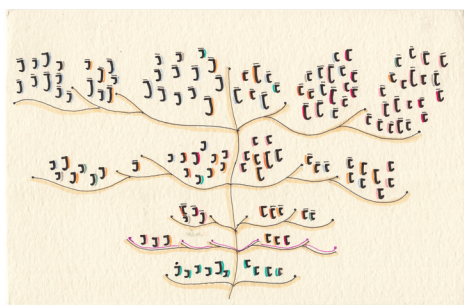


Cours de data visualisation

Laurent Politis

Contents

Syllabus	1
Présentation générale	1
Thèmes abordés	2
Références et conseils de lecture	3



Syllabus

Présentation générale

L'objectif du cours est d'utiliser le langage de programmation statistique R et d'outils en ligne simple afin de mettre en valeur vos résultats d'analyses statistiques. Le cours est constitué d'exemples de codes R et d'exercices que l'étudiant appliquera directement sur son ordinateur et sa console R sous cette forme :

```
print('Hello world !')
```

```
## [1] "Hello world !"
```

Dans un premier temps, nous allons découvrir comment importer et explorer différentes données sous R. Puis nous allons apprendre comment préparer les données afin de créer de magnifiques graphiques et applications interactives. Ces outils acquis durant le cours vous permettront de mettre en valeur vos études ou reportings, qui vous seront utiles dans un contexte académique ou professionnel.

Plan

Ce cours est un TD informatique constitué d'exemples et d'exercices que l'étudiant reproduira sur son ordinateur. Durant ces 15 heures de cours les points abordés sont :

1. Extraction et manipulation de données
2. Représentations graphiques classiques et améliorations
3. Représentations graphiques interactives (html widget, plotly ...)
4. Shiny et R markdown création d'application et de rapport interactif

Mode d'évaluation

L'évaluation de ce cours consiste en la création d'un programme R sur la mise en forme des données utilisées pour votre mémoire. Le projet peut être mené individuellement ou en binôme.

Thèmes abordés

Présentation du cours et de R

Le 14 mars, 10h00/13h00 * 1h cours magistral*

Au début du cours seront présentés la data visualisation en général et le logiciel R.

≈ 2h travaux dirigés

Un rappel sur l'utilisation de la console et de l'environnement de développement intégré (IDE) R studio sera effectué au début de cette première partie. Des fonctions de base sous R permettant d'importer les données depuis un fichiers texte, csv, excel et même directement depuis le web seront présentées. On examinera comment R stocke les données sous forme de variables, comment les manipuler et de créer des graphiques. . .

Visualiser ces données

Le 13 avril, 10h00/13h00 = 3h travaux dirigés

Dans les cours du précédent semestre vous avez appris les fonctions R de méthodes statistiques descriptives et de modélisation de base, tels que `mean()`, `median()`, `sum()`, `lm()`, `summary()`... Il existe des plots aussi de base qui permettent d'examiner les différents jeux de données et modèles statistiques. Ces graphiques constituent des arguments et des preuves pour mettre en évidence les résultats que vous avancez et leurs pertinences de manière simple et synthétique. Les graphiques que nous allons créer sont entre autres :

- Les boîtes à moustaches
- Les histogrammes
- Les courbes de densité
- Les fameux camembert (et les limites de leurs utilisations)

Dans le cours du semestre dernier vous avez appris comment modéliser des relations entre différents jeux de données à l'aide de modèle linéaire `lm()` et ARIMA... Nous allons découvrir les diverses graphiques diagnostiques de ces modélisations statistiques qui permettent entre autres de repérer les meilleures variables explicatives, de rejeter les mauvaises et de vérifier ces résultats de modélisation.

- Nuage de points, qqplots, séries temporelles ...

Graphiques Interactifs et applications

= 6h travaux dirigés

Nous avons passé en revue différentes manières de présenter nos données sous forme d'illustration statistique. Dans cette partie, les outils de visualisation interactive faisant appel à des librairies codées en java et en html 5 seront présentés (`html widget`, `shiny` et `r markdown`). Les outils comme `shiny`, `r markdown` et `plotly` permettent de créer de véritables petites applications de visualisations et d'applications de vos résultats. Par exemple, le syllabus que vous lisez est entièrement fait sous R à l'aide de `r markdown`.

Travaux pratiques

= 3h tp

Le but de ce TP est de commencer à travailler sur les données de votre mémoire à l'aide d'un programme R et des notions apprises.

Références et conseils de lecture

Sur la visualisation des données

- <http://www.r-bloggers.com/> un blog contenant des articles divers et variés sur les nouveautés R.
- <http://vissoc.co/> cours de Data Visualisation avec R et en particulier avec ggplot2
- <https://sites.google.com/a/analytics.usfca.edu/msan-622-experimental-design-and-visualization/> cours de data visualisation avec R ggplot2 et shiny
- <http://flowingdata.com/> site compilant de très belles illustrations faites en grande partie avec R
- <http://www.htmlwidgets.org/>

Sur les bases de R

- <https://openclassrooms.com/courses/effectuez-vos-etudes-statistiques-avec-r>
- <http://www.statmethods.net/> pour des exemples rapides en R
- <https://www.datacamp.com/> des cours de base gratuits sur R interactifs

Autres

- <https://www.kaggle.com/>