

CS 636: Data Analytics with R Programming

Summer 2020

Schedule: Saturday 9:00 am - 1:00 pm, WebEx

Instructor: David Li, email: dli@njit.edu

TA: Jianlan Ren, email: jr689@njit.edu Hours: Friday 1-2pm, or by appointment in other days

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 enrollment and student background and will adjust according to the progress.

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. In general, students who attend class regularly perform much better than those who come only occasionally. If you miss one class be sure to watch the recorded video and get notes, exercises, assignments, deadlines and announcements.

Textbooks (helpful but not required):

- R Programming for Data Science, by Roger D. Peng, <https://leanpub.com/rprogramming>
- Using R for Introductory Statistics, by John Verzani, Chapman & Hall/CRC, 2004, ISBN 1584884509
- Advanced R, by Hadley Wickham, ISBN 9781466586963.

Grading:

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

Homework, computing lab exercise (10%)

Quiz (20%)

Term Project (20%)

Midterm (20%)
Final Exam (30%)

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 topics (Subject to changes according to progress)

1. Class overview and R basics.
2. Quiz 1 & Advanced Data Structures, Functions, Vectorization.
3. Functional Programming with R.
4. Quiz 2 & Manipulate Dataframe and Visualization with Plotly.
5. Big Data & Graph Analytics with Spark.
6. Midterm Exam & Probability and Statistics for Data Science.
7. Text Analytics, NLP & Similarity.
8. Quiz 3 & Data Clustering.
9. Linear models and metrics.
10. Quiz 4 & Feature Selection & Other Machine Learning Algorithms.
11. Final Exam