BIODIVERSITY ANALYSIS

Biodiversity analytical data prepared for the National Parks Service for recommendations on resource allocation for species protection.

PROVIDED INFORMATION

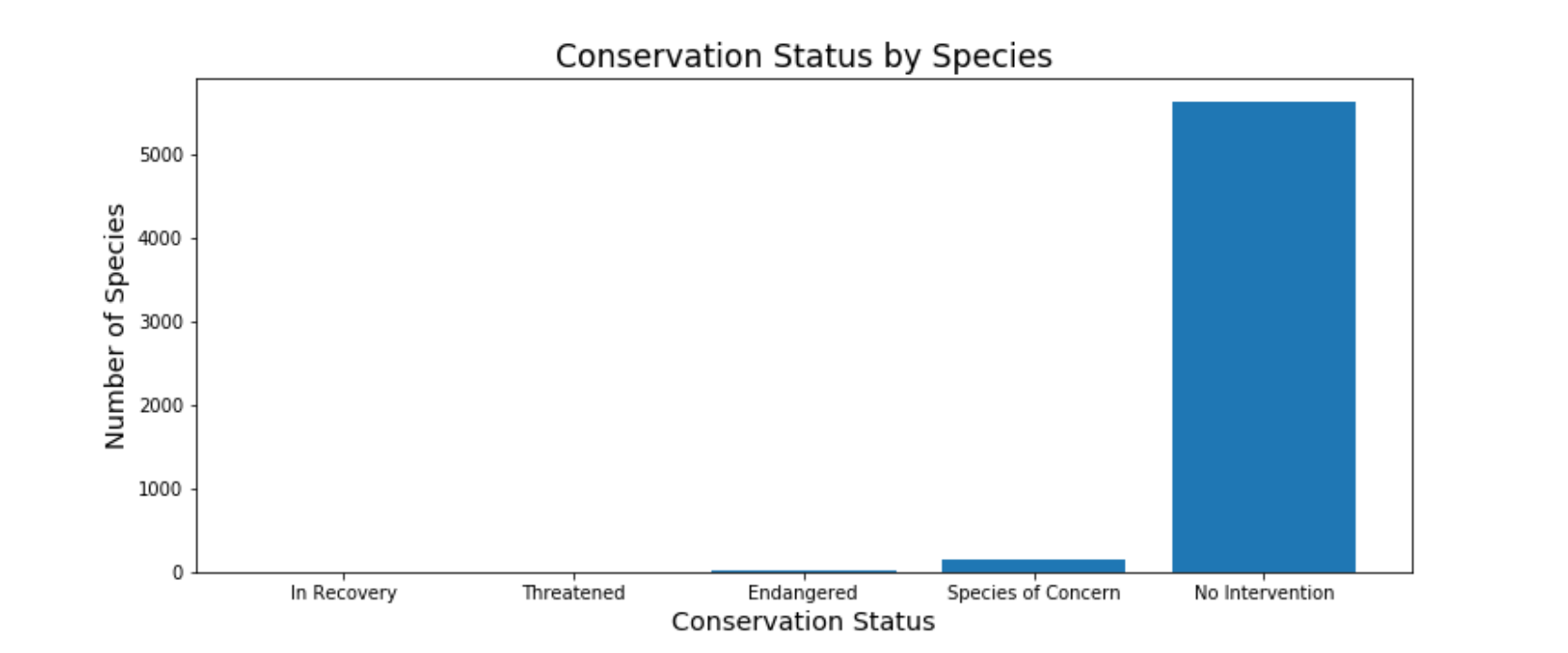
From database, species_info.csv

- Data file contains scientific name of species, common names and conservation status.

Data Deduction

- A large percentage of species have no conservation status. We assume these are species that are not at risk of being endangered, and have had no intervention to date.
- There are a total of 5541 species: Mammals, Birds, Reptiles, Amphibian, Fish, Vascular Plants & Nonvascular Plants.
- Conservation statuses: Not specified, Species of Concern, Endangered, Threatened, In Recovery

After identifying the blank conservation statuses as ‘No Intervention’ we plotted the relative frequencies of each conservation status, for all species.



- This plot raises the question: Which species are more likely to be endangered?
- If we can identify these ‘at risk’ species, the appropriate intervention can be taken.

STATISTICAL ANALYSIS

1. **The classes were separated into groups of ‘Protected’ and ‘Not-Protected’ and the counts are displayed in the following table.**
2. **Mammals and birds significantly stand out as being the most protected. Are mammals more likely to be endangered than birds?**
3. **We employed a Chi Square test to answer this question. The results of a Chi Square test indicate ‘IF’ there is a significant difference between the two phylums.**

	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

- Test results for Mammals and birds indicated no significant differences between the phylums, in terms of if one was more likely to be endangered or not (p value = 0.69).
- Test was run again for Reptiles and Mammals. Result (p value = 0.038) indicates significant difference.

ANALYZING THE RESULTS

The statistical analysis indicates that Mammals are more at risk of being endangered than reptiles.

We recommend that further analysis of data takes place to ensure the best allocation of resources for the best outcome for vulnerable species.

The following are our recommendations.

1. Allocate more resources to the protection of mammals than to reptiles.
2. Verify that the conservation statuses in the provided file are valid. This would require further statistical analysis. A starting point would be to obtain historical data of the number of species each year. An analysis of the change of species numbers over time would shed insight into the true stability of their respective populations.
3. Provide financial data, to verify that the available resources are being allocated in a manner consistent with the findings here.

PROVIDED INFORMATION

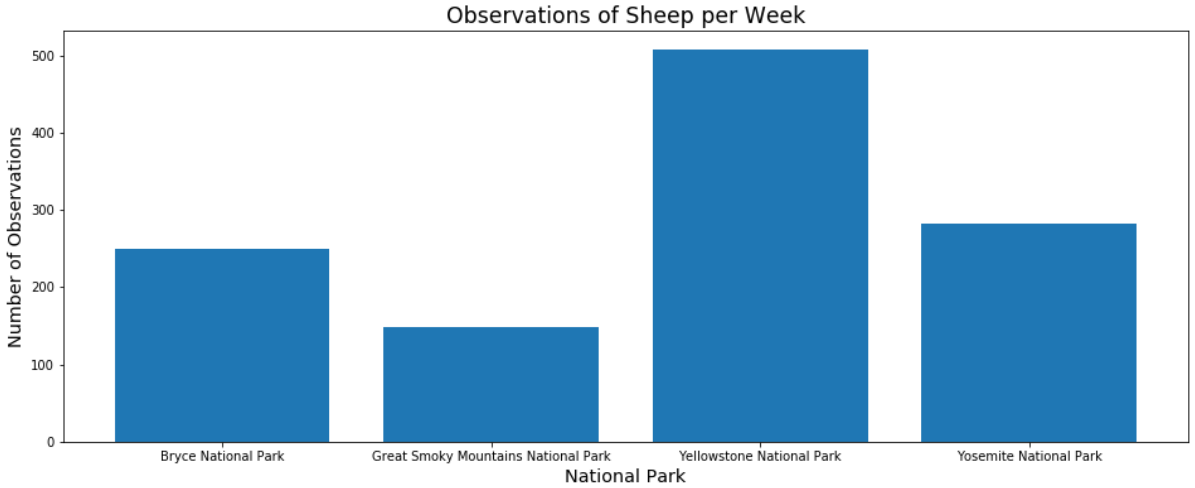
From Database, observations.csv

- Data file contains scientific name of species, park name, and number of sightings.

Notable features

- Assumes 15% of sheep have foot and mouth disease
- Park rangers want to be able to show if their disease reduction program is working or not. Our goal is to determine the sample size necessary for the results to show a detectable effect.

The data in observations.csv was arranged so that we obtained counts of the sheep sightings per week at each park.



Given the numbers of sheep sightings at each park, we can calculate the minimum number of weeks necessary to accumulate a sufficient sample size, for which to report the percentage of foot and mouth disease in the sheep population.

SAMPLE SIZE DETERMINATION

Minimum Detectable Effect: 33%

Baseline: 15% (based on Bryce National Park)

Using a standard sample size calculator we determined that 520 sheep observations would be necessary in order to determine the percentage with foot and mouth disease in the sheep population as a whole.

Based on the number of sightings per week, we can determine the duration for which the experiment should run.

- Bryce National Park: 3 weeks (750 Sheep)
- Yellowstone National Park: 2 weeks (1014 Sheep)

If the sheep are observed for this amount of time, sufficient data can be collected in order to determine if the disease reduction program is working as intended.