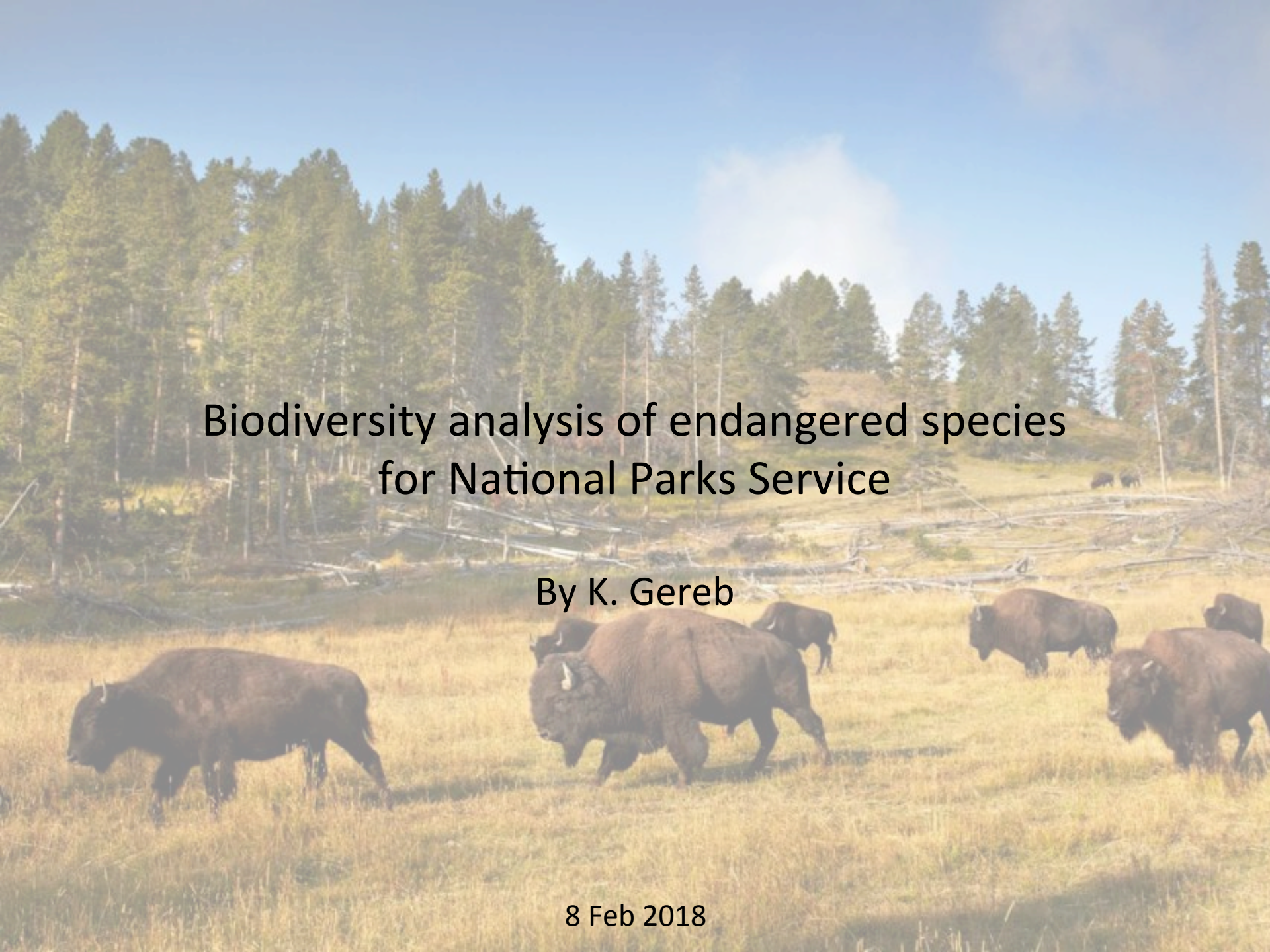


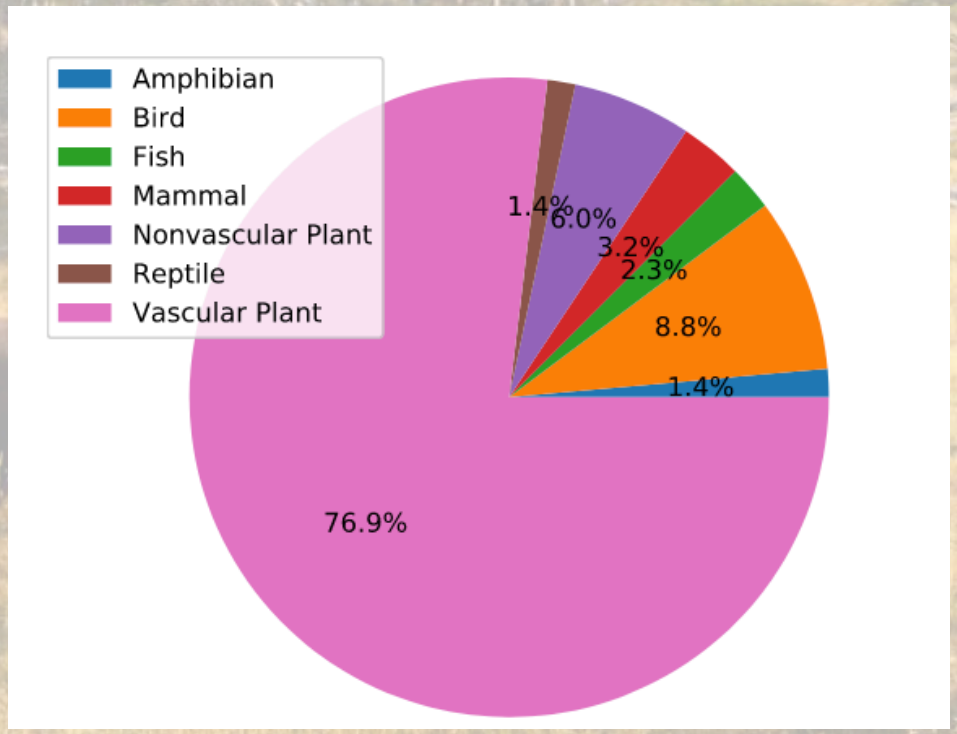
Biodiversity analysis of endangered species for National Parks Service

By K. Gereb

8 Feb 2018

Introduction

- The data provided by National Parks Service (species_info.csv) contains information on different species found at the park, the categories they belong to, and their conservation status.
- To summarize, there are 5541 different species belonging to 7 categories: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Non-vascular plant. A pie chart of these categories can be seen below. Vascular Plants make up ~77% of the categories, after which birds are the most numerous.

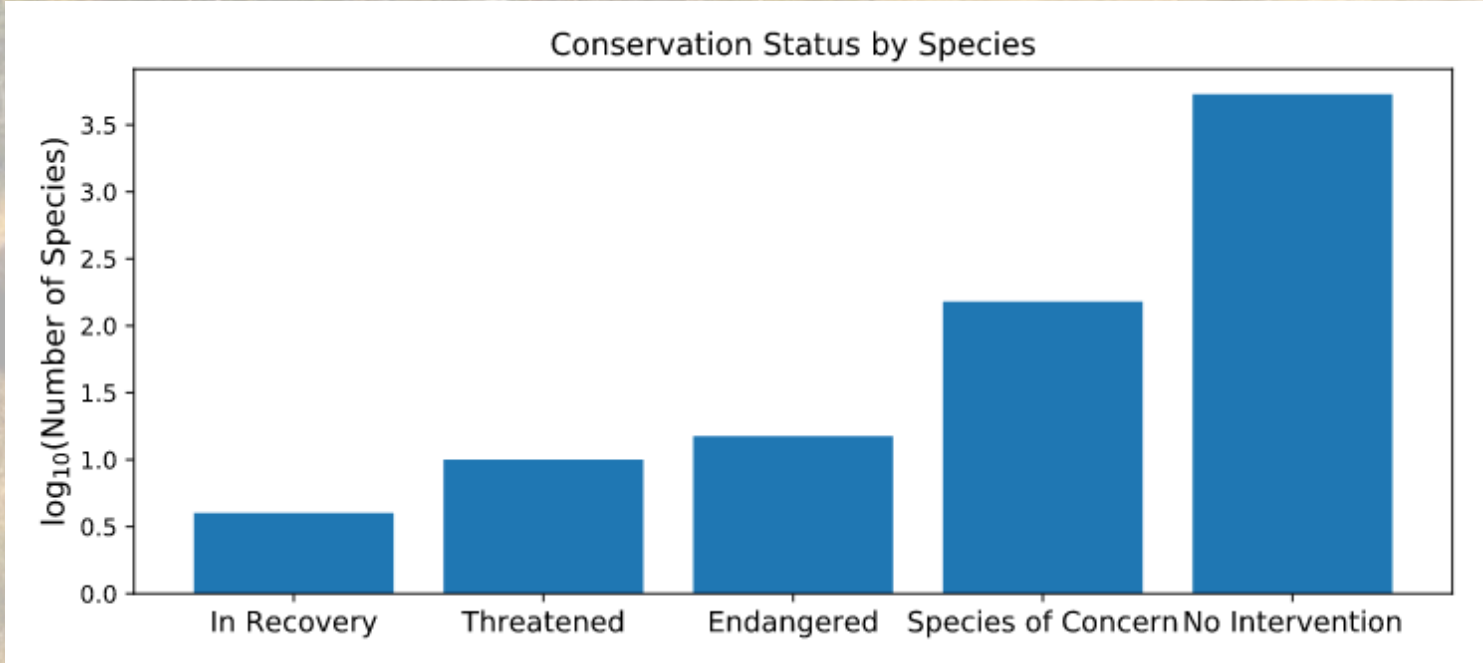


Endangerment analysis

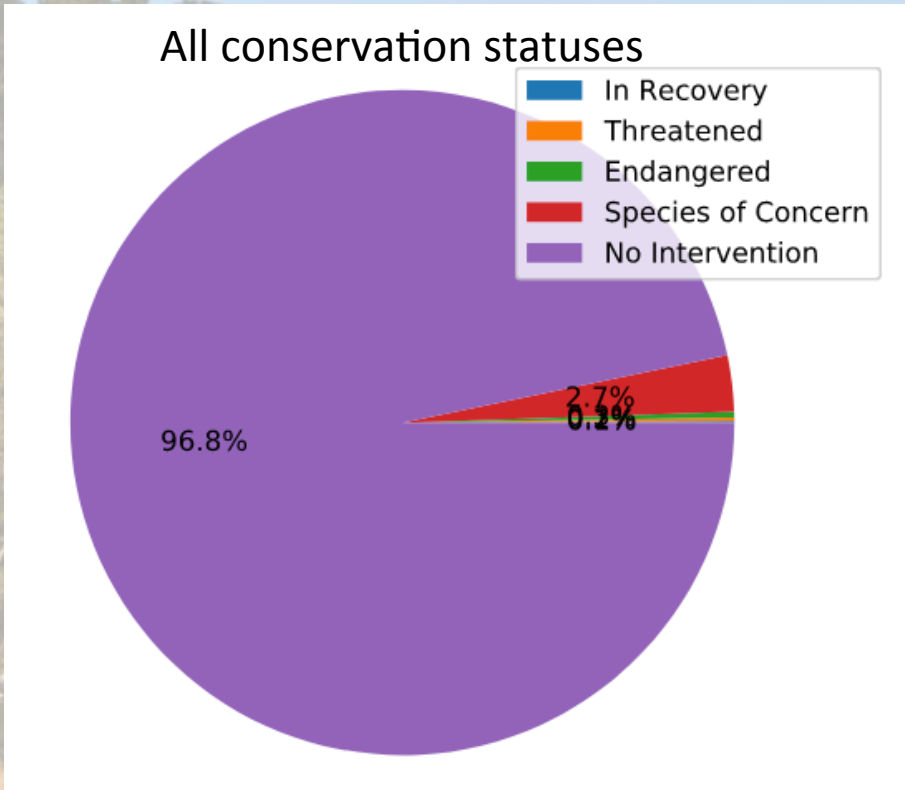
The conservation status of different species are:

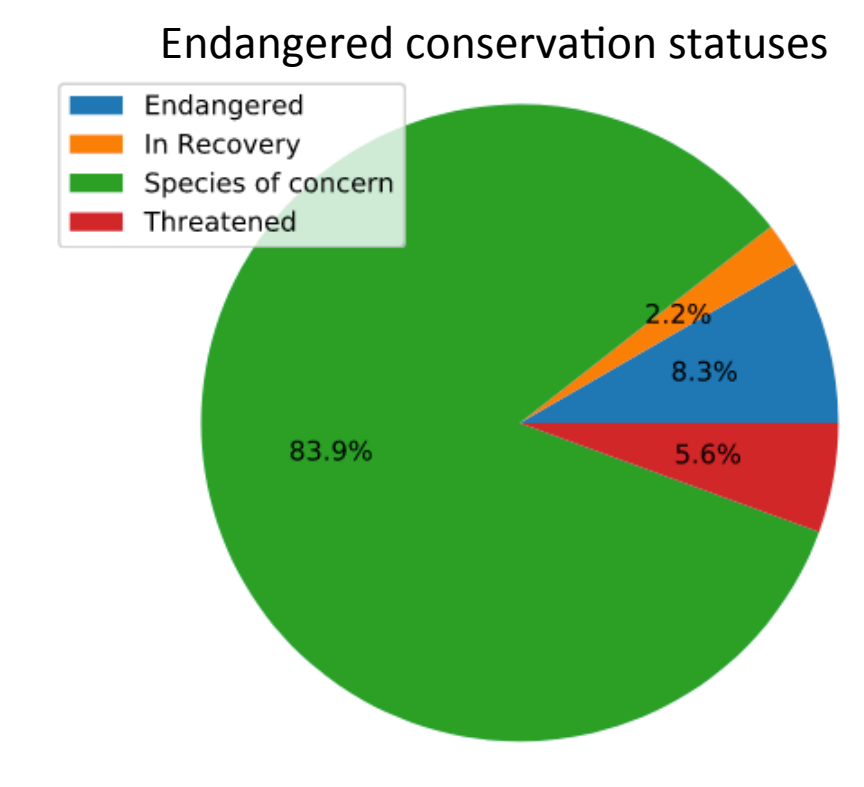
- Species of concern (151) – declining population, in need of conservation
- Endangered (15), seriously at the risk of extinction
- Threatened (10), vulnerable to endangerment in the future
- In Recovery (4), formerly endangered, but currently not in danger of extinction
- No intervention (5363): no conservation status is specified, no intervention is needed

More than 5000 species require no intervention (great!). The bar chart below (note log scale on y axis!) shows that this category outnumbers those that do need protection… see more details on the next slide.



Endangerment analysis





- 96.8% of the species requires no intervention, great!
- The remaining 3.2% (about 200 species) do need some sort of protection.
- The pie chart on the right focuses on the protected category, showing that “Species of concern” are the most numerous, making up 83.9% of the endangered species.

Endangerment analysis

Category	No protection needed	Protection needed	Percent_endangered
Amphibian	72	7	8.9%
Bird	413	75	15.4%
Fish	115	11	8.7%
Mammal	146	30	17.0%
Nonvascular Plant	328	5	1.5%
Reptile	73	5	6.4%
Vascular Plant	4216	46	1.1%

Are certain types of species more likely to be endangered?

- In the last column we show what percentage of a certain category needs protection. It looks like mammals have the highest percentage, but is this significant compared to other categories?

- We carry out a chi2 statistical test based on the data of mammals and birds. The null hypothesis is: the difference is due to chance. The chi2 test returns a p-value of 0.68, thus we can not reject the null hypothesis. There is no significant difference in the endangerment status of mammals and birds.

Endangerment analysis

Category	No protection needed	Protection needed	Percent_endangered
Amphibian	72	7	8.9%
Bird	413	75	15.4%
Fish	115	11	8.7%
Mammal	146	30	17.0%
Nonvascular Plant	328	5	1.5%
Reptile	73	5	6.4%
Vascular Plant	4216	46	1.1%

Are certain types of species more likely to be endangered?

- We carry out a chi2 statistical test for mammals and reptiles. The null hypothesis again is: the difference is due to chance. The chi2 test returns a p-value of 0.04, thus we can reject the null hypothesis. There is a significant difference in the endangerment status of mammals and reptiles, mammals being more endangered than reptiles.

Conclusions and advice for National Parks Service

- Earlier we could see that 96.8% of the species requires no intervention, which is great!

- "Species of concern" are the most numerous within the endangered category, these are the declining population, in need of conservation. The best would be to interfere before the decline happens.

- Some species are significantly more endangered than others, see mammals vs. reptiles; Perhaps National Parks Service should focus their efforts on mammals and birds first.


Analysis of sheep counts at Yellowstone National Park

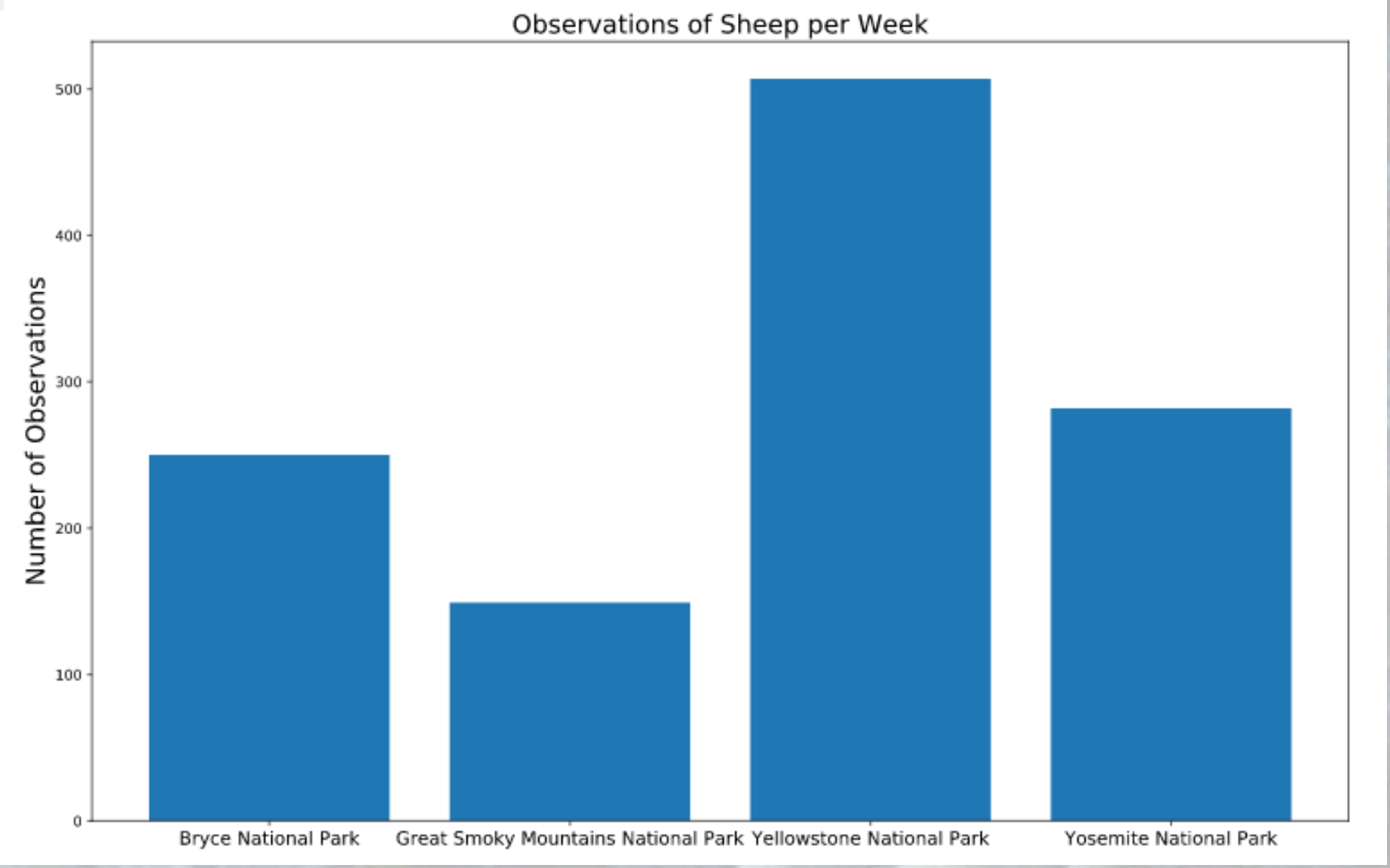
Introduction

- In this project we are carrying out an analysis for park rangers at Yellowstone National Park who want to know what percentage of sheep have foot and mouth disease.

- From our species_info.csv table we select species which have “sheep” in their common names. We merge these data with the observations.csv table to find out in which parks were these sheep seen, and how many were observed.

- We are going to use these data from previous years collected at 4 different parks for our analysis. This is illustrated in the bar chart on the following slide.

Analysis



The number of observations are the highest at Yellowstone National Park, where 507 sheep are seen per week.

Sample size determination

- In order to determine the percentage of sheep with foot and mouth disease with a 90% significance, we calculate the baseline percentage and minimum detectable effect. Then we plug these numbers in the sample size calculator to find out how big of a sample we need to observe.

- We know that 15% of sheep at Bryce had the disease last year, so this will be our baseline.

- The rangers want to be able to detect reductions of 5 percentage points. The Minimum detectable effect is a percentage of this number and the baseline. We want to observe a 5% change, thus the Minimum detectable effect = 100*5% / baseline. This gives 33%.

- Based on these numbers we'll need to observe at least 510 sheep according to the calculator. At Yellowstone we can do this in ~1 week, at Bryce it will take 2 weeks.

Conclusions

- Based on the sample size calculator, we need to observe 510 sheep to know what percentage has foot and mouth disease with a 90% significance.
- At Yellowstone National Park this can be done in about 1 week.
- Observations would take the longest at Great Smoky Mountains, almost 3 and a half weeks to collect all the data.