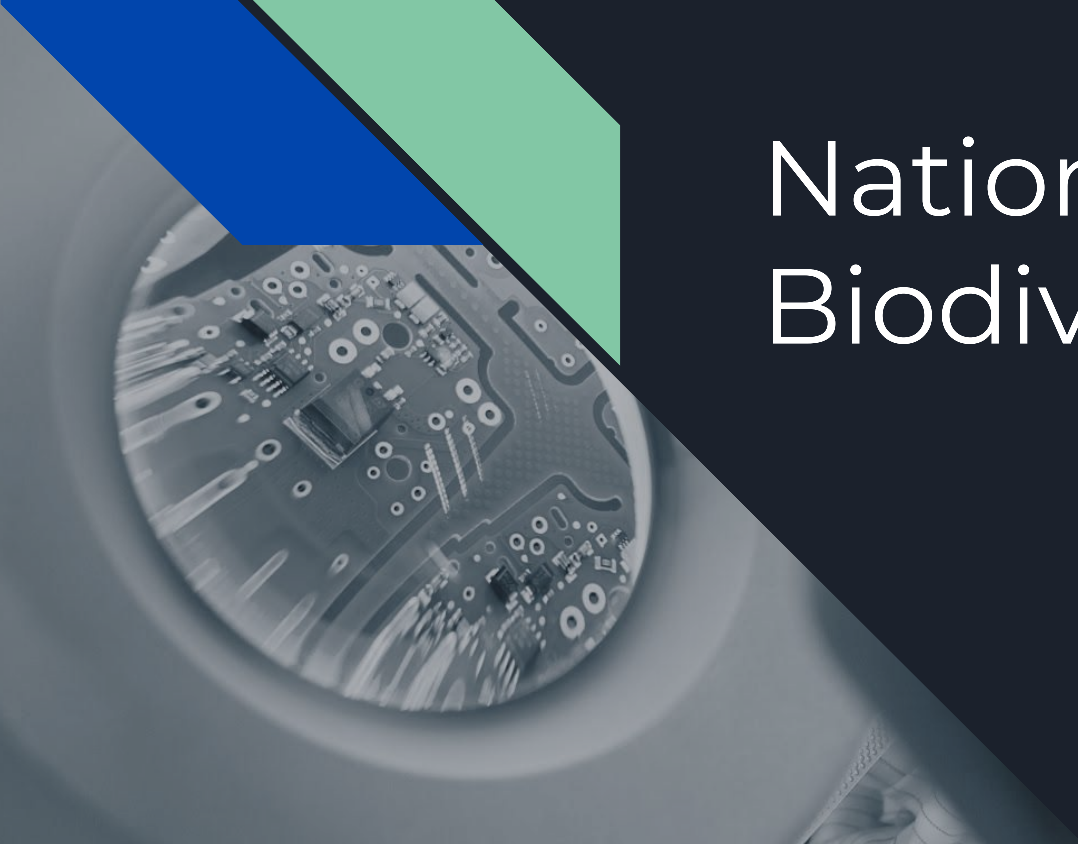


National Parks' Biodiversity

Endangered Species Data Analysis

By Daniel Kimm, Biodiversity Analyst
National Parks Service

National Parks Data Information

Data contains 5,541 unique species found across different national parks with the following:

4 Data Categories:

Category (Specie Type)

Scientific Name

Common Name

Conservation Status

7 Category of Species:

Mammal

Bird

Reptile

Amphibian

Fish

Vascular Plant

Nonvascular Plant

5 Conservation Status:

Species of Concern

Endangered

Threatened

In Recovery

No Intervention (species with no conservation status)

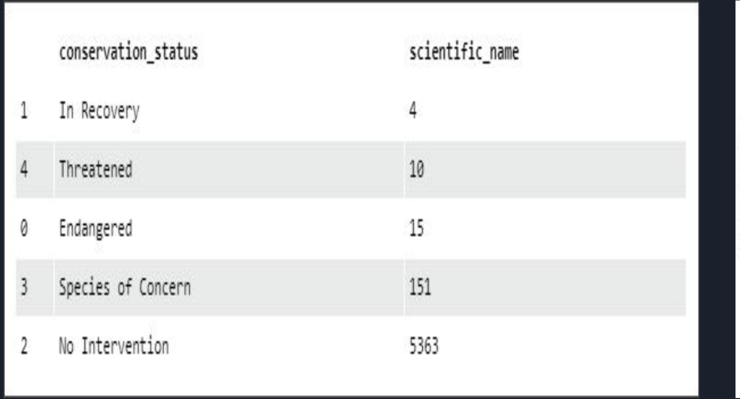
Key Findings of Species’ Conservation Status

- Table and graphic below represents counted number of species by their conservation status, and are arranged from the least to greatest

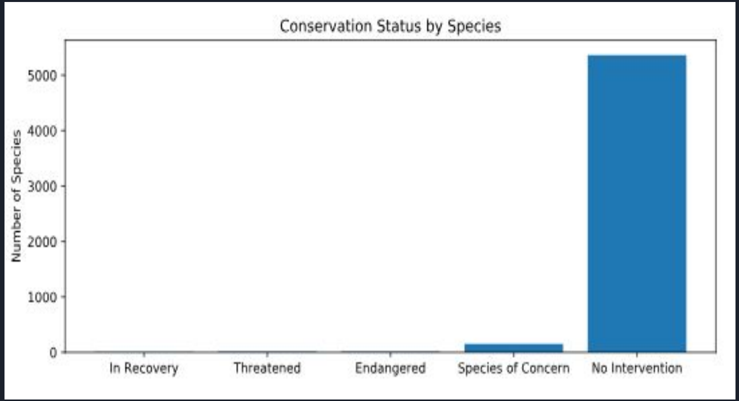
- Majority of our species require no protection (5,363)

- 4 species are in recovery

- 151 species that may be in need of conservation (species of concern), and 25 species face risk of extinction (threatened or endangered)



	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	15
3	Species of Concern	151
2	No Intervention	5363



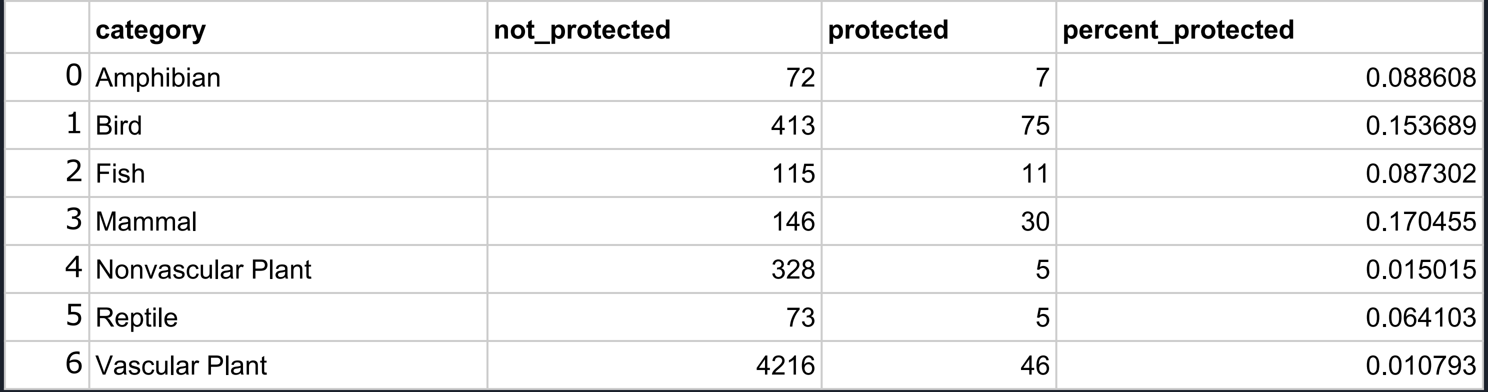
Analysis of Endangered Species

Are certain types of species more likely to be endangered?

- A pivot table used to group by specie category, protection status, and percent of unique species within a category that were protected.

- “not_protected” column: count of unique species that have “no intervention”

- “protected” column: count of unique species with conservation status not equal to “ no intervention”



	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

Species' Protected Status Comparison Analytic Methodology

How to determine if there is significant difference between different category of species and their protected status numbers?

Using **Chi-Squared test** can help understand if the numerical difference between distributions of categorical data have statistical significance or were attributed to chance (null hypothesis).

Two chi-squared test were performed:

- Mammal vs. Birds

- Mammal vs. Reptile

Chi-Squared Test #1: Mammal vs. Birds

- Data observation: Mammals are more likely to be endangered than Birds

- 17% of species in Mammal category are protected vs. 15% of species in Bird category

- Is the difference due to chance?

- Null hypothesis: no significant difference between the mammal dataset and the bird dataset

    - p-value < 0.05 means the null hypothesis is rejected and there is significant difference

- A contingency table was created and the chi2_contingency() function from scipy.stats was used to generate p-value.

- Chi-squared test reveal p-value = 0.688; p-value > 0.05

Conclusion: Null hypothesis cannot be rejected and there is no significant difference between Mammal and Bird.

Chi-Squared Test #2: Mammal vs. Reptile

- Data observation: Mammals are more likely to be endangered than Reptiles

- 17% of species in Mammal category are protected vs. 6% of species in Reptile category

- Is the difference due to chance?

- Null hypothesis: no significant difference between the mammal dataset and the bird dataset

    - p-value < 0.05 means the null hypothesis is rejected and there is significant difference

- A contingency table was created and the chi2_contingency() function from scipy.stats was used to generate p-value.

- Chi-squared test reveal p-value = 0.038; p-value < 0.05

Conclusion: Null hypothesis is rejected and there is significant difference between Mammal and Reptile.

Recommendation for Conservationists concerned about Endangered Species

From chi-squared test results, mammal and bird categories may not have a significant difference between them, but they are the top two categories of species more likely to become endangered than the other specie categories.

- This conclusion is based on percent_protected values: the higher the percent value, the more likely the specie will become endangered.
    - Vascular and nonvascular plant species are least likely to become endangered.

BONUS: Revisiting Question: Are certain types of species more likely to be endangered?
Based on the two chi-squared tests, YES! Certain types of species are more likely to be endangered than others!

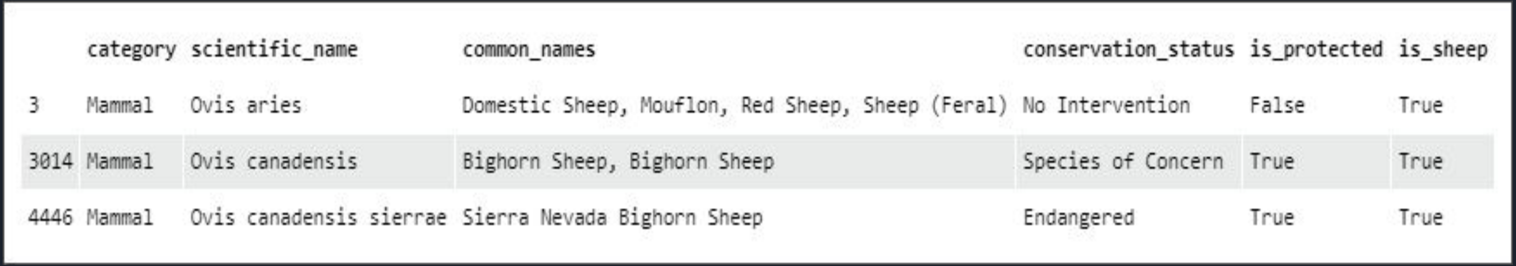
Sheep Observations Dataset

Conservationists have been recording sightings of different species at several national parks for the past 7 days.

An “is_sheep” data column was added to collected dataset and data was filtered to only include mammal sheep species

Data contains 3 sheep species: Ovis aries, Ovis canadensis, and Ovis canadensis sierrae

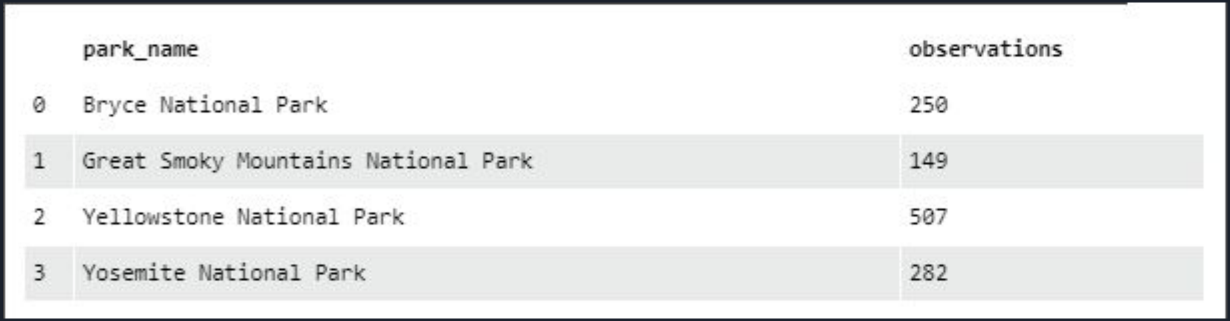
Analysis of sheep dataset provided the following:



	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

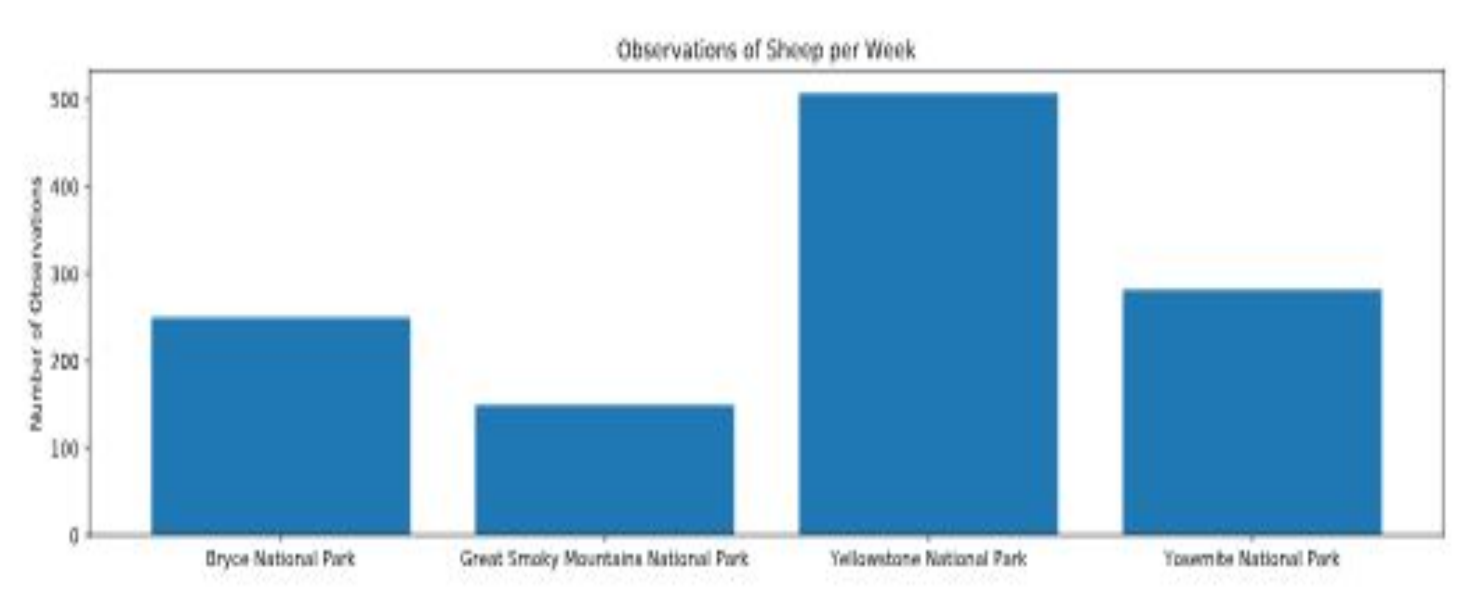
Count of Sheep per National Park

The observations data was then merged with our previous species data. The following is a total sheep observations across the 3 species, grouped by the national park:



	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

Chart of Sheep Sightings at Four National Parks



Foot and Mouth Disease Evaluation

- Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park.

- Objective is to find out whether or not this program is working (A/B Test), and to detect reductions of at least 5 percentage points.

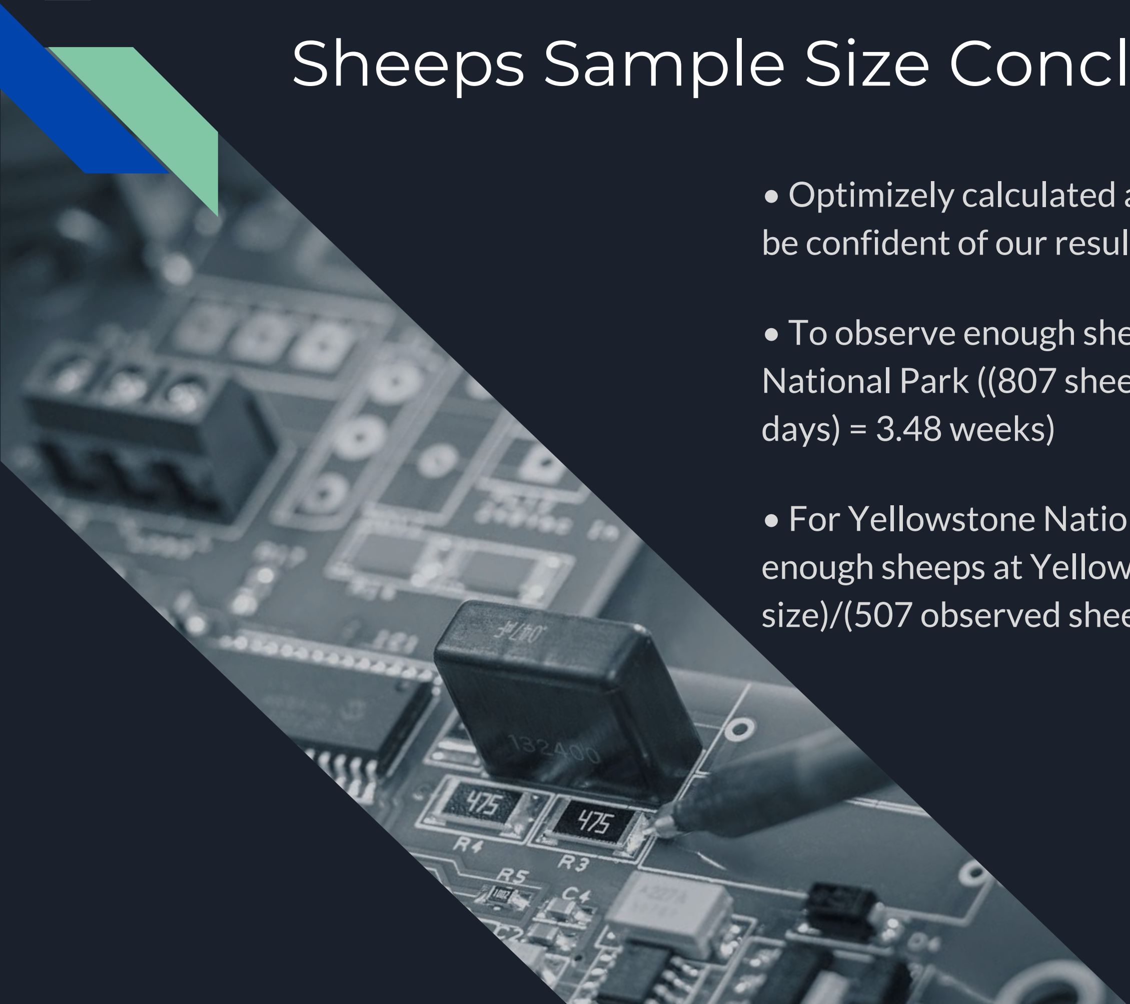
Known data:

- Last year, 15% of sheep at Bryce National Park have foot and mouth disease.

Determining Sample Size for A/B Test

- Optimizely was utilized to determine the sample size needed for A/B test, or the number of sheep observations that needed to be made at each park.
- For the calculator:
    - Baseline conversation rate: 15%
    - Minimum Detectable Effect: 33.33%
    - Statistical Difference: 90

Sheeps Sample Size Conclusion



- Optimizely calculated a sample size of 807 sheeps is needed in order to be confident of our results for the foot and mouth disease study.

- To observe enough sheep, we would need about 3.5 weeks at Bryce National Park ((807 sheep sample size)/(250 observed sheeps during 7 days) = 3.48 weeks)

- For Yellowstone National Park, we need about 2 week to observe enough sheeps at Yellowstone National Park ((807 sheep sample size)/(507 observed sheeps sighting during 7 days) = ~1.72 weeks)