

VISUALISATION DES DONNEES HISTORIQUES

Comment analyser facilement, d'un seul coup d'oeil, des données historiques en comparaison avec les données d'aujourd'hui ?

Par exemple, comment situer la température sous Louis XIV par rapport à nos normales saisonnières ?

La première solution est d'exhumer les données numériques et de se livrer à de patients calculs statistiques dont la justesse et la pertinence sont difficiles à évaluer. Une erreur dans une formule est si vite arrivée, et si subtile à détecter.

Une deuxième solution consiste à visualiser graphiquement les données, à condition d'organiser les graphiques pour ne pas s'y perdre. Une erreur dans le titre ou dans la légende d'un graphe est si vite arrivée, et si subtile à détecter.

Malgré tout, l'analyse graphique est efficace quand elle organise visuellement des données en grand nombre. Par exemple, comment repérer 5 événements de même nature dans un lot de plusieurs milliers.

Exemple de graphe à 8 dimensions :

Le graphe ci-dessous permet de suivre la progression de deux camions sur une autoroute équipée de station de mesures de vitesse et de longueur de tous les véhicules tous les 500m sur 8 km.

- En abscisse, le temps sur une échelle de trente minutes par défaut
- En ordonnées, un graphe par station
- Chaque graphe de station est subdivisé avec un sous-graphe pour chaque voie

Chaque sous-graphe représente, sur l'échelle des temps, le passage d'un véhicule, sous forme d'un trait vertical :

- l'espace entre 2 traits verticaux représente l'intervalle temporel entre véhicules
- la hauteur est proportionnelle à la vitesse du véhicule (150km/h à pleine échelle par défaut)
- la couleur est fonction de la classe de longueur du véhicule (10 classes réglables)
- l'épaisseur est proportionnelle au temps d'occupation du capteur par le véhicule.

Au total : intervalle, vitesse, longueur, temps d'occupation, voie de passage, heure de passage, station de passage, soit 7 dimensions. auxquelles on a ajouté les états du système de régulation de vitesse affichant la limitation de vitesse sur le portique surplombant chaque voie de chaque station de mesure.

Ce graphique affiche environ 10 000 mesures brutes. Leur représentation visuelle, sous réserve que l'analyste se familiarise avec les codes visuels proposés, est immédiatement efficace, à comparer avec une analyse statistique, dont on est jamais certain de la fiabilité, tant les sources d'erreurs possibles sont nombreuses (sur quels jeux de données, comment sont filtrées les valeurs nulles ou aberrantes ?). L'analyse visuelle a l'avantage de piéger les incohérences, les jeux de données fallacieux...

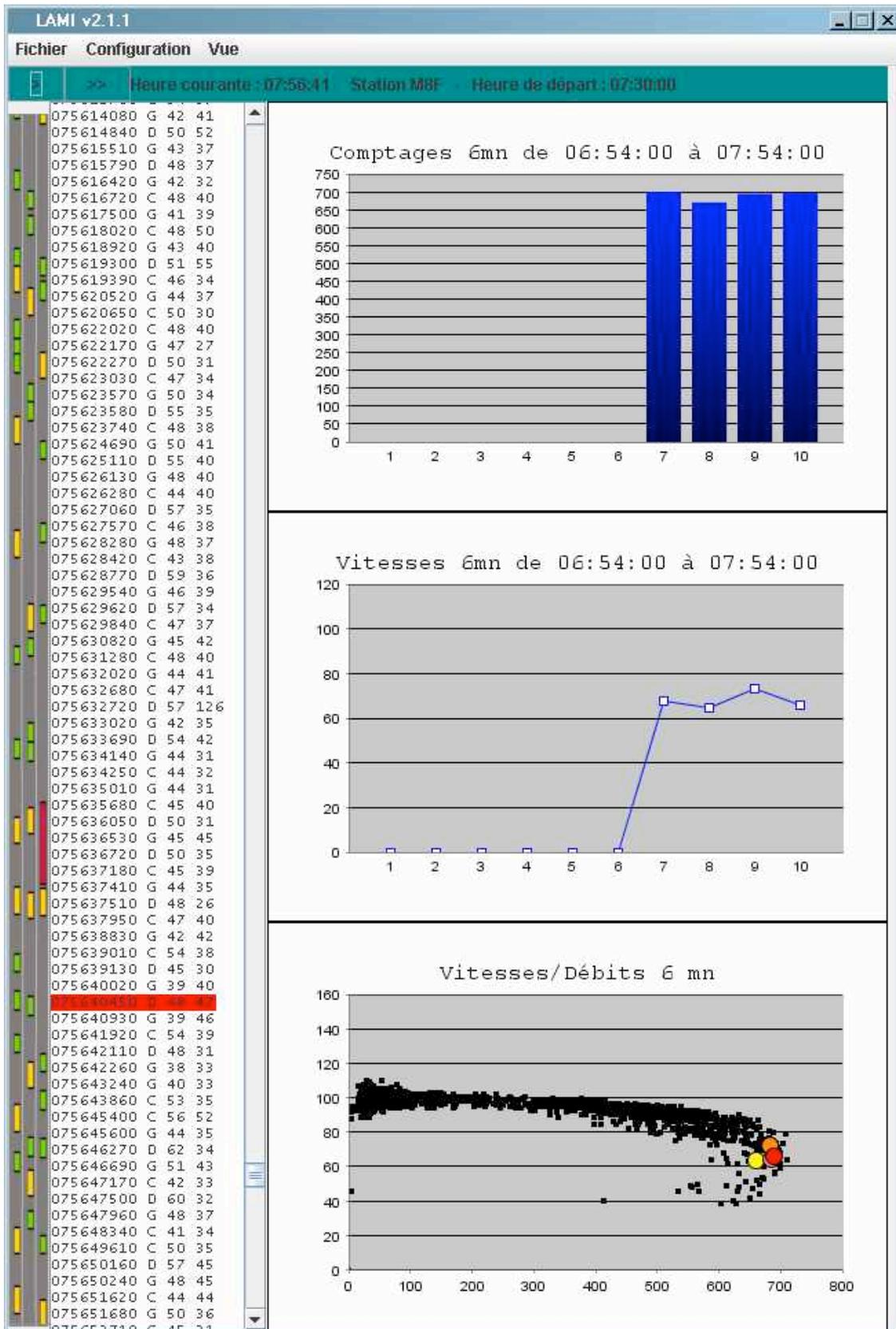


D'autres dimensions sont implicites :

- Un sous-graphe vide renseigne sur une panne de détecteur
- Un vide de données au même moment sur les 3 voies renseigne sur un défaut de transmission
- Un vide de données au même moment sur plusieurs stations renseigne sur un défaut de réseau
- Un graphe vide renseigne sur une panne de la station
- Une incohérence de couleurs entre deux stations successives qui mesurent globalement les mêmes véhicules renseigne sur un défaut de configuration des capteurs
- Une incohérence sur la progression des pelotons renseigne sur un défaut de l'horloge interne

Graphe dynamique

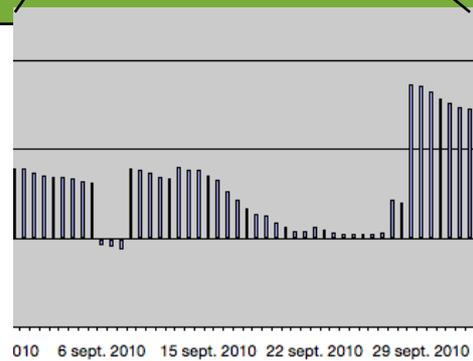
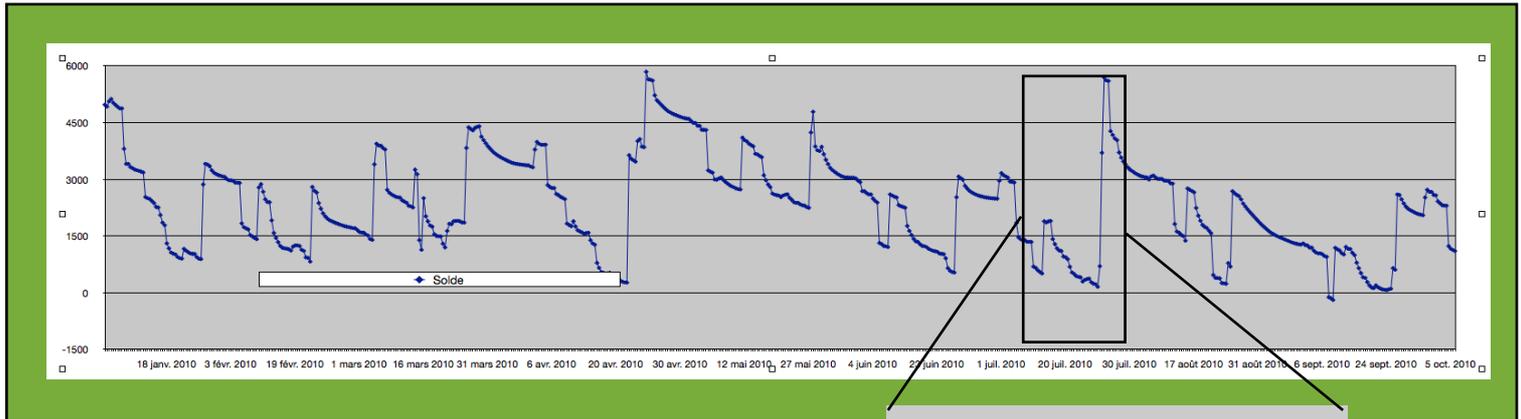
Le graphe précédent est statique. Il peut-être intéressant de représenter le trafic sur une seule station de façon dynamique, comme sur l'exemple ci-dessous. L'autoroute est simulée sur 300m par défaut. Chaque icône de véhicule est injectée à son heure de passage avec sa vitesse de déplacement rapportée à la longueur d'autoroute simulée. L'icône a une longueur déduite de la vitesse et du temps d'occupation et une couleur selon sa classe de longueur



La simulation dynamique peut visualiser plusieurs stations de suite. Le recouvrement en cascade permet de mieux repérer les pelotons.



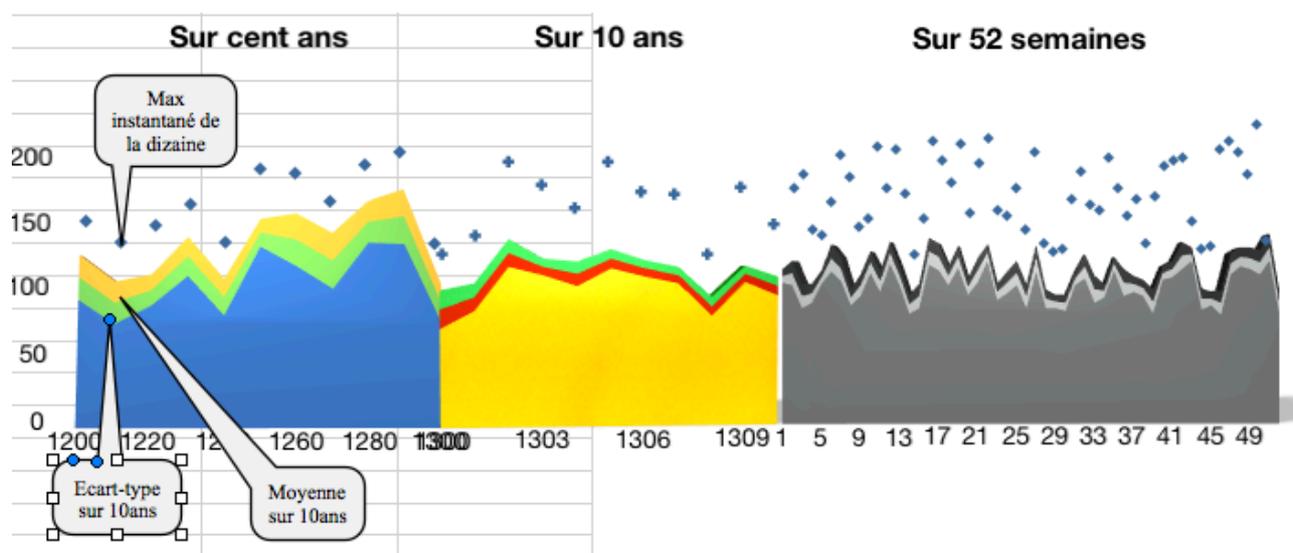
Graphes avec curseur-loupe



Les graphes avec curseur-loupe présentent l'intérêt de visualiser une longue période tout en accédant à volonté aux valeurs individuelles.

Graphe au temps logarithmique

Les données très nombreuses et très étalées sur l'échelle des temps ne sont pas directement visualisables. Par exemple, il est intéressant de disposer d'un même coup d'oeil des mesures fines récentes et des moyennes avec minimas et maximas plus anciens.



L'exemple s'arrête à la semaine, il pourrait être enrichi de la même manière avec les données de la semaine la plus récente, puis avec les 240 données des dernières 24h, si les données brutes sont recueillies toutes les minutes.

Il faut noter que les données mensuelles sont dangereuses, car elle représentent une période variant de 28 à 31 jours (10%), avec un nombre de WE variable. Toutes les données mensuelles liées aux activités humaines sont sujettes à ces variations. Il vaut mieux 52 semaines que 12 mois.

Sillages météo

Les animations des cartes météo représentent des états successifs discontinus. Si chaque perturbation laissait derrière elle son sillage comme un bateau, l'information serait plus riche.

Les spécifications sont sur http://ertia2.free.fr/Pages_liees/Sillages_meteo.pdf

Unités

Regrettons que Talleyrand n'ait pas su négocier avec les anglais quant au système métrique. Le monde a donc su resté poétique, avec des pouces et des décimètres. Fahrenheit et Celsius sont restés célèbres. Les livres et le rôl nous rappellent que l'Angleterre est une île, tandis que les gallons des militaires, nous permettent de ne pas faire le lien direct entre le prix du pétrole au sortir du derrick et le prix du litre d'essence que nous mettons dans notre voiture.