Biodiversity
Capstone project

NIKITA BOYKO
Data analyst

Source data and purpose of research

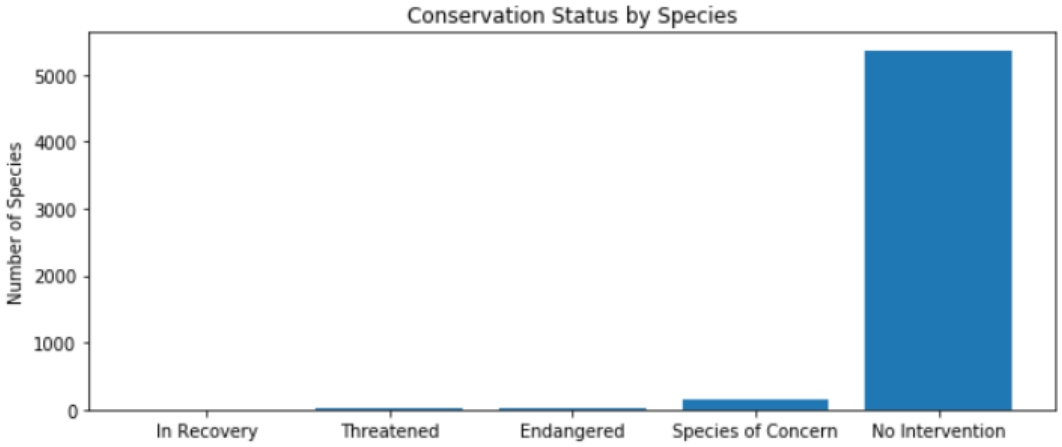
National Park Service provided a spreadsheet to analyze the variety of species and check if there are any of them in danger.

Species (5824, 5541 unique) are divided into categories (7), each of them has common and scientific name. Also some of species have an attribute ‘conservation status’ which reflects a risk 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), Dom...	NaN
3	Mammal	Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	NaN
4	Mammal	Cervus elaphus	Wapiti Or Elk	NaN

How many species with each conservation status?

For the most it’s ‘No intervention’



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

Which particular types of species are more likely to be endangered?

Plants have to be need protection more than other categories

category	not_protected	protected	percent_protected
Amphibian	72	7	0.088608
Bird	413	75	0.153689
Fish	115	11	0.087302
Mammal	146	30	0.170455
Nonvascular Plant	328	5	0.015015
Reptile	73	5	0.064103
Vascular Plant	4216	46	0.010793

Chi squared test:

no significant difference between Mammals and Birds (pval > 0.05)
significant difference between Mammals and Reptiles (pval < 0.05)

Loading an additional spreadsheet which contains recording sightings of species for the last 7 days:

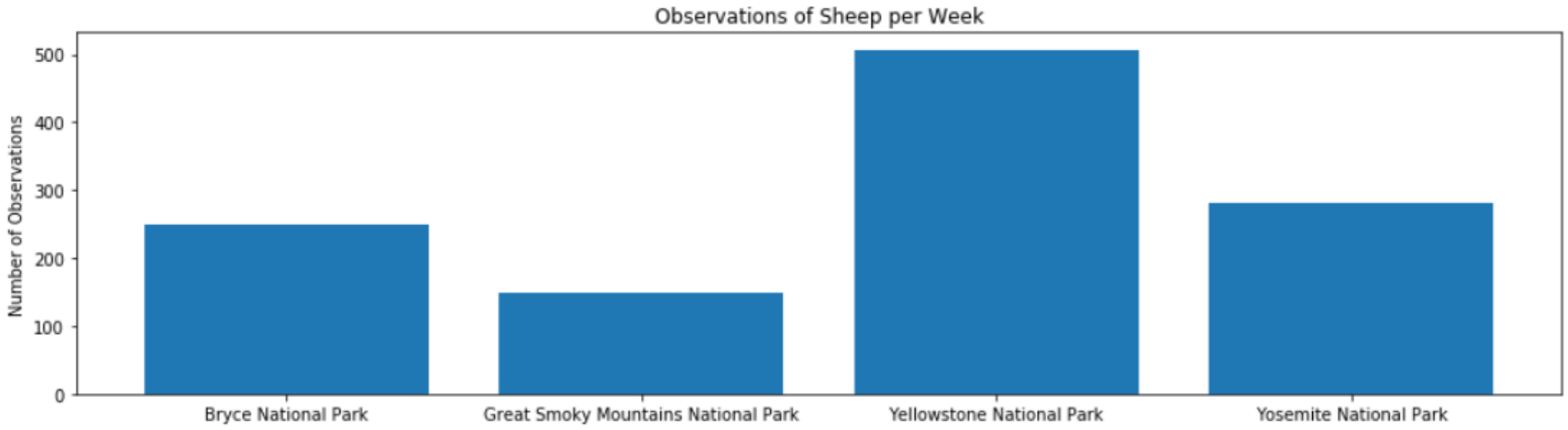
	scientific_name	park_name	observations
0	Vicia benghalensis	Great Smoky Mountains National Park	68
1	Neovison vison	Great Smoky Mountains National Park	77
2	Prunus subcordata	Yosemite National Park	138
3	Abutilon theophrasti	Bryce National Park	84
4	Githopsis specularioides	Great Smoky Mountains National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

We're really interested what's wrong with the sheep. Let's merge all the data for sheep into one spreadsheet:

	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

In which park you could observe a sheep with the highest probability?

Yellowstone National Park



park_name	observations
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

Using sample size calculator

minimum_detectable_effect = 100 * 0.05 / 0.15 = 33.3
baseline conversion rate 15%
level of significance 90%

Some sheep have diseases. How many sheep we have to observe to make sure their foot and mouth percentages are significant?

510

How many weeks will it take for Bryce and Yellowstone?

Bryce Park ~2 weeks
Yellowstone Park ~1 week

Thanks for your attention!