Bryan L. Mack – Week 4 Assignment.

Attached are six datasets, a single column of numbers.  Import these into RStudio.

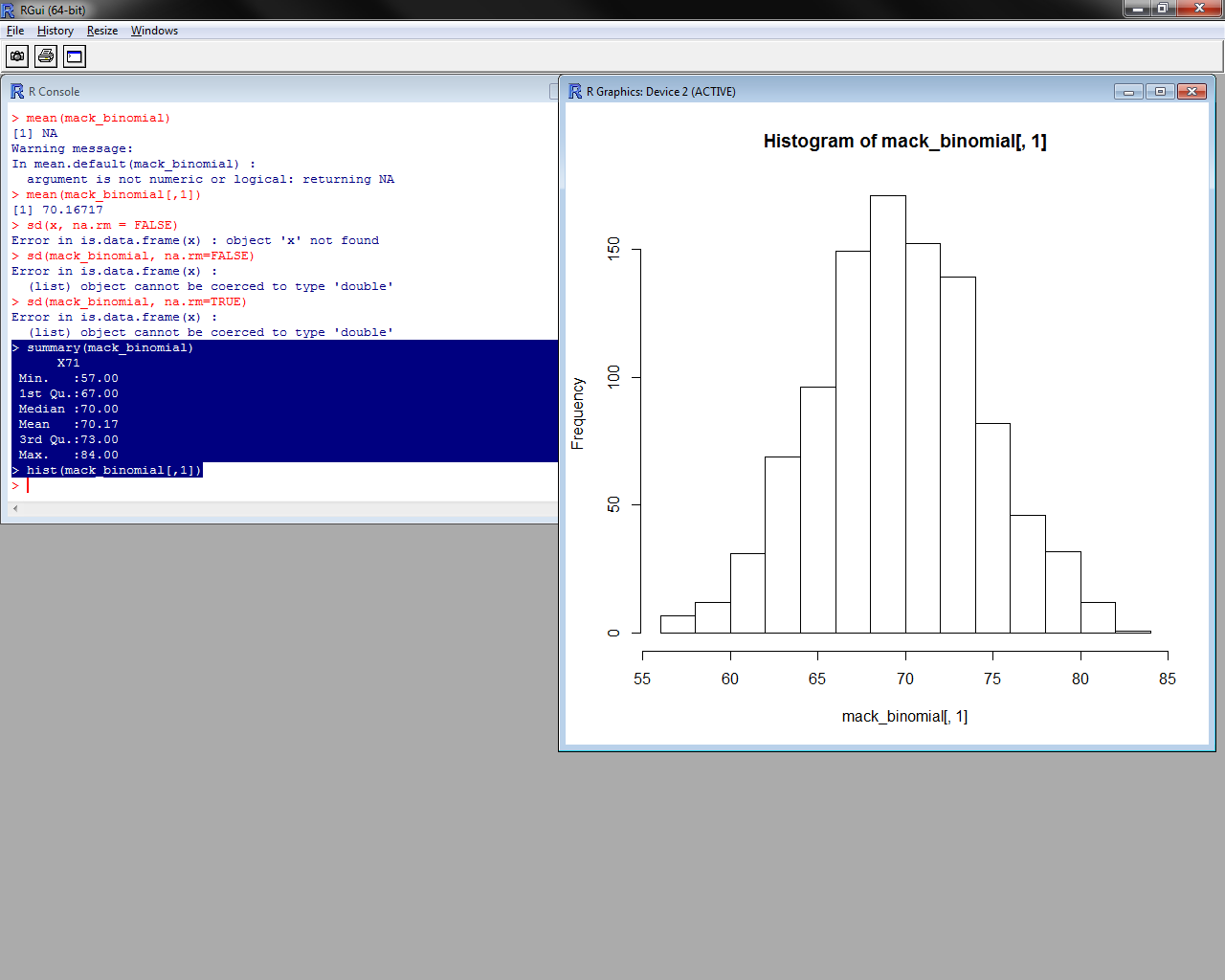
> setwd("C:/Users/bmack")  
> getwd()  
[1] "C:/Users/bmack"  
> mack\_binomial = read.csv("Binomial.csv")  
> mack\_BN1 = read.csv("BN1.csv")  
> mack\_BN2 = read.csv("BN2.csv")  
> mack\_N1 = read.csv("N1.csv")  
> mack\_N2 = read.csv("N2.csv")  
> mackln = read.csv("ln.csv")

Compute a set of descriptive statistics for each; including mean, standard deviation, minimum, maximum, a histogram plot, and any other descriptive statistic you might find meaningful.

**Binomial.csv**

> summary(mack\_binomial)  
 X71   
 Min. :57.00   
 1st Qu.:67.00   
 Median :70.00   
 Mean :70.17   
 3rd Qu.:73.00   
 Max. :84.00   
> mack\_binomial\_matrix <- as.matrix(mack\_binomial)  
> sd(mack\_binomial\_matrix[,1])  
[1] 4.691599

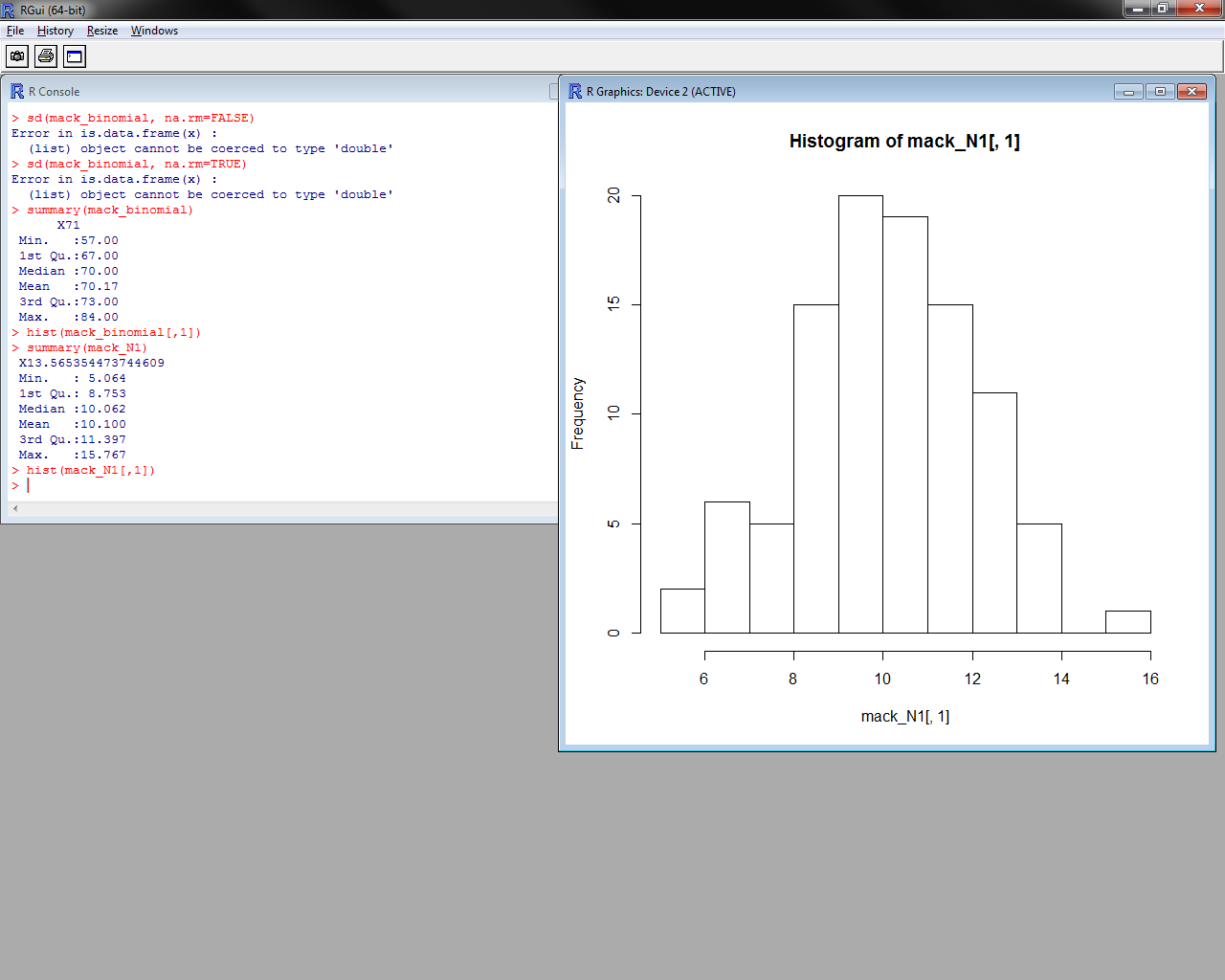
> hist(mack\_binomial[,1])



**N1:**

> summary(mack\_N1)  
 X13.565354473744609  
 Min. : 5.064   
 1st Qu.: 8.753   
 Median :10.062   
 Mean :10.100   
 3rd Qu.:11.397   
 Max. :15.767

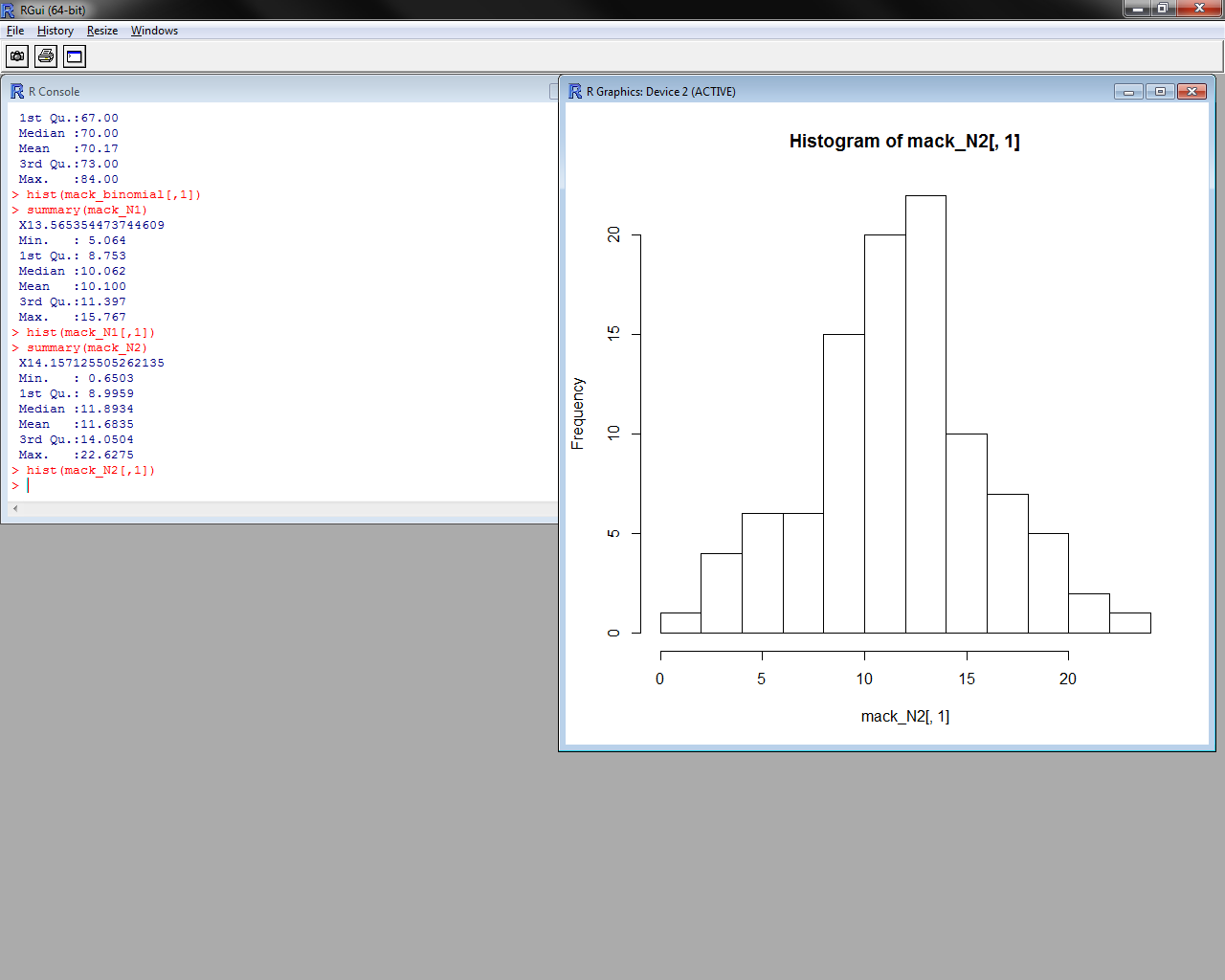
> mack\_n1\_matrix <- as.matrix(mack\_N1)  
> sd(mack\_N1[,1])  
[1] 1.975384



**N2:**

> summary(mack\_N2)  
 X14.157125505262135  
Min. : 0.6503   
 1st Qu.: 8.9959   
 Median :11.8934   
 Mean :11.6835   
 3rd Qu.:14.0504   
 Max. :22.6275

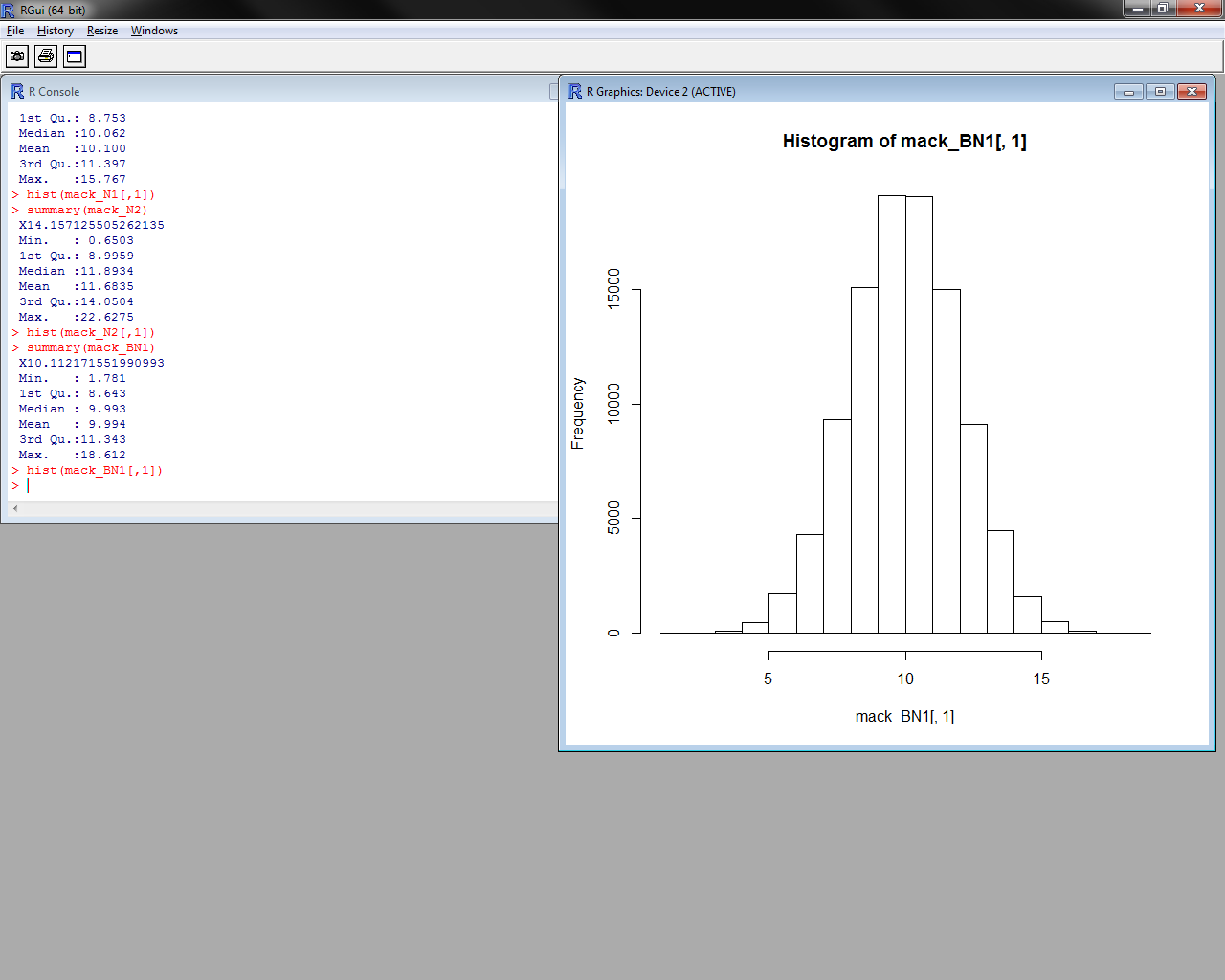
> mack\_n2\_matrix <-as.matrix(mack\_N2)  
> sd(mack\_n2\_matrix[,1])  
[1] 4.305943



**BN1:**

> summary(mack\_BN1)  
 X10.112171551990993  
 Min. : 1.781   
 1st Qu.: 8.643   
 Median : 9.993   
 Mean : 9.994   
 3rd Qu.:11.343   
 Max. :18.612

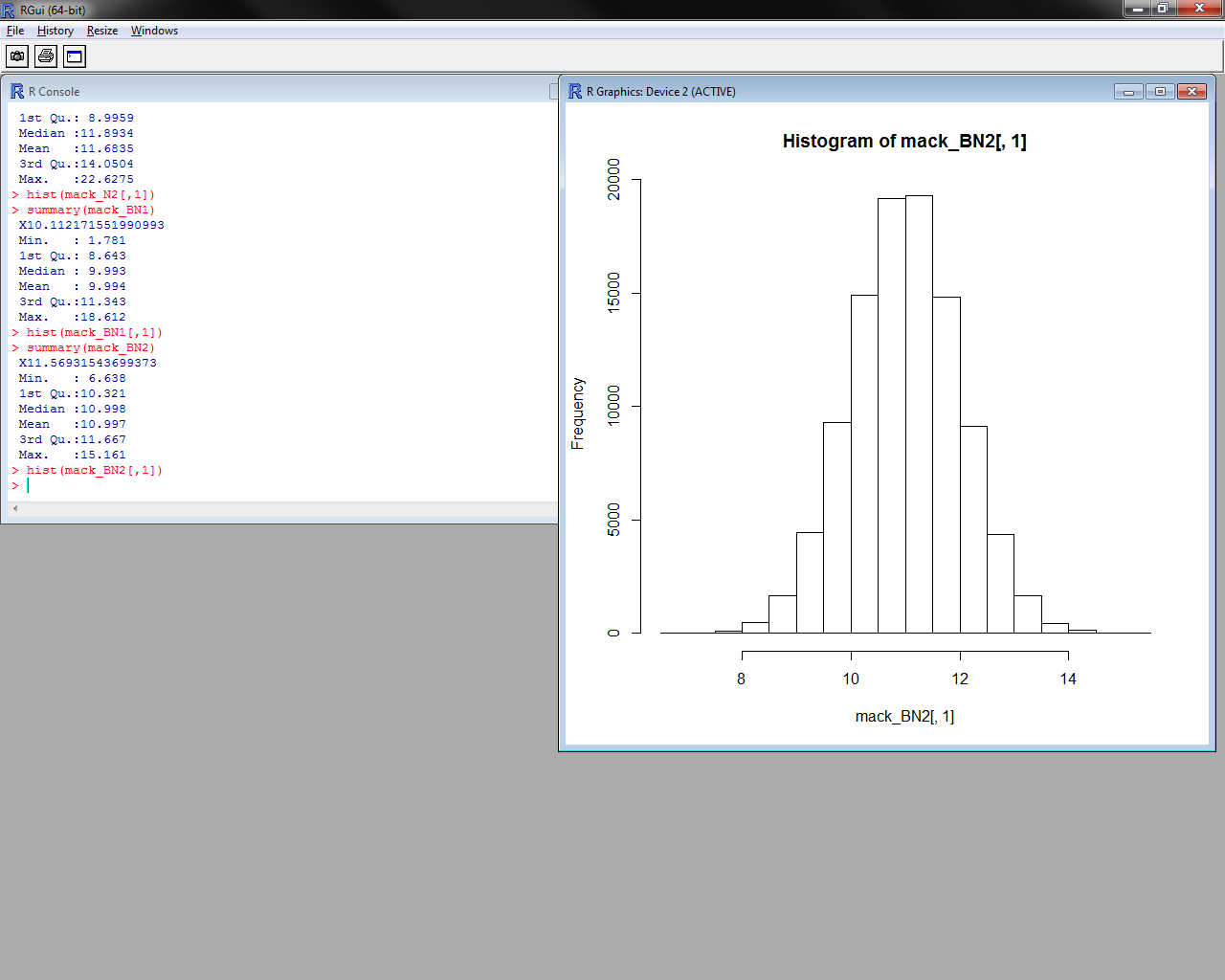
> mack\_bn1\_matrix <-as.matrix(mack\_BN1)  
> sd(mack\_bn1\_matrix[,1])  
[1] 2.000367



**BN2:**

> summary(mack\_BN2)  
 X11.56931543699373  
 Min. : 6.638   
 1st Qu.:10.321   
 Median :10.998   
 Mean :10.997   
 3rd Qu.:11.667   
 Max. :15.161

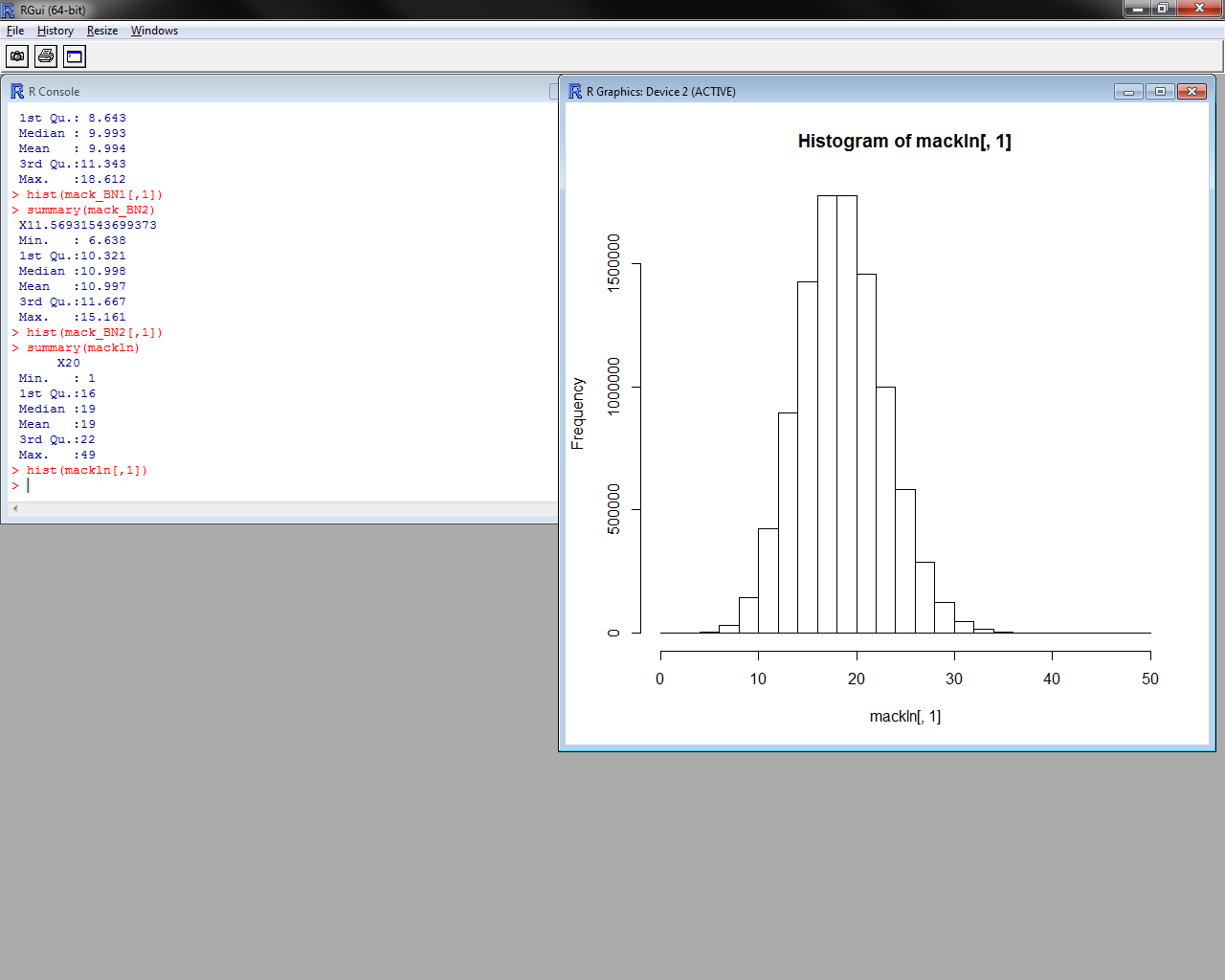
> mack\_bn2\_matrix <-as.matrix(mack\_BN2)  
> sd(mack\_bn2\_matrix[,1])  
[1] 0.9994748



**Ln:**

> summary(mackln)  
 X20   
 Min. : 1   
 1st Qu.:16   
 Median :19   
 Mean :19   
 3rd Qu.:22   
 Max. :49

> mack\_ln\_matrix <- as.matrix(mackln)  
> sd(mack\_ln\_matrix[,1])  
[1] 4.357854



Consider the two pairs of data sets N1 and N2 along with BN1 and BN2. What, if anything, can be said about the differences between N1 and N2?  Similarily, what, if anything, can be said about the differences between BN1 and BN2?

**N1 vs. N2:**

These two have the same sample sizes of 100, yet N2 has a much higher standard deviation (double). N2 contains more extreme outliers on either ends of the 1st & 3rd quartiles, and thus, a higher standard deviation. So while the averages are similar, the range or “responses” is wider.

**BN1 vs. BN2**

The situation is similar here, the number of samples is the same at 100k, yet BN1 has nearly double the standard deviation of BN2, yet again, the two samples have similar mean values. This indicates a wider range of responses with approximately the same average.