BIODIVERSITY IN NATIONAL PARKS

INTRODUCTION

This project involves the analysis of data on a variety of species observed in four different U.S. National Parks. Here are a few question that this project aims to answer:

- *What is the distribution of conservation status for each species?*
- *Which type of species is most prevalent and what is their distribution amongst the four parks?*
- *Are certain types of species more likely to be endangered?*

DATA SOURCE

The data used is provided by Codecademy.com. It is inspired by real data from the National Parks Service, however, it's mostly fictional.

Data Sets Used

species_info.csv with data about different species, their scientific names, and their conservation status.

observations.csv with recorded sightings of different species at four different national parks.

SPECIES DATA

The species data contains information on the species observed in the national parks included.

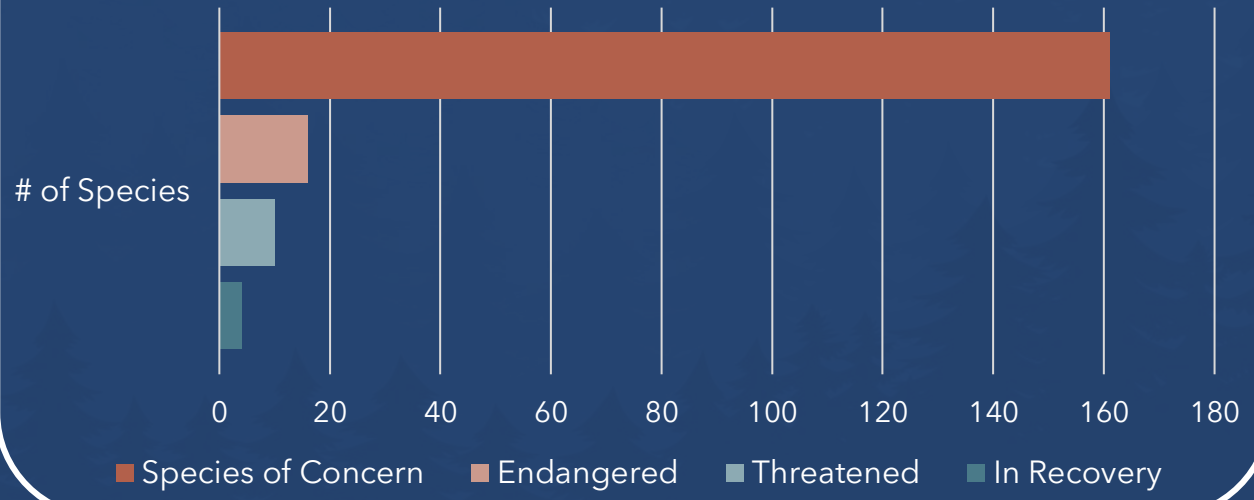
There were four conservation statuses that each species could fall into:

1. Species of Concern: species at low risk
2. Endangered: species facing high risk of extinction
3. Threatened: species likely to become endangered
4. In Recovery: formerly endangered species that have increased in population

~3.28% of species have a conservation status

CONSERVATION STATUS	NUMBER OF SPECIES
No Status/Not at Risk	*5333*
Species of Concern	161
Endangered	16
Threatened	10
In Recovery	4

Number of Species With Each Status



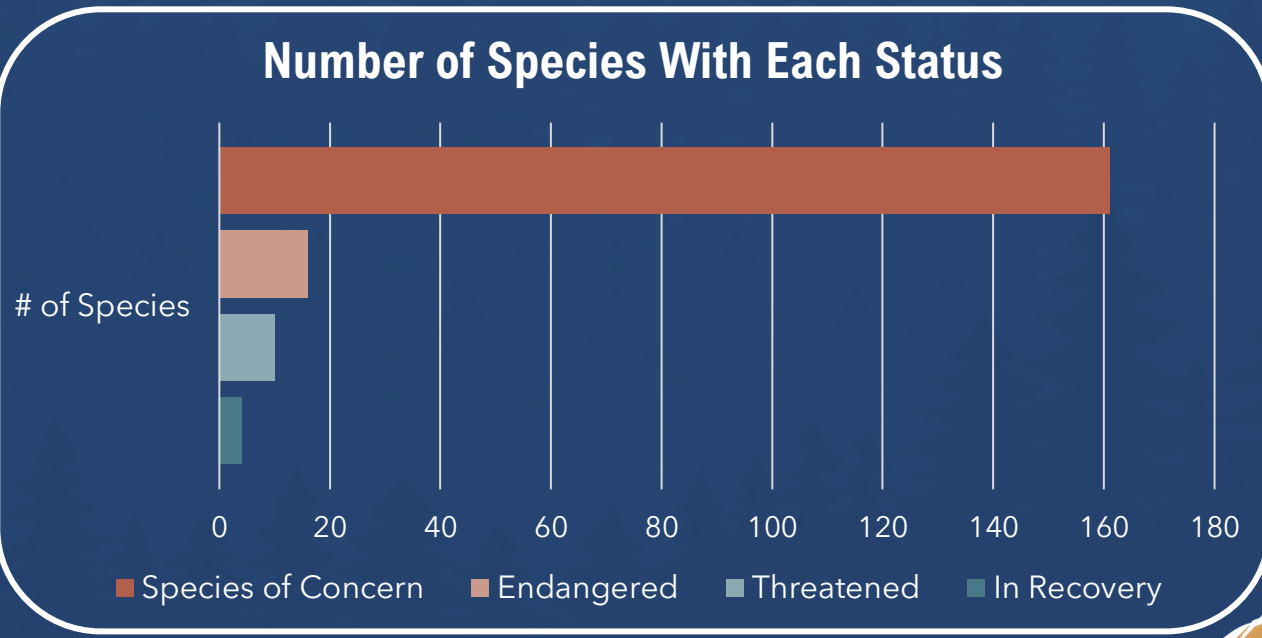
Most of the observed species have no conservation status, meaning they're not at risk.

Only 191 species, about 3.28% of those observed, have a conservation status. Of those:

- 84% are species of concern
- 8% are endangered
- 5% are threatened
- 2% are in recovery

~3.28% of species have a conservation status

CONSERVATION STATUS	NUMBER OF SPECIES
No Status/Not at Risk	*5333*
Species of Concern	161
Endangered	16
Threatened	10
In Recovery	4

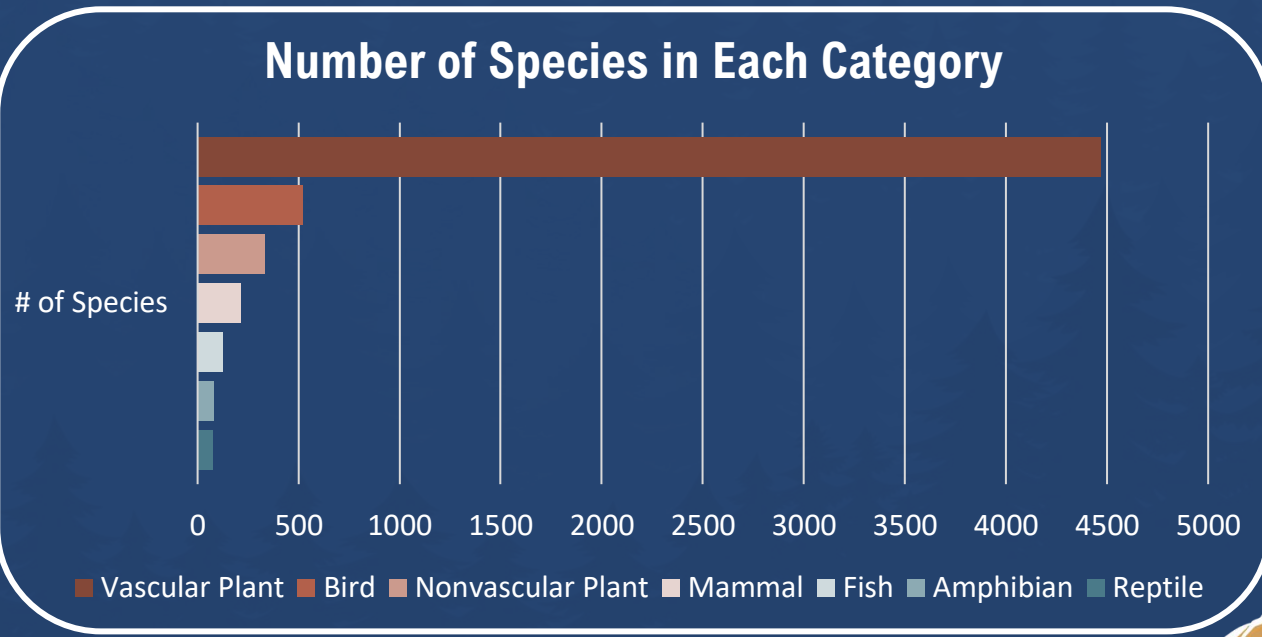


There are seven species categories: vascular plant, bird, nonvascular plant, mammal, fish, amphibian, and reptile.

Vascular plants, by far, have the most listed species with 4470 species – 77% of the species in the species data. The rest are as follows:

- 9% species of bird
- 6% species of nonvascular plant
- 4% species of mammal
- 2% species of fish
- 1% species of amphibian
- 1% species of reptile

SPECIES TYPE	NUMBER OF SPECIES
Vascular Plant	4470
Bird	521
Nonvascular Plant	333
Mammal	214
Fish	127
Amphibian	80
Reptile	79



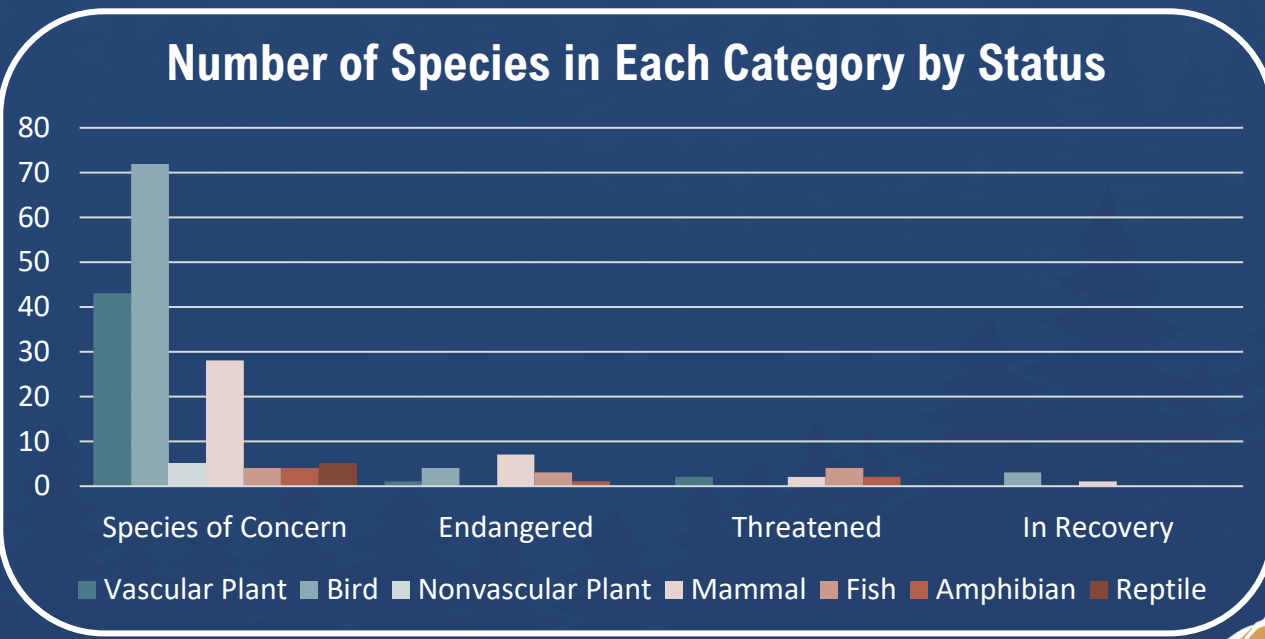
Looking at the distribution of species category in each conservation status, we can see that species of bird are the most common amongst species of concern as well as the species in recovery.

Mammal species are the most common in the endangered status and fish are the most common under the threatened status.

Does this mean that an observed mammal is more likely than other species categories to be endangered?

The Bird Species Category is most common in Species of Concern and In Recovery Statuses.

CONSERVATION STATUS	PREDOMINANT TYPE
Species of Concern	Bird
Endangered	Mammal
Threatened	Fish
In Recovery	Bird

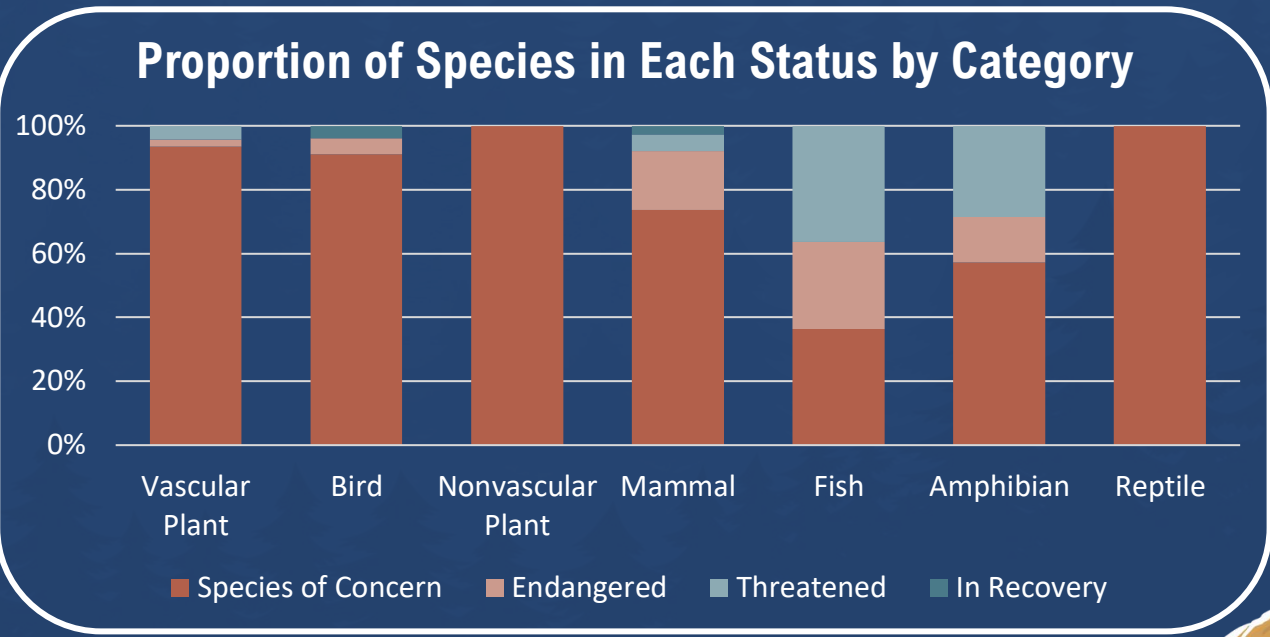


Looking at the distribution this way shows that, even though mammals have the highest number of species with the endangered status, the percent of animal species that are endangered is actually quite low.

Fish, however, have the largest percentage of species within the endangered category. This means that if you were to see a fish in one of the four national parks included in this data, it has a higher chance than any other species type of having an endangered status.

What about rate of protection? Are there certain species more likely to be protected than others (more likely to have a conservation status)?

SPECIES TYPE	PREDOMINANT STATUS
Vascular Plant	Species of Concern
Bird	Species of Concern
Nonvascular Plant	Species of Concern
Mammal	Species of Concern
Fish	Species of Concern/Threatened
Amphibian	Species of Concern
Reptile	Species of Concern

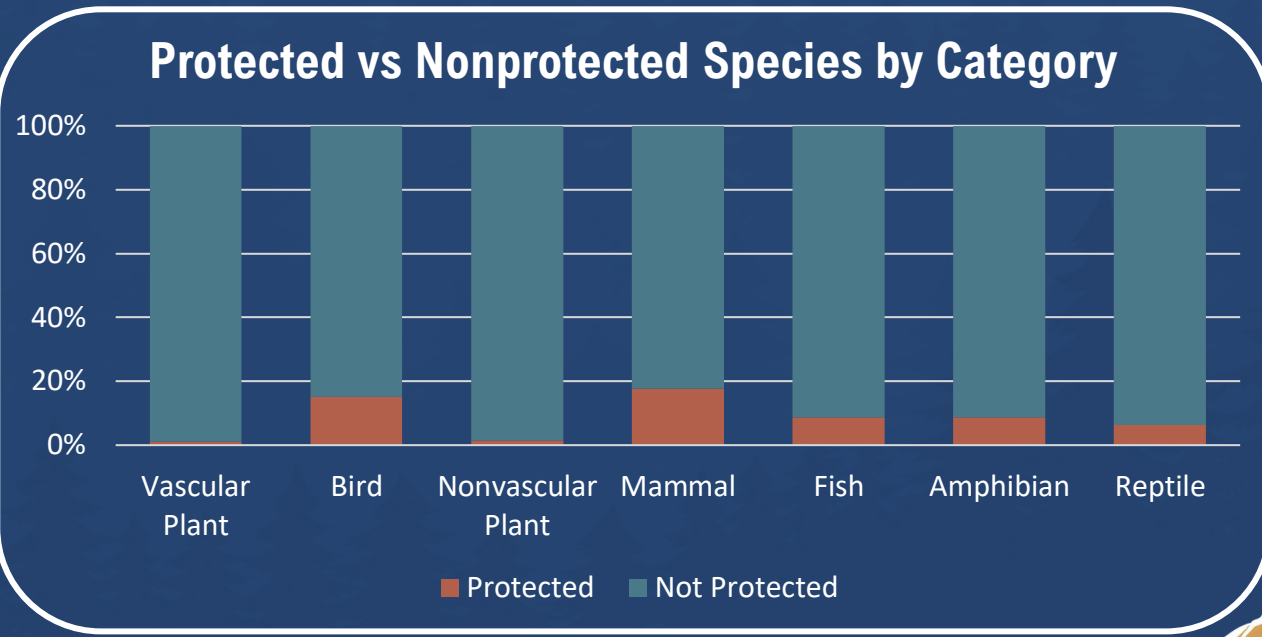


In terms of numbers, Bird and Vascular Plant are the categories with the most protected species. However, looking at proportions, we can see that Mammal and Bird are the categories with the highest percentage of protected species.

Mammal is the top category for proportion of protected species with about 17.8% of all observed mammal species being protected. The Bird category is close behind with about 15.2% of all observed bird species being protected.

The Vascular Plant category is the lowest with only about 1% of all observed Vascular Plant species being protected species.

CATEGORY	PROTECTED	NOT PROTECTED
Vascular Plant	46	4424
Bird	79	442
Nonvascular Plant	5	328
Mammal	38	176
Fish	11	116
Amphibian	7	73
Reptile	5	74



Is there any statistical significance to these differences?

Running a Chi-Square test to find whether there is any significance to the difference between each species' likelihood to be protected gives us a P-value of $3.1 * 10^{-98}$. This means that the difference between all types of species and their rates of protection is significant.

There are four other Chi-Square test results to the right. These tests tell us that the difference between rate of protection for the Mammal category and Bird category is not significantly different. However, the difference between rate of protection for the Bird category and Vascular Plant category *is* significantly different.

CATEGORY	PROTECTED	NOT PROTECTED
Mammal	38	176
Bird	79	442

P-value: 0.45 | NOT Significant

CATEGORY	PROTECTED	NOT PROTECTED
Bird	79	442
Vascular Plant	46	4424

P-value: $9 * 10^{-84}$ | Significant

CATEGORY	PROTECTED	NOT PROTECTED
Bird	39	442
Amphibian	7	73

P-value: 0.18 | NOT Significant

CATEGORY	PROTECTED	NOT PROTECTED
Bird	79	442
Fish	11	116

P-value: 0.08 | NOT Significant

The last two Chi-Square tests look at the difference between the rate of protection for the Bird category and the Amphibian and Fish categories.

The Amphibian and Fish categories were tied for third in rate of protection at about 8.7%.

The result of the third Chi-Square test, based off the P-value, tells us that the difference between the rate of protection for the Bird category and the Amphibian category is *not* significantly different.

The final Chi-Square test's P-value shows us that the difference between the rate of protection for the Bird category and the Fish category is also *not* significantly different.

CATEGORY	PROTECTED	NOT PROTECTED
Mammal	38	176
Bird	79	442

P-value: 0.45 | NOT Significant

CATEGORY	PROTECTED	NOT PROTECTED
Bird	79	442
Vascular Plant	46	4424

P-value: $9 * 10^{-84}$ | Significant

CATEGORY	PROTECTED	NOT PROTECTED
Bird	39	442
Amphibian	7	73

P-value: 0.18 | NOT Significant

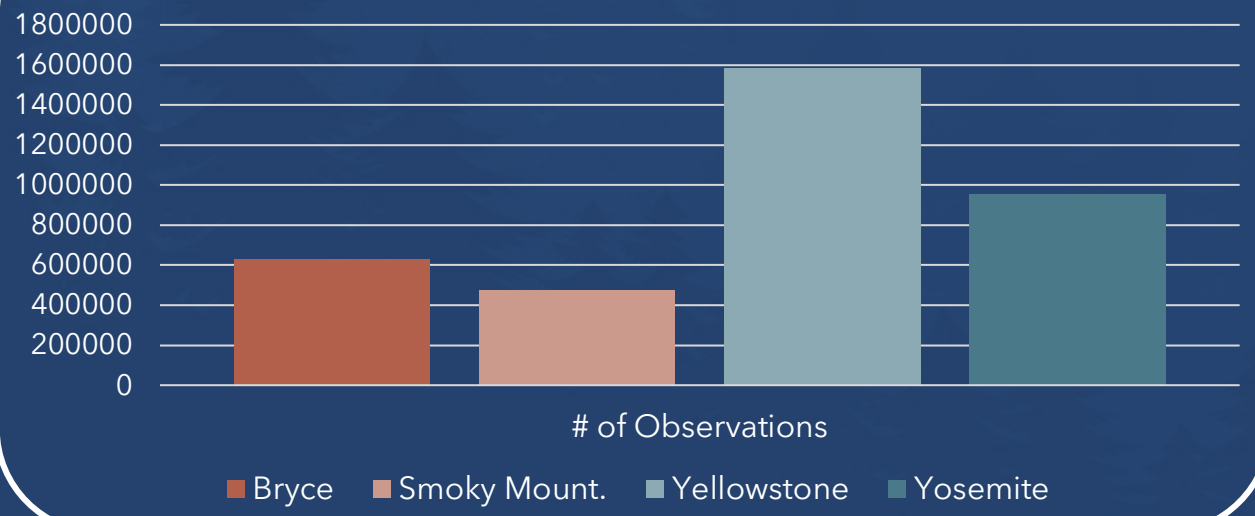
CATEGORY	PROTECTED	NOT PROTECTED
Bird	79	442
Fish	11	116

P-value: 0.08 | NOT Significant

Observations made at Yellowstone National Park make up 44% of recorded observations.

NATIONAL PARK	NUMBER OF OBSERVATIONS
Bryce National Park	633,043
Great Smoky Mountains	473,979
Yellowstone National Park	1,584,890
Yosemite National Park	948,460

Number of Observations in Each Park

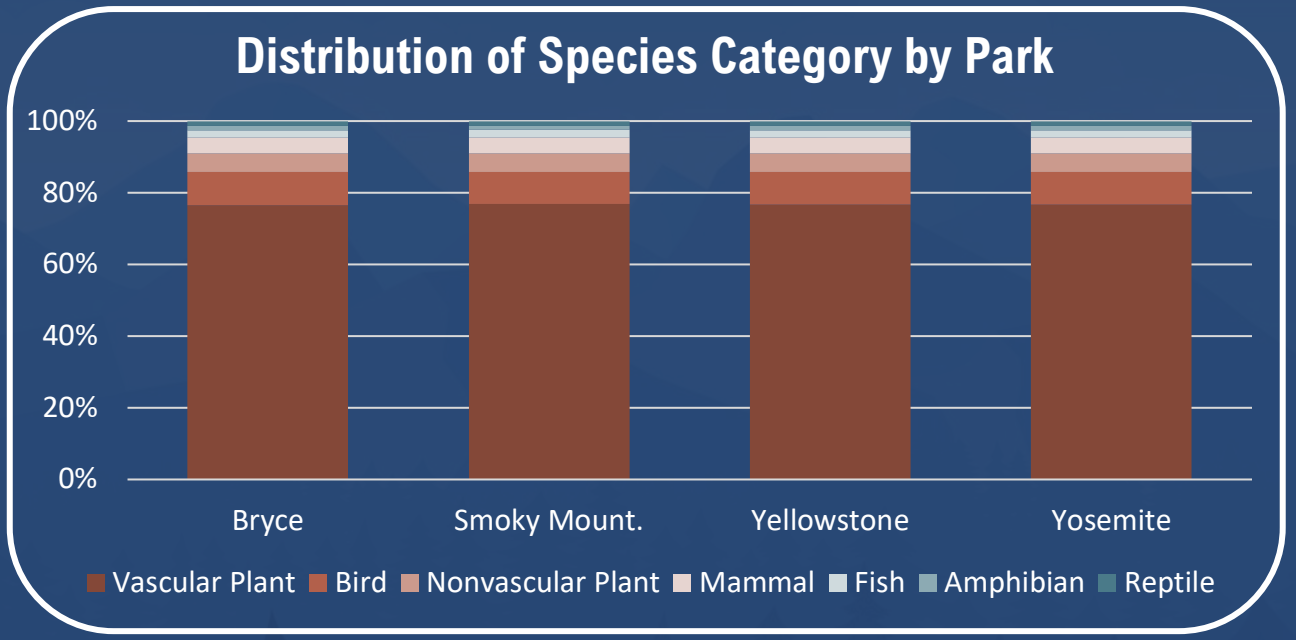


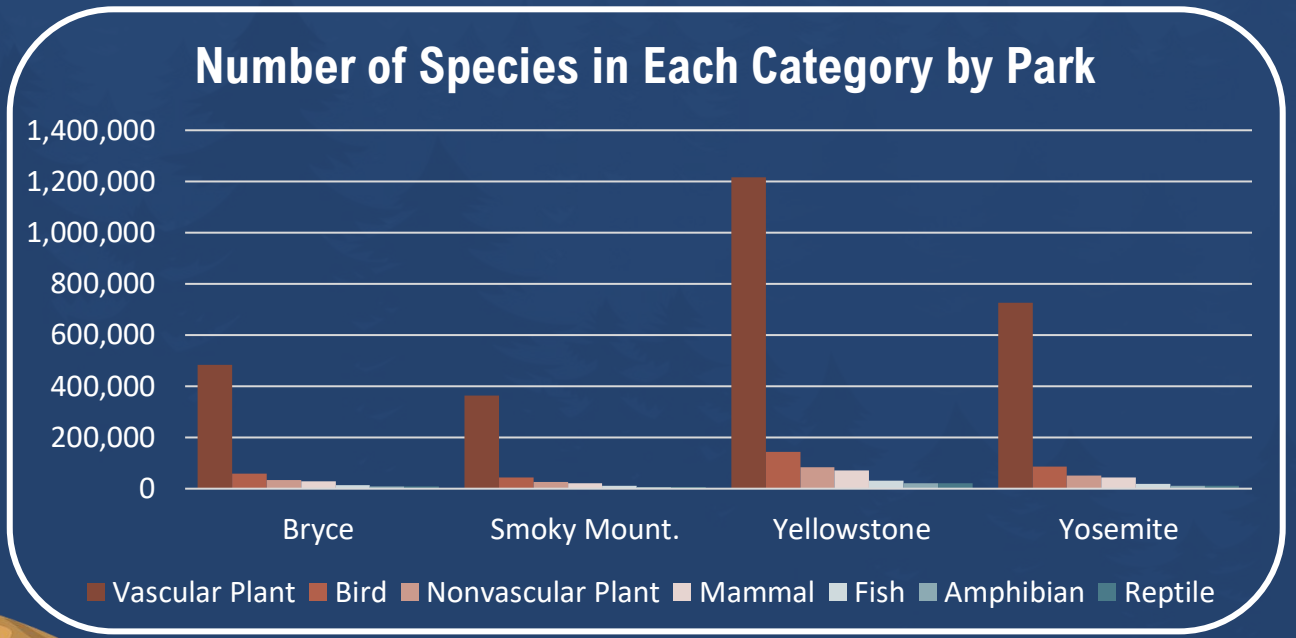
OBSERVATIONS DATA

The observations data contains recorded sightings of various species in four different national parks.

These four national parks are Yellowstone, Yosemite, Great Smoky Mountains, and Bryce National Parks.

Yellowstone has the greatest number of observations, though this may be due to the fact that it is the largest park included, covering a total of 2,221,766 acres of land. The next largest park is Yosemite with only 748,542 acres of land.





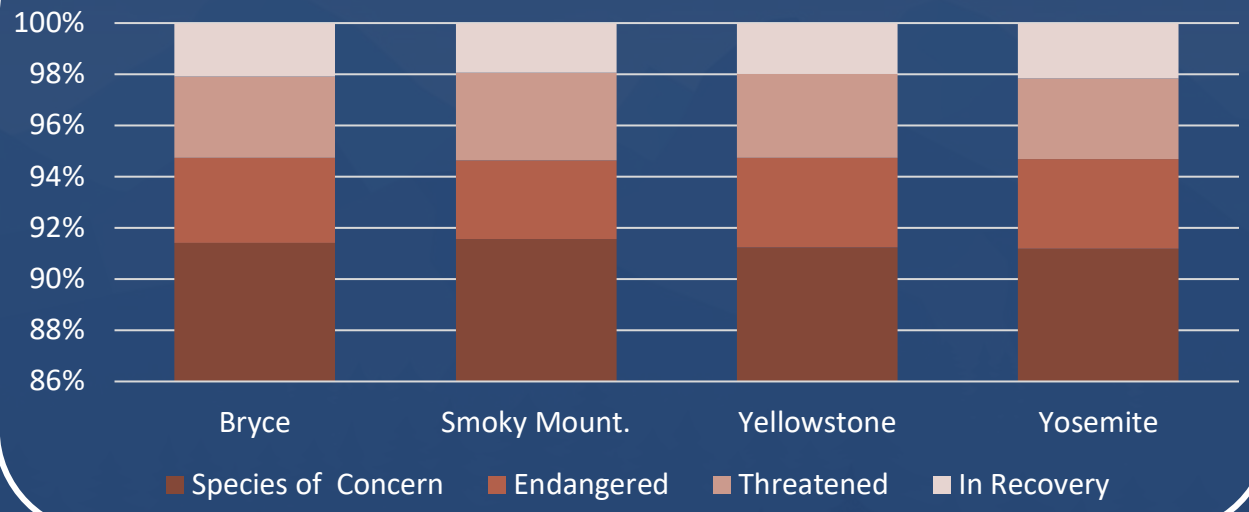
The distribution of species throughout each park is very similar.

Vascular Plant species account for most observations in all four parks, meaning someone is most likely to observe a Vascular Plant species while in any of these four parks.

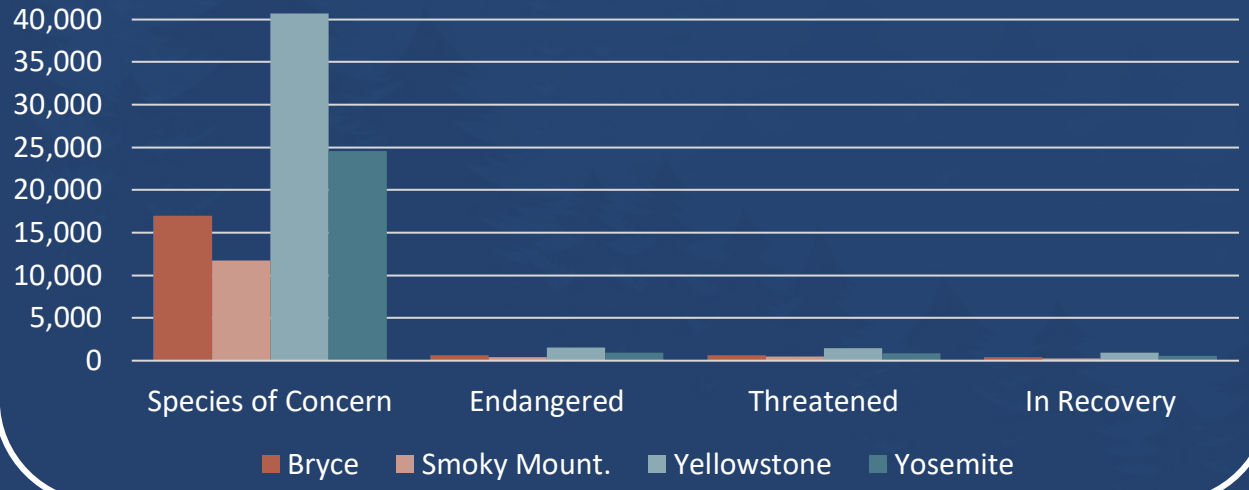
Reptile species are consistently observed least, meaning of any type of species you could observe in these four parks, you're least likely to observe Reptile species.

Are you more likely to find endangered species at a certain park?

Distribution of Conservation Status by Park



Number of Species With Each Status by Park



Each conservation status was observed most often in Yellowstone, which is to be expected since Yellowstone has the most observations overall.

However, from the proportional bar chart, we find that Yellowstone and Yosemite give someone the same likelihood of observing an endangered species.

In Yellowstone, Endangered species make up 3.5% of all observed species. In Yosemite, Endangered species make up the same proportion of observed species.

In Great Smoky Mountain National Park, Endangered species make up the smallest proportion of observations.

CONCLUSIONS

- *A majority of the observed species in this data set have no conservation status.*
- *Observed Fish species have the highest proportion of endangered speces, though they're more likely to be labeled 'Species of Concern' or 'Threatened'.*
- *Birds make up the majority of species observed with the Conservation Status 'Species of Concern'* and *'In Recovery'.*
- *The species category with the most species under protection is the Mammal category.*
- *You're most likely to observe vascular plant species and least likely to observe reptile species in any of the four national parks listed.*
- *You're most likely to observe endangered species in Yellowstone and Yosemite National Parks. You're least likely to observe endangered species in Great Smoky Mountains National Park.*