BiodiversityProject

July 23, 2025

[4]:
```python
from matplotlib import pyplot as plt
import pandas as pd
species = pd.read_csv('species_info.csv')
```

[5]:
```python
species = pd.read_csv('species_info.csv')
```

[6]:
```python
species.head()
```

[6]:
```
  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
```

[7]:
```python
species.scientific_name.nunique()
```

[7]:
```
5541
```

[8]:
```python
species.category.unique()
```

[8]:
```
array(['Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant',
       'Nonvascular Plant'], dtype=object)
```

[9]:
```python
species.conservation_status.unique()
```

[9]:
```
array([nan, 'Species of Concern', 'Endangered', 'Threatened',
       'In Recovery'], dtype=object)
```

```
[10]: species.groupby('conservation_status').scientific_name.nunique().reset_index()
```

```
[10]:   conservation_status  scientific_name
      0          Endangered               15
      1         In Recovery                4
      2  Species of Concern              151
      3          Threatened               10
```

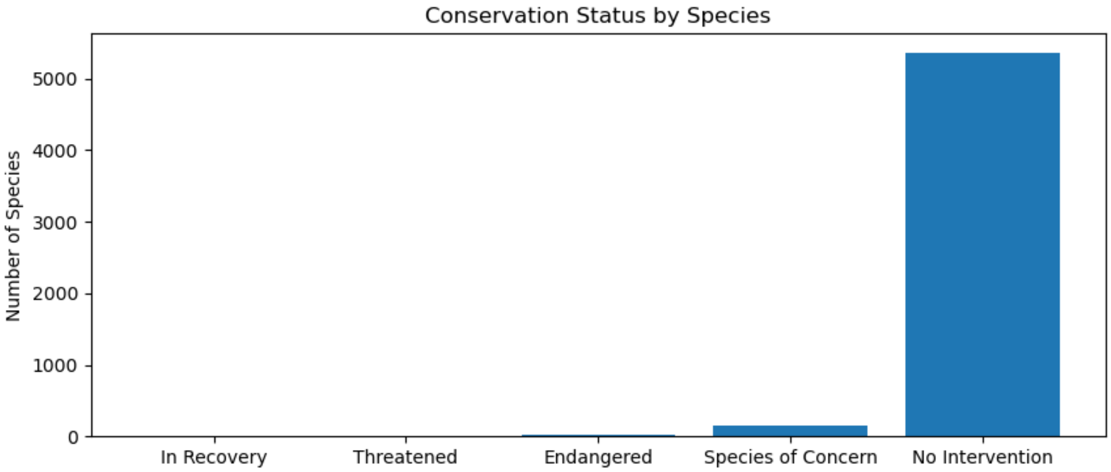
```
[11]: species.fillna('No Intervention', inplace=True)
```

```
[12]: species.groupby('conservation_status').scientific_name.nunique().reset_index()
```

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

```
[14]: protection_counts = species.groupby('conservation_status')\
          .scientific_name.nunique().reset_index()\
          .sort_values(by='scientific_name')
```

```
[15]: plt.figure(figsize=(10, 4))
      ax = plt.subplot()
      plt.bar(range(len(protection_counts)),
              protection_counts.scientific_name.values)
      ax.set_xticks(range(len(protection_counts)))
      ax.set_xticklabels(protection_counts.conservation_status.values)
      plt.ylabel('Number of Species')
      plt.title('Conservation Status by Species')
      plt.show()
```



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

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

```
[19]: category_counts.head()
```

```
[19]:     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
```

```
[20]: category_pivot = category_counts.pivot(columns='is_protected',
                                             index='category',
                                             values='scientific_name')\
                                      .reset_index()
```

```
[21]: category_pivot
```

```
[21]: is_protected            category  False  True
      0                      Amphibian     72     7
      1                           Bird    413    75
      2                           Fish    115    11
      3                         Mammal    146    30
      4              Nonvascular Plant    328     5
      5                        Reptile     73     5
      6                 Vascular Plant   4216    46
```

[22]:
```
category_pivot.columns = ['category', 'not_protected', 'protected']
```

[23]:
```
category_pivot['percent_protected'] = category_pivot.protected / \
                                      (category_pivot.protected +␣
 ↪category_pivot.not_protected)
```

[24]:
```
category_pivot
```

[24]:
```
            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
```

[25]:
```
contingency = [[30, 146],
               [75, 413]]
```

[26]:
```
from scipy.stats import chi2_contingency
```

[27]:
```
chi2_contingency(contingency)
```

[27]:
```
Chi2ContingencyResult(statistic=0.1617014831654557, pvalue=0.6875948096661336,
dof=1, expected_freq=array([[ 27.8313253, 148.1686747],
       [ 77.1686747, 410.8313253]]))
```

[28]:
```
contingency = [[30, 146],

                [5, 73]]

chi2_contingency(contingency)
```

[28]:
```
Chi2ContingencyResult(statistic=4.289183096203645, pvalue=0.03835559022969898,
dof=1, expected_freq=array([[ 24.2519685, 151.7480315],
       [ 10.7480315,  67.2519685]]))
```

[29]:
```
observations = pd.read_csv('observations.csv')

observations.head()
```

[29]:

```
            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
```

[30]:

```python
# Does "Sheep" occur in this string?
str1 = 'This string contains Sheep'
'Sheep' in str1
```

[30]:

```
True
```

[31]:

```python
# Does "Sheep" occur in this string?

str2 = 'This string contains Cows'

'Sheep' in str2
```

[31]:

```
False
```

[32]:

```python
species['is_sheep'] = species.common_names.apply(lambda x: 'Sheep' in x)

species.head()
```

[32]:

```
  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     No Intervention
1                              American Bison, Bison     No Intervention
2  Aurochs, Aurochs, Domestic Cattle (Feral), Dom…     No Intervention
3  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)     No Intervention
4                                      Wapiti Or Elk     No Intervention

   is_protected  is_sheep
0         False     False
1         False     False
2         False     False
3         False      True
4         False     False
```

[33]:

```
species[species.is_sheep]
```

[33]:

```
            category              scientific_name  \
3             Mammal                   Ovis aries   
1139  Vascular Plant             Rumex acetosella   
2233  Vascular Plant           Festuca filiformis   
3014          Mammal              Ovis canadensis   
3758  Vascular Plant             Rumex acetosella   
3761  Vascular Plant            Rumex paucifolius   
4091  Vascular Plant                 Carex illota   
4383  Vascular Plant  Potentilla ovina var. ovina   
4446          Mammal      Ovis canadensis sierrae   

                                           common_names conservation_status  \
3     Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)     No Intervention   
1139                        Sheep Sorrel, Sheep Sorrell     No Intervention   
2233                              Fineleaf Sheep Fescue     No Intervention   
3014                       Bighorn Sheep, Bighorn Sheep  Species of Concern   
3758  Common Sheep Sorrel, Field Sorrel, Red Sorrel,…     No Intervention   
3761   Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock     No Intervention   
4091                       Sheep Sedge, Smallhead Sedge     No Intervention   
4383                                   Sheep Cinquefoil     No Intervention   
4446                        Sierra Nevada Bighorn Sheep          Endangered   

      is_protected  is_sheep  
3            False      True  
1139         False      True  
2233         False      True  
3014          True      True  
3758         False      True  
3761         False      True  
4091         False      True  
4383         False      True  
4446          True      True  
```

[34]:

```
sheep_species = species[(species.is_sheep) & (species.category == 'Mammal')]

sheep_species
```

[34]:

```
     category          scientific_name  \
3      Mammal               Ovis aries   
3014   Mammal          Ovis canadensis   
4446   Mammal  Ovis canadensis sierrae   

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

      is_protected  is_sheep
3            False      True
3014          True      True
4446          True      True
```

[35]:

```
sheep_observations = observations.merge(sheep_species)

sheep_observations
```

[35]:

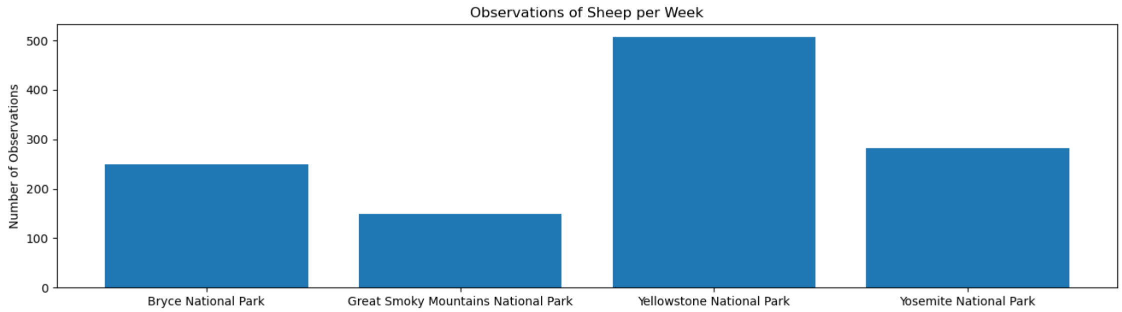
```
             scientific_name                            park_name  \
0            Ovis canadensis            Yellowstone National Park
1            Ovis canadensis                  Bryce National Park
2            Ovis canadensis               Yosemite National Park
3            Ovis canadensis  Great Smoky Mountains National Park
4    Ovis canadensis sierrae            Yellowstone National Park
5    Ovis canadensis sierrae               Yosemite National Park
6    Ovis canadensis sierrae                  Bryce National Park
7    Ovis canadensis sierrae  Great Smoky Mountains National Park
8                 Ovis aries               Yosemite National Park
9                 Ovis aries  Great Smoky Mountains National Park
10                Ovis aries                  Bryce National Park
11                Ovis aries            Yellowstone National Park

    observations category                                       common_names  \
0            219   Mammal                       Bighorn Sheep, Bighorn Sheep
1            109   Mammal                       Bighorn Sheep, Bighorn Sheep
2            117   Mammal                       Bighorn Sheep, Bighorn Sheep
3             48   Mammal                       Bighorn Sheep, Bighorn Sheep
4             67   Mammal                        Sierra Nevada Bighorn Sheep
5             39   Mammal                        Sierra Nevada Bighorn Sheep
6             22   Mammal                        Sierra Nevada Bighorn Sheep
7             25   Mammal                        Sierra Nevada Bighorn Sheep
8            126   Mammal  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)
9             76   Mammal  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)
10           119   Mammal  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)
11           221   Mammal  Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)

   conservation_status  is_protected  is_sheep
0   Species of Concern          True      True
1   Species of Concern          True      True
2   Species of Concern          True      True
3   Species of Concern          True      True
```

```
4           Endangered          True      True
5           Endangered          True      True
6           Endangered          True      True
7           Endangered          True      True
8      No Intervention         False      True
9      No Intervention         False      True
10     No Intervention         False      True
11     No Intervention         False      True
```

```
[36]: obs_by_park = sheep_observations.groupby('park_name').observations.sum().
       ↪reset_index()

      obs_by_park
```

```
[36]:                              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
```

```
[37]: plt.figure(figsize=(16, 4))

      ax = plt.subplot()

      plt.bar(range(len(obs_by_park)),

              obs_by_park.observations.values)

      ax.set_xticks(range(len(obs_by_park)))

      ax.set_xticklabels(obs_by_park.park_name.values)

      plt.ylabel('Number of Observations')

      plt.title('Observations of Sheep per Week')

      plt.show()
```



```
[38]: minimum_detectable_effect = 100 * 0.05 / 0.15

      minimum_detectable_effect
```

```
[38]: 33.333333333333336
```

```
[39]: baseline = 15
```

```
[40]: sample_size_per_variant = 870

      # Note: This could be 890 if you used 33% for the "Minimum Detectable Effect" instead of 33.33%.  That's fine.
```

```
[41]: sample_size_per_variant = 870

      # Note: This could be 890 if you used 33% for the "Minimum Detectable Effect" instead of 33.33%.  That's fine.
```

```
[42]: bryce = 870 / 250.

      yellowstone = 810 / 507.

      # Approximately 3.5 weeks at Bryce and 1.5 weeks at Yellowstone.
```

```
[ ]:
```