Biodiversity Capstone Project

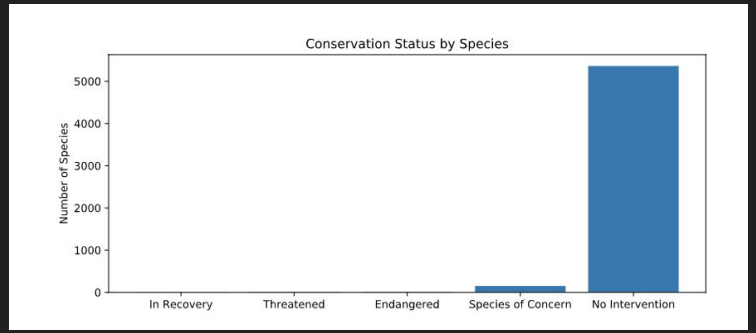
Investigating Protected Species

In a world…

...described by the species_info.csv…

...one man has run the numbers

Our hero has discovered that It's a dire situation for many plants and animals in the nation's greatest parks. There are 151 "species of concern," 10 that are threatened, and 15 that are endangered, while only 4 species are in recovery. (It's not the worst looking when graphed against the species with no intervention, but that's what the lesson called for!)



Which animals are under the greatest threat?

Of the animals, some kinds have it worse than others: 17% percent of mammals are protected, compared to 6% of reptiles. A chi-squared test confirmed our hero's worst suspicions: that this difference is *statistically significant!**

*with a p-value of ~4%

What now?

After exhaustively exploring the data, our hero has made the shocking discovery that mammals and birds are under the greatest threat, but he’s only a humble student of data science. It’s up to conservationists to fight for the rights of poor defenseless endangered animals, but at least now they are armed with the statistically significant information that might show them where to focus their efforts!

But what about the sheep?

There have been efforts to treat foot-and-mouth disease among the sheep in national parks. To determine how long it would take to discover if 5% reductions in cases are statistically significant, our hero took the baseline rate of the disease's prevalence in Bryce National Park (15%) calculated the minimal detectable effect (33%), and sample size at 90% significance (870). From there it was a simple matter of calculating how many weeks it would take, given the rate of sheep observed:

Yellowstone: 1.7 weeks

Bryce: 3.5 weeks.