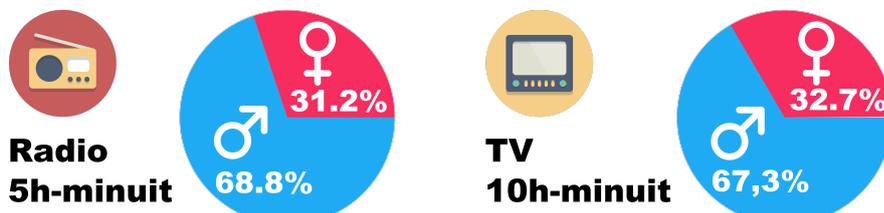


Traitement des données : temps de parole des femmes à la télé et à la radio



- À partir du site www.data.gouv.fr, rechercher puis télécharger les données sur le « **Temps de parole des hommes et des femmes à la télévision et à la radio** ». Plusieurs fichiers sont disponibles. Choisir celui sur les « **Moyennes par heure, année et chaîne** ». Il s'agit d'un fichier CSV.
- Ouvrir le fichier avec **LibreOffice Calc** (éventuellement à l'aide d'un clic droit).
Dans la fenêtre de dialogue d'ouverture du fichier :
 - Vérifier que le **Jeu de caractères** est bien **Unicode (UTF-8)** ;
 - Vérifier que **Virgule** (et rien d'autre) est sélectionné comme **séparateur** ;
 - Pour les colonnes contenant des nombres (les trois dernières), on observe que ceux-ci ont pour séparateur de décimale un point au lieu d'une virgule.
Changer le **Type de colonne** en « **Anglais US** » pour chacune d'elles (on doit le faire l'une après l'autre). Attention à bien afficher toutes les colonnes de droite, par exemple à l'aide de l'ascenseur horizontal.
 - Cliquer ensuite sur **OK**.
 Une fois le fichier ouvert, vérifier que les nombres sont bien écrits avec des virgules dans les colonnes **F**, **G** et **H**, sinon recommencer l'étape précédente en fermant puis en ouvrant à nouveau le fichier.
- Fixer la première ligne (**Affichage > Fixer des cellules > Fixer la première ligne**).
- Mettre en place le filtre dans le tableau (**Données > AutoFiltre**).
- Enregistrer le fichier dans vos documents. Attention à bien choisir le format ODF (.ods).
On pourra par exemple l'enregistrer sous le nom **parole.ods**.
- Répondre aux questions suivantes dans des cellules des premières lignes, à droite du tableau.
 - À quelle heure et sur quelle chaîne le taux d'expression des femmes a-t-il fait ses 3 meilleurs scores ?
 - Ses 3 plus mauvais scores ?
- Un exemple de réutilisation de cette table de données est présenté dans [cette page de la revue des medias](#) du site de l'INA. En particulier, l'auteur écrit :

L'analyse massive des fonds collectés de 2010 à 2018 dresse un état des lieux caractérisé par un fort déséquilibre entre le temps de parole utilisé par les hommes et les femmes. Les prises de parole des femmes à la télévision représentent moins d'un tiers du temps de parole alloué (32,7 %). Ce constat est encore plus frappant à la radio où il n'est que de 31,2 %, comme le montre la figure ci-dessous.



Source: David Doukhan, "À la radio et à la télé, les femmes parlent deux fois moins que les hommes", INA, mars 2019

On souhaite retrouver les pourcentages de temps de parole pour les femmes à la radio et à la télévision entre 2010 et 2018.

 Faire la moyenne des pourcentages de temps de parole des femmes n'a pas de sens (Pourquoi?). Il va donc falloir faire quelques manipulations.

- (a) Faire un clic droit sur la colonne **H** et choisir **Insérer des colonnes avant**.

Dans la cellule en tête de la colonne obtenue, écrire le titre « **Heures de parole** ».

Écrire dans la cellule juste en dessous une formule utilisant les valeurs de ses voisines de la colonne du nombre d'heures analysées (`nb_hours_analyzed`) et de celle de la proportion de parole (`speech_rate`, qui est en pourcentage) et qui donne effectivement le nombre d'heures de parole.

- (b) Insérer de même encore une colonne à gauche de la colonne **G** nouvellement créée.

Ajouter le titre « **Heures de paroles des femmes** » à cette colonne.

Écrire une formule dans la cellule en dessous qui calcule ce nombre d'heures à l'aide des valeurs la colonne des heures de paroles et de celle de la proportion de parole des femmes (`women_expression_rate`, elle aussi en pourcentage).

- (c) Sélectionner exactement les deux cellules contenant les formules écrites dans les deux questions précédentes, puis double-cliquer dans le coin inférieur droit pour étirer ces formules tout en bas du tableau.

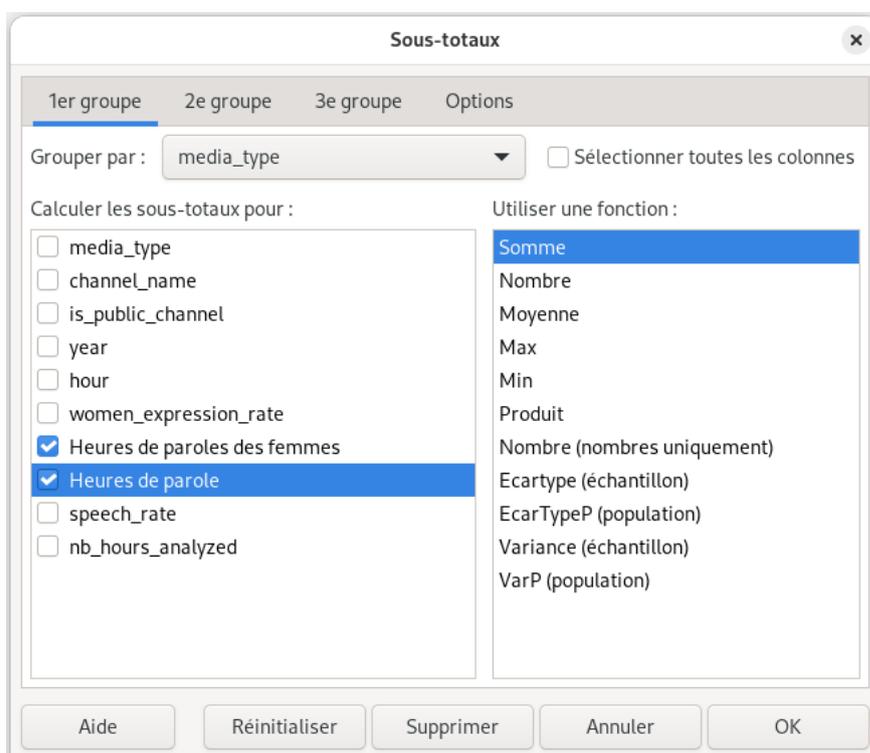
- (d) À l'aide de l'AutoFiltre, filtrer les années pour n'avoir que celles de 2010 à 2018.

- (e) Aller dans **Données > Sous-totaux...**

Uniquement dans l'onglet **1er groupe**, grouper par `media_type`.

Cocher les cases « **Heures de paroles des femmes** » et « **Heures de parole** ».

Vérifier que pour les deux, c'est bien la fonction **Somme** qui est sélectionnée :



Puis cliquer sur **OK**.

- (f) À gauche est apparue une colonne en dehors de la feuille avec des boutons carrés numérotés de **1** à **3**.

Cliquer sur le bouton **2** pour réduire l'affichage aux sommes calculées pour les deux médias, radio et télévision, ainsi qu'à la somme totale.

	A	B	C	D	E	F	G	H
1	media_tyt	channel_name	is_public_chann	yes	hol	women_expression_rate	Heures de	Heures de
6987	radio	Résultat					12596.56	39633.24
11510	tv	Résultat					20775.72	68446.11
11511		Grande somme					33372.28	108079.35

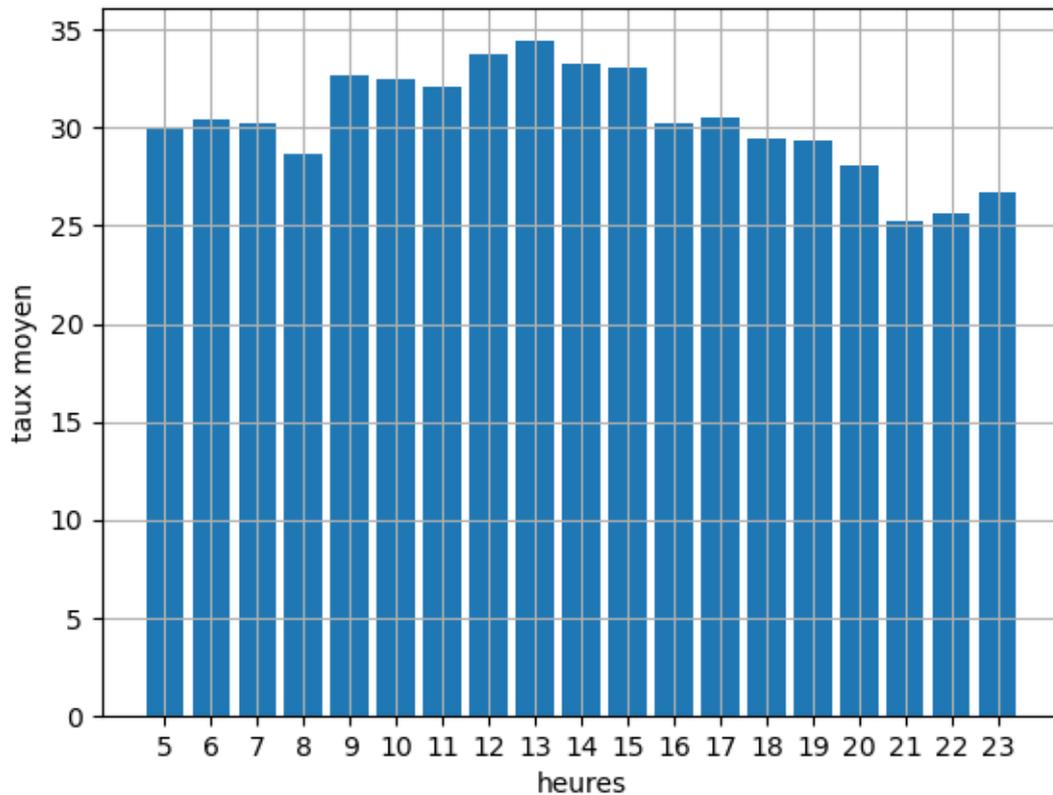
(g) Écrire dans la cellule de la ligne du total pour la radio et d'une colonne à droite du tableau une formule qui permet de donner le taux de parole des femmes en pourcentage.

Étirer cette formule vers le bas pour avoir les taux pour la télévision et pour l'ensemble.

On observera des différences avec les valeurs du site données plus haut.

8. On souhaite obtenir un diagramme en barre représentant le temps d'expression moyen des femmes en fonction de l'heure de la journée, tous médias confondus.

Le graphique doit pouvoir ressembler à ceci :



On souhaite utiliser le langage de programmation Python pour le faire.

On utilisera pour cela le fichier disponible sur Moodle.

Suivre les instructions du fichier, données en commentaires.

9. (Question bonus)

Si nous voulions obtenir ce même diagramme avec le tableur, comment le ferions-nous ?

Envoyer le fichier .ods contenant le diagramme et les données permettant de l'obtenir.