Biodiversity in National Parks

Data Analysis and Conclusions

Introduction

- This analysis uses data from four national parks recording observations of 5541 species over 7 days, as well as data on each species' common names and endangered status. I also independently added data on the size of each park.
- Questions answered in analysis:
  - Which park is home to the most endangered species?
  - Which park has the most biodiversity?
  - Is there a relationship between endangered species and observation level at the parks?
  - Which genus (type of species) is most in need of protection?

Data Review

- Observations Data
  - 23,296 rows and three columns: species scientific name, park name, and number of observations
  - No null values
  - No duplicate rows, but some duplicate observations entries (i.e. the same species list more than once for the same park, with different observation counts)
- Species Data
  - 5,824 rows and four columns: category, scientific name, common names, and conservation status
  - No null values
  - No duplicate rows, but some duplicate scientific names (two entries for the same scientific name, with different common names)

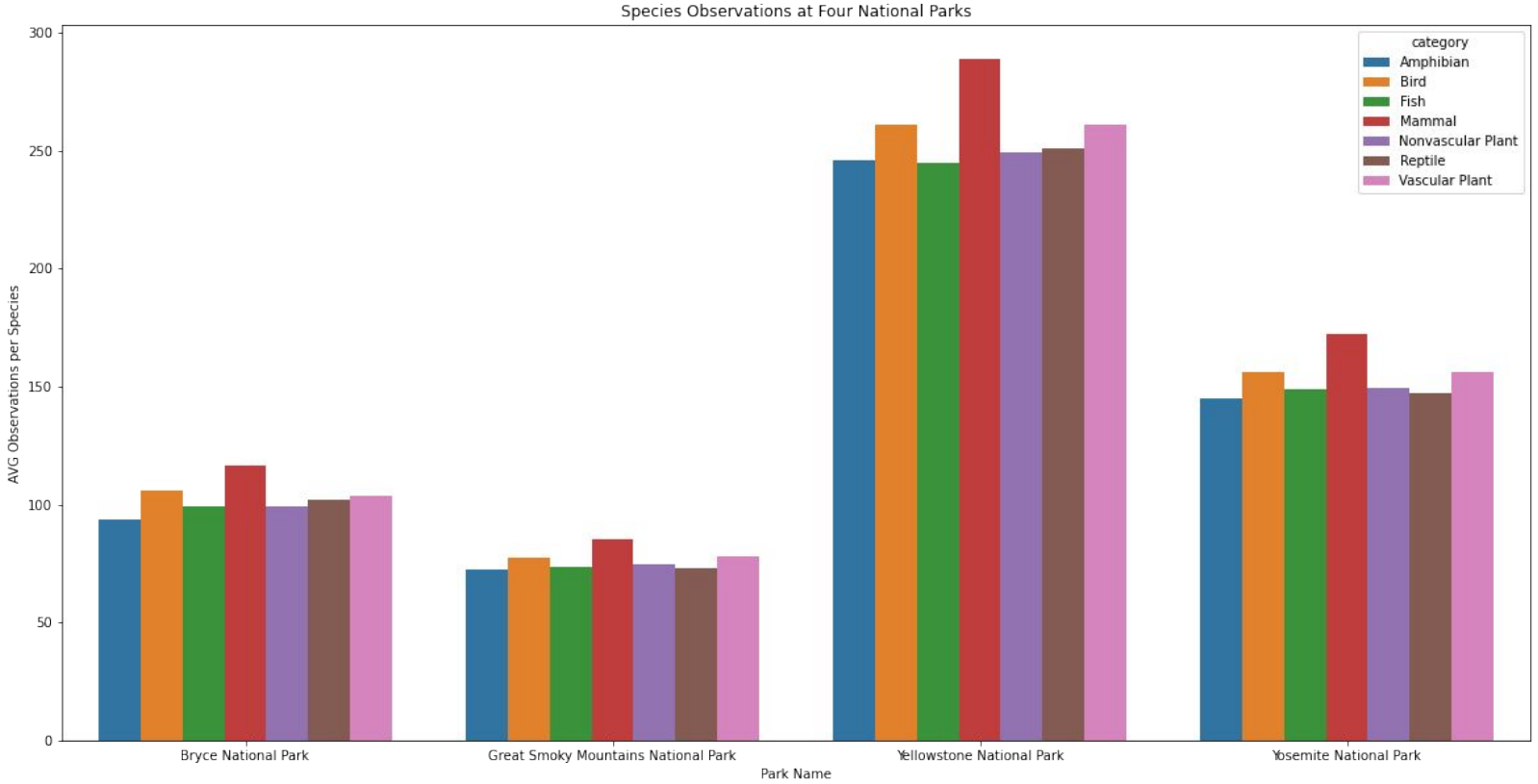
Which park is home to the most endangered species?

Park Name	Endangered Obs	Total Obs	% Endangered
Bryce	16,099	576,025	2.8%
Great Smoky Mountains	11,177	431,820	2.6%
Yellowstone	38,663	1,443,562	2.7%
Yosemite	23,356	863,332	2.7%

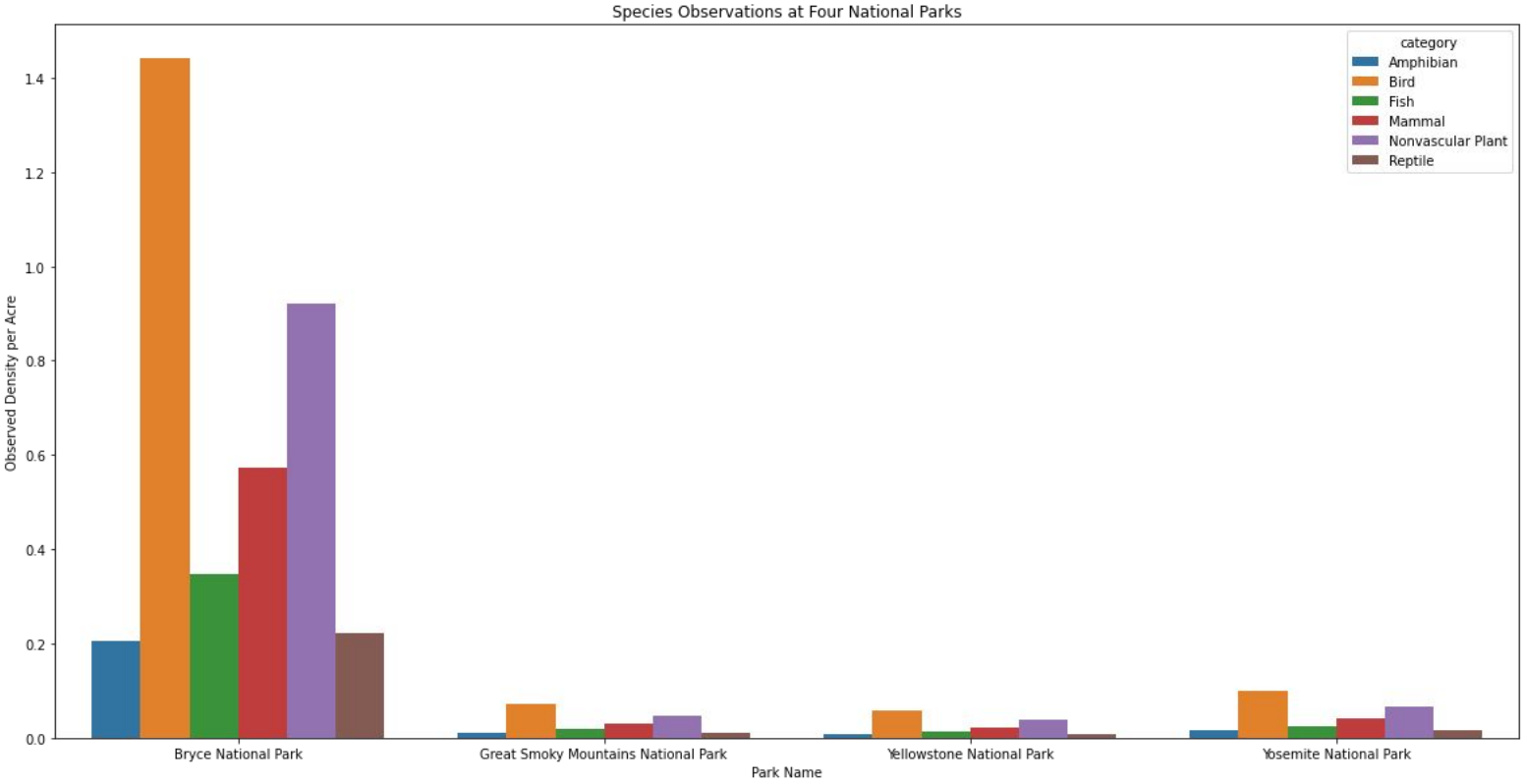
All parks had observations of all 178 endangered species, but Yellowstone National Park had the highest number of endangered species observations. However, it also had the highest number of total observations, and all parks had approximately the same percentage of observations of endangered species.

Further research (see appendix) revealed that Yellowstone is by far the largest of the parks, so the high number of observations makes sense.

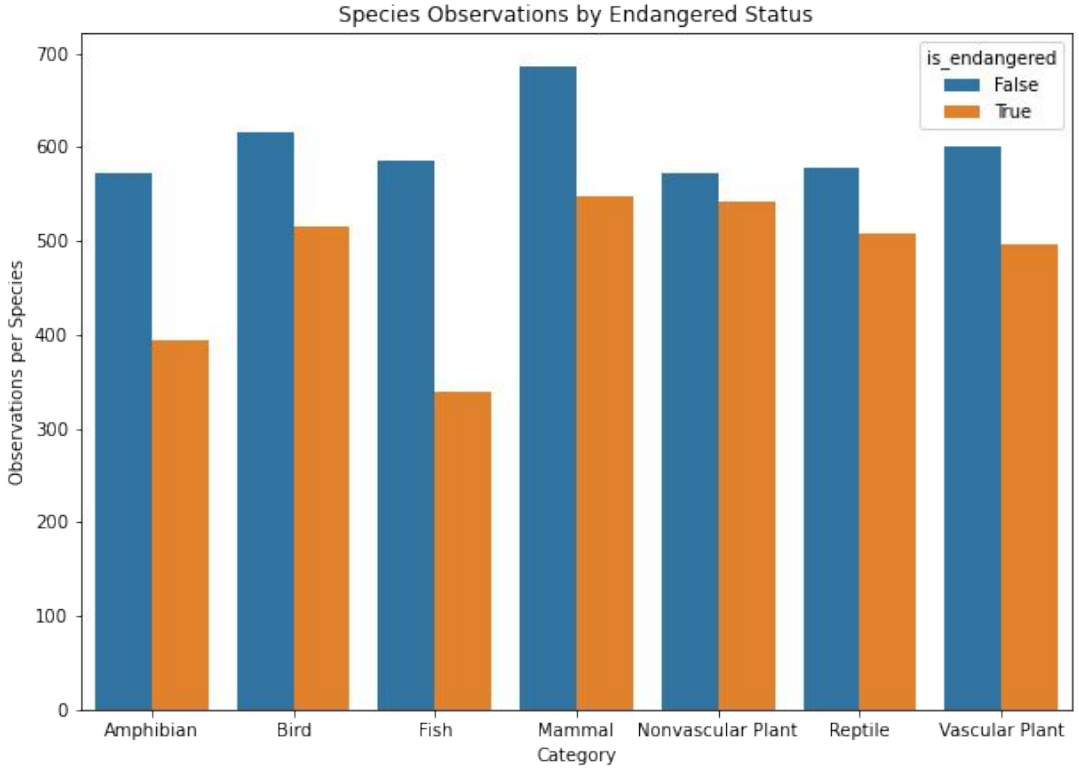
Which park has the most biodiversity?



Which park had the highest density of observations?



Are endangered species less likely to be observed?



For all species categories, there were more observations per species for non-endangered species versus endangered species.

Which type of species is most in need of protection?

Genus Name	Category	Common Name(s)	Species Count
Myotis	Mammal	Bat	12
Falco	Mammal	Falcon	4
Canis	Bird	Coyote, Wolf	3
Rana	Amphibian	Frog	3

The genus with the most endangered species was Myotis, or bats. Next up were falcons, wolves/coyotes, and frogs.

Answers to Questions

- Which park is home to the most endangered species?
  - Yellowstone is home to the most endangered species, with about 38 thousand observations in the week I had data for. This is likely due to its large size compared to the other parks.
- Which park has the most biodiversity, in plants and animals?
  - All parks had observational data for all the animals on the list, both endangered and not endangered. Bryce National Park had the greatest density of species per acre, and vascular plants had the greatest number of species observed per acre for all four parks.
- Is there a relationship between endangered species and observation level at the parks?
  - There is a statistically significant difference in observations per species between endangered and non endangered species. Endangered species are less likely to be observed than non endangered species.
- Which genus (type of species) seems to be most in need of protection?
  - The Myotis genus (bats) is most in need of protection, with 12 entries on the endangered species list. The next most endangered were Falco (falcons), Canis (wolves and coyotes), and Rana (frogs).

Further Research and Suggestions

- The observational data used in this analysis was for a seven-day period. Using observation data for many periods over the course of a year would give a better idea about the behaviors of these species.
- We know that Bryce has the highest observed density of species. Is this because it is truly more dense in wildlife, or because it is smaller and therefore easier to observe in?
- Is there observational data for these species from areas around the parks with similar geography? If so, we would be able to see if the protected status of the parks lead to more observations per species or per acre.

Appendix

- Source for park size in acres: nationalparkobsessed.com/national-parks-by-size/
- The observations and species_data csv files are available in the project repository
- The ipynb file I used to analyze the data and draw these conclusions is also in the project repository, labeled “biodiversity”

- To determine statistical significance for observations per species (endangered vs non-endangered), I conducted one-sided t-tests for all species categories except amphibians, non-vascular plants, and reptiles. These categories did not have enough species on the endangered species list to create a good sample size.
- For the remaining categories, all p-values were under 1%, except for fish, which was at 5.01%
    - Since there were only 11 endangered fish species across four parks, this was a pretty small sample size and likely responsible for the high p-value.