BIOL 3295: Population and Evolutionary Ecology, Winter 2023

Amy Hurford

2023-01-19

Contents

Chapter 1

Syllabus

1.1 Instructor Information

Instructor: Dr. Amy Hurford
Office: CSF 4338
Email: ahurford@mun.ca
I will try to reply to emails within 24 hours (excluding evenings, weekends and holidays). Office hours: Tuesday 1-2pm; Thursday 1-2pm

1.2 Course Information

TR 12.00-12.50pm
F 1-1.50pm
Classroom: SN3060 (unless stated otherwise on the schedule)

All Course Announcements will be made on BrightSpace. Should lectures be remote a WebEx link will be provided on BrightSpace.

Course description:
Population and Evolutionary Ecology is an introduction to the theory and principles of evolutionary ecology and population dynamics. Pre-requisites: BIOL 2600; at least one of BIOL 2010, 2122 or 2210.

Course format:
The course consists of lectures, 4 data analysis assignments, 2 exams and a final exam.

Course expectations:
Please attend lectures and respect the learning environment of other students. If you have COVID-19 please follow university and provincial public health

guidelines.

Learning goals:
The course content emphasizes a deeper understanding of fewer concepts. You
have seen much of the course material in pre-requisite courses. In this course, I
will revisit the models, clarify the assumptions and when they are appropriate,
and we will fit the models to data to estimate parameters. By the end of the
course, I hope that if you were given population data, that you would know the
key quantities that you might estimate, and could complete the analysis.

Required Text and Resources:
The course materials are online at https://ahurford.github.io/biol-3295-winter-
2023/index.html.

Most readings are assigned from two textbooks that are available electronically
from the library:

- Vandermeer, J.H., Goldberg, D.E., 2013. Population Ecology: First Prin-
  ciples (Second Edition). Princeton University Press, Princeton, United
  States. Link

- Otto, Sarah P., and Troy Day. 2007. A Biologist's Guide to Mathematical
  Modeling in Ecology and Evolution, Princeton University Press. Link

If you wish to use your own computer for assignments you should install `R` and
`RStudio` (see also here).

1.3   Method of Evaluation

- 4 Assignments - 20%

- 2 Exams - 40%

- Final Exam - 40%

Late assignments and missed exams, and final exams will be accommodated as
described by University Regulation 6.7.3 and 6.7.5 (see https://www.mun.ca/r
egoff/calendar/sectionNo=REGS-0474 for Regulations). Please discuss missed
assignments and exams with me. To accommodate the absence an assignment
may be modified or exempted and re-weighted in the grading scheme.

1.4   Additional Policies

1.4.1 Accommodation of students with disabilities

Memorial University of Newfoundland is committed to supporting inclusive education based on the principles of equity, accessibility and collaboration. Accommodations are provided within the scope of the University Policies for the Accommodations for Students with Disabilities see www.mun.ca/policy/site/policy.php?id=239. Students who may need an academic accommodation are asked to initiate the request with the Glenn Roy Blundon Centre at the earliest opportunity (see www.mun.ca/blundon for more information).

1.4.2 Academic misconduct

Students are expected to adhere to those principles, which constitute proper academic conduct. A student has the responsibility to know which actions, as described under Academic Offences in the University Regulations, could be construed as dishonest or improper. Students found guilty of an academic offence may be subject to a number of penalties commensurate with the offence including reprimand, reduction of grade, probation, suspension or expulsion from the University. For more information regarding this policy, students should refer to University Regulation 6.12.

1.4.3 Equity and Diversity

A safe learning environment will be provided for all students regardless of race, colour, nationality, ethnic origin, social origin, religious creed, religion, age, disability, disfigurement, sex (including pregnancy), sexual orientation, gender identity, gender expression, marital status, family status, source of income or political opinion.

You should not photograph or record myself, teaching assistants, or other students in the class without first obtaining permission. Accommodation will be made for students with special needs.

The sound should be turned off on phones and computers during class.

1.5 Additional Supports

Resources for additional support can be found at:

- www.mun.ca/currentstudents/student/
- https://munsu.ca/resource-centres/

Chapter 2

Schedule

All lectures are in SN 3060 unless otherwise stated

- Thurs Jan 5: Introduction
- Fri Jan 6: Population biology with discrete and continuous variables
- Tues Jan 10: —
- Thurs Jan 12: **CSF 2218** Introduction to Rmarkdown and tidyverse **Assignment 1 is assigned**
- Fri Jan 13: Geometric growth
- Tues Jan 17: Exponential growth
- Thurs Jan 19: **CSF 2218** Numerical solutions and graphing population data **Assignment 2 is assigned**
- Fri Jan 20: Exponential growth
- Tues Jan 24: Density dependence and logistic growth
- Thurs Jan 26: Density dependence and logistic growth **Assignments 1 & 2 are due**
- Fri Jan 27: Density dependence and logistic growth
- Tues Jan 31: Discrete time density dependence
- Thurs Feb 2: **EXAM I** (all material covered to date)
- Fri Feb 3: Age-structured models
- Tues Feb 7: Stage-structured models
- Thurs Feb 9: Stage-structured models
- Fri Feb 10: Stage-structured models
- Tues Feb 14: **CSF 2218** Numerical analysis of stage-structured models **Assignment 3 is assigned**
- Thurs Feb 16: Density dependence in stage-structured models
- Fri Feb 17: Metapopulation models

WINTER BREAK

- Tues Feb 28: Continuous space models **Assignment 3 is due**
- Thurs Mar 2: Spatially explicit models in population biology

- Fri Mar 3: Population dynamics in a warming world
- Tues Mar 7: Spatially explicit population dynamics in a warming world
- Thurs Mar 9: Disease dynamics
- Fri Mar 10: The net reproduction number
- Tues Mar 14: Overview of models in population biology
- Thurs Mar 16: **EXAM II** (All material since Exam I)
- Fri Mar 17: What is evolutionary ecology?
- Tues Mar 21: Haploid selection model
- Thur Mar 23: Selection coefficients for COVID-19 variants
- Fri Mar 24: **CSF 2218** Estimating selection coefficients **Assignment 4 is assigned**
- Tues Mar 28: The evolutionary ecology of pathogens
- Thurs Mar 30: The evolutionary ecology of COVID-19
- Fri Mar 31: The evolutionary ecology of hosts **Assignment 4 is due**
- Tues Apr 3: The evolution of reproductive effort in plants
- Thurs Apr 5: Evolutionarily stable and convergent stable strategies
- Fri Apr 6: Review

TBD **FINAL EXAM** (all course material)

Chapter 3

Jan 5: Introduction

- Survey of student computer preferences

3.1   Some questions

- What is a population?
- What are some definitions of a population that are given in textbooks?
- In research studies, how are populations discussed in the *Discussion*?
- How are individuals that comprise the sample selected in the *Methods* of a research study?
- List some potential differences between how populations are defined and discussed and the research methods?
- Why does the definition of a population matter?

3.2   References

Vandermeer, J.H., Goldberg, D.E., 2013. Population Ecology: First Principles (Second Edition). Princeton University Press, Princeton, United States. Link

The Princeton Guide to Ecology, edited by Simon A. Levin, et al., Princeton University Press, 2009. ProQuest Ebook Central, Link

Sacchi, R., Gentilli, A., Razzetti, E., Barbieri, F., 2002. Effects of building features on density and flock distribution of feral pigeons Columba livia var. domestica in an urban environment. Can. J. Zool. 80, 48-54. Link

Chapter 4

Jan 6: Discrete and continous variables

Reading: Otto, Sarah P., and Troy Day. 2007. A Biologist's Guide to Mathematical Modeling in Ecology and Evolution, Princeton University Press. Link **pages 33-38 in Section 2.3**

- Parameters versus variables
- Fitted versus independently estimated parameters

Chapter 5

Jan 12: ASSIGNMENT Rmarkdown and tidyverse

(Dates changed owing to the university closure for a snow day)

ASSIGNMENT 1 due Jan 26.

PART I is to reproduce a figure and the figure caption of a plot in Vandermeer and Goldberg 2013 or another textbook or a published paper in Rmarkdown and as an output file: .hmtl, .pdf, or .docx. Please choose a figure to reproduce in the area of Population Biology or Evolutionary Ecology.

You can simplify a complex figure if necessary. Your figure should be made in `ggplot()` and have:

- the title (if there is one),

- axes labels,

- points or lines or both,

- approximately the same data as the original figure,

- the correct axes limits.

The objective is for you to learn how to use Rmarkdown to make a synthetic write-up that includes code, a figure and text. Your completed output should have:

1. A brief text description of where I can find the figure you reproduced.
2. Code that makes a figure that is suppressed in the output file.
3. The reproduced figure (or simplified figure).
4. The actual figure

You are to hand-in the `.Rmd` file and an output file (`.html`, `.pdf`, or `.docx`)

Here, is an example of what a completed PART 1 looks like (as an .html output).

_____

The graph is Figure 3.1 from Population Ecology: First Principles - Second Edition (Vandermeer and Goldberg) on p67.

FIGURE 3.1 Difference in per capita egg production between the O lines and B lines from Rose and Charlesworth's (1981) experiment.

FIGURE 3.1. Difference in per capita egg production between the O lines and B lines from Rose and Charlesworth's (1981) experiment.

Instructions to complete PART I are here.

PART II of this assignment is to clean up messy data. As a biologist, much of my coding work involves getting data into the right format to use in functions, this might be a plot function like `ggplot()`, which you used in PART I, or a statistical function like `lm()`, which will perform a regression, t-test, or analysis of variance. The objective of PART II is to practice cleaning messy data into a useable format.

1. You are to clean the messy data from here enough so you can make a plot using `ggplot()`. Some helpful instructions for how to do this are here. The code that you write **must use the `select()`, `mutate()`, and `ggplot()` functions**. The graph that you make must be different than the example given in the instructions - for example, you might plot the same variables but for Plot 2.

(perhaps this sounds easy - this data is pretty messy - I found it quite hard!)

TO HAND IN

1. Hand in an `.Rmd` file and an output file (`.html`, `.docx`, or `.pdf`) with the solutions to both PART I and PART II of this assignment. Each part is 10 marks for a total of 20 marks.

Chapter 6

Jan 13: Geometric growth

6.1 Reading

Vandermeer, J.H., Goldberg, D.E., 2013. Population Ecology: First Principles
(Second Edition). Princeton University Press, Princeton, United States. **p1-3.**
Link

- Von Foerster human population become effectively infinite on Nov 13 2026

- Lilly pads replicate once per week. If it take a year for 1/2 a pond to
  be covered, when will it be completely covered? See also the wheat and
  chessboard

- $N_{t+1} = \lambda N_t$ equation 3 in Vandermeer and Goldberg. For what values of
  λ will the population size, N_t, grow?

- $N_t = N_0 \lambda^t$ is equation 4 in Vandermeer and Goldberg (but written more
  generally). With $N_0 > 0$ sketch a graph of N_t for different values of λ. If
  $N_0 = 0$, sketch a graph of N_t.

- If $N_0 = 1.1$ individuals per km^2, $\lambda = 2$, what is is the population size at
  time $t = 10$?

- Consider population growth of pheasants on Protection Island. If we were
  to apply the geometric growth model to the pheasant population, what
  are some assumptions? How might this affect our parameterization (i.e.,
  our estimate of λ) for the pheasant population?

- How can we understand what λ is in a population that has births and
  deaths?

Chapter 7

Jan 17: Geometric growth

7.1 Reading

Download the .pdf of the MSc thesis below and read the Abstract (the first two pages prior to the title page). Pay specific attention to the number of pheasants at different points in time, these might be N_{t+1} and N_t in the geometric growth model formula; and the number of births and deaths that occur, these may help you estimate λ in the geoemtric growth formula. Pay attention to the length of time that births and deaths are reported over, and what time of the year the population size is reported.

Newcomb, HR. 1940. Ring-necked pheasant studies on Protection Island in the Strait of Juan de Fuca, Washington. MS thesis. Oregon State University. [two pages prior to the title page]

Noteably,

    a. Pheasant chicks are born during the summer.
    b. In May 1937, 10 pheasants were introduced to the island. Before the next breeding season there were 35.
    c. November 10, 1938 a census estimated 110 pheasants.
    d. October 13, 1939 a census estimated 400 pheasants.
    e. Between the 1938 and 1939 censuses, Newcomb observed that 17 adult birds died.
    f. During the 1938 nesting season there were 5.86 eggs/nest. 83.57% of eggs hatched.
    g. During the 1939 nesting season there were 8.73 eggs/nest. 64.58% hatched.
    h. During the 1939 nesting season: Average number of chicks per clutch was 6.93.[1]
    i. You can assume the sex ratio is 50:50 male to female. Pheasants are a sexually reproducing species.

[1] Note that g. and h. appear to be contradictory.

7.2 Questions

This approach is called independent parameter estimation because we will estimate the birth and mortality rates independently of the population size data for different years.

1. $b > 0$ is the per capita number of births each year. The estimation of b for a geometric growth model is more subtle. First, b is estimated as the number of births (occurring between t and $t + 1$), divided by the number of individuals that could have given birth, N_t. You might average this value across multiple years if sufficient data are available. Furthermore, to correctly project the future population size, we should consider what we have assumed about survival of the pheasant chicks, given the time step of our model. **Given a.-i. estimate b. Write down any assumptions you have made.**

2. Is the probability that pheasants survive from one time step to the next. Estimate d.

3. What is the value of λ given your estimate of b and d from previous questions? Is this population is expected to grow over time?

4. Lets assume that the pheasant population on Protection Island grows geometrically (i.e. exponentially but for a discrete time model) where the geometric growth rate, λ, is the value that you estimated in question 3. Let $N_0 = 10$ and let t be the number of years since May 1937. Recall that when a population grows geometrically,

$$N_t = N_0 \lambda^t$$

Use the formula and your answer to 3. to predict the number of pheasants in May 1938, May 1939, May 1940, and May 1950.

Chapter 8

Jan 19: ASSIGNMENT Estimating the geometric growth rate

ASSIGNMENT 2 due Jan 26.

You are to write an Rmarkdown report that estimates the geometric growth rate, λ, for the Protection Island pheasant population. Your report needs to consider two methods for estimating the geometric growth rate: fitting; and estimation from independent data.

The code you will need to complete this assignment is below. You may copy and paste this code into your Rmarkdown report. Sometimes you will need to change the values.

The requirements of the report follow the code.

8.1 Code

8.1.1 Loading and plotting the data

Load the data. I copied the data on to a website, so it can be loaded with the command below. Click here to view the code on the website.

```
data = read.csv("https://raw.githubusercontent.com/ahurford/biol-4605-data/main/data/protection-i
```

Let's plot the data. We select only columns 1 and 2 of the data because column 3 contains comments. We plot `year` on the horizontal (x-) axis and `size` on the (y-) vertical axis.

```
require(ggplot2)
protection.island <- data[,1:2]

g1 = ggplot(data = protection.island, aes(x = year, y = size)) +
  geom_point() +
  xlab("year")+
  ylab("population size")+
  ggtitle("Protection Island")
g1
```


8.1.2 Function for geometric growth

Below is the definition of a geometric growth function with time, t, specifically defined for the number of years of Protection Island pheasant data, and the initial population size, N_0, defined specifically for the Protection Island data. You need to give this function in your code *before* you call the function.

```
geo.pred <- function(lambda){
  t = protection.island$year - protection.island$year[1]
  N0 = protection.island$size[1]
  size = N0*lambda^t
  pred = data.frame(year = protection.island$year, size = size)
}
```

The function is called by running `geo.pred(lambda)` in the console, where you enter a specific numerical value for `lambda`.

After you have run the code (i.e. in the Console) that defines the
`geo.pred(lambda)` function (above), try `lambda = 3` as:

```
result = geo.pred(3)
result
```

```
##   year size
## 1 1936   10
## 2 1937   30
## 3 1938   90
## 4 1939  270
```

Remember, negative values of `lambda` are non-sensical. (*For what values of*
lambda (or λ*), does the population grow, again? Not sure? You can try some*
values in the `geo.pred(lambda)` function to figure this out by experimentation.)

Question 1 The `geo.pred()` function works by running the lines of code inside
the `geo.pred()` function definition for the value of `lambda` that you supply
(inside the parentheses in the function call). In the function definition, what
are the lines of code that are pasted below doing? (Hint: What is the data
frame `protection.island`? Note that `data[1]` selects the first value of a list
of values and the `$` selects a particular column of a data frame).

```
t = protection.island$year - protection.island$year[1]
N0 = protection.island$size[1]
```

8.1.3 Fitting lambda

We have defined a function that will predict the population size of pheasants on
Protection Island for different user supplied values of `lambda`. But what value
of `lambda` is most likely given the data?

To answer this question we will use a statistical method known as maximum
likelihood.

Our first step is to define a function that quantifies the fit of a given `lambda`
value. This function assumes that deviations of the recorded data from the
model-predicted values follow a Poisson distribution:

```
geofit <- function(lambda){
  pred=geo.pred(lambda=lambda)
  Ypred = pred$size
  -sum(dpois(protection.island$size, Ypred, log=T))
}
```

After running the `geo.fit(lambda)` function (in the Console), lets try to use
the function and get some values of the negative log likelihood (i.e. the fit):

```
geofit(3)
```

```
## [1] 41.64846
```

```
geofit(1)
```

```
## [1] 1280.129
```

This result tells us that given the data, `lambda = 3` is much more likely than `lambda = 1` because the negative log likelihood value (41.65) is much smaller.

But what value of `lambda` is most likelihood given the data? Equivalently, for what value of `lambda` is the negative log likelihood minimized? To answer this question we need to call a function that will perform an optimization. This requires the `mle2` function from the `bbmle` package, and you will need to install this package prior to using this function.

```
library(bbmle)
```

```
## Loading required package: stats4
```

```
fit.geo <- mle2(geofit, start=list(lambda=3))
summary(fit.geo)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = geofit, start = list(lambda = 3))
##
## Coefficients:
##        Estimate Std. Error z value      Pr(z)
## lambda 3.408974   0.053472  63.752 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 24.32396
```

```
confint(fit.geo)
```

```
##    2.5 %    97.5 %
## 3.304256 3.513845
```

The output above tells us that the maximum likelihood estimate of `lambda` is 3.41 and that the 95% confidence interval is [3.30, 3.51].

8.1.4 Plotting the fit

Finally, we would like to use our estimate values of `lambda` in the geometric growth function and compare the fitted values with the observed data.

In the code below, `lambda = 3` is the estimated lambda value, and the 95% confidence interval is [2,4] (as example values). To use this code for your assignment you will need to substitute different values.

```
fit.predictions = geo.pred(3)$size
lower.fit = geo.pred(2)$size
upper.fit = geo.pred(4)$size
```

We had already made a plot of the data and we named our graph `g1`. We can now add some more layers to the graph as shown below. Note that the line is the value of `fit.predictions` as defined above (i.e., set `=3` as an example), and the shaded ribbon spans from `lower.fit` to `upper.fit` (i.e., set to `2` and `4` as an example).

```
g2 = g1 +
  geom_line(aes(y=fit.predictions)) +
  geom_ribbon(aes(ymin = lower.fit, ymax = upper.fit), alpha = .2)
```

```
g2
```
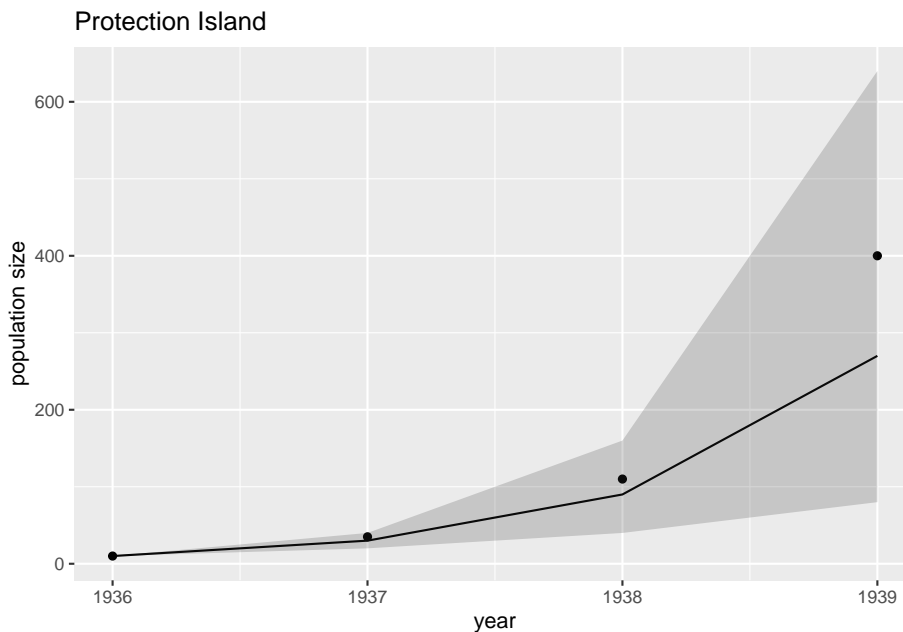

(The value of `alpha = 0.2` makes the ribbon transparent)

8.2 To hand in

1. Answer Question 1 which appears in bold in the *Function for geometric growth* section. [3 marks]

2. Write an Rmarkdown report that estimates λ (i.e. `lambda`) in the geometric growth rate function using maximum likelihood fitting.

You must include a graph that shows:

- The data (shown as dots);

- The predicted values for the maximum likelihood estimate of `lambda` (shown as a line);

- The predicted values for the 95% confidence interval for the estimate of `lambda` (shown as a shaded region);

You must include a figure caption that explains the main point of your graph, and what the symbols are.

Your Rmarkdown file must contain a sequence of R commands that produces the graph. [10 marks]

3. Estimate λ from data independent of the time series describing the population size of pheasants on Protection Island. This means that you cannot use more than one value of the population size to estimate one value of a quantity (you should be using the population size in a given year to calculate per capita rates by dividing only). You goal is to keep your parameter estimation method independent from the time series of pheasant population size so that you can *validate* the assumptions of your geometric growth model. In your Rmarkdown report you should:

- Derive a formula for λ in terms of b, the per capita birth rate, and the probability of mortality d each year. Do this based on the class discussion on January 17 or by reading Otto and Day, 2007 *Section 2.5.1 Discrete-Time Models* p47-50 and omit migration (i.e. $m = 0$) for the application to the Protection Island pheasant population. [2 marks]

- State some important assumptions of your λ formula. [3 marks]

- Estimate b and d using some of the information given here. State assumptions that you have made when data is lacking or information is contradictory. Calculate λ given the formula you derived. [4 marks]

- Use the `geo.pred(lambda)` function to predict the population size of pheasants on Protection Island for your independently estimated λ value. The code to do this must be included in your Rmarkdown file. [2 marks]

- Make a plot of your predictions relative to the reported data. Include a figure caption that describes the main point of your figure and defines all the symbols. [3 marks]

PART I - Instructions

For general instructions on installing `R`, `RStudio` and installing packages see the Quantitative Training Manual.

- Install the `Rmarkdown` package and all dependencies.
- Install `tinytex`. In the past we have had some problems with this on PCs. If your `tinytex` installation fails, what you might try is a package manager for Windows, i.e. `Chocolatey` or `Scoop`. See here for details. You are unsuccessful at installing `tinytex` that is okay, this package is only necessary to produce a .pdf output. You can complete your assignment as a .docx output or .html output.

Why use R markdown?

- Integrate code and write-up to avoid mistakes moving between `.R` (or other software) for analysis and `.docx` for write-up.

- It is easier to find all your work when everything is in one file (or linked to from one file).

- Run code in the background of your write-up so that if something changes the write-up automatically updates in all the relevant places. The reduces the chances of errors in your write-up.

- Publish your work as a website. This facilitates hyper-linking, you can update your work at any time, avoiding emailing your work keeps email storage free, and your work can be easily shared (i.e., in conversation I might say 'that analysis is linked off my faculty website').

- Include math symbols quickly because your hands don't leave the keyboard to make selections from drop-down menus.

- If your analysis is time-consuming you might not want the calculations in your write-up, slowing the compilation of your write-up. In this case you might have a separate `.R` analysis file that outputs your results as a `.csv` or plot. You can read these in automatically to your write-up by specifying the path to the `.csv` or plot.

1. In `R Studio`, select `File > New Project... > R Markdown`. Give the file a name, etc.

2. The default `.Rmd` opens already with some code to help you. With the default `.Rmd` opened, there should be a `Knit` button at the top and center of the Editor pane. Click the `Knit` button to knit to `.pdf`, `.html`, or `.docx` output. Alternatively, do `Cmd/Ctrl + Shift + K`.

(If this did not work, perhaps you have not installed the rmarkdown or tinytex packages)

3. Beside the `Knit` button is an arrow. You have the option to knit to `.pdf`, `.html`, or `.docx` output. Try producing other outputs.

(For me, producing a .docx opened Skype (clearly a bug). This was fixed by using Finder (on my Mac) to find the .docx file that I made, and selecting Open With > Microsoft Word)

4. Below are some things to try, that will help you to complete PART I. Type the code, then `Knit` to see what happens.

- Include variables in-text by enclosing in `$x$`, i.e. this renders as x, which is italicized to indicate in your writing that x is a variable rather than a letter.

- Load data using R commands. *(If you want to do this quickly copy and paste the command at the end of this section)*

- Hide the code that loads the data in the output. i.e., read about the options for r code chunks: `echo`, `include`, `message`, `warning`, `eval`, and `results`. Print the data in your output. Show both the code and the output. Try it all!

- Show only your code print out. Can you do this?

```
##   Psoil Pcorn
## 1     1    64
## 2     4    71
## 3     5    54
## 4     9    81
## 5    13    93
## 6    11    76
```

- Include code in-text as `` `r x ` ``. This renders as 80 because in a hidden coding block I loaded data and assigned `x<-mean(data$Pcorn)`. Therefore, the reported value of $x = 80$ is the mean phosphorous in the soil for the data I loaded in the background. If the data change, the mean reported in this document will automatically change too.

- Make headings, subheadings, bold font, etc.

- Make a hyperlink.

- Use latex commands to include in-text equations, i.e., `$y = \beta_0 + \beta_1 x$` renders as $y = \beta_0 + \beta_1 x$; `$\frac{dy}{dt} = e^x$` renders as $\frac{dy}{dt} = e^x$. *(You may need to type `require(tinytex)` in the Console to get this to work. 'The website Detexify is fun for identifying latex commands for different symbols (Some advanced symbols may require packages that you haven't installed and therfore won't work)).*

- Try some more complicated Latex.

```
data <- read.csv('https://raw.githubusercontent.com/ahurford/biol-4605-data/main/data/corn.csv',
```

- If you would like a more structured introduction to R Markdown you can read R Markdown: the definitive guide.

- This R markdown cheat sheet is helpful.

- Some more advanced skills you might learn are making alert boxes, or changing some of the options in the YAML. The alert boxes in this document are made as `div class="alert alert-info"` between `< >`, then the text, and closed with `/div` between `< >`.

- My experience making tables in `.Rmd` has not been good. Usually, I make the table in `.docx`, print to `.pdf`, take a screenshot and import the `.png` to `.Rmd` or `.tex`.

5. For your PART I specifically, you need to make a figure in `ggplot`. For `ggplot` you need your data as a data frame. The code that I used in the example was:

```
age = seq(1,25)
difference = c(-20, -26, -24, -20, -22, -20, -10, 0,
               20, 23, 9, 5, 12, 14, 16, 10, 8, 6, 7, 11, 5, 6, 4, 5, 10)
data = data.frame(age=age, difference = difference)
```

As you can see, I have guessed the values in the plot and entered them manually. This is okay for the purposes of completing your assignment. (Extra for experts - try a package like `scrapR` or `digitize`).

6. You need to install `ggplot2`. You also need to load that package because we are going to use functions from it now (do this as `require(ggplot2)`). The code that I used to make my `ggplot` was:

```
g1=ggplot(data = data, aes(x = age, y = difference)) +
  geom_point() +
  geom_line() +
  xlab("Adult Age, Days")+
```

```
  ylab("Difference")
g1
```
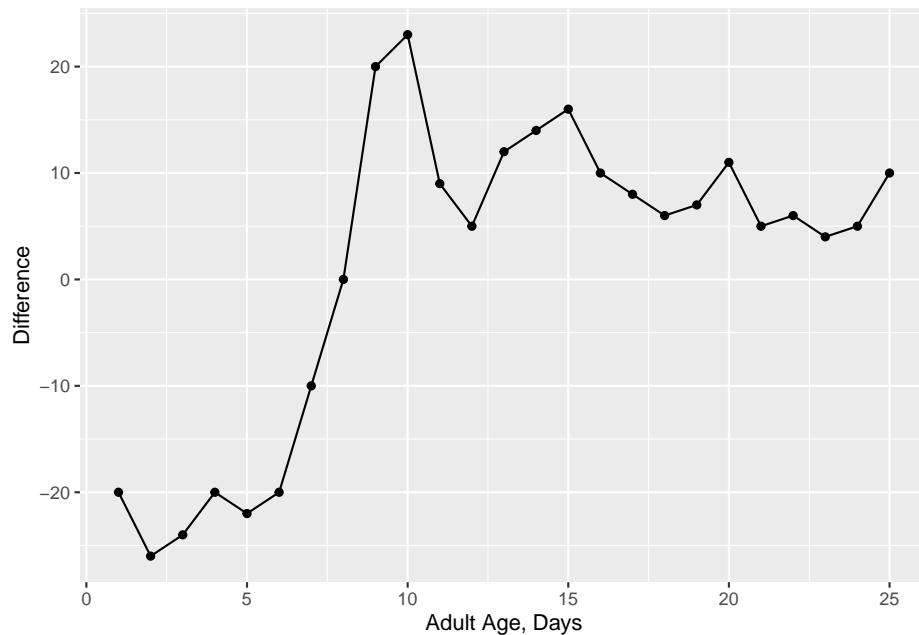

If you need to add a title, you can add a layer `+ ggtitle("Your title)`, to control the axis limits, use `+ xlim(c(-10,10))` (with the values you need). Generally, you should be able to use an internet search to find what you need. You can also read more about `ggplot` here.

7. To make your figure caption, just type below where your figure prints. To get bold text, use ** bold text ** (but without space between the ** and the text).

8. The last thing we need is to take a screenshot of the figure you are trying to reproduce, and to include it as a figure. I like to put all my figures in a folder named figure. You can read about including a figure that is a `.png` or other format. Of the options, I find using knitr::include_graphics() within a code chunk best because it seems easier to control the figure size. The code that I used in my example of how I completed PART I was:
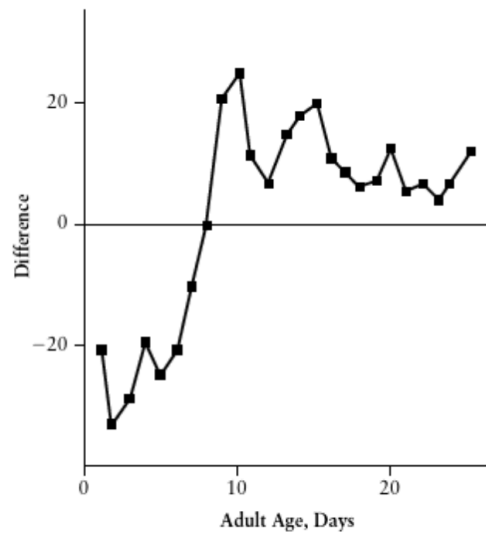
FIGURE 3.1. Difference in per capita egg production between the O lines and B lines from Rose and Charlesworth's (1981) experiment.

Now you have all the information you need to complete PART I, you just need to put the pieces together. If you are stuck, ask me or a classmate.

PART II - Instructions

- Install the `dplyr` package, all dependencies, and load the package.

- For instructions to clean data using tidyverse see here

First you need to load the messy data from here.

It may be helpful to view the data in Excel to understand what it looks like before you import it to `R`.

Use `head(data)` in the Console, or `data` to view your data (where `data` is the name I gave my data).

```
head(data)
```

```
##                    X          X.1        X.2       X.3    X.4 X.5              X.6
## 1                                                             NA
## 2 Data for Site 7                                             NA
## 3                                                             NA
## 4        Plot: 1                                              NA          Plot: 2
## 5  Date collected       Family      Genus  Species Weight  NA Date collected
## 6        01/09/14 Heteromyidae Dipodomys merriami     40  NA        01/08/14
##           X.7     X.8       X.9   X.10 X.11           X.12                X.13
## 1                                      NA
## 2                                      NA
## 3                                      NA
## 4                                      NA        Plot: 3
## 5     Family    Genus  Species Weight  NA Date collected             Species
## 6 Cricetidae Neotoma albigula   -999  NA            1/8 Dipodomys ordii*
##     X.14
## 1
## 2
## 3
## 4
## 5 Weight
## 6     42
```

These data are very messy indeed! A helpful command is to know the column

names:

```
colnames(data)
```

```
##  [1] "X"    "X.1"  "X.2"  "X.3"  "X.4"  "X.5"  "X.6"  "X.7"  "X.8"  "X.9"
## [11] "X.10" "X.11" "X.12" "X.13" "X.14"
# Extract a row using tidyverse commands
dataX =select(data, "X")
# This is base R syntax to extract specifically rows 6 to 14
date.collected = dataX$X[6:14]
# as.Date() is needed for R to treat this variable as a date
date.collected = as.Date(date.collected, format = "%m/%d/%y")
# This prints to the output, so you can see what I have done
date.collected
```

```
## [1] "2014-01-09" "2014-01-09" "2014-01-09" "2014-01-09" "2014-01-20"
## [6] "2014-01-20" "2014-03-13" "2014-03-13" "2014-03-13"
```

I will aim to make a data frame with "date collected" and "weight" for Plot 1.
Inspecting the data, weight is `"X.4"` for Plot 1.

```
weight = select(data, "X.4")
# as.numeric() is needed because otherwise R doesn't recognize these data as numbers -
weight = as.numeric(weight$X.4[6:14])
# Make this into a data frame so I can plot using ggplot
cleaned.data = data.frame(date.collected, weight)
# add a column that is a mutated column
cleaned.data = mutate(cleaned.data,weight.kg = weight/1000)
# print the cleaned data so we can see what it looks like
cleaned.data
```

```
##   date.collected weight weight.kg
## 1     2014-01-09     40     0.040
## 2     2014-01-09     36     0.036
## 3     2014-01-09    135     0.135
## 4     2014-01-09     39     0.039
## 5     2014-01-20     43     0.043
## 6     2014-01-20    144     0.144
## 7     2014-03-13     51     0.051
## 8     2014-03-13     44     0.044
## 9     2014-03-13    146     0.146
```

Now the data is in a format that I can make a plot:

```
require(ggplot2)
g1=ggplot(data = cleaned.data, aes(x = date.collected, y = weight.kg)) +
  geom_point() +
  geom_line() +
  xlab("Date collected")+
```

```
  ylab("Weight in kg")
g1
```