Capstone Project: Biodiversity

Eric Kinley

Introduction to Data Analysis

Situation

- The National Parks Service maintains a database of species called 'species_info.csv'
- The data details the scientific names, common names, categories, and conservation status
- The following analysis was done:
  - Graphed the conservation status by species
  - Determined whether certain categories of species are more likely to be endangered
  - Graphed the observations of sheep per week by park
  - Conducted sample size determination for foot and mouth disease analysis

Considerations

- The species database includes multiple instances of some species (for example the Zizia aptera), so when conducting a count of the species only unique instances of scientific names were included

	category	scientific_name	common_names	conservation_status
632	Vascular Plant	Zizia aptera	Golden Alexanders	No Intervention
3347	Vascular Plant	Zizia aptera	Heartleaf Alexanders, Heart-Leaf Alexanders, M...	No Intervention

Conservation Status By Species

Conservation Status by Species
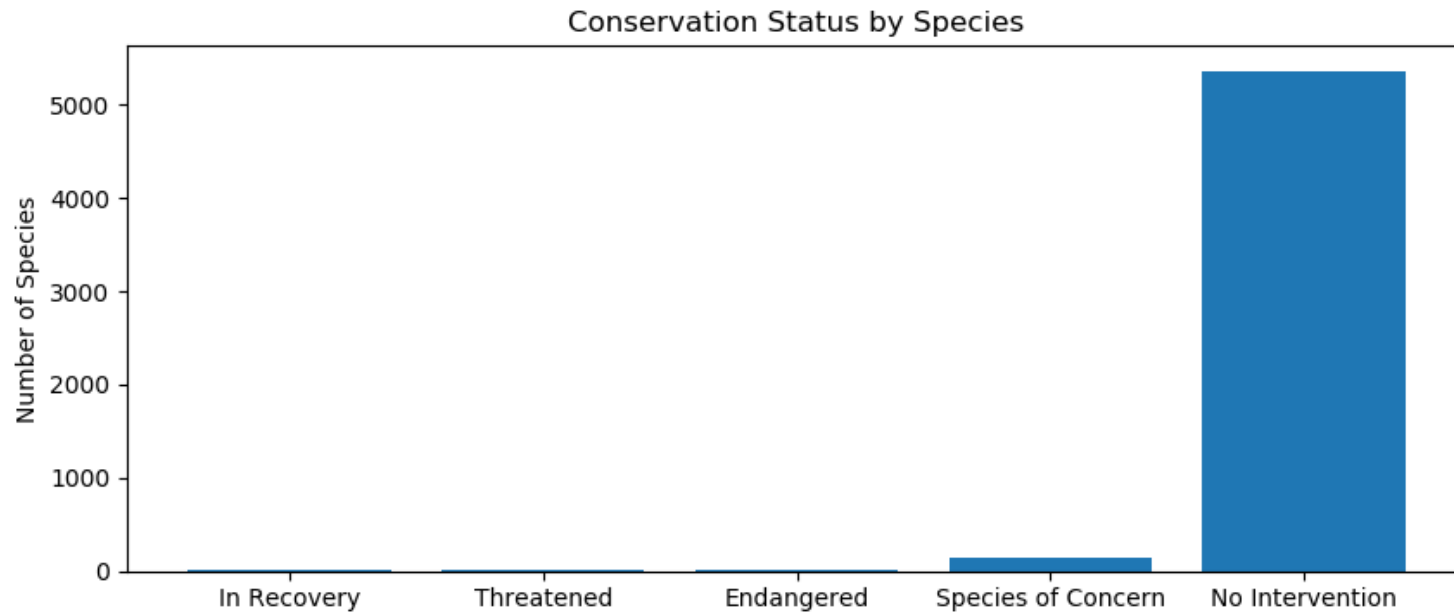
- Thankfully most species are not threatened

Endangered Species -- Analysis

- Tests of significance were done to assess the endangered status between different categories of species
  - Chi square analysis was conducted because the data is categorical
  - A p-value of less than 5% is considered significant
- Observations:
  - the difference between the percentage of <u>endangered mammals</u> and <u>endangered birds</u> is **not** statistically significant (calculated p-value 0.6876)
  - the difference between the percentage of <u>endangered mammals</u> and <u>endangered</u> reptiles is statistically significant (calculated p-value 0.0384)

Endangered Species -- Recommendation

- Recommend that conservationists focus their protection efforts on the categories of birds and mammals which have the highest percentage of species with a recorded conservation status

Observations of Sheep per Week

Observations of Sheep per Week
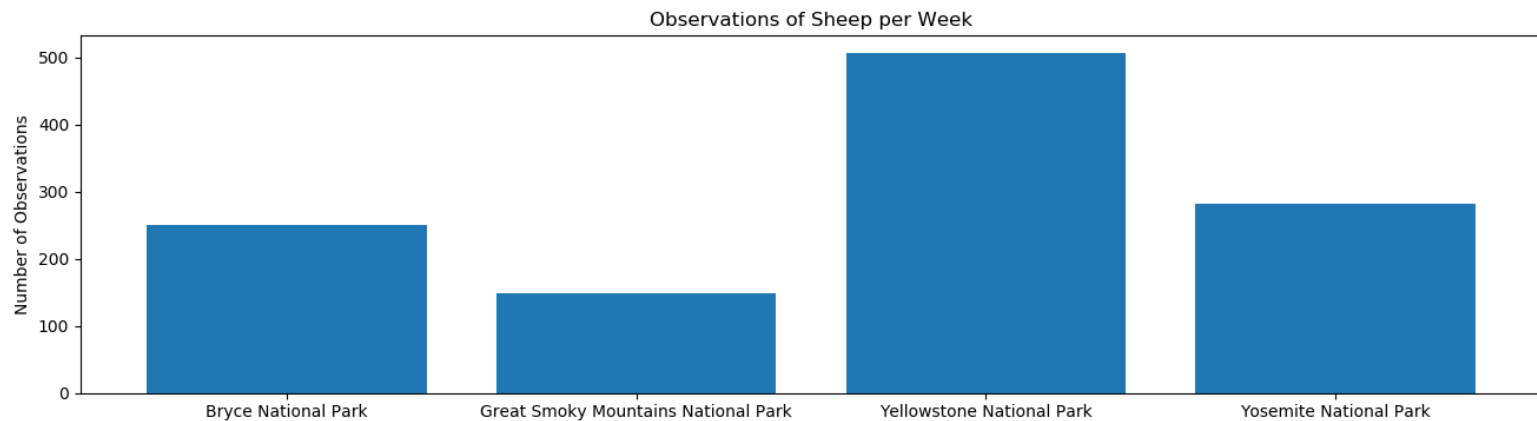
- Largest number of sheep sightings at Yellowstone National Park

Foot and Mouth Disease

- Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease. Currently 15% of sheep at Bryce National Park have foot and mouth disease. Scientists would like to know whether the program is able to reduce this occurrence to 10%.
- Calculations:
  - Minimum detectable effect: 5%/15% = ~33%
  - Required sample size was determined to be 520
  - In order to gather sufficient observations at each park it will take:
    - 2.08 weeks at Bryce National Park
    - 1.03 weeks at Yellowstone National Park