



For the Love of Mammals!

Kristin Leach
Capstone Project

About the Data

- The csv file provided for this capstone project lists 5541 different species of plants and animals all categorized into five animals: mammals, birds, reptiles, amphibians, fish, and two plant types: vascular and non-vascular plants. The columns of this data frame include the category (previously described), the scientific name, common names, and the conservation status.

Significance Calculations

- Significance calculations have meaning if you have a null hypothesis first. Our null hypothesis in this case is that there is no statistical significance between species types and endangered status of species categories. (ex: Vascular plants are just as likely to be endangered as mammals. However, based on the results of the chi squared test, there are some significant differences.
- First, why the chi 2 test? Since our results are categorical, a chi test is called for (is protected or not?). By comparing two species categories, we can determine if there is a significant difference based on the p value. (significant if $p < 0.05$ generally)

Recommendations for Conservationists

- Two chi test comparisons were made:
- 1. Reptiles and mammals
- 2. Mammals and fish
 - Although the mammal and fish comparison was not significant, the reptiles and mammals were! Recommendations to the conservationists would be to focus particularly on mammals and fish, and reptiles would be less of a concern. However, any new data should be continuously monitored and appropriate efforts should be able to be redirected as needed.

```
Out[32]:
```

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088805
1	Bird	413	75	0.152819
2	Fish	116	11	0.087902
3	Mammal	146	30	0.170455
4	Nonvascular Plant	320	5	0.015615
5	Reptile	78	6	0.064103
6	Vascular Plant	4216	46	0.010935

```
In [34]: contingency = [[30, 146], [75, 413]]

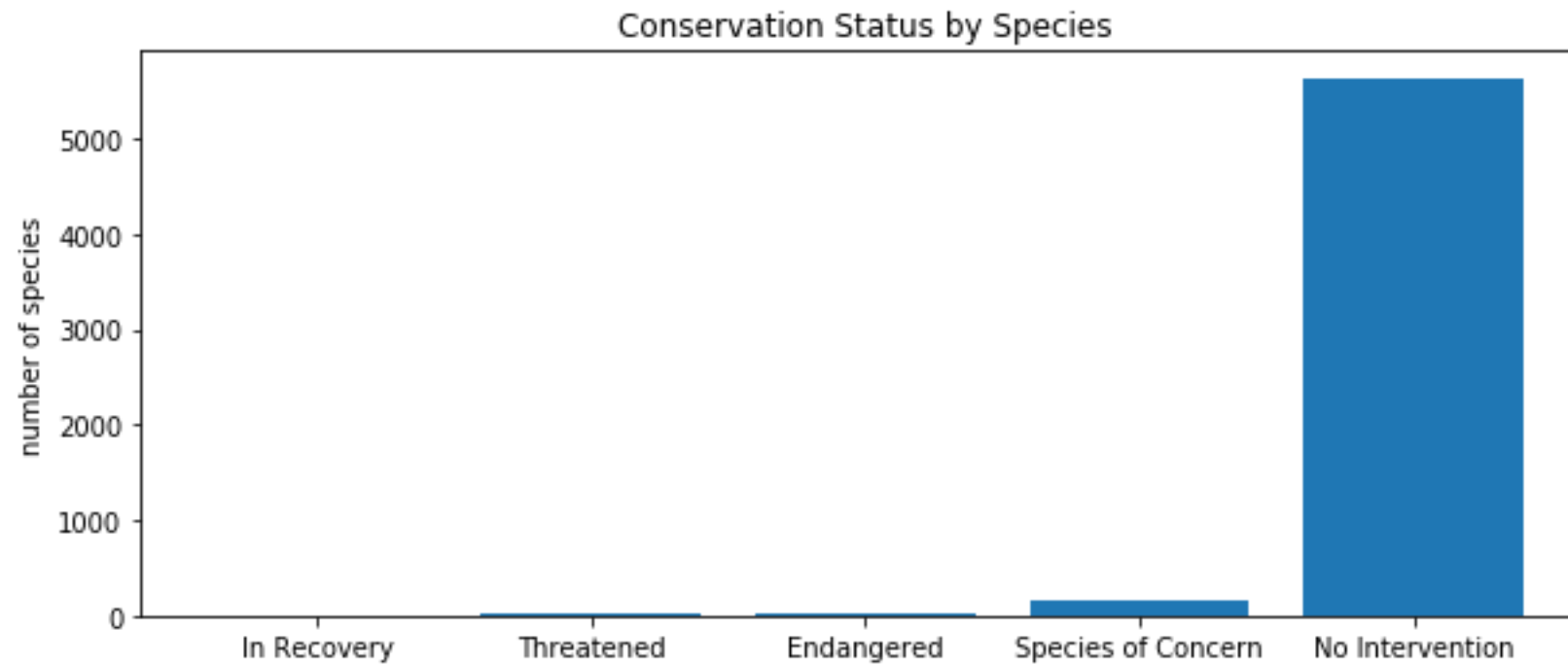
In [35]: from scipy.stats import chi2_contingency

In [36]: chi2_contingency(contingency)

Out[36]: (0.1617014831654857, 0.6875948096661336, 1, array([[ 27.8313253, 148.1688747],
               [ 77.1688747, 410.8313253]]))

In [37]: contingency = [[30, 146],
                        [5, 73]]
chi2_contingency(contingency)

Out[37]: (4.209183096203645, 0.03835559022969890, 1, array([[ 24.2519685, 151.7480315],
               [ 10.7480315,  67.2519685]]))
```



- Bar graph showing the total species count for each conservation status. Luckily, the ‘No intervention” is the highest by more than an order of magnitude. Species of concern is the next highest, but that seems to be most of the information that can be gleaned from this graph.

Sample Size Determination

- 15% of sheep at BNP have foot and mouth disease (from the scientists). To detect at least 5 percentage points, a minimum detectable effect must be calculated, and then the numbers applied to the Optimizely website to determine sample size. 90% is the default level of significant and that will remain the same. The baseline conversion rate is the starting percent (15%), and all that is left is figuring out the minimum detectable effect.
- That is figured out by the following formula:
 - $(100 * 0.05) / (15\% - 10\%)$ (which is the detection level desired) / the baseline conversion
- In this case, the minimum detectable effect is 33.33. When that number is typed in, the required sample size is 520.

Observations of Sheep per Week

