



# Expérimentation pédagogique sur le thème

## Vers le Big Data

### « Répartition des richesses »

Utiliser des données pour créer une activité en maths complémentaires

Lycée

Testée sur deux années en maths complémentaires.

## 1. Présentation de l'activité

L'objectif de cette activité est l'utilisation de données afin d'introduire le thème « répartition des richesses et inégalités » en maths complémentaires. Il s'agit donc de chercher des données sur un site, de comprendre ces données et de les « digérer » pour créer des problèmes à proposer aux élèves.

## 2. La recherche des données.

Dans les programmes, le thème « répartition des richesses » est introduit par la phrase suivante : « l'étude de la répartition de richesses dans la population d'un pays, des salaires dans une entreprise, etc., et la comparaison des différentes répartitions sont des occasions de réinvestir des connaissances antérieures de statistique descriptive et de construire de nouveaux outils d'analyse faisant intervenir les fonctions d'une variable (notamment des fonctions de répartition) et le calcul intégral ». L'une des difficultés est donc de trouver des données exploitables. Après discussion avec quelques collègues de SES, certains sites ont été indiqués. La difficulté reste de trouver des données brutes compréhensibles pour un professeur de mathématiques. L'un des sites possibles est le site de l'Organisation de coopération et de développement économiques (OCDE ou OECD en anglais). Il permet d'avoir accès à des statistiques sur plusieurs pays et [aux data base](#). Celle trouvée se présente comme ceci :

Key indicators on the distribution of household disposable income and poverty, 2007, 2019 and 2020 or most recent year																				Indicateurs clés de la distribution des res								
Values refer to income earned in the year																				6 juin 2023								
Country	Gini coefficient		S80/S20 income share ratio				Income share in total income					Poverty rate (relative threshold)					Poverty rate (threshold "anchor" in 2005)		Coefficient de Gini			Ratio interdéclies S80/S20						
	2007	2019	2007	2019	2020 or latest available year	Bottom 10%	Bottom 20%	Bottom 40%	Top 40%	Top 20%	Top 10%	Total	By age group, latest available year				2007	2019	2020 or latest available year	2007	2019	2020 or latest available year						
						2019 or latest available year (%)					2007	2019	2020 or latest available year	Children (< 18)	Youth (18-25)	Adult (26-65)	Elderly (> 65)	Working poor										
1 Australia	0.338	0.325	0.318	5.8	5.6	2.6	7.2	19.7	63.2	40.0	24.8	14.9	12.4	12.6	13.3	8.2	10.4	22.6	5.1	7.2	6.6	Australie	0.338	0.325	0.318	5.8	5.6	5.6
7 Austria	0.284	0.274	0.272	4.4	4.2	3.2	8.6	22.8	58.9	35.9	21.3	9.7	10.0	9.6	12.0	9.4	8.6	10.6	7.0	9.7	7.5	Autriche	0.284	0.274	0.272	4.4	4.2	4.2
8 Belgium	0.277	0.262	0.248	4.1	3.8	3.6	4.0	9.8	24.0	57.3	34.1	20.1	8.3	8.1	7.3	5.0	5.9	6.9	8.6	3.6	8.3	Belgique	0.277	0.262	0.248	4.1	3.8	3.8
10 Canada	0.317	0.300	0.290	5.3	4.8	4.2	3.5	8.8	22.5	59.7	36.7	22.1	10.9	11.5	8.6	7.3	8.5	9.1	12.1	4.7	11.3	Canada	0.317	0.300	0.290	5.3	4.8	4.2
11 Chile	0.480	0.454	0.460	11.8	10.0	10.3	1.9	5.1	14.4	72.1	52.2	38.7	17.8	16.1	16.5	21.5	16.5	14.0	17.8	13.8	14.0	Chili	0.480	0.454	0.460	11.8	10.0	10.3
12 Colombia	0.478	0.487	0.487	12.7	13.3	1.3	4.0	12.2	74.9	53.8	36.4	19.9	20.3	27.4	17.4	17.4	22.4	15.7	3.3	1.3	1.6	Colombie	0.478	0.487	0.487	12.7	13.3	13.3
13 Costa Rica	0.259	0.248	0.255	3.6	3.5	3.6	4.1	9.9	24.3	57.9	35.5	21.8	5.5	5.6	5.3	7.8	5.3	4.5	5.1	3.1	3.3	Costa Rica	0.259	0.248	0.255	3.6	3.5	3.6
15 Denmark	0.244	0.258	0.258	3.4	3.8	3.6	3.8	9.4	23.4	56.5	36.0	22.3	5.9	6.5	6.9	4.8	20.3	3.2	4.3	4.7	4.7	Danemark	0.244	0.258	0.258	3.4	3.8	3.8
16 Estonia	0.313	0.305	0.305	5.2	5.1	5.2	2.8	7.3	20.2	61.7	37.7	24.1	13.9	14.9	15.8	10.8	13.5	11.7	34.6	9.2	4.4	Estonie	0.313	0.305	0.305	5.2	5.1	5.2
17 Finland	0.269	0.273	0.265	3.9	3.9	3.7	4.1	9.7	23.7	58.5	36.2	22.3	7.8	6.4	5.7	2.4	16.8	4.5	6.9	3.3	6.4	Finlande	0.269	0.273	0.265	3.9	3.9	3.7
18 France	0.295	0.292	0.292	4.4	4.4	4.4	3.4	8.7	22.3	69.2	38.1	24.0	7.6	8.4	8.4	11.7	13.2	7.7	4.4	7.1	6.8	France	0.295	0.292	0.292	4.4	4.4	4.4
19 Germany	0.285	0.296	0.296	4.3	4.6	4.6	3.2	9.2	21.8	69.4	36.0	23.5	9.6	10.9	10.9	11.7	18.6	9.5	10.9	4.6	8.8	Allemagne	0.285	0.296	0.296	4.3	4.6	4.6
20 Greece	0.329	0.312	0.320	5.8	5.1	5.6	2.3	7.0	20.0	62.3	38.1	24.2	13.2	11.5	13.0	16.3	15.6	13.2	9.3	10.2	11.3	Grèce	0.329	0.312	0.320	5.8	5.1	5.6
21 Hungary	0.257	0.286	0.286	3.7	4.4	4.3	3.2	8.6	22.5	59.5	36.6	22.3	6.5	9.2	8.7	10.2	10.9	8.7	6.1	7.7	7.2	Hongrie	0.257	0.286	0.286	3.7	4.4	4.4
22 Iceland	0.285	0.284	0.250	4.1	3.8	3.5	4.1	10.0	24.6	57.3	35.0	21.3	8.6	6.0	4.9	5.4	5.0	5.1	3.1	4.4	3.7	Islande	0.285	0.284	0.250	4.1	3.8	3.5
23 Ireland	0.304	0.293	0.282	4.6	4.3	4.1	3.7	9.0	22.6	59.9	37.1	22.9	9.6	8.1	7.7	7.4	3.6	8.6	14.7	2.8	7.3	Irlande	0.304	0.293	0.282	4.6	4.3	4.1
24 Israel	0.365	0.342	0.340	7.5	6.4	6.1	2.4	6.5	18.6	64.0	39.7	23.7	11.3	17.3	16.9	20.1	19.3	13.9	17.0	6.4	8.8	Israël	0.365	0.342	0.340	7.5	6.4	6.1
26 Italy	0.313	0.304	0.311	6.2	6.8	6.0	2.1	8.8	16.6	63.1	40.8	24.0	11.0	14.6	14.6	17.3	14.7	14.6	10.9	10.6	10.6	Italie	0.313	0.304	0.311	6.2	6.8	6.0

Et en zoomant :

Indicateurs clés de la distribution des revenus disponibles et de la pauvreté monétaire de													
Les valeurs correspondent à l'année des r													
1	2	3	Gini		Ratio interdéciles S80/S20			Part du revenu total, par groupe de revenu					
4	5	6	7	8	9	10	11	12	13	14	15	16	17
18 Dec. 2020	Updated countries and values are highlighted in light blue	2018 ou + récent	2007	2017	2018 ou + récent	10% les + pauvres	20% les + pauvres	40% les + pauvres	40% les + riches	20% les + riches	10% les + riches	2018 ou + récent (%)	
2007													
7	Australia	0,325	5,8	e 5,5	5,6	2,7	7,2	19,6	63,4	40,4	25,5	14,9	14,9
8	Austria	0,280	4,4	e 4,2	4,3	3,1	8,5	22,7	59,4	36,6	22,3	9,7	9,7
9	Belgium	0,258	4,1	e 4,0	3,8	3,8	9,2	23,3	58,1	34,8	20,7	9,3	9,3
10	Canada	0,303	5,3	e 5,1	4,9	2,9	7,7	21,0	61,3	38,1	23,2	12,9	12,9
11	Chile	0,460	11,8	e 10,0	10,3	1,9	5,1	14,4	72,1	52,2	36,7	17,8	17,8
12	Colombia	..	..	e ..	..	..	..	..	..	..	..	..	..
13	Czech Republic	0,249	3,6	e 3,5	3,5	4,1	9,9	24,3	57,5	34,7	20,5	5,5	5,5
14	Denmark	0,264	3,4	e 3,7	3,8	3,8	9,5	23,6	58,3	35,7	21,9	5,9	5,9
15	Estonia	0,305	5,2	e 5,2	5,2	2,8	7,2	20,1	61,8	37,6	21,9	13,9	13,9
16	Finland	0,269	3,9	e 3,8	3,9	3,9	9,4	23,4	58,8	36,3	22,3	7,8	7,8
17	France	0,301	4,4	e 4,4	4,6	3,4	8,6	21,9	60,8	39,0	25,0	7,6	7,6
18	Germany	0,289	4,3	e 4,6	4,5	3,3	8,4	22,2	60,1	37,5	23,2	9,0	9,0
19	Greece	0,306	5,6	e 5,4	5,0	2,9	7,7	20,9	61,5	38,2	23,4	13,2	13,2

Elle est relativement compliquée à lire pour quelqu'un d'aussi peu expérimenté que moi. Après réflexion, je décide d'interpréter certaines données et de les proposer sous un autre format.

### 3. Exploitation des données brutes

1<sup>er</sup> temps : l'exemple de la France.

Les données qui m'intéressent sont celles qui proposent certains déciles pour l'année 2018 : 10%, 20%, 40% les plus pauvres puis 40%, 20%, 10% les plus riches. Dans un premier temps, j'ajuste ces données pour les faire correspondre à une vision plus classique en France. La source est indiquée mais les valeurs sont arrondies à 0.01 près :

Un 1<sup>er</sup> tableau : répartition de la richesse en France

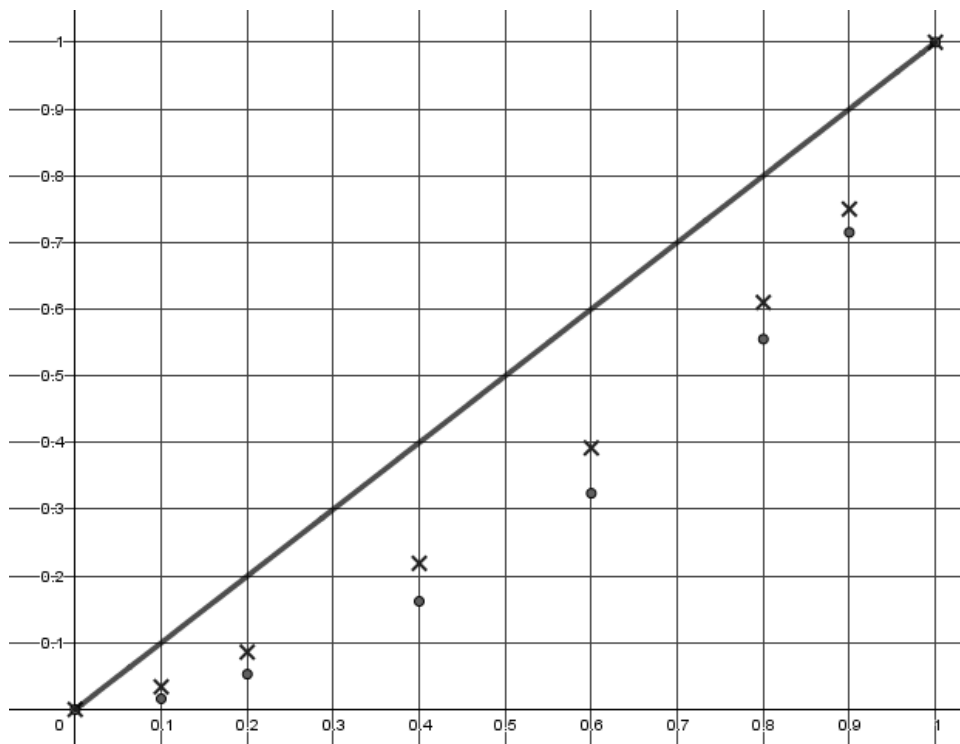
Part de la population Française en %	Distribution des revenus disponibles en France (2018) (source : base de donnée de l'OCDE)
0 %	0 %
10 %	3 %
20 %	9 %
40 %	22 %
60 %	39 %
80 %	61 %
90 %	75 %
100 %	100 %

Ce premier travail facilite la compréhension, notamment pour les élèves n'ayant pas la spécialité SES.

Plusieurs questions sont posées, un graphique est demandé et la 1<sup>ère</sup> bissectrice est tracée. Cette première activité se termine sur la question : comment interpréter cette droite d'équation  $y=x$  dans le contexte de l'exercice ? (Réponse : elle indique une répartition des richesses strictement égalitaire.)

2<sup>ème</sup> temps : comparaison pour différents pays.

Il s'agit de comparer la répartition des richesses aux Etats-Unis et en France. Un nouveau graphique est proposé. C'est une satisfaction de constater que ce graphique est proche de celui indiqué par les élèves. On retrouve les données de la France (points en forme de croix). Des points correspondant aux données pour les États-Unis sont rajoutés.



Un tableau complète ce graphique. Il est fourni afin de parvenir plus vite au cœur du nouveau problème.

Déciles	Part de la population Française en %	Distribution des revenus disponibles en France ( 2018).	Distribution connue en 2018 des revenus disponibles au Etats-Unis.
	0	0 %	
D <sub>1</sub>	10 %	3 %	
D <sub>2</sub>	20 %	9 %	
D <sub>4</sub>	40 %	22 %	
D <sub>6</sub>	60 %	39 %	
D <sub>8</sub>	80 %	61 %	
D <sub>9</sub>	90 %	75 %	
	100 %	100 %	

L'objectif est de faire intervenir de nouveaux outils mathématiques pour le traitement. Il est alors demandé de proposer une courbe passant par tous les points afin de déterminer d'autres valeurs. Certains élèves se rappellent que de telles courbes sont des courbes de Lorenz (ceux en spécialité SES).

La courbe tracée à partir des points correspondants à la France, nous réfléchissons collectivement aux propriétés de la fonction associée à cette courbe. Ce travail permet de retrouver la notion de continuité et la notion de convexité.

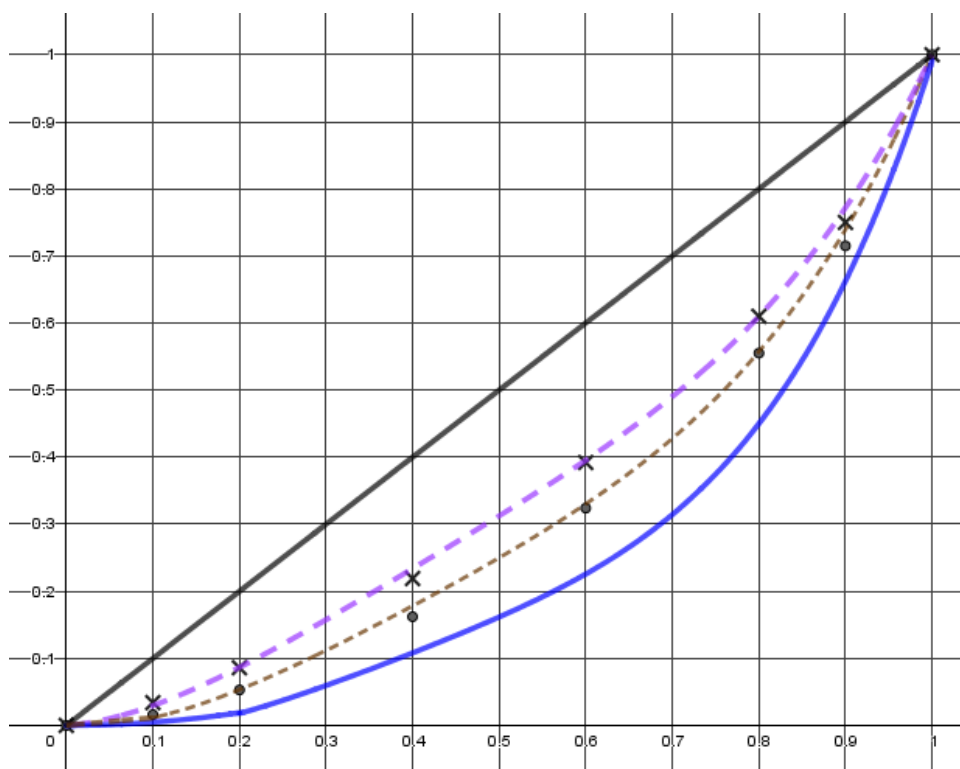
Le logiciel geogebra est alors fortement exploitée par le professeur pour déterminer une fonction polynomiale possible (ce qui est plus pratique pour le calcul d'aire sous la courbe – voir partie 3). Ce travail permet de reparler du thème « modèle défini par une fonction d'une variable ». Il s'agit de savoir si la fonction proposée répond bien aux propriétés trouvées à savoir :

- l'image de 0 est 0,
- l'image de 1 est 1,
- la fonction est continue sur  $[0 ; 1]$ ,
- la fonction est convexe sur  $[0 ; 1]$ .

Il est à signaler que cette dernière notion n'avait pas été introduite la deuxième année. C'est donc cette activité qui a véritablement permis de travailler ce concept en maths complémentaires puisqu'un exemple concret était obtenu : une courbe de Lorenz ne peut pas être la courbe représentative d'une fonction non convexe sur  $[0 ; 1]$ . Outre la propriété des fonctions convexes via la position des tangentes par rapport à la courbe de la fonction, ce travail a été prolongé par la mise en place de la propriété liant le signe de la dérivée seconde et la convexité de la fonction, I. ( $f: x \mapsto 2.2x^4 - 3.6x^3 + 2.3x^2 + 0.1x$ )

3<sup>ème</sup> temps : le coefficient de Gini.

Les données obtenues sur le site de l'OCDE donnent un nouvel indicateur utilisé en économie à savoir le coefficient de Gini qui est un indicateur synthétique permettant de rendre compte du niveau d'inégalité pour une variable et sur une population donnée. La dernière partie explique donc comment calculer ce coefficient et donc de retrouver l'indicateur (ou une valeur proche suivant la fonction retenue lors de l'approximation).



Pour cette partie, un nouveau pays est introduit ainsi que de nouvelles propositions de fonctions répondant aux 4 critères trouvées dans la partie 2. Certaines de ces fonctions sont définies par morceaux et donc non dérivables en tout point de  $[0 ; 1]$  (la propriété est alors admise à partir du graphique – aucun élève n’est revenu à la définition d’une fonction convexe). Ce travail permet de revenir sur le calcul intégral (et donc le thème « calcul d’aires ») et, pour certains, d’avoir un exemple concret pour utiliser certaines propriétés du calcul intégral telles que la relation de Chasles.

## 4. Conclusion

L’exemple proposé est un exemple d’exploitation de données brutes en mathématiques. Pour le professeur, cela demande du temps de recherche, du temps de compréhension et du temps pour construire un problème global cohérent. Néanmoins, cela a permis d’introduire parfois certaines notions (convexité), ou d’améliorer la compréhension d’autres (continuité, calcul intégral). Pour les élèves, ceux ayant choisi la spécialité SES, cela a permis de créer une bonne passerelle entre les deux matières et donc de mieux comprendre certains indicateurs statistiques utilisés en économie (calculs des indices de Gini de la France, des États-Unis et éventuellement de la Chine). La grande majorité des élèves a trouvé la situation motivante et a pu aller à son rythme (différenciation possible en ajoutant des pays). La question reste posée quant à l’éventualité de proposer les données brutes et de demander aux élèves de les exploiter (activité 1).