Analyzing Biodiversity trends at National Parks

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.

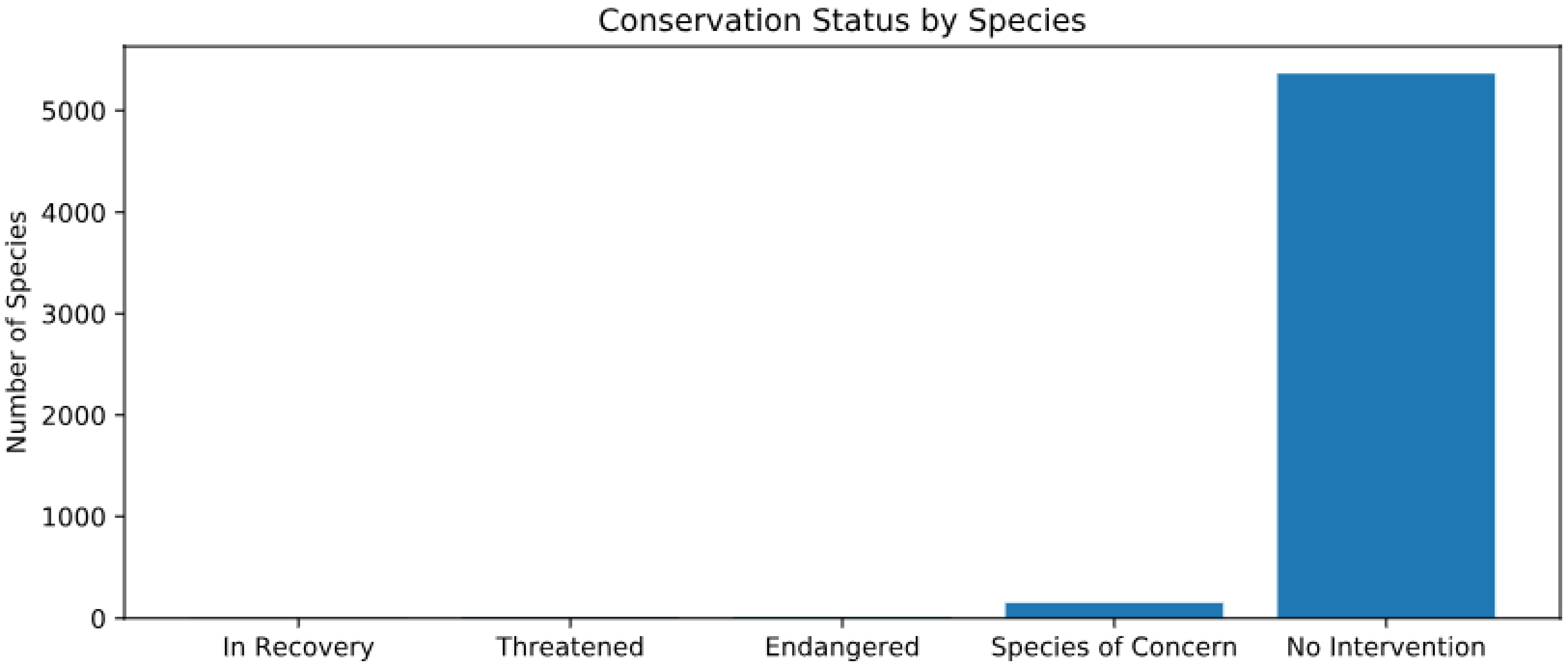
In comparing the proportions of endangered birds and mammals there was no significant difference found. Using a Chi-squared test the p-value turned out to be ~0.688.

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 ~0.038.

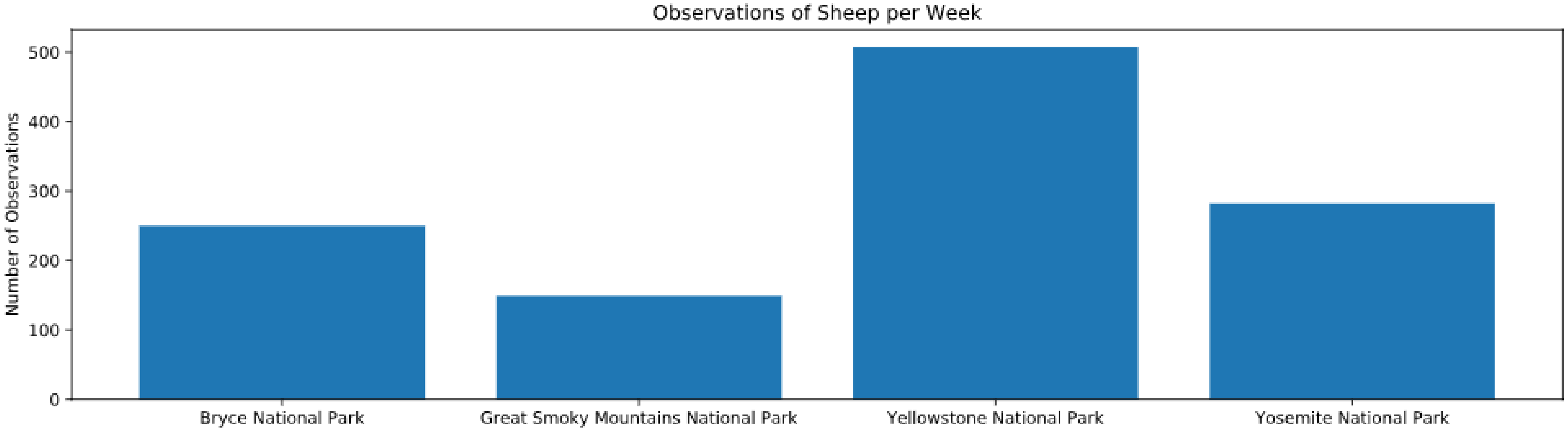
The proportion of species in the fish and amphibian categories that are endangered are nearly identical, both being a little under 9%. Running a Chi-squared test the obtained p-value is ~0.825, no significant difference even closely detected when comparing these two categories.

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.

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.



Conservation Status by Species


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park