Biodiversity analyst working for the National Parks Service.

Goal is to analyze some data about species at various national parks.

BIODIVERSITY PROJECT

PROVIDED INFORMATION

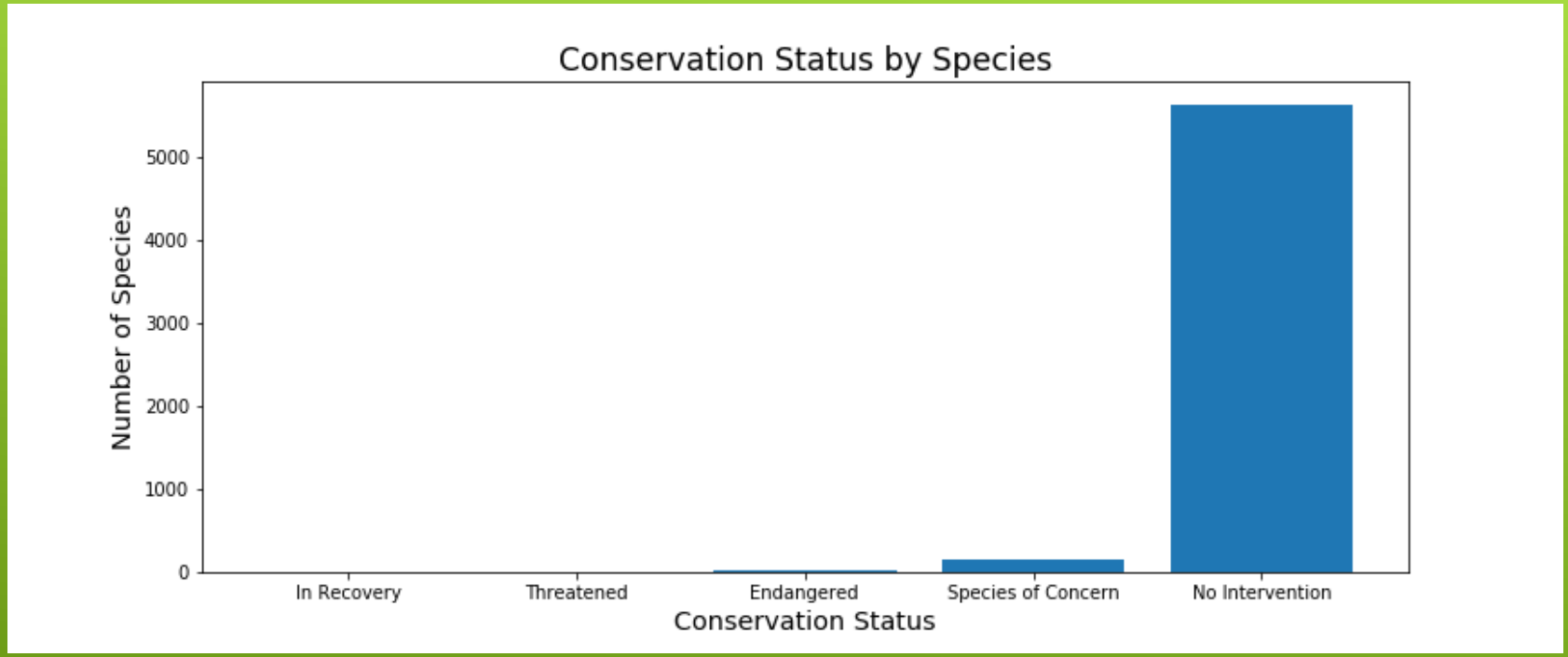
species_info.csv

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

Notable features

- Large percentage of species have no conservation status. Assume these are species that are not at risk of being endangered, and have had no intervention to date.
- 5541 total species: Mammals, Birds, Reptiles, Amphibian, Fish, Vascular Plant, Nonvascular Plant.
- 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

- The species were separated into groups of 'Protected' and 'Not-Protected' and the counts are displayed in the following table.
- Mammals and birds stand out as being the most protected. Are mammals more likely to be endangered than birds?
- 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 species.

	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 species, in terms of if one was more likely to be endangered or not (p value = 0.69).
- Test was rerun for Reptiles and Mammals. Result (p value = 0.038) indicates significant difference.

INTERPRETING THE RESULTS

- Our statistical analysis indicated that Mammals are more at risk of being endangered than reptiles.
- Where we go from here is up to the National Parks.
  - Option 1: Allocate more resources to the protection of mammals than to reptiles.
  - Option 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.
  - Option 3: Provide financial data, to verify that the available resources are being allocated in a manner consistent with the findings here.

PROVIDED INFORMATION

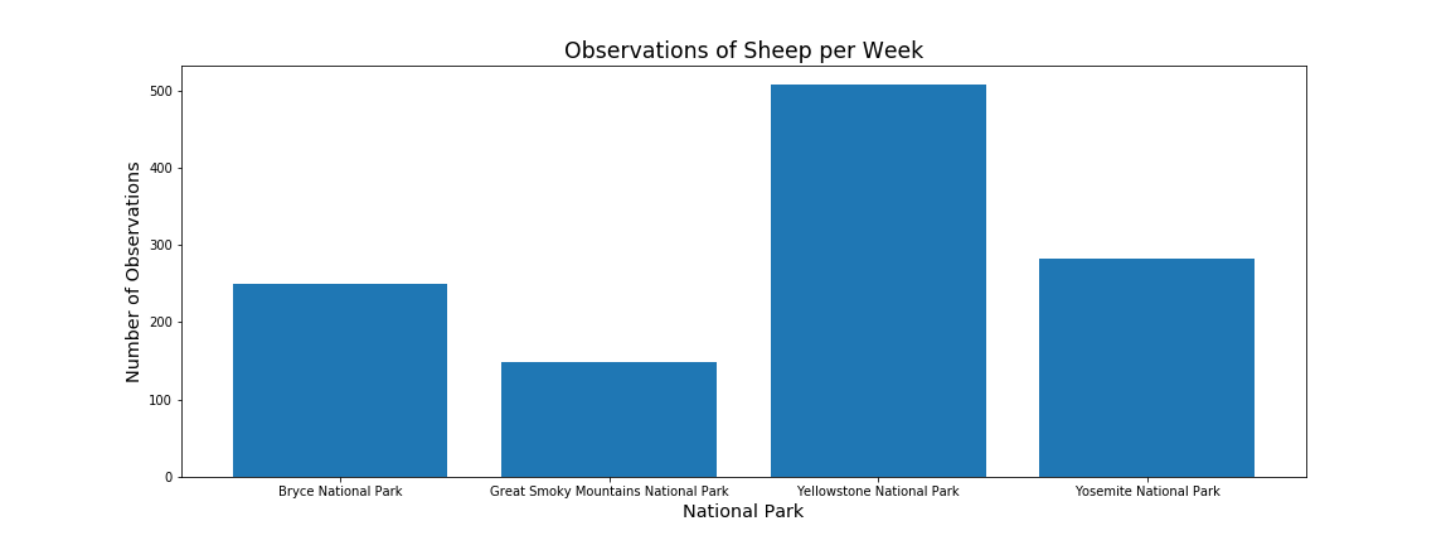
observations.csv

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

Notable features

- Assume 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 be a detectable effect.

- The data in observations.csv was sorted so that we obtained counts of the sheep sightings per week, at each of the parks.



- 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 (test for foot and mouth disease) would be necessary in order to speak about 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
  - Yellowstone National Park: 2 weeks
- If the observed sheep are tested for this amount of time, it is expected that sufficient data will be collected in order to determine if the disease reduction program is working as intended.