Written By: Acyuta Raman
Date: 8/5/2024

Biodiversity in National Parks: Analysis Report

Introduction

The purpose of this analysis was to check the conservation statuses of different types of animals, as well as the prevalence of each animal type in various parks. The data used for this analysis was originally from the National Parks Service, but it was directly received from Codecademy.
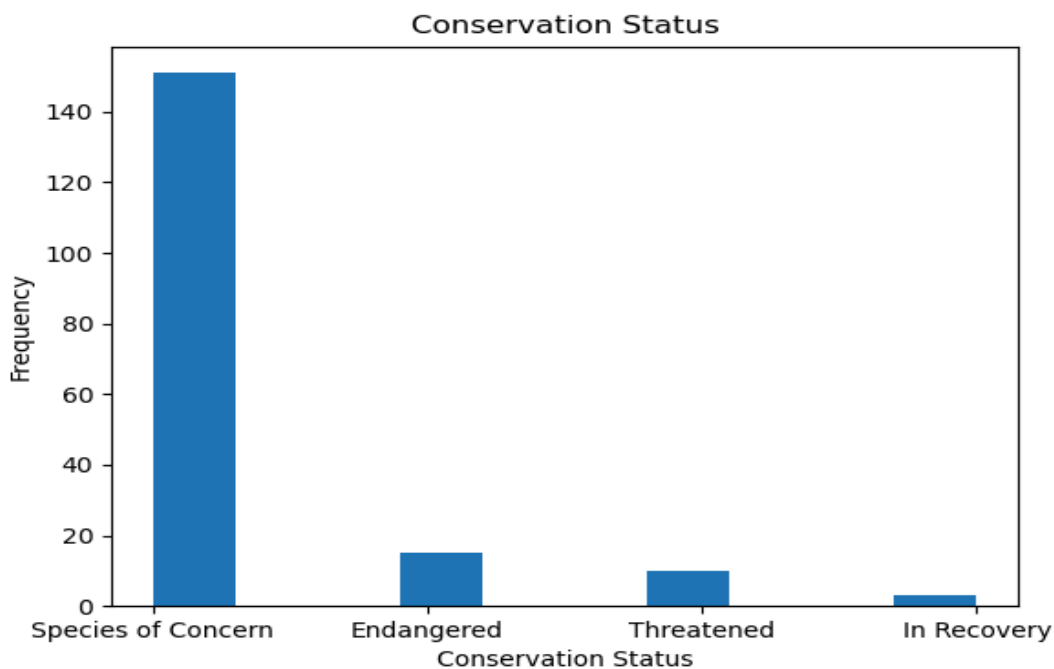
Body

Figure 1: Conservation Status Frequencies

The vast majority of species listed in the dataset did not have a conservation status, meaning their populations are healthy. Additionally, the vast majority of species who had a conservation status were species of concern. Relatively few were endangered, threatened, or in recovery.
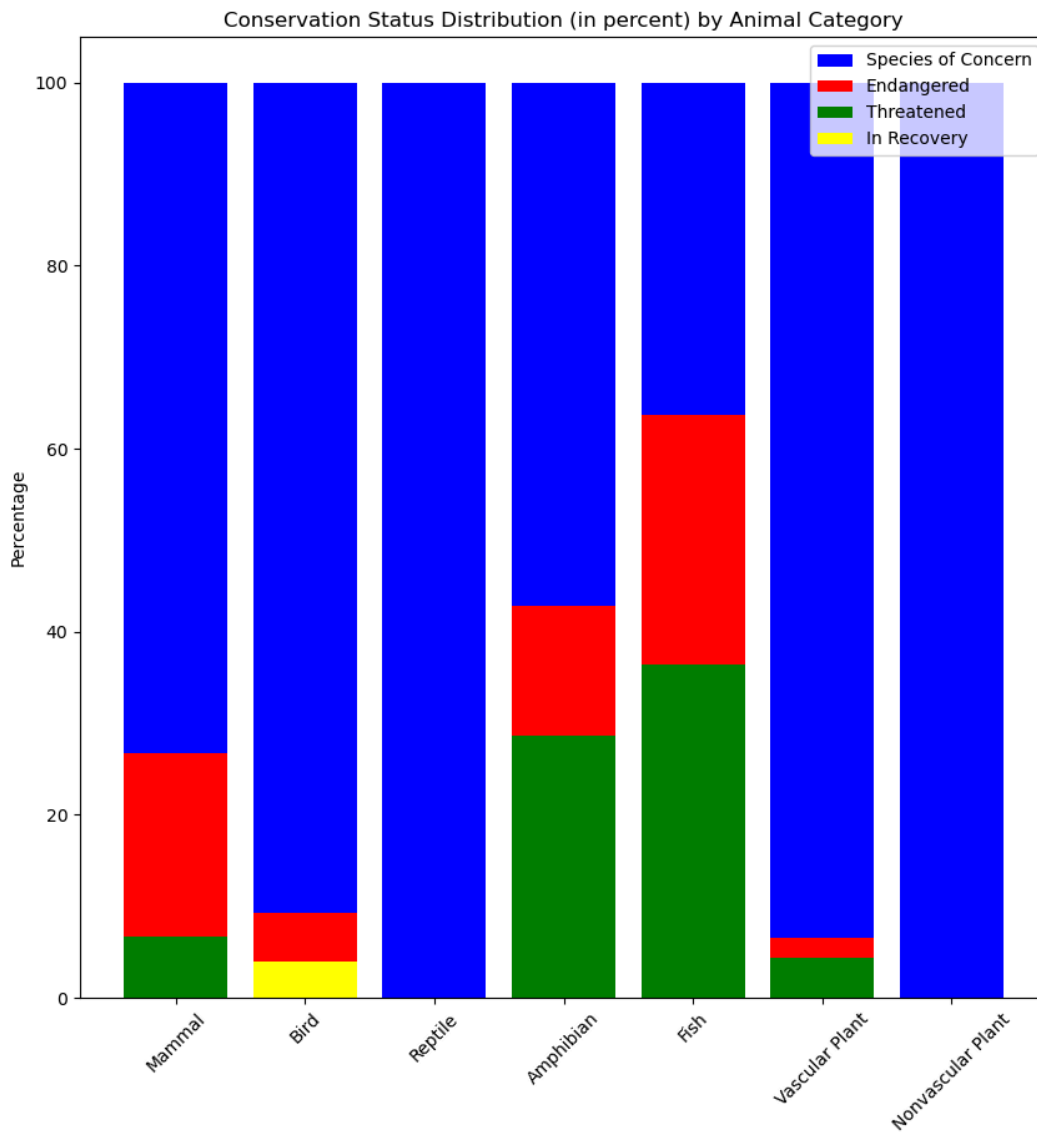
Figure 2: Percentages of Each Animal Category that Belong in Each Status

This stacked bar graph visualizes the proportion of species in each animal category that belong to a defined conservation status. The majority of these species, for most animal categories, are species of concern. The only exception is for fish, which has a larger proportion of species that are endangered or threatened. Note that these proportions do not represent those out of all species of each animal category but rather out of the ones with a defined conservation status.
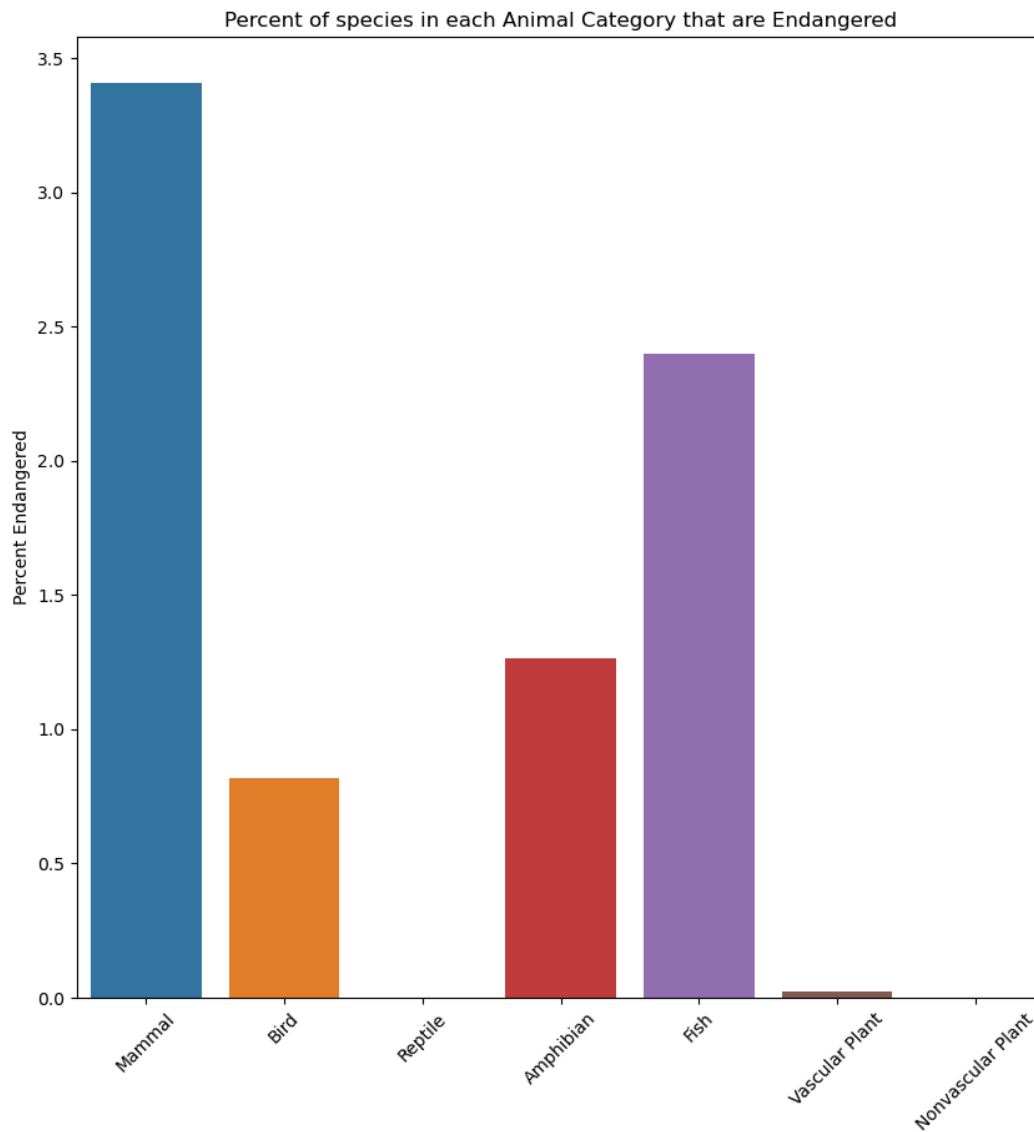
Figure 3: Percent of species in Animal Categories that are endangered.

This bar graph visualizes the percentages of all species in each category that are endangered. The percentage is zero for both Reptiles and Nonvascular Plants. Also, a relatively low percentage (approximately 0.023%) of vascular plant species are endangered. Mammals have the highest percentage of endangered species with a value of approximately 3.41%. Additionally, 2.4% of fish species are endangered.

	Endangered	In Recovery	Safe	Species of Concern	Threatened	Totals
Amphibian	1	0	72	4	2	79
Bird	4	3	413	68	0	488
Fish	3	0	115	4	3	125
Mammal	6	0	146	22	2	176
Nonvascular Plant	0	0	328	5	0	333
Reptile	0	0	73	5	0	78
Vascular Plant	1	0	4216	43	2	4262
Totals	15	3	5363	151	9	5541

Figure 4: Contingency Table of Species Categories and Conservation Status

The vast majority of species in this contingency table fall into the "Safe" category. This category represents species that were not assigned a conservation status in the original dataset. Only two categories of species contained species that were in recovery. The results of a Chi-Squared Test on this contingency table were a Chi-Squared Statistic of approximately 570.019 and a P-value of approximately 4.38 x 10^{-105}. This means that if a species' category and conservation status were independent, there is a probability of 4.38 * 10^{-105} chance that a result at least as extreme as this would occur. Therefore, we can conclude that a species' conservation status is correlated with its category. This P-value is significantly lower than even a standard significance level of 0.05.
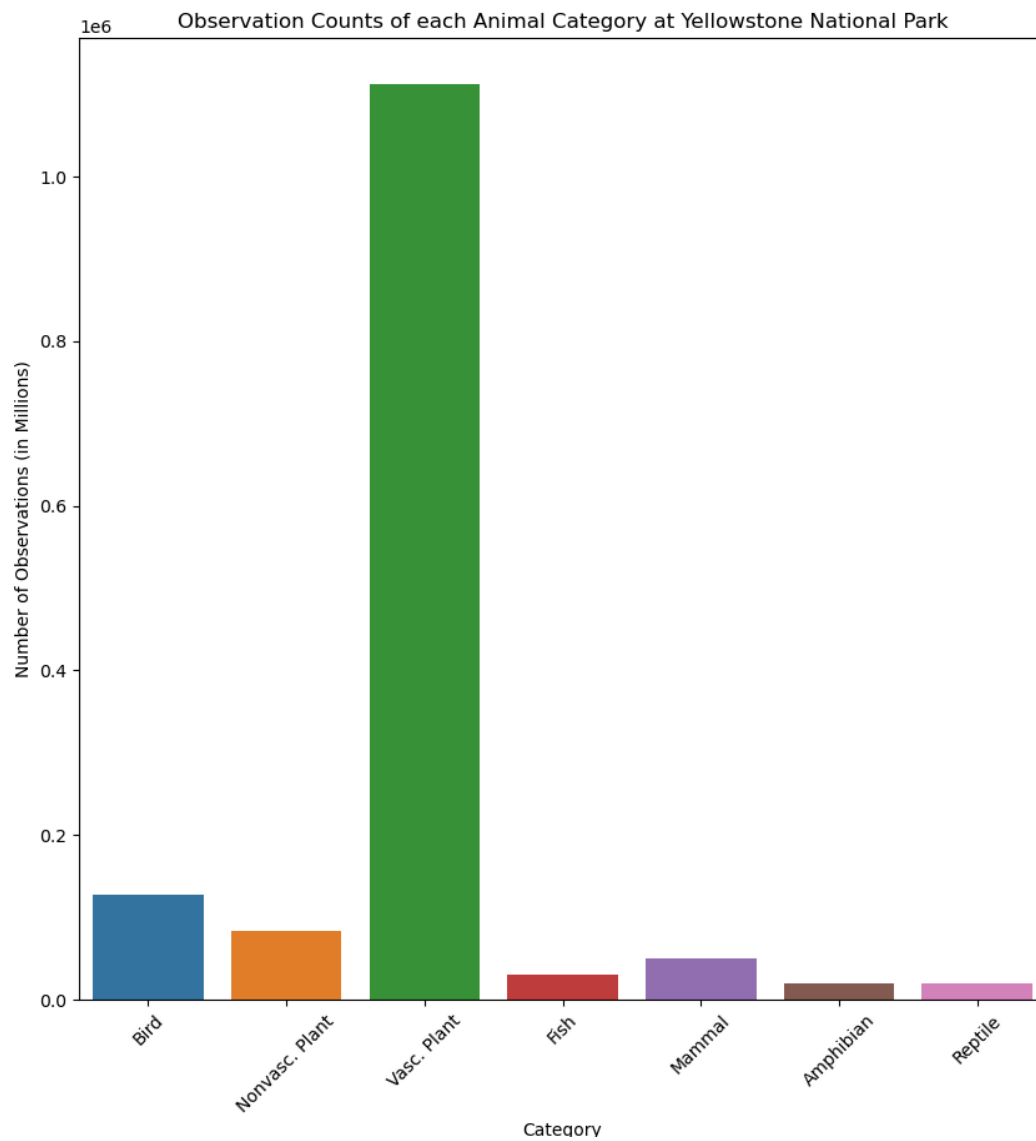
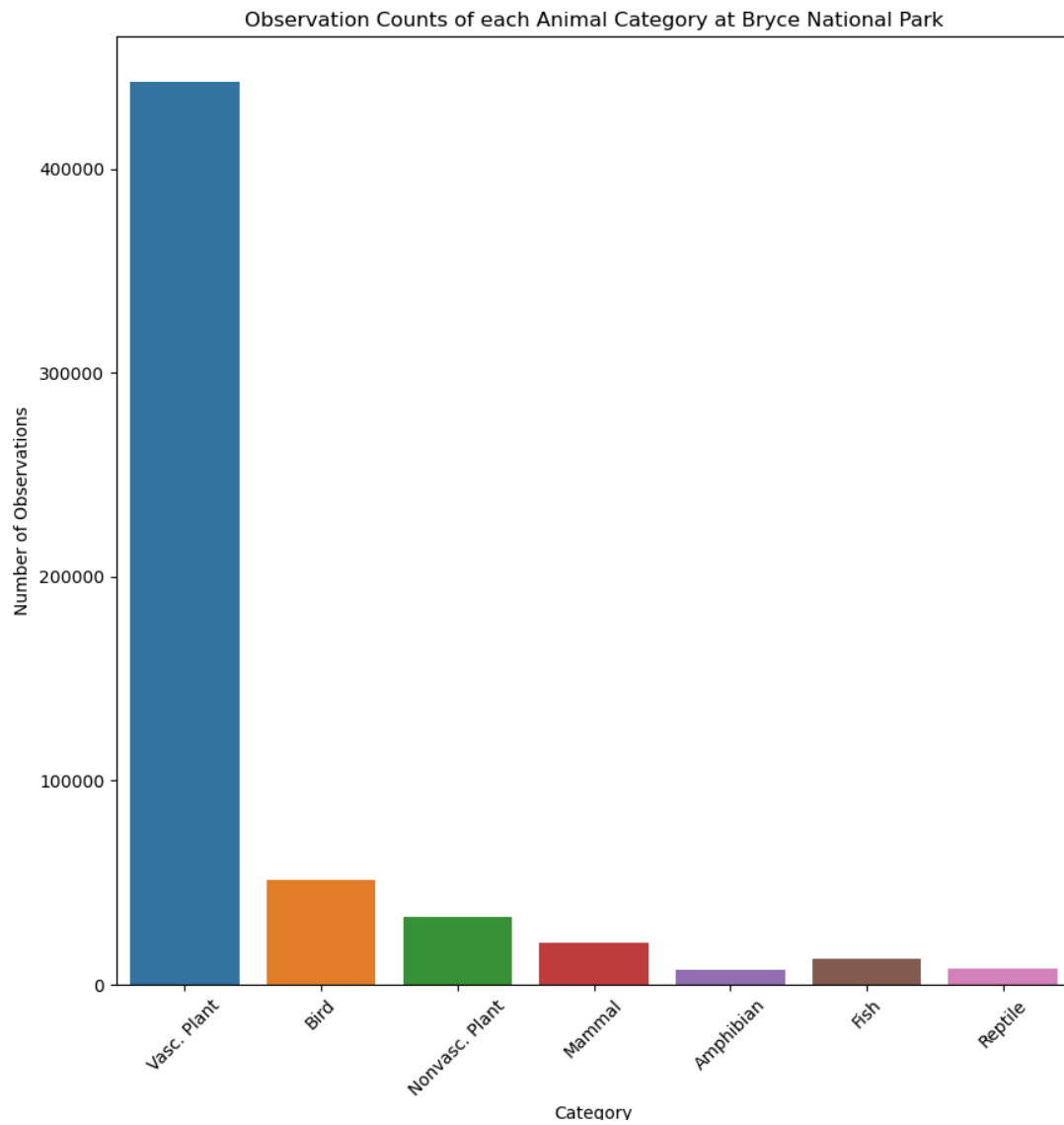*Figure 5: Total Observation Counts, for animals of each category, at Yellowstone National Park*

Figure 6: Total Observation Counts, for animals of each category, at Bryce National Park
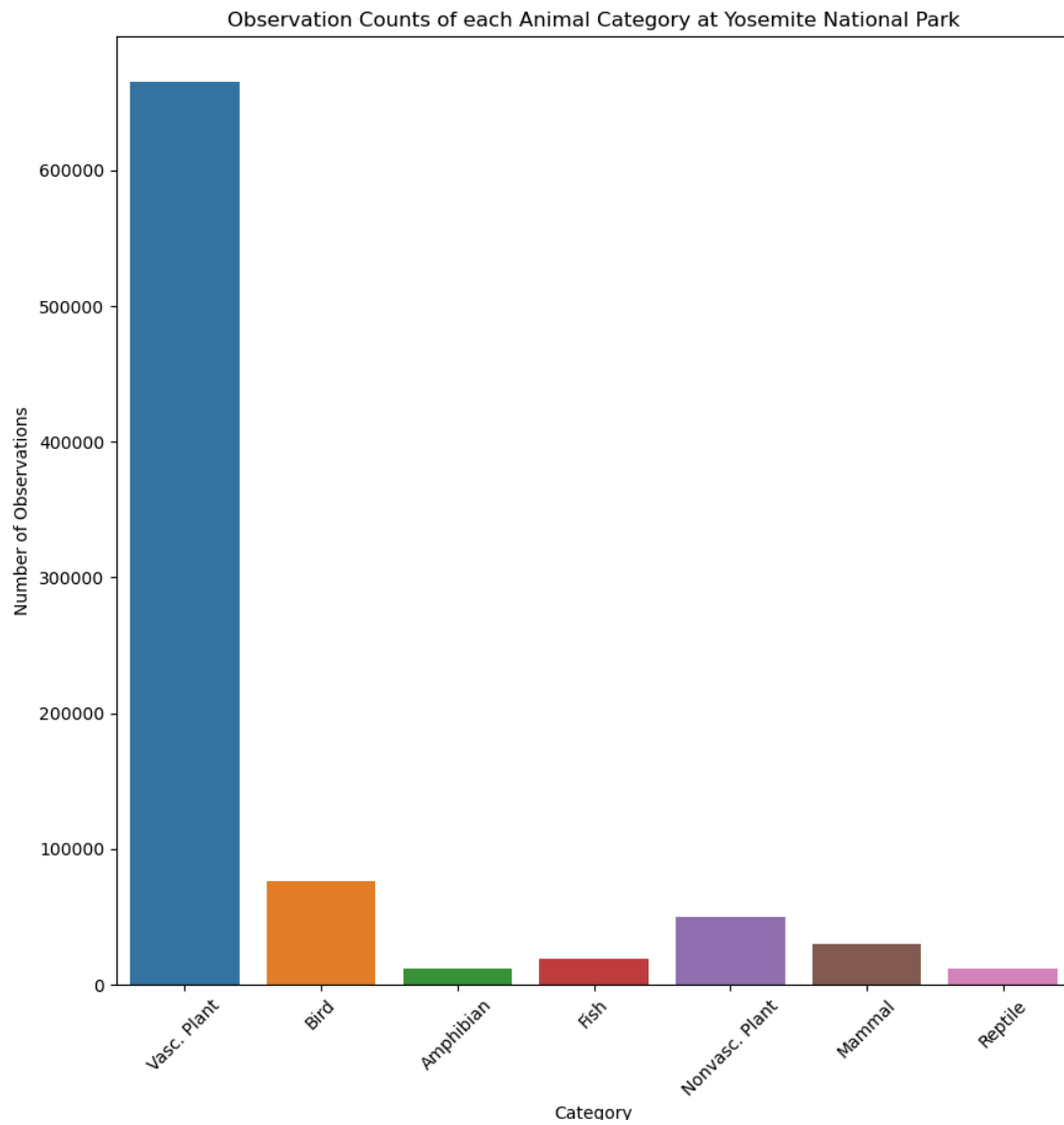
Figure 7: Total Observation Counts, for animals of each category, at Yosemite National Park
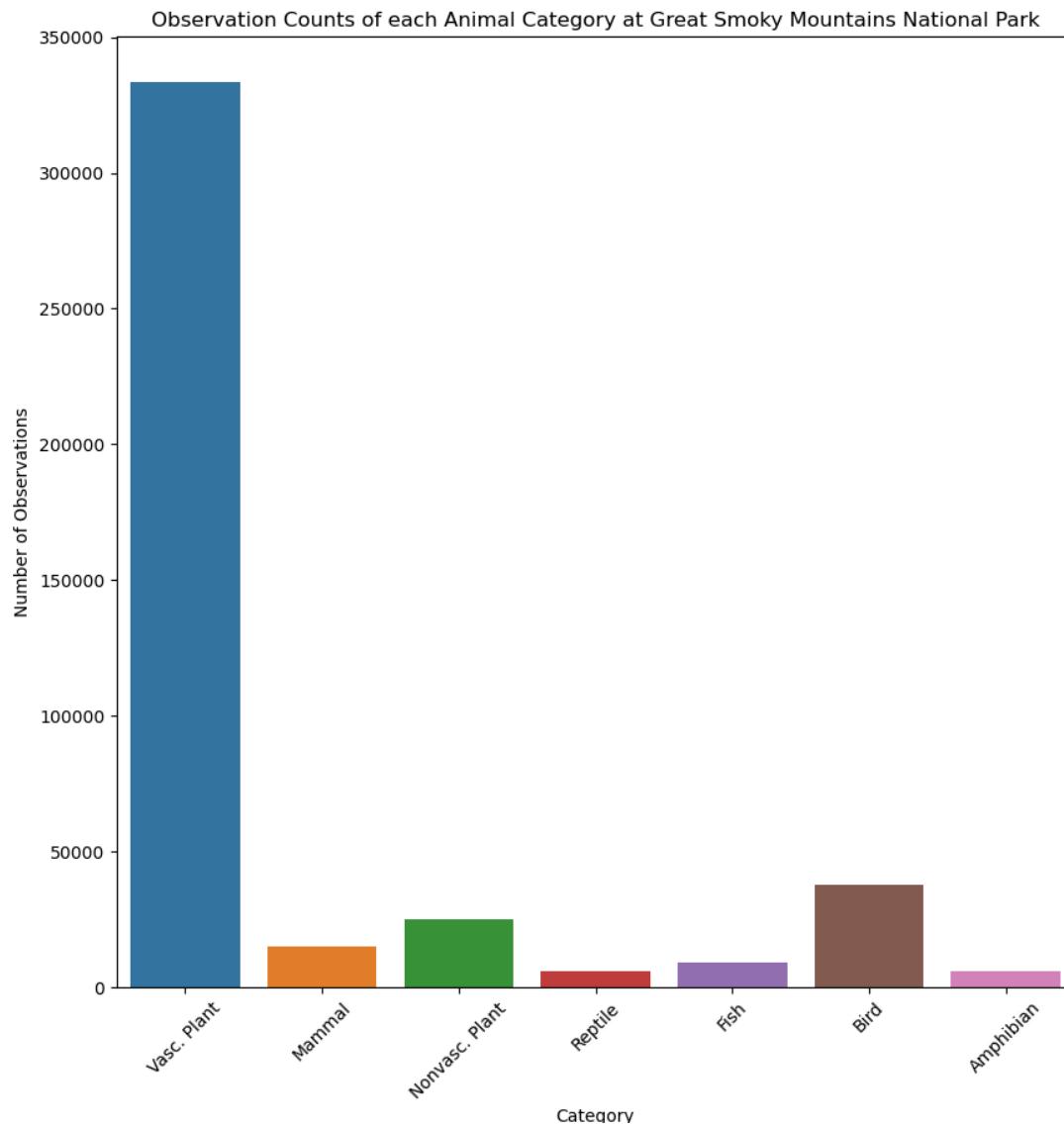
Figure 8: Total Observation Counts, for animals of each category, at Great Smoky Mountains National Park

The vast majority of species present at each national park are vascular plants. There are relatively lesser observations of other species categories. This could be due to the tree-filled environments of these national parks.

Conclusion

There is strong evidence, as shown by the contingency table, chi-squared test, and bar graphs, that species of certain groups are more likely to be endangered. Particularly, mammals and fish have a higher likelihood of being endangered than species of other kinds. This could be due to factors such as global warming, poachers, or hunters. The vast majority of species at each national park were vascular plants. This could be due to the plant-filled environment of these national parks.

Appendix

Chi-Squared-Test: A test between two categorical variables to assess whether or not they are independent. The expected values for each combination of both categorical variables are enumerated in a contingency table where the both categories are independent.

Chi-Squared Statistic: The sum of the squared differences between observed and expected values for each cell in the contingency table divided by the expected values.

Chi-Squared P-Value: The chance of a result at least as extreme as observed if the two categories in question are independent. The probability of receiving the given result by random chance.