



# BIODIVERSITY FOR THE NATIONAL PARKS

Capstone Project 2

04/09/2018

Le Tran

# OBSERVATIONS FROM DATA

- Used Pandas to load species\_info.csv
- Observations from species\_info.csv:
  - Data includes a list of species that were observed during a 7-day period
  - Each species include a scientific name, the type of species it is (category), and its conservation status

	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
5	Mammal	Odocoileus virginianus	White-Tailed Deer	nan

# OBSERVATIONS FROM DATA

- Calculations:

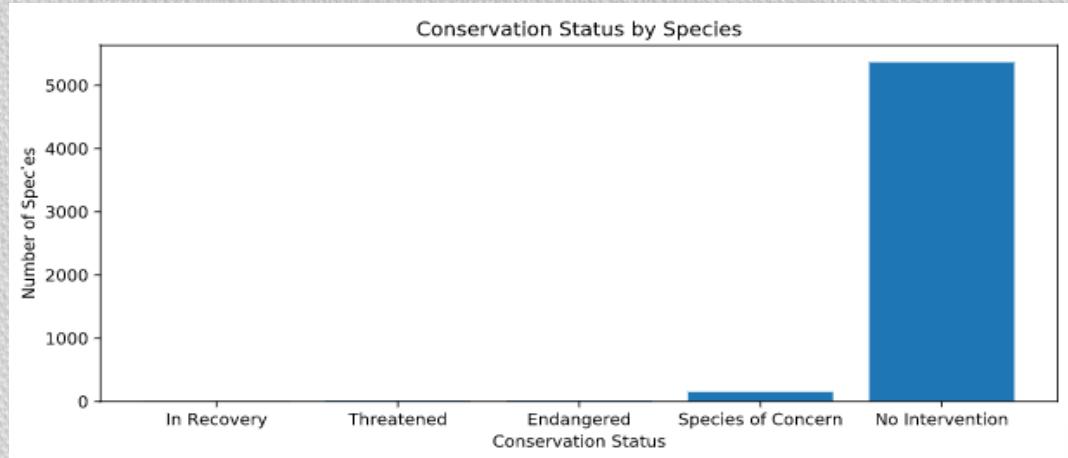
- Using different commands to count the various values of species and types of species and found...
- There were 5,541 different species from the data
- The species came from 7 different types of species (categories)
  - Mammal, bird, reptile, amphibian, fish, vascular plant, nonvascular plant

```
8
9  species_count = species.scientific_name.nunique()
10 print(species_count)
11
12 species_type = species.category.unique()
13 print(species_type)
14
```

- There are 5 different conservation statuses in the data. “No Intervention” was added to fields that were null to make the data more uses.
  - Endangered, In Recovery, No Intervention, Species of Concern, Threatened

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

# OBSERVATIONS FROM DATA



- There are 151 species that are of concern due to a decreasing population. Other species requiring protection were minimal. Most species do not require any sort of protection (5,363)
  - We further delved into the number of different types of species that require and do not require protection
  - Even though birds, vascular plant and mammals have higher number of species protected, the data does not tell us much about endangered species
- |   | 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          |

# STATISTICAL ANALYSIS

- Chi-Squared test was used to test for significance
  - Null hypothesis: difference is due to chance
  - Contingency table was used to test the difference between mammals and other types of species

```
38 contingency = [[30, 146], [75, 413]]
39 chi2, pval, dof, expected = chi2_contingency(contingency)
40 print pval
41
42 contingency_rep = [[30, 146], [5, 73]]
43 chi2, pval, dof, expected = chi2_contingency(contingency_rep)
44 print pval
45 pval_reptile_mammal = 0.0383555902297
46
47 contingency_fish = [[30,146], [11, 115]]
48 chi2, pval, dof, expected = chi2_contingency(contingency_fish)
49 print pval
50
51 contingency_amph = [[30,146], [7, 72]]
52 chi2, pval, dof, expected = chi2_contingency(contingency_amph)
53 print pval
```

- If  $p\text{-val} < 0.05$ , we are reasonably confident that the two datasets are significantly different and would therefore reject the null hypothesis

- Results:

0.687594809666	Mammal & Bird
0.0383555902297	Mammal & Reptile
0.0561483484489	Mammal & Fish
0.127576696932	Mammal & amphibian

- No significant difference between mammals and birds, mammals and fish, and mammals and amphibians. They were by chance.
- There is a significant difference between mammals and reptiles.

# RECOMMENDATION FOR CONSERVATIONISTS

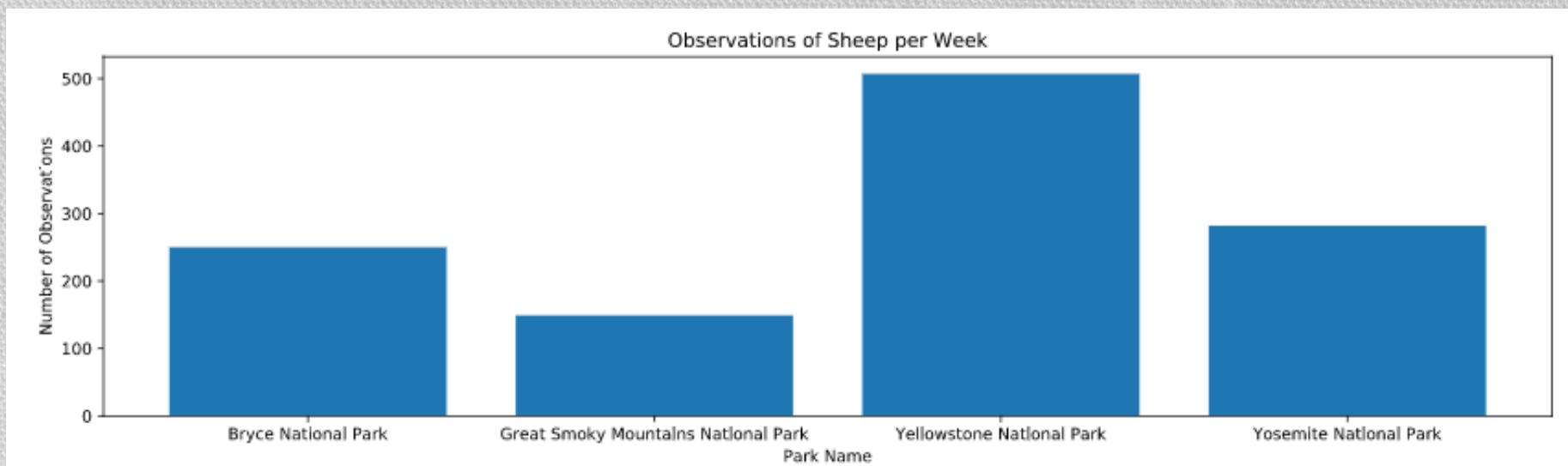
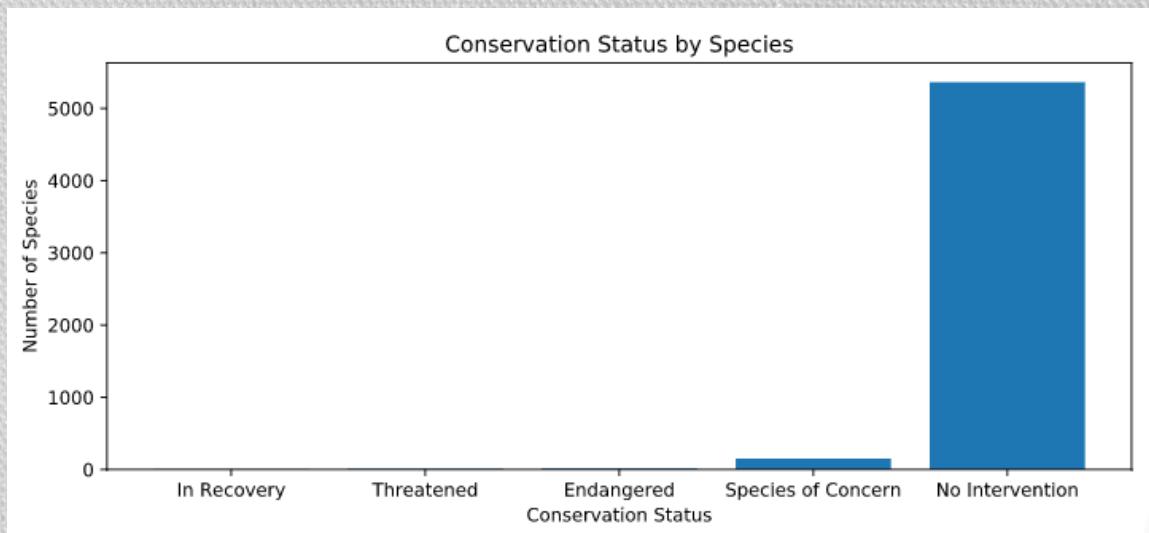
- Based on the significance test calculations performed, some types of species are more likely than other types of species to be endangered.

# FOOT & MOUTH DISEASE STUDY

- Determining Sample Size

- To determine the sample size using the sample size calculator, we first had to determine the baseline conversion rate, the level of significance we want, and the minimum detectable effect.
- We chose a baseline conversion rate of 15%. Which was the percentage of sheep at Bryce National Park that had foot and mouth disease the previous year.
- The test we were doing wanted a 90% level of significance
- To calculate the minimum detectable effect, we divided 5% by the baseline 15% and multiplied by 100. The 5% was the reduction amount that the park rangers wanted to detect. Minimum detectable effect = 33%
- Using the numbers we have, plugging the results into the sample size calculator will give us a sample size of 890.

# GRAPHS CREATED IN CAPSTONE 2 PROJECT





# THE END

Capstone Project 2

04/09/2018

Le Tran