

# Supplemental Material

*CBE—Life Sciences Education*

Guzman *et al.*

## Appendix 1 - Homework assignments

The homework assignments were designed to (i) reduce the extraneous load, (ii) reduce the intrinsic load and (iii) increase the germane load.

### *(i) Reducing the extraneous load*

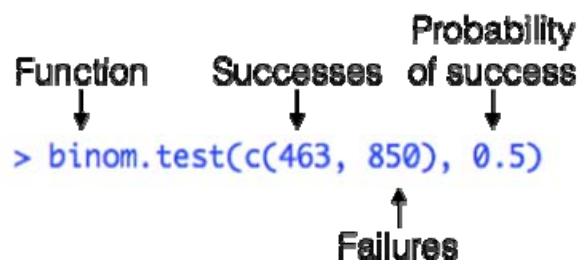
#### *Split attention effect*

Code is often presented separate from its explanation, for example:

```
binom.test(c(463, 850), 0.5)
```

Where 0.5 is the probability of success, 463 is the number of successes, 850 is the number of failures and binom.test is the function to run a binomial test.

We minimized the split attention effect for the students learning new code by providing code with the english explanations of each part:



And by providing explanations of the code output:

## Variable

↓  
> grouped\_titanic  
Source: local data frame [1,313 x 7]  
Groups: Sex, Survive [4]

	Passenger.Class	Name	Age
	<fctr>	<fctr>	<dbl>
1	1st	Allen, Miss Elisabeth Walton	29.0000
2	1st	Allison, Miss Helen Loraine	2.0000

*Worked example effect:*

For conceptual questions (in Biostatistics) and for programming questions, we included a worked example in every assignment. For each worked example, we also highlighted the different parts that must be completed to do a question correctly.

Conceptual question:

### From Whitlock and Schluter:

Ch 3: 15 The data in the accompanying table are from an ecological study of the entire rainforest community at El Verde in Puerto Rico (Waide and Reagan 1996). Diet Breadth is the number of types of food eaten by an animal species. The number of animal species having each diet breadth is shown in the second column. The total number of species listed is  $n = 127$ .

- Calculate the median number of prey types consumed by animal species in the community.

**First**, order your list of numbers from smallest to largest.

Luckily, this step has already been completed for us in the table. The lowest diet breadth is 1 and the greatest diet breadth is >20. Keep in mind that this is a condensed list as multiples of each diet breadth are counted up in the

frequency columns. If we wrote out this list with full, it would begin with twenty-one “1”s followed by eight “2”s and so on.

**Second**, determine how you will obtain or calculate your median number.

Because this list contains an odd number of observations, our median will be

equal to  $Y_{\left(\frac{[n+1]}{2}\right)}$  meaning that the median will be equal to  $Y_{64}$ .

**Third**, find the 64th number in your list. This will be your median.

Add together the multiples of each diet breadth starting at “1” until you reach 64.

$$21 + 8 = 29$$

$$21 + 8 + 9 = 38$$

$$21 + 8 + 9 + 10 = 48$$

$$21 + 8 + 9 + 10 + 8 = 56$$

$$21 + 8 + 9 + 10 + 8 + 3 = 59$$

$$21 + 8 + 9 + 10 + 8 + 3 + 4 = 63$$

$$21 + 8 + 9 + 10 + 8 + 3 + 4 + 8 = 71$$

From these calculators we know that numbers 64 - 71 in our list are “8”s. Therefore our median is 8 prey species.

Programming question:

**Q1** Using the “titanic” dataset mentioned above, create three histograms: one for “Passenger Class”, one for “Age” and one for “Survive”.

**First**, we need to read the “titanic” data frame into R. This is done using the following line of code:

```
> titanic <- read.csv(file = 'titanic.csv')
```

**\*\*Note**, after uploading your csv, you should see the item, “titanic”, under the “Data” heading in the “Environment” tab of the upper right panel of your R console. If you would like to view the data frame in a spreadsheet format, simply click on this item or employ the View() function in the following line of code:

```
> View(titanic)
```

**Second**, now that we have read our data into R, we can use “ggplot2” to create our histograms. Start by using the library() function to load this package onto your console.

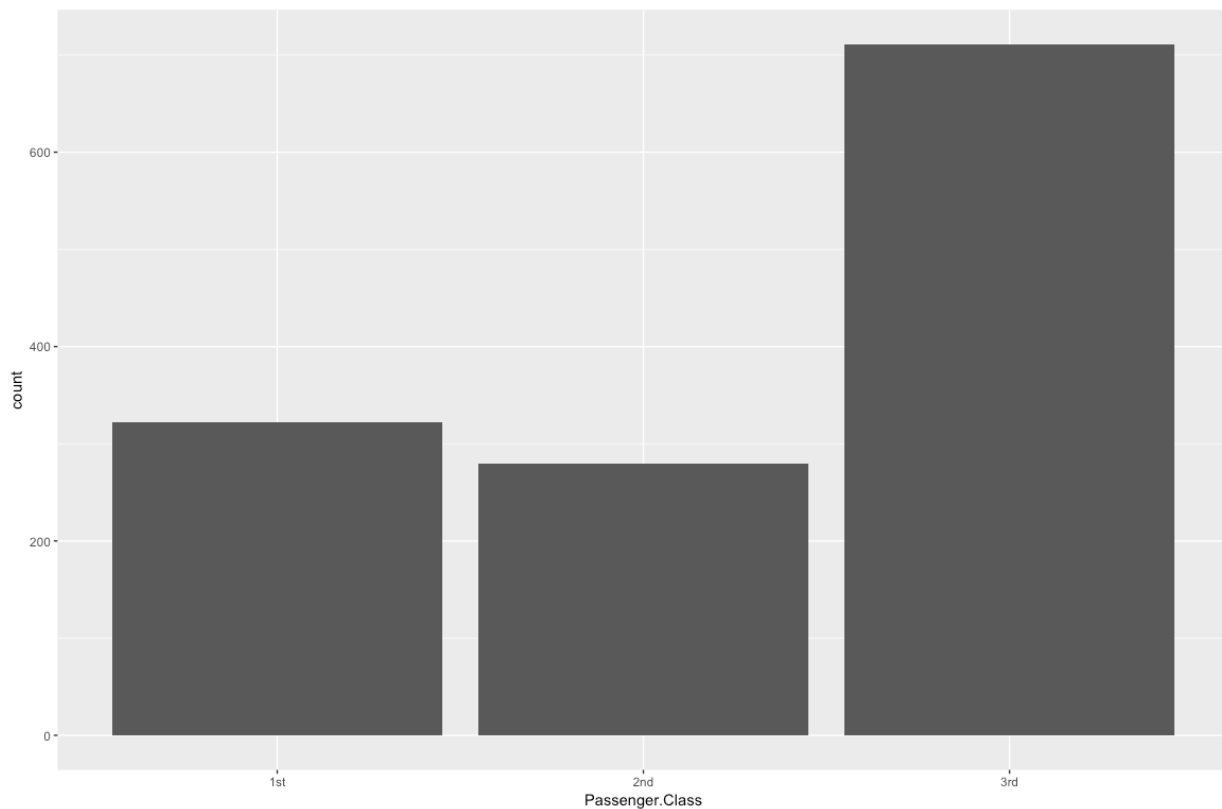
```
> library(ggplot2)
```

**Third**, construct your first histogram using the “Passenger Class” variable. This can be done using a single line of code. Please return to the previous sections if you do not remember what the different parts of the following code represent.

```
> ggplot(data = titanic, aes(x = Passenger.Class))+  
geom_histogram(stat= 'count')
```

**For you to think:** What kind of data is “Passenger Class”? Why do we need to include the ‘count’ statistic in this line of the code?

**Fourth**, view your graph. It should look like this:



*Partially completed problem:*

Similar to the worked example problems, we included partially completed questions in conceptual (Biostatistics) and programming questions. We scaffolded both the different parts to a question and the steps needed for each part of the question. As the assignment progressed, both types of scaffolding was reduced.

Conceptual question:

**From Whitlock and Schluter:**

Ch 3: 21 Researchers have created every possible “knockout” line in yeast. Each line has exactly one gene deleted and all the other genes present (Steinmetz *et al.* 2002). The growth rate--how fast the number of cells increase per hour--of each of these yeast lines has also been measured, expressed as a multiple of the growth rate of the wild type that has all the genes present. In other words, a growth rate greater than 1 means that a given knockout line grows faster than the wild type, whereas a growth rate less than 1 means it grows more slowly. Below is the growth rate of a random sample of knockout lines:

0.86, 1.02, 1.02, 1.01, 1.02, 1, 0.99, 1.01, 0.91, 0.83, 1.01

- a) What is the mean growth rate of this sample of yeast lines?

**First**, review the formula for calculating the mean of a list of numbers.

$$\text{mean} = \frac{\sum_{i=1}^n Y_i}{n}$$

**Finally**, use this formula to calculate the mean

- b) What is the median growth rate of this sample?

**First**, order your list of numbers from smallest to largest

**Second**, count how many observations are in your list.

**Finally**, based on the results from the previous step, determine what your median is.

Programming question:

**Question:** Construct a boxplot plotting the “growthexpt2” column against the “fertexpt2” column in the "fertilized\_block" data set.

**First,** make sure that you have ggplot2 installed by using the library() function.

**Second,** decide which column is your independent and dependent variable.

Independent: \_\_\_\_\_      Dependent: \_\_\_\_\_

**For you to think:** Can you explain how you made this distinction?

**Third,** construct your box plot using ggplot. Fill in the blanks in the following code to do so:

```
> ggplot(data = _____, aes(x = _____, y = _____))  
+ _____()
```

### ***(ii) Reducing the intrinsic load***

We reduced the element interactivity of the material by presenting (i) only one way to do one job and (ii) presenting only the functions that were needed for that assignment. In R, there are many synonymous ways to do the same task. For example, if we wanted to select one column out of a table these are the possible ways of doing it:

```
my_table$column1
```

```
my_table[,1]
```

```
select(my_table, column1)
```

We decided to choose only one option and consistently use it throughout the course.

Additionally, in every assignment the students learned only a handful of new functions, but throughout the course they gained practice in a diverse assortment of functions. For

example, they learned to input data into R, graph data, manipulate data and do multiple statistical tests.

***(iii) Increasing the germane load***

Both in worked examples and in partially completed problems we asked the students to reflect on a part of the question to engage in germane load activities such as self-explaining.

*Self-explanation questions:*

**For you to think:** What kind of variable is “fertexpt2” and what kind of variable is “growthexpt2”?



## Biostatistics:

1. What year of your undergraduate degree are you in?
  - 1
  - 2
  - 3
  - 4
  - 5+

Before this course started.

2. My programming skills (in any programming language) were
3. My skills in any statistical software (JMP, SPSS, etc) were
  - Excellent
  - Above Average
  - Average
  - Below Average
  - Very Poor
4. Before this course started I used R...
  - Daily
  - Weekly
  - Monthly
  - Yearly
  - Never
5. Before this course started, I was interested in learning a programming language
  - Strongly Agree
  - Agree
  - Neutral / Undecided
  - Disagree
  - Strongly Disagree

During this course I felt that the ...

6. statistics concepts were easy to understand (Normal distribution, t-test, ANOVA, regression)
7. programming concepts in R were easy to understand (Data input, visualization, writing code)
8. statistical software JMP was easy to understand
9. statistical software (R or JMP) sessions were helpful to understand the statistics

- Strongly Agree
- Agree
- Neutral / Undecided
- Disagree
- Strongly Disagree
- NA

During this course, the level of difficulty of was...

The concept refers to the material you saw in the lecture and the application refers to doing the analysis in the software (either R or JMP)

- 10. Probability (Concept)
- 11. Probability (Application)
- 12. Contingency analysis (Concept)
- 13. Contingency analysis (Application)
- 14. t-test (Concept)
- 15. t-test (Application)
- 16. ANOVA (Concept)
- 17. ANOVA (Application)
  - Too high
  - High
  - Right
  - Low
  - Too low

I felt ...

Mark all that apply

- 18. While working on the conceptual parts of the statistics assignments
- 19. While using the statistical software R
- 20. While using the statistical software JMP
  - Happy
  - Excited
  - Motivated
  - Supported
  - Overwhelmed
  - Anxious
  - Bored

- Frustrated
- Stressed
- Angry
- Annoyed
- Proud
- Scared
- NA

I used ...

- 21. R outside of class
- 22. JMP outside of class
- 23. R in the laboratories JMP in the laboratories
- 24. R in the lectures
- 25. JMP in the lectures
  - Daily
  - Weekly
  - Monthly
  - Once in the term
  - Never

Having completed this course I would ...

- 26. rate my programming proficiency as high
- 27. put the ability to use R as a skill on my CV
- 28. continue using R in my own projects for my undergraduate or graduate school
  - Extremely likely
  - Very likely
  - Moderately likely
  - Slightly likely
  - Not at all likely

Open response questions:

- 29. Do you have any other comments about the course?
- 30. If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

31. If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?

**Eco-Methods:**

1. What year of your undergraduate degree are you in?
  - 1
  - 2
  - 3
  - 4
  - 5+
2. Who was your instructor?
  - MKT
  - DSS

Before this course started.

3. My programming skills (in any programming language) were
4. My skills in any statistical software (JMP, SPSS, etc) were
  - Excellent
  - Above Average
  - Average
  - Below Average
  - Very Poor
5. Before this course started I used R...
  - Daily
  - Weekly
  - Monthly
  - Yearly
  - Never
6. Before this course started, I was interested in learning a programming language
  - Strongly Agree
  - Agree
  - Neutral / Undecided
  - Disagree

- Strongly Disagree

During this course I felt that the ...

7. statistics concepts were easy to understand (ANOVAS, regression, ordination, etc)
8. programming concepts in R were easy to understand (Data input, visualization, writing code)
9. statistical software JMP was easy to understand
10. statistical software (R or JMP) sessions were helpful to understand the statistics
  - Strongly Agree
  - Agree
  - Neutral / Undecided
  - Disagree
  - Strongly Disagree
  - NA

During this course, the level of difficulty of was...

The concept refers to the material you saw in the lecture and the application refers to doing the analysis in the software (either R or JMP)

11. Two way ANOVA (Concept)
12. Two way ANOVA (Application)
13. Randomized block ANOVA (Concept)
14. Randomized block ANOVA (Application)
15. Multiple regression (Concept)
16. Multiple regression (Application)
17. Ordination (Concept)
18. Ordination (Application)
  - Too high
  - High
  - Right
  - Low
  - Too low

I felt ...

Mark all that apply

19. While working on the conceptual parts of the statistics assignments

20. While using the statistical software R

21. While using the statistical software JMP

- Happy
- Excited
- Motivated
- Supported
- Overwhelmed
- Anxious
- Bored
- Frustrated
- Stressed
- Angry
- Annoyed
- Proud
- Scared
- NA

I used ...

22. R outside of class

23. JMP outside of class

24. R in the laboratories JMP in the laboratories

25. R in the lectures

26. JMP in the lectures

- Daily
- Weekly
- Monthly
- Once in the term
- Never

Having completed this course I would ...

27. rate my programming proficiency as high

28. put the ability to use R as a skill on my CV

29. continue using R in my own projects for my undergraduate or graduate school

- Extremely likely
- Very likely
- Moderately likely

- Slightly likely
- Not at all likely

Open response questions:

30. Do you have any other comments about the course?

31. If you could change anything about the way the statistical software (R or JMP) was taught, what would it be?

32. If I could keep anything about the way the statistical software (R or JMP) was taught, what would it be?