BIODIVERSITY

Ying-Chih Chiang
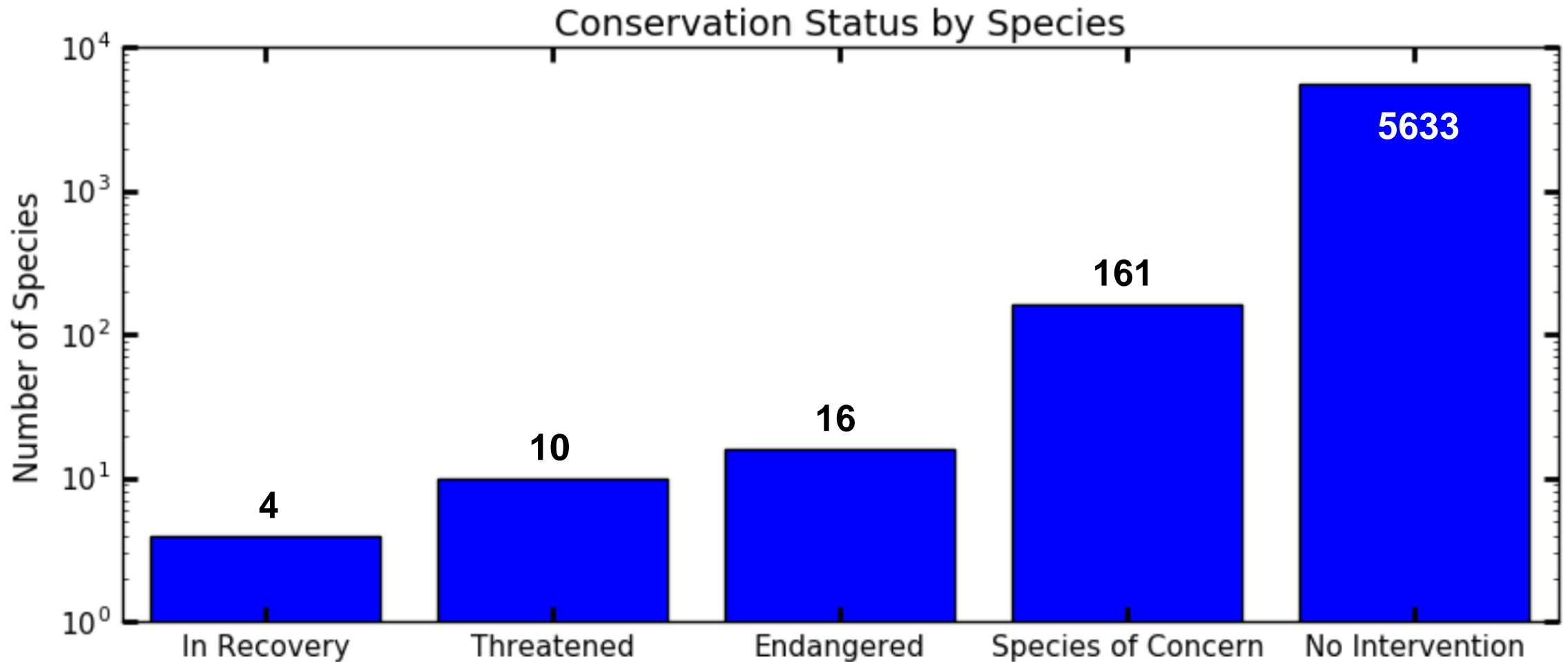
- **Conservation status**
  - ➢ Over 5541 unique species (5824 entries with repetition).
  - ➢ Over 7 different categories.

- **Observation of sheep in four national parks**
  - ➢ Number of sheep observed in each park per week.
  - ➢ Calculate the size of samples and number of weeks required to collect the samples.

CONSERVATION STATUS BY SPECIES

- Most species do not require any attention.
- 16 species are endangered (seriously at risk of extinction).
- 10 species are threatened (vulnerable to endangerment soon).
- 161 species are of concern (declining or in need of conservation).
- 4 species are in recovery (no longer endangered).

Conservation Status by Species

- Bird and Mammal have the highest percentage of species being protected. Their percentage difference is insignificant, i.e. the chi2_contingency test returns a p-value 0.446 (>> 0.05).

- Nonvascular Plant and Vascular Plant have the lowest percentage being protected. Their difference is also not significant, i.e. the chi2_contingency test returns a p-value 0.593 (>> 0.05).

| Category	Not protected	Protected	Percent Protected
Amphibian	73	7	0.087500
Bird	442	79	0.151631
Fish	116	11	0.086614
Mammal	176	38	0.177570
Nonvascular Plant	328	5	0.015015
Reptile	74	5	0.063291
Vascular Plant	4424	46	0.010291

Same level of percent protected

Same level of percent protected

SHEEP OBSERVATION

- The table below shows the number of sheep observed in each national park within a week.

- Yellowstone National Park has more sheep observed in one week, in comparison with other national parks.

- Great Smoky Mountains has least number of sheep observed.

Observations of Sheep per Week
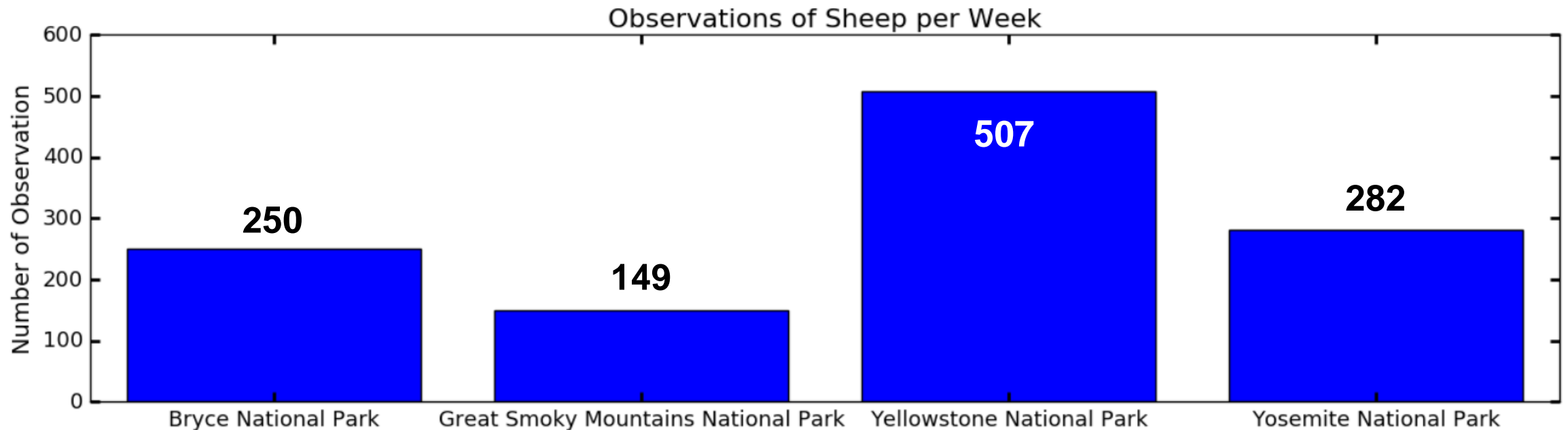
Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease and want to test whether the program is working, i.e. detecting reductions of 5 percentage points. *How long will they need to perform their observation?*

- **Baseline: 15%** of sheep at Bryce National Park have foot and mouth disease. *We assume that the rate of disease in Yellowstone National Park is the same as in Bryce National Park.*

- **Minimum Detectable Effect: 33.33 %.** To be able to detect a 5 percentage point means that we want to see the drop of rate from 15% to 10%, i.e. 1/3 less than before. This gives a minimum detectable effect 33.33%.

- **Level of significance: 90%** (default choice).

Using the sample size calculator at Optimizely, we found that the required number of sheep samples is 510. Given that one can observe 507 sheeps in Yellowstone National Park in one week, the park rangers only need 1.01 weeks to collect sufficient sample of sheep. Interestingly, if other parks adopt the same protocol, they will need a longer period of time to collect sufficient samples, e.g. 2.04 weeks for park rangers at Bryce National Park.

CONCLUSION

- By using data analysis we can better understand how the species are protected, and we can also determine the necessary sample size for monitoring whether the control of the foot mouth disease is successful.

- Although the data analysis skill empowers us to do amazing jobs, **we should pay attention to the _approximation_ made during the analysis.** For instance, **there is no obvious reason why the baseline of disease rate should be identical in two different national parks**, as the sheep from two parks are not living together. Therefore it is likely the baseline disease rate in Yellowstone National Park is completely different from the known rate in Bryce National Park! Consequently, even if we follow the "correct" math to evaluate the sample size, we will still end up with a wrong sample size and make the Park rangers fail their jobs!