Géomarketing - Statistiques et Cartographie

Serge Lhomme

Maître de conférences en Géographie
http://sergelhomme.fr/
serge.lhomme@u-pec.fr

24/09/2025 Statistiques et cartographie (1/3)

01/10/2025 Statistiques et cartographie (2/3)

08/10/2025 Statistiques et cartographie (3/3)

15/10/2025 Modèles géographiques pour le géomarketing (1/3)

22/10/2025 DS Statistiques et cartographie

05/11/2025 Modèles géographiques pour le géomarketing (2/3)

12/11/2025 Modèles géographiques pour le géomarketing (3/3)

19/11/2025 Les concepts fondamentaux du géomarketing (1/3)

26/11/2025 DS Modèles géographiques pour le géomarketing

03/12/2025 Les concepts fondamentaux du géomarketing (2/3)

10/12/2025 Les concepts fondamentaux du géomarketing (3/3)

17/12/2025 DS Les concepts fondamentaux du géomarketing

Introduction

Le marketing se fixe pour objectif principal la satisfaction du client. Le marketing ne doit pas être confondu avec l'approche commerciale. Le marketing n'est pas synonyme de publicité. Le marketing vise néanmoins de différentes manières à influencer le comportement du consommateur.

Le géomarketing se focalise sur les questions spatiales relatives au marketing comme par exemple : Où est-il pertinent d'implanter un magasin ? Où résident et travaillent les clients ? C'est pourquoi la cartographie tient une place centrale dans le domaine du géomarketing.

Attention

Le géomarketing ne s'applique pas uniquement aux entreprises privées. Les questions centrales du géomarketing se transposent effectivement aux entreprises publiques, collectivités, établissements de santé... Où implanter un hôpital ? Où résident et travaillent les usagers ?

Les fondements de la cartographie

La subjectivité

On appelle carte toute représentation graphique partielle ou complète dans le plan d'un objet plus complexe. Dans les domaines de la géographie et de l'aménagement, cet objet est un territoire représenté selon une "vue de dessus".

Cette représentation graphique est établie par un auteur, à un moment donné, sur un espace donné. Ce caractère interprétatif (voire subjectif) de la conception cartographique est largement accepté par les géographes.

Naturellement, aucun cartographe digne de ce nom ne cherche à tromper ses lecteurs, mais force est de constater que le travail du cartographe résulte de multiples choix d'application, de conventions, de plaisirs esthétiques qui ne sont pas toujours explicités, ni justifiés.

Surtout, une carte est toujours "synthétique", elle ne peut à elle seule refléter toute la complexité du problème traité, elle porte un message.

Les fondements de la cartographie

L'art et le langage cartographique

La cartographie peut être définie comme l'« *ensemble des études et des opérations scientifiques, artistiques et techniques intervenant à partir des résultats d'observations directes ou de l'exploitation d'une documentation, en vue de l'élaboration de cartes et autres modes d'expression, ainsi que dans leur utilisation* ».

A l'instar des mathématiques, la cartographie tend à être un langage universel. Ainsi quelle que soit la langue utilisée, une carte est théoriquement compréhensible par tout le monde, même si la légende est parfois nécessaire pour saisir des points de détails.

Pour cela, ce langage doit respecter les règles de lisibilité, de clarté, d'intelligibilité et d'enchaînement logique, inhérentes à tout langage humain.

Les fondements de la cartographie

L'objectif du cartographe

Outil de communication par l'image, la carte doit être perçue avec un minimum de biais, dans la mesure où le concepteur a su prendre en compte les lois de la perception visuelle, du pouvoir intégrateur et séparateur de l'œil, des contrastes de couleurs, et des règles typographiques concernant les écritures.

Entre la subjectivité de l'objet qu'il réalise et la possibilité d'en faire un objet d'art, le cartographe ne doit jamais oublier que sa première mission est de se faire comprendre pour apporter l'information souhaitée.

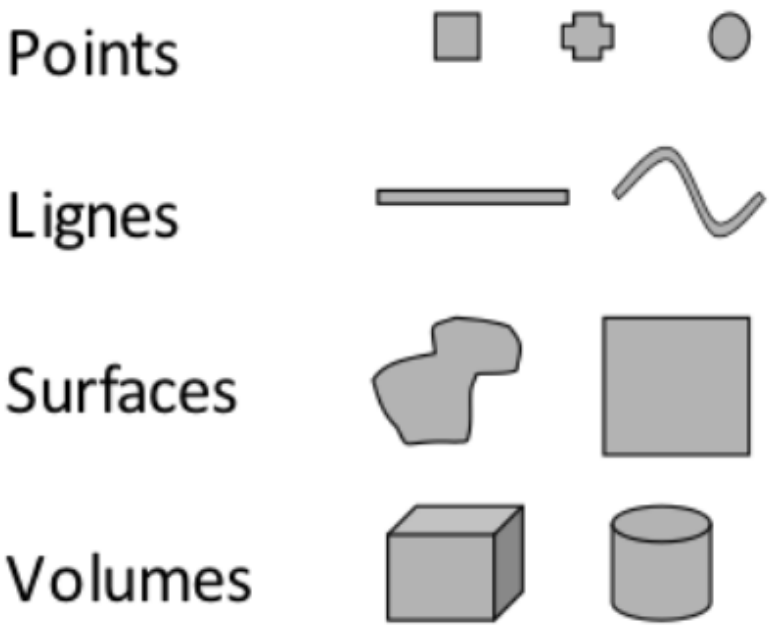
Une carte doit s'adapter pour cela au public visé.

Comme il est toujours difficile de se faire comprendre, il ne faut jamais oublier de légender sa carte.

Les fondements de la cartographie

Primitives, structures visuelles

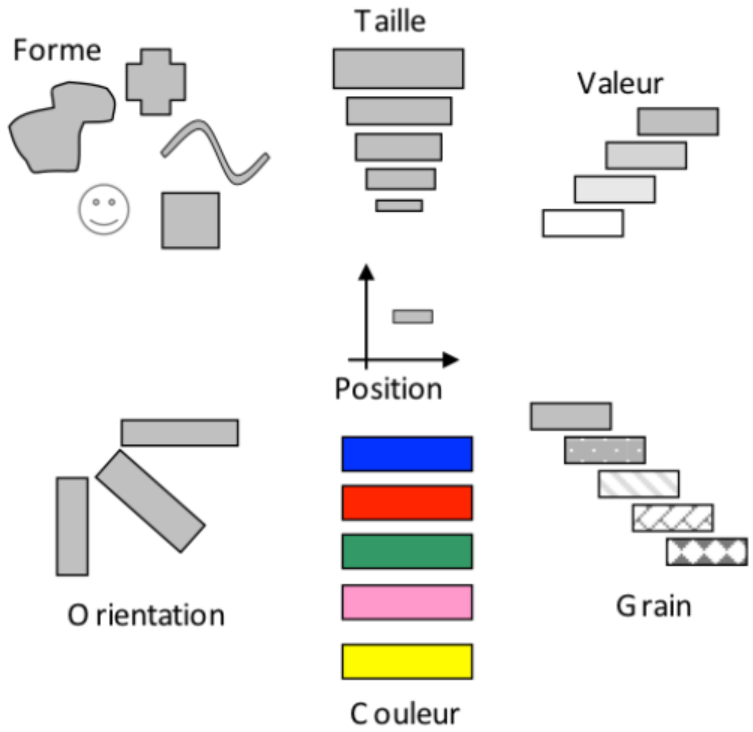
Un objet géographique matériel prend la forme d'une structure visuelle (ou primitive) en fonction de l'échelle de représentation : à l'échelle du 1 : 1 000 000 une ville est représentée par un point, alors qu'à l'échelle du 1 : 100 000, elle occupe un périmètre plus ou moins nettement délimité.



Les fondements de la cartographie

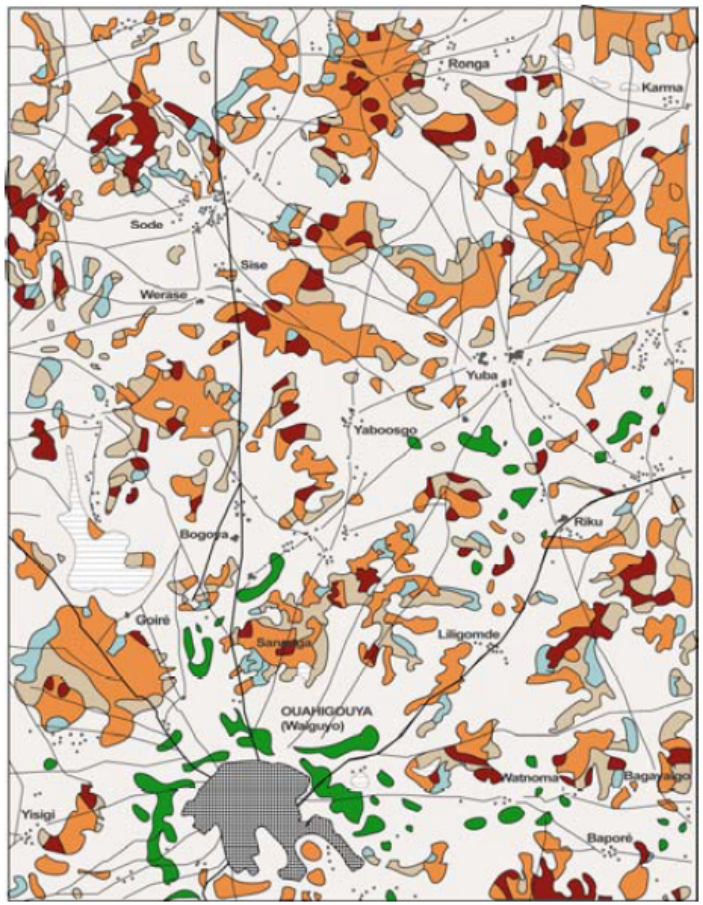
Coder les informations

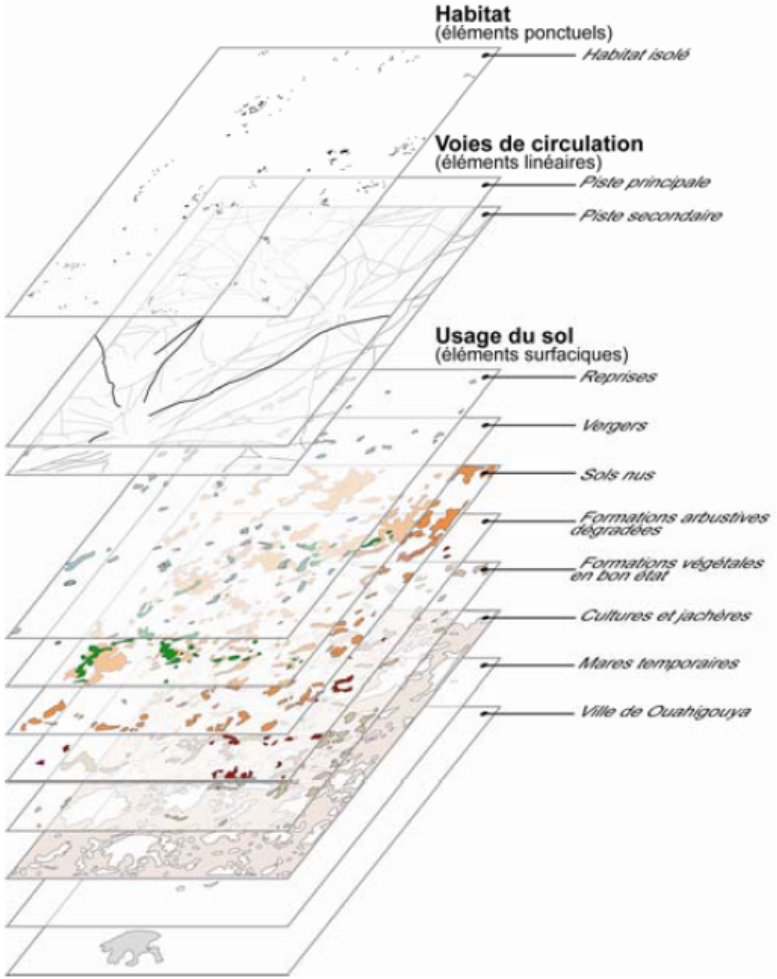
Pour coder des informations, il est possible de faire varier certaines propriétés graphiques de ces structures visuelles (par exemple la forme). Les variations possibles sur les structures visuelles sont regroupées par type et sont nommées « variables visuelles ».



Les fondements de la cartographie

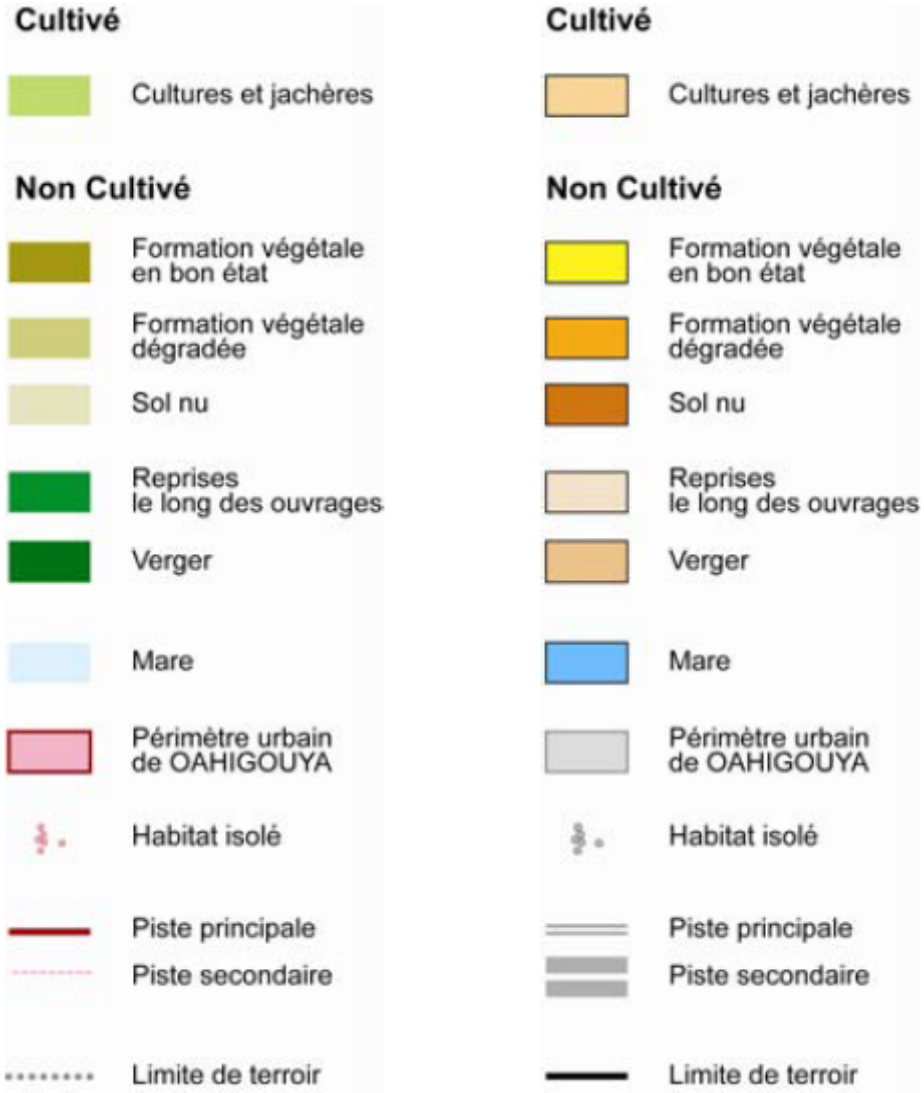
La carte est un ensemble de calques





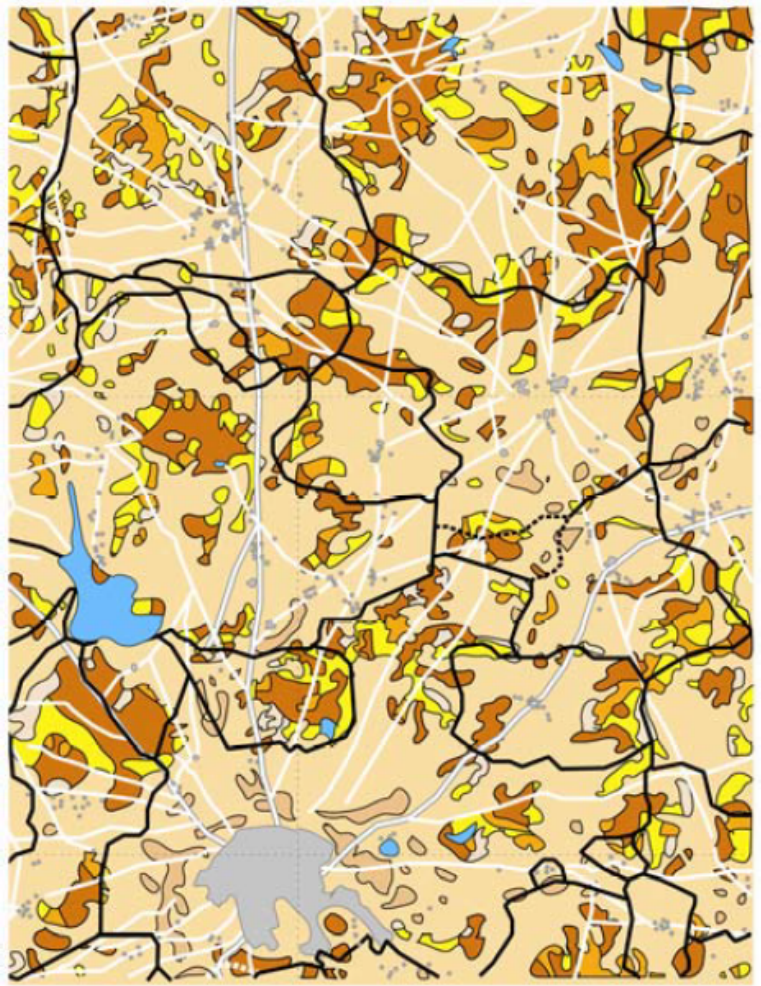
Les fondements de la cartographie

La carte parfaite n'existe pas



Les fondements de la cartographie

La carte parfaite n'existe pas, mais les cartes fausses oui



Les fondements de la cartographie

Variables et valeurs

Les attributs décrivant un objet géographique peuvent être quantitatifs ou qualitatifs. On parle aussi de variables, de caractères, de catégories...

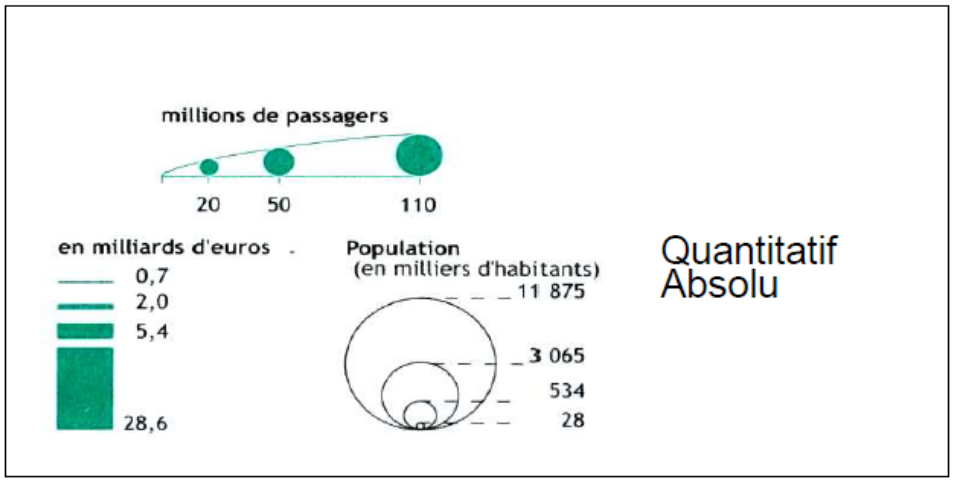
Une variable qualitative est composée de valeurs qui ne sont pas des quantités (ou des rapports de quantités). Ces valeurs peuvent être simplement nominales (lettres, mots, chiffres) ou ordinales c'est-à-dire qu'on peut les ordonner par ordre croissant (les classements : 1er, 2ème, 3ème, ... ; les jugements : bons, moyens, mauvais...). Ces variables peuvent être numériques (numéro de département).

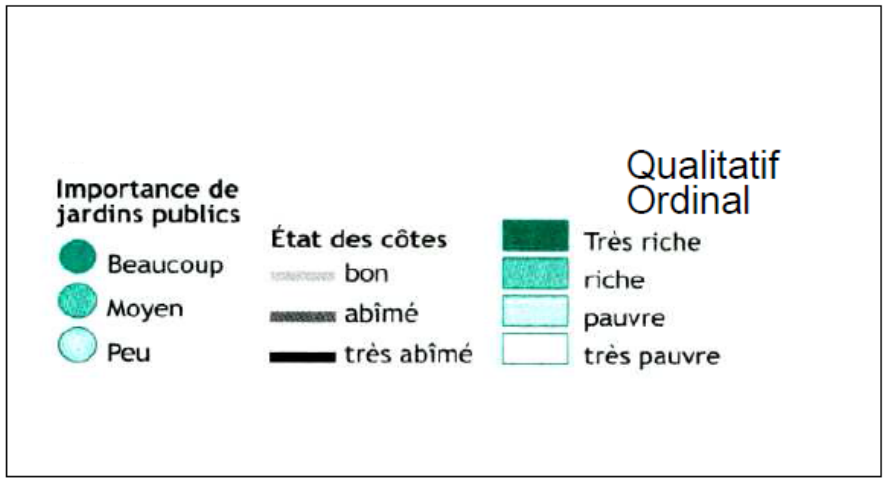
On va dans ce cours plutôt se focaliser sur des variables quantitatives.

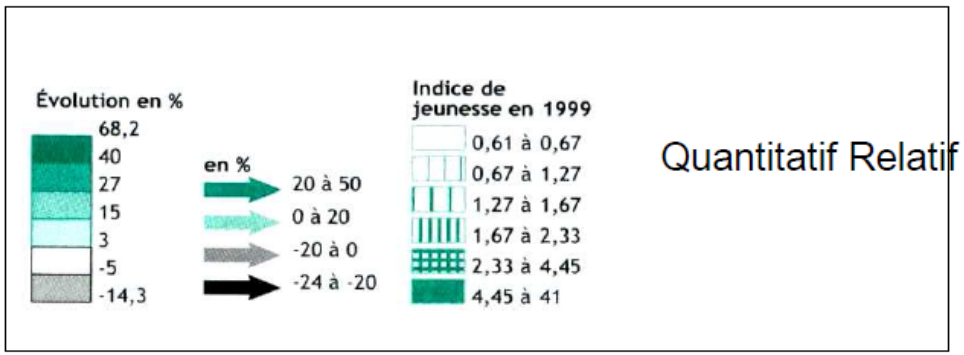
En cartographie, on différencie notamment les variables quantitatives de stock et celles de taux (ou de rapport). On peut aussi distinguer des variables de flux (tableau d'échange).

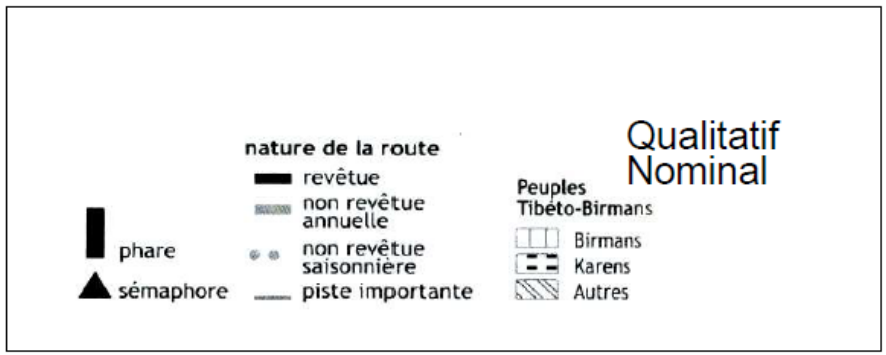
Les fondements de la cartographie

Résumé









Les fondements de la cartographie

Ce qu'il est nécessaire de connaitre avant de commencer une carte :

- Bien appréhender l'objectif de la carte ;
- Identifier la cible ;
- Identifier l'information à cartographier (collecte et traitement) ;
- Adapter le fond de carte (projection, généralisation) ;
- Choisir le langage cartographique.

Ce qu'il ne faut pas oublier :

- D'indiquer le nord ;
- De préciser l'échelle ;
- De mettre la légende ;
- De mettre un titre ;
- De préciser les sources ;
- De mettre son nom et la date de réalisation.

Discrétisation

Simplifier l'information pour faciliter le message à faire passer

La discrétisation consiste à regrouper dans des classes des objets relativement similaires. Elle s'applique à une seule variable. Chaque classe regroupe ainsi plusieurs objets possédant des valeurs proches.

C'est une opération de simplification nécessaire en cartographie quand celle-ci vise à aider à prendre des décisions.

Pour réaliser une discrétisation, il faut déterminer le nombre de classes et les bornes des classes.

Intuitivement, un bon découpage correspond à des classes homogènes et séparées, ce qui correspond respectivement aux notions statistiques de faible variance intraclasse et de forte variance interclasse.

Et oui, pour réaliser une simple carte représentant une seule variable quantitative de taux ou de rapport, il faut des connaissances statistiques solides !

Discrétisation

Tableau d'information géographique

Variables Caractères

Entités
Individus
Objets
Unités

	AGRI	ARTI	CADRE	PRO INT	EMPLOYE	OUVRIER	RETRAITE	AUTRES
0	4067	16745	36426	70663	77349	78998	30417	29910
1	5201	11414	18629	48889	71578	78906	32063	42108
2	6159	9396	11842	31102	46196	42101	24218	21344
3	2027	6387	6951	16140	20885	15560	10687	10278
4	2040	5286	5883	15687	19707	12415	8865	7222
5	1918	39514	73884	117652	160574	87227	54801	68990
6	4132	10758	12461	31859	39088	37199	21660	19063
7	3362	6520	9771	25487	36326	42618	16464	23226
8	2348	4888	5537	14442	20788	16635	10379	9741
9	4786	7198	12499	28889	40121	43224	18616	18373
10	5474	12045	13486	31504	47297	34501	25584	25482
11	11100	10116	9937	25717	33970	29390	18550	13039
12	4508	52145	135584	225526	268192	168920	100790	155952
13	6081	18540	36155	72740	94384	82643	41776	34559
14	7431	4976	4547	12517	19508	17183	10386	7786
15	5756	10236	14308	32982	46169	46763	24653	20416
16	9061	21025	24906	57199	84800	65577	46140	35151
17	4025	8183	12967	29897	42604	39209	22057	18361
18	4930	7553	9273	23538	33050	28048	16824	12970
19	5352	13503	32185	62097	71854	61993	30098	23303
20	12517	17359	24766	54660	70213	68805	43023	29034
21	5259	3696	3669	9387	16039	12764	9811	6780
22	7694	15460	14125	34308	54103	47943	31651	24558
23	4123	12196	29138	57994	65018	77588	28449	27692
24	6106	14998	23027	52929	59618	57491	28522	30398
25	4064	14930	28312	62716	74944	85833	34937	33452
26	3848	10110	22675	47678	58632	57416	25285	22370
27	11054	23043	46057	94797	116573	101817	59116	46699

Valeur
Modalité

Discrétisation

Les résumés statistiques classiques

La moyenne :

$$\bar{X} = \frac{1}{N} \times \sum_{i=1}^{N} X_i$$

L'écart-type :

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2}$$

Les caractéristiques de position fondées sur l'ordonnancement (les modes) :

Minimum ; Maximum ; Médiane ; Quantiles ; Déciles...

Discrétisation

Les règles

Comme il existe de nombreuses méthodes de discrétisation, le cartographe va devoir choisir la meilleure représentation et cela dépendra de l'objectif poursuivi.

L'objectif est bien souvent de conserver la forme de la distribution et de conserver au mieux la structure interne : Jenks ou Seuils naturels.

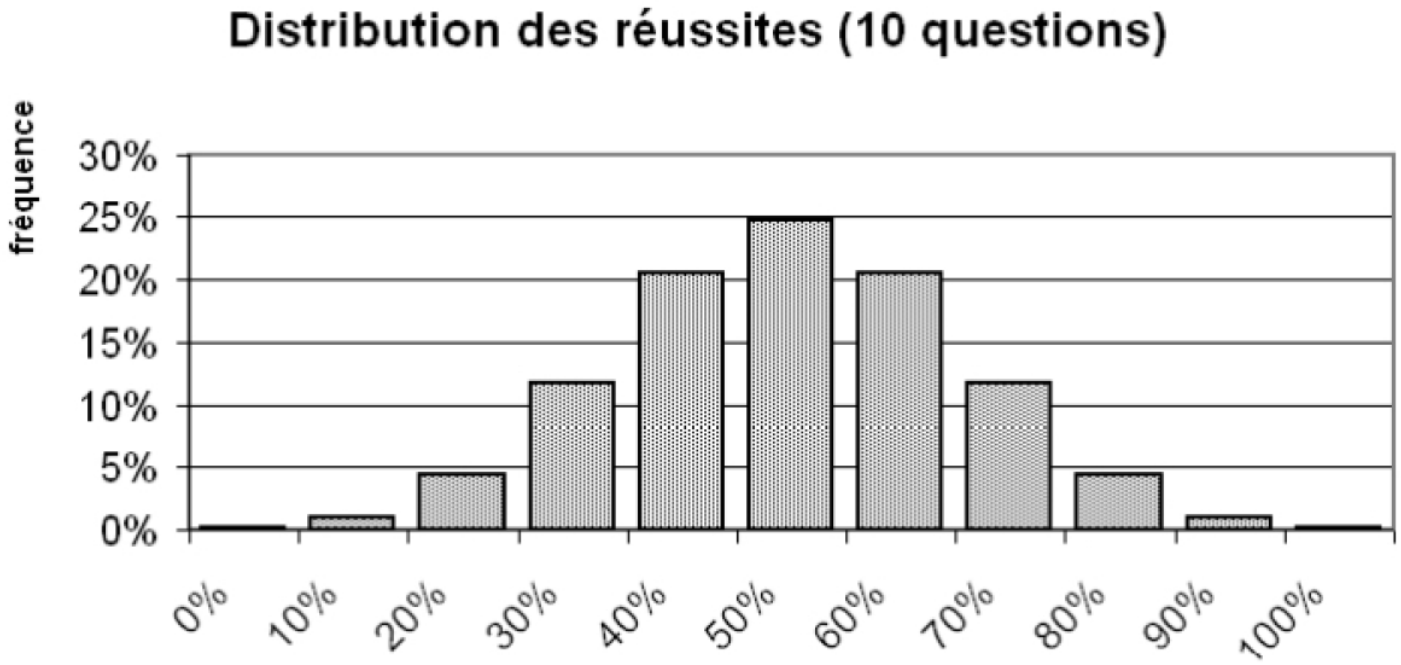
Pour illustrer la dispersion des variables étudiées, il faut tout simplement choisir une classification par amplitude égale (qui permet en fait d'obtenir une carte de la distribution statistique).

Pour comparer la position de certains lieux en fonction de différentes variables (sur plusieurs cartes), il faudra utiliser des méthodes faisant référence à des paramètres statistiques (moyenne, médiane, EcT...) : effectifs égaux (EF) ; Q6 ; ST.

Discrétisation

L'arme absolue : la distribution statistique

La distribution statistique (distribution des fréquences) est un tableau qui associe des classes de valeurs obtenues lors d'une expérience à leurs fréquences d'apparition. Ce tableau de valeurs est modélisé en théorie des probabilités par une loi de probabilité. Il peut notamment se représenter sous la forme d'un histogramme.



Discrétisation

Le nombre de classes

Choisir le nombre de classes d'une discrétisation est très problématique.

Le nombre de classes est en théorie proportionnel au nombre d'individus observés. Ainsi, il peut paraitre excessif de créer 8 classes pour seulement 15 départements. De même, il peut paraitre trop simpliste de classer dans uniquement deux classes plus de 200 pays.

A priori, en matière de discrétisation, il n'est pas pertinent de dépasser 8 classes, l'oeil ne pouvant alors faire une différence nette entre certains dégradés de couleurs.

Il existe certains indices permettant de connaitre le nombre idéal de classes.

$$Ncl = 1 + 3,3log(N) \text{ Huntsberger}$$

$$Ncl = 5 \times log(N) \text{ Brooks-Carruthers}$$

TP1

Une entreprise souhaite identifier les départements où se concentrent les professions intermédiaires et les cadres pour prendre des décisions d'implantation.

1. Dans Philcarto, testez différentes représentations obtenues par différentes discrétisations des taux de professions intermédiaires. Comment choisir la meilleure représentation ?
2. A l'aide d'Excel, réalisez un histogramme de la distribution statistique des taux de professions intermédiaires. Testez pour cela plusieurs amplitudes de classe. Quelle est la forme de cette distribution statistique ?
3. A partir de l'analyse de cette distribution, choisissez la meilleure méthode de discrétisation et le bon nombre de classes afin d'obtenir la meilleure représentation possible, puis terminez cette carte à l'aide d'Inkscape.
4. Reprenez les questions précédentes afin d'étudier la répartition des taux de cadres.

Corrélation et régression linéaire

Les statistiques multivariées

Définition

En statistiques, les analyses multivariées ont pour caractéristique de s'intéresser à la distribution conjointe de plusieurs variables. Les analyses bivariées sont des cas particuliers à deux variables.

Les analyses multivariées sont très diverses selon l'objectif recherché ou la nature des variables. On peut identifier deux grandes familles :

- celle des méthodes descriptives visant à structurer et résumer l'information ;
- celle des méthodes explicatives visant à expliquer une ou des variables dites "dépendantes" (variables à expliquer) par un ensemble de variables dites "indépendantes" (variables explicatives).

Corrélation et régression linéaire

Définitions

Corrélation

Etudier la corrélation entre deux ou plusieurs variables, c'est étudier l'intensité de la liaison qui peut exister entre ces variables.

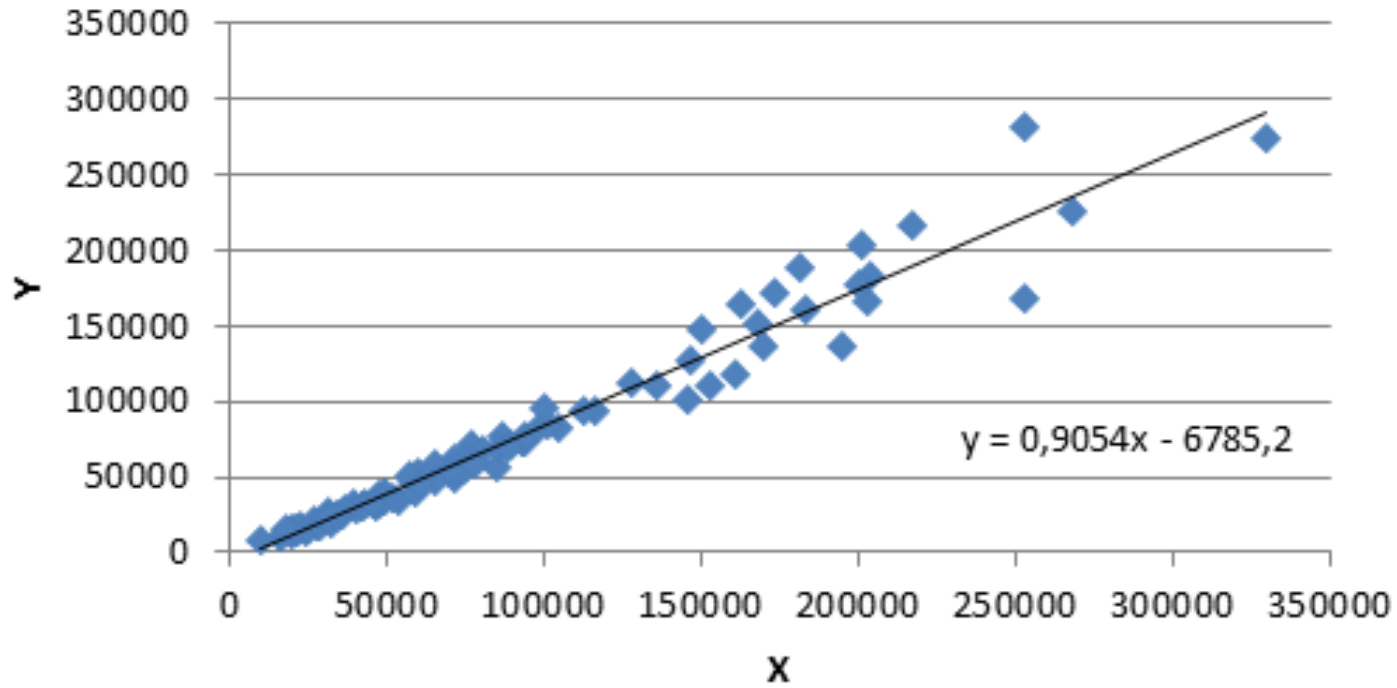
Régression linéaire

La régression est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres.

Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

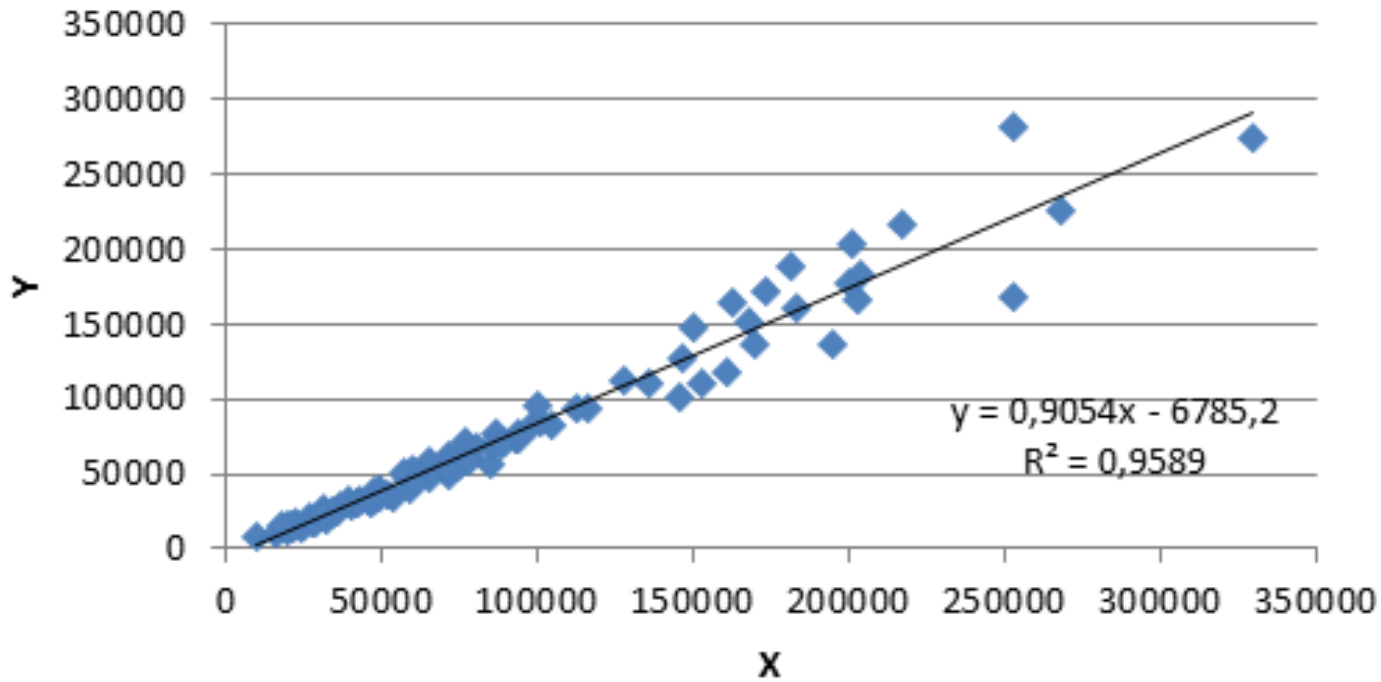
Pour faire simple, lorsque l'on étudie deux variables quantitatives, on peut produire un "nuage de points", une régression linéaire vise alors à résumer ce nuage de points par une forme plus simple à interpréter : une droite.



Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

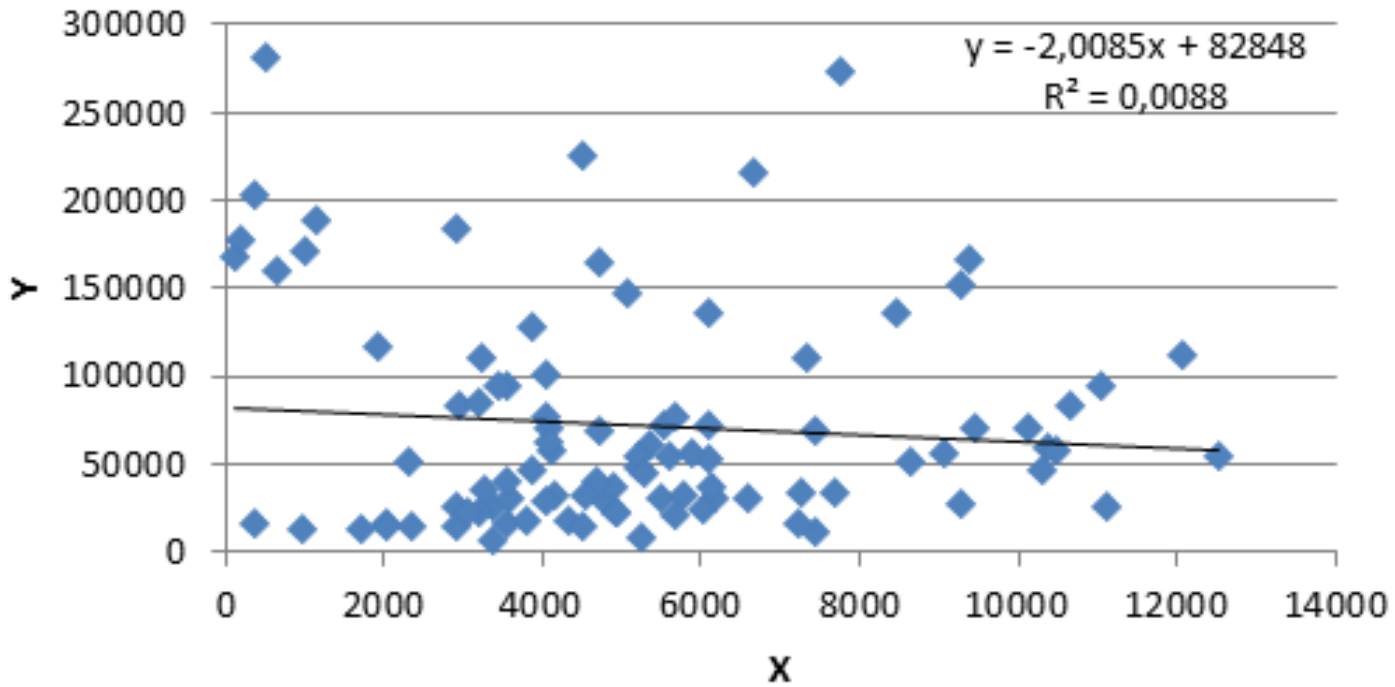
C'est le coefficient de corrélation ou le coefficient de détermination qui nous permet de dire si cette régression est "juste" :



Corrélation et régression linéaire

Des graphiques plutôt que des définitions ou des calculs pour comprendre

C'est le coefficient de corrélation ou le coefficient de détermination qui nous permet de dire si cette régression est "juste" ou pas du tout :



Corrélation et régression linéaire

La significativité

Dans les faits, il est important de savoir si le coefficient calculé est significativement différent de ce que l'on pourrait obtenir par hasard entre deux variables aléatoires de même taille.

Pour cela, il faut comparer la valeur t obtenue avec celle du tableau de Student pour n - 2 degrés de liberté :

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995	0.9975	0.9995
1	0.1584	0.3249	0.5095	0.7265	1	1.3764	1.9626	3.0777	6.3137	12.706	31.821	63.656	127.32	636.58
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.92	4.3027	6.9645	9.925	14.089	31.6
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408	7.4532	12.924
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.941	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041	5.5975	8.6101
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.015	2.5706	3.3649	4.0321	4.7733	6.8685
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168	5.9587
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.896	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995	4.0294	5.4081
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.306	2.8965	3.3554	3.8325	5.0414
9	0.1293	0.261	0.3979	0.5435	0.7027	0.8834	1.0997	1.383	1.8331	2.2622	2.8214	3.2498	3.6896	4.7809
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.5868
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.201	2.7181	3.1058	3.4966	4.4369
12	0.1283	0.259	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.681	3.0545	3.4284	4.3178

Corrélation et régression linéaire

La significativité

	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995	0.9975	0.9995
1	0.1584	0.3249	0.5095	0.7265	1	1.3764	1.9626	3.0777	6.3137	12.706	31.821	63.656	127.32	636.58
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.92	4.3027	6.9645	9.925	14.089	31.6
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408	7.4532	12.924
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.941	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041	5.5975	8.6101
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.015	2.5706	3.3649	4.0321	4.7733	6.8685
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168	5.9587
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.896	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995	4.0294	5.4081
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.306	2.8965	3.3554	3.8325	5.0414
9	0.1293	0.261	0.3979	0.5435	0.7027	0.8834	1.0997	1.383	1.8331	2.2622	2.8214	3.2498	3.6896	4.7809
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.5868
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.201	2.7181	3.1058	3.4966	4.4369
12	0.1283	0.259	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.681	3.0545	3.4284	4.3178

40	0.1265	0.255	0.3881	0.5286	0.6807	0.8507	1.05	1.3031	1.6839	2.0211	2.4233	2.7045	2.9712	3.551
41	0.1264	0.255	0.388	0.5285	0.6805	0.8505	1.0497	1.3025	1.6829	2.0195	2.4208	2.7012	2.967	3.5443
42	0.1264	0.255	0.388	0.5284	0.6804	0.8503	1.0494	1.302	1.682	2.0181	2.4185	2.6981	2.963	3.5377
43	0.1264	0.2549	0.3879	0.5283	0.6802	0.8501	1.0491	1.3016	1.6811	2.0167	2.4163	2.6951	2.9592	3.5316
44	0.1264	0.2549	0.3878	0.5282	0.6801	0.8499	1.0488	1.3011	1.6802	2.0154	2.4141	2.6923	2.9555	3.5258
45	0.1264	0.2549	0.3878	0.5281	0.68	0.8497	1.0485	1.3007	1.6794	2.0141	2.4121	2.6896	2.9521	3.5203
46	0.1264	0.2548	0.3877	0.5281	0.6799	0.8495	1.0482	1.3002	1.6787	2.0129	2.4102	2.687	2.9488	3.5149
47	0.1263	0.2548	0.3877	0.528	0.6797	0.8493	1.048	1.2998	1.6779	2.0117	2.4083	2.6846	2.9456	3.5099
48	0.1263	0.2548	0.3876	0.5279	0.6796	0.8492	1.0478	1.2994	1.6772	2.0106	2.4066	2.6822	2.9426	3.505
49	0.1263	0.2547	0.3876	0.5278	0.6795	0.849	1.0475	1.2991	1.6766	2.0096	2.4049	2.68	2.9397	3.5005
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778	2.937	3.496
60	0.1262	0.2545	0.3872	0.5272	0.6786	0.8477	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603	2.9146	3.4602
70	0.1261	0.2543	0.3869	0.5268	0.678	0.8468	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479	2.8987	3.435
80	0.1261	0.2542	0.3867	0.5265	0.6776	0.8461	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387	2.887	3.4164
90	0.126	0.2541	0.3866	0.5263	0.6772	0.8456	1.0424	1.291	1.662	1.9867	2.3685	2.6316	2.8779	3.4019
100	0.126	0.254	0.3864	0.5261	0.677	0.8452	1.0418	1.2901	1.6602	1.984	2.3642	2.6259	2.8707	3.3905
110	0.126	0.254	0.3863	0.5259	0.6767	0.8449	1.0413	1.2893	1.6588	1.9818	2.3607	2.6213	2.8648	3.3811
120	0.1259	0.2539	0.3862	0.5258	0.6765	0.8446	1.0409	1.2886	1.6576	1.9799	2.3578	2.6174	2.8599	3.3734
130	0.1259	0.2539	0.3862	0.5257	0.6764	0.8444	1.0406	1.2881	1.6567	1.9784	2.3554	2.6142	2.8557	3.367
140	0.1259	0.2538	0.3861	0.5256	0.6762	0.8442	1.0403	1.2876	1.6558	1.9771	2.3533	2.6114	2.8522	3.3613
infini (loi normale)	0.1257	0.2533	0.3853	0.5244	0.6744	0.8416	1.0364	1.2816	1.6449	1.96	2.3264	2.5759	2.8072	3.2908

Si $| t | > t_{seuil}$, on rejette l'hypothèse nulle pour le risque choisi.

Corrélation et régression linéaire

Les résidus

Définition

Un résidu est dans une régression le terme qui n'est pas expliqué par la ou les variables explicatives.

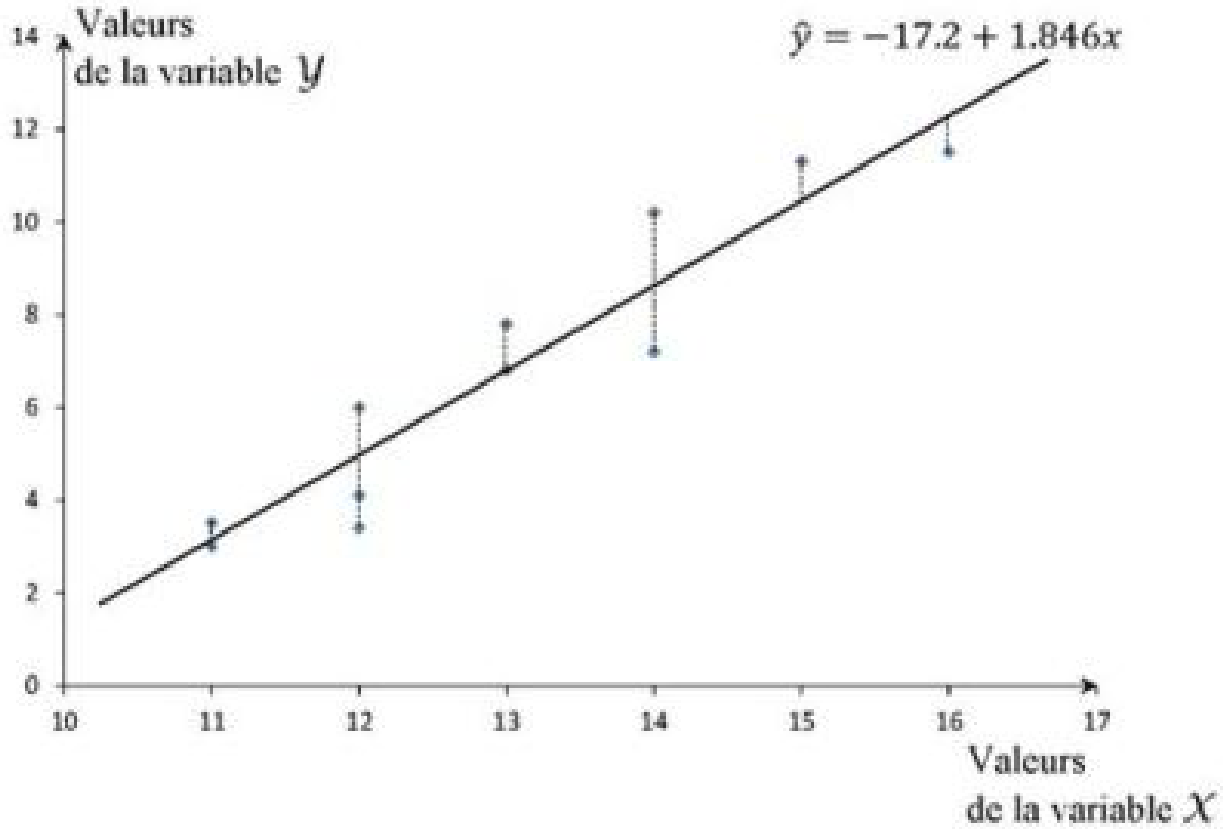
Il se calcule simplement en calculant l'écart entre la valeur réelle de y et la valeur théorique de y (obtenue à partir de l'équation déterminée par la régression linéaire) :

$$e_i = Yi - \hat{Y}i$$

$$\hat{Y}i = aXi - b$$

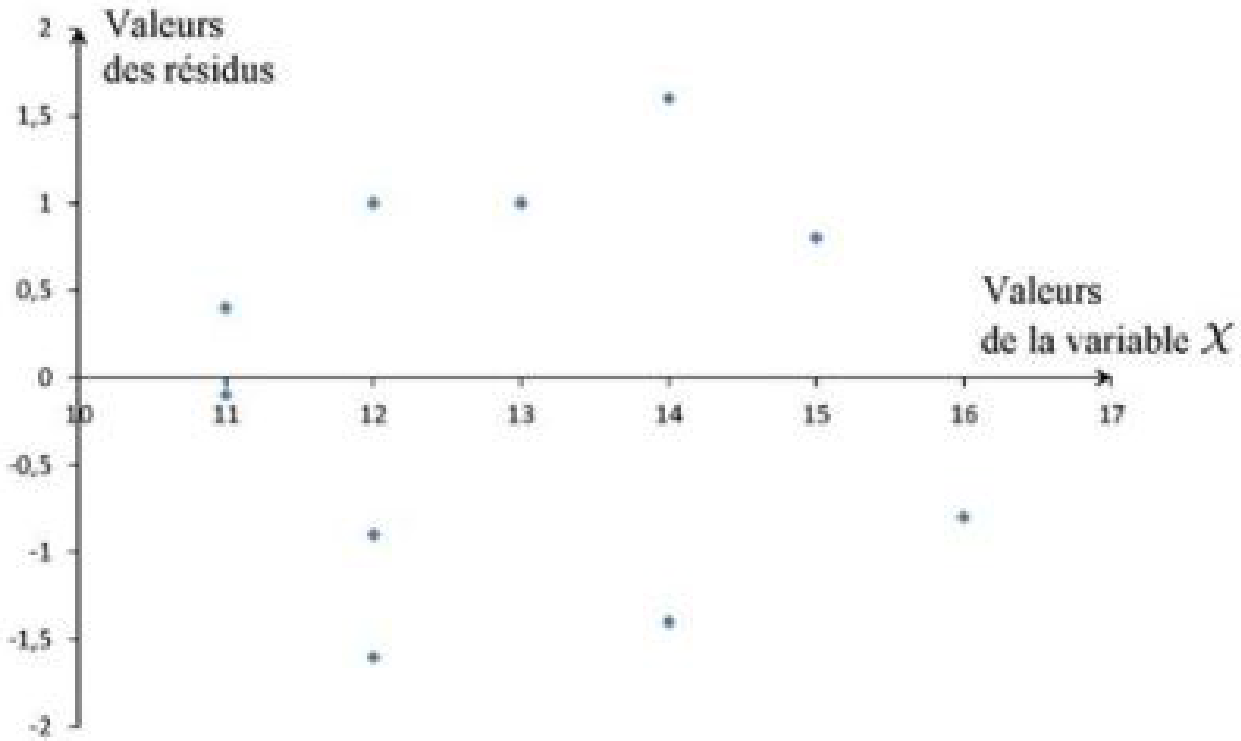
Corrélation et régression linéaire

Les résidus



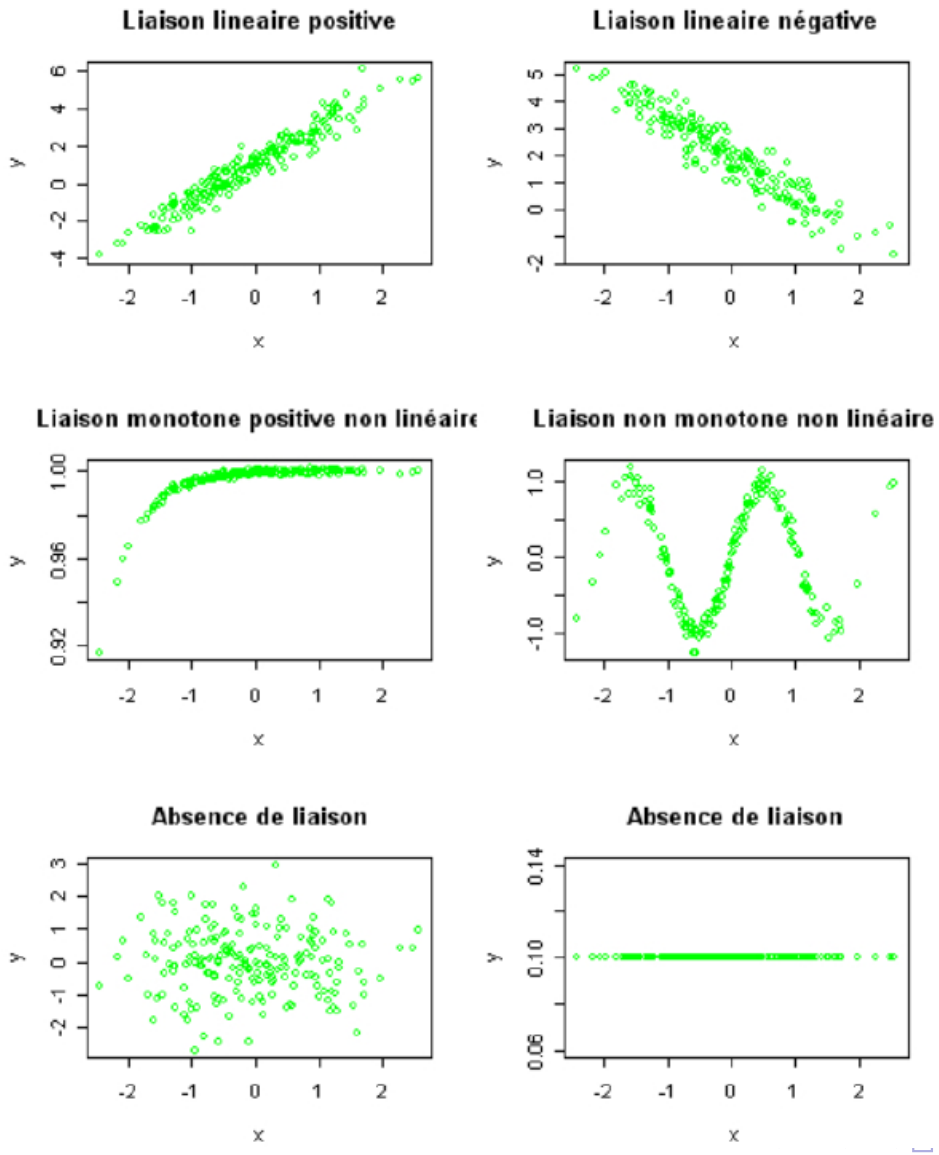
Corrélation et régression linéaire

Les résidus



Corrélation et régression linéaire

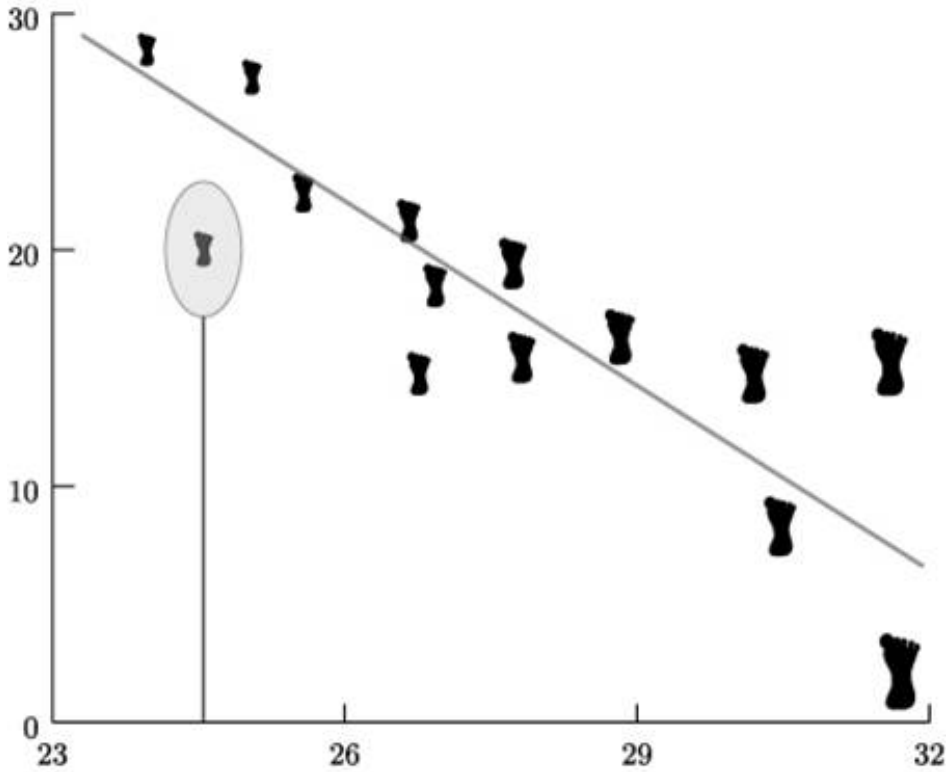
Différentes formes de nuages de points



Corrélation et régression linéaire

Les pièges à éviter : Des relations problématiques

Nombre de fautes d'orthographe en fonction de la pointure. Les élèves ayant les plus grands pieds font moins de fautes.



Corrélation et régression linéaire

Les pièges à éviter : Attention à l'erreur écologique

En géographie, l'étude des corrélations se fait à travers l'analyse d'un ensemble de lieux, de territoires au sein desquels on a des agrégats (des quantités).

Ainsi lorsque les variables décrivant ces lieux sont des attributs sociaux décrivant des habitants, il faut toujours faire attention au fait qu'une corrélation établie au niveau des lieux n'implique pas forcément une corrélation au niveau des individus.

Une étude menée au niveau des individus (sociologique) peut montrer que le taux de criminalité est plus élevé chez les autochtones que chez les étrangers. Pourtant, cette étude au niveau des quartiers (géographique) peut montrer une corrélation parfaite entre la proportion d'étrangers des quartiers et leur taux de criminalité.

Il faut faire attention à ne pas "individualiser" une corrélation issue d'un agrégat.

TP2

Une entreprise s'intéressant à toutes les CSP demande de lui synthétiser des résultats avec un minimum de cartes, mais aussi de produire des analyses fines présentant certaines spécificités de ce territoire.

1. Calculez la corrélation entre la variable à expliquer "Employé" et la variable explicative "Profession Intermédiaire". Faites de même entre la variable à expliquer "Employé" et la variable explicative "Agriculteur". Commentez les résultats obtenus. Qu'est-il possible de faire cartographiquement ?
2. Calculez les résidus de la relation Employé, Profession Intermédiaire. Représentez ces résidus sur une carte après les avoir interprétés.
3. A partir des données du TP1, calculez la corrélation entre la variable à expliquer "Tx_Employé " et la variable explicative "Tx_Profession Intermédiaire". Commentez les résultats obtenus.

CAH

La classification ascendante hiérarchique

Nous l'avons vu, une discrétisation (une classification) consiste à regrouper entre eux les éléments qui ressemblent le plus (en classes homogènes).

Si la question de la ressemblance est triviale lorsque l'on étudie une seule variable quantitative, cela devient moins évident lorsque l'on étudie plusieurs variables.

Le premier enjeu est donc ici de mesurer la ressemblance (ou au contraire la dissemblance) entre les objets étudiés.

La CAH est simplement une méthode permettant de regrouper les objets entre eux à partir d'un tableau de ressemblance (dissemblance).

CAH

La classification ascendante hiérarchique

	Xi (km)	Yi (km)
Paris	600	2428
Marseille	846	1815
Saint-Etienne	760	2050
Bordeaux	369	1986
Reims	723	2474
Lyon	794	2087

$$Distance(euclidienne) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

$$Dist_{(Paris-Marseille)} = \sqrt{(600 - 846)^2 + (2428 - 1815)^2} = 660$$

CAH

La classification ascendante hiérarchique

Distance euclidienne : $\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$

$De_{(P-M)} = \sqrt{(600 - 846)^2 + (2428 - 1815)^2} = 660$

Distance de Manhattan : $\mid X_1 - X_2 \mid + \mid Y_1 - Y_2 \mid$

$Dm_{(P-M)} = \mid 600 - 846 \mid + \mid 2428 - 1815 \mid = 246 + 613 = 859$

Distance de Tchebychev : $Max[(X_1 - X_2); (Y_1 - Y_2)]$

$Dt_{(P-M)} = Max[(600 - 846); (2428 - 1815)] = Max[246; 613] = 613$

CAH

La classification ascendante hiérarchique

	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0

CAH

La classification ascendante hiérarchique

	Variable 1	Variable 2	Variable 3	Variable 4
Objet 1	5	2	6	4
Objet 2	2	5	2	4

$$Dist_{(Paris-Marseille)} = \sqrt{(5-2)^2 + (2-5)^2 + (6-2)^2 + (4-4)^2}$$

CAH

La classification ascendante hiérarchique

La classification ascendante hiérarchique (CAH) est une méthode de classification itérative dont le principe est simple :

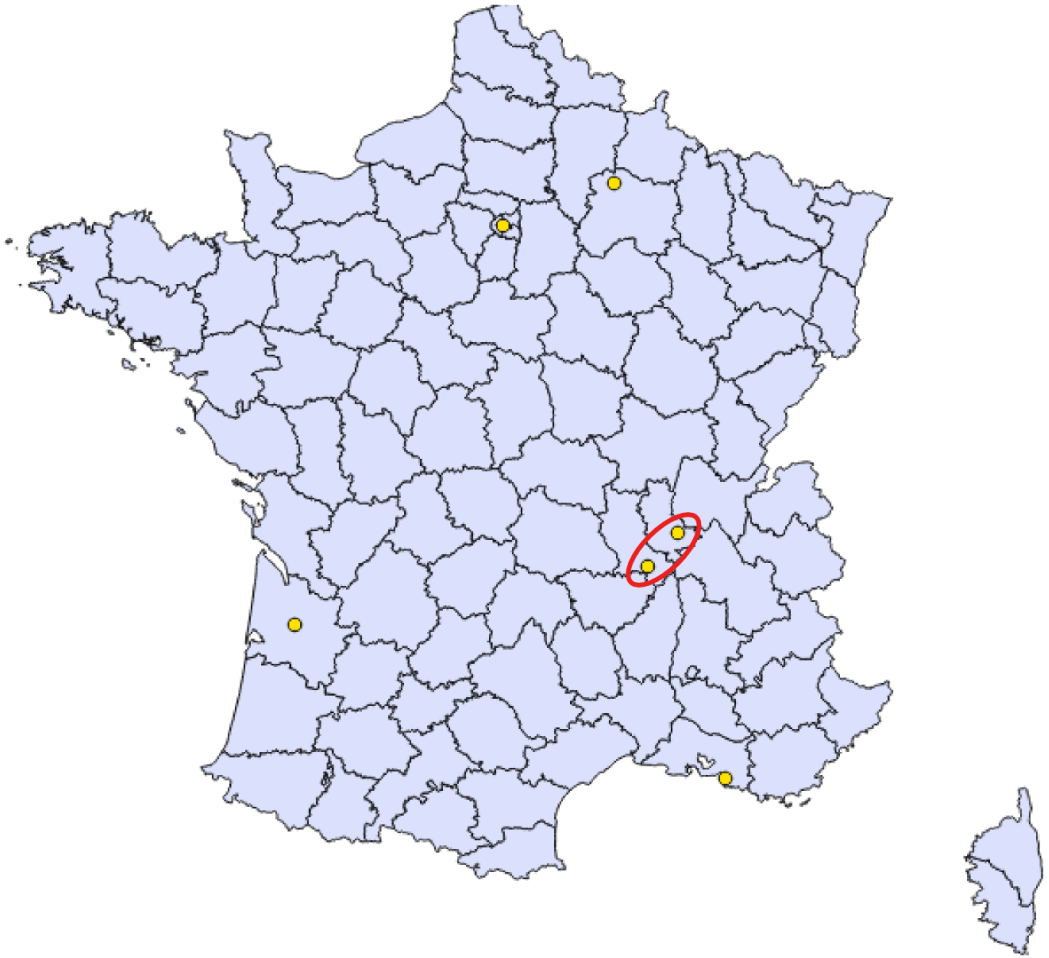
- On commence par calculer la dissimilarité entre les N objets.
- Puis on regroupe les deux objets dont le regroupement minimise un critère d'agrégation donné, créant ainsi une classe comprenant ces deux objets.
- On calcule ensuite la dissimilarité entre cette classe et les N-2 autres objets en utilisant le critère d'agrégation. Puis on regroupe les deux objets ou classes d'objets dont le regroupement minimise le critère d'agrégation.

On continue ainsi jusqu'à ce que tous les objets soient regroupés.

CAH

La classification ascendante hiérarchique

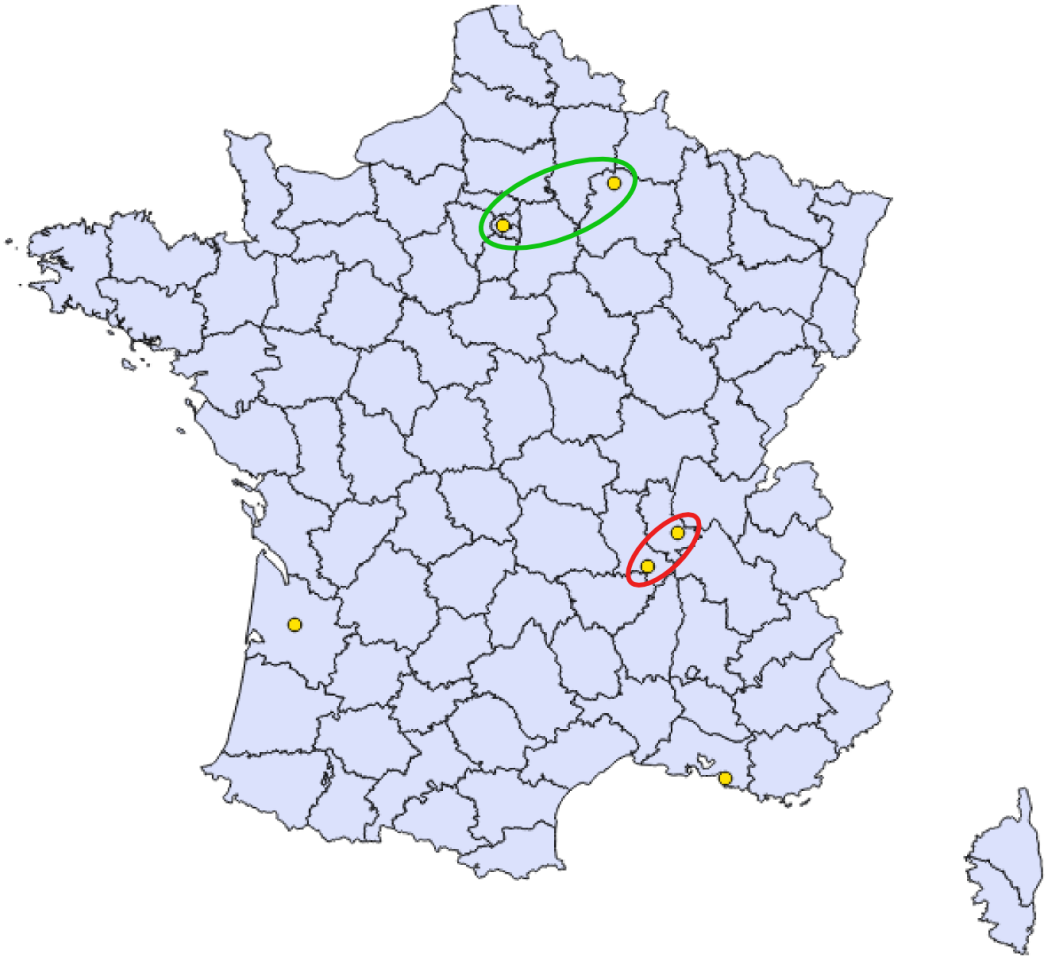
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0



CAH

La classification ascendante hiérarchique

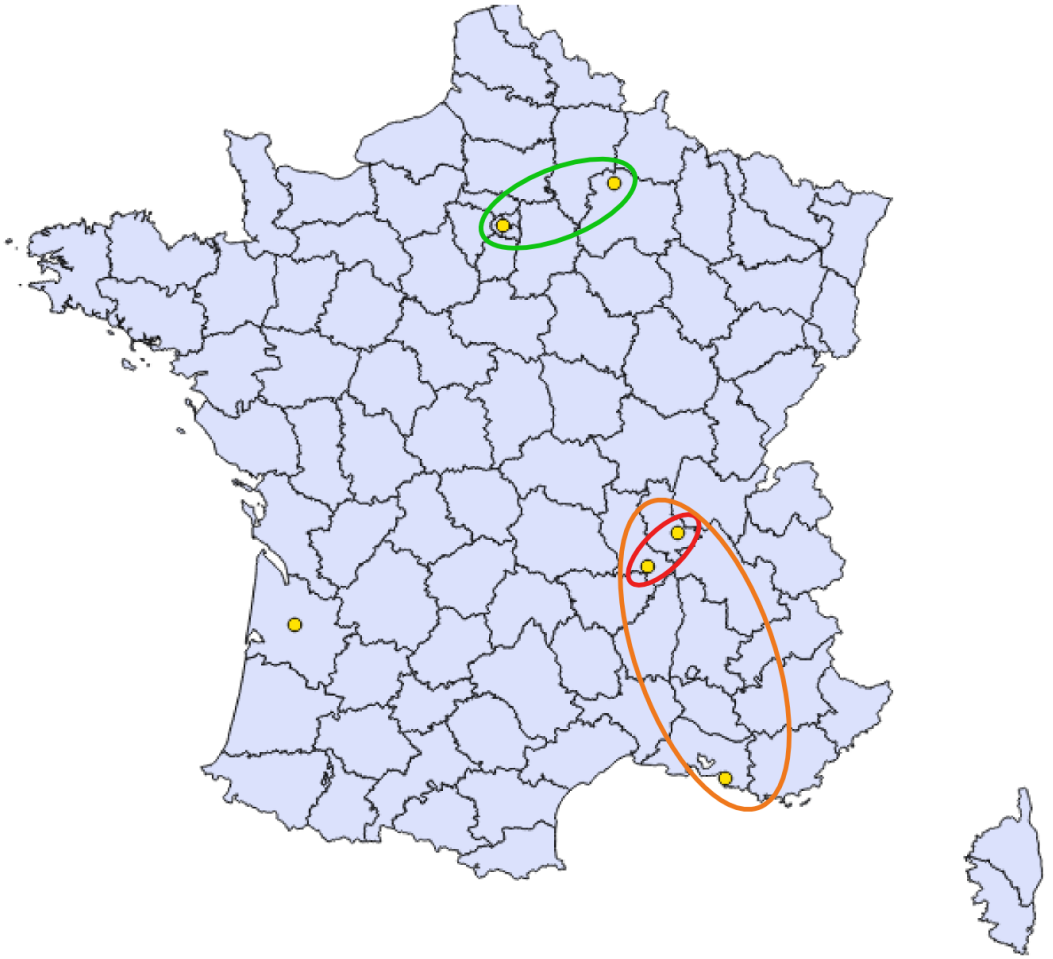
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0



CAH

La classification ascendante hiérarchique

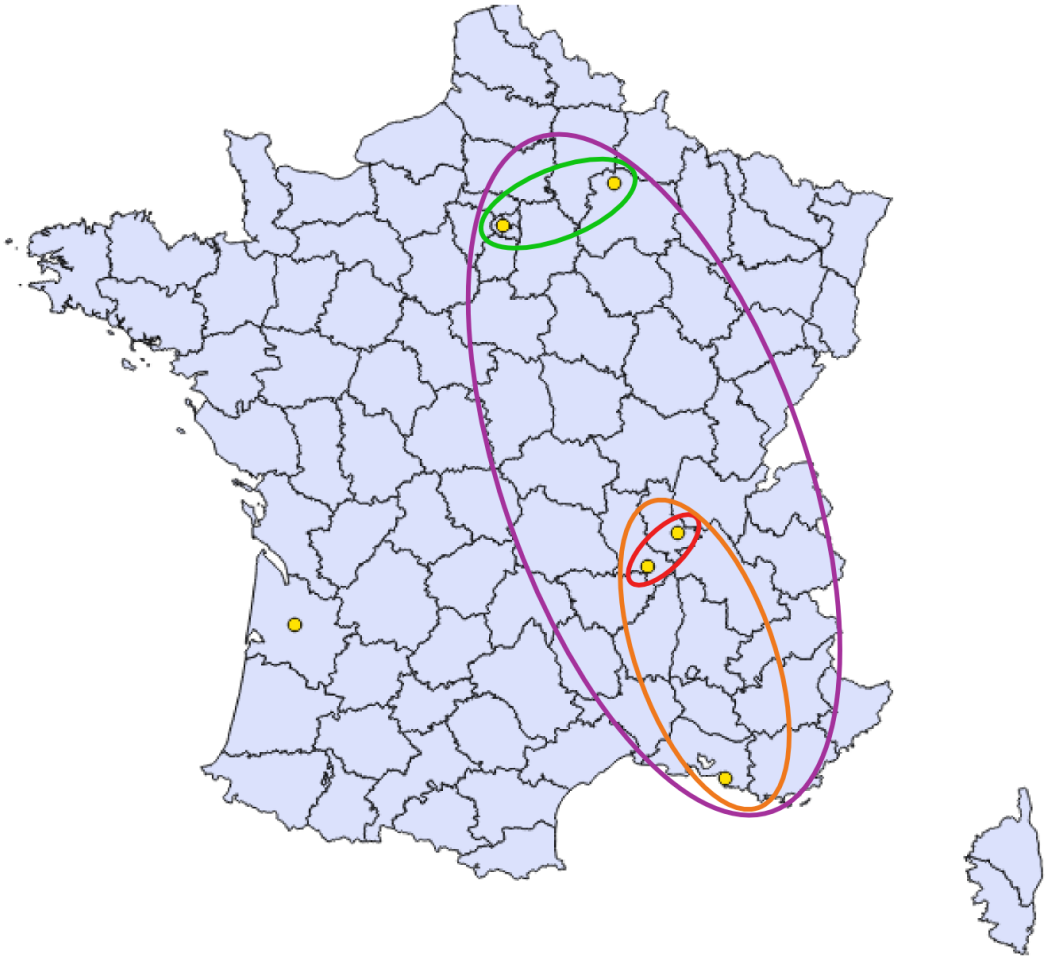
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0



CAH

La classification ascendante hiérarchique

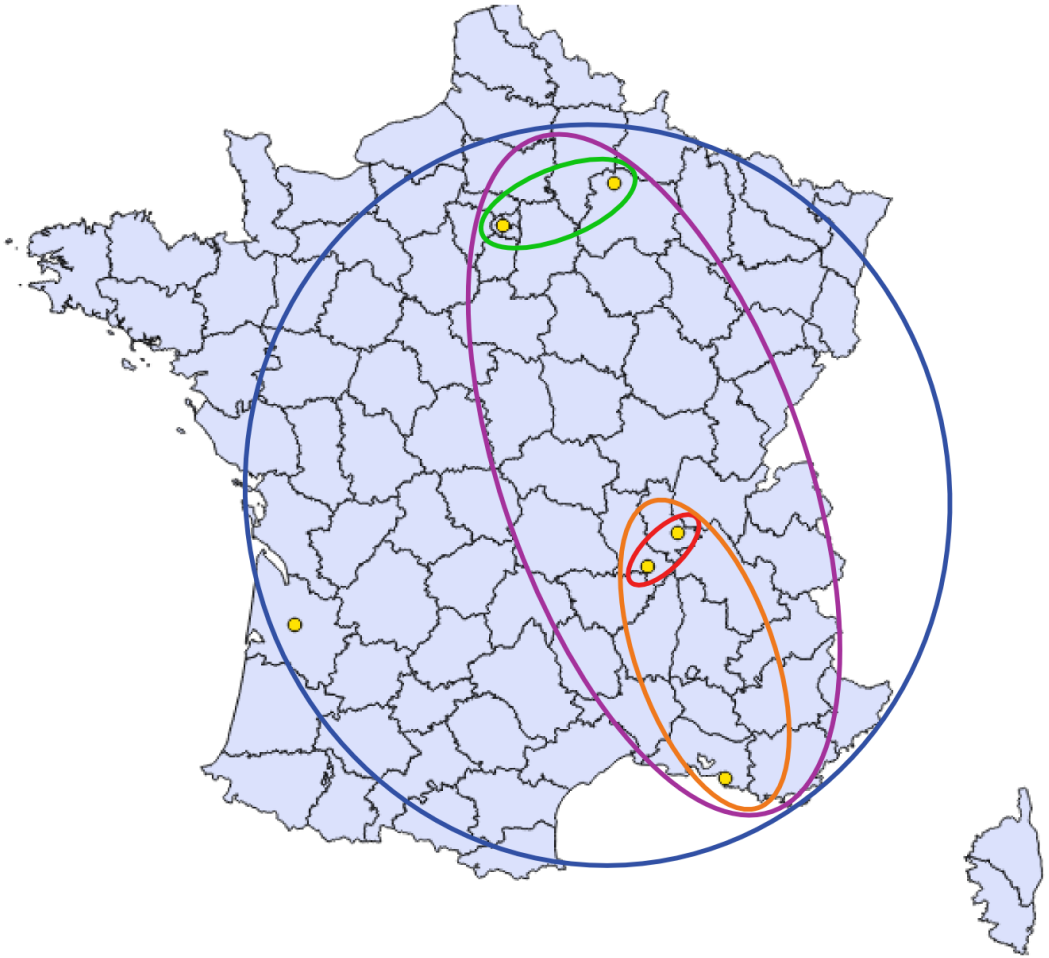
	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0



CAH

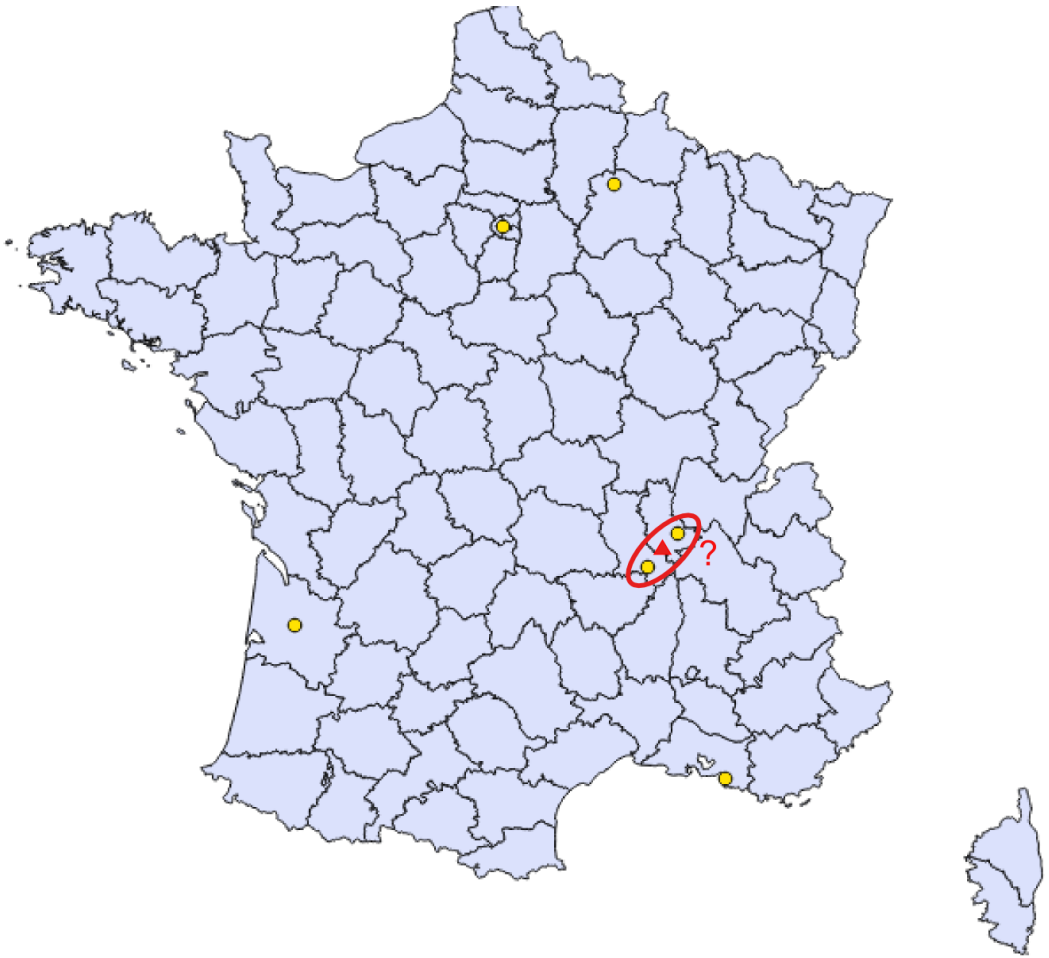
La classification ascendante hiérarchique

	Paris	Marseille	Saint-Etienne	Bordeaux	Reims	Lyon
Paris	0	660	410	498	131	392
Marseille	660	0	250	506	670	276
Saint-Etienne	410	250	0	396	425	50
Bordeaux	498	506	396	0	602	436
Reims	131	670	425	602	0	393
Lyon	392	276	50	436	393	0



CAH

La classification ascendante hiérarchique



CAH

La classification ascendante hiérarchique

Le saut minimum retient le minimum des distances entre individus de C1 et C2. C'est ce critère qu'on a appliqué précédemment.

Le saut maximum est la dissimilarité entre les individus de C1 et C2 les plus éloignés.

Le lien moyen consiste à calculer la moyenne des distances entre les individus de C1 et C2.

La distance de Ward vise à maximiser l'inertie inter-classe.

TP3

Une entreprise s'intéressant à toutes les CSP demande de lui synthétiser les résultats sur une seule carte.

1. A l'aide de Philcarto, produisez une CAH portant sur les différentes CSP.
2. Commencez par interpréter les résultats obtenus pour un partition en deux classes. Allez comme cela jusqu'à une partition en six classes.
3. Retenez la partition qui semble la plus pertinente. Puis a l'aide d'Inkscape produisez une cartographie finale.