



Biodiversity for the National Parks

Capstone Project Option 2

Codecademy Pro Intensive: Introduction to Data Analysis

Tayler Scory, M.Sc.

Data Description: **species_info**

- Variables in species_info.csv:
 - Category
 - Mammal, bird, reptile, amphibian, fish, vascular and nonvascular plant
 - Scientific name and common name of each species
 - 5541 different species
 - Conservation status of each species
 - Species of concern, endangered, threatened, in recovery, no intervention (NaN)

Data Description: **species_info**

Table 1: Number of species that fall into each conservation status

| Conservation Status | Number of Species |
|---------------------|-------------------|
| Endangered | 15 |
| In Recovery | 4 |
| Species of Concern | 151 |
| Threatened | 10 |
| No Intervention | 5363 |

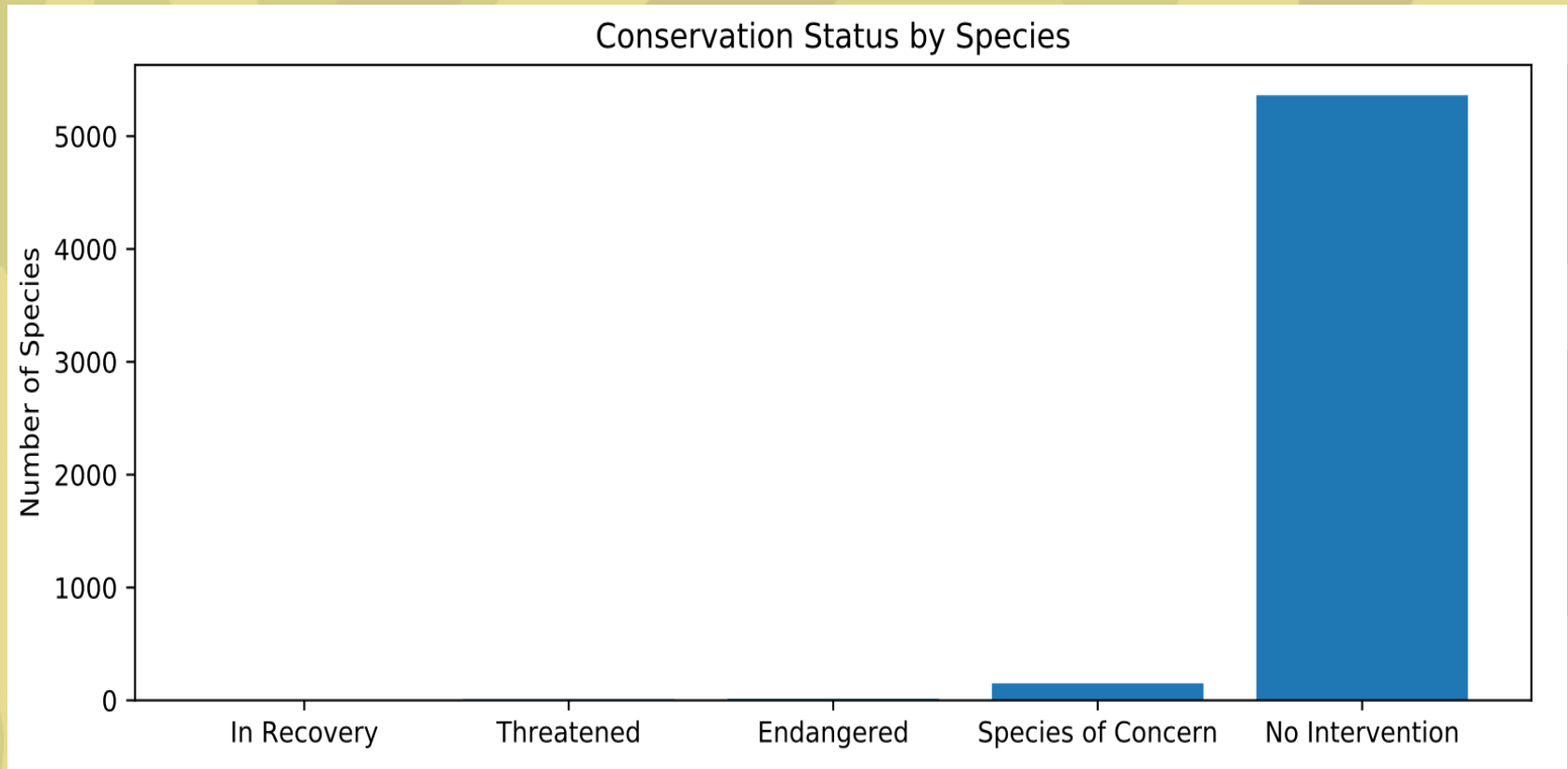


Figure 1: Number of species that belong to each conservation status

Are certain types of species more likely to be endangered?

Table 2: Counts of protected and non-protected animals, and the percentages of protected animals in each category

| Category | Non-Protected | Protected | % Protected |
|-------------------|---------------|-----------|-------------|
| Amphibian | 72 | 7 | 8.8608 |
| Bird | 413 | 75 | 15.3689 |
| Fish | 115 | 11 | 8.7302 |
| Mammal | 146 | 30 | 17.0455 |
| Nonvascular Plant | 328 | 5 | 1.5015 |
| Reptile | 73 | 5 | 6.4103 |
| Vascular Plant | 4216 | 46 | 1.0793 |

Are certain types of species more likely to be endangered?

H_0 : There is no difference between mammals and birds

H_A : Mammals are more likely to be endangered than birds

| | Protected | Not Protected |
|--------|-----------|---------------|
| Mammal | 30 | 146 |
| Bird | 75 | 413 |

P-value ≈ 0.687594809666

The difference between the percentages of protected birds and mammals is **not statistically significant** (i.e. due to random chance)!

H_0 : There is no difference between mammals and reptiles

H_A : Mammals are more likely to be endangered than reptiles

| | Protected | Not Protected |
|---------|-----------|---------------|
| Reptile | 5 | 73 |
| Mammal | 30 | 146 |

P-value ≈ 0.0383555902297

The difference between the percentages of protected reptiles and mammals is **statistically significant!**
Certain types of species are more likely to be endangered!

Recommendation

- The chi-square test of significance showed that certain species are more likely to be endangered than others with a p-value of ~ 0.0388
 - ★ A chi-square test was used because there are two or more categorical datasets to compare
- It is recommended to increase conservation efforts for species that are more likely to become endangered

Tracking Sheep Locations

Table 3: Species of sheep in the dataset, including scientific and common names, and conservation status

| Scientific Name | Common Names | Conservation Status | Protected? |
|-------------------------|---|---------------------|------------|
| Ovis aries | Domestic sheep, Mouflon, Red Sheep, Sheep (Feral) | No Intervention | No |
| Ovis canadensis | Bighorn Sheep | Species of Concern | Yes |
| Ovis canadensis sierrae | Sierra Nevada Bighorn Sheep | Endangered | Yes |

Tracking Sheep Locations

Table 4: Total sheep sightings (across all three species) at each national park over the past 7 days

| Park Name | Observations |
|-------------------------------------|--------------|
| Bryce National Park | 250 |
| Great Smoky Mountains National Park | 149 |
| Yellowstone National Park | 507 |
| Yosemite National Park | 282 |

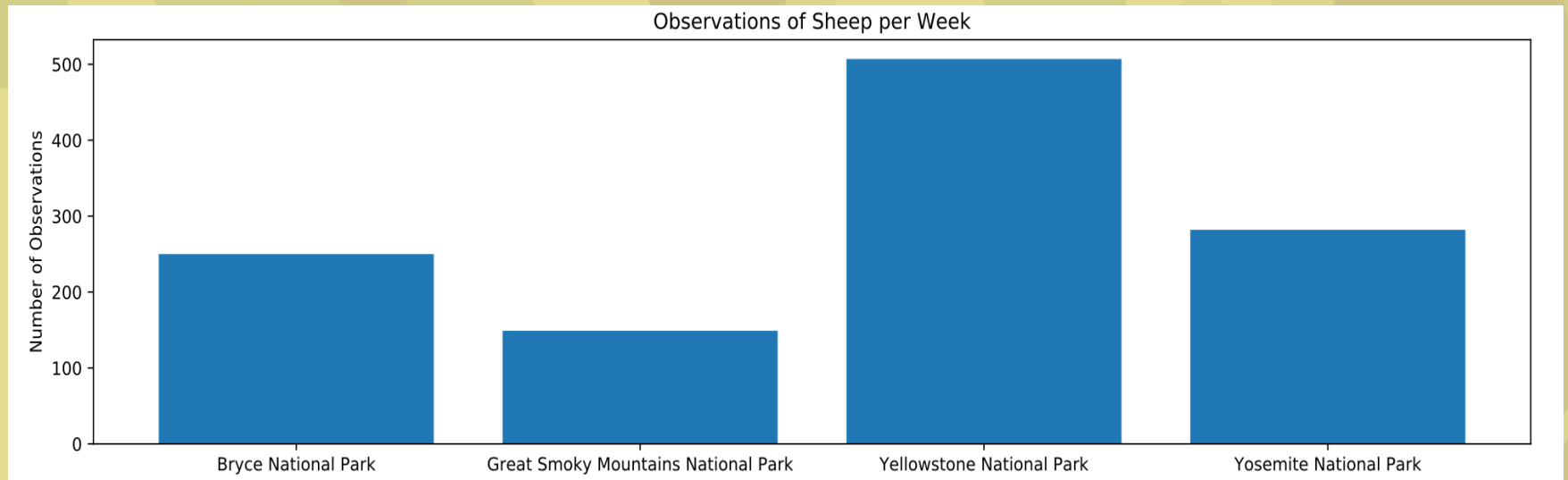


Figure 2: Bar chart of total sheep sightings in each national park over the past 7 days

Foot and Mouth Disease Program

- The goal is to determine whether the foot and mouth disease reduction program at Yellowstone National Park is working
- Last year, 15% of sheep at Bryce National Park had foot and mouth disease
- The park rangers want to be able to detect at least a 5% reduction in foot and mouth disease
- How many sheep should we observe from each park to make sure the program is working?

Sample Size Determination

- Baseline percentage: 15%
- Minimum detectable effect:
 $(100.0 * 5\%) / 15\% \approx 33.3\%$
- Significance level: 90%
- The sample size needed to observe at least a 5% reduction with 90% confidence is **510** (per national park)

Sample Size Determination

- The sample size needed to observe at least a 5% reduction with 90% confidence is **510** (per national park)
- The scientists would need to spend approximately **1 week at Yellowstone National Park** to observe enough sheep
 - $(510 \text{ sheep} / 507 \text{ observed at Yellowstone}) \approx 1.00591716$
- The scientists would need to spend approximately **2 weeks at Bryce National Park** to observe enough sheep
 - $(510 \text{ sheep} / 250 \text{ observed at Bryce}) \approx 2.04$



The End!