ID de aportación : **61** Tipo: **Oral**

# Integrating R Software as a Teaching Resource for STEM Education: A Multidisciplinary Approach

In the context of higher education, teaching Science, Technology, Engineering and Mathematics (STEM) disciplines constantly faces the challenge of staying relevant and effective in a world characterized by rapid technological advancements and socioeconomic changes. In particular teaching advanced statistics and data science or machine learning have many challenges due to the complexity of the methods and the limitations for formal demonstrations in non-maths carrears. In response to this demand, this project proposes an innovative methodology that uses R software and some of their specific libraries as the main teaching resource in advanced statistical and data science training for STEM higer education.

R software [1], widely used in the scientific and business community for data analysis and visualization, offers a rich and versatile environment for exploring and applying key concepts in STEM degrees. By adopting R as a pedagogical tool as it is shown in [2], the aim is not only to teach students the technical skills necessary for data manipulation and statistical analysis but also to foster a deeper understanding of the underlying principles in areas such as data science, mathematical modeling, and programming.

The proposed methodology is based on a practical and experiential team work approach, where students learn by doing as discussed in [3].

The methodology is based on building wide groups of students to face long term project for an entire term where they will ve asked to apply the theoretical knowledge gained in their own real data, enabling them to find the difficulties of real data and assisting them on overcomming the challenges week by week. This contributes to

develop transferable skills and a more solid understanding of concepts. Every week the theoretical class introduces a concept from data collection, preprocessing, modelling or validation and the corresponding practical class consists on applying the new knowledge to their project. Lecturers provides a basic script in R, RMarkdown or Shiny depending on the week . The script contains some errors and they have to tune it to their particular data by correcting the mistakes as well . Sometimes the lecturer runs the script with some academic data while in parallel they try to run with their own data. Thanks to this pedagogic technique, the students learn to solve mistakes in the code, gain practical skills on real applications and connect what they do with the theoretical concepts. Additionally, the focus on collaboration and teamwork promotes peer learning and fosters a supportive and collective environment in the classroom. Lecturers answer questions and doubts during lab sessions and guides the group as far as they can go.

The projects are presented in public sessions to the entire class twice or 3 times along the course so that the evolution of the projects is progressive.

The common discussion is part of the training programm as well. Each dataset has different requirements at the level of preprocessing, modelling or interpretation and sharing the work to the entire class allows to learn from the experience of other groups as well. This is relevant as the knowledge corpus of these courses is huge and this is a way to cover more topics inside a term.

One of the distinctive aspects of this methodology is its ability to integrate multiple STEM disciplines into a single course or project. By using examples and applications that span diverse areas such as biology, engineering, economics, and social sciences, students have the opportunity to explore the interconnections between different fields and understand how they can apply their skills in a variety of contexts.

However, despite the many potential benefits that this methodology offers, it also presents unique challenges. The initial learning curve for students can be steep, especially for those with no prior experience in programming or data analysis. Additionally, the effective integration of R into the curriculum requires careful planning and adequate resources, including teacher training and access to suitable hardware and software.

As a conclusion, this project proposes an innovative methodology for teaching STEM disciplines that uses R software as the main learning engine. By providing students with a practical and multidisciplinary experience, it is hoped that this methodology will not only prepare students for real-world challenges but also inspire a

lasting passion for learning and exploration in the field of STEM.

**Palabras claves**. Education, R, Statistics methods, Machine Learning

References
[1] de Micheaux, P. L., Drouilhet, R., & Liquet, B. (2013). The R software. Fundamentals of Programming and Statistical Analysis, 978-1.
[2] Koretsky, M., Keeler, J., Ivanovitch, J., & Cao, Y. (2018). The role of pedagogical tools in active learning: a case for sense-making. International journal of STEM education, 5(1), 1-20.
[3] Thompson, P. (2010). Learning by doing. Handbook of the Economics of Innovation, 1, 429-476

## ¿Presentas la comunicación a premio?

Premio estudiante (grado, máster, doctoral)

## Afiliación (del autor)

KEMLG at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC)

**Autor primario:** ANGERRI, Xavier (KEMLG at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC))

**Coautores:** GIBERT, Karina (KEMLG at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC)); RAMIREZ-MITJANS, Sergi (KEMLG at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC))

**Presentador:** ANGERRI, Xavier (KEMLG at Intelligent Data Science and Artificial Intelligence (IDEAI-UPC))

**Clasificación de la sesión:** Sesión premio predoctoral

**Clasificación de pistas:** Estadística