

TP4 (version Python)

INF8808: Visualisation de données

Département de génie informatique et génie logiciel



POLYTECHNIQUE MONTREAL

Auteure: Olivia Gélinas

Chargé de lab: Hellen Vasques

Objectifs

L'objectif de ce travail pratique est de créer un graphique à bulles (*bubble chart*) interactif à l'aide de données ouvertes en format JSON.

Avant de commencer, nous vous recommandons d'avoir effectué les lectures suivantes et d'avoir effectué les exercices suivants:

- Bubble Charts

<https://plotly.com/python/bubble-charts/>

- Intro to Animations

<https://plotly.com/python/animations/>

Lectures :

- Advanced Callbacks

<https://dash.plotly.com/advanced-callbacks>

- Discrete colors

<https://plotly.com/python/discrete-color/>

Exercices : TP4 exercices : 1, 2, 3, 4

Introduction

Un graphique à bulles est un type de graphique utilisé pour visualiser des données possédant de nombreux paramètres et où chaque point est représenté par un cercle. C'est une combinaison d'un nuage de points et un graphique à aires proportionnelles. Avec ce type de graphique, nous pouvons encoder jusqu'à quatre dimensions: la position en x, la position en y, la couleur et l'aire des cercles.

Dans ce travail pratique, vous allez créer un graphique à bulles interactif en utilisant des données sur des émissions de CO2 par habitant et sur revenus par habitant fournis par le portail de données ouvertes de la Banque mondiale [1]. Vous aurez accès aux données des années 2000 et 2015.

Avant de compléter le code du graphique à bulles, vous allez également implémenter une version simplifiée de celui-ci à l'aide d'un cahier Jupyter. La source du cahier, que vous allez devoir compléter, vous a été fournie. Ce cahier représente une étape initiale importante dans la création de la plupart des visualisations de données, c'est-à-dire l'exploration des données et des possibilités de design.

Description

Dans ce travail pratique, vous devrez compléter le code Python en utilisant Plotly et Dash pour afficher un graphique à bulles représentant la relation entre le PIB et les émissions de CO2. Dans un premier temps, vous implémenterez une application interactive simple à l'aide de Dash et Plotly dans un cahier Jupyter. Ensuite, vous réutiliserez et complèterez le code de votre cahier en complétant le code de l'archive fournie avec ce TP.

Dans le graphique à bulles, la couleur de chaque bulle représente le continent et sa taille représente la population. Pour rendre le graphique à bulles interactif, un bouton permet de basculer entre les vues des données des années 2000 et 2015. Lorsque les données sont mises à jour à l'aide du bouton, les cercles se déplacent progressivement vers leurs nouvelles positions avec une animation.

Les sous-sections suivantes présentent les différentes parties que vous aurez à compléter pour ce travail pratique. Nous vous recommandons de compléter l'exploration des données dans le cahier Jupyter en premier. Ensuite, ce code vous aidera à réaliser le prétraitement des données, suivi de l'implémentation du graphique à bulles animé. Nous vous recommandons d'implémenter le code de l'info-bulle comme dernière étape.

Structure des fichiers

Pour réaliser ce travail, vous devrez remplir les différentes sections **TODO** dans les fichiers de l'archive fournie pour le travail pratique. Les commentaires dans le code expliquent plus en détail les étapes à suivre. Les scripts à utiliser sont situés dans le répertoire **assets** de l'archive fournie pour le travail pratique.

Dans ce travail pratique, nous vous fournissons 5 fichiers Python utilisés pour accomplir la visualisation souhaitée:

- **app.py**: Ce fichier génère la structure HTML de la page Web et orchestre les étapes requises pour créer la visualisation. Vous n'avez pas besoin de le modifier.
- **bubble.py**
- **hover_template.py**
- **preprocess.py**

- `server.py`: Ce fichier est utilisé pour lancer l'application. Vous n'avez pas besoin de le modifier.

Données

L'ensemble de données se trouve dans le répertoire `src/assets/data/` dans l'archive fournie pour le travail pratique. L'ensemble de données contient un tableau d'objets pour les années 2000 et 2015. Chaque objet contient les clés suivantes :

- `Country Name`: Le nom du pays.
- `PIB`: PIB par habitant en dollars américains courants.
- `CO2`: les émissions de CO2 par habitant en tonnes métriques.
- `Population`: la population du pays.
- `Continent`: Le continent du pays.

Exploration des données

Pour cette partie, utilisez le cahier Jupyter fourni avec ce travail pratique.

Dans le cahier, vous allez implémenter le code pour créer des nuages de points pour les données de 2000 et 2015. Ces graphiques seront affichés dans une application interactive Dash. Les directives spécifiques concernant les nuages de points sont contenues dans le cahier Jupyter. Comme vous le verrez, les nuages de points sont interactifs: lorsqu'une bulle est survolée dans l'un des deux graphiques, les informations sur le pays donné pour les deux années sont affichées dans le tableau en bas de page.

Pour plus d'indications, la figure 1 ci-dessous affiche à la fois les nuages de points et le comportement de l'application lorsqu'une bulle est survolée.



Figure 1: Les nuages de points résultant de l'exploration des données. La bulle pour le pays Trinidad et Tobago en 2015 est survolée par le curseur.

Prétraitement des données

Pour cette partie, vous devrez modifier le code fourni dans l'archive du travail pratique. Pour commencer, vous devrez prétraiter les données que nous vous fournissons. Les données contenu dans le fichier JSON sont brutes, il est donc nécessaire de réorganiser certaines parties de celles-ci afin qu'elles puissent être correctement utilisés par la bibliothèque Plotly. Pour ce faire, vous devez compléter le fichier `preprocess.py`.

Les étapes sont:

1. Arrondissez les nombres dans les données vers le bas afin qu'ils aient moins de décimales à afficher (fonction `round_decimals`). Vous pouvez réutiliser une partie de votre code de votre cahier pour cette étape.
2. Obtenez la plage de valeurs possibles pour les axes x et y (fonction `get_range`). Vous pouvez réutiliser une partie de votre code de votre cahier pour cette étape.
3. Combinez les données dans une structure plus facilement lisible par Plotly (fonction `combine_dfs`)

4. Triez les données par année et par continent pour simplifier l'affichage (fonction `sort_dy_by_yr_continent`)

Graphique à bulles animé

Pour cette troisième partie, le code à modifier se trouve dans l'archive du travail pratique. Plus précisément, il se trouve dans le fichier `bubble.py`. Cette partie consiste à générer la partie principale de la visualisation des données. Pour cette partie, vous allez d'abord générer le graphique à bulles avec une animation vis-à-vis l'année. Les axes x et y sont sur une échelle logarithmique. Les axes x et y représentent respectivement le PIB et les émissions de CO2 par habitant. La couleur de chaque cercle représente son continent, tandis que sa taille représente sa population. Une fois que le graphique à bulles suit ces spécifications, vous complétez son animation en simplifiant le menu qui y est associé et en vous assurant que le modèle (*template*) de survol est défini pour chaque trame de l'animation. Enfin, vous effectuerez quelques ajustements visuels, tels que la définition des étiquettes des axes, du modèle (*template*) et de la légende de la figure.

Pour résumer, les étapes de cette partie sont les suivantes:

1. Dessinez le graphique à bulles animé (fonction `get_plot`). Vous pouvez réutiliser une partie de votre code de votre cahier pour cette étape.
2. Mettez à jour le modèle de survol pour qu'il apparaisse sur toutes les trames de l'animation (fonction `update_animation_hover_template`)
3. Simplifiez l'affichage du menu d'animation (fonction `update_animation_menu`)
4. Mettez à jour certains éléments visuels de la figure, tels que les étiquettes des axes, le modèle et la légende (fonctions `update_axes_labels`, `update_template` et `update_legend`)



Figure 2: Le graphique à bulles avec des données de 2000



Figure 3: Le graphique à bulles avec des données de 2015

Info-bulle

Pour cette quatrième partie, vous définirez un modèle permettant d'afficher une info-bulle qui apparaît lorsque le curseur survole une bulle. L'info-bulle doit contenir le nom du pays, la population, le PIB par habitant et les émissions de CO2 par habitant, dans cet ordre. Chaque information doit être précédée d'une étiquette correspondante et suivie, le cas échéant, de son unité de mesure. L'intégralité du code de cette section peut être écrite dans la fonction `get_bubble_hover_template` dans le fichier `hover_template.py`. Assurez-vous que vous avez également géré correctement le modèle (*template*) de survol lorsque vous avez créé le graphique à bulles.

Voir la figure 4 pour le résultat attendu.



Le graphique à bulles pour 2015 avec la bulle pour la Chine survolée

Soumission

Les instructions pour la soumission sont:

1. Vous devez placer le code de votre projet dans un fichier ZIP compressé nommé [Matricule1_Matricule2_Matricule3.zip]
2. Le travail pratique doit être soumis avant le [20 Mars 23:59]

Évaluation

Dans l'ensemble, votre travail sera évalué selon la grille suivante. Chaque section sera évaluée sur l'exactitude et la qualité du travail.

Exigence	Points
Exploration des données	7
Prétraitement des données	3
Diagramme à bulles animé	7
Info-bulle	2
Qualité globale et clarté de la soumission	1
Total	20

Références

[1] The World Bank, "DataBank ," The World Bank. Available: <https://databank.worldbank.org/home> [Accessed 01 09 2020].