

# EEB 174 Homework week 2

## Preliminaries

Create a new repo in your `eeb-177` Desktop folder called `homework`. Follow the [instructions](#) from week 0 for creating a new git repo.

Create a file called `week-2-hw-eeb174.txt` in your `homework` directory.

Use this file for your homework answers.

## Question 1.

Lets revisit the `Pacifici2013_data.csv` file from class today (in `CSB/unix/data`). Show the Order of the first 5 species in the data set.

(hint: you will need `cut`, `l`, `tail`, and `head`)

## Question 2.

How many unique families are in this data set? Append the line "There are X unique families:" (fill in the value for x) to your exercise file. Then append the list of unique families to your hw file.

**hint:** you will need the `uniq` command. Also, you will to sort your data before you apply `uniq` as `uniq` only collapses non-unique values when they are contiguous.

## Question 3.

Work through the Allesina example 1.7.3 that introduces the `tr` command and shows how the commands you have learned can be used to subset the `Pacifici2013_data.csv`. Append your command history **for 1.7.3 only** to your hw file.

## Question 4.

What is the 10th largest species in the database? What is the 10th smallest?

## Question 5.

Work through section 1.7.4 in the Allesina text on `grep`. Append your command history **for 1.7.3 only** to your hw file.

## Question 6.

Do exercise 1.11.1 in Allesina.

**hint 1:** you will need to use the squeeze option (`-s`) to deal with repeat spaces for #5.

**hint 2:** #7 requires numeric sorting on the number of reads. How could you get a column of numreads only and then how could you cut the number from the text "numread"?

## Question 7

Read the software carpentry section on loops in the shell [here](#). Then do the following.

- create a directory called `temp` in `CSB/unix/sandbox`
- `cd` to the `Saavedra2013` directory in `CSB/unix/data`



- **write a loop** that will copy all of the text files in this directory to your newly created `temp` directory. The copied files should all be renamed to include "edited-" in front of the original file name. Append this loop to your HW file.


**hint:** you will need to give the `cp` command the relative (or absolute) path to the target directory



**hint** use the `echo` command to see the new file name in the loop before you copy as in the *Nelle's Pipeline: Processing Files* section of the loops page.



## Question 8



Work through the software carpentry exercise on finding things [here](#). Answer the questions in the orange boxes at the end of this section in your HW file.



 Using grep 

 find Pipeline Reading Comprehension 

 Matching and Subtracting 

 Tracking a Species 

 Little Women 

 Finding Files With Different Properties 

---

Commit all of your work as you go.

## Due Dates

Homework must be committed by **Monday, January 23rd at 5:00 PM.**