Homogeneity Analysis of Pavings

Jan de Leeuw
UCLA Statistics

http://gifi.stat.ucla.edu/pub/roll-call.pdf

- A *paving* is a set of subsets of a *carrier*, a set of *objects*. These subsets are the *data*.

- Each subset is a *characteristic*. Characteristics *cover* objects, objects *belong to* characteristics.

- A characteristic can be *coded* as an *indicator function*, i.e. a function of the set of objects to {0,1}.

- The *carrier* and the number of characteristics are not necessarily finite.

- Objects have *profiles*.

$$O : \Lambda \to 2^{\Omega} \quad \text{and} \quad \mathcal{O} = O(\Lambda)$$

$$L : \Omega \to 2^{\Lambda}$$

$\Lambda = \{1, 2, 3, 4\}$

$\Omega = \{\text{John,Paul,George,Ringo,Yoko,Linda}\}$

$O(1) = \{\text{John,Paul,George,Ringo}\}$

$O(2) = \{\text{John,Paul}\}$

$O(3) = \{\text{John,Yoko}\}$

$O(4) = \{\text{Ringo}\}$

$L(\text{John}) = \{1, 2, 3\}$ $\quad L(\text{Paul}) = \{1, 2\}$

$L(\text{George}) = \{1\}$ $\quad L(\text{Ringo}) = \{1, 4\}$

$L(\text{Yoko}) = \{3\}$ $\quad L(\text{Linda}) = \emptyset$

John	1	1	1	0
Paul	1	1	0	0
George	1	0	0	0
Ringo	1	0	0	1
Yoko	0	0	1	0
Linda	0	0	0	0

- legislators and votes
- students and items
- animals and morphology
- plants and transects
- artefacts and graves
- interviewees and attitude questions

- A *representation* is a mapping of the carrier into a *representation space*. Often metric, often Euclidean.

- In this space a notion of the *size* of a subset is defined. Often a gauge or norm. Often a function of the distances.

- representation

$$\phi : \Omega \rightarrow \mathcal{X}$$

For each characteristic we define its *homogeneity*

$$\mathbf{hom}(\phi, O) = \frac{\mathbf{size}(\phi(O))}{\mathbf{size}(\phi(\Omega))}$$

and we aggregate homogeneities over characteristics

$$\mathbf{hom}(\phi) = \frac{\mathbf{ave}\{\mathbf{size}(\phi(O)) \mid O \in \mathcal{O}\}}{\mathbf{size}(\phi(\Omega))}$$

Usually size() is an *outer measure*, i.e.

$$i) \quad \mathbf{size}(X) \geq 0 \quad \forall X \in \mathcal{X}$$

$$ii) \quad \text{if } X \subseteq Y \text{then } \mathbf{size}(X) \leq \mathbf{size}(Y) \quad \forall X, Y \in \mathcal{X}$$

$$iii) \quad \mathbf{size}(X \cup Y) \leq \mathbf{size}(X) + \mathbf{size}(Y) \quad \forall X, Y \in \mathcal{X}$$

And usually ave() satisfies (for all sequences of reals)

$$i) \quad \min(X) \leq \mathbf{ave}(X) \leq \max(X)$$

$$ii) \quad \text{if } X \leq Y \text{ then } \mathbf{ave}(X) \leq \mathbf{ave}(Y)$$

Representation on the line: possible sizes

- the range
- sum of squares around the mean
- mean deviation around mean or median
- Gini’s mean difference
- other robust measures of scale

Representation space is Euclidean: possible sizes

- length of the MST
- size of convex hull
- size of Weber star
- size of Gifi star
- length of traveling salesman tour
- size of smallest enclosing sphere/ellipsoid

Algorithm

Consider the problem of maximizing

$$\lambda(x) = \frac{\alpha(x)}{\beta(x)}$$

where both numerator and denominator are gauges (homogeneous convex functions). We use Robert's algorithm

$$y^{(k)} \in \partial\alpha(x^{(k)}) \quad x^{(k+1)} \in \partial\beta^{\circ}(y^{(k)})$$

Polar and Subdifferential

$$\eta^\circ(y) = \inf\{\mu \geq 0 \mid x'y \leq \mu\eta(x)\ \forall x\}$$

$$\eta^\circ(y) = \max_{x \neq 0} \frac{x'y}{\eta(x)}$$

$$\partial\eta(x) = \{y \mid \eta(z) \geq \eta(x) + y'(z - x)\ \forall z\}$$

Dedoublement

- For each characteristic we can add its complement to the data
  $$O \in \mathcal{O} \iff \Omega - O \in \mathcal{O}$$
- This means we do not only try make the size of the representation of the characteristic small, but also the size of its complement.
- Generalization of this: missing data.
- Related to this: homogeneity analysis of partitionings (see below).

Order Theorems

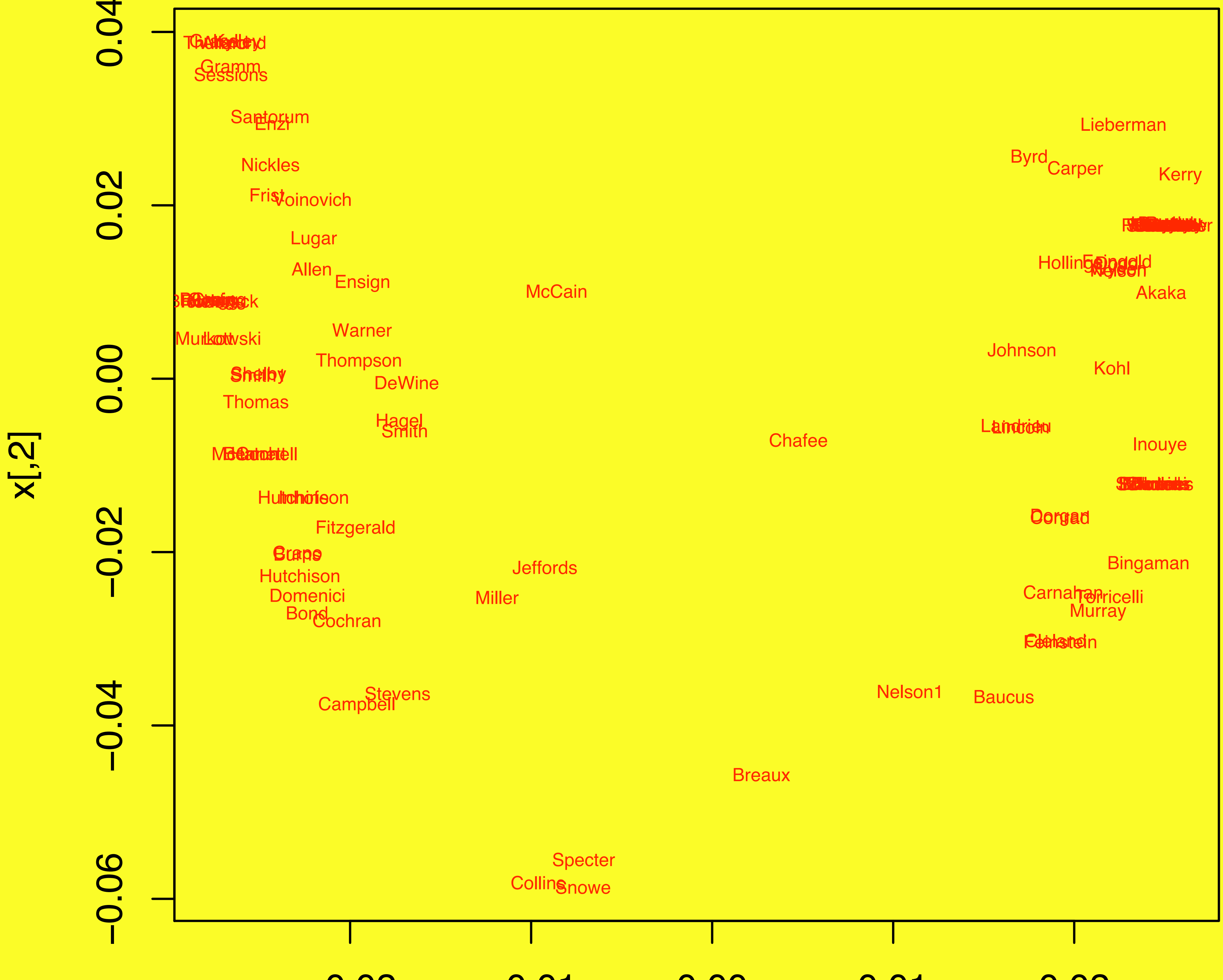
- If we use dedoublement and the least squares size measure, then order is recovered for *monotone* items (Guttman, 1941)
- If we do *not* use dedoublement and the least squares size measure, then order is recovered for *single-peaked* items (Mosteller, 1942).
- Does this generalize to other size measures ?

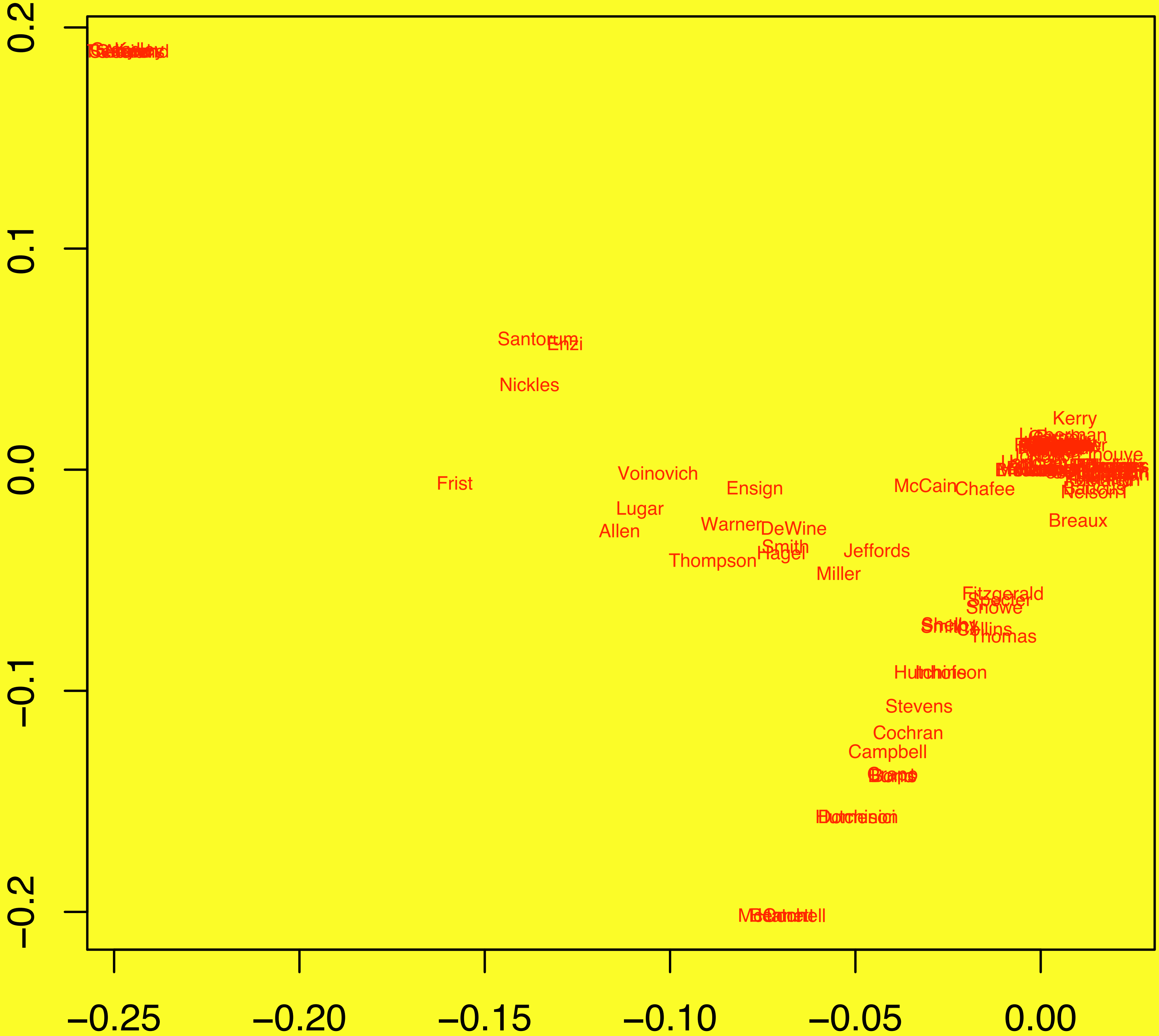
Further Developments

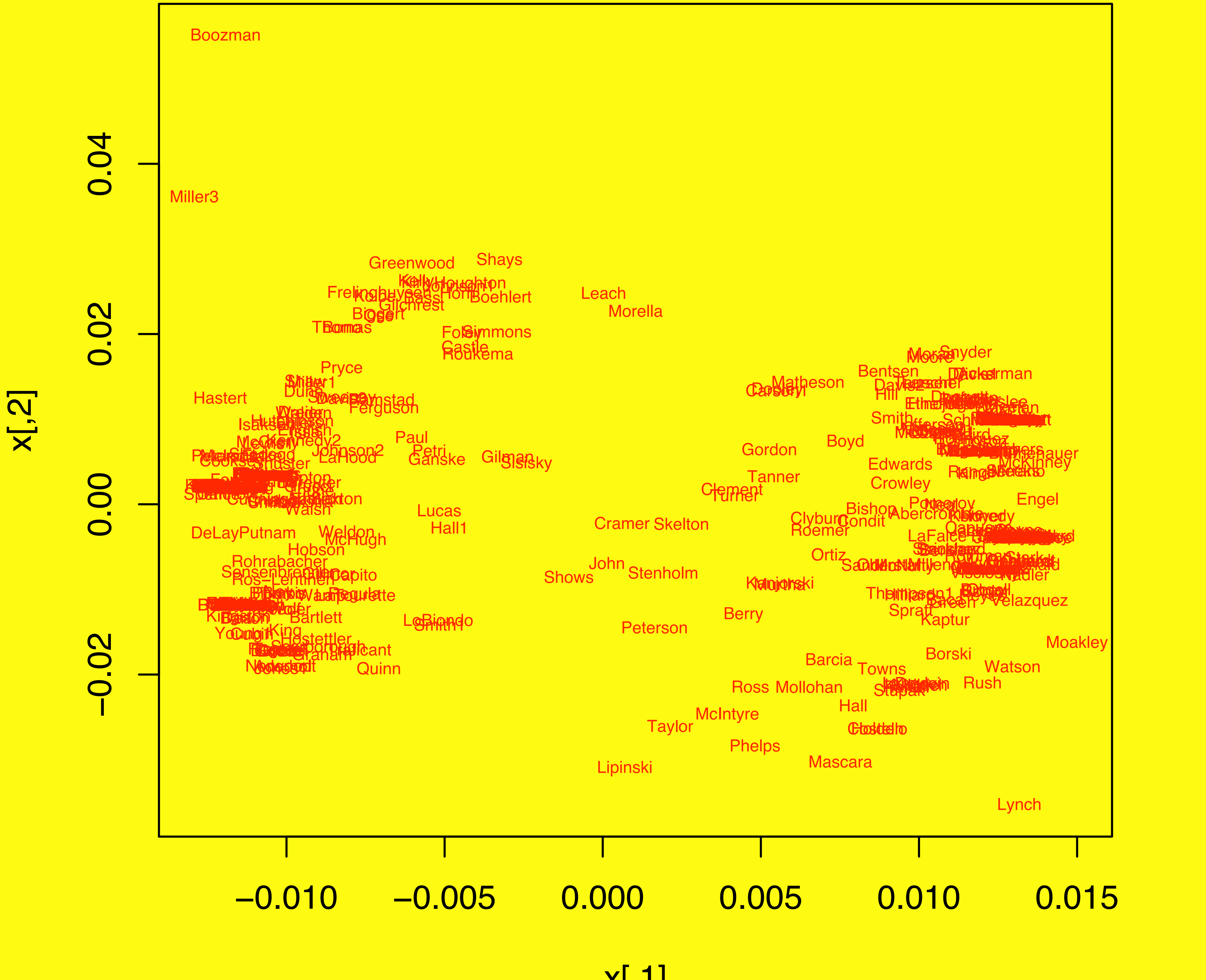
- Fuzzy Pavings. These are mappings of the objects into [0,1] instead of {0,1}.
- Partitions. This generalizes dedoublement to more than two subsets.
- Fuzzy Partitions. For each object we use a partition of unity.

Example from ADA

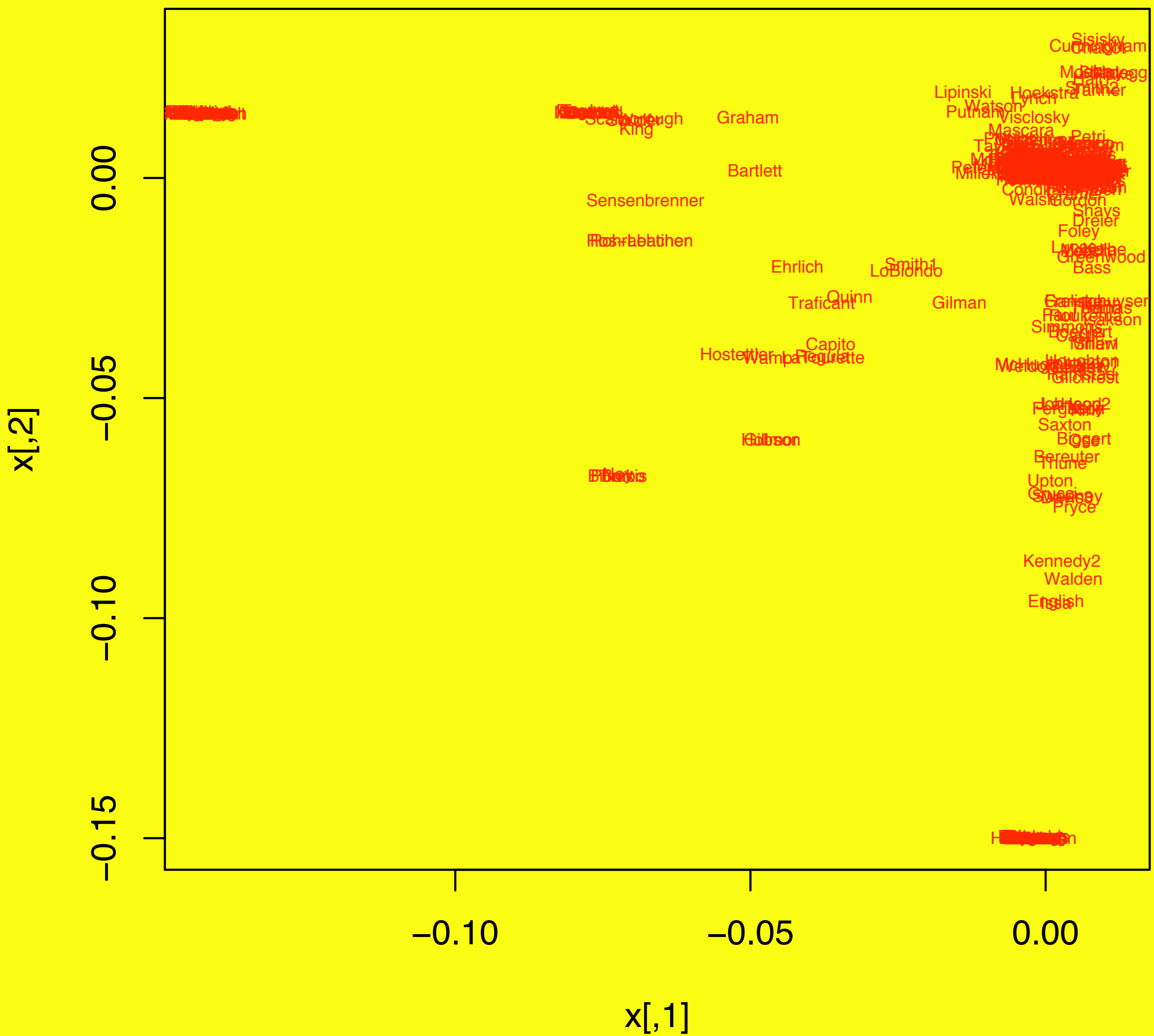
- House and Senate 2000.
- With and without dedoublement.
- Using Gifi stars (i.e. size is squared distance to the centroid).


x[,2]
0.04
0.02
0.00
−0.02
−0.04
−0.06
Gramm
Sessions
Santorum
Enzi
Nickles
Frist
Voinovich
Lugar
Allen
Ensign
McCain
Warner
Thompson
DeWine
Thomas
Hagel
Smith
Chafee
Fitzgerald
Hutchison
Domenici
Bond
Cochran
Jeffords
Miller
Campbell
Stevens
Breaux
Specter
Collins
Snowe
Lieberman
Byrd
Carper
Kerry
Akaka
Johnson
Kohl
Inouye
Bingaman
Torricelli
Murray
Nelson1
Baucus


Santorum
Enzi
Nickles
Frist
Voinovich
Ensign
McCain
Chafee
Kerry
Breaux
Lugar
Allen
Warner
DeWine
Smith
Hagel
Thompson
Jeffords
Miller
Fitzgerald
Specter
Snowe
Shelby
Collins
Thomas
Stevens
Cochran
Campbell
Nelson
0.2
0.1
0.0
−0.1
−0.2
−0.25
−0.20
−0.15
−0.10
−0.05
0.00



x[,2]
x[,1]
0.04
0.02
0.00
−0.02
−0.010
−0.005
0.000
0.005
0.010
0.015
Boozman
Miller3
Greenwood
Shays
Leach
Morella
Boehlert
Gilchrest
Roukema
Castle
Pryce
Hastert
Ferguson
Paul
Petri
Ganske
Gilman
Sisisky
Gordon
Boyd
Matheson
Bentsen
Snyder
Moore
Ackerman
Edwards
Crowley
Tanner
Clement
Turner
Engel
Bishop
Condit
Roemer
Lucas
Hall1
DeLayPutnam
Weldon
McHugh
Hobson
Rohrabacher
Ros−Lehtinen
Capito
Cramer Skelton
John
Shows
Stenholm
Ortiz
Berry
Peterson
Bartlett
LoBiondo
Quinn
Barcia
Towns
Borski
Watson
Rush
Moakley
Ross Mollohan
Hall
McIntyre
Taylor
Phelps
Mascara
Lipinski
Lynch
Kaptur
Velazquez
Spratt
Nadler
McKinney
Walsh
Smith
Hill



x[,2]
0.00
−0.05
−0.10
−0.15
x[,1]
−0.10
−0.05
0.00
Sisisky
Graham
King
Bartlett
Sensenbrenner
Lipinski
Hoekstra
Watson
Putnam
Visclosky
Mascara
Petri
Walsh
Gordon
Shays
Dreier
Foley
Greenwood
Bass
Ehrlich
Smith1
LoBiondo
Quinn
Traficant
Gilman
Capito
Hostettler
Saxton
Bereuter
Upton
Pryce
Kennedy2
Walden
English
Gilchrest