

English version available at: http://ricardo.medialab.sciences-po.fr/Girardetal_ricardo_dh2016_en.pdf

RICardo (*Research on International Commerce*) est une base de données rassemblant des flux bilatéraux de commerce international extraits d'un grand nombre de sources statistiques historiques sur la période 1787-1938. Les principes de construction de cette base sont décrits dans un working paper (Dedinger et Girard 2015) soumis à la revue *Historical Methods*. Cette base de données a commencé à être élaborée par des chercheurs en économie et en histoire en 2004 dans le but de renouveler la recherche sur l'histoire de la mondialisation commerciale. En 2013, une orientation entièrement nouvelle a été donnée au projet RICardo. Économistes et historiens se sont associés à des ingénieurs développeurs et designers afin de créer un outil numérique totalement inédit susceptible d'être utilisé à la fois pour l'enseignement et la recherche.

Explorer le commerce international du XIX^e siècle

Il n'existe pas à ce jour de ressource numérique équivalente sur une période aussi ancienne et aussi longue que celle couverte par la base RICardo. Les seuls outils comparables sont les applications web du FMI (data.imf.org), de l'OMC (stat.wto.org) et des Nations-Unies (comtrade.un.org) qui proposent des données et des visualisations du commerce international sur la période post-1950. Comme nous le montrerons, les problèmes méthodologiques posés par les données commerciales historiques (XIX^e-XX^e siècles) sont plus complexes. Le premier résultat de notre travail est une application web mise à disposition de tous à l'adresse : <http://ricardo.medialab.sciences-po.fr>¹. La base de données sera pour sa part rendue publique en 2017 à l'occasion d'une conférence internationale marquant le deux centième anniversaire de la publication de l'œuvre principale de Ricardo (Ricardo 1817).

Nous discuterons des choix méthodologiques faits lors de la création de cet outil d'exploration de données – ce que nous appelons *datascape* (Latour et al. 2012) – et nous présenterons rapidement les perspectives de recherche et d'enseignement qu'ouvre cet outil, avant de conclure sur la méthode de travail transdisciplinaire employée.

Représenter les données dans leur incertitude

Dans RICardo, l'unité d'information élémentaire est un flux de commerce (exportation ou importation) entre une entité *reporting* et une entité *partner*. Dans sa version actuelle, la base comprend 267 000 flux. Les *partners* sont les partenaires commerciaux enregistrés dans les relevés annuels statistiques du commerce des pays *reporting*. La large couverture spatiale et temporelle de la base soulève un certain nombre de problèmes. Tout d'abord, les entités couvrent des réalités très hétérogènes : les *partners* peuvent être des pays ("United Kingdom"), des groupes de pays ("United Kingdom & Ireland") ou des aires géographiques ("British colonies"). Ensuite, la disponibilité des statistiques du commerce avant la fin du XIX^e siècle est très problématique (Dedinger et Girard 2015), ce qui se traduit par l'absence dans la base d'un nombre non négligeable (17 000) de flux au travers de la période considérée. Le défi que nous avons donc dû relever a consisté à agréger des données hétérogènes sans écraser la complexité du corpus et en tenant compte des données manquantes dans les séries. Nous tentons de répondre à cet épineux problème par l'analyse exploratoire de données (Tukey 1977).

¹ Le code source de l'outil est disponible à l'adresse <http://github.com/medialab/ricardo>

3 niveaux d'exploration

L'interface d'exploration développée propose trois niveaux d'entrée du plus global au plus local : *world view*, *country view* et *bilateral view*. Chaque vue propose un point de vue particulier sur les données de la base. La complexité s'accroît à mesure que l'on passe du point de vue mondial (*world*) au point de vue national (*bilateral*). Pour accompagner l'exploration à travers ces différents niveaux, nous avons choisi de définir une structure commune à ces trois vues. Chacune s'ouvre sur une courbe principale représentant les flux totaux par année (représentation du commerce mondial pour la *world view*, du commerce total d'un pays pour la *country view*, du commerce total entre deux pays pour la *bilateral view*) ainsi que les limites temporelles. C'est une courbe brisée qui s'interrompt lorsqu'il n'y a pas de valeur connue. En plus des interruptions de courbes, un axe temporel au-dessous indique l'absence de données sous la forme d'une projection de la courbe sur l'axe des abscisses. Cet objet graphique sert également d'interface de filtrage temporel. Il est ainsi possible de sélectionner graphiquement une sous-période à explorer plus précisément (Becker et Cleveland 1987). En pied de page de chaque vue, un tableur permet de naviguer dans les données sources visualisées et de les exporter au format tableur (CSV). L'utilisateur peut ainsi poursuivre l'analyse dans un outil de statistique externe. Au centre de cette structure commune, chaque vue propose une exploration plus détaillée.

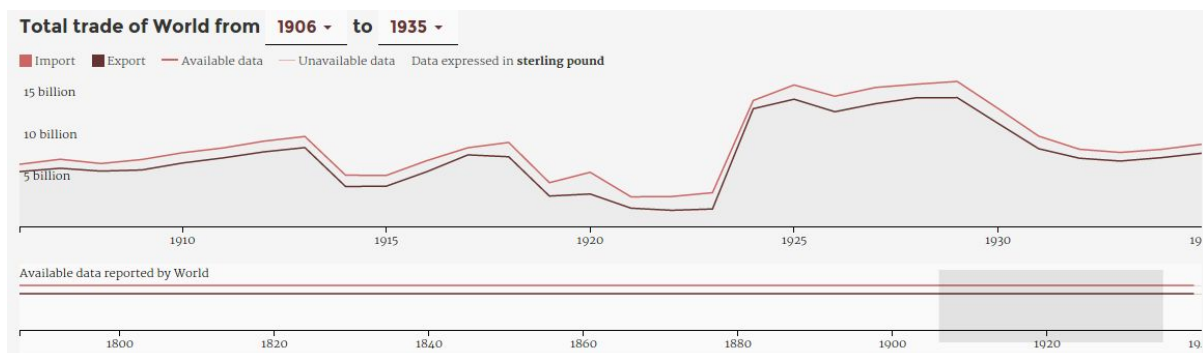


figure 1 : filtrage temporel

Étudier un pays au prisme de ses partenaires

La vue pays propose à l'utilisateur de centrer son analyse sur un pays *reporting* en représentant les balances commerciales par année de ses *partners* sous la forme d'une série d'histogrammes. Cette utilisation de *small multiples* (Tufte 1990) permet de faciliter les comparaisons entre partenaires. Le choix de détailler la balance par année tient au fait qu'une agrégation sur l'ensemble de la période aurait introduit un biais en masquant les années pour lesquelles il manque des données. Par ailleurs l'ordre dans la longue liste des partenaires correspond à la part annuelle moyenne de chaque partenaire dans le commerce total du reporting sur les années disponibles. Cette métrique, représentée sous la forme d'un cercle en début de ligne, permet d'une part de mettre en évidence les partenaires par importance décroissante et d'autre part de faire ressortir des entités *partner* similaires. La figure 3 montre ainsi qu' "Ireland & United Kingdom" et "United Kingdom" étaient les 2^e et 3^e partenaires des États-Unis au XIX^e siècle. Ce sont deux entités différentes mais en réalité un

partenaire semblable : le Royaume-Uni (le recouvrement des années 1864-1876 s'explique par l'inclusion de deux sources différentes).

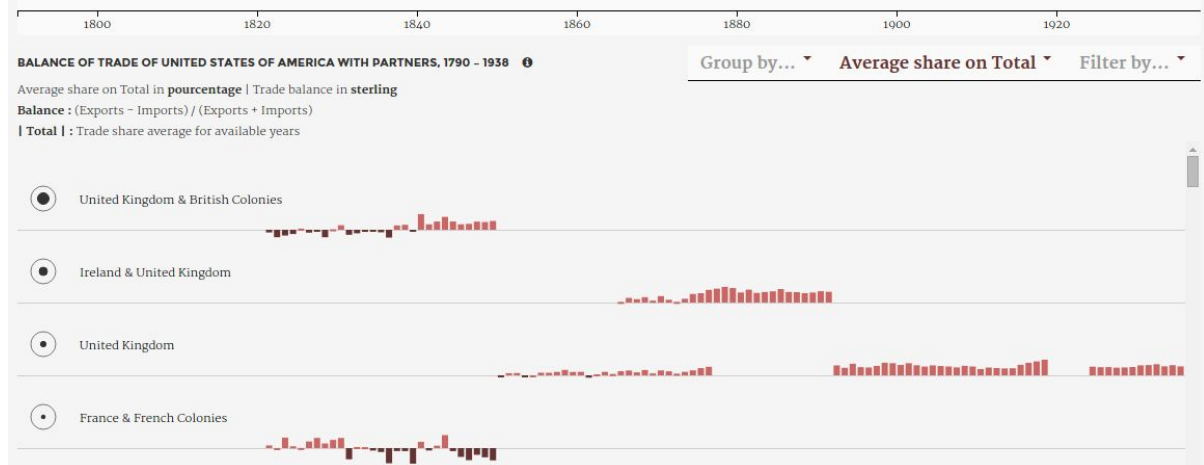


figure 2 - les balances commerciales des partenaires

Le miroir déformant des flux bilatéraux

Dans la vue bilatérale, l'application propose d'interroger un couple de pays *reportings* en utilisant une représentation de leurs flux miroirs. C'est l'une des grandes forces de la base RICardo : un même flux bilatéral est déclaré comme une exportation par l'un des pays et comme une importation par l'autre, or ces deux enregistrements statistiques sont rarement identiques (Dedinge et Girard 2015). Grâce au calcul d'un indicateur (Dedinge 2012), la vue bilatérale offre une vision immédiate de l'écart de ces flux miroirs et de ses fluctuations sur la période choisie.

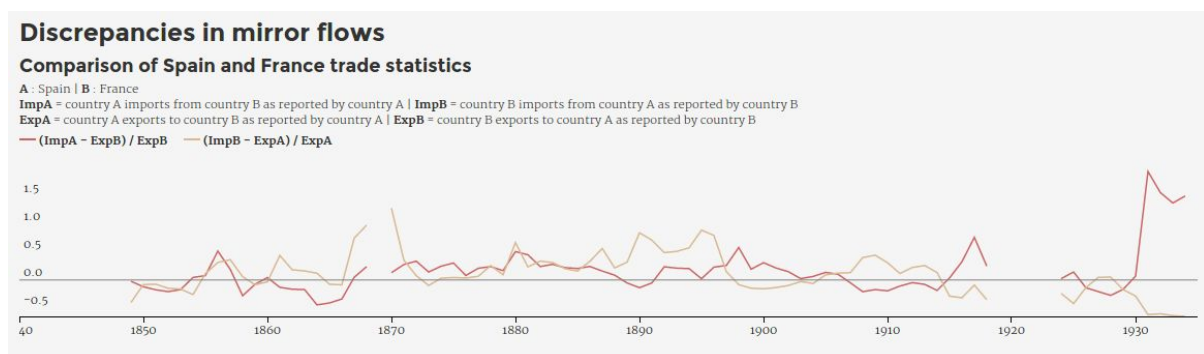


figure 3 - la vue bilatérale

Un outil pour chercheurs et étudiants en SHS

La mise en forme exploratoire de données apporte de puissants outils pour aider la recherche en histoire économique. C'est un excellent moyen de détecter d'éventuelles incohérences dans les données (Leclercq et Girard 2013) afin d'en déterminer les causes. Pour arriver à un tel résultat, notre méthode a consisté à considérer les visualisations de données tant comme résultat que comme moyen de recherche (Stefaner 2010), à ponctuer et dynamiser la collaboration de long terme par un atelier intensif d'une semaine mêlant chercheurs, ingénieurs et designers, ainsi qu'à interroger la pertinence de chaque

visualisation en articulant les contraintes méthodologiques de l'Histoire et de l'Économie, les idées et principes du Design et la faisabilité technique.

Les chercheurs ont à leur disposition un outil incomparable pour approfondir l'analyse de la fiabilité des statistiques commerciales (vue bilatérale), composer des monographies sur l'histoire commerciale des pays du monde (vues pays), étudier l'historique de la mondialisation commerciale depuis le début du XIX^e siècle (vue monde) et pour enseigner l'histoire commerciale des pays et l'évolution historique de la configuration des entités du monde.

Bibliographie

(Becker et Cleveland 1987) Becker, Richard A., and William S. Cleveland, 'Brushing Scatterplots', *Technometrics*, 29 (1987), 127 <<http://dx.doi.org/10.2307/1269768>>

(Dedinger 2012) Dedinger, Béatrice, 'The Franco-German Trade Puzzle: An Analysis of the Economic Consequences of the Franco-Prussian war1: THE FRANCO-GERMAN TRADE PUZZLE', *The Economic History Review*, 65 (2012), 1029–54
<<http://dx.doi.org/10.1111/j.1468-0289.2011.00604.x>>

(Dedinger et Girard 2015) Dedinger, Béatrice, and Paul Girard, 'Visualizing Trade Globalization in the Long Run : The RICardo Project', Working Paper Submitted to *Historical Methods*, 2015
<http://ricardo.medialab.Sciences-po.fr/Dedinger_Girard_RICardo_HistoricalMethods.pdf>

(Latour et al. 2012) Latour, Bruno, Pablo Jensen, Tommaso Venturini, Sébastien Grauwin, and Dominique Boullier, "'The Whole Is Always Smaller than Its Parts" - a Digital Test of Gabriel Tarde's Monads', *The British Journal of Sociology*, 63 (2012), 590–615
<<http://dx.doi.org/10.1111/j.1468-4446.2012.01428.x>>

(Leclercq et Girard 2013) Leclercq, Christophe, and Paul Girard, 'The Experiments in Art and Technology Datascape', in *Collections électroniques de l'INHA. Actes de Colloques et Livres En Ligne de l'Institut National D'histoire de L'art (INHA, 2013)* <<http://inha.revues.org/4926>> [accessed 27 October 2015]

(Ricardo 1817) Ricardo, David, *On the Principles of Political Economy and Taxation*, 1817

(Stefaner 2010) Stefaner, Moritz, 'Bootstrapping - Use Visualizations to Create Visualizations' (presented at the VisualEyes / The Role of Design in Data, Information and Knowledge Visualization, Politecnico Milano, Italy, 2010)
<<http://fr.slideshare.net/MoritzStefaner/bootstrapping-use-visualizations-to-create-visualizations>> [accessed 27 October 2015]

(Tufte 1990) Tufte, Edward, *Envisioning Information* (Cheshire, CT, USA: Graphics Press, 1990)

(Tukey 1977) Tukey, John Wilder, Exploratory Data Analysis (Addison-Wesley Publishing Company, 1977)