

Visualisation de données multi-dimensionnelles, axes flexibles, géométrie et combinatoire

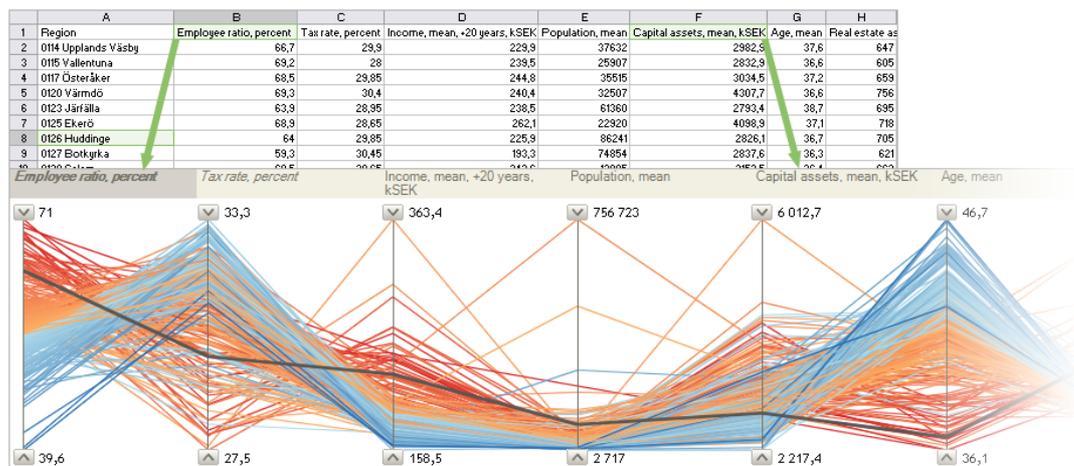
Encadrants: Bruno Pinaud (et Guy Melançon)

Equipe d'accueil: MaBioVis, thème Visualisation, équipe projet INRIA GRAVITE

Lieu: LaBRI

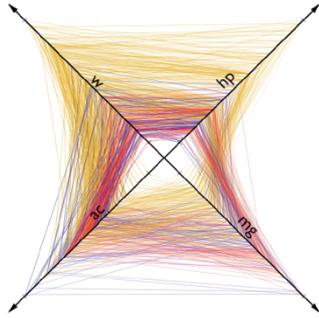
Description du sujet

La visualisation reste une approche populaire pour mettre à jour la structure inhérente à un ensemble de données multi-dimensionnelles. L'une des visualisations maintenant classique sont les *coordonnées parallèles*.

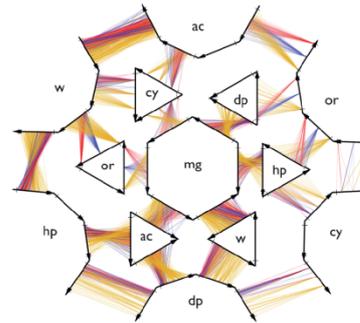


L'idée de cette visualisation, due à (Inselberg 1985) consiste à appliquer une transformation affine pour disposer les axes d'un repère orthonormé dans un plan 2D de manière parallèle. Un point en N dimensions est ensuite matérialisé par une ligne brisée croisant chacun des axes. On utilise cette visualisation de manière classique pour rechercher des corrélations possibles entre des variables. Les coordonnées parallèles montrent donc N-1 corrélations : N variables disposées le long d'un axe. (L'exploration interactive d'un tel diagramme permet toutefois de faire émerger des corrélations ou des motifs entre variables qui ne sont pas directement voisines.)

Un travail récent de (Claessen et van Wijk 2011) propose de disposer les axes librement, donnant lieu à des configurations parfois excentriques. Leur idée suit en cela celle de (Hoffman 1999) (que ne cite pas Claessen et van Wijk) et d'autres auteurs (cités par Claessen et van Wijk).



Radial chart



Many-to-many PCP

La liberté gagnée sur la disposition des axes permet en outre de montrer plus de corrélations entre des variables. Un graphe permet de capturer les relations de voisinage entre variables montrées dans un diagramme. Ce graphe est alors une chaîne à N sommets dans le cas des coordonnées parallèles classiques ; le cas du radial chart est un cycle sur 4 sommets ; le cas des Many-to-many PCP est une clique sur 7 sommets. Ce graphe est appelé *ARGOI* par (Claessen et van Wijk 2011) (Attribute Relation Graphs of Interest).

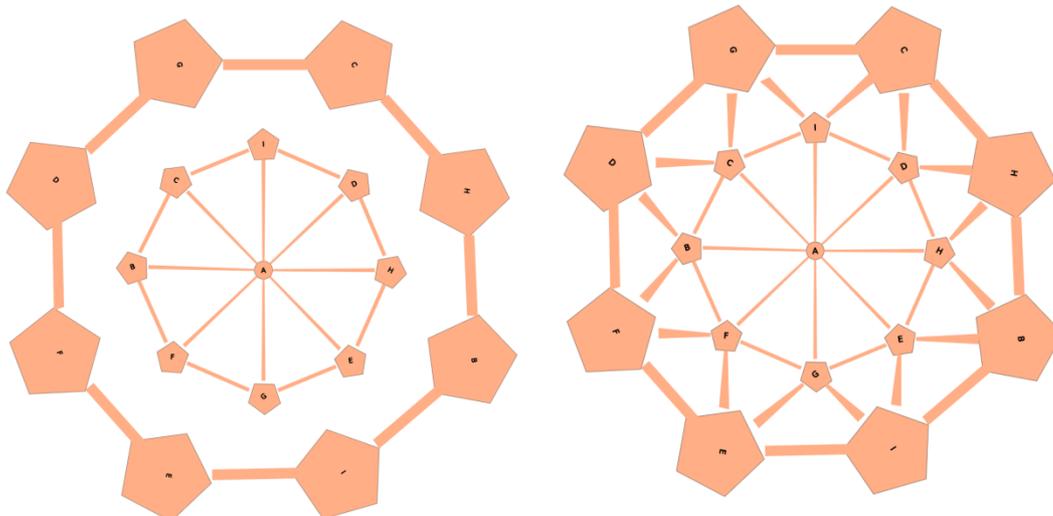
Le défi

Le but du stage est d'arriver à donner un algorithme permettant de visualiser les relations entre certaines variables capturées par un graphe *ARGOI* quelconque.

On pourra débuter par chercher une solution pour le cas où l'*ARGOI* est une clique de taille quelconque (même si on peut penser que cela ne soit raisonnable que pour une clique de taille modeste). Cela revient à visualiser toutes les relations possibles entre N variables.

Ce premier problème admet une solution triviale qui vaut la peine d'être explicitée. Il est possible de recouvrir une clique par un ensemble de chaînes : la solution exige de bien définir des chemins qui sillonnent la clique en zigzag. Ainsi, si on produit pour chaque chaîne un diagramme en coordonnées parallèles classiques, on visualise bien toutes les relations. Cette piste a été explorée par (Wegman 1990).

Si on ferme ces chaînes pour en faire des cycles, on produit alors des diagrammes circulaires que l'on peut emboîter (figure suivante, schéma de gauche, auquel il manquerait encore un ruban de variable. Noter que l'on s'épargne un peu de travail en plaçant l'une des variables au centre du schéma). Si on s'autorise à visualiser des relations entre variables voisines placées sur des rubans concentriques distincts, on diminue le nombre de ruban nécessaire pour visualiser l'ensemble des relations de la clique, et on produit un schéma qui approche le Many-to-many PCP illustré plus haut. Le schéma Many-to-many PCP, encore plus économe, est une solution ad hoc pour le cas $N = 7$. Il est économe notamment parce qu'il matérialise certaines variables à l'aide de triangle.



Le défi compte une conjecture que l'on devrait pouvoir établir et qui mérite d'être portée à l'attention de la communauté : on peut donner un algorithme qui réalise toutes les relations entre N variables à l'aide de rubans de pentagones disposés adéquatement.

Cette solution doit pouvoir être exploitée pour traiter le cas d'un ARGOI quelconque, en jouant sur le nombre de rubans leur disposition. La forme pentagonale semble naturelle ici puisqu'elle permet de matérialiser les relations entre couches de variables.

Le stage devrait permettre d'explorer pleinement ce type de visualisation, et idéalement d'établir la robustesse de la piste évoquée, voire de concevoir une toute autre stratégie pour matérialiser un ARGOI quelconque.

Travail attendu.

Le stage exige de prendre en main les articles de (Claessen et van Wijk 2011), et accessoirement d'explorer le travail de (Hoffman 1999).

L'utilisation de la plate-forme Tulip permettra d'implémenter les solutions et d'expérimenter de nouvelles pistes.

Références

Claessen, J. H. T. and J. J. van Wijk (2011). "Flexible Linked Axes for Multivariate Data Visualization." *Visualization and Computer Graphics*, IEEE Transactions on 17(12): 2310-2316.

Hoffman, P. E. (1999). *Table Visualizations: A Formal Model and Its Applications*. Computer Science, Institute for Visualization and Perception Research, University of Massachusetts: 239.

Inselberg, A. (1985). "The Plane with Parallel Coordinates." *The Visual Computer* 1: 69-91.

Wegman, E. J. (1990). "Hyperdimensional Data Analysis Using Parallel Coordinates." *Journal of the American Statistical Association* 85(411): 664-675.