Codecademy Biodiversity Project Presentation

Tim Humphreys
April 2018

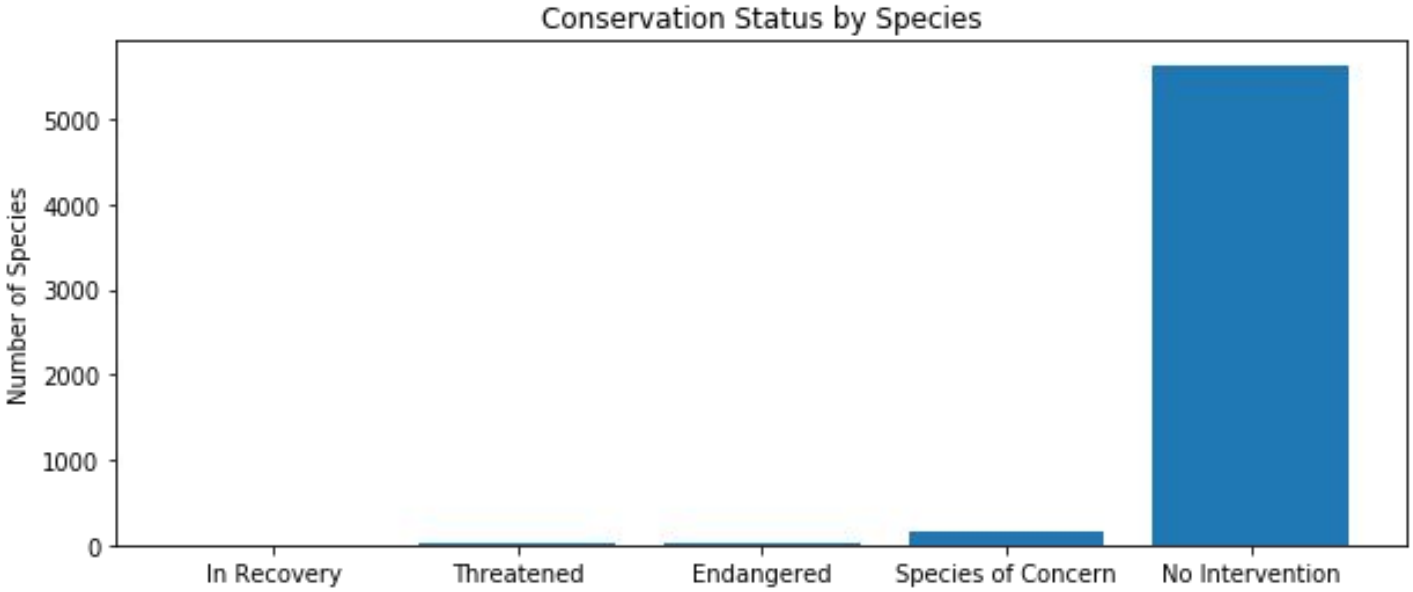
Data from species_info.csv

- We were given one dataset species_info.csv which contained the following data for various species in US Nattional Parks:
  - The scientific name of each species
  - The common names of each species
  - The species conservation status

- There were 5541 species in the Parks, which were classified into the following categories: Mammal, Bird, Reptile, Amphibian, Fish, Vascular Plant, Nonvascular Plant

- Each of these were categorised depending on their vulnerability, into 'No Intervention' 'Species of Concern',,'Endangered' ', Threatened', 'In Recovery'

Data from species_info.csv



- The vast majority of species were not endangered, but 180 were

Data from species_info.csv

	category	not_protected	protected	percent_protected (as decimal)
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

- Mammals and birds are most likely to be endangered, though the difference wasn't significant (pval of 0.687 in Chi squared test)
- Mammals are much more likely to be endangered than Reptiles, (pval of 0.038 in Chi squared test)
- Animals are much more endangered than plants

Recommendations for Conservationists

- Good Quality data is CRUCIAL for understanding and directing resources to the areas in highest need
- Focus Conservation efforts on Animals rather than plants
- Mammals and birds are by far the highest risk categories, so focus on these two
    - Get more species protected
- There are 151 “Species of Concern” - need to ensure these do not become endangered as only 29 are. Would increase workload significantly if a number of these became endangered
    - Act now to ensure risk is minimised
    - Preventative work rather than “fire-fighting” endangered species

Chi Square Test

What test to use?

	Numerical	**Categorical**
Sample vs. Known Quantity	1 Sample T-Test	Binomial Test
2 Samples	2 Sample T-Test	Chi Square
More Than 2 Samples	ANOVA and/or Tukey	Chi Square

- Because the data we had was Categorical and not Numerical a Chi Square Test was the most appropriate
- We were comparing 2 samples of data: Protected and Not Protected

observations.csv

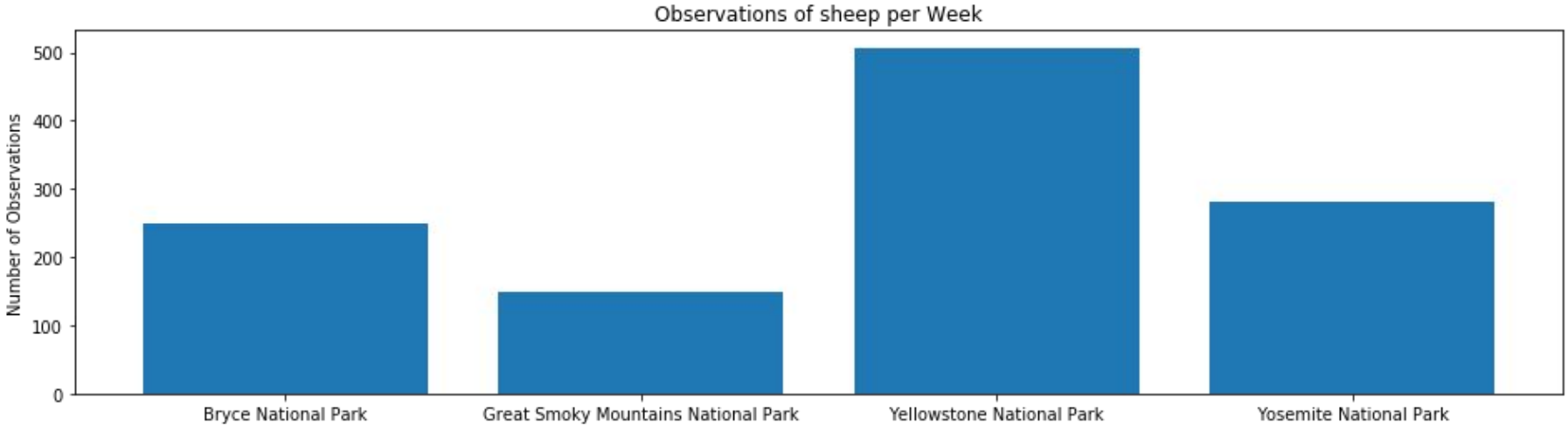
- We were also lucky to be able to access a second set of data that detailed the sightings of different species at several national parks for the past 7 days
- We were able to study the data and focus our attention on the sheep species within the Parks:

Common_names	conservation_status	is_protected
Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False
Bighorn Sheep, Bighorn Sheep	Species of Concern	True
Sierra Nevada Bighorn Sheep	Endangered	True

observations.csv

- We were able to see where the sheep had been spotted over the past week

Park_name	observations
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282



Sample size determination

Our scientists know that 15% of sheep at Bryce National Park have foot and mouth disease. Park rangers at Yellowstone National Park have been running a program to reduce the rate of foot and mouth disease at that park. The scientists want to test whether or not this program is working. They want to be able to detect reductions of at least 5 percentage point. For instance, if 10% of sheep in Yellowstone have foot and mouth disease, they'd like to be able to know this, with confidence.

- We used the sample size calculator at Optimizely to calculate the number of sheep that they would need to observe from each park
- We used the following inputs:
    - Baseline Conversion Rate is 15%
    - Minimum Detectable Effect is 33.333% (5% as a percentage of 15%)
    - Significance level 90%
- This resulted in a sample size of 510 sheep needed to observe change
- Focussing on two National parks, this number of observations would take this many weeks:

Number of weeks for Bryce observations is 2.04
Number of weeks for Yellowstone observations is 1.01

Thankyou!

timothydhumphreys@gmail.com

Sydney, Australia