

Data. Sont-elles représentatives ? Sont-elles fiables ?

1. Objectif initial

On peut utiliser la base de données de généalogie mathématique de l'université du Dakota du Nord (dite MGP, pour "Mathematical Genealogy Project")¹ pour essayer de se faire une petite idée de la répartition des mathématiciennes et mathématiciens sur le globe, par le pays de leur thèse, en faisant une recherche en ne renseignant que le champ pays.

On obtient ces données pour l'Europe :

Pays	Nb. Enreg. dans MGP	Pays	Nb. Enreg. dans MGP
Allemagne	38883	Pays bas	23719
Royaume Uni	18359	France	16342
Suisse	7883	Espagne	4760
Italie	4004	Pologne	3045
Belgique	2667	Autriche	2520
Suede	2500	Ukraine	2364
Finlande	1328	Roumanie	1157
Danemark	1007	Portugal	1003
Serbie	829	Irlande	791
Norvege	777	Hongrie	760
Grece	689	Slovenie	369
Slovaquie	355	Lituanie	329
Croatie	253	Bulgarie	241
Armenie	204	Moldavie	144
Estonie	143	Bielorussie	125
Luxembourg	125	Georgie	108
Bosnie Herzegovine	73	Lettonie	24
Islande	19	Montenegro	11
Albanie	7	Andorre	—
Azerbaïdjan	—	Chypre	—
Lichtenstein	—	Kazakhstan	—
Macedoine du nord	—	Malte	—
Monaco	—	Saint Morin	—
Tchequie	—	Turquie	—
Vatican	—		

Soit un total de 137917 enregistrements sur 335441 enregistrements recensés (en date du 14 novembre 2025 ; on n'a considéré que les lignes pour lesquelles les noms étaient écrits en alphabet latin, et non pas en idéogrammes, chinois par exemple, alors que la base peut peut-être en contenir). Ce qui correspondrait à un ratio de 41 % des enregistrements dont les thèses auraient été obtenues en Europe.

Et on obtient ces retours sur certains autres pays.

1. <https://genealogy.math.ndsu.nodak.edu/>

Pays	Nb. Enreg. dans MGP	Pays	Nb. Enreg. dans MGP
Etats Unis	137653	Chine	3629
Inde	3823	Nouvelle Zelande	448
Japon	1816	Vietnam	164
Russie	5894	Iran	1679
Canada	9220	Arabie Saoudite	159
Liban	44	Mexique	1052
Venezuela	144	Argentine	869
Uruguay	46		

On ne sait pas si ces données sont effectivement représentatives, i.e. si beaucoup de mathématiciennes et mathématiciens mondiaux auraient renseigné les fiches de manière à rendre la base de données la plus exhaustive et représentative de la réalité possible.

Pour les thèses soutenues en France, toujours selon cette base de données MGP², on essaye de comptabiliser les enregistrements correspondant à des thèses soutenues par des femmes, on utilise un fichier de recensement de 1524 prénoms féminins³.

On obtient ces ratios :

identif. de l'enreg. dans MGP	Nombre de fichiers	Nb. Fm.	ratio
entre 0 et 100000	3126	348	11.13 %
entre 100000 et 200000	4985	541	10.85 %
entre 200000 et 300000	4740	699	14.74 %
entre 300000 et 335441	1598	241	15.08 %
tous enreg.	14449	1829	12.65 %

On peut aussi utiliser les données ouvertes fournies par Data.gouv⁴, qui fournissent entre autres le titre de toutes les thèses soutenues en France de 1985 à 2025, et recenser par la même méthode les thèses féminines parmi celles dont le nom du domaine commence par “Mathématiques” ou “Logique” (on trouve 12878 enregistrements en date du 18 novembre 2025). En faisant cela, et en compulsant également les thèses en informatique ou théorie de l’information, on réalise que les thèses d’informatique sont plus nombreuses pour la période 1985-2025 que les thèses de mathématiques (18998 enregistrements en date du 18 novembre 2025). Par l’analyse des prénoms, on trouve 2285 thèses féminines, soit un ratio de 17.74%.

Dans la page des “Female mathematicians” du site aussi exhaustif qu’il est possible Mac-Tutor, ici <https://mathshistory.st-andrews.ac.uk/Biographies/category-women/>, on trouve 372 pages consacrées à des mathématiciennes ; mais dans la page consacrée aux mathématiciens et mathématiciennes français, seulement 2% environ des entrées renvoient vers des pages de mathématiciennes.

2. <https://www.mathgenealogy.org/>. Ce site utilisant des cookies google, ses concepteurs peuvent s’ils le souhaitent savoir quelles sont les pages les plus visitées.

3. On trouve tout sur la toile : dans le palmarès des prénoms les plus donnés en France depuis 1900, mon prénom, Denise, est à la 19^{ième} place où il s’avère avoir été le plus donné de 1925 à 1926.

4. Voir à cette page <https://www.data.gouv.fr/datasets/theses-soutenues-en-france-depuis-1985/>

Concernant les thèses, on ne parvient pas à établir un pourcentage précis : les données sont trop délicates à répertorier : il y a des erreurs potentielles dans la liste des prénoms (oublis, ou “faux positifs”). Il y a les prénoms “épiciens”, pour utilisables indifféremment pour les deux genres (comme Claude, ou Dominique). Il y a toute la problématique actuelle autour du genre à considérer, même si cela ne doit pas servir de prétexte du style “on ne peut rien mesurer du tout”, qui profite toujours aux dominants. Il me semble qu’un pourcentage à la louche compris entre 10 % et 15 % semble avoisiner assez bien le réel ; prenons-le de 12 %, ce qui correspond à plus de 5 mathématiciens français pour une mathématicienne française.

Un autre problème, et de taille, est qu’en comparant les thèses recensées dans MGP et celles postérieures à 1985 recensées dans data.gouv (cette dernière source étant à considérer comme une source fiable, dans la mesure où les thèses sont soumises à obligation de recensement, ainsi qu’à nécessité d’accepter que les données soient publiques, avec impossibilité de les retirer une fois publiées - on trouve cette information dans la notice associée à la page des thèses, dans data.gouv), il y a la broutille de 1500 et quelques thèses qu’on trouve dans MGP et pas dans data.gouv. S’agit-il toutes de thèses publiées avant 1985 (cela semble un nombre énorme cependant, compte-tenu de l’historique assez bien connu du nombre de thèses féminines) ? S’agit-il de leurres ?

Si le problème de la représentation des mathématiciennes est avéré dans les données data.gouv, aurait-on des idées de solutions pour corriger le désastre ? Si l’origine de la conception d’un célèbre réseau social par un jeune informaticien, telle qu’elle est présentée dans le film *Social network*, est véridique, (il s’agissait de noter les étudiantes de son université par une simple comparaison physique “plus belle”, “plus moche”), on comprend qu’un tel concept peut décourager toute femme de mettre sa photo sur les réseaux. Pourtant, c’est peut-être là que l’invisibilisation commence : essayez de rechercher certains mathématiciens contemporains sur la toile et vous serez étonnés du nombre de leurs photos. Peut-être que le souhait initial à l’origine de cette profusion était une volonté de rendre plus visibles des mathématiques : “il faudrait que les mathématiques aient leurs “stars””, de même qu’en ont les sports collectifs, les arts, la chanson, la politique. Il faudrait maintenant tenter d’égaliser cette “visibilisation des mathématiciens”, cependant.

On tient à attirer l’attention ici sur le fait qu’on trouve très malheureux d’avoir à faire ce genre de calculs, destinés à évaluer au mieux la représentation des mathématiciennes de tel ou tel pays, dans cette base de données très particulière qu’est la base de données MGP. Le but des concepteurs de cette base était extrêmement louable (outre le fait que c’est amusant de réaliser qu’un chemin relie une tutrice de thèse à Euler !). D’autre part, permettre à tous de la mettre à jour, pour équilibrer au mieux la portion de travail à effectuer par chacun(e), est une sacrément bonne idée. Cependant cette base, comme toute base de données d’ailleurs, peut être utilisée à diverses fins, selon l’intention de celui ou celle qui en a connaissance. On insiste sur le fait que le but ici était d’essayer, autant que faire se peut, de calculer la représentation des mathématiciennes françaises au sein des mathématiques françaises, en utilisant la base de données MGP.

2. Rectification le 27 novembre 2025

On a réalisé que la liste des prénoms n’allait pas du tout. Elle est trop courte, trop de thèses sortent de la raquette. On récupère de ce fait une nouvelle liste de prénoms de l’Insee beaucoup

plus conséquente, contenant au départ 28080 prénoms féminins, sans doublons. Malheureusement, on s'aperçoit, et cela est d'ailleurs expliqué en accompagnement des données Insee, qu'un certain nombre d'exceptions ou d'erreurs administratives font que des prénoms masculins ont parfois été attribués à des naissances de filles dans les prénoms donnés dans les actes de naissances de l'état civil.

Il faut bien mesurer tout ce que la présente étude peut avoir de vain, voire de franchement (c'est le cas de le dire) non convenable, dans ces temps où l'on se bat pour plus d'inclusion, plus de genres, autant de causes que beaucoup devraient défendre. Mais vaille que vaille, le but restant de mesurer la représentation des mathématiciennes dans la communauté mathématique française, en supprimant (avec une certaine tristesse symbolique) les prénoms qui "tout de même semblent bien être des prénoms masculins", on arrive à un ratio féminin de 3611 thésardes ayant obtenu leur thèse en France depuis 1985 sur 18999 thèses obtenues en France depuis 1985, selon le fichier datagouv, ce qui correspond à 19 %. C'est mieux qu'une fourchette entre 10 et 15 %, mais ça reste inférieur à 1/5.

3. Résultat intermédiaire le 30 novembre 2025

Après un dernier traitement aussi précis qu'il est possible (voir l'annexe B qui explique ce traitement du genre) compte-tenu des données (pour calculer le ratio, on fixe qu'un prénom est un prénom féminin ou masculin suivant qu'il a été attribué de plus nombreuses fois à une petite fille ou à un petit garçon à la naissance, dans les données Insee depuis 1900 fournies à cette page,

<https://www.insee.fr/fr/statistiques/8595130?sommaire=8595113>

voir notamment le fichier qui fournit pour chaque prénom combien de fois il a été attribué telle ou telle année, et on calcule des cumuls toutes années confondues). On obtient le dernier ratio qu'on calcule au sujet de la représentation des mathématiciennes thésardes parmi l'ensemble des thésards et thésardes français recensés dans la base de données MGP :

$$\frac{2739}{15913} \approx 17.21 \%$$

4. Suppression des doublons, le 6 décembre 2025 et tentative d'extrapoler la liste des prénoms féminins à d'autres pays pour estimer sa couverture

On fournit ci-dessous par pays les nombres de thésards et thésardes recensés par pays dans la base MGP en première colonne numérique, le nombre de thésardes avec genre fixé comme précédemment, en utilisant à nouveau la base la plus complète des prénoms de l'Insee. Comme attendu, cette liste semble inadapté à certains pays ; cependant, les nombres obtenus étant semblable à ceux obtenus pour la France, beaucoup de prénoms fréquents dans ces pays doivent déjà faire partie de la liste Insee (du fait des mouvements de personnes). On ne sait pas si ce serait un fait connu que des pays tels que la Roumanie ou l'Italie fasse la part belle aux thésardes en mathématiques. Pour l'Italie, on trouve cependant une explication qui fait un peu baisser le ratio, celui-ci restant élevé à 26.9 % : le prénom Daniele est un prénom qui, sans son accent, est un prénom féminin en France, mais masculin en Italie ; cela a eu comme conséquence que le système a compté davantage de thésardes qu'il n'y en a vraisemblablement ; de même pour le prénom Michele, souvent masculin en Italie.

Une professeure d’université nous fait remarquer que l’énorme différence entre le nombre de thèses en Allemagne et le nombre de thèses en France est étonnant. En étudiant les données, on réalise que les thèses allemandes (et il est bien spécifié dans les missions de la base généalogique mathématique MGP que cette base se veut très ouverte - notamment à l’application des mathématiques à d’autres sciences comme les sciences dites “naturelles”) contiennent de nombreuses thèses de biologie (le diplôme a alors son intitulé qui se termine par “rer.nat”, de philosophie (l’intitulé se termine par “phil”) : on ne conserve que les diplômes d’intitulé Ph.D. ou approchant (PhD par exemple) ; les nombres de thèses des différents pays sont en accord avec leur nombre d’habitants sauf pour les Pays-Bas qu’il faudra aussi analyser plus avant.

Concernant les thèses des Pays-Bas (Netherlands), leur grand nombre inexplicable par rapport à la taille du pays se justifie à l’étude par le fait qu’ont été incluses dans la base des thèses de médecine. On ne les inclut pas dans le calcul du total pour les pays d’Europe à nombre de mathématicien(ne)s supérieur à 1111.

Pays	Nb d’enreg dans le pays	Nb de femmes selon occu. liste Insee	Ratio
Allemagne	15262	1918	12.5 %
Autriche	2279	338	14.8 %
Belgique	2470	404	16.3 %
Espagne	4612	881	19.1 %
Finlande	1214	124	10.2 %
France	15173	2586	17.0 %
Italie	3579	1053	29.4 %
Pologne	2925	476	16.2 %
Roumanie	1080	303	27.8 %
Royaume Uni	16615	2218	13.3 %
Suede	2306	295	12.8 %
Suisse	7212	1175	16.2 %
Pays Bas	22271	1983	8.9 %
Total avec Pays Bas	96998	13754	14.1 %
Total sans Pays Bas	74727	11771	15.7 %
Japon	1684	67	3.9 %
Canada	8844	1159	13.0 %
Inde	3621	258	7.1 %

5. Résumé des dernières statistiques obtenues dans la base de données MGP

Le premier tableau fournit les résultats par pays dont les noms sont triés alphabétiquement.

Le second tableau fournit les mêmes données mais avec les pays triés selon leur ratio nombre de thèses associées à un prénom vraisemblablement féminin sur nombre de thèses.

Pays	These de prenom feminin	Nb de theses	ratio
Afrique du Sud	94	693	13.5 %
Algerie	140	336	41.6 %
Allemagne	1918	15262	12.5 %
Argentine	252	833	30.2 %
Armenie	24	189	12.6 %
Australie	432	3168	13.6 %
Autriche	338	2279	14.8 %
Belgique	404	2470	16.3 %
Bielorussie	25	118	21.1 %
Bosnie	25	73	34.2 %
Bresil	651	3670	17.7 %
Bulgarie	50	230	21.7 %
Canada	1159	8844	13.1 %
Chine	190	3253	5.8 %
Croatie	76	252	30.1 %
Danemark	85	922	9.2 %
Espagne	881	4612	19.1 %
Estonie	20	130	15.3 %
Finlande	124	1214	10.2 %
France	2586	15173	17.0 %
Georgie	16	100	16 %
Grece	86	670	12.8 %
Hongrie	56	664	8.4 %
Inde	258	3621	7.1 %
Iran	177	1602	11.0 %
Irlande	154	759	20.2 %
Islande	2	18	11.1 %
Israel	321	2235	14.3 %
Italie	1053	3579	29.4 %
Japon	67	1684	3.9 %
Lettonie	2	18	11.1 %
Lituanie	48	326	14.7 %
Luxembourg	13	109	11.9 %
Maroc	61	368	16.5 %
Mexique	175	1017	17.2 %
Moldavie	20	138	14.4 %
Montenegro	2	10	20 %
Norvege	81	714	11.3 %
Nouvelle Zelande	69	434	15.8 %
Pays Bas	1983	22271	8.9 %
Pologne	476	2925	16.2 %
Portugal	270	889	30.3 %
Roumanie	303	1080	28.0 %
Royaume Uni	2218	16615	13.3 %
Russie	684	4688	14.5 %
Serbie	206	806	25.5 %
Slovaquie	39	244	15.9 %
Slovenie	75	363	20.6 %
Suede	295	2306	12.7 %
Suisse	1175	7212	16.2 %
Tunisie	117	294	39.7 %
Ukraine	383	2195	17.4 %

Pays	These de prenom feminin	Nb de theses	ratio
Algerie	140	336	41.6 %
Tunisie	117	294	39.7 %
Bosnie	25	73	34.2 %
Portugal	270	889	30.3 %
Argentine	252	833	30.2 %
Croatie	76	252	30.1 %
Italie	1053	3579	29.4 %
Roumanie	303	1080	28.0 %
Serbie	206	806	25.5 %
Bulgarie	50	230	21.7 %
Bielorussie	25	118	21.1 %
Slovenie	75	363	20.6 %
Irlande	154	759	20.2 %
Montenegro	2	10	20.0 %
Espagne	881	4612	19.1 %
Bresil	651	3670	17.7 %
Ukraine	383	2195	17.4 %
Mexique	175	1017	17.2 %
France	2586	15173	17.0 %
Maroc	61	368	16.5 %
Belgique	404	2470	16.3 %
Suisse	1175	7212	16.2 %
Pologne	476	2925	16.2 %
Georgie	16	100	16.0 %
Slovaquie	39	244	15.9 %
Nouvelle Zelande	69	434	15.8 %
Estonie	20	130	15.3 %
Autriche	338	2279	14.8 %
Lituanie	48	326	14.7 %
Russie	684	4688	14.5 %
Moldavie	20	138	14.4 %
Israel	321	2235	14.3 %
Australie	432	3168	13.6 %
Afrique du Sud	94	693	13.5 %
Royaume Uni	2218	16615	13.3 %
Canada	1159	8844	13.1 %
Grece	86	670	12.8 %
Suede	295	2306	12.7 %
Armenie	24	189	12.6 %
Allemagne	1918	15262	12.5 %
Luxembourg	13	109	11.9 %
Norvege	81	714	11.3 %
Islande	2	18	11.1 %
Lettonie	2	18	11.1 %
Iran	177	1602	11.0 %
Finlande	124	1214	10.2 %
Danemark	85	922	9.2 %
PaysBas	1983	22271	8.9 %
Hongrie	56	664	8.4 %
Inde	258	3621	7.1 %
Chine	190	3253	5.8 %
Japon	67	1684	3.9 %

Annexes

A. Liste des prénoms féminins utilisés dans les premiers traitements effectués pour distinguer, autant que faire se peut, les thèses féminines dans l'ensemble des thèses

On a dans un premier temps utilisé une base de prénoms féminins compilée à partir de diverses listes trouvées sur la toile et qui contenait 1500 prénoms environ. À l'avenir, il faudrait utiliser la base des prénoms fournis sur datagouv et qui contient 96184 prénoms féminins et 87644 prénoms masculins utilisés en France de 1900 à aujourd'hui.

B. Ajout d'un champ genre aux entrées de la base MGP

Je remercie Jacques qui m'a aidée pour cette partie du traitement de la base de données.

Les champs standard du catalogue MGP sont complétés avec un nouveau champ nommé Genre, qui est déduit du prénom.

L'Insee fournit des données sur les prénoms attribués aux enfants nés en France depuis 1900 aux niveaux national, régional et départemental. Une liste des prénoms par sexe attribués au moins 3 fois depuis 1900 est également proposée.

Certains prénoms figurant dans ces listes sont attribués aux deux sexes. Il peut s'agir d'erreurs administratives, de choix des parents, ou de prénoms épiciènes, c'est-à-dire qui s'écrivent de la même manière au masculin et au féminin, comme Alex, Claude ou Sacha.

Pour déterminer le genre le plus probable en cas d'ambiguïté, on dénombre toutes les occurrences d'un prénom attribuées à chaque sexe depuis 1900, et on retient le genre présentant le plus d'occurrences.

C. Liste des domaines des thèses à dominante Mathématiques et/ou Informatique

On trouvera la liste des domaines des thèses en mathématiques ici et celle des domaines des thèses en informatique là.

D. Traduction de la page de la “Déclaration de mission”, de l'équipe de la base de données MGP (Mathematics Genealogy Project) consultable ici <https://www.mathgenealogy.org/mission.php>.

Il est noté dans la page News que le projet avait été initié par Harry B. Coonce, décédé le 14 février 2025.

Ce projet vise à compiler des informations sur tous les mathématiciens du monde. Nous sollicitons instamment des informations auprès de tous les établissements d'enseignement participant au développement de la recherche en mathématiques et de toute personne susceptible de détenir les informations recherchées.

Remarque : Dans le cadre de ce projet, les termes “mathématiques” ou “mathématicien” sont employés au sens large. Ainsi, toutes les données pertinentes en didactique des mathématiques, en statistiques, en informatique ou en recherche opérationnelle sont les bienvenues.

Dans les paragraphes suivants, nous présenterons nos objectifs et la philosophie qui sous-tend ce projet de généalogie. Notre objectif est de recenser toutes les personnes titulaires d'un doctorat en mathématiques. Pour chaque personne, nous prévoyons d'indiquer :

- Le nom complet de la personne titulaire du diplôme
- Le nom de l'université ayant délivré le diplôme
- L'année d'obtention du diplôme
- Le titre complet de la thèse

- Le(s) nom(s) complet(s) du/des directrice(s) ou directeur(s) de thèse

Remarque : Pour les périodes plus anciennes, la relation directeur de thèse/étudiant était probablement beaucoup moins formelle qu'aujourd'hui. Par conséquent, les liens présentés pour ces périodes peuvent refléter une relation mentor/étudiant quelque peu différente de celle des décennies plus récentes. N'oubliez pas : nous cherchons à retracer l'histoire intellectuelle de notre discipline. De plus, nous reconnaissons que le modèle utilisé peut être anachronique pour les périodes plus anciennes.

Si l'histoire du **Mathematics Genealogy Project** vous intéresse, les *Notices of the AMS* ont publié un article pertinent d'Allyn Jackson⁵ en septembre 2007.

Problèmes potentiels

Chacun des cinq éléments peut poser problème. Prenons l'exemple du nom. La plupart de nos données proviennent soit de l'université, soit des résumés de thèses. Aucune de ces sources n'est parfaite. De plus, il arrive que le nom enregistré dans les archives au moment de l'obtention du diplôme ne soit pas celui sous lequel la personne est connue aujourd'hui. Lorsque nous avons connaissance d'un tel changement, par exemple suite à un mariage, à une modification des préférences de la personne ou encore à une révision de l'orthographe, nous essayons d'en tenir compte. Parfois, par souci d'exhaustivité historique, nous choisissons d'afficher le nom complet. Voir, par exemple, E. H. Moore, dont la page affiche également Eliakim Hastings. Si une personne utilise couramment un deuxième prénom au lieu du premier et que nous en sommes informés, nous l'affichons, comme dans l'exemple suivant : C. Felix Klein, dont la page affiche Christian. Consultez le paragraphe ci-dessous pour des conseils de recherche.

Pour les personnes dont la langue maternelle utilise l'alphabet latin étendu, nous privilégions cette méthode à l'utilisation de l'anglais.

Nous souhaiterions afficher l'année d'obtention du diplôme. Cependant, certaines sources indiquent l'année de réception de la notification. Dès que nous en serons informés, nous nous efforcerons de corriger ce type d'erreur.

Nous aimerions fournir le titre complet de chaque thèse. Toutefois, certaines contraintes techniques nous en empêchent actuellement. Une fois ces problèmes résolus, les titres pourront être plus détaillés. Merci de nous signaler tout titre incomplet, contenant du code TeX non décodé ou d'autres erreurs. Puisque nous utilisons désormais MathJax pour l'affichage des titres de thèse, merci de nous signaler les titres qui devraient inclure des formules mathématiques formatées en TeX/LaTeX en nous soumettant une mise à jour.

Commentaires sur la recherche

Nous nous efforçons d'afficher les noms complets. Si toutefois le nom sous lequel vous connaissez une personne n'apparaît pas lors d'une recherche par nom et prénom, essayez de le saisir comme deuxième prénom. Nous essayons d'inclure tous les noms qu'une personne a pu utiliser au cours de sa vie dans le champ "Deuxième prénom". Si le nom d'une personne contient une lettre autre que les vingt-six lettres courantes, nous nous efforçons de l'afficher correctement. Notre moteur de recherche est assez performant : saisir "Muller" trouvera également "Müller". Les correspondances partielles sont également possibles ; ainsi, saisir "ller" vous permettra de trouver Muller, Mueller et Müller, mais aussi d'autres noms comme Keller. De même, si vous vous souvenez qu'une personne a fréquenté une université dont le nom contient "south", mais que vous ne vous rappelez plus s'il s'agissait de la Caroline du Sud, de la Floride du Sud ou de l'Afrique du Sud, saisissez simplement "south" comme nom de l'établissement. Vous obtiendrez bien sûr de nombreux résultats ; c'est pourquoi un nom, même partiel (prénom, nom ou deuxième prénom), peut aider à affiner la recherche.

5. consultable à l'adresse : <https://www.ams.org/notices/200708/tx070801002p.pdf>.