Biodiversity for the National Parks

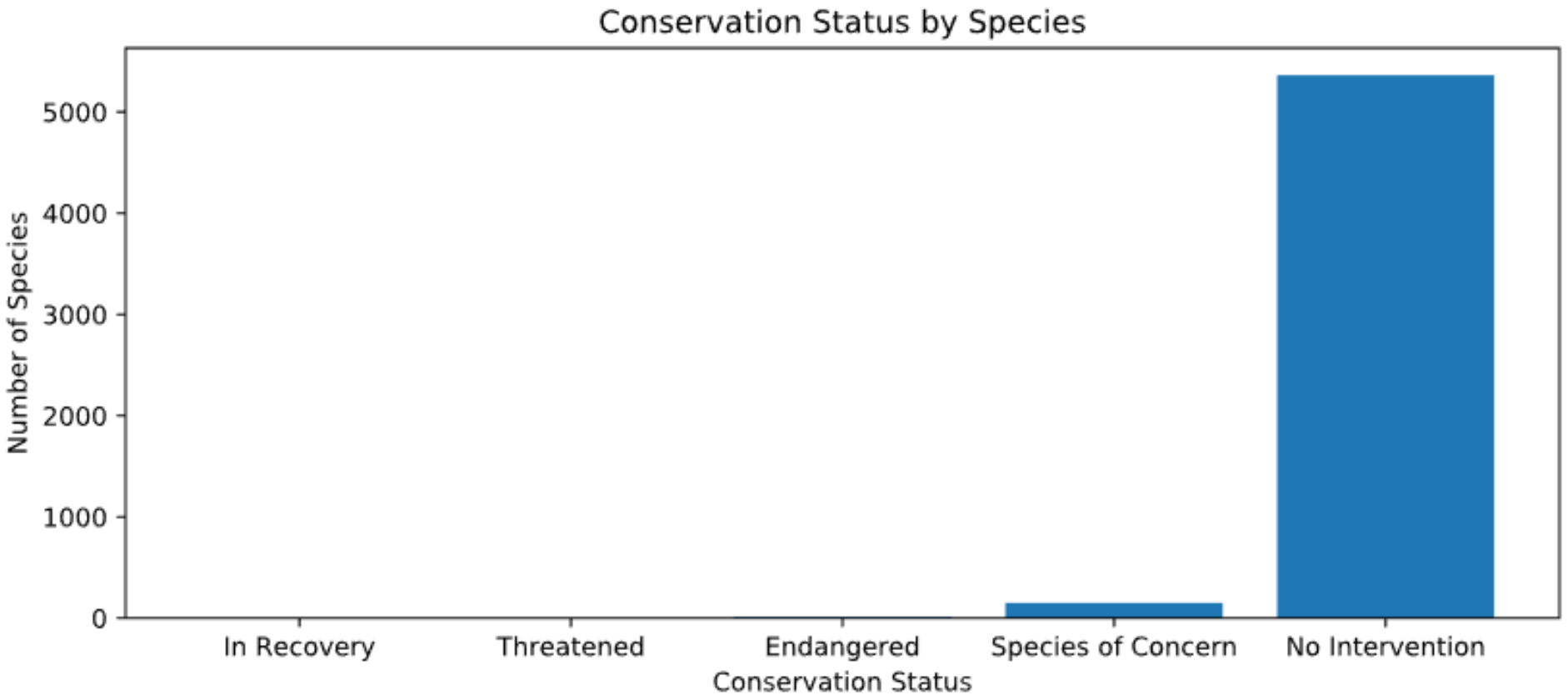
Data Analysis approach on National Parks species

Species in National Park

Data frame contains information about the following:

- 5541 variety of species
- Scientific name of each species (data type: string; e.g. Cervus elaphus)
- Common names of each species (data type: string; e.g. Wapiti or Elk),
- Each species fall into 1 of 7 different categories: Amphibian, Bird, Fish, Mammal, Nonvascular Plant, Reptile, or Vascular Plant (data type: string)
- Each species are labeled as 1 or 5 different conservation status: Species of Concern, Endangered, Threatened, In Recovery, No Intervention (data type: string)

Number of Species within Each Conservation Status



Conservation Status	No. of species
In Recovery	4
Threatened	10
Endangered	15
Species of Concern	151
No Intervention	5363

Among the species that are marked as needing intervention, majority has the conservation status of "Species of Concern"

Percent of Endangered Species by Category

Category	Protected	Not Protected	Percent Protected
Amphibian	7	72	8.86%
Bird	75	413	15.37%
Fish	11	115	8.73%
Mammal	30	146	17.05%
Nonvascular Plant	5	328	1.50%
Reptile	5	73	6.41%
Vascular Plant	46	4216	1.08%

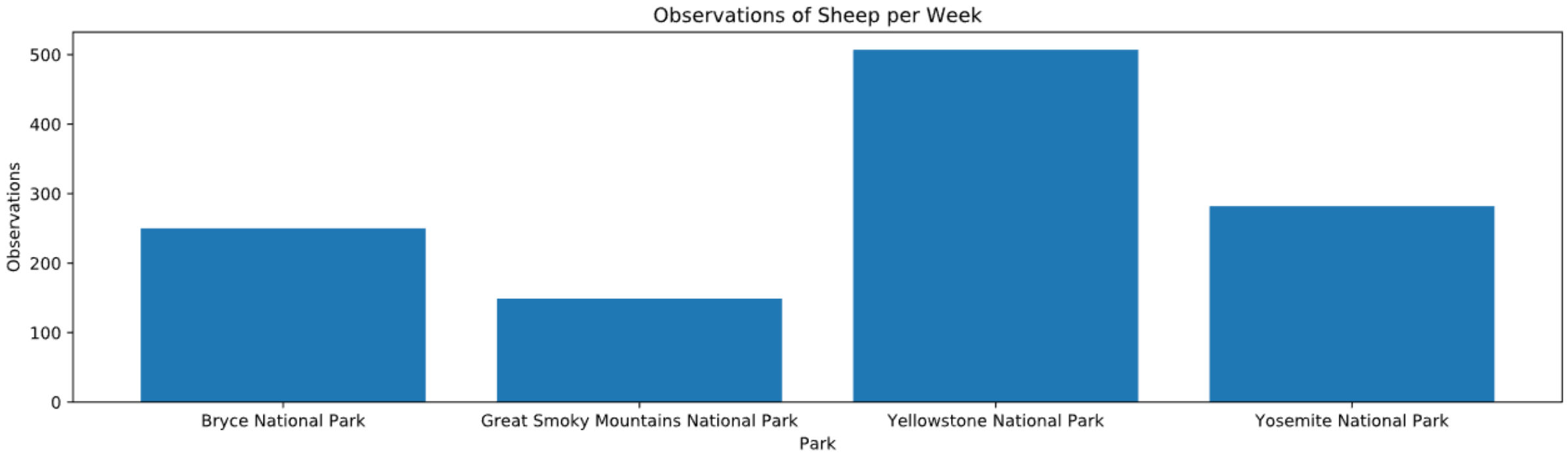
After observing a slight difference in the percentages of birds and mammals that fall into a protected category, a chi-squared test was run to determine whether the difference was significant or a result of chance.

Differences observed in percentages between different category of species that fall into the protected category

Significance Calculation between Species

- The chi-squared test between mammals and birds found a p-value of ~ 0.688. Therefore, there is not enough evidence to reject the null hypothesis that the difference in percentages of birds and mammals that fall into the protected category is by chance. The difference is not significant.
- However, a chi-squared test run between reptiles and mammals did prove that the difference between reptiles and mammals is significant. The chi-squared test found a p-value of ~0.038. Therefore, the null hypothesis can be rejected and it can be concluded that certain types of species are more likely to be endangered than others.
- Based on this calculation, our **recommendation** to conservationists concerned about endangered species is species that are more likely to be endangered must be prioritized and need more immediate attention than other species that are less likely to be endangered.

Observations of Sheep per week by park



Yellowstone National Park has the highest observed sheep per week compared to other parks

Sample size for Foot and Mouth Disease Study

- To find a reduction of at least 5 percent point, given a baseline of 15%, the minimum detectable effect is 33.33%.
- To be able to measure at least 5 percent drop in observed cases of foot and mouth disease in the sheep at Yellowstone and the result to be significant at 90% confidence level, the scientists would have to observe **at least 870 sheep**.
- The scientist will need to spend **1.7 weeks in Yellowstone** National Park to observe enough sheep
- If the study is repeated in **Bryce** National Park, scientists will need to spend **3.48 weeks** to observe enough sheep. The time of observation is longer in Bryce since previous data showing observed sheep shows fewer observed sheep in Bryce compared to Yellowstone (250 observed sheep vs. 507 observed sheep, respectively)

Thank You