ENS 495 Independent Data Analysis Assignment 2017

**Today:**

1. Adapt script template to your work **and upload draft to D2L** (see below for further details)
2. Identify and prepare your final dataset **and upload it to D2L**
3. Finish working on a draft of your data dictionary **and upload it to D2L**
4. Make exploratory graphs of your data.  **Upload at least one to D2L** (see below also)
5. Explore how to make refinements to your graph, such as colors, labels, etc. Your figure uploaded to D2L should have **at least 1 refinement**.
6. Attempt to adapt code in the script to build a model using lm(). The script you upload to D2L should contain this attempt; it doesn’t have to work, but you have to try.

Everyone should fall into one of the following categories:

1. **You have your own dataset** 
   1. …and you know how to set it up and work with it. Awesome!
      1. You can proceed with making a data dictionary, making graphs and model. Awesome!
   2. …and you aren’t sure what to do with it yet. **We’ll finalize that today!**
      1. ~~We’ll work on reshaping the data, though this might have to be finished out of class.~~
      2. **Today you can probably start working on identifying the variables most important to your analysis and drafting a data dictionary , etc. We’ll try to make as much progress as possible.**
2. **I am providing you with a dataset from other researchers or my own work**
   1. ~~I’ll help you today figure out how to proceed. This might require finalizing the question, subsetting that data, isolating columns, or in a few cases collecting a bit more data.~~
   2. **Today you can probably start working on identifying the variables most important to your analysis, drafting a data dictionary, and beginning exploratory graphs**
   3. **You should be able to build the basic structure of a model using lm, even if it doesn’t work**
3. **You are working with Breeding Bird Survey (BBS) or other USGS Data**
   1. **BBS data**
      1. If you want to work with BBS data you should chose the form of the data you want to work on.
         1. ~~1st, whether you want to do a~~ **~~species-level analysis~~** ~~(~~**~~population~~**~~) or~~ **~~community-level analysis~~** ~~(species richness).~~
         2. 2nd, whether you want to look at how things change with variation in **landcover** (as we have done in class) OR how things change over **time**.
      2. **SPECIES-LEVEL CHANGES**
         1. If you haven’t already, **select a species**.
            1. Songbird species will work best; raptors, ducks, waterfowl etc not so much.
            2. An excellent website is <https://www.allaboutbirds.org/>
            3. You need to figure out the four-letter species code (eg, the Louisiana waterthrush is LOWA, the Robin is AMRO for American Robin, the Cardinal is NOCA for Northern Cardinal. Normally is the 1st two letters of the first and 2nd part of the name, but not always)
            4. You need to figure out the numeric code for the species.
            5. You can usually figure out the four-letter code by googling
            6. There is a dataframe in the *wildlifeR* package that contains all of these four-letter codes and numeric codes. You can find them using which() statements; I am also working on a function to access them easily.

Note that the numeric species code IS NOT the same as the ROW NUMBER where the species occurs in the AOU species code database.

* + - 1. If you will be examining how the species varies with **LANDCOVER**, you will be following the general workflow we have done in class for the Scarlet Tanager (SCTA) and today for the Pileated Woodpecker.
         1. For your species, you need to pick out two types of landcover that you think impact the abundance or you species. This should be informed by the ecology of the species. See the sheet of USGS landcover codes. For the SCTA example, we looked at total forest landcover, which was the sum of 3 separate USGS landcover classes. For this assignment, I could look at 2 of these 3 landcover classes.
         2. Create a dataframe with the BBS data for you species and the 3 landcover variables.
         3. Create a data dictionary for your species.
      2. If you want to study the species over **TIME**, you will use the function **sample\_BBS\_routes()** to select a subset of BBS routes for analysis.
         1. You should select as species and create a datasubset. This is given to **sample\_BBS\_routes()** as “dat = your.data.subset”
         2. You can choose the years you want to consider. It will probably work well to choose three years spaced 5 to 10 years apart, each 1995, 2000, 2005, or 1990,2000, and 2010
         3. sample\_BBS\_routes() will randomly sample and approximately equal number of routes from each of the years you specify, and guarantee that a route only shows up in one year. One of your tasks will be to figure out why it does this!
         4. You can save the data using write.csv()
         5. Create a data dictionary for the data.
    1. **COMMUNITY-LEVEL CHANGES**
       1. I have not finalized the setup of the community-level data. If you are interest in working with community data let me know and I will finalize it ASAP.
  1. **Woodcock**
     1. Woodcock data are collected by the USGS in a manner similar to the BBS routes but I do not have landcover data available for these particular routes. Analysis of woodcock data will therefore look at changes over time similar to what is outlined above for the BBS data
     2. Woodcock call surveys often locate few birds, and the abundance of birds varies a lot between states. PA does not actually yield data that is very good for this project, but Michigan does.
     3. There is an example dataframe from Michigan that I have generated. You can use this as the basis to begin your data dictionary.
     4. I will provide a finalized version of the data for you by the next lab session that you can use to explore the data. I will hopefully be able to give you the option to select data from other states, but need to check things out first.
  2. **Mourning Doves**
     1. Mourning dove data can be obtained from the BBS\_PA dataset.