

Integrating the use of RStudio in Field-Based Courses: Creating Human Survivorship Curves

Allyson Degrassi

Shenandoah University, Departments of Biology and Environment and Society, 1460 University Drive Gregory Building, Winchester, VA 22601, USA
(adegrass@su.edu)

Many ecology courses are primarily field-based and students may miss an opportunity to learn the necessary data analysis, scripting, and manuscript writing skills used by many professional ecologists. The purpose of this laboratory exercise is to demonstrate the techniques needed to analyze changes in human survivorship as a tool to make critical observations regarding human demography. Students used vital statistics collected from tombstones to create Type I survivorship curves. Students incorporated age-structure models to calculate the generation time, intrinsic rate of increase, and net reproductive rate of the population using RStudio. Students learned to 1) collect and use data, 2) input data into a CSV file using Microsoft Excel, 3) analyze data and calculate the generation time, intrinsic rate of increase, and net reproductive rate in RStudio, and 4) produce graphs and tables reflecting a Type I survivorship curve. This 3-hour laboratory exercise was conducted by undergraduate Juniors and Seniors majoring in Biology, Ecology, and Environmental Studies after several ecology labs using RStudio.

Keywords: human demography, R Statistical Software, R Markdown

Link To Supplemental Materials: <https://doi.org/10.37590/able.v42.sup26>

Introduction

Many ecology courses are primarily field-based and students may miss an opportunity to learn the necessary data analysis, scripting, and manuscript writing skills used by many professional ecologists. Through data collection, computer modeling, and simulation, the integration of R Statistical Software and field-based laboratory exercises, instructors improve their demonstration of the scientific process and better prepare students for graduate school and/or careers involved with data collection and analysis.

It is essential to integrate ecology with human studies, as humans have directly or indirectly impacted almost every ecosystem on earth and the planet as a whole. By combining these two studies,

factors that influence human demography and ecological processes can be better understood. The purpose of this laboratory exercise is twofold: 1) to demonstrate the techniques needed to analyze changes in human survivorship as a tool to make critical observations regarding human demography, and 2) to introduce the use of R as a tool for data analysis.

In this exercise, students learned how to 1) collect demographic data in the field, 2) transpose and organize data from collection sheets into useable long-format data in spreadsheet CSV file, 3) write script to calculate survivorship curves, 4) make graphical representations of Type I Survivorship curves using R Studio, and 5) write a full manuscript-style report using R Markdown. In particular, students described age-structured populations, compare and explain differences in the survivorship curves and life-

tables in a community with respect to implied gender and historical trends, and incorporate age-structure to estimate generation time, intrinsic rate of increase, and net reproductive rate of the local community.

This exercise has been most successful when conducted under particular circumstances. This laboratory exercise was intended for undergraduate Juniors and Seniors majoring in Biology, Ecology, or Environmental Studies and when students have some base knowledge and practice with introductory laboratory exercises in Excel, RStudio, and R-Markdown, which are highly recommended before

beginning this exercise. Overall, there was very little preparation time required for this exercise if the instructor intends to use the lab manual (Appendix A), RStudio script, and the RMarkdown student report template. If the associated materials are utilized, the only preparation prior to the data collection would be uploading associated files for student use. If the instructor does not use the associated material, I would allocate approximately 10 hours to write a manual and to produce code/script.

Student Outline

Objectives

A complete laboratory manual with purpose, scenario, directions, definitions, R Script, and R Markdown template are given to students prior to data collection. Please see the laboratory manual (Appendix A) for full directions and scripting directions as shortened version is presented here.

The objectives of this lab are to:

1. Collect demographic data and learn to describe age-structured populations.
2. Compare and explain differences in the survivorship curves and life-tables in a community with respect to gender and historical trends.
3. Incorporate age-structure to estimate generation time, intrinsic rate of increase, and net reproductive rate of the local community.
4. Consider forces that shape human populations and how these compare to other age-structured populations.

Introduction

Today, you will be looking at local human demographics and population dynamics by collecting information from tombstones. Humans experience age-structured population growth, meaning birth rates and death rates depend on the age of the individual. In exponential and logistic growth models, we assumed that birth rates and death rates are constant with age. Now, we will incorporate age-structure to estimate generation time, intrinsic rate of increase, and net reproductive rate. From this information, you will be able to identify some basic population data as well as compare characteristics of age-classes or cohorts. For our graveyard study, our age-classes will consist of 10-year intervals, for which we will collect birth and death dates from tombstones in a local cemetery.

You will investigate the hypothesis that there are differences in survivorship of individuals who lived in the 19th century (1801-1900) compared to individuals who lived in the 20th century (1901-2000). You will also examine the data to see if sex (assumed by gender) plays a role in survivorship both as a whole and between the two time periods. Please note, *sex* is a biological concept regarding gametes at birth, while *gender* is a social concept and refers to the social and cultural differences a society assigns to people based on their sex or an individual assign themselves. We will be categorizing individual's gender based on their names only as we do not have information about their biological condition.

Background

Demography is the descriptive study of biological populations, including human populations. Demography includes the study of birth rates, death rates, immigration, and emigration patterns (sometimes referred to as vital statistics). Ecologists have perfected the analysis of demography because the vital statistics of a population tell us if the population is growing or declining, what will be the stable age distribution of the population, and how natural selection might work on individuals of each age class.

Human population dynamics can be particularly interesting because of how events or trends in human history affect them. Everyone has heard the term "Baby Boomers", which refers to the generation that was born after WWII when the economy was strong and soldiers returning from the war were getting married or were reunited with their spouses. This socio-political-economic event is observable in records of human populations via increased birthrate, life expectancy, increased total population, and changes in sex ratios in countries affected by the war. Other global events that are evident in population data are the flu pandemic of 1918, which killed between 50 and 100 million people in only 18 months. In the US, this was followed shortly by the Great Depression, which added to the dip in population growth.

There are many such events in human history that have impacted populations, some global and some local. An event that changes a particular population could be as seemingly insignificant as the closing of a mine in a small West Virginia town, closing of a tobacco farm in Virginia, or increasing housing prices in inner city Boston. Advances

in healthcare have also significantly altered demographics in populations. In the early 1800s, one of every five women died during childbirth, and infant mortality was much higher than it is today. Fortunately, childbirth is not nearly as fatal today. These changes have been at least partially attributed to medical breakthroughs such as the discovery of antibiotics in the 1930s and the polio vaccine in the 1950s.

Survivorship Curves

Survivorship curves are representations of population mortality over a lifespan. They illustrate the proportion of an initial cohort that is alive at age x . There are three common forms that these curves can take that represent different life history strategies. It is easy to forget that demography includes all biological populations, not just humans.

1. **Type I survivorship** is characterized by relatively low mortality rates for young age classes, but high mortality rates for old age classes. Species that follow this type of curve often have few offspring and provide parental care for the younger age classes. Large-bodied animals and ones in the higher trophic levels, such as lions, whales, and apes, often exhibit Type I survivorship. In developed nations with good health care, human populations also have survivorship curves that appear most like Type I. Survivorship curves are often described as properties of a species, but they actually depend on the local conditions in a particular environment. In many places in the world, human infant mortality is still high, and the survivorship curve will not look like an idealized Type I curve.
2. **Type II survivorship** is characterized by nearly constant mortality over all age classes, leading to a linear decline in survivorship with age. This type of curve is fairly uncommon in nature, because most species and populations exhibit some differences in mortality rates as they age.
3. **Type III survivorship** is characterized by relatively high mortality rates for young age classes, but low mortality rates for old age classes. In this kind of life history, organisms produce many offspring and provide little or no parental care. As a consequence, mortality in young age classes is very high, and very few offspring will survive to reproduce. Some examples of these types of species might include organisms such as octopi, which can lay up to 200,000 eggs. Many plant species also exhibit Type III survivorship curves. Maple trees, for example, can drop thousands of seeds of which only a handful may ever make it to reproductive maturity. However, once they get past the recruitment stage, the survivorship rate of large old trees is relatively high.

Graveyards

Graveyards are interesting in themselves because they reflect social change and attitudes toward death. Prior to the Revolutionary War era, many European and New England gravestones were decorated with grinning skulls and cross-bones, stressing the inevitability of death and the relatively brief time of life. Cemeteries were muddy and unkempt, reinforcing this grim message. By the late 1700s, death was viewed more as a release from earthly suffering, and white marble, a symbol of purity, became common for grave markers. They were often shaped like headboards of beds and the size indicated the age of the deceased.

In the 1830s, cemeteries began to be run by towns rather than churches and it became necessary to buy plots. Grave markers became more ornate, and, for wealthy families, mausoleums were popular. By the late 1800s, family memorials had a central marker for the head of the family (usually an adult male), with smaller stones around it representing family members.

At this time, one out of five women died during childbirth, and many children died before the age of five. Thus, death in the immediate family was fairly frequent. Because cemeteries held those recently deceased individuals, families would often make weekend excursions to the cemetery and took along picnic lunches. The atmosphere of graveyards reflected such social patterns, and graveyards were landscaped and groomed for a park-like effect. During this century, the trend is for spouses to have equal and side-by-side headstones stressing the marital bond, rather than the supremacy of the husband.

I am sure you do not have to be reminded, but cemeteries are places of honor for those who went before us. Please remember to show respect for the dead and for any other people who may be visiting their loved ones in the cemetery while we gather the data. Please respect the privacy of any people visiting graves by avoiding those areas.

Methods and Data Collection

Part A: Location and Team Work

Today, we will be working in a local cemetery with 19th and 20th century gravestones. I have provided directions below.

Work in pairs to collect data in the area of the graveyard you are assigned. Record the year of birth, death, and assumed gender, and gravestone observation notes (e.g large, decorated, worn, difficult to read, etc..) from at least 50 individuals from each century cohort.

Part B: Standardized Data Collection

Each group will collect data from 50 gravestones; 25 from each treatment (the 19th century and 25 from the 20th century). Data collection best practices for this lab:

1. If the gender of the name is too ambiguous to infer sex, omit that individual.
2. Be sure not to focus exclusively on large tombstones, because tombstone size is influenced by wealth and selecting large (or small) tombstones will bias the data by introducing differences in quality of health care.
3. Choose data for individuals who lived primarily in one century or the other to avoid overlap of the two categories.
4. If the name or date on a gravestone is illegible, make a rubbing using blank paper and the side of a dark crayon. Rubbings can greatly enhance legibility.

Part C: Creating Your Working Directory and Preparing Data in Your csv. File

By now, you should be comfortable with creating a working directory for this project (recall 'setwd' function). Now, before you start analyzing your data, you first need to format it in a way that R can read it. For this lab, you need to input your data into Excel first. Please see associated '.csv' file. You should have 6 columns that **MUST BE LABELED THE SAME WAY OR THE SCRIPT WILL NOT WORK!** There is a template for you. Once you are finished entering your data, save your Excel document as a '.csv' file (Comma Separated).

Part D: Importing Data

Here you are importing your dataset from your '.csv' file to your R script. There are TWO csv files needed for this script.

1. The class raw data (we will name this csv file "survival"). This dataset has all observations you recorded from the tombstones (gender, year of birth, year of death, notes).
2. The birth schedule (we will name this csv file "birth") is provided as a download. This information was made based off of the historical census data, and provides a typical birth schedule (average number of offspring per female of each age class) for the 19th and 20th centuries.

Part E: Attach Primary Datasets in RStudio and Manipulate Datasets

You have seen before how you can extract a column from a dataset and save it as an object in R. This can be tedious when you have datasets with several columns. This also creates several objects in R that just use extra

space. R has a nice function called ``attach``. This function allows you to attach a 'dataframe' and then call each column later as if it was an independent object. At the end of your script, you can just ``detach`` your dataframe, and these columns will no longer be recognized as separate objects. You have already attached the two csv files ("survival" and "birth"), but we should recall them using the ``attach`` (). Now it is time to start working with the data you have collected from the Cemetery. We have a lot of calculations to do so make sure you follow the lab manual the RScript Chucks, and your textbook (Gotelli 2008)!

1. Calculate age of death (life span)
2. Calculate Cohort Survival (S_x)
3. Calculate Survivorship Schedule (l_x)
4. Calculate Survival Probability (g_x)
5. Calculate $l_x * b_x$
6. Calculate $l_x * b_x (x)$
7. Calculate Net Reproductive Rate (R_0)
8. Calculate Generation Time (G_x)
9. Estimate the Intrinsic Rate of Increase (r)
10. Plot Survivorship Curves

Part F: Plot Survivorship Curves

That was a lot of calculation, but now we have all the data we need to plot survivorship curves (see lab manual and RStudio Chunks). On the x-axis we will need the ages (Ages) and on the y-axis we need the logarithm of the elements in the Survivorship Schedule (l_x). We also need four lines on this graph (19th century females, 19th century males, 20th century females, 20th century males. First, we plot the data for the 19th Century Females with the ``plot`` () function. Then, we add the other lines with the ``point``() function.

1. Male Survivorship Curve in 19th century (blue dashed)
2. Male Survivorship Curve in 20th century (blue solid)
3. Female Survivorship Curve in 19th century (pink dashed)
4. Female Survivorship Curve in 20th century (pink solid)

Part G: Interpret Your Graph and Write Your Paper

Now it's time to ask questions that will help you write your paper (meaning, the following questions should be answered in your paper). You should check and follow the rubric for every paper before you start writing.

1. Compare generation time (G), intrinsic rate of increase (r), and net reproductive rate (R_0) of males and females from the 19th and 20th century.
2. What biological mechanism(s) may cause these values to differ?
3. Which survivorship schedule curve (Type I, Type II, or Type III) do your data most closely match?
4. Are the survivorship curves different between centuries?
5. Are the survivorship curves different between gender and what can you infer about sex?
6. What biological, historical, or cultural mechanism(s) could explain the patterns you observe in your data?

Part H: Report Guidelines

Abstract (~300 words): The abstract is the summary of the entire paper, so it is best to write it last. The abstract should address:

1. What is the primary purpose of the study?
2. What are the biological, alternative, and null hypotheses?
3. How were data collected?
4. What methods were used to analyze the data?
5. Was there a statistical difference between times or between genders?
6. Was the null hypothesis accepted or rejected?

7. What is the overall conclusion?

Introduction (1-2 pages): Here you should look up some papers on survivorship curves in different species. Give examples from all III curves. Discuss human demography in the US and what you expect to see from your sample. You need to also clearly state your biological, alternate, and null hypothesis in this section. Also include the purpose of this study. Hint: look at the questions above to help you search for papers.

Methods (~1 page): This will be a short section and should include the following:

1. Where was the study conducted?
2. Where were data collected?
3. How were data collected?
4. How were data analyzed?

Results (~ ½ - 1 page of text, 2 tables, 1 graph): In the results section, describe the patterns in the graphs and tables that you have created. Make sure that you specifically cite each table or graph and tell what the result is. Just report the results in this section. You will interpret them in the discussion section. What does your table of values show? What patterns are present for the survivorship curves of males and females in the 19th and 20th centuries?

Discussion (~1-2 pages): This will be a short section and should address the following:

1. What can you say about the demography the area during the 19th and 20th century between males and females?
2. Are you going to accept or reject your Null Hypothesis?
3. Why or why not?
4. What does your study say about human demography and using gravestones to collect this type of data?
5. What are the potential sources of error and uncertainty in your estimates of the survivorship and birth schedules?

Cited References

Gotelli, N. A Primer of Ecology. 4th Edition. Sutherland, MA: Sinauer Associates, Inc Publisher; 2008.

Materials

A computer with Internet access to download associated material is need. The lab manual (Appendix A), field data collection sheet (Appendix B), “birth” rate data csv, the student’s combined “survival” data collected from the cemeteries, R Statistical Software (<https://www.r-project.org/>), and the RStudio Integrated Development Environment Software (<https://www.rstudio.com/products/rstudio/>) will be downloaded prior to the start of the exercise. The above software and csv files are required for each student to have on their computer. All R software are opensource products and are free.

For data collected in the field, students will need either a printed version of their data collection sheet (Appendix B), a clipboard, and writing utensils or they will need to download the field data collection sheet as a pdf and write on their tablets.

Notes for the Instructor

Students do not always see how the end product of a manuscript is connected to field data collection and data analysis at the beginning of the scientific process, especially when they are not introduced to the primary tool reasonable for analyzing data and displaying results. As field-based instructors, we face many challenges are responsible for helping students making those connections.

One challenge I see is with “canned” labs, because they are often designed to teach a specific skill (which definitely has its uses). The challenge to overcome is making sure students understand that the scientific process does not end after data collection. Rather, field data collection is one of the early steps in a long process of interpreting our data. Another challenge is remembering that students are here to learn skills that will aid them in securing a position in the workforce or graduate school upon completion of their undergraduate degree. Staying current and teaching students to use discipline appropriate tools (such as R and R Studio in ecology and environmental science) will not only give them an extra line on their CV, but may also influence potential employers and graduate advisor take a second look at our students.

The most time-consuming part of this exercise is the scripting. Students that pay attention and follow the lab manual instructions and scripting directions, usually do not require additional help.

However, some students, and I believe they are the minority of the class, make easily correctable mistakes. The most common mistakes that require attention are small errors in the syntax (e.g., missing commas, parenthesis, etc....). By this stage in the course, most students are comfortable and competent in setting up their data which is often another place where errors occur.

There are four laboratory exercises involving the teaching of R, RStudio, and RMarkdown prior to this exercise. In each exercise, students learn new functions and scripting techniques that build on the previous labs, in addition to the scientific method process and ecological analysis. For example, the first laboratory exercise is “Introduction to R, Studio, and R Markdown” where the instructor describes the software, layout, and uses of the different software. The students are given script and instructions to follow. In the second lab, students create a working directory and are given data that they must input into as a csv file and attach to R. This process is then repeated throughout all other labs. Each lab is made so that students keep building on their data management, scripting techniques, and ecological skills and knowledge. The rubric used for these laboratory reports are more-or-less a simple checklist to make sure they are reporting the elements of a manuscript. The rubrics and the graded elements change from a presence/absence style to content and demonstration of understanding the materials. Although I have not conducted an experiment to evaluate the effectiveness of this teaching and learning model, there is some evidence of its effectiveness on student engagement.

A study conducted in Florida on selected Junior level students showed that the integration of statistical software and use of “real data” provided a highly effective teaching-learning model for biology and statistics-based courses at university level (Gomez 2010). I have not yet collected official data on the student response to these exercises, but I will attempt to provide my interpretation of the student feedback regrading this lab and the others involving R.

In general, students seem to have a positive reaction to using R and RStudio. It is a very steep learning curve, but they seem to take pride in their work, especially when their graphs work! They find it very challenging, but also rewarding. Students have seen the benefit of using R even outside my classroom. I have more students approach me after class seeking help in applying R to other courses,

projects, and their research, than I have students openly discontent about using R. In my course evaluations, I usually have one or two students that say they do not see a need for it, but the majority of students like using RStudio and find it a skill worth learning. As a side note, I had multiple students say they believe their experience with R contributed to their success of being accepted into graduate school. Once they communicated with their potential graduate advisors that they were familiar with RStudio, the quality and enthusiasm of correspondence between them increased.

Cited References

Gomez R. 2010. Innovations in teaching undergraduate statistics courses for biology students. RHESL; 3/7:8-13.

Acknowledgments

Thank you to Nick Gotelli at the University of Vermont who contributed the backbone of this lab and encouraged me to create ecology labs using R and RStudio for his classes, and thank you to all of the BIO 321 students at Shenandoah University who have helped improve this laboratory exercise over the last two years.

About the Authors

Allyson Degrassi has been an Assistant Professor at Shenandoah University since 2019, where she teaches courses in ecology, wildlife management, and environmental science and sustainability.

Appendix A
Lab Manual – pdf

Appendix B
Example of Field Data Collection Sheet

Data Collection Sheet for Graveyard Lab									
Date: _____					Group: _____				
19th Century: 1801-1899					20th Century: 1901-1998				
Sample	Gender (F or M)	Birth Year	Death Year	Notes	Sample	Gender3 (F or M)	Birth Year	Death Year	Notes
1					1				
2					2				
3					3				
4					4				
5					5				
6					6				
7					7				
8					8				
9					9				
10					10				
11					11				
12					12				
13					13				
14					14				
15					15				
16					16				
17					17				
18					18				
19					19				
20					20				
21					21				
22					22				
23					23				
24					24				
25					25				

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit <http://www.ableweb.org/>.

Advances in Biology Laboratory Education is the peer-reviewed publication of the conference of the Association for Biology Laboratory Education. Published articles and extended abstracts are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board. Published abstracts are evaluated and selected by a committee prior to presentation at the conference.

Citing This Article

Allyson Degrassi. 2022. Integrating the use of RStudio in Field-Based Courses: Creating Human Survivorship Curves . Article 26 In: Boone E and Thuecks S, eds. *Advances in biology laboratory education*. Volume 42. Publication of the 42nd Conference of the Association for Biology Laboratory Education (ABLE). <https://doi.org/10.37590/able.v42.art26>

Compilation © 2022 by the Association for Biology Laboratory Education, ISBN 1-890444-17-0. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner. ABLE strongly encourages individuals to use the exercises in this volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.