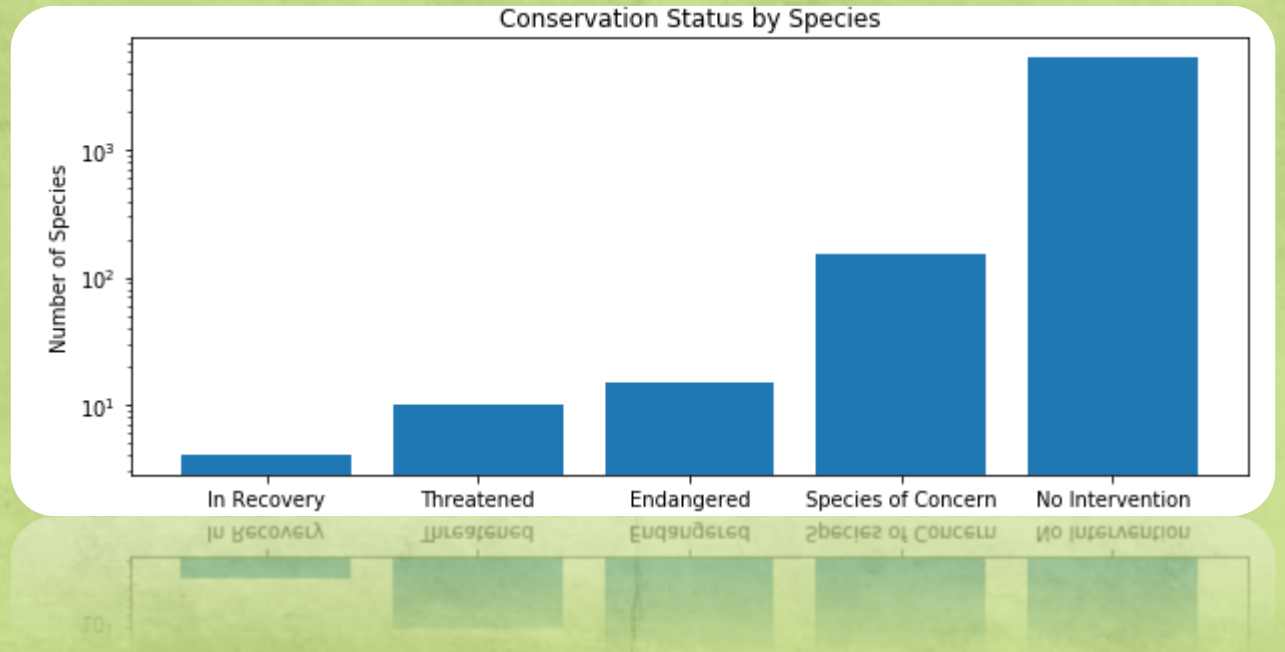
Biodiversity

Codecademy Introduction to Data Analysis Capstone Project

Species Info

- Dataset includes **5541** unique species, divided into 7 categories of plants and animals
- **180** species are subject to intervention
  - **Mammals** and **birds** are most commonly subject to intervention
  - **Vascular** and **nonvascular plants** are least commonly subject to intervention
- Note that the dataset is incomplete both in exclusion of categories (e.g. arthropods) and in enumeration (there are ~391,000 species of vascular plants)

Frequency of Intervention Among All Species

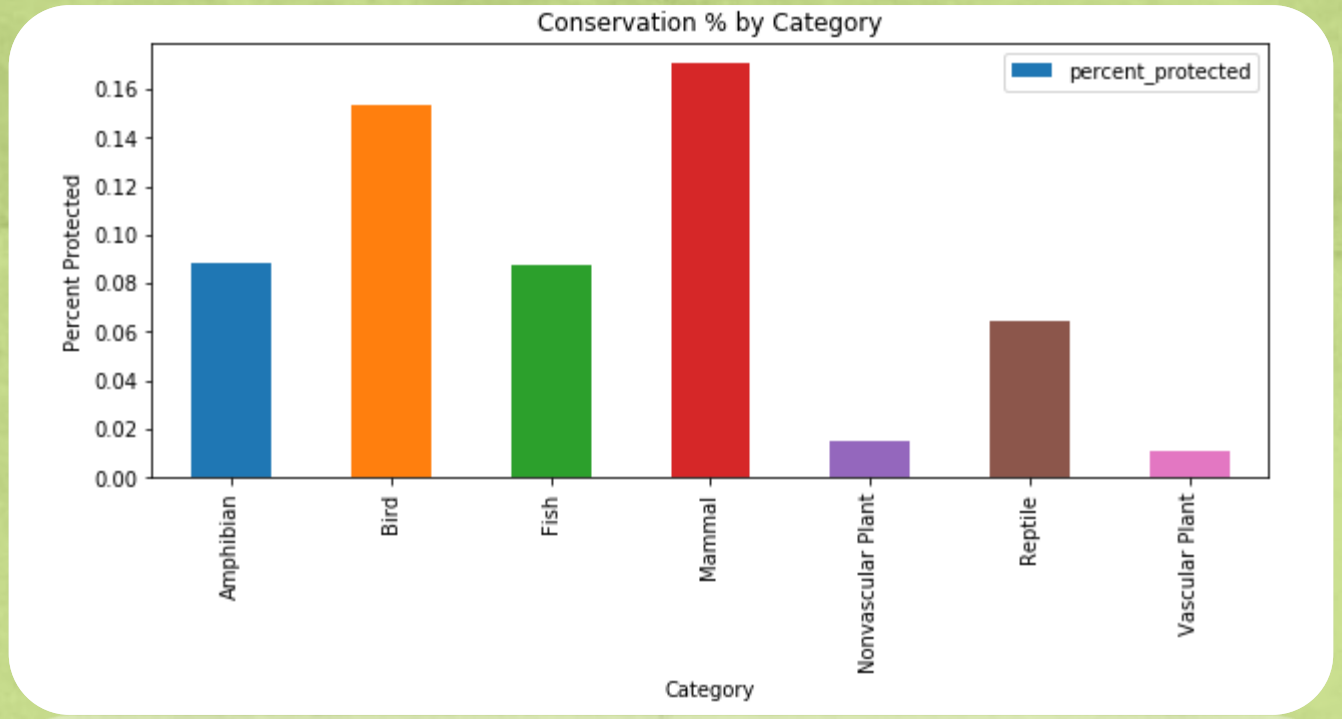


Significance of Categorical Conservation Status

- Using chi2 contingency, the difference in intervention rate between **mammals** and **birds** has a p-value of **0.69**, statistically insignificant.
- The difference in intervention rates **can** be explained by random chance alone.

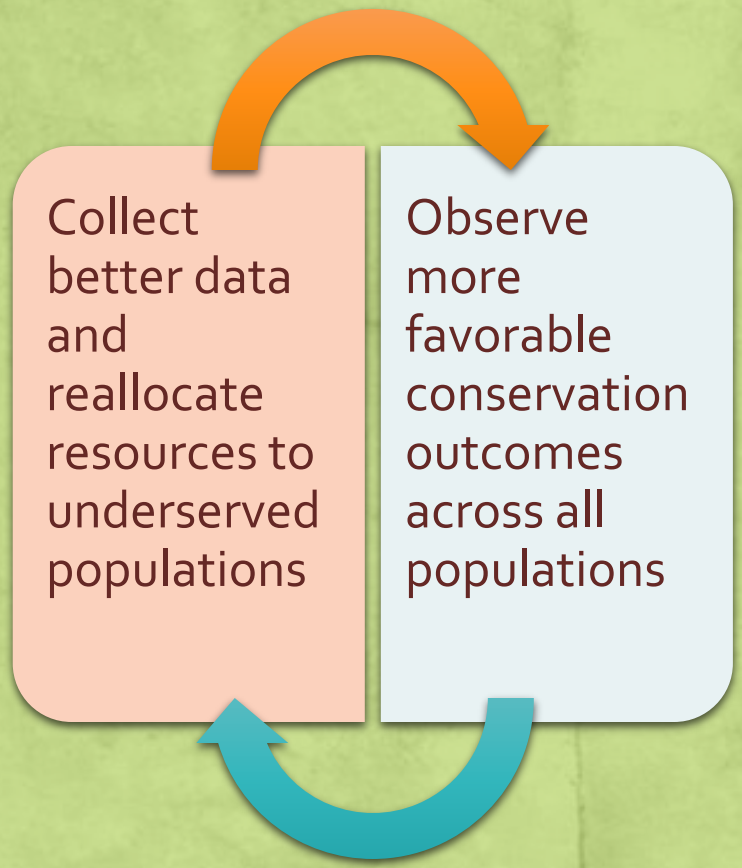
- Using chi2 contingency, the difference in intervention rate between **mammals** and **reptiles** has a p-value of **0.04**, statistically significant.
- The difference in intervention rates **cannot** be explained by random chance alone.

Categorical Conservation Rates



Recommendation for Conservationists

- Strongly consider using a more robust dataset to plan policy decisions
- Determine if additional resources should be allocated to interventions for underserved species categories:
  - Plants
  - Fish
  - Reptiles
  - Amphibians



Sample Size Determination for Impact Evaluation

Scenario Overview:

- Observe 3 species of sheep across 4 national parks and evaluate the impact of foot and mouth disease reduction initiatives
- Baseline rate: **15%**
- Minimum detectable effect: **1/3**
- Level of significance: **90%**
- This yields a sample size of **890** observations.

Time Needed to Conduct Evaluation:

- If weekly observations hold steady at each site, researchers would need to spend:
  - **1.8 weeks** observing at Yellowstone National Park
  - **3.6 weeks** observing at Bryce National Park

Observation Frequency Across Locations