

Visualisation Interactive de Données (Dataviz)

MOS 5.5 — École Centrale de Lyon

Romain Vuillemot

Version 0.1 du 2022/01/03 à 12:50:50

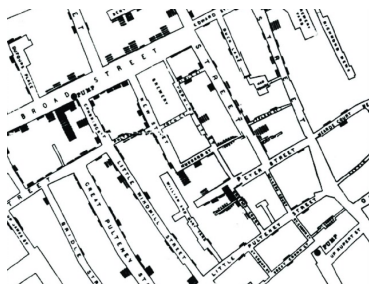
Table des matières

1	Exploration visuelle de données	5
1.1	Checklist d'exploration de données	5
1.2	Vocabulaire de visualisation	9
1.3	Outils de visualisation	12
1.3.1	Tableau Software	12
1.3.2	Vega Charts	12
1.3.3	Observable Plots	12
1.4	Exemples et exercices	12
2	Dessiner une visualisation interactive	16
2.1	Marques et propriétés graphiques	16
2.2	D'une visualisation statique à dynamique	18
2.3	Repère de coordonnées et échelles	18
2.4	Data binding	19
2.5	Exercices	19
3	Dessiner les graphiques de base en visualisation	20
3.1	Bar chart	20
3.2	Line chart	21
3.3	Scatterplot	21
4	Rendu multi-vues interactives	22
4.1	Encapsuler les visualisations	22
4.2	Interactions	22
4.3	Assembler en vues coordonnées et tableau de bord	23
4.4	Décoration et layouts	23
4.5	Animation et transitions animées	23
5	Graphiques avancés	25
5.1	Visualisation de trajectoires spatio-temporelles	25
5.2	Visualiser les données temporelles	27
5.3	Visualisation de cartes géographiques	28
5.4	Visualiser les graphes et les arbres	28
A	Projets	30
B	Questions fréquemment posées	32

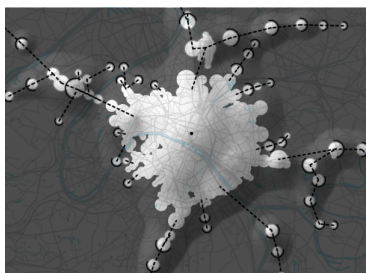
Préambule

Le but de ce cours est d'apprendre à **créer ses propres visualisations interactives** de manière dynamique à partir de données. Il s'agit d'une étape essentielle du domaine plus large de la *data science* (science des données) dont le but est la découverte de connaissance à partir de données. Dans ce domaine, la visualisation est souvent nécessaire pour plusieurs raisons : déterminer les propriétés des données et aider à la modélisation, comprendre un phénomène de manière objectif, et enfin communiquer efficacement des résultats. Ainsi la visualisation ne se destine pas à seulement réaliser de belles images, mais de donner des points de vues sur les données en fonction du contexte (données, utilisateurs, tâches). Dans ce cours nous aurons une approche orientée *informatique* sur toute la chaîne de vie d'une donnée, depuis sa captation, stockage, pré-traitement, mais nous verrons aussi quelques bases de design graphique. Enfin même si l'utilisation d'outils de visualisation sera faite (ex. Tableau) le coeur du travail sera de créer ses propres visualisations personnalisées et déployables sous forme d'applications web.

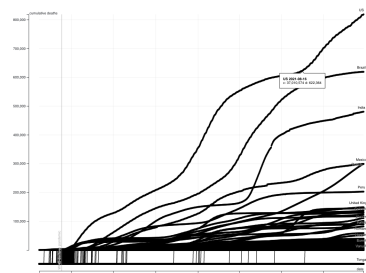
Ci-dessous quelques exemples démontrant la diversité des représentations possibles, qui ont été conçues depuis plusieurs siècles sous forme papier au début pour identifier les cas de choléra (a) ¹ et plus récemment la création de cartes avancées de mobilité urbaine (b) ² et plus récemment le nombre de cas quotidiens du COVID-19 (c) ³. A la fin de ce cours vous serez en mesure de créer des visualisations similaires avec le jeu de données de votre choix.



(a) Cas de choléra XIXème siècle



(b) Carte de mobilité urbaine



(c) Nombre de cas du Covid-19

Organisation du cours, préparation aux séances et rendus

Le cours comporte au total 28 heures de face à face divisé en 7 séances de 2h de cours et 2h de TP. Les séances ont pour objectif d'introduire progressivement les compétences aussi bien techniques en programmation et traitement de données, que de design relatives à la conception de visualisation. Toutes les séances sont indispensables et il est important de bien respecter l'ordre des séances.

Vendredi	07/01/2022	13:30	Cours 1 : Exploration visuelle de données
Vendredi	07/01/2022	15:45	TP : Tableau, Observable et Vega/Plots
Vendredi	14/01/2022	13:30	Cours 2 : Dessiner une visualisation
Vendredi	14/01/2022	15:45	TP : Une visualisation interactive
Vendredi	21/01/2022	13:30	Cours 3 : Graphiques de base
Vendredi	21/01/2022	15:45	TP : Bibliothèque de visualisations
Vendredi	28/01/2022	13:30	Cours 4 : Multi-vues
Vendredi	28/01/2022	15:45	TP : Tableau de bord
Vendredi	04/02/2022	13:30	Cours 5 : Graphiques avancés
Vendredi	04/02/2022	15:45	TP : Graphe, layouts
Vendredi	11/02/2022	13:30	Cours 6 : Applications
Vendredi	11/02/2022	15:45	TP : Evaluation inter-groupes de projets

1. <https://www.ph.ucla.edu/epi/snow/mapsbroadstreet.html>
2. <https://observablehq.com/@romsson/isochrone>
3. <https://observablehq.com/@fil/covid-19-derived-chart>

Vendredi	16/02/2022	13:30	AUTONOMIE
Vendredi	16/02/2022	15:45	AUTONOMIE
Vendredi	04/03/2022	13:30	Soutenance projets
Vendredi	04/03/2022	15:45	Soutenance projets

Le rendu principal du cours est un projet sous forme de web application interactive à réaliser seul ou en binôme. Les objectifs, règles de travail et conseils d'un bon projet sont présentés en annexe de ce cours. Une évaluation écrite sera aussi à réaliser individuellement sous format papier début mars 2022. Enfin une note de bonus participation et suivi aux cours sera attribuée en fonction de l'activité pendant et entre les séances de cours. Vous pouvez poser les questions dans le GitHub du cours sous forme de tickets avec le label questions <https://github.com/romsson/visualisation-interactive>

Concernant la préparation aux séances de cours, **ce cours est conçu pour être enseigné sous forme de classe inversée**. Cela implique du travail en amont de chaque séance, principalement lire le chapitre du cours correspondant et s'auto-former si certaines connaissances vous manquent. Les séances de cours ne servent principalement qu'à passer en revue le contenu et à répondre aux questions. Le cours aura aussi principalement pour but de réaliser les exercices qui sont proposés. Normalement cette préparation doit vous pendre au maximum une heure, peut être plus si vous avez besoin de vous mettre à jour sur certains aspects techniques (programmation, etc.).

Quelques rappels de règles de base :

- Tout travail individuel (en groupe) doit rester du travail individuel (ou en groupe) ; pas de travail inter-groupes ou inter-personnes (si travail individuel)
- Si vous utilisez du code ou des données produit par d'autres personnes, donnez crédit
- Toute absence doit être notifiée la veille du cours au plus tard à romain.vuillemot@ec-lyon.fr

Pré-requis pour le cours

Il est indispensable avant de commencer le cours d'avoir un ordinateur portable configuré/fonctionnel et pensez à venir avec un chargeur pendant le cours. Enfin si les conditions sanitaires imposent à nouveau un travail à distance, essayez dans la mesure du possible de trouver un endroit calme pour travailler, aussi bien pour préparer la séance que pendant la séance (prévoyez aussi webcam et micro).

Installez un outil de gestion de données (suggestion : Excel ou WPS) afin de charger et naviguer dans des données que vous venez de télécharger (souvent en .xls mais parfois texte ou csv). Éventuellement nous ferons quelques calculs et visualisations en utilisant ces outils, mais aucune opération avancée ne sera nécessaire.

Installez un éditeur de texte et IDE (suggestion VSCode ou SublimeText) pour la partie programmation du cours (éventuellement aussi pour inspecter des jeux de données brut). Vous passerez la plus grande partie de votre temps sur ces outils pour le projet, donc essayez de faire un choix adapté à vos connaissances. Il existe également des outils spécifiques aux systèmes d'exploitation comme Notepad++. Ces logiciels sont très utiles pour la programmation, avec du surlignage en cas d'erreur syntaxique et éventuellement des suggestions de correction. Certains IDEs permettent éventuellement la compilation du code.

Créez un compte GitHub et notebooks Observable qui seront les plateformes que nous utiliserons pour stocker, partager les données (GitHub <https://github.com/>) mais aussi pour effectuer des calculs sur les données (Observablehq <https://observablehq.com/> avec identification en utilisant votre compte GitHub). Les notebooks Observable sont des outils de travail collaboratif qui permettent de partager des données entre plusieurs personnes de manière similaire aux Jupyter Notebooks, mais qui utilisent

JavaScript. Le site web d'Observable contient de nombreux tutoriels, en voici un spécifique au cours : <https://observablehq.com/@john-guerra/observable-ieeevis-tutorial>.

Outil de Visualisation Tableau. Enfin installer Tableau Software qui est disponibles pour Mac et Windows <https://public.tableau.com/fr-fr/s/download>. Pour les utilisateurs de Linux hélas il n'y a pas de support, alors utiliser la version en ligne <https://online.tableau.com/> ou une version allégée (<http://vega.github.io/polestar/>). Suivre un premier tutorial de présentation de Tableau <https://www.youtube.com/watch?v=TPM1ZxRRaBQ>.

Installez un serveur web local (optionnel). Si vous souhaitez développer en local votre projet, vous pouvez utiliser un serveur web local (exemple : <http://localhost:8080/>). Cela vous permet de tester votre code sans avoir besoin d'être connecté à Internet. Il vous faut pour cela soit configurer votre machine afin d'activer le serveur web, soit utiliser un serveur web déjà installé (exemple : Apache ou Nginx) ce qui est le cas pour Mac et Linux. Sous Windows, il faut installer un serveur web (exemple : WAMP).

Maîtrise de langages de programmation web et traitement de données. Dans le cadre de ce cours vous devez absolument maîtriser JavaScript et la programmation web en générale. Pour les étudiants de l'ECL il s'agit de revoir ce qui a été abordé en INF TC3 (programmation web et traitement de données) car dans ce cours nous mettrons à nouveau en oeuvre les notions acquises. Voici les liens à lire et auxquels se référer lors de la partie programmation (il faudra beaucoup de pratique pour bien tout assimiler, donc il faut certes tout lire mais aussi s'y référer régulièrement) :

- HTML et CCSS <https://www.w3.org/Style/Examples/011/firstcss.en.html>
- JavaScript <https://javascript.info/> ou <https://exploringjs.com/impatient-js/toc.html>
- JavaScript et science des données <http://js4ds.org/>
- SVG https://oreillymedia.github.io/Using_SVG/guide/markup.html
- Manipulation du DOM en JS https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Client-side_web_APIs/Manipulating_documents

Enfin quelques liens utiles autour du traitement de données

- Glossaire des données <https://github.com/ramiroaznar/data-glossary>
- Format de données Tidy Data <https://www.jstatsoft.org/article/view/v059i10>

Les principaux livre sont les suivants (en particulier vous êtes invité à lire le premier livre par Scott Murray qui reprend tous les concepts qui seront abordés dans ce cours) :

- *Interactive Data Visualization for the Web* par Scott Murray
- *Visualization Analysis and Design* par Tamara Munzner
- *Design for Information* par Isabel Meirelles
- *Making Data Visual : A Practical Guide to Using Visualization for Insight* par Danyel Fisher et Miriah Meyer