Biodiversity Analysis of National Parks

Lucas Mejia

Data provided

A CSV file named species_info.csv was given. It contains:

a. The scientific name of each species
b. The common names of each species
c. The species conservation status

The first five rows of the file look like this:

```
  category               scientific_name  \
0   Mammal  Clethrionomys gapperi gapperi
1   Mammal                      Bos bison
2   Mammal                      Bos taurus
3   Mammal                      Ovis aries
4   Mammal                  Cervus elaphus

                                       common_names conservation_status
0                           Gapper's Red-Backed Vole                 NaN
1                            American Bison, Bison                  NaN
2  Aurochs, Aurochs, Domestic Cattle (Feral), Dom...                 NaN
3  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)                 NaN
4                                      Wapiti Or Elk                 NaN
```

Quick overview of species_info.csv

- Total number of species: 5541
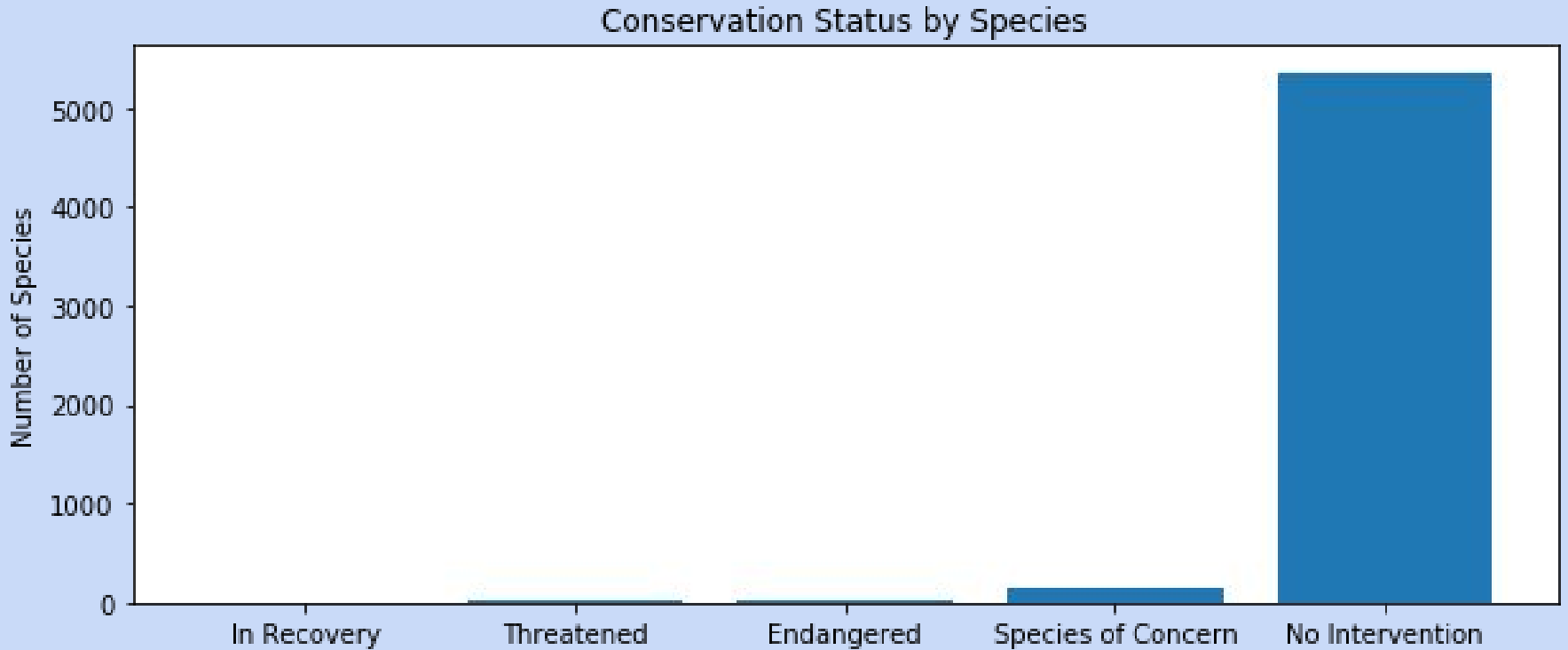- Types of species: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant

Conservation Status of Species in Parks

- Out of the 5541 species, 180 are endangered
- Following is a table summarizing number of endangered species and their conservation status (i.e how endangered they are)

Conservation Status	Number of Species
Endangered	15
In Recovery	4
Species of Concern	151
Threatened	10

Visualization of Conservation Status

Conservation Status by Species

Percent of protected species by category

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

Chi squared test to see if percentages of different categories are related or are a result of chance

Null hypothesis: The percentage of protected species in one category is independent of the percentage of protected species in other categories
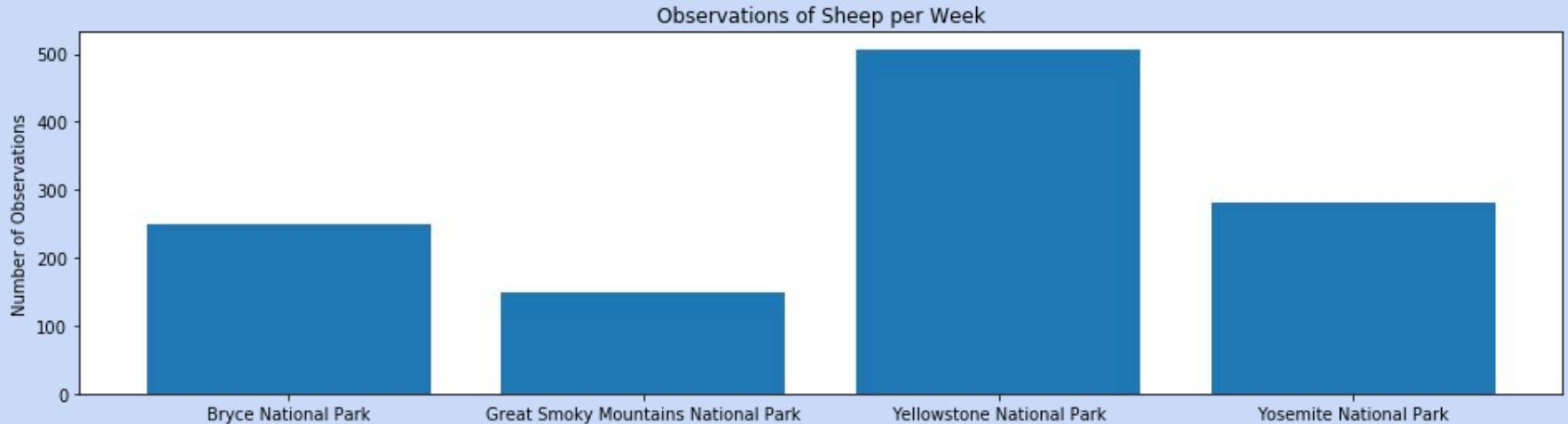
- Pvalue for test between birds and mammals: 0.69
  - The percentages seem to be independent
- Pvalue for test between reptiles and mammals: 0.038
  - The percentages are probably dependent

Recommendation based on significance results

Mammals' and reptiles' percentages of endangerment are coupled. These two categories probably share ecological factors that are being disrupted by man-made or natural causes. I recommend looking at:

- Shared areas where most of the species in these categories live (too much hunting or too many fires)
- Critters/plants that species from both categories eat (a common organism down the food chain is becoming endangered, less available food for both categories)

Observations of Sheep per Week

Sample size determination for foot and mouth study

- Baseline: 15%
- Significance: 90%
- Minimum detectable effect: 33.33% (percent to be detected/baseline)
- Sample size: **510**
- To observe in Yellowstone: 1 week
- To observe in Bryce: 2 weeks