

Species Analysis of National Parks

By Conor Murphy

Raw Data

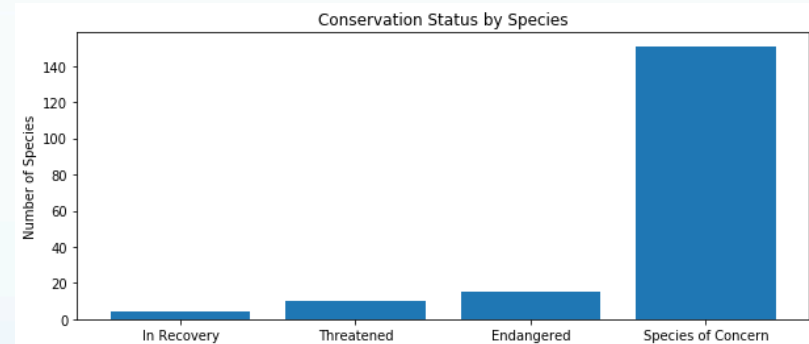
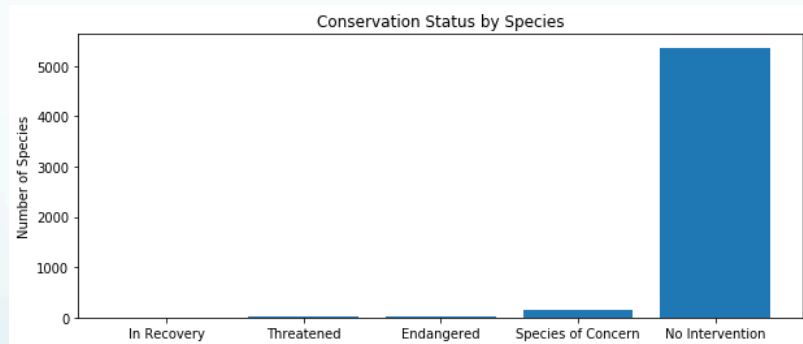
- 'species_info.csv' was provided by the National Parks Service.
- It contained:
 - The species' scientific name,
 - The species' common name,
 - The species' conservation status.

- Initial points from the data:
 - There are seven species categories; Mammal, bird, reptile, amphibian, fish, vascular plant and nonvascular plant.
 - There are 5541 individual species documented.
 - Of these, only 15 are currently listed as Endangered.

Analysis of Conservation Status

Over 96% of species need no intervention

Of species under watch, 8.3% are endangered



- A protected species is defined as one having a conservation status other than 'No Intervention'.
- Below is a table showing the percentage of protected species per category.
- Mammals have the highest percentage of protected species with vascular plants the least.

ID	Category	Not Protected	Protected	Percent Protected
0	Amphibian	72	7	8.8608
1	Bird	413	75	15.3689
2	Fish	115	11	8.7302
3	Mammal	146	30	17.0455
4	Nonvascular Plant	328	5	1.5015
5	Reptile	73	5	6.4103
6	Vascular Plant	4216	46	1.0793

Are some specie categories inherently in more danger?

- We can see from our current data that some categories are in more need of protection than others.
- But is any of this significant? Are the differences the result of chance or do they tell us something about the categories?
- To check for significance, we will perform a chi-squared test.

Bird-mammal chi-squared

- Mammals needed the most protection with birds in second.

ID	Category	Not Protected	Protected	Percent Protected
1	Bird	413	75	15.3689
3	Mammal	146	30	17.0455

- Are mammals significantly more likely to be endangered?
- We perform a chi-squared test to find out.
 - p-value = 0.687594809666
- Our p-value is greater than 0.05. This indicates that we cannot reject the null hypothesis that it is due to chance.
- Therefore, mammals are not significantly more likely to be endangered.

Reptile-mammal chi-squared

- 6.4% of Reptile species are in need of protection. Is mammals' 17% a significant increase?

ID	Category	Not Protected	Protected	Percent Protected
3	Mammal	146	30	17.0455
5	Reptile	73	5	6.4103

- Let's find out with a chi-squared test.
 - p-value = 0.0383555902297
- In this case, the p-value is less than 0.05. This means that we can reject the null hypothesis and therefore mammals *are* significantly more likely to be endangered.

What's the bottom line?

- The previous result is important. It confirms that some species are indeed more likely to be endangered than others.
- With this knowledge, the approach to specie protection can be tailored to certain categories.
- For example, more resources can be focused on endangered mammals compared to reptiles or both plant categories.

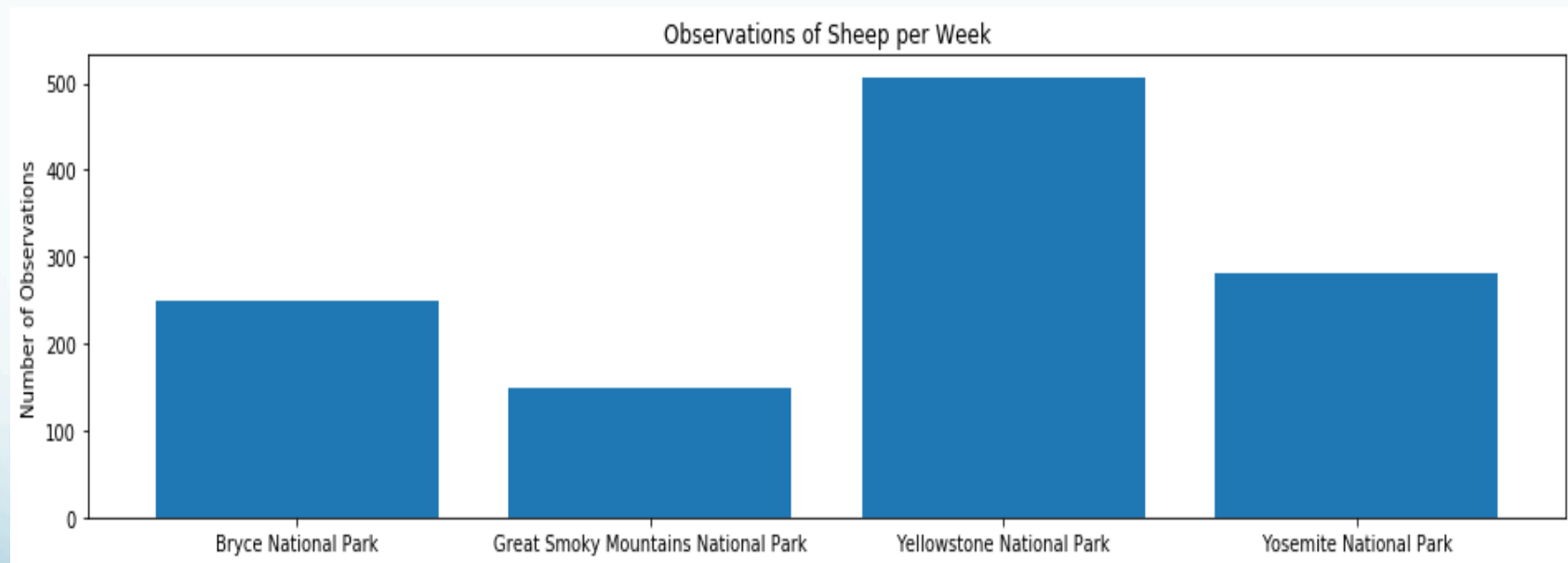
Foot & Mouth Reduction Project

- The goal of the new program is to reduce the cases of foot & mouth disease in sheep by 5%.
- National Parks Service have provided another data set, 'observations.csv', tracking the observations of each species across the national parks.
- It contains:
 - The name of the species,
 - The park where it was observed,
 - The number of times it was observed in this park.

- Cross-referencing the sheep species with 'species.csv', the below chart indicates the number of times a species of sheep was observed in each park per week:

ID ▼	Park Name ▼	Observations ▼
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

- The previous table is also displayed graphically below:



Determining Sample Size

- In order to test that our new foot & mouth regime is effective, we must ensure that the sheep sample size is sufficient.
- The calculator that helps determine sample size needs three inputs:
 - Baseline rate: Our current rate of foot & mouth among sheep,
 - Minimum Detectable rate: The minimum percentage change from the baseline that we require to see in order to deem the experiment a success,
 - Statistical Significance: The probability that we can reject the null hypothesis that any perceived change is due to chance.

- Last year, 15% of the observed sheep were reported to have foot & mouth disease. Therefore,

$$\text{Baseline} = 15\%$$

- We said earlier that we wanted a 5% drop in reported cases. The minimum detectable rate is a percentage of the baseline so

$$\text{MDR} = (5/15) * 100 = 33\%$$

- For this test, we set our statistical significance to 90%, so

$$\text{Statistical Significance} = 90\%$$

- Using the sample size calculator and inserting the values from the previous slide, the minimum sample size to be used in this experiment is 890 sheep.
- Therefore, given the number of sheep observed in each park, the below table indicates how many weeks the experiment must run to reach the sample size:

ID	Park	Weeks of Experiment
0	Bryce Nation	3.56
1	Great Smoky	5.973154
2	Yellowstone	1.755424
3	Yosemite Na	3.156028

Thanks for reading!