



# Biodiversity for the National Parks

DYLAN MOSQUERA

# Species\_info.csv

- ▶ The Species\_info.csv contains information on different species in the national parks, including their Species category, scientific name, common name, and their conservation status. In below table, “nan” means no conservation effort is necessary at this time.

	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

# Species\_info.csv breakdown

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

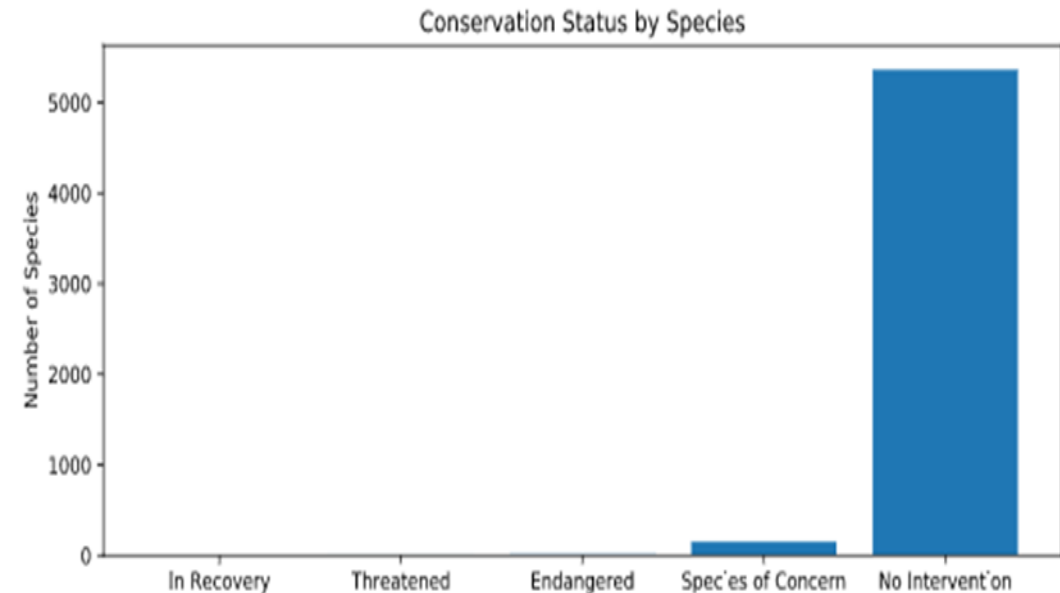
**Endangered** – Seriously at risk of extinction

**In recovery** – Formerly **Endangered**, but out of danger currently

**No intervention** – Is perfectly safe currently

**Species of Concern** – Declining population or appears in need of conservation efforts

**Threatened** – Vulnerable to **endangerment** in near Future



- ▶ As we can see from this bar chart, overwhelmingly in **no intervention** (thankfully!). (Note: updated all nan to **No Intervention**)
- ▶ With the table we can see the exact values of each species
- ▶ Species of Concern are most prevalent protected group, meaning many species populations are declining at concerning rate.

# Further Breakdown: Danger by Species Type

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	0.088608
1	Bird	413	75	0.153689
2	Fish	115	11	0.087302
3	Mammal	146	30	0.170455
4	Nonvascular Plant	328	5	0.015015
5	Reptile	73	5	0.064103
6	Vascular Plant	4216	46	0.010793

- ▶ Above Chart is breakdown of danger to species by Category.
- ▶ Protected contains all conservation statuses but **No Intervention**
- ▶ The higher the percent protected, the more species in that group need protection, indicating they are generally at greater risk as a whole than other categories of being **endangered**, and need more focus to help that category as a whole.
- ▶ In order to determine if a group is more in need of protection than another, we must see if the differences above are a result of random chance.

# Is there Significant difference of need of protection between Category?

- ▶ By utilizing the `chi2_contingency` function from `scipy.stats`, we can analyze if the difference between any 2 categories are significantly different from each other.
- ▶ If we get a  $pval < .05$ , the difference in protect is greater than random chance.
- ▶ Pvalues:
  - ▶ Bird(~15%) vs Mammal(~17%): 0.696 : Not significant
  - ▶ Reptile(~6.4%) vs Mammal(~17%): 0.038: Significant Difference
- ▶ We have shown that some categories have significant difference in risk of being **endangered** than others, meaning some groups are at greater risk than others for reasons other than random chance.

Note: Percent next to category is percent protected

# Recommendation to Conservationists

- ▶ Based on the previous calculations, certain categories of species are in greater danger than others.
  - ▶ My recommendation to conservationist is to focus their efforts mainly on combating Mammal and Bird population decline, using protection and repopulation efforts, as these are groups seeing the greatest risk of **extinction**.
  - ▶ The next group of animals need moderate monitoring and intervention: Amphibian, Fish, Reptile
  - ▶ Plants, be they Nonvascular or Vascular, have the lowest rates of endangerment in the national parks.
- ▶ For all these categories, it is obviously important to not ignore the endangered species on the lower risk groups. They are only at less risk as a whole
- ▶ Conservationist should look at possible root causes that make Mammals and Birds more in danger, such as cross-species diseases or human interference with their habitats.



# Foot and Mouth Disease: Sample Size Determination

- ▶ An outbreak of Foot And Mouth Disease is effecting the sheep population of the National Parks, and we need to observe the sheep to see at what rate we are seeing the disease and if Conservation efforts are helping stop infections
- ▶ Scientist need to know how long they need to go out to observe the Sheep populations to get reliable data on if reduction efforts are working. They would like to be able to measure with confident a reduction of 5 percentage points of infection.
  - ▶ Variables:
    - ▶ Baseline = 15% (Percent of Sheep last year with Foot and Mouth Disease)
    - ▶ Minimum detectable Effect =  $100 * \text{Reduction} / \text{baseline} = 100 * 5 / 15 = 33.33\%$
    - ▶ Statistical Significance = 90%

# Sample Size

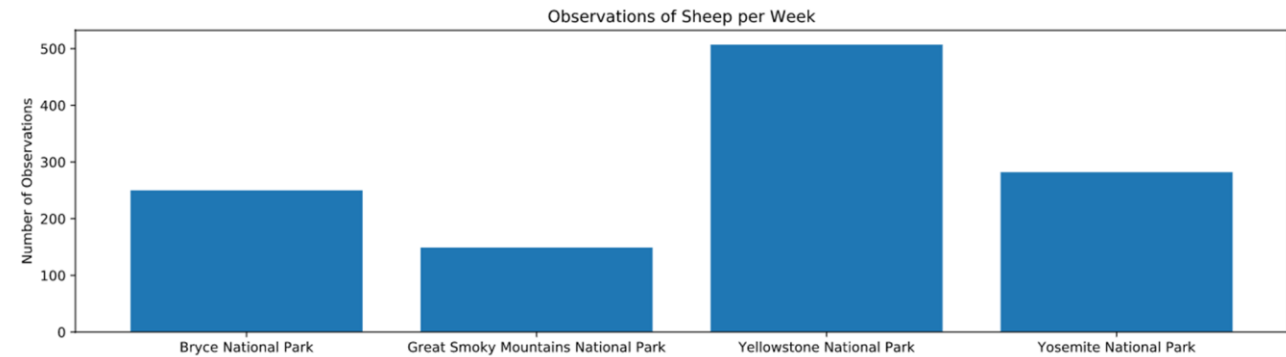
- ▶ Plugged all the previous variables into this calculator
- ▶ Got Sample Size of 870
- ▶ This means in order to be confident that our observations are correctly 5 percentage point lower, we need to observe 870 sheep.

Baseline conversion rate: 15 %  
Statistical significance: 85% 90% 95%  
Minimum detectable effect: 33.33 %  
Sample size: 870



# Trial Durations

	park_name	observations
0	Bryce National Park	250
1	Great Smoky Mountains National Park	149
2	Yellowstone National Park	507
3	Yosemite National Park	282



- ▶ Above table and chart is the number of sheep observations in each National Park per week.
- ▶ In order to find how long scientist need to be confident in their results, need to divide sample size by the number of sheep observed at each park.
  - ▶ Bryce:  $870/250 = 3.48$  weeks
  - ▶ Great Smoky Mountains:  $870/149 = 5.84$  weeks
  - ▶ Yellowstone:  $870/507 = 1.72$  weeks
  - ▶ Yosemite:  $870/282 = 3.09$  weeks

# Foot and Mouth Final Thoughts

- ▶ After scientists observe for the allotted period in their parks, if they get a 5% difference from last year, they can be confident those results are accurate with how many sheep they observed, and not a result of taking too few observations and accidentally getting a non representative pool of sheep.