

Codecademy - Introduction to data analysis

Jonathan Svärdh

Data description of species_info.csv

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

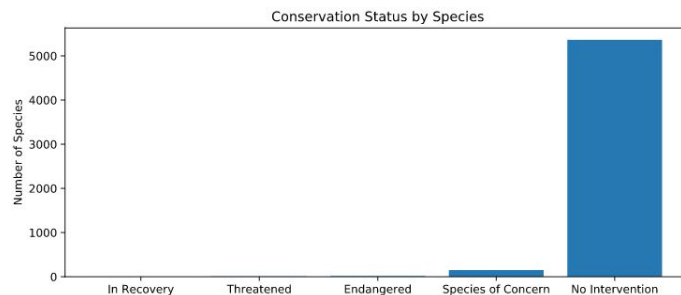
5541 different species have been observed in the dataset.

7 different categories that cover different types of animals and plants observed.

	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

4 different conservation statuses: Species of Concern, Endangered, Threatened, In Recovery and No intervention. Would like to have a “Recovered” status. This for tracking and measuring and also seeing possible progress in helping the wildlife.

15 species are Species of concern with the Categories Reptiles and Nonvascular plants being the category that is least protected today by quantity.



Mammals and Birds seem to be the species that are most likely to be endangered. Since the pval between Reptiles and Mammals was less than 0.05 we can conclude that some species are more likely to be endangered.

Recommendation to Conservationists:

Focus on Mammals and birds as far as prevention work, but also work on Reptiles and Non vascular plants since these are the most endangered in amount.

Sample size (foot and mouth disease)

It is calculated that approx 890 sheep observations need to be made to conclude with singificance

I would say that the sample is too low. This due to the nature of the disease. F&M is highly contagious and should atleast have a statistical significance of 95% rather than 90%. In a perfect world they should test all animals.

One thought would be to start with Yellowstone National Park since it has the most sightings,fastest to get observations and probably the most sheep i.e the most sheep in danger.

