

20019803_A3

Code ▾

Assignment:

1. A code chunk loading the required packages.

Hide

```
library(dplyr)
library(ggplot2)
library(ggpubr)
```

Note: I am using ggpubr for making the composite plot rather than cowplot. Cowplot's installation was giving me alot of problems, and ggpubr's ggarrange function proved intuitive for creating composite graphs.

- 2/3) import and check the data frame with the standard functions

Hide

```
loosDat<-read.csv("/Users/kelli/Desktop/Biol343/Assignment_3/loosestrife.csv")
summary(loosDat)
```

```
      site      latn      lonw      region      release
B1.PLD : 1   Min.    :42.58   Min.    :-81.52   Central:20   Nonrelease:32
B2.NUM : 1   1st Qu.:44.10   1st Qu.: -79.83   East  :18   Release  :20
B3.LDY : 1   Median :44.56   Median : -79.27   Other :14
B4.SCUG: 1   Mean    :44.41   Mean    :-78.40
B5.CAL : 1   3rd Qu.:44.73   3rd Qu.: -76.25
B6.DUF : 1   Max.    :45.74   Max.    :-75.67
(Other):46
      stemdens      stemheight_cm      inflorescences      damage
Min.    :    0.0   Min.    : 42.0   Min.    : 0.100   Min.    :0.3464
1st Qu.:   28.0   1st Qu.:106.5   1st Qu.: 2.525   1st Qu.:2.0282
Median :   86.0   Median :121.0   Median : 4.400   Median :2.3726
Mean    :  138.3   Mean    :118.8   Mean    :14.901   Mean    :2.5564
3rd Qu.:  177.5   3rd Qu.:135.5   3rd Qu.: 6.000   3rd Qu.:3.1445
Max.    :1000.0   Max.    :171.0   Max.    :515.000   Max.    :4.4505
NA's    :1       NA's    :1       NA's    :1
```

Hide

```
head(loosDat)
```

site <fctr>	latn <dbl>	lonw <dbl>	region <fctr>	release <fctr>	stemd... <int>	stemheight_cm <dbl>	infloresce
1E10.MTALB	44.20960	-79.33869	Central	Nonrelease	191	130	
2E11.JP	44.32528	-79.37285	Central	Nonrelease	NA	117	

site <fctr>	latn <dbl>	lonw <dbl>	region <fctr>	release <fctr>	stemd... <int>	stemheight_cm <dbl>	infloresce
3E12.HWY8.1	44.22103	-79.44601	Central	Nonrelease	96	101	
4E13.HWY8.2	44.28527	-79.44429	Central	Nonrelease	86	111	
5E14.HWY48	44.32518	-79.23592	Central	Nonrelease	15	121	
6E16.ROG2	44.08084	-79.45420	Central	Nonrelease	166	152	

6 rows

Hide

```
tail(loosDat)
```

site <fctr>	latn <dbl>	lonw <dbl>	region <fctr>	release <fctr>	stemd... <int>	stemheight_cm <dbl>	inflorescences <dbl>
47 W7.DDAM	44.50415	-81.33283	Other	Release	3	54.21053	0.6315789
48 W8.RPK	43.40403	-80.36287	Other	Release	3	123.33333	11.6666667
49 W9.HWY17	43.53402	-80.48205	Other	Release	11	89.00000	3.5000000
50 B5.CAL	43.81927	-79.94017	Other	Release	2	115.00000	0.3000000
51 B7.LEN	43.41978	-80.32984	Other	Release	5	102.00000	NA
52 B8.WPND5	42.92883	-80.31860	Other	Release	20	145.00000	4.2500000

6 rows

Hide

```
class(loosDat)
```

```
[1] "data.frame"
```

Hide

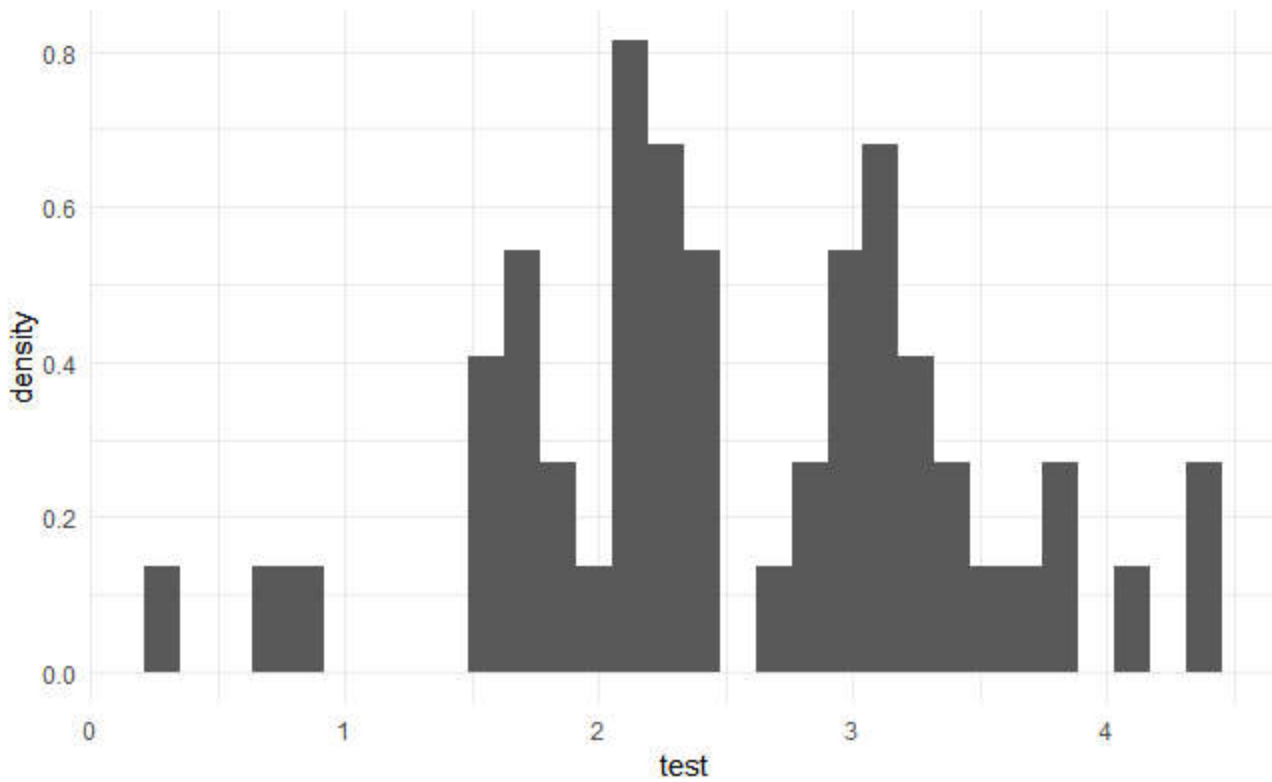
```
dim(loosDat)
```

```
[1] 52  9
```

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

Hide

```
loosDat <- loosDat %>% mutate(test = damage)
ggplot(loosDat) + geom_histogram (aes(x=test, y = ..density..))+
  theme_minimal()
```


[Hide](#)

```
#ggplot(loosDat) + geom_histogram (aes(x=log10(1+test), y = ..density..)) + theme_minimal()
```

5. A section of text summarizing what you've discovered about your data and any changes or transformations you decided to make.

There is a significant outlier in the inflorescences data that is skewing the data and making it illegible on the histogram. This outlier will be removed. Additionally, as stem density is heavily skewed I've log transformed it to make it a more normally distributed dataset. Finally, I have created a new dataframe without the "other" region level. This will enable easier testing of hypothesis 1, which focuses specifically on the eastern and central regions, while maintaining the data from "other" regions for use in testing the other hypothesis.

6. Code that creates transformed versions of any variables that require transformation, and makes any other necessary changes to the data.

[Hide](#)

```
loosDat$inflorescences[which(loosDat$inflorescences>100)] = NA
loosDat <- loosDat %>% mutate(logDense = log10(1 + stemdens))
regDat <- loosDat %>% filter(region != "Other")
```

7. A plot that contrasts distributions of damage between regions. The plot code must include both scale and theme functions.

Hide

```
ggplot(regDat) + geom_boxplot (aes(x=region, y= damage), position = "identity", alpha
= 0.5) + scale_y_continuous(name = "Damage index") + scale_x_discrete(name ="Region")
+ theme_minimal()
```

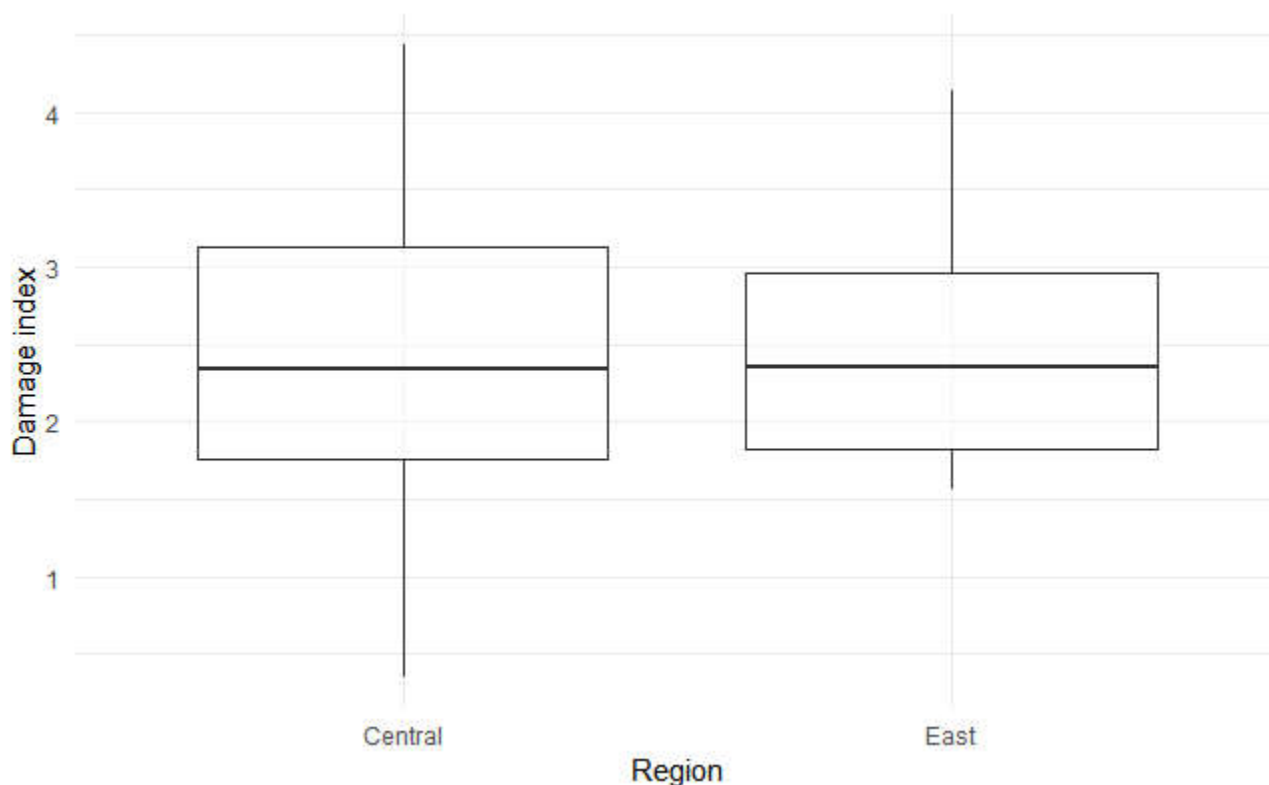


Figure 1. Damage index by site location in ontario. Site included were either found in the central or eastern regions of ontario. Scattered sites from other regions of the province have been omitted.

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!

Hide

```
cv <- function(i){
  sd(i)/mean(i)*100
}
regDat %>% group_by(region) %>% summarise(n = sum(!is.na(damage)), mean = mean(damage, na.rm = T), cv = cv(damage))
```

region <fctr>	n <int>	mean <dbl>	cv <dbl>
Central	20	2.373849	44.55229
East	18	2.487880	30.13860

2 rows

9. A plot that contrasts the distribution of damage between release and nonrelease sites. This plot code must include both scale and theme functions.

Hide

```
ggplot(loosDat) + geom_boxplot(aes(x=release, y=damage)) + theme_minimal() + scale_y_continuous(name = "Damage index") + scale_x_discrete(name="Beetle release status")
```

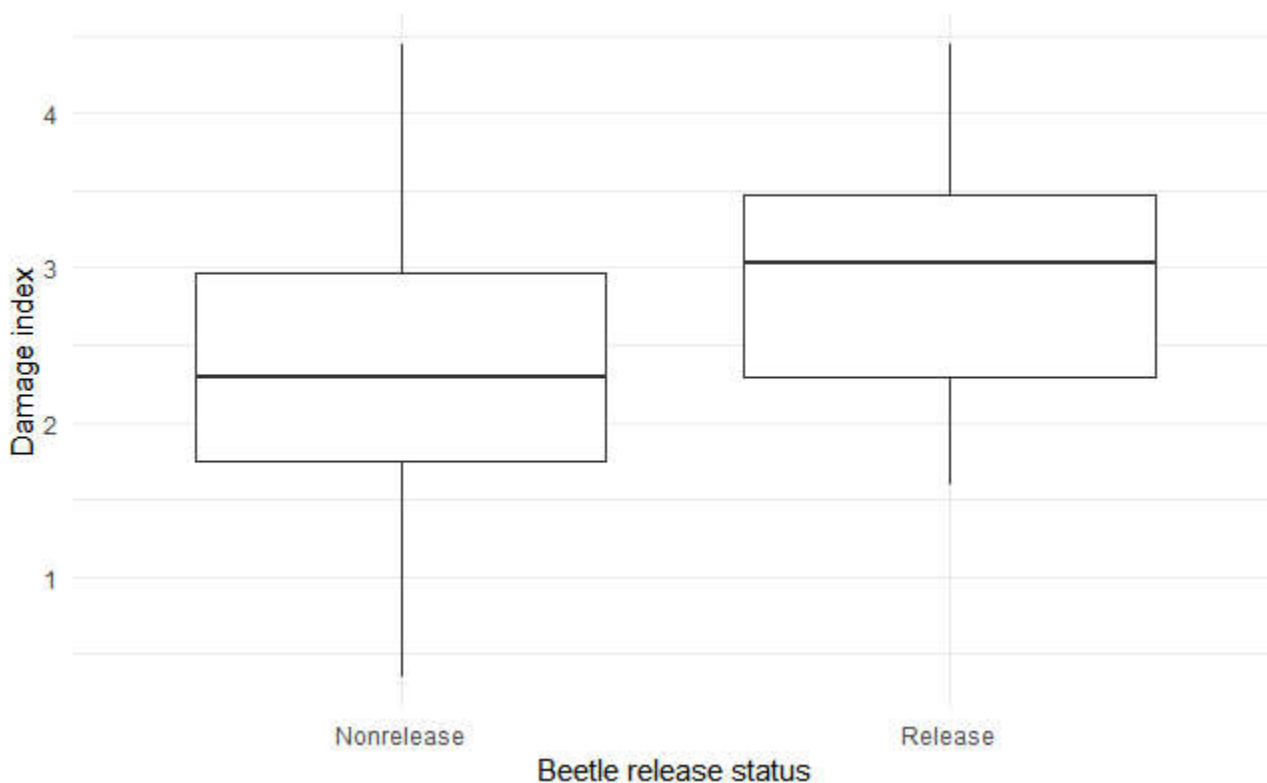


Figure 2. Damage index distributions across sites where beetles were intentionally released and not-released.

10. Code and output that compares the mean and variability of damage between release and nonrelease sites. See comments in task #8.

Hide

```
loosDat %>% group_by(release) %>% summarise(n = sum(!is.na(damage)), mean = mean(damage, na.rm = T), cv = cv(damage))
```

release <fctr>	n <int>	mean <dbl>	cv <dbl>
Nonrelease	32	2.299327	38.15632
Release	20	2.967664	25.41741

2 rows

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.

Hide

```
den <- ggplot(loosDat, aes(y = logDense, x = damage)) + geom_point(aes(colour = release)) + scale_color_manual(breaks = c("Release", "Nonrelease"), values=c("red","blue")) + labs(x = "", y = expression(paste("Stems per ",m^2," (",log[10],")"), colour = "")) +
  theme(axis.title.y= element_text(size = 9.5)) + geom_smooth(method = lm, se = FALSE) #+ theme_minimal()
hei <- ggplot(loosDat, aes(y = stemheight_cm, x = damage)) + geom_point(aes(colour = release)) + scale_color_manual(breaks = c("Release", "Nonrelease"), values=c("red","blue")) + labs(x = "", y = "Stem height (cm)", colour = "") +
  theme(axis.title.y= element_text(size = 9.5,vjust = 2, hjust = .5) ) + geom_smooth(method = lm, se = FALSE) #+ theme_minimal()
inf <- ggplot(loosDat, aes(y = inflorescences, x = damage)) + geom_point(aes(colour = release)) + scale_color_manual(breaks = c("Release", "Nonrelease"), values=c("red","blue")) +
  theme(axis.title.y= element_text(size = 9.5)) + labs(x = "", y = "Inflorescences", colour = "") + geom_smooth(method = lm, se = FALSE) #+ theme_minimal()
ffigure <- ggarrange(den,hei,inf, ncol = 1, nrow = 3, common.legend = T)
```

Removed 1 rows containing non-finite values (stat_smooth).Removed 1 rows containing missing values (geom_point).Removed 1 rows containing non-finite values (stat_smooth).Removed 1 rows containing missing values (geom_point).Removed 1 rows containing non-finite values (stat_smooth).Removed 1 rows containing missing values (geom_point).Removed 2 rows containing non-finite values (stat_smooth).Removed 2 rows containing missing values (geom_point).

Hide

```
annotate_figure(ffigure,
  bottom = text_grob("Damage index", rot = 0, size = 9.5, vjust = -1.5))
```

Figure 3. Damage index plotted against (from top to bottom) stem density, stem height, and inflorescences. Sites where beetles were intentionally released are highlighted in blue, while non-release sites appear in red. The blue line represents a general linear trend for each relationship.

12. An interpretation of how the results support or fail to support the three hypotheses you are testing.

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 supports hypotheses 2, and partially supports 3, but fails to support hypothesis 1.

Region data shows that there is no difference between mean or variability across regions of high (eastern) or

low (central) plant density (figure 1.)

However, when the damage is plotted against site release status (figure 2.) we observe an increase in the mean damage, and a reduction in the variance, supporting hypothesis 2.

Finally, when damage is plotted against the aspects of plant growth (figure 3.) we can observe a negative correlation between damage and plant density. However, there appears to be little if any correlation between damage and the plant growth features height and inflorescence number. This partially supports hypothesis 3, as beetle damage appears to induce a loss of plant density, however it seems to have little impact on the growth for of *L. salicaria*.

13. Each figure requires a formal caption that fully describes the components of the figure so that your reader can understand it.