Biodiversity for the National Parks

Capstone Project

Tim Ezzy - August 2018

PART ONE: ENDANGERED STATUS OF ANIMAL SPECIES

Background: National Parks Service maintains a register of different species in a number of National Parks.
This register (Species_info.csv) contains information on:

- The Scientific Name of each species
- The common name of each species
- The species conservation status

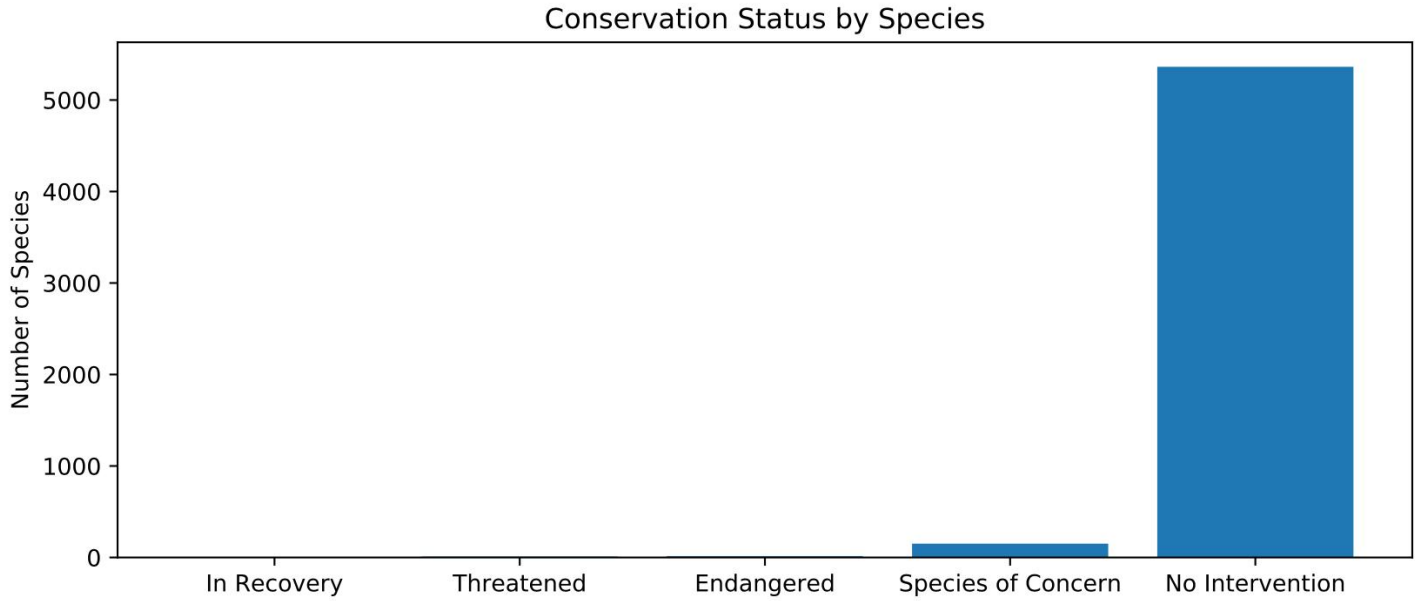
Analysis of the register showed that:

- There are 5541 individual species
- There are seven species types: Mammal; Bird; Reptile; Amphibian; Fish; Vascular Plant; Nonvascular Plant
- Registered conservation statuses:

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	Species of Concern	151
3	Threatened	10

Where conservation status was blank, a status of “no intervention” was applied. The updated conservation status:

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10



Research Question 1: Are certain types of species likely to be more endangered?

Methodology: Apply a significance test to compare different species and whether they are likely to be more endangered.

What significance test? We know the data are categorical and there are two or more samples, so the ***Chi Squared test*** was selected

Data: The species dataset was analysed to determine the number of protected and non-protected species per category of animal. The results of this data analysis are presented in Table 1.

Table 1: Protected Status of Animal Species in National Parks

Category	Not Protected Status	Protected Status	Percent Protected
Amphibian	72	7	0.088
Bird	413	75	0.153
Fish	115	11	0.087
Mammal	146	30	0.170
Nonvascular plant	328	5	0.015
Reptile	73	5	0.064
Vascular plant	4216	46	0.010

Results:

1. A Chi Squared test comparing whether mammals are likely to be more endangered than birds produced a p value of 0.688 which is **not significant**. Therefore our null hypothesis “that the difference in endangered status of these two species is by chance” holds true.

2. A second Chi Squared test was undertaken comparing reptiles and mammals. This test returned a p value of 0.038, which **is significant.**

Conclusion:

The research question of whether some species are likely to be more endangered than others can be answered as **“yes, some species are more endangered than others”**

Recommendation:

Now that we have rejected our null hypothesis, and we know with confidence that some species are more endangered than others, we need to undertake a more rigorous statistical analysis of all remaining species to evaluate which require increased conservation efforts.

PART TWO: ANALYSIS OF SHEEP OBSERVATIONS WITHIN NATIONAL PARKS TO PROVIDE INSIGHTS INTO FOOT AND MOUTH RESEARCH TRIALS

Background: Conservationists have recorded observations of particular species at several National Parks over the past 7 days. This data has been compiled into a database. Analysis of this database has the potential to provide important insights for our research into sheep foot and mouth disease.

The data are stored with the animal's scientific name, the name of the park where it was sighted and the actual number of observations over the 7 day period.

The table below provides a few records to illustrate the data recorded

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85
5	Elymus virginicus var. virginicus	Yosemite National Park	112

One of the challenges with this observation data was that only the scientific names were recorded. As we are interested in sheep we needed to analyse the common name of each species. To do this the observations database was merged with the National Parks species database, so that each instance of scientific name in the observations database now included in a common name.

Searching for all instances of “sheep“ in the merged database provided the following table. Note that some plant species have the name “sheep” so the data had to be further grouped to only include mammals.

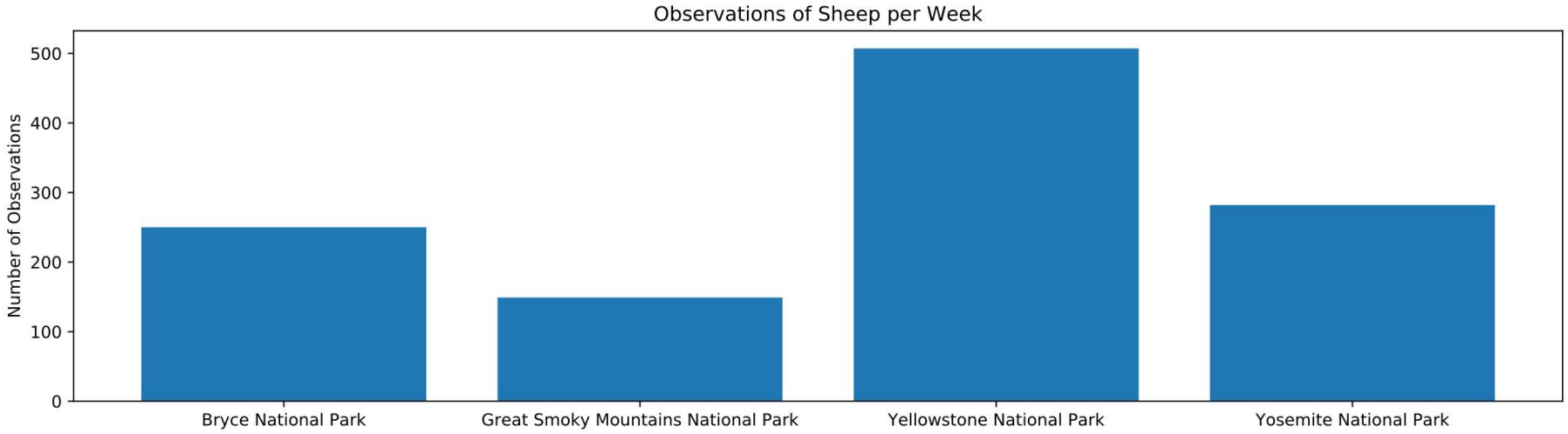
	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
1139	Vascular Plant	Rumex acetosella	Sheep Sorrel, Sheep Sorrell	No Intervention	False	True
2233	Vascular Plant	Festuca filiformis	Fineleaf Sheep Fescue	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3758	Vascular Plant	Rumex acetosella	Common Sheep Sorrel, Field Sorrel, Red Sorrel, Sheep Sorrel	No Intervention	False	True
3761	Vascular Plant	Rumex paucifolius	Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock	No Intervention	False	True

Further analysis, including grouping the sheep observation data by park name and a count of the number of observations at each park revealed the following dataset:

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

This dataset provides us with critical information on the numbers of sheep we can expect to observe at each park on a given week.

The sheep observation data has been plotted on a bar chart below to allow for easy comparison between each park. It is evident that **Yellowstone National Park** has the highest number of sheep observations in the trial week.



Research Question Two: How many sheep do we need to observe to statistically prove a reduction in foot and mouth disease.

Research Question Three: Which National Park is the best site for the NPS foot and mouth trial?

Methodology: Use a sample size calculator to calculate the minimum number of sheep observations required to detect a >5% reduction in foot and mouth disease with 90% confidence

Data:

1. Baseline conversion rate = 15% of sheep have foot and mouth disease
2. Minimum detectable effect = 5% / baseline conversion rate
   = 33.33%
3. Statistical Significance = 90% confidence

Results:

1. Sample size required = 870 sheep to detect at least a 5% reduction in foot and mouth

2. Time to survey at Yellowstone = 870 sheep / 507 observed per week
   = 1.72 weeks
3. Time to survey at Bryce = 870 sheep / 250 observed per week
   = 3.48 weeks

Conclusions:

1. 870 sheep would need to be observed to detect at least a 5% reduction in foot and mouth disease
2. Yellowstone National Park has the highest observations of sheep per week and a sufficient sample size could be obtained in < 2 weeks

Recommendation:

Undertake a two week field campaign at Yellowstone National Park

Note: completed code in the learning environment