

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

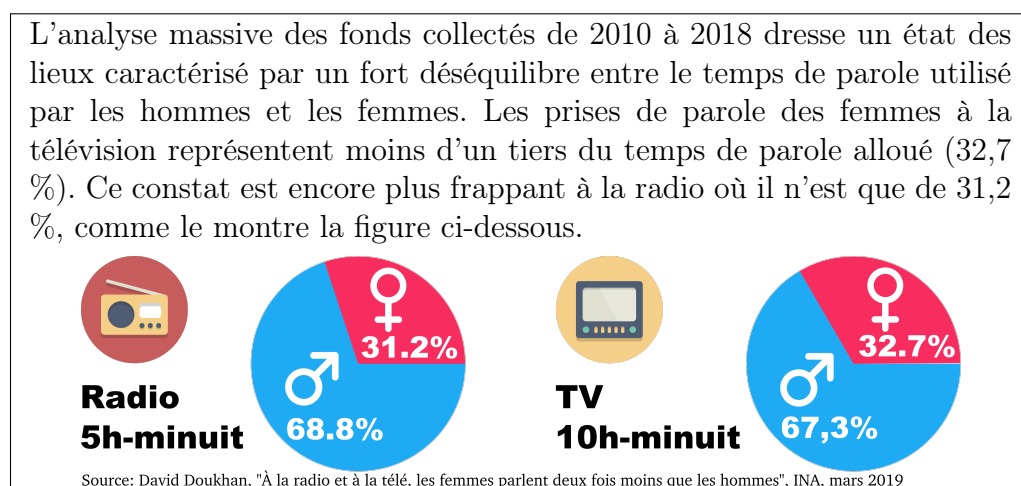


1. À partir du site www.data.gouv.fr, télécharger des données ouvertes sur le « **Temps de parole des hommes et des femmes à la télévision et à la radio** ». Plusieurs fichiers sont disponibles. Choisir celui sur les « **Moyennes par heure, année et chaîne** ». Il s'agit d'un fichier CVS.
2. Ouvrir le fichier avec LibreOffice Calc. Dans la fenêtre de dialogue :
 - vérifier que la virgule est sélectionnée 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 ces colonnes pour « Anglais US ».
 - Cliquer ensuite sur OK.



Par la suite, si vous enregistrez votre fichier, ne pas écraser le fichier .csv initial, mais enregistrer sous un nouveau fichier (au format ODF, autrement dit .ods de préférence).

3. 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 ?
Pour répondre à cette question on pourra utiliser l'**AutoFiltre** obtenu à partir du menu « Données ».
4. 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 :



- (a) Retrouve-t-on (à peu près) les données présentées par l'auteur sur le temps de parole moyen des femmes à la télévision et à la radio ?

Pour cela, on fera dans un premier temps attention à filtrer les données en ne gardant que les années 2010 à 2018.

Pour obtenir des données regroupées (et des calculs de sommes ou de moyennes), on pourra également utiliser la fonctionnalité des **Sous-totaux** du menu « Données ».

- (b) En fait, si c'est ce que vous avez fait à la question précédente, faire la moyenne des taux d'expressions des femmes n'a pas de sens ; Il faut absolument prendre en compte les nombres d'heures analysées et la proportion de parole dans ces heures. Ainsi, pour ce calcul de moyenne de la proportion du temps de parole des femmes, il faut calculer la proportion du temps de parole des femmes sur le temps total de parole. Pour cela :
- Insérer une colonne « Heures de paroles » et écrire dans celle-ci une formule (à étirer) dépendant des valeurs de la colonne du nombre d'heures analysées et de celles de la proportion de parole.
 - Insérer une colonne « Heures de paroles des femmes » et écrire dans celle-ci une formule dépendant des valeurs de la colonne « Heures de paroles » précédente et de celles de la proportion de parole des femmes.

Avec les sous-totaux, on peut alors faire la somme sur ces deux colonnes.

Obtenir alors finalement les moyennes pour la radio et la télé en faisant les calculs dans des cellules situées sous le tableau.

On observera des différences avec les simples (mais incorrectes) moyennes des fréquences.

- (c) Si nous voulions représenter le temps d'expression moyen des femmes en fonction de l'heure de la journée, comment le ferions-nous avec le tableau ?

5. On souhaite obtenir le graphique évoqué précédemment à l'aide de Python, pour automatiser davantage les choses. On utilisera pour cela le fichier disponible sur Moodle. Suivre les instructions du fichier, données en commentaires.

Le graphique doit pouvoir ressembler à ceci :

