Biodiversity for the National Parks

Capstone Project for Introduction to Data Analysis

Sample Data used

Using the file ‘species_info.csv’, the following was extrapolated.

- 5541 different species in the file.
- 7 different categories (Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, and Nonvascular Plant)
- 5 values for conservation status, NaN values were converted to ‘No Intervention’ (No Intervention, Species of Concern, Endangered, Threatened, In Recovery

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
1	Mammal	Bos bison	American Bison, Bison	NaN
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN

Significance calculations

From our file consisting of 5824 distinct rows, we want to count the number of species that meet each criteria. With this we can get a general idea of the number of species requiring some form of protection. Since Null values are not counted, we filled the NaN values with the text 'No Intervention' and sorted.

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	16
3	Species of Concern	161
2	No Intervention	5633

A new column is created, which is True if conservation_status is not equal to No Intervention, and False otherwise. We group and pivot and have the following

	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

Recommendations

Recommend that more bird and mammal species be put on the protected list as they are more likely to be endangered.

Little significance was found between them. This was tested using a chi square test with a return value of: 0.687594809666

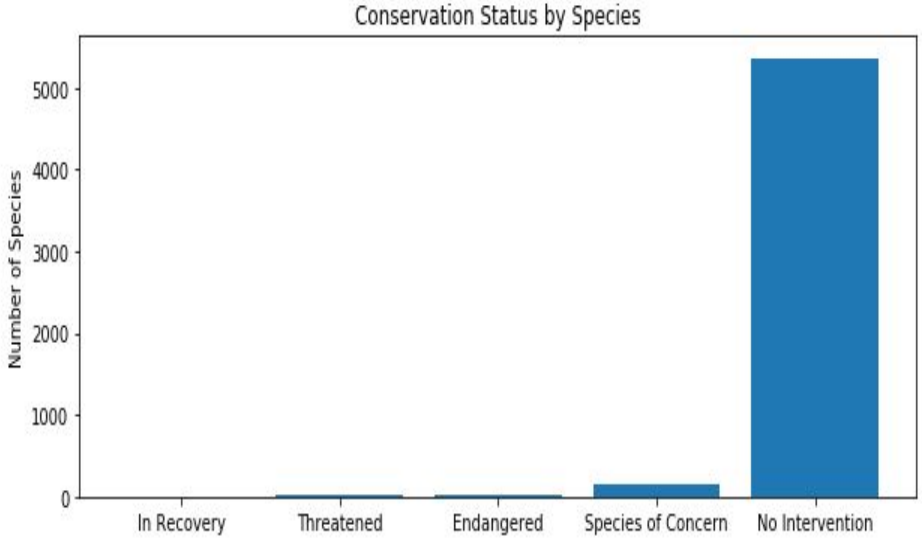
	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015

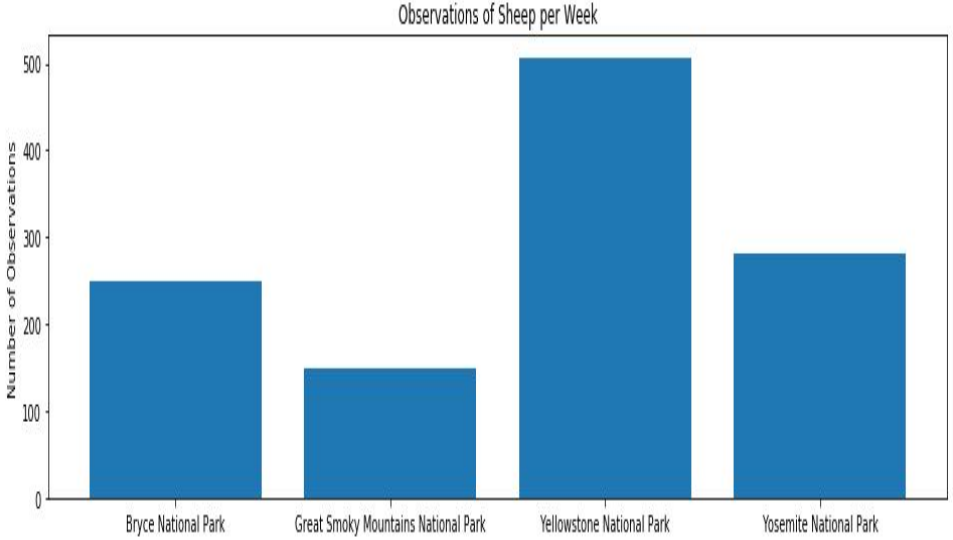
Sample Size Determination

The sample size calculator Optimizely was used to calculate the number of sheep that they would need to observe from each park. A sample size per variant (amount of sheep to be observed) was found of 510.

Using the following for bryce (510 / 250) and for yellowstone (510 / 507) we calculate that approximately 2 weeks at Bryce and 1 week at Yellowstone would be needed to observe sheep for foot and mouth disease.

Graphs


Conservation Status by Species
Number of Species
0
1000
2000
3000
4000
5000
In Recovery
Threatened
Endangered
Species of Concern
No Intervention


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park