

Exploration Interactive de Données Spatio-Temporelles Brutes

Romain Vuillemot
LIRIS / Ecole Centrale de Lyon

Ma recherche s'intéresse au **design** et à l'**évaluation** de méthodes visuelles d'**exploration de données**, en particulier spatio-temporelles.

Doctorat (2006-2010)

INSA Lyon / LIRIS

Chercheur invité Université du Maryland

Post-Doctorat (2011-2013)

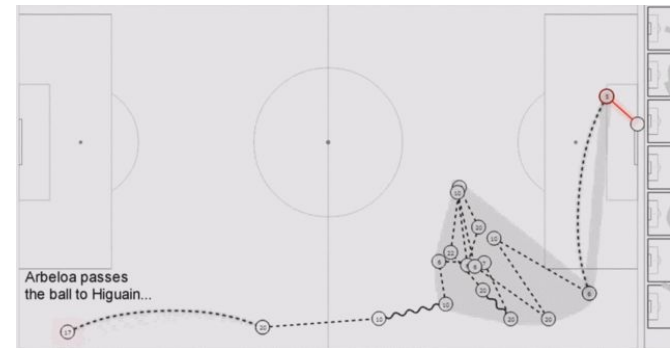
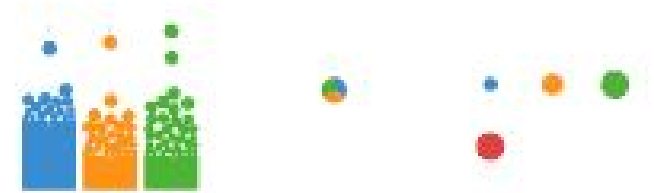
Equipe AVIZ - INRIA Saclay

Research Fellow (2013-2016)

Université d'Harvard

Enseignant-Chercheur (2016)

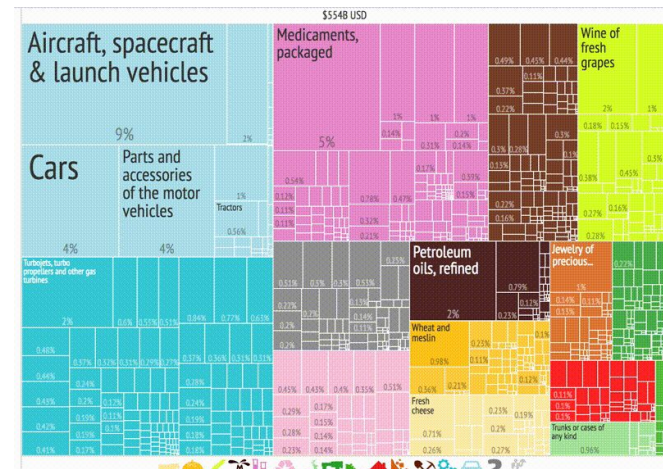
Ecole Centrale de Lyon / LIRIS



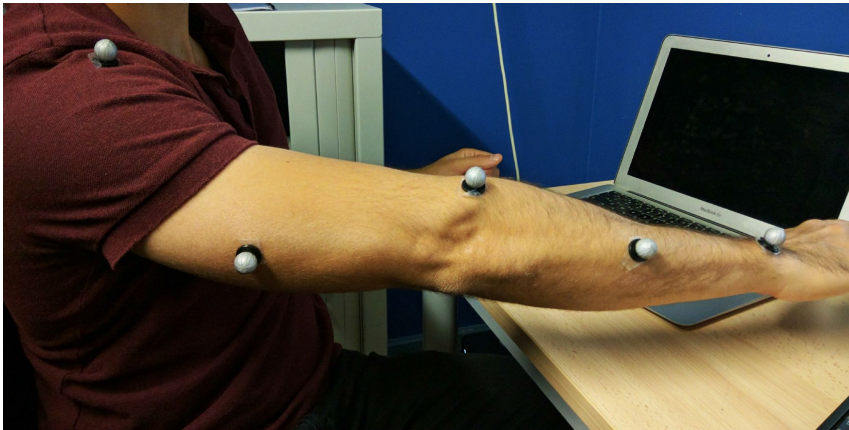
Journées VISU 2017 !

<http://visu2017.liris.cnrs.fr/>

7 et 8 juin 2017 à l'IFP



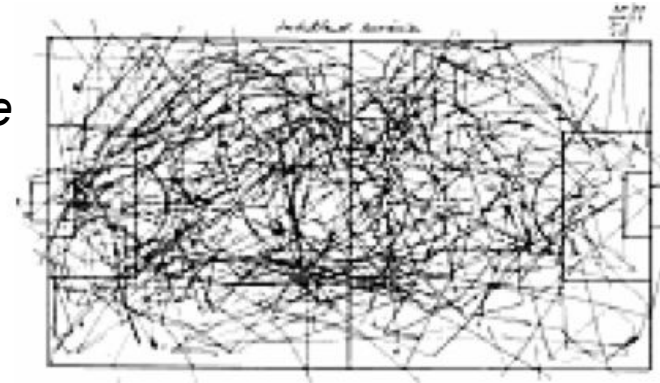
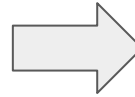
Les données spatio-temporelles sont de différentes natures. Nous les définissons comme toutes données avec **une position** (x, y) qui change au **fil du temps** (t)



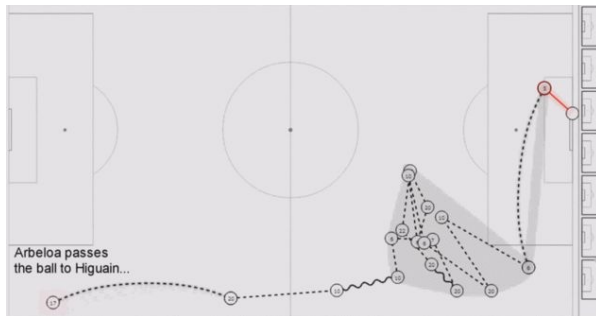
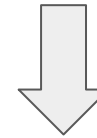
Les données sont cependant **très volumineuses**, avec une composante spatiale et temporelle, mais aussi beaucoup de dimensions supplémentaires



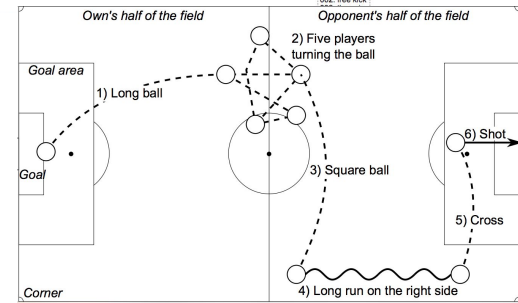
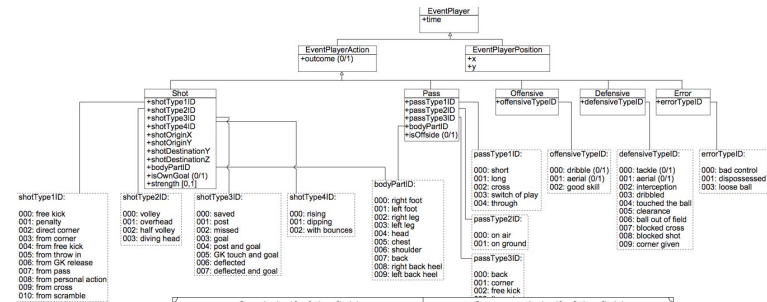
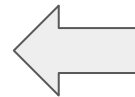
collecte



modélisation

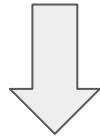


communication

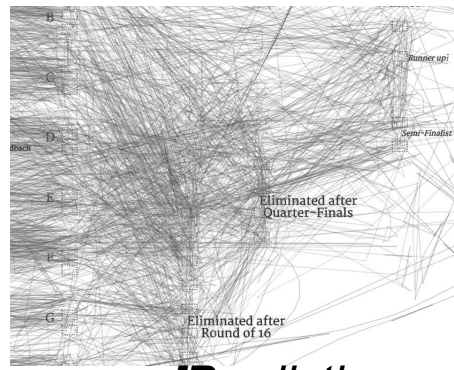
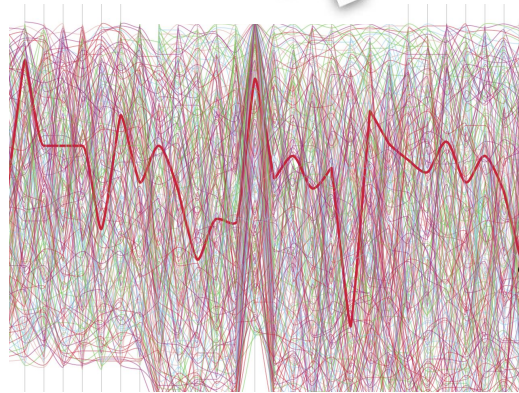


[*SoccerStories, InfoVis 2013*]

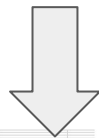
Il existe des méthodes de **conception et d'évaluation** d'interfaces interactives pour analyser les données **brutes**, les données **transformées**, et les **communiquer** efficacement.



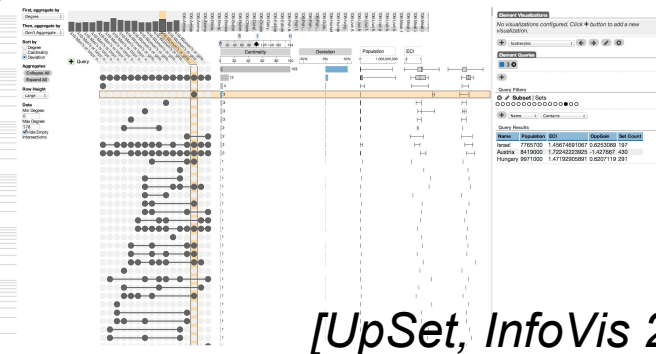
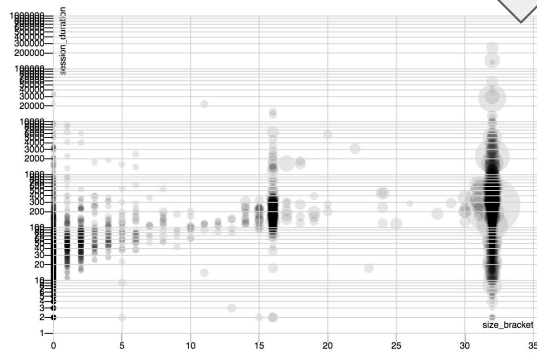
données brutes



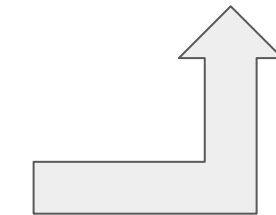
[Predictions, CG&A 2016]



traitement



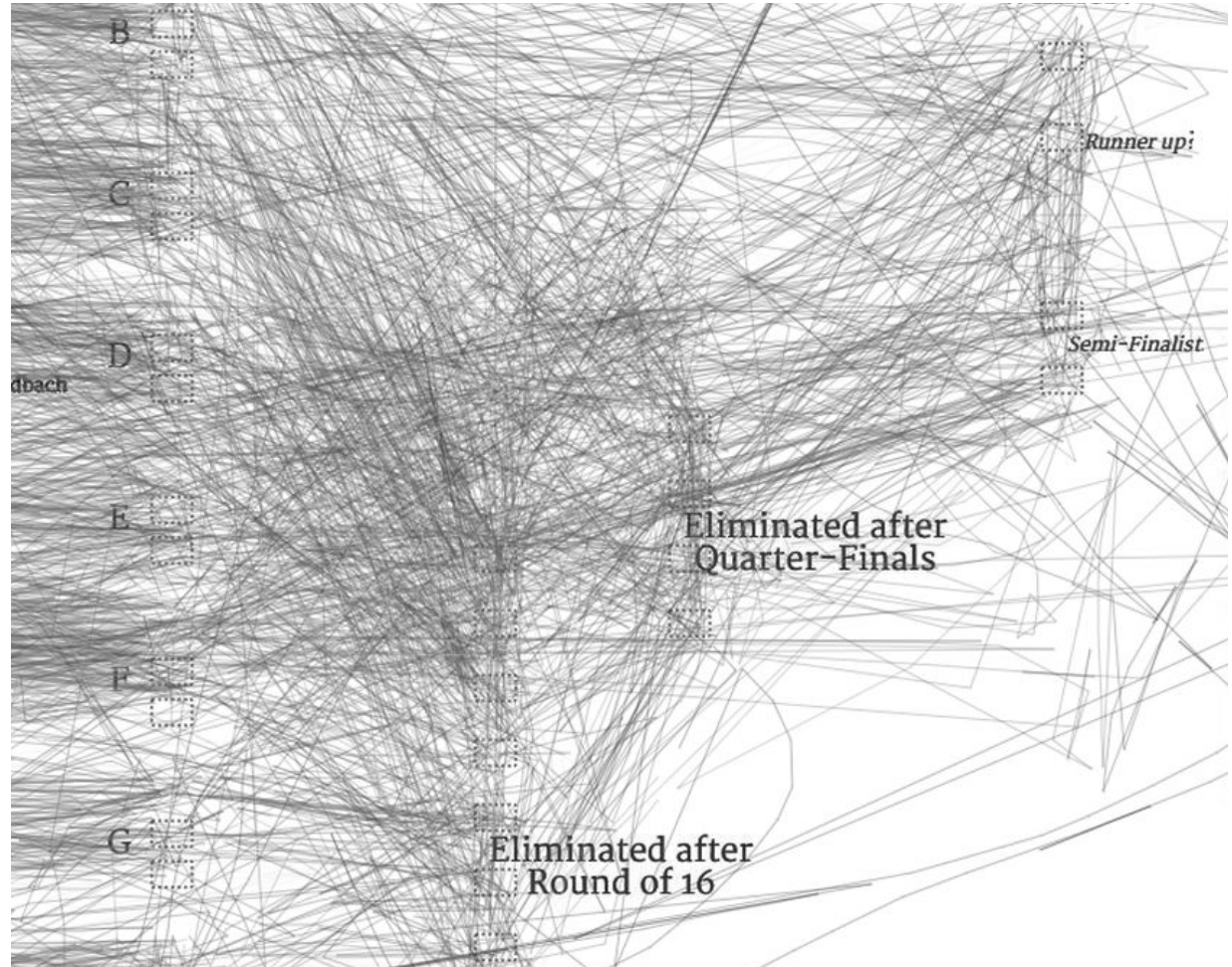
[UpSet, InfoVis 2014]



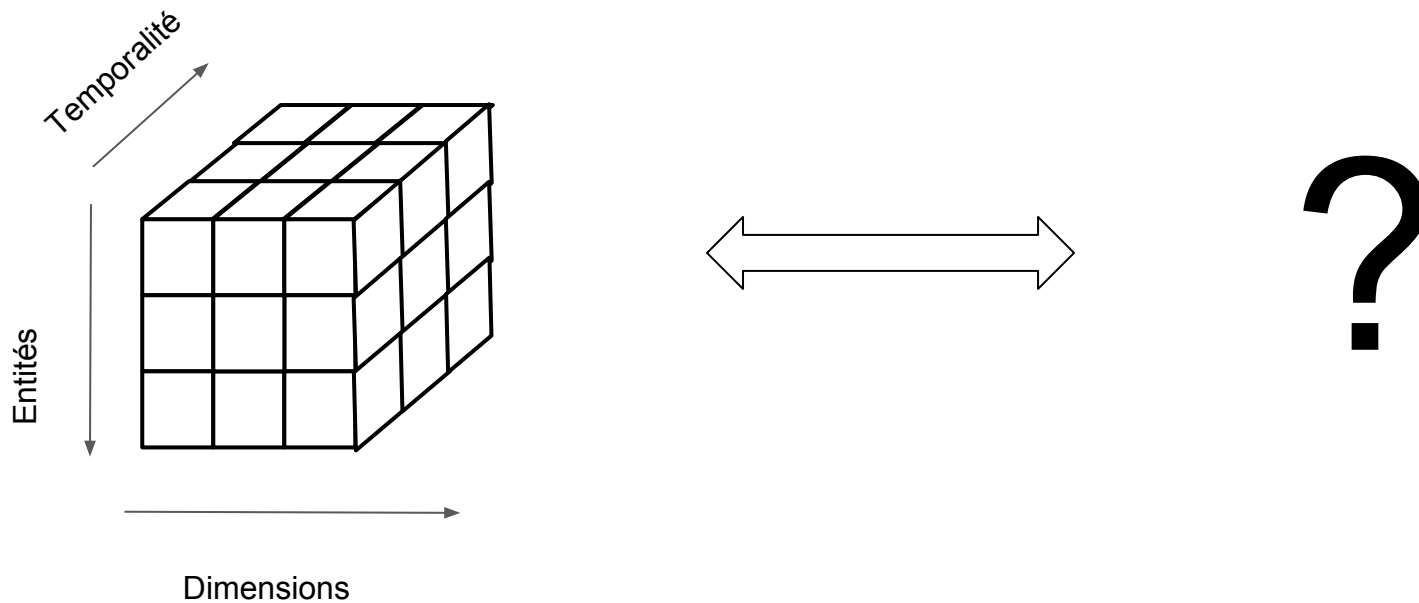
*point de vue sur
les données
[Tabard, 2014]*



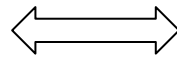
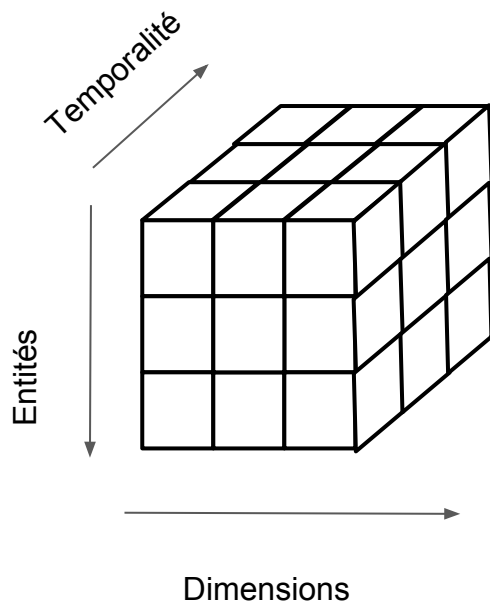
Problématique : comment explorer les données spatio-temporelles brutes sans a priori sur leur volume, distribution et tâches ?



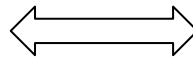
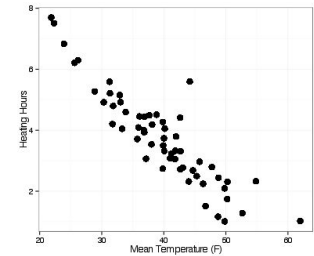
Approche : la création **d'interfaces visuelles interactives de données abstraites**, dont l'objectif est de permettre l'extraction de connaissance à partir de ces données.



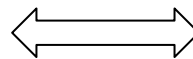
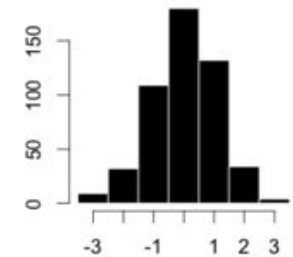
Des **visualisations simples** (scatterplot, histogrammes) permettent de répondre à de nombreuses questions en rendant accessible des statistiques descriptifs.



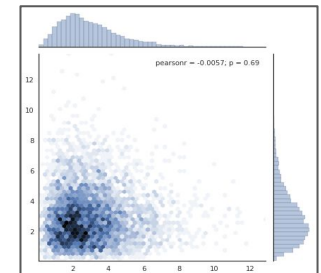
Détection de patterns : outliers, trends, etc.



Aperçu des caractéristiques globales des données : distribution, max/min, etc.

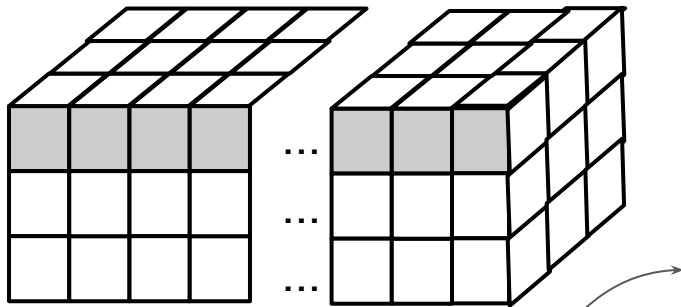


Exploration multidimensionnelle

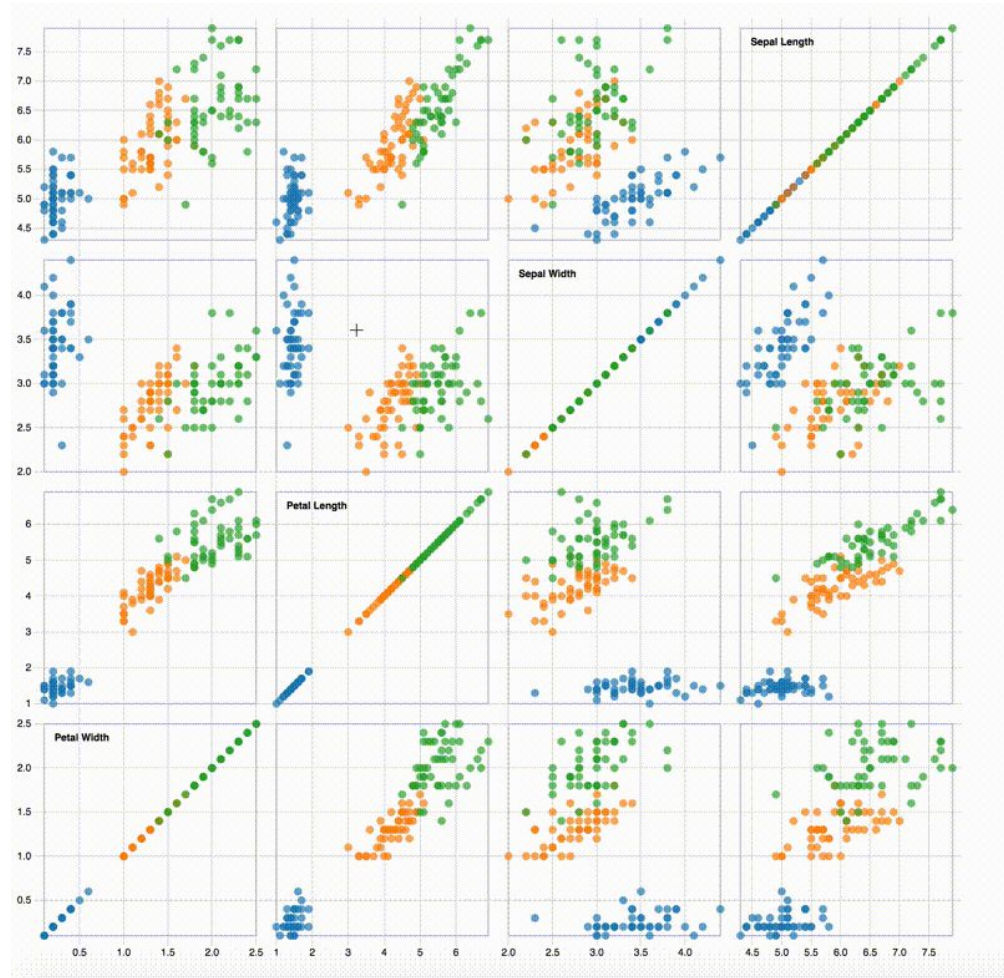


La combinaison de **permutations de données** et **interactions simples** permettent de comprendre des jeux de données multidimensionnels

Aperçu de toutes les dimensions
en même temps?

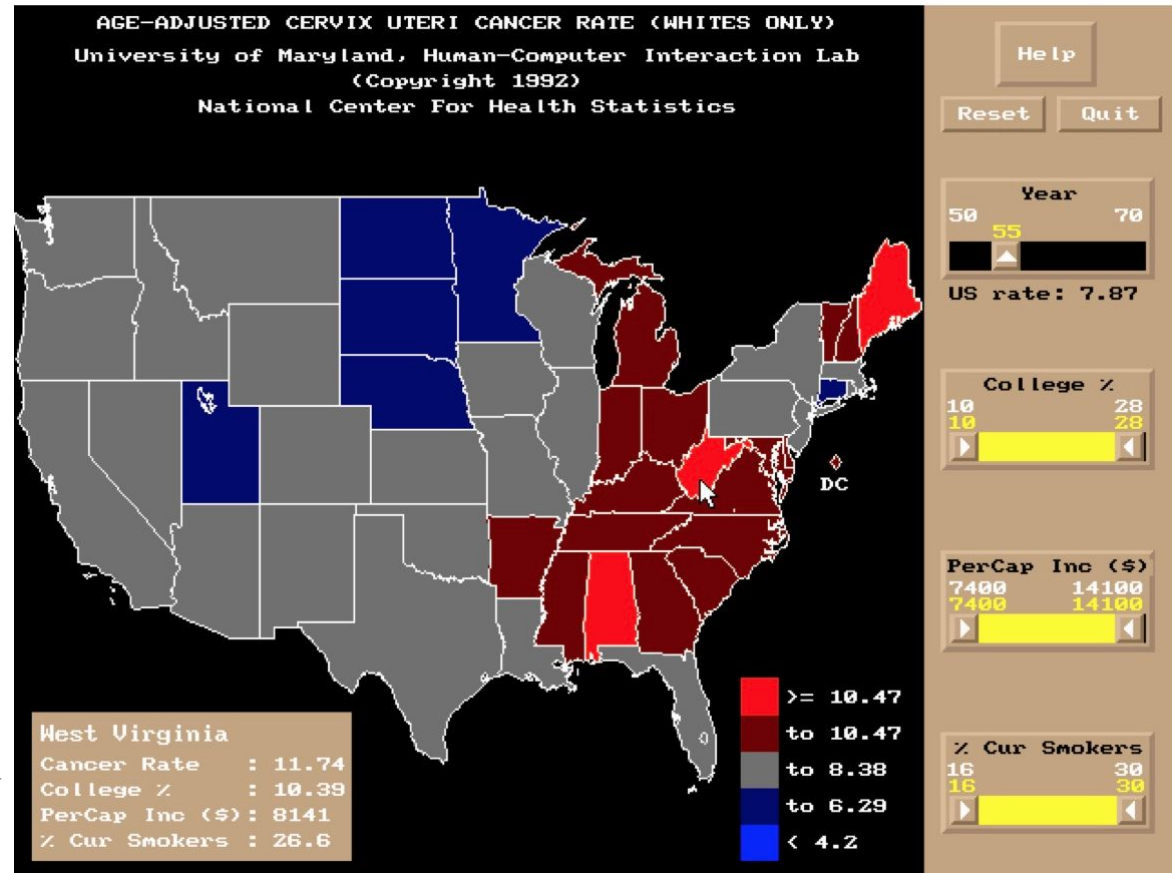
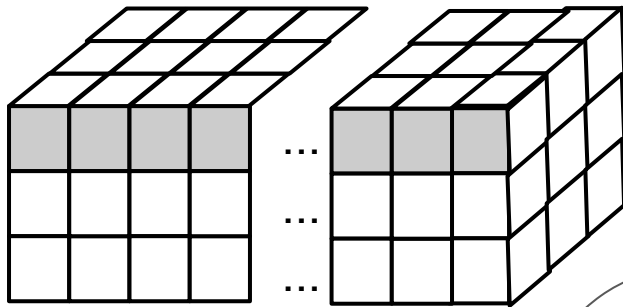


Permutation des dimensions
gènèrent un aperçu des graphiques



La combinaison de **permutations de données** et **interactions simples** permettent de comprendre des jeux de données multidimensionnels

Aperçu de toutes les dimensions
en même temps?

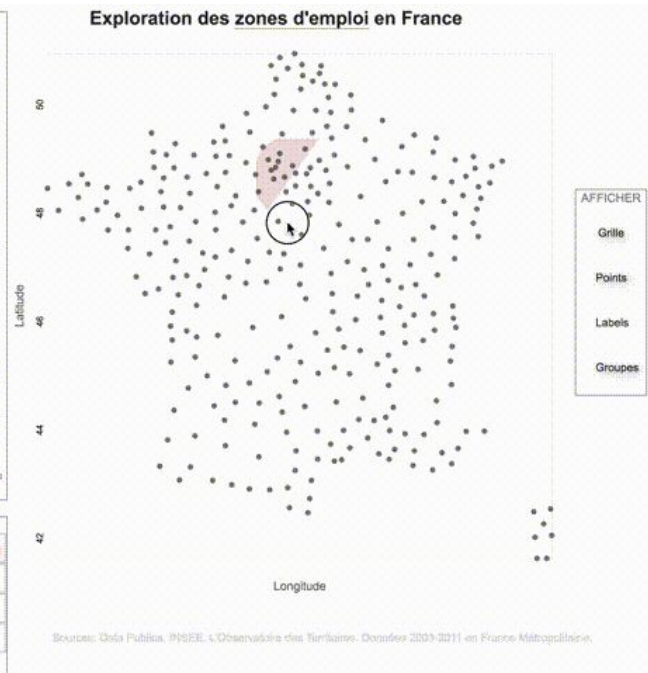
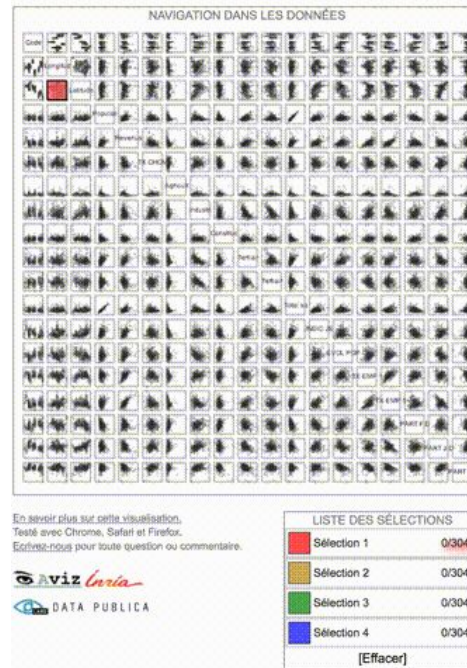
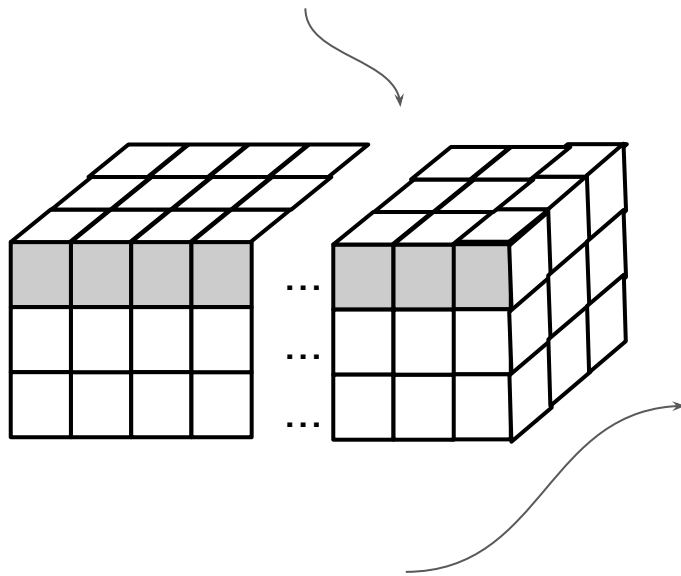


Permutation des dimensions
gènèrent un aperçu des graphiques

La combinaison de **permutations de données** et **interactions simples** permettent de comprendre des jeux de données multidimensionnels

The grand tour : exploration systématique de données multidimensionnelles

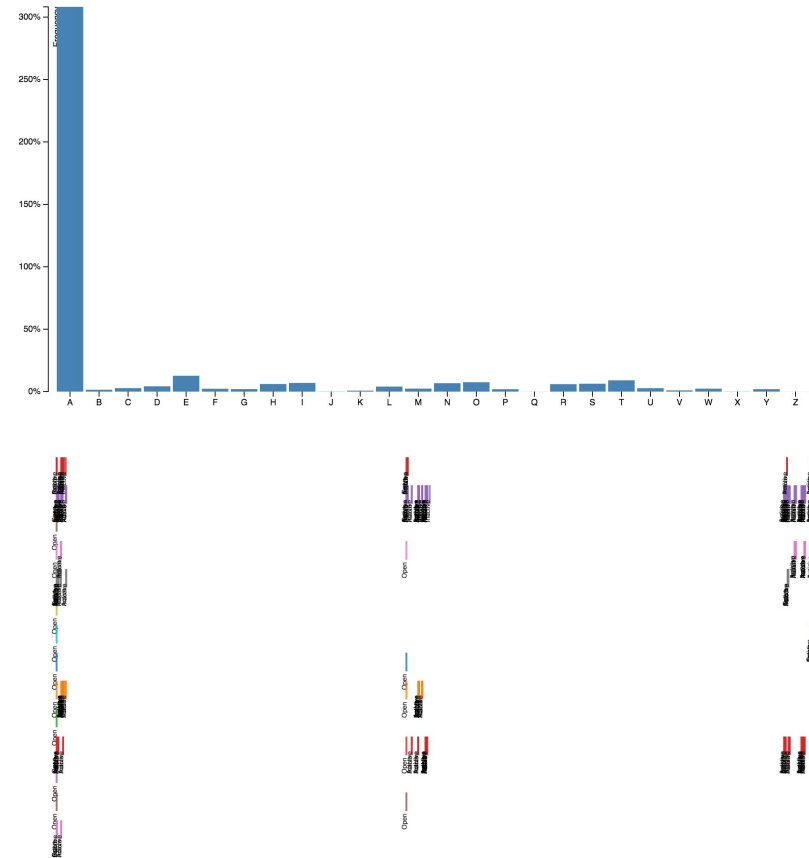
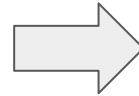
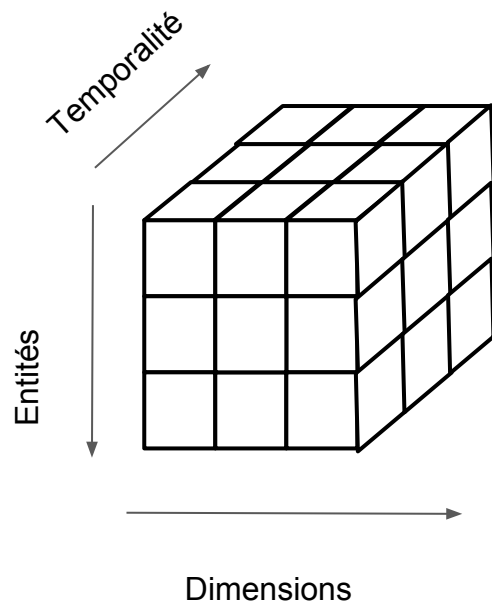
Aperçu de toutes les dimensions
en même temps?



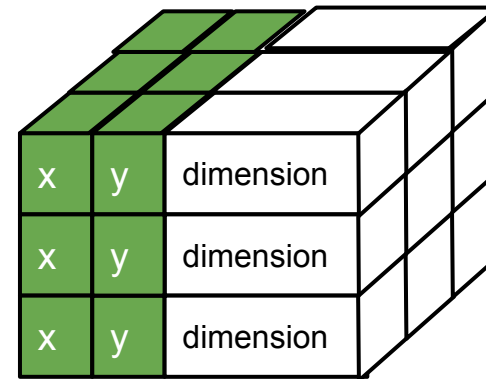
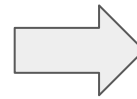
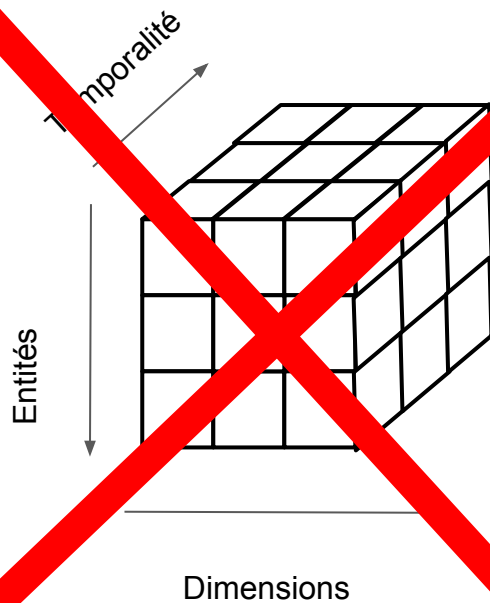
Permutation des dimensions
gènèrent un aperçu des graphiques

Asimov, Daniel. (1985). The grand tour: a tool for viewing multidimensional data. *SIAM journal on scientific and statistical computing*, 6(1), 128-143.

Verrou 1 : les données ne sont jamais distribuées de façon idéale et une approche générique est souvent inutilisable due aux **grandes variations** et à la **disparité** des données



Verrou 2 : dans le cas des données spatio-temporelles, les composantes X et Y sont fortes et nécessitent leur représentation dans le plan cartésien. Il ne reste donc plus d'autres encoding que la taille, couleur, etc.. pour les autres dimensions !



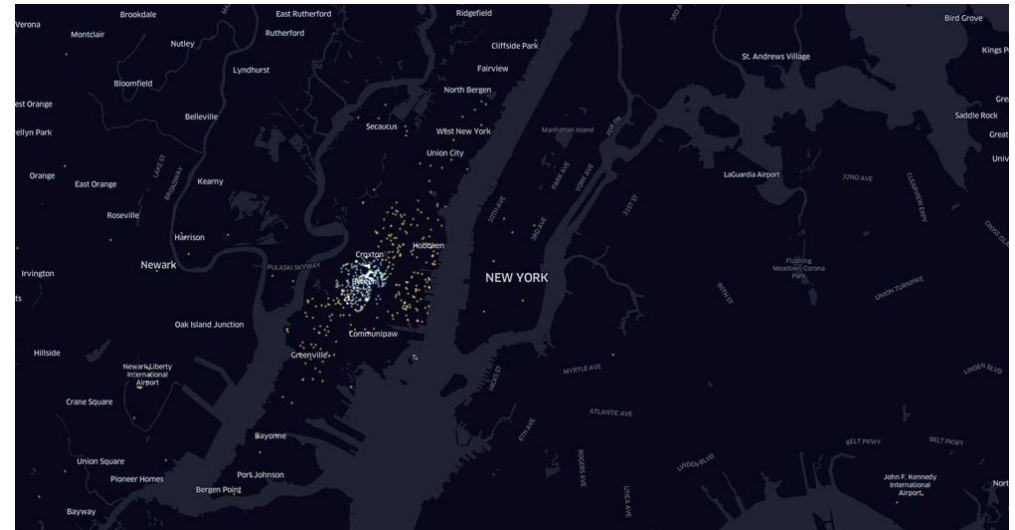
Verrou 3 : il reste très difficile de générer automatiquement des graphiques avec plus de 2 ou 3 variables et en utilisant la couleur et la taille.



Solution : augmenter la bande passante visuelle avec une très forte interactivité et densité d'information

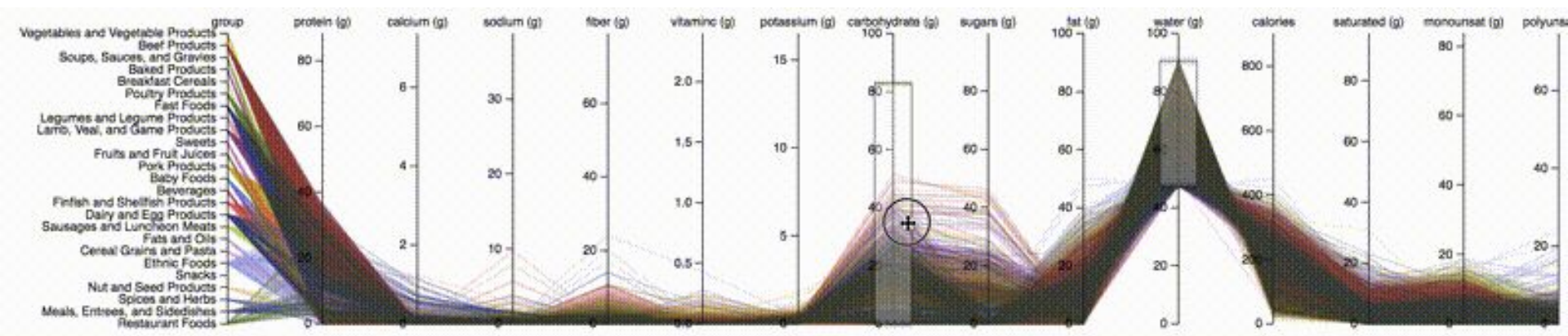
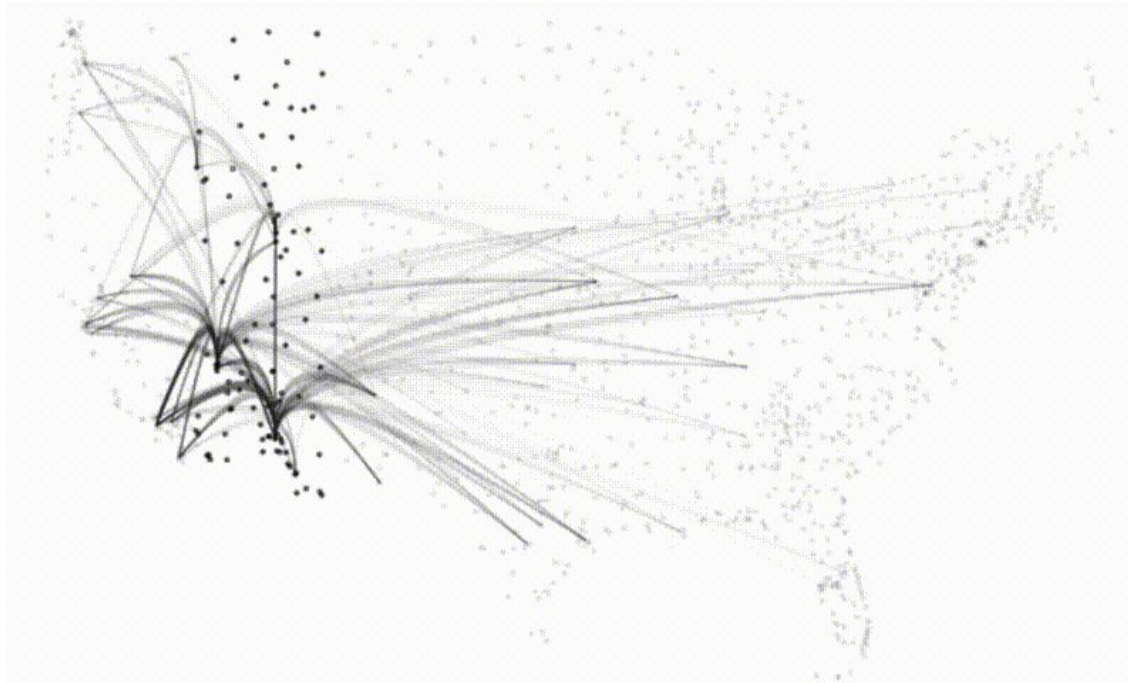


MapBox

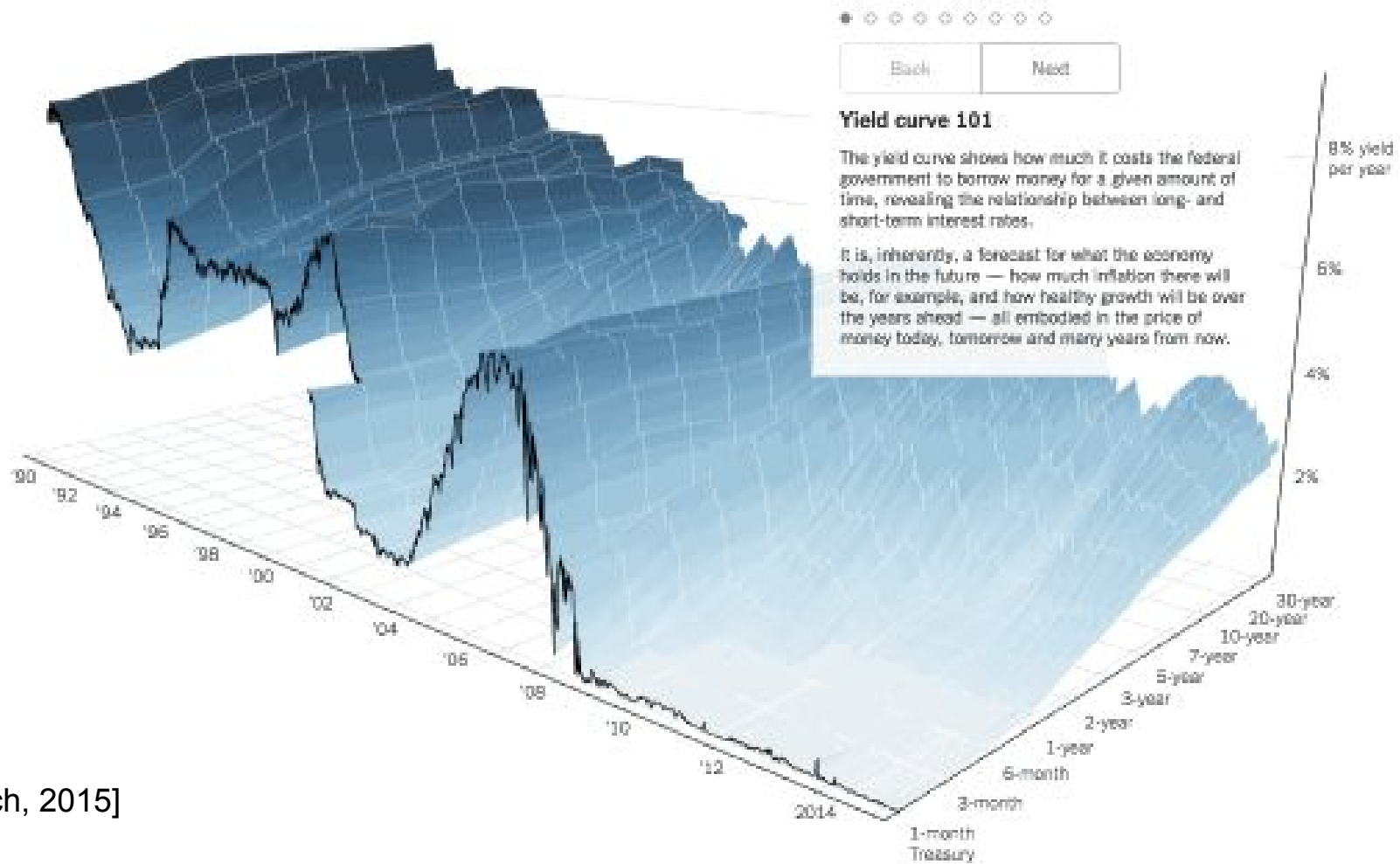


Uber

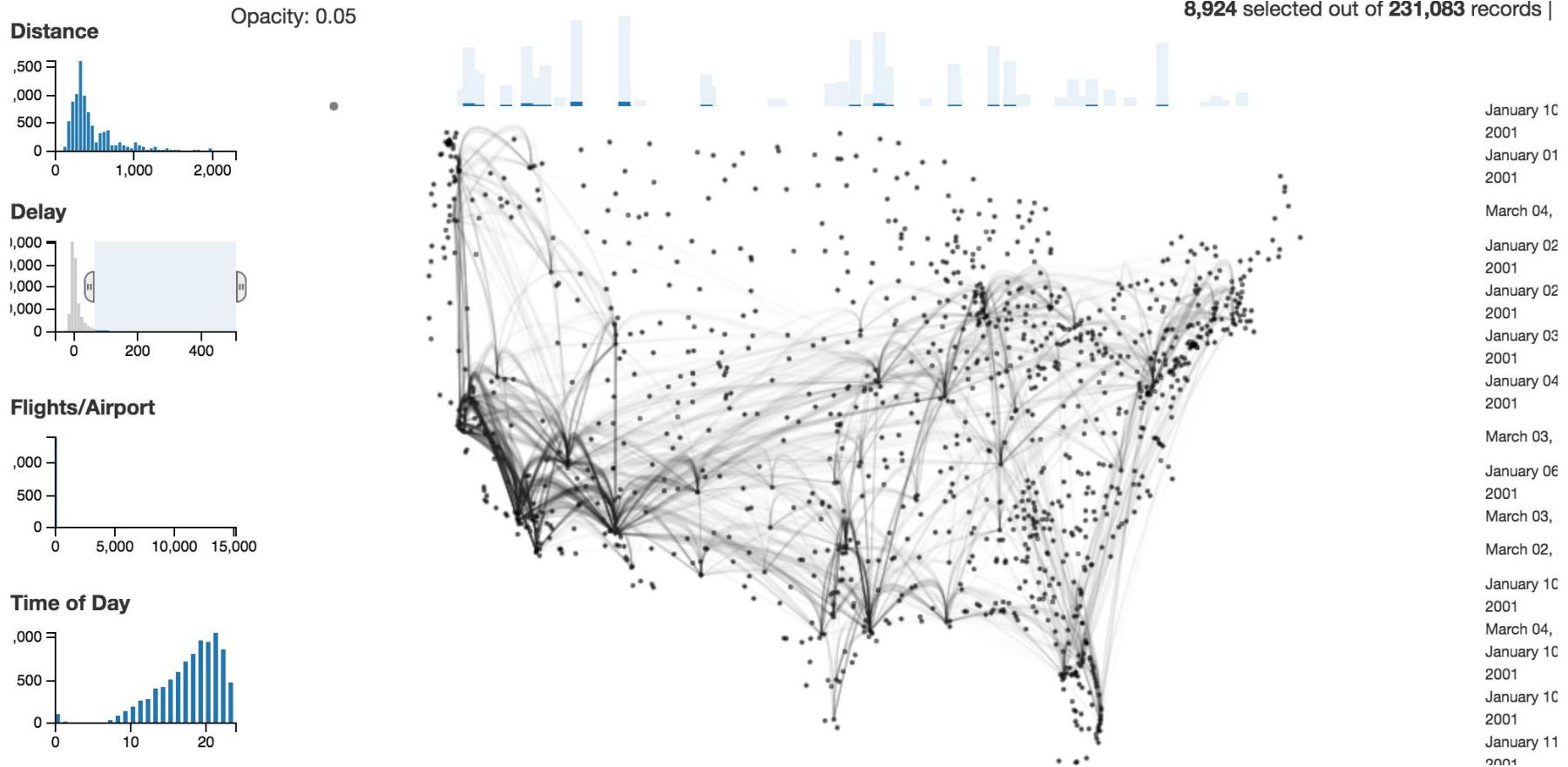
Solution : effectuer un rendu progressif des données afin de permettre de garantir la réactivité



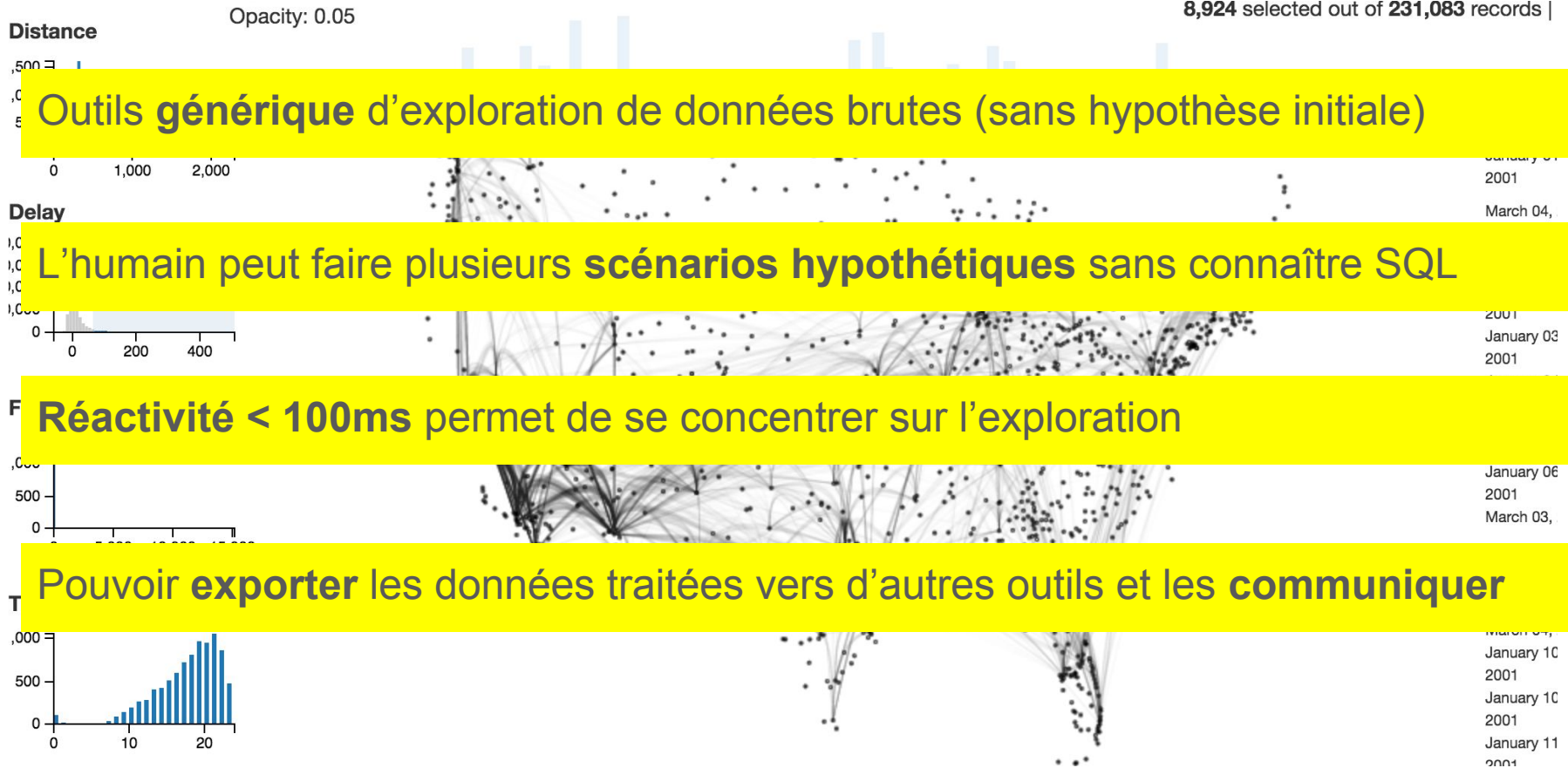
Solution : ajouter une 3ème dimension



Un exemple de **traces d'activité** sont les trajets en avion. Jeu de données issu du ASA Data Expo dataset. Échantillon de **231 083 trajets** du 1er janvier au 30 avril 2001 (4 mois).



Un exemple de **traces d'activité** sont les trajets en avion. Jeu de données issu du ASA Data Expo dataset. Échantillon de **231 083 trajets** du 1er janvier au 30 avril 2001 (4 mois).



Conclusion

Résumé des verrous scientifiques : concernant l'avancée de la science nécessitent de nouvelles approches (théorie, algorithmique, ..)

- 🕒 Cartes de densités (points, trajectoires) sans over plotting
- 🕒 Processus d'analyse visuelle et de guidage d'interaction
- 🕒 Visualisation des composantes fortement multidimensionnelles et coordination
- 🕒 Combinaison avec des traitements automatiques, méthodes de nettoyage et de reconstruction de données (trajectoires, ..)

Résumé des verrous techniques : concernent des obstacles étant donnée une technologie particulière.

- 🕒 Très grands volumes de données dans le navigateur
- 🕒 Support de nouveaux dispositifs d'entrée et de sortie (écrans, interaction post-wimp, gestes)

Merci pour votre attention

<http://romain.vuillemot.net/>

Journées VISU 2017 !

<http://visu2017.liris.cnrs.fr/>

7 et 8 juin 2017 à l'IFP

