Analysis of Species at US National Parks

CODECADEMY INTRODUCTION TO DATA ANALYSIS INTENSIVE

'BIODIVERSITY' CAPSTONE PROJECT SUMMARY

BY PAUL CARBONI

Species Data

- The data we worked with (from species_info.csv) included four columns: Category of species, scientific name, common name(s), and conservation status.
- This data included information on 5,541 unique species.
- Species were categorized as being mammals, birds, reptiles, amphibians, fish, vascular plants, or nonvascular plants.
- The included conservation statuses were species of concern, endangered, threatened, in recovery.
- The vast majority (over 95%) of species in the CSV did not have a status, indicating that these species exhibit healthy populations.

Relative Likelihood of Endangered Status between Categories of Species

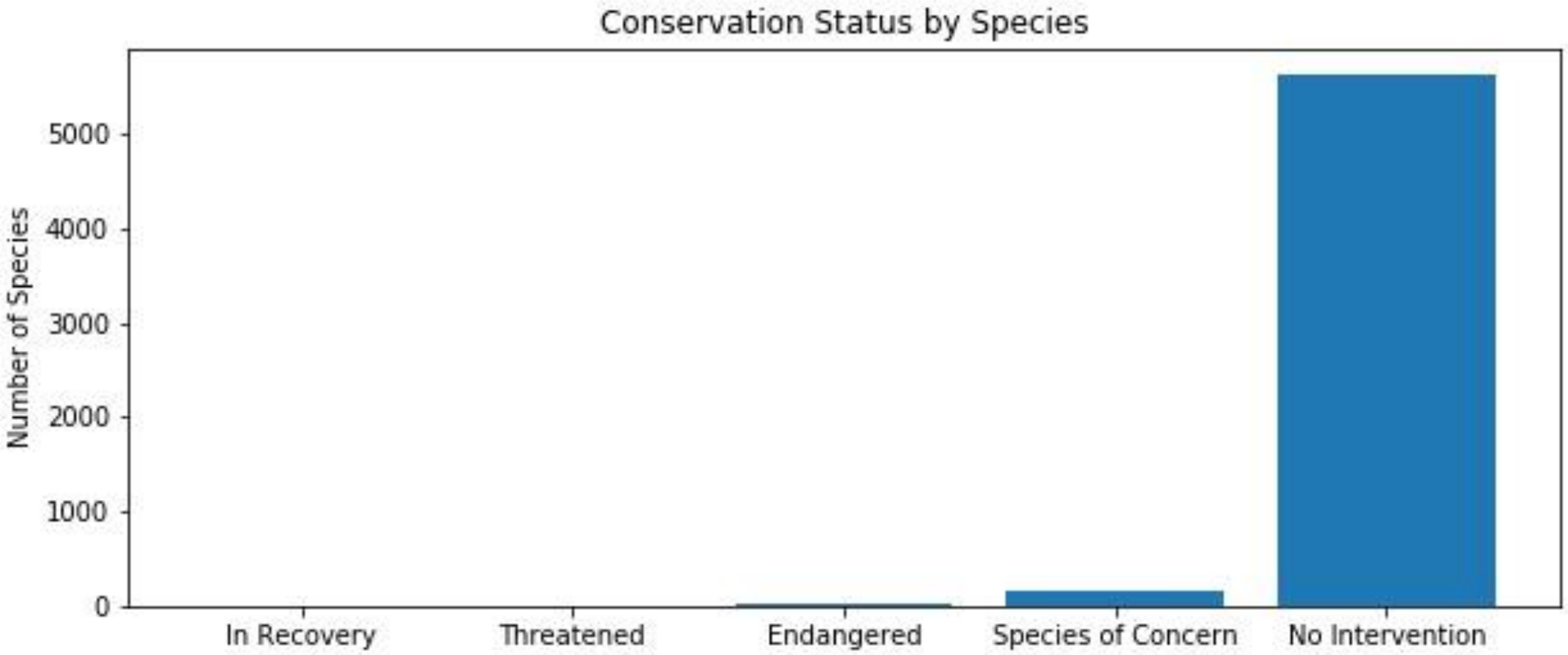
- We wanted to determine if some categories of species were significantly more likely to have higher numbers of protected species.
- The data we compared was categorical (protected or not protected) and we compared two categories of species. Because of this, a chi squared test was the appropriate hypothesis test to use.
- We first compared mammals and birds, but found that the test returned a p-value higher than 0.05. We were unable to reject the null hypothesis.
- In our second comparison however, we found that mammals are signifcantly more likely to be endangered than reptiles due to the chi squared test returning a p-value less than 0.05 when comparing these categories

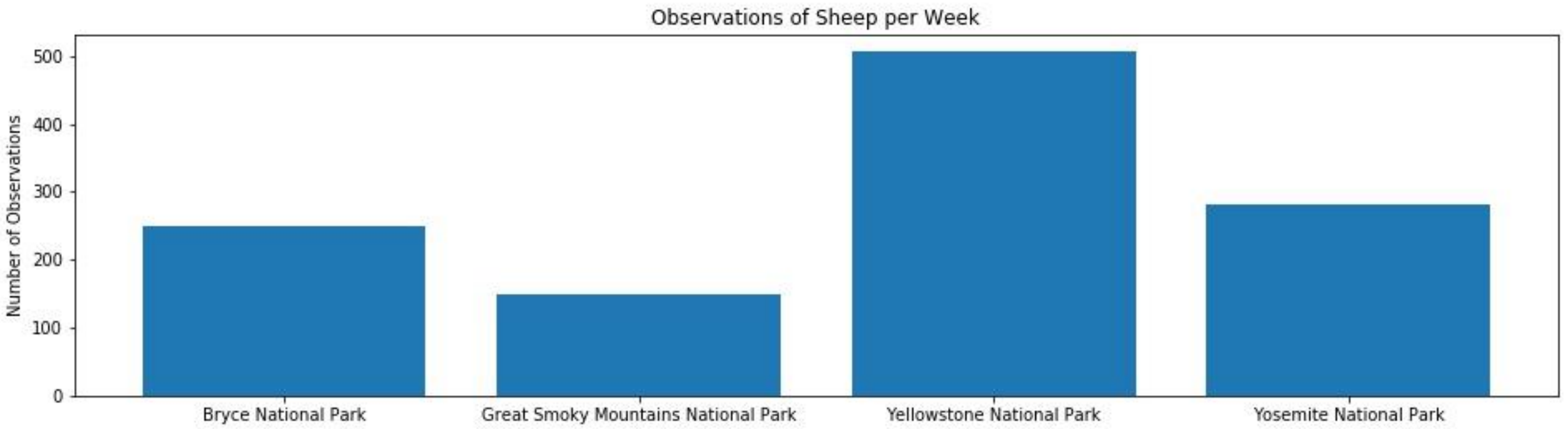
Recommendations to Conservationists

- Based on our findings we recommend that when deciding on whether to commit resources to protecting mammal or bird species, conservationists should make decisions based on factors other than category. Project cost, duration, and potential benefits should be weighted more heavily than a consideration of whether one category of animal is more likely to be endangered.
- However, when deciding between mammal and reptile species, we recommend prioritizing mammal species preservation projects due to the increased likelihood of mammal species being or becoming endangered.
- Additionally, conservationists may consider a study into the factors that lead mammals to be more likely to face endangered status.

Foot and Mouth Disease Sample Size Determination

- To determine the sample size needed for the study on new programs to prevent foot and mouth disease in sheep species at national parks, we started with data from Bryce National Park about the rate of foot and mouth disease among their sheep: 15%. This is our baseline conversion rate.
- Park rangers told us they wished to reduce the rate of foot and mouth disease by at least five percentage points. Dividing five by fifteen yields a minimum detectable difference of 33%.
- Using a 90% level of statistical significance, we ran these numbers through a sample size calculator to arrive at a sample size of 520.


Conservation Status by Species
Number of Species
5000
4000
3000
2000
1000
0
In Recovery
Threatened
Endangered
Species of Concern
No Intervention


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park