# Biodiveristy Capstone Project

04/04/18

*Investigating Protected Species*

# Description of data in species_info.csv

- A section describing the data in **species_info.csv**. Be sure to include some (or all) of what you noticed while working through the notebook.

- the information contained is the dataframe **species_info.csv** :
    - index category,
    - scientific_name,
    - common_names
    - conservation_status

- There are 5541different species in the species DataFrame ➜ species_count = 5541.

# Description of data in species_info.csv 2

- There are 7 categories ( species type) in the DataFrame species, which are :
    - Mammal
    - Bird
    - Reptile
    - Amphibian
    -  Fish
    - Vascular Plant
    - Nonvascular Plant

- The different values of conservation status are : nan,  'Species of Concern' 'Endangered', 'Threatened' and 'In Recovery'

# Species Conservation Status Analysis

- Number of scientific_names for each conservation_status criteria are:

| Conservation_status | Conservation_counts |
|---|---|
| Endangered | 15 |
| In Recovery | 4 |
| Species of Concern | 151 |
| Threatened | 10 |

We notice that we have a conservation status for only 180 species, but we found that there far more than that. Only a small number of species are categorized as needing some sort of protection. The rest have conservation_status equal to None, or NaN.

# Conservation_counts_fixed

|   | conservation_status | scientific_name |
|---|---------------------|-----------------|
| 1 | In Recovery | 4 |
| 4 | Threatened | 10 |
| 0 | Endangered | 15 |
| 3 | Species of Concern | 151 |
| 2 | No Intervention | 5363 |



Conservation Status by Species
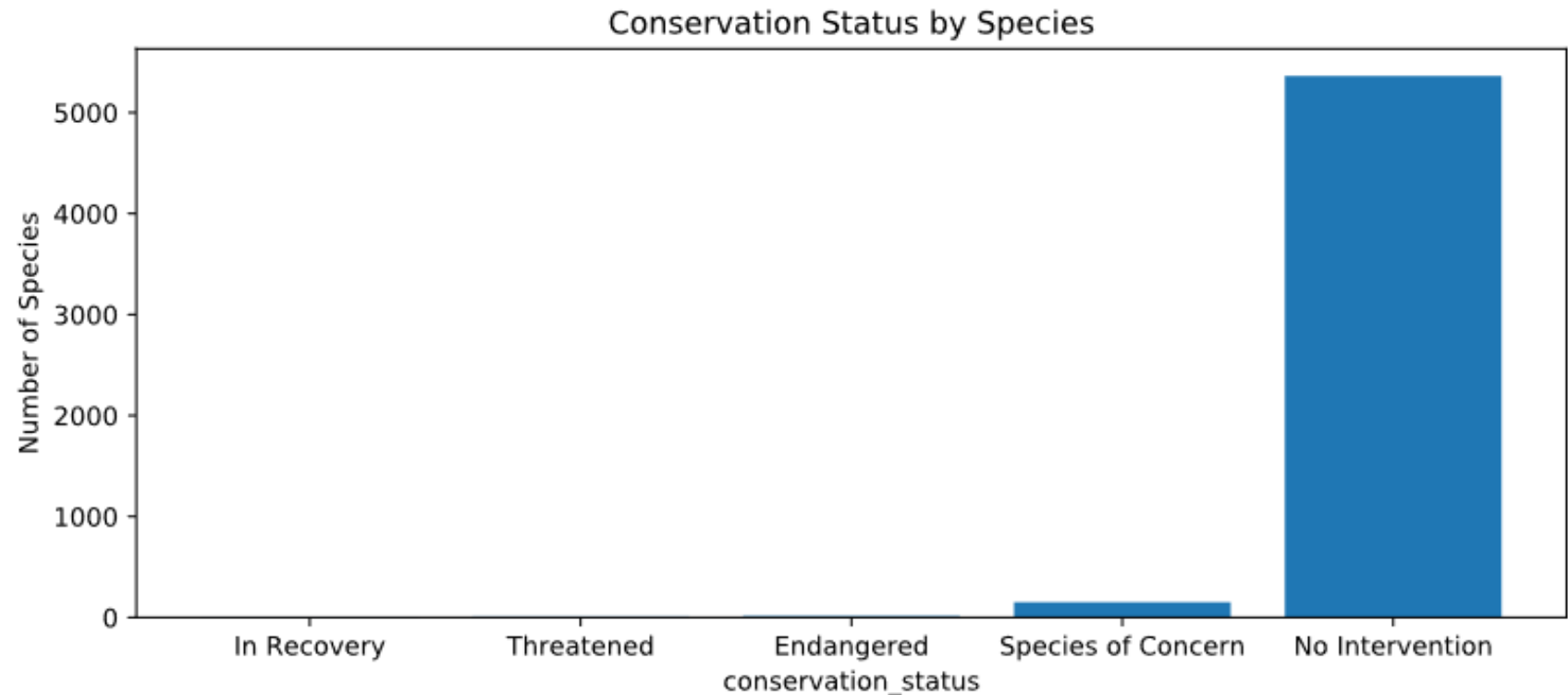
# Significance calculations for endangered status between different categories of species

- Mammals 17% and birds 15% are on the top of the categories that are protected ➜ it seems that both of them are likely to be more endangered.

| category | not_protected | Protected | percent_protected |
|---|---|---|---|
| Amphibian | 72 | 7 | 0.088608 |
| Bird | 413 | 75 | 0.153689 |
| Fish | 115 | 11 | 0.087302 |
| Mammal | 146 | 30 | 0.170455 |
| Nonvascular Plant | 328 | 5 | 0.015015 |
| Reptile | 73 | 5 | 0.064103 |
| Vascular Plant | 4216 | 46 | 0.010793 |

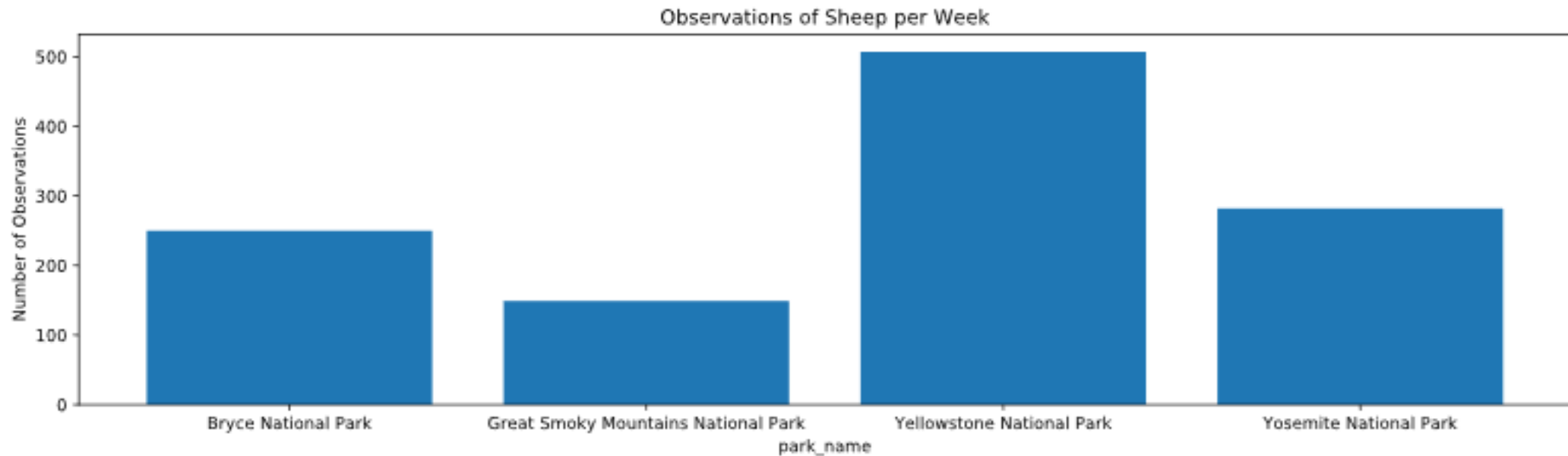# Recommendation for conservationists concerned about endangered species

- pval_mammals_birds = 0.688 > 0.05 ➔ we fail to reject the null hypothesis there is no significant difference between Mammals and Birds : the difference is due to chance.

- pval_reptile_mammal = 0.038<0.05➔ we reject the null hypothesis there is a significant difference between Mammals and reptiles, the difference is not due to chance.

- pval_reptile_mammal = 0.7>0.05, then we fail to reject the null hypothesis there is no significant difference between fish and reptiles, the difference is due to chance.

- pval_nonvascular_reptile = 0.033<0.05, then we reject the null hypothesis there is a significant difference between non vascular and reptiles, the difference is not due to chance.

- pval_fish_amphibian = 0.82 > 0.05, then we fail to reject the null hypothesis there is no significant difference between non fish and Amphibians, the difference is due to chance

    ➔ Conservationists should focus on both Mammals and Birds regarding the protection in comparison with other categories. They are more likely to be endangered.

    ➔ Only small proportion of species are protected; percent_protected values are less than 15,50%

# Sample size of Foot and Mouth Reduction Effort

- baseline = 15

- minimum_detectable_effect = 33.33
- sample_size_per_variant = 870
- yellowstone_weeks_observing = 870 / 507
- at yellowstone_weeks_observing the scientists need 1.7 weeks; 12 days
- bryce_weeks_observing = 870 / 250
- at bryce_weeks_observing the scientists need 3.5 weeks; 24.5 days

# Observations of sheep per week

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



Observations of Sheep per Week

# Thanks