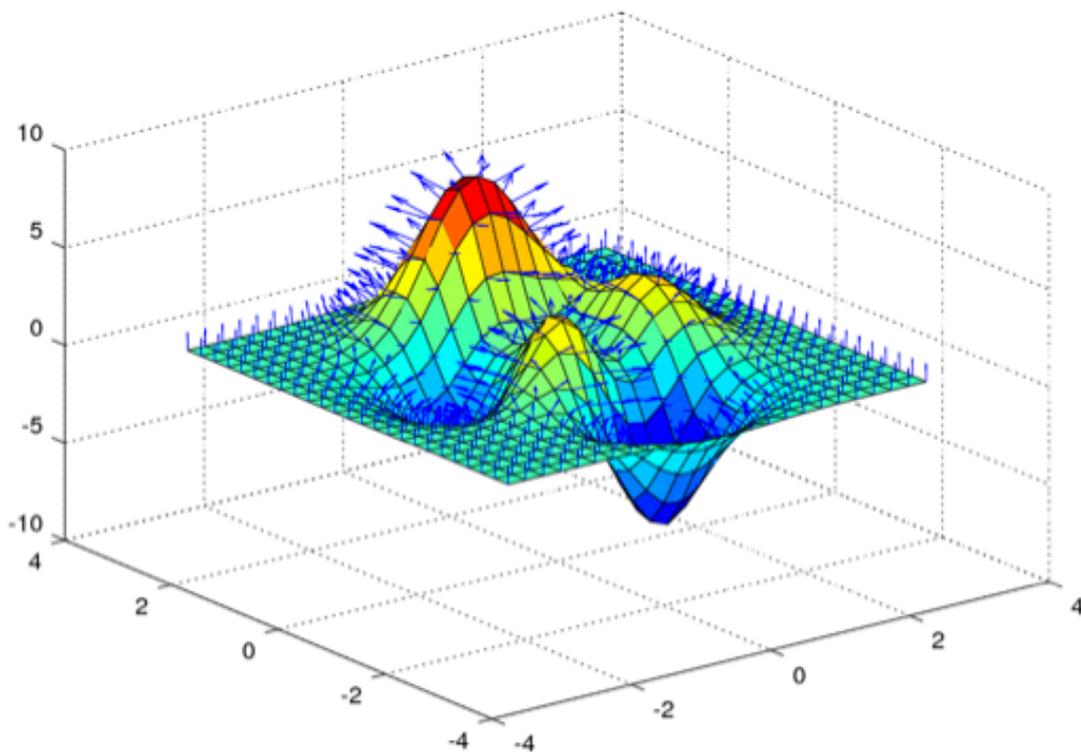


Représentations graphiques multidimensionnelles

Comment représenter des données en n dimensions ?

La plupart des analyses multifactorielles sont exprimées à l'aide de graphes en 2 dimensions, souvent par méconnaissance des possibilités de regroupement visuel de plus de 2 variables ou par crainte d'incompréhension par le lecteur. Les scientifiques savent utiliser les graphes à 3 dimensions, agrémentés parfois d'une «plume» figurant une 4ème dimension.

quiver3_2_1.png



source : http://octave.sourceforge.net/compare_plots/

Par exemple, comment, d'un seul coup d'oeil, visualiser température, vitesse et direction du vent, humidité, pression,.. en même temps sur un graphe jour/année, ou mettre ensemble visuellement les variables d'une évolution financière, afficher la répartition des accidents routiers dans l'espace-temps, avec leurs caractéristiques, éditer le résultat d'un sondage, rechercher les environnements favorables à la flore ou à la faune, ou les critères qui fondent la mixité sociale dans l'habitat...

Les tableurs classiques proposent des graphes en 3 dimensions, soit en perspective, soit en graphe à plat avec colorié selon la valeur de la donnée. Il est possible d'ajouter d'autres dimensions afin de corrélérer plusieurs variables entre elles, sans forcément exiger du lecteur un grand niveau d'abstraction.

Exemple : Production solaire

La production d'une toiture solaire en fonction de la période de la journée et du jour dans l'année.

		2013 Mars																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
4:00																																	
4:10																																	
4:20																																	
4:30																																	
4:40																																	
4:50																																	
5:00																																	
5:10																																	
5:20																																	
5:30																																	
5:40																																	
5:50																																	
6:00																																	
6:10																																	
6:20																																	
6:30																																	
6:40																																	
6:50																																	
7:00																																	
7:10																																	
7:20																																	
7:30																																	
7:40																																	
7:50																																	
8:00																																	
8:10																																	
8:20																																	
8:30																																	
8:40																																	
8:50																																	
9:00																																	
9:10																																	
9:20																																	
9:30																																	
9:40																																	
9:50																																	
10:00																																	
10:10																																	
10:20																																	
10:30																																	
10:40																																	
10:50																																	
11:00																																	
11:10																																	
11:20																																	
11:30																																	
11:40																																	
11:50																																	
12:00																																	
12:10																																	
12:20																																	

La représentation est selon le principe :

L'échelle des valeurs est entre le noir (production nulle la nuit) et le rouge (production voisine du maximum possible). En abscisse les jours et en ordonnées les tranches de 10mn de la journée de 4h du matin à 22h.

Sur http://ertia2.free.fr/Niveau2/Projets/PV/Graphes_Photovoltaiques.html, on peut voir 4 années de données superposées, représentant 108*365*4 soit 150 000 valeurs. Le fait de représenter les années en alignement vertical consitue en soi une dimension supplémentaire, puisque le graphe permet de comparer les données sur plusieurs années. On peut aussi distinguer, en partie inférieure de chaque graphique la courbe des productions de chaque jour et en dessous, les hauteurs de pluie, soit deux dimensions supplémentaires facilement visibles.

L'ensemble est un graphe à 9 dimensions :

kWh/10mn, jour, mois, trimestre, année, cumul jour, cumul mois, cumul trimestre, pluie, heures de soleil/jour, auxquelles il convient d'ajouter une dimension implicite : l'indisponibilité de la donnée.

Exemple : Analyse du trafic routier

http://siredo.free.fr/pages_menu/zoom.htm

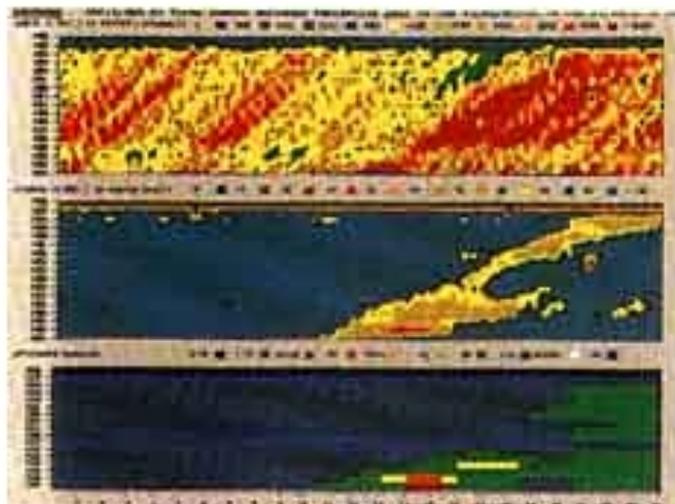
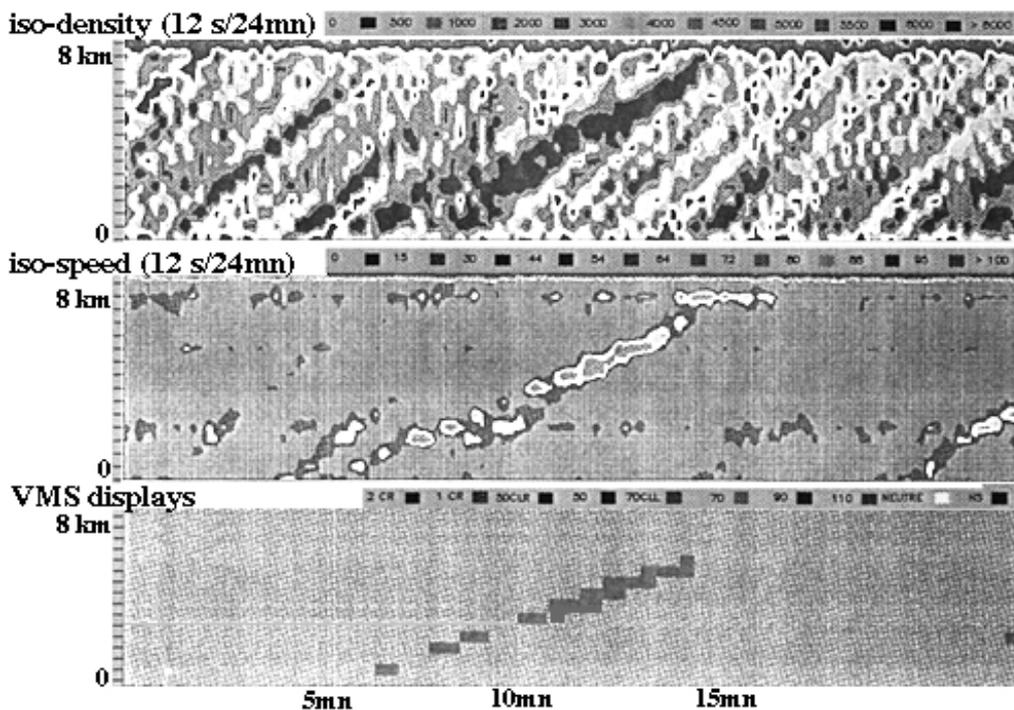
Ce graphe à n dimensions (espace ; temps ; voie ; vitesse - intervalle - longueur - durée d'occupation) qui représente environ 8000 données individuelles, est dérivé des écrans du système d'exploitation des autoroute de Marseille (MARIUS), réalisé vers 1991.

http://siredo.free.fr/pages_menu/Dernieres_mesures.htm

Ce groupe de graphes représente une situation animée virtuellement à partir des données réelles listées. Une dimension supplémentaire est le nuage de point caractéristique de la section d'autoroute concerné, sur lequel se superpose les données de la tranche de temps considérée.

Graphes iso-traffic de l'autoroute de Marseille

7 dimensions : Espace ; temps ; densité ; vitesse ; affichage des limitations de vitesse induite ; accidents (*information externe, superposée sur le graphe VMS, pas d'occurrence sur le graphe présenté*) ; état OK/HS des équipements (*absences de données, pas d'occurrence dans le grpahe présenté*)

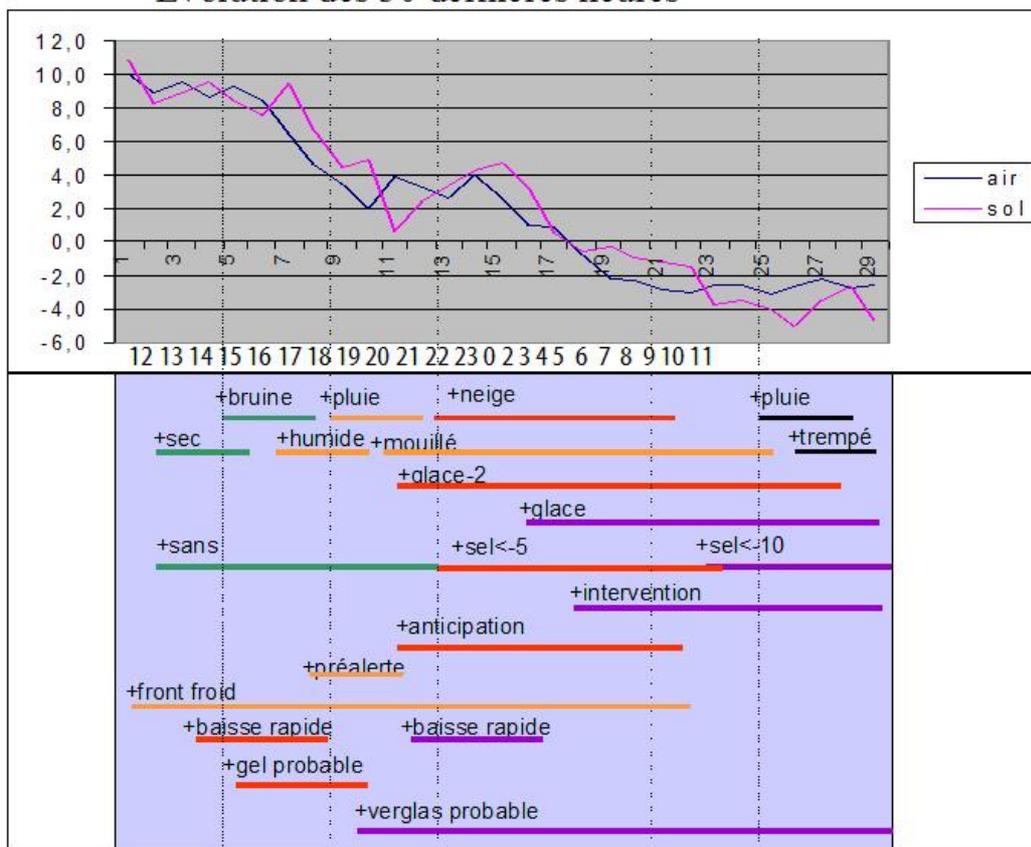


Graphisme d'évolution des vitesses sur autoroute. De haut en bas, courbes isodensités et isovitesses et affichage sur les PMV, pour une période de 24 minutes (en abscisse), sur l'A7, près de Marseille.

Exemple : Météorologie multidimensionnelle



Evolution des 30 dernières heures



Outre l'évolution des températures de l'air et du sol, le graphe inférieur renseigne les différents états de la chaussée et de l'air au cours du temps, soit une corrélation entre 14 critères.

Exemple : Suivi de la température

La température mesurée par une station météo peut être présentée de telle façon qu'au premier coup d'oeil on identifie les périodes de gel ou de canicule au cours de la journée et sur un trimestre entier, soit 48 X 90 jours =4320 données.



A ces données brutes, on peut associer, sur le graphe du dessous le cumul thermique de la journée. Le graphe en haut à droite donne le cumul thermique selon la demi-heure de la journée. Le graphe inférieur donne les minima et maxima de chaque journée, la moyenne de la journée, la moyenne entre le minima et le maxima et l'écart entre ces deux moyennes. Les petits graphes sont la distribution des minima et des maxima et la distribution des écarts max-min de chaque jour.

Cette visualisation graphique propose une quinzaine de dimensions.

Exemple : Suivi du vent

La composition suivante donne, jour par jour sur un trimestre, la distribution des rafales selon leur direction (échelle de gris), l'intensité des rafales (échelles de couleur), le cumul venteux de chaque jour, la distribution des direction des rafales pour chaque mois et un graphe multiple avec les maxima, minima, moyenne de la journée et moyenne mini/max.

Une telle composition est à première vue "désordonnée". Elle nécessite une compréhension préalable. Ces différents graphiques ont pour vocation de s'intégrer dans une visualisation de l'année entière sur plusieurs années, à l'instar des graphes de suivi de production photovoltaïque.



Exemple : Suivi d'une éolienne

<http://ertia2.free.fr/Niveau2/Projets/Meteo/Eolienne/Eolienne.html>

Cette représentation est une maquette réalisable à partir des données existantes (qui pourrait faire l'objet d'un projet pour étudiant en informatique)

Conclusion

Visualiser plusieurs milliers de données multidimensionnelles est possible. Cette approche visuelle est complémentaire de l'approche statistique et souvent utile à la vérification de la cohérence des résultats des calculs abstraits sujets à erreurs cachées de programmation.

Voir aussi :

http://ertia2.free.fr/Niveau2/Blogrinages/Blogrinages_ici_et_la/Donnees_bien_donnees.pdf

http://ertia2.free.fr/Pages_liees/Transparence_%20donnees%20routieres_v2.pdf

Annexe

Dans un graphe à plat avec couleur, chaque cellule prend une couleur selon la donnée qu'elle contient.



Le nuancier est à choisir en fonction de l'objectif recherché. Par exemple dans le suivi de la production photovoltaïque, le noir et le gris sont utilisés pour illustrer la nuit et les faibles luminosités. Pour d'autres applications, le noir et le blanc peuvent être exclus du nuancier pour servir à d'autres figurations, comme la panne du capteur ou la donnée douteuse.

Pour ajouter une 4ème dimension, on peut jouer sur la forme de la cellule : carré, triangle, cercle, polygone, flèche, icône.... Cette forme est déterminée par la valeur de la donnée. Chaque forme peut représenter une caractéristique particulière, par exemple, une station météo en bord de mer est à différencier d'une station météo en altitude, ou appartenant à un même réseau.

Pour une 5ème dimension, on peut jouer sur la taille de la forme. Ce type de variation n'est envisageable que sur un jeu de données restreint à afficher sur des cases suffisamment grandes pour que les variations soient faciles à distinguer.

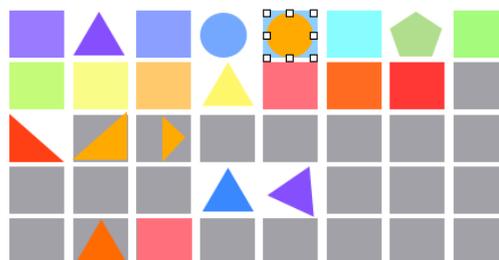
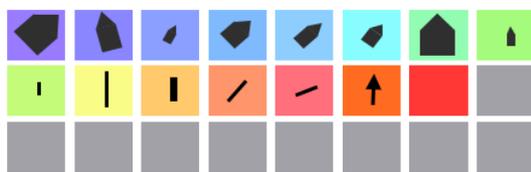
Pour une 6ème dimension, on peut jouer sur l'orientation de la forme, pour autant que celle-ci soit identifiable. Un triangle équilatéral et un triangle rectangle n'ont pas les mêmes possibilités d'orientations.

Pour une 7ème dimension, on peut superposer un objet de taille variable dans la cellule, par exemple un trait d'épaisseur et/ou de longueur variable. Encore faut-il que la forme de la cellule permette cette inclusion.

Pour une 8ème dimension, on peut orienter le trait de 0 à 180°.

Pour une 9ème dimension, on peut ajouter au trait ou au rectangle une flèche à l'une ou l'autre des extrémités.

Il est possible d'ajouter d'autres indices, mais peu à peu la lisibilité du graphe diminue. Il s'agit donc, en fonction des objectifs d'analyse visuelle recherchés de limiter la variabilité de chaque dimension.



Développement

L'application consiste à éditer plusieurs tableaux superposés de n lignes et n colonnes ou chaque case contient un objet dont la représentation varie selon la forme, la couleur, la taille ou la direction selon les valeurs contenues dans les tables correspondantes.

Les données sont prises dans différentes feuilles de calcul (.xls ou .csv ou .odt), à raison d'un indice/type de donnée par feuille. Les données sont des valeurs numériques ou des chaînes de caractère appartenant à une liste pré-définie par l'utilisateur.

9	4	7	7	7	0
0	7	9	5	1	2
2	8	3	5	5	2
1	3	7	9	3	3
2	7	1	10	10	4
3	3	8	3	1	9
6	7	6	2	2	7
8	8	3	5	6	3
0	4	7	6	3	5

Sable	terre	calcaire	granite	foret	urbain

Les données de la feuille 1 déterminent le nombre de lignes et de colonnes du graphe.

S'agissant d'une représentation graphique, le nombre de lignes et de colonnes, c'est à dire le nombre de case doit être compatible avec la définition de l'écran ou du support de présentation. Par exemple, une feuille A3 format horizontal peut proposer un graphe 3D de 400x290 valeurs, avec des cases de 1mm de coté. En passant à 4 dimensions, les cases doivent faire au moins 4mm sur 3mm. Le nombre de valeurs, soit 100 colonnes de 4mm sur 90 lignes de 3mm est alors de 900 + 900.

Les données de la feuille f sont prises dans les feuilles suivantes, avec le même nombre de lignes et de colonnes.

L'application superpose sur la grille initiale un calque par indice.

Le choix d'un indice détermine les données à représenter sur le calque correspondant.

L'ordre des indices est choisi pour que les représentations visuelles ne se masquent pas les unes les autres.

L'application permet de sélectionner un type d'indice par feuille ainsi que les bornes de changement de l'indice.

Il est souhaitable que tous les choix soient accessibles dans une seule fenêtre pour un meilleur confort du paramétrage, avec le nom de la feuille en regard de chaque indice.

Les listes des bornes sont dans des tableaux à 2 lignes : la représentation sur la première et les bornes ou les chaînes de caractère correspondantes.

Indice	par défaut	Paramètres
Couleur	aucune	- Choix du nombre de nuances - Choix du nuancier - choix des bornes de changement de nuance
Forme de base	cellule entière	- Choix de la forme : - rectangle horizontal pleine largeur x h (%) - rectangle vertical pleine hauteur x l (%) - carré pleine largeur ou hauteur - cercle plein largeur ou hauteur - Triangle équilatéral/isocèle - pentagone - autre (icone...)
Taille de la forme de base	100	- Choix du nombre de tailles différentes - Choix des tailles en % de la cellule
Orientation de la forme de base	vertical	360°/n où n est le nombre d'orientations possibles
Forme incluse	vide	- rectangle, carré, cercle, triangle, carré
Taille de la forme incluse	100 %	- l%xl%
Orientation de la forme incluse	horizontal	- 360°/p où p est le nombre d'orientations possibles
Fléchage	aucune	- trait avec flèche début et/ou fin
Orientation de la flèche	vertical	360°/t où t est le nombre d'orientations possibles

La validation du paramétrage déclenche la mise en graphique des données sur les calques correspondants, avec une ligne de titre contenant un lien sur la légende des différents symboles utilisés. La légende peut éventuellement être incluse au milieu du graphe en un plusieurs morceaux de positions et de tailles ajustables (auto-suffisance du graphe)

Le résultat graphique est imprimable et exportable en .pdf, .jpg, .png (par défaut sous le nom du fichier source des données).

Ces spécifications sont un exemple. C'est au concepteur, selon le jeu de données et selon les objectifs de son illustration, de trouver l'ergonomie visuelle la mieux adaptée et la plus digeste pour soi-même et surtout pour les autres. Il convient de vérifier que la définition visuelle de l'image d'origine ne soit pas dégradée lors du transfert sur un diaporama ou sur sa version papier. Les outils informatiques sont parfois facétieux ! L'aide à un (bon) ergonomiste peut être utile, sur le fond comme sur la forme. De plus, le public ciblé peut avoir des préjugés face à ce qu'il ne comprend pas d'emblée. Une approche pédagogique, dimension par dimension, est souhaitable.