Biodiversity Capstone Project

Investigating Protected Species

Author: Bastian Weber

Describe the data in species_info.csv

	category	scientific_name	common_names	conservation_status
1	Mammel	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	NaN
2	Mammel	Bos bison	American Bison, Bison	NaN
3	Mammel	Bos taurus	Aurochs, Aurochs, Domestic Cattle (Feral), Dom...	NaN
4	Mammel	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN

There are 5.541 different species in the DataFrame that can be categorized into 7 distinct species types.

```
['Mammal' 'Bird' 'Reptile' 'Amphibian' 'Fish' 'Vascular Plant'
 'Nonvascular Plant']
```

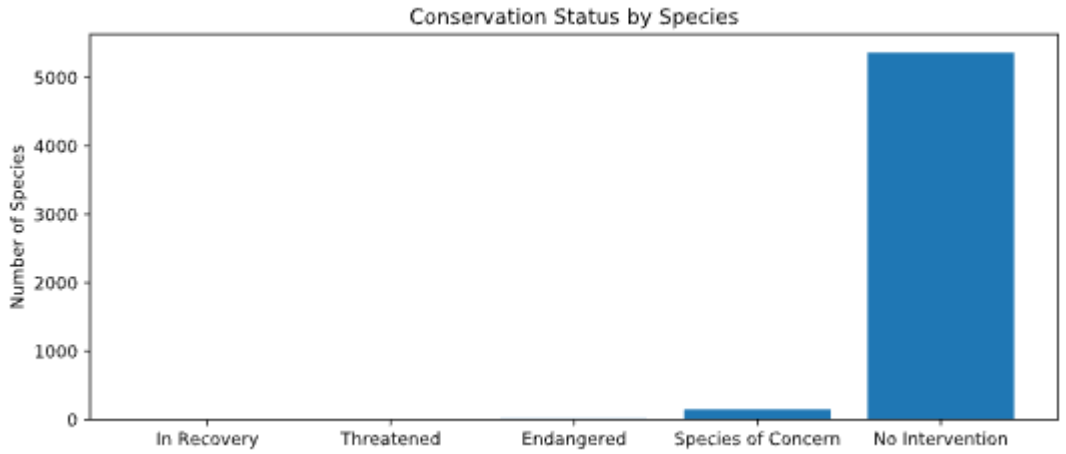
Species are grouped into 5 types by conservation status:

```
['No Intervention' 'Species of Concern' 'Endangered' 'Threatened'
 'In Recovery']
```

Describe the data in species_info.csv

The aim of the this project is to get insights about endangered animals. The table below displays the sum of scientific names grouped by the conservation status.

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



It shows that a big amount of animals come with an „No Intervention“ status. That is, only a small number of them are categorized as needing some sort of protection.

Describe the significance calculations that you did for endangered status between different categories of species

However, when pivoting the endangered species data with percent protected estimation per category, more valuable information is gathered.

	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

The table shows that Mammals are more likely to be endangered than other categories.

The Chi square test says that birds are almost in the same endangered level as Mammals as there's no significant difference between the datasets (p-value of 0.688 and hence > 0.05). Significance is given, however, when comparting Mammals with Reptiles (p-value of 0.038 and thus < 0.05).

```
0.687594809666
0.0383555902297
```

Give recommendations for conservationists concerned about endangered species, based on your significance calculations

Conservationists are worried as they have been recording sightings of different species at several national parks for the past 7 days. The dataset they provided is found below:

	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

The number of sheep sightings at different national parks are studied. By using lambda and group functions, a table with all observations per park is displayed below. Yellowstone National Park registered the highest amounts of sheep observations.

	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

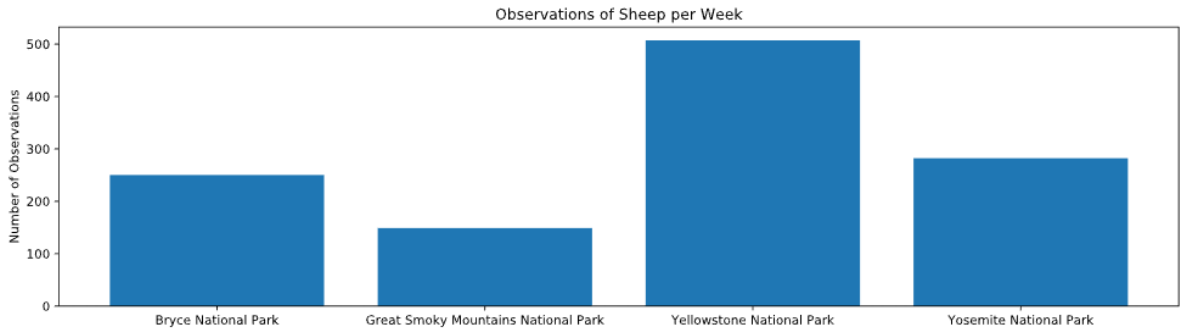
Give recommendations for conservationists concerned about endangered species, based on your significance calculations

Conservationists are worried as they have been recording sightings of different species at several national parks for the past 7 days. The dataset they provided is found below:

	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

The number of sheep sightings at different national parks are studied. By using lambda and group functions, a table with all observations per park is displayed below. Yellowstone National Park registered the highest amounts of sheep observations.

	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



Describe the sample size determination that you did for the foot and mouth disease study

Scientists found a 15% occurrence of foot and mouth disease in sheep at Bryce National Park. If the scientists wanted to be sure that a >5% drop in observed cases of foot and mouth disease in the sheep at Yellowstone was significant they would have to observe at least 870 sheep.

This would approximately take one week of observing in Yellowstone to see that many sheep, or approximately two weeks in Bryce to see that many sheep.

```
print minimum_detectable_effect     33.3333333333
print sample_size_per_variant       870
print yellowstone_weeks_observing   1.71597633136
print bryce_weeks_observing         3.48
```