**Project Summary**

The global decline in insect populations, termed the "Insect Apocalypse," poses a significant threat to ecosystem health and biodiversity, driven by factors such as human development, pesticide use, and climate change. The California Insect Barcoding Initiative (CIBI), spearheaded by the Natural History Museum of Los Angeles County, is addressing this issue by aiming to barcode every insect species in California, creating a database to track population changes. Our project complements CIBI’s efforts by utilizing the Global Biodiversity Information Facility (GBIF) database, funded by governments worldwide to provide access to data about life on Earth, to analyze the distribution of mantises (superfamily: Mantodea) in California, identifying patterns and potential gaps in their distribution. We will examine various factors including ecoregion, human population density, and climate data to predict mantis habitats and understand their ecological roles, ultimately aiding in the conservation efforts and efficient resource allocation for insect population monitoring.

In our analysis, we focus on four main data types: biodiversity, ecoregion, climatological, and human population density, all disaggregated by counties to facilitate association between datasets. The choice of counties as the primary key is due to their relatively uniform size and manageable computational requirements across California. We sourced our county data from the Database of Global Administrative Areas, ensuring high accuracy and resolution for our geographical data. This structured approach lays a solid foundation for our subsequent analysis, aiming to provide valuable insights for the CIBI and contribute to broader insect conservation initiatives.

**Biodiversity Update**

To synthesize the biodiversity of mantises in each county (Figure 1), we used geographic data for California county borders combined with GBIF data with coordinates of each mantis observation in California, containing 4,556 records of mantis observations including species name, latitude, longitude, date, and elevation of the place and time of the observation. Using coordinates from the GBIF data, we mapped each observation to a county and filtered for only unique species to see how many distinct species are found in each county, and potentially identify where more observations might be needed.

On initial observation, it appears that heavily populated areas (near Los Angeles, San Diego, and San Francisco) have more mantis diversity, which we will continue to investigate as the project progresses. We may find that this is an artifact of sampling bias.

|  |  |
| --- | --- |
| ***Figure 1:*** *A heatmap of the diversity of mantis (Mantodea) species in California counties.* | ***Figure 2:*** *Level 3 Ecoregions of California disaggregated by county* |

**Ecoregion Update**

The California ecoregion data, which comes from the US Environmental Protection Agency (EPA), contains level 3 and level 4 shapefiles which differ in granularity. After plotting both shapefiles, it appears that the more broadly defined level 3 ecoregions (*Figure 2*) would be better suited for analysis; the level 4 data splits California into an excessive number of ecoregions which do not relate well to counties.

After looking at the ecoregion data in Figure 2, we wanted to integrate it with county data by combining county and ecoregion data by intersection points and created a table with county (primary key) and a list of that county’s ecoregions. We link county data across different datasets (human population, biodiversity, climatological).

**Climatological Update**

Daymet, hosted by the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC), is a data product supported by NASA and the U.S. DOE that provides accurate, high-resolution (1km x 1km) daily weather parameters across North America. The dataset uses advanced algorithms to interpolate and extrapolate daily meteorological observations, creating detailed grids of weather parameters such as temperature, precipitation, and vapor pressure.

Daymet data is accessible through the Thematic Real-time Environmental Distributed Data Services (THREDDS) Data Server. We used “Monthly Climate Summaries on a 1-km Grid for North America, Version 4 R1” dataset, choosing to use total precipitation, maximum temperature, and minimum temperature metrics as they are the most biologically relevant to mantises. Following visualization to confirm data downloaded correctly (*Figure 3*), the data was clipped to county (*Figure 4*), and averaged across counties, producing a table with county as the primary key.

|  |  |
| --- | --- |
| ***Figure 3****: Visualization of Daymet monthly average of maximum temperature (°C) in the year 1980 across North America.* | ***Figure 4****: Example visualization of Daymet monthly average of maximum temperature (°C) in the year 1980 clipped to the boundary of Alameda County, California.* |

**Human Population Density Update**

We used human population density data from the US Census Bureau 2020 American Community Survey, focusing on the total population of each county in California. A heatmap of the population for each county can be seen below in *Figure 5*. The most populated county by a large margin was Los Angeles, followed by a few counties in Southern California and the Bay Area.

|  |
| --- |
| ***Figure 5****: Population density of California by county* |

**Next Steps**

Having successfully acquired and organized our diverse datasets (biodiversity data, ecoregion data, climatic data, and human population data) into structured 3NF tables with 'county' as a common primary key, we are now set to create an Entity-Relationship (ER) diagram. This diagram will visually represent the relationships and dependencies between these datasets, serving as a crucial tool for our subsequent analysis to identify patterns that could predict the presence of mantises across different counties in California. Utilizing machine learning models such as the Apriori Algorithm and Decision Trees, we aim to uncover meaningful relationships between mantis presence and various ecological, climatic, and anthropogenic factors. These insights will not only enhance our understanding of mantis distribution but also help identify gaps in their observed presence, providing valuable recommendations to the California Insect Barcoding Initiative (CIBI) for future sampling efforts and contributing to broader biodiversity conservation and monitoring goals.