



Codecademy - Introduction to Data Analysis - Feb 12, 2019 - May 7, 2019

Capstone Option 2: Biodiversity for the National Parks

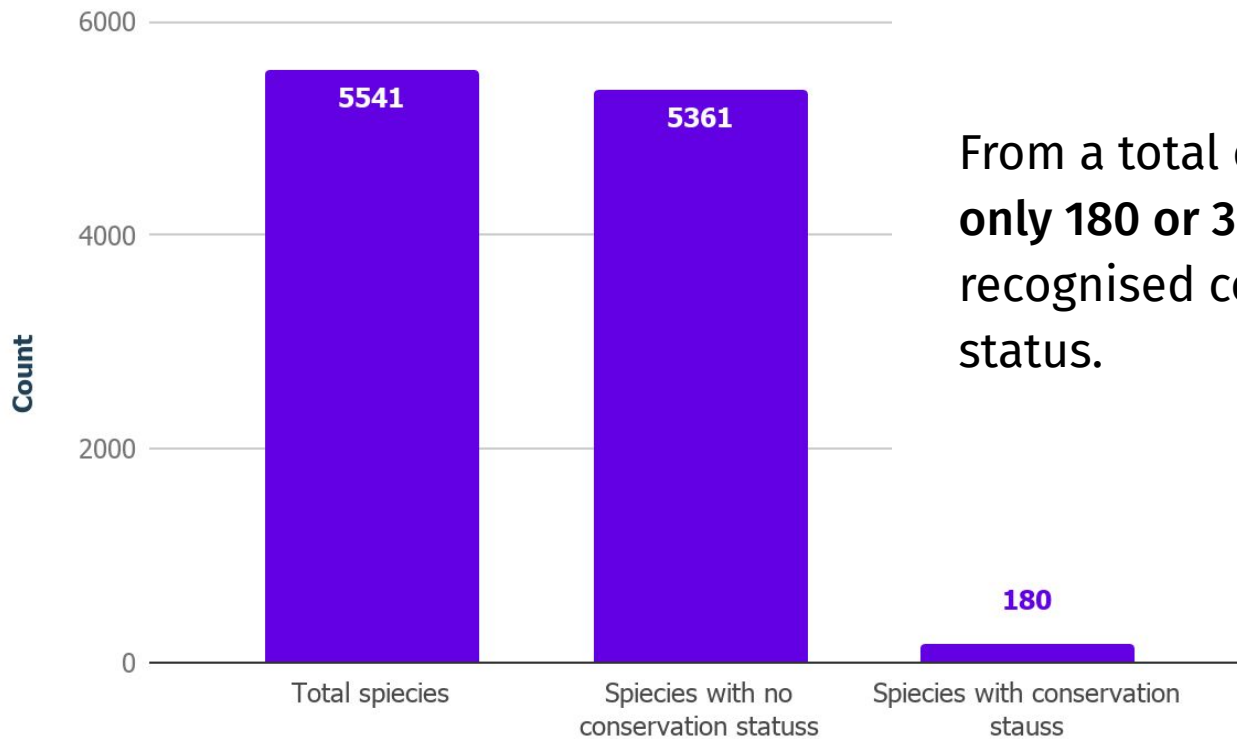
Ruta Fridrihsone

30.04.19

Species_info.csv *Data frame description*

- **Species_info.csv** from National Parks Service includes data about species in National parks and their conservation statuses.
- **Data frame includes a description of 5541 species across classes (types of species) like:** *Mammals, Birds, Reptiles, Amphibians, Fish, Vascular Plants and Nonvascular Plants*
- **Data frame covers conservations statuses like:** *No Intervention, Species of Concern, Endangered, Threatened and In Recovery*

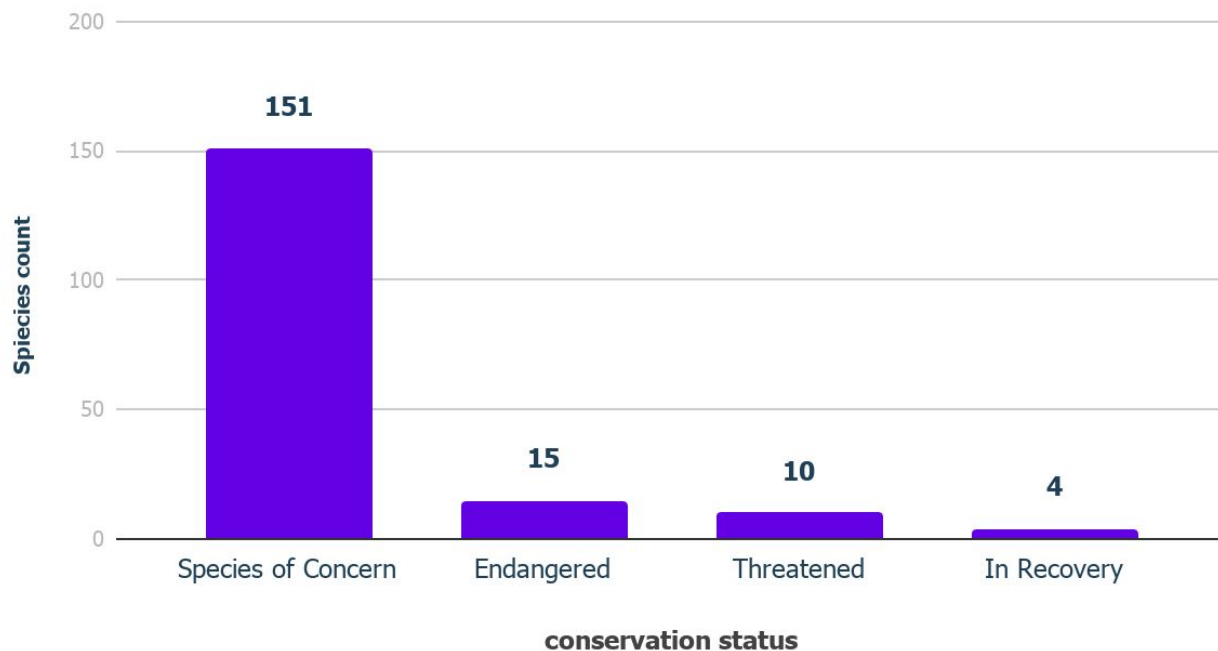
Species_info.csv Species in the Conservation process



From a total of 5541 species, **only 180 or 3.24%** has a recognised conservation status.

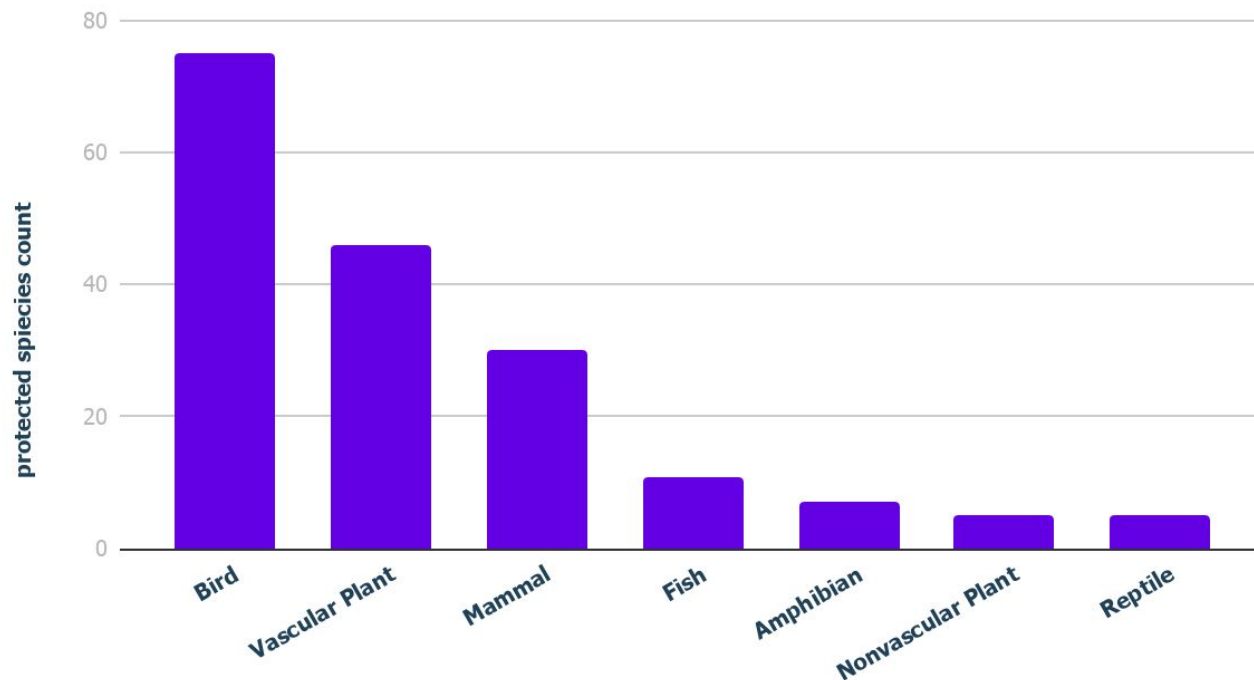
Species_info.csv Species in the Conservation process

Active Conservation status distribution

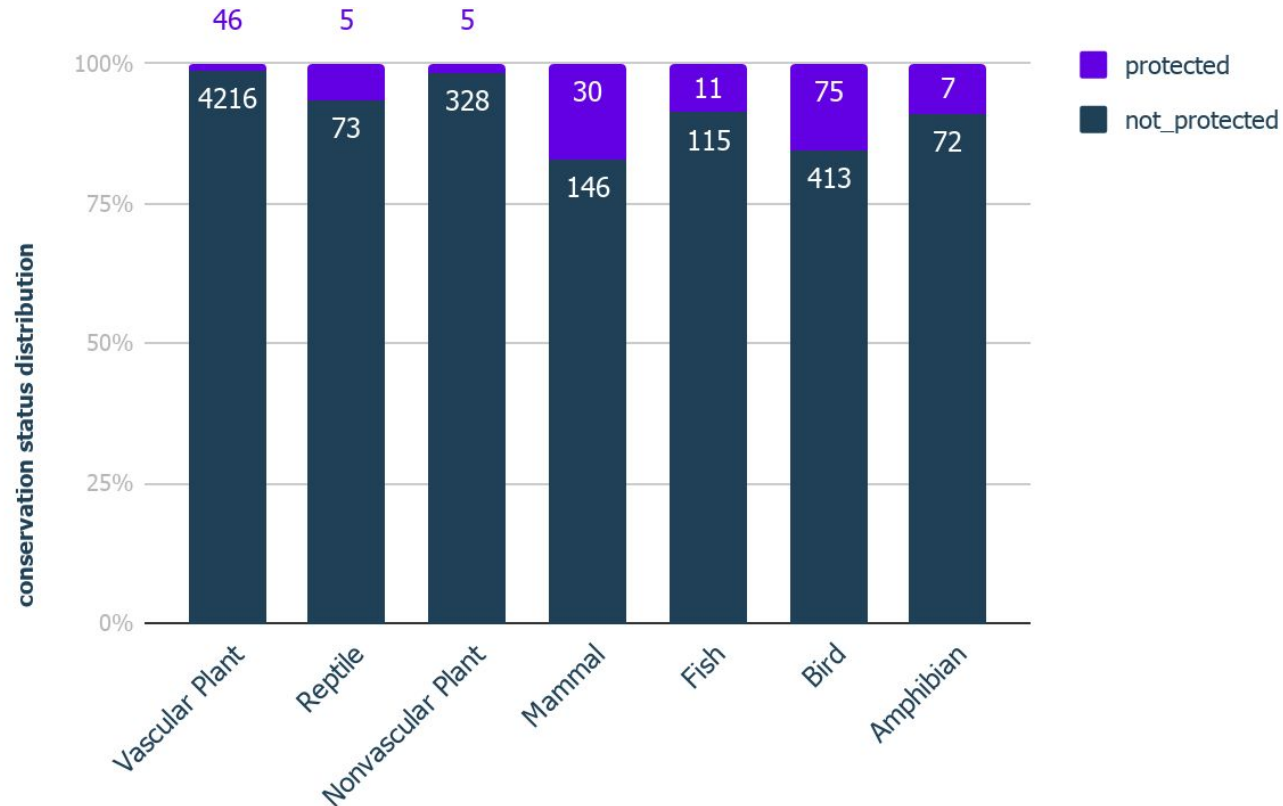


Species_info.csv Conservation status by classes, protected species

Conservation status distribution by classes



Species_info.csv Conservation status within classes



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

Significance calculations

Are certain types of species more likely to be endangered and is the difference significant?

*We can do a significance test to see if the null hypothesis:
'this difference is due to chance'.*

Significance calculations

To test the null hypothesis, we will use the **chi-squared test** based on data from **Species_info.csv**

category	not_protected	protected	percent_protected
Amphibian	72	7	0.088608
Bird	413	75	0.153689
Fish	115	11	0.087302
Mammal	146	30	0.170455
Nonvascular Plant	328	5	0.015015
Reptile	73	5	0.064103
Vascular Plant	4216	46	0.010793

Significance calculations

To run a **chi-squared test**, we need to compare several pairs of species types in the **contingency table** to see if the difference is significant or due to chance

species types	protected	not-protected
Mammal	30	146
Bird	75	413

species types	protected	not-protected
Reptile	5	73
Mammal	30	146

Significance calculations

Chi-squared test results return P-value that helps us conclude is the difference of significance or do to chance.

To conclude that difference is of significance and proving our null hypothesis wrong **P-value has to be below 0.05**

Mammals-Birds Test: P-value = ~ 0.687 (difference isn't significant!)

Reptiles-Mammals Test: P-value = ~ 0.038 (difference is significant!)

Significance calculations

Conclusion:

As the difference between types of species being endangered at least in one pair is significant, we can conclude that:

Certain types of species are more likely to be endangered.

Recommendation for conservationists

As analysing data about endangered species in National parks concludes that some species are more likely to be endangered:

I would recommend doing a more detailed analysis to learn which type of species are more likely to need protection and research reasons behind these tendencies.

Once we can qualify the conditions causing the issues National parks conservationists can decide on actions that can be taken to reduce conditions harming specific types of species.

Sample size determination

Approach:

To determine the sample size needed to test if the Yellowstone National Park program to reduce the rate of foot and mouth disease is effective, we can use existing data from past year recording of cases of illness and expected results as a guideline.

Sample size determination

Available data and expected results:

- Last years records show that 15% of sheep at Bryce National Park have foot and mouth disease.
- To conclude that program is working it's necessary to detect at least a 5% decrease amount of sheep having the disease
- Statistical significance has to be at 90%

Sample size determination

Identifying the necessary variables to calculate the sample size:

- Baseline conversion rate: 15%
- Statistical significance: 90%
- Minimum detectable effect: $100 \cdot 5. / 15 = 33 \%$

Using the Sample size calculator:

Sample size: 890

Task description:

Add all of the graphs that you created in the notebook

Comment:

I completed the project online and you most probably have access to graphs created for the project, anyhow i will add the chart files and screenshot for you to access from here

Conservation Status by Species

