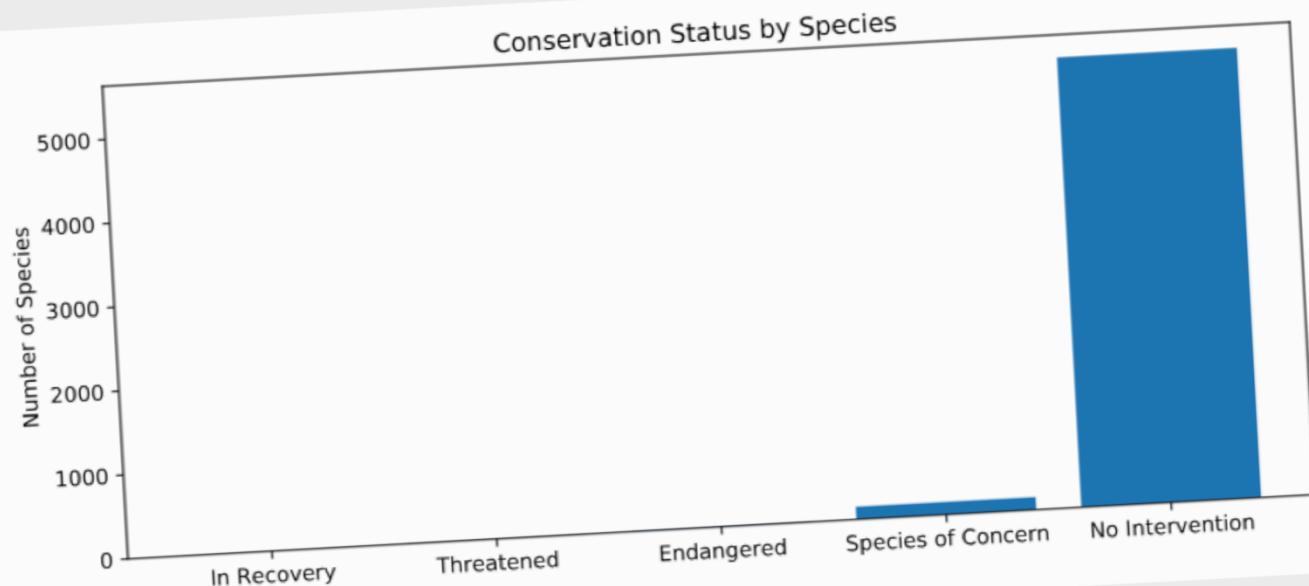


Investigating Protected Species

Josh Casiano

Part I: Species Info

- The CSV file ‘species_info’ contains information about 5541 different species.
- The organisms are categorized into the following seven categories: mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant.
- Organisms which need some protection are listed as species of concern, threatened, endangered, or in recovery.



Most species in the data provided required no intervention

- Although it is good news that most species are doing well, further examination reveals important information to consider.
- When broken down by category, it becomes clear that plants not only make up the vast majority of species represented but are proportionally less at risk than some types of animals.
- This is good news for plants, but other organisms aren't as lucky...

Category	# Not Protected	# Protected	% Protected (rounded)
Amphibian	72	7	8.9%
Bird	413	75	15.4%
Fish	115	11	8.7%
Mammal	146	30	17.0%
Nonvascular Plant	328	5	1.5%
Reptile	73	5	6.4%
Vascular Plant	4216	46	1.1%

- Not including the plants, each category has at least 6.4% of species needing some protectional designation.
- Excluding plants, the average percentage of species with a designated conservation level is 11.28%.
- Birds and mammals have the highest percentage of protected species.
- A further look at the data will help us determine if certain types of species are likely to be endangered.

- A chi-squared test can determine if there is a significant difference between the percentages or if the differences can be attributed to chance.
- The chi-squared test between mammals and birds gives a p-value of 0.687594809666, so these groups do not have a significant difference.
- A chi-squared test between mammals and reptiles, however, gives a p-value of 0.0383555902297.
- This means that there are indeed certain types of species which are more likely to be endangered than others.

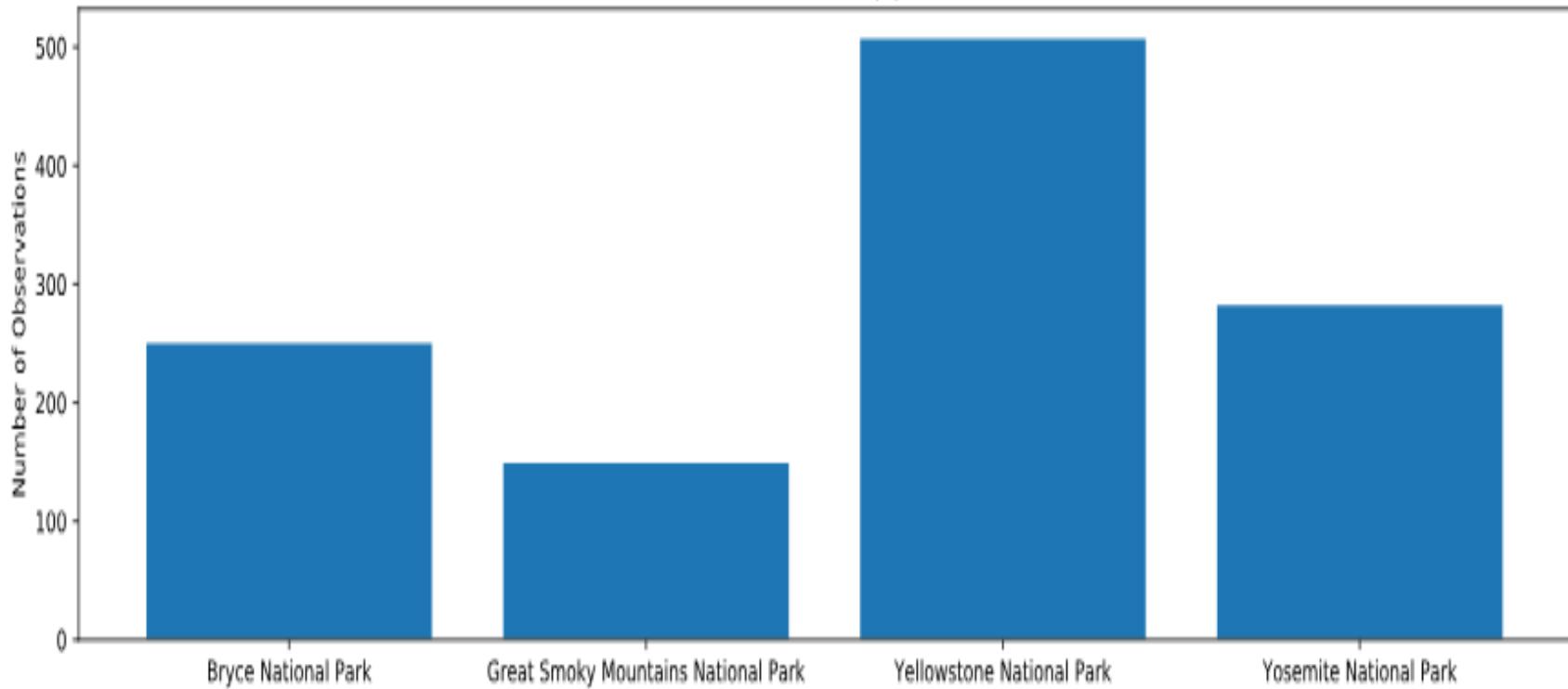
- Knowing that some types of species are more likely to be endangered leads to several questions conservationists may wish to pursue:
- What pressures are acting on species with the highest percentage of protected species?
- What factors, if any, are helping the species which are doing best? Can these be translated to other species?
- Mammals have the highest percentage of protected species. Can the fact that humans are also mammals be used to grow public support for conservation?

Part II: Observations

DataFrame

- The CSV file ‘observations’ contains records about animals sightings at four different national parks.
- The number of sightings each type of animal is reported per park over a seven day period.
- Each entry also notes if the animals holds a protected status.
- Analysis was performed on sheep populations.

Observations of Sheep per Week



- Bars are combined numbers of all sheep species.

- The Park Rangers at Yellowstone have been conducting a program to reduce the instances of foot and mouth disease. To determine an appropriate sample size we can use the following information:
- The baseline conversion rate (from Bryce National Park) is 15%
- The ranges would like to detect a 5% decrease, so we calculate the minimum detectable effect with the following calculation: $100*5/15$
- The standard 90% statistical significance is used here
- This gives a sample size of 870.

- Combining our knowledge of observation numbers and our calculated sample size, we can determine how long it will take to collect data that is statistically relevant.
- Given that over one week 507 sheep were seen in Yellowstone, it would take about 1.7 weeks of observation.
- In Brice, where 250 sheep were seen, it would take 3.48 weeks of observation.
- Best of luck to the rangers with their programs!