

VÉRONIQUE BEAULIEU

ÉTUDE DE LA VISUALISATION GÉOGRAPHIQUE DANS UN ENVIRONNEMENT D'EXPLORATION INTERACTIVE DE DONNÉES GÉODÉCISIONNELLES

Adaptation et améliorations

Mémoire présenté

à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en sciences géomatiques
pour l'obtention du grade de maître ès sciences (M. SC.)

DÉPARTEMENT DES SCIENCES GÉOMATIQUES
FACULTÉ DE FORESTERIE ET DE GÉOMATIQUES
UNIVERSITÉ LAVAL
QUÉBEC

2009

Résumé

La visualisation géographique de l'information est un aspect important du SOLAP. Elle est le troisième élément clé pour supporter une analyse facile, rapide et intuitive des données prônée par l'approche SOLAP. Cependant, à ce jour, aucune étude ne s'est intéressée spécifiquement à cet aspect du SOLAP et les stratégies de visualisation utilisées traditionnellement dans les SIG ne conviennent pas à l'approche interactive et multidimensionnelle. Ce mémoire présente une recherche exploratoire proposant premièrement de placer la visualisation géographique au cœur de la relation entre l'utilisateur et la base de données. Cette stratégie fait ressurgir plusieurs besoins et possibilités explorées au cours de ce mémoire. Le respect de la flexibilité d'analyse SOLAP, les divers types de visualisation nécessaires pour rendre les différents types de requêtes, puis les manipulations graphiques qui peuvent s'opérer dans un environnement SOLAP font l'objet de propositions. Deuxièmement, ce mémoire propose une méthode permettant une relation directe entre les données et leur visualisation, sans l'intermédiaire de l'utilisateur. La solution théorique résultante respecte la philosophie d'exploration interactive prônée par l'approche SOLAP. Elle suggère l'intégration de l'expertise en visualisation géographique à l'intérieur même de la technologie SOLAP par l'ajout de métadonnées et d'une base de connaissances sémiologiques. Ce premier pas vers une amélioration du processus de visualisation avec SOLAP ouvre la porte à plusieurs recherches ultérieures.

Avant-propos

Mon objectif académique, depuis que j'étais toute petite, était d'obtenir un diplôme de maîtrise. C'est donc avec une grande fierté que je dépose ce mémoire, le fruit de mon travail ces deux dernières années. Lorsque j'ai entrepris cette maîtrise, je n'avais qu'une vague idée à quoi ressembleraient mes deux prochaines années. L'aventure qui me guettait est quasiment indescriptible. Certains la décrivent comme une montagne russe, une corde raide à l'équilibre fragile, un inconnu dans lequel on bâtit son chemin. Pour ma part, je dirais que c'est un peu tout à la fois. Mon aventure a été empreinte de hauts et de bas, d'émotivité, de fébrilité et d'engagement. Au-delà des connaissances théoriques et des nombreux échanges, parfois enflammés, la maîtrise a été, pour moi, un moment pour me dépasser personnellement, intellectuellement et professionnellement. Le chemin que j'y ai tracé devient aujourd'hui une partie intégrante de moi-même.

Mes premiers remerciements vont à mon compagnon de vie, Philippe, qui au cours de ces deux dernières années a su me supporter, me deviner et m'épauler autant dans les moments joyeux que dans ceux qui l'étaient moins. Merci pour ton accompagnement, ta patience, ton écoute, ta lecture attentive et les distractions nécessaires à une bonne santé mentale. Le mémoire que je remets aujourd'hui n'aurait pas été celui-ci sans ton aide.

Merci également à Mathieu Bertrand et Ève Grenier, mes compagnons d'aventure. Embarqués dans le même bateau, ils m'ont aidé à tenir le cap, à réparer la barque lorsque c'était nécessaire et à désamorcer quelques bombes. Les échanges, les rires et les motivations de part et d'autre ont contribué à l'avancement de cette maîtrise, mais aussi au relâchement de la pression et au façonnement de bons souvenirs.

Merci à mon directeur, Yvan Bédard pour m'avoir accueilli dans son équipe. Le défi qui m'a été confié était motivant et l'environnement, agréable. Merci Yvan pour tes conseils qui font réfléchir et qui rend tout l'« anormal », « normal » en situation de maîtrise. Merci aussi de m'avoir accordé ta confiance et de m'avoir permis de réaliser une conférence à l'étranger.

Merci à l'équipe de professionnels de la chaire de recherche industrielle en base de données géodécisionnelles. Ce mémoire est le résultat des collaborations et des idées échangées ces derniers deux ans.

Merci à ma famille, Papa, maman et à tous ceux qui ont déjà traversé l'aventure de la maîtrise. Merci de vos conseils et de vos encouragements.

Finalement, merci aux organismes qui ont subventionné cette recherche : la chaire de recherche en bases de données géodécisionnelles et ses partenaires, et le Conseil de recherche en sciences naturelles et génie qui m'a accordé une bourse d'étude supérieure. Le support financier m'a permis de me concentrer à 120% sur ma recherche et de la mener à terme sans souci.

*À mon compagnon de vie, Philippe
et à la vie, tout simplement*

Table des matières

Résumé	i
Avant-propos	ii
Table des matières	v
Liste des tableaux	vii
Liste des figures	viii
CHAPITRE 1 - Introduction.....	1
1.1 Mise en contexte	1
1.2 Problématique	5
1.3 Solution proposée.....	7
1.4 Motivations et objectifs.....	9
1.5 Méthodologie	12
1.6 Synopsis	17
CHAPITRE 2 - Revue de littérature	18
2.1 Les données, l'information et les connaissances	18
2.1.1 La hiérarchie DIKW	19
2.1.2 Des données à l'information aux connaissances	22
2.2 La pensée visuelle (<i>visual thinking</i>)	23
2.2.1 La vision et le fonctionnement perceptuel	25
2.2.2 La visualisation de l'information	28
2.2.3 La perception visuelle : pouvoir et limites	29
2.3 Découvrir l'inconnu	31
2.3.1 La découverte de connaissances dans les données (<i>KDD</i>).....	32
2.3.2 L'exploration des données (<i>EDA</i>)	34
2.4 La géovisualisation.....	35
2.4.1 La cartographie traditionnelle et la géovisualisation	36
2.4.2 La géovisualisation et la sémiologie graphique.....	39
2.4.3 Au-delà de Bertin : les méthodes, les techniques et les outils	44
2.5 Synopsis	50
CHAPITRE 3 - SOLAP et Géovisualisation.....	51
3.1 SOLAP et l'exploration interactive des données	51
3.1.1 La flexibilité d'analyse des données : facile, rapide et intuitive.....	51
3.1.2 Les variables et les requêtes de l'information	53
3.1.3 La navigation dans les données.....	57
3.2 SOLAP et la visualisation géographique : les besoins	58
3.2.1 La flexibilité de la visualisation géographique : facile, rapide et intuitive	59

3.2.2	Les types de visualisation géographique.....	60
3.2.3	Les manipulations graphiques.....	61
3.3	SOLAP et la visualisation géographique : les améliorations potentielles.....	63
3.3.1	L'automatisation de la visualisation géographique.....	63
3.3.2	L'accès à différents types de visualisation géographique adaptés	64
3.3.3	La diversité des manipulations graphiques.....	69
3.4	Synopsis	75
CHAPITRE 4 - Proposition d'une méthode de définition de la visualisation géographique des données		76
4.1	La méthode.....	76
4.1.1	La taxonomie des concepts	77
4.1.2	L'intégration des éléments de la méthode : les métadonnées et la base de connaissances sémiologiques	99
4.1.3	La base de connaissances sémiologiques.....	104
4.2	Les validations préliminaires	113
4.2.1	Les tests de faisabilité	113
4.2.2	Les expérimentations exploratoires.....	119
4.3	L'analyse critique	137
4.3.1	Les points forts	137
4.3.2	Les points faibles	139
CHAPITRE 5 - Conclusion et perspectives.....		142
5.1	Retour sur les objectifs de recherche.....	142
5.2	Perspectives d'avenir.....	145
Bibliographie		148
Annexe 1 - Taxonomie des concepts		165
Annexe 2 - Modèle Conceptuel de la base de connaissances sémiologiques.....		167
Annexe 3 - Description des modèles de représentation		170

Liste des tableaux

Tableau 4-1 : Exemple de description de la distribution	114
Tableau 4-2 : Exemple de description du changement.....	115
Tableau 4-3 : Exemple de description des relations entre les mesures	115
Tableau 4-4 : Caractéristiques des données de la requête 1.....	121
Tableau 4-5 : Les propositions de la base de connaissances sémiologiques	125
Tableau 4-6 : Caractéristiques des données de la requête 2.....	127
Tableau 4-7 : Exemples possibles de modèles de représentations formant la composition répondant à la requête 2.....	128
Tableau 4-8 : Exemples de superposition multidimensionnelle possible répondant à la requête 2	130
Tableau 4-9 : Caractéristiques des données de la requête 3.....	132
Tableau 4-10 : Exemples de représentations multivariées pour la requête 3.....	136
Tableau 4-11 : Exemples de représentation simples pour la requête 3	136

Liste des figures

Figure 1.1 : Exemple d'un cube à 3 dimensions de la superficie de territoire en culture	2
Figure 1.2 : Stratégie d'exploration d'une approche SIG. Inspiré de [Keim et <i>al</i> , 1996]	6
Figure 1.3 : Stratégie d'une approche d'exploration de données. Inspiré de [Keim et <i>al</i> , 1996] (1) premier volet de la solution et (2) deuxième volet de la solution	8
Figure 1.4 : Motivations et objectif général de la recherche	11
Figure 1.5 : Méthodologie de recherche.....	16
Figure 1.6 : Positionnement des chapitres selon la nouvelle stratégie de visualisation proposée ..	17
Figure 2.1 : La hiérarchie DIKW. Traduit de [Ackoff, 1989].....	19
Figure 2.2 : La hiérarchie DIKW ou le continuum de la compréhension. Traduit de [Bellinger et <i>al</i> , 2004] et [Clark, 2004]	21
Figure 2.3 : Association des différents types de systèmes d'information et des concepts DIKW. Traduit de [Rowley, 2007]	22
Figure 2.4 : Les hémisphères cérébraux et leurs principales fonctions. Inspiré de [UCNAS Canada, 2007].....	24
Figure 2.5 : Fonctionnement perceptuel. Traduit de [Ware, 2008, p.9].....	26
Figure 2.6 : Cognition interne et externe	27
Figure 2.7 : Modèle de la visualisation. Source [Card, MacKinlay et Shneiderman, 1999]	29
Figure 2.8 : Différentes phases dans la méthode scientifique et les modes de raisonnement qui s'y appliquent. Traduit de [Gahegan et <i>al</i> , 2001]	33
Figure 2.9 : Le rôle de la carte dans le processus de construction des connaissances spatiales. Traduit de [Kraak, 2006]	37
Figure 2.10 : Types de représentations visuelles	38
Figure 2.11 : Définition de la cartographie selon les modes d'utilisation (<i>cartography</i>) ³ . Source [MacEachren, 1994b, p.6]	38
Figure 2.12 : Classement des variables visuelles selon le niveau d'organisation de la perception selon [Bertin, 1973; MacEachren, 1994a; Slocum et <i>al</i> , 2005]	40
Figure 2.13 : Perception a) associative, b) dissociative, c) sélective, d) ordonnée et e) quantitative	41
Figure 2.14 : Inclusion des attitudes perceptives. Source [Bertin, 1973]	42
Figure 2.15 : La synoptique des constructions graphiques. Source [Bertin, 1977, p.29].....	44
Figure 2.16 : Exemples de représentation de données multivariées et multidimensionnelles en une seule image a) <i>SOM</i> de couleur encodée, b) <i>parallel coordinate plot</i> , c) cube spatiotemporel, d) <i>pixel-dense map</i> , e) <i>star plot map</i> et f) cartogramme de Dorling.....	45
Figure 2.17 : Exemples de représentation des données en plusieurs images a) composition, b) répétition et c) réexpression	46
Figure 2.18 : Exemples de techniques d'interaction a) image originale b) agrégation des données, c) filtrage de données et d) élagage (<i>brushing</i>) de données.....	47
Figure 2.19 : Autres exemples d'interaction a) synchronisation de la sélection (<i>linking</i>) b) triage (<i>ordering</i>) c) présentation graduelle (<i>successive build-up</i>)	48

Figure 2.20 : Exemples d'outils de géovisualisation a) CommonGIS b) GeoViz Toolkit c) GeoDa	49
Figure 3.1 : Les trois types de dimensions spatiales supportées par le SOLAP	54
Figure 3.2 : Les types de requêtes a) la requête simple, b) la requête multivariée, c) la requête avec multi-sélections dans 2 et 3 dimensions et d) la requête multi-niveau.....	56
Figure 3.3 : Exemples de forage a) l'image initiale, b) le forage par membre sur l'Afrique, c) le forage par niveau et d) l'ouverture. Source [Bédard et al, 2008]	57
Figure 3.4 : Exemples d'une opération de pivot sur un tableau et sur une carte. Source [Bédard et al, 2008].....	58
Figure 3.5 : Visualisation géographique des mesures et de leur combinaison : a) carte multivariée et b) répétition de carte simple.....	65
Figure 3.6 : Composition selon 3 dimensions	65
Figure 3.7 : Répétition selon 3 dimensions : Provenance Mondiale, Programme de DEC, Temps...	66
Figure 3.8 : Exemple d'une superposition de diagrammes sur une carte	67
Figure 3.9 : Carte multi-niveau du nombre de ménages au niveau des municipalités et des cellules et le diagramme équivalent.	68
Figure 3.10 : Comparaison de données d'évaluation de performance sportive (cf. projet en kinésiologie de M. Lambert et J.P. Veilleux) a) Diagramme et b) Carte multiple relative à la moyenne.....	69
Figure 3.11 : Exemple de comparaison interactive. Source [Fraunhofer IAIS, 2007]	71
Figure 3.12 : Changement de précision symbolique a) dans un nuage de point, b) dans un histogramme	72
Figure 3.13 : Regroupement des profils de valeurs des mesures	74
Figure 4.1 : Les caractéristiques intrinsèques aux données	77
Figure 4.2 : Les variables de la structure multidimensionnelle	78
Figure 4.3 : Les références	79
Figure 4.4 : Les niveaux d'organisation.....	80
Figure 4.5 : Les caractéristiques de la relation entre les mesures et les dimensions.....	81
Figure 4.6 : La distribution	82
Figure 4.7 : Le changement.....	82
Figure 4.8 : Modèle de données selon le changement et la distribution. Traduit de [MacEachren, 1994a].....	83
Figure 4.9 : Le type d'agrégation	84
Figure 4.10 : Les caractéristiques graphiques.....	85
Figure 4.11 : Les attitudes perceptives	86
Figure 4.12 : Les variables visuelles	87
Figure 4.13 : Les implantations graphiques	88
Figure 4.14 : Relation entre l'implantation graphique et l'implantation géométrique.....	89
Figure 4.15 : Carte de livraison utilisant la taille comme variable visuelle a) en implantation ponctuelle et b) en implantation linéaire.....	90
Figure 4.16 : Relation entre l'implantation graphique et la géométrie d'un point de vue de visualisation de l'information.....	90
Figure 4.17 : Les arrangements graphiques.....	91

Figure 4.18 : Définition des arrangements graphiques a) Axe Y, b) Axe X, c) Plan, d) Superposition et e) le figuré	92
Figure 4.19 : Arrangement graphique a) d'une carte multivariée et b) d'une multicarte	93
Figure 4.20 : Les caractéristiques d'ensemble	93
Figure 4.21 : Les constructions graphiques.....	94
Figure 4.22 : Exemples de diagrammes a) orthogonal, b) rectiligne, c) circulaire et d) polaire	94
Figure 4.23 : Les relations entre les variables de même type.....	95
Figure 4.24 : Les niveaux de lecture.....	96
Figure 4.25 : Synthèse de la taxonomie des concepts	98
Figure 4.26 : Cas d'utilisation de l'approche SOLAP proposée	101
Figure 4.27 : Exemple de modèles de représentation et de la description de leur combinaison a) une mesure quantitative thématique sur une carte choroplèthe et b) un groupe de membres temporels dans un diagramme à lignes.	102
Figure 4.28 : Diagramme UML de collaboration	103
Figure 4.29 : Diagramme d'activité lorsqu'une nouvelle requête est exécutée.....	105
Figure 4.30 : Diagramme d'activité lorsqu'un changement est apporté à la symbologie par l'utilisateur.....	106
Figure 4.31 : Modèle conceptuel de la base de connaissances sémiologiques.....	108
Figure 4.32 : Regroupement des variables d'une requête multivariée selon les types de visualisation : a) une visualisation multivariée et b) une répétition	110
Figure 4.33 : Regroupement des variables d'une requête avec multi-sélections dans deux dimensions selon les types de visualisation : a) une visualisation par composition, b) une répétition et c) une superposition.	111
Figure 4.34 : Métadonnée de visualisation.....	114
Figure 4.35 : Description d'une carte choroplèthe conventionnelle	117
Figure 4.36 : Description d'un diagramme en pyramide	118
Figure 4.37 : Description d'une carte utilisant la taille et la forme d'un symbole sur une implantation zonale.....	122
Figure 4.38 : Représentation de la requête 1 avec une carte utilisant la taille et la forme d'un symbole sur une implantation zonale	122
Figure 4.39 : Description d'une carte choroplèthe de couleurs divergentes	123
Figure 4.40 : Représentation de la requête 1 avec une carte choroplèthe de couleurs divergentes	124
Figure 4.41 : Description d'un diagramme comparatif.....	124
Figure 4.42 : Représentation de la requête 1 avec un diagramme comparatif.....	125
Figure 4.43 : Exemple d'une composition pour la requête 2.....	128
Figure 4.44 : Exemple d'une représentation de la requête 2 par la répétition de carte.....	130
Figure 4.45 : Exemple d'une représentation pour la requête 2 par une superposition multidimensionnelle.....	131
Figure 4.46 : Description d'une carte de symboles proportionnels superposés	133
Figure 4.47 : Exemple de représentation pour la requête 3 avec une carte de symboles proportionnels superposés	133

Figure 4.48 : Description d'une carte avec camemberts superposés	134
Figure 4.49 : Exemple de représentation de la requête 3 avec une carte de camemberts superposés	135
Figure 4.50 : Exemple de représentation de la requête 3 à l'aide de camemberts supersposés...	135
Figure 5.1 : Stratégie d'exploration interactive des données. Traduit de [Keim et al, 1996]	142

CHAPITRE 1 - Introduction

1.1 Mise en contexte

L'approche SOLAP (*Spatial On-Line Analytical Processing*) est arrivée parmi les systèmes d'aide à la décision dans le milieu des années 1990. Cette approche a permis de jeter un nouveau regard sur l'analyse de données spatio-temporelles. Celle-ci ouvre davantage la porte aux analyses spatio-temporelles dans les niveaux tactiques et stratégiques; ces analyses étant autrefois d'efficacité limitée et réservées aux spécialistes des systèmes d'information géographique (SIG).

SOLAP est le résultat de la combinaison de fonctions SIG pour les requêtes spatiales aux technologies OLAP (*On-Line Analytical Processing*) pour l'analyse décisionnelle (laquelle est typiquement temporelle et multi-échelles). Il permet aux experts de domaine d'analyser eux-mêmes leurs données. SOLAP est défini comme « *un logiciel de navigation facile et rapide dans les bases de données spatiales qui offre plusieurs niveaux de granularité d'information, plusieurs thèmes, plusieurs époques et plusieurs modes de visualisation synchronisés ou non : cartes, tableaux et diagrammes* » [Bédard, 2004]. Ainsi, en quelques clics de souris, un utilisateur compose sa requête, obtient instantanément une réponse et la visualise sous différentes formes. Celui-ci possède également la liberté de naviguer dans les données avec des opérateurs simples mais nouveaux tels le forage, le remontage et le pivot.

La simplicité et la rapidité du SOLAP pour l'interrogation des bases de données spatiales résident dans la structuration multidimensionnelle des données (Figure 1.1). Cette structure s'appuie sur cinq concepts principaux : les dimensions, les membres, les mesures, les faits et le cube. Les dimensions représentent les axes d'analyse. Une dimension contient des membres organisés en hiérarchie et chaque membre appartient à un niveau particulier de cette hiérarchie. Par exemple, pour une dimension nommée *Temps*, les niveaux de la hiérarchie peuvent être année, mois et jour et les membres respectifs, 2008, décembre 2008 et 17 décembre 2008. Les mesures sont les observations qui nous intéressent, par exemple, le montant total des ventes, la population, la vitesse, etc. Ces mesures se rapportent à un fait unique identifié par le croisement des membres des différentes dimensions de la base de données. Ainsi, un fait présente un état d'un sujet d'analyse en fonction des caractéristiques que l'on choisit, et ceci à l'échelle d'analyse de son choix pour chaque dimension. La Figure 1.1 présente le fait suivant : il y avait 70 millions d'acres en

culture bovine en Amérique du Nord en 1998.¹ Finalement, le cube se définit comme le regroupement des faits ou comme l'ensemble des mesures organisées selon un ensemble de dimensions.² Il faut noter ici que le terme précis est « hypercube » puisqu'une base de données multidimensionnelle peut avoir plus que 3 dimensions. La littérature a cependant consacré le terme « cube de données » et c'est l'expression que nous utilisons dans le présent mémoire.

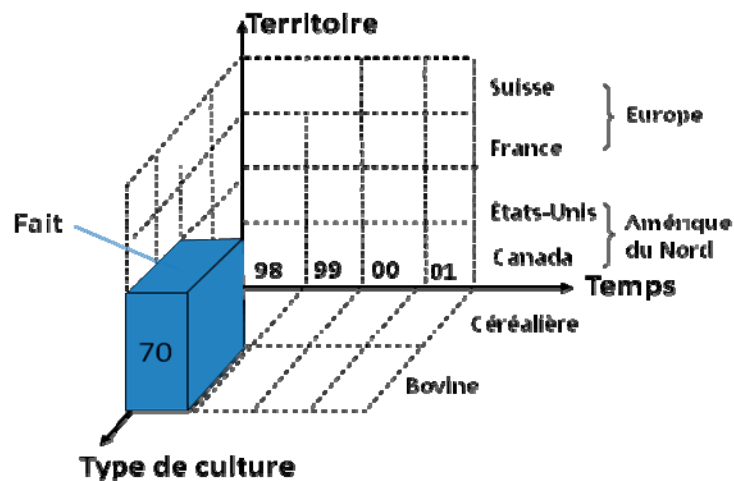


Figure 1.1 : Exemple d'un cube à 3 dimensions de la superficie de territoire en culture

Cette organisation de données simplifie grandement l'exploration des données spatiales. Premièrement, l'utilisateur bénéficie d'une grande flexibilité d'analyse multithématique, multi-échelle et multi-temporelle. Ensuite, la présence d'opérateurs de navigation créés spécialement pour exploiter cette organisation de données facilite les manipulations. La structure interne des données est également beaucoup plus rapide que les structures traditionnelles, i.e. transactionnelles normalisées. L'utilisateur conserve donc son flux de pensée à travers l'exploration des données. Tout ceci permet à l'utilisateur de procéder à des cycles de questions/réponses très rapide, de concentrer ses efforts sur l'objet de sa recherche (le « quoi ») plutôt que sur les moyens pour l'exécuter (le « comment ») et ainsi de mieux supporter la découverte de connaissances spatiales (*geographic knowledge discovery*). En effet, par ces avantages, la structure multidimensionnelle utilisée dans une approche SOLAP répond à deux des trois caractéristiques essentielles pour supporter la découverte de connaissances [Bédard, Proulx et Rivest, 2005; Caron, 1998]. (1) Le temps de réponse est très court, sous la bande cognitive de

¹ Ces données sont purement fictives.

² La définition de concepts est ici très succincte. Pour plus d'informations, vous pouvez consulter les ouvrages suivants : Bédard et Han (2008); Rivest et al (2005) et Bédard, Proulx et Rivest (2005).

Newell de 10 secondes [Newell, 1990 dans Rivest et *al*, 2001]. La structure multidimensionnelle et le pré-calcul de nombreux agrégations et croisements de membres permettent de supporter cette première caractéristique. (2) Ensuite, la requête est facile à composer. L'utilisation d'un langage de requête tel que le SQL implique que l'utilisateur connaisse ce langage et la structure de la base de données. Au niveau décisionnel des organisations, l'utilisateur n'est souvent pas familier avec les concepts de bases de données et encore moins avec les langages de requête. En plus de cacher la complexité du langage, la structure multidimensionnelle des données ne restreint pas l'utilisateur à quelques requêtes prédéfinies. Tous les croisements possibles sont disponibles pour ses analyses par quelques clics de souris. Par l'utilisation de la structure multidimensionnelle, la technologie SOLAP répond bien à ces deux premières caractéristiques tel qu'abondamment discuté dans la littérature SOLAP [Bédard, 1997; Rivest et *al*, 2001, 2005; etc.]. À celles-ci s'ajoute une troisième caractéristique essentielle pour la découverte de connaissance: (3) la visualisation géographique de l'information. Le pouvoir des cartes pour l'analyse spatiale et celui de la visualisation en général n'est plus à faire [Bertin, 1973; Larkin et Simon, 1987; Fortin et Rousseau, 1992; Tufte, 1990; MacEachren, 1995; Caron, 1998; Card, Mackinlay et Shneiderman, 1999; Kraak et Ormeling, 2003; Buzan et Buzan, 2003; Slocum et *al*, 2005; Ware, 2005; National Research Council, 2006]. Elle agit à deux niveaux. Elle permet la présentation des résultats, mais elle offre surtout une autre option d'exploration, l'exploration visuelle des données. Cette troisième caractéristique est peu discutée dans la littérature SOLAP. Sporadiquement, on retrouve la manifestation de quelques désirs en matière de visualisation [Bédard, Proulx et Rivest, 2005; Rivest et *al*, 2005; Bédard 1997; Marchand, 2004; etc.], mais aucune étude ne s'est encore spécifiquement intéressée à ce volet particulièrement important du SOLAP.

Les technologies OLAP ont la particularité de produire des graphiques de toutes sortes en quelques secondes selon la volonté de l'utilisateur pour l'exploration interactive des données. Cependant, celles-ci ne prennent pas en compte les particularités de la visualisation des données géographiques. Bien que l'affichage et la navigation cartographique soient les caractéristiques qui distinguent la technologie SOLAP des technologies OLAP, dans un contexte d'exploration interactive de données, la visualisation géographique va bien au-delà du simple besoin d'affichage. Celle-ci doit supporter la pensée par la manipulation flexible des tableaux, des cartes et des graphiques qui rendent visibles les relations, les structures et les tendances dans les données. Lors de l'exploration des données, « *un graphique n'est jamais une fin en soi. C'est un moment dans le*

processus de la décision » [Bertin, 1977, p16]. Ce sont des images à partir desquelles on déclenche un vaste éventail d'associations et à partir desquelles de multiples questions peuvent être soulevées ou des conclusions, tirées instantanément.

SOLAP propose une approche particulière pour l'analyse des données spatio-temporelles, différente des SIG conventionnels. L'utilisateur n'est pas expert en SIG et encore moins en cartographie; les analyses sont multiples, simples ou complexes, à différents niveaux de granularités et à différentes échelles, puis les représentations sont nombreuses : cartes de toutes sortes, symbologie diversifiée, diagrammes variés. Malgré un effort pour harmoniser la symbologie des différentes méthodes de visualisation, la définition et le choix des symbologies dans les affichages reste un défi important pour l'utilisateur. Ce choix nécessite d'ailleurs une attention particulière dans le cas des méthodes de visualisation géographique. C'est pour mieux répondre aux besoins d'analyse spatio-temporelle décisionnelle que le volet de la visualisation géographique d'un environnement d'exploration interactive des données géodécisionnelles fait l'objet de la présente étude.

Avant d'aller plus loin dans ce mémoire, il convient d'éclaircir deux termes du domaine cartographique qui seront largement utilisés tout au long de ce mémoire. Dans la littérature cartographique, une certaine confusion règne concernant les termes « symbologie » et « sémiologie ». Ces deux termes sont souvent utilisés, à tort ou raison, de façon interchangeable dans la littérature française et anglaise. Dans les dictionnaires populaires français le terme « symbologie » n'existe pas, alors que le terme « semiology » est généralement utilisé uniquement lorsque l'on réfère spécifiquement aux travaux de Bertin. En conciliant les définitions françaises et anglaises dans les différentes littératures, on arrive à ces définitions. La symbologie est la science d'associer un signifiant à un signifié, c'est-à-dire la science des symboles et de leur sémantique attribuée. La sémiologie graphique, quant à elle, est la science des signes et de leur articulation dans la pensée [Le Petit Robert, 2000]. Dans le cadre de ce mémoire, le terme « symbologie » sera utilisé pour désigner l'exercice technique d'association d'un signe ou d'un style à une donnée tel qu'utilisé par l'*Open Geospatial Consortium* [OGC, 2006] (ex. les données de 0 à 4 sont de couleur rouge). Le terme « sémiologie » réfèrera aux connaissances perceptuelles d'un style (ex. un ordre peut s'exprimer par la valeur). On détermine donc la symbologie d'une carte à l'aide des connaissances sémiologiques.

1.2 Problématique

Un nouveau regard est jeté sur l'exploration interactive des données spatio-temporelles par l'avenue de l'approche SOLAP. La navigation facile et intuitive et l'interaction rapide permettent à l'utilisateur d'analyser les données au rythme de sa pensée. Cependant, pour supporter la pensée de l'utilisateur, l'interaction et la navigation facile et rapide dans les données ne sont pas suffisantes pour faire ressortir l'information, les structures et les tendances insoupçonnées et provoquer la découverte des connaissances spatiales. Un troisième volet essentiel est la façon dont on regarde les données. Dans un contexte de découverte de connaissances, les cartes et les diagrammes vont bien au-delà du simple affichage des données, ces outils sont des instruments actifs dans le processus de pensée de l'utilisateur [Keim, Panse et Sips, 2005; MacEachren et Kraak, 2001; Gahegan, 1999] et ils incarnent la porte d'entrée sur les données. Bien que la définition SOLAP comprenne un affichage en cartes, tableaux et diagrammes, le premier problème réside dans le fait qu'**aucune étude n'a cherché à tirer profit des particularités de la visualisation géographique dans un contexte SOLAP**. En plus d'un temps de réponse très court et de l'abstraction de la structure interne des requêtes (langage de requête et structure des données), la visualisation des données représente le troisième élément clé pour l'exploration et l'analyse des données spatiales.

Contrairement aux tableaux de bord, aux outils de *reporting* ou aux applications de cartographie en ligne, SOLAP est un environnement d'exploration interactive des données ou celle-ci se réalise autant par la recherche dans les données que par la recherche dans les représentations tabulaires, graphiques et cartographiques. Les représentations ne sont pas prédéterminées et varient au choix de l'utilisateur. Bien que SOLAP se distingue des SIG par la façon simple et rapide d'accès aux données, le volet visualisation des applications SOLAP développées jusqu'à maintenant et présentées dans la littérature utilise une stratégie principalement basée sur la cartographie traditionnelle. On utilise une symbologie prédéfinie, modifiable par l'utilisateur via un gestionnaire des symbologies [Rivest et al, 2001; Rivest et al, 2005; Kheops Technologie et CRG, 2004]. Comme à l'utilisation des SIG, l'utilisateur est celui qui détermine les caractéristiques de la représentation des données. Il se situe au centre des interactions entre les données, la visualisation et l'analyse (Figure 1.2). En plus de chercher une réponse à sa question, de choisir la façon dont il veut consulter ses données, il doit savoir définir visuellement leur représentation. Cette dernière responsabilité peut se révéler très fastidieuse dans certains cas et peut donner lieu à des

représentations inappropriées pour le phénomène à représenter et ainsi « nuire » à la découverte de tendances et des relations, voire même conduire à de fausses pistes.

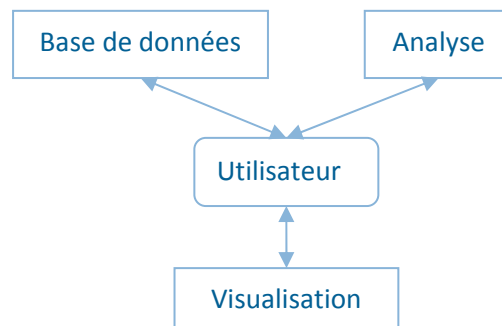


Figure 1.2 : Stratégie d'exploration d'une approche SIG. Inspiré de [Keim et al, 1996]

Cette stratégie répond mal à l'environnement flexible, interactif et multidimensionnel et au support cognitif recherché pour le SOLAP. L'objectif d'exploration poursuivi par l'environnement géodécisionnel est de permettre à l'utilisateur de mettre qu'un minimum d'énergie pour trouver « comment faire » afin obtenir ce qu'il veut, ceci pour mieux se concentrer sur « quoi observer ». Un environnement comme SOLAP nécessite moins d'intervention de l'utilisateur pour transformer les données visuellement tout en permettant plus de flexibilité dans la navigation entre les différents modes de représentations. L'utilisateur doit avoir accès à plusieurs représentations tabulaires, graphiques et cartographiques des données sans avoir à les bâtir lui-même. Le sujet d'analyse et le choix des représentations doivent rester sous la responsabilité de l'utilisateur pour supporter son processus de pensée. Cependant, les détails de construction d'une méthode de visualisation doivent lui être transparents tout comme le langage de requête et la structure des données le sont. Or, la stratégie de visualisation utilisée typiquement par les SIG requiert au contraire que l'utilisateur soit l'intermédiaire entre les données et la visualisation et qu'il s'implique davantage dans la construction graphique des données. Pour cette raison, une deuxième problématique guide cette recherche : **les stratégies usuelles de visualisation des SIG (appliquée actuellement) ne conviennent pas à l'environnement interactif d'exploration de données SOLAP**. À cause de son côté dynamique, ses croisements multiples, ses différents niveaux de granularité et ses utilisateurs cibles, la visualisation géographique de l'information avec SOLAP doit être étudiée pour mieux assister l'utilisateur dans sa découverte de connaissances.

1.3 Solution proposée

L'exploration interactive des données avec SOLAP comportent plusieurs implications à prendre en compte pour l'élaboration d'une solution. Premièrement, lors de l'exploration de données, le but recherché est souvent vague ou mal défini [Keim, Panse et Sips, 2005; Mirel, 1998]. Le but premier est de faire ressortir des hypothèses qui peuvent ensuite être validées (cf. Figure 2.8). Il est donc difficile d'établir préalablement le ou les principaux types idéaux de visualisation. Deuxièmement, l'exploration et l'analyse sont des activités privées qui servent la réflexion d'une ou quelques personnes. Ces activités sont efficacement réalisées par les analystes et les experts du domaine d'application. Les enjeux, les variables non tangibles et les connaissances tacites font partie de l'expertise d'un décideur ou d'un expert de domaine utile à une analyse efficace des données. Cependant, ceux-ci possèdent a priori peu de connaissances de bases de données spatiales et de visualisation. L'affichage des données sert donc typiquement des utilisateurs non experts en visualisation. Ensuite, la technologie SOLAP s'appuie sur des principes de facilité et de rapidité concernant la création des requêtes traditionnellement complexes. À notre avis, ces deux principes doivent également être inclus dans la réflexion concernant la visualisation géographique. Afin d'adhérer pleinement à l'approche SOLAP, il faut étendre le succès actuel de création de requêtes à la création des meilleures symbologies. Dans un tel environnement, l'utilisateur ne devrait plus se demander "comment représenter l'information pour améliorer l'efficacité de mes analyses?" tout comme il ne se demande déjà plus "comment créer ma requête pour effectuer mon analyse?". En d'autres termes, il faut idéalement atteindre le même niveau de transparence pour représenter l'information que ce qui est déjà atteint pour obtenir l'information. De plus, des analyses très complexes peuvent être composées en quelques clics de souris dans un environnement SOLAP. Toutes les données requêtées doivent être représentées pour permettre l'extraction des structures et tendances par l'analyste tout en réduisant la lourdeur de la présentation. En tout temps, un utilisateur doit avoir accès à des cartes à voir plutôt que des cartes à lire afin de percevoir facilement l'organisation de l'information dans les données, développer ses hypothèses et faire avancer sa découverte de connaissances spatiales.

La visualisation géographique des données avec SOLAP représente donc un défi de taille. Les connaissances et les responsabilités de la construction de la visualisation ne peuvent être laissées totalement à l'utilisateur comme c'est souvent le cas aujourd'hui. Celui-ci doit conserver la fluidité de sa pensée sur l'objet de son analyse et non sur la façon de représenter ses données.

Inversement, l'utilisateur est le mieux placé pour savoir quelles méthodes graphiques vont l'aider à découvrir des connaissances. En ce sens, il doit bénéficier d'une certaine liberté de choix entre différentes méthodes de visualisation adéquate tout en évitant de chercher "comment" produire ces choix. Là est le but ultime du SOLAP ainsi que son éternel défi: combiner facilité et performance avec flexibilité.

La définition des représentations ne peut non plus être laissée à l'administrateur des données. Ceci impliquerait de définir préalablement tous les types de représentation désirés pour chacun des croisements possibles de dimensions et de connaître les intentions de l'utilisateur. C'est ce qui existe présentement dans les applications de type tableau de bord spatial où facilité et performance priment clairement sur la flexibilité. Contrairement aux tableaux de bord spatiaux, les solutions SOLAP visent justement à ne pas circonscrire l'utilisateur à un nombre fini de requêtes ou de modes de représentation et à lui donner toute la flexibilité d'analyse et de représentation requises pour la libre exploration interactive de données spatiales.

La solution proposée est donc, dans un premier temps, **(1) d'adapter la stratégie de visualisation géographique à une stratégie d'exploration interactive multi-niveaux, multi-époques et multi-thèmes de données** en plaçant la visualisation au centre de la stratégie d'exploration de données interactive et , dans un deuxième temps, **(2) d'intégrer une intelligence en visualisation géographique à l'intérieur même de la technologie SOLAP** en enrichissant la relation entre les données et leur visualisation (Figure 1.3).

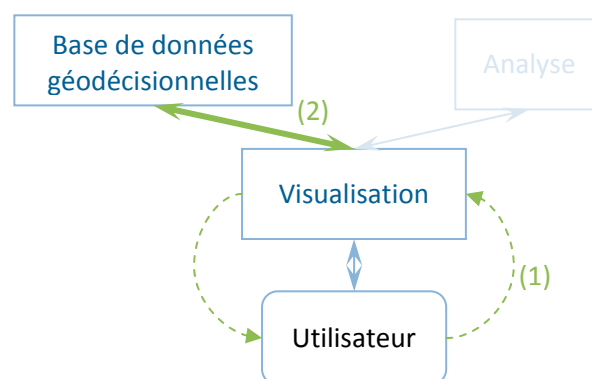


Figure 1.3 : Stratégie d'une approche d'exploration de données. Inspiré de [Keim et al, 1996] (1) premier volet de la solution et (2) deuxième volet de la solution

La solution développée dans ce mémoire s'appuie principalement sur les théories et les concepts du domaine de la géovisualisation³ comme moyen d'améliorer l'exploration interactive des données avec SOLAP.

1.4 Motivations et objectifs

L'idée de départ de cette étude était d'explorer le potentiel de visualisation géographique des données avec la technologie SOLAP. La technologie devenant de plus en plus mature, le troisième élément clé pour la découverte de la connaissance (i.e. la visualisation) se montrait maintenant déficient et il devenait critique d'en établir les bases. De plus, plusieurs avancements dans la communauté scientifique de géovisualisation montraient des progrès intéressants pour la communauté SOLAP. Plusieurs motivations ont guidé les premières étapes de cette recherche.

Dans un contexte d'aide à la décision, les meilleures représentations visuelles doivent être immédiatement accessibles à des utilisateurs non-experts en visualisation de données et permettre de discerner d'un coup d'œil les informations utiles, présentes dans les données. La **première motivation** consistait donc à explorer comment, dans un outil SOLAP, supporter la découverte de la connaissance en facilitant la **perception rapide** des structures, tendances et relations cachées dans les données.

Deuxièmement, peu de travaux ont été réalisés jusqu'à ce jour pour adapter la flexibilité de visualisation des données à la navigation et l'interaction de la technologie SOLAP. Cependant, plusieurs applications sont actuellement en fonction et la commercialisation de la technologie en 2005 a forcé la mise en place d'une stratégie qui s'est révélée lacunaire sur cet aspect. La **deuxième motivation** consistait donc à adapter la **gestion des symbologies** pour la technologie SOLAP.

Finalement, la carte n'est maintenant plus perçue que comme un outil de communication. En devenant interactive elle est maintenant imbriquée dans le processus cognitif de découverte de connaissance et sert à stimuler visuellement la réflexion tout comme les graphiques dans les

³ Pour éviter la confusion entre le domaine scientifique et l'action de visualisation géographique, le terme « géovisualisation » sera utilisé pour désigner le domaine de recherche qui entoure la visualisation géographique et le terme « visualisation géographique » sera utilisé pour désigner le fait de visualiser les données géographiquement à l'aide de cartes.

domaines des statistiques et des sciences. Ainsi, la géovisualisation se développe depuis un peu plus de deux décennies par l'interaction et les différents types de visualisation visant à faciliter la construction de la connaissance. Les concepts et les méthodes en découlant deviennent très intéressants pour les outils d'aide à la décision. **L'intégration des concepts et des méthodes de géovisualisation** avec les technologies SOLAP forme donc **la troisième motivation de cette recherche.**

Ces trois motivations constituent la base de cette recherche exploratoire. L'accomplissement de ces trois orientations implique des exigences beaucoup trop fastidieuses à combler dans le cadre d'un mémoire de maîtrise. Par exemple, faciliter la perception rapide des structures, tendances et relations implique une étude de l'efficacité de différents types de visualisation géographique. Actuellement, les tests d'utilisabilité dans le monde de la visualisation sont à leur début [Fuhrmann et al, 2005] et l'environnement SOLAP contient trop de paramètres pour que cette exigence soit remplie dans le cadre de ce mémoire. La gestion des symbologies ne constitue pas, non plus, l'objet de ce mémoire bien qu'elle y soit intimement liée. Le développement d'un gestionnaire de symbologie nécessite avant tout de solidifier la stratégie de visualisation et les concepts de base de symbologie et de sémiologie graphique pour un environnement SOLAP. Finalement, cette recherche ne prétend pas intégrer tous les concepts utiles de la géovisualisation applicables à la technologie SOLAP. La géovisualisation est un domaine encore jeune qui foisonne de nouvelles idées et de nouvelles théories. L'objectif général de cette recherche se situe donc à l'intersection de ces motivations et vise à répondre partiellement à celles-ci (Figure 1.4). **L'objectif général est donc d'améliorer la façon de visualiser les données géographiquement dans le but de faciliter la découverte de connaissances spatiales avec la technologie SOLAP.**

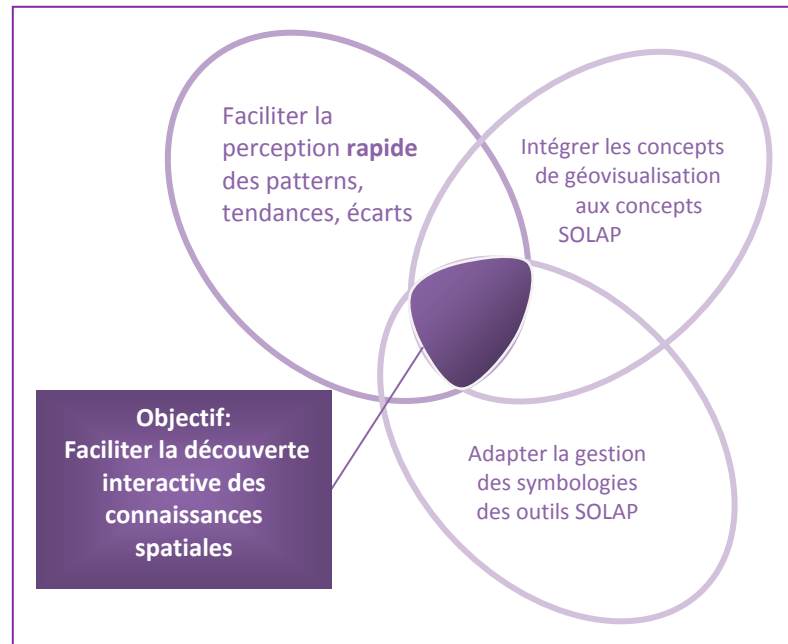


Figure 1.4 : Motivations et objectif général de la recherche

Cet objectif général sera atteint principalement par la concrétisation de trois objectifs spécifiques.

1. **Approfondir les connaissances en géovisualisation pour la technologie SOLAP.** La géovisualisation compte maintenant une communauté scientifique nombreuse et multidisciplinaire. Plusieurs réalisations et méthodes ont été mises en place durant la dernière décennie. La prise de connaissance de la philosophie et des méthodes qui sous-tendent ces réalisations permettra de clarifier la place de la visualisation géographique dans un contexte géodécisionnel et d'extraire les concepts applicables à la technologie SOLAP.
2. **Adapter la stratégie de visualisation géographique du SOLAP à l'exploration interactive et proposer des améliorations suivant cette stratégie.** La stratégie actuelle de visualisation SOLAP étant incomplète, l'approche de la visualisation doit être examinée. La première partie de la solution proposée est d'adapter la stratégie de visualisation géographique avec SOLAP à une stratégie d'exploration interactive de données. Ce changement dans l'organisation des composantes du système fait resurgir certains besoins. En réponse à ces besoins, des améliorations seront soulevées.
3. **Concevoir une méthode de visualisation géographique des données respectant la philosophie d'exploration interactive des données géographiques.** Le changement de

stratégie de visualisation implique d'enrichir le lien entre la base de données et la visualisation. Une méthode de communication entre les deux composantes devra être élaborée.

Cette recherche vise en premier lieu à mieux définir la place et l'impact de la visualisation géographique dans un contexte SOLAP. Il en résultera la mise en place de la nouvelle stratégie et la proposition d'une solution théorique pour la visualisation géographique des données dans un environnement SOLAP.

1.5 Méthodologie

Cette recherche s'inscrit dans une démarche exploratoire où, à chaque itération, les résultats obtenus se rapprochent un peu plus de l'accomplissement des objectifs proposés. Elle se déroule également en collaboration avec l'équipe de professionnels de la chaire industrielle en bases de données géodécisionnelles qui ont l'expérience requise pour réagir aux résultats proposés. Trois projets expérimentaux, dont deux avec des données d'organismes externes, soulèveront questions et réponses tout au long de cette recherche. La méthodologie est présentée à la Figure 1.5 (p. 16). Elle se divise en trois blocs principaux correspondant aux trois sous-objectifs énumérés dans la section précédente.

La première partie de cette recherche consiste à approfondir la connaissance des concepts théoriques en géovisualisation et en technologie décisionnelle. Trois grandes actions ont été entreprises afin de mieux saisir les enjeux liés à la problématique de cette étude. Premièrement, une **revue de littérature** est couverte comprenant deux groupes de thèmes : la géovisualisation avec ses disciplines connexes et les concepts OLAP/SOLAP. La géovisualisation représente un domaine d'étude depuis un peu plus de deux décennies. Une communauté scientifique de plus en plus grandissante et multidisciplinaire s'intéresse à cette spécialité. Celle-ci regroupe plusieurs axes de recherche tels que les sciences cognitives, la perception de l'information, la visualisation scientifique et de l'information, l'interface homme-machine, les statistiques et les SIG [Dyke, MacEachren et Kraak, 2005]. Plusieurs ateliers (GIScience 2006, 2008, in Helinski 2007, Visualization Summit 2007, AGILE 2008), plusieurs publications et une commission (*ICA Commission on Geovisualization*, [Gennady et Dykes, 2008]) alimentent les défis de cette communauté. La participation à une de ces conférences (GeoVisual Analytics Workshop à GIScience 2008) a d'ailleurs permis de constater que la présente recherche vise à utiliser les

concepts développés en géovisualisation pour améliorer la technologie à l'étude et ne prétend pas apporter de nouveaux concepts au domaine de la géovisualisation. L'étude des théories et méthodes qui appuient les développements scientifiques de cette communauté s'étend à plusieurs disciplines. La cartographie thématique, la psychologie de la perception, la sémiologie graphique, la visualisation de l'information, le design d'information et l'interface homme-machine forment la majorité des sujets de lecture concernant la géovisualisation. Un cours supervisé par le département des arts visuels de l'Université Laval m'a également permis d'enrichir ma formation théorique en établissant l'état des lieux en visualisation de l'information alliant la perception et le design. La littérature OLAP et SOLAP forme le deuxième groupe de littérature explorée. Les concepts SOLAP, développés depuis une quinzaine d'années, gagnent en maturité et les publications sont de plus en plus nombreuses. Un cours de notions avancées en bases de données SIG a aussi été complété pour parfaire la connaissance de ces concepts. La bibliographie SOLAP⁴ s'élargit avec le développement et la maturité de cette approche, cependant l'approche ayant été développée par l'équipe du Dr. Bédard, la majorité de la littérature étudiée provient de l'équipe de recherche dans laquelle s'effectue cette étude. Quelques publications externes à l'équipe ont été consultées [Bimonte, 2007; Bimonte et al, 2006; Malinowski et Zimanyi, 2008; Parmanto et Scotch, 2005; Thomsen, 2002; Kimball et Ross, 2002]. Cependant, les concepts rencontrés ne sont pas révélés utiles à cette recherche. La revue de littérature est une étape importante qui permet d'extraire les concepts et les théories qui sont le fondement des communautés scientifiques étudiées.

Parallèlement à la revue de littérature, des **essais technologiques** ont été réalisés. Parmi les technologies étudiées, on retrouve des technologies de cartographie automatique (Carthema, Carte et données et cartographie thématique à l'aide de Google maps et Google Earth), des technologies de géovisualisation proprement dites (GeoDA, GeoVista Studio, GeoViz Toolkit et CommonGIS) et des technologies de visualisation décisionnelle géographique (Tableau Software et JMAP Solap). Les essais technologiques permettent d'extraire les principes, les méthodes et les techniques qui seraient d'intérêt pour l'approche SOLAP. Ils permettent également d'explorer comment la théorie se traduit en pratique.

⁴ Pour davantage de références, le lecteur est invité à consulter le portail Spatial OLAP : <http://www.spatialbi.com/>

Finalement, pour mieux comprendre les enjeux de cette recherche, une **collaboration** à deux projets de la chaire industrielle en bases de données géodécisionnelles a eu lieu. Premièrement, l'examen de problèmes d'échelle et de surpopulation de la scène visuelle a été réalisé sur un projet d'application du SOLAP à la gestion des emprises de lignes à haute tension d'Hydro-Québec. Ensuite, un projet d'application du SOLAP sur des données de transport multimodal des marchandises en collaboration avec le Ministère des Transports a permis d'explorer les alternatives géovisuelles des données du type origine/destination et les problèmes qui en résultent. L'implication au niveau de la visualisation de l'information à ces projets a soulevé beaucoup plus de questions qu'elle n'en a répondues. Ceci a permis d'éclaircir les besoins spécifiques de la technologie SOLAP en matière de visualisation géographique et les problèmes reliés à l'utilisation des stratégies de cartographie traditionnelles ont été ciblés.

La première étape de cette recherche s'est déroulée en itérations successives avec les suivantes. La deuxième étape se découpe en deux actions principales. Premièrement, le premier bloc d'action fait surgir les problèmes, les besoins et les possibilités intéressantes en matière de visualisation géographique pour le SOLAP. Au fil de l'approvisionnement et de la connaissance des théories, des méthodes et des outils de géovisualisation et du SOLAP, **l'analyse des besoins** permet de dégager les caractéristiques et les fonctionnalités nécessaires pour l'optimisation de la visualisation géographique dans un contexte d'exploration de données géodécisionnelles. Ces caractéristiques permettent de **cibler les améliorations potentielles** à apporter en matière de flexibilité d'analyse, de types de visualisation et de manipulations graphiques. Ces deux actions interdépendantes construisent la vision du SOLAP idéal en matière de visualisation géographique. Elles permettent d'évaluer l'impact du passage d'une stratégie de visualisation SIG (Figure 1.2) et une stratégie d'exploration interactive de données géodécisionnelles (Figure 1.3). Ce bloc d'actions répond au second sous-objectif.

Lors des itérations entre les blocs 2 et 3, plusieurs questions ont été soulevées et plusieurs décisions ont été prises. Parmi les améliorations soulevées à la deuxième étape, plusieurs tentatives de conception ont été réalisées pour concrétiser ces améliorations. À la suite des limites et des contraintes rencontrées, il a été choisi de s'attarder à l'amélioration de la relation entre les données et leur visualisation. Pour ce faire, la **conception d'une méthode de visualisation géographique** a été réalisée en trois étapes principales. Premièrement, une taxonomie des concepts de géovisualisation utiles à un environnement SOLAP est élaborée. Ensuite, la définition

des rôles de chacun des acteurs dans la définition des symbologies est établie introduisant deux nouveaux éléments : des métadonnées et une base de connaissances sémiologique. Finalement, la base de connaissances est conçue. La réalisation de ces trois étapes s'est faite en collaboration avec un troisième projet de l'équipe de recherche : l'amélioration de la gestion des symbologies de la technologie JMap SOLAP. Bien que la solution développée par la présente recherche soit théorique, la participation à cette activité a permis d'enrichir la conception de la méthode en l'ancrant à la réalité d'un environnement SOLAP. La méthode de géovisualisation développée est ensuite **validée de façon préliminaire** par des tests de faisabilité et des expérimentations exploratoires. La conception de la méthode de géovisualisation complète le troisième sous-objectif de cette recherche.

À l'issue de ces trois étapes et des nombreuses itérations qui les auront raffinées, une solution théorique de visualisation géographique applicable à un environnement SOLAP est exposée dans le présent mémoire. Une analyse critique de la solution est ensuite réalisée pour conclure cette étude.

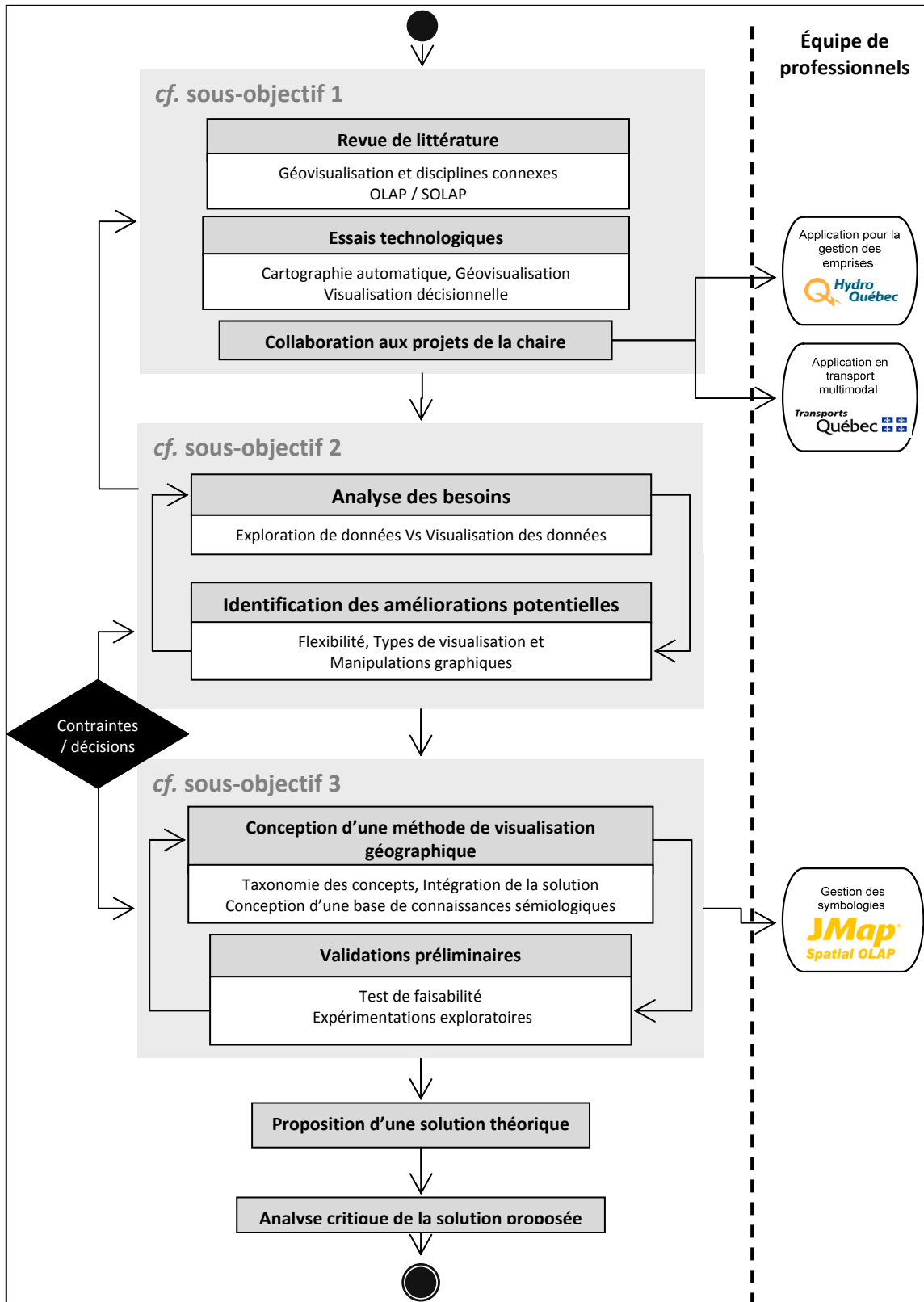


Figure 1.5 : Méthodologie de recherche

1.6 Synopsis

Ce premier chapitre a d'abord exposé le contexte de cette étude. SOLAP est une technologie nouvelle de laquelle un élément clé, la visualisation géographique de l'information, n'avait pas été étudié. La problématique a été explicitée révélant une lacune importante de la stratégie actuellement utilisée. En réponse à ces problématiques, une solution théorique est proposée en deux temps. Premièrement, une approche de visualisation géographique adaptée à un environnement SOLAP est avancée. Ensuite, une méthode de communication entre les données et leur visualisation est conçue. Les objectifs qui mèneront à cet accomplissement ont été énumérés et la méthodologie employée pour les réaliser, précisée.

La suite de ce mémoire se divise en trois chapitres principaux (Figure 1.6). Le chapitre 2 expose les théories et concepts entourant la visualisation de l'information en général. Les explications se rattachent davantage à la relation entre l'utilisateur et la visualisation et la visualisation en tant que tel. Le changement de stratégie fait ressurgir de nouveaux problèmes, des besoins et des possibilités. Le chapitre 3 expose les caractéristiques de l'exploration de données interactive avec SOLAP, les besoins qu'elles engendrent en matière de visualisation géographique et les améliorations proposées pour respecter la philosophie d'exploration. Finalement, le chapitre 4 s'attarde à définir une solution théorique pour améliorer la relation entre les données et leur visualisation. Ce mémoire se conclut par un bref retour sur les objectifs et une longue liste de perspectives de recherche concernant la visualisation géographique des données avec SOLAP.

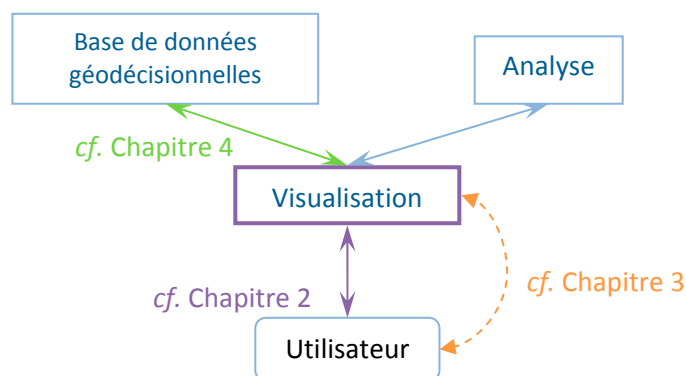


Figure 1.6 : Positionnement des chapitres selon la nouvelle stratégie de visualisation proposée

CHAPITRE 2 - Revue de littérature

Le chapitre précédent a exposé la problématique de la visualisation géographique dans un environnement géodécisionnel, multi-échelle et interactif tel que SOLAP. Avant de pouvoir proposer une solution à ce problème, il est nécessaire d'explorer les travaux et les approches des différents domaines entourant cette problématique. Ce chapitre présente premièrement les concepts de bases en sciences de l'information. Les concepts de données, information et connaissances sont clarifiés pour saisir la portée de chacun de ces concepts. Ensuite, le concept de pensée visuelle est explicité et son rôle pour le traitement de l'information est mis de l'avant. Le lecteur est alors initié à la nouvelle réalité du traitement de l'information, c.-à-d. beaucoup de données et peu d'information acquise de ces données. Deux domaines de recherche récents visant à extraire l'information des données pour mieux supporter la construction de connaissances (c.-à-d. l'exploration de données (EDA) et la découverte de connaissances à partir des données (KDD)) sont décrits à la section suivante. À ces deux domaines de recherche s'ajoute la géovisualisation. Celle-ci s'intéresse particulièrement aux traitements de l'information à référence spatiale par la visualisation géographique des données. La dernière section de ce chapitre présente brièvement l'origine, les fondements théoriques, les méthodes, les techniques et les outils de ce domaine de recherche.

2.1 Les données, l'information et les connaissances

Depuis quelques décennies, nous assistons à une explosion d'information et de moyens de communication. Que ce soit l'internet ou les communications mobiles, nous sommes sans cesse en échange d'information. De plus, les technologies de l'information et de communication forment la nouvelle économie qui supplante les industries de transformation. Nous passons de l'ère industrielle à l'ère d'information [Davis, 2001] où le produit principal n'est plus la matière et le produit transformé, mais l'information et les services. Les données représentent un précieux capital duquel de nombreuses informations et connaissances peuvent être dérivées. Elles sont une monnaie d'échange. Elles sont manipulées, analysées et intégrées pour en comprendre les relations et répondre aux besoins grandissants d'information de notre société. De plus en plus, différentes sphères de la société ont recours aux systèmes d'aide à la décision pour édifier leur savoir et éclairer leur décision. L'aide à la décision repose largement sur les connaissances d'une

partie de la réalité et conséquemment sur l'information qu'on s'efforce d'assembler [Bédard, 1986]. L'information et les connaissances sont au cœur des systèmes d'information. Dans cette section, les concepts de données, information et connaissances et le passage de l'un à l'autre sont définis.

2.1.1 La hiérarchie DIKW

La définition des termes donnée, information et connaissance et des relations qui les unissent sont abondamment discutées dans la littérature par plusieurs auteurs avec différentes subtilités. Une étude, auprès de 45 experts des sciences de l'information, a inventorié 130 définitions des mots *data*, *information* et *knowledge* pouvant se regrouper en cinq modèles différents [Zins, 2007]. Aujourd'hui, la théorie généralement acceptée dans le domaine des sciences de l'information pour définir ces trois concepts et les relations qui les unissent est la hiérarchie DIKW (*Data, Information, Knowledge, Wisdom* : Figure 2.1) [Ackoff, 1989; Bellinger et al, 2004; Rowley, 2007]. Cette théorie pousse la théorie de la triade donnée-information-connaissance un peu plus loin [Machlup, 1983 dans Bédard, 1986] en formulant un concept supplémentaire, la sagesse. La hiérarchie est souvent représentée par une pyramide (Figure 2.1). Celle-ci rappelle que chaque niveau s'appuie et se construit sur le niveau précédent.

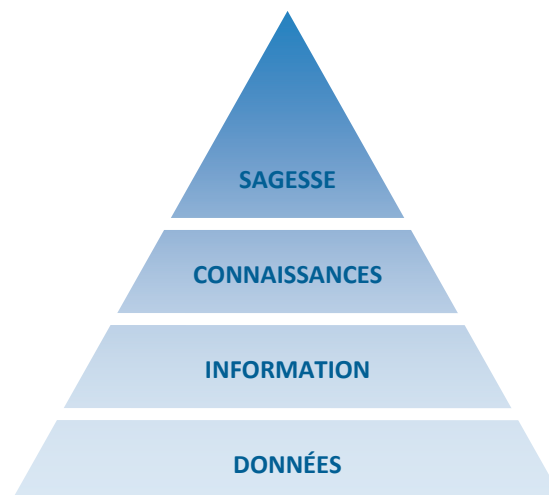


Figure 2.1 : La hiérarchie DIKW. Traduit de [Ackoff, 1989]

Les définitions retrouvées dans la littérature sont souvent fonction des autres concepts de la hiérarchie. Leur description est donc souvent circulaire. Par exemple, on définit l'information comme l'interprétation, l'organisation ou la vue des données ou encore comme la source de connaissances [Spence, 2001; Bonham-Carter, 1994; Bédard, 1986]. Aussi, la structure des

différents concepts est largement acceptée, mais la distance entre ceux-ci et le changement de terme dans le continuum qui lie chacun des concepts est difficile à situer. Dans un tel contexte, il est difficile d'obtenir un consensus pour les définitions intrinsèques à chacun des termes. Dans le cadre de ce mémoire, les définitions succinctes de [Ackoff, 1989] et [Bellinger et al, 2004] sont adoptées.

- Une **donnée** est une information brute, sans signification autre que son existence. Elle est le produit de l'observation de la réalité.
- L'**information** contient des descriptions et un contexte qui nous permettent de répondre aux questions qui, quoi, quand, où et combien. Ce sont des données avec une signification attribuée par l'intermédiaire des relations ou d'une organisation.
- La **connaissance** est une collection d'informations utiles qui répond à la question comment. Elle peut être acquise d'un autre individu, par des instructions ou par l'expérience. La connaissance n'est pas seulement le résultat de la pensée, mais celle-ci est alimentée par des intrants [Bédard, 1986], c'est-à-dire les informations.
- La **sagesse** est le point culminant de la pyramide. Elle implique le jugement, la morale et les codes éthiques. Elle se construit à partir des connaissances acquises. Elle permet de différencier le bien du mal, le bon choix du mauvais.

En montant les niveaux de la pyramide, la contribution humaine est de plus en plus importante. Les facultés cognitives de réfléchir et de comprendre sont mises à profit. L'ascension de la pyramide implique une transition à travers les concepts de données, information, connaissances et sagesse. Selon [Bellinger et al, 2004] et [Clark, 2004], c'est la compréhension qui supporte la transition entre chaque concept (Figure 2.2). [Bellinger et al, 2004] suggèrent que le passage entre la donnée et l'information implique de comprendre les relations. Le passage entre l'information et les connaissances implique de comprendre les structures. Finalement, le passage entre les connaissances et la sagesse implique de comprendre les principes. Cette progression est aussi décrite par [Clark, 2004] qui la nomme le continuum de la compréhension (*continuum of understanding*).

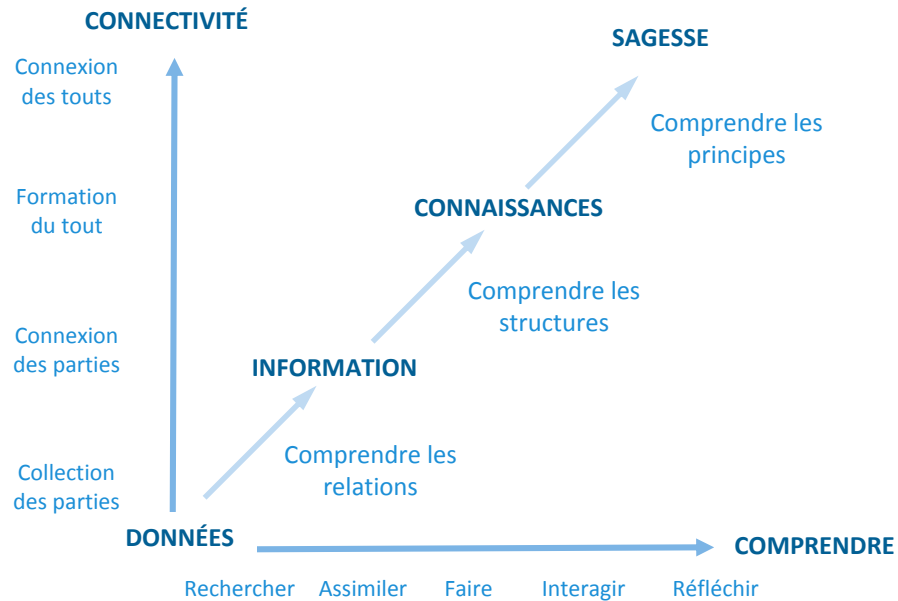


Figure 2.2 : La hiérarchie DIKW ou le continuum de la compréhension. Traduit de [Bellinger et al, 2004] et [Clark, 2004]

Cette théorie est à la base des technologies de l'information et du développement des sciences de l'information. Alors qu'il y a quelques années, les développements en sciences de l'information s'attardaient à la saisie et la production de données; aujourd'hui, l'attention est dirigée vers les systèmes qui stockent et génèrent de l'information (SIGéographique, SIGestion, SIManagement, etc.). Plus récemment, ces systèmes ont évolué vers les systèmes visant la production de connaissances (cf. section 2.3). [Rowley, 2007] a proposé une pyramide qui associe les types de systèmes à la hiérarchie DIKW (Figure 2.3). Avec l'avancement de la technologie, l'objectif des systèmes d'information remonte la hiérarchie DIKW. SOLAP est une approche se situant quelque part entre les systèmes d'information et les systèmes d'aide à la décision. Il permet d'accéder à l'information directement sans intermédiaires et la présentation graphique aide à faire ressortir les relations et tendances dans les données pour éclairer les décisions.

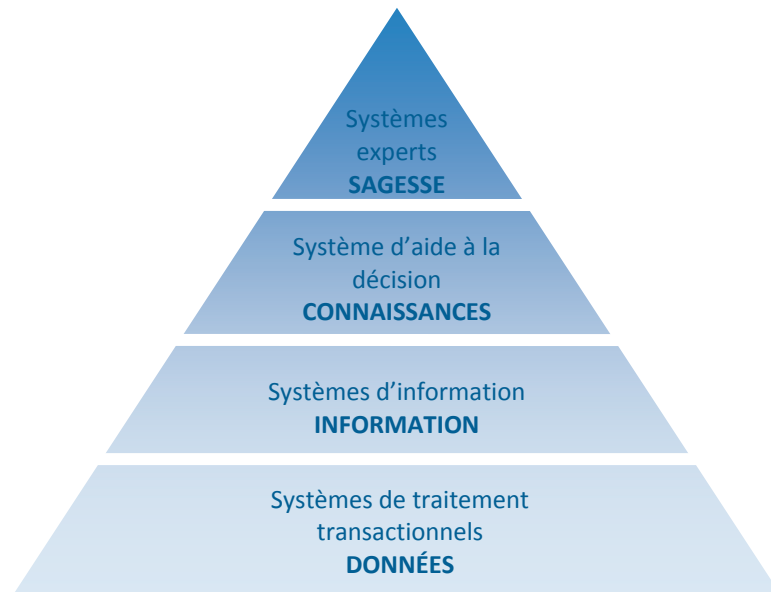


Figure 2.3 : Association des différents types de systèmes d'information et des concepts DIKW.
Traduit de [Rowley, 2007]

Les concepts présentés de donnée, information et connaissances spatiales sont rarement abordés selon la hiérarchie DIKW dans la littérature cartographique, géographique ou géomatique. Leur particularité spatiale leur vaut un intérêt particulier. Il est intéressant de constater par contre que cette hiérarchie et le changement d'accent sur les systèmes s'opèrent également dans le domaine de la cartographie et de la géovisualisation (voir section 2.4.1).

2.1.2 Des données à l'information aux connaissances

Les données sont le produit de l'observation de la réalité. Elles sont un chiffre ou une qualité que l'on extrait de son contexte. Le passage des données à l'information se produit à l'extraction des données par les systèmes d'information. À l'intégration de celles-ci dans les bases de données, les données sont structurées, mises en contexte et leurs relations, modélisées. Les bases de données créent donc l'information grâce à l'organisation des données. Plusieurs systèmes et technologies ont été développés pour rendre l'information accessible. Parmi ces technologies, on peut citer les modèles relationnels et les modèles multidimensionnels. Les modèles relationnels de données définissent explicitement les relations dans la base de données. Les informations peuvent donc être facilement reconstituées à partir de ces tables de relations. Les modèles multidimensionnels, quant à eux, stockent directement les informations de contexte et les données (mesures) dans une table de fait. L'information est donc directement accessible.

Une fois que l'information est accessible, une collection appropriée d'informations amène à la construction des connaissances. L'efficacité des systèmes d'aide à la décision dépend énormément de la facilité avec laquelle on acquiert ces connaissances. Le regroupement d'informations pertinentes est alors essentiel pour permettre de reconnaître une partie de la réalité [Bédard, 1986]. Les systèmes d'aide à la décision ont pour objectif de faciliter ces regroupements pour permettre une découverte des connaissances efficace. La façon et les moyens pour faciliter ces regroupements et édifier les connaissances sont discutables. Le *data mining*, l'exploration de données, la visualisation scientifique et de l'information sont tous des moyens explorés pour faire ressortir les connaissances à partir des données. Ces moyens seront discutés à la section 2.3. La section 2.2 présente la pensée visuelle. Ce mode de raisonnement par l'image présente plusieurs caractéristiques intéressantes pour un passage rapide de l'information aux connaissances.

2.2 La pensée visuelle (*visual thinking*)⁵

Il est reconnu que l'humain n'utilise qu'une faible partie des capacités du cerveau. La pensée visuelle est un de ces modes de raisonnement sous-utilisé et pourtant si puissant. Elle se définit comme un raisonnement par l'image s'appuyant sur la perception visuelle [Wikipédia, 2008]. Ce mode de pensée utilise la partie droite du cerveau pour raisonner de façon créative et imaginative. Elle traite l'information de façon parallèle permettant l'extraction de nombreuses informations rapidement [MacEachren, 1995, Ware, 2008]. La vision globale est employée pour déclencher un vaste éventail d'associations. La partie droite du cerveau est aussi le siège des émotions, de l'art et de l'intuition.

La pensée visuelle inclut également la spatialisation de l'information ou la pensée spatiale. La pensée spatiale utilise les représentations pour aider la remémoration, la compréhension, le raisonnement et la communication des propriétés et des relations entre les objets représentés dans l'espace, qu'ils soient ou non des objets intrinsèquement spatiaux [National Research Council, 2006]. Elle permet de reconnaître l'environnement comme un tout, une globalité et tire les relations et les associations entre les positions et les caractéristiques des différents objets d'une scène. Plus spécifiquement, la pensée spatiale met l'accent sur la compréhension des

⁵ Pour plus d'informations sur la pensée visuelle et sur son fonctionnement, la lecture de [WARE, 2008] est fortement recommandée.

structures et des fonctions reconnues et découvertes dans une scène visuelle. Les structures se définissent comme l'organisation des parties entre elles; elles sollicitent la dimension spatiale. Les fonctions caractérisent l'impact des actions d'une partie sur un autre objet; celles-ci sollicitent la dimension spatio-temporelle [National Research Council, 2006]. La carte est la représentation graphique qui sert la pensée spatiale depuis des siècles. Qu'on parle de carte topographique, de carte mentale, de carte thématique, la carte est la spatialisation de l'information.

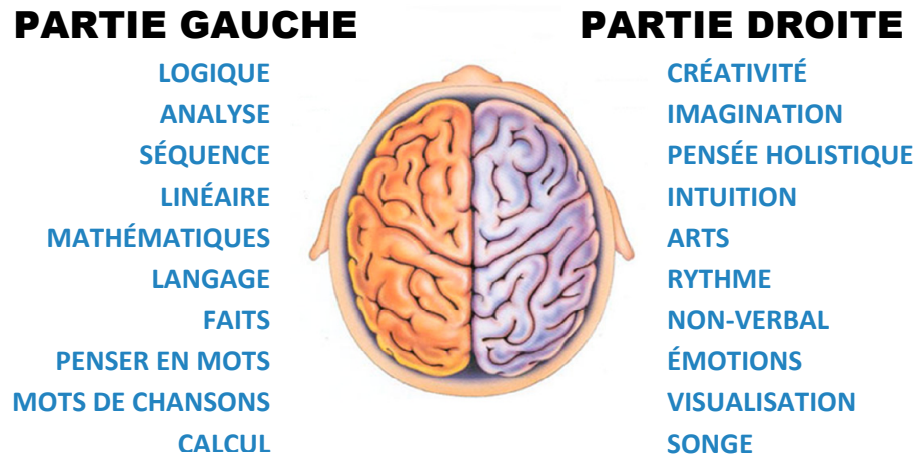


Figure 2.4 : Les hémisphères cérébraux et leurs principales fonctions. Inspiré de [UCNAS Canada, 2007]

De l'autre côté, la pensée verbale est celle avec laquelle la plupart d'entre nous sont familiers. Par contraste avec la pensée visuelle, la pensée verbale utilise la partie gauche du cerveau pour traiter l'information en série et raisonner par objet, de façon logique et linéaire. Elle utilise la connaissance du langage et les facultés mathématiques. À l'âge adulte, on a habituellement appris à développer ce mode de pensée [National Research Council, 2006]. On communique par le langage, on analyse avec des chiffres et on raisonne de façon logique. Cependant, le raisonnement logique montre parfois ses limites. Lorsque le problème devient très complexe et que le nombre de données ou d'objet à analyser est important, le raisonnement verbal n'arrive plus à gérer la quantité d'informations et devient rapidement surchargé [Miller, 1956; Wurman, 1989]. Plusieurs exemples du passé montrent une résolution de problème plus efficace et plus rapide par la pensée visuelle (ex. Dr. Snow et l'épidémie de choléra, la marche de Napoléon) [Mackinlay, 2008]. Toute la puissance de la pensée visuelle réside dans le fait que toutes les informations saisies par notre système visuel sont traitées en parallèle. La collaboration active des hémisphères cérébraux droit et gauche s'occupe ensuite de trier et de raisonner cette information. Les deux hémisphères

cérébraux, bien que distincts, sont en constante communication favorisant une collaboration serrée entre les deux modes de pensée⁶. L'emploi de notre pensée visuelle est donc une voix exploratoire intéressante pour surpasser notre capacité d'analyse actuelle des données.

Cette section vise à comprendre le processus de la pensée visuelle. Pour ce faire, la section 2.2.1 explique le fonctionnement perceptuel visuel. Les moyens pour l'exploiter sont exposés dans la section 2.2.2. Puis, les avantages et limites de ce mode de raisonnement sont décrits à la section 2.2.3.

2.2.1 La vision et le fonctionnement perceptuel

La pensée visuelle s'appuie sur les capacités de la vision. La vision est notre principal mode d'acquisition d'information provenant de l'extérieur [Davis, 2001; Card, Mackinlay et Shneiderman, 1999]. Nos yeux nous permettent d'apprécier le monde qui nous entoure et d'enregistrer une quantité phénoménale d'information. Notre système visuel est très puissant et la perception nous permet de détecter l'information utile. En effet, il nous permet de reconnaître un individu dans une foule ou de percevoir une maille irrégulière dans un tricot. Pour mieux comprendre la puissance de notre système visuel, il convient d'expliquer davantage le processus de la perception visuelle.

La perception se caractérise par deux processus : des traitements dirigés par les données (*bottom-up processes*) et des traitements dirigés par les concepts (*top-down processes*) (Figure 2.5). La perception est constamment influencée par ces deux processus et l'un influence inmanquablement le comportement de l'autre. [Ware, 2008; MacEachren, 1995].

La première étape du processus « *bottom-up* » vise à extraire les informations de la scène visuelle. Il se nomme la vision pré-attentive ou « *early-vision* » [Ware, 2000]. L'extraction des caractéristiques visuelles se produit de façon parallèle. Simultanément, les formes et les orientations, les contrastes rouge-vert et jaune-bleu, les mouvements et la profondeur sont extraits. Ce qui sera détecté dépend du contexte [Ware, 2000]. Les contrastes, les mouvements et la profondeur sont le résultat des différences avec leur voisin dans la scène visuelle. Ceci est très

⁶ Il est à noter que la pensée verbale et la pensée visuelle ne sont pas les seuls modes de pensée que l'humain utilise. Ces deux modes de pensée sont par contre souvent présentés en opposition pour mettre en évidence leurs caractéristiques propres et le pouvoir de leur alliance.

important puisque la couleur rouge ne garantit pas d'attirer l'œil si elle se situe dans un contexte similaire (ex. rose ou orangé). Ensuite, dans une étape intermédiaire, les caractéristiques extraites sont utilisées pour organiser l'image de façon à reconnaître les contours, les formes, les structures qui s'y trouvent. C'est à cette étape qu'interviennent les principes de la Gestalt (globalité : le tout est plus grand que la somme de ses parties) depuis longtemps étudiés en psychologie de la perception [Fortin et Rousseau, 1992]. Les objets se forment et sont ensuite maintenus dans la mémoire visuelle active. Les objets formés interpellent les concepts autrefois enregistrés dans la mémoire long terme qui leur sont pertinents. Très peu d'éléments peuvent être maintenus en mémoire visuelle active simultanément, soit un maximum de trois objets à la fois [Ware, 2008]. Cependant, le contenu de la mémoire visuelle est très dynamique et change constamment.

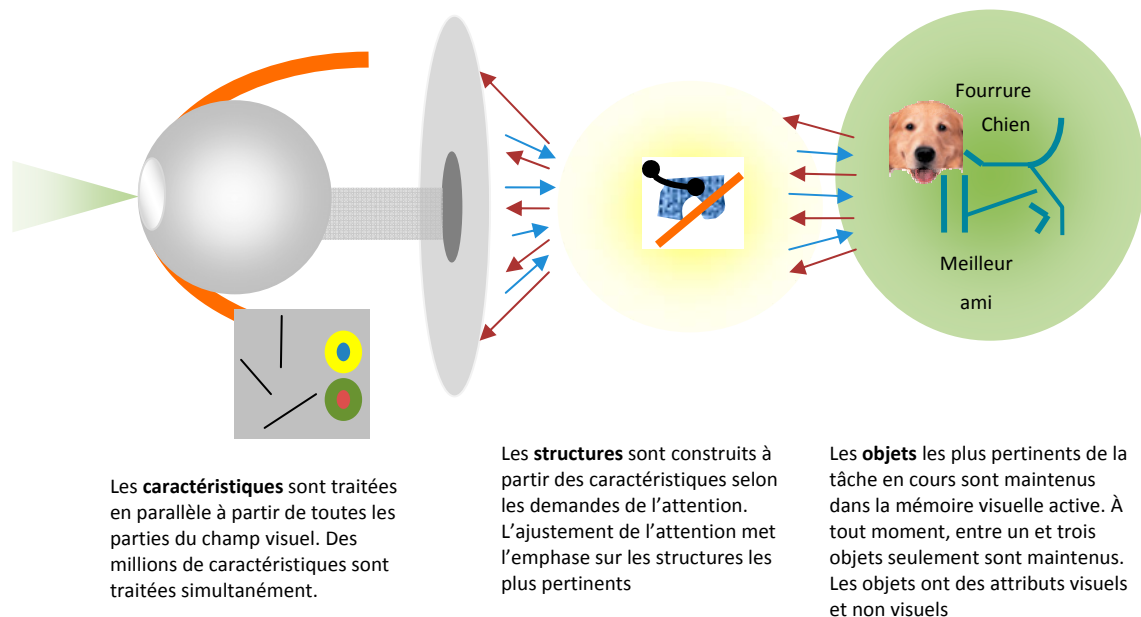


Figure 2.5 : Fonctionnement perceptuel. Traduit de [Ware, 2008, p.9]

Le second processus de la perception est le traitement attentionnel ou dirigé par les concepts (*top-down*). Celui-ci trouve son origine dans l'attention sélective. L'attention cherche à tirer de la scène les informations utiles à un certain but. Il influence la perception en liant et reliant les informations visuelles de façon à tirer une information. Il influence aussi la vision pré-attentive en

introduisant un biais sur une caractéristique particulière recherchée. C'est ainsi qu'on « arrive à lire des mots même si l'ordre des lettres est mélangé en autant que la première et la dernière lettre soient à la bonne place⁷ » [Fortin et Rousseau, 1992]. Les lettres sont décodées et les mots que nous avons appris correspondants à ces lettres sont apportés en mémoire active. Les mots sont alors décodés selon le sens et les lettres lues.

Le fonctionnement perceptuel est un processus à deux sens : on obtient des caractéristiques du monde extérieur et on modifie, on met en valeur, on interprète ces informations selon l'attention du moment. L'incroyable puissance de la pensée visuelle réside dans l'échange entre l'information extérieure et l'information intérieure (*external and internal cognition*) (Figure 2.6) [Chabris et Kosslyn, 2005, Scaife and Rogers, 1996]. On utilise les données de l'extérieur pour alimenter notre raisonnement intérieur et notre raisonnement peut être transféré sur les supports visuels. Pour l'analyse des données, l'intervention humaine est essentielle pour pouvoir bénéficier de cet échange.

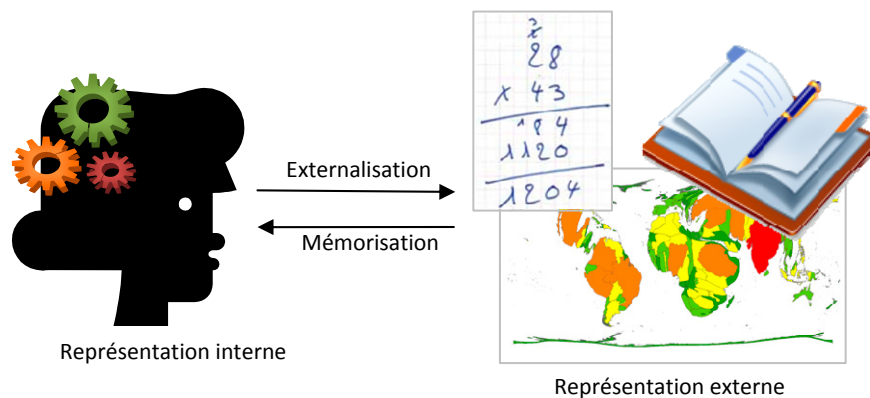


Figure 2.6 : Cognition interne et externe

Jusqu'à maintenant, on arrive à imiter, jusqu'à un certain point, la pensée verbale et le traitement sériel des données à l'aide d'algorithmes et de la puissance de calcul des ordinateurs beaucoup plus forte que le cerveau humain. La pensée visuelle par contre ne trouve pas d'égal. D'un autre côté, la mémoire visuelle est très limitée. Pour exploiter ce mode de pensée, des supports visuels doivent être fournis. C'est d'ailleurs la création de ceux-ci qui aide l'idéation.

⁷ Le mélange des lettres est volontaire. Ceci est pour démontrer que l'influence du processus attentionnel sur notre perception.

The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities. How have we increased memory, thought, and reasoning? By the inventions of external aids: It is things that make us smart. Donald Norman [Tricot, 2008]

La création des supports visuels constitue le domaine de la visualisation de l'information.

2.2.2 La visualisation de l'information

[Card, MacKinlay et Shneiderman, 1999] a défini la visualisation de l'information comme suit : « *The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.* » Visualiser l'information consiste donc à utiliser une représentation graphique interactive de données ou de concepts pour amplifier notre capacité à raisonner [Ware, 2000; Rhyne, 2003]. Elle cherche à optimiser les représentations graphiques pour collaborer avec nos habiletés cognitives afin de faciliter la découverte d'information. Dans cette définition, deux concepts principaux supportent la visualisation. Premièrement (1), cette discipline s'intéresse au processus externe à l'humain qui vise à transformer les données en une forme compréhensible visuellement. Elle crée les supports visuels à la pensée. Dans un but d'améliorer le raisonnement, ces outils doivent être vus comme des extensions à notre habileté cognitive (cognition externe [Tricot, 2008]). « *Le monde est sa propre mémoire* » [Kevin O'Regan dans Ware, 2008, p.2; traduction libre]. En s'appuyant sur cette citation, on peut comprendre l'importance de fournir les représentations visuelles pour prolonger la mémoire visuelle. En effet, la vision est sans cesse en quête d'information. Cette information est tirée directement des scènes qui lui sont présentées [Ware, 2005]. On ne connaît rien du monde qui nous entoure si ce n'est de chaque petit fragment perçu instantanément à chaque requête visuelle⁸. Les représentations visuelles fournissent directement les relations, les structures et les tendances. Il suffit alors de les percevoir pour les découvrir. Deuxièmement (2), les supports visuels, en plus de prolonger la mémoire, doivent être interactifs. La pensée visuelle est bidirectionnelle. Elle tire de l'information d'une scène visuelle, mais elle en demande également (*demand-driven*). Dans un contexte de recherche d'informations, la vision est un processus actif utilisé autant pour recevoir l'information que pour questionner ce

⁸ Le concept de requête visuelle ne sera pas élaboré dans le cadre de ce mémoire. Pour plus d'informations, il est fortement recommandé de se référer à [Ware, 2005].

qui nous apparaît et chercher l'information utile. Pour permettre d'exploiter le côté actif de la vision, l'interactivité avec les données est primordiale (Figure 2.7). L'utilisateur doit manipuler au gré de sa pensée les interfaces pour augmenter sa capacité de découverte d'information dans les données.

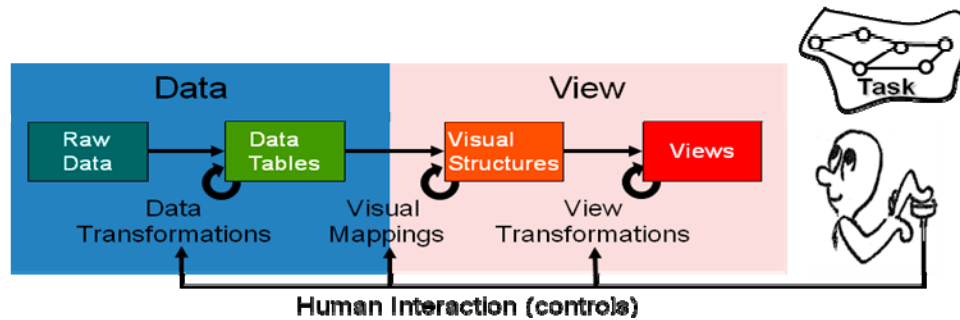


Figure 2.7 : Modèle de la visualisation. Source [Card, MacKinlay et Shneiderman, 1999]

Organiser l'information de façon visuelle présente plusieurs avantages [Ware, 2000; Card, Mackinlay et Shneiderman, 1999; Larkin et Simon, 1987]. (1) La visualisation permet de comprendre une grande quantité de données simultanément et rapidement. Elle augmente la mémoire et les ressources de traitement. (2) Elle permet de percevoir des relations entre les données qui n'auraient pu être anticipées profitant de la capacité visuelle pour la détection des structures. (3) Les inférences perceptuelles sont exécutées plus facilement. Par exemple, les problèmes attribuables à la qualité ou aux données mêmes deviennent rapidement visibles. (4) La visualisation permet de ramener les phénomènes de très grandes ou très petites échelles à une échelle compréhensible. (5) Elle facilite la formation ou la génération d'hypothèses. (6) Finalement, elle réduit la recherche d'informations en regroupant l'information pertinente ensemble. Elle accélère par le fait même notre capacité à trouver un sens aux données.

La visualisation de l'information est un moyen d'étendre nos capacités cognitives au-delà des limites imposées par la mémoire. Cette science s'appuie sur les connaissances en psychologie cognitive pour exploiter les capacités extraordinaires du système visuel.

2.2.3 La perception visuelle : pouvoir et limites

L'utilisation de la perception visuelle pour supporter la pensée possède évidemment des avantages et des limites. Parmi les avantages, on peut compter (1) un meilleur support à l'analyse. [Le Robert, 2002] définit l'analyse comme une « opération intellectuelle consistant à décomposer un tout en ses éléments constituants et d'en établir les relations ». L'opération de décomposition

d'un tout est élaborée de façon logique et très linéaire. L'édification des relations par contre nécessite de percevoir le tout. C'est à ce point que le raisonnement par l'image prend toute son importance. Le traitement parallèle des données permet d'extraire une grande quantité d'information simultanément et les relations pertinentes sont maintenues pour un court laps de temps en mémoire active pour permettre les associations. Des structures qui peuvent être perçues en un seul instant avec une image peuvent prendre des heures ou même ne jamais être trouvées avec un texte ou des chiffres. Ensuite (2), l'imagination et la créativité sont éveillées par le traitement de l'image. Ces caractéristiques propres à l'humain sont sollicitées et l'émergence d'associations nouvelles et inédites est davantage possible. Newton n'a pas découvert la gravité par une déduction logique. C'est d'ailleurs dans des situations où l'on est forcé de penser non traditionnellement que les découvertes les plus créatives ont eu lieu [Finke et al, 1992 dans Kraak, 2003]. C'est précisément ce côté unique à l'humain que l'on désire exploiter par l'utilisation de la pensée visuelle pour l'analyse des données. Le raisonnement logique et l'analyse usuelle déductive peuvent devenir très ou trop lourds pour l'analyse de certains phénomènes et ce mode de raisonnement ne permet pas d'utiliser la flexibilité et la créativité de l'esprit pour découvrir de nouvelles connaissances.

Les connaissances dans le domaine de la psychologie de la perception visent à faciliter la pensée visuelle, favorisant la vision active. Dans un souci d'ergonomie cognitive, les designs d'information tentent d'exploiter le mieux possible les capacités de la perception et la pensée visuelle pour explorer, analyser et transmettre l'information. Cependant, la pensée visuelle possède aussi des limites. (1) La perception est fortement biaisée par l'attention sélective et les tâches en cours. Le processus *top-down* de la pensée visuelle est interne à l'utilisateur. Il est impensable de contrôler cet aspect de la vision active. Deuxièmement (2), la mémoire visuelle active est très limitée. Cette limite est par contre repoussée par le design des outils de visualisation interactifs. Ceux-ci doivent mettre en valeur les relations, les structures dans les données tout en laissant libre cours à l'imagination et la créativité pour la découverte de nouvelles connaissances. Le biais doit par contre être celui de l'utilisateur et non celui du designer. Les manipulations visuelles permettent de répondre à la nature dynamique de la vision active et d'exploiter au maximum la flexibilité de l'esprit.

Le pouvoir de la pensée visuelle n'est plus à discuter. « Une image vaut mille mots » est un dicton connu qui révèle bien la force de la pensée visuelle par rapport à la pensée verbale pour la compréhension des structures et des relations. La vision active et l'interactivité des systèmes peuvent augmenter davantage ce pouvoir. Shneiderman l'exprime bien par ces mots : « *If a picture is worth a thousand words, then an interface is worth a thousand pictures.* » [Shneiderman, 2001 dans Plaisant, 2005]

2.3 Découvrir l'inconnu

Actuellement, nous possédons plus d'informations potentielles que nous pouvons en découvrir. En effet, la quantité de données dont nous disposons aujourd'hui est phénoménale. Des chercheurs de l'Université de Berkeley estiment que chaque année, environ deux exaoctets (10^{12} Mo) de données sont générés, dont la majeure partie est numérique, avec un taux de croissance annuel de 50 % [Keim, Panse et Sips, 2005]. La plupart de ces données sont acquises automatiquement par des récepteurs ou des systèmes de suivi. Nous n'avons qu'à penser que chaque transaction ou chaque appel téléphonique que nous complétons sont des données supplémentaires dans une base de données pour s'apercevoir de l'incroyable flux de données se déroulant chaque seconde. Les données sont non seulement plus importantes en nombre d'enregistrements, mais aussi en nombre d'attributs pour un seul enregistrement. Ceci crée donc d'énormes bases de données cachant une grande quantité d'informations.

La majorité de ces données ont une référence spatiale, par exemple, une adresse, un indicatif régional ou des coordonnées géographiques. En effet, il est reconnu qu'environ 80 % de l'information stockée dans les bases de données ont une composante spatiale [Franklin, 1992 dans Rivest et al, 2001]. Ce chiffre pourrait être revu à la hausse compte tenu de la démocratisation des outils de localisation spatiale (GPS, Google Earth, etc.) et de la préoccupation courante de la dimension géographique pour la planification et l'aide à la décision.

Les données spatiales et non spatiales sont omniprésentes et, par le fait même, l'information cachée est disponible et les connaissances accessibles. Les différents domaines d'études, qui devaient autrefois composer avec une quantité de données limitée, possèdent maintenant un environnement riche en données de toutes sortes. Ces environnements sont aujourd'hui tellement riches que de larges volumes de données sont négligés et l'information potentielle

qu'elles contiennent reste inexplorée [MacEachren et al, 1999]. Les données deviennent alors inutiles et les bases de données, des dépotoirs de données [Keim, Panse et Sips, 2005].

Plusieurs disciplines dans différents domaines sont nées de cette problématique. La découverte de connaissances dans les données en informatique (section 2.3.1) et l'exploration des données en statistique (section 2.3.2) en sont des exemples. Ces domaines cherchent à découvrir l'inconnu à partir des données et à faciliter la production de connaissances insoupçonnées. Dans cette section, ces deux approches sont expliquées.

2.3.1 La découverte de connaissances dans les données (KDD)

Les nombreuses données que nous acquérons visent à parfaire notre connaissance du monde. Traditionnellement, lorsqu'on vise à résoudre un problème par la méthode scientifique, on pose une hypothèse, on acquiert les données, on les analyse et finalement on en tire des conclusions. Actuellement, de grands volumes de données sont disponibles comportant l'information potentielle conduisant à de nouvelles connaissances. Cependant, de larges volumes de données sont négligés et l'information potentielle qu'elles contiennent reste inexplorée [MacEachren et al, 1999]. Le domaine de recherche KDD (*knowledge discovery in databases*) est une réponse à cette problématique. La découverte de connaissance à partir des données se définit comme « *the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data* » [Fayyad, Piatetsky-Shapiro et Smyth, 1996]. La KDD vise à extraire des connaissances à partir de très grandes bases de données faiblement comprises [Gahegan et al, 2001]. La découverte de connaissances se déroule au tout début de la méthode scientifique ou de la construction de la connaissance (Figure 2.8), avant même qu'une hypothèse ne soit posée.

Le processus de construction des connaissances selon la méthode scientifique a été décrit en plusieurs étapes par [Gahegan et al, 2001]. Celui-ci le décrit en quatre étapes itératives (Figure 2.8) passant de l'abduction à la déduction par l'induction. La première étape est l'exploration, la fouille de données. À ce point, les données sont peu ou pas connues. On cherche à raffiner notre connaissance des données jusqu'à ce que ces dernières s'organisent et aient un sens qui permet de pouvoir émettre des hypothèses. Lorsqu'une hypothèse guide notre recherche, on entre dans un raisonnement inductif duquel on cherche à découler des règles et des

comportements. On analyse et on traite ensuite de façon déductive les données pour valider les informations trouvées pour ensuite les évaluer et les présenter.

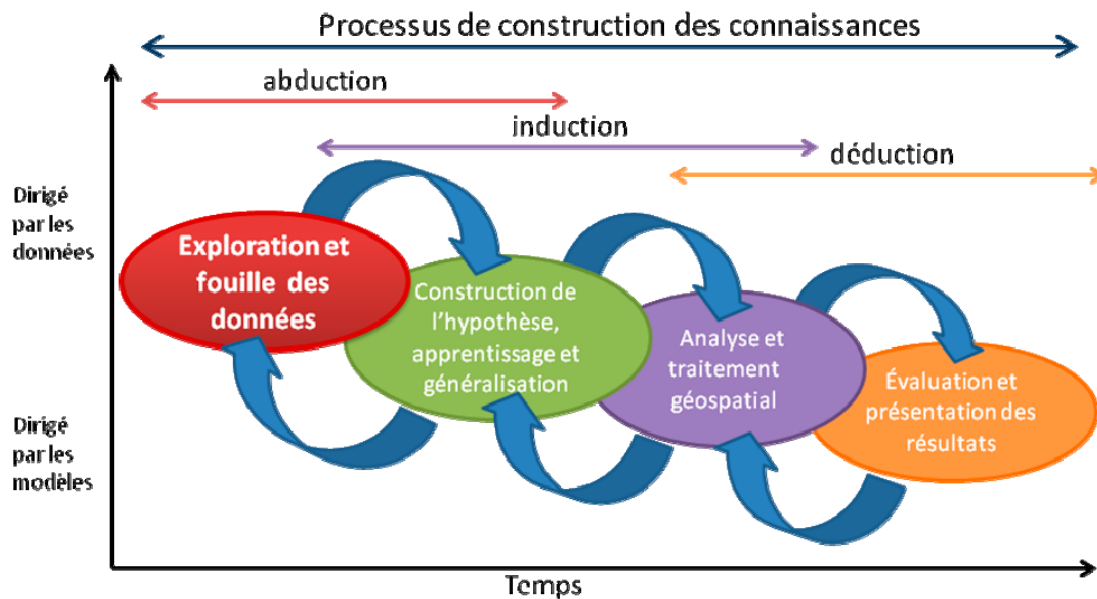


Figure 2.8 : Différentes phases dans la méthode scientifique et les modes de raisonnement qui s'y appliquent. Traduit de [Gahegan et al, 2001]

Le processus de construction de la connaissance est hautement itératif. On passe constamment de la donnée à l'information à la connaissance et l'inverse. Le processus abductif qui guide les premières étapes du processus de construction de la connaissance est primordial pour la découverte de nouvelles connaissances. Le manque de connaissances des données permet de procéder à une recherche aléatoire, non dirigée d'informations permettant l'extraction de tendances et de relations inédites ou insoupçonnées. L'imagination, la créativité et la flexibilité de l'esprit humain sont alors mises à profit.

La KDD est souvent associée à la fouille de données automatique (*data mining*). Celle-ci y joue un rôle central [MacEachren et al, 1999]. La fouille de données automatique est un processus automatique d'identification ou de découverte de structures utiles dans les données [Fayyad et Grinstein, 2002]. Il se base sur l'application d'algorithmes plus ou moins complexes pour révéler les structures et les relations présentes dans les données, c.-à-d. l'information cachée [Miller et Han, 2001]. La fouille de données automatique spatiale (*geographic data mining*) se spécialise dans l'extraction des propriétés spatiales des objets tout comme les relations spatio-temporelles entre les objets [Miller et Han, 2001]. Différentes classifications des algorithmes ont été proposées

[MacEachren et *al*, 1999; Miller et Han, 2001]. Les algorithmes permettent de gérer les grandes quantités de données trop difficiles à gérer pour l'analyse statistique manuelle traditionnelle. Cependant, cette approche garde l'utilisateur-explorateur à l'extérieur du processus d'analyse.

2.3.2 L'exploration des données (EDA)

L'exploration des données (*exploratory data analysis* ou *EDA*) est une seconde réponse au besoin d'extraire l'information potentiellement cachée dans les larges bases de données. Plus précisément, c'est une philosophie, une approche, dérivée des statistiques, d'accoutumance de l'utilisateur avec les données [Andrienko et Andrienko, 2005]. Cette idée a été définie par John Tukey [Tukey, 1977] dans le but d'édifier les hypothèses plutôt que de les tester. Elle vise la détection des structures, tendances et relations intéressantes dans les données pour générer des hypothèses. Bien qu'elle appartienne au domaine des statistiques, elle présente des différences fondamentales avec celles-ci. Par exemple, lorsqu'on réalise une analyse statistique, on vise souvent à répondre à une question précise (ex. quelles sont les variables les plus importantes influençant la migration des jeunes?). Une analyse exploratoire elle est guidée par l'inconnu, c'est-à-dire la méconnaissance de ce que les données pourraient apporter.

Il est difficile de croire qu'une analyse ne soit complètement pas dirigée par une question quelconque. Pour réaliser une analyse, il doit y avoir une motivation initiale. [Andrienko et Andrienko, 2005] propose donc de définir l'exploration des données par ces trois caractéristiques. (1) L'exploration est dirigée par une intention d'investigation. Cette intention se traduit par une ou des questions plutôt générales. (2) Lors de l'exploration, les questions varient en niveau de généralité. La découverte d'un fait peut susciter un intérêt qui amène l'utilisateur à formuler des questions de plus en plus spécifiques pour en connaître davantage sur ce fait. (3) La majorité des questions surviennent lors de l'exploration. La motivation de départ change souvent avec les découvertes et les connaissances que l'on acquiert lors de l'exploration de données. Les questions sont aussi flexibles et dynamiques que l'est la pensée. Le concept d'exploration de données peut facilement être résumé par le « *Information Seeking Mantra* » [Shneiderman, 1996] : « *Overview first, zoom and filter, and then details-on-demand* ». On passe d'un aperçu global, à un groupe d'items d'intérêt pour ensuite d'insister sur un item pour en connaître le détail, et ce, en boucle perpétuelle.

La variété des questions constituant une exploration de données et l'aspect dynamique de celles-ci font en sorte qu'il est difficile de déterminer un seul outil ou une seule technique efficace pour supporter cette approche. L'application de la philosophie d'exploration passe par un ensemble de méthodes et de techniques disponibles en combinaison et de façon hautement interactive pour supporter le fil de pensée lors de l'exploration. SOLAP est une technologie utilisant une approche d'exploration de données. Ses caractéristiques répondent parfaitement à cette approche.

2.4 La géovisualisation

Jusqu'ici les concepts présentés se réfèrent aux sciences de l'information en général. La géovisualisation est le domaine qui s'intéresse aux particularités des données, de l'information et des connaissances géographiques. Ce domaine d'étude intègre les approches de visualisation de l'information, d'exploration de données et de découverte de connaissances avec celles de la cartographie et des SIG [MacEachren et Kraak, 2001]. Dans un tel contexte, les cartes sont utilisées pour stimuler la pensée visuelle à propos des relations, des tendances et des structures géospatiales [Kraak, 2007] et supporter la construction des connaissances géographiques. La géovisualisation vise donc le développement des outils cognitifs pour aider et rehausser la compréhension de l'information géographique.

L'aspect géographique des données tout comme l'aspect temporel devient de plus en plus une facette importante d'analyse pour plusieurs phénomènes. Aussi, lorsque l'aspect spatial⁹ est exploité, il représente généralement la principale dimension d'analyse [MacEachren et al, 1999]. La dispersion des polluants, la déforestation, la planification du territoire et la gestion des transports sont quelques exemples d'applications où la dimension spatiale est au premier plan.

La section 1.3 du précédent chapitre a initié la définition de ce qu'est la géovisualisation. Dans la présente section, les fondements théoriques de cette discipline seront détaillés. Premièrement, le foyer de naissance de cette communauté scientifique est abordé. Ensuite, les théories de la sémiologie graphique et de la synoptique des constructions graphiques appuyant ce domaine d'étude sont explicitées. Finalement, au-delà des théories, les méthodes, les techniques et les outils principalement utilisés dans cette communauté sont brièvement présentés. La

⁹ Dans le cadre de mémoire, les termes « géographique », « géospatial » et « spatial » sont utilisés comme synonymes tout au long de ce mémoire

géovisualisation constitue une communauté de recherche de plus en plus importante et inclut des collaborateurs de plusieurs spécialités. Les descriptions fournies ci-dessous résument de façon très succincte les travaux de ce domaine. Le lecteur est invité à consulter la littérature citée pour plus d'informations.

2.4.1 La cartographie traditionnelle et la géovisualisation

Le terme géovisualisation vient de la contraction des termes « visualisation » et « géographique ». Dans la littérature, d'autres appellations telles que visualisation cartographique, cartographie analytique ou cartographie de traitement ont aussi été retrouvées [Bertin, 1977; MacEachren et Kraak, 1997]. Cette discipline découle d'un changement de paradigme en cartographie. Autrefois, la production de carte relevait exclusivement du domaine de la cartographie. Ce type de document, alors, visait majoritairement à communiquer un message. Le cartographe récupérait les données, les analysait et représentait le résultat à communiquer sur une carte de la meilleure façon possible. Le cartographe s'imposait pour la production de ce document et le rôle de la carte se résumait généralement à **présenter les résultats de l'analyse, c'est-à-dire l'information**. Cette approche est d'ailleurs encore très présente dans la littérature [Dent, 1999; Slocum et al, 2005; Kraak et Omerling, 2003; Béguin et Pumain, 2007; MacEachren, 1994a; Zanin et Trémélo; 2003, etc.]. Avec l'arrivée des technologies interactives, des interfaces graphiques, des systèmes d'aide à la décision et des énormes bases de données, le rôle de la carte a été redéfini pour inclure la fonction **d'outil de réflexion, d'exploration et de raisonnement**. Le rôle de la carte s'étend alors sur tout le processus de construction des connaissances (Figure 2.9). Au lieu de viser la production de données, l'accent est mis sur l'exploration interactive des données pour faciliter la découverte de connaissances. Des recherches présentant ce nouveau point de vue ont émergé dans la littérature rejoignant les concepts élaborés en statistique et en informatique (EDA et KDD). Dès lors, des approches de plusieurs disciplines telles que la visualisation scientifique et de l'information, la cartographie, le traitement d'images, l'exploration de données (EDA) et les systèmes d'information géographique (SIG) se sont regroupées pour développer des théories, des méthodes et des outils pour l'exploration visuelle, l'analyse, la synthèse et la présentation des données géospatiales [MacEachren et Kraak, 2001]. Ce domaine d'étude est nommé la géovisualisation.

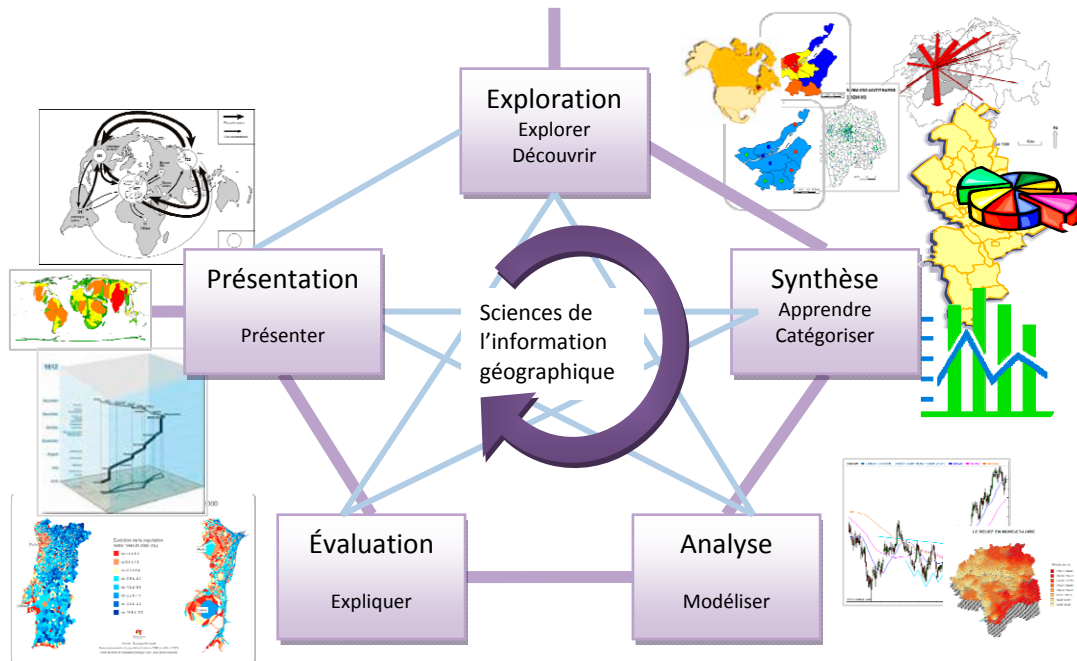


Figure 2.9 : Le rôle de la carte dans le processus de construction des connaissances spatiales.
Traduit de [Kraak, 2006]

Pour supporter toutes les étapes de la construction des connaissances, il existe deux types de représentations visuelles des données correspondant aux deux principales fonctions de la carte : la visualisation des données et la présentation des données [Mackinlay, 2008] (Figure 2.10). Ces deux types servent des buts distincts, mais la limite entre les deux n'est pas si clairement définie. Déjà, [Bertin, 1977] opposait ces deux approches. Il les nommait graphique de traitement (visualisation) et graphique de communication (présentation). La graphique de traitement « sert à découvrir les relations que les données vont révéler ». La graphique de communication « est un moyen de fixer et de dire aux autres ce que l'on a découvert » [Bertin, 1977, p.22].

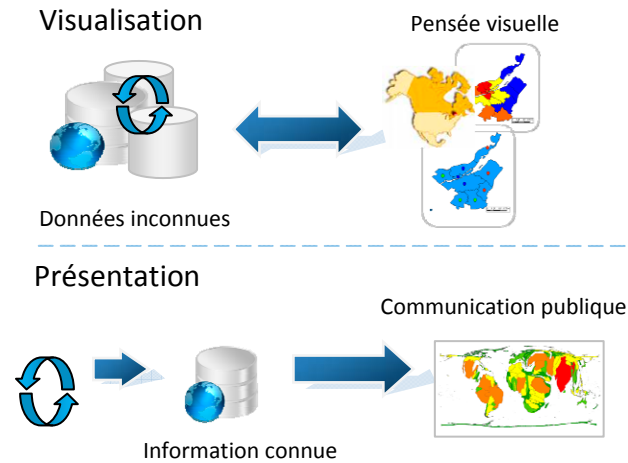
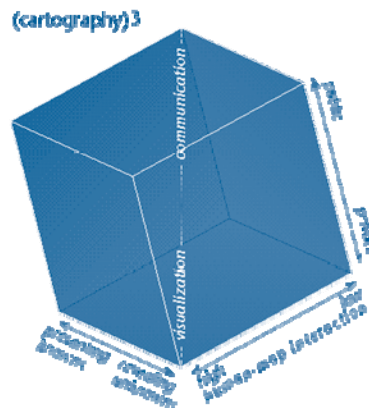


Figure 2.10 : Types de représentations visuelles

MacEachren, quant à lui, différencie la visualisation de la présentation de l'information en les contrastant selon leur utilisation. Cette définition s'appuie sur le concept de cartographie au cube ((*cartography*)³) (Figure 2.11) [MacEachren, 1994b]. Cette idée définit l'utilisation de la carte selon trois axes : l'interaction entre l'homme et la carte, le niveau de connaissance des données et le type d'utilisateur. Une définition de la cartographie comme outil de visualisation peut donc être formulée comme suit : une activité privée avec laquelle les inconnus sont révélés dans un environnement hautement interactif. Dans un même ordre d'idée, la cartographie comme outil de présentation des données, quant à elle, se définit comme une activité publique selon laquelle le savoir est connu et présenté dans un environnement peu interactif [MacEachren, 1994b].

Figure 2.11 : Définition de la cartographie selon les modes d'utilisation (*cartography*)³. Source [MacEachren, 1994b, p.6]

La géovisualisation englobe autant la cartographie comme outil de visualisation ou de traitement et comme outil de présentation ou de communication. Le rôle et l'utilisation de la carte sont donc

étendus à tout le processus de la construction de connaissances. Ces dernières années, l'intérêt est cependant davantage pointé sur l'utilisation de la carte à l'amont du processus de construction des connaissances. La plupart des recherches abordent la géovisualisation comme un outil de visualisation et cherchent à améliorer cette fonction moins habituelle de la cartographie. Avec la montée de ce domaine de recherche, la cartographie n'est plus seulement l'art et la science de présenter l'information sur une carte. Elle doit également soutenir l'exploration interactive des données réalisée par des experts de domaine; c'est-à-dire, les environnementalistes, les décideurs politiques, les gestionnaires d'hôpitaux, les analystes en transport, etc. Le domaine de la cartographie se transforme permettant dorénavant à des utilisateurs non experts d'en tirer directement profit comme outil d'analyse.

2.4.2 La géovisualisation et la sémiologie graphique

La géovisualisation, bien que différente de la cartographie traditionnelle dans son utilisation, se base sur les principes de sémiologie graphique pour la construction des cartes. Les bases de la sémiologie graphique ont été établies par [Bertin, 1973]. Peu de ces principes ont été validés expérimentalement [Nelson, 2000; Ware, 2000], cependant les principes de Bertin restent toujours une référence pour la représentation visuelle de l'information. La sémiologie graphique est définie comme l'« étude des signes graphiques, de leurs propriétés et de leurs rapports avec les éléments de l'information qu'ils expriment » [OQLF, 1973]. La géovisualisation tout comme la cartographie a tout intérêt à utiliser ces principes afin de transmettre des messages pertinents. La carte est le principal outil de représentation des données géographiques et la sémiologie graphique y est particulièrement importante. En effet, puisque l'information sur la localisation monopolise les deux dimensions du plan, la représentation de l'information thématique en 2D¹⁰ doit disposer d'une troisième dimension : la dimension visuelle [Bertin, 1973]. Les grands principes de la sémiologie graphique permettent alors de rendre sur une carte l'information le plus fidèlement possible. Ainsi, pour construire une carte qui soit visuellement efficace lors de l'exploration, le savoir-faire en sémiologie graphique est primordial.

¹⁰ Bien que les avancées technologiques permettent de rendre à l'écran des représentations en 3D, les représentations traitées dans ce mémoire se limitent au plan (2D).

La sémiologie graphique

La sémiologie graphique consiste à attribuer aux signes, aux symboles ou aux taches du dessin, le style qui permettra d'exprimer l'information désirée. Le style est appliqué selon l'**implantation graphique**, c.-à-d. de façon ponctuelle, linéaire ou zonale. La sémiologie établit donc les liens entre les caractéristiques d'un objet et leur perception en passant par l'attribution de variables visuelles selon une certaine implantation graphique. [Bertin, 1973] a défini huit **variables visuelles**. Ces variables sont la position dans la première dimension du plan (X), la position dans la deuxième dimension (Y), la taille, la valeur, le grain, la couleur, l'orientation et la forme. Ces huit variables visuelles appartiennent à un certain **niveau d'organisation de la perception** et chacune de ces variables rend avec plus ou moins de fidélité une certaine perception tel que présenté par la Figure 2.12. Les plus gros signes dans la Figure 2.12 représentent une meilleure association variable/perception et l'encadré respecte les recommandations de Bertin. Ce qui se trouve à l'extérieur de l'encadré concorde avec d'autres auteurs.


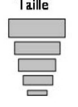
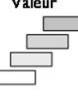
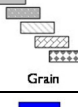



Niveau d'organisation	Associative	Sélective	Ordonnée	Quantitative
Position 	A	S	O	Q
Taille 		s	O	Q
Valeur 		s	O	Q
Grain 	A	s	O	Q
Couleur 	A	S	O	Q
Orientation 	A	s		Q
Forme 	A			

Figure 2.12 : Classement des variables visuelles selon le niveau d'organisation de la perception selon [Bertin, 1973; MacEachren, 1994a; Slocum et al, 2005]

La perception d'une variable peut être associative ou dissociative, sélective, ordonnée ou quantitative (Figure 2.13). La perception associative permet d'ignorer la variation d'une variable visuelle pour reconnaître les similarités d'une autre variable. Elle permet de regrouper les éléments d'une image. Par exemple, à la Figure 2.13a, on ignore la variation de forme pour regrouper tous les signes ayant une couleur similaire. La variable forme prend dans cet exemple une attitude perceptive associative. Une variable visuelle peut avoir une perception associative si elle n'interfère pas avec la perception des autres variables en place. La variation de taille à la Figure 2.13b domine l'image. Elle ne permet pas de regrouper tous les signes de même forme et interfère donc avec la perception de la forme. La taille est donc une variable dissociative. La perception sélective permet à l'œil d'isoler l'information et de l'abstraire de l'image. Par exemple, on peut facilement isoler les signes verts de la Figure 2.13c. On cherche à comparer les rangs avec la perception ordonnée. Naturellement, on établira les voisins thématiques (plus petit, plus grand) d'une certaine valeur. Par exemple, la Figure 2.13d s'ordonne du gris pâle au gris foncé. Un autre ordre ne serait pas intuitif. La perception quantitative cherche à définir les rapports entre les signes ou à regrouper les signes qui ont une faible distance quantitative. Un signe est non seulement plus gros, il est deux fois ou trois fois plus gros (Figure 2.13e). Ainsi, si l'on désire convier un message quantitatif, on choisira de faire varier la taille des symboles selon la quantité attribuée. Si, par contre, on veut simplement différencier des types de sols (perception sélective), on choisira de faire varier la couleur des parcelles.

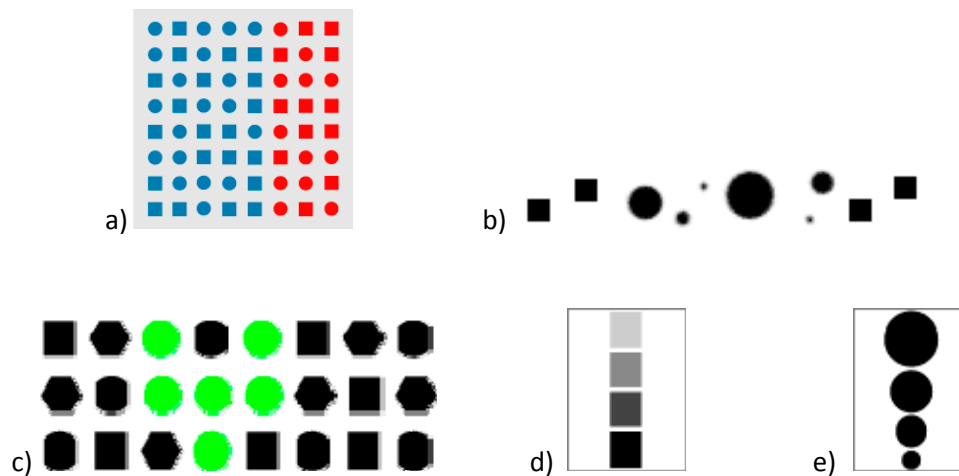
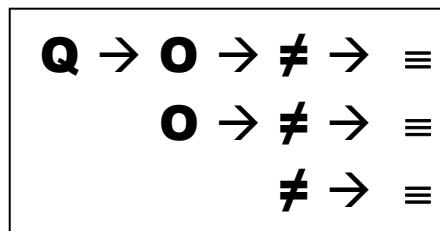


Figure 2.13 : Perception a) associative, b) dissociative, c) sélective, d) ordonnée et e) quantitative

Le choix du niveau de la perception est grandement dépendant du **niveau d'organisation des données**. Selon [Bertin, 1973], les données sont qualitatives, ordonnées ou quantitatives.

L'attitude perceptive à adopter pour un groupe de données doit avoir un niveau d'organisation au moins égale à celui des données (Figure 2.14). Par exemple, les données qualitatives seront exprimées en premier lieu selon un niveau de perception associative ou sélective. Les données ordonnées peuvent être perçues selon une attitude ordonnée, sélective ou associative. Finalement, les données quantitatives sont associées premièrement au niveau quantitatif, mais peuvent aussi adopter une attitude perceptive ordonnée, sélective ou associative. Par contre, une donnée qualitative ne peut être perçue de façon quantitative.



Q : quantitative, O : ordonnée, ≠ : sélective et ≡ : associative.

Figure 2.14 : Inclusion des attitudes perceptives. Source [Bertin, 1973]

En plus de convier un message, les variables visuelles peuvent se combiner de façon redondante ou de façon significative. Dans une situation de combinaison redondante, les deux variables combinées encodent la même information et varient ensemble. Ce type de combinaison est généralement utilisé pour augmenter la séparation visuelle entre les différents paliers. La combinaison significative implique que les deux variables visuelles encodent deux informations différentes. Les variables varient donc indépendamment l'une de l'autre. Les combinaisons de variables visuelles peuvent également être intégrables ou séparables. Une combinaison intégrable entraîne la fusion de la perception des variables. Par exemple, on ne peut dissocier la couleur de la valeur. Une combinaison séparable permet de percevoir chacune des variables indépendamment l'une de l'autre. Par exemple, la taille et la couleur sont séparables.

Avec l'avenue des nouvelles technologies et l'étude de la psychologie de la perception, certaines variables se sont ajoutées et d'autres sont critiquées. Bien qu'un consensus soit loin d'être établi dans ce domaine, l'utilisation de variables telles que l'animation ou la dynamique, le clignotement, l'ombrage, le mouvement, la transparence, le focus et la hauteur (pour les représentations volumiques) devient plus courante. La variable grain, quant à elle, apparaît sous plusieurs formes (texture, structure, espacement, arrangement, séparation, trame) et est discutée par plusieurs auteurs [Slocum et al, 2005; MacEachren, 1994a; Béguin et Pumain, 2003; Denègre, 2005; etc.].

La théorie de la sémiologie graphique de Bertin est généralement acceptée dans ses grandes lignes. Cependant, des préoccupations émergent de la nouvelle réalité du traitement de l'information. Les bases de données à analyser sont de plus en plus importantes en quantités de données et en dimensionnalité. Les données sont intégrées, croisées et comparées et la synoptique des constructions graphiques devient un défi important afin de retirer une idée globale de cet ensemble de données.

La synoptique des constructions graphiques

La géovisualisation de l'information amène de nouvelles exigences à combler qui vont au-delà des principes de la sémiologie. La géovisualisation traite des données **multivariées** et **multidimensionnelles**¹¹ dont l'accent est mis sur les dimensions spatiale et temporelle. Les multiples dimensions et variables à traiter simultanément compliquent les représentations graphiques. [Bertin, 1977] traitait déjà de ce problème par la **synoptique**¹² des constructions graphiques. L'image que procure une construction graphique inclut la vision d'ensemble des différentes variables représentées. Selon [Bertin, 1977], la barrière infranchissable de la synoptique est de trois éléments (Figure 2.15). C'est-à-dire qu'un maximum de trois éléments peut être représenté simultanément dans une construction graphique pour percevoir l'ensemble. Lorsque l'on traite des données à plus de trois variables (n), deux options peuvent permettre de transcrire l'information exhaustivement : (1) construire plusieurs diagrammes et perdre les relations d'ensemble entre les différentes variables (composition) ou (2) construire des matrices et découvrir les relations par des manipulations (répétition et interactivité). [Bertin, 1977; Mackinlay, 2008]. Ces règles, aussi simples soient-elles, ont guidé les développements de méthodes, de techniques qui appuient les outils présents dans la communauté de géovisualisation.

¹¹ Il existe une confusion dans la littérature concernant les données d'appellation multivariée et multidimensionnelle. Dans le cadre de ce mémoire, nous adoptons les définitions suivantes selon [Wong et Bergeron, 1997]

- Les données multivariées réfèrent à l'analyse des relations entre les variables dépendantes.
- Les données multidimensionnelles réfèrent à l'analyse des relations entre les variables indépendantes

¹² image, aperçu global

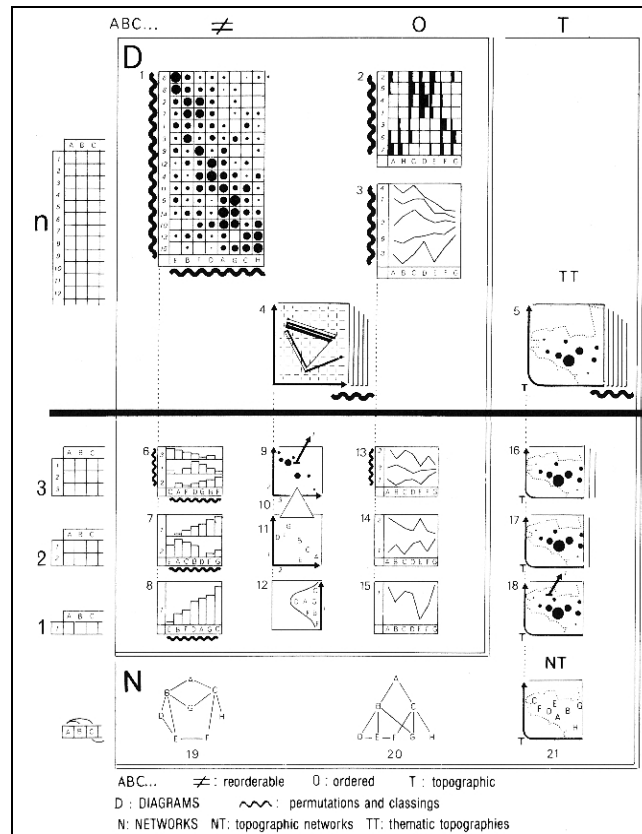


Figure 2.15 : La synoptique des constructions graphiques. Source [Bertin, 1977, p.29]

2.4.3 Au-delà de Bertin : les méthodes, les techniques et les outils

Les méthodes de représentation avancées

Dans la communauté de géovisualisation, la représentation de quantités importantes de données multivariées et multidimensionnelles est un problème encore très présent. Pour y remédier, plusieurs méthodes sont élaborées. Premièrement, de **nouvelles méthodes de représentation des données** ont été élaborées par les membres de cette communauté scientifique pour repousser la barrière de la synoptique. Ces représentations, plus ou moins complexes, permettent de percevoir les relations d'ensemble en conjuguant dans une seule image la totalité des données selon les variables étudiées (ex. *self-organizing map (SOM)*, *parallel coordinate plot*, *Cube spatiotemporel*¹³;

¹³ Le terme cube spatiotemporel prend un sens différent dans la communauté de géovisualisation de celui de l'intelligence d'affaire décisionnelle. Le cube spatio-temporel pour la communauté de géovisualisation est une représentation des données selon l'espace et le temps tel que présenté à la Figure 2.16c). Le cube spatio-temporel dans la communauté de l'intelligence d'affaire décisionnelle est utilisé pour illustrer la structure de données multidimensionnelles. Ces deux concepts ne sont en aucun point liés.

pixel-dense maps, star plot, etc.)(Figure 2.16) [Guo et al, 2005; Guo et al, 2006; Keim et Kriegel, 1999; Kraak, 2002; Dykes, MacEachren et Kraak, 2005; Card, Shneiderman et Mackinlay, 1999]. Que ce soit des méthodes novatrices ou l'application d'anciennes méthodes d'une nouvelle façon, les méthodes de représentation n'ont comme limite que l'imagination humaine. Elle tente de présenter l'information en une seule image permettant d'extraire instantanément l'information. Il est à noter par contre que l'utilité de ces méthodes avancées et parfois surprenantes est questionnée pour les utilisateurs non experts [Pendse N. et C. Bange, 2007].

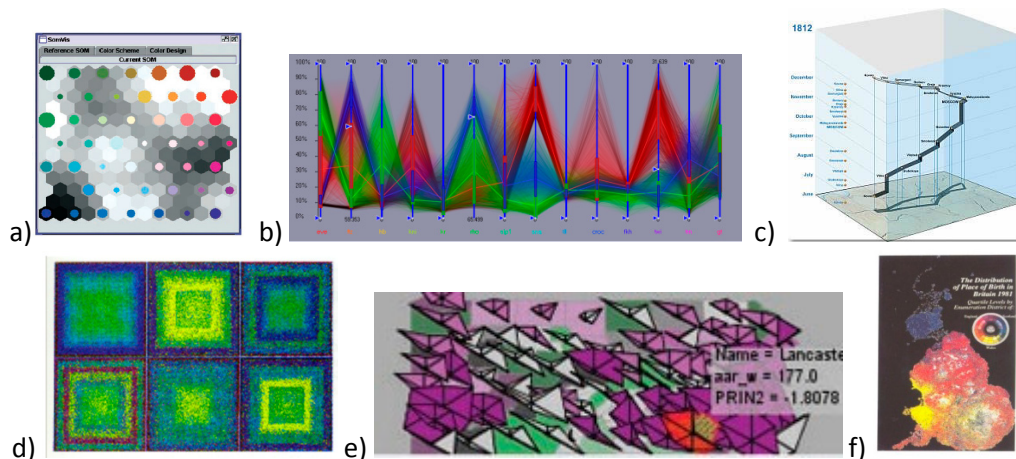


Figure 2.16 : Exemples de représentation de données multivariées et multidimensionnelles en une seule image a) SOM de couleur encodée, b) *parallel coordinate plot*, c) cube spatiotemporel, d) *pixel-dense map*, e) *star plot map* et f) cartogramme de Dorling

Les techniques

De l'autre côté de la barrière de la synoptique, trois techniques de représentation des données sont mises de l'avant pour permettre la perception des relations d'ensemble dans les données multivariées et multidimensionnelles (Figure 2.17). (1) La première, recommandée par Bertin, est la **composition**. Certaines recherches visent à augmenter la capacité d'intégration et de synchronisation de plusieurs diagrammes simultanément. L'image perçue intègre alors plusieurs représentations mettant en évidence diverses relations (ex. *GeoViz Toolkit*, [GeoVista Center, 2008], *JMapSOLAP* [Kehops Technologies, 2008]; [Guo et al, 2006]). (2) Une autre technique est celle de la **répétition** d'une image sur laquelle varie un élément [Tufté, 1990; Bertin, 1973; Guo et al, 2005]. Cette technique est généralement recommandée puisque la comparaison est facilitée. (3) Finalement, pour permettre de percevoir les différentes relations, la dernière technique est celle de la **réexpression** [Slocum et al, 2005; Kraak et Omerling, 2003]. On représente les mêmes données sous des formes différentes permettant de mettre en évidence différentes relations.

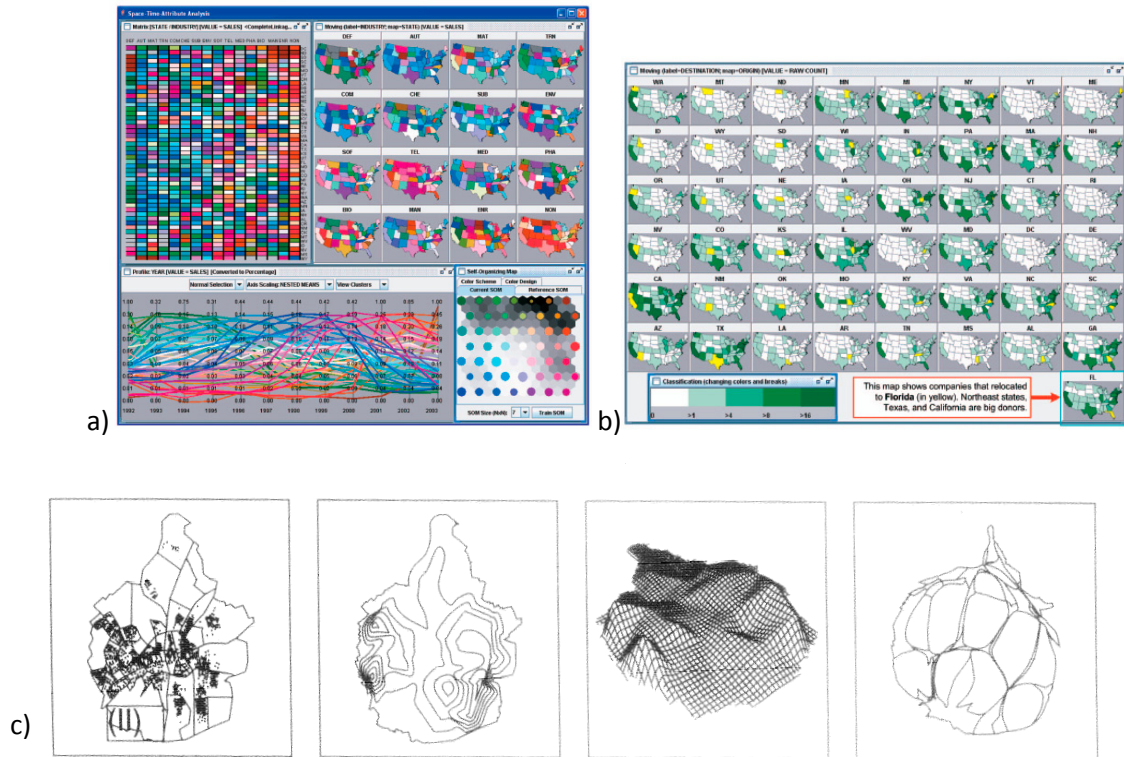


Figure 2.17 : Exemples de représentation des données en plusieurs images a) composition, b) répétition et c) réexpression

Pour permettre la découverte des relations, l'interactivité entre l'utilisateur et la représentation graphique est très importante. Aussi, la communauté de géovisualisation a développé plusieurs techniques hautement interactives permettant à l'utilisateur d'entrer dans des cycles questions/réponses très rapides. Certaines recherches se sont intéressées à **l'automatisation des règles de sémiologie graphique** (ex. *A presentation tool* [Mackinlay, 1986]; CommonGIS [Andrienko et Andrienko, 1997, 1998, 1999, 2002, 2003]; Polaris [Stolte, Tang et Hanrahan, 2002a, 2002b, 2002c]). Ceci facilite l'accès aux données d'une façon visuelle en évitant le questionnement dû à la représentation des données. D'autres études ont proposé de **nouvelles techniques d'interaction** (ex. filtrage (*filtering*), élagage (*brushing*), synchronisation (*linking*), triage (*ordering*) et présentation graduelle (*successive build up*), etc.) (Figure 2.18 et Figure 2.19) [Mackinlay, 2008; Slocum et al, 2005; Kraak et Omerling, 2003].

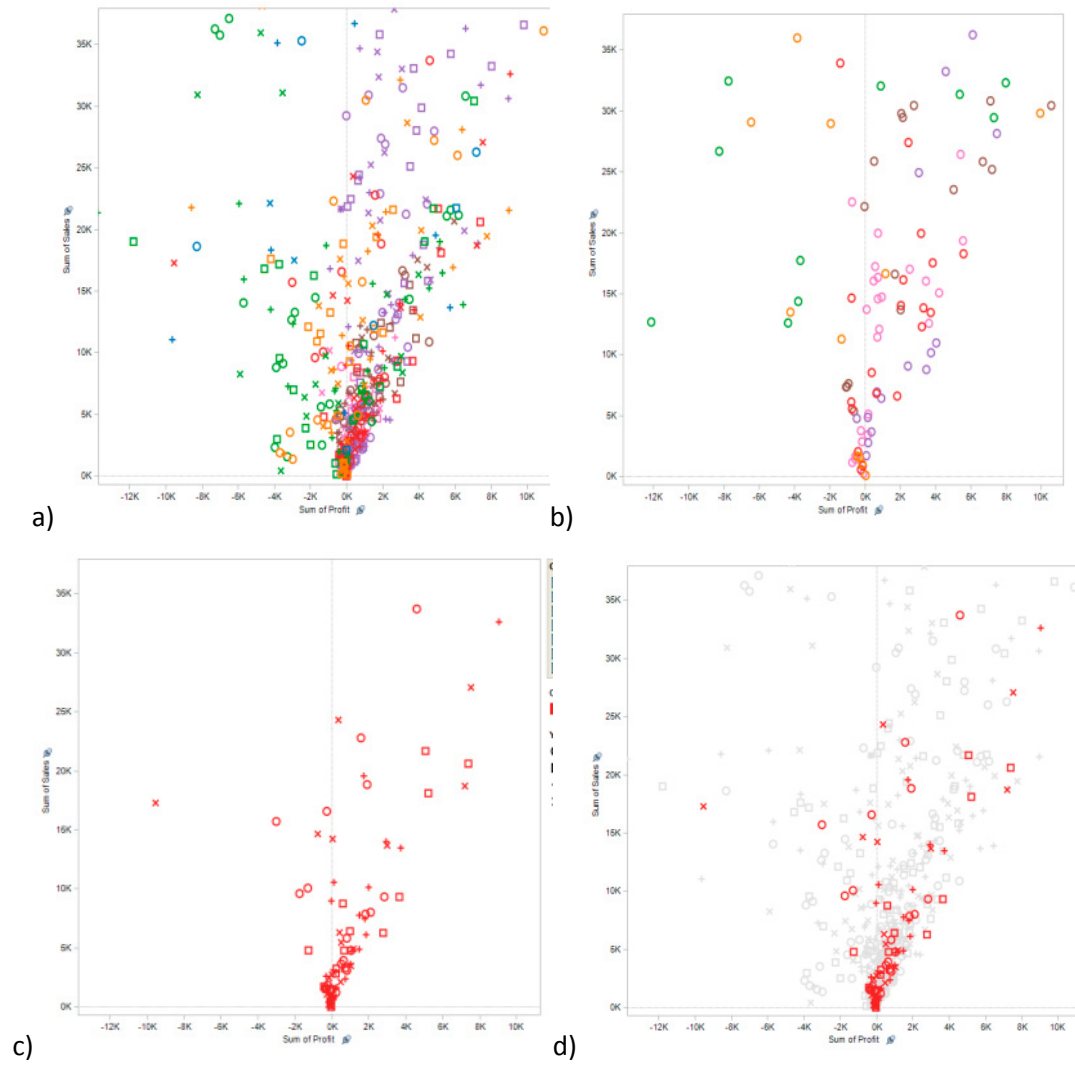


Figure 2.18 : Exemples de techniques d'interaction a) image originale b) agrégation des données, c) filtrage de données et d) élagage (*brushing*) de données

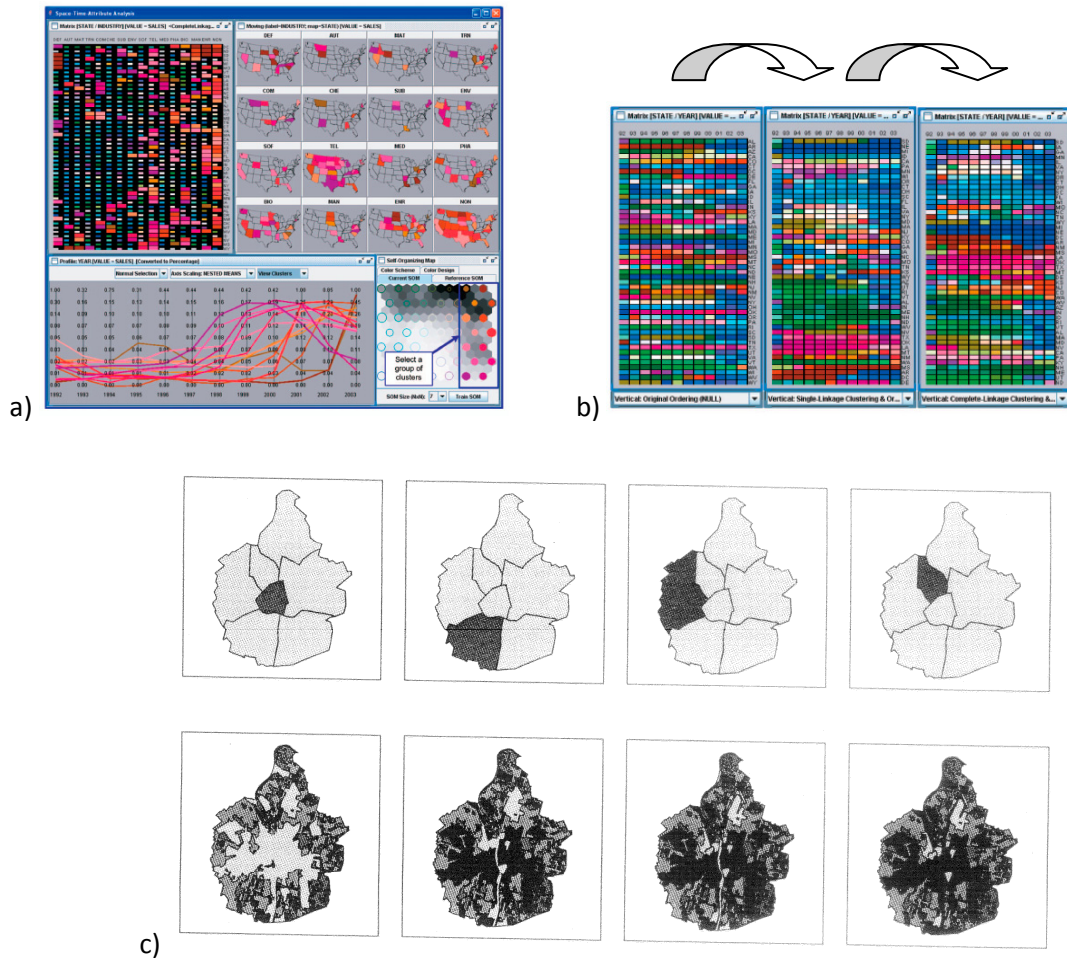


Figure 2.19 : Autres exemples d'interaction a) synchronisation de la sélection (*linking*) b) triage (*ordering*) c) présentation graduelle (*successive build-up*)

Les méthodes et les techniques présentées jusqu'ici sont une synthèse des efforts réalisés dans le domaine de la géovisualisation pour améliorer la construction de connaissances à partir des données. Ces méthodes se traduisent dans l'emploi d'outils et la réalisation d'applications diverses.

Quelques outils

Les outils développés par la communauté scientifique de géovisualisation regroupent certaines de ces méthodes et de ces techniques pour soutenir la construction de connaissances. Il est à noter par contre qu'il n'existe pas de méthode généralement acceptée pour l'exploration des données spatiales [Slocum et al, 2005] et que les outils sont des regroupements partiels de ces fonctionnalités [Keim, Panse et Sips, 2005]. Parmi les outils de géovisualisation offerts, on retrouve

CommonGIS [Fraunhofer IAIS, 2008], GeoViz Toolkit [GeoVista Center, 2008] et GeoDa [Spatial Analysis Lab, 2008] (Figure 2.20).

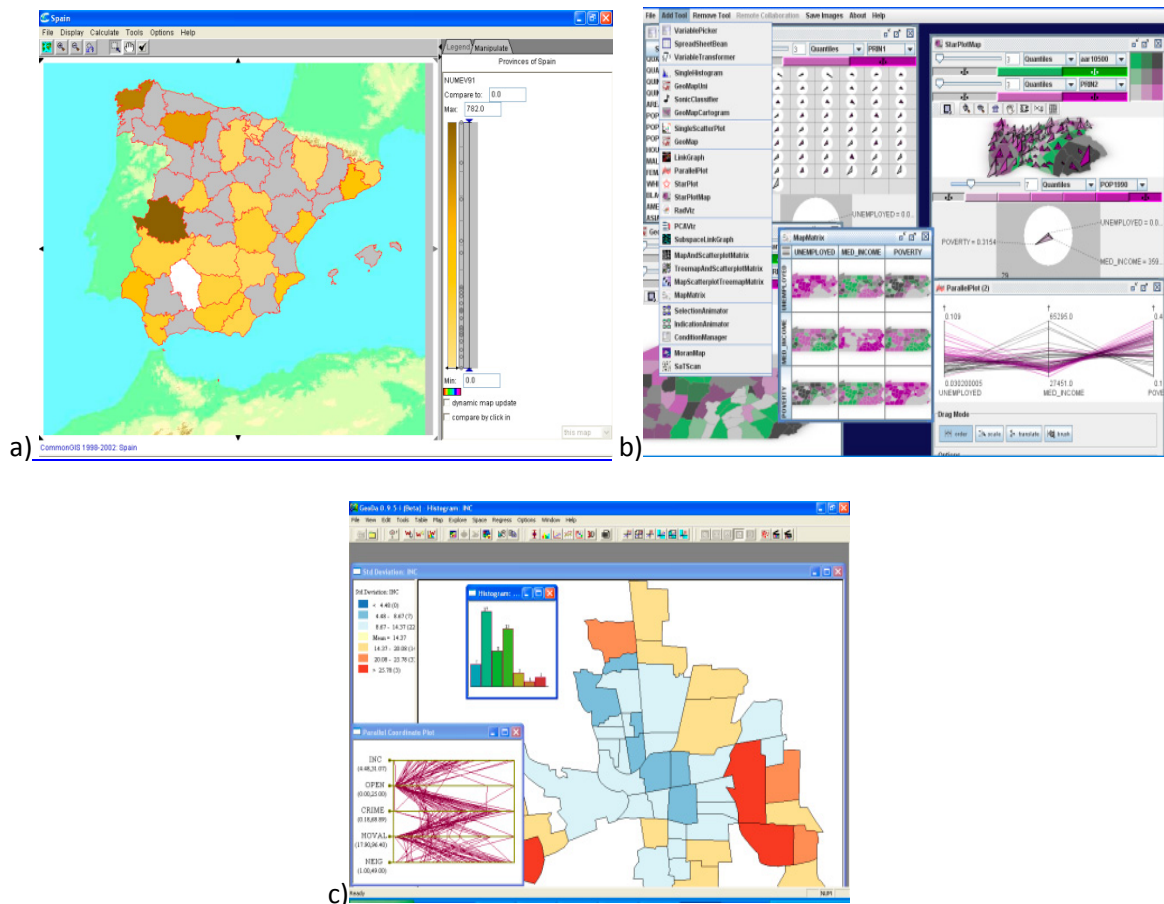


Figure 2.20 : Exemples d'outils de géovisualisation a) CommonGIS b) GeoViz Toolkit c) GeoDa

La communauté de géovisualisation est aujourd'hui très importante et bénéficie d'une collaboration serrée avec les domaines qui l'inspirent (visualisation scientifique, cartographie, traitement d'images, visualisation de l'information, cognition, exploration de données et systèmes d'information géographique, data mining). La quantité et la complexité des données représentent un grand défi pour les sciences de l'information en général. Les dimensions spatiale et temporelle et les relations particulières qu'elles contiennent amplifient ce défi. Des méthodes, des techniques et des outils sont maintenant proposés pour rehausser la compréhension de l'information géospatiale et stimuler la pensée visuelle. Cette section résume un ensemble de méthodes et de techniques et présente brièvement quelques outils tirés de la communauté de géovisualisation. Pour plus de détails, le lecteur est invité à consulter [Card, Mackinlay et Shneiderman, 1999] et [Dykes, MacEachren et Kraak, 2005].

2.5 Synopsis

Ce chapitre introduit l'approche et la philosophie sur lesquelles s'appuie l'exploration de la visualisation géographique présentée dans ce mémoire. Aucune étude sur le sujet particulier n'ayant été réalisée jusqu'à maintenant, cette étude doit explorer les approches entourant la problématique pour inspirer une solution adéquate. Dans la première section 2.1, les concepts de données, information et connaissances sont démêlés. Ceci permet de clarifier la portée de chacun de ces termes et leur interrelation. La section suivante (2.2) introduit le raisonnement par l'image, la pensée visuelle. Notre cerveau a la capacité d'analyser une grande quantité de données et d'informations pour édifier nos connaissances. Cette section met donc en évidence l'extraordinaire puissance et l'attrait de la pensée visuelle pour l'exploration interactive de données. La section 2.3 traite de deux domaines de recherche s'intéressant à la capacité limitée des technologies traditionnelles à extraire l'information des bases de données grandissantes. Le premier domaine, l'exploration de données (EDA), émerge des statistiques et exploitent principalement des moyens graphiques pour construire les connaissances. La découverte de connaissances à partir des données (KDD), quant à elle, provient du domaine de l'informatique et s'appuie principalement sur des méthodes algorithmiques (*data mining*) pour extraire, les structures, les relations et les tendances dans les données. La dernière section (2.4) de ce chapitre aborde le concept de géovisualisation dans son ensemble. Le changement de paradigme en cartographie duquel il est né, les fondements théoriques sur lesquels il s'appuie et les méthodes, les techniques et les outils de ce domaine de recherche y sont brièvement présentés. Ce chapitre initie le lecteur à l'approche utilisant la visualisation géographique pour découvrir l'inconnu et à ses concepts de base. Cette vue d'ensemble guide et soutient le développement de cette recherche.

CHAPITRE 3 - SOLAP et Géovisualisation

Le chapitre 2 introduit les différentes approches pour le traitement de l'information dans un but d'exploration interactive de données. SOLAP aborde les données selon une approche d'exploration combinant les capacités de requêtage et d'analyse visuelle spatiale (SIG) et de l'analyse décisionnelle (OLAP). En ralliant l'analyse spatiale des données, on introduit un nouvel aspect généralement absent des technologies OLAP, soit la visualisation géographique. La nécessité de l'intégration de cet aspect a d'ailleurs été soulevée dans le monde décisionnel [Meehan, 2007a, 2007b]. L'art et la science de la visualisation géographique dans un contexte d'exploration et d'analyse de données géodécisionnelles est au cœur du système d'analyse et fait l'objet de ce chapitre.

SOLAP possède déjà plusieurs caractéristiques et fonctionnalités pour l'exploration interactive des données. Cette philosophie doit maintenant s'étendre à la stratégie de visualisation géographique de l'information. Dans ce chapitre, la première section présente un tour d'horizon des caractéristiques et particularités en place de l'approche SOLAP. L'introduction de la visualisation au centre des interactions entre la base de données et l'utilisateur (*cf.* Figure 1.3) soulève ensuite des besoins spécifiques à la technologie SOLAP. Dans la section suivante, une réponse est fournie à ces besoins par la proposition d'améliorations.

3.1 SOLAP et l'exploration interactive des données

L'exploration des données avec SOLAP est définie dans cette section selon trois aspects. Premièrement, l'environnement SOLAP doit permettre une grande flexibilité d'analyse. Ensuite, l'approche multidimensionnelle se compose de différents types d'information. Finalement, des outils permettent de naviguer dans les données au fil du raisonnement.

3.1.1 La flexibilité d'analyse des données : facile, rapide et intuitive

Un aspect important recherché par l'approche SOLAP est la flexibilité d'analyse. L'approche SOLAP vise « à assister l'utilisateur dans son analyse en lui facilitant l'exploration de ses données [spatiales et non spatiales] et en lui donnant la possibilité de le faire rapidement » [Bédard, 2007].

Généralement, l'analyste est un expert de domaine. Celui-ci a une plus-value lors de l'exploration et l'analyse des données. Il connaît les enjeux, les variables non tangibles et possède des connaissances tacites enrichissant l'analyse. Par contre, ce dernier ne possède souvent pas les compétences des experts en technologies de l'information. Pour l'assister, la technologie SOLAP doit donc être facile, rapide et intuitive.

La **facilité** d'utilisation se traduit par deux caractéristiques importantes contrastant l'utilisation des SIG traditionnels. Premièrement, l'utilisateur n'a pas à maîtriser de langage de requête tel le SQL (Sequential Query Language) généralement utilisé dans les SIG. SOLAP offre une interface simple de type « *Keyboardless* » qui permet de requêter les données par de simples clics de souris. Deuxièmement, la navigation est aussi facilitée par l'utilisation d'opérateurs simples (section 3.1.3). L'utilisateur a donc accès aux données détaillées et aux données agrégées en interagissant directement avec celles-ci sans manipulation complexe ni calcul. [Bédard, 2007]

La **rapidité** d'utilisation se traduit de deux façons. Premièrement, la structure multidimensionnelle des données utilise la redondance et stocke les données pré-agrégées permettant l'accès aux données directement. L'approche SOLAP respecte les délais de la bande cognitive de Newell de 10 secondes [Newell, 1990 dans Rivest et al, 2005] incluant les opérations de requête et de réponse, mais aussi les opérations d'affichage permettant à l'utilisateur de maintenir son flux de pensée. Deuxièmement, la courbe d'apprentissage est très courte. En effet, l'utilisateur d'une technologie SOLAP devient rapidement initié et opérationnel, même s'il n'est pas expert en SIG [Bédard, 2007].

L'approche SOLAP est aussi **intuitive**. Celle-ci s'appuie sur une structure multidimensionnelle des données. La structure hiérarchique permet d'explorer les données à différents niveaux de granularité. Ceci respecte le mantra d'exploration de données de Shneiderman « *Overview first, zoom and filter, then details-on-demand* » [Shneiderman, 1996] et permet de suivre le mode de raisonnement naturel lors de l'exploration des données. De plus, l'organisation des données en thèmes d'analyse respecte la catégorisation sémantique qui s'opère lors de la mémorisation [Fortin et Rousseau, 1992; MacEachren, 1995]. Ceci facilite la navigation rendant l'environnement d'exploration de données familières à l'utilisateur et les regroupements significatifs. Finalement, la fonctionnalité de synchronisation des modes de représentation des données permet de préserver la cohérence dans les différentes informations présentées [Rivest et al, 2005].

3.1.2 Les variables et les requêtes de l'information

Les variables de l'information

L'expression de l'image créée par une construction graphique doit respecter deux règles très importantes : (1) toute l'information doit être représentée et (2) seulement l'information [Mackinlay, 1986]. L'expression d'informations additionnelles pourrait induire une faute. Par exemple, une donnée qualitative exprimée selon une attitude perceptive ordonnée induit un ordre à quelque chose qui n'en a pas nécessairement ou une pointe de tarte infère un tout alors que les parties ne s'additionnent pas nécessairement. Afin de pouvoir créer les images les plus expressives et les plus exactes possibles, la construction graphique doit exprimer l'ensemble des faits requêtés et leur relation. Un fait est la combinaison unique d'une mesure avec les membres des dimensions. Ainsi, les informations qui décrivent les faits se regroupent sous trois grands thèmes : les mesures, les membres des dimensions et les niveaux de granularité.

La **mesure** est la donnée dépendante, l'attribut ou la caractéristique [Andrienko et Andrienko, 2005]. On peut la questionner seule ou en combinaison avec d'autres mesures. Elle est souvent quantitative puisqu'elle résulte de l'observation de la réalité. Cependant, elle n'y est pas limitée. La mesure peut également être qualitative ou ordonnée (ex. le type de sol ou le niveau de risque). SOLAP définit aussi la mesure spatiale. Celle-ci peut prendre une forme géométrique résultant d'opérations spatiales ou peut également être le résultat d'opérations topologiques ou métriques [Rivest et al, 2001]. Les mesures sont le résultat, le sujet de l'analyse (ex. les accidents au Québec, l'occupation du sol, la vitesse de développement et le nombre de personnes atteintes). Cette variable prend donc une place importante dans la visualisation de l'information.

Les **membres des dimensions** sont les données indépendantes ou les référents [Andrienko et Andrienko, 2005]. Ils donnent le contexte de la mesure. Il existe trois types de dimensions auxquelles se rattachent les membres : les dimensions thématiques, temporelles et spatiales. Les dimensions thématiques peuvent être de type qualitatif (ex. les espèces de bois), ordonné (ex. l'âge) ou quantitatif (ex. le montant des budgets alloués). Les dimensions spatiales peuvent être géométriques (tous les membres de cette dimension possèdent une géométrie), non-géométriques (aucun des membres de cette dimension possède une géométrie) ou mixte, soit un

mélange des deux dimensions précédentes (Figure 3.1). Les dimensions géométriques peuvent prendre une couverture de type vectoriel ou raster [McHugh, 2008]. Les membres d'une même dimension forment un groupe similaire et ne peuvent être croisés entre eux.

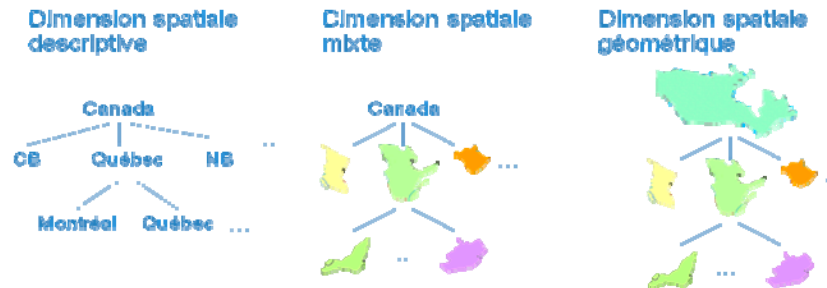


Figure 3.1 : Les trois types de dimensions spatiales supportées par le SOLAP

Le troisième type d'information est le **niveau de granularité**. Le niveau de granularité définit le détail de l'information requêtée. Par exemple, dans une dimension de produits, on peut questionner les données au niveau « catégorie de produit » (alimentaire, vestimentaire, hygiénique, etc.) ou au niveau du nom du produit (banane, pomme, pantalon, shampoing, etc.). Le niveau de granularité définit les groupes homogènes de membres permettant d'établir des comparaisons.

Les requêtes

La flexibilité d'analyse avec SOLAP donne à l'utilisateur le choix de ses requêtes. De la simple requête à une requête multivariée de données croisées et agrégées à différents niveaux, tout est possible à l'intérieur du cube. Les constructions graphiques sont par contre dépendantes des variables de la requête. Elles doivent exprimer toute l'information et seulement l'information (mises à part les informations de repérage comme les fonds de carte). Afin de démêler les cas possibles du SOLAP, simples et complexes, différentes requêtes typiques ont été identifiées.

La requête simple est probablement la requête utilisée la majorité du temps. Le résultat de cette requête est une série linéaire de cellule du cube (Figure 3.2a). Elle comprend une mesure et un groupe de membres sur un seul niveau hiérarchique. Par exemple, le total des ventes de produits à travers les municipalités de la région de Québec. Cette requête implique une donnée dépendante (le total des ventes) selon un groupe de données indépendantes sur un seul niveau de granularité (les municipalités de la région de Québec).

La requête multivariée implique de comparer plusieurs variables, c'est-à-dire plusieurs mesures entre elles. Elle résulte en une série linéaire de cellules dans le cube où chaque cellule porte plusieurs mesures (Figure 3.2b). Par exemple, je veux comparer le total des ventes et le revenu des produits selon les différentes municipalités. Cette requête compare deux mesures (le total des ventes et le revenu) selon un groupe de membres à un niveau de granularité (les municipalités).

La requête comprenant des multi-sélections dans deux dimensions ou plus implique de représenter explicitement les croisements entre les dimensions, c'est-à-dire qu'un même membre est associé à deux ou plusieurs membres d'une autre dimension résultant en une seule valeur de mesure par association. Le résultat de cette requête est illustré par une tranche d'un cube pour une multi-sélection dans deux dimensions différentes ou un sous-ensemble du cube si plus de 2 dimensions comportent des multi-sélections (Figure 3.2c). Par exemple, on cherche le total des ventes selon les catégories de produits à travers les municipalités. Cette requête implique donc que pour chaque municipalité, chaque catégorie de produits doit être représentée. Dans cet exemple, une requête de ce type implique une seule mesure et deux ou plusieurs groupes de membres (deux groupes de membres : catégories de produits et municipalités) à un seul niveau de granularité (les membres de chacun des groupes se trouvent au même niveau).

La requête multi-niveau implique que l'on demande simultanément de l'information à plusieurs niveaux de granularité dans une même hiérarchie. Le résultat de la requête est représenté par un ensemble détaillé de cellules du cube et leur cellule englobante (Figure 3.2d). Par exemple, on cherche à comparer de total de vente de chaque municipalité par rapport à leur région. Cette requête implique une seule mesure (le total des ventes) d'un groupe de membre (les municipalités) à deux niveaux de granularité (les municipalités et la région). Cette requête implique un aspect relatif parent-enfant important pour l'expressivité de l'image.

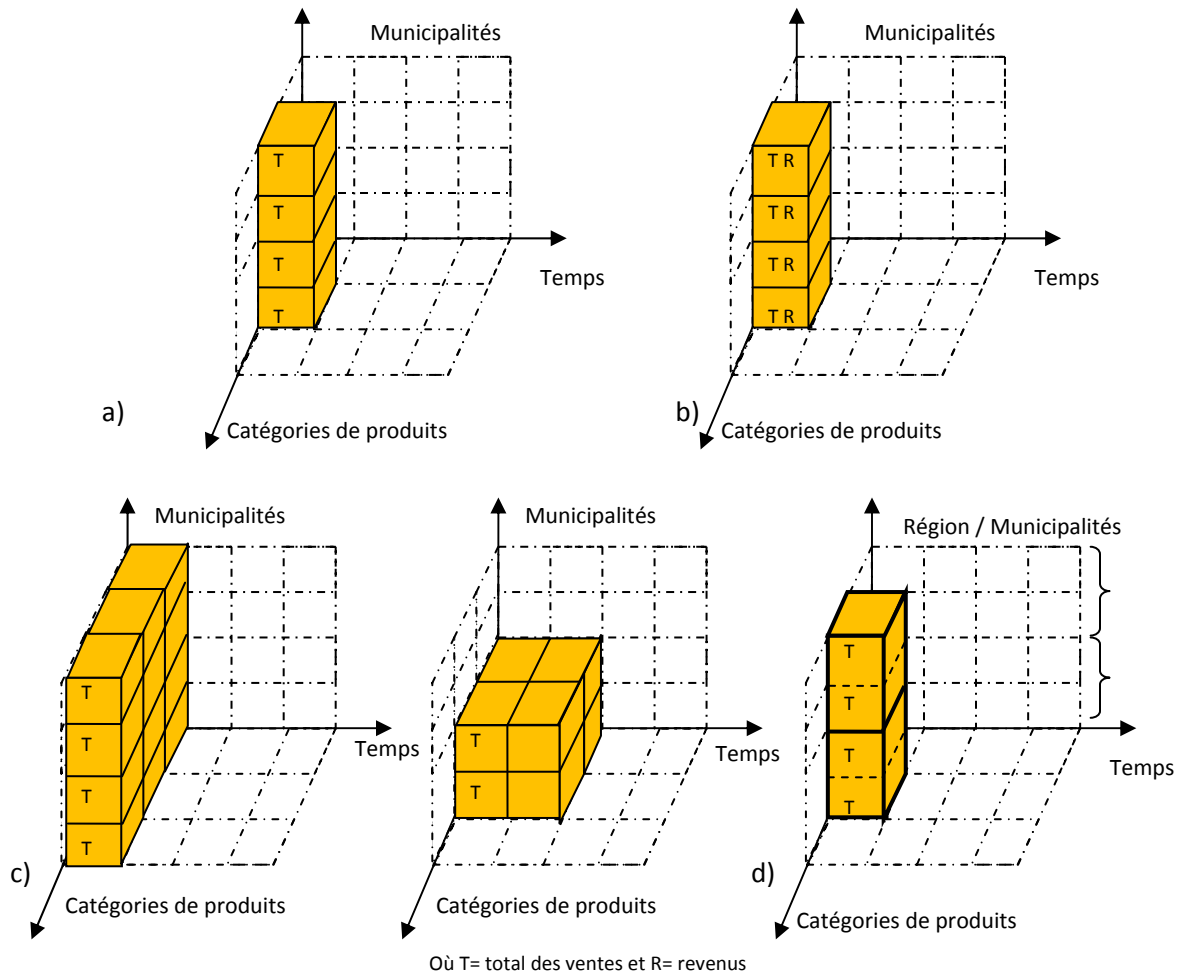


Figure 3.2 : Les types de requêtes a) la requête simple, b) la requête multivariée, c) la requête avec multi-sélections dans 2 et 3 dimensions et d) la requête multi-niveau.

Dans un environnement tel que SOLAP, les requêtes peuvent se combiner l'une à l'autre pour répondre à certaines questions complexes. Dans le plus complexe des cas, une requête multivariée, avec des multi-sélections dans plusieurs dimensions et multi-niveau pourrait être composée. Par exemple, on veut savoir le nombre d'hectares de forêts coupées et le volume de bois récolté selon les familles et les espèces de bois, entre 1997 et 2007 selon les unités administratives de la Baie James. Cette requête questionne deux mesures (nombre d'hectares et le volume de bois), trois groupes de membres (les espèces de bois, la période d'années et les unités administratives) et un groupe de membre se situe à deux niveaux de granularité (les familles et les espèces de bois).

3.1.3 La navigation dans les données

Dans un environnement SOLAP, la navigation dans les données se fait à l'aide d'opérateurs simples. Ces opérateurs s'opèrent de la même façon directement dans tous les types d'affichages. Ces outils sont les suivants : le forage, le remontage et le pivot. Le **forage** (*drill-down*) présente les données d'un niveau supérieur à un niveau plus détaillé. Il peut être appliqué sur un seul membre ou sur un niveau (Figure 3.3b et c). Le forage peut aussi être appliqué latéralement. Cette opération permet d'afficher, à partir d'un groupe ou d'un seul membre, tous les membres du même niveau hiérarchique. Finalement, le forage peut également être opéré de façon sélective. Dans ce cas, il se nomme l'ouverture (Figure 3.3d). Cette opération résulte en un forage sur un seul objet ou un groupe d'objet sélectionné. Seul l'objet ou la sélection révélera son niveau inférieur. Tous les autres membres resteront au niveau hiérarchique initial.

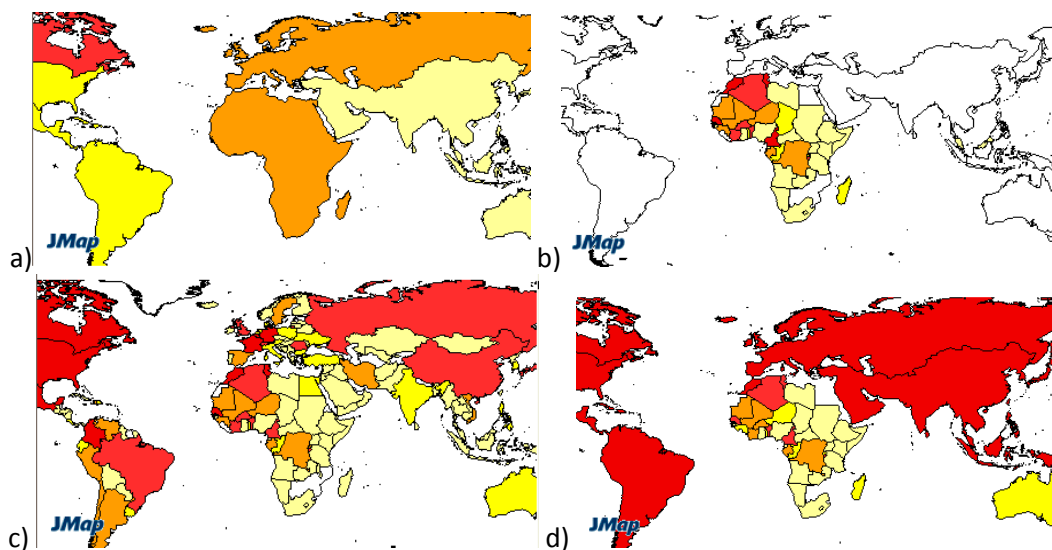


Figure 3.3 : Exemples de forage a) l'image initiale, b) le forage par membre sur l'Afrique, c) le forage par niveau et d) l'ouverture. Source [Bédard et al, 2008]

Le **remontage** (*roll-up*) est l'opération inverse au forage. Il permet de passer d'un niveau détaillé de données à un niveau plus général. Il peut également être appliqué par membre ou par niveau. Le **pivot** (*swap*) est une opération de transposition des données. Cette opération est plus facilement illustrée à l'aide d'un tableau. Un pivot sur un tableau permute les données en X pour les présenter en Y et vice versa. Sur une carte ou un diagramme, les dimensions sont également

permutées. Par exemple sur la Figure 3.4, l'information qui crée la répétition de la carte est transposée sur une seule carte.

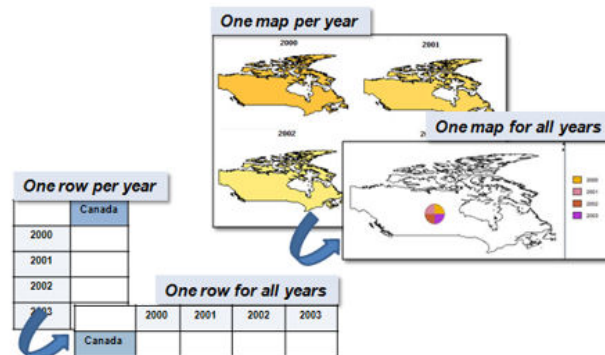


Figure 3.4 : Exemples d'une opération de pivot sur un tableau et sur une carte. Source [Bédard et al, 2008]

L'environnement SOLAP permet de questionner les données d'une façon quasi illimitée et la navigation permet d'avoir un lien direct avec les données. Pour enrichir l'exploration de données, la visualisation de l'information doit respecter la philosophie d'exploration, fournir les réponses pertinentes et augmenter les capacités d'analyse.

3.2 SOLAP et la visualisation géographique : les besoins

SOLAP est un environnement d'exploration et interactif. Pour respecter cette philosophie, la première partie de la solution est d'adapter la stratégie de visualisation à une stratégie d'exploration interactive des données, c.-à-d. que l'utilisateur doit utiliser la visualisation pour accéder aux données (cf. Figure 1.3). En modifiant le système de la sorte, plusieurs conséquences émergent. Premièrement, la visualisation géographique doit respecter la flexibilité d'analyse recherchée dans un environnement SOLAP. Ensuite, les variables particulières à cette approche doivent être fidèlement rendues par le processus de visualisation et toutes les requêtes possibles et disponibles à l'utilisateur doivent être supportées. Finalement, pour conserver le sentiment de dialogue entre les données et l'utilisateur, des manipulations graphiques, aussi essentielles que les opérations de navigation dans les données, doivent être intégrées. Les « besoins » exprimés dans cette section ne sont pas énoncés a priori, ils résultent des itérations réalisées dans le cadre de ce mémoire pour mieux répondre sur le plan cognitif aux exigences d'une exploration interactive de données spatiales.

3.2.1 La flexibilité de la visualisation géographique : facile, rapide et intuitive

La flexibilité d'analyse concerne la façon de requêter les données, mais aussi l'affichage de celles-ci. Pour respecter la flexibilité d'analyse promu par l'approche SOLAP, la visualisation géographique doit être **facile**. Ainsi, il ne doit pas être nécessaire de connaître un langage de construction graphique pour afficher les données. Les connaissances en langage graphique (sémiologie) doivent être transparent à l'utilisateur tout comme l'est le langage de requête de données. De plus, toute option graphique adéquate doit être à portée de main et peut être accédée à l'aide de quelques clics de souris.

La visualisation géographique doit aussi être **rapide**. La rapidité peut se traduire de plusieurs façons. Premièrement, elle se traduit par la rapidité d'affichage. La représentation de l'information doit se faire à l'intérieur de 10 secondes pour préserver le flux de pensée de l'utilisateur (*cf.* bande cognitive de Newell). Deuxièmement, elle se traduit par la rapidité de perception de l'information. L'utilisateur doit avoir la possibilité de percevoir l'information rapidement grâce à l'optimisation de la visualisation et ainsi de développer ses connaissances. Enfin, la rapidité d'accès aux différents modes de représentations est la troisième façon de respecter la rapidité d'analyse d'un environnement SOLAP.

Finalement, la visualisation géographique doit respecter la caractéristique d'**intuitivité**. La visualisation fait appel à l'intuition en soi, par contre certaines caractéristiques peuvent être mises de l'avant pour augmenter son pouvoir intuitif (tout comme certaines caractéristiques peuvent nuire au pouvoir intuitif). (1) Premièrement, les types de visualisation disponibles doivent être facilement interprétables. L'efficacité des différents types de visualisation avancée est questionnée [Pendse et Bange, 2007]. Les cartes et les graphiques utilisés doivent être familiers et permettre une interprétation initiale globale plutôt qu'une lecture détaillée de l'affichage. (2) Aussi, une certaine cohérence de la représentation à travers la navigation doit être préservée. SOLAP est un environnement très dynamique qui utilise la visualisation géographique comme un outil de réflexion et d'exploration de données. Chaque représentation graphique ne sert alors qu'un court moment dans le processus de la pensée de l'utilisateur. Lors de la navigation, une certaine cohérence doit par contre être conservée entre les représentations successives pour

mieux supporter le flux de pensée. Une attention particulière est portée vers les changements de niveau de granularité tant spatiale que temporelle et thématique. SOLAP supporte plusieurs structures de données géométriques et, d'un niveau de granularité à l'autre, la géométrie (point, ligne, polygone) diffère fréquemment. L'implantation de la symbologie est souvent dépendante de la géométrie originale des objets. Par exemple, si à un certain niveau, la géométrie des objets est polygonale et qu'on utilise la taille pour symboliser la mesure, le mode de représentation peut prendre la forme d'un cartogramme, d'une texture (en implantation de surface) ou encore d'une série de symboles proportionnels situés au centroïde de l'objet (en implantation ponctuelle). Supposons qu'à un niveau inférieur de granularité, la géométrie des objets soit ponctuelle. La représentation prend alors la forme d'une série de symboles proportionnels. Le passage d'un niveau où la géométrie des membres est polygonale à un niveau où la géométrie est ponctuelle entraîne un changement dans la représentation même si l'analyse reste la même. Afin de préserver le flux de pensée de l'utilisateur, la continuité dans la représentation est essentielle. (3) Ensuite, l'affichage doit s'adapter à l'échelle de représentation. Des options de généralisation cartographique et de représentations multiples sont souhaitables pour clarifier l'affichage et laisser libre cours à l'intuition.

3.2.2 Les types de visualisation géographique

Les différents types de visualisation géographique doivent représenter fidèlement les variables et leur organisation. Toute l'information doit être représentée et seulement l'information [Mackinlay, 1986]. Les types de visualisation géographiques disponibles doivent donc être en mesure de fournir les réponses pertinentes et répondre efficacement à tous les types de requêtes possibles du SOLAP.

(1) Ils doivent premièrement être en mesure de rendre expressivement et efficacement **les mesures et leur combinaison** de façon visuelle. Les mesures caractérisent le sujet d'analyse et la valeur de celles-ci, selon le contexte, doit généralement être l'information de premier plan. La représentation utilisée doit être fidèle au niveau d'organisation de la mesure choisie pour induire la bonne information (cf. section 2.4.2). Aussi, la combinaison des mesures peut créer certains conflits de représentation et les combinaisons de variables visuelles ne sont pas toutes efficaces.

Les conflits doivent être gérés de façon adéquate et transparente à l'utilisateur et les combinaisons efficaces doivent être proposées à ce dernier.

(2) Les types de visualisation géographique doivent aussi permettre l'affichage adéquat selon plusieurs groupes de membres simultanément. La requête incluant des multi-sélections dans plusieurs dimensions engage **le croisement de ces différents membres dans la représentation**. La représentation doit donc être fidèle et doit présenter la ou les mesures selon tous les croisements requêtés. Par exemple, si l'on veut connaître le niveau de scolarité principal (secondaire, collégiale ou universitaire) selon l'âge (jeune, adulte, âgé), le sexe (homme et femme) à travers les régions administratives de la province, on croise trois groupes de membres ou trois dimensions dans une seule représentation. Cette multiplicité de données doit être présentée de façon synoptique pour conserver une image globale de l'information ou donner l'opportunité à l'utilisateur d'analyser toutes les facettes de sa requête.

(3) Finalement, les types de visualisation disponible doivent gérer la possibilité d'un affichage à **plusieurs niveaux de granularité simultanément**. Par exemple, la vitesse moyenne des membres d'une équipe peut être comparée à la vitesse moyenne de cette même équipe. La visualisation géographique avec SOLAP doit gérer les propriétés de dépendance parentale non seulement dans les données, mais dans l'affichage de celles-ci également. Dans un environnement SOLAP, la relation parent-enfant est déterminée par les opérateurs d'agrégation utilisés pour composer le cube. Les opérateurs d'agrégation sont généralement des opérateurs mathématiques simples : somme, moyenne, maximum, minimum, compte, etc. Les propriétés inclusives ou relatives présentes dans la relation parent-enfant doivent aussi être rendues fidèlement à l'écran.

3.2.3 Les manipulations graphiques

Pour augmenter les capacités d'analyse à l'aide de la visualisation, l'approche SOLAP doit inclure des manipulations graphiques. Les opérateurs de navigation dans les données (c.-à-d. le forage, le remontage présenté à la section 3.1.3) permettent à l'utilisateur d'avoir un contact direct avec les données. De la même façon, parce que les représentations visuelles rendent l'information directement disponible à l'utilisateur (cf. section 2.2.2), l'interaction avec celles-ci permet d'avoir un contact direct pour rechercher l'information cachée. Les manipulations graphiques présentées

dans cette section ne sont donc pas des opérations de requêtage sur les données comme le forage ou le remontage. Ces manipulations permettent de changer la façon dont on voit les données sans changer les données présentées.

Les manipulations graphiques sont des changements apportées à la représentation sans modification aux données de la requête. Plusieurs représentations différentes peuvent être adoptées pour un même groupe de données. Les différentes représentations permettent de visualiser les données et d'organiser les variables de façon différente afin de percevoir différentes informations. Les manipulations graphiques se divisent en trois groupes : l'interaction, les modifications et la réexpression. (1) **L'interaction graphique** consiste à exécuter des requêtes visuelles à l'aide d'opérateurs simples tels que le triage et le filtrage de la légende. Lors de ces opérations, l'organisation des données dans l'affichage diffère. (cf. Figure 2.18 et Figure 2.19) (2) **Les modifications graphiques** impliquent un choix de l'utilisateur sur la symbologie des données. Le style des données se définit par l'attribution d'une variable visuelle, d'une classification et d'une implantation graphique. Les modifications n'apportent aucun changement à l'organisation structurelle des données entre elles. Elles permettent simplement de percevoir différemment les mêmes données en mettant l'emphase sur certaines caractéristiques du spectre de données. Par exemple, en changeant la classification de classes égales à une classification par moyenne emboîtée les extrêmes seront mieux révélés. (3) Le troisième groupe de manipulations graphiques est la **réexpression**. Ce concept, emprunté au domaine de la géovisualisation, consiste à représenter la même information organisée différemment (cf. Figure 2.17c). Par exemple, passer d'une carte à un diagramme, ou d'un diagramme à un tableau croisé, ou d'une carte multivariée à une répétition d'une carte simple ou encore d'une carte avec diagrammes superposés à une composition carte-diagramme.

Dans cette section, les besoins en matière de visualisation dans un environnement d'Exploration interactive des données géodécisionnelles sont énumérés. Ces besoins peuvent prendre différentes formes. Des améliorations concrètes sont présentées à la section suivante.

3.3 SOLAP et la visualisation géographique : les améliorations potentielles

Suivant les besoins du SOLAP en matière de visualisation géographique soulevés par la nouvelle stratégie, certaines améliorations potentielles ont été identifiées. Il est à noter par contre que cette liste n'est pas exhaustive car une telle exhaustivité est difficile, voire impossible à atteindre et ne faisait d'ailleurs pas partie des conditions d'atteinte des objectifs de cette recherche de maîtrise. Cette liste est toutefois le reflet de la situation et du travail exécuté selon la méthode de recherche expliquée au chapitre 1 (*cf.* section 1.5 et Figure 1.5). Des idées y sont soulevées s'appuyant sur la littérature explorée et les essais technologiques exécutés combinés aux besoins de visualisation géographique du SOLAP identifiés.

3.3.1 L'automatisation de la visualisation géographique

Un des aspects les plus importants de l'approche SOLAP est la flexibilité d'analyse qu'elle promeut. Cette flexibilité engendre plusieurs besoins concernant la visualisation géographique (*cf.* section 3.2.1). Dans le but de respecter la philosophie d'exploration SOLAP, la relation entre les données et la visualisation doit être indépendante de l'utilisateur. Ce dernier pour conserver son flux de pensée doit intervenir le moins possible dans la construction des méthodes de visualisation tout en conservant une grande flexibilité dans le choix de représentation des données. Une automatisation de la symbologie selon les règles sémiologiques permettrait donc à l'utilisateur de bénéficier d'une expertise en visualisation géographique sans implication de sa part. Certains auteurs ont déjà tenté, avec succès, de relever le défi de l'automatisation du design de visualisation. [Mackinlay, 1986] propose un système expert qui permet d'évaluer des critères d'expressivité et d'efficacité pour proposer le graphique le plus pertinent aux données d'entrée. [Stolte, Tang et Hanrahan, 2002a, 2002b, 2002c] ont développé un algèbre qui permet de construire et de jouer avec les diagrammes très librement à partir de groupes de données. La particularité géographique et dynamique de la visualisation avec SOLAP pousse le défi de l'automatisation un peu plus loin. Une solution théorique pour permettre l'automatisation de la visualisation géographique à l'intérieur du cadre sémiologique est proposée au chapitre 4.

Malgré une automatisation du processus d’affichage des données, la facilité, la rapidité et l’intuitivité de l’approche SOLAP dépend aussi d’une réponse géovisuelle adéquate à toutes les requêtes de données et des manipulations graphiques disponibles à l’utilisateur pour approfondir son exploration.

3.3.2 L’accès à différents types de visualisation géographique adaptés

Les types de visualisation géographique à inclure dans une approche SOLAP sont nombreux. Certains exemples sont présentés ici répondant aux trois besoins identifiés à la section 3.2.2 : les mesures et leur combinaison, le croisement des dimensions et l’affichage de plusieurs niveaux de granularité.

(1) Dans le but de supporter la représentation des mesures et de leur combinaison, on peut utiliser premièrement la **visualisation multivariée** où l’on attribue un style par mesure et qu’on combine les styles pour former la symbologie des objets (Figure 3.5a). Les différentes combinaisons de variables peuvent être intégrables ou séparables. Dans le but de comparer deux mesures différentes, le choix des combinaisons doivent être séparables (*cf.* section 2.4.2). Des combinaisons telles que taille en X -taille en Y, taille-valeur, couleur-orientation et couleur-taille [Ware, 2000; Bertin, 1973; Béguin et Pumain, 2003; Hearnshaw, 1994; MacEachren, 1995] sont recommandables pour ce type de représentation. La combinaison de mesures est aussi bien rendue par la **répétition** (Figure 3.5b). Elle permet de séparer les deux mesures tout en facilitant leur comparaison. Ce genre de représentation est souvent rencontré dans la littérature cartographique et plusieurs le recommandent pour un petit nombre de répétitions.

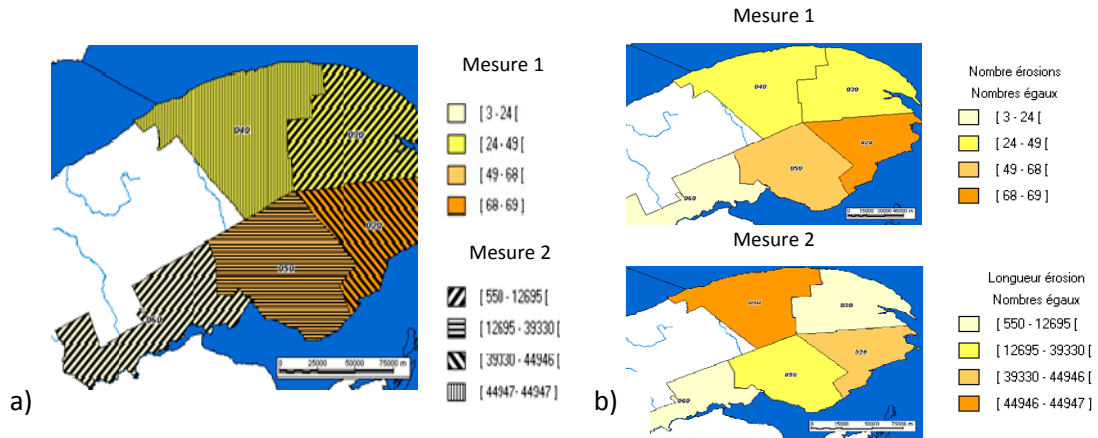


Figure 3.5 : Visualisation géographique des mesures et de leur combinaison : a) carte multivariée et b) répétition de carte simple

(2) Le croisement de dimensions peut être rendu de différentes façons : la composition, la répétition et les superpositions de diagrammes. La **composition** permet de représenter les données à l'aide de plusieurs graphiques présentant les différents contextes. Par exemple, on peut vouloir croiser les dimensions temps, provenance mondiale et programme de DEC dans un cube de gestion des inscriptions universitaires pour connaître le nombre d'inscrits. La collection de diagrammes serait composée d'une carte pour représenter la dimension géographique combinée à un diagramme à ligne représentant le temps et d'un histogramme présentant les inscriptions selon le programme de DEC (Figure 3.6). Chacun des graphiques faisant abstraction des deux autres dimensions. La relation d'ensemble entre les dimensions est perdue, mais une image globale de chacune des dimensions est présentée simultanément. L'utilisateur peut alors naviguer, filtrer et raffiner sa requête pour obtenir des réponses plus détaillées, surtout lorsque la technologie SOLAP permet de forer dans chacune des vues tout en les gardant synchronisées si désiré.

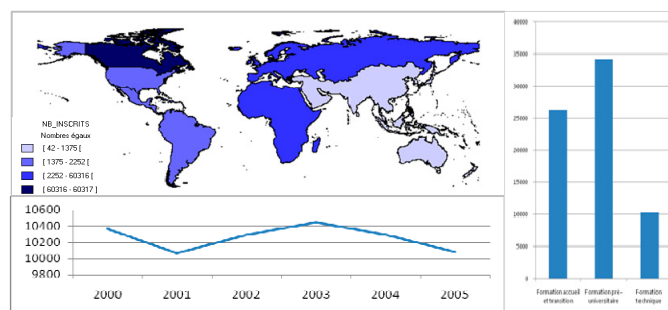


Figure 3.6 : Composition selon 3 dimensions

Une autre façon de visualiser les croisements de données est d'utiliser la **répétition**. La répétition donne une image globale de chacun des croisements, mais néglige la relation d'ensemble sur chacune des dimensions. La répétition consiste à répéter l'image d'une dimension autant de fois qu'elle est croisée avec une autre dimension. La répétition peut être définie comme une matrice selon les deux axes du plan : le croisement d'une dimension sur l'axe X et le croisement d'une autre dimension sur l'axe des Y (Figure 3.7). À cause des limites du plan, le nombre de dimensions croisées pouvant être représentée ainsi est de trois. La répétition est recommandée pour un petit nombre de croisements. L'image globale de la situation pourra facilement être reconstituée mentalement avec des opérateurs d'interaction graphique.

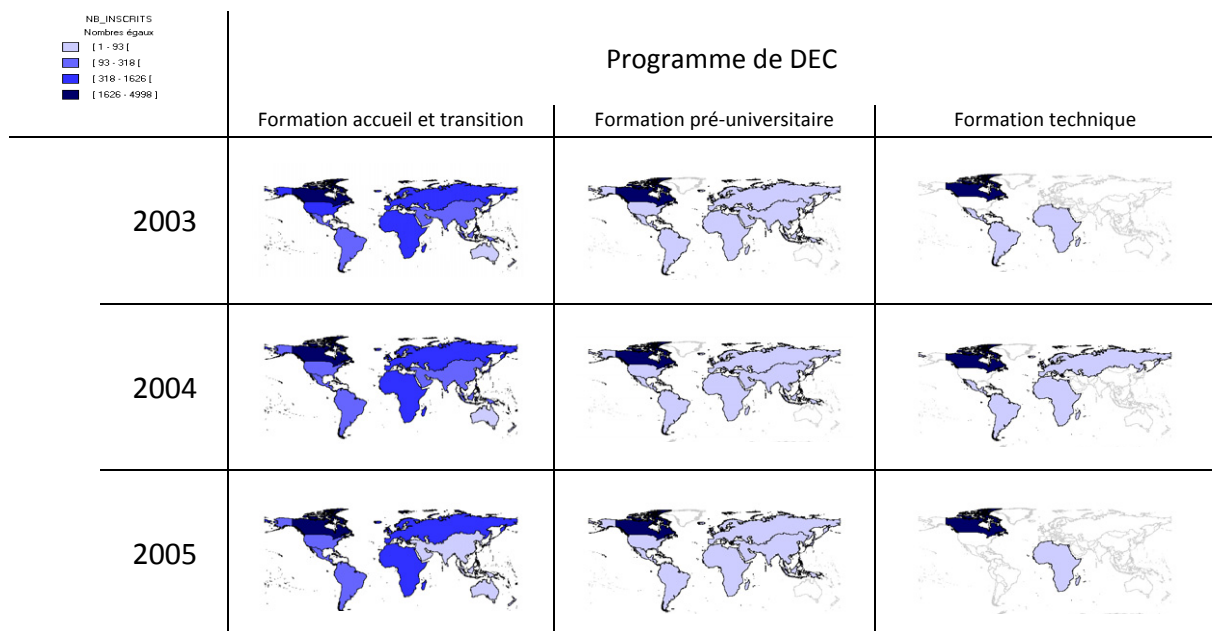


Figure 3.7 : Répétition selon 3 dimensions : Provenance Mondiale, Programme de DEC, Temps

La troisième façon de représenter les croisements est la **superposition de diagrammes** (Figure 3.8). Habituellement, ce type de représentation entre dans la catégorie des graphiques de traitements. Une image globale est plutôt difficile à tirer de ce type de représentation. Par contre, toute l'information est disponible pour une étude plus attentive de la situation. On aperçoit souvent ce type de représentation dans la cartographie statistique et les diagrammes superposés peuvent devenir complexes. En fait, tout diagramme pouvant être composé seul doit pouvoir être superposé à une carte pour évaluer la distribution géographique. La superposition de diagrammes peut aussi être disponible dans les graphiques et même dans les tableaux.

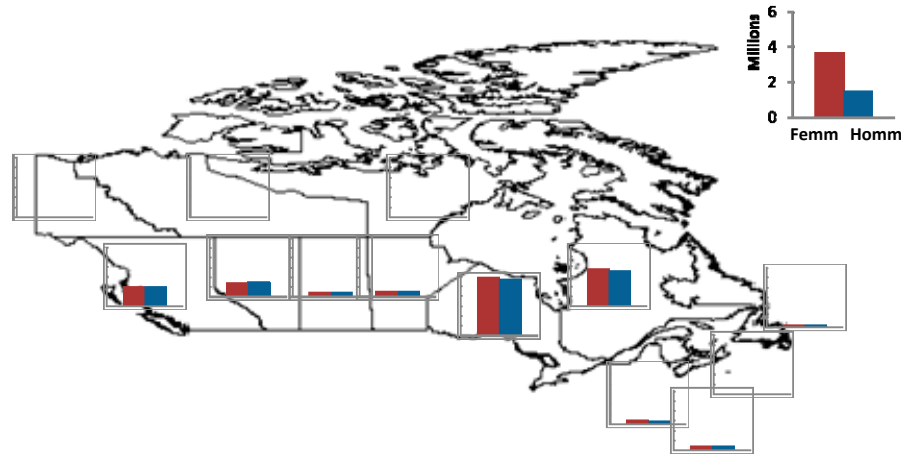


Figure 3.8 : Exemple d'une superposition de diagrammes sur une carte

(3) Supporter l'affichage de plusieurs niveaux de granularité simultanément et rendre visible la relation parent-enfant dans la visualisation géographique est importante pour une requête multi-niveaux. Un parent n'est pas une entité indépendante de ses enfants, il se compose d'eux. La relation de composition peut être inclusive ou relative. Ceci peut se traduire de différentes façons : une attention particulière à la symbologie pour créer des familles, des diagrammes comparatifs ou des groupements. Un premier exemple de visualisation géographique multi-niveaux est lorsque l'opérateur d'ouverture est utilisé. Cette région spécifiquement affichera les données au niveau inférieur et toutes les autres régions resteront au niveau initial. Par exemple sur la Figure 3.9a, trois municipalités voisines de Québec ont été forés pour découvrir la valeur du nombre de ménage pour toutes les cellules qui les composent. Toutes les autres municipalités affichent leur nombre de ménages agrégés. Sur une carte où la relation entre les niveaux hiérarchiques est inclusive (opérateur d'agrégation somme), on ne peut comparer les valeurs du niveau parent avec les valeurs du niveau enfant puisqu'elles se situent à des échelles de valeurs différentes. Une autre légende doit donc être créée pour le niveau enfant (Figure 3.9a). Sur une carte, la relation entre le membre parent et les enfants est évidente à cause de la localisation géographique. Dans un diagramme, par contre, la relation d'appartenance doit être préservée par un **groupement de données**. Dans l'histogramme de la Figure 3.9b, chaque barre représente une municipalité et le détail des celles des trois municipalités montré sur la carte est représentée par un cumul des cellules.

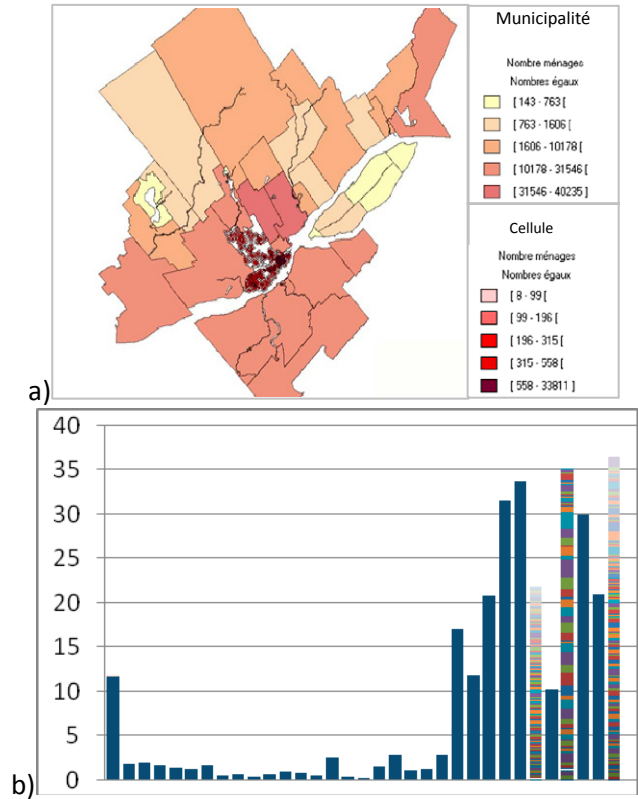


Figure 3.9 : Carte multi-niveau du nombre de ménages au niveau des municipalités et des cellules et le diagramme équivalent.

Un autre type de relation est la relation relative. Par exemple, trois athlètes cherchent à savoir comment leur vitesse moyenne se compare par rapport à la vitesse moyenne de l'équipe. Dans ce cas, la vitesse de l'équipe se situe dans le même ordre de valeur que la vitesse de chacun des athlètes. Cependant, elle ne se réfère pas au même contexte et se compare à chacune des valeurs. Un **diagramme comparatif** s'appuyant sur la valeur du membre parent (Figure 3.10a) et/ou une **symbolologie particulière** délimitant la valeur de chacun des individus par rapport à celle de l'équipe (Figure 3.10b) rendent la relation parent-enfant visible.

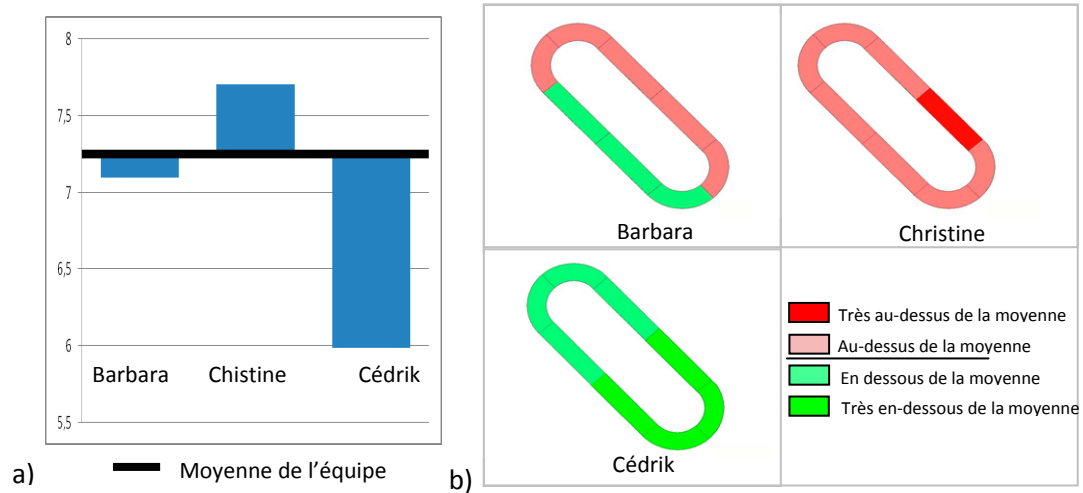


Figure 3.10 : Comparaison de données d'évaluation de performance sportive (cf. projet en kinésiologie de M. Lambert et J.P. Veilleux) a) Diagramme et b) Carte multiple relative à la moyenne

« Cent constructions pour une information » [Bertin, 1973]. Plusieurs types de visualisations différents peuvent être fournis pour représenter la même information. Il n'existe pas de représentation idéale pour une situation, mais certains types de représentations sont mieux adaptés pour une situation donnée. Quelques exemples de cartes et diagrammes et les spécificités de ces modes de représentation ont été présentés dans cette section pour répondre au besoin de la visualisation géographique des données avec SOLAP. La visualisation est cependant plus qu'une appréciation statique des données. Pour augmenter la capacité d'analyse, l'utilisateur doit être en mesure de manipuler facilement les affichages pour répondre aux questions du moment.

3.3.3 La diversité des manipulations graphiques

Les manipulations graphiques potentielles sont nombreuses. Elles sont ici regroupées selon les besoins d'interaction graphique, de modifications graphiques et de réexpression.

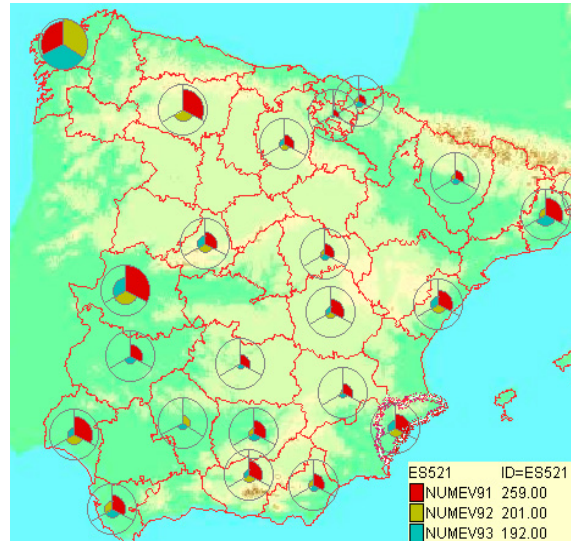
L'interaction graphique

L'interaction graphique consiste à permettre à l'utilisateur d'exécuter ses requêtes visuellement. Comme pour l'interaction avec les données, l'interaction graphique permet à l'utilisateur de manipuler directement l'affichage des données. Un premier opérateur de sélection lui permettrait de mettre certaines données en évidence reculant à l'arrière plan les données non sélectionnées.

L'opération se nomme **élagage** (traduction libre pour *brushing*). Un second opérateur de sélection, **le filtrage**, pourrait lui permettre de choisir des données spécifiques éliminant les données non sélectionnées de l'affichage. Le filtre peut s'appliquer sur les valeurs de la mesure représentée ou sur les membres. Le **trilage** est un troisième opérateur qui permettrait de réorganiser les données ordonnables entre elles selon, entre autres, la valeur du sujet d'analyse.

Dans une composition de diagrammes, l'élagage, le triage et le filtrage peuvent se propager aux autres affichages par une opération de **synchronisation des données**. Ceci permet de connaître la réaction des données dans les différents affichages et selon les différents contextes affichés. Un autre type de synchronisation est la **synchronisation de la symbologie**. Celle-ci permet de distribuer la légende à tous les affichages d'une composition. La synchronisation de la symbologie est parfois délicate et n'est pas toujours souhaitable. Certains conflits ou mauvaises représentations peuvent résulter de cette opération. Par exemple, en utilisant une composition de carte multi-niveaux et histogramme (Figure 3.9), la légende s'applique aux valeurs de la mesure et plus aucune différenciation n'est possible pour identifier les différentes cellules des municipalités en ouverture. Elle est par contre fort utile lorsque plusieurs représentations présentant la même information en premier plan (généralement la mesure) avec une même échelle de valeur. Elle permet à l'utilisateur de ne mémoriser qu'une seule fois la légende et d'effectuer rapidement les associations à travers la collection de représentations (Figure 2.19a).

Finalement, parmi les opérations d'interaction graphique, la **comparaison interactive** doit permettre d'établir directement une référence pour les différentes comparaisons, soit par la manipulation d'un seuil ou par le choix visuel de l'objet de référence (Figure 3.11).



Les modifications graphiques

Les modifications graphiques sont une modification au style des données (mesure ou membres) sans modifier les données de la requête ou leur organisation visuelle. Plus simplement, les modifications graphiques sont un changement à la légende de la représentation. Les légendes sont souvent nécessaires dans les représentations cartographiques pour définir la symbologie. Comme la position des objets occupe les deux dimensions du plan, la symbologie permet de visualiser l'information selon une troisième dimension, la dimension visuelle. La légende donne une signification à la symbologie. Modifier la symbologie est une opération strictement visuelle et si une caractéristique de la symbologie devait être changée, les modifications devraient pouvoir être opérées directement dans la légende. La légende, en plus d'afficher la symbologie, deviendrait un **outil interactif** de définition des styles. Ceci permettrait à l'utilisateur une plus grande facilité d'interaction avec la représentation. De plus, les changements apportés dans la légende devront se refléter automatiquement dans l'affichage. L'utilisateur bénéficierait donc de la possibilité de constater le changement à l'écran immédiatement.

Dans un autre ordre d'idée, un **remontage sur la légende** permettrait d'accéder à différents niveaux de précision de la symbologie. Habituellement, la représentation affiche le niveau de détail déterminé par le niveau de granularité des données de la requête. Ces données peuvent être très détaillées et leur regroupement visuel pourrait alléger la représentation et permettre

certaines associations. Par exemple, la Figure 3.12a montre un nuage de points affichant la relation entre deux mesures et la Figure 3.12b affiche la valeur d’une mesure par rapport à une dimension quelconque. Les figures de gauche montrent un niveau de précision visuelle correspondant au niveau de granularité de la requête. Après une opération de remontage sur la légende, les données sont groupées visuellement selon le niveau supérieur, mais le niveau de détail de la représentation reste égal au niveau de granularité de la requête. Quelquefois certains détails peuvent être omis visuellement et leur omission permet d’éclaircir la représentation pour une meilleure compréhension des données. L’opération de remontage sur la légende permet d’abstraire visuellement certains détails (le modèle de voiture dans l’exemple de la Figure 3.12) tout en conservant toutes les données sans agrégation.

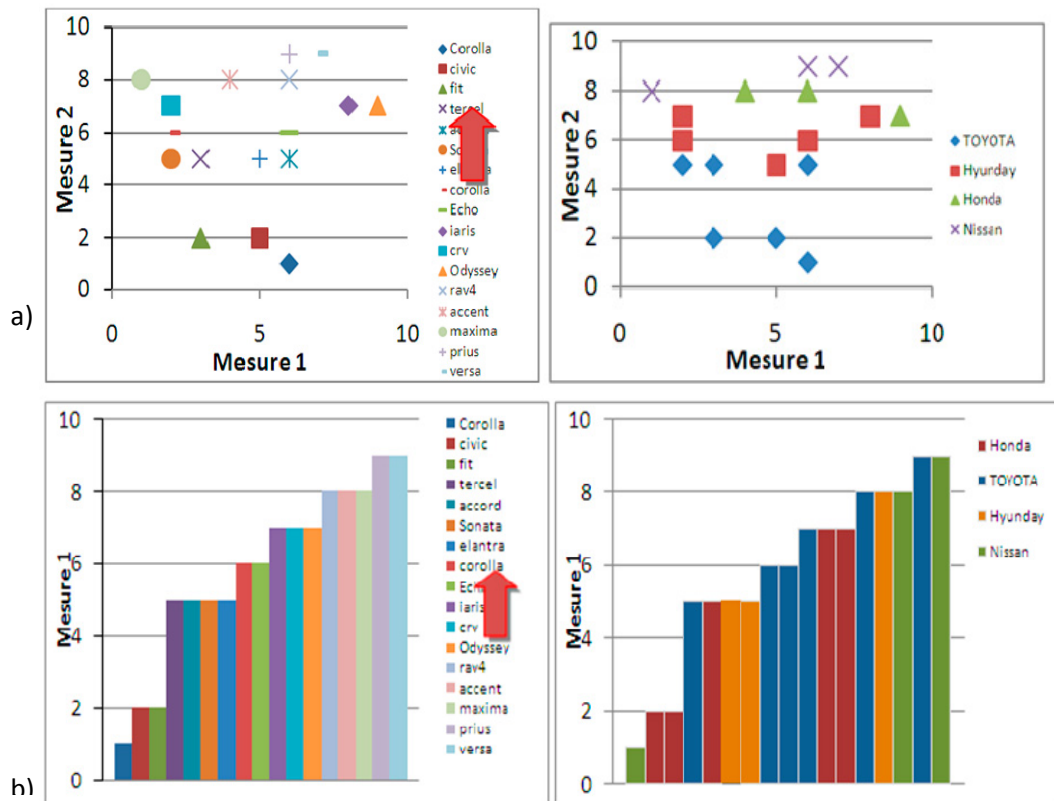


Figure 3.12 : Changement de précision symbolique a) dans un nuage de point, b) dans un histogramme

La réexpression

La réexpression est le troisième groupe de manipulations graphiques qui améliorerait potentiellement la visualisation géographique d'un environnement SOLAP. La réexpression consiste à offrir un choix multiple de représentation de la même information. Aucune représentation n'est idéale pour toutes les situations. La réexpression permet donc à l'utilisateur de naviguer dans les représentations pour trouver celles qui lui permettront de raffiner ses connaissances des données. La première forme de réexpression est la possibilité de **basculer d'une représentation discrète à une représentation continue du territoire**. Cette possibilité est d'abord dépendante de la structure de données. [McHugh, 2008] a défini les concepts de base de l'intégration du matriciel dans un environnement SOLAP. Cependant, des options de transfert à la volée d'une couverture vectorielle à matricielle et l'inverse pourrait permettre ce changement à l'affichage sans modification à la structure des données. L'utilisation de ce modèle de données pourrait également servir les différentes représentations des phénomènes pour s'abstraire des limites visuelles des objets. Par exemple, la représentation des ventes par unité administrative nous oblige à considérer le regroupement défini des données par unité administrative. Une représentation raster permettrait de voir le territoire davantage comme une grille s'abstrayant des limites politiques si nécessaire.

Ensuite, parmi les opérations de réexpression, une **liste géovisuelle** adéquate et variée pourrait être fournie à l'utilisateur lui permettant de changer de construction graphique aisément tout en conservant le sujet de son analyse. Les différentes constructions fournies mettent en évidence différentes informations qui lui seront accessibles en fouillant dans les différentes représentations.

Finalement, la réexpression peut aussi prendre la forme d'une **abstraction visuelle** des données. Un exemple est la possibilité de regrouper les faits selon la valeur de plusieurs mesures plutôt que selon les membres qui définissent leur contexte. Les regroupements des faits selon la valeur des mesures se basent sur les ressemblances de leur profil. Pour illustrer cette technique, la Figure 3.13 montre une requête multivariée comprenant quatre mesures. Au lieu de caractériser chacun des faits par quatre symbologies différentes correspondant à chacune des quatre mesures, on peut regrouper les faits selon leur profil de valeurs de mesures et leur attribuer un seul style. Par exemple, le côté gauche de la figure montre quatre symbologies différentes (forme, couleur,

valeur et taille) qui devront se combiner pour chacun des faits requêtés. La lecture de l'image résultante est donc une lecture présentant les détails de chacune des mesures. Une option d'abstraction pourrait permettre de regrouper les faits selon le profil des valeurs des mesures. Le côté gauche de l'image présente un diagramme parallèle (*parallel plot*) illustrant le profil des valeurs de mesures. Chaque fait appartient alors à un groupe caractérisé par le profil des mesures. Ainsi, les faits reliant les mêmes classes de mesures se verront attribués le même style. Une seule symbologie définit les groupes de mesure et la légende est donnée par un profil plutôt que par des valeurs de mesures. La représentation omet alors le détail de chacune des mesures pour faire ressortir les différences entre chacun des groupes. La lecture de l'image passe d'un niveau détaillé à un niveau plus abstrait.

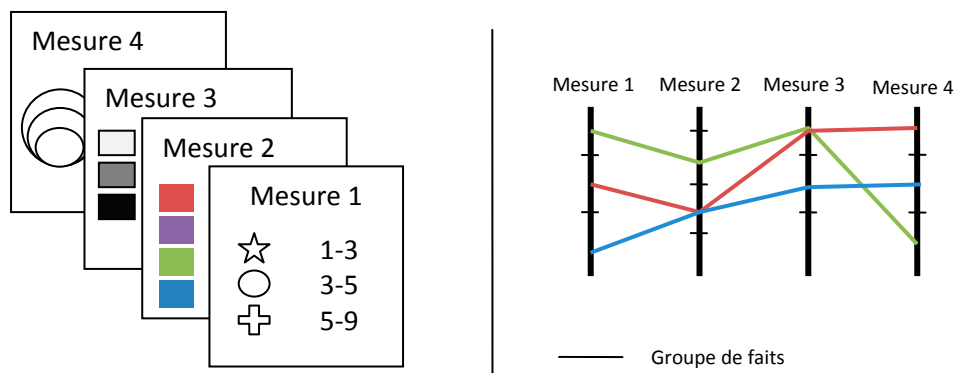


Figure 3.13 : Regroupement des profils de valeurs des mesures

Cette technique de réexpression des données pourrait ouvrir la porte à toutes sortes de manipulations graphiques respectant la philosophie d'exploration SOLAP. Par exemple, un forage sur les mesures pourrait détailler la représentation en passant d'un nombre réduit à un nombre plus grand de classes. Aussi, un forage sur les groupes de faits révélerait la distribution de ce groupe de faits dans les classes de chacune des mesures. Ainsi, une hiérarchie basée sur le regroupement des valeurs de mesures plutôt que sur le regroupement de membres de façon sémantique pourrait offrir une autre façon d'explorer les données augmentant ainsi la capacité d'analyse de l'approche SOLAP.

3.4 Synopsis

SOLAP est une approche interactive, dynamique et flexible réduisant la complexité de l'exploration des données géographiques. En introduisant la dimension géographique aux analyses multidimensionnelles, SOLAP doit composer avec la complexité de la représentation géographique des données pour supporter la découverte des connaissances. Dans ce chapitre, les caractéristiques des données, des requêtes et de la navigation propres au SOLAP ont d'abord été expliquées. Pour respecter la philosophie SOLAP, la stratégie de visualisation doit correspondre à une stratégie d'exploration interactive des données. La visualisation agit alors comme l'interface entre l'utilisateur et les données. Ceci engendre plusieurs besoins en matière de visualisation géographique. Pour répondre à ces besoins, quelques idées ont été présentées pour améliorer l'exploration des données par la visualisation.

CHAPITRE 4 - Proposition d'une méthode de définition de la visualisation géographique des données

Le chapitre précédent a soulevé plusieurs idées permettant de supporter une stratégie de visualisation géographique adaptée à l'environnement d'exploration interactive des données SOLAP. La concrétisation de chacune de ces idées dépasse largement les limites de ce mémoire. Un premier pas vers l'amélioration de la visualisation géographique avec SOLAP est réalisé dans ce chapitre. Celui-ci décrit une méthode d'intégration de l'expertise en géovisualisation à l'intérieur même de la technologie SOLAP. Cette méthode s'appuie sur deux éléments essentiels: des métadonnées de visualisation et une base de connaissances sémiologiques. La méthode proposée permet d'établir les bases de l'automatisation du design de cartes, de diagrammes et de tableaux adéquats tout en laissant une grande latitude à l'utilisateur.

4.1 La méthode

Comme exprimée à la section 3.3.1, l'automatisation de la visualisation géographique permettrait à l'utilisateur de se concentrer sur son analyse plutôt que sur la façon de représenter ces données. Ce bout de chemin permettrait à l'utilisateur de vraiment utiliser la visualisation comme un outil d'analyse en bénéficiant d'une grande flexibilité et d'une assistance intelligente. L'automatisation complète de la visualisation est bien sûr utopique, puisqu'une grande partie de la visualisation de l'information appartient à l'utilisateur c'est-à-dire ses préférences, ses limites, sa capacité de lire des diagrammes complexes, etc. Le but de la solution est donc de réduire l'intervention de l'utilisateur dans la construction des représentations par l'intégration d'une expertise en visualisation géographique tout en conservant une grande flexibilité dans la façon de visualiser les données géographiquement.

La solution proposée dans ce chapitre prend la forme d'une méthode d'intégration de l'expertise en visualisation géographique à l'intérieur même de l'approche SOLAP. Cette solution constitue la deuxième partie de la solution globale présentée dans ce mémoire (cf. Figure 1.3). Elle vise à renforcer le lien entre les données et la représentation de celles-ci sans l'intermédiaire de l'utilisateur. Ceci permettrait, lorsque pertinent, de fournir des diagrammes comparatifs,

agrégatifs et des images globales tout comme détaillées de l'information requête. Pour ce faire, une taxonomie des concepts permet de rassembler les renseignements pertinents et nécessaires pour le design d'information dans un contexte géodécisionnel. Ensuite, les éléments de la méthode sont présentés, c.-à-d. les métadonnées et la base de connaissances sémiologiques. La troisième partie de ce chapitre détaille la méthode d'intégration de l'expertise en géovisualisation proposée dans ce mémoire.

4.1.1 La taxonomie des concepts

La taxonomie des concepts permet de rassembler les renseignements pertinents au design d'information. Bien que les cartes, les tableaux et les diagrammes soient dynamiques et éphémères dans un environnement SOLAP, l'information qu'ils dévoilent doit être complète et vraie à chaque affichage. Les bonnes pratiques pour la représentation de l'information sont largement discutées dans la littérature [Few, 2006; Slocum et al, 2005, MacEachren, 1994a; Bertin, 1973; Tobelem-Zanin et Trémélo, 2003; etc.]. Dans chacune de ces références, de nombreux concepts sont énumérés et mis en relation. Cependant, il est rare de voir l'explication de ces concepts faite de façon générique pour aider à structurer l'intelligence qui supporte ces bonnes pratiques. La taxonomie présentée dans la présente section a pour but de structurer les concepts nécessaires afin de définir une intelligence générique pour n'importe quelle requête.

Les renseignements nécessaires à une bonne construction graphique se divisent en quatre groupes : les caractéristiques intrinsèques aux données, les caractéristiques de relation entre mesure et dimension, les caractéristiques d'affichage et les caractéristiques d'ensemble. Dans cette section, un résumé de ces concepts est présenté.

Caractéristiques intrinsèques aux données

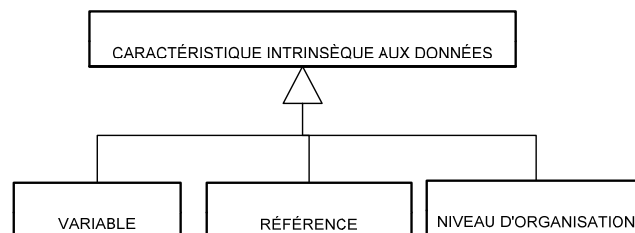


Figure 4.1 : Les caractéristiques intrinsèques aux données

Les caractéristiques intrinsèques aux données s'appliquent à tous les types de données présents dans un environnement SOLAP. Elles permettent de définir **la nature de l'information à représenter ou les variables de l'image**. Les caractéristiques se divisent en trois concepts : les variables, les références et le niveau d'organisation.

Variable

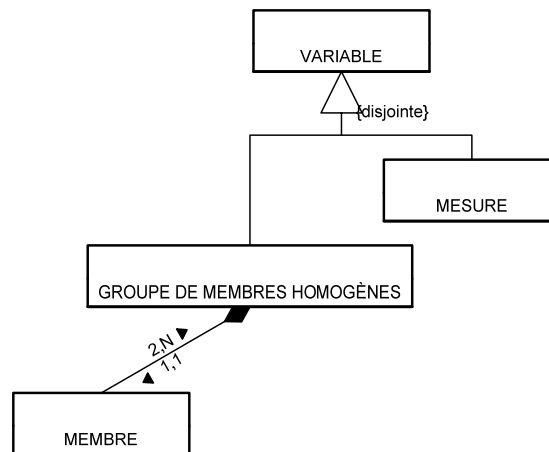


Figure 4.2 : Les variables de la structure multidimensionnelle

Dans un environnement géodécisionnel, les variables à représenter dans une image prennent le nom de mesures ou de membres. Leur combinaison (mesures + membres) permet de représenter les faits. Les **mesures** sont les données dépendantes et elles apparaissent généralement comme le sujet d'analyse et l'information de premier plan. Elles varient en fonction d'un ou de plusieurs groupes de membres.

Un **groupe de membres homogènes** représente la variable indépendante. Pour être considérée comme une variable, un groupe de membre doit varier. Une distinction importante doit être apportée ici entre le groupe de membres en tant que variable de l'image et les membres qui composent les faits. Dans un environnement géodécisionnel, un fait se compose du croisement d'un membre de chacune des dimensions et des mesures. Par conséquent, au moins un membre par dimension doit être sélectionné. Dans le cas d'une variable de représentation, si un seul membre est sélectionné dans la dimension, ce membre ne représente pas une variable, mais une constante. Ainsi, le groupe de membres en tant que variable de l'image doit se composer d'au moins deux membres distincts appartenant au même niveau hiérarchique de la même dimension.

Par exemple, à partir d'un cube de données où l'on retrouve une mesure (total des ventes de voitures) et trois dimensions (marque, lieu de vente et année), on veut connaître le total des ventes de voitures de marque Yaris, Aveo et Fit en 2008 au Québec. Le groupe de membres qui varient dans la requête se compose de trois membres (Yaris, Aveo et Fit) tandis que les membres des autres dimensions sont constants (lieu de vente= Québec et Année = 2008). Il n'y a donc qu'un seul groupe de membres dans cet exemple.

L'identification des variables dépendantes et indépendantes est à la base du design de visualisation. Ces informations permettent d'organiser l'information pour faire ressortir visuellement les relations et les structures qui existent dans les données. Ce concept est donc le premier en importance dans la définition des concepts de visualisation.

Référence

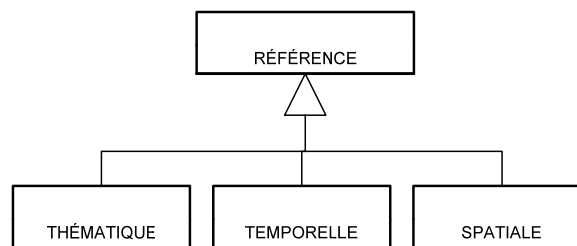


Figure 4.3 : Les références

Les données se caractérisent également par le second concept soit les références. Trois types de références ont été identifiés: **spatiale, temporelle et thématique**. Les données ayant une référence spatiale possèdent généralement une géométrie vectorielle et/ou matricielle. Au minimum pour être représentable cartographiquement, les données spatiales doivent comporter des coordonnées (directement ou indirectement). La notion de référence se rapporte aux concepts actuels du SOLAP. Les dimensions et les mesures peuvent être spatiales, temporelles ou thématiques (ex. un type de berge peut prendre une valeur telle que dune, marais, falaise, plage et talus).

Il est important de connaître la référence des variables à représenter. Ceci permettra de mieux choisir le modèle de représentation pour révéler les relations spatiales ou temporelles qui pourraient exister dans les faits à représenter. Par exemple, la carte n'est possible que si l'on inclut

un groupe de membres spatiaux dans la requête et l'animation est fortement recommandée dans le cas d'une variable indépendante temporelle.

Niveau d'organisation

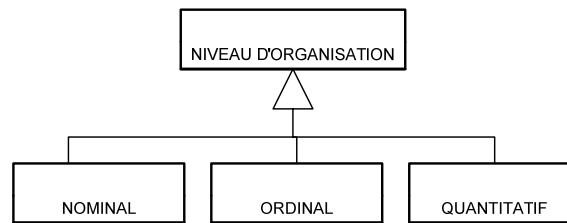


Figure 4.4 : Les niveaux d'organisation

Le niveau d'organisation des données est le point de départ du design d'information. Il permet d'identifier la nature première des données à représenter. Les données sont nominales, ordinales ou quantitatives.

Un niveau d'organisation **nominal** inclut des catégories non ordonnées. Les catégories sont, par définition, équidistantes les unes des autres et aucune n'est plus importante que les autres.

Un niveau d'organisation **ordinal** inclut toutes les catégories qui se définissent dans un ordre particulier. Ses catégories s'ordonnent d'une seule manière et sont équidistantes. Une échelle nominale devient ordinale lorsqu'on lui attribue un ordre particulier. Par exemple, si on ordonne des membres selon la valeur de la mesure associée, les membres s'ordonnent d'une seule façon et sont perçus comme ordinaux.

Le niveau **quantitatif** regroupe les catégories qui se mesurent par une série de nombres desquels on peut déduire une distance entre les catégories. Certaines distinctions peuvent être faites parmi les différents types de niveau quantitatif (intervalle versus ratio, absolu versus relatif, taux, indice, etc.). Ces différentes façons de classer les données quantitatives résultent souvent d'une transformation mathématique (ex. densité de population = population/superficie). La nécessité d'utiliser ces distinctions est souvent illustrée par l'exemple suivant. Une population est une donnée brute et doit être représentée par un symbole proportionnel alors que la carte choroplèthe doit représenter la densité de population pour minimiser l'impact visuel des grandes superficies. Cette règle est généralement acceptée par la communauté cartographique. Bien que

la densité de population est une donnée relative et que la population soit une donnée absolue, lorsque la donnée relative ne se rapporte pas à la superficie, est-ce tout aussi pertinent de faire le même choix? Voici un autre exemple d'une donnée relative versus une donnée absolue, le nombre de personnes sans emploi et le nombre de personne sans emploi/la population active. Le nombre de personnes sans emploi représenté par une carte en symbole proportionnel est un choix très valable. Cependant est-ce toujours aussi valable de représenter la proportion de la population active sans emploi par une carte choroplèthe « pour diminuer l'impact visuel des grandes superficies »? Si l'on transforme la carte en cartogramme de population, le choix est tout aussi justifié puisque la superficie égale la population. Dans l'autre cas, une interrogation reste. Dû à la complexité de gérer les subtilités des données relatives et leur référence, la distinction entre les différents types de niveaux d'organisation quantitatifs n'est pas prise en compte dans cette taxonomie. Des études supplémentaires pourraient pousser davantage cet aspect de la visualisation géographique dans un environnement décisionnel. Celles-ci pourraient cibler l'impact de chacun des types de données, établir une classification et définir leur utilisation.

Malgré que le concept « relatif » et « absolu » ne soit pas pris en compte ici, certaines transformations mathématiques peuvent être opérées à l'affichage automatique sur demande de l'utilisateur ou de l'administrateur. Par exemple, dans le cas d'une mesure de population, il peut être spécifié que la mesure doit être divisée par la superficie avant l'affichage de la carte, tout comme une proportion est automatiquement calculée avant l'affichage d'un camembert.

Caractéristiques de relation entre mesure et dimension

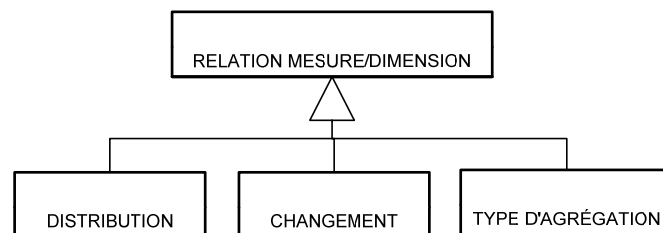


Figure 4.5 : Les caractéristiques de la relation entre les mesures et les dimensions

Le deuxième groupe de concepts détermine la relation entre les mesures et les dimensions (cf. Figure 4.25). Dans un environnement de données géodécisionnelles, lors de l'analyse, la ou les mesures varient selon un ou plusieurs groupes de membres (2 à N membres; les dimensions ne

possédant qu'un seul membre sélectionné sont des constantes et non des variables indépendantes). Une mesure peut varier de différentes façons selon un groupe de membres ou un autre. Ainsi, la relation entre chacune des mesures et chacun des groupes de membres doit être décrite. La relation entre les mesures et les groupes de membres prend trois formes: la distribution, le changement, et le type d'agrégation.

Distribution

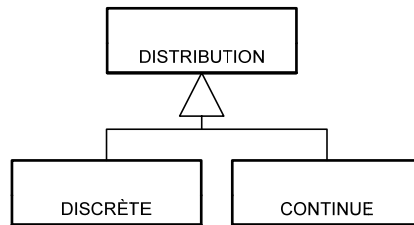


Figure 4.6 : La distribution

La distribution correspond à la façon dont les données couvrent une dimension, c'est-à-dire la distribution d'une mesure sur un groupe de membres. Par exemple, une mesure « température » pourrait être **discrète** dans un groupe de membres spatiaux si l'on considère les stations météorologiques ou **continue** si l'on considère des régions contiguës. La distribution est aussi valable pour une référence temporelle. Par exemple, un nombre d'accidents est discret dans le temps (en 2002, il y a eu 238 accidents tandis qu'en 2006, il y en a eu 195) et le risque d'avoir un accident varie de façon continue (le risque d'accident est passé de 3 à 6% 2002 et 2006). Finalement, la référence thématique peut également être décrite par une distribution. Par exemple, l'utilisation du sol possède un domaine de valeur fini et donc une distribution discrète tandis qu'une valeur marchande est continue, mais délimité par des bornes (ex. $0 < \text{valeur} < 1$ million).

Changement

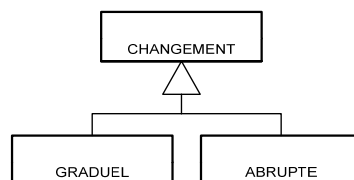


Figure 4.7 : Le changement

Le changement est la distance entre des valeurs contigües dans un groupe de membres. Un changement **abrupt** présente une coupure nette du phénomène entre deux membres. On peut facilement tracer la limite entre les catégories. Une carte d'utilisation du sol en est un bon exemple. Lors d'un changement **graduel** du phénomène, il est correct de supposer que toutes les valeurs sont possibles entre les points de mesure. La direction du vent et la température sur le territoire en sont un bon exemple. La limite entre les classes est floue.

Le changement et la distribution sont souvent intimement liés l'un à l'autre. Ces deux concepts permettent d'inférer visuellement de meilleures relations par rapport au couple mesure/membres. Ces concepts sont par contre difficiles à capturer puisqu'ils dépendent du niveau de granularité de l'analyse et sont limités par la géométrie implantée pour une dimension spatiale. Dans une application idéale où la géométrie des objets pourrait être transformée à la volée, il pourrait être intéressant d'offrir à l'utilisateur de basculer d'une représentation discrète à une représentation continue et vice et versa pour offrir davantage de flexibilité.

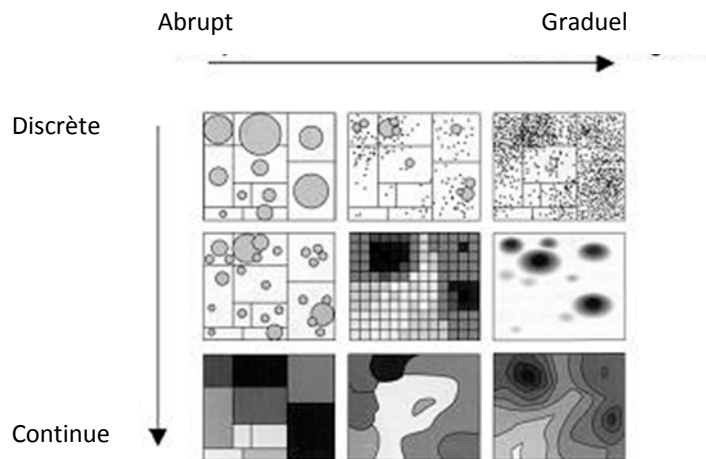


Figure 4.8 : Modèle de données selon le changement et la distribution. Traduit de [MacEachren, 1994a]

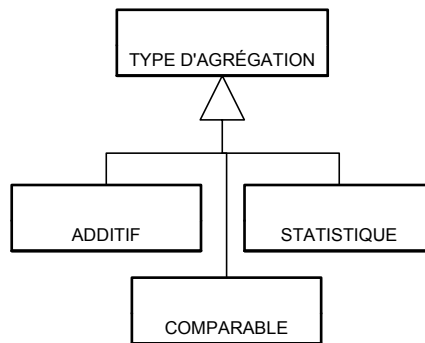
Type d'agrégation

Figure 4.9 : Le type d'agrégation

Le type d'agrégation se définit par la question suivante : selon ce groupe de membres, la valeur de la mesure s'agrège de façon : **additive, statistique** ou ne s'agrège pas visuellement (**comparable**). Une agrégation additive ou statistique indique qu'il est pertinent de faire l'agrégation visuellement des valeurs de la mesure dans ce groupe de membres pour des fins de comparaison (ex. les ventes de différents produits sur les jours de la semaine : Lundi -> 5k\$; mardi -> 2K\$, mercredi -> 1.5K\$, le début de la semaine = $5+2+1.5 = 8.5K\$$). Une relation comparable considère chaque valeur de mesure pour chaque membre comme indépendante (ex. la population des 5 dernières années : 2002 -> 2 000, 2003 -> 2500, 2004 -> 2200). Un type d'agrégation existe pour chaque association d'une mesure avec chacune des dimensions du cube. Par exemple, une mesure « population » peut s'additionner sur les régions du Québec, mais ne s'additionne pas sur la dimension temporelle. (Selon le vocabulaire multidimensionnel, la mesure est dite « additive » sur le premier groupe de membres et « non-additive » sur le second [Kimball et Ross, 2002]). Les relations additive et statistique présente des caractéristiques différentes. Chacune d'entre elles est dépendante de l'opérateur d'agrégation utilisé sur les dimensions. La relation additive correspond à l'opérateur somme, tandis que la relation statistique peut correspondre à des opérateurs tels que la moyenne, le minima/maxima, l'écart-type, la variance, etc. Visuellement, l'additivité permet de cumuler les informations tandis que la relation statistique implique de comparer le groupe à un chiffre particulier (ex. la moyenne du groupe) (cf. Figure 3.10).

Le type d'agrégation est important pour caractériser la façon dont les faits se regroupent entre eux. Il permet de mettre en évidence le cumul ou la position de chacun des faits par rapport aux

autres. Ce concept est souvent implicite pour le designer. Il révèle l'organisation des données entre elles.

D'autres caractéristiques d'agrégation pourraient être ajoutées à ce concept puisque l'agrégation ne se limite pas aux opérations mathématiques décrites ci-dessus. Cependant, celles présentées sont les plus fréquentes et couvrent les situations rencontrées lors des expérimentations exploratoires.

Caractéristiques graphiques

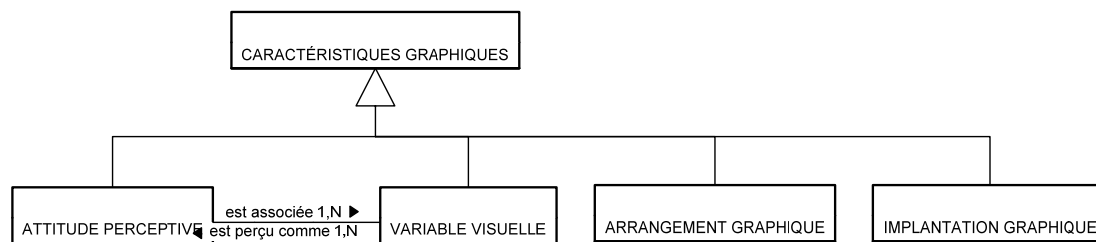


Figure 4.10 : Les caractéristiques graphiques

L'affichage des données est complexe puisque l'image nous donne plusieurs options pour encoder l'information. Avant d'aller plus loin dans la description des caractéristiques graphiques, une brève explication des composantes du langage cartographique est nécessaire. Premièrement, le langage cartographique utilise trois éléments de base, le point, le trait et la tache. Ces éléments se combinent pour former des figurés cartographiques. Les figurés portent une certaine signification dans l'image. Par exemple, le camembert est un figuré graphique qui se compose de taches de différentes couleurs et de différentes tailles. Le figuré est ensuite appliqué selon une implantation graphique. L'implantation peut être ponctuelle (carte de symbole), linéaire (carte de flux) ou zonale (carte choroplèthe). Les variables visuelles, maintenant, sont une caractéristique du figuré qui varie pour encoder une information. Une description plus détaillée de chacun de ces concepts peut être retrouvée dans [Béguin et Pumain, 2003, p. 41].

Les caractéristiques graphiques consistent donc à déterminer comment chacune des variables de la requête est encodée dans l'image. Parmi ces caractéristiques, on retrouve l'attitude perceptive, les variables visuelles, l'implantation graphique et l'arrangement graphique. Les trois premiers concepts s'appuient fortement sur la théorie de la sémiologie graphique pour représenter les

données. Ces concepts sont généralement bien connus des designers d'informations et leur application se produit systématiquement. Le dernier a été ajouté à la définition des caractéristiques graphiques pour définir explicitement les différents endroits dans l'image où les données peuvent être encodées. Ces choix sont habituellement implicites par le designer qui possède déjà en tête le modèle de représentation choisis pour ses données. Dans le cas présent, le modèle reste à définir. Ce concept est donc nécessaire à la définition des caractéristiques graphiques dans le but de déterminer un modèle de représentation.

Attitude perceptive

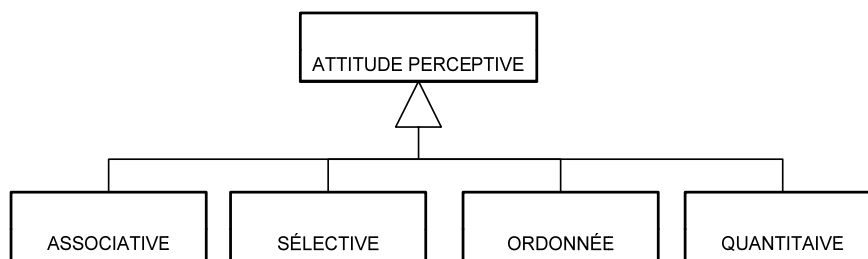


Figure 4.11 : Les attitudes perceptives

L'attitude perceptive correspond à l'information que l'on veut retirer de l'image qui illustre une variable. Par exemple, naturellement, l'utilisation de symboles de formes différentes nous permet de faire des associations entre différents types de phénomènes sur un territoire (ex. un marais (🌿) vs un terrain de camping (⛺) ou une école (🏫) vs une église (⛪) sur une carte topographique), mais la taille nous indique tout de suite un ordre ou une quantité (ex. Une carte de population en symbole proportionnel, plus le symbole est grand plus la population est grande). L'attitude perceptive peut être associative, sélective, ordonnée ou quantitative. Ce concept se base entièrement sur la théorie de Bertin. Les définitions de chacun de ces concepts sont détaillées à la section 2.4.2.

Les attitudes perceptives sont intimement liées au niveau d'organisation de la donnée. Idéalement, les données quantitatives devraient toujours être associées une attitude quantitative, les données ordinales à une attitude ordonnée et les données nominales à une attitude sélective ou associative selon la quantité d'information combinée dans l'image. Cependant, il n'est pas toujours possible de conserver ces associations intactes dans la présentation des données. Lorsqu'une limite est rencontrée, le principe de l'inclusion des attitudes perceptives s'applique (cf.

Figure 2.14). Par exemple, lorsque trois mesures quantitatives sont combinées dans une seule carte, il n'est pas possible d'utiliser la taille (une des variables qui permet l'attitude perceptive quantitative) pour représenter les trois mesures; un choix s'impose. On peut alors choisir de faire varier la valeur pour représenter l'une des variables et la taille en X et en Y pour faire varier les deux autres.

Les attitudes perceptives permettent de coordonner l'affichage à la nature même des données. Elles sont également intimement liées au concept de variables visuelles (cf. Figure 4.25). L'attitude perceptive permettra de sélectionner les variables visuelles les plus adaptées à représenter la donnée (cf. Figure 2.12).

Variable visuelle

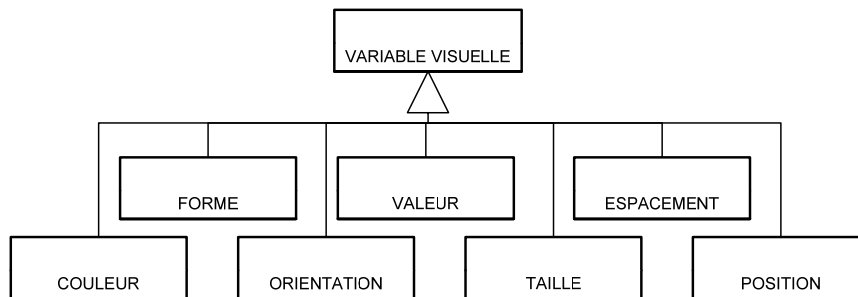


Figure 4.12 : Les variables visuelles

Les variables visuelles déterminent l'encodage de l'information sur le figuré. Elles représentent la caractéristique identifiée qui varie en fonction des données. Les variables visuelles identifiées sont les suivantes : couleur, forme, orientation, valeur, taille, espacement et position. La plupart des variables choisies correspondent aux variables de [Bertin, 1973] excepté la variable « espacement ». Cette dernière, identifiée dans quelques livres de cartographie [Béguin et Pumain, 2003; Slocum et al, 2005], rend, à mon avis, de plus grands services que le « grain » de Bertin. Ces différentes façons d'encoder l'information sont intimement liées avec les attitudes perceptives à rendre. Toutes les variables ne sont pas associées à toutes les attitudes perceptives. Par exemple, une couleur ne peut pas prendre une perception quantitative. Les règles décrivant les contraintes entre les variables visuelles et les attitudes perceptives sont décrites à la section 2.4.2.

Les variables visuelles ont un grand rôle à jouer principalement pour la représentation cartographique de l'information. Elles permettent d'encoder l'information supplémentaire dans le plan sans surcharger l'image. Elles se combinent pour encoder différentes informations (combinaison significative : ex. la taille encode la population et la couleur, l'origine dominante) ou pour augmenter la séparabilité visuelle des classes (combinaison redondante : ex. on fait varier la couleur et la forme ensemble pour mieux distinguer les classes de population). Les variables visuelles infèrent également une image globale et la détection automatique d'extrêmes. Bien utilisées, elles peuvent rendre de très grands services pour la découverte de connaissances.

Implantation graphique

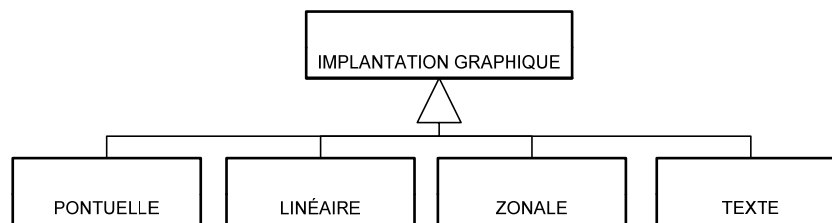


Figure 4.13 : Les implantations graphiques

L'implantation graphique du figuré détermine la forme que prend l'objet à représenter sur l'image. À l'instar des formes géométriques, l'implantation graphique peut être ponctuelle, linéaire ou zonale. L'implantation graphique « texte » a également été considérée puisque certaines données sont régulièrement affichées en texte (ex. les étiquettes sur une carte, les valeurs de la mesure dans un tableau ou les membres dans un histogramme). Il est à noter que l'implantation graphique correspond à l'application du figuré sur l'image. Ainsi, pour une structure de données matricielle, chaque cellule devient un petit polygone sur lequel on encode l'information de façon ponctuelle, linéaire ou zonale.

Le choix de l'implantation graphique sur une carte est souvent, au premier instinct, dépendant de la structure géométrique de la dimension géographique. Une correspondance exacte entre la géométrie et le type d'implantation graphique est souvent désirable, mais c'est aussi une pratique contraignante. Par exemple, la population des régions du Québec serait bien représentée par une couleur de remplissage dans les polygones des régions (la population est alors transformée en densité de population), mais peut aussi être représentée par des symboles proportionnels

(implantation graphique ponctuelle au centroïde des polygones régionaux). Également, lorsque plusieurs mesures doivent être affichées, il peut être utile de changer l'implantation graphique pour permettre l'encodage de plusieurs informations (par exemple, utiliser la couleur du polygone pour une mesure et la couleur d'un symbole pour une seconde mesure). Par conséquent, il doit toujours être possible de représenter les objets par plusieurs implantations graphiques indépendamment de leur géométrie (Figure 4.14). Par exemple, l'utilisation d'un symbole au centroïde d'un polygone (implantation ponctuelle d'une géométrie polygonale), l'utilisation de la bordure du polygone (implantation linéaire d'un polygone) ou même le trajet entre deux points (implantation linéaire d'une géométrie ponctuelle) sont différentes implantations qui ne correspondent pas à la géométrie de la dimension spatiale.

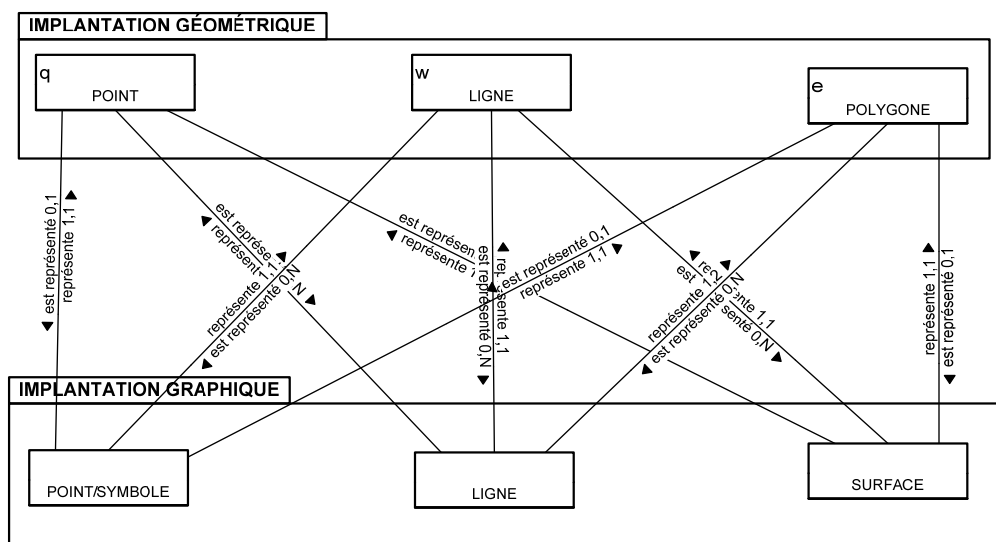


Figure 4.14 : Relation entre l'implantation graphique et l'implantation géométrique

Bien qu'en pratique, toutes les implantations graphiques soient possibles pour n'importe quelle géométrie, du point de vue de la visualisation de l'information, certaines associations sont moins souhaitables (ex. représenter un polygone par sa bordure ou représenter une ligne par les zones qu'elle croise). Cependant, d'un point de vue d'analyse spatiale, la possibilité de voir les trajets (implantation linéaire) entre des points (géométrie ponctuelle) peut être fort intéressante. Par exemple, sur la Figure 4.15, le lien entre l'origine et la destination et la qualité des livraisons qui transigent entre les deux est plus claire sur la carte en implantation linéaire.

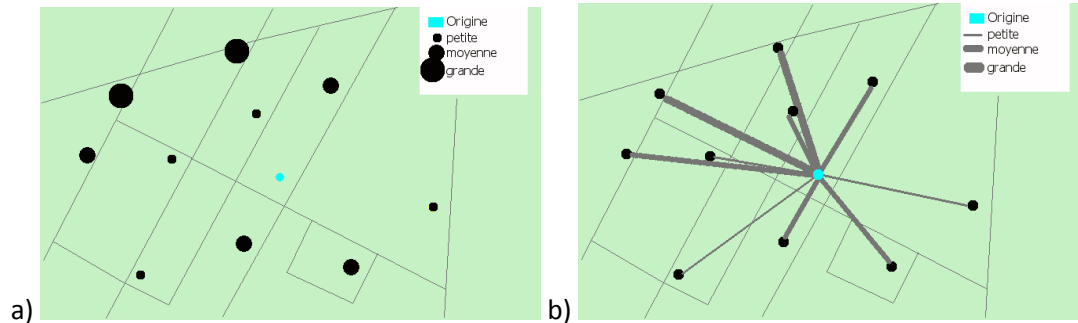


Figure 4.15 : Carte de livraison utilisant la taille comme variable visuelle a) en implantation ponctuelle et b) en implantation linéaire

La visualisation est toujours une histoire de compromis et de négociation entre les différentes informations à représenter et ce qu'on veut voir. L'implantation graphique peut être un outil supplémentaire pour faire ressortir certaines associations dans les données. En se tenant uniquement au point de vue de la visualisation géographique, les associations souhaitables entre une géométrie et son implantation se limitent à une implantation graphique de dimension égale ou inférieure (0D, 1D, 2D). Ainsi, l'utilisation d'une implantation graphique de dimension inférieure peut permettre d'encoder davantage d'information sur une même carte sans diminuer visuellement l'information encodée. Par exemple, une carte choroplèthe peut avoir un symbole proportionnel superposé pour ajouter une information supplémentaire.

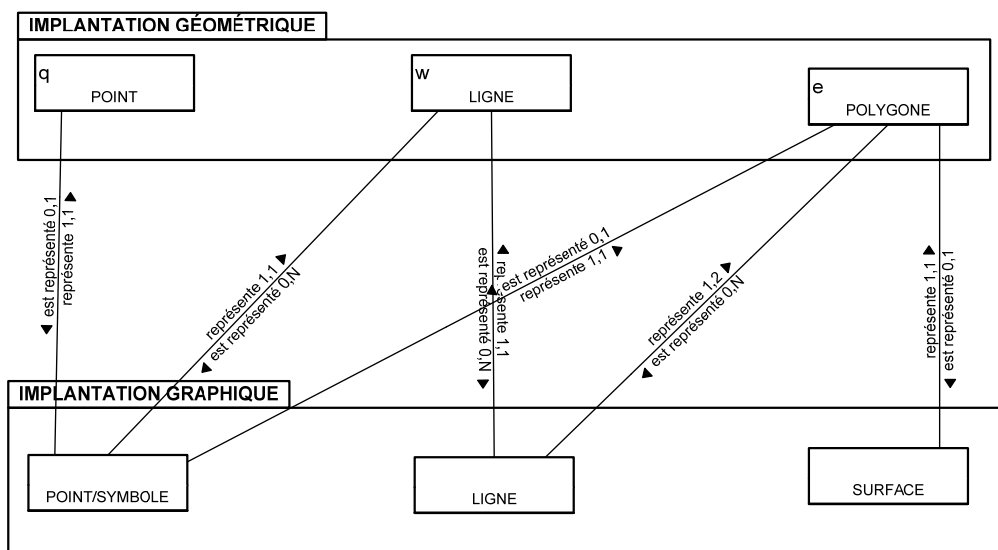


Figure 4.16 : Relation entre l'implantation graphique et la géométrie d'un point de vue de visualisation de l'information

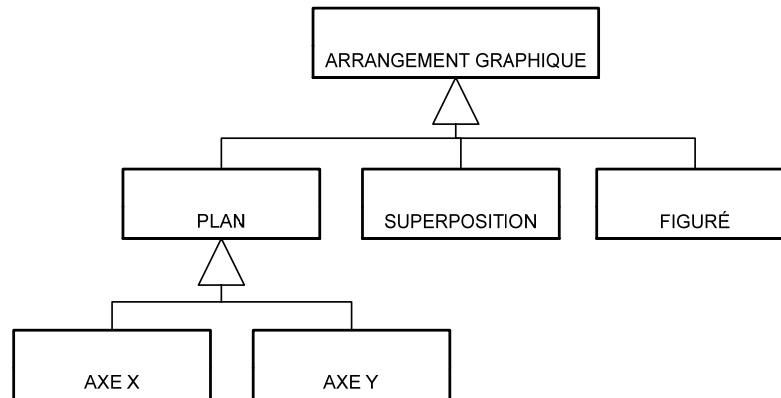
Arrangement graphique

Figure 4.17 : Les arrangements graphiques

L'arrangement graphique est un concept qui est discuté implicitement dans la littérature pour définir un modèle de représentation (ex. camembert, un histogramme, une carte). Dans le but de développer une méthode générique, ce concept a été inclut parmi les caractéristique graphique afin de déterminer à quel endroit dans le plan de l'image, les variables de l'information sont encodées. Les concepts découlant de l'arrangement graphique sont les axes X et Y, la superposition et le figuré (Figure 4.18). Ceux-ci définissent tous les endroits où il est possible de faire varier une caractéristique. Premièrement, l'information peut varier selon l'axe horizontal ou l'axe vertical du plan. Les axes X et Y ont été regroupés sous le concept « plan » puisqu'une seule variable peut monopoliser les deux axes du plan simultanément (ex. les membres spatiaux dans une carte). La superposition est un troisième endroit où l'information peut être encodée. La variable se place alors sur un axe transversal au plan de l'écran. Le résultat de la superposition prendre la forme d'une animation séquentielle ou une superposition d'images (ex. un histogramme superposé). Finalement, le figuré est l'endroit où les variables visuelles s'appliquent. Il se constitue des caractéristiques visuelles de l'objet.

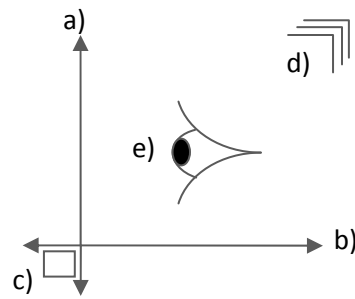


Figure 4.18 : Définition des arrangements graphiques a) Axe Y, b) Axe X, c) Plan, d) Superposition et e) le figuré

L'arrangement graphique est un concept un peu abstrait, mais très utile dans la définition des modèles de représentation. Par exemple, un histogramme à barres verticales est défini par les valeurs de la mesure sur l'axe Y et les membres (variable indépendante) sur l'axe X. À l'inverse, un histogramme à barre horizontale place les valeurs de la mesure sur l'axe X et les membres sur l'axe Y. Elle permet également de différencier une carte multivariée et une multicarte (Figure 4.19). La carte multivariée comporte plusieurs variables placées sur le figuré (ex. le nombre d'inscrit -> couleur du polygone et la cote confidentielle -> taille du symbole) et une seule dans le plan (les régions du Québec). La multicarte, à l'inverse, comporte une seule variable sur le figuré (le nombre d'inscrit OU la cote confidentielle -> couleur du polygone) et les autres sur le plan (les régions du Québec), en X et/ou en Y (les mesures).

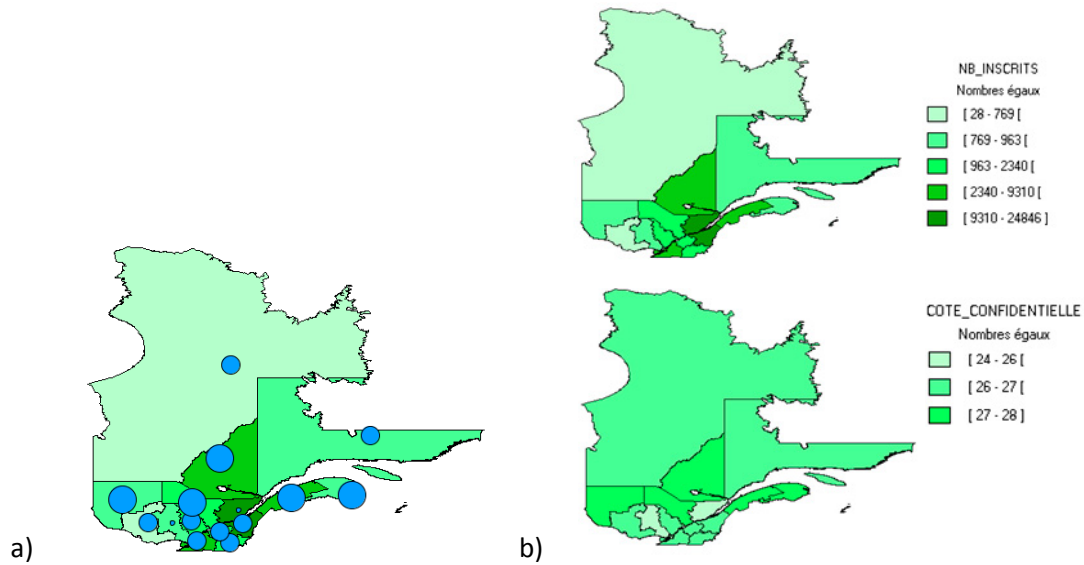


Figure 4.19 : Arrangement graphique a) d'une carte multivariée et b) d'une multicarte

Caractéristiques d'ensemble

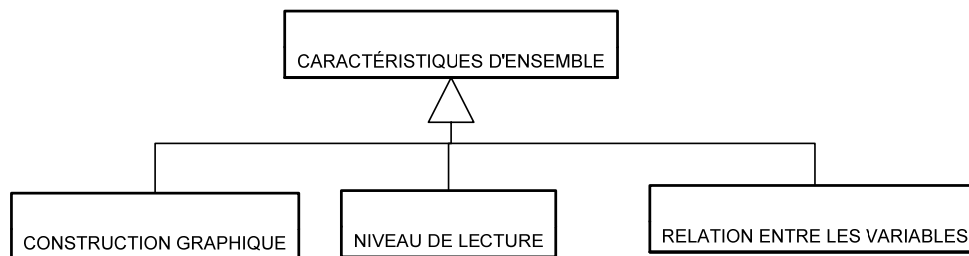


Figure 4.20 : Les caractéristiques d'ensemble

Pour arriver à représenter un groupe de faits, il est nécessaire d'avoir au minimum une mesure et un groupe de membres (requête simple, cf. section 3.1.2) et un membre constant pour les autres dimensions¹⁴. Jusqu'ici les concepts présentés décrivent les caractéristiques de la représentation d'une seule variable (caractéristiques intrinsèques et caractéristiques graphiques) ou d'une association simple d'une variable dépendante à une variable indépendante (caractéristiques de

¹⁴ Les sélections uniques dans chacune des dimensions représentent des membres constants de la requête. Ces membres déterminent le contexte. Par exemple, une requête comprenant les membres suivant : année 2008, sexe féminin, tous les types de maladies, 5 régions du Québec et la mesure nombre de personnes affectées, la variable indépendante est : les régions du Québec, la variable dépendante : le nombre de personnes affectées et le contexte est déterminé par les membres 2008 et femmes pour toutes les maladies. Les membres constants n'entre pas dans la définition des modèles de représentation.

relation). Cependant, dans un environnement multidimensionnel, il n'est pas rare de regrouper plusieurs variables indépendantes (requête avec multi-sélections dans plusieurs dimensions, cf. section 3.1.2) et/ou dépendantes (requête multivariée, cf. section 3.1.2) dans la même requête pour afficher le résultat. Lorsque l'on regroupe toutes ces variables pour former l'image, des caractéristiques supplémentaires doivent être documentées. Ces caractéristiques sont décrites par les caractéristiques d'ensemble.

Construction graphique

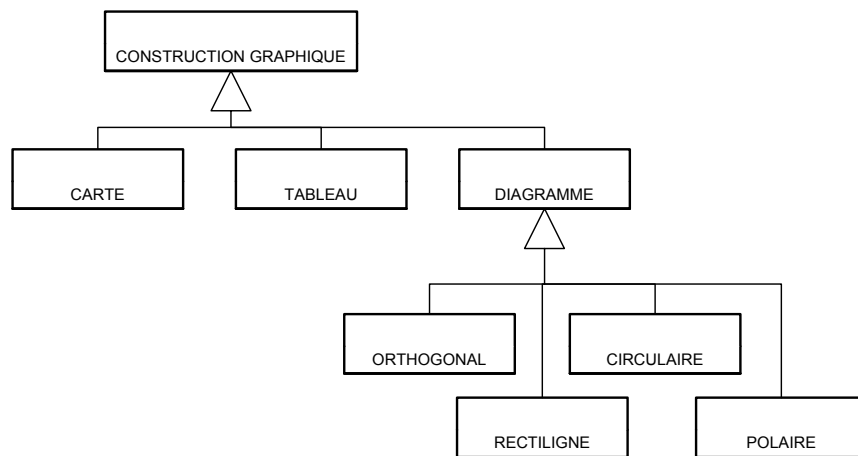


Figure 4.21 : Les constructions graphiques

Les constructions graphiques regroupent les différents types de représentation possibles de l'information : les cartes, les tableaux et les diagrammes. Les diagrammes se divisent en quatre catégories : orthogonal, rectiligne, circulaire et polaire (Figure 4.22). Ces catégories décrivent la forme générale et peuvent inclure plusieurs types de diagrammes. Par exemple, parmi les diagrammes orthogonaux, on retrouve le diagramme à ligne, le diagramme à barres horizontales et verticales, le nuage de points, etc. Les catégories regroupent les diagrammes selon l'orientation de leurs axes [Bertin, 1973].



Figure 4.22 : Exemples de diagrammes a) orthogonal, b) rectiligne, c) circulaire et d) polaire

Les constructions graphiques offrent des formats prédéterminés dans lesquels les variables seront représentées. Différentes variantes permettent d'encoder plus ou moins d'information tout en respectant le format de base. Par exemple, une carte demande au minimum une variable dépendante et une variable indépendante spatiale (le nombre d'inscrit et les régions du Québec). Cependant, on peut ajouter un autre groupe de membres (le sexe, femme et homme) et former une multicarte ou une autre mesure (le nombre d'inscrit et la cote confidentielle) et former une carte multivariée. Les constructions graphiques permettent à l'utilisateur de percevoir la même information sous différentes formes lui permettant de découvrir différentes structures dans les données. Par exemple, le nombre d'inscrits à l'université selon les sexes et les régions du Québec peut être vu selon un histogramme, un camembert, une barre superposée à une carte, en multicarte ou encore à l'aide d'un histogramme cumulé.

Relation entre les variables de même type

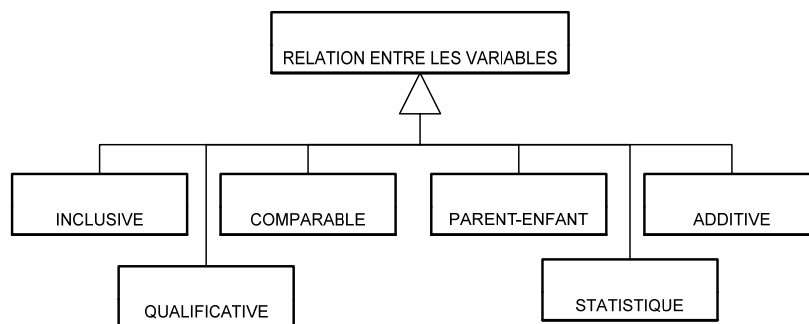


Figure 4.23 : Les relations entre les variables de même type

La relation entre les variables de même type décrit le lien entre les différentes mesures principalement, mais également entre les différents groupes de membres homogènes. La liste des relations identifiées est la suivante : inclusive, qualificative, comparable, statistique, additive et parent-enfant. La relation inclusive implique qu'une mesure soit incluse dans l'autre (ex. le nombre de naissance et la population). Une mesure peut également identifier la qualité d'une autre mesure (ex. le taux d'incertitude et le nombre de malades prévus dans un scénario d'épidémie). Deux mesures reliées statistiquement ont une relation statistique (ex. la vitesse moyenne et l'écart type). La relation additive indique qu'il est cohérent de cumuler visuellement les deux mesures (ex. le nombre de morts et le nombre de blessés). La relation comparable signifie que l'ordre de grandeur des variables associées est comparable et qu'il est justifier de les

comparer parallèlement (ex. le nombre de décès et le nombre de naissances). Dans le cas d'une relation entre deux groupes de membres, seules les relations comparables et parent-enfant sont possibles. De plus, elles sont complètes et mutuellement exclusives; c.-à-d. que la relation est soit comparable (ex. les sexes et les régions), soit parent-enfant (ex. les équipes et les individus de ces équipes). Dans le cas de la relation parent-enfant, l'opérateur d'agrégation correspondant (ex. la moyenne ou la somme) déterminera l'organisation visuelle que ces deux groupes de membres peuvent avoir entre eux.

Le choix de ces relations découle des nombreux tests effectués sur les données (cf. Figure 1.5). Elles sont le reflet des situations rencontrées lors des expérimentations exploratoires. La liste n'est pas exhaustive et il est possible que d'autres relations doivent être ajoutées. Cependant, dans le cadre de cette maîtrise, elle couvre l'ensemble des relations rencontrées dans les données.

La description de ces relations est nécessaire pour connaître quelles relations peuvent être mise de l'avant dans la représentation des données et obtenir la plus grande expressivité possible. Les relations entre les différentes mesures permettent de choisir les combinaisons significatives de variables visuelles les plus appropriées. Par exemple, lorsqu'une mesure qualifie une autre, il est pertinent d'utiliser une variable de valeur pour la mesure qualificative afin de diminuer l'impact visuel des données de moins bonne qualité. La relation entre les différents groupes de membres permet d'obtenir des diagrammes statistiques lorsque c'est pertinent (relation parent-enfant utilisant la moyenne comme opérateur d'agrégation); par exemple, une symbologie utilisant deux couleurs pour indiquer les valeurs au-dessus et en dessous de la moyenne.

Niveau de lecture

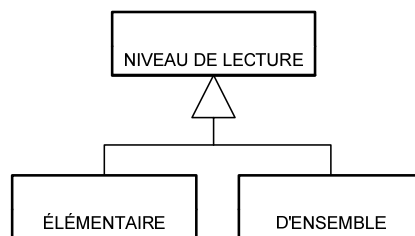


Figure 4.24 : Les niveaux de lecture

Le niveau de lecture établit la différence entre une carte à voir et une carte à lire [Bertin, 1973]. La carte à voir permet d'abstraire certains détails pour faire émerger une meilleure vision d'ensemble [Tufté, 2006]. Quand plusieurs informations sont combinées dans la même image, il devient parfois difficile de distinguer une tendance générale sans une analyse approfondie de la représentation. Passer d'un niveau de lecture élémentaire à un niveau d'ensemble permet alors de négliger visuellement les détails tout en les gardant accessibles au besoin. L'abstraction peut se faire de trois façons : par généralisation, par interprétation et par regroupement des profils de mesure (cf. Figure 3.13). La première façon concerne généralement la représentation cartographique des données; on doit pouvoir s'abstraire de certains détails de localisation pour percevoir la tendance (ex. la concentration de données à certains endroits par rapport à l'ensemble du territoire crée une superposition qui empêche d'apprécier la tendance thématique). La deuxième concerne l'identification automatique des zones possédant des caractéristiques particulières (homogènes, extrêmes, semblables, etc.). Finalement, la troisième produit une abstraction sur les valeurs de mesures en regroupant les profils semblables (cf. section 3.3.3). L'intégration de ce concept au SOLAP nécessite plusieurs avancées dans différents domaines, notamment des avancées en généralisation cartographique dans un environnement géodécisionnel pour permettre l'éclaircissement de l'information thématique dans les situations cartographiques conflictuelles et en *data mining* géographique pour repérer automatiquement les zones de ressemblances ou de différences. À cause de la complexité de ce concept, celui-ci n'est pas intégré pour la suite de la méthode proposée.

Synthèse

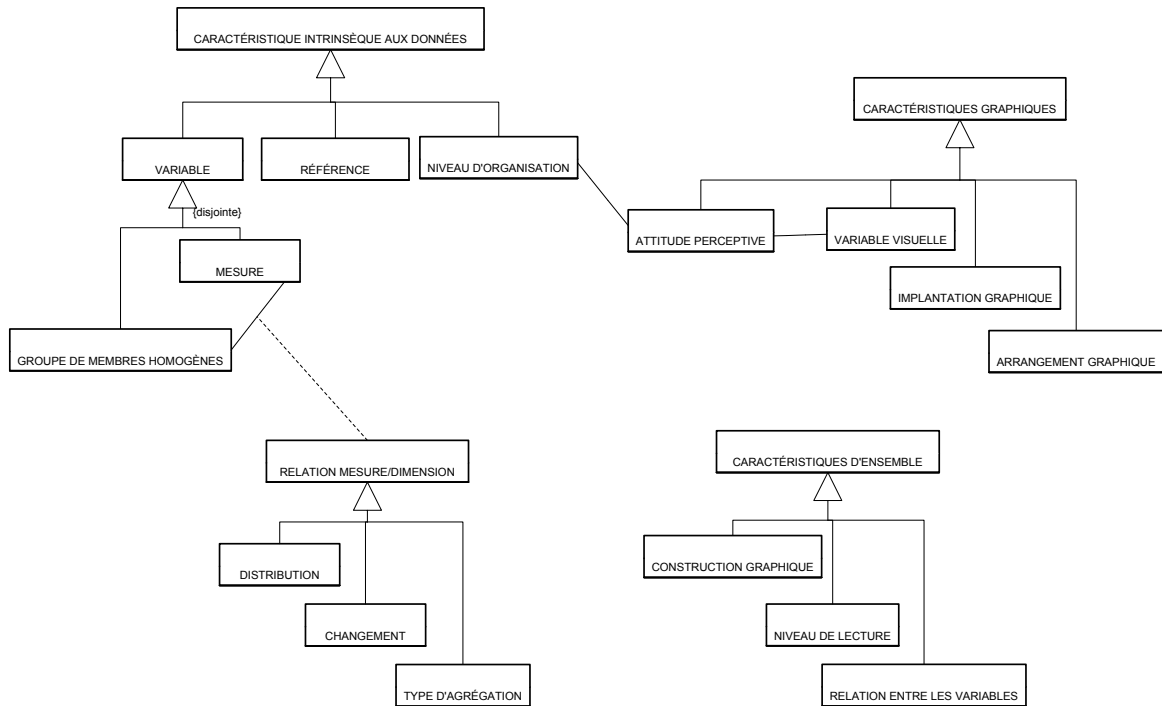


Figure 4.25 : Synthèse de la taxonomie des concepts

La grande quantité de concepts présentés dans la présente taxonomie indique bien la quantité d'information traitée par un designer ou un cartographe lors du design d'information¹⁵. C'est pourquoi l'intelligence cartographique est un défi à reproduire automatiquement. Chacun de ces concepts possède leur importance dans une situation ou une autre. Par exemple, les variables visuelles permettent l'ajout d'information sur l'image à l'aide de la dimension visuelle. On peut alternativement faire varier l'implantation graphique pour augmenter la quantité d'information représentée sans nécessairement surcharger l'image. Il est également possible d'utiliser les axes du plan pour établir des comparaisons directes. Le type d'agrégation et les relations entre les variables permettent de pousser un peu plus loin l'intelligence en visualisation géographique pour rendre visuellement les liens entre les données lorsqu'ils sont connus. Chacun de ces concepts interagit avec les autres pour déterminer les méthodes de visualisation appropriées pour chacune des requêtes. Par exemple, le niveau d'organisation est directement lié avec l'attitude perceptive

¹⁵ La taxonomie complète est présentée à l'Annexe 1.

qui influence le choix des variables visuelles. En même temps, chaque modèle de représentation est spécifique et ne s'applique que dans certains cas (ex. un camembert représente des proportions). Le design de l'information est une science de compromis et de négociation entre ces différents concepts. On regarde premièrement les données, leurs caractéristiques intrinsèques et leurs relations. Ensuite, on peut jouer avec les différentes caractéristiques graphiques et d'ensemble pour déterminer les modèles de représentation adéquats. Dans un but de simplification de la construction de la visualisation géographique avec SOLAP, les premiers concepts d'importance du côté graphique sont les variables visuelles, le niveau d'organisation et l'implantation graphique. Cependant, les autres concepts apportent de la précision, de la justesse et de la flexibilité à l'utilisateur.

La sélection des différents concepts a été réalisée au cours des nombreuses itérations qui ont mené à cette solution. L'influence de chacun des concepts sur un modèle de visualisation sera explicitée lors des expérimentations exploratoires à la section 4.2.2. La taxonomie présentée dans cette section constitue une boîte à outils. Il suffit ensuite de les utiliser adéquatement pour construire les modèles de visualisation.

4.1.2 L'intégration des éléments de la méthode : les métadonnées et la base de connaissances sémiologiques

Les concepts annoncés par la taxonomie de la section précédente n'ont rien d'innovateur. Les cartographes et designers d'information regroupent la plupart de ces renseignements afin de produire les représentations de données les plus expressives et efficaces possibles dans chacun des cas. L'exercice habituellement exécuté par un expert en visualisation doit être transféré dans le client SOLAP d'une certaine façon. Le design d'information et la cartographie sont, certes, à la fois un art et une science qu'aucune machine n'arrivera à égaler. Par contre, cette science possède aussi une certaine redondance qu'il est possible de reproduire, car dans plusieurs cas, les mêmes types d'information mèneront aux mêmes designs. C'est cette redondance que l'on cherche à exploiter pour l'automatisation de la visualisation géographique avec SOLAP. À partir des informations sur les données (ex. niveau d'organisation, référence et relation mesure/dimension),

un système devrait être en mesure de trouver une grande partie des solutions graphiques documentées pour cette situation.

Pour intégrer l'intelligence cartographique, la taxonomie présentée à la section précédente était une étape importante. Elle fait l'inventaire des caractéristiques qui décrivent les données. Pour le développement de l'automatisation de la visualisation géographique, plusieurs options sont possibles. Parmi elles, une série d'affichages par défaut peuvent être définis dans le client SOLAP. Cette option contraint par contre l'utilisateur à se servir de ces affichages par défaut avant de personnaliser son choix de visualisation s'il y a lieu. Cette pratique permet une intégration rapide et facile des bonnes pratiques de visualisation, mais elle limite, par le fait même, les choix offerts à l'utilisateur, oblige le passage par une représentation par défaut et ne traite pas a priori plusieurs des relations plus subtiles dans les données. Elle réduit alors la flexibilité d'analyse recherchée dans un environnement d'exploration interactive de données géodécisionnelles.

L'idée principale qui soutient la solution de la présente recherche est d'intégrer une base de connaissances sémiologiques à l'approche SOLAP. Cette base de connaissances contient la description générique des modèles de représentation supportés et les fournit au client SOLAP (Figure 4.26 A). Son peuplement relève du cartographe ou du designer d'information (Figure 4.26 B). Pour permettre le lien entre les données du cube et la base de connaissances sémiologiques, des métadonnées de visualisation doivent également être intégrées à l'approche SOLAP (Figure 4.26 C). La responsabilité de celles-ci relève de l'administrateur de données. Ces métadonnées décrivent la nature des données de l'application tandis que la base de connaissances sémiologiques contient la description générique des modèles de visualisation et des données pour lesquelles ces modèles sont adaptés. Les nouveaux éléments ajoutés sont présentés en pointillé dans le cas d'utilisation décrivant l'approche SOLAP (Figure 4.26).

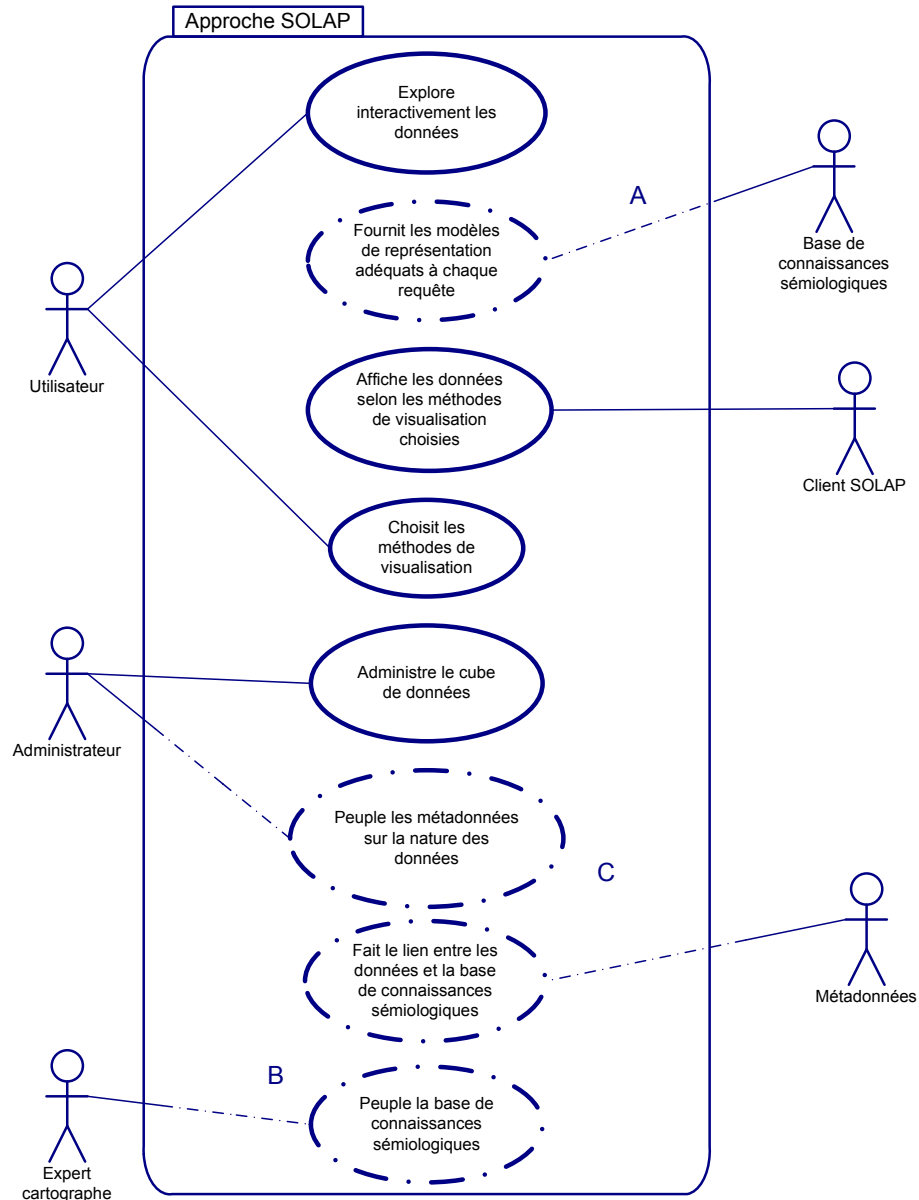


Figure 4.26 : Cas d'utilisation de l'approche SOLAP proposée

La base de connaissances sémiologiques organise les concepts, présentés à la section précédente, de façon à décrire les modèles de représentation et les données pour lesquelles ils sont adaptés. Ces modèles sont décrits d'une façon générique. Différents modèles posséderont différentes combinaisons de concepts. Par exemple, une mesure (type de variable) quantitative (niveau d'organisation) thématique (référence) est bien représentée par une variable ordonnée (attitude perceptive) comme la couleur (variable visuelle) en implantation zonale (implantation graphique) sur le figuré (arrangement graphique). Cette combinaison décrit une carte choroplèthe si cette

mesure est combinée à un groupe de membres spatiaux (Figure 4.27a). Similairement, un groupe de membres (type de variable) temporels (référence) ordinaux (niveau d'organisation) prend une perception ordonnée (attitude perceptive) par une variable position (variable visuelle) qui varie sur l'axe des X (arrangement graphique) de façon linéaire (implantation graphique). On trouve alors un diagramme à ligne si l'on place la mesure associée sur l'axe Y (Figure 4.27b). Ces combinaisons de concepts pour la représentation d'une variable rappellent la structure des bases de données multidimensionnelles dans laquelle chacun des membres est croisé avec les membres des autres dimensions (c'est pourquoi la base de connaissances est représentée par un cube dans la Figure 4.28). Sa structure, son fonctionnement et son rôle sont explicités à la section suivante.

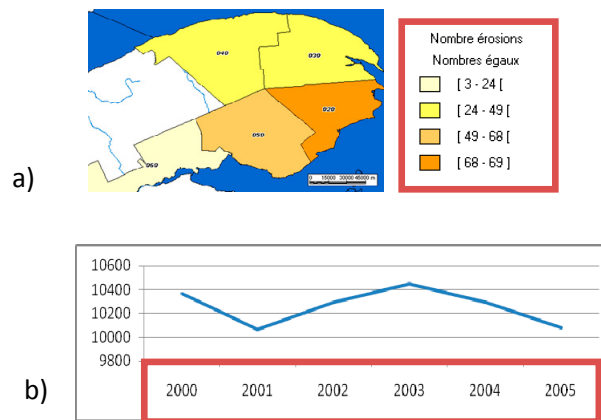


Figure 4.27 : Exemple de modèles de représentation et de la description de leur combinaison a) une mesure quantitative thématique sur une carte choroplèthe et b) un groupe de membres temporels dans un diagramme à lignes.

Ensuite, pour permettre la correspondance des données de chacune des applications avec la base de connaissances sémiologiques générique, des métadonnées sur la nature même des données doivent être ajoutées aux cubes pour déterminer les caractéristiques et les limites de ces données. Les métadonnées à ajouter incluent, les caractéristiques intrinsèques aux données (le type de variables, la référence et le niveau d'organisation des données) ainsi que les relations entre chacune des mesures et des dimensions (le type d'agrégation, la distribution et le changement) et une caractéristique d'ensemble (les relations entre les variables de même type). Le peuplement des métadonnées relève généralement du concepteur du cube de données, bien que certaines informations puissent être déduites par le système. SOLAP permettant non seulement d'explorer les données d'un cube, mais de les visualiser sous différentes formes, le concepteur d'un cube doit prévoir dans sa modélisation l'intégration des métadonnées de visualisation.

L'intégration de ces nouveaux éléments à l'approche SOLAP permet une collaboration plus serrée entre les données et leur visualisation. Le diagramme UML de collaboration (Figure 4.28) montre les échanges imaginés pour l'intégration de la solution. L'utilisateur saisit une requête (1). L'application SOLAP filtre les données du cube (2) et reçoit les métadonnées associées aux données de la requête (3). Les métadonnées permettent de filtrer la base de connaissances (4) pour ne retirer que les modèles de représentation adéquats (5). Ceux-ci sont proposés à l'utilisateur (6) qui choisit (7) et reçoit l'affichage demandé (8). Contrairement à une application qui définit les affichages par défaut, l'utilisateur SOLAP obtient alors plusieurs choix au bout des doigts et peut donc basculer facilement d'une représentation à une autre pour visualiser ses données.

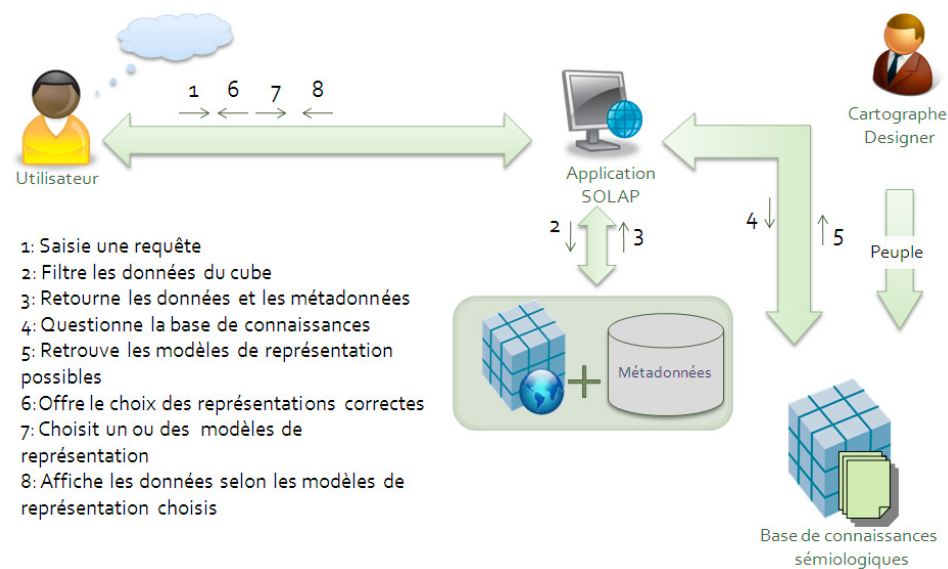


Figure 4.28 : Diagramme UML de collaboration

L'introduction de métadonnées et d'une base de connaissances sémiologiques dégage l'utilisateur et l'administrateur de la responsabilité d'un affichage vrai et complet des données. En d'autres termes, aucun affichage erroné sur le plan sémiologique ne sera proposé à l'utilisateur, ce qui est une amélioration importante en comparaison à la situation actuelle où l'utilisateur peut produire n'importe quel affichage sans même savoir s'il est approprié. Il suffit à l'application SOLAP de faire correspondre les métadonnées aux connaissances sémiologiques pour déterminer la ou les représentations s'appliquant à la requête.

4.1.3 La base de connaissances sémiologiques

Son rôle

Le rôle de la base de connaissances sémiologiques s'étend sur deux volets. Elle documente tous les modèles de représentations applicables dans un environnement SOLAP pour offrir une flexibilité de visualisation encadrée à l'utilisateur. Elle guide aussi les modifications possibles à la symbologie.

L'idée derrière la base de connaissances sémiologiques est avant tout de documenter les différents modèles de représentation selon la combinaison minimale des concepts nécessaires à leur construction. Par exemple, une carte thématique nécessite un groupe de membres spatiaux combinés à une mesure thématique quantitative (ex. la population dans les régions administratives). Ceci permet de décrire la correspondance parfaite entre les données et les modèles de représentation (ex. une carte avec des symboles proportionnels). Une fois les modes de représentation décrits, il suffit alors de connecter les propriétés des données aux informations requises pour obtenir tous les modes de représentations possibles pour une requête. Tout comme dans le cas des bases de données multidimensionnelles, l'information stockée dans la base de connaissances n'a qu'à être filtrée selon les métadonnées accompagnant la requête pour rendre disponible à l'utilisateur les modèles de représentations appropriés (Figure 4.29). Contrairement aux applications où les options d'affichage sont définies par défaut, l'utilisateur peut alors choisir un ou plusieurs modèles de représentation, l'afficher et basculer d'un modèle à un autre pour percevoir différentes relations et ceci, au bout des doigts.

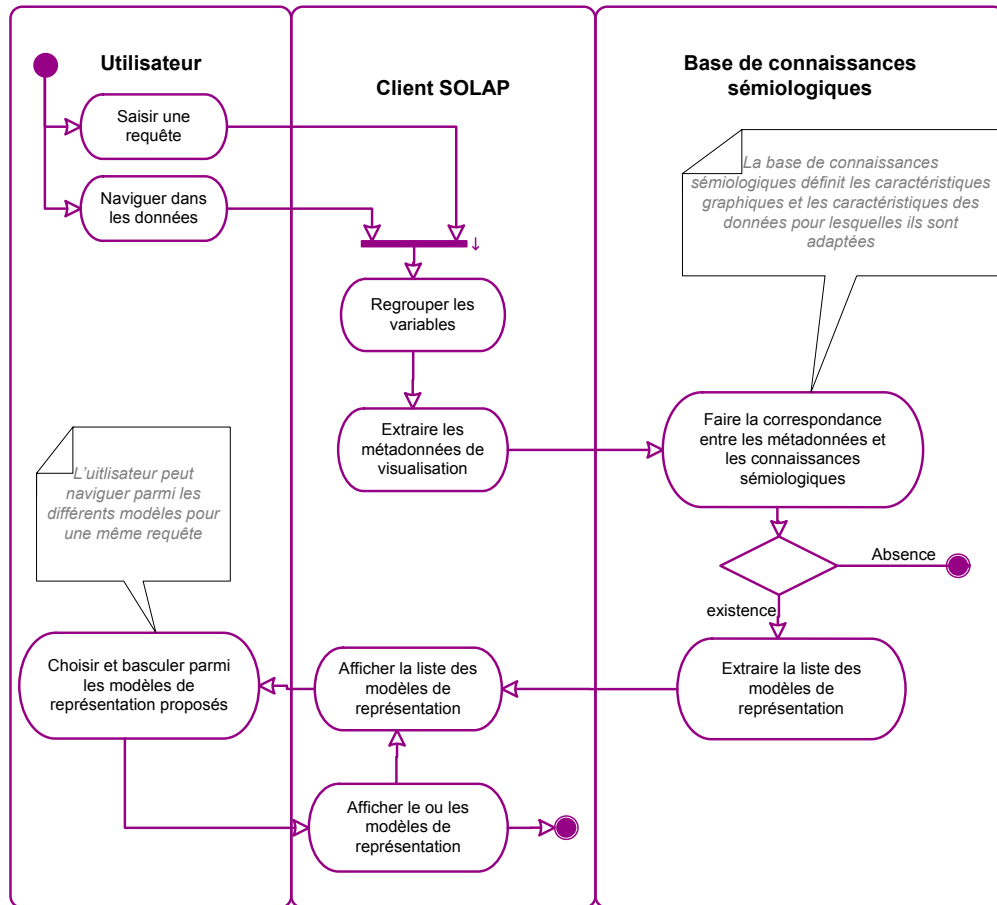


Figure 4.29 : Diagramme d'activité lorsqu'une nouvelle requête est exécutée

La base de connaissance permet également de guider l'utilisateur pour la modification de la symbologie. Dans un SOLAP idéal, l'utilisateur devrait avoir la possibilité de modifier la symbologie à sa guise. Cependant, pour éviter que des erreurs surviennent lors de ces manipulations, la base de connaissances sémiologiques doit encadrer les choix disponibles au niveau de la symbologie. Les différents modèles de représentation stockés dans la base de connaissances permettent donc de limiter les options à celles adaptées aux données (Figure 4.30).

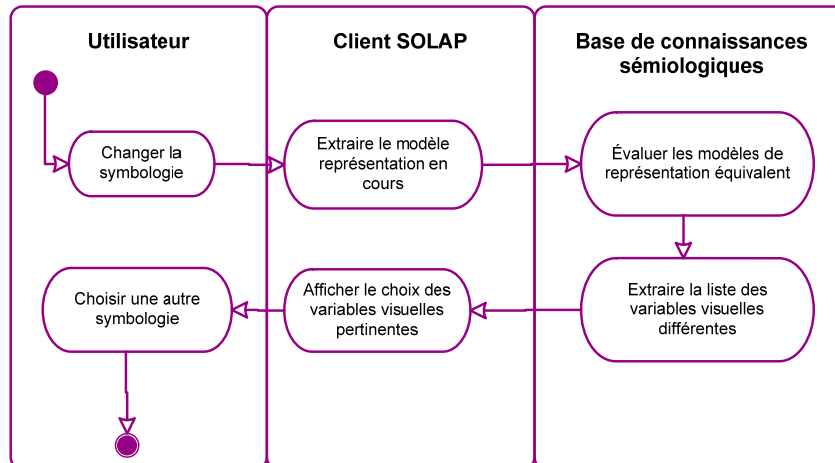


Figure 4.30 : Diagramme d'activité lorsqu'un changement est apporté à la symbologie par l'utilisateur

L'intégration de la base de connaissances sémiologiques dans un environnement SOLAP permet d'utiliser de façon transparente les règles de sémiologies graphiques au sens large (c.-à-d. en incluant les règles de construction graphiques et non seulement les règles régissant l'utilisation des variables visuelles). L'action de l'utilisateur se limite alors à requêter, choisir et naviguer parmi les options offertes. Des changements peuvent alors être rapidement et facilement apportés à la représentation des données tout en restant dans un cadre sémiologique approprié pour l'analyse en cours. De plus, plus la base de connaissances sémiologiques sera enrichie, plus les options offertes à l'utilisateur seront intelligentes et lui permettront de voir les données sous des formes qu'il n'aurait même pas pensé lui-même (ex. la possibilité de voir des « *sparklines* » sur une carte ou dans un tableau pour l'évolution temporelle ou spatio-temporelle [Tufté, 2006]). L'intégration de la base de connaissance sémiologique ajoute donc l'intelligence du « comment représenter les données » avec toutes les options qu'elle comporte directement dans l'application SOLAP.

Sa conception

La base de connaissances sémiologique organise les concepts présentés à la section précédente (section 4.1.1) de façon à décrire les modèles de représentation. La Figure 4.31 illustre le modèle conceptuel de la base de connaissances sémiologiques. Le modèle regroupe tous les concepts énumérés et les relations entre ceux-ci.

La description d'un modèle de représentation passe premièrement par la description des variables qui le composent. Ainsi, au centre du modèle, on retrouve les variables dépendantes et

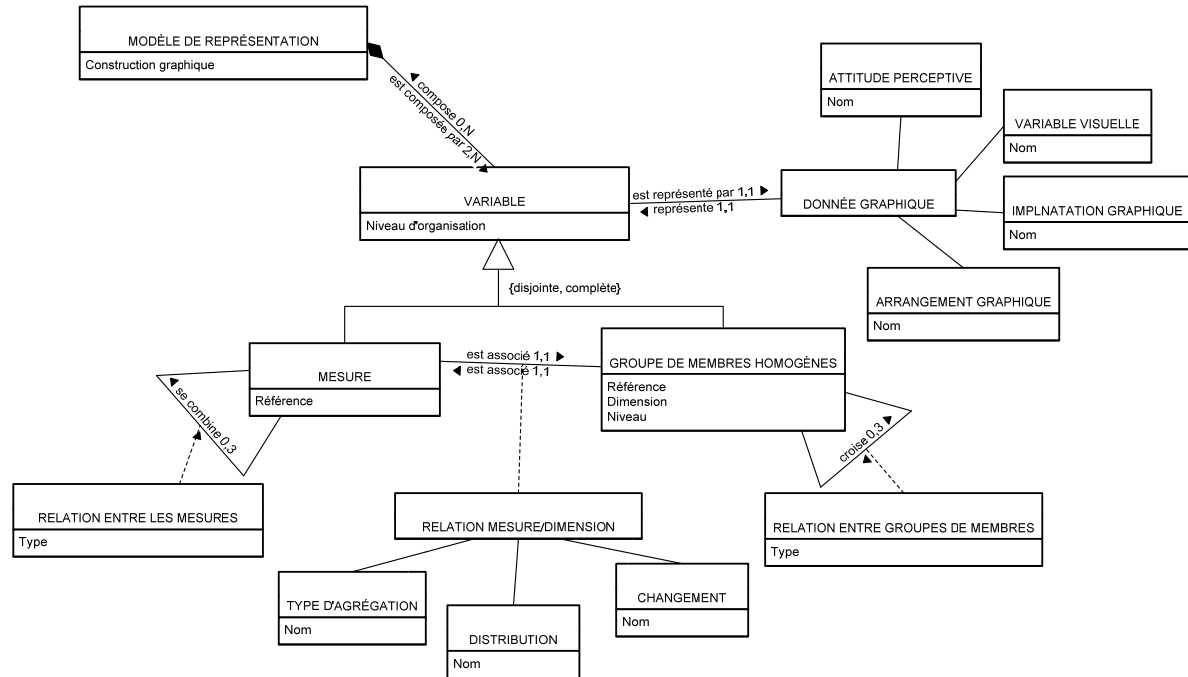
indépendantes d'un environnement multidimensionnel avec leurs caractéristiques. Une variable est premièrement décrite par son niveau d'organisation. Ensuite, cette variable est soit une mesure, soit un groupe de membres homogènes. Dans un environnement SOLAP, une mesure ou un groupe de membre possède une référence qui peut être thématique, temporelle ou spatiale. Un groupe de membre appartient aussi à une seule dimension et un seul niveau hiérarchique. Ces concepts sont regroupés sous le concept « caractéristiques intrinsèques des données ».

Deuxièmement, certaines caractéristiques des modèles de représentation passent par la description de la relation entre chacune des mesures et chacun des groupes de membres. Par exemple, les valeurs d'une mesure additive sur une dimension pourront être illustrées par un camembert (ex, le total des ventes sur une semaine) alors que les valeurs d'une mesure statistique pourront être comparées à leur ensemble (ex. la vitesse moyenne de trois athlètes). La relation entre une mesure et une dimension est décrite de trois façons : la distribution, le changement et le type d'agrégation. Ces concepts sont regroupés sous le concept « relation mesure/dimension ».

Ensuite, les données graphiques décrivent comment chacune des variables est encodée visuellement. Parmi, les données graphiques, on retrouve l'attitude perceptive, les variables visuelles, l'implantation graphique et l'arrangement graphique. Chacun de ces concepts se croise les uns les autres, en respectant la théorie de la sémiologie, pour définir la représentation d'une donnée. Ces concepts se regroupent sous le concept « caractéristiques graphiques ».

Les modèles de représentation comptent au minimum un groupe de membres (variable indépendante) et une mesure (variable dépendante) pour composer une construction graphique de base. Ainsi, un modèle de représentation se compose d'au moins deux variables. Par contre, ceci n'est qu'un minimum. Lors des requêtes multivariées, multi-niveaux ou ayant des multi-sélections dans plusieurs dimensions, les mesures se combinent et les membres des différents groupes se croisent entre eux. Il est proposé dans le modèle actuel de limiter les combinaisons entre les mesures et entre les groupes de membres à trois, suivant la recommandation de plusieurs auteurs [Bertin, 1973; Béguin et Pumain, 2003; MacEachren, 1995]. La combinaison des différentes variables comporte des subtilités à rendre visuellement. Celles-ci sont décrites par les relations entre les variables de même type. Par exemple, une relation parent-enfant peut se représenter par un cumul dans un histogramme (si l'opérateur utilisé est somme) ou une relation

comparable permet d'utiliser la même échelle dans les diagrammes orthogonaux. La relation entre les variables de même type et la construction graphique sont des concepts regroupés sous le concept « caractéristiques d'ensemble ».



* Les domaines de valeurs associés ne sont pas présentés ici par souci de concision. Ils sont présentés à l'annexe 2.

Figure 4.31 : Modèle conceptuel de la base de connaissances sémiologiques

La Figure 4.31 est le résultat de plusieurs itérations entre les validations préliminaires et la conception de la solution (cf. section 1.5, Figure 1.5) tel que décrit par la méthodologie de la présente recherche. Le modèle présenté ci-dessus, bien que plus complexe, peut s'apparenter à une structure multidimensionnelle en étoile. Il résulte d'ailleurs de l'évolution antérieure d'une structure en étoile (Annexe 2). Au centre du modèle, on trouve la « table de faits » qui décrit la représentation des variables par les différentes caractéristiques. Les caractéristiques correspondent aux différentes dimensions qu'aurait un modèle en étoile conventionnel. Les dimensions sont les données graphiques (l'attitude perceptive, les variables visuelles, l'implantation graphique et l'arrangement graphique), les caractéristiques des données (le niveau d'organisation, la référence et le type de variable) et les relations entre chacune des mesures et chacun des groupes de membres (la distribution, le changement et le type d'agrégation). Les caractéristiques d'ensemble (les relations entre les variables de même type et la construction

graphique) décrivent davantage le modèle de représentation plutôt que chacune des variables. Une relation N à N lie le modèle de représentation et chacune des variables. En effet, les variables doivent d'abord se combiner par nombre de deux (au minimum) pour former un modèle de représentation. Les caractéristiques d'ensemble (c.-à-d. la relation entre les variables de même type et la construction graphique) sont donc décrites à un second niveau dans la base de connaissances sémiologiques.

La comparaison du modèle conceptuel présenté avec une structure en étoile illustre les croisements entre les différents concepts de visualisation utilisés pour définir les différents modèles de représentation. Habituellement, dans un cube de données, tous les croisements entre les différentes dimensions sont possibles. Dans la base de connaissances, la combinaison de chacun des concepts n'est pas toujours possible. Elles sont limitées entre autres par la théorie de la sémiologie graphique. Ces limites sont en quelques sortes des contraintes d'intégrité supplémentaires. La formalisation de ces règles pourrait d'ailleurs faire l'objet de recherches ultérieures utilisant le langage proposé par [Salehi, 2009]. Par exemple, une attitude perceptive « quantitative » ne peut être combinée à une variable visuelle « couleur » et une donnée « nominale » ne peut prendre une attitude perceptive « quantitative » (même avec toute la latitude de la théorie sémiologique). Seules les combinaisons menant à une représentation adéquate sont stockées. Des tests en cours de conception de la solution comprenant certaines dimensions (type de variable, référence, niveau d'organisation, distribution, changement, type d'agrégation, implantation graphique, variable visuelle, perception et arrangement graphique) comptent environ 4400 combinaisons valides sur une possibilité d'environ 147 000. Les faits valides ne représentent donc que 3% de toutes les combinaisons possibles. La complexité de la diversité de la visualisation à l'intérieur du cadre sémiologique est donc gérée par l'application de contraintes. Les combinaisons résultantes sont en quelque sorte les associations intelligentes dans un cadre de conception graphique. Elles sont stockées de façon générique et s'appliquent donc à toute application SOLAP. De plus, chacune des combinaisons n'a qu'à être stockée qu'une seule fois dans la base de connaissances donnant accès à une expertise en visualisation géographique par n'importe quelle application pour n'importe quelle requête dans un cube de données.

Son fonctionnement

Selon le diagramme de collaboration présenté (Figure 4.28), la base de connaissances est interrogée par le client SOLAP lors d'une requête. Avant d'interroger la base de connaissances, le client SOLAP doit regrouper les variables qui composeront la cellule de l'image (Figure 4.29). La façon de regrouper les variables et de les organiser dans l'image dépend du type de visualisation (visualisation multivariée, répétitive, par composition et par substitution, cf. section 3.3.2). Lors des requêtes simples (ex. la population des régions administrative), on dispose d'un seul groupe de membres (les dimensions à sélection unique déterminant des membres constants, cf. section 4.1.1) et d'une seule mesure. La base de connaissances est donc interrogée directement pour connaître les représentations adéquates à cette combinaison de variables. Lors des requêtes multiples (ex. le total des ventes par année et par catégorie de produits ou le nombre de naissances et le nombre de décès sur quatre ans), le type de visualisation guide les regroupements des variables avant qu'une requête ne soit envoyée à la base de connaissances. La requête multivariée peut s'exprimer selon une visualisation multivariée ou une répétition (cf. section 3.3.2). La première combine les deux mesures et le groupe de membres en une seule image. La seconde sépare les variables multiples pour présenter deux ou plusieurs cartes ou diagrammes simples (Figure 4.32).

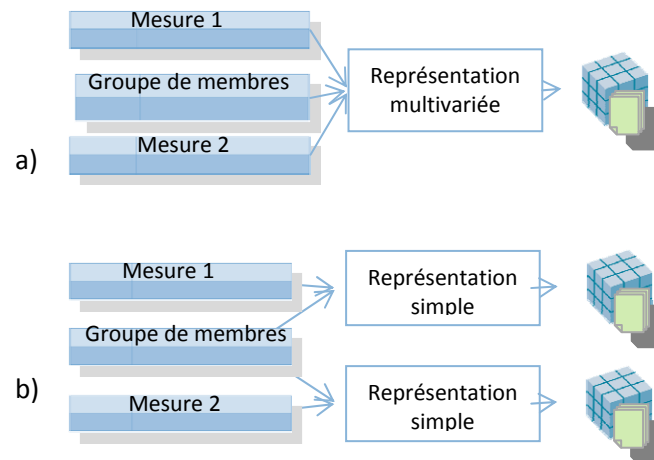


Figure 4.32 : Regroupement des variables d'une requête multivariée selon les types de visualisation : a) une visualisation multivariée et b) une répétition

Pour les requêtes multidimensionnelles, la visualisation peut être faite par composition, par répétition et par superposition (Figure 4.33). La composition regroupe chacune des dimensions requêtées avec la mesure indépendamment des autres dimensions. La répétition applique au moins un groupe de membres sur un axe du plan et réunit le groupe de membres restant avec la mesure pour questionner la base de connaissances. La superposition regroupe les variables de façon très similaire au regroupement pour la répétition. La différence notée est que les membres exclus du regroupement en premier sont spatiaux.

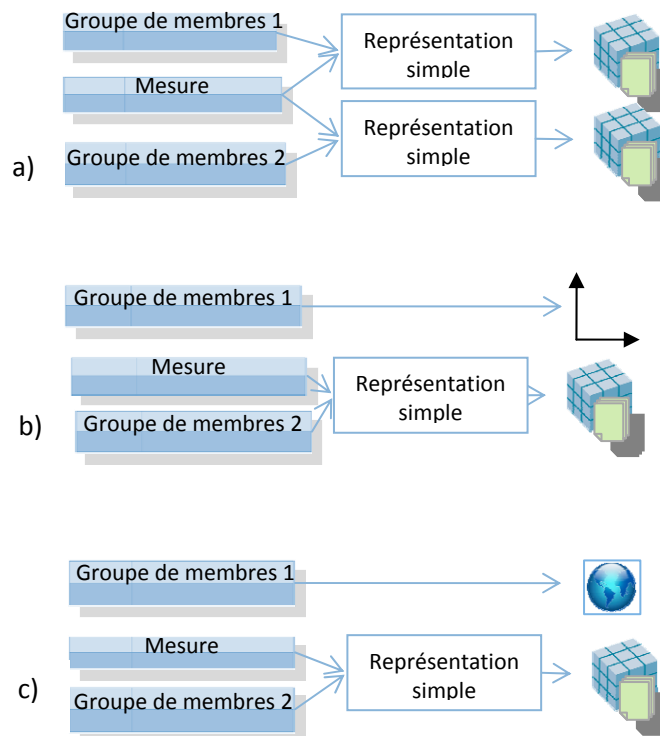


Figure 4.33 : Regroupement des variables d'une requête avec multi-sélections dans deux dimensions selon les types de visualisation : a) une visualisation par composition, b) une répétition et c) une superposition.

SOLAP permet également des requêtes multi-niveaux qui nécessitent des constructions graphiques particulières. Des diagrammes comparatifs et des regroupements devront être encodés dans la base de connaissances. Dans ces cas, les regroupements de variables dépendent de la relation parent-enfant ou de l'opération d'agrégation les reliant. La relation additive permet de définir un membre du niveau supérieur comme la somme de ses enfants. Une relation statistique doit par contre montrer la position du parent relativement aux enfants. Ces cas plus particuliers ne sont pas traités dans la présente proposition.

En résumé, cette première section du chapitre présente l'élaboration d'une méthode pour améliorer le lien entre les données et leur visualisation géographique avec SOLAP, le but étant d'intégrer l'expertise en visualisation géographique à l'intérieur même de la technologie SOLAP pour permettre une meilleure flexibilité de représentation à l'utilisateur. Premièrement, une taxonomie des concepts fait l'inventaire de tous les renseignements pertinents à la visualisation de l'information dans une approche SOLAP. Ensuite, la méthode proposée introduit deux éléments dans le système SOLAP : des métadonnées de visualisation et une base de connaissances sémiologiques. Un ensemble de métadonnées concernant les caractéristiques intrinsèques aux données, les relations entre chaque mesure et chaque dimension et les relations entre les variables de même type décrivent les données d'un cube. Puis, ces métadonnées pourront être utilisées pour établir le lien entre les données du cube et la base de connaissances sémiologiques. Cette dernière décrit tous les modèles de représentation supportés par une technologie SOLAP en spécifiant leur construction et les catégories de variables pour lesquelles ces modèles sont adaptés. Les informations stockées dans la base de connaissances permettent donc d'établir la correspondance entre les métadonnées du cube et les modèles de visualisation adaptés. Ainsi, les connaissances sémiologiques sont directement utilisables par l'application SOLAP évitant à l'utilisateur et à l'administrateur, la gestion de l'affichage.

Dans cette section, certains fondements doivent être considérés. Premièrement la théorie de la sémiologie n'est pas une science exacte. Elle est interprétée selon le sens critique de chacun des designers d'information (ou cartographes). Ainsi, le peuplement de la base de connaissances sémiologiques peut être fidèle aux règles strictes de la sémiologie graphique tout comme il est possible de supporter une certaine latitude. Par exemple, la théorie de Bertin indique que la couleur ne peut être perçue comme un ordre. Cependant la couleur est régulièrement utilisée dans les atlas pour représenter l'élévation du terrain puisque la distance de chacune des couleurs à l'œil est différente [Ware, 2000]. Aussi, certaines couleurs évoquent culturellement un ordre. Par exemple, la gradation de vert à jaune à rouge indique dans les pays industrialisés un niveau de danger ou de risque attribué aux feux de circulation [Few, 2006]. Le choix d'inclure ou d'exclure certaines règles sémiologiques est laissé à la discrétion du cartographe ou designer responsable du peuplement de la base de connaissances sémiologiques.

Ensuite, comme la visualisation de l'information s'applique différemment selon le contexte des données et de l'analyse qu'on veut y faire, créer une seule base de connaissances sémiologiques s'appliquant à tous les cubes de données est très ambitieux, peut rapidement devenir très ou trop lourd à gérer et certaines contradictions peuvent apparaître. Ainsi, la base de connaissances sémiologiques telle que proposée dans ce mémoire peut être vue comme un entrepôt de données duquel on crée des marchés de données personnalisables appliqués aux cubes de données. Pour un même cube, on peut alors créer deux « marchés de connaissances sémiologiques » différents pour deux types d'analyse différents. La base de connaissances sémiologiques génériques (l'entrepôt) peut alors regrouper tous les modèles de représentation complètement ou partiellement fidèle à la théorie de Bertin sans contradiction.

4.2 Les validations préliminaires

La solution proposée pour permettre de lier les données à leur visualisation sans l'intermédiaire de l'utilisateur est conceptuelle. L'implantation de la solution sous quelque forme que ce soit dépasse largement les limites de ce mémoire. Des validations préliminaires sont par contre réalisées tout au long de la conception pour faire évoluer la solution et ressortir les points forts et les points faibles de la solution. Des tests de faisabilité et quelques expérimentations exploratoires ont été exécutés. Des exemples de ces validations sont présentés dans cette section. Des vérifications similaires ont été réalisées tout au long de l'élaboration de cette solution.

4.2.1 Les tests de faisabilité

La solution proposée se divise en deux parties. La première est l'intégration des métadonnées à la conception du cube et la deuxième, le peuplement de la base de connaissances par la prise en compte des concepts. Les tests de faisabilité ont été réalisés parallèlement à la conception de la solution et ont aidé à raffiner la solution à son état actuel.

Les métadonnées de visualisation

Les métadonnées de visualisation sont les renseignements sur un cube de données nécessaires au design adéquat de visualisation pour chacune des requêtes. L'intégration des métadonnées dans le système demande au concepteur du cube de préciser ces renseignements. Plusieurs de ces

métadonnées peuvent être dérivées de la structure du cube. Par exemple, le type de variables, les références et le type d'agrégation découlent directement de la conception du cube. Ces renseignements sont donc explicitement stockés. Par contre, d'autres doivent être précisés par le concepteur du cube. Par exemple, le niveau d'organisation des données, la distribution et le changement et les relations entre les mesures devront être spécifiés (Figure 4.34).

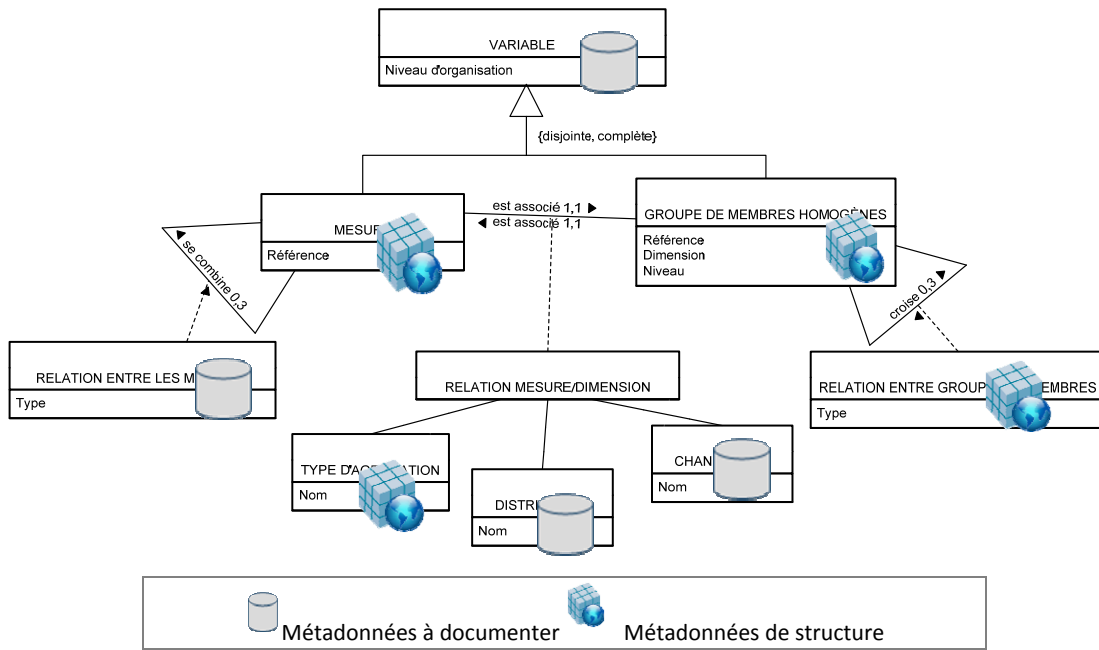


Figure 4.34 : Métadonnée de visualisation

Le niveau d'organisation doit être défini pour chacune des variables d'un cube. La distribution et le changement, eux, sont déterminés par la couverture de chaque mesure sur la dimension spatiale et temporelle. Ces renseignements peuvent être définis par l'utilisation d'une matrice formée des mesures en rangée et des dimensions en colonne (Tableau 4-1 et Tableau 4-2). Les valeurs présentées dans les cellules de ces tableaux et des suivants correspondent au niveau feuille de la taxonomie des concepts (cf. section 4.1.1) ou encore au domaine de valeurs du modèle conceptuel (Annexe 2).

Tableau 4-1 : Exemple de description de la distribution

Mesures	Dimension spatiale	Dimension temporelle
Nombre de mm de précipitation	Discrète	Discrète
Vitesse du vent	Continue	Continue

Tableau 4-2 : Exemple de description du changement

Mesures	Dimension spatiale	Dimension temporelle
Nombre de mm de précipitation	Abrupt	Graduel
Vitesse du vent	Graduel	Abrupt

La relation entre les variables de type « mesure » doit aussi être précisée. Cette relation est définie par une matrice symétrique formée par toutes les mesures (Tableau 4-3). Contrairement aux relations entre mesures, la relation entre les groupes de membres peut être négligée. En effet, les différents groupes de membres appartenant à des dimensions différentes sont comparables entre eux et la relation parent-enfant entre ceux-ci est dérivée directement de la structure hiérarchique du cube de données. Cette relation est donc dans la catégorie des métadonnées de structure (Figure 4.34).

Tableau 4-3 : Exemple de description des relations entre les mesures

Mesures	Population	indice de fiabilité	Nombre de naissances	Taux de croissance
Population		Qualificatif	Inclusif	Statistique
indice de fiabilité	Qualificatif		Non-comparable	Non-comparable
Nombre de naissances	Inclusif	Non-comparable		Non-comparable
Taux de croissance	Statistique	Non-comparable	Non-comparable	

La définition des métadonnées n'est jamais une tâche simple. Les renseignements doivent être complets et précis. Les renseignements décrivant la nature et les limites des données sont limités à leur minimum dans la solution proposée. Le concepteur du cube de données doit par contre bien connaître les phénomènes à représenter pour pouvoir les décrire.

Le peuplement de la base de connaissance

L'implantation de la base de connaissances peut prendre plusieurs formes : système expert, algèbre de composition, entrepôt de données, etc. Son implantation complète est, par contre, largement en dehors du cadre de ce mémoire. Les tests de faisabilité visent à confirmer qu'il est possible de décrire et de différencier les différents modèles de représentation selon les concepts et la structure proposés. La description de plus de cinquante modèles de représentation a été réalisée regroupant des représentations simples, multivariées, multidimensionnelles et multi-

niveaux en carte, diagramme ou tableau (Annexe 3). Les caractéristiques du même type y sont regroupées sous forme de tableau. Le processus de peuplement pour décrire chaque modèle de représentation est un processus *top-down*, c.-à-d. qu'on commence par identifier le modèle de représentation, puis on décrit ses caractéristiques jusqu'à identifier les données pour lesquelles ce modèle est adapté¹⁶. Les modèles de représentation stockés dans la base de connaissances sémiologiques sont tous les modèles supportés par l'application. En autant que possible, les principes de sémiologie graphique sont respectés. Cependant, comme la sémiologie graphique n'est pas une science exacte, le cartographe ou le designer (responsable du peuplement, cf. Figure 4.26) possède la pleine liberté de la conception des modèles de représentation.

Deux exemples de modèles de représentation sont présentés ci-dessous. Un symbole générique composé des caractéristiques principales du modèle (ex. une carte, un polygone et un spectre de valeur) est utilisé pour identifier le modèle de représentation. Pour le premier exemple, une carte choroplèthe conventionnelle est décomposée et le croisement de chacun des concepts est explicité dans la Figure 4.35. La carte choroplèthe entre dans la catégorie des cartes (construction graphique) ayant une variable indépendante ordonnée (attitude perceptive) en polygone (implantation graphique) dans le plan (arrangement graphique) par sa position (variable visuelle) et une variable dépendante colorée (arrangement graphique) par un dégradé (attitude perceptive) de valeur (variable visuelle) sur les polygones (implantation graphique). Cette représentation convient à un groupe de membres (variable) spatiaux (référence) ordonné (niveau d'organisation) associée à une mesure (variable) thématique (références) au minimum¹⁷ ordonnée (niveau d'organisation) continue dans l'espace (distribution). Pour ce modèle de représentation, le changement et le type d'agrégation n'ont aucun impact. Ils ne sont donc pas spécifiés.

¹⁶ Lors de l'utilisation de la base de connaissances, le processus est inverse; c.-à-d. qu'on observe d'abord les données pour trouver quel modèle de représentation y sont adaptés.

¹⁷ c.-à-d. qu'elle peut aussi être quantitative, cf. le principe de l'inclusion perceptive, Figure 2.14

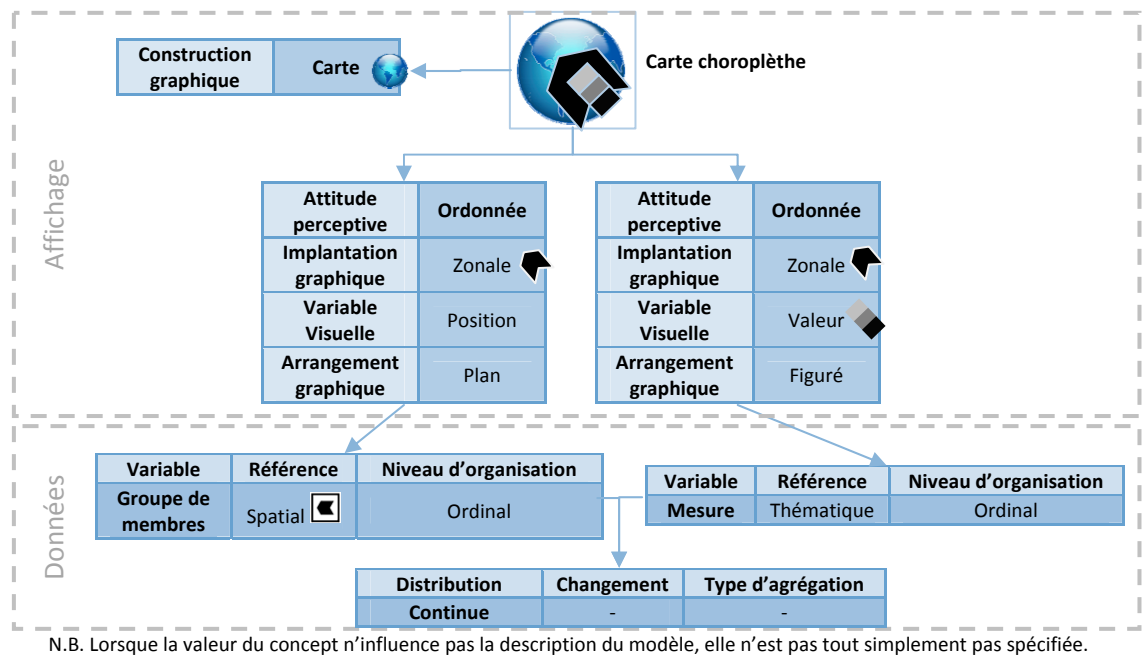


Figure 4.35 : Description d'une carte choroplèthe conventionnelle

Un deuxième exemple, plus complexe, décrit un diagramme en pyramide (Figure 4.36). Le diagramme en pyramide est un diagramme orthogonal représentant le croisement de deux variables indépendantes, dont une des variables indépendantes contient exactement deux membres, et une variable dépendante. La première variable indépendante est positionnée ponctuellement de façon ordonnée sur l'axe des Y. La deuxième est située de part et d'autre de l'axe des Y sur l'axe des X permettant de distinguer les membres de la variable. La variable dépendante, elle, est alignée pour représenter la quantité de part et d'autre de l'axe des Y sur l'axe des X. Cette représentation est adéquate pour une mesure thématique quantitative. Les variables indépendantes, elle, sont respectivement un groupe de membre thématique ordinal et un groupe de membre thématique qualitatif contenant exactement deux membres. Les relations mesure/dimension inférées par le modèle sont une distribution discrète, un changement abrupt et un type d'agrégation comparable. Aucune relation spécifique entre les deux groupes de membres n'est suggérée par le diagramme. Il est à noter que la seule variable visuelle utilisée pour illustrer ce diagramme est la variable de position à différents endroits dans le plan. Une combinaison redondante pourrait ajouter une variable visuelle (couleur, valeur) pour représenter la mesure ou un groupe de membres si une synchronisation de la symbologie est nécessaire avec d'autres diagrammes.

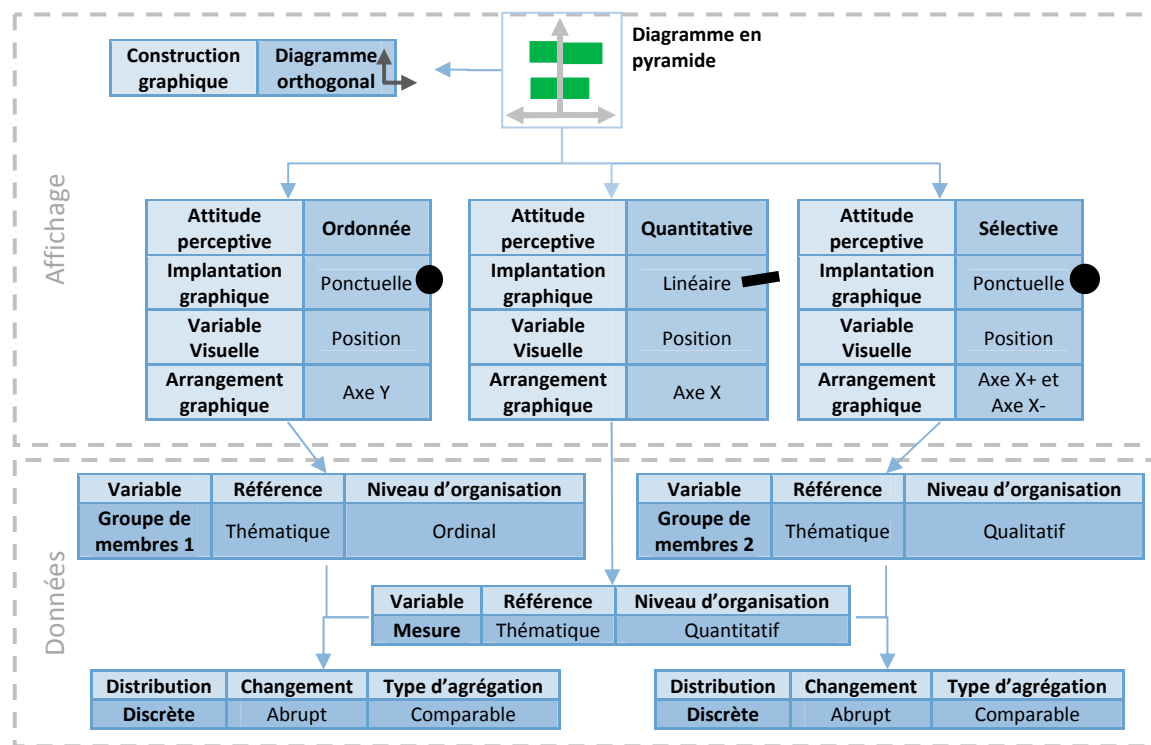


Figure 4.36 : Description d'un diagramme en pyramide

La cinquantaine de modèles de représentation de toutes sortes décrits à l'annexe 3 prouve qu'il est faisable de peupler une base de connaissances, c.-à-d. de décrire les modèles de représentation, à partir des concepts rassemblés selon le modèle conceptuel (Figure 4.31). La description du cadre d'application de chacun des modèles de représentation, tel que présenté à l'annexe 3, se compare au développement d'un système expert similaire à ce que [Mackinlay, 1986] propose. Une fois le cadre d'application de chacun des modèles décrit, il est possible de questionner la base de connaissance en commençant par le type de données pour retrouver toutes les options de représentation pouvant s'appliquer à une requête (de bas en haut).

Ces modèles ne sont qu'un résumé succinct de toutes les variantes possibles des modèles de représentation. Chacune des variantes supportées par l'application SOLAP devra être décrite de cette façon pour exploiter la solution proposée dans ce mémoire. Dans une structure multidimensionnelle, on évalue le nombre d'occurrences de la table de faits en multipliant le nombre de membres feuille de chacune des dimensions. Si l'on considère seulement les dimensions graphiques telles que les variables visuelles (7), l'implantation graphique (4), l'arrangement graphique (4) et l'attitude perceptive (4), il est possible de définir $7*4*4*4= 448$

combinaisons. Chacune de ces combinaisons étant jumelée l'une à l'autre pour définir un modèle de représentation. Évidemment, toutes les combinaisons ne sont pas stockées puisque toutes ne respectent pas la théorie de la sémiologie graphique. Par exemple, un groupe de membres spatiaux ou temporels (référence) sont ordinaux (niveau d'organisation) par définition (l'espace et le temps sont toujours ordonnés) et une attitude perceptive associative ne peut être rendue par la variable visuelle taille puisque celle-ci est dissociative (cf. section 2.4.2). Toutefois à l'intérieur du cadre des connaissances sémiologiques et à différents degrés d'intégrité, l'imagination est la seule limite pour la définition des modèles de représentation. Ainsi, de nombreux autres modèles pourraient être ajoutés. D'un autre côté, trop de diversité peut aussi emmener un éparpillement de l'information et une surcharge de choix pour l'utilisateur. Il est donc recommandé de n'ajouter que les modèles de représentation familiers dans le domaine d'analyse pour garder la base de connaissances sémiologiques plus générique. Des domaines d'analyses plus spécialisés demandant de nouvelles représentations pourront à leur tour enrichir la base de connaissances pour leur besoin. Ainsi, mis à part la base de connaissance générique, une base de connaissance appliquée¹⁸ peut être développée pour chaque application spécifique selon la méthode proposée dans le présent mémoire. L'échantillon préparé à l'annexe 3 représente les modèles de représentation les plus communs pour l'analyse statistique ou spatiale.

En résumé, la description des métadonnées et la possibilité de détailler les différents concepts attribués aux différents modèles de représentation prouvent qu'il est faisable d'intégrer une expertise en géovisualisation à une technologie SOLAP, toute limite ou considération technologique exclue.

4.2.2 Les expérimentations exploratoires

Tout au long de l'élaboration de la méthode, les différentes expérimentations ont permis de raffiner celle-ci jusqu'à son état actuel. L'intégration de l'expertise en visualisation géographique à l'intérieur de la technologie SOLAP a été expérimentée de façon exploratoire à l'aide d'applications SOLAP concrètes :

¹⁸ Des "marches de connaissances sémiologiques" cf. discussion de la base de connaissances sémiologiques Versus l'entrepôt de données, p.113

- Projet en kinésiologie de M. Lambert et J.P. Veilleux [Lambert, 2006];
- Projet pour la gestion des données du registraire 2000-2006 [Proulx et Bédard, 2004];
- Projet du cours SCG-66124 avec les données du recensement du Canada;
- Projet de génie MTQ Érosion pour la gestion des sites en érosion de R.-M. McHugh et F. Bilodeau [Mc Hugh, Bilodeau, Rivest et Bédard, 2006];
- Projet Muscamags 2008;
- Projet pour la gestion des emprises de lignes à haute tension avec Hydro-Québec et
- Projet pour la gestion du transport multimodal avec MTQ [Proulx, Rivest et Bédard, 2008].

Tous ces projets ont servi à tester les différents concepts et les différents besoins en matière de visualisation des applications SOLAP. Les cubes de données ont été requêtés à l'aide de la technologie JMapSOLAP. Étant donné que ces expérimentations exploratoires se sont tenues itérativement avec la conception de la méthode, il est impossible de présenter dans le présent document toutes les expérimentations exécutées. Deux mises en situation démontrant la validité de la solution pour les requêtes simples, multivariées et avec des multi-sélections dans plusieurs dimensions serviront d'exemples pour démontrer l'applicabilité de la solution. Ces mises en application sont tirées du projet en kinésiologie [Lambert, 2006] et du projet réalisé dans le cadre du cours SCG-66124 avec les données du recensement du Canada.


L'analyse de performances sportives (cf. projet en kinésiologie de M. Lambert et J.P. Veilleux)

La première expérimentation exploratoire est réalisée à l'aide d'un cube de données de performance sportive. Ce cube de données avait été développé dans le cadre d'un projet de maîtrise précédent [Lambert, 2006] à partir de mesures GPS traitées dans le cadre d'un autre projet de maîtrise [Veilleux, 2007]. Les mesures disponibles sont la vitesse moyenne, la vitesse maximale, la vitesse minimale et la somme des durées. Les dimensions sont le sexe, les athlètes (équipe et athlète), le temps (année, mois et jour), les tours et les portions de piste (piste complète, segments, sous-segments et segments atomiques de 70cm). Les requêtes testées avec ce cube sont une requête simple (requête 1) et une requête avec multi-sélections dans plusieurs dimensions (requête 2).

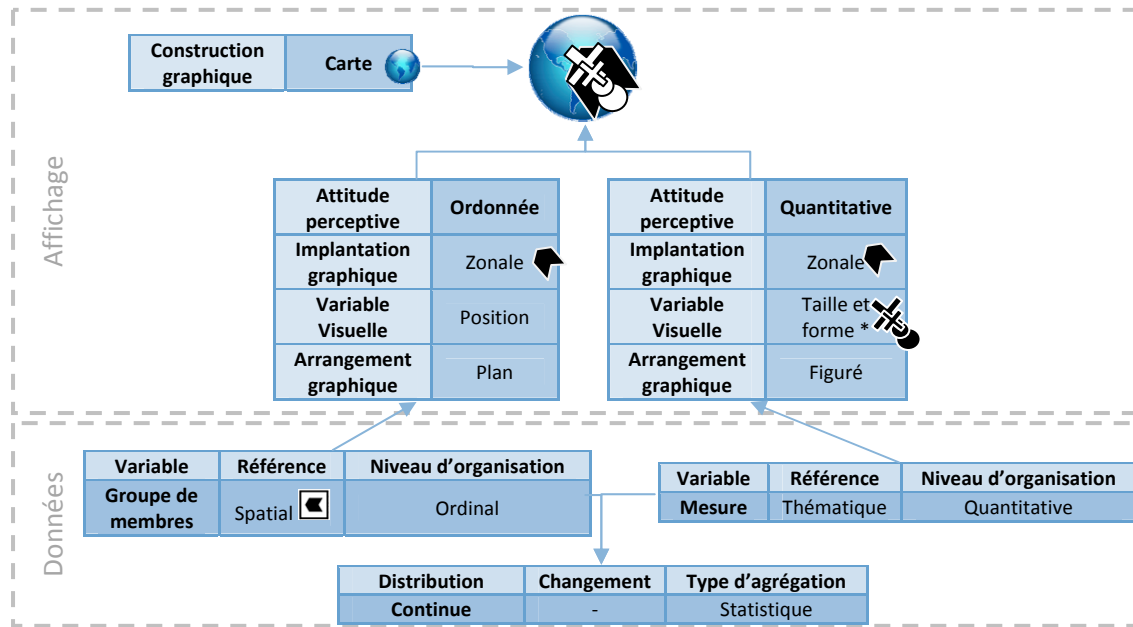
Requête 1 : Quelle est la vitesse moyenne de l'équipe des rapidos sur certains segments de piste sélectionnés?

La première étape lors d'une requête est de rassembler les métadonnées. Les caractéristiques des données de la requête sont listées dans le Tableau 4-4. La requête est simple puisqu'elle se compose d'une seule mesure et d'un seul groupe de membres. Toutes les autres dimensions composant les faits résultant de la requête ont une valeur de membre constante (ex. niveau « Tous » ou le « premier tour » de la dimension tour). Les relations importantes à tenir en compte dans cet exemple sont la relation spatiale, la relation statistique, la relation continue et graduelle de la mesure associée à ce groupe de membre.

Tableau 4-4 : Caractéristiques des données de la requête 1

	Variable	Référence	Niveau d'organisation		Distribution	Changement	Type d'agrégation
Segments de piste	Groupe de membres	 Spatial	Ordinal	}	Continue	Graduel	Statistique
Vitesse moyenne	Mesure 1	Thématique	Quantitative				

Une fois les métadonnées rassemblées, il suffit de parcourir la base de connaissances sémiologiques et récupérer les modèles de représentation correspondants. Un extrait de la base de connaissances comptant une cinquantaine de modèles de représentation est présenté à l'annexe 3. En premier lieu, la base de connaissances est filtrée selon le processus inverse au peuplement (*cf.* p.116), on recherche tous les modèles adaptés aux caractéristiques des données. Par exemple, une carte utilisant la taille d'un symbole sur une implantation graphique zonale répond quasi parfaitement aux métadonnées de la requête (Figure 4.37). Seul le changement diffère entre les métadonnées de la requête et la description de ce modèle. En fait, le changement n'est pas une caractéristique distinctive de ce modèle de représentation puisqu'il n'est pas spécifié. (Dans une structure de cube, ceci indiquerait le membre racine, « Tous », de la dimension.) Le changement graduel serait respecté si la taille s'étalait sur une gamme naturelle. Ce modèle respecte donc toutes les relations présentes dans les données. Le choix de la carte permet de respecter la représentation des relations spatiales, le choix de l'implantation graphique zonale permet de respecter la distribution continue et le choix de la combinaison des variables visuelles taille et forme permettent de respecter la relation statistique tout comme la perception quantitative de la mesure.



*La combinaison de la taille et la forme permet d'établir une symbologie divergente

Figure 4.37 : Description d'une carte utilisant la taille et la forme d'un symbole sur une implantation zonale

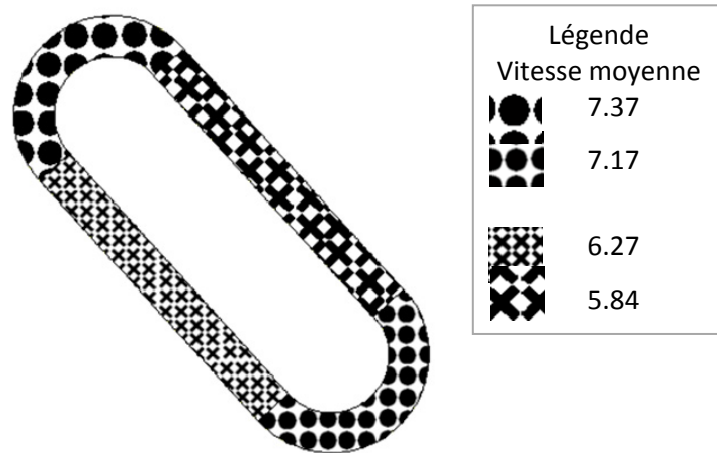
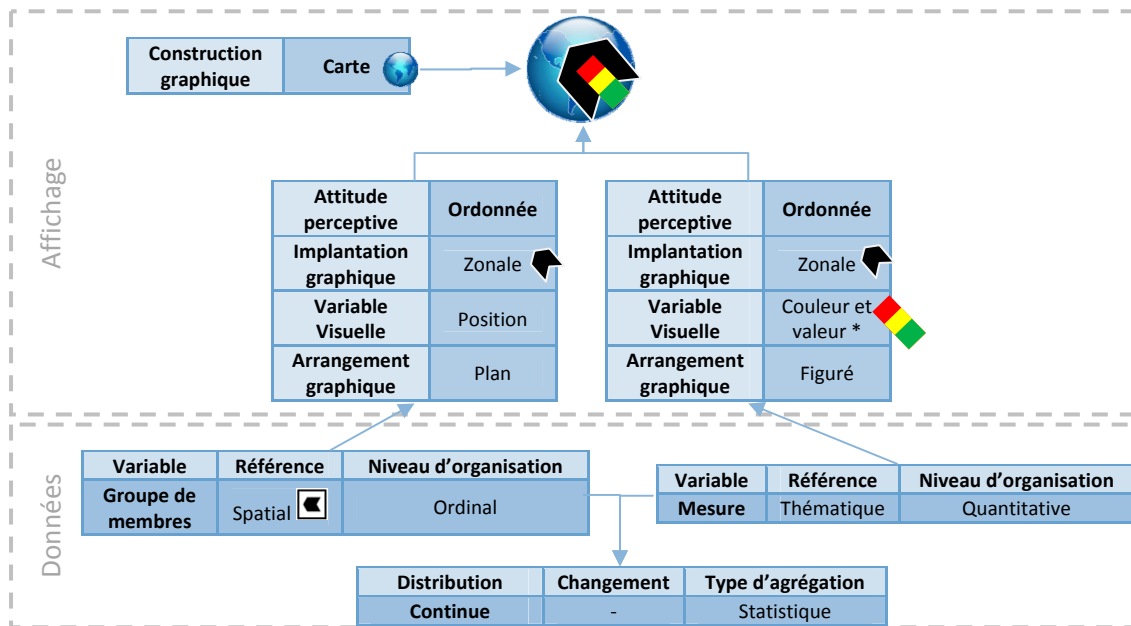


Figure 4.38 : Représentation de la requête 1 avec une carte utilisant la taille et la forme d'un symbole sur une implantation zonale

La visualisation géographique des données dépend généralement du niveau de familiarité de l'utilisateur avec les différentes symbologies. Même si la carte précédente (Figure 4.38) répond aux règles sémiologiques le plus fidèlement possible, elle peut ne pas convenir à l'utilisateur. De plus, même si toutes les relations sont exprimées dans un modèle de représentation, un autre pourrait mieux faire apparaître certaines relations ou structures dans les données. Plusieurs

options de visualisation doivent donc être fournies à l'utilisateur. Une autre option serait d'utiliser une carte choroplèthe avec un spectre de couleur divergent. La sélection de cette option par la base de connaissance implique que la représentation ne traduise plus la perception quantitative de la mesure, mais une perception ordonnée. Habituellement, ce compromis n'a pas de grandes conséquences sur la lecture d'ensemble d'une image. En fait, la perception quantitative relève davantage de la lecture élémentaire de l'image puisque le lecteur doit porter attention à un symbole en particulier pour le comparer à un autre pour en déduire la distance entre les deux. Plusieurs travaux mettent d'ailleurs en doute la capacité de l'humain à percevoir les quantités visuellement [Dent, 2005]. Dans ce cas, une carte de couleurs divergentes utilisant la valeur statistique comme seuil serait une bonne option, sinon la meilleure (dépendamment à quel point on se colle à la théorie de la sémiologie) (Figure 4.39).



*Le spectre de couleur doit être divergent. Le seuil représente la statistique à laquelle la valeur de la mesure est comparée.

Figure 4.39 : Description d'une carte choroplèthe de couleurs divergentes

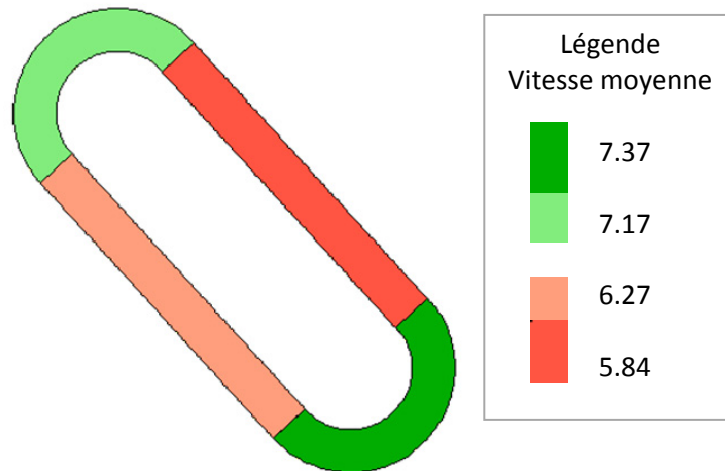


Figure 4.40 : Représentation de la requête 1 avec une carte choroplèthe de couleurs divergentes

En plus des cartes, des diagrammes peuvent aussi être utilisés en combinaison avec la carte pour faire ressortir la relation statistique de la mesure. Toutefois, à l'utilisation du diagramme seulement, on perd la relation spatiale entre les membres bien que la relation statistique soit rehaussée. Pour obtenir le meilleur des deux mondes, il est possible d'afficher côte à côte les deux représentations, la carte et le diagramme. Un histogramme comparatif peut donc être également intéressant pour l'utilisateur (Figure 4.41).

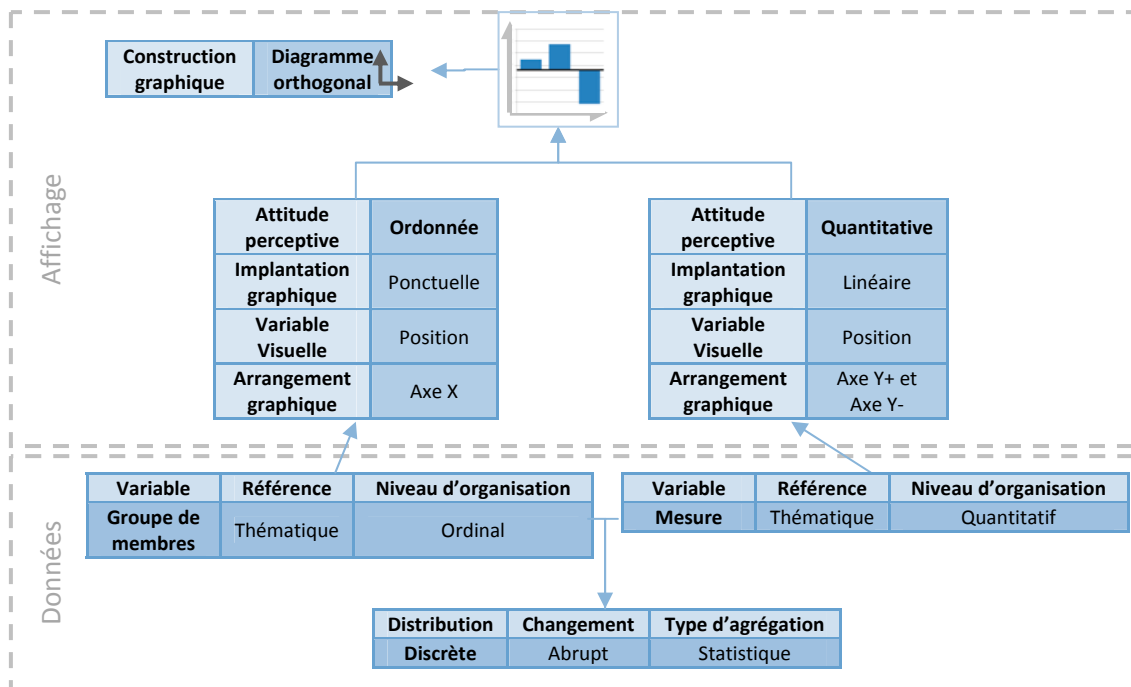


Figure 4.41 : Description d'un diagramme comparatif

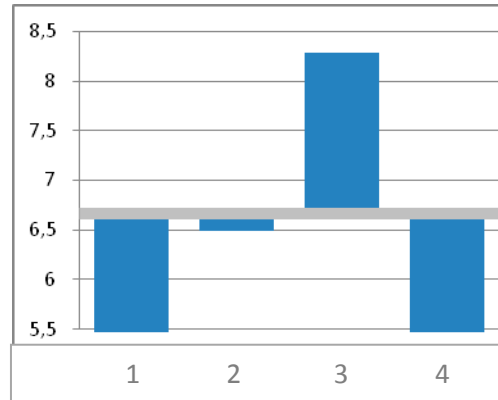


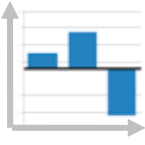

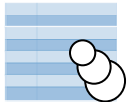





Figure 4.42 : Représentation de la requête 1 avec un diagramme comparatif

D'autres propositions peuvent être incluses dans la liste de choix à l'utilisateur. Ces propositions comprennent les propositions qui conviennent le mieux aux métadonnées de la requête jusqu'à des propositions plus génériques. La liste des propositions retenue pour la *requête 1* est décrite au Tableau 4-5. Sur la première rangée du tableau, on trouve les propositions qui représentent le plus fidèlement la plupart des relations présentes dans les données (c.-à-d. les relations spatiales introduites par le groupe de membres spatiaux et la relation statistique). Cependant, pour certains utilisateurs des représentations plus génériques peuvent être suffisantes à leur exploration. C'est pourquoi la liste contenu dans le tableau suivant comprend également des propositions plus génériques qui n'introduisent pas de fausses relations, mais ne permettent pas de voir toutes les relations (ex. tableau figuré, carte choroplèthe de valeur et l'histogramme).

Tableau 4-5 : Les propositions de la base de connaissances sémiologiques

<p>Carte choroplèthe de taille</p> 	<p>Carte choroplèthe avec couleurs divergentes</p> 	<p>Diagramme comparatif</p> 	<p>Tableau régulier avec couleur divergente</p> 
<p>Tableau figuré de taille</p> 	<p>Carte choroplèthe de valeur</p> 	<p>Diagramme à ligne avec seuil</p> 	<p>Histogramme</p> 

La liste de modèles de représentation pour une requête simple et commune peut être longue, puisque certains modèles spécifiques (la carte avec couleur divergente et le diagramme comparatif) s'appliquent autant que les modèles plus génériques (histogramme, tableau simple, carte choroplèthe). « Cent constructions pour une information » [Bertin, 1973] est une phrase bien adaptée dans ces situations. Le nombre de propositions doit tout de même être restreint pour ne pas surcharger l'utilisateur d'options de visualisation. La règle de sept plus ou moins d'eux s'appliquerait ici [Miller, 1956; Wurman, 1989]. Le choix des premiers modèles de représentation proposé à l'utilisateur peut à ce moment être défini par des règles de priorisation établis, par exemple, à partir des choix précédents de l'utilisateur. La priorisation des modèles de représentation pourra faire l'objet de recherches ultérieures pour mieux assister l'utilisateur.

La seconde requête est une requête avec des multi-sélections dans trois dimensions engageant de rendre visuellement le croisement de trois groupes de membres.

Requête 2 : Comment varie la vitesse moyenne sur les différents tours dans les différents segments de la piste pour les trois mois de pratique.

De la même façon que lors de la requête précédente, les métadonnées de la requête sont rassemblées. Le

Tableau 4-6 liste les métadonnées de cette requête. La requête regroupe quatre variables (une mesure et trois groupes de membres). Les relations importantes traduites dans les métadonnées sont la relation statistique de l'association de la mesure avec chacun des groupes de membres, la relation temporelle continue et graduelle et la relation spatiale continue et graduelle. Les relations spatiale et temporelle ont une distribution continue puisque des données sont disponibles à un niveau de granularité plus fin dans les deux dimensions. Elles ont également toutes les deux un changement graduel puisque, dans le temps comme dans l'espace, la vitesse ne varie pas par bonds saccadés. Ces deux caractéristiques sont très relatives à l'application. C'est pourquoi ces métadonnées doivent être documentées par l'administrateur du système qui connaît bien le contexte dans lequel se tiendra l'interprétation des données. Comme différentes applications pourraient nécessiter différentes métadonnées, différentes bases de connaissances pourraient être conçues pour un même cube (*cf.p.* 113).

Tableau 4-6 : Caractéristiques des données de la requête 2

	Variable	Référence	Niveau d'organisation		Distribution*	Changement*	Type d'agrégation
Vitesse moyenne	Mesure 1	Thématique	Quantitative		Continue	Graduel	Statistique
Segments de piste	Groupe de membres 1	Spatial	Ordinal		-	-	Statistique
Tours	Groupe de membres 2	Thématique	Ordinal		Continue	Graduel	Statistique
3 mois	Groupe de membres 3	Temporel	Ordinal				
Relation entre les dimensions							
Comparable							


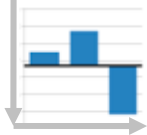
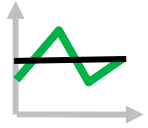




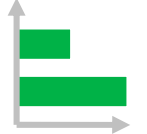
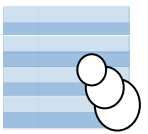
*La distribution et le changement ne sont valables que pour les dimensions de référence temporelle ou spatiale dans cette requête.

Pour une telle requête, plusieurs types de visualisation différents (cf. section 3.3.2 et section 4.1.3) peuvent mener à des représentations différentes. Premièrement, la **composition** organise les variables de façon à ce que chaque groupe de membres se regroupe avec la mesure afin de trouver une représentation simple pour chaque association (Figure 4.33). Dans cet exemple, trois représentations composent l'image. Une première comprend la combinaison de la mesure « vitesse moyenne » avec « les segments de piste ». La deuxième comprend la « vitesse moyenne » et les « tours ». La troisième comprend la « vitesse moyenne » et les « mois ». La base de connaissances sémiologiques est interrogée indépendamment pour chacun de ces groupes. Une liste de choix correspondant aux métadonnées est présentée dans le Tableau 4-7. Chacun des modèles présentés ci-dessous est décrit à l'annexe 3. Par souci de concision, la description de ces modèles ne sera pas explicitée dans cette section.

En lien avec la requête 1, on remarque que chacune des combinaisons « vitesse moyenne/groupe de membres » donne lieu à une représentation simple. Les premières représentations simples (première ligne du tableau) montrent clairement la relation statistique. Les autres représentations suggèrent de négliger certaines relations pour en dégager d'autres. Par exemple, l'histogramme pour représenter la combinaison « vitesse moyenne/mois » néglige la relation de continuité dans le temps et, par conséquent, met l'accent sur la perception quantitative de la mesure. L'utilisation de la visualisation géographique dans un environnement interactif d'exploration de données est en fait une série de compromis que l'on fait pour arriver à trouver une réponse, une tendance ou

une structure dans les données. Ces compromis sont en fait une série de choix que l'utilisateur doit faire [Kraak et Omerling, 2003]. Ces choix sont soumis à l'utilisateur par l'intermédiaire d'une base de connaissances sémiologiques.

Tableau 4-7 : Exemples possibles de modèles de représentations formant la composition répondant à la requête 2

Vitesse moyenne/ segment de piste	Vitesse moyenne/ Tours	Vitesse moyenne / mois
Carte choroplèthe avec couleur divergente 	Diagramme comparatif 	Diagramme à ligne avec seuil 
Carte choroplèthe avec valeur 	Histogramme 	Histogramme 
Carte de symboles proportionnels 	Histogramme à barres horizontales 	Tableau figuré de taille 

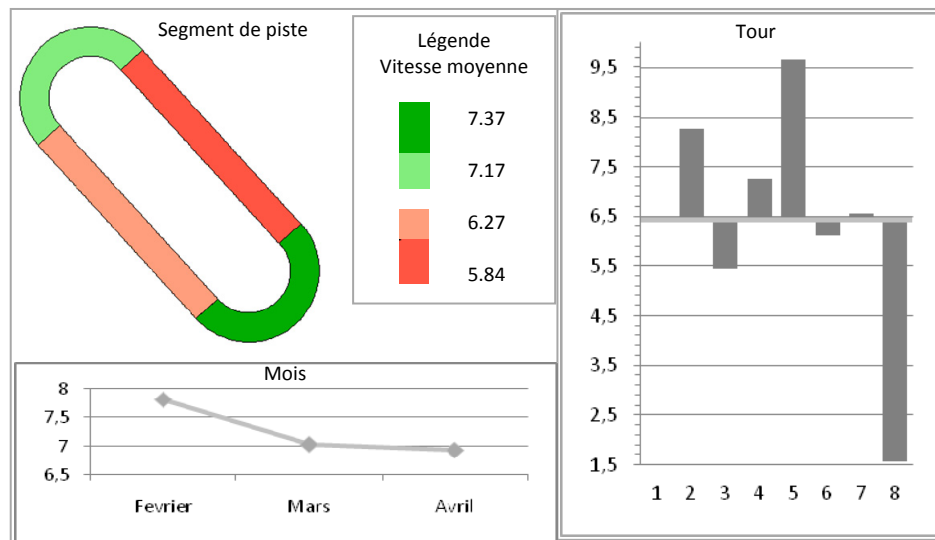


Figure 4.43 : Exemple d'une composition pour la requête 2

La composition n'est pas la seule façon de représenter une requête ayant des multi-sélections dans plusieurs dimensions. Celle-ci peut également être vue par une **répétition** d'images. Pour ce faire, deux variables sont affectées aux axes du plan. La troisième variable associée à la mesure formera l'image (Figure 4.33). Une seule requête est donc envoyée à la base de connaissances.



Dans l'exemple ci-dessus, le client SOLAP peut proposer de disposer les « tours » sur l'axe Y et les « mois » sur l'axe X (Figure 4.44). Ce choix est totalement arbitraire et l'utilisateur pourra modifier cette orientation avec les outils de navigation tels que le pivot (*cf.* Figure 3.4). La dimension spatiale (segment de piste) et la mesure (la vitesse moyenne) formeront la combinaison de variable d'entrée pour la base de connaissances sémiologiques. Comme la requête avec multi-sélection est réduite à une requête simple à l'entrée de la base de connaissances sémiologiques, la liste des modèles de représentation extraite de la base de connaissances sémiologiques sera la même que dans le Tableau 4-7, excepté que seuls les modèles sous le titre « vitesse moyenne/segment de piste » seront proposés à l'utilisateur. Dans le cas où ce dernier choisirait de mettre le groupe de membres spatiaux sur un des deux axes du plan (ex. en Y), le groupe de membres remplacé par la dimension spatiale (dans ce cas-ci, les mois) composerait alors la cellule du plan formé par les groupes de membres en X et en Y. La liste se composerait alors des modèles valables pour la combinaison « vitesse moyenne/mois » dans le Tableau 4-7.

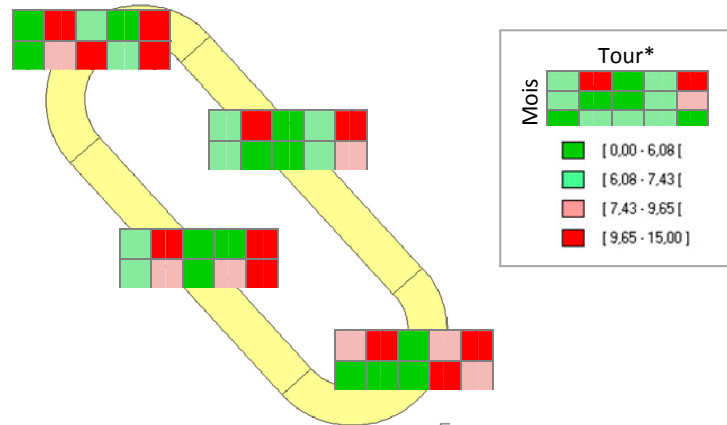
	Février	Mars	Avril
1 ^{er} tour			
2 ^e tour			
3 ^e tour			
4 ^e tour			
5 ^e tour ...			

Figure 4.44 : Exemple d'une représentation de la requête 2 par la répétition de carte.

Une autre option pour la visualisation des requêtes ayant des multi-sélections est la **superposition**. À ce moment, la dimension spatiale sera exclue des variables d'entrée pour monopoliser le plan (Figure 4.33). Un diagramme multidimensionnel comprenant les « tours » et les « mois » sera superposé à chaque membre spatial. Dans l'exemple ci-dessus, ce type de visualisation est difficilement réalisable dans un but de visualisation globale et rapide. En effet, on compte 9 tours et 3 mois différents. 27 valeurs de vitesse moyenne doivent donc être encodées par membre spatial. Bien que ceci soit possible, le but de la visualisation est d'accélérer l'exploration et l'analyse des données dans un environnement SOLAP. Une représentation de la sorte nécessiterait beaucoup d'attention et de mémorisation de la part de l'utilisateur. Elle ne serait donc pas le meilleur choix à faire. Si une solution doit absolument être fournie, la superposition d'un tableau figuré de taille ou de couleur sur les membres spatiaux pourrait répondre à cette requête.

Tableau 4-8 : Exemples de superposition multidimensionnelle possible répondant à la requête 2

Vitesse moyenne/ Tours / mois
Tableau figuré de couleur divergente 
Tableau figuré de taille 



*Pour une meilleure lecture le nombre de tour a été réduit à 5.

Figure 4.45 : Exemple d'une représentation pour la requête 2 par une superposition multidimensionnelle

Évidemment, plus le nombre de variables à représenter est important, plus la représentation de l'information sera complexe et plus le nombre d'options sera faible. Selon la barrière infranchissable de la synoptique (cf. section 2.4.2), un maximum de trois composantes peut être encodée dans une image. Au-delà de cette barrière, il faut user de ruse et d'imagination pour créer des représentations efficaces. L'abstraction et l'interaction sont deux moyens de permettant surpasser cette barrière.

L'analyse des données de recensement (cf. Projet du cours SCG-66124)

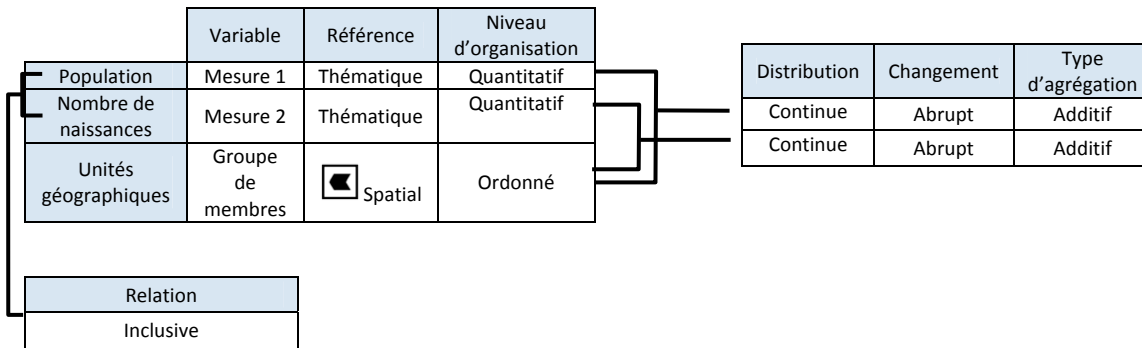
Dans le cadre du cours de notions avancées en bases de données SIG (SCG-66124), un cube de données comprenant les données du recensement du Canada est créé par les étudiants. Ce cube sert également la deuxième expérimentation exploratoire présentée ici. Le cube comprend trois mesures (population, nombre de naissances et nombre de décès) et cinq dimensions (unités géographiques, statut, année, sexe et classes d'âge). La requête testée dans cet exemple est multivariée.

Requête 3 : Quelle est le nombre de naissances et la population à travers le territoire?

La requête 3 comprend deux mesures et un groupe de membre. Les caractéristiques de cette requête sont présentées au Tableau 4-9. Les relations importantes à rendre sur l'image sont la

relation inclusive entre les deux mesures, la relation additive de chacune des mesures sur le groupe de membres, la relation spatiale, la distribution continue et le changement abrupt.

Tableau 4-9 : Caractéristiques des données de la requête 3



Lors d'une requête multivariée, l'utilisateur dispose de deux options (la visualisation multivariée et la répétition, cf. section 3.3.2). La première est d'intégrer toutes les variables dans l'image pour obtenir une **visualisation multivariée**. Toutes les variables constituent donc les données d'entrée pour l'interrogation de la base de connaissances sémiologiques. Un premier modèle de représentation des données est une carte de symboles proportionnels superposés (Figure 4.46). Deux symboles proportionnels peuvent être superposés puisque la première mesure (population) inclut l'autre (les naissances). Le premier symbole sera nécessairement toujours plus grand que le deuxième. Ce modèle respecte la relation spatiale et la relation d'inclusion entre les mesures. Par contre, la distribution continue, le changement graduel et la relation additive sont négligés.

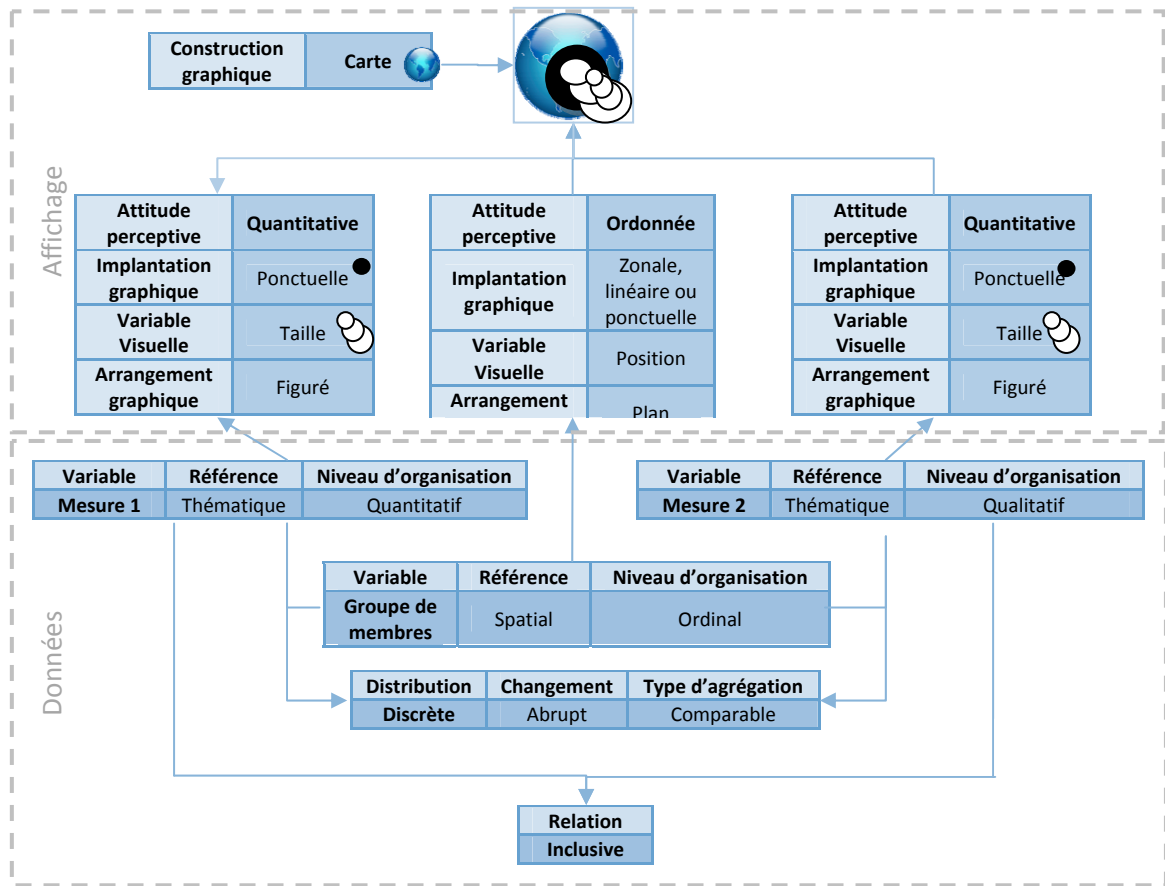


Figure 4.46 : Description d'une carte de symboles proportionnels superposés

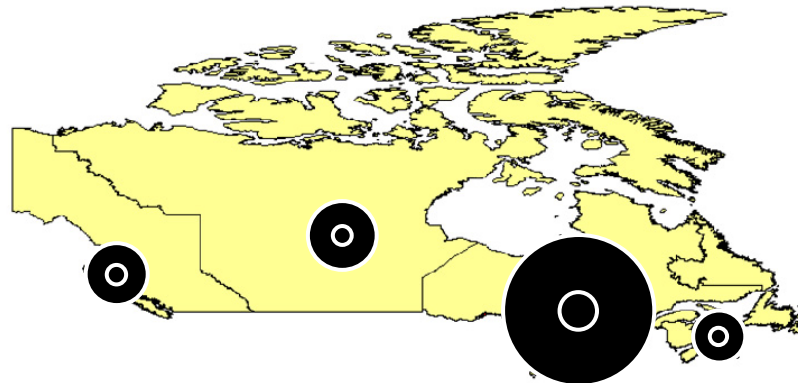


Figure 4.47 : Exemple de représentation pour la requête 3 avec une carte de symboles proportionnels superposés

Un autre modèle proposé similaire au précédent est la carte avec camemberts proportionnels (Figure 4.48). Considérant que la deuxième mesure est incluse dans la première, elle représente donc une proportion de la première. Cependant, l'utilisation de ce symbole est pertinente

uniquement si la proportion de la deuxième mesure (nombre de naissances) est plus grande que 2% de la première mesure (population) pour tous les membres spatiaux de la requête. Dans le cas contraire, la seconde mesure ne sera pas visible puisque la part du camembert sera trop petite ($<10^\circ/360^\circ$). Ce modèle respecte alors la relation spatiale, et la relation d'inclusion entre les mesures.

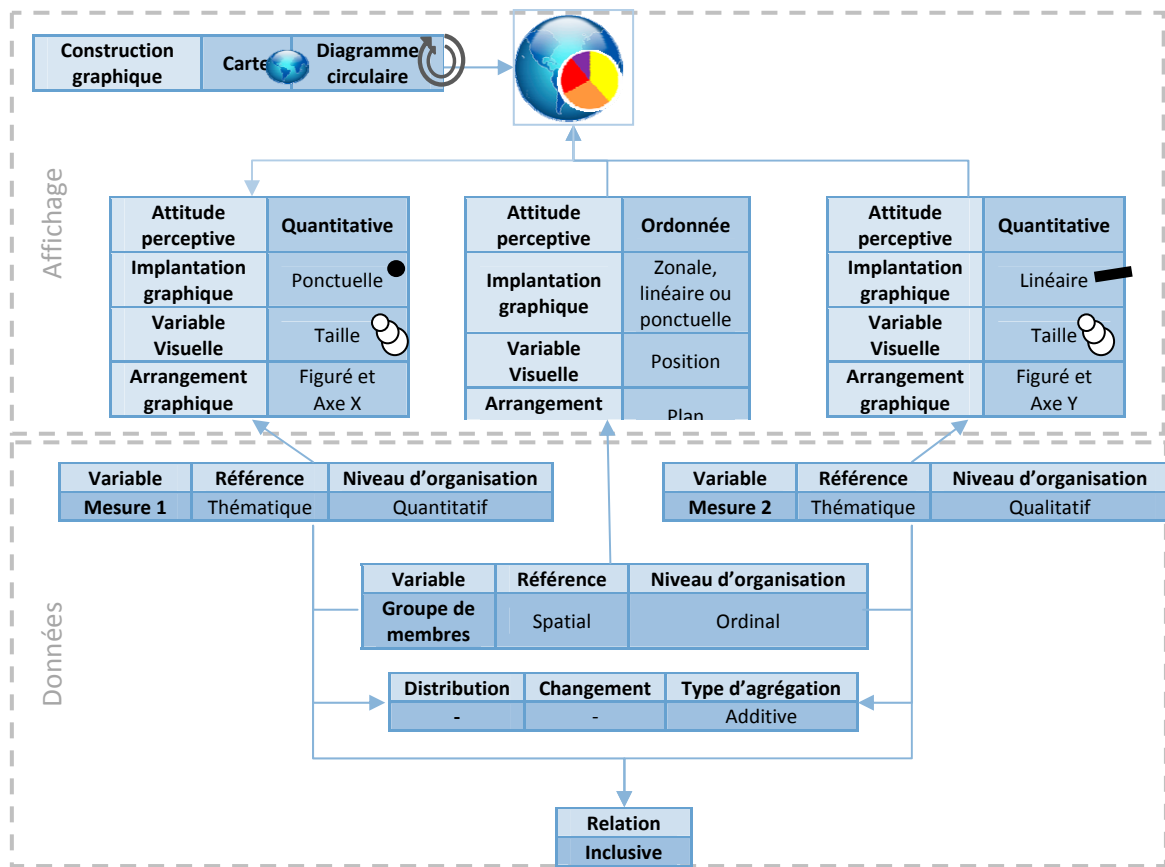


Figure 4.48 : Description d'une carte avec camemberts superposés

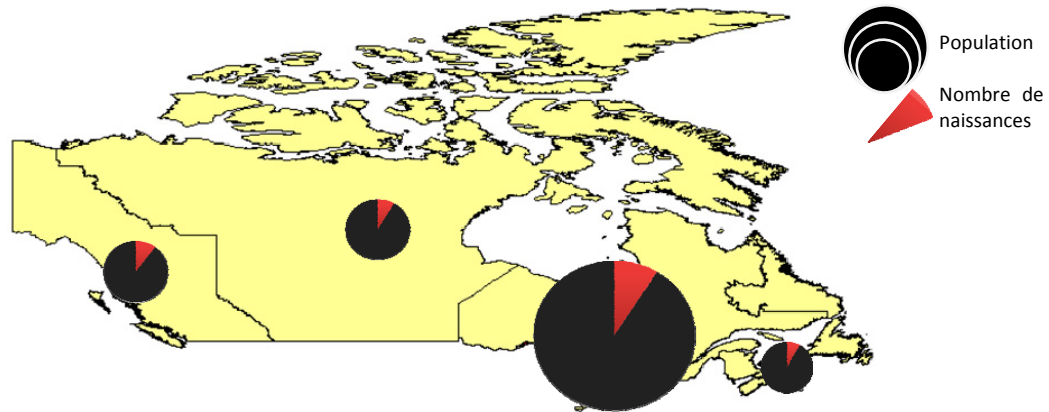


Figure 4.49 : Exemple de représentation de la requête 3 avec une carte de camemberts superposés

Dans les cas de la visualisation multivariée, un choix entre la relation entre les mesures (inclusion) et la relation des mesures dans le groupe de membres (addition) doit être fait. Il est très difficile voir impossible d’allier ces deux relations dans une seule image. De plus, dans cet exemple, il est impossible de rendre la relation d’additivité et la relation spatiale sur le même modèle. Ainsi, un autre modèle comme des camemberts superposés peut être proposé à l’utilisateur pour voir la relation d’additivité des deux mesures par rapport aux membres spatiaux (Figure 4.50).

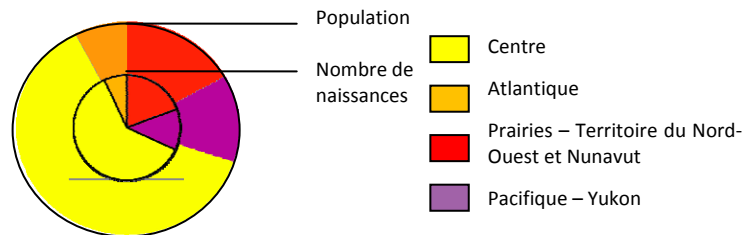





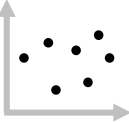


Figure 4.50 : Exemple de représentation de la requête 3 à l’aide de camemberts superposés

D’autres modèles de représentation s’appliquant aux variables de cet exemple sont présentés dans le



Tableau 4-10 . La définition de chacun de ces modèles de représentation est fournie à l’annexe 3.

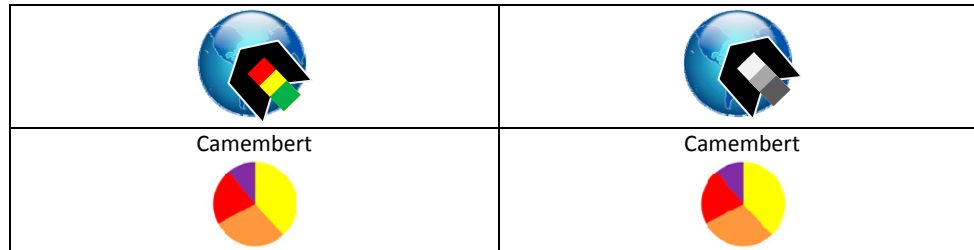
Tableau 4-10 : Exemples de représentations multivariées pour la requête 3

<p>Carte de symboles proportionnels superposés</p> 	<p>Histogramme superposé</p> 
<p>Carte avec camemberts proportionnels</p> 	<p>Tableau multivarié avec couleur et texte</p> 
<p>Camembert superposé</p> 	<p>Nuage de point</p> 

La deuxième option pour la représentation des requêtes multivariées est de **répéter** la représentation pour chacune des mesures. (À ce moment la relation d'inclusion entre les deux mesures est négligée.) Pour ce faire, la variable « unités géographiques » est alors jumelée à chacune des mesures avant de requêter la base de connaissances sémiologiques (Figure 4.32). Une demande à la base de connaissances sémiologiques est envoyée pour les couples de variables « population/unité géographique » et « nombre de naissances/unité géographique ». Une liste de modèles de représentation est alors extraite de la base de connaissances (Tableau 4-11). Les deux mesures étant inclusives, il serait inadéquat d'utiliser une seule légende pour représenter ces deux mesures à travers les différentes représentations. En effet, leur ordre de grandeur peut différer considérablement. Les deux affichages seront donc indépendants au niveau de la symbologie.

Tableau 4-11 : Exemples de représentation simples pour la requête 3

Population/ unités géographique	Nombre de naissances / unités géographiques
Carte de symboles proportionnels	Carte de symboles proportionnels
	
Carte choroplèthe de couleur	Carte choroplèthe de valeur



Les différentes mises en contexte présentées valident d'une façon préliminaire la solution. Il est théoriquement possible d'associer des métadonnées à une base de connaissances sémiologiques pour trouver les meilleurs modèles de représentation supportés pour une requête ou une analyse. Certains modèles feront ressortir certaines relations dans les données alors que d'autres sont plus généraux et s'appliquent à toutes les situations. Par exemple, un camembert fait ressortir une relation de proportion alors qu'un histogramme sépare les données pour les comparer entre elles. La solution proposée permet donc d'intégrer d'une façon générique une expertise en visualisation géographique par l'intermédiaire d'une base de connaissances sémiologique. Celle-ci s'appuie sur la taxonomie des concepts de visualisation qui sont à la base de design graphique.

4.3 L'analyse critique

La solution proposée présente plusieurs avantages, mais possède aussi des limites et des inconvénients. À la lumière des validations préliminaires, une analyse critique de la solution est réalisée pour mettre en évidence les points forts et les points faibles de la solution.

4.3.1 Les points forts

Le premier point fort qui ressort de cette solution est l'existence d'une taxonomie de concepts de visualisation. Les bonnes pratiques et la théorie de la sémiologie graphique sont grandement documentées dans la littérature cartographique. Cependant, les principes sont souvent appliqués dans un contexte particulier. La taxonomie telle que présentée à la section 4.1.1 permet de regrouper et d'organiser les concepts qui soutiennent la cartographie et le design d'information de façon générique. Avec la méthode proposée, il devient alors possible de former une base de connaissances sémiologiques qui puisse gérer les subtilités de la conception cartographique et ce,

selon la discrétion du cartographe ou du designer d'information. Ainsi l'art peut se marier à la science pour produire de meilleurs résultats.

Comme présentée lors des expérimentations exploratoires, la visualisation géographique, en plus d'être une discipline plus ou moins définie, est une histoire de compromis. Il est souvent impossible de représenter toutes et chacune des relations entre les variables lors de requêtes plus complexe. Certaines relations sont mises en évidence et d'autres sont cachées ou absentes. La méthode proposée permet de sortir plusieurs représentations pertinentes à chacune des relations décrites dans les données. Cette liste d'options transmise à l'utilisateur lui donne donc l'opportunité de chercher à travers différentes représentations de façon flexible et rapide (considérations technologiques exclues) dans le but de trouver des relations, des structures et des tendances dans ses données. Ceci permet de répondre au besoin de réexpression.

Ensuite, la solution permet d'accéder à différents types de visualisation. En effet, les données peuvent être regroupées selon une visualisation multivariée, répétitive, en composition ou en superposition. Les différents regroupements permettent de changer le focus des relations à percevoir tout en respectant le cadre sémiologique établi par le cartographe. Par exemple, passer d'une représentation en composition à une représentation répétitive permet de percevoir la globalité de chacune des dimensions ou de percevoir chacun des croisements.

La solution proposée permet également d'encadrer les différentes modifications graphiques que pourrait apporter l'utilisateur (Figure 4.30). En effet, en disposant d'une série de représentations adéquates, l'application SOLAP dispose de toute l'information nécessaire pour encadrer et limiter la diversité des symbologies possibles (implantation graphique et variables visuelles) pour une requête particulière.

Finalement, la synchronisation des symbologies entre les différents tableaux, cartes et diagrammes d'un collection peut être respectée (cf. p.118). Celle-ci se traduit par l'ajout d'une combinaison redondante, lorsque possible. L'utilisation d'une ou de plusieurs légendes pour plusieurs mesures peut être gérée à partir des relations entre les mesures. La base de connaissances sémiologiques fournit toute l'information nécessaire pour savoir ce qui est possible et ce qui ne l'est pas.

Au-delà des besoins d'un environnement SOLAP, la base de connaissances sémiologiques est ouverte et plusieurs modèles de représentation peuvent être ajoutés. Certaines applications peuvent nécessiter certaines visualisations particulières qui impliquent de nouveaux modèles de représentation. En proposant une structure de la sorte, ces nouveaux modèles pourront être ajoutés au fil des développements soit par la mise à jour de la base de connaissances génériques soit par la création d'un « marchés de connaissances sémiologiques » pour l'application (cf.p.113). La méthode proposée permet de développer une base de connaissances appliquée pour compléter la base de connaissances générique dans un contexte particulier.

4.3.2 Les points faibles

Bien que la solution proposée permette de répondre à plusieurs besoins d'un environnement SOLAP, elle possède également quelques points faibles. Le premier est que la définition de tous les modèles de représentation possibles est lourde et fastidieuse. Le peuplement de la base de connaissances est donc un travail imposant qui demande une bonne connaissance du cadre sémiologique et de la visualisation en général. Le point positif de cet inconvénient est que le travail n'est réalisé qu'une seule fois pour la base de connaissance générique. Une fois la base de connaissances peuplée, c'est elle qui gère l'organisation des variables et leur affichage. Cependant, pour des modèles de représentation plus particuliers, le travail devra être réalisé pour peupler les bases de connaissances sémiologiques appliquées.

Ensuite, la solution fait abstraction de certains concepts multidimensionnels. Par exemple, lors de l'analyse multidimensionnelle, la visualisation des données se base régulièrement sur une structure en tableau. Les concepts de « rangée » et de « colonne » sont souvent utilisés pour bâtir la représentation. Ces concepts sont omis dans la solution proposée laissant l'organisation des données à la base de connaissances sémiologiques. Le silence sur ces concepts peut introduire des limites. Par exemple, la fonction de pivot qui s'appuie sur ces concepts doit être modifiée. En revanche, ces concepts de rangée et de colonne causent une confusion pour plusieurs utilisateurs, entre autres lors de représentations cartographiques. Leur omission peut donc représenter une amélioration face à ce problème.

Un autre point à améliorer serait la priorisation des modèles. Comme seulement sept plus ou moins deux modèles de représentation devraient être suggérés à l'utilisateur au premier abord, une discrimination doit être faite parmi tous les modèles répondant à l'appel de la base de connaissances sémiologiques. Cette priorisation peut être établie selon le nombre de relations représentées, le nombre de fois que l'utilisateur a choisi ce modèle, le lien entre les différentes représentations (ex. un histogramme et un histogramme cumulé pour la requête subséquente) et aussi selon le choix de l'administrateur ou les besoins de synchronisation entre les différentes fenêtres ouvertes. La solution présentée ne répond pas à ce besoin de l'exploration SOLAP.

De plus, des contraintes de visualisation doivent également être décrites d'une certaine façon. Par exemple, dans l'exemple de la requête 2 (p.134) la proportion de nombre de naissances par rapport à la population devait être au minimum de 2% pour tous les membres spatiaux pour être visible dans un camembert. Ces contraintes sont généralement bien documentées dans la littérature traitant de la perception visuelle. Elles devront par contre être intégrées à la solution pour ne pas permettre les solutions qui sont invisibles à l'écran.

La visualisation géographique de l'information dans un environnement telle que SOLAP est très importante puisqu'elle constitue l'aboutissement de l'exploration des données. Celle-ci peut également être révélatrice tout comme frustrante pour un utilisateur. La complexité de la visualisation est difficile à circonscrire dans un modèle puisqu'à chaque subtilité, sa solution. La meilleure représentation pour un utilisateur n'est pas nécessairement la plus complète. Par exemple, un utilisateur pourrait préférer regarder ses données selon un histogramme même si une relation d'additivité existe dans ses données et que le camembert serait « mieux » en termes de design. Cette subjectivité doit être laissée à l'utilisateur. C'est pourquoi des options génériques doivent toujours être proposées à l'utilisateur même si ces options ne rendent pas toutes les relations présentes dans les données.

Aussi, la complexité de la visualisation augmente avec le nombre de variables à représenter. Plusieurs mesures ou croisements dans une seule image peuvent rapidement surcharger une image. Arranger l'information pour percevoir une image globale représente alors un défi de taille. La solution proposée ne gère qu'une quantité limitée de croisements et de mesures. Les problèmes résultant d'un trop grand nombre de faits ne sont pas non plus résolus par la méthode

proposée. L'abstraction devient souvent nécessaire pour représenter ces informations complexes. Des processus d'abstraction visuelle tels que l'interprétation automatique, la généralisation cartographique ou le regroupement par profil de mesures seront des options à explorer pour permettre l'abstraction visuelle.

CHAPITRE 5 - Conclusion et perspectives

5.1 Retour sur les objectifs de recherche

SOLAP se définit comme suit : « un logiciel de navigation facile et rapide dans les bases de données spatiales qui offre plusieurs niveaux de granularité d'information, plusieurs thèmes, plusieurs époques et plusieurs modes de visualisation synchronisés ou non : cartes, tableaux et diagrammes » [Bédard, 2004]. La visualisation géographique de l'information prend une place importante dans cette approche. Cependant, aucune étude, précédant ce mémoire, n'avait cherché à tirer profit des particularités de la visualisation géographique dans un contexte SOLAP. De plus, la stratégie de visualisation généralement utilisée par les SIG ne convient pas à l'environnement interactif et multidimensionnel qu'est SOLAP.

Dans un tel environnement, l'utilisateur s'appuie sur la visualisation de l'information pour explorer interactivement les données. La visualisation est donc au cœur de la relation entre les données et l'utilisateur. La stratégie de visualisation doit être adaptée à une stratégie d'exploration interactive des données (Figure 5.1). De plus, ne maîtrisant pas les concepts de bases de données et les processus d'analyse spatiale comme un expert SIG, l'utilisateur SOLAP doit avoir confiance en l'information visuelle qui lui est fournie. De la même façon que les données résultantes d'une requête doivent être exactes, l'information visuelle fournie doit aussi rendre les relations et structures de façon exacte. La relation entre les données et la visualisation doit donc être enrichie pour répondre adéquatement et efficacement à l'exploration de l'utilisateur.

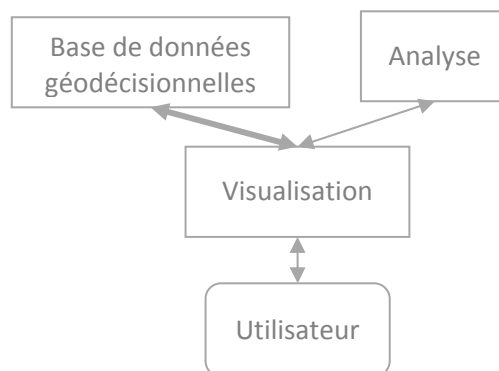


Figure 5.1 : Stratégie d'exploration interactive des données. Traduit de [Keim et al, 1996]

L'objectif général énoncé de cette recherche était d'améliorer la façon de visualiser les données géographiquement dans le but de faciliter la découverte de connaissances spatiales avec la technologie SOLAP. La réalisation de cet objectif passe par l'accomplissement de trois objectifs spécifiques.

Le premier objectif était *d'approfondir les connaissances en géovisualisation pour la technologie SOLAP*. Dans le chapitre 2, plusieurs théories soutenant la relation entre la visualisation et l'utilisateur ont été étudiées. En partant du domaine des sciences de l'information en général et en passant par l'étude de la cognition visuelle et du principe de découverte de connaissances, la revue de littérature aboutit en expliquant brièvement les théories et les principes de la géovisualisation à partir de ses origines jusqu'aux techniques développées actuellement. Cette revue met l'emphase sur l'importance de la visualisation dans le processus de réflexion d'un décideur.

Le deuxième objectif spécifique consistait à *adapter la stratégie de visualisation géographique du SOLAP à l'exploration interactive et proposer des améliorations suivant cette stratégie*. Cette stratégie implique d'introduire la visualisation au centre des composantes du système. L'utilisateur accède donc aux données par l'intermédiaire de la visualisation (Figure 5.1). Après avoir étudié l'état actuel de l'exploration des données avec SOLAP (*cf.* section 3.1), l'introduction de la visualisation géographique au cœur du processus d'exploration soulève plusieurs besoins et plusieurs possibilités (*cf.* section 3.2). À la suite de l'identification des besoins, des idées ont été proposées afin d'améliorer la visualisation géographique des données tout en respectant la philosophie d'exploration des données SOLAP (*cf.* section 3.3). La visualisation doit être facile, rapide et intuitive; elle doit répondre aux différentes requêtes spécifiques au SOLAP et elle doit également permettre de questionner visuellement les données.

La stratégie d'exploration des données demande qu'un lien plus serré soit établi entre les données et leur visualisation. Ce lien doit donc être mieux défini par une méthode de visualisation qui permettra la fluidité de l'information à travers les données, la visualisation et finalement, l'utilisateur (Figure 5.1). Le troisième objectif était donc de *concevoir une méthode de visualisation géographique des données respectant la philosophie d'exploration des données géographiques*. Le chapitre 4 décrit l'élaboration du deuxième volet de la solution théorique proposée par le présent

mémoire. La proposition consiste à concevoir une méthode dans le but d'automatiser le design d'information. L'automatisation de la visualisation géographique des données permet de dégager l'utilisateur et l'administrateur de la responsabilité du design d'information. Ceci nécessite par contre que des renseignements sur les données soient stockés et que les théories sémiologiques soient accessibles. La méthode suggère donc premièrement d'ajouter des métadonnées de visualisation à la conception du cube et d'intégrer une base de connaissances sémiologiques à la technologie SOLAP. Ces deux nouveaux acteurs collaborent ensuite pour déterminer les choix disponibles à l'utilisateur parmi les « *cent constructions pour une information* » [Bertin, 1973, p100].

L'atteinte des objectifs spécifiques permet de conclure que l'objectif général de ce mémoire a été atteint. Cette recherche facilite la découverte de la connaissance par la visualisation géographique :

- **En plaçant la visualisation au centre de la relation entre les données et l'utilisateur.** L'adoption de cette stratégie permet d'exploiter la pensée visuelle et toute sa puissance.
- **En intégrant les concepts de l'exploration interactive des données à la visualisation** (*cf.* section 3.2 et 3.3). L'utilisateur peut participer activement à la réflexion et à la recherche d'information non seulement en naviguant dans les données, mais aussi de façon visuelle.
- **En intégrant une expertise en visualisation géographique** (*cf.* CHAPITRE 4 -). Les principes de la sémiologie graphique sont alors respectés et les relations, les structures et les tendances présentées par la ou les scènes visuelles sont exactes et adaptées aux données de la requête.
- **En supportant la phase abductive du processus de construction de connaissances** (*cf.* Figure 2.8) L'utilisateur peut utiliser la visualisation géographique de façon non dirigée, tout lui est disponible en n'importe quel instant indépendamment de l'application. La relation directe entre les données et leur visualisation établie par la méthode proposée (*cf.* CHAPITRE 4 -) permet à l'utilisateur de mettre sa créativité et son intuition à profit.

5.2 Perspectives d'avenir

Cette recherche exploratoire a fait ressortir l'importance de la visualisation géographique avec SOLAP et a soulevé plusieurs questions et idées. Les itérations de cette recherche ont raffiné la solution jusqu'au moment de l'écriture de ce mémoire. Le travail est cependant loin d'être terminé.

Premièrement, des tests plus approfondis devront être réalisés pour solidifier la méthode proposée au chapitre 4. À ce point, la solution est théoriquement possible, mais est-elle réalisable pour une technologie SOLAP? Et si oui, jusqu'à quel niveau? Des essais technologiques pourraient préciser la faisabilité de la méthode proposée et valider le respect des critères d'exploration de données d'un environnement SOLAP : facilité, rapidité et intuitivité. Il serait intéressant de considérer l'algèbre de composition proposé par [Stolte, Tang et Hanrahan, 2002a, 2002b, 2002c] pour l'étendre aux dimensions géographiques et temporelles.

Ensuite, le potentiel de l'interactivité entre l'utilisateur et la visualisation doit être visité plus concrètement. Par exemple, les différentes manipulations graphiques proposées pour un environnement SOLAP dans le chapitre 3 pourrait être évaluées. Dans la littérature, beaucoup d'études font la preuve que plusieurs de ces manipulations augmentent le pouvoir de la découverte de la connaissance par la visualisation géographique. Est-ce vraiment applicable à l'environnement SOLAP? L'environnement SOLAP doit-il viser à supporter de façon universelle tous ces types de manipulations pour la découverte de la connaissance? Si elles sont utiles, quelles formes peuvent-elles prendre parmi les fonctionnalités déjà mises à place par cette approche?

Troisièmement, les utilisateurs SOLAP ont une grande importance dans le développement de l'approche visuelle. Cette dernière permet la communication entre les données et la pensée de l'utilisateur. Des études sur la pertinence des modèles visuels complexes et sur la perception des utilisateurs dans certains domaines d'application du SOLAP pourraient permettre un meilleur enrichissement des techniques de visualisation du SOLAP. Quels genres de requêtes sont réalisés? À quel point certains graphiques sont trop complexes pour un utilisateur? Est-ce que les techniques avancées de visualisation de l'information peuvent être interprétées convenablement et rapidement? À quoi un utilisateur s'attend-t-il de l'interface après une requête? Toutes ces

questions centrées sur l'utilisateur SOLAP guideraient d'une meilleure façon les solutions visuelles du SOLAP. Certains auteurs mettent d'ailleurs en doute l'efficacité des graphiques complexes proposés par les technologies modernes de visualisation et OLAP.

"You might argue that BI software takes care of effective data presentation for us by providing functionality that makes choosing the right design easy. For example, BI software usually provides formatting defaults that adorn our tables and graphs with just the right look. I wish this were true. In fact, BI software has contributed as much to the problem of bad data presentation as it has to its solution. In their frantic efforts to compete, software vendors can easily lose sight of what really matters — and what really works... [For example] in almost every case, the use of 3D actually makes data harder to read. Such charts offer a telling example of what Tufte calls "chartjunk." [Few, S. 2004].

Quatrièmement, le potentiel de l'abstraction visuelle pourrait aussi être évalué. SOLAP est une approche qui permet de passer d'un niveau détaillé dans les données à des niveaux plus abstraits. De la même façon, les représentations des données doivent pouvoir passer d'une lecture détaillée de l'image (où les questions précises trouvent des réponses exactes) à une lecture d'ensemble (où le détail de la donnée importe peu, mais la tendance générale est affichée). Trois types d'abstraction ont été identifiés : la généralisation, l'interprétation automatique et le regroupement par profil. La généralisation fait déjà l'objet de travaux de recherche au sein de la chaire industrielle en bases de données géodécisionnelles. Les autres types d'abstraction nécessitent plusieurs nouveaux développements technologiques. Un regard de ce côté pourrait être intéressant pour l'analyse de données complexes ou de grands volumes de données dans une approche SOLAP.

Finalement, la solution présentée dans ce mémoire s'applique à des données déjà transformées. Les transformations visuelles des données thématiques, géographiques ou temporelles n'ont pas été abordées. Par exemple, dans le cas des données origine/destination les données sont des origines et des destinations, mais l'information visuelle à représenter est le lien entre les deux. La transformation de données ponctuelles à des courbes de niveau nécessite également des opérations de transformation à la volée. Ce genre de visualisation pourrait faire l'objet de nouvelles investigations.

Le pouvoir de l'approche SOLAP réside dans l'intelligence et l'utilisabilité des outils visuels disponibles à l'utilisateur [Vinnik et Mansmann, 2006]. La visualisation de l'information est l'interface directe entre l'information contenue dans les données et les connaissances construites à partir de ces informations. Le présent mémoire constitue un modeste pas vers l'intégration d'une meilleure stratégie de visualisation géographique de l'information avec SOLAP.

Bibliographie¹⁹

- *Ackoff R.L. (1989) From data to wisdom, *Journal of Applied Systems Analysis*, 16, pp.3–9.
- *Andrienko G. et N. Andrienko (2005) *Exploratory Analysis of Spatial and Temporal data: a systematic approach*, Springer-Verlag, Berlin, 703p.
- *Andrienko G. L. et N.V. Andrienko (1997) IRIS: an Intelligent Tool Supporting Visual Exploration of Spatially Referenced Data, [En ligne] <http://ui4all.ics.forth.gr/UI4ALL-97/andrienko.pdf> (Page consultée à l'été 2008)
- *Andrienko G. L. et N.V. Andrienko (1998) Intelligent Visualization and Dynamic Manipulation: Two Complementary Instruments to Support Data Exploration with GIS, In: T. Catarci, M. F. Costabile, G. Santucci, and L. Taranfino (Eds) *Proceedings of the working conference on Advanced visual interfaces*, L'Aquila, Italy, May 24 - 27, AVI '98, ACM, New York, pp. 66-75
- *Andrienko G. L. et N.V. Andrienko (1999) Interactive maps for visual data exploration, *International Journal of Geographical Information Science*, 13(4), pp 355-374
- *Andrienko G. L. et N.V. Andrienko (2002) Intelligent Support of Visual Data Analysis in Descartes, *Symposium on Smart Graphics, ACM*, Juin 11-13, Hawthorne, NY, USA, pp 21-26
- *Andrienko G. L. et N.V. Andrienko (2003) Informed spatial decisions through coordinated views, *Information visualization*, 2, pp270-285
- Andrienko G., N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren et S. Wrobel (2007) Geovisual analytics for spatial decision support : Setting the research agenda, *International Journal of Geographic Information Science*, 21(8), pp. 839-857

¹⁹ Les références avec un astérisque (*) sont référencées dans le présent ouvrage, les autres ont été lues mais n'ont pas fait l'objet de référence.

-
- *Bedard Y. (1986) *A study of the nature of data using a communication-based conceptual framework of land information systems*, Thèse de doctorat, Dep. Civil Engineering University du Maine à Orono, 259 p.
- *Bédard Y. (1997) Spatial OLAP, Vidéoconférence In *2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible*, (Montréal, 13-14 novembre), Conférencier invité
- *Bédard Y. (2004) Amélioration des capacités décisionnelles des SIG par l'ajout d'un module SOLAP, *Laboratoire des Sciences de l'Information et des Systèmes (LSIS)*, Centre de Mathématiques et Informatique, Université de Provence, Marseille, 8 avril
- *Bédard Y. (2007) " OLAP et SOLAP", *Notes de cours SCG-66124 Notions avancées de base de données SIG*, Document inédit, Université Laval
- *Bédard Y. et Han J. (2008) Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, Miller H. J. Et J. Han (Eds) *Geographic Data Mining and Knowledge Discovery*, 2e édition, Chap.3, Taylor & Francis, sous presse
- *Bédard Y., E. Bernier, S. Larrivée, M. Nadeau, M.-J. Proulx et S. Rivest (2009) Spatial OLAP, [En ligne]. <http://www.spatialbi.com/> (page consulté à l'automne 2008)
- *Bédard Y., M.-J. Proulx, S. Rivest (2005) Enrichissement du OLAP pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques, *1re Journée Francophone sur les Entrepôts de données et l'Analyse en ligne (EDA, 2005)*, Université de Lyon-2, Lyon, France, 10 juin, Conférencier invité
- Bédard Y., P. Gosselin, S. Rivest, M.-J. Proulx, M. Nadeau, G. Lebel et M.-F. Gagnon (2003) Integrating GIS components with knowledge discovery technology for environmental health decision support, *International Journal of Medical Informatics*, 70, pp.79-94
- Bédard Y., S. Rivest & M.-J. Proulx (2007) Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective, Wrembel R. et C. Koncilia (Eds) *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Chap. 13, Idea Group, pp.298-319

-
- *Béguin M., Pumain D. (2003) *La représentation des données géographiques: statistique et cartographie*, Armand-Colin, Paris, 192p
- *Bellinger G., D. Castro et A. Mills (2004) *Data, Information, Knowledge and Wisdom*, [En ligne]. <http://www.systems-thinking.org/dikw/dikw.htm> (Page consultée en décembre 2008)
- *Bertin J. (1973) *Sémiologie graphique les diagrammes, les réseaux, les cartes*, Deuxième édition, Éditions Gauthier- Villars, Paris, 431p.
- *Bertin J. (1977) *La graphique et le traitement graphique de l'information*, Flammarion, France, 273 p.
- *Bimonte S. (2007) *Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation*, Thèse de Doctorat, Lyon, Institut National des sciences appliquées, 207p.
- *Bimonte S., P. Wehrle, A. Tchounikine et M. Miquel (2006) *GEWOLAP: A Web Based Spatial OLAP Proposal*, In *On the Move to Meaningful Internet Systems 2006 : OTM 200 Workshops, Lecture Notes in Computer Science 4278*, Springer Berlin, pp. 1596-1605
- *Bonham-Carter G.F. (1994) *Geographic Information systems for geoscientists: Modelling with GIS*, Pergamon, New York, 398p.
- Buja A., D. Cook et D.F. Swayne (1996) *Interactive High-Dimensional Data Visualization*, *Journal of Computational and Graphical Statistics*, 5(1), pp.78-99
- *Buzan T. et B. Buzan (2003) *Mind map: Dessine moi l'intelligence*, 2e édition, Éditions d'Organisation, Paris, 325p.
- *Card S. K., Mackinlay J. D. et Shneiderman B. (1999) *Readings in Information Visualization Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, 686p.
- *Caron P.-Y. (1998) *Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle*, Mémoire de maîtrise, Québec, Université Laval, 129 p.

-
- *Chabris C. F. et S. M. Kosslyn (2005) Representational Correspondence as a Basic Principle of Diagram Design, S.-O. Tergan et T.Keller (Eds) *Knowledge and Information Visualization*, Chap. 2, LNCS 3426, pp.36-57
- *Chrisman N. (1997) *Exploring geographic information systems*, John Wiley & Sons, New York, 298p.
- *Clark D. (2004) "Understanding" In Performance, Learning, Leadership, & Knowledge, [En ligne]. <http://www.nwlink.com/~Donclark/performance/understanding.html> (Page consultée en décembre 2008)
- *Davis B. E. (2001) *GIS: A visual approach*, 2e édition, OnWord Press, New York, 438p.
- *Denègre J. (2005) *Sémiologie et conception cartographique*, collection ENSG –IGN, Lavoisier, Paris, 274p.
- *Dent B.D. (1999) *Cartography : Thematic Map Design*, 5e edition, McGraw-Hill Companies, Boston, 417 p.
- DiBiase D. (1990) Visualization in the Earth Sciences, *Earth and Mineral Sciences*, 59 (2), [En ligne]. <http://www.geovista.psu.edu/publications/others/dibiase90/swoopy.html> (Page consultée à l'été 2007)
- *Dykes J., A.M. MacEachren, M.- J. Kraak (2005) *Exploring Geovisualization*, Elsevier Ltd., Oxford, 710p.
- Egbert S. L. et T.A. Slocum (1992) EXPLOREMAP: An Exploration System for Choropleth Maps, *Annals of the American Geographers*, 82(2), pp.275-288
- Eick S.G. (2000) Visualizing Multi-Dimensional Data, *ACM SIGGRAPH Computer Graphics*, 34(1), pp.61-67
- *Fayyad U. et G. G. Grinstein (2002) Introduction, In Fayyad U., G. G. Grinstein et A. WIERSE (2002) *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers, San Francisco, pp.1-17
- *Fayyad U., G. G. Grinstein, A. Wierse (2002) *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers, San Francisco, 407 p.

-
- *Fayyad U., G. Piatetsky-Shapiro, P. Smyth (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Communications of the ACM*, 39(11), pp.27-34
- *Fayyad U., G. Piatetsky-Shapiro, P. Smyth (1996) The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Communications of the ACM*, 39(11), pp.27-34
- *Few S. (2004) Common mistakes in Data Presentation, *Intelligent enterprise*, [En ligne] http://www.intelligententerprise.com/info_centers/bi/showArticle.jhtml?articleID=26100530 (Page consultée au printemps 2009)
- Few S. (2006) *Information Dashboard Design: The effective Visual Communication of Data*, O'Reilly Media, 223 p.
- Few S. (2007) Three Blind Men and an Elephant: The Power of Faceted Analytical Displays, *Perceptual Edge*, [En ligne]. http://www.perceptualedge.com/articles/Whitepapers/Three_Blind_Men.pdf (Page consultée à l'automne 2008)
- Filippakopoulou V., B. Nakos, E. Michaelidou et L. Stamou (2008) Evaluation of the Selectivity of Visual Variables, [En ligne]. http://www.ntua.gr/eseve/Vasikh_Ereyna/Thalis/Thalis_projects_English_summaries/Fillipakopoulou.pdf (Page consultée à l'été 2008)
- *Finke R. A., T. B. Ward et S. M. Smith (1992) *Creative Cognition: Theory, Research, and Applications*, The MIT Press, Cambridge, 239p.
- *Fortin C., R. Rousseau (1992) *Psychologie cognitive, une approche de traitement de l'information*, Télé-université, Sainte-Foy, 434p.
- *Franklin C. (1972) *An Introduction to Geographic Information Systems: Linking Maps to Databases*, *Database*, 15(2), pp.13-21
- *Fraunhofer IAIS (2007) CommonGIS, In: Fraunhofer IAIS, [En ligne] <http://www.iais.fraunhofer.de/index.php?id=1863&L=1> (Page consultée à l'hiver 2008)

-
- *Fuhrmann S., P. Ahonen-Rainio, R. M. Edsall, S. I. Fabrikant, E. L. Koua, C. Tóbon, C. Ware et S. Wilson (2005) Making Useful and Usable Geovisualization: Design and Evaluation Issues, In Dykes J., A.M. MacEachren, M.- J. Kraak (Eds) *Exploring Geovisualization*, Chap. 28, Elsevier Ltd., Oxford, pp. 553-566
- *Fuhrmann S., P. Ahonen-Rainio, R. M. Edsall, S. I. Fabrikant, E. L. Koua, C. Tóbon, C. Ware et S. Wilson (2005) Making Useful and Usable Geovisualization: Design and Evaluation Issues, In Dykes J., A.M. MacEachren, M.- J. Kraak (Eds) *Exploring Geovisualization*, Chap. 28, Elsevier Ltd., Oxford, pp. 553-566
- Gahegan M. (1999) Four barriers to the development of effective exploratory visualisation tools for the geosciences, *International journal of geographical information science*, 13(4), pp. 289-310.
- Gahegan M. (2007) Multivariate Geovisualization, J.P. Wilson et A.S. Fotheringham (Eds) *Handbook of Geographic Information Science*, Chap. 16
- Gahegan M., M. Takatsuka, M. Wheeler et F. Hardisty (2000) GeoVISTA Studio: A Geocomputational Workbench, [En ligne]. <http://www.geovista.psu.edu/publications/geocomp2000/FinalStudioPaper.pdf> (Page consultée à l'été 2008)
- *Gahegan M., M. Wachowicz, M. Harrower et T.-M. Rhyne (2001) The Integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation, *Cartography and Geographic Information Science*, 28(1): 29-44
- *Gennady A. et Dykes (2008) International Cartographic Association : Commission on Geovisualization, [En ligne] <http://geoanalytics.net/ica/> (page consultée à l'automne 2008)
- *GeoVista Center (2008) GeoViz Toolkit, In PennState University, [En ligne]. <http://www.geovista.psu.edu/geoviztoolkit/index.html> (Page consultée à l'automne 2007)
- Gilmartin P. (1992) Twenty-five Years of Cartographic Research: A Content Analysis, *Cartography and Geographic Information Systems*, 19(1), pp. 37-47

-
- Grinstein G., J. C. Sieg, S. Smith et M.G. Williams (1992) Visualization for Knowledge Discovery, *International Journal of intelligent systems*, 7, pp. 637-648
- Guo D. (2007) Visual Analytics of Spatial interaction patterns for pandemic decision support, *International Journal of Geographical Information Science*, 21(8), pp 859-877
- *Guo D., J. Chen, A.M. MacEachren, K. Liao (2006) A Visualization System for Space-Time and Multivariate Patterns (VIS-STAMP), *IEEE Transactions on Visualization and Computer Graphics*, 12(6), pp1461-1474
- *Guo D., M. Gahegan, A. M. MacEachren, B. Zhou (2005) Multivariate Analysis and Geovisualization with an Integrated Geographic Knowledge Discovery Approach, *Cartography and Geographic Information Science*, 32(2), pp. 113-132
- Harris R. L. (1999) *Information Graphics: A comprehensive Illustrated Reference, Visual Tools for Analyzing, Managing and Communicating*, Oxford University Press, New York, 448p.
- *Hearnshaw H. (1994) Psychology and display in GIS, Hearnshaw H. et D.J. Unwin (Eds) *Visualization in geographical Information Systems*, The Association for geographic information, John Wiley & Sons, Chichester, pp. 193-200
- ISO (2003) *ISO 19115: Geographic information – Metadata*, Switzerland, 140p.
- *Keim D. A. et H.-P. Kriegel (1999) VisDB: Database Exploration Using Multidimensional Visualization, In: Card S. K., Mackinlay J. D. et Shneiderman B. (Eds) *Readings in Information Visualization Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, pp.126-139
- *Keim D. A. et H.-P. Kriegel (1999) VisDB: Database Exploration Using Multidimensional Visualization, In: Card S. K., Mackinlay J. D. et Shneiderman B. (Eds) *Readings in Information Visualization Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, pp.126-139
- *Keim D. A. et H.-P. Kriegel (1999) VisDB: Database Exploration Using Multidimensional Visualization, In: Card S. K., Mackinlay J. D. et Shneiderman B. (Eds) *Readings in Information Visualization Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, pp.126-139

-
- *Keim D. A. et H.-P. Kriegel (1999) VisDB: Database Exploration Using Multidimensional Visualization, In: Card S. K., Mackinlay J. D. et Shneiderman B. (Eds) *Readings in Information Visualization Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, pp.126-139
- *Keim D. A., C. Panse, M. Sips (2005) Information Visualization: Scope, Techniques and Opportunities for Geovisualization, In Dykes J., A.M. MacEachren, M.- J. Kraak (Eds) *Exploring Geovisualization*, Chap. 2, Elsevier Ltd., Oxford, pp.23-52
- Keim D. A., C. panse, M. Sips et S.C. North (2003) PixelMaps: A New Visual Data Mining Approach for Analysing Large Spatial Data Sets, *ICDM 2003. Third IEEE International Conference on Data Mining*, 19-22 novembre, pp. 565-568
- *Keim D.A., J. P. Lee, B. Thuraisingham et C. Wittenbrink (1996) Database Issues for Data Visualization: Supporting Interactive Database Exploration, *Database Issues for Data Visualisation, Lecture Notes in Computer Sciences*, Springer Berlin/Heidelberg, pp. 12-25
- *KHEOPS Technologies (2008) Jmap Spatial OLAP, In KHEOPS Technologies, [En ligne] <http://www.kheops-tech.com/en/jmap/solap.jsp> (Page consultée à l'hiver 2008)
- *KHEOPS Technologies Technologies et Centre de recherche en géomatique (2004) Formation administrateur SOLAP, Formation offerte dans le cadre du projet BIC-SOLAP, Ministère des Transports du Québec, 11 juin
- *Kimball, R. and M. Ross (2002) The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (2nd Edition), Wiley, 464 pages
- Koedinger K.R. et J.R. Anderson (1990) Abstract Planning and Perceptual Chunks: Elements of Expertise in Geometry, *Cognitive Science*, 14, pp. 511-550
- Koua E. L., A. MacEachren et M.-J. Kraak (2006) Evaluating the usability methods in an exploratory geovisualization environment, *International Journal of Geographical Information Sciences*, 20(4), pp. 425-448
- Kraak M.-J. (2003) Geovisualization Illustrated, *Journal of Photogrammetry and Remote Sensing*, 57, pp. 390-399

-
- *Kraak M.-J. (2003a) The Space-Time Cube Revisited from a Geovisualization Perspective, *Proceedings of the 21st International Cartographic Conference (ICC)*, Durban, Afrique du Sud, 10-16 Août 2003, pp.1988-1996
- *Kraak M.-J. (2003b) Geovisualization illustrated, *Journal of Photogrammetry & Remote Sensing (ISPRS)*, 57, pp.390-399
- *Kraak M.-J. (2006) Beyond Geovisualization, *IEEE Computer Graphics and Applications*, July/August pp. 6-9
- *Kraak M.-J. (2007) Is there a need for embedded cartographic knowledge in geovisualization software? In International Institute for Geo-Information Science and Earth Observation (ITC), [En ligne]. <http://www.itc.nl/personal/kraak/publications/2007/2007-wroclaw.pdf> (Page consultée à l'automne 2008)
- Kraak M.-J. et A. MacEachren (1999) Visualization for exploration of spatial data, *International Journal of Geographical Information Sciences*, 12(4) pp.285-287
- *Kraak M.-J., Omerling F. (2003) *Cartography : Visualization of Geospatial Data*, 2 édition, Pearson Education Ltd, Essex, 205 p.
- *Lambert M. (2006) *Développement d'une approche pour l'analyse SOLAP en temps réel: Adaptation aux besoins des activités sportives en plein air*, Mémoire de M.Sc., Dép. Sciences géomatiques, Québec, Université Laval, 139p.
- *Larkin J.H., H.A. Simon (1987) Why a Diagram is (Sometimes) Worth Then Thousand Words, *Cognitive Science*, 11, pp. 65-99
- *Le Petit Robert (2000) *Sémiologie*, p. 2318
- *Le Robert (2002) *Analyse*, p.58
- *MacEachren A. M. (1994b) Visualization in Modern Cartography: Setting the agenda, In MacEachren A. M., D.R.F. Taylor (Eds) *Visualisation in modern cartography*, Pergamon, Oxford, pp.1-12

-
- MacEachren A. M. et J. H. Ganter (1990) A Pattern Identification Approach to Cartographic Visualization, *Cartographica*, 27(2), pp. 64-81
- MacEachren A.M. (1992) Application of Environmental Learning Theory to Spatial Knowledge Acquisition from Maps, *Annals of the Association of the American Geographers*, 82(2), pp. 245-274
- *MacEachren A.M. (1994a) *Some Truth with Maps: A Primer on Symbolization & Design*, Association of the American Geographer (AAG), Washington D.C., 129 p.
- *MacEachren A.M. (1995) *How maps work: representation, Visualization and Design*, The Guilford Press, New York, 513p.
- *MacEachren A.M. et M. J. Kraak (1997) Exploratory cartographic visualization: advancing the agenda, *Computers & geosciences*, 23(4), pp 335-343
- *MacEachren A.M. et M.-J. Kraak (2001) Research Challenges in Geovisualization, *Cartography and Geographic Information Science*, 28(1)
- MacEachren A.M., C. Polsky, D. Haug, D. Brown, F. Boscoe, J. Beedasy, L. Pickle et M. Marrara (1997) Visualizing Spatial Relationships Among Health, Environmental, and Demographic Statistics: Interface Design Issues, *Proceedings of the 18th International Cartographic Conference*, Stockholm, 23-27 Juin
- *MacEachren A.M., M. Gahegan, W. Pike, I. Brewer, G. Cai et E. Lengerich (2004) Geovisualization for Knowledge Construction and Decision Support, *IEEE Computer Graphics and Applications*, 24(1), pp. 13-17
- *MacEachren A.M., M. Wachowicz, R. Edsall, D. Haug, R. Masters (1999) Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with discovery in database methods, *International Journal of Geographical Information Science*, 13(4), pp. 311-334
- *MacKinlay J. (1986) Automating the Design of Graphical Presentations of Relation Information, *ACM transactions on Graphics*, 5(2) pp.110-141

-
- *MacKinlay J. (2008) "Designing Great Visualisations" In Tableau Software, [En ligne]. <http://www.tableausoftware.com/videos/designing> (Page consultée à l'automne 2008)
- *Mackinlay J.D., P. Hanrahan et C. Stolte (2007) Show Me: Automatic Presentation for Visual Analysis, *IEEE Transactions on visualization and computer graphics*, 13(6), pp.1137-1144
- *Malinowski E. et E. Zimanyi (2008) *Advances Data Warehouse Design: From Conventional to Spatial and Temporal Applications*, Springer, 435p.
- *Marchand P. (2004) *The spatio-temporal topological operator dimension, a hyperstructure for multidimensional spatio-temporal exploration and analysis*, Thèse de doctorat, Québec, Dép. Sciences géomatiques, Université Laval, 108p.
- *McHugh R.-M. (2008) *Intégration de la structure matricielle dans les cubes spatiaux*, Mémoire de M.Sc., Québec, Dép. Sciences géomatiques, Université Laval, 126 p.
- *Mc Hugh R.-M., F. Bilodeau, S. Rivest, Y. Bédard (2006) *Analyse du potentiel d'une application SOLAP pour une gestion efficace de l'érosion des berges en Gaspésie Îles-de-la-Madeleine*, Géomatique 2006, 25-26 octobre 2006, Montréal, Canada.
- *Meehan A. (2007a) Performance Management with Spatial Analysis (Part 1), Directions Magazine, 4 avril 2007, [En ligne]. http://www.directionsmag.com/article.php?article_id=2441 (Page consultée à l'été 2007)
- *Meehan A. (2007b) Performance Management with Spatial Analysis (Part 2), Directions Magazine, 12 avril 2007, [En ligne]. http://www.directionsmag.com/article.php?article_id=2444&trv=1 (Page consultée à l'été 2007)
- *Miller H. J. et J. Han (2001) *Geographic Data Mining and Knowledge Discovery*, Taylor & Francis, London, 372p.
- *Miller H. J. et J. Han (2001) Geographic data mining and knowledge discovery: an overview, In Miller, H. J. et J. Han (Eds) *Geographic data mining and knowledge discovery*, Taylor & Francis, London, pp. 3-32

-
- Miller H.J. (2007) Geographic Data Mining and Knowledge Discovery, J.P. Wilson et A.S. Fotheringham (Eds) *Handbook of Geographic Information Science*, Chap. 19
- *Mirel B. (1998) Visualization for data exploration and Analysis : A Critical Review of Usability Research, *Technical Communication*, 45 (4), pp. 491-509
- Murray B.S. et E. MacDaid (1993) Visualizing and representing knowledge for the end user: a review, *International Journal of Man-Machine Studies*, 38, pp. 23-49
- *National Research Council (2006) *Learning How to Think Spatially*, The National Academies Press, Washington D.C., 313 p.
- *Nelson E. S. (2000) The Impact of Bivariate Symbol Design on Task Performance in a Map Setting, *Cartographica*, 37(4), pp.61-77
- *Newell A. (1990) *Unified Theories of Cognition*, Harvard University Press, Chambridge, 572 p.
- *NIST/SEMATECH (2006) "What is EDA?" In Engineering Statistics Handbook, [En ligne]. <http://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm> (Page consultée à l'automne 2008)
- *OGC (2006) OpenGIS Symbology Encoding Implementation Specification, In Open GIS Standards, [En ligne] <http://www.opengeospatial.org/standards/symbol> (Page consultée en décembre 2008)
- *OQLF (1973) Sémiologie graphique: [En ligne] www.granddictionnaire.com
- Patil G.P., J. Balbus, G. Biging, J. Jaja, W.L. Myers et C. Taillie (2004) Multiscale advanced raster map analysis system: Definition, design and development, *Environmental and Ecological Statistics*, 11, pp. 113-138
- *Pendse N. et C. Bange (2007) The missing 'Next Big Thing', In: The OLAP Report, [En ligne] <http://www.olapreport.com/Faileddozen.htm> (Page consultée à l'hiver 2008)
- Peterson M.P. (2008) International Perspectives on Maps and the Internet, Lecture Notes in Geoinformation and Cartography, Springer-Verlag Berlin Heidelberg, New York, 441p.

-
- Peuquet D. J. et M.-J. Kraak (2002) Geobrowsing: creative thinking and knowledge discovery using geographic visualization, *Information Visualization*, 1, pp.80-91
- Peuquet D.J. (2001) Making Space for Time: Issues in Space-Time Data Representation, *GeoInformatica*, 5(1) pp. 11-32
- *Plaisant C. (2005) Information Visualization and the Challenge of Universal Usability, Dykes J., A.M. MacEachren, M.- J. Kraak (Eds) *Exploring Geovisualization*, Chap. 3, Elsevier Ltd., Oxford, pp.53-82
- Proulx M.-J. et Y. Bédard (2004a) Le potentiel de l'approche multidimensionnelle pour l'analyse de données géospatiales en comparaison avec l'approche transactionnelle des SIG, *Colloque géomatique 2004- Un choix stratégique! Montréal, 27-28 octobre, Montréal, Canada*
- *Proulx, M.J. et Y. Bédard (2004b) *Analyses décisionnelles géospatiales des données du registre de l'Université Laval*, Rapport de recherche pour Bureau de planification des études institutionnelles, Université Laval, Octobre, 22p.
- *Proulx, M.J., S. Rivest et Y. Bédard (2008) *Développement d'un prototype d'outil d'analyse de données à référence spatiale sur le transport multimodal des marchandises*, rapport d'activité pour le Ministère des transports du Québec, Université Laval, Avril, Version 1.2, 54 pages
- *Rhyne T.M. (2003) Does the difference between Information and Scientific Visualization Really Matter?, *IEEE Computer Graphics and Applications*, May/June 2003, pp6-8
- Rhyne T.M., A. MacEachren et J. Dykes (2006) *Exploring Geovisualization*, IEEE Computer Graphics and Applications, Juillet/Août, pp. 20-21
- *Rivest S., Y. Bédard, M.-J. Proulx, M. Nadeau, F. Hubert et J. Pastor (2005) SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data, *Journal of Photogrammetry & Remote Sensing (ISPRS)*, 60, pp. 17-33
- *Rivest S., Y. Bédard, P. Marchand (2001) Toward Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP), *Geomatica*, 55(4), pp. 539-555

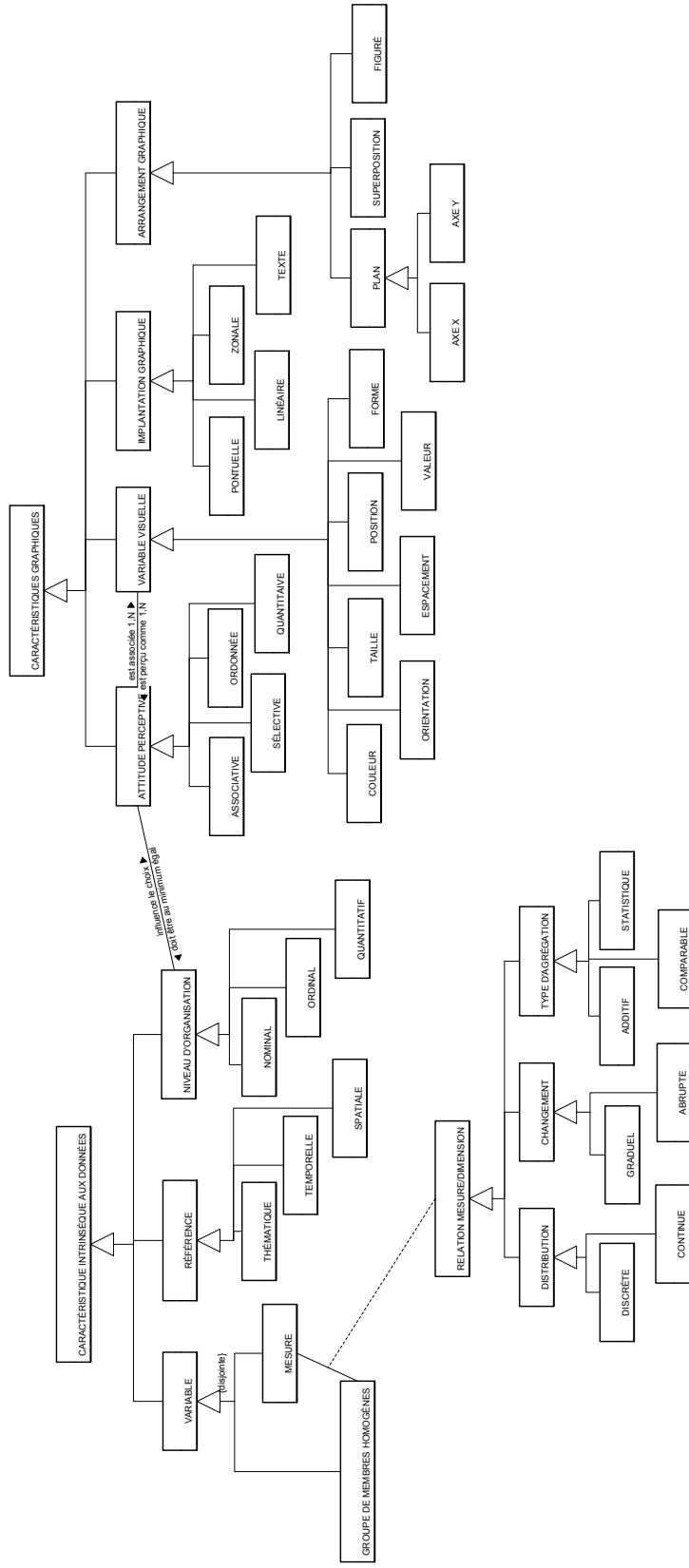
-
- *Rowley J. (2007) The wisdom hierarchy: representations of the DIKW hierarchy, *Journal of Information Science*, 33(2), pp163-180
- *Salehi M. (2009) *Developing a Model and a Language to Identify and Specify the Integrity Constraints in Spatial Datacubes*, Thèse de Doctorat, Québec, Dép. Sciences géomatiques, Université Laval
- *Scaife M. et Y. Rogers (1996) External Cognition: How do graphical representations work? , *International Journal of Human-Computer Studies*, 45, pp185-213
- *Scotch M. et B. Parmanto (2005) SOVAT: Spatial OLAP Visualization and Analysis Tool, *Proceedings of the 38th Annual Hawaii International Conference on System Sciences HIC SS'05*, January 3 -6th
- *Shneiderman B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, *Proceedings of the IEEE Symposium on Visual Languages, Boulder*, 3-6 Sept., pp. 336-343
- *Shneiderman B. (2001) Supporting Creativity with Advanced Information-Abundant User Interfaces, In Earnshaw r., R. Guedj, A. Van Dam et J. Vince (Eds) *Human-Centered Computing, Online Communities, and Virtual Environments*, Springer-Verlag, Londres, pp.469-480
- Sips M., J. Schneidewind et D.A. Keim (2007) Highlighting space-time patterns: Effective visual encodings for interactive decision-making, *International Journal of Geographical Information Science*, 21(8), pp 879-893
- *Slocum T.A., R.B. McMaster, F.C. Kessler, H.H. Howard (2005) *Thematic Cartography and geographic visualization*, 2e edition, Pearson Prentice Hall, Upper Saddle River, 518 p.
- *Spatial Analysis Lab (2006) GeoDA: An Introduction to Spatial Data Analysis In University of Illinois [En ligne] <https://www.geoda.uiuc.edu/> (Page consultée à l'hiver 2008)
- *Spence R. (2001) *Information Visualization*, ACM Press, Essex, 206 p.
- Stasko J. (2006) Note de cours CS 7450 Information Visualization, Georgia Institute of Technology, [En ligne]. <http://www.cc.gatech.edu/~john.stasko/7450/syllabus.html> (Page consultée à l'été 2008)

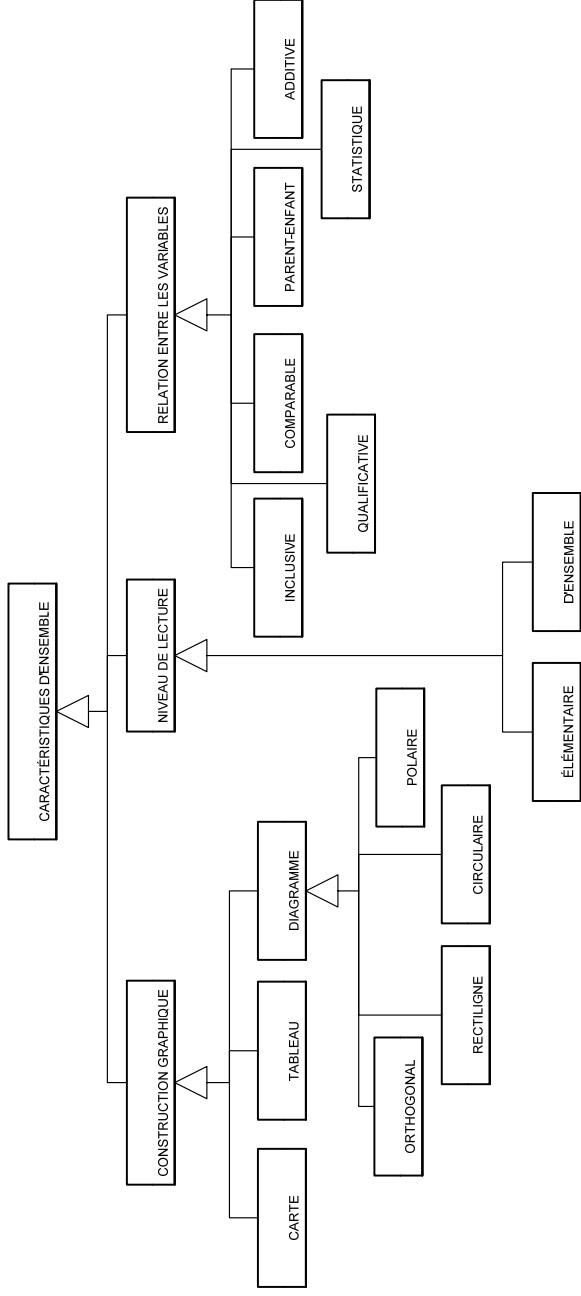
-
- Stefan H., S. René, R. Marianne et H. Lorenz (2008) Multivariate Mapping in High Quality Atlases [En ligne]. http://www.atlasderschweiz.ch/pdf_publi/ICC07_HuberSieber.pdf (Page consultée à l'été 2008)
- Stolte C. (2008) The Zen of Visual Analysis In Tableau Software, [En ligne]. <http://www.tableausoftware.com/videos/zen> (Page consultée à l'automne 2008)
- *Stolte C., D. Tang et P. Hanrahan (2002a) Polaris: A system for query, Analysis, and Visualization of Multidimensional Relational Databases, *IEEE Transactions on Visualization and Computer Graphics*, 8(1), pp.52-65
- *Stolte C., D. Tang et P. Hanrahan (2002b) Query, Analysis, and visualization of Hierarchically Structured Data using Polaris, *Proceedings of the ACM SIGKDD '02*, Edmonton, Alberta, Juillet
- *Stolte C., D. Tang et P. Hanrahan (2002c). Multiscale Visualization Using Data Cubes, *Proceedings of the IEEE Symposium on information Visualization (infovis'02)*, 28 - 29 Octobre, IEEE Computer Society, Washington, DC, pp. 7-14
- *Tergan S.-O. et T.Keller (2005) *Knowledge and Information Visualization*, Springer-Verlag Berlin Heidelberg, 384p.
- Thomas J.J. and K.A. Cook (2005) *Illuminating the path: The Research and Development Agenda for Visual Analytics*, National Visualization and Analytics Center, IEEE, Los Alamitos, 190 p.
- *Thomsen E. (2002) *OLAP Solutions: Building Multidimensional Information Systems*, 2e édition, Wiley, 688p.
- *Tobelem-Zanin C., M.-L. Trémélo (2003) *Savoir faire une carte: aide à la conception et à la réalisation d'une carte thématique univariée*, Belin, Paris, 199p.
- Tomaszewski B.M., C. Weaver, A.C. Robinson, M. Stryker et A.M. MacEachren (2007) Geovisual Analytics and Crisis Management, B. Van de Walle, P. Burghardt and C. Nieuwenhuis (Eds) *Proceedings ISCRAM2007*, pp. 173-179

-
- *Tricot C. (2008) Principe de l'information cartographique In knowledge-mapping.net, [En ligne].
http://www.knowledge-mapping.net/index.php?option=com_content&task=view&id=26&Itemid=1 (Page consultée à l'été 2008)
- *Tufte E. (1990) *Envisioning Information*, Graphics Press, Cheshire, 126p.
- Tufte E. (1997) *Visual Explanation: Images and Quantities, Evidence and Narrative*, Graphics Press, Cheshire, 156p
- Tufte E. (2001) *The Visual Display of Quantitative Information*, 2e édition, Graphics Press, Cheshire, 197p.
- *Tufte E. (2006) *Beautiful Evidence*, Graphics Press LLC, Cheshire, 213 p.
- *Tukey J.W. (1977) *Exploratory Data Analysis*, Addison-Wesley Reading, 688p.
- *UCMAS Canada (2007) Brain development, [En ligne]
http://www.ucmascanada.com/page/Brain_Development/rmenu_id,20/ (Page consultée à l'automne 2008)
- *Veilleux J.-P. (2007) Détermination optimale des positions, des vitesses et des accélérations d'un athlète par mesures GPS, Mémoire de M.Sc., Québec, Dép. Sciences géomatiques, Université Laval, 125 p.
- *Vinnik S. et F. Mansmann (2006) From Analysis to Interactive Exploration: Building Visual Hierarchies from OLAP Cubes, *Advances in database technology -EDBT 2006, Lecture notes in Computer Sciences*, 3896, pp.496-514
- *Ware C. (2000) *Information Visualization: perception for Design*, Academic Press, London, 438p.
- *Ware C. (2005) Visual Queries: The Foundation of Visual Thinking, S.-O. Tergan et T.Keller (Eds) *Knowledge and Information Visualization*, Chap. 1, LNCS 3426, pp. 27-35
- *Ware C. (2008) *Visual Thinking for Design*, Elsevier, Burlington, 197p.

-
- *Wikipédia (2008) Pensée visuelle, [En ligne] http://fr.wikipedia.org/wiki/Pens%C3%A9e_visuelle (Page consultée à l'automne 2008)
- *Wong P. C. et R.D. Bergeron (1997) 30 Years of Multidimensional Multivariate Visualization, In G. M. Nielson, H. Hagen, et H. Müller (Eds) *Scientific Visualization, Overviews, Methodologies, and Techniques*, IEEE Computer Society, Washington, pp. 3-33
- Wright P. et K. Fox (1970) Presenting information in tables, *Applied Ergonomics*, 1(4), pp 234-242
- *Wurman R. S. (1989) *Information Anxiety*, Doubleday, 356 p.
- Xie Z., S. Huang, M.O. Ward et E.A. Rundensteiner (2006) Exploratory Visualization of Multivariate Data with Variable Quality, *IEEE Symposium on Visual Analytics Science and Technology 2006*, 31 Octobre, Baltimore, E.-U., pp. 183-190
- Zhan F. B. et B. P. Buttenfield (1995) Object-Oriented knowledge-based symbol selection for visualizing statistical information, *International Journal of Geographical Information Systems*, 9(3), pp. 293-315
- *Zins C. (2007) Conceptual Approaches for Defining Data, Information and Knowledge, *Journal of the American Society for Information Science and Technology*, 58(4), pp. 479-493

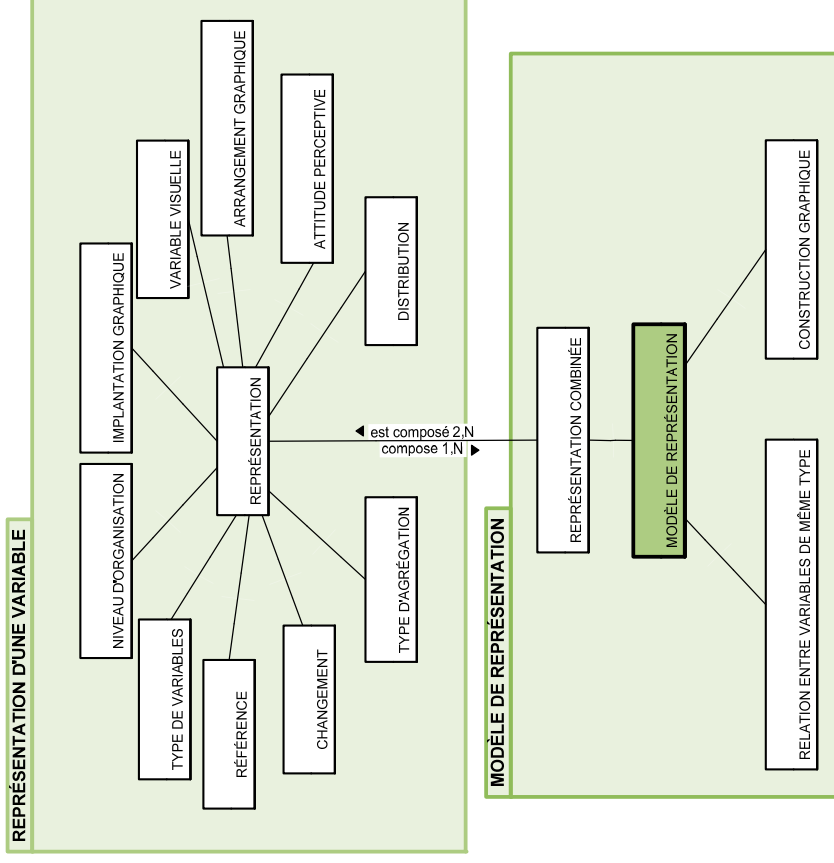
Annexe 1 - Taxonomie des concepts



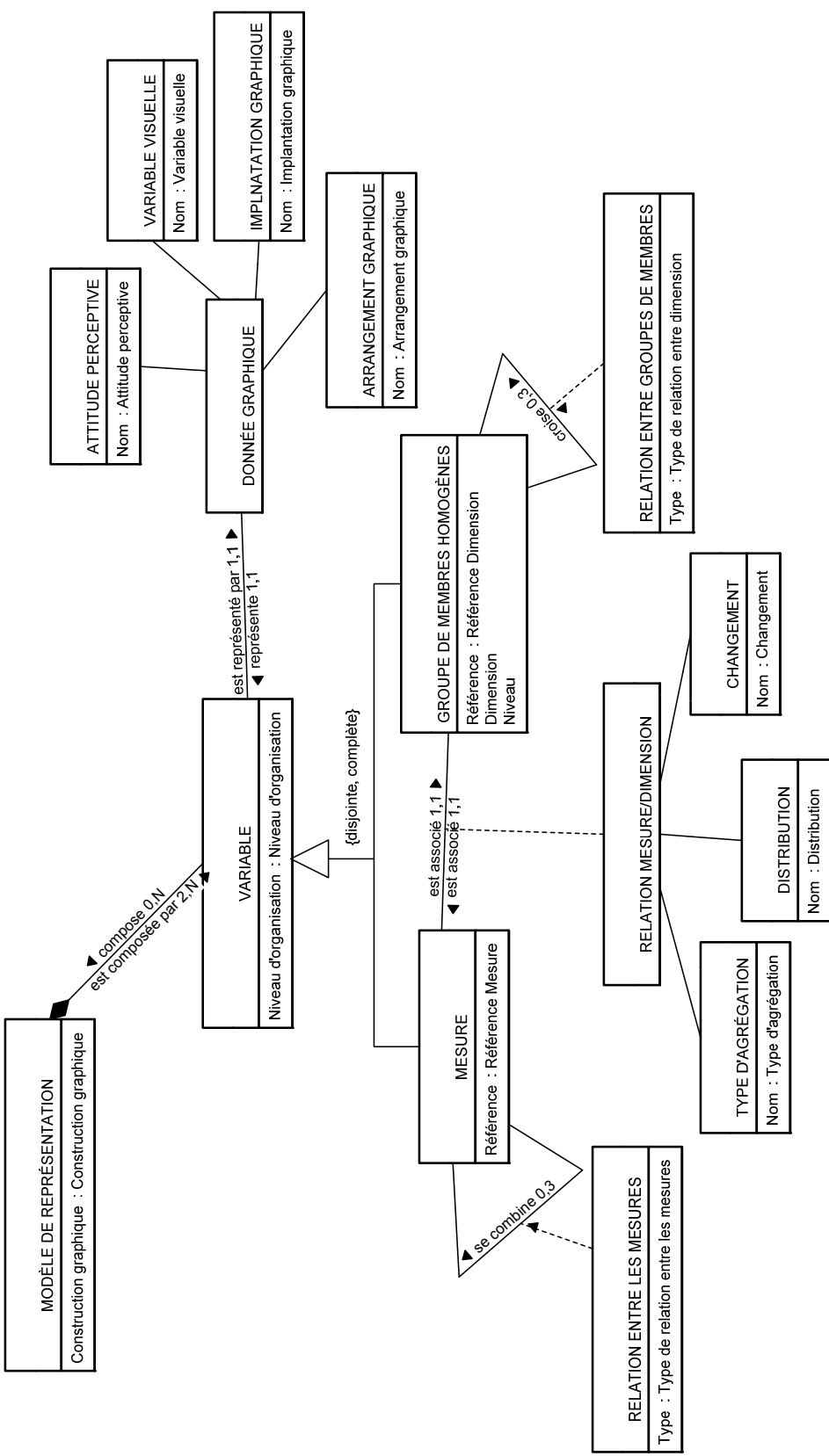


Annexe 2 - Modèle Conceptuel de la base de connaissances sémiologiques

Version antérieure



Version récente



<<ÉNUMÉRATION>> NIVEAU D'ORGANISATION
1 = Nominal 2 = Ordinal 3 = Quantitatif

<<ÉNUMÉRATION>> RÉFÉRENCE MESURE
1 = Thématique 2 = Spatiale

<<ÉNUMÉRATION>> RÉFÉRENCE DIMENSION
1 = Spatial 2 = Temporel 3 = Thématique

<<ÉNUMÉRATION>> ATTITUDE PERCEPTIVE
1 = associative 2 = sélective 3 = ordonnée 4 = quantitative

<<ÉNUMÉRATION>> IMPLANTATION GRAPHIQUE
1 = Ponctuelle 2 = Linéaire 3 = Zonale 4 = Texte

<<ÉNUMÉRATION>> VARIABLE VISUELLE
1 = Couleur 2 = Forme 3 = Valeur 4 = Taille 5 = Orientation 6 = Espacement 7 = Position

<<ÉNUMÉRATION>> ARRANGEMENT GRAPHIQUE
1 = Axe X 2 = Axe Y 3 = Axes 4 = Superposition 5 = Figuré

<<ÉNUMÉRATION>> CHANGEMENT
1 = Graduel 2 = Abrupt

<<ÉNUMÉRATION>> DISTRIBUTION
1 = Continue 2 = Discrete

<<ÉNUMÉRATION>> TYPE D'AGRÉGATION
1 = Comprable 2 = Additive 3 = Statistique





<<ÉNUMÉRATION>> TYPE DE RELATION ENTRE DIMENSION
1 = parent-enfant 2 = Non-Comparable





<<ÉNUMÉRATION>> TYPE DE RELATION ENTRE LES MESURES
1 = inclusive 2 = Comparable 3 = Qualificative 4 = Statistique 5 = Additive






<<ÉNUMÉRATION>> CONSTRUCTION GRAPHIQUE
1 = Carte 2 = Tableau 3 = Diagramme orthogonal 4 = Diagramme Rectiligne 5 = Diagramme circulaire 6 = Diagramme polaire







Annexe 3 - Description des modèles de représentation

A3.1 Légende des symboles utilisés

Référence : Spatiale - polygone , ligne , point , Temporelle - 

Implantation graphique : zonale , linéaire , ponctuelle  et texte **A** 

Variable visuelle : taille , valeur , couleur , orientation , forme 

Construction graphique : Carte , Diagrammes orthogonal , linéaire , circulaire  et polaire , tableau 

A3.2 Représentations simples




La description des différents modèles de représentation est présentée sous forme de tableau afin de réduire l'espace demandé pour chacune de ces descriptions. En premier lieu, un logo illustre le modèle de représentation décrit, puis, les caractéristiques intrinsèques des données, les relations entre mesure et dimension, les caractéristiques graphiques et les caractéristiques d'ensemble sont présentées en bloc. Les symboles présentés ci-dessus sont utilisées pour décrire les caractéristiques. La figure ci-dessous explique la lecture du tableau.




LOGO	Caractéristiques intrinsèques	Relation mesure/dimension	Caractéristique d'affichage	Caractéristiques d'ensemble
Modèle de représentation générique	Description des variables pour lesquelles le modèle de représentation est adapté.	Description de la relation entre chaque couple de variable dépendante/indépendante. Comme ces caractéristiques sont en relation 1,2 avec les variables, les cases sont grisées pour la variable de référence.	Description des caractéristiques d'affichage. Les caractéristiques d'affichage sont en relation 1,1 avec les variables à représenter.	Description de la construction graphique qui s'applique pour tous les modèles de représentation (relation 1, N) et de la relation entre les variables de même type qui s'applique seulement lorsque plusieurs variables indépendantes ou dépendantes sont présentes (relation 1,2).




I) Cartes simples




Les cartes simples encodent une mesure et un groupe de membres spatial.





Cartes choroplèthe





Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal			Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Quantitative	Continue	-	Quantitative		Taille	Figuré		

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal			Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Ordinal	Continue	-	Ordonnée		Valeur	Figuré		





Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal			Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Ordinal	Continue	-	Ordonnée		Couleur	Figuré		

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal			Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Qualitatif	Continue	Abrupt	Selective		Orientation	Figuré		

	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation Externe
Groupe de membres		Ordinal					Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Qualitatif	Continue	Abrupt	-		Associative		Forme	Figuré		





	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation Externe
Groupe de membres		Spatial	Ordinal				Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Quantitative	Discret	-	Statistique		Ordonnée		Couleur et valeur*	Figuré		





*La couleur et la valeur forment une symbologie divergente.





	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation Externe
Groupe de membres		Spatial	Ordinal				Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Quantitative	Continue	-	statistique		Quantitative		Taille et forme*	Figuré		





*Les variables visuelles forment une symbologie divergente.

Cartes linéaires





	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation Externe
Groupe de membres		Spatial	Ordinal				Ordonnée		Position	Plan	Carte	
Mesure 1	Thématique	Quantitatif	Discrete	-	-		Quantitative		Taille	Figuré		



	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal					Ordonnée		Position	Plan	Carte
Mesure 1	Thématique	Ordinal	Discrète	-	-	-	Ordonnée		Valeur	Figuré		



	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal					Ordonnée		Position	Plan	Carte
Mesure 1	Thématique	Ordinal	Discrète	-	-	-	Ordonnée		Couleur	Figuré		



	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal					Ordonnée		Position	Plan	Carte
Mesure 1	Thématique	Qualitatif	Discrète	Abrupt	-	-	Associative		Forme	Figuré		



Cartes de symboles

	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptuelle	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal					Ordonnée		Position	Plan	Carte
Mesure 1	Thématique	Quantitatif	Discrète	-	-	-	Quantitative		Taille	Figuré		



Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres Mesure 1	 Thématique	Ordinal	Discrete	-	Ordonnée	 Position	Position	Plan	Carte

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres Mesure 1	 Thématique	Ordinal	Discrete	-	Ordonnée	 Position	Position	Plan	Carte

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres Mesure 1	 Thématique	Ordinal	Discrete	-	Ordonnée	 Position	Position	Plan	Carte

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres Mesure 1	 Thématique	Ordinal	Discrete	Abrupt	Ordonnée	 Position	Position	Plan	Carte

Cartes descriptives (implantation texte A)

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	 Thématique	Ordinal				Ordonnée	 Position	Position	Plan	Carte

Mesure 1	Thématique	Quantitatif	Discrète	Abrupt	-	Quantitative	A	Taille	Figuré	Caractéristiques d'ensemble
Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage		Arrangement graphique		Construction graphique	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Relation
Groupe de membres		Ordinal				Ordonnée		Position	Plan	Carte
Mesure 1	Thématique	Ordinal	Discrète	Abrupt	-	Ordonnée	A	Couleur	Figuré	

Exemple de cartes avec une combinaison redondante de variables visuelles

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Relation
Groupe de membres		Ordinal				Ordonnée		Position	Plan	
Mesure 1	Thématique	Qualitatif	Discrète	Abrupt	Comparable	Sélective	•	Forme et couleur	Figuré	Carte

Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Relation
Groupe de membres		Ordinal				Ordonnée		Position	Plan	
Mesure 1	Thématique	Qualitatif	Continue	Abrupt	-	Sélective		Orientation et couleur	Figuré	Carte





Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble		
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Relation
Groupe de membres		Ordinal				Ordonnée		Position	Plan	
Mesure 1	Thématique	Ordinal	Continue	-	-	Ordonnée		Taille et espacement	Figuré	Carte

L'utilisation des combinaisons redondantes de variables visuelles est suggérée lorsque le nombre de classes est plus grand que six. Pour plusieurs variables visuelles, la combinaison redondante permet d'augmenter la séparation visuelle des classes. Elle trouve toute son utilité lorsqu'une attitude perceptive sélective est mise de l'avant. Comme les deux variables visuelles varient ensemble lors d'une combinaison redondante, il importe peu que la combinaison soit intégrable ou séparable. Toutes les combinaisons de variables visuelles sont donc possibles.





II) Cartes simples particulières

Les cartes simples particulières transforment l'implantation géométrique des membres spatiaux pour révéler l'information différemment.

Cartogramme





	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble				
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal				Ordonnée		Position	Plan	Carte	
	Mesure 1	Thématique	Quantitatif	Continue	-	-	Quantitatif		Taille	Figuré		

Carte d'isolignes

	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal				Ordonnée		Position	Plan	Carte	
	Mesure 1	Thématique	Quantitatif	Continue	Graduel	-	Quantitative		Position	Figuré		

Les cartes d'isolignes nécessitent un calcul d'interpolation entre les valeurs discrètes et la représentation continue. Ce calcul doit être fait à la volée lors de chaque requête.

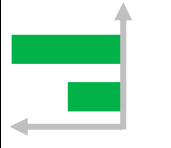
Carte par semis de points

	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Ordinal				Ordonnée		Position	Plan	Carte	
	Mesure 1	Thématique	Quantitatif	Discrete	Graduel	-	Quantitative		Position	Figuré		

Les cartes par semis de points sont utiles lors de l'étude d'un phénomène détaillé à petite échelle. Comme chaque point représente une valeur fixe, un calcul de regroupements (*clustering*) des données doit être effectué à la volée pour la localisation des points dans l'image. Lorsque la valeur d'un objet dépasse la valeur attribuée au point, plusieurs points représenteront l'objet. Un calcul de déplacement est alors nécessaire pour rendre visible tous les points représentant la valeur de l'objet.

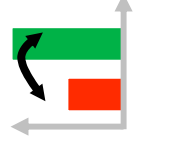
III) Diagrammes simples

Histogramme

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal	-	-	-	Ordonnée	■	Position	Axe X	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative	■	Position	Axe Y		

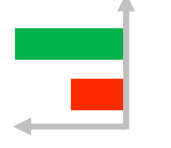
Si la distance entre les classes sur l'axe horizontal est importante, on considère l'attitude perceptue quantitative. Si la distance n'est pas importante, l'attitude perceptue est ordonnée. Les données qualitatives doivent être ordonnées d'une certaine façon pour permettre la perception ordonnée.

Histogramme ordonné

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Qualitatif	-	-	-	Ordonnée	■	Position et Couleur	Axe X	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative	■	Position	Axe Y		

La couleur peut être utilisée pour encoder de l'information dans le diagramme. Dans le diagramme ci-dessus, la couleur varie selon les membres représentés.

Histogramme

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal	-	-	-	Sélective	■	Couleur et Position	Axe X	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative	■	Position	Axe Y		

De la même façon que pour les cartes simples, des combinaisons de variables redondantes (position et couleur) peuvent aider la sélection visuelle.

Histogramme horizontal

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal				Ordonnée			Axe Y	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	--	-	Comparable	Quantitative			Position	Axe X	

Diagramme à ligne

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres		Quantitatif				Quantitative			Axe X	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	Continue	-	Comparable	Quantitative			Position	Axe Y	

Barre cumulée

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal				Ordonnée			Axe Y	Diagramme Rectiligne	
	Mesure 1	Thématique	Quantitatif	-	-	Additif	Quantitative			Position	Axe Y	

Diagramme rectiligne

	Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage			Caractéristiques d'ensemble		
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Qualitatif				Associative			Axe X	Diagramme Rectiligne	
	Mesure 1	Thématique	Quantitatif	Discrete	-	-	Quantitative			Position	Axe X	

Camembert ou pointe de tarte


	Caractéristiques intrinsèques			Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal				Sélective	◀	Position et Couleur	Axe X	Diagramme Circulaire	
	Mesure 1	Thématique	Quantitatif	-	-	Additif	Quantitative	▬	Taille	Axe Y		

Diagramme en secteur


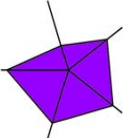

	Caractéristiques intrinsèques			Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Qualitatif				Sélective	◀	Position et couleur	Axe X	Diagramme Polaire	
	Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative	▬	Position	Axe Y		

Diagramme en Étoile


	Caractéristiques intrinsèques			Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal				Associative	▬	Position	Axe X	Diagramme polaire	
	Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative	▬	Position	Axe Y		

Histogramme comparatif

	Caractéristiques intrinsèques			Relation mesure/dimension		Caractéristique d'affichage			Caractéristiques d'ensemble			
	Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
	Groupe de membres	Thématique	Ordinal				Ordonnée	▬	Position	Axe X	Diagramme orthogonal	
	Mesure 1	Thématique	Quantitatif	-	-	Statistique	Quantitative	▬	Position	Axe Y+ et Y-		

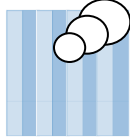
IV) Tableaux simples

Tableau simple



Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble		
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective	A	Position	Axe X	Tableau	
Mesure 1	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	-	Sélective	A	Forme	Figuré		

Tableau de taille



Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble		
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective	A	Position	Axe X	Tableau	
Mesure 1	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	-	Quantitative	•	Taille	Figuré		

Tableau simple avec couleur




Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble		
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective	A	Position	Axe X	tableau	
Mesure 1	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	-	Ordonnée	A	Position et couleur	Figuré		


Les tableaux peuvent aussi utilisés des combinaisons de variables redondantes entre autres pour la synchronisation des symboliques.


V) Représentations multivariées



Cartes multivariées




Les cartes multivariées implique d'utiliser deux variables visuelles pour encoder deux informations différentes. La combinaison des variables est donc significative. Pour une perception sélective les combinaisons doivent être séparables (taille et valeur, taille et orientation, orientation et valeur, taille et couleur, couleur et forme). Pour une perception associative, les combinaisons doivent être intégrables (taille en X et en Y, orientation et couleur).

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal				Ordonnée	◀	Position	Plan	Carte	Tous
Mesure 1	Thématique	Quantitatif	Continue	Abrupt	-	Quantitative	◀	Taille	Figuré		
Mesure 2	Thématique	Ordinal	Continue	-	-	Ordonnée	◀	Valeur	Figuré		

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal				Ordonnée	◀	Position	Plan	Carte	Tous
Mesure 1	Thématique	Quantitatif	Continue	-	-	Quantitative	◀	Taille	Figuré		
Mesure 2	Thématique	Qualitatif	Continue	Abrupt	-	Sélective	◀	Orientation	Figuré		




Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Ordinal				Ordonnée	◀	Position	Plan	Carte	Comparable, statistique, qualitative
Mesure 1	Thématique	Ordinal	Continue	-	-	Ordonnée	◀	Couleur	Figuré		
Mesure 2	Thématique	Qualitatif	Continue	Abrupt	-	Sélective	◀	Orientation	Figuré		

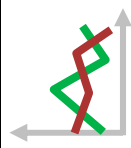
Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Mesure 1	Thématique	Quantitatif	Discret	-	Quantitative	•	Taille	Figuré		
Mesure 2	Thématique	Quantitatif	Discret	-	Quantitative	•	Taille	Figuré	Carte	Inclusive
Groupe de membres		Ordinal			Ordonnée		Position	Plan		





Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Mesure 1	Thématique	Quantitatif	-	-	Quantitative	•	Taille	Figuré et Axe X		
Mesure 2	Thématique	Quantitatif	-	-	Quantitative		Taille	Figuré et Axe Y	Diagramme Circulaire	Inclusive
Groupe de membres		Ordinal			Ordonnée		Position	Plan	Carte	

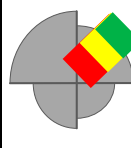
VI) Diagrammes multivariés




Les diagrammes multivariés consistent à encoder une information supplémentaire dans un diagramme simple. Cette information peut être encodé par l'ajout d'une nouvelle variable visuelle sur le figuré ou, si la relation le permet, à superposer les données.


Variable	Caractéristiques intrinsèques		Relation mesure/dimension		Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Ordinal			Ordonnée		Position	Axe X		
Mesure 1	Thématique	Quantitatif	Discrète	Abrupt	Quantitative		Position	Axe Y	Diagramme orthogonal	Tous
Mesure 2	Thématique	Ordinal	-	-	Ordonnée		Valeur	Figuré		






Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres		Quantitatif				Ordonnée		Position	Axe X	Diagramme orthogonal	Comparable
Mesure 1	Thématique	Quantitatif	Continue	-	Comparable	Quantitative		Position	Axe Y		
Mesure 2	Thématique	Quantitatif	Continue	-	Comparable	Quantitative		Position	Superposition		

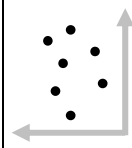





Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique					Sélective		Position	Axe X	Diagramme Polaire	Tous
Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative		Position	Axe Y		
Mesure 2	Thématique	Ordinal	-	-	Comparable	Ordonnée		Couleur	Figuré		



Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Ordinal				Ordonnée		Position	Axe X	Diagramme orthogonal	Inclusive
Mesure 1	Thématique	Quantitatif	-	-	Comparable	Quantitative		Position	Axe Y		
Mesure 2	Thématique	Quantitatif	-	-	Comparable	Quantitative		Position	Superposition		




Nuage de points




Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Qualitatif				Associative		Position	Figuré	Diagramme orthogonal	Tous
Mesure 1	Thématique	Quantitatif	-	-	-	Quantitative		Position	Axe X		
Mesure 2	Thématique	Quantitatif	-	-	-	Quantitative		Position	Axe Y		



Le nuage de point néglige l'information du groupe de membres pour mettre à l'avant plan la relation entre les deux mesures.

Diagramme parallèle

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective		Couleur	Figuré	Diagramme orthogonal	Tous
Mesure 1	Thématique	Quantitatif	-	-	-	Quantitative		Position	Axe Y		
Mesure 2	Thématique	Quantitatif	-	-	-	Quantitative		Position	Axe Y		

VII) Tableaux multivariés

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective	A	Position	Axe X		Tous
Mesure 1	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	-	Sélective	A	Forme	Figuré	Tableau	
Mesure 2	Thématique	Ordinal	-	-	-	Ordonnée		Couleur	Figuré		

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif				Sélective	A	Position	Axe X		-
Mesure 1	Thématique	Quantitatif	-	-	-	Quantitative		Taille	Figuré		
Mesure 2	Thématique	Ordinal	-	-	-	Ordonnée		Couleur	Figuré	tableau	

A3.3 Représentations multidimensionnelles

VIII) Cartes multidimensionnelles

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres 1		Ordinal	Continue	-	-	Ordonnée		Position	Plan	Carte	Tous
Groupe de membres 2		Ordinal	Continue	Graduel	-	Ordonnée		Position	Superposition		
Mesure 1	Thématique	Quantitative				Quantitative		Taille	Figuré		

IX) Diagrammes multidimensionnels

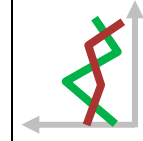
Pyramide

Variable	Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Ordinal	-	-	Comparable	Ordonnée		Position	Axe Y	Diagramme orthogonal	Tous
Groupe de membres 2	Thématique	Qualitatif	-	-	Comparable	Qualitative		Position	Axe X		
Mesure 1	Thématique	Quantitatif				Quantitative		Position	Axe X+ et Axe X-		

Pour obtenir ce type de diagramme le groupe de membres 2 doit avoir exactement 2 membres.

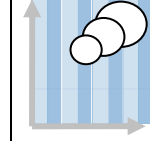
Histogramme à barres cumulées

Caractéristiques intrinsèques		Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble		
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Ordinal	-	-	Comparable	Ordonnée	■	Position	Axe Y	Diagramme orthogonal	Tous
Groupe de membres 2	Thématique	Ordinal	-	-	Additif	Ordonnée	■	Position et Couleur	Axe Y		
Mesure 1	Thématique	Quantitatif				Quantitative	■	Position	Axe Y		



Caractéristiques intrinsèques				Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation	
Groupe de membres 1	Thématique	Quantitatif	Continue	-	Comparable	Quantitative	■	Position	Axe X	Diagramme orthogonal	Tous	
Groupe de membres 2	Thématique	Qualitatif	-	-	Comparable	Sélective	■	Couleur	Figuré			
Mesure	Thématique	Quantitatif				Quantitative	■	Position	Axe Y			

X) Tableaux multidimensionnels



Caractéristiques intrinsèques			Relation mesure/dimension			Caractéristique d'affichage				Caractéristiques d'ensemble	
Variable	Référence	Niveau d'organisation	Distribution	Changement	Type d'agrégation	Attitude perceptive	Implantation graphique	Variable visuelle	Arrangement graphique	Construction graphique	Relation
Groupe de membres	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	Comparable	Ordonnée	■	Position	Axe X	Tableau	Tous
Groupe de membres 2	Thématique	Quantitatif, Ordinal ou Qualitatif	-	-	Comparable	Ordonnée	■	Position	Axe Y		
Mesure 1	Thématique	Quantitatif				Quantitative	●	Taille	Figuré		

