Biodiversity Capstone Project

Codeacademy

Part 1:

A first look at the Species table data

	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

Key points about the Species dataset

- the dataset contains information on **5541** unique species
- For each species entry, the dataset includes a scientific name, a list of common names and a **conservation status**
- The dataset also assigns each species to one of 7 **categories**

Cleaning up and aggregating the dataset

	category	scientific_name	Percentage of species
0	Amphibian	79	1.425735
1	Bird	488	8.807075
2	Fish	125	2.255910
3	Mammal	176	3.176322
4	Nonvascular Plant	333	6.009746
5	Reptile	78	1.407688
6	Vascular Plant	4262	76.917524

	conservation_status	scientific_name
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

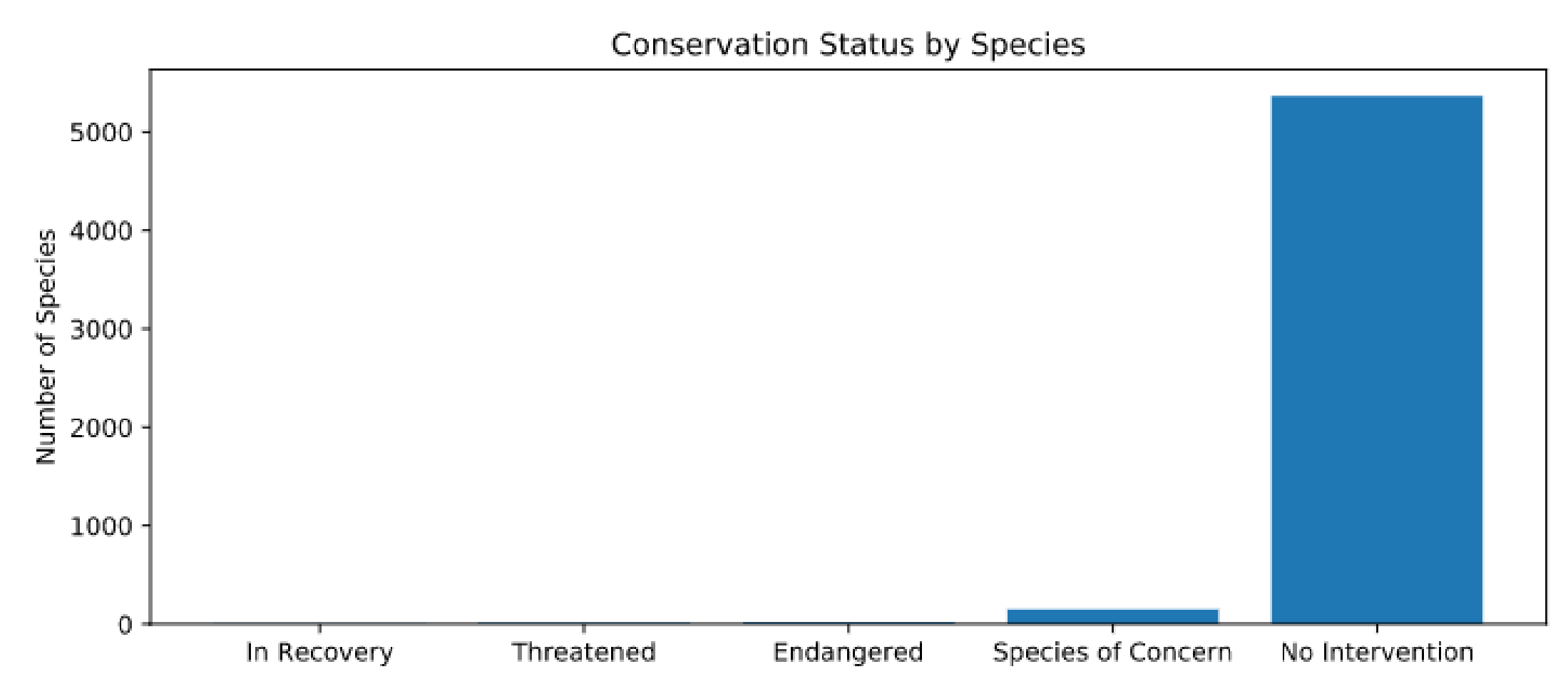
The Species by 'Category'

- We can see that more than ¾ of species in the dataset comprise of Vascular plants (78%)
- Reptiles and Amphibians comprise **less than 3%** of 'Animal' species featured in the dataset

Grouping by Conservation status

- There are **5** distinct Conservation statuses that species could be assigned
- For the purposes of the analysis, all species with no conservation status were assigned a '**No intervention**' status
- Approximately 3.2% of all species in the dataset have some kind of protected status
- For the 180 species with Protected status only 4 are in recovery, implying that for every **1 recovering species there are 45 species at risk** of declining posing serious implications for biodiversity

Visualising Conservation Status across species in the dataset



Does species category impact the likelihood of a species having protected status?

	category	not_protected	protected	percent_protected
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

Analysis

- To determine whether species in particular categories were more or less likely to have protected status we grouped all species by category and calculated the number of species with protected status as a percentage of total number of species in that category

Main observations

- Mammal species have the highest proportion of protected species among all categories at **17%** followed closely by Bird species at **15.3%** implying that species in these categories *seem* to be more likely to have protected status
- Plant species, both Vascular and Non-Vascular have the lowest proportion of protected species among categories
- Collectively, 'animals' are significantly more likely to have protected status than 'plants'

Further investigation

- We will conduct a test to assess whether the observed differences in likelihood of having protected status between the categories are **statistically significant** or are likely due to random chance

Statistical significance of differences in protection status between species categories

Chi-Squared test	Probability Statistical Significance threshold	P-value	Outcome
Bird – Mammal	> 5%	68.8%	Accept Null Hypothesis
Reptile – Mammal	> 5%	3.84%	Statistically significant

- To determine whether different categories of species were more or less likely to have protected status we conducted two Chi-Squared tests comparing the proportion of protected species of two categories. We set a probability threshold (p-value) of less than 5% as a benchmark for statistical significance.

Test results

- **Mammal and Bird**: our Chi-squared test results show that observed differences in the rate of protected status among birds and mammals can occur 68.8% of the time implying that differences are likely due to random chance and so the likelihood of having protected status is similar for species in both categories
- **Mammals and Reptiles**: our Chi-squared test results show that observed differences in the rate of protected status among mammals and reptiles have a probability of 3.85% implying that differences in likelihoods of protected status are statistically significant and that mammal species are indeed more likely to be at risk compared to reptile species

Recommendations for conservationists

- Observation data and statistical tests have shown that Mammal and Bird species are **more likely** than other categories of species to have protected status and therefore be **at risk**
- This data also shows that a greater proportion of Mammal and Bird species have protected status compared to other categories which could possibly be due to an accelerating trend of Mammals and Birds becoming endangered

Recommendations

1. Conservationists should increase observation data collection efforts for Mammal and Bird species with an emphasis on broadening the fields of data collected for each species so that deeper insights can be yielded from further analysis e.g. include data on geographic location, average temperature, proximity to nearest human settlement or infrastructure etc
2. Conservationists should attempt to collate and compile historic observation data on Mammal and Bird species so that a **time series** can be created which would allow the tracking of rates of change in protected status
3. Conservationists should step up efforts to increase awareness and increase the level of resources for the protection Mammal and Bird species, e.g. they can **reallocate conservation budgets** to reflect the higher levels of risk to Mammal and Bird species

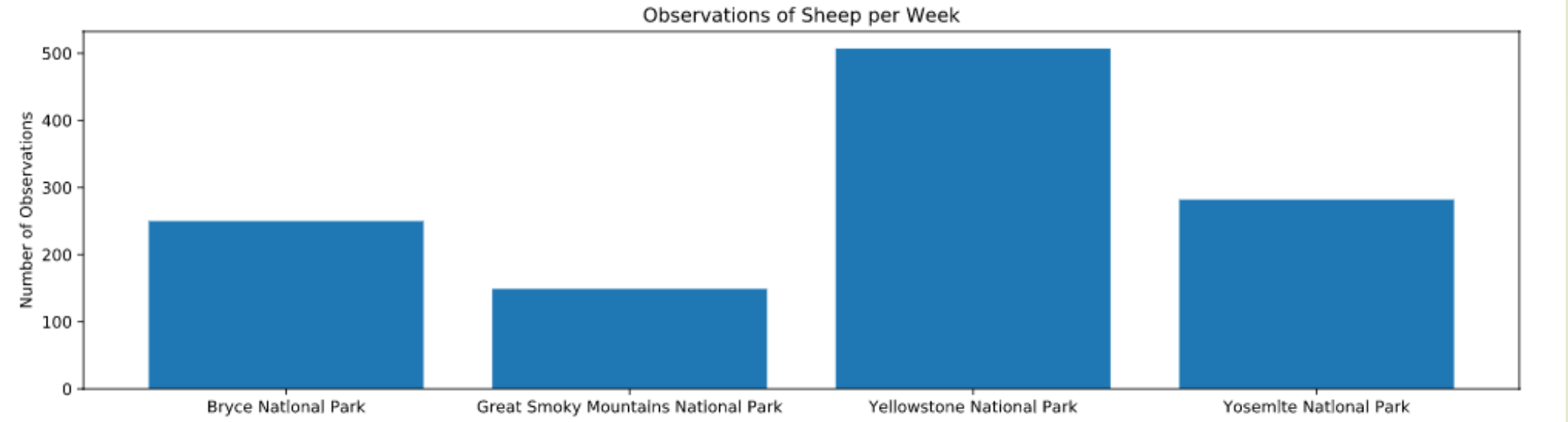
Part 2:

Foot & Mouth Sample Size Determination

Objective

- We wish to collate and group all the observation data available on Sheep species by location and compare observations across different parks.
- We wish to determine the correct sample size to assess the effectiveness of a program designed to reduce the incidence of Foot & Mouth disease among sheep population
- We wish to know how many weeks of observation would be required to collect enough samples from different parks

Sheep observations by park



	park_name	observations
0	Bryce National Park	576025.0
1	Great Smoky Mountains National Park	431820.0
2	Yellowstone National Park	1443562.0
3	Yosemite National Park	863332.0

Calculating Sample Size for Foot & Mouth disease study

Inputs to the sample size calculation

- Currently available information tells us that in previous year 15% of sheep observed in Bryce Park suffered from Foot & Mouth disease, this will be used as our **baseline conversion**.
- Park rangers have specified that they would need to see at least a 5% reduction in this rate of infection to judge the program as effective implying that we require a sample size sufficient to detect a 33.3% fall in observable cases of Foot & Mouth, this will be our **minimal detectable effect**
- We wish to have a high degree of certainty in results and so need a sample size that can give us results that are at least 90% **statistically significant**

Sample size calculator		
Baseline Conversion	Minimum Detectable effect	Statistical significance of result
33.3%	5%	90%
Sample size = 870 Sheep		
Number of observation weeks (Yellowstone) = **1.71 weeks**		
Number of observation weeks (Bryce Park) = **3.48 weeks**		