



SesIndexCreator


Création et visualisation d'indices socioéconomiques

Benoît Lalloué,

S. Deguen, J.-M. Monnez, C. Padilla,
W. Kihal, D. Zmirou-Navier & N. Le Meur



EHESP

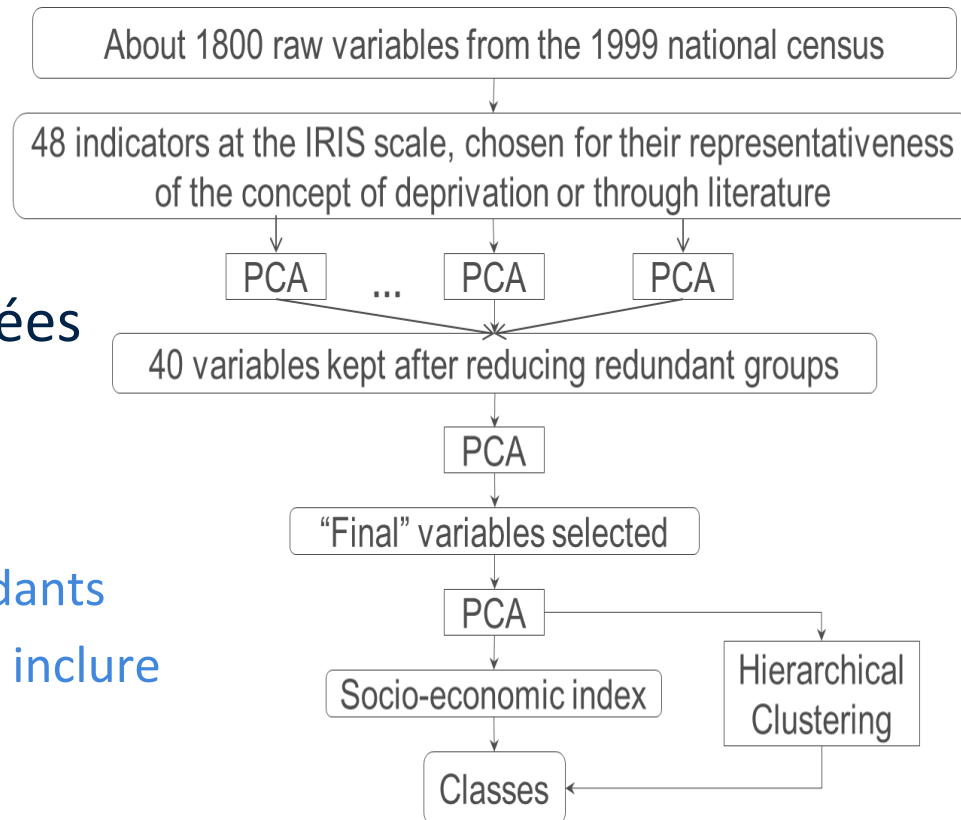


Statut socio-économique

- - Outil fréquent de l'analyse d'inégalités sociales de santé : le **statut socio-économique (SES)**
- Notion complexe, multidimensionnelle : revenu, emploi, éducation, famille, logement, ...
- Une solution pour le prendre en compte : utiliser un **indice socioéconomique** synthétisant l'information
- De nombreux indices socio-économiques existent déjà : Townsend, Carstairs, ...

Une nouvelle procédure

- Méthodologie statistique adaptée et flexible
- Basée sur l'analyse de données
- 3 phases :
 - Réduire les groupes redondants
 - Sélectionner les variables à inclure
 - Synthétiser : Créer l'indice



Lalloué et al.: A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. *International Journal for Equity in Health* 2013 12:21.



Pourquoi un package R ?

- Offrir une intégration sans a priori des variables par rapport aux autres indices
- Faciliter la reproductibilité
- Faciliter l'utilisation pour des non-statisticiens
- Étendre les zones et les contextes d'application

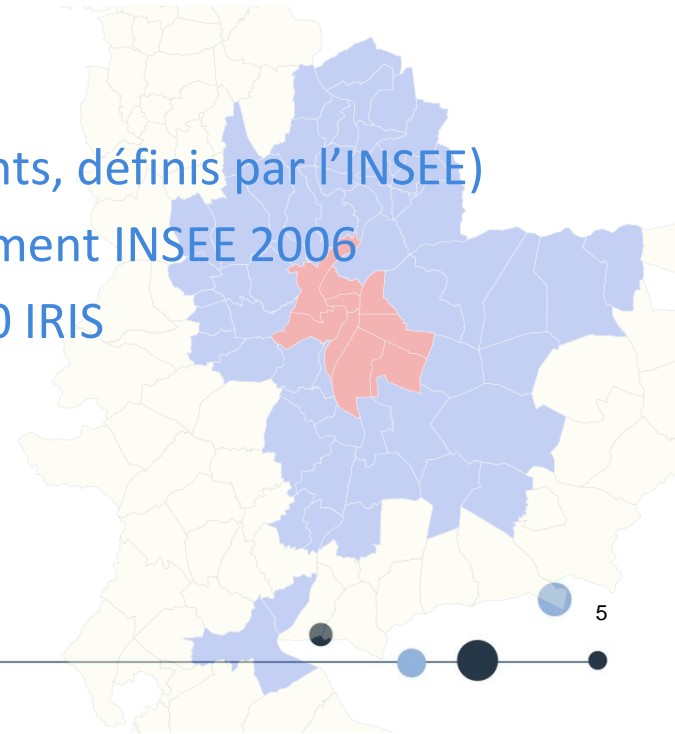
SesIndexCreator

- Disponible sur CRAN et www.equitarea.org

- Dépendances : FactoMineR, class, R ($\geq 3.0.0$)

- Données d'exemple :

- Échelle : IRIS (« quartiers », ~2000 habitants, définis par l'INSEE)
- 54 variables socio-économiques, recensement INSEE 2006
- Agglomération de Lyon (Grand Lyon) : 510 IRIS





SesIndexCreator: Les fonctions

Function	Description
<code>ClassifHC</code>	Internal function: Classification with Hierarchical Clustering (HC)
<code>ClassifInt</code>	Internal function: Classification by intervals
<code>ClassifQuant</code>	Internal function: Classification by quantiles
<code>plot.SesClassif</code>	Plot the results of the classification of a socioeconomic index
<code>plot.SesIndex</code>	Plot the results of the construction of a socioeconomic index
<code>print.SesClassif</code>	Print the classification of a socioeconomic index results
<code>print.SesIndex</code>	Print the creation of a socioeconomic index results
<code>SelectVar</code>	Internal function: Selection of variables
<code>SesClassif</code>	Create categories from a socioeconomic index
<code>SesIndex</code>	Creation of a Socio-Economic Index
<code>SesReport</code>	Creation of a report for <code>SesIndex</code> and <code>SesClassif</code> functions
<code>SesStep1</code>	Internal function: perform the first step of the creation of the socioeconomic index

Exemple d'application

```
> library(SesIndexCreator)
>
> #Import et préparation des données
> SesData = read.table("Lyon_2006_revenu-comp.txt", header=TRUE, sep="\t", row.names=1)
> varnames <- colnames(SesData)[3:ncol(SesData)]
> group1 <- grep("+Chomeur", colnames(SesData), value=TRUE)
> group2 <- grep("+active", colnames(SesData), value=TRUE)
> groupvarnames <- list(group1, group2)
> illus <- rownames(SesData[SesData[, "TYP_IRIS"] %in% c("A", "D"),])
>
> # Application de la procédure
> index <- SesIndex(SesData, varnames=varnames, groupvarnames=groupvarnames, sup=illus)
> index
**Procedure of creation of a socio-economic index**
*The results are available in the following objects:
```

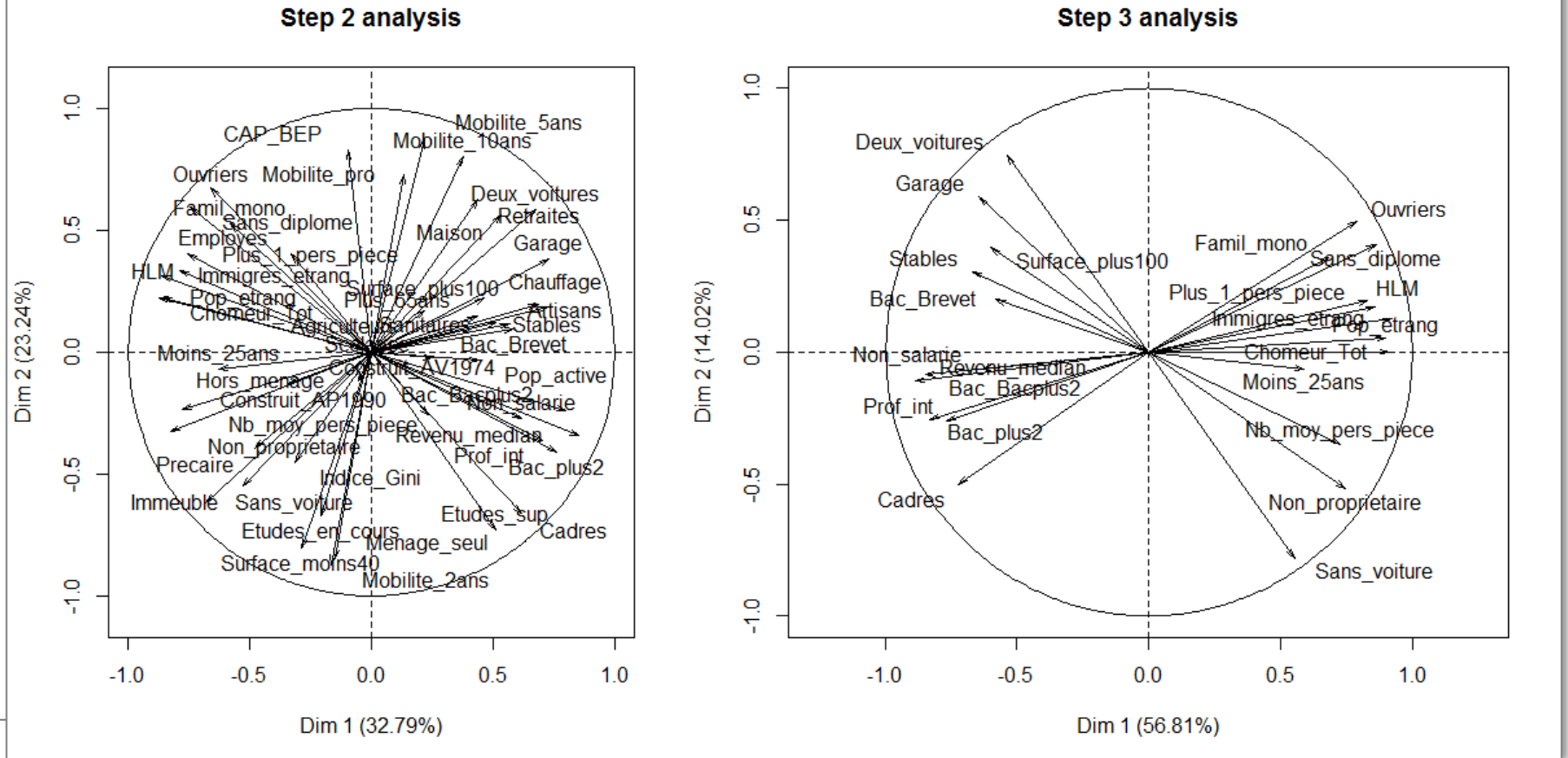
	Name	Description
1	"\$step1"	"Results of the first step"
2	"\$step1\$analysis"	"Detailed results of the PCA on each redundant group"
3	"\$step1\$selection"	"Names of the variable selected in each group"
4	"\$step2"	"Results of the second step"
5	"\$step2\$analysis"	"Detailed results of the selection step (PCA)"
6	"\$step2\$selection"	"Names of the selected variables"
7	"\$step3"	"Results of the third step"
8	"\$step3\$analysis"	"Detailed results of the final PCA"
9	"\$step3\$indices"	"Original dataset and the computed socio-economic index"
10	"\$call"	"Arguments passed in the call of the function"

```
> |
```

Exemple – SesIndex

```
> plot(index, choice="var")
```

Summary of the creation procedure - Variables



Exemple – SesIndex (2)

Variables sélectionnées lors de la réduction des groupes redondants

Variables sélectionnées à l'étape 2 pour constituer l'indice

Coordonnées sur les 2 premiers axes des variables de l'indice (i.e. corrélations variables/indice)

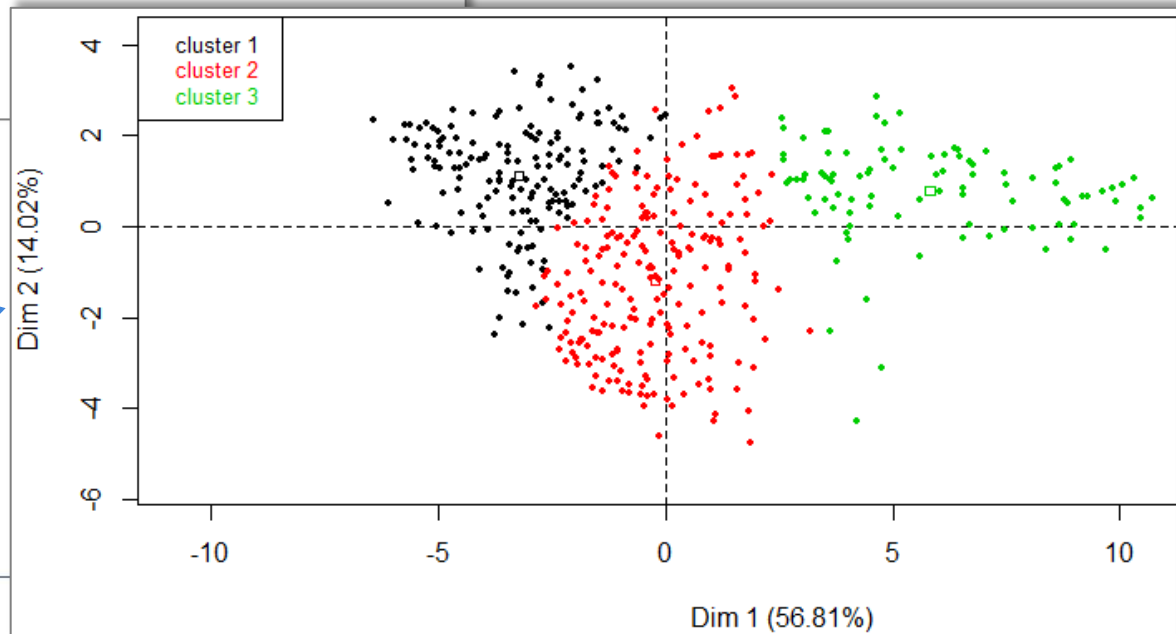
```
> index$step1$selection
[1] "Chomeur_Tot" "Pop_active"
> index$step2$selection
 [1] "Moins_25ans"      "Pop_etrang"      "Chomeur_Tot"
 [4] "Non_salarie"     "Stables"         "Sans_diplome"
 [7] "Bac_Brevet"      "Bac_plus2"       "Bac_Bacplus2"
[10] "Garage"          "Non_proprietaire" "Surface_plus100"
[13] "Sans_voiture"    "Deux_voitures"  "HLM"
[16] "Immigres_etrang" "Revenu_median"  "Plus_1_pers_piece"
[19] "Nb_moy_pers_piece" "Cadres"          "Prof_int"
[22] "Ouvriers"        "Famil_mono"
> round(index$step3$analysis$var$coord[,c(1,2)],2)
      Dim.1 Dim.2
Moins_25ans    0.59 -0.06
Pop_etrang     0.90  0.05
Chomeur_Tot   0.91  0.00
Non_salarie   -0.64 -0.06
Stables       -0.67  0.31
Sans_diplome   0.86  0.41
Bac_Brevet    -0.58  0.20
Bac_plus2     -0.83 -0.25
Bac_Bacplus2 -0.85 -0.09
Garage        -0.65  0.59
Non_proprietaire 0.74 -0.52
Surface_plus100 -0.60  0.40
Sans_voiture   0.55 -0.78
Deux_voitures -0.54  0.74
HLM            0.83  0.19
Immigres_etrang 0.92  0.13
Revenu_median -0.89 -0.11
Plus_1_pers_piece 0.86  0.17
Nb_moy_pers_piece 0.73 -0.35
Cadres         -0.72 -0.50
Prof_int       -0.77 -0.26
Ouvriers       0.79  0.50
Famil_mono     0.68  0.36
```

Exemple – SesClassif

```
> categories <- sesClassif(index, nb=3)
> categories
**Classification of a socio-economic index**
 3 classes were created using: HC+knn

*The results are available in the following objects:
  Name      Description
1 "$analysis" "Detailed results of the HC"
2 "$table"   "Results of the classification"
3 "$call"    "Arguments passed in the call of the function"

*Number of units in each class:
  1  2  3
178 225 103
> |
```



```
> plot(categories$analysis, choice="map",
label="none", draw.tree=F)
```

Exemple – SesReport

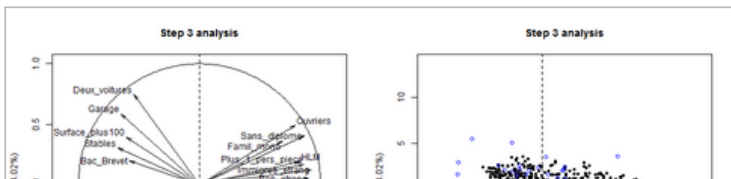
> SesReport(categories)

Step 3: Creation of the index

The final step of the creation perform a Principal Component Analysis on the variables selected in Step 2. The socio-economic index is therefore the first component of the final PC interpreted as a socio-economic component.
The results of the final analysis are :

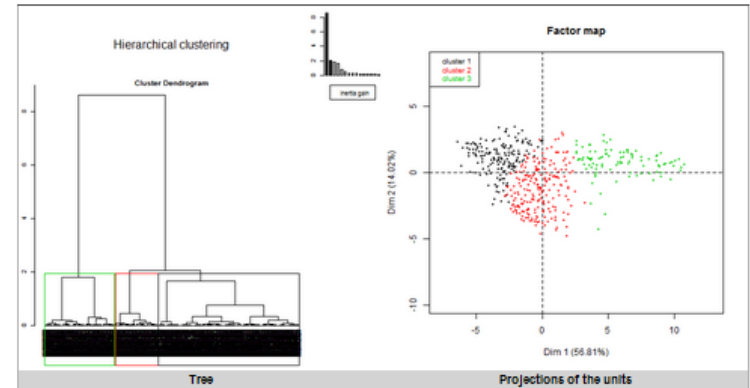
10 first eigenvalues			
	Eigenvalue	Percentage of variance	Cumulative percentage of variance
comp 1	13.07	56.81	56.81
comp 2	3.22	14.02	70.82
comp 3	2.14	9.32	80.14
comp 4	1.07	4.67	84.81
comp 5	0.81	3.53	88.34
comp 6	0.63	2.74	91.08
comp 7	0.39	1.69	92.77
comp 8	0.27	1.17	93.94
comp 9	0.23	1.01	94.94
comp 10	0.19	0.82	95.77

Summary of the two first components						
Variable	Coord 1	Cos2 1	Contrib 1	Coord 2	Cos2 2	Contrib 2
Moins_25ans	0.59	0.35	2.67	-0.06	0	0.12
Pop_etrang	0.9	0.81	6.21	0.05	0	0.09
Chomeur_Tot	0.91	0.84	6.4	0	0	0
Non_salarie	-0.64	0.4	3.09	-0.06	0	0.12
Stables	-0.67	0.45	3.42	0.31	0.09	2.91
Sans_diplome	0.86	0.74	5.69	0.41	0.17	5.13
Bac_Brevet	-0.58	0.34	2.57	0.2	0.04	1.22
Bac_plus2	-0.83	0.69	5.28	-0.25	0.06	2.01
Bac_Bacplus2	-0.85	0.72	5.5	-0.09	0.01	0.23
Garage	-0.65	0.42	3.19	0.59	0.35	10.72
Non_proprietaire	0.74	0.55	4.24	-0.52	0.27	8.28
Surface_plus100	-0.6	0.36	2.77	0.4	0.16	4.86
Sans_voiture	0.55	0.31	2.35	-0.78	0.61	18.91
Deux_voitures	-0.54	0.29	2.22	0.74	0.55	17.2
HLM	0.83	0.69	5.3	0.19	0.04	1.17
Immigres_etrang	0.92	0.85	6.52	0.13	0.02	0.49
Revenu_median	-0.89	0.79	6.02	-0.11	0.01	0.38
Plus_1_pers_piece	0.86	0.74	5.68	0.17	0.03	0.91
Nb_moy_pers_piece	0.73	0.53	4.04	-0.35	0.12	3.74
Cadres	-0.72	0.52	4	-0.5	0.25	7.8
Prof_int	-0.77	0.59	4.5	-0.26	0.07	2.09
Ouvriers	0.79	0.63	4.79	0.5	0.25	7.6
Famil_mono	0.68	0.47	3.56	0.36	0.13	4.02



Classification

After the creation of the socioeconomic index, a classification is performed using HC+knn. 3 classes are created using this technic.



The following tables contain descriptions of each classe by the variables selected in the socioeconomic index (only variables with a significant difference between the class and the whole sample are shown):

Description by the variables of the class 1				
Variable	Mean in category	Overall mean	Sd in category	Overall Sd
Deux_voitures	0.43	0.26	0.15	0.17
Surface_plus100	0.37	0.2	0.19	0.18
Garage	0.76	0.55	0.14	0.23
Revenu_median	22632.43	18606.23	4081.71	5163.72
Bac_Bacplus2	0.28	0.24	0.06	0.07
Non_salarie	0.12	0.09	0.06	0.05
Stables	0.73	0.68	0.09	0.09
Bac_plus2	0.13	0.11	0.04	0.04
Cadres	0.18	0.14	0.09	0.09
Bac_Brevet	0.15	0.13	0.05	0.04
Prof_int	0.21	0.19	0.04	0.06
Famil_mono	0.08	0.09	0.03	0.04
Moins_25ans	0.31	0.34	0.05	0.07
Ouvriers	0.1	0.14	0.05	0.08
Sans_diplome	0.11	0.16	0.05	0.11
HLM	0.07	0.21	0.08	0.25
Plus_1_pers_piece	0.04	0.07	0.02	0.05
Pop_etrang	0.04	0.09	0.02	0.07
Immigres_etrang	0.07	0.13	0.03	0.09
Chomeur_Tot	0.07	0.12	0.03	0.06
Nb_moy_pers_piece	0.61	0.68	0.05	0.1
Sans_voiture	0.11	0.24	0.06	0.15
Non_proprietaire	0.31	0.54	0.15	0.23

Description by the variables of the class 2				
Variable	Mean in category	Overall mean	Sd in category	Overall Sd



Conclusions

- - Facilité d'utilisation : il faut tout de même connaître R
- Facilité d'interprétation : il faut connaître l'ACP
- Procédure de création d'indices statistiquement justifiée, flexible, reproductible à différentes échelles et sur différentes zones
- Package « tout en un » comprenant l'ensemble des outils pour l'appliquer et l'interpréter

Perspectives

- Autres méthodes de classifications ?
- Nouveaux outils d'aide à l'interprétation ?
- Autres moyens de visualisation ?
- Autres possibilités basées sur les retours ...

