# Checks

All checks on the data- missing values, data types (integers in taxon id, and numeric in locations), and valid values (only positive values for taxon ID)- passed for all values

# Finding relationships between species

I used three different methods of clustering to find samples that may co-exist across the world.

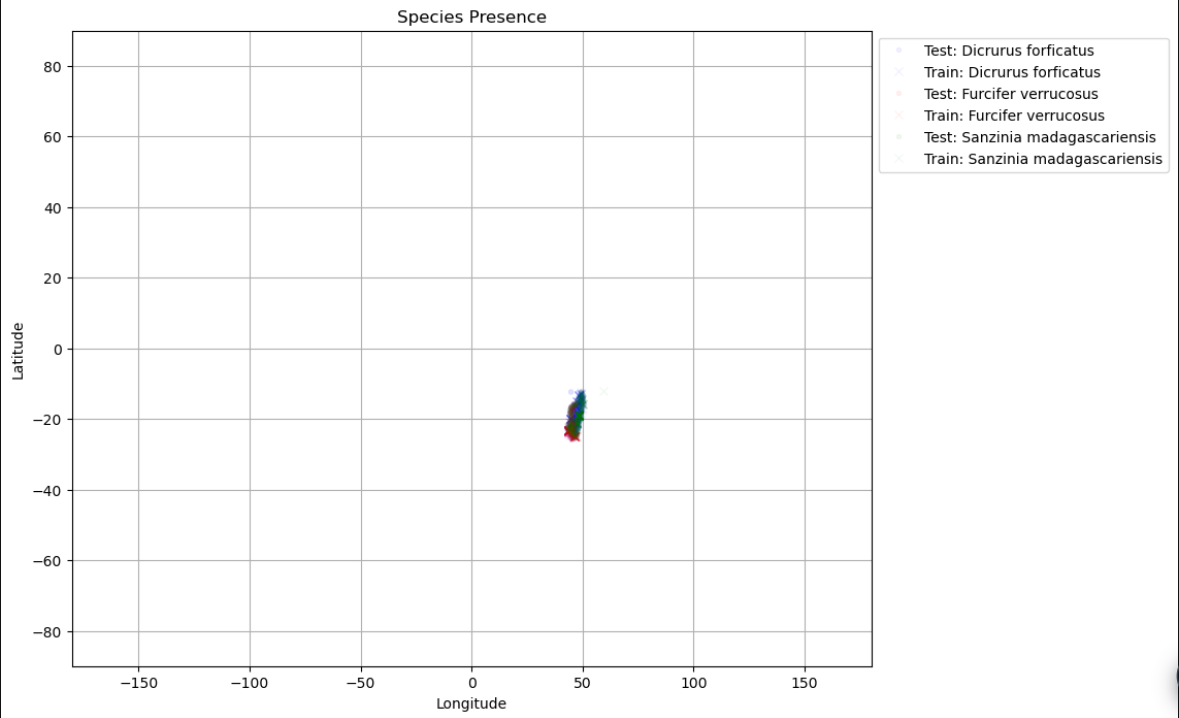
The methods used were:

DBSCAN

Agglomerative Clustering (n=10)

K-Means Clustering (n=10)

One such example is listed below, where the species Dicrurus forficatus, Furcifer verrucosus, and Sanzinia madagascariensis have highly overlapping regions. This graph is generated using Agglomerative Clustering.



Over the next week, I will be comparing the efficacy of these three methods, both within themselves by varying the number of clusters, and across the three different clustering algorithms. The algorithms will be scored according to Silhouette Scores, and Davies Bouldin Indices.

# Species Co-occurrence in relevance to project

A main critique of translocation projects involves an inability to predict how the existing fauna of the new region would be affected. Knowing which species tend to cluster together in a given habitat allows for relocation of multiple species, and may be able to give us a better idea of what target fauna in the relocation zone best looks like