CODECADEMY BIODIVERSITY

Capstone Project Option 2

By Warren Slatem

SUPPLIED DATA: NATIONAL PARK SPECIES INFORMATION [SPECIES_INFO.CSV]

- The data base includes a detailed list of flora and fauna species found in various national parks

Fauna Categories:

- Mammals
- Birds
- Fish
- Reptiles
- Amphibians

Flora Categories:

- Vascular plants
- Nonvascular plants

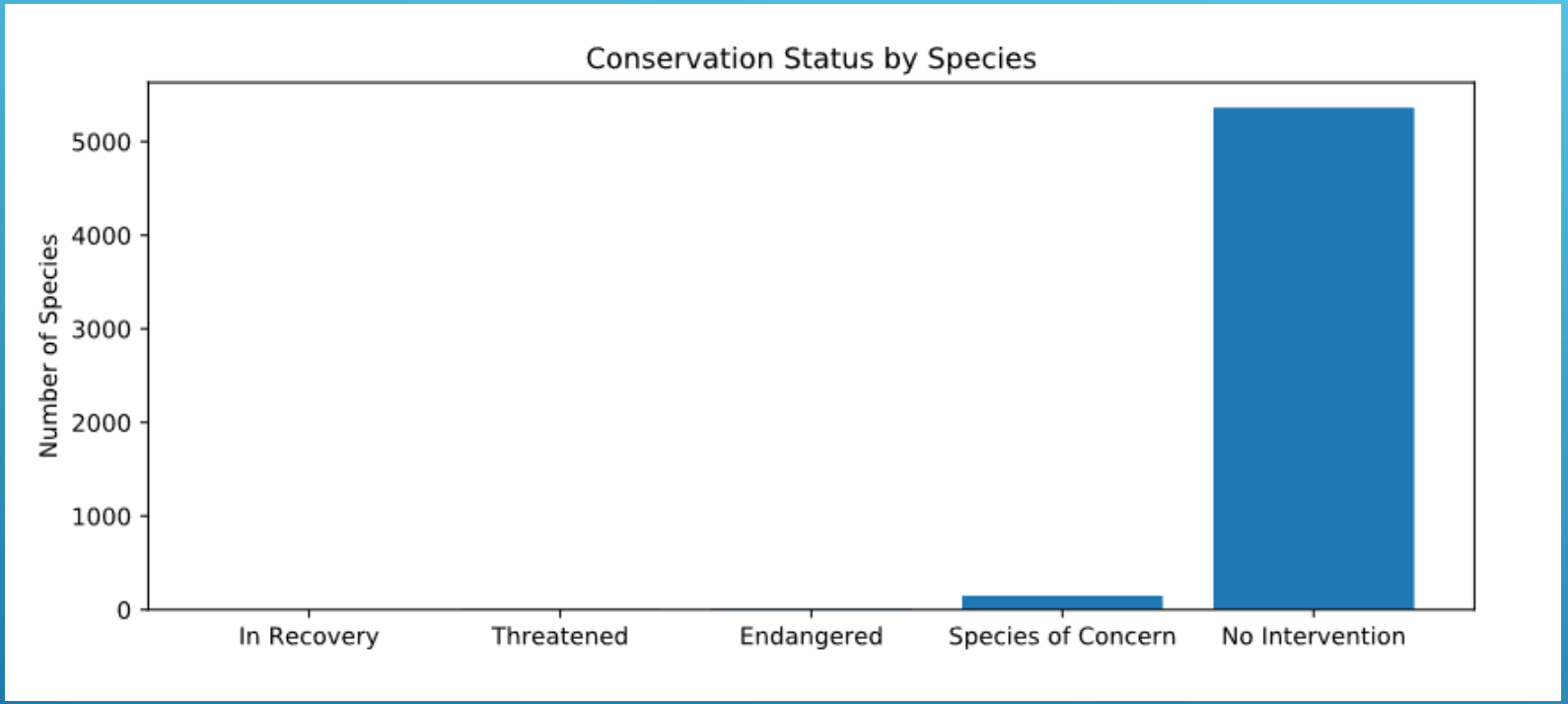
A deeper look at the Species_Info.csv database we find that each of the entries lists the following:

- Category
- Scientific Name
- Common Name
- Conservation Status

Further More, our Conservation Status has the following states:

- In Recovery
- Threatened
- Endangered
- Species of Concern
- No Intervention

VISUALISATION OF CONSERVATION STATUS OF ALL SPECIES



Using the data provided we generated a bar graph to visualise the significance of all the data. As one can see the graph on 'No Intervention' is the leading result to the extent that it dwarfs the other categories. This does not mean the other data is insignificant.

We turn to analytics to make sense of the smaller ratios and where the issue lie.

TESTING FOR RESULT SIGNIFICANCE

category	not_protected	protected	percent_protected
Amphibian	73	7	0.0875
Bird	442	79	0.151631
Fish	116	11	0.086614
Mammal	176	38	0.17757
Nonvascular Plant	328	5	0.015015
Reptile	74	5	0.063291
Vascular Plant	4424	46	0.01029

First we make sense of the large database of data and sum all the protected and not protected species of each category.
We also calculate the percentage of these value to give an idea of the significance to try start making sense of the information

Using the previous generated table, we can run the pairs of data of each category through a Chi2 Contingency Test to generate an answer to a proposed null hypothesis

We first ask if the data on Mammals and Birds are significant:

This returns a result of: 0.687.
This indicates there is no significance, proving this hypothesis false

We next then try compare Mammals and Reptiles.
This returns a result of: 0.038

This indicates a significant result and proving our hypothesis true.

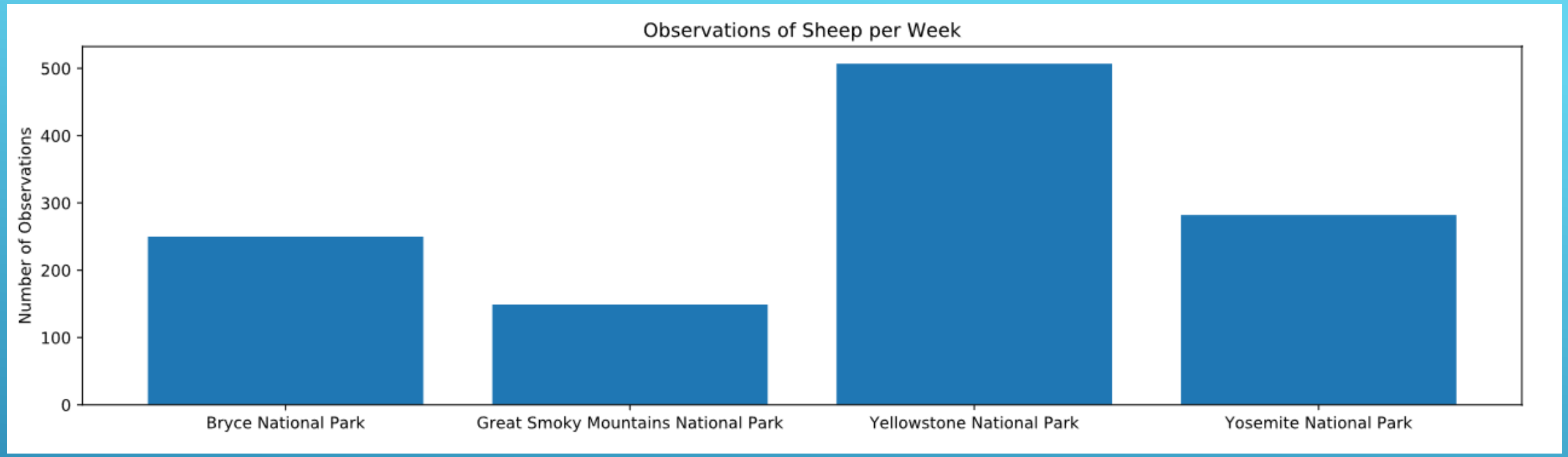
This concludes that some species are more like to be endangered than others, specifically reptiles are more likely to be endangered than birds (It must be due the flight advantage, but we need more data to prove that!)

STUDY: FOOT AND MOUTH DISEASE

THE NATIONAL PARKS SUBMITTED US WITH OBSERVATION ON SHEEP IN THE VARIOUS NATIONAL PARKS

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

In the table above we are shown the quantity of sheep observed in each national park in a week.



We generated a bar graph to make it simpler to visualise the information that we have been given from each park and it quickly give us an idea of the ratio's of sheep observations in each park

THE PARK RANGERS WOULD LIKE TO KNOW THE REQUIRED SAMPLE SIZE OF THEIR PROGRAM TO REDUCE FOOT AND MOUTH DISEASE AT YELLOWSTONE NATIONAL PARK

The rangers are looking to reduce the disease by 5%, given that 10% of the sheep at Yellowstone likely have the disease.
A study last year at Bryce National park recorded that 15% of the sheep observed where infected.
We need to find the required population size to be observed, with 90% confidence, to be able to make an accurate measurement of infected sheep at Yellowstone.

Using the provided information we determined the sample size required to make an accurate assumption of disease carries

Baseline – 15% (The study done at Bryce National Park)

Statistical significance – 90% (Our target significance)

Minimum Detectable Effect – 33.33% (New Target [5%[/ acquired data [15%])

The result is a required **Sample Size of 870**

Given that 507 sheep are sighted in a week in Yellowstone, it would take just over 12 days to acquire the correct sample size.
While at Bryce it would take just over 24 days to acquire the same sample size