Data Analysis about Biodiversity

Shuhei Ishida

Table of Contents

- Description of the data.
- Significance calculations for conservation status.
- A recommendation for conservationists.
- Description of foot and mouth disease study.

Description of the Data -summary

- “species_info.csv” is the Data about different species in U.S. National Parks.
- A table below shows a summary of the data.
  - As this shows, most of the species are not conserved (Conservation_status is NaN.).

	category	scientific_name	common_names	conservation_status
Count	5824	5824	5824	191
Unique	7	5541	5504	4
Top	Vascular Plant	Streptopelia decaocto	Brachythecium Moss	Species of Concern
Freq	4470	3	7	161

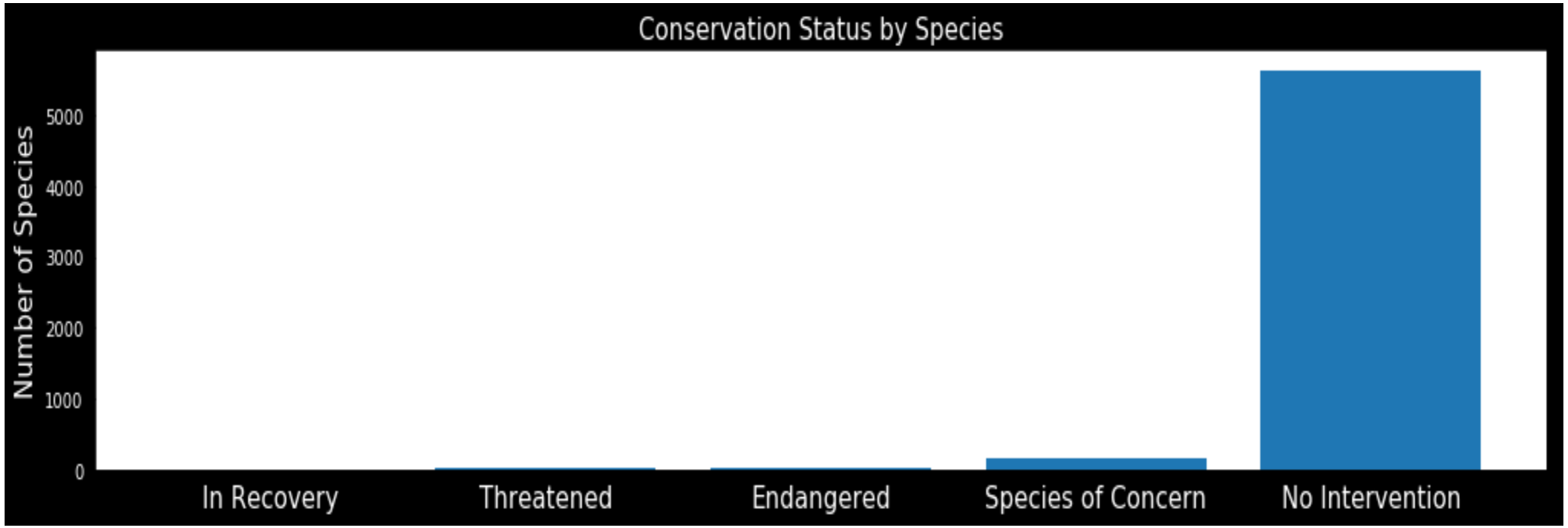
Significance calculations for conservation status

- Firstly, scientific names are counted for each “conservation status”.
- A table below counts how many “scientific name” meet each of “conservation status”(ref. appendix 1).

conservation_status	scientific_name_count
In Recovery	4
Threatened	10
Endangered	16
Species of Concern	161
No Intervention	5,633

Significance calculations for conservation status 2

- As this graph shows, most of the species are “No Intervention”.
  - Then, what kind of species are more likely to be protected?



Significance calculations for conservation status 3

- A table below shows protected percent for each category.
- This shows that “Mammal” is protected most often. Then, Mammal is the most important category? or Bird is also? Is there any significant difference?

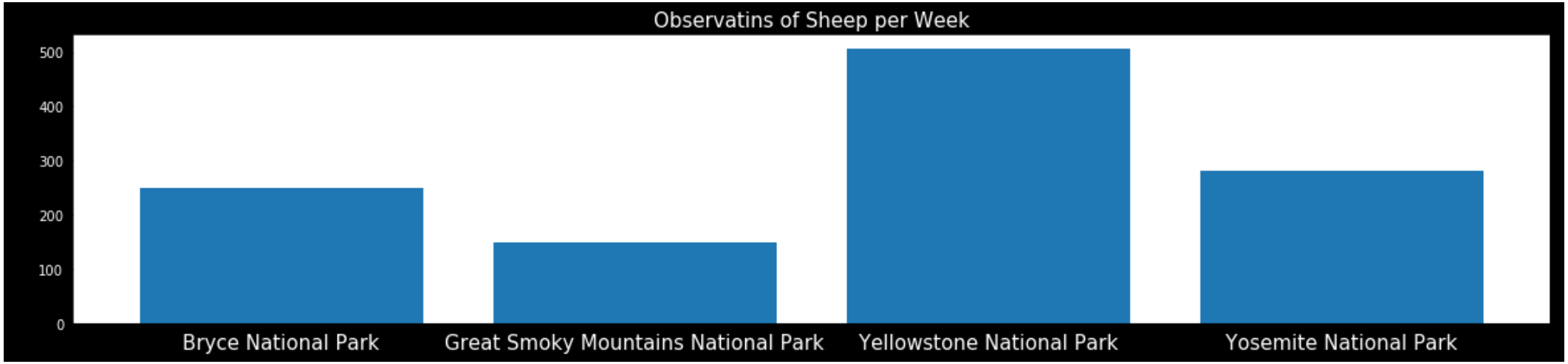
	category	not_protected	protected	percent_protected
0	Amphibian	73	7	8.750000
1	Bird	442	79	15.163148
2	Fish	116	11	8.661417
3	Mammal	176	38	17.757009
4	Nonvascular Plant	328	5	1.501502
5	Reptile	74	5	6.329114
6	Vascular Plant	4424	46	1.029083

A recommendation for conservationists

- In conclusion, Mammal and Bird are protected most often. Conservationists should take steps to make them more safe.
  - The result of chi-square test for the difference between **Mammal**'s protected ratio and **Bird**'s protected ratio shows that the difference is **not significant**(P-value is 0.4459).
  - Another test for the difference between that of **Amphibian** and **Mammal** shows that there might be a **significant** difference(P-value is 0.0842).
  - The third test for the difference between that of **Reptile** and **Mammal** shows that there is a **significant** difference(P-value is 0.0234).

Description of foot and mouth disease study 1

- This graph shows observations of sheep per week for each National Parks.



Description of foot and mouth disease study 2

- Under the condition below, it takes 2 or 3 weeks to take enough sample for A/B test.
  - Background : Park rangers want to reduce the rate of foot and mouth disease at that park.
  - Sample size needed is 520
    - 15% of sheep at Bryce National Park have foot and mouth disease.
      - Baseline disease rate is 15%
    - Scientists want to be able to detect reductions of at least 5 percentage point.
      - Minimum detectable effect is 33% (= 5% / 15% * 100)
    - The level of significance is 90%.
  - How many weeks?
    - Bryce National Park : 520/250 = 2.08
    - Yellowstone National Park : 520/507 = 1.03

Thank you

Appendix

Appendix 1

Explanation of “Conservation Status”

- "Species of Concern"
  - declining or appear to be in need of conservation
- "Threatened“
  - vulnerable to endangerment in the near future
- "Endangered“
  - seriously at risk of extinction
- "In Recovery“
  - formerly "Endangered", but currnetly neither in danger of extinction throughout all or a significant portion of its range
- "No Intervention“
  - Endangered