Biodiversity

In the National Parks

Species Dataset

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	No Intervention
1	Mammal	Bos bison	American Bison, Bison	No Intervention
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	No Intervention
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention
4	Mammal	Cervus elaphus	Wapiti Or Elk	No Intervention

In this dataset we have all the species in the National Parks organized by:

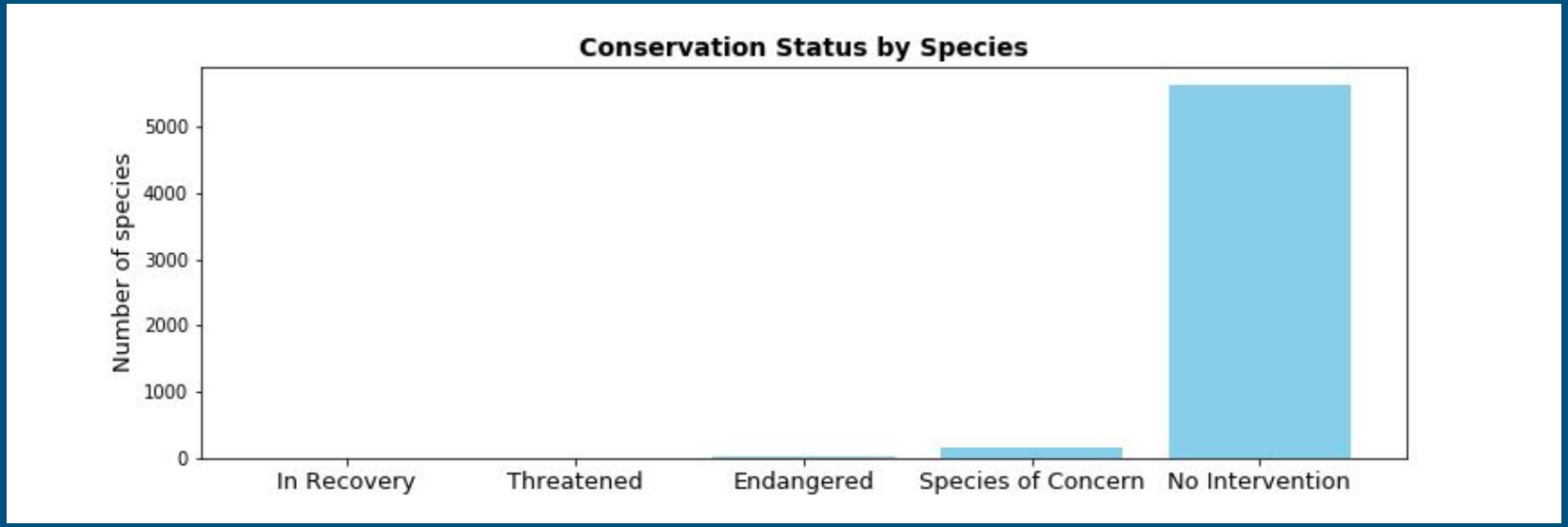
- Category
- Scientific name
- Common name
- Conservation status

➢ The total number of 5541 different species fall into such categories:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

➢ Based on the conservation status they're divided into:

- Species of Concern: declining or appear to be in need of conservation
- Threatened: vulnerable to endangerment in the near future
- Endangered: seriously at risk of extinction
- In Recovery: formerly Endangered, but currently neither in danger of extinction throughout all or a significant portion of its range
- No Intervention: don't need protection at the moment



	conservation_status	scientific_name
0	In Recovery	4
1	Threatened	10
2	Endangered	16
3	Species of Concern	161
4	No Intervention	5633

Number of species grouped by a conservation status and a bar chart that represents these numbers.

From previous numbers and a chart we can see that only a small percentage of species need protection.

But to see what types of species are more likely to be endangered we need to examine this table:

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	8.86
1	Bird	413	75	15.37
2	Fish	115	11	8.73
3	Mammal	146	30	17.05
4	Nonvascular Plant	328	5	1.50
5	Reptile	73	5	6.41
6	Vascular Plant	4216	46	1.08

Based on protected numbers, in the Bird category, 75 species are endangered.

Looking at the percentage protected we can see that it's actually Mammals who are more likely to be endangered, almost 2% higher than Birds.

Significance Test

To find out if the difference between Mammal and Bird categories is significant we need to perform Chi Square Test.

First we create a contingency table with protected and not protected number of species.

```
contingency = [[30, 146],
               [75, 413]]
```

--- Mammal
--- Bird

After we run our test we get a (rounded) pvalue of 0.69. Because if higher than 0.05 it means that our difference isn't significant.

We run one more test to compare Mammal and Reptile difference.

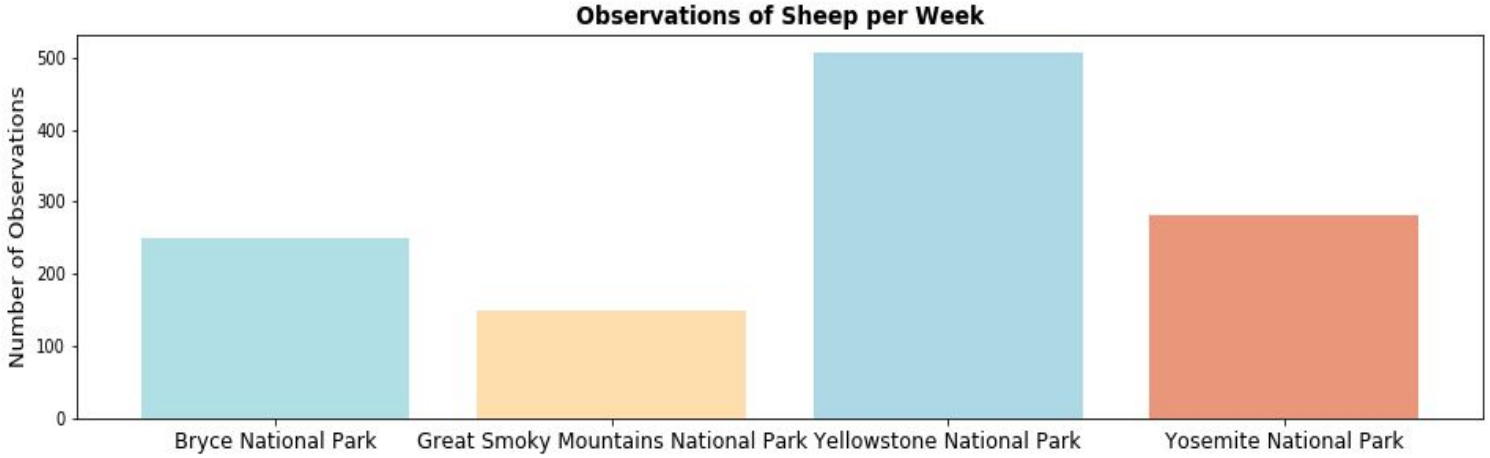
```
contingency_2 = [[30, 146],
                 [5, 73]]
```

--- Mammal
--- Reptile

This time we got a (rounded) pvalue of 0.04. Which means that the difference is significant.

We can conclude that both categories, Mammal and Bird, need higher protection.

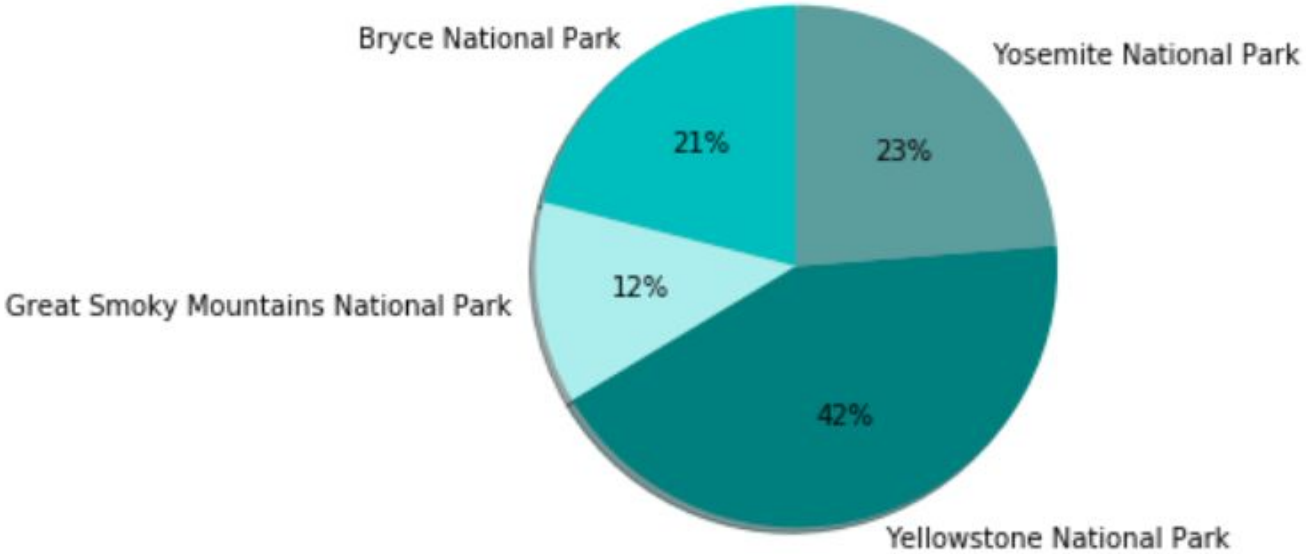
Sheep Observations



	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

Total number of sheep observed in each park over the past 7 days.

Sheep Observations



Sample Size Determination

To calculate the number of sheeps needed to be observed, in order to find out if the program works against foot and mouth disease, we need to have:

- Baseline Conversion Rate
- Minimum Detectable Effect
- Statistical Significance

➢ Our baseline is 15% (% of sheep at Bryce National Park that have disease) and we need to detect 5% of reduction.

➢ Our Minimum Detectable Effect: (5/15) * 100 = 33.33

➢ Our Statistical Significance: 90%

Based on these numbers we need 510 sheep to observe for each park.

For Bryce National Park we need to observe sheep for 2 weeks and for Yellowstone National Park - 1 week.