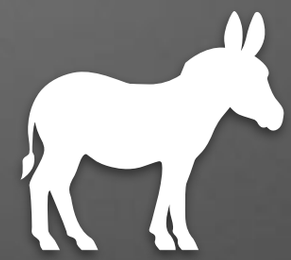
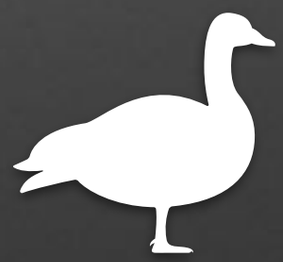
Biodiversity Research Analysis

Codecademy Capstone Project
Zicheng Wei

Holistic Species Data Summary

- In the data set, the conservation situation of different kind of species had been listed. Accordingly, different species' category, scientific name, common name and conservation status are included in the data set.

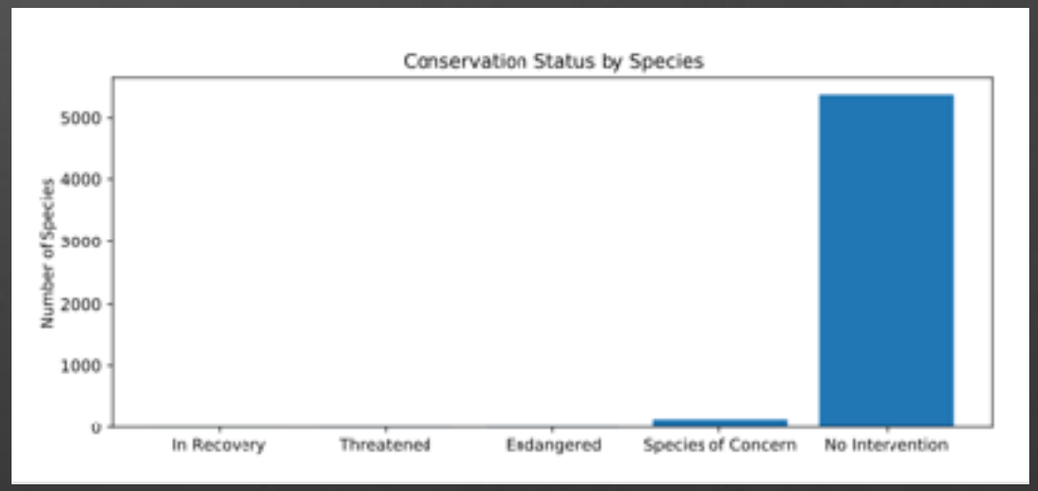




Data Classification

- Over all, 5541 kinds of species had been involved. Including mammals, birds, reptiles, amphibians, fishes and vascular plants etc. Through these species, 5363 kinds of species are not intervened, for the further information refer to the table and chart on the right.

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10



Conservation Status Count

Are Certain Types of Species More Likely to be Endanger

- According to the table below, for each category species, it's easily to notice that mammals are protected best although the protected mammal only occupied 17%. More species needs more attentions on protection, such as amphibians and fishes only get around 8% protected percent. Extremely, vascular plants only get 1.1% protected percentage. Which means, there are reasons to believe that certain species are more likely to be endanger. In order to be much more accurate, a chi-square test is necessary to be processed.

Category	Not Protected	Protected	Percent Protected
Amphibian	72	7	0.088608
Bird	413	75	0.153689
Fish	115	11	0.087302
Mammal	146	30	0.170455
Nonvascular Plant	328	5	0.015015
Reptile	73	5	0.064103
Vascular Plant	4216	46	0.010793

Chi-square Test

- Sample size is a crucial factor of research's accuracy, it's not pretty much accurate if only consider the protected percentage due to the sample size of different species are not same. In order to avoid this error, a chi-square test should be processed.

Are Certain Types of Species More Likely to be Endanger

- Through the chi-square test, the null hypothesis always is that the difference between protected percent of different species are duo to chance. For the test between birds and mammals, the p-value is 0.688 which is higher than 0.05. It indicated that, the similarity between protected percent of birds and mammals is due to chance. But running the on mammals and reptiles, the p-value is 0.038, which indicated the difference between the mammals and reptile is significant.

- To conclude, certain types of species more likely to be endanger than others.

Observations Through Different Parks

- In the file observations.csv, different species observed in different parks had been recorded. For different species or specific animal or plant, its distribution can be explored.

- For example, it is easy to conclude the distribution mammal which common name include sheep along the observation of scientist.

Distribution of Specific Species

- In this case, the distribution of mammals which's common name contain sheep had been concluded. Through the graph, it is easily to get the result, Yellowstone National Park has the most observation amount.

Recommendation

- Due to certain types of species are more likely to be endanger, scientists would like to allocating resource depending on the order of severity.
- Refer to the result of distribution observation, scientists can find out specific species’ observation amount in different parks. Depending on this information, scientists may explore the distribution of the species they focus on. To do so, scientists can allocate their resource depending on the distribution of the species.

Foot and Mouth Disease Study on Sample Size

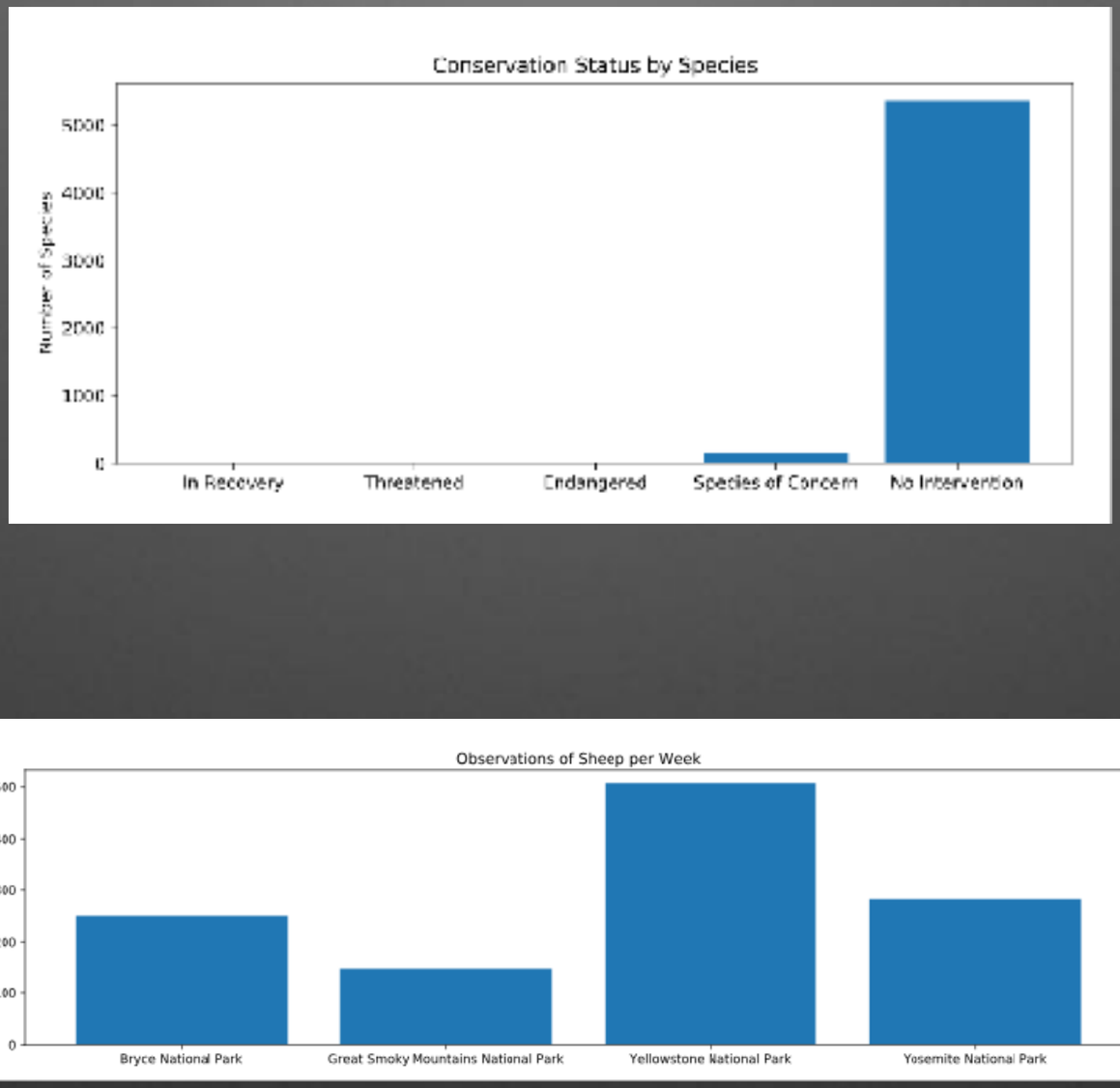
- In order to observe accurate data to speculate results but do not occur unnecessary wasting of recourses, a feasible sample size of research is crucial.

- Sample size determination depending on base line conversion rate, statistical significance and minimum detectable effect.

- In foot and mouth disease case, 15% sheep at Bryce National Park have foot and mouth disease as result the baseline conversion rate would be set as 15%. As long as scientist wants to detect 5% reduction, the detectable effect would be calculated as detection demand (5%) / baseline conversion rate (15%) which is equal to 33.3%. Due to the demand statistical significance is 90%, sample size would be set as 890.

Foot and Mouth Disease Study on Sample Size

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

Thanks

–Zicheng Wei