Biodiversity for US National Parks

IDA Intensive Capstone

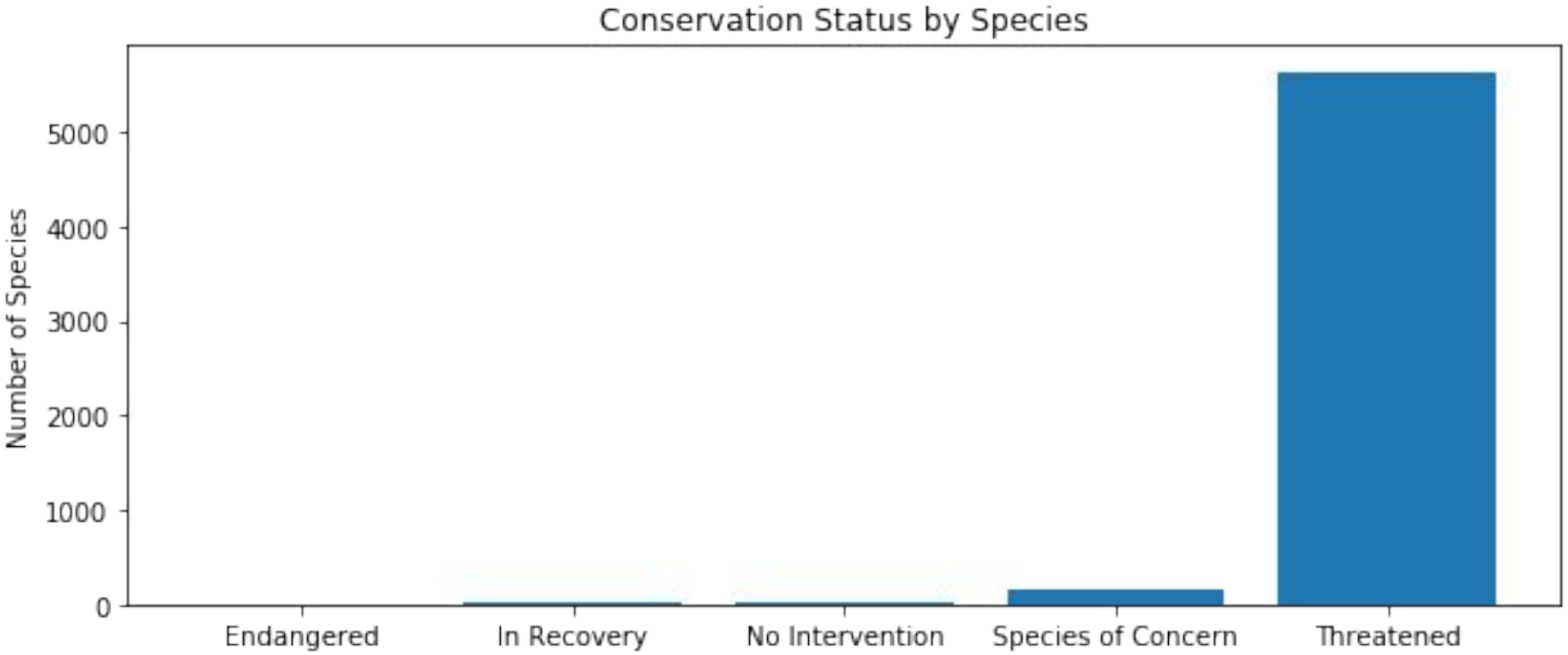
Natural Selection

- The data provided has over 5800 entries with around 5500 unique species in the catalog
- Categories include Mamalian, Avian, Reptilian, Amphibian, Pescataerian, and Plantae entries
- Species have been grouped into four categories
    - “Species of Concern”, “Endangered”,”Threatened”, and “In recovery”
    - All null values are grouped into “No Intervention”

Counting Categories

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

Visualizing the Biology



Statistical Significance Tests

- Ultimately, one can group the categories of species between protected or not protected, filtering and narrowing down to two choices
- Contingency tables were constructed to compute Chi Squared Tests
  - Tabulated values contain counts of each category of species and protection
  - Scipy Module used to compute with Python

Results:

Mammal v Bird	Reptile v Mammal	Bird v Fish v Mammal v Plant
False	True	True

0.08%

Average species protected among all cateogries

Recommendations

A higher percentage of mammalian species are protected than reptilian species, focus efforts on research and developing services.

How to determine sample size for a Disease study

Find Baseline percentage

-15% from known study at Bryce National Park

Determine Minimum Detectable Effect

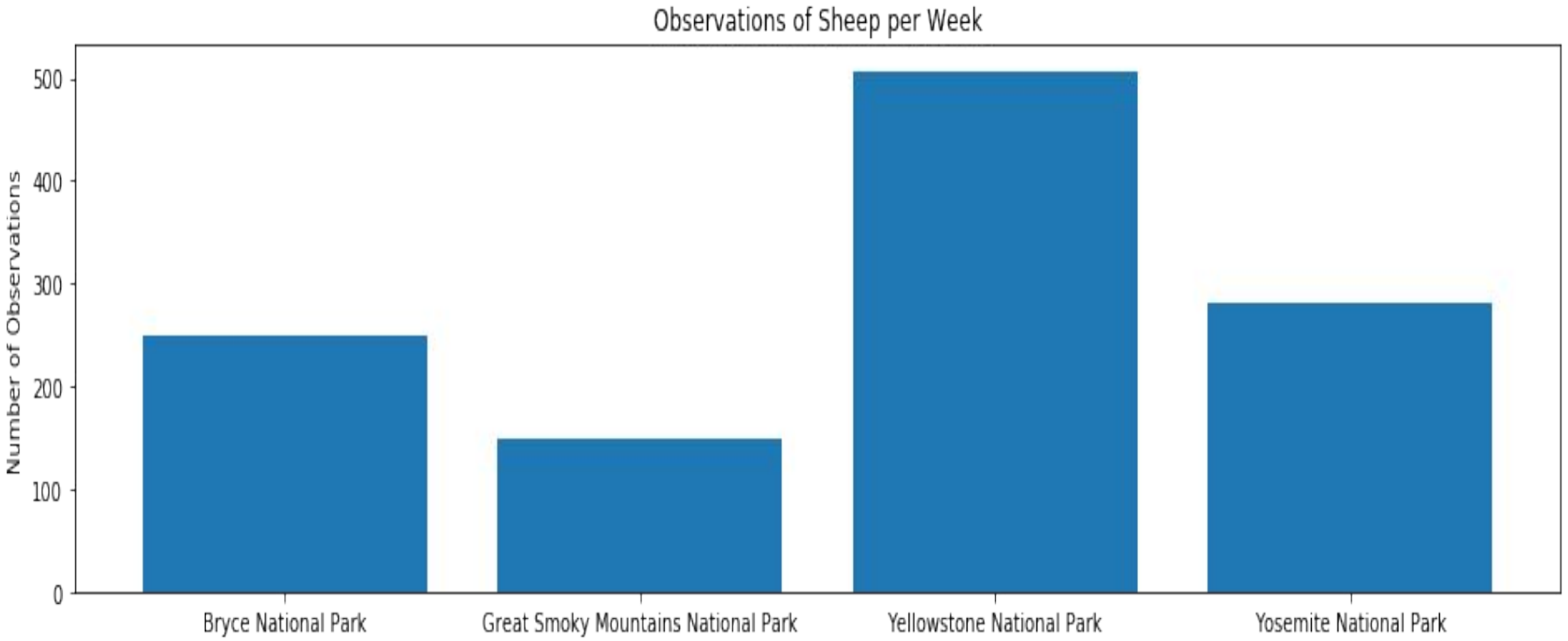
Select a level of Significance

Calculate using Codecademy sample size calculator web link

Sample Size needed: 870

Weeks to observe full sample size at Bryce National Park and Yellowstone National Park based on weekly observations: 4 weeks at Bryce, 2 weeks at Yellowsone.

Conservationists at work



Thank you, notebook uploaded to Codecademy online