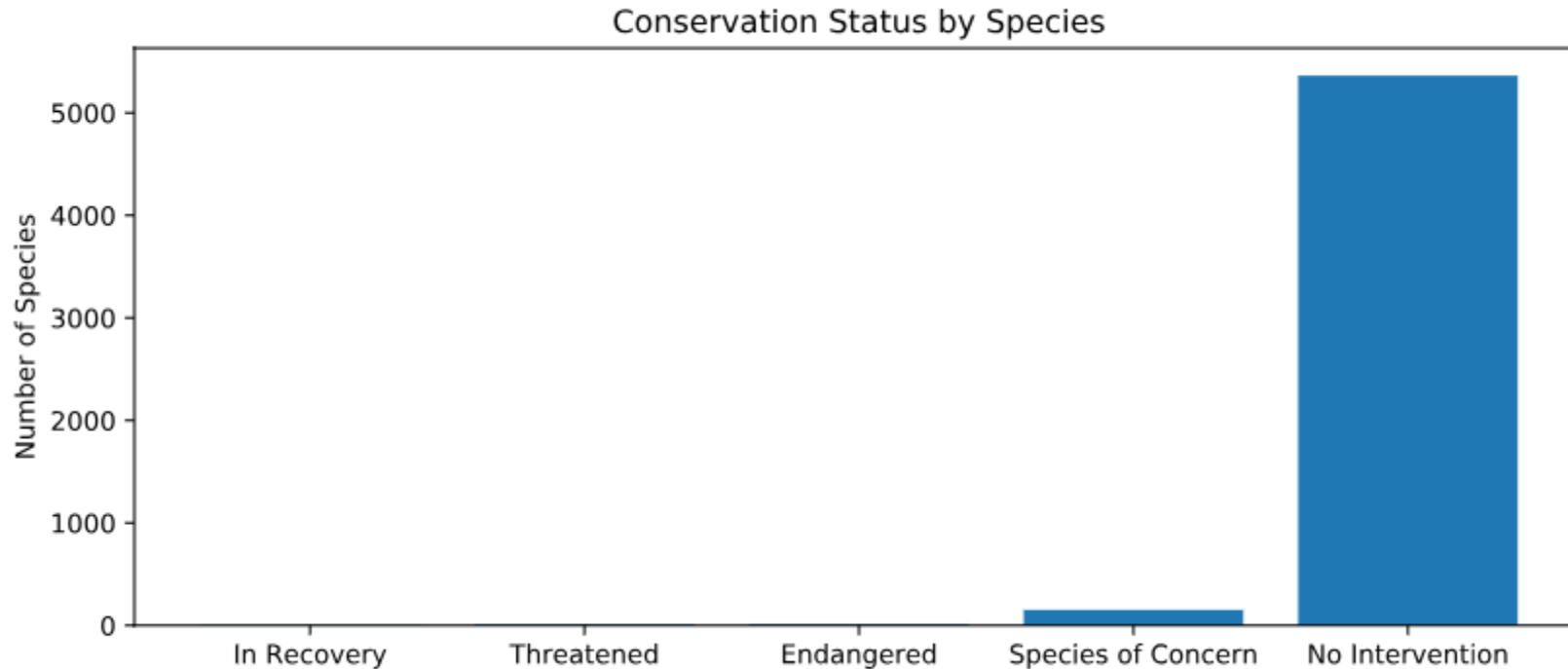


# Analyzing Biodiversity Trends at National Parks

# Introduction

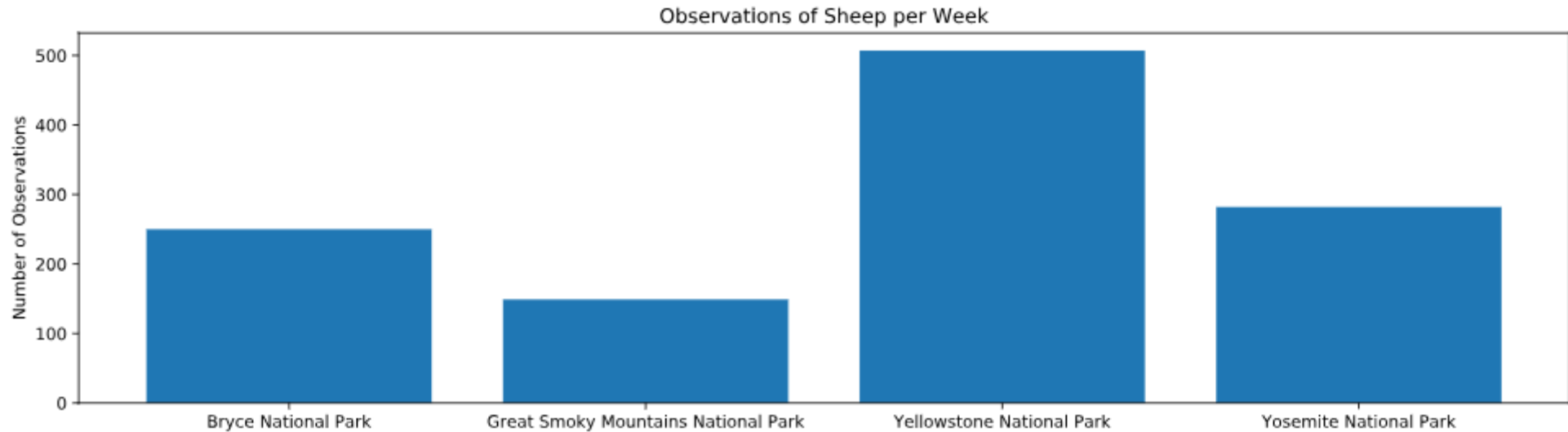
- The goal of the project is to examine the situation and trends with regards to different species at several major national parks. The project will examine which species are endangered and compare which groups of species are more endangered than others. The project will also examine the foot and mouth disease among sheep and will calculate a necessary sample size of the sheep to be selected to detect whether a program against the disease is working or not. In comparing the proportions of endangered species among different groups of species the chi-squared test will be used. There are many things to be learned from the dataset given and the project will not cover everything that can be inferred as the amount of inferences that can be made from the dataset is so vast.

# Number of Species Falling in Each Conservation Status



As can be seen from the graph the overwhelming majority of species are not assigned a conservation status. The percentage of species that are assigned into one of the first three categories is tiny as to not be noticeable on the graph.

# Number of Sheep Observed in a Week in Four Major National Parks:



The graph shows that the Yellowstone National Park has clearly the biggest amount of sheep observed out of all the four parks - which is around five-hundred, and the Great Smoky Mountains National Park has the least, being less than two hundred.

# Exploring the Dataset for the Project

The species\_info dataset gives a category and protection status to 5824 different species that have been observed at one of several national parks.

Out of the thousands of species in the dataset 180 have conservation statuses, indicating that their population is at some form of risk.

The seven categories of species in the dataset are mammal, bird, reptile, amphibian, fish, vascular plant and nonvascular plant.

Fifteen species are at risk of extinction while the overwhelming majority were not given any conservation status.

Mammals have the highest percentage of protected species out of all categories while vascular plants have the lowest.

# Comparing Proportions of Endangered Species in Different Categories of Species

It is possible to obtain the proportions of endangered species within each group but it can be hard to tell whether the differences in proportion are due to sampling variation or due to there being a difference in the population proportions. To determine this we make use of the chi-squared test with a 0.05 alpha level. If the p-value of the test for two groups is less than 0.05 it will be determined that the categories differ in their proportions of endangered species. This test will help make our comparisons accurate rather than just eyeball intuitively to determine if the sample proportions are different to a significant extent.

In comparing the proportions of endangered birds and mammals there was no significant difference found. Using a Chi-squared test the p-value turned out to be  $\sim 0.688$ .

When the Chi-squared test was used to compare the proportions of reptiles and mammals the difference in the proportions was found to be significant with a p-value of  $\sim 0.038$ .

The proportions of species in the fish and amphibian categories that are endangered are nearly identical, both being a little under 9%. Running a Chi-squared test the obtained p-value is  $\sim 0.825$ , no significant difference even closely detected when comparing these two categories.

# Observations and Recommendations

Some categories have significantly more species that are endangered than others. The mammal category has the highest percentage of species that are in danger. Animal categories have much higher percentages of species that are endangered than plant categories. Identify the causes that are making foremostly the animals endangered and come up with effective measures to make sure the populations are in healthy conditions.

# Determine Sample Size to Detect Whether Intervention Program is Effective

To determine the necessary sample size for detecting if their program is working for reducing the amount of foot and mouth disease among sheep the baseline was set to 15%. Since scientists want to detect changes of at least 5% the minimum detectable effect was calculated as  $(100 * 5\%) / 15\%$  which comes out to 33 and 1/3. Plugging these values into the calculator and setting the statistical significance level at 90% the obtained sample size for the task is 870. Given the sample size that is necessary for the purpose and the amount of sheep detected at each of the four major parks it is easy to calculate how many weeks scientists need to spend at each park to attain the necessary sample size with the minimum amount being at Yellowstone National Park - just a little under two weeks, and the greatest at Great Smoky Mountains National Park - close to six weeks.



# Conclusion

This project examined several issues including comparing proportions of endangered species among different categories, determining the necessary sample size of sheep to determine if an intervention is helpful and other statistics regarding the biodiversity at several major national parks. In comparing proportions of endangered species we concluded that the proportions do differ across categories of species. This proportion is highest in the mammal category and lowest in the vascular plant category. To determine the sample size of sheep for detecting if the foot and mouth disease intervention is effective it was decided that the sample size should be 870. The amount of time spent to obtain this sample size varies depending on the park where the sheep would be sampled with the least amount of time being at the Yellowstone National Park and the greatest at the Great Smoky Mountains National Park.