Problem Set #6: Multivariate Stats

(1) (10 pts) Mark Steele and colleagues studied communities of coral reef fishes in the Bahamas. Fishes were surveyed along 5-8 transects at each site. The file BahamasFish.*csv*, includes densities of only the four most abundant species. Convert data to z-scores and use PCA to evaluate whether density variation of the four species can be summarized by new, derived variables. How much variation in density is explained by the first two principal components?

PC1 explains 38% and PC2 explains 27.8%. Together they explain **65.8%.**

```
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4
Standard deviation     1.2153385 1.0395696 0.9331931 0.6761008
Proportion of Variance 0.3801226 0.2781226 0.2241157 0.1176392
Cumulative Proportion  0.3801226 0.6582451 0.8823608 1.0000000
```

(2) (10 pts) Make a biplot of PC1 vs PC2. Add vectors showing the relationship between fish species and these two components. How is density of each species related to PC1 and PC2?

The squirrelfish and redband parrotfish are most related to PC1, and the samples in this group have **lower densities**. For PC2 the squirrelfish density has a higher density while the redband parrotfish has lower density.
The doctorfish and stripped parrotfish are most related to PC2 but the **doctorfish have lower density** while the **stripped parrotfish have a higher density.** These two fish also show high density in relation to PC1.
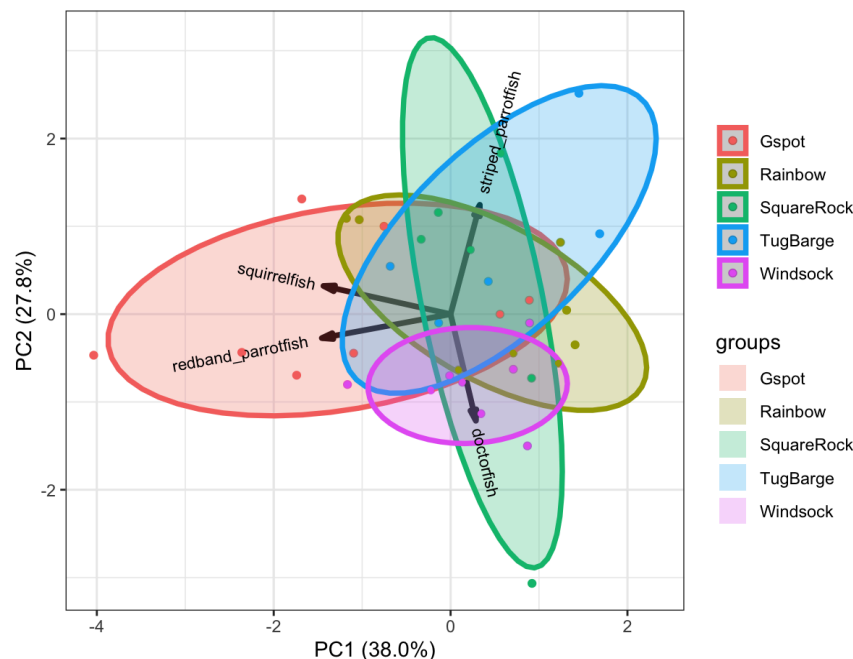
(3) (10 pts) The Steele lab wanted to determine whether fish communities differed among 5 study sites. Use ANOVA on PC1 scores from the Bahamas fish community to determine whether fish communities differ among sites. The independent transects at each site will serve as replicates.

Anova on PC1 scores shows a **significant difference** in fish community among the 5 study sites **(F= 3.8; p<0.05)**.

(4) (10 pts) Jeanne Robertson is examining lizards (*Liolaemus darwinii*) living on different types of soils. She wants to know if lizard color is adapted to the color of the soil on which they're found. She collects several lizards from areas with black, white, and red soils. Then she measures the color of each lizard using three metrics. The first is brightness, which distinguishes dark from light colors (e.g., black vs. white). The second is hue, which is best at distinguishing red from black. The third is saturation, which is best at distinguishing red from white. Using all three of these metrics, she can quantify the color of each lizard. Instead of quantifying hue, saturation, and brightness as continuous variables, she creates eight bins for each variable and measures the percentage of pixels that fall into each bin. So pure red lizards would have pixels all in bin 1 of hue. Pure black lizards would have all pixels in bin 8 for brightness and bin 8 for hue. This dataset has 24 dependent variables: 8 bins for each of hue, saturation, and brightness. The data are in the file lizardcolor.csv. Use discriminant function analysis to determine how well lizards can be assigned to their original collection location, based on their color. What was the percent success for each soil type?

Black is 62.5%, Red is 55.6% and White is 95.7%. Overall is 80%

```
> diag(prop.table(assignmatrix, 1))
    Black       Red     White
0.6250000 0.5555556 0.9565217
> sum(diag(prop.table(assignmatrix))) # Gives total correct assignment
[1] 0.8
>
```

(5) (10 pts) One cold February day in 1898, an "uncommonly severe" storm passed over New England. After the storm, the zoologist Hermon Bumpus collected 136 house sparrows that had been brought down by the storm in the vicinity of his laboratory at Brown University in. More than half of the birds recovered, but the rest died from exposure. Bumpus took this as an opportunity to study natural selection in action, and measured a number of skeletal features on all the birds, as well as recording whether they survived the storm, their sex, and (in the males) whether they were adults or yearlings. Bumpus's data are in the file bumpus.csv.
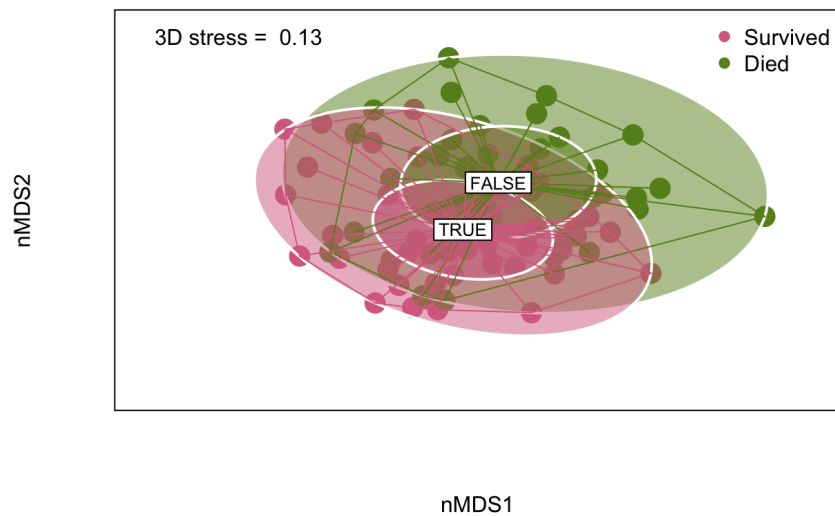Bumpus measured (in order on the file: total length, alar extension (wingspan), weight, head length, humerus length, femur length, tibiotarsus length, skull width, and keel to sternum length. We already know that house sparrows are sexually dimorphic in overall size, so split the data by sex and do separate analyses on males and females. Use **perMANOVA** to determine whether the birds that **survived** the storm had **different morphologies than birds that died**. **What trait(s)** distinguish these two groups of birds the most? Do the **results depend on sex**? Include **a NMDS plot** for each sex.

perManova shows **<u>a significant effect of morphology on storm survival in male birds (p<0.05)</u>** and the permutation dispersion test shows very similar average distance to the median. Further analysis by adding coefficients shows that **<u>the most important trait for male survival is Length with the highest absolute value.</u>**

For females perManova **does not show a significant effect of morphological traits on storm survival** and the dispersion test shows **different average distances to the median (True = 0.008299; False = 0.010693).** Further analysis still shows **length as the most important trait.**
We can state that Length is the most important trait but the result does not truly depend on sex.

Male nMDS Plot

Female nMDS plot