# National Parks Biodiversity

Christina Stefaniuk

# Data in species_info.csv

- There are 5541 different species in the data set provided

- There are 7 different species types in the data set provided they are:
  Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant and Nonvascular Plant

- There are 5 conservation status types, they are:
  - nan (renamed to No Intervention) – 5363 Scientific Names fall into this category
  - Species of Concern – 151 Scientific Names fall into this category
  - Endangered – 15 Scientific Names fall into this category
  - Threatened – 10 Scientific Names fall into this category
  - In Recovery – 4 Scientific Names fall into this category

# Observations of data set

▶ By observing the grouping of conservation statuses it can be seen that ~3.2% of the species within the data set currently require some degree of intervention in order to ensure that they remain thriving. 96.8% of them currently require no intervention.
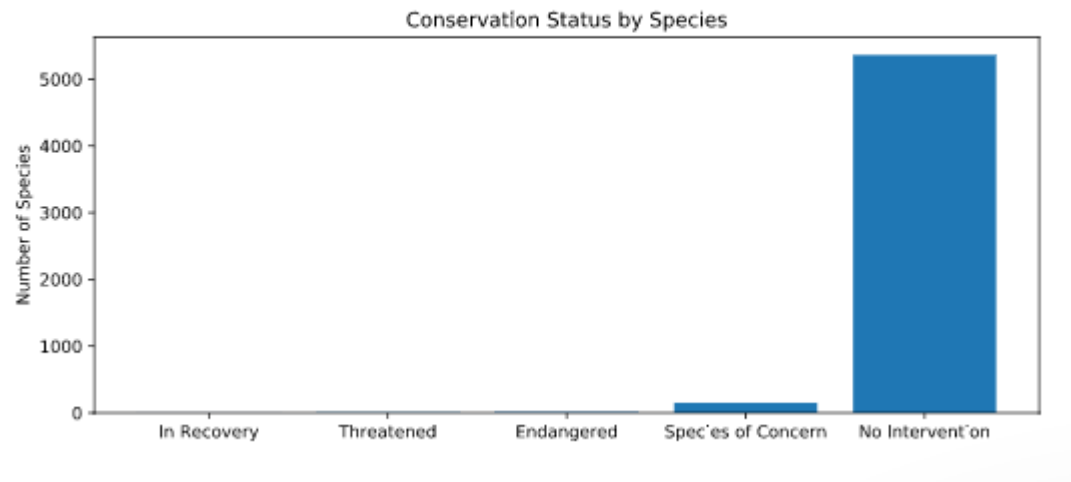
# Observations of data set

- Below is a ranking of the species types in the order of highest percentage of protected species within species types:
    - 1. Mammal -17% (30/146)
    - 2. Bird – 15% (75/413)
    - 3. Amphibian – 8.8% (7/72)
    - 4. Fish – 8.7% (11/115)
    - 5. Reptile – 6.4% (5/73)
    - Nonvascular Plant – 1.5% (5/328)
    - Vascular Plant – 1% (46/4216)
    - Plants seem to be the least susceptible to danger, with cold blooded species types being the next least likely and warm blooded species type being the most likely to have a higher majority of protected species.
    - However the sample sizes of the various different species types vary greatly. Not sure if this is significant. Further testing of significance required.

# Chi Squared Test

- Based on the data in the data set mammals appeared to be more likely to be endangered than birds. Created a null hypothesis that checked if this difference is due to chance and performed a chi squared test on it. As the p-value turned out to be 0.687. Due to this we accept the null hypothesis that the result is due to chance as the p-value is higher than 0.05.

- Based on the data in the data set mammals appeared to be significantly more likely to be endangered than reptiles. Created a null hypothesis that checked if this difference is due to chance and performed a chi squared test on it. As the p-value turned out to be 0.038. Due to this we reject the null hypothesis that the result is due to chance as the p-value is lower than 0.05 and that the difference is significant.

- Conclusion certain types are definitely more likely than others to be protected.

# Recommendations for Scientists

► As there is a very large significance difference in the likelihood of certain species types becoming protected species faster scientists should first focus on:

► 1. Mammals and Birds (There is no significant difference in the % between these two and as such should both be assigned high priority to distribute resources available to)

► 2. Amphibian, Fish and Reptile species as a secondary priority

► 3. Plants as a third priority

# Plotting Sheep Sightings
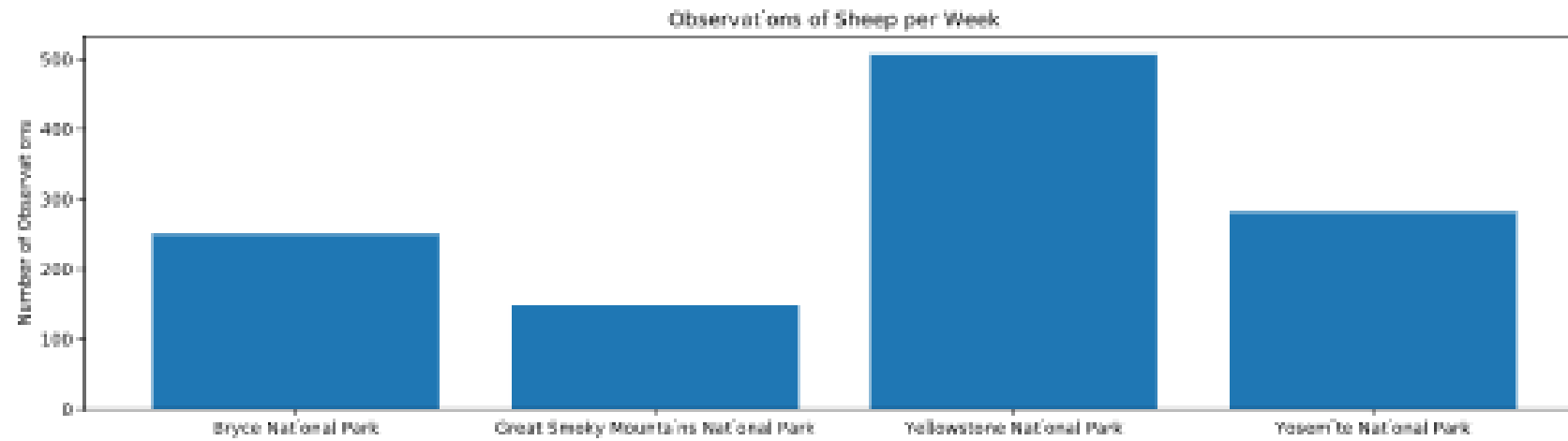
# Sample Size Determination

Baseline of sample size determination is the 15% contamination of foot and mouth disease that occurred last year.

To be able to see their measures are working with confidence they need a minimum detectable difference of 33.3333%.

Using the sample size calculator with the following inputs: Baseline Conversion Rate (15%) Statistical Significantce (90%) Minimum detectable effect (33.3333%) the resulting sample size ended up being 870.

Divide the sample size by variant by the available number of observations per park to determine that it would take 1.7 weeks to observe Yellowstone park and 3.5 weeks to observe Bryce Park