



# Biodiversity for the National Parks

By Ali-John Mirsepassi

As part of the CodeAcademy Intensive course for  
Data Analysis

# Project Aims

The National Parks Service has requested that we undertake Data Analysis for conservation statuses on a number of species and to investigate if there are any patterns or themes to the types of species that become endangered. We will show the number of species that should be of concern to conservationists and whether a particular type of species (Mammal, Plants, etc) needs to have more due care and attention focused on it than others.

We shall also look at sheep species in the national parks and the rates of foot and mouth disease. Scientists have been working on whether their recent experiments have resulted in a significant reduction or not, and so we will say what sample sizes are needed and how long some parks will need to hold their experiments for.

# The Species in these National Parks.

There are (counting the Scientific Names and not the Common Names) 5541 distinct species in these National Parks. Of these they have been sorted into seven categories.

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

# The Species in these National Parks.

Species we should be concerned about are in one of four conservation statuses.

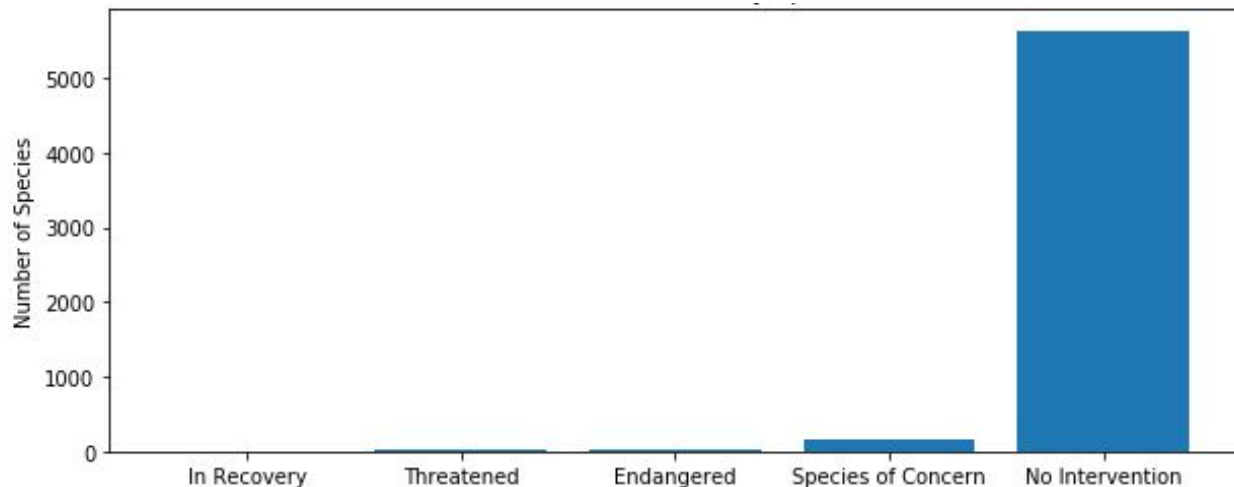
- Species of Concern: Declining or appear to be in need of conservation
- Threatened: Vulnerable to endangerment in the near future
- Endangered: Seriously at risk of extinction
- In Recovery: formerly Endangered, but currently neither in danger of extinction throughout all or a significant portion of its range

There is also a fifth value in the table: None. This means that there is no issues with their conservation species and thus no action to protect them need be taken.

We mark this in our data as “No Intervention”

# Conservation Status By Species

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



Summing all the categories that are not in the category “No Intervention”, we find that there are 191 species in total that we should be currently concerned about.

# Conservation Status By Species

The following table breaks down whether a species is protected (requires no intervention) or not by the type of category that the species falls under, alongside the percentage of species in that category that are currently protected.

Category	Is Not Protected	Is Protected	Percent Protected (2 decimal places)
Amphibian	7	73	91.25
Bird	79	442	84.83
Fish	11	116	91.34
Mammal	38	176	82.24
Nonvascular Plant	5	328	98.50
Reptile	5	74	93.67
Vascular Plant	46	4424	98.97

# The Null Hypothesis

Before we continue, it would be best if we just make a note about the definition of the null hypothesis.

The Null Hypothesis (in the context of Data Analysis) is the statement that any observed differences between two or more sets of data is by pure chance. To reject the Null Hypothesis is to say that there is a reason why any observed differences occur.

To see whether we can assume the Null Hypothesis is valid, we make a Hypothesis Test on the data which produces a p-value. This p-value is the probability that the Null Hypothesis is valid (so will be a value between 0 and 1)

# Comparison between species

Take two categories on the table: Mammals and Birds

Bird	79	442	84.83
Mammal	38	176	82.24

There appears to be a greater rate of Mammal species in our data risking extinction than there are Birds, and so it is in our interest to see if this is indeed the case or if the Null Hypothesis is True. Hence we should perform a significance test that will produce a p-value and if this p-value is below 0.05 or 5% (an arbitrary number that we have decided in advance) then it is likely that the Null Hypothesis is not true, that the data differences is significant, and that there could be much more risk to mammals than there are to birds.

Because we are comparing categorical values from a table (specifically the middle two columns) we should use a Chi Square Test to get our p-value



# Comparison between species

After calculating the Chi Square Contingency test, we find that the P-Value is 0.445901703047 or 44.59%

Hence, because the p-value is almost 40 points above the threshold we have set for ourselves, we can assume the differences of the rates of protection and non-protection is not significant. Hence, we can be confident in saying to conservationists that extra work on conservation for mammals over birds is not needed.

# Comparison between species

We will now do a similar test between Reptiles and Mammals. As a reminder for what the rates are for “Is Not Protected” and “Is Protected” with percentages for the latter.

Mammal	38	176	82.24
Reptile	5	74	93.67

At a glance, Mammals are much more vulnerable The p-value calculated is 0.0233846521487 or 2.34%. Because it is below the threshold by half, it appears that the difference in data is significant. Note that **this is not absolute proof** that the Null Hypothesis is wrong and that conservationists should give more due care to Mammals as opposed to Reptiles, but that more research should be done about these rates and what the causes of the difference in rates of protection and non-protection are. From a more detailed analysis, conservationists should be able to determine exactly how much more due care is owed to Mammals if any increase is needed at all.

A photograph of a white sheep standing in a lush green field. The sheep is facing the camera, and its woolly texture is clearly visible. In the background, there is a wire fence and more greenery. A semi-transparent grey box is overlaid on the image, containing the title and a paragraph of text.

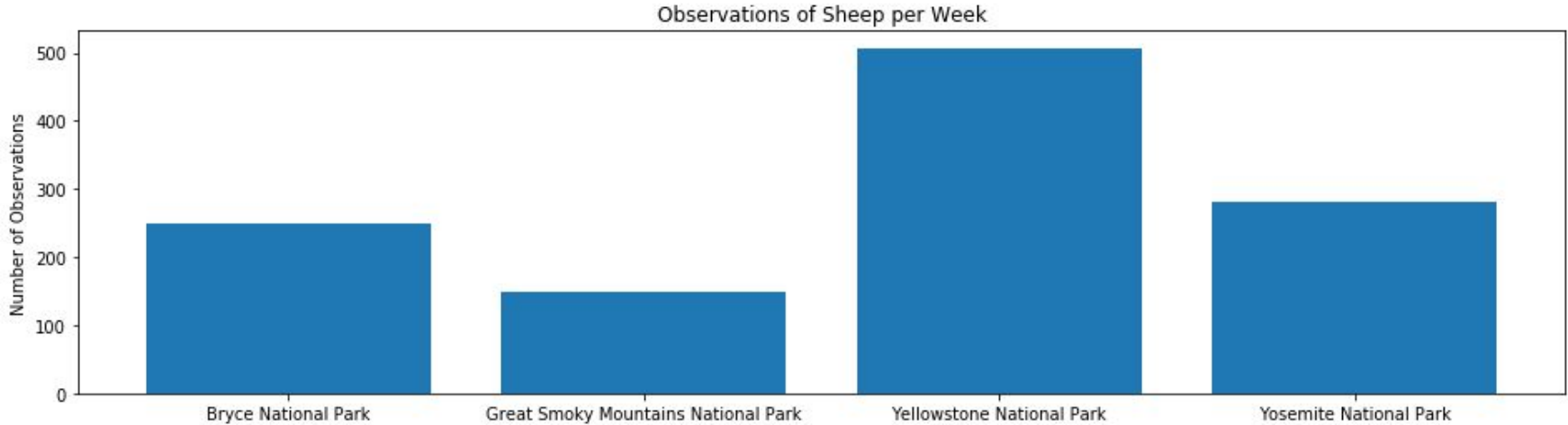
# The Status of Sheep in the National Parks

This section will be about Sheep in the National Parks, how many species of sheep there are per National Parks and the rates of Foot and Mouth in some of these parks.

# Sheep in the National Parks

We have been sent data from observations on the sightings of different species in four American National Parks (Bryce, Great Smoky Mountains, Yellowstone, Yosemite). This data was gathered over a period of one week. We have then merged the data with the number of sheep species that we know are in the parks to produce the following chart and table about the numbers of sheep observed per National Park.

# Sheep in the National Parks



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

# Sheep in the National Parks: Foot and Mouth Disease

We have been informed by scientists that at 15% of sheep in Bryce National Park. In Yellowstone, there is an attempt by Park Rangers to use a program to reduce the rates of F&M. The scientists are interested in knowing if it is working, and seek to detect reductions of 5 percentage points.

In the case of Bryce National Park, with its 15% rate, this would mean a minimal detectable effect (The minimum relative change in conversion rate you would like to be able to detect. [Source: [Optimizely](#)]) of 33.3%\* which we use when calculating the necessary sample size of our experiments.

\*33.3 recurring.

# Sheep in the National Parks: Foot and Mouth Disease

Using the [A/B Test Sample Size Calculator on the website Optimizely](#), we have calculated that we need the sample size for this experiment to be 510. From the data gathered on the number of sheep observed, we need to observe sheep for two weeks at Bryce and one week at Yellowstone for us to get the required sample size.

Baseline Conversion Rate

15 %

Your control group's expected conversion rate. [\[?\]](#)

Minimum Detectable Effect

33.33 %

The minimum relative change in conversion rate you would like to be able to detect. [\[?\]](#)

Statistical Significance

90%

[EDIT](#)

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [\[?\]](#)

Sample Size per Variation

510

# In Conclusion

Analysing the data from the rates are conservation we can see that the fact that a larger percentage of mammal species having conservation issues than birds does not seem to be significant, but mammals having a larger percentage than reptiles is. Hence more due care may not need to be given to mammals over birds, but further analysis should be done to see why birds species are less likely to be at risk.

On the issue of Foot and Mouth Diseases in Sheep, we have found that in order detect where there is a five point reduction in the rates of F&M (from 15% to 10%), we need to have a sample size of 510 sheep, which can be gathered over a period of one week at Yellowstone National Park, or two weeks at Bryce National Park



A photograph of a forest path. The path is covered in fallen leaves and leads into a dense forest of tall, thin trees. Sunlight filters through the canopy, creating a warm, golden glow. The grass is green and lush. The text "Thank you for reading" is overlaid in the center in a large, white, sans-serif font.

# Thank you for reading