Biodiversity Capstone Project:
Review and Analysis of National Park Service Data

Daniel Berger

Capstone Project for CodeAcademy:
Introduction to Data Analysis

NPS Conservation Species Data

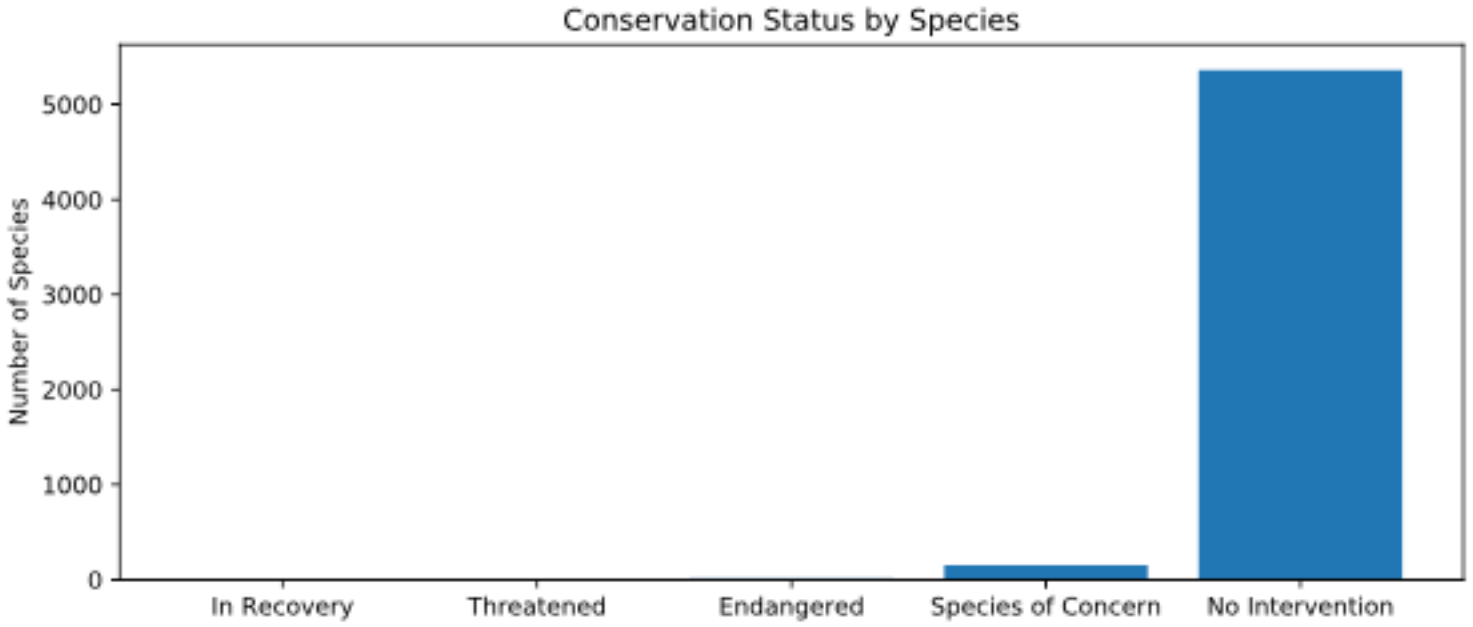
For more than 5500 species the raw data contains:

- Category (7): Amphibian, Bird, Fish, Mammal, Nonvascular Plant, Reptile, and Vascular Plant
- Scientific names
- Common names
- Conservation Status (4): Endangered, In Recovery, Species of Concern, and Threatened
- The data set needed to be massaged as there were duplicates and missing values in the data
- Species with missing Conservation Status were assigned the status: “No Intervention”

Conservation Status Counts by Species

Conservation Status	Count
Endangered	15
In Recovery	4
No Intervention	5363*
Species of Concern	151
Threatened	10

* ~97% have no protected status



Percent Protected by Category

category	% protected
Mammal	17.04
Bird	15.36
Amphibian	8.86
Fish	8.73
Reptile	6.41
Nonvascular Plant	1.50
Vascular Plant	1.07

Analysis Between Categories

If we group the categories into percentage levels:

```
Level 1 (<1.5%): Nonvascular & Vascular Plants
Level 2 (<6.41%): Reptile
Level 3 (<8.86%): Amphibian, Fish
Level 4 (<17.04%): Mammal, Bird
```

Using Chi-squared analyses there are significant (pval <.05) differences between:

- Level 1 (plant categories) and all other levels
- Any levels 2 “steps” apart from each other
- For example, level 2 & level 3 have no significant difference, but levels 2 & 4 do. (except for birds & reptiles which have a pval slightly greater than .05)

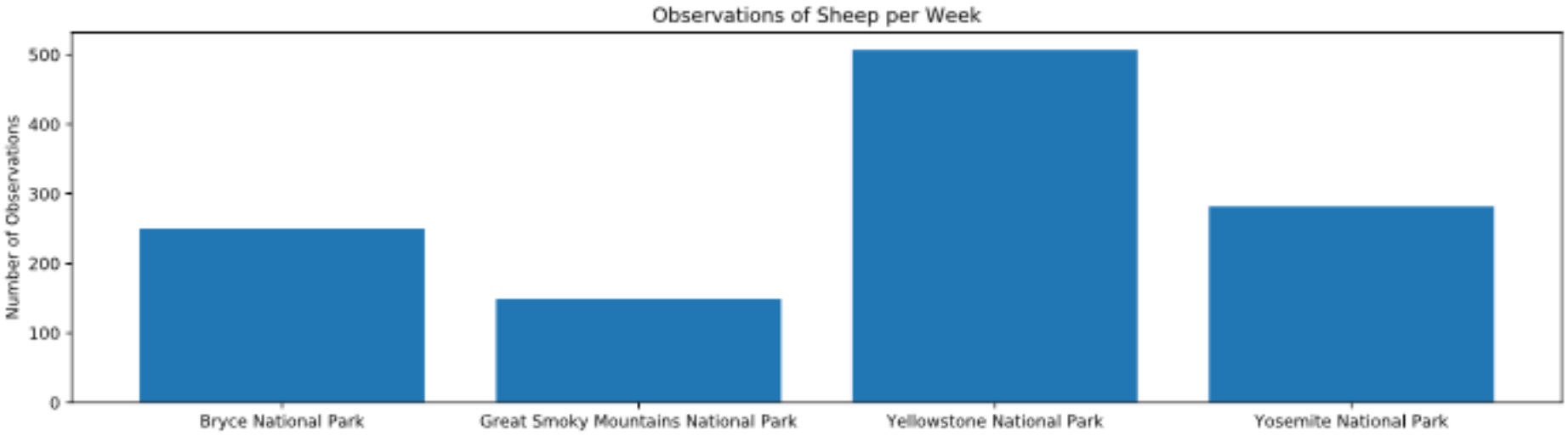
Recommendations on Endangered Species

- Thoroughly observe and count Mammals & Birds as they are significantly more likely to be endangered
- Similarly Amphibians and Fish should be carefully monitored as their level of endangerment is also significant
- These analyses do not consider the potential “whys” for the significant differences and may required further investigation:
  - Are the categories of concern more vulnerable to being endangered from human activities?
  - Are the categories of concern more observed or observable than the others?
  - In other words might the other categories be just as endangered, but are not being monitored to the same degree?

Effectiveness of the Foot & Mouth Reduction Program on Sheep (Data Prep)

The following data preparation steps were taken to determine the required sample size of sheep observations:

1. Joined the species & observation data sets by species name
2. Extracted the (mammal) sheep species from the combined data set through their common names
3. Determined the number of the sheep observed in each park



Effectiveness of Foot & Mouth Reduction Program on Sheep (Sample Size)

The following values were used to determine the Sample Size

- Baseline: 15% - the occurrence rate of F&M in Bryce last year
- Minimum Detectable Affect: 33% - a decrease of at least 5 percentage points was needed to show effectiveness – meaning a decrease from 15% to 10%.
- Statistical significance: 90%

Sample size required: 870 observations*

*Unfortunately the sample size calculator was not working correctly and the correct sample size should be 510

Number of Weeks needed for Foot & Mouth Sheep Observations

For a sample size of 870, sheep will need to be observed for the following number of weeks in each park:

Park	Observations/wk	Total Wks
Bryce	250	3.5
Great Smoky Mountains	149	5.8
Yellowstone	507	1.7
Yosemite	282	3.1