Coastal Solutions Workshop
MODULE 1: Introduction to R – Extra Practice
February 20, 2023 2:30-5:30 PM

---------- **Extra practice!** ----------

We'll start by using the worldfloras dataset to get some practice with the following:

- Writing an organized script from scratch.
- Using indexing, `apply()`, and `which()` to access and subset data.
- Analyzing data with a simple linear model.
- Plotting

Keep your code in an organized script with each task labeled. Our hope is that you can use this script as a reference later on! So let's get set up:

- Open up a new script and give it a title. Don't forget to save your script along the way.
- Set your working directory.
- Load the worldfloras dataset.
- To keep this tidy, we suggest organizing your script with headers for each question number. Something like this:

```
#### ------------------------------------------------------------ ####
#### Script title
#### ------------------------------------------------------------ ####

#### Set working directory

#### Load data

#### Question 1

## [if needed, include subheader within a question]
```

First, let's refamiliarize ourselves with the data.

1. Glimpse the first and last 6 rows of the dataset using the functions `head()` and `tail()`.

2. What is the class of each column?

3. How many continents are in this dataset? *Hint: ?unique()*.

4. Calculate the mean, median, and sd for the population column.

5. Plot a simple histogram of the endemism column. What happens when you change the "breaks" argument?

6. Create a new column in the dataset which shows the number of species per unit area for each country. What is the mean of this column?

Okay, time to get some practice with `which()` and `apply()`!

7. Use the function `summary()` to examine the $Flora column. Something seems off here... how can you have -1 species in your country? How many times does this error occur?

8. Looks like your data entry team entered a -1 when they should have entered NA. Replace these values with NA. *Hint: Try indexing out the -1's, then use <− NA to replace*.

9. Which country has the lowest population?

10. What is the mean endemism rate in the dataset? *Warning: looks like there's an issue – some of the data values are NA! Use ?mean() to find a solution.*

11. What is the average endemism rate for each continent?

12. How many countries does this dataset have for each continent? Hint: *?table()*

13. Of the Asian countries, how many have >10% endemism and >3000 species?

    Challenge question: Calculate the average latitude for this subset of countries.

Let's run a simple linear model.

14. Use indexing to remove all countries with an area > 3000 and save the result as a new object. Why is it good practice to create a new object when filtering the dataset?

    How many countries were removed by this filtering?

15. Using this filtered dataset, run a simple linear model to test the hypothesis that the number of species increases with increasing area.

    What is the mean coefficient estimate for the effect of area?

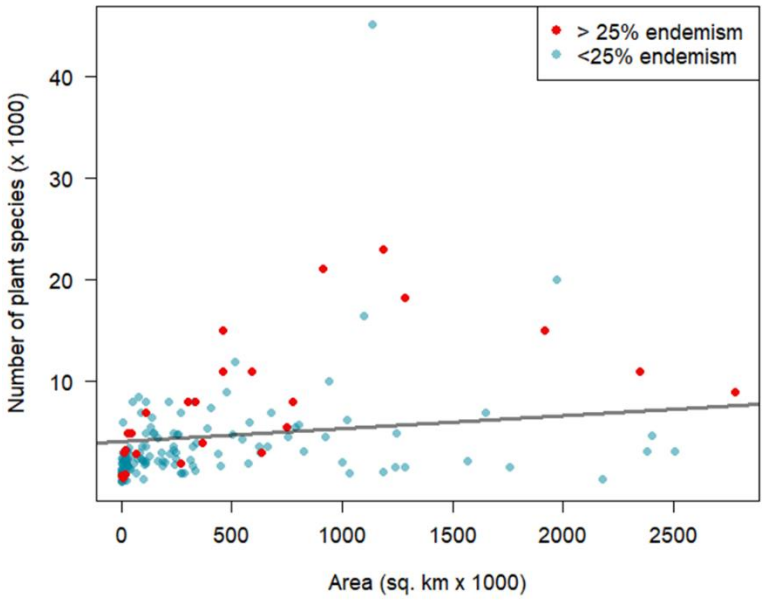
    What is the adjusted R^2 value for this model?

Into the weeds with plotting.

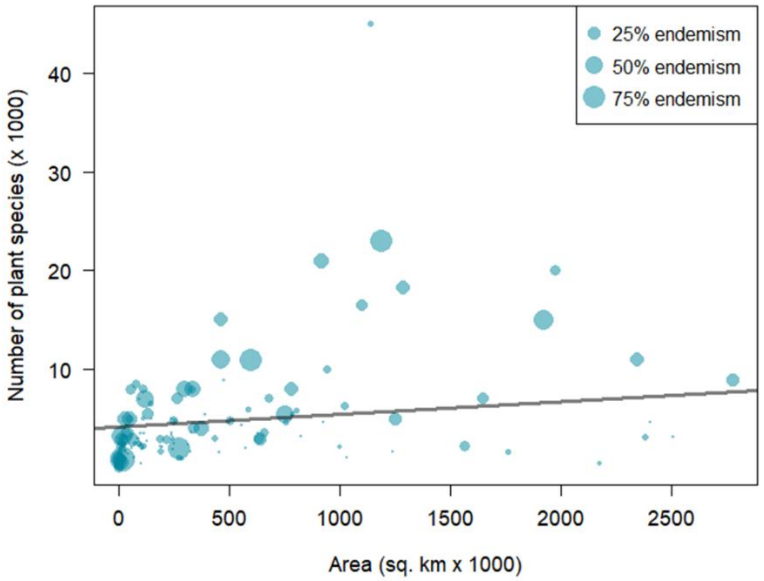
16. Great work on that model. Now we need to make a pretty plot of the results. Using your knowledge of plotting and R's help files, see how close you can come to replicating this plot of the results!

    Hints

- *Start by setting some plotting parameters with `par()`. The "cex" series of arguments controls the relative size of things in the plot.*
- *Our suggestion is to begin by plotting an empty plot window using the argument type="n". Then layer on addition parts of the plot. In creating this example, we used the lower-level functions `axis()`, `points()`, and `abline()`.*
    - *The two point colors in this plot are "#0087A080" and "#EC0909"*
- *Last, add a legend.*



Optional plotting challenge: can you play with the "cex" argument to make the point size vary by endemism rate?



---------- **Analyzing woodpecker home ranges** ----------

Explore the data

Recall that this dataset includes home range estimates from 32 radio-tagged Black-backed Woodpeckers in California. There are 5 columns:

BirdID = A number assigned to each tagged bird.
HR_size = Home range area (ha).
Dead_tree_BA = Basal area of dead trees within the home range (m^2/ha).
Elevation = Mean elevation (m).
Nat_forest = National forest where tracking occurred.

1) What is the ID number for the bird with the largest home range? How many hectares is it?

2) What is the median elevation of field sites within each national forest?

3) What is the mean home range size for just the birds in Lassen National Forest?

4) How many birds did they track in each national forest?

Analyze the data

5) Run a test to see if there is a correlation between (a) home range size and elevation, and (b) home range size and dead tree basal area. Which of the environmental covariates has a stronger correlation based on Pearson's correlation coefficient?

6) Run a simple linear model to examine the effect of *both* elevation and dead tree basal area on home range size. Your model will follow the formula:

```
HR_size ~ Dead_tree_BA + Elevation
```

7) Print a summary of this model. What are the coefficient estimates for the two covariates? Do either of these covariates have a statistically significant relationship to home range size based on $\alpha = 0.05$?

Create your own plot

Time to get creative! Your task is to make your own plot to visualize the model results while showcasing your new plotting skills. There is no single right answer here. We're looking for a plot with the following:

- Include 2 panels, one showing the relationship between dead tree basal area and home range size, and the other showing the relationship between elevation and home range size. *Hint: within the par() function, set mfrow=c(2,1).*
- Include a line showing the linear relationship between X and Y for each variable.

- Use shape, size, color, or text to show additional information beyond a simple scatterplot. For example, you use different colors or point shapes for different national forests. Don’t forget the legend!
- Give the plot a nice, professional look by giving it descriptive axis labels and other modifications that look good to you.