**Title**: Investigating the effects of urbanization on bird biodiversity: Testing three biodiversity hypotheses using citizen science data

**The Ecological Question:**  How does urbanization affect bird biodiversity?

**Ecological Content:** biodiversity, Shannon’s biodiversity index, land use, urbanization, productivity hypothesis, ecosystem stress hypothesis, intermediate disturbance hypothesis

**What Students Do:**

-Read excerpts of primary literature in order to define three biodiversity hypotheses

 -Generate testable predictions for three biodiversity hypotheses and draw graphs that would support each hypothesis/ prediction

 -Analyze a large database of citizen science collected bird counts for 14 different geographic locations at 2 different time points 10 years apart to generate 26 Shannon’s biodiversity indices

 -Relate biodiversity with land use

 -Relate changes in biodiversity with changes in land use

 -Perform basic calculations from a large data set, generate graphs, and use appropriate statistical analyses to generate a p-value

 -Use statistics to evaluate predictions and hypotheses

**Student-active Approaches:** scientific method, hypothesis array, hypothesis testing, graphing, data analysis

**Skills:**

-Making testable predictions based on hypotheses

-Drawing prediction graphs

-Using spreadsheets

-Making graphs

-Evaluating hypotheses using data, graphing, and basic statistics

**Assessable Outcomes:** The deliverables are: graphs generated by students (figure), a figure legend for the graph with statistics, and a 250-word abstract for their study. The basics of the scientific method, generating testable predictions and evaluating hypotheses using data, are at the core of the student assessment.

**Source**:

* Sullivan, B.L., C.L. Wood, M.J. Iliff, R.E. Bonney, D. Fink, and S. Kelling. 2009. eBird: a citizen-based bird observation network in the biological sciences. Biological Conservation 142: 2282-2292.
* eBird Basic Dataset. Version: EBD\_relMay-2013. Cornell Lab of Ornithology, Ithaca, New York. May 2013.

## Kovacs, J.L**.**, G.N.H. Franklin, N. Flowers & E. Gaillard. In prep for Biology Open. Long-term effects of land use changes on bird biodiversity in the southeast United States

**Author:** Jennifer L Kovacs

**Institution:** Spelman College, Atlanta, GA

**Acknowledgements:** This project came about as part of my Research in Urban Ecology course at Spelman. I’d like to thank the undergraduate students and researchers who helped with the ideas, data collection, and data analysis for this project.

**Relevant Cover Image:** <include an image which complements or in some way helps to illustrate the data set; send as separate image file>

**2. Overview**

Urbanization can affect biodiversity in a variety of ways, including habitat loss and fragmentation and the introduction of non-native species. In this module, students will use bird counts (eBird) and land use data (LandSat) to test three hypotheses about urbanization and its effect on biodiversity. Students first generate testable predictions for the productivity hypothesis (Luck et al. 2004), the ecosystem stress hypothesis (Rapport et al 1985), and the intermediate disturbance hypothesis (Connell 1978). Using Google Sheets, students then calculate Shannon’s Diversity Indices for 13 different locations at two different time points ten years apart. By combining these data with measures of land use for each location at each time point, students are able to graphically and statistically evaluate the three assigned biodiversity hypotheses. This module is presented as a guided inquiry module, though the provided dataset could serve as the basis for more open-ended inquiry and extended projects.

**3. The Data Sets**

*Submit as a separate email attachment:*

Two Excel files of your data: one for faculty use that will include figures, and the other for student use, which do not include figures. Make sure that there is at least one data set that students can work up in a single lab session; additional data for more extended work or advanced students can also be included.

*On this form:*

Florida\_biodiversity\_students.xls

Florida\_biodiversity\_instructor.xls

*Jennifer Kovacs gives permission for this data set to be posted and distributed on the TIEE website. The bird count data has been modified from data provided by project eBird and the terms of use are also included with these files.*

*eBird Basic Dataset. Version: EBD\_relAug-2014. Cornell Lab of Ornithology, Ithaca, New York. Aug 2014*

**4. Student Instructions: (starts on next page)**

# How does urbanization affect biodiversity?

Stop for a minute and picture yourself in the middle of a city. When you look around what do you see? Any plants? Any animals? What living things are there around you? How many different species do you think you’d find?

Now picture a place about 10 or 15 miles outside the city. What’s different? Are there more plants? More animals? Do you think there are more species here? Are they different from the ones you imagined in the city?

What do you think that spot 10 to 15 miles outside the city looked like 20 years ago? Were there more plants and animals? Were they the same or different?

Humans can rapidly change the landscape around them. We disturb ecosystems through processes such as urbanization and farming. By building roads, factories, farm fields, parking lots, and houses, as well as going about our everyday modern human lives, we fragment and destroy natural habitats, introduce invasive species, and remove native species. You can easily imagine how some of these changes in land use (e.g. from undisturbed forest to paved road) can affect the plants and animals native to those areas.

Think back to those spaces you imagined earlier, the city, the space outside the city, and that same space 20 years ago. You imagined the types of plants and animals there. You probably also imagined they were different in the three different places.

How could you quantify the differences? Well, for one you could count the number of species in each location. This would provide us with measurements of ***species richness***. You could also count how many individuals of each species there were in each location. This would give us an idea of ***species evenness*** in each space. Together these would give us measures of the ***biodiversity*** for each location. We could then compare them. This would give us a much better idea of how land use and changes in land use over time affect the plants and animals in a space.

Based on your three imaginary locations, you might predict that as urbanization increases, biodiversity decreases. You probably imagined fewer animals and plants in your city than 20 miles away. But, it’s not always clear how changes in land use, and urbanization in particular, affect biodiversity.

**PART 1: HYPOTHESES**

Below is an excerpt from Lepczyk et al. 2008 Conservation Biology outlining three different hypotheses about how urbanization affects biodiversity. We will be testing these three hypotheses throughout this exercise.

Read the excerpt below and finish the three sentences below. Remember to be clear and concise in your answers.

Increased human domination of the Earth’s ecosystems (Vitousek et al. 1997) and intensifying human land use (Foley et al. 2005) raise the question of how biodiversity will be affected by these factors. Currently, the observed patterns of association between humans and biodiversity are described by 3 hypotheses (Fig. 1). First, the productivity hypothesis states that more productive systems support more species and more people—a covariation manifesting as a positive correlation between species richness and human population. The positive correlation is thought to be due to productivity gradients caused by varying energy availability (Gaston 2005), with more productive landscapes attracting both humans and other species. Empirical evidence that species richness increases with human population density in Africa, Europe, and North America (Balmford et al. 2001; Hawkins et al. 2003; Gaston & Evans 2004; Luck et al. 2004) appears to support this productivity hypothesis. It remains unclear, however, whether this correlation reflects spatial congruence stemming from selection of high productivity sites by people and wildlife or a causal link, whereby human activities (supplemental feeding, irrigation, plantings) elevate resources and support a more diverse biota (Rapport et al. 1985).

Second, the ecosystem-stress hypothesis states that humans are detrimental to species diversity because they remove habitat and resources of most species and thus predicts a negative relationship between species richness and human influence (Rapport et al. 1985). Evidence that bird and anuran diversity decreases with increases in urbanization (Clergeau et al. 1998; Cam et al. 2000; Genet et al., in Press) supports the ecosystem-stress Hypothesis.

Third, the intermediate-disturbance hypothesis (Connell 1978) states that landscapes under moderate levels of human impact have higher habitat and resource diversity compared with pristine or human-dominated landscapes.
This higher resource diversity thus leads to higher species diversity and forms a negative quadratic relationship (i.e., a hump-shaped relationship) between species richness and human influence (McDonnell & Pickett 1990). Such
patterns have been observed in a number of plant and animal taxa (reviewed in McKinney 2002), typically along rural-to-urban gradients (Blair 1996).”

The **PRODUCTIVITY HYPOTHESIS** states:

The **ECOSYSTEM STRESS HYPOTHESIS** states:

The **INTERMEDIATE DISTURBANCE HYPOTHESIS** states:

Are these hypotheses mutually exclusive?

Now that we’ve identified and summarized the three different hypotheses we will be working with for the rest of this exercise, let’s think about what testable predictions we can make about them.

For all three of these hypotheses can we identify a single independent variable?

How about a dependent variable?

**PART 2: PREDICTIONS**

For each of the three hypotheses draw a graph of the relationship you would predict to see between the independent and the dependent variable. Remember to label your axes.

Under the **PRODUCTIVITY HYPOTHESIS,** we’d predict:

Under the **ENVIRONMENTAL STRESS HYPOTHESIS,** we’d predict:

Under the **INTERMEDIATE DISTURBANCE HYPOTHESIS,** we’d predict:

**PART 3: QUANTIFYING BIODIVERSITY**

When we talk about biodiversity we don’t want to just talk about it in relative terms of more or less, we want to quantify it. Scientists have devised several different ways of doing this, though nearly all of them start with counting the number of species in an area (species richness) and the number of individuals in each species (evenness). With those metrics in hand as well as what species are in each location, we can quantify not only the biodiversity in a spot, but we can start thinking about comparing communities or looking at how biodiversity and communities change over time.

For this exercise, we’re going to focus on a single measurement of biodiversity known as the Shannon’s biodiversity Index (also known as the Shannon’s Index or Shannon entropy). Here’s the formula:



Where *pi* is the proportion of individuals belonging to the *i*th species in group *R*

So, let’s talk through an example. You have two sampling locations A & B, and you go to count birds in the two areas.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Location A |  |  |  | Location B |  |  |  |
| Robins | 4 |  |  |  | 20 |  |  |  |
| Blue Jays | 8 |  |  |  | 2 |  |  |  |
| Cardinals | 12 |  |  |  | 2 |  |  |  |
| Crows | 3 |  |  |  | 2 |  |  |  |
| Hawks | 0 |  |  |  | 1 |  |  |  |
|  |  |  |  |  |  |  |  |  |

You saw a total of 27 birds at Location A and another 27 at Location B. Now, you need to find the proportion for each bird species at each location. So, for example for the Robins at Location A, divide 4 by 27.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Location A | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  | Location B | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  |
| Robins | 4 | 4/27 = 0.15 |  |  | 20 |  |  |  |
| Blue Jays | 8 | 0.30 |  |  | 2 |  |  |  |
| Cardinals | 12 | 0.44 |  |  | 2 |  |  |  |
| Crows | 3 | 0.11 |  |  | 2 |  |  |  |
| Hawks | 0 | 0 |  |  | 1 |  |  |  |
|  |  |  |  |  |  |  |  |  |

Now, find the natural log (ln) for each proportion and multiply that by the proportion.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Location A | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  | Location B | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  |
| Robins | 4 | 4/27 = 0.15 | -1.91 | -0.28 | 20 |  |  |  |
| Blue Jays | 8 | 0.30 | -1.22 | -0.36 | 2 |  |  |  |
| Cardinals | 12 | 0.44 | -0.81 | -0.36 | 2 |  |  |  |
| Crows | 3 | 0.11 | -2.20 | -0.24 | 2 |  |  |  |
| Hawks | 0 | 0 |  |  | 1 |  |  |  |
|  |  |  | - Sum = Shannon’s |  |  |  |  |  |

Now sum those numbers for each location. This should be a negative number, now find the opposite of that. That’s it.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Location A | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  | Location B | p*i* |
| Robins | 4 | 4/27 = 0.15 | -1.91 | -0.28 | 20 |  |
| Blue Jays | 8 | 0.30 | -1.22 | -0.36 | 2 |  |
| Cardinals | 12 | 0.44 | -0.81 | -0.36 | 2 |  |
| Crows | 3 | 0.11 | -2.20 | -0.24 | 2 |  |
| Hawks | 0 | 0 |  |  | 1 |  |
|  |  |  | - Sum = Shannon’s | 1.25 |  |  |

Now do that in a spreadsheet using formulas for Location B.

We can now compare the biodiversity between the two sampling locations. Which has the higher Shannon’s Biodiversity Index (this is a unitless measure) and therefore the most biodiversity? Look at the sampling can you see why that would be the case intuitively?

**PART 4**: **MAKING CITIZEN SCIENCE WORK FOR YOU**

So, we’ve thought about how urbanization might affect biodiversity, and we’ve identified some hypotheses, identified independent and dependent variables, made testable predictions for each hypothesis, and even identified a way to quantify the concept of biodiversity.

Now it’s time to do some data collection. We now know that in order to quantify biodiversity we have to go out and count the number of species in an area and the number of individuals in each species. As you can imagine, it takes a lot of effort and time to get large enough sample sizes to start to really meaningfully address issues around biodiversity.

Lucky for us we live in the digital age of citizen science. Citizen science is science done by the general public often in collaboration with professional scientists. Several exemplary and longest running citizen science projects are those run by the Cornell Lab of Orinthology (http://www.birds.cornell.edu). These include Project FeederWatch, The Great Backyard Bird Count, and eBird. What all of these projects have in common is that they recruit and rely upon amateur birdwatchers to record and report the birds they see. We will be using some their data for this exercise. eBird is a giant database of birdwatcher’s checklists from all over the world. Participants submit their lists with all of the relevant data (location, type of birdwatching activity-- casual observation, survey, etc.), and the database is updated daily.

There’s a lot of data in eBird. A lot. More than we could easily download and work with. What you will be working with is a very small (relatively) and curated subset of data for 13 locations in Florida. We’ve chosen these locations because they have bird count records from the same location, in the same season (in the autumn), ten years apart. Also, the bird counts were done in the same way and often by the same person. We’re going to use this subset of data to get measurements of biodiversity.

 So, we have biodiversity. What about urbanization? How can we get information about land use? Never fear, we have some of that. For this exercise, we’re going to use satellite data collected by the LandSat program (<https://landsat.usgs.gov/>). LandSat is the longest-running satellite collection of land use data. It is jointly run by NASA and the USGS. We will be using data collected from the compiled satellite images from the 2001 and 2011 and made publicly available through the Multi-Resolution Land Characteristics Consortium (<https://www.mrlc.gov/>). We will discuss the land use data for this exercise a bit more in the next section.

**PART 5: TIME FOR SOME DATA**

Open up the excel workbook entitled “FL\_bird\_counts\_land\_use” and look at sheets “1:locations”, “2:rawbirdcounts”, and “3:landuse”.

What information are you looking at in each of the sheets? How are they connected?



In sheet “1:locations” we are looking at a rundown of the thirteen locations in Florida used in this study (see map below). Included are the county, latitude, and longitude points for each location. What do the timepoints correspond to? HINT: It has to do with the bird count and land use data sheets.

Map of 14 Florida Locations

Open up sheet “2:rawbirdcounts”. What is the data contained in this sheet? How could you use this data to answer questions about biodiversity?

Open up sheet “3:landuse”. The data in this sheet comes from an analysis of images from the LandSat program described above. Below is an example of one of the images that was analyzed for land use for one of our 13 sampling sites.

Each pixel represents a 3x3 m square of space and has been categorized into one of 20 land use categories (see the legend below). We have used a software program called QGIS to quantify the amount of developed land (combined categories 21,22,23, & 24) in a 42.25 km2 area around each of the 13 locations in 2001 and 2011.





How could you use this data to address the three biodiversity hypotheses?

**PART 6:** **TIME TO TEST THOSE HYPOTHESES**

Go back to those prediction graphs you made before.

What did you have on your x-axis?

What did you have on your y-axis?

It may be that we don’t have the exact independent and dependent variables that you labelled your graphs with earlier , but we do have data that could be used to address those three hypotheses.

Now that you’ve taken a look at the data we *do* have, what do you need to do to take our data and make a graph like the ones in your predictions?

What do we have to calculate? What’s your new independent and dependent variables? Note: there’s no right or wrong answer here. There are multiple ways to do this analysis. Think about what makes the most sense to you. Can you explain and justify your thinking? Then you’re doing it right!

Label the axes on the graph below.

What kind of graph are you going to make?

What kind of statistic would you use for that kind of graph?

**PART 7: TIME TO MAKE THE GRAPHS!**

Now that you’ve decided what your graph is going to look like, it’s time to make it!

You’ll probably want to start a new sheet in the Google spreadsheet to put all of your analyzed data in and to build your graph (you could also do this in a new spreadsheet).

Remember to label your columns and to be careful when copying and pasting.

Once you’ve created your graph, don’t forget to run a quick statistical analysis to see if there is a significant relationship between your two variables.

***Your assignment*** is to write a 250-word science style abstract that outlines what you did in this project. This should include a few sentences of background, your hypothesis/es and prediction/s, a brief bit about your methods, your results, and a sentence of conclusion.

Also, turn-in your graph. Remember to label all of your axes! You’ll also need a title and figure legend for you graph, just like one you would see in a journal.

**FURTHER EXPLORATION:**

Now that you’ve finished your first analysis and see what the data has revealed to you, take a minute and think about what other factors might affect biodiversity?

What other questions could you ask? What about the species composition of the different locations? What other metrics could we use?

Is there other data we could look at or include?

**5. Faculty Notes:**

The goal of this module is to have students generate testable predictions for several mutually exclusive hypotheses and to analyze a real-world dataset to see whether the data supports any of these predictions.

Students should work in small groups (~2 students/ group) to encourage discussion of answers, and the case could be done in an interrupted case study style. Have student work on a section as a small group and then discuss as a whole class. In a smaller class, student groups could work straight through with the instructor checking each group’s progress as they go and then have each group report out at the very end.

PART 1:

To do that, we first have students read an excerpt from the primary literature to introduce them to three biodiversity hypotheses, the productivity, ecosystem stress, and intermediate disturbance hypotheses.

These three hypotheses all make different predictions about how urbanization should affect the biodiversity in an ecosystem.

Under the productivity hypotheses biodiversity increases with urbanization due to increased access to food, shelter, and other resources that may be more scarce in more natural habitats.

Under the ecosystem stress hypotheses we predict a decrease in biodiversity with an increase in urbanization due to a reduction in resources that is associated with increased human disturbance.

The third hypothesis to be discussed in this module is the intermediate disturbance hypothesis which states that biodiversity is lowest at high levels of disturbance and low levels of disturbance and highest somewhere in the middle (like Goldilocks).

Once students understand the very basics of these three hypotheses and that they are really pretty much mutually exclusive, have them start thinking about what the independent and dependent variables are that are shared by all three hypotheses.

There really are no absolute right answers here. Ideally, urbanization (or some measure of it) would be named the independent variable and biodiversity the dependent. However, as you will see later these are rather vague variables, but for now that’s OK we’ll get more specific about what we’re defining as “urbanization” and “biodiversity” later.

PART 2:

Have students take the time to label the graphs and draw their predictions. I have found that if students take the time to do this step here, data analysis and interpretation later in the module is much easier for them.

**PRODUCTIVITY HYPOTHESIS,** we’d predict:

Biodiversity

Urbanization

**ENVIRONMENTAL STRESS HYPOTHESIS,** we’d predict:

Biodiversity

Urbanization

**INTERMEDIATE DISTURBANCE HYPOTHESIS,** we’d predict:

Biodiversity

Urbanization

Once students have these predictions in hand we can start thinking about how we’re going to define our variables.

PART 3:

For this module we’re going to quantify biodiversity using the Shannon’s Biodiversity index. There are many ways to define biodiveristy, but Shannon’s is relatively easy to understand and compute and is included in most intro bio textbooks, so that’s what we’ll be using in this module.

Having students walk through the example on paper and then using excel (or google sheets) is very important at this step. When we get to data analysis, they’re going to need to be able to calculate Shannon’s from a much larger data set. Having the formulas and layout already done for a simpler example will make that step much easier.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Location A | p*i* | ln(p*i*) | p*i*\* ln(p*i*)  | Location B | p*i* | ln(p*i*) | p*i*\* ln(p |
| Robins | 4 | 4/27 = 0.15 | -1.91 | -0.28 | 20 | 0.74 | -0.30 | -0.22 |
| Blue Jays | 8 | 0.30 | -1.22 | -0.36 | 2 | 0.07 | -2.60 | -0.19 |
| Cardinals | 12 | 0.44 | -0.81 | -0.36 | 2 | 0.07 | -2.60 | -0.19 |
| Crows | 3 | 0.11 | -2.20 | -0.24 | 2 | 0.07 | -2.60 | -0.19 |
| Hawks | 0 | 0 |  |  | 1 | 0.04 | -3.30 | -0.12 |
|  |  |  | - Sum = Shannon’s | 1.25 |  |  |  | 0.92 |

So, even though there are more species in location B (5) than in location A (4), the biodiversity is high in location A due to there being more species evenness in location A (i.e. the species are more evenly represented in sample A, while they are more skewed towards robins in sample B).

PART 5:

Note: I have highlighted any differences between the teacher’s excel sheets and the student’s excel sheets in yellow. Generally these are calculations which the students could do.

Here we start getting much more into the raw data itself. The excel sheet provided has three sheets in the workbook. The first sheet “1:locations” provides a basic description of the sites identified and used in the study. We have simplified the data here a bit by only presenting data from only one day for each of the two time periods.

Sheet 2 “2:rawbirdcounts” is a large sheet. It has a simplified and modified version of the raw data counts from eBird from two time points (10 years apart) from the Florida locations. In the instructor version, I have calculated Shannon’s for these sites in sheets 4 & 5.

In sheet 3 “3:landuse” we provide students with the land use numbers that we calculated in 2011 and 2001 using QGIS. In the teacher’s excel sheet you will see that we’ve done some extra calculations looking at change in development (both in % and in square kilometers). While again there is no one “right” way to analyze this dataset, some students may want to look at how change in urbanization levels over time affects biodiversity over time. If they were to do that, they would need those calculations. I have also added in the Shannon’s from sheet 2 to this datasheet in order to make graphs. Graphs are shown in sheet 6.

PART 6 & 7:

Here’s where the students start to use the data themselves. We’re now going to revisit those prediction graphs we made earlier. Now that we have specific variables to work with, students will need to decide what specific variable they are going to use for the more general variables like “biodiversity” or “urbanization” that we were using earlier.

Really push your students to be specific here. What exactly are they going to put on the X and the Y in their graph. By doing this they will be defining what calculations they will need to do to make the graphs, so it is really important that they are clear on that and don’t just start calculating things without a plan.

There are several ways to analyze the data that would address the 3 hypotheses that we are testing in this module. I can think of at least 3 different ways to analyze the data that would be appropriate, and I’m sure that students can come up with a couple of others too!

One way would be just look at landuse (X) and biodiversity (Y) for all the data all together (regardless of year) and for our purposes here, that would be fine.

Another way would be to break it up by year (do 2001 and 2011) separately (which is a bit more accurate).

Another way would be to look to see if changes in urbanization correlate with changes in biodiversity.

For students, all of these are legitimate ways to analyze the data. All would require a line graph to be made and a correlation tested (which can be done in excel or google sheets).

Though, I am sure that some students will come up with other ways to analyze the data and that is great! Just make sure they can explain it!

Once students have identified what they are going to graph, it’s time for them to do the calculations themselves. Most groups should be calculating Shannon’s for the sites. Remember to remind them that they have already made a spreadsheet with formulas that will help them make this much bigger sheet. Graph making can be really challenging for some students, so make sure that there’s time in class to have students work through how to make the graph. This really is the most important step!