



COURSE SYLLABUS

I-Course Information :

Title of the course	--R Data Analysis
Department	Information Management and Data Sciences
Specialization	Data Science
Academic Year	2019-2020
Semester Duration	
Credit Number	
Number of Hours	55 hours
Course Instructor	Dr.Maguy Medlej
Email	Maguy.medlejkhoury@hotmail.com
Office Hours	0

II- Course Description

This course covers the essential data analysis techniques applied in R. These techniques are typically applied before formal modeling commences and can help inform the development of more complex statistical models. Exploratory techniques are used for eliminating or sharpening potential hypotheses about the world that can be addressed by the data. We cover in detail the plotting systems in R as well as some of the basic principles of constructing data graphics. We also cover some of the common data manipulation, data correlation, data clustering and dimensionality reduction techniques.

III-Course Objectives

This course aims to provide students with/allows students to:

- ❖ Making exploratory data analysis
- ❖ Principles of analytic graphics
- ❖ Plotting systems and graphs in R
- ❖ Data Manipulation in R
- ❖ Correlation Analysis in R
- ❖ Clustering methods
- ❖ Dimension reduction techniques



❖ Practical use case implementation

In order to achieve the objectives of the course, the following requirements must be met by the student:

- Scientific activities
- Use case implementation
- Academic regulations
- Academic work and perseverance
- 70% attendance and participation
- Project implementation : in the project, you will be asked to construct or reproduce certain plots, to analyze your data through the covered techniques . You will be evaluated on the plot that you produce, the analysis output and the code that you write.

IV- Course Outline and Calendar

Learning Objective	Learning Content	Pedagogical Strategy& Tools
Data Analysis Overview	• General Introduction • Why Data Analysis? • What is Data Analysis? • Data Analysis Tool • Data Analysis: What Kind of Data? • Types of Data Analysis: Techniques and Methods • Data Analysis Process	Lecture R environment preparation Assignment: Data Collection
Exploratory Data Analysis in R	• Formulate your question • Read in your data • Check the packaging ▶ Run str() ▶ Look at the top and the bottom of your data • Check your “n”s • Validate with at least one external data source • Try the easy solution first • Challenge your solution • Follow up questions	Lecture Lab R Assignment



Principles of Analytic Graphics	• Show comparisons • Show causality, mechanism, explanation, systematic structure • Show multivariate data • Integrate evidence • Describe and document the evidence Content	Lecture Practical R Exercises Task Assignment
Exploratory Graphs in R	• Characteristics of exploratory graphs • Use Case: Air Pollution in the United States • Getting the Data • Simple Summaries: One Dimension • Five Number Summary • Boxplot • Histogram • Overlaying Features • Barplot • Simple Summaries: Two Dimensions and Beyond • Multiple Boxplots • Multiple Histograms • Scatterplots	Lecture Lab Task Assignment
Plotting System in R	• The Base Plotting System • The Lattice System • The ggplot2 System • Simple Base Graphics • Some Important Base Graphics Parameters • Base Plotting Functions • Base Plot with Regression Line • Multiple Base Plots	Lecture Lab Task Assignment
Data Manipulation in R	• Data Already in R • Examples of Reading and Formatting Datasets • Major tasks in Data Preprocessing • Data transformation and Data Discretization • Manipulating Data with dplyr • Tidying Data with tidyr	Lecture Practical R Exercises Task Assignment



	• Using dplyr • Using tidy • Correlation Analysis	
Data Clustering	• Clustering techniques • Illustrating the K-means algorithm • Stopping the algorithm • Using the kmeans() function • Building heatmaps from K-means solutions	Lecture Practical R Exercises Task Assignment
Working with large Datasets: Dimension Reduction	• Matrix data . • Patterns in rows and columns . Principle components analysis • Dealing with missing values	Lecture Lab R Assignment
Use Case implementation in R	• Implementing practical use case	Data Analysis Case Study

P.S.: The course outline above may be modified during the semester.

V- Teaching approach & methods

Slides, Practical exercises, assignments, documents reading, project implementation

VI- Assessment Methods

The success of the student in this course is based on:

- 1- His/her active participation in class.
- 2- His/her scientific contribution through research done outside the classroom.
- 3- His/her reading notes, research reports and practical applications as requested by course instructor.
- 4- Assignments delivery
- 5- Project implementation

The assessment of the student is carried out according to the following grid:

Assessment Type	Final Grade Weighting (%)	Dates/Deadlines
Participation (projects, research paper, oral presentation, etc.)	20%	



Midterm Exam – Practical Assignment	20%	
Final exam	60%	
Total	100%	

VII- Specific Course Requirements

1- Regulations concerning deadlines for written and applied work submissions

The teacher has the right to:

- Refuse to receive written applied work outside the deadlines.
- Give the student additional time after omitting up to 25% of the mark awarded to the concerned activity.

2- Academic measures in case of plagiarism

The student shall be liable to the following penalties in case of any attempt of illicit scientific citation:

- Cancellation of the student's contribution, activity or scientific article and reduction of the final mark.
- If the same offense is reported in other courses, cancellation of the final exam and failure in the relevant course(s).

VIII- Reference Book

- **Beginning Data Science in R**, Thomas Mailund, 2nd edition
- **R and Data Mining: Examples and Case Studies**, Yanchang Zhao. yanchang@rdatamining.com, <http://www.RDataMining.com>, October 20, 2015
- **R Programming for DataScience**, Roger D. Peng, 2015
- **Exploratory Data Analysis with R**, Roger D. Peng, 2015

IX- Bibliography& Additional Sources

- <http://cran.r-project.org/doc/manuals>
- <http://www.r-project.org/doc/bib/R-books.html>
- Chambers (2008). Software for Data Analysis, Springer.
- Chambers (1998). Programming with Data, Springer.
- Murrell (2005). R Graphics, Chapman & Hall/CRC Press.
- <http://www.r-project.org/doc/bib/R-books.html>