Biodiversity in National Parks

Caitlin Cassidy

Introduction to this Project

The United States has 62 National Parks. The National Park Service has provided data on 56 of these park, which details the biodiversity within the parks. The data contain a catalog of species that live within the parks, information on the conservation status of many of these species, as well as general information regarding the parks' locations and size.

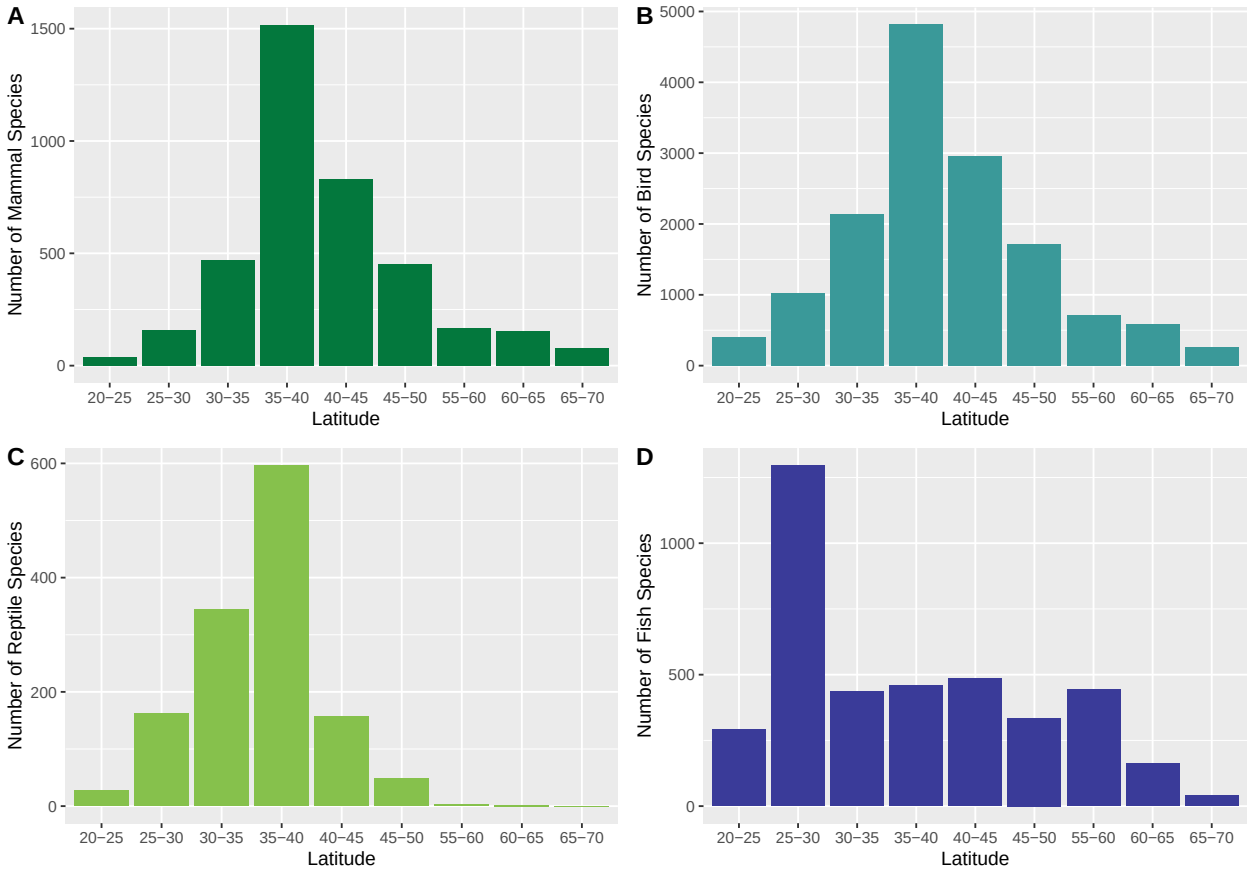
Because of the varying geography of the U.S., the parks vary greatly in terms of their biodiversity. Biodiversity refers to the variety and variability of species in an ecosystem. Biodiversity is important for maintaining healthy ecosystems. It is important that we keep records of park biodiversity in order to understand the ecology and health of these ecosystems.

In this project, we will examine this data in an effort to uncover patterns and relationships in the data. We will also generate linear models and suggest efforts for future data collection and analysis efforts in the National Park Service.

Variability of Number of Species Latitude

The U.S. National Parks are home to a variety of different species of mammals, birds, reptiles, and fish, among many other types of organisms. Since these parks are located all over the U.S., they vary in terms their wildlife.

One interesting pattern that emerges in the data is the number of unique species at different latitudes. The following bar charts depict the number of unique species of mammals (A), birds (B), reptiles (C), and fish (D), at different latitudes. The latitude categories are defined in intervals of 10 degrees latitude.



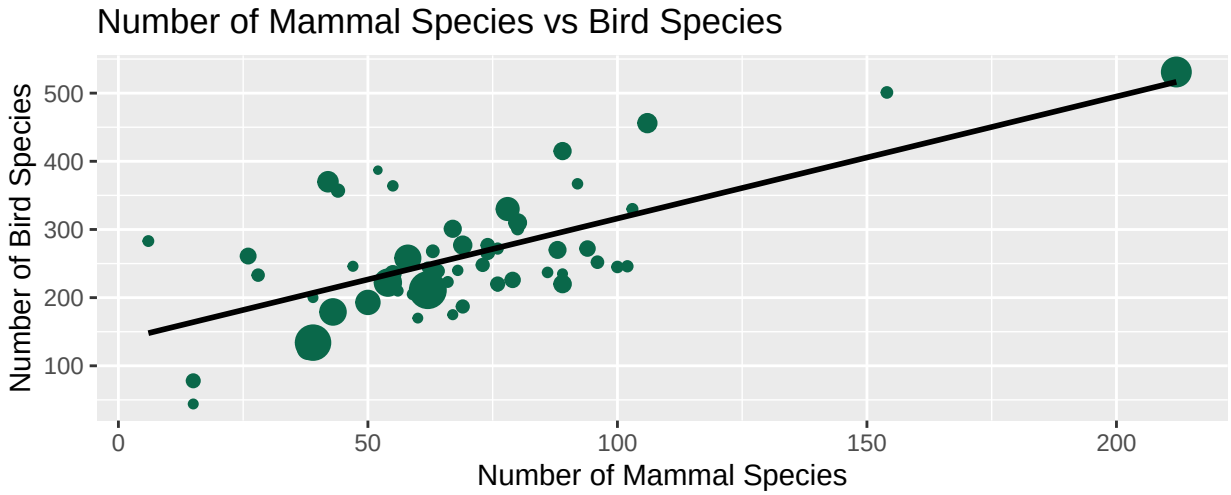
By looking at these bar charts, we can see that there are greater numbers of unique mammal, bird and reptile species that live in latitudes in the middle latitudes. Very few mammal, bird and reptile species live in low latitudes or high latitudes. The climate conditions at the low and high latitudes are more extreme than climates in the middle latitudes. Therefore, this data suggests that middle latitudes provide a better climate for mammals, birds and reptiles to live. This pattern makes sense. It is also interesting to note that the distributions for mammals and birds are extremely similar. I will follow up on this relationship later in this discussion.

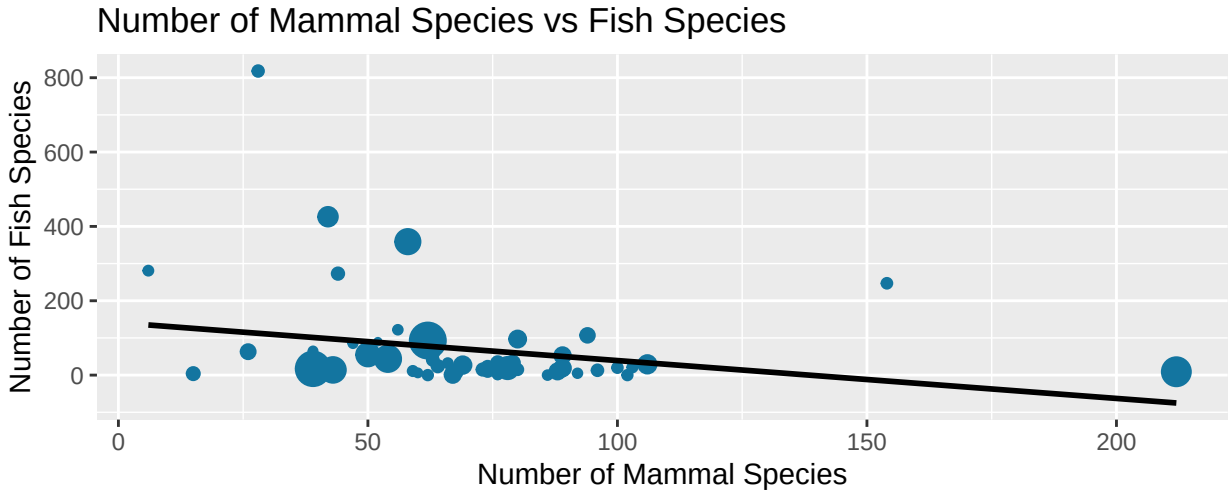
The number of unique fish species, however, does not follow this same pattern. While more fish species live in low latitudes than middle or high latitudes, the distribution is more even across the categories of latitude than the distribution of the other species.

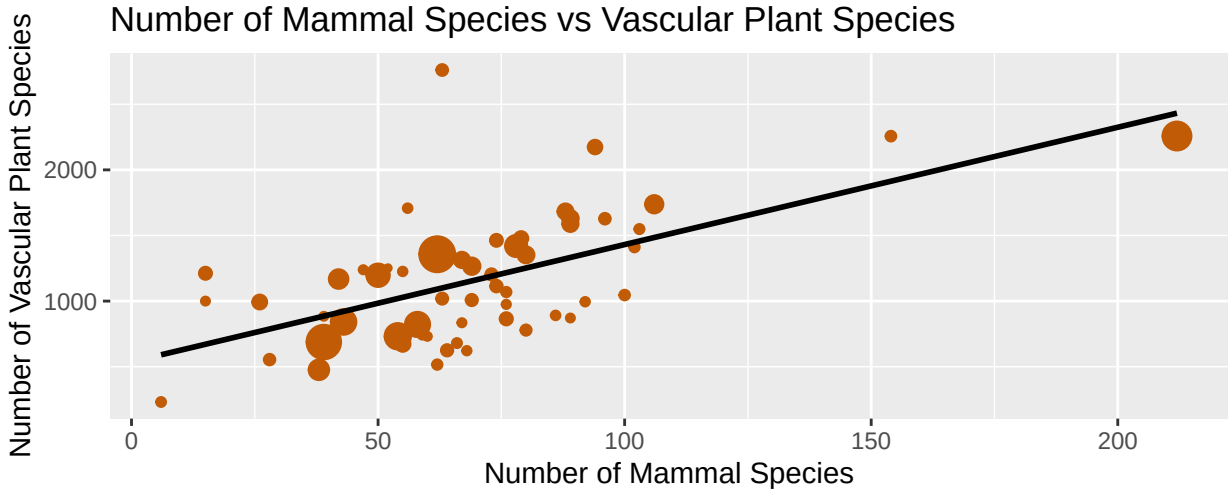
In general, from these figures we can see that parks at middle latitudes have a great number of unique species and thus greater biodiversity than parks at low and high latitudes.

Relationships Between Species

From the previous discussion, we noticed that the distributions of the number of unique species of birds and mammals by latitude were very similar. I decided to see there a pattern between the number of species of mammals and birds exists in the parks. I also wanted to see if patterns existed between the number of species of mammals and the number of species of other organisms.







These scatterplots illustrate correlations between mammals and other species in the parks. The different size of the points represents the acreage of the park (larger points correspond to larger acreage).

The first scatterplot shows a positive association between the number of mammal species and the number of bird species in the parks. The second scatterplot shows a negative association between the number of mammal species and the number of fish species in the parks. The third scatterplot shows a positive association between the number of mammal species and the number of bird species in the parks.

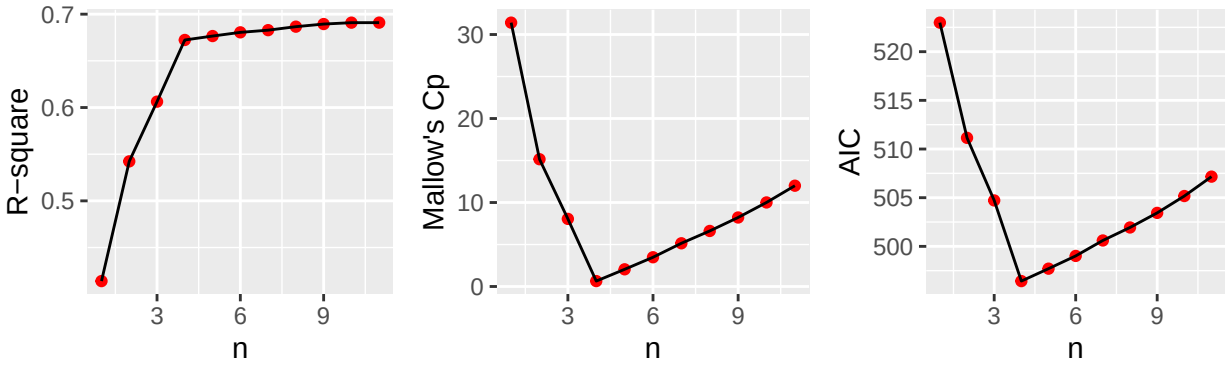
I chose to include these plots in this report because these relationships results in the most interesting patterns in the scatterplots. Other correlations do exist. For instance, there was evidence of a weak positive correlation between the number of mammal species and the number of nonvascular plant species. There was also evidence of a correlation between the number of mammal species and the number of amphibian species. For some comparisons, no correlation appear to exist. For instance, there was no pattern in the relationship between mammal species and fungi species.

From these plots, we can see that there are some associations between the number of mammal species and the number of other types of species. This information is useful for a variety of reasons. For example, it may be possible to estimate the number of mammal species in a park for which the number of another type of species is known.

Can we predict the number of mammal species in a park?

Is it possible to predict the number of mammal species in a park?

After observing the simple linear relationship between the number of mammal species and the number of other species, I built multiple linear regression models to explore this question. I used the best subsets regression technique to compare all possible models to predict the number of mammal species. I created a few plots from the resulting models to illustrate different parameters of models created by the best subsets regression.

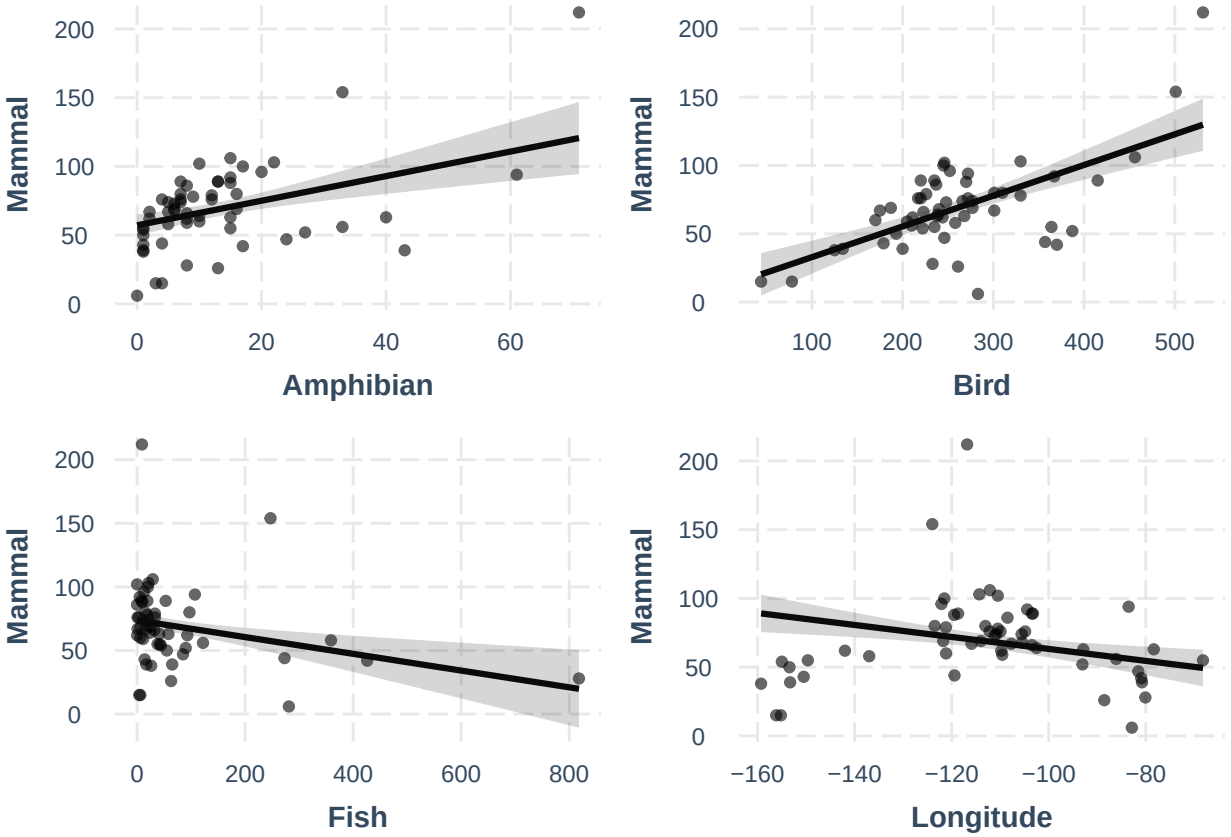


These plots show the values of R-square, Mallow's Cp and AIC for the models generated by the best subsets regression for n = 1 to n = 11 variables.

From these plots, we can see that the increase in the R-square value is very minimal for each model term added after the 4th variable. In general, for this problem, we want to select the model with a high R-square with the fewest number of terms. Minimizing Mallow's Cp and AIC are two other model selection criteria used. The model with four variables minimizes both of these measures.

Therefore, I will choose the model with four variables as the best model to predict the number of mammal species in the parks.

The four variables in this model are: the number of amphibian species, the number of bird species, and the number of fish species and longitude of the park. The scatterplots below illustrate the number of unique mammal species vs each variable along with a regression line and a prediction interval for each variable.



The following table lists each predictor in the model along with the beta estimate and p-value. Each predictor is significant at the alpha = 0.05 level.

Table 1: Model selected

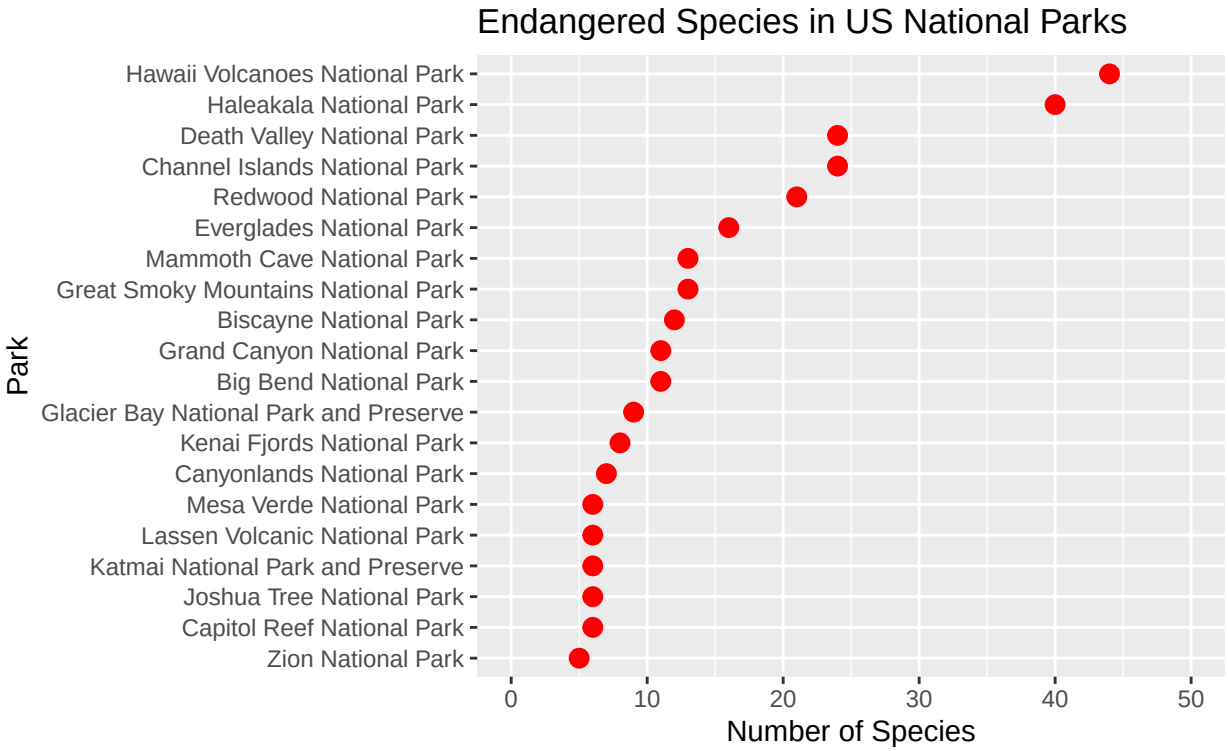
predictors	estimates	p
Intercept	-46.09	0.026
Bird	0.22	<0.001
Fish	-0.07	0.002
Amphibian	0.89	<0.001
Longitude	-0.44	0.002

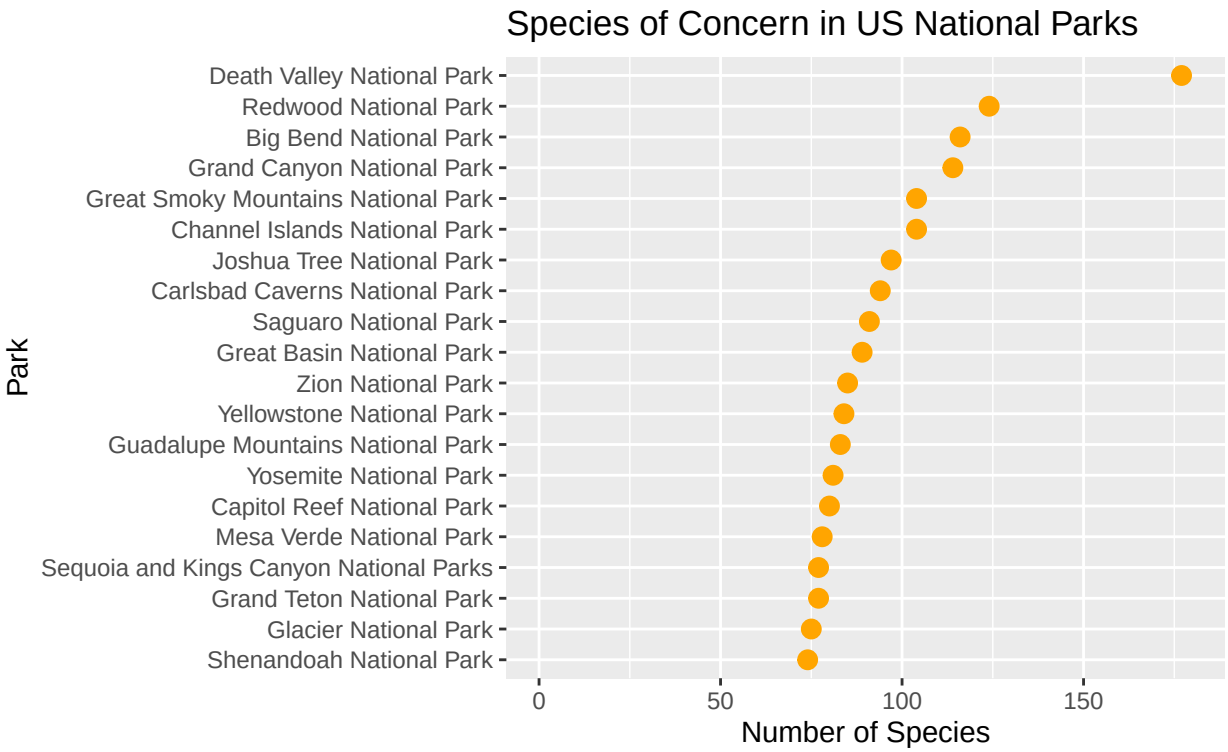
The R-square value for this model is approximately 0.67 which is not very high. It is possible that this model could still be useful in predicting the number of unique mammal species in a park if we know the number of unique bird, fish and amphibian species and the longitude of the park.

Our sample size is 56, which is fairly small. Therefore, we don't have enough data to truly claim that these variables are good predictors of the number of unique mammals in a park.

Conservation Status of Species

The data also give information about the conservation status of every species that is located in the parks. Some examples of the conservation statuses are: endangered, extinct, threatened, species of concern or no concern. Let's look at the number of endangered species and the number of species of concern at each of the parks. The following dot plots show the 20 parks with the most endangered species and the 20 parks with the most species of concern.





The Hawaii Volcanoes National Park has the most endangered species of any U.S. National Park with 44 endangered species. There are 51 parks total that reported at least 1 endangered species.

Death Valley National Park has the most species of concern of any U.S. National Park with 177 species of concern. Every single park reported species of concern with the smallest number of species reported being 21.

There are hundreds of species that are listed has having some sort of concern for conservation.

Future Work

The Biodiversity in National Parks data sets contain many pieces of useful information. I think that it will be beneficial to continue and expand data collection efforts within the National Park system. The data sets that were used for this project contained information on the number of unique species in the parks, but did not contain information about the number of unique organisms in the parks. I think that efforts should be undertaken to sample each of the National Parks to collect data on the individual organisms. This type of data collection would be a massive undertaking and would likely need to be a long term effort. I think it would also be very useful to track changes in populations over time. This would allow park rangers and the public to be aware of changing animal and plant populations. As more and more species are becoming endangered each year, tracking the number of individual organisms over time could allow us to determine if conservation efforts are successful or to predict how many years are left until another species is extinct.

Further analysis can also be conducted on the existing data sets. We could explore more model building techniques to find an improved model for predicting the number of mammal species. We could also explore if useful models can be generated to make other predictions, such as predicting the number of plants, the number of endangered species, or even the size of the park. We could further analyze the species data to find patterns between the number of specific species, such as the relationship between the number of preditor species and prey species. There are several possibilities for further analysis.