Capstone Project

Analysis of Biodiversity for the National Parks by Collin Morgan

The dataset

Context

The data within this dataset represents a breadth of information around the species located within the National Parks. We can use this information to better understand the work that needs to be done by the National Parks Service to preserve and protect the species within the National Parks.

Content

Initial information provided within the dataset:

- The scientific name of each species
- The common names of each species
- The species conservation status

Info

The data can be used for a wide range of analysis including:

- Conservation status by species
- Likelihood of certain species being more susceptible to becoming endangered
- Observations of specific species by park

Observations - Conservation Status by Species

Conservation Status by Species
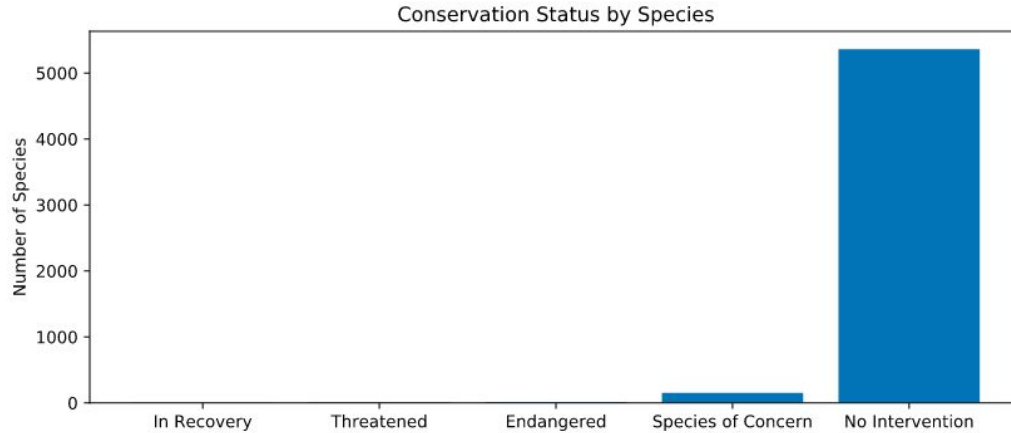
- There are 5,543 recorded species within the National Parks
- Less than 1% of recorded species are considered 'Endangered'
- Most species in the National Parks fall under 'No Intervention'
- The next highest percentage after 'No Intervention' is 'Species of Concern' at 2.7%

Observations - Endangered Status between 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

Table review

- category = Group By species type
- not_protected = conservation_status
- protected = all other conservation_status items
- percent_protected = protected divided by protected + not_protected

Chi2 Testing

- Test between **birds** and **mammals** showed the difference in status **was not** statistically significant
- Test between **reptiles** and **mammals** showed the difference in status **was** statistically significant

Review

- Based on this information there is no significant difference in endangered status between mammals and birds in the National Parks
- Mammals are significantly more likely to be endangered in the National Parks than reptiles

Observations - Foot and Mouth Reduction Effort
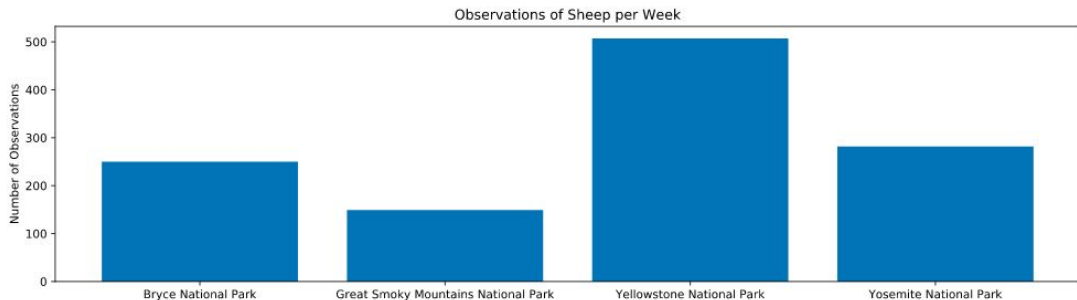

Observations of Sheep per Week

Table review

- The table above shows the observed number of sheep per week for each National Park
- We will use this information to understand the number of sheep we need to observe and for how long in order to see disease reduction

Determining sample size

- baseline = 15
- minimum_detectable_effect = 100*5./15
- sample_size_per_variant = 870
- yellowstone_weeks_observing = sample_size_per_variant/507.
- bryce_weeks_observing = sample_size_per_variant/250.

Review

- Given a baseline of 15% disease occurrence at Bryce National Park, if the scientists wanted to be sure that a >5% drop in observed cases disease in the sheep at Yellowstone was significant they would have to observe at least 510 sheep
- This would take approximately one week of observing in Yellowstone to see that many sheep, or approximately two weeks in Bryce to see that many sheep

"Completed Code in the Learning Environment"
Link to learning environment -
https://www.codecademy.com/courses/biodiveristy-capstone/lessons/protected-status-analysis/exercises/load-and-inspect-dataframe
Link to user profile - https://www.codecademy.com/collinkmorgan