Introduction to Data Science (DSC 105) Syllabus

DSC105 Introduction to Data Science - Syllabus

Marcus Birkenkrahe

August 13, 2023

Contents

1 General Course Information

- Course title: Introduction to data science
- Course number and section: DSC 105.01

- Meeting Times: Mon-Wed-Fri from 13:00-13:50 am
- Meeting place: Derby Science Building computer lab room 209
- Professor: Marcus Birkenkrahe
- Professor's Office: Derby Science Building 210
- Phone: (870) 307-7254 (office) / (501) 422-4725 (private)
- Office hours: Mon/Wed 2-3pm, Tue 4-4.45pm, Thu 9.30-10:45am
- Textbook: The Book of R (Part I, ch. 1-8) by T Davies (NoStarch, 2016) - reserved copy available in the library

2 Objectives

Data science is concerned with getting data to work for us, to give us its (presumed) hidden treasures. Data science has been called "the sexiest job of the 21st century". Even if you don't want to become a professonial data scientist, it's helpful to master the basic concepts if you want to succeed in today's data-driven business.

This course focuses on data science basics, simple visualization and productivity tools using the statistical programming language R as our starting point.

3 Target audience

The course is for anyone who is interested in becoming more data literate in their own field of interest - be it languages, theatre, biology, psychology or exercise science - and growing their personal skill stack.

Visualization of data-driven insights and improved productivity when working with data and media is a concern for any professional.

4 Student Learning Outcomes

Students who complete "Introduction to data science" (DSC 105) will be able to:

- Learn computer and data science principles by playing with data

- Acquire basic statistical programming skills using R
- Apply literate programming principles to their work with Org-mode
- Use infrastructure including command line, Jupyter notebooks, GitHub
- Understand the relationship of humans, machines, and data
- Develop their critical thinking skills
- Know how to effectively present assignment and project results

Students, who complete DSC 105 will have fulfilled the prerequisites for DSC 205, Introduction to advanced data science, which focuses on advanced practices in data exploration, visualization, and integration with databases and other languages through R and Python packages and APIs.

5 Course requirements

Formal prerequisites: Introduction to Programming (either CSC100 or CSC115 or CSC109, and MTH101 (College Algebra).

Otherwise no prior knowledge required. Some knowledge of, and experience with computers is useful but not critical. Curiosity is essential. You will gain data literacy skills by taking this course. The course will prepare you for further studies in computer and data science, or in other disciplines that use modern computing, i.e. every discipline, from accounting to zoology).

6 Grading

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

Notes:

- To pass: 60%
- DataCamp assignments: there are 15 assignments spread out over 3 courses. Each assignment contributes 1.6667% (25/15) to your final grade. Late assignments are counted as 60% 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.

7 Dates and class schedule

Week	Date	Datacamp assignments	Project
1	Aug 21-Aug 25		
2	Aug 28-Sep 01		
3	Sep 04-Sep 08		
4	Sep 11-Sep 15		1st sprint review
5	Sep 18-Sep 22		
6	Sep 25-Sep 29	Introduction to R	
7	Oct 02-Oct 06		
8	Oct 09-Oct 13		2nd sprint review
9	Oct 16-Oct 20		
10	Oct 23-Oct 27		
11	Oct 30-Nov 03	data.table in R	
12	Nov 06-Nov 10		3rd sprint review
13	Nov 13-Nov 17		
14	Nov 20-Nov 24		
15	Nov 27-Dec 01	Python for R users	
16	Dec 04-Dec 08		4th sprint review

- NO CLASSES: Aug 21, Oct 9 (Fall break), Nov 22 + 24 (Thanksgiving). See 2023-2024 academic calendar.
- ONLINE CLASSES: Sept 15 + 22.

8 Learning management system

- We use Lyon's Canvas installation for this course.
- The home page contains: assignments, grades, pages, people, syllabus, quizzes, Google Drive, Course evaluation and Zoom.
- The Zoom page includes cloud recordings of all past sessions.
- Recorded sessions will be deleted after the last class.

9 DataCamp

The course includes a free subscription to the DataCamp classroom at datacamp.com for further study, and for the opportunity to earn certificates. DataCamp is a popular data science online learning platform.

10 GitHub

All course materials are available in a public GitHub repository (github.com/birkenkrahe/ds1). GitHub is the worldwide largest online platform for software development.

11 Lyon College Standard Policies (Fall 2023)

Online: `https://tinyurl.com/LyonPolicyF23`, see also Class Attendance