Biodiversity in National Parks

Capstone project for Introduction for Data Analysis

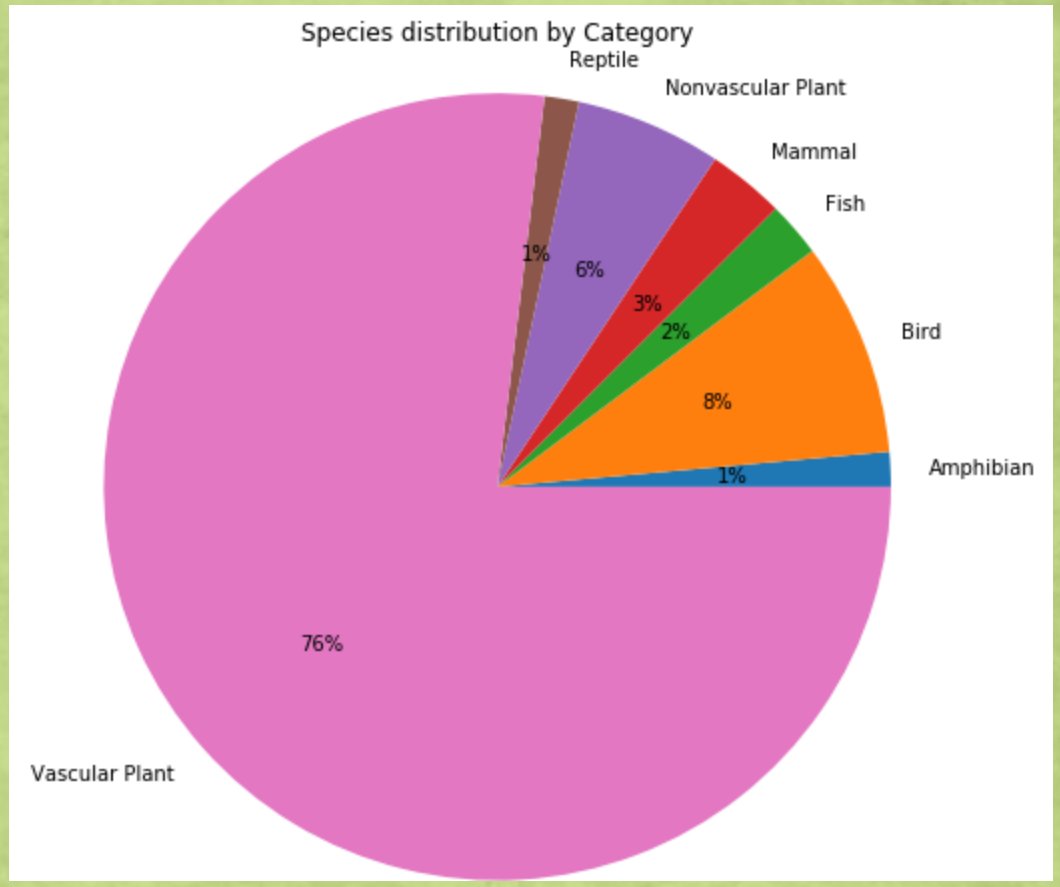
Introduction

This presentation aims to present data analysis on:

- Conservation status of species in numerous National Parks
- Investigate for patterns to the types of species that are endangered
- Assist with determining the required sample size of observations of sheep for foot and mouth program

Species Data – initial observations

- The species data provided has information on the name (common and scientific), Category (e.g. Mammal, Bird, Reptile etc.) and the Conservation Status of the species.
- There are 5541 different species in 7 Categories



Species Data – initial observations

There are 4 populated Conservation Statuses

- Species of Concern
- Threatened
- Endangered
- In Recovery

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	16
3	Species of Concern	161
2	No Intervention	5633

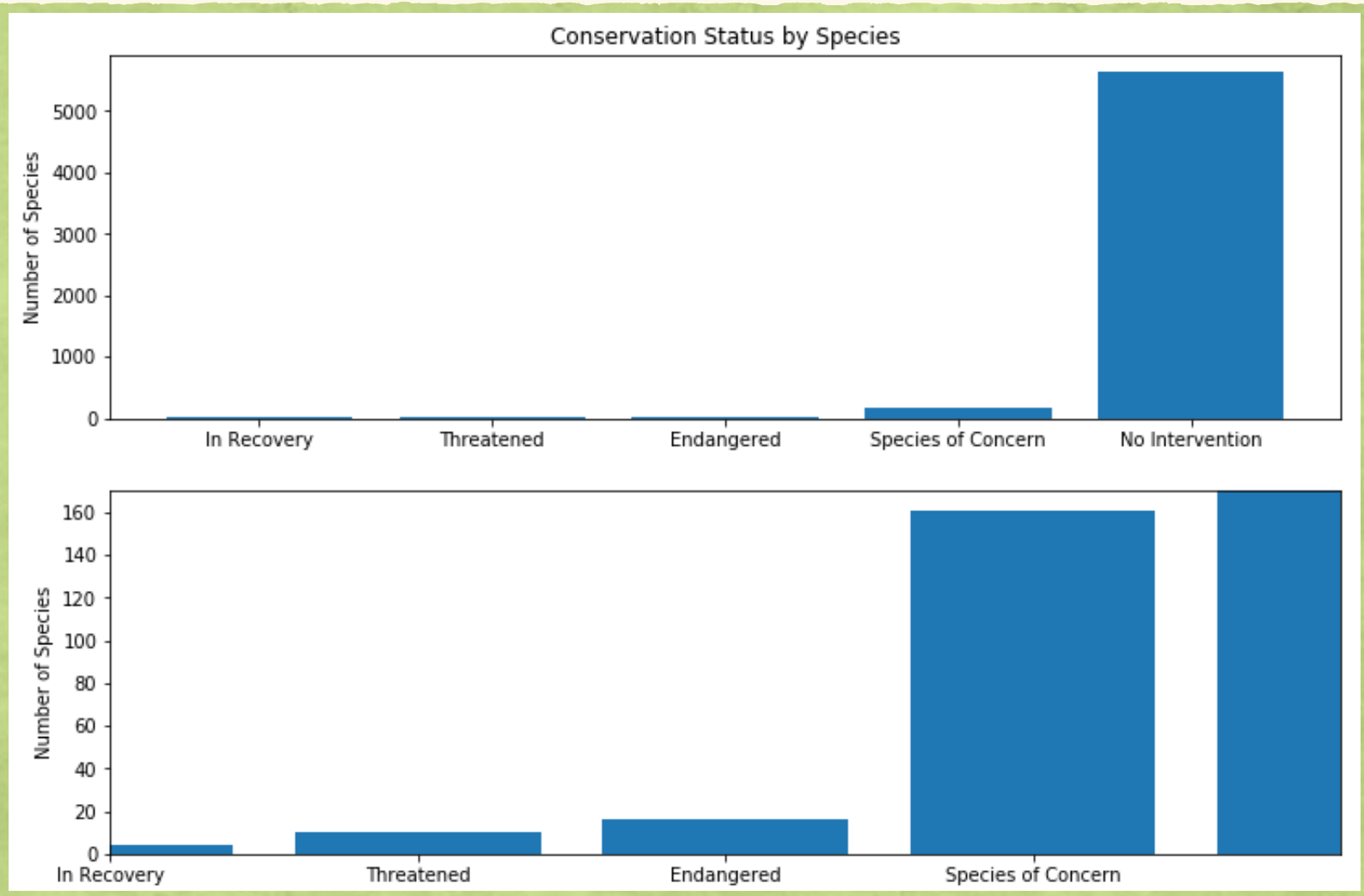
- 5633 records in the data have a null value for Conservation Status

As shown in the table above, in this analysis an assumption is made that NULL in this instance means that no intervention is required.

Species Data – Conservation Status

This bar chart shows the full distribution of species across all Conservation Statuses. Due t the high number of species in the No Interventions category it is difficult to discern the differences in the smaller categories.

This bar chart restricts the range of the Y axis to 160 to show the differences in the smaller categories more clearly.



Are some types of species more likely to be endangered?

To get the table to the right

- *the data was categorized as protected or not protected based on the values in the conservation status field*
- *It was grouped by category and protection status*
- *Finally, a percentage of the species that is protected in each category was calculated*

The lower the percentage here the less likely species in that category are endangered.

So it birds (15.3%) may be less endangered than mammals (17%).
Similarly, reptiles (6.4%) may be even less likely to be endangered than mammals.

In order to test these theories chi-squared tests were used as there are more than two categorical datasets to compare.

	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

Chi Square test results (using a significance threshold of 0.05)

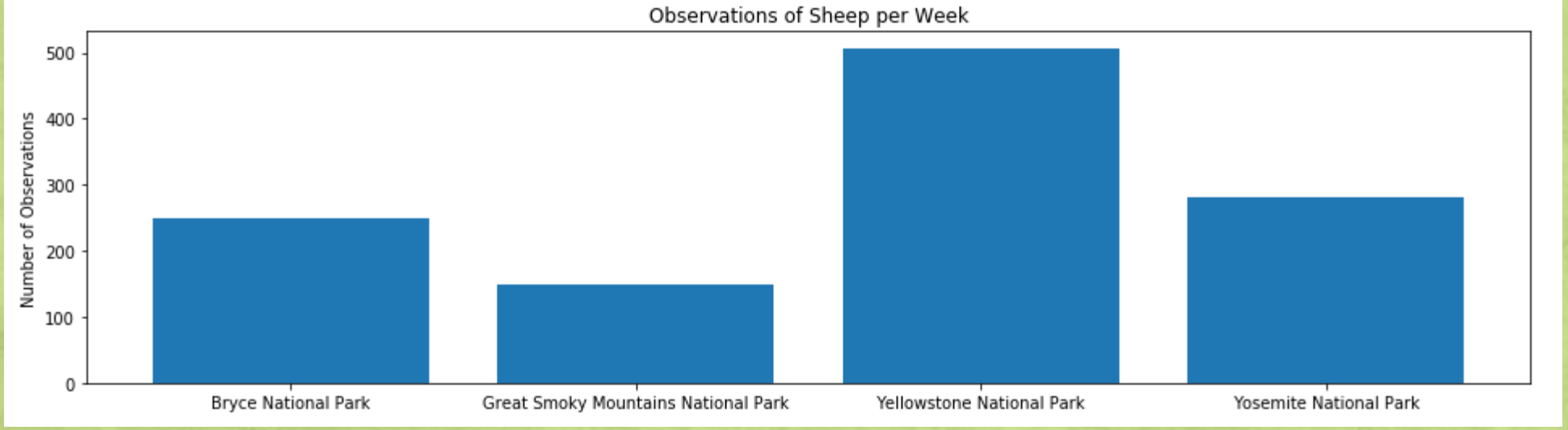
Birds and Mammals gave a p value of 0.6876 determining that the difference is not significant.

Mammals and reptiles however gave a p value of 0.0384 indicating a significant difference.

Species of mammal are more likely to be endangered than species of reptile, perhaps more effort should be concentrated on mammal conservation.

Sample size determination for Foot and mouth disease study in sheep

- Observations data gathered over the past 7 days from 4 National Parks was provided
- As scientists are interested in performing sheep studies, the sightings of mammals with the term 'sheep' in their name were extracted and counted for each National Park



Sample size selection

- 15% of sheep at Bryce National Park have Foot and Mouth disease (Baseline)
- Yellowstone National Park are running a program to reduce foot and mouth and want to be able to detect reductions of at least 5 percentage point (reduction to detect)
- Minimum detectable effect is

100 * (detect_reduction/baseline)

100 * (0. 05/0.15) = 33.33

- 15% Baseline conversion rate
- 33.33 minimum detectable effect
- 90% Statistical Significance

Required Sample Size = 510

Based on current data it will take just over 2 weeks for Bryce And just over 1 week for Yellowstone to reach this sample size

	park_name	observations	weeks_to_sample510
0	Bryce National Park	250	2.040000
1	Great Smoky Mountains National Park	149	3.422819
2	Yellowstone National Park	507	1.005917
3	Yosemite National Park	282	1.808511