# 4DVST (Data Visualization Tools)

Jean-Baptiste, Miguel, Olivier

Table des matières

[4DVST (Data Visualization Tools) 1](#_Toc162561747)

[Contexte du projet 3](#_Toc162561748)

[Description du cahier des charges 3](#_Toc162561749)

[Description du dataset 3](#_Toc162561750)

[Champs du dataset 4](#_Toc162561751)

[Composition de l’équipe 5](#_Toc162561752)

[Choix des technologies 5](#_Toc162561753)

[Réponses aux questions 5](#_Toc162561754)

## Contexte du projet

Dans la continuation de notre formation SUPINFO en Master 1 en spécialisation Data Engineer, il nous a été demandé d’accomplir un travail de visualisation de données, à la suite du cours qui nous a été dispensé à ce sujet.

## Description du cahier des charges

Nous avons reçu un cahier des charges assez précis, mais qui laisse tout de même place à de la créativité et à de l’interprétation.

Nous avions quatre visualisations à faire, qu’il faut regrouper dans un dashboard, puis il a fallu faire un data storytelling qui interprète les visualisations de données.

## Description du dataset

Un dataset nous à été fourni dans le contexte de ce projet. Il s’agit de données synthétiques de la production scientifique mondiale de 1996 à 2014.

### Champs du dataset

Le dataset comporte neuf champs qui vont être détaillés ci-dessous.

* Documents

Le nombre de documents produit durant l’année.

* Citable.documents

Le nombre de documents citables produit durant l’année.

* Citations

Le nombre de citations de documents durant l’année.

* Self.Citations

Le nombre de citations interne au pays (citation au sein d’un pays pour un document produit dans le même pays).

* Citations.per.Document

Le nombre moyen de citations d’un document durant l’année.

* H.index

Indice qui permet de quantifier la production scientifique d’un scientifique. Il est calculé comme ceci : un chercheur avec un indice de *h* a publié *h* articles qui ont été cités au moins *h* fois.

Il s’agit ici d’une moyenne du hindex de chaque scientifique d’un pays pour une année.

* Year

L’année.

* Rank

Rang du pays en termes de production de document par rapport aux autres pays du dataset pour une année.

* Country

Le nom du pays.

## Composition de l’équipe

L’équipe initiale était constituée de Jean-Baptiste et de Miguel, elle a été rejointe par Olivier en cours de projet.

En termes de distribution des rôles, Jean-Baptiste et Miguel ont collaborés sur les quatre visualisations ainsi que le dashboard. Olivier, quant à lui, avant de rejoindre l’équipe, à fait les quatre visualisations de son coté, pour ensuite rejoindre l’équipe et partager son travail avec les autres afin de s’en inspirer (ce travail n’a pas pu être intégré au projet car il est basé sur une technologie différente).

Le data storytelling ainsi que la documentation est le fruit de la collaboration des trois membres de l’équipe.

## Choix des technologies

Comme mentionné précédemment, il y a deux versions du projet, en ce qui s’agit des visualisations.

Les visualisations faites par Olivier, qui ne constituent pas ceux du projet final, ont été fait en Python avec les librairies numpy, pandas, geopandas et matplotlib. Ces technologies sont bien implantées dans le monde de la visualisation de données, ce qui fait qu’elles sont très bien documentées. En revanche, leur bémol est leur interface peu esthétique et la difficulté de la mise en place de dashboards.

Les visualisations gardées pour la version finale du projet ont été réalisés avec Python et les librairies streamlit, pandas et geopandas. En comparaison avec matplotlib, strealit propose une interface web épuré et donne la posibilité de regrouper les visualisations en dashboards facilement. Il nous a donc parut pertinent de garder cette version du projet plutôt que celle utilisant matplotlib.

Nous utilisons le gestionnaire de dépendances Poetry, qui nous permet de facilement déployer un projet identique sur différentes machines et de créer un manifest du projet qui liste les dépendances.

## Réponses aux questions

**Question 1 : Quels pays disposent des scientifiques les plus performants et pourquoi‎ ?**  
Les Pays les Plus Performants

Selon les données, les États-Unis sont en tête quand on parle de la performance des scientifiques, suivis de loin par des pays comme la Chine, le Royaume-Uni, l'Allemagne, et le Japon. Ces pays ont produit le plus grand nombre de documents scientifiques et ont des indices H (H-index) élevés, ce qui montre que leurs recherches sont souvent citées par d'autres scientifiques.

Pourquoi Sont-ils Performants ?

Beaucoup de Recherches : Ces pays font beaucoup de recherches. Par exemple, les États-Unis ont le plus grand nombre de documents publiés, ce qui montre qu'ils ont beaucoup de scientifiques qui travaillent sur plein de sujets différents.

Recherches Citées : Leur travail est bien reconnu. L'indice H élevé des États-Unis signifie que leurs recherches sont souvent utilisées comme référence par d'autres scientifiques dans le monde. Ça veut dire que ce qu'ils découvrent est important et utile pour plein de gens.

Basé sur les Données :

En regardant les chiffres dans le document, on voit que les pays avec le plus de documents et les indices H élevés investissent probablement beaucoup dans la science. Ils ont peut-être de bons programmes pour soutenir les scientifiques, comme des financements pour les recherches ou des universités et des labos de pointe.

Ces pays montrent aussi qu'ils peuvent attirer des scientifiques talentueux, peut-être parce qu'ils offrent de bonnes conditions de travail ou parce qu'ils ont une longue tradition de recherche scientifique. Le fait d'avoir beaucoup de documents publiés et un indice H élevé dans le document suggère que ces pays sont vus comme des leaders dans le domaine de la science.

**Question 2 : Qu'est-ce qui justifierait que certains pays aient plus de scientifiques que d'autres‎ ?**  
Les pays avec une grande production de documents scientifiques et un indice H élevé sont probablement ceux qui investissent significativement dans la recherche et le développement. Cela peut être dû à des budgets de recherche élevés, à des politiques gouvernementales favorisant la science, et à des infrastructures de recherche de haute qualité.

La présence de ces pays en tête de liste peut aussi refléter une tradition longue et établie de recherche scientifique ainsi que l'existence de nombreuses institutions d'enseignement supérieur et de recherche de renommée internationale.

La différence entre les pays en termes de production scientifique et d'impact pourrait aussi être influencée par la taille de leur population, leur économie, et leur niveau d'accès aux ressources internationales de collaboration et de publication.

**Question 3 : Comment revitaliser les‎ ‎‎5 pires pays ?**  
  
"Investissement dans la recherche et développement (R&D) :

Encourager les gouvernements à investir davantage dans la recherche scientifique et technologique. Cela pourrait inclure l'allocation de fonds spécifiques, la création de subventions pour les chercheurs et les institutions de recherche, et le financement de projets de recherche innovants.",

"Amélioration de l'infrastructure de recherche :

Développer et moderniser les infrastructures de recherche, telles que les laboratoires, les centres de recherche et les universités. Cela pourrait impliquer la construction de nouvelles installations de recherche, l'acquisition d'équipements de pointe et la mise en place de collaborations internationales.",

"Éducation et formation :

Renforcer les programmes d'éducation scientifique et technologique à tous les niveaux, de l'enseignement primaire à l'enseignement supérieur. Encourager la formation continue des chercheurs et des professionnels de la science pour maintenir leurs compétences à jour et favoriser l'innovation.",

"Promotion de la collaboration internationale :

Encourager la collaboration et les échanges scientifiques internationaux en facilitant les partenariats entre les chercheurs, les institutions de recherche et les organisations internationales. Cela pourrait favoriser le partage des connaissances, des ressources et des meilleures pratiques.",

"Création d'incitations :

Mettre en place des incitations pour encourager la recherche et l'innovation, telles que des prix, des subventions de recherche, des crédits d'impôt pour la recherche et le développement, et des politiques de propriété intellectuelle favorables à l'innovation.",

"Promotion de l'entrepreneuriat scientifique :

Encourager la création d'entreprises innovantes dans le domaine des sciences et de la technologie en fournissant un soutien financier, une assistance technique et des infrastructures adaptées.",

"Transfert de technologie :

Faciliter le transfert de technologie et de connaissances scientifiques des institutions de recherche vers le secteur industriel afin de stimuler l'innovation et la croissance économique."