With JMP® 12, Correspondence Analysis Goes to Multiple

Yves Gueniffey, PhD, Assistant Professor, École des Mines de Nancy, France

ABSTRACT

Correspondence analysis, single and multiple (CA and MCA), is a well-known method among statisticians trained in the "French school" of data analysis. Developed in the 1960s by Jean-Paul Benzécri and his team, it took some time before being accepted by our Anglo-Saxon colleagues; for example, PROC CORRESP appeared only with SAS® 6, released in the 1990s. With the current version of JMP (11.2.0), we have only single correspondence analysis, hidden in the options available in the ***Contingency Analysis*** platform. **JMP®** 12 offers a new platform that enables us to perform correspondence analysis, single and multiple, with (almost) all options and aids in the interpretation of the results that the French statisticians are accustomed to. I'm going to have a look at this new platform in my presentation.

.

INTRODUCTION

This article offers no theoretical presentation of CA or MCA, nor does it explain the historical origins. For historical aspects, one can, if one reads French, consult articles that Jean-Paul Benzécri himself devoted to the *Histoire et préhistoire de l'analyse des données*, in the first 1976 and 1977 issues of the *Cahiers de l'Analyse des données* [6], online at http://www.numdam.org/numdam-bin/feuilleter?j=CAD, and in English, the article by Michel Tenenhaus and Forrest W. Young appeared in Psychometrika in 1985 [4].

For these historical aspects and a very accessible theoretical presentation, we will choose the works of Michael J. Greenacre [2, 3]. According to Michael J. Greenacre,

"Correspondence analysis (CA) has long and interesting history of being defined, rediscovered, and redefined over decades. Statisticians who have contributed to its origins in the first half of the 20th century have been H.O. Hartley (Hirschfeld), R.A. Fisher, and Louis Guttman, among others. In the second half of the century, this method sprung up fairly independently in Japan, France, and Holland, respectively, guided by Chikio Hayashi (who saw it as a method of categorical data scaling), Jean-Paul Benzécri (who saw it as a method of data visualization), and Jan de Leeuw (who saw it as integrating categorical data into classical interval-level multivariate analysis)." [3]

and :

« Correspondence analysis (CA) is a method of data visualization that applies to cross-tabular data such as counts, compositions, or any ratio-scale data where relative values are of interest. [...]

Correspondence analysis (CA) offers the remarkable feature of jointly representing individuals and variables. As a result of such analyses, not only does one gain insight in the relationship amongst individuals and amongst variables, but one can also find an indication of which variables matter in the description of which individuals. It is therefore natural to develop clustering algorithms based on the coordinates of a CA."

Our purpose will be limited to illustrate the new JMP® 12 ***Multiple Correspondence Analysis*** (MCA) platform.

We will show in this paper that the new MCA platform enables:

- to properly perform a CA on a contingency table with the possibility to use additional illustrative rows and columns;
- to perform a MCA on a table crossing individuals and more than two active categorical variables; also with the possibility of considering individuals and/or illustrative categorical variables;
- to support both continuous and categorical variables in the same analysis;
- and finally to perform a clustering of individuals that takes into account two types of variables by simply performing clustering on the factorial axis coordinates of the previous MCA (which are themselves continuous).

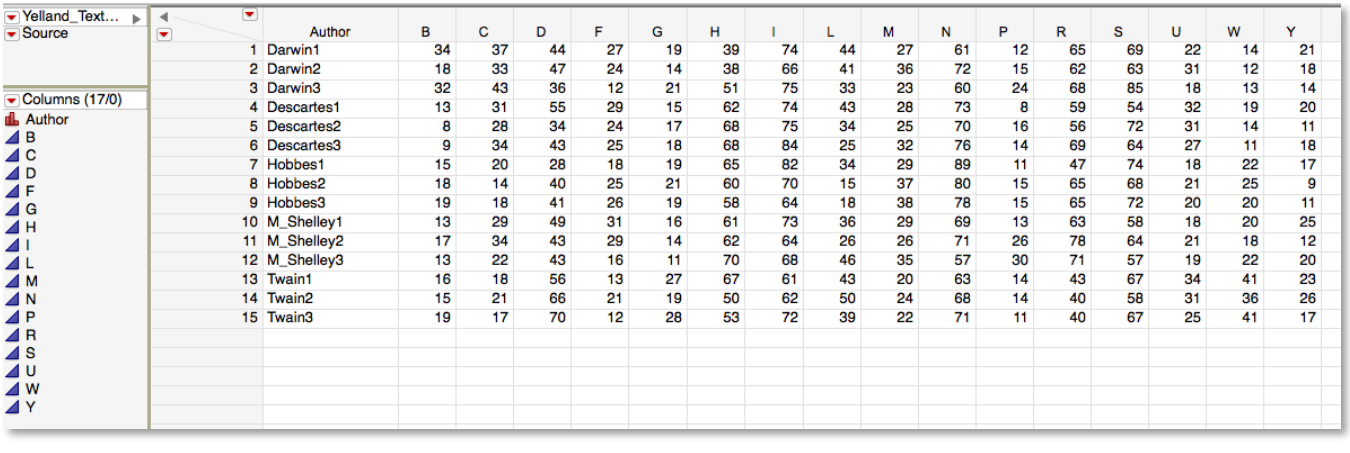
CORRESPONDENCE ANALYSIS : A NEW TOOL FOR SHERLOCK HOLMES (THE USE OF ADDITIONAL ILLUSTRATIVE ROWS AND COLUMNS)

CA enables us to interpret the projections of the points on factor planes (factorial maps), specifically, to interpret the proximity of the profile-line points between them, and the proximity of the profile-column points between them, but also the proximity of the profile-line points and profile-column points between them, insofar as the quality of their representations allows (**Squared Cosines**).

To illustrate these properties, and make a small nod to the early work of Jean-Paul Benzécri on textual statistics, we borrow this simplified example from Phillip M. Yelland [5].
Given the corpus of the five classics Charles Darwin, Descartes, Thomas Hobbes, Mary Shelley and Mark Twain: Ph. Yelland randomly sampled three distinct samples of 1000 characters from each of these authors and noted occurrences of the letters B, C, D, F, G, H, I, L, M, N, P, R, S, U, W, and Y. We then get the contingency table (15x16) as follows:

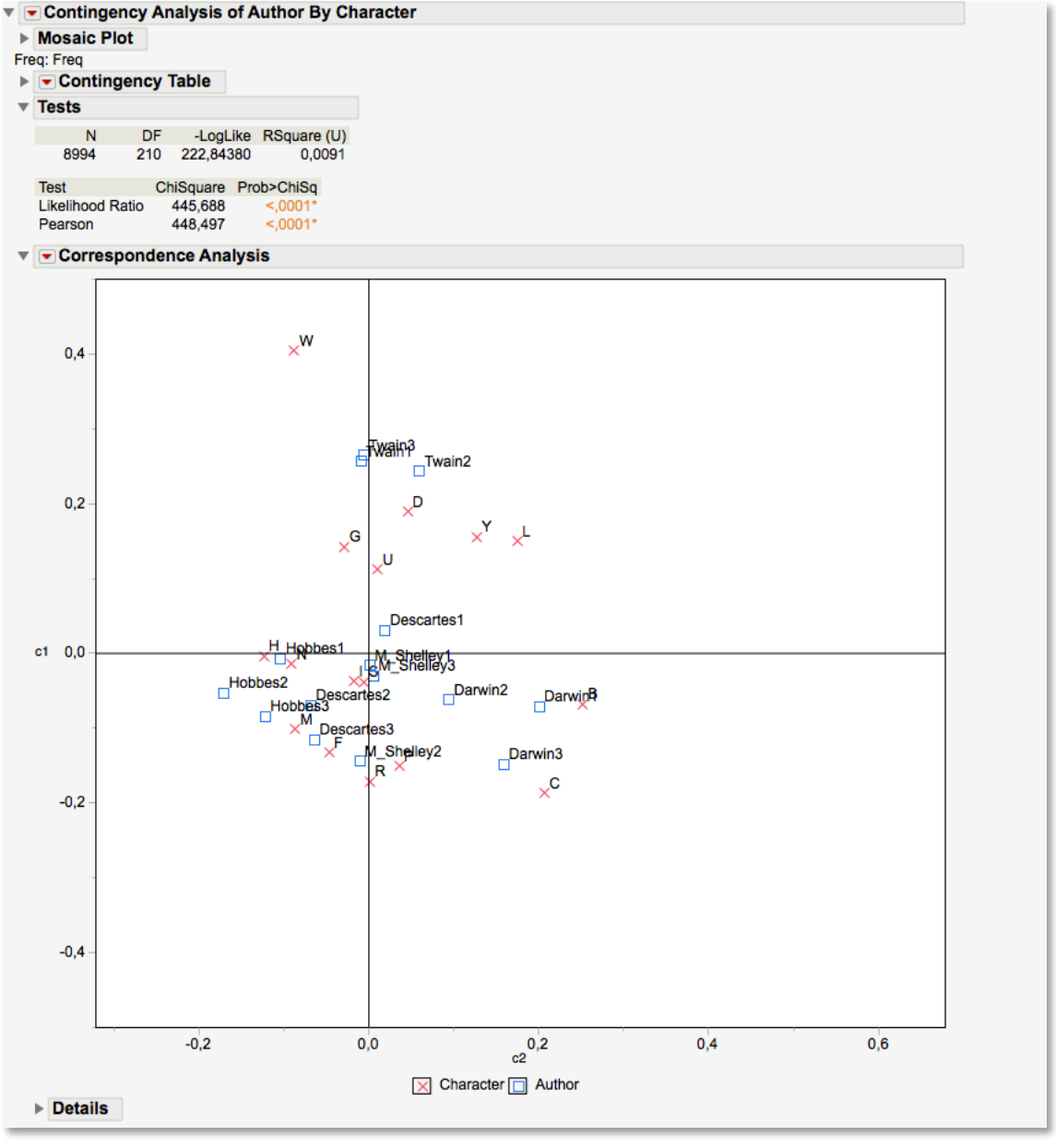
Author	B	C	D	F	G	H	I	L	M	N	P	R	S	U	W	Y	Total
Darwin1	34	37	44	27	19	39	74	44	27	61	12	65	69	22	14	21	609
Darwin2	18	33	47	24	14	38	66	41	36	72	15	62	63	31	12	18	590
Darwin3	32	43	36	12	21	51	75	33	23	60	24	68	85	18	13	14	608
Descartes1	13	31	55	29	15	62	74	43	28	73	8	59	54	32	19	20	615
Descartes2	8	28	34	24	17	68	75	34	25	70	16	56	72	31	14	11	583
Descartes3	9	34	43	25	18	68	84	25	32	76	14	69	64	27	11	18	617
Hobbes1	15	20	28	18	19	65	82	34	29	89	11	47	74	18	22	17	588
Hobbes2	18	14	40	25	21	60	70	15	37	80	15	65	68	21	25	9	583
Hobbes3	19	18	41	26	19	58	64	18	38	78	15	65	72	20	20	11	582
M_Shelley1	13	29	49	31	16	61	73	36	29	69	13	63	58	18	20	25	603
M_Shelley2	17	34	43	29	14	62	64	26	26	71	26	78	64	21	18	12	605
M_Shelley3	13	22	43	16	11	70	68	46	35	57	30	71	57	19	22	20	600
Twain1	16	18	56	13	27	67	61	43	20	63	14	43	67	34	41	23	606
Twain2	15	21	66	21	19	50	62	50	24	68	14	40	58	31	36	26	601
Twain3	19	17	70	12	28	53	72	39	22	71	11	40	67	25	41	17	604
Total	259	399	695	332	278	872	1064	527	431	1058	238	891	992	368	328	262	8994

and the JMP table:



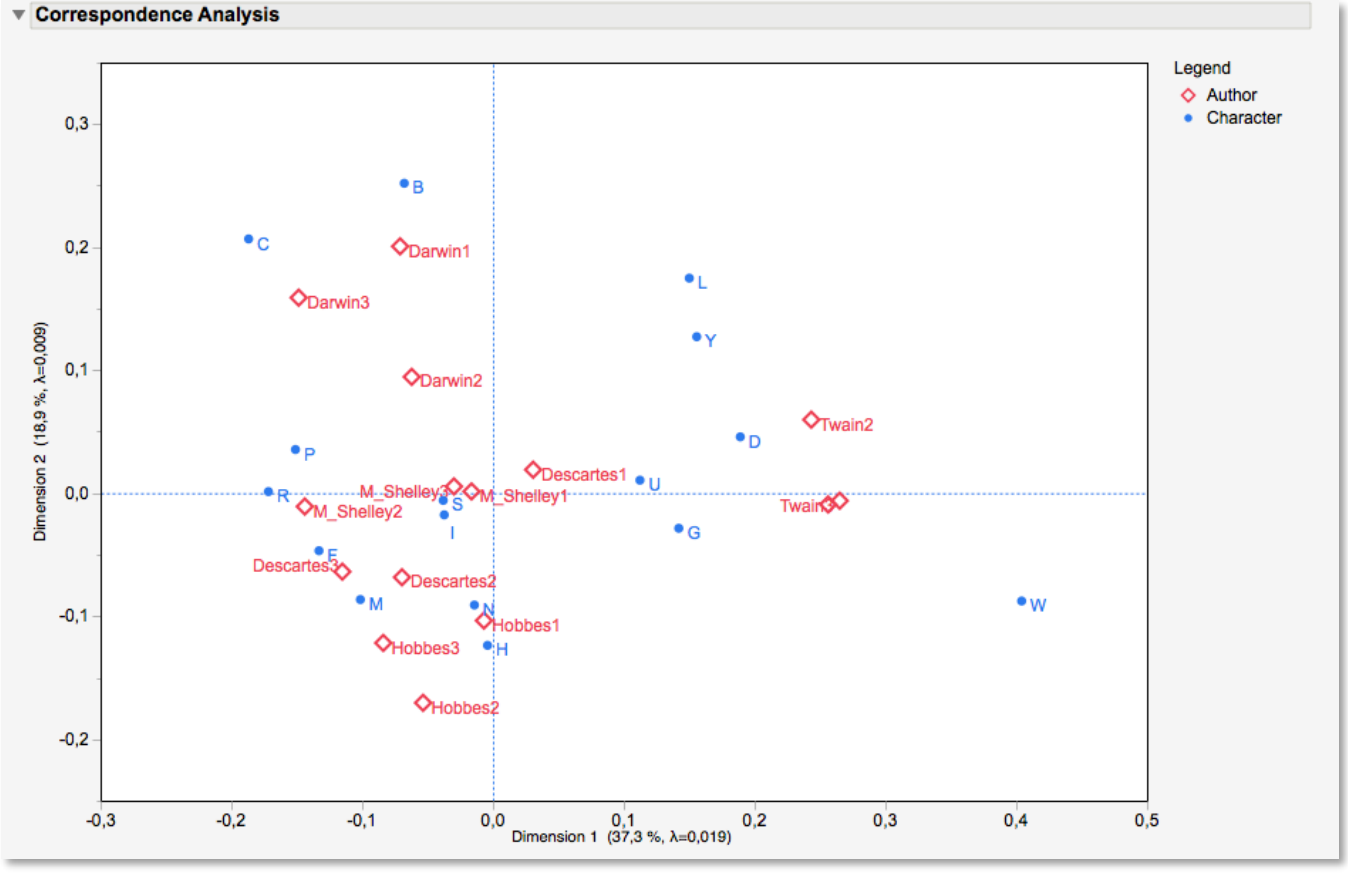
	Author	B	C	D	F	G	H	I	L	M	N	P	R	S	U	W	Y
1	Darwin1	34	37	44	27	19	39	74	44	27	61	12	65	69	22	14	21
2	Darwin2	18	33	47	24	14	38	66	41	36	72	15	62	63	31	12	18
3	Darwin3	32	43	36	12	21	51	75	33	23	60	24	68	85	18	13	14
4	Descartes1	13	31	55	29	15	62	74	43	28	73	8	59	54	32	19	20
5	Descartes2	8	28	34	24	17	68	75	34	25	70	16	56	72	31	14	11
6	Descartes3	9	34	43	25	18	68	84	25	32	76	14	69	64	27	11	18
7	Hobbes1	15	20	28	18	19	65	82	34	29	89	11	47	74	18	22	17
8	Hobbes2	18	14	40	25	21	60	70	15	37	80	15	65	68	21	25	9
9	Hobbes3	19	18	41	26	19	58	64	18	38	78	15	65	72	20	20	11
10	M_Shelley1	13	29	49	31	16	61	73	36	29	69	13	63	58	18	20	25
11	M_Shelley2	17	34	43	29	14	62	64	26	26	71	26	78	64	21	18	12
12	M_Shelley3	13	22	43	16	11	70	68	46	35	57	30	71	57	19	22	20
13	Twain1	16	18	56	13	27	67	61	43	20	63	14	43	67	34	41	23
14	Twain2	15	21	66	21	19	50	62	50	24	68	14	40	58	31	36	26
15	Twain3	19	17	70	12	28	53	72	39	22	71	11	40	67	25	41	17

After transforming the data (**Tables→*Stack***), we can perform a CA using the ***Fit Y by X*** platform to get this mapping:



It was the only way to achieve a CA with JMP11; it is still available in the new version JMP12.

But it is much more satisfying to use the new platform ***Multiple Correspondence Analysis***, available from the menu ***Analyze→ Consumer Research***. Indeed, a CA is a special case of MCA, and according to the principle "who can do more can do less", one can also perform this analysis on this new platform.



Using the principles of interpretation outlined above, we can see that on the first factorial plane which combines 56,17% of the inertia of the cloud of points:

- Works by the same author are projected in the same area;
- Twain's works are distinguished from works by other authors;
- Darwin's works are distinguished well from those by Hobbes;
- It seems more difficult to distinguish between the works by Descartes and those by Mary Shelley.

One may further clarify these remarks with the traditional interpretation aids:

- Lines contributions (authors) and columns contributions (letters) of the contingency table to the interpretation of each axis;
- The quality of representation of the lines and columns projections points on the factor axes.

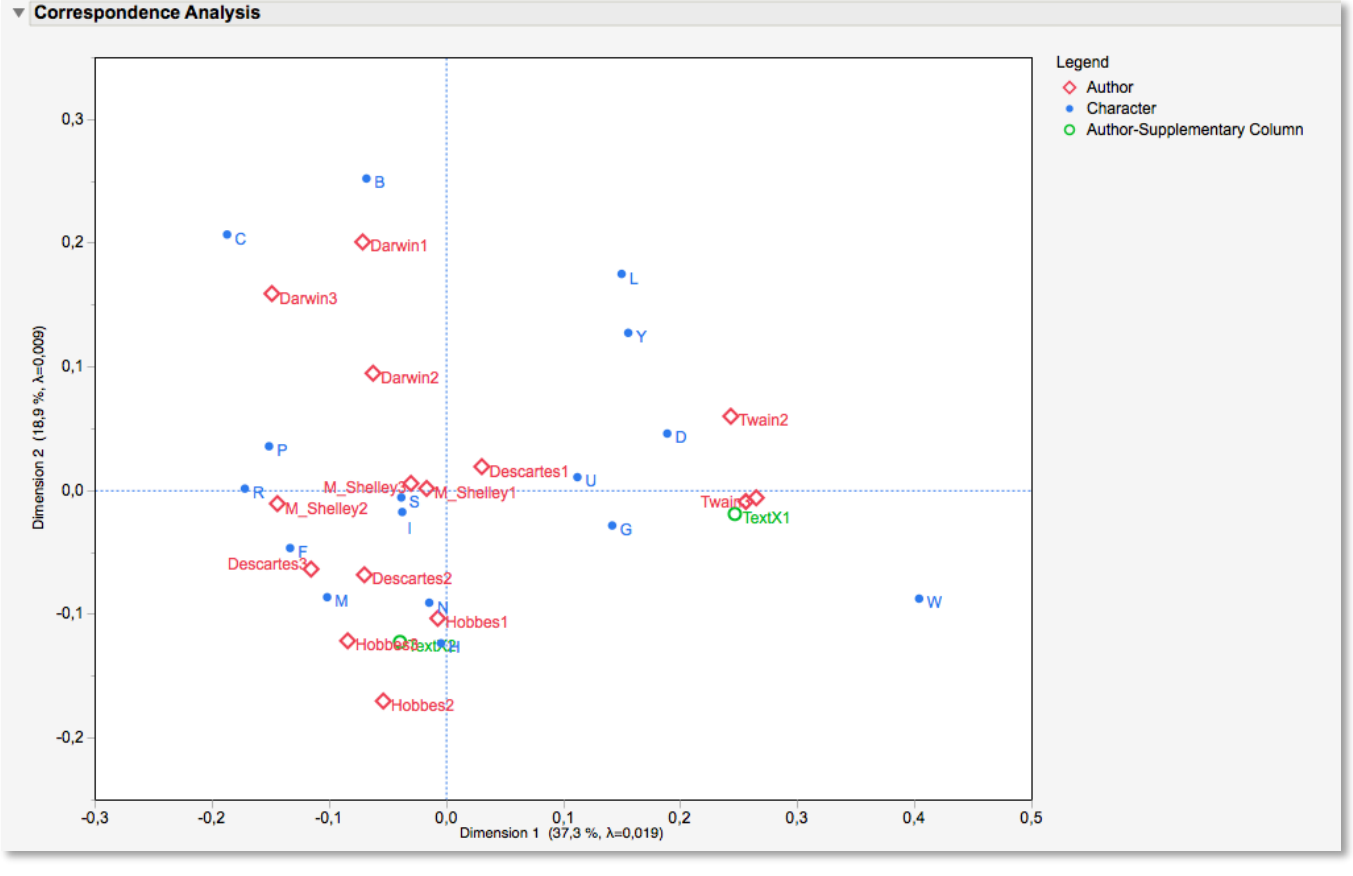
Better yet, we can interpret the position of the additional lines points. Phillip M. Yelland has also collected two additional samples of 1,000 letters in the works by two of the authors mentionned above, but without noting their names. Can we identify these authors from the single frequency letters used earlier in the series?
The contingency table is now supplemented by two new lines:

Yelland_Text…
Source

Columns (17/0)
Author
B
C
D
F
G
H
I
L
M
N
P
R
S
U
W
Y

	Author	B	C	D	F	G	H	I	L	M	N	P	R	S	U	W	Y
1	Darwin1	34	37	44	27	19	39	74	44	27	61	12	65	69	22	14	21
2	Darwin2	18	33	47	24	14	38	66	41	36	72	15	62	63	31	12	18
3	Darwin3	32	43	36	12	21	51	75	33	23	60	24	68	85	18	13	14
4	Descartes1	13	31	55	29	15	62	74	43	28	73	8	59	54	32	19	20
5	Descartes2	8	28	34	24	17	68	75	34	25	70	16	56	72	31	14	11
6	Descartes3	9	34	43	25	18	68	84	25	32	76	14	69	64	27	11	18
7	Hobbes1	15	20	28	18	19	65	82	34	29	89	11	47	74	18	22	17
8	Hobbes2	18	14	40	25	21	60	70	15	37	80	15	65	68	21	25	9
9	Hobbes3	19	18	41	26	19	58	64	18	38	78	15	65	72	20	20	11
10	M_Shelley1	13	29	49	31	16	61	73	36	29	69	13	63	58	18	20	25
11	M_Shelley2	17	34	43	29	14	62	64	26	26	71	26	78	64	21	18	12
12	M_Shelley3	13	22	43	16	11	70	68	46	35	57	30	71	57	19	22	20
13	Twain1	16	18	56	13	27	67	61	43	20	63	14	43	67	34	41	23
14	Twain2	15	21	66	21	19	50	62	50	24	68	14	40	58	31	36	26
15	Twain3	19	17	70	12	28	53	72	39	22	71	11	40	67	25	41	17
16	TextX1	24	26	80	17	32	91	86	54	32	91	19	58	93	50	58	30
17	TextX2	19	33	35	22	40	96	116	39	40	129	17	72	104	30	25	24

The correspondence analysis performed with the new platform enabling the introduction of additional lines then gives:



Of course we see that the analysis is not different from the one previously carried out: the additional points are projected on factorial planes calculated from the only texts identified thanks to their authors. But this is where lies the value of the transaction. Now we can reasonably identify unknown authors: Text1 is doubtlessly by Twain and Text2 probably by Hobbes.

THE TITANIC CASE: WOMEN AND CHILDREN FIRST? (MCA AT WORK)

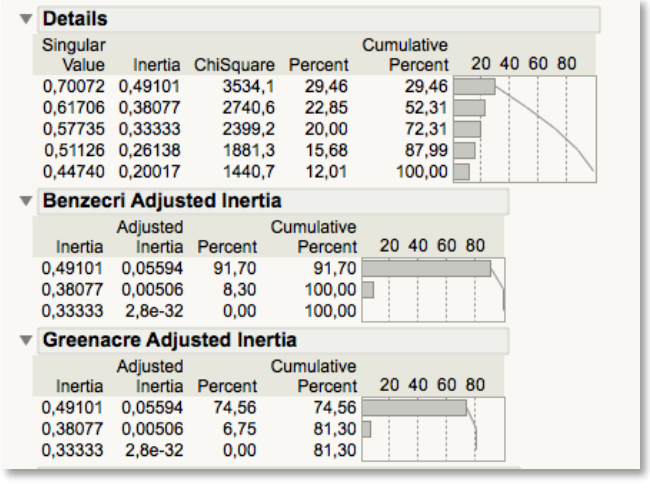
MCA borrows from both the PCA because it addresses an array crossing individuals and variables, but also from CA because these variables are categorical ones (as opposed to PCA regarding continuous variables, at least for those who are "active", i.e. those that determine the factor planes).

Everyone knows the story of the *Titanic*, at least since James Cameron's movie ... April 15, 1912, the wonder of *The White Star Line* sank with 2,201 people on board; it was her inaugural cruise.

We'll see to what extend the popular saying "Women and children first" was respected, and that even with a very interesting feature of the new JMP MCA platform, which enables to introduce additional “illustrative” variables.

The JMP **Titanic** table provided in the Help examples offer the first 3 variables we take as active variables: Class (first, second, third, crew) Age (adult, child) and Sex (male, female). The last variable, Survived (yes, no) will be taken as an additional variable, which will be projected on the factor planes established with the first 3 variables.

We therefore carry a MCA on the table with 2201 rows and 3 + 1 variables previously mentioned.



Details

Singular Value	Inertia	ChiSquare	Percent	Cumulative Percent
0,70072	0,49101	3534,1	29,46	29,46
0,61706	0,38077	2740,6	22,85	52,31
0,57735	0,33333	2399,2	20,00	72,31
0,51126	0,26138	1881,3	15,68	87,99
0,44740	0,20017	1440,7	12,01	100,00

Benzecri Adjusted Inertia

Inertia	Adjusted Inertia	Percent	Cumulative Percent
0,49101	0,05594	91,70	91,70
0,38077	0,00506	8,30	100,00
0,33333	2,8e-32	0,00	100,00

Greenacre Adjusted Inertia

Inertia	Adjusted Inertia	Percent	Cumulative Percent
0,49101	0,05594	74,56	74,56
0,38077	0,00506	6,75	81,30
0,33333	2,8e-32	0,00	81,30

The number of eigenvalues is less than or equal to the total number of active categories minus the number of variables, here is 4 + 2 + 2 - 3 = 5.
The sum of the the eigenvalues (trace, total inertia) equals the total number of categories divided by the number of variables minus 1, or (8/3) - 1 = 1.6667.
The percentage of the total inertia ed by each factor axis, initially too pessimistic, is corrected by the Benzécri’s adjustment (which considers here that the first two factorial axes are sufficient) and the Greenacre’s adjustment, slightly less optimistic, but which also validates here the choice of only the first two factorial axes.

Partial Contributions to Inertia for the Column Points

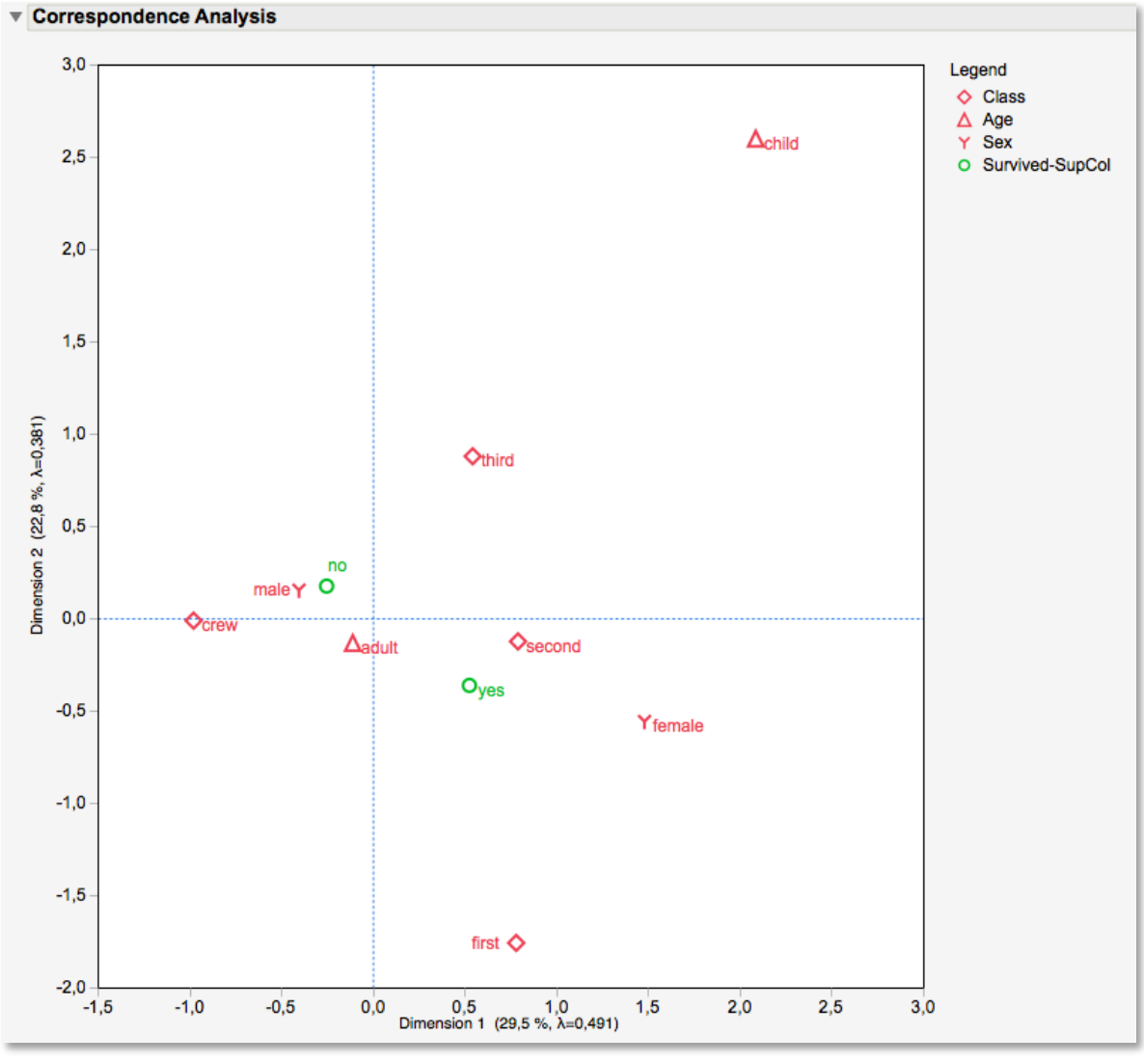
Y	Category	Dimension 1	Dimension 2	Dimension 3
Class	crew	0,26039	6,25e-5	0,00230
Class	first	0,06129	0,40004	0,11172
Class	second	0,05510	0,00179	0,74511
Class	third	0,06462	0,21612	0,14087
Age	adult	0,00763	0,01523	0,00000
Age	child	0,14653	0,29237	0,00000
Sex	female	0,31807	0,05850	0,00000
Sex	male	0,08636	0,01589	0,00000

Squared Cosines for the Column Points

Y	Category	Dimension 1	Dimension 2	Dimension 3
Class	crew	0,64151	0,00012	0,00384
Class	first	0,10592	0,53613	0,13108
Class	second	0,09323	0,00235	0,85595
Class	third	0,14015	0,36345	0,20739
Age	adult	0,22709	0,35137	0,00000
Age	child	0,22709	0,35137	0,00000
Sex	female	0,59575	0,08498	0,00000
Sex	male	0,59575	0,08498	0,00000

Let’s note first that the second class is poorly represented on the first factorial plane (the cosine square of its angle with the third axis is much more important than with the first two axes).

Here is the first factorial plane (1x2):



The first dimension shows that there are many survivors among women and children, and very few among the crew, quite in keeping with the spirit of the saying (except for the *Concordia* crew...).
The second dimension shows that the survivors are on the side of the first class rather than that of children and the third class. We have to explain this paradoxical situation of children.

The Burt table of this MCA shows that out of the 109 children on board, 79 were in the third class, and 57 out of 109 survived.

Burt Table

Count		Class				Age		Sex		Total
Total %		crew	first	second	third	adult	child	female	male	
Class	crew	885 4,47	0 0,00	0 0,00	0 0,00	885 4,47	0 0,00	23 0,12	862 4,35	2655 13,40
	first	0 0,00	325 1,64	0 0,00	0 0,00	319 1,61	6 0,03	145 0,73	180 0,91	975 4,92
	second	0 0,00	0 0,00	285 1,44	0 0,00	261 1,32	24 0,12	106 0,54	179 0,90	855 4,32
	third	0 0,00	0 0,00	0 0,00	706 3,56	627 3,17	79 0,40	196 0,99	510 2,57	2118 10,69
Age	adult	885 4,47	319 1,61	261 1,32	627 3,17	2092 10,56	0 0,00	425 2,15	1667 8,42	6276 31,68
	child	0 0,00	6 0,03	24 0,12	79 0,40	0 0,00	109 0,55	45 0,23	64 0,32	327 1,65
Sex	female	23 0,12	145 0,73	106 0,54	196 0,99	425 2,15	45 0,23	470 2,37	0 0,00	1410 7,12
	male	862 4,35	180 0,91	179 0,90	510 2,57	1667 8,42	64 0,32	0 0,00	1731 8,74	5193 26,22
Total		2655 13,40	975 4,92	855 4,32	2118 10,69	6276 31,68	327 1,65	1410 7,12	5193 26,22	19809 100,00

Contingency Table: Supplementary Columns

Count		Survived		Total
Total %		no	yes	
Class	crew	673 10,19	212 3,21	885 13,40
	first	122 1,85	203 3,07	325 4,92
	second	167 2,53	118 1,79	285 4,32
	third	528 8,00	178 2,70	706 10,69
Age	adult	1438 21,78	654 9,90	2092 31,68
	child	52 0,79	57 0,86	109 1,65
Sex	female	126 1,91	344 5,21	470 7,12
	male	1364 20,66	367 5,56	1731 26,22
Total		4470 67,70	2133 32,30	6603 100,00

If we compare these results with those of Burt table obtained after MCA performed on the subarray of individuals of the only first and second classes, we find that the 30 children of both classes have all survived, i.e. 100% . And therefore only 57 - 30 = 27 third-class children on board 79 survived, i.e. only 34%.

Burt Table

Count		Class		Age		Sex		Total
Total %		first	second	adult	child	female	male	
Class	first	325 5,92	0 0,00	319 5,81	6 0,11	145 2,64	180 3,28	975 17,76
	second	0 0,00	285 5,19	261 4,75	24 0,44	106 1,93	179 3,26	855 15,57
Age	adult	319 5,81	261 4,75	580 10,56	0 0,00	237 4,32	343 6,25	1740 31,69
	child	6 0,11	24 0,44	0 0,00	30 0,55	14 0,26	16 0,29	90 1,64
Sex	female	145 2,64	106 1,93	237 4,32	14 0,26	251 4,57	0 0,00	753 13,72
	male	180 3,28	179 3,26	343 6,25	16 0,29	0 0,00	359 6,54	1077 19,62
Total		975 17,76	855 15,57	1740 31,69	90 1,64	753 13,72	1077 19,62	5490 100,00

Contingency Table: Supplementary Columns

Count		Survived		Total
Total %		no	yes	
Class	first	122 6,67	203 11,09	325 17,76
	second	167 9,13	118 6,45	285 15,57
Age	adult	289 15,79	291 15,90	580 31,69
	child	0 0,00	30 1,64	30 1,64
Sex	female	17 0,93	234 12,79	251 13,72
	male	272 14,86	87 4,75	359 19,62
Total		867 47,38	963 52,62	1830 100,00

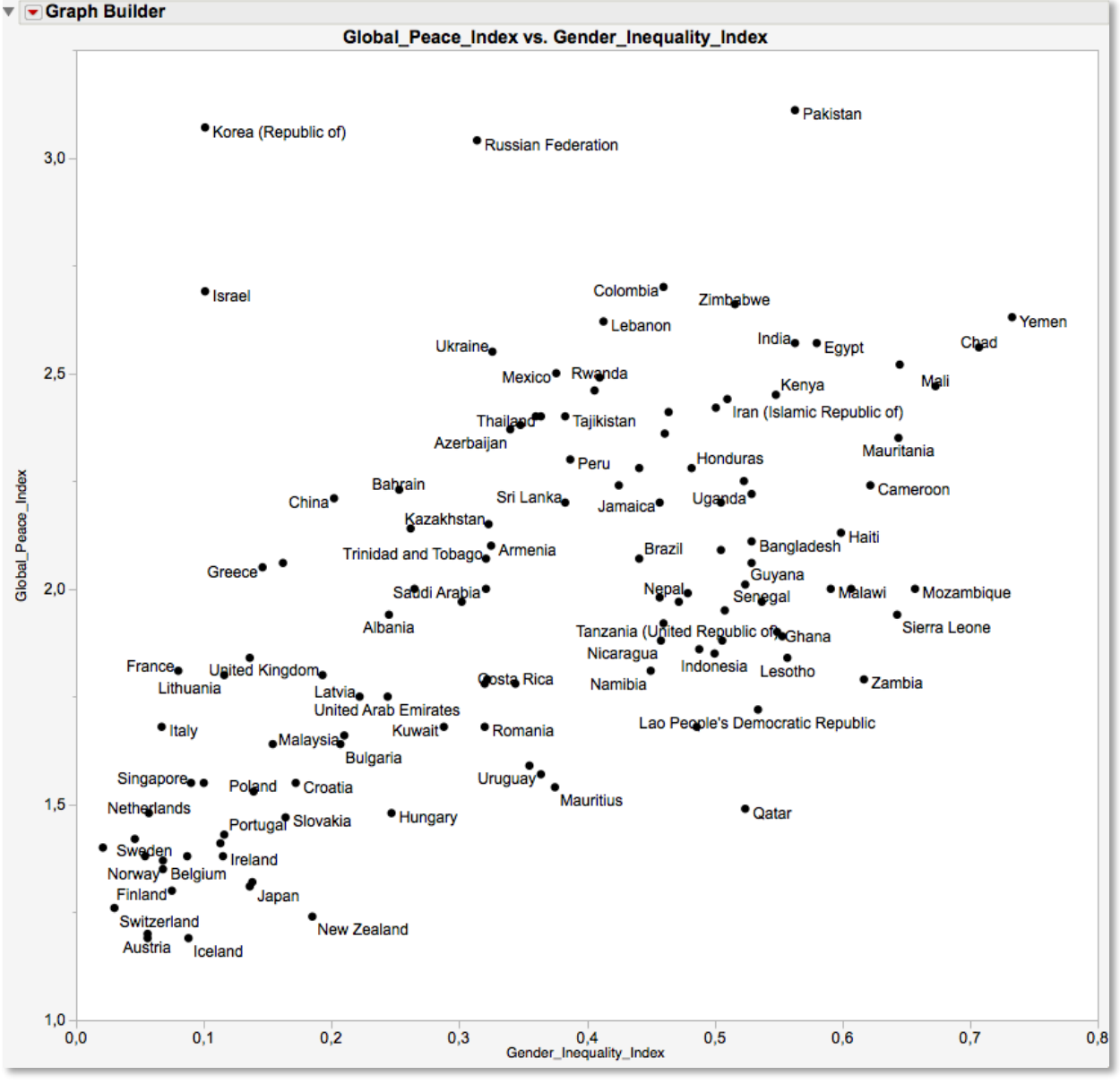
This explains the paradox found in the factorial plane 1-2: children who appear on axis 1 are those of "luxury" classes and those who are less fortunate on axis 2 are those of the third class. MCA enables to clarify the interpretation of the saying: *women and children first*, certainly, but it is also better to be in the proper class ...

THE COMPETITION AMONG INDICES (MCA AND NON LINEARITIES)

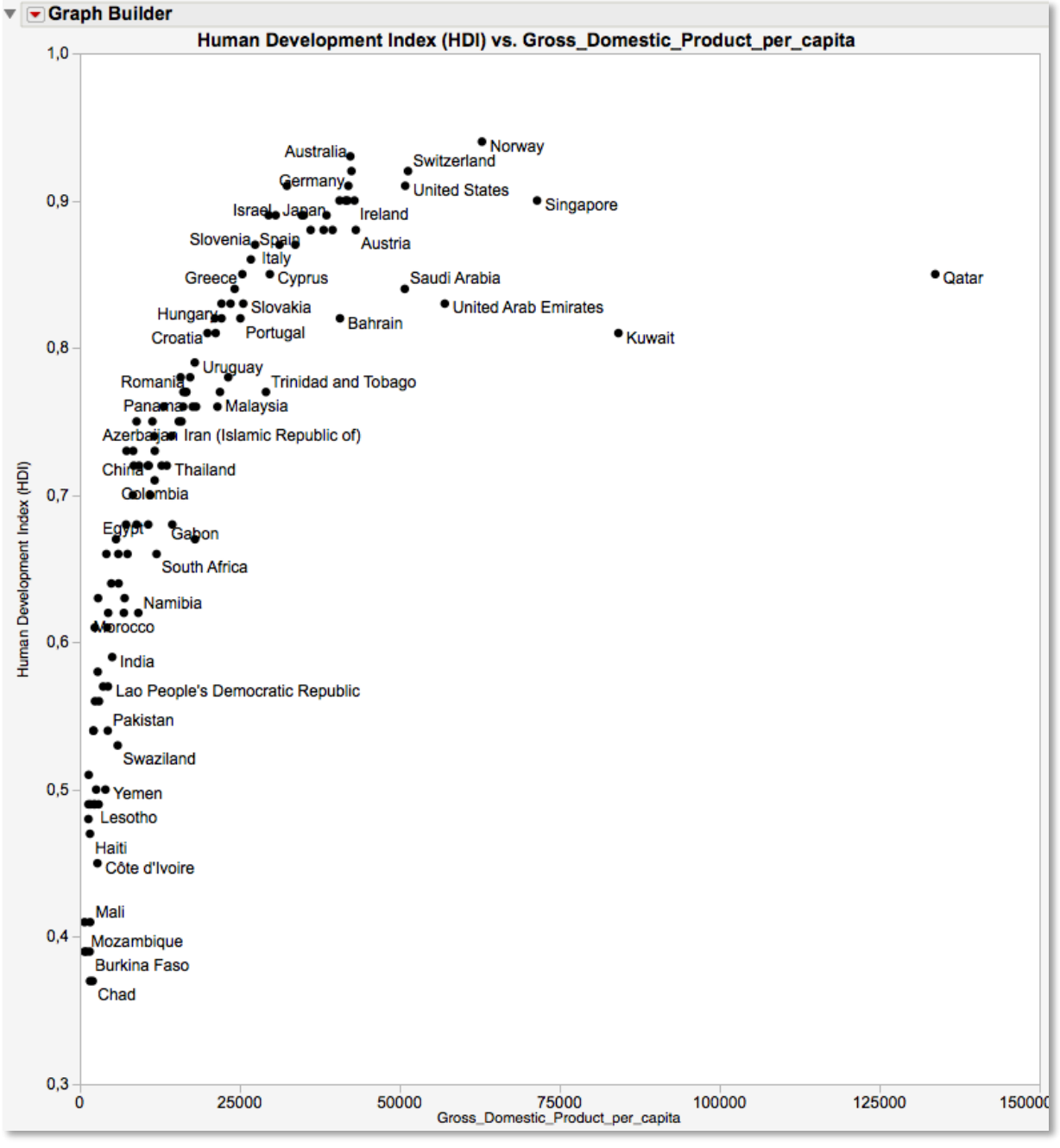
Wealth is by excellence a multidimensional scale, be it that of an individual or of a nation. Yet it is still too often measured by GDP alone. The research initiated in the early 1990s by the United Nations Development Programme team (UNDP) contributed to changing thoughts about it. The paradox is that UNDP has had to publish himself as "his" index –one dimension- the Human_Development_Index (HDI), so the media and politicians can contrast it with the almighty GDP.

Many NGOs are now publishing indices of all kinds. In the JMP **Indices120** table, we have collected for 120 countries some of these indices, starting with the HDI, but also Life_expectancy_at_birth the Mean years of schooling, the Overall_life_satisfaction_Index the Global_Peace_Index (GPI), the Global_Competitiveness_Index (GCI), the Gender_Inequality_Index (G2I), and of course the Gross_Domestic_Product_per_capita (GDP).

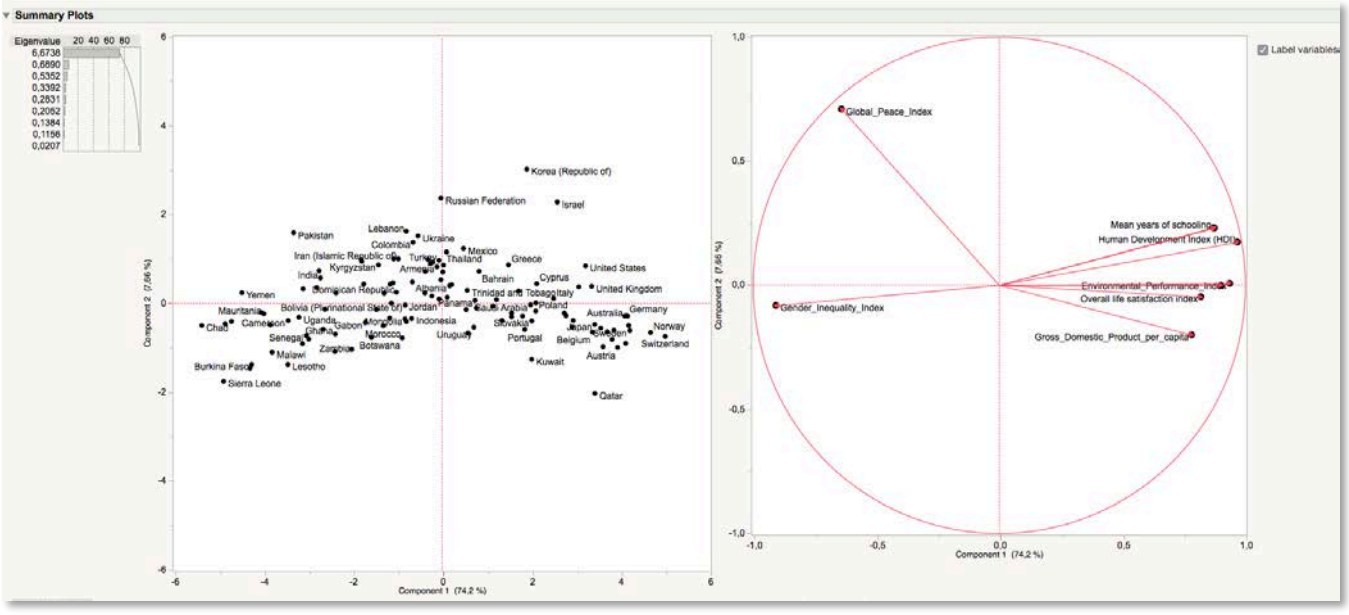
The relationship between these indices are sometimes reasonably "linear", such as the interesting link between Global_Peace_Index and Gender_Inequality_Index: most conscious gender equality countries would be less quarrelsome …



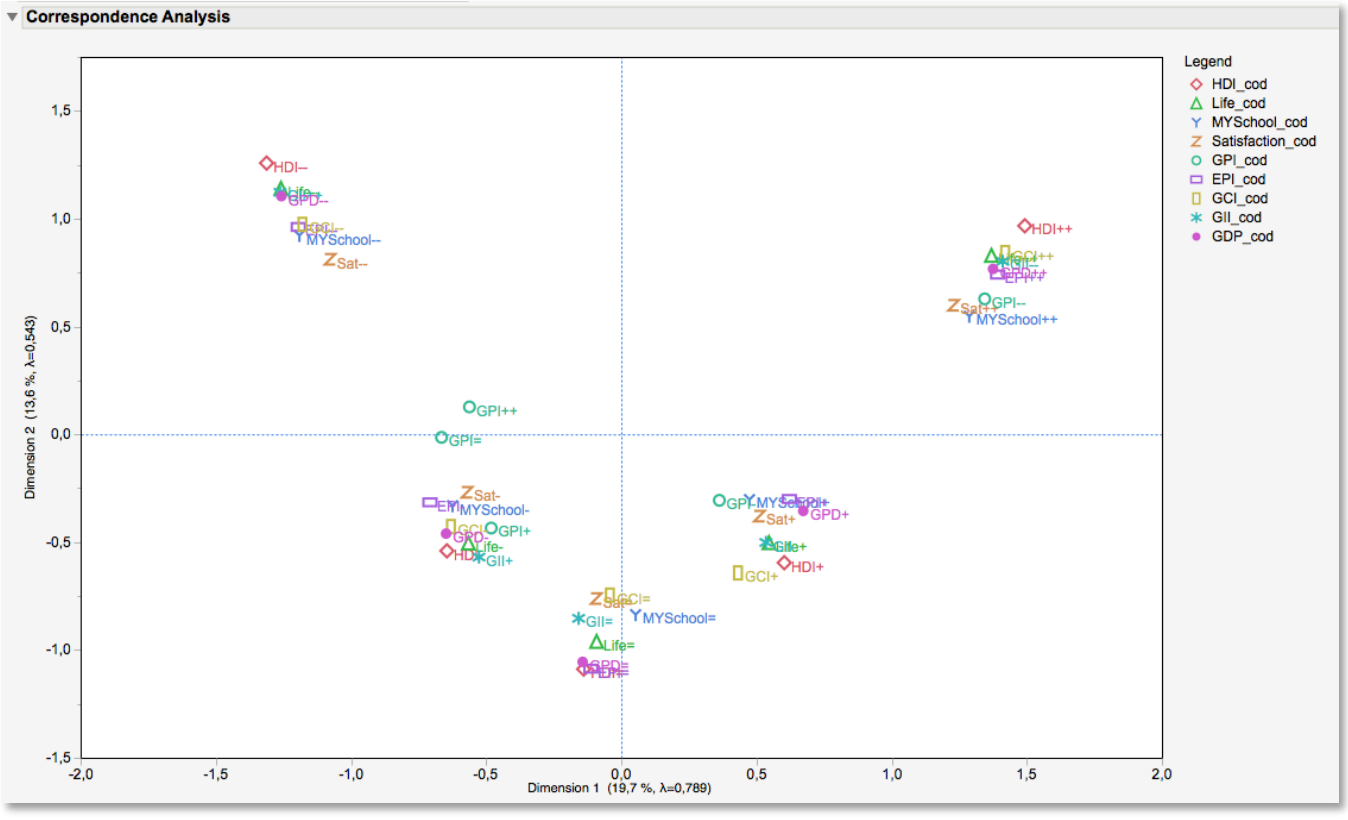
But most of the time this linearity is at fault, such as between HDI and GDP: a very important wealth does not necessarily lead to an important human development.



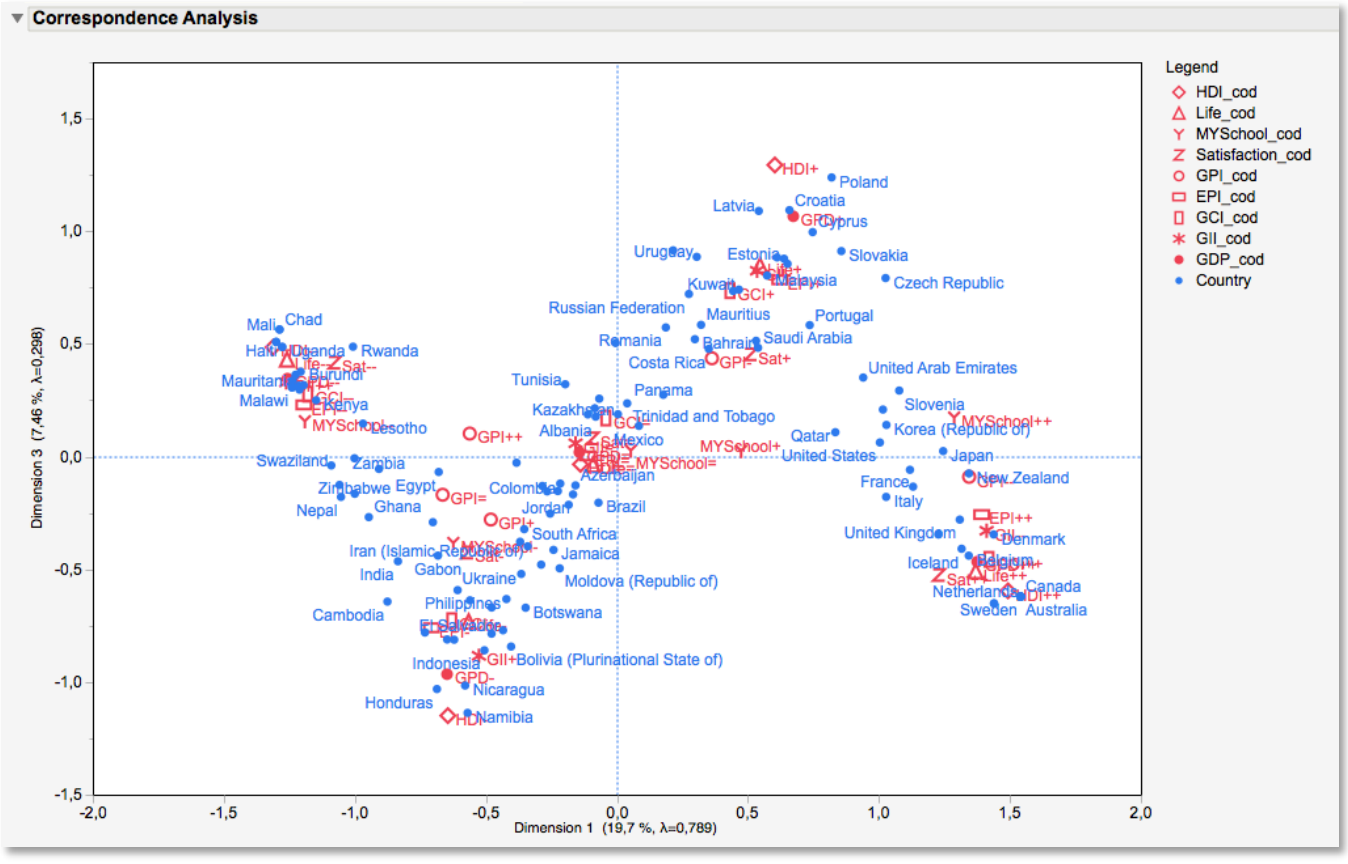
The principal component analysis performed on the table shows that the indices are highly correlated with the first component, except Global_Peace_Index. This first major component exhibits what is called a "size" effect, that is to say here, a gradient of quality indices.



But we can verify this collinearity by performing a MCA on these variables, previously transformed into categorical variables. To do this, we choose a division into 5 bins of quasi equal sample size, denoted --, -, =, +, ++, from the lowest level to the highest level, for each variable. We are seeing a Guttman effect quite exemplary, where the second-factor is a quadratic function of the first factor and the third-factor a function of degree 3 of the first factor.
This effect is observed especially when the categories of each of the categorical variables are ordered *a priori*, which is our case, since the variables come from the recoding up continuous variables (variables of the first "size effect" PCA factor).


Correspondence Analysis
Legend
HDI_cod
Life_cod
MYSchool_cod
Satisfaction_cod
GPI_cod
EPI_cod
GCI_cod
GII_cod
GDP_cod
Dimension 1 (19,7 %, λ=0,789)
Dimension 2 (13,6 %, λ=0,543)


Correspondence Analysis
Legend
HDI_cod
Life_cod
MYSchool_cod
Satisfaction_cod
GPI_cod
EPI_cod
GCI_cod
GII_cod
GDP_cod
Country
Dimension 1 (19,7 %, λ=0,789)
Dimension 2 (13,6 %, λ=0,543)


Correspondence Analysis
Legend
HDI_cod
Life_cod
MYSchool_cod
Satisfaction_cod
GPI_cod
EPI_cod
GCI_cod
GII_cod
GDP_cod
Country
Dimension 3 (7,46 %, λ=0,298)
Dimension 1 (19,7 %, λ=0,789)
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
2,0
HDI+
Poland
Latvia
Croatia
Cyprus
Uruguay
Estonia
Slovakia
Kuwait
Malaysia
Czech Republic
GCI+
Russian Federation
Mauritius
Portugal
Romania
Bahrain
Saudi Arabia
Costa Rica
GPI-
Sat+
United Arab Emirates
Mali
Chad
Rwanda
Tunisia
Panama
Slovenia
Malawi
Kenya
Kazakhstan
Trinidad and Tobago
MYSchool++
Lesotho
GPI++
Albania
Mexico
Korea (Republic of)
MYSchool+
Qatar
Swaziland
Zambia
United States
Japan
MYSchool=
Azerbaijan
France
New Zealand
Zimbabwe
Egypt
GPI=
Colombia
Italy
Nepal
Ghana
Jordan
Brazil
EPI++
GPI+
United Kingdom
Denmark
South Africa
Iran (Islamic Republic of)
Jamaica
Iceland
Belgium
India
Gabon
Ukraine
Moldova (Republic of)
Netherlands
Canada
Cambodia
Philippines
Botswana
Sweden
Australia
Indonesia
GII+
Bolivia (Plurinational State of)
GDP-
Nicaragua
Honduras
Namibia
HDI-

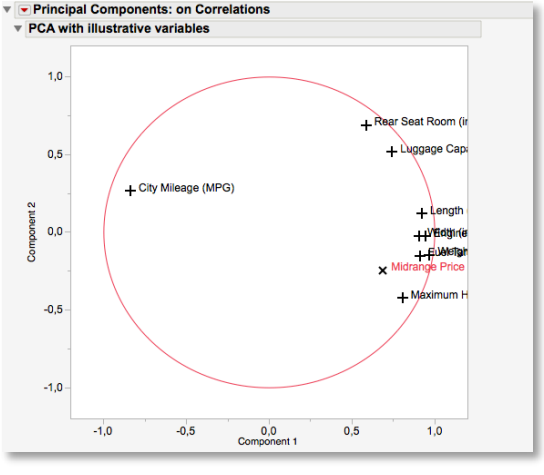
THE CARS 1993 DATA REVISITED (HOW TO SUPPORT BOTH CONTINUOUS AND CATEGORICAL VARIABLES IN THE SAME ANALYSIS)

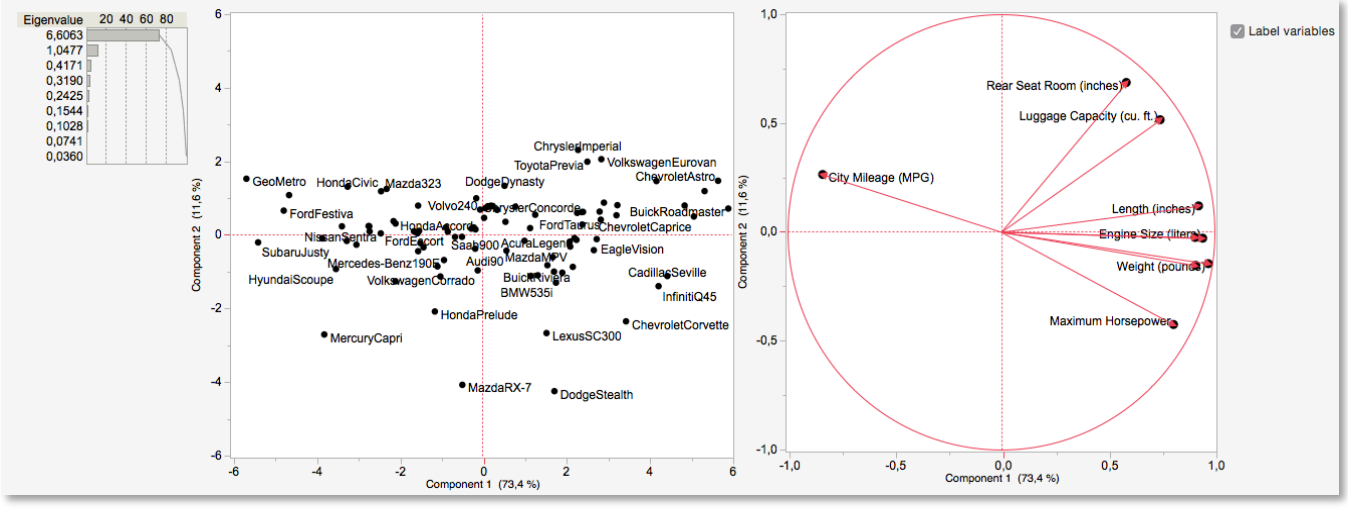
Let's consider the JMP table obtained by selecting some of the variables in the **Cars 1993** table (in data samples of JMP Help).
The variables are of two types:

- The continuous variables: Midrange Price ($1000), City Mileage (MPG), Engine Size (liters), Maximum Horsepower, Fuel Tank Capacity, Length (inches), Width (inches), Luggage Capacity (cu. ft.), Weight (pounds), Rear Seat Room (inches),
- The categorical variables (nominal): Manufacturer, Model, Car_ID, Vehicle Category, Drive Train Type, Manual Transmission Available, Domestic Manufacturer, Air Bags Standard.

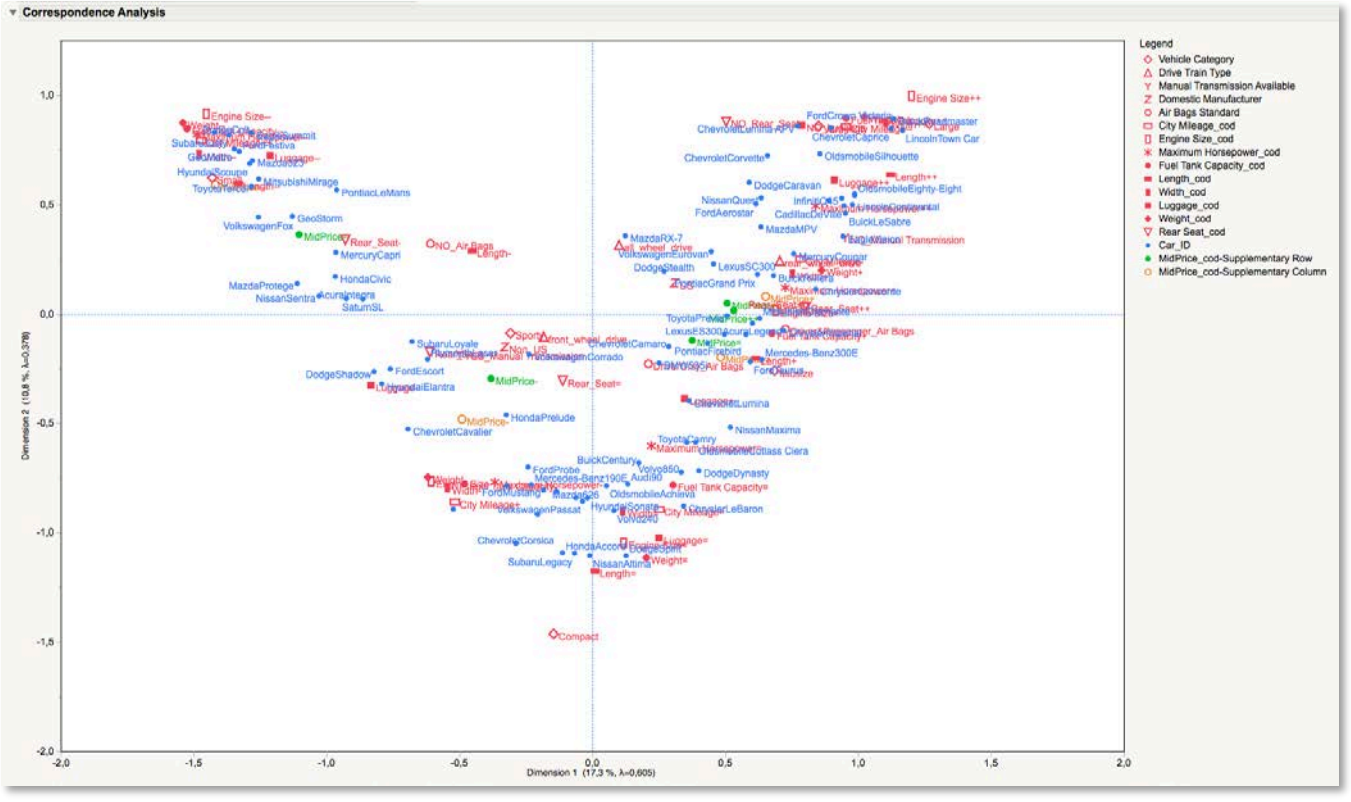
If we wish to perform a multidimensional analysis of this table, for example we can make a couple PCA+Clustering, but only on continuous data.
MCA enables us to take into account two types of data, provided the continuous variables are recoded into categorical variables. As with the previous example, we choose a division into 5 bins of quasi equal sample size, denoted --, -, =, +, ++, from the lowest level to the highest level, for each variable.
To illustrate this point, we chose to take the price variable (Midrange Price) as an illustrative additional variable in both the PCA and the MCA. Let's consider a regression modeling where average price of a car is a function of explanatory variables (continuous and categorical): these two factorial analyses enable us to "X-ray" candidate variables to explain the price and identify potential collinearity problems.

For PCA, on continuous variables (Midrange Price variable being taken as illustrative, –thanks to the Add-in **PCA with illustrative variables** by Florence Kussener of JMP-France):





And for MCA, with all variables (initial categorical variables and transformed variables) :



Once more, we find a Guttman effect, less pronounced than in the previous example, simply because *a contrario* of transformed variables, initial categorical variables do not have ordered categories.

One must also consider that the MCA enables us to achieve a clustering that takes into account all the variables (continuous and categorical), unlike the usual **Clustering** that supports the continuous variables. Just do that **Clustering** with individuals coordinates (cars) on factorial axes of the MCA, previously saved (note that this backup is not that easy in the MCA platform since it requires displaying more than three columns, and then a **Make into Data Table**).

CONCLUSION

In this presentation, we have illustrated some possibilities of the new JMP® 12 ***Multiple Correspondence Analysis*** (MCA) platform. This new platform enables us to perform correspondence analysis, single and multiple, with all options and aids in the interpretation of the results that the French statisticians are accustomed to. This is very good news.
Perhaps still can one ask a little more :

- Why not a PCA that enables now supplementary individuals and variables?
- Why not the very nice **Select Dimension** of the MCA plots also in the PCA?
- Why the lack of dynamic graphics in this MCA platform? (If we click on a point, it is not selected in the other windows; maybe it's difficult because of the mixing of the two spaces on the factorial maps)
- Why not a **Save Row and Column Coordinates** in MCA, the same way we have a **Save Principal Components** in PCA?
- Maybe it's worth reconsidering the place of MCA: why not in the **Analyze→Multivariate Methods**?
- Why keep the old JMP® 11 CA?

REFERENCES

1. Jean-Paul Benzécri, *Correspondence Analysis Handbook*, New York: Marcel Dekker, 1992.
2. Michael J. Greenacre. *Correspondence Analysis in Practice.* 2nd edn. 2007. Chapman&Hall/CRC, London.
3. Michael J. Greenacre. Correspondence analysis Focus Article WIREs Computational Statistics, Vol. 2, Issue 5, 2010.
4. Michel Tenenhaus and Forrest W. Young. An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods for Quantifying Categorical Multivariate Data. *Psychometrika*, Vol. 50, O. 1, 91-119,1985.
5. Phillip M. Yelland, An Introduction to Correspondence Analysis, *The Mathematica Journal*, Vol. 12, 2010.
6. *Cahiers de l'Analyse des données*, 1976-1996, archived on http://www.numdam.org/numdam-bin/feuilleter?j=CAD,
7. CARME-N.ORG http://www.carme-n.org/
8. Global Peace Index Report 2014, in www.visionofhumanity.org
9. Human Development Report 2014, United Nations Development Programme, in http://hdr.undp.org/en