Reviewable project – Capstone Option 2

Biodiversity in American National Parks

Data analyst: Marius Razem

Submission Deadline: 27th November 2018

codecademy

Data Analysis

Guiding Questions

Conservation Analysis

→ What is the conservation status of the species living in the national parks?

→ Are there some specimen especially being endangered?

→ Which differences between each specimen can be observed?

Observation Analysis – Focus: Sheep Populations

→ How many sheep populations have been observed in all park areas?

→ How many weeks do sheep have to be observed to clarify the question if foot and mouth diseases have successfully been reduced during the disease reduction program?

Data Structure

- Given DataFrame Structure based on **species_info.csv**:

Index / number	category	scientific_name	common_names	conservation_status
…	…	…	…	…

Total distinct specimen: 5541

Animal Categories:

- Mammal
- Bird
- Reptile
- Amphibian
- Fish
- Plant (Vascular / Nonvascular)

Possible conservation status / state:

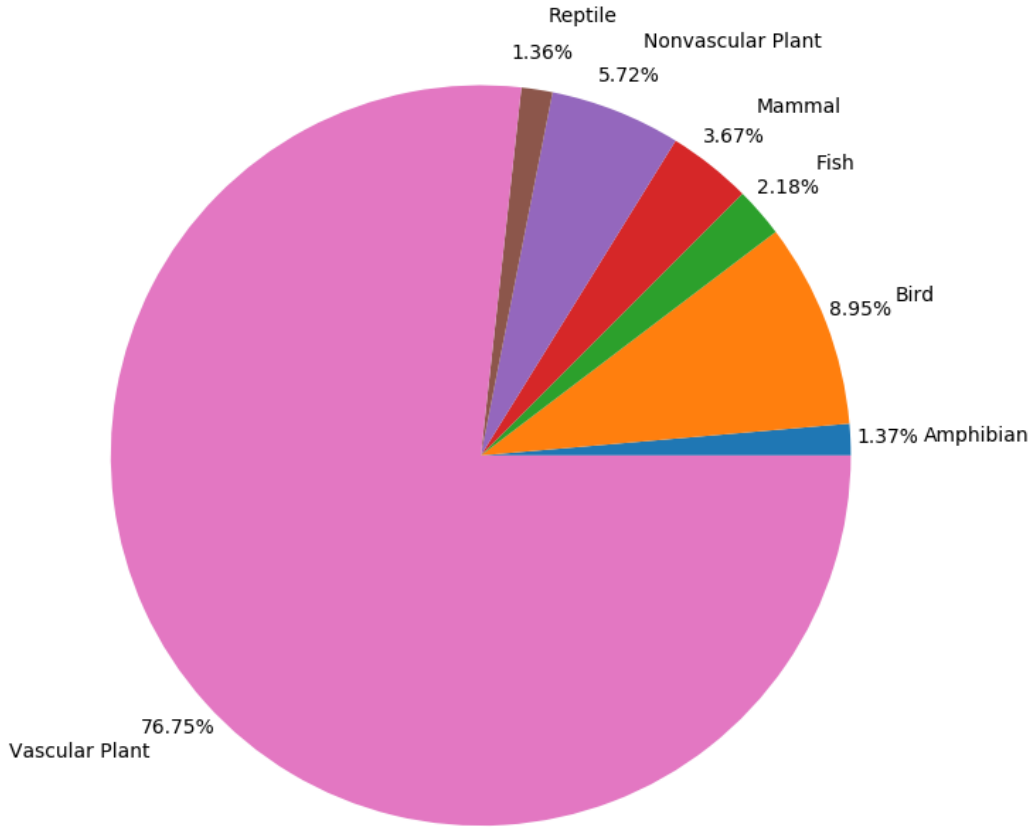
a) Species of Concern
b) Threatened
c) Endangered
d) In Recovery
e) No Intervention (NaN)

Data Analysis – Conservation Analysis

Animal Distribution

– Basic information about existing animals in all national parks:

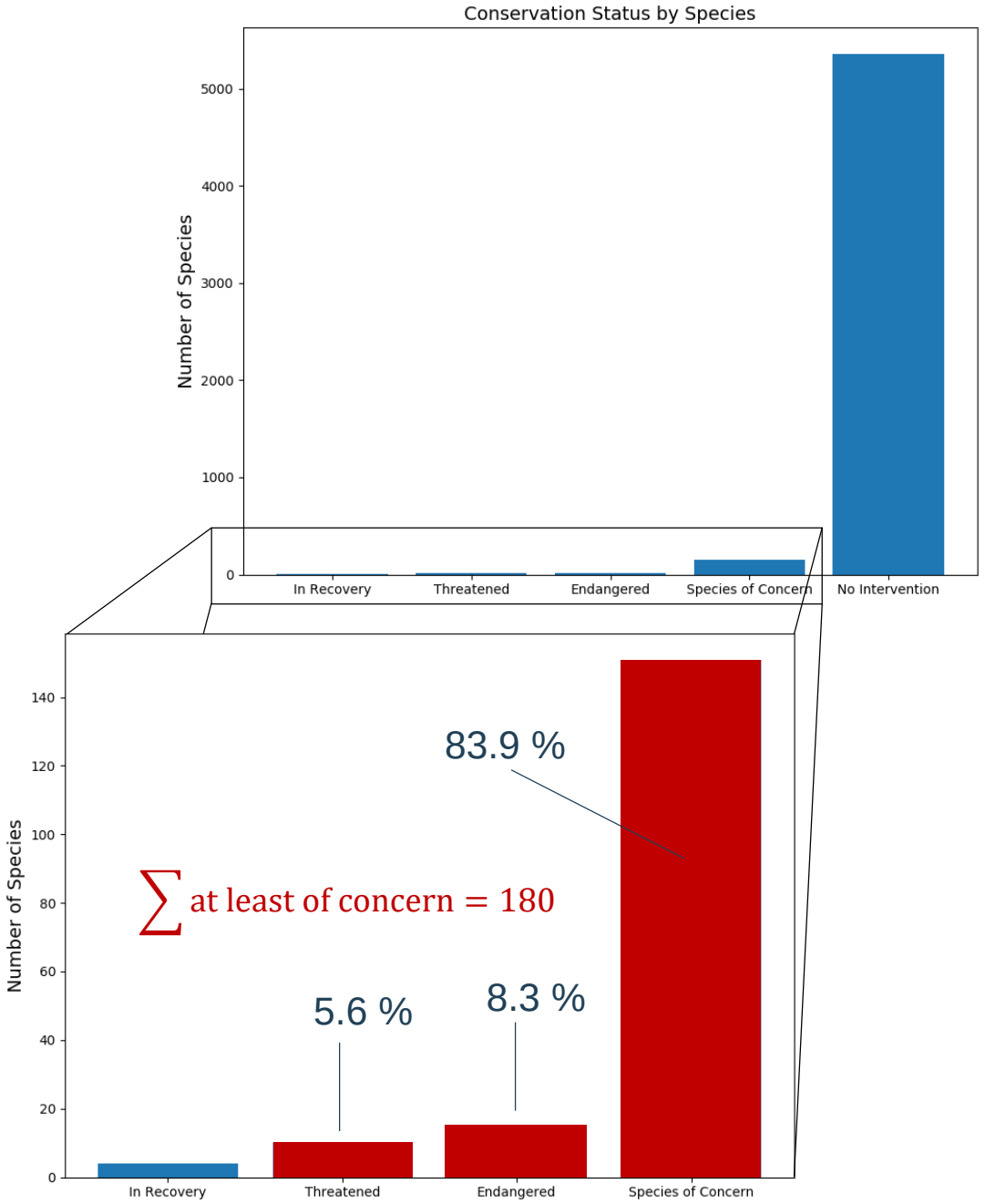
Category	Number of specimen (all parks)	% of total
Amphibian	80	1,37 %
Bird	521	8,95 %
Fish	127	2,18 %
Mammal	214	3,67 %
Nonvascular Plant	333	5,72 %
Reptile	79	1,36 %
Vascular Plant	4470	76,75 %



Conversation Status

Conversation Status	Number of specimen (% of total number)
Species of Concern	151 (2.72)
Threatened	10 (0.18)
Endangered	15 (0.27)
In Recovery	4 (0.07)
No Intervention	5363 (96.75)

- about 96.75 % of all specimen do not require any intervention
- *only* about 0.07 % of all specimen are in recovery so far
- 3.18 % of all specimen are *at least species of concern* (180 specimen)
- considering those 180 specimen, 83.9 % are of concern, 5.6 % are at least threatened and 8.3 % are endangered

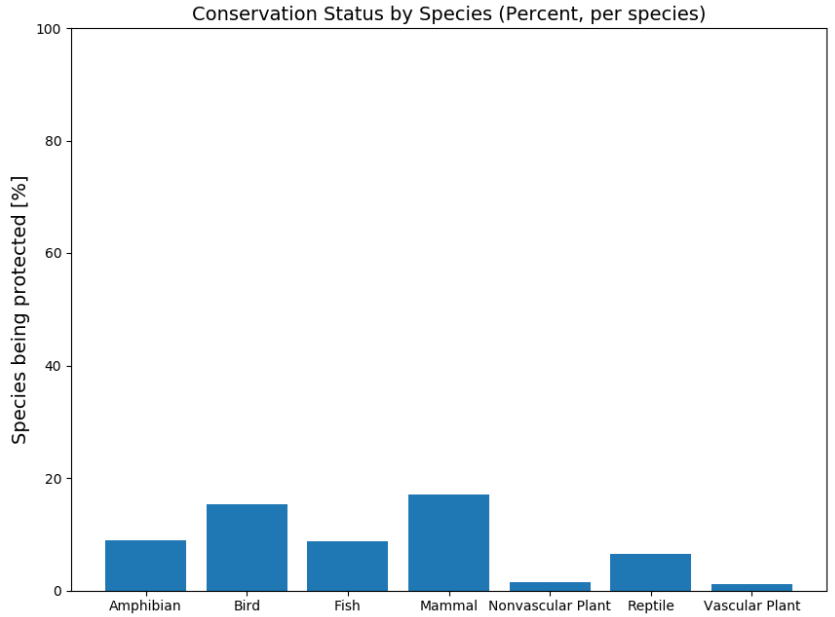


Protected Specimen and Endangerment Potentials

Category	Protected	Not Protected	Protected [%]
Amphibian	7	72	8.86
Bird	75	413	15.37
Fish	11	115	8.73
Mammal	30	146	17.04
Nonvascular Plant	5	328	1.5
Reptile	5	73	6.4
Vascular Plant	46	4216	1.07

Specimen requiring special attention

- Birds and mammals show endangerment potentials with protection percentages > 15 %
- Amphibians, fishes and reptiles are characterized by protection levels between 5 and 10 %
- Plants (no matter if vascular or nonvascular) show very small protection percentages of < 1.5 %



Significance Testing between each specimen (I/III)

- Two specimen are compared per test
- Assumption of the following hypotheses (*categorical chi-squared test*):

H_0: *There is <u>no</u> significant (endangerment potential) difference between the two specimen.*

H_1: *There is a significant (endangerment potential) difference between the two specimen.*

Data Analysis – Conservation Analysis

Significance Testing between each specimen (II/III)

– Comparison matrix (general):

Specimen with higher protection

Specimen with lower protection

	Amphibian	Bird	Fish	Mammal	Nonvascular Plant	Reptile	Vascular Plant
Amphibian	---	n.d. (17.59 %)	n.d. (82.48 %)	n.d. (12.76 %)	s.d. (0.18 %)	n.d. (78.15 %)	s.d. (0.00 %)
Bird	n.d. (17.59 %)	---	n.d. (7.67 %)	n.d. (68.76 %)	s.d. (0.00 %)	n.d. (5.31 %)	s.d. (0.00 %)
Fish	n.d. (82.48 %)	n.d. (7.67 %)	---	n.d. (5.61 %)	s.d. (0.05 %)	n.d. (74.07 %)	s.d. (0.00 %)
Mammal	n.d. (12.76 %)	n.d. (68.76 %)	n.d. (5.61 %)	---	s.d. (0.00 %)	s.d. (3.84 %)	s.d. (0.00 %)
Nonvascular Plant	s.d. (0.18 %)	s.d. (0.00 %)	s.d. (0.05 %)	s.d. (0.00 %)	---	s.d. (3.36 %)	n.d. (66.23 %)
Reptile	n.d. (78.15 %)	n.d. (5.31 %)	n.d. (74.07 %)	s.d. (3.84 %)	s.d. (3.36 %)	---	s.d. (0.01 %)
Vascular Plant	s.d. (0.00 %)	s.d. (0.00 %)	s.d. (0.00 %)	s.d. (0.00 %)	n.d. (66.23 %)	s.d. (0.01 %)	---

n.d. = no difference, s.d. = significant difference

Significance Testing between each specimen (III/III)

– Comparison matrix (simplified to specimen with more protection):

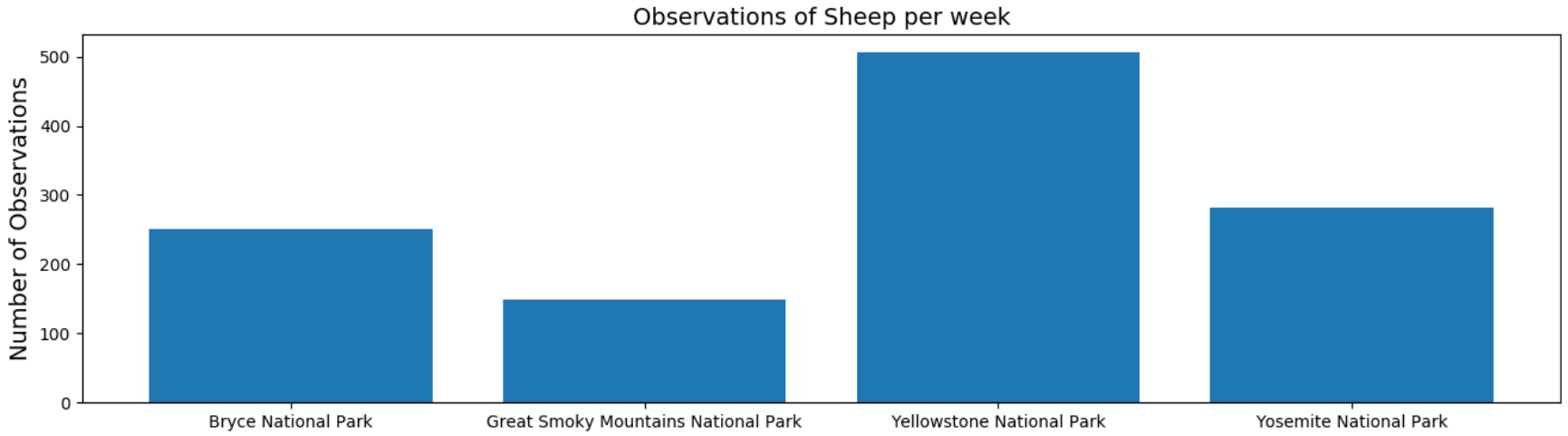
	Amphibian	Bird	Fish	Mammal	Reptile
Amphibian	---	n.d. (17.59 %)	n.d. (82.48 %)	n.d. (12.76 %)	n.d. (78.15 %)
Bird	n.d. (17.59 %)	---	n.d. (7.67 %)	n.d. (68.76 %)	n.d. (5.31 %)
Fish	n.d. (82.48 %)	n.d. (7.67 %)	---	n.d. (5.61 %)	n.d. (74.07 %)
Mammal	n.d. (12.76 %)	n.d. (68.76 %)	n.d. (5.61 %)	---	s.d. (3.84 %)
Reptile	n.d. (78.15 %)	n.d. (5.31 %)	n.d. (74.07 %)	s.d. (3.84 %)	---

n.d. = no difference, s.d. = significant difference

⇒ *among all protected specimen, reptile and mammals show a significant difference and can therefore be expected to have a different endangerment potential (this should be investigated)*

Sheep Observations per park area (weekly obervations)

Park Name	No. of observed sheep
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

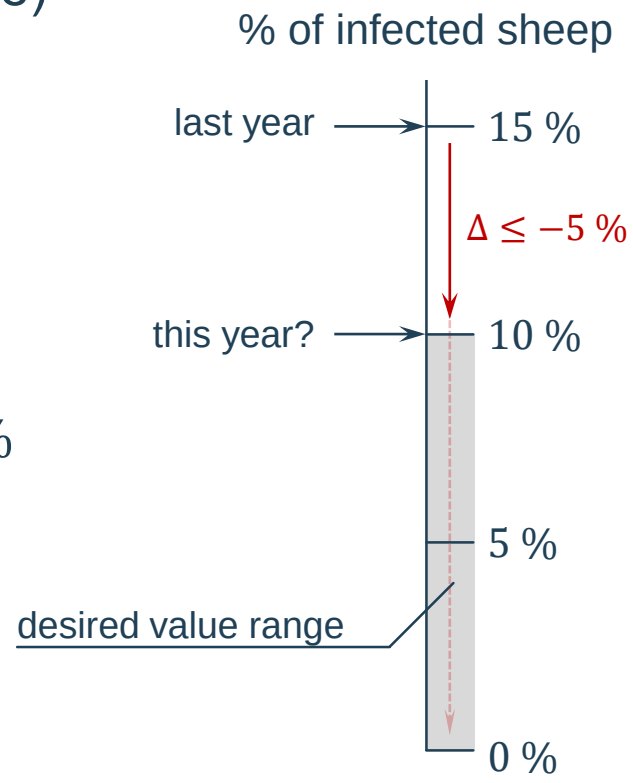


Foot and Mouth Disease Reduction Program

- Development investigation for the foot and mouth disease programs in all national parks
- Given requirements of the park rangers:

1. Initial baseline estimation: $B = 15\ \%$ (last year's value)
2. Level of significance: $90\ \%$
3. Disease target value $10\ \%$
4. Lift value for significant difference: $\Delta B = (-)\ 5\ \%$ (absolute)

$$L = \frac{\Delta B}{B} = (-)\frac{5\ \%}{15\ \%} = (-)\frac{0.05}{0.15} \approx (-)0.33 = (-)33.33\ \%$$



- Necessary sample size (number of observation): $S = 870$ sheep

Sample Size Determination and Observation Time Estimation

– Assuming the analyzed, weekly observation numbers per park:

Park Name	No. of observed sheep	Minimum Observation Time (rounded values)
Bryce National Park	250	3.48 (3.5) weeks
Great Smoky Mountains National Park	149	5.83 (6) weeks
Yellowstone National Park	507	1.71 (2) weeks
Yosemite National Park	282	3.08 (3.5) weeks