STAT 217: Project 2
Due Friday, January 30 in class

You must work in a group of 2-3. There will be a point deduction if you do not work in a group.

From *The Statistical Sleuth*:

> In an 1898 lecture at Woods Hole, Massachusetts, Hermon Bumpus reminded his audience that the process of natural selection for evolutionary change was an unproved theory: "Even if the the theory of natural selection were as firmly established as Newton's theory of attraction of gravity, scientific method would still require frequent examination of its claims." As evidence in support of natural selection, he presented measurements on house sparrows [found near his laboratory] after an uncommonly sever winter storm. Some of these birds had survived and some had perished. Bumpus asked whether those that perished did so because they lacked physical characteristics enabling them to withstand the intensity of that particular instance of selective elimination.

We have measurements of the humerus (arm bone) lengths for 24 adult, male sparrows. To answer Bumpus's question of whether the two groups had different average humerus lengths, we will compare the humerus lengths for birds that survived to those that perished. Conduct a permutation test to address this question; make sure to include a confidence interval to estimate the size of the effect. Summarize your results in a written report following the **Project Writing Guidelines** posted on D2L. The following R code will get you started.

Include one additional section in your report and label it *Additional Section*. In this section, conduct a two-sample t-test and write a conclusion with your results from the t-test. Compare these results to the permutation test results.

Before running the R code, load the bumpus data into R just as we did in Project 1. Make sure it is called `bumpus`. You will also need to make sure you have installed the `mosaic` and `beanplot` R packages. To do this, run `install.packages("mosaic")` and `install.packages("beanplot")`. Once you've done this, the package is on installed on your computer and you do not need to do it again. Note that the lines below preceded with # are comments and are there to help you understand what the code is doing. You can copy and paste the following code into your R script and run it one line at a time.

R Code

```
# make sure the data are loaded into R in an object called bumpus

require(mosaic)
favstats(Humerus ~ Status, data = bumpus)
Tobs <- diffmean(Humerus ~ Status, data = bumpus)
```

```r
# plots
require(beanplot)
par(mfrow = c(1, 2))
beanplot(Humerus ~ Status, data = bumpus, method = "jitter",
         log = "", col = 7, ylab = "Humerus Length")
boxplot(Humerus ~ Status, data = bumpus, ylab = "Difference")
par(mfrow = c(1, 1))

# Permutation test
B <- 1000
Tstar <- matrix(NA, nrow = B)
for(b in 1:B) {
  Tstar[b] <- diffmean(Humerus ~ shuffle(Status), data = bumpus)
}
hist(Tstar, xlab = "", main = "Permuted Distribution of Differences")
abline(v = Tobs, col = "red", lwd = 2)
# pvalue
pdata(abs(Tobs), abs(Tstar), lower.tail = F)

# Use for CI and t-test
t.test(Humerus ~ Status, data = bumpus, var.equal = T)
```