BIOL343 – Assignment #2

Fun with ggplot

Assigned Sunday 13 January
Due Saturday 19 January 1159pm

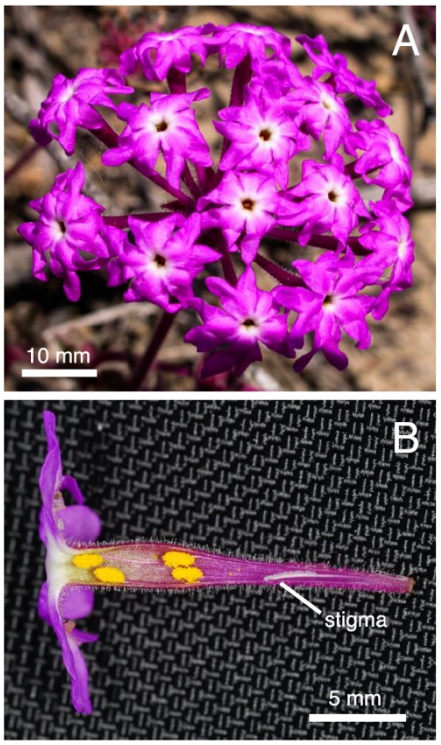

FIGURE 1. Morphology of inflorescences (A) and flowers (B) of *Abronia umbellata* var. *umbellata* (Nyctaginaceae).

This assignment will give you some practice with using the **ggplot2** package for data visualization. You will use a dataset from a study of variation in floral morphology across the geographic range of the Pacific coast dune endemic plant *Abronia umbellata* (Nyctaginceae). These were collected as part of my former student Laura Doubleday's research. For more info on Laura's field work on the west coast go here: https://youtu.be/TNllnUEbVr8

Across its coastal range *Abronia umbellata* exhibits striking variation in self-incompatibility (SI), a physiological response that prevents a plant being fertilized by its own pollen. Because plants are usually hermaphroditic, SI has evolved to prevent the production of genetically inferior selfed offspring. However, under certain ecological conditions (e.g. scarce pollinators, low density of compatible conspecific mates) self-fertilization may be favoured by natural selection, resulting in the evolution of self-compatible (SC) populations. The dataset you will analyze contains data on flower and inflorescence morphology measured on a random sample of plants from 12 sites, from just north of Los Angeles California to the species northern range limit in southern Oregon. Most of the populations contain SI plants but some contain SC plants.

By graphing the data in various ways are going to evaluate two hypotheses:

(1) SC evolves towards geographical range limits where mate density and possibly pollinator service is particularly low.
(2) Once SC evolves, selection favours reduced investment in floral traits that attract pollinators because a selfing individual does not require assisted pollination.

You will hand in your work as an R Notebook that includes all the code required to address the questions below. All the graphs for this assignment – in fact all graphs you make for BIOL343 – must be outfitted with axis titles that include appropriate units. Axis titles are sentence case (i.e. Capitalize the first word of the title only, unless proper nouns are involved). Each figure must also have figure caption that adequately describes the figure and any special symbols used. See an example to the right.

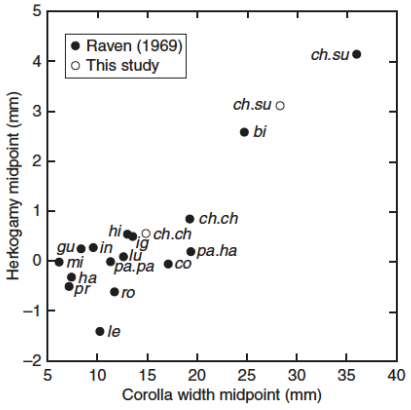

FIG. 8. Range of evolutionary diversification in floral morphology within the genus *Camissoniopsis*. This figure shows covariation between corolla width and herkogamy among taxa. Filled circles are the midpoints between phenotypic values provided by Raven (1969), open circles are midpoints calculated from data collected in this study. The acronyms indicating species identity are as follows: *ch.su*, *C. cheiranthifolia suffruticosa*; *ch.ch*, *C. cheiranthifolia cheiranthifolia*; *bi*, *C. bistorta*; *pa.ha*, *C. pallida hallii*; *pa.pa*, *C. pallida pallida*; *co*, *C. confusa*; *ig*, *C. ignota*; *lu*, *C. luciae*; *in*, *C. intermedia*; *hi*, *C. hirtella*; *gu*, *C. guadalupensis*; *mi*, *C. micrantha*; *ha*, *C. hardhamiae*; *pr*, *C. proavita*; *ro*, *C. robusta*.

Here's what's in the data set:

- **self_inc** = An indication of whether the plant comes from a self-incompatible (SI) or self-compatible (SC) population
- **site** = The code for the costal site where the data were collected
- **latitude_N** = The latitude of the site (in ° N). Note that all plants at the same site have the same value for latitude.
- **umbel_angle_deg** = Umbel angle measured in degrees. This measure indicates how spread out the flowers in an inflorescence are (see photo above). The greater the angle, the more spread out and available to pollinators the individual flowers in an inflorescence are.
- **flower_number** = The number of flowers in an inflorescence. More flowers make an inflorescence more attractive to pollinators, so we might expect flower number to decline in conjunction with the evolution of selfing.
- **face_diam_mm** = Flowers are tubular with a wide face, the diameter of which is measured here. In the picture of the inflorescence above the flowers are facing you. Larger faces are expected to be more attractive for pollinators.
- **tube_length_mm** = You can see in the cross-sectional photo that flowers have a long tube, which is measured here from the opening at the face to the bottom of the tube.

(1) Import the data into R and check it over using the set of 4 functions that you used in last week's assignment. This is standard procedure.

(2) Plotting histograms of the data is an excellent way to detect typographical errors that might have occurred during data entry and/or strange-but-true biological outliers. Use `ggplot()` to make a histogram of floral face diameter separately for SI and SC plants, and stack these histograms on top of each other. Make sure that you modify with width of the bars in each histogram to display the distribution of the data with a reasonable level of resolution. Once you've checked the first variable, you would normally repeat these steps for the other three variables, but you don't need to do this for the purposes of this assignment.

(3) Pipe together the **dplyr** functions required to calculate the sample size and mean for each of the four traits for SI vs. SC plants separately. This information will appear in your notebook below the appropriate code chunk. Write the code so that it deals appropriately with missing values if there are any. These summary statistics will give an idea of whether the traits differ between SI and SC plants in the expected way.

(4) Graphically contrast the distribution of umbel angle among SC vs. SI sites using boxplots. Fill the boxplots for SI and SC sites with different colours and arrange sites along the x-axis from left to right in order of increasing north latitude.

(5) Repeat the boxplots in task #4 for the other three variables. You don't have to show the figures in your R Notebook but simply describe the results in the text. Which variables show the largest contrasts between SI vs. SC populations?

(6) Interpret the results of #4 and #5 in terms of the two hypotheses you are testing.

(7) Using piped **dplyr** functions, calculate the site means for latitude and the 4 floral traits, and put these means in a new dataframe. Within the new dataframe, sites should be organized by mating system. You will use this dataset in the task below.

(8) On one graph, plot the relation between floral face diameter (x-axis) and tube length (y-axis). Use different coloured points for each site. These points are individual plants. Draw a simple linear regression line through the data for each site. These regression lines should be coloured by site, just like the data points. Now comes the tricky part: On the same graph, plot the population means for these two floral traits and colour those means by site as you did for the individual data points.

(9) The graph you just made shows the relation between these two main components of floral morphology. We might expect this relation to change during an evolutionary shift from outcrossing to selfing. For instance, selection may favour an uncoupling of tube length and face diameter in outcrossing populations because these traits have quite different functions. Tube length determines contact between pollinator and the floral sex organs (anthers and stigma, see the photo above), whereas the face diameter influences the attractiveness of the flower. In selfing populations, however, both traits are relatively unimportant components of overall flower size, which we expect to be reduced in tandem. What prediction emerges from this hypothesis? Does your graph support this prediction? Explain?

For this assignment, you will submit a **PDF version** of your .html R notebook document called "StudentNumber_A2.pdf", where the file name starts with your student number. Remember, it is your responsibility to make sure that the PDF version of your R Notebook is as nicely formatted as possible and includes all the required components. So, please look carefully at your PDF file before you submit it. You should upload your file to the Assignment #2 OnQ dropbox by **Saturday 18 January at 1159pm**.

As always, Hana and I will be in tutorial to provide advice.
Have fun!!