

Conférence "Mathématiques et Grandes Dimensions"

Explorer des données désagrégées pour un non-statisticien

Frédéric Vautrain

vautrain@isthma.com

isthma

Les 3 V + 2 du big data

« Les entreprises les plus chanceuses sont capables d'analyser 7 % de leurs données tout au plus » (IBM)

Volume (nombre de lignes)

Variabilité (dimensions et référentiels)

Vélocité (granularité temporelle)

Visualisation (« Simplexifier »)

Valeur (tous les acteurs)



Data deluge - The Economist

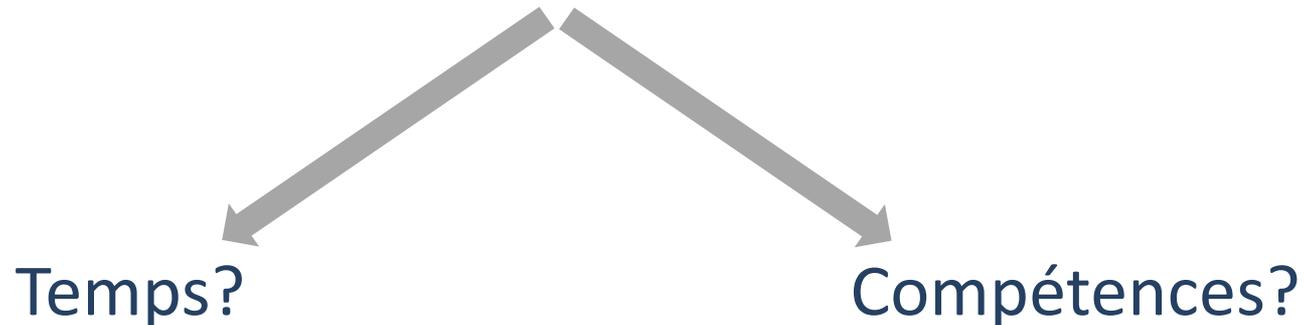


Les métiers face aux données

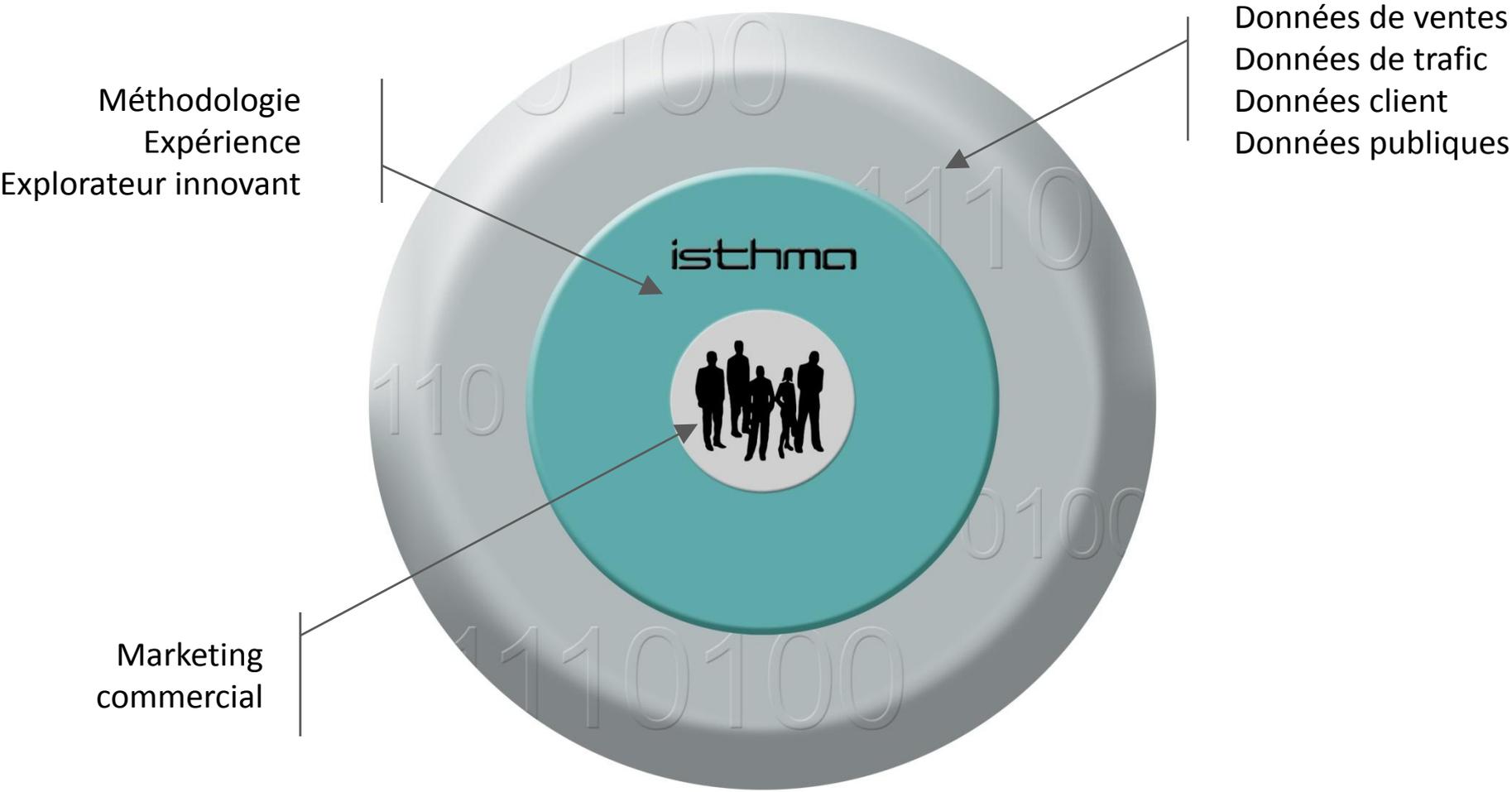
Des aides à la décision, stratégique et tactique

- Client (CRM) et produit (prix, cross selling)
- Segmentation, prévision...

Une appétence croissante à la donnée



Un promoteur de l'analyse de données



Notre approche: la BI Analytics

Prendre des décisions: pertinentes et rapides

- Une triple compétence
- Une suite logicielle DeltaMetric

Etre pro-actif sur les données

Rendre plus autonome nos clients

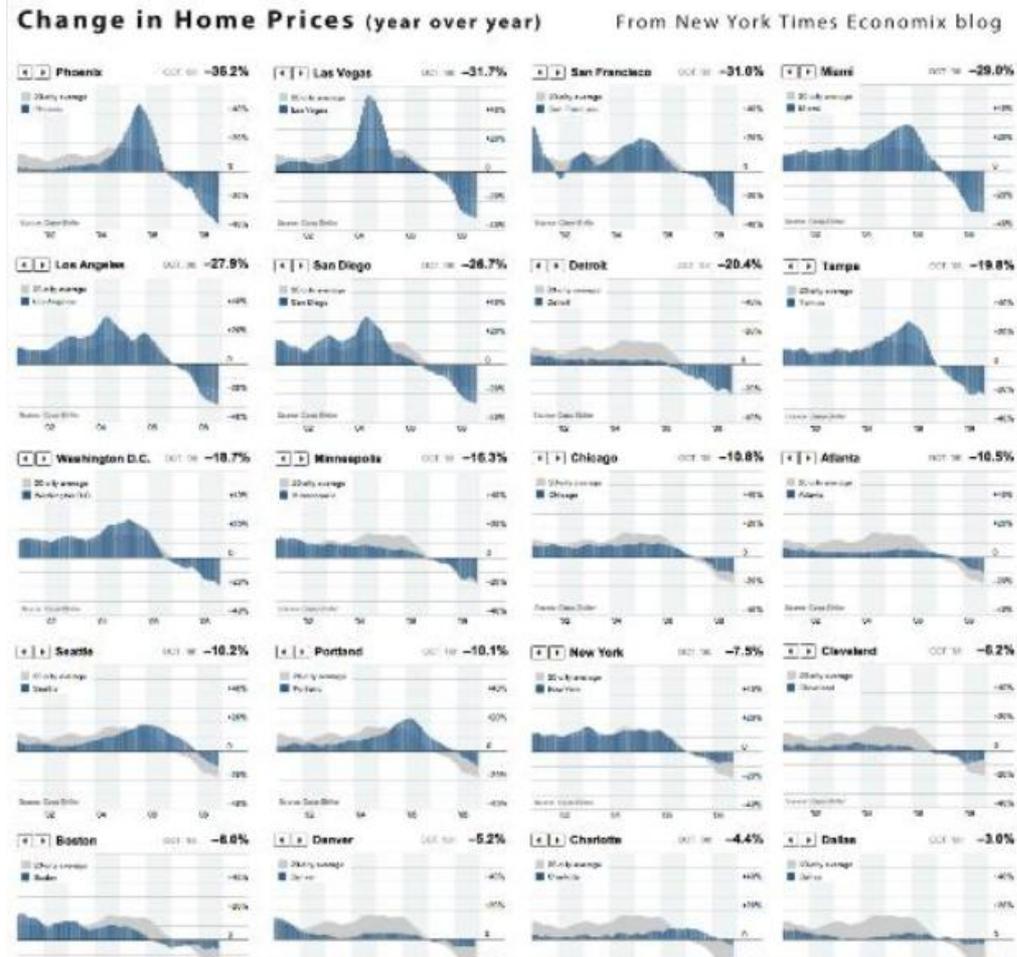


Les matrices de graphes

Densifier l'information pour éviter
l'agrégation



Les matrices de graphes: visualisation + restitution



Les small multiples (E. tuffte) dans le NYT

Pourquoi les small multiples?

Une étude: IEEE Symposium on Information Visualization 2008

« *Effectiveness of Animation in Trend Visualization* », Microsoft research, 2008

Contexte et objectif

- Données Banque Mondiale / ONU
- 3 visuels différents (Animation, traces, small multiples)
- 18 personnes questionnées: « *select 3 countries whose rate of energy consumption was faster than their rate of GDP per capita growth* »
- 3 indicateurs séparés: précision, vitesse et satisfaction ressentie
- 2 objectifs: présentation et analyse

Les types de restitution proposés

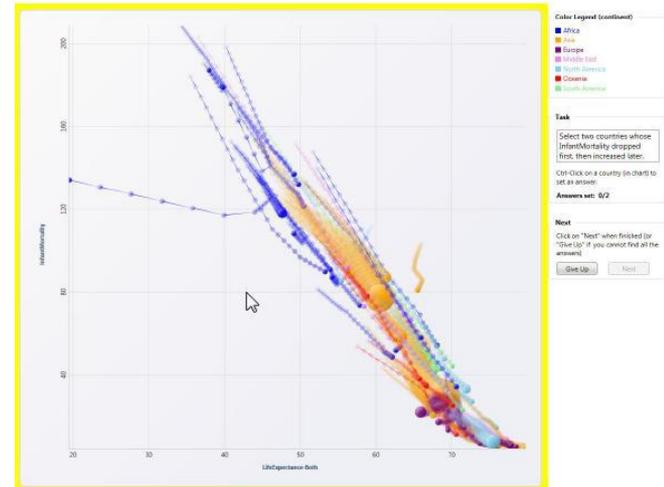
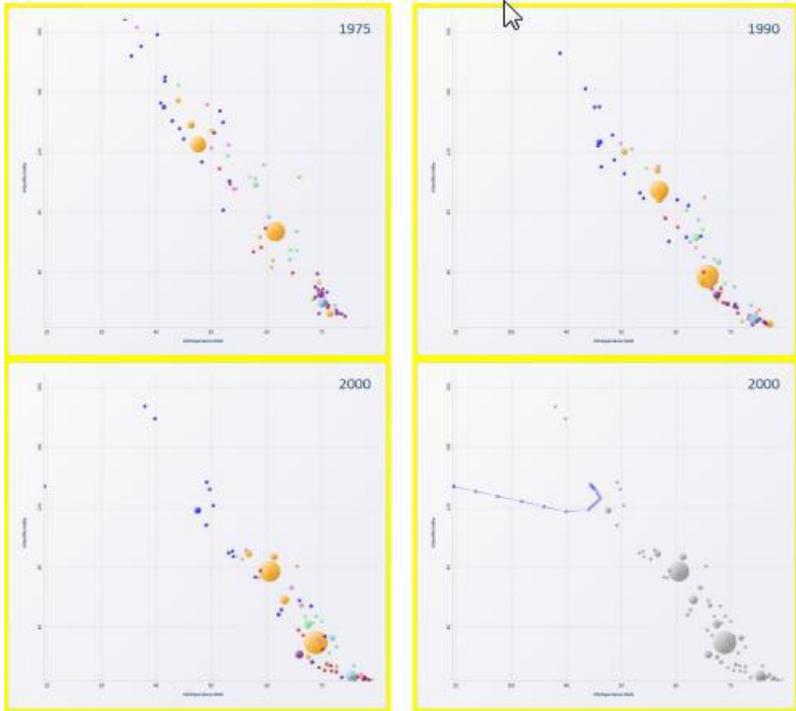


Fig. 2. Traces Visualization shows all trace lines simultaneously.

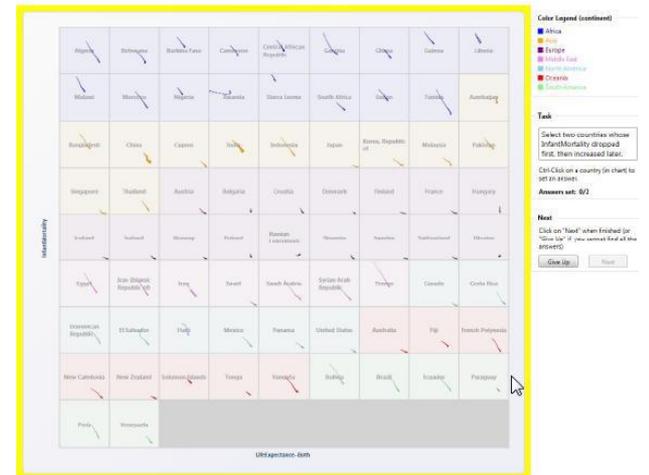


Fig. 3. Small Multiples Visualization shows trace lines for each country separately.

Résultats

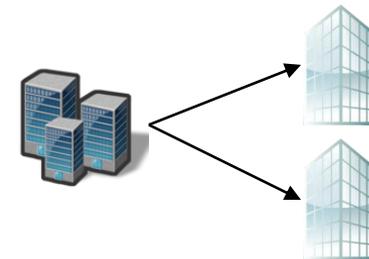
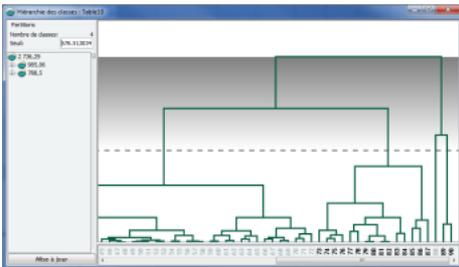
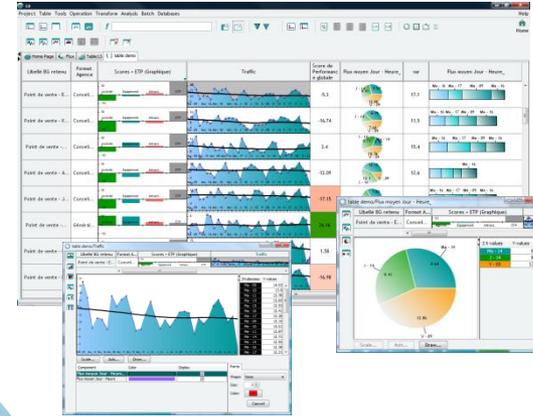
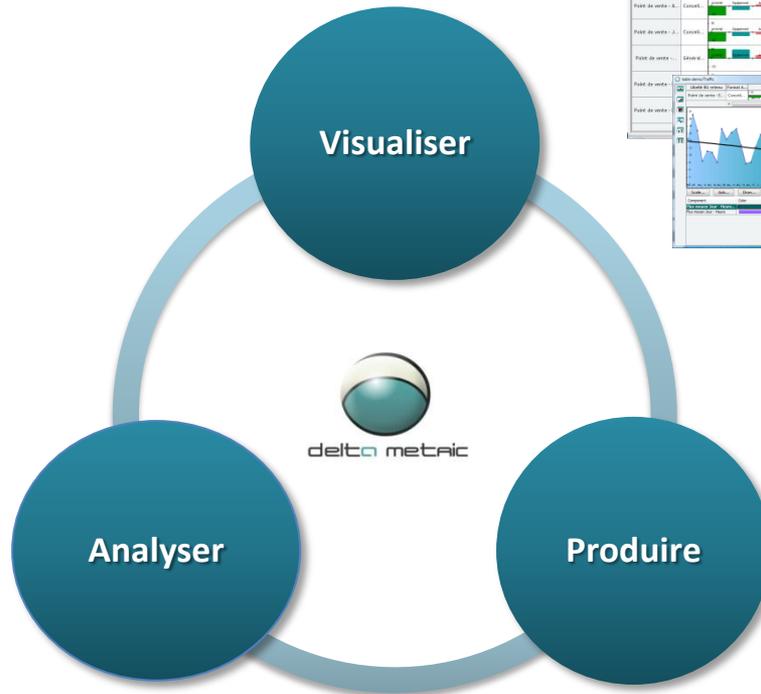
L'animation est la meilleure pour la présentation

L'usage des Small Multiples est meilleur pour l'analyse

Le visual mining avec DeltaMetric®

Explorer rapidement

Créer un dialogue avec le métier



Étudier la variation des données sur de multiples dimensions

Index	SHOP ID	PRODUCT ID	DATE	VALUE	REGION	POPULATION	EMPLOYEES	AREA (m ²)
00 001	Shop1	Product A	01/11/2011	277	Basse Normandie	1-low	7	2 100
00 002	Shop1	Product B	01/11/2011	243	Basse Normandie	1-low	7	2 100
00 003	Shop1	Product C	01/11/2011	758	Basse Normandie	1-low	7	2 100
00 004	Shop10	Product A	01/11/2011	127	Provences Alpes ...	1-low	7	2 100
00 005	Shop10	Product B	01/11/2011	313	Provences Alpes ...	1-low	7	2 100
00 006	Shop10	Product C	01/11/2011	249	Provences Alpes ...	1-low	7	2 100
00 007	Shop100	Product A	01/11/2011	408	Franche Comte	2-medium	7	2 100
00 008	Shop100	Product B	01/11/2011	255	Franche Comte	2-medium	7	2 100
00 009	Shop100	Product C	01/11/2011	381	Franche Comte	2-medium	7	2 100
00 010	Shop101	Product A	01/11/2011	251	Provences Alpes ...	2-medium	7	2 100
00 011	Shop101	Product B	01/11/2011	681	Provences Alpes ...	2-medium	7	2 100
00 012	Shop101	Product C	01/11/2011	373	Provences Alpes ...	2-medium	7	2 100
00 013	Shop102	Pr				1-low	7	2 100
00 014	Shop102	Pr				1-low	7	2 100
00 015	Shop102	Pr				1-low	7	2 100
00 016	Shop103	Pr				3-high	7	2 100
00 017	Shop103	Pr				3-high	7	2 100
00 018	Shop103	Pr				3-high	7	2 100
00 019	Shop104	Product A	01/11/2011	71	Pays de la Loire	1-low	7	2 100
00 020	Shop104	Product B	01/11/2011	147	Pays de la Loire	1-low	7	2 100
00 021	Shop104	Product C	01/11/2011	105	Pays de la Loire	1-low	7	2 100
00 022	Shop105	Product A	01/11/2011	168	Pays de la Loire	2-medium	7	2 100
00 023	Shop105	Product B	01/11/2011	212	Pays de la Loire	2-medium	7	2 100

30000 lignes x 8 colonnes

magasins

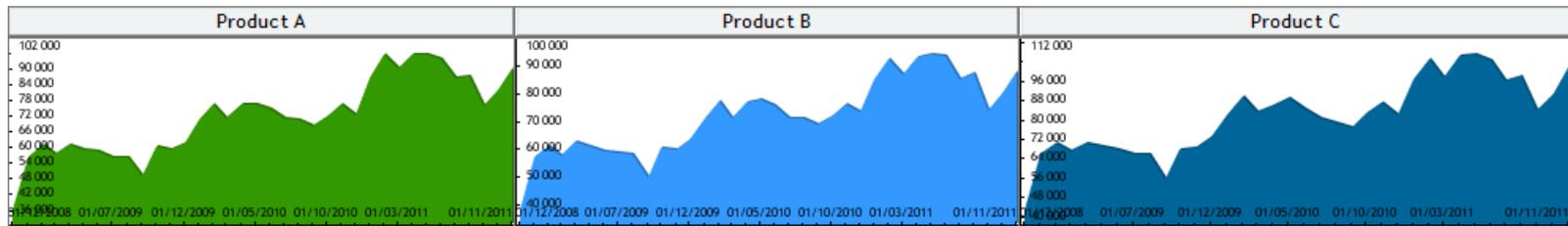
produits

temps

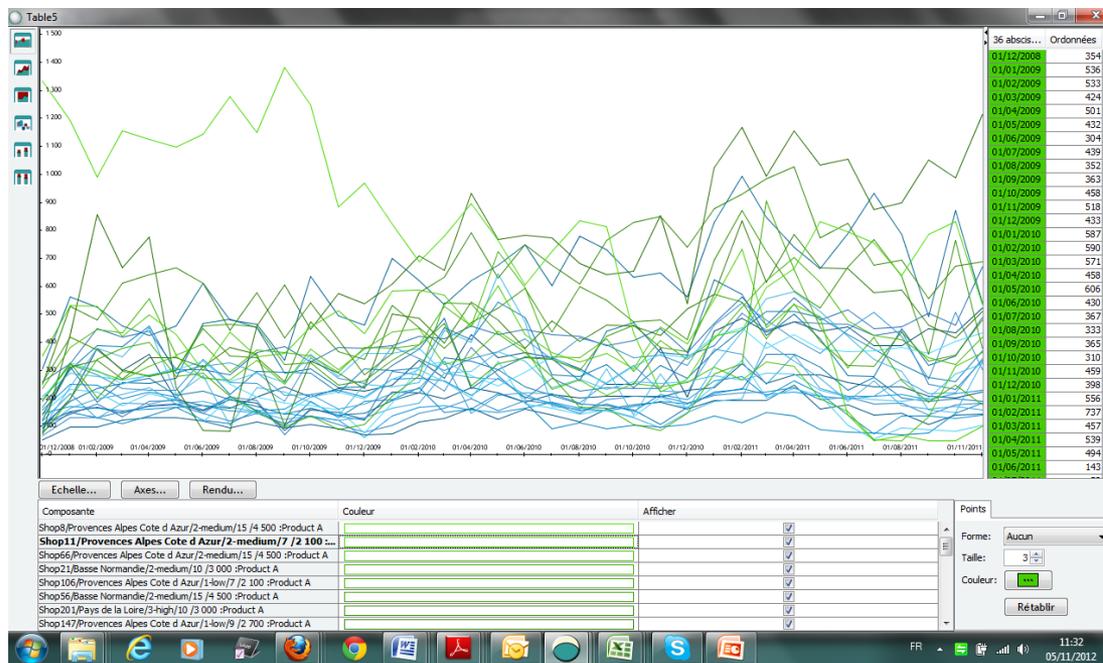
indicateur

Socio-démo

Comment explorer toutes les situations



Total vente par produit: tout va-t-il très bien?



Affichage de 10 magasins sur 280...

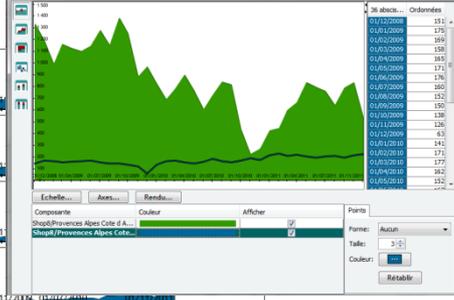
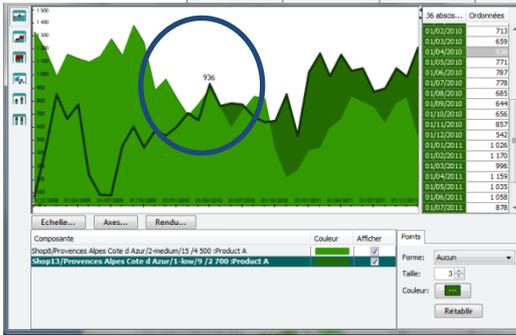
Densifier l'information dans les cellules



Extraire des indicateurs et les interpréter

Cannibalisation et ventes croisées

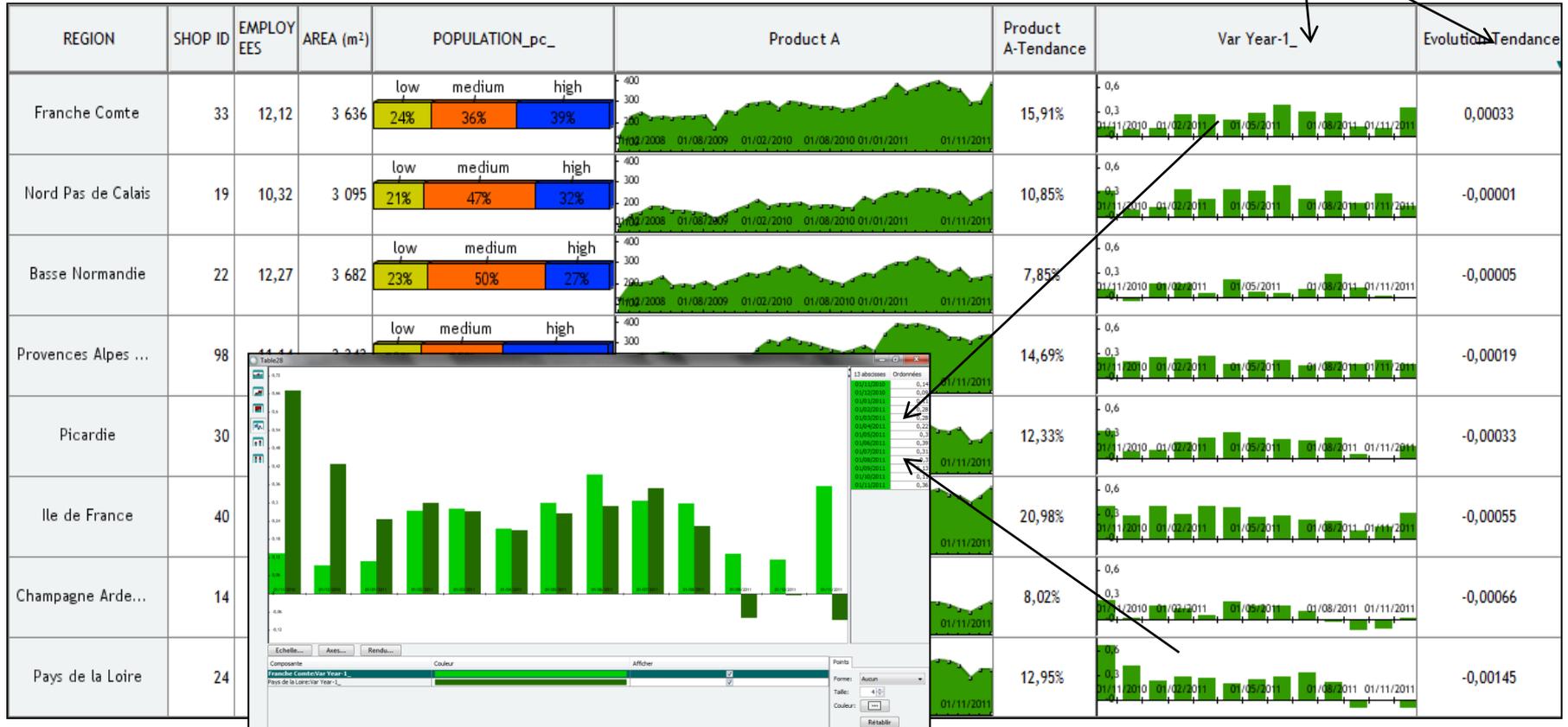
SHOP ID	REGION	POPULATION	EMPLO YEES	AREA (m ²)	Product A	Product A-Moyenne des effectifs	Product A-Tendance	Product B	Product C	Corr A et B
Shop8	Provence...	2-medium	15	4 500		839	-0,7			-0,601
Shop11	Provence...	2-medium	7	2 100		399	-0,3			0,946
Shop200	Provence...	3-high	10	3 000		169	-0,2			0,927
Shop6	Basse No...	1-low	15	4 500		205	-0,1			0,972
				4 500		288	-0,1			0,956
				2 100		228	-0,1			0,969
				2 700		136	-0,1			0,992
				4 500		747	0,4			
Shop129	Provence...	2-medium	9	2 700		450	0,5			
Shop49	Provence...	3-high	12	3 600		580	0,5			
Shop248	Ile de Fr...	3-high	12	3 600		458	0,5			
Shop55	Franche...	3-high	12	3 600		376	0,6			0,981
Shop63	Franche...	1-low	15	4 500		539	0,6			0,962
Shop13	Provence...	1-low	9	2 700		722	0,7			0,947



Une vision régionale

Valeur, tendance et évolution

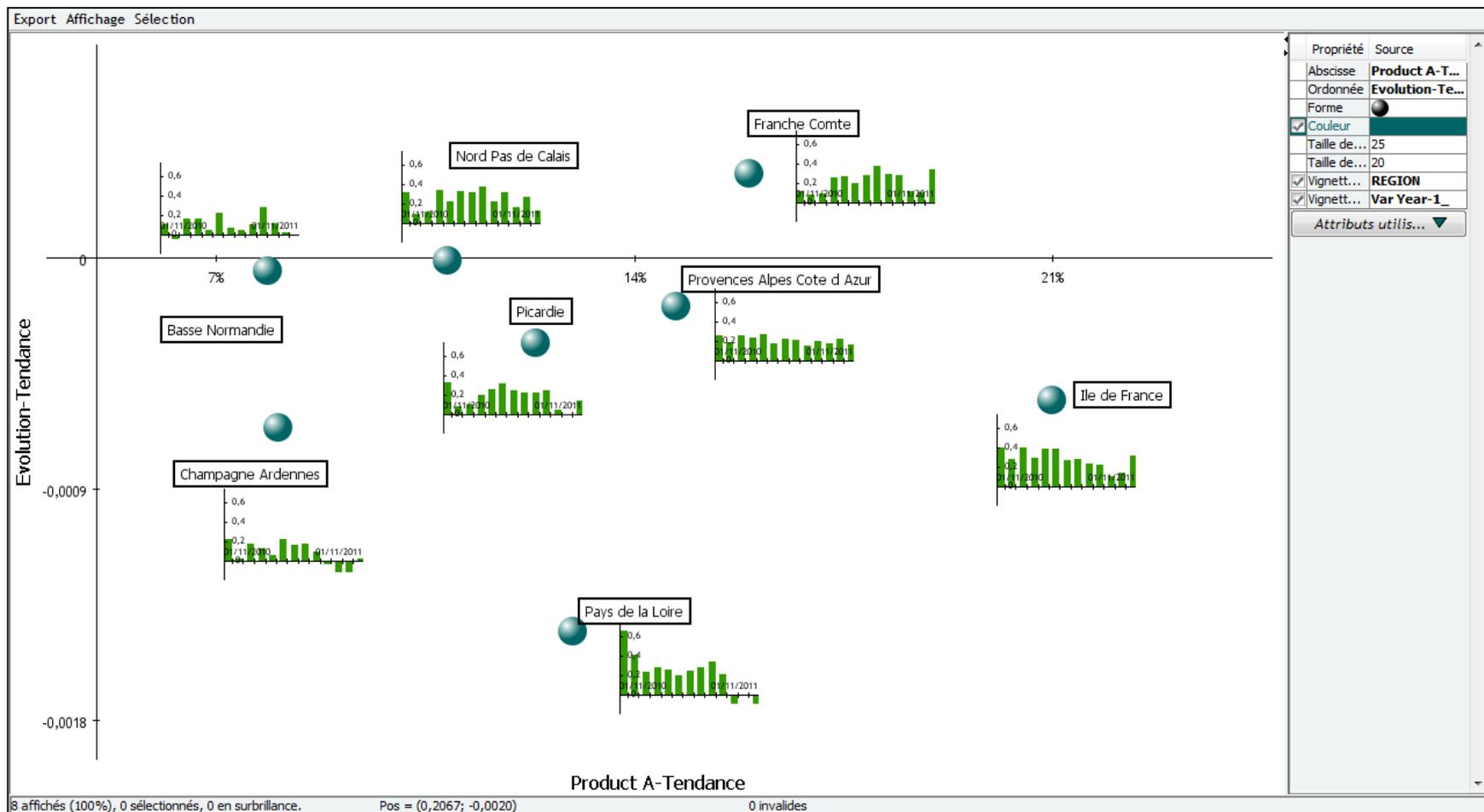
Nouveaux indicateurs



F. Comté et P. de Loire: 2 situations différentes

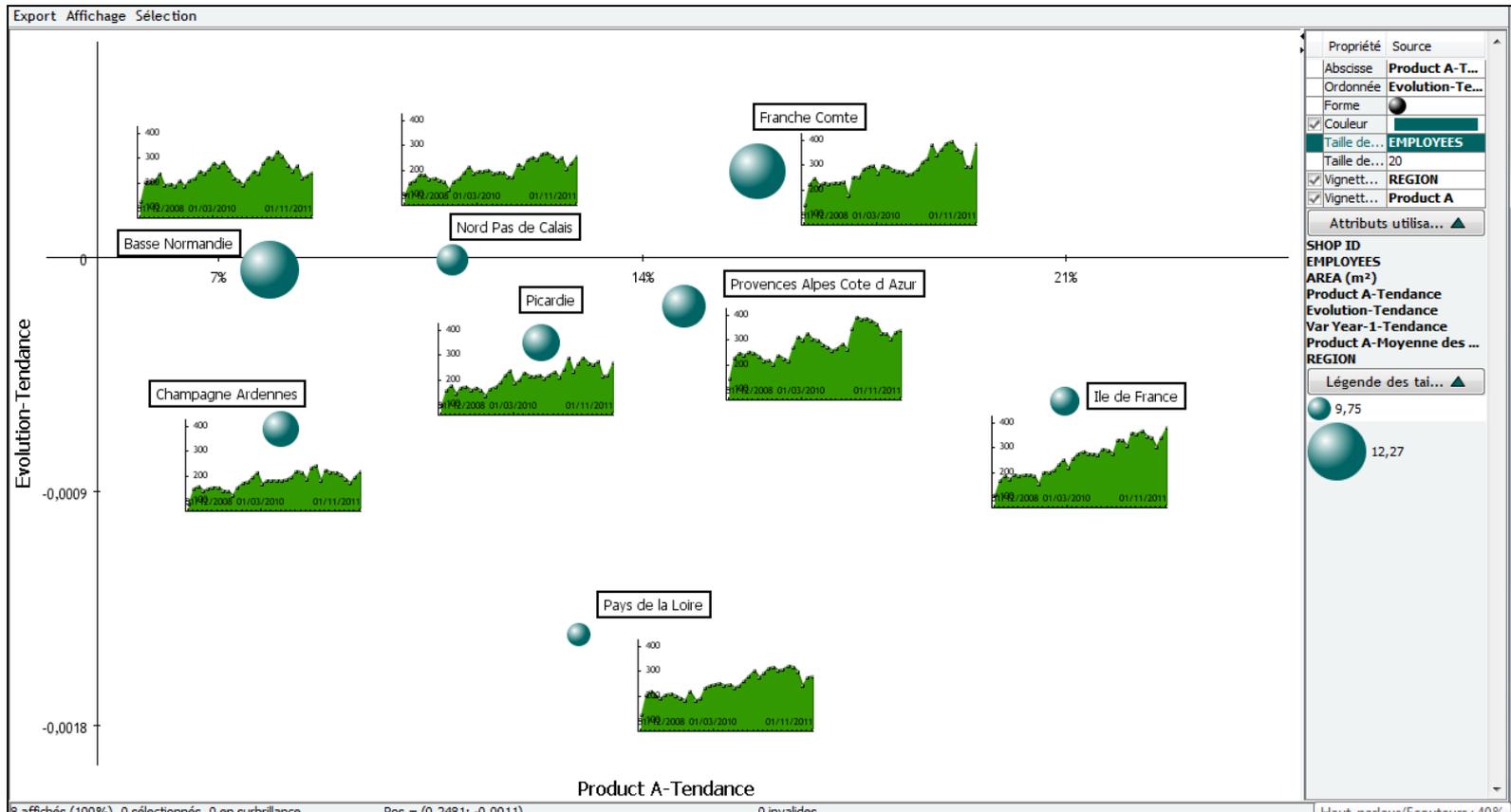
Une vision régionale

Valeur, tendance et évolution



Une vision régionale

Valeur, tendance et évolution



Nombre d'employés semble avoir un effet sur l'érosion de la croissance