

# Description des disciplines principales en thèse en France

Sébastien Mertès

24 mai 2023

# Table des matières

<b>1. Introduction</b>	<b>1</b>
<b>2. Présentation des données</b>	<b>1</b>
<b>3. Méthode d'extraction des disciplines les plus représentées</b>	<b>1</b>
<b>4. Sélection des disciplines principales</b>	<b>2</b>
<b>5. Mise en forme des données pour une représentation graphique en air</b>	<b>3</b>
<b>6. Présentation des différentes disciplines à la lecture d'un graphique en air</b>	<b>3</b>

## **Résumé**

Nous avons étudié de 1985 à 2018, l'évolution des disciplines les plus représentatives dans les soutenances en thèse. L'ensemble des disciplines analysées, représente 40 à 60 % des soutenances en thèse de 1985 à 2000. Il en ressort qu'après une forte croissance de 1985 à 2000, la médecine connaît une chute vertigineuse pour pratiquement disparaître, tout comme la biologie, au profit de l'informatique et du droit.

## 1. Introduction

Pour observer l'évolution des proportions des différentes disciplines au fil des ans de 1985 à 2018, nous montrerons la méthode utilisée pour extraire les disciplines les plus représentées dans les soutenances sur la période étudiée. Nous concluons par la description d'un graphique représentant visuellement l'évolution de la proportionnalité de chaque discipline par année.

## 2. Présentation des données

Les données que nous allons utiliser afin de représenter les proportions des différentes disciplines en thèse, sont extraites d'un dataset au format CSV dont les variables sont listées dans le Tableau 1. Pour cette étude, nous nous concentrerons sur les variables, "Discipline" et "Date de soutenance" afin de calculer le nombre de soutenances par discipline et d'en déduire leur proportionnalité sur l'année puis sur toute la période.

	Variable	non nulle	Type
1	Auteur	448047	object
2	Identifiant auteur	317700	object
3	Titre	448040	object
4	Directeur de these	448034	object
5	Directeur de these (nom prenom)	448034	object
6	Identifiant directeur	448047	object
7	Etablissement de soutenance	448046	object
8	Identifiant etablissement	430965	object
9	Discipline	448047	object
10	Statut	448047	object
11	Date de premiere inscription en doctorat	64331	object
12	Date de soutenance	390961	object
13	Year	390961	float64
14	Langue de la these	448047	object
15	Identifiant de la these	448047	object
16	Accessible en ligne	448047	object
17	Publication dans theses.fr	448047	object
18	Mise a jour dans theses.fr	447870	object

TABLEAU 1 – Variables du dataset.

## 3. Méthode d'extraction des disciplines les plus représentées

Afin de connaître les disciplines les plus soutenues en thèse, nous avons préalablement calculé, le nombre de chaque discipline et leur proportion sur chaque année et pour l'ensemble de la période. Le résultat des disciplines les plus représentées sur toute la période est indiqué dans le Tableau 2. Par souci de lisibilité sur le graphique à venir, nous fusionnerons les disciplines voisines comme le droit et les renommerons ainsi que d'autres dont le nom est trop long.

Discipline	Soutenances	Proportion(%)
MEDECINE	25520	6.73
PHYSIQUE	13446	3.55
INFORMATIQUE	12059	3.18
SCIENCES ECONOMIQUES	8819	2.32
SCIENCES BIOLOGIQUES ET FONDAMENTALES APPLIQUEES. PSYCHOLOGIE	8750	2.31
CHIMIE	8710	2.30
SCIENCES APPLIQUEES	8365	2.21
HISTOIRE	6801	1.79
PHARMACIE	6792	1.79
SCIENCES DE GESTION	6184	1.63
DROIT PRIVE	5127	1.35
DROIT PUBLIC	5118	1.35
SOCIOLOGIE	5097	1.34
PSYCHOLOGIE	4863	1.28
MATHEMATIQUES	4510	1.19
GEOGRAPHIE	4188	1.10
DROIT	4145	1.09
PHILOSOPHIE	4028	1.06
TERRE, OCEAN, ESPACE	3284	0.87
SCIENCES DE L'EDUCATION	3260	0.86

**TABLEAU 2 – Nombre et proportion des soutenances par discipline sur la période 1985-2018.**

## 4. Sélection des disciplines principales

Afin de sélectionner un nombre limité de disciplines mais choisies selon leur représentativité, nous prenons par convention les disciplines dont la proportion est supérieure à 1 % (voir Tableau 1) du total du nombre de soutenances sur la période étudiée. Le Tableau 3 montre les disciplines les plus représentatives.

Disciplines représentatives
PHILOSOPHIE
DROIT
GEOGRAPHIE
MATHEMATIQUES
PSYCHOLOGIE
SOCIOLOGIE
DROIT PUBLIC
DROIT PRIVE
SCIENCES DE GESTION
PHARMACIE
HISTOIRE
SCIENCES APPLIQUEES
CHIMIE
SCIENCES BIOLOGIQUES ET FONDAMENTALES APPLIQUEES. PSYCHOLOGIE
SCIENCES ECONOMIQUES
INFORMATIQUE
PHYSIQUE
MEDECINE

**TABLEAU 3 – Disciplines les plus soutenues en thèse sur la période 1985-2018.**

## 5. Mise en forme des données pour une représentation graphique en air

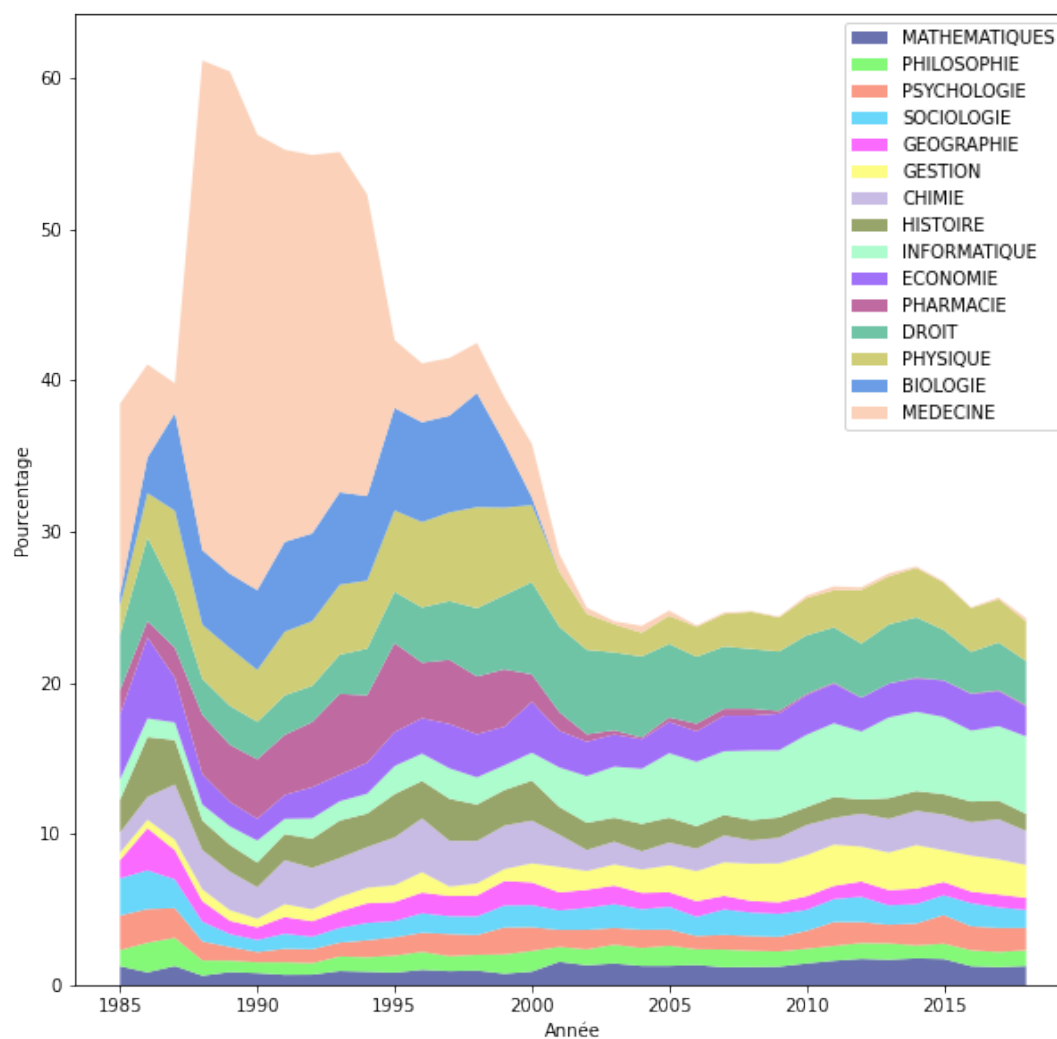
Le choix de la production d'un graphique en air nécessite une mise en forme des données au préalable afin de représenter les valeurs maximales de la proportion de soutenances par discipline dans l'ordre croissant afin de minimiser la répercussions des valeurs les plus extrêmes sur les autres disciplines. Le Tableau 4, montre les disciplines dans l'ordre croissant de leur proportionnalité sur l'ensemble de la période.

Discipline	Valeur maximale(%)
MATHEMATIQUES	1.78
PHILOSOPHIE	1.96
PSYCHOLOGIE	2.29
SOCIOLOGIE	2.56
GEOGRAPHIE	2.79
GESTION	2.88
CHIMIE	3.66
HISTOIRE	3.95
INFORMATIQUE	5.34
ECONOMIE	5.35
PHARMACIE	5.89
DROIT	6.10
PHYSIQUE	6.69
BIOLOGIE	7.54
MEDECINE	33.21

**TABLEAU 4 – Proportion maximale des principales disciplines sur la période 1985-2018.**

## 6. Présentation des différentes disciplines à la lecture d'un graphique en air

Sur la Figure 1 on visualise les proportions en pourcentage des principales disciplines étudiées en thèse en France. Ces disciplines ont été étudiées sur la période de 1985 à 2018 ou chacune d'entre elles est représentée par une couleur dont la surface indique une mesure quantitative des soutenances dans cette discipline. Ainsi, de 1985, jusqu'au début des années 2000, les disciplines indiquées représentaient plus de 40 à 60 % des soutenances avec une forte proportion pour la médecine qui représentait, environ, à elle seule, plus de 35 % des soutenances à la fin des années 80 et début 90. Cette dernière décline fortement après 1995 pour finalement devenir négligeable tout comme en biologie. Après les années 2000, le droit et l'informatique semble être les disciplines les plus représentées en thèse après les années 2000 avec plus ou moins 5 % des soutenances chacune.



**FIGURE 1 – Proportion des disciplines des soutenances en thèse en France de 1985 à 2018**