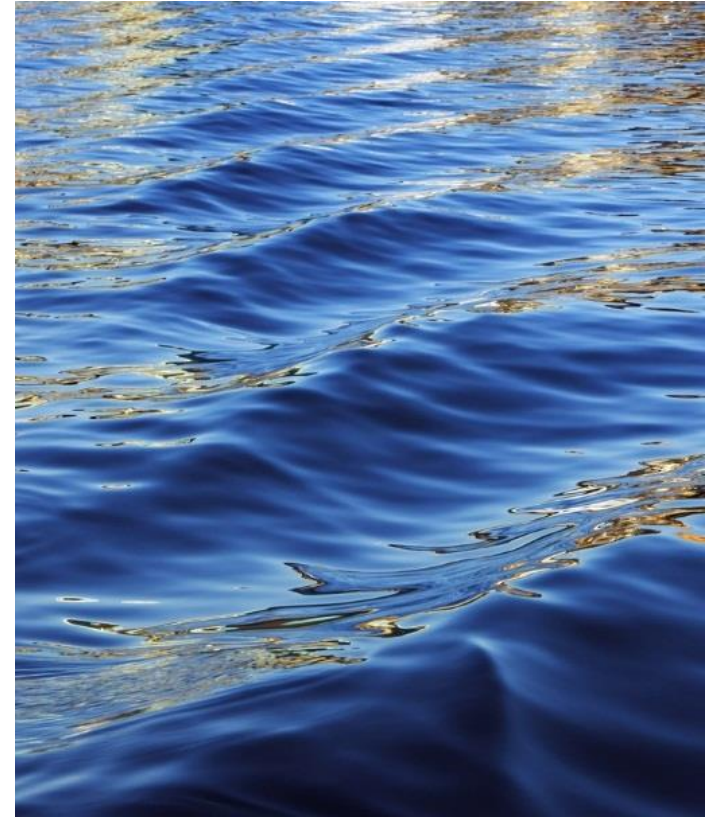# Biodiversity Capstone Project - Investigating Protected Species

Brandon Artura

Species Information Analysis

The Species dataset includes the scientific names and common names of each species and their conservation status. It contains 5,824 records.

- The dataset contains 4 fields: Category, Scientific Name, Common Names and Conservation Status

- Based on the unique number of Scientific Names, there are 5,541 unique species

- There are 7 unique categories in the dataset, they include
  - Mammal
  - Bird
  - Reptile
  - Amphibian
  - Fish
  - Vascular Plant
  - Nonvascular Plant

- There are 5 conservation statuses for the species, NaN values have been changed to "No Intervention"
  - No Intervention
  - Species of Concern
  - Endangered
  - Threatened
  - In Recovery

Species Information Analysis

species_info.csv sample, first 5 rows

	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

Endangered status between different categories of species

Based on the analysis of the dataset, there are 179 protected species and 5363 non protected species. The table below shows the distribution in each category.

Category pivot table, Protected species vs. Non protected

	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

Based on information from the table it seems that Mammals are more likely to be endangered than other species. In order to confirm this, we need to run some significance tests.

Endangered status between different categories of species

The top 4 endangered species from the table are: Mammals, Birds, Amphibians and Reptiles. Using the Chi Square Test for Independence with SciPy we can validate the following hypothesis.

- Mammals are more likely to be endangered than Birds
  - The results demonstrate no significant difference
  - P-value: 0.688

- Mammals are more likely to be endangered than Amphibians
  - The results demonstrate no significant difference
  - P-value: 0.128

- Mammals are more likely to be endangered than Reptiles
  - The results demonstrate that there is a significant difference
  - P-value: 0.038

Based on the results of the significance test, the hypothesis that Mammals are more likely to be endangered than Birds has been accepted (p-value: 0.687). Based on these results, we can conclude that further efforts and resources should be allocated to protecting Mammals.

Foot and Mouth Disease Study

Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage points. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

Based on the results of a sample calculation with a 90% significance
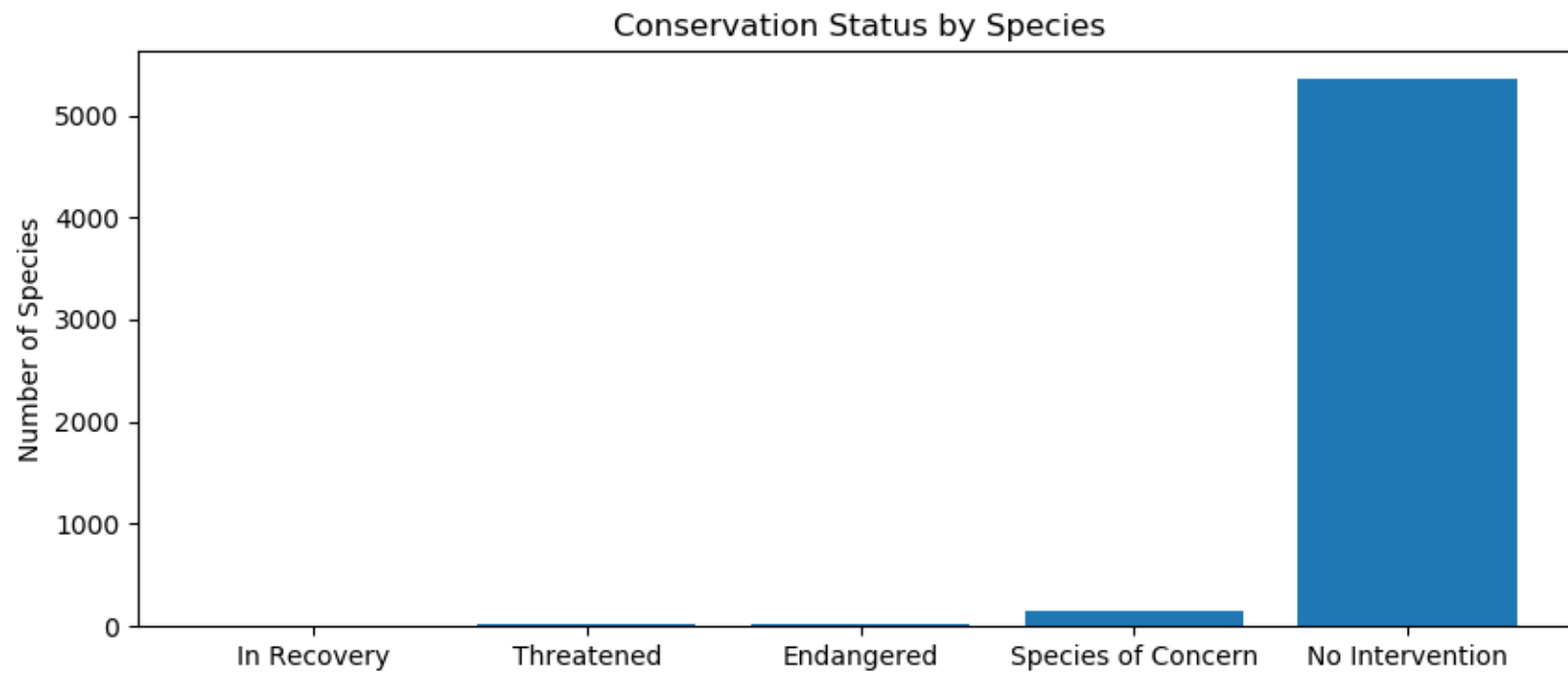
The Minimum Detectable Effect is equal to 33.33%, this is translated to a sample size of 870 sheep

In order to reach the desired level of observations in each park selected:

It will take **1.7** weeks at **Bryce National Park** based on **250 sheep** observations per week

It will take **3.5** weeks at **Yellowstone National Park** based on **507 sheep** observations per week

Conservation Status By Species

Observations of Sheep Per Week