Fév - Mars 2019 | IMT Mines Alès | Spécialisation Informatique et IA

Module Visualisation & REStitution

Compétences acquises : à la fin du module, les étudiants auront des bases en visualisation de données : principaux types de graphiques, comment les choisir et principales erreurs. Ils seront capables d’explorer efficacement un jeu de données grâce à ggplot2, créer un rapport reproductible avec R Markdown et le partager grâce à Github.

# Introduction A la visualisation de données (2h / Cours)

## Qu’est-ce que la visualisation de données

Définition. A quoi cela sert. Différence exploration vs communication. Quels sont les différents types de rendu. Quelques exemples connus. Quels sont les principaux outils existant : Excel vs outil click bouton vs langage de programmation vs visualisation pour le web. Ressources. Auteurs connus.

## Vue d’ensemble des différents types de graphique

## Partie basée sur le site [data-to-viz.com](https://www.data-to-viz.com/). Rappels des types de graphiques les plus connus, puis découverte des moins connus. Montrer le lien donné d’entrée – graphique possible. Exemple d’analyse. Partie interactive : montrer un jeu de données, demander ce que l’on pourrait faire avec.

## Principales erreurs

## Description des 10 principales erreurs rencontrées en visualisation de données. Partie interactive : montrer un graphique qui ne va pas, et discuter pourquoi il ne va pas. Voir [ici](https://www.data-to-viz.com/caveats) pour les erreurs qui seront discutées.

## Prérequis : installer R + R Studio + ggplot2

# INTRODUCTION à ggplot2 (3h / TP)

## Qu’est-ce que ggplot2

## Définition. Pourquoi c’est génial. Qui l’a inventé. Qu’est-ce que le Tidyverse. Quelques exemples

## Construction d’un premier graphique

## Construction d’un premier graphique sur le jeu de données iris, en temps réel chacun sur sa machine. Détails des différentes étapes nécessaires : data, Aesthetics, Shapes, et customisation.

## Exercice

## Liste de ~ 4 jeux de données simples à analyser chacun à son rythme. Permet de découvrir les différentes fonctionnalités de ggplot2 et les différents types de graphique. Le but est aussi de montrer aux étudiants que toutes les réponses sont sur le web.

## Graphiques interactifs avec ggplotly

## Si suffisamment de temps : introduction aux graphiques interactifs avec ggplotly

## Prérequis : arriver avec un jeu de donné de son choix chargé dans R

# MISE en application (2H / TP)

## Option 1 : Travail sur leur projet de Module

## Si j’ai bien compris les étudiants ont un projet à présenter à la fin du module. Ce créneau permettrait donc de discuter de leur projet et de les aider pour la partie visualisation.

## Option 2 : Travail sur un jeu de donné réel

## Leur demander de venir avec un jeu de donné trouvé qqpart et l’analyser en temps réel.

## Prérequis : installer les packages rmarkdown et knitr. Avoir au moins 2 graphiques de l’étape précédente terminés avec une histoire à raconter.

# Restituer ses resultats avec R Markdown (1h30 / TP)

Introduction à R Markdown. Tutorial. Application aux résultats de l’étape précédente. A la fin de la séance, les étudiants ont transformé leurs quelques graphiques en un rapport reproductible contenant leur code et explications

## Prérequis : créer un compte GitHub, avec sa photo de profil et ses renseignements basic.

# PArtager ses resultats grace à github (1h30 / TP)

Introduction à github, création d’une premier repo, premier commit. Transfert du rapport précédent sur github. Puis processus de mise en ligne : leur rapport devient un site web qui peut donc être partagé avec un URL.   
Si assez de temps : soumettre une issue et un pull request au travail de son voisin.