

Énoncé TP2

Cours :	INF 735 – Entrepôt et forage de données
Trimestre :	Automne 2018
Professeur :	Robert J. Laurin

1. Directives

- TP à faire individuellement ou en groupe de trois (3) étudiants au maximum
- Date de remise : *lundi le 3 décembre* – 18h45 (aucun travail en retard ne sera accepté)
- La remise de la version électronique doit se faire à michel.hebert3@usherbrooke.ca (la réception de la version électronique sera la preuve de remise du travail)
- Le plagiat, en totalité ou partie, ne sera pas toléré et sera sanctionné.
- Ce TP vaut 15% de la note finale

2. Livrables

Application MS Power BI (format pbix) contenant le résultat de votre visualisation des données.

Document (Format MS Word ou PDF) contenant les explications de la visualisation des données.

3. Mise en Contexte

Un chercheur universitaire a trouvé sur le Web des fichiers contenant des données statistiques sur les pays de la planète, entre 1960 et 2016. Ces fichiers contiennent beaucoup d'informations, mais il est très difficile de les comprendre simplement dans un tableur. Sachant votre intérêt pour l'analyse de données, il vous demande de l'aide pour visualiser certaines informations qui lui semblent pertinentes pour ses recherches. Il vous remet donc une copie des fichiers ainsi que des croquis et explications sommaires de ce qu'il désire voir comme résultats.

4. Démarche

Pour ce travail, vous devez réaliser, à l'aide de l'outil **Power BI** de Microsoft, les visualisations correspondantes aux exigences du chercheur. Ce genre de visualisation doit cependant toujours être accompagné d'une documentation claire expliquant d'où proviennent les données, les transformations qui leur sont appliquées et comment utiliser correctement la visualisation interactive. Vous devez donc produire cette documentation.

Récupérer les sources de données

Trois fichiers de données sont disponibles sur Moodle :

- **WPP2017_F01_LOCATIONS.xlsx** : il s'agit d'une liste des pays produite par un département des Nations Unies¹. La liste contient à la fois les pays, les continents et les régions. Les pays ont aussi quelques attributs sociaux économiques qui peuvent être utilisés comme dimensions d'analyse
- **Popular Indicators.csv** : est un fichier de données produit par la Banque mondiale² qui inclut plusieurs mesures par pays, entre 1960 et 2016.
- **Popular Indicators – Définitions.xlsx** est un complément au fichier précédent. Comme son nom l'indique, il contient la description des mesures incluses dans le fichier csv.

Charger les données dans Power BI

Vous devez charger ces données dans Power BI. Vous devez éditer le chargement pour bien organiser les données (tel qu'enlever les lignes d'entêtes superflues dans un fichier) dans le but de produire les visualisations proposées dans les croquis.

Modéliser les données chargées

Typiquement, on doit indiquer à Power BI comment les tables de données sont reliées entre elles afin de permettre la création de mesures, la navigation et les filtres.

¹ <https://esa.un.org/unpd/wpp/Download/Standard/Population/>

² <http://databank.banquemondiale.org/data/>

Créer les visualisations

Votre travail final doit comprendre les visualisations demandées par le chercheur et incluses à la suite de ce document. Les croquis sont des suggestions, vous pouvez proposer une visualisation différente pourvu qu'elle soit basée sur les mêmes mesures. Le résultat final doit être soigné (choix judicieux de la mise en page) car le chercheur va s'en servir lors d'une conférence mondiale devant des centaines de collègues.

Ajouter la navigation

Des points supplémentaires vous seront accordés si le rapport permet le forage (*drill down*). Par exemple, dans un graphique illustrant une valeur par continent, on peut voir les valeurs de ses pays en cliquant sur le continent.

Documenter le travail

Le document d'accompagnement doit contenir, au minimum, les informations suivantes :

- Objectifs de cette visualisation, sujet traité ce que l'on peut observer;
- Description des sources des données;
- Liste et justification des transformations apportées dans les données;
- Modélisation et définitions des mesures affichées;
- Explication sur l'utilisation du composant (filtres et navigation).

5. Autres consignes

La documentation doit être en français.

Les étiquettes dans l'application doivent être en français, mais il n'est pas nécessaire de traduire le contenu.

Si vous utilisez des composants complémentaires, la documentation doit en donner la source précise et ils doivent être gratuits.

6. Évaluation

L'évaluation du travail sera sur 100 points selon les critères suivants :

Application Power BI	Chargement : chargement réussi, nettoyage approprié, utilisation correcte des techniques.	20
	Visualisation : Choix approprié des composants visuels selon l'objectif et les données, esthétisme de la mise en page.	25
	Filtre : choix judicieux des filtres, les filtres fonctionnent sur les composants visés.	5
	Forage : forage fonctionnel et approprié au sujet et aux données.	10 (bonus)
Documentation	Objectifs : Les objectifs sont bien expliqués	5
	Sources : La description des sources de données est précise	5
	Transformations : Les transformations sont correctement détaillées et justifiées.	5
	Modélisation et mesures : Le modèle des données est présenté. Les mesures calculées sont bien définies.	10
	Utilisation : La navigation de l'application Power BI est bien expliquée de même que l'utilisation des filtres.	10
	Mise en page : La mise en page est cohérente et facilite la lecture.	5
	Français : Il n'y a pas de phâtes d'orthographes ☺, les phrases sont faciles à lire, les paragraphes bien découpés et enchainés.	10

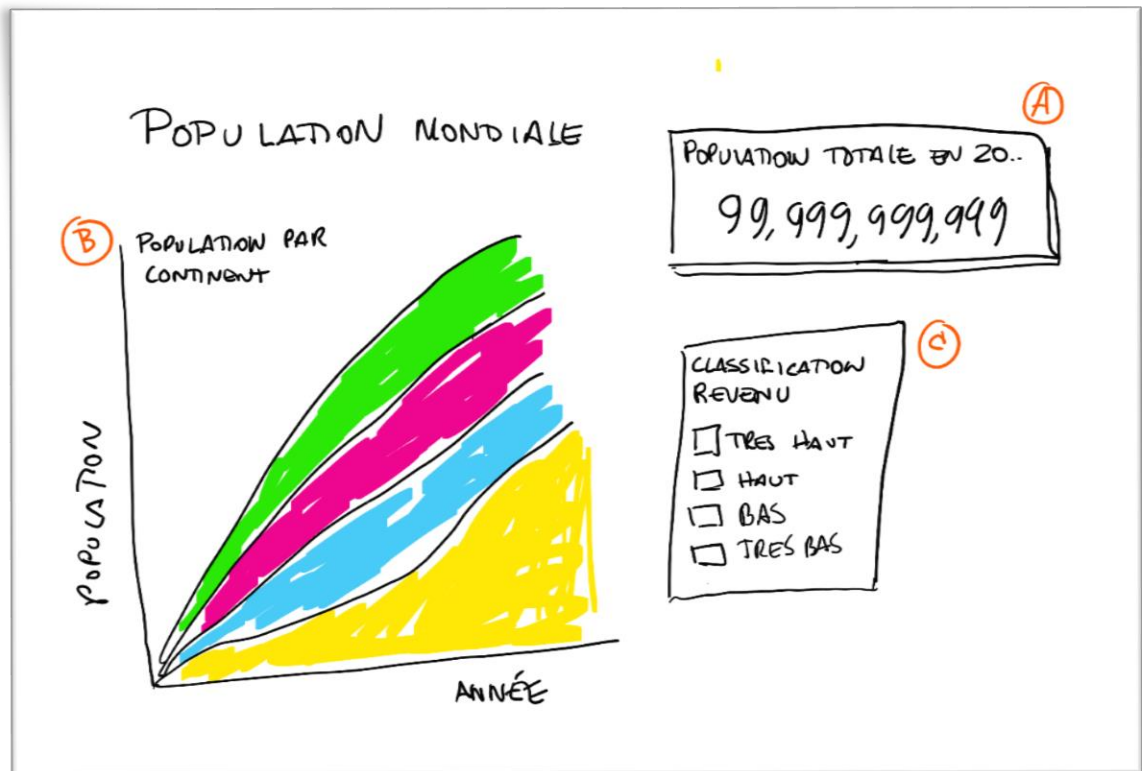
La note sera rapportée sur la valeur du travail pour le cours.

Bon TP.

Croquis du chercheur

Le chercheur vous propose trois croquis abordant chacun un thème distinct à partir des données dans les fichiers trouvés. Chaque croquis doit être présenté dans une page du résultat dans Power BI.

Population mondiale



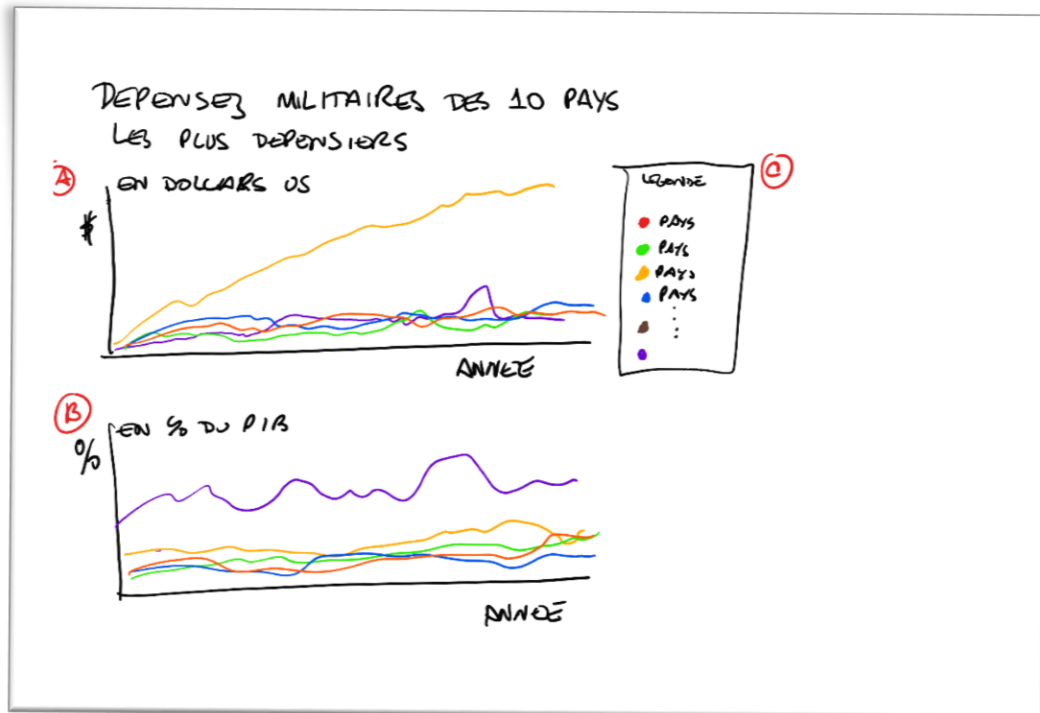
Cette page illustre l'évolution de la population mondiale dans les données recueillies.

Zone A : Il s'agit d'afficher le total de la population dans la dernière année disponible dans le fichier. Attention, ce nombre ne doit pas changer lorsqu'on change le filtre.

Zone B : Le graphique montre le total de la population par continent (Afrique, Amérique du Nord, etc.). Il doit y avoir une légende pour les couleurs utilisée.

Zone C : Le fichier des pays attribue une classification de revenu à chacun des pays, intitulé *Income Groups*. Il y a 5 niveaux de classification, mais les niveaux « moyen haut » et « moyen bas » sont des subdivisions du niveau « moyen ». Il y a donc 4 niveaux mutuellement exclusifs. Il serait intéressant de filtrer le graphique selon cette classification.

Dépenses Militaires



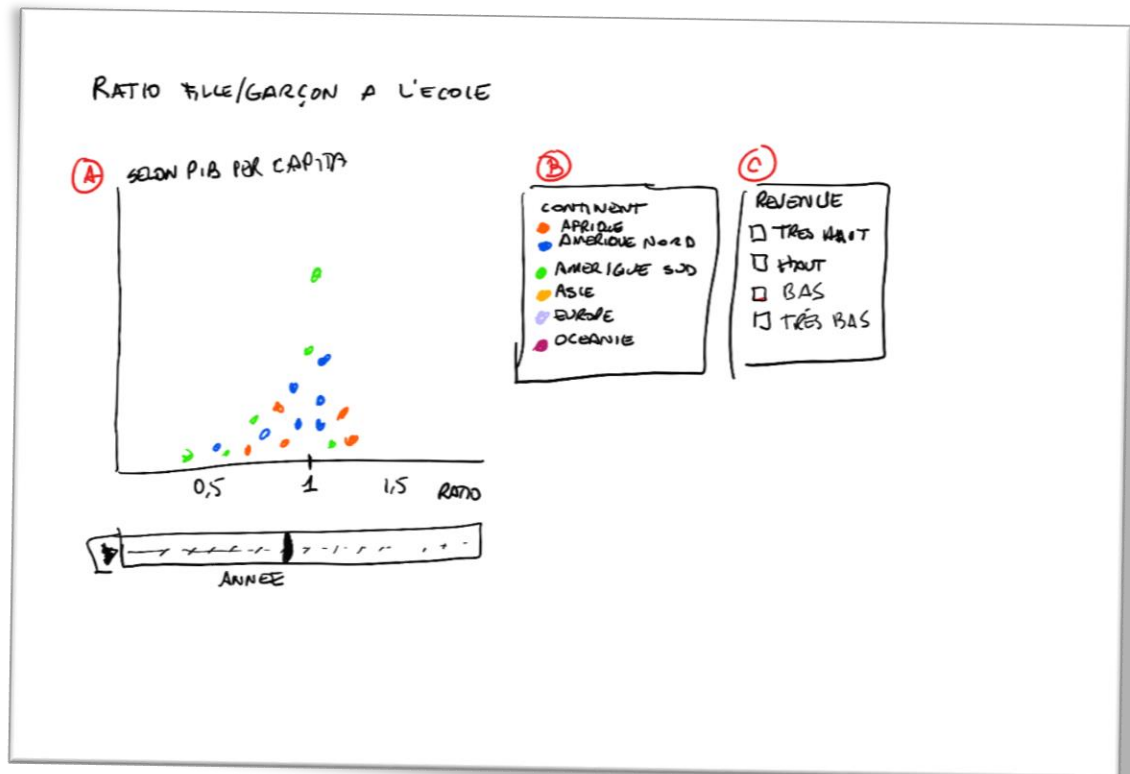
Dans cette page, le chercheur se concentre sur les 10 pays ayant les plus fortes dépenses militaires totales (en dollars américains) dans la dernière année.

Zone **A** : Ce graphique illustre l'évolution des dépenses militaires totales pour les 10 pays retenus.

Zone **B** : Ce graphique illustre l'évolution des dépenses militaires exprimées en pourcentage du produit intérieur brut (PIB) du pays.

Zone **C** : La légende des couleurs des deux graphiques peut servir de filtre si on désire consulter qu'un sous-ensemble de courbes.

Ratio filles – garçons à l'école



Le chercheur s'interroge aussi sur l'impact de l'éducation des filles dans le succès économique des pays. Le fichier des mesures inclut un ratio d'inscription au primaire et au secondaire basé sur le sexe des enfants.

Zone A : ce graphique pointe la relation, pour chaque pays, entre le ratio fille- garçon et le produit intérieur brut (PIB) per capita. L'année des données sert de filtre d'animation.

Zone B : La couleur des points doit correspondre au continent du pays et la légende sert aussi de filtre.

Zone C : Il est aussi possible de filtrer le graphique à partir de la classification de groupe de revenus du pays (même filtre que la zone C sur le croquis Population).