TIEE

Teaching Issues and Experiments in Ecology - Volume 19, October 2023

ISSUES: DATA SET

Exploring species diversity across space and time with data from the National Ecological Observatory Network

Lauren Holian*, Cleber Ten Caten*, and Tad A Dallas

*Both of these authors contributed equally

Department of Biological Sciences, University of South Carolina, Columbia, SC, 29208

Corresponding author: Lauren Holian (lholian@boataid.us)

A representation of the global diversity in species and ecosystems

THE ECOLOGICAL QUESTION:

What is ecological diversity and how do we interpret its variation?

In the face of climate change and increased human pressures on wildlife and natural areas, conserving species and ecosystems has become a critical goal. But how do we decide what areas to conserve? Given that resources are limited, we might first aim to preserve areas that are ecologically important, such as the Amazon rainforest, coral reefs, or African savannas. Amongst other attributes, these ecosystems are notable in that they represent hotspots of biological *diversity*. However, diversity can be measured in a number of different ways. **What are some of the most common ways to measure diversity and how do we interpret them?**

FOUR DIMENSIONAL ECOLOGY EDUCATION (4DEE) FRAMEWORK

- **Core Ecological Concepts:**
    - Community
        - Species diversity
        - Biodiversity
- **Ecology Practices:**
    - Quantitative reasoning and computational thinking
        - Data skills – inputting and data-mining / meta-analysis/ data visualization
        - Data analysis and interpretation
    - Designing and critiquing investigations
        - Study design, familiarity with basic modes of ecological inquiry (description, comparison, experimentation, modeling)
- **Human-Environment Interactions:**
    - How humans shape and manage resources/ecosystems/the environment
        - Conservation Biology
- **Cross-cutting Themes:**
    - Spatial & Temporal
    - Scales
        - Stability & change

WHAT STUDENTS DO:

During this lab students will use small mammal data from the National Ecological Observatory Network (NEON) to develop data manipulation skills, explore measures of ecological diversity (alpha & beta diversity) across space and time, and consider the meaning of these diversity metrics in a real-world context. The appropriate data for each measure is provided in a comma-separated value (CSV) file and can be manipulated in Microsoft Excel, Google sheets, or programmatically through computing languages (e.g., R, Python, julia). Included are reference materials for students working with Excel that introduce the basic functionality of Excel and fundamental computational skills (IF, AND, SUM functions). For each diversity measure, students will have to consider the provided information (alpha-definition, beta-equation), format the data appropriately, calculate the diversity metric, create a graph representing their findings and have group discussions about their results. At the end of each part – Estimating diversity across space and Estimating diversity across time – students will answer questions that assess their understanding and prompt them to consider how each diversity metric might be used in a real-world context (i.e., conservation decisions).

STUDENT-ACTIVE APPROACHES:

Guided inquiry, open-ended inquiry, predict-observe-explain, small group discussion

STUDENT ASSESSMENTS:

In this lesson, students can be assessed on the learning objectives through several means, including quantitative outputs (e.g., biodiversity statistic calculation), visualizations (e.g., plots of richness), and textual exercises or critical thinking extension activities (e.g., discussion answers). The lesson includes several discussion questions. Faculty can opt to assess student knowledge based on participation in the discussion or have these be a written assessment. Upon completing this activity, students will be able to:

- Explore how ecological data are structured and the importance of data management. Assessment through oral/written responses to Discussion Questions.
- Create a spreadsheet data table for transcription of field collected data. Assessment through written exercises.
- Understand the differences between alpha and beta diversity. Assessment through guided exercises.
- Calculate species diversity indices. Assessment through guided exercises.
- Produce scatter plots to assess how diversity changes across the geographic and temporal space. Assessment through guided exercises.
- Gain an appreciation for how species diversity might change over short timescales. Assessment through guided exercises.

CLASS TIME:

This lab is designed for undergraduate students in a 2.5 hour lab period, but is flexible for students at various stages and may be condensed/extended to accommodate alternate schedules.

COURSE CONTEXT:

Upper-level courses (e.g., General Ecology or Community Ecology) in Biology and related majors (e.g., Environmental Science). This module should be introduced in the first half of the course when students are learning concepts of species and diversity.

SOURCES:

Small mammal box trapping dataset available in the National Ecological Observatory Network (2016). Data Product DP1.10072.001 which can be accessed online at https://data.neonscience.org/ from Battelle, Boulder, CO, USA.

ACKNOWLEDGEMENTS:

This work has been performed with funding to Tad Dallas from the National Science Foundation (NSF-DEB-2017826) Macrosystems Biology and NEON-Enabled Science program. This lab contributes to a body of teaching materials exploring ecological concepts using the NEON dataset (Styers et al., 2021a,b), we complement these previous labs by considering both spatial as well as temporal patterns of biodiversity in our teaching module. We thank Robbie Richards and Kristiaan Merritt for their thoughtful comments and conversations.

OVERVIEW OF THE ECOLOGICAL BACKGROUND

Understanding and describing how diversity changes across geographic space and over time is a major goal of ecologists and has direct implications for conservation decisions. There are several ways in which diversity can be measured, and every diversity metric has its weaknesses and strengths. Alpha diversity and beta diversity are among the most common diversity metrics used by ecologists, where alpha diversity is often defined as the number of species occurring in a specific area while beta diversity describes the fraction of unshared species between two distinct areas (Whittaker 1972). Thus, alpha diversity describes the diversity of a site in terms of number of species but ignores the identity of species and whether the same or different species are found in different sites, a facet of biodiversity that is captured by the beta diversity metric. Although alpha diversity and beta diversity are commonly taught in ecology courses, generally not enough time is devoted to explaining how these diversity metrics change across the geographic and temporal (time) space, which is challenging to convey without considering a case study.

This lab explores concepts of ecological diversity using the small mammal data available in the NEON dataset and it introduces students to data manipulation and challenges them to interpret their results under specific conditions. This lab is designed for undergraduate students in a 2.5 hour lab period, but is flexible for students at various stages and may be condensed/extended to accommodate alternate schedules. Below is a schedule for completing the lab in a 2.5 hour time period.

Part 1: Estimating diversity across geographic space

(Expected duration: 75 minutes)

1. Activity 1: Alpha diversity (Expected duration: 20-25 minutes)
    a) Calculate alpha diversity
    b) Visualize alpha diversity
    c) Explore alpha diversity and latitude
2. Activity 2: Beta diversity (Expected duration: 20-25 minutes)
    a) Calculate beta diversity
    b) Visualize beta diversity
    c) Beta diversity and geographic distance
3. Activity 3: Questions (Expected duration: 25 minutes)

Part 2: Estimating diversity across time

(Expected duration: 75 minutes)

1. Activity 1: Alpha diversity for sites JORN, MOAB, BART (Expected duration: 20-25 minutes)
    a) Calculate alpha diversity
    b) Visualize alpha diversity
2. Activity 2: Beta diversity for sites JORN, MOAB, BART (Expected duration: 20-25 minutes)
    a) Plot the mean beta diversity for a site in 2016 vs time 2017, 2016 vs 2018 and 2016 vs 2019
    b) Make a scatterplot between sites temporal beta diversity and their latitude
3. Activity 3: Questions (Expected duration: 25 minutes)

Part 3: Discussion/Interpretation

LEARNING OBJECTIVES

Upon the completion of this module, each student should be able to:

- Explain the concept and importance of biodiversity
- Understand the different ways in which biodiversity can be measured
- Understand that biodiversity changes across geographic space and over time
- Clean and organize biodiversity data
- Be able to handle biological data and visualize results

DATA SETS AND RELATED FILES

The NEON represents a long term and geographically widespread effort to understand how ecosystems are changing across the United States. The NEON is a research initiative where organismal data are collected in 47 terrestrial sites and 34 aquatic sites spread throughout different environmental domains of the United States. In terrestrial sites, where site size can range from 5 up to 215km^2, data is collected for seven taxa (small mammals, breeding landbirds, ground beetles, mosquitoes, ticks, plants and soil microbes) across different habitat types present in these sites so that species occurring in different habitats have the chance of being collected. Similarly, data is collected for 5 taxa (aquatic microbes, phytoplankton, macroinvertebrates, zooplankton and fishes) across 1-km-long transects in wadeable streams, lakes and rivers in aquatic sites. For the present activity, we are going to use the small mammal data present in the NEON dataset to answer diversity-related questions using alpha and beta diversity metrics. Small mammals are defined as mammals that are non-volant (mammals that cannot fly), nocturnal, forage mostly above ground and weights between 5 to 500 g in the NEON dataset. One of the reasons that small mammals are an ideal group to study diversity patterns is because this taxonomic group has several species with different ecological characteristics. For example, species may be geographically widespread or restricted, and many species specialize on foraging on different types of habitats. This wide range of ecological characteristics and level of specialization of small mammals make them an ideal group to evaluate alpha and beta diversity patterns.

The data used in this project can be assessed through the following link: https://doi.org/10.48443/j1g9-2j27

Files used by faculty and students: Spatial.csv and Distances_KM.csv for the activities in Part 1 and Temporal_JORN.csv, Temporal_MOAB.csv, and Temporal_BART.csv for the activities in Part 2. Faculty can also use any of the additional files provided in this teaching module that have temporal data for all sites sampled in NEON in case they wish to focus more on the temporal aspect of this module. Data files are named according to the following convention Temporal_SITENAME.csv.

The file Hw2.csv is to be used by faculty. This file has all the formulas and calculations used to get to the answers for all questions asked in this teaching module as well as the plots associated with each question.

STUDENT INSTRUCTIONS

Before class

Watch introductory video "NEON: Open Data to Understand our Changing Ecosystems." https://www.youtube.com/NEONScience

What is NEON?

The National Ecological Observatory Network (NEON) is a Long-Term Ecological Research (LTER) station where organismal data are continuously collected in 47 terrestrial sites (Thorpe et al. 2016) and 34 aquatic sites (Goodman et al. 2015) spread throughout different environmental domains of the United States (Figure 1). In terrestrial sites – varying in area from 5 up to 215km^2 – data are collected for seven group of species (small mammals, breeding landbirds, ground beetles, mosquitoes, ticks, plants and soil microbes) across different habitat types present in these sites, so that species occurring across all available habitats have the chance of being collected. Similarly, data are collected for five groups of species (aquatic microbes, phytoplankton, macroinvertebrates, zooplankton and fishes) in aquatic sites across 1-km-long transects in wadeable streams, lakes and rivers. Data collected by LTER stations are important because there are ecological questions that cannot be answered with short-term observations or experiments, and this type of research is key to understanding what factors lead to change in populations (i.e., individuals of the same species living in the same location) and communities (i.e., the set of species in a geographic location) over time as well as to test ecological theory.

For the present activity, we are going to answer diversity-related questions, specifically alpha and beta diversity, using the small mammal data present in the NEON dataset. NEON defines small mammals as mammals that are non-volant (mammals that cannot fly), nocturnal, forage mostly above ground and weigh between 5 to 500 g. Small mammals are ideal for our investigation of diversity because they have a wide range of ecological characteristics, such as their range size (some species are widespread and some are restricted), diet/foraging preferences, and specific habitat requirements. This broad range of ecological characteristics makes them an ideal group to evaluate patterns of diversity through the lenses of alpha and beta diversity. But before evaluating these diversity patterns, we first need to define and understand what exactly alpha and beta diversity are.

TIEE

Teaching Issues and Experiments in Ecology - Volume 19, October 2023

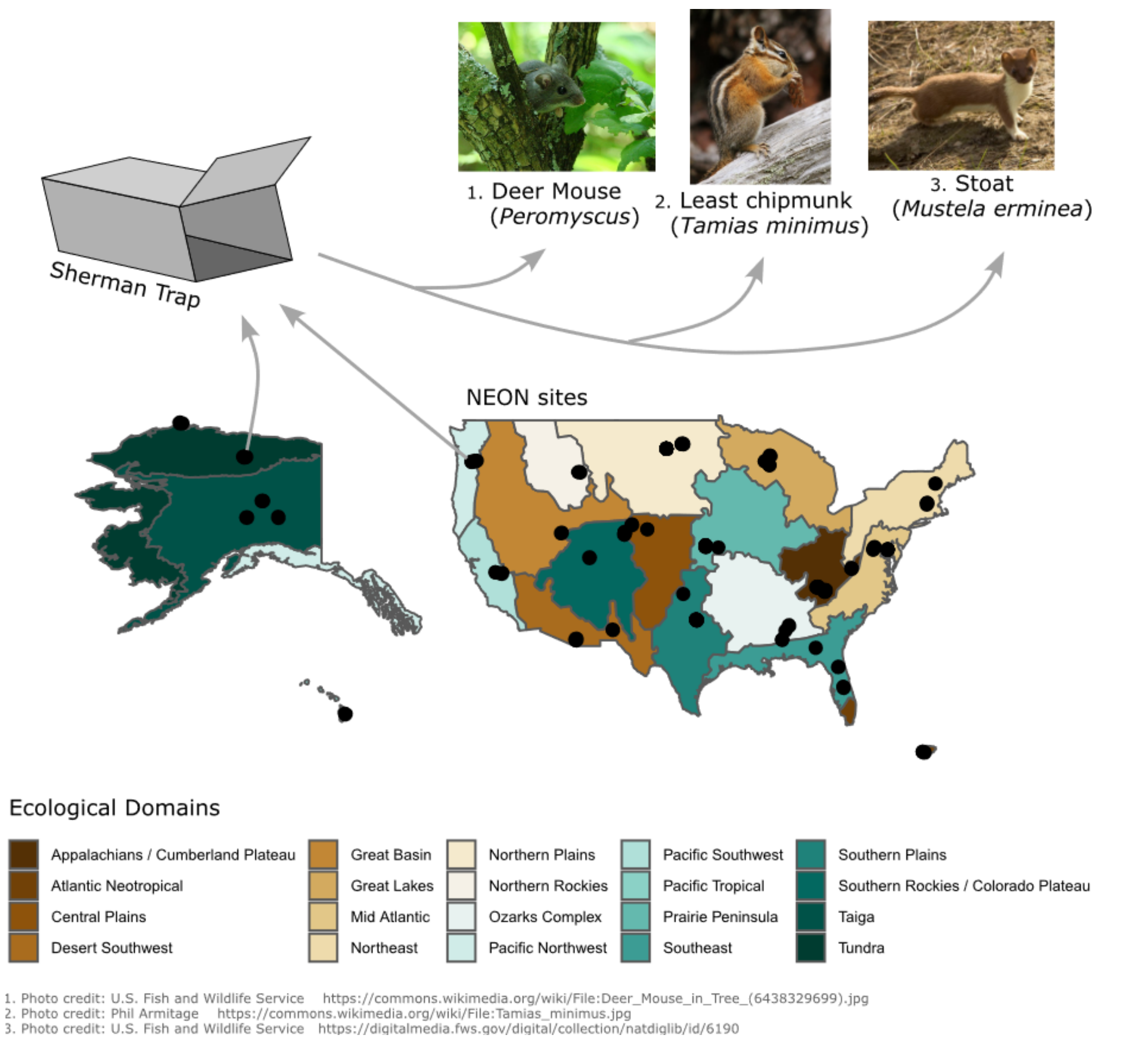


Figure 1: NEON small mammal sampling across ecological domains in North America. These ecological domains represent the different environment types that span North America. Black circles indicate NEON sites where small mammals are sampled. Sampling is conducted via Sherman traps. These traps are non-lethal and allow for a diversity of small mammals to enter for bait and be identified and counted by scientists several hours later.

Alpha diversity

When ecologists refer to the alpha diversity (also referred to as *species richness*) of a geographic area (or *site*), they are concerned with the number of species occurring in that specific location (Figure 2, Whittaker (1972)). Although alpha diversity can also refer to the *evenness* (distribution of abundances) of species in a geographic area, for this lesson we will focus on the aspect of species richness. In that case, for example, the term biodiversity hotspot means that an area has greater species richness than most other areas being considered. Some examples of areas that have particularly high alpha diversity are the Amazonian rainforest or the coral reefs of Polynesia. There are thousands of species occurring in these areas and the high diversity of such areas has been linked to ecosystem functioning (e.g., forests might remove carbon dioxide from the air and wetlands might filter out heavy metals from water) and ecosystem resilience (e.g., an ecosystem's ability to recover from a disturbance such as a wildfire). Given the importance of diversity and its implications for conservation and human processes, it is important to understand how diversity changes over the geographic and temporal space, so we can take effective conservation measures.

If our aim was to conserve/preserve the most biodiverse areas, where would we focus our attention?

In the case of alpha diversity, areas that have more species would be considered to have a greater diversity than locations that have fewer species.

Alpha diversity is a great first step at estimating species diversity across geographic space and through time and is particularly useful since the only information needed is the number of species occurring (though even estimating those species can be complex and there is a whole subfield of ecology tackling issues of species occurrence). However, while alpha diversity can be a powerful tool, it does not consider the identity of species, so how would we compare sites that all have the same species richness?

Alpha diversity is the number of species (i.e., species richness) occurring in a given location.

Beta diversity

Beta diversity is a measure of biodiversity that takes into account both the number of species found in a location and how different the species occurring in a location are from species occurring in other locations (Figure 2, Whittaker (1972)). An important difference between alpha and beta diversity is that beta diversity is a pair-wise measure that compares the diversity of two different sites while alpha diversity measures the diversity of a single site. Thus, beta diversity captures a facet of diversity that is ignored by alpha diversity, i.e., measuring how unique an area is in terms of the community of species compared to a different area.

If our aim was to conserve/preserve the most biodiverse areas, where would we focus our attention? In the case of beta diversity, locations that have the same species would be considered low diversity, whereas locations with unique species would be considered high diversity. As you can see, the answer to the "focus of our attention" might differ depending on the metric of diversity considered. This highlights the importance of considering diversity from multiple angles and understanding the implications of the different metrics.

An important goal of ecology is understanding why different locations have different communities of species (the set of species in a geographic location). Among the factors that lead to differences in species' occurrences are the environmental conditions and accessibility of particular locations. Differing environmental conditions can lead to different communities because species vary in their environmental requirements (i.e., polar bears occur in the arctic and require sea ice for hunting grounds but black bears would be poorly suited for hunting on the sea ice). However, even when two locations have similar environmental conditions, there can still be different communities. This could happen, for example, when the two locations are far from each other. In other words, the spatial distance between environmentally similar locations might explain why they have different species. This can arise when species are unable to access suitable habitats from limitations in their ability to move there or from geographic barriers such as rivers, mountains, or cities.

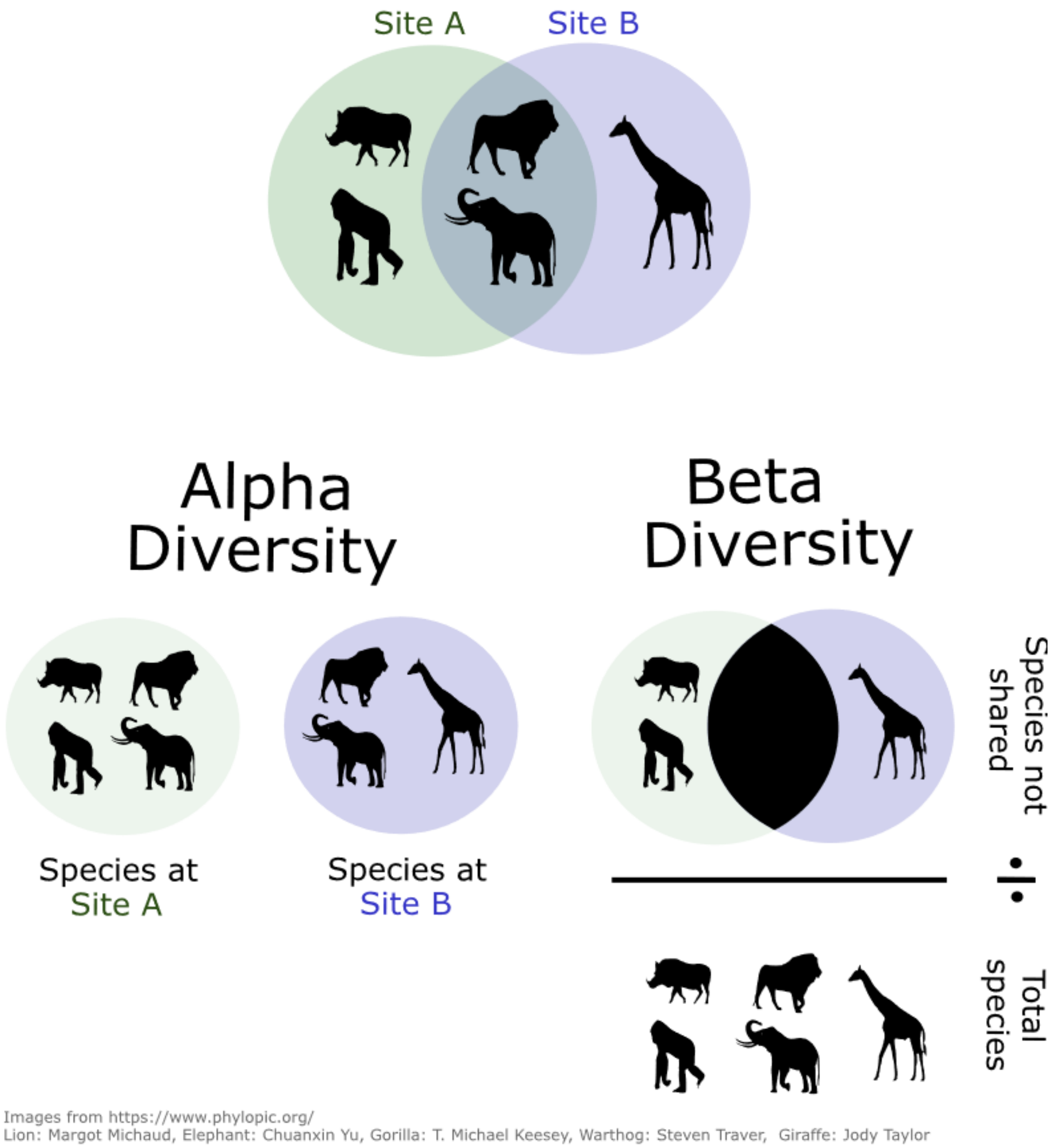


Figure 2: Consider a landscape where the circles represent different sites (A or B), and the animals that can be found at each site. Though some species are present in both, several species are unique to either site A or site B (top Venn diagram). The diversity of these sites can be measured as the **alpha diversity** of each site (bottom left Venn diagram), and the **beta diversity** across sites (bottom right). The simplest method of quantifying alpha diversity is simply the number of species present at a site (i.e., 4 = site A and 3 = site B). Beta diversity accounts for differences between sites in terms of the species they contain. The simplest method for calculating beta diversity is using the Jaccard index, where we take the number of species that are not between the sites (there are 3 unshared species between site A and B) and divide it by the total number of unique species present at the sites (5 species present).

Alternative uses of alpha and beta diversity

Throughout this lesson alpha and beta diversity are going to be estimated considering species richness and number of unshared species, but these diversity metrics can also be estimated considering the characteristics of the species being observed in an area. In these cases, alpha and beta diversity are still going to be estimated following the same concepts presented earlier, but different factors will be considered in the calculations. For example, alpha and beta diversity can be estimated considering the phylogenetic history of the species found in a location, such that phylogenetic alpha diversity is higher when more distantly related species are occurring in a given site (Swenson et al. 2012). Similarly, the functional alpha and beta diversity can also me estimated considering the functional traits (e.g., body size or litter size) of the species occurring in a site, such that functional alpha diversity is higher when more different species are occurring in an area (Swenson et al. 2012). Nonetheless, the focus of this lesson is to estimate alpha and beta diversity considering species richness and amount of unshared species across sites.

Diversity across space

In order to understand why some places are hotspots of biodiversity and why some places have less biodiversity, it is useful to think about how diversity varies across geographic space. In particular, it might be useful to take a step back and look at this at the global scale (Figure 3).

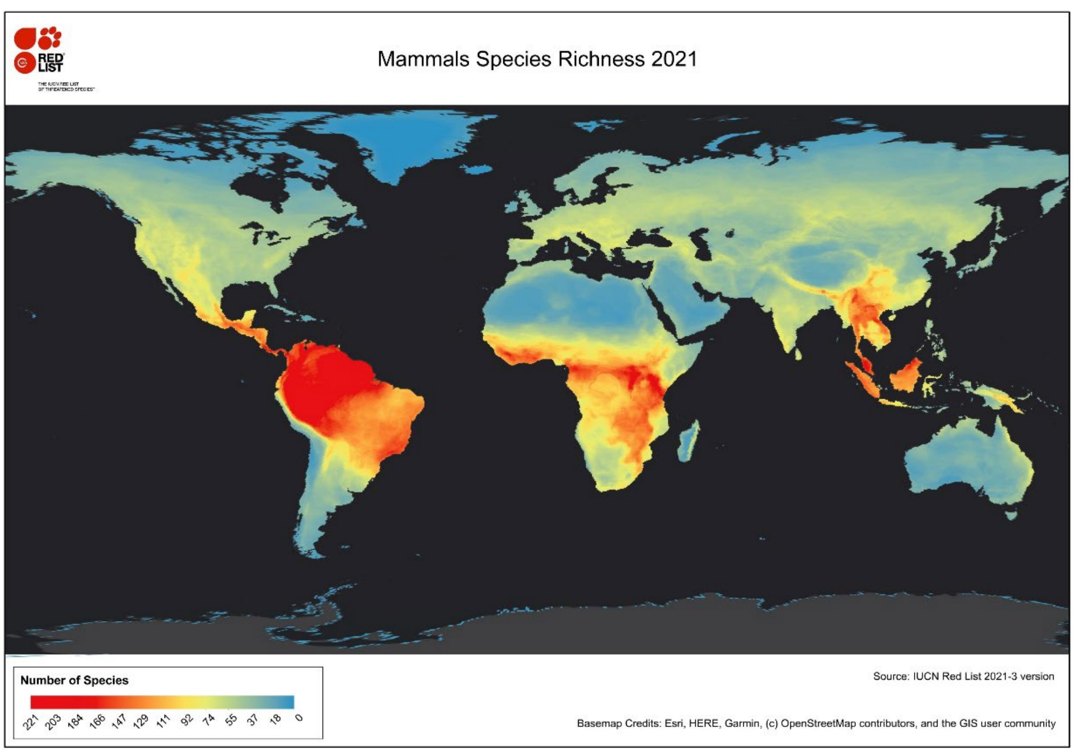


Figure 3: This is the International Union for Conservation of Nature map of non-marine mammal species richness (or alpha diversity) for 2021. The variation in color represents the number of mammal species potentially occurring in a geographic location, with areas in red indicating high species richness and areas in blue indicating relatively low species richness. Do you see patterns in areas of low and high diversity? Why do you think that is?

https://www.iucnredlist.org/resources/other-spatial-downloads

You may have noticed that areas closer to the equator tend to have more species than those farther away. This is referred to as the *latitudinal diversity gradient* and is considered a particularly interesting pattern of diversity. The *latitudinal diversity gradient* describes how regions closer to the tropics have more species than regions that are farther from the tropics (Pianka 1966). Thus, there is a negative relationship between absolute latitude and number of species found in an area, and this diversity pattern has been described for most aquatic and terrestrial species.

Several climate-related hypotheses have been proposed to explain the latitudinal diversity gradient (Pianka 1966). Among them is the hypothesis that tropical regions have more species because they have a higher energy than the other regions of the globe. In this case the higher solar energy in the tropics leads to an increase in the net primary productivity (i.e., plants store more energy that can be used by the organisms that consume them) in these regions, which make it possible for more individuals of different species to be supported in such areas. On the other hand, the climate stability hypothesis posits that there are more species in the tropics because the climatic conditions are more stable in those regions (i.e., differences in temperature/precipitation from day to night & across seasons). In this case it is suggested that higher climatic stability would allow species to specialize in specific resources, which would make it possible for more species to coexist in the same location.

There are also evolutionary and historical hypotheses that have been proposed to explain the latitudinal diversity gradient (Pianka 1966). An example of an evolutionary hypothesis is the evolutionary speed hypothesis that suggests that there are more species in the tropics because the species occurring there have shorter generation times (i.e., there is a shorter average time between two generations) and this would lead speciation rates (the rate at which new species are formed) to be higher in the tropics. This would then culminate in the tropics having more species than other regions of the globe. On the other hand, the historical perturbation hypothesis posits that there are more species in the tropics because there has not been enough time since historical perturbations for species to colonize areas outside the tropics. Glaciations are an example of such perturbations that often occur in the temperate zones and lead to the extinction of species in these regions. Glaciations could then lead to relatively lower species richness in temperate versus tropical regions. Thus, there are several different mechanisms that could have led to the often-observed latitudinal diversity gradient and it is currently thought that several of these mechanisms might have played a role in the causation of the latitudinal diversity gradient.

Diversity across time

Patterns of species diversity might not only change over geographic space, they can also change over time (Dornelas et al. 2014). Changes in environmental conditions over the years, such as the ones caused by climate change, can lead locations to become environmentally unsuitable for species that were already present and/or can make locations suitable for new species. For example, climate change and land-use change have been causing the extinction of rare species that depend on specific environmental conditions to survive. On the other hand, there is also evidence that extinctions caused by human activities such as deforestation and waste management can be offset by new species colonizing those locations. Thus, patterns of temporal diversity can be very complex, and some locations might experience species gains, increasing their alpha diversity over time whereas other places might experience species loss and have their alpha diversity decrease over time. As with diversity across space, it is important to consider multiple types of diversity, and the identity of species gains and losses (i.e., if rare species tend to be lost, what kinds of species might be gained?).

Part 1: Estimating species diversity across geographic space

In your reading you learned about alpha and beta diversity and some potential drivers of the latitudinal diversity gradient. Now, you will use a real dataset of ecological samples collected across the continental United States to investigate these measures of diversity for small mammals. You will use the Spatial.csvfile to answer the activities in this section.

Activity 1: Alpha diversity

For this activity you will calculate the alpha diversity (species richness) for all the sites in the Spatial.csv file and create a plot comparing the alpha diversity of the considered sites. Reference the background information and Excel guide as needed to help you. It may help to think about alpha diversity first and how you will calculate it, then consider the data you have and how to get it to where it needs to be (this will require some manipulation of the data, Figure 4).

Below are guidelines for how to approach the activity.

Calculate alpha diversity for all sites found in the Spatial.csv file

1. Format the data into a presence/absence format using the **IF()** function. This means that each time a species occurs it counts as 1 (presence) and zero each time it is absent.

Rows	Column A	Column B	Column C	Column D	Column E
1	0	0	0	0	6
2	1	1	0	0	8
3	2	4	7	1	3
4	3	8	0	1	10

Rows	Column A	Column B	Column C	Column D	Column E
1	0	0	0	0	1
2	1	1	0	0	1
3	1	1	1	1	1
4	1	1	0	1	1

Using the **IF()** function we can set conditions for what excel should do under different circumstances. The bottom table assigns each cell a “1” or “0” depending on whether the corresponding cell in the top table has a value greater than zero.
Use this function by typing **=IF()** and then provide the condition you want. The setup is **=IF(condition, what to do if true, what to do if false)**.

Example in the formula bar:

=IF(F12>0, 1, 0)

Figure 4: Example illustrating the usage of the **IF()** statement in excel to transform count data into binary (i.e., presence and absence) data for the NEON sites where species were observed. This is an important step to calculate the alpha and beta diversity of sites.

2. Ensure that all values are either zero or one and that they correctly correspond to the species of the Spatial.csv datasheet.
3. Calculate alpha diversity at each site.

Visualize alpha diversity across sites

Make a scatter plot showing the relationship between the alpha diversity of a site and site latitude. Which variables should be on the Y and X axes of this plot? (Hint: the Y axis should be the dependent or response variable and the X axis should be the independent or predictor variable). Add a trend line to the figure.

1. Create a third datasheet and copy the last column (alpha diversity column) to this new datasheet. Note that when copying it may copy the formula rather than the value, so make sure you right click and then select values (clipboard icon with numbers) under paste options.
2. Rename the datasheet "AlphaPlot"
3. Copy the site names and corresponding latitudes from Latitudes.csvfiles and paste it alongside the alpha diversity column.
4. Highlight the totals and the appropriate column names and select "Insert" and

then "scatter" under "Charts".

Activity 2: Beta diversity

For this activity you will calculate the beta diversity between the sites BART, BLAN, and HARV found in the Spatial.csv file and create a plot of the beta diversity and distance between sites. Distances between sites in kilometers can be found in the Distances_KM.csv file. Reference the background information and Excel guide as needed to help you.

Calculate the beta diversity between sites: BART, BLAN, and HARV

1. Copy the converted table of presences and absences to a new sheet and rename it "Betadiversity"
2. Use the beta diversity formula (representing the Jaccard index) below to calculate the differences between each site.

Beta diversity = Total unshared species/total species = (Species at site X - Shared species between sites X and Y) + (Species at site Y - Shared species between sites X and Y) / Total unique species of site X and Y

Note that this equation produces values between 0 and 1, where values of zero indicate sites that share all of their species and values of 1 indicate sites where no species are shared.

To put into more mathematical terms, we use *set theory*, where the total unshared species are those that are unique to site X plus those that are unique to site Y. This would be represented as the species found at site X minus the shared species ($X \cap Y$), where $\cap$ refers to the 'intersection' of X and Y. The total unique species between the two sites is ($X \cup Y$), referring the 'union' of X and Y, where we make sure to not "double count" species which occur in both sites.

- Create a new table for the number of species shared between each set of sites (BART ∩ BLAN), (BART ∩ HARV), (BLAN ∩ HARV)
- Fill the cells with the number of species shared between each set of sites (Consider using an IF statement and consult excel cheat sheet provided)
- Use these values to fill out the beta diversity equation shown above for each combination of sites (HINT: Each site should be compared to the two other sites in this assignment. Thus, each site will have two values of beta diversity)

Visualize the relationship between beta diversity and distance

Make a scatterplot showing the beta diversity between sites and the distance between them (this information can be obtained from the Distances_KM.csv file). Which variables should be on the Y and X axes?

Activity 3: Discussion

Alpha Diversity

1. Which site is the most diverse and the least diverse? What could be driving this?
2. Which sites should be targeted for conservation according to the alpha diversity metric?
3. Is there any apparent relationship between site alpha diversity and latitude? Why do you think that happened? Potential reference: https://www.neonscience.org/field-sites/about-field-sites

Beta Diversity

1. Which sites are the most similar and the most different to each other?
2. *Blarina brevicauda* is a species found in all three sites whereas *Zapus hudsonius* is found only in the BLAN site. How do you think these species affected the calculation of beta diversity between these sites?
3. Are sites that are spatially closer more similar to themselves? What do you think is causing this?
4. Group question: Suppose you have a limited budget and your task is to preserve the most unique NEON site out of all sites considered in this activity, which site would you preserve and why?

Part 2: Exploring diversity through time

In Part 1, you calculated alpha and beta diversity at different sites and considered how distance or geographic location might affect these measures, but what about how these measures might change across time? To answer the activities in this section you will be using the Temporal_JORN.csv, Temporal_MOAB.csv, and Temporal_BART.csv files.

Activity 1: Alpha diversity across time for sites JORN, MOAB, and BART

In this activity you will calculate alpha diversity across years (2014-2019). To do so, you will repeat the procedures from Part 1, activity 1 (Alpha Diversity) but for the remaining years. When you are done with this portion you should have a measure of alpha diversity for sites JORN, MOAB, and BART for each year (2014,2015,2016,2017,2018,2019). Information on each of these sites can be obtained in the Temporal_JORN.csv, Temporal_MOAB.csv, and Temporal_BART.csv files.

Plot the relationship between alpha diversity and years

Make a scatter plot for each site (JORN, MOAB, and BART) showing the relationship between alpha diversity and year (HINT: each site should have its own plot showing its alpha diversity over the years). Which variables should be in the Y and X axes? Make sure to include axis titles.

Activity 2: Beta diversity across time for sites JORN, MOAB, and BART

Now, you will calculate the temporal beta diversity of sites JORN, MOAB, and BART. In part 1, beta diversity was calculated considering two different sites, here you will calculate the beta diversity of the same site considering different time periods. Specifically, you will calculate the temporal beta diversity of each site comparing the year of 2016 against 2017, 2018, and 2019.

Plot the relationship between beta diversity and years

In this plot, you will show the temporal beta diversity over the years. Which variables should be in the Y and X axes? (HINT: each data point in this plot should represent the comparison between the year of 2016 to 2017, 2018, and 2019. Thus, there will be three data points in the plot for each site)

Activity 3: Discussion

Alpha Diversity

1. Which site has the most stable diversity value across time? The most varied?
2. Do sites tend to be increasing, decreasing, or remaining the same in diversity across time?
3. What might be causing species losses and gains across sites?

Beta Diversity

1. Which site changes the most over time?
2. What does it mean when a site's beta diversity is changing over time but its alpha diversity is remaining the same?
3. In general, was beta diversity higher comparing 2016 to 2019 than comparing it to 2017 and 2018? What if we waited for 20 more years and calculated the beta diversity between 2016 and 2042, what do you think would happen?

Synthesis: Looking at your findings for small mammals across the United States, which sites would you prioritize for conservation and why? What information supports your decision and what are the drawbacks? Is there more information that you would want when weighing these options?

REFERENCES

Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. Assemblage time series reveal biodiversity change but not systematic loss. Science 344:296–299.

Goodman, K. J., S. M. Parker, J. W. Edmonds, and L. H. Zeglin. 2015. Expanding the scale of aquatic sciences: the role of the National Ecological Observatory Network (NEON). Freshwater Science 34:377–385.

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. The American Naturalist 100:33–46.

Styers, D. M., J. L. Schafer, M. B. Kolozsvary, K. M. Brubaker, S. E. Scanga, L. J. Anderson, J. J. Mitchell, and D. Barnett. 2021a. Calculating and comparing biodiversity materials across sites using NEON herbaceous vegetation data - a tutorial. QUBES Educational Resources. doi:10.25334/DKBX-8394

Styers, D. M., J. L. Schafer, M. B. Kolozsvary, K. M. Brubaker, S. E. Scanga, L. J. Anderson, J. J. Mitchell, and D. Barnett. 2021b. Developing a flexible learning activity on biodiversity and spatial scale concepts using open-access vegetation datasets from the National Ecological Observatory Network. Ecology and Evolution 11:3660–3671.

Swenson, N. G., D. L. Erickson, X. Mi, N. A. Bourg, J. Forero-Montaña, X. Ge, R. Howe, J. K. Lake, X. Liu, K. Ma, N. Pei, J. Thompson, M. Uriarte, A. Wolf, S. J. Wright, W. Ye, J. Zhang, J. K. Zimmerman, and W. J. Kress. 2012. Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. Ecology 93:S112–S125.

Thorpe, A. S., D. T. Barnett, S. C. Elmendorf, E.-L. S. Hinckley, D. Hoekman, K. D. Jones, K. E. LeVan, C. L. Meier, L. F. Stanish, and K. M. Thibault. 2016. Introduction to the sampling designs of the National Ecological Observatory Network terrestrial observation system. Ecosphere 7:e01627.

Whittaker, R. H. 1972. Evolution and measurement of species diversity. Taxon 21:213–251.

NOTES TO FACULTY

Classroom Management

This exercise has the goal of helping students develop a deeper understanding of the concepts of diversity metrics used in ecology while also practicing their ecological data management and visualization skills. This exercise provides general background on the concepts and estimation of alpha and beta diversity and we suggest the instructors use this exercise when these concepts are being introduced in the classroom. We highlight the effects of space and time on these diversity metrics and we provide real-world data that can be used to measure and test some hypotheses that have been proposed to explain these diversity patterns.

This activity can be completed individually or in small groups of students working together depending on the level of familiarity of the students with Excel. The Student Instructions have been written to provide general directions on what the students should do to be able to answer a specific question. However, the instructions can be modified to provide more or less details to the students. For example, in part 1 the students are instructed to transform the data into presences (1) and absences (0) using the IF command in Excel, but if the students are not familiar with Excel a step-by-step of how to do this transformation is also provided and can be given to the students.

Some of the questions asked during this activity are focused on a subset of three sites that are sampled in the NEON project. However, we make available data for all 47 sites that could be used for this activity, which gives the instructors more options as to how they want to run the lab. For example, more sites can be used to explore in a more comprehensive manner the relationship between geographic distance between sites and beta diversity (i.e., part 1). On the other hand, the instructor can also narrow down the number of sites used in the activity if, for example, time is a limiting factor in the lab. Similarly, we also provide data for two more years that the instructor could choose to use during part 2 in case the instructor wants to focus on the temporal aspect of alpha and beta diversity. Similarly, parts 1 and 2 of this teaching module can be done separately. If the instructor desires to focus solely on the spatial patterns of diversity, then only part 1 of this module can be done. Likewise, if the interest of the instructor is on the temporal aspects of diversity, then they can do only part 2 of this module.

Part 1

The first part of this activity the students will need to manipulate the data that are available in an abundance format to transform it into a presence/absence format. We provide background information, and an Excel cheat sheet that can be used by the students to help them transform this data. If the instructor wants to focus on the data manipulation aspect of the lab, then they can opt to provide a shorter version of the cheat sheet to the students.

The overall goal of part 1 is to emphasize how diversity metrics can vary spatially and that sites that are close to each other are going to tend to be more similar in terms of species richness (alpha diversity) and species composition (beta diversity). Here it is important to highlight the differences in which alpha and beta diversity are calculated such that alpha diversity is a site measure whereas beta diversity is a pairwise (i.e., a comparison between sites) metric. The figures that students plot are an important component of this lesson because it is through them that the students will be able to visualize the relationship between alpha diversity and latitude and beta diversity between sites and their distance from each other. The instructor can also opt to add more sites in the activity so that the students can explore more thoroughly how these diversity metrics vary spatially.

In activity 3, we present seven potential questions that the instructor could ask the students about the spatial patterns of alpha and beta diversity. To be able to answer the first three questions the students will need to have done activity 1 whereas the last four questions rely on the students having done activity 2. Thus, the first two activities can be done independently from each other, which gives the instructor the option of focusing on any of those two activities if desired. For example, if the instructor wants to focus on the spatial aspects of beta diversity they can tell the students to do only activity 2 and to answer only the questions about beta diversity.

The instructor can also select only a few questions to be answered during activity 3. For example, some of the questions asked are more straightforward to answer whereas others are more thought-provoking and might require more time to be answered. Examples of straightforward questions are the first questions in the alpha diversity and beta diversity subsections while the last few questions in both sections are more complex. Thus, depending on the goal of the lab, the instructor can ask the students to answer all questions or only a subset of questions that they think is suitable for the purpose of their class. For example, if this lab is being taught to an upper-level class, then the instructor might prefer that the students answer all questions whereas if the lab is being taught in an introductory class, then the instructor might prefer to ask the students to only answer the more straightforward questions.

Part 2

In the second part of this activity the students will need to take similar steps as in part 1, where they will be given data in an abundance format and they will need to transform it into a presence/absence format. In this case the biggest difference lies in terms of how the data are organized. While in part 1 we had species in columns and sites in rows, in this part we still have species in columns but now we have year in rows. Thus, the students will be calculating temporal alpha and beta diversity of the same site at different time periods. When students are calculating temporal alpha diversity of a site, they should simply sum the number of unique species that occur in a site over specific years. So, if at time *i* a given site has species A, B and C and at a time *i*+1 the same site has species B, D and E, then the temporal alpha diversity of the site 3 for each considered year (i.e., B is counted independently for each year). Beta diversity will be calculated in the same way as in part 1, the only difference is that the students will compare the year of 2016 to 2017 and 2018.

The goal of this part is to highlight that alpha and beta diversity patterns not only vary across the space, but that they can also vary over time. This temporal variation in the diversity of a site might be less intuitive to students, so we suggest that the instructor remind the students of why the diversity of a site could change over time. Some examples as to why the diversity of a site would change over time could be the species phenology (i.e., the timing of seasonal events like flowering in plants or hibernation) or the seasonal changes in climatic conditions that might render the environment unsuitable for a species occurrence in a given location during specific times of the year.

The activities in part 2 are structured similarly to part 1, where students will calculate the temporal alpha and beta diversity of sites and plot their results. However, in this part we only have three questions for each diversity metric. To answer the first three questions the students will need to have completed activities 1 whereas the last three questions rely on the students having done activities 2. As in the previous part, these sets of activities can be done independently, and thus the instructor can opt to focus on temporal alpha or beta diversity if desired. While a few of the questions asked in this part are straightforward to answer, we believe that in general these questions might be a little bit more challenging to the students. Three out of the six questions are going to ask the students to explain their answer, which might not be as intuitive as the questions in part 1. For example, the third question in the alpha diversity subsection question might require the students to consider the idea of different levels of environmental stability across latitudes. For that reason, the instructor might opt to leave some of the discussion questions out depending on the level of the lab being taught.

Potential Extensions

Here we present some extra topics that could be taught during this lab in case the instructor wants to cover patterns of biodiversity in more detail.

Species-area relationship

The species-area relationship is a topic that could also be addressed by the instructor during this lab given that this relationship might directly affect the observed patterns of biodiversity. The species-area relationship links the species richness of a site with its area, such that locations with larger areas will have higher species richness. Thus, the alpha diversity of a sampled site is linked to the area that was sampled, and comparing the alpha diversity of a sampled site that have significantly different areas might pose an additional challenge when trying to disentangle the effects of climate, phenology, and area on species richness.

While the effects of sampled area on beta diversity might be more uncertain, one could expect that when sites are closer, a larger sampled area will lead to a lower beta diversity between the sites. Conversely, if the sampled sites are far from each other, a larger sample might lead to a higher beta diversity between those sites. Thus, it might be a good question to ask students what they would think would happen with the beta diversity of two sites if they have large areas sampled.

Rarefaction curves

A question that students might ask during this lab is how we know that we have sampled all (or most) of the species that occur in a site. This is an extremely important question because incomplete sampling between sites is going to bias the results of these diversity metrics. In this case, the instructor can introduce the topic of rarefaction curves to the class. Rarefaction curves allow one to evaluate whether most of the species that occur in a community have been sampled. The way in which rarefaction curves work is that one can calculate the number of species that are added to a community every time a new individual is sampled in that community. Thus, one only needs the sampled data to calculate a rarefaction curve for a site. The plot for a rarefaction curve will have species richness on the Y axis and number of individuals sampled on X axis.

An extreme example when calculating a rarefaction curve might be when there have been 99 individuals sampled in a site and those individuals represent 3 species (each species having 33 individuals). To calculate a rarefaction curve in this case, we could sample for example 60 individuals and see how the species richness would change as individuals are sampled. It is likely that after having sampled a handful of individuals we will have sampled all the three species in the community, and thus, our curve will get flat because sampling more individuals will not lead to more species being found in that community. When a flat curve like this hypothetical example is observed when calculating rarefaction curves, it means that we have successfully sampled most of the species in a community. On the other hand, if those 99 individuals represent 99 different species, then every time a new individual is sampled a new species added to the community, and in this case the rarefaction curve will not be flat, which indicates that there are likely more species in that site than what we have sampled. Thus, one can use these rarefaction curves to establish whether they have sampled most of the species in different communities and whether it makes sense comparing the alpha diversity of these communities.

Beta diversity metrics that consider species abundance

One of the first steps that students should take when doing the activities we propose is to transform the species data into a presence/absence format. While this is required to calculate the alpha diversity and the beta diversity metrics that we introduce in this teaching module, there are also other beta diversity metrics that take into account the abundance of each species in a site to measure the level of dissimilarity between sites. A strength of the methods that take into account species abundance when calculating the dissimilarity between sites is that they do not treat common and rare species equally, which leads to more reliable dissimilarity estimates. Thus, it might of interest of the instructor to use abundance-based dissimilarity metrics when data on the species abundance are available.

Among the methods that have been proposed to calculate abundance-based dissimilarity metrics the Bray-Curtis index and the Chao index are among the most commonly used metrics. The Bray-Curtis index is a variation of the Sorensen index that takes species abundance into account but it might have a poor performance if the sampling fractions between the compared sites are not similar. The Chao index is a more sophisticated probabilistic-based approach that takes into account the effects of unseen shared species when estimating dissimilarity between sites. This approach provides reliable dissimilarity estimates and has been extensively used in ecological research since its proposal.

Questions and answers

Here we provide a rubric of potential answers to each student activity discussed above.

Part 1: Activity 3

Alpha Diversity

1. Which site is the most diverse and the least diverse? What could be driving this?

A: The most diverse site is SRER with 16 species and the least diverse sites are BARR and DELA with only one sampled species each. SRER has the most species likely because it is found in one of the lowest latitudes in the U.S. whereas BARR is found at one of the highest latitudes. Although DELA is in Alabama, it could have fewer species sampled that year because it was sampled less often than the other sites.

2. Which sites should be targeted for conservation according to the alpha diversity metric?

A: SRER should be preserved because it is the most diverse site in terms of alpha diversity.

3. Is there any apparent relationship between site alpha diversity and latitude? Why do you think that happened? Potential reference: https://www.neonscience.org/field-sites/about-field-sites

A: There is only a very weak relationship between alpha diversity and latitude. This likely happened because a wider latitudinal extent needs to be sampled to detect latitudinal patterns of species richness. For example, in Figure 3 we can see that the species richness pattern in the U.S. is relatively constant across the geographic space, indicating a low variability in species richness. However, if we sampled more points at lower latitudes (e.g. Central America) and also more points at higher latitudes (e.g. Alaska) a striking latitudinal pattern in species richness would be detected.

Beta Diversity

1. Which sites are the most similar and the most different to each other?

A: BART and HARV are the most similar sites with a beta diversity of 0.4 while BART and BLAN are the most different sites with a beta diversity of 0.7.

2. *Blarina brevicauda* is a species found in all three sites whereas *Zapus hudsonius* is found only

in the BLAN site. How do you think these species affected the calculation of beta diversity between these sites?

A: *Blarina brevicauda* is contributing to all sites being more similar to each other whereas *Zapus hudsonius* is contributing to BLAN site being more different than the other sites.

3. Are sites that are spatially closer more similar to themselves? What do you think is causing this?

A: Yes, sites that are closer to each other are more similar than sites that are distant from each other. This probably occurs because of spatial autocorrelation, where things that are closer to each other tend to be more similar to each other than things that are far from each other. For example, the climatic conditions in BART and HARV are more similar to each other than to BLAN, which allow a similar set of species to occupy those 2 sites.

4. **Group question:** Suppose you have a limited budget and your task is to preserve the most unique NEON site out of all sites considered in this activity, which site would you preserve and why?

A: If we don't care about number of species and only in level of uniqueness of a site, we should preserve BLAN because it is the site that has the most species that only occurs in it.

Part 2: Activity 3

Alpha Diversity

1. Which site has the most stable diversity value across time? The most varied?

A: The most stable site was BART as the species richness remained relatively constant across the years. MOAB was the least stable site as its species richness varied widely across the sampled years.

2. Do sites tend to be increasing, decreasing, or remaining the same in diversity across time?

A: The temporal patterns in species richness varied across the considered sites. In general, BART and JORN tended to have relatively stable species richness over time, but species richness tended to decrease over time for MOAB.

3. What might be causing species losses and gains across sites?

A: Several factors can lead to species losses and gains over time. Change in environmental conditions, for example, can make a site suitable or unsuitable for species over time and can lead to species gain and losses. Similarly, changes in resource availability over time can favor some species while others might be negatively affected by it. Competition for resources can also change community composition over time such that some species can outcompete others for resources, which can lead some species to be excluded from a community. Finally, stochasticity (i.e., random events that affect population and community dynamics) can also lead to species gains and losses over time.

Beta Diversity

1. Which site changes the most over time?

A: The site that changed the most over time was JORN, where the beta diversity between the years 2016 and 2019 was 0.73, indicating a drastic change in community composition over time.

2. What does it mean when a sites' beta diversity is changing over time but its alpha diversity is remaining the same?

A: This means that the species are not simply being lost in a site, but that species are being replaced over time.

3. In general, was beta diversity higher comparing 2016 to 2019 than comparing it to 2017 and 2018? What if we waited for 20 more years and calculated the beta diversity between 2016 and 2042, what do you think would happen?

A: Yes, beta diversity was considerably higher for all sites when comparing 2016 to 2019 than when comparing it to 2017 or 2018. If we compared 2016 to 2042 it is likely that beta diversity would be extremely high, given that so much time would have passed and several species would have colonized or would have been extinct from a site. Thus, the longer the time period between two sampling occasions in a specific site, the more different the site will be when comparing the species composition between the two sampling events.

R files

We provide to the instructor two files, Getting_Mammals.R and MammalsData.RData that can be used to programmatically access the NEON small mammal data. R is an open-source software environment that is widely used for data management and statistical analyses. The file MammalsData.RData contains all the raw small mammal data from NEON and file Getting_Mammals.R is a script that has all the steps necessary to manage the raw data in MammalsData.RData and generate all the data in the format that is available in this teaching module. Before running the Getting_Mammals.R script, the following command should be run: *install.packages(c('plyr', 'dplyr', 'stringr', 'ISOweek', 'stringi', 'geosphere'))* to install all the packages that are used in the R script to manage the raw data in MammalsData.RData.

SUPPLEMENTARY MATERIALS

Below is a brief reference guide for students with varying degrees of experience using Excel. The guide outlines the steps within Excel that are necessary for students to complete the assignments. Not all aspects of this guide may be necessary for students and certain aspects may be excluded or serve as review. This guide was written for Excel version 2211 using an English keyboard in Windows, if certain functions are not operating as intended then it may be useful to look up if the functions have changed since this version/have difference across OS systems or keyboards. As most of the functions used are rather basic, we do not anticipate substantial differences across versions or systems.

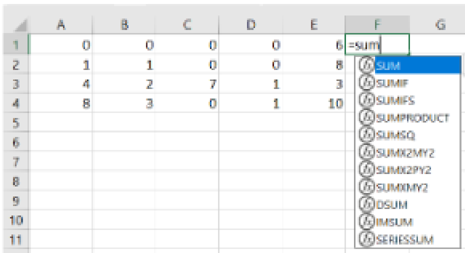
Excel Guide

Below are descriptions and examples of how to use two useful functions in Excel (**SUM** and **IF**). While the exercises may be completed in different ways, implementing these functions will be helpful in efficiently completing the exercises.

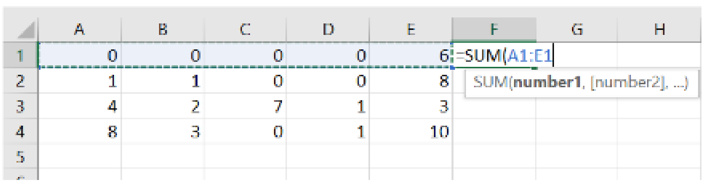
TIEE

Teaching Issues and Experiments in Ecology - Volume 19, October 2023

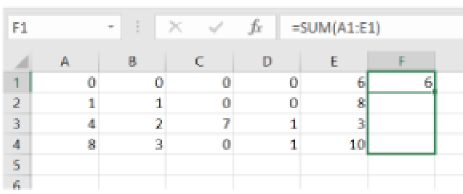
The **SUM function** takes the total (sum) across highlighted cells (cell 1 + cell 2 + cell 3…), this can be done across rows and columns.



To get the sum of a row of cells, begin by typing =sum in an empty cell (the total will appear in this cell) and then select the SUM

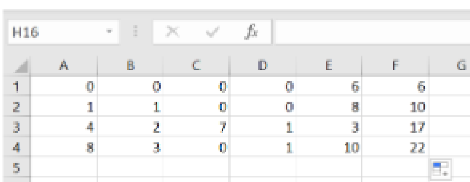


Next, the function will allow you to select the cells you want the total from =SUM(First cell:last cell), seen in blue on left. These can be selected via click and drag or by typing the cell names into the parentheses



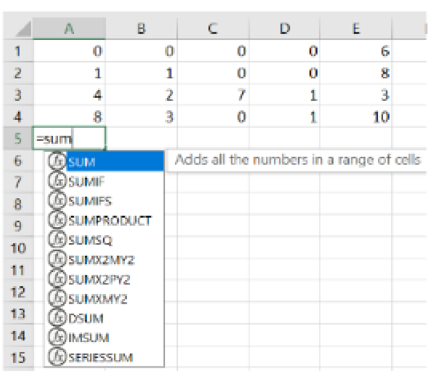
Once the parentheses are closed (either by typing or clicking enter after click & drag), the total should appear in the cell.

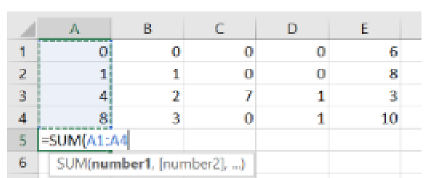
The same structure of the function can then be extended by clicking on the bottom right corner of the cell and dragging down.

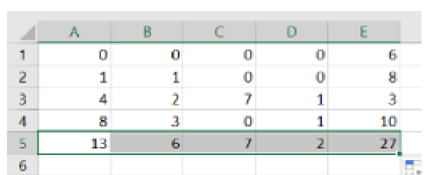


This replicates the above procedure for the rows below. Be careful when using this method though and make sure that each sum is for the correct cells. This can be done by clicking on the total in each cell and it will highlight the cells it is summing across and display the function =SUM(cellx:celly)

The same procedure as above can also be done for a column of cells (see below)







TIEE

Teaching Issues and Experiments in Ecology - Volume 19, October 2023

The **IF function** can enter specific values for different circumstances. For example, the table below has a range of numbers (0-10) but if we only want to know whether the cells contain a zero or another integer, we can use an IF statement to create another table reflecting whether cells contain a zero or another integer.

	A	B	C	D	E	F	G
1	0	0	0	0	6		=if
2	1	1	0	0	8		
3	4	2	7	1	3		
4	8	3	0	1	10		

The process starts the same as sum, begin by typing =if in an empty cell (the total outcome will appear in this cell) and then select the IF function from the menu.

	A	B	C	D	E	F	G
1	0	0	0	0	6		=if(A1>0,1,0)
2	1	1	0	0	8		
3	4	2	7	1	3		
4	8	3	0	1	10		

Now we enter the conditions we want for a given cell. First we select a cell and the rule (in this ex. if the value is greater than 0). Next, we tell the function what to enter into the empty cell if this is true and what to enter if it is false.

IF function general setup: =IF(**condition**, **what to do if true**, **what to do if false**).

	A	B	C	D	E	F	G
1	0	0	0	0	6		0
2	1	1	0	0	8		
3	4	2	7	1	3		
4	8	3	0	1	10		
5							

As with SUM, the appropriate outcome should appear in the cell once the function is completed, and the function can then be extended to other rows/columns.

Now we have all the original cells represented as 0 or 1 in the new cells depending on the condition we set (if the cell value > 0).

	A	B	C	D	E	F	G	H	I	J	K
1	0	0	0	0	6		0	0	0	0	1
2	1	1	0	0	8		1	1	0	0	1
3	4	2	7	1	3		1	1	1	1	1
4	8	3	0	1	10		1	1	0	1	1