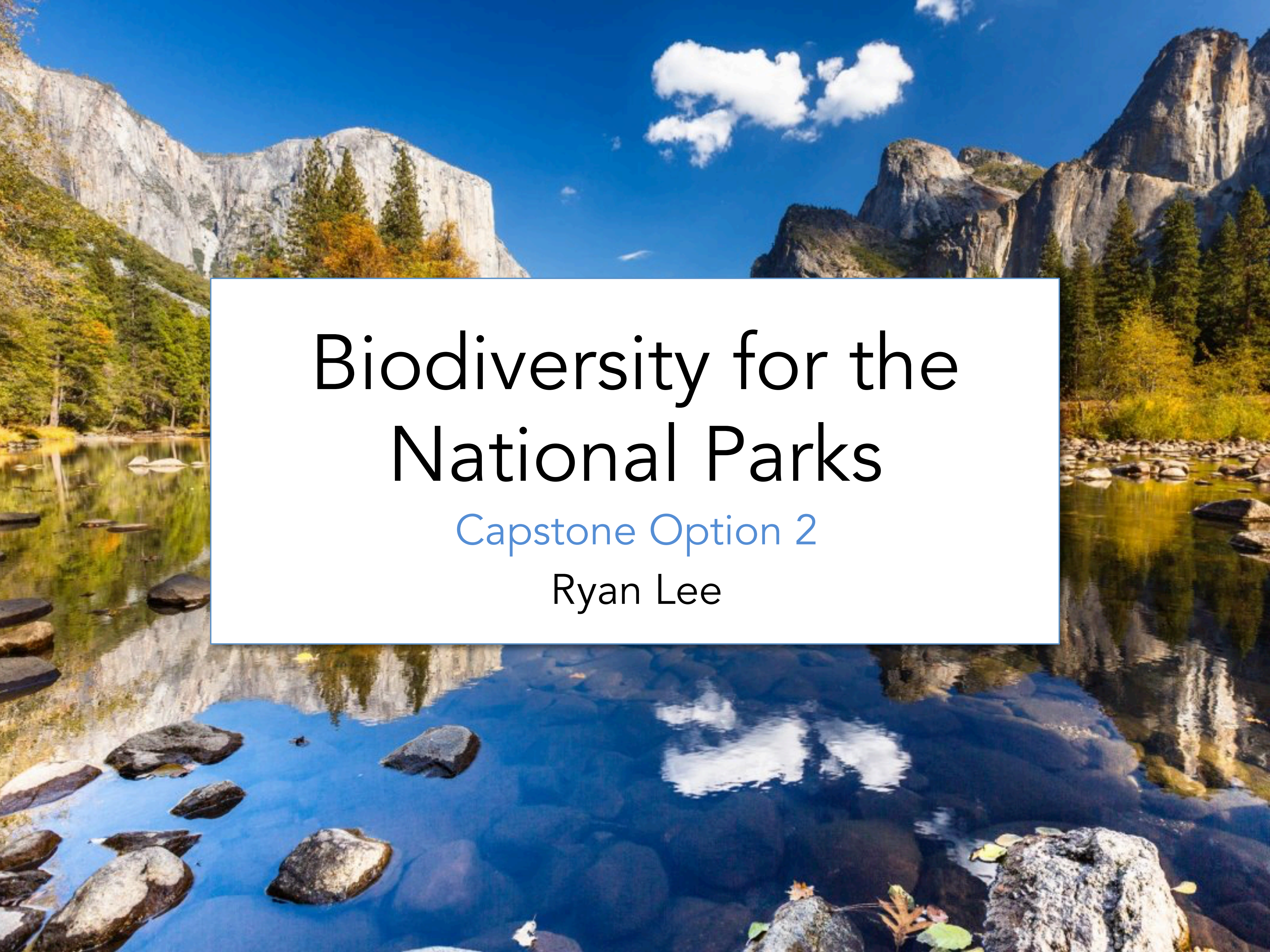


Biodiversity for the National Parks

Capstone Option 2

Ryan Lee

Getting to know the Species Dataframe

1 The Species Dataframe contains 4 columns of information:

- Category
- Scientific Name
- Common Names
- Conservation Status

2 There are 5,541 different species included in the dataframe

3 7 different categories of species

Mammal
Bird
Reptile
Fish
Vascular plant
Nonvascular plant

4 5 different types of "conservation statuses"

Endangered
In recovery
No Intervention
Species of Concern
Threatened

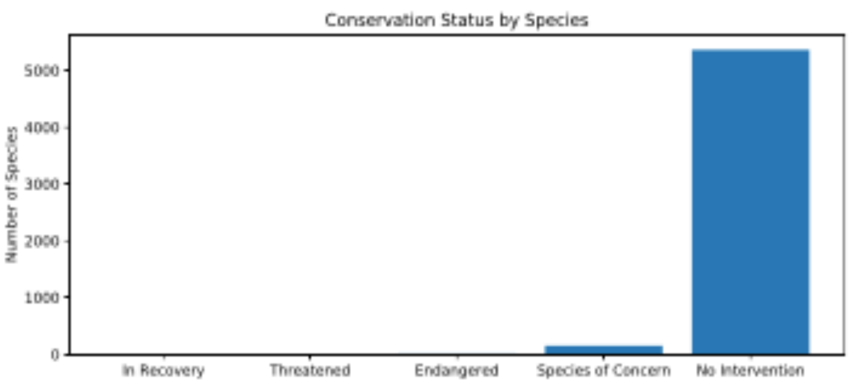
5 5,363 species do not need any protection right now ("No Intervention")

In other words, about 97% of the species in this dataframe have "no intervention"

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

6

The bar chart shows the stark difference in the amount of species that have “no intervention” vs. the other categories



7

According to the table, there are significantly more species “not protected” vs. “protected” for each category of species

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

The highest percent of category that is protected vs. not protected are Mammals at .17%, while Vascular Plants are the least at .01%

According to this data, of the categories in the Species dataframe, Mammals *seem to be* the most likely to be endangered

Significance Calculations

Mammals vs. Birds

- Based on the table in the previous slide, it looked like Mammals were more likely to be endangered than Birds, but we wanted to determine if this was a significant difference
  - *We used a Chi-Squared Test to determine whether this statement was true*
  - *In our test, our null hypothesis is that this difference is due to chance*
- The Chi-Squared Test resulted in a pvalue that was greater than .05 (actual = ~0.688)
  - *So, we can say that there is no significant difference; the relationship between Mammals and Birds, in this case, was due to chance*

Mammals vs. Reptiles

- Next, we wanted to know if the difference between Reptiles and Mammals was significant
- The Chi-Squared Test resulted in a pvalue that was less than .05 (actual = ~.038)
  - *So, we can say there is a significant difference, or "proof" of a difference between Mammals and Reptiles*

So what can we conclude from these findings?

We can conclude that certain types of species are more likely to be endangered than others

Recommendation for Conservationists

Based on my significance calculations, I would recommend that conservationists spend more time, attention, and resources protecting the species who prove to be more endangered. In this case, we found that mammals are more endangered than reptiles, so, I recommend that conservationists aim to protect mammals more than reptiles at this time.

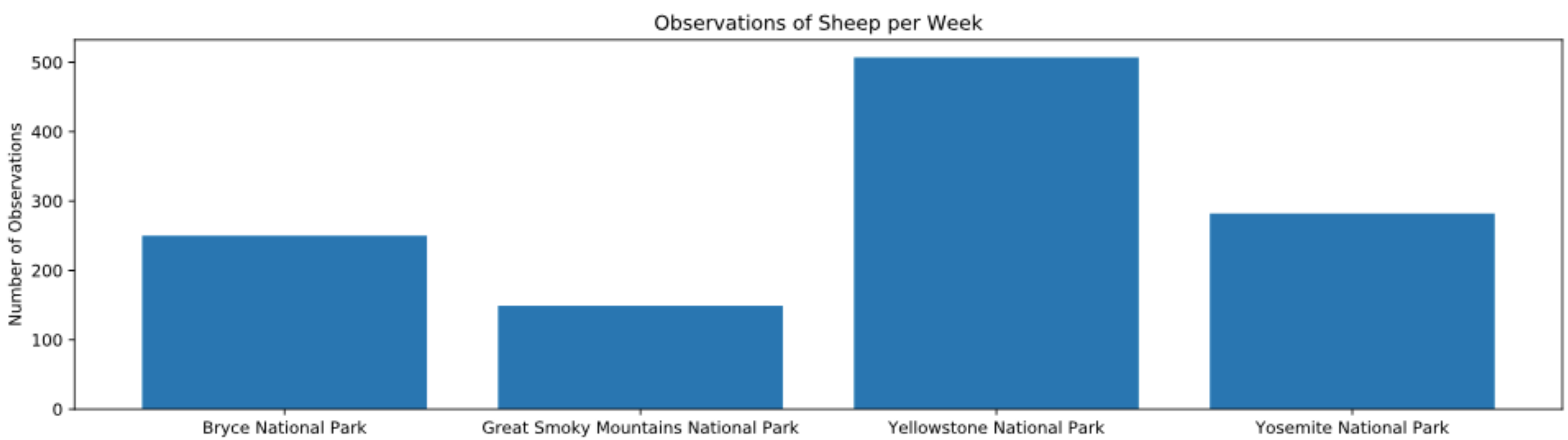
Out of my own curiosity, I performed chi squared tests on mammals vs. all other species in the species dataframe to see if the pvalue proved significant against more than just reptiles. The pvalues I calculated are below. I found that no other relationship with mammals proved to be significant.

Mammals vs. Fish pvalue = .0561
Mammals vs. Nonvascular Plant pvalue = 1.481
Mammals vs. Vascular Plant pvalue = 1.44
Mammals vs. Amphibian pvalue = .128

What about the Sheep?

Park Rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage points. The only information that the scientists had was that last year it was recorded that 15% of sheep at Bryce National Park had foot and mouth disease. Using this value and the sample size calculator provided, I needed to calculate the number of sheep that they would need to observe from each park to make sure their foot and mouth percentages are significant. I used the default level of significance (90%)

First, I created a bar chart to show the number of sheep sightings at each of the four national parks under investigation (below).



Sheep Cont'd

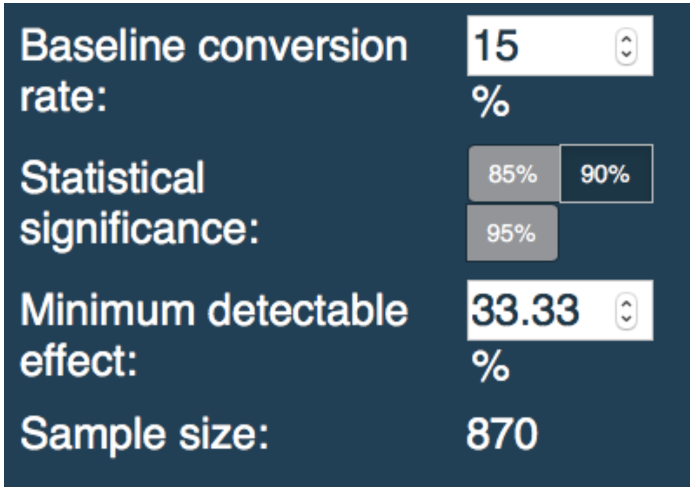
In order to ultimately determine how many weeks the scientists would need to spend at Yellowstone National Park and Bryce National Park, I needed

5 pieces of information:

1. Baseline
2. Minimum Detectable Effect
3. Level of Significance
4. Sample Size per Variant (determined by sample size calculator)
5. # of Observations at each park (data from bar chart)

To the right is a picture of the sample size calculator which determined the sample size necessary for each park (870)

- Baseline = 15%
- Statistical Signif. = 90%
- Minimum detectable effect = 33.33% (100*5/15)



Using the information from the previous slide, we had one last calculation to do to finally answer:

How many weeks would the scientists need to spend at Yellowstone National Park and Bryce National Park to observe enough sheep?

Sample Size (870)
/ # of Yellowstone observations (507)
= 1.71

Sample Size (870)
/ # of Bryce observations (250)
= 3.48

Yellowstone National Park
= ~1-2 weeks

Bryce National Park
= ~3 weeks

Thank you!