Conservation Status Trends Amongst National Parks Species

Including Analysis for Ensuring Viability of Sheep Foot and Mouth Reduction Program

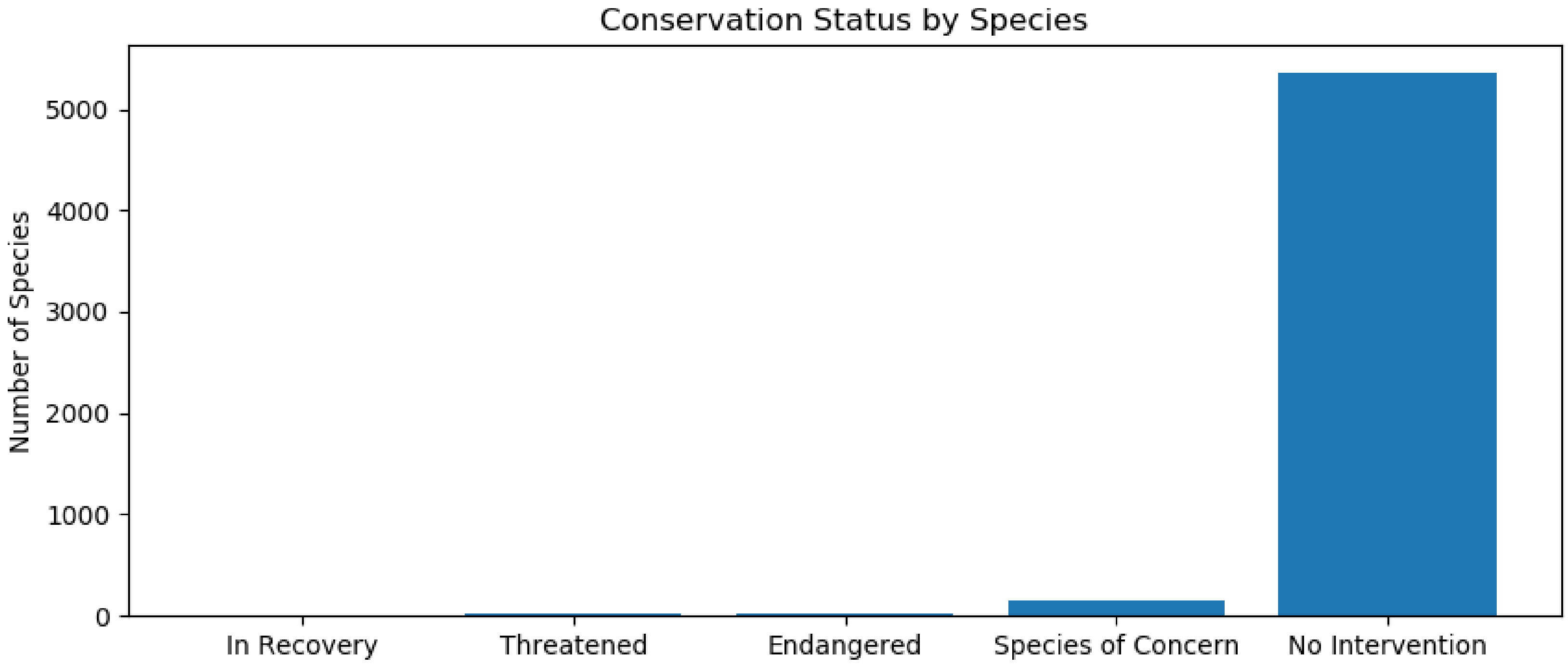
By

Vivek Menon

Observations about Species_Info.csv

- Contains 5541 different unique species which inhabit our national parks.
- Each species is defined by a set of properties. The properties are category, scientific name, common names, and conservation status.
- Category values are mammal, bird, reptile, amphibian, fish, vascular and nonvascular plants.
- The conservation status values are “no intervention”, “species of concern”, “threatened”, “endangered”, and “in recovery”.
- The majority of the species in our park have a conservation status of no intervention.
- We currently have 15 endangered species, 10 threatened species, 4 in recovery, and 151 species of concern.

Conservation Status By Species



Are some species more endangered than others? (Part I)

- Using species data, I tested to see if some species are more likely to be endangered than others.
- First, I added a column to determine if the species was protected (ie, the “conservation_status” value is not “No Intervention”)
- Second, I grouped the data into a pandas dataframe which displayed, for each category, the total number of protected and not protected species.
- Third, I applied a pivot to make the data more readable and added a column for the percentage of each category that is protected.
- From initial observations it looked like some species are more at risk than others, but we needed to determine if this was based on random chance or is the difference significant.

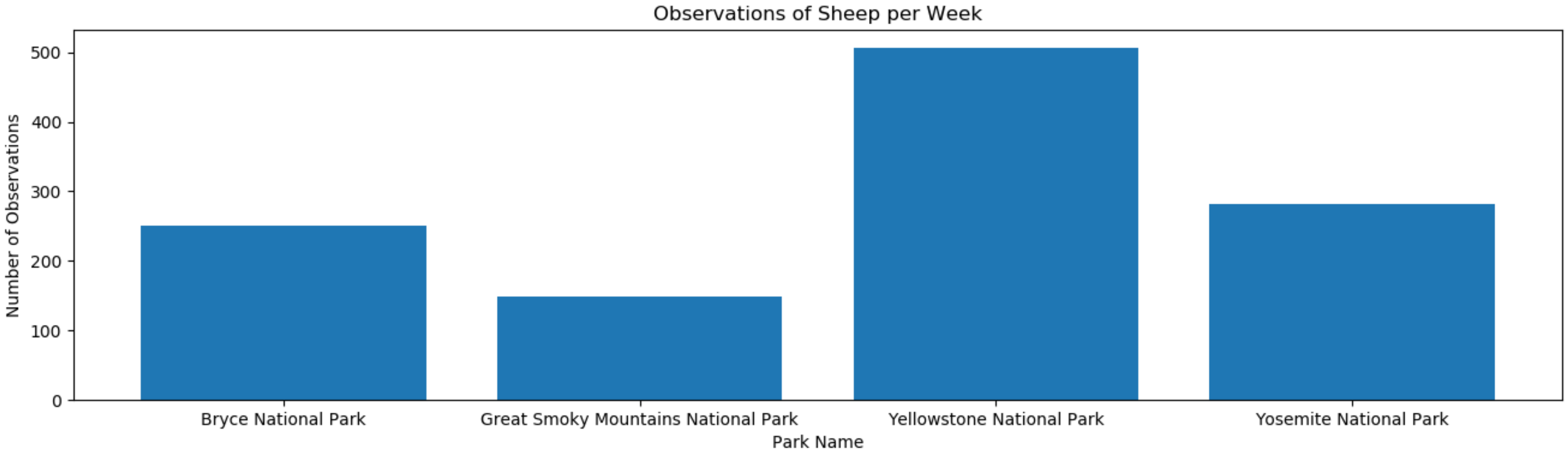
Are some species more endangered than others? (Part II)

- Since our dataframe on species protection counts is categorical with multiple samples, it was best to test the significance of our data with a Chi Squared Test.
- If our test resulted in a pvalue < .05, I could be confident that the difference is significant and reject the null hypothesis (the difference is due to random chance). A pvalue > .05 showed that we could not reject the null hypthoses.
- I first tested if the difference between our bird and mammal populations was significant. I created a contingency table from the “protected”, and “not_protected” values of both groups and passed it to the scipy.stats.chi2_contingency function. The resulting pvalue of .6875 showed we could not reject the hypothesis that, at least for birds and mammals, the difference was due to chance.
- Testing if there was a significant difference between mammals and reptiles did result in a much lower pvalue of .0383, meaning that the risk faced by mammals is significantly different than for reptiles.
- **We can assume, with at least 95% confidence, that birds and mammal species are more likely to be endangered than others. So this detail must be included while designing conservation tactics for species in our national parks.**

Viability of Sheep Foot and Mouth Reduction Program
Part I –Sheep Counts Per Park

- I took the data from our observations dataframe, featuring the weekly park wise sightings of each park species and merged it with our species dataframe.
- I filtered thorough the combined data to return only those that referred to actual sheep.
- I then grouped the data and summed the values to return an approximate number of total sheep in each national park. **See figure on next slide.**
- Knowing the value of sheep per park, once we determine the necessary sample size for our foot and mouth reduction program, we can determine the number of weeks needed to obtain the sample size for each park.

Park-wise Observations of Sheep Per Week



Viability of Sheep Foot and Mouth Reduction Program
Part II – Determining Sample Size

- In order to determine a sample size, I need to determine a baseline conversion rate, a minimum detectable effect, and a level of significance
- Using historical data, my **baseline is 15%**, the number of sheep which had the disease last year
- In order for the program to be successful, we need to detect a reduction of at least 5% (ie can confidently detect if %10 of sheep in Yellowstone are affected). Hence our **minimum detectable effect is 33.33** (using mde = 100.0 * (min detectable reduction as percent / baseline as percent))
- I chose a **statistical significance level of 90%,** meaning that the probability of error is less than 10%.
- Plugging in the above values to our sample size calculator, I determined the **required sample size of 870** for each group to ensure our foot and mouth disease program is viable is.
- Knowing the total number of sheep at each of our national parks, **it will take us about 1.72 weeks of observations at Yellowstone and 3.48 weeks of observations at Bryce to reach our sample size for each park.**