

# BIODIVERSITY

Capstone Project 2  
Marisha Eygelaar  
10/04/2018





# OBJECTIVE

- An analysis of the conservation statuses of endangered species from several different parks for the National Park Service.
- To investigate patterns and or themes pertaining to the types of species that become endangered.
- An analysis of the observations of sheep at several different parks in order to determine the correct sample size to test whether a foot and mouth disease program is successful in one of the parks.



# OBSERVATIONS FROM THE DATA

TWO FILES WERE PROVIDED WITH DATA GATHERED FOR THE STUDY:

1. DATA ON CONSERVATION STATUS OF EACH SPECIES AT VARIOUS NATIONAL PARKS
2. DATA ON THE NUMBER OF OBSERVATIONS OF EACH SPECIES AT VARIOUS NATIONAL PARKS



# OBSERVATIONS FROM THE DATA:

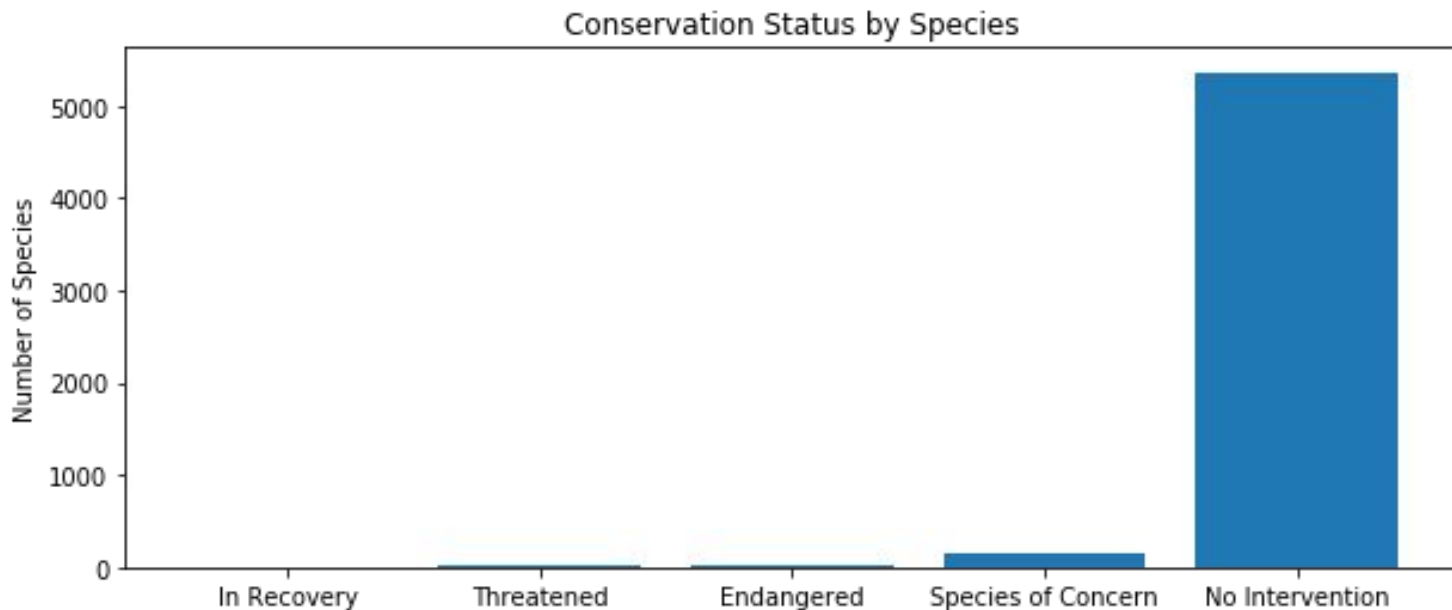
## 1.Data on species

- The data included a total of 5541 different species
- The species could be classified according to the following different categories
  - Mammal ; Bird ; Reptile ; Amphibian ; Fish ; Vascular Plant ; Nonvascular Plant
- The conservation status of species needing some sort of protection could be categorized as follows:
  - Endangered ; In Recovery ; Species of Concern ; Threatened
- Species that did not need any sort of protection were categorized under
  - No Intervention
- The bar chart that follows shows the number of species in each category of conservation-status



# OBSERVATIONS FROM THE DATA:

## 1.Data on species





# RESULTS FROM STATISTICAL ANALYSIS

- The conservation status of each CATEGORY of species was analyzed in order to establish if certain types of species are more likely to be endangered
- The conservation status was grouped into 2 categories:
  1. Protected - All the species categorized under Endangered ; In Recovery ; Species of Concern and Threatened and therefore needing some sort of protection.
  2. Not Protected - Species categorized under 'No Intervention'
- The grouped data was then analyzed in terms of category of species protected vs not protected and a percentage protected per category was calculated
- The table below displays the statistical data:



# RESULTS FROM STATISTICAL ANALYSIS

	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



# RESULTS FROM STATISTICAL ANALYSIS

- One could say that species in category Mammal are more likely to be endangered than species in category Bird which is why more percent of mammals are protected.
- In order to be sure that this statement is definitely true and not because of certain differences in the mean of the data that is being compared we did a significance test.
- The significance test between Mammal and Bird yielded a p-value of 0.6875948096661336
- This means that there is no significant difference in the mean of two datasets which is being compared and the null hypothesis is therefore rejected. The comparative results can be accepted as True.
- The significance test between Mammal and Reptile yielded a p-value of 0.03835559022969898
- This means that there is significant difference in the data which is being compared and the null hypothesis is therefore accepted. The result might be a false positive and further investigation would be needed if we want to be very sure that the results are not just due to chance.





# **OBSERVATIONS FROM THE DATA:**

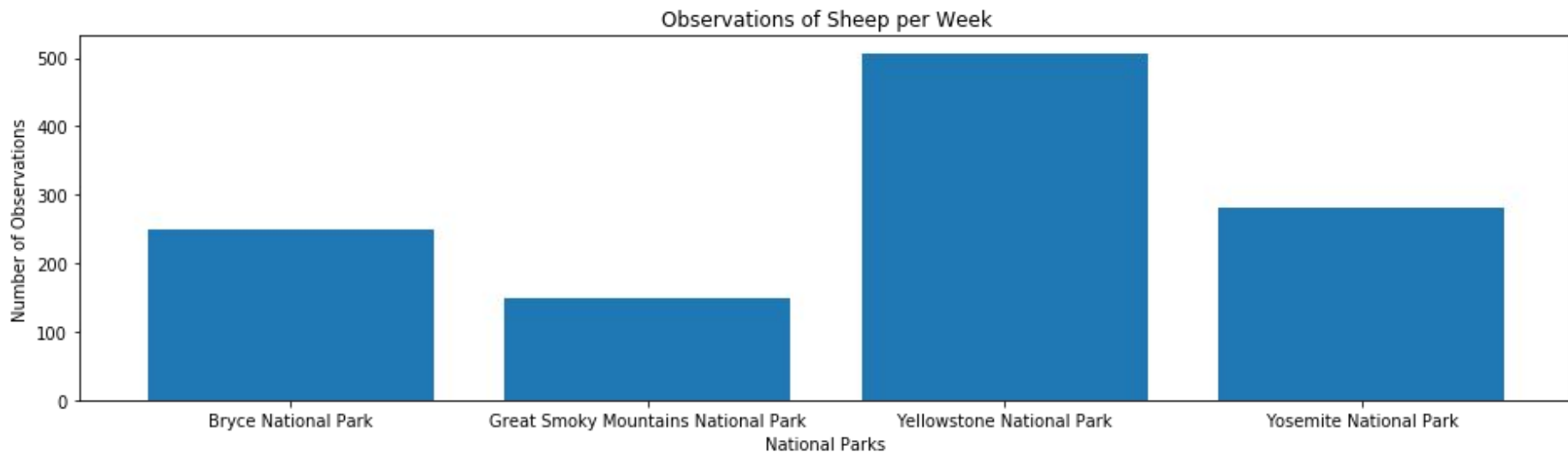
## **2.Data on number of observations**

- The observations data was merged with the species data in order to get a dataframe with the number of observations of a certain species.
- This data could be used to further establish whether the sample size of each species was correct in order to reject a null hypothesis in any given situation.
- The observations of Sheep per week at each National Park was investigated in order to establish whether a program to reduce foot and mouth disease in sheep has been successful in Yellowstone National Park.
- The chart below shows the number of sheep observed per week at each national park



# OBSERVATIONS FROM THE DATA:

## 2.Data on number of observations





# RESULTS FROM STATISTICAL ANALYSIS

- A sample size calculator was used to determine the number of sheep they would need to observe at each park in order to be confident about the results of their tests.

## HOW THE SAMPLE SIZE WAS DETERMINED:

- Our scientists knew that 15% of sheep at Bryce National Park have foot and mouth disease - this was therefore the BASELINE
- In order to calculate the minimum detectable effect we used the fact that our scientists wanted to be able to detect reductions of at least 5 percentage point. Minimum detectable effect is a percent OF the baseline. So we calculated 5% of 15% which resulted in 33.33%
- A default level of significance of 90% was used



# RESULTS FROM STATISTICAL ANALYSIS

- 510 sheep would need to be observed at each park
- In Bryce National Park we would need approximately 2 weeks in order to observe enough sheep
- In Yellowstone Park we would need approximately 1 week.



# RECOMMENDATIONS

- One could use both of the data files provided and merge the data files from the start.
- It would then be easier to compare results between species and parks as one could calculate a sample size for each comparative test , take enough time to study the required sample sizes in order to attain results where the null hypothesis can be rejected.
- The results would therefor be more trustworthy if one takes the above approach.



**THE END**

**Thank you very much.**