## Capstone Option 2:

# Biodiversity for the National Parks

Ernest Lindner

10 April 2018

Performed on Codecademy Website

# Observations on species_info.csv Data

- An Extensive Database with
  - 5541 Lines of Data / Species
  - 7 Categories of Wildlife (Plants and Animals)
- Aspects of Database (Column Names) Include:
  - Category (category)
  - Scientific Name (scientific_name)
  - Common Names (common_name)
  - Conservation Status (conservation_status)
- Seven Categories of Wildlife Include:
  1. Mammal
  2. Bird
  3. Reptile
  4. Amphibian
  5. Fish
  6. Vascular Plant
  7. Nonvascular Plant

# Observations on species_info.csv #2

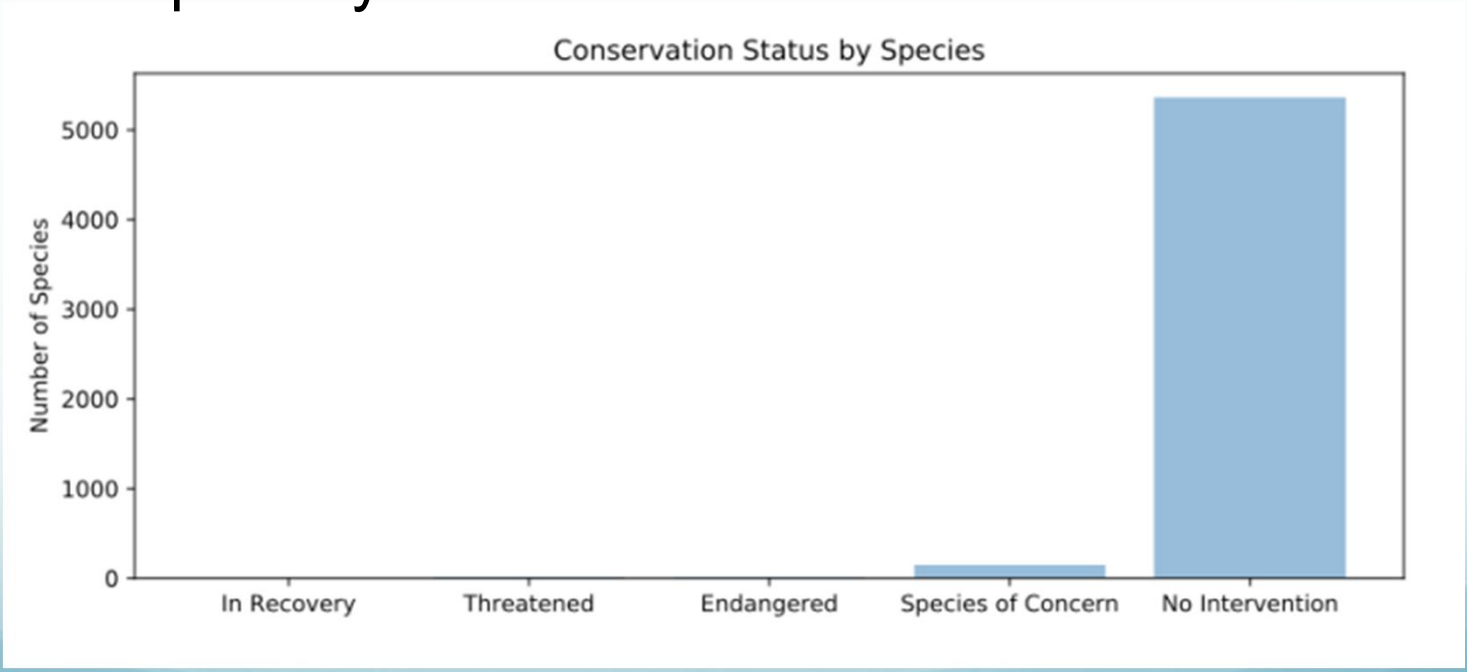
- Items Under Scientific Names Included Genus and Species
- Several Relevant Items Appeared Under Common Names
- Conservation Status Data (Originally) was Divided into Four Groups (with Some Items Unclassified) (NAN):
  - Endangered (Seriously at risk of extinction)
  - Threatened (Vulnerable to endangerment in the near future)
  - Species of Concern (Declining population or appears to be in need of conservation)
  - In Recovery (Formerly Endangered, but currently not in danger of extinction)

# Observations on species_info.csv #3

- However, the Unclassified Conservation Status Items were Converted to “No Intervention”
- Therefore, Conservation Status Values Are:
  - Endangered 15
  - In Recovery 44
  - Threatened 10
  - Species of Concern 151
  - No Intervention 5363

# Observations on species_info.csv. #4

- Most are Listings are Identified with “No Intervention”;
- Less than 0.3% Identified as “Endangered” and 2.7% Are Considered “Species of Concern”. Graphically:



# Significant Calculations #1

- Breakdown on Protected Species*

| Category | Not Protected | Protected | Percent Protected |
|---|---|---|---|
| Amphibian | 72 | 7 | 8.9% |
| Bird | 413 | 75 | 15.4% |
| Fish | 115 | 11 | 8.7% |
| Mammal | 146 | 30 | 17.0% |
| Nonvascular Plant | 328 | 5 | 1.5% |
| Reptile | 73 | 5 | 6.4% |
| Vascular Plant | 4216 | 46 | 1.1% |

* Protected status includes Conservation Status items of Endangered, In Recovery, Threatened, and Species of Concern.

# Significant Calculations #2

- Based on a Chi-Squared Independence Test, Some Differences in Table Are NOT Significant
  - e.g., Difference between Mammals and Birds was Assessed with a P-value of 0.6876, i.e., *Not Significant* ( > 0.05)
- However, Some Differences ARE Significant
  - e.g., Difference between Mammals and Reptiles was Assessed with a P-value of 0.0384 and was Considered *Significant* ( < 0.05)
- Therefore, Certain Types of Species Are More Likely to be Endangered
- Birds and Mammals are at the Top of the List

# Recommendations

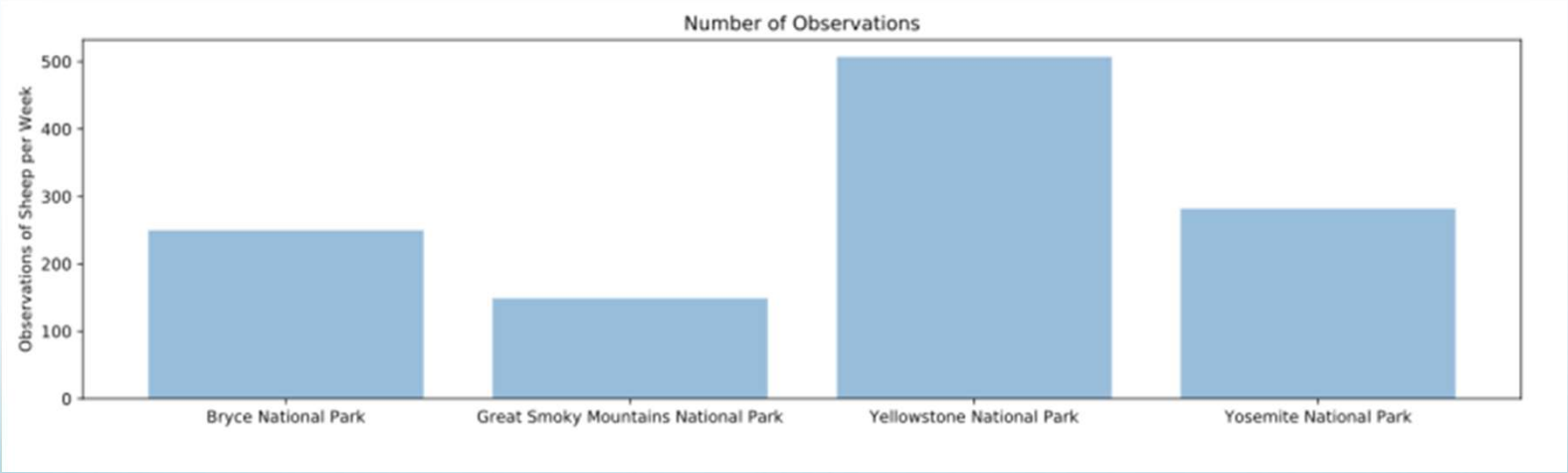
- Based on Calculations, Concerns Should Prioritize the Protection of Mammal and Birds
- Less Focus on Vascular and Nonvascular Plants

# Foot and Mouth Study

- Created a New Database on Sheep Observations
- Database is a Composite of ‘observations.csv’ and 'species_info.csv’
- Focus is on Sheep (Animal) Population
  - Excludes Plants with a Common Name “Sheep”
- Based on Observations at Bryce National Park, Great Smoky Mountains National Park, Yellowstone National Park and Yosemite National Park

# Sheep Sightings* in Each Park

| Park | Observations |
|---|---|
| Bryce National Park | 250 |
| Great Smoky Mountains National Park | 149 |
| Yellowstone National Park | 507 |
| Yosemite National Park | 282 |




* Sightings per Week

# Parameters for Study

To Determine the Effectiveness of Program:

- A Sample Size Required for Study is 890*
- The Time to Conduct a Study (Based on Observation History):
  - At Yellowstone 1.76 Weeks
  - At Bryce 3.56 Weeks

* Size needed to determine if change is significant (as due to treatment). Sample size was computed using a level of significance of 90%, a baseline conversion rate of 15%, and a minimum detectable effect of 33% (a 5% drop).

# Graphs Created in Project