Université de Montréal

Faculté des arts et des sciences
Département de démographie

Plan de cours

DMO 3900 Informatique pour la démographie (3 cr.)

Hiver 2024

Plage horaire : MARDI 8h30-11h30
Local : B-XXXX, pavillon 3200 Jean-Brillant

CHARGÉ DE COURS :
Pietro VIOLO
C-5044, pavillon Lionel-Groulx
pietro.violo@umontreal.ca
Disponibilité : MERCREDI 11h30-12h30

AUXILIAIRE D'ENSEIGNEMENT :
Alexis NADEAU
C-5045, pavillon Lionel-Groulx
alexis.nadeau@umontreal.ca
Disponibilité : MARDI 12h10-13h10

1 Objectifs du cours

Ce cours offre d'abord une formation approfondie en R, puis introduit d'autres langages de programmation tels que SAS et STATA, qui sont couramment utilisés dans les milieux professionnels et académiques. L'apprentissage des spécificités de ces langages, ainsi que des bonnes pratiques de codage et de leur mise en œuvre dans des projets, permettra aux étudiants de développer des compétences informatiques solides, incluant :

- la conception et la mise en oeuvre d'algorithmes pour résoudre des problèmes concrets ;
- l'adoption de bonnes habitudes en matière de programmation informatique, une activité se déroulant le plus souvent en groupe qu'en solo ;
- la familiarisation avec les structures conditionnelles et les boucles itératives.
- la visualisation des données, un atout extrêmement recherché sur le marché du travail et en recherche.

Tous les exemples d'application s'inscriront dans le domaine de la démographie et mettront à profit des connaissances théoriques et méthodologiques déjà acquises par les étudiant.e.s dans d'autres cours de démographie (e.g., manipulation de données, calculs d'indicateurs, représentations graphiques pour la visualisation de données et de résultats).

Grâce aux compétences informatiques générales nouvellement développées dans ce cours, les étudiant.e.s seront ainsi plus à même d'apprendre et de maîtriser plus rapidement d'autres langages de programmation (tels que Python, C++, Matlab, etc.).

2 Méthodes pédagogiques

Les séances de cours se dérouleront en présentiel dans une salle informatique (B-XXXX). Les différents sujets seront d'abord introduits sous forme d'exposés magistraux, suivis de démonstrations faites par le chargé de cours avec le logiciel à l'étude. Afin de maximiser la consolidation des acquis, chaque séance de cours se clôturera par des exercices pratiques supervisés, à résoudre par les étudiant.e.s avec l'aide du chargé de cours et de l'auxiliaire d'enseignement, qui circuleront dans la salle informatique afin de répondre aux questions. À l'occasion, de courtes capsules vidéo pédagogiques préenregistrées portant sur des sujets ciblés pourraient être conçues et mises à la disposition des étudiant.e.s afin de compléter les contenus des séances de cours en présentiel.

Il convient de souligner que **l'apprentissage de nouveaux langages de programmation nécessite de la pratique sur une base régulière**. Entre deux séances de cours successives, les étudiant.e.s devront ainsi s'assurer de prendre le temps de mettre en pratique tous les sujets abordés en classe lors de la plus récente séance complétée. La maîtrise de tous les sujets précédents est incontournable pour l'acquisition de ceux de la semaine suivante.

3 Évaluation

Examen intra (à livres ouverts, sans ordinateur) – **35 %**
20 février 2024, 8h30-10h30
C-5143, pavillon Lionel-Groulx
Sujet : Programmation sous R

TABLE 1 – Calendrier des TP et Devoirs

Type	Travail	Sujet	Date d'annonce	Date de remise
TP (**30%**)	TP1 (**6%**)	Hygiène et environnement	8 janvier	15 janvier
	TP2 (**8%**)	Programmation sous R	15 janvier	22 janvier
	TP3 (**8%**)	Programmation sous SAS et Stata	26 février	18 mars
	TP4 (**8%**)	Programmation sous R	18 mars	1er avril
Devoirs [1] (**35%**)	Devoir 1 (**17,5%**)	Programmation sous R	22 janvier	12 février
	Devoir 2 (**17,5%**)	Projet de recherche	18 mars	29 avril

4 Plagiat et intégrité intellectuelle

À l'Université de Montréal, le plagiat est sanctionné par le Règlement disciplinaire sur le plagiat et la fraude concernant les étudiant.e.s du premier cycle et les étudiant.e.s des cycles supérieurs. L'Université met à votre disposition d'excellentes ressources pour prévenir le plagiat et la fraude dans vos travaux et vos examens, consultez-les ici. L'Université de Montréal interdit l'utilisation de **ChatGPT**, et toute utilisation suspectée sera **signalée**. Vous ne pouvez utiliser que les notions apprises en classe.

1. Les devoirs peuvent être faits en équipe de deux, les TP sont des travaux individuels.

5 Calendrier du cours

Semaine 1 (8 janvier)

- — Présentation et discussion du plan de cours, des méthodes pédagogiques et de l'évaluation
- — Hygiène informatique : les principes de base
- — Qu'est-ce que la programmation informatique et qu'entend-on par ses « bonnes pratiques » ?
- — Le logiciel R et le convivial RStudio : description sommaire du logiciel R, anatomie de la session de travail, rubriques d'aide, etc.
- $\star$ **Travail pratique #1** : remise de l'énoncé

Semaine 2 (15 janvier)

- — Scalaires, vecteurs et matrices : définitions et opérations de base
- — Programmation sous R, partie 1 : premiers pas (affectation, objets, fonctions internes, sauvegarde, ...)
- — Bonnes pratiques (plan, terminologies, présentation, commentaires)
- $\star$ **Travail pratique #2** : remise de l'énoncé
- $\star\star\star$ **Travail pratique #1** : remise des copies

Semaine 3 (22 janvier)

- — Programmation sous R, partie 2 : structures de données sous R, importation et exportation de données
- — Bonnes pratiques (séquence d'exécution)
- $\star\star$ **Devoir#1** : remise de l'énoncé
- $\star\star\star$ **Travail pratique #2** : remise des copies

Semaine 4 (29 janvier)

- — Programmation sous R, partie 3 : installation et utilisations de *packages*, conditions, boucles itératives (et à pré-condition), *arrays*
- — Bonnes pratiques (présentation, commentaires)

Semaine 5 (5 février)

- — Programmation sous R, partie 4 : outils graphiques dans R de base
- — Bonnes pratiques (visualisation)
- — Exporter les visualisations

Semaine 6 (12 février)

- — Programmation sous R, partie 5 : créer ses propres fonctions, révision en vue de l'examen intra
- — Bonnes pratiques (organisation)
- $\star\star$ **Devoir #1** : remise des copies

Semaine 7 (19 février)

- **Examen intra (35%)** : 20 février 8h30 à 10h30

Semaine 8 (26 février)

— Programmation sous Stata (Introduction)

⋆ **Travail pratique #3** : remise de l'énoncé

Semaine 9 (4 mars) ***** Pas de cours : période d'activités libres *****

Semaine 10 (11 mars)

— Programmation sous SAS (Introduction)

Semaine 11 (18 mars)

— Programmation sous R : L'Univers Tidyverse

— *tidyr, dplyr, ggplot2, stringr, etc.*

⋆ **Travail pratique #4** : remise de l'énoncé

⋆⋆ **Devoir #2** : remise de l'énoncé

⋆⋆⋆ **Travail pratique #3** : remise des copies

Semaine 12 (25 mars)

— Programmation sous R : Création de cartes

— Manipuler des *shapefiles*

— Comprendre les coordonnées géographiques

— Base de données *OpenStreetMap*

— Bonnes pratiques (choix de couleurs, design)

Semaine 13 (1er avril)

— Programmation sous R : Webscraping et analyse de texte

— Enjeux éthiques, bonnes pratiques

— Extraire des données de l'Internet

— Nettoyage des données et analyse des sentiments

⋆⋆⋆ **Travail pratique #4** : remise des copies

Semaine 14 (8 avril)

— Autres outils informatiques essentiels

— Github et RStudio

— RMarkdown

— Overleaf (LaTeX)

— Photoshop/Inkscape

⋆⋆ **Devoir #2** : remise des copies avant la fin de session (29 avril)

6 Ressources utiles

CODY, Ron (2018). **Learning SAS by example : A programmer's guide**, Second Edition, Cary (NC), SAS Institute, 536 pages.

DELWICHE, Laura D. et SLAUGHTER, Susan J. (2012). **The little SAS book : A primer**, Fifth Edition, Cary (NC), SAS Institute, 380 pages.

KABACOFF, Robert I. (2015). **R in action**, Second Edition, Shelter Island (NY), Manning Publications Co., 579 pages.

LONG, J.D et TEETOR, Paul (2019). **R Cookbook**, 2nd Edition, O'Reilly Media, 413 pages. En accès libre ici.

MUENCHEN, Robert A. (2011). **R for SAS and SPSS Users**, Second Edition, New York (NY), Springer Series in Statistics and Computing, Springer, 686 pages. Une ancienne version intégrale est disponible ici.

PRESSAT, Roland (1983). **L'analyse démographique**, Paris, Presses universitaires de France, 4^{e} édition, 295 pages.

PRESTON, Samuel H., HEUVELINE, Patrick et GUILLOT, Michel (2000). **Demography. Measuring and modeling population processes**, Oxford, Blackwell Publishers, 291 pages.

R Core Team (2016). **R Data Import/Export**, Vienne, R Foundation for Statistical Computing. Une version électronique intégrale est disponible ici.

VENABLES, W. N., SMITH, D. M. et le R Core Team (2018). **An introduction to R**, 99 pages. En accès libre ici.

WACHTER, Kenneth W. (2014). **Essential demographic methods**, Cambridge, Harvard University Press, 288 pages.

WICKHAM, Hadley et GROLEMUND, Garrett (2016). **R for Data Science**, O'Reilly Media, 474 pages. En accès libre ici.