

Introduction à R en une heure chrono



Ateliers du PerSciF FPSE, 13 mars 2023

Prof Philippe Grosjean

Université de Mons
STAT for U

UMONS

[https://web.umons.ac.be/STATforU/
statforu@umons.ac.be](https://web.umons.ac.be/STATforU/statforu@umons.ac.be)

STAT for **U**

Introduction

Qui suis-je ?



- Professeur en Faculté des Sciences, service d'Écologie numérique
- Enseigne la science des données de la Bab2 à la Ma2 au cursus de Biologie
- Développeur R depuis 30 ans, traducteur principal de R en français
- Fondateur avec Kathy Huet de la plateforme d'appui en Statistique et Science des Données à l'UMONS, **STAT for U**

STAT for **U**

STAT for U

Conseiller, former, outiller et partager autour des statistiques et science des données
Consultance pour orienter, et même depuis peu pour analyser des données
(communauté UMONS et extérieur)



UMONS – Université de Mons

MENU Services Equipe

STAT for U

Projets Contact

FR

**Plateforme d'appui
en Statistique et
Science des Données
à l'UMONS**

Une structure institutionnelle transdisciplinaire de support en Statistique et plus généralement en Science des Données à l'UMONS pour la recherche, l'enseignement et les services à la société.

STAT for U

Qu'est-ce que R ?

R (<https://www.r-project.org>) est un logiciel pour la **visualisation** et l'**analyse des données**.

Puissant, Open Source, disponible pour Windows, MacOS et Linux. Voir <https://youtu.be/XcBLEVknqvY>.



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Conferences](#)

[Search](#)

[Get Involved: Mailing](#)

[Lists](#)

[Get Involved:](#)

[Contributors](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- [R version 4.2.3 \(Shortstop Beagle\) prerelease versions](#) will appear starting Sunday 2023-03-05. Final release is scheduled for Wednesday 2023-03-15.
- [R version 4.2.2 \(Innocent and Trusting\)](#) has been released on 2022-10-31.
- [R version 4.1.3 \(One Push-Up\)](#) was released on 2022-03-10.

Que peut-on analyser avec R ?

Tout ! La quasi-totalité des méthodes statistiques existantes sont disponibles dans R.

Le site <https://cran.r-project.org> propose près de 20,000 packages avec des extensions pour tout faire.



CRAN
[Mirrors](#)
[What's new?](#)
[Search](#)
[CRAN Team](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Task Views](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

[A3](#)
[AalenJohansen](#)
[AATools](#)
[ABACUS](#)
[abbreviate](#)
[abbyyR](#)
[abc](#)
[abc.data](#)
[ABCRAP](#)
[ABCanalysis](#)
[abclass](#)
[ABCOptim](#)
[ABCp2](#)
[abcrf](#)
[abcrlda](#)
[abctools](#)
[abd](#)
[abdiv](#)
[abe](#)
[abess](#)
[abglasso](#)
[ABHgenotypeR](#)
[abind](#)
[abjData](#)
[abjutils](#)
[ABM](#)
[abmR](#)

Available CRAN Packages By Name

[A](#)[B](#)[C](#)[D](#)[E](#)[F](#)[G](#)[H](#)[I](#)[J](#)[K](#)[L](#)[M](#)[N](#)[O](#)[P](#)[Q](#)[R](#)[S](#)[T](#)[U](#)[V](#)[W](#)[X](#)[Y](#)[Z](#)

Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
Conditional Aalen-Johansen Estimation
Reliability and Scoring Routines for the Approach-Avoidance Task
Apps Based Activities for Communicating and Understanding Statistics
Readable String Abbreviation
Access to Abbyy Optical Character Recognition (OCR) API
Tools for Approximate Bayesian Computation (ABC)
Data Only: Tools for Approximate Bayesian Computation (ABC)
Array Based CpG Region Analysis Pipeline
Computed ABC Analysis
Angle-Based Large-Margin Classifiers
Implementation of Artificial Bee Colony (ABC) Optimization
Approximate Bayesian Computational Model for Estimating P2
Approximate Bayesian Computation via Random Forests
Asymptotically Bias-Corrected Regularized Linear Discriminant Analysis
Tools for ABC Analyses
The Analysis of Biological Data
Alpha and Beta Diversity Measures
Augmented Backward Elimination
Fast Best Subset Selection
Adaptive Bayesian Graphical Lasso
Easy Visualization of ABH Genotypes
Combine Multidimensional Arrays
Databases Used Routinely by the Brazilian Jurimetrics Association
Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association
Agent Based Model Simulation Framework
Agent-Based Models in R

- RStudio permet de travailler confortablement dans R
- Il permet d'éditer des documents **R Markdown** compilables en différentes formes (rapports, présentations, tableaux de bord interactifs, publications scientifiques, ouvrages ...)



PRODUCTS ▾

SOLUTIONS ▾

LEARN &
SUPPORT ▾

EXPLORE
MORE ▾

PRICING 🔍

DOWNLOAD
RSTUDIO

TOP FEATURES

Go further with RStudio Desktop Pro



R par l'exemple

Matériel disponible

Le matériel relatif à cette présentation est disponible ici :

<https://github.com/STATforU/introR1h>

Analyse d'un jeu de données

Grand sondage mondial concernant la protection de l'environnement en 1993.
Données d'Allemagne de l'Ouest.

The screenshot shows the GESIS website interface. At the top, the GESIS logo is on the left, and navigation links for 'Deutsch', 'Contact', and 'Help' are on the right. Below the header is a large blue banner with the ISSP logo and the text 'International Social Survey Programme'. A search bar is located below the banner. The main navigation menu includes 'Home', 'Modules', and 'Search and Data Access'. The breadcrumb trail reads: 'You are here: ISSP > Modules > ISSP Modules by Topic > Environment > 1993'. The main heading is 'ISSP 1993 - "Environment" - ZA No. 2450'. Below this, a section titled 'Latest news on this module' contains the text: 'The latest data set version (1995/ 13.04.2010) : ZA2450 (v1.0.0), doi:10.4232/1.2450'. There are two sidebars: 'Data Access' on the left and 'Related Materials' on the right. The 'Data Access' sidebar contains a link to 'Download of data and documents' and a note that 'Registration is required for data download.' The 'Related Materials' sidebar contains three links: 'ISSP 1993 source questionnaire (PDF)', 'ISSP 1993 codebook (PDF)', and 'Errata and earlier dataset versions' with a note that they are provided on the corresponding study page.

GESIS Leibniz Institute for the Social Sciences

Deutsch Contact Help

ISSP International Social Survey Programme

Search GESIS...

Home Modules Search and Data Access

You are here: [ISSP](#) > [Modules](#) > [ISSP Modules by Topic](#) > [Environment](#) > 1993

ISSP 1993 - "Environment" - ZA No. 2450

Latest news on this module

The latest data set version (1995/ 13.04.2010) : ZA2450 (v1.0.0), doi:10.4232/1.2450

Data Access

- Download of data and documents

[Registration](#) is required for data download.

Related Materials

- [ISSP 1993 source questionnaire](#) (PDF)
- [ISSP 1993 codebook](#) (PDF)
- [Errata and earlier dataset versions](#) are provided on the corresponding study page.

Note: The data sets from **Austria** and **Switzerland** are not included in the integrated file because they were archived too late. These data can be however downloaded as separate data files under the study numbers ZA2793 and ZA3024.

Perception de la science

Quatre questions concernent la manière dont la science est perçue par les répondants en relation avec la protection de l'environnement :

- A. Les gens croient trop souvent à la science, et pas assez aux sentiments et à la foi.**
- B. En général, la science moderne fait plus de mal que de bien.**
- C. Tout changement dans la nature apporté par les êtres humains risque d'empirer les choses.**
- D. La science moderne va résoudre nos problèmes relatifs à l'environnement sans faire de grands changements à notre mode de vie.**

Échelle de Likert à 5 modalités : 1 = tout à fait d'accord à 5 = pas du tout d'accord

Variables explicatives

Trois autres variables sont enregistrées :

- le **genre** (**sex**, 1 = homme, 2 = femme)
- l'**âge** (**age**, 1 = 18-24, 2 = 25-34, 3 = 35-44, 4 = 45-54, 5 = 55-64, 6 = 65+)
- le **niveau d'éducation** (**edu**, 1 = primaire, 2 = second. partim, 3 = secondaire, 4 = univ. partim, 5 = univ. cycle 1, 6 = univ. cycle 2+)

Chargement des données et réencodage

Données au format CSV, fichier `wg93.csv` dans le sous-dossier `data`.

```
wg93 <- read.csv(here::here("data", "wg93.csv"))  
knitr::kable(head(wg93))
```

A	B	C	D	sex	age	edu
2	3	4	3	2	2	3
3	4	2	3	1	3	4
2	3	2	4	2	3	2
2	2	2	2	1	2	3
3	3	3	3	1	5	2
3	4	4	5	1	3	2

Réencodage des modalités pour 'sex'

- sex => 1 = "H" et 2 = "F"
- Utilisation d'outils d'aide dans RStudio : **QuestionR**, **Levels recoding**

```
# Recoding wg93$sex
wg93$sex <- wg93$sex %>%
  as.character() %>%
  fct_recode(
    "H" = "1",
    "F" = "2"
  )
```

- Récupération de l'instruction équivalente

Réencodage des modalités de toutes les variables

```

wg93 <- wg93 %>%
  mutate(
    A = fct_recode(as.character(A), "++" = "1", "+" = "2", "0" = "3", "-" = "4"),
    B = fct_recode(as.character(B), "++" = "1", "+" = "2", "0" = "3", "-" = "4"),
    C = fct_recode(as.character(C), "++" = "1", "+" = "2", "0" = "3", "-" = "4"),
    D = fct_recode(as.character(D), "++" = "1", "+" = "2", "0" = "3", "-" = "4"),
    sex = fct_recode(as.character(sex), "H" = "1", "F" = "2"),
    age = fct_recode(as.character(age), "18-24" = "1", "25-34" = "2", "35-44" = "3",
      "45-54" = "4", "55-64" = "5", "65+" = "6"),
    edu = fct_recode(as.character(edu), "primaire" = "1", "sec. part" = "2",
      "univ. part" = "4", "univ. cycle 1" = "5", "univ. cycle 2" = "6")
  )

```

```
## Warning: Unknown levels in `f`: 1, 2
```

Tableau réencodé

`head(wg93, 15)`

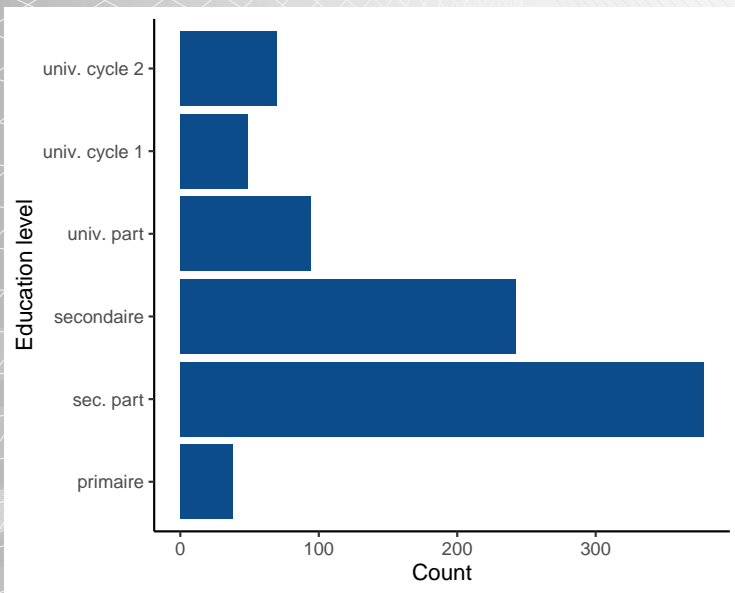
A	B	C	D	sex	age	edu
+	0	-	0	F	25-34	secondaire
0	-	+	0	H	35-44	univ. part
+	0	+	-	F	35-44	sec. part
+	+	+	+	H	25-34	secondaire
0	0	0	0	H	55-64	sec. part
0	-	-	-	H	35-44	sec. part
0	-	+	-	F	55-64	sec. part
0	-	-	+	H	35-44	secondaire
0	+	+	++	H	35-44	sec. part
0	0	+	+	H	35-44	sec. part
-	-	0	-	F	25-34	univ. cycle 2
-	-	++	-	F	25-34	secondaire
0	-	0	+	H	55-64	sec. part
-	-	+	+	H	45-54	sec. part
0	-	0	-	H	35-44	univ. cycle 2

Niveau d'éducation des répondants

- Outil **Esquisse** dans RStudio.
- Graphique en barres horizontales.

```
# Niveau d'éducation  
# ... instructions à ajouter pour créer le graphique...
```


Graphique obtenu



Question B en fonction de edu

B. En général, la science moderne fait plus de mal que de bien.

Tableau de contingence entre question B et edu avec `table()` :

```
wg93_edu <- table(wg93$B, wg93$edu)
knitr::kable(wg93_edu)
```

	primaire	sec. part	secondaire	univ. part	univ. cycle 1	univ. cycle 2
++	6	34	19	6	4	2
+	10	93	47	12	5	7
0	11	95	55	18	11	15
-	7	112	82	37	16	27
-	4	44	39	21	13	19

Dépendance entre question B et edu ?

Test d'indépendance du χ^2 avec `chisq.test()`

```
chisq.test(wg93$B, wg93$edu)
```

```
## Warning in chisq.test(wg93$B, wg93$edu): Chi-squared approximation may be
## incorrect
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

```
## data:  wg93$B and wg93$edu
```

```
## X-squared = 42.764, df = 20, p-value = 0.002196
```

Il y a une dépendance significative au seuil α de 5% entre la réponse à la question B et le niveau d'éducation.

Détails question B *versus* edu

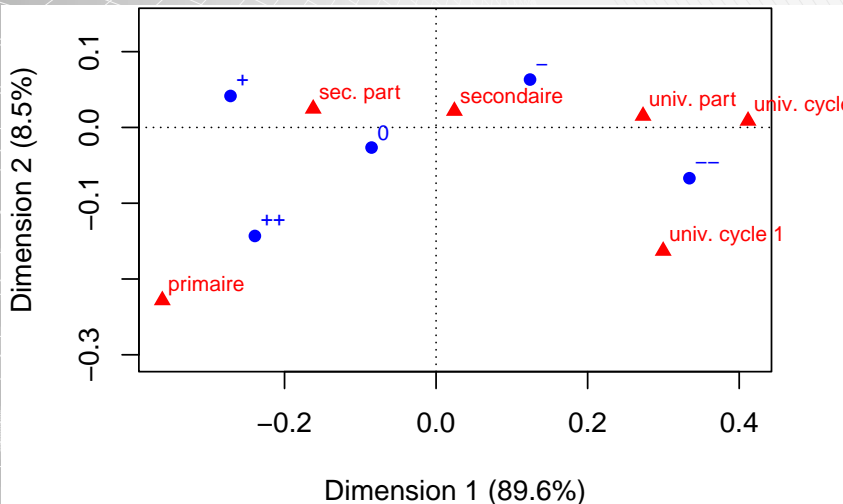
Une **analyse factorielle des correspondances** donne plus de détails (fonction `ca()` du package `{ca}`) :

```
library(ca)
wg93_b_edu <- ca(data = wg93, ~B + edu)
summary(wg93_b_edu, rows = FALSE, columns = FALSE)
```

```
##
## Principal inertias (eigenvalues):
##
## dim      value      %   cum%   scree plot
## 1      0.043989  89.6  89.6  *****
## 2      0.004191   8.5  98.1  **
## 3      0.000914   1.9 100.0
## 4      4e-06000   0.0 100.0
##
## -----
## Total: 0.049097 100.0
```

Graphique de l'AFC - question B *versus* edu

```
par(bg = "white", mar = c(9, 4, 0, 2) + .1)
plot(wg93_b_edu)
```



Dépendance entre question D et edu ?

D. La science moderne va résoudre nos problèmes relatifs à l'environnement sans faire de grands changements à notre mode de vie.

```
chisq.test(wg93$D, wg93$edu)
```

```
## Warning in chisq.test(wg93$D, wg93$edu): Chi-squared approximation may be
## incorrect
```

```
##
```

```
## Pearson's Chi-squared test
```

```
##
```

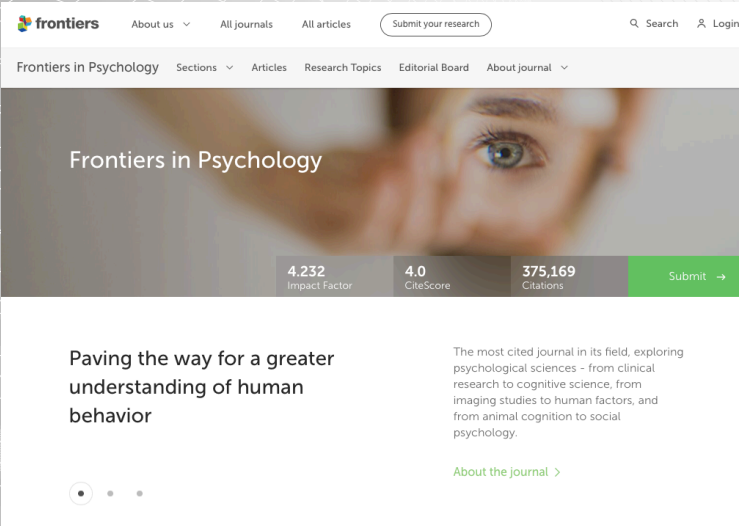
```
## data: wg93$D and wg93$edu
```

```
## X-squared = 25.366, df = 20, p-value = 0.1878
```

Pas de dépendance significative au seuil α de 5% entre la question D et le niveau d'éducation.

Écriture du manuscrit

■ Choix de la revue : **Frontiers in Psychology**



frontiers About us ▾ All journals All articles Submit your research Search Login

Frontiers in Psychology Sections ▾ Articles Research Topics Editorial Board About journal ▾

Frontiers in Psychology

4.232 Impact Factor	4.0 CiteScore	375,169 Citations	Submit →
------------------------	------------------	----------------------	----------

Paving the way for a greater understanding of human behavior

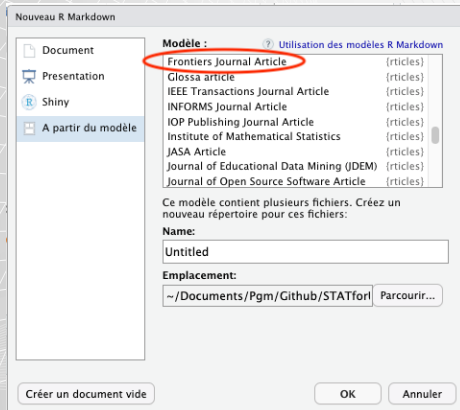
The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

[About the journal >](#)

STAT for U

Formattage du manuscrit pour Frontiers in Psychology

- Template R Markdown pour ce manuscrit (cf dossier docs/frontiers_in_psychology)



Analyse reproductible

- Workflow classique : on **copie-colle** depuis le logiciel de stat vers Word. **Non reproductible !**



- Workflow R Markdown : les instructions qui génèrent les résultats (tableaux, graphiques...) sont dans le rapport directement. **Parfaitement reproductible.**
-> Science Ouverte.

Changement d'idée

- On (re)soumet à la revue **Psychological Science**

The screenshot shows the SAGE Journals website interface. At the top, there's a search bar with the text "Search this journal" and "Enter search terms...". To the right of the search bar are icons for "Access/Profile" and "Cart". Below the search bar, there are dropdown menus for "Browse by discipline" and "Information for". The main header area features the "Psychological Science" title in large white text on a dark red background. Below this, the APS (Association for Psychological Science) logo is visible on the left, and "Journal indexing and metrics" is in the center. To the right of the logo are buttons for "JOURNAL HOMEPAGE" and "SUBMIT PAPER". The main content area is divided into two columns. The left column features a thumbnail of the journal cover and a description: "Psychological Science, the flagship journal of the Association for Psychological Science, is the highest ranked empirical journal in psychology and is truly a leader in the field. The journal publishes cutting-edge research articles and ... | [View full journal description](#)". Below this, it states "This journal is a member of the [Committee on Publication Ethics \(COPE\)](#)". The right column is titled "Browse journal" and contains links for "Current issue", "OnlineFirst", "All issues", and "Free sample".

- Reformattage complet nécessaire (y compris format de bibliographie)

Pas besoin ! Il suffit juste de changer de template (voir dossier docs/psychological_science).

Ressources

- **Contenu de la présentation** : <https://github.com/STATforU/introR1h>
- **SaturnCloud** : <https://saturncloud.io>
- **R** : <https://www.r-project.org>
- **RStudio** : <https://posit.co/products/open-source/rstudio/>
- **STAT for U** : <https://web.umons.ac.be/statforu/>