

# Biodiversity for National Parks

Introduction to Data Analytics Capstone Project

Zach Hewlin

# Species\_info.csv example

	category	scientific_name	common_names	conservation_status
1	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
2	Mammal	Bos bison	American Bison, Bison	NaN
3	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
4	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN

The dataframe categorizes National Park species by

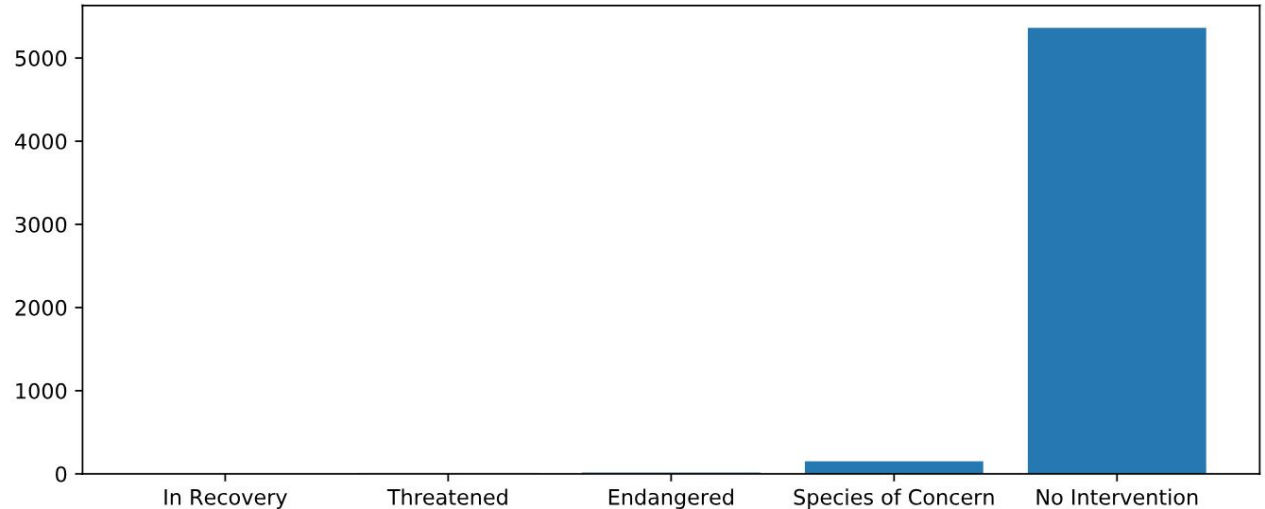
Scientific Name | Common Name | Conservation Status

The column of Conservation Status is further broken down into

Species of Concern | Threatened | Endangered | In Recovery

# Data Points of Interest

- Conservation Status contained many nan values which required some editing to make the data more easily legible
- Below is a visualization of 'Conservation Status by Species' after the nan data was altered



# Endangered Status Significance Calculations

Species Category by percentage protected:

- Found Mammals (17.1%) and Fish (15.4%) to be the most protected species types

Chi Squared Significance tests:

- **Goal:** Answer the question, “Are certain species types more likely to be endangered than others?”
- **Found** p-value **significance** in chi2 test of reptiles and mammals
- **Showcases**, some species types are more likely to be endangered than others

# Recommendations

## Focus Efforts by Species Type

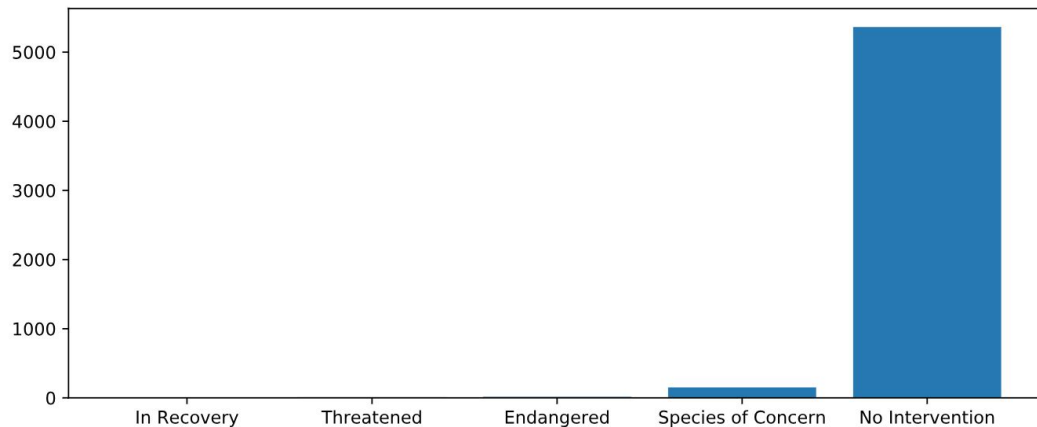
- a. Based upon the prevalence of different protection rates by Species Type it is clear that conservation efforts should be tailored to the Species Types of greatest risk.
- b. This is supported by the chi-squared significance test showing certain Species Types are more likely to be endangered than others.

# Foot & Mouth Disease Study

## Sample Size Determination

- Using Baseline Conversion Rate (15%), Statistical Significance Level (90%), and Minimum Detectable Difference (5%) I found that the study would require a sample size of 39000.
- Based upon this required sample size I determined that it would require almost 77 weeks to collect data in Yellowstone National Park, while requiring 156 weeks in Bryce National Park.

# Collected Graphs



Observations of Sheep per Week

