

Outils numériques de recherche

Adrien Cloutier

19 septembre 2024

Section 1

Les langages de programmation

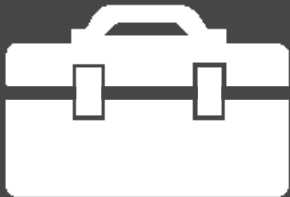
Les langages de programmation



Section 2

Retour sur le cours 2

Logiciel libre et code source ouvert



Le logiciel libre et le code source ouvert

- ❶ Quelle est la différence entre logiciel libre et le code source ouvert?
- ❷ Sont-ce des outils?
- ❸ Des exemples de logiciels libres?
- ❹ Méthodes vs outils?
- ❺ Qu'est-ce que la License publique général (GNU)?

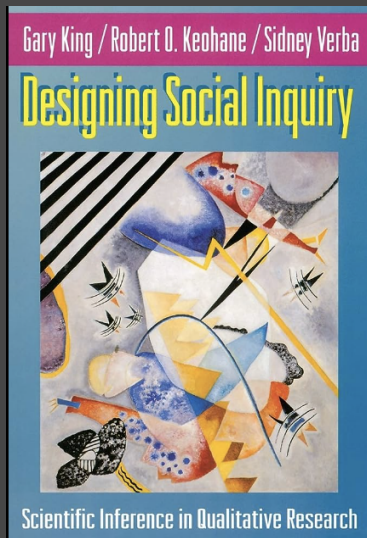
License publique général (GNU)

- ❶ Création en 1983 par Stallman.
- ❷ Permet l'arrivée de Firefox, OpenOffice, Linux. . .
- ❸ Valeur: entraide, collaboration, partage.
- ❹ Pour protéger les droits des utilisateurs, pas les droits des auteurs.

Citation inspirante

« Seulement de cette façon, l'utilisateur peut contrôler le logiciel, plutôt que de se faire contrôler par ce dernier (Stallman, 1986). »

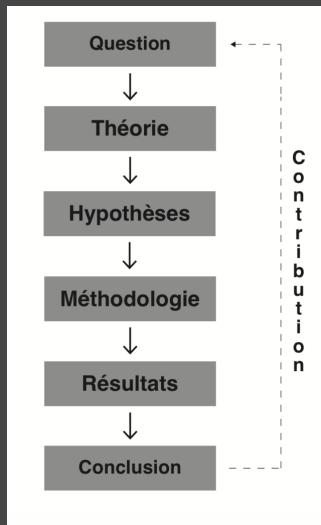
Designing Social Inquiry, King et al. (2021)



4 critères de la recherche scientifique selon KKV

- ❶ Le but est l'inférence;
- ❷ Les procédures sont publiques;
- ❸ Les conclusions sont incertaines;
- ❹ Le contenu est la méthode.

Cours basé sur le cycle de la recherche



Avant la recherche

- ❶ « Dépôt » GitHub pour rendre son code accessible;
- ❷ Enregistrement des hypothèses de recherche.

Pendant la recherche

- ❶ Cueillette de données (R)
- ❷ Analyse de données (R)
- ❸ Visualisation de données (R)
- ❹ Rédaction (LaTeX, Zotero)

- ❶ « Open science » (science ouverte)
 - Partage des données
 - Article accessible gratuitement à tous

Critères de sélection des outils du cours

- ➊ Accessibilité (Gratuit ou peu dispendieux)
- ➋ Existence d'une communauté d'utilisateurs
- ➌ Popularité dans le champ
- ➍ Compatibilité avec d'autres outils
- ➎ Transparence et répliquabilité
- ➏ Adaptabilité et flexibilité

Section 3

R comme outil de recherche

R comme outil de recherche

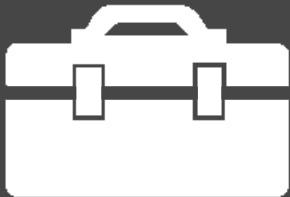
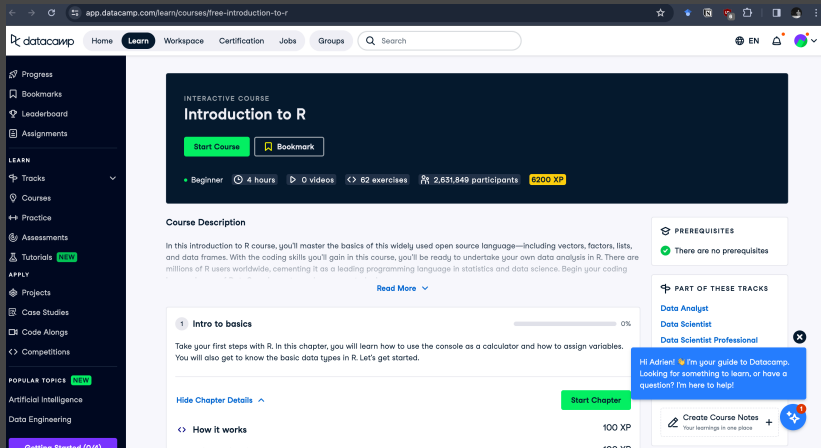


Table 2.1: Résumé des principaux outils de programmation pour l'analyse de données

Critères	Logiciels sous licence (SAS, SPSS, STATA)	Python	Julia	R
Accessibilité (Gratuit ou peu dispendieux)	Non	Oui	Oui	Oui
Existence d'une communauté d'utilisateurs	Modérée	Très élevée	En croissance	Très élevée
Popularité dans le champ	Élevée dans certains secteurs	Très populaire	Modérée	Très populaire en sciences sociales et statistiques
Compatibilité avec d'autres outils	Bonne	Excellente	Bonne	très bonne
Transparence et répliquabilité	Faible	Bonne	Bonne	Très élevée
Adaptabilité et flexibilité	Limitée	Très flexible	Très flexible	Très flexible

Datacamp (5%!)



The screenshot displays the DataCamp website interface. At the top, the navigation bar includes links for Home, Learn, Workspace, Certification, Jobs, and Groups, along with a search bar and user profile options. The left sidebar lists various features like Progress, Bookmarks, Leaderboard, and Assignments, as well as a 'LEARN' section with Tracks, Courses, Practice, Assessments, and Tutorials. The main content area features the 'Introduction to R' course, labeled as an 'INTERACTIVE COURSE'. It includes a 'Start Course' button, a 'Bookmark' button, and course details: Beginner level, 4 hours duration, 62 videos, 62 exercises, 2,631,849 participants, and 6200 XP. Below this is the 'Course Description' section, which states that the course covers the basics of R, including vectors, factors, lists, and data frames. A 'Read More' link is provided. The course is part of several tracks: Data Analyst, Data Scientist, and Data Scientist Professional. A 'Start Chapter' button is visible for the first chapter, 'Intro to basics'. A chat bubble from 'Adrien' is also present, offering assistance. The bottom of the page shows the 'Gettina Started (0/4)' status.

Qu'est-ce?

- 1 Une fonction?
- 2 Un package?
- 3 base R?
- 4 Tidyverse?
- 5 La différence entre R et RStudio?

Posez beaucoup de questions!!

Les logiciels statistiques

- SPSS (1968)
- SAS (1976)
- STATA (1985)

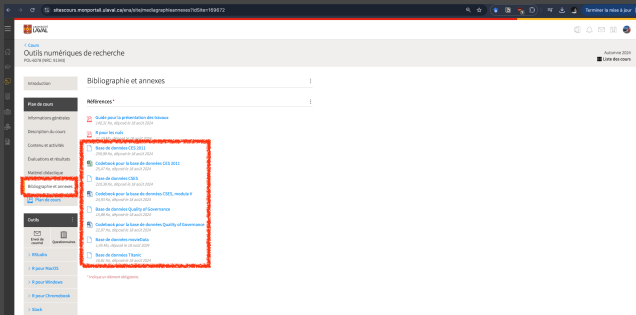
Émergence du langage R

- Avec la philosophie du logiciel libre et de l'open source
- Université d'Auckland, Nouvelle-Zélande
- Construit sur le langage S
- Gratuit, flexible, extensible

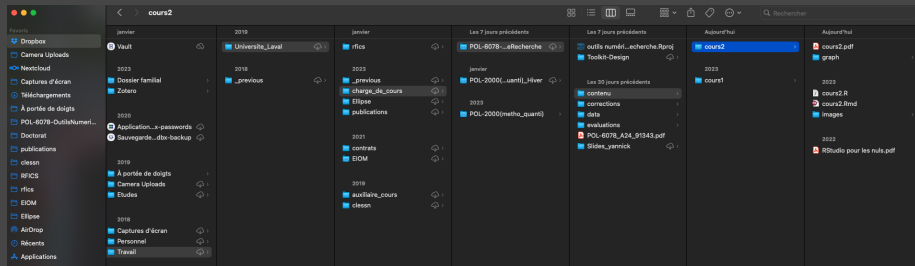
- Comprehensive R Archive Network (CRAN)
- Entrepôt de packages
- Politique stricte

Télécharger les données

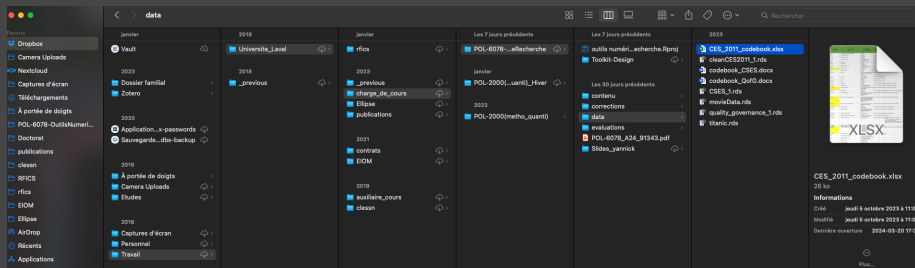
- 1 Créez un nouveau dossier nommé « POL6078 » sur votre ordinateur (je conseille de toujours tout stocker dans un cloud, comme Dropbox)
- 2 Créez un sous-dossier nommé « données » ou « data » (toujours mieux sans accents et sans espace)
- 3 Téléchargez movieData.rds, et autres rds (et les codebooks!)
- 4 Une fois téléchargés, placez-les dans le sous-dossier « data »



Chemin d'arborescence (le fameux « path »)



Stocker ses données



Qu'est-ce qu'un codebook?

- Le « dictionnaire » qui comprend toutes les informations sur le jeu de données
- ESSENTIEL à consulter et à bien comprendre pour un aperçu rapide des données

Vos meilleurs amis

- Google
- Stackoverflow
- r-bloggers
- DataCamp
- Chaines Slack r_et_datacamp
- ... ChatGPT..!

Prochain cours

- Outils de gestion de projet
- Datacamp: Introduction to R: Data frames + Lists
- Installation de Notion.so!
- Création de votre compte GitHub!
- Lectures:
 - Chapitre 2: Les langages de programmation
 - R for Data Science de Hadley Wickam, chapitre 4 (Workflow: code style)

On ouvre RStudio

The screenshot shows the RStudio IDE with a R Markdown file open. The source editor on the left contains R code for a presentation, including comments in French and LaTeX commands for including images. The viewer on the right shows the rendered presentation slides. The console at the bottom displays the command used to run the presentation with Beamer and the resulting output.

```
547  
548 - ## L'ordinateur  
549  
550 \begin{center}  
551 \includegraphics[width=0.6\textwidth]{images/annexe_ordi.png}  
552 \end{center}  
553  
554  
555 - ## Logiciel libre et code source ouvert  
556  
557  
558 \begin{center}  
559 \includegraphics[width=0.6\textwidth]{images/chapitre1.png}  
560 \end{center>  
561  
562  
563 - ## Panique?  
564  
565 - Un cours à la fois  
566 - Je m'occupe de rendre ça concret  
567 - En échange: lecture + relecture + Datacamp + R + installations + pratique  
568 - Autodidaxie et indépendance  
569  
570  
571 - ## À la semaine prochaine!  
572  
573 \begin{center}  
574 \includegraphics[width=0.6\textwidth]{images/rstudio.png}  
575 \end{center>  
576  
577
```

Environment is empty

Files Plots Packages Help Viewer Presentation

Name	Size	Modifié
previous		
Rhistory	0 B	May 3,
cours1.Rmd	19.1 KB	Sep 5,
images		
cours1.pdf	17.4 MB	Sep 5,

Console Terminal Render Background Jobs

```
.../cours1.Rmd  
  
/Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/pandoc -RTS -KS12e -RTS cours1.knit.md --to beamer --from markdown+autolink_bare_uris+tex_math_single_backslash --output cours1.tex --lua-filter /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/library/markdown/markdown.lua/page break.lua --lua-filter /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/library/markdown/markdown.lua/latex-div.lua --variable theme=Madrid --variable colortheme=beaver --variable fonttheme=serif --highlight-style tango --pdf-engine pdflatex --include-in-header /var/folders/3b/f41xb9d3331tp_5p9lvc8cW00p/T/RtmpLH1Et/markdown-str4db30c1e5af.html  
  
processing file: cours1.Rmd
```