# Biodiversity for the National Parks

**Submitted by:** Tracy Laranjo
**IDA:** 4/9 Cohort

# Species DataFrame Observations

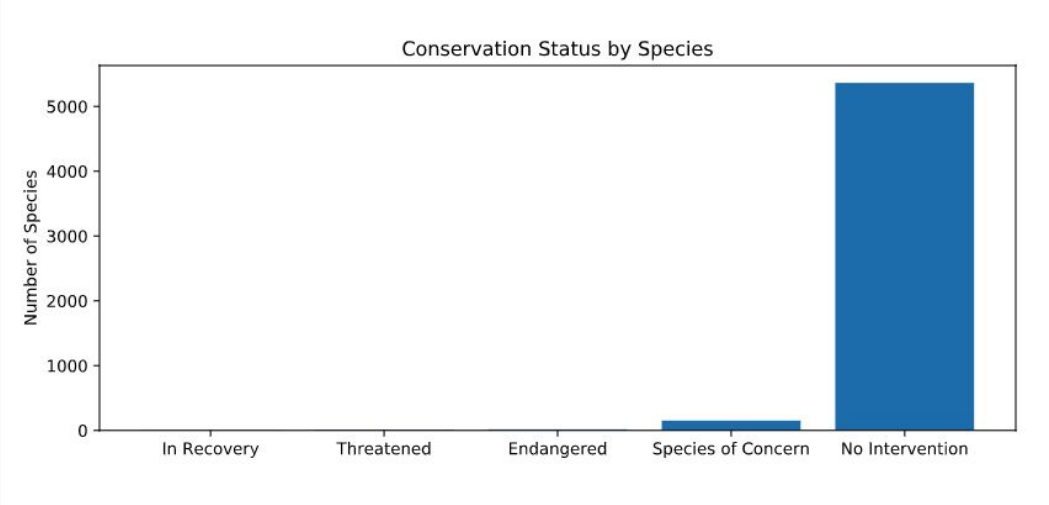
**Data:** species_info.csv

# Species DataFrame Observations

- There are **5,541** different species that fall under **7** unique category values: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant
- There are **5** unique conversation statuses: Species of Concern, Endangered, Threatened, In Recovery, NaN (No Intervention)
    - NaN (“Not a Number”) represents a null value and the species that require no intervention. We will replace NaN values with ‘No Intervention’ using `.fillna`.
- After replacing NaN values, we can see that the majority of species (**96.8%**) do not require intervention, which is a good thing!

# Species DataFrame Observations

We can see these results visually through a bar chart:



# Species DataFrame Observations

- The remaining **3.2%** are protected species.
- We want to find out if certain categories of species are more likely to be protected. This is accomplished by creating a pivot table grouped by category and protection status. By adding a `percent_protected` column, we are able to break down how many species of each category are have a protected status:
  - **8.9%** of amphibian species are protected.
  - **15.4%** of bird species are protected.
  - **8.7%** of fish species are protected.
  - **17.0%** of mammal species are protected.
  - **1.5%** nonvascular plant species are protected.
  - **6.4%** of reptile species are protected.
  - **1.1%** of vascular plant species are protected.

***Preliminary observations show that mammal and bird species are the most likely to be protected.***

# Significance Calculations

# Significance Calculations

- Mammals appear to be more likely than birds to be protected species but we cannot confirm that the **1.6%** difference is significant until we perform a chi-squared test.
    - Chi-squared tests are ideal for testing this hypothesis because we have two categorical datasets that we want to compare.
- The difference between mammals and birds is **not significant** because the p-value of 0.69 calculated from the chi-squared test is greater than 0.05. This is a result of chance so we cannot say with confidence that either one is more likely to be protected.
- The difference between mammals and reptiles is **significant** because the p-value of 0.04 calculated from the chi-squared test is less than 0.05. From this, we can conclude that some species, like mammals and birds, are more likely to be protected than others.

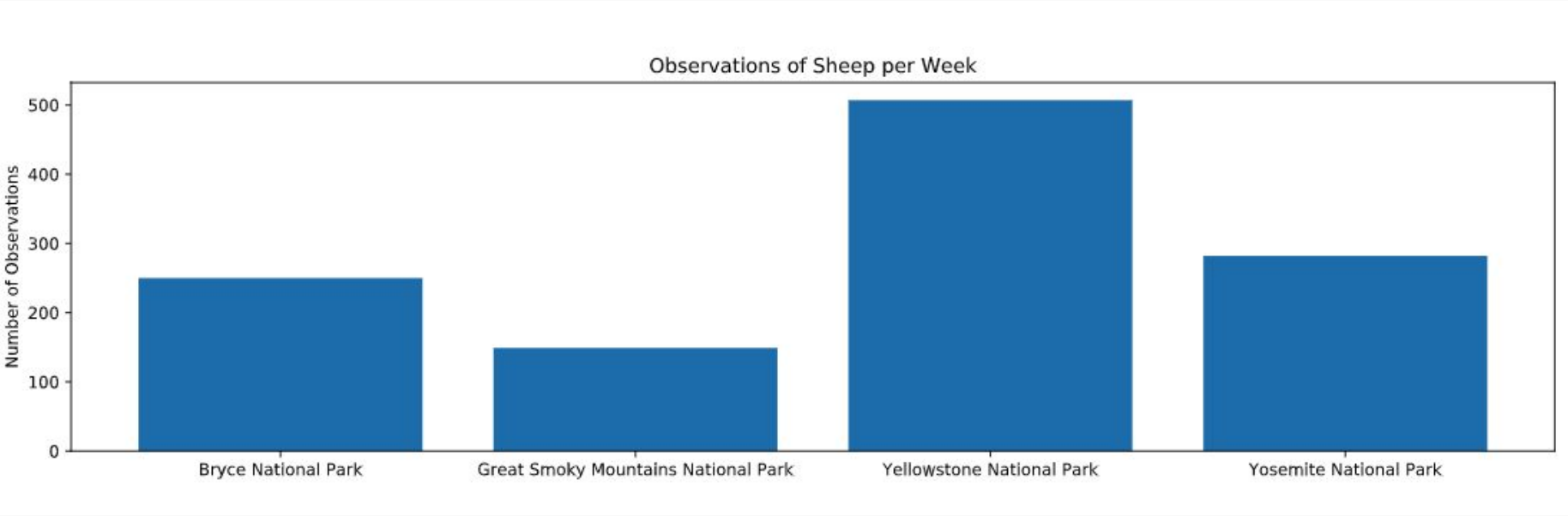
# Recommendations

# Recommendations

- Prioritize the conservation of the 15 endangered and 10 threatened species.
- Closely monitor the 4 species in recovery and 151 species of concern.
- Pay special attention to mammals and birds. They are the species that are most likely to have a protected status.
- The species with the lowest risk of protected status are vascular plants and nonvascular plants. Monitor - but do not prioritize - the conservation of these species.

# Foot and Mouth Disease Study

# Foot and Mouth Disease Study

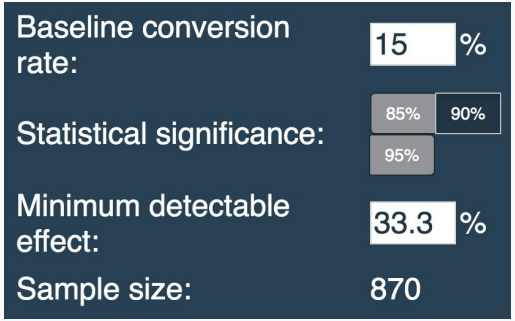


# Foot and Mouth Disease Study

- The baseline percentage would be **15%** as this was the recorded percentage of sheep with foot and mouth disease at Bryce National Park. Yellowstone National Park can use this as their baseline.
- The statistical significance specified in the instructions is **90%**.
- We want to see a 5% reduction so this is how we would calculate the minimum detectable effect, with decimals for accuracy:

```
minimum_detectable_effect = 100*5./15 = 33.3333333333 (33.3%)
```

- After plugging the minimum detectable effect in the sample size calculator, we come up with **870** sample size per variant.



# Foot and Mouth Disease Study

As we learned in task #12, scientists at Yellowstone National Park can observe **507 sheep** in 1 week. If we want to observe 870 sheep to hit our sample size, they would need **1.7 weeks** to observe enough sheep. This is the ideal place to observe sheep as it is the park that would take the least amount of time to hit our sample size.

```
yellowstone_weeks_observing = sample_size_per_variant/507.
```

Brice Park can observe **250 sheep** in 1 week, so they would need **3.5 weeks** to hit sample size.

```
bryce_weeks_observing = sample_size_per_variant/250.
```