

# 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

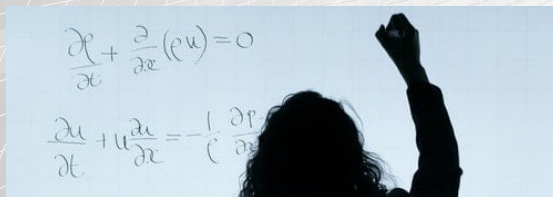
## Qui suis-je ?

- 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

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)



# Qu'est-ce que R ?

R est un logiciel pour la visualisation et l'analyse des données.

Puissant, Open Source, disponible pour Windows, MacOS et Linux

<https://youtu.be/XcBLEVknqvY>

# 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 (Leibniz Institute for the Social Sciences) 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 positioned below the banner. The main navigation menu includes 'Home', 'Modules', and 'Search and Data Access'. A breadcrumb trail indicates the current location: '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. The left sidebar, 'Data Access', contains a link 'Download of data and documents' and a note that 'Registration is required for data download.' The right sidebar, 'Related Materials', 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.

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

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

```
wg93 <- read.csv(here::here("data", "wg93.csv"))  
str(wg93)
```

```
## 'data.frame':    871 obs. of  7 variables:  
## $ A   : int  2 3 2 2 3 3 3 3 3 3 ...  
## $ B   : int  3 4 3 2 3 4 4 4 2 3 ...  
## $ C   : int  4 2 2 2 3 4 2 4 2 2 ...  
## $ D   : int  3 3 4 2 3 5 4 2 1 2 ...  
## $ sex: int  2 1 2 1 1 1 2 1 1 1 ...  
## $ age: int  2 3 3 2 5 3 5 3 3 3 ...  
## $ edu: int  3 4 2 3 2 2 2 3 2 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(educ), "primaire" = "1", "sec. part" = "2",  
      "univ. part" = "3", "univ. cycle 1" = "4", "univ. cycle 2" = "5")  
  )
```

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

```
wg93
```

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

# Niveau d'éducation des répondants

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

```
# Niveau d'éducation
```

## 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  $\chi^2$  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  $\alpha$  de 5% entre la réponse à la question B et le niveau d'éducation.*

## Détails question B *versus* edu

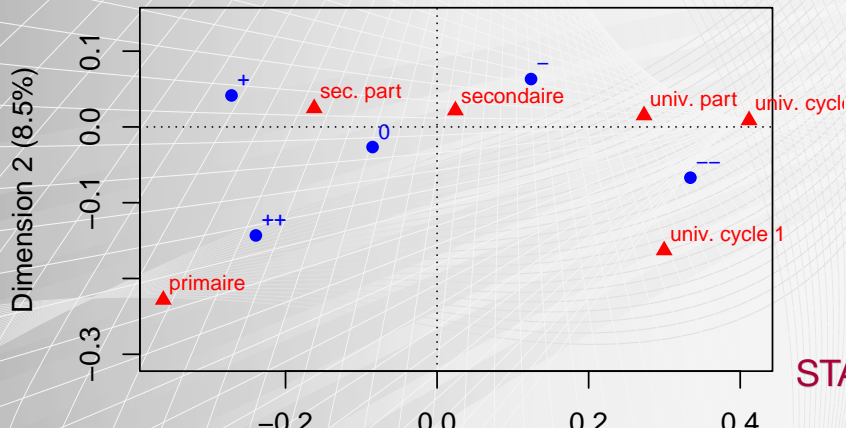
Une **analyse factorielle des correspondances** donne plus de détails (fonction `ca::ca()`) :

```
wg93_b_edu <- ca::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-0600    0.0 100.0
##      -----
## Total: 0.049097 100.0
```

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

```
plot(wg93_b_edu)
```



## Détails question D *versus* 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  $\alpha$  de 5% entre question D et edu.

## Analyse factorielle des correspondances ente D et edu

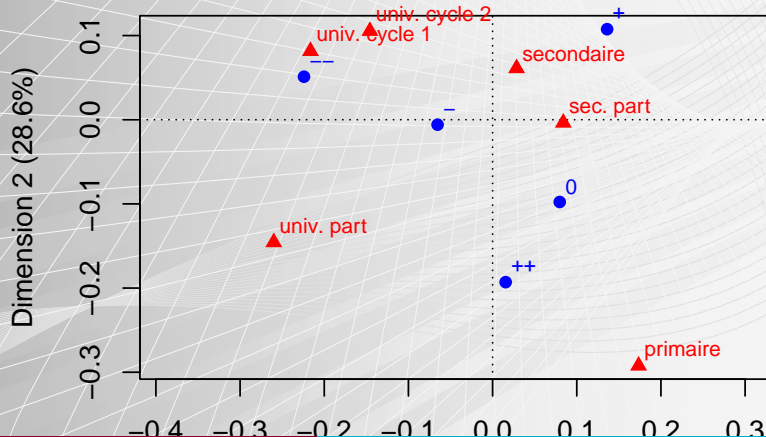
```
wg93_d_edu <- ca::ca(data = wg93, ~D + edu)
summary(wg93_d_edu, rows = FALSE, columns = FALSE)
```

```
##
## Principal inertias (eigenvalues):
##
## dim      value      %   cum%   scree plot
## 1      0.016235  55.7  55.7  *****
## 2      0.008325  28.6  84.3  *****
## 3      0.004048  13.9  98.2   ***
## 4      0.000515   1.8 100.0
##      -----
## Total: 0.029123 100.0
```



# Visualisation graphique avec l'AFC

```
plot(wg93_d_edu)
```



## Qu'en est-il sur l'ensemble des 4 questions ?

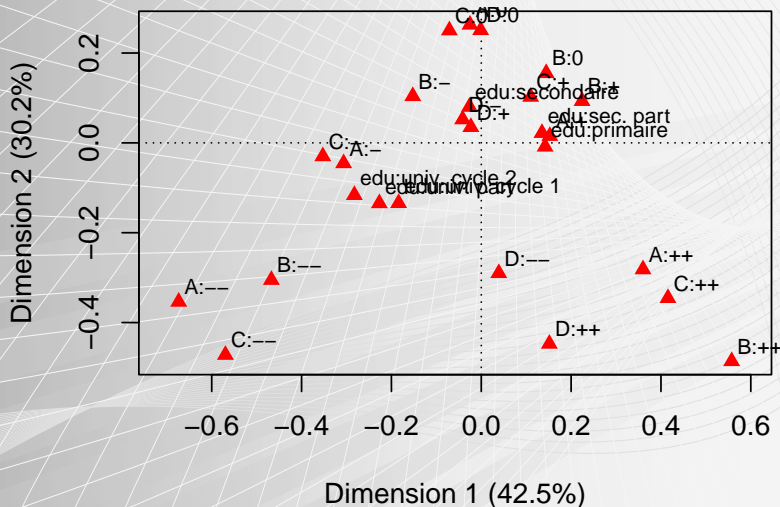
Une **analyse factorielle des correspondances multiples** permet une vue globale sur les 4 questions avec `ca::mjca()` :

```
wg93_ad_edu <- ca::mjca(wg93[, c("A", "B", "C", "D", "edu")], lambda = "adj")
summary(wg93_ad_edu, columns = FALSE)
```

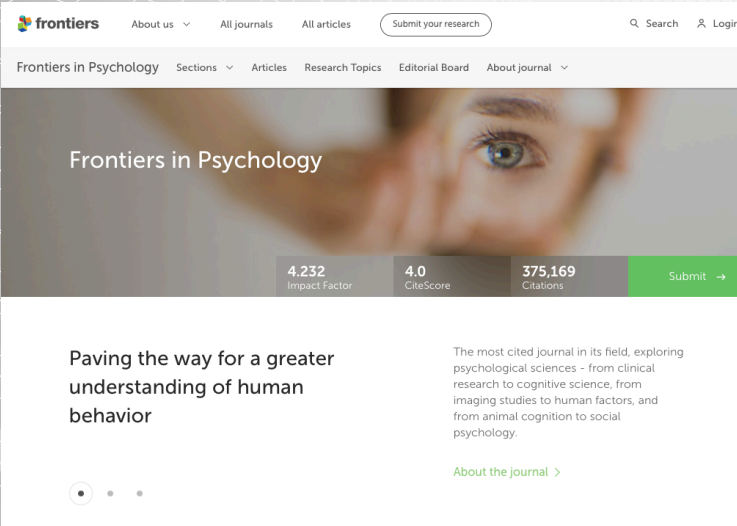
```
##
## Principal inertias (eigenvalues):
##
## dim      value      %   cum%   scree plot
## 1      0.051171  42.5  42.5  *****
## 2      0.036374  30.2  72.7  *****
## 3      0.006103   5.1  77.8  **
## 4      0.004668   3.9  81.7  *
## 5      0.001249   1.0  82.7
## 6      0.000768   0.6  83.4
## 7      0.000254   0.2  83.6
## 8      1.1e-050   0.0  83.6
## 9      1.1e-050   0.0  83.6
## -----
## Total: 0.120367
```

## Graphique des 4 questions

```
par(mar = c(9, 4, 0, 2) + .1)  
plot(wg93_ad_edu)
```



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



The screenshot shows the homepage of the journal 'Frontiers in Psychology'. The header includes the 'frontiers' logo, navigation links for 'About us', 'All journals', 'All articles', and a 'Submit your research' button. There are also search and login icons. Below the header, a secondary navigation bar lists 'Frontiers in Psychology', 'Sections', 'Articles', 'Research Topics', 'Editorial Board', and 'About journal'. The main banner features a close-up of a person's eye and the journal title 'Frontiers in Psychology'. Below the banner, three boxes display key metrics: '4.232 Impact Factor', '4.0 CiteScore', and '375,169 Citations'. A green 'Submit' button with a right arrow is positioned to the right of these metrics. The main content area has the headline 'Paving the way for a greater understanding of human behavior' and a descriptive paragraph: '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.' A link 'About the journal >' is located below the paragraph. At the bottom left, there are three small circular icons, with the first one being filled.

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 >](#)



# Formattage du manuscrit pour Frontiers in Psychology

- Template R Markdown pour ce manuscrit (cf dossier docs/frontiers\_in\_psychology)
- Analyse reproductible

# Changement d'idée


## ■ On (re)soumet à la revue **Psychological Science**


 Search this journal ▾ Enter search terms...  [Advanced search](#)

Access/Profile  Cart 

Browse by discipline ▾ Information for ▾

# Psychological Science

 [Journal indexing and metrics](#) [JOURNAL HOMEPAGE](#) [SUBMIT PAPER](#)



*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](#)

This journal is a member of the [Committee on Publication Ethics \(COPE\)](#).

### Browse journal

- Current Issue
- OnlineFirst
- All issues
- Free sample

### Journal information

- Journal description
- Aims and scope
- Editorial board

### Browse by

- Most recent
- Most read
- Most cited
- Trending

Articles most recently published online for this journal.

SaturnCloud

- ...
- ...
- ...