Biodiversity Analysis for National Parks

By Fred Hertlein

Overview of Species Data

- Species data file has been stored in a CSV with the following columns and values:
  - category (Amphibian, Bird, Fish, Mammal, Nonvascular Plant, Reptile, and Vascular Plant
  - scientific_name
  - commons_names
  - conservation_status (NaN, Endangered, In Recovery, Species of Concern, and Threatened)
    - ❖ Changes were made to the data file to replace NaN or null with No Intervention

- A total of 5541 unique species have their data stored in the CSV

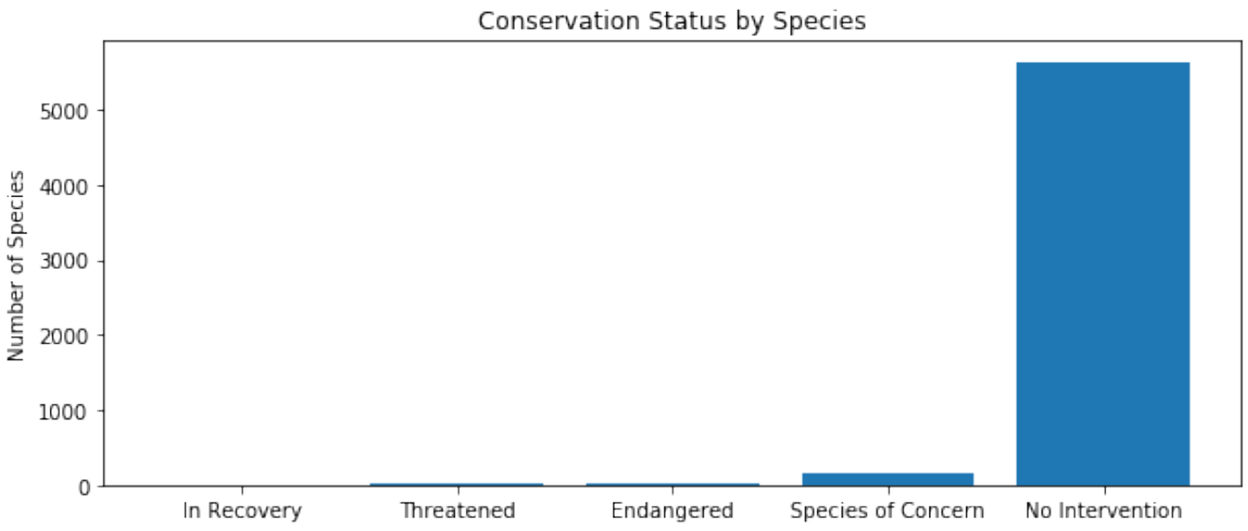
Calculations of Endangered Species by Category

1. Conservation status of all species was compared

2. An additional boolean identifier of protection status was assigned to each species

3. All species were grouped by category and sorted to compare total number of protected vs non-protected species

4. Finally this information was placed in a pivot table to with an additional column showing the percentage of protected species in each category



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

Significance Calculations of Endangered Species

- To compare two categorical datasets we used chi squared tests

- In our first comparison we made an assumption that mammals are more likely to be endangered compared to birds
  - Our chi squared test resulted in a p-value of roughly 0.69 which shows that there is NOT a significant difference in likelihood of endangerment between mammals and birds

- The second comparison assumed that mammals are more likely to be endangered compared to reptiles
  - Chi squared test resulted in a p-value of roughly 0.04 which proves there is a significant difference and thus mammals are more likely to be endangered than reptiles

- As such it is recommended that conservationists focused on endangered species should focus more of their efforts on mammals and birds than reptiles

Sample Size Determination

- An online sample size calculator was used to determine the number of samples that needed to be taken in order to detect a reduction in foot and mouth disease found in the sheep at different National Parks
  - Baseline conversion rate: 15% (current percentage of sheep with disease at Bryce)
  - Minimum detectable effect: 33% (5% change detected of 15% is 1/3 or 33%)
  - Statistical significance: 90% (confidence in results)

- From Optimizely's sample size calculator this results in 520 samples needed

- From 'observations.csv' we found that 250 observations occur weekly at Bryce National Park and 507 weekly observations are made at Yellowstone National Park
  - ➡ Roughly 2 weeks to collect the necessary samples at Bryce and 1 week at Yellowstone