

# Biodiversity in National Parks

Data Analysis Report on the Endangered Species in Four National Parks



# Introduction

---

The goal of this project is to analyze data from the National Parks Service about endangered species in different parks. To accomplish this, data will be prepared, analyzed and plotted in order to better understand the conservation status of the species.

Note: The data used in this project is made to resemble real data, but it is fictitious. Both datasets were provided by [Codecademy](#) as part of the course material.

# Scope of the project

The National Parks Service wants to ensure the survival of at-risk species in order to maintain biodiversity in their parks.

Therefore, the main objectives of the analysis will be to understand the characteristics of the species and their conservation status, as well as their relationship to the national parks.

# Questions we set out to answer

1. What is the distribution of conservation status for the parks?
2. Are certain types of species more likely to be endangered?
3. Are the differences between species and their conservation status significant?
4. Which birds are the most prevalent and what is their distribution amongst parks?

# Data

# Two Datasets

## Species\_info.csv

Contains data about different species and their conservation status.

- **category**: class of animal
- **scientific\_name**: the scientific name of each species
- **common\_name**: the common names of each species
- **conservation\_status**: each species' current conservation status

Rows: 5,824

## Observations.csv

Holds recorded sightings of different species at several national parks for the past 7 days.

- **scientific\_name**: the scientific name of each species
- **park\_name**: park where species were found
- **observations**: the number of times each species was observed at park

Rows: 23,296

# Species Dataset

Contains information about **5,824 distinct species** in **7 unique categories** that include both animals and plants.

The most populous one is the **Vascular Plant** category, with 4,470 species, and the least represented is the Reptile category, with only 79 species.

Vascular Plant	4470
Bird	521
Nonvascular Plant	333
Mammal	214
Fish	127
Amphibian	80
Reptile	79

# Observations Dataset

It contains a total of 3,314,739 observations of 5,541 unique species across 4 unique parks.

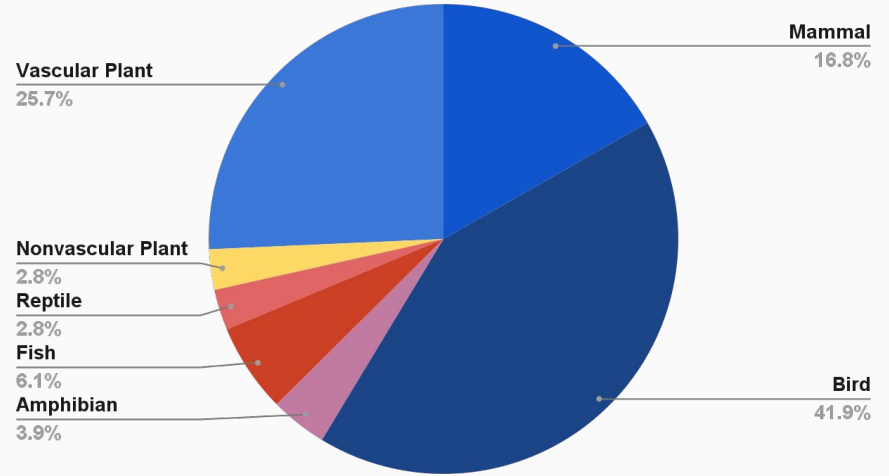
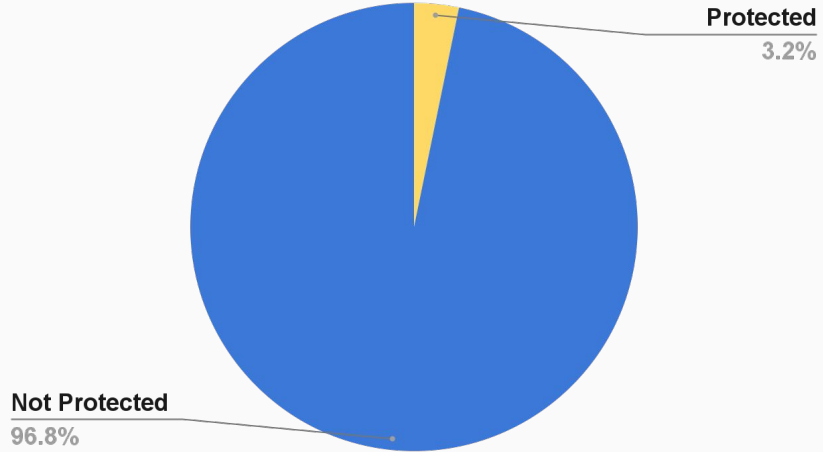
The most observed species was **Streptopelia decaocto** with a count of **1,782** observations during the 7 days.

The most observations were made in the **Great Smoky Mountains National Park**.

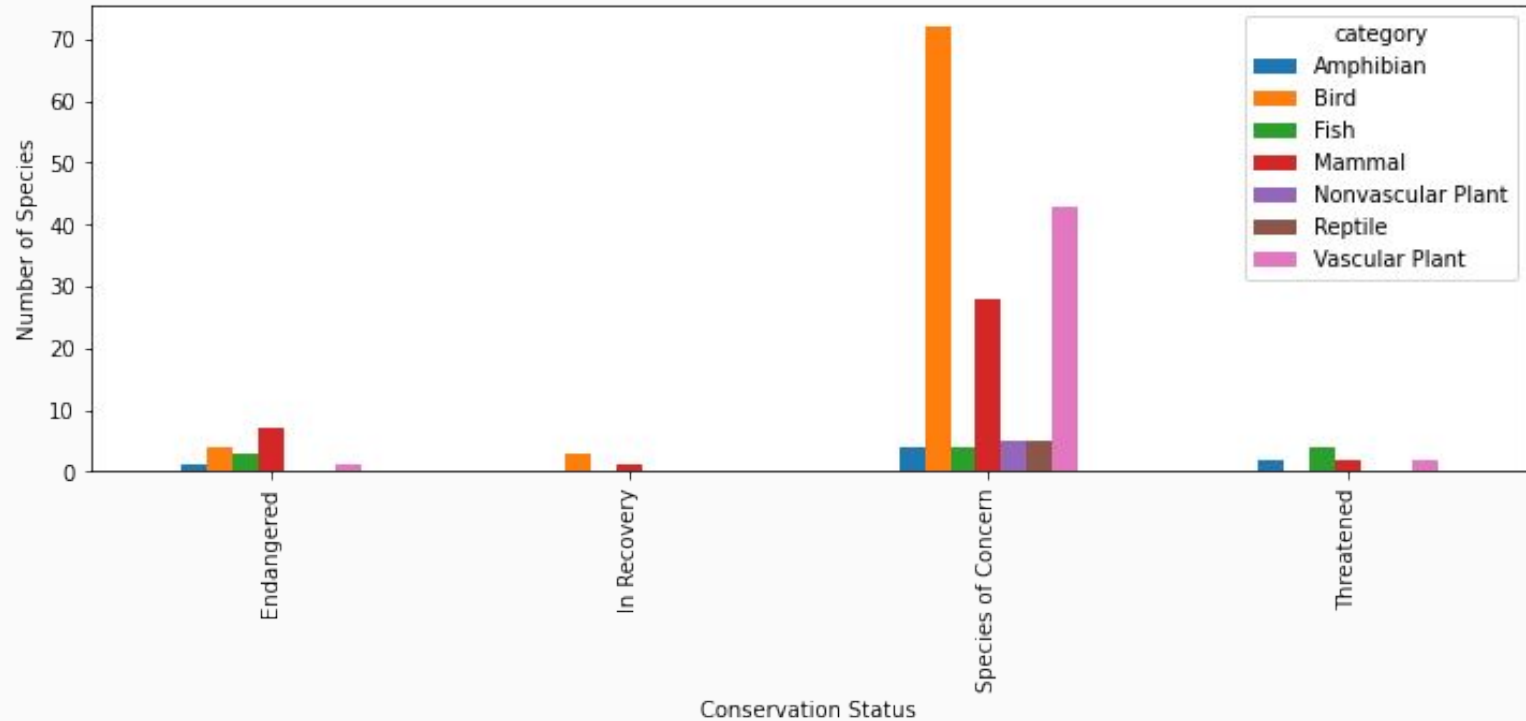


# Analysis

# Protected status



# Category breakdown of Conservation Status



# Statistical difference in conservation status between categories

Pearson's chi-squared test to determine if the difference is significant.  
Significance threshold: 0.05.

	Not Protected	Protected
Bird	413	75
Mammal	146	30

P-value: 0.6875948096661336

	Not Protected	Protected
Bird	413	75
NonVascular Plants	328	5

P-value: 1.0546306904975004e-10

# Most prevalent bird species

We have found that the most prevalent bird species in our dataset are **Warblers**, with **59 mentions**, followed by Sparrows, with 29 mentions.

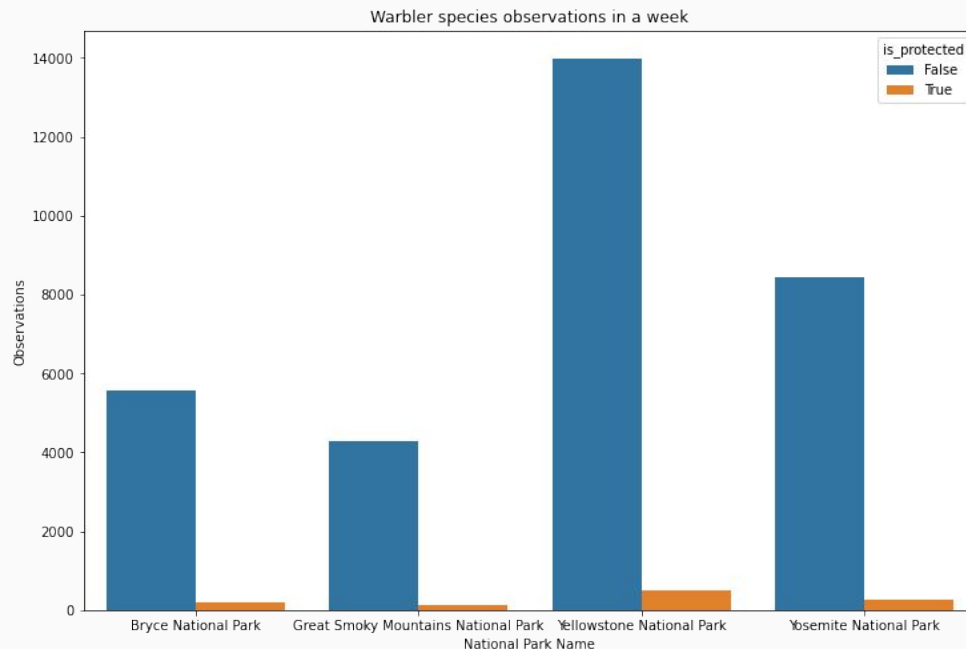
Various Passeriformes are commonly referred to as warblers. They are not necessarily closely related to one another, but share some characteristics, such as being fairly small, vocal, and insectivorous.



**Bachman's warbler** (*Vermivora bachmanii*)  
Live bird photographed by Jerry A. Payne in 1958

# 33,394 sightings in 7 days

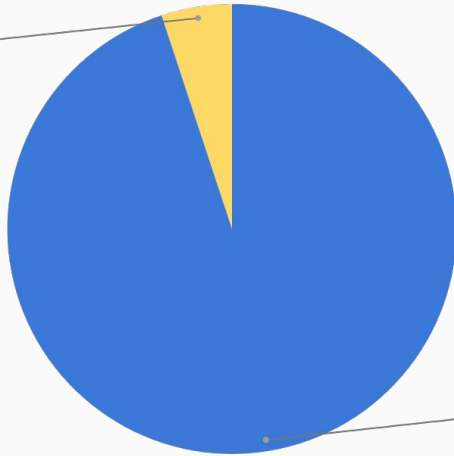
Most observations were made in the **Yellowstone National Park** with a total of **14,487** sightings, **492** of which were of protected Warbler species.



# Species vs Observations

Warblers Species Status

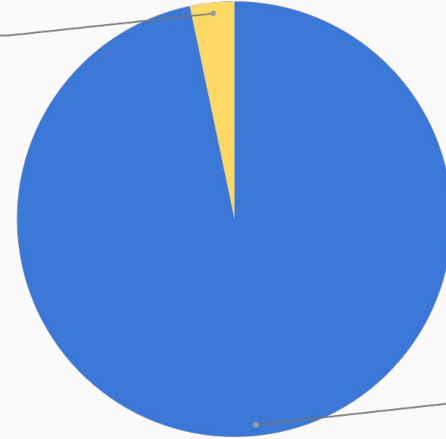
**Protected**  
5.1%



**Not Protected**  
94.9%

Warblers Observed in 7 Days

**Protected**  
3.3%



**Not Protected**  
96.7%

# Results



- What is the distribution of conservation status for species?
  - The vast majority of species were not part of conservation.(5,363 vs 179)
- Are the differences between species and their conservation status significant?
  - While mammals and birds did not have significant difference in conservation percentage, mammals and nonvascular plants exhibited a statistically significant difference.
- Are certain types of species more likely to be endangered?
  - Mammals and Birds had the highest percentage of being in protection.
- Which birds are the most prevalent and what is their distribution amongst parks?
  - The study found that Warblers occurred the most number of times and they were most likely to be found in Yellowstone National Park.