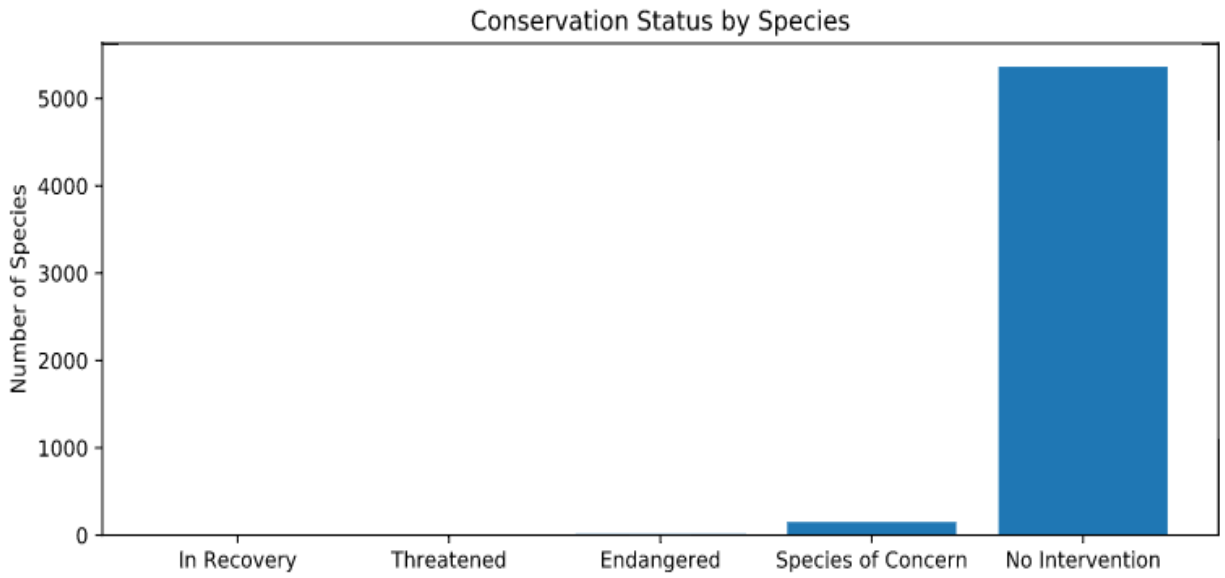
Biodiversity for the National Parks

Beata Mako (beata.mako@yahoo.com)

1. Conservation status by species

- After performing some data analysis on the species dataframe, I found that there are 5541 species across all national parks.
- 15 species fall into the endangered conservation status and 5363 species require no intervention. You can see the results in the table and the graph below.
- In order to get the results shown below, I counted the unique values in the scientific_name, category and conservation_status columns of the species dataframe. Then grouped the results by conservation_status and made sure that the null values are displayed as 'No Intervention'.

Conservation status	Number of species
Endangered	15
In recovery	4
No intervention	5363
Species of concern	151
Threatened	10



Are certain types of species more likely to be endangered?

- According to my initial analysis, it looks like mammals and birds are more likely to be endangered than any other types of species.
- 17% of mammals and 15% of birds are protected which means that these types of species need more protection than any other types of species shown in the table below.
- The breakdown of each category is listed in the table below.

Category	Not protected	Protected	% protected
Amphibian	72	7	9%
Bird	413	75	15%
Fish	115	11	9%
Mammal	146	30	17%
Nonvascular Plant	328	5	2%
Reptile	73	5	6%
Vascular Plant	4216	46	1%

Is the difference a result of chance?

In order to find out if the difference is a result of chance, I performed a couple of chi-squared tests.

a) chi-squared test comparing the percentages of mammals and birds

p value = ~ 0.688 = **not significant**

b) chi-squared test comparing the percentages of mammals and reptiles

p value = ~0.038 = **significant**

CONCLUSION: Certain types of species are more likely to be endangered than others.

2. Observations of sheep by national park

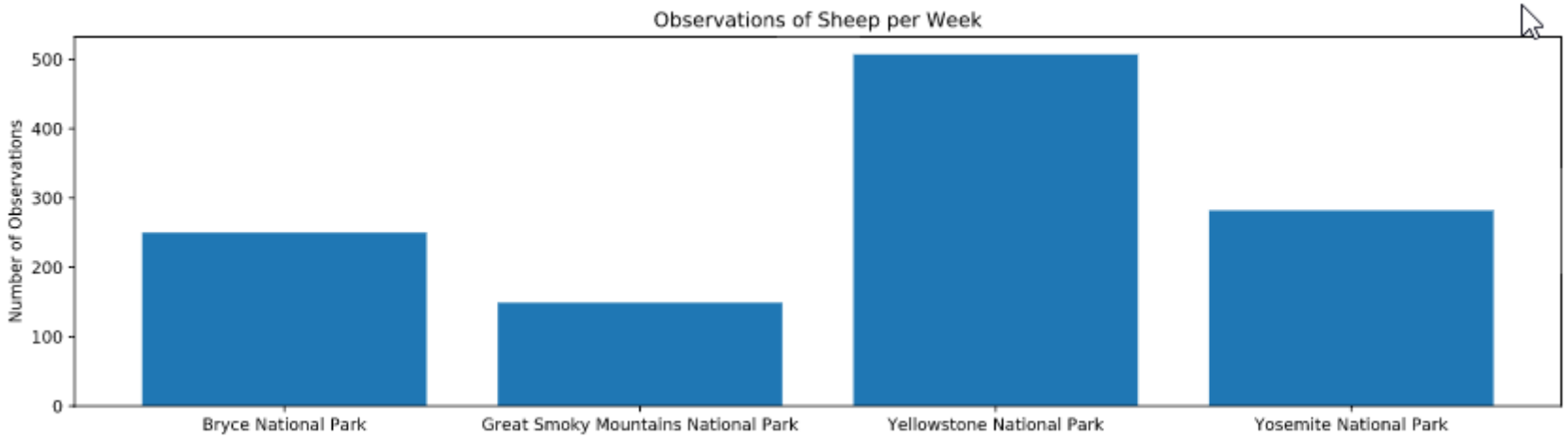
- In order to find out the number of observations of sheep across national parks, I performed some analysis on the observations dataframe.

- I used a lambda function to find the word ‘Sheep’ in the common_names column of the species dataframe and selected the corresponding rows.

- Then I merged the species dataframe with the observations dataframe. To get the total number of sightings of sheep at each national park, I grouped the data by park_name.

The table and graph on this slide show the number of observations of sheep at each national park per week.

Park name	Number of observations (per week)
Bryce National Park	250
Great Smokey Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282



Program to reduce foot and mouth disease

How many sheep should be observed to make sure that a 5% drop in observed cases of foot and mouth disease in the sheep was significant?

I have performed some calculations with Optimizely's A/B test sample size calculator to determine the sample size.

Baseline = 15%

Minimum detectable effect = 33%

Statistical significance = 90%

Sample size = 520

As per my calculations, the scientists would need to observe **520** sheep from each park in order to make sure that the foot and mouth percentages are significant.

How many weeks should scientists spend at each national park?

The table shows how many weeks the scientists would need to spend at each park at order to observe enough sheep.

Park Name	Number of weeks
Bryce National Park	2 weeks (14 days)
Great Smokey Mountains National Park	3.5 weeks (24 days)
Yellowstone National Park	1 week (7 days)
Yosemite National Park	1.8 weeks (13 days)