

Visualisation d'information

Jean-Yves Antoine

<http://www.info.univ-tours.fr/~antoine/>

Visualisation d'information

Chapitre 1.3 – Techniques de visualisation avancées

... ou comment quitter les histogrammes et les camemberts Excell pour entrer dans la 5^o dimension

PRINCIPES - Objectifs

1.3.1. Notions

1.3.1.1. Information et incertitude de données

1.3.1.2. Représentation des données multi-dimensionnelles

1.3.1.3. Représentations avancées pour les données hiérarchiques

1.3.2. Pratiques

1.3.2.1. Représentation de l'incertitude

UNCERTAINTY REPRESENTATION

Uncertainty : multiple sources and causes

Source: kind of uncertainty	Possible cause
Limited accuracy	Limitations of measuring instruments or computational processes
Missing data	Non-representative sample size
Inconsistency	Conflict between data from several sources
Noise	Random noise
Bias	Personal idiosyncrasy, systematic measure error
Ambiguity	Linguistic data

Managing uncertainty

- Direct influence on data understanding and decision making
- A complete correction / reduction of uncertainty is impossible
- Modeling (statistics, probability, fuzzy set...) and reasoning with uncertainty
- **Few works on the visualization of uncertainty** – needs for :
 - Seamless integration of information and uncertainty (restricted cognitive load)
 - Flexible representation of uncertainty : handling easily multiples uncertainty models

UNCERTAINTY MODELS

Probability

- Likelihood of an event to occur / a data to be observed

Example - Bayesian probability $P(\text{StateA} \mid \text{ObservationB})$
- Descriptive probability $P(x = A)$

↳ Paramétric model : normal distribution (mean, standard deviation)

Provability

- Certitude Facteurs : -1 = surely false, 1 = surely true
- Belief function : Dempster-Shafer Theory

Example : $BF(A)$, $BF(B)$, $BF(C)$, $BF(\{A,C\})$, $BF(\{A,B\})$, etc...

Possibility

- Alternative matches for an observation

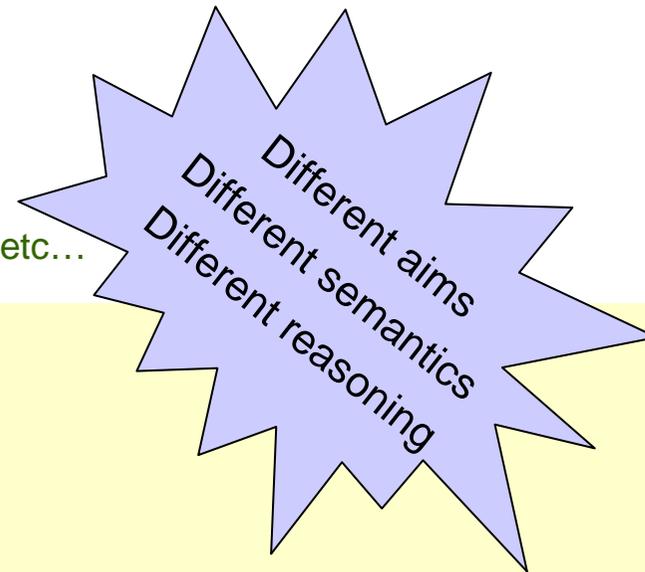
Example : range of errors on an measure

Membership

- Fuzzy sets : degree of membership to a specific set

Degree of confidence

Degree of membership

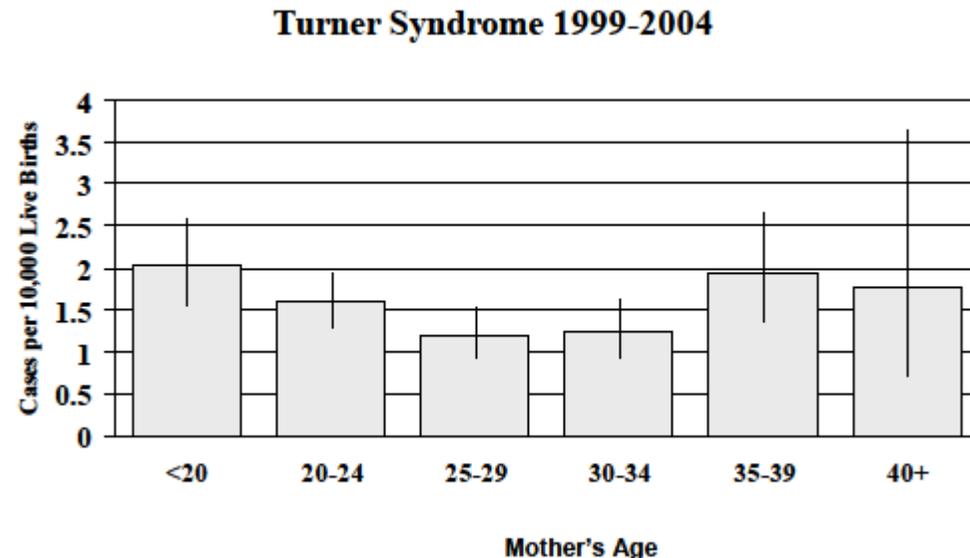
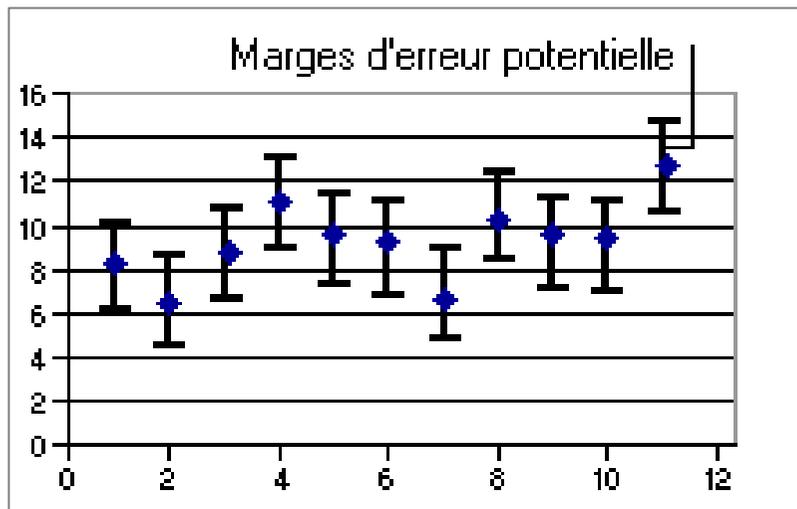


UNCERTAINTY REPRESENTATION

- Visualization techniques better adapted to specific uncertainty models
- Two main approaches : additional information vs. integrated information

Additional information : extrinsic representation

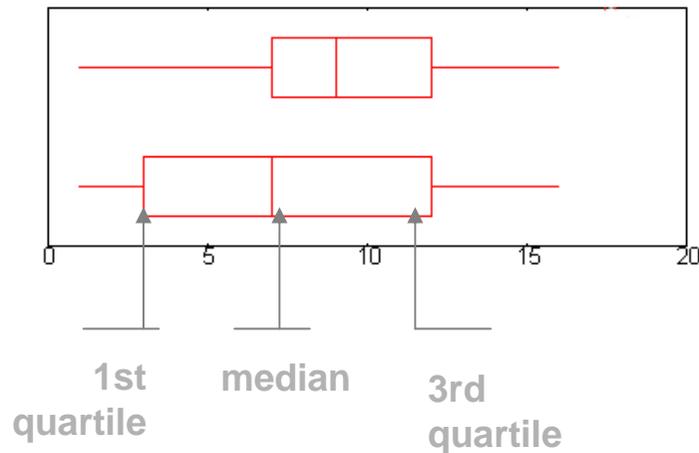
- Additional object to represent a specific parameter / data
- **Error bar** –
 - represent standard deviation or a range of values (confidence interval).
 - works with point chart or bar charts



UNCERTAINTY REPRESENTATION

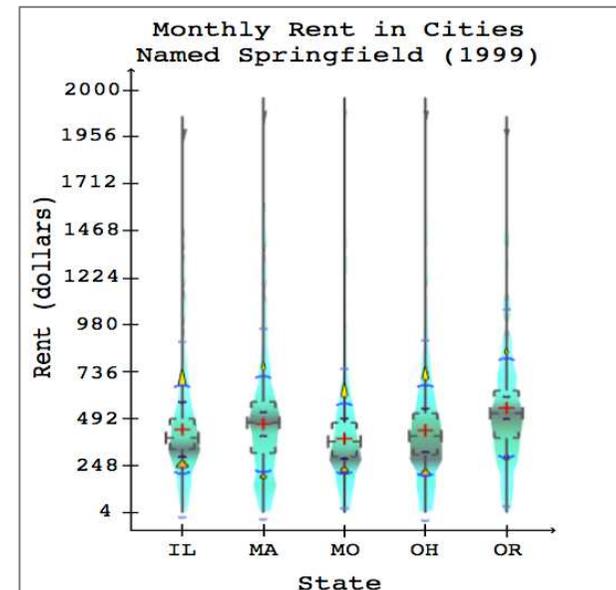
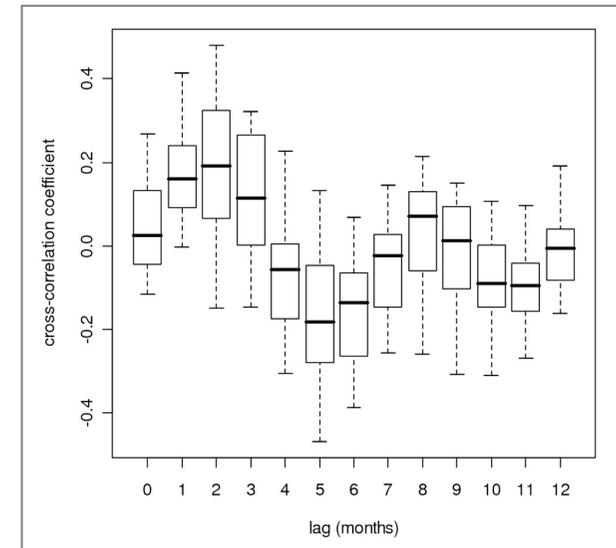
Additional information : extrinsic representation

- **Boxplot, or box and whisker plot (*boite à moustache*)** – to represent visually a distribution (quartile or decile, extreme values)



- **Summary plot** – explicit representation of the distribution (thickness of the line)

[Potter and al. 2011]



UNCERTAINTY REPRESENTATION

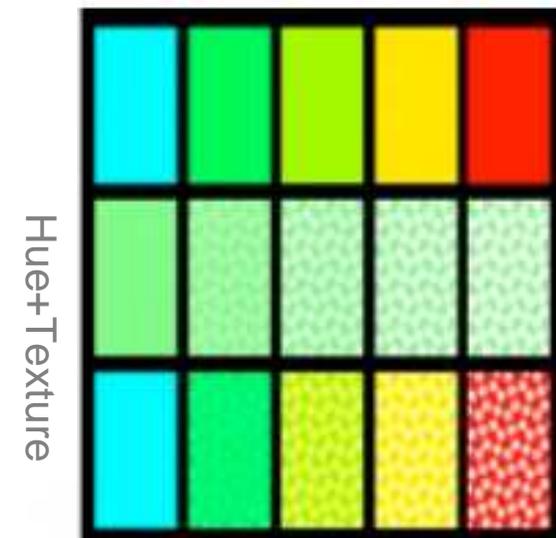
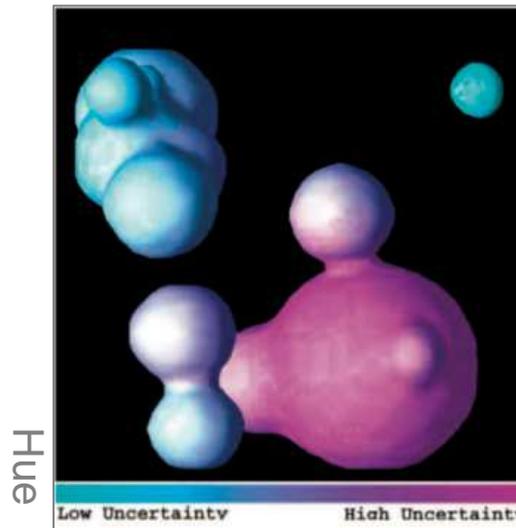
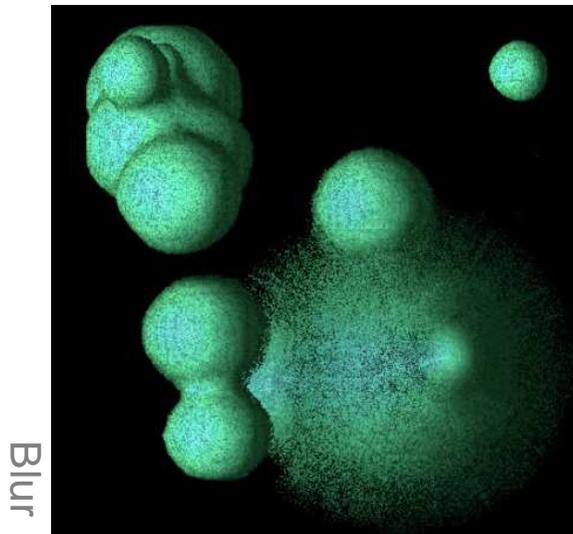
Integrated information : intrinsic representation

One visual feature dedicated to uncertainty representation

Blur

- Easy interpretation of the uncertain nature of the data
- Impossibility to estimate the degree of uncertainty

[Grigoryan, Rheingans 2004]



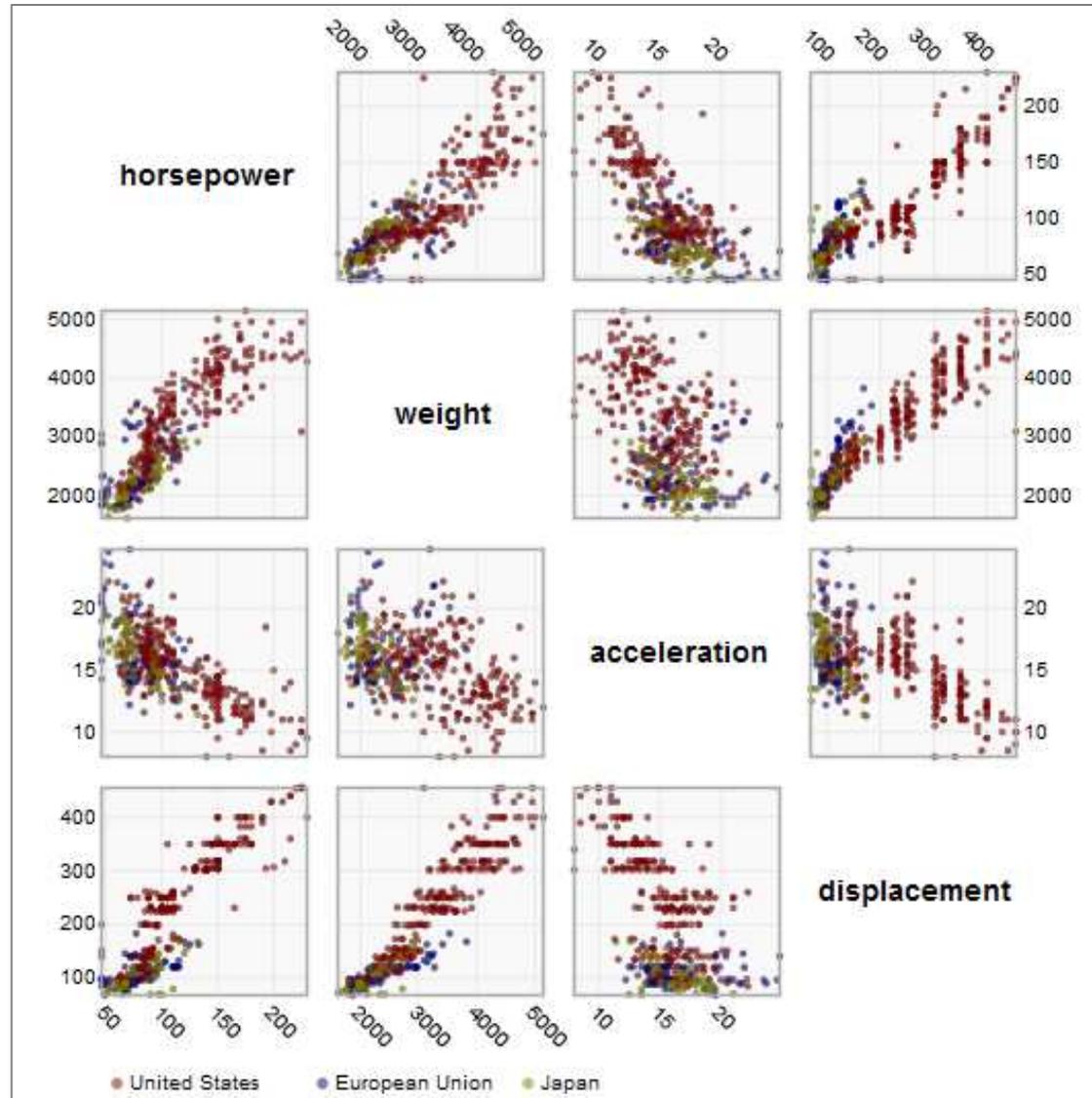
Hue, luminance, saturation, texture

N-DIMENSIONAL DATA

Scatter Plot Matrix (SPLOM)

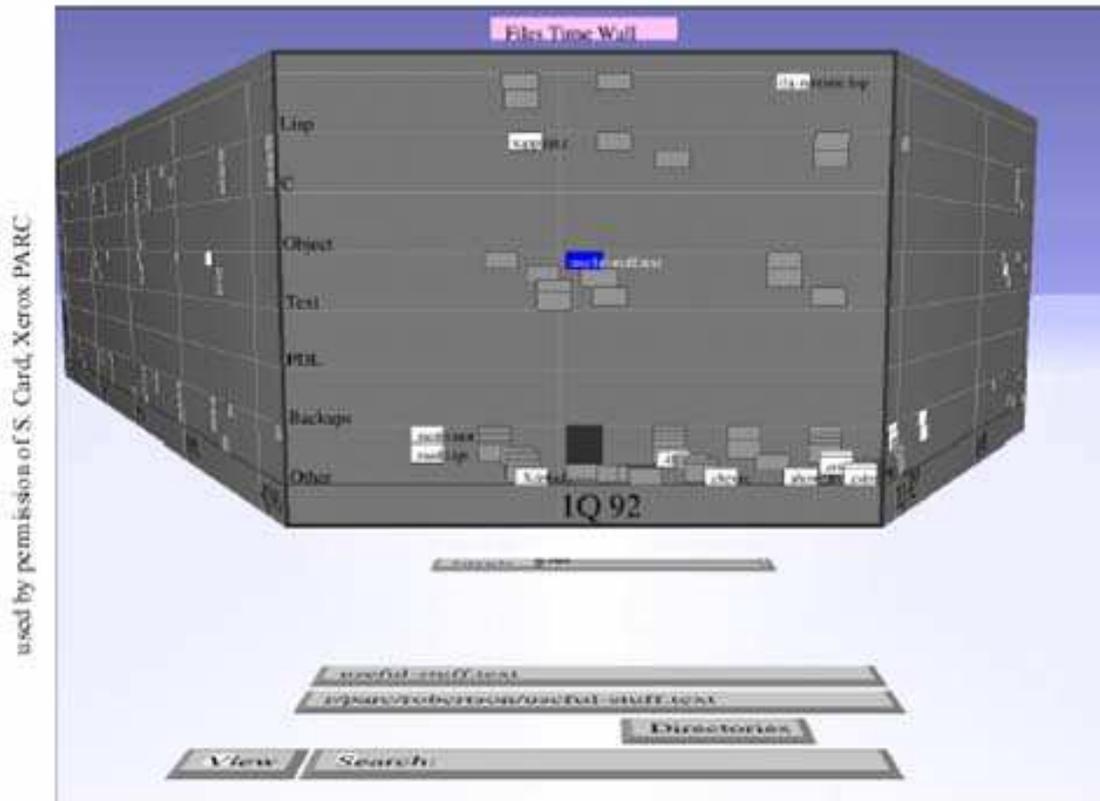
[Heer and al. 2010]

Principle – use small multiples of scatter plots showing a set of pair-wise relations among variables



DONNEES A N-DIMENSIONS

Murs perspectifs



documents
arranged on a
perspective wall

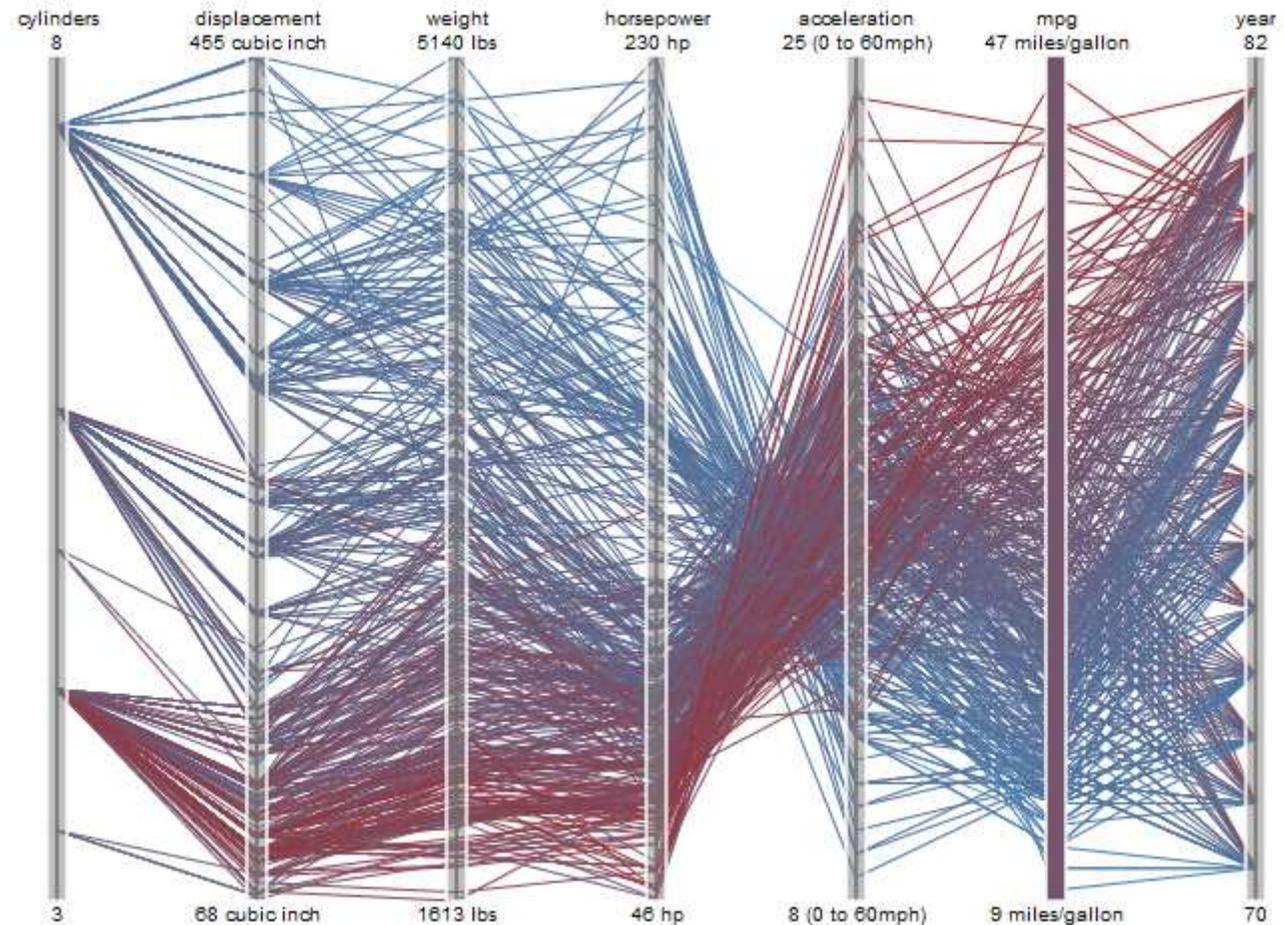
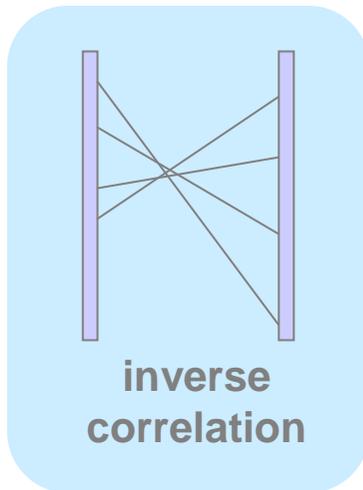
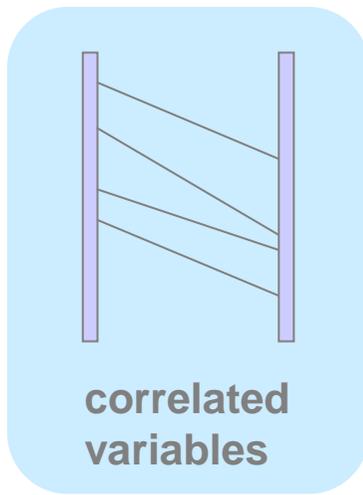
DONNEES A N-DIMENSIONS

Parallel coordinates

[Heer and al. 2010]

Principle – Variables are plotted on parallel axes. A poly-line = a row of the DB

Interactive visualization – You can change the order of the variables



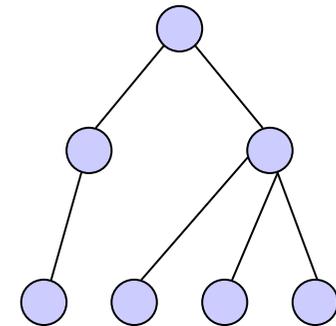
DONNEES HIERARCHIQUES

Deux propriétés visuelles pour guider la visualisation

- Connectivité
- Clôture

Connectivité : arbres hiérarchiques

- Représentation la plus courante
- Des limitations toutefois [Card *et al.* 1999]
 - peu utilisable à mesure que l'arbre croît
 - arbre non équilibré occupation non optimale de l'espace
 - valeurs associées au nœuds : attributs quantitatifs non intégrables



Clôture : *treemaps*

- Technique de remplissage d'espace
- *Treemap* : principes généraux [Schneidermann 1992]
 - relations hiérarchiques représentées par une inclusion d'espaces
 - aire des rectangles proportionnelles aux attributs
- Intérêt
 - attributs quantitatifs intégrés en propre
 - espace consommé indépendant de la profondeur de la hiérarchie

DONNEES HIERARCHIQUES

Approche surfacique : *Treemaps*

- **Partionnement d'espace** – chaque rectangle représente un nœud de taille proportionnelle à sa valeur d'attribut. Sous-nœuds : partition du rectangle père.
- Plusieurs algorithmes de découpage : propriétés différentes
 - ✓ Ratio – rectangles plus ou moins proche du carré (le plus lisible)
 - ✓ Ordre – ordre des sous-rectangles proche de celui des données initiales
 - ✓ Stabilité – stabilité du découpage quand les données changent.

Algorithme	Ratio	Ordre	Stabilité
Arbre binaire	+	+	++
Mixed treemap	-	++	++
Ordered	+	+	+
Slice & Dice	++	++	+
Squarified	-	-	+
Strip	+	++	+



- **Critères perceptifs**

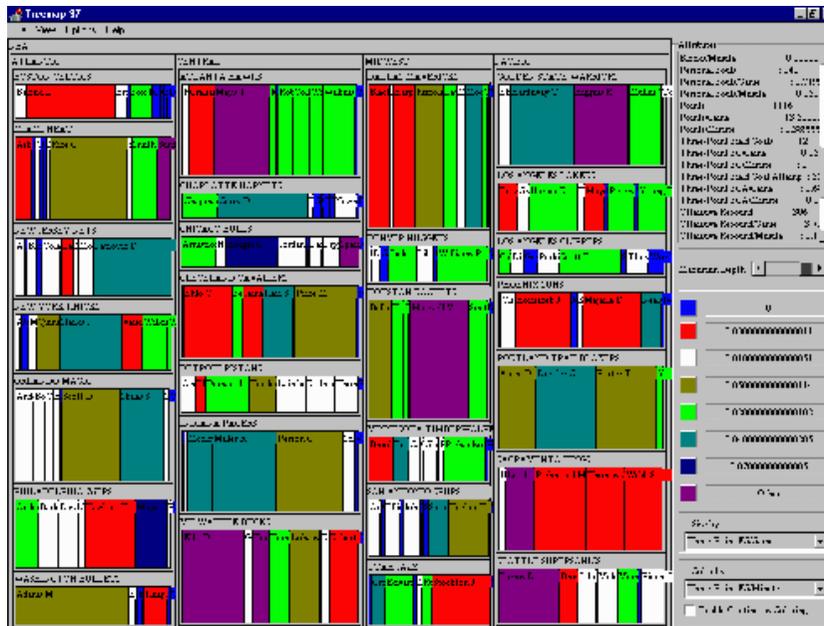
- ✓ Attention au ratio largeur / longueur (se rapprocher du carré)
- ✓ Stabilité – stabilité du découpage quand les données changent.

DONNEES HIERARCHIQUES

Approche surfacique : Treemaps

- Recommandations ergonomiques

- ✓ Ratio longueur / largeur : éviter les zones trop minces
- ✓ Couleurs – Pas plus de deux couleurs, la perception devient vite trop dure sinon
- ✓ Adjacences – Ne pas utiliser les treemaps si adjacences non interprétables



[Schneiderman 1992]



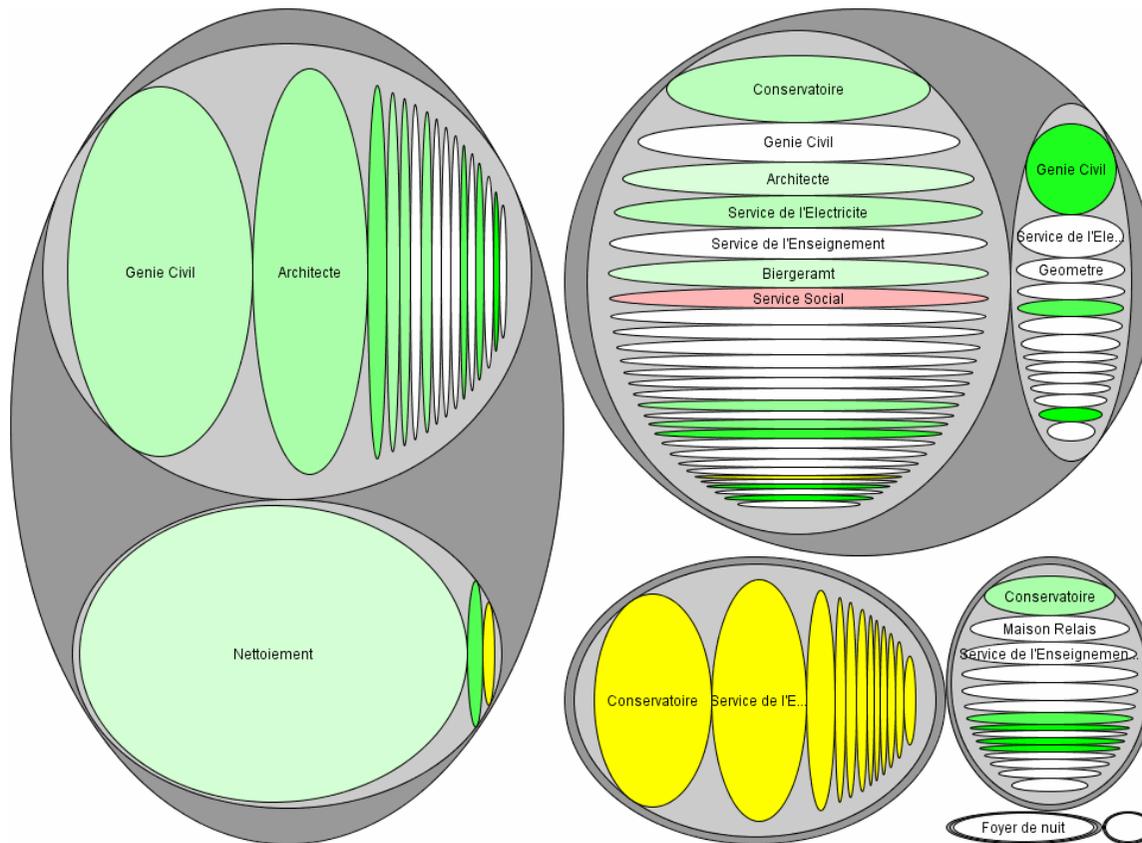
SmartMoney (1999)

DONNEES HIERARCHIQUES

Ellimaps

[Otjacques *et al.* 2008]

- Même principe que les *squarified treemaps* mais avec des ellipses
- Gestalt : meilleur respect du principe contour et de continuité

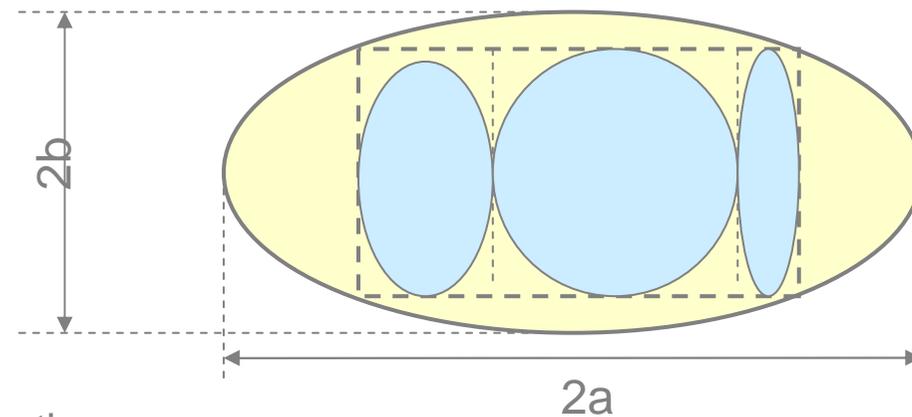


DONNEES HIERARCHIQUES

Ellimaps :algorithme de construction (version initiale)

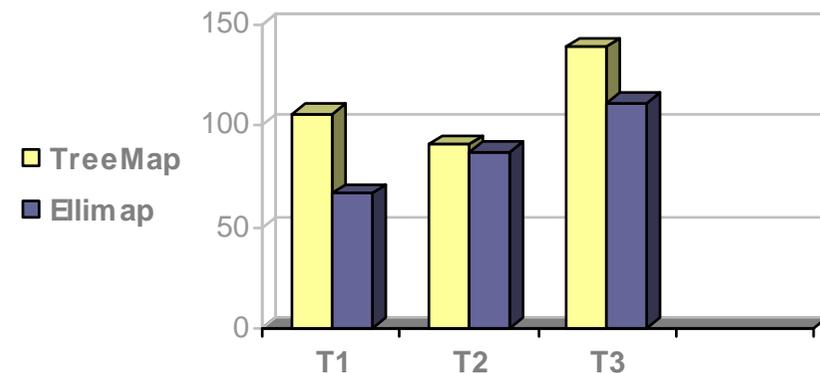
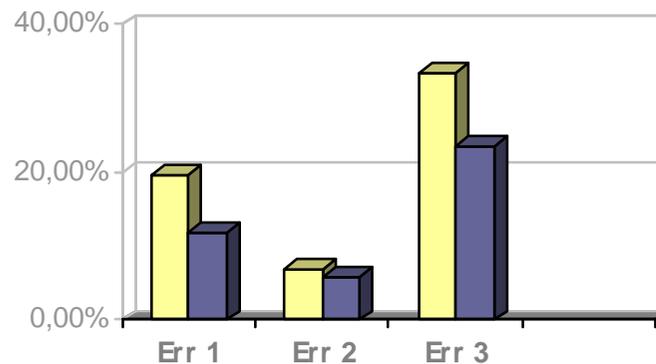
Adaptation aux ellipses des *squarified treemaps*

- Découpage du rectangle inscrit dans l'ellipse de ratio $\phi = a/b$
- Aire des ellipses proportionnelle à la valeur de l'attribut
- Découpage vertical ou horizontal suivant le plus grand côté du rectangle)



Evaluation

- Trois tâches de recherche d'information
- Temps et erreurs



DONNEES MULTI-DIMENSIONNELLES

Entrepôts de données et cubes de données

Cubes de données: intérêt

- Cubes de données: ensemble des agrégations réalisables sur les données
- Pré-calcul réduisant le requêtage OLAP

Cubes de données: limitations

- Exploration fastidieuse des cubes de données du fait de leurs très grandes dimensions

Exemple – 1 000 000 tuples \Leftrightarrow > 100 000 000 cubes

- A l'heure actuelle, exploration avec des techniques inadaptées (SQL, représentation graphiques simples) ou des requêtes OLAP pré-programmées : aucune interactivité

Solution : visualisation de données

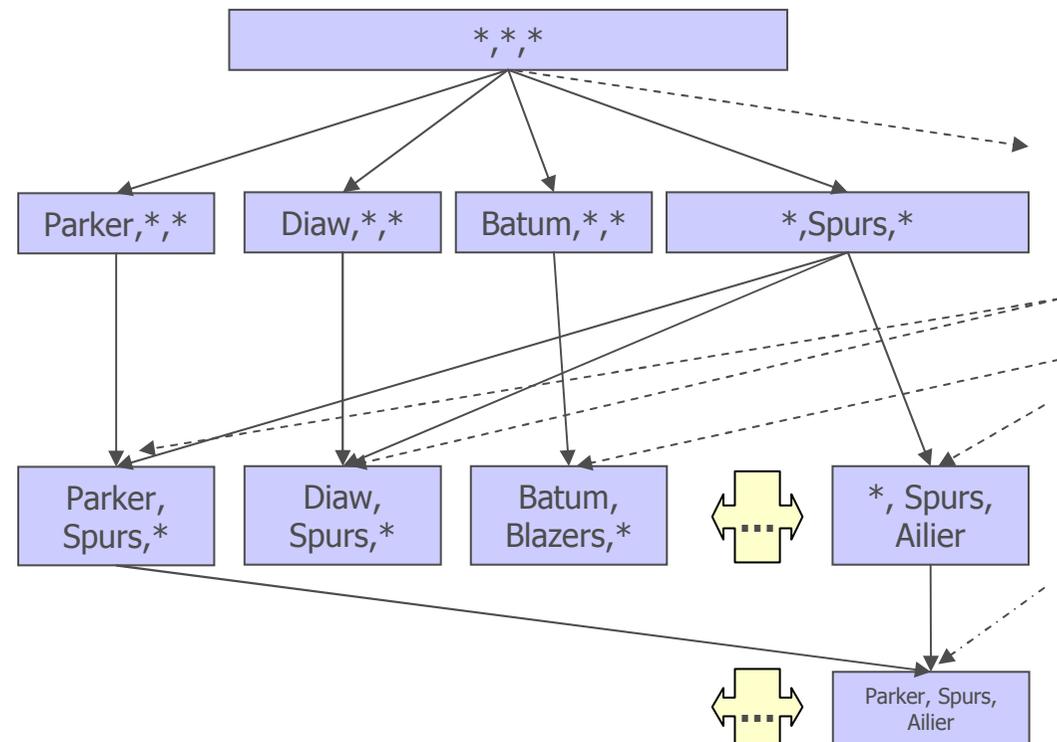
- Techniques de visualisation avancées permettant une réelle exploration **interactive**

DONNEES MULTI-DIMENSIONNELLES

Semi-treillis des valeurs des cuboïdes

- Graphe acyclique orienté permettant de représenter les attributs
- Taille des nœuds proportionnelle à la valeur de la variable d'agrégation

Nom	Equipe	Poste	Count
ALL	ALL	ALL	3
Parker	ALL	ALL	1
Diaw	ALL	ALL	1
Batum	ALL	ALL	1
ALL	Spurs	ALL	2
ALL	Blazers	ALL	1
ALL	ALL	Meneur	1
ALL	ALL	Ailier	2
Parker	Spurs	ALL	1
...
ALL	Blazers	Ailier	1
...
Batum	Blazers	Ailier	1



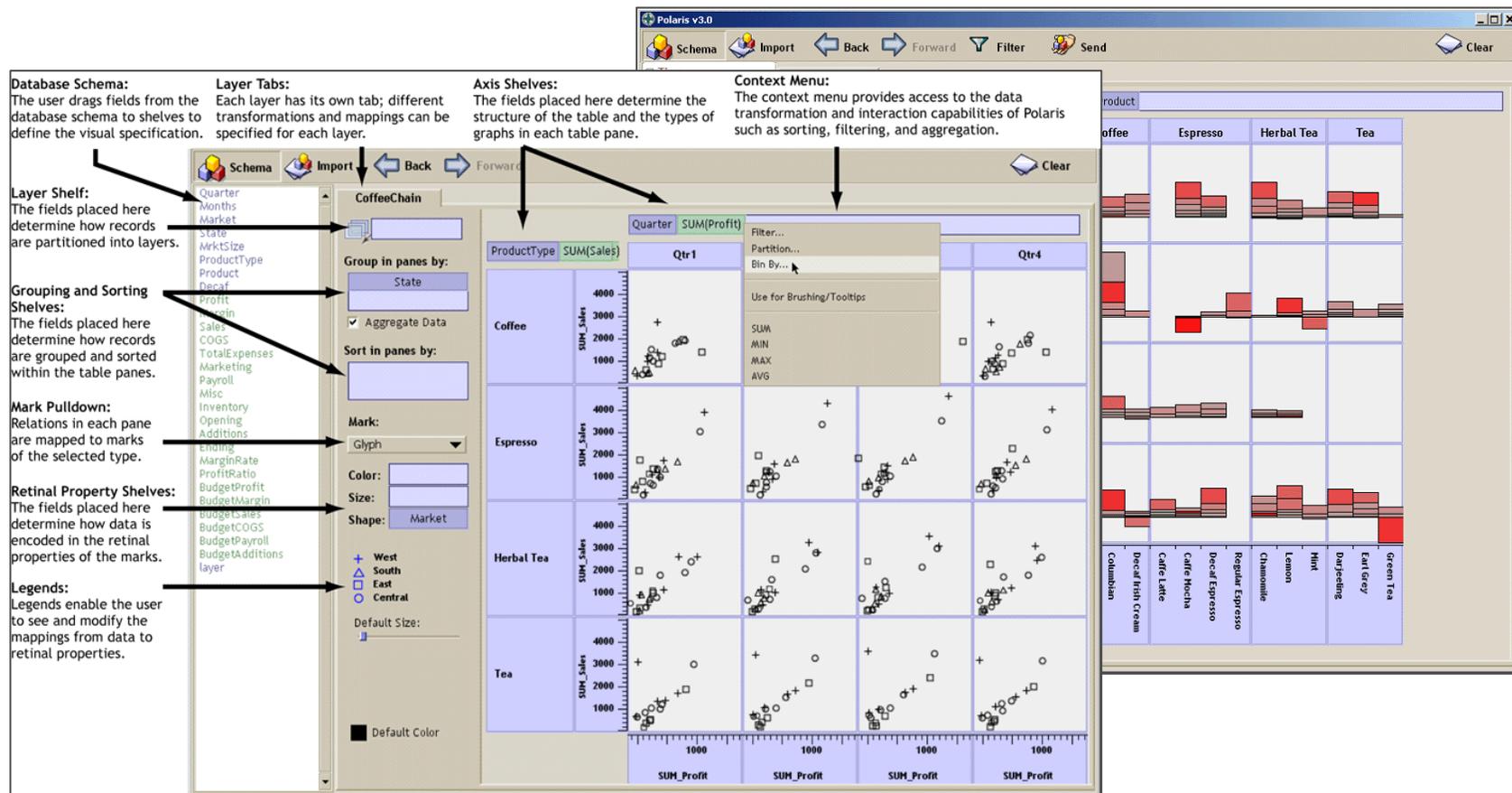
- **Limitation** – Visualisation difficile au-delà de 1000 nœuds

DONNEES MULTI-DIMENSIONNELLES

Projections multiples

Polaris [Stolte *et al.*, 2002]

- Projection 2D, chaque dimension d'affichage correspondant à des agrégations
- Plusieurs formes de représentation (courbes, histogramme, camemberts)
- Plusieurs projections affichables simultanément.

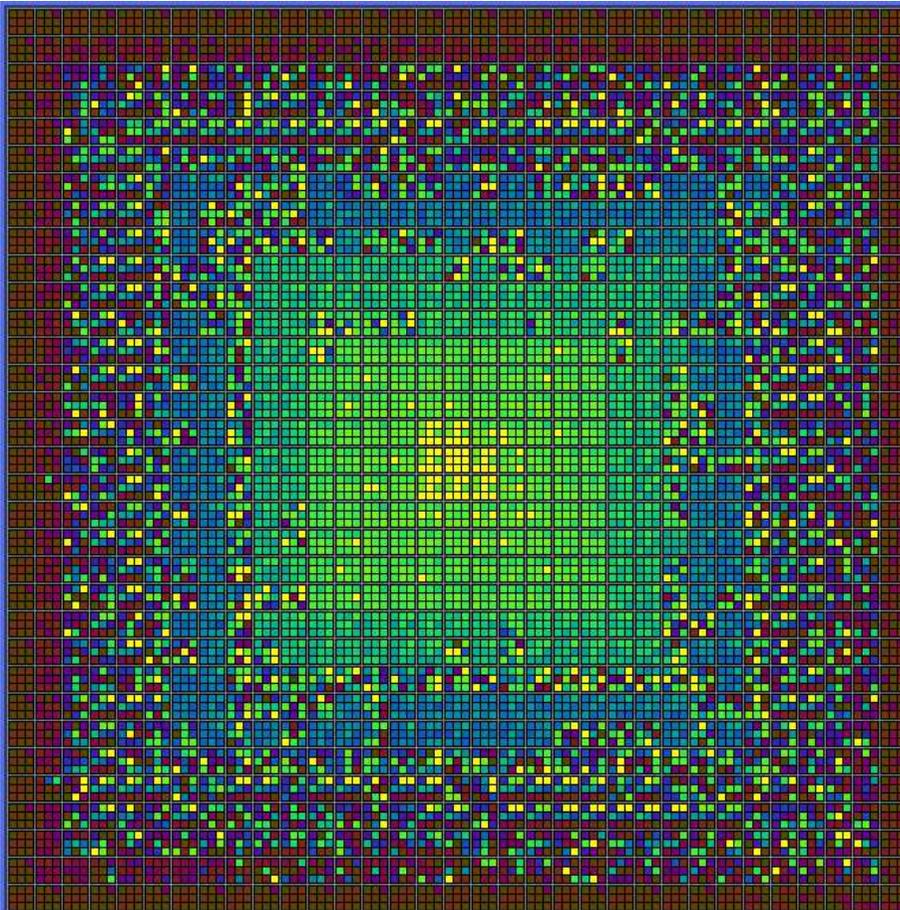


- **Limitation** – Efficace, mais le nombre de projection n'est pas infini: exploration orientée par les possibilités d'affichage, nécessite une étude préalable fine des besoins

DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel VisDB [Keim, Kriegel, 2000] d3Viz [Novelli, Auber, 2008]

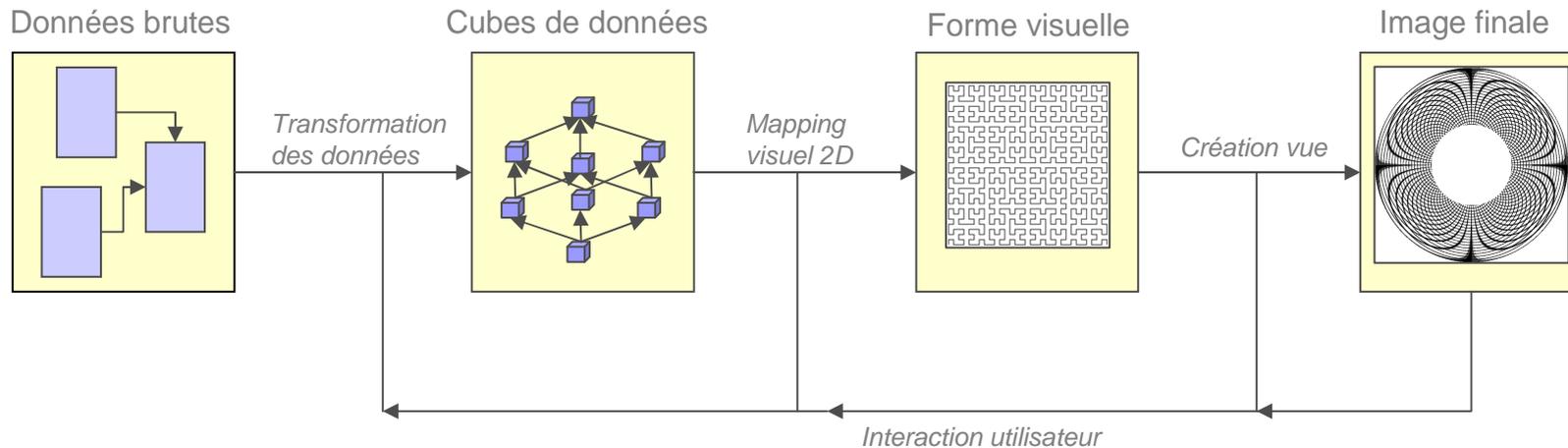
- Visualisation très dense de l'information : un pixel = une donnée
- Jeu sur la couleur et la position des pixels pour interpréter l'image visuelle présentée
- Fouille visuelle sans perte de contexte: zooms de type glyphs fish-eye



DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel

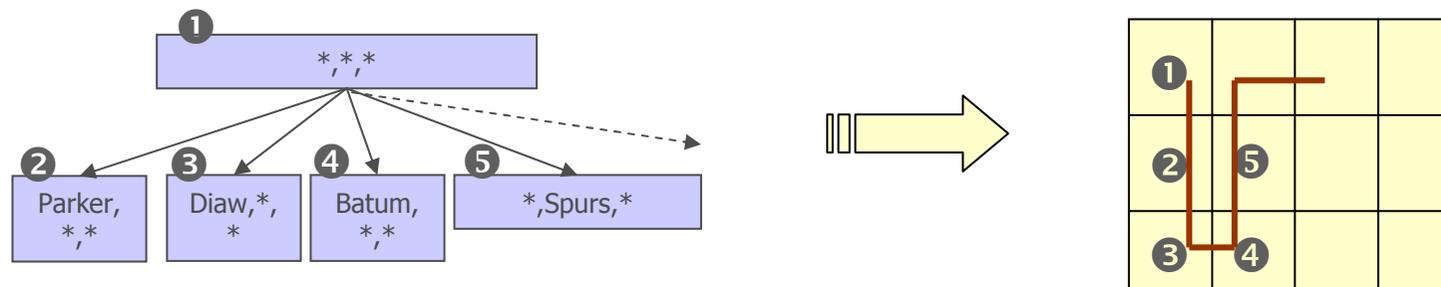
- Pipeline de visualisation



Visualisation orientée pixel : mapping visuel

- Principe

- Associer un pixel à une donnée sur l'image, en respectant leur proximité originelle
- Plusieurs courbes de remplissage d'espace envisageables

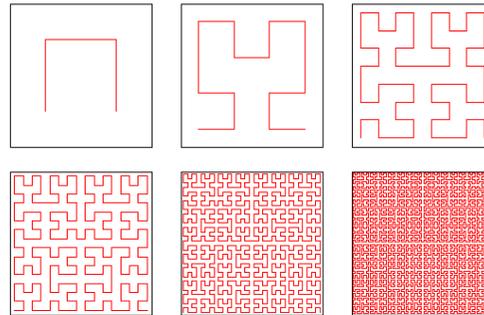


DONNEES MULTI-DIMENSIONNELLES

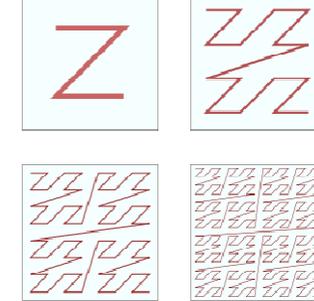
Visualisation orientée pixel : mapping visuel

- Courbes fractales

Peano-Hilbert

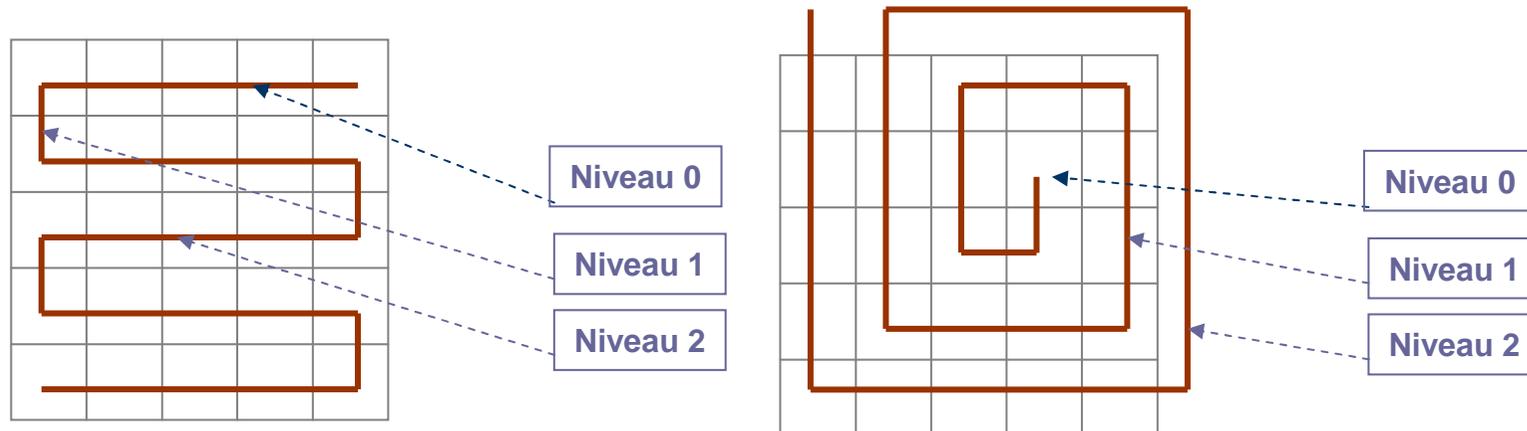


Morton (ordre Z)



Niveaux du semi-treillis non respectés : exploration difficile

- Courbes en scanline et spirale



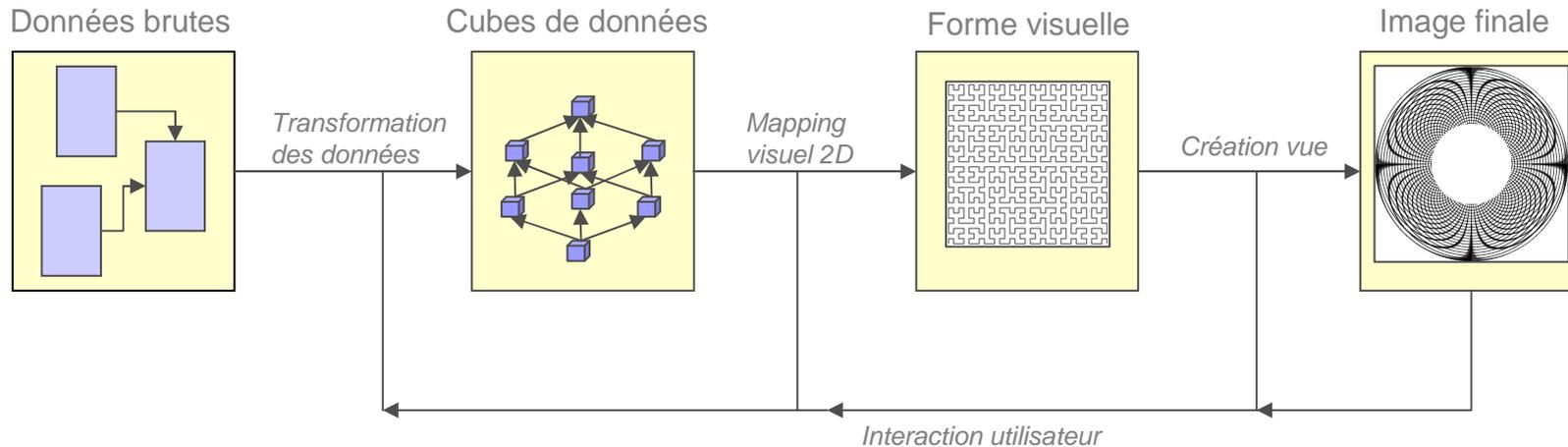
- Exploration

- *Observation* : rotation, zoom de type fisheye etc...
- *Action* : recalcul de la vue suivant de nouvelles dimensions

DONNEES MULTI-DIMENSIONNELLES

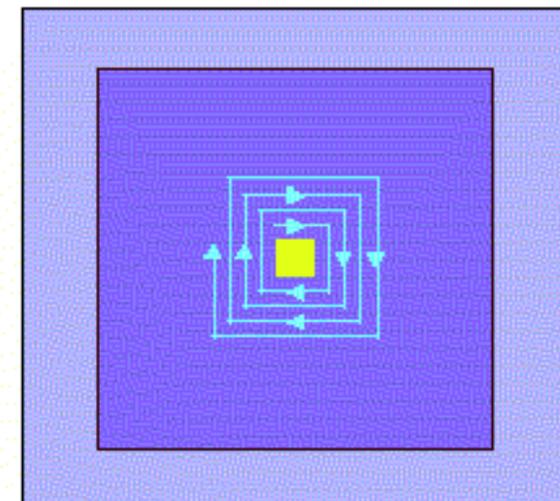
Visualisation orientée pixel

- Pipeline de visualisation



Visualisation orientée pixel : mapping visuel et requêtage

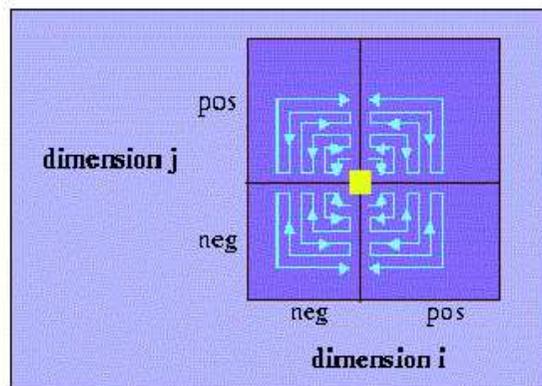
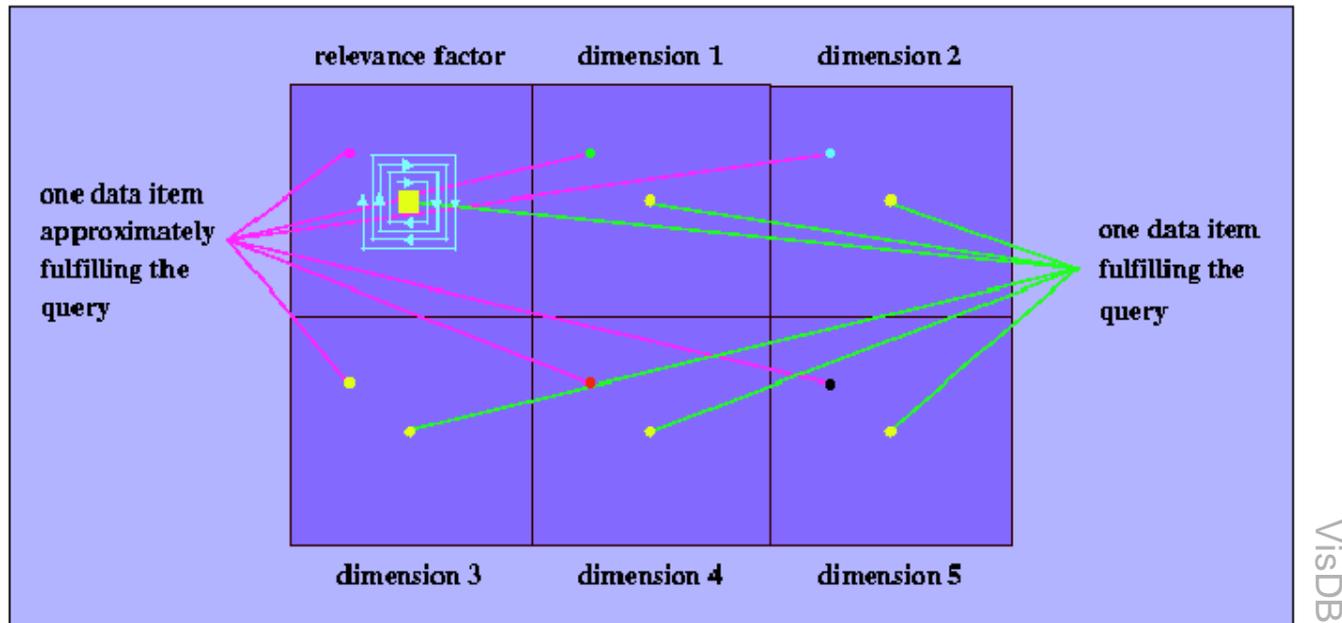
- Associer un pixel à une donnée
- **Flexibilité d'interrogation** : pas de filtrage, mais degré de pertinence des données
- **Visualisation** : couleur et proximité des données en fonction de leur pertinence
- Plusieurs courbes de remplissage d'espace (spirale...)



DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel : visualisation multi-dimensionnelle

- Une fenêtre pour la visualisation des résultats de la requête (facteur de pertinence)
- Autres fenêtres pour projection sur différentes dimensions

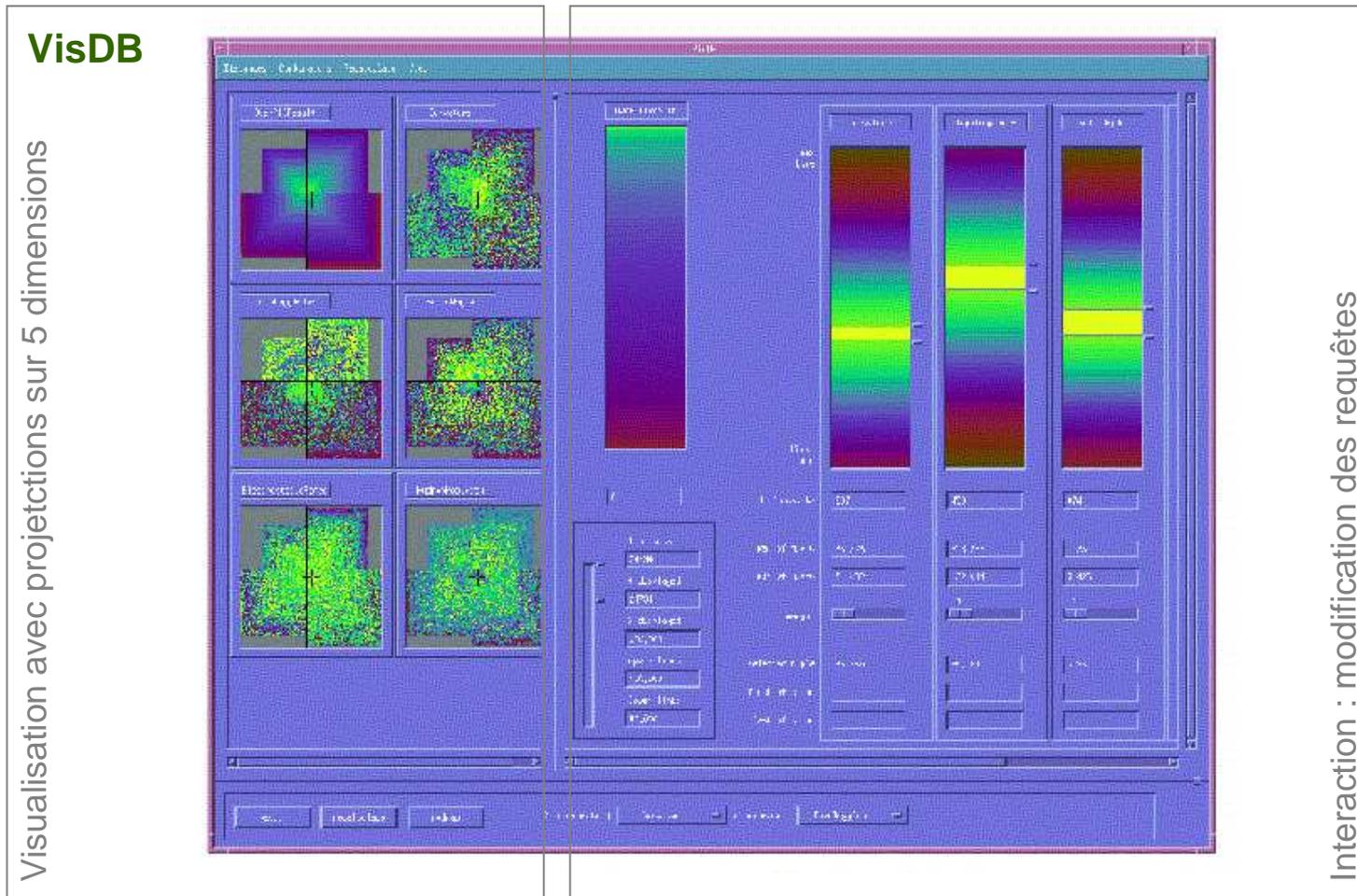


Visualisation 2D – association de 2 dimensions aux axes + remplissage par courbe fractale

DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel : interaction

1. L'utilisateur définit initialement ses requêtes
2. Possibilité de les modifier interactivement dans l'interface au vu des résultats



Limitations – Filtrage, projection gérées, mais pas de zoom sur les données brutes

RESEARCH

Main scientific challenges in current researches

- Real-Time visualization of huge datasets
- Virtual or augmented reality for enhanced interaction (immersive displays)
- Multimodal visualization : interaction in a seamless way (reduced cognitive load)
- Collaborative visualization
- **HCI : design and ergonomic guidelines for information visualization**

⇒ « *Highest priority task* »

[Zudilova and al. 2009, chapter 1]

BIBLIOGRAPHIE

Ressources en ligne : concepts

Semiologie graphique

www.sciences-po.fr/cartographie/semio/

Treemap

www.sc.umd.edu/hcil/treemap/

Ressources en ligne : outils

- **30 freeware pour la visualisation d'information (post en anglais et reprise en français)**

- [www.computerworld.com/s/article/9214755/Chart and image gallery 30 free tools for data visualization and analysis](http://www.computerworld.com/s/article/9214755/Chart_and_image_gallery_30_free_tools_for_data_visualization_and_analysis)

- www.lemondeinformatique.fr/actualites/lire-22-outils-gratuits-pour-visualiser-et-analyser-les-donnees-1ere-partie-47241.html

- **Toolkits opérationnels pour découvrir la visualisation interactive**

- InfoVis

<http://philogb.github.com/jit/>

(freeware)

- VTK (Kitware Visualization Toolkit)

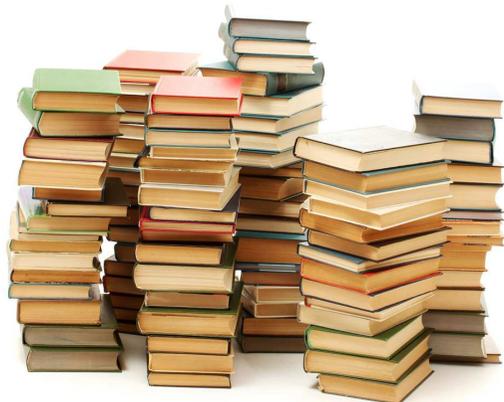
<http://www.vtk.org/>

(visualisation 3D)

- Prefuse Information Visualization Toolkit

<http://prefuse.org/>

(freeware: graphs, trees)



BIBLIOGRAPHIE

Ouvrages de référence : concepts et techniques

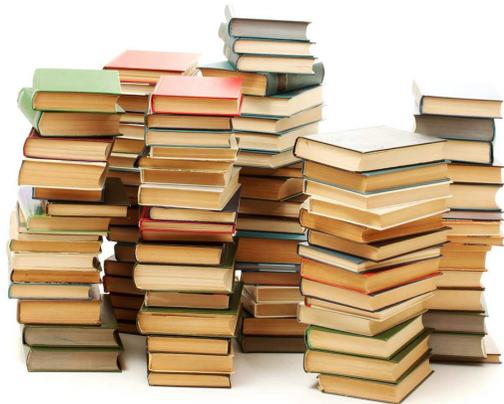
Card S., Mackinlay J., Schneidermann B. (1999) *Readings in Information Visualization, using vision to thing*. San Francisco, Morgan Kaufmann Publ.

Schneiderman B. (1992) Tree visualization with tree maps: 2-D space filling approach. *ACM Trans. on Computer Graphics*. 11(1), 92-99.

Ouvrages de référence : recherche & veille technologique

Von Landesberger T., Kuijper A., Schreck T., Kohlhammer J., van Wijk J.J., Fekete J.-D., Fellner D.W. (2011) Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Computer Graphics*. **30**(6). 1719-1749.

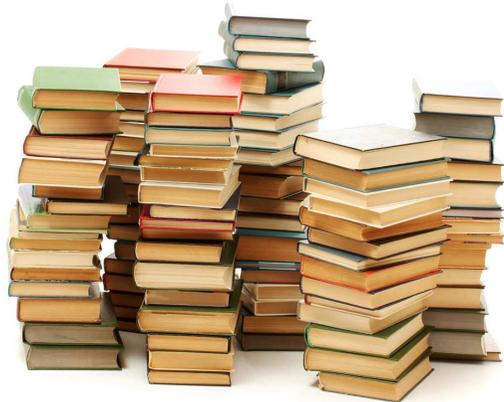
Zudililova-Seintra E., Adriaansen T., van Liere R. (2009) *Trends in interactive visualization : state of the art and survey*. Springer, London, OK. Advanced Information and Knowledge Processing Series



BIBLIOGRAPHIE

Travaux cités

- Camacho M.J., Steiner B.A., Berson B.L. (1990) Icons versus alphanumerics in pilot-vehicles interfaces. Actes Human Factors and Ergonomics Society Annual Meeting.
- Grigoryan G., Rheingans P (2004) Point-Based Probabilistic Surfaces to Show Surface Uncertainty, *IEEE Transactions on Visualization and Computer Graphics*, **10**(5), 546—573.
- Keim D.A., Kriegel H.P. (2000) Designing pixel-oriented visualization techniques: theory and application. *IEEE Trans. Viz. Computer Graphics*. **6**(1), 59-78.
- Keim D., Andrienko G., Fekete J.-D., Görg C., Kohlhammer J., Melançon G. (2008). Visual Analytics: Definition, Process, and Challenges, *In. Information Visualization, Lectures Notes in Computer Sciences. LNCS 4950*:154-175.
- Novelli N., Auber D. (2008) Calcul et fouille visuelle orientée-pixel de cubes de données. *RSTI-RIA*. **22/2008**:329-352.
- Potter K., Kniss J., Riesenfeld R., Johnson C.R. Visualizing Summary Statistics and Uncertainty. *CGF* 29(3)



BIBLIOGRAPHIE

Travaux cités (2)

Otjacques B., Collin P., Noirhomme M., Feltz F. (2008) Ellimaps, une technique basée sur la loi d'inclusion pour représenter des hiérarchies avec nœuds pondérés. *RSTI-RIA*. **22/2008**:301-327.

Pylyshyn Z. (1994) Some primitive of spatial attention. *Cognition*. **50**:363-384.

Steim D.A. (2002) Information visualization and visual datamining. *IEEE Trans. Visualization and Computer Graphics*. **7(1)** : 100-107. <http://nm.merz-akademie.de/~jasmin.sipahi/drittes/images/Keim2002.pdf>

Stolte C., Tang D., Hanrahan P. (2002) Multiscale vizualisation using data cubes. *InfoVis'2002*, 7-14.

Stolte C., Tang D., Hanrahan P. (2002) Polaris, a system for query, analysis and vizaulisation of multidimensional relational databases. *IEEE Trans. Vis. Computer. Graphics*, **8(1)**, 52-65..

Travaux cités en ligne

Heer J., Bostock M., Ogievetsky V. (2010) A tour through the vizualisation zoo. [<http://hci.stanford.edu/jheer/files/zoo/>]

