Stephen Posek
10/20/2018

Capstone Option 2: Biodiversity for the National Parks

Part 1: Protected species data from the parks

The Data: species_info.csv

- The columns of species_info.csv consisted of the following values:
  - ‘category’
    - This column grouped each species by their categorical type
  - ‘scientific_name’
    - This column contained the scientific names of each animal found in the parks
  - ‘common_names’
    - This column contained the common names of each animal found in the parks
  - ‘conservation_status’
    - This column grouped each animal by a conservation status

The Data: Category

- The ‘category’ column contained the following classifications:
  - Mammal
  - Bird
  - Reptile
  - Amphibian
  - Fish
  - Vascular Plan
  - Nonvascular Plant

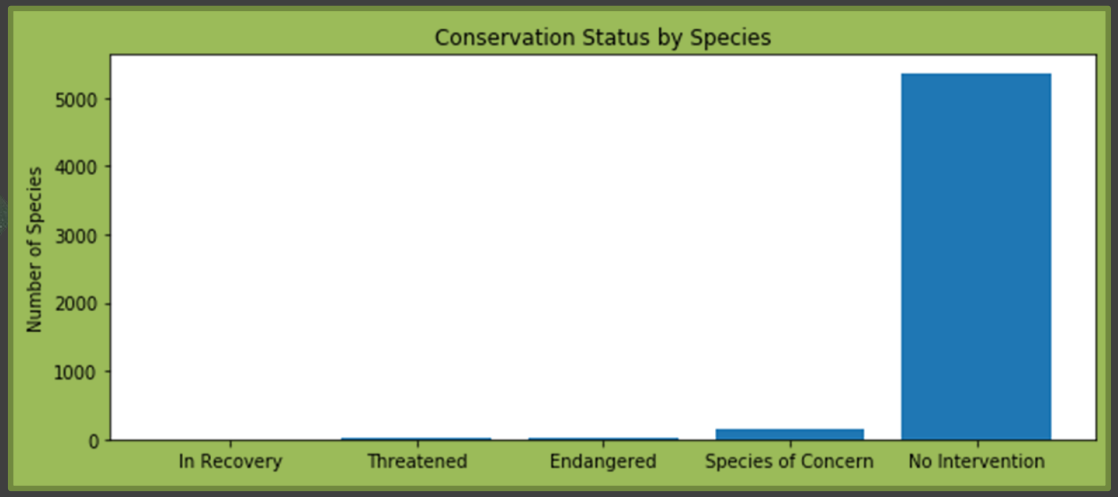
The Data: Conservation Status

- The conservation statuses were categorized as follows:
  - ‘Nan’
    - No conservation status listed on the animal
  - ‘Species of Concern’
    - Declining or appear to be in need of conservation
  - ‘Endangered’
    - Seriously at risk of extinction
  - ‘Threatened’
    - Vulnerable to endangerment in the near future
  - ‘In Recovery’
    - Formerly Endangered, but currently neither in danger of extinction throughout all or a significant portion of its range

The Totals

conservation_status	scientific_name
Endangered	15
In Recovery	4
No Intervention	5,363
Species of Concern	151
Threatened	10

Visualizing The Data



Key Calculations: Percent Protected Per Species

category	not_protected	protected	percent_protected
Amphibian	72	7	.088608
Bird	413	75	.153689
Fish	115	11	.087302
Mammal	146	30	.170455
Nonvascular Plant	328	5	.015015
Reptile	73	5	.064103
Vascular Plant	4216	46	.010793

Key Calculations: What does it tell us?

- By looking at the data on the previous slide, we see that Mammals and Birds respectively hold the two highest percentages for species categorized as protected
- In order to determine if certain types of species are more likely to be protected than others, we created contingency tables consisting of mammals and another species, and compared them using a Chi-squared significance test

Is it Significant: Mammals vs Birds

- We used the following contingency table to perform a Chi-squared test for Mammals & Birds

	protected	not protected
Mammal	30	146
Bird	75	413

- From the test, we were able to determine a p-value of ~68% meaning the difference is not significant and is the result of chance

Is it Significant: Mammals vs Reptiles

- We used the following contingency table to perform a Chi-squared test for Mammals & Reptiles

	protected	not protected
Mammal	30	146
Reptiles	5	73

- From the test, we were able to determine a p-value of ~3.8% making the difference significant and validating the statement “Mammals are more likely to be protected than reptiles.”

Conservation Recommendation

- Based on the significance calculations we performed, we can make two recommendations regarding conservation efforts:
  - The conservation efforts for Mammals and Birds should be of equal priority
  - The conservation efforts for Mammals should be prioritized over that of the conservation efforts of Reptiles

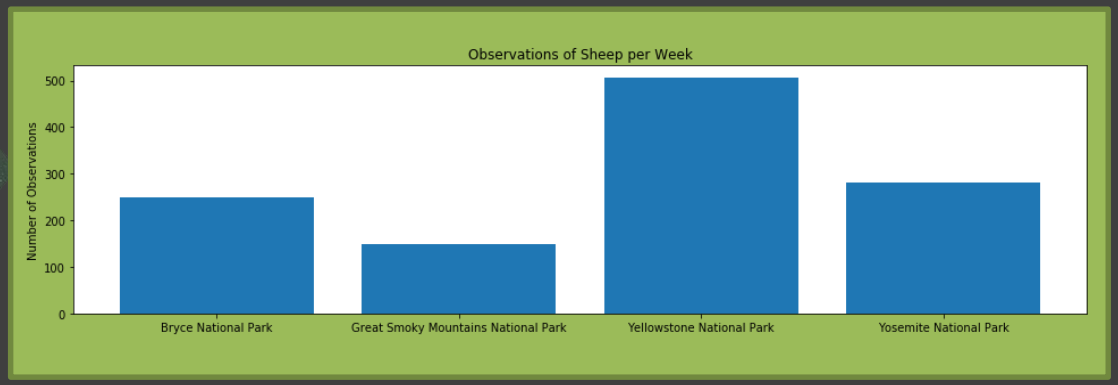
Part 2: Sheep Sightings & Foot and Mouth Disease Sample Size Determination

The Data: Sheep Sightings

- Given a dataset by the National Parks Service titled 'observations.csv' we were asked to help analyze sheep sightings at several national parks over the past 7 days
  - With the provided observations dataset combined with the data from our 'species_info.csv,' we were able to determine the number of sheep sightings in each park over a 7 day period

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

Visualizing The Data: Sheep Sightings



Foot and Mouth Reduction Effort

- Background:
  - The Yellowstone National Park Rangers have been running a program in an effort to reduce the rate of foot and mouth disease at the park and asked for our help in determining the necessary sample size
- They wanted to be able to detect reductions of at least 5 percentage points. With a given baseline of 15%, a level of significance equal to 90%, and our data from sheep sightings we were able to determine the following:
  - Minimum detectable effect = 33%
  - Sample size per variant = 890
  - Weeks of observation required for Yellowstone = 1.75
  - Weeks of observation required for Bryce = 3.56