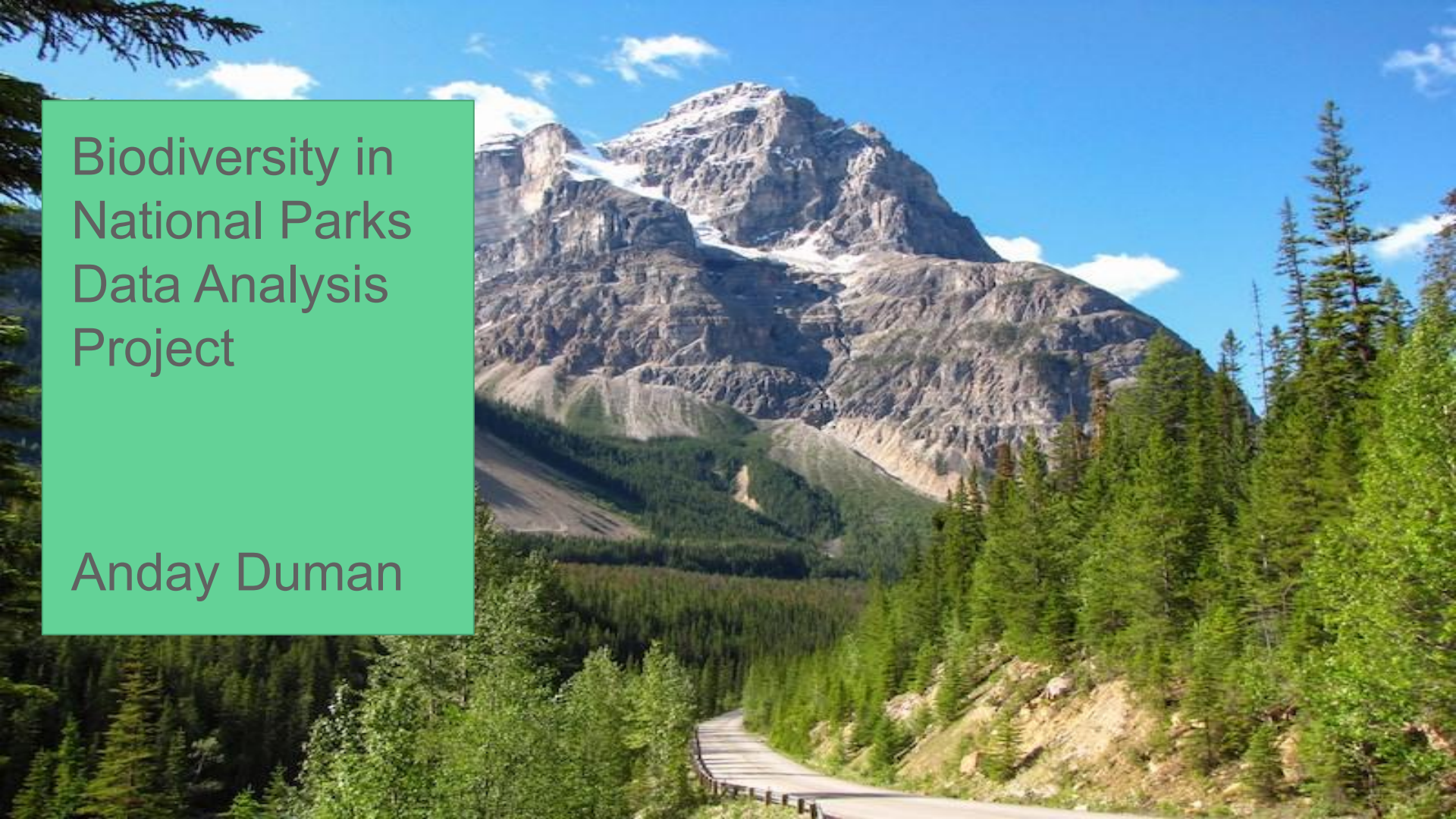


Biodiversity in National Parks Data Analysis Project

Anday Duman

Species Dataset Analysis

The following is relevant information found from species_info.csv file containing species data from United States national parks:

- There is a total of 5824 entries in the dataset
- The dataset contains 5541 unique observed species of organisms.
- Each entry has the following information: Category, Scientific Name, Common Names, Conservation Status
- Different categories of species in the dataset include:
    - Mammal
    - Bird
    - Reptile
    - Amphibian
    - Fish
    - Vascular Plant
    - Non Vascular Plant
- These are the different conservation statuses that can be assigned to an entry in the dataset:
    - Species of Concern
    - Endangered
    - Threatened
    - In Recovery

	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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN
5	Mammal	Odocoileus virginianus	White-Tailed Deer	NaN
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	NaN
7	Mammal	Canis latrans	Coyote	Species of Concern
8	Mammal	Canis lupus	Gray Wolf	Endangered
9	Mammal	Canis rufus	Red Wolf	Endangered

Table 1: Sample entries of species dataframe

Protection of Categories of 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 2: Percent protected of each category

This table presents a breakdown of the percentages of protected and non protected species within each category of organisms in the dataset.

- Lowest protected category is Vascular Plant at 1.07%.
- Highest protected category is Mammal at 17.05%.

Although Mammals appear to be the most at risk, significance tests need to be applied to this information to validate whether the amount of risk of Mammals is significant to the next highest endangered categories.

Evaluating and Comparing Endangered Species

To confirm the hypothesis that mammals are the most endangered category, the chi-squared test for independence can be used to compare mammals against the next two endangered categories: Bird and Amphibian.

- Test #1: Determining significance between endangerment of mammals and birds:
  - Calculated p-value: 0.6875948096661336
  - Since the p-value > 0.05, there is no significance between mammals and birds (meaning they are equally susceptible to endangerment).

- Test #2: Determining significance between endangerment of birds and amphibians:
  - Calculated p-value: 0.0383555902297
  - Since the p-value > 0.05, there is no significance between birds and amphibians (meaning they are equally susceptible to endangerment).

Recommendation for Endangered Species

As there is no significant difference between the level of endangerment to the three highest protected species, I would recommend that special attention be put towards the Mammal, Bird and Amphibian categories to ensure that species within these categories do not risk being endangered.

Analysis of Sheep Populations

The group of scientists collecting this data also emphasized studying sheep populations.

Through filtering entries in the species dataframe and creating a new table, important information about the different species observed in the study.

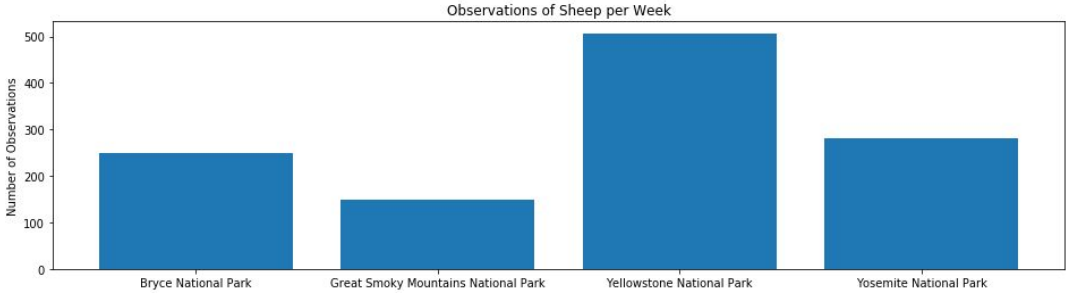
	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
5	Ovis canadensis sierrae	Yosemite National Park	39	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
6	Ovis canadensis sierrae	Bryce National Park	22	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
7	Ovis canadensis sierrae	Great Smoky Mountains National Park	25	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
8	Ovis aries	Yosemite National Park	126	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
9	Ovis aries	Great Smoky Mountains National Park	76	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
10	Ovis aries	Bryce National Park	119	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
11	Ovis aries	Yellowstone National Park	221	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True

Table 3: Sheep observations from Species dataframe

Analysis of Sheep Populations Contd.

	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

Table 4: Table Showing Total Sheep Populations at Parks



Graph 1: Observations of Sheep per Week at Each Park Recorded in Study

Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage point. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

The number of observations needed to reduce disease rates can be determined (with a default level of significance of 90%):

With the given baseline conversion rate of 15% and a calculated minimum detectable effect of 33.33%, the number of observations required is 870 observations.

In order to complete all 870 observations, for each park it would take:

- Bryce National Park: 3.5 weeks
- Yellowstone National Park: 1.8 weeks
- Yosemite National Park: 3.1 weeks
- Great Smoky Mountains National Park: 5.9 weeks