

2ème Année Cycle Supérieur (2CS SID)

## Rapport du TP 02

---

Représentation graphique avec D3.js

---

**Réalisé par:**

Adimi Alaa Dania

**Présenté à:**

Madame. Fadloune

2022/2023

# Table des matières

<b>Introduction</b>	<b>3</b>
<b>Partie I</b>	<b>4</b>
Diagramme de zone	4
Focus-plus-Context	5
Travail demandé	5
Implémentation	5
Données utilisées	5
Code	6
Démonstration	7
<b>Partie II</b>	<b>8</b>
Diagramme de lignes	8
Mantras de Ben Shneiderman	8
Implémentation	9
Données utilisées	9
Code	11
Démonstration	11
Choix des couleurs	11
Représentation obtenue	12
<b>Conclusion</b>	<b>14</b>
<b>Bibliographie</b>	<b>14</b>

# Introduction

La visualisation des données désigne le fait de représenter visuellement ses data pour pouvoir déceler et comprendre des informations, les données brutes étant difficilement interprétables et exploitables. Ce processus se fait par des outils analytiques spécifiques et se matérialise par des tableaux (type Excel), des graphiques, des cartes visuelles ou même des infographies regroupées dans des dashboards (tableaux de bord).

Data-Driven Documents ou simplement D3.js, est une bibliothèque de javascript qui a connu un grand développement grâce à la participation de nombreux développeurs et testeurs. Cette bibliothèque nous permet de créer des représentations graphiques interactives sur le web.

Dans ce rapport je vais présenter les deux exercices faits lors du TP02 intitulé “Représentation graphique avec D3.js”. Ce rapport est donc divisé en deux parties essentielles, chacune traitant un exercice.

Tout les liens des datasets utilisés ainsi que les sites utilisés sont disponibles dans la section bibliographie.

# Partie I

Dans cette partie, je vais présenter l'implémentation de la première figure qui représente un **diagramme de zone** avec **focus-plus-context**. Il s'agit de suivre les étapes fournies, et ajouter des données réelles afin d'obtenir une représentation fonctionnelle. Mais avant d'aller plus loin, qu'est ce qu'un diagramme de zone? et qu'est ce que ça veut dire focus-plus-context?

## Diagramme de zone

Les diagrammes de zone (appelés aussi diagrammes de surface) sont principalement utilisés pour représenter des données qui suivent une relation de série chronologique. L'échelle d'un remplissage de couleur indique l'ampleur de la valeur correspondante. Un graphique de surface trace des informations sur l'axe des x et des valeurs de données sur l'axe des y en reliant les points de données par des segments de ligne solides. Lorsque l'ampleur de la tendance doit être communiquée, plutôt que les valeurs individuelles, les diagrammes de surface sont les plus efficaces.

## Focus-plus-Context

“Focus-plus-context” est un des principe de la visualisation des informations. En termes simples, ce principe repose sur le fait que l'utilisateur doit avoir à la fois une vue d'ensemble (contexte) et d'informations détaillées (focus).

# Travail demandé

Le but du premier exercice est d'apprendre comment on peut créer un diagramme de zone avec context-plus-focus. Les étapes ont été déjà bien expliquées, et donc, le travail demandé est de suivre les étapes et d'ajouter des données réelles afin d'avoir une représentation fonctionnelle.

## Implémentation

### Données utilisées

Le dataset utilisé dans cette partie représente l'historique des prix de stock (5 dernières années) pour toutes les sociétés figurant actuellement dans l'indice S&P 500. Le dataset se trouve sur Kaggle sous le nom "**S&P 500 stock**". Il contient les informations suivants:

- **Date** - format yy-mm-dd
- **Open** - prix de l'action à l'ouverture du marché (il s'agit de données NYSE, donc tout en USD)
- **High** - Prix le plus élevé atteint dans la journée
- **Low Close** - Prix le plus bas atteint dans la journée
- **Volume** - Nombre d'actions échangées
- **Name** - le nom du symbole boursier

J'ai effectué un prétraitement sur ces données afin de pouvoir les exploiter pour notre représentation. La phase de prétraitement consisté à:

- Traité les valeurs manquantes et les valeurs abberantes.

- Transformé le format de l'attribut "Date" de yy-mm-dd à yymmdd afin de pouvoir l'utiliser dans notre représentation.
- Changé le type des données au type approprié.

Le prétraitement a été fait en utilisant les deux fameuses bibliothèques de Python: **Pandas** et **Numpy**. Pour notre représentation, j'ai utilisé les deux attributs suivants: "Date" et "Open". Vous trouveriez dans le fichier "**stocks.csv**" les données utilisées dans la représentation.

## Code

Comme déjà mentionné, le code a été déjà fourni. Donc j'avais juste à ajouter les données nettoyées avec des petites modifications. Vous pouvez consulter le fichier "**index.html**" qui se trouve dans le dossier "**Figure 01**" pour plus de détails.

## Démonstration

La représentation que j'ai obtenu est fonctionnelle et similaire à celle qui se trouvent dans l'énoncé du TP. Voici quelques captures illustrant le résultat obtenu.

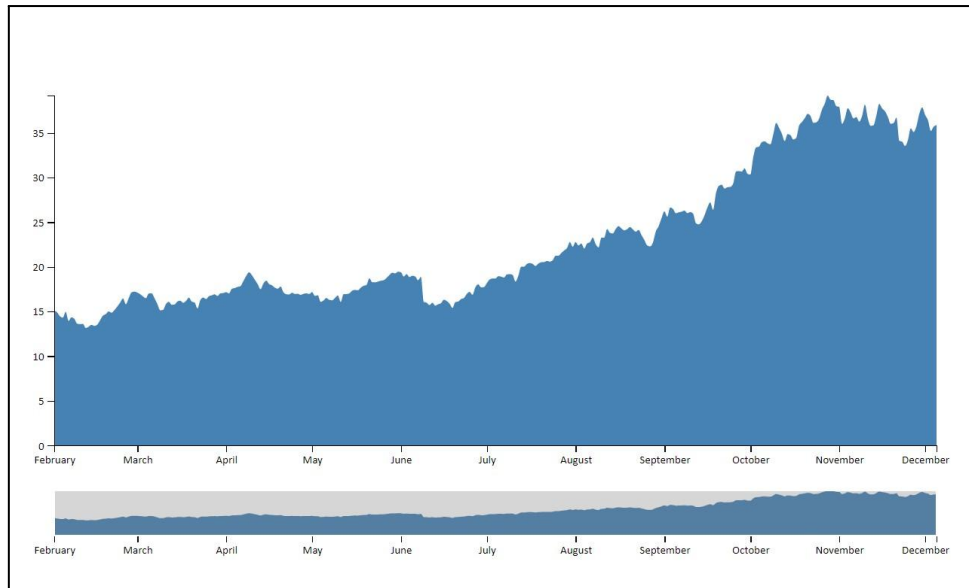


Figure 01 - Diagramme de zone avant interaction

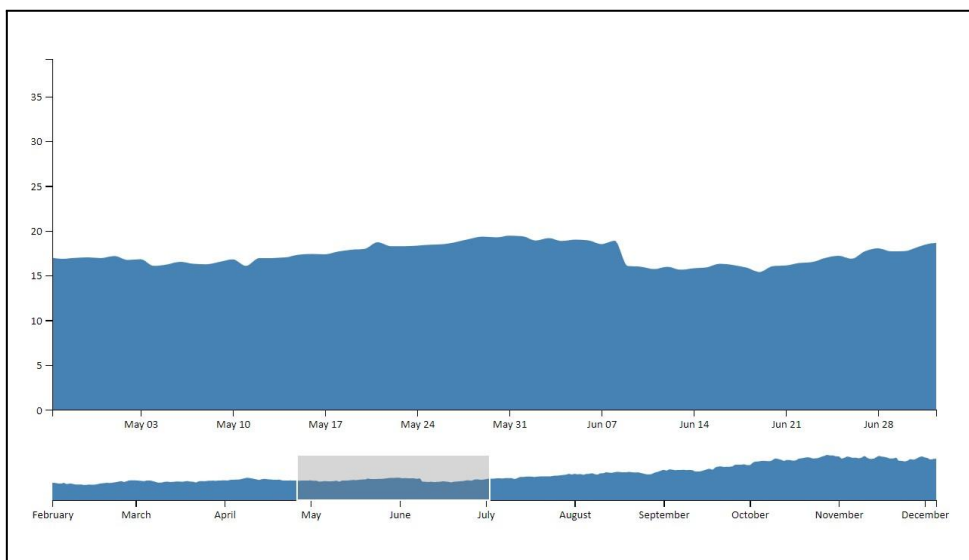


Figure 02 - Diagramme de zone après une interaction de l'utilisateur

## Partie II

Dans le deuxième exercice du TP02, on nous a été demandé de créer un diagramme de lignes qui utilise le principe de focus-plus-context, et tout en respectant les mantras de Ben Shneiderman. On nous a été

demandé aussi de justifier le choix du dataset ainsi que le choix des couleurs.

Le but de cet exercice est d'utiliser tous les concepts déjà vu en cours dans la pratique.

## Diagramme de lignes

Le diagramme de lignes est un des représentations visuelles les plus utilisées. Généralement, ce diagramme fait référence à une situation qui évolue dans le temps (années, mois, jours, etc.).

## Mantras de Ben Shneiderman

Ben Shneiderman est un scientifique américain ayant une forte expertise dans le domaine de l'interaction homme machine, bon nombre de ses travaux sont des fondamentaux de l'interaction homme machine d'aujourd'hui, on peut par exemple citer la création des "Treemap". Shneiderman a proposé une collection de ces principes dérivés de son expérience, qui, après avoir été perfectionnés, développés et expliqués, sont applicables aux systèmes interactifs. Les principes les plus connus sont les suivants:

- **“Overview”** - Il est recommandé de visualiser l'ensemble de données donné de manière simple sans entrer dans trop de détails. Les utilisateurs doivent avoir d'abord une vue globale sur les données.
- **“Zoom and Filter”** - Il faut donner aux utilisateurs la possibilité d'interagir avec la représentation, en zoomant ou en filtrant les données utilisées.



- **“Details on Demand”** - Il faut donner aux utilisateurs la possibilité d’avoir plus de détails sur les données. Par exemple, l’utilisation des tooltips.

## Implémentation

### Données utilisées

Le dataset utilisé dans cette partie représente les tendances de musiques sur Spotify (année 2017). Ce dataset vient de Kaggle et il contient les informations suivants:

- **Date** - format yy-mm-dd
- **Shape of You** - Nombre de fois que quelqu'un a entendu “Shape of You” pendant un jour.
- **Despacito** - Nombre de fois que quelqu'un a entendu “Despacito” pendant un jour.
- **Something Just Like you** - Nombre de fois que quelqu'un a entendu “Something Just Like you” pendant un jour.
- **Humble** - Nombre de fois que quelqu'un a entendu “Humble” pendant un jour.
- **Unforgettable** - Nombre de fois que quelqu'un a entendu “Unforgettable” pendant un jour.

J’ai effectué un prétraitement sur ces données afin de pouvoir les exploiter pour notre représentation. Cette étape est très importante, il n’est pas possible d’utiliser des données brutes directement, la qualité des

données doit être vérifiée d'abords, par exemple on peut avoir des données erronées. Le prétraitement consisté à:

- Traité les valeurs manquantes et les valeurs abberantes en utilisant plusieurs techniques.
- Transformé le format de l'attribut "Date" de yy-mm-dd à yymmdd afin de pouvoir l'utiliser dans notre représentation.
- Changé le type des données au type approprié.
- Changé le nom des colonnes afin d'avoir une structure uniforme.

Le prétraitement a été fait en utilisant les deux fameuses bibliothèques de Python: **Pandas** et **Numpy**. Le dataset prétraité est disponible dans le dossier partagé sous le nom "**Spotify.csv**".

J'ai choisi de travailler avec ce dataset car il contient des données réelles, et ces données là peuvent être très bien représentées par un diagramme de lignes, car il s'agit de l'évolution des tendances de musique dans le temps. À travers cette représentation, on peut facilement comparer entre l'évolution des cinq chansons dans le temps.

## Code

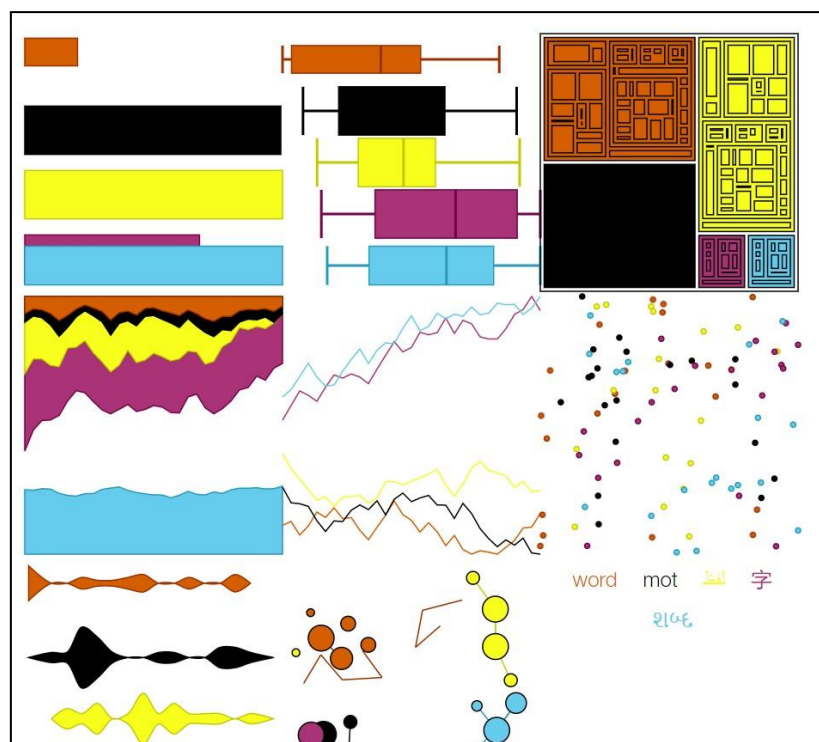
Pour la programmation, j'ai utilisé les concepts déjà acquises dans le premier exercice (comme la création du context, la création des axes, etc), ainsi que les concepts vu dans le TP01. Comme demandé, j'ai utilisé un diagramme de lignes tel que l'axe horizontal représente le temps, et l'axe vertical présentant le nombre de fois que quelqu'un a entendu une chanson.

Et afin de différencier les chansons, j'ai utilisé une couleur unique pour chaque chanson.

## Démonstration

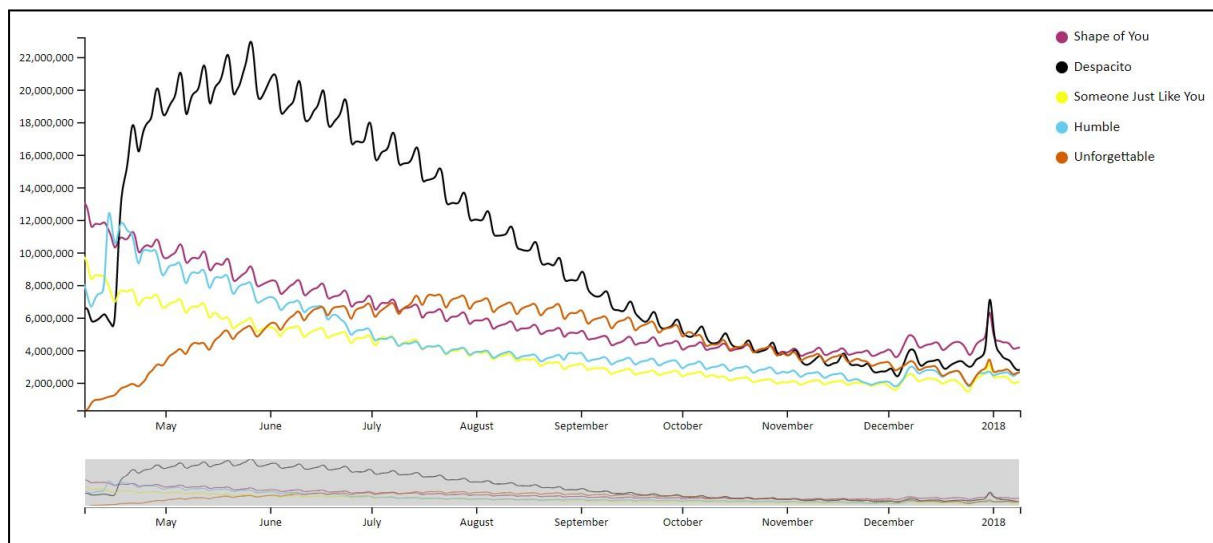
### Choix des couleurs

Les couleurs utiliser pour représenter les 5 états doivent être séparables, il faut aussi tenir compte des personnes daltoniennes tel que deux couleurs qui semblent distinctes pour une personne ayant une vision normale des couleurs peuvent sembler identiques pour une personne daltonienne. Pour choisir les couleurs qui conviennent, j'ai utilisé le site web **VIZ PALETTE** qui fournie 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. Voici la palette que j'ai utilisé:

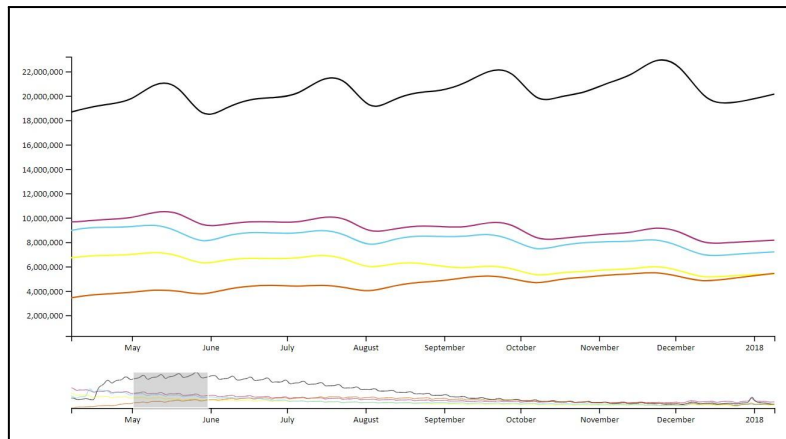


Les couleurs choisies peuvent être facilement différenciées par les personnes ayant deutéranomalie et autres. Vous pouvez consulter la palette via le lien suivant: [PALETTE](#).

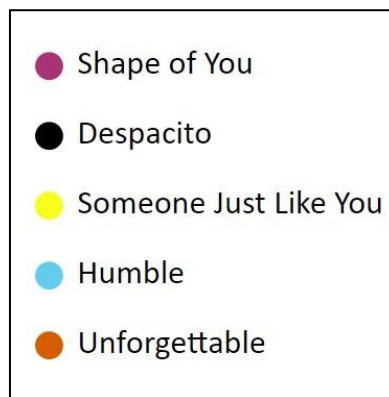
## Représentation obtenue



- La représentation contient un **context**. Ce context contient l'ensemble complet de donnée avec un seul axe qui représente le temps (l'axe des x). Le context permet l'utilisateur d'avoir un vue globale sur les données.
- La représentation contient un **focus**. Ce focus qui contient la partie sélectionnée de données (bruch) en haut de contexte avec deux axes.
- L'utilisateur peut voir l'évolution de tendances de musique sur une certaine période. Par exemple, entre Mai 2017 et Juin 2017. Cela peut être fait avec deux façons différentes:
  - Réduire la taille du bruch et le faire déplacer.
  - Zoomer directement sur le context.



- La représentation dispose d'une **légende** qui aide les utilisateur à différencier entre les chansons.



- La représentation donne la possibilité à l'utilisateur d'avoir une vue globale sur l'ensemble de données "**Overview**". Elle lui permet aussi de zoomer et voir plus de détails sur les données. Cette représentation peut être améliorée en ajouter des tooltips et des boutons radio pour filtrer les données. Ce qui va satisfaire tous les mantras de Ben Shneiderman.
  - "**Tooltip**" afin de permettre l'utilisateur de voir la valeur exacte d'un point sur la ligne.

- “**Boutons radio**” afin de permettre l'utilisateur de filtrer les données (par exemple, visualisation de l'évolution de deux chansons seulement).

## Conclusion

À travers ce travail pratique j'ai pu comprendre la visualisation de l'information pour partager de précieuses informations de manière conviviale. Je me suis également familiarisé avec D3.js qui est très utilisé pour la visualisation de données.

Les représentations fournies peuvent être améliorées afin de satisfaire tous les principes de Ben Shneiderman. Ce travail m'a motivé d'apprendre plus sur la visualisation de données en général, et la représentation graphique avec D3.js en particulier.

# Bibliographie

- Dataset des Stocks:  
<https://www.kaggle.com/datasets/camnugent/sandp500?resource=download>
- Dataset de Spotify:  
<https://www.kaggle.com/code/alexisbcook/line-charts/data?select=spotify.csv>
- Mantras de Ben Shneiderman:  
<https://jtr13.github.io/cc21/ben-shneidermans-visualization-mantra.html>
- VIZ Palette: <https://projects.susielu.com/viz-palette>