

# Title Slide

## Codecademy Presentation

**Course:** Introduction to Data Analysis

**Project:** Biodiversity Project

**Student name:** Rogier Eduardus Laurentius Hetem

**Total page count:** 10

**Submit date:** May 10, 2018

# Content

## Part 1

- Data description of 'species\_info.csv' (general)
- Data description of 'species\_info.csv' (into detail)
- Description of the significant calculations
- Recommendation for conservationists

## Part 2

- Sample size determination

## Part 3

- Graphs

## Data description of 'species\_info.csv'

### General information

- We got a csv file containing data of all animals living in 4 different National Parks
- The file holds 5543 different species of animals (rows)
- The file gives us 4 different pieces of information per animal (columns)
- Such as; category, name (scientific and common) and conservation status

### First three rows of the data file

ID	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), Domesticated Cattle	nan

## Data description of 'species\_info.csv'

### Into detail (1)

#### Description of conservation\_status

<b>Species of Concern:</b>	Declining population or appears to be in need of conservation.
<b>Threatened:</b>	Vulnerable to endangerment in the near future.
<b>Endangered:</b>	Seriously at risk of extinction.
<b>In Recovery:</b>	Formerly Endangered, but currently not in danger of extinction throughout all or a significant portion of its inhabitable range.
<b>nan:</b>	No Intervention

#### Number of species per conservation status

ID	Status	Number of Species
0	Endangered	15
1	In Recovery	4
2	No Intervention	5363
3	Species of Concern	151
4	Threatened	10

## Data description of 'species\_info.csv'

### Into detail (2)

After categorizing the data on 'Category' (type of species) and labeling each animal; either protected or not protected based on their 'Status' (only 'nan' = not protected). This is depicted in the table below.

We asked ourselves the question: "Are certain types of species more likely to be endangered?".

ID	Category	Not protected	Protected	Percent protected
0	Amphibian	73	7	0.0875
1	Bird	442	79	0.1516
2	Fish	116	11	0.0866
3	Mammal	176	38	0.1776
4	Nonvascular Plant	328	5	0.0150
5	Reptile	74	5	0.0633
6	Vascular Plant	4424	46	0.0103

## Description of the significant calculations

- (1) To answer the question whether certain types of species are more likely to be endangered we performed a Chi-Square test on the 6 different types of species on their status (protected vs not protected). See table on page 5. After this analysis we also performed:
- (2) A Chi-Square test on 2 types of species (Mammal and bird)
- (3) A Chi-Square test on 2 types of species (Mammal and Reptile)
- (4) A Chi-Square test on 2 types of species (Bird and Nonvascular plant)

### Results:

- (1) We found a significant interaction ( $X^2 (6) = 514.38, p < .001$ ) between all different types of species. Meaning the species are independent and some species are in fact more likely to be endangered than others.
- (2) We found no significant interaction between the type of species Mammal and bird
- (3) We found a significant interaction ( $X^2 (1) = 5.13, p < .05$ ) between the type of species Mammal and reptile. Meaning Mammals are more likely to be endangered than reptiles.
- (4) We found a significant interaction ( $X^2 (1) = 41.23, p < .001$ ) between the type of species Bird and Nonvascular plant. Meaning Birds are more likely to be endangered than Nonvascular plants

## **Recommendation for conservationists**

We recommend paying extra attention to both Birds and Mammals. They seem to be the two types of species that are most likely to be endangered.

To do so, we suggest (partly) cutting the budget intended for researching the conservation of plants, especially vascular plants. As the number of vascular plant that are endangered is very low. This money can be used for the extra research on both Birds and Mammals.

## Sample Size Determination

Here we check whether the program is successful;  
to reduce the rate of foot and mouth disease, that been running in the Yellowstone  
National Park, by the Park Rangers.

Baseline = 15%

The baseline is based on the recorded foot and mouth disease in Bryce National Park.

Minimum detectable effect = 33%

The Rangers are expecting a drop of at least 5%-points compared to the baseline. This means a drop  
of 33%. Anything less than this number is not-of-interest.

Significance level = 90%

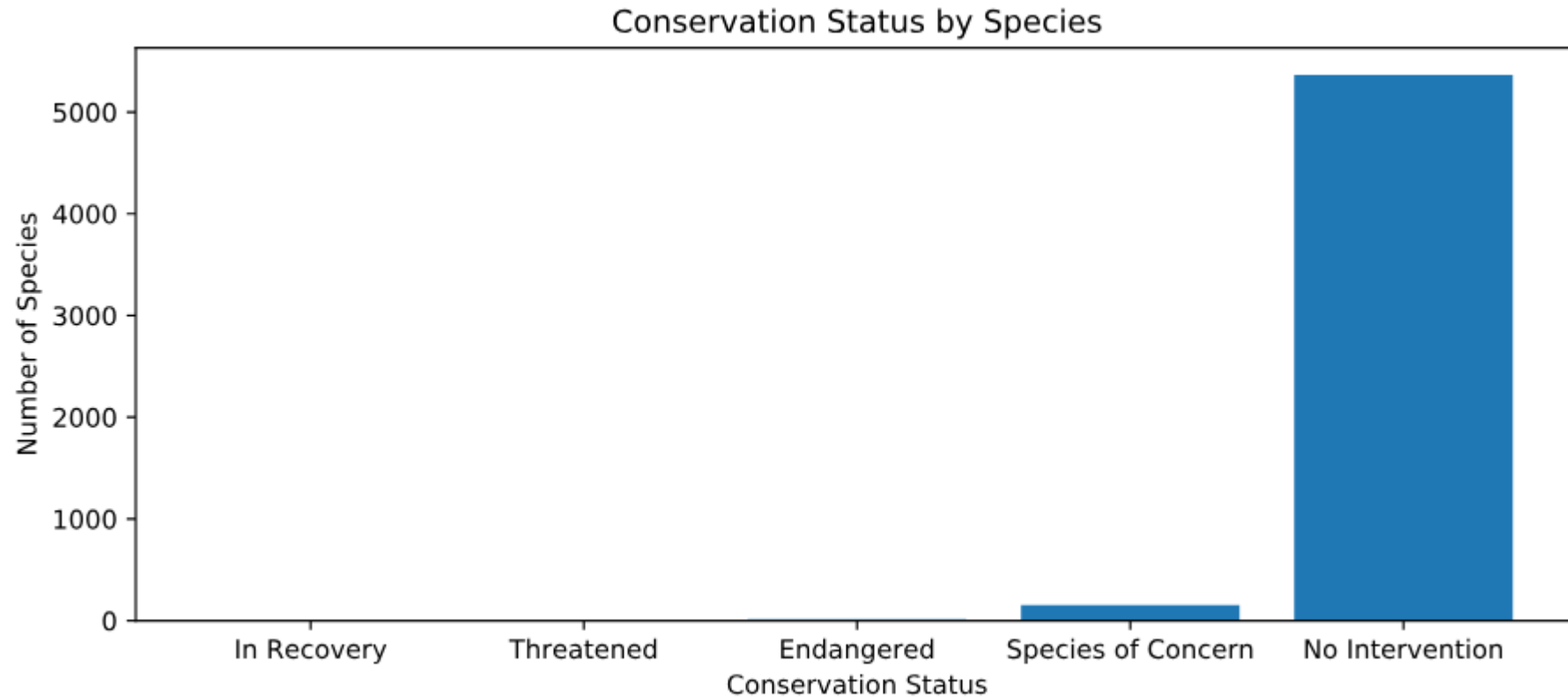
We will use a standard significance level of 90%.

Sample size = 890 per sheep variant

Based on the baseline, minimum detectable effect and the significance level. This will mean that the  
Yellowstone National Park has to do approximately 2 weeks of observations to gather enough data.



## Graphs



## Graphs

