Codecademy Capstone Project
Biodiversity of the National Parks

Barry Cahill

08.09.2018

1. species_info.csv – about the data-set

- The data set contains the conservation status of 5824 species in 7categories.
  - Breakdown of the data by category:

Amphibian	80
Bird	521
Fish	127
Mammal	214
Nonvascular Plant	333
Reptile	79
Vascular Plant	4470

Each entry in the database contains a category, a scientific name and a common name and the conservation status.

2. species_info.csv - data-set summary

- Note that the majority of species in the dataset have no intervention status (= species not currently at risk of endangerment).
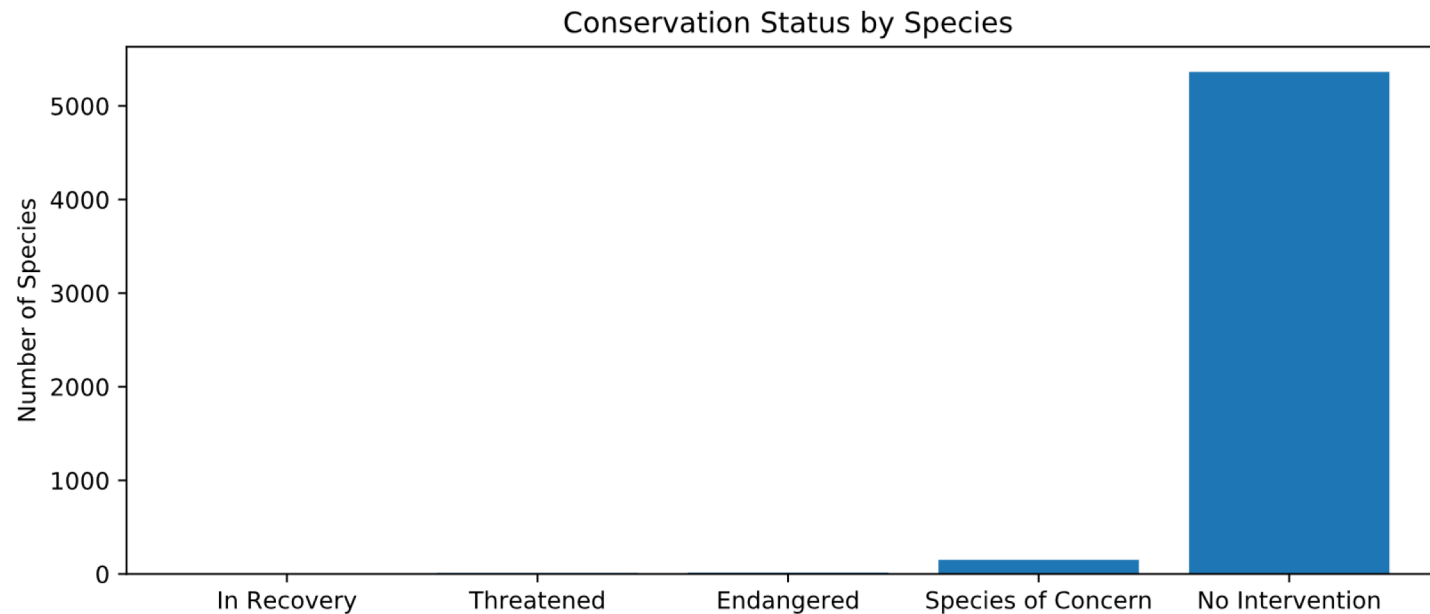
3. species_info.csv - data-set summary

- Percentage of each species which is categorised as protected :

| Category	not_protected	protected	percent_protected
Amphibian	72	7	8.86%
Bird	413	75	15.37%
Fish	115	11	8.73%
Mammal	146	30	17.05%
Nonvascular Plant	328	5	1.50%
Reptile	73	5	6.41%
Vascular Plant	4216	46	1.08%

4. species_info.csv – significance calculations

- According to our chi-squared tests, we can conclude:
  - The difference between the percentage of birds protected (15.37%) and mammals protected (17.05%) is not stastically significant [p-value = 0.6875]
  - The difference between the percentage of reptiles protected (6.41%) and mammals protected (17.05%) is statistically significant [p-value = 0.0383]

Conclusion: Some species are more likely to be endangered than others

Recommendation: Focus on protection of mammals & birds, as they are at highest likelihood of endangerment

5. Sheep Observations at national parks

Observations of Sheep per Week
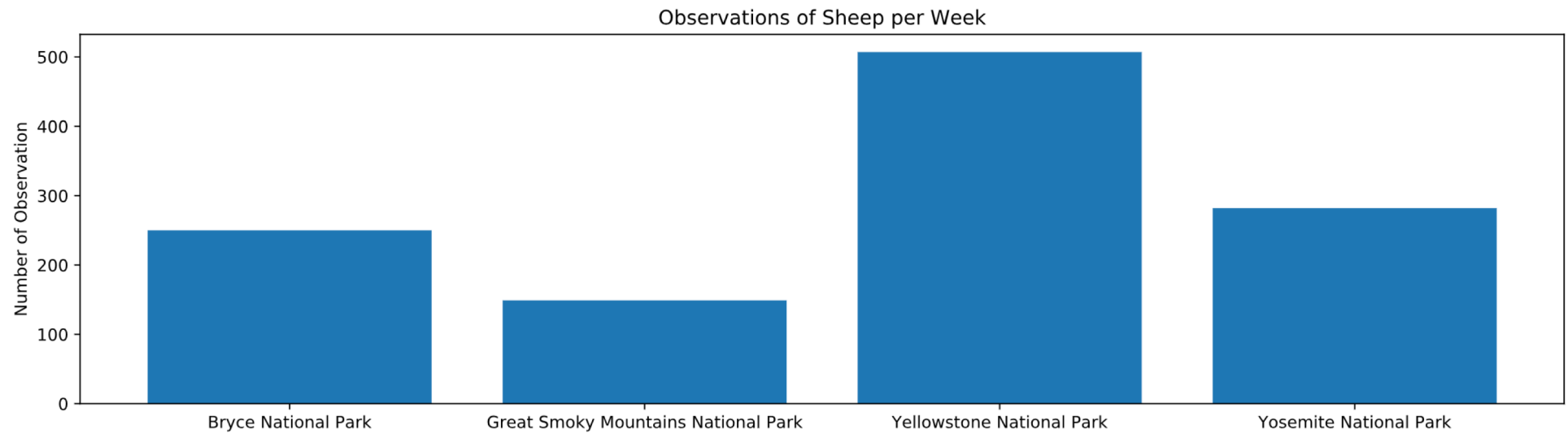
6. Sample size determination - foot & mouth

- We have a baseline of 15% occurrence of foot & mouth disease in the sheep (as reported by the rangers at Bryce national Park).

- We want to detect 5% changes

- Out minimum detectable effect is 33.35% (5% desired detection of change / 15% baseline)

- As such, sample size must be **870.**

Baseline conversion rate:	15	%
Statistical significance:	85%	90%
	95%	
Minimum detectable effect:	33.3	%
Sample size:	870	

Conclusion: Researchers must observe 870 sheep in order to have a statistically significant sample

7. Sample size determination - foot & mouth

- Yellowstone observes 507 sheep/week, so it can reach this sample in approx. 1.7 weeks

- Bryce national park observes 250 sheep/week, and can reach this sample size in 3.5 weeks.