

INTRODUCTION TO DATA ANALYSIS

Biodiversity for the National Parks

PRESENTATION BY OSCAR INOWE

“DESCRIPTION”

- Analyze data for the National Park Service.
- Data from endangered species from distinct parks.
- Evaluate the conservation status of the species.
- Investigate for patterns.

species_info.csv

Use Pandas to read the .csv.

Learn about the different species, their categories and conservation status.

Use functions such as groupby to determine the species that match certain criteria.

Create charts to represent this information.

Determine which are the species that are more likely to be endangered by pivoting the table and getting percentages.

Run chi square tests to determine if there are any significant differences between species.

biodiversity.csv

Determine the species that refer to sheep.

Apply lambda functions to investigate which are the species that are sheep, adding a new column to the table.

Merge this table with observations previously had.

Determine the number of observations per park.

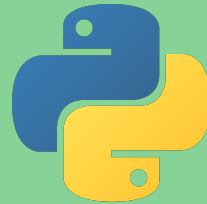
Plot this information and determine the weeks needed to observe sheep at Bryce National Park and Yellowstone National Park in order to determine if the program that is currently running to reduce the rate of foot and mouth diseases works.

SOFTWARE

Tools used



*PyCharm
Community
Edition*



*Python
Version
2.7*



*Matplotlib
Pandas
Chi Contingency*

species_info.csv

Total Species: 5541

```
groupby('category').category.count()
```

Category	Count
Amphibian	80
Bird	521
Fish	127
Mammal	214
Nonvascular Plant	333
Reptile	79
Vascular Plant	4470

Values of conservation status

- Species of Concern
- Endangered
- Threatener
- In Recovery

Conservation Status	Scientific Name Count
In Recovery	4
Threatened	10
Endangered	16
Species of Concern	161
No Intervention	5633

Hyphotesis

Category	Not Protected	Protected	Percent
Amphibian	72	7	8.86%
Bird	413	75	15.3%
Fish	115	11	8.73%
Mammal	146	30	17.04%
Nonvascular Plant	328	5	1.5%
Reptile	73	5	6.41%
Vascular Plant	4216	46	1.07%

*“Conservation actions should be into **Mammals** & **Birds** taking in consideration the results shown on the table and the Chi2 test.*

Reptiles are not in the need of these actions in comparison to these other species”

First Chi2 Test

Not significant > 0.05
p-value: 0.687594809666

“Cannot reject null hypothesis”

Mammal vs. Bird

Second Chi2 Test

Significant < 0.05
p-value: 0.038355590229

“Can reject null hypothesis”

Mammal vs. Reptile

observations.csv

Determine which of the species are “sheep”
Break into the different sightings of the parks
Determine the weeks needed to observe the
number of sheeps with foot & mouth disease

```
species['is_sheep'] = species.common_names.apply(lambda x: 'Sheep' in x)
sheep_species = species[(species.is_sheep) & (species.category == 'Mammal')]
sheep_observations = observations.merge(sheep_species)

obs_by_park =
sheep_observations.groupby('park_name').observations.sum().reset_index()
```

Common Name	Status	Park Name	Scientific Name Count
Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	Bryce National Park	250
		Great Smoky Mountains National Park	149
Bighorn Sheep, Bighorn Sheep	Species of Concern	Yellowstone National Park	507
Sierra Nevada Bighorn Sheep	Endangere	Yosemite National Park	282

Baseline: 15%
Minimum Detectable Effect: 33.33%

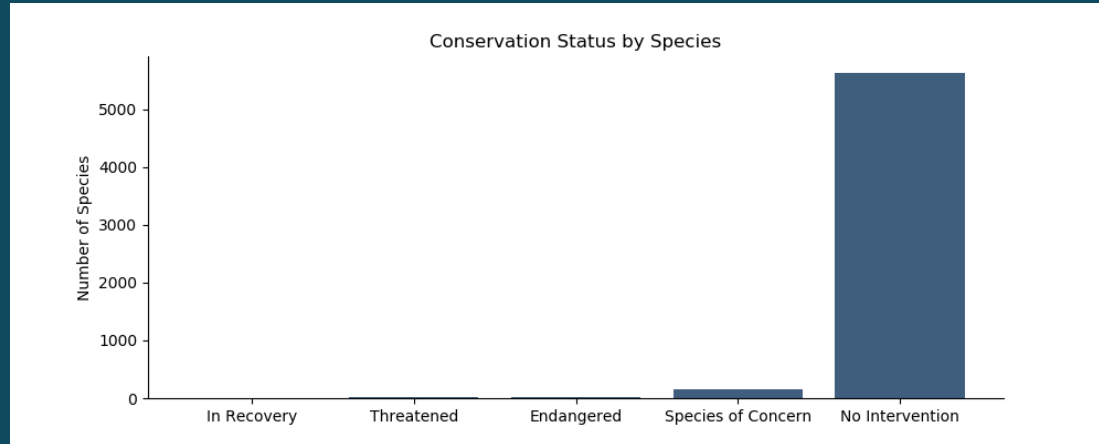
Sample Size per Variation: 510

Bryce: 2 weeks (510 / 250)

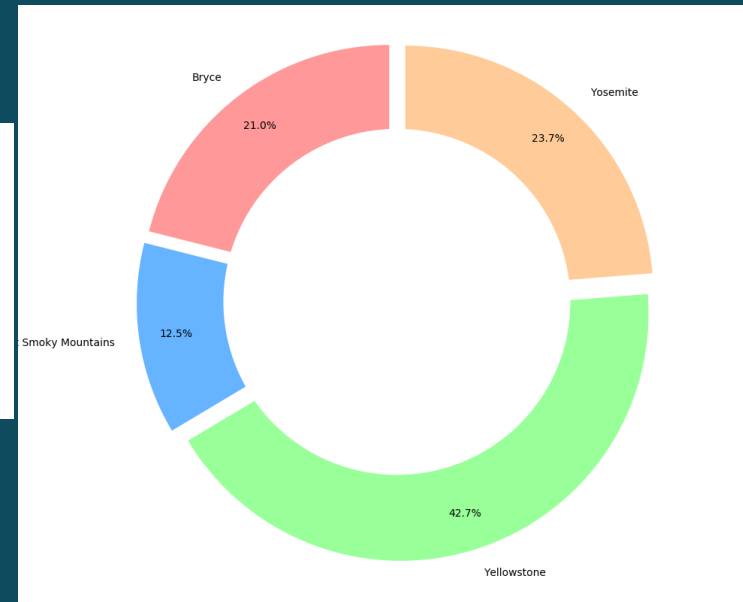
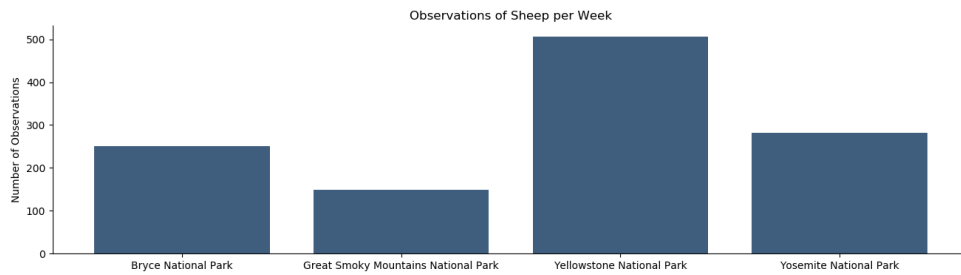
Yellowstone: 1 week (510 / 507)

Charts

species_info



observations



Thank you for
your
attention.

The End