United States Park Service Database:
An Investigation into Species Conservation and Ruminate Hoof and Mouth Disease

Thomas Walsh

Codecademy Capstone Project

June 7, 2018

A Glance into the Data

- species_info.csv

  – Database provided by the United States National Park Service

  – Contains information on plants and animals found throughout the various parks

  – Information includes:

    - Scientific Name

    - Common Name(s)

    - Conservation Status

  – Additionally, each organism is grouped into a category based on its classification (i.e. reptile, amphibian, bird, etc.)

Conservation Data from speces_info.csv

Category	Total No. Species	No. Protected Species	Percent Protected
Reptiles	78	5	6.4%
Mammals	176	30	17.0%
Birds	488	75	15.4%
Fish	126	11	8.7%
Amphibians	79	7	8.9%
Vascular Plants	4262	46	1.1%
Nonvasc. Plants	333	5	1.5%
Total Species	5542	179	3.3%

- A higher percentage of animal species have protected status than plants [13.5% to 1.1%]

- Among animals, birds and mammals have the highest likelihood of needing a protected status [15.4% and 17.0% respectively]

Overall Species Conservation Status

Conservation Status by Species
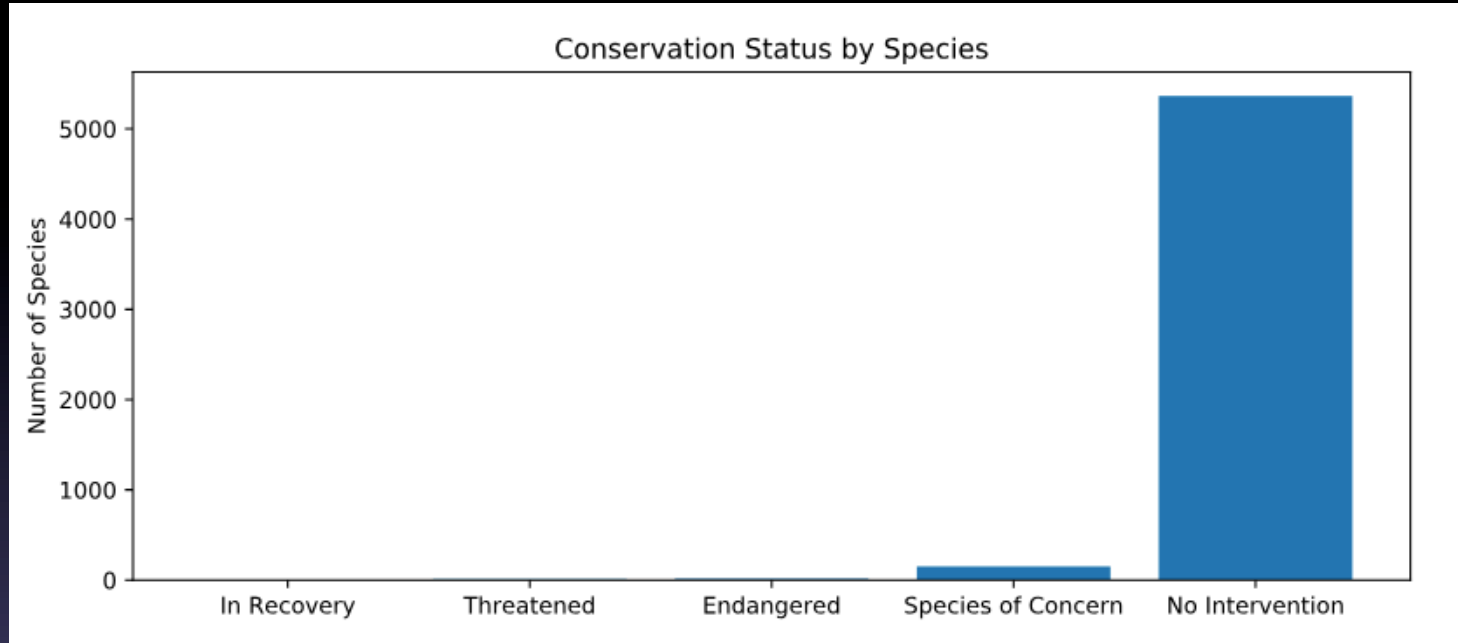
Total Species	No Intervention	Concern	Endangered	Threatened	In Recovery
5542	5363	151	15	10	4

Are Some Categories More Likely to be Endangered?

- Based on the conservation status data from species_info.csv, it appears that certain categories (birds and mammals in particular) are more likely to be endangered

- Are these trends statistically significant and should they be used to formulate US Park Service conservation strategies going forward?

Mammals vs. Birds

	Protected	Non Protected
Mammals	30	146
Birds	75	413

- Percentage of bird species with a protected status = 15.4%
- Percentage of mammal species with a protected status = 17.0%
- Using the above contingency table and Chi-Squared Test we can determine significance
- Null Hypothesis is that the difference between birds and mammals is due to chance.
- P Value of Birds vs Mammals = 0.688
- We CANNOT reject the null hypothesis, the difference in protected status percentage between birds and mammals is a result of chance

Mammals vs. Reptiles

	Protected	Not Protected
Mammals	30	146
Reptiles	5	73

- Percentage of reptile species with a protected status = 6.4%
- Percentage of mammal species with a protected status = 17.0%
- Using the above contingency table and Chi-Squared Test we can determine significance
- Null Hypothesis is that the difference between reptiles and mammals is due to chance.
- P Value of Birds vs Mammals = 0.038
- We CAN reject the null hypothesis, the difference in protected status percentage between reptiles and mammals is not simply due to chance

Conservation Analysis Final Conclusions

- Certain categories of species are more likely to be endangered than others
  - Animals are impacted more heavily than plants
  - Birds and Mammals are impacted more heavily than fish, reptiles and amphibians

- With this knowledge in mind, conservationists need to focus efforts on understanding how animals (particularly birds and mammals) are being differentially impacted and search for methods to reduce the stress on these populations.

Ruminate Hoof and Mouth Disease and Sample Size Analysis

- In addition to the DataFrame species_info.csv, a DataFrame containing information on observations of various animals in the US State Parks was provided.

- The DataFrame observation.csv contains:
  - Scientific Name
  - State Park where observations occurred
  - Number of observations

- These two DataFrames were then joined to allow for searching of animals with sheep as a common name while retaining Park locations and number of observations

Sheep Observations by State Park
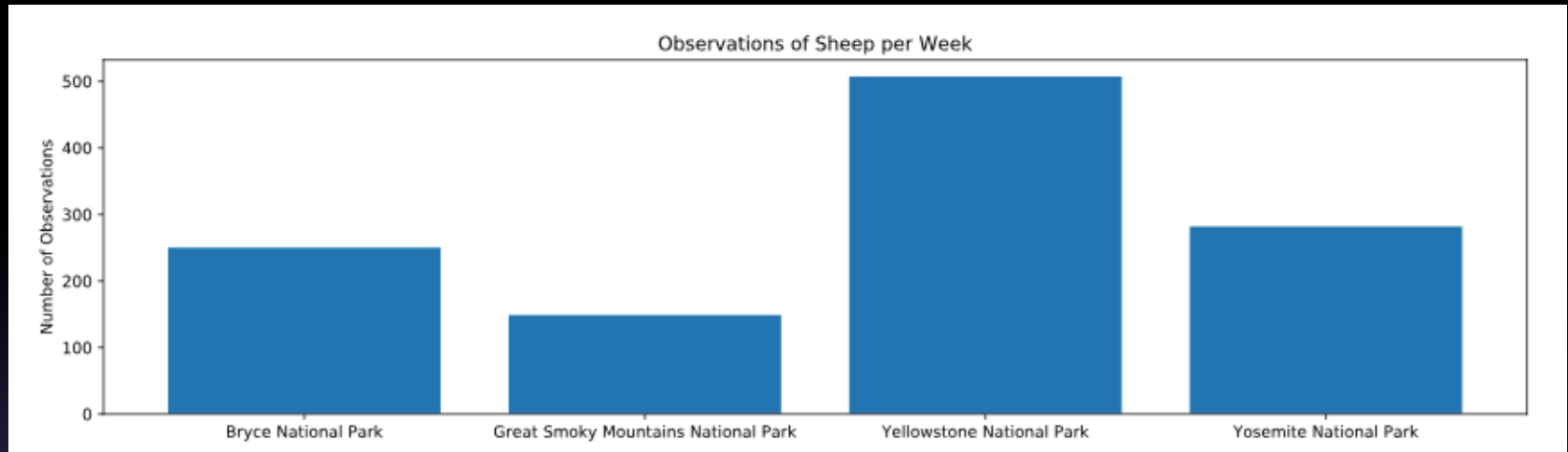
- Using the joined DataFrames (species_info.csv and observations.csv) we can generate a plot of sheep observations per week for the four national parks shown above

- This observation per week data can then be used to in conjunction with sample size calculations to determine the approximate number of weeks of observation is needed in each park

Sample Size Determination

- Goal: Scientists want to accurately test whether an on-going program to reduce hoof and mouth disease is working. They wish to be able to detect reductions of at least 5% compared to previous years data

- Baseline Conversion Rate = 15%

- Statistical Significance = 90%

- Minimum Detectable Effect = [100 * 5/15] OR 33.3%

- Using these inputs, the sample size needed to detect 5% reductions in hoof and mouth disease is 870 observations

Sheep Observation Schedules

- Bryce National Park
  - 870 observations / 250 obs per week ~ 3.5 weeks
- Great Smoky Mountains National Park
  - 870 observations / 149 obs per week ~ 5.8 weeks
- Yellowstone National Park
  - 870 observations / 507 obs per week ~ 1.7 weeks
- Yosemite National Park
  - 870 observations / 282 obs per week ~ 3.1 weeks