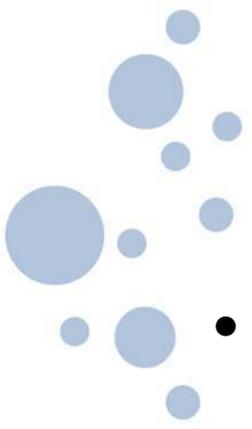# A statistical method to create a neighborhood deprivation index for health inequalities analysis

**B.Lalloué**, J.-M.Monnez, C.Padilla, D.Zmirou-Navier & S.Deguen

# Social health inequalities

- Well documented: the burden of disease is **different between deprived and affluent** populations.

- An issue for many health outcomes: Pregnancy outcomes and infant mortality, Cardiovascular diseases, Respiratory diseases, Cancers, Mental health, …

- Deprivation, a **complex notion**:

  *« The concept of deprivation covers the various conditions, independent of income, experienced by people who are poor, while the concept of poverty refers to the lack of income and other resources which makes those conditions inescapable or at least highly likely »* (**Gordon**, 1995)
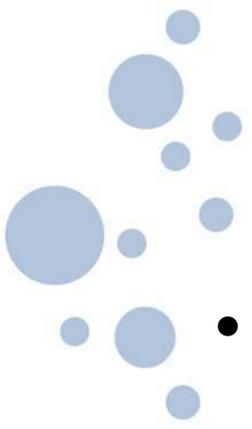
# Socioeconomic status (SES)

- Most frequently used: an **individual** socioeconomic information such as occupation, income, educational level …
  But often difficult to obtain more than a few variables.

- One alternative: use **contextual** SES. Not equivalent to individual SES but often easier to collect and to have more variables (census, …).

- When both are available, contextual SES remains **significant** even after adjustment on individual SES.
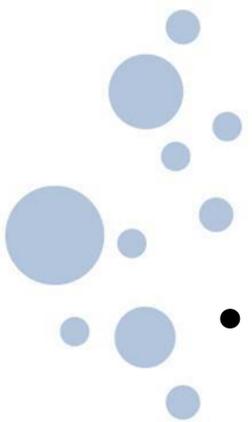
# Deprivation index

- Deprivation is **complex**:
  - One variable seems too few to represent it,
  - Too many variables could cause statistical issues.

- A solution: build a **deprivation index** (or SES index)

- Many indices already exist: Townsend, Carstairs, ...
  **Issues**: arbitrary choice of the variables, small number of variables, very simple "aggregation" methods, ...

- Our purpose : use **more variables**, have an **adapted methodology** to select and aggregate them.
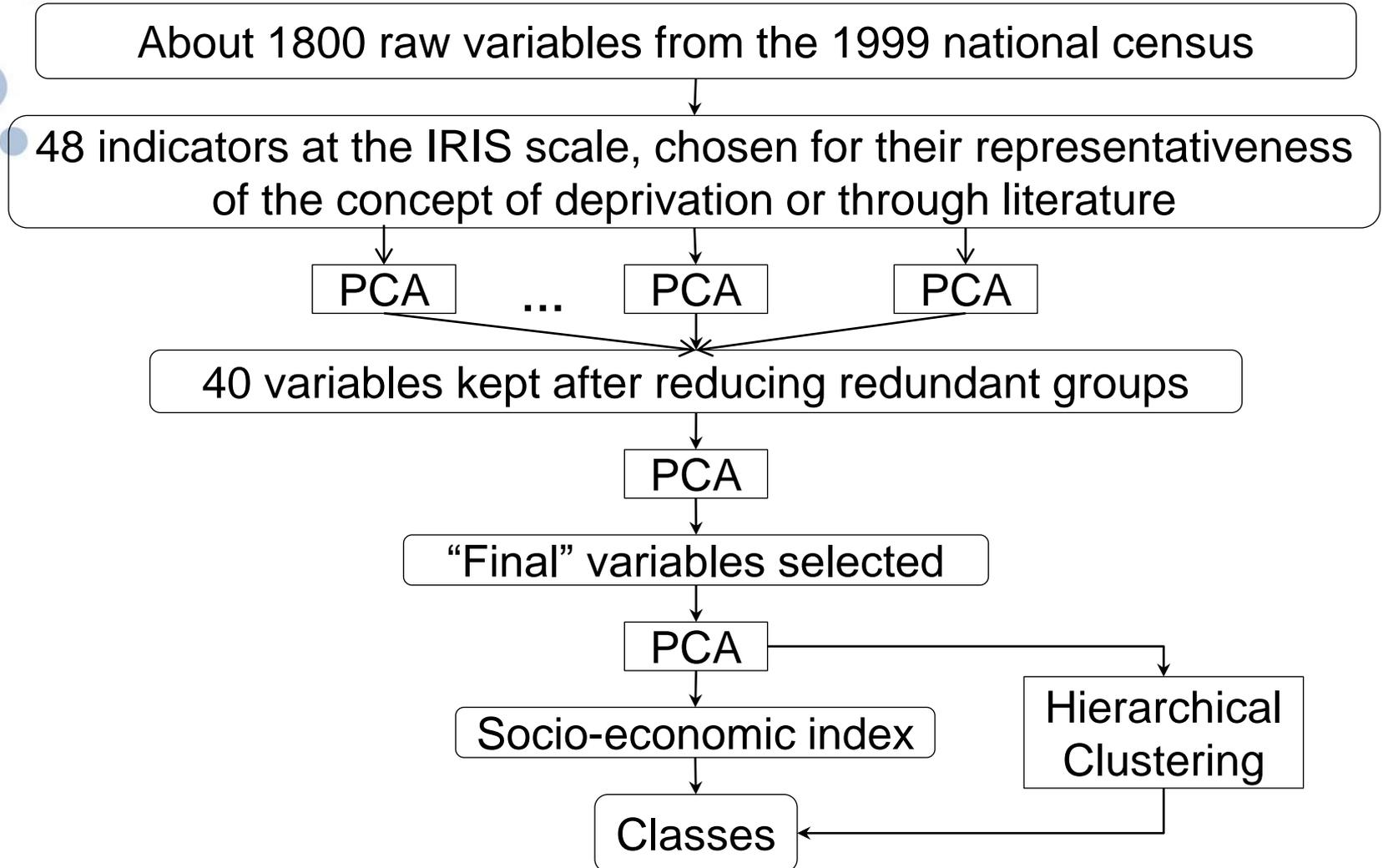
# Study setting & data

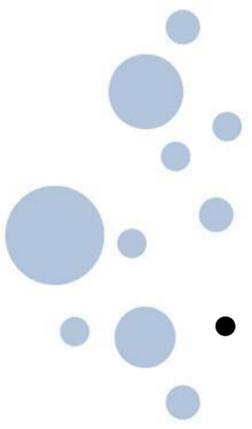- **Lille Métropole**, **Grand Lyon** and **Aix-Marseille** metropolitan areas.

|  | Lille Métropole | Grand Lyon | Aix-Marseille |
|---|---|---|---|
| Population in 1999 (inhabitants) | 1,091,438 | 1,193,384 | 1,349,772 |
| Population in 2007 (inhabitants) | 1,106,885 | 1,260,348 | 1,434,845 |
| Number of municipalities | 85 | 58 | 38 |
| Number of census blocks | 506 | 510 | 630 |
| Number of housing blocks | 475 | 465 | 563 |
| Area (km²) | 611.45 | 527.15 | 1289.59 |

- Statistical unit: **IRIS** (~2000 inhabitants).
  3 distinct types: **housing**, *activity, miscellaneous.*

- **48 indicators** at the IRIS scale from the **1999 national census.**

# Creation procedure

About 1800 raw variables from the 1999 national census

48 indicators at the IRIS scale, chosen for their representativeness of the concept of deprivation or through literature

PCA     ...     PCA     PCA

40 variables kept after reducing redundant groups

PCA

"Final" variables selected

PCA

Socio-economic index

Hierarchical Clustering

Classes

# Deprivation classes

- In applications: often needed to use **deprivation classes** (mapping, potential non-linear relationships, …) created from the index

- **Hierarchical classification** (HC) is used. In our case, two possibilities :

  - Classes distributed on *more than one dimension*: the index alone cannot be used, reduction of the number of classes.

  - Classes distributed *along the first axis of the PCA*: determination of thresholds allowing to create classes based only on the index.
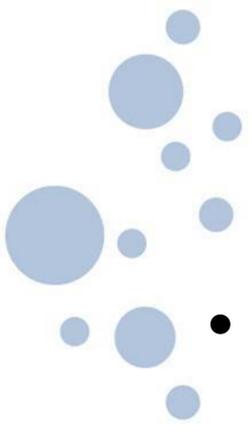
# Results: creation & variances

- Procedure applied on each metropolitan area separately and on the whole set: **4 indices** created.

- Large part of the total variance explained by the index (first component of the PCA):

| | Lille Métropole | Grand Lyon | Marseille urban unit | Global |
|---|---|---|---|---|
| 1$^{st}$ component | 60.73% | 57.79% | 57.29% | 57.22% |
| 2$^{nd}$ component | 12.13% | 16.71% | 14.66% | 11.45% |

- For each metropolitan area, the first component reveals an **opposition** between "deprivation variables" and "affluent variables".

# Variables selections
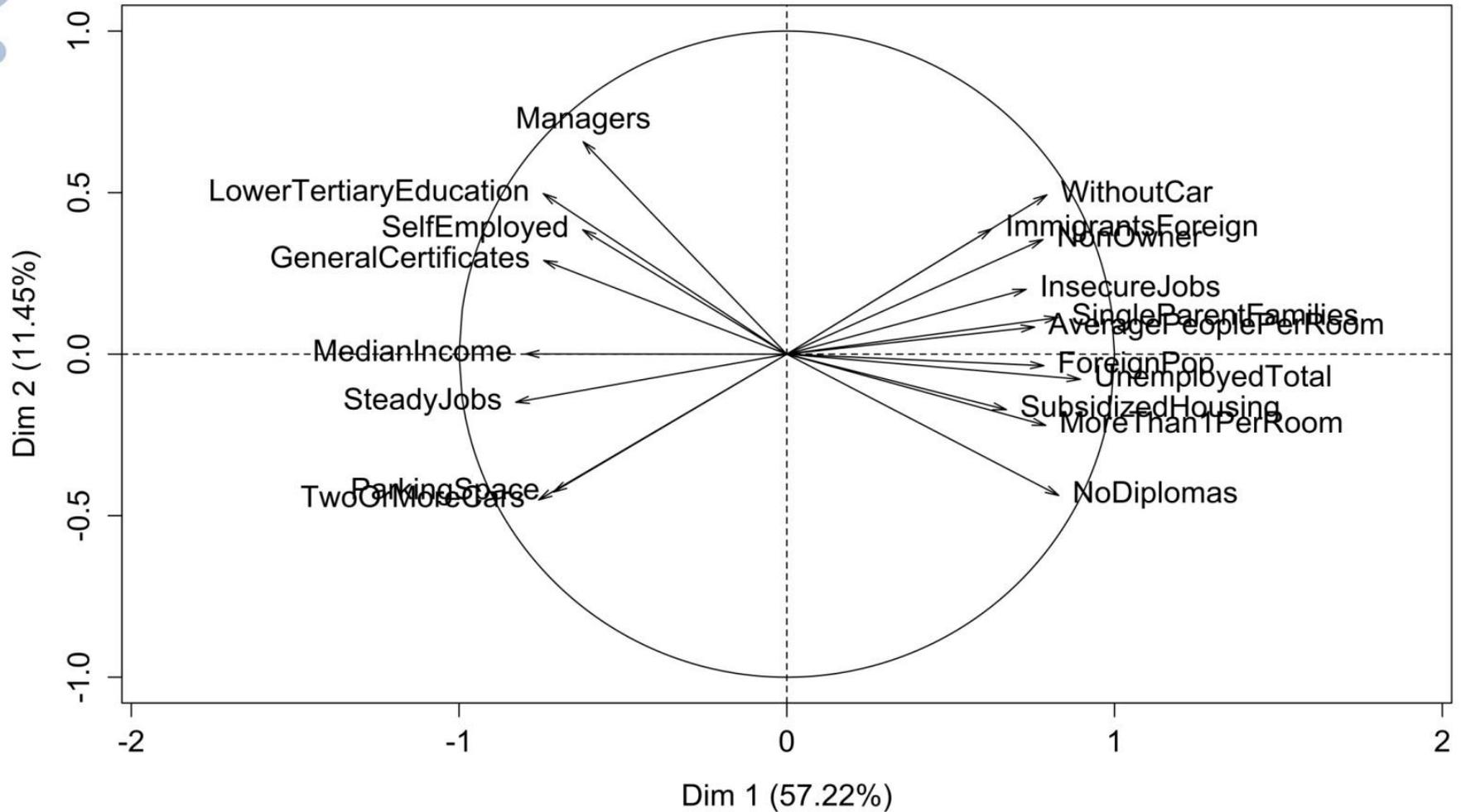
# Variables selections (2)

**Family/Household:** Single-parent families
**Immigration/mobility:** Foreign people, Foreign immigrant since last census
**Education:** Without diploma, With Bac-Brevet, With Bac+2
**Employment/Income:** Self-employed, Steady jobs, Unstable jobs, Median income
**Housing:** Non-owner occupied, More than one person/room, Average number of people/room, Without a car, With 2 or more cars

| *Global* | *Lille* | *Lyon* | *Marseille* |
|---|---|---|---|
| HLM | HLM | HLM | Parking space |
| Parking space | Parking space | Total unemployment | Total unemployment |
| Total unemployment | Unemployment (for more than one year) | Blue-collar workers | Managers |
| Managers | Under the age 25 | Managers | Individual houses |
| | Individual houses | | Multiple dwelling units |
| | Multiple dwelling units | | |

* Excepted for median income and the average number of persons/room, all the variables are proportions

# Comparison between indices

- High correlations between:
  - Each metropolitan area index and the global index.
  - Our index and the Carstairs and Townsend indices.

|  | Lille Métropole | Grand Lyon | Marseille urban unit | Global |
|---|---|---|---|---|
| Global [a] | 0.99 | 1 | 0.99 | / |
| Carstairs | 0.92 | 0.96 | 0.91 | 0.94 |
| Townsend | 0.98 | 0.94 | 0.96 | 0.96 |

[a] When comparing the global index with a city index, global index is restricted to that very city census blocks.



A. Global
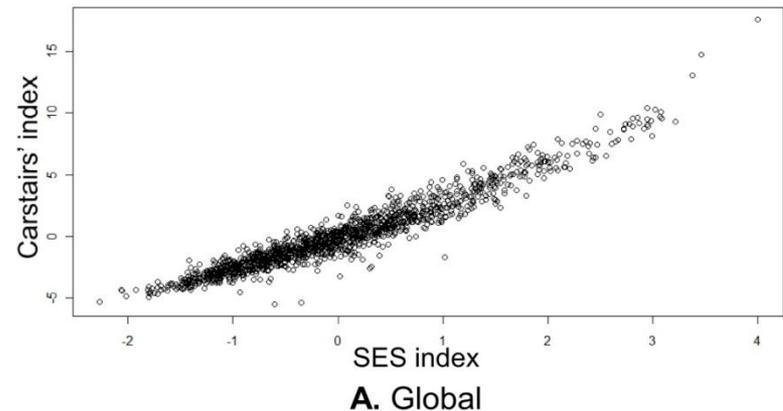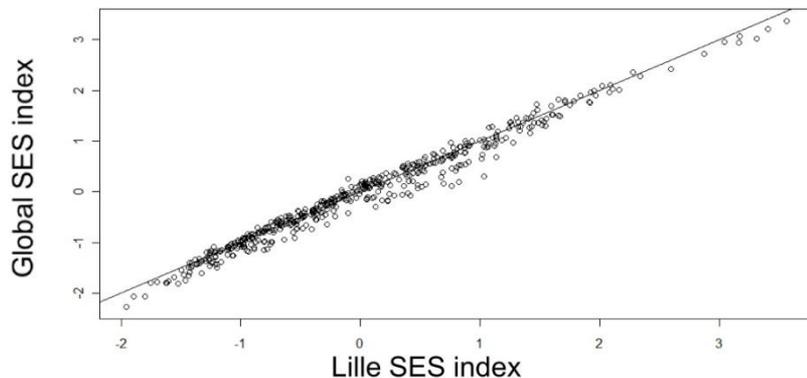
# Classifications

- With HC, optimal partition in **3 classes** along the first axis.

- Comparison of the **concordance rate** between the technics :

| | HC (5) [a] vs. quintiles [b] | HC (3) [a] vs. tiertiles [b] | HC (3) [a] vs. optimal thresholds [b] | Optimal thresholds vs. tiertiles [b] | Carstairs | | Townsend | |
| | | | | | HC (5) [a] vs. quintiles [b] | Optimal thresholds vs. tiertiles [b] | HC (5) [a] vs. quintiles [b] | Optimal thresholds vs. tiertiles [b] |
|---|---|---|---|---|---|---|---|---|
| Lille Métropole | 41% | 78% | 98% | 79% | 38% | 70% | 42% | 78% |
| Grand Lyon | 48% | 74% | 93% | 78% | 47% | 77% | 40% | 75% |
| Marseille urban unit | 48% | 69% | 97% | 67% | 51% | 67% | 50% | 69% |
| Global | 63% | 71% | 97% | 72% | 57% | 70% | 55% | 71% |

[a] Hierarchical Clustering using Principal Components (in parenthesis, the number of classes chosen)

[b] Concordance rate (percent of census blocks categorized into the same class using the two different clustering schemes).
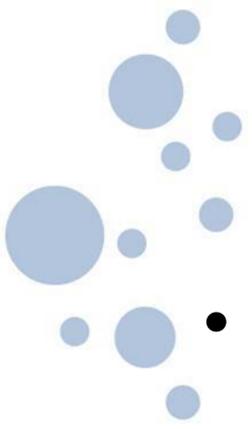
# Health application

- Infant mortality between 2000 and 2009 in the Lyon metropolitan area. Relative rates adjusted by SES.

- Bayesian hierarchical (BYM) model

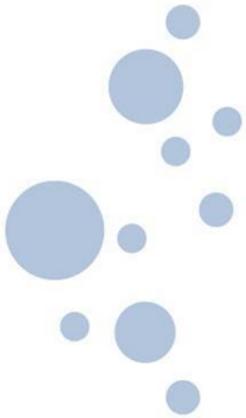| Model | | RR | 95% CI | DIC[b] |
|---|---|---|---|---|
| **Quantitative index** | | 1.24 | 1.14, 1.35 | 1486.99 |
| **Index in 3 classes by terciles** | *Category 1[a]* | 1.00 | - | |
| | *Category 2* | 1.57 | 1.24, 2.00 | 1487.64 |
| | *Category 3* | 1.76 | 1.39, 2.23 | |
| **Index in 3 classes by optimal thresholds (via HC)** | *Category 1[a]* | 1.00 | - | |
| | *Category 2* | 1.58 | 1.29, 1.96 | 1483.17 |
| | *Category 3* | 1.91 | 1.50, 2.43 | |
| **Index in 5 classes by quintiles** | *Category 1[a]* | 1.00 | - | |
| | *Category 2* | 0.72 | 0.52, 1.01 | |
| | *Category 3* | 1.26 | 0.94, 1.70 | 1485.29 |
| | *Category 4* | 1.32 | 0.99, 1.77 | |
| | *Category 5* | 1.53 | 1.14, 2.05 | |

[a] Used as referent category

[b] Deviance Information Criterion
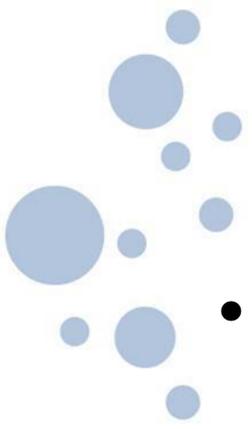
# Conclusion

- Creation procedure of a SES index statistically justified.

- Application on three contrasted French metropolitan areas.

- High number of common variables: common determinants of the deprivation.

- High number of variables: possibilities of interpretation increased.

- Classification should be done carefully in practice.

Thank you for your attention.