

Rapport du TP2 InfoVis

Réalisé par :
MAHMAHI Anis
Ilhem Bekkar

Encadré par :
Mme. FADLOUN Samiha

Table des Matière:

1. Introduction	3
2. Contexte d'étude	3
2.1. Objectif	3
2.2. Description de données	4
2.3. Codage visuel	4
2.4. Perception visuelle	5
3. Conclusion	5

Tableau des figures :

Figure 1 : Vue globale	2
Figure 2 : Défaut de couleur	3
Figure 3 : Deutéranomalie	4
Figure 4 : Protanomalie	4
Figure 5 : Protanopia	5
Figure 6 : Deutéranopie	5
Figure 7 : Overview	6
Figure 8 : Zoom and filter	6
Figure 9 : Détails en demande	7

1. Introduction :

La visualisation d'information est l'utilisation de représentations graphiques pour rendre les données plus compréhensibles et plus faciles à interpréter. Les visualisations peuvent inclure des graphiques, des diagrammes, des cartes, et des images qui permettent de mettre en évidence les tendances, les relations et les anomalies dans les données.

La visualisation d'information est utilisée dans de nombreux domaines, tels que la science des données, la recherche, l'analyse financière et la prise de décision d'affaires. Elle permet de transformer des données complexes en informations claires et utilisables pour aider les utilisateurs à comprendre rapidement les insights cachés dans les données et à prendre des décisions éclairées.

2. Contexte d'étude:

2.1. Objectif :

Au cours de ce TP, nous allons mettre en pratique le processus vu en InfoVis pour créer de visualisations d'information à partir de données brutes. Il est important de comprendre l'utilité de ce processus car il peut être difficile la plupart du temps d'analyser des données sous forme abstraite et en visualisant ces informations, nous facilitons la compréhension et l'analyse de ces données.

2.2. Description de données :

A. Format : tableau CSV

1. La première colonne représente le temps : des années de 2009 à 2020.
2. Trois colonnes représentent les langages : **Python**, **R**, **Matlab**.

B. Type : quantitative

1. Interval : qui représente le temps (les années).
2. Valeurs discrètes : qui représentent le nombre des questions posées sur Stack Overflow sur la data science.

C. La taille :

1. L'intervalle de temps comprend les années de 2009 à 2020, l'analyse des valeurs sur des périodes courtes peut être difficile, donc on peut penser à utiliser une représentation basée sur *context-focus*, cela nous aidera à bien visualiser le global, ainsi que les détails (**Overview, zoom-Filtre & details on demand**)
2. Les valeurs des 3 attributs varient entre 100 et 24000, par conséquent un axe de pas égale à 2000 sera largement suffisant pour contenir ces valeurs.

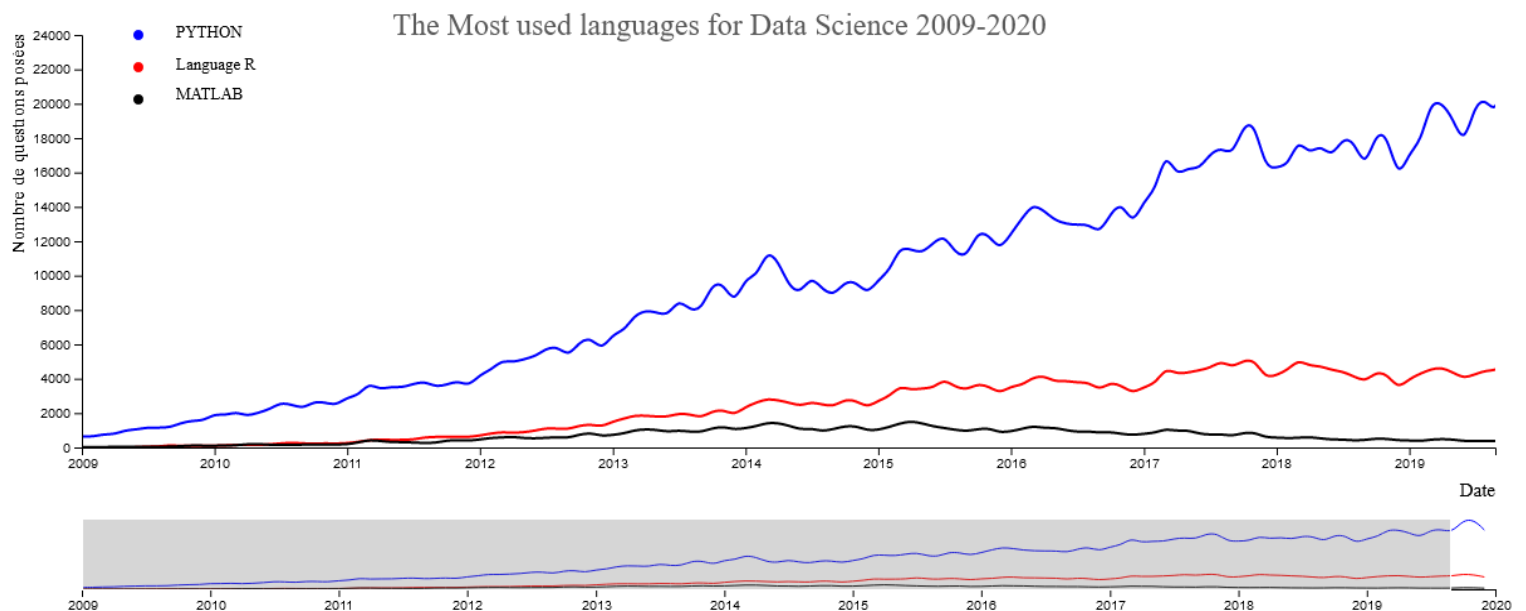
D. Indépendance (corrélation):

Il n'y a pas de relation entre les données des 3 langages, ie : pas de relation entre les valeurs correspondantes entre eux, par conséquent les graphes et les schémas entités-association ne seront pas des bonnes représentations visuelles pour notre cas.

L'aspect de la comparaison est présent, mais à cause de la présence de facteur temps le Pie Chart n'est plus une option de représentation.

*D'après les points abordés ci-dessus, un bon choix de représentation sera : un diagramme de ligne (Line Chart), où l'axe des X représente les années, et Y représente le nombre de questions dans chaque langage.
Cette courbe est visualisée par la représentation basée sur context-focus.*

Vue globale :



2.3. Codage visuel:

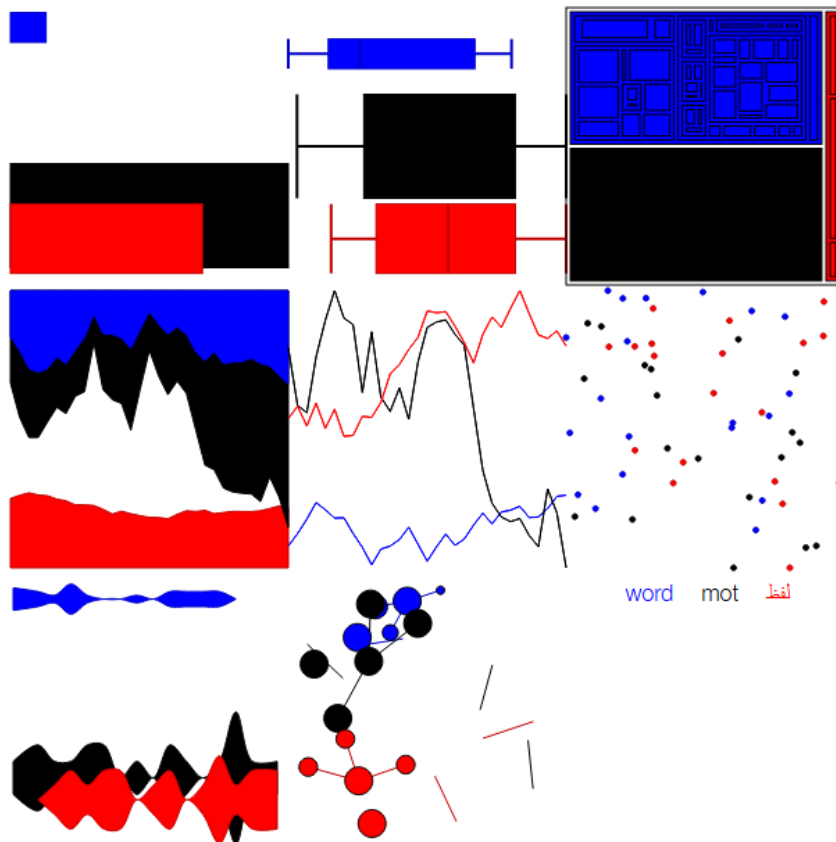
Comme on a déjà mentionné précédemment qu'il n'y a aucune similarité entre les 3 attributs étudiés, par contre ils servent beaucoup plus la comparaison, et pour cela on doit montrer **le contraste** dans le choix des couleurs utilisées d'une part, et d'autre part il faut tenir compte des personnes daltoniennes, donc pour nous aider à choisir les couleurs qui conviennent, le site web VIZ PALETTE fournit la possibilité de choisir un certain nombre de couleurs et de simuler à quoi ressemblent ces couleurs pour une personne daltonienne et voir est-ce qu'il peut y exister un conflit de couleurs.

Les couleurs choisies sont :

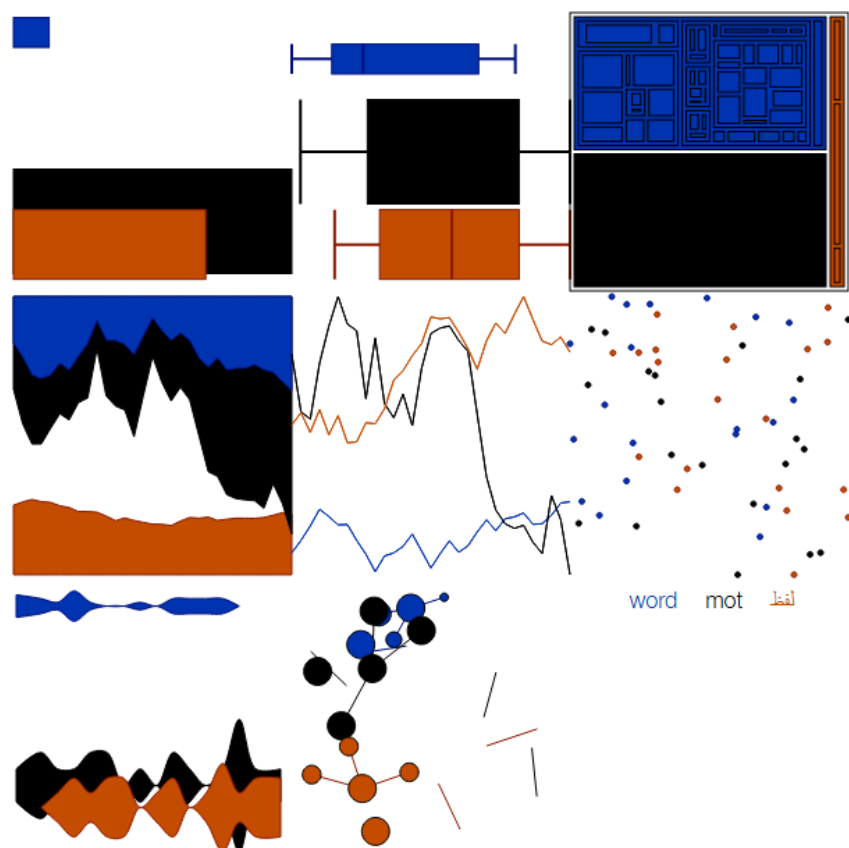
1. Le blanc `rgb(255, 255, 255)` pour le background.
2. **Le rouge** `rgb(255, 0, 0)` pour la 1ère variable.
3. Le noir `rgb(0, 0, 0)` pour la 2ème variable et les axes et les texts.
4. **Le bleu** `rgb(0, 0, 255)` pour la 3ème variable.

Voici ci-dessous les captures d'écran issues du site web VIZ PALETTE :

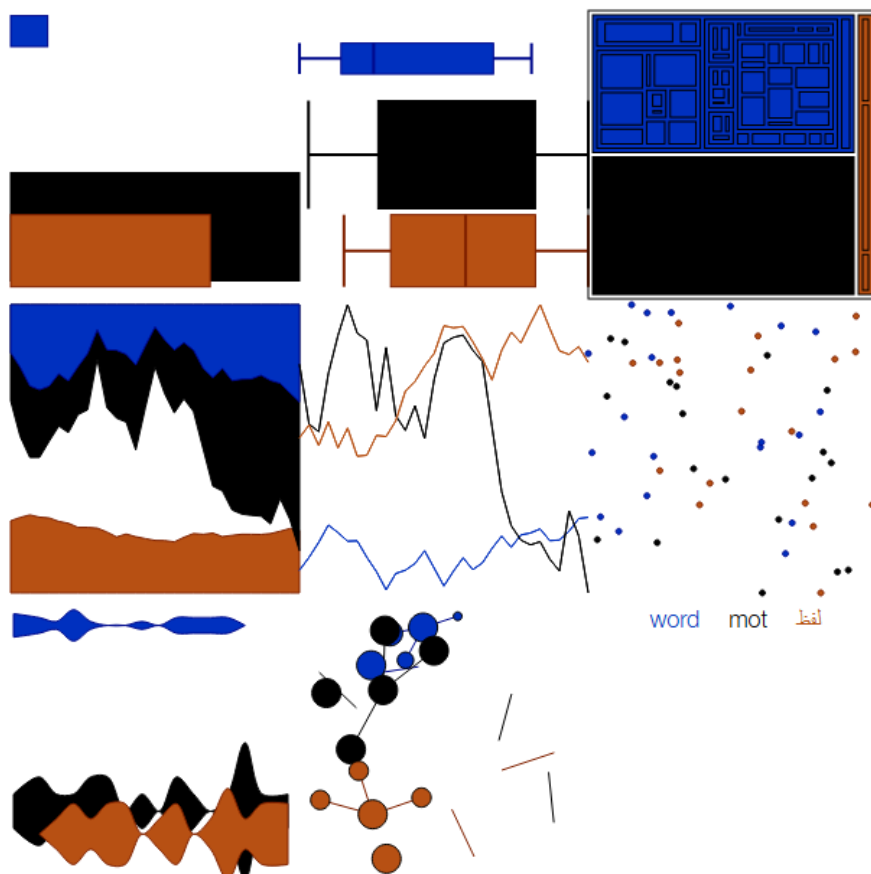
Aucun défaut de couleur - 96% :



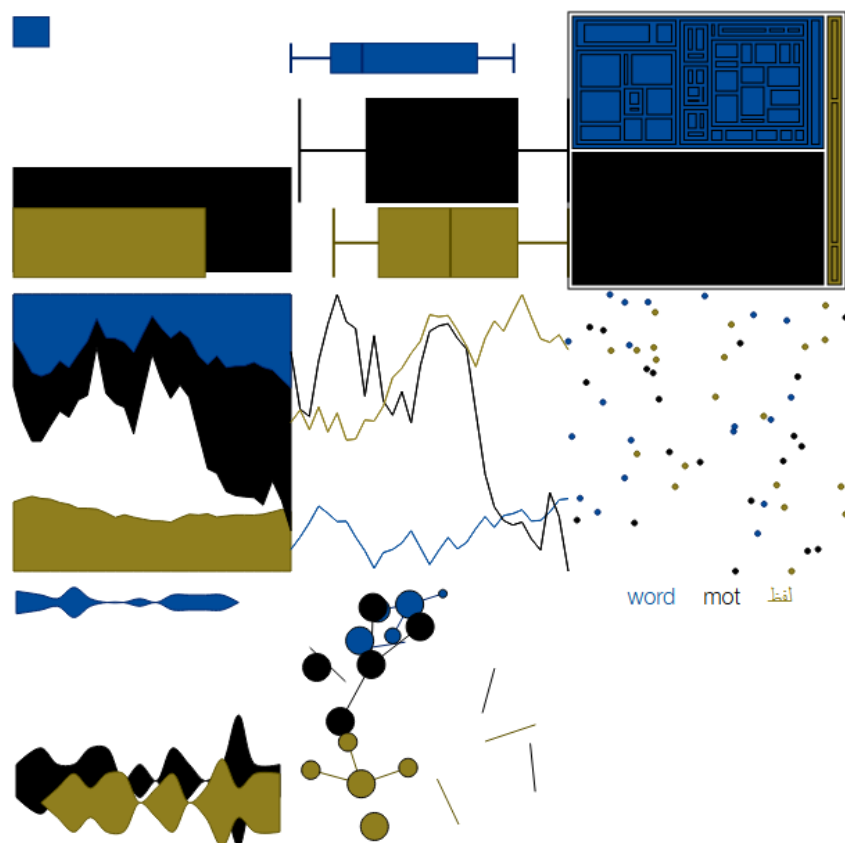
Deutéranomalie - 2,7 % :



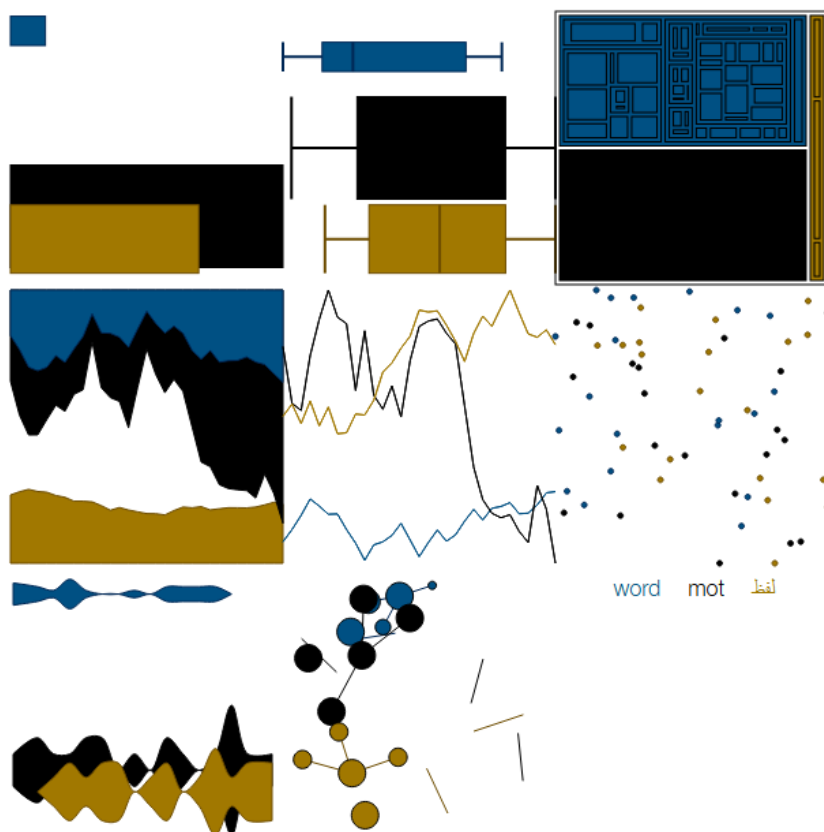
Protanomalie - 0,66 % :



Protanopia - 0,59 %



Deutéranopie - 0,56 %



On peut conclure depuis les captures ci-dessus que les couleurs choisies donnent une bonne perception visuelle pour tous les types de daltoniens.

Le texte est une variable visuelle qui sera consacrée pour la définition des axes, le titre général, légende, ainsi les valeurs représentées sur les axes (les années et le nombre des questions).

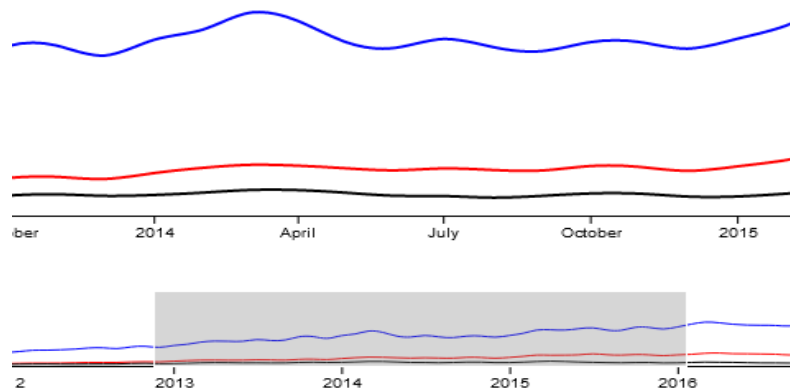
2.4. Perception visuelle :

Les principes de la perception visuelle sont utilisés pour concevoir des visualisations efficaces qui permettent aux utilisateurs de rapidement identifier les tendances et les relations importantes dans les données.

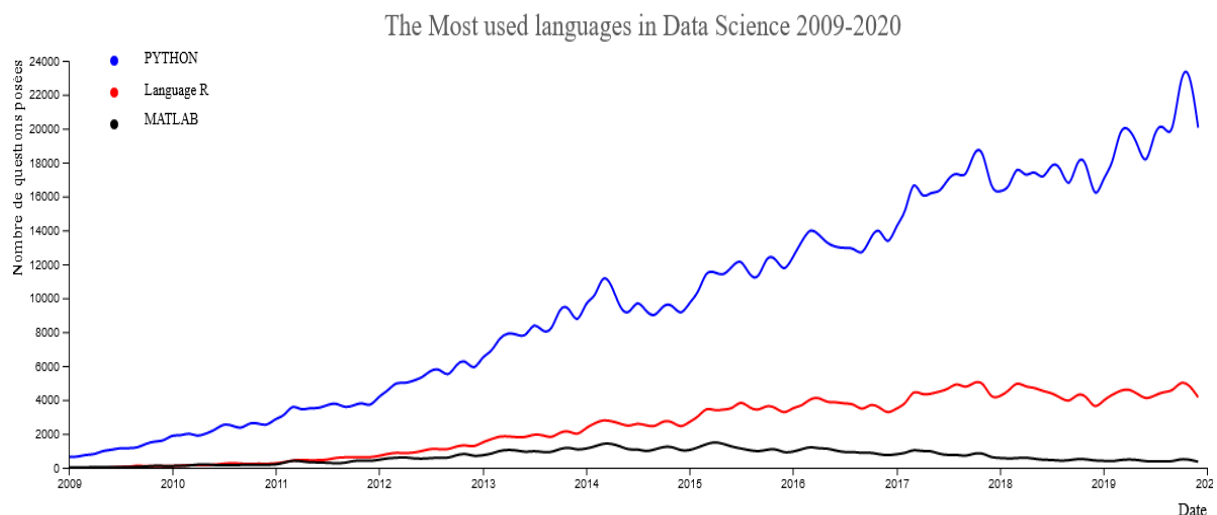
En plaçant le graphe au centre de la fenêtre, l'utilisateur peut le percevoir plus facilement car la plupart du temps, il se concentre d'abord sur le centre. Et cela justifie la position choisie pour le graphe.

2.5. Mantras de Ben shneiderman

1. Overview : c'est le contexte en bas de la visualisation.



2. Zoom and filter : c'est le focus en haut de la visualisation.



3. Détails en demande : c'est les cases à cocher en haut-gauche

Python: ☒ Language R: ☒ Matlab: ☒

3. Conclusion :

Les différentes méthodes de visualisation utilisées ont montré leur efficacité pour éclaircir les données et montrer également qu'elle est un outil puissant pour aider à la prise de décisions éclairées. Il est important de continuer à explorer de nouvelles méthodes et techniques pour améliorer l'analyse des données à l'avenir.