Biodiversity for the National Parks

Capstone Project

Data in species_info.csv

Species data

There are 5,541 species in the National Parks.

The csv file contains this kind of information about species living in National Parks:

1. category: categorisation of the animal as amphibian, bird, fish, mammal, nonvascular plant, reptile, vascular plant.
2. scientific_name: the scientific name of a species in Latin.
3. common_name: the name a species is commonly referred to.
4. conservation_status: either, there is no need for concern about a species survival, then this field has no value (NaN); otherwise, we can see if the species is a species of concern, threatened, endangered, or in recovery.

Conservation Status I

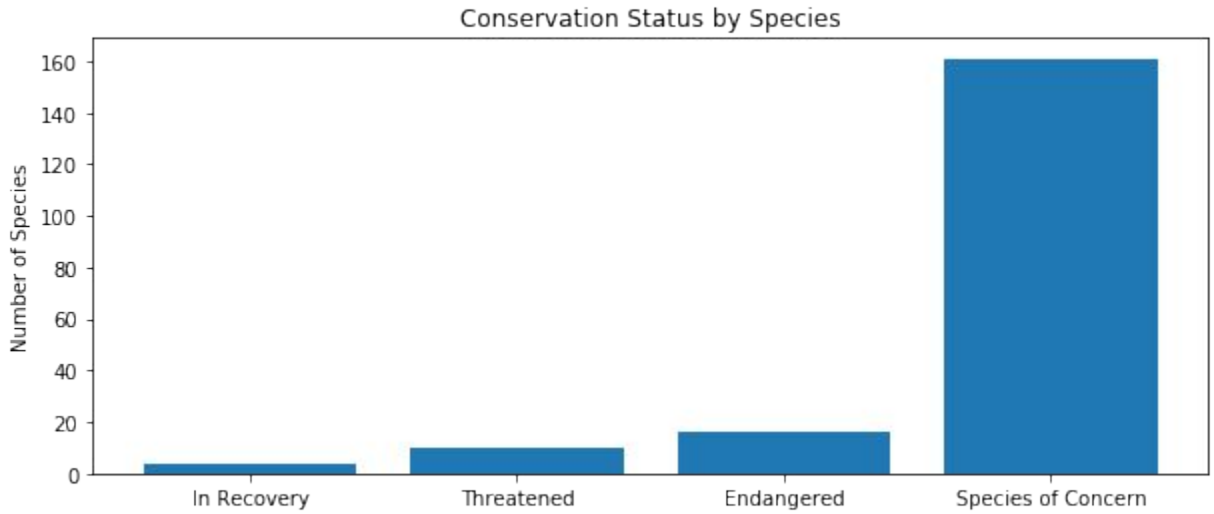
There are 180 species that need to be watched for their conservation status.

151 of the 180 are species of concern.

10 are threatened and 15 endangered.

Only four species are recovering from endangerment.

Conservation Status II



Conservation status by species category

Conservation status by species category: Overview

Category	Not protected	Protected	Percent Protected
Amphibian	72	7	8.86%
Bird	413	75	15.37%
Fish	115	11	8.73%
Mammal	146	30	17.05%
Nonvascular Plant	328	5	1.5%
Reptile	73	5	6.41%
Vascular Plant	4,216	46	1.08%

Conservation status by species category: Mammals vs Birds

Different species have different risks of endangerment. Are these differences significant?

The percent of mammal species in need of protection is higher than that of birds: 17% for mammals vs 15% for birds.

To see if this is a significant difference, we can perform a Chi Square test, as we compare categorical datasets. Our null-hypothesis is that there is no significant difference between the two datasets.

The p-value in this case - comparing birds to mammals - is 68.76%, so we cannot reject the null hypothesis. The difference is not significant, meaning it is due to chance.

Conservation status by species category: Mammals vs Reptile

The percent of protected reptile species is even lower than that of birds. Is the difference significant, when we compare reptiles to mammals?

We can again perform a Chi Square test to compare the two categorical datasets.

In this case, the p-value is 3.84%. So we can say with more than 95% certainty that the two datasets are significantly different.

Over all, the risk of endangerment for different categories of species living in the National Parks shows a significant difference: National Park conservationists should focus their work on mammals, as their need for protection is significantly higher than that of reptiles.

Foot and mouth disease study: Sample size determination

Foot & mouth disease study: Counting sheep I

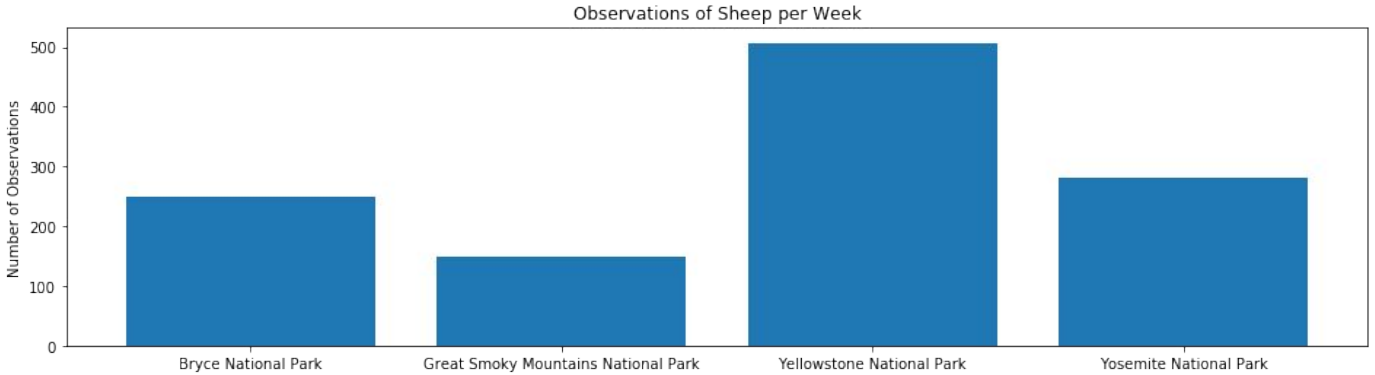
The sheep population in the National Parks suffer from foot and mouth disease. For example, around 15% of the sheep at Bryce National Park are affected.

When we want to know, if the sheep population is e.g. recovering from the disease, we need to know, how many sheep we have to observe to make a significant statement.

We have a list containing the sightings of animals by scientific name for each park. The scientific name is not helpful here, but we can use the species list to see which species are commonly called sheep and also are mammals. With the latter condition, we exclude plants like the Sheep Sorrel.

Foot & mouth disease study: Counting sheep II

Now, we have the scientific names of all sheep species in the National Parks. With this data, we can see, how many sheep were observed per park in a week from the observation list.



Foot & mouth disease study: Calculating the sample size I

We know that 15% of the sheep in Bryce National Park have foot and mouth disease. This is our baseline.

We want to see if the rate of sick sheep has change more than 5%. Our minimum detectable effect is 100 * 5 (minimum detectable percent points) / 15 (baseline) = 33%

Putting this into the sample size tool, we get a result of 890. This means, we need to observe 890 sheep per park to see if the rate of sheep affected by foot and mouth disease has change for more than 5% from the baseline of 15%.

Foot & mouth disease study: Calculating the sample size II

Comparing that information to the observations of sheep per week, we can calculate, that we will need:

- Almost two week (1.76) at Yellowstone National Park
- Over three and a half weeks (3.56) at Bryce National Park
- Over three weeks (3.16) at Yosemite National Park
- Almost six weeks (5.97) at Great Smoky Mountains National Park