

Spring 2020

CS 636-1J4: Data Analytics with R Programming

Yao Shen

Follow this and additional works at: <https://digitalcommons.njit.edu/cs-syllabi>

Recommended Citation

Shen, Yao, "CS 636-1J4: Data Analytics with R Programming" (2020). *Computer Science Syllabi*. 97.
<https://digitalcommons.njit.edu/cs-syllabi/97>

This Syllabus is brought to you for free and open access by the NJIT Syllabi at Digital Commons @ NJIT. It has been accepted for inclusion in Computer Science Syllabi by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

Course Syllabus

CS636 – Spring 2020 - Data Analytics with R Programming

Class Schedule: Monday 6pm-8:50pm (Jersey City Campus)

Instructor: Yao Shen, email: yshen@njit.edu

Office hour: 4:50pm-5:50pm MR

TA: Le Gao, email: lg282@njit.edu

TA remote office hour by appointment

Course description and format:

This course will teach how to program in R and how to use R for effective data analysis. The students will learn how to install and configure R necessary for an analytics programming environment and gain basic analytic skills via this high-level analytical language. The course covers fundamental knowledge in R programming. Popular R packages for data science will be introduced as working examples. The format of the course will include lectures by the instructor, computing labs, class discussion, directed reading, and student presentation or project. The exact format will depend on the size of enrolment and student background.

Prerequisite:

Some basic knowledge of programming, probability and statistics. If in doubt about the prerequisites, please consult with the instructor for permission to take the class.

Attendance:

You are supposed to attend all the classes. Participation is highly encouraged to make the class more interactive. Class attendance and participation are taken into consideration by the instructor for the evaluation of the students. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class be sure to consult one of your classmates about the content of the lecture and visit the course web page and Canvas (canvas.njit.edu) to get notes, exercises, assignments, deadlines and announcements.

Textbooks:

1. R programming for data science, by Roger D. Peng

<https://www.cs.upc.edu/~robert/teaching/estadistica/rprogramming.pdf>

2. simpleR - Using R for introductory statistics, by John Verzani, Chapman & Hall/CRC, 2014, ISBN 1466590734

<https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

3. Using R for Introductory Statistics, by John Verzani, Chapman & Hall/CRC, 2004, ISBN 1584884509

<http://www.alexdevri.es/wp-content/uploads/Using-R.pdf>

4. An Introduction to Statistical Learning with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
<http://www-bcf.usc.edu/~gareth/ISL/>

5. Advanced R, by Hadley Wickham, ISBN 9781466586963
<https://adv-r.hadley.nz/>

Grading:

The requirements of this course will consist of participating in lectures, homework, in class computing lab assignments, two exams, five quizzes and two projects. The grading breakdown is the following:

- Homework and Computing Lab Exercise (10%)
- Quiz (20%)
- Term Projects (10%)
- Midterm (20%)
- Final Exam (35%)
- Attendance (5%)

Collaboration and Honor Code:

Students may discuss problems together but must write up their own solutions. When writing up the solutions, students should write the names of people, if any, with whom they discussed the assignment. Note in particular that copying homework or programming assignments, in full or in part is forbidden. Students found cheating or plagiarizing will be immediately referred to the Dean of Students and the NJIT Committee on Professional Conduct and subject to Disciplinary Probation, a permanent marking on the record, possible dismissal, and an “F” grade in the course. All submitted assignments will be checked for similarities, and plagiarism and guilty students identified.

Tentative course schedule (Subject to changes according to progress):

Week	Date	Topic	Readings	Assignments	Due
1	1/27/2020	R introduction: overview of R, getting started with R	R Programming: Pg. 4-12, 23-36; simpleR: Pg. 1-7	Lab 1	
2	2/3/2020	R Nuts and Bolts	R Programming: Pg. 12-23, 36-49	Lab 2 HW 1	
3	2/10/2020	Control structures, loop functions, and R functions	R Programming: Pg. 12-23, 36-49, 62-108	Lab 3 Quiz 1	HW1 due
4	2/17/2020	Data manipulation and data simulation	R Programming: Pg. 49-62, 123-131; simpleR: Pg. 41-54	Lab 4 HW 2	
5	2/24/2020	Data explorative analysis and data visualization	simpleR: Pg. 8-24, 54-59	Quiz 2	HW2 due
6	3/2/2020	Probability distributions	simpleR: Pg. 59-77,	Lab 5	

			89-93	Project 1 assign	
7	3/9/2020	Probability distributions and statistical tests	simpleR: Pg. 59-77, 89-93	Quiz 3	
8	3/16/2020	Spring Break			
9	3/23/2020	Midterm			
10	3/30/2020	Statistical tests	Statistical_usingR: Pg. 233-253, Pg. 325-336, Pg. 353-379	Lab 7	
11	4/6/2020	Unsupervised learning	An introduction: Pg. 373-419	Lab 8 Project 2 assign	Project 1 due
12	4/13/2020	Supervised learning, Model evaluation and tree based method	An introduction: Pg.15-42, 59-127, 175-203; simpleR: Pg. 77-89	Quiz 4	
13	4/20/2020	Linear Regression, Logistic Regression	An introduction: Pg. 203-265; 127-175, 265-303	Lab 9	
14	4/27/2020	SVM, PCA	An introduction: Pg. 303-337	Lab 10	
15	5/4/2020	Neural Networks, Ensemble models, Final Review	An introduction: Pg. 337-373	Quiz 5	
16	5/11/2020	No class			Project 2 due
17		Final Exam			