



Biodiversity in our National Parks



By: Lindsay Hutchins



Problem Statement

The National Parks Service needs some data analysis performed on the **conservation statuses** of a selection of species. First, I must investigate if there are **any patterns or themes** to the types of species that become endangered.

During this project I analyzed, cleaned, and visualized data. I then posed questions and sought to answer them in a meaningful way.

Data in species_info.csv

Contains data about different species in our National Parks, including:

- The scientific name of each species
- The common names of each species
- The species conservation status

Specific Columns:

1. Category
2. Scientific_Name
3. Common_Names
4. Conservation_Status

Top 10 Rows of species_info.csv

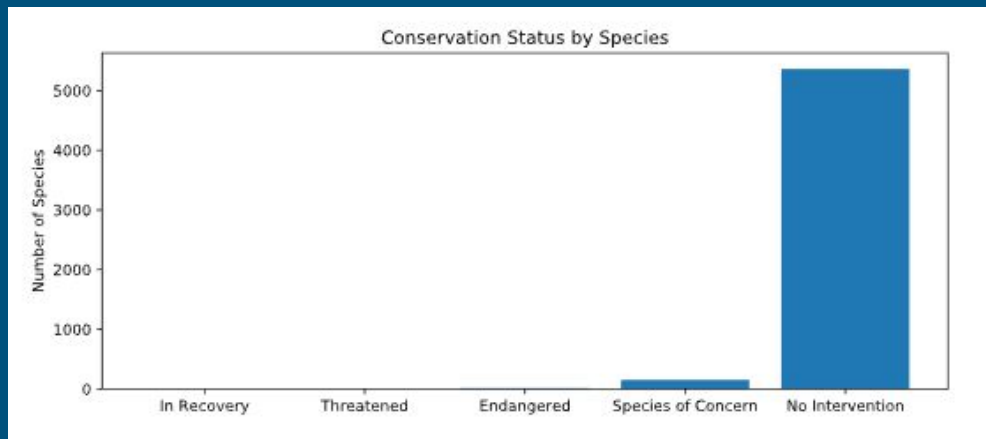
ind	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan
5	Mammal	Odocoileus virginianus	White-Tailed Deer	nan
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	nan
7	Mammal	Canis latrans	Coyote	Species of Concern
8	Mammal	Canis lupus	Gray Wolf	Endangered
9	Mammal	Canis rufus	Red Wolf	Endangered

Number of **unique species** in data: **5,541**

Types of species included: Mammals, Birds, Reptiles, Amphibians, Fish, Vascular Plants, Nonvascular Plants.

Possible **Conservation Statuses:** **Endangered**, **Threatened**, **Species of Concern**, **In Recovery** or **No Intervention**.

Conservation Status	Number of Species
Endangered	15
Threatened	10
Species of Concern	151
In Recovery	4
No Intervention	5,363



Distribution of protected species by Category

Here, “Protected” means all those with a status that is *not* No Intervention.

Category	Count Protected	Count <u>not</u> Protected	Percent Protected
Amphibian	7	72	~8.9%
Bird	75	413	~15.4%
Fish	11	115	~8.7%
Mammal	30	146	~17%
Nonvascular Plant	5	328	~1.5%
Reptile	5	73	~6.4%
Vascular Plant	46	4216	~1.1%

Question: Are certain types of species more likely to be endangered?

From the previous pivot table, it looks like mammals are most likely to be endangered, but is the difference significant enough to prove this point?

Methodology: Chi-squared Test

Null hypothesis: This difference is due to chance.

Significance Defined as: A value < 0.05

Mammals v. Birds: There was no significant difference - this is most likely a result of chance.

Mammals v. Reptiles: There was a significant difference - therefore certain types of species are more likely to be endangered.

Recommendation: Further examine specific types of species, and what risk factors could uniquely affect them on a type level.

Data in observations.csv

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

Columns:

1. Scientific_name
2. Park_name
3. Observations (count)

Top 3 Rows:

scientific_name	park_name	observations
Vicia benghalensis	Great Smoky Mountains National Park	68
Neovison vison	Great Smoky Mountains National Park	77
Prunus subcordata	Yosemite National Park	138

Foot and Mouth Reduction Effort

Problem Statement: 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.

Last year it was recorded that **15%** of sheep at **Bryce National Park** have foot and mouth disease.

How many sheep need to be observed from each park to make sure recorded foot and mouth percentages are significant?

Sample size determination

15% of sheep at Bryce National Park have foot and mouth disease, so **our baseline conversion rate is 15%.**

Our **minimum detectable effect** (meaning the value we must achieve to reach significance) is $100 * 5$ (the value change we want - 5% reduction) / 15 (baseline) = **33.3%**

This means with the default level of significance (90%), we require a **sample size of 870.**

Estimated Work Effort

- For **Yellowstone National Park**, we would need **~1.7 weeks** of data to get this sample size (at 507 observations from week)
- For **Bryce National Park**, we would need **~3.5 weeks** of data to get this sample size (at 250 observations/week)
- This data was gleaned from our conservationist observations by just looking at sheep by park:

