

Course Syllabus – BIO 200U-J1 / CSC 200U-J1

Working Smarter, Not Harder: Using R to Visualize, Analyze, and Communicate Science

Instructor: John Egner

Contact: jegner77@gmail.com, 815-557-8470

Office Hours: M, T, W, R: 4:30 PM to 6:30 PM and by appointment, location TBD

Date: 01/09/2019 – 01/31/2019

Time: Monday – Friday, 1 PM to 4 PM

Place: Clausen Center 108

Credit: 4 credit

Examination: in-class and out-of-class exercises, in-class assessments (quizzes), and research project comprised of a written component and oral presentation.

Course Description:

Ninety percent of all data was collected in the past two years, at a rate of 2.5×10^{15} GB of data per day (DOMO). This age of big data has influenced several fields of science including biology, chemistry, and the psychological sciences. As a result, data scientists were ranked as the number 1 best job in U.S.A. in 2016, 2017, and 2018 (Glassdoor). The purpose of this course is to provide students with a strong foundation in R programming that will enable them to do data science. Upon completion of this course, students will be able to transform, visualize, model, and communicate (scientific) data. Important tools for data visualization, linear/non-linear regression, and model discrimination will be demonstrated through in-class tutorials and exercise. Most class sessions are comprised of discussion in combination with tutorials and assessments. Example datasets ranging from biology, chemistry, physics, pop culture, and history will be included. The culmination of skills learned in this course will be assessed through a research project comprised of a written and oral presentation component on a dataset of the student's choice with instructor approval.

Student Learning Objectives/Outcomes:

Students will gain critical and creative thinking skills and be challenged to think of solving problems as a scientist. Understanding and implementation of important concepts for data visualization, regression, and modeling will be assessed through in-class exercises, active participation, and weekly quizzes. A final research project will assess student's skills for using R in data science. The research project will consist of students selecting a unique dataset, visualizing relationships between variables, and communicating their findings through a written report (R script included) and an oral presentation.

- 1) Upon completion of this course, students will be able to learn and apply data science concepts to a wide array of problems/datasets and strengthen their abilities as a scientist.
 - a) Participate during in-class exercises and tutorials.
 - b) Complete out-of-class exercises (Homework).
 - c) Think critically of data and demonstrate ability to interpret different datasets.
 - d) Develop learning skills and approaches to troubleshooting problems.
- 2) Upon completion of this course, students will be able to communicate data relationships, scientific findings, and share R scripts.
 - a) Develop good coding practices and etiquette.
 - b) Produce a minimal reproducible example (reprex) of a section of R code.
 - c) Solve coding errors and problems with available resources.
 - d) Deliver an oral presentation on a research project

- 3) Upon completion of this course, students will be able to use R in data analysis and modeling.
 - a) Identify variables and observations that comprise a particular dataset.
 - b) Use appropriate methods for modeling and statistics for model discrimination.
- 4) Upon completion of this course, students will be able to apply R to make publishable quality figures, which will benefit students short-term on their senior thesis and long-term in any scientific career.
 - a) Select appropriate figures for presentations and interpretations.
 - b) Generate publishable high-quality figures with proper formatting, labels, and annotations.

Prerequisites:

The major prerequisite is to be excited about learning and science. Technical prerequisites are having completed one of the following courses or receiving departmental approval.

BIO 1110 – Molecules, Cells, and Organisms (LAB SCI)
 BIO 1120 – Organisms, Populations, and Systems (LAB SCI)
 CHM 1020 – General Chemistry I (LAB SCI)
 PYC 1500 – Introduction to Psychological Science (SOC)
 MTH 1120 – Calculus I (MTH)
 CSC 1100 – Introduction to Computing (CSC)

Grading Policy:

Students will be evaluated through the following mechanisms:

a) In-class participation and exercises	30%
b) Out-of-class exercises (Homework)	15%
c) In-class assessments (Quizzes)	15%
d) Research project	
i. R script and written report	20%
ii. Oral presentation	20%

Resources:

The required textbook is *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* by Hadley Wickham & Garrett Grolemund (1st edition), which is freely available at <http://r4ds.had.co.nz/>. Reading assignments will include scientific research papers and sections from the textbook. All handouts of lectures and oral presentations will be provided to students and available on GitHub at <http://github.com/jegner77/r-jterm>.

Technical Requirements:

R and RStudio, an integrated development environment (IDE) to simplify data science for R users. All students will require a PC running Windows OS or a MAC running OS X and an internet connection for downloading R and RStudio. All resources are freely available. Hyperlinks are provided below for downloading R and RStudio and other helpful resources.

R software	– https://ftp.ussg.iu.edu/CRAN/
RStudio	– https://www.rstudio.com/products/rstudio/download/#download
Q&A and troubleshooting issues	– https://stackoverflow.com/
Blog specific to R topics	– https://www.r-bloggers.com/
Helpful cheat sheets for using R	– https://www.rstudio.com/resources/cheatsheets/
Code repository	– https://github.com/

Course Evaluations:

All students completing the course for credit are strongly encouraged to complete an online course evaluation through google docs and in accordance with Carthage College.

Makeup Policy:

Any absences or missed work must be discussed with course instructor prior to incidence or as soon as possible. Attendance policy is in accordance with Carthage College.

Course Schedule (subject to change):

Date – Jan	Topic	Recommended Reading (Chapters in r4ds) before class	Misc.
W – 9	What is data science, R, and the Tidyverse?		
R – 10	Data visualization: Grammar of graphics	1, 2, 3	
F – 11	Data visualization: Which way to plot your data?	7	
M – 14	Exploratory data analysis and how to solve R problems	4, 8, 9	
T – 15	Generating a minimal reproducible example (reprex)	Handout	Quiz
W – 16	Data visualization: Plotting statistics	22, 23	
R – 17	Data visualization: Modeling	24, 25	
F – 18	Data visualization: Collecting data and generating hypotheses	26, 28	
M – 21	Data wrangling: Import, Tidy, and Transform	9, 10, 11, 12	Research Project approved
T – 22	Data wrangling: Import, Tidy, and Transform continued	9, 10, 11, 12	Quiz
W – 23	Data wrangling: Handling dates, times, and categorical data	13, 14, 16	
R – 24	Modeling basics: “All models are wrong, but some are useful”	15, 16	
F – 25	Model building and model discrimination	24, 25 + Handout	
M – 28	Research project work day		
T – 29	Journal Club: Examples of R in primary literature	Assigned journal article	Quiz
W – 30	Oral presentations on research projects		Written reports due
R – 31	Oral presentations on research projects		Complete course evaluation