Project Topic Proposed Workflow

2024-11-12

I. Literature review

1. Introduce the problem and explain why: Wild and managed pollinators benefit society in various ways, including bolstering food security, supporting livelihoods, enriching cultural values, and sustaining biodiversity and ecosystem health (Potts et al., 2016), which is why the study of them is a prominent subject in ecology. Pollinator populations are experiencing significant declines due to habitat loss and land-use changes such as deforestation and agricultural expansion, as well as wider "threats to global food supply and the stability of wider pollination systems in non-crop vegetation. . . .[and] disruptions of numerous plant-pollinator systems" (Mader et al., 2011). This hypothesis tests how different land-use types–mixed forests and grasslands–impact pollinator diversity. Understanding which land-use types most support pollinator diversity is critical to understanding how habitat changes affect pollinator populations because it provides greater insight into which habitats pollinators rely most heavily on. Identifying which habitat types best support pollinator diversity can inform land management strategies to protect these populations because it can provide insight into which land-use types are most critical for pollinator diversity.

2. Past work and data available: Similar questions have been raised and answered in different regions around North America. For example, in Vermont, USA, Richardson et al. found that greater species diversity was correlated with grasslands and lower was correlated with deciduous and mixed forest cover (Richardson et al., 2019). In Southeastern United States, pine forests and extensive wetlands correlated negatively with bee richness, while open canopies correlated positively (Ulyshen et al., 2024). Our project looks to answer similar questions, specifically exploring Oregon. In addition to the dataset from OBA, we will be using the National Land Cover Database (NLCD), which is a dataset that provides land cover information from across the country.

3. Purpose of the study: We are going to overlay the species from the OBA dataset on the NLCD. We will differentiate between different ranges of elevation, using elevation data in meters provided by the OBA dataset. Within each elevation range, we will assess species diversity between Mixed Forest and Grassland land cover types using A/B testing. This will allow us to answer the question of how species diversity varies with land cover types specifically in Oregon state. Additionally, this method will allow for systematic assessment of these differences while controlling for elevation across these ranges.

4. Hypotheses/questions: The primary question we want to explore with our project is how does pollinator diversity change between grassland versus forested areas within specific elevational bands in Oregon? For our hypothesis, we predict that Grassland areas have greater bee species diversity and heterogeneity compared to forested areas within the same elevational band. The null hypothesis is that there is no variation in pollinator species diversity between grassland and forested areas in Oregon. Additionally, we would like to explore the effects of elevation on bee species diversity by comparing the diversity between elevation ranges.

II. Dataset identification

We will utilize two primary datasets to address our research question on how mix forest and grassland land-use types impact bee species diversity in Oregon. Below are the details of each dataset, including metadata, collection methods, and relevant attributes.

1. **Oregon Bee Atlas (OBA)**

   - **Data Source**: Oregon Bee Atlas (OBA)
   - **Description**: The OBA dataset is compiled by a collaboration of citizen scientists and professional entomologists throughout Oregon. The dataset includes extensive records on bee species, along with their geographic and environmental context.
   - **URL**:
   - **Metadata**:
     - `Species`: The taxonomic identification of each bee recorded.
     - `Elevation`: The elevation (in meters) at which each bee was collected.
     - `Coordinates`: Latitude (`Dec.Lat.`) and longitude (`Dec.Long.`) in decimal degrees.

2. **National Land Cover Database (NLCD) - Oregon**

Data Source: National Land Cover Database (NLCD) - Oregon

Description: This dataset is part of the Oregon GIS Framework and has been specifically clipped to the Oregon state boundary. The dataset is projected in Oregon Lambert (EPSG:2992) and provides comprehensive classifications of land cover types, including grassland and forested areas.

- **URL**: https://geohub.oregon.gov/documents/9bbaa64718774bbfbf5c6ade0edf86d3/about

Metadata: - Provides detailed land cover classifications, with attributes such as:

- **Grassland/Herbaceous**:
  - *Enumerated Domain Value*: `71`
  - *Definition*: Areas dominated by graminoid or herbaceous vegetation, with greater than 80% total vegetation cover. These areas are generally not intensively managed (e.g., not tilled) but may be used for grazing.
  - *Source*: NLCD Legend Land Cover Class Descriptions.
- **Mixed Forest**:
  - *Enumerated Domain Value*: `43`
  - *Definition*: Areas characterized by trees taller than 5 meters, with neither deciduous nor evergreen species exceeding 75% of total tree cover.
  - *Source*: NLCD Legend Land Cover Class Descriptions.
- **Publication Information**:
  - *Published*: January 2019
  - *Updated*: November 2023
  - *Publisher*: U.S. Geological Survey (USGS)
  - *References*:
    Yang, L., et al. (2018). *A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies.* ISPRS Journal of Photogrammetry and Remote Sensing, 146: 108-123.

III. Workflow plan

Overview

The goal of this study is to assess the impact of habitat type—specifically grasslands versus mixed forests—on bee species diversity within different elevational bands in Oregon. To achieve this, we will use data from the Oregon Bee Atlas (OBA) and the National Land Cover Database (NLCD) for Oregon. This workflow outlines the steps for cleaning and processing the datasets, performing statistical tests, and visualizing the results to test our hypothesis that grassland habitats support greater bee diversity than forested areas within similar elevational bands.

Step 1: Data Cleaning

We will begin by importing and cleaning the raw datasets. This includes:

- **Loading the datasets**: The OBA dataset will be loaded as a CSV file, and the NLCD dataset will be imported as an img file. We will then convert the OBA CSV file to a spatial object and ensure that their formats are compatible, converting their attributes as needed and checking the CRS for both.
- **Cleaning the OBA dataset**: The columns containing species, genus, elevation, latitude, and longitude data will be cleaned by removing extra spaces and excluding any null or blank entries. Column names will be standardized for consistency:
  - Rename `Dec..Lat.` to `Lat`
  - Rename `Dec..Long.` to `Long`
  - Rename `Elevation..m` to `Elevation` Rows without a valid entry for Elevation will be dropped.
- **Creating identifiers**: To ensure unique species identification, we will combine the `Species` and `Genus` columns into a new column called `GenusSpecies`.

These initial cleaning steps are essential to remove any inconsistencies in the raw data, thereby improving the accuracy of subsequent analyses.

Step 2: Data Integration and Preparation

Once the data has been cleaned, we will integrate the OBA dataset with the NLCD data to assign each bee occurrence record to a specific land cover type. This will involve:

- **Spatial overlay**: Using the pollinator samples and their corresponding coordinates from the OBA dataset, we will overlay the bee occurrence points on the NLCD land cover map. This process will classify certain observations as occurring within grassland or mixed forest land cover types.
- **Elevation categorization**: We will divide the samples into distinct elevational bands. We will do so by finding the lower quartile (Q1), mean (Q2), and upper quartile (Q3) values for all OBA samples.
  - Low (Average elevation for a selected region is no less than 50 meters under and no more than 50 meters above Q1)
  - Middle (Average elevation for a selected region is no less than 50 meters under and no more than 50 meters above Q2)
  - High (Average elevation for a selected region is no less than 50 meters under and no more than 50 meters above Q3)

This will ensure that we can control for elevation while focusing on the effects of land type.

Step 3: Statistical Testing and Hypothesis Evaluation

To test whether there is a significant difference in bee species diversity between grassland and forested areas, we will perform an A/B test that involves the following key steps:

- **Hypothesis**: Grassland habitats support greater bee species diversity than mixed forests within the same elevational band. - **Null Hypothesis:** Assumes no difference in diversity between grassland and mixed forest habitat types.
- **Diversity metrics**: To quantify species diversity, we will calculate the *proportion of unique species*. This will be calculated by dividing the number of unique species by the total number of observations. This will account for variation in sample sizes in each region.
- **Setting up the A/B test**:

  1. We will compute the observed difference in pollinator species diversity between grassland and forest habitats.
  2. First, we will shuffle the labels (Mixed Forest vs Grassland) on our original dataset, and calculate the test statistic for this simulated dataset, which relates to the proportion of species diversity within each specified region.
  3. We will then compare the observed difference to this null distribution to calculate a p-value. If the observed difference falls outside the 95% confidence interval of the null distribution, we will reject the null hypothesis, indicating a significant effect of habitat type on species diversity.

Step 4: Visualization and Data Analysis

The following types of visualizations will effectively display our findings: - **Geographic Plot**: Plot the relevant OBA data points overlayed on the land cover regions. - **Boxplots**: To compare unique species proportion across the two habitat types within each elevational band. - **Venn diagrams**: To highlight the overlaps in pollinator diversity and species found at each habitat type.

Step 5: Interpretation and Implications The final step will involve interpreting the results of our analyses:

- If the hypothesis is supported, we expect to see a statistically significant difference indicating that grassland habitats have higher species richness and heterogeneity compared to mixed forests.
- These findings would suggest that grassland areas are more critical for supporting pollinator diversity, which could have implications for conservation strategies and land management practices. Identifying the habitats that best support pollinator populations is essential for protecting these crucial species amid ongoing habitat loss and land-use changes.

By following this detailed workflow, we aim to provide a robust analysis of how land cover types affect pollinator diversity in Oregon, ultimately contributing to informed conservation efforts.

IV. Partner contributions

Eliza:

-Identified the additional datasets we will be using
-Completed the first step of the literature review: "Introduce the problem and explain why" -Wrote most steps in the Workflow section

Emmie:

-Found articles covering relative past work and data available and wrote the paragraph on this.
-Wrote the paragraph on Part 3, which is the Purpose of the Study.
-Wrote out research questions and hypotheses into paragraph form for Paragraph 4.
-Added steps to A/B testing in the Workflow section. -Added the sources

V. Sources

New, T.R. Eric Mader, Matthew Shepherd, Mace Vaughan, Scott Hoffman Black and Gretchen LeBuhn: attracting native pollinators. Protecting North America's bees and butterflies. The Xerces Society Guide. J Insect Conserv 15, 611–612 (2011). https://doi.org/10.1007/s10841-011-9409-4

Potts, S., Imperatriz-Fonseca, V., Ngo, H. et al. Safeguarding pollinators and their values to human well-being. Nature 540, 220–229 (2016). https://doi.org/10.1038/nature20588

Richardson, L.L., McFarland, K.P., Zahendra, S. et al. Bumble bee (Bombus) distribution and diversity in Vermont, USA: a century of change. J Insect Conserv 23, 45–62 (2019). https://doi.org/10.1007/s10841-018-0113-5

Ulyshen, Michael, et al. "Spatiotemporal Patterns of Forest Pollinator Diversity across the Southeastern United States." Diversity and Distributions, vol. 30, no. 8, 2024, pp. 1–13. JSTOR, https://www.jstor.org/stable/48780052. Accessed 13 Nov. 2024.