

# Biodiversity for the National Parks

Capstone Project for Introduction to Data Analysis

*Vina Noviana Limanto*

# Overview of presentation

- **Part 1: Overview of species dataframe**
- **Part 2: Investigating endangered species**
  - Count of species by conservation status
  - Differences in conservation status by species category
  - Testing for significance in differences in conservation status by category
  - Recommendations
- **Part 3: Foot and mouth disease reduction effort**
  - Sample size determination
  - Recommendations

## PART 1: OVERVIEW OF SPECIES DATAFRAME

# Inspecting the species dataframe

### Overview of species dataframe:

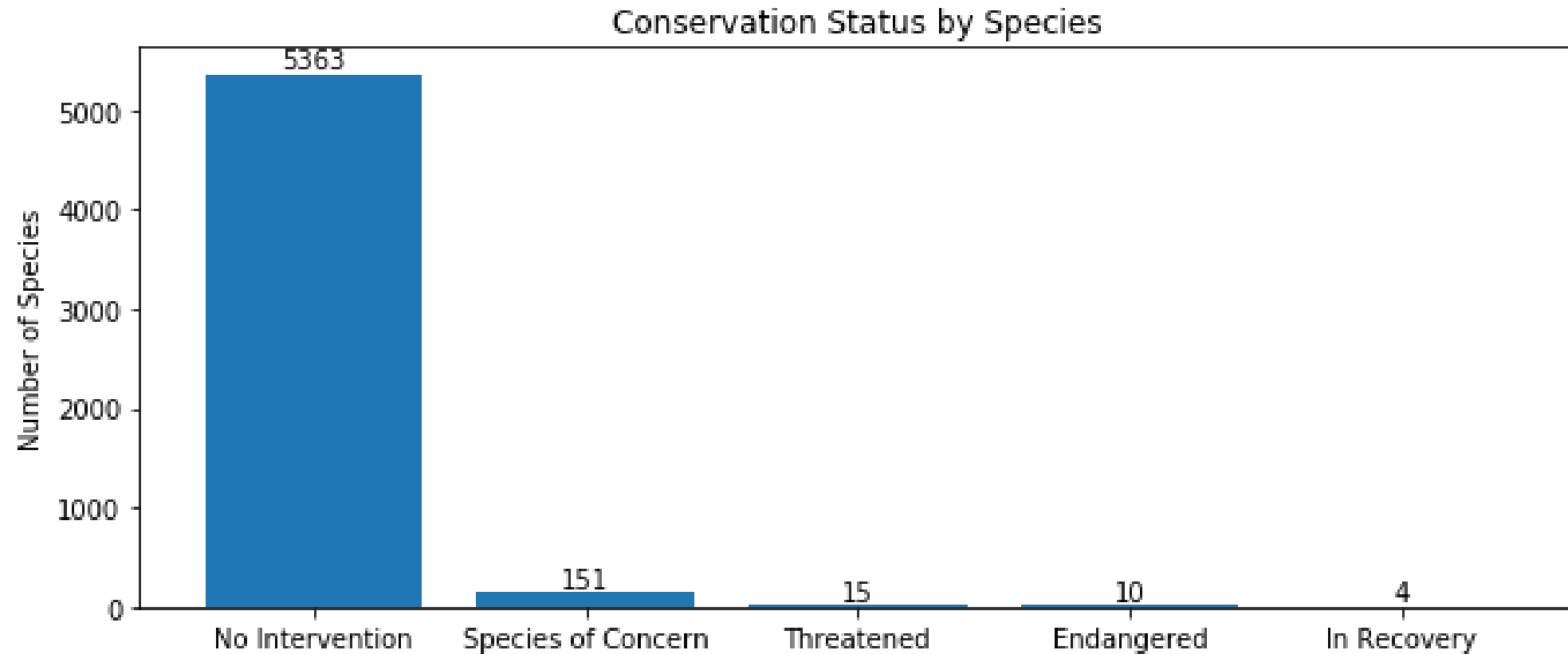
	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon,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

### The species dataframe contains:

- 5541 unique species
- 7 species categories, incl. Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant
- 5 conservation statuses, incl. Species of Concern, Endangered, Threatened, In Recovery, and no status (i.e. no intervention/ protection required)

## PART 2: INVESTIGATING ENDANGERED SPECIES

# Count of species by conservation status



# Conservation status by species category

Category	not_protected	protected	percent_protected
Mammal	146	30	0.170455
Bird	413	75	0.153689
Amphibian	72	7	0.088608
Fish	115	11	0.087302
Reptile	73	5	0.064103
Nonvascular Plant	328	5	0.015015
Vascular Plant	4216	46	0.010793

- There are differences in the percentage of protected species across categories
- However, we need to test if these observed differences are significant, i.e. not due to chance
- Chi-squared test is most suitable to test significance, given that we are dealing with categorical data and more than one set of data

## PART 2: INVESTINGATING ENDANGERED SPECIES

# Conservation status by species category

Chi-Squared pval	Mammal	Bird	Amphibian	Fish	Reptile	NVC Plant	VC Plant
Mammal	N/A	0.45	0.09	0.03	0.02	<0.01	<0.01
Bird	0.45	N/A	0.18	0.08	0.05	<0.01	<0.01
Amphibian	0.09	0.18	N/A	0.82	0.78	<0.01	<0.01
Fish	0.03	0.08	0.82	N/A	0.74	<0.01	<0.01
Reptile	0.02	0.05	0.78	0.74	N/A	0.03	<0.01
NVC Plant	<0.01	<0.01	<0.01	<0.01	0.03	N/A	0.66
VC Plant	<0.01	<0.01	<0.01	<0.01	<0.01	0.66	N/A

Legend:

pval < 0.05

pval > 0.05

### Key observations:

- 1 Animals are more likely to be endangered than Plants
- 2 Mammals are more likely to be endangered than Fish and Reptile

# Recommendations

- As animals are more likely to be endangered than plants, conservation efforts (including preventive efforts) need to be focused on animals rather than plants
- Conservation and preventive efforts need to be further focused on mammals, which are more likely to be endangered than other animal categories such as fish and reptiles

## PART 3: FOOT & MOUTH DISEASE REDUCTION EFFORT

# Sample size determination

Sample size of 510 is needed in each park, based on the following assumptions:

- Baseline conversion rate of 15%, based on previous year's record of foot and mouth disease occurrence
- Minimum detectable effect of 33.3%, given that scientists aim to be able to detect reductions of at least 5%, which is 33.3% of the 15% baseline rate
- Statistical significance of 90%

Baseline Conversion Rate

15 %

Your control group's expected conversion rate. [\[?\]](#)

Minimum Detectable Effect

33.3 %

The minimum relative change in conversion rate you would like to be able to detect. [\[?\]](#)

Statistical Significance

90%

[EDIT](#)

95% is an accepted standard for statistical significance, although Optimizely allows you to set your own threshold for significance based on your risk tolerance. [\[?\]](#)

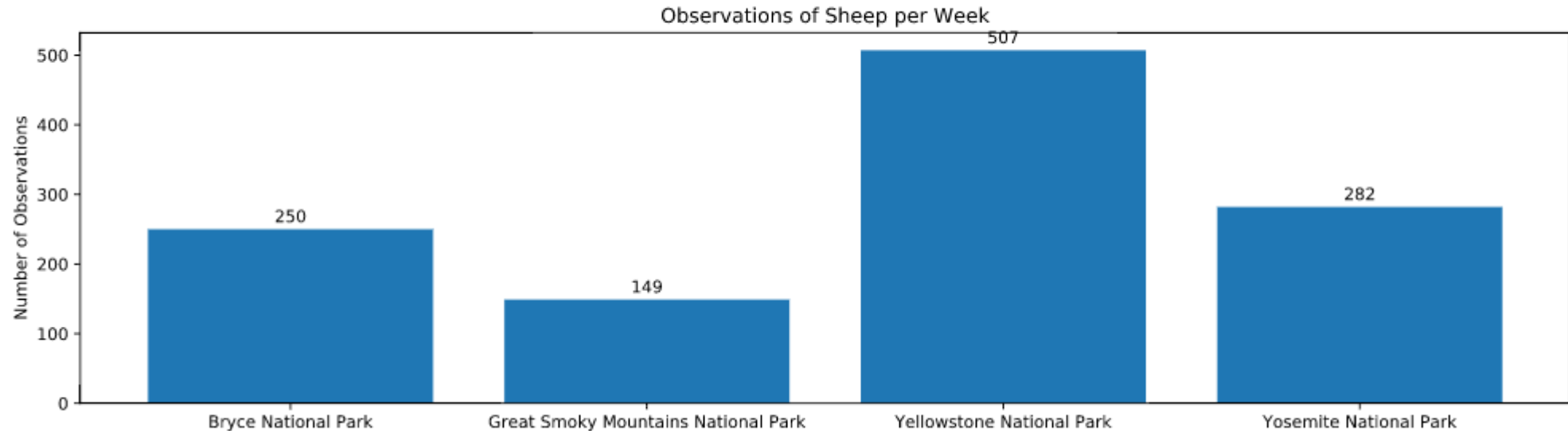
Sample Size per Variation

510



## PART 3: FOOT & MOUTH DISEASE REDUCTION EFFORT

# Recommendations



Est. weeks:

2.04

3.42

1.00

1.81

- To ensure that a >5% drop in observed cases of foot and mouth disease is significant, National Parks scientists need to observe at least 510 sheep in each park
- Based on observations of sheep per week in each park, this is estimated to take 1 – 4 weeks