Biodiversity for the National Parks: Examined

Jeff Waltz

Goal

- Are there patterns or themes to the types of species that become endangered?

Data – Overview

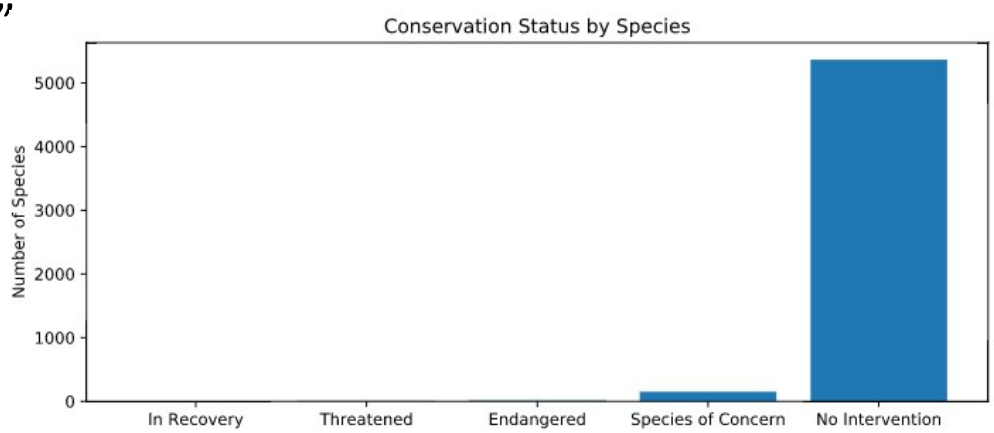
- National Parks Service maintained an extensive database on the conservation status of over 5500 plants and animals spanning 7 species.
- While the dataset was compressive, it only contained complete entries for animals with a labeled conservation status (analysis required filling in the incomplete data).
- Example of raw data—

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Data – Rearranged

- While our data was largely complete, it wasn't organized in a sufficient manner for our analysis.
- After creating new tables we learn that the National Park Service describes four conservation statuses and can now view how many different plants and animals fall into each category.
- Plotting this data we see the scale to which most plants and animals have a conservation status of "no intervention"

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



Analysis – More Data Manipulation

- As we continue to synthesize the data by creating new tables and data points (i.e. “protected” vs “not protect”), we begin to represent the data categorically, implying possible relationships.
- In the below table we observe large discrepancies in the percent of animal and plants that are, and are not, protected, when grouped by species.

	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

Data Driven Hypothesis

- With this insight we can now propose a meaningful hypothesis – a particular species is more likely to be endangered than another.
- We use a Chi Square Test to test our hypothesis based on the type of data we are comparing – numerical, two data points associated with each category of species.
- When compare the reptiles to mammals we our hypothesis fails producing a P value of .68 well about .05 to invalid our null hypothesis.
- However, when we compare reptiles to mammals we find that we do have a significant result, a P value of 0.038, disproving our null hypothesis and confirming that, at least in this comparison – mammals when compared to reptiles – are more likely to become endangered.

Further Testing

- While we can demonstrate that particular combinations of species are more likely to become endangered than others based on Chi Tests I don't believe we should conclude the obvious – some species are more likely to become endangered than others.
- After Chi testing all species combinations, a very strong trend emerged – Vascular Plants and Non Vascular Plants showed statistically significant differences when compared against each species of animal (confirming our hypothesis), while only one animal to animal species comparison yielded significant results.
- Upon separating plants from our dataset, a total of 14 combinations of species / Chi Tests will only yield one significant result – mammals and reptiles.

Recommendations for Conservationists

- Examining species according to protected and non protected status does yield interesting results, though it would appear meaningful trends cannot be properly assessed with such limited data points.
- Separating plant and animal conservation statuses might be helpful in further studies as the datasets were statistically significant when compared against any animal species.

Foot and Mouth Disease – Determining Sample Size

- Since our future Foot and Mouth test is trying to determine whether the program is working, the data that we will be collecting is similar to an A/B Test – Yes or No. These types of sample size calculations do not involve margin of error inputs since we are making a comparison of two constant values.
- By using the calculator we know can calculate our ideal sample size given the following parameters:
  - Baseline: is the approximate percent of the population who have Foot and Mouth disease – last year recorded at 15%.
  - Statistical significance: is given at 90%, which indicates the probability that the margin of error does not contains our findings.
  - The minimum detectable effect is the percentage off the baseline that our test wants to be able to statistically capture, in our example – 5% change off a baseline value of 15% is 33.33%.
- Inputting these variables into the calculator we find that the ideal sample size for our purposes is 510 sheep.

Graphs Created