meanline = list(visible = T)) %>%

layout(title = "Violin Chart of Aneurisms\_q1")

p

api\_create(p, filename = "HW4Q3")

# https://plot.ly/~WeiH/1/