Capstone Project: Biodiversity for the National Parks

Pui Yan WU
17th February 2019

Table of content

Descriptions on data species_info.csv

Total number of unique species in the dataset - 5541 (based on scientific names)

Columns in the dataset include: “category”, “scientific_name”, “common_names” and “conservation_status”

Species categories: “Mammal”, “Bird”, “Reptile”, “Amphibian”, “Fish”, “Vascular Plant” and “Nonvascular Plant”

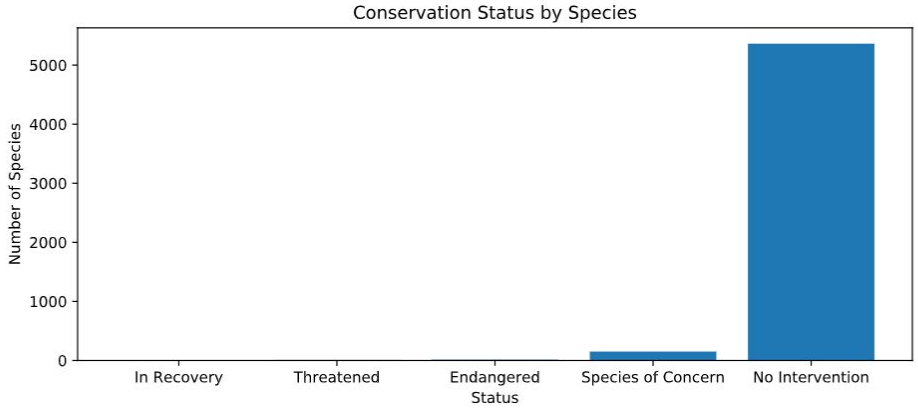
Categories of conservation status and quantity of each category grouped by unique scientific names -

1. Endangered: 15
2. In Recovery: 4
3. Species of Concern: 151
4. Threatened: 10
5. No Intervention: 5363*

*There was initially no category for the 5363 species and their conservation status was “None” or “NaN”. To make comparison between categories easier, this additional category was created for species with the “NaN”. However this may cause errors since we are unsure if they actually require no intervention or if it is due to lack of information.

According to the chart below, about 96.79%* of all species require no intervention, about 3.18% are endangered, threatened or are species of concerns. Unfortunately, there are 0.09% of species are in recovery. These figures also suggest that more protections should be done to help endangered species to recover.

* all numbers in this ppt are rounded to the closest 2 decimals.



To find out more on conservation status within each species category, a new column called “is_protected” is created. For species with status “No intervention”, they are categorized as “not protected” while all other categories are grouped as “is protected”.

Afterwards a pivot table is created to show the number and percentage of each species category whether they are protected or not.

Based on the figures below, apparently “Mammal” and “Bird” have a higher percentage of being protected, 17.05% and 15.37% respectively as against “Nonvascular Plant” (1.50%) and “Vascular Plant” (1.08%). Overall animals, “Amphibian”, “Bird”, “Fish”, “Mammal” and “Reptile”, seem to have a higher percentage of being protected (average 11.29%) in comparison to “Nonvascular Plant” and “Vascular Plant” (average 1.29%). For animals, they all have 5 or more % being protected whereas plants have less than 2%.

	Category	False	True	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

To further analyse if there is significant difference in terms of protection status between categories (here “Mammal” is used to compare with other animal species), chi-square tests are performed on the following pairs with a 0.05 as the significance level. Below shows the findings.

1. “Mammal” and “Bird” - 0.69, > than 0.05 and is not significantly different
2. “Mammal” and “Amphibian” - 0.04, < than 0.05 and is significantly different
3. “Mammal” and “Fish” - 0.06, > than 0.05 and is not significantly different
4. “Nonvascular Plant” and “Vascular Plant” - 0.66, > than 0.05 and is not significantly different
5. Animals and Plants - 3.20175187422e-85, < than 0.05 and is significantly different
    a. Animals: “Amphibian”, “Bird”, “Fish”, “Mammal”, “Reptile”
    b. Plants: “Nonvascular Plant”, “Vascular Plant”

Conclusions - For “Mammal” and “Amphibian”, “Animals” and “Plants”, we can conclude that certain types of species are more likely to be endangered than others. And such difference is influenced by some non-random cause instead of result of chance.

Some recommendations on protecting endangered species

- Pay extra attention to educating the public on the values of species which are less protected, in particular reptiles and plants, such as making TV programs, films or advertisements
- Introduce volunteer programs to children/teenages, or as family activities or short terms internships to work at national parks
- Go around schools and have exhibitions/presentations on the endangered species

Descriptions on sample size determination for the foot and mouth disease study

- Baseline conversion rate - 15%
- Minimum detectable effect - 100*(20-15.)/15 = 33.3%
- Statistical significance - 90%
- Sample size per variant - 870
- Minimum required weeks of observations at Yellowstone National Park - 870/507 = 1.72 or rounded to 2 weeks
- Minimum required weeks of observations at Bryce National Park - 870/250 = 3.48 or rounded to 4 weeks

Appendix: Sample size determination graph