



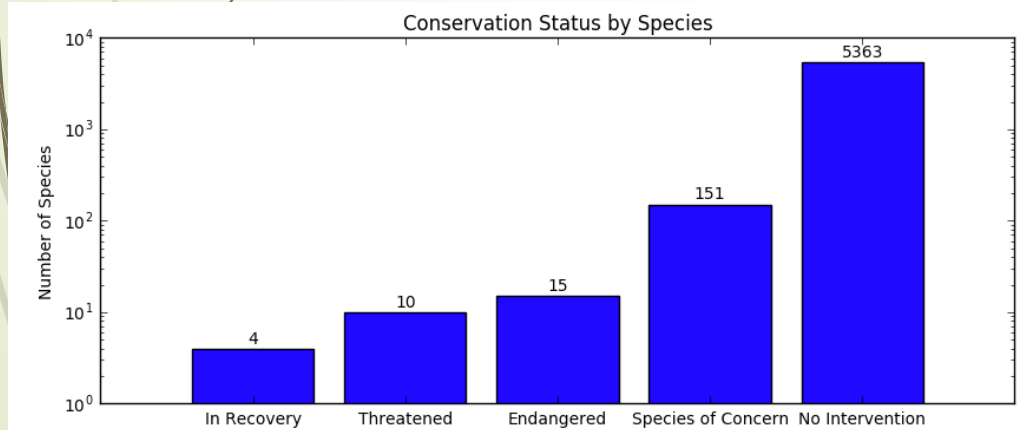
Biodiversity for the National Parks

Annie Frazer

Species Info

category	scientific_name	common_names	conservation_status
Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	No Intervention
Mammal	Bos bison	American Bison, Bison	No Intervention
Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	No Intervention
Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention

- The Species dataset that was provided includes the category, scientific name, common names and conservation status of a number of species found in national parks.
- In total there were 5,543 unique species in the dataset (a number of the records contained duplicate species)
- The chart to the left shows the number of species in each conservation status. The data is plotted using a logarithmic scale, given the large disparity in magnitude between the group with No Intervention and all of the other groups.



Are certain types of Species more likely to be endangered?

category	not_protected	protected	percent_protected
Amphibian	72	7	0.089
Bird	413	75	0.154
Fish	115	11	0.087
Mammal	146	30	0.170
Nonvascular Plant	328	5	0.015
Reptile	73	5	0.064
Vascular Plant	4216	46	0.011

- In order to determine if certain types of species are more likely to be endangered, we started by creating a pivot table that summarized the percent of species in each category that have a protected status.
- We could then compare various categories to determine if their rates of protection status were significantly different.

Hypothesis Tests and Recommendations

	Numerical	Categorical
Sample vs. Known Quantity	1 Sample T-Test	Binomial Test
2 Samples	2 Sample T-Test	Chi Square
More Than 2 Samples	ANOVA and/or Tukey	Chi Square

➤ Because we were dealing with categorical data, we used a chi-square test to determine:

1. If mammals are more likely to be endangered than birds

➤ Pvalue = 0.6875

2. If mammals are more likely to be endangered than reptiles.

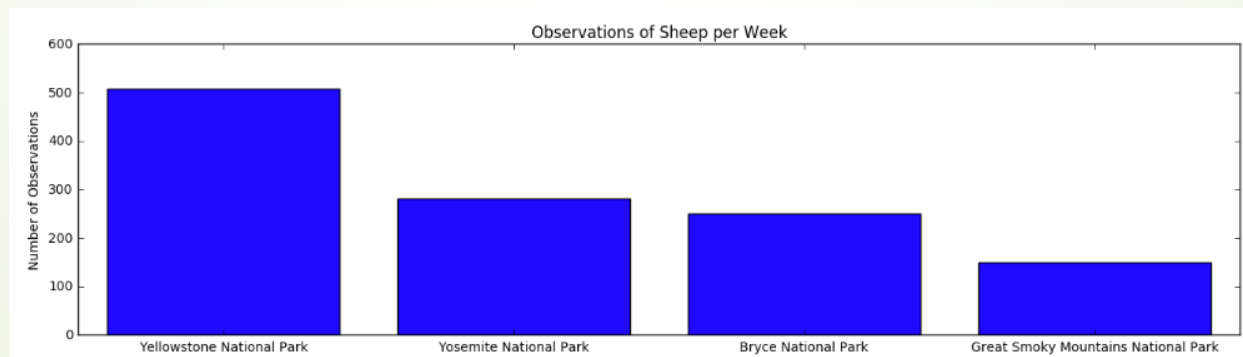
➤ Pvalue = 0.0384

- The results of the first test were not significant at the .05 level, therefore we can not say with certainty that rates of protected status between Mammals and Birds is different. The observed difference could just be mere chance.
- On the contrary the results of the mammal/reptile test were significant at the .05 level. Therefore we reject the null hypotheses; we can be reasonably sure that the protected status rates between the two are different.
- Given these results, conservationists concerned about endangered species should focus their efforts around categories of species such as mammals, over species such as reptiles or plants.

Observations of Sheep

park_name	observations
Yellowstone National Park	507
Yosemite National Park	282
Bryce National Park	250
Great Smoky Mountains National Park	149

- Observations.csv contains data regarding the number of times certain species were observed in 4 national parks over a 1 week period.
- Merging this data with data from the species dataset, we were able to summarize the number of times any species of sheep was observed in each park (chart below)



Sample Size

Foot and Mouth Disease Study

Baseline conversion rate: 15 %
Statistical significance: 85% 90% 95%
Minimum detectable effect: 33.3 %
Sample size: 870

- To determine the number of sheep that need to be observed to detect a reduction of at least 5 percentage points in the rate of foot and mouth disease, we first have to determine the Minimum Detectable Effect.
 - In this case it would be 5% of the 15% known baseline, or 33.33%
- Using the online sample size calculator with a 90% level of significance, the necessary sample size is 870 observations.
- Given the observations per week at the different parks on the previous slide, we would need to observe sheep for approximately 3.5 weeks at Bryce National Park and just under 2 weeks at Yellowstone