

PLAN DE COURS

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

Sigle : SOC8655 **Groupe :** 10

Titre : Méthodologies quantitatives avancées en sociologie

Session : Automne 2020
Période : Lundi de 14 h à 17 h

Enseignant : Vissého Adjiwanou
Téléphone : (514) 987-3000 poste 0900
Bureau : A-5260
Courriel : adjivanou.vissého@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 2020
<u>Période:</u>	
Cours :	Lundi, 14:00 – 15:00
Lab.:	Lundi, 15:00 – 17:00

Évaluations

Six devoirs (dont 5 retenus):	Voir date de remise des travaux sur le calendrier
Examen final:	21 décembre

Pondération:

Devoirs (5):	50%
Exercices maison, module en ligne et participation	15%
Examen final:	35%

Professeur:	Dr. Vissého Adjiwanou
Email:	adjiwanou.visseho@uqam.ca
Aide:	Mercredi, 10:00 – 12 :00 (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 chercheur en sciences sociales est amené 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 étudiants. 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 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 et leurs extensions;
- vii. Comprendre et utiliser les nouvelles approches d'analyse sur les 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 avoir avant le début du cours et qui sont décrit 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 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.

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.

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.

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. Veuillez-vous inscrire à partir de ce lien :

https://datacamp.com/groups/shared_links/0eddbd267a31311936bff247a1a7e139714410dca5f40562c7a8aacb6d82cbf3

Je vous **encourage vivement** (dans la mesure de vos possibilités) à compléter le premier chapitre "Introduction à R" avant le début des classes. Cela vous prendra 8h, mais c'est un 8h bien investi. Ce premier cours vous introduit à ce qui est communément appelé « Base R ». Nous utiliserons davantage dans ce cours la nouvelle manière d'écrire des codes dans R qui s'appelle « Tidyverse ». N'hésitez pas à me contacter si vous avez des questions.

3.4. Perusall

J'ai créé un compte en ligne sur Perusall pour vos lectures. Perusall est une plate-forme avec annotation collaborative qui vous permet de poser et de répondre aux questions directement dans le texte que vous lisez ; des questions sur les parties du texte que vous ne comprenez pas. D'autres

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

étudiants du cours peuvent y répondre. Donc, poser de bonnes questions ne vous aide pas seulement, cela aide aussi vos camarades de classe. Si vous connaissez la réponse à une question posée par un autre étudiant, veuillez apporter une contribution à la classe et essayez d'y répondre! C'est la première fois que j'utilise cette ressource, j'en ai entendu plein de bonnes choses.

Pour accéder aux textes en ligne, vous devez créer un compte sur <https://perusall.com/> et joindre le cours avec ce code VISSEHO-RVPTJ.

3.5. 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/soc8655a20/shared_invite/zt-gedk79yv-fqQieJ4lt6bkGxSxf~7ZQ

3.6. Références

3.6.1. Obligatoire

Les trois références obligatoires du cours sont :

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 sociologue sérieux 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. Donc, en travaillant avec les codes du livre, ce serait une bonne idée de traduire les codes en version "Tidyverse".
3. Wickham, Hadley & Grolemond, 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.

3.6.2. Autres références

En statistiques:

- Jean-Herman, Guay. 2014. Statistiques en sciences humaines avec R, 2e édition. Presses de l'Université de Laval.
- Kosuke Imai, Quantitative social science: An introduction. Princeton University Press (2017).
- 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.
- 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

3.6.3. 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:

- <https://juba.github.io/tidyverse/index.html>
- <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://jrmold.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 Lectures d'articles et travaux pratiques sur Datacamp (15%)

Avant chaque cours, je vous conseille vivement de lire les articles ou sections 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. De même, je vais vous assigner des capsules vidéo à écouter sur Datacamp chaque semaine, qui seront évaluées. Quinze pour cent (15%) de la note finale est attribuée à votre engagement avec ces matériels.

4.2 Travaux de maison (50%)

Vous aurez six (6) devoirs à faire à la maison et à rendre à la veille de chaque nouveau chapitre. Les devoirs vous seront donnés le matin du cours, et vous devez les remettre au bout de 13 jours, c'est-à-dire à la veille du nouveau chapitre. 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. Je prendrai seulement la note de vos cinq meilleurs devoirs. 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 atteindre. 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 étudiants qui rendront des travaux semblables seront sévèrement pénalisés.

4.3 Examen final

Finalement, vous aurez un examen final sur table qui consistera en une partie théorique et une partie pratique. Cet examen final comptera pour 35% de la note finale.

4.4. Note finale

Votre note finale sera une moyenne pondérée de votre note de l'examen final sur table (35%), des séries de devoirs (50%), et de vos lectures et travaux pratiques (15%). Je me réserve le droit de fournir un bonus pour une participation active à l'intérieur et à l'extérieur de la classe. Par exemple, un étudiant 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. Déroulement du cours en ligne

Ce cours est donné entièrement en ligne. Voici comment nous allons nous organiser. Ceci peut changer en tout moment. Vous comprenez que nous sommes en mode essai et le but est de trouver la meilleure manière de vous engager pour le cours et de rendre votre apprentissage des plus satisfaisantes.

- i. **Avant le cours** : Poste de vidéo de cours avant chaque séance : Je vous enverrai une courte introduction sur les notions que je vais couvrir. Cette introduction sera de 1h. Vous devez écouter cette vidéo avant le début des cours. Vous devez obligatoirement lire aussi les sections des livres du cours qui sera assignée au cours de la semaine. Pour me montrer que vous avez lu ces matériels, vous devez les annoter sur Perusall. Je mettrai aussi à votre disposition les PowerPoint du cours. Finalement, s'il y a des exercices, vous devez les compléter et les soumettre au plus tard à la veille du cours.
- ii. **Durant le cours** : J'ai subdivisé les heures en deux parties (uniquement à partir de la séance 4, avant cela, on sera en ligne tous ensemble à partir de 1hh). De 14h00 à 15h00, vous pouvez me poser des questions sur les difficultés que vous avez eues, sur les notions qui ne sont pas claires, etc... Cette partie du cours est individuelle. Vous n'êtes pas nombreux, on peut se permettre de faire un suivi individuel. Dans la deuxième partie du cours, de 15h00 à 17h00, je donnerai le cours sous forme de résolution de problèmes. Pour bien suivre cela, vous devez bien sûr lire le cours à l'avance. Je supposerai acquis les notions, c'est pourquoi je consacrerai la première partie du cours à répondre à vos questions. Cette seconde partie du cours sera enregistrée et déposée sur Moodle. J'exigerai que tout le monde mette sa vidéo.

6. Comment apprendre dans ce cours

Si vous trouvez ce cours difficile, vous n'êtes pas seul. 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 parler 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.

7. 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. Commentaires quotidiens

Après chaque cours, je vais faire circuler une carte de correspondance et vous demander de noter quelque chose à propos de la classe. Vous pouvez écrire quelque chose que vous avez aimé ou que vous n'avez pas aimé. Quelque chose que vous voulez mieux comprendre ou que vous voulez en savoir plus. Vous voudrez peut-être savoir comment un élément du cours est relié aux objectifs plus larges de la classe. Vous pouvez même simplement dessiner un visage souriant. Je vais répondre aux questions soit en classe ou par mail ou sur Slack.

iii. Lectures et diapositives

Si vous étudiez seul 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.

iv. 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.

v. 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 3 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.

vi. 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 serait 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.

vii. Quelques conseils qui peuvent vous être utiles

Chaque année, je demande aux étudiants de donner des conseils aux futures générations d'étudiants. Voici quelques conseils d'anciens étudiants de premier cycle qui ont répondu «Quel conseil donneriez-vous à un autre étudiant 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 :

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/reglements/>
- Pour les cycles supérieurs: Règlement 8 de l'UQAM, article 9.4
<https://instances.uqam.ca/reglements/>

8. Calendrier

Je vous donnerai plus d'indications sur les lectures plus tard.

Se m	Date	Séance	Chapitre	Exercices / Devoirs
PREMIÈRE PARTIE: Introduction à R et notions fondamentales				
1	14 sept	<p>Séance 1: Introduction</p> <ul style="list-style-type: none"> S1.1 : Introduction (Vidéo & PowerPoint) S1.2 : Présentation du cours (Vidéo & PP) Signature de l'entente 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>Cours</p> <ul style="list-style-type: none"> Salganik (https://www.bitbybitbook.com/fr/1st-ed/introduction/) Fox(p.1-32, 47-52) Wickham (https://r4ds.had.co.nz/introduction.html) Wickham (https://r4ds.had.co.nz/r-markdown.html) 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>En ligne</p> <ul style="list-style-type: none"> https://rmarkdown.rstudio.com/lesson-1.html https://juba.github.io/tidyverse/01-presentation.html https://juba.github.io/tidyverse/02-prise_en_main.html 	Exercice 1
2	21 sept	<p>Séance 2 : Survol de l'analyse descriptive avec tidyverse et summarystools</p> <ul style="list-style-type: none"> 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>Cours</p> <ul style="list-style-type: none"> Fox (p33-46, 53-106) Wickham (https://r4ds.had.co.nz/transform.html) Lecture : Pierotti, Rachel. (2013). "Increasing Rejection of Intimate Partner Violence: Evidence of Global Cultural Diffusion" American Sociological Review, 78: 240-265. (A discuter 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>En ligne</p> <ul style="list-style-type: none"> https://cran.r-project.org/web/packages/summarytools/vignettes/Introduction.html https://juba.github.io/tidyverse/03-premier_travail.html https://juba.github.io/tidyverse/06-tidyverse.html https://juba.github.io/tidyverse/07-import.html 	Exercice 2
3	28 sept	<p>Section 3 : Visualisation avec ggplot</p> <ul style="list-style-type: none"> S3.1: Visualisation de statistiques 	<p>Cours</p> <ul style="list-style-type: none"> Kieran (https://socviz.co/lookatdata.html#what-makes-bad-figures-bad) 	Devoir 1

		descriptives univariées <ul style="list-style-type: none"> S3.2 : Visualisation de statistique descriptive bivariée Labo 3 	<ul style="list-style-type: none"> Fox(p123-172) Wickham(https://r4ds.had.co.nz/data-visualisation.html) Wickham(https://r4ds.had.co.nz/exploratory-data-analysis.html) <p>En ligne</p> <ul style="list-style-type: none"> https://juba.github.io/tidyverse/08-ggplot2.html 	
4	5 oct	Séance 4 : Causalité <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 	<ul style="list-style-type: none"> Cours Salganik(https://www.bitbybitbook.com/en/1sted/running-experiments/) Gerber, Alan S. and Donald P. Green, Field Experiments: Design, Analysis, and Interpretation. Chapters 1 and 2. Stephen, Morgan & Chistopher Winship. 2018. Counterfactuals and Causal Inference: methods and principles for Social research, 2nd edition. Chapter 1. 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 discuter en classe) Shelley, Clark et al. 2019. "The Impact of Childcare on Poor Urban Women's Economic Empowerment in Africa". Demography, 56, 4. (Optionnel) 	
5	12 oct	Action de Grâce	<ul style="list-style-type: none"> Séance de soutien (facultatif) 	
DEUXIÈME PARTIE: Méthodes usuelles d'analyse des données				
6	19 oct	Séance 6 : Régression linéaire	Cours <ul style="list-style-type: none"> Fox(p173-190, 197-207, 235-241, 243-246, 254-255) Malhotra, N., Margalit, Y. and Mo, C.H., 2013. "Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact." American Journal of Political Science , Vol. 38, No. 3, pp. 393-433. <p>En ligne</p> <ul style="list-style-type: none"> https://r4ds.had.co.nz/functions.html 	Devoir 2
7	26 oct	Semaine de relâche		
8	2 nov	Séance 8 : Régression linéaire (Fin)	Fox (p190-197, 207-224, 232-235, 246-254, 258-270) Fox (p385-417, 429-436) <p>En ligne</p> <ul style="list-style-type: none"> https://purrr.tidyverse.org/ 	Devoir 3

9	9 nov	Séance 9 : Régression logistique	Fox (p271-296)	
10	16 nov	Séance 10 : Régression logistique	Fox (p309-314 Fox (p417-425)	Devoir 4
TROISIÈME PARTIE: Données digitales et analyses textuelles				
11	23 nov	Séance 11 : Données digitales Labo11.1 : Collecte de données digitales Labo 11.2 : données textuelles	Cours <ul style="list-style-type: none"> • Fox (107-117) • Wickham (https://r4ds.had.co.nz/strings.html) • David Lazer and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data". Annual Review of Sociology 43:19–39. • Christopher Bail. 2014. "The cultural environment: measuring culture with big data". Theory and Society 43 (3-4): 465–482. • https://cran.r-project.org/web/packages/stringr/vignettes/stringr.html 	
12	30 nov	Séance 12 : Analyse textuelle S12.1: Analyse de sentiment (Sentiment analysis)	Cours <ul style="list-style-type: none"> • Silge (https://www.tidytextmining.com/sentiment.html) • Evans, James, and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." Annual Review of Sociology 42(1): 21-50. • 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. • Wallach, Hanna. 2018. "Computational Social Science 6 = Computer Science + Social Data." Communications of the ACM 61(3): 42-44. • DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." Big Data & Society 2(2): 1-5. 	Devoir 5
13	7 déc	Séance 13 : Analyse textuelle S13.1: Analyse des terms (topic and structural topic modelling)	Cours <ul style="list-style-type: none"> • Slige (https://www.tidytextmining.com/topicmodeling.html) • Blei, David M. 2012. "Probabilistic Topic Models." Communications of the ACM 55(4): 77-84. • Nelson, Laura K. 2017. "Computational Grounded Theory: A Methodological Framework." Sociological Methods & Research . • 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. 	

14	14 déc	Séance 14 : Analyse textuelle S13.2 : Extension	Cours <ul style="list-style-type: none"> • Silge (https://www.tidytextmining.com/nasa.html) • Silge (https://www.tidytextmining.com/usenet.html) 	Devoir 6
15	21 déc	Séance 15 : Conclusion	Examen final	

8. Inspiration

Ce cours s'est appuyé en grande partie sur le cours de méthodes que j'ai enseigné à la University of Cape Town au cours des 5 dernières années. Il a été révisé plusieurs fois et est toujours dans son développement. Les changements de ce cours et l'utilisation de R ont été influencés par un certain nombre de livres que j'ai lu ou par des discussions que j'ai eues avec des enseignants passionnés: Matthew Salganik, Kosuke Imai, Brandon Steward, Chris Bail, et Thomas LeGrand.

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 liées 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