DSC 205.01/CSC 482.02 Spring 2023 Syllabus

Marcus Birkenkrahe

January 11, 2023

1 General Course Information

- Meeting Times: Tuesday/Thursday, 13:00-13:50 hrs

- Meeting place: Lyon Building Computer Lab 104

- Professor: Marcus Birkenkrahe

- Office: Derby Science Building 210

- Phone: (870) 307-7254 (Office) / (501 422-4725 (Private)

- Office Hours: Mon/Wed/Fri 16:15-16:45, Tue/Thu 16:00-16:30

- Textbook: Book of R, Davies (2016), Part II, R Programming

2 Standard and course policies

- **Standard Lyon College Policies** are incorporated into this syllabus and can be found at: lyon.edu/standard-course-policies.

- The **Assignments and Honor Code** and the **Attendance Policy** are incorporated into this syllabus also and can be found at: tinyurl.com/LyonPolicy.

3 Objectives

This course continues the journey into data science using the functional, object-oriented statistical programming language R, begun in DSC 105, "Introduction to data science". It includes calling and writing functions, conditional and looping statements. We will also explore data science using command line UNIX tools.

4 Student learning outcomes

Students who complete dsc 205, "Advanced introduction to data science", will be able to:

- Import data into R, store them, and transform them for analysis

- Visualize data as part of advanced explorative data analysis

- Understand basic predictive modeling strategies and methods

- Master statistical programming in R using the "Tidyverse" package

- Master the infrastructure for advanced statistical computing

- Know how to effectively present assignment results

- Improve data literacy

- Be ready for advanced data science courses like data visualization (DSC 302) and machine learning (DSC 305)

- Research and present a project as a team

5 Course requirements

Introductory knowledge of R as taught in DSC 105 or obtained independently by completing the DataCamp online course "Introduction to R" or "fasteR: Fast Lane to Learning R!" (chapters 1-15 only, freely available on GitHub), or Davies, The Book of R (NoStarch, 2016, Part I only). Basic R concepts are repeated and practiced at the start of the term.

6 Grading system

You should be able to see your current grade at any time using the Canvas gradebook for the course.

REQUIREMENT	UNITS	PPU	TOTAL	% of TOTAL
Final exam	1	100	100	20.
Home assignments	10	10	100	20.
Class assignments	10	10	100	20.
Project sprint reviews	5	20	100	20.
Multiple-choice tests	10	10	100	20.
TOTAL			500	100.

7   Grading table

This table is used to convert completion rates into letter grades. for the midterm results, letter grades still carry signs, while for the term results, only straight letters are given (by rounding up).

%	MIDTERM GRADE	FINAL GRADE
100-98	A+	
97-96	A	A (PASSED -
95-90	A-	VERY GOOD)
89-86	B+	
85-80	B	B (PASSED -
79-76	B-	GOOD)
75-70	C+	
69-66	C	C (PASSED -
65-60	C-	SATISFACTORY)
59-56	D+	
55-50	D	D (PASSED)
49-0	F	F (FAILED)

8   Schedule and session content

For **important dates**, see the 2022-2023 Academic Calendar at: catalog.lyon.edu/202223-academic-calendar

Lectures and lab sessions are aligned with the content of the 10 Data-Camp lessons that need to be completed in the course of the term. Short introductions to text mining, machine learning, and data science on the command line are included, time permitting.

WEEK	DATE	TOPICS & **ASSIGNMENTS**	TESTS
1	Jan 11,13	Calling functions	
2	Jan 18,20	**Intermed R: Conditionals**	Test 1
3	Jan 23,25,27	**Intermed R: Loops**	Test 2
4	Jan 30, Feb 1,3	**Intermed R: Functions**	Test 3
5	Feb 6,8,10	Writing functions	
6	Feb 13,15,17	**Intermediate R: apply**	Test 4
7	Feb 20,22,24	**Intermed R: Utilities**	Test 5
8	Mar 1,3	**Introduction to Bag-of-Words**	Test 6
9	Mar 6,8,10	Natural language processing	
10	Mar 13,15,17	**Introduction to data.table**	Test 7
11	Mar 27,19,31	**Importing and exporting data**	Test 8
12	Apr 3,5	Introduction to shell	Test 9
13	Apr 10,12,14	**Downloading data on shell**	
14	Apr 17,19,21	**Data cleaning and munging**	Test 10
15	Apr 24,26,28	Machine learning	
16	May 1, 3	Project presentations	