BIODIVERSITY CAPSTONE PROJECT - INVESTIGATING PROTECTED SPECIES

Lauren Bakker, 6/12/19

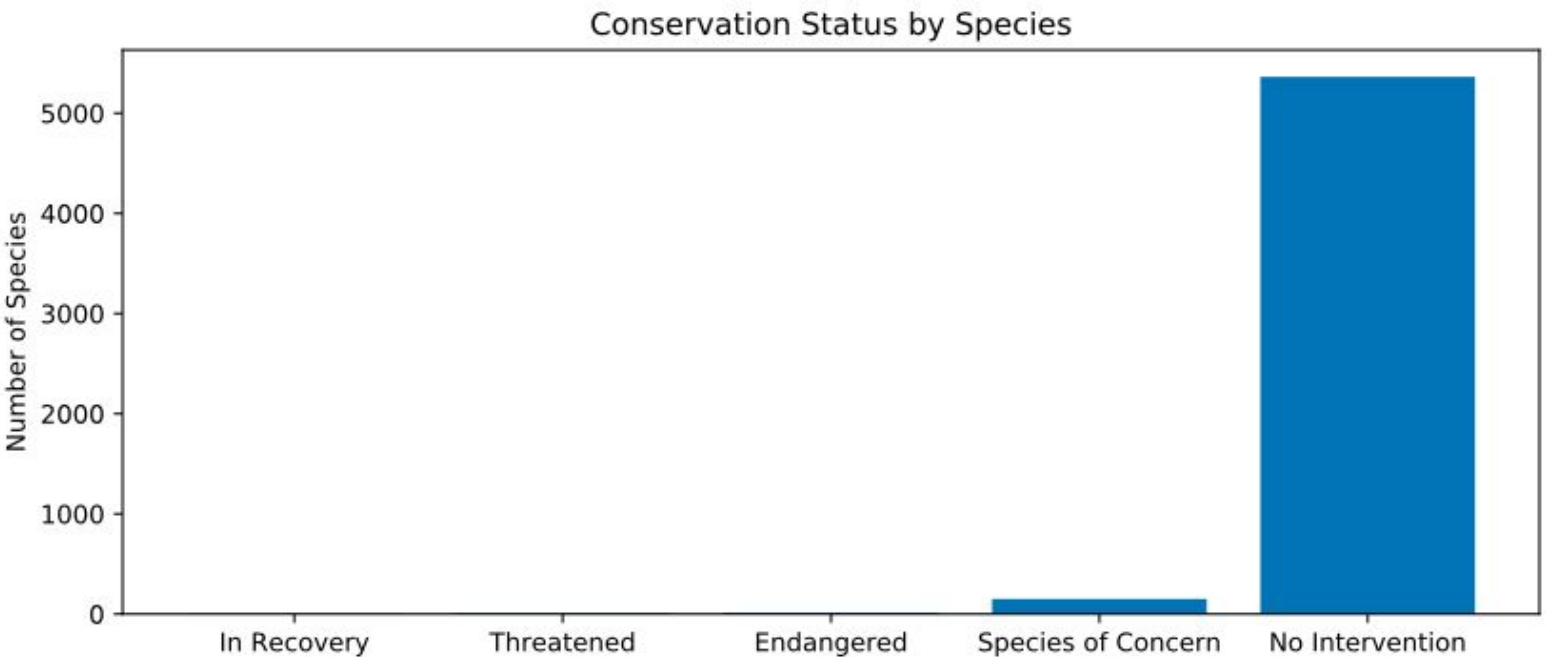
Initial Observations of the “species_info.csv” Data

- The DataFrame includes the following columns: id, category, scientific_name, common_name, and conservation_status.
- The DataFrame includes 5,541 unique species.
- The species category values are: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, and Nonvascular Plant.
- The conservation status values are: Species of Concern, Endangered, Threatened, In Recovery.
- Several rows have a conservation status of “None.”

Analysis

Number of unique species in each conservation status, including “No Intervention” (see visual on next slide):

- Endangered: 15
- In Recovery: 4
- No Intervention: 5363
- Species of Concern: 151
- Threatened: 10


Conservation Status by Species
Number of Species
0
1000
2000
3000
4000
5000
In Recovery
Threatened
Endangered
Species of Concern
No Intervention

Significance Testing

The following analysis was performed to determine if certain species were significantly more likely have have protected statuses. To start, the data was grouped by category and whether the status was protected or “No Intervention.” This generated the percentages of protected species in each category:

Category	**Percent Protected**
Amphibian	8.75
Bird	15.16
Fish	8.66
Mammal	17.76
Nonvascular Plant	1.50
Reptile	6.33
Vascular Plant	1.03

Several chi-squared tests were run to determine if some species were significantly more likely than others to have a protected statuses, the null hypothesis being that the differences in percent protected (previous slide) were due to chance. These significance tests revealed the following:

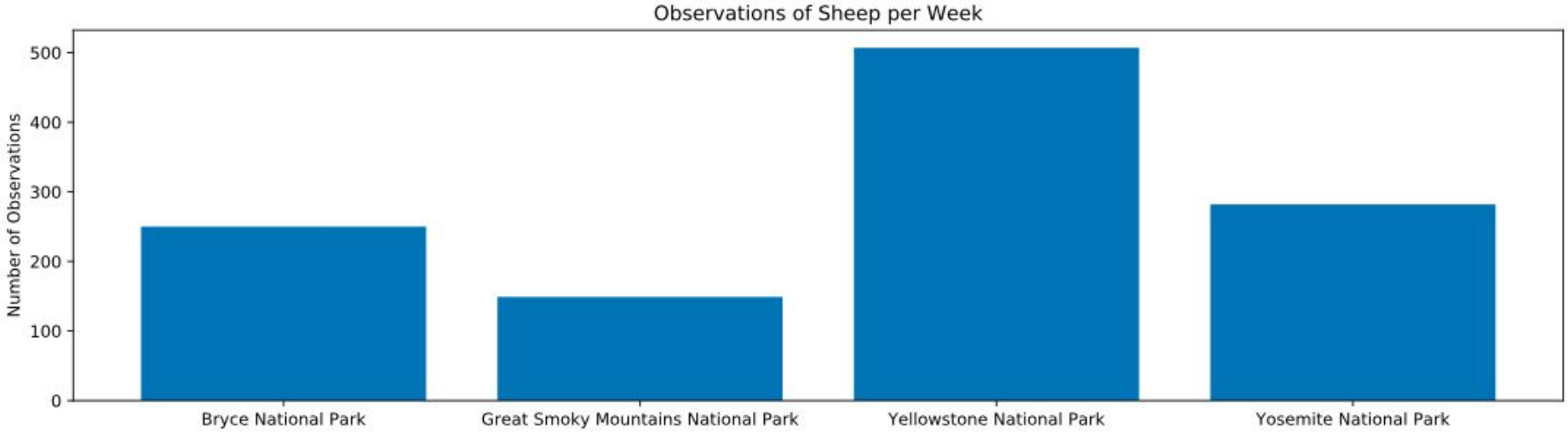
- There is no significant difference in the rate of protected status between birds and mammals. (The p value for this chi-squared test was 0.69.)
- Mammals are, however, significantly more likely than reptiles to have a protected status. (That p-value was 0.38.)
- Mammals are significantly more likely to have protected statuses than the following species categories: Reptile, Amphibian, Nonvascular Plant, Vascular Plant
- Vascular plants are significantly *less* likely to have protected statuses than all other species categories except nonvascular plants.

Based on these tests, it is recommended that conservation resources and funding for mammals be no more than those for birds. However, more effort should be put towards mammal conservation than towards that for reptiles, amphibians, nonvascular plants, and vascular plants. Vascular plants require fewer conservation efforts than all other species except nonvascular plants.

Foot and Mouth Disease Study: Sample Size Determination

To generate per-week sheep observations in various national parks, a lambda function was first used to create a new column in the “species” DataFrame, indicating whether a given species was a sheep or not. Then, all rows for sheep were selected and cleaned up by selecting just those rows that where the category was mammal. The results were saved as “sheep_species.”

Next, “sheep_species” was merged with “observations” to get a DataFrame of sheep observations. The groupby function was then used to get the sum of all sheep sightings for each park. The results are shown on the next slide.


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park

To determine the needed sample size for the foot and mouth disease study, the following information was plugged into the sample size calculator:

- Baseline: 15% [last year's recorded rate of foot and mouth disease in Bryce National Park]
- Statistical significance: 90% [default level of significance]
- Minimum detectable effect: 33.33% [the percent change from the baseline needed to observe minimum desired disease reductions]

Sample size: 870

Based on the needed sample size above and the sheep-per-week observation rate, for scientists to observe a foot and mouth disease reduction of five percentage points or more in Yellowstone National Park, the study would need to last just under two weeks (870 / 507 = 1.7). In Bryce National Park, the same study would need to last three and a half weeks (870 / 250 = 3.5).