# Introduction to Data Analysis: Biodiversity Capstone Project

Abigail Crawford, 21 June 2019

# The data

## species_info.csv

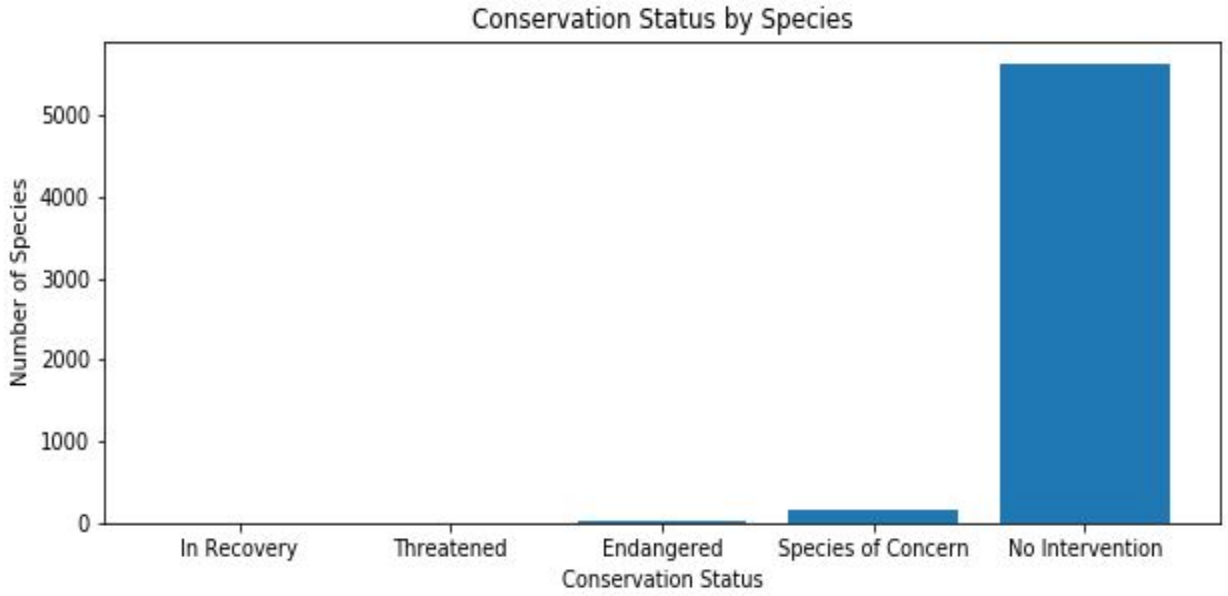
- 5824 records of species in American National Parks
- **Category**
  - Mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant
- **Scientific Name**
- **Common Names**
- **Conservation Status**
  - Species of concern, endangered, threatened, in recovery

## observations.csv

- 23296 records accounting for wildlife observations made over 7 contiguous days
- **Scientific Name**
- **Park Name** (4 parks)
  - Bryce NP
  - Great Smoky Mtns NP
  - Yellowstone NP
  - Yosemite NP
- **Observations**

# Key Takeaways from Species_info.csv

- Most of the species in the database are plants rather than animals
- The majority of species require no intervention with respect to conservation



# Statistical Significance

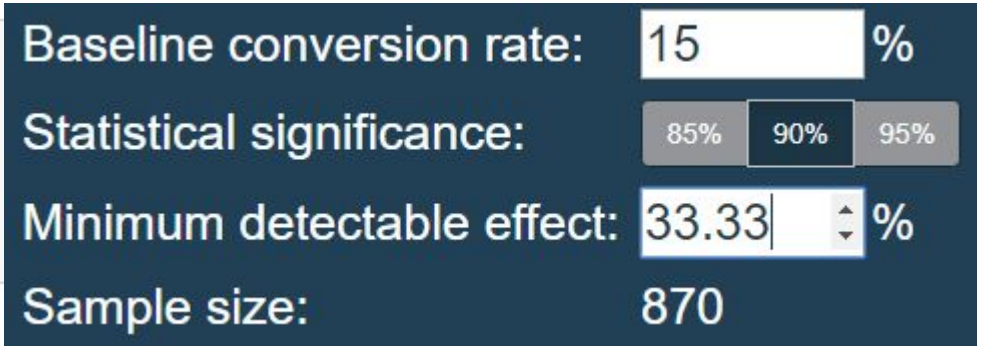
- Of all species categories, mammals have the largest percentage of species that are protected, followed by birds
- A Chi square test reveals that mammals are statistically more likely than reptiles and fish to be endangered.
    - **This finding has implications for conservation-related policies; efforts should be put into protecting mammals, who are more at risk of becoming endangered than other types of species.**
- Statistical analysis shows that there is no significant difference between the percentage of mammals who are endangered, and that of birds and amphibians.

# Foot and Mouth Disease in Sheep

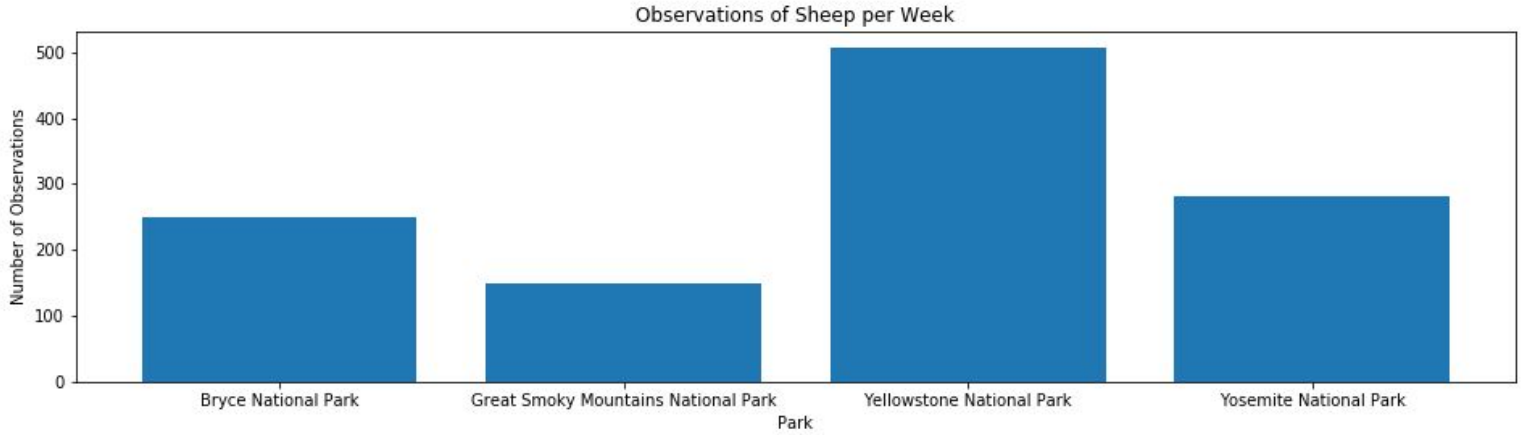
- Using Bryce Canyon NP's population of sheep with Foot and Mouth Disease (15%), a statistical significance of 90%, and a minimum detectable effect of 33.33%, I used the Sample Size Calculator to determine that each of the four national parks should observe 870 sheep in order to determine whether the occurrences of Foot and Mouth Disease have varied by at least 5%.

```
baseline = 15
statistical_significance = 90

minimum_detectable_effect = (100 * .05) / 0.15
print(minimum_detectable_effect)

sample_size = 870.0
```

```
33.3333333333
```



# Sheep Sightings per Week



# Expected Timeframe for Sheep Study

- Given the established rate of sheep observation determined in observations.csv and shown on the previous slide (250/wk at Bryce Canyon NP, 507/wk at Yellowstone NP), I anticipate the Foot and Mouth Disease data for the sheep populations in Bryce Canyon NP and Yellowstone NP will be complete roughly 3.5 weeks from when the work begins.
- Yellowstone's data should be complete within 2 weeks of beginning data collection, but Bryce Canyon NP will take somewhat longer, due to that park's slower rate of sheep sightings.

# Thank you!