

BIO500 : Méthodes en écologie computationnelle

Victor Cameron

Département de biologie
Université de Sherbrooke
Local D8-0012
819-821-8000 #61928

victor.cameron@usherbrooke.ca

Hiver 2023

Renseignements généraux ¹

Nombre de crédits	2
Cours préalables	Aucun
Lieu du cours	D7-2023
Jours et heures des cours	14 février, 7, 14, 21, 28 mars, 4 et 11 avril Tous les cours sont le mardi de 8h30 à 11h30
Session	Hiver 2023
Date de début	14 février
Date de fin	11 avril
Date de remise de l'évaluation finale	22 avril à 16h00
Date limite de retrait	6 mars
Date limite d'abandon	28 mars
Assistant à l'enseignement	Benjamin Mercier Local D8-0022 benjamin.b.mercier@usherbrooke.ca

Objectif général

Les outils informatiques sont utilisés de façon croissante en écologie, que ce soit pour la réalisation d'analyses spatiales, statistiques ou pour la gestion de bases de données. On exige de plus en plus la transparence et la reproductibilité des études scientifiques et d'évaluations environnementales.

Au terme de ce cours, l'étudiant sera en mesure de réaliser l'ensemble de la séquence d'une étude d'écologie en respectant les standards de gestion, d'analyse et de présentation des données. Le cours portera sur la réalisation d'un projet intégrateur, de la collecte des données à la production du rapport final.

Objectifs spécifiques

Au terme de ce cours, l'étudiant sera en mesure de :

- Planifier une base de données et la préparation de formulaires pour leur acquisition ;

1. Ce plan de cours est soumis au *Règlement facultaire d'évaluation des apprentissages des étudiantes et des étudiants* de la faculté des sciences de l'Université de Sherbrooke et y est conforme.

- Programmer et interroger une base de données relationnelle ;
- Compiler et exécuter un projet au moyen de la librairie targets ;
- Représenter visuellement les données au moyen de R ;
- Préparer un rapport d'étude au moyen de RMarkdown ;
- Utiliser un système de contrôle de version pour le suivi des modifications sur du code ;
- Porter un regard critique sur la reproductibilité et la transparence d'études scientifiques ;

Pré-requis

Un ordinateur portable personnel est requis pour ce cours.

Ce cours obligatoire est offert aux étudiants en fin de programme de baccalauréat en biologie, concentration écologie. Le cours *BIO109 : Introduction à la programmation scientifique* est préalable à celui-ci.

Approche pédagogique

Les séances seront constituées de courtes leçons magistrales sur des notions de bases sur les différents outils utilisés, entre-coupées d'exercices spécifiques destinés à pratiquer les éléments enseignés. Les séances seront complémentées de discussions sur les enjeux de la reproductibilité en science. Les séances se conclueront sur la réalisation d'un exercice intégrateur à compléter à la maison. L'apprentissage portera sur la réalisation d'un projet de session où les étudiants seront responsables de l'ensemble des étapes d'une étude en écologie. Le travail sera réalisé par blocs, au fur et à mesure de la présentation du matériel.

L'ensemble du matériel du cours sera disponible sur un dépôt git à l'adresse :

<https://github.com/EcoNumUdS/BIO500.git>

Contenu

Dates	Contenus	Activités	Évaluations
Bloc 1 14 février au 20 mars Planification et organisation de la collecte et organisation des données	Introduction	Séance 1 (14 février) <i>Travaux</i> - Commencer à récolter des données <i>Lectures</i> - Baker. 2016. Is there a reproducibility crisis ?	

Dates	Contenus	Activités	Évaluations
	La gestion des données biologiques <ul style="list-style-type: none"> - Types de données - Formulaire de saisie - Base de données relationnelles SQL 	Séance 2 (7 mars) Discussion sur la reproductibilité <i>Travaux</i> <ul style="list-style-type: none"> - Collecter les données pour constituer un réseau - Concevoir et scripter une base de données - Collecter des données - Préparer les arguments pour le débat <i>Lectures</i> <ul style="list-style-type: none"> - Poisot et al. 2014. Moving toward a sustainable ecological science : don't let ecological data go to waste ! - Mills et al. 2015. Archiving Primary Data : Solutions for Long-term Studies. 	
	La gestion des données biologiques <ul style="list-style-type: none"> - Bases de données relationnelles SQL - Requêtes 	Séance 3 (14 mars) Débat sur le partage de données <i>Travaux</i> <ul style="list-style-type: none"> - Installer Git - Scripter le traitement, la création de la base de données et l'injection de données - Préparer les requêtes et les résultats - Calculer les propriétés du réseau <i>Lectures</i> <ul style="list-style-type: none"> - Stokel-Walker & Van Noorden. 2023. What ChatGPT and generative AI mean for science. - Nature Machine Learning. 2023. The AI writing on the wall. 	
Bloc 2 21 au 27 mars Outils pour une science reproductible	Les outils pour la reproductibilité <ul style="list-style-type: none"> - Système de contrôle de version Git 	Séance 4 (21 mars) Débat sur l'impact d'algorithmes comme ChatGPT sur la science et sa transparence <i>Lectures</i> <ul style="list-style-type: none"> - Milcu et al. 2018. Genotypic variability enhances the reproducibility of an ecological study 	

Dates	Contenus	Activités	Évaluations
	Les outils pour la reproductibilité <ul style="list-style-type: none"> - Cahier de laboratoire avec RMarkdown - Librairie targets 	Séance 5 (28 mars) Discussion sur les liens entre la variabilité et la reproductibilité <i>Travaux</i> <ul style="list-style-type: none"> - Créer un cahier de laboratoire et le sauver sur GitHub - Construire un script target pour suivre l'évolution du travail 	Essai (25%) Rédaction d'un essai sur les enjeux de reproductibilité en science à remettre le 22 avril à 16h sur Moodle
Bloc 3 28 mars au 10 avril Visualisation des données	Visualisation des données au moyen de R <ul style="list-style-type: none"> - Fonctions graphiques de base et paramètres graphiques - Librairies R spécialisées 	Séance 6 (4 avril) <i>Travaux</i> <ul style="list-style-type: none"> - Identifier les questions de recherche - Illustrer le réseau de collaboration - Compléter l'analyse au moyen de visualisations <i>Lectures</i> <ul style="list-style-type: none"> - Pennisi. 2020. Spider biologist denies suspicions of widespread data fraud in his animal personality research - Lawkowski. 2020. What to do when you don't trust your data anymore 	
Bloc 4 11 au 22 avril Communication scientifique au moyen de RMarkdown	Documents dynamiques avec RMarkdown <ul style="list-style-type: none"> - Rédaction de rapports scientifiques - Gestion des références 	Séance 7 (11 avril) Discussion sur la fraude en science et sa prévention	Évaluation terminale (75%) Travail de session : Écrire un rapport de votre analyse des données de collaboration entre étudiants de la classe sous forme d'article scientifique Tous les scripts sont à remettre sous la forme d'un dépôt GitHub

Évaluation

L'évaluation porte sur la réalisation d'un travail de session (75%), réalisé en équipe de 4 personnes. Le travail sera divisé en 3 étapes réparties au cours de la session. L'évaluation finale (25%) portera sur la rédaction d'un essai de 1200 mots sur les enjeux de reproductibilité en science expérimentale. Le travail de session et l'essai doivent être déposés au plus tard le vendredi 22 avril 2020 à 16 :00. La pénalité est de 10% par journée de retard.

Modalités de remise

Les travaux devront tous être remis sur Moodle. Aucun travail ne sera accepté par courrier électronique.

Modalités de correction et de notation pour l'évaluation terminale

La note obtenue pour l'ensemble des travaux sera convertie en fonction des cotes proposées par la Politique d'évaluation de l'Université de Sherbrooke. La notation définitive sera exprimée en conformité avec le règlement de la Faculté des sciences de l'Université de Sherbrooke, soit à partir du tableau suivant :

Excellent A+, A, A-	Très bien B+, B, B-	Bien C+, C, C-	Passable D+, D	Échec E
W : échec par abandon	AB* : abandon	IN** : Incomplet	R : réussite	

* La mention AB est consignée seulement si l'étudiante ou l'étudiant abandonne le cours avant la date limite d'abandon. Si la date limite d'abandon est dépassée, la mention au relevé de notes de l'étudiante ou de l'étudiant sera W.

** La mention IN est utilisée au relevé de notes pour les activités pédagogiques lorsque, pour des motifs acceptés par la faculté ou le centre universitaire de formation, l'étudiante ou l'étudiant n'a pas satisfait à toutes les exigences. Est remplacée par la note W (échec par abandon) au relevé de notes du trimestre au cours duquel prend fin le délai accordé si l'activité n'a pas été complétée.

La note finale du cours sera remise au plus tard deux semaines après le dépôt de l'évaluation finale.

Appréciation de la qualité de la langue

En conformité avec l'article 17 du règlement facultaire d'évaluation des apprentissages des étudiantes et des étudiants, la qualité du français écrit dans l'évaluation peut être prise en compte. Tout travail non conforme aux exigences quant à la qualité du français écrit et aux normes de présentation peut retourner à l'étudiante ou à l'étudiant et peut aussi entraîner la perte de points pour une mauvaise qualité du français écrit. La qualité du français peut compter jusqu'à 5% des points alloués à l'évaluation.