Biodiversity for National Parks

Codecademy Pro Intensive Capstone

The Species 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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN
5	Mammal	Odocoileus virginianus	White-Tailed Deer	NaN
6	Mammal	Sus scrofa	Feral Hog, Wild Pig	NaN
7	Mammal	Canis latrans	Coyote	Species of Concern
8	Mammal	Canis lupus	Gray Wolf	Endangered
9	Mammal	Canis rufus	Red Wolf	Endangered
10	Mammal	Urocyon cinereoargenteus	Common Gray Fox, Gray Fox	NaN
11	Mammal	Vulpes fulva	Black Fox, Cross Fox, Red Fox, Silver Fox	NaN
12	Mammal	Vulpes vulpes	Red Fox	NaN
13	Mammal	Felis concolor	Mountain Lion	NaN
14	Mammal	Felis silvestris	Wild Cat, Wildcat	NaN
15	Mammal	Lynx rufus	Bobcat	NaN
16	Mammal	Puma concolor	Panther (Mountain Lion)	NaN
17	Mammal	Mephitis mephitis	Striped Skunk	NaN
18	Mammal	Spilogale putorius	Eastern Spotted Skunk	NaN
19	Mammal	Lontra canadensis	River Otter	NaN

The information from the table above was provided by the National Parks Service. A quick look at the first few rows shows that, while it is extensive, this data can be redundant and perhaps a bit hard to understand. For example, I don't need the scientific names for my purposes. Removing this column would make the table easier to understand. Additionally, if we want to look at all data as a whole, it would be more useful to see each category or conservation status as one entry, rather than list each species in each category or conservation status.

The Condensed Species Data

	category	common_names	conservation_status
0	Mammal	Gapper's Red-Backed Vole	No Intervention
1	Mammal	American Bison, Bison	No Intervention
2	Mammal	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	No Intervention
3	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention
4	Mammal	Wapiti Or Elk	No Intervention
5	Mammal	White-Tailed Deer	No Intervention
6	Mammal	Feral Hog, Wild Pig	No Intervention
7	Mammal	Coyote	Species of Concern
8	Mammal	Gray Wolf	Endangered
9	Mammal	Red Wolf	Endangered
10	Mammal	Common Gray Fox, Gray Fox	No Intervention
11	Mammal	Black Fox, Cross Fox, Red Fox, Silver Fox	No Intervention
12	Mammal	Red Fox	No Intervention
13	Mammal	Mountain Lion	No Intervention
14	Mammal	Wild Cat, Wildcat	No Intervention
15	Mammal	Bobcat	No Intervention
16	Mammal	Panther (Mountain Lion)	No Intervention
17	Mammal	Striped Skunk	No Intervention
18	Mammal	Eastern Spotted Skunk	No Intervention
19	Mammal	River Otter	No Intervention

By getting rid of the unnecessary information and renaming our “NaN” values, the table is less cluttered and can be sifted through more easily.

Being able to read data is great, but until we understand how to manipulate and apply what we are reading we will not realize the true potential of the data. In the following slides we will look at a couple ways we were able to apply tests to the data to determine if what we were looking at is significant, or simply the product of chance.

What are the chances?

	category	not_protected	protected	percent_protected
0	Amphibian	72	7	8.860759
1	Bird	413	75	15.368852
2	Fish	115	11	8.730159
3	Mammal	146	30	17.045455
4	Nonvascular Plant	328	5	1.501502
5	Reptile	73	5	6.410256
6	Vascular Plant	4216	46	1.079305

Birds and Mammals stick out as the two most protected categories. But can we confidently say mammals are more likely to be protected than birds?

According to our Chi-squared test, there is a 68% chance that mammals being more protected is the result of chance based on the number of observations and the spread.

```
contingency = [[30,146], [75,413]]
chi2, pval, c, d = chi2_contingency(contingency)
print(pval)
```

```
0.6875948096661336
```

```
reptile_mammal = [[5,73], [30,146]]
e, pval_reptile_mammal, f, g = chi2_contingency(reptile_mammal)
print(pval_reptile_mammal)
```

```
0.03835559022969898
```

However, there is only a 3% chance that reptiles are more protected than mammals. We can say with confidence that mammals are more protected.

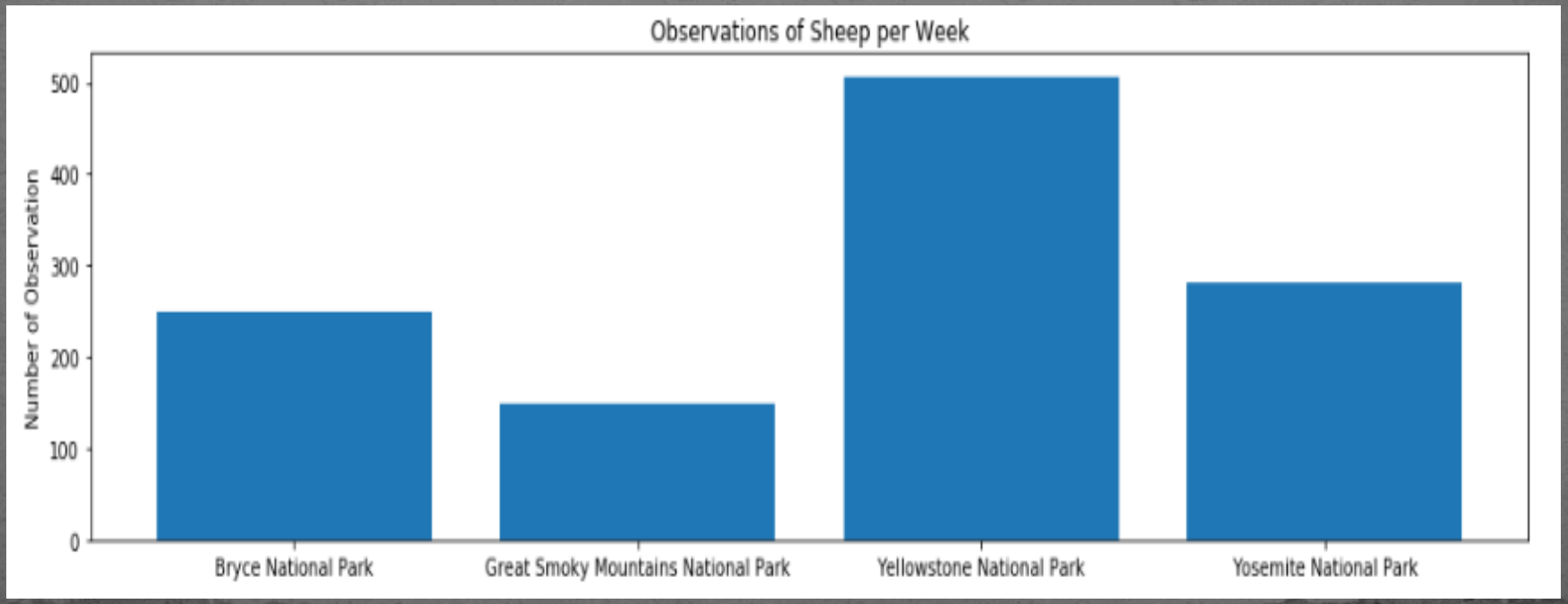
Recommendations to Conservationists

The Chi-squared significance tests gives us a lot of good data to interpret. We can see there is a significant amount of birds and reptiles that are protected compared to other species. Just as important is that there is a significant amount of plants, both vascular and non-vascular that are not protected. Resources and efforts might be best spent focusing on protecting birds and mammals, and studying how plants have been able to survive so well.

Foot and Mouth Disease in Sheep

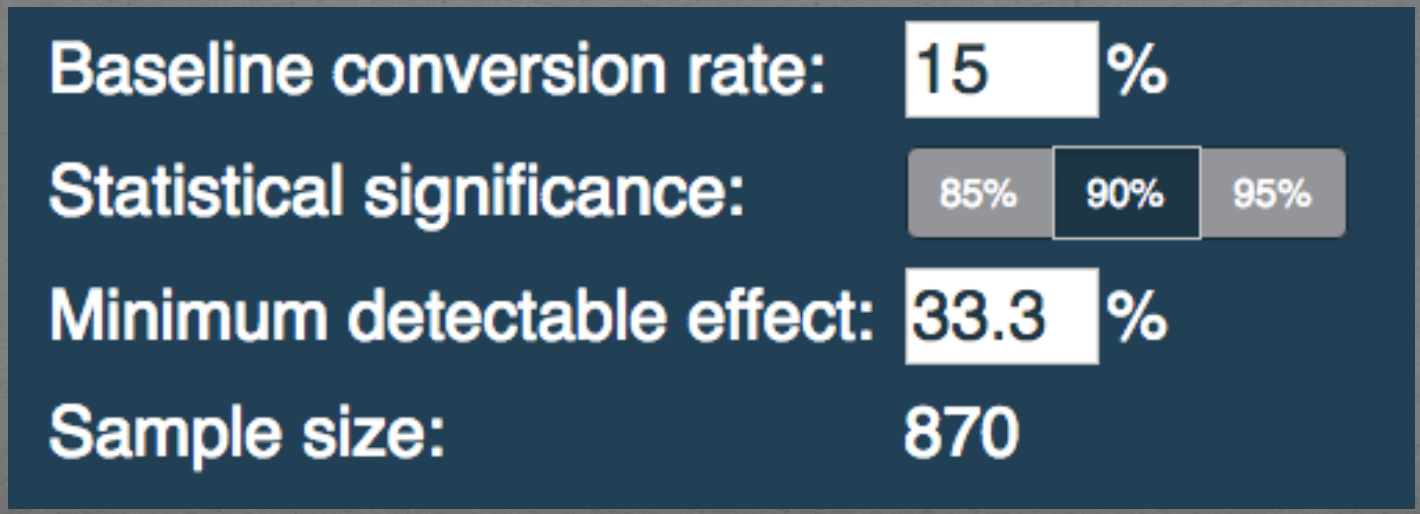
While the last tests let us know on a general level which species could use more attention, narrowing our scope could give us more actionable intelligence. The chart below shows the number of sheep observed by a group of scientist in different parks over the last week.

	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



Foot and Mouth Disease in Sheep

These scientist would like to know how many sheep need to be observed to determine is a certain disease is becoming less prevalent. This is a simple task with a sample size calculator as long as you know what each input is. According to the information we have been given, 15% of sheep in Bryce National Park had Foot and Mouth Disease last year. This is out baseline. The Park Rangers would like to be able to detect differences of at least 5%. This means our minimum detectable effect will be 33% (5% is 1/3 of 15%, or 33%). When we use a standard statistical significance of 90% we get a sample size of 870 sheep. In other words, scientist would have to observe at least 870 sheep before they could say with 90% confidence that the number sheep with this disease has decreased by 5%.



Recommendation to Conservationists

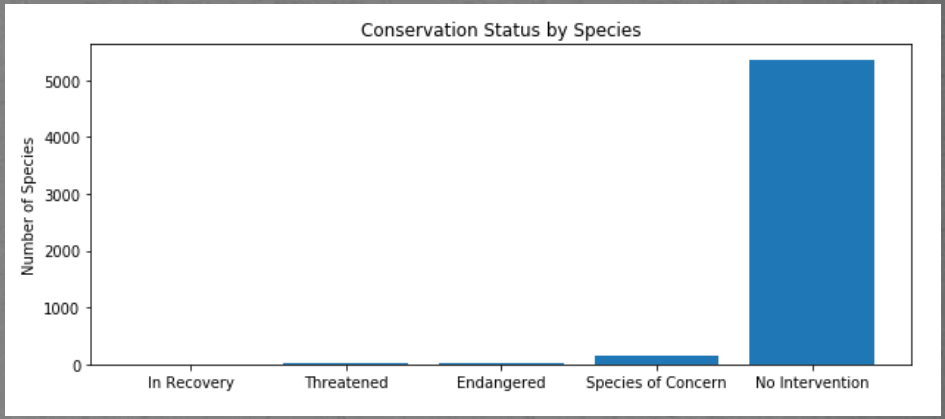
If we know that the Sierra Nevada Bighorn Sheep is more at risk than other species of sheep it might be worth while to focus our observations on those sheep in particular.

	category	common_names	conservation_status
3	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention
3014	Mammal	Bighorn Sheep, Bighorn Sheep	Species of Concern
4446	Mammal	Sierra Nevada Bighorn Sheep	Endangered

While this disease may not be the only danger plaguing this species, and while it will without a doubt take longer to collect this data, it is my opinion that focusing on the sheep that are endangered would do the most good.

List of Chart created:

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



	category	not_protected	protected	percent_protected
0	Amphibian	72	7	8.860759
1	Bird	413	75	15.368852
2	Fish	115	11	8.730159
3	Mammal	146	30	17.045455
4	Nonvascular Plant	328	5	1.501502
5	Reptile	73	5	6.410256
6	Vascular Plant	4216	46	1.079305

	is_protected	is_sheep	park_name	observations
0	False	True	Yosemite National Park	126
1	False	True	Great Smoky Mountains National Park	76
2	False	True	Bryce National Park	119
3	False	True	Yellowstone National Park	221
4	True	True	Yellowstone National Park	219
5	True	True	Bryce National Park	109
6	True	True	Yosemite National Park	117
7	True	True	Great Smoky Mountains National Park	48
8	True	True	Yellowstone National Park	67
9	True	True	Yosemite National Park	39
10	True	True	Bryce National Park	22
11	True	True	Great Smoky Mountains National Park	25

Baseline conversion rate: 15 %

Statistical significance: 85% 90% 95%

Minimum detectable effect: 33.3 %

Sample size: 870

	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