

Master Mathématiques appliquées, statistique

Données : manipulation et visualisation

Responsable	Descriptions	Informations
Pierre PUDLO pierre.PUDLO@univ-amu.fr	Code : SMSAU24 Nature : Unité d'enseignement Domaines : Sciences et Technologies	Composante : Faculté des Sciences

LANGUE(S) D'ENSEIGNEMENT

Français

CONTENU

Le CM s'attachera à présenter différents concepts de gestion des données tabulaires qui sont au cœur de nombreuses technologies de manipulation et traitement dans les langages spécialisés comme R, Python, SAS, Spark, SQL,... mais également les tableaux. La partie du CM portant sur la visualisation présentera les différents types de graphiques possibles et proposera quelques lectures critiques de graphiques.

- Données tabulaires : variables et individus ; typage d'une variable
- Variable répartie sur plusieurs colonnes suivant la date, ou autre : redressement sous forme de tableau long ; opération inverse
- Opérations de filtre, group by et de jointure
- Notion de calcul à virgule flottante, et données manquantes dans les opérations algébriques
- Identification et traitement de problèmes : données manquantes, incohérentes, doublons, erreurs de rapprochement,...
- Les différents types de représentation graphique
- Critique d'une représentation graphique

L'UE « Données : manipulation et visualisation » fait l'objet d'un complément spécifique à chaque parcours.

- DS : Découverte des outils de Business Intelligence comme Power BI et/ou Tableau
- IMSA : Applications actuarielles en SAS et Python pour la construction d'indicateurs, la gestion de bases de stock et de flux
- MASS POP : Ce complément s'ajoute au travail effectué dans l'UE de projet intégré sur le traitement de données catégorielles

COMPÉTENCES À ACQUÉRIR

- Importer des données complexes dans un logiciel-langage de science des données
- Remettre en forme, gérer les données manquantes, transformer des variables
- Trier et filtrer les individus, calculer de nouvelles variables ou en supprimer
- Utiliser des jointures pour compléter un tableau de données
- Calculer des statistiques par groupe (group by)
- Visualiser des données univariées et bivariées

Les TP seront

- sous Python (NumPy, Pandas, Matplotlib, Seaborn) en Data Science
- sous R (base, tidyverse,...) en IMSA
- sous SAS (étape DATA et GPLOT) en MASSPop

BIBLIOGRAPHIE, LECTURES RECOMMANDÉES

Cady, F. (2017). The data science handbook. John Wiley & Sons.

Cody, R. (2018). Learning SAS by example: a programmer's guide. SAS institute.

Cleveland, W. S. (1985) The Elements of Graphing Data. Wadsworth: Monterey, CA, USA.

Delwiche, L. D., & Slaughter, S. J. (2019). The little SAS book: a primer. SAS institute.

Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media.

Healy, K. (2018). Data visualization: a practical introduction. Princeton University Press.

McDaniel, S., & Hemedinger, C. (2007). SAS for Dummies. John Wiley & Sons.

McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media.

Mount, J., & Zumel, N. (2019). Practical data science with R. Simon and Schuster.

VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media.

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). R for data science. O'Reilly Media.

VOLUME HORAIRE

- Volume total: 24 heures
- Cours magistraux: 12 heures
- Travaux pratiques: 12 heures

CODES APOGÉE

- SMSAU24C [ELP]

M3C

Aucune donnée M3C trouvée

POUR PLUS D'INFORMATIONS

[Aller sur le site de l'offre de formation...](#)



Dernière modification le 19/06/2024