Capstone Project: Biodiversity in National Parks

J. ANTHONY ACKER

Agenda

- Description of The Data in Species_info.csv
- Significance Calculations
- Sample Size Determination For Analysis of Effectiveness of Hoof and Mouth Disease Reduction Program




Description of The Data in Species_info.csv

Shape of Species_info.csv

```
species.head()
```

	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

```
species.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5824 entries, 0 to 5823
Data columns (total 4 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   category             5824 non-null   object
 1   scientific_name      5824 non-null   object
 2   common_names         5824 non-null   object
 3   conservation_status  191 non-null    object
dtypes: object(4)
memory usage: 182.1+ KB
```

```
print(species.shape)
```

```
(5824, 4)
```

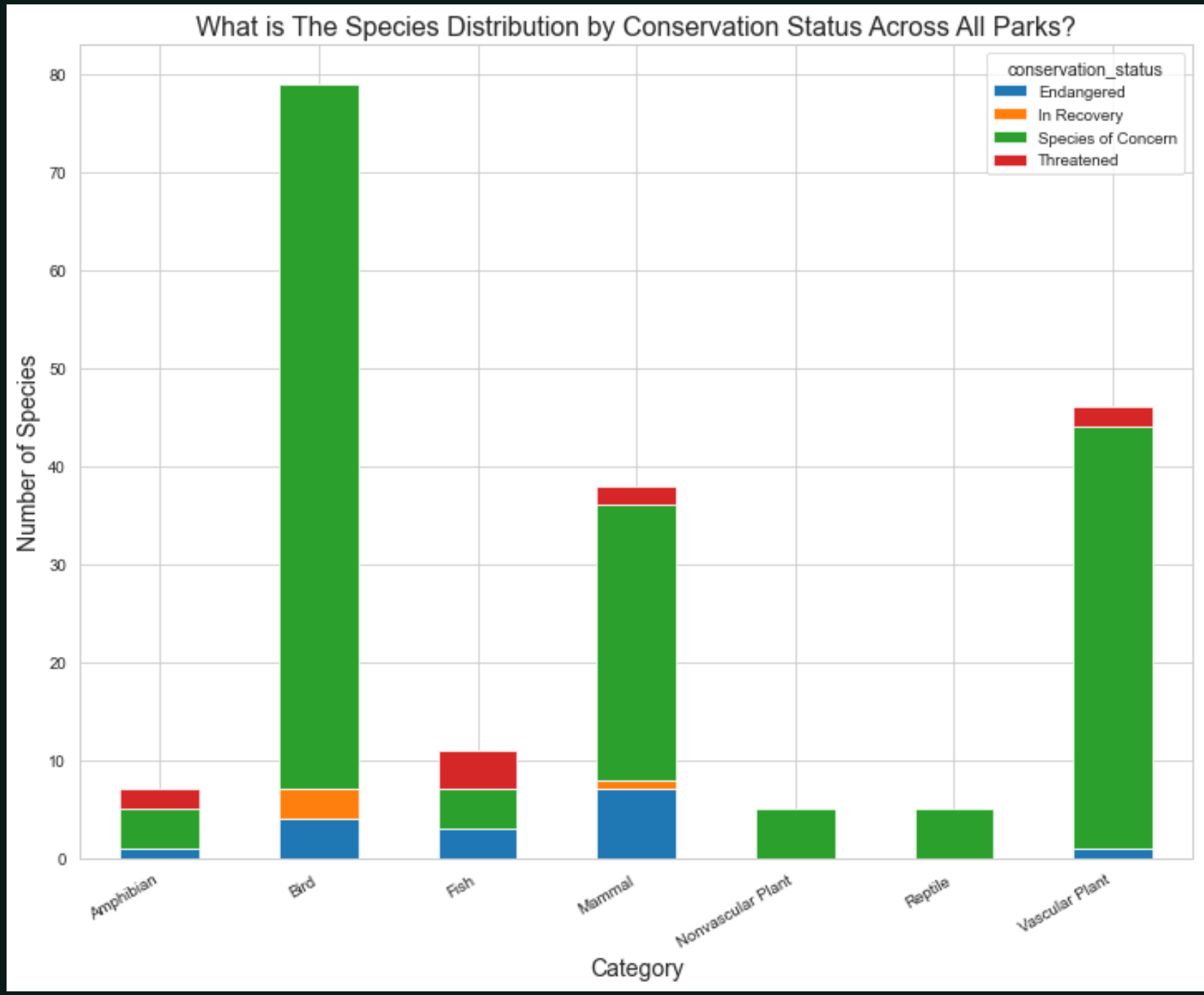
Details of Species_info.csv

How many different species are in the dataset?	5541
What are the different values of category in species?	['Mammal' 'Bird' 'Reptile' 'Amphibian' 'Fish' 'Vascular Plant' 'Nonvascular Plant']
What are the different values of conservation_status?	[nan 'Species of Concern' 'Endangered' 'Threatened' 'In Recovery']

What is The Distribution of Category Across Conservation_status?

conservation_status category	Endangered	In Recovery	Species of Concern	Threatened
Amphibian	1	0	4	2
Bird	4	3	72	0
Fish	3	0	4	4
Mammal	7	1	28	2
Nonvascular Plant	0	0	5	0
Reptile	0	0	5	0
Vascular Plant	1	0	43	2

What is The Species Distribution by Conservation_status Across All Parks?

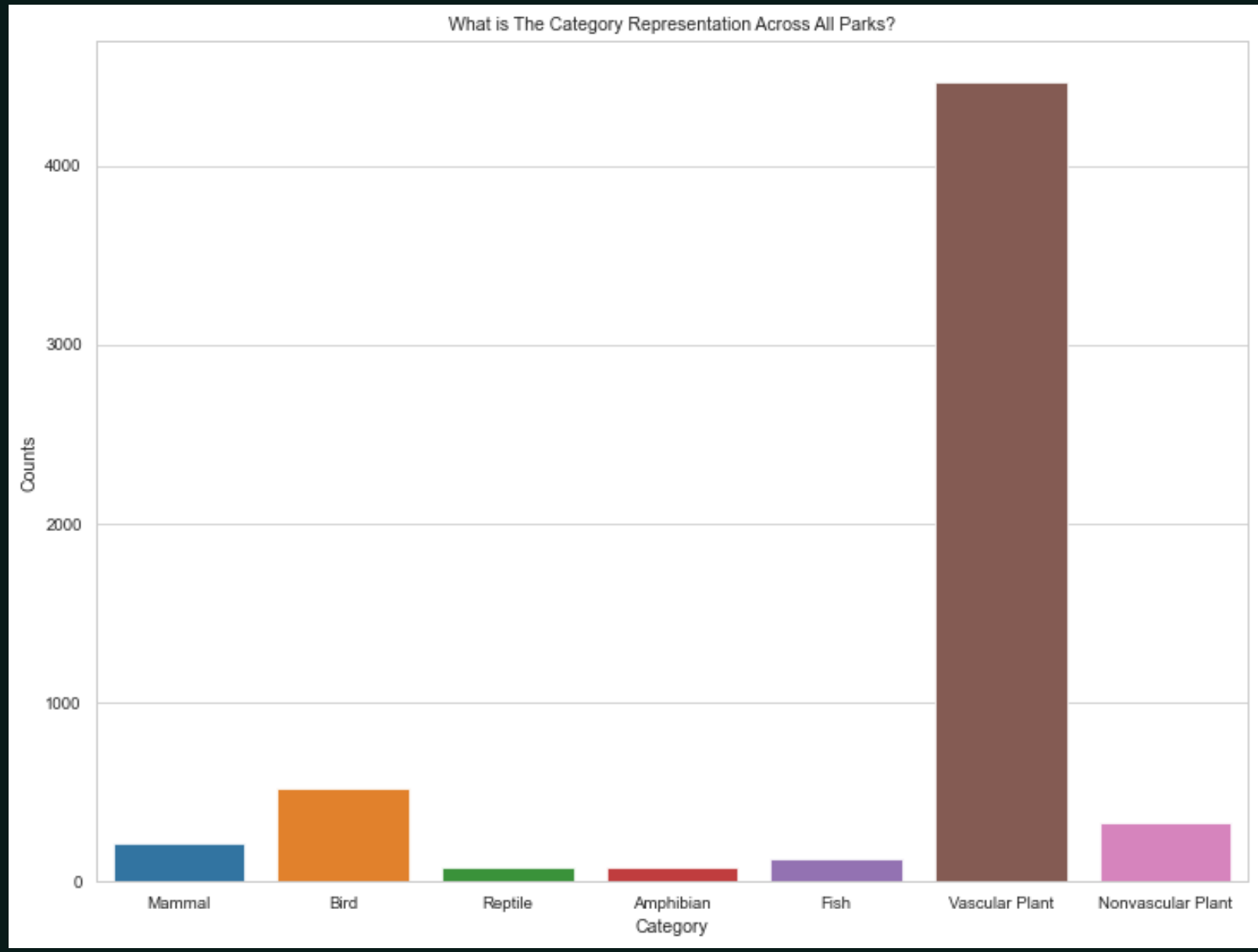


Observations

- Birds are vastly the more represented species of those considered protected species.
- Birds have the most species in the 'Species of Concern' conservation_status. Followed by 'Vascular Plants'.
- Birds are the species most likely to be 'In Recovery'.
- Mammals have the most species that are 'Endangered'.
- Fish are the species most likely to be 'Threatened'.
- All categories of species have a high amount of 'Species of Concern'. In fact, it is the number one or two most likely conservation_status for a given category of species. In the case of the 'Nonvascular Plant' and the 'Reptile' categories, it is the only conservation_status for those species.

What is The Category Representation Across All Parks?

Observation: 'Vascular Plant' is the most represented species in the dataset.



How Many Species Are Protected?

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

Out[12]:

	conservation_status	scientific_name
0	Endangered	16
1	In Recovery	4
2	Species of Concern	161
3	Threatened	10

As we saw before, there are far more than 200 species in the species table. Clearly, only a small number of them are categorized as needing some sort of protection. The rest have conservation_status equal to None. Because groupby does not include None, we will need to fill in the null values. We can do this using .fillna. We pass in however we want to fill in our None values as an argument.

In [13]:
```
species.fillna('No Intervention', inplace=True)
```

In [14]:
```
species.groupby(['conservation_status'])['scientific_name'].count().reset_index()
```

Out[14]:

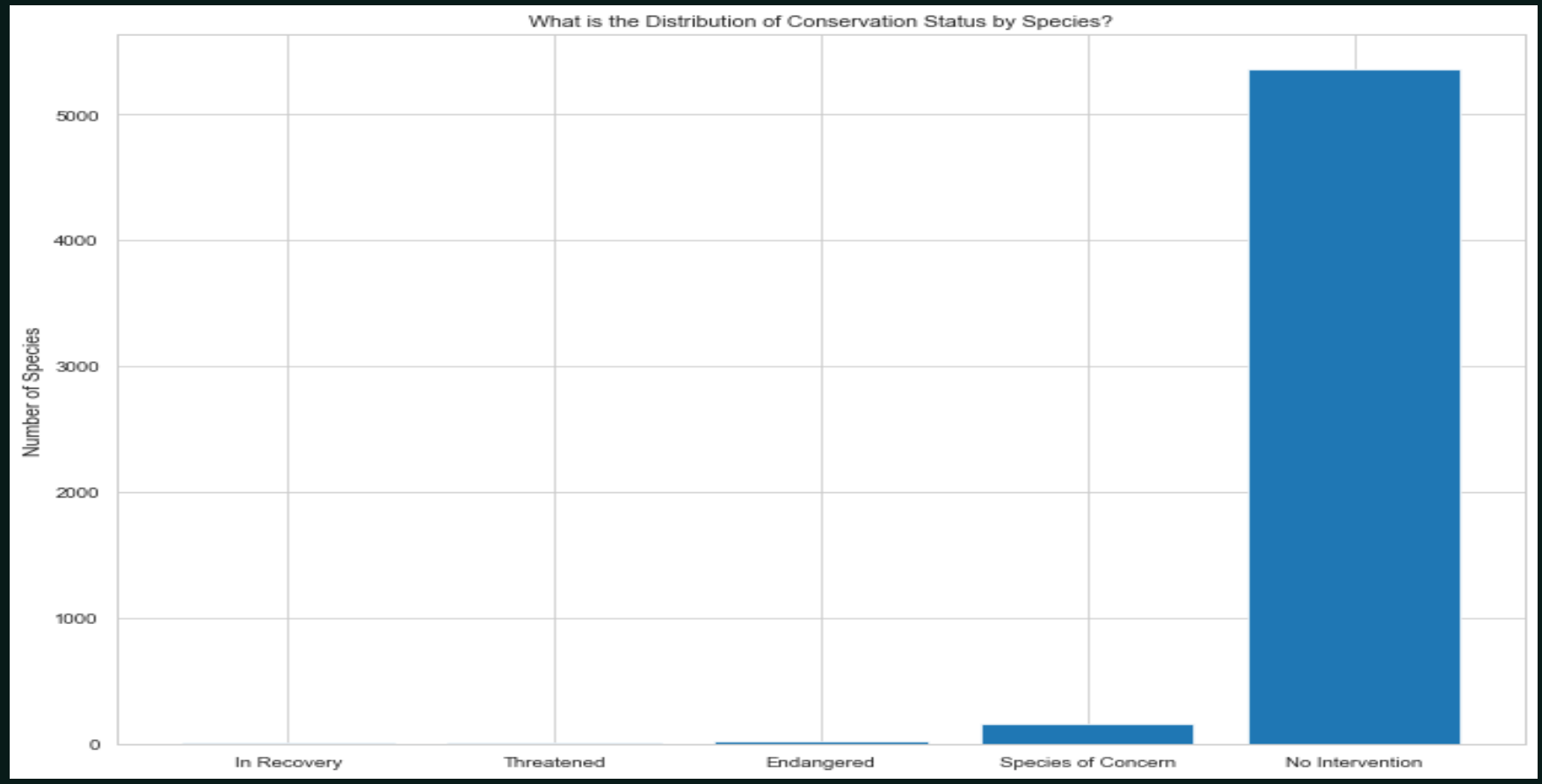
	conservation_status	scientific_name
0	Endangered	16
1	In Recovery	4
2	No Intervention	5633
3	Species of Concern	161
4	Threatened	10

In [15]:
```
no_intervention_percentage = 5633 / 5824 * 100
no_intervention_percentage
```

Out[15]:
```
96.72046703296702
```

Observation: 'No Intervention' is the most represented conservation_status in the dataset. 96.72% of the species in the dataset are not protected.

What is the Distribution of Conservation_status by Species?



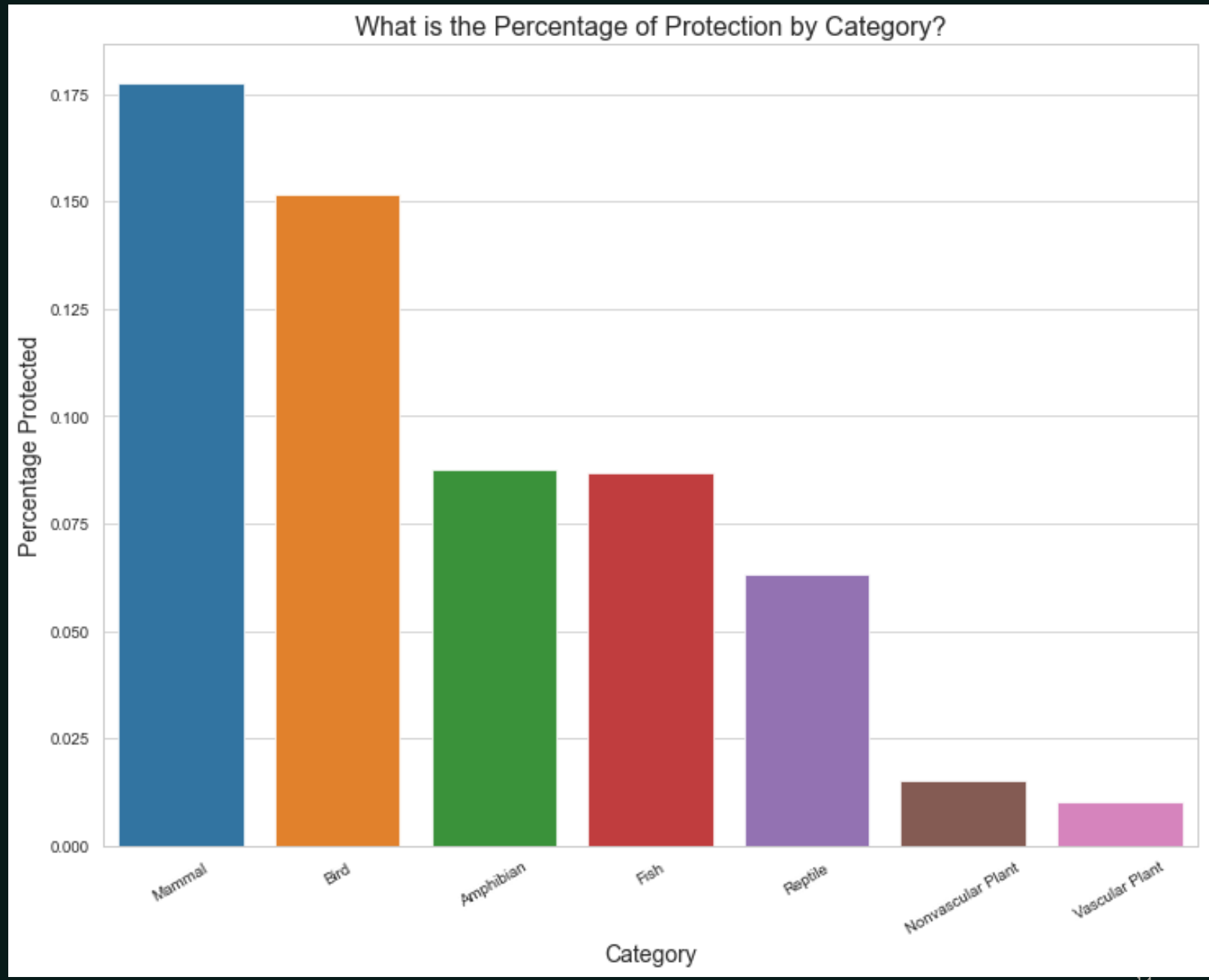
Are Certain Types Of Species More Likely To Be Protected?

	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

What is the Percentage of Protection by Category?

It looks like species in category 'Mammal' are more likely to be protected than species in 'Bird'. We're going to do a significance test to see if this statement is true.

The data is categorical and there are two pieces of data to test. I would recommend the chi squared test.



Chi Square Test

Are Certain Types Of Species More Likely To Be Protected?

category	protected	not_protected
Mammal	38	176
Bird	79	442

Pvalue = 0.445901703047197

It looks like this difference isn't significant!

category	protected	not_protected
Mammal	38	176
Reptile	5	74

Pvalue = 0.023384652148715 47

It looks like there is a significant difference between Reptile and Mammal!

Sample Size Determination For Analysis of Effectiveness of Hoof and Mouth Disease Reduction Program

Are We Interested in Everything Called 'Sheep'?

Not all species with 'Sheep' in their name are the mammals we are interested in studying.

	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,...	No Intervention	False	True
3761	Vascular Plant	Rumex paucifolius	Alpine Sheep Sorrel, Fewleaved Dock, Meadow Dock	No Intervention	False	True
4091	Vascular Plant	Carex illota	Sheep Sedge, Smallhead Sedge	No Intervention	False	True
4383	Vascular Plant	Potentilla ovina var. ovina	Sheep Cinquefoil	No Intervention	False	True
4446	Mammal	Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered	True	True

Observation: Many of the results are actually plants.

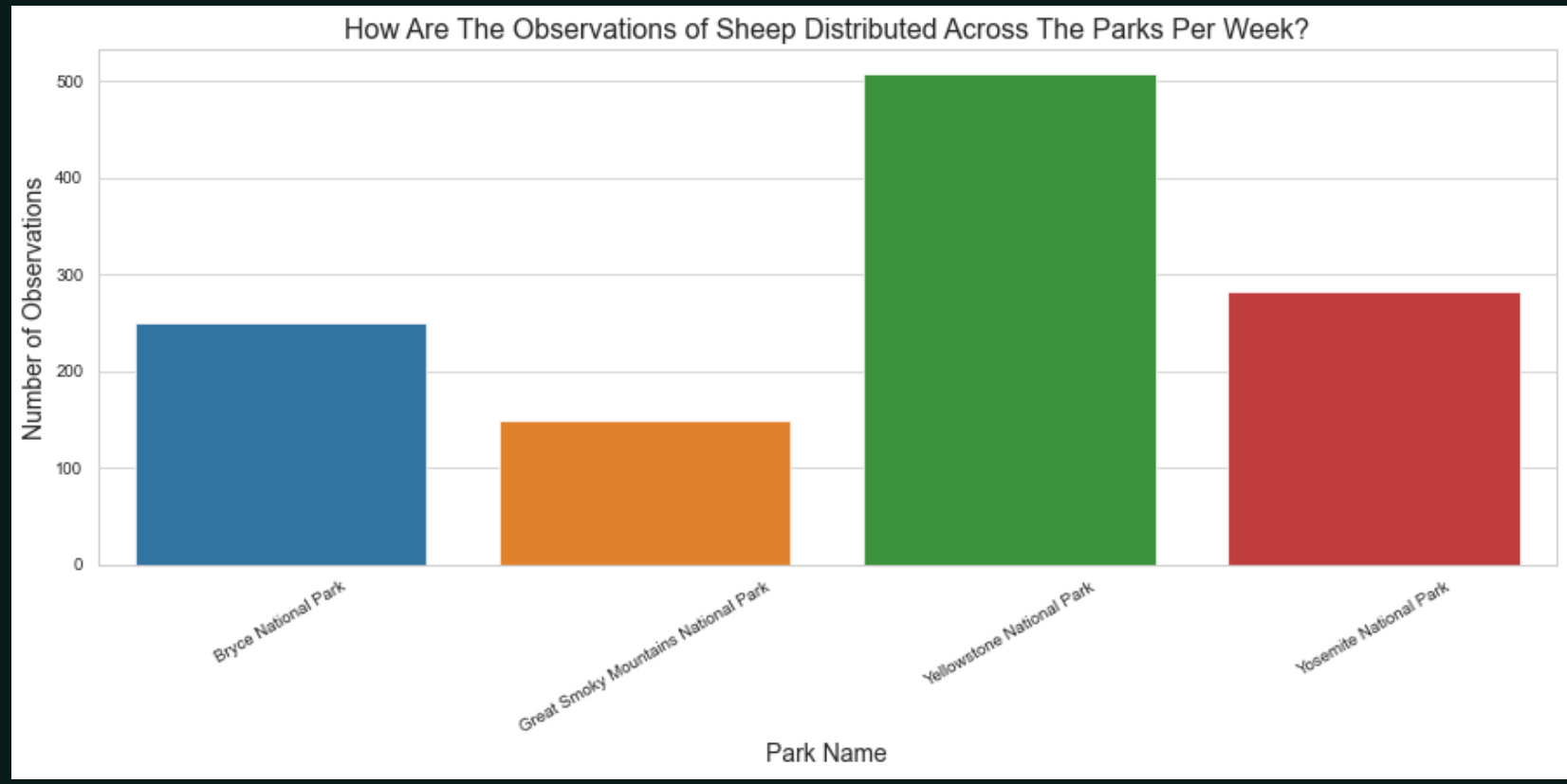
```
sheep_species = species[(species.is_sheep) & (species.category == 'Mammal')]
sheep_species
```

	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
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

What Are The Sheep Observations Across All Parks?

	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
5	Ovis canadensis sierrae	Yosemite National Park	39	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
6	Ovis canadensis sierrae	Bryce National Park	22	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
7	Ovis canadensis sierrae	Great Smoky Mountains National Park	25	Mammal	Sierra Nevada Bighorn Sheep	Endangered	True	True
8	Ovis aries	Yosemite National Park	126	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
9	Ovis aries	Great Smoky Mountains National Park	76	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
10	Ovis aries	Bryce National Park	119	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True
11	Ovis aries	Yellowstone National Park	221	Mammal	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention	False	True

How Are The Observations of Sheep Distributed Across The Parks Per Week?

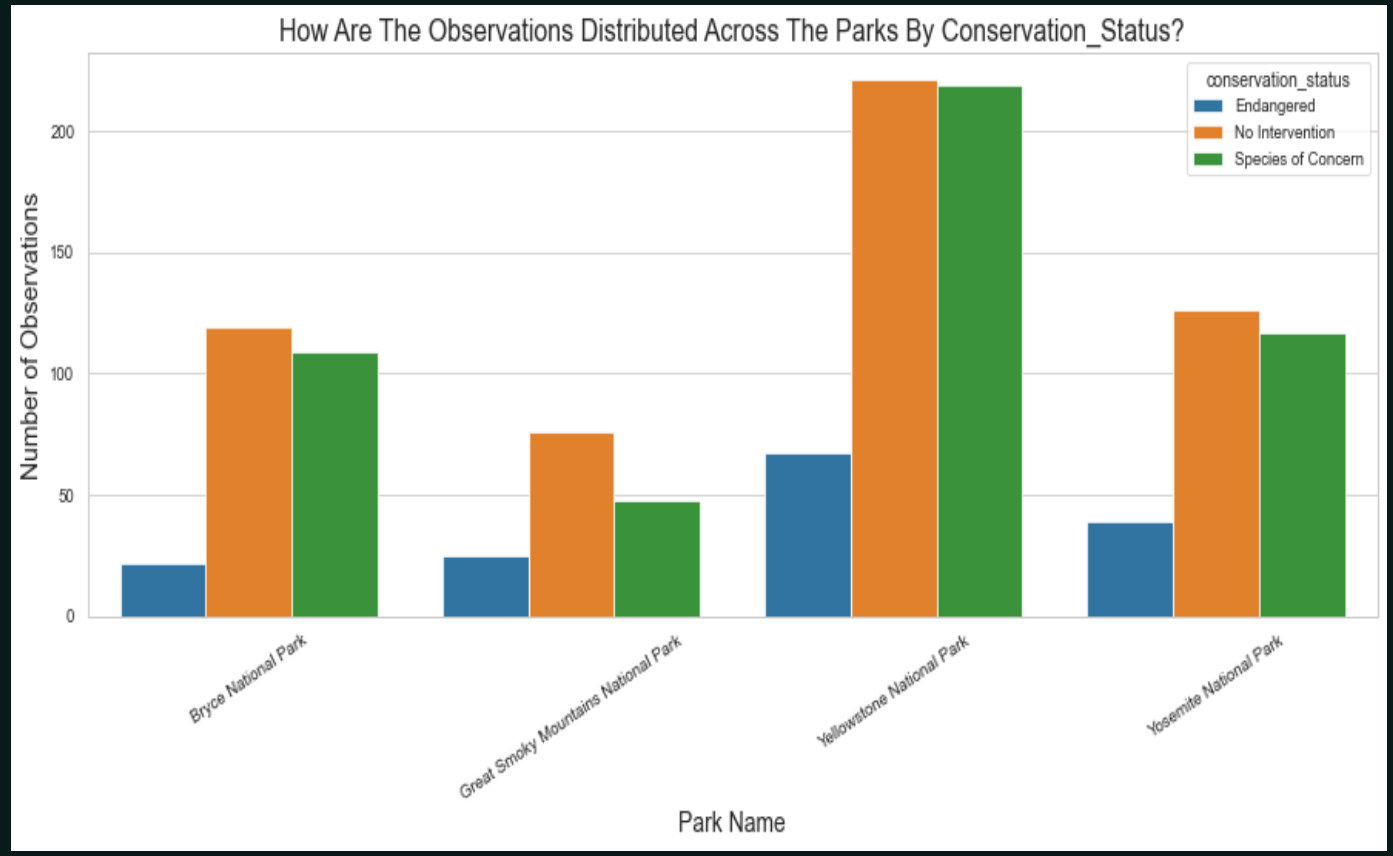


YELLOWSTONE NATIONAL PARK HAS THE MOST OBSERVATIONS.

How Are The Observations Distributed Across The Parks By Conservation_status?

Observations:

- Most of the sheep observed did not have protected status.
- Most of the protected sheep are 'Species of Concern'.
- The balance of the protected sheep are 'Endangered'.
- Yellowstone National Park has the most observations of sheep and the highest number of protected sheep.



```
minimum_detectable_effect = 100 * 0.05 / 0.15

baseline = 15
sample_size = 870
minimum_detectable_effect
```

```
33.333333333333336
```

How many weeks would be needed to observe sheep at Bryce National Park in order to observe enough sheep? How many weeks would be needed to observe at Yellowstone National Park to observe enough sheep?

```
weeks_to_observe_bryce = sample_size / 250
weeks_to_observe_yellowstone = sample_size / 507

print("weeks_to_observe_bryce = {:.2f}".format( weeks_to_observe_bryce ))
print("weeks_to_observe_yellowstone = {:.2f}".format( weeks_to_observe_yellowstone ))
```

```
weeks_to_observe_bryce = 3.48
weeks_to_observe_yellowstone = 1.72
```

Prepare to Analyze Effectiveness of Hoof and Mouth Disease Reduction Program

Recommendation: Bryce National Park should collect sheep data from the park for about 3.48 weeks and at Yellowstone National Park should collect sheep data for roughly 1.72 weeks.

Thank you

J. Anthony Acker

codingsexagenarian@gmail.com