Biodiversity for National Parks

Analysis of Observations and Species

Presented by: U.Ujam

Executive Summary

- Two data sets were analysed
  - Endangered Species
  - Sheep – Foot and Mouth Reduction
- Some living organisms appear to be more protected than others. The data shows Mammals are better protected than plants
- Some initial observations could be explained by chance. But we found certain species are more likely to be endangered than others.
- The sample size required to determine a reduction in Foot and Mouth amongst sheep was 870 (510)

Endangered Species

- Mammals 17.8% appear to be the most protected species followed closely by Birds.
- Least protected animals are Vascular Plants at 1.03%. Plants are not protected very much.

- Are certain types of species more likely to be endangered?

- We initially saw that there was a slight difference in the percentages of birds and mammals that fall into a protected category. Our null hypothesis here is that this difference was a result of chance.

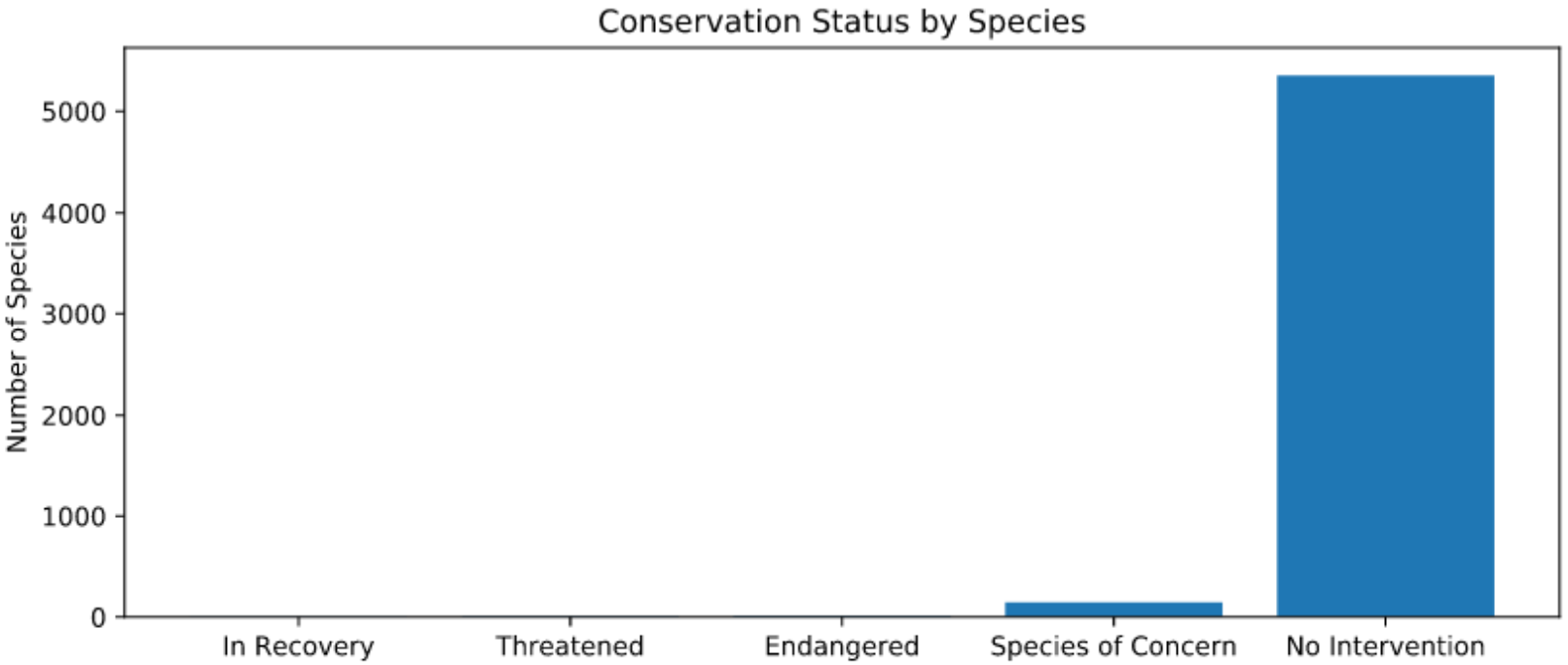
Endangered Species

- When we ran our chi-squared test, we found a p-value of ~0.688, so we can conclude that the difference between the percentages of protected birds and mammals is not significant and is a result of chance.

- But, when we compared the percentages of protected reptiles and mammals and ran the same chi-squared test, we calculated a p-value of ~0.038, which is significant.

- Therefore, we can conclude that certain types of species are more likely to be endangered than others.

Conservation Status by Species



species_info.csv

- Species_info.csv consisted of four main columns:
  - Category
  - scientific_name
  - common_names
  - conservation_status
- Conservation Status provided information as to the likelihood of a species going extinct

species_info.csv

- The entire dataset had to be cleaned up. Especially where conservation_status cells were left empty or null
- Majority (5363) of the species required no intervention or protection
- Birds were the most protected, by number. However, the chi square analysis showed this was not statistically significant when compared to Mammal.

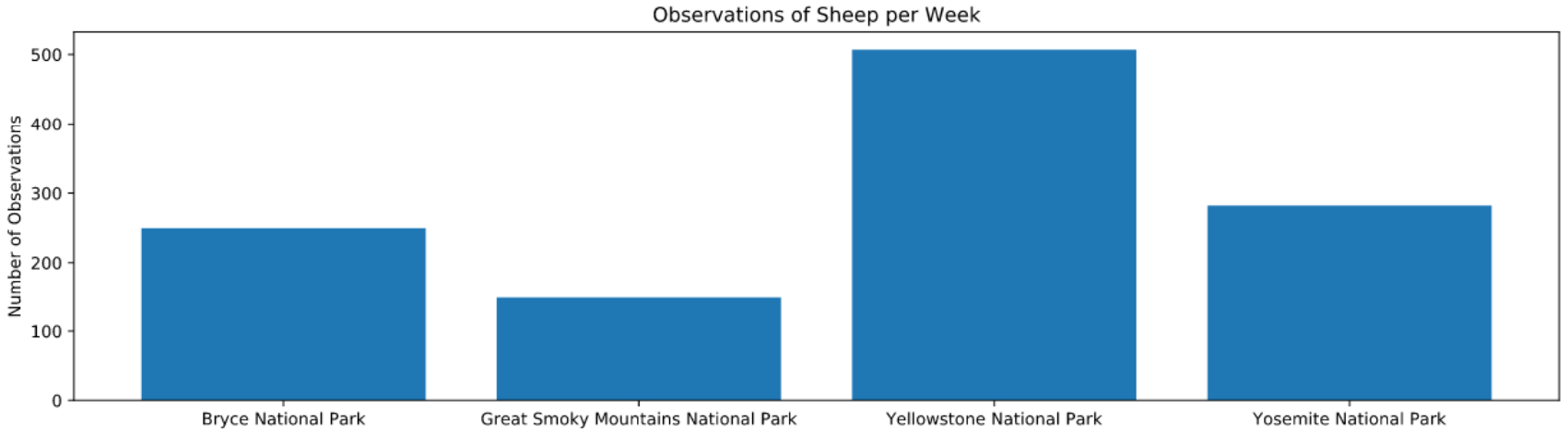
Observation

- National Parks Observation exercise for the reduction of the occurrences of Foot and Mouth disease in sheep.
- For this exercise data was provided in csv format – observations.csv
- To glean insights into the data, the comma separated information had to be filtered for just sheep

Observations

- Observations data was merged with species data to generate more accurate dataset of sheep varieties and sightings across various National Parks.
- The baseline for the previous year’s reduction in Foot and Mouth was 15%
- As a result, 15% was used as the baseline for this data set

Observations of Sheep per Week



Observations

- The National Park wanted to detect reductions of at 5 percentage points
- MDE – Minimum Detectable Effect (or Lift) had to be calculated in order to determine the Sample Size.
- MDE formula→ 100 * ((0.2 – 0.15) / 0.15) which yields a value of 33.3%
- Sample size was a bit of a challenge with the codecademy site giving 870 whilst optimizely gives 510

Observations – A/B Testing

- The upshot? The number of weeks the scientists will have to observe sheep in the Yellowstone National Park was 1 week depending on the sample size used
- Sheep in Bryce National Park will have to be observed for 2 weeks approximately (or 870 / 250 = 3.48 weeks)