



Biodiversity Capstone Project

Grace Taiwo

Investigating Protected Species

Protected Species



INTRODUCTION TO THE DATA

- When beginning this project, I was given a .csv file called 'species_info.csv' to load into the programme
- The file stored 5541 different species that could found in the national parks involved.
- Below is a sample of what could be found in the .csv table produced when the programme runs:

ID	category	scientific_name	common_name	conservation_status
0	Mammal	Bos Bison	Bison, American Bison	Nan



INTRODUCTION TO THE DATA CONTINUED

The possible categories of the species include:

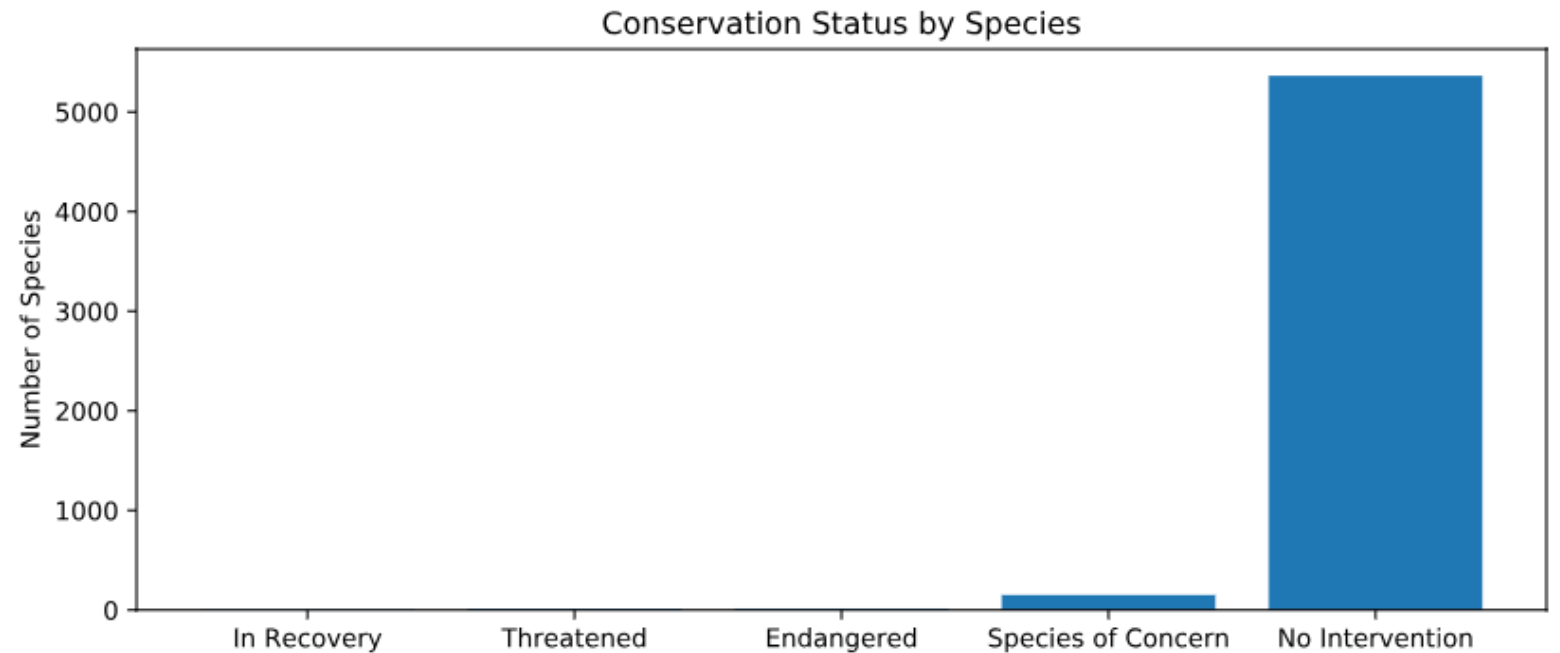
- Mammals
- Birds
- Reptiles
- Amphibians
- Fish
- Vascular Plants
- Non-Vascular Plants

– Possible values for the conservation status of species:

- nan (null)
- Species of Concern
- Endangered
- Threatened
- In Recovery


DATA OBSERVATIONS

The graph adjacent shows the distribution of species and their conservation status. From the data it's pretty clear, that most species need no prevention.



FURTHER OBSERVATIONS

- To fully understand the data, I calculated the percentages of each specie category that were or weren't under a protected status. The results made me question whether there was a relationship between different species' likelihood of it's protection status.



Category	protected	not_protected	percent_protected
Amphibian	7	72	0.088608
Bird	75	413	0.153689
Fish	11	115	0.087302
Mammal	30	146	0.170455
Nonvascular Plants	5	328	0.015015
Reptiles	5	73	0.064103
Vascular Plants	46	4216	0.010793

Value 1 points to the 'protected' column header.

Value 2 points to the 'not_protected' column header.

- As there were two categorical values to be compared, using Chi-Squares to test the significance was the best option



THE SIGNIFICANCE CALCULATIONS

- After creating contingency tables for each test, I plugged the table into an expression which gave me these results:
 - When looking at **Birds** and **Mammals**, the difference between the number of protected birds and mammals was **insignificant** (0.688)
 - Suggesting the null hypothesis could be correct as the difference was a result of chance
 - **Reptiles** and **Mammals** on the other hand, did have a **significant** relationship (0.038)
 - We can therefore conclude that certain categories of species are more likely to be endangered than others.



MY RECOMMENDATIONS

- Based on the calculations made, I know that some categories of animals are more likely to have a protected status compared to others
- This suggests that further action must be taken to fully understand the reasons for this (more research and analysis), to find the causes and provide a plan on how to resolve the issues at hand.

Investigating Protected Species

Foot and Mouth Reduction Effort



DISEASEINTRODUCTION TO THE DATA

- For the second section, I was given another .csv file called ‘observations.csv’ to load into the programme
- The file stored where they’ve been spotted and the amount of observations of the species that week:

ID	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68



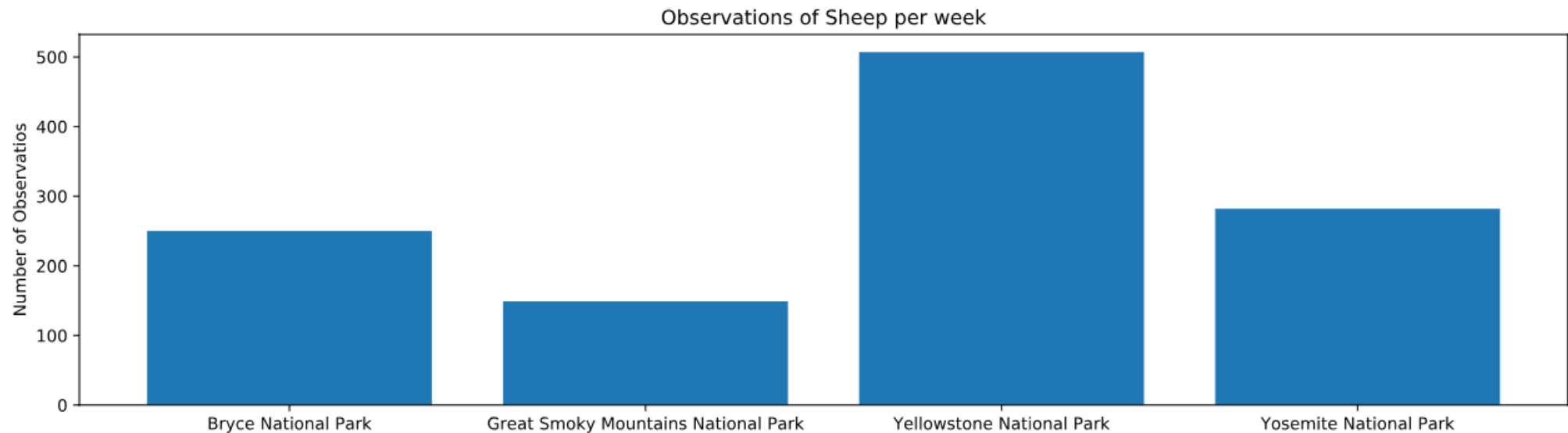
CREATION OF SHEEP TABLES

- Because of the focus requested on sheep in the parks, a new table was created with data from both .csv files for just the sheep.
- Below is the result of the two merged tables (species and observations)

ID	category	scientific_name	common_name	conservation_status	is_protected	park_name	observations
0	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	Yosemite National Park	126
1	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	Great Smoky Mountains National Park	76
2	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	Bryce National Park	119
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	Yellowstone National Park	221
4	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	Yellowstone National Park	219

SHEEP OBSERVATION GRAPH

- Since four of the parks were under investigation, I created a bar chart to show the distribution of sheep in each park.





SAMPLE SIZE DETERMINATION – FOOT AND MOUTH

- As the Park Rangers wanted to reduce the amount of Foot and Mouth disease, I calculated a few things to get them on their way
- After working out the baseline percentage and the minimum detectable effect, I was able to work out the sample size of each variant should be **890**.
- This meant it would take the following amount of weeks for each park to collect the amount of sheep needed to make sure their foot and mouth percentages were significant

Park Name	Weeks
Bryce National Park	3.56
Great Smoky Mountain National Park	5.97
Yellowstone National Park	1.75
Yosemite National Park	3.15

Thank you for your time

