



# **Rapport de Visualisation sur les Données des Thèses de Doctorat Nationales et les Vols vers et depuis la Russie**

Ru WANG

**Encadrant Dr. MATHIEU CISEL**

**29.05.22**

## Résumé

La visualisation permet de piloter les données plus efficacement en les représentant a de façon simple et compréhensible. Dans ce rapport, trois graphiques sont présentés en utilisant deux jeux de données, dont deux sont basés sur les données relatives aux thèses de doctorat en France et l'autre est basé sur les données des vols vers et depuis la Russie au mois de Février. Avec ces deux types de données, trois stratégies différentes de représentations graphiques ont été utilisées pour mieux caractériser les résultats.

.

Mots-clefs : science des données visuelles, visualisation des données, analyse visuelle

## 1. Introduction

La visualisation des données (ou dataviz ou représentation graphique de données) est un ensemble de méthodes permettant de résumer de manière graphique des données statistiques qualitatives et surtout quantitatives afin de montrer les liens entre des ensembles de ces données. Cette visualisation fait partie de la science des données [1].

La visualisation des données permet à toute organisation de piloter son activité plus efficacement en adoptant une stratégie data-driven (pilotée par les données) et agile. En représentant les données de manière simple et compréhensible, la visualisation de données permet de faire ressortir des tendances, des phénomènes et des corrélations qui peuvent ensuite être utilisés stratégiquement. Elle raconte en quelque sorte une histoire qu'il convient ensuite d'intégrer à ses plans d'action.

Le type de graphique est un critère de choix en visualisation des données [2]. Il existe les graphiques simples (histogrammes, diagrammes circulaires, courbes, etc.), qui permettent de rentrer de plain-pied dans ce qui constitue l'essence de la visualisation des données. Il existe aussi des graphiques interactifs qui permettent à l'utilisateur de devenir l'acteur de ses découvertes basées sur les données.

Dans ce rapport, deux bases de données sont utilisées, une est tirée du site *theses.fr* et constitue les données de thèses doctorales en France, présentée au chapitre 2.1, et l'autre contient les données de vols vers et depuis la Russie au mois de février 2022, présentée au chapitre 2.2. Trois figures ont été réalisées, deux concernent les données de thèses et la troisième représente les vols vers et depuis la Russie.

## 2. Visualisation de Données

Premièrement, deux visualisations des données de thèses doctorales en France sont exposées, avec des représentations non-interactive et interactive respectivement. Ensuite, une visualisation des données de vols vers et depuis la Russie illustre une technique de représentation spatiale.

### 2.1 Visualisation de Données de Thèses Doctorales En France

Au sein de la base de données des thèses doctorales en France, on trouve les informations sur les années de soutenance de thèses, 15 disciplines séparant les domaines de thèses, le genre de l'auteur, etc. La Figure 1 montre l'évolution du nombre de thèses sur différentes disciplines au cours du temps.

## Evolution du nombre des thèses sur différentes disciplines

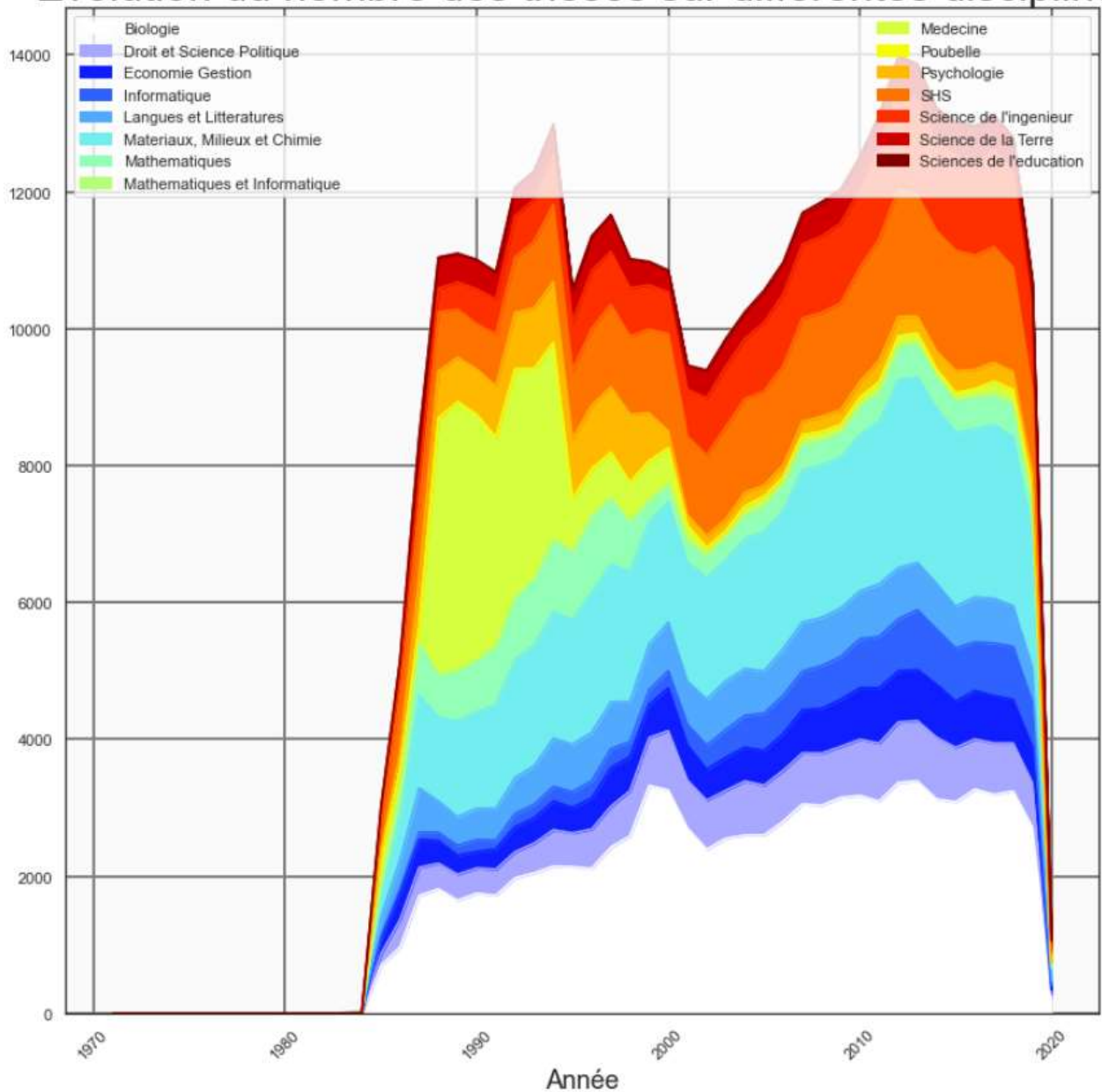
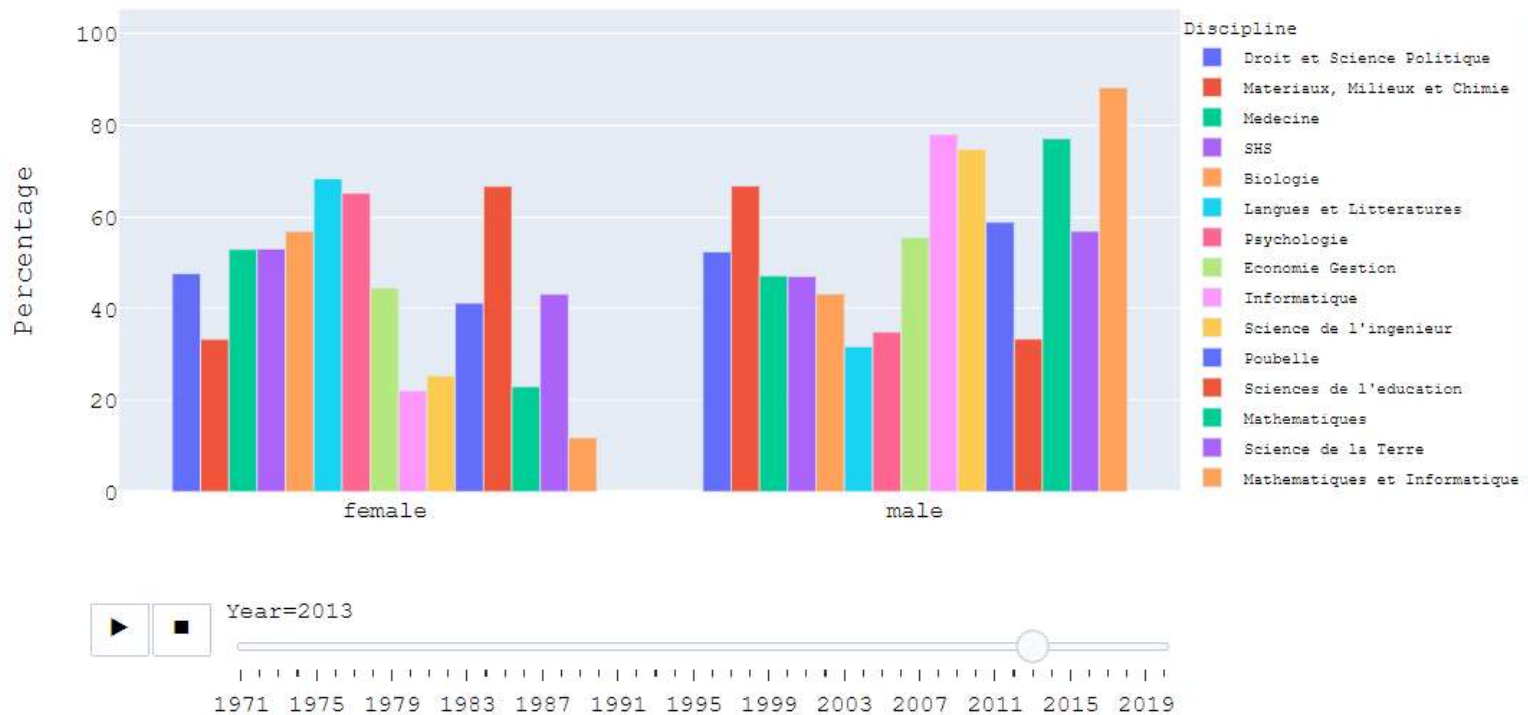


Figure 1 : L'évolution du nombre des thèses dans 15 disciplines de 1970 à 2020

Le Figure 1 nous montre comment le nombre de thèses réparties en 15 disciplines évolue pendant la période 1970 - 2020. Avant 1983, le nombre de thèses dans toutes les 15 disciplines est négligeable (quelques unités). Après cette année, ce nombre total augmente jusqu'à 11 000 en 5 ans et fluctue autour de ce chiffre jusqu'en 2019. Il chute à zéro en 2020 par effet de bord. Cette tendance peut être observée dans les 15 disciplines. Celles-ci sont différenciées par un code couleur. Il permet de voir que la répartition du nombre de thèses entre ces disciplines évolue au fur et à mesure des années. Certaines disciplines ont un

nombre de thèses plus important que les autres comme *la Biologie, les Matériaux, Milieux et Chimie*, certaines disciplines comme *Informatique* et *SHS* voient leur importance augmenter (nombre de thèses relatif), tandis que d'autres décroissent comme la *Médecine*,



sans doute en raison d'un numerus clausus constant, qui fixe le nombre de thèses annuel en médecine, tandis que la demande industrielle tend à augmenter la production de thèses dans les autres domaines.

Figure 2 - Capture d'une visualisation de la répartition du nombre de thèses par genre de l'auteur dans 15 disciplines au fil du temps (1971 à 2019).

La Figure 2 est une visualisation d'un graphique interactif qui représente le pourcentage de thèses par genre de l'auteur dans 15 disciplines en 2013. Sur cette figure, le pourcentage de femmes qui soutiennent une thèse est inférieur à celui des hommes dans la majorité des disciplines. On peut notamment relever une forte majorité masculine dans les domaines *Matériaux, Milieux et Chimie*, *Informatique*, *Science de l'ingénieur* et *Mathématiques et Informatique*. Par contre, la majorité des auteurs est féminine dans le domaine *Langues et Littératures*. Les autres domaines sont à peu près équilibrés. Ces résultats sont cohérent avec les injonctions sociétales qui éloignent les femmes des métiers technico-scientifiques.

## 2.2 Visualisation des Données des Vols vers et depuis la Russie

Au sein de la base de données des vols internationaux vers et depuis la Russie au mois de Février 2022, on trouve les informations sur les dates des vols, les deux points de coordonnées (latitude et longitude) correspondant à l'origine et la destination des vols. Pour

contraster les différences des vols liées à la guerre, j'ai choisi deux dates : 16 Février et 28 Février qui maximisent la différence.

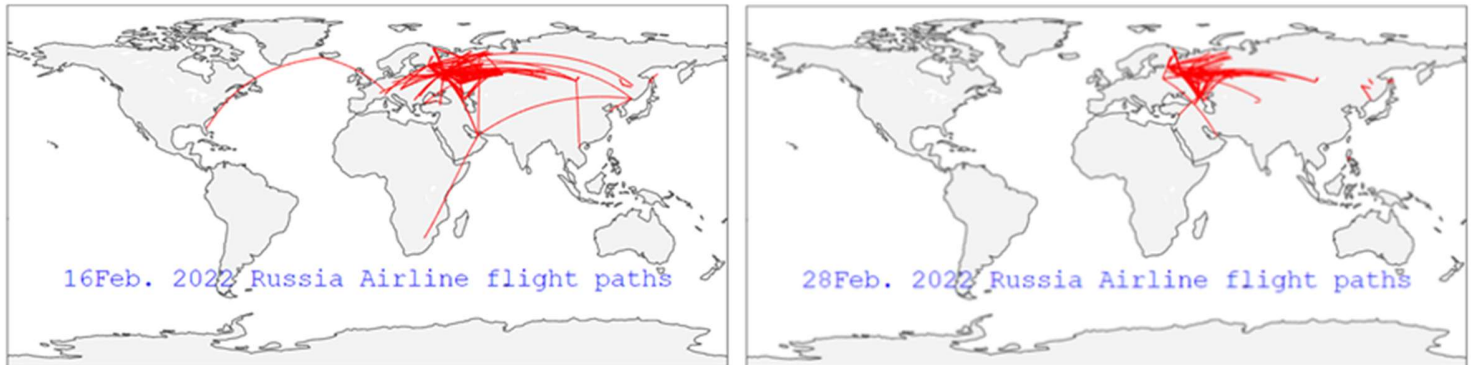


Figure 3 Les vols vers et depuis la Russie le 16 (à gauche) et le 28 février (à droite) 2022.

La Figure 3 représente les vols vers et depuis la Russie avant (gauche) et après (droit) la guerre. Avant la guerre, les vols sont centrés sur la Russie et s'étendent partout dans le monde, contrairement à après la déclaration de guerre, où les vols sont concentrés à l'intérieur de la Russie, suite aux sanctions géopolitiques prises à l'égard de la Russie après la déclaration de guerre à l'Ukraine.

### 3. Conclusion

En utilisant deux base de données, trois représentations graphiques ont été produites. Avec la base de données des thèses doctorales en France, on a montré l'évolution de nombre de thèses réparties dans 15 disciplines et le pourcentage de thèses par genre de l'auteur dans 3 disciplines sous la forme d'un graphique interactif. Il est évident que le graphique interactif est plus avantageux quand les informations sont multi-dimensionnelles, car il concentre plus d'informations dans peu d'espace et les illustre de façon plus intuitive. Par contre, le graphique simple présente une version synthétique des résultats nécessitant une analyse logique pour en extraire les tendances et faire parler les chiffres. La visualisation des données des vols vers et depuis la Russie est représentée par un graphique géographique qui correspond au contexte spécifique de l'étude et rend la première conclusion plus facile à faire. Pour aller plus loin, il est possible à partir des données actuelles de définir et calculer des indicateurs permettant la validation des hypothèses d'interprétation de ces mêmes données. Par exemple, l'étude du ratio entre les vols nationaux et les vols internationaux validerait la fermeture des échanges aériens à partir de la déclaration de guerre. De même, le ratio entre les vols vers et depuis les pays de l'OTAN et les vols internationaux vers et depuis la Russie montrerait l'influence de la géopolitique sur le trafic aérien. Enfin, une étude des vols intérieurs au cours du temps pourrait mettre en évidence la gestion aérienne de l'effort de guerre.

## Annexe

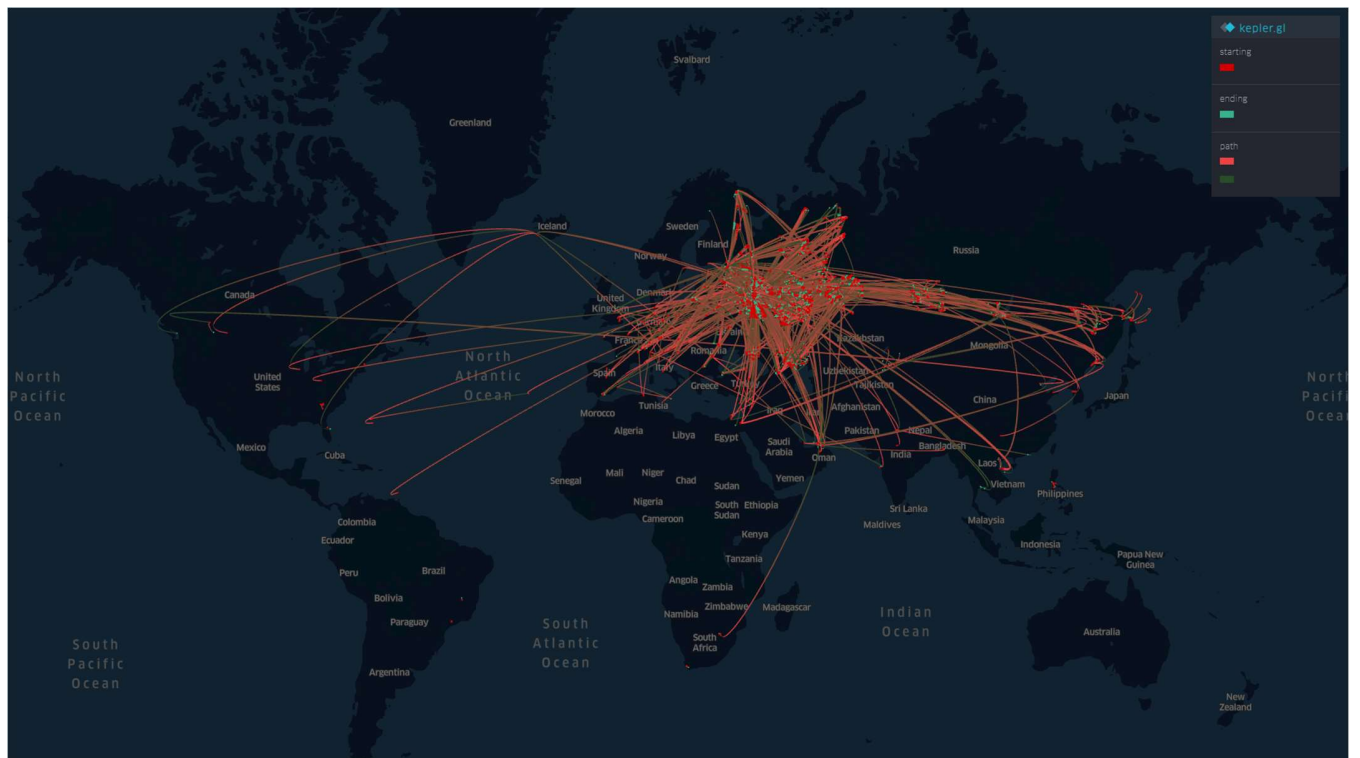


Figure 4 une capture d'écran pour les données des vols vers et depuis la Russie réalisée par le site kepler.gl

## Bibliographie

1. R. Rossi, Data Science in Perspective, Information Society Conference (i-Society), 121-126,2021
2. Schmidt, Johanna. Visual Data Science. 10.5772/intechopen.97750,2022