



## **Introduction to Data Analysis**

Capstone Project: **Biodiversity for the National Parks**

Kathrin Aurelio

## Initial situation

- A section describing the data in **species\_info.csv**. Be sure to include some (or all) of what you noticed while working through the notebook.
- Are there any patterns or themes to the types of species that become endangered?
- Basis of the analysis: data given from the National Park Service containing:
  - 5541 Different species in our National Parks and their categories
  - Their different common names and unique scientific names
  - Their conservation status

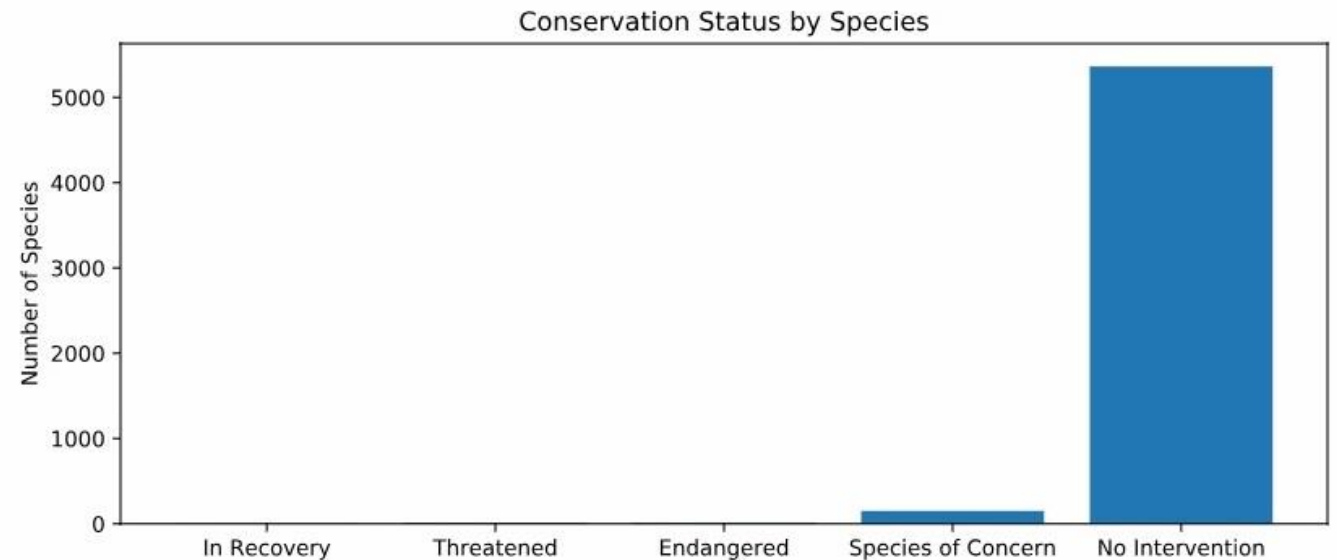
The distribution of the species looks like following:

Category	Number of species
Amphibian	79
Bird	488
Fish	124
Mammal	176
Nonvascular Plant	333
Reptile	78
Vascular Plant	4262

## Initial situation

The distribution of the conservation status looks like following:

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



## **Are certain types of species more likely to be endangered?**

We separated the conservation status in not protected and protected:

not protected      = no intervention

protected          = endangered, in recovery, species of concern,  
threatened

**Are certain types of species more likely to be endangered?**

Category	Is not protected	Is protected	Protected percent
Amphibian	72	7	9%
Bird	413	75	15%
Fish	113	11	9%
Mammal	146	30	17%
Nonvascular Plant	328	5	2%
Reptile	73	5	6%
Vascular Plant	4216	46	1%

**→ It seems, that Mammals or Birds are more likely to be endangered.**

## **Are certain types of species more likely to be endangered?**

To make sure, that there is an actual difference in probability, we checked the significance:

- Method: chi-squared test
- Null hypothesis: difference is a result of chance
- Categories, we compared:
  - Birds & mammals
  - Reptiles & mammals
  - Amphibians & mammals

**Are certain types of species more likely to be endangered?**

Birds (15%) & mammals (17%):

→ P-value: 0.69 – no significant difference

Reptiles (6%) & mammals (17%):

→ P-value: 0.038 – **significant difference here**

Amphibians(9%) & mammals (17%):

→ P-value: 0.12 – no significant difference

**→ Yes, certain types of species are more likely to be endangered!**



**Mammals**, for example, are more likely to be endangered than reptiles. With regard to conservatory aspects it might make sense to **focus** on this category for more success:

- **Research:**  
Are there reasons for the increased risk to be endangered in these species, their environment and their interaction?
- **Prevention:**  
By knowing those reasons, can those species be protected better to not get endangered?
- **Focus of the conservatory work:**  
Focus on mammals for a greater influence and better chances to preotect more species.

## Is the program to reduce the rate of the Foot and Mouth Disease in the Yellowstone National Park working?

The scientists want a reduction of at least 5% and the sample size needs to be big enough to deliver a significant result.

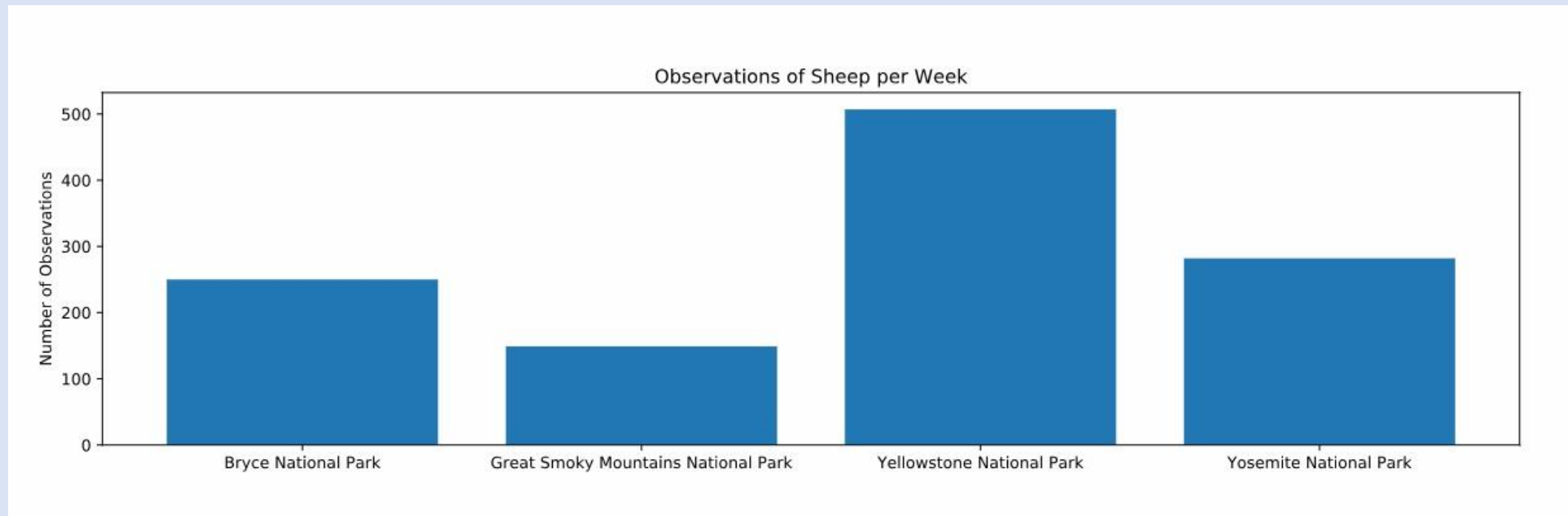
Which sample size do we need? How long do they need to observe enough sheep?

- We know that last year 15% of the sheep in Bryce National Park suffered from Foot and Mouth Disease.  
→ our **baseline is 15%**.
- For a change of 5% we need a **minimum detectable effect of 33.33%**.

**Is the program to reduce the rate of the Foot and Mouth Disease in the Yellowstone National Park working?**

- We want a statistical **significance of 90%**.
- Our **sample size** needs to be at least **870 sheep**.

**Is the program to reduce the rate of the Foot and Mouth Disease in the Yellowstone National Park working?**



## Is the program to reduce the rate of the Foot and Mouth Disease in the Yellowstone National Park working?

- Given the 507 sheep sightings in **Yellowstone** last week, the scientist would approximately need **2 weeks** to observe enough sheep.
- In **Bryce National Park** there were only 250 sightings last week. Therefore the scientist would approximately need **4 weeks** to observe the needed 870 sheep.