

# Biodiversity for the National Parks

Author: Bogdan Protsenko

# Introduction

The National Parks Service would like you to perform some data analysis on the conservation statuses of endangered species and to investigate if there are any patterns or themes to the types of species that become endangered.

# Study

The previously collected data consists of two files. The first one is general information about species with their conservation statuses. The second table contains observation of these species in different national parks.

# Dataset summary

The general species information is stored in 'species\_info.csv'. This csv file includes the information about species category, scientific name, common names and conservation status. The number of rows is 5824. However, the count of unique scientific names is a bit less - only 5541. There are only two records with the same scientific name, category and common names but with different conservation status. Species are divided in seven categories: mammal, bird, reptile, amphibian, fish, vascular plant and nonvascular plant. The conservation status may not be defined if the particular species is not endangered; otherwise, one of the fourth conservation status should be specified for the species: in recovery, threatened, endangered or species of concern.

The 'observations.csv' file stores observations information at national parks for the past 7 days. It has three columns: scientific name, park name and the number of observations. Similar to 'species\_info.csv', it has 5541 unique scientific name. Overall, this csv contains 23296 rows.

# Hypotheses

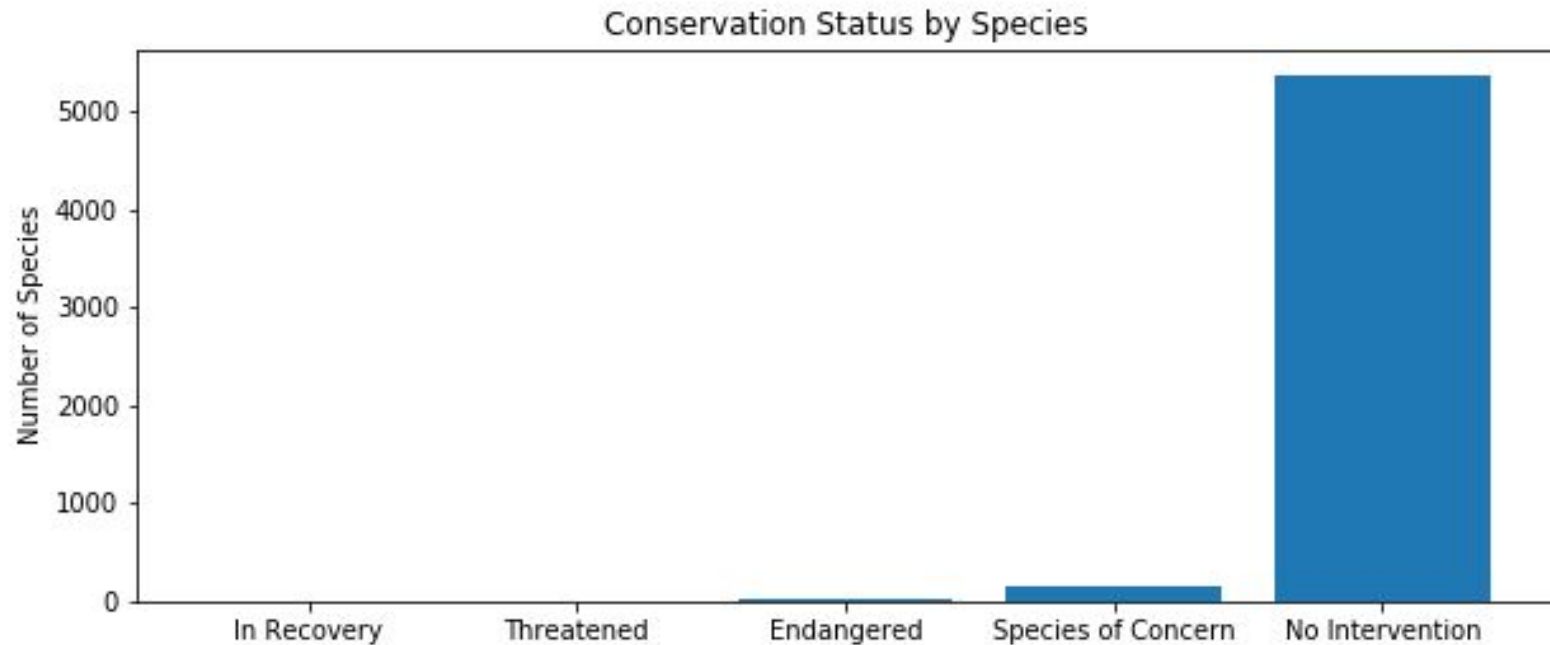
1. Certain types of species are more likely to be endangered than others.
2. A program (to reduce the rate of foot and mouth disease) performs well.

# Species conservation statuses

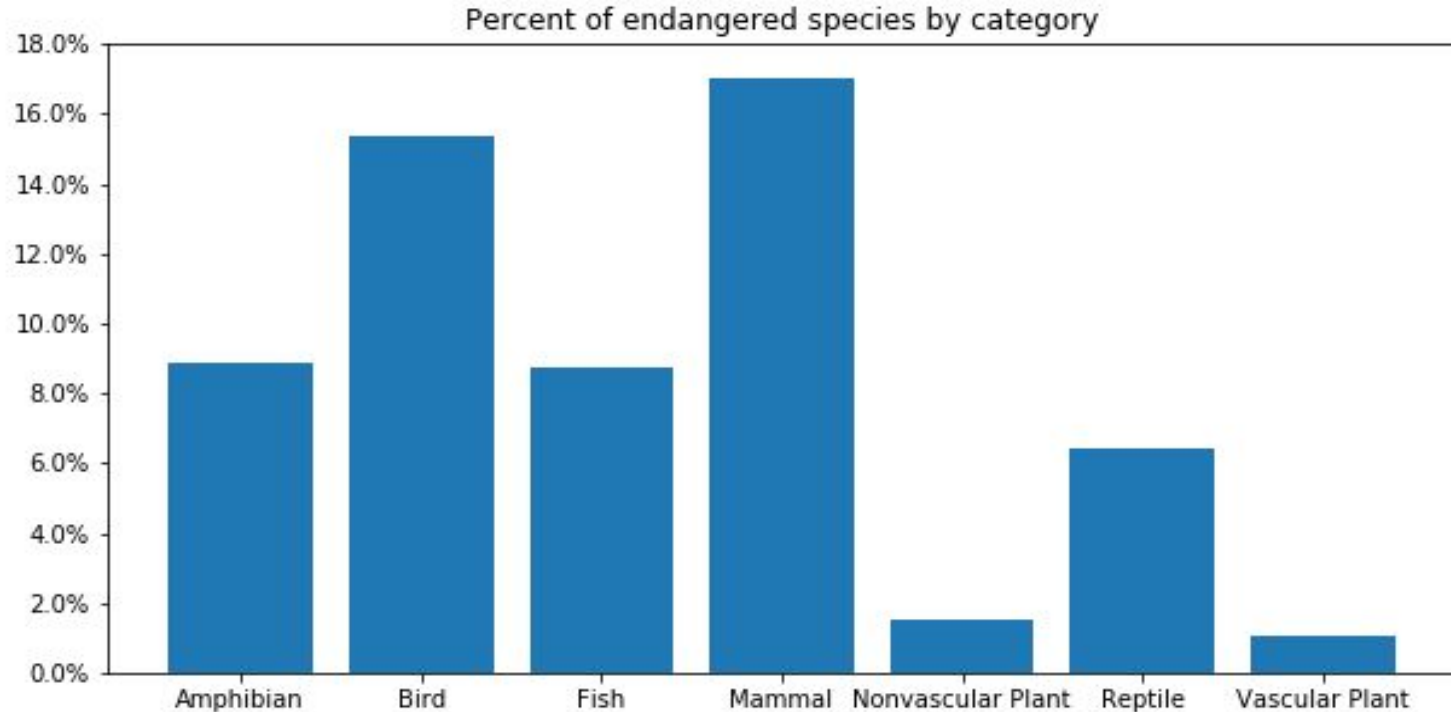
Conservation status	Number of species
In Recovery	4
Threatened	10
Endangered	15
Species of Concern	151
No Intervention	5363

The most of the species are not endangered.

# Species conservation statuses



# Are certain types of species more likely to be endangered?





# Are certain types of species more likely to be endangered?

**The most endangered categories are birds and mammals.** Chi-squared test p-value indicates that there is no significant difference between these two categories, while any of these categories has a significant difference with any of the rest categories.

Amphibians, fishes and reptiles have the average percent of endangered species. Chi-squared test indicates the relatively high p-value between any of those categories, but not for the rest of categories, if tested with these average ones.

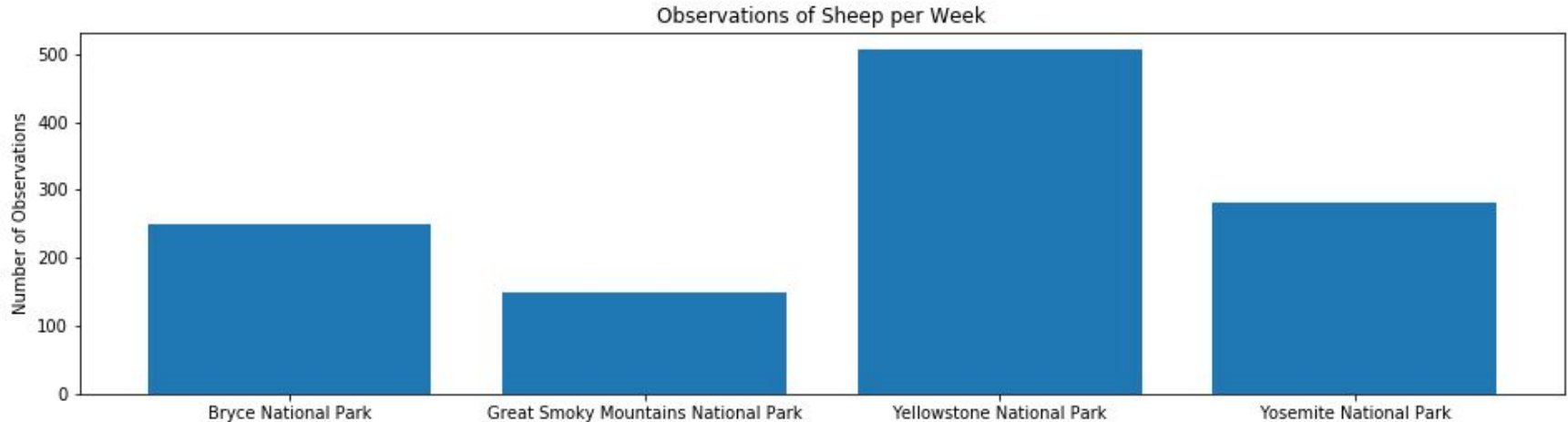
The most safe categories are vascular and nonvascular plants. Proved by Chi-squared test.

# Foot and Mouth disease reduction program description

Park Rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage point. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

The only information that the scientists currently have is that last year it was recorded that 15% of sheep at Bryce National Park have foot and mouth disease.

# Sheep observations



The Yellowstone National Parks is the best candidate for sheeps observations, while the number of sheeps in Bryce National park is relatively close to the average.

# Calculating the sample size

Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. They want to be able to detect reductions of at least 5 percentage point. In other words, the minimum detectable effect is 33.33% of the baseline. Employing this information, observing the sample size of **510** sheep allows us to indicate the effectiveness of foot and mouth disease reduction program with 90% significance.

# Weeks to observe sheep

We already established the sample size of 510 sheep. The previous week observations identifies that there were 507 sheep in Yellowstone National Park and 250 sheep in Bryce National Park. Thus, Yellowstone National Park allows to check the foot and mouth disease reduction program performance in one week, while the same check in Bryce National Park should take two weeks.

# Inferences

1. Mammals and birds are the most endangered species. For sure they deserve the special attention.
2. It is highly recommended to spend one week testing the foot and mouth disease reduction program in Yellowstone National Park.

# Thank you!

Bogdan Protsenko

[bprotsenko@gmail.com](mailto:bprotsenko@gmail.com)