CS614 – Prof. Berardi
Homework #3 (Lecture 5)

This homework assignment will ask you to perform some preliminary visualization tasks in Base R. The built-in `possum` data set from the `DAAG` package will be used. A description of the data can be found here.

Perform the following tasks:

1. Produce two histograms of the `age` variable, one using the default bin boundaries and another using the boundaries $\{0, 1.5, 3, \ldots, 7.5, 9\}$. Can you explain why you see such dramatic differences in the two figures. (The function `table(cut(<vec>))` might be useful). Also produce a kernel density estimate (KDE) - what are the pros/cons of the KDE versus the histogram?

2. A researcher produces a histogram of the `earconch` variable and notices that it is bimodal. Based on this, she hypothesizes that ear conch lengths differ by `sex`. First reproduce the original histogram. Then create a side-by-side boxplot of ear conch length. Compare and contrast the two boxplots. How well do the figures support the hypothesis that ear conch length varies by sex?

3. Produce a 10x10 grid of scatter plots comparing all combinations of the numeric variables in the `possums` data frame, excluding `case` and `site`. Pick the one combination that you feel has the strongest linear relationship and create a single scatter plot. Add to this plot a single red triangle at the mean x and mean y location. You'll likely need to look into the `pch` option.