**כריית נתונים ב- R**

***מיני תרגיל מספר 3***

**שאלות:**

בתרגיל זה נתשמש במסד Cross Auctions.csv.

1. קרא את הקובץ לתוך מסד ב- R
2. הצג את ההתפלגות של StartPrice בעזרת היסטוגרמה. שנה את מספר השנתות כדי לאפשר ויזואליזציה נהירה.
3. הצג את ההתפלגות של StartPrice בעזרת boxplot. במידת הצורך שנה את סקאלת הצירים לסקאלה לוגריתמית.
4. הצג גרף פיזור של ClosePrice כפונקציה של StartPrice.
   1. הוסף מידע לגבי SellerAboutMePage.
   2. שנה את סקאלת הצירים (שניהם!) לסקאלה לוגריתמית.

**פתרון:**

setwd("C:/Users/user/Desktop/R")

ebay.df <- read.csv("Cross Auctions.csv", stringsAsFactors = TRUE)

## StartPrice

hist(ebay.df$StartPrice,

main = "Histogram of Start Price",

xlab = "Start Price",

br = 50)

boxplot(ebay.df$StartPrice, ylab = "StartPrice", log = 'y')

## Close Price ~ Start Price

plot(ClosePrice ~ StartPrice, data = ebay.df)

# color by aboutme page

plot(ClosePrice ~ StartPrice, data = ebay.df,

col = ifelse(ebay.df$SellerAboutMePage=="FALSE", "blue", "red"),

pch = ifelse(ebay.df$SellerAboutMePage=="FALSE", 1, 2))

legend("bottomright",

legend = c("No AboutMe Page", "AboutMe Page"),

col = c("blue", "red"),

pch = c(1, 2))

# color by aboutme page

plot(ClosePrice ~ StartPrice, data = ebay.df,

col = ifelse(ebay.df$SellerAboutMePage=="FALSE", "blue", "red"),

pch = ifelse(ebay.df$SellerAboutMePage=="FALSE", 1, 2),

log = 'xy')

legend("bottomright",

legend = c("No AboutMe Page", "AboutMe Page"),

col = c("blue", "red"),

pch = c(1, 2))