Multiple categorical variables

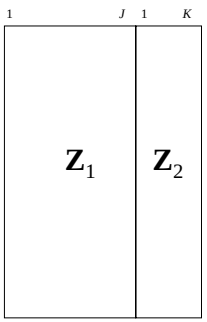
Multiple Correspondence Analysis

Supplementary points

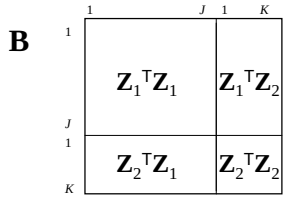
Missing and "middle" responses

Supplementary points

Suppose J substantive categories and K demographic groups



Burt matrix



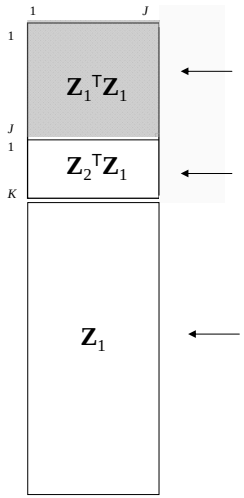
Supplementary points

Burt matrix

B

Suppose J substantive categories and K demographic groups

Indicator matrix of individual respondent (case) data



CA of **B** (or adjusted): standard coordinates the same

shows each demographic category, at the average of the cases in this category

shows each case as a point, at the average of his/her responses

Data set "women94"

Substantive variables: *Do you strongly agree/ agree/ neither...nor.../ disagree/ strongly disagree to these statements...*

A: a working mother can establish a warm relationship with her child
B: a pre-school child suffers if his or her mother works
C: when a woman works the family life suffers
D: what women really want is a home and kids
E: running a household is just as satisfying as a paid job
F: work is best for a woman's independence
G: a man's job is to work; a woman's job is the household
H: working women should get paid maternity leave

Demographic variables

g: gender (1=male, 2=female)
m: marital status (1=married/living as married, 2=widowed, 3=divorced, 4=separated, but married, 5=single, never married)
e: education (0=no formal education, 1=lowest education, 2= above lowest education, 3=higher secondary completed, 4=above higher secondary level, below full university, 5=university degree completed
a: age (1=16-25 years, 2= 26-35, 3=36-45, 4=46-55, 5=56-65, 6=66 and older)

Sample: Spanish sample (year 2002); N=2471 (including missing values)

ISSP 1994 survey on Family and Changing Gender Roles

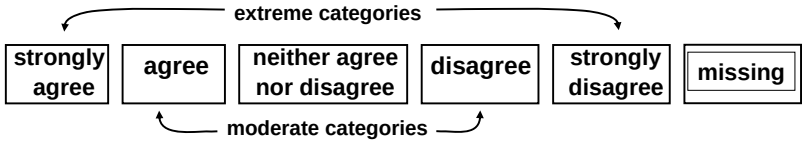
***A* [+]**	***A working mother can establish just as warm and secure a relationship with her children as a mother who does not work***
***B* [–]**	***A pre-school child is likely to suffer if his or her mother works***
***C* [–]**	***All in all, family life suffers when the woman has a full-time job***
***D* [–]**	***A job is all right, but what most women really want is a home and children***
***E* [?]**	***Being a housewife is just as fulfilling as working for pay***
***F* [+]**	***Having a job is the best way for a woman to be an independent person***
***G* [?]**	***Most women have to work these days to support their families***
***H* [+]**	***Both the man and woman should contribute to the household income***
***I* [–]**	***A man's job is to earn money; a woman's job is to look after the home and family***
***J* [?]**	***It is not good if the man stays at home and cares for the children and the woman goes out to work***
***K* [?]**	***Family life often suffers because men concentrate too much on their work***

+ in favour of working women – against women working ? not clear

24 countries (N=33,590)

Data – middle alternative

Common to both surveys is the measurement scale, a 5-point bipolar scale:



We are particularly interested in the middle alternative, how it associates

– with other middle alternatives

– with other response categories

– with the demographic covariates

Of course, there are also missing values, but since we will analyse the data at the nominal level, a missing value is just an additional category

Background & previous research

Presser & Schuman (1980)

The measurement of a middle position in attitude surveys

The Public Opinion Quarterly.
Revised version in *Schuman & Presser, Questions & Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context.* Sage, 1996

Use contingency tables and χ^2 tests in 5 split-ballot experiments.

They assess the consequences of offering/omitting "a **logical middle position**, for example whether one is liberal or conservative could be answered by 'middle-of-the-road' "…. "Although there is a very slight decrease in the proportion of spontaneous 'don't know' responses when the middle alternative is offered, almost all the change in the middle position comes from a decline in the polar positions."

Background & previous research

Andrich, de Jong & Sheridan (1997)

Diagnostic opportunities with the Rasch model for ordered response categories.

In: *Applications of latent trait and latent class models in the social sciences*

Use Rasch modelling.

"…the middle category designated as Neutral, Not Sure or Undecided in the Likert-style response format … should not be treated as an attitude more or less somewhere between a negative and a positive attitude."

Background & previous research

González-Romá and Espejo (2003)

Testing the middle response categories «Not sure», «In between» and «?» in polytomous items.

Psicotema

Use Bock's Nominal Model to verify ordering.

They use different wordings for the middle category, and find that the ordering is verified only for the wording *"in between".*

Hernández, Espejo and González-Romá (2006)

The functioning of central categories Middle Level and Sometimes in graded response scales: Does the label matter?

Psicotema

Background & previous research

O'Muircheartaigh, Krosnick & Helic (2000)

Middle alternatives, acquiescence, and the quality of questionnaire data.

Working paper posted on web.

Use response counts and structural equation modelling

'Approximately half the respondents … were asked to select an answer from among the following five alternatives: strongly agree, agree to some extent, neither agree nor disagree, disagree to some extent, strongly disagree. These responses were coded 5, 4, 3, 2, and 1, respectively. The other half of the respondents … were asked to select an answer from a set of four, omitting the "neither agree nor disagree" option. Answers were coded 5, 3.66, 2.33, and 1, respectively.'

…"We also found evidence of acquiescence response bias in answers to the agree/disagree items..."

Methods

We use correspondence analysis (CA) and several of its variants to visualize and interpret the relative positions of the response categories, at a country level, a respondent level and demographic-subgroup level.

1. Simple CA provides maps of aggregated or count data, for example proportions of question responses for the set of countries.

2. Subset correspondence analysis permits focusing on a particular set of response categories so that we can eliminate missing responses, for example, or restrict our attention to specific categories such as the middle responses.

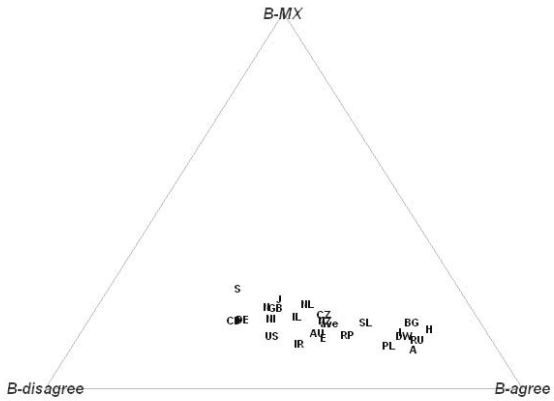
3. Multiple correspondence analysis (MCA) and its refinement joint correspondence analysis (JCA) provide maps of individual-level data, concentrating on the two-way relationships between the questions. In this respect MCA functions like an exploratory factor analysis for categorical data (on nominal scales). Usually we are not interested in individual case points but in mean points for the demographic and other external variables (e.g., age, level of interest…)

4. Canonical correspondence analysis (CCA) focuses on (or partials out) external variables – thus we can explore variation in the responses that is focused on interest, for example, or eliminate aquiescence effects.

B: A pre-school child is likely to suffer if his or her mother works

	B-agree	*B-MX*	*B-disagree*
AU	0.497	0.149	0.354
DW	0.684	0.141	0.175
DE	0.325	0.186	0.490
GB	0.377	0.215	0.408
NI	0.382	0.187	0.431
US	0.406	0.142	0.452
A	0.720	0.106	0.174
H	0.727	0.160	0.113
I	0.671	0.152	0.177
IR	0.472	0.120	0.407
NL	0.439	0.225	0.336
N	0.359	0.218	0.424
S	0.272	0.267	0.461
CZ	0.487	0.197	0.315
SL	0.584	0.177	0.238
PL	0.664	0.115	0.220
BG	0.682	0.178	0.140
RU	0.717	0.130	0.153
NZ	0.498	0.182	0.320
CD	0.306	0.183	0.511
RP	0.564	0.144	0.292
IL	0.434	0.193	0.373
J	0.373	0.238	0.389
E	0.516	0.137	0.346
ave	*0.514*	*0.171*	*0.316*

Question B: ***A pre-school child is likely to suffer if his or her mother works***

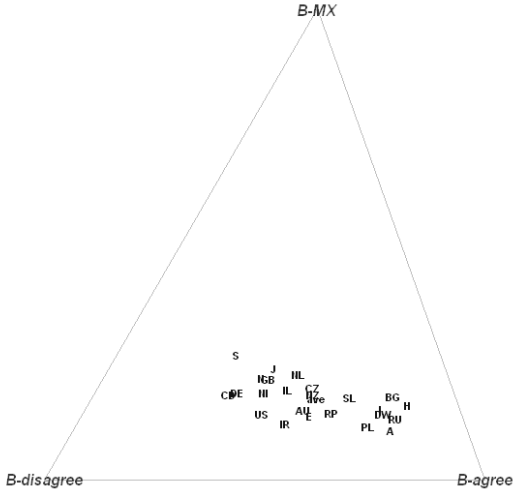


Ternary coordinates: Euclidean distance

B: A pre-school child is likely to suffer if his or her mother works

	B-agree	B-MX	B-disagree
AU	0.497	0.149	0.354
DW	0.684	0.141	0.175
DE	0.325	0.186	0.490
GB	0.377	0.215	0.408
NI	0.382	0.187	0.431
US	0.406	0.142	0.452
A	0.720	0.106	0.174
H	0.727	0.160	0.113
I	0.671	0.152	0.177
IR	0.472	0.120	0.407
NL	0.439	0.225	0.336
N	0.359	0.218	0.424
S	0.272	0.267	0.461
CZ	0.487	0.197	0.315
SL	0.584	0.177	0.238
PL	0.664	0.115	0.220
BG	0.682	0.178	0.140
RU	0.717	0.130	0.153
NZ	0.498	0.182	0.320
CD	0.306	0.183	0.511
RP	0.564	0.144	0.292
IL	0.434	0.193	0.373
J	0.373	0.238	0.389
E	0.516	0.137	0.346
ave	*0.514*	*0.171*	*0.316*

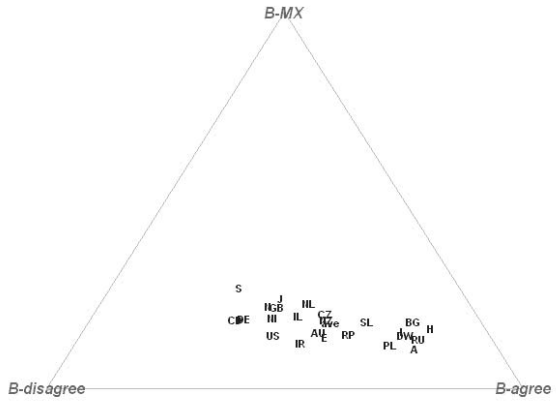
Question B: ***A pre-school child is likely to suffer if his or her mother works***



Stretched ternary coordinates: chi² distance

Animation achieved by saving 101 frames in R of the ternary diagram as the metric smoothly changes from equal weighted Euclidean distance to differentially weighted chi-square distance, then frames saved into a GIF file

Question B: ***A pre-school child is likely to suffer if his or her mother works***



From Euclidean to chi² distance (regular to irregular simplex)

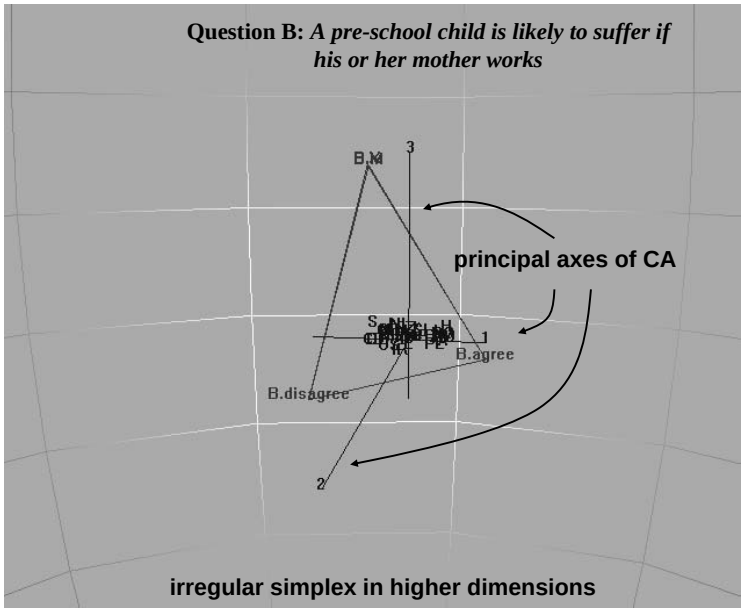
From 3 to 4 categories…

Middle response* (M) *and missing response* (X) *separated

Question B: ***A pre-school child is likely to suffer if his or her mother works***

	B-agree	B-M	B-disagree	B-X
AU	0.497	0.140	0.354	0.009
DW	0.684	0.104	0.175	0.037
DE	0.325	0.150	0.490	0.036
GB	0.377	0.176	0.408	0.040
NI	0.382	0.151	0.431	0.036
US	0.406	0.121	0.452	0.021
A	0.720	0.088	0.174	0.018
H	0.727	0.146	0.113	0.014
I	0.671	0.138	0.177	0.015
IR	0.472	0.082	0.407	0.038
NL	0.439	0.197	0.336	0.028
N	0.359	0.179	0.424	0.039
S	0.272	0.220	0.461	0.047
CZ	0.487	0.178	0.315	0.020
SL	0.584	0.144	0.238	0.033
PL	0.664	0.068	0.220	0.047
BG	0.682	0.067	0.140	0.110
RU	0.717	0.086	0.153	0.044
NZ	0.498	0.157	0.320	0.026
CD	0.306	0.160	0.511	0.022
RP	0.564	0.143	0.292	0.001
IL	0.434	0.174	0.373	0.019
J	0.373	0.199	0.389	0.039
E	0.516	0.090	0.346	0.047
ave	*0.514*	0.138	*0.316*	0.033

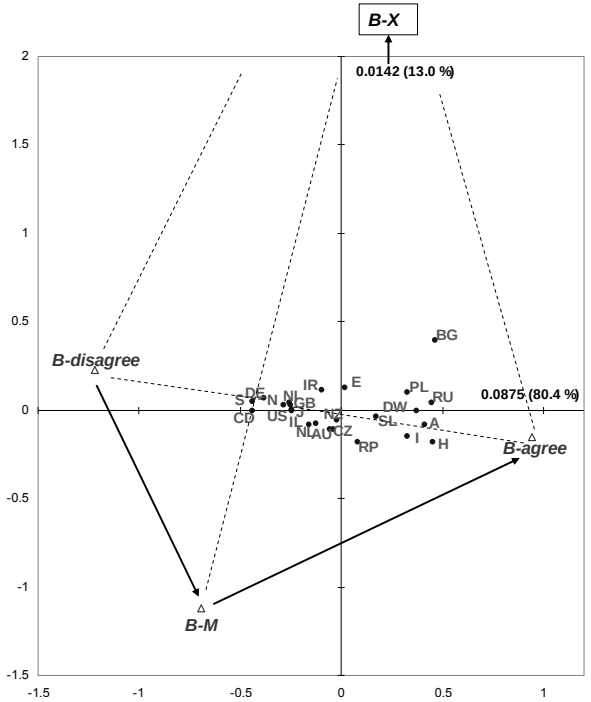
Rotation in three dimensions of the country profiles within a tetrahedron, starting and ending with middle (M) and missing (X) categories lined up (i.e., the two-dimensional map seen previously. Chi² distance evens out the contributions by the categories



Two-dimensional **CA** map, with "*asymmetric scaling*", i.e.

- rows (countries) in principal coordinates as the projections of the profiles
- columns (response categories) in standard coordinates as the projections of the corners of the simplex

93.4% inertia explained



1994: response proportions, 24 x 66 table

Question A
"A working mother can establish just as warm and secure a relationship with her children as a mother who does not work"

	A1	A2	AM	A4	A5	AX
AU	0.183	0.349	0.092	0.282	0.085	0.009
DW	0.369	0.356	0.040	0.158	0.040	0.037
DE	0.627	0.275	0.017	0.050	0.008	0.022
GB	0.176	0.443	0.120	0.188	0.043	0.030
NI	0.155	0.468	0.085	0.201	0.054	0.037
US	0.289	0.409	0.050	0.189	0.048	0.015
A	0.524	0.230	0.033	0.150	0.041	0.021
H	0.327	0.197	0.194	0.161	0.107	0.013
I	0.215	0.409	0.100	0.179	0.094	0.003
IR	0.183	0.428	0.057	0.215	0.100	0.017
NL	0.208	0.490	0.098	0.155	0.026	0.023
N	0.118	0.415	0.129	0.254	0.052	0.032
S	0.222	0.432	0.131	0.152	0.034	0.030
CZ	0.212	0.261	0.096	0.263	0.157	0.012
SL	0.169	0.430	0.077	0.268	0.038	0.018
PL	0.173	0.319	0.065	0.327	0.074	0.041
BG	0.293	0.218	0.071	0.139	0.207	0.072
RU	0.250	0.406	0.069	0.185	0.039	0.050
NZ	0.134	0.402	0.093	0.285	0.059	0.028
CD	0.308	0.412	0.069	0.160	0.035	0.015
RP	0.054	0.568	0.145	0.213	0.018	0.001
IL	0.232	0.429	0.092	0.166	0.064	0.016
J	0.515	0.158	0.147	0.073	0.077	0.030
E	0.141	0.406	0.036	0.337	0.053	0.027

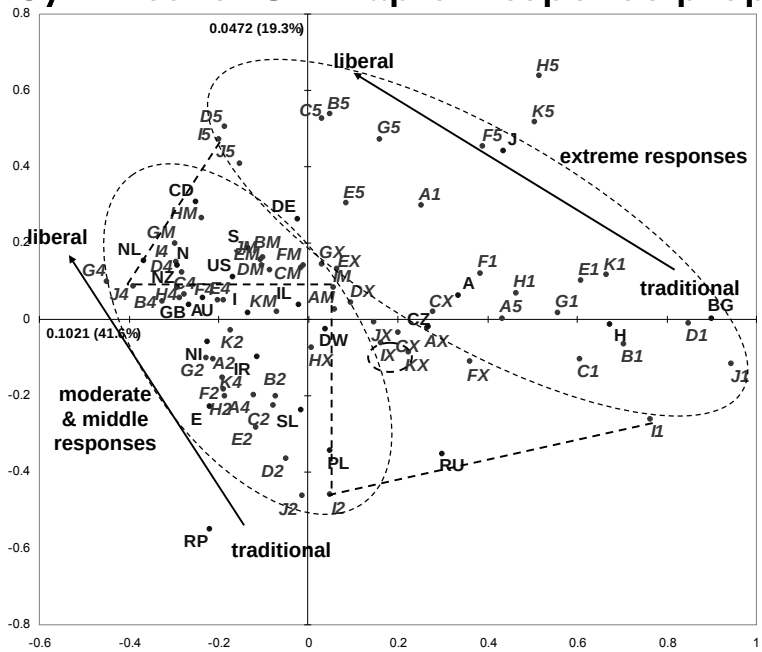
.....

Question K
"Family life often suffers because men concentrate too much on their work"

K1	K2	KM	K4	K5	KX
0.127	0.608	0.138	0.112	0.013	0.003
0.117	0.488	0.148	0.142	0.028	0.077
0.099	0.427	0.180	0.180	0.042	0.072
0.058	0.544	0.180	0.172	0.021	0.025
0.068	0.495	0.181	0.212	0.005	0.040
0.084	0.480	0.198	0.168	0.028	0.041
0.259	0.418	0.126	0.127	0.039	0.032
0.384	0.349	0.171	0.055	0.025	0.017
0.117	0.542	0.154	0.145	0.028	0.014
0.135	0.596	0.071	0.132	0.033	0.032
0.049	0.531	0.217	0.155	0.012	0.036
0.090	0.570	0.179	0.107	0.012	0.043
0.067	0.352	0.286	0.189	0.041	0.065
0.196	0.386	0.220	0.140	0.040	0.019
0.098	0.523	0.181	0.130	0.019	0.048
0.098	0.519	0.122	0.147	0.023	0.091
0.477	0.253	0.083	0.045	0.036	0.105
0.111	0.295	0.243	0.224	0.050	0.077
0.107	0.606	0.138	0.119	0.011	0.018
0.099	0.476	0.210	0.163	0.031	0.020
0.046	0.468	0.213	0.251	0.021	0.002
0.200	0.494	0.148	0.103	0.023	0.032
0.245	0.264	0.138	0.081	0.235	0.037
0.078	0.510	0.088	0.226	0.021	0.076

e.g., Spain
14.1% strongly agree to Qu.A (average across countries: 24.9%)
7.8% strongly agree to Qu.K (average across countries: 13.6%)

"Symmetric" CA map of response proportions



1 **strongly agree**
2 **agree**
M **middle**
4 **disagree**
5 **strongly disagree**
X **missing**

Qu.I
"Man's job is to earn money; woman's job is to look after home and family"

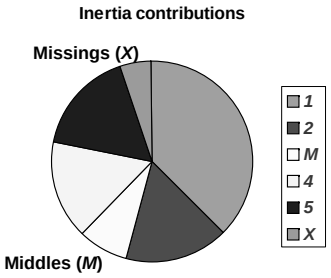
23-dimensional

Total inertia = 0.2453

60.9% inertia explained in 2-d CA map

Decomposition of inertia across categories

Category	Inertia
1	0.0921
2	0.0400
M	0.0197
4	0.0394
5	0.0418
X	0.0123
total	0.2453



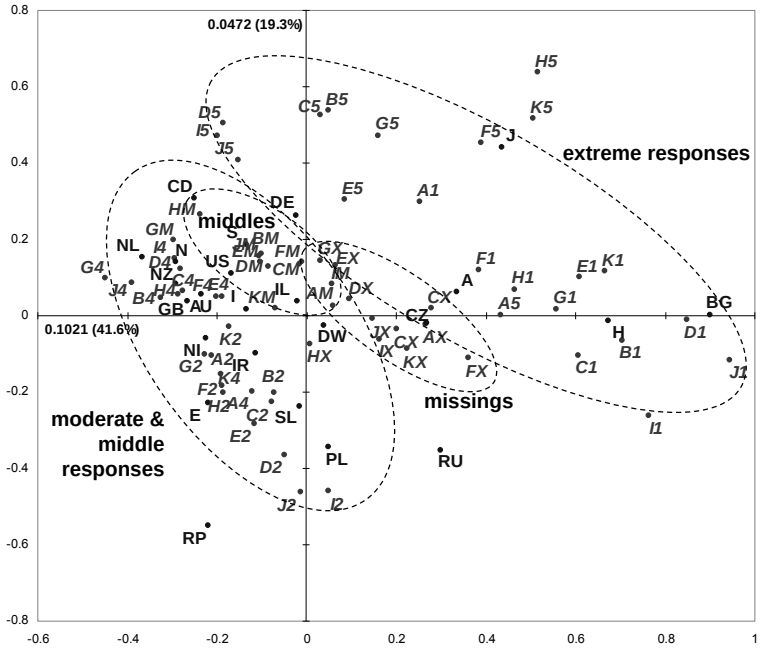
middles & missings account for 0.03200 of the total inertia of 0.2453, i.e. , 13.0%; "extreme" responses account for 0.1339, i.e. 54.6%.

Methods

We use correspondence analysis (CA) and several of its variants to visualize and interpret data at a country level as well as respondent and demographic-subgroup level.

1. **Simple CA provides maps of aggregated or count data, for example proportions of question responses for the set of countries.**
2. **Subset correspondence analysis permits focusing on a particular set of response categories so that we can eliminate missing responses, for example, or restrict our attention to specific categories such as the middle responses (Greenacre & Pardo, *SMR*, 2006)**
3. **Multiple correspondence analysis (MCA) and its refinement joint correspondence analysis (JCA) provide maps of individual-level data, concentrating on the two-way relationships between the questions. In this respect MCA functions like an exploratory factor analysis for categorical data (on nominal scales). Usually we are not interested in individual case points but in mean points for the demographic and other external variables (e.g., age, level of interest…)**
4. **Canonical correspondence analysis (CCA) focuses on (or partials out) external variables – thus we can explore variation in the responses that is focused on interest, for example, or eliminate aquiescence effects.**

CA of proportions of response categories

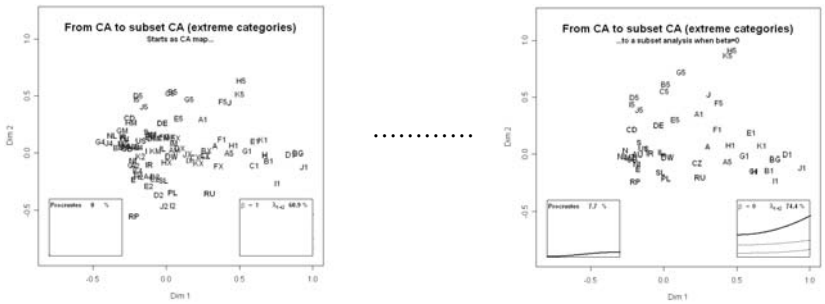


Animation achieved by changing weight

The animations which link methods are achieved by either reducing the mass of certain points or transferring mass between points.

For example, to show the missings and all moderate responses in this example) we multiply these dummy variables by β where β starts at 1 (i.e., the regular MCA) and decreases in small steps of 0.01 until 0 (i.e., the subset MCA). At each step a hybrid of a regular MCA and a subset MCA is performed, maintaining the margins of the table constant.

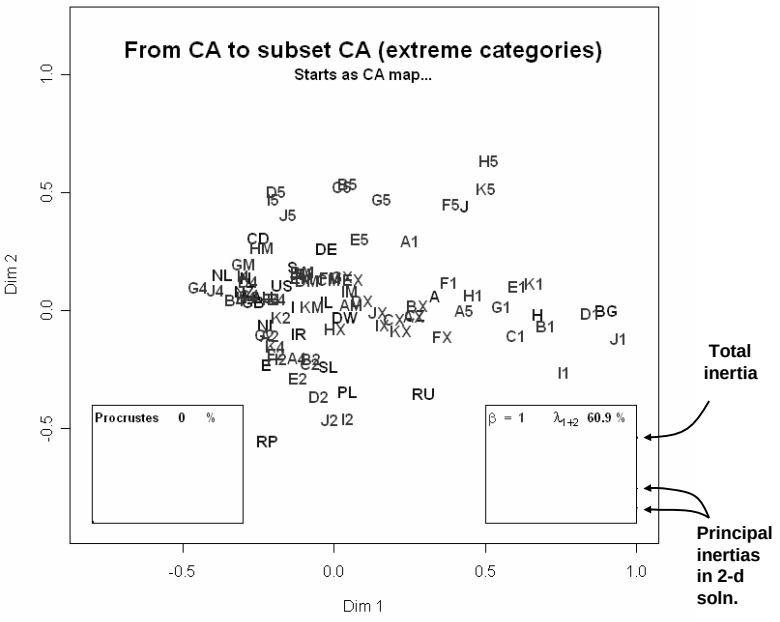
β = 1, 0.99, 0.98, 0.97, ……, 0.04, 0.03, 0.02, 0.01, 0



CA to subset CA

Data: **WOMEN WORKING, stacked frequencies, *N*=33590**

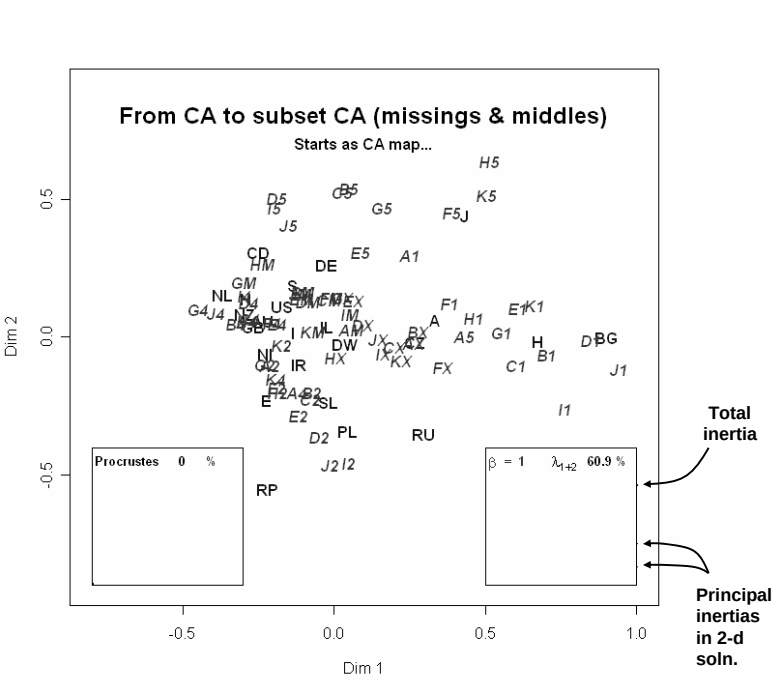
Method: **Contribution of categories not in subset reduced by factor β, from 1 (CA) to limiting case of 0 (subset CA), always using original masses for centring and weighting)**



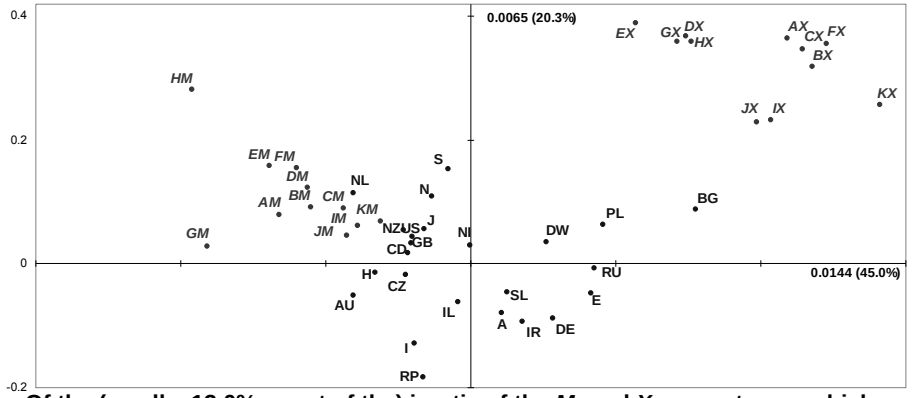
CA to subset CA

Data: **WOMEN WORKING, stacked frequencies, *N*=33590**

Method: **Contribution of categories not in subset reduced by factor β, from 1 (CA) to limiting case of 0 (subset CA), always using original masses for centring and weighting)**



Subset CA of proportions of middles and missings



Of the (small – 13.0% – part of the) inertia of the *M*- and *X*-percentages, which is contained in a 22-dimensional space, 65.3% is explained in this map. The fact that all the *M*'s are together and all the *X*'s together and separate, does not reflect category associations at an individual level. For example, the proximity of Spain (E) and Russia (RU) means that their percentages of *M*- and *X*-responses are similar (less *M*'s than average, more *X*'s than average)

Methods

We use correspondence analysis (CA) and several of its variants to visualize and interpret data at a country level as well as respondent and demographic-subgroup level.

1. **Simple CA provides maps of aggregated or count data, for example proportions of question responses for the set of countries.**
2. **Subset correspondence analysis permits focusing on a particular set of response categories so that we can eliminate missing responses, for example, or restrict our attention to specific categories such as the middle responses.**
3. **Multiple correspondence analysis (MCA) and its refinement joint correspondence analysis (JCA) provide maps of individual-level data, concentrating on the two-way relationships between the questions. In this respect MCA functions like an exploratory factor analysis for categorical data (on nominal scales). Usually we are not interested in individual case points but in mean points for the demographic and other external variables (e.g., age, level of interest…)**
4. **Canonical correspondence analysis (CCA) focuses on (or partials out) external variables – thus we can explore variation in the responses that is focused on interest, for example, or eliminate aquiescence effects.**

Looking at respondent-level data

To investigate response behaviour at individual respondent level we pass from CA to multiple correspondence analysis (MCA). The classic definition of MCA is the CA of the data coded in an indicator matrix, i.e., as dummy variables, one variable for each response category

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
1	3	3	5	5	1	1	1	5	5	4
9	2	2	2	2	4	4	4	2	2	2
.										
.										
.										

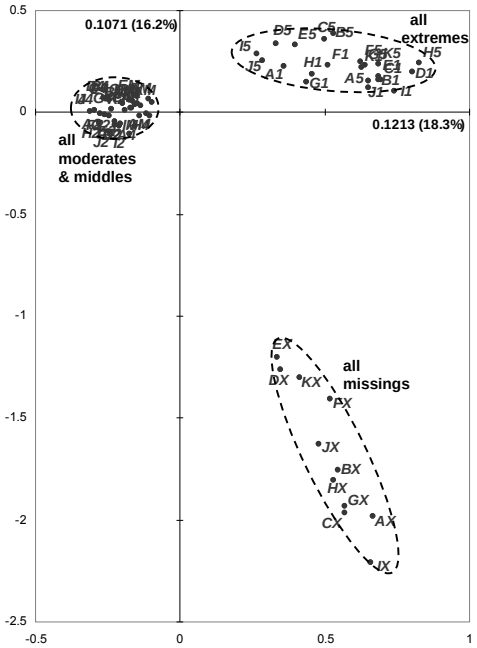
Original responses ($Q = 11$)

A1	*A2*	*AM*	*A4*	*A5*	*AX*	*B1*	*B2*	*BM*	...
1	0	0	0	0	0	0	0	1	...
0	0	0	0	0	1	0	1	0	...
			.						
			.						
			.						

Dummy variables ($J = 66$)

Dimensionality is 66 **less the** 11 **linear restrictions on the columns, i.e.** 55

MCA of all response categories



1 **strongly agree**
2 **agree**
M **middle**
4 **disagree**
5 **strongly disagree**
X **missing**

Total inertia = 0.6626

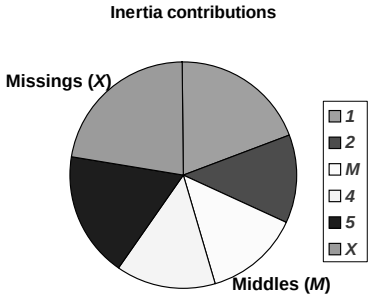
34.5% inertia explained

Adjusted value:

59.9% inertia explained

Decomposition of inertia across categories in MCA

Category	Inertia
1	0.1297
2	0.0827
M	0.0881
4	0.0928
5	0.1207
X	0.1485
total	0.6626

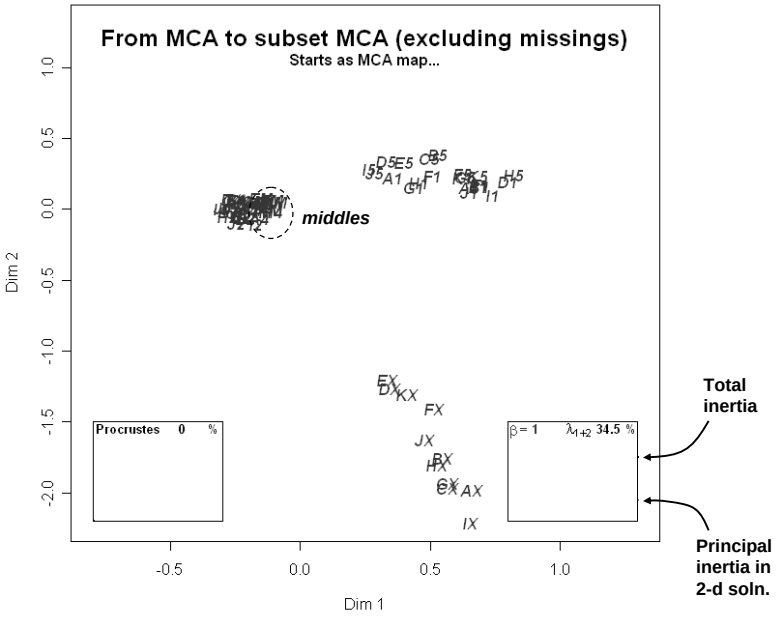


middles & missings account for 0.2366 of the total inertia of 0.6626, i.e. , 35.7% ; missings account for largest part of inertia

MCA to subset MCA

Data: **WOMEN WORKING, Burt matrix, *N*=33590**

Method: **Contribution of categories not in subset reduced by factor β, from 1 (CA) to limiting case of 0 (subset CA), always using original masses for centring and weighting)**

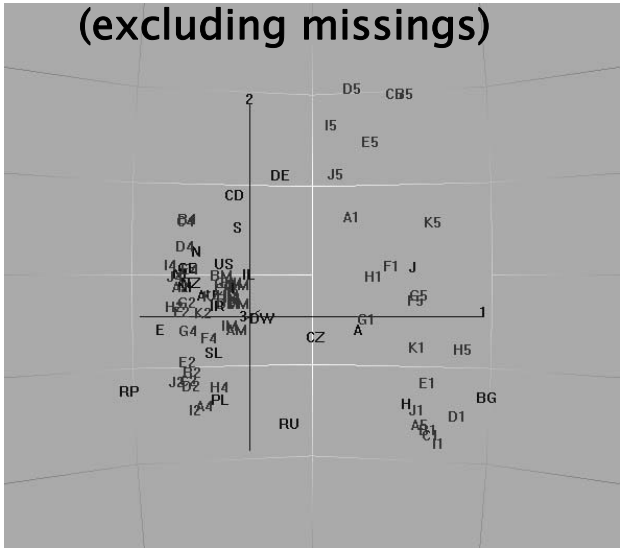


Rotating the subset MCA solution (excluding missings)

Data: **WOMEN WORKING, stacked frequencies, *N*=33590**

Using ca package in R.

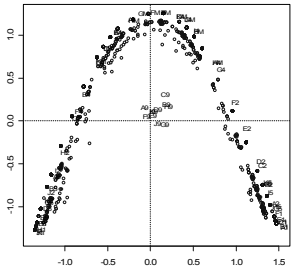
Also showing the country points in "symmetric" scaling so they are more spread out in the visualization for ease of interpretation



Reference: Nenadić, O. & Greenacre, M.J. (2007). Correspondence analysis in R, with two- and three-dimensional graphics: the **ca** package. *Journal of Statistical Software 20(3).* URL **`http://www.jstatsoft.org/v20/i03/.`**

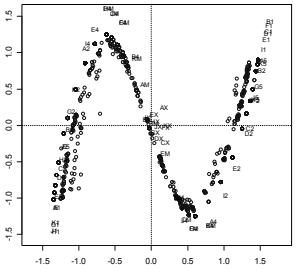
What does a "perfect unidimensional model" look like in an MCA?

Data that follow a perfect traditional-to-liberal scale were generated, reversing the scales of oppositely worded statements, and randomly adding 3% missing responses. Here are different MCA maps of the data:



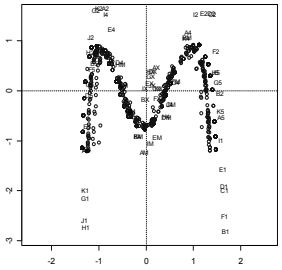
In first two dimensions:

Parabola:
the "horseshoe"/ "arch"/
"Guttman" effect



In dimensions 1 and 3:

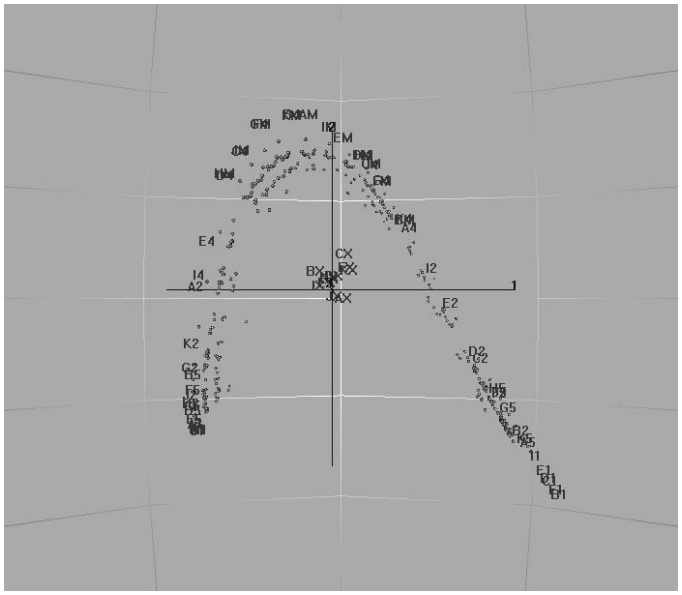
Cubic



In dimensions 1 and 4:

Quartic, etc....

What does a "perfect unidimensional model" look like in an MCA?



Looking by country: Spanish and West German data

- **Previous MCA maps based on all 33590 respondents from 24 countries.**
- **We have already seen that there are inter-country differences in their overall levels of middle and missing responses.**
- **Our aim is to investigate how respondents use the middle responses and if there any associations with demographic variables. To avoid inter-country differences in our results we concentrate (separately) on two countries: Spain (N = 2494) and West Germany (N = 2324) – as we shall see, they present contrasting results.**
- **We will also introduce the following demographic variables into our study:**

  ***Gender* (2 categories)**

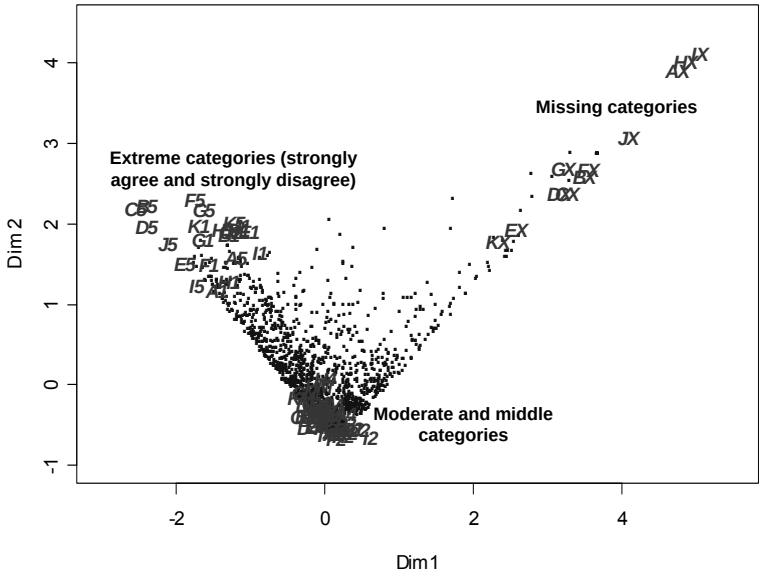
  ***Age* (6 categories)**

  ***Marital status* (5 categories)**

  ***Education* (7 categories)**

  (Education not available in Spanish 1994 sample)

MCA of Spanish "women working" data (N = 2494)

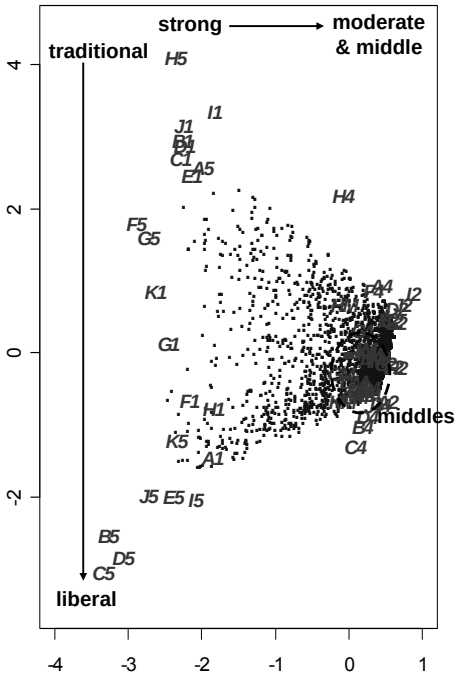


Subset MCA of Spanish "women working" data

(N = 2494)

missing responses excluded from subset

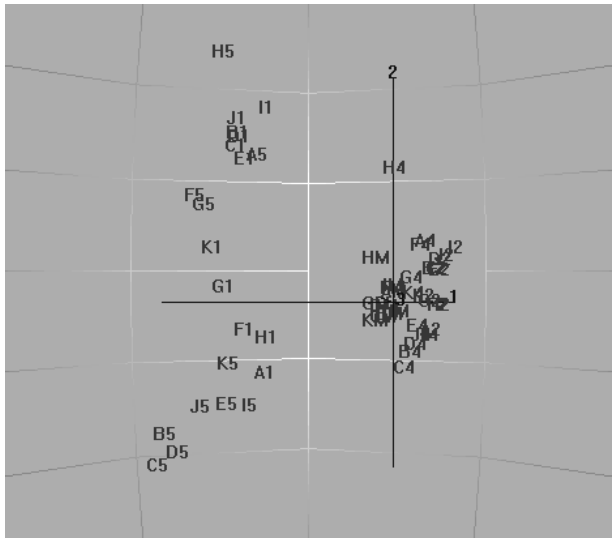
middles amongst the moderate responses in this two-dimensional view

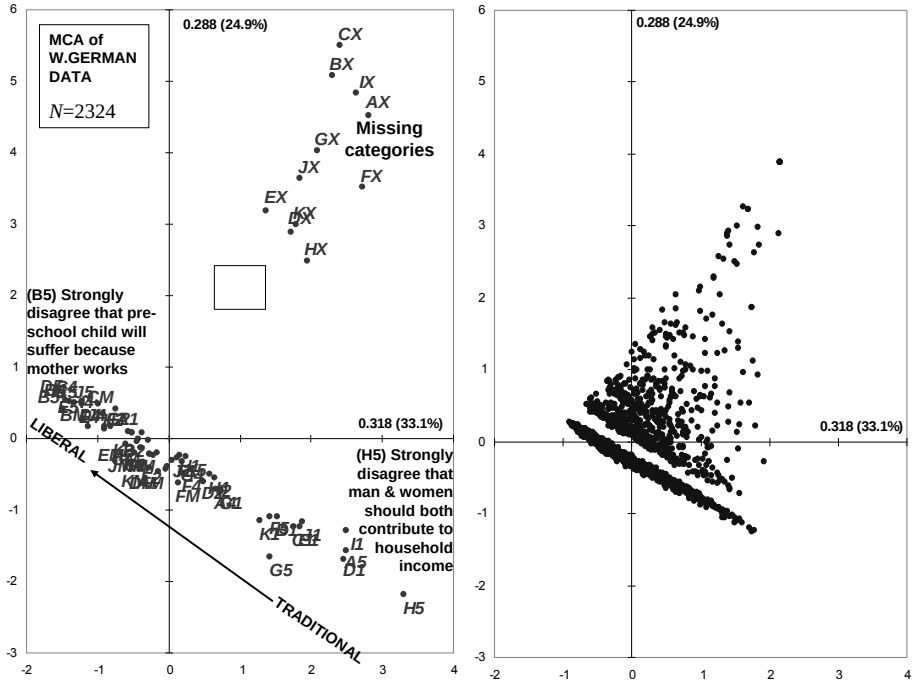


Rotating the subset MCA solution (subset excluding missings)

Data: **WOMEN WORKING, Spanish data, subset MCA, *N*=2494**

Method: **Rotation about 2nd axis, recorded using `plot3d` function in ca package for R**

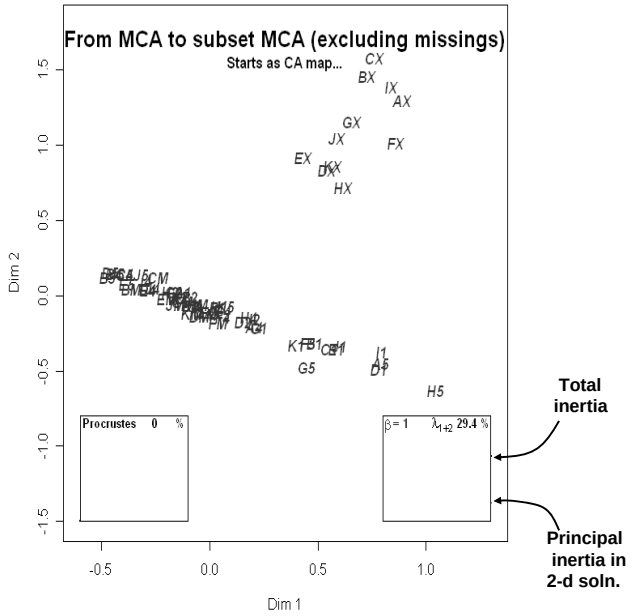




MCA to subset MCA

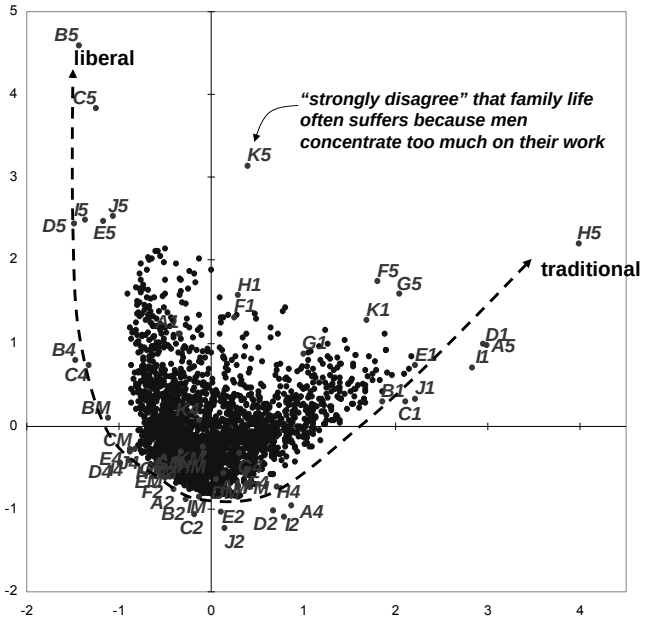
Data: **WOMEN WORKING, W.German sample, *N*=2324**

Method: **Contribution of categories not in subset reduced by factor β, from 1 (CA) to limiting case of 0 (subset CA), always using original masses for centring and weighting)**



Subset MCA of W German "women working" data

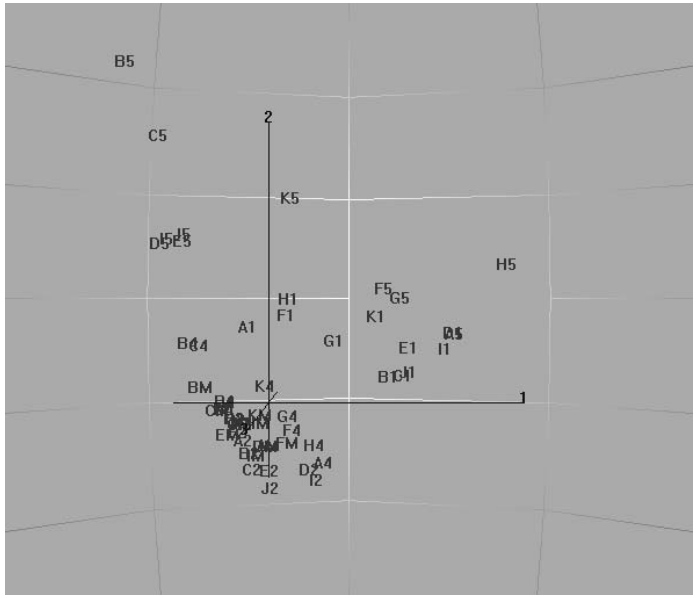
(N = 2324)
missing responses excluded from subset



Rotating the subset MCA solution (excluding missings)

Data: **WOMEN WORKING, W.German data, subset MCA, *N*=2324**

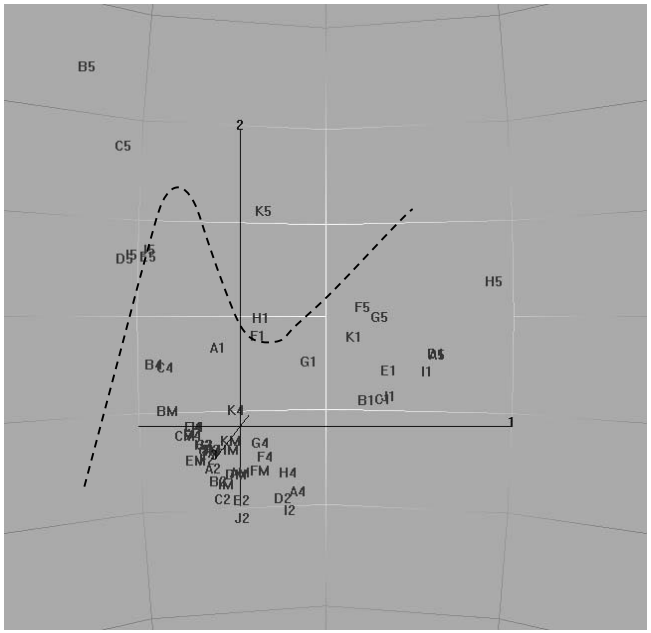
Method:
Rotation about 2nd axis, recorded using `rgl.snapshot` in rgl package and `plot3d.ca` in ca package for R



Rotating the subset MCA solution (excluding missings)

Data: **WOMEN WORKING, W.German data, subset MCA, *N*=2324**

Method:
Rotation about 1st axis, recorded using `rgl.snapshot` in rgl package and `plot3d.ca` in ca package for R



Response sets in Spanish sample

Looking more closely at the individual Spanish data, we discover the following response sets (*figures for W.Germany for comparison*):

		(*WG*)
"strongly agree" to all questions	– 4 respondents	(***2***)
"agree" to all questions	**– 20 respondents**	(***6***)
"neither/nor" to all questions	**– 6 respondents**	(***0***)
"disagree" to all questions	**– 2 respondents**	(***0***)
"strongly disagree" to all questions	**– 0**	(***0***)
missing values for all questions	**– 18 respondents**	(***6***)
	50 out of the 2494 respondents, i.e. 2%	(½%)

The "middle" and "missing" response sets accentuate the association within these categories

The "strongly agree" and "agree" response sets (categories 1 and 2) to questions which have reverse orientations will tend to bring the opposite poles closer than they would otherwise have been.

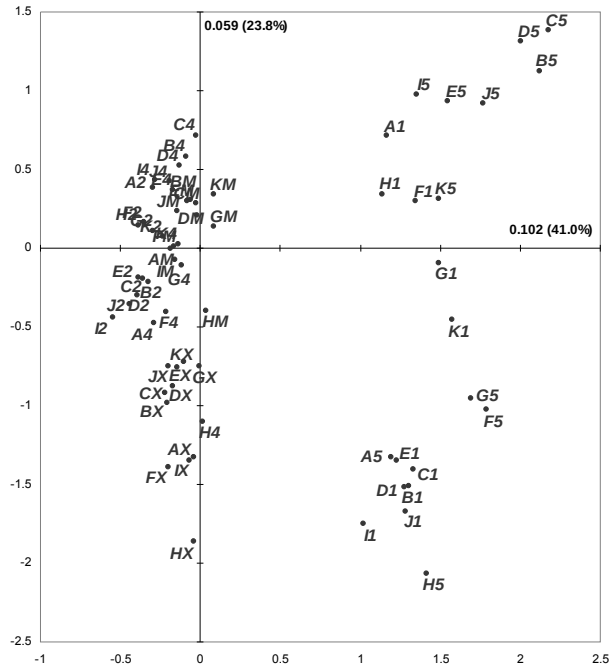
We now remove all these response sets, all the features previously seen are still there, just their ordering on the principal axes changes: e.g., the group of "missings" is now on the 3rd dimension and the "middles" on the 5th.

MCA of Spanish data without response sets

Data: **WOMEN WORKING, Spanish sample, *N*=2444**

(without 50 response sets)

Dimensions 1 and 2

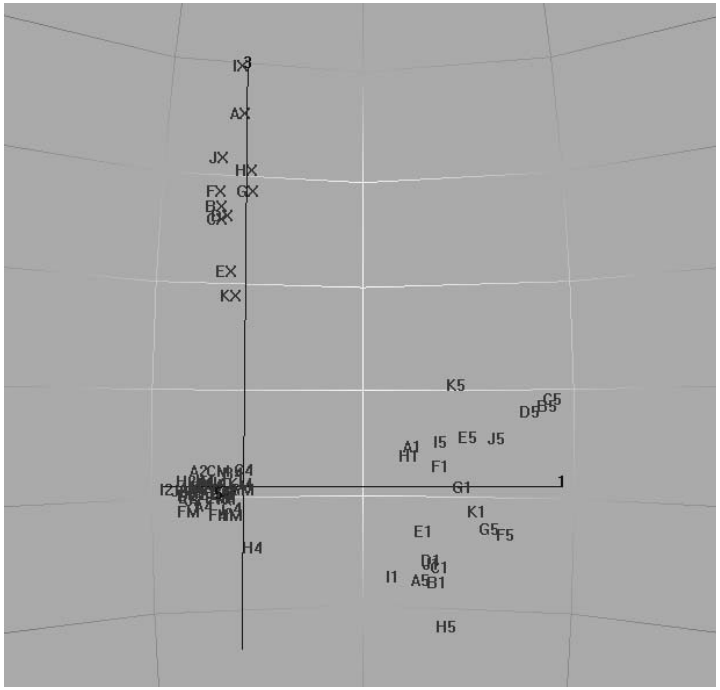


MCA of Spanish data without response sets

Data: **WOMEN WORKING, Spanish sample, *N*=2444**

(without 50 response sets)

Dimensions 1, 3 and 5

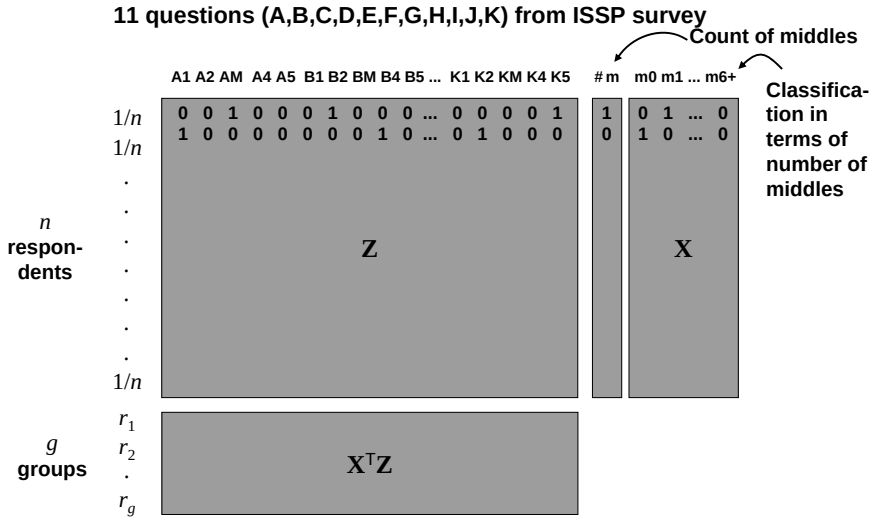


Focusing the display on the number of middle responses

11 questions (A,B,C,D,E,F,G,H,I,J,K) from ISSP survey

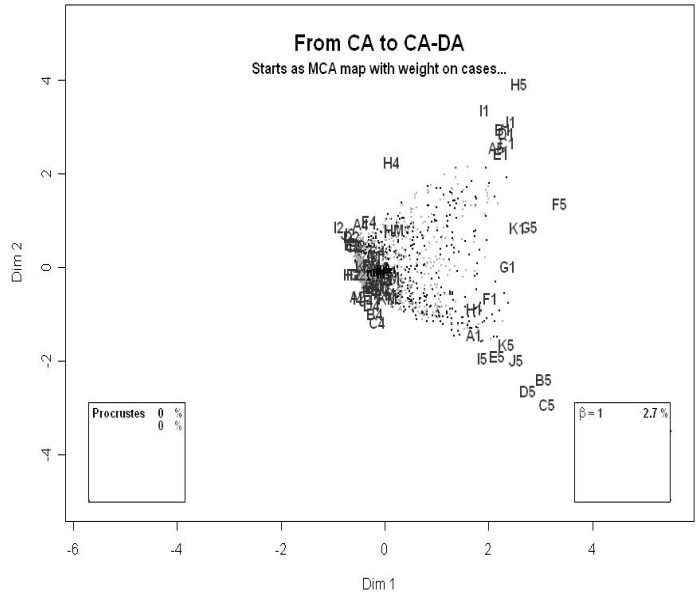
Count of middles

Classification in terms of number of middles

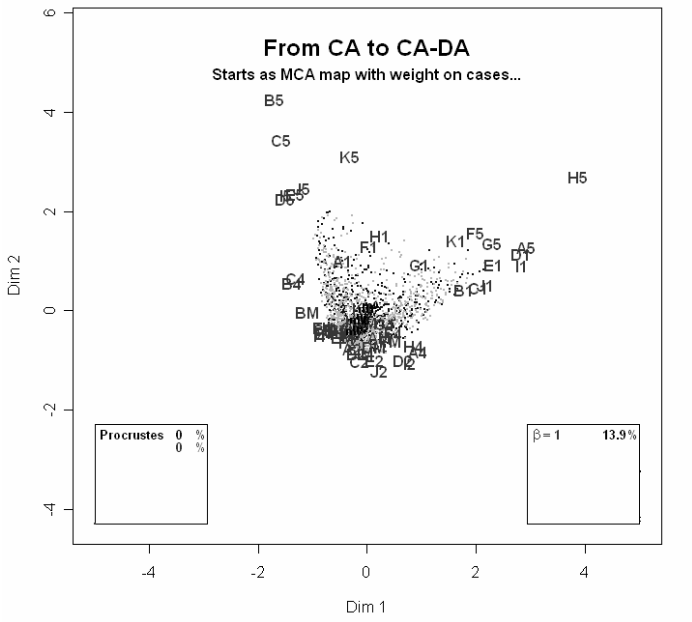


Pass the weight smoothly from the respondents to the group centroids:

MCA → MCA-Discriminant analysis (middle groups): Spain



MCA → MCA–Discriminant analysis (middle groups): W. Germany



Methods

We use correspondence analysis (CA) and several of its variants to visualize and interpret data at a country level as well as respondent and demographic-subgroup level.

1. **Simple CA provides maps of aggregated or count data, for example proportions of question responses for the set of countries.**
2. **Subset correspondence analysis permits focusing on a particular set of response categories so that we can eliminate missing responses, for example, or restrict our attention to specific categories such as the middle responses.**
3. **Multiple correspondence analysis (MCA) and its refinement joint correspondence analysis (JCA) provide maps of individual-level data, concentrating on the two-way relationships between the questions. In this respect MCA functions like an exploratory factor analysis for categorical data (on nominal scales). Usually we are not interested in individual case points but in mean points for the demographic and other external variables (e.g., age, level of interest…)**
4. **Canonical correspondence analysis (CCA) focuses on (or partials out) external variables – thus we can explore variation in the responses that is focused on interest, for example, or eliminate aquiescence effects.**

Partialling out acquiescence effects

O'Muircheartaigh, Krosnick & Helic (2000)

'We estimated a model in which all items were allowed to load on the same latent factor representing attitude toward science, plus a second latent factor intended to represent acquiescence. All items were constrained to load equally on this latter factor, an assumption required to identify the model. This is reasonable, because acquiescence is defined as a tendency to agree with any item regardless of its content, so it should account for the same amount of variance in responses to all the items. The acquiescence factor was constrained to be uncorrelated with the factor representing attitudes toward science, another assumption required in order to identify the model.'

Identifying or partialling out "acquiescence effects"

11 questions (A,B,C,D,E,F,G,H,I,J,K) from ISSP Family and Changing Role survey II (1994).

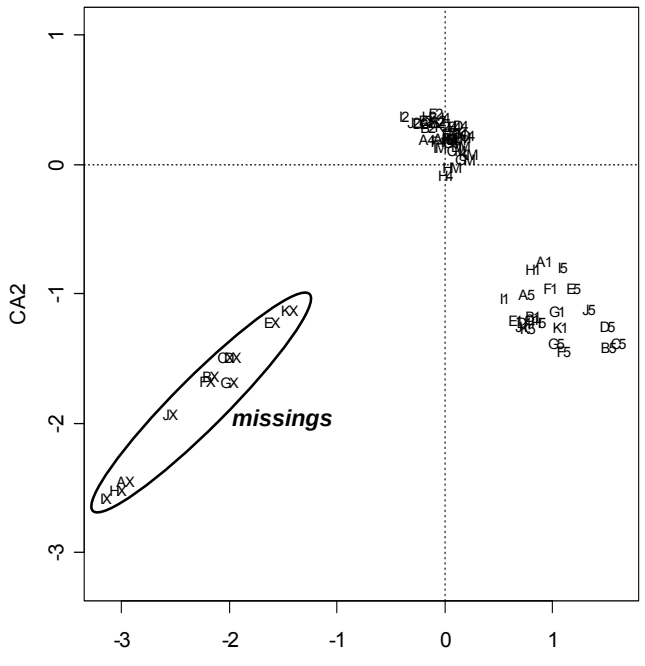
counts of number of responses in each category

A1	A2	AM	A4	A5	AX	B1	B2	BM	B4	B5	BX	…	K1	K2	KM	K4	K5	KX	1	2	M	4	5	X
0	0	1	0	0	0	0	1	0	0	0	0	…	0	0	0	0	1	0	2	3	1	3	1	1
1	0	0	0	0	0	0	0	0	0	1	0	…	0	0	0	0	0	1	3	1	0	2	2	3
Z																			X					

The counts in the matrix X are variables which quantify tendencies to use the same response category. CCA can restrict the MCA solution to be linearly related to any subset of these external variables. This is a projection in the MCA space. We can also partial out acquiescence effects by looking in the space orthogonal to the restricted space (this is an alternative strategy to subset CA).

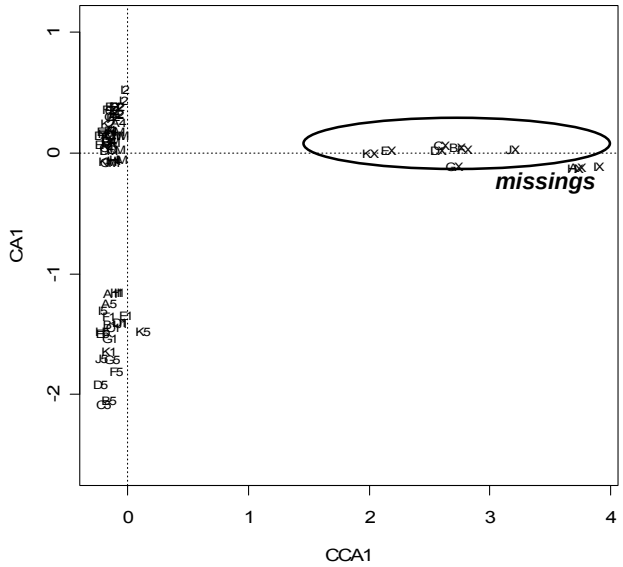
MCA

Data:
WOMEN WORKING, Spanish sample, *N*=2494



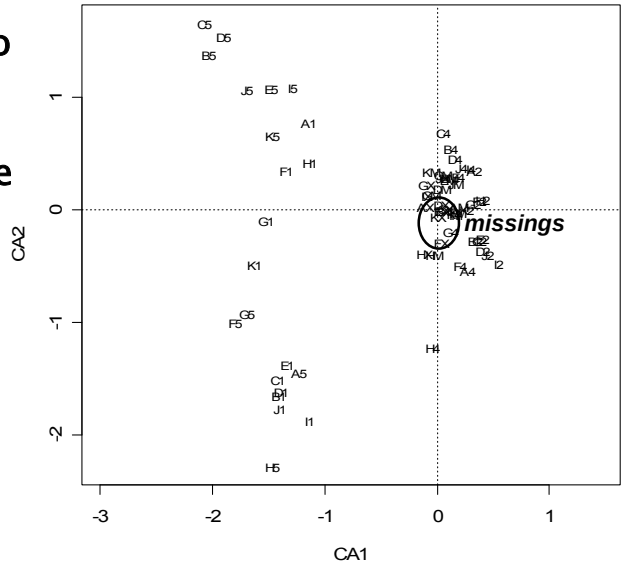
CCA counts of missings as external variable which restricts the solution

Data:
WOMEN WORKING, Spanish sample, *N*=2494



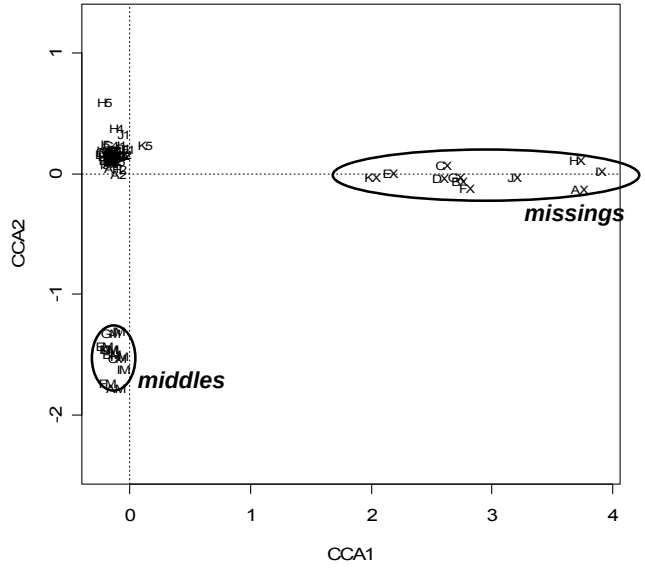
Partial CCA solution orthogonal to one that restricts the solution to be linearly related to missing counts

Data:
WOMEN WORKING, Spanish sample, *N*=2494



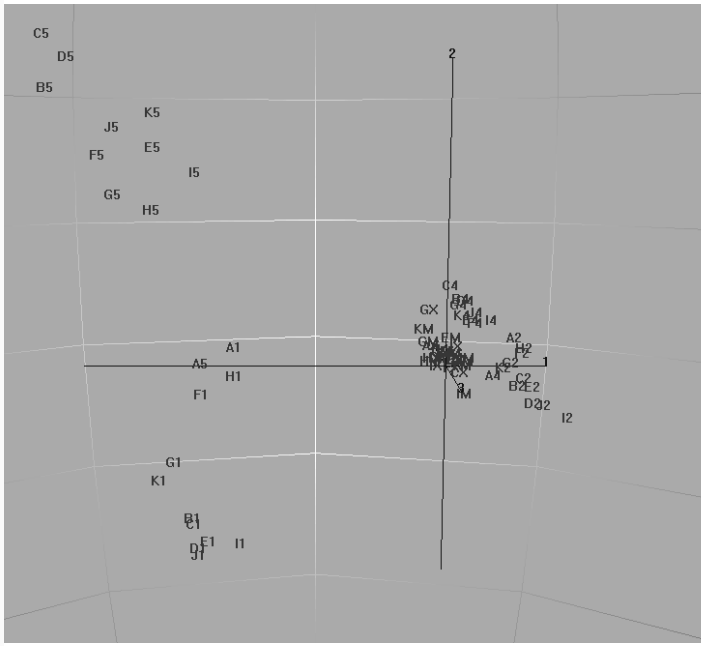
CCA counts of missings and middles as external variable which restricts the solution

Data:
WOMEN WORKING, Spanish sample, *N*=2494



CCA solution partialling out the middles and missings and restricted to the extremes and moderates

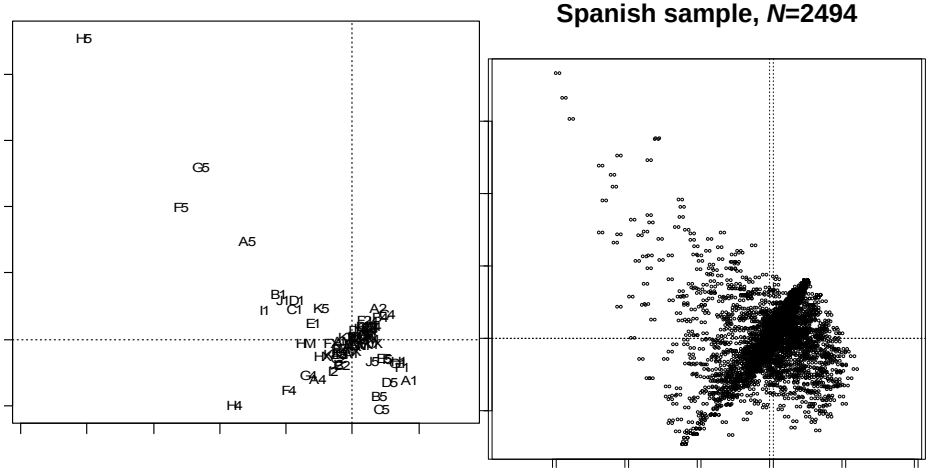
Data: **WOMEN WORKING, Spanish sample, *N*=2494**



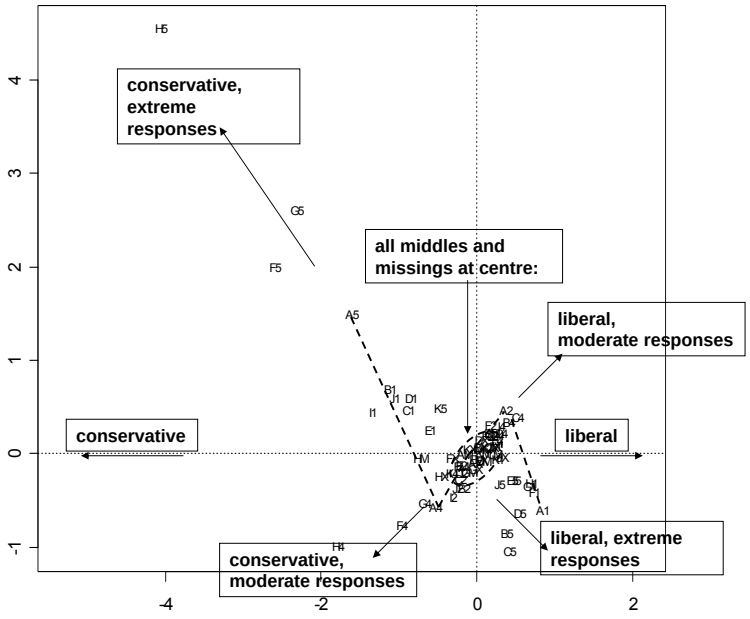
Partial CCA

solution orthogonal to one that restricts the solution to be linearly related to all acquiescence effects

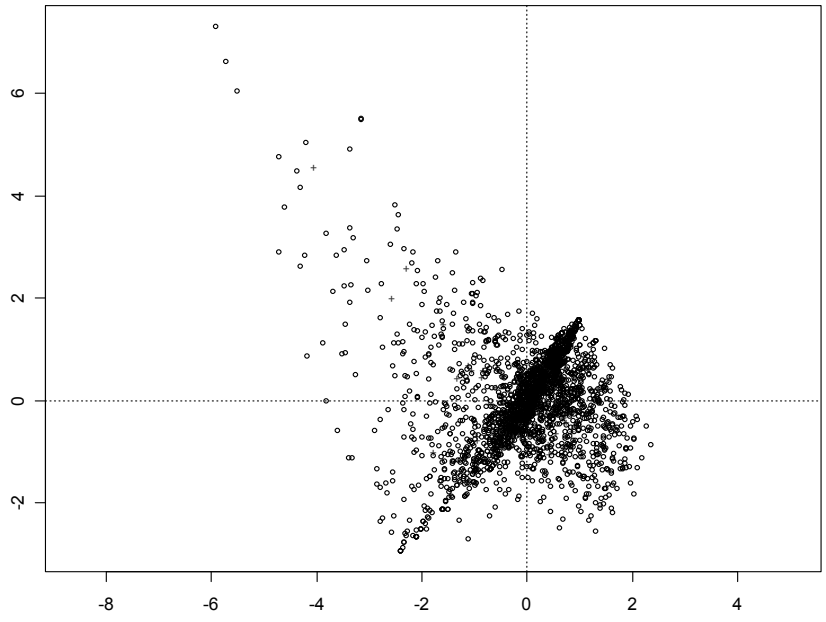
Data: **WOMEN WORKING, Spanish sample, *N*=2494**



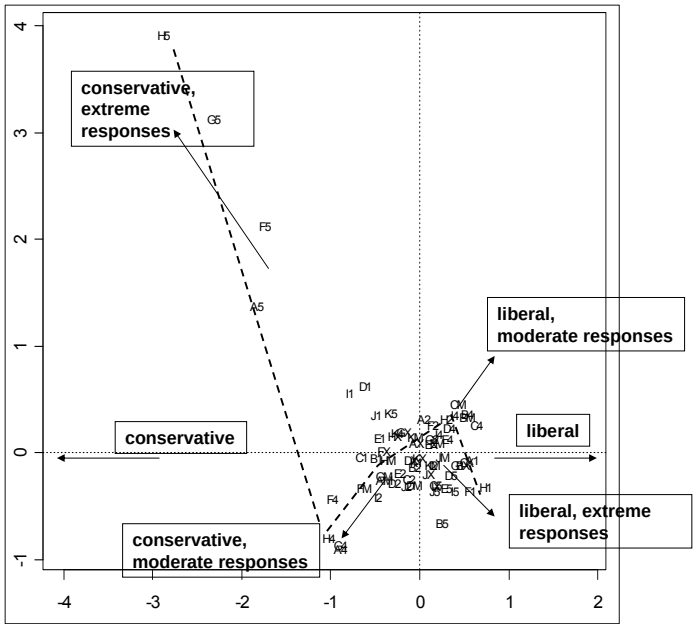
CCA of Spanish "women working" data:categories



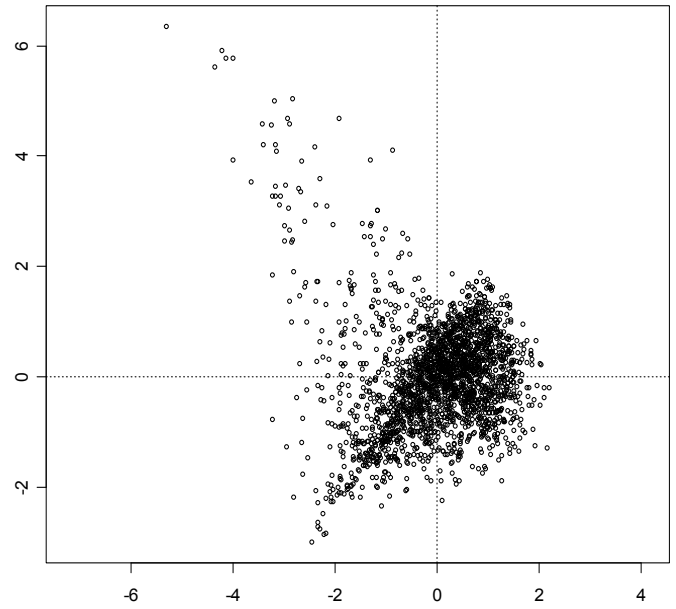
CCA of Spanish "women working":cases



CCA of W.German "women working":categories



CCA of W.German "women working":cases

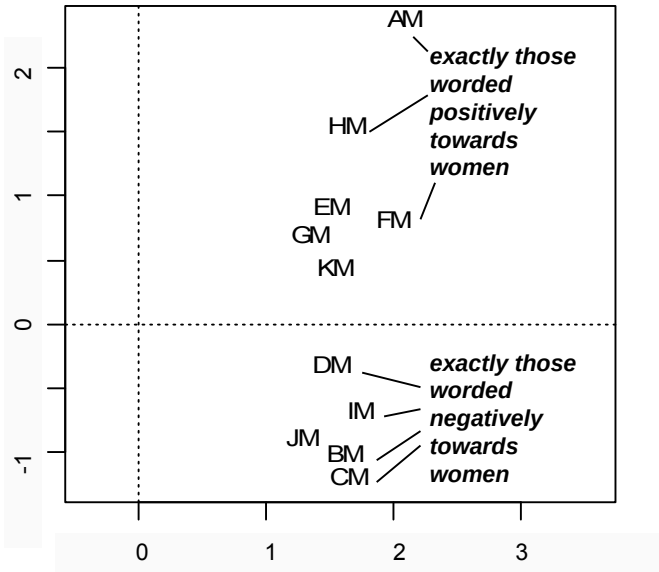


Demographic categories

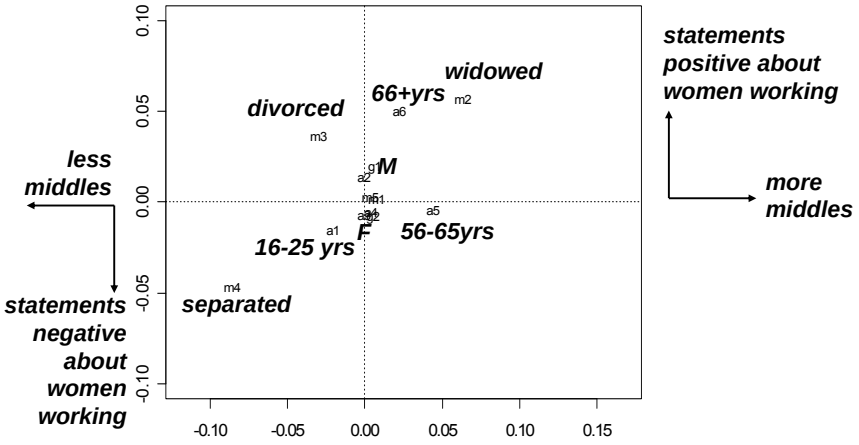
- *Gender : g1 (male), g2 (female)*
- *Age 6 groups: a1 (up to 25), a2 (26-35), a3 (36-45)*
  a4 (46-55), a5 (56-65), a6 (66 and over)
- *Marital status 5 groups: m1 (married), m2 (widowed), m3 (divorced),*
  m4 (separated), m5 (single)
- *Education 7 groups: e1 (none), e2 (incomplete primary),*
  e3 (primary), e4 (incomplete secondary),
  e5 (secondary), e6 (incomplete tertiary),
  e7 (tertiary)

(Education not available for Spanish sample in 1994)

Subset MCA of middle categories – "women working", Spain



Demographic averages on subset MCA map of middle responses – "women working", Spain



Checking the MCA results on the data: calculating averages per respondent