

visualisation de données





Introduction

La visualisation de données désigne la représentation graphique d'informations et de données. À l'aide d'éléments visuels comme les graphiques et les cartes, une visualisation de données permet de voir et de comprendre des tendances ou des valeurs inhabituelles dans les données, de manière très accessible.

Dans le monde du Big Data, les outils et technologies de visualisation de données sont indispensables pour analyser d'énormes volumes d'informations et prendre des décisions en s'appuyant sur les données.

Avantages d'une visualisation de données efficace

- Nous sommes attirés par les couleurs et les schémas répétitifs.
- Nous pouvons facilement distinguer les couleurs, les formes, etc.
- Il existe une culture du visuel (cinéma, publicité, etc.)
- La visualisation permet d'intégrer une information rapidement.
- La visualisation découle d'un but précis. C'est une forme de récit de données.
- Nous avons tous déjà essayé de lire une feuille de calcul sans y repérer la moindre tendance.



Big data

- Donner du sens à des milliards de données par jour.
- Une visualisation pertinente donne du sens et de la cohérence aux données en éliminant les données non pertinentes dans les représentations.
- Il est nécessaire d'opter pour une visualisation efficace et pertinente.
- Le graphique le plus simple peut faire passer un message totalement inaperçu ou à l'inverse un message percutant.
- L'association analyse et récit associé sont la clé d'une visualisation réussie.

Visualisation

- La visualisation est nécessaire pour tous les types de données métier
- Chaque secteur d'activité gagne en bénéfices en améliorant l'interprétation de ses données.
- Dans les domaines scientifiques et techniques, dans l'administration, dans le sport, dans la finance, dans le marketing, la visualisation de données devient incontournable
- Il est nécessaire de savoir exploiter les données pour prendre des décisions et savoir utiliser des éléments visuels pour faire passer un message pertinent.

Type de visualisation

Il existe de nombreuses méthodes de visualisation permettant de présenter les données de manière efficace et attrayante.

Types de visualisations les plus courants :

- Diagrammes
- Tables
- Graphiques
- Cartes
- Infographies
- Tableaux de bord

Exemples de visualisations plus spécifiques :

- Graphique par aires
- Graphique à barres
- Graphiques à barres d'erreur
- Nuage de points
- Graphique à puces
- Cartographie
- Vue circulaire
- Carte de distribution de points
- Diagramme de Gantt
- Carte de chaleur
- Table surlignée
- Histogramme
- Matrice
- Réseau
- Graphique polaire
- Nuage de points (2D ou 3D)
- Graphique de flux
- Tableaux de texte
- Chronologie
- Arborescence
- Diagramme circulaire à plusieurs niveaux

Exemple de données

Description d'un jeux de données

World University Rankings

<https://www.kaggle.com/mylesoneill/world-university-rankings>

Le classement des universités est une pratique difficile, politique et controversée. Il existe des centaines de systèmes de classement universitaire nationaux et internationaux différents, dont beaucoup sont en désaccord les uns avec les autres. Cet ensemble de données contient trois classements mondiaux d'universités très différents.



Données de classement des universités

Le Times Higher Education World University Ranking est largement considéré comme l'une des mesures universitaires les plus influentes et les plus largement observées. Fondé au Royaume-Uni en 2010, il a été critiqué pour sa commercialisation et pour avoir porté atteinte aux institutions non anglophones.

Le classement académique des universités mondiales, également connu sous le nom de classement de Shanghai, est un classement tout aussi influent. Il a été fondé en Chine en 2003 et a été critiqué parce qu'il se concentrerait sur le pouvoir de recherche brut et décrédibiliserait les sciences humaines ainsi que la qualité de l'enseignement.

Le Center for World University Rankings, est une liste moins connue qui vient d'Arabie saoudite, elle a été fondée en 2012.

Exemple de questionnaire

1. Comment ces classements se comparent-ils les uns aux autres?
2. Les diverses critiques formulées contre ces classements sont-elles justes ou non?
3. Comment votre alma mater s'en sort-elle avec le monde?

Pour étendre davantage vos analyses, il y a deux ensembles de données supplémentaires.

Le premier d'entre eux est un ensemble de données sur le niveau d'instruction dans le monde. Il provient de la Banque mondiale de données et comprend des informations de l'Institut de statistique de l'UNESCO et de l'ensemble de données Barro-Lee. Comment le niveau d'éducation national est-il lié à la qualité des universités de chaque nation?

Le deuxième ensemble de données supplémentaires contient des informations sur les dépenses publiques et privées directes consacrées à l'éducation dans tous les pays. Ces données proviennent du National Center for Education Statistics. Il représente les dépenses en pourcentage du produit intérieur brut.

- Est-ce que dépenser plus pour l'éducation mène à de meilleurs classements universitaires internationaux?

Plusieurs approches pour traiter des données

1. On possède un jeu de données, on va essayer d'extraire des formes de données identifiables (clusters de données) et en déduire une interprétation : on analyse des données sans avoir de questions à priori. Ce sont des comportements, des analyses qui sont fréquentes dans les milieux scientifiques/techniques.
2. Une entreprise se pose des questions face à des comportements sociétaux, des tendances, etc. et l'analyse des données va permettre de répondre à ces questions. On extrait donc des données, uniquement celles qui sont pertinentes pour répondre à une problématique.