

# EEB 174 Homework week 8

March 2nd, 2017

## Preliminaries

Create a new ipython notebook called python-week-8-hw-eeb174.ipynb and save your answers in this notebook. Commit the notebook to your homework repo. Also create an R Markdown document called r-week-8-hw-eeb174.Rmd

**Commit all of your work as you go.**

## Part 1: Canids continued

### 1a. Using the python code we generated in Thursday's class as a start do the following:

1. Remove quotes from species names (and any other fields that re written to the output file).
2. Sort the output so that species are listed alphabetically.
3. Write a function that does the following:
  1. takes a file name as argument
  2. opens the file that is formatted like the cleaned version of the canid file we made from the shell in class.
  3. extracts the species name for each species with a latin binomial.
  4. ignores records that have only the genus recorded.
  5. populates a dictionary where the species are keys and the values are the ages of fossils known for each species
  6. returns this dictionary
4. Write a second function that does the following:
  1. takes a dictionary and output file name as arguments
  2. opens a file for writing given the file name in the argument.
  3. determines the minimum and maximum fossil age for each species in the passed dictionary
  4. writes to the outfile a csv with the following fields: genus, species, minage, maxag
5. Write a third function that finds the species with the longest fossil range given a dictionary of species ranges. #####1b.  
Visualization
6. Use the output of your python program above to plot the stratigraphic ranges of the canids in ggplot following the example code from lecture.
7. In your R markdown document write a short report with the following information about your figure:
  1. What taxonomic group is displayed?
  2. How many unique fossil species are contained within it?
  3. How many unique fossil genera?
  4. What species has the longest
8. Repeat parts 1a and 1b with a new family or order from the [PBDB](#). You may use the taxonomic group you have chosen for your final project or another group. Make sure that whatever group you decide to use has at least 100 unique species. **Note that will need to check the *Show accepted names only* box next to *Select by taxonomy* to ensure that the field order is the same as in the example file.**
9. (Extra Credit) Write a set of python functions that will take species level fossil occurrence data such as that used in parts 1 and 2 and produce the minimum and maximum age of each **genus** represented in the file. Then use R to plot the stratigraphic ranges of each genus.

## Part 2: Data Frames

In your R Markdown document complete the following exercises.

1. [Exploring Data Frames](#) Show your work for the examples and answer Challenges 1-8.
2. [Subsetting Data](#). Show your work for the examples and answer Challenges 1-8.
3. [Creating Publication Quality Graphcs](#). Show your work for the examples and answer Challenges 1-5.

## Here are some useful ggplot resources

[R Markdown cheatsheet](#)

[ggplot cheatsheet](#)

[intro to r graphics with ggplot](#)