

# Projet Scin1B 2019

Si vous avez des questions, mail [stephanie.bricq@u-bourgogne.fr](mailto:stephanie.bricq@u-bourgogne.fr)

Projet à réaliser en binôme.

Vous devez rendre un rapport (max 10 pages) contenant les éléments suivants

- Explications de votre code,
- Images des résultats obtenus
- Manuel utilisateur indiquant les commandes à utiliser pour exécuter vos programmes

Rapport papier à rendre lors de la démonstration de projet (créneau de TD de la semaine du 16 décembre)

Vous devez également rendre le code de votre projet lors de la démonstration.

La démonstration aura lieu sur le créneau du dernier TD (semaine du 16 décembre) et durera 10 minutes par binôme. Les salles et créneaux horaires vous seront communiqués ultérieurement. Vous devez être présents 10 minutes avant le début de votre passage pour vous installer.

Les données sur lesquelles vous allez travailler sont issues d'un jeu de données du Metropolitan Museum of Art<sup>1</sup> dont on a gardé seulement une partie des champs.

Les différents champs sont séparés par des **tabulations**.

## Champs

1. Object number : numéro de l'objet (peut contenir toutes sortes de caractères)
2. Is public domain : indique si l'objet est du domaine public ou non (valeur : true ou false)
3. Object ID : identifiant de l'objet (nombre unique)
4. Department : département d'affectation
5. Object Name : nom de l'objet
6. Title : description de l'objet
7. Artist Display Name : nom de l'artiste
8. Object Date : date de l'objet (date ou intervalle, ou indication de siècle, n.d. signifie no date)
9. Medium : matière
10. Dimensions : dimensions de l'objet
11. Country : pays d'origine
12. Classification : catégorie de classement (il peut y avoir plusieurs catégories séparées par | )

Certains champs peuvent être vides si les données ne sont pas connues.

Un jeu de données `donnees_test1` vous est fourni.

Un autre jeu de données `donnees_test2` vous sera fourni le jour de la présentation. Vos programmes devront également fonctionner sur ce jeu de test.

---

<sup>1</sup> <https://github.com/metmuseum/openaccess/>

- 1) Dans un premier temps, on souhaite garder uniquement les lignes pour lesquelles l'identifiant de l'objet (Object ID) est un nombre. Créer un programme prenant en paramètre le fichier de données. Ce programme doit créer un fichier de sortie *nomFichier\_valide* contenant uniquement les lignes valides. Il devra également afficher sur le terminal :
  - Le nombre de lignes valides
  - Le nombre d'objets du domaine public (parmi les lignes valides)
  - Le nombre d'objets dont la date est connue (parmi les lignes valides)
  - Le nombre d'objets dont le nom (Object name) commence par un P (parmi les lignes valides)

Dans la suite on travaillera sur le fichier *nomFichier\_valide* créé à la question 1 qui sera passé en paramètre de chaque programme.

- 2) On souhaite ensuite regrouper dans un fichier *nomFichier\_sans\_dim* tous les objets dont la dimension n'est pas indiquée ou n'est pas disponible. Votre programme devra également afficher le nombre d'objets sans dimension trouvé.
- 3) On souhaite ensuite créer un fichier *nomFichier\_dep* indiquant le nombre d'objets des départements suivants:
  - American Decorative Arts
  - Drawings and Prints
  - Medieval Art
  - Egyptian Art
  - Autres

Le fichier créé contiendra le nom du département suivi d'une tabulation, suivi du nombre d'objets trouvés.

En utilisant gnuplot, représenter les résultats sur un même graphique sous forme d'histogrammes (créer un programme gnuplot qui sera réutilisable pour un autre jeu de données).

- 4) On souhaite ensuite créer un fichier *nomFichier\_class* indiquant le nombre d'objets ayant les classifications suivantes:
  - Metal
  - Silver
  - Prints
  - Drawings
  - Ceramics
  - Ephemera
  - Autre
  - Nul (non renseigné)

Le fichier créé contiendra le nom de la classification suivi d'une tabulation, suivi du nombre d'objets trouvés. (Attention, si un objet a plusieurs classifications, il doit être compté dans chacune des classifications).

En utilisant gnuplot, représenter les résultats sur un même graphique sous forme d'histogrammes (créer un programme gnuplot qui sera réutilisable pour un autre jeu de données).

- 5) On souhaite maintenant créer 2 fichiers *nomFichier\_print* et *nomFichier\_drawings* contenant les aires en cm2 des objets dont le nom contient Print, respectivement Drawings.

Chaque fichier créé contiendra le numéro de l'objet suivi d'une tabulation suivi de l'aire (en cm2)

En utilisant gnuplot, représenter les résultats sous forme de boxplots (créer un programme gnuplot qui sera réutilisable pour un autre jeu de données).