Biodiversity for the National Parks

Capstone Option 2

Amanda Olavessen

Based off of the information provided from the species_info.csv I noticed that at least 180 species was in some stage of intervention.

Biodiversity Capstone Project

script.py

```
import pandas as pd
from matplotlib import pyplot as plt

species = pd.read_csv('species_info.csv')
print species.head()

species_count = species.scientific_name.nunique()

species_type = species.category.unique()

conservation_statuses = species.conservation_status.unique()

conservation_counts =
species.groupby('conservation_status').scientific_name.nunique().res
et_index()

print conservation_counts

species.fillna('No Intervention', inplace = True)

conservation_counts_fixed =
species.groupby('conservation_status').scientific_name.nunique().res
et_index()
print(conservation_counts_fixed)
```

Run

```
  category                scientific_name  \
0   Mammal  Clethrionomys gapperi gapperi
1   Mammal                      Bos bison
2   Mammal                     Bos taurus
3   Mammal                     Ovis aries
4   Mammal                  Cervus elaphus

                                    common_names conservation_status
0                         Gapper's Red-Backed Vole                NaN
1                            American Bison, Bison                NaN
2  Aurochs, Aurochs, Domestic Cattle (Feral), Dom...                NaN
3  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)                NaN
4                                   Wapiti Or Elk                NaN
  conservation_status  scientific_name
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
```

Bar Graph of Conservation Status by Species:

Biodiversity Capstone Project

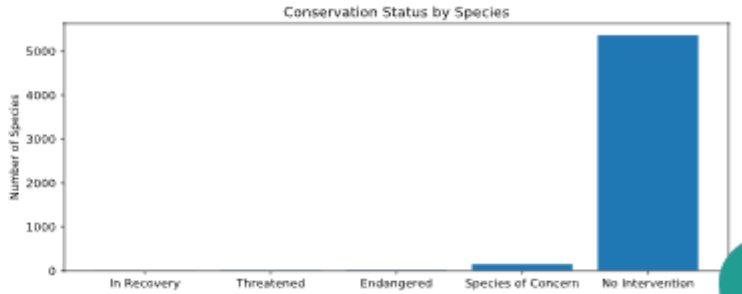
script.py

```
species.fillna('No Intervention', inplace = True)

protection_counts = species.groupby('conservation_status')\
    .scientific_name.nunique().reset_index()\
    .sort_values(by='scientific_name')
print(protection_counts)

scientific_name = [4, 10, 15, 151, 5363]
plt.figure(figsize = (10, 4))
ax = plt.subplot()
plt.bar(range(len(protection_counts)), scientific_name)
ax.set_xticks([0, 1, 2, 3, 4])
ax.set_xticklabels(['In Recovery', 'Threatened',
'Endangered', 'Species of Concern', 'No Intervention'])
plt.ylabel('Number of Species')
plt.title('Conservation Status by Species')
plt.show()
```

Run

Type URL here

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



Pivot Table of the different species and if they are protected or not and the percentages of each species protected.

Biodiversity Capstone Project

script.py

```
species['is_protected'] = species.conservation_status
!= 'No Intervention'

category_counts = species.groupby(['category',
'is_protected']).scientific_name.nunique().reset_index
()

category_pivot = category_counts.pivot
(columns='is_protected',
                       index='category',
                       values='scientific_name')\
                       .reset_index()

category_pivot.columns = ['category',
'not_protected', 'protected']
category_pivot['percent_protected'] =
category_pivot.protected / (category_pivot.protected
+ category_pivot.not_protected)

print category_pivot
```

	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

To determine if there are certain species that are more likely to be endangered than other species I performed a Chi-Squared Test. My first comparison was Mammals and Birds. This did not show much of a significance. My next comparison was Mammals and Reptiles. This did show a significant difference that some species are more likely to be endangered than others. My recommendation for conservationists concerned about endangered species is to keep track of the species that have the least amount of protection, because they will need the most help with intervention.

Biodiversity Capstone Project

script.py

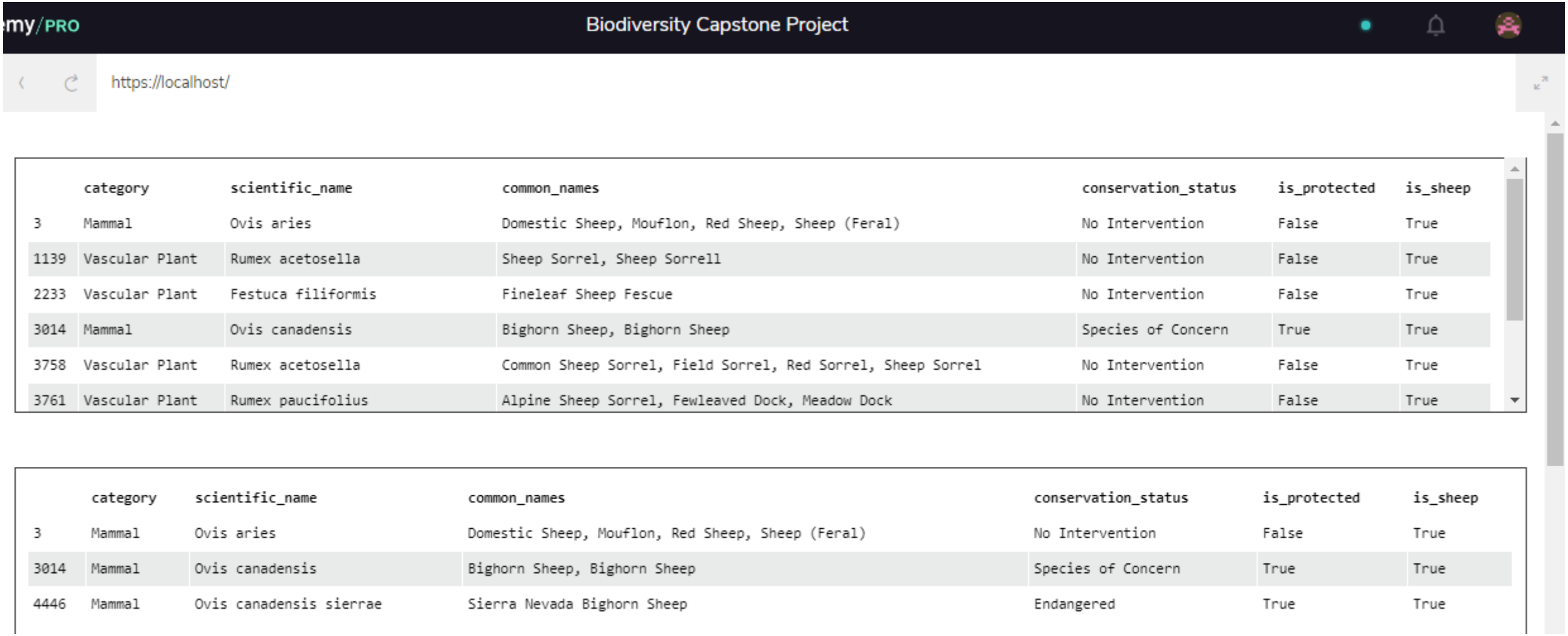
```
# print category_counts.head()

category_pivot =
category_counts.pivot(columns='is_protected',
index='category', values='scientific_name').reset_index()

category_pivot.columns = ['category', 'not_protected',
'protected']

category_pivot['percent_protected'] =
category_pivot.protected / (category_pivot.protected +
category_pivot.not_protected)

print category_pivot

contingency = [[38, 176],
     [79, 442]]
chi2, pval, dof, expected = chi2_contingency(contingency)
print pval
contingency2 = [[38, 176],
     [5, 74]]
chi2, pval_reptile_mammal, dof, expected =
chi2_contingency(contingency2)
print(pval_reptile_mammal)
```

Run

	category	not_protected	protected	percent_protected
0	Amphibian	73	7	0.087500
1	Bird	442	79	0.151631
2	Fish	116	11	0.086614
3	Mammal	176	38	0.177570
4	Nonvascular Plant	328	5	0.015015
5	Reptile	74	5	0.063291
6	Vascular Plant	4424	46	0.010291

```
0.445901703047
0.0233846521487
```

Break down list of species with “Sheep” in the name, and list of Sheep species.



	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
1139	Vascular Plant	Rumex acetosella	Sheep Sorrel, Sheep Sorrell	No Intervention	False	True
2233	Vascular Plant	Festuca filiformis	Fineleaf Sheep Fescue	No Intervention	False	True
3014	Mammal	Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern	True	True
3758	Vascular Plant	Rumex acetosella	Common Sheep Sorrel, Field Sorrel, Red Sorrel, Sheep Sorrel	No Intervention	False	True
3761	Vascular Plant	Rumex paucifolius	Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock	No Intervention	False	True

	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

Sheep Observation list and Observations by Park list.

emy/PRO Biodiversity Capstone Project

	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

	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

Bar Graph: Observations of Sheep per Week

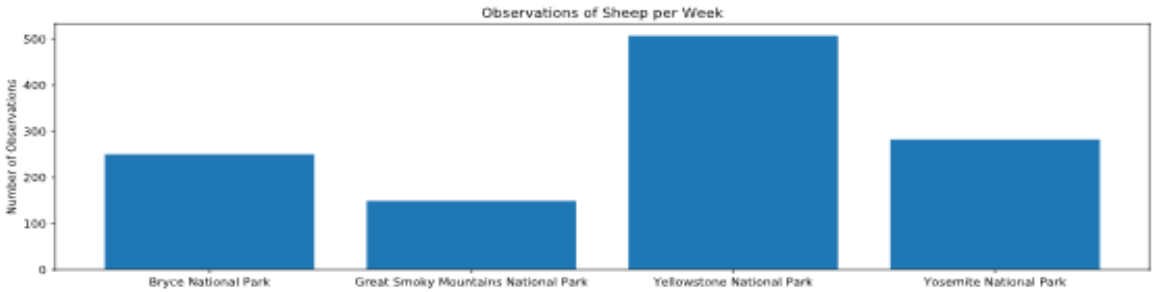
Biodiversity Capstone Project

script.py

```
8
9   observations = pd.read_csv('observations.csv')
10
11  sheep_observations = observations.merge(sheep_species)
12
13  obs_by_park = sheep_observations.groupby
    ('park_name').observations.sum().reset_index()
14
15  observations = [250, 149, 507, 282]
16  plt.figure(figsize = (16, 4))
17  ax = plt.subplot()
18  plt.bar(range(len(obs_by_park)), observations)
19  ax.set_xticks([0, 1, 2, 3])
20  ax.set_xticklabels(['Bryce National Park', 'Great Smoky
    Mountains National Park', 'Yellowstone National Park',
    'Yosemite National Park'])
21  plt.ylabel('Number of Observations')
22  plt.title('Observations of Sheep per Week')
23  plt.show()
```

Run

Type URL here



Based on the information given, the scientists would need to spend more weeks at the Great Smoky Mountain National Park, 5.84 weeks, and less time at Yellowstone National Park, 1.72 weeks, to determine how many of their sheep have foot and mouth disease.