

# Biodiversity for the National Parks

Adrian's DA Capstone Project



# Understanding the data: species\_info.csv contents

- The species\_info.csv file consists of 5,824 rows pertaining to 5,541 unique species
- Each row contains a category (Mammal, Fish, etc.), scientific name, common name, and conservation status
- While the most common conservation status listed is 'Species of Concern,' far and away the most common status is none at all, which is to say, unprotected (see Fig. 1, Fig. 2 below)
- Further, vascular plants make up by far the largest proportion (77%), and Reptiles and Amphibians the smallest (both at 1.4%)

# Endangered status, difference among species

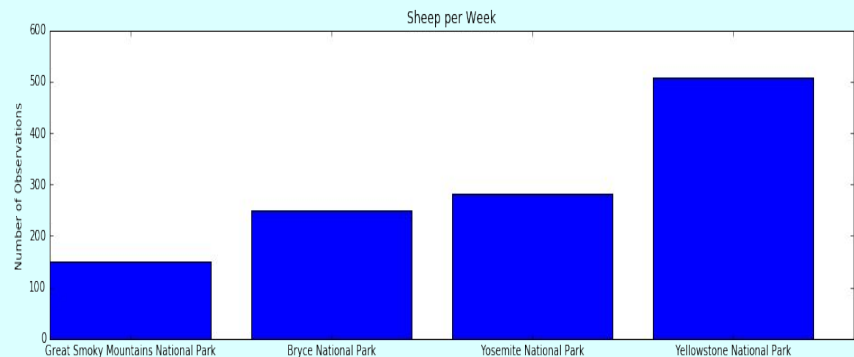
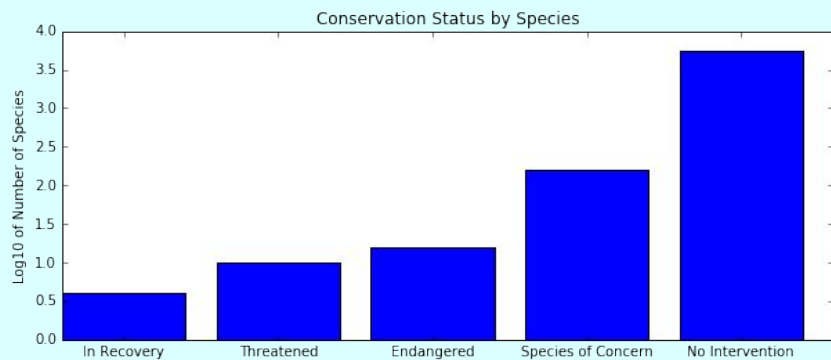
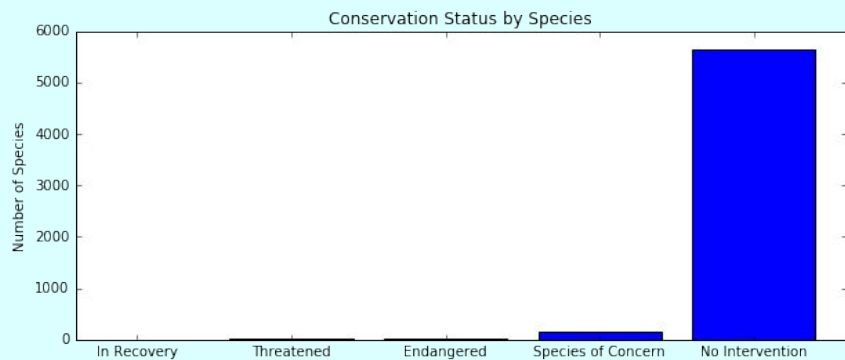
- One question thus naturally arises - are certain categories protected less than others?
- To answer this question, we must choose a test appropriate to our data - here, we have two categorical datasets, namely species A and B, both with a protected and unprotected total
- Further, we use the standard .05 p value; we want to be 95% certain of the falsehood of the Null Hypothesis
- So, employing a chi-squared test, we find that there's no significant difference in protection levels between Mammals and Birds
- But Mammals, it turns out, are significantly more likely to be protected than Lizards, meaning there is a statistically significant difference between some species re: protection

# Recommendation to Conservationists

- One broad conclusion from our data is that certain species are more likely to be protected than others
- Thus, perhaps you could focus more resources on the less-protected categories
- To that end, you could use the protected percentage of each category as a guideline, i.e. look at the species whose protected percentages are below the average value of 8.4% (Reptiles, and Vascular and Nonvascular Plants)

# Sample Size determination for sheep disease study

- After attempting to remedy the sheep foot-and-mouth disease issue, we'd like to know if the treatment had a significant (and presumably positive) impact on the population
- To determine this, we used a sample size calculator to determine the appropriate sample size, using the original disease rate (15%), a predetermined target of 10%, or a 33% drop from the original 15%, and a significance of 90%
- The calculator determined that we'd need to observe 890 sheep in order to be certain enough that our program has worked or not
- Given the observation rates at each park, that would amount to a ~4 week period at Bryce National Park, or a ~2 week period at Yellowstone National Park (see Fig. 3)



Top Left: Figure 1, Conservation Status by Species

Top Right: Figure 2, Conservation Status by Species with base-10 log of total as y-axis

Bottom Left: Figure 3, Number of Sheep observed per week, at each park