Species Status in Our National Parks

Including an examination of the program to reduce foot & mouth disease in sheep species in four parks

The Data

We have two datasets: Species and Observations. There are 5, 541 species in the data we collected. This breakdown includes:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Vascular Plant
- Nonvascular Plant

The second dataset, discussed later on, is Observations of those species in different national parks taken over the course of 7 days.

Determining Conservation Status

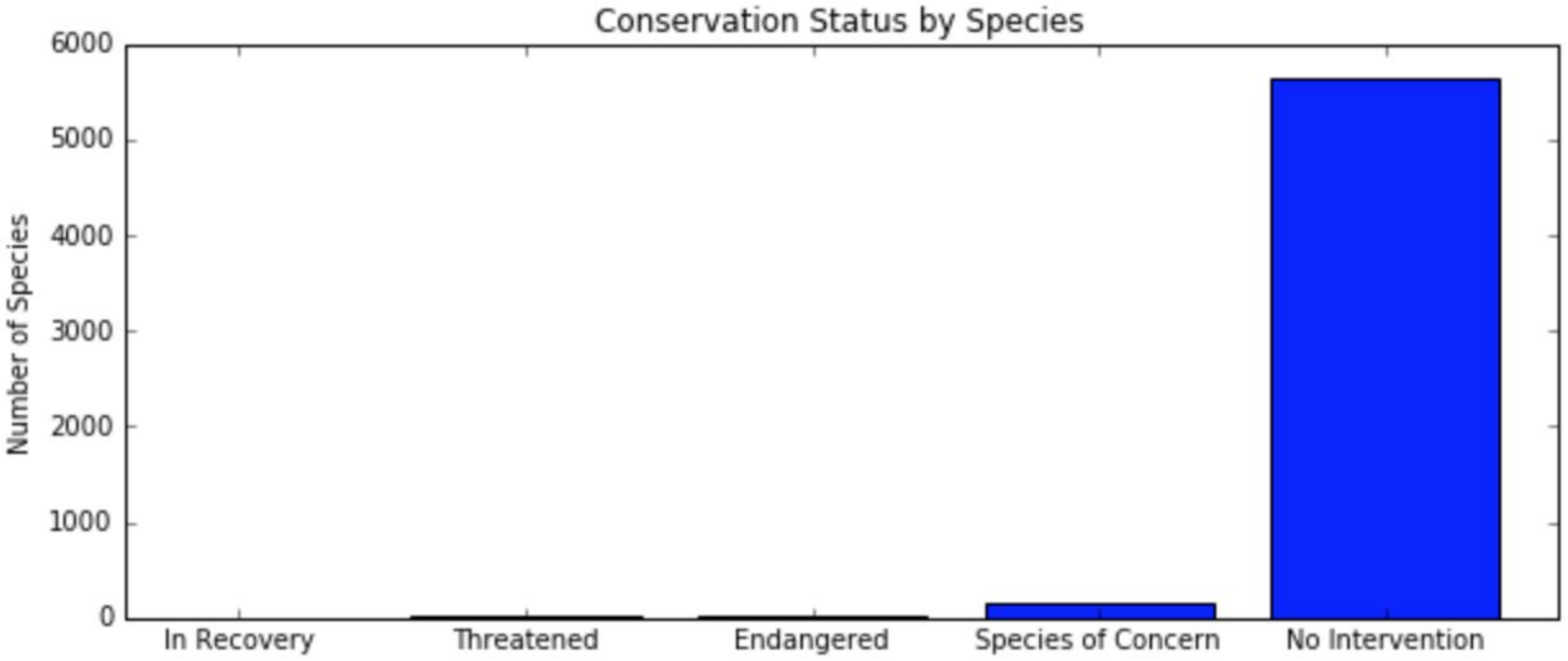
We needed a further breakdown of the statuses and number of species in each. The column conservation_status has several possible values:

- Species of Concern: declining or appear to be in need of conservation
- Threatened: vulnerable to endangerment in the near future
- Endangered: seriously at risk of extinction
- In Recovery: formerly Endangered, but currently neither in danger of extinction throughout all or a significant portion of its range
- No Intervention: no conservation status

So, we grouped the different species by their Status:

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

Conservation Status by Species



Species and Protection Status

What species are more likely to have a protected status? Comparing species at a glance, it looks like **mammals** are more likely to be protected than **birds**. So, let's see if it's a Significant difference….

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

Significance Testing

Because it's categorical data, we ran a Chi-Square test comparing Mammals to Birds and found a pvalue of greater than 0.05 which is not a significant difference.

contingency = [[30, 146], [75, 413]]

chi2, pval, dof, expected = chi2_contingency(contingency)

pval = 0.6875

We then compared Reptile & Mammal, and found the difference to be significant:

contingency = [[5, 73], [30, 146]]

chi2, pval, dof, expected = chi2_contingency(contingency), **pval = 0.0383**

Recommendations Based on Significance

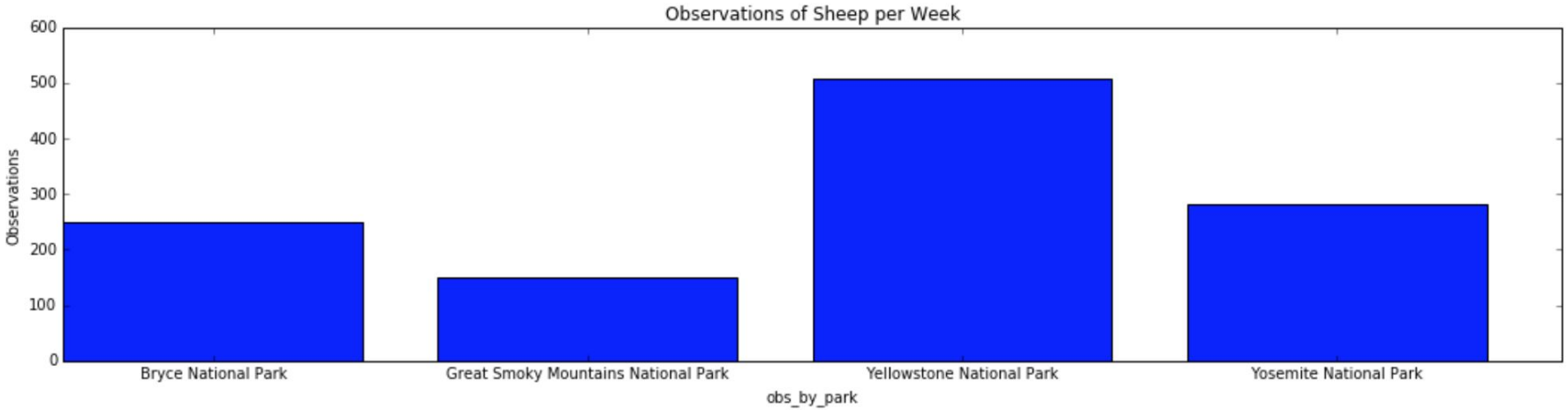
Based on our tests of significance between Mammals/Birds and Mammals/Reptiles, we found that Reptiles are more likely to be endangered. Therefore, efforts should be focused on programs to get them to a lesser, or lower status, with the ultimate goal of “In Recovery” status and, if we’re lucky, a “Not Protected” status.

Observations in Each Park

The second dataset, Observations, is a record of species that have been spotted in parks over the period of 7 days. In it, we singled out the observations of sheep in each of the parks as follows:

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Observations of Sheep Per Week in Each Park



How We Determined A Sample Size For Study

Our scientists know that 15% of the sheep at Bryce National Park have the highly contagious viral foot and mouth disease. 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*. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

We used the sample size calculator at Optimizely to calculate the number of sheep that they would need to observe from each park. So...

How We Determined Sample Size, pt. 2

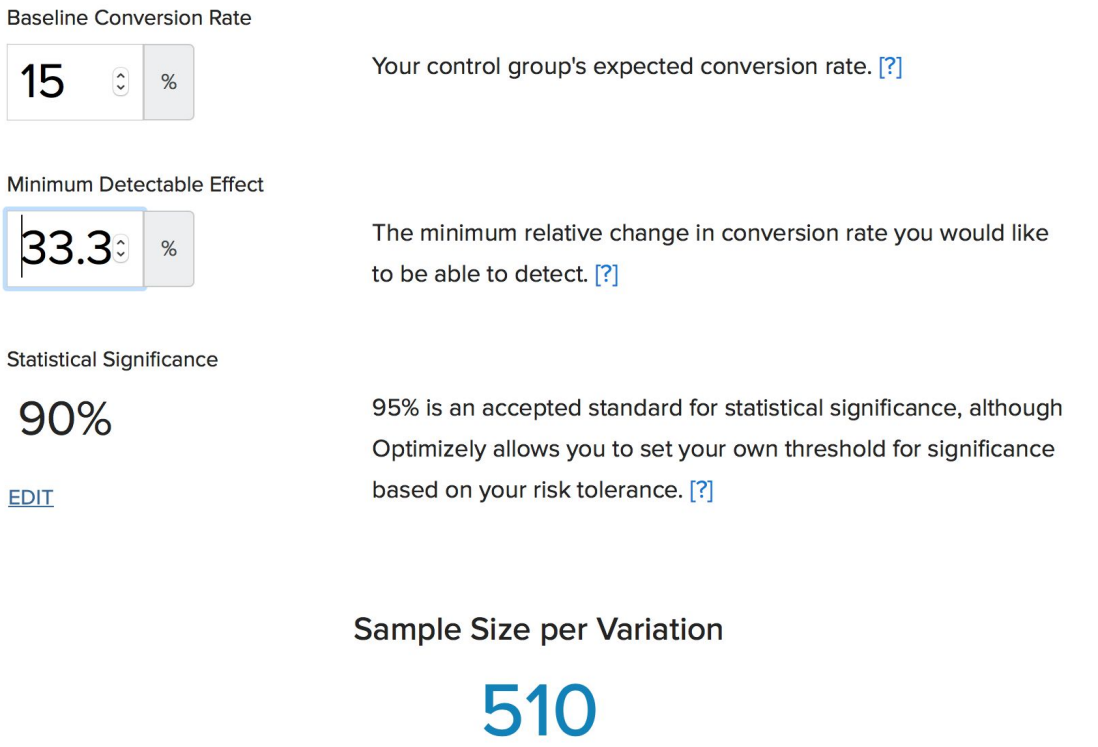
We used this formula to determine the Minimum Detectable Effect (MDE), which is a percentage of our baseline (15%):

minimum_detectable_effect =

100 * 0.05 / 0.15

Which is 33.3333333

So, our sample size is 510



Observations

Based on the sample size of 510, and the number of sheep sightings at each park per week, the number of weeks for observation at each park is as follows:

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282

Bryce = 510/250 , or, approx. 2 weeks

Yellowstone = 510/507 , approx. 1 week

Smokeys = 510/149 approx. 3 weeks

Yosemite = 510/282 , almost 2 weeks, a little over 1.5 weeks.

Let’s get out there and see if there’s a reduction in numbers & see if the program is working!