

BANA 8090 Special Topic: Data Wrangling with R Spring (2nd half) 2018

Class Information

Instructor: Dr. Bradley C. Boehmke
E-mail: boehmkbc@wfu.edu
Class Days: Saturdays
Class time: Sat 1:00-4:50pm
Class Location: Lindner 110
Class Webpage: http://uc-r.github.io/data_wrangling
Office Hours: TBD

Welcome to Data Wrangling with R! This course provides an intensive, hands-on introduction to the R programming language. You will learn the fundamental skills required to acquire, munge, transform, manipulate, and visualize data in a computing environment that fosters reproducibility.

Course Objectives

Upon successfully completing this course, you will be able to:

- Perform your data analysis in a literate programming environment
- Import and manage structured and unstructured data
- Manipulate, transform, and summarize your data
- Join disparate data sources
- Methodically explore and visualize your data
- Develop your own functions
- And perform basic predictive analytic modeling

... all with R!

Class Structure

Each week you will read and work through selected tutorials on specific data wrangling activities in R. In this class I blend external interactive learning modules via Datacamp.com where you will complete specified modules prior to each Saturday session. Then in each Saturday class I'll spend the first part of class reviewing the data wrangling activity and answer any burning questions. Then you will break up into defined small groups and work together to complete a data wrangling problem prior to the end of class. Thus, the majority of class time will be spent practicing and applying what you learned outside of the classroom.

The purpose for this course structure is multi-dimensional:

- It will teach you to read and learn R programming tutorials and techniques on your own.
- The out-of-class modules will force you to come to each class prepared and these modules will also prepare you for your final project.
- The in-class small group work will teach you to work on a coding task collaboratively and within a constrained time limit and also teach you to assess other people's code.

Material

All required classroom material will be provided in class or online. Any recommended yet optional material will also be provided in the classroom notes.

Performance Evaluation

Your final course grade will be determined according to the following requirements and their respective weights.

1. Datacamp Modules	20%
2. Mid-term Project Evaluation	20%
2. Final Project	50%
3. Participation	10%
TOTAL	100%

Final grades will be distributed according to the following cutoffs:

A	94 – 100 %
A-	90 – 93 %
B+	87 – 89 %
B	83 – 86 %
B-	80 – 82 %
C+	77 – 79 %
C	73 – 76 %
C-	70 – 72 %
D & F	Hopefully None!

Datacamp Modules

During the class term, you will have free access to Datacamp learning modules. I have selected specific modules that you will need to complete prior to each Saturday class. I will track on-time, late, and missed completions of these modules. Instructions to access these modules will be provided at the start of class.

Mid-term Project Evaluation

Throughout the term you will progressively create your final project. You will submit your final project progress midway through the course for a progress evaluation. This progress checks will provide you direction for final completion. Further guidance regarding expectations for this midway project evaluation will be provided in class.

Final Project

The purpose of the final project is to import an existing data set on the web, clean and tidy the data, and perform exploratory data analysis; all while using R Markdown to produce an HTML report that is fully reproducible. This project puts to work the tools and knowledge that you gain throughout this course and will provide you with multiple benefits:

- It will provide you with more experience using data wrangling tools on real life data sets.
- It helps you become a self-directed learner. As a data scientist, a large part of your job is to self-direct your learning and interests to find unique and creative ways to find insights in data.
- It starts to build your data science portfolio. Establishing a data science portfolio is a great way to show potential employers your ability to work with data.

The course is structured in a way that allows you to work on your project as you progress through the weeks. Thus, you should not have to cram during the last two weeks of the term to complete your project. Rather, I plan to have you work on the project and use some of the in-class time to do peer evaluation of your code.

Engagement

Engagement will include participating in class discussions, small group work, providing quality peer review of code, and being involved in discussions on Slack. At the end of the term I will have all students perform a peer assessment of their small group members. This assessment will rate each member from 0-5 in several areas regarding:

- Reviewing and providing constructive feedback regarding your code
- Engaging in quality discussions with the group to improve code knowledge
- Working well as a team member in the small group activities
- etc.

I will use this feedback to help in determining your level of engagement.

Midterm & Final Exams

The midterm and final exams have been replaced by the final project and mid-term project evaluation.

Software

We will use the following software during the course. Plan on bringing a computer to each class meeting.

- R and RStudio will be used to perform all programming activities, assignments, and the final project. You can find details on how to download these here.
- Slack will replace e-mail and Blackboard for our course. You will receive an invitation to the UC R slack team. You may wish to install one of the apps; more information will be provided.

Schedule (tentative)

<u>Week</u>	<u>Topic</u>	<u>Datacamp Tutorial</u>	<u>Due</u>
1	Introduction & Reproducibility <ul style="list-style-type: none"> • Intro to the course, R, and RStudio • Managing workflow & reproducibility 	Intro to R (Ch. 1) Rstudio IDE Part 1 Reporting w/R Markdown	
2	First Date Guidelines for Data <ul style="list-style-type: none"> • Importing data • Getting to know your data 	Importing data in R (part 1) Importing data in R (part 2)	
3	Data Structures & Cleaning <ul style="list-style-type: none"> • Understanding data structures • Tidying & preparing data for analysis 	Intro to R (Ch 2-6) Cleaning data in R	
4	Data Transformation <ul style="list-style-type: none"> • Data manipulation • Relational data 	Data manipulation w/dplyr Joining data in R with dplyr	Mid-term Project Evaluation
5	Data Visualization <ul style="list-style-type: none"> • Data visualizations 	Data visualization w/ggplot 1 Data visualization w/ggplot 2	
6	Creating Efficient Code <ul style="list-style-type: none"> • Control statements and iterations • Writing functions 	Intermediate R Writing functions in R	
7	Introduction to Applied Modeling <ul style="list-style-type: none"> • Unsupervised learning • Supervised learning 	Machine learning toolbox	
8	Finals Week		Final Project Due

Fine Print

Academic Integrity: As with all Lindner College of Business efforts, this course will uphold the highest ethical standards, which are critical to building character. LCB instructors are required to report ANY incident of academic misconduct (e.g., cheating, plagiarism) to the college review process, which could result in severe

consequences, including potential dismissal from the college. For further information on Academic Misconduct or related university policies and procedures, please see the UC Code of Conduct

All academic programs at the Lindner College of Business apply a “Two Strikes Policy” regarding Academic Integrity. Any student who has been found responsible for two cases of academic misconduct may be dismissed from the College. The “Two Strikes Policy” supplements the UC Student Code of Conduct.

All cases of academic misconduct (e.g., cheating, plagiarism, falsification) will be formally reported by faculty. Students will be afforded due process for allegations, as outlined in the policy. If a student is found guilty of academic misconduct in two instances, the student may be dismissed from the Lindner College of Business. The “Two Strikes Policy” is now in effect.

Disability: Students with disabilities who need academic accommodations or other specialized services while attending the University of Cincinnati will receive reasonable accommodations to meet their individual needs as well as advocacy assistance on disability-related issues. Students requiring special accommodation must register with the Disability Services Office.

Attendance: Your attendance is expected at every meeting. If you must be absent, I request that you notify me in advance of the class meeting.

Grade appeals: If you think the grade of your work (homework, peer reviews, participation) is miscalculated, you have the right to appeal. The appeal must be done (through email) within 7 calendar days since the grade is released/posted. After that, your grade is final and will not be changed.