Biodiversity Capstone Project

Explaining the data in **species_info.csv**

In the csv dataset, theres:

-5541 different species

-in over 7 different of categories of species;

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

-There's different values of conservation_status:

No Intervention, Species of Concern, Endangered, Threatened, In Recovery

Explaining the data in **species_info.csv**

-There's different values of conservation_status, and a vast majority of the values are at No intervention

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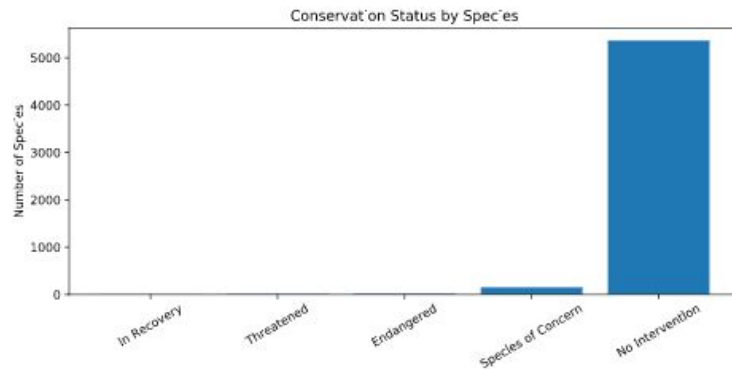
	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	15
3	Species of Concern	151
2	No Intervention	5363

Explaining the data in **species_info.csv**

Are certain types of species more likely to be endangered?

Below is the spread across different species. We can use this data even further on the next slide to answer the above question

	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

Explaining the data in **species_info.csv**

Are certain types of species more likely to be endangered?

Are Mammals more likely to be endangered than Birds?

Is the difference between Mammals and Birds significant?

= No it's not. pvalue:0.687594809666

This means that Mammals are not more likely to be endangered than Birds.

Are Reptiles more likely to be endangered than Mammals?

Is the difference between Reptile and Mammal significant?

=Yes it is pvalue:0.0383555902297 This means Reptiles are more likely to be endangered than Mammals

The methods used to produce these numbers are:

A chi-squared test are performed to find the pvalue, that indicates if a null hypothesis is by chance or significant if it's above or below 0.05
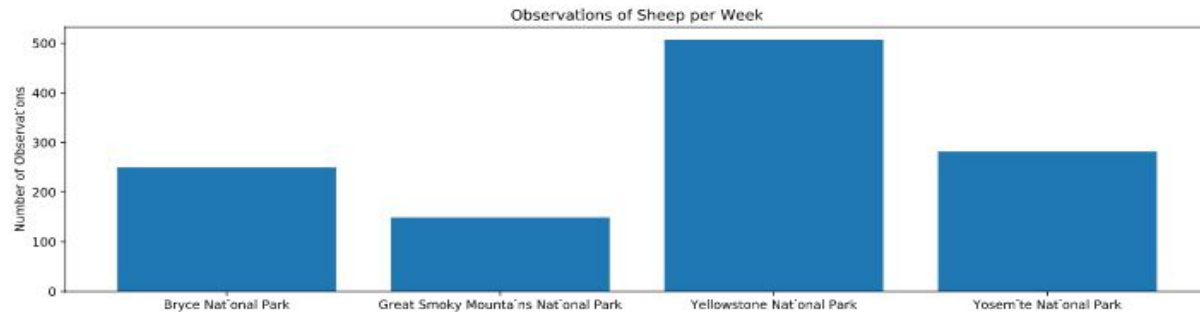
To be able to perform a chi-squared test, we need to create a contingency table with the actual values about if the species are protected or not as mention in table at the previous slide.

Sheep sightings across different parks

	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

Foot and Mouth Reduction Effort

Baseline conversion rate:	15	%
Statistical significance:	85% **90%** 95%	
Minimum detectable effect:	33.33	%
Sample size:	870	

To be able to find and calculate the numbers shown above, we use Sample Size Determination as explained to the right

Baseline conversion rate:

We knew that last year; 15% of sheep at Bryce National Park have foot and mouth disease

90% were used as level of significance.

Minimum detectable effect:

The desired percent point wanted to detect changes. In this case it's 5%

The formula is: x% change / Baseline gives:

5.0 / 15 * 100 = 33.3%

Sample size:

The amount needed to detect statistical significance.