Biodiversity for the National Parks

Analyst: Yiying Zheng

Part I

Data Description

- The CSV file **species_info.csv** with data about different species in the National Parks, including (the printed first 5 rows see below):
  - The category of each species
  - The scientific name of each species
  - The common names of each species
  - The species 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

Data Description

- In the CSV file **species_info.csv**:

- There are 5541 different species in the species DataFrame.
- There are 7 types of species in the DataFrame: ‘Mammal’, ’Bird’, ‘Reptile’, ‘Amphibian’, ‘Fish’, ’Vascular Plant’, and ‘Nonvasular Plant’.
- There are 5 different values of ‘Conservation_status’: ‘Species of Concern’, ‘Endangered’, ‘Threatened’, ‘In Recovery’, and nan.

Species Conservation Status Analysis

- There are 15, 4, 151, and 10 ‘scientific_name’s fall into ‘conservation_status’of ‘Endangered’, ‘In Recovery’, ‘Species of Concern’, and ‘Threatened’, respectively.
- Compared with the total total number of species, there is only a small number of the species (180) are categorized as needing some sort of protection.

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	Species of Concern	151
3	Threatened	10

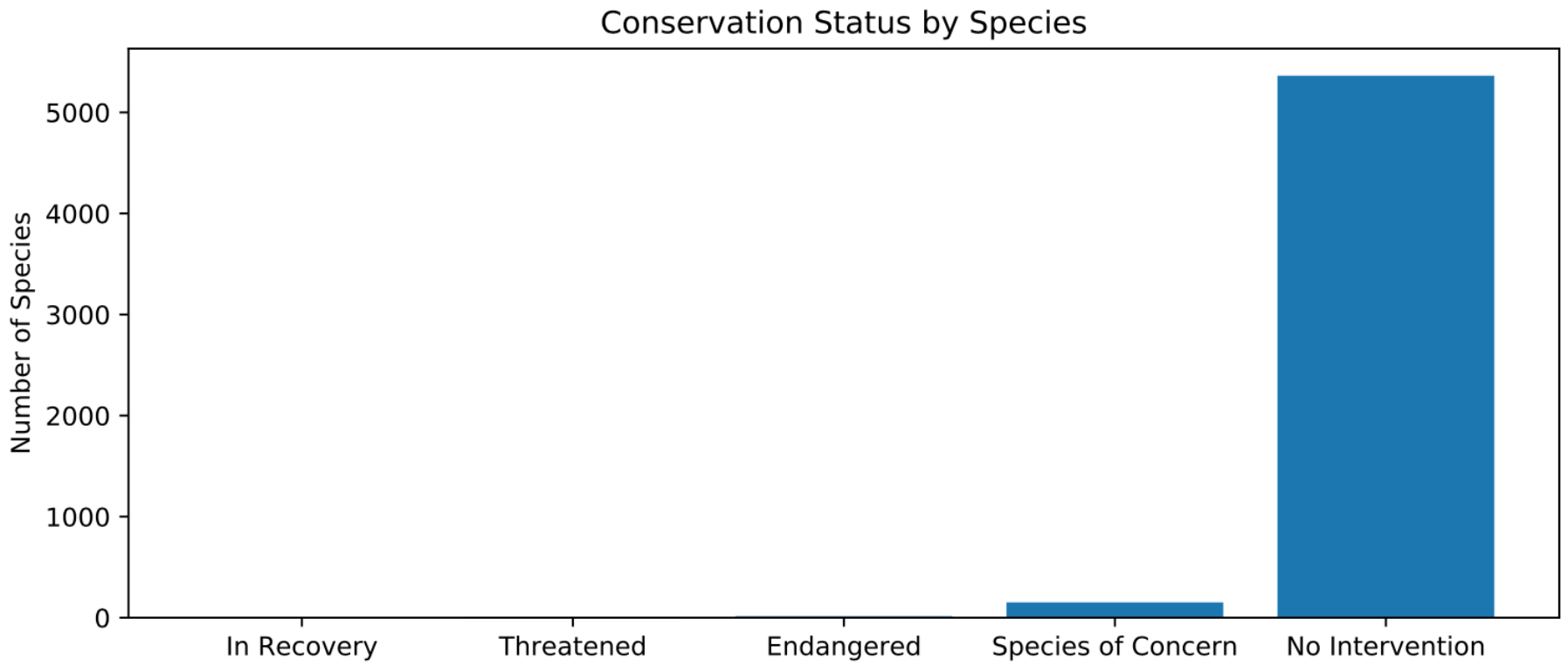
Species Conservation Status Analysis

- After replaced NaN with ‘No Intervention’, we got a fixed conservation counts as below.
- This shall give an accurate representation of the species conservation status.

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

Conservation Status by Species Plot

- Here we make this data visual by plotting the conservation status by species.



Endangered Species Investigation

- We pivoted the species conservation data and get a tale as follow:

	category	not_protected	protected
0	Amphibian	72	7
1	Bird	413	75
2	Fish	115	11
3	Mammal	146	30
4	Nonvascular Plant	328	5
5	Reptile	73	5
6	Vascular Plant	4216	46

Chi-Squared Test for Significance

- From the table, it looks like Mammals are more likely to be endangered than Birds. Then I did a significant text to see if this statement is true.
- The null hypothesis is this difference is due to chance.
- Then I chose to do a chi-squared test.
- The calculated p-value is ~0.688. The difference between the percentages of protected birds and mammals is not significant and is a result of chance.

- Then I also tested between Reptile and Mammal significant.
- The calculated p-value is only ~0.038. This difference is significant!
- **Some certain types of species are more likely to be endangered than others.**

Part II

Observations DataFrame

- The CSV file **observations.csv** with data about record sightings of different species at several national parks for the past 7 days, including (the printed first 5 rows see below):
  - The scientific name of each species
  - The park names
  - The observations of each species
  - The category of each species
  - The common names of each species
  - The species conservation status
  - The protection status
  - Whether the species is sheep

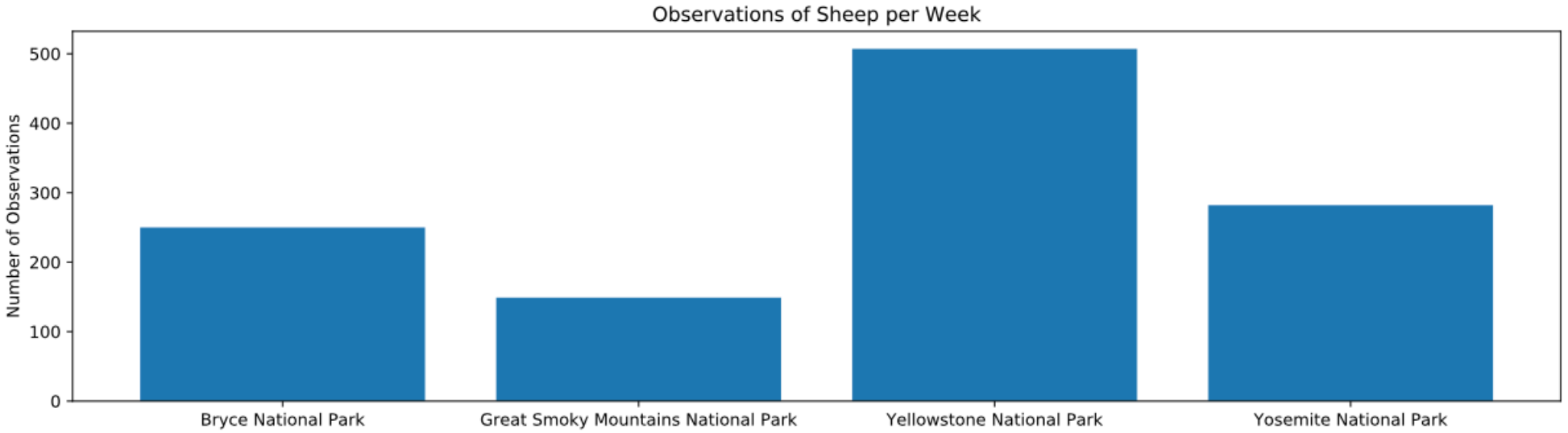
In Search of Sheep

- A team of ruminant-enthused scientists has been tracking the movements of various species of sheep across different national parks.
- I am asked to analyze the observation and species DataFrames to help track sheep locations.
- Below is the sheep observations by park names.

	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 Plot

- Below is the graph of the sheep observation data:



Foot and Mouth Reduction Effort - Sample Size Determination

- 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 points.
- Last year it was recorded that 15% of sheep at Bryce National Park have foot and mouth disease.

- The baseline percentage of this sample size determination is 15.
- The “Minimum Detectable Effect” is 33.3%.
- Statistical significance is 90%.
- Therefore, we need **510 samples** of sheep the scientists would need to observe from each park to make sure their foot and mouth percentages to be significant.
- Scientists need to spend approximately1 and 2 weeks at Yellowstone ans Bryce National Parks, respectively, to observe enough sheep.

Thank You!