DSC 205.01/CSC 482.02 Spring 2024 Syllabus

Marcus Birkenkrahe

January 4, 2024

1 General Course Information

- Meeting Times: Mon-Wed-Fri, 14:00-14:50 hrs
- Meeting place: Derby Science Center Computer Lab room 209
- Professor: Marcus Birkenkrahe
- Office: Derby Science Center 210
- Phone: (870) 307-7254 (Office) / (501 422-4725 (Private)
- Office Hours: Mon/Wed/Fri 10-10.50 & 15-15:50, Tue/Thu 16:15-16:45

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 begun in DSC 105 using both the R and Python programming languages. It includes writing functions, natural language processing, importing and exporting data, cleaning data, and working on the command line. Student projects cover R packages and Python libraries as part of an interactive session.

4 Student learning outcomes

Students who complete DSC 205, "Introduction to advanced data science", will be able to:

- Import data into R and Python, store 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 and Python
- Master the infrastructure for advanced statistical computing
- Know how to effectively present assignment results
- Be ready for advanced data science courses like data visualization (DSC 302) and machine learning (DSC 305)
- Research and present a capstone project
- Apply the agile Scrum methodology to project management

5 Course requirements

Formal prerequisites (2024 catalog): DSC 105 (intro to data science) or CSC 100/115/109 (intro to programming). May be waived if relevant knowledge can be demonstrated.

Introductory knowledge of R as taught in DSC 105 or obtained independently by completing the DataCamp online course “Introduction to R” and "Python for R Users", and "Introduction to Python", or “fasteR: Fast Lane to Learning R!” (chapters 1-15 only, freely available on GitHub). For R: Davies, The Book of R (NoStarch, 2016, Part I only). For Python: Matthes, Python Crash Course (NoStarch, 2023, Chapter 1-7), Basic R and Python 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.

WHEN	DESCRIPTION	IMPACT
Weekly	DataCamp assignments	25%
Monthly	Sprint review presentations	25%
Weekly	Tests	25%
TBD	Final exam	25%

Notes:

- To pass: 60% of all available points.
- DataCamp assignments: there are 16 assignments spread out over 4 courses. Each assignment contributes 1.5625% (25/15) to your final grade. Late assignments are counted as 50% complete only.
- Sprint review presentations: a customer-focused team effort resulting in a project presentation, with 4 Scrum sprint reviews.
- Tests: weekly online quizzes, which are previewed and reviewed in class.
- Final exam: selection of the most challenging weekly quiz questions.
- You only have to write the final exam if you want to improve your grade at the end of the course. If the final exam result is below your final grade average up to this point, it will be ignored.
- There will be bonus programming assignments to make up for other categories.

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). This table is also used in Canvas to compute your grades. Transitions between letter grades are rounded (e.g. 89.5% to 90%).

Percentage	LETTER GRADE*
100% to 89.5%	A (very good)
< 89.5% to 79.5%	B (good)
< 79.5% to 69.5%	C (satisfactory)
< 69.5% to 59.5%	D (passed)
< 59.5% to 0%	F (FAILED)

8 Schedule and session content

For **important dates**, see the 2023-2024 Academic Calendar at: catalog.lyon.edu.

Lectures and lab sessions are aligned with the content of the 10 DataCamp 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	TOPICS	DataCamp
1	Introduction to R: Basics and Calling Functions	Intermediate R (C)
2	Introduction to Python: Basics and Calling Functions	Intermediate Python (C)
3	Intermediate R: Conditionals and Loops	
4	Intermediate Python: Conditionals and Loops	
5	Writing Functions in R 1st Sprint Review	
6	Writing Functions in Python	
7	Intermediate R: Utilities and apply functions	
8	Intermediate Python: Data Manipulation with Pandas	
9	Natural Language Processing in R 2nd Sprint Review	Bag of Words
10	Natural Language Processing in Python	WordClouds
11	R: Importing and Exporting Data	Import data into R
12	Python: Importing and Exporting Data	Importing data
13	Data science on the command line 3rd Sprint Review	Data Processing in Shell
14	Integration of SQL with R and Python	SQLite in R/SQLAlchemy
15	Integration of C++ with R	Optimizing R code with Rcpp
16	Project Presentations 4th Sprint Review	

DataCamp assignments: C = course, L = Lesson, T = Tutorial, P = Project

Workload: approx. 4.25 hours per week.

1. Class time = 16 * 3 * 50/60 = 40 hours
2. Tests (home) = 16 * .5 = 8 hours
3. DataCamp assignments = 20 * 1 = 20 hours

No class on: Monday 19 Feb (Mental-Health Monday) - 18,20,22 March (Spring break) - Friday 29 March (Easter break)