

L'administration française rend disponible publiquement de plus en plus de données via le site web [data.gouv.fr](http://data.gouv.fr). Vous allez travailler sur certains de ces jeux de données pour manipuler les filtres shells via les tubes.

Pour mémoire voici une liste de filtres intéressants (qui ne dispense pas de se référer aux pages correspondantes dans le manuel, bien au contraire) :

<b>cat</b>	retourne les lignes lues sans modification.
<b>cut</b>	ne retourne que certaines parties de chaque lignes lues.
<b>grep</b>	retourne uniquement les lignes lues qui correspondent à un modèle particulier ou qui contiennent un mot précis.
<b>head</b>	retourne les premières lignes lues.
<b>sort</b>	trie les lignes lues.
<b>tail</b>	retourne les dernières lignes lues.
<b>tee</b>	envoie les données lues sur la sortie standard ET dans un fichier passé en paramètre.
<b>tr</b>	remplace des caractères lus par d'autres.
<b>uniq</b>	supprime les lignes identiques.
<b>wc</b>	retourne le nombre de caractères, mots et lignes lus.
<b>sed</b>	édite le texte lu (requêtes <b>ed</b> comme avec la commande <b>ed</b> ).

### **Exercice 1 : Recherche de données dans l'annuaire de l'éducation**

Le jeu de données utilisé dans cet exercice est l'[annuaire de l'éducation](#).

Le fichier **annuaire.csv** contient ces données. C'est un fichier dont chaque ligne contient un *enregistrement* : les données pour un établissement d'éducation. Chaque ligne est structurée de manière identique en plusieurs colonnes. La première ligne du fichier donne le nom de chaque colonne. Les colonnes sont séparées les unes des autres par le caractère ;.

Le fichier **annuaire.csv.gz** est une version comprimée du fichier **annuaire.csv**. Pour afficher son contenu sur la sortie standard il suffit d'utiliser la commande `zcat annuaire.csv.gz`.

**Q 1.** Copiez le fichier `/home/public/m1101/tp/tubes/annuaire.csv.gz` dans votre répertoire `~/asr/tubes` (que vous devez donc créer d'abord).

**Q 2.** Exécutez la commande suivante dans votre terminal :

```
source /home/public/m1101/tp/tubes.env
```

**Q 3.** À l'aide d'une ligne de commande utilisant des tubes, filtres et redirection, créez un fichier **annuaire.colonnes** contenant une ligne numérotée pour chaque entête de colonne du fichier **annuaire.csv**.

**Q 4.** Quelle ligne de commande permet d'afficher toutes les lignes du fichier **annuaire.csv** sauf la première.

**Q 5.** Créez une commande shell **annuaire** qui affiche toutes les lignes du fichier **annuaire.csv** sauf la première. Pour cela il vous suffit de mettre les commandes qui vous semblent nécessaire dans le fichier **annuaire** (avec l'éditeur de textes de votre choix) puis d'exécuter la commande suivante :

```
chmod u+x annuaire
```

**Q 6.** Utilisez cette commande connectée à d'autres filtres via des tubes pour effectuer les actions suivantes :

1. afficher le **nombre d'établissements**; 66 456
2. afficher les valeurs utilisées pour la colonne **Statut\_public\_privé**;  
 Les valeurs à trouver sont :
  - une valeur vide
  - -
  - Public
  - Privé
3. afficher le **nombre d'établissements publics**; 53 665
4. afficher le **nombre de Lycée**; 5 714
5. afficher le **nombre de lycées publics**; 3 478

- |  |        |
|--|--------|
| 6. afficher le <b>nombre de noms différents</b> parmi tous les établissements ;                                    | 37 114 |
| 7. afficher le <b>nombre d'établissement</b> portant le nom de BRASSENS ;  | 134    |
| 8. afficher l' <b>identifiant du lycée dans lequel vous avez préparé votre terminale</b> ;                         |        |
| 9. afficher le <b>nombre de lycée publics dans le nord</b> (département de code 59) ;                              | 108    |
| 10. afficher le nombre de lycée publics dans le nord (département de code 59) et leur nom dans un second terminal. |        |

## Exercice 2 : Les prénoms en France

Le jeu de données utilisé dans cet exercice est [le fichier des prénoms de 1900 à 2019](#).

Le fichier `prenoms` est un fichier qui contient le nombre de naissance déclaré par prénom et département pour toutes les années entre 1900 et 2019. Chaque ligne est structurée de manière identique en plusieurs colonnes. La première ligne du fichier donne le nom de chaque colonne. Les colonnes sont séparées les unes des autres par le caractère `:`. Pour la colonne du sexe, 1 correspond aux garçons et 2 aux filles.

Le fichier `prenoms.gz` est une version comprimée du fichier `prenoms`. Pour afficher son contenu sur la sortie standard il suffit d'utiliser la commande `zcat prenoms.gz`.

**Q 1.** Copiez le fichier `/home/public/m1101/tp/tubes/prenoms.gz` dans votre répertoire `~/asr/tubes`.

**Q 2.** Exécutez la commande suivante dans votre terminal :

```
source /home/public/m1101/tp/tubes.env
```

**Q 3.** Créez une commande shell `prenoms` qui affiche toutes les lignes du fichier `prenoms` sauf la première.

**Q 4.** À l'aide d'une ligne de commandes utilisant des tubes, filtres et redirection, créez un fichier `prenom.colonnes` contenant une ligne numérotée pour chaque entête de colonne du fichier `prenoms`.

**Q 5.** Combien le fichier `prenoms` contient-il de lignes (sans compter la première) ? 3 676 682

**Q 6.** Connectez des filtres via des tubes pour effectuer les actions suivantes :

- |   |        |
|---|--------|
| 1. afficher le nombre de prénom de filles différents ;  | 19 784 |
| 2. afficher le nombre de prénom de garçons différents en 2000 ;   | 1 521  |
| 3. afficher la liste des prénoms de filles différents en 2000 dans le département du nord (59) et du pas de calais (62) ; |        |
| 4. afficher le prénom utilisé dans le plus de département pour les garçons en 1900.                                       | PAUL   |

**Q 7.** En utilisant le filtre `somme`, qui fait la somme des nombres lus sur l'entrée standard, trouver la ligne de commande permettant d'afficher le nombre d'enfants déclarés sous le nom `GAEL` dans le nord (59) avant 2010. 1 008