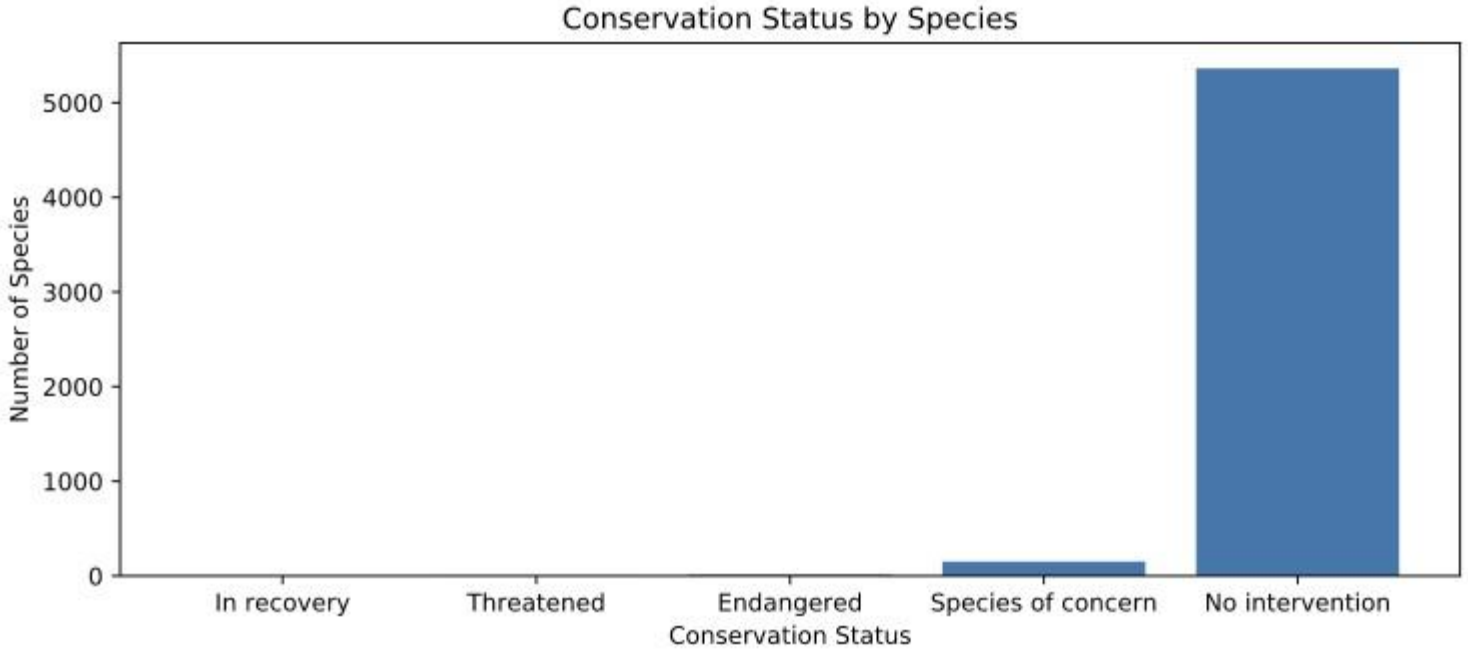
Capstone Option 2: Biodiversity for the National Parks

By Nana Adu-Poku

Regarding the CSV file species_info.csv ...

- This csv contained information about different species, including: their scientific names, their category (Mammal, Bird, etc) their common names, and their conservation status (protected, endangered, etc)
- There are 5541 different species in this file
- The different categories are: mammal, bird, reptile, amphibian, fish, vascular plant, and non-vascular plant
- The different conservation statuses are: no intervention, species of concern, endangered, threatened, and in recovery
- 5363 species are classified as “no intervention”, 4 are “in recovery”, 151 are classified as “species of concern”, 10 are “threatened”, and 15 are “endangered”,


Conservation Status by Species
Number of Species
5000
4000
3000
2000
1000
0
In recovery
Threatened
Endangered
Species of concern
No intervention
Conservation Status

Significance calculations for endangered status between different categories of species:

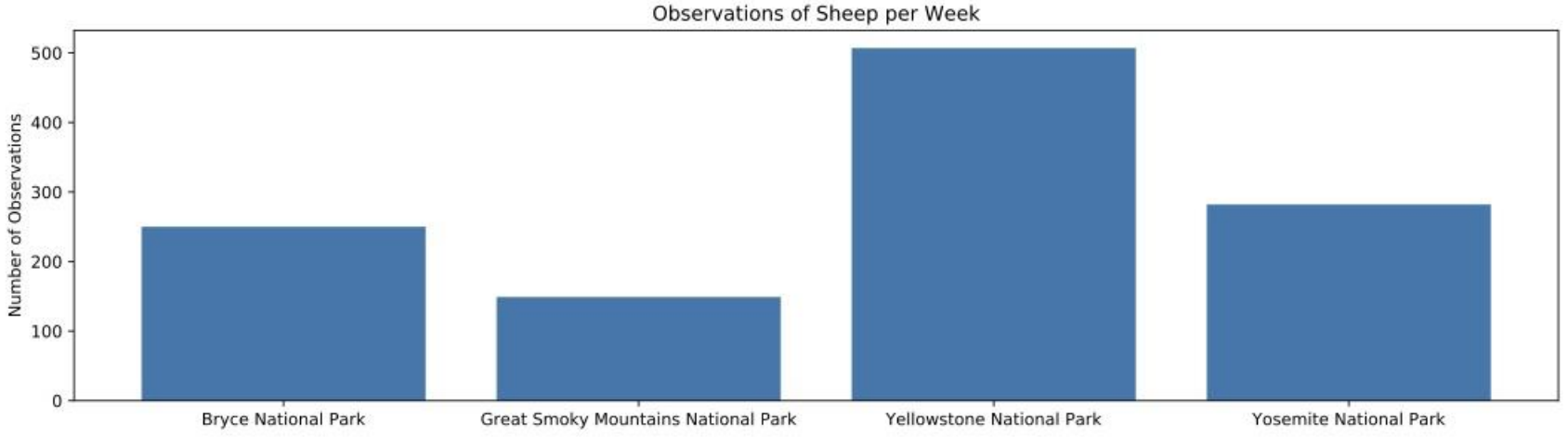
- I created a pivot table with values for each species
- The species in the table are grouped by their category or type (amphibian, bird, fish, mammal, nonvascular plant, reptile,and vascular plant)
- The table contains the number of each species that’s protected vs not protected, and the % of each category that is protected
- The data suggests that over 80% of these species are protected vs not protected across all categories, with plants having more protected species (by percentage of the whole) than animals

Chi-Squared test for significance of species data

- I ran a Chi-Squared test to see if the there was a significant difference between the likelihood that mammals are more endangered than birds.
  - This test showed that the difference between the two categories was not statistically significant, so mammals are no more likely to be endangered than birds (the test gave us a pvalue of 0.688 according to the provided data)
  - However, there is a significant difference between the likelihood that reptiles are endangered compared to mammals, suggesting that reptiles are more likely to be endangered (pvalue of 0.038)
  - This shows us that certain types of species are more likely to be endangered than others

Sample size determination for foot and mouth study

- I was tasked with determining the correct sample to use to study species of sheep across the National Parks for foot and mouth disease, as the rangers are conducting a program to reduce incidence of the disease
- I joined the table of species observations across the parks with the species_info csv to determine the common names and conservation status of several kinds of sheep
    - I cleaned up the sheet by checking to see which rows contained the word sheep, removing the ones that referred to plants
- I used this data to determine that there were 1188 sheep sighting across the 4 National Parks in the past 7 days.


Observations of Sheep per Week
Number of Observations
0
100
200
300
400
500
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park

Sample size determination for foot and mouth study (cont’d)

- Using these sighting numbers, I was able to determine that the best sample size for this study is 870 sheep
- The baseline rate for the study was 15%, as this many sheep had the disease last year.
- The minimum detectable effect for the study was 33.3% (a reduction of at least 5% in foot and mouth disease incidence) because they’d like to show that 10% of sheep have the disease, with confidence.
- Using these values, and the statistical significance value of 90% significance, I determined that they would need to observe 870 sheep for this study.
- This means that Yellowstone will need to observe for just over a week (1.7 wks) and Bryce will need to observe for a little over 3 weeks (3.48 wks)