
Biodiversity Capstone Project

Into to Data Analysis

Created By: Megan A. Myers



Conservation Concerns



What Are We Trying to Answer?

- ❖ What species and types of animals species are within the data available to us?
- ❖ Are certain types of species more likely to be endangered?
- ❖ What level of confidence can we have in our answer? Is the answer significant?
- ❖ What can we do with the data to assist with conservation efforts?

Information Contained Within 'species_info.csv'

- ❖ Within this CSV there were several different data items available for review
 - ❖ Species type / category:

1. Mammal	5. Fish
2. Bird	6. Vascular Plant
3. Reptile	7. Nonvascular Plant
4. Amphibian	
 - ❖ Scientific Name
 - ❖ Common Name
 - ❖ Conservation Status

1. Nan/Null	3. Endangered	5. In Recovery
2. Species of Concern	4. Threatened	

Use of Data in CSV

- ❖ While working with this data several things were able to be observed:
 - ❖ Number of species within the data: 5,541
 - ❖ Number of species within each category / type
 - ❖ Number of species that fell within conservation
 - ❖ How many of each species were protected
 - ❖ Which species was more likely to be endangered than others

On the following slides are graphs and charts and discuss the information above

	category	scientific_name	common_names	conserv
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

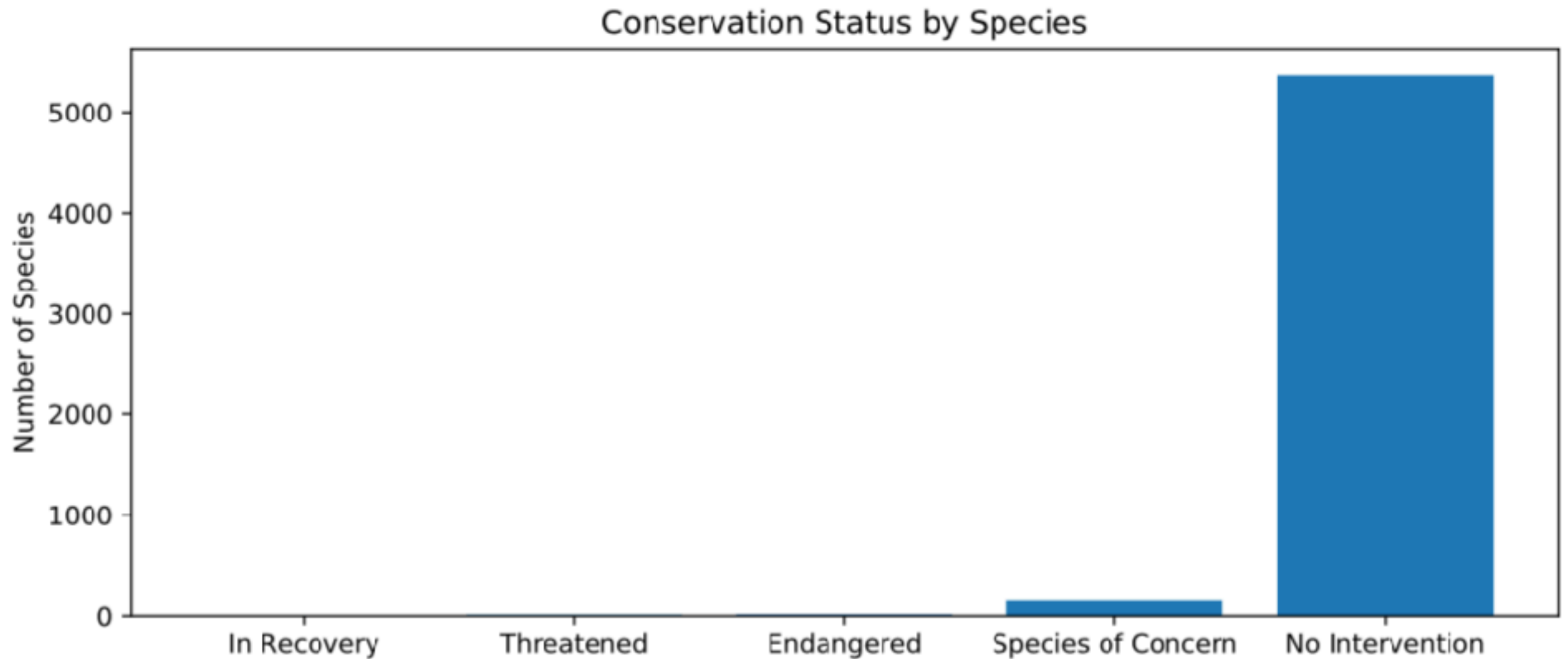
Overview of first lines of chart showing species within species_info.csv.

	Conservation Status	Number
Conservation Counts	Endangered	15
	In Recovery	4
	No Intervention	5,363
	Species of Concern	151
	Threatened	10

There are five (5) categories of conservation. The chart to the left shows the breakdown of the number of species within each conservation category.

Overwhelmingly the species observed require no intervention.

Approximately 180 species within the available data require some sort of intervention.



The above bar graph breaks down the numbers for each conservation category. While it is a visual aid, the chart shows more information due to the large spread of data and the fact that there is a much larger portion of the population that requires no intervention.

	Species / Category Type	Protected?	Number
Category Counts	Amphibian	False	72
	Amphibian	True	7
	Bird	False	413
	Bird	True	75
	Fish	False	115
	Fish	True	11
	Mammal	False	146
	Mammal	True	30
	Nonvascular Plant	False	328
	Nonvascular Plant	True	5
	Reptile	False	73
	Reptile	True	5
	Vascular Plant	False	4,216
	Vascular Plant	True	46

The chart to the left shows a breakdown of the different species / categories and the numbers that are either protected / True (and therefore require intervention) or are not protected / False (and no intervention is required).

This chart is not the most user friendly, so a more user friendly chart is on the next slide.

The chart to the right is a more user friendly version of the chart on the previous slide.

It displays the same data but groups the types together and then states how many of the species within that type are either protected or not protected.

	Species Type	Not Protected	Protected
Category Counts	Amphibian	72	7
	Bird	413	75
	Fish	115	11
	Mammal	146	30
	Nonvascular Plant	328	5
	Reptile	73	5
	Vascular Plant	4,216	46

	Species Type	Not Protected	Protected	Pop. Protected (%)
Category Counts	Amphibian	72	7	8.86%
	Bird	413	75	15.37%
	Fish	115	11	8.73%
	Mammal	146	30	17.05%
	Nonvascular Plant	328	5	1.50%
	Reptile	73	5	6.41%
	Vascular Plant	4,216	46	1.08%

The chart to the left shows the same information as the previous chart but adds in the percentage of the populations of each type of animal that are protected and require intervention.

	Species Type	P Value	Significant?
Results of Chi Square Test	Mammal vs. Bird	~0.68759	No, Chance
	Reptile vs. Mammal	~0.03836	Yes

Above is a chart that shows the results of the Chi Square tests regarding mammals vs. birds and then also reptiles vs. mammals to see if:

1. The chances of being an endangered species was by chance

- ❖ Based off of the samples of the population observed

OR

- ❖ The information available

2. If they are significant

- ❖ The information supports that a type of species is more likely than another to be endangered

AND

- ❖ The results were not just due to random chance based on information available or sample

Observations Based on Data Analysis

- ❖ Are certain types of species more likely to be endangered?
 - ❖ Yes, some types of animals and the species within them are more likely to become endangered.
- ❖ What level of confidence can we have in our answer? Is the answer significant?
 - ❖ With a 90% level of confidence we are able to tell if our findings are significant or more along the lines of chance. This was done by doing Chi Square Testing.
 - ❖ The findings between mammals and birds was determined to not be significant. Mammals do not have a higher chance of being endangered than birds.
 - ❖ The findings between reptiles and mammals was determined to be significant. Mammals have a higher chance or likelihood of becoming endangered than reptiles.
- ❖ What can we do with the data to assist with conservation efforts?
 - ❖ This data shows that mammals and birds are more likely to become endangered than reptiles, amphibians, fish, or plants (either vascular or nonvascular)
 - ❖ Conservationists may need to focus efforts and preventative measures on mammals and birds to find out what is causing those types and their associated species to become endangered at a higher likelihood than

New/Additional Data:
Observations on Animals within
National Parks

Tracking of Sheep

What Questions Are We Trying to Answer?

- ❖ What are the best locations (by park) to find the sheep populations being observed?
- ❖ Where are the different sheep species located (by park)?
- ❖ What sample size of the populations within Yellowstone and Bryant National Parks are needed to determine significant drops within the sheep populations of foot and mouth diseases?

Additional Data (observations.csv)

- ❖ Within this data we were provided with:
 1. Scientific Name
 2. Park Name
 3. Number of Observations
- ❖ Using the additional information from Observations and the previous information from Species it was attempted to track various species of sheep.

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

The above chart is the head view of the sheep within the different National Parks for which we had data.

The information we received included the scientific name of the sheep, the park in which it was observed, and the number of observations of that species within the associated park.

In order to determine which scientific names are associated with sheep information that was previously received within our endangered species data (species_info.csv) was utilized to assist in cross referencing the common names and their associated species. This allowed for the sheep themselves to be isolated from the species and observation data.

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
1139	Vascular Plant	Rumex acetosella	Sheep Sorrel, Sheep Sorrell	No Intervention	False	True
2233	Vascular Plant	Festuca filiformis	Fineleaf Sheep Fescue	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3758	Vascular Plant	Rumex acetosella	Common Sheep Sorrel, Field Sorrel, Red Sorrel, Sheep Sorrel	No Intervention	False	True
3761	Vascular Plant	Rumex paucifolius	Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock	No Intervention	False	True

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

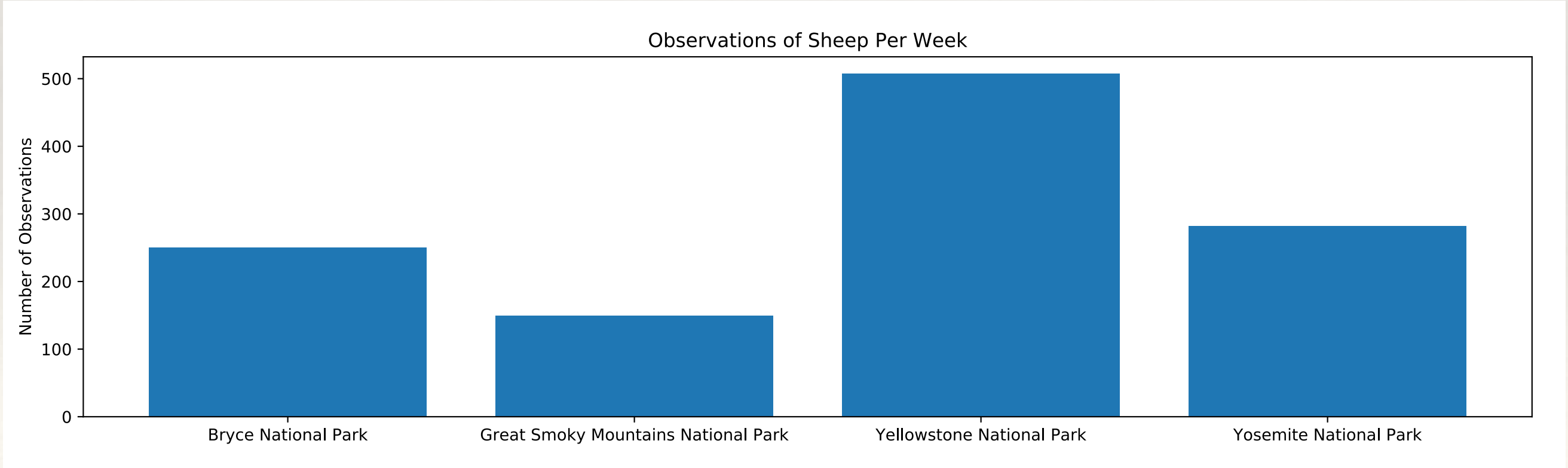
	category	scientific_name	common_names	conservation_status	is_protected	is_sheep	park_name	observations
0	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yosemite National Park	126
1	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Great Smoky Mountains National Park	76
2	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Bryce National Park	119
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True	Yellowstone National Park	221
4	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True	Yellowstone National Park	219

The above charts show the elimination of other non-sheep related data and a snap shot of the animal type/ category, scientific name, common name, conservation status, park observed within, and number of observation in addition to other data that is not particular related to the questions we are trying to answer.

	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

Above is a more streamlined chart than the three on the previous slide that shows the park name and the number of observations of sheep within those parks.

Below is a bar chart that visual shows the number of sheep observations that occurred within each park over a week/seven day period.



How Many Sheep to Seek to Determine Percentage of Population Affected by Mouth and Foot Diseases?

- ❖ Yellowstone is interested in determining how effective their program to reduce mouth and foot diseases within their sheep population has been.
- ❖ Bryant has previous information that shows that 15% of their sheep population is affected by those disease types.
- ❖ Using the previous historical data we calculated the appropriate sample size for both Yellowstone and Bryant (Bryant wanted to repeat the survey) to determine if the program has been successful in reducing disease.

Baseline conversion rate:	<input type="text" value="15"/>	%
Statistical significance:	<input type="button" value="85%"/> <input type="button" value="90%"/> <input type="button" value="95%"/>	
Minimum detectable effect:	<input type="text" value="33.333"/>	%
Sample size:	870	

Using the previous information, the fact that Yellowstone was looking to see if there was at least a 5% reduction in disease, and that they wanted to be 90% confident in their findings from the survey it was determined that a base sample size of 870 would be needed.

How Long Will It Take to Reach the Sample?

- ❖ Based on the number go observations of sheep within the different parks the following was determined:
 - ❖ Yellowstone: Observation will need to occur for a little over one week due to the amount of sheep typically observed.
 - ❖ Bryant: Observation will take a little over two weeks.

What Was Able to Be Done With The Data?

- ❖ Based on the information received regarding species and observations we were able to assist with conservation efforts as well as track and assist in determining the sample size of sheep regarding disease numbers within national parks.
- ❖ Using a combination of analytical tools the data was able to be used and combined to present a fuller picture of the national parks, their conservation efforts, what is required to determine disease prevention success, and other information that will be useful for the parks moving forward.