Capstone: Biodiversity for the National Parks

Introduction to Data Analysis

Léon Junique, 15th June 2018

codecademy

Outline

- About the Project
- Description of the data
- Conservation Status by Species
- Protected percentage of each specie category
- Recommendation for conservationists
- Sheeps in the different parks
- Sample size determination for the foot and mouth disease study

About the project

The National Parks Service would like to perform some data analysis on the conservation statuses of endangered species and to investigate if there are any patterns or themes to the types of species that become endangered.

Description of the data

Species_info.csv

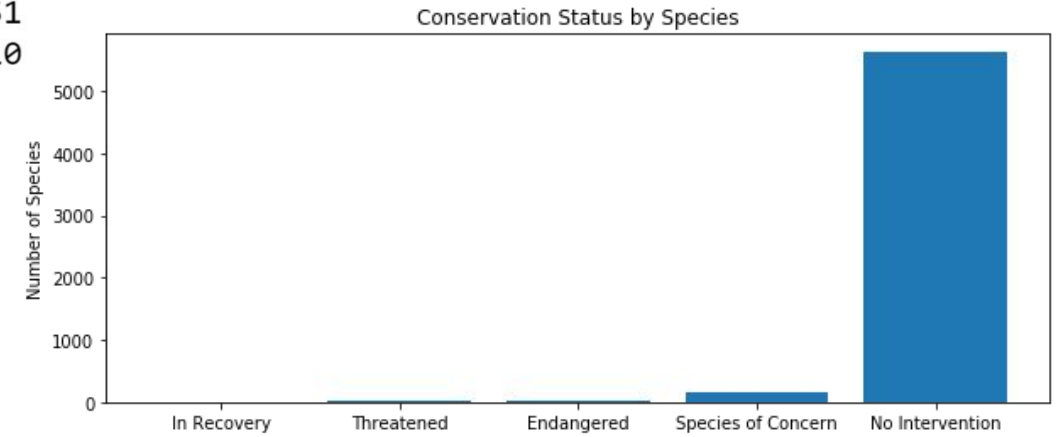
Contains data about different species in the National Parks, including:

- **The scientific name of each species** (5541 different species in total)
- **The common names of each species** (severals common names for each scientific name)
- **The species conservation status** (5 in total: ‘nan’ ; 'Species of Concern' ; 'Endangered' ; 'Threatened' ; 'In Recovery')
- **The category of the specie** (7 in total: 'Mammal' ; 'Bird' ; 'Reptile' ; 'Amphibian' ; 'Fish' ; 'Vascular Plant' ; 'Nonvascular Plant')

Conservation Status by Species

A significant number 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



Protected percentage of each specie category

Mammal (17%) and Bird (15%) are the most protected species. The difference between those two is not significant (Chi squared test).

Vascular Plant (1%) and Nonvascular Plant (1.5%) are the least protected species → Invest in new conservation projects for these 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

Recommendation for conservationists

A need to protect better the other species as well.

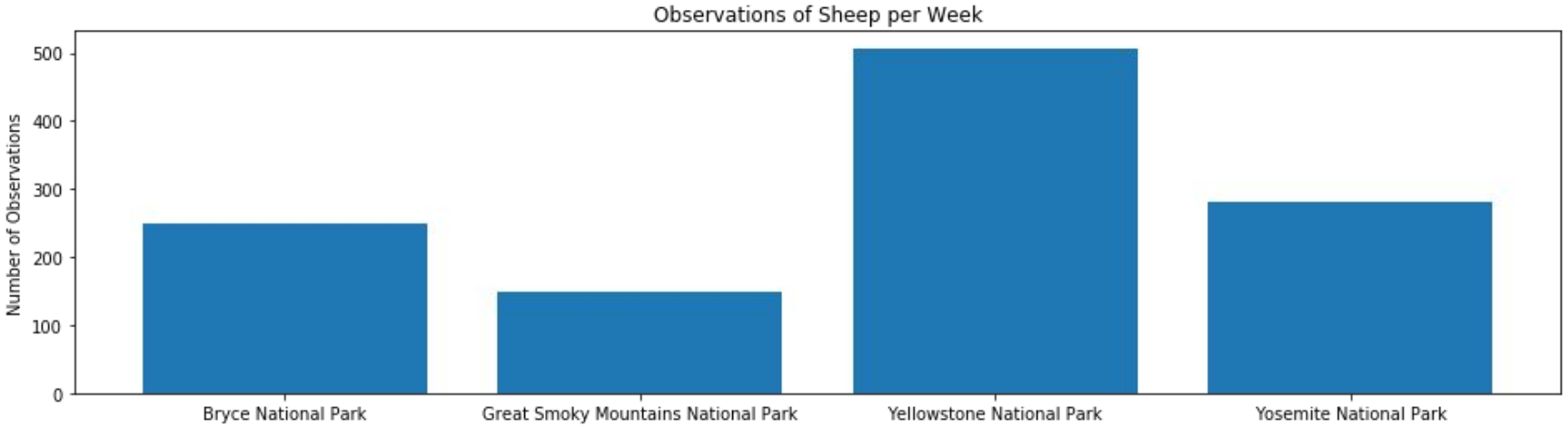
Remember species form a chain together. The disappearing of one in the bottom like the Plants will affect also the other species. Conservationists must look at the big picture.

	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

Sheeps in the different parks

Different types of sheeps across the parks.

The chart shows the observations of sheep per week in the different parks



Sample size determination: foot & mouth disease

Given a baseline of 15% occurrence of foot and mouth disease in sheep at Bryce National Park, if the scientists wanted to be sure that a >5% drop in observed cases in the sheep at Yellowstone was significant they would have to observe at least 870 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.