

PLAN DE COURS
Département de sociologie
Université du Québec à Montréal

Sigle : SOC-8695 **Groupe :**
Titre : **École d'été en** Sciences sociales computationnelles
Session : Été 2023
Enseignant : Vissého Adjivanou
Téléphone : (514) 987-3000 poste 0900
Bureau : A-5260
Courriel : adjivanou.visseho@uqam.ca

AVIS À TOUTES LES ÉTUDIANTES ET TOUS LES ÉTUDIANTS

Boîte courriel UQAM

La boîte courriel est activée automatiquement **dans les 48 heures suivant l'inscription à au moins un cours.**

Pour y accéder aller au :

<https://servicesinformatiques.uqam.ca/services/Courriel%20%C3%A9tudiant>

Vous trouverez votre NIP sur votre relevé facture ou votre bulletin d'inscription

1. Description

Le paysage de l'analyse quantitative en sciences sociales tel que nous le connaissons est en train de changer en profondeur devant nous avec d'une part, la multiplicité des données, notamment des données numériques et administratives qui deviennent de plus en plus disponibles, et d'autre part, avec l'intégration de nouvelles méthodes d'analyses issues des sciences computationnelles (informatiques) pour traiter ces données. Le cours SOC Sciences Sociales Computationnelles ambitionne d'introduire les participant.es à ces changements et de les former à prendre plein contrôle de ces nouvelles méthodes. Les étudiants hors sciences sociales auront l'occasion de voir où leurs compétences en informatique et en statistique peuvent aller, avec des applications innovantes à des problèmes d'intérêt sociétal massif. Pour les étudiants en sciences sociales, ce sera l'occasion de développer de nouvelles compétences en statistiques et en analyse des données pour apporter des contributions nouvelles et passionnantes dans leur recherche. Ils développeront aussi des outils qui façonneront l'originalité et la puissance de leur travail pour les années à venir.

A la différence d'un cours magistral, cette école d'été favorise une prise en main complète des étudiant.es à travers des travaux pratiques et en groupe.

2. Objectifs

De manière spécifique, cette école vise à :

1. Former les étudiant.es et chercheur.es en sciences sociales (sociologie, démographie, sciences politiques, anthropologie...) au domaine émergent mais croissant des sciences sociales computationnelles
2. Former les étudiant.es et chercheur.es des sciences des données sur les thématiques de sciences sociales (causalité, expérimentations, biais...);
3. Favoriser une communication entre les étudiants de différentes disciplines ;
4. Familiariser les étudiant.es sur les récents développements dans le domaine des sciences sociales computationnelles et les possibilités de recherche et de carrière ;
5. Créer une communauté de jeunes étudiant.es et chercheur.es qui deviendront de futurs leaders dans le domaine ;
6. Développer des projets novateurs à l'intersection des sciences sociales et des sciences computationnelles sur des sujets sociaux contemporains au Nord et au Sud.

3. Résultats attendus

À la fin de cours / cette école le/la participant sera capable de :

- Maîtriser le nouvel environnement de production et de développement des données ;
- Maîtriser la collecte des données digitales en utilisant plusieurs techniques ;
- Maîtriser les approches traditionnelles d'analyse de données, leurs limites et leur extension ;
- Maîtriser le traitement et l'analyse de données textuelles ;
- Maîtriser l'analyse de réseaux.
- Comprendre les techniques d'apprentissage automatiques et les questions éthiques qui leurs sont associées ;
- Maîtriser et utiliser R pour l'analyse des données.

4. Principaux publics cibles

Ce cours / cette formation est ouverte aux étudiant.es des deuxièmes et troisièmes cycles ainsi qu'à des chercheur.es postdoctoraux et aux chercheur.es juniors (moins de 5 ans après le doctorat). Au total, 20 à 30 participant.es seront recrutés en provenance des sciences sociales (sociologie, démographie, économie, sciences politiques, médias...) ou sciences computationnelles (informatiques, statistiques, mathématiques appliquées...). Le nombre de places disponibles pour les étudiant.es qui veulent se faire créditer le cours est limité à 20.

4. Descriptions des activités

La partie présentielle de l'école d'été se déroulera les semaines 2 et 3. La deuxième semaine est destinée à l'enseignement des méthodes des sciences sociales computationnelles et la troisième semaine au développement de projets en groupes.

Pendant la deuxième semaine de l'école d'été, les journées sont divisées en trois parties principalement. Premièrement, le formateur du jour ou moi-même fournira une introduction à une partie spécifique des sciences sociales computationnelles. Ces enseignements porteront sur plusieurs sujets dont la collecte de données digitales, l'analyse textuelle et de réseaux, l'apprentissage automatique, l'expérimentation (numériques) sur le terrain et l'échantillonnage non probabiliste entre autres. Dans l'après-midi, les étudiants travaillent en groupe sur une activité liée à la formation du matin. Enfin, la journée se termine avec un conférencier invité qui est un chef de file dans le domaine. Ainsi, tout au long de cette deuxième semaine, des chercheurs renommés internationaux en sciences sociales computationnelles animeront des présentations thématiques portant sur leurs travaux spécifiques et sur les nouveaux développements de la discipline.

Notre expérience avec les écoles d'été nous indique que certains des meilleurs apprentissages se produisent dans le processus de recherche réelle. Par conséquent, au cours de la troisième et dernière semaine, les participants vont participer à des projets de recherche de groupe facultatifs dans lesquels ils feront de la recherche réelle dans un cadre de collaboration avec d'autres participants sous la supervision des organisateurs. Tout au long des deux premières semaines, les étudiants sont encouragés à réfléchir à des sujets de projet de recherche de groupe possibles pendant les exercices de l'après-midi et via des conversations avec les professeurs invités (soit en personne ou via un canal Slack qui relie tous les participants). Nous encouragerons également les professeurs invités à identifier des orientations prometteuses pour l'avenir du domaine au cours de leurs conférences qui pourraient être poursuivies par les équipes de recherche au cours de la dernière semaine. L'objectif global de ces projets de groupe est que les participants acquièrent des compétences essentielles dans la conception de la recherche, la collecte, le nettoyage et le traitement des données par eux-mêmes, car nous pensons que ces compétences fondamentales sont trop souvent exclues de la formation en sciences sociales computationnelles. Ces groupes travailleront ainsi sur leurs propres données, ou à partir des données que nous mettrons à leur disposition. À cet effet, nous solliciterons des collaborations avec le milieu non académique, avec les presses, et autres institutions génératrices de données pour voir de quelles manières nous pouvons bénéficier de leurs données dans le cadre de l'école d'été. Dans le même temps, nous encourageons les étudiants à viser des recherches qui seront éventuellement publiées.

5. Matériels

5.1. Préalable :

Vous devez avoir suivi un cours avancé en statistiques (par exemple SOC8655) ou des cours équivalents pour pouvoir suivre cette formation. Vous devez avoir une bonne connaissance de la régression linéaire et une connaissance de base de R.

5.2. Logiciels

Le cours utilise intensivement R. R est un langage de programmation libre et largement utilisé en sciences sociales dans les principales universités américaines. Pour pouvoir utiliser ce langage, nous allons recourir à R Studio qui est un environnement de développement intégré. R Studio permet de sortir les documents en version Word ou PDF. Pour ce faire, vous aurez besoin aussi d'installer sur votre machine latex. Nous n'allons pas apprendre à utiliser Latex, mais une connaissance antérieure est un atout. Voici la procédure à suivre pour l'installation de R, de RStudio, et de Latex.

5.2.1. Installation de R

Suivez les instructions selon votre système d'exploitation : <http://cran.cnr.berkeley.edu/>

5.2.2. Installer RStudio (Desktop)

- Sur le site de RStudio <https://www.rstudio.com/>
- Ou directement à partir de ce lien : <https://www.rstudio.com/products/rstudio/download/>

5.2.3. Installer Latex

- Pour les utilisateurs de Windows : installez **la complète version de MiKTeX** : <https://miktex.org/2.9/setup>:
 - Cliquez sur Net Installer
 - Téléchargez la version
 - Installez la en double-cliquant sur le fichier téléchargé et suivez les instructions.
- Pour les utilisateurs de Mac, vous pouvez utiliser TeXShop : <https://pages.uoregon.edu/koch/texshop/>
- Vous pouvez aussi utiliser MiKTeX.

5.2.4. Datacamp

Datacamp est une plateforme d'apprentissage en ligne qui sera utilisé tout au long de la formation pour appuyer votre apprentissage. Vous n'avez pas à payer pour les cours qui y sont dessus, je m'en suis déjà chargé. Veuillez-vous inscrire à partir de ce lien pour accéder aux cours :

https://www.datacamp.com/groups/shared_links/9a097f100397384b4f659fcbddb786e9e16ff9b2

Je vous **encourage vivement** (dans la mesure de vos possibilités) à compléter les deux premiers chapitres "Introduction à R" et "Introduction to the Tidyverse" avant le début des classes. N'hésitez pas à me contacter si vous avez des questions. S'il y a d'autres cours sur la plateforme qui vous intéresse, faite-moi savoir pour que je les inclus aussi. Il peut s'agir de logiciels ou de notions qui vous intéressent personnellement.

5.3. Références

Plusieurs références seront utilisées dans le cadre de cette formation. Il s'agit à la fois de livre de cours et d'articles.

- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691207544/text-as-data>
- Salganik, Matthews. (2017). Bit by bit: Social research in the digital age. Princeton University Press.
- McLevey, J. (2022). Doing Computational Social Science. SAGE Publications Inc. <https://us.sagepub.com/en-us/nam/doing-computational-social-science/book266031>
- Kosuke Imai. (2017). Quantitative social science : An introduction. Princeton University Press.
- Jean-Herman, Guay. 2014. Statistiques en sciences humaines avec R, 2^e édition. Presses de l'Université de Laval.
- John, Fox & Sanford, Weisberg. 2019. An R Companion to Applied regression, 3rd edition. Sage.
- Wickham, Hadley & Golemund, Garrett. 2017. R for Data Science: Import, Tidy, Transform, Visualize, and Model data. Boston. O'Reilly. Pp.492
- Babbie, Earl. 2015. The Practice of social research. 14th Edition. Belmont, CA: Wadsworth. <http://www.bitbybitbook.com/en/observing-behavior/> - version française en ligne.
- Kennedy, Peter. 2008. A Guide to Econometrics. Malden, MA: Blackwell Publishing.

5.4. Ressources en ligne

En français:

- <https://juba.github.io/tidyverse/index.html>: Une très bonne ressource en français
- <http://dimension.usherbrooke.ca/dimension/v2ssrcadre.html>

En anglais:

- <https://clanfear.github.io/CSSS508/>: Vous avez une série de cours et de code ici
- <https://jrnold.github.io/qss-tidy/>: tous les codes du livre de Kosuke Imai sont réécrits ici en version tidyverse.
- <https://r4ds.had.co.nz/>: le livre de Wickham en ligne
- <https://www.bitbybitbook.com/en/1st-ed/preface/>: le livre de Salganik en ligne
- <https://data.hypotheses.org/>

Comme vous vous en rendez compte assez rapidement, il y a des millions de ressources en ligne. Votre tâche souvent va consister à trouver la bonne ressource pour la tâche à faire. Aussi, google.com vous sera d'un atout formidable.

6. Mode d'évaluation

Deux modes d'évaluation sont mis en place dans le cadre du cours : des lectures obligatoires suivi de travaux pratiques, et la rédaction d'un mini-projet, ou de répliation d'un article. Une seule approche n'est pas susceptible de vous faire comprendre l'ensemble des méthodes que nous allons voir dans le cadre de ce cours. Quand les travaux de groupe et individuels vous permettent de travailler chaque méthode, ils ne vous permettent pas de conduire une étude indépendante par vous-mêmes. De même, la répliation d'articles vous permettra uniquement de mieux comprendre une seule méthode et de ne pas avoir une vue globale de l'ensemble des méthodes. Aussi, la combinaison de ces méthodes est pour vous permettre d'avoir une vue globale de l'ensemble des méthodes en même temps que de vous permettre d'investir du temps pour mieux assimiler la méthode qui est plus proche de votre projet de recherche.

6.1. Lectures d'articles et travaux pratiques

Avant chaque cours, je vous conseille vivement de lire les articles ou sessions de livres assignés à ce cours avant de venir en classe (voir section sur Calendrier). Cela facilitera largement votre compréhension. Vous devez obligatoirement lire ces matériels et produire un résumé succinct d'une page. Au cours des deux premières semaines, vous devez compléter les travaux pratiques commencés en classe. Ces travaux pratiques vont être constitués d'une partie à faire en groupe et d'une partie individuelle. Cela va compter pour 50% de la note finale.

6.2 Projet final

Le dernier mode d'évaluation sera de travailler sur un projet ou de répliquer un article qui vous sera assigné. Vous aurez à présenter votre projet le dernier jour de l'école. Le projet final doit être soumis 15 jours après la fin de l'école d'été. Les indications concernant la répliation de l'article vous seront données en classe. Ce projet comptera pour 50% de la note finale, soit 15% pour la présentation orale et 35% pour le projet écrit.

6.3. Note finale

Votre note finale sera une moyenne pondérée de votre projet (50%), et de vos lectures et travaux pratiques (50%).

6. Calendrier

Semaine 1 : Fondation

À la fin de cette semaine, vous devriez être en mesure de faire ce qui suit :

- Savoir comment télécharger, installer et utiliser R, Rstudio et Rmarkdown ;
- Exécuter des commandes shell de base pour interagir avec un ordinateur à partir de la ligne de commande ;
- Utiliser github pour suivre les modifications apportées à un projet ;
- Savoir-faire ou approfondir l'analyse de données numériques (description, analyse multivariée, visualisation) ;
- Savoir travailler avec les chaînes de caractères et comprendre les données textuelles

Jour 1 à 6 (lundi à vendredi)	9h00 – 12h30	12h30 – 13h30	13h30 – 16h00
Lundi	Introduction à R, RStudio et RMarkdown	Lunch	Exercices en groupe
Mardi	Transformation des Données avec dplyr	Lunch	Exercices en groupe
Mercredi	Visualisation avec ggplot	Lunch	Exercices en groupe
Jeudi	Modélisation	Lunch	Exercices en groupe
Vendredi	Programmation avec R	Lunch	Exercices en groupe

- Il y aura deux pauses, une pause pour le déjeuner de 12h30 à 13h30 et une pause-café de 15h à 15h30.
- Les étudiant.es doivent soumettre à la fin de chaque journée le travail commencé en équipe.

Lectures

- Sciences sociales computationnelles –

- Matthew Salganik: <https://www.bitbybitbook.com/fr/1st-ed/introduction/>
- Joshua Blumenstock : [Predicting poverty and wealth from mobile phone metadata](#)
- David Donoho. 2015. “50 Years of Data Science”. In Paper Presented at the Tukey Centennial Workshop. Princeton, NJ. <http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf>.
- Michael F. Schober, Josh Pasek, Lauren Guggenheim, Cliff Lampe, and Frederick G. Conrad. 2016. “Social Media Analyses for Social Measurement”. *Public Opinion Quarterly* 80 (1): 180–211. doi:10.1093/poq/nfv048.

- Causalité

- Gerber, Alan S. and Donald P. Green, *Field Experiments: Design, Analysis, and Interpretation*. Chapters 1 and 2.
- Stephen, Morgan & Christopher Winship. 2018. *Counterfactuals and Causal Inference: methods and principles for social research*, 2nd edition. Chapter 1.
- Stephen, Morgan & Christopher Winship. 2018. *Counterfactuals and Causal Inference: methods and principles for social research*, 2nd edition. Chapter 2.
- Pierotti, Rachel. (2013). “Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion.” *American Sociological Review*, 78: 240-265.
- Shelley, Clark et al. 2019. “[The Impact of Childcare on Poor Urban Women’s Economic Empowerment in Africa](#)”. *Demography*, 56, 4.

- Codes

- <https://jrnold.github.io/qss-tidy/discovery.html#textual-data>
- <https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html>

Semaine 2 : méthodologie des sciences sociales computationnelles (en présentiel)

À la fin de cette semaine, vous devriez être en mesure de faire ce qui suit :

- Décrire les principaux composants du traitement du langage naturel.
- Savoir faire le prétraitement des données textuelles
- Savoir-faire l'analyse de sentiments, la modélisation des termes et la modélisation structurelle des termes
- Expliquer comment l'analyse des réseaux sociaux et la pensée relationnelle diffèrent des approches des sciences sociales quantitatives
- Comparer les différentes manières de représenter les relations dans les données de réseau (par exemple, dirigées ou non dirigées, signées ou non signées).
- Travailler avec des données relationnelles ;
- Expliquer les défis du consentement éclairé dans les sciences sociales informatiques.
- Décrire les tensions entre les principes éthiques de la vie privée et de la transparence.
- Expliquer comment les biais algorithmiques et les ensembles de données d'entraînement biaisés peuvent amplifier et exacerber les inégalités.
- Identifier les types de sciences sociales computationnelles que vous ferez et ne ferez pas.

Jour	9h00 – 12h30	12h30 – 13h30	13h30 – 16h00	16h00 – 17h30
Lundi	Collecte de données digitales	Lunch (formation flash)	Travail en groupe	Conférencier invité
Mardi	Analyse de textes	Lunch (formation flash)	Travail en groupe	Conférencier invité
Mercredi	Analyse de textes	Lunch (formation flash)	Travail en groupe	Conférencier invité
Jeudi	Enquête à l'ère digitale + Éthique	Lunch (formation flash)	Travail en groupe	Conférencier invité
Vendredi	Apprentissage automatique	Lunch (formation flash)	Travail en groupe	Conférencier invité
Samedi	Analyse de réseaux	Lunch (formation flash)	Travail en groupe	Conférencier invité

Lectures

- David Lazer and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data". Annual Review of Sociology 43:19–39. doi:10.1146/annurev-soc-060116-053457.
- Christopher Bail. 2014. "The cultural environment: measuring culture with big data". Theory and Society 43 (3-4): 465–482. doi:10.1007/s11186-014-9216-5.

- <https://jrnold.github.io/qss-tidy/discovery.html#textual-data>
- <https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html>

Semaine 3 : Projet de recherche (en présentiel)

Les participants vont participer à des projets de recherche de groupe ou individuels dans lesquels ils feront de la recherche réelle dans un cadre de collaboration avec d'autres participants sous la supervision des organisateurs. Tout au long des deux premières semaines, les étudiants vont réfléchir à des sujets de projet de recherche de groupe possibles pendant les exercices de l'après-midi et via des conversations avec les professeurs invités (soit en personne ou via un canal Slack qui relie tous les participants). Nous encourageons également les professeurs invités à identifier des orientations prometteuses pour l'avenir du domaine au cours de leurs conférences qui pourraient être poursuivies par les équipes de recherche au cours de la dernière semaine. L'objectif global de ces projets de groupe est que les participants acquièrent des compétences essentielles dans la conception de la recherche, la collecte, le nettoyage et le traitement des données par eux-mêmes. Ces groupes travailleront ainsi sur leurs propres données, ou à partir des données que nous mettrons à leur disposition. Nous encourageons les étudiants à viser des recherches qui seront éventuellement publiées.

	9h00 – 12h30	12h30 – 13h30	13h30 – 16h00	16h00 – 17h30
Lundi	Formation des groupes		Projet de groupe	Projet de groupe
Mardi/ven dredi	Projet de groupe	Lunch	Projet de groupe	Projet de groupe
Vendredi	Présentation des travaux de groupes	Clôture		

- Conférenciers confirmés

	First Name	Last Name	Email	Institution	Place
1	Matthews	Salganik	mjs3@princeton.edu	Princeton	USA
2	Yannik	Dufresne	yannick.dufresne@pol.ulaval.ca	Université Laval (Québec)	Québec

3	Benjamin	Guinaudeau	https://benguinaudeau.com/	University of Konstanz (Germany)	Québec
4	Nicole	Denier	nicole.denier@ualberta.ca	Université d'Alberta	Edmonton
5	Hervé	Mensanh		Autodesk	Québec
6	Catherine	Ouellet		Université de Montréal	Québec
7	Evelyne	Brie		Western Ontario	

PRÉCISIONS AU SUJET DE L'ENTENTE D'ÉVALUATION À L'UQAM

Une entente est signée par l'enseignant. et les étudiant.e.s du cours dans les deux premières semaines du cours. Cette entente porte uniquement sur les aspects suivants :

1. Le nombre et les échéances des évaluations ;
2. La pondération respective des contenus ou objets d'évaluation dans l'évaluation globale du cours

Pour tout renseignement supplémentaire, il est possible de consulter les documents suivants :

- Pour le premier cycle : Règlement 5 de l'UQAM, article 7.9 https://instances.uqam.ca/wp-content/uploads/sites/47/2017/12/REGLEMENT_NO_5.pdf
 - Pour les cycles supérieurs : Règlement 8 de l'UQAM, article 9.4 https://instances.uqam.ca/wp-content/uploads/sites/47/2017/12/REGLEMENT_NO_8.pdf
-