Biodiversity Capstone Project

by Werner Wasmayr

1. Data in „*species_info.csv*"

The data in the file '*species_info.csv* contains **5541** species.

There are seven types of species: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant' and 'Nonvascular Plant'

In the conservation status there are four values present when beginning the analysis: 'Endangered', 'In Recovery', 'Species of Concern' and 'Threatened'

2. Objective I: Analyze species conservation status

The objectives of the first part of the project are to visualize the conservation status by species.

After investigating the data we will run a chi-squared test for significance, comparing results of Mammals versus Birds, Mammals versus Reptile...

Finally, there will be a recommendation for conservationists concerned about endangered species.

3. Graph I: Conservation Status by Species

Conservation Status by Species
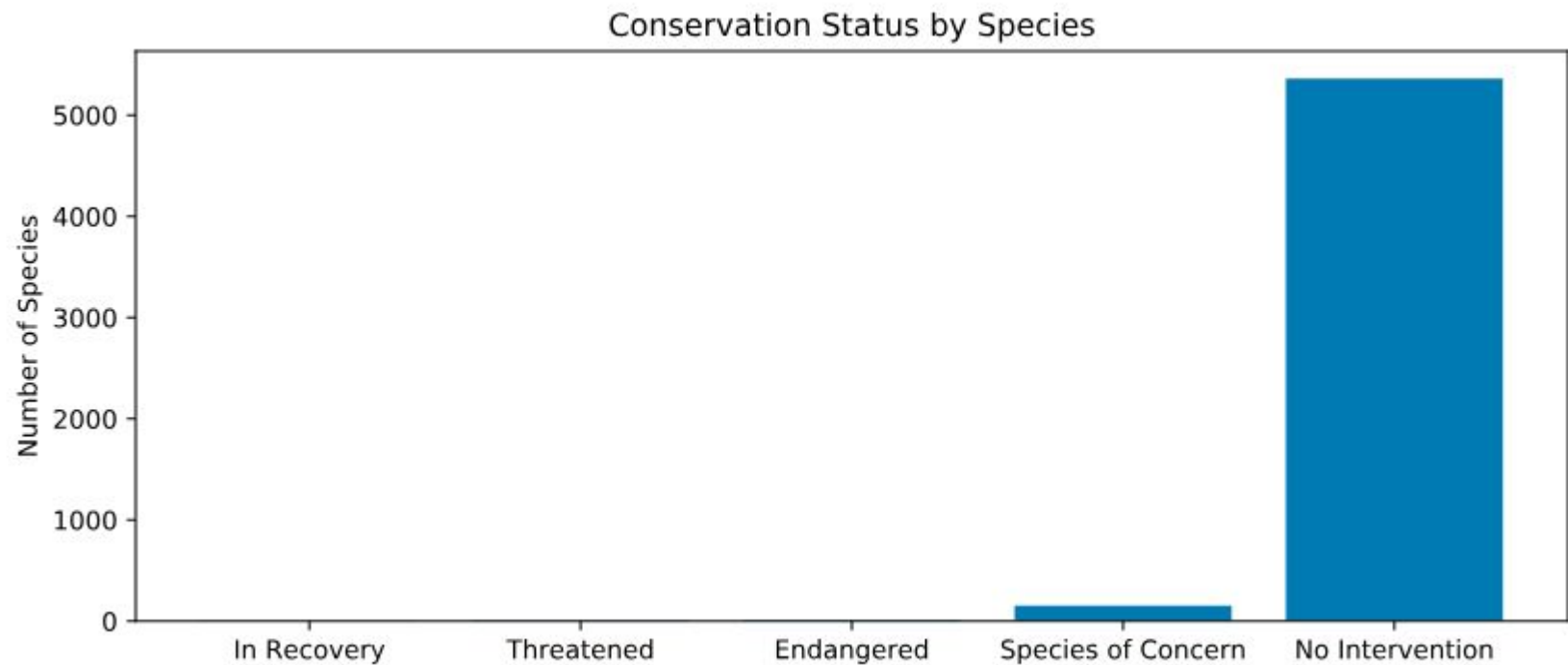
4. Significance calculations for endangered status between different categories of species

Category	Not Protected	Protected	Percent Protected
Amphibian	72	7	0,088608
Bird	413	75	0,153689
Fish	115	11	0,0873
Mammal	146	30	0,170455
Nonvascular Plant	328	5	0,015015
Reptile	73	5	0,064103
Vascular Plant	4216	46	0,010793

5. Recommendation for conservationists concerned about endangered species

The question was: "are certain types of species more likely to be endangered?".

According our chi-squared test, the answer is **yes**! We calculated a p-value of ~0.038, which **is** significant.

That means, that certain types of species **are** more likely to be endangered than others.

The recommendation is to focus on highly endangered species **Mammals** and **Birds**.

6. Objective II: Foot and mouth disease analysis

The objectives of the second part of the project are to visualize the number of sightings of sheep at each of the four national parks under investigation.
Based on the observed data we will calculate a proper sample size and the needed time to observe the sheep.

The data in the file '*observations.csv*' contains **1188** sightings observed in each of the national parks over the past 7 days.

7. Graph II: Observation of Sheep per Week
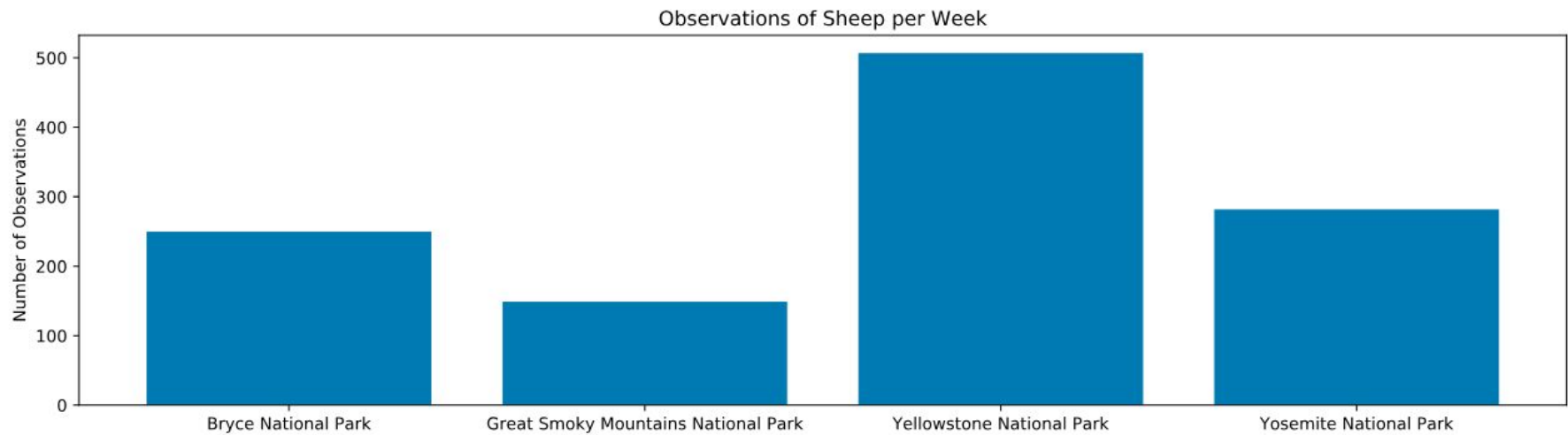
8. Sample size determination for the foot and mouth disease study

The sample size was calculated with the following parameters:

- Baseline conversion rate: 15%
- Statistical significance: 90%
- Minimum detectable effect: 33%

The result is a sample size of: **890**

9. Number of weeks needed to observe sheep

With the sample size of 890 and the sightings in the parks, here are the results of weeks needed for observation:

- Bryce National Park: 890 / 250 = approximately **3.5 weeks**

- Yellowstone National Park: 890 / 507 = approximately **1.75 weeks**