

# BIO500 : Méthodes en écologie computationnelle

Victor Cameron

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

[victor.cameron@usherbrooke.ca](mailto:victor.cameron@usherbrooke.ca)

Hiver 2024

## Renseignements généraux <sup>1</sup>

Nombre de crédits	2
Cours préalables	Aucun
Lieu du cours	D7-2016
Jours et heures des cours	13 et 20 février, 12, 19 et 26 mars, 2 et 9 avril Tous les cours sont le mardi de 8h30 à 11h30
Session	Hiver 2024
Date de début	13 février
Date de fin	9 avril
Date de remise de l'évaluation finale	21 avril à 16h00
Date limite de retrait	19 février
Date limite d'abandon	19 mars
Assistant à l'enseignement	Benjamin Mercier Local D8-0012 <a href="mailto:benjamin.b.mercier@usherbrooke.ca">benjamin.b.mercier@usherbrooke.ca</a>

## 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.

---

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.

## 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 ;
- 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
<b>Bloc 1</b> 13 février au 18 mars  Planification et organisation des données	<b>Introduction</b>	<b>Séance 1</b> (13 février)  <i>Travaux</i> - Explorer et choisir un jeu de données - Préparer les arguments pour le débat  <i>Lectures</i> - Baker. 2016. Is there a reproducibility crisis ? - 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.	

Dates	Contenus	Activités	Évaluations
	<b>La gestion des données biologiques</b> <ul style="list-style-type: none"> <li>- Types de données</li> <li>- Formulaires de saisie</li> <li>- Base de données relationnelles SQL</li> </ul>	<b>Séance 2</b> (20 février) <b>Débat sur le partage de données</b>  <i>Travaux</i> <ul style="list-style-type: none"> <li>- Plan d'assemblage des données</li> <li>- Nettoyer et valider les données</li> <li>- Concevoir et scripter une base de données</li> </ul>	<b>Évaluation formative</b> Évaluation par les pairs des scripts d'assemblage, de nettoyage et de validation des données à remettre le 11 mars sur Moodle
	<b>La gestion des données biologiques</b> <ul style="list-style-type: none"> <li>- Bases de données relationnelles SQL</li> <li>- Requêtes</li> </ul>	<b>Séance 3</b> (12 mars) Reproductibilité  <i>Travaux</i> <ul style="list-style-type: none"> <li>- Scripter le traitement, la création de la base de données et l'injection de données</li> <li>- Préparer les requêtes et les résultats</li> </ul> <i>Lectures</i> <ul style="list-style-type: none"> <li>- Stokel-Walker &amp; Van Noorden. 2023. What ChatGPT and generative AI mean for science.</li> <li>- Nature Machine Learning. 2023. The AI writing on the wall.</li> </ul>	
<b>Bloc 2</b> 19 mars au 1 avril  Outils pour une science reproductible	<b>Les outils pour la reproductibilité</b> <ul style="list-style-type: none"> <li>- Système de contrôle de version Git</li> <li>- Gestion des conflits</li> </ul>	<b>Séance 4</b> (19 mars) Débat sur l'impact d'algorithmes comme ChatGPT sur la science et sa transparence  <i>Lectures</i> <ul style="list-style-type: none"> <li>- Milcu et al. 2018. Genotypic variability enhances the reproducibility of an ecological study</li> </ul>	
	<b>Les outils pour la reproductibilité</b> <ul style="list-style-type: none"> <li>- Cahier de laboratoire avec RMarkdown</li> <li>- Librairie targets</li> </ul>	<b>Séance 5</b> (26 mars) <b>Discussion sur les liens entre la variabilité et la reproductibilité</b>  <i>Travaux</i> <ul style="list-style-type: none"> <li>- Créer un cahier de laboratoire et le sauver sur GitHub</li> <li>- Construire un script target pour suivre l'évolution du travail</li> </ul>	<b>Évaluation formative</b> Évaluation par les pairs des scripts de création de la base de données à remettre le 1er avril sur Moodle  <b>Essai (25%)</b> Rédaction d'un essai sur les enjeux de reproductibilité en science à remettre le 21 avril à 16h sur Moodle

Dates	Contenus	Activités	Évaluations
<b>Bloc 3</b> 2 au 8 avril  Visualisation des données	<b>Visualisation des données au moyen de R</b> - Fonctions graphiques de base et paramètres graphiques - Librairies R spécialisées	<b>Séance 6</b> (2 avril)  <i>Travaux</i> - Identifier les questions de recherche - Compléter l'analyse au moyen de visualisations  <i>Lectures</i> - 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	
<b>Bloc 4</b> 9 au 21 avril  Communication scientifique au moyen de RMarkdown	<b>Documents dynamiques avec RMarkdown</b> - Rédaction de rapports scientifiques - Gestion des références	<b>Séance 7</b> (9 avril) <b>Discussion sur la fraude en science et sa prévention</b>	<b>Évaluation terminale (75%)</b> Travail de session : Écrire un rapport de votre analyse des données 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 sur les enjeux de reproductibilité en science. Le travail de session et l'essai doivent être déposés au plus tard le vendredi 21 avril 2024 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 pris 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.