

## TRAVAIL À RÉALISER

Ce travail a pour objectif **d'étudier** un échantillon de données fourni dans le fichier jfreechart-stats.csv. Cet échantillon donne les valeurs de métriques collectées par les classes du dossier main du JFreechart:

- NoCom : nombre de *commits* (combien de fois la classe a été changée) dans l'historique git de la classe
- NCLOC : nombre de lignes de code qui ne sont pas ni vides ni commentaires
- DCP : densité de commentaires (CLOC/LOC) donnée en pourcentage

Les tâches à effectuer :

- T1. (30%) **Visualisez** chacune des métriques de l'échantillon en créant les boîtes à moustaches. Calculez les informations pertinentes selon les définitions du cours (diapositives 6,7) et décrivez les distributions.
- T2. (30%) **Étudiez les corrélations** entre NoCom et NCLOC, NoCom et DCP. Visualisez les données, calculez les droits de régression, et les coefficients de corrélation qui ont du sens.
- T3. (40%) Nous voulons évaluer l'hypothèse : « *les classes qui ont été modifiées plus de 10 fois sont mieux commentées que celles qui ont été modifiées moins de 10 fois* ». **Décrivez** la conception d'une quasi-expérience qui vous permettra de le faire. Ensuite, **évaluez l'hypothèse**, discutez les résultats et décrivez vos conclusions.  
Rappelez-vous que selon les diapositives du cours, les étapes à suivre sont : choix d'étude, énoncé des hypothèses, définition des variables, interprétation et généralisation des résultats, discussion des menaces à la validité.

## PRÉCISIONS GLOBALES

1. Travail à remettre le **vendredi 18 novembre 23h59** via StudiUM. Aucun retard ne sera accepté.
2. Le livrable le plus important est le **rapport**, en format PDF. Assurez-vous de communiquer clairement en français, la qualité de votre rapport est très importante. Un mauvais rapport pourrait causer une déduction très significative.
3. Travaillez en groupes de 2. **Aucune soumission individuelle ne sera acceptée.** Soumettez un ZIP nommé comme suit : `prenom1_nom1_prenom2_nom2.zip`
4. Le membre de l'équipe qui ne soumet pas le fichier Zip principal de la soumission doit soumettre un zip (même nom) avec le fichier readme.txt, qui devrait contenir les noms de l'équipe.
5. Les rapports doivent compter au maximum 4 pages, y compris toutes les figures et références. Rapports de plus de 4 pages vont être éliminés d'office (par défaut).
6. Votre fichier ZIP doit aussi contenir un readme.txt avec liens vers tous ressources nécessaires (y compris votre repository git).
7. Comme en TP1 et TP2, vous devez utiliser un **repository git** pour stocker tout code, script, etc (idéalement, le même repository que TP1, TP2). Vous pouvez utiliser n'importe quel service gratuit comme Github, Bitbucket, et autres (quelques-uns vous permettent de créer des comptes académiques avec votre courriel @umontreal.ca). Utilisez le repository pour collaborer avec votre coéquipier.
8. Vous pouvez utiliser n'importe quel logiciel qui appuie les opérations statistiques requises (Excel, R, Octave, SPSS, Python, Java, ...). Cependant, nous n'offrons pas de conseils sur l'utilisation des logiciels statistiques.
9. Vous devez mettre dans votre repository git tout script, ou fichier de configuration que vous avez utilisé. En fait, votre repository doit contenir assez des détails pour nous convaincre que vous avez vraiment effectué le travail vous-même.
10. Nous allons examiner l'historique de votre repository pour nous assurer que tous les deux coéquipiers ont travaillé sur le TP et que votre code n'est pas plagié. Un historique de commit plausible devrait contenir de nombreux petits commit, chacun avec un message de commit approprié. **Faire juste quelques commit massives proche à la date limite pourrait entraîner une déduction considérable.**