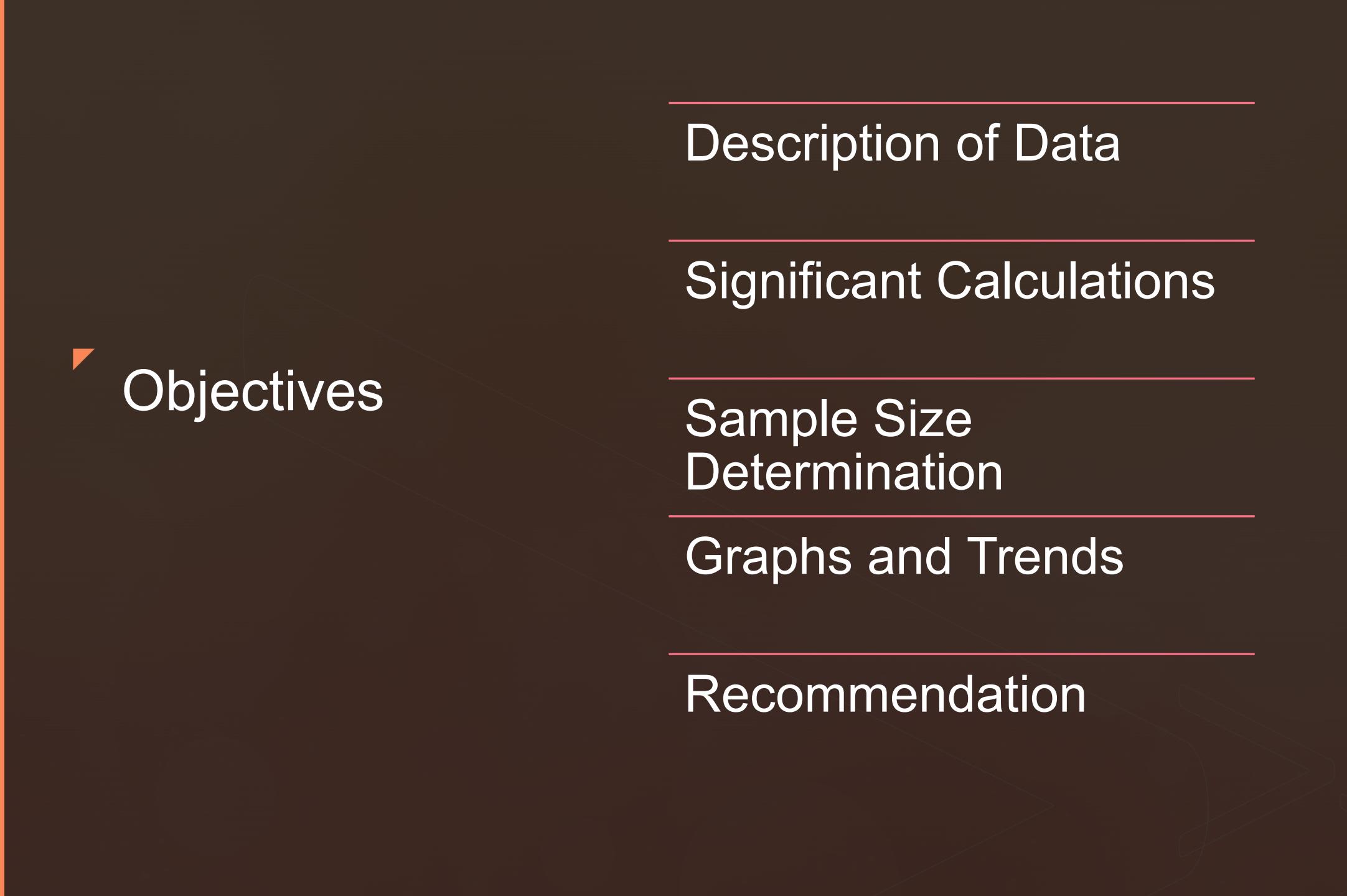


Presented by: Mahmoud Shihab

Codecademy Capstone: National Park Services Biodiversity Analysis



Objectives

Description of Data

Significant Calculations

Sample Size
Determination

Graphs and Trends

Recommendation

The Data: What's in it? (species_info.csv)

- 4 Columns with 5,824 entries
- category: [String: 7 options]
Gives the animal class of animal in question:
Amphibian, Bird, Fish,
Mammal, Nonvascular Plant,
Reptile, and Vascular Plant
- scientific_name: [String]
Gives the scientific name of an animal species
- common_names:[String]
The name most people know the animal by
- conservation_status: [String: 5 options]
States the conservation state of the animals: Blank,
Endangered, In Recovery,
Species of Concern,
Threatened

The Data: What's in it? First 10 rows: (species_info.csv)

category	scientific_name	common_names	conservation_status
Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	
Mammal	Bos bison	American Bison, Bison	
Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	
Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	
Mammal	Cervus elaphus	Wapiti Or Elk	
Mammal	Odocoileus virginianus	White-Tailed Deer	
Mammal	Sus scrofa	Feral Hog, Wild Pig	
Mammal	Canis latrans	Coyote	Species of Concern
Mammal	Canis lupus	Gray Wolf	Endangered

The Data: What's in it? (observations.csv)

- 3 Columns with 23,295 entries
- scientific_name: [String]
Gives the scientific name of an animal species
- observation:[INT]
The number of observations of the animal per park
- park_name: [String: 4 options]
States the name of the park where the observation took place: Bryce National Park, Great Smoky Mountains National park, Yellowstone National park, Yosemite National Park

The Data: What's in it? First 10 rows: (observations.csv)

scientific_name	park_name	observations
<i>Vicia benghalensis</i>	Great Smoky Mountains National Park	68
<i>Neovison vison</i>	Great Smoky Mountains National Park	77
<i>Prunus subcordata</i>	Yosemite National Park	138
<i>Abutilon theophrasti</i>	Bryce National Park	84
<i>Githopsis specularioides</i>	Great Smoky Mountains National Park	85
<i>Elymus virginicus</i> var. <i>virginicus</i>	Yosemite National Park	112
<i>Spizella pusilla</i>	Yellowstone National Park	228
<i>Elymus multisetus</i>	Great Smoky Mountains National Park	39
<i>Lysimachia quadrifolia</i>	Yosemite National Park	168

Significant Calculations

- The following picture shows the significance of “Is species x more likely to be endangered than species y?” from most significant to least.

```
Pval (    Bird vs Vascular Plant    ): 4.6123e-79. Significant
Pval (    Mammal vs Vascular Plant  ): 1.4405e-55. Significant
Pval (    Fish vs Vascular Plant    ): 1.4868e-12. Significant
Pval (    Bird vs Nonvascular Plant ): 1.0546e-10. Significant
Pval (    Mammal vs Nonvascular Plant): 1.4819e-10. Significant
Pval (    Amphibian vs Vascular Plant): 1.0441e-08. Significant
Pval (    Reptile vs Vascular Plant ): 1.4505e-04. Significant
Pval (    Fish vs Nonvascular Plant ): 4.9606e-04. Significant
Pval (Amphibian vs Nonvascular Plant): 1.7786e-03. Significant
Pval (Nonvascular Plant vs Reptile  ): 3.3627e-02. Significant
Pval (      Mammal vs Reptile      ): 3.8356e-02. Significant
```

Sample Size Determination

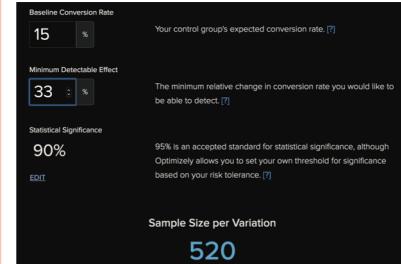
- Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. They want to be able to detect reductions of at least 5 percentage point:

- $$MDE = 100 * \frac{\text{Decitible Perct Point}}{\text{Baseline}}$$

$$MDE = 100 * \frac{5}{15} = 33.33\%$$

- Recommendation per Park: Number of weeks the parks would need to observe sheep in order to observe enough sheep for "Foot and Mouth Disease"

Baseline: 15.00%
MDE: 33.33%
Significance: 90.00%
Sample Size: 520

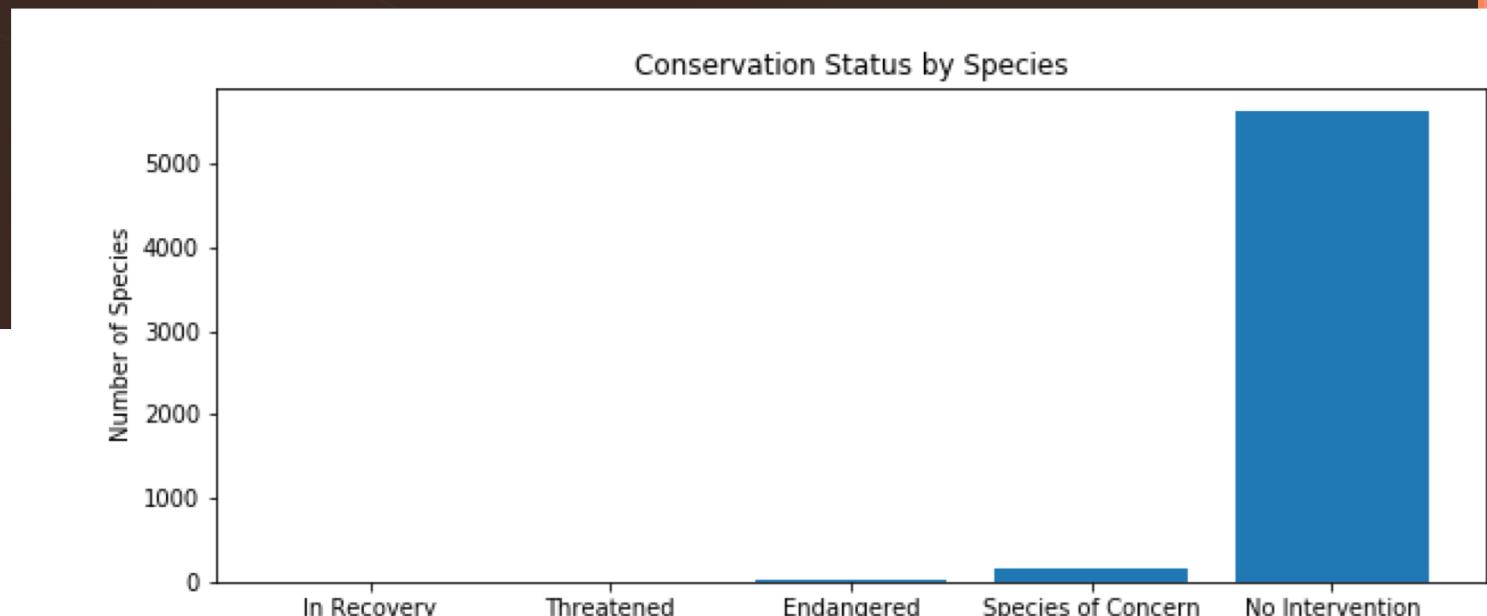


Bryce National Park: 2.08 weeks
Great Smoky Mountains National Park: 3.49 weeks
Yellowstone National Park: 1.03 weeks
Yosemite National Park: 1.84 weeks

Graphs and Trends

- It's clear that a lot of animals need no intervention, but there are still a big portion that are of concern.
- Recommendation: Some species are in 'Threatened' or "Endangered". Focusing on getting these animals into Recovery would be a good start.

	conservation_status	scientific_name
1	In Recovery	4
4	Threatened	10
0	Endangered	16
3	Species of Concern	161
2	No Intervention	5633



Graphs and Trends

- Recommendation: Link the parks together to avoid ‘conflicting data’
- Yellowstone seems to be a park with the most observations, which can sometimes lead to conflicting observations:

```
Canis Lupus: Gray Wolf      - Endangered
Canis Lupus: Gray Wolf, Wolf - In Recovery
Canis Lupus: Gray Wolf, Wolf - Endangered

Onchorhynchus mykiss: Rainbow Trout - Blank
Onchorhynchus mykiss: Rainbow Trout - Threatened
```

	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

Observations of Sheep per Week

