codecademy/PRO

Introduction to Data Analysis, Intensive course

Capstone project: Biodiversity for the National Parks

Dmitry Khochanskiy

Goals

- to analyze data to draw a conclusion on the protection state of species and possible interventions measures
- To analyze sheep observation data from different parks and estimate time period, required to confirm any change in prevalence of foot and mouth disease among them

Part one

Species evaluation

Species status in national park

- We have **5541** different species in 'Mammal‘, 'Bird‘, 'Reptile‘, 'Amphibian‘, 'Fish‘, 'Vascular Plant‘, 'Nonvascular Plant’ categories combined
- The most numerous are **Vascular Plants**
- Of animals the most numerous are **Birds**

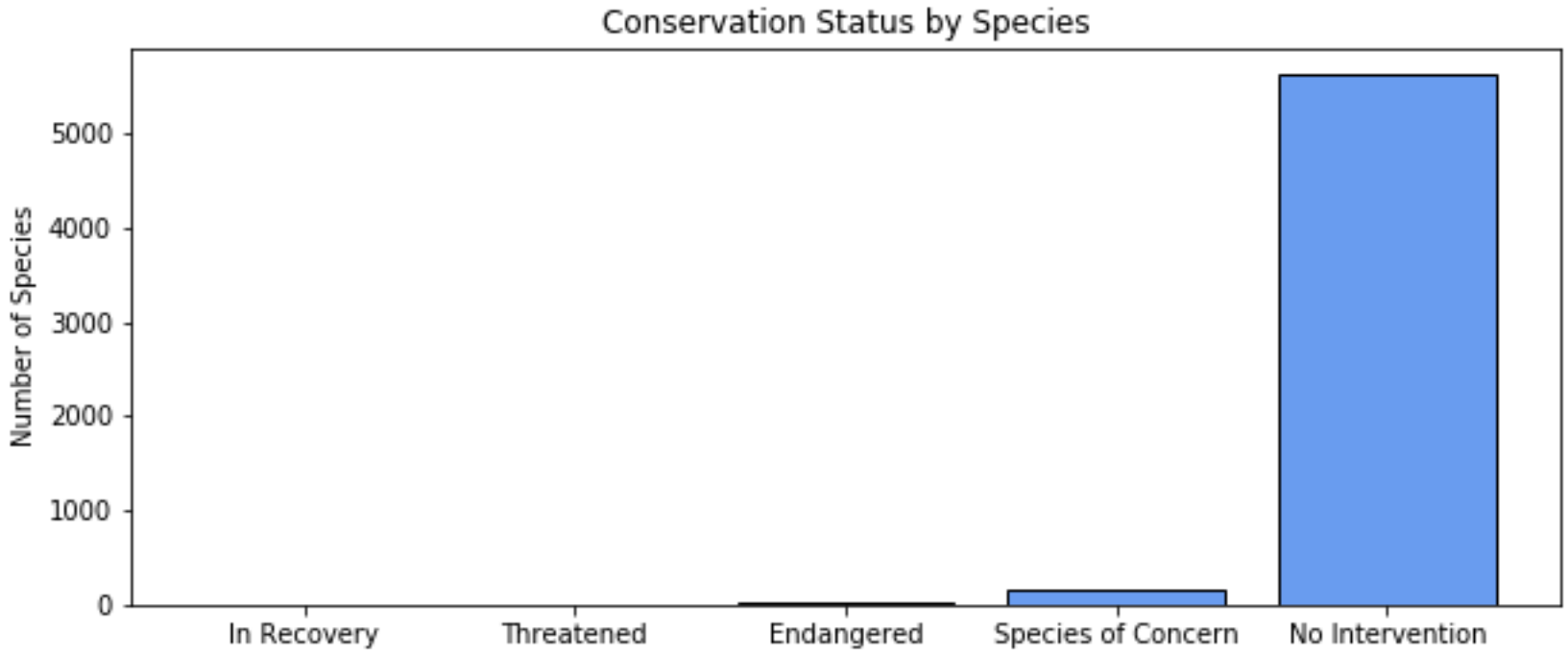
Category	Species number
Reptile	78
Amphibian	79
Fish	125
Mammal	176
Nonvascular Plant	333
Bird	488
Vascular Plant	4262

Species status in national park

- On basis of their protection status animal and plant species can be grouped into **'Species of Concern‘, 'Endangered‘, 'Threatened‘, 'In Recovery'** and **'No intervention'** categories
- Most species have no need of protection, though some do

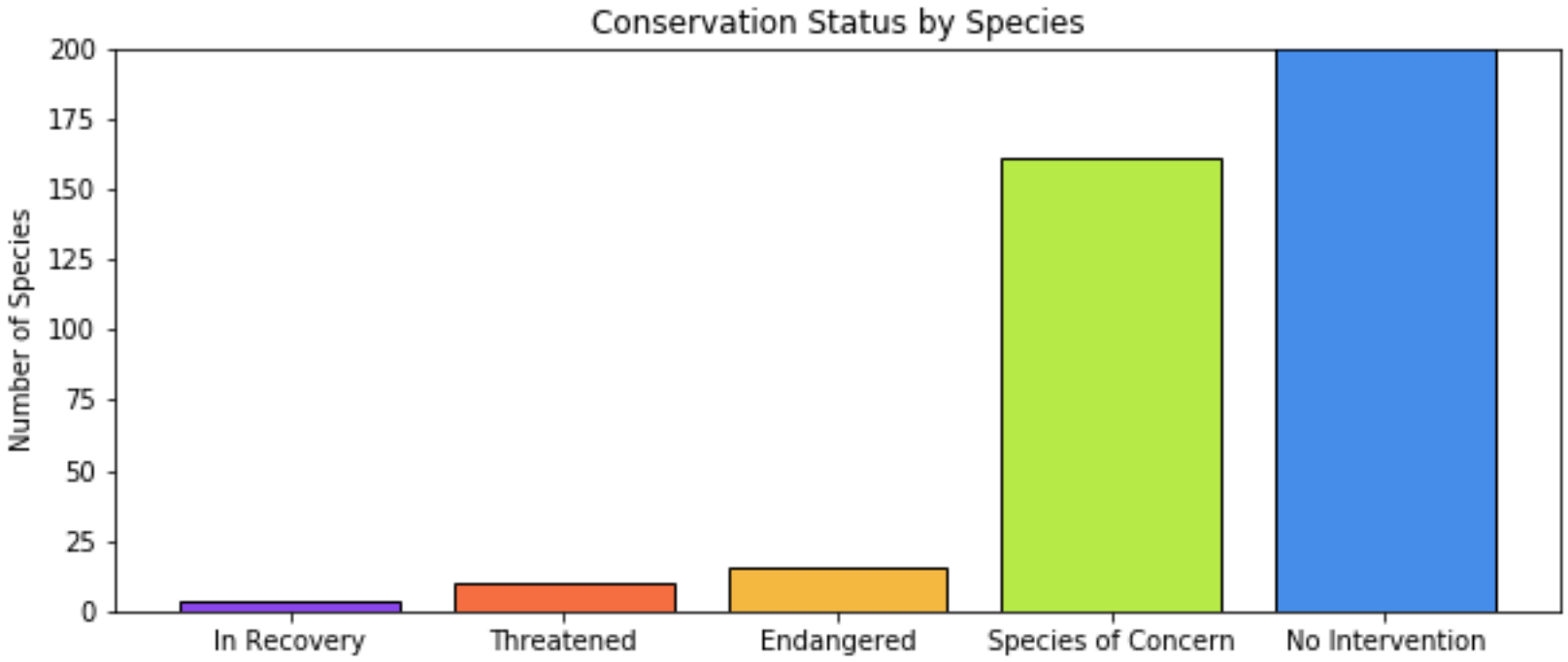
Conservation status	Species number
In Recovery	4
Threatened	10
Endangered	15
Species of Concern	151
No Intervention	5363

Species status in national park



- On basis of their protection status they can be grouped to **'Species of Concern‘, 'Endangered‘, 'Threatened‘, 'In Recovery'** and **'No intervention'** categories
- Most species have no need of protection, though some do

Species status in national park



- Most species have no need of protection, though some do
- **'Species of Concern‘** is the second most numerous category
- All other conservation status categories are far smaller

Species status in national park

- There is a significant difference among categories in terms of protection status
- In absolute numbers **Vascular Plants** hold the second biggest amount of species in need of protection, yet in relative term they are the last
- **Birds** and **mammals** are close if we count the relative quantity of species, which are needed to be protected

Category	Not protected	Protected	%
Amphibian	72	7	8.86
Bird	413	75	15.37
Fish	115	11	8.73
Mammal	146	30	17.05
Nonvascular Plant	328	5	1.50
Reptile	73	5	6.41
Vascular Plant	4216	46	1.08

Species status comparison

This is confirmed by **the chi-square test** if we compare two categories **pairwise**:

- p-value of bird-mammal comparison is 0.688 – **not significant**
- p-value of reptile-mammal comparison is 0.038 – **significant**

If we use **multiple comparison** with **the chi-square** test for these three categories **mammal/bird/reptile** – we get p=0.0752 – **not significant** if we assume 0.05 to be a break even for significance. We may review our data/analysis methods or relax requirements for the level of significance

Recommendations

- The most vulnerable categories – **mammals** and **birds**. Amphibians, reptiles and fish species are less numerous overall and seem to be less in need of protection. Nevertheless all protected species should be taken into consideration
- More information is needed to address why some species are more vulnerable – size, migration behavior, food source etc. The protection status itself may be biased – **mammals** and **birds** are easier to observe and count, they are a common target for protective efforts. People usually pay less attention to small cryptic flowering plants or nocturnal newts. This needs to be accounted for in future research

Part two

Sheep evaluation

Sheep in parks

- Of 176 mammal species three belong to sheep species
- Excluding feral sheep two species are need for protection
- When calculating the disease prevalence we must account for domestic sheep as well so we can implement prevention measures in endangered species

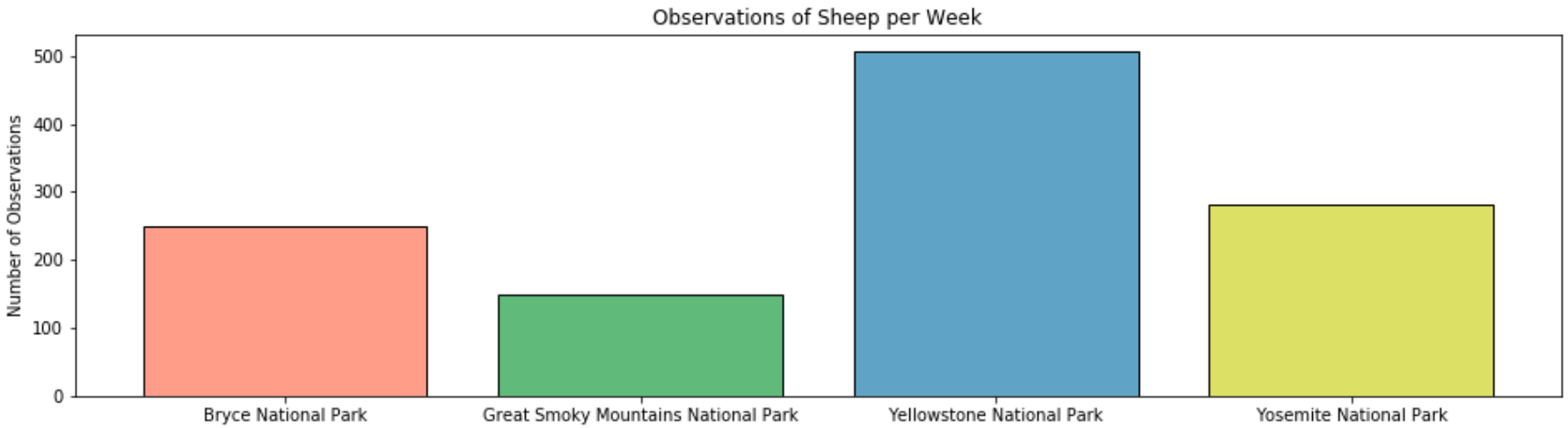
Scientific name	Common name	Protection status
Ovis aries	Domestic Sheep, Mouflon, Red Sheep, Sheep (Feral)	No Intervention
Ovis canadensis	Bighorn Sheep, Bighorn Sheep	Species of Concern
Ovis canadensis sierrae	Sierra Nevada Bighorn Sheep	Endangered

Sheep in parks

- Yellowstone National Park holds the maximum observation number
- Great Smoky Mountains National Park holds the minimum
- This may be due to different size of parks, ease of observation, sheep number etc.

Park name	Observations
Bryce National Park	250
Great Smoky Mountains National Park	149
Yellowstone National Park	507
Yosemite National Park	282

Sheep in parks



- Yellowstone National Park holds the maximum observation number per week
- Great Smoky Mountains National Park holds the minimum
- This may be due to different size of parks, ease of observation, sheep number etc.

Foot and mouth disease of sheep

- 15% of sheep at Bryce National Park have the disease **– baseline**
- Decrease by 5% - thus **Minimum Detectable Effect** - 33.3%
- the default **level of significance** - 90%
- With sample calculator this gives a **sample size** of 870 animals

It will take:

- **3.5 weeks** in Bryce National Park
- **1.7 weeks** in Yellowstone National Park

to observe this amount of sheep to see the decrease

Better to invest in UAVs with FLIR and sheep recognition software to speed up the process!

Thank you for attention!