

SYLLABUS DE COURS

Année d'étude - parcours	M2 Economie de l'innovation et veille stratégique
Intitulé du cours	Introduction to Data Science
Numéro de l'UE	UE 4.6
Volume horaire (en heures maquettes)	15H
Objectifs et compétences développées	The most underrated task in data science is arguably data preparation. From accessing the data, to importing it, to cleaning it, to combining it, and finally to creating the set of variables we want. That's easily 80% of the data scientist' job. And unfortunately, the road of data prep' is fraught with pitfalls. In this course, we will see how to develop a professional workflow by efficiently working with data.
Contenu et moyens pédagogiques	We will learn data science with the free statistical software R. We'll cover (provisional outline): • How to organize your workflow and make it reproducible • Data handling with data.table • Use functions! • Creating complex variables • Data science problems: what it is and how to avoid them The course will strive to be as hands on as possible. We will deal together with crappy real world data to illustrate the concepts. Students will be required to make short presentations at the beginning of the class (to be defined).
Pré-requis	None, although a basic knowledge of R would be best.
Modalités d'évaluation	One group project and one sitting exam.
Références bibliographiques	• Introduction to R, by Robert J. Hijmans • Grant McDermott's Big Data in Economics