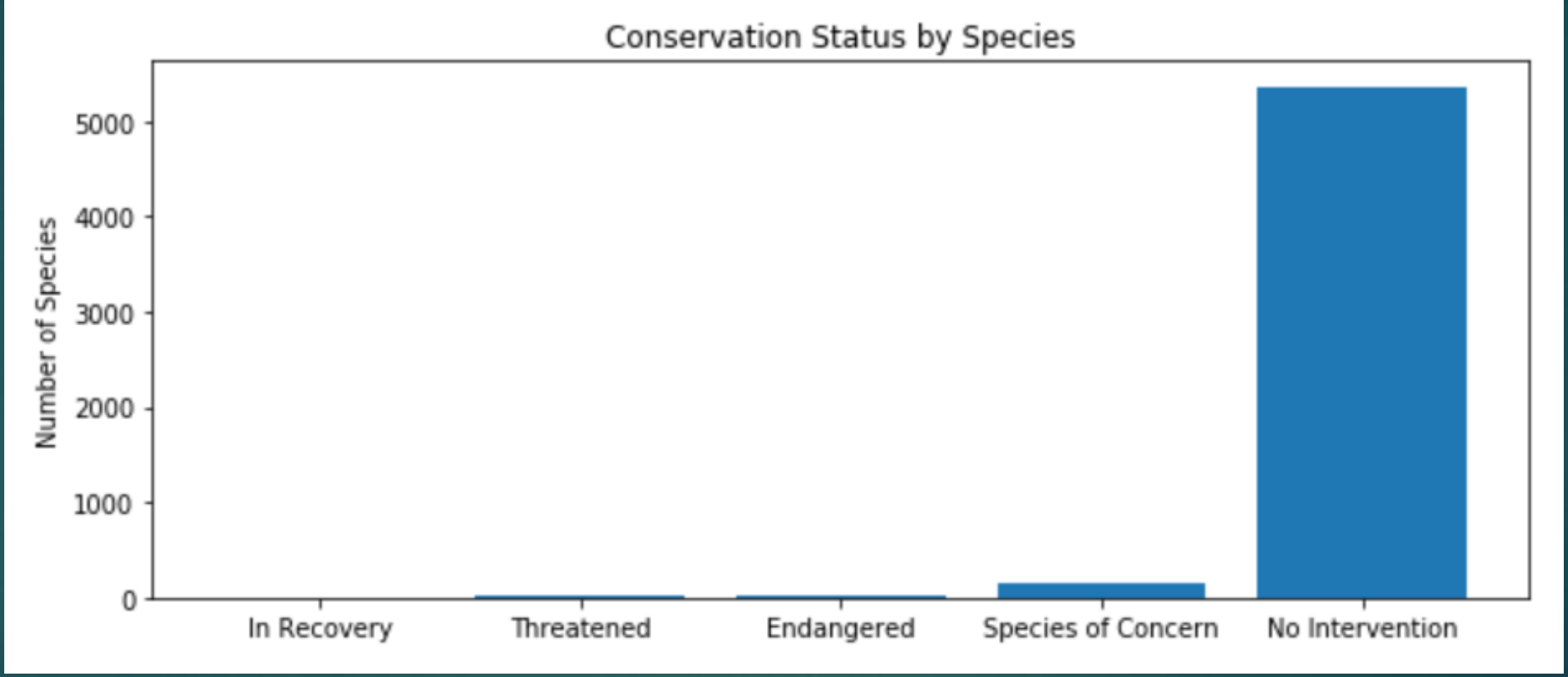
Codecademy :: Capstone Project

BIODIVERSITY

Objective 1 :: Initial Observations and Deductions

- The database contains 5824 records, of which, 5541 correspond to a unique scientific name (which will heirin be referenced to as ‘species’).
- The 5541 species fall into 7 different categories.
- Upon some slight data re-arranging, you can immediately tell that a very small population of the species are being monitored by the Natural Park Service due to conservation concerns: 3.21% to be exact (5,363).
- View exhibit 1 on the following page for visual.

Exhibit #1


Conservation Status by Species
Number of Species
5000
4000
3000
2000
1000
0
In Recovery
Threatened
Endangered
Species of Concern
No Intervention

Objective 1:: Initial Observations and Deductions Cont.

- If we slice up the data a little more we can see a clear disproportionate lack of conservation *need* for the vascular plant species.
- 78.61% (4,216) of the species which are not under a conservation status are vascular plants.
- Now, this in and of itself isn't necessarily significant (there could just be a lot more vascular plants than other species), however, the fact that relative to the number of species which have a conservation status, the plant species still comes in last with only 1.07% requiring protection.
- *We are making the assumption that the Park Service always designates a conservation status when appropriate. In other words, the 1.07% is a good thing, and not due to a lack of empathy or bias.

Exhibit #2

	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

Objective 2 :: Determining statistical significance

- On the opposite end of the spectrum are the mammals, with 17.04% (30) of their species under a conservation status; this is the highest of the species.
- For perspective these protected mammals make up .54% of the total species (30/5,541).
- In second place are the birds with 15.63% protected.
- Now, what we want to determine is if there is any relationship between the species and the conservation status. In other words, are the 2 variables associated at all? We need to see if there's a statistical significance.

Objective 2 :: Determining statistical significance Cont.

- In order to determine if there is a statistical significance between the two categories, we ran a Chi-Squared test.
- Our Null Hypothesis is that the difference in % is simply due to chance, and the 2 variables we are testing (species + conservation status) have no association
- Our test returns a p-value of .687.
- *As everyone knows*, there is no significance amongst test samples unless the p-value returns < .05
- Due to this, we accept our Null Hypothesis and the notion that the mammals aren't necessarily more likely to be endangered than the birds.

Objective 2 :: Determining statistical significance Cont.

- What if we ran the same test for reptiles (6.41%) and mammals though? Would we come to the same conclusion?
- When we run a Chi-Squared test against these two, we get a p-value of .038 which is significant!
- This indicates to us that there is an association between the species and endangerment!
- I'm sure if we ran the test against mammals and vascular plants, we would see a giant significance!

Objective 3 :: Recommendation to Conservationists

- Based on our Chi Test we determined that certain species are more likely than others to become endangered/need protection.
- Due to this finding, we would advise conservationists that there is a negative tangible effect primarily influencing mammals and birds which is leading to un-proportionate need for protection of these 2 species.
- In other words, the mammals and birds are not "within the margin of error" or just an anomaly – rather – there is a statistical connection which should be investigated further in order to identify the cause and curb the issue.

Objective 4 :: Sample Size Determination – Foot & Mouth Disease

- As we know, conservationists have been studying the number of sheep sightings throughout 4 parks over the past 7 days to see if their program to reduce the Foot and Mouth Disease has been effective.
- However, they were not sure as to how many sheep they needed to examine in order to get a good enough sample size which would render their findings accurate.
- The scientists sent us their observations and asked us to determine this.

Objective 4 :: Sample Size Determination – Foot & Mouth Disease Cont.

- After manipulating the species data a bit we are able to figure out that there were 3 types of sheep out of the 5,824 records in the file.

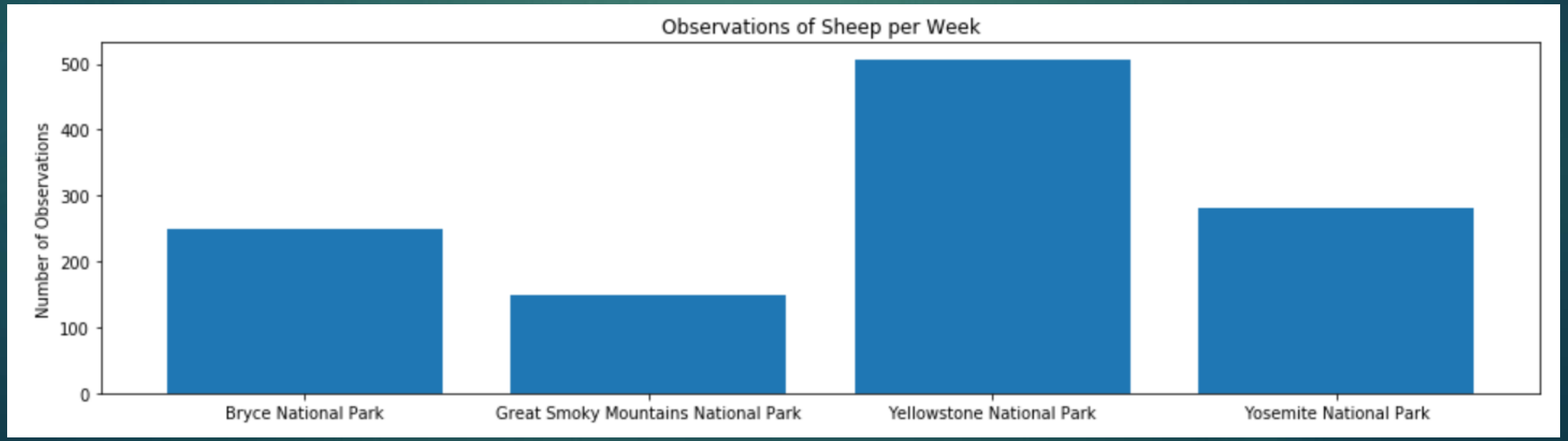
Exhibit #3

	category	scientific_name	common_names
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep

Objective 4 :: Sample Size Determination – Foot & Mouth Disease Cont.

- We then merge this data with our observations data from the scientists. Initially we get 12 records (3 species x 4 parks). But we can group the data by park to get the total observations across all 3 sheep species. Below is our results

Exhibit #4



Objective 4 :: Sample Size Determination – Foot & Mouth Disease Cont.

- We now have the # of sheep observed per park per week
- In addition to this information we know that in Bryce park, 15% of the sheep have Foot & Mouth Disease. We will use this for our "Baseline Conversion Rate"
- Using the scientists program in Yellowstone, they are hoping to bring this down to 10% which is a drop of 33.3%. We also call will call this the "Minimum Detectable Effect"
- To better guarantee accuracy, we will use a Statistical Significance level of 90%

Objective 4 :: Sample Size Determination – Foot & Mouth Disease Cont.

- Using the data in the last slide we can calculate that we will need to observe 510 sheep to have a large enough sample size to be confident in whatever results we find.
- Based on this it would take approximately 2 weeks at Bryce Park and 1 week at Yellowstone