Discussion 2 - Ishita Dutta

Problem 1

I. Enter the following code into R to load the Lynx Dataset:

```
library(datasets)
X = as.numeric(lynx)
#?lynx
```

(a) Provide a brief description of the dataset. *Solution:* Annual numbers of lynx trappings for 1821–1934 in Canada. Taken from Brockwell & Davis (1991), this appears to be the series considered by Campbell & Walker (1977).

(b) How many values are in the data? *Solution:* There are 114 values in the lynx dataset.

Problem 2

II. Perform the following:

(a) Find the average of X. *Solution:* The average is:1538.0175439.
(b) Find the minimum and maximum of X. *Solution:* The minumum is:39. The maximum is:6991.
(c) Subset the data to get all values that are less than the average. Call this Y1.
(d) Subset the data to get all values that are more than the average. Call this Y2.

```
Y1 = X[X < mean(X)]
Y2 = X[X > mean(X)]
```

(e) Find the 5th, 23rd, 27th, and 88th percentiles of Y1. *Solution:* The 5th percentile is: 63.05 The 8th percentile is: 76.64 The 23rd percentile is: 211.44 The 27th percentile is: 229 The 88th percentile is: 987.96

(f) Find the average of the deviations of Y2 from its mean. *Solution:* The standard deviation(average of deviations from mean) is: 1341.3051301
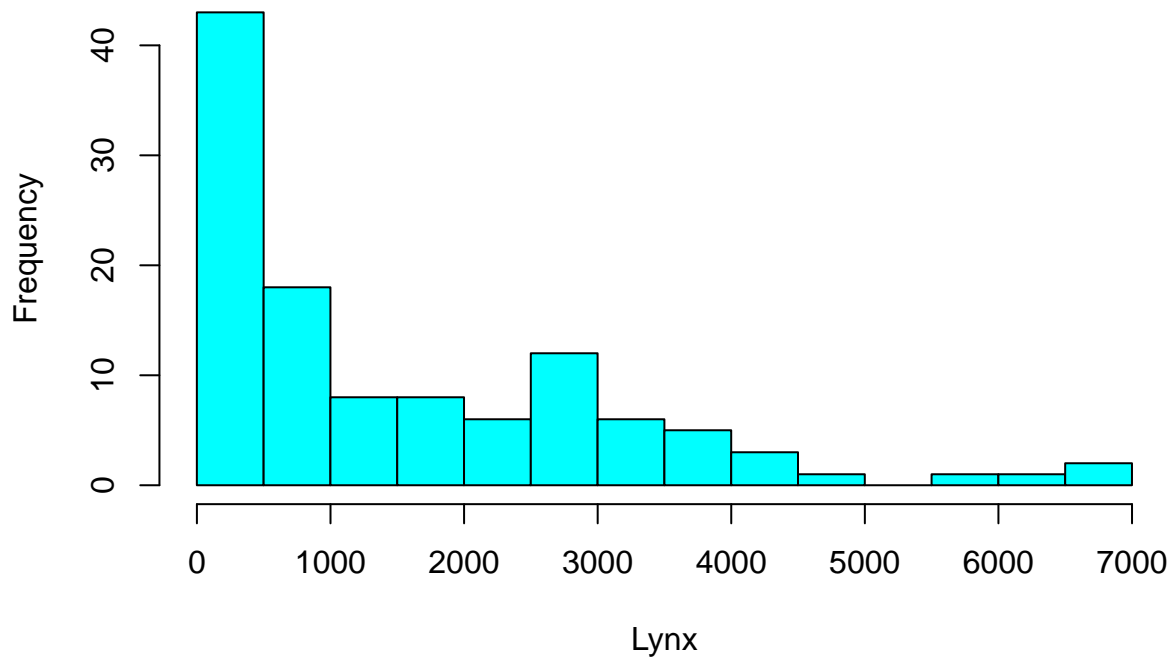
Problem 3

III. Review histograms and boxplots to do the following:

(a) Create a histogram for X making sure to label the plot and appropriate axes.

```
library(MASS)
truehist(X, nbins = 10, xlab = "Lynx", ylab = "Frequency", prob = FALSE)
title("Histogram for Lynx")
```

Histogram for Lynx

(b) What is the purpose of a histogram? Describe the histogram from part (a). *Solution:* Illustrate the dataset, showing what the shape of the data is. In the case of the histogram from part(a) is skewed right.

(c) Create a horizontal boxplot for X making sure to label the plot and appropriate axes.

If you give boxplot a vector of TRUE's and FALSE's, it treats them as a grouping variable. Then you can use the command
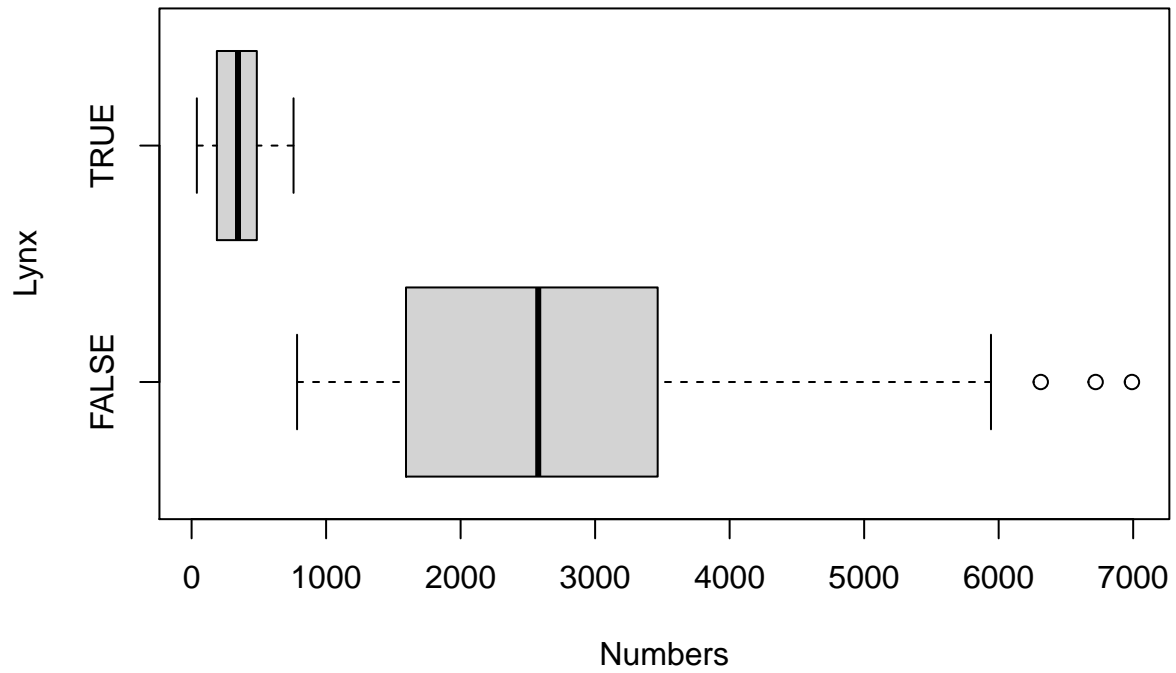
```
boxplot(X ~ Group)
```

```r
boxplot(X, horizontal = TRUE, xlab = "numbers", ylab = "Lynx")
title("Boxplot of Lynx", sub = "True: bigger than mean, False: smaller than mean")
```

Boxplot of Lynx

numbers
True: bigger than mean, False: smaller than mean
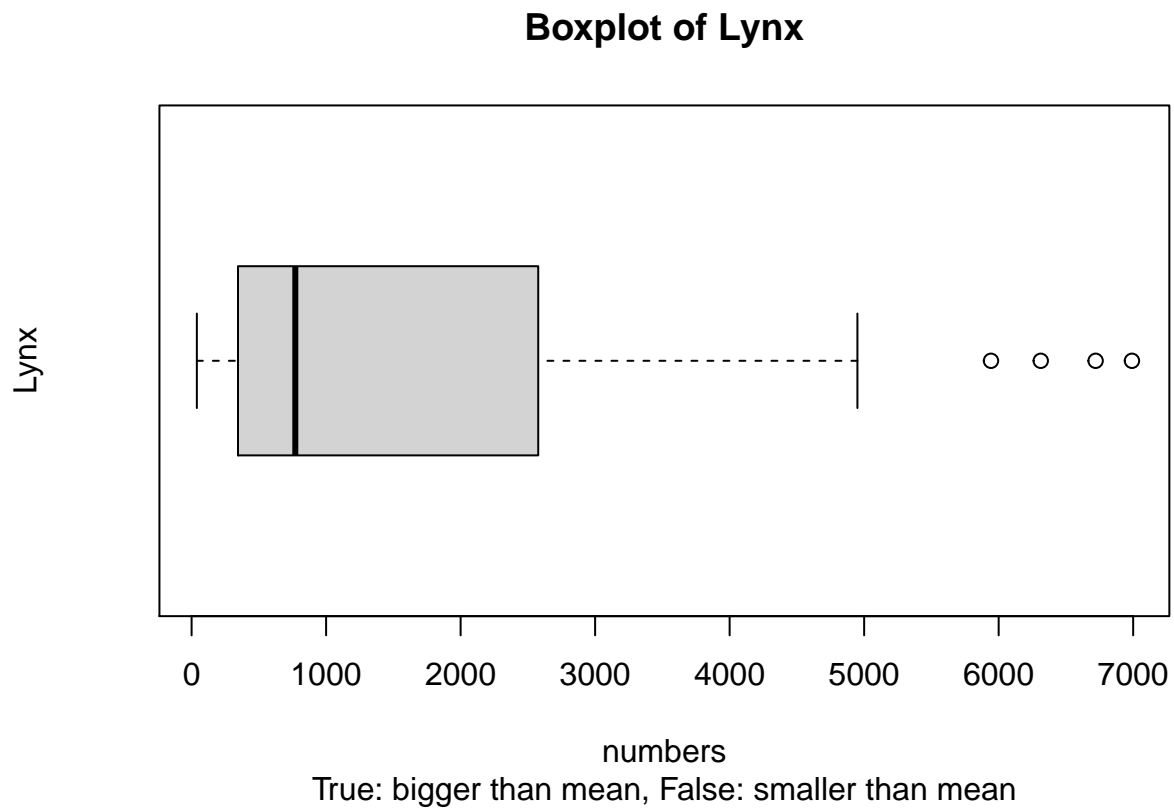
(d) Create aGroupvector that has TRUE if the corresponding value of X is less than the median, and FALSE otherwise, and plot the two boxplots.

```r
boxplot(X ~ X < median(X), horizontal = TRUE, xlab = "Numbers", ylab = "Lynx")
title("Boxplot of Lynx", sub = "True: bigger than median, False: smaller than median")
```

Boxplot of Lynx

Numbers
True: bigger than median, False: smaller than median