The Biodiversity Project

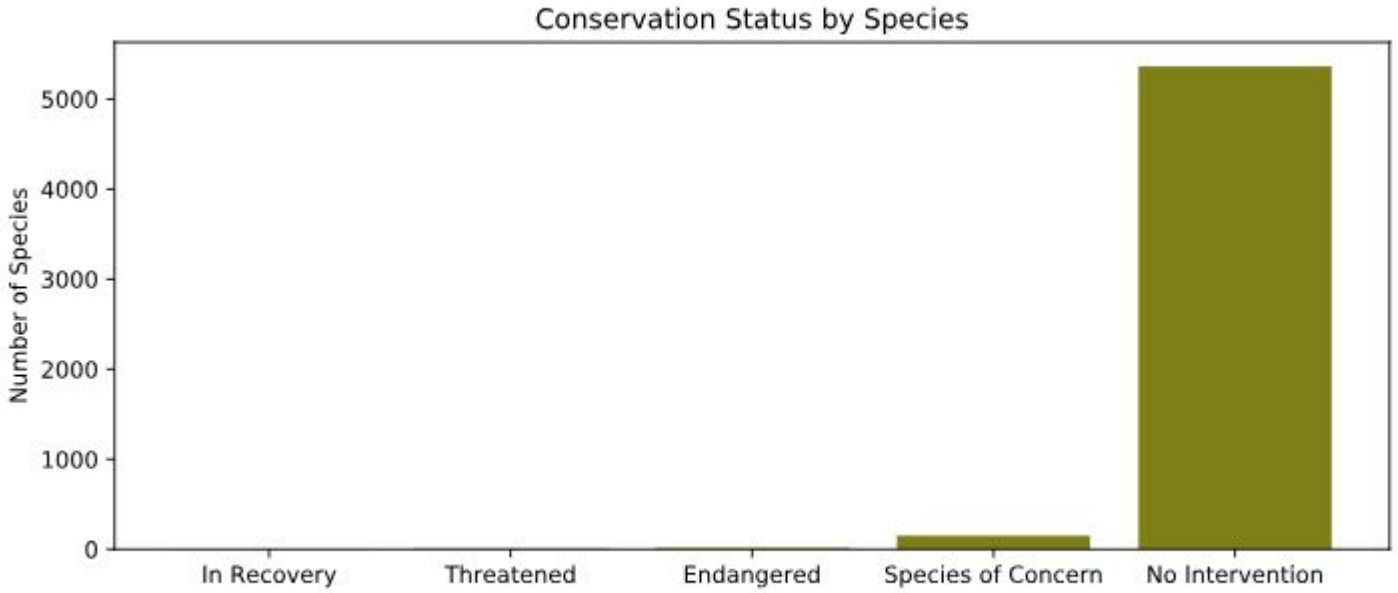
Annabel

About the DataFrame

- 5541 different species
- 7 Different Category Types: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, and Nonvascular Plant
- There are 5 Conservation Statuses: nan (or null), Species of Concern, Endangered, Threatened, and In Recovery
- Using nunique() - I was given the number of distinct scientific_name(s) that fell within each conservation status category. 5363 needed no intervention.

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	Species of Concern	151
3	Threatened	10

- This shows that a lot of species are of concern but also a lot of them are missing which we later changed to no intervention.



No intervention is the largest category, completely outweighing the others.

Pivot

	category	is_protected	scientific_name
0	Amphibian	False	72
1	Amphibian	True	7
2	Bird	False	413
3	Bird	True	75
4	Fish	False	115

is_protected	category	False	True
0	Amphibian	72	7
1	Bird	413	75
2	Fish	115	11
3	Mammal	146	30
4	Nonvascular Plant	328	5
5	Reptile	73	5
6	Vascular Plant	4216	46

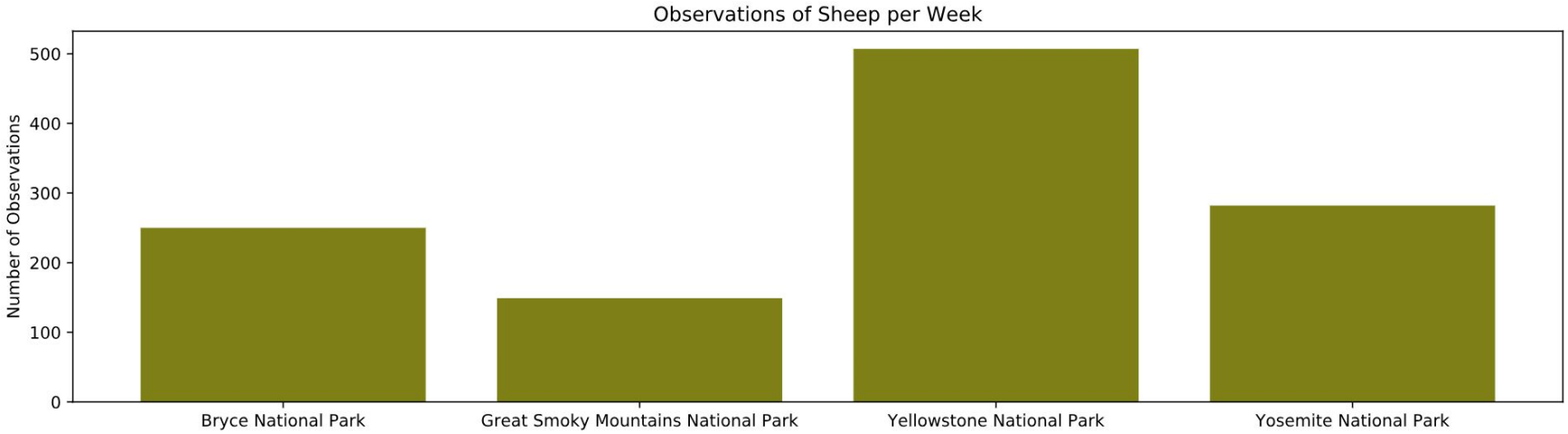
This is the pivot of the data to make it more intuitive to view. It looks like there might be a significant difference between the rate of extinction and certain categories.

Percentage Protected

	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

What does the new `percent_protected` column seem to indicate? That they're not evenly protected and therefore there's likely to be a difference. After running the chi-squared tests we determined that there was a significant difference between mammals and reptiles, but not mammals and birds. Based on this mammals are more cautiously watched because they had a statistical significance with the proportion of the category that is protected.

- A section describing the sample size determination that you did for the foot and mouth disease study



A section describing the sample size determination that you did for the foot and mouth disease study

baseline = 15

minimum_detectable_effect = 100*5/15

sample_size_per_variant = 890

yellowstone_weeks_observing = 1.755

bryce_weeks_observing = 3.56

Based on the numbers given I plugged in the baseline conversion of 15, minimum effect of 33, and statistical significance of 90% to get to 890. Then compared it to the previous observations seen in the data.