

Chiffres ANIMÉS

LAURENT.Kling@epfl.ch, EPFL-STI



La civilisation moderne exige la manipulation et l'analyse de grandes quantités de données. Une manière élégante est de créer un graphique, qui comme le veut l'adage, vaut plus de 1000 mots. Avec les outils informatiques actuels, il est aisé de réaliser un histogramme, un camembert ou une autre représentation graphique.



UN COCKPIT D'AVION POUR UN SIMULATEUR D'INTERVENTION TECHNIQUE

Par contre, son interprétation peut laisser à désirer. La facilité de manipulation obtenue est généralement faible, par exemple pour la cotation d'Apple dans la Bourse américaine avec CNN Money.

Chaque nouvelle version nécessite l'entrée de la consigne suivie de la génération et de l'affichage du graphique.



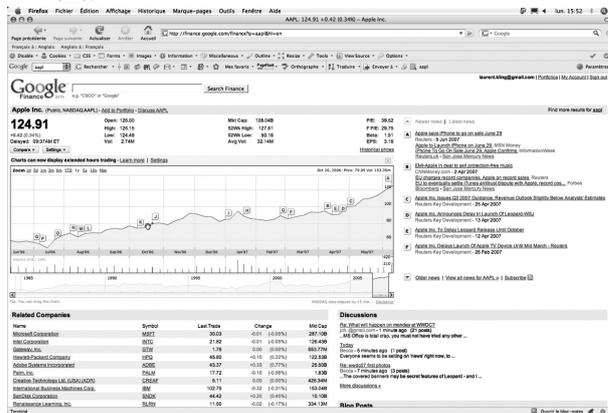
UN COURS SPÉCIFIQUE EN 3 ÉTAPES

UNE NOUVELLE DIMENSION, L'INTERACTIVITÉ

Le principal intérêt de la technologie Web 2.0 est de proposer pour l'utilisateur la possibilité d'interagir directement sur les données. Par exemple, **GoogleMap** ou **Google Document** et **Tableur** utilisent ce genre de technique. **Google Finance** est confronté au même problème d'affichage des données de la Bourse américaine, mais offre en plus :

- l'affichage direct de la valeur, du volume et de la date des transactions ;

- la mise en relation du cours avec des dépêches ;
- la modification et déplacement de la fenêtre temporelle.

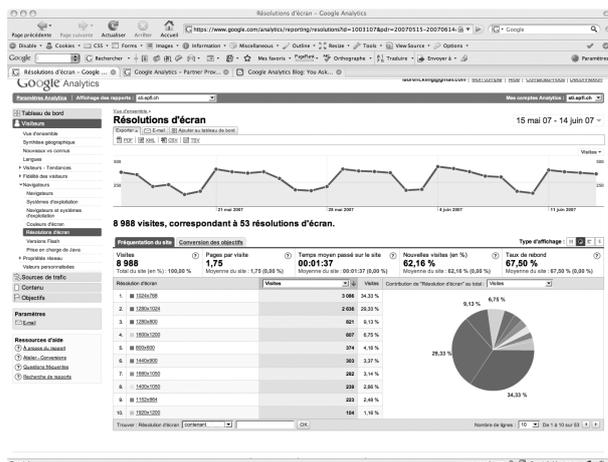


UN GRAPHIQUE DIRECTEMENT MODIFIÉ PAR L'USAGER

L'influence de la manipulation directe avec une souris renvoie à des temps préhistoriques les interfaces découpant le processus de visualisation des données en trois phases séparées :

- consignes ;
- calcul des données ;
- affichage.

GRAPHIQUES MULTI VALEURS

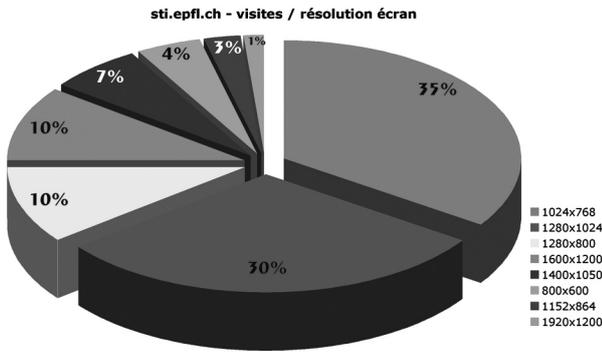


LA DISTRIBUTION DE LA TAILLE DES ÉCRANS DES USAGERS DU SITE **sti.epfl.ch** SUR UNE PÉRIODE D'UN MOIS AVEC GOOGLE ANALYTICS

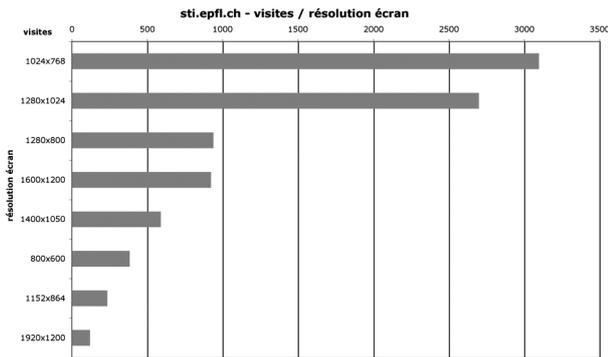
L'idée immédiate qui vient à l'esprit quand on souhaite comparer deux séries de valeurs entre elles est d'utiliser une représentation cartésienne, les abscisses et les ordonnées représentant respectivement chacune des catégories. Si cette figuration est familière, elle peut vite tourner au cauchemar si le même raisonnement est appliqué à l'utilisation de l'espace pour trois séries de valeur. Comme exemple, je désire répondre à la question suivante: *Quelle est la résolution la*

plus adaptée pour la mise en page d'un futur site Web de la Faculté STI?

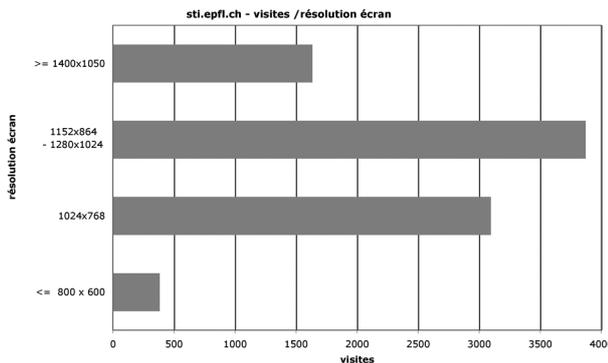
Pour éviter tout *a priori*, je me base sur des données statistiques fournies gratuitement par Google Analytics. Les lecteurs intéressés par cet outil peuvent parcourir cet article (ditwww.epfl.ch/SIC/SA/SPIP/Publications/spip.php?article1190). Malheureusement, les informations brutes ne me sont pas très utiles, car trop détaillées. Utilisant un outil familier à beaucoup d'usagers, Excel, voici quelques tentatives de présentation.



CAMEMBERT 3D, TOUTE RESSEMBLANCE AVEC DES GRAPHIQUES DÉJÀ OBSERVÉS EST PUREMENT FORTUITE



X-Y TRIÉS PAR LE NOMBRE DE VISITES



X-Y REGROUPÉS PAR CATÉGORIE

Le premier graphique correspond à un camembert, à part une discussion sur les couleurs, il est difficile de faire un choix. La représentation circulaire est pernicieuse, car il est délicat de mettre en relations des valeurs angulaires. Pour corser le tout, j'ai rajouté une 3e dimension avec une projection axonométrique qui n'amène aucune lisibilité supplémentaire.

Le deuxième est une représentation classique en abscisse-ordonnée (x-y) avec un tri basé sur le nombre d'usagers. Dans cette figure, je serais tenté de choisir une résolution de 1024x768. La vue linéaire permet de facilement comparer des valeurs entre elles. Dans ce cas, on peut être induit en erreur par l'ordre utilisé. Par exemple, où se trouve la ligne des écrans 800x600?

Pour limiter l'étendue des données, un regroupement en 4 catégories d'écrans est créé :

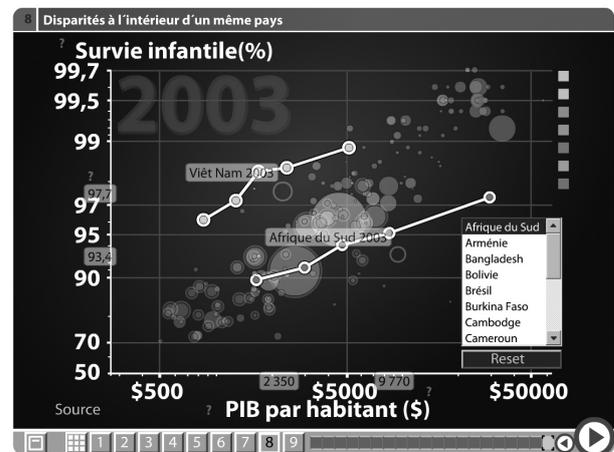
- Petits, jusqu'à 800x600
- Moyens, 1024x768
- Grands, 1280x1024
- Très grands, 1400x1050 et au-delà...

Ce dernier exemple est probablement le plus objectif, la résolution de travail doit se situer entre 1280x1024 et 1024x768.

Malgré les apparences, les données sont les mêmes entre ces 4 images, uniquement le mode de représentation et le regroupement changent !

Avec une mise en page Web fluide qui s'adapte à la résolution, incorporant le confort du choix de la taille des caractères pour l'utilisateur, ce choix de taille d'écran devient théorique. En réalité, la partie visible dans le navigateur ne représente que 80% de la surface disponible sur l'écran.

GAPMINDER



EN 2003, AVEC UN REVENU PAR HABITANT INFÉRIEUR DE TROIS QUARTS, 80 % DE LA POPULATION DU VIËT NAM POSSÈDE UN NIVEAU DE SOIN SUPÉRIEUR À CELUI DE L'AFRIQUE DU SUD

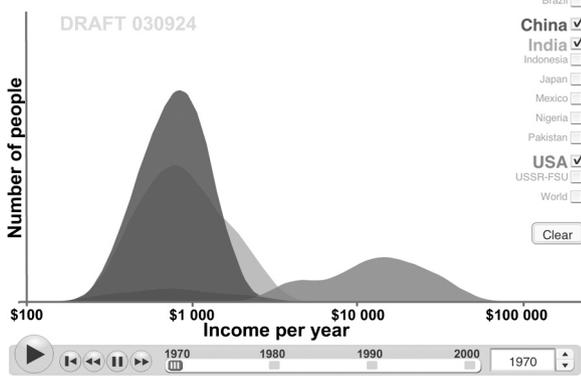
En politique, pour probablement compenser l'absence d'idées fortes, on assène souvent des chiffres sans préciser le contexte. Pour tenter de clarifier le débat dans le domaine de la santé publique, Hans Rosling, Professeur à l'université médicale de Stockholm (*Karolinska Institutet*), a développé une série de représentations dynamiques des données statistiques par pays.

Un travail disponible est la présentation de l'évolution humaine, réalisé en 2005 sur la base du rapport du PNUD (*Programme des Nations Unies pour le Développement*): hdr.undp.org/reports/global/2005/francais/. La version en technologie Flash avec Gapminder: www.gapminder.org/GapminderMedia/GapTools/HDT05L/application.swf.

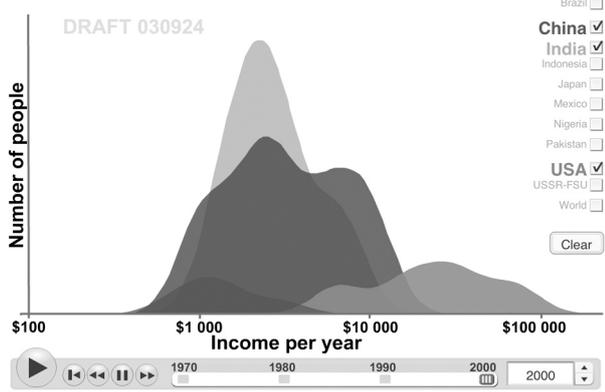
La visualisation de cette animation est particulièrement instructive, en particulier par la capacité de comparer les pays entre eux et d'explorer les zones géographiques pour présenter les disparités régionales.

Il est également intéressant de voir la variation de distribution des revenus entre 1970 et 2000: www.gapminder.org/Gapminder/Media/GapTools/IncomeDistribution2003/swfVersion/incomeDistrGraph_oct7_03.swf.

Income distribution 1970



Income distribution 2000



GAPMINDER CHEZ GOOGLE

Le logiciel utilisé par **Gapminder**, **Trendalyzer**, a été acquis par Google en mars 2007 (tools.google.com/gapminder/). Ainsi, on peut visualiser et comparer les statistiques de développement humain pour l'ensemble de la planète. Une présentation de cet outil est disponible en vidéo: video.google.com/googleplayer.swf?docId=7996617766640098677&hl=en.

L'animation de la comparaison entre deux critères de développement, par exemple entre le produit de la valeur ajoutée par habitant (GNP) et l'espérance de vie est déjà particulièrement intéressante. Si on rajoute la capacité de sélectionner certains pays et d'observer la trace qu'ils laissent sur le graphique, le tout animé sur une base temporelle, cela devient magique.

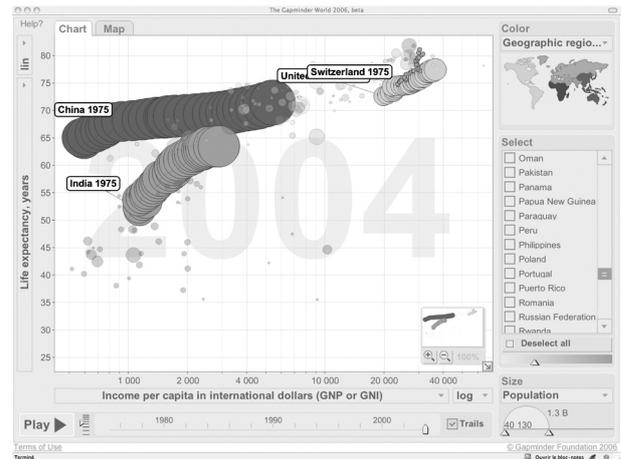
Sur un autre exemple, je désire comparer quelques pays de l'Europe et de l'Asie avec une population autour de 10 millions. Avec la recherche Google et le mot clé population, on me propose directement un lien parrainé sur Gapminder: tools.google.com/gapminder/index.html?i=SP.POP.TOTL.

Pour la Suisse, la Belgique, la Hongrie et le Kazakhstan, je désire comparer l'espérance de vie, le revenu par habitant, la taille des populations et leurs évolutions de 1975 à 2006, tools.google.com/gapminder/index.html?c=KAZ,CHE,HU,N,BEL. Plutôt que de séparer les pays par leur continent (c'est le même), je choisis de présenter un quatrième critère, le nombre d'utilisateurs du téléphone (fixe et mobile) pour 1000 habitants. Pour améliorer la vision, uniquement une zone significative du graphique est affichée. Il est intéressant d'observer les conséquences de la chute du mur dans deux pays du COMECON, en particulier la régression simultanée de l'espérance de vie et du revenu pour le Kazakhstan. Le revenu a rattrapé son niveau antérieur, mais l'espérance de vie est inférieure à son niveau de 1990!

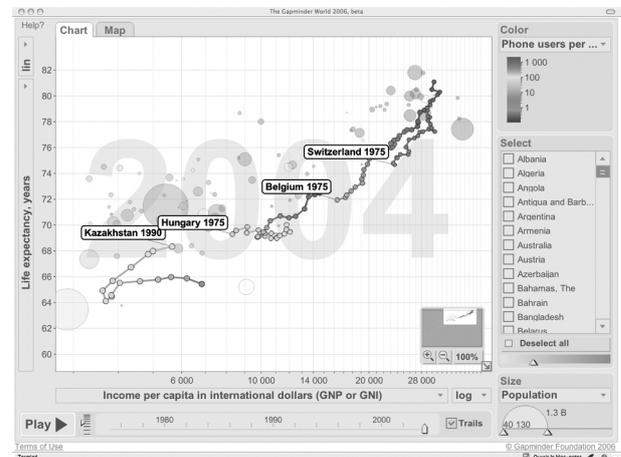
Autre point remarquable, pour la Belgique, le niveau de vie et de la santé ont crû plus rapidement que celui de la Suisse.

Sur une autre thématique, comparez le développement économique avec l'émission de dioxyde de carbone et la proportion de population urbaine.

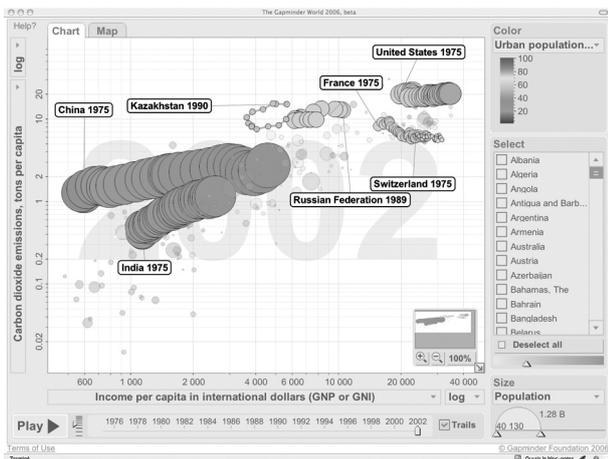
On voit que les USA sont bien le plus gros pollueur de la planète, que la réduction de CO₂ en Suisse est faible comparée à la France (probablement un effet du parc nucléaire français).



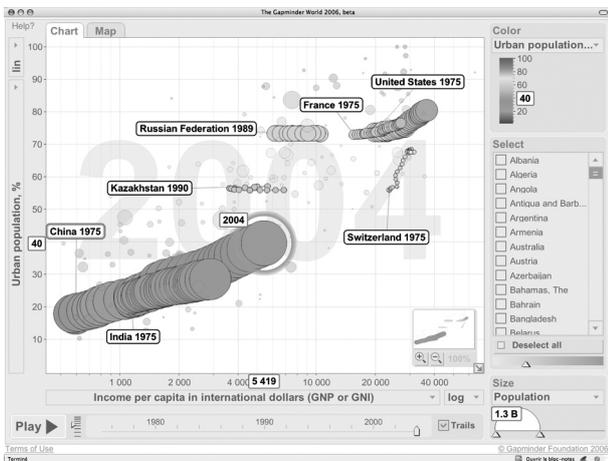
COMPARAISONS DE L'ESPÉRANCE DE VIE DES USA, SUISSE, INDE ET CHINE



DÉVELOPPEMENT ÉCONOMIQUE ET ESPÉRANCE DE VIE POUR LA SUISSE, LA BELGIQUE, LA HONGRIE ET LE KAZAKHSTAN



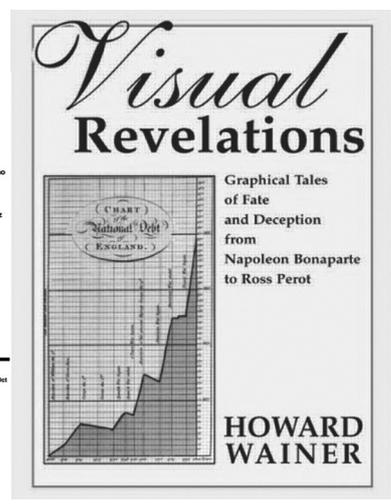
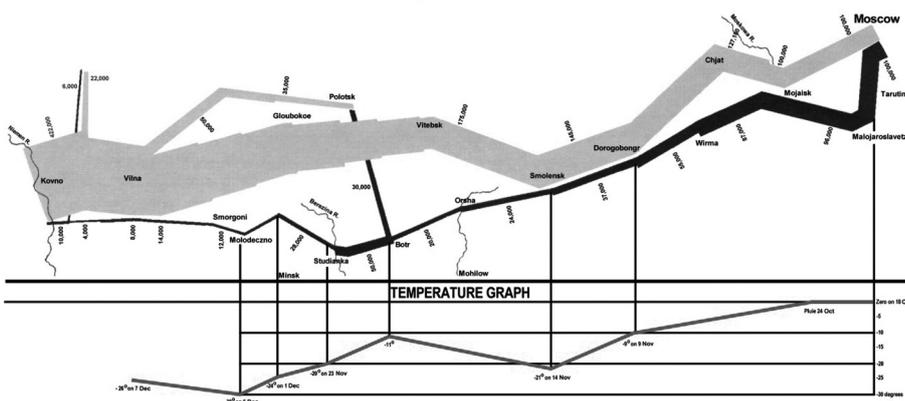
développement économique, production de CO₂, déplacement des populations vers les villes



développement économique et population urbaine

Pour le Kazakhstan, on observe également la conséquence de l'effondrement de l'URSS et la nette reprise de l'activité depuis 6 ans. La baisse du CO₂ entre ces deux périodes est le résultat de la perte d'industrie lourde, malheureusement compensée par un développement important de la circulation automobile.

Pour terminer, vérifions la relation entre le développement économique et la population urbaine. Contrairement à un *a priori*, l'exode rural est plus important en France qu'en Suisse, mais la tendance suisse est inquiétante. Les USA et sa civilisation de la voiture possèdent un développement urbain important. Par contre, la Fédération russe et le Kazakhstan



restent pratiquement à leur niveau respectif. Pour les deux plus grands, la Chine a dépassé l'Inde pour le développement économique et son exode rural. Gapminder représente en deux dimensions plus de 4 informations simultanément :

- le pays avec un cercle proportionnel à la taille de sa population,
- la relation d'un pays avec deux données statistiques,
- une autre donnée statistique par la couleur du cercle.

Et l'animation temporelle qui représente une 5^{ème} dimension ! Pour rendre possible la visibilité de l'évolution des années, la trace du cheminement des valeurs, présente sur un graphique l'ensemble de ces données et leurs interrelations.

Gapminder propose une souplesse de manipulation remarquable, l'usage de l'animation lui permet de produire des graphiques significatifs malgré la complexité des données.

GARDER UN ESPRIT CRITIQUE

Cependant, il convient de mettre en relation les informations, car il est aisé de faire passer une hérésie en innovation. Par exemple, la commune de Morges a eu l'idée d'utiliser l'eau du lac pour arroser le cimetière, si cette idée paraît lumineuse, son retour sur investissement est hasardeux. 4000 m³/an à frs. 1.20/m³, soit 4'800.-/an; avec l'eau du lac, le prix de revient baisse de moitié 4000 m³/an à frs. 0.60/m³, soit 2'400.-/an. Cette démarche semble logique jusqu'ici, mais avec le perfectionnisme qui nous caractérise, le coût total de l'installation est de frs. 275'000.-. Le temps de retour sur l'investissement est donc de 114 ans ! Il est certain qu'il faudra changer cette installation avant qu'elle soit amortie...

Pour finir cet article, je vous propose de lire l'excellent livre de **Howard Wainer**, *Visual Revelations: Graphical Tales of Fate and Deception From Napoleon Bonaparte To Ross Perot*, ISBN 0-387-94902-X.

De ce livre, un magnifique exemple français, de Charles Joseph Minard sur la campagne de Russie de Napoléon où un unique graphique est capable de représenter: la position des troupes sur une carte géographique, la quantité des troupes, les pertes subies, l'arrivée à Moscou, la température extérieure et la retraite qui suivit avec les ravages sur les soldats provoqués par le terrible hiver russe. ■