Code Academy Capstone

Intro to Data Analysis
(Nickelous Teixeira)

Species_info

Looking at this data frame, we can see that there are 4 main columns which describe the category, scientific name of a species, the common name of the species, and the conservation status.

While analysing this data, we could see that the majority of species had no conservation status. This presented a skewed bar graph * and resulted in extraneous data not relevant to our analysis. By creating another column and finding only those species that had an identified conservation status, we could narrow down our analysis to relevant species

*Figure 1

Significance & Recommendation

We grouped the data based upon category, and if species within that category were either protected or not. We created a pivot table of this information, by added an addition column to determine the percent of species in a given category that were protected.

We then ran a chi squared test comparing Mammals to Birds and Mammals to Reptiles to determine significance. Significance was only found in comparing Mammals to Reptiles.

--Based on these resulted we can determine that if the focus is on protecting the most species, then the focus should be on Birds and Vascular plants. However if the intent is to protect the largest percentage of species in a given category, the focus should be on Birds & Mammals--

Sheep Foot & Mouth Disease (sample determination)

For this section, after organizing the data, I created a bar graph representing the sheep observations per week in 4 national parks.**

I then did my own minimal detectable effect equation to result in a 33% baseline, which I then used Optimizely to put in the baseline conversion(15%), minimal detectable effect,(33%)and the statistical significance(90%.) This resulted in a sample size of 520 needed for each park.

**See Figure 2

Figure 1

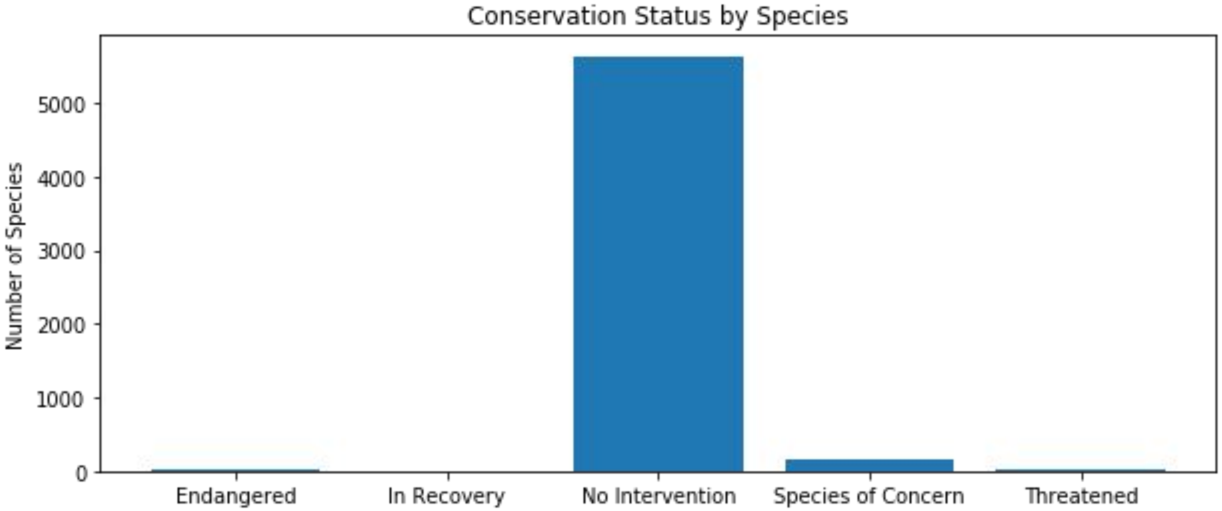
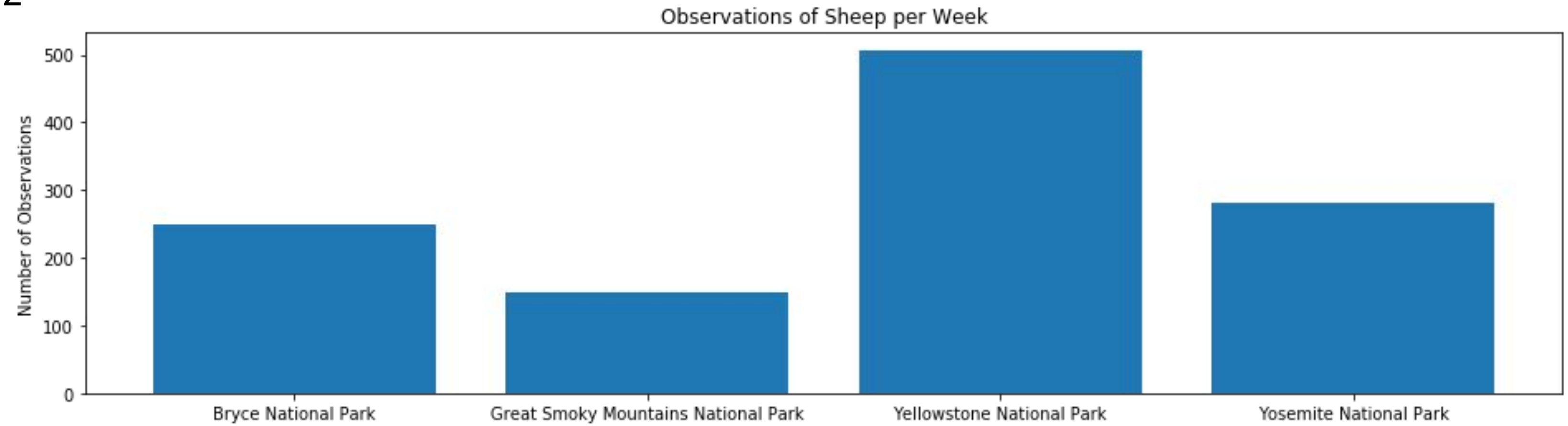


Figure 2


Observations of Sheep per Week
Number of Observations
Bryce National Park
Great Smoky Mountains National Park
Yellowstone National Park
Yosemite National Park