

Géomatique des transports. Une approche transversale entre «géographie» et «informatique et création numérique (ICN)»

Samuel Coulon (histoire-géographie) et David Roche (ICN), lycée général, académies de Paris et Grenoble

● Niveau et thèmes du programme

Géographie, 1^{re} L, ES :

Thème 2 - Aménager et développer le territoire français.

Informatique et création numérique (ICN), 1^{re} S, ES, L :

Visualisation graphique de données : exemple avec des données géolocalisées.

● Problématique à traiter

Big Data et *Open Data* : quels usages pour la géographie et plus particulièrement pour une géographie des transports ?

● Objectifs de contenu

Vocabulaire : datavisualisation, cartogramme, carte choroplèthe, bibliothèque.

Notions : *Big Data*, *Open Data*.

● Objectifs méthodologiques

Comprendre les enjeux liés au phénomène du *Big Data* et de l'*Open Data*

Réfléchir sur l'intérêt, les atouts et les limites de la datavisualisation

Prendre en main une bibliothèque *JavaScript* pour réaliser des graphiques, cartes ou cartogrammes

● Ressources et outils numériques utilisés

D3.js <https://d3js.org/> (une bibliothèque *JavaScript*), le portail européen des données www.europeandataportal.eu, SNCF Open Data <https://data.sncf.com>, Rails & histoire (Association pour l'histoire des chemins de fer) www.ahicf.com

● Cadre de référence des compétences numériques

1.3. Traiter des données.

2.3. Collaborer.

3.2. Développer des documents visuels et sonores.

● Organisation de l'ensemble de la séquence

1. Collecter des informations dans un « océan ».

2. La « datavisualisation » ou l'art de représenter des données de façon visuelle.

3. Faire parler les chiffres... et les élèves. Réflexion sur les modes de représentations graphiques.

● Pistes d'évaluation

Intégration à un « projet » d'ICN réalisé en classe de première par groupes de deux ou trois. Ce projet doit être finalisé par une « réalisation » qui peut faire l'objet d'une évaluation.



17 h

L'ère numérique se caractérise par une augmentation exponentielle du nombre de données et par une accélération croissante de leur duplication et de leur diffusion. Avec la révolution numérique, les données apparaissent comme «le nouveau carburant de l'économie». L'expression anglaise «*data fuel the economy*» souligne le parallèle entre une économie industrielle alimentée par le pétrole (fuel), et une économie de services dont l'énergie serait devenue les «données». La question du *Big Data* et de l'*Open Data* trouve toute sa place dans le cadre de l'enseignement d'«informatique et création numérique». Il s'agit d'un nouvel enseignement d'exploration pour la classe de seconde depuis la rentrée 2015 et d'une option facultative pour les élèves des séries ES, L et S, qui pourra se prolonger en terminale ES et L à la rentrée 2017. Le préambule de son programme, paru au BO n° 29 du 21 juillet 2016 (www.education.gouv.fr/pid285/bulletin_officiel.html?cid_bo=104657f), souligne que «*face à l'accroissement considérable des volumes de données, les représentations visuelles, parfois interactives (graphiques, diagrammes, cartogrammes, etc.) permettent de mieux appréhender ces données et d'en percevoir les enjeux*». Dans une logique transversale et pluridisciplinaire, le programme d'ICN offre de nombreuses possibilités aux professeurs qui enseignent l'informatique de construire avec leurs collègues géographes des «projets» communs s'inscrivant dans une démarche ayant recours à la géomatique. À titre d'exemple, le programme d'ICN propose un projet intitulé «*Visualisation graphique de données : exemple avec des données géolocalisées*». C'est cet exemple que nous mettons en œuvre ici, autour de la question des «*Mobilités, flux et réseaux de communication dans la mondialisation*».

Big Data et Open Data. Quels usages pour la géographie ?

Il s'agit ici, de conjuguer les capacités du programme de géographie de 1^{re} ES et L avec les attentes du programme d'ICN sur la datavisualisation (représentation graphique de données). Cette séquence pédagogique propose d'initier les élèves à la programmation à partir du traitement d'une série de données ouvertes (*Open Data*) sélectionnées en cours de géographie sur la question des mobilités, des flux et des réseaux de communication en France et en Europe. Dans le cadre du cours d'ICN, ils exportent et préparent ensuite leurs bases de données pour pouvoir les visualiser sous différentes formes. Pour ce faire, il existe aujourd'hui de nombreuses «bibliothèques» *JavaScript* qui permettent de diversifier les modes de représentation et d'intégrer des fonctionnalités interactives pour mieux appréhender un phénomène (diagrammes, cartogrammes ou cartes choroplèthes, c'est-à-dire des cartes statistiques utilisant une variation de tons gradués pour caractériser quantitativement un phénomène dans un espace). À la fin de la séquence, les productions font l'objet d'une analyse critique en cours de géographie. Placée après l'étude de l'aéroport de

Roissy, cette séquence n'est pas destinée à remplacer le cours de géographie consacré au sujet, mais peut servir d'exemple ou de point d'appui pour approfondir certaines notions et s'interroger sur les atouts, les contraintes et les limites de la datavisualisation.



1 h

séance

1 Collecter des informations dans un «océan» (géographie)

• **Big Data, Open Data : quelle différence pour le géographe ?**

Bien que souvent associés, le *Big Data* et l'*Open Data* sont deux notions très différentes. Le *Big Data* concerne les méthodes et les technologies de collecte, de traitement et d'analyse en temps réel d'un très grand nombre de données. L'approche *Big Data* cherche à détecter des relations significatives entre ce que l'on étudie et les éléments de l'écosystème qui l'entoure, afin de prévoir une évolution plutôt que d'en expliquer ses causalités. L'*Open Data*, ou donnée ouverte, n'a pas véritablement de corrélation avec le *Big Data*. C'est un mode d'accès à l'information. Les données ouvertes mises à disposition du public et des professionnels sans aucune contrepartie, peuvent être utilisées, analysées et redistribuées librement. Pour des raisons évidentes liées à la complexité des techniques du *Big Data*, cette séquence de travail ne s'intéresse qu'à l'*Open Data*.

• **Chercher, identifier et sélectionner des données pertinentes**

Le volume des données est aujourd'hui tel, qu'il devient difficile de trouver des informations utiles, au bon format, sans y consacrer beaucoup de temps. Le professeur doit par conséquent réaliser lui-même une présélection de quelques sites avant de choisir en classe les données à exploiter avec ses élèves. C'est l'objet de cette première séance de travail. Dans le cadre d'un cours dialogué à l'aide d'un vidéoprojecteur, ou en salle informatique, il s'agit d'identifier les principaux sites d'*Open Data* dans un premier temps (www.data.gouv.fr/fr, <https://data.sncf.com>, www.europeandataportal.eu, etc.) puis de repérer des données pertinentes ayant un lien avec la question des mobilités, des flux et des réseaux de communication en France et en Europe, et enfin de vérifier l'exploitabilité des formats proposés pour des opérations de datavisualisation. Nous n'exploiterons ici que des données sur le transport ferroviaire :

- les données Eurostat sur le transport ferroviaire de passagers (organisme dépendant de la Commission européenne chargée de produire des statistiques officielles de l'UE), <http://ec.europa.eu/eurostat/tgm/table.do?t=ab=table&init=1&language=fr&pcode=ttr00015&plugin=1> ;
- les données de la SNCF sur la durée moyenne des trajets sur les lignes TGV ;
- les chiffres sur la durée moyenne des trajets avant la mise en place des LGV disponibles sur « Rails & histoire », www.ahicf.com/les-rails-de-l-histoire,96.html, un site entièrement consacré à l'histoire du chemin de fer.

• **L'Open Data, un « bien commun » dont la diffusion sert l'intérêt général**

Cette activité doit s'accompagner d'une réflexion sur la mise à disposition des données publiques qui s'inscrit, depuis le début des années 2000, dans le cadre d'une volonté d'optimisation constante des prestations de service dans la gouvernance des territoires. Ces informations diffusées gratuitement par les États, collectivités territoriales ou des entreprises, dans des formats standards peuvent être utilisées, analysées et redistribuées librement. Pour la Commission européenne l'ouverture des données publiques est un « moteur pour l'innovation, la croissance et la transparence des gouverne-

ments» (2011). Dans ce cadre elle a créé son propre portail *Open data* (www.europeandataportal.eu/), qui collecte les données fournies par les administrations nationales, locales et européennes.



15 h

séance

2 La « datavisualisation » ou l'art de représenter des données de façon visuelle (informatique et création numérique)

- **D3.js, une bibliothèque JavaScript pour manipuler des données**

S'il existe mille façons d'effectuer des opérations de datavisualisation, les «bibliothèques» restent les meilleures, pour peu que l'on soit en mesure de se familiariser avec le langage *JavaScript*. C'est l'objet de cette série d'une quinzaine d'heures d'activités guidées.

Écrire un code *JavaScript* est parfois fastidieux, d'où l'émergence de bibliothèques qui proposent de nouvelles fonctionnalités très pratiques pour simplifier le développement du code. C'est le cas de *D3.js* www.ahicf.com/les-rails-de-l-histoire,96.html, qui n'est pas un logiciel mais une bibliothèque créée en 2011 qui permet l'affichage de données numériques sous la forme d'objets vectoriels (SVG) intégrés à une page web et autorise l'ajout d'effets dynamiques (transition, infos bulles, etc.), de styles (couleurs, etc.) et de fonctionnalités interactives.

- **Série statistique, algorithmes et production graphique**

Par deux, les élèves s'initient dans un premier temps aux principes de base de la programmation (affectations, conditions et boucles), puis prennent en main dans un second temps la bibliothèque *D3.js* en réalisant une quinzaine d'activités accessibles sur le site informatiquelycee.fr. Dans un troisième temps, ils conçoivent leurs propres graphiques, cartes ou cartogrammes en s'appuyant sur les compétences acquises préalablement (reformatage des données sélectionnées si nécessaire, choix du mode de représentation, réalisation de la représentation graphique).

Les possibilités sont quasiment illimitées, ce qui fait de *D3.js* une bibliothèque incontournable. Nous proposons ici à titre d'exemples, un cartogramme interactif sur la durée moyenne des trajets en train depuis Paris avant et après la mise en place des LGV (figures 1 et 2), et deux cartes «choroplèthes» (figures 3 et 4) conçues avec les mêmes données Eurostat sur le transport ferroviaire de passagers (millions passagers-km).

Cartogramme interactif des LGV depuis Paris réalisé avec les données SNCF et du site « Rail & histoire ». La longueur du trait depuis Paris est proportionnelle à la durée du trajet avant et après la mise en place des LGV.

Fig. 1 : Trajets en 1980

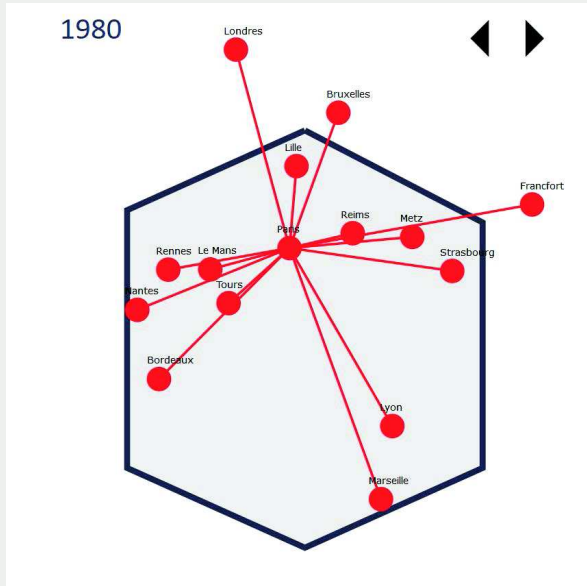
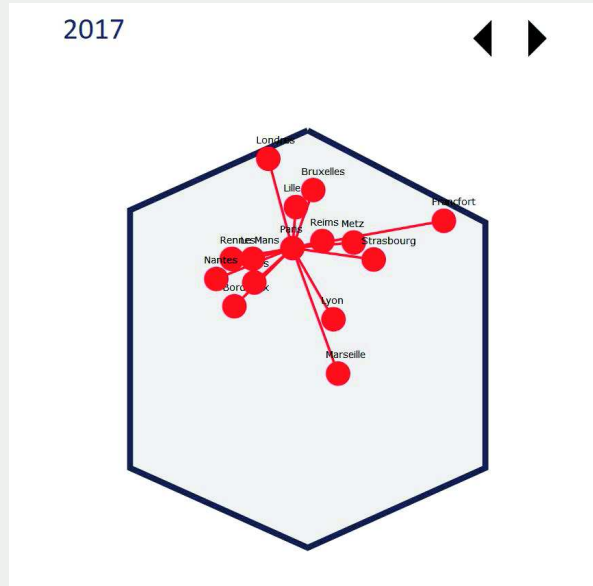


Fig. 2 : Trajets en 2017



Les productions citées en exemples sont accessibles à l'adresse suivante : <http://ftpweb12.scola.ac-paris.fr/>

Productions graphiques réalisées avec les données Eurostat sur le transport ferroviaire de passagers en 2014 (millions passagers-km).

Fig. 3 : Carte choroplèthe n°1 (Découpage en 7 classes)

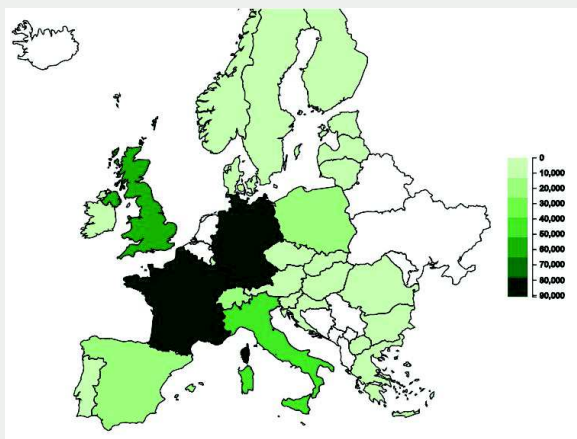
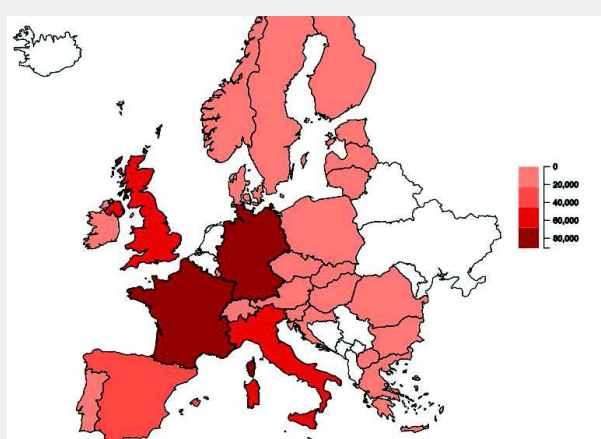


Fig. 4 : Carte choroplèthe n°2 (Découpage en 4 classes)



Les productions citées en exemples sont accessibles à l'adresse suivante : <http://ftpweb12.scola.ac-paris.fr/>



1 h

séance

3

Faire parler les chiffres... et les élèves. Réflexion sur les modes de représentations graphiques (géographie)

Il s'agit de clore la séquence de travail par une mise en perspective des productions d'élèves avec la question de géographie traitée et par une analyse critique autour des modes de représentation utilisés par les élèves. Chaque groupe présente ses productions au reste de la classe (cartogramme, *tree-map*, *bubble chart*, carte choroplèthe, etc.) qui exprime à son tour une série d'observations. Pour guider les élèves, le professeur peut proposer quelques pistes de questionnement : l'indicateur pour la représentation graphique est-il pertinent ? Le graphique est-il complet, lisible et compréhensible sans ambiguïté ? Le sens ou le message porté par le graphique est-il clair ? La nature de l'information est-elle adaptée au mode de représentation choisi ? La discrétisation opérée pour élaborer des cartes choroplèthes est-elle pertinente (échelle, nombre de classes, seuils ...) ? Le graphique apporte-t-il une aide efficace pour la compréhension de phénomènes géographiques ? Etc.

- **Des représentations graphiques pour comprendre les territoires**

Le cartogramme interactif réalisé sur les LGV (figures 1 & 2) met en évidence plusieurs phénomènes : l'inégale desserte du territoire français par les lignes LGV, une configuration radioconcentrique du réseau LGV centré sur Paris (on peut regretter que les axes transversaux n'apparaissent pas sur le cartogramme), l'intégration précoce (Eurostar) et progressive (Thalys, LGV Est) du réseau LGV et des principales métropoles françaises à l'Europe vers le Nord et vers l'Est.

La seconde carte choroplèthe (figure 4), grâce à son découpage en quatre classes et ses couleurs contrastées, rend plus accessible l'analyse géographique du territoire européen que la première carte constituée de sept classes (figure 3). Elle met ici en exergue les grandes inégalités entre les territoires européens. Deux ensembles s'opposent sur le continent : une Europe du Nord-Ouest avec un trafic passager sur rail relativement dense, et une Europe scandinave, méditerranéenne et centre orientale où le transport ferroviaire de voyageurs est moins développé. On regrettera néanmoins l'absence de données pour un certain nombre de pays situés au cœur de l'Europe (Pays-Bas, Belgique, etc.).

- **Engager une réflexion sur les apports et les limites de la représentation graphique de données**

Si la datavisualisation est une solution efficace pour affronter la complexité et démocratiser l'accès à l'information, il convient de ne jamais oublier que c'est avant tout un outil de communication. Il est donc nécessaire de s'interroger sur ses pièges et limites : visualisations déséquilibrées (trop ou pas assez de données), effet brouillard (éléments qui perturbent la bonne compréhension d'un graphique), mauvaise gestion des échelles graphiques (discrétisation maladroite, utilisation de périodes insuffisamment larges pour dégager des tendances par exemple), non-respect des règles de représentation (distorsions, problèmes de perception des données), inadéquation des formes graphiques aux données représentées, prouesse graphique privilégiée pour impressionner son auditoire au détriment de la lisibilité des informations, corrélations trompeuses, etc. Il convient par ailleurs de ne jamais oublier que la datavisualisation, comme toute autre représentation graphique, est porteuse d'un « effet vérité » alors qu'elle procède de la part de son auteur d'une élaboration, de choix, de décisions. Le graphique ou le cartogramme n'offre que ce que son auteur veut montrer.

LES APPORTS DU NUMÉRIQUE DANS LES APPRENTISSAGES

Cette séquence pédagogique se situe au cœur de la « géomatique » (contraction de géographie et informatique), une discipline qui regroupe l'ensemble des outils et méthodes permettant d'acquérir, de représenter, d'analyser, de traiter et d'intégrer les informations géographiques d'origines diverses. Les compétences acquises en programmation par les élèves permettent d'exploiter avec plus de précision leurs bases de données qu'avec de simples logiciels grand public. Elles leur procurent une plus grande souplesse quant aux choix des modes de représentation et permettent de personnaliser leurs réalisations par l'intégration de fonctionnalités interactives. Si le codage peut sembler réhibitoire ou inaccessible pour les néophytes, l'option ICN offre la possibilité de prendre son temps pour acquérir les compétences de base en la matière et de réaliser des travaux plus ou moins complexes selon le niveau atteint par les élèves.

La compréhension du phénomène du *Big Data* et de l'*Open Data*, l'identification des ressources disponibles en ligne pour la collecte de données statistiques, la compréhension des enjeux liés aux représentations graphiques, contribuent à l'enseignement des bases de la littératie informationnelle. Dans un monde où les techniques de visualisation graphique des données rencontrent un véritable succès, les objectifs sont dans le cas présent, autant géographiques que civiques.
