# Capstone 2: Biodiversity Project

By Alicia (Ali) Alvarez

# Agenda

- Species Raw Data
- Endangered Status per Category
- Recommendation for conservationists
- Sample size determination for the foot and mouth disease study
- Graphics

# Species Raw Data

The **species_info.csv** contains 4 columns: category, scientific_name, common_names, conservation_status

5541 Unique Species [scientific_name]

7 Unique Categories

| category | scientific_name |
|---|---|
| Vascular Plant | 4262 |
| Bird | 488 |
| Nonvascular Plant | 333 |
| Mammal | 176 |
| Fish | 125 |
| Amphibian | 79 |
| Reptile | 78 |

5 Unique Conservation Status

| conservation_status | scientific_name |
|---|---|
| No Intervention | 5363 |
| Species of Concern | 151 |
| Endangered | 15 |
| Threatened | 10 |
| In Recovery | 4 |

# Endangered Status per Category

Calculations made for endangered status between different categories of species

1) Determined which species require intervention

```
species.fillna('No Intervention', inplace=True)
```

2) Grouped by Conservation Status to evaluate how many species required conservation

| conservation_status | scientific_name |
|---|---|
| No Intervention | 5363 |
| Species of Concern | 151 |
| Endangered | 15 |
| Threatened | 10 |
| In Recovery | 4 |

# Endangered Status per Category

Calculations made for endangered status between different categories of species

3) Calculated new field is_protected by using a lambda formula:

```
my_lambda = lambda x: True if x == 'No Intervention' else False
species['is_protected']=species.conservation_status.apply(my_lambda)
#checking it worked
species.head(10)
```

4) Determined the % of protected species per category, by using a pivot table and a new calculation:

```
category_pivot = category_counts.pivot(columns='is_protected',
                                       index='category',
                                       values='scientific_name')\
                                .reset_index()
```

```
category_pivot.columns = ['category', 'not_protected', 'protected']
category_pivot.head(5)
```

| | category | not_protected | protected | percent_protected |
|---|---|---|---|---|
| 0 | Amphibian | 7 | 72 | 0.911392 |
| 1 | Bird | 75 | 413 | 0.846311 |
| 2 | Fish | 11 | 115 | 0.912698 |
| 3 | Mammal | 30 | 146 | 0.829545 |
| 4 | Nonvascular Plant | 5 | 328 | 0.984985 |
| 5 | Reptile | 5 | 73 | 0.935897 |
| 6 | Vascular Plant | 46 | 4216 | 0.989207 |

# Endangered Status per Category

Calculations made for endangered status between different categories of species

## 5) Found significant difference between categories by performing chi study on all values:

```
In [12]: import pandas as pd
         import numpy as np
         from scipy.stats import chi2_contingency

In [45]: def load_data():
             df = pd.DataFrame([
                 ['Amphibian',7,72,0.911392],
                 ['Bird',75,*413,0.846311],
                 ['Fish',11,*115,0.912698],
                 ['Mammal',30,146,0.829545],
                 ['Nonvascular Plant',5,328,0.984985],
                 ['Reptile',*5,73,0.935897],
                 ['Vascular Plant',46,4216,0.989207] ],
                 columns=['category','not_protected','protected','percent_protected'])
             return df

In [46]: def chi(cat1, cat2,contingency):
             chi2, pval, dof, expected = chi2_contingency(contingency)
             if pval<0.05:
                 print("significant difference exists between", cat1," and ", cat2)

In [47]: #_main_
         df=load_data()
         for row in range (0,len(df)-1):
             cat1=df.loc[row][0]
             for row2 in range(1,len(df)):
                 cat2=df.loc[row2][0]
                 # calculating contingency
                 contingency = [df.loc[row][1],df.loc[row][2]],[df.loc[row+1][1],df.loc[row+1][2]]
                 chi(cat1,cat2,contingency)
```

### Results of Chi study:

```
significant difference exists between Mammal  and  Bird
significant difference exists between Mammal  and  Fish
significant difference exists between Mammal  and  Mammal
significant difference exists between Mammal  and  Nonvascular Plant
significant difference exists between Mammal  and  Reptile
significant difference exists between Mammal  and  Vascular Plant
significant difference exists between Nonvascular Plant  and  Bird
significant difference exists between Nonvascular Plant  and  Fish
significant difference exists between Nonvascular Plant  and  Mammal
significant difference exists between Nonvascular Plant  and  Nonvascular Plant
significant difference exists between Nonvascular Plant  and  Reptile
significant difference exists between Nonvascular Plant  and  Vascular Plant
significant difference exists between Reptile  and  Bird
significant difference exists between Reptile  and  Fish
significant difference exists between Reptile  and  Mammal
significant difference exists between Reptile  and  Nonvascular Plant
significant difference exists between Reptile  and  Reptile
significant difference exists between Reptile  and  Vascular Plant
```

Note: Calculations of this slide are  included in further_exploration.py

# Recommendations

Recommendation for conservationists

- Increase number of protected species that are either Mammal or Bird since these two represent 59% of the unprotected species.
- Increase efforts to study the three species categories that are more likely to be endangered than the others, these are: Mammal, Nonvascular Plant and Reptile.

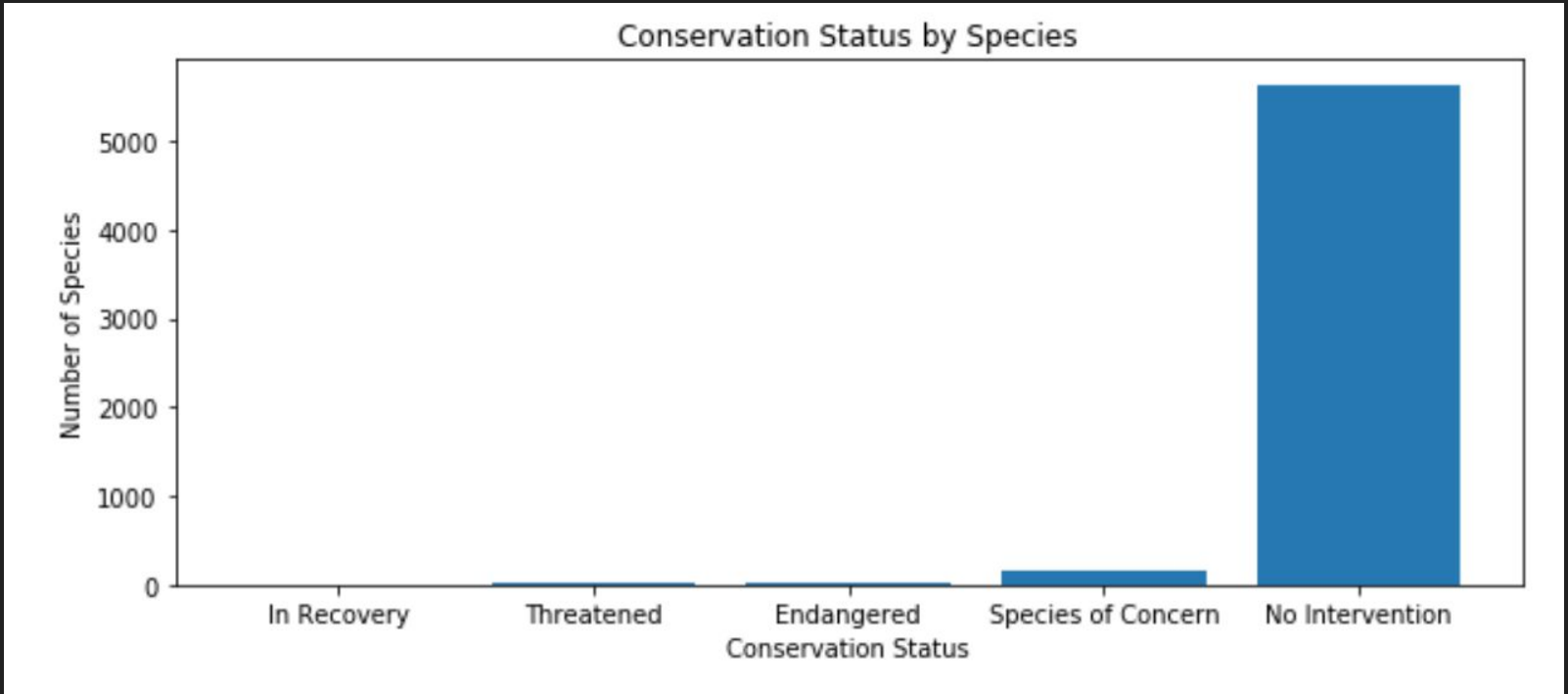
# Sample Size For Foot & Mouth Disease Study

The calculated sample size per variant is 520. This value is the result of using the following parameters:

- minimum_detectable_effect = 33%
- baseline = 15%
- Statistical Significance: 90%

# Graphics

## Conservation Status by Species



# Graphics

## Observations of Sheep per Week