

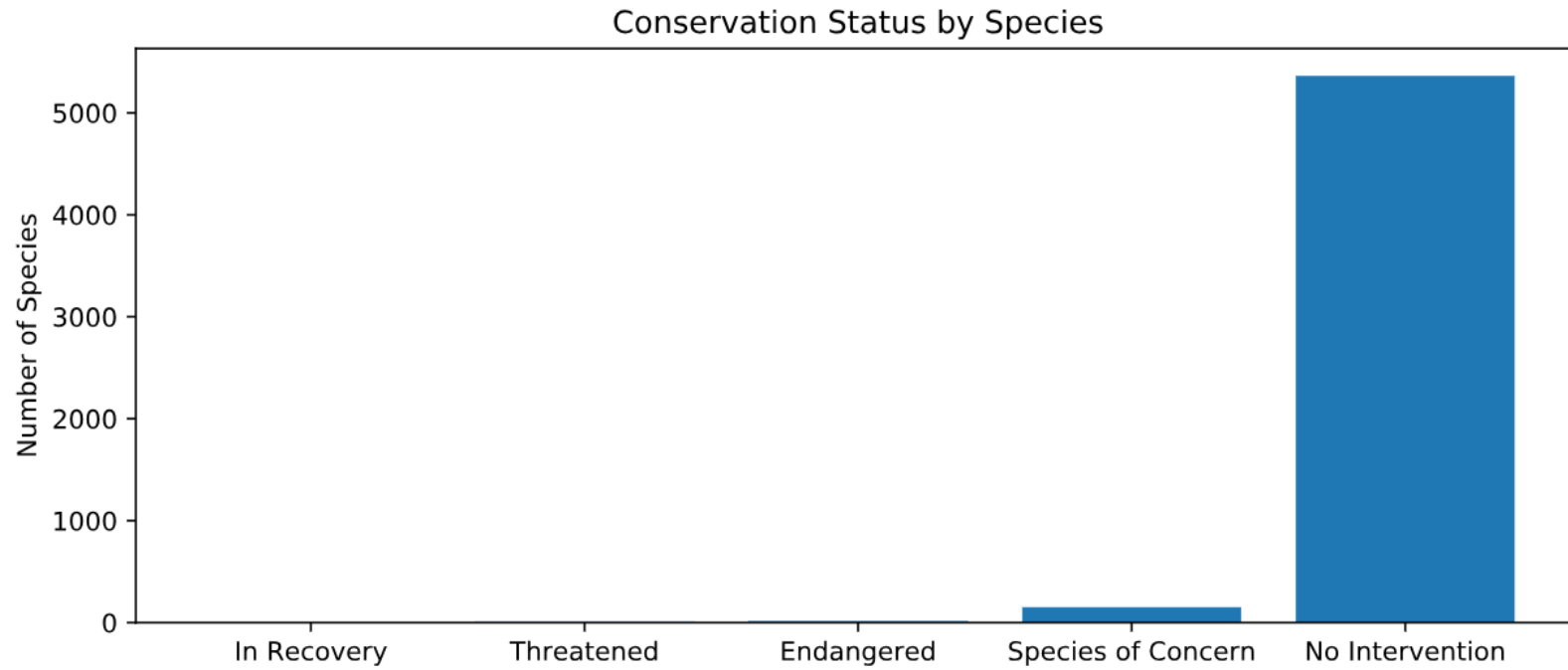
# Biodiversity of the National Parks Capstone Project

Thomas Donnelly

# Species Info

- Species\_info.csv contains information on different species of plant and animals: category, scientific name, common names, and conservation status
- Species are categorized as either Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, or Nonvascular Plant
- Species conservation status is given as Species of Concern, Endangered, Threatened, In Recovery or nan which can be considered as No Intervention
- There are far more species requiring no Intervention than any other conservation action, this is visualized on graph 1

# Graph 1



# Significance Calculations

- It was investigated whether certain types of species more likely to be endangered than others.
- Species were divided between protected and unprotected based on their conservation status and then a percent of the species that is protected was calculated.
- It was found that some categories of species had a higher percentage of species that were protected than others, with mammals having the highest at 17%
- Significance in this difference needed to be tested and so a chi squared test was performed by creating contingency table of two categories with columns of the number of protected and unprotected species.
- When chi squared test was run on mammals and birds the p value was 0.68 showing no significance, however when the test was run on mammals and reptiles there was a p value of 0.03.
- As this P Value was below 0.05, it disapproves the null hypothesis that there is no significance difference in protected species based on their category.
- Thus there is a significant differences in the protected status of species based on their category.

# Recommendation

- Conservationists concerned about endangered species should focus their efforts on Mammals and Birds as they are more likely to be endangered than other species.
- Conservationists should also study the resistances of plants and reptiles as they are less likely to be in need of protection. Possibly these findings to better protect other species.

# Sample Size Determination

- The sample size needed to detect a 5% reduction in foot and mouth disease was determined to be 870 sheep.
- The baseline was determined to be 15% from previous data.
- Minimum detectable effect was determined through the equation  $100 \times 5 / 15$ . Where 5 is the wanted % observed change with confidence and 15 is the baseline. Resulting in a 33.3% minimum detectable effect
- Statistical significance of 90% was used as a default.
- The parameters were entered into the sample size calculator to calculate the final required sample size.

# Sample Size Determination

- Species with the common name sheep were selected in data from species\_info.csv. This data was merged with data from observations.csv to select just the observations of sheep at the national parks.
- This data was then grouped by the total observations of sheep over the past week at each national park. This data is plotted on graph 2.
- By dividing the required sample size by the observations of sheep at each park, it was determined that it would take 1.7 weeks for scientists at Yellowstone National Park to observe enough sheep and 3.48 weeks for scientist at Bryce National Park to do the same.

# Graph 2

