# Stat 528 HW1

Dongyang Wang

2023-01-15

## Question 1

```r
rm(list=ls())
df = read.csv("CHSdataEx1.csv")
```

## Question 2

### Question 2.1

```r
library(corrplot)
library(xtable)

df_exer = na.omit(df[,c("exint0", "exint3", "block0", "block3", "kcal0", "kcal3")])
corrplot.mixed(cor(df_exer),
               lower = "number",
               upper = "circle",
               tl.col = "black")

boxplot(df$exint0, df$exint3, main = "Boxplot for Exercise Intensity",
        xlab = "Baseline and 3 Years After",
        ylab = "Intensity")
tab1 <- summary(df$exint3- df$exint0)

boxplot(df$block0, df$block3, main = "Boxplot for Blocks Walked",
        xlab = "Baseline and 3 Years After",
        ylab = "Blocks")
tab2 <- summary(df$block3- df$block0)

boxplot(df$kcal0, df$kcal3, main = "Boxplot for kCal Expended",
        xlab = "Baseline and 3 Years After",
        ylab = "Kilocalories")
tab3 <- summary(df$kcal3- df$kcal0)

table1 <- rbind(tab1, tab2, tab3)
rownames(table1) <- c("Exercise Intensity", "Blocks Walked", "kCal Expended")
table1
```

### Question 2.2

```r
baseline = c("season", "gender", "age", "weight", "weight50", "grade", "arth", "sbp", "pkyrs", "diab",
exercise = c("exint0", "exint3", "block0", "block3", "kcal0", "kcal3")
df_clean = na.omit(df[,c(baseline, exercise)])
```

```
corrplot.mixed(cor(df_clean),
               lower = "number",
               upper = "circle",
               tl.col = "black")
```

## Question 2.3

```
library(jtools)
df_reg = df[,c("mortality", baseline)]
lm1 = lm(mortality ~ . , data = df_reg)
summary(lm1)
summ(lm1)
```

## Question 2.4

```
df_reg2 = df[,c("mortality", baseline, exercise)]
lm2 = lm(mortality ~ . , data = df_reg2)
summary(lm2)
summ(lm2)
```