Preserving Biodiversity in the National Parks

By Soren Rehn, Codecademy Data Intensive, Winter 2018

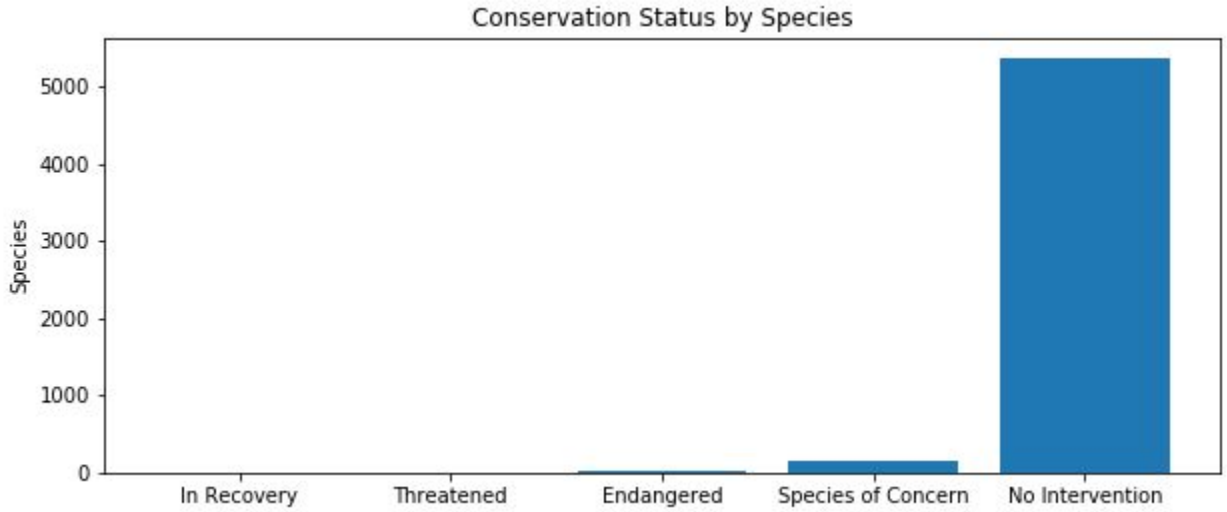
The Data

Understanding the nature of the species that are threatened is critical to designing strategies to protect them.

To that end, we're exploring a set of species. Let's look at the data:

Species Data Summary

- Data gathered contained 5541 unique species, providing their common and scientific names, species category, and conservation status.
- The categories were widely recognizable divisions - Mammal, Amphibian, Bird, Fish, Reptile, Vascular and Nonvascular Plants.
- In the data, Vascular plants dominated the sample at 76.9%, but that is actually in line with general species data according to UN data from 2004 [1]



Looking into the conservation status of the species, we find that the vast majority are not affected or needing special status.

Phew. Most species are ok!

What is under threat?

With that in mind, we wanted to understand if some species categories were more likely to require special status.

Methodology

- First, we calculated the percentages of each category that is under a protected status.
- Then, we compared the categories to see if there were significant differences in the likelihood that a species from a given category would be protected.
- To achieve this, we ran a chi-squared test to compare the categorical status data, comparing Mammals to Birds and Reptiles

category	not_protected	protected	percent_protected
Amphibian	72	7	8.860759
Bird	413	75	15.368852
Fish	115	11	8.730159
Mammal	146	30	17.045455
Nonvascular Plant	328	5	1.501502
Reptile	73	5	6.410256
Vascular Plant	4216	46	1.079305

Summary of Protection Status by Category

0.68

The Mammals/Birds contingency test produced a very high p-value – so neither of these categories are significantly more likely to be protected.

0.038

The Mammals/Reptile contingency test had very different results, producing a p-value that shows Mammals are significantly more likely to be protected.

As conservationists, we need to focus our efforts on understanding why Mammals and Birds are more at risk.

Further Study

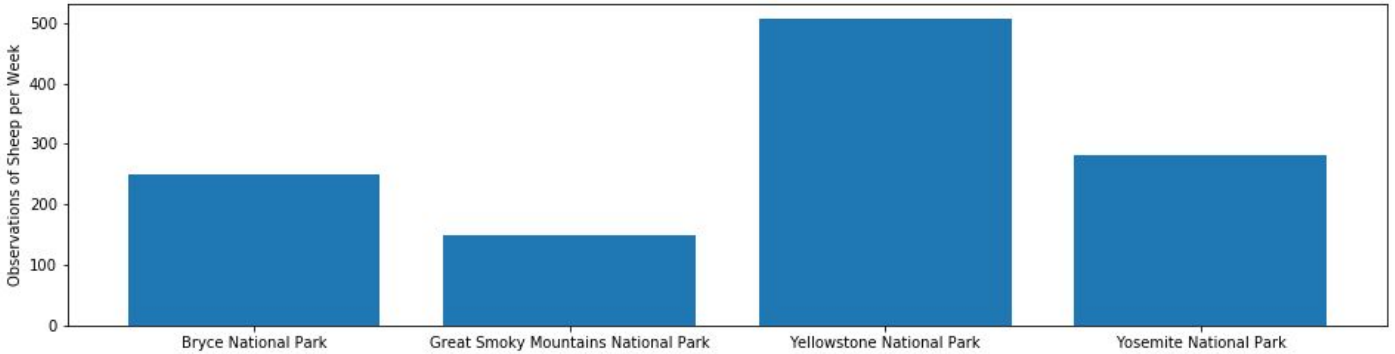
With the data in hand, we've taken initial steps to understand Foot and Mouth disease in sheep in our National Parks.

Sample Size for Disease Study

- The goal is to determine the effect of the rangers' program in Yellowstone to reduce Foot and Mouth disease.
- We are comparing this to the sheep population in Bryce
- The benchmark is a 15% disease rate in Bryce, and we need to detect if the program is achieving at least a 5 percentage point reduction in the disease in Yellowstone.
- To detect this, we'd need a sample size of 520 observations in each park.

Data for Disease Study

- Looking at the sightings of “sheep” species across the parks we have data for:



- We’d need to observe sheep for about 1.02 weeks in Yellowstone, and 2.08 weeks in Bryce.

Thank you!