Biodiversity For The National Parks

Introduction to Data Analysis / Codecademy Pro Intensive

Joshua Wallin

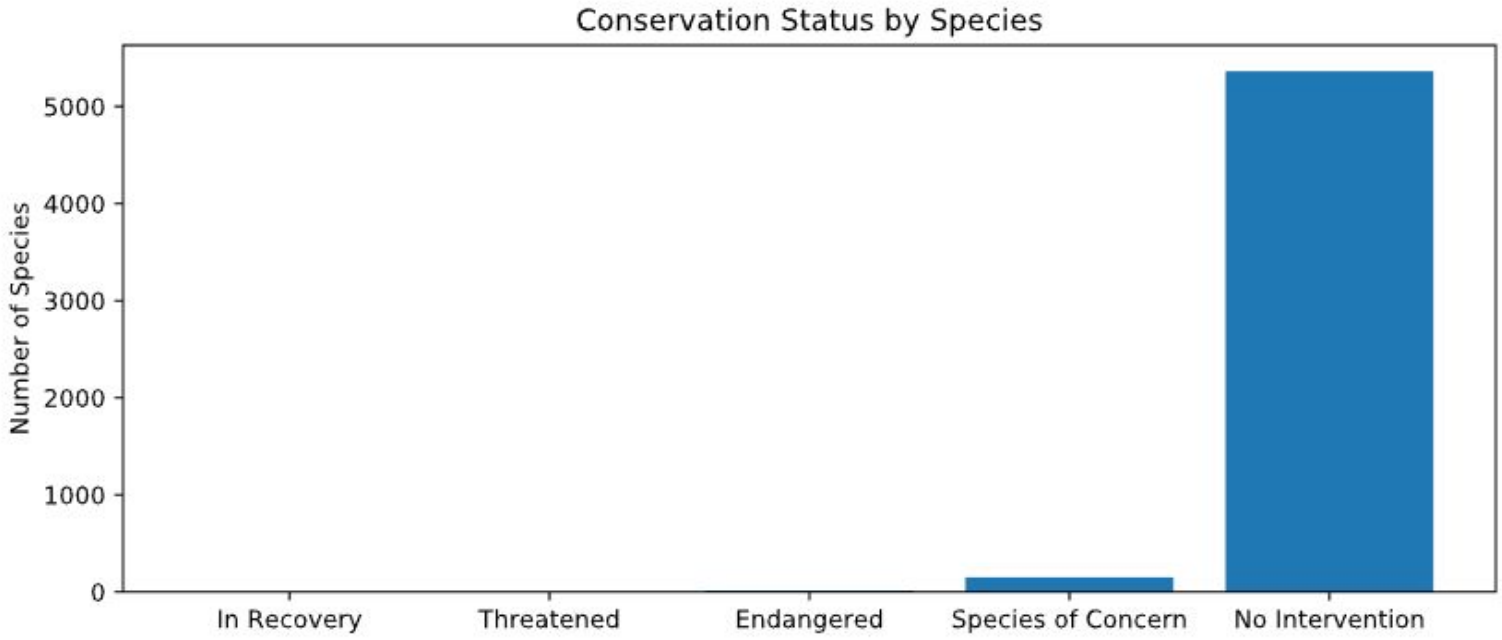
Information On The Various Species Within The National Parks

*CSV file species_info.csv provided by the National Parks Service

The provided data contains information, including **Category, Common & Scientific Names, & Conservation Status**, on 5541 species found within the National Park System.

Each species is categorized as **Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, or Nonvascular Plant**. The status of each species is listed as **Endangered, In Recovery, No Intervention, Species of Concern, or Threatened**.

Interestingly, Vascular Plants have the largest number of species at 4258. Birds have 488, Nonvascular Plants 333, Mammals 176, Fish 126, Amphibians 79, & Reptiles 78 species.



When the data is grouped by Conservation Status, **97%, or 5363**, of the combined species are listed as **No Intervention**.

Endangered Species

When the number of Protected species is compared to the species total per category, **Mammals and Birds have the highest Protected percentage at 17% and 15%** respectively. **Amphibians & Fish have 9%, Reptiles 6%, Nonvascular Plants 2%, and Vascular Plants 1%.**

Although the difference between Mammals & Birds, Amphibians & Fish, and Vascular & Nonvascular Plants is not statistically significant, **the difference between Mammals & Reptiles is significant**. Given that the difference, it is reasonable to assume that **certain species are more likely to become endangered**.

Recommendation For Conservationists

Equipped with this knowledge, it may be prudent for conservationists concerned about the endangered status of these species to prioritize their efforts and focus on the categories with the highest percentage of Protected species:

1) **Birds & Mammals**
2) **Amphibians & Fish**
3) **Reptiles**
4) **Nonvascular & Vascular Plant**s

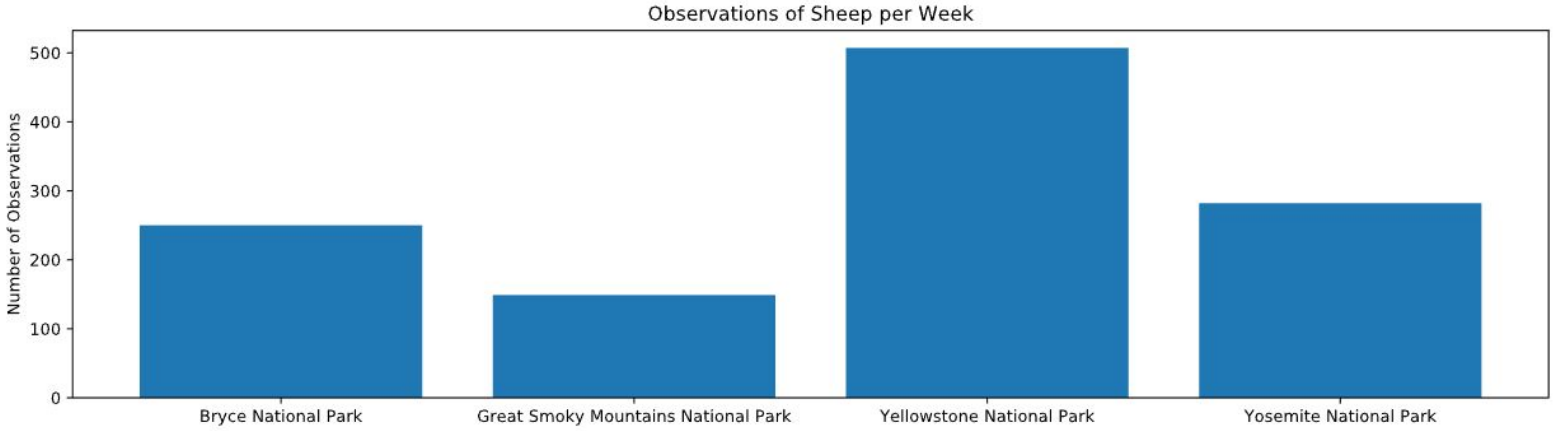
Species Observation

*CSV file observations.csv provided by National Park Service

Utilizing the observation data provided by the Park Service and merging it with the previous data, the location for each species within the system can be tracked. That data can assist park rangers in their attempt to curb foot and mouth disease among specific sheep species within each park.

Filtering the data for weekly sheep sightings produces the following results:

1) **Bryce National Park has 250,**
2) **Great Smoky Mountains National Park has 149,**
3) **Yellowstone National Park has 507**
4) **Yosemite National Park has 282**



Estimated weekly sheep observations per park

Determining The Correct Sample Size

That observation data can be plugged into a sample size calculator to determine the correct sample size needed to detect the desired change among the infected sheep. **In the case of Yellowstone National Park, the desired result is a minimum reduction of 5% in foot and mouth disease among the infected 15% of the sheep population.**

With the aid of a sample size calculator, **a sample size of 890 sheep is needed to give the rangers 90% confidence in determining a 5% reduction**. Knowing that there are an estimated 507 weekly sheep sightings in Yellowstone National Park, it can be assumed that i**t would take around 1.75 weeks to collect the needed samples.**

Sample Size Calculator provided by Codecademy

Determining The Correct Sample Size

If the infection data is consistent among the sheep species in the remaining parks:

1) **Bryce National Park requires a sample size of 890 and 3.56 weeks to collect the needed samples.**
2) **Great Smoky Mountains National Park requires 890 samples and 5.97 weeks to collect the needed samples.**
3) **Yosemite National Park requires 890 samples and 3.04 weeks to collect the needed samples.**