Code Academy
- Biodiversity Project

James Hodge – November 2018

Data – species_info.csv

- The file species_info.csv contained a list of 5824 animals & plants and described their category, scientific name, common name and conservation status. A sample of the data is shown below;

	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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN

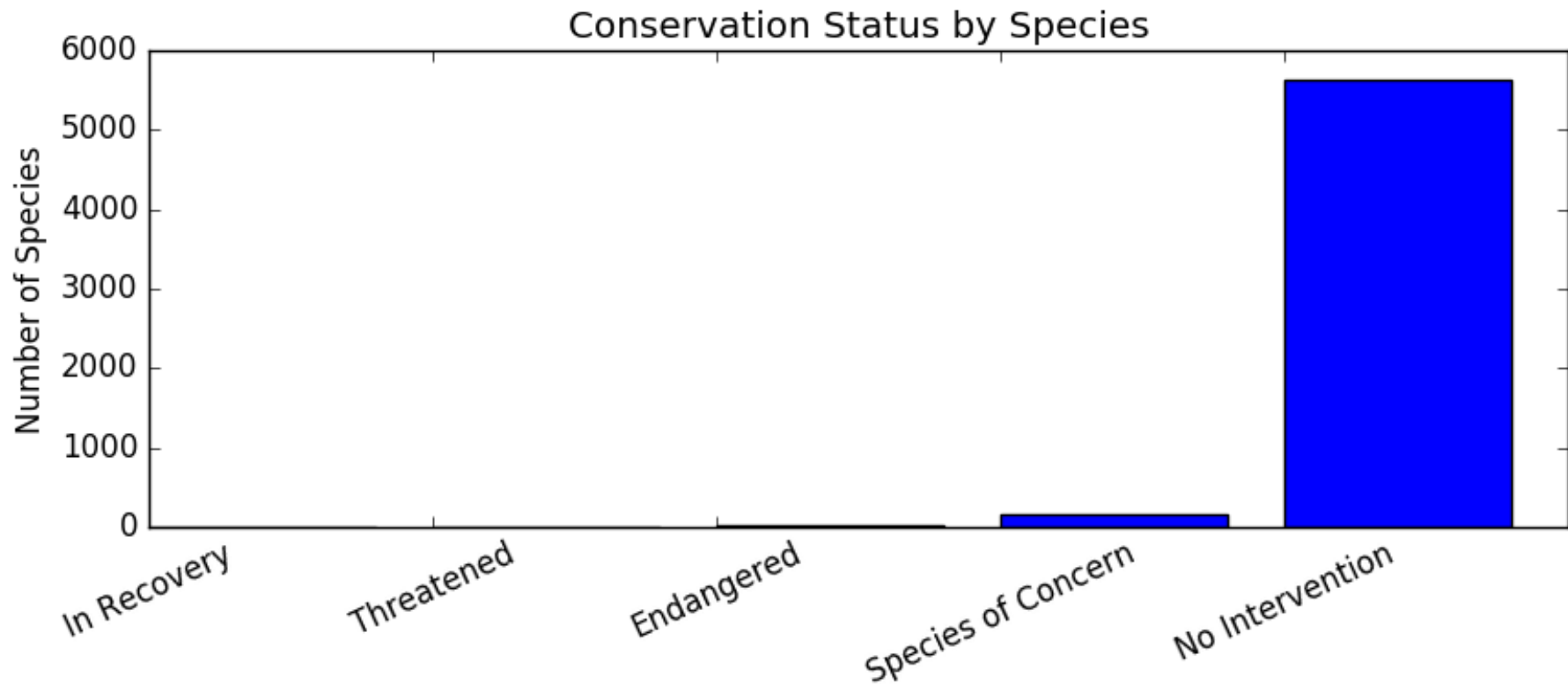- There were 7 types of specie category;

  Mammal, Bird, Reptile, Amphibian, Fish, Vascular plant,
  Nonvascular plant

- There were 5 types of conservation status;

  None, Species of concern, Endangered, Threatened, In Recovery

Data – species_info.csv

- The chart below shows the number of species in each conservation status

Conservation Status by Species

Significance calculations

- The below table shows by species the percentage that is protected.

	category	not_protected	protected	percent_protected
0	Amphibian	73	7	8.750000
1	Bird	442	79	15.163148
2	Fish	116	11	8.661417
3	Mammal	176	38	17.757009
4	Nonvascular Plant	328	5	1.501502
5	Reptile	74	5	6.329114
6	Vascular Plant	4424	46	1.029083

- Chi-squared significance tests were conducted;
  - Firstly to establish whether the difference in percentage protected where significantly different between mammals and birds
  - With a p-value of 0.44 this was proven not to be significant

  - Secondly the significance of the difference in percentage protected between mammals and reptiles
  - With a p-value of 0.02 this was shown to be significant

- I would recommend not only focusing on mammals which would seem to be the most protected but also birds as the significance tests show that the difference is not significant.

Observation data

■ Observations.csv was analysed this contained sightings of different species in the last 7 days across 4 national parks. Data from the species info file was filtered for sheep only and merged with the sightings data, sample below

	scientific_name	park_name	observations	category	common_names	conservation_status	is_protected	is_sheep
0	Ovis canadensis	Yellowstone National Park	219	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
1	Ovis canadensis	Bryce National Park	109	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
2	Ovis canadensis	Yosemite National Park	117	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3	Ovis canadensis	Great Smoky Mountains National Park	48	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4	Ovis canadensis sierrae	Yellowstone National Park	67	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True

Observations

- The sheep observations data was summarised in the table and chart below

	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
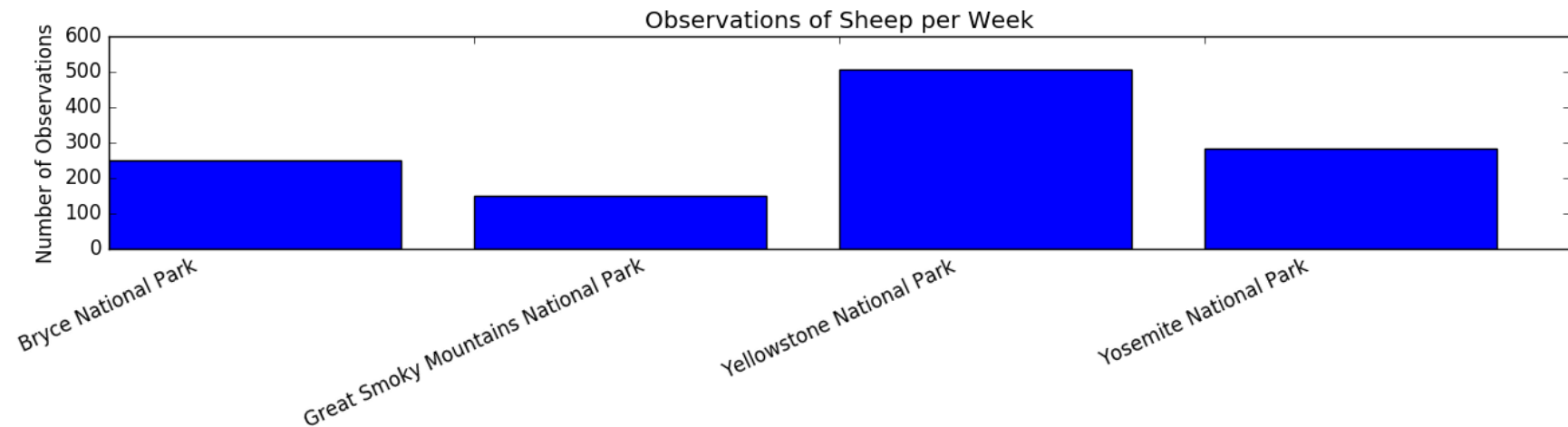
Sample size determination

- The sample size calculator was used to find the required sample size to determine with 90% confidence that the foot and mouth disease had reduced to 10% from 15%.

Baseline conversion rate: 15 %

Statistical significance: 85% **90%** 95%

Minimum detectable effect: 33.3333 %

Sample size: 870

- Given the number of sightings in the last 7 days this would take approximately 3.5 weeks at Bryce and just over 1.5 weeks at Yellowstone National Park to determine.

Thanks

James Hodge – November 2018