



CAPSTONE 2: BIODIVERSITY PROJECT

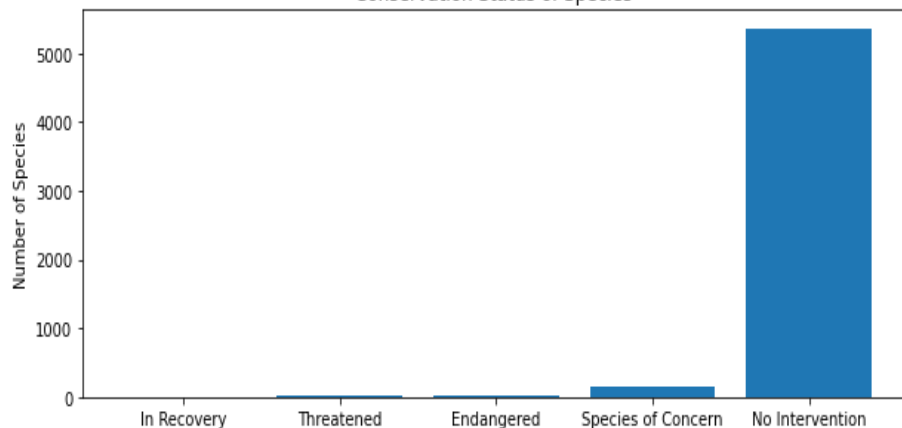
Judy Ping McCormick

Python, Pandas, Matplotlib, Scipy,
chi2_contingency

16th Jan 2021

category	scientific_name
Vascular Plant	4262
Bird	488
Nonvascular Plant	333
Mammal	176
Fish	125
Amphibian	79
Reptile	78

Conservation Status of Species



1. INTRODUCTION

Biodiversity analysis about species at various national parks

- Analysis Based on 2 Original CSV DataSets: species & observations
- 5541 different types of species with 7 categories
- Majority needs no intervention

2. PERCENTAGE OF PROTECTED SPECIES

- 17% of Mammals are protected
- Just over 1% of Vascular Plant have protection

category	not_protected	protected	percent_protected
Mammal	146	30	17.05
Bird	413	75	15.37
Amphibian	72	7	8.86
Fish	115	11	8.73
Reptile	73	5	6.41
Nonvascular Plant	328	5	1.5
Vascular Plant	4216	46	1.08

	protected	Not_protected
Mammal	30	146
Birds	75	413

3. CHI SQUARE CONTINGENCY TEST

Mammal & Bird

- Statement I:

Species in category `Mammal` are more likely to be endangered than species in `Bird`.

- Conclusion I:

p-value=0.687595, significance=0.05

At 0.05 level of significance, we accept the null hypotheses. They are independent.

It looks like this difference isn't significant!

Mammal & Reptile

- Statement II:

Species in category `Mammal` are more likely to be endangered than species in `Reptile`.

- Conclusion II:

p-value=0.038356, significance=0.05

At 0.05 level of significance, we reject the null hypotheses and accept H1. They are not independent.

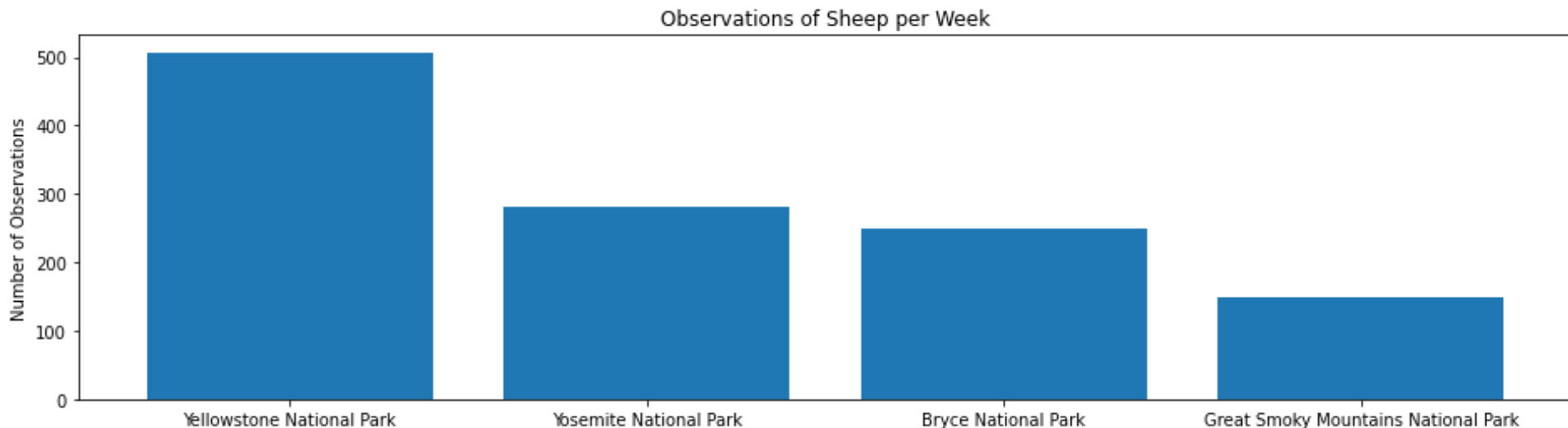
Yes! It looks like there is a significant difference between `Reptile` and `Mammal`!

	protected	not protected
Reptile	5	73
Mammal	30	146

4. SHEEP OBSERVATIONS AT NATIONAL PARKS

Based on the original data set 'observations.csv'

- Mammals that have 'sheep' in its 'common-names' are classified as 'sheep'
- Yellowstone National Park has the highest number of 507 observations;
- Great Smoky Mountains National Park has only 149 observations per week



5. FOOT & MOUTH DISEASE SAMPLE SIZE

Situation:

- 15% of sheep at Bryce National Park have foot and mouth disease
- Yellowstone National Park want to be able to detect reductions of at least 5 percentage point
- With 90% significance level

Suggestion:

- Minimum Detectable Effect is 33.33
- With Sample size of 870
- Bryce National Park needs approx.3 weeks, while Yellowstone National Parke needs about 2 weeks, in order to observe enough sheep.



EXECUTIVE SUMMARY

Biodiversity analysis about species at various national parks

- 5541 different types of species with 7 categories
- 17% of Mammals are protected, while Just over 1% of Vascular Plant have protection
- Species in category `Mammal` are more likely to be endangered than species in `Bird`.
- Species in category `Mammal` are not more likely to be endangered than species in `Reptile`.
- Yellowstone National Park has the highest number of 507 sheep observations;
- Great Smoky Mountains National Park has only 149 sheep observations per week
- Programs aim to reduce 'Foot & Mouth' disease, Bryce National Park needs approx.3 weeks, while Yellowstone National Parke needs about 2 weeks, in order to observe enough sheep.