Data Tutorial Outline:

I. Introduction Blurb: In order to address research questions about mosquito populations in the United States, we need the requisite skills and data to carry out analysis. This tutorial will focus on obtaining data from the NEON data portal and the Global Historical Climatology Network, provide methods for composing data frames with relevant information, and a preliminary exploration of various research questions related to mosquito data.

II. Lesson Summary:

1. Learning Objectives:
   1. Download mosquito trapping, identification, and sorting information from the NEON data portal
   2. Download precipitation and temperature data from Global Historical Climatology Network
   3. Merge data frames to create one unified data frame with relevant variables to address research questions
   4. Subset data by year
   5. Use ggplot2 in R to create visualizations of data trends and maps
2. Things You’ll Need to Complete this Lesson:
   1. Some previous R experience
3. Libraries to Install:
   1. Tidyverse
   2. Plyr
   3. Mosaic
   4. Lme4 (maybe)
   5. metScanR

III. Research Question(s)/ Motivation for Tutorial:

To many, mosquitos are nothing more than a seasonal annoyance. However, mosquitos are important for their role in the transmission of human diseases. For example, the species *Aedes albopictus* is notable in North America for its ability to spread the West Nile Virus and its association with human inhabited urbanized environments (Rochlin et al. 2013). As the environment changes in response to global warming, it is likely that the ranges and habits of mosquitos like *Aedes albopictus* will also change. Differences in mosquito habits over time could put an increased number of people at risk for contracting mosquito-borne diseases, so it is important to investigate the various facets of mosquitos interacting with their environments.

This tutorial will provide a preliminary look at three different avenues of investigation. First, we will examine whether there is a relationship between mosquito species richness and latitude in the United States. The latitudinal diversity gradient (LDG) has been explored in other species such as shore fishes and African birds, but little work has been done on LDGs for mosquitos. Findings seem to suggest that factors such as the mid-domain effect and average temperature were useful predictors for LDGs. A study on African birds also found that examining species richness in latitudinal bands could be misleading on a continental scale (Jetz and Rahbek 2001, Mora et al. 2005). In this tutorial we will transform NEON data to construct preliminary models of mosquito species richness in relation to latitude. A second aspect of mosquito data under consideration is the Julian date of first occurrence for certain species of mosquitos. In particular in this tutorial we will be examining the species *Aedes albopictus* and *Culex tarsalis* because of their abilities to transmit diseases and their relative presence in the NEON data. Part of our analysis will focus on whether these two species of mosquitos are invading new areas of the United States and visualizing our results. Our third vignette will then focus on the abundance of mosquitos over time. As the environment changes in response to rises in temperature or changes in land use, it is possible that the number of mosquitos in these areas could also change. So to investigate this possible phenomenon, we will be creating preliminary models and visualizations.

1. Background and Context of why studying mosquitos is important
   1. Inclusion of some of literature review findings.
2. Introduction of Vignette Topics
   1. Richness vs. gradient
   2. Julian date of first occurrence vs. lat
   3. Mapping occurrence over U.S
   4. TBD

IV. Obtaining the Data:

1. Obtaining Precip and Temperature Data (More detail needed)
   1. Install the package metScanR
   2. Use metScanR to look for weather stations where:
      1. Weather station is within 150km of site of study
      2. Active precip and temp data within last ten years
      3. Within 300 m of the site’s elevation
2. Obtaining NEON Mosquito Data
   1. TBD (I am unsure of how the data will be obtained via the online data portal)

V. Cleaning the Data Frames:

A. Read in data frames, libraries and requisite functions:

a. Review the read.csv and source function

B. Obtain latitude and longitude information

a. Creation of a for loop to scrape lat and long data from APIs

i. Explanation of getNEONlocation function

c. Append lat lon vectors to unique plot ids to create a new df

d. Replace all 1s in lat lon data with NAs

1. Merging dataframes
   1. Explanation and documentation of the merge function
   2. Merging the lat lon df with the trap data frame by plotID
   3. Creation of secondary lat lon variable to use lat lon data in trap dataframe
      1. Explanation of why lat lon data in different areas
      2. Review of ifelse statements and variable creation
   4. Create new data frame containing only unique plot ids and updated lat lon variables
   5. Merge with ID dataframe
   6. Create date and year variables as well as site id variable
   7. Merge with precip and temp data
2. Creation of Time difference variable
3. Create subsets of id data frame based on year

VI. Vignette 1: Richness vs. Gradient

1. Using ddply to create richness dataframe
   1. Introduction to ddply
2. Merge with unique trap data frame to get lat lon information
3. Create subsets based on year
4. Visualizations
   1. How to use ggplot2 for creating scatterplots
   2. TBD relevant visualizations

VII. Vignette 2: Julian date of first occurrence vs Lat

1. PROCESS TBD

VIII. Vignette 3: Mapping Occurrence over a continental map of U.S

1. PROCESS TBD

IX. Vignette 4: Trends in abundance over time

Misc Notes: <http://neondataskills.org/R/COOP-precip-data-R> <- good for explaining precip data

Further Actions: TBD