STT 180 Project
Sloth Endangerment

Kaitlyn Szarenski

Sakshi Daduvai

Divyalakshmi Varadha Rajan Prem Sudha

Melvin Lawpan Toe

Badri Aiman Khan bin Badrul Zeman Khan

Sloth Species Dataset

- **Categorical Variables**
  - **Endangerment Level: Critically endangered, vulnerable, or least concern**
  - Species Type: Two-toed or three-toed sloth
  - Sub-species Type: Hoffman's two-toed sloth, Linnaeus's two-toed sloth, pale-throated sloth, or brown-throated sloth

- **Numerical Variables**
  - Claw Length: The length in cm of the sloth's claws
  - Sloth size: The size in cm of the sloth from head to tail
  - Tail Length: Length in cm of the sloth's tail
  - Total Weight: Weight in kgs of the sloth

Question:

Which factors have the highest impact on a sloth's endangerment level?

Plan of Approach

1.) We **loaded in the dataset**, removed NA values
2.) We **glimpsed the dataset** to look at the variables and the data types
3.) We brainstormed the types of plots. Then created the plots and visualized the data using **violin plots, stacked histograms, and several scatterplots**
4.) We found **several patterns using the graphs** that showed to have correlation to show us the correlations between the variables
5.) We v**erified the numeric correlation** between each of the variables and the endangerment level using the **corr() function**
6.) We **developed a linear model** to discover which characteristics affected the endangerment of the sloth the most.
   > ***Foundation for our next models** because the accuracy of the plots was determined by the residual model
7.) We concluded that the **subspecies pygmy three-toed sloth is the most endangered species.** The **most impactful category is the sloth's size with the weight of the sloth,** as there was a direct correlation to the size of the sloth.

Amount of sloths in each endangerment level

The bars represent the number of sloths in each endangerment category divided by the subspecies. We can see from this plot that the sub-species that is most common would be brown-throated sloth while the least common would be pygmy three-toed sloth
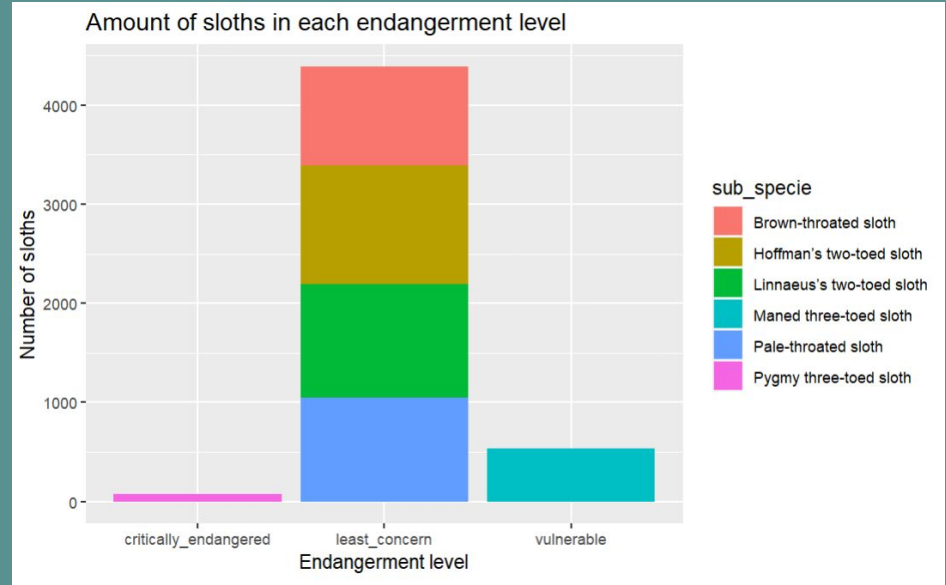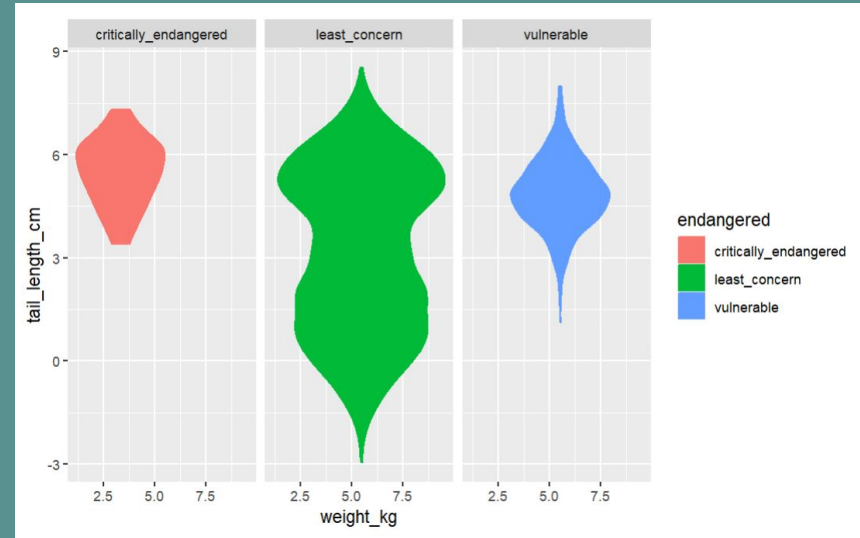
Sub species of sloths

This graph shows how the two main species of sloths: Two toed and three toed sloths are divided into smaller species.
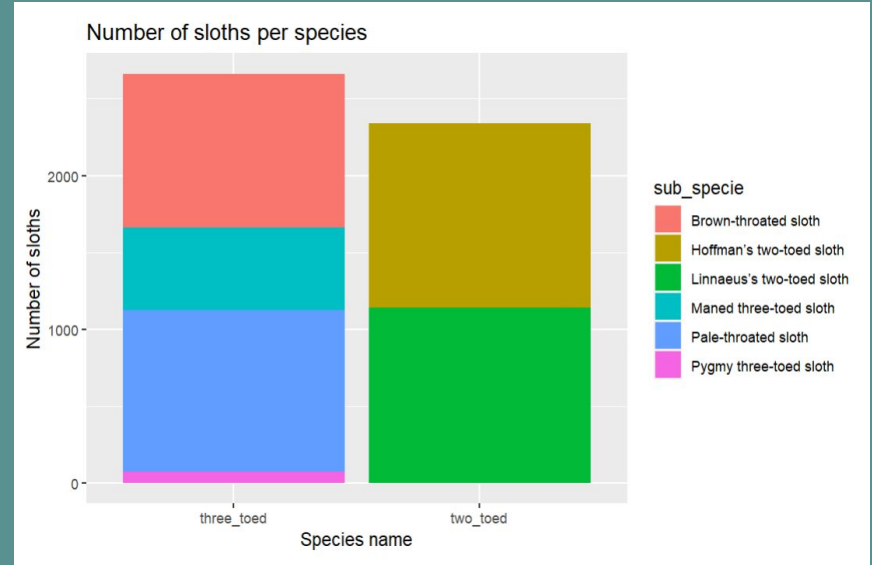
Tail length by weight separated by endangerment level

This graph shows the graph of weight against tail length for the three endangerment levels. We can see from the plot that the plot with the biggest range of length would be those of least concern

Conclusion

We concluded that the **subspecies pygmy three-toed sloth is the most endangered species**.

The **most impactful category is the sloth's size** with the weight of the sloth having a direct correlation to the size of the sloth

Adjusted R^2 values that signify how good of a predictor the variable is. The values are for:

```
[1] 0.03324664
[1] 0.03344706
[1] 0.1214873
[1] 0.2314013
```

weight,
weight and claw length,
weight, tail length,claw length,
weight claw length,tail length, and size

Limitations/Challenges:

- Originally, we **struggled to create a linear model** between the variables weight, subspecies, and endangered.
  - We found that this was because we did not take into account the endangerment and subspecies variables are **categorical variables**, meaning that they **could not directly run in the linear regression** function in r.

- When using the correlation function (corr), we found that it was **difficult to directly determine the relationship** between one or more variables which had **categorical values**.
  - So, for that function, **we only looked at numeric variables** (claw length, sloth size, tail length, total weight).

References

https://www.kaggle.com/datasets/bertiemackie/sloth-species

https://www.statology.org/stepwise-regression-r/