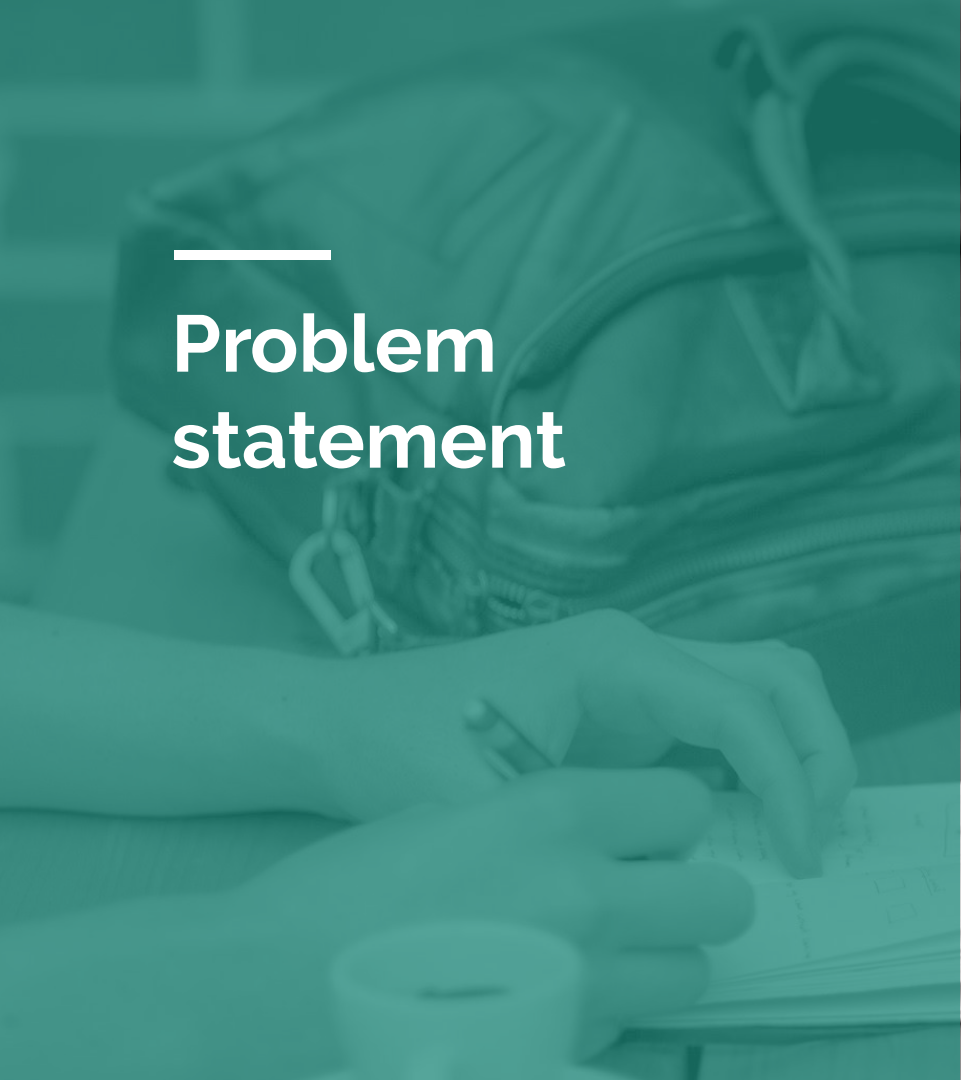
Capstone Project: Biodiversity

National Parks Service Project

Sergio Garcia Puga - November 2019



Problem statement

Analyze data on endangered species from several different parks.

We are analyzing the conservation statuses of the species to see is there are any patterns to the species that become endangered..

Data Description

We are analysing the conservation status of 5541 different species.

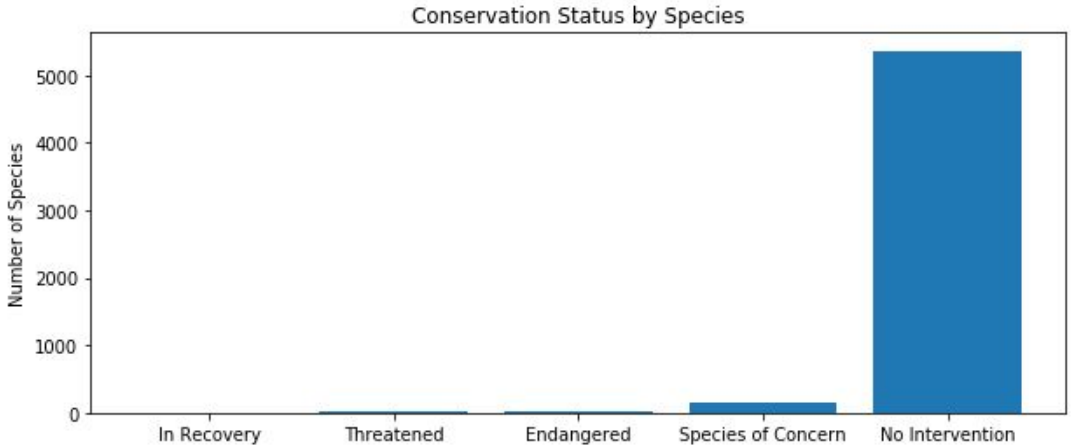
This species belong to 7 categories:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

1. Conservation Status by Species

We are studying 5 different statuses: In Recovery, Threatened, Endangered, Species of Concern and No intervention. We clearly see that most of the species do not need any intervention.

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10



2. Are certain types of Species more likely to be endangered?

After studying more in detail our dataset, we see that most of the species not in danger belong to Vascular Plant and only 1% are in danger.

On the other hand, we see that Mammal and Bird have the highest percent of species in danger: 17% and 15%. Are Mammal more likely to be endangered than Bird?

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

2. Are certain types of Species more likely to be endangered? (continued)

Are Mammal more likely to be endangered than Bird? Is the difference between Mammal and Reptile significant? We have we have performed a Chi Square Test.

In [26]:
```
chi2, pval, dof, expected = chi2_contingency(contingency)
print("P-value for our Chi Square Test between Mammal and Bird is " + str(pval))
```

```
P-value for our Chi Square Test between Mammal and Bird is 0.6875948096661336
```

In [28]:
```
chi2, pval, dof, expected = chi2_contingency(contingency)
print("P-value for our Chi Square Test between Mammal and Reptile is " + str(pval))
```

```
P-value for our Chi Square Test between Mammal and Reptile is 0.038355590229698 98
```

3. Should be conservationist concerned?

We started with a data divided in 5 categories and then we have divided this categories in 2. Our approach discarding those species not in danger it is correct. However, we are aggregating all the rest of species in 1 group so our analysis is not considering any trend: for example those species in recovery or those species that move down from one category to another.

Omitting all the above and based on our results, we can see that Mammal and Bird are the species that conservationist should be more focused.

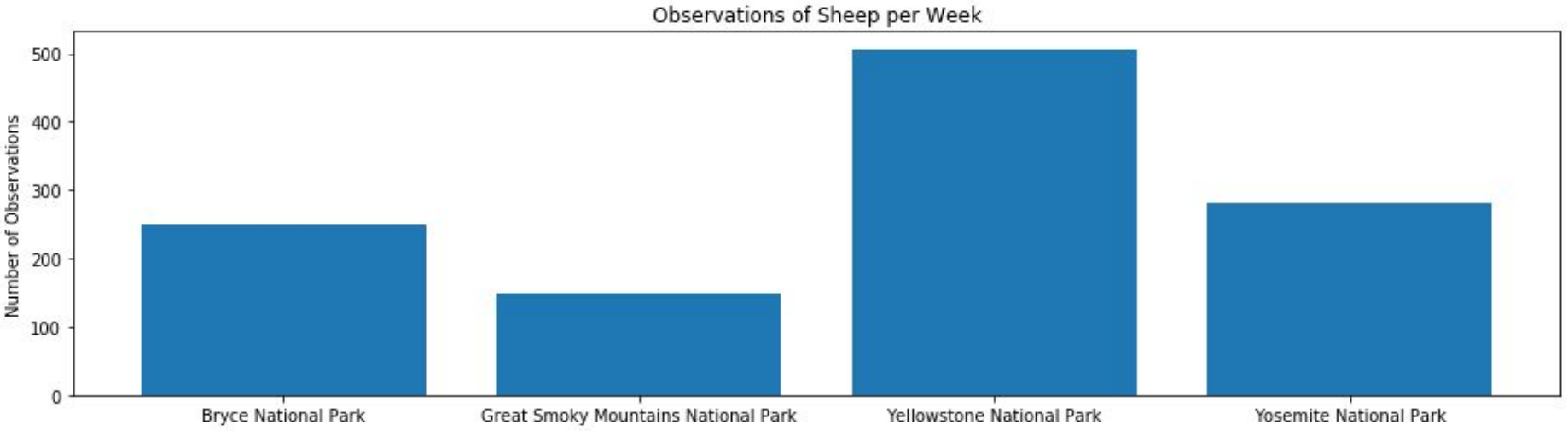
Observation Analysis

Sheeps observations: foot and mouth disease study.

Yellowstone National Park have been running a program to reduce this disease. We want to know how many sheeps we need to observe in order to know if this program is working with a 90% of confidence.

1. Sheep observations

Below are the Sheep sightings at different national parks:



2. What is the sample size we need?

We know that 15% of sheeps at Bryce Park have foot and mouth disease. We want to detect reductions of at least 5 percent. Using a sample size calculator we know that we need a sample 870 observations.

This will take:

- 3.5 weeks at Bryce Park
- 1.7 weeks at Yellowstone Park
- 5.8 weeks at Great Smoky Mountains Park
- 3 weeks at Yosemite Park