BIOL343 – Assignment #3

Does the efficacy of biocontrol increase with regional host abundance? A case study with the invasive wetland plant purple loosestrife.

Assigned Sunday 19 January
Due Saturday 25 January 1159pm

Biological control is a potentially cost-effective and powerful management tool for controlling the spread and impact of established, widespread exotic species. Developing successful biocontrol programs depends upon understanding the factors that influence control agent success following establishment. Among populations of the target species, the survival, reproduction and feeding activity of control agents may vary with a variety of abiotic or biotic factors. At the landscape level, habitat connectivity and target host distribution are thought to play a significant role in dictating the spread of a control agent from its initial release sites and hence its broader impact on the regional dynamics of the target species. This expectation derives from metapopulation theory, which predicts that shorter dispersal distances between target host patches increase the rate of patch colonization. However, the influence of host distribution at regional scales on the efficacy of biological control is virtually unstudied.

A better understanding of the landscape-level factors that influence the success of biological control can only be gained through adequate post-release monitoring of control agent performance, yet this is rarely performed. Between 1991–1996, 70% of invasive plant biocontrol agents introduced globally established effectively. Only 24% of these established agents were subjected to post-release monitoring to determine their efficacy. A review of the primary literature indicates that of those 16 biocontrol agents introduced to Canada during this time period, only 59% were subjected to post-release monitoring and most of these were monitored at release sites only. The extent of spread to nonrelease sites was only determined for 24% of agents. This is problematic because several studies have found that the probability of biocontrol agents being present in stands of their invasive host plant decreases with increasing distance from release sites. Hence monitoring only at release sites may substantially overestimate the regional efficacy of biocontrol.

The Case of Purple Loosestrife

For this assignment, you are going to analyze data collected as part of a study investigating the influence of host distribution on the success of the purple loosestrife (*Lythrum salicaria* L., Lythraceae) biocontrol program in Ontario, Canada. *Lythrum salicaria* is a perennial wetland plant that was introduced from Eurasia to central North America via ship ballast and horticulture in the

early 1800s. During the 20th century the plant spread across the continent and established dense populations in many regions.

A biological control program was initiated in the USA and Canada in 1992; two Chrysomelid beetle species, *Galerucella pusilla* and *G. calmariensis*, were widely introduced to populations throughout North America and more specifically in southern Ontario. Adults and larvae of these beetles eat leaves thereby reducing the production and storage of photosynthate, and larvae also consume the apical meristem thereby reducing vertical stem growth and inducing branching. In Ontario, > 320,000 beetles were released between 1992-1997 at > 200 sites. Province-wide monitoring of documented release sites only was conducted in 1997 and again in 2004. Little impact was detected in 1997 but by 2004 beetles were present at 90% of release sites, and had imposed some degree of control at 66%. Anecdotal evidence suggested that the beetles had dispersed widely to *L. salicaria* stands where they had not been released but no one has tested this.

Extensive field surveys and analysis of herbarium records indicates that stands of *L. salicaria* are historically much more abundant in eastern Ontario than central Ontario. If stands are fewer and further between in central than eastern Ontario, *Galerucella* beetles may disperse less frequently among central Ontario stands yielding lower levels of plant damage on average. Longer distances between host stands may also reduce the metapopulation "rescue effect" in which immigration protects small, newly established populations of the control agent from extinction.

Objectives

In this assignment, you will be testing three hypotheses:

(1) Beetle damage index is lower and more variable in regions where the host plant is more sparsely distributed.
(2) Damage index is higher and less variable at sites where the biocontrol agents was deliberately released than sites where they were not.
(3) Damage by beetles reduces the local abundance and alters the growth form of *L. salicaria*.

The Data

Angela Boag, an undergraduate thesis student, and I surveyed beetle damage and plant traits correlated with damage at each of 52 stands of *L. salicaria* across Ontario, Canada. Among these sites, 18 were in the high-density eastern Ontario region and 20 were in the lower density central Ontario region. Beetles were intentionally released at 20 sites and 32 sites had no known history of beetle release.

You have been provided with a .csv file (loosestrife.csv) containing the following columns:
(1) site = A unique identification code for each site sampled
(2) latitude in °N
(3) longitude in °W
(4) region = a factor indicating whether the site was in eastern Ontario (East), central Ontario (Central) or elsewhere in the province (Other). Testing the regional density hypothesis will only involve the East and Central sites.

(5) release = a factor indicating whether beetles had been intentionally released (Release) or not (Nonrelease)
(6) stemsper50m2 = the density of *L. salicaria* stems measured from a 50 m^2 belt transect at each site. This is our best measure of the local abundance of *L. salicaria*.
(7) stemheight_cm = the height of the primary stem (in cm) for a sample of 20 randomly selected plants per site. Because the biocontrol agents destroy apical meristems, stem height should be lower in sites more heavily damaged by the biocontrol agents. A measure of growth form.
(8) inflorescences = the number of inflorescences per plant borne by a sample of 20 randomly selected plants per site. Because the biocontrol agents destroy apical meristems thereby increasing branching, the number of inflorescences should be higher in sites more heavily damaged by biocontrol agents. Another measure of growth form.
(9) damage = leaf damage index: herbivore damage subjectively scored using a semi-quantitative 0–5 scale, where 0 denoted no diagnostic evidence of *Galerucella* herbivory, while 5 denoted severe damage. Subjective visual measurement of leaf damage on *L. salicaria* correlates very strongly with direct quantitative measurement of leaf damage.

The Assignment

With these data, you will test the three hypotheses stated above using graphical analysis and summary statistics. You should complete your assignment as an R Notebook and submit it to onQ as a PDF.

Each graphical display of data should be suitable for publication with proper axis titles and a descriptive figure caption. The third hypothesis is best evaluated with a composite graph that includes three separate panels. There are several ways of achieving this in **ggplot**, but we will use the `plot_grid()` function in the **cowplot** package.

Of course, the first step is to check the dataframe and deal with any issues in the data. If you find a data point outlier, explain and provide a rationale for how you recoded the suspicious data point. All modifications of the original data must be done in R and recorded in your R Notebook. Use all the **dplyr** and **ggplot** tricks that you know to make your Notebook look as professional as possible. Make sure your code is as concise and readable as possible.

So, here's a summary of what Hana and I are expecting from you:

1. A code chunk loading the required packages.
2. Code importing the dataframe.
3. Code chunks checking the dataframe with the standard 4 functions.
4. Code that plots the distribution of the main continuous variables, with which you can identify outliers or mistakes and you can assess whether any of the variables are log-normally distributed. Show only one example graph. No figure caption required.
5. A section of text summarizing what you've discovered about your data and any changes or transformations you decided to make.
6. Code that creates transformed versions of any variables that require transformation, and makes any other necessary changes to the data.
7. A plot that contrasts distributions of damage between regions. The plot code must include both scale and theme functions.

8. Code and output that compares the *mean* and *variability* of damage between eastern and central regions. You should use the *coefficient of variation* as your measure of variability. This will require that you make your own function and use it in a set of piped **dplyr** functions. Please do not calculate summary statistics individually. You can calculate all of them at once using a single chain of **dplyr** commands using `group_by()` and `summarise()`. Remember: it is a virtue to write concise, compact code!
9. A plot that contrasts the distribution of damage between release and nonrelease sites. This plot code must include both scale and theme functions.
10. Code and output that compares the *mean* and *variability* of damage between release and nonrelease sites. See comments in task #8.
11. A composite plot that explores whether damage index is related to the density of stems at a site and the aspects of plant growth form. Colour the points for release sites "blue" and nonrelease sites "red. Pass a simple linear regression line through these points. This plot code must include both scale and theme functions.
12. An interpretation of how the results support or fail to support the three hypotheses you are testing.
13. Each figure requires a formal caption that fully describes the components of the figure so that your reader can understand it.

A couple of technical challenges:

(1) You will be plotting $\log_{10}$-transformed stems per m^2 and must use a suitable axis title for this. Notice that the text for the axis label has one subscript (the "10") and one superscript (the "2" on "m"). You can modify the text of axis labels in this way using the expression() function, but I always find it kind of tricky to figure out the code for any given example. Google is your friend here.

(2) Making composite plots using ggplot is tricky because ggplot has no built-in capacity for composite plotting except maybe the facet functions, which don't always give you what you want. There are several packages such as **grid** and **gridExtra** that contain some appropriate functions but things get technical quickly. I suggest that you use `plot_grid()` in the **cowplot** package. To get the hang of this, check out the tutorial at:
https://wilkelab.org/cowplot/articles/plot_grid.html
Please make sure that the axes of the three plots are perfectly aligned in your composite figure. This will require that you understand the various arguments in the `plot_grid()` function.

You will name the PDF version of your R notebook .html file "StudentNumber_A3.pdf" (as in assignment #2), and upload it to the Assignment #3 OnQ dropbox by **Saturday 25 January at 1159pm**.

Remember to carefully check the resulting PDF file before submitting it.