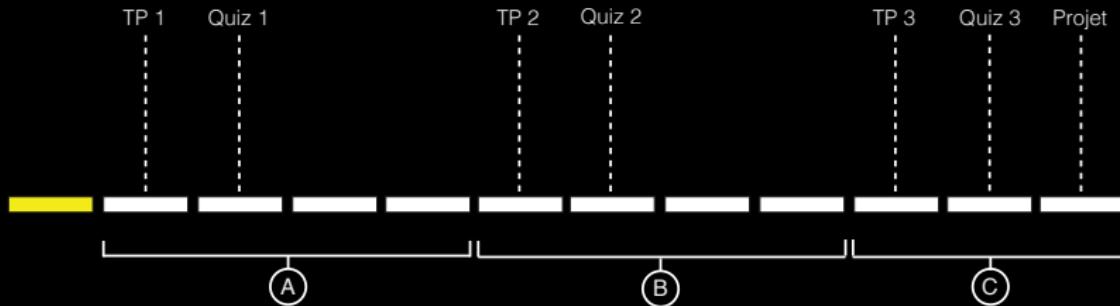


Prélude

POL7004 - Analyse quantitative
Université Laval



Instructeurs

- Enseignant: Justin Savoie

Instructeurs

- ▶ Enseignant: Justin Savoie
- ▶ Assistants: Laurence-Olivier M. Foisy, Hubert Cadieux

Slack



Slack: Popularité fulgurante



- ▶ Lancement en 2013: 8000 utilisateurs
- ▶ Avril 2015: 750 000 utilisateurs
- ▶ 2017: > 5 000 000 utilisateurs
- ▶ 2019: > 10 500 000 utilisateurs
- ▶ 2021: > 19 200 000 utilisateurs
- ▶ 2023: > 32 300 000 utilisateurs
- ▶ et ça continue...

Slack: Outil académique



airbnb

TIME

Pinterest

BuzzFeed



PANDORA



HARVARD
UNIVERSITY

LinkedIn

SAMSUNG

Ogilvy

ebay

AUTODESK

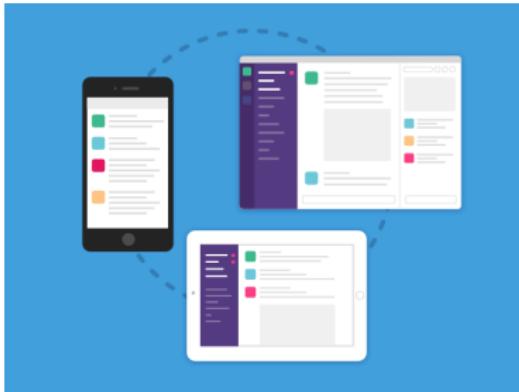
ticketmaster®

CONDÉ NAST

URBAN OUTFITTERS

- ▶ Lancement en 2013: 8000 utilisateurs
- ▶ Avril 2015: 750 000 utilisateurs
- ▶ 2017: > 5 000 000 utilisateurs
- ▶ 2019: > 10 500 000 utilisateurs
- ▶ 2021: > 19 200 000 utilisateurs
- ▶ 2023: > 32 300 000 utilisateurs
- ▶ et ça continue...

Slack: Efficacité



- ▶ Réponse plus rapide
- ▶ Permet la vidéo
- ▶ Création de sous-groupes privés
- ▶ ...et beaucoup plus

Slack

POL7004 - Anal...

Jump to...

All Unreads

All Threads

Channels

_général

- # _random
- # datacamp
- # diapositives
- # lectures
- # projet-final
- # quiz
- # tp

Direct Messages

- slackbot
- yannick.dufresne (you)
- Alexis Bibeau
- Alexis Bibeau, Nadjim F...
- Camille Tremblay
- Jean-Charles Del Duchi...
- Laura Pelletier-Bonnier
- Marielle Papin-Manjarr...
- Nadine Boucher

#_général

☆ | 57 | 0 | Company-wide announcements and work-related discussions

Today

joined #_général along with Lucas Moreau.

Bonne question!

DKN-5418 (https://www.fss.ulaval.ca/sites/fss.ulaval.ca/files/fss/science-politique/Horaires/Hor_23_A18_20180827.pdf)

Adrien Cloutier 10:38
joined #_général.

Nadjim Fréchet 11:38
Non, il n'y a pas de portail de cours pour POL-7004, nos interactions auront lieu sur slack 😊 !

1

joined #_général. 11:42

joined #_général. 11:45

Ou est la salle de cours? 12:03

joined #_général.

yannick.dufresne 13:18
: J'ai envoyé un courriel à tous. L'as-tu reçu?

14:25

Oui, desoler, jai envoyé la question avant de regarder mes courriels 17:06

joined #_général.

+ Message #_général

Au menu cette semaine

Au menu cette semaine



Survol du cours

Au menu cette semaine



Survol du cours



Choix des lectures

Au menu cette semaine



Survol du cours



Choix des lectures



RStudio et R ... L^AT_EX et RMarkdown

Au menu cette semaine



Survol du cours



Choix des lectures



RStudio et R ... L^AT_EX et RMarkdown



DataCamp

Au menu cette semaine



Survol du cours



Choix des lectures



RStudio et R ... L^AT_EX et RMarkdown



DataCamp



Trucs de la semaine

Au menu cette semaine



Survol du cours



Choix des lectures



RStudio et R ... L^AT_EX et RMarkdown



DataCamp



Trucs de la semaine

?

Objectifs du cours



Objectifs du cours

1. Technique



Objectifs du cours

1. Technique
2. Analytique



Objectifs du cours

1. Technique
2. Analytique
3. Application



A - Technique

1. Programmation R
2. Intro au monde *Open source* scientifique
3. Gérer des données réalistes



B - Analytique

1. Connaissances des méthodes statistiques
2. Interprétation des résultats



C - Application

1. Organiser un travail en analyse quantitative avec R Markdown... et L^AT_EX
2. Gérer les problèmes typiques liés à l'analyse de données
3. Exposition aux nouvelles méthodes quantitatives de recherche



Trop?



Proposition

1. Moins de maths
2. Plus de technique



Fonction mathématique

\sqrt{x}

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$



Fonction en \mathbb{R}



```
1 var <- function (x, y = NULL, na.rm = FALSE, use)
2 {
3     if (missing(use))
4         use <- if (na.rm)
5             "na.or.complete"
6         else "everything"
7     na.method <- pmatch(use, c("all.obs", "complete.obs", "pairwise.complete.obs",
8         "everything", "na.or.complete"))
9     if (is.na(na.method))
10        stop("invalid 'use' argument")
11     if (is.data.frame(x))
12         x <- as.matrix(x)
13     else stopifnot(is.atomic(x))
14     if (is.data.frame(y))
15         y <- as.matrix(y)
16     else stopifnot(is.atomic(y))
17     .Call(C_cov, x, y, na.method, FALSE)
18 }
```



Fonction en R : Utilisation

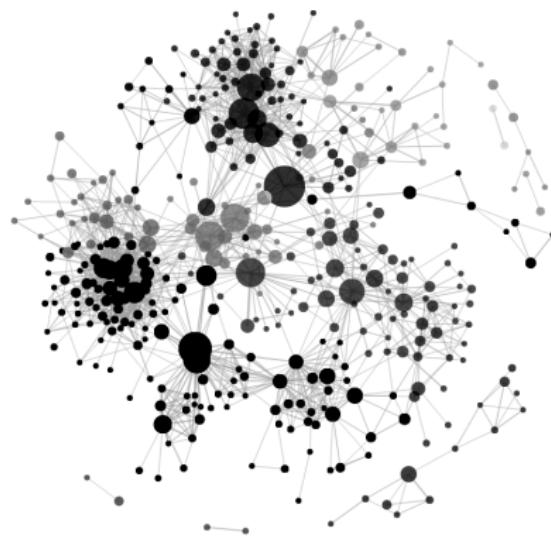
<...>

```
1 var(votreVariable)
```





Royaume vaste et confus



Mon rôle

- ▶ Présenter de manière organisée les ★ de la littérature

Mon rôle

- ▶ Présenter de manière organisée les ★ de la littérature
- ▶ Partager des trucs pratiques du monde académique

Mon rôle

- ▶ Présenter de manière organisée les ★ de la littérature
- ▶ Partager des trucs pratiques du monde académique
- ▶ Inciter à faire les lectures et pratiquer
 - ▶ Choix des lectures

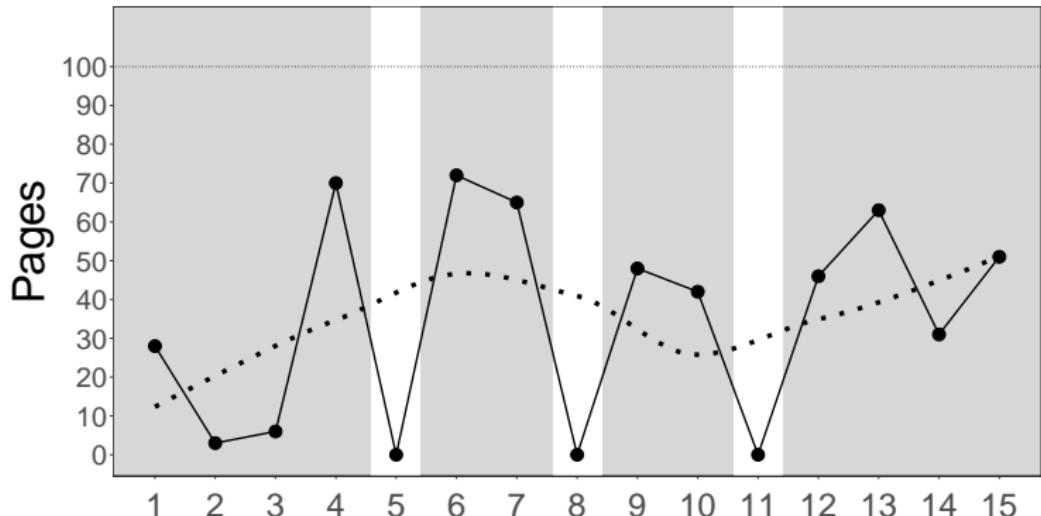
Incitatif: 100 pages max.



Ma philosophie: Lire moins, lire mieux.
Lisez lentement. Chaque mot.

Apprendre le “jargon” de la recherche scientifique est la clé.
C'est d'autant difficile que ce jargon est pour la plupart du temps
développé en anglais.

Incitatif: 100 pages max.



Incitatif: Pertinence et diversité

Méthode de sélection des textes

Incitatif: Pertinence et diversité

Méthode de sélection des textes

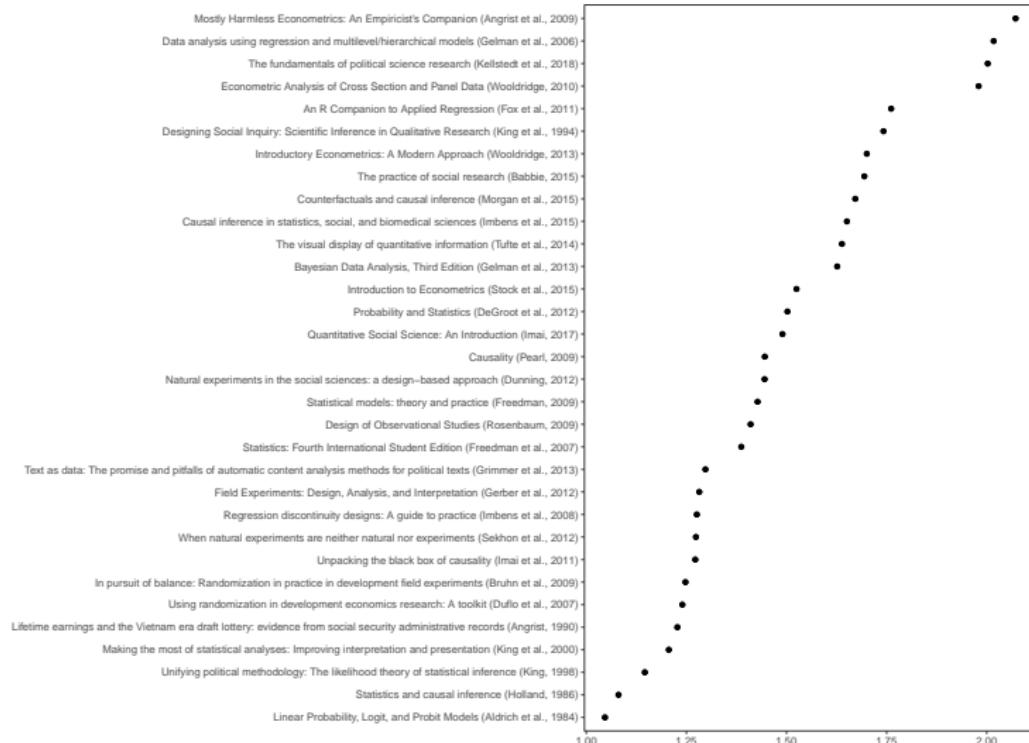
- Base d'ouvrages et d'articles issus de >30 syllabus d'Universités de prestige ayant des objectifs similaires au cours *Analyse quantitative*

Incitatif: Pertinence et diversité

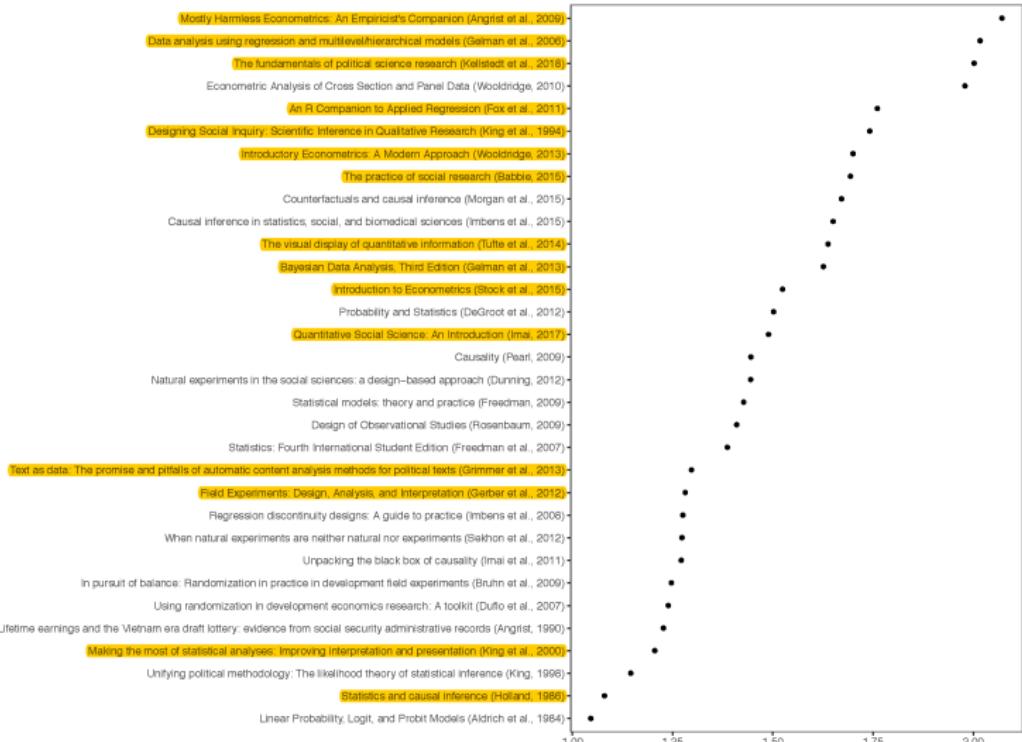
Méthode de sélection des textes

- ▶ Base d'ouvrages et d'articles issus de >30 syllabus d'Universités de prestige ayant des objectifs similaires au cours *Analyse quantitative*
- ▶ Pondération
 - ▶ Nombre de citations sur Google Scholar
→ Importance en science
 - ▶ Nombre d'utilisations dans les syllabus sélectionnés
→ Importance dans le champ
 - ▶ Années
→ Actualité

Incitatif: Pertinence et diversité



Incitatif: Pertinence et diversité



Sur la pertinence



Les lectures ★ sont difficiles. Mais vous ne perdez pas votre temps.

Ces ouvrages sont situés au cœur de la discipline.

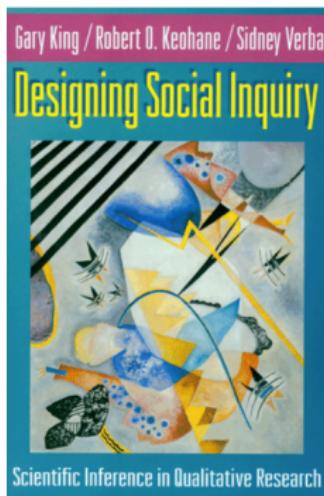
On pourrait presque construire un cours complet sur une seule de ces lectures.

Incitatif: Recueil agréable



Dufresne, Yannick. *Sélection de textes: POL7004.* (2020).

Première lecture: KKV

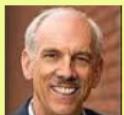


King, Keohane et Verba. *The Science in Social Science* (1994).

Première lecture: KKV



Gary King (Harvard University) — 51398 citations



Robert Keohane (Princeton University) — 35000
citations



Sydney Verba (Harvard University) — 50000 citations

Lectures et pratique



Essentielles.

L'analyse quantitative doit être vue comme l'apprentissage d'une nouvelle langue.

Lectures et pratique



Mieux de pratiquer un peu chaque jour que beaucoup un seul jour.

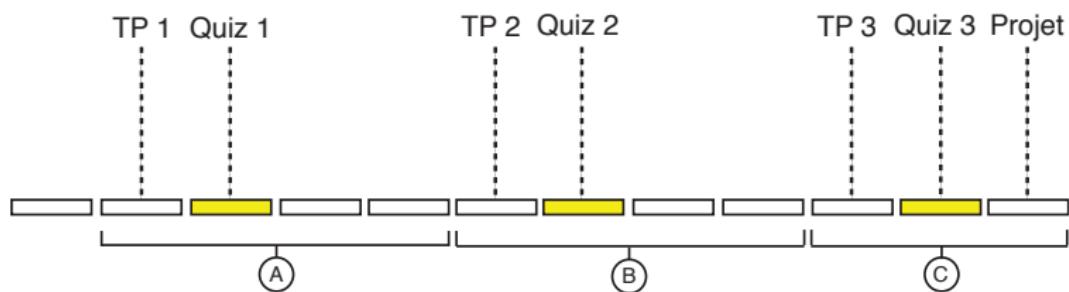
Évaluations

Évaluations

- Travaux pratiques (3): 30%
- Quiz (3): 30%
- Projet final: 30%
- Participation: 10%

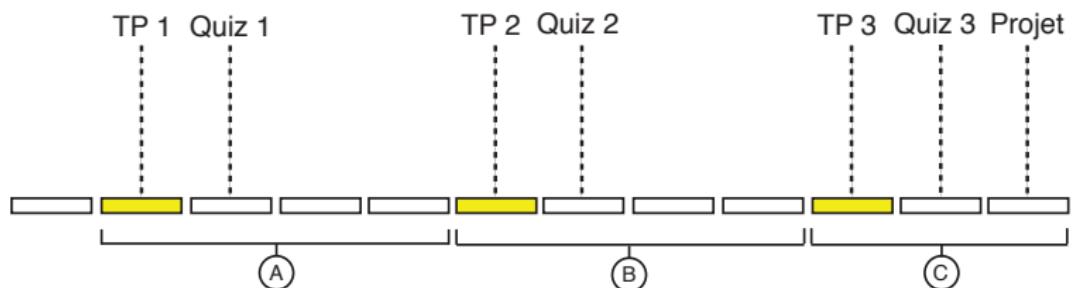
Quiz

► 3 Quiz



TP

► 3 TP



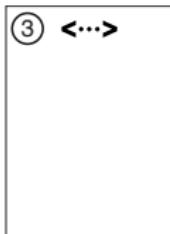
TP: Démarche évolutive

TP1

CodeR-TP1.R



CodeLaTeX-TP1.tex



TP: Démarche évolutive

TP1

CodeR-TP1.R

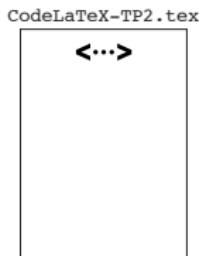
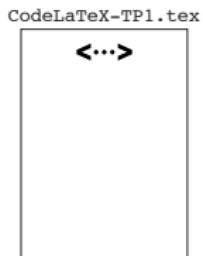
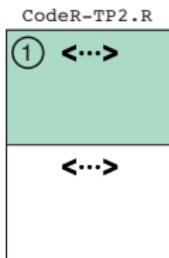


CodeLaTeX-TP1.tex



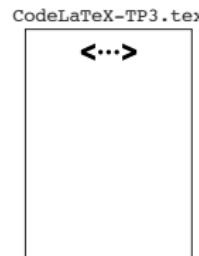
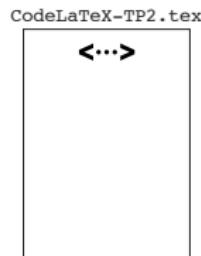
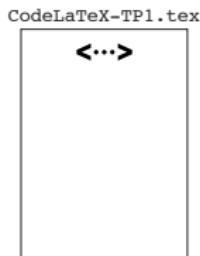
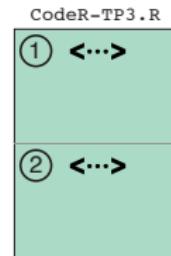
TP: Démarche évolutive

TP1 → TP2



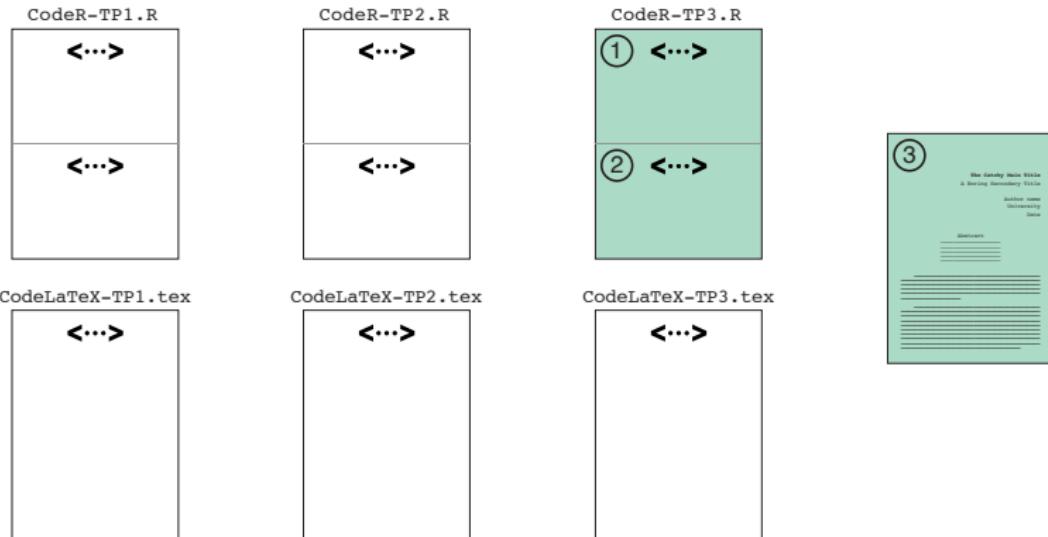
TP: Démarche évolutive

TP1 → TP2 → TP3

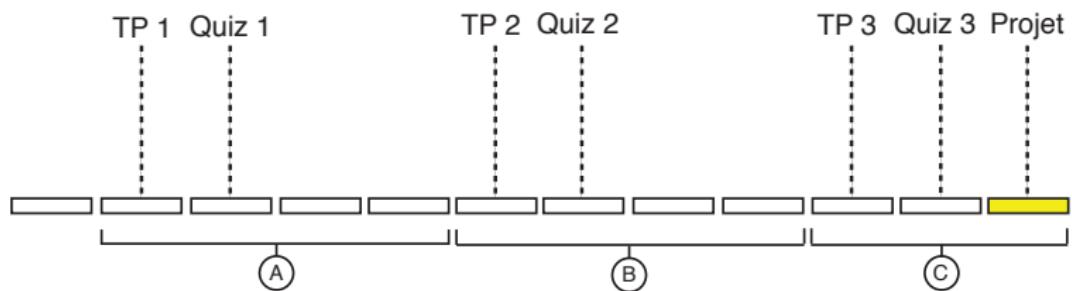


TP: Démarche évolutive

TP1 → TP2 → TP3 → Projet

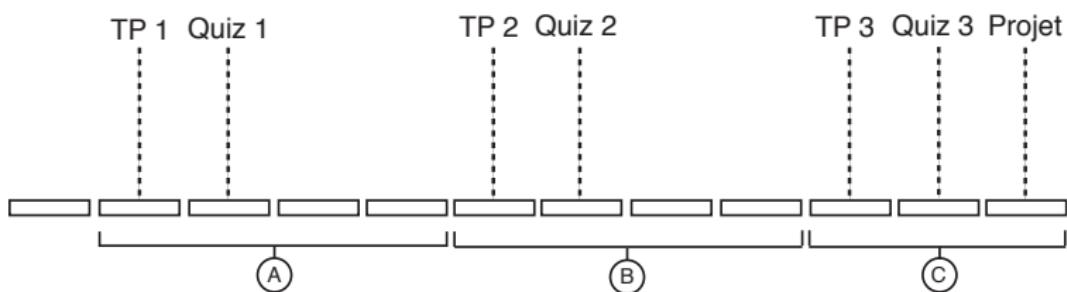


Projet



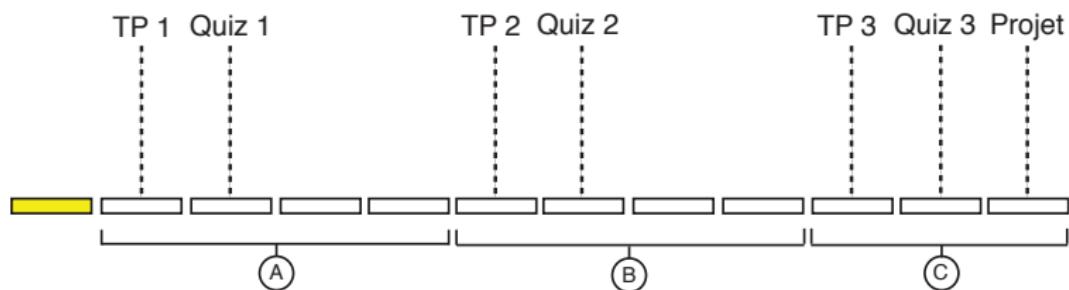
Structure du cours

Cours



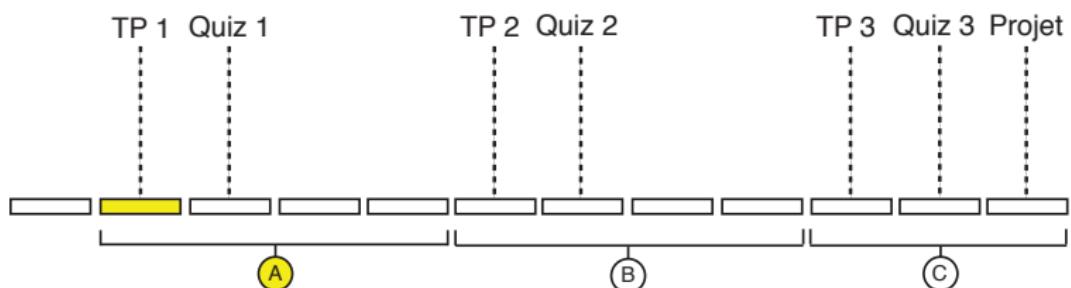
Cours

Cours 1: Prélude



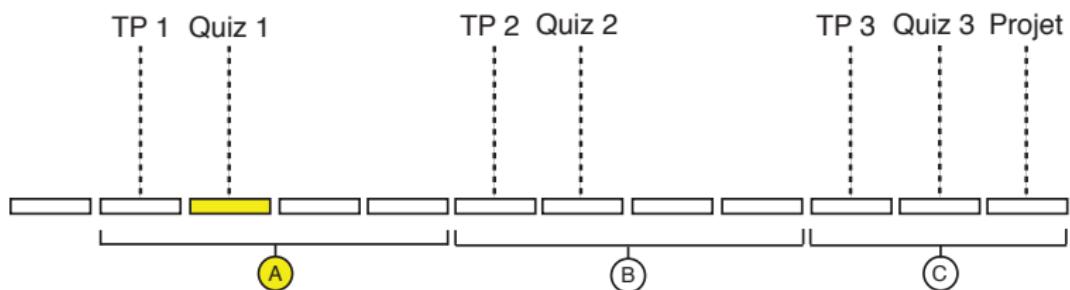
Objectif A - Technique

Cours 2: Sciences sociales informatiques



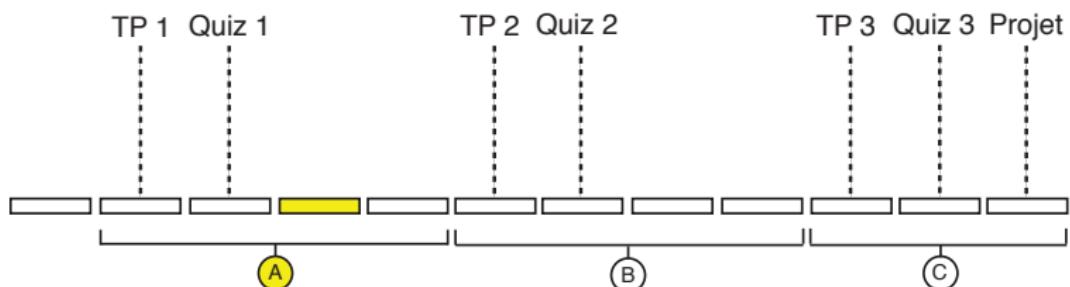
Objectif A - Technique

Cours 3: Domestication des données



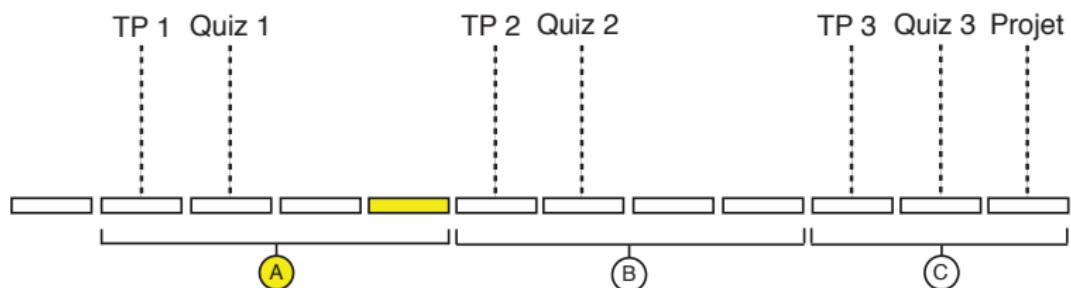
Objectif A - Technique

Cours 4: Échelles de mesure



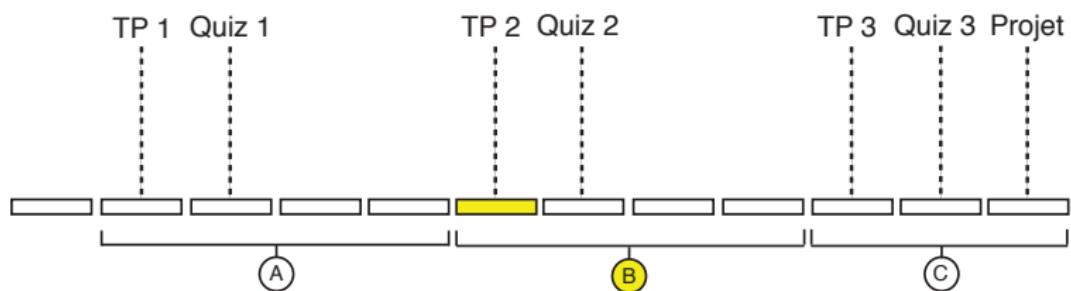
Objectif A - Technique

Cours 5: Description univariée



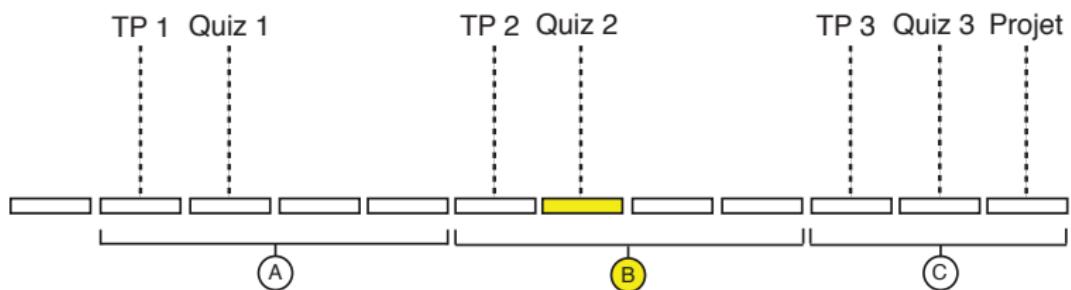
Objectif B - Analyse

Cours 6: Analyse bivariée et inférence statistique



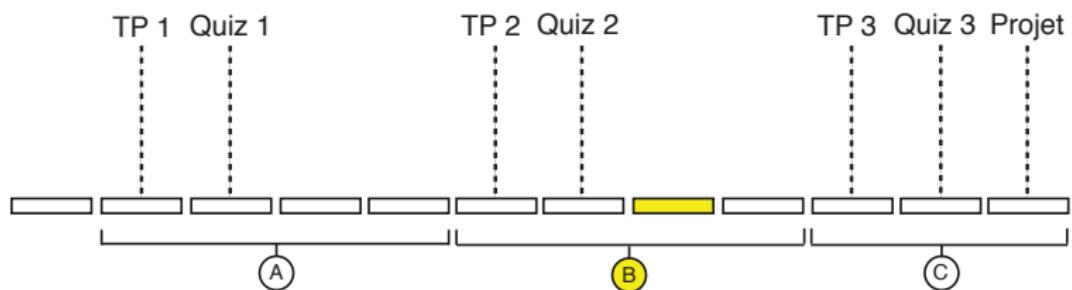
Objectif B - Analyse

Cours 7: Régression linéaire simple



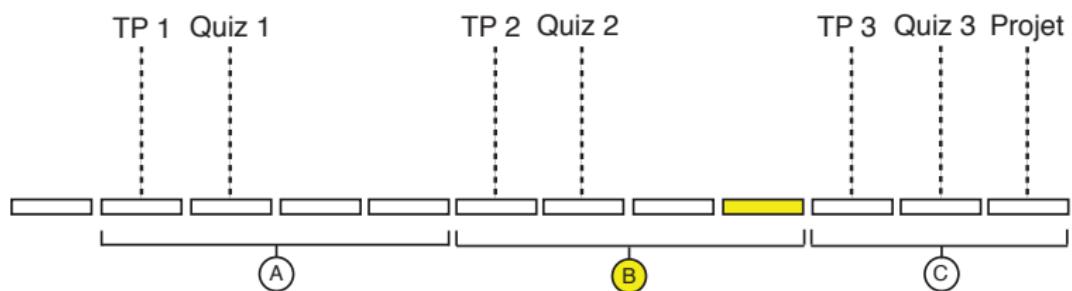
Objectif B - Analyse

Cours 8: Régression linéaire multiple



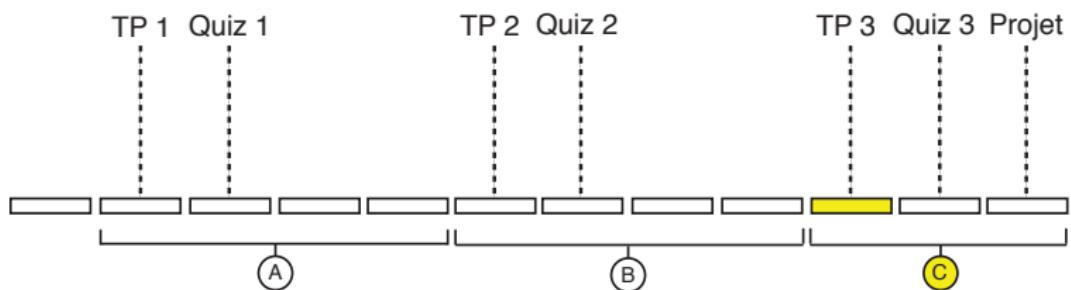
Objectif B - Analyse

Cours 9: Régression logistique



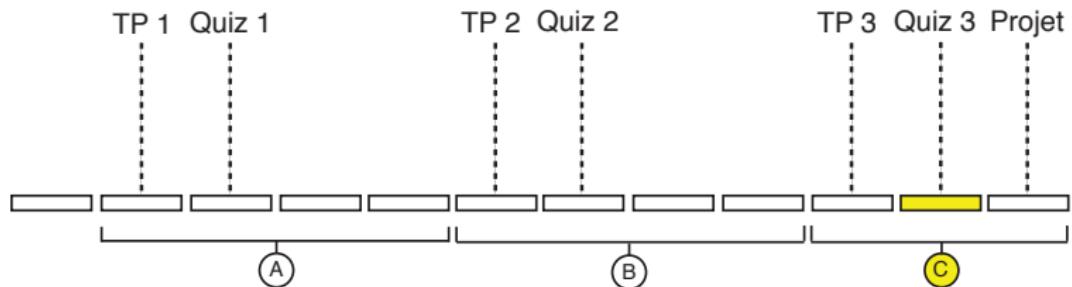
Objectif C - Application

Cours 10: Visualisation graphique avancée



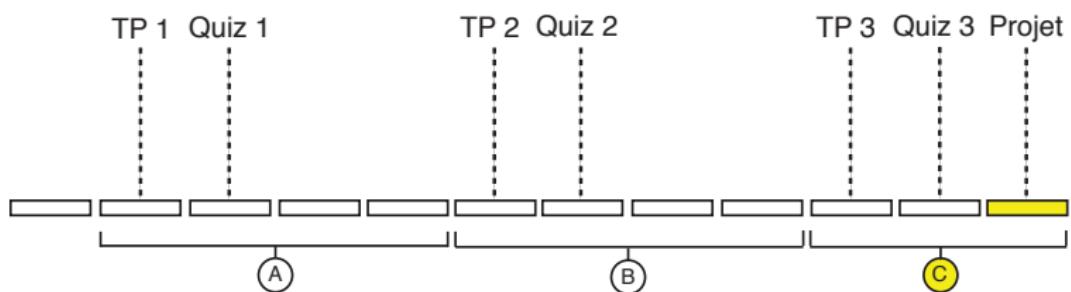
Objectif C - Application

Cours 11: Données textuelles et intelligence artificielle



Objectif C - Application

Cours 12: Méthode expérimentale et inférence





Langages et logiciels

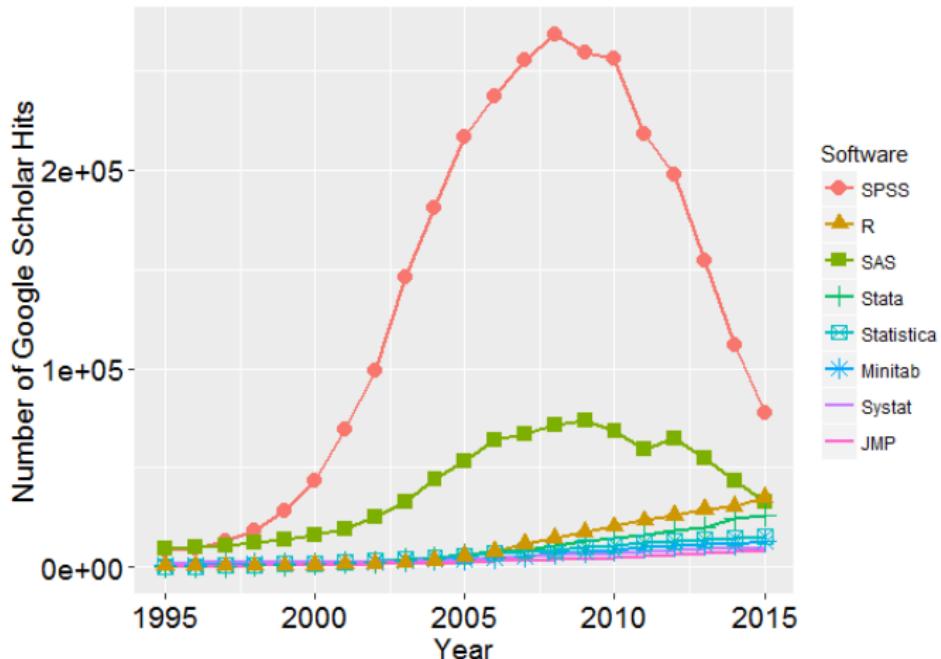
1. $\mathbb{R} \rightarrow \text{RStudio}$

Langages et logiciels

1. R → RStudio
2. R Markdown... L^AT_EX → RStudio

Pourquoi R ?

Pourquoi R ?



Pourquoi R ? Raisons d'adorer

1. Gratuit
2. Disponible pour toutes les plateformes
3. Visualisation graphique + R Markdown et L^AT_EX
4. Diffusion + packages
5. *Open source*: développement par des scientifiques

Pourquoi R ? Raisons de détester

1. Programmation: Côte d'apprentissage abrupte
2. Développement éclectique, voire anarchique

Pourquoi L^AT_EX?

Pourquoi L^AT_EX? Raisons d'aimer

- ▶ Bibliographies: BIBT_EX
- ▶ Table des matières, tableaux, etc.
- ▶ Renomme automatiquement les tableaux, figures, etc.
- ▶ Modèles (“Templates”)
- ▶ Code + *Open source* = Communauté d'experts sur le web

Pourquoi L^AT_EX? Raisons de détester

- ▶ Difficile à apprendre: mais la base est facile
- ▶ Incompatibilité avec Word
- ▶ ~~Pas d'auto-correcteur~~
- ▶ Pas de “Suivre les corrections” et autres trucs du genre
- ▶ On ne voit pas le résultat en temps réel
- ▶ Certaines revues n'acceptent pas... mais d'autres l'imposent

LAT_EX: Un beau tableau

Tableau 1. Grandeur des bananes et des pommes

Quantile	Bananes	Pommes
0%	59	44
50%	69	64
100%	77	71

LATEX: Le code du beau tableau

```
\begin{table}
  \centering
  \caption{Grandeur des bananes et des pommes}
  \begin{tabular}{lrr}
    Quantile & Bananes & Pommes\\ \hline
    0\% & 59 & 44 \\
    50\% & 69&64 \\
    100\% & 77 & 71 \\
  \end{tabular}
  \label{tab:bananespommes}
\end{table}
```

LATEX: S't'une joke?!



LATEX: Non



LATEX

Tableau 1: Tests des hypothèses

	Vote pour le NPD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Évaluation du chef NPD					3.87*** (0.22)	3.81*** (0.24)	3.17*** (0.52)
Droite idéologique			-2.86*** (0.46)	-3.24*** (0.53)			-2.66*** (0.57)
Québec	0.69*** (0.09)	0.61*** (0.16)		0.92** (0.34)		0.56** (0.17)	0.93** (0.35)
Femme		0.05 (0.09)		-0.08 (0.19)		-0.03 (0.10)	-0.08 (0.20)
Francophone		-0.02 (0.17)		-0.37 (0.35)		-0.29 (0.18)	-0.63 (0.37)
allophone		-0.17 (0.15)		-0.38 (0.34)		-0.18 (0.17)	-0.22 (0.36)
Moins de 34 ans		-0.03 (0.15)		-0.17 (0.34)		-0.13 (0.16)	-0.26 (0.36)
Plus de 55 ans		-0.23* (0.10)		-0.33 (0.21)		-0.24* (0.11)	-0.23 (0.22)
Haut revenu		-0.33** (0.12)		-0.36 (0.24)		-0.30* (0.13)	-0.32 (0.25)
Faible revenu		0.30* (0.15)		0.33 (0.31)		0.40* (0.17)	0.49 (0.33)
Pas de diplôme secondaire		-0.23 (0.15)		0.04 (0.36)		-0.12 (0.17)	0.03 (0.38)
Diplôme universitaire		0.13 (0.10)		-0.61** (0.21)		-0.12 (0.11)	-0.79*** (0.22)
-constante	-1.05*** (0.05)	-0.86*** (0.11)	0.34 (0.20)	0.96** (0.35)	-3.17*** (0.15)	-2.95*** (0.19)	-1.21* (0.51)
N	2,745	2,464	655	610	2,636	2,381	602
Log Likelihood	-1,650.11	-1,487.30	-383.02	-346.16	-1,412.88	-1,276.31	-317.77
AIC	3,304.22	2,996.60	770.04	716.31	2,829.77	2,576.62	661.54

Source : Étude électorale canadienne, 2011.

Note : Régression logistique binomiale.

*p<0.05 ; **p<0.01 ; ***p<0.001

LATEX: Partie 1 du code

<...>

```
1 % Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac
2 % Date and time: Wed, Jan 07, 2015 - 22:20:00
3 \begin{table}[-1.8ex]{!htbp} \centering
4   \caption{Tests des hypoth\`eses}
5   \label{}
6 \scriptsize
7 \begin{tabular}{@{\extracolsep{5pt}}lcccccc}
8 \hline
9 & \multicolumn{7}{c}{Vote pour le NPD} \\
10 & (1) & (2) & (3) & (4) & (5) & (6) & (7) \\
11 \hline
12 \\'{E}valuation du chef NPD & & & & 3.87$^{***}$ & 3.81$^{***}$ & 3.17$^{***}$ \\
13 & & & & (0.22) & (0.24) & (0.52) \\
14 Droite id\'eologique & & & & -$2.86$^{***}$ & -$3.24$^{***}$ & -$2.66$^{***}$ \\
15 & & & & (0.46) & (0.53) & (0.57) \\
16 Qu\'ebec & 0.69$^{***}$ & 0.61$^{***}$ & 0.92$^{**}$ & 0.56$^{**}$ & 0.93$^{**}$ \\
17 & (0.09) & (0.16) & (0.34) & (0.17) & (0.35) \\
18 Femme & 0.05 & -$0.08 & -$0.03 & -$0.08 \\
19 & (0.09) & (0.19) & (0.10) & (0.20) \\
20 Francophone & -$0.02 & -$0.37 & -$0.29 & -$0.63 \\
21 & (0.17) & (0.35) & (0.18) & (0.37) \\
22 Allophone & -$0.17 & -$0.38 & -$0.18 & -$0.22 \\
23 & (0.15) & (0.34) & (0.17) & (0.36)
```

LATEX: Partie 2 du code

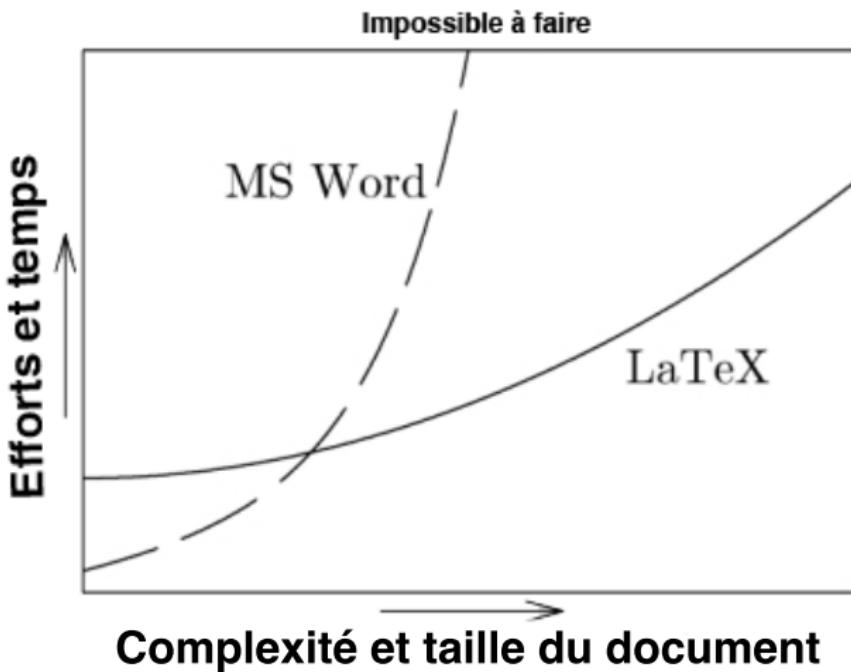
<...>

```
1 Moins de 34 ans & & $-$0.03 & & $-$0.17 & & $-$0.13 & $-$0.26 \\
2 & & (0.15) & & (0.34) & & (0.16) & (0.36) \\
3 Plus de 55 ans & & $-$0.23$^{*} \$ & & $-$0.33 & & $-$0.24$^{*} \$ & $-$0.23 \\
4 & & (0.10) & & (0.21) & & (0.11) & (0.22) \\
5 Haut revenu & & $-$0.33$^{**} \$ & & $-$0.36 & & $-$0.30$^{*} \$ & $-$0.32 \\
6 & & (0.12) & & (0.24) & & (0.13) & (0.25) \\
7 Faible revenu & & 0.30$^{*} \$ & & 0.33 & & 0.40$^{*} \$ & 0.49 \\
8 & & (0.15) & & (0.31) & & (0.17) & (0.33) \\
9 Pas de diplôme secondaire & & $-$0.23 & & 0.04 & & $-$0.12 & 0.03 \\
10 & & (0.15) & & (0.36) & & (0.17) & (0.38) \\
11 Diplôme universitaire & & 0.13 & & $-$0.61$^{**} \$ & & $-$0.12 & $-$0.79$^{***} \$ \\
12 & & (0.10) & & (0.21) & & (0.11) & (0.22) \\
13 \_constante & $-$1.05$^{***} \$ & $-$0.86$^{***} \$ & 0.34 & 0.96$^{**} \$ & $-$3.17$^{***} \$ & $-$1.276 3 \\
14 & (0.05) & (0.11) & (0.20) & (0.35) & (0.15) & (0.19) & (0.51) \\
15 N & 2,745 & 2,464 & 655 & 610 & 2,636 & 2,381 & 602 \\
16 Log Likelihood & $-$1,650.11 & $-$1,487.30 & $-$383.02 & $-$346.16 & $-$1,412.88 & $-$1.276 3 \\
17 AIC & 3,304.22 & 2,996.60 & 770.04 & 716.31 & 2,829.77 & 2,576.62 & 661.54 \\
18 \hline \\[-1.8ex] \\
19 \multicolumn{8}{l}{\emph{Source}: Étude électorale canadienne, 2011.} \\
20 \multicolumn{8}{l}{\emph{Note}: Régression logistique binomiale.} \\
21 \multicolumn{8}{l}{\$^{*} \$p\$ < \$0.05; \$^{**} \$p\$ < \$0.01; \$^{***} \$p\$ < \$0.001} \\
22 \end{tabular} \\
23 \end{table}
```

LATEX



LATEX



R et LATEX

<...>

```
1 stargazer(modele1, modele2, modele3, modele4, modele5, modele6, modele7)
```

R et LATEX

Tableau 1: Tests des hypothèses

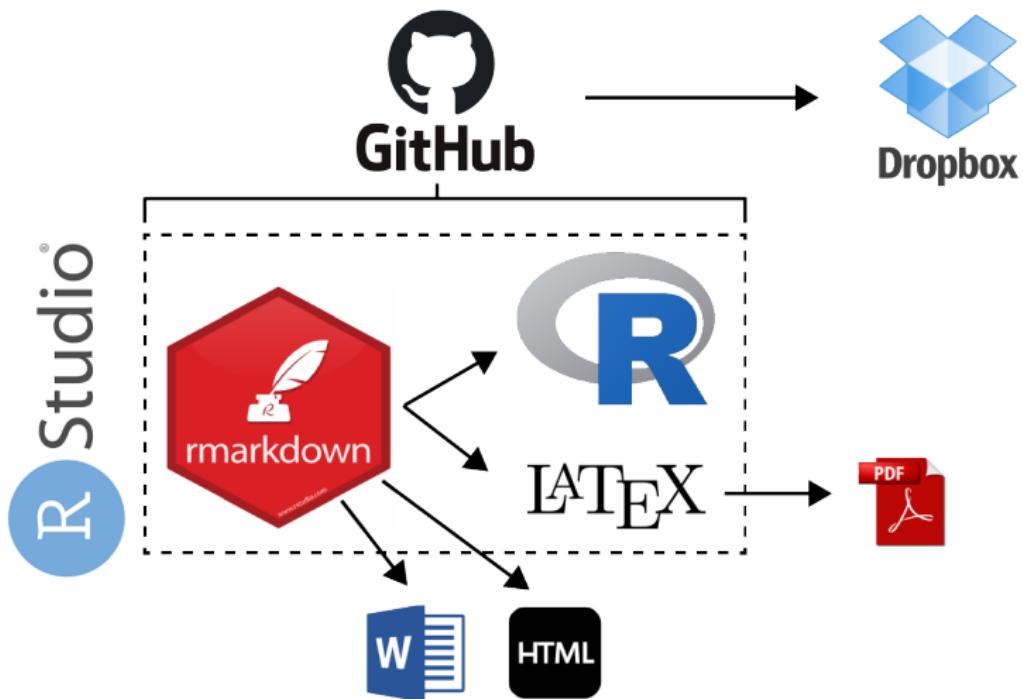
	Vote pour le NPD						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Évaluation du chef NPD					3.87*** (0.22)	3.81*** (0.24)	3.17*** (0.52)
Droite idéologique			-2.86*** (0.46)	-3.24*** (0.53)			-2.66*** (0.57)
Québec	0.69*** (0.09)	0.61*** (0.16)		0.92** (0.34)		0.56** (0.17)	0.93** (0.35)
Femme		0.05 (0.09)		-0.08 (0.19)		-0.03 (0.10)	-0.08 (0.20)
Francophone		-0.02 (0.17)		-0.37 (0.35)		-0.29 (0.18)	-0.63 (0.37)
allophone		-0.17 (0.15)		-0.38 (0.34)		-0.18 (0.17)	-0.22 (0.36)
Moins de 34 ans		-0.03 (0.15)		-0.17 (0.34)		-0.13 (0.16)	-0.26 (0.36)
Plus de 55 ans		-0.23* (0.10)		-0.33 (0.21)		-0.24* (0.11)	-0.23 (0.22)
Haut revenu		-0.33** (0.12)		-0.36 (0.24)		-0.30* (0.13)	-0.32 (0.25)
Faible revenu		0.30* (0.15)		0.33 (0.31)		0.40* (0.17)	0.49 (0.33)
Pas de diplôme secondaire		-0.23 (0.15)		0.04 (0.36)		-0.12 (0.17)	0.03 (0.38)
Diplôme universitaire		0.13 (0.10)		-0.61** (0.21)		-0.12 (0.11)	-0.79*** (0.22)
-constante	-1.05*** (0.05)	-0.86*** (0.11)	0.34 (0.20)	0.96** (0.35)	-3.17*** (0.15)	-2.95*** (0.19)	-1.21* (0.51)
N	2,745	2,464	655	610	2,636	2,381	602
Log Likelihood	-1,650.11	-1,487.30	-383.02	-346.16	-1,412.88	-1,276.31	-317.77
AIC	3,304.22	2,996.60	770.04	716.31	2,829.77	2,576.62	661.54

Source : Étude électorale canadienne, 2011.

Note : Régression logistique binomiale.

*p<0.05 ; **p<0.01 ; ***p<0.001

Pourquoi R Markdown?



TP1 • Démo



DataCamp

The screenshot shows the DataCamp website with a dark blue header. On the left is the DataCamp logo. To the right are navigation links: 'Learn ▾', 'For Business', and a user profile section showing '15,528 XP' and a small icon. Below the header, a large banner for a 'FREE COURSE' titled 'Introduction to R' is displayed. It features a yellow 'Continue Course' button. To the right of the title is a circular badge with a stylized 'R' and the text 'INTRODUCTION TO R'. At the bottom of the banner, course statistics are listed: '4 hours', '0 Videos', '62 Exercises', '930,327 Participants', '6,200 XP', and a link to 'Download the app' with icons for iOS and Android.

Course Description

In this introduction to R, you will master the basics of this beautiful open source language, including factors, lists and data frames. With the knowledge gained in this course, you will be ready to undertake your first very own data analysis. With over 2 million users worldwide R is rapidly becoming the leading programming language in statistics and data science. Every year, the number of R users grows by 40% and an increasing number of organizations are using it in their day-to-day activities. Leverage the power of R by completing this free R online course today!



1 Intro to basics

13%

In this chapter, you will take your first steps with R. You will learn how to use the console as a calculator and how to assign variables. You will also get to know the basic data types in R. Let's get started!

[VIEW CHAPTER DETAILS ▾](#)

[Continue Chapter](#)

This course is part of these tracks:

[Data Analyst with R](#)

[Data Scientist with R](#)

[R Programmer](#)

[R Programming](#)



Jonathan Cornelissen





La principale difficulté du TP1 sera de comprendre l'arborescence des dossiers (“Directory tree”)



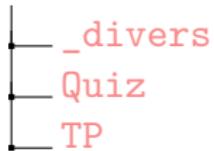
Mémorisez l'arborescence de votre dossier DropBox

Arborescence DropBox

POL7004-NomPrenom

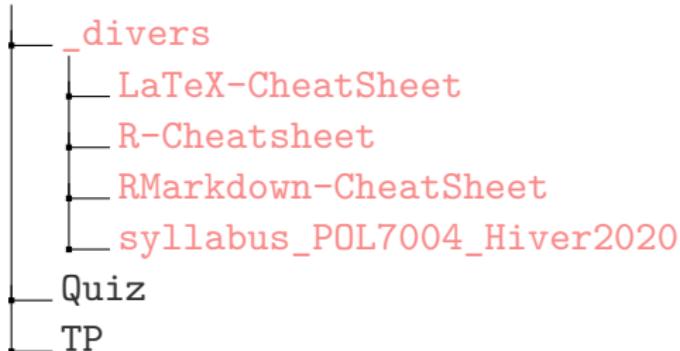
Arborescence DropBox

POL7004-NomPrenom



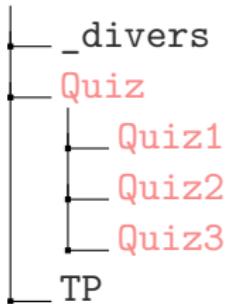
Arborescence DropBox

POL7004-NomPrenom



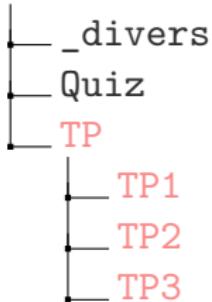
Arborescence DropBox

POL7004-NomPrenom



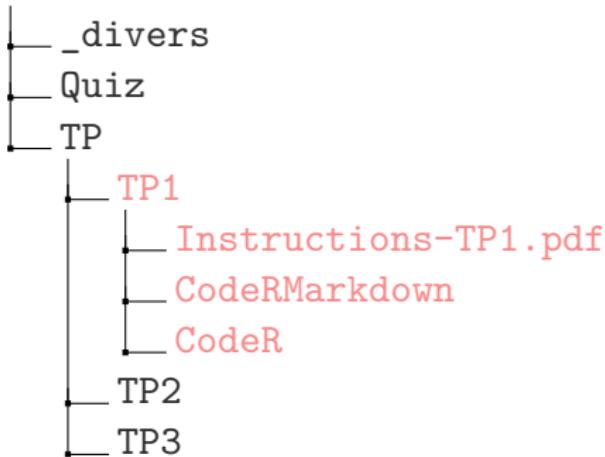
Arborescence DropBox

POL7004-NomPrenom



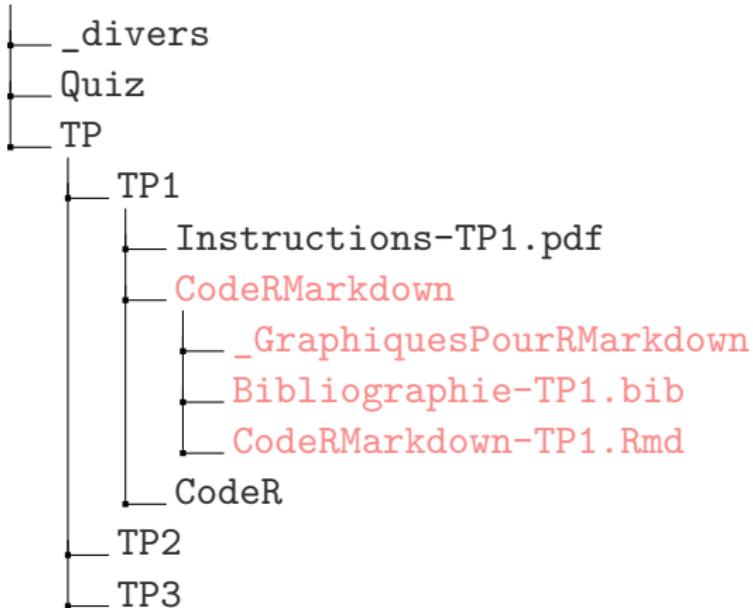
Arborescence DropBox

POL7004-NomPrenom



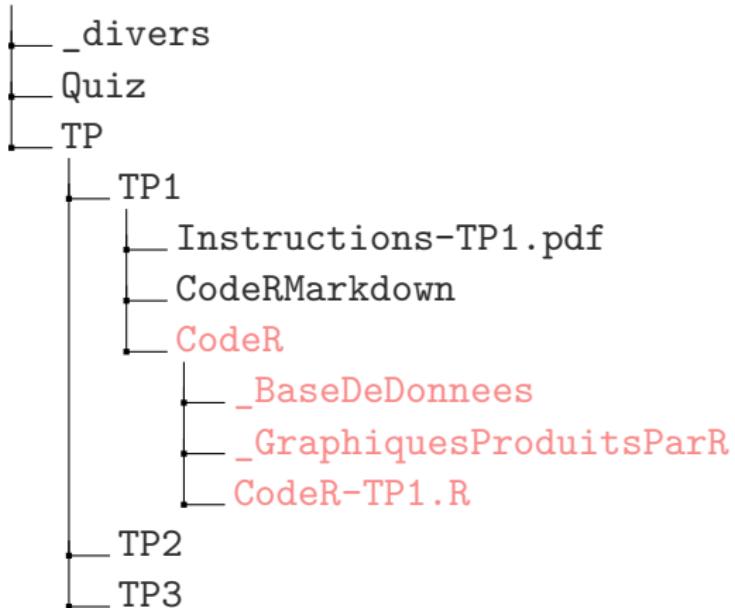
Arborescence DropBox

POL7004-NomPrenom



Arborescence DropBox

POL7004-NomPrenom



Quoi faire quand ça ne marche pas?



Demander à la grande communauté

Chercher des solutions sur le web fait partie de l'apprentissage de l'univers *Open source*.

Stackoverflow deviendra rapidement votre ami intime.

Quoi faire quand ça ne marche pas?



Demander à notre petite communauté POL-7004
Demander sur Slack

Pour finir...

Résumé

1. ★+ 
2. Accepter l'invitation DropBox
3. Suivre les étapes du TP1

Fin.