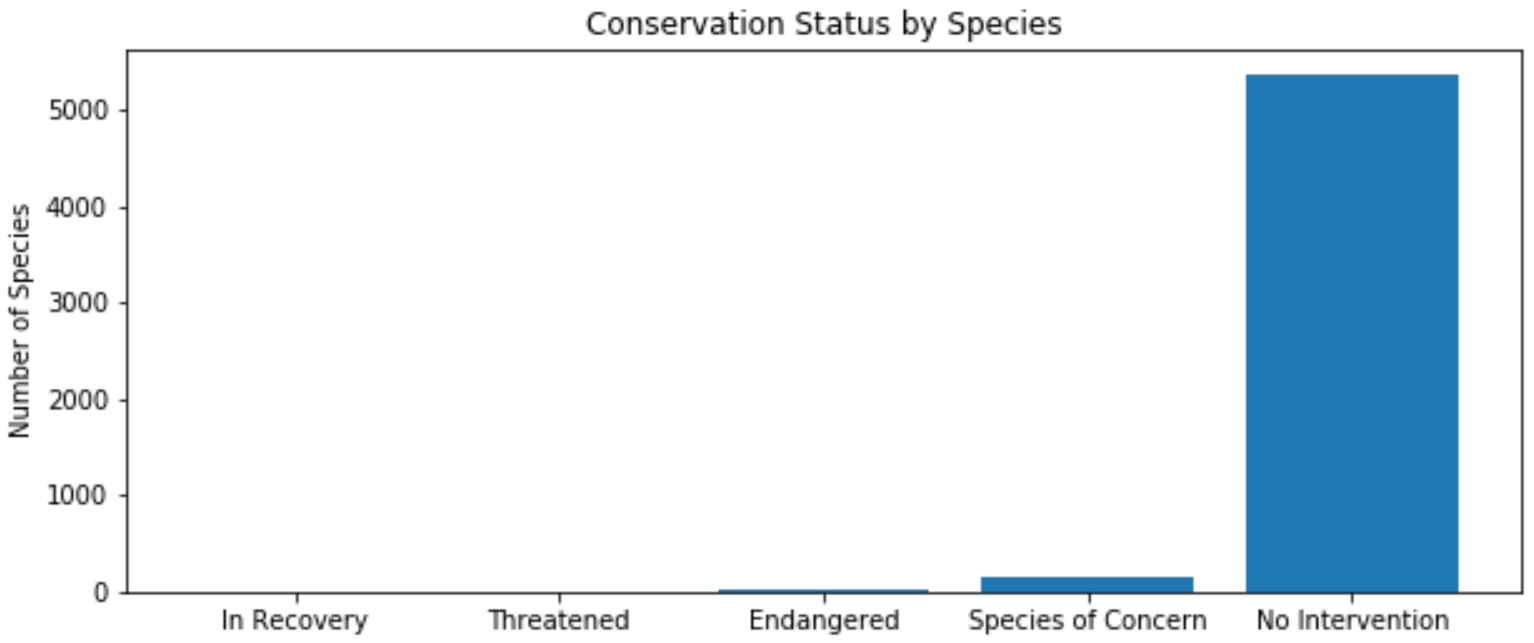
Introduction to Data Analysis - Biodiversity Capstone project

CodeAcademy username: homunculus631

Description of species

- **Dataframe with 5823 rows & 4 columns:**
  - **Category**
  - **scientific_name**
  - **common_names**
  - **conservation_status**
- **5541 unique values for scientific_name (i.e., species)**
- **Seven values for category**
- **Four (non-NaN) values for conservation_status**

Bar Graph of species by conservation status



Describe the significance calculations for endangered status between different categories of species

- The problems was to test if statistical significance existed between two different categorical variables (e.g. ‘protected’ (or not) categories of mammals versus birds or reptiles). Chi-squared test are appropriate for such tests.
- The required input to the Chi-Squared function, i.e., the contingency table, was directly provided in the species dataframe

Recommendation for conservationists concerned about endangered species, based on significance calculations

- The result showed that while the protected status difference of between mammal and birds was not significant, the difference between mammal and reptiles was significant

Description of sample size determination for the foot and mouth disease study

- Merging observation and species dataframes allowed for compiling sheep observations at the National Parks