
Visualisation de données

Diplôme Universitaire Data Analyst enseigné
par **Dr Matthieu Cisel**

Version 3.0

6 juin 2022

Table des matières

1	Esthétique	3
1.1	Choix des disciplines	3
1.2	Analyse	3
1.3	Interprétation	4
2	Figure interactive	5
2.1	Analyse	5
2.2	Interprétation	5
3	Données spatialisées	6
3.1	Choix des dates	6
3.2	Analyse	6
3.3	Représentation des vols	7
3.4	Source de données	8
3.5	Interprétation	8

Liste des figures

1	Évolution du nombre de thèses soutenues entre 1985 et 2018 pour les disciplines biologie, sciences de la terre et médecine dans le jeu de données theses.fr	3
2	Entre 1985 et 2018, par disciplines regroupées en domaine, pourcentage de thèses rédigées en anglais dans le jeu de données theses.fr, la discipline au plus fort pourcentage est en haut du graphique. N compris entre 1 et 3926.	5
3	Évolution du nombre de vols civils depuis et vers la Russie du 20 au 28 février 2022	6
4	Représentation du trafic aérien civil en partance et à destination de la Russie lundi 21 février et lundi 28 février 2022	7

Liste des tableaux

1	Regroupement des 15 disciplines en six grands domaines	3
---	--	---

1 Esthétique

1.1 Choix des disciplines

Afin de dégager des tendances générales, les 15 disciplines ont été regroupées en six grands domaines. Ce regroupement est visible dans le Tableau 1 et sur les figures présentées dans ce rapport. Chaque domaine comprend entre une et cinq disciplines. Le regroupement a pour effet de faciliter la lecture et l'analyse des figures. Le choix esthétique de la Figure 1 est fait en faveur de l'accessibilité par le choix d'une palette de couleurs adaptée aux daltoniens.

Discipline	Domaine
Biologie	Sciences naturelles
Médecine	Sciences naturelles
Science de la Terre	Sciences naturelles
Sciences Humaines et Sociales	Sciences humaines
Psychologie	Sciences humaines
Sciences de l'éducation	Sciences humaines
Matériaux, Milieux et Chimie	Sciences de l'ingénieur
Science de l'ingénieur	Sciences de l'ingénieur
Mathématiques	Sciences de l'ingénieur
Informatique	Sciences de l'ingénieur
Mathématiques et Informatique	Sciences de l'ingénieur
Langues et Littératures	Langues et Littératures
Droit et Science Politique	Economie et politique
Économie Gestion	Economie et politique
Divers	Divers

TABEAU 1 – Regroupement des 15 disciplines en six grands domaines

Le Tableau 1 représente le regroupement des 15 disciplines en six grands domaines. Cinq disciplines sont regroupées dans le domaine "sciences de l'ingénieur". Trois disciplines sont regroupées dans chacun des domaines : "sciences naturelles" et "sciences humaines". Deux disciplines sont regroupées dans le domaine "économie et politique". Les disciplines "Langues et littératures" et "Divers" constituent chacune un domaine à part entière.

1.2 Analyse

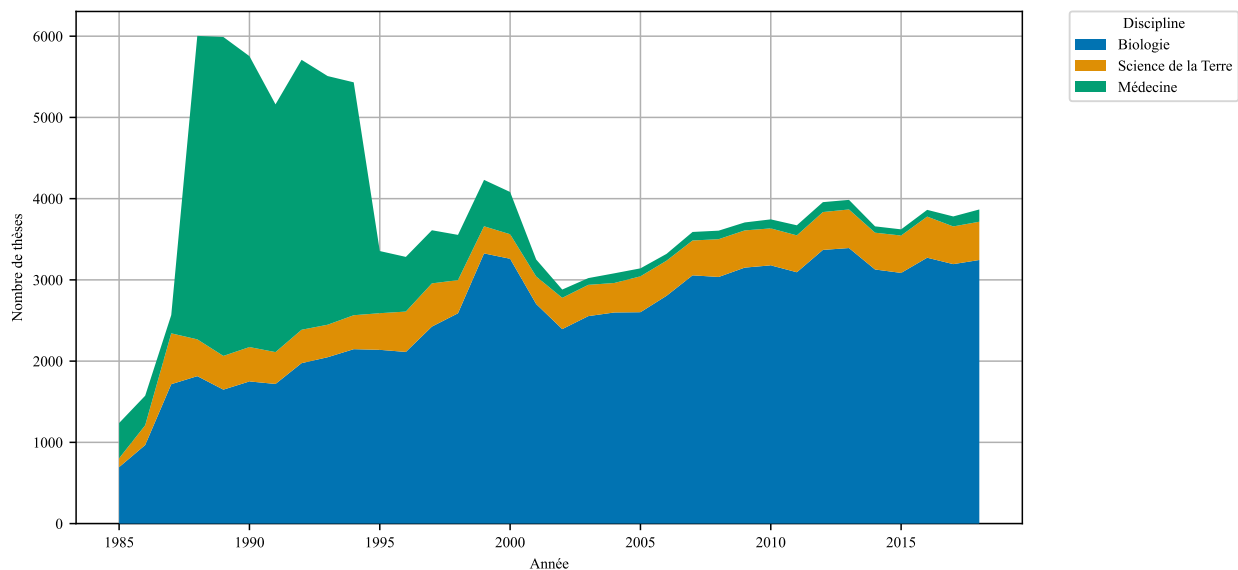


FIGURE 1 – Évolution du nombre de thèses soutenues entre 1985 et 2018 pour les disciplines biologie, sciences de la terre et médecine dans le jeu de données theses.fr

La Figure 1 représente l'évolution entre 1985 et 2018 du nombre de thèses pour les trois disciplines du domaine des sciences naturelles : "Biologie", "Science de la terre" et "Médecine". La Figure 1 permet de voir une augmentation du nombre de thèses soutenues en biologie où le nombre de thèses passe de 695 en 1985 à 3245 en 2018. L'augmentation est rapide de 1985 à 1988 (1815). Elle se poursuit moins rapidement jusqu'en 1998 (2139). Un pic est observé entre 1999 (3326) et 2001 (2703). La tendance à l'augmentation reprend plus doucement entre 2002 (2394) et 2018 (3245). On note un petit creux en 2011 (3094) suivi d'un petit pic en 2012 (3369) et 2013 (3392). Depuis 2016 la tendance semble être à la stagnation. Concernant le nombre de thèses soutenues en sciences de la terre, la Figure 1 permet de voir une augmentation rapide du nombre de thèses entre 1985 (109) et 1987 (626), année du plus grand nombre de thèses. Le nombre de thèses oscille ensuite autour d'une moyenne annuelle de 421 avec un minimum à 302 en 2000 et un maximum à 532 en 1997. Pour la médecine, on observe un volume annuel moyen de 341 entre 1985 et 1987. Le volume augmente considérablement entre 1988 (3736) et 1994 (2867). Ce volume décroît ensuite brutalement : 765 en 1995 et continue à décroître jusqu'en 2001 (209). A partir de 2002 et jusqu'en 2018, le volume oscille autour d'une moyenne de 104 avec un minimum de 76 (en 2015) et un maximum de 151 (en 2018).

1.3 Interprétation

La biologie est la discipline comptant le plus grand nombre de thèses avec une moyenne annuelle de 2535. Le volume des thèses soutenues dans cette discipline a fortement augmenté depuis 1985 et jusqu'en 2016. Le pic observé entre 1999 et 2001 pourrait être lié à la mise en oeuvre d'un plan stratégique de développement des sciences du vivant. L'[archive "Recherche ; Direction générale de la recherche et de la technologie ; Mission scientifique universitaire ; Département biologie, médecine, santé \(1993-1998\)"](#) mentionne que « Le département scientifique pédagogique et technique dénommé "Biologie, médecine, santé" eut pour but, entre 1993 et 1998, de réunir et de coordonner les éléments nécessaires à l'élaboration d'un plan stratégique de développement des sciences du vivant, en concertation avec les organismes de recherche et les établissements d'études supérieures concernés. » Ce programme ainsi que différentes mesures européennes comme la [Directive 98/44/CE du parlement européen et du conseil du 6 juillet 1998 relative à la protection juridique des inventions biotechnologiques](#) ainsi que les intérêts de grands groupes internationaux implantés en France Sanofi, Pfizer,... pourraient avoir favorisé la recherche biologique dans cette période. Le volume semble stagner à partir de 2016. Le « [Rapport d'information sur la recherche publique et privée en France face au défi international](#) » déposé en 2004 à l'assemblée nationale par la commission des affaires culturelles, familiales et sociales, souligne les difficultés de la recherche française dans le contexte international et notamment la fuite vers les pays étrangers des cerveaux faisant face à des conditions de travail peu valorisantes en France. Mais aussi la fuite des entreprises, par exemple le déménagement des centres de recherche du laboratoire Pfizer de France vers le Royaume-Uni. Ces éléments pourraient mettre en danger l'activité de recherche française. L'évolution du nombre de thèses soutenues dans cette discipline est un indicateur à surveiller dans les prochaines années. Il serait notamment intéressant d'étudier la part de recherche publique et privée en biologie et la part privée financée par les grands groupes internationaux.

Pour les Sciences de la Terre, le volume observé est relativement faible (moyenne annuelle de 421 thèses soutenues) et stable dans le temps. Les enjeux environnementaux actuels devraient conduire à l'augmentation des travaux doctoraux dans cette discipline, même si le lien avec d'autres disciplines peut conduire les travaux doctoraux à une classification dans une discipline alternative. Ainsi le [travail doctoral](#) de Alice Guille des Buttes portant sur l'« Optimisation conjointe de la consommation d'essence et des émissions de polluants réglementés pour un véhicule hybride essence-électrique d'architecture parallèle » est classé de façon justifiée dans la discipline des "Sciences de l'ingénieur", mais il inclut une dimension environnementale.

Concernant la médecine, les études de médecine se terminent par une courte thèse d'exercice qui dure habituellement entre six mois et un an. La Figure 1 permet d'émettre l'hypothèse selon laquelle les courtes thèses soutenues entre 1988 et 1994 ont été enregistrées sur le site theses.fr. Ces thèses constituent un travail moindre par rapport aux thèses produites dans le cadre d'un doctorat. Ce volume important de thèses a donc un effet de faux positif. Il conviendrait de remplacer le pic artificiel par une moyenne annuelle recalculée et faisant le lien entre le volume observé en 1987 et le volume observé en 1995. Entre 2002 et 2018, le nombre annuel moyen de thèses en médecine est de 104. Ce volume semble représenter le nombre réel de thèses soutenues dans cette discipline. L'étude du [guide des "Thèses autres écrits académiques identification et localisation"](#) publié en 2005 par la BnF et des ressources référencées permettrait de compléter et fiabiliser les données du site theses.fr antérieures à 2011. Le travail nécessite la consultation de nombreux ouvrages et micro fiches et semble toutefois fastidieux. Ce travail constitue-t-il lui-même un sujet de doctorat ?

2 Figure interactive

2.1 Analyse

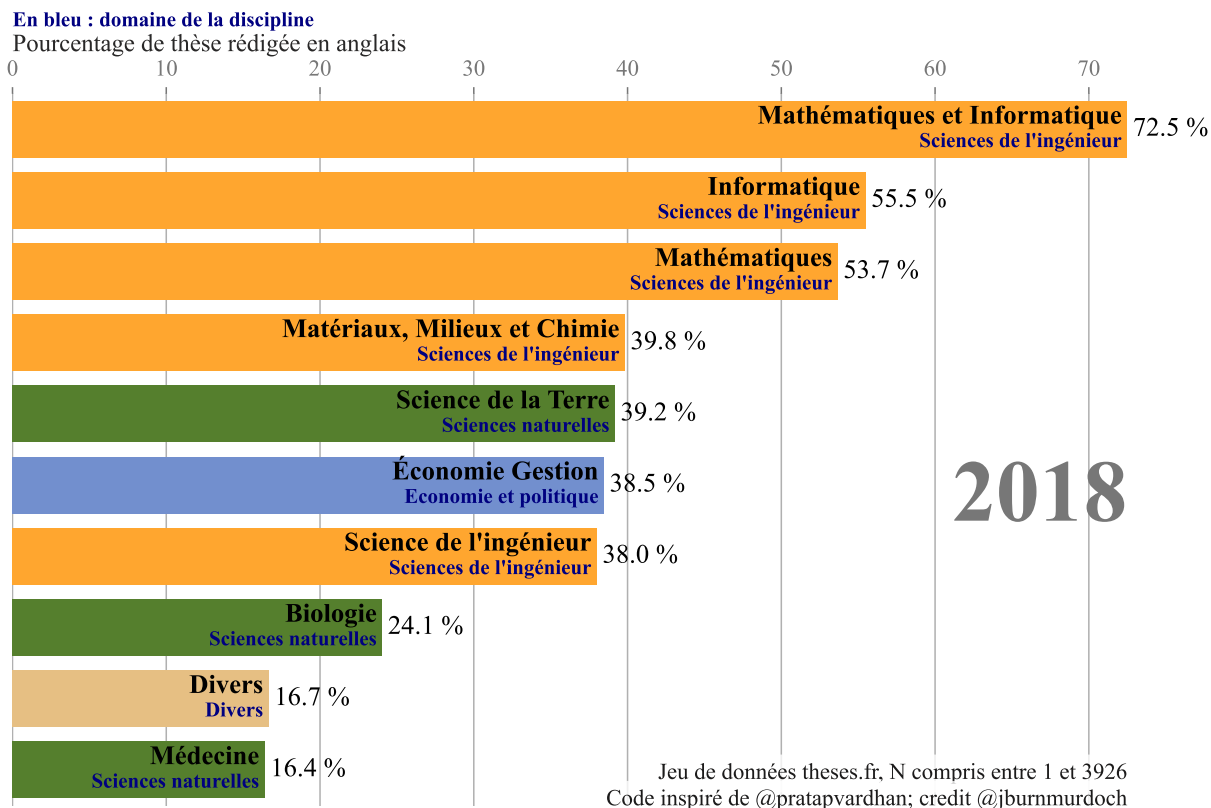


FIGURE 2 – Entre 1985 et 2018, par disciplines regroupées en domaine, pourcentage de thèses rédigées en anglais dans le jeu de données theses.fr, la discipline au plus fort pourcentage est en haut du graphique. N compris entre 1 et 3926.

La Figure 2 représente le pourcentage de thèses rédigées en anglais par discipline et chaque année entre 1985 et 2018. Les disciplines sont regroupées en domaine pour analyse de tendances macroscopiques. 14 années sur 34 années observées, les trois disciplines pour lesquelles le plus fort pourcentage de thèses rédigées en anglais est observé appartiennent exclusivement au domaine des sciences de l'ingénieur. Pour 18 autres années, les trois premières disciplines appartiennent soit au domaine sciences de l'ingénieur, soit au domaine sciences naturelles. Les deux années restantes font figure d'exception : en 2009, la discipline Divers arrive en deuxième position avec un pourcentage de 25% (N = 4) et en 2003, la discipline économie et politique arrive en troisième position avec un pourcentage de 4,6% (N = 699). Le plus fort pourcentage (72%) est observé en 2018 pour la discipline mathématiques et informatique (N = 40).

2.2 Interprétation

La Figure 2 permet de vérifier que la rédaction de thèses en anglais est pratiquée principalement dans les domaines des sciences de l'ingénieur et des sciences naturelles. Comme indiqué dans l'article "[Le français perd-il du terrain dans les thèses de doctorat ?](#)" publié le 6 novembre 2016 par The conversation, la publication en anglais est un des facteurs de visibilité pour les docteurs. Il semble plus naturel que les travaux relatifs aux domaines scientifiques aient un intérêt à l'échelle internationale tandis que les travaux du domaine langues et Littératures restent relatifs à la langue française. Ces données sont toutefois à examiner avec prudence. Suivant les années et les disciplines, les pourcentages sont relatifs à des échantillons de tailles disparates et parfois en nombre très limité. N est compris entre 1 et 3926. Les analyses menées sur des échantillons de taille réduite manquent de fiabilité. Les comparaisons entre échantillons de tailles disparates peuvent induire des écarts non significatifs. Il serait pertinent de limiter les comparaisons à des échantillons respectant une taille minimale. N minimum = 100 par exemple.

3 Données spatialisées

Dans ce chapitre sont observés les vols civils à destination et depuis la Russie dans la période qui entoure le 24 février 2022, date du début de la guerre déclarée par la Russie à l'Ukraine. L'objectif est d'étudier un impact éventuel du début de la guerre sur le trafic aérien. Le premier travail consiste à trouver les dates qui vont servir pour la comparaison.

3.1 Choix des dates

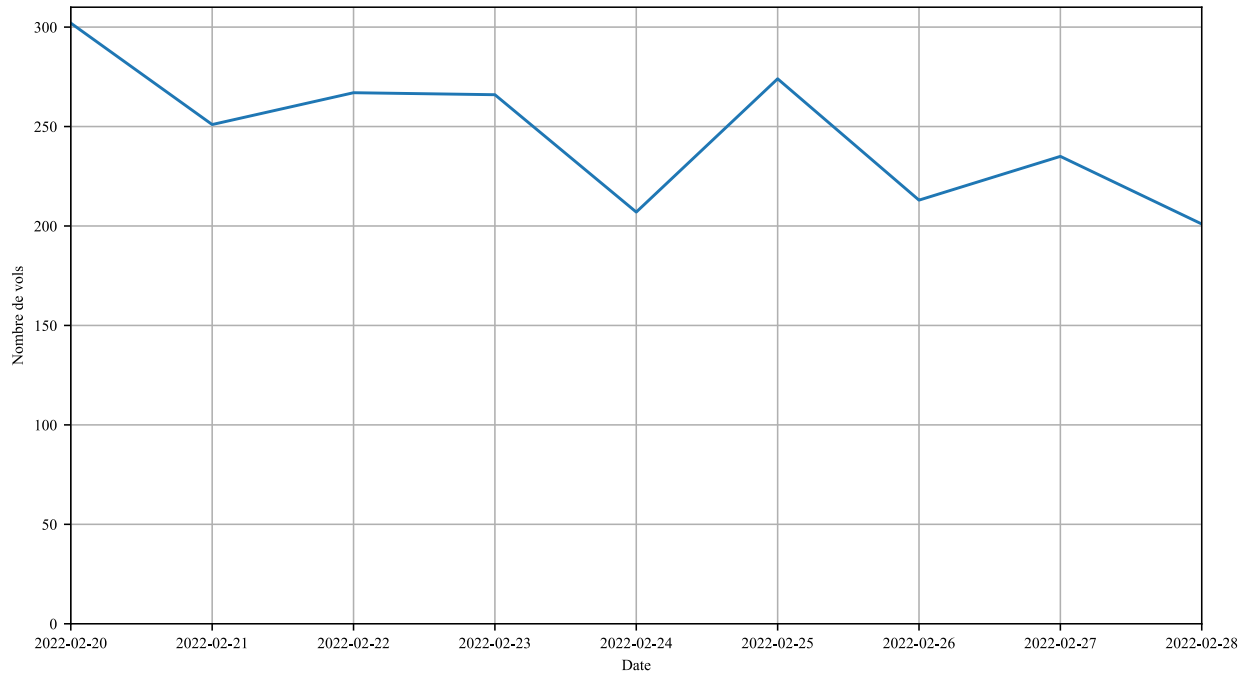


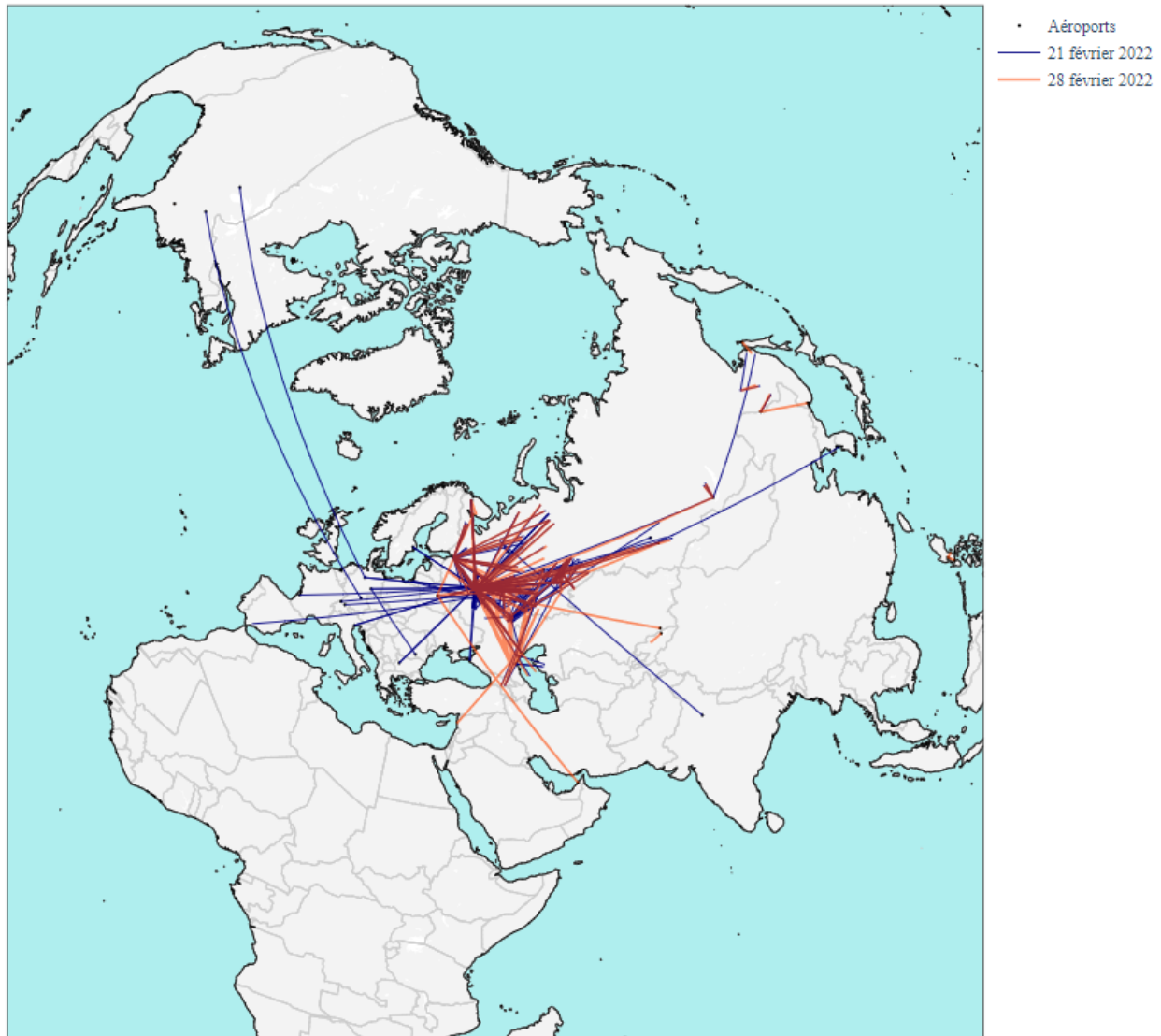
FIGURE 3 – Évolution du nombre de vols civils depuis et vers la Russie du 20 au 28 février 2022

La Figure 3 permet de voir que le volume quotidien des vols subit de légères variations sur la période du 20 au 28 février 2022 avec une tendance à la baisse. Sur cette période, le nombre de vols quotidiens est compris entre 300 dimanche 20 février 2022 et 200 lundi 28 février. On observe un premier creux (250) lundi 21 février, puis un nouveau creux (208) jeudi 24 février, suivi d'un pic vendredi 25 février (272). Lundi 28 février (200 vols), on observe une diminution de 20% du volume des vols observé lundi 21 février (250 vols).

3.2 Analyse

Le jour de la semaine peut impacter le trafic aérien civil. On observe par exemple des pics dans le trafic aérien les vendredis et les dimanches. Pour cette raison, on choisit de comparer le trafic aérien sur le même jour de la semaine. Pour limiter les variations liées aux déplacements des weekends, on évite de comparer le volume observé les vendredis, samedis et dimanches. Le jeu de données fourni couvre la période qui va jusqu'au lundi 28 février. Ainsi les dates de comparaison retenues sont les deux lundis les plus proches du 24 février 2022, soient le lundi 21 et le lundi 28 février 2022.

3.3 Représentation des vols



Utiliser la souris pour afficher les informations des aéroports

FIGURE 4 – Représentation du trafic aérien civil en partance et à destination de la Russie lundi 21 février et lundi 28 février 2022

La Figure 4 permet d'observer une limitation dans les aéroports en partance et à l'origine des vols russes le 28 février 2022 en comparaison avec les aéroports en partance et à l'origine des vols russes le 21 février 2022. 14 aéroports situés à l'ouest de la Russie sont la destination ou la ville d'origine d'un vol à destination ou en partance d'un aéroport russe lundi 21 février 2022 : Chicago, Colombus, Anderson, Stockholm, Bordeaux, Zurich, Bâle, Hanovre, Leipzig, Nuremberg, Sofia, Bucharest, Venise et une ville du sud de l'Espagne. Ces aéroports ne figurent pas dans la liste des aéroports à destination ou en partance d'un aéroport russe lundi 28 février 2022. Lundi 21 février, New Delhi est en relation avec un vol à destination ou en partance d'un aéroport russe contrairement à lundi 28 février. Lundi 28 février, à l'exception de trois vols en partance ou à destination de Minsk (Biélorussie), l'ensemble des vols est à destination ou en partance d'un aéroport situé en Russie, à l'Est ou au sud de la Russie. Lundi 28 février, Dubaï, Almaty (Kazakhstan), Balyktchy (Kirghizstan), une ville de Syrie et deux villes de Géorgie sont en relation avec un vol à destination ou en partance d'un aéroport russe. Les autres vols sont internes à la Russie.

3.4 Source de données

Le jeu de données fourni résulte d'un tri réalisé dans les données aériennes. De grandes tendances sont observées. Il reste des particularités : deux vols depuis ou vers la mer caspienne et un vol qui semble relier la ville d'Anderson à elle-même. Ces étrangetés mériteraient une analyse plus poussée des données fournies et invitent à la prudence sur les détails observés.

3.5 Interprétation

La Figure 3 met en évidence une diminution de 20% du volume des vols entre le 21 février et le 28 février 2022. De plus la Figure 4 met en évidence une diminution des vols occidentaux en partance et en provenance de la Russie entre le 21 février et le 28 février 2022. Une des sanctions prises par les pays occidentaux lors de l'annonce de la guerre de la Russie à l'Ukraine a été la fermeture des espaces aériens occidentaux aux avions russes. Cette sanction a été suivie de la fermeture miroir par la Russie de son espace aérien aux avions des pays occidentaux. La presse a relayé ces informations dès les premiers jours après le début de la guerre comme nous pouvons le voir dans [cet article du 27 février 2022 paru dans Le Monde](#). La diminution observée des vols civils entre les pays occidentaux et la Russie est donc une conséquence directe du début de la guerre menée par la Russie en Ukraine. Malgré l'absence de vol entre l'Inde et la Russie le 28 février 2022 alors qu'un vol est visible entre ces deux pays le 21 février, le site numérama indique dans [cet article paru le 16 mars 2022](#) qu'«il y a toujours des liaisons aériennes entre New Delhi et Moscou».

Parmi les impacts du conflit, ce [même article](#) documente en images les détours rendus obligatoires par la fermeture de l'espace aérien russe. Les avions sont forcés d'emprunter, dans leurs liaisons avec l'Asie de l'Est, de nouvelles routes contournant la Russie par le sud. Ces nouveaux trajets ont des impacts sur l'allongement des parcours, le coût des billets et l'affluence dans le ciel aérien. L'ensemble des impacts est à suivre à l'avenir avec plus de recul et de nouvelles données.