

Department of Microbiology, University of Manitoba
MBIO7040 Module –
Statistical Analysis and Visualization of Data in R
Spring Term 2020

Course Details

Course Schedule: Wednesday, January 8 & Friday, January 10; 9:30-16:30
Location for Lectures: 311 Machray Hall
Course website: <http://home.cc.umanitoba.ca/~gersteia/MBIO7040/>

Instructor Contact Information

Instructor: Aleeza Gerstein
Office Location: 364 Machray Hall
Phone: (204) 272-1681
Email: Aleeza.Gerstein@umanitoba.ca
Office Hours: By appointment

General Goals for this Course

This module will be focusing on learning how to use the R Programming Language in R Studio for data analysis and visualization. It will serve as a refresher course for introductory statistics and visualization techniques, and provide students with some ‘best practices’ for data storage and file organization.

Textbook and Other Materials

Textbook: There is no formal textbook for this course. My course notes will draw in part from the following textbooks available online.

- Claus Wilke, Fundamentals of Data Visualization, <https://serialmentor.com/dataviz/>
- Chester Ismay and Albert Y. Kim, An Introduction to Statistical and Data Sciences via R <https://moderndive.com/>
- Susan Holmes and Wolfgang Huber, Modern Statistics for Modern Biology <http://web.stanford.edu/class/bios221/book/>

Online resources: The following are very useful online resources for R:

- <http://www.rstudio.com> (for downloading R Studio)
- <http://rmarkdown.rstudio.com> (to learn about R Markdown)
- <https://www.tidyverse.org/> (to learn about the tidyverse packages)
- <http://r-pkgs.had.co.nz/style.html>
(a simple R style guide that is generally good to follow)

Outline of Covered Topics

Topic	Learning Objectives
Spreadsheets	Implement best practices in data table formatting Identify and address common formatting mistakes
Introduction to R Studio	Describe the purpose of different window panes Organize files and directories for a set of analyses Use built-in interface to search for help
Introduction to R	Assign values to objects in R Learn how to name objects Solve simple arithmetic operations in R Inspect the content of vectors and manipulate their content Call functions and use arguments to change their default options
Starting with Data	Describe what the here package does Read in tibbles Extract values from tibbles Perform basic operations on columns in a tibble
Manipulating data with dplyr	Describe what the dplyr package in R is used for Apply common dplyr functions to manipulate data in R Employ the ‘pipe’ operator to link together a sequence of functions Employ the ‘mutate’ function Employ the ‘split-apply-combine’ concept Convert tibbles from long to wide and back again
Statistics bootcamp	Describe the difference between populations and samples Distinguish between data that follows and deviates from a normal distribution Define the difference between null and alternate hypothesis Define type I and type II error Describe the relationship between alpha, beta and power
Exploratory data analysis	Produce histograms, barplots, boxplots, and scatterplots using ggplot Describe what faceting is and apply faceting in ggplot Modify the aesthetics of an existing ggplot Build complex and customized plots from data in a data frame Be wowed by the power of R and compelled to keep using it after the workshop

Assignments, Test and Exams

Each student will pick a dataset that they have already collected in the lab, or ask a lab-mate or friend for a suitable dataset. They will then do a statistical analysis on this dataset and create a visualization that displays the information in a ‘publication quality’ figure. Everything will be recorded in an R Script that is properly commented at each step so that the analysis is reproducible. The goal of the assignment is that whatever is done should be actually useful.

Assignments should be emailed to Aleeza in the form of: 1) the dataset in excel/.csv/.tab format, 2) the R script that records all step and the results of the statistical analysis, 3) a pdf version of the figure.

Due date: Monday, January 20 at 5pm EST.

Grading rubric: Will be provided at the beginning of the module.

Grading timeline: Under normal circumstances, assignments should be graded a week after submission.

Late policy: No late submissions will be accepted.

Technology in the Classroom

This module requires all students to bring a laptop with them. It is the University of Manitoba policy that technology resources are to be used in a responsible, ethical and legal manner. Students should restrict their use of technology to those approved by the instructor *for educational purposes only*. Electronic messaging, email, social networking, gaming, etc. should be avoided during class time. Cell phones should be on silent. If a student is on call for emergencies, their cell phone should be on vibrate mode and the student should leave the classroom before using it.

Class Communications

The University requires all students to activate an official U of M email account, which should be used for all communications between yourself and the university (including all your instructors). All these email communications should comply with the University’s policy on electronic communication with students, which can be found at: http://umanitoba.ca/admin/governance/governing_documents/community/electronic_communication_with_students_policy.html

Copyrighted Materials: We will use copyrighted content in this course. The content I use is appropriately acknowledged and is copied in accordance with copyright laws and University guidelines. Copyrighted works, including those created by me, are made available for private study and research and must not be distributed in any format without permission. More details are available at <http://umanitoba.ca/copyright/>.

Lectures: No audio or video recording of lectures or presentations is allowed in any format, openly or surreptitiously, in whole or in part without my permission.

Academic Integrity

Students are encouraged to discuss course material, including the final project. However, each student must hand in his or her own copy of each assignment/project with personalized solutions, including comments, discussions, explanations and interpretations, and R code. Copying from anywhere, including other students, books, or the internet constitutes a case of academic dishonesty and could have serious consequences.

The goal in this class (as in all academic pursuits) is to learn. If you are unclear on what is acceptable, please ask me or visit the Faculty of Science page on cheating and plagiarism at: <http://umanitoba.ca/faculties/science/undergrad/resources/webdisciplinedocuments.html>

ROASS Schedule A

Schedule “A” of the *Responsibilities of Academic Staff with regards to Students* (ROASS) policies of the University of Manitoba lists resources and policies for students. It is important that you familiarize yourself with these resources and policies. This document is available from the Department of Statistics web page at: <http://umanitoba.ca/science/statistics/>.

University of Manitoba Acknowledgement of Traditional Territories

The University of Manitoba campuses are located on original lands of Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples, and on the homeland of the Métis Nation.

We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of reconciliation and collaboration.