BIODIVERSITY ANALYSIS

Bernardo Guerra

THE SPECIES DATA

- The data describes the different species of animals and plants in the:

1. Bryce National Park;
2. Yellowstone National Park;
3. Yosemite National Park;
4. Great Smoky Mountains National Park.

	category	scientific_name	common_names	conservation_status	is_protected
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	No Intervention	False
1	Mammal	Bos bison	American Bison, Bison	No Intervention	False
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	No Intervention	False
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False
4	Mammal	Cervus elaphus	Wapiti Or Elk	No Intervention	False
5	Mammal	Odocoileus virginianus	White-Tailed Deer	No Intervention	False
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	No Intervention	False
7	Mammal	Canis latrans	Coyote	Species of Concern	True
8	Mammal	Canis lupus	Gray Wolf	Endangered	True
9	Mammal	Canis rufus	Red Wolf	Endangered	True
10	Mammal	Urocyon cinereoargenteus	Common Gray Fox, Gray Fox	No Intervention	False
11	Mammal	Vulpes fulva	Black Fox, Cross Fox, Red Fox, Silver Fox	No Intervention	False
12	Mammal	Vulpes vulpes	Red Fox	No Intervention	False
13	Mammal	Felis concolor	Mountain Lion	No Intervention	False
14	Mammal	Felis silvestris	Wild Cat, Wildcat	No Intervention	False
15	Mammal	Lynx rufus	Bobcat	No Intervention	False

Part of the species.csv dataset

- Proceeding with the analysis of the dataset, it was noticed that although Vascular Plants, Birds and Non-Vascular Plants had the 1st, 2nd and 3rd most abundant number of species, respectively, did not have the biggest percentage of species of their category protected by the park

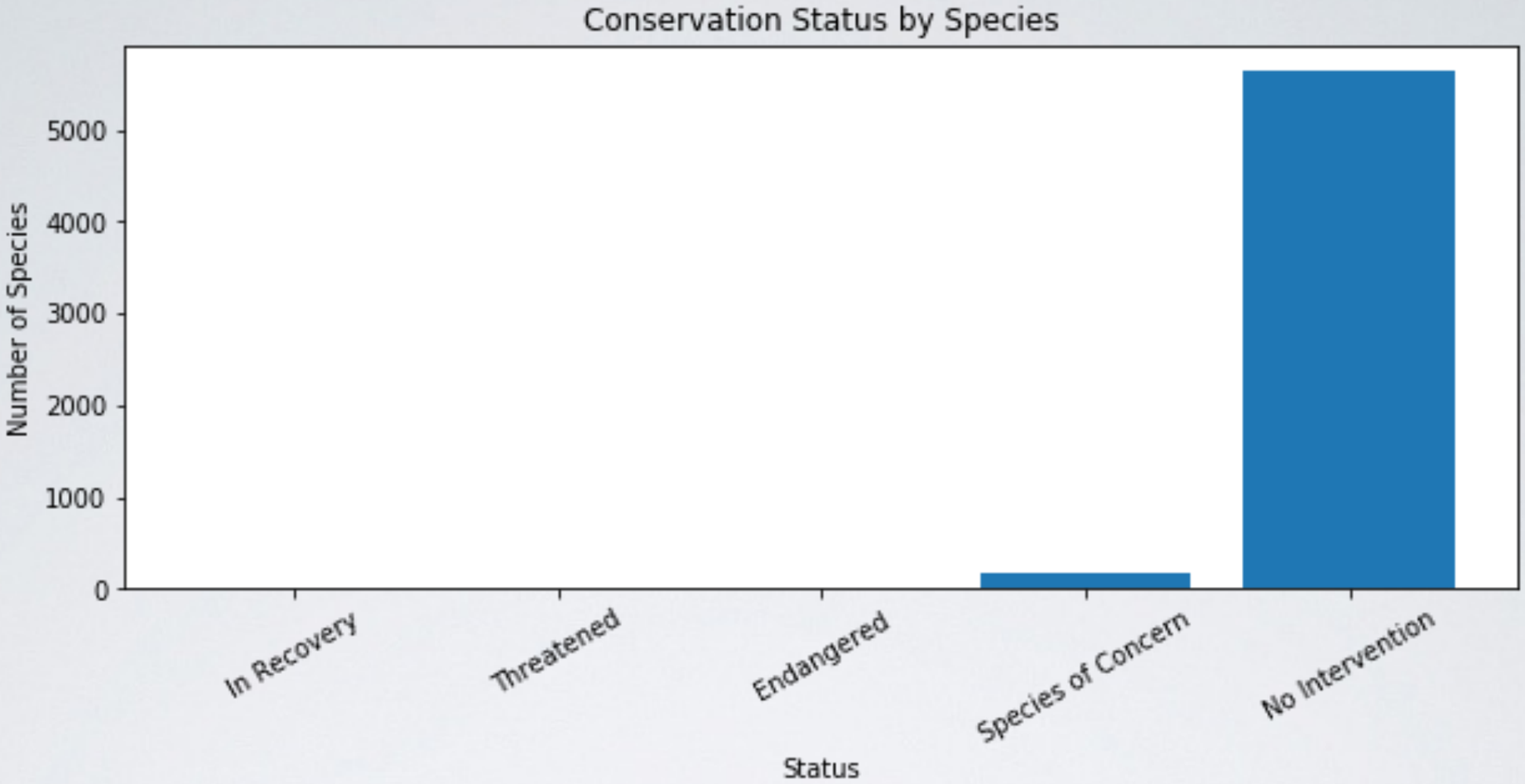
- In this matter, Mammals occupied the 1st place with 17% species protected, followed by birds with 15.4% and Amphibians with 8.9%.

- This fact seemed to indicate that the species of the Mammal category were more likely to be endangered than birds, amphibians and reptiles for example.

- In relation to birds, amphibians and fish, this was proven incorrect with a p-value from a chi-square test of about roughly 0.69, 0.12 and 0.056, respectively, disproving the previous hypothesis.

- However, when run for the other categories, the hypothesis proved correct, with p-values as in the table below:

Reptiles - Mammals	Non-vascular Plants - Mammals	Vascular Plants - Mammals
0.038	1.48E-10	1.44E-55



- However, when analyzing the overall number of species in the different status, it was noticeable that the species that were at no risk of danger, and therefore required no intervention, were far more abundant (5363)
- Species in risk of becoming endangered (Concern) followed in the second most abundant with 151 total species
- Endangered, Threatened and Recovering species followed behind with 15, 10 and 4 species, respectively
- This shows that although many species are in a state of no intervention, there is a high number of species of concern which may become endangered, worsening the scenario, considering the small number of recovering species.

RECOMMENDATIONS

- In order for the recovering of more of the endangered species and the protection of species of concern, the following may be incorporated in the park:

    A. Implementation of a more advanced security system in the entrance of the Parks. These may include:

    1. Metal detectors to prevent entrance with weapons and other prohibited tools, permitting only tools necessary for enjoying the park and in aid in a difficult situation

    2. Electric fence around the entirety of the Park in collaboration with private investors and other donors

    3. Keep a book of reports which prevents visitors who have brought prohibited items or entered the park illegally to prevent further access.

RECOMMENDATIONS

B. Open Conservation Events:

1. Can be done through the opening of the park to the public in specific events which welcome any volunteers and focus on attracting veterinarians and biologists for aid in conserving and protecting the area

2. Credit system for any individual who aids the conservation and protection of the park in any way (removing trash, cleaning, helping an animal, etc) which could be used for purchase of items in the park

*The recommendations were based solely on the species dataset provided and calculations done upon it.

SIGHTINGS AND OBSERVATIONS DATA

- Along with the species data, a dataset with the compilation of various sightings around the parks in the period of a week was also provided

- The dataset was divided into species and parks where they were seen

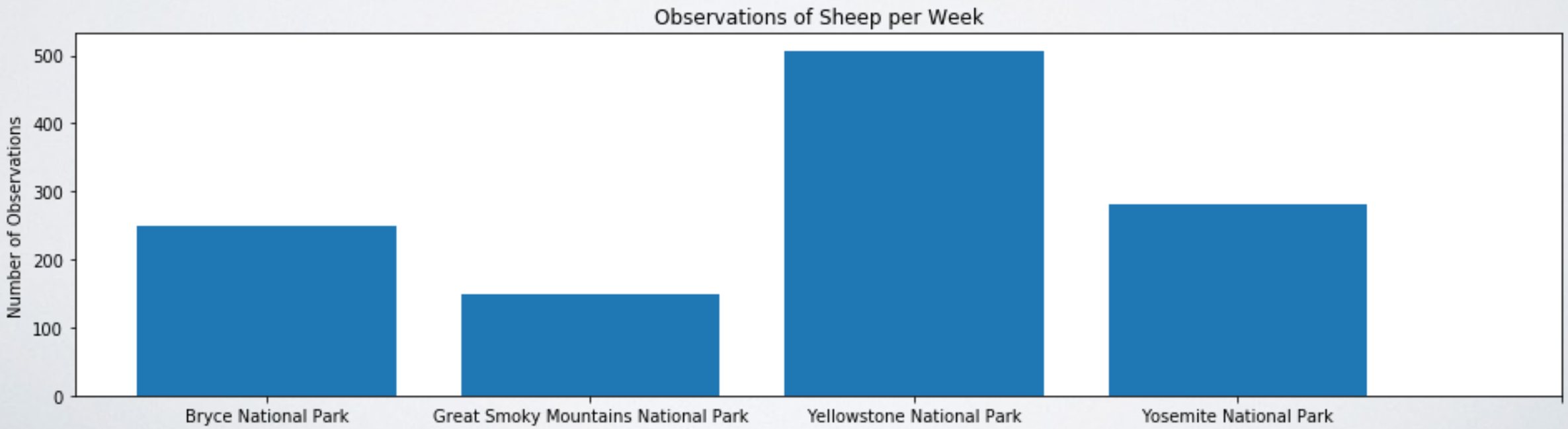
- This dataset was used to help study the endangered species observed and test for the foot and mouth disease in species of sheep in the parks

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

Part of the observations.csv dataset

PROCESSING THE DATASET

- The observations dataset was processed and grouped by the species and selecting only those that belonged to the sheep family.
- Then, the sightings were separated and counted for each national park, as in the graph below



FOOT AND MOUTH DISEASE

- Based on the results, a sample size determination calculation was done for finding how many sheep should be studied for significant results in the foot and mouth disease study
- The following values were used:
    1. Level of significance of 90%
    2. Minimum detectable effect of 33.3%
    3. Baseline Conversion rate of 15%
- The result returned was a sample size of 510 sheep

CONCLUSIONS

- Due to the results received, a calculation against the different parks was done to find the time it would take on average to find the 510 samples needed. The results are as below:

    1. Bryce National Park - 2 weeks
    2. Yellowstone National Park - 1 week
    3. Yosemite National Park - 2 weeks
    4. Great Smoky Mountains National Park - 3.5 weeks

- Due to these results, the study should be done in the Yellowstone National Park, for the abundance of observations of the different species of sheep or at Bryce National Park or Yosemite National Park.