Biodiversity Capstone Project

A analysis of species conservation status at National Parks

By Karthik Bhupatiraju

A Codecademy Project

Thoughts on ‘species_info.csv’… part 1

- The first six in the file:

	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
5	Mammal	Odocoileus virginianus	White-Tailed Deer	nan

- The conservation statuses:

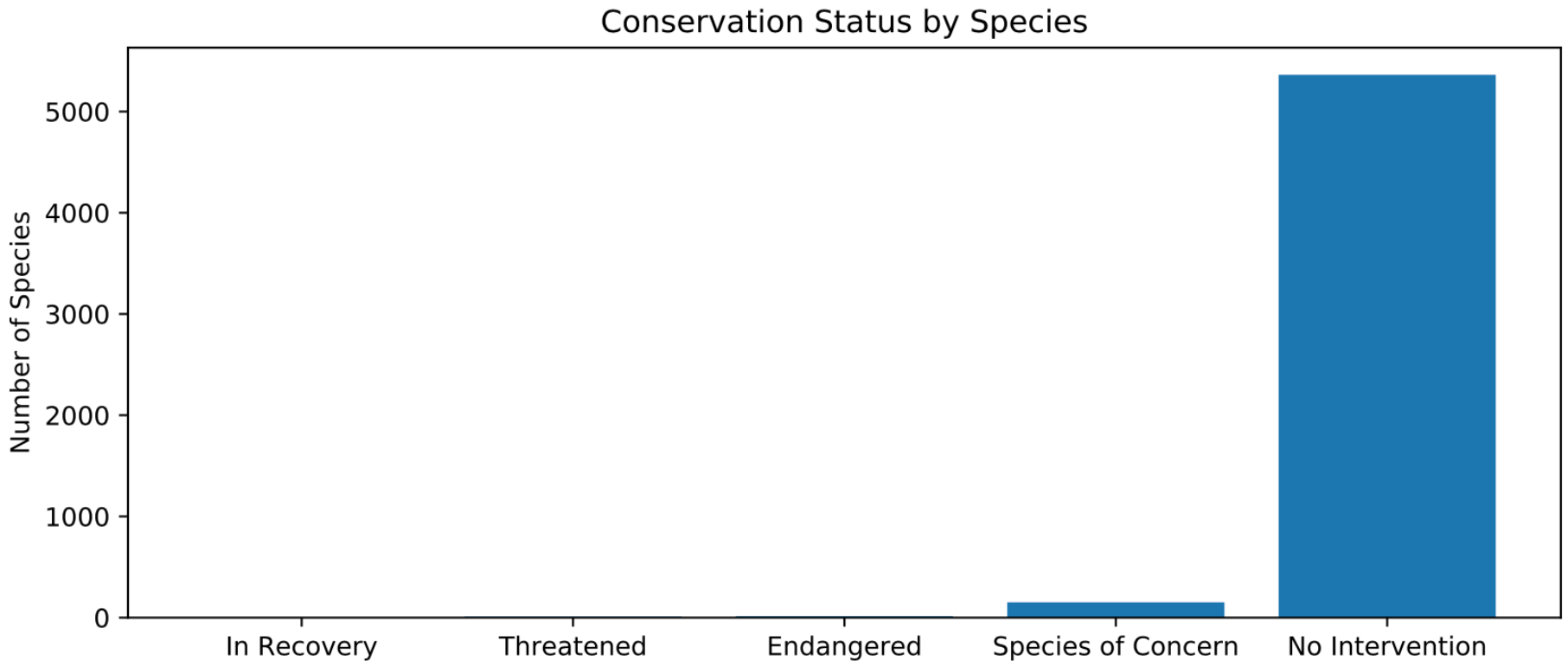
	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	Species of Concern	151
3	Threatened	10

Thoughts on ‘species_info.csv’… part 2

At first glance:

- Total of 5823 species listed in data
- Most of the data is “NaN” or “No Intervention” status with a total of 5363 occurances
- Some species have different names
- The data is sorted in an odd pattern:
  - Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant, Mammal, Bird, etc…
  - Maybe based on different national park data sorted in a similar manner

Thoughts on ‘species_info.csv’… part 3



Calculations and Comparisons… part 1

- In order to compare which type of species are the most likely to be endangered, we calculated the current percentages.
- Total table:

	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

Calculations and Comparisons… part 2

- The data suggests that mammals are most likely to be endangered, followed closely by birds.
- Vascular plant have the least likelihood.
- It is important for us to test whether the data is actually significant.
  - For example: We want to understand if the endangered mammals and endangered birds have any correlation or if those percentages occurred based solely by chance.
  - The null hypothesis suggested that the percentages are based on chance.
  - When comparing Mammals to Birds, the chi-test returned a p-value of 0.688 which means there is 68.8% chance that the percentages are based on chance, so the relationship is not significant and the null hypothesis is correct.
  - When comparing reptiles and mammals, the p-value was 0.038, which is less than 0.05, and means there's only a 3.8% chance the percentages are based on chance.
    - The result is significant and the null hypothesis is wrong.

Calculations and Comparisons… part 3

- Our findings suggest that mammals are in fact more likely to be endangered.
- I believe conservationists should considered spending more resources towards protecting mammals since they are most at harm.

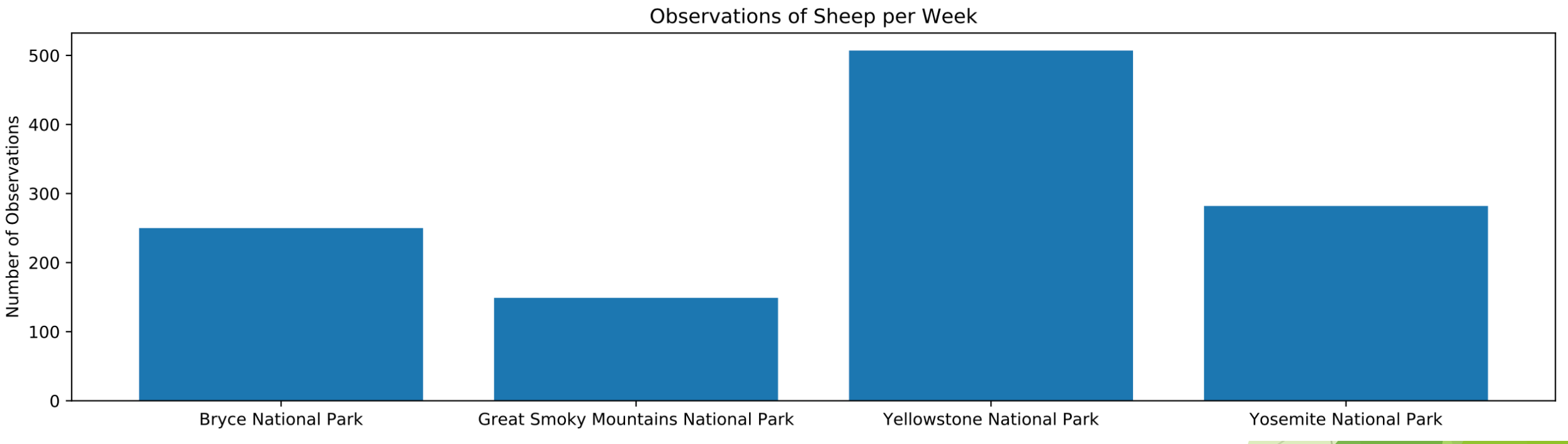
In Search of Sleep… part 1

- Scientists wanted to observe the movements of sheep in national parks.
- First we had to isolate the sheep data in the file “species_info.csv”.
- After finding that info, it was important to sort the data based on which national parks the observations were made.

	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

	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

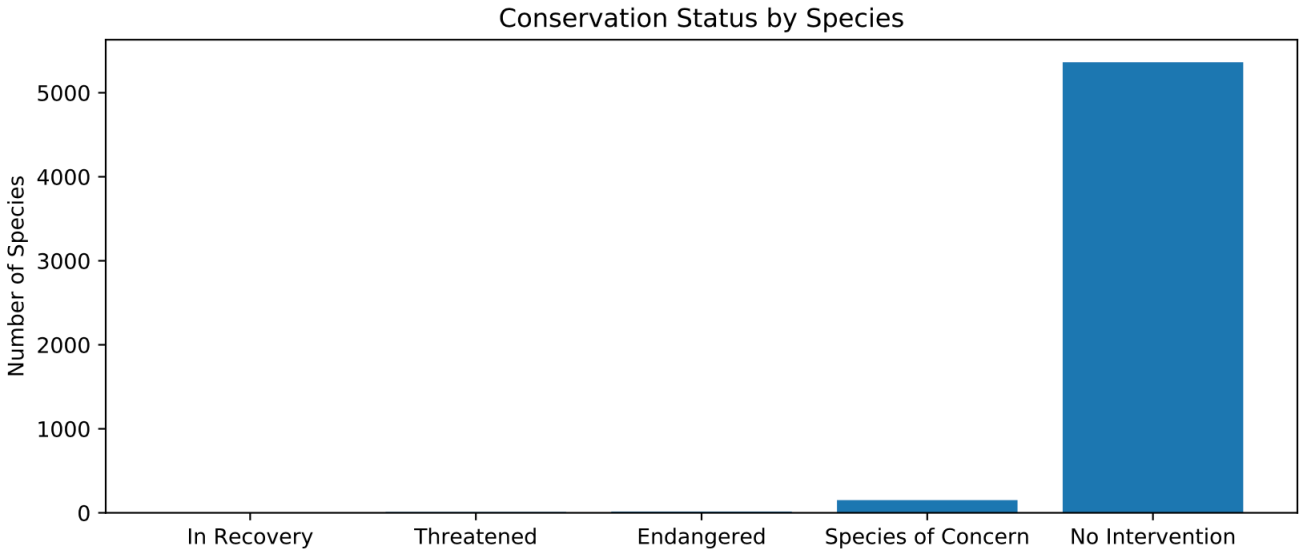
In Search of Sleep... part 2

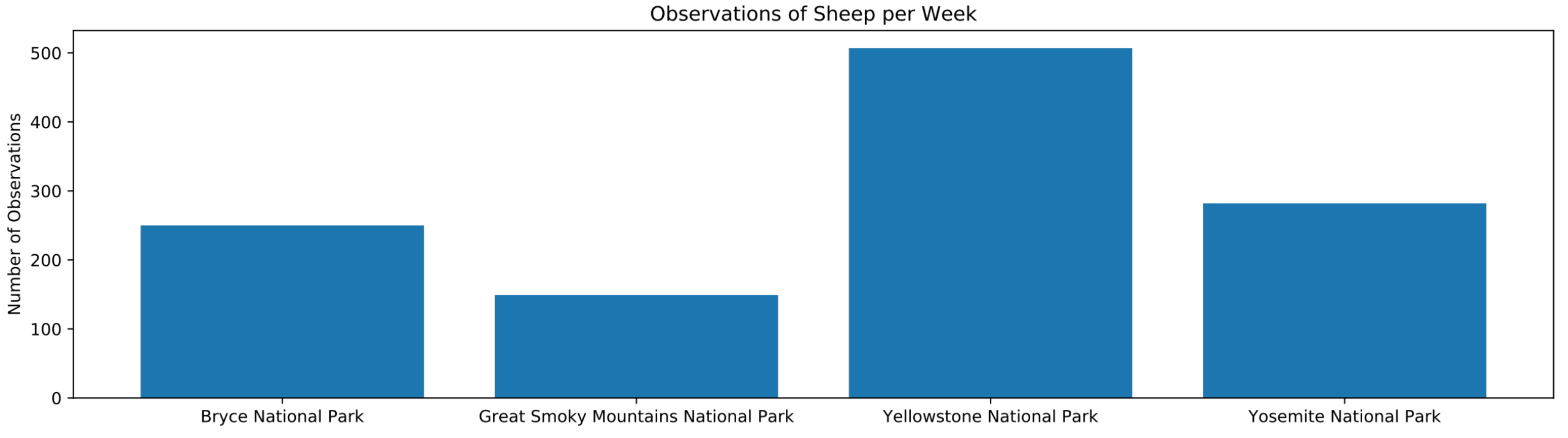


In Search of Sleep… part 3

- In order to find the required sample size of sheep to examine at Bryce National Park:
  - Had to use the baseline (15%) and the minimum detectable effect ((100*5)/15 = 33.33%) after setting the significance to 90%.
  - Found the sample size per variant to be 870 sheep.
- Then after dividing 870 by the total number of observations in each national park:
  - Bryce needs 3.48 weeks to see that many sheep.
  - Yellowstone needs 1.71 weeks to see that many sheep.

Graphs


Conservation Status by Species
Number of Species
5000
4000
3000
2000
1000
0
In Recovery
Threatened
Endangered
Species of Concern
No Intervention


Observations of Sheep per Week
Number of Observations
500
400
300
200
100
0
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park