

# FAS1002 – Initiation à la programmation en sciences sociales

Automne 2021

Samuel Guay

Horaire du cours: Mardi, 16:00 - 18:59

Local du cours: C-4019, pavillon Lionel-Groulx ou en ligne

Bureau: En ligne

Disponibilité: Prise de rendez-vous au [rencontre.samuelguay.ca](https://rencontre.samuelguay.ca)

Courriel: [samuel.guay@umontreal.ca](mailto:samuel.guay@umontreal.ca)

---

*En raison de l'incertitude liée à la situation sanitaire, ce plan de cours pourrait être amené à changer avant le début de la session. À noter que peu importe la décision et les recommandations du gouvernement et de l'Université, la **présence physique au cours est optionnelle**. La planification du cours a été pensée de façon à ce que tout le monde soit à l'aise de participer peu importe le mode de communication employé. De plus, nous avons la chance d'avoir un sujet et des d'apprentissage qui ont été conçus, entre autres, pour le travail collaboratif à distance.*

## Prérequis

Le cours ne nécessite aucune connaissance préalable en programmation ou statistique. Les étudiant-e-s de tous horizons sont les bienvenu-e-s. Ce cours d'initiation à la programmation se veut accessible pour tout individu désirant développer des connaissances en science des données sociales.

## Introduction et objectifs

Ce cours est une introduction à la programmation en sciences sociales avec R, un langage de programmation et un environnement statistique les plus populaires dans le monde pour l'analyse de données. Principalement convoité pour ses fonctionnalités statistiques avancées, R s'avère autant utilisé en recherche qu'en industrie, ce qui confère un avantage concret à son apprentissage. À ce sujet, R se prête bien à l'apprentissage de la programmation en raison de:

1. sa nature [libre](#) et gratuite;
2. sa polyvalence de fonctionnalités;
3. sa popularité et sa variété de domaines d'application;
4. la diversité de ressources et d'aide disponible gratuitement sur le Web;
5. sa vibrante communauté reconnue pour son inclusivité et son support bienveillant.

Ce cours offre un apprentissage axé sur l'acquisition de compétences pratiques et réelles qui sont devenues essentielles en science des données reproductibles. Dans la première partie du cours, les concepts et techniques de base en programmation informatique seront enseignés pour se familiariser avec l'environnement et développer ses compétences qui faciliteront la manipulation, et l'analyse des données avec R. La deuxième partie du cours couvre des fonctions et des procédures couramment utilisées dans R et les aspects de travail collaboratif. La troisième partie du cours propose un survol des différents dialectes ou styles de syntaxe en R, du principe de données bien rangées ou *Tidy data* et de nettoyage de données en découvrant les expressions régulières. Enfin, le dialecte du *Tidyverse* et d'autres packages facilitant la vie des analystes seront abordés avant de clore la session avec des trucs et astuces pour déboguer son code, l'importance de la reproductibilité et de la science ouverte.

## Objectifs

L'objectif principal du cours *FAS1002 - Initiation à la programmation en sciences sociales* est d'offrir une première exposition positive à la programmation et de développer une culture de l'informatique dans le but d'acquérir des compétences techniques et des compétences générales essentielles en science des données. Pour ce faire, une approche axée sur la pratique sera mise de l'avant afin de se familiariser avec plusieurs concepts et outils de base en informatique qui permettront une meilleure compréhension et utilisation de R et son univers en vue de l'application de méthodes statistiques avancées dans les cours suivants. Plus spécifiquement, les séances abordent différents aspects techniques qui permettront d'apprendre graduellement à développer du code afin d'analyser des données, produire des rapports interactifs et bien plus encore.

Les compétences techniques qui seront apprises comprennent:

- Maîtriser l'environnement de RStudio;
- Connaître les types et structures des données;
- Manipuler et nettoyer des jeux de données;
- Conceptualiser et créer des boucles pour réaliser différents types d'analyses;
- Utiliser des fonctions internes et créer des fonctions originales;
- Manipuler et transformer des chaînes de caractères des bases de données;
- Construire des tableaux et des figures pour réaliser des analyses statistiques simples;
- Connaître les différents packages du Tidyverse;
- Être en mesure d'appliquer les bonnes pratiques reconnues en contexte de partage de code et de travail collaboratif.

L'individu complétant le cours avec assiduité sera également en mesure de transférer et d'appliquer ses connaissances dans d'autres environnements puisque des compétences générales seront acquises:

- Capacité à chercher de l'aide et à bien poser ses questions;
- Connaître les concepts de base d'informatique et de programmation;
- Élaborer la structure de scripts à l'aide de pseudocode et décomposition;
- Capacité à résoudre des problèmes concrets à l'aide de la programmation;
- Savoir choisir les outils ou programmes optimaux pour manipuler et analyser des données;
- Comprendre l'importance de la reproductibilité et de la science ouverte dans le contexte de programmation et de données.

## Qui bénéficiera le plus du cours:

(En ordre décroissant)

- L'individu qui n'a aucune notion de programmation ou du langage R;
- L'individu qui a appris de façon autodidacte et qui souhaite consolider ses bases;
- L'individu qui connaît R et qui souhaite apprendre les méthodes de travail collaboratif.

À noter que l'emphase est mise avant tout sur l'apprentissage des notions pour les néophytes en la matière, donc une personne ayant déjà des connaissances pourrait trouver que le cours n'avance pas assez vite. Ceci étant dit, le cours a été pensé afin d'offrir une flexibilité pour les individus plus rapides ou avancés qui leur permet d'approfondir des concepts liés au cours ou d'assister leurs collègues.

Enfin, il importe de préciser que ce cours ne prétend pas rendre la personne qui complète le cours experte en programmation ou en analyse de données. Ce cours offre plutôt la possibilité d'apprendre et de mettre en pratique les bases de la programmation avec R dans un environnement structuré qui favorise l'autonomie de sa propre courbe d'apprentissage en fonction de ses intérêts.

## Pédagogie

Ce cours est interactif et se déroule principalement, sans grande surprise, sur votre ordinateur! La meilleure façon d'apprendre des outils de programmation et d'analyse est de travailler directement avec des données et mettre du temps à essayer de construire des choses. Ainsi, adoptant une méthode d'enseignement analogue à une classe inversée, la majorité de la théorie est enseignée à

partir d'une plateforme Web interactive, alors que la majorité du temps en classe sera consacrée à la révision de concepts, à répondre aux questions et à échanger, à des sessions de codage en direct et à avancer les travaux. Plusieurs modalités d'enseignement seront utilisées afin d'agrémenter l'apprentissage des concepts et techniques en vue de les appliquer pour résoudre des problèmes réels et concrets. Contrairement aux langages de programmation à usage général tels que Python, R a été spécialement conçu pour les statistiques. Nous apprendrons à faire des choses plus générales de programmation avec, mais l'utilité principale de R demeure vraiment l'analyse de données. Cependant, les compétences enseignées dans ce cours sont conçues pour être transférables à d'autres langages comme Python.

La [plateforme Web Ed](#) est au coeur de la pédagogie et du bon déroulement du cours. La plateforme permet de diffuser la théorie, les quiz et les ateliers pratiques, de poser des questions avec du code, de développer du code en collaboration avec tous les membres de la classe. De plus, elle offre un espace de travail similaire à un [Jupyter Notebook](#) pour développer ses travaux. Pour la pratique, elle sera davantage utilisée en première partie du cours puisqu'une migration progressive vers l'utilisation de [RStudio](#) et [Git\(Hub\)](#) se fera au courant de la session. Une instance de RStudio est toutefois disponible sur la plateforme, ce qui permet aux gens avec un Chromebook roulant ChromeOS de se familiariser avec RStudio, outil largement utilisé dans la réalité. Une période d'installation et de débogage des différents logiciels est prévue à l'horaire lors de la deuxième semaine puisqu'il y a souvent des erreurs. Les bogues sont nos amis, ils offrent des opportunités en or pour apprendre. L'enseignement dans ce cours consiste, avant tout, à soutenir l'apprentissage.

En plus des lectures et de la théorie enseignée, il y a plusieurs méthodes d'évaluation servant à mesurer la progression des apprentissages tout au long du semestre. Aucun examen en classe n'est prévu, seulement des travaux de différentes ampleurs. Les évaluations servent avant tout d'aide pour compléter et appliquer avec assiduité la matière vue dans le cours.

## Évaluations

Sur un total de 100 points:

- **10 points** sont alloués pour la **participation observée** autant sur les plateformes que dans les séances en direct. J'estime qu'être engagé-e dans le cadre du cours ne signifie pas nécessairement de prendre la parole lors des séances synchrones. Vous pouvez faire preuve de participation par le biais d'entraide avec vos collègues lors des sessions de codage, à réagir aux commentaires des autres, tenter de répondre aux questions ou suggérer des améliorations sur le code des autres, etc.
- **24 points** sont alloués pour **8 quiz** dispersés tout au long de la session qui doivent être complétés avant le début de la séance en direct. Ces quiz doivent être complétés avec un score de 100 % pour obtenir tous les points, mais ils peuvent être repris autant de fois que nécessaire. Le principal objectif de ces quiz est de motiver l'étudiant-e à compléter le contenu à temps et mettre en évidence les concepts essentiels de la semaine.
- **16 points** sont alloués pour la **présentation d'un package du Tidyverse**. Tidyverse est une collection avisée de packages R conçus pour la science des données. Tous les packages partagent une philosophie de conception, une syntaxe et des structures de données communes. Selon le nombre d'étudiant-e-s inscrit-e-s au cours, la présentation se fera individuellement ou en équipe.
  - Dates de présentation: 9, 16 et 23 novembre.
- **20 points** sont alloués pour le **travail de mi-session** qui évaluera les compétences enseignées jusqu'à la semaine de lecture. Des instructions détaillées sur ce qui est attendu dans le travail seront fournies lors des premières semaines du semestre. Ce travail mettra à l'épreuve concepts de bases en programmation, gestion de version de contrôle et collaboration.
  - Date limite pour la remise: [22 octobre 2021 - 23:59 - Anywhere On Earth \(AOE\)](#).
- **30 points** sont alloués pour le **travail final** qui ressemblera au travail de mi-session, mais qui deviendra un vrai rapport de recherche complet et automatisé. Tous les concepts abordés dans le cours devront être approfondis lors de sa création.
  - Date limite pour la remise: [22 décembre 2021 - 23:59 - Anywhere On Earth \(AOE\)](#).

**Calendrier :**

7 septembre	Semaine 1	<b>Présentation du plan de cours et d'éléments d'informatique pour la programmation</b> Tous les liens vous seront partagés sur <a href="#">la page d'accueil du cours sur StudiUM</a> avant la première semaine. Lectures: <b>Module 1</b> Évaluation: <b>Quiz 1 à compléter avant le cours</b>
14 septembre	Semaine 2	<b>Introduction et installation de R et RStudio, aide, et débogage</b> Lectures: <b>Module 2</b> Évaluation: <b>Quiz 2 à compléter avant le cours</b>
21 septembre	Semaine 3	<b>Bases de la programmation et R</b> Lectures: <b>Module 3 et 4</b> Évaluation: <b>Quiz 3 à compléter avant le cours</b>
28 septembre	Semaine 4	<b>Bases de la programmation et R</b> Lectures: <b>Module 3 et 4</b> Évaluation: <b>Quiz 4 à compléter avant le cours</b>
05 octobre	Semaine 5	<b>Structure et manipulation des données</b> Lectures: <b>Module 5</b> Évaluation: <b>Quiz 5 à compléter avant le cours</b>
12 octobre	Semaine 6	<b>Scripts, Rmarkdown, Git et Github</b> Lectures: <b>Module 6</b> Évaluation: <b>Quiz 6 à compléter avant le cours</b>
19 octobre	NA	<b>Pas de cours - Semaine de travaux et d'étude</b>
26 octobre	Semaine 7	<b>Styles de syntaxe en R et utilisation des bibliothèques et des packages et bonnes pratiques de programmation</b> Lectures: <b>Module 7</b> Évaluation: <b>Quiz 7 à compléter avant le cours</b>
2 novembre	Semaine 8	<b>Principes de Tidy data et de nettoyage de jeux de données et manipulation de caractères</b> Lectures: <b>Module 8</b> Évaluation: <b>Quiz 8 à compléter avant le cours</b>
9 novembre	Semaine 9	<b>Tidyverse I (tibble, magrittr, dplyr)</b> Lectures: <b>Module 9</b> Évaluation: <b>Présentations des étudiant-e-s 1</b>
16 novembre	Semaine 10	<b>Tidyverse II (tidyr, forcats, stringr)</b> Lectures: <b>Présentations des étudiant-e-s 2</b> Évaluation: <b>Module 9</b>
23 novembre	Semaine 11	<b>Tidyverse III (ggplot2) et autres packages utiles</b> Lectures: <b>Présentations des étudiant-e-s 3</b> Évaluation: <b>Module 9</b>
30 novembre	Semaine 12	<b>Débogage, reproductibilité et science ouverte</b> Lectures: <b>Module 10</b>
7 décembre	Semaine 13	<b>La semaine manquante - Méli-mélo de trucs et astuces qu'on n'enseigne jamais</b> (mais qu'on devrait peut-être enseigner) Évaluation: <b>Pré-soumission du travail final sur GitHub.</b>

## Ouvrages

Toutes les lectures utilisées sur la plateforme du cours sont disponibles gratuitement et librement sur internet. Elles seront ajoutées à même la plateforme pour faciliter leur lecture. L'ouvrage principal sur lequel le cours est basé est [Programmer avec R par Vincent Goulet \(2021\)](#).

## Rappel de règlements pédagogiques

Veillez prendre note que le trimestre se termine le 22 décembre 2021 (incluant la période des examens). Aucune demande d'examen différé ne sera acceptée sans motif valable. Nous entendons par motif valable, un motif indépendant de la volonté de l'étudiant-e, tel que la force majeure, le cas fortuit ou une maladie attestée par un certificat de médecin.

### Absence à un examen

L'étudiant-e doit motiver, en remplissant le formulaire disponible dans le Centre étudiant, toute absence à une évaluation ou à un cours faisant l'objet d'une évaluation continue dès qu'il est en mesure de constater qu'il ne pourra être présent-e à une évaluation et fournir les pièces justificatives dans les sept jours suivant l'absence.

### Délais pour la remise d'un travail

L'étudiant-e doit motiver, en remplissant le formulaire disponible dans le Centre étudiant, toute demande de délai pour la remise d'un travail dès qu'il ou elle est en mesure de constater qu'il ou elle ne pourra remettre à temps le travail et fournir les pièces justificatives.

La pénalité imposée pour les retards dans la remise des travaux est de 10 points de pourcentage par jour à moins d'indication contraire. Cette pénalité est calculée en déduisant 10 points de pourcentage à la note obtenue pour le travail en question. Il s'agit de la politique du Département.

### La prévention du plagiat

Le Département porte une attention toute particulière à la lutte contre le plagiat, le copiage ou la fraude lors des examens. Le plagiat consiste à utiliser de façon totale ou partielle, littérale ou déguisée le texte d'autrui en le faisant passer pour sien ou sans indication de référence à l'occasion d'un travail, d'un examen ou d'une activité faisant l'objet d'une évaluation. Cette fraude est lourdement sanctionnée.

Tous les étudiant-e-s sont invités à consulter le site web <http://www.integrite.umontreal.ca/> et à prendre connaissance du Règlement disciplinaire sur le plagiat ou la fraude concernant les étudiant-e-s. Plagiat peut entraîner un échec, la suspension ou le renvoi de l'Université.

### Le harcèlement, y compris à caractère sexuel

Il incombe à chaque membre de la communauté universitaire de se conduire avec respect en tout temps envers tout le monde. En particulier, le Département de science politique de l'Université de Montréal s'engage à créer un milieu accueillant et sécuritaire pour toutes et tous, quelle que soit leur identité.

Les documents suivants ont des démarches pratiques à suivre :

- [Si vous pensez que vous vivez du harcèlement](#)
- [Si on s'est confié à vous ou si vous êtes témoin de harcèlement](#)

Pour toute autre question : <http://harcelement.umontreal.ca/>

### Besoin d'écoute? Situation de détresse?

Comme étudiant-e, vous pouvez faire appel à plusieurs lignes d'écoute ou d'urgence. Tous les liens sont rassemblés sur la page [Ça va aller](#).

Les étudiant-e-s ont accès à un [service 24 heures/7 jours](#) offert par l'Alliance pour la santé étudiante au Québec. Le numéro est le suivant : 1-833-851-1363.

Vous pouvez aussi faire appel à une sentinelle. La sentinelle est un-e employé-e de l'UdeM, formé-e et disponible pour vous accueillir, vous écouter et vous orienter vers les bonnes ressources. Son accueil est spontané, respectueux et confidentiel. Voir le [Bottin des sentinelles](#) au besoin.