Biodiversity for the National Parks

Capstone Project

Pro Intensive: Introduction to Data Analysis

(30th Oct 2018 – 30th Jan 2019)

Created by Tanisha Perry-Warner

Dataset 1: species_info.csv

- Contains the following fields of data for different species in our National Parks:
  - The specific name of each species
  - The common names of each species
  - The species conservation status

- High level data summary:
  - Types of specie = 5541
    - Categories of specie = 7
      - Conservation status types = 5

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

Above: data sample

Focus:
Endangered status by specie category

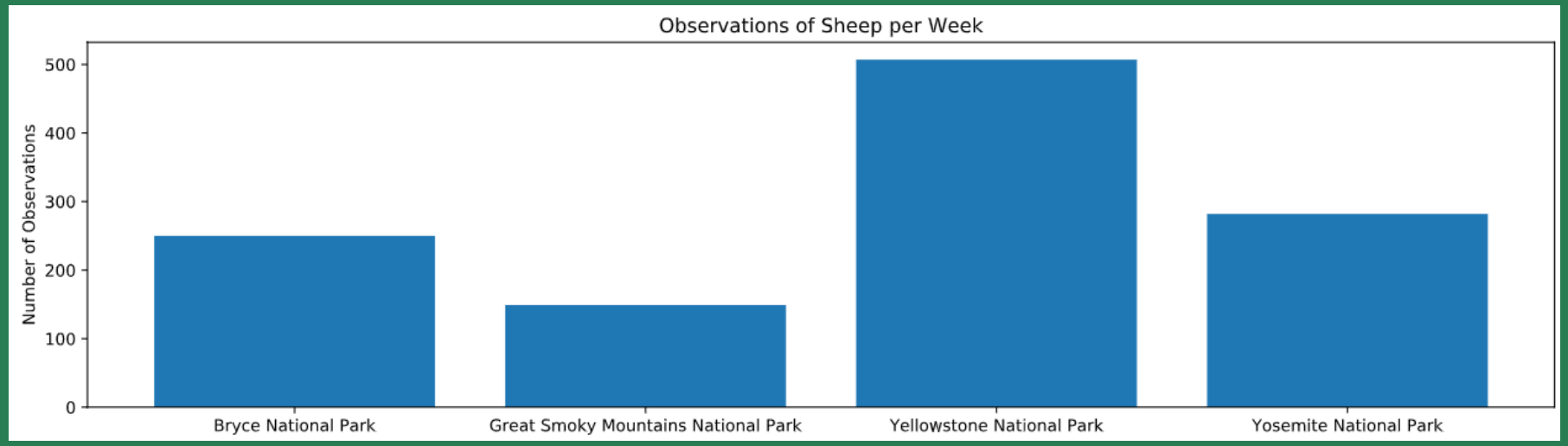
- When comparing the protected conservation status of specie categories, it is apparent that some species are more likely to be endangered than others.
- On concluding a Chi-Squared Test it can be said that a significant relationship exists between the percentage of protected reptiles and mammals (~p-value of 0.038).

<u>Recommendation for conservationists</u>

- The factor(s) affecting the protection of mammals and reptiles show a significance. Perhaps they are the same factor(s)?
- Addressing this could help to improve the protection status of both specie categories.

Dataset 2: Specie Observations
Foot and Mouth Disease Study

- A 7 day observations of sheep at National Parks showed that Yellowstone National Park had the highest recordings. This is the best park to observe sheep.



- After carrying our a disease reduction study for confident measurements (>5% drop in observed cases), the necessary sample size is 870. It would take Yellowstone Park approximately 1 week to complete this.

Additional Graphs