install.packages("ggplot2")

# prepare ggplot library

library(ggplot2)

# set working directory and import US Data

setwd("//myhome.itap.purdue.edu/puhome/pu.data/Desktop/mayer15/STAT 350/Lab 02")

USData <- read.table("USData\_Spring.txt", header=TRUE, sep="\t")

# clean US Data

USData\_clean <- USData[complete.cases(USData),]

# use clean dataset

attach(USData\_clean)

### PART B: TestScore ###

# print five-number summary

FNS <- fivenum(TestScore)

# print 1.5 IQR limits and find outliers

IQR <- FNS[4] - FNS[2]

IF\_U <- FNS[4] + 1.5\*IQR # upper limit

IF\_L <- FNS[2] - 1.5\*IQR # lower limit

IF\_U

IF\_L

Outlier\_Index <- which(TestScore < IF\_L | TestScore > IF\_U)

Outliers <- TestScore[Outlier\_Index]

Outliers

# make modified boxplot

windows()

ggplot(USData\_clean, aes(x = "", y = TimeToStart)) +

stat\_boxplot(geom = "errorbar") +

geom\_boxplot() +

ggtitle("Test Scores throughout the US") +

stat\_summary(fun.y = mean, col = "black", geom = "point", size = 3)