Biodiversity Capstone

A look at species, national parks, and diseases

What is in species_info.csv?

Data about different species in our National Parks, including:

- The category of each species
- The scientific name of each species
- The common names of each species
- The species conservation status

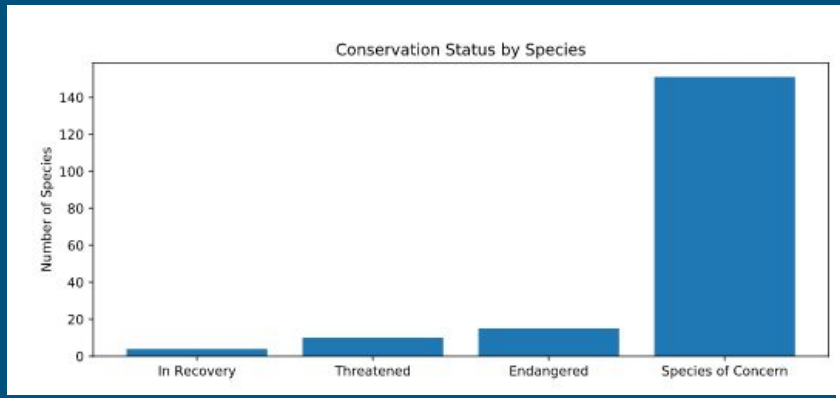
	category	scientific_name	common_names	conservation_status
0	Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	nan
1	Mammal	Bos bison	American Bison, Bison	nan
2	Mammal	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Domesticated Cattle	nan
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	nan
4	Mammal	Cervus elaphus	Wapiti Or Elk	nan

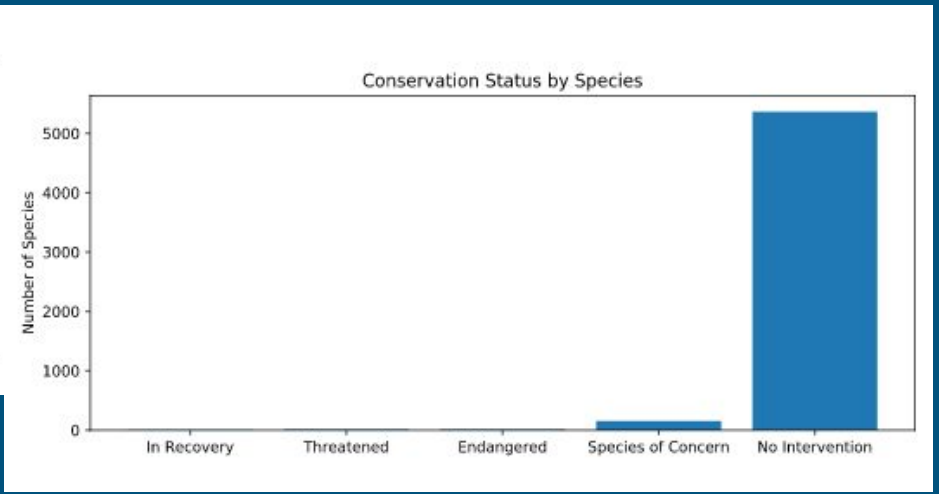
Conservation Status with None Intervention

	conservation_status	scientific_nam
0	Endangered	15
1	In Recovery	4
2	Species of Concern	151
3	Threatened	10

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

Conservation Status Bar Charts





Are certain types of species more likely to be endangered?

	category	is_protected	scientific_name
0	Amphibian	False	72
1	Amphibian	True	7
2	Bird	False	413
3	Bird	True	75
4	Fish	False	115

	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

Mammals And Birds are significantly more likely to be endangered than reptiles.

script.py

```
import codecademylib
import pandas as pd
from matplotlib import pyplot as plt
from scipy.stats import chi2_contingency

contingency = [[30, 146],
               [75, 413]]

pval = chi2_contingency(contingency)[1]
print(pval)
# No significant difference because pval > 0.05

contingency_reptile_mammal = [[30, 146],
                              [5, 73]]

pval_reptile_mammal = chi2_contingency(contingency_reptile_mammal)[1]
print(pval_reptile_mammal)
# Significant difference! pval_reptile_mammal < 0.05
```

```
0.687594809666
0.0383555902297
```

Recommendations

Since mammals and birds are more likely to be endangered, focus should be on methods that could help mammals and birds.

Observations

```
1
2
3  import codecademylib
4  import pandas as pd
5  from matplotlib import pyplot as plt
6
7  observations =
   pd.read_csv('observations.csv')
8  print observations.head()
```

	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National Park	85

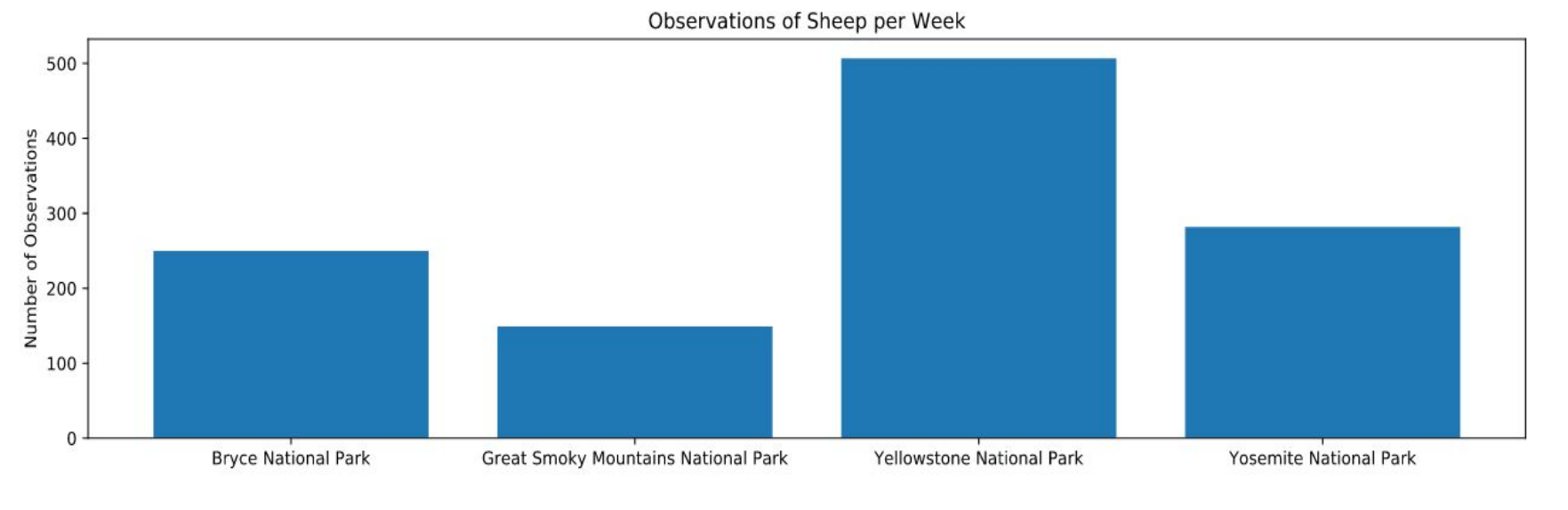
Mammal Sheep

	category	scientific_name	common_names	conservation_status	is_protected	is_sheep
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

Total number of sheep by park

	park_name	observations	is_protected	is_sheep
0	Bryce National Park	250	2.0	3.0
1	Great Smoky Mountains National Park	149	2.0	3.0
2	Yellowstone National Park	507	2.0	3.0
3	Yosemite National Park	282	2.0	3.0

Observations of Sheep per Week bar chart



Foot and Mouth Reduction Effort - Sample Size Determination

```
1  baseline = 0.15
2  minimum_detectable_effect = 100 * 0.05 / 0.15
3  print minimum_detectable_effect
4  sample_size_per_variant =  870
5  yellowstone_weeks_observing = float(870)/507
6  print yellowstone_weeks_observing
7  bryce_weeks_observing = float(870)/250
8  print bryce_weeks_observing
```

Run

```
33.3333333333
1.71597633136
3.48
```

Sample size determination

Sample Size: 870

Yellowstone: Observe 1.71 weeks

Bryce: Observe 3.48 weeks