BIODIVERSITY

An Analysis of Species Conservation Status in the American National Parks

THE BIODIVERSITY PROJECT

The Biodiversity project is an effort to map wildlife for the National Parks Service. The project aims to answer the following questions:

1. *What species data is available and what is the conservation status of different species?*
2. *Are there any patterns explaining why some species are endangered?*
3. *On which species should conservationists concentrate their efforts?*
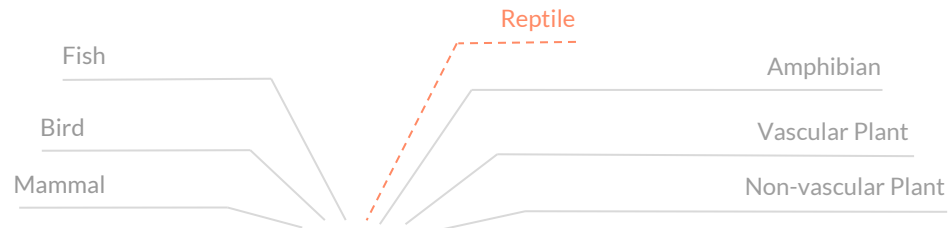
Further, departing from the recent Foot and Mouth disease reduction program, specific study guidelines will be presented:

4. *What sample sizes are needed to make reliable estimates of disease within a population?*

ANALYSIS OF AVAILABLE DATA

1. *What species data is available and what is the conservation status of different species?*

Above: the columns of species.csv to the left and an example of a row to the left.

A CLOSER LOOK AT THE SPECIES DATA

5541 total species in **7** categories

176	488	125	78	79	4262	333
Mammals	Birds	Fish	Reptiles	Amphibians	Vascular Plants	Nonvascular Plants

CONSERVATION STATUS BY SPECIES GROUP

5541 total species

5365
No Intervention

4
In Recovery

151
Species of
Concern

15
Threatened

10
Endangered
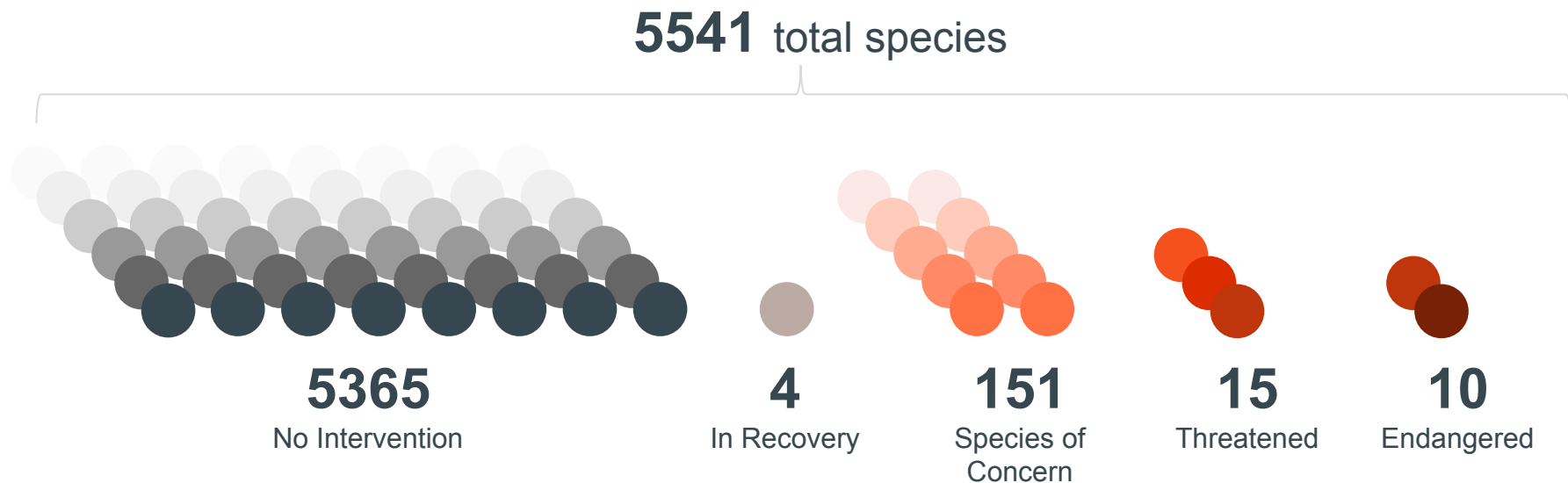
The matplotlib plot of the conservation status by species can be found in the appendix,

IDENTIFYING PATTERNS

2. *Are there any patterns explaining why some species are endangered?*

PROTECTION STATUS PER SPECIES CATEGORY

% protected per category

17.0	**15.4**	**8.7**	**6.4**	**8.9**	**1.5**	**1.1**
30 out of **146** protected	**75** out of **413** protected	**11** out of **115** protected	**5** out of **73** protected	**7** out of **72** protected	**46** out of **4216** protected	**5** out of **328** protected

A WORD ON STATISTIC SIGNIFICANCE

- When comparing features between different populations, one needs to determine if any observed difference is real or due to chance.
- Generally, a difference is deemed significant if the calculated p-value is lower than 0.05. This means:
  - that 1 in 20 such results would be due to chance
  - the difference observed is likely to be due to a real difference between the populations

- *Our researchers have used a statistical hypothesis test called the Chi-square test to determine whether observed difference in the protection status between mammals, birds and reptiles are due to chance or not.*
  - *The standard p-value of 0.05 has been used in the test.*

PROTECTION STATUS: BIRDS VS MAMMALS

p-value = **0.69**

Not Significant

15.4
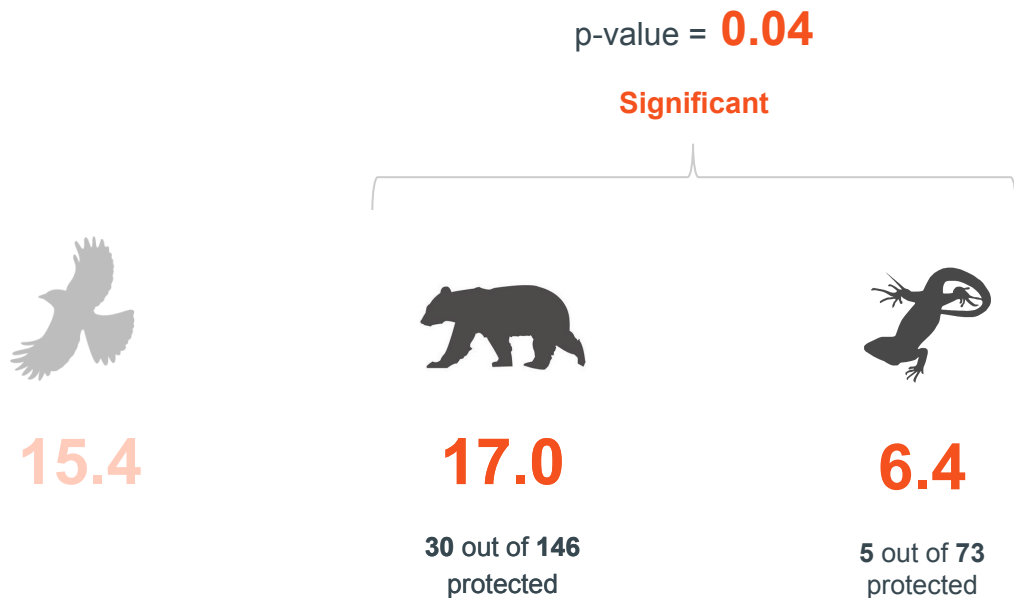
17.0

6.4

75 out of **413**
protected

30 out of **146**
protected

*Although 17 is more than 15.4%, both mammals and birds
seem to be equally likely to become endangered. The
difference is most likely due to chance.*

PATTERNS IDENTIFIED

- It seems mammals are more vulnerable compared to reptiles, while mammals and birds are equally likely to require protection

- Based on the available data, it would seem that birds and mammals are the most vulnerable while vascular and nonvascular plants are the most robust. *To confidently draw such conclusions, further chi-square testing of the remaining categories must be done.*

- It's further recommended to conduct chi-testing with extended categories.
  - Instead of protected/not protected one could use the original five protection status categories.
  - Thus, one could *determine if a certain category of species is more likely to be in a more critical protection status, i.e. 'endangered' or 'threatened' as opposed to 'species of concern'*

RECOMMENDATIONS

3. *On which species should conservationists concentrate their efforts?*

WHERE SHOULD EFFORTS BE CONCENTRATED?

- Based on initial findings, mammals and birds should be prioritized while reptiles can be given less priority
- These are initial findings and more testing needs to be performed before any decisions are made. One suggestion would be to conduct new chi-square testing where the 'protected'-group is split into the actual conservation status categories.

| Nonvascular Plants | Vascular Plants | Reptiles | Fish | Amphibians | Birds | Mammals |

Initial Recommendation for Protection Priority

Low   High

SAMPLE SIZE RECOMMENDATIONS

4. *What sample sizes are needed to make reliable estimates of disease within a population?*

ASSESSING IMPACT OF A PROGRAM

- Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease, which is a disease that affects sheep
    - 15% of sheep in the Bryce national park carry the disease.
    - Researchers want to be able to detect reductions of at least 5 % in the Yellowstone park
- To assess the effects of that study, one needs to know what sample sizes are needed to draw reliable conclusions. Three parameters are relevant for determining sample sizes:

Baseline Conversion Rate

- The control group's expected conversion rate, i.e. 15% in Bryce National Park

Minimum Detectable Effect

- A percentage of the baseline
- equals approx 33.33% with the baseline conversion rate of 15 %
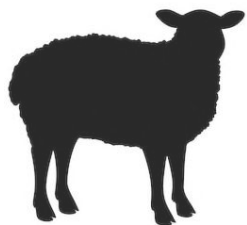
Statistic Significance

- The probability that the observed result is due to change. p = 0.1 will be used

Based on above data, the sample size for the study should be **520** *sheep per park*

HOW LONG WILL THE STUDY TAKE?

- Researchers want to know how much time a study of the impact of the foot and mouth disease reduction program would take
- Along with the sample size, the following information is relevant to determine the time of study:
  - The observation frequency of sheep in the Yellowstone park (study group)
  - The observation frequency of sheep in the Bryce park (control group)

- Observation data is available through the 'observations.csv' file

OBSERVATION DATA EXAMPLE

One of several types of sheep

Yosemite National Park

Bryce National Park

Great Smoky Mountains National Park

Yellowstone National Park

Scientific Name: Ovis Canadensis

Park Names: Yosemite National Park

Observations: 117

Unique value

Above: the columns of observations.csv to the left and an example of a row to the left.

TOTAL SHEEP OBSERVATIONS PER PARK AND DATA EXAMPLE

Observations of Sheep per Week
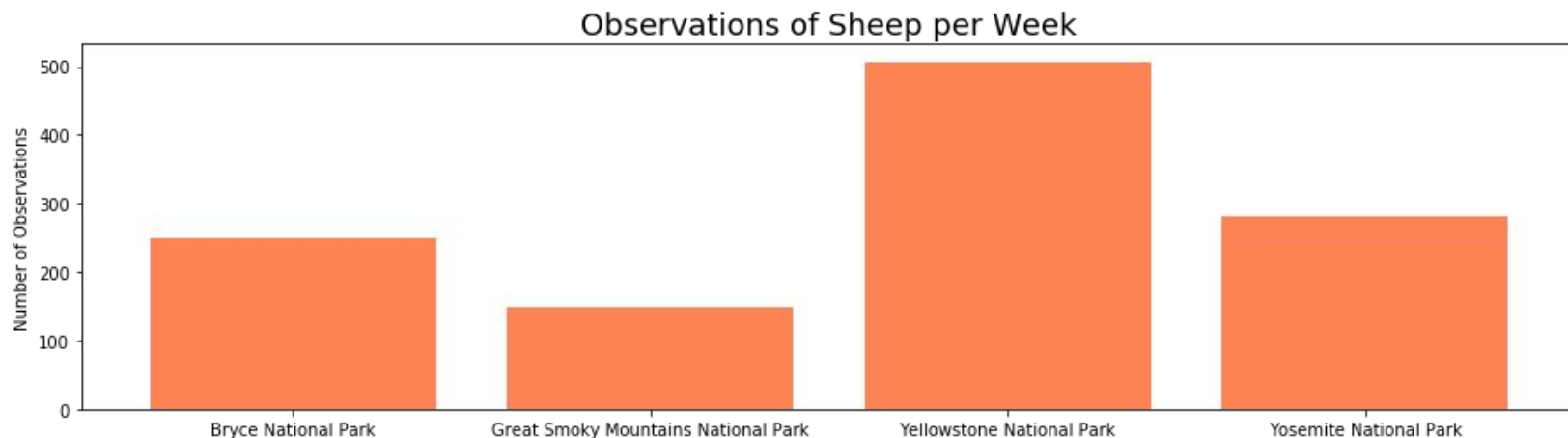
Bryce National Park: **250**

Yellowstone National Park: **507**

Great Smoky Mountains National Park: **149**

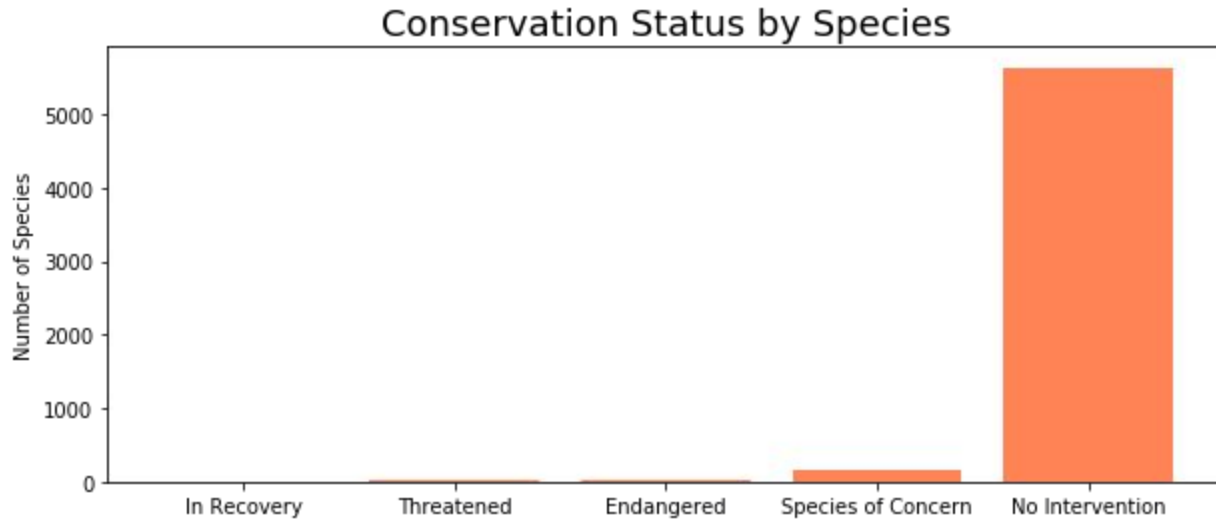Yosemite National Park: **282**

○

CONCLUSIONS

- Based on the following data:
  - Sample size: **520** sheep per park
  - Observation frequency: **507** sheep per week in Yellowstone National Park
  - Observation frequency: **250** sheep per week in Bryce National Park

The study will take 507 / 520 ≈ 1 week in Yellowstone National Park

and 250 / 520 ≈ 2 weeks in Bryce National Park

THANK YOU

APPENDIX

Bar chart of conservation status by specie, see slide 6. It's not included in the actual presentation since it was too hard to read.