Biodiversity Data Analysis

By Chunjin Gao

03/22/2018

Whats in Species.csv

- The table contains 4 columns:

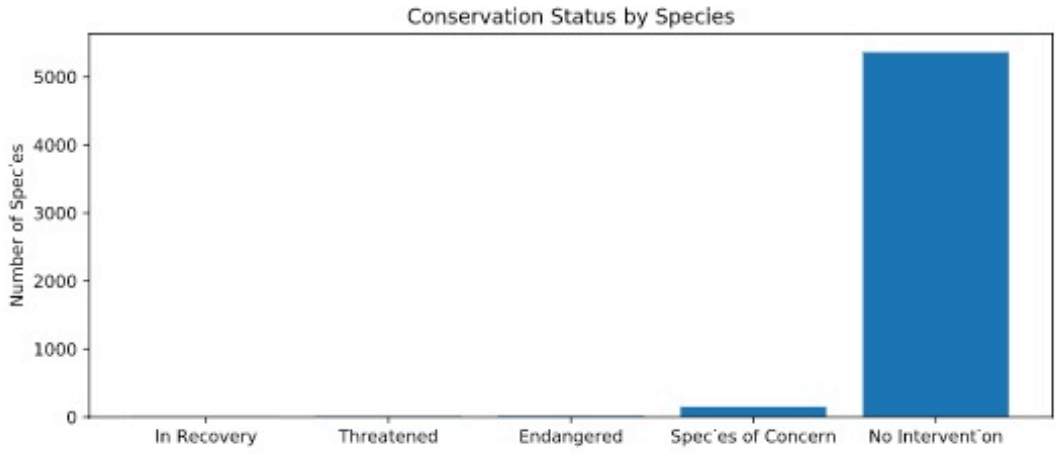
- 1. Category - mammal, Bird, Reptile, Amphibian, etc.
- 2. Scientific Names of species habitats in the National Parks
- 3. Common Names of species that we commonly refer to
- 4. Conservation Status - whether the species is endangered or concerned

Findings from Tables

- Most wolfs are endangered species and most hawks are species of concerned.
- All categories has some species that are either endangered or species of concerned.
- Most Species are neither species of concerned nor endangered species

Statistic of Species Status

Conservation Status	Scientific Name
Endangered	15
In Recovery	4
Species of Concerned	151
Threatened	10



Conservation Statistics By Category

Conservation Status are separated and grouped by different categories

	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

Chi Square Test On different Categories

- Null Hypothesis: Difference on protected percentages between mammals and birds is due to chance
- P-value = 0.6876 >0.05 There is no significant difference, Null is rejected
- Null Hypothesis: Difference on protected percentages between mammals and reptile is due to chance
- P-value = 0.0384 <0.05 There is significant difference, Null is accepted
- ✓ Conclusion: Certain type of species are more likely to be endangered than others.

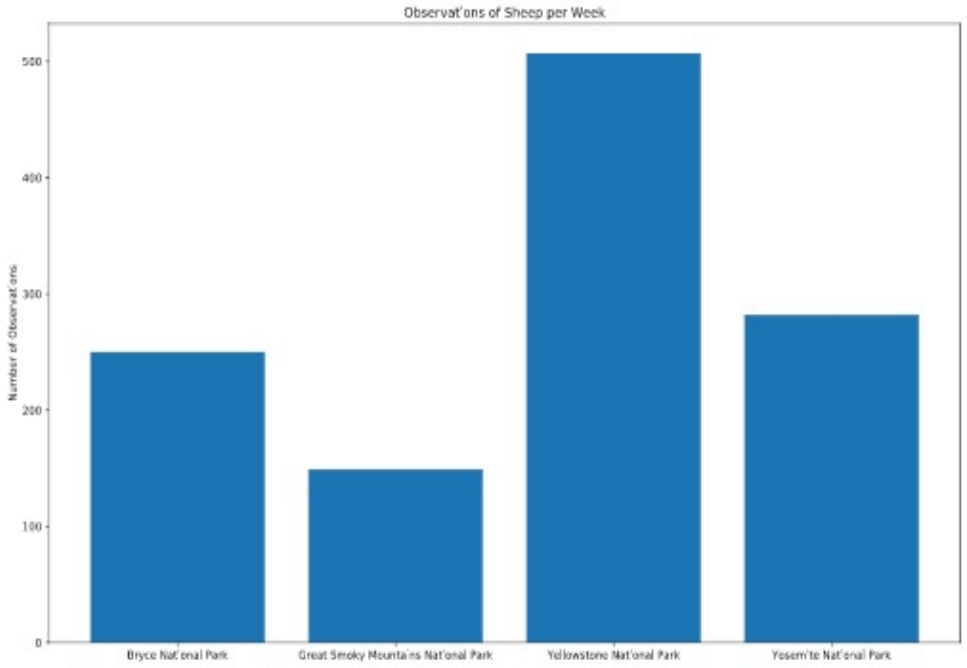
Recommendation

- Authorities need to pay more attention to species that are easier to be endangered and those species of concern such as mammals and birds
- Amphibians and reptiles are relatively less possible to be endangered or species of concern but they have smaller numbers of species existed in this area. Therefore, they need to be carefully monitored because they plays significant parts in eco system and small change in these categories will cause large effects

Finding ‘Sheep’ in Observaitions.csv

- The table consist of names of national parks and number of sheeps observed in a week in each park

	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



Foot & mouth Disease

- There are 15% of diseased sheep at the Yellow Stone National Park
- After running a program, the scientists expect to see reduction of at least 5%
- Following data is entered in to sample size calculator:
- Baseline conversion rate: 15%
- Statistical significance: 95%
- Minimum detectable effect: 33.3% (5%/15%)
- Sample Size = 870

How long?

- Further calculation reveled that 1.72 weeks is needed to sight 870 sheep in Yellowstone National Park (870/507)

- In Bryce National Park, 3.84 week is needed to generate 870 sample size (870/250)

The End – Questions?