

# HomeWork 4

Chris Humphreys & Adam Chisholm  
CS455 - Data Mining  
University of Tennessee at Martin

02/25/19

## 1.

In the below screenshots, you can see our initial setup and usage of R. See figure 1 for initial setup. See figure 2 for the histogram of the Sepal Lengths.

## 2.

Figure 3 is a screenshot of our initial setup of Weka. In this screenshot we have created the histogram for the sepal length, but have not applied any filters to it. You will notice that the histograms in figure 2 and figure 3 are different although they come from the same data set. This is due to the fact that in RStudio (figure 2) the histogram starts at 4 and ends at 8, and is composed of 8 bins, each 1/2 cm wide, while in Weka (figure 3) the bins start at the minimum sepal length of 4.3 and go up to the maximum sepal length of 7.9. Additionally, the Weka histogram is composed of only 7 bins that are a bit over .5 cm each.

## 3.

Figure 4 shows the same sepal data again, but this time after being filtered as directed in the assignment in question 3. This time the histogram still looks different than the histogram in figure 2. There are a number of reasons for the two histograms looking different. First is the fact that although we have 8 bins in both, the first bin in figure 2 contains all the sepal lengths from 0 to 4.5, while the first bin in figure 4 contains all the sepal lengths from 0 to 4.75. From these different starting points of 4.5, and 4.75, figure 2 creates a new bin every .5 cm, while figure 4 creates a new bin every .45 cm. This difference does not seem like much initially, but as can be seen in the figures below, it can result in a significant difference in the final histogram.

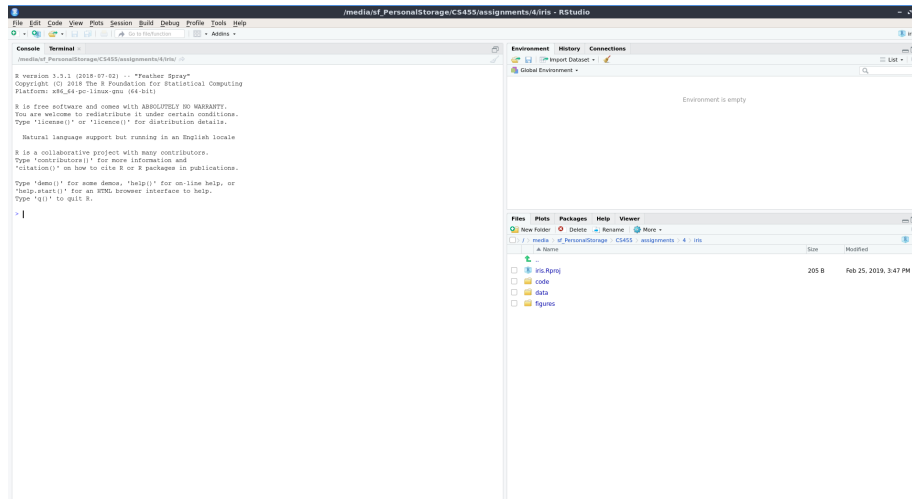


Figure 1: Initial setup of R

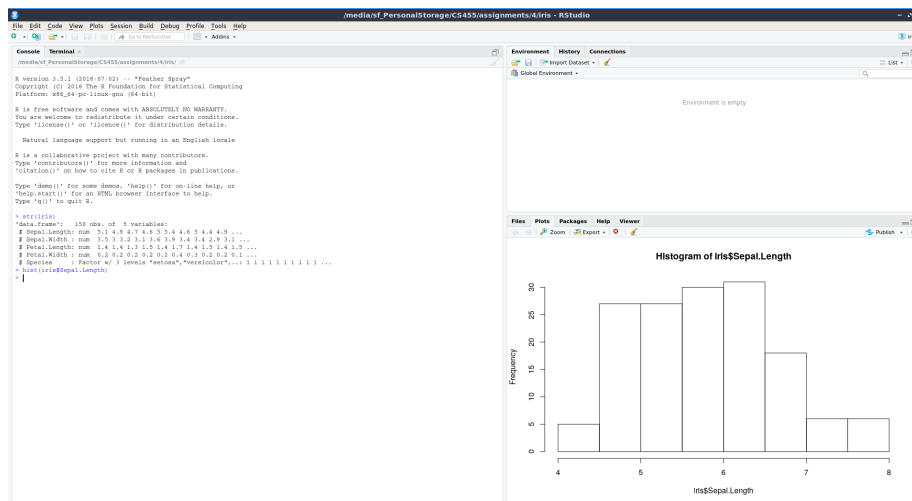


Figure 2: Iris histogram in RStudio

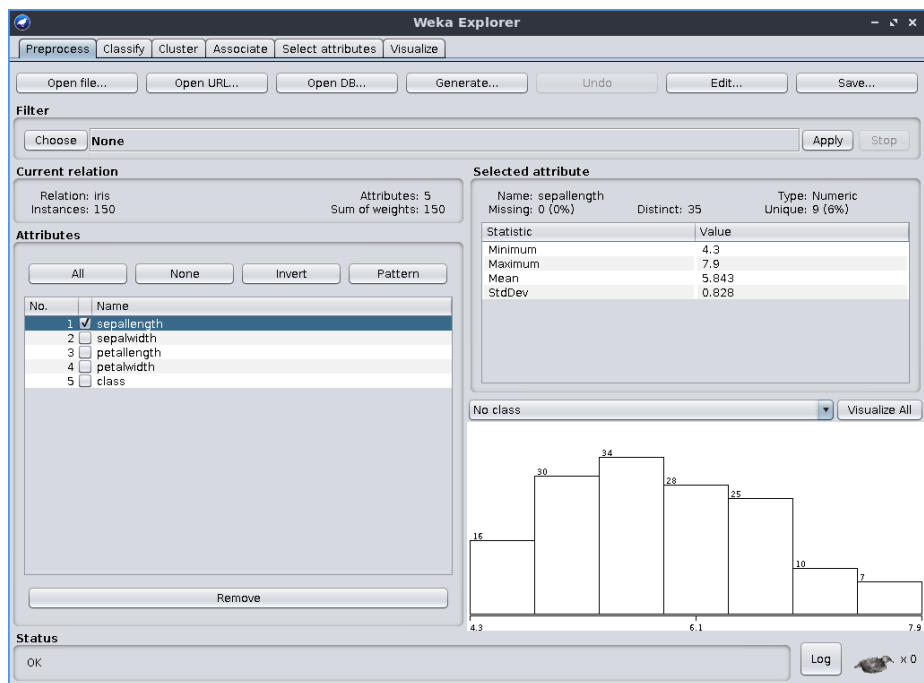


Figure 3: Iris histogram in Weka, no filters

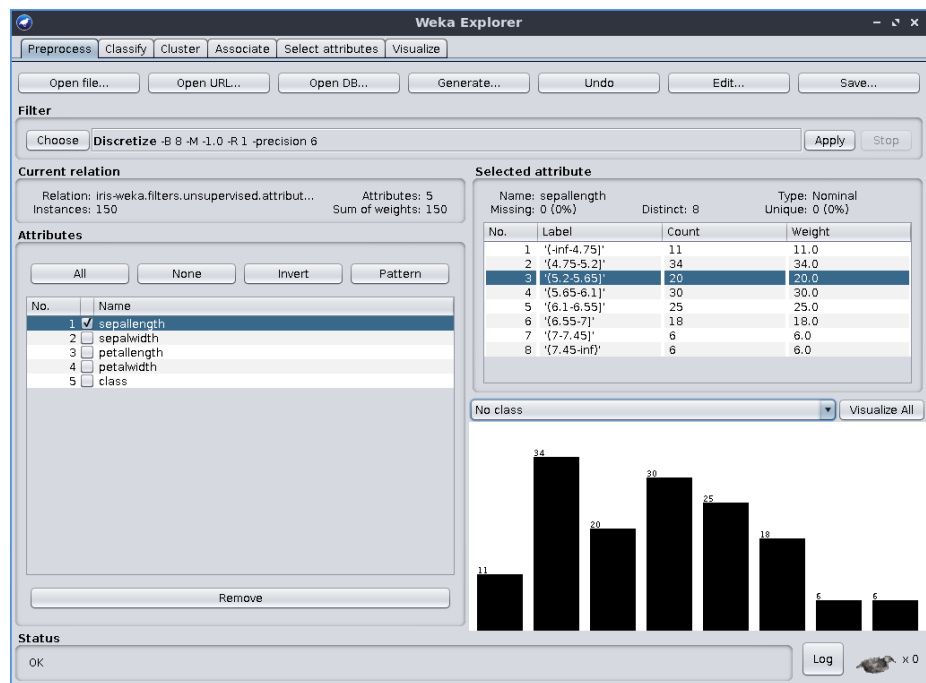


Figure 4: Iris histogram in Weka with filters