Capstone Project 2: Biodiversity for the National Parks

Rakshak Lodha

18-Mar-2018

Data – Species

Data in the ‘Species’ file contains following details about various species:

- Category
- Scientific name
- Common names
- Conservation status

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Analyzing species conservation status

Species data grouped by conservation status:

The data in the table below indicates the number of species against each ‘conservation status’ (including ‘No Intervention’ status)

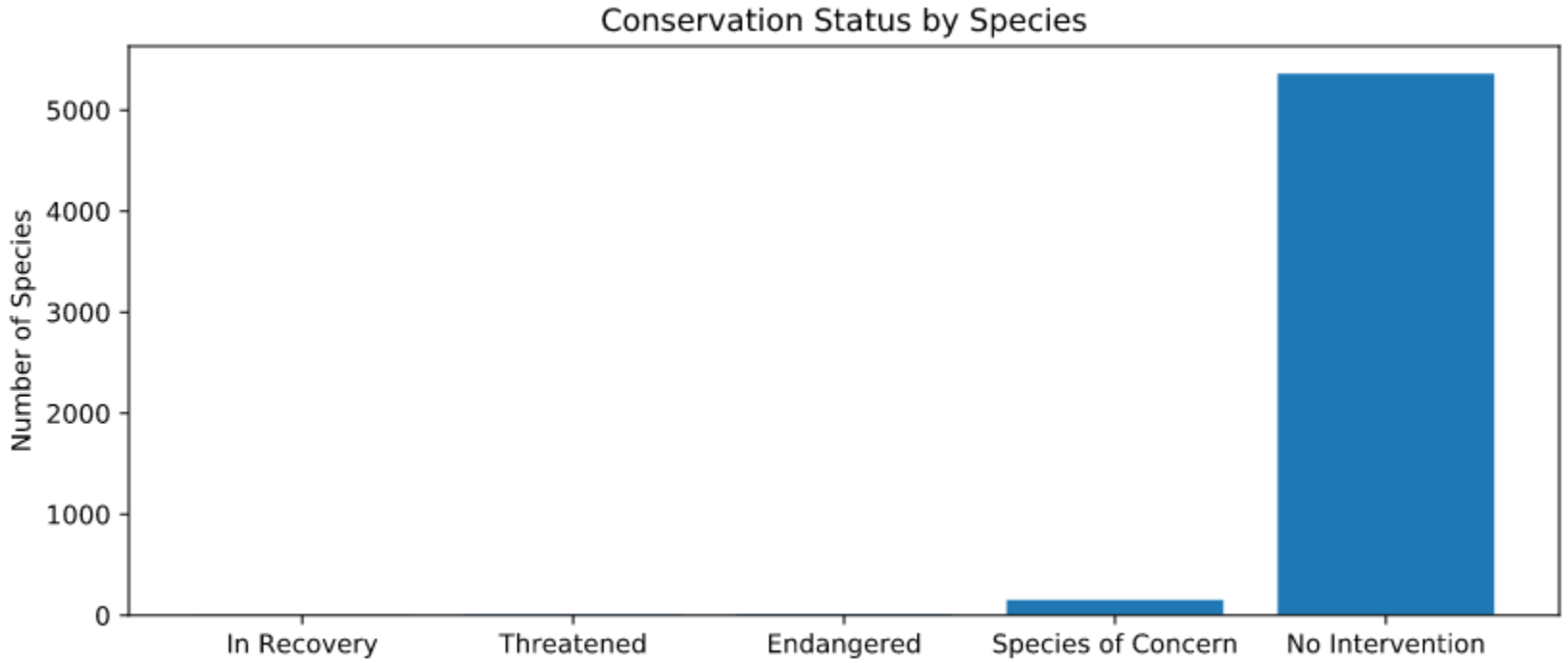
	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

Analyzing species conservation status

Cont…

Bar chart indicating species data grouped by conservation status:

The data in the bar chart below compares the number of species against each ‘conservation status’ (including ‘No Intervention’ status)



Investigating endangered species

Table indicating species data grouped by category and conservation status:

- Species with ‘No Intervention’ conservation status are considered ‘not_protected’
- Species with all other types of conservation status are considered ‘protected’
- Last column indicates percentage of species with ‘protected’ status for each category
- Percentage of protected / endangered species varyies across multiple categories

is_protected	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

Investigating endangered species Cont…

Chi-square test for checking difference in likeliness to be ‘Endangered’ among various types of species:

- Null hypothesis: Difference in percent of endangered species across multiple categories is due to chance
- Test 1 - Comparison of data of ‘Mammal’ and ‘Bird’: p-value > 0.05 indicates that no significant difference is present between these two categories
- Test 2 - Comparison of data of ‘Mammal’ and ‘Reptile’: p-value < 0.05 indicates that significant difference exists between these two categories.
- Conclusion: Null hypothesis has been proved incorrect. Certain category of species are more likely to be endangered than others

Sheep sightings

Analysis of specific type of species: Sheep

- New dataframe ‘sheep_species’ created from ‘species’ containing all the species with common names containing ‘Sheep’ and having category ‘ Mammal’
- Merged the two datasets: ‘sheep_species’ and ‘observations’ in order to generate the table containing sheep related observations
- Grouped the resulting data by each national park

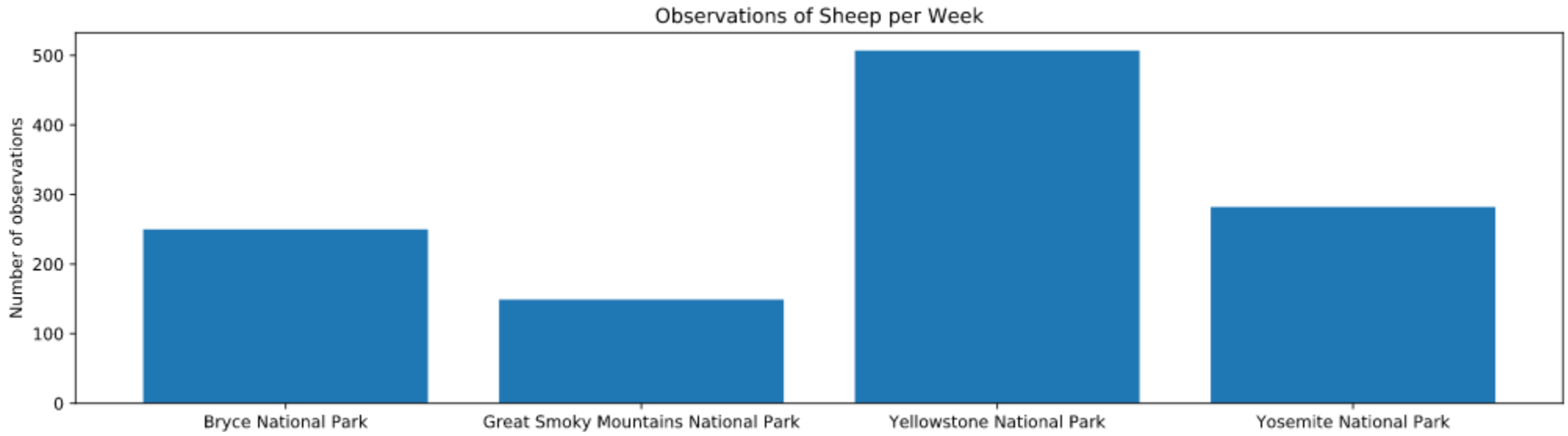
	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Sheep sightings

Cont...

Analysis of specific type of species: Sheep

- Bar chart indicating number of observations of ‘Sheep’ in each national park



Foot and mouth disease reduction effort

Sample Estimation

- Baseline percentage = 15%
- Minimum Detectable Effect = 5*100/15 = 33.33%
- Statistical significance assumed = 90%
- Estimated requisite sample size = 870 (as per sample size calculator)
- Number of weeks required at Yellowstone National Park = 870 / 507 = 1.72
- Number of weeks required at Bryce National Park = 870 / 250 = 3.48

Thank You!