

Biodiversity in National Parks

Data Analysis Portfolio Project



Table of Contents

- 1 Description of the available data from the National Parks
- 2 Distribution of the conservation status of the species across all parks
- 3 Identification of families of species more likely to be endangered
- 4 Exploration of possible association between the species' families and its conservation status
- 5 Analysis of observations of species in each park
- 6 Appendices

Description of the available data from the National Parks

Observations database¹ ('observations.csv')

- Number of total entries: 23,269

scientific_name	park_name	observations
Vicia benghalensis	Great Smoky Mountains National Park	66
Neovison vison	Great Smoky Mountains National Park	77
Frunus subcordata	Yosemite National Park	138
Abutilon theophrasti	Bryce National Park	84
Githopsis specularioides	Great Smoky Mountains National Park	85

- Max. number of obs./entry: 321
- Min. number of obs./entry: 9

- Number of unique status: 5
- Name of unique status: 'No Intervention', 'Species of Concern', 'Endangered', 'Threatened', 'In Recovery'

- Number of unique species: 5,541

- Number of unique parks: 4
- Name of unique parks: 'Great Smoky Mountains National Park', 'Yosemite National Park', 'Bryce National Park', 'Yellowstone National Park'

- Total number of observations: 3,314,739

- Number of unique species: 5,541

- Number of total entries: 5,824

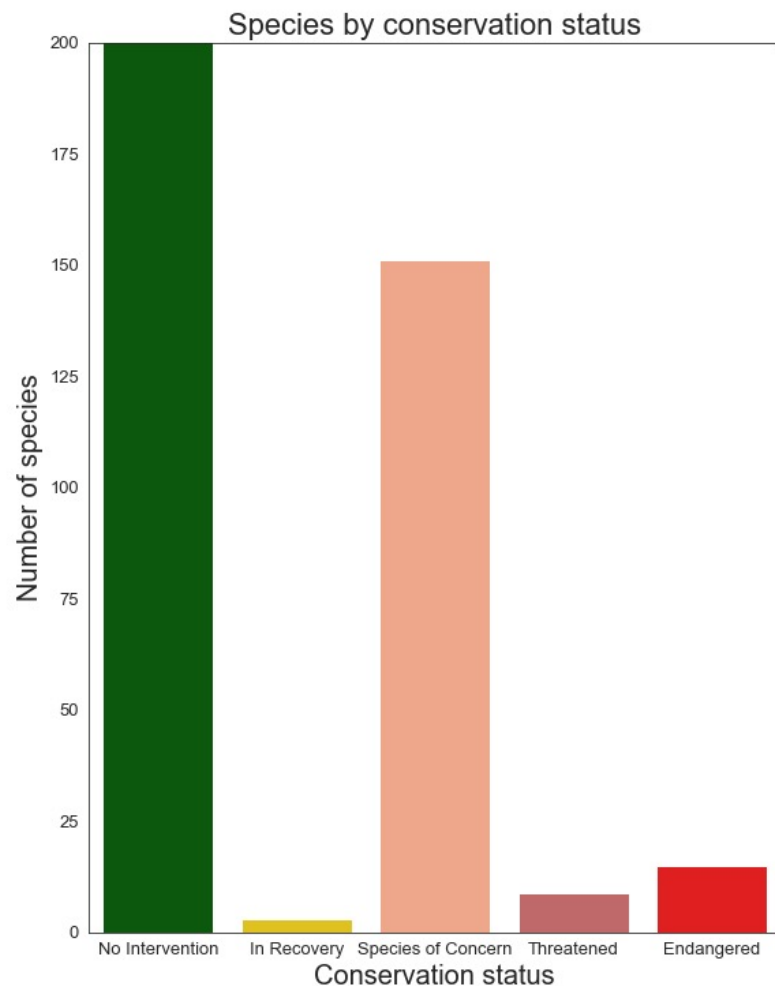
Species database¹ ('species_info.csv')

category	scientific_name	common_name	conservation_status
Mammal	Clethrionomys gapperi gapperi	Gapper's Red-Backed Vole	No Intervention
Mammal	Bos bison	American Bison, Bison	No Intervention
Mammal	Bos taurus	Aurochs	No Intervention
Mammal	Ovis aries	Domestic Sheep	No Intervention
Mammal	Cervus elaphus	Wapiti Or Elk	No Intervention

- Number of unique categories: 7
- Name of unique parks: 'Mammal', 'Bird', 'Reptile', 'Amphibian', 'Fish', 'Vascular Plant', 'Nonvascular Plant'

1. Showing first 5 dataframe rows

Distribution of the conservation status of the species across all parks



of species per category classified by status (excl. 'No Intervention')¹

Cons. Status	Amphibian	Bird	Fish	Mammal	Nonvascular Plant	Reptile	Vascular Plant	Total
In Recovery	0	3	0	0	0	0	0	3
Sp. of Concern	4	68	4	22	5	5	43	151
Threatened	2	0	3	2	0	0	2	9
Endangered	1	4	3	6	0	0	1	15
Total	7	75	10	30	5	5	46	178

Conclusions drawn from working with the data:

- There are 178 species 'at some degree of danger' (excluding 'No Intervention'). In this group, 151 are 'Species of Concern' (85%). The endangered species – the most critical group – represent 15 species in total, or 8% of the total species 'at some degree of danger'
- The birds are the largest category (75 species in total), most of them being 'Species of Concern' (68 in total, or 91% of the category). The second largest group are the vascular plants (46 species in total), most of them also in the 'Species of Concern' group (43 out of 46 in total, or 93%)
- For the mammals (30 species 'at some degree of danger'), 8 species are 'Threatened' or 'Endangered' (27% of this subset), making it the most concerning category

1. Refer to Appendix 1 for visual breakdown

Identification of families of species more likely to be endangered

of species per category classified by Protected /Not Protected

Category	Not protected	Protected	% Protected
Amphibian	72	7	10%
Bird	413	75	18%
Fish	115	10	9%
Mammal	146	30	21%
Nonvascular Plant	328	5	2%
Reptile	73	5	7%
Vascular Plant	4216	46	1%
Total	5363	178	32%

‘Protected’ = ‘In Recovery’ + ‘Species of Concern’ + ‘Threatened’
 + ‘Endangered’
‘Not Protected’ = ‘No Intervention’

Conclusions drawn from the data:

- The species in the ‘No Intervention group’ are the ones with a certain level of risk because currently there is no protection program established for them (the parks assume that this protection is currently not needed), yet this still must be checked
- The category which has the largest number of species under protection are the vascular plants (46 species), yet the mammals and the birds are the category with the biggest percentage of species currently protected (21% and 18% respectively)
- The plants (vascular and nonvascular) have less than 2% of the total species under some protection program, yet since they are not animals, they appear to be less affected by external effects in their surroundings

Exploration of possible association between the species' families and its conservation status

Results of p-values for test between 'level of protection' and category (all possible combinations)

Test Parameters:

- Type of test: chi-square test
- Variables involved: 'Category' and 'Level of protection'¹
- P-value threshold: 0.05 (less indicates significant association between variables, in **red**)

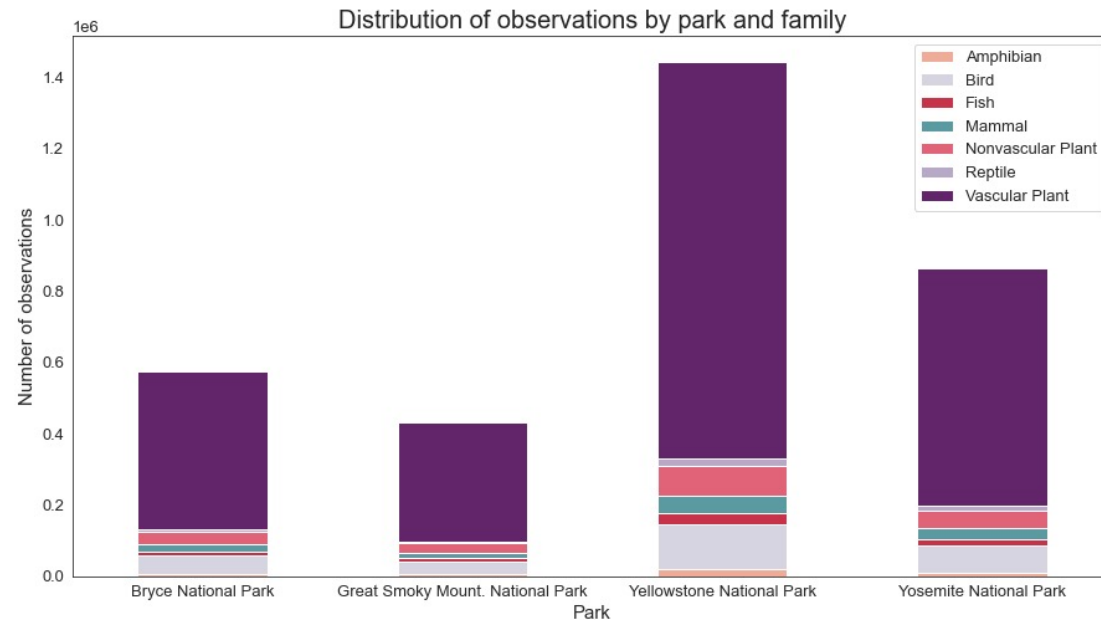
Category	Amphibian	Bird	Fish	Mammal	Non Vasc. Plant	Reptile	Vascular Plant
Amphibian		0.1759	0.9654	0.1276	0.0018	0.7815	0.0000
Bird	0.1759		0.0475	0.6876	0.0000	0.0531	0.0000
Fish	0.9654	0.0475		0.0352	0.0014	0.8844	0.0000
Mammal	0.1276	0.6876	0.0352		0.0000	0.0384	0.0000
Non Vasc. Plant	0.0018	0.0000	0.0014	0.0000		0.0336	0.6623
Reptile	0.7815	0.0531	0.8844	0.0384	0.0336		0.0001
Vasc. Plant	0.0000	0.0000	0.0000	0.0000	0.6623	0.0001	

Conclusions drawn from the data:

- The plants (both vascular and non-vascular) show the smallest p-values in tests against animal categories (therefore making a case for a strong association between its level of protection compared to other species). This makes sense since they are less susceptible to changes in their environment, therefore requiring less protection (around 1% - 2% from the previous slide)
- Regarding animals, the fishes show the smallest p-values when compared to other animal species (vs. mammals and birds). One theory to explain why they are less protected than other species can be its reproduction patterns and variety (usually fishes lay eggs by the thousands)

1. Level of protection: number of species per category either 'Protected' or 'Not protected' as table from Slide 5

Analysis of observations of species in each park



Most spotted species in each park

park	name	scientific_name	number_obs	category
Bryce Nat. Park	Rock Dove	Columba livia	339	Bird
Great Smoky Mount. Nat. Park	Eurasian Collared-Dove	Streptopelia decaocto	256	Bird
Yellowstone	Common Velvetgrass	Holcus lanatus	805	Vascular Plant
Yosemite	Spotted Cat's Ear	Hypochaeris radicata	505	Vascular Plant

Conclusions drawn from working with the data:

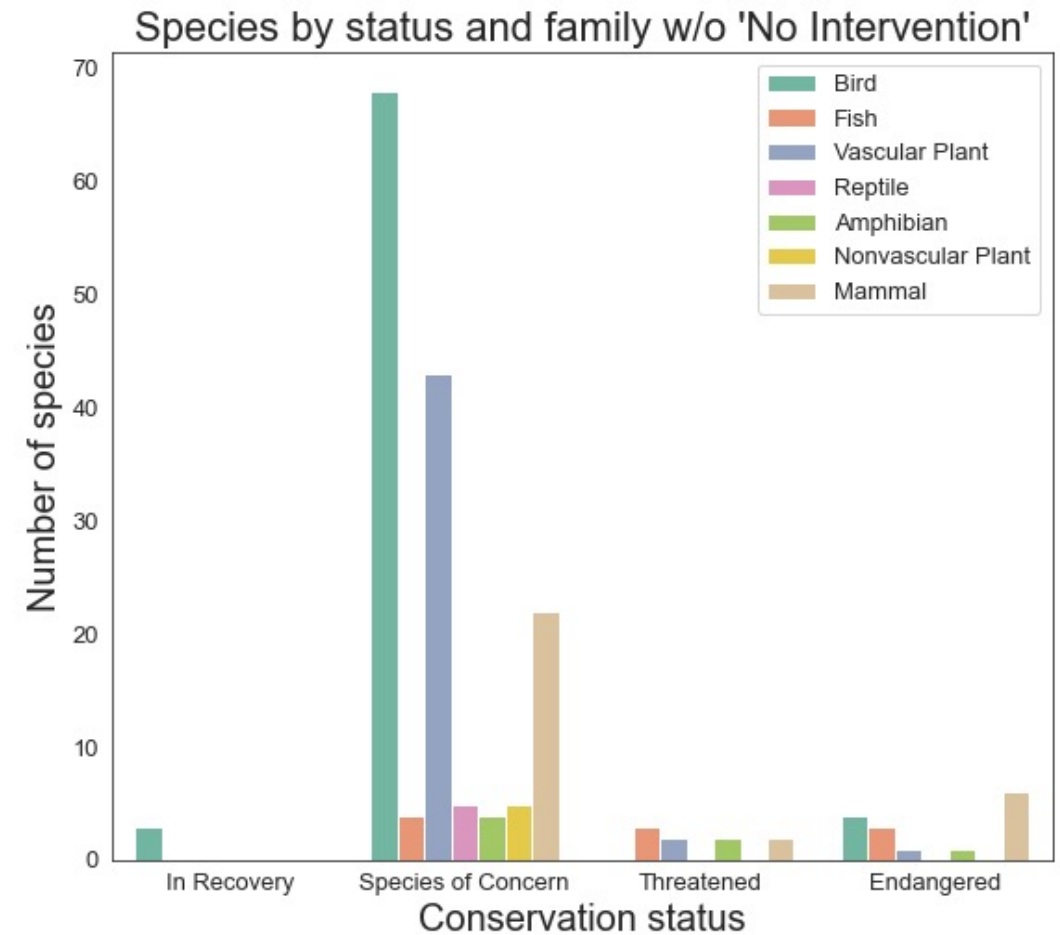
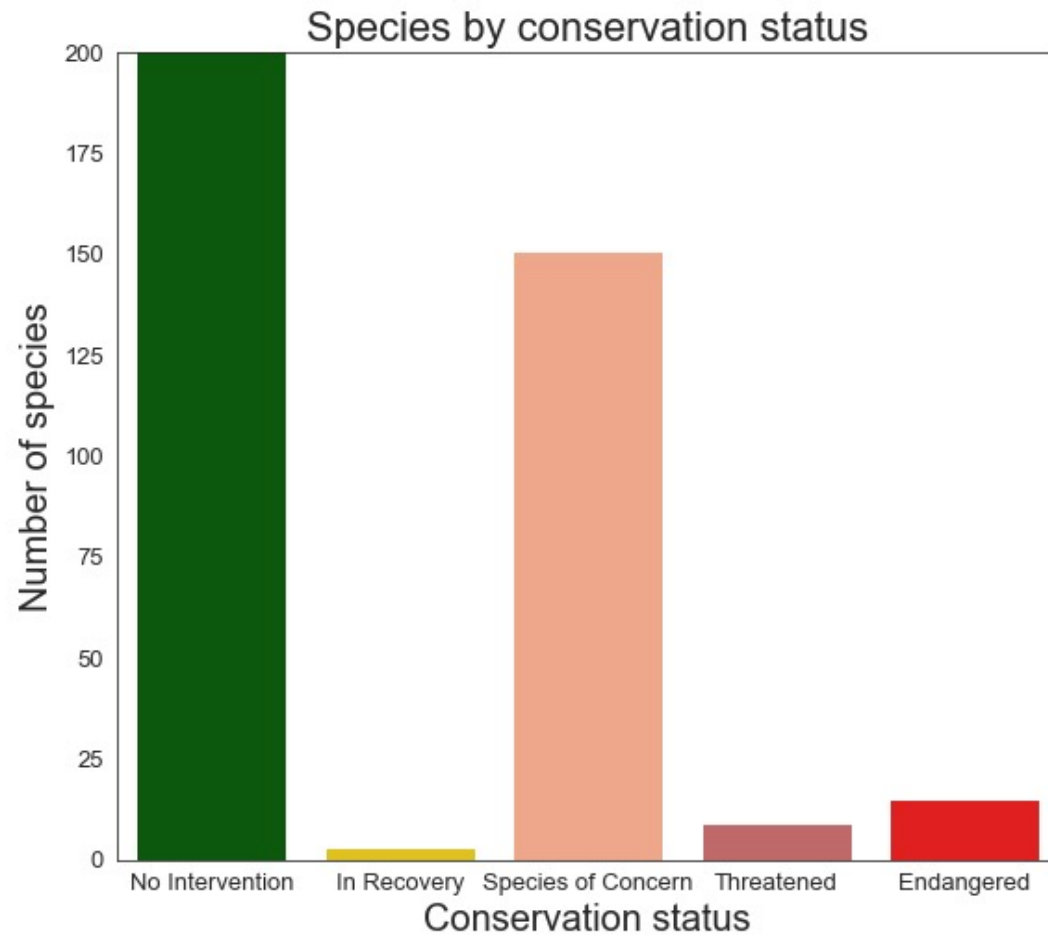
- From the graph: Yellowstone is the park one with the largest number of observations by far (1,443,562). The family that has the largest number of observations (consistently across all 4 parks) are the vascular plants, followed by birds and nonvascular plants
- The velvetgrass is the MSS¹ and appears in the Top 10 of all 4 parks. The Rock Dove is the MSS in the Bryce National Park, and appears in the Top 10 of Yellowstone and Yosemite parks; the Eurasian-Collared Dove (MSS in the Great Smoky Mountains Park) appears in all 4 Top 10; while the Cat's Ear Plant (MSS in the Yosemite Park) appears in 3 out of 4 Top 10 (does not appear in the Great Smoky Mountains Park Top 10)²

1. MSS: Most-spotted species
 2. Refer to the Appendix 2 for some the detailed Top 10 of MSS per park

Appendices

1. Distribution of species by conservation status
2. List of the Top 10 most-spotted species per park

Appendix 1: Distribution of species by conservation status (related to section 2)



Appendix 2.1: List of the Top 10 most-spotted species per park (related to section 5)

Bryce National Park

sci_names	num_obs	names	category
Columba livia	339	Rock Dove	Bird
Puma concolor	311	Panther (Mountain Lion)	Mammal
Avena sativa	303	Cultivated Oat	Vascular Plant
Streptopelia decaocto	301	Eurasian Collared-Dove	Bird
Myotis lucifugus	298	Little Brown Bat	Mammal
Holcus lanatus	296	Common Velvet Grass	Vascular Plant
Hypochaeris radicata	294	Cat's ear	Vascular Plant
Castor canadensis	278	American Beaver	Mammal
Viola arvensis	277	Field Pansy	Vascular Plant
Amaranthus albus	268	White Pigweed	Vascular Plant

Great Smoky Mountains National Park

sci_names	num_obs	names	category
Streptopelia decaocto	256	Eurasian Collared-Dove	Bird
Procyon lotor	247	Common Raccoon	Mammal
Castor canadensis	243	American Beaver	Mammal
Puma concolor	239	Panther (Mountain Lion)	Mammal
Fragaria virginiana	217	Virginia Strawberry	Vascular Plant
Holcus lanatus	216	Common Velvet Grass	Vascular Plant
Microtus longicaudus	212	Long-Tailed Vole	Mammal
Apocynum androsaernifolium	210	Spreading Dogbane	Vascular Plant
Polygonum aviculare	206	Bird Knotweed	Vascular Plant
Sus scrofa	205	Feral Hog	Mammal

Appendix 2.2: List of the Top 10 most-spotted species per park (related to section 5)

Yellowstone National Park

sci_names	num_obs	names	category
Holcus lanatus	805	Common Velvet Grass	Vascular Plant
Streptopelia decaocto	771	Eurasian Collared-Dove	Bird
Puma concolor	753	Panther (Mountain Lion)	Mammal
Procyon lotor	745	Common Racoon	Mammal
Hypochaeris radicata	726	Cat's ear	Vascular Plant
Columba livia	722	Rock Dove	Bird
Castor canadensis	703	American Beaver	Mammal
Myotis lucifugus	679	Little Brown Bat	Mammal
Bidens tripartita	602	Stick-Tight	Vascular Plant
Sorex hoyl	573	Pygmy Shrew	Mammal

Yosemite National Park

sci_names	num_obs	names	category
Hypochaeris radicata	505	Cat's ear	Vascular Plant
Castor canadensis	501	American Beaver	Mammal
Holcus lanatus	463	Common Velvet Grass	Vascular Plant
Streptopelia decaocto	457	Eurasian Collared-Dove	Bird
Procyon lotor	453	Common Racoon	Mammal
Columba livia	423	Rock Dove	Bird
Puma concolor	408	Panther (Mountain Lion)	Mammal
Myotis lucifugus	396	Little Brown Bat	Mammal
Nepeta cataria	387	Catnip	Vascular Plant
Agrostis variabilis	382	Mountain Bentgrass	Vascular Plant