Biodiversity Capstone Project

Introduction to Data Analysis
Codecademy

Biodiversity Capstone Project

The **NATIONAL PARKS SERVICE** would like to investigate if....

❖ Exist any **PATTERNS or THEMES** to the types of species that become **ENDANGERED**.

❖ Track **SHEEP LOCATIONS**.

❖ Test if the program to reduce the rate of **FOOT AND MOUTH DISEASE** at parks is working.

Our resources are two CVS datafiles from National Parks Services: ⟫⟫ **species_info.CSV**

observations.CSV

Goal: ANALYZE and PLOT data information from CSV datafiles to answer the questions

Biodiversity Capstone Project

PATTERNS or THEMES to the types of species that become ENDANGERED.

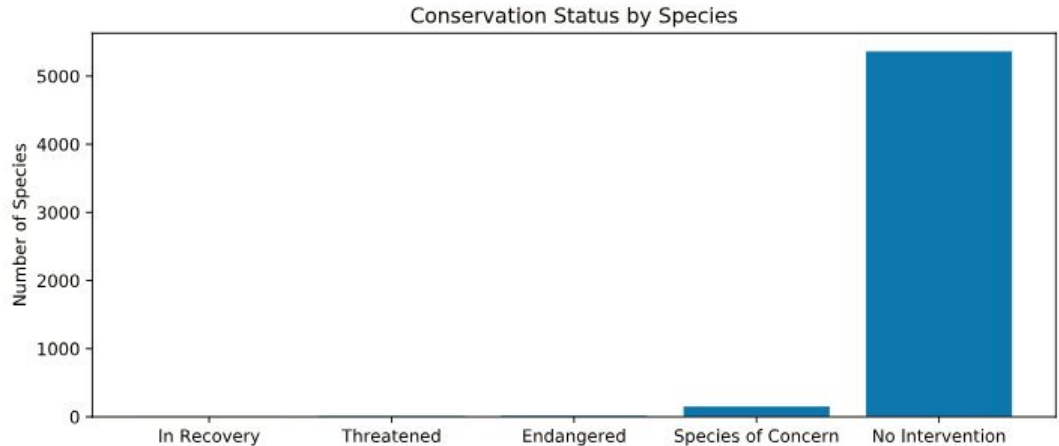According our information[1] we found several possible status of conservation to different species.

Species of Concern(151)
Threatened(10)
Endangered (15)
In Recovery (4)
No Intervention (5363)

Conservation Status by Species

Are certain types of species more likely to be endangered?

[1] Information from Natural Parks Services. Data files "**species_info.csv**" and "**observations.csv**"

Biodiversity Capstone Project

Are certain types of species more likely to be endangered?

```
   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
```

Are Mammals more likely to be endangered than Bird?

CHI-SQUARED TEST
null HYPOTHESYS = difference is a result of a chance

Mammal vs Bird: difference isn't significant

Mammal vs Reptile: difference is significant

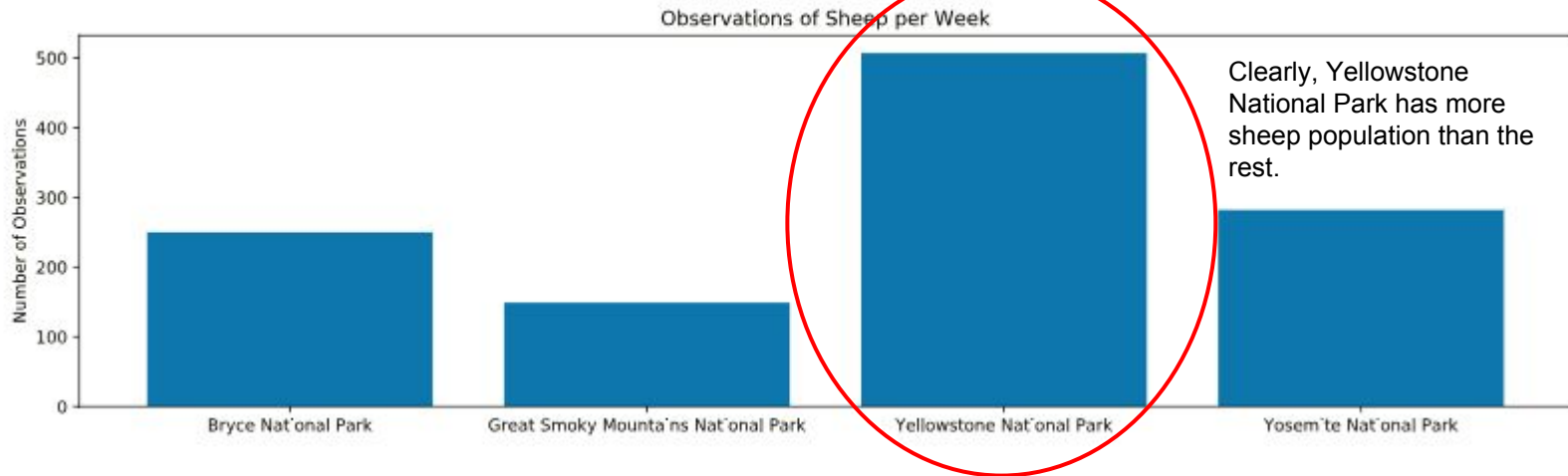# Biodiversity Capstone Project

Conclusions:

"Certain types of species, like Mammals and Birds, are more likely to be endangered than others"

The national parks service should develop protection programs for Birds and Mammals and investigate what may be the causes that these species are more likely to be endangered.

Biodiversity Capstone Project

Track SHEEP LOCATIONS

We have information about recording sightings of different species at several national parks for the past 7 days Need to merge species_info.CSV and observations.CSV to get information with observations of sheep.

Observations of Sheep per Week

Clearly, Yellowstone National Park has more sheep population than the rest.

Biodiversity Capstone Project

Program to reduce the rate of FOOT AND MOUTH DISEASE at parks.

The program is running at Yellowstone National Park.
The scientists want to be able to detect reductions of at least 5 percentage points.
The only information that the scientists currently have is that last year it was recorded that **15%** of sheep at **Bryce National Park** have foot and mouth disease.

Need to calculate the number of sheep that we would need to observe from each park to make sure our foot and mouth percentages are significant.

We need to calculate:
- Baseline
- Minimum Detectable Effect (Lift)
- Level of significance (by default is 90%)

Biodiversity Capstone Project

Baseline ==> 15

Our baseline is 15 (The only information that the scientists currently have is that last year it <mark>was recorded that 15% of sheep</mark> at **Bryce National Park** have foot and mouth disease)

Minimum Detectable Effect (called LIFT) ==> 33,33

$$\frac{100*(\text{old value} - \text{new value})}{\text{old value}}$$

OLD VALUE is 15, is our only information about foot and mouth disease
NEW VALUE is 10, because is a result of reduce 5 percentage points.

Level of significance ==> 90 (by default)

The sample size to our study of foot and mouth disease is 870

Biodiversity Capstone Project

How many total sheep sightings were made at each national park?

Park_name	Observations	Weeks to observe 870
Bryce National Park	250	3,48
Great Smoky Mountains National Park	149	5,84
Yellowstone National Park	507	1,52
Yosemite National Park	282	3,09

How many weeks would we need to spend at Bryce National Park to observe enough sheep?
How many weeks would we need to spend at Great Smoky Mountains National Park to observe enough sheep?
How many weeks would we need to spend at Yellowstone National Park to observe enough sheep?
How many weeks would we need to spend at Yosemite National Park to observe enough sheep?

Biodiversity Capstone Project

We found that if we want to be sure that a >5% drop in observed cases of foot and mouth disease in the sheep at parks is significant we would have to observe at least 870 sheep

We need approximately one week of observing in Yellowstone to see that many sheep, or approximately three weeks in Bryce to see that many sheep.