Biodiversity Capstone Project

By Bohdana Golubiev

Species Info

The information in species_info.csv specifies the category, scientific name, common names and conservation status of different animal species. The following presentation summarizes the analysis and conclusions drawn from further introspection into this data.

	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), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Species Information

- There are 5541 species listed

- The species listed are: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', 'Nonvascular Plant'

- The Conservation Status values are 'Species of Concern', 'Endangered', 'Threatened', 'In Recovery' as well as 'No intervention'.

Conservation Status by Species
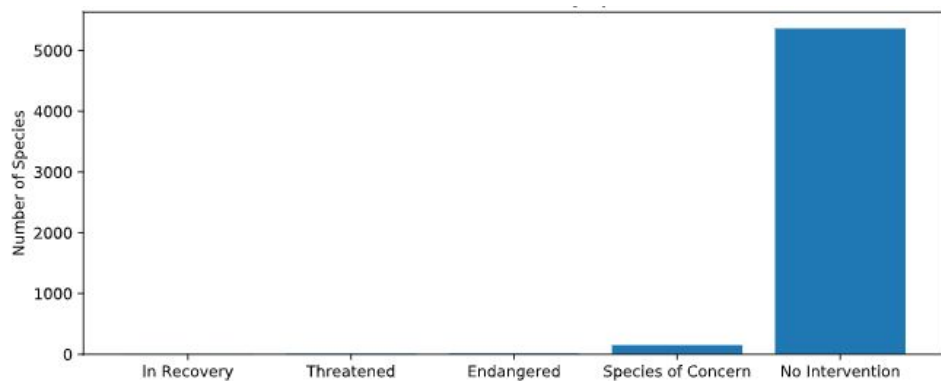
Species that require no intervention (count 5363) are by far outnumbering other categories.

However, species of concern (count: 151), endangered (count: 15), threatened (count: 10) and species in recovery (count: 4) also require our attention.

Percent protected

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

It seems that certain species are endangered to a greater degree than other types of species, as indicated by the column 'Percent Protected'. But a more thorough look will help us determine if that is the case.

A Chi-Squared test would help determine if there is a significant difference between the percentages in certain species and if such conclusions can be made when data is extrapolated.The data we have is categorical and we are comparing two groups with two conditions each. The result revealed no significant difference in the endangerment status of Mammals and Birds, but a significant difference between Reptiles and Mammals.

We draw these conclusions as the p-value for the first comparison is greater than 0.05 (~0.68, difference observed in this sample is due to chance) and it is smaller than 0.05 for the second comparison (~0.038, difference observed in this sample is NOT due to chance).

We can conclude that certain types of species are indeed more likely to be endangered than others. Specifically, according to our analysis, Mammals are more endangered than Reptiles. Conservationists should pay attention to all the species, but Mammals would require closer attention than Reptiles.

In Search of Sheep

We have selected sheep species and examined in which parks they are locating from the data provided.

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Observations of Sheeps per Week
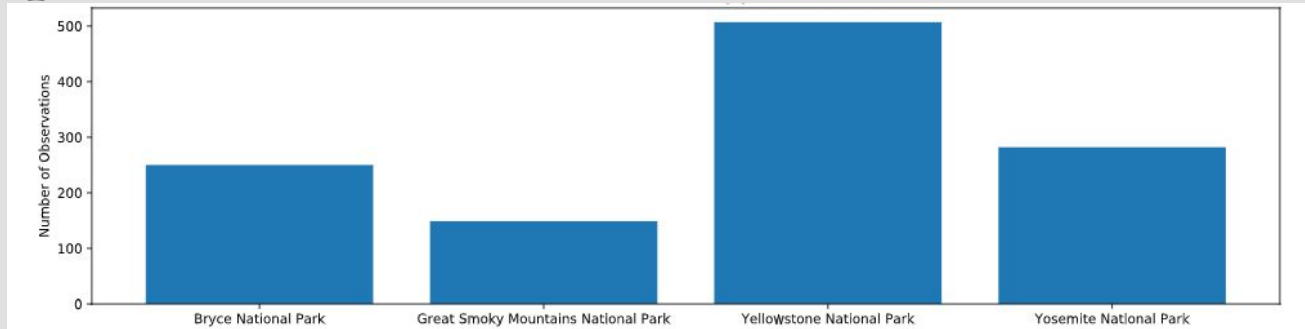
Foot and Mouth Disease Sample Size Determination

-Baseline Conversion Rate: 15%, as it is the percentage that was recorded at Bryce National Park and therefore the estimate that we consider.
-Minimum Detectable Effect: 33%, as it is calculated 100*5/15.
-The significance level is set to 90%.

Given these parameters, using the sample size calculator, it was determined that there needs to be a sample size of 870 sheep in Yellowstone Park in order to conclude that a drop of over 5% is a significant result.

It would take 1.7 weeks for Yellowstone Park to observe the necessary amount of sheeps (870 / 507)
It would take 3.5 weeks for Yellowstone Park to observe the necessary amount of sheeps (870 / 250)