

PLAN DE COURS

**Département de sociologie
Université du Québec à Montréal**

Sigle : SOC8655 **Groupe :** 010

Titre : Méthodologies quantitatives avancées en sociologie

Session : Automne 2023
Période : Lundi de 14 h à 17 h

Enseignant : Vissého Adjiwanou
Téléphone : (514) 987-3000 poste 0900
Bureau : A-5260
Courriel : adjiwanou.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

SOC8655 : Méthodes (computationnelles et) quantitatives avancées en Sociologie

Vissého Adjiwanou, PhD.

<u>Année:</u>	Automne 2023
<u>Période:</u>	
Cours :	Lundi, 14:00 – 16:00
Lab.:	Lundi, 16:00 – 17:00
<u>Évaluations</u>	
Quatre devoirs	Voir date de remise des travaux sur le calendrier
Participation	
Examen final	18 décembre
<u>Pondération:</u>	
Devoirs (4):	50% (10%, 10%, 15%, 15%)
Participation	10%
Projet final:	40%
Professeur:	Dr. Vissého Adjiwanou
Email:	adjiwanou.visseho@uqam.ca
Aide:	Mercredi, 4pm – 6pm (sujet à changement)
Assistant d'enseignement:	
Email:	

1. Description

Le paysage de l'analyse quantitative en sciences sociales tel que nous le connaissons est en train de changer drastiquement devant nous avec d'une part, la multitude des données, notamment des données numériques et administratives qui deviennent de plus en plus disponibles, et d'autre part, le développement de nouvelles méthodes d'analyses issues des sciences computationnelles pour traiter ces données. Ces changements conduisent à une redéfinition de la nature même des recherches que le/la chercheur.e en sciences sociales est amené.e à faire. Le cours SOC8655 - Méthodes quantitatives avancées en Sociologie - ambitionne de vous introduire à ces changements et de vous former à prendre plein contrôle de ces nouvelles méthodes. Il vise à renforcer l'utilisation des données sociales et démographiques, quelle que soit leur nature, à des fins d'analyse explicative.

Ce cours est structuré en trois parties. La première partie introduit le logiciel R et l'interface RStudio aux étudiant.es. Il s'agit du logiciel qui sera utilisé tout au long du cours. Cette partie peut vous sembler un peu difficile si vous n'avez jamais travaillé avec un logiciel statistique (de programmation). Je vous conseille de vous mettre à jour rapidement avant le début des cours avec les formations en ligne que je vous ai indiqué au point 3.2.4. La deuxième partie présente les concepts fondamentaux de la modélisation statistique et traite de leurs problèmes (causalité, mesure, échantillonnage) et de leurs solutions. Elle se poursuit par la présentation des notions de relation entre les variables et de leur évaluation. Je fais un survol rapide de l'analyse descriptive, suivi d'une série de modèles de régression et de leurs extensions. Dans la dernière partie du cours, je présente de nouveaux types de données disponibles pour les chercheurs en sciences sociales, principalement les données de Twitter et de

Facebook et quelques exemples d'analyses pouvant être conduites avec ces données (analyses textuelles, analyses de réseaux).

2. Objectifs

À la fin du cours, l'étudiant.e sera capable de :

- i. Comprendre quels sont les types de données en sociologie et en démographie et les problèmes qui s'y rattachent ;
- ii. Comprendre les problèmes conceptuels et empiriques présents dans les études empiriques ;
- iii. Appliquer un jugement critique et objectif sur la validité de la méthodologie d'un article scientifique ;
- iv. Comprendre la différence entre corrélation et causalité ;
- v. Conduire une analyse descriptive avec une visualisation appropriée des résultats ;
- vi. Savoir utiliser les méthodes statistiques les plus usuelles utilisées en sociologie (régression linéaire, régression logistique) et leurs extensions (modèles multiniveaux, modèles de survie...) ;
- vii. Comprendre et utiliser les nouvelles approches d'analyse des données numériques en sciences sociales ;
- viii. Utilisez R pour l'analyse des données.

3. Matériels

Vous trouverez à ce [lien](#), un récapitulatif de tout ce que vous devez installer sur votre machine avant le début du cours et qui sont décrits en détail ci-après.

3.1. Préalable:

Vous devez avoir suivi le cours SOC2206, ou le ou les cours équivalents pour pouvoir suivre ce cours. Ce cours suppose donc beaucoup de connaissances au préalable. Si vous pensez avoir oublié les notions de base, je vous conseille de lire rapidement le livre : Fox, W. 1999. Statistiques sociales. Les Presses de l'Université Laval. Traduit de l'Anglais et adapté par L.M. Imbeau, disponible à la bibliothèque de l'UQAM. Par ailleurs, vous allez vous en rendre compte tout au long de la session, je vous apporterai le soutien dont vous avez besoin pour réussir ce cours, du moment où vous faites l'effort nécessaire. Vous pouvez accéder à ce cours par ces deux liens ;

- <https://ena01.uqam.ca/enrol/instances.php?id=42197>
- <https://ena01.uqam.ca/course/view.php?id=42197>

3.2. Logiciels

La meilleure façon et souvent la seule, d'apprendre de nouvelles procédures statistiques est de les appliquer. Le cours fera recours au logiciel R. R est un logiciel libre et largement utilisé en sciences sociales dans les grandes universités. R est probablement le logiciel statistique le plus utilisé. Nous allons utiliser R avec RStudio. Je vous ai inscrit à un cours en ligne pour apprendre les rudiments de ce logiciel (voir point 3.2.4).

Pour pouvoir utiliser ce logiciel, nous allons recourir à RStudio qui est un environnement de développement intégré. Finalement, RStudio 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 LaTeX¹. Toutefois, vous pouvez aussi tout simplement installer le package tinytex pour être en mesure de sortir vos résultats en version pdf : <https://bookdown.org/yihui/rmarkdown/installation.html>.

3.2.1. Installation de R

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

3.2.2. 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/>

3.2.3. 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.

Les étudiants de mon cours SOC2206 ont utilisé cette année le logiciel BlueSkyStatistics qui est une interface intégrée (Graphical User Interface) pour utiliser R (similaire à SPSS). Vous pouvez l'installer à partir de ce lien : <https://www.blueskystatistics.com/Articles.asp?ID=317>. Si son utilisation est simple, il ne nous permet d'aller au-delà de l'analyse quantitative pour couvrir l'analyse de texte. Aussi, allons-nous utiliser directement RStudio dans ce cours.

3.3. DataCamp

Datacamp est une plateforme d'apprentissage en ligne qui sera utilisée tout au long de la session pour appuyer votre apprentissage. Vous n'avez pas à payer pour les cours qui y sont dessus, je m'en suis déjà chargé.

Vous allez recevoir ou avez probablement reçu un mail pour vous inscrire, suivez les instructions dans ce mail pour le faire (regarder dans votre « courriel indésirable » si vous n'avez pas reçu ce mail). Vous pouvez aussi utiliser ce lien avec les comptes (courrier.uqam.ca, uqam.ca ou gmail.com) :

https://www.datacamp.com/groups/shared_links/52bbf27885a77b61cfe0261318f98b52833044be94de019f28a6c130c36ae317

Une fois inscrit, compléter le chapitre “Introduction to R”. N'hésitez pas à me contacter si vous avez des questions. Cela vous prendra 8h pour le compléter, mais c'est un 8h bien investi. Faites-le à votre rythme, **mais assurez-vous d'avoir compris ce que vous faites**. Réessayer au besoin plusieurs fois. Ce que vous

¹ Tout ceci vous sera précisé à nouveau lors du premier cours, si vous n'avez pas pu les installer

allez apprendre dans ce premier cours va vous suivre tout au long de la session. Si vous avez des lacunes, vous aurez de la difficulté tout le long. Il y a des étudiant.es qui se permettent de ne pas suivre cette formation en ligne. La conséquence est qu'il finisse rarement la session. Alors, je vous conseille vivement de le faire et plusieurs fois jusqu'à ce que vous compreniez la logique de R.

Remarque : Datacamp met à votre disposition une plateforme pour écrire vos commandes. En classe, nous allons utiliser RStudio pour le faire.

3.4. Slack

Slack est un forum de discussion en classe où vous pouvez poser des questions sur le matériel, interagir avec vos camarades. Vous pouvez aussi poser directement vos questions sur Moodle. Sachez que vos questions peuvent aider d'autres personnes aussi à comprendre la matière. De même, d'autres personnes peuvent avoir de meilleures réponses à vos questions. Slack dispose d'une interface bureau ainsi que d'une application pour votre téléphone. Voici le lien pour joindre le Slack du cours :

https://join.slack.com/t/soc8655a23/shared_invite/zt-21vvzp6aa-Wzqw0UTz8Wp9LR1Pyce7Kw

3.5. Références

3.5.1. Obligatoire

Les trois références obligatoires du cours sont :

<https://www.sea-stat.com/wp-content/uploads/2020/12/Springer-Texts-In-Statistics-Richard-A.-Berk-Statistical-Learning-From-A-Regression-Perspective-2020-Springer-libgen.lc.pdf>

1. Salganik, Matthews. 2017. "[Bit by bit: Social research in the digital age](#)". Princeton University Press. Vous pouvez accéder gratuitement à ce livre en ligne. Bien que nous ne couvrions pas l'ensemble des chapitres de ce livre, c'est en soit un livre à lire complètement pour tout.e sociologue sérieux/sérieuse qui conduit de la recherche dans cette ère digitale. Vous pouvez l'acheter ou le louer sur Perusall. Vous pouvez le lire aussi directement sur Perusall, j'ai mis les chapitres là-bas pour vous, si vous ne voulez pas l'acheter.
2. John, Fox & Sanford, Weisberg. 2019. An R Companion to Applied regression, 3rd edition. Sage. Ce livre utilise "Base R" pour les codes alors que dans le cours, je vais vous apprendre la version tidyverse, une version plus moderne d'écritures de codes. Donc, en travaillant avec les codes du livre, ce serait une bonne idée de traduire les codes en version "Tidyverse".
3. Wickham, Hadley & Grolemund, Garrett. 2017. "[R for Data Science: Import, Tidy, Transform, Visualize, and Model data](#)". Boston. O'Reilly. Pp.492. Ce livre est aussi gratuitement accessible en ligne à l'adresse suivante : <https://r4ds.had.co.nz/>. Dans ce livre, tous les codes sont écrits en tidyverse.

3.5.2. Autres références

En statistiques:

- Vincent Arel-Bundock. 2021. Analyse causale et méthodes quantitatives : Une introduction avec R, Stata et SPSS. Presses de l'Université de Montréal.
- Jean-Herman, Guay. 2014. Statistiques en sciences humaines avec R, 2e édition. Presses de

l'Université de Laval.

- Kosuke Imai. 2017. Quantitative social science : An introduction. Princeton University Press.
- Tabachnick, B. et Fidell, L. 2013. Using Multivariate Statistics, 6th Edition. Pearson.
- Babbie, Earl. 2015. The Practice of social research. 14th Edition. Belmont, CA: Wadsworth. Version française en ligne.
- Kennedy, Peter. 2008. A Guide to Econometrics. Malden, MA: Blackwell Publishing.

En sciences des données

- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as Data*. Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691207544/text-as-data>
- Silge, Julia, and David Robinson. 2017. Text Mining with R: A Tidy Approach. O'Reilly Media. [disponible à <https://www.tidytextmining.com/>]
- Healy, Kieran. 2018. Data Visualization: A Practical Introduction. Princeton University Press. Disponible gratuitement en ligne ici : <https://socviz.co/>

En Sociologie digitale :

- Noortje Marres. 2017. Digital Sociology: The Reinvention of Social Research Polity, Cambridge: Policy press
- Neil Selwyn. 2019. What is Digital Sociology? Wiley

3.5.3. Ressources en ligne

Vous vous en rendrez compte assez rapidement qu'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, une simple recherche dans google.com vous sera d'un formidable atout. Par exemple, écrire dans le moteur de recherche "create new variable, tidyverse" vous indiquera des pages où vous trouverez des réponses à votre question. Ne vous attardez pas à tout lire. Chercher d'abord à lire les exemples pour vous s'assurer que c'est exactement ce que vous voulez.

En français :

Les séries intéressantes de Larmarange :

- <https://larmarange.github.io/analyse-R/presentation-et-philosophie.html>
- <https://larmarange.github.io/guide-R/>
- <https://larmarange.github.io/webin-R/seances.html>
- <https://juba.github.io/tidyverse/index.html>
- <http://dimension.usherbrooke.ca/dimension/v2ssrcadre.html>
- Vincent Arel-Bundock : Analyse causale et méthodes quantitatives Une introduction avec R, Stata et SPSS : <https://www.pum.umontreal.ca/catalogue/analysecausaleetmethodesquantitatives>

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.

4. Mode d'évaluation

Trois modes d'évaluation sont mis en place dans le cadre du cours : des lectures obligatoires suivies de travaux pratiques, des devoirs de maison, et l'examen final.

4.1 Participation (10%)

Vous avez à faire le résumé de deux lectures assignées à vous et le soumettre la veille de chaque cours (voir section sur Calendrier). Cela facilitera largement votre compréhension. Vous devez obligatoirement lire ces matériels. Dix pour cent (10%) de la note finale est attribuée à votre engagement avec ces matériels.

4.2 Travaux de maison (50%)

Vous aurez quatre (4) devoirs à faire à la maison et à rendre au bout de deux semaines (sauf le dernier, voir calendrier). Les devoirs seront un mélange de problèmes analytiques, de simulations informatiques et d'analyse de données. Chaque devoir sera marqué de 0 à 100. Les devoirs n'auront pas forcément la même pondération. Les devoirs doivent être complétés dans R Markdown, ce qui vous permet d'afficher à la fois vos réponses et le code que vous avez utilisé pour les obtenir. Ne vous inquiétez pas si vous ne connaissez pas RMarkdown, je vais vous montrer comment cela fonctionne. Pour les curieux, voici une description succincte ici : <https://data.hypotheses.org/1144>

Je répondrai à toutes vos questions concernant les devoirs au cours de la première semaine de remise de ces devoirs. Vous pouvez aussi discuter avec vos camarades à propos des devoirs, mais comprenez que c'est un travail personnel et vous devez rendre un travail personnel, dénué de plagiat et qui reflète votre propre effort. A cet effet, je vous conseille de poser des questions ouvertes à vos camarades (exemple : peux-tu me ré expliquer cette partie du code, cours, etc...) et pas des questions très spécifiques (exemple : comment avez-vous répondu à la question x ?). Si vous avez demandé de l'aide à quelqu'un de la classe, vous devez indiquer le nom de ces personnes sur votre copie. Vous devez indiquer aussi les noms de ceux que vous avez aidés. Les étudiant.es qui rendront des travaux semblables seront sévèrement pénalisés.

4.3 Projet final (40%)

Finalement, vous aurez à soumettre un projet final de 7500 mots sur un sujet de votre choix. Idéalement, ce projet doit couvrir les deux parties du cours. Vous aurez jusqu'au **5 janvier** pour soumettre ce projet. Ce projet peut être fait individuellement ou par binôme. Voici les étapes et échéances à respecter pour la soumission du projet :

1. Choix du sujet : Vous
2. Problématique
3. Brève revue de littérature
4. Méthodologie
5. Analyse des données
6. Présentation des résultats

Ce projet final comptera pour 40% de la note finale.

4.4. Note finale

Votre note finale sera une moyenne pondérée de votre note du projet final (40%), des séries de devoirs (50%), et de vos lectures, présentations et travaux pratiques (10%). Je me réserve le droit de donner un bonus pour une participation active à l'intérieur et à l'extérieur de la classe. Par exemple, un.e étudiant.e

qui aide activement ses camarades de classe, en répondant aux questions ou qui s'engage de manière productive en classe pourrait avoir droit à une petite prime.

5. Comment apprendre dans ce cours

Si vous trouvez ce cours difficile, vous n'êtes pas seul.e. Les statistiques peuvent être difficiles et nous couvrons beaucoup de matières. Cependant, je suis convaincu que vous pouvez y faire face. Dans cette section du syllabus, je vais donner des détails sur certaines des formes de soutien que je propose dans cette classe.

Votre principale responsabilité dans cette classe est de **travailler dur** et de **communiquer avec moi** au sujet de vos besoins. Vous ne pouvez pas apprendre si vous ne prenez pas le temps nécessaire pour étudier. Je ne peux pas vous aider si vous ne me posez pas de questions ou me parlez de vos défis pour le cours.

Le cours est conçu pour fournir tous les outils auxquels je peux penser pour vous aider à apprendre les matériels. Si vous êtes prêt à y investir du temps, je veux vous assurer que ce temps est utilisé le plus efficacement possible.

6. Ressources pour obtenir de l'aide

Il existe quelques sources principales de soutien dans la classe.

i. Classe et cours en ligne sur datacamp

Je vous encourage fortement à participer activement en classe et à suivre les capsules vidéo en ligne qui vous seront assignées. Posez des questions pendant le cours si vous ne comprenez pas ce qui se passe. Il n'est même pas nécessaire de poser une question précise : il suffit de lever la main et de me faire savoir que vous ne suivez pas. Je suis toujours content d'arrêter et de revenir en arrière.

ii. Lectures et diapositives

Si vous étudiez seul.e et que vous rencontrez quelque chose que vous ne comprenez pas, votre premier réflexe devrait être **d'étudier les notes de cours et les diapositives**. Il y a beaucoup d'information dans les diapositives et ils sont destinés à être examinés plusieurs fois, pas seulement vus une fois pendant le cours.

iii. Heures de bureau

Mes heures de bureau sont chaque fois que ma porte est ouverte. Vous pouvez venir me parler à tout moment de la classe. Si vous avez besoin de planifier une heure particulière pour me voir, contactez-moi par courrier électronique et nous trouverons une heure qui convient.

iv. Devoir de maison

Dès que le devoir est publié en ligne, n'attendez pas avant de commencer. Et ne perdez pas trop de temps à chercher seul ce que vous ne comprenez pas. Au début, je vous conseille vivement de me poser vos questions une fois que vous êtes bloqués plus de 30 minutes sur des problèmes de codes qui ne fonctionnent pas. N'oubliez pas de vous référer au point 2 ci-dessus. Les notions que vous apprenez dans un devoir peuvent revenir dans un autre devoir, apprenez à maîtriser ce que vous faites, et à refaire les devoirs si possible.

v. Séminaires

J'organiserai des séminaires méthodologiques les vendredis (ou jeudi dans la soirée) tout au long de la session pour vous apprendre la programmation sous R ou d'autres compétences utilisées en méthodes quantitatives. J'inviterai aussi d'autres professeurs ou chercheurs à venir animer ces séminaires. Ce seront des séminaires pratiques. Si vous jugez que vous avez des compétences que vous voulez partager avec nous, contactez-moi pour que je vous inscrive pour ce séminaire.

C'est beaucoup de ressources, mais si vous pouvez penser à autre chose qui vous serait utile pour passer du bon moment ensemble, je vous encourage à venir me parler. Encore une fois, si vous êtes prêt à vous y investir, je m'efforcerai à vous accompagner du mieux de mes compétences.

vi. Laboratoire de méthodologie

Le laboratoire est fonctionnel l'automne 2023. Il y a un étudiant qui assure la permanence qui peut aussi répondre à vos multiples questions.

vii. Quelques conseils qui peuvent vous être utiles

Chaque année, je demande aux étudiant.es de prodiguer des conseils aux futures générations d'étudiant.es. Voici quelques conseils d'ancien.nes étudiant.es de premier cycle qui ont répondu «Quel conseil donneriez-vous à un.e autre étudiant.e qui envisage de suivre ce cours?». Je pense que le conseil est excellent et qu'il peut être utile à vous aussi. Ces conseils proviennent de mon autre cours. À la fin de la session, je vous poserai la même question.

- Être prêt à passer beaucoup de temps sur la matière.
- Posez des questions si vous ne savez pas ce qui se passe.
- Étudiez et travaillez fort, passez en revue les diapositives.
- Investir un temps considérable pour se familiariser avec R et ses divers outils sera rentable à long terme.
- Passez en revue les diapositives de la matière chaque semaine.
- C'est un défi, mais très faisable et gratifiant si vous y consacrez du temps. Il existe de nombreuses ressources à exploiter.
- Prenez ce cours, vous ne le regretterez pas.
- Ce sera difficile mais vous apprendrez beaucoup.
- Veuillez suivre ce cours pour intégrer les statistiques aux sciences sociales. J'ai trouvé le cours très pratique (beaucoup de R) et complet. La petite taille des classes est également une excellente ressource pour obtenir une assistance individuelle dans un cours très quantitatif. De plus, si vous envisagez des études supérieures ou de la recherche, ce cours est efficace pour intégrer des questions de recherche tout au long.
- Ce cours est très stimulant mais a grandement contribué à ma compréhension des statistiques sociales. Si vous êtes vraiment investi dans le sujet et êtes prêt à travailler (plus que ce à quoi vous vous attendiez) éventuellement), ce sera l'un des meilleurs cours que vous aurez suivi.
- C'est un cours dans lequel vous apprendrez beaucoup et passerez le plus clair de votre temps à faire les codes.

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

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

- le nombre et les échéances des évaluations;
- 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 les cycles supérieurs: Règlement 8 de l'UQAM, article 9.4
<https://instances.uqam.ca/reglements/>

7. Calendrier

Je vous donnerai plus d'indications sur les lectures plus tard. Chaque étudiant aura à faire un résumé de deux articles chaque cours. Cette programmation est ce que je compte couvrir dans le cadre de ce cours. Je sais que c'est assez de matériels. Aussi, suis-je assez flexible pour l'amender en fonction de vos besoins et de votre compréhension de la matière. Je vous conseille vivement de travailler en binôme pour vous faciliter la tâche.

Sem	Date	Séance	Chapitre	Exercices* / Devoirs
PREMIÈRE PARTIE: Introduction à R et notions fondamentales				
1	11 sept	<p>Séance 1: Introduction</p> <ul style="list-style-type: none"> S1.1 : Introduction (Vidéo & PowerPoint) S1.2 : Présentation du cours (Vidéo & PP) S1.3 : Introduction à R Labo 1.1 : Installation des différents logiciels (A faire avant le cours) Labo 1.2 : Introduction à R et à RMarkdown (Cour sur Datacamp) 	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p.1-32, 47-52) Salganik (https://www.bitbybitbook.com/fr/1st-ed/introduction/) Joshua Blumenstock : Predicting poverty and wealth from mobile phone metadata (A discuter en classe) 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/do/cs/50YearsDataScience.pdf. (Optionnel) <p>Applications</p> <ul style="list-style-type: none"> Wickham (https://r4ds.had.co.nz/introduction.html) Wickham (https://r4ds.had.co.nz/r-markdown.html) https://socviz.co/gettingstarted.html https://rmarkdown.rstudio.com/lesson-1.html https://juba.github.io/tidyverse/01-presentation.html https://juba.github.io/tidyverse/02-priseenmain.html 	Exercice 1
2	18 sept	<p>Séance 2 : Survol de l'analyse descriptive avec tidyverse et summarystools</p> <ul style="list-style-type: none"> Signature de l'entente S2.1 : Introduction à Tidyverse (Cour sur Datacamp) S2.2 : Analyse univariée (tableau de fréquence) S2.3 : Paramètres de tendance centrale et de dispersion Labo 2 : Article de Pierotti 	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p33-46, 53-106) https://www.bitbybitbook.com/en/1st-ed/asking-questions/ (A présenter en classe) Lecture : Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion" American Sociological Review, 78: 240-265. (A présenter en classe) 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. (Optionnel) <p>Applications</p> <ul style="list-style-type: none"> Wickham (https://r4ds.had.co.nz/transform.html) https://cran.r-project.org/web/packages/summarytools/vignettes/Introduction.html https://juba.github.io/tidyverse/03-premiertravail.html https://juba.github.io/tidyverse/06-tidyverse.html https://juba.github.io/tidyverse/07-import.html 	Exercice 2

3	25 sept	<p>Section 3 : Visualisation avec ggplot</p> <ul style="list-style-type: none"> S3.1: Visualisation de statistiques descriptives univariées S3.2 : Visualisation de statistique descriptive bivariable Labo 3 	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p123-172) Kieran (https://socviz.co/lookatdata.html#what-makes-bad-figures-bad) (A présenter en classe) <p>Applications</p> <ul style="list-style-type: none"> https://juba.github.io/tidyverse/08-ggplot2.html Wickham(https://r4ds.had.co.nz/data-visualisation.html) Wickham(https://r4ds.had.co.nz/exploratory-data-analysis.html) 	Soumission de votre sujet de recherche (1 page d'introduction – 10%)
3	25 sept	Séance 3 bis : Mesure : réduction dimensionnelle	<p>http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/</p> <p>http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/76-afdm-analyse-factorielle-des-donnees-mixtes-avec-r-l-essentiel/</p>	
4	2 oct	<p>Séance 4 : Causalité</p> <ul style="list-style-type: none"> S4.1 : A partir de données expérimentales Labo 4.1 : Discrimination S4.2 : A partir de données observationnelles Labo 4.2 : Salaire minimum et chômage 	<p>Lectures</p> <ul style="list-style-type: none"> Salganik(https://www.bitbybitbook.com/en/1sted/running-experiments/) (A présenter en classe) Athey, Susan, 2017. Beyond prediction: Using big data for policy problems. Science, 355(6324), pp.483-485 Shmueli, G. (2010). To Explain or to Predict? Statistical Science, 25(3), 289–310. https://doi.org/10.1214/10-STS330 (A présenter en classe) Mirna Safi et Patrick Simon. 2013. "Les discriminations ethniques et raciales dans l'enquête Trajectoires et Origines: représentations, expériences subjectives et situations vécues." Économie et Statistiques, No 464-465-466. (A présenter en classe) Shelley, Clark et al. 2019. "The Impact of Childcare on Poor Urban Women's Economic Empowerment in Africa". Demography, 56, 4. (A présenter en classe) 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. Ellen L. Hamaker , Jeroen D. Mulder, Marinus H. Van IZendoorn, 2020, Description, prediction and causation: Methodological challenges of studying child and adolescent development Ressource: http://arelbundock.com/acmq.html 	Devoir 1
5	9 oct	Action de Grâce		
DEUXIÈME PARTIE: Méthodes usuelles d'analyse des données				
6	16 oct	Séance 6 : Régression linéaire	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p173-190) 	

			<ul style="list-style-type: none"> Fox (197-207) Fox (235-241) Fox (243-246) Fox (254-255) <p>Applications</p> <ul style="list-style-type: none"> https://r4ds.had.co.nz/functions.html 	
7	23 oct	Semaine de relâche		
8	30 oct	Séance 8 : Régression linéaire (Fin)	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p190-197) Fox (207-224) Fox (232-235) Fox (246-254) Fox (p258-270) Fox (p385-417) Fox (p429-436) <p>Applications</p> <ul style="list-style-type: none"> https://purrr.tidyverse.org/ 	Soumission de votre revue de littérature (15%)
9	6 nov	Séance 9 : Régression logistique	<p>Lectures</p> <p>Fox (p271-296)</p>	Devoir 2
10	13 nov	Séance 10 : Régression logistique (fin)	<p>Lectures</p> <ul style="list-style-type: none"> Fox (p309-314) Fox (p417-425) Adjiwanou, V. al. 2019. The effect of partners' education on women's reproductive and maternal health in developing countries. <i>Social. Social Science & Medicine</i>, V. 197, January 2018, Pages 104-115 doi: 10.1016/j.socscimed.2017.11.054 (A présenter en classe) 	
11	20 nov	Séance 11 : Modèle de régression logistique multinomial et ordonné		
12	27 nov	<p>Séance 12 : Extension des modèles de régressions classiques</p> <ul style="list-style-type: none"> Modèles multiniveaux Modèles de risques et durée Modèles basés sur l'intelligence artificielle 	<p>Lectures</p> <ul style="list-style-type: none"> Athey, S., & Imbens, G. W. (2017). The State of Applied Econometrics: Causality and Policy Evaluation. <i>Journal of Economic Perspectives</i>, 31(2), 3–32. https://doi.org/10.1257/jep.31.2.3 Salganik, M, al. 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. <i>PNAS</i> 117 (15) 8398-8403 https://doi.org/10.1073/pnas.1915006117 Bruno Arpino, Marco Le Moglie, Letizia Mencarini, (2022); What Tears Couples Apart: A Machine Learning Analysis of Union Dissolution in Germany. <i>Demography</i> 1 February 2022; 59 (1): 161–186. doi: https://doi.org/10.1215/00703370-9648346 Hernán, M. A., Hsu, J., & Healy, B. (2019). A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. <i>CHANCE</i>, 32(1), 42–49. 	Devoir 3

			https://doi.org/10.1080/09332480.2019.1579578 Application https://app.datacamp.com/groups/soc8655_a22/assignments/185330 https://app.datacamp.com/groups/soc8655_a22/assignments/185332 https://app.datacamp.com/groups/soc8655_a22/assignments/185331	
TROISIÈME PARTIE: Données digitales et analyses textuelles				
13	4 déc	Séance 11 : Données digitales Labo 11.1 : Collecte de données digitales Labo 11.2 : données textuelles	Lectures <ul style="list-style-type: none"> • Fox (107-117) • Ollion, É., & Boelaert, J. (2015). Au delà des big data. Les sciences sociales et la multiplication des données numériques. Sociologie, 6(3), 295–310. • Vidéo à suivre: https://www.youtube.com/watch?v=sOaPTrhBuU David Lazer and Jason Radford. 2017. “ Data ex Machina: Introduction to Big Data ”. Annual Review of Sociology 43:19–39. (optionnel) Applications <ul style="list-style-type: none"> • https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html • Wickham (https://r4ds.had.co.nz/strings.html) 	Soumission de votre méthodologie et plan d’analyse (15%)
14	11 déc	Séance 12 : Analyse textuelle S12.1: Analyse de sentiment (Sentiment analysis) Évaluation du cours	Lectures <ul style="list-style-type: none"> • Evans, James, and Pedro Aceves. 2016. “Machine Translation: Mining Text for Social Theory.” Annual Review of Sociology 42(1): 21-50 (A présenter en classe) • Grimmer, Justin, and Brandon Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Documents.” Political Analysis 21(3): 267-97 • Grimmer, Justin. 2015. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.” PS: Political Science & Politics 48(1): 80-83 (optionnel) • Wallach, Hanna. 2018. “Computational Social Science 6 = Computer Science + Social Data.” Communications of the ACM 61(3): 42-44 (optionnel) • DiMaggio, Paul. 2015. “Adapting Computational Text Analysis to Social Science (and Vice Versa).” Big Data & Society 2(2): 1-5 (optionnel) Applications	Devoir 4

			<ul style="list-style-type: none"> Silge (https://www.tidytextmining.com/sentiment.html) 	
15	18 déc	<p>Séance 15 : Analyse textuelle</p> <p>S13.1: Analyse des terms (topic and structural topic modelling)</p> <p>Évaluation du cours</p>	<p>Lectures</p> <ul style="list-style-type: none"> Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." Sociological Methods & Research. Vol.49(1), 3-42. (A presenter en classe) Baumer, Eric P. S., David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. "Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence?" Journal of the Association for Information Science and Technology 68(6): 1397-1410. (A presenter en classe) Blei, David M. 2012. "Probabilistic Topic Models." Communications of the ACM 55(4): 77-84. <p>Applications</p> <ul style="list-style-type: none"> Slige (https://www.tidytextmining.com/topicmodeling.html) 	Soumission du draft du projet

* Les exercices ne sont pas notés. Il s'agit de pratique pour vous permettre de comprendre ce que j'attends de vous.

8. Informations concernant le projet final (voir détails sur le github du cours) <https://github.com/visseho/SOC8655>

8.1 Sujets

- Gender - Worldbank

<https://blogs.worldbank.org/opendata/how-have-womens-legal-rights-evolved-over-last-50-years>

8.2 Base de données existantes

8.3 Collecter vos propres données

PLAGIAT

Règlement no 18 sur les infractions de nature académique

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes;
- l'utilisation totale ou partielle du texte d'autrui en le faisant passer pour sien ou sans indication de référence;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant;
- l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche;

Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Pour plus d'information sur les infractions académiques et comment les prévenir :

www.integrite.ugam.ca