



National Parks' Biodiversity

Data analysis on endangered species – review by Marco Cavalazzi



Data considered

In the National Parks' Database, we find the following information about each animal.

- Category (Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant)
- Scientific Name
- Common Name
- Conservation Status (No Intervention, Species of Concern, Endangered, Threatened, In Recovery)

The current situation

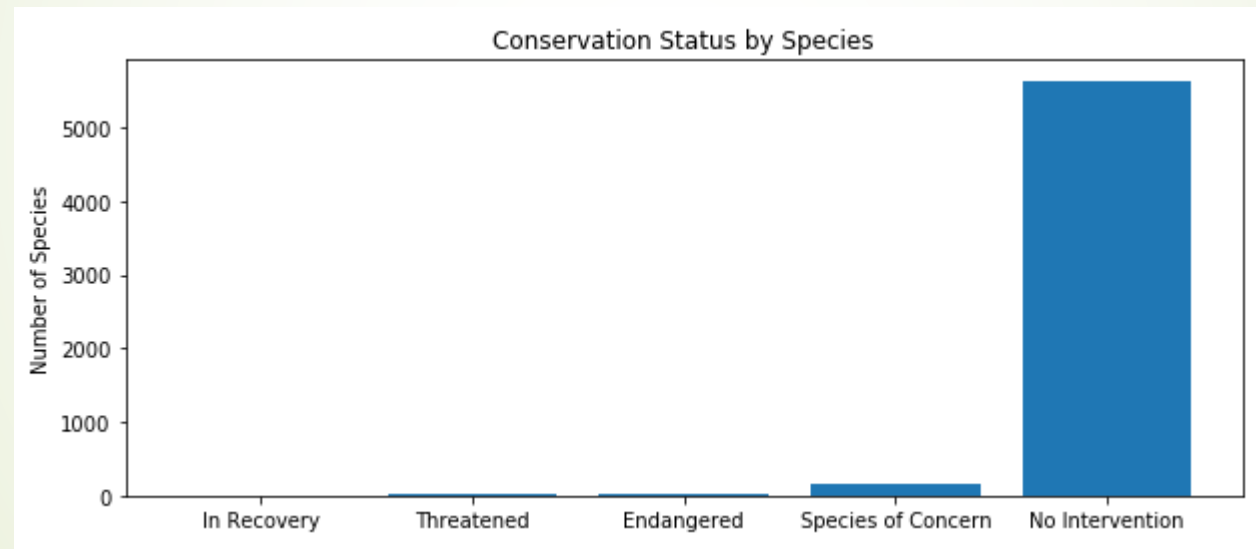
Examining the data, we find that:

- There are 5824 species in the parks
- Between those, the number of endangered species is:

Conservation status	Count
In Recovery	4
Threatened	10
Endangered	16
Species of Concern	161
No Intervention	5633

The current situation

The following bar chart gives a graphical explanation of the data collected:



The current situation

The data also show that some categories are more likely to include endangered species.

Category	Not protected	Protected	Percent protected
Amphibian	73	7	0.087500
Bird	442	79	0.151631
Fish	116	11	0.086614
Mammal	176	38	0.177570
Nonvascular Plant	328	5	0.015015
Reptile	74	5	0.063291
Vascular Plant	4424	46	0.010291



Our approach

In order to best evaluate the results of our analysis, we decided to study the differences between the groups. Using the **Chi Square** test we can understand if the numerical differences had statistical significance or were the result of mere coincidence.

A Chi Square (X^2) statistic is used to investigate whether distributions of categorical variables differ from one another.

Data Type	Question Type	Possible Responses
Categorical	What is your sex?	male or female
Numerical	Discrete - How many cars do you own?	two or three
Numerical	Continuous - How tall are you?	72 inches



Findings

Studying the difference between protected *mammals* and *birds* we find that the Chi Square test returns a p-value of 0.445901703047, which is way higher than 0.05, so there is no significant difference between the two groups. They both present almost the same percentage of endangered species.

On the other hand, confronting *mammals* and *reptiles* we find a p-value of 0.0233846521487. This tells us that there is a significant difference between the groups of animals. In particular, there is a significantly higher percentage of endangered mammals than reptiles.

Findings

Conservationists have been recording sightings of different species at several national parks for the past 7 days. An analysis of the data regarding sheep tells gives us the following table:

Category	Scientific name	Common names	Conservation status	Is protected	Is sheep
Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True



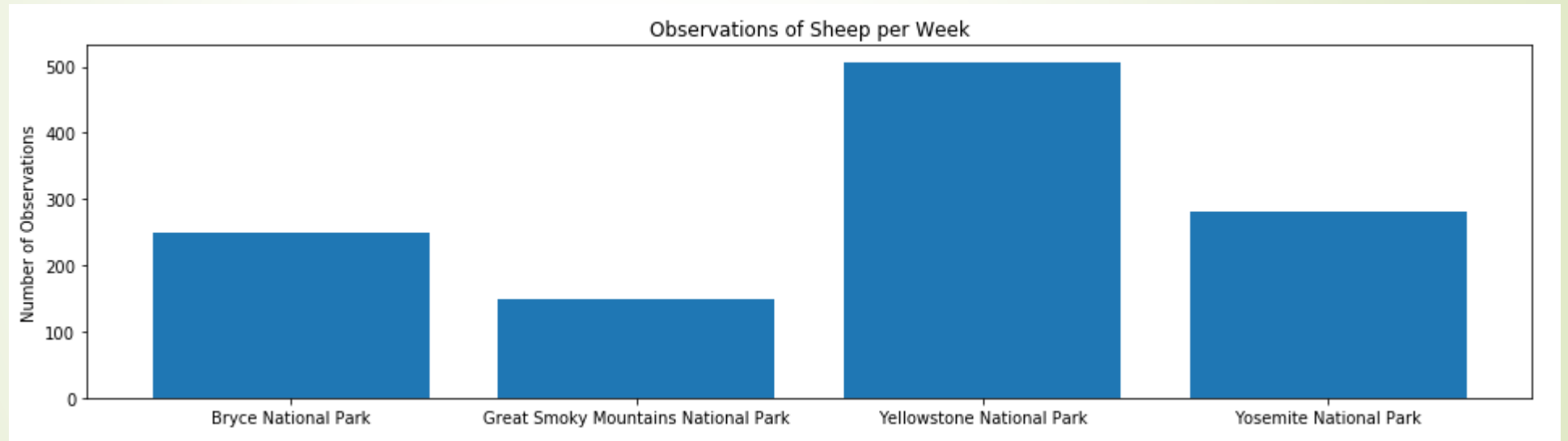
Findings

The number of observations for all three sheep's species for each National park is:

Park name	Observations
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

Findings

The equivalent bar chart is:





Problem

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.

We want to test whether or not this program is working and we want to be able to detect reductions of at least 5 percentage points.



Conclusions

According to the data, the categories at risk, and thus to keep under strict observation, are mammals and birds.

In particular, regarding sheeps we can devise through Optimize the number of sheeps to observe in order to be confident in our results for the foot and mouth disease, which is 510.

According to the data previously shown, in order to have enough data for the Bryce National Park we will need to collect data for 2 weeks (2.04 weeks).

For the Yellowstone National Park, instead, 1 week will suffice (1.005917 weeks).