

## Didacticiel –Microsoft Power BI

Cours :	INF 735 – Entrepôt et forage de données
Trimestre :	Automne 2018
Enseignant :	Robert J. Laurin (Chargé de cours) et Michel Hébert (Chargé de laboratoires)

Ce didacticiel vous permettra d'explorer **Microsoft Power BI** comme outil analytique ou « front-end » d'un entrepôt de données.

### Contenu

Didacticiel –Microsoft Power BI.....	1
Prérequis.....	2
Qu'est-ce que Power BI ?.....	2
Étapes d'utilisation.....	3
Se connecter aux données.....	3
Contexte du laboratoire.....	4
Fichier de données.....	4
Visualisations demandées.....	5
Chargement des données.....	5
Retrait des lignes superflues.....	7
Filtrer les données.....	8
Division des faits.....	8
« Unpivot » de la population.....	9
Élimination des données superflues.....	10
Modélisation des données.....	11
Tables de dimensions.....	11
Relations.....	11
Visualisation.....	13
Distribution de la population par tranche d'âge.....	13
Distribution de la population par province.....	16
Création de la mesure PopulationCanadienne.....	18
Filtre contextuel.....	19
Visualisation de l'âge moyen.....	19
Analyse du revenu moyen par ménage selon la taille du ménage.....	21
Autres sources d'informations sur Power BI.....	23

## Prérequis

Pour ce didacticiel vous aurez besoin de :

- Microsoft PowerBI
- Les fichiers de données du laboratoire

## Qu'est-ce que Power BI ?

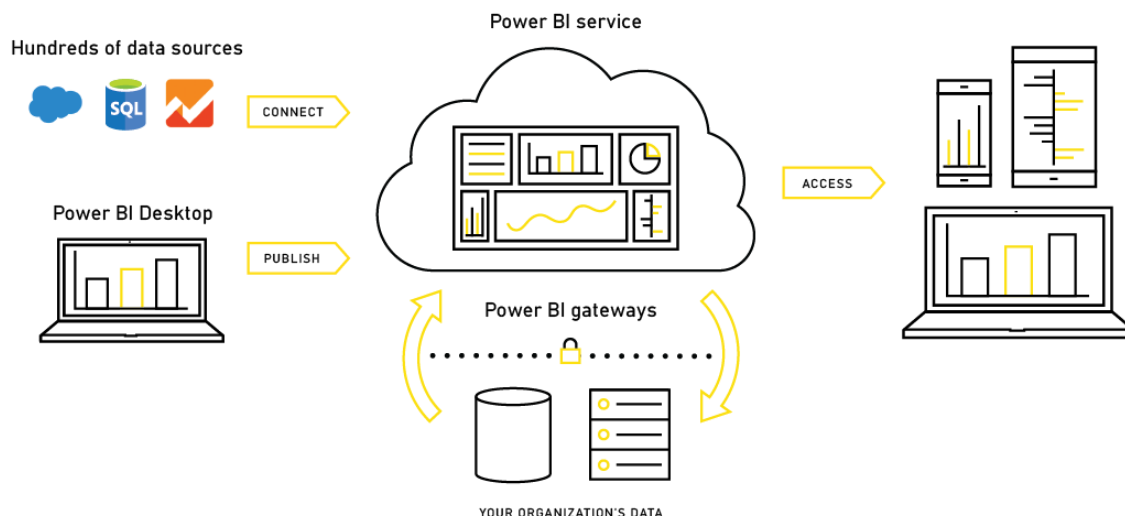
*Power BI est une suite d'outils d'analyse marketing offrant des insights dans l'ensemble de votre organisation. Connectez-vous à des centaines de sources de données, simplifiez la préparation des données et générez des analyses ad hoc. Produisez de magnifiques rapports, puis publiez-les pour que votre organisation puisse les utiliser sur le web et les appareils mobiles. Tout le monde peut créer des tableaux de bord personnalisés, avec une vue unique à 360° de l'activité. Bénéficiez d'une scalabilité à l'échelle de l'entreprise, avec une gouvernance et une sécurité intégrée.*

Note : cette section est fortement inspirée du site officiel de Power BI :

<https://powerbi.microsoft.com/fr-fr/>

La suite Power Bi comporte différents modules qui répondent à des besoins spécifiques :

- **Power BI Desktop** est un outil de création d'applications web hybride de données et de rapports riches en fonctionnalités.
- Utilisez le **service Power BI** pour publier des rapports en toute sécurité à votre organisation et configurer une actualisation automatique des données afin que tout le monde dispose des informations les plus récentes.
- Vous pouvez accéder à vos données et rapports où que vous soyez, à l'aide des applications **Power BI Mobile**, qui se mettent à jour automatiquement avec toutes les modifications apportées à vos données.
- Power BI peut unifier toutes les données de votre organisation, qu'elles se trouvent dans le cloud ou qu'elles soient stockées localement. Les **passerelles Power BI** vous permettent de connecter des bases de données SQL Server, des modèles Analysis Services, ainsi que de nombreuses autres sources de données à vos tableaux de bord dans Power BI.



Pour le cours 735, nous allons nous concentrer sur **Power BI Desktop**.

C'est un outil gratuit que vous pouvez installer à partir du site de Microsoft ou de son magasin (*app store*). Si vous avez Windows 10, il est recommandé d'installer Power BI Desktop depuis le magasin.

## Étapes d'utilisation

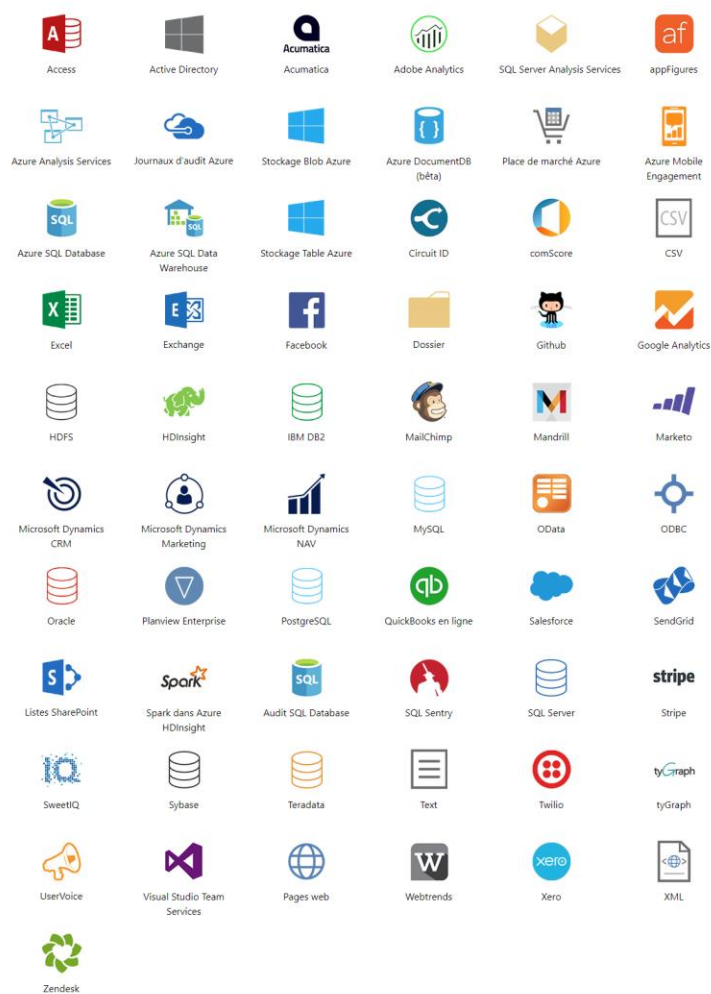
L'utilisation de Power BI desktop se divise en quatre étapes :

- 1- Extraire les données des sources
- 2- Modéliser les données extraites
- 3- Concevoir les pages de visualisations
- 4- Publier les résultats.

Pour ce didacticiel, nous allons nous concentrer sur les étapes d'extraction (1) et de conception (3).

## Se connecter aux données

Power Bi peut se connecter à plusieurs dizaines de sources de données distinctes, et la liste s'allonge à toutes les nouvelles versions.



## Contexte du laboratoire

Nous allons produire différentes visualisations à partir du profil de recensement Canada 2016<sup>1</sup>, en particulier sur la distribution de la population par tranche d'âges et la correspondance du revenu à la taille de la famille.

## Fichier de données

Le fichier de données téléchargé depuis le site de Statistiques Canada fait plus de 1,3 Go. Il est trop gros pour le laboratoire, nous en avons donc extrait un sous-ensemble. C'est à partir de ce fichier que nous allons travailler. Voici les fichiers de données :

**INF735 – Labo2\_définitions.txt** : définition des mesures et d'autres explications sur les données.

**INF735 - Labo2\_Data.xlsx** : fichier contenant les données de recensement :

	A	B	C	D	E	F	G	H	I
1					<b>Profil du recensement Canada 2016</b>				
2		Téléchargé de:			<a href="http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=F">http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=F</a>				
3					<b>Sous-ensemble des données préparé par Michel Hébert pour INF735</b>				
4									
5	annee_d	code_geo	niveau_gé	nom_géo	mesuredesc	profil	Total	total_masculin	total_feminin
6	2016	1	0	Canada	Revenu total médian des ménages en 2015 (\$)	727	70336	NULL	NULL
7	2016	1	0	Canada	Revenu après impôt médian des ménages en 2015 (\$)	728	61348	NULL	NULL
8	2016	1	0	Canada	Revenu total médian des ménages comptant une personne en 2015 (\$)	730	35174	NULL	NULL
9	2016	1	0	Canada	Revenu après impôt médian des ménages comptant une personne en 2015 (\$)	731	31446	NULL	NULL
10	2016	1	0	Canada	Revenu total médian des ménages comptant deux personnes ou plus en 2015 (\$)	733	88249	NULL	NULL
11	2016	1	0	Canada	Revenu après impôt médian des ménages comptant deux personnes ou plus en 2015 (\$)	734	76419	NULL	NULL
12	2016	1	0	Canada	Revenu total médian des familles économiques en 2015 (\$)	777	88306	NULL	NULL
13	2016	1	0	Canada	Revenu après impôt médian des familles économiques en 2015 (\$)	778	76372	NULL	NULL
14	2016	1	0	Canada	Taille moyenne des familles économiques	779	3	NULL	NULL
15	2016	1	0	Canada	Revenu total médian des familles économiques comptant un couple sans enfants ni autres pers	781	78976	NULL	NULL
16	2016	1	0	Canada	Revenu après impôt médian des familles économiques comptant un couple sans enfants ni autr	782	68309	NULL	NULL
17	2016	1	0	Canada	Taille moyenne des familles économiques comptant un couple sans enfants ni autres personne	783	2	NULL	NULL
18	2016	1	0	Canada	Revenu total médian des familles économiques comptant un couple avec enfants en 2015 (\$)	785	113465	NULL	NULL
19	2016	1	0	Canada	Revenu après impôt médian des familles économiques comptant un couple avec enfants en 20	786	96608	NULL	NULL
20	2016	1	0	Canada	Taille moyenne des familles économiques comptant un couple avec enfants	787	4	NULL	NULL
21	2016	1	0	Canada	Revenu total médian des familles économiques monoparentales en 2015 (\$)	789	53193	NULL	NULL
22	2016	1	0	Canada	Revenu après impôt médian des familles économiques monoparentales en 2015 (\$)	790	48856	NULL	NULL
23	2016	1	0	Canada	Taille moyenne des familles économiques monoparentales	791	2,7	NULL	NULL
24	2016	1	0	Canada	Revenu total médian en 2015 pour les personnes hors famille économique âgées de 15 ans et p	793	31890	34483	29801
25	2016	1	0	Canada	Revenu après impôt médian en 2015 pour les personnes hors famille économique âgées de 15 a	794	28959	30746	27495
26	2016	10	1	Terre-Neuve-et-Labrador 0 à 4 ans		10	22365	11430	10930
27	2016	10	1	Terre-Neuve-et-Labrador 5 à 9 ans		11	26040	13430	12605
28	2016	10	1	Terre-Neuve-et-Labrador 10 à 14 ans		12	26035	13340	12700
29	2016	10	1	Terre-Neuve-et-Labrador 15 à 19 ans		14	27255	13980	13275
30	2016	10	1	Terre-Neuve-et-Labrador 20 à 24 ans		15	27700	13915	13785

Tel que l'on peut observer dans la figure, le fichier contient quelques lignes titres avant la section des données. Les colonnes de données sont les suivantes :

Titre de colonne	Description
<b>annee_de_recensement</b>	L'année du recensement, soit 2016 pour toutes les lignes
<b>code_geo</b>	Code de 1 à 4 chiffre identifiant la zone géographique. 1 correspond au Canada. Les valeurs entre 10 et 99 sont pour les provinces. Les valeurs plus grandes sont des régions spécifiques des provinces (régions, centres urbains, etc.)
<b>niveau_géo</b>	Le niveau hiérarchique de la zone géographique : 0 est le pays, 1 les provinces, etc.
<b>nom_géo</b>	Nom de la région géographique
<b>mesuredesc</b>	Description de la mesures
<b>profil</b>	Code identifiant la mesure
<b>Total</b>	Valeur de la mesure. Pour les mesures de population, total des hommes et des femmes.
<b>total_masculin</b>	Pour les mesures de population, valeur pour les hommes, NULL lorsque non applicable.
<b>total_feminin</b>	Pour les mesures de population, valeur pour les femmes, NULL lorsque non applicable.

<sup>1</sup> Disponible à [http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page\\_dl-tc.cfm?Lang=F](http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/download-telecharger/comp/page_dl-tc.cfm?Lang=F)

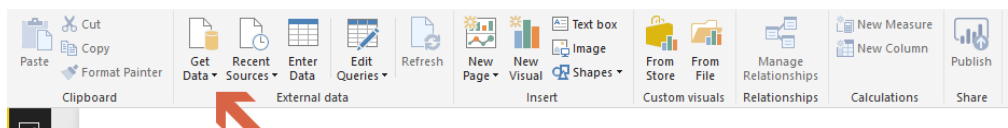
## Visualisations demandées

On nous demande d'abord une analyse de la distribution de la population par tranche d'âge. On veut distinguer la distribution pour l'ensemble du Canada ou par province et comparer la population des hommes et des femmes. On veut aussi voir l'âge moyen des femmes et des hommes par provinces.

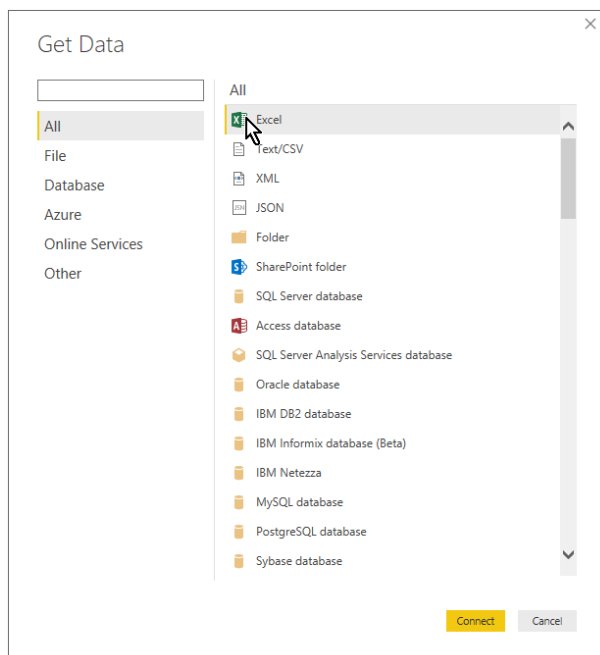
Enfin, il serait intéressant de voir une comparaison, pour chaque province, du revenu moyen par ménage par la taille du ménage.

## Chargement des données

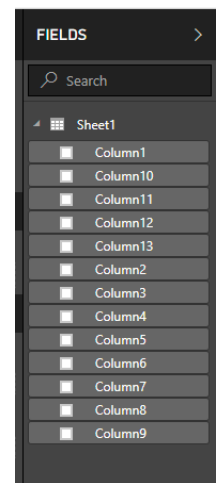
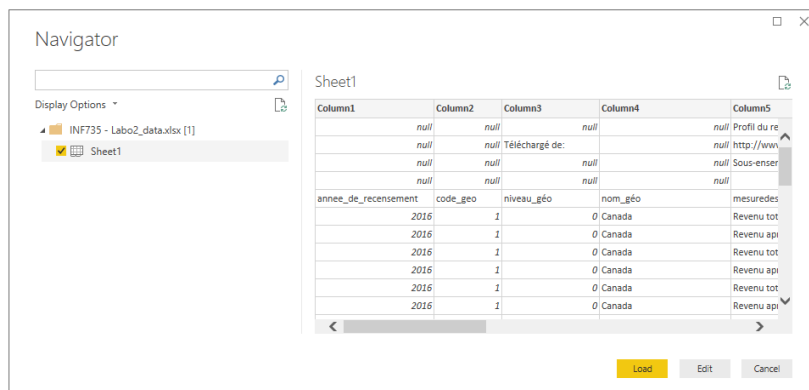
Après avoir démarré Power BI desktop, cliquer sur le bouton **Get Data** pour charger le fichier de données.



Choisir Excel dans la liste, puis cliquer sur **Connect**. Vous devrez alors préciser l'emplacement du fichier sur votre ordinateur. Prendre note que le fichier ne peut pas être ouvert dans Excel ou une autre application.

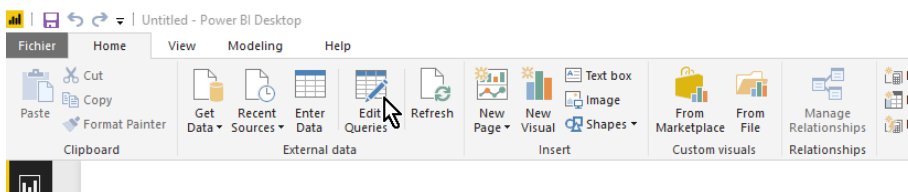


Power BI analyse alors le fichier de données et identifie les onglets qu'il contient. Dans notre cas, le fichier ne contient qu'un seul onglet. Vous devez le choisir en cochant la case correspondante. Lorsque le fichier contient plusieurs onglets, il est possible d'en choisir plus d'un. Power BI fera alors une « table » pour chaque onglet. Cliquer sur **Load** pour démarrer le chargement.

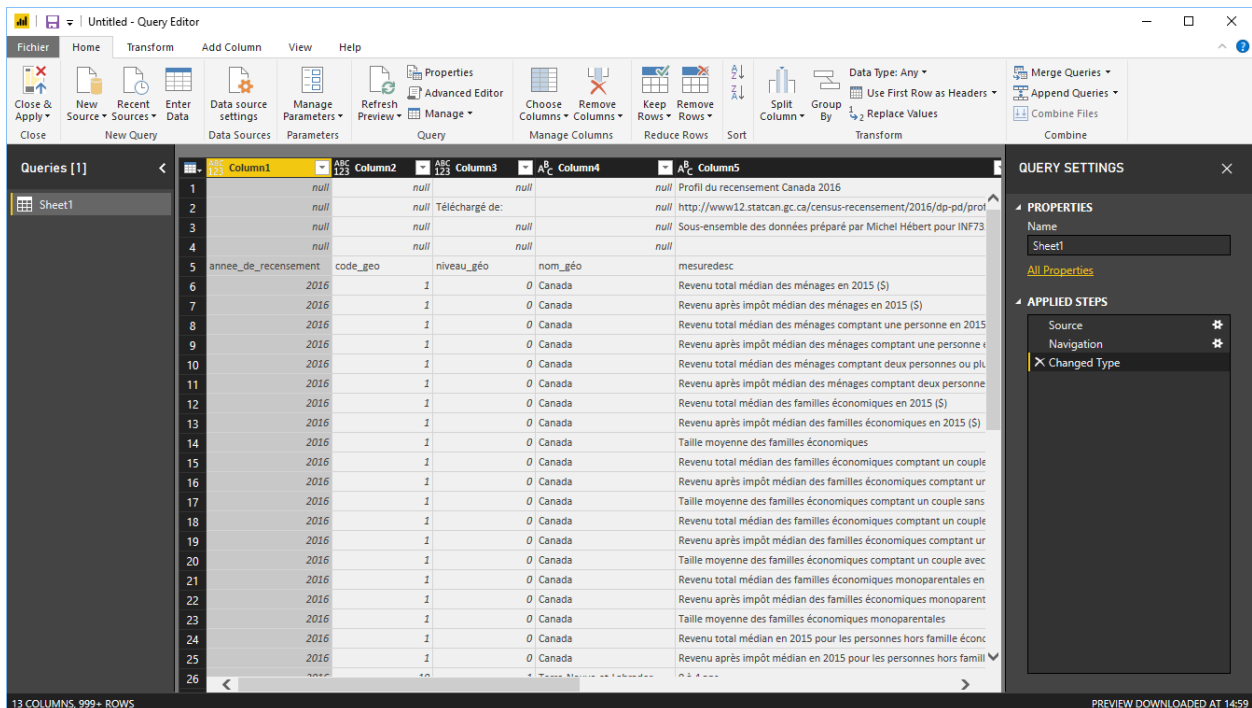


Après le chargement, on peut voir dans l'explorateur de PowerBi cette nouvelle table. On doit faire quelques manipulations dans l'éditeur de requête (*query editor*) pour mieux organiser l'information.

Dans la barre de menu, cliquer sur **Edit Queries** pour démarrer l'éditeur de requêtes.



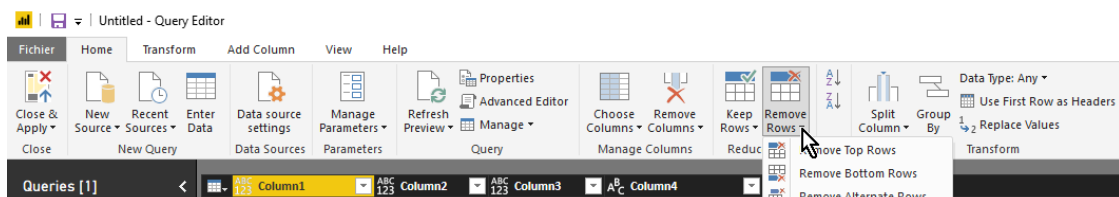
Une nouvelle fenêtre s'ouvre avec les requêtes déjà définies dans notre environnement. Il n'y en a qu'une seule pour le moment dans notre cas.



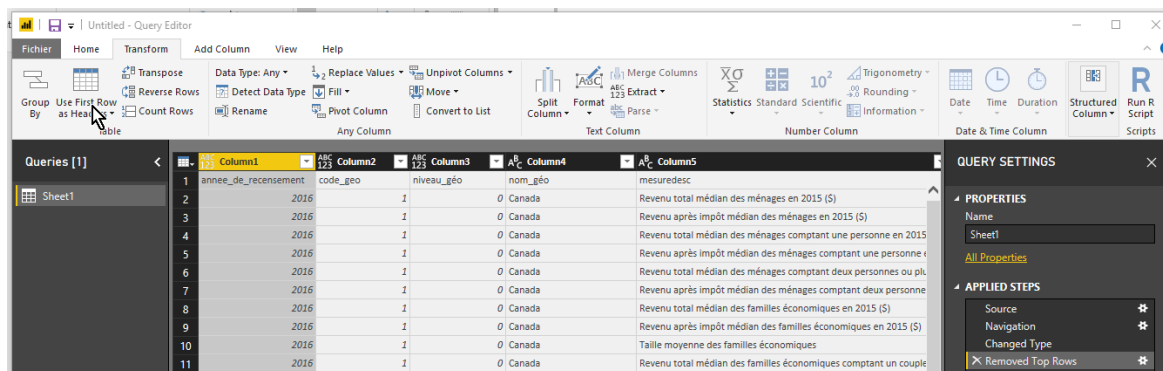
L'éditeur de requêtes possède une collection de rubans avec les commandes nous permettant d'ajuster le résultat des requêtes. Il est possible de faire plusieurs transformations sans écrire de programme. L'éditeur contient à gauche un panneau de navigation dans les requêtes, un panneau central qui affiche le résultat courant de la requête active et, à droite, un panneau avec l'historique des transformations appliquées. Il est possible d'effacer des transformations dans ce dernier panneau.

## Retrait des lignes superflues

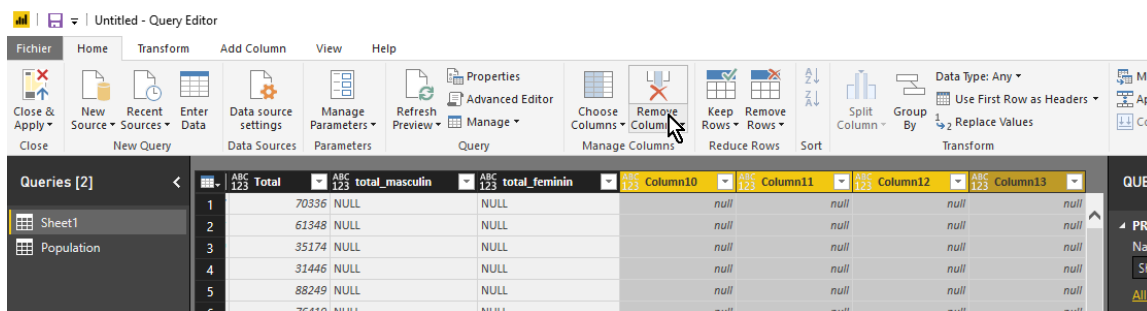
La première étape est d'éliminer les lignes titres qui ne contiennent pas d'information et d'indiquer que la première ligne des données contient en fait le nom des colonnes. Cliquer sur **Remove Rows** puis, dans la liste des commandes, choisir **Remove Top Rows**.



Power BI vous demandera combien de lignes doivent être enlevées. Inscrire 4 puis cliquer sur Ok. On doit maintenant indiquer que la première rangée contient en fait le nom des colonnes. Dans le ruban **Transform**, cliquer sur **Use First Row as Header**.

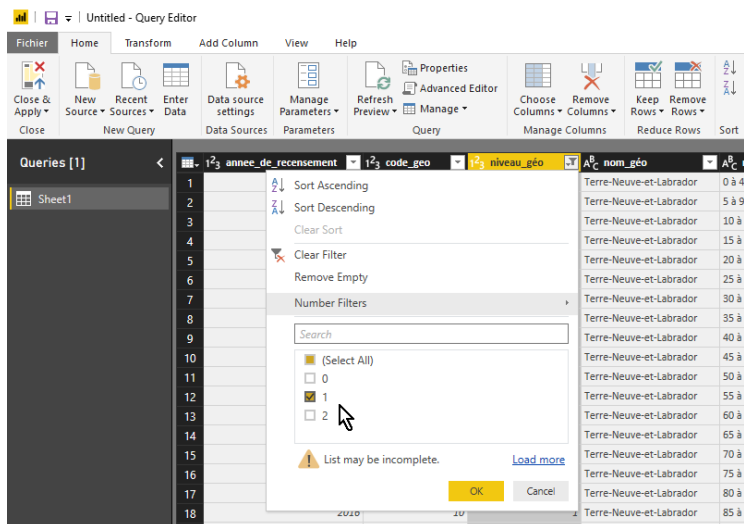


Si on observe la table, on remarque que la requête a conservé des colonnes superflues à la fin de chaque ligne. Ces colonnes sont inutiles et il serait préférable de les enlever. Choisir les colonnes à enlever (avec contrôle-clic) puis cliquer sur **Remove Columns** (attention, il ne faut pas choisir **Remove Other Columns**).



## Filtrer les données

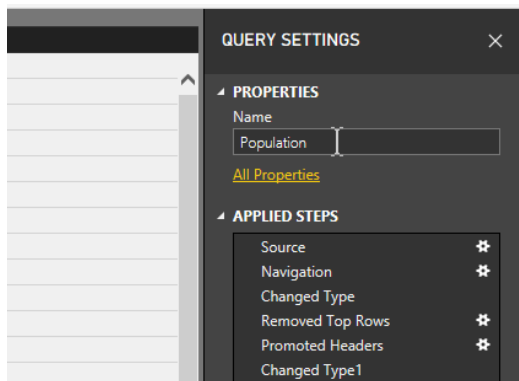
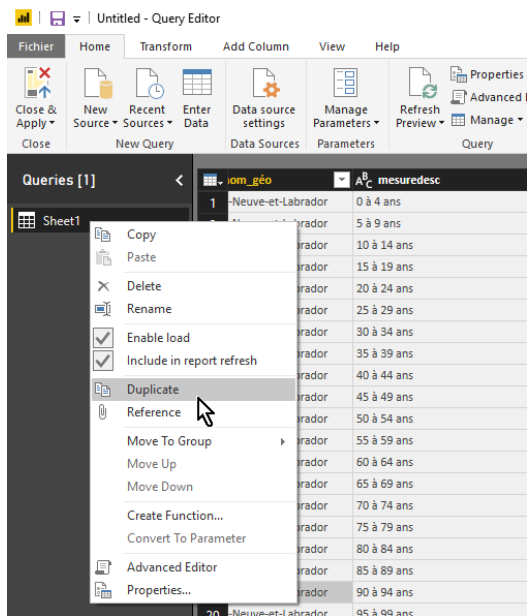
Puisque les visualisations qu'on nous demande ne concernent que les provinces, on peut filtrer les rangées afin de ne conserver que les données pertinentes. Il s'agit en fait des lignes dont le niveau géographique est 1. Cliquer sur le bouton de la colonne **Niveau\_Geo**, puis ne cocher que la valeur 1 dans la liste des valeurs. Cliquer sur Ok. La table est maintenant filtrée.



## Division des faits

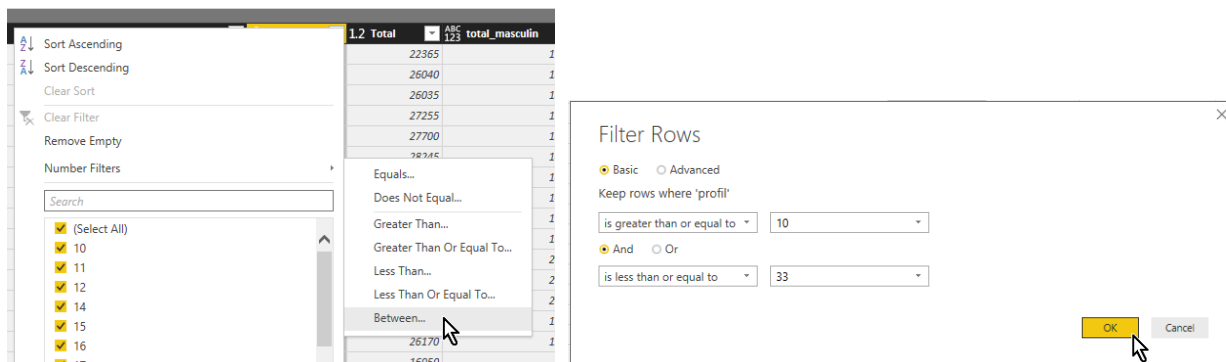
La table contient trois types de faits : la population par tranche d'âge, des données sur le revenu familial et des mesures sur la taille moyenne des familles. Pour l'analyse des données, il est plus simple de diviser ces faits dans trois tables distinctes.

La première étape est de doubler la table source. Cliquer avec le bouton de droite sur le nom de la table (sheet1) dans l'explorateur puis choisir la commande **Duplicate**. Cliquer ensuite sur le nom de la nouvelle table – sheet1 (2) – dans le panneau de droite et changer son nom pour **Population**.





Il faut maintenant filtrer les lignes de cette nouvelle table pour ne conserver que les mesures de population. Il s'agit en fait des mesures dont le code de profil est entre 10 et 33. Appliquer ce filtre sur la colonne **profil**.



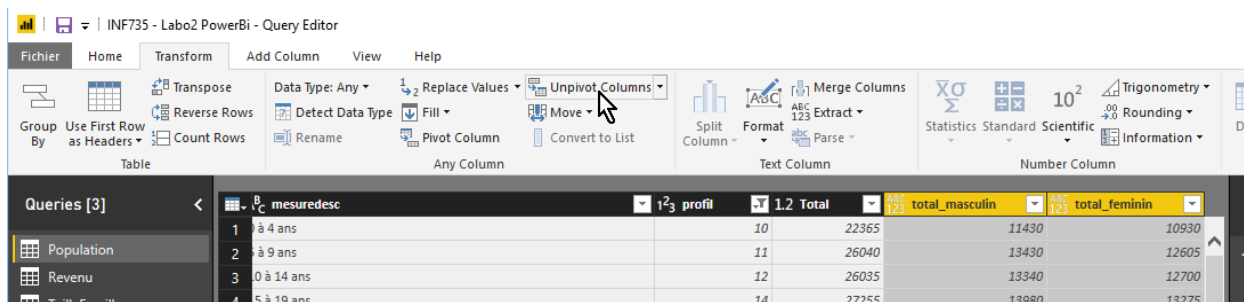
Faire les mêmes opérations pour les deux autres types de mesures :

- Revenu (description des mesures commence avec « Revenu » )
- TailleFamille (description des mesures commence avec « Taille » )

On peut maintenant retirer la table Sheet1 de la liste des requêtes. Cliquer avec le bouton de droite sur le nom de la table Sheet1 puis choisir la commande **Delete**.

## « Unpivot » de la population

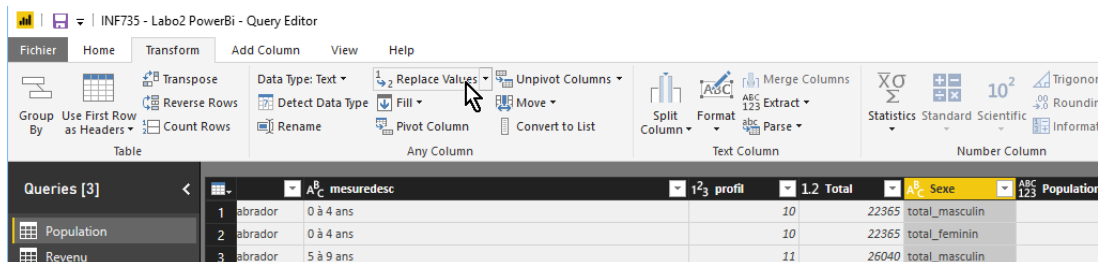
Dans la table Population, nous avons trois colonnes avec des mesures : total, total\_masculin et total\_Feminin. Il serait plus facile pour l'analyse d'avoir une colonne Population et de doubler les lignes pour distinguer les hommes des femmes. Nous allons utiliser la fonction *Unpivot* pour y arriver. Assurez-vous que la table Population est choisie, sélectionnez les colonnes total\_masculin et total\_feminin puis cliquer sur le bouton **Unpivot Columns** du ruban **Transform**.



Power BI remplace alors ces deux colonnes par une colonne **Attribute** qui contient le nom de l'une ou l'autre des anciennes colonnes, et une colonne **Value** qui contient la valeur correspondante à l'attribut. Le nombre de rangée a aussi été doublé.

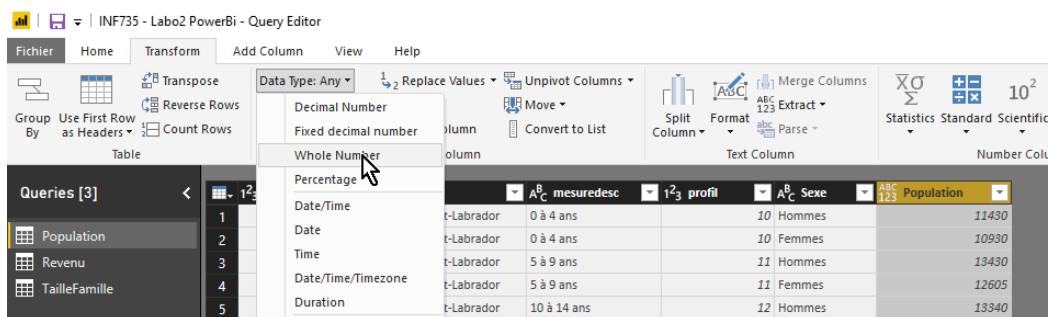
Afin de simplifier notre tâche, nous allons renommer la colonne **Attribute** à Sexe et la colonne **Value** à Population. Il suffit de cliquer avec le bouton de droite sur le nom de la colonne, choisir la commande **Rename...** puis inscrire le nouveau nom.

Enfin, nous allons corriger les valeurs de sexe pour remplacer total\_masculin par homme et total\_feminin par femmes. Choisir la colonne Sexe, puis cliquer sur **Replace Values** dans le ruban **Transform**.



Une fenêtre s'ouvre pour inscrire le texte à chercher et le texte de remplacement. Inscrire total\_masculin dans le texte à rechercher et Hommes dans le texte de remplacement puis cliquer sur **Ok**. Refaire la même opération pour remplacer total\_feminin par Femmes.

Finalement, on doit préciser que la colonne Population est une valeur numérique. Choisir la colonne Population, puis changer le type de données (ruban **Transform**) pour **Whole Number**.



## Élimination des données superflues

Pour faciliter le travail d'analyse, il est préférable d'enlever les colonnes inutiles, mais ce n'est pas une obligation. Pour enlever une colonne dans une table, il suffit de cliquer avec le bouton de droite sur le nom de la colonne, choisir la commande **Remove**.

Enlever les colonnes suivantes dans chacune des tables

Population	Revenu	TailleFamille
Année_de_Recensement	Année_de_Recensement	Année_de_Recensement
Niveau_Géo	Niveau_Géo	Niveau_Géo
Total		Total_masculin
		Total_Feminin

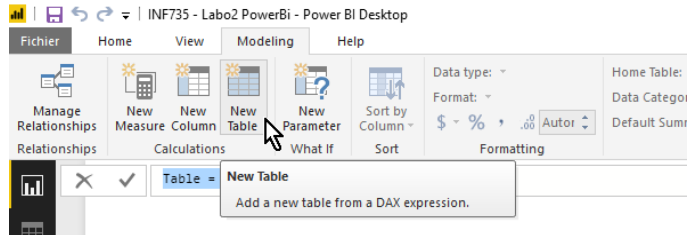
Cette étape est maintenant terminée, cliquer sur le bouton **Close & Apply** sur le ruban **Home** pour mettre à jour les données dans Power Bi et fermer l'éditeur.

## Modélisation des données

Lorsqu'on veut offrir à l'utilisateur des visualisations des filtres combinés, il est plus simple si les dimensions qui serviront de filtres sont dans leur propre tables et que les tables de faits sont reliées à ces dimensions. Il faut donc dans un premier temps créer ces tables de faits puis établir les relations.

### Tables de dimensions

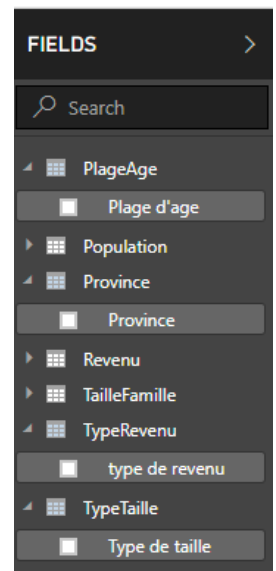
Nous allons commencer par faire une table pour la liste des provinces. Dans Power BI, cliquer sur le bouton **New Table** du ruban **Modeling**.



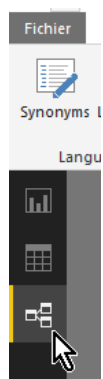
Une zone d'édition de formule s'affiche avec le texte par défaut : **Table =** . Remplacer ce texte par cette formule : **Province = DISTINCT(Population[nom\_géo])**. Cette formule extrait toutes les valeurs distinctes de la colonne nom\_géo de la table Population. L'équivalent SQL serait une vue basée sur cette requête : *Select distinct nom\_géo from Population*.

Créer une table pour les plages d'âge (PlageAge), une pour la dimension des revenus (TypeRevenu) et une autre pour la dimension taille de famille (TypeTaille) à partir des valeurs distinctes de la colonne MesureDesc des tables Population, Revenu et TailleFamille respectivement.

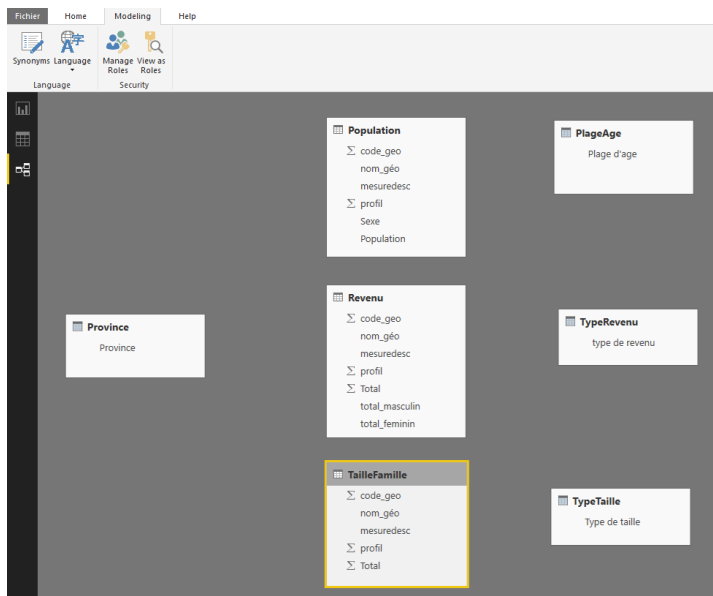
Dans l'explorateur des tables, exploser les nouvelles tables et renommer les champs (cliquer avec le bouton de droite sur le nom du champ, puis choisir la commande **Rename**) selon les noms dans l'illustration.



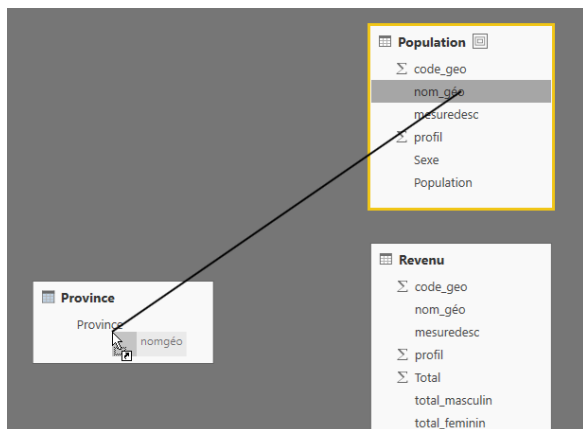
### Relations



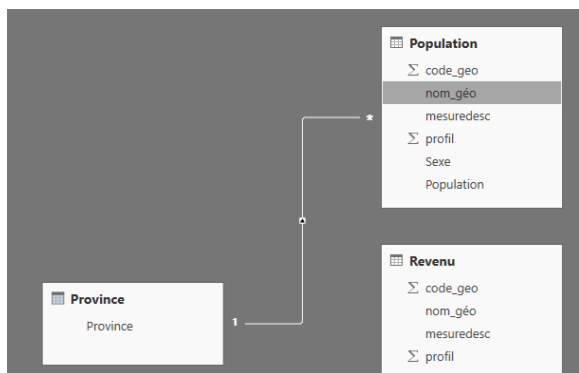
Cliquer sur le bouton **Relationship** dans le navigateur de gauche. L'espace de travail vous présentera un modèle contenant les 7 tables de notre environnement. Ces tables peuvent être déplacées et leur taille ajustée. Organiser les tables de la manière illustrée.



Pour créer une relation, il suffit de cliquer sur une colonne qui correspond à une clef étrangère (par exemple nom\_géo dans la table population, puis de glisser vers la table primaire en maintenant le bouton de la souris. Un trait sera alors dessiné entre les tables.

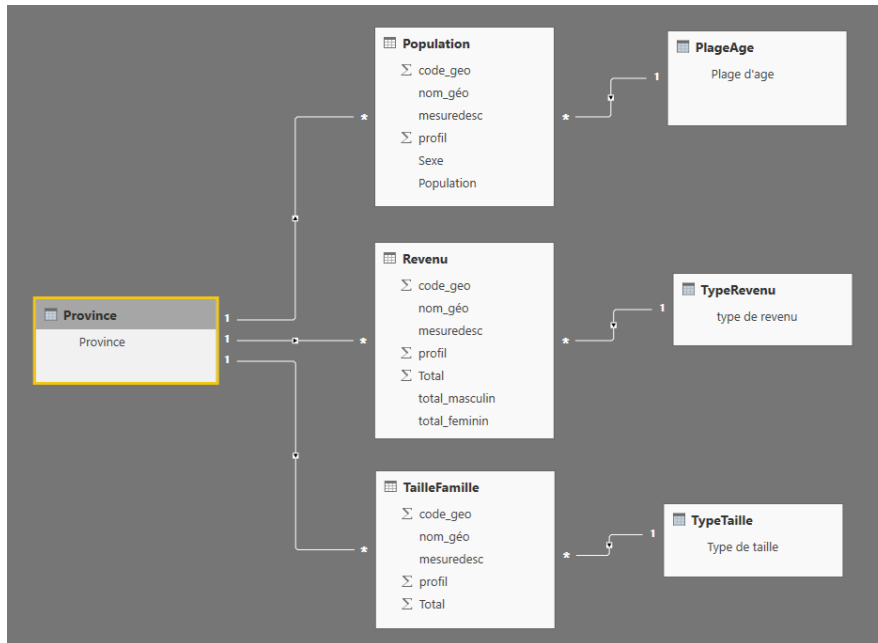


Relâcher le bouton de la souris et la relation sera créée :



Ajouter les relations entre nom\_géo de toutes les tables et la table Province, de même qu'entre les colonnes mesureDesc des tables et leur dimension correspondantes.

Le résultat final devrait ressembler à ceci :



Noter que le diagramme illustre la cardinalité des relations (1 à plusieurs).

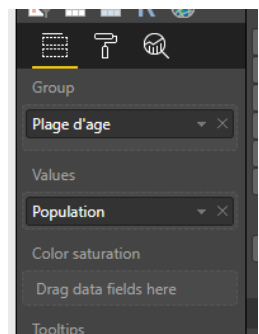
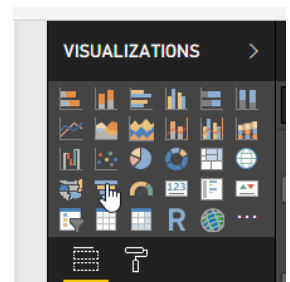
## Visualisation

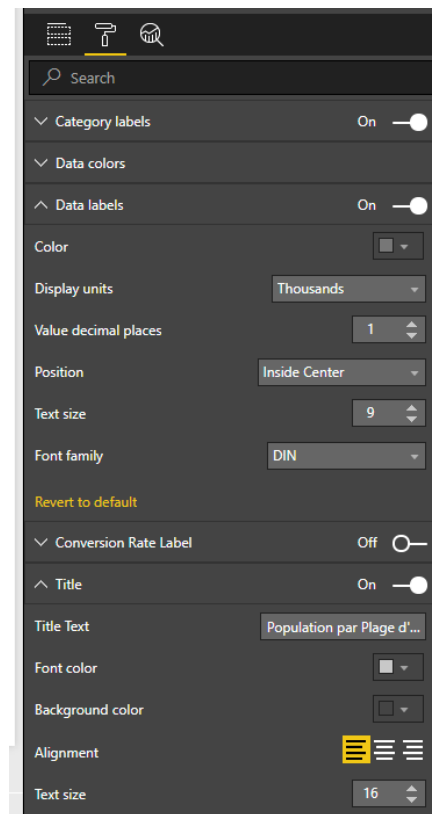
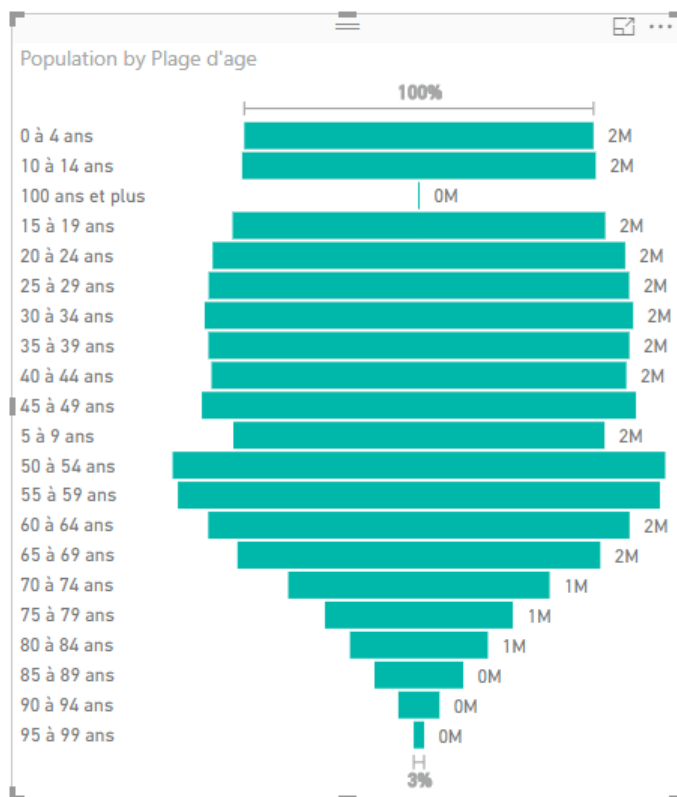
### Distribution de la population par tranche d'âge.

Nous allons maintenant créer notre premier composant visuel pour illustrer la distribution de la population par tranche d'âge. Le composant approprié est le diagramme entonnoir (*funnel*). Cliquer le bouton correspondant à ce composant. Une zone d'ancrage apparaît dans la feuille de travail; vous pouvez la déplacer et ajuster sa taille selon vos besoins.

Il suffit maintenant de glisser les éléments d'information dans les cases appropriées pour afficher les données dans le composant. Glisser d'abord l'élément **Population** de la table Population dans la case **Values** puis l'élément **Plage d'âge** de la table PlageAge dans la case **Group**.

Voilà votre premier composant visuel :





On peut maintenant ajuster certaines caractéristiques visuelles pour améliorer l'aspect du graphique. Cliquer sur le pinceau pour afficher les propriétés de mises en pages.

Dans la section **Data Labels**, choisir **Thousands** comme unités d'affichage avec 1 décimale, et **Inside Center** comme position.

Dans la section **Title**, inscrire le titre en français et ajuster la taille à 16 points.

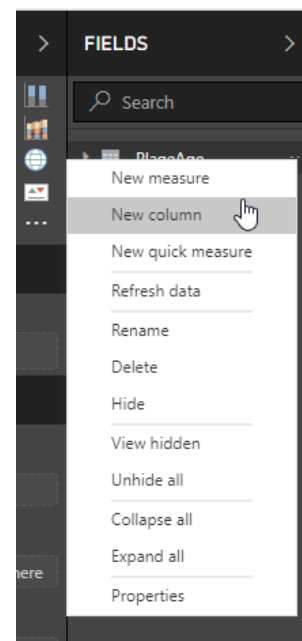
Mettre la section **Conversion Rate Label** à **off**

Il reste un problème à corriger : les données sont triées en ordre alphabétique des étiquettes de plage d'âge plutôt qu'en ordre numérique. Nous devons créer une nouvelle colonne qui permettra le tri dans le bon ordre. Plusieurs stratégies sont possibles. Pour cette démonstration, nous allons simplement ajouter des espaces devant l'étiquette.

Cliquer avec le bouton de droite sur le symbole ... à droite du nom de la table PlageAge et choisir la commande **New Column**.

Une zone d'édition s'ouvre sous le ruban de commande. Cette zone contient un début d'expression : **Column =**.

Remplacer ce texte par l'expression suivante :



```
PlageAgeTri = REPT(" ";4-find(" "; PlageAge[Plage d'age]))&PlageAge[Plage d'age]
```

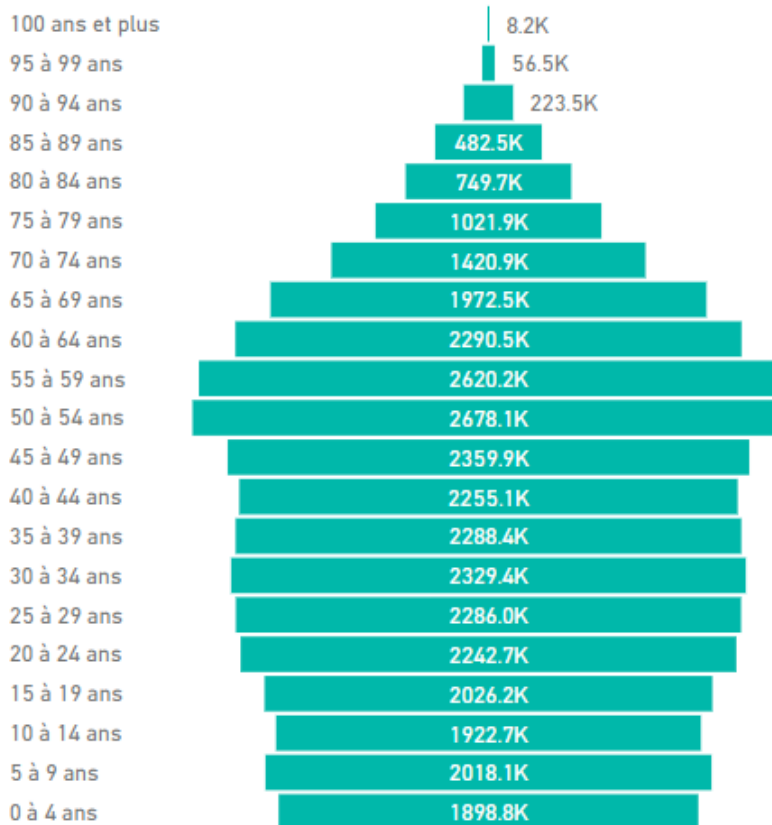
La fonction **FIND( )** retourne la position de la première occurrence d'une chaîne de caractère, un espace dans notre exemple.

La fonction **REPT( )** retourne une répétition d'une chaîne de caractère.

L'expression détermine donc le nombre d'espace à ajouter devant l'étiquette pour pouvoir les trier correctement.

Il faut maintenant remplacer Plage d'age par PlageAgeTri dans la case **Group** des propriétés du composant. Les données sont maintenant triées correctement. Si les données sont en ordre inverse, il suffit de cliquer sur ... dans la barre titre du graphique puis choisir de trier sur PlageAgeTri, ceci change l'ordre de croissant à décroissant.

## Population par Plage d'age



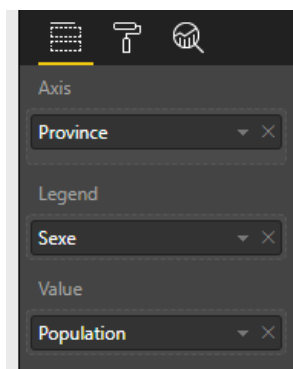
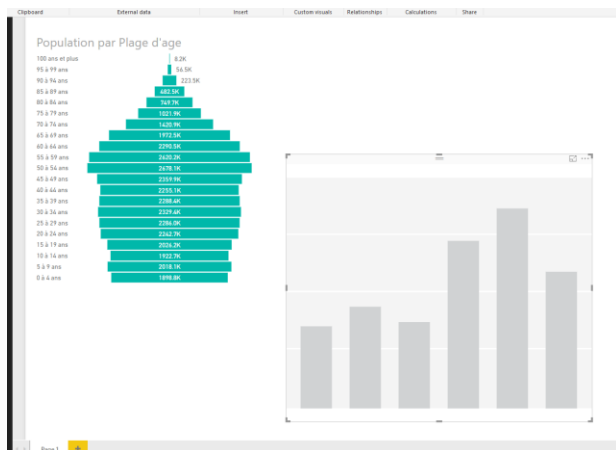
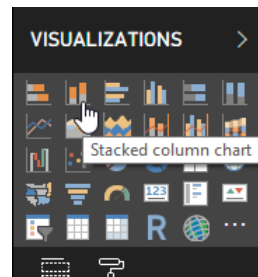
Voir ce site pour une comparaison avec les données mondiale :

<https://www.populationpyramid.net/>

## Distribution de la population par province.

Nous allons maintenant ajouter un composant visuel pour illustrer la distribution de la population des hommes et des femmes par province. Le composant approprié est le graphique par colonne empilée (*stacked column chart*).

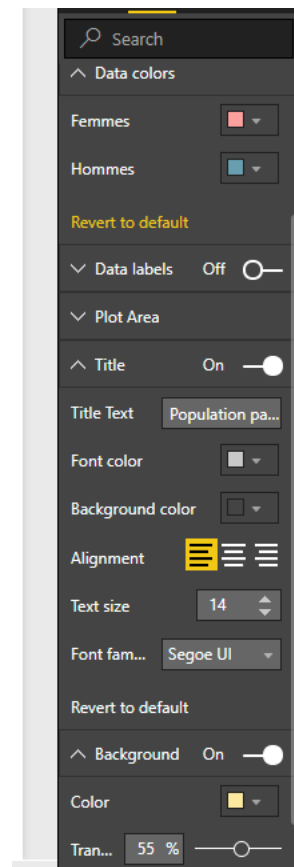
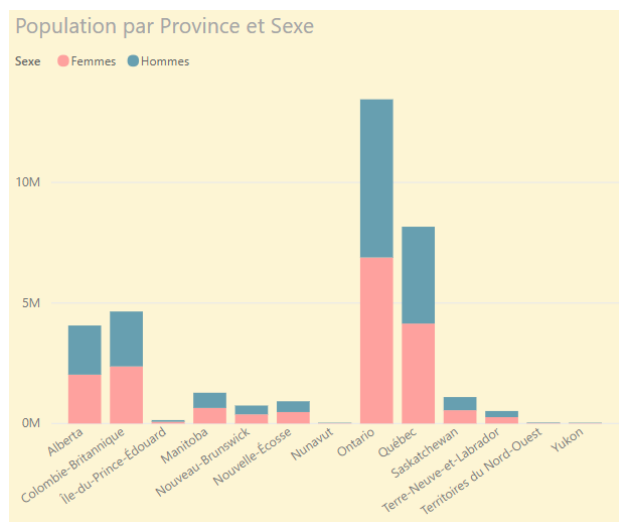
Assurez-vous qu'aucun contrôle de la page n'est sélectionné puis cliquer sur le bouton **Stacked column chart**. Une zone d'ancrage pour le graphique est ajoutée à la page. Déplacer et ajuster la taille de cette zone tel qu'illustré.



On doit maintenant compléter les propriétés de données pour le graphique :

- Glisser **Province[Province]** dans la case **Axis**
- Glisser **Population[Sexe]** dans la case **Legend**
- Glisser **Population[Population]** dans la case **Value**

Enfin, on peut ajuster la mise en page du graphique. Choisir des couleurs plus appropriées pour les données (hommes en bleu, femmes en rose), corriger le titre et sa taille et choisir une couleur de fond.



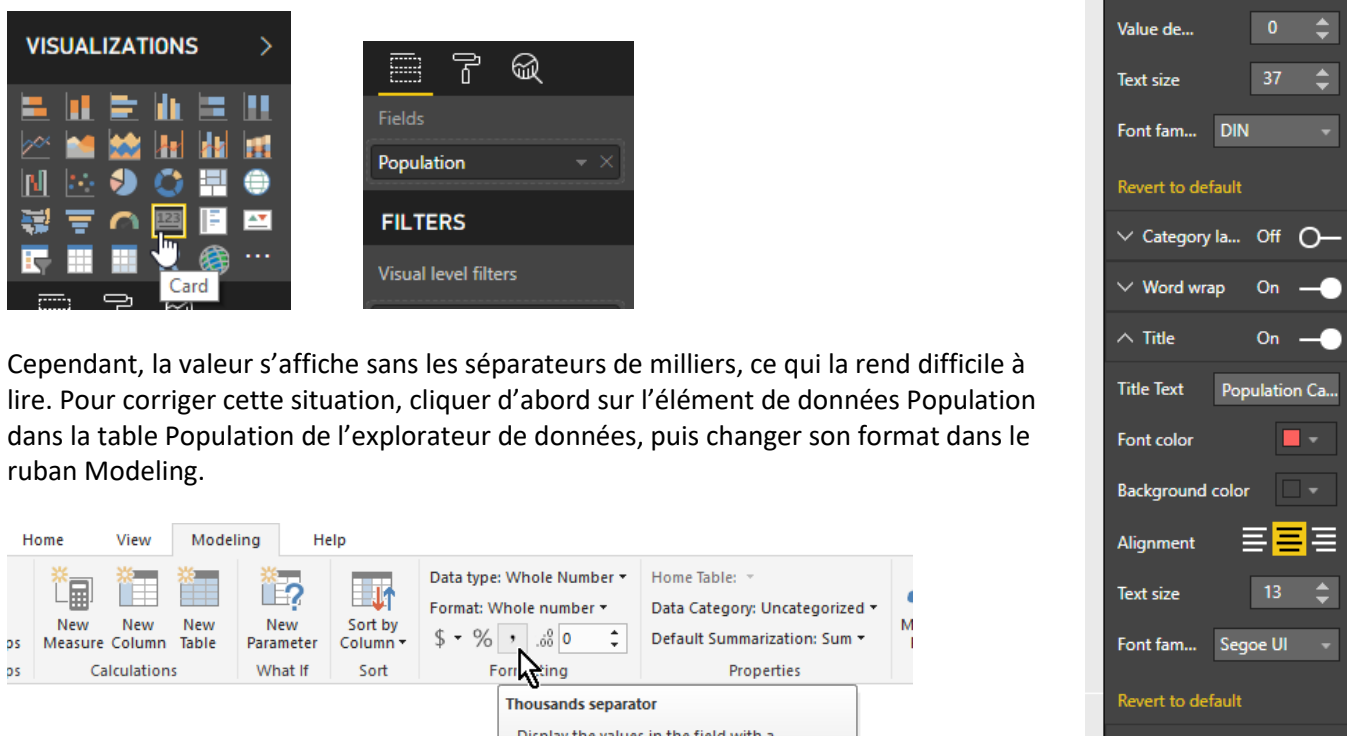


Ce serait intéressant de filtrer les deux graphiques sur le sexe. Il suffit d'ajouter un composant filtre (*slicer*). Assurez-vous qu'aucun contrôle de la page n'est sélectionné puis cliquer sur le bouton **Slicer**. Une zone d'ancrage pour le graphique est ajoutée à la page. Déplacer et ajuster la taille du composant, puis glisser **Population[Sexe]** dans la case **Field**.

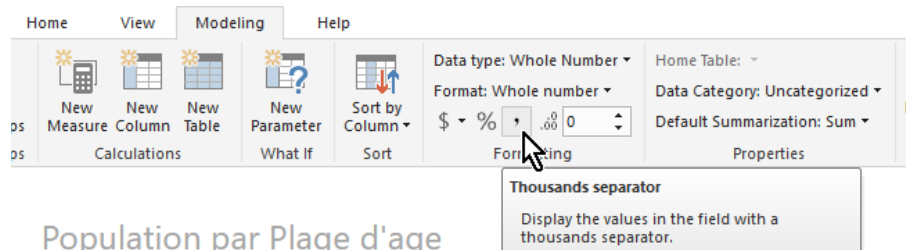


Explorer comment se comporte les graphiques lorsque vous interagissez avec eux et avec le filtre.

Nous allons maintenant ajouter un contrôle simple pour voir la population totale canadienne. Assurez-vous qu'aucun contrôle de la page n'est sélectionné puis cliquer sur le bouton **Card**. Encore une fois, une zone d'ancrage est ajoutée. Déplacer cette zone dans le haut de la page. Glisser **Population[Population]** dans la case **Fields**. La population totale est maintenant visible dans le composant. On peut ajuster la mise en page du composant.

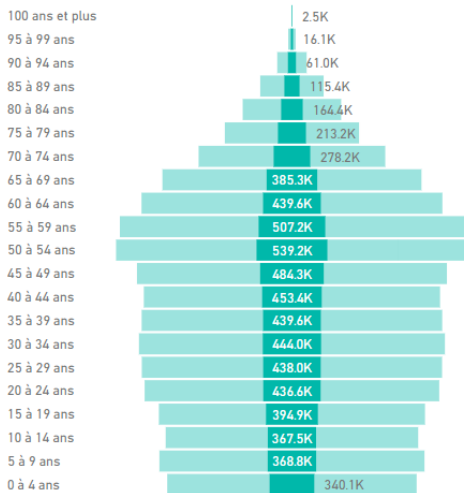


Cependant, la valeur s'affiche sans les séparateurs de milliers, ce qui la rend difficile à lire. Pour corriger cette situation, cliquer d'abord sur l'élément de données Population dans la table Population de l'explorateur de données, puis changer son format dans le ruban Modeling.



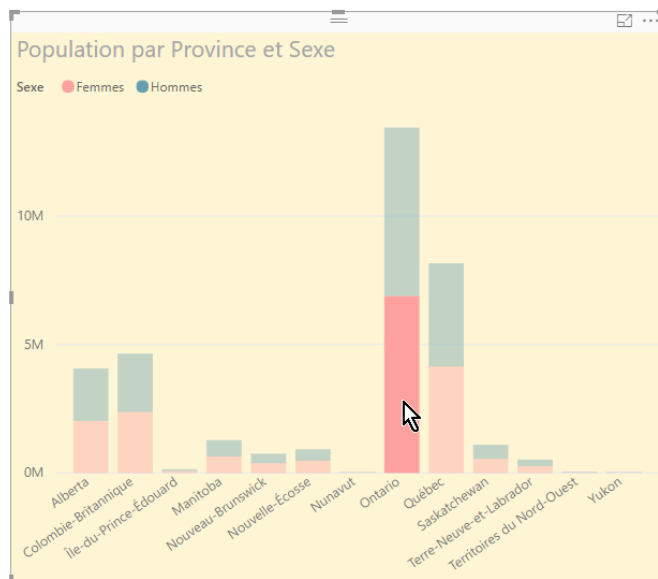
Nous observons cependant un problème. Lorsqu'on clique sur un des composants, la valeur affichée dans la carte change. Elle ne correspond plus à la population canadienne, mais plutôt au total de ce qui est choisie dans les composants.

Population par Plage d'âge



Population Canadienne Totale

6,889,115



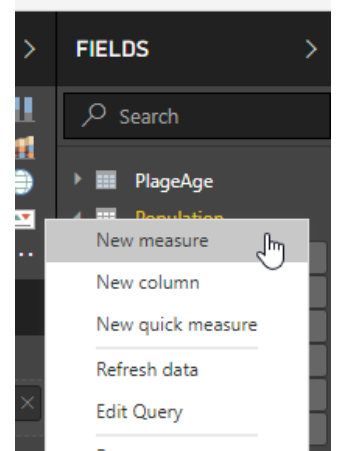
Il y a différentes façons de corriger cette situation, nous allons apprendre à faire une **mesure**.

## Création de la mesure PopulationCanadienne

Les colonnes calculées s'ajoutent comme nouvel élément de données sur chaque rangée de la table. Si on imagine la table comme une feuille de calcul Excel, une colonne calculée correspond à une formule que l'on copie sur toutes les lignes de la feuille de calcul.

Les mesures sont des formules d'agrégation de données. Elles peuvent être basées sur les éléments d'une ou plusieurs tables. On les insère habituellement dans la table qui contient les données sources principales de la mesure, mais ce n'est pas une obligation.

Pour ajouter une mesure, cliquer droit sur le nom de la table désirée, puis choisir **New measure**.



Une formule s’affiche alors dans la zone d’édition. Remplacer cette formule par celle-ci :

```
PopulationCanadienne = sum(Population[Population])
```

Glisser maintenant **PopulationCanadienne** dans la zone **Fields** de la fiche Population Canadienne Totale. Malheureusement, la valeur change toujours selon les choix fait dans les autres composants visuels.

## Filtre contextuel

Il faut comprendre que Power BI possède une notion de filtre contextuel et que toutes les valeurs affichées dans une page y sont assujetties par défaut. Nous devons donc modifier la formule de notre mesure de manière à ignorer le filtre contextuel de la page. Pour ce faire, nous allons utiliser la fonction **Calculate()**. Cette fonction prend deux paramètres : la fonction de calcul de la mesure, et le filtre à appliquer. Modifier la formule de **PopulationCanadienne** pour celle-ci :

```
PopulationCanadienne = CALCULATE(sum(Population[Population]);ALL(Population))
```

Vous aurez devinez que la fonction **ALL()** enlève tous les filtres appliqués sur la table mentionnée, soit Population dans ce cas. Cette mesure retourne donc la somme de Population pour toutes les lignes de la table.

Après avoir appliquer le changement, remarquer que la valeur de la population canadienne totale ne varie plus lorsqu’on filtre les autres composants.

## Visualisation de l’âge moyen

Nous voulons maintenant afficher l’âge moyen par province. Pour calculer l’âge moyen avec précision, nous devrions faire la somme de l’âge de tous les résidents de la province divisé par le nombre de résident dans la province. Nous n’avons pas les données pour faire ce calcul. Cependant, il est possible d’avoir une approximation de l’âge moyen à partir du nombre de résident par tranche d’âge, en leur attribuant l’âge médian de la tranche d’âge.

$$\frac{\sum (Population \times AgeMedian)}{\sum Population}$$

Il faut procéder par étape. La première étape est d’ajouter une nouvelle colonne dans la table **PlageAge** pour calculer l’âge médian. Cliquer droit sur le nom de la table, choisir **New column** puis inscrire cette formule :

```
AgeMedian = VALUE(left(PlageAge[Plage d'age];FIND(" ";PlageAge[Plage d'age] )-1))+2
```

Cette formule extrait le premier « mot » de la chaîne **Plage d’âge**, le convertit en valeur numérique et lui additionne 2 pour trouver l’âge médian sur des tranches de 5 ans.

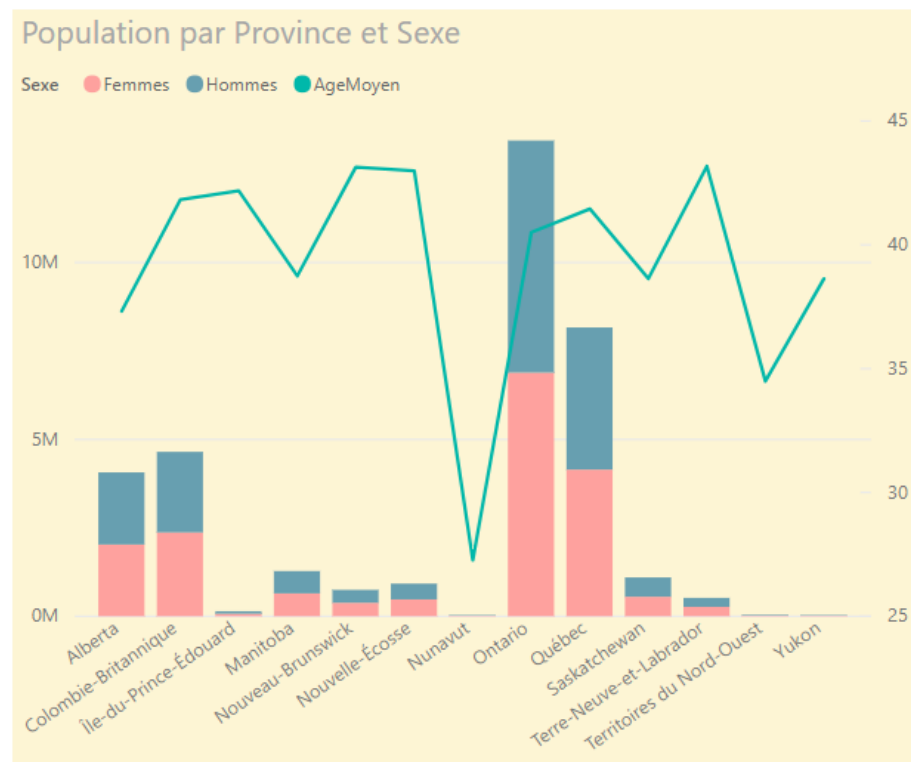
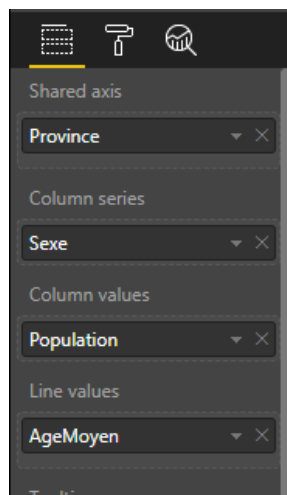
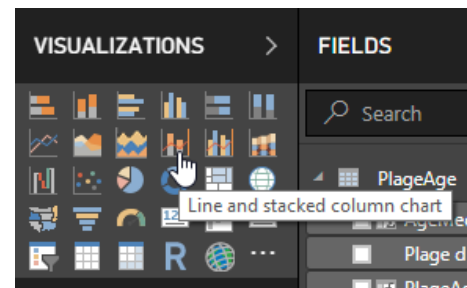
Puisque la fonction **SUM()** n'accepte pas d'expression, nous devons ajouter une colonne pour calculer le numérateur de la formule. Cliquer droit sur le nom de la table **Population** et choisir **New column**. Inscrire la formule suivante. Cette formule utilise la fonction **Related()** pour aller chercher la valeur de **ageMedian** dans la table liée **plageAge**.

```
PopulationAgeMedian = Population[Population] * RELATED(plageAge[ageMedian])
```

Nous pouvons maintenant calculer l'âge moyen. Cliquer droit sur le nom de la table **Population** puis choisir **New measure**. Inscrire la formule suivante pour cette mesure.

```
AgeMoyen = sum(Population[PopulationAgeMedian]) / sum(Population[Population])
```

Pour voir l'âge moyen par province, nous allons modifier le graphique de la population par province pour y ajouter une ligne. Choisir le graphique dans la zone de travail, puis cliquer sur le bouton du composant **Line and Stacked column chart**. Ceci modifie le composant tout en conservant les valeurs des propriétés communes. Il suffit maintenant de glisser **Population[AgeMoyen]** dans la case Line values pour voir le résultat.



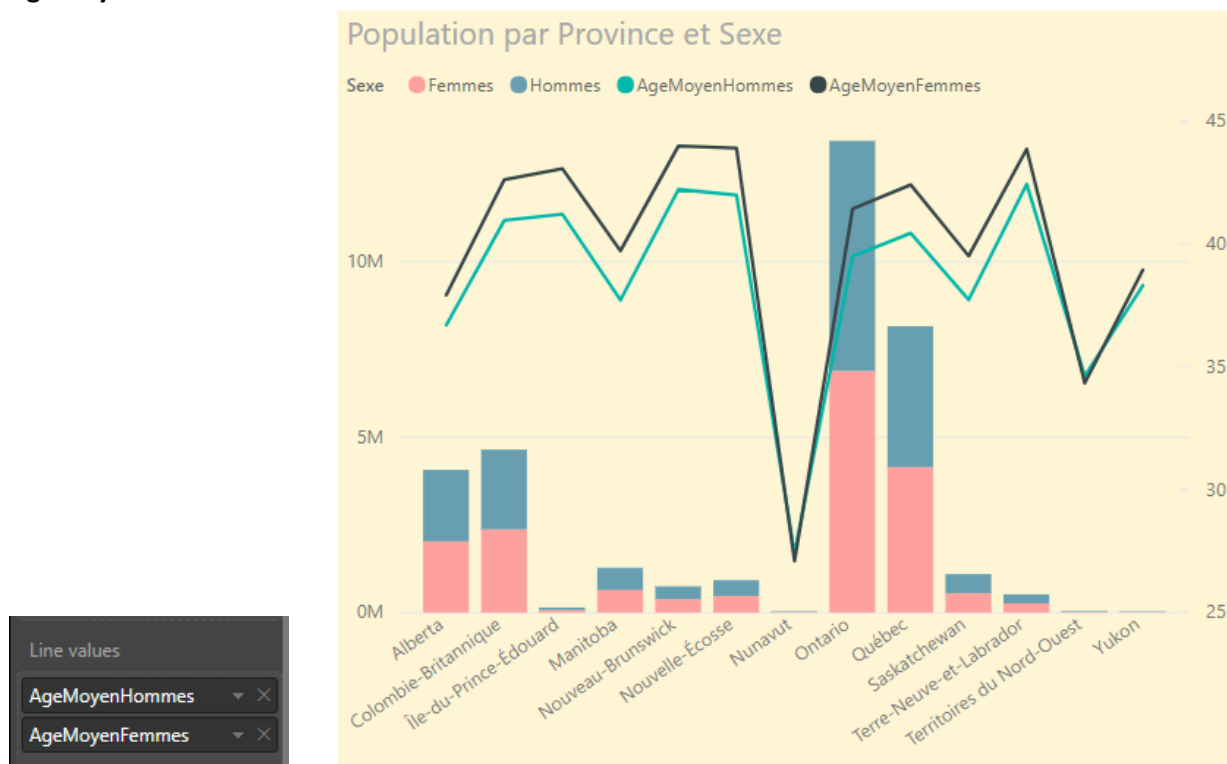
Il serait intéressant de distinguer l'âge moyen des femmes et des hommes. Nous avons tous les éléments de base requis pour créer deux nouvelles mesures :

```
AgeMoyenFemmes = CALCULATE(sum(Population[PopulationAgeMedian])/sum(Population[Population]);
Population[Sexe]="Femmes")
```

```
AgeMoyenHommes = calculate(sum(Population[PopulationAgeMedian])/sum(Population[Population]);
Population[Sexe]="Hommes")
```

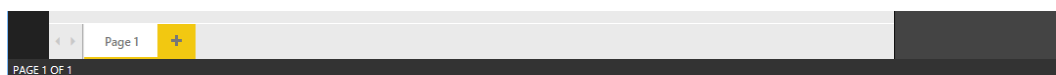
*Truc : pour les formules qui sont longue, cliquer sur le chevron à l'extrême droite de la zone d'édition pour avoir plus de lignes. Vous pouvez utiliser les touches alt-enter pour insérer un changement de ligne.*

Dans les propriétés de données du graphique, remplacer **AgeMoyen** par **AgeMoyenFemmes** et **AgeMoyenHommes**.

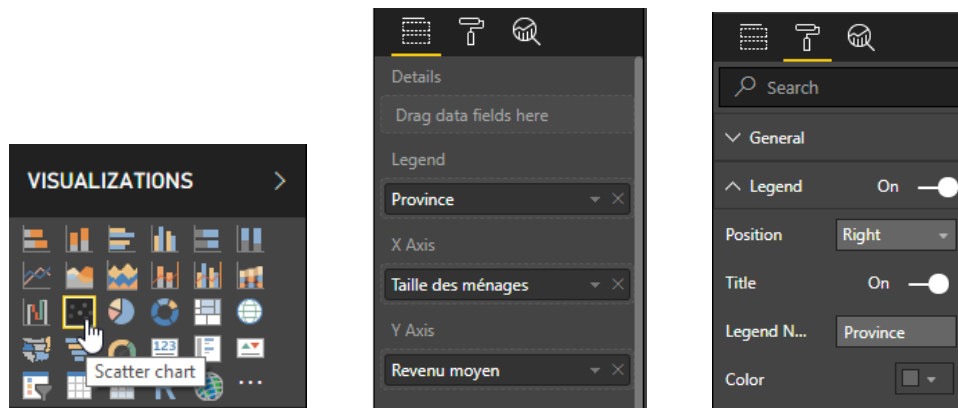


## Analyse du revenu moyen par ménage selon la taille du ménage

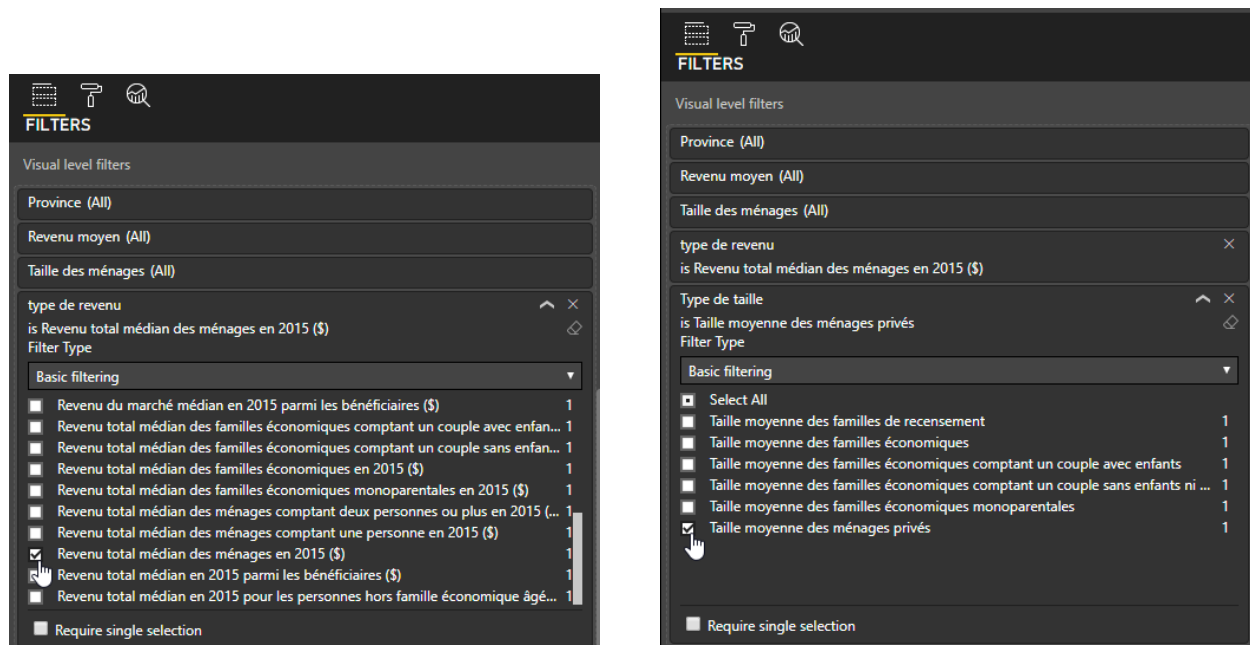
Pour terminer cet atelier, nous allons utiliser un graphique de dispersion pour illustrer la relation entre la taille moyenne et le revenu moyen des ménages par province. Cliquer sur le + à droite de l'onglet Page 1 pour ajouter une nouvelle page à l'environnement de travail.



Cliquer sur le bouton **Scatter chart** pour ajouter ce composant. Vous pouvez ajuster la position et la taille du composant pour occuper l'espace de la page. Glisser **Province**[Province] dans la case **Legend**. Glisser **Taille**[Total] dans la case **X Axis** puis renommer **Taille des ménages**. Glisser **Revenu**[Total] dans la case **Y Axis**, puis renommer à **Revenu Moyen**. Comme il y a plusieurs provinces, il est difficile de voir la légende dans le haut du composant. Nous allons donc la déplacer à la droite en choisissant **Right** dans la case **Position** du groupe **Legend** de l'onglet des propriétés de mise en page.

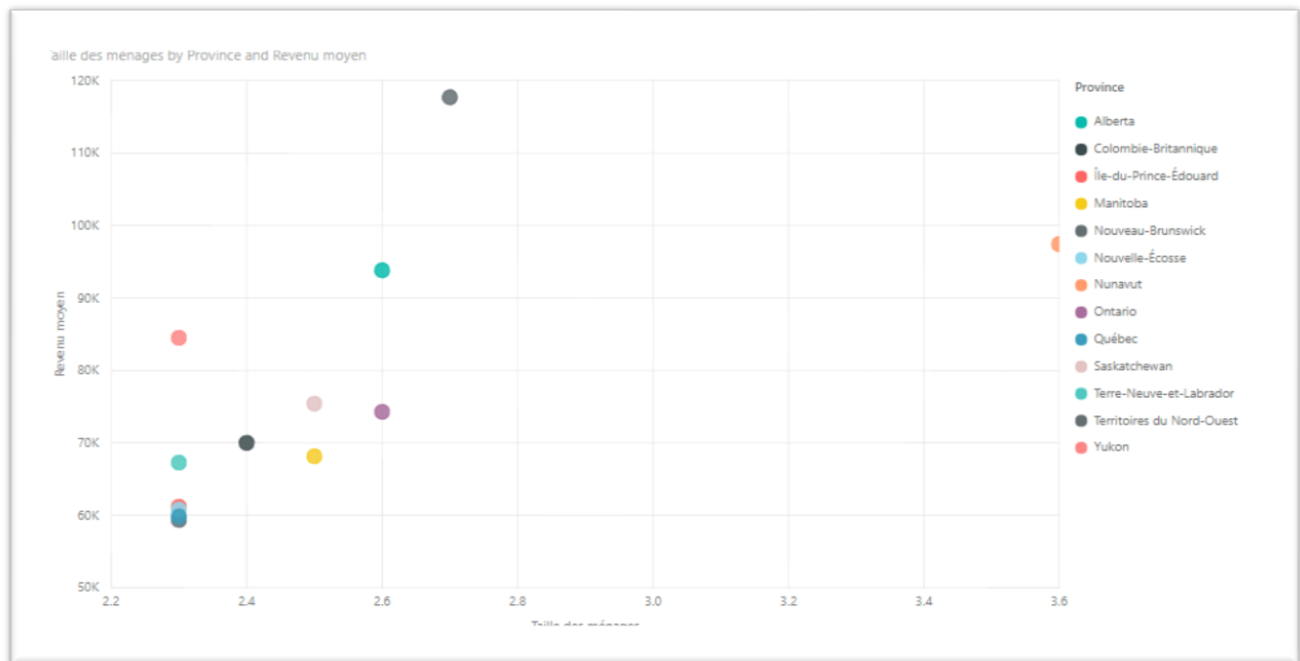


Il reste une étape importante. Les tables Revenu et Taille comporte en fait plusieurs mesures. En ce moment, le graphique affiche le total de ces mesures. Nous devons donc ajouter un filtre explicite dans le composant. Glisser **TypeRevenu**[Type de revenu] dans la case **Visual level filters** puis choisir la valeur *Revenu total médian des ménages en 2015*. Glisser **TypeTaille**[Type de taille] dans la même case et choisir *Taille moyenne des ménages privés*.



*Truc : élargissez la zone des propriétés pour faciliter la recherche des valeurs. Vous pouvez aussi cliquer sur le chevron à côté du nom de l'élément de données pour fermer ou ouvrir le détail.*

Voici le résultat que vous devriez avoir :



Le graphique de dispersion possède d'autres propriétés de données qui permettent d'ajuster la taille des pastilles (**Size**) ou l'intensité de leur couleur (**Color saturation**) en fonction d'une mesure. La propriété **Play axis** est très utile pour animer le graphique en fonction d'une dimension temporelle.

Ces présentations TED utilisent ces techniques pour illustrer l'évolution de mesures sur la démographie :

<https://www.youtube.com/watch?v=jbkSRLYSojo>

[https://www.ted.com/talks/hans\\_rosling\\_shows\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen?language=fr](https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?language=fr)

[https://www.ted.com/talks/hans\\_rosling\\_on\\_global\\_population\\_growth?language=fr](https://www.ted.com/talks/hans_rosling_on_global_population_growth?language=fr)

## Autres sources d'informations sur Power BI

Voici quelques liens pour compléter votre apprentissage sur Power BI (il en existe plusieurs autres...)

<https://powerbi.microsoft.com/fr-fr/>

Site principal de Power BI où vous trouverez des outils de formation et la communauté des utilisateurs.

<https://www.sqlbi.com/> (en anglais seulement)

Société italienne spécialisée en Power BI qui offre des vidéos de formation sur leur site Web, dont plusieurs sont gratuits, en plus d'articles sur des trucs et astuces.

<https://www.edx.org/course/analyzing-visualizing-data-power-bi-microsoft-dat207x-8> (en anglais)

Cours en ligne sur l'utilisation de Power BI