Coding Curriculum Outline

Amy Kendig, Whalen Dillon

5/15/2019

## Learning Goals

1. Understand how data collected from experiments leads to conclusions
2. Gain confidence in ability to code
3. Learn about uncertainty/variation/the limits of scientific inference

**Work backwards from data they will work with to skills for visualization**

## Overview

1. Time: 1:15 pm – 4:pm
2. Location: Architecture 116
3. Lessons
   1. Introduction to R
   2. Analyzing data: Biodiversity and its benefits
   3. Analyzing data: Invasive species and infections
   4. If time: Using publicly available data to answer your own questions
4. Refer to the full-science-curriculum document for experiment details relevant to b and c
5. Lessons will involve walking through concepts while students follow along on their computers and then an opportunity for them to try concepts on their own.
6. URL for students to download materials: <https://minhaskamal.github.io/DownGit/#/home?url=https://github.com/aekendig/CodingEnvSci_UF_FieldTrip/tree/master/for_students> or tinyurl.com/UFEnvCode

## Introduction to R

1. Overview of what R is
   1. A free, open-source (define) software for analyzing data and making scientific graphs
   2. R Studio is the program we’ll use to organize all of our tools in R
   3. We’ll use some interactive documents within R studio to learn new concepts
2. Basic concepts that will be needed for the next sections
   1. Simple operations
   2. Variable naming - NOSPACES
      1. Commonly used: variablename, variable\_name, variable.name, variableName, VariableName
      2. Recommend: variable\_name
   3. Commenting code: e.g., #this is a comment, so won’t be read by R
   4. Dataframe
      1. Data types? numeric, character, factor
   5. Plots
   6. Packages?
      1. ggplot
      2. dplyr? magrittr? depending on data manipulations

## Analyzing data: Biodiversity and its benefits

1. Import data
   1. read.csv, readr::read\_csv, readxl::read\_excel
2. Plot biomass vs. species richness
   1. Point plots
3. On their own: plots of pollinators or infection severity vs. species richness
4. Not sure how far into the stats to go. Options:
   1. Mini lesson on linear regression
   2. ggplot linear regressions: ‘lm’ vs. ‘lowess’ method
   3. lm function
5. Translate into a verbal conclusion, talk about variation and next steps

## Analyzing data: Invasive species and infectious disease

1. Import data
2. Make new variables:
   1. LMA = weight / leaf area
   2. Infection severity = area infected / leaf area
   3. Add to a data frame: data\_frame$LMA = weight / leaf\_area
3. 3-D plot of plant traits with points colored by species and shapes by origin
4. On their own: change the color palette or shapes
5. Plots of infection severity vs. leaf traits and origin
   1. Point plots
   2. Bar graphs or boxplots
6. Build upon stats concepts from last lesson because there are multiple predictor variables
7. On their own: write their conclusions in words following the model from the last lesson

### Using publicly available data to answer your own questions

1. Given some dataset options that could include similar analyses to those above
2. In groups of 2-3, write their own question and use the methods they learned to answer it

## Potentially useful resources

### Introduction to R

* <https://www.rforcats.net/>
* <http://stat545.com/topics.html>
* <https://datacarpentry.org/R-ecology-lesson/01-intro-to-r.html>
* Hour of code Python: <https://www.sololearn.com/Play/Python/hoc>
  + This doesn’t seem super helpful, but I like that it asks questions and gives feedback
* <https://swirlstats.com/students.html>
  + Interactive (is within R), but not as visually appealing as the Hour of Code programs
* <https://www.datacamp.com/courses/free-introduction-to-r>
  + Easy and interactive, but requires personal information to log on

### Public datasets

* <https://github.com/awesomedata/awesome-public-datasets>
* <https://www.kaggle.com/datasets> (non-ecology)
  + <https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018>
  + <https://www.kaggle.com/city-of-seattle/seattle-library-collection-inventory>
  + <https://www.kaggle.com/aaronschlegel/austin-animal-center-shelter-intakes-and-outcomes#aac_intakes.csv>
  + <https://www.kaggle.com/cityofLA/la-restaurant-market-health-data>
* <https://datadryad.org/> (FL ecology?)
  + <https://datadryad.org/resource/doi:10.5061/dryad.2ck58>