

A herd of bison is grazing in a vast, open grassy field. The bison are dark brown with thick, shaggy fur. The field is covered in tall, dry grass. In the background, there are rolling hills under a bright blue sky with scattered white clouds. The text "OUR NATIONAL PARKS BIODIVERSITY" is overlaid in white, sans-serif capital letters.

OUR NATIONAL PARKS BIODIVERSITY

Presented by: Bryan Leung
Codecademy Data Analysis Capstone Project

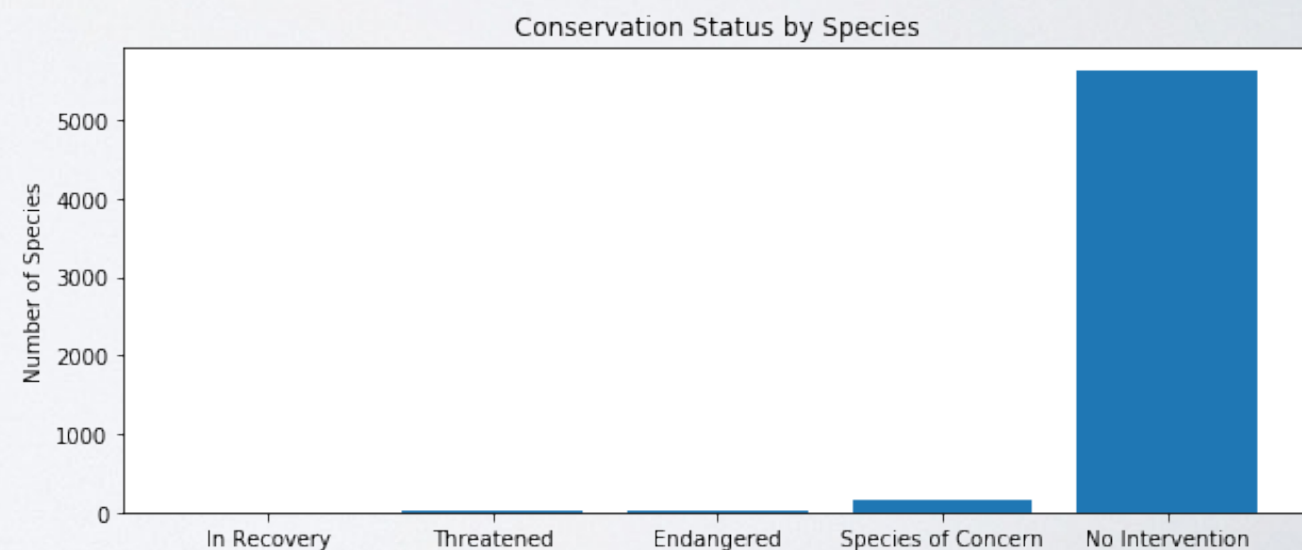
SPECIES INFO

- From *species_info.csv*, we have data on 5541 unique species across our various national parks.
- They are tabulated by *common name*, *scientific name*, *species category*, and *conservation status*.
- There are 7 species categories: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', and 'Nonvascular Plant'.
- There are 5 categories of conservation status: 'No Intervention', 'In Recovery', 'Species of Concern', 'Threatened', and 'Endangered'.

CONSERVATION STATUS BY SPECIES

- We counted the number of species by their conservation status and categorized them from greatest to least concern.
- A majority of our species require no protection (5363) or are in recovery (4).
- However, 151 species may be in need of conservation and 25 species are threatened or endangered.

Conservation Status	Scientific Name
Endangered	15
Threatened	10
Species of Concern	151
In Recovery	4
No Intervention	5363



COMPARING SPECIES BY CATEGORY

- Are certain types of species more likely to be endangered?
- To answer this question, we created a pivot table (next slide) and grouped species by their category and protection status:
 - The '*Protected*' column displays the count of unique species requiring intervention (status \neq '*No Intervention*').
 - The '*Not Protected*' column displays the count of unique species not requiring intervention (status = '*No Intervention*').

SPECIES BY CATEGORY AND PROTECTED STATUS

	CATEGORY	PROTECTED	NOT PROTECTED	PERCENT PROTECTED
1	Amphibian	7	73	8.75%
2	Bird	79	442	15.16%
3	Fish	11	116	8.66%
4	Mammal	38	176	17.76%
5	Nonvascular Plant	5	328	1.50%
6	Reptile	5	74	6.33%
7	Vascular Plant	46	4424	1.03%

COMPARING SPECIES CATEGORIES

- How can we determine if there is a significant difference between two categories of species and their protection status?
- For example, 17% of *Mammals* are protected while 15% of *Birds* are protected. We'd like to know if mammals are more likely to be endangered than birds.
- To compare differences in categorical data (presented with our species categories data), we'll conduct two chi squared tests.

CHI SQUARED TEST #1: COMPARING MAMMALS AND BIRDS

- Null Hypothesis: there is no significant difference between the mammal and bird dataset.
- To reject this hypothesis, we need to look for a p-value of less than 5%.
- Our results came out to a 44.59% margin or error, and we can confidently say there is no significant difference between the protection of birds and mammals.

CHI SQUARED TEST #2: COMPARING MAMMALS AND REPTILES

- Is the difference between protected Reptiles (6%) and protected Mammals (17%) significant?
- Null Hypothesis: there is no significant difference between the mammal and reptile dataset.
- Once again, to reject this hypothesis, we need to look for a p-value of less than 5%.
- With a new contingency table and using the `chi2` contingency function, we received a p-value of 2.33%, less than 5%.
- Therefore, we reject our null hypothesis, and see there IS a significant difference in protected species of reptiles and mammals.

OBSERVATIONS OF PARK ENDANGERED SPECIES

- While there may not be a significant difference between the data of Mammals and Birds, they are trending to be species that need our attention the most for protection.
- We can calculate these results by observing our '*Percentage Protected*' values. The higher they are, then it's likely the species will be threatened and endangered.
- Amphibians and Fish are species we'll need to keep a watch on with Reptiles not far behind. Vascular and Non-Vascular appear to be the least threatened category overall.

SHEEP SIGHTINGS

- Conservationists have been recording sightings of different species of sheep at several national parks for the past seven days. Data is collected from the provided datafile, *observations.csv*.
- We combined data from our species data and filtered the categories to '*Mammal*' and data in '*Common Name*' contains '*Sheep*'.

SHEEP SIGHTINGS

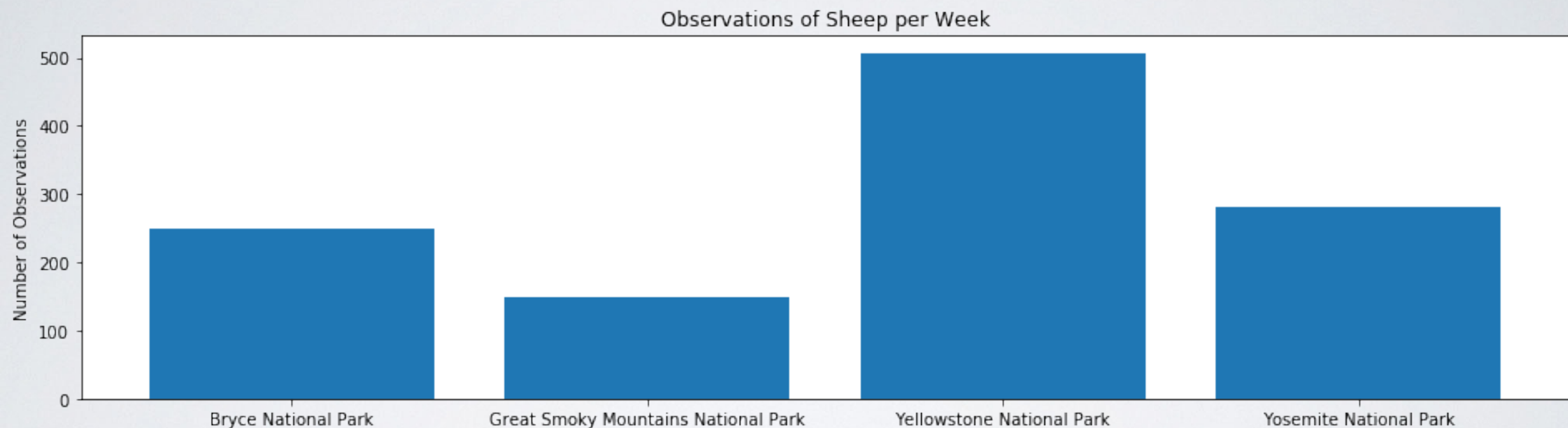
- Our data shows that there are three unique sheep species in our national parks listed:
- Domestic Sheep (*Ovis Aries*)
- Bighorn Sheep (*Ovis Canadensis*)
- Sierra Nevada Bighorn Sheep (*Ovis Canadensis Sierrae*)

SHEEP OBSERVATIONS BY PARK

- We grouped the total sheep sightings (across all three species) from each national park by grouping the sum of observations by each park name.

- | | PARK NAME | OBSERVATIONS |
|---|-------------------------------------|--------------|
| 1 | Bryce National Park | 250 |
| 2 | Great Smoky Mountains National Park | 149 |
| 3 | Yellowstone National Park | 507 |
| 4 | Yosemite National Park | 282 |

SHEEP OBSERVATIONS BY PARK (CONTINUED)



- Yellowstone National Park had the most sheep sightings (507) out of the four national parks collected from our data.

FOOT AND MOUTH DISEASE STUDY AMONGST PARK SHEEP

- Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease.
- Park rangers at Yellowstone National Park are running a program to reduce the rate of foot and mouth disease at that park.
- Our scientists want to know whether or not this program is working by performing an A/B test; they want to be able to detect reductions of at least five percentage points. For instance, if 10% of sheep in Yellowstone have the disease, they'd like to know this with confidence.

FINDING THE SAMPLE SIZE FOR A/B TEST

- We used a sample size calculator from Optimizely to calculate the number of sheep that we need to observe from each park.
- Our data for the calculator:
- Baseline conversion rate: 15%
- Minimum detectable effect: 100×5 percentage points/baseline = 33.33%
- Statistical significance: 90%

SHEEP SAMPLE CONCLUSIONS

- We calculated that a sample size of 510 sheep from each park need to be observed to detect reductions of at least 5 percentage points.
- To calculate how long it would take to observe enough sheep, we divided the sample size per park by its sheep observations.
- Therefore, we need 2 weeks at Bryce National Park ($510/250$) and about 1 week at Yellowstone National Park ($510/507$) to observe enough sheep.