

Biplots in Practice

MICHAEL GREENACRE

Professor of Statistics at the Pompeu Fabra University

Chapter 10 Offprint

Multiple Correspondance Analysis Biplots II

First published: September 2010
ISBN: 978-84-923846-8-6

Supporting websites:
<http://www.fbbva.es>
<http://www.multivariatestatistics.org>

© **Michael Greenacre, 2010**
© **Fundación BBVA, 2010**

Multiple Correspondence Analysis Biplots II

Multiple correspondence analysis (MCA) is the CA of a special type of concatenated table, where a set of variables is cross-tabulated with itself. Therefore we can say that, whereas in Chapter 9 the concatenation was made *between* two sets of different variables, in this chapter the concatenation is made of variables *within* one single set. This concatenated matrix has identical rows and columns and is thus square symmetric. It includes cross-tabulations between each variable and itself, which are diagonal matrices of perfect association. These perfect associations are impossible to represent in a low-dimensional display, so any biplot as described in Chapter 9 of this particular concatenated matrix would be degraded as far as overall quality of data representation is concerned. This problem is avoided with a simple adjustment of the solution (or by using an alternative approach called joint correspondence analysis). In this chapter we show some possibilities for biplots in this MCA context, where the rows and columns are identical, and also how individual case points, or group averages of cases, can be displayed.

Contents

Burt matrix	99
Indicator matrix	101
Indicator matrix biplot	102
Adjusted inertias	103
Burt matrix biplot	104
MCA contribution biplot	104
Supplementary points	105
SUMMARY: Multiple Correspondence Analysis Biplots II	108

Using the same “women” data set for Spain as in Chapter 9, Exhibit 10.1 shows part of the concatenated table of all the cross-tabulations of the questions with one another (here, for the moment, we do not consider the demographic variables). This is a square symmetric block matrix, since the cross-tabulation of variable q with variable s is the transpose of that of variable s with variable q . This ma-

[Burt matrix](#)

Exhibit 10.1:

Part of the Burt matrix of the eight variables of the “women” data set, showing the first three variables cross-tabulated with one another, including the cross-tabulations of perfect association between each variable and itself down the diagonal blocks

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	C1	C2	C3	C4	C5	
a1	397	0	0	0	0	19	113	42	132	91	37	91	45	154	70	...
a2	0	932	0	0	0	18	362	126	405	21	21	405	128	359	19	...
a3	0	0	91	0	0	2	44	22	21	2	8	44	25	13	1	...
a4	0	0	0	598	0	40	411	42	101	4	48	422	36	88	4	...
a5	0	0	0	0	89	45	26	5	6	7	51	26	3	6	3	...
b1	19	18	2	40	45	124	0	0	0	0	80	34	4	3	3	...
b2	113	362	44	411	26	0	956	0	0	0	52	673	65	154	12	...
b3	42	126	22	42	5	0	0	237	0	0	12	82	82	59	2	...
b4	132	405	21	101	6	0	0	0	665	0	8	182	79	378	18	...
b5	91	21	2	4	7	0	0	0	0	125	13	17	7	26	62	...
c1	37	21	8	48	51	80	52	12	8	13	165	0	0	0	0	...
c2	91	405	44	422	26	34	673	82	182	17	0	988	0	0	0	...
c3	45	128	25	36	3	4	65	82	79	7	0	0	237	0	0	...
c4	154	359	13	88	6	3	154	59	378	26	0	0	0	620	0	...
c5	70	19	1	4	3	3	12	2	18	62	0	0	0	0	97	...
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:

trix is called a *Burt matrix*, after the psychometrician Cyril Burt, who first considered this type of data structure.

Like the concatenated table of Chapter 9, this one has the property that each subtable has a sum equal to the sample size, and the row margins are all the same across the variables in each row block and the column margins are all the same across the variables in each column block. Hence, the inertia of the Burt matrix of, in general, Q variables is the average of the inertias in the Q^2 subtables. In this eight-variable example there are eight cross-tabulations down the diagonal, each variable cross-tabulated with itself, which are diagonal matrices of perfect association, each having an inertia of 1 less than the number of categories, i.e. 4 in this example (a 5×5 diagonal matrix has four principal inertias, each equal to 1). These high values inflate the total inertia considerably, so we do not include them in the computation of the total inertia, preferring to define the total inertia to be explained as the average of all the other (“off-diagonal”) subtables, of which there are $Q^2 - Q = Q(Q - 1)$. This is called the *adjusted inertia* in MCA. In fact, since the matrix is symmetric we can compute the adjusted inertia as the average of the upper or lower triangle of subtables, of which there are $\frac{1}{2}Q(Q - 1)$. If the usual total inertia of the complete Burt matrix, $\text{inertia}(\mathbf{B})$, is available, the constant amounts due to the problematic diagonal matrices can be simply subtracted, and the adjusted inertia can be shown to be:

$$\text{adjusted inertia of Burt matrix} = \frac{Q}{Q-1} \left(\text{inertia}(B) - \frac{J-Q}{Q^2} \right) \quad (10.1)$$

where Q = number of variables, J = total number of categories in all the variables. For example, in this example the total inertia of $\mathbf{B}$ is equal to 0.6776, but when we make the calculation removing the contributions of the eight diagonal matrices it reduces to (remembering that we combined two categories of H , so there are 39 categories in total):

$$\text{adjusted inertia of Burt matrix} = \frac{8}{7} \left(0.6776 - \frac{39-8}{64} \right) = 0.2208$$

(If one computes the individual inertias in the $\frac{1}{2} \times 8 \times 7 = 28$ cross-tabulations between pairs of variables and averages them, the result is identical.)

In Exhibit 10.2, on the left, the first five records of the original respondent-level data are shown, first the response categories to the eight questions, then the demographic groups for gender, marital status, education, age and the gender-age combination. On the right we see an alternative way of coding the data, where the columns are *dummy variables*, one column for each category of response, and coded with a 1 to indicate the response category, otherwise 0. The matrices on the right are called *indicator matrices*. The indicator matrices can be used to construct concatenated tables: for example, if we denote the 39-column indicator matrix for the eight questions A to H as $\mathbf{Z}$, then the Burt matrix $\mathbf{B}$ is simply:

$$\mathbf{B} = \mathbf{Z}^T \mathbf{Z} \quad (10.2)$$

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>g</i>	<i>m</i>	<i>e</i>	<i>a</i>	<i>ga</i>	<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	<i>A5</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>B5</i>	...	<i>g1</i>	<i>g2</i>	<i>m1</i>	<i>m2</i>	<i>m3</i>	<i>m4</i>	<i>m5</i>	...	
2	4	3	3	4	1	4	1	1	1	3	4	4	0	1	0	0	0	0	0	0	0	1	0	...	1	0	1	0	0	0	0	...
3	2	2	3	4	1	3	2	2	2	1	6	12	0	0	1	0	0	0	0	1	0	0	0	...	0	1	0	1	0	0	0	...
2	4	4	4	2	2	5	1	1	5	4	2	2	0	1	0	0	0	0	0	0	0	1	0	...	1	0	0	0	0	0	1	...
1	3	2	3	2	4	4	2	2	1	5	5	11	1	0	0	0	0	0	0	0	1	0	0	...	0	1	1	0	0	0	0	...
2	4	2	3	4	4	5	1	2	1	5	3	9	0	1	0	0	0	0	0	0	0	1	0	...	0	1	1	0	0	0	0	...
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	...	:	:	:	:	:	:	:	...	

Exhibit 10.2:
Data for first five respondents (out of 2107) in the “women” data set, showing on the right the corresponding indicator coding of some of the variables

The total inertia of an indicator matrix of Q variables with a total of J categories can be shown to be equal to a constant which depends only on J and Q :

$$\text{total inertia of indicator matrix} = (J - Q)/Q \quad (10.3)$$

In the present example, this inertia would be equal to $(39 - 8)/8 = 3.875$.

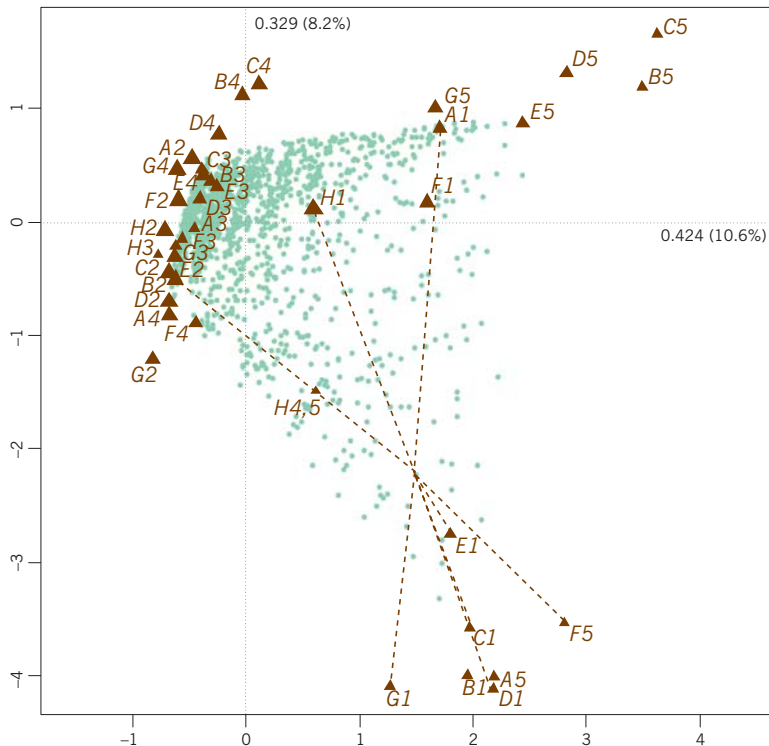
Indicator matrix biplot

All the above results are relevant to understanding biplots for multiple correspondence analysis (MCA). MCA is classically defined as the CA algorithm applied to either the indicator matrix $\mathbf{Z}$ or the Burt matrix $\mathbf{B}$. It is well-known that the two alternatives lead to exactly the same standard coordinates for the variable categories, while the singular values in the CA of $\mathbf{B}$ are the squares of those in the CA of $\mathbf{Z}$. What interests us here is how to biplot the results meaningfully so that data are successfully recovered in the graphical representation. For example, consider the asymmetric map of the indicator matrix, with rows in principal coordinates and columns in principal coordinates, shown in Exhibit 10.3.

In contrast to the biplots of Chapter 9, which showed relationships between the question responses and the demographic categories, this biplot shows the relationships amongst the response categories themselves. All the extreme categories (the 1's and 5's) are on the right and all the moderate categories (2's and 4's) and

Exhibit 10.3:

Asymmetric map/biplot of the 2107×39 indicator matrix of the eight questions of the "women" data set. Each respondent point is at the average of the corresponding eight response categories—an example is shown of a respondent linked to her responses A1, B2, C1, D1, E1, F5, G1, H1



middles (3's) are on the left. The main feature of the data is thus the opposition between respondents giving extreme opinions on the issue of working women and those with moderate attitudes. The conservative-liberal dimension in the responses is reflected in the vertical spread of the response categories, with conservative views at the bottom and liberal views at the top, on both extreme and moderate sides of the map.

Each row (respondent) is displayed as a green dot, and because each row of the indicator matrix consists of 0s apart from eight 1s in the response category positions, its profile consists of 0s except for the value $1/8$ in those positions. The barycentric (or weighted average) property in the asymmetric map implies that each row point is at the ordinary average position of its eight response categories, for example the point shown in Exhibit 10.3, which is situated towards the extreme conservative region of the map. It is clear that in such a biplot there is no question of calibrating axes or trying to recover the 0/1 data—this is borne out by the fact that percentages of inertia are very low (10.6% and 8.2% respectively on the axes), which is typical if the indicator matrix is analyzed.

While it is interesting to see the whole cloud of respondent points, the biplot of Exhibit 10.3 can be made more meaningful by representing subgroups of points, for example the demographic groups such as male/female, or high/low education. This is explained below using the supplementary point idea, but basically the idea is to remove the individual case points and rather display average points for individuals in specified demographic groups.

A biplot based on the Burt matrix is similar to the one we had for concatenated tables in Chapter 9, except for two major differences: the row and column variables are the same, and the tables of perfect association down the diagonal of the Burt matrix need to be avoided in some way. Our objective is to achieve a biplot that reconstructs the profiles of all tables apart from those in the diagonal blocks, and we have already proposed an adjusted total inertia that omits these cross-tabulations of perfect association. A simple adjustment of the singular values turns out to be just what is necessary so that the MCA solution best fits the off-diagonal tables. If λ_k denotes the k -th principal inertia of $\mathbf{B}$, then only those axes for which $\sqrt{\lambda_k}$ are larger than $1/Q$ are retained (notice that the $\sqrt{\lambda_k}$'s are exactly the principal inertias of the CA of the indicator matrix). The adjustment is as follows:

Adjusted inertias

$$\frac{Q}{Q-1} \left(\sqrt{\lambda_k} - \frac{1}{Q} \right), \quad k=1, 2, \dots, \quad \text{for } \sqrt{\lambda_k} > \frac{1}{Q} \quad (10.4)$$

The values in (10.4) effectively replace the singular values in the MCA and their squares are the *adjusted principal inertias*, which can be expressed relative to the adjusted total inertia of (10.1) to obtain percentages of inertia.⁶

Burt matrix biplot

We now have all the results necessary to define a biplot of the Burt matrix. First, we perform the MCA of the complete Burt matrix, giving us the standard coordinates identical to those used to plot the categories in Exhibit 10.3. The number of principal inertias λ_k that satisfy $\sqrt{\lambda_k} > 1/8$ is equal to 9. Applying the adjustment (10.4), the first two are equal to:

$$\frac{8}{7}\left(\sqrt{0.1801} - \frac{1}{8}\right) = 0.3422 \quad \text{and} \quad \frac{8}{7}\left(\sqrt{0.1079} - \frac{1}{8}\right) = 0.2326$$

These two values replace the singular values in the MCA and then the asymmetric biplot in Exhibit 10.4 can be drawn, with the rows, say, in principal coordinates (standard coordinates multiplied by the above adjusted singular values) and the columns in standard coordinates. Their squared values quantify the amount of inertia accounted for by each axis: $0.3422^2 = 0.1171$ and $0.2326^2 = 0.0541$, and relative to the adjusted total inertia of 0.2208 calculated previously, they explain 53.0% and 24.5% respectively.

One might ask what the benefit is of representing the categories twice, since the row principal coordinates are at the same relative positions along the principal axes as the column standard coordinates. The answer is that, like in any biplot, a biplot axis can be drawn through the point C5, for example, in standard coordinates, and then the profile values of all other categories on C5—except for the categories of the variable C itself—can be lined up by their projections onto that axis. So the fact that categories C5 and B5 are close means that all other response categories have similar profile values on these two categories. A possibly more interesting biplot of the two sets of identically labelled points is provided by the contribution biplot—since we are only interested in the directions of the biplot axes, we can change the lengths of the set of points in standard coordinates to reflect the contributions to the principal axes.

MCA contribution biplot

Since the row configuration in principal coordinates gives us the essential information for interpreting inter-category associations, we can use the column con-

6. The set of dimensions for which $\sqrt{\lambda_k} > 1/Q$ will usually account for less than 100% of the inertia in the off-diagonal cross-tabulations. To account for 100% of the inertia, another form of MCA called *joint correspondence analysis* needs to be used (not treated in this book).

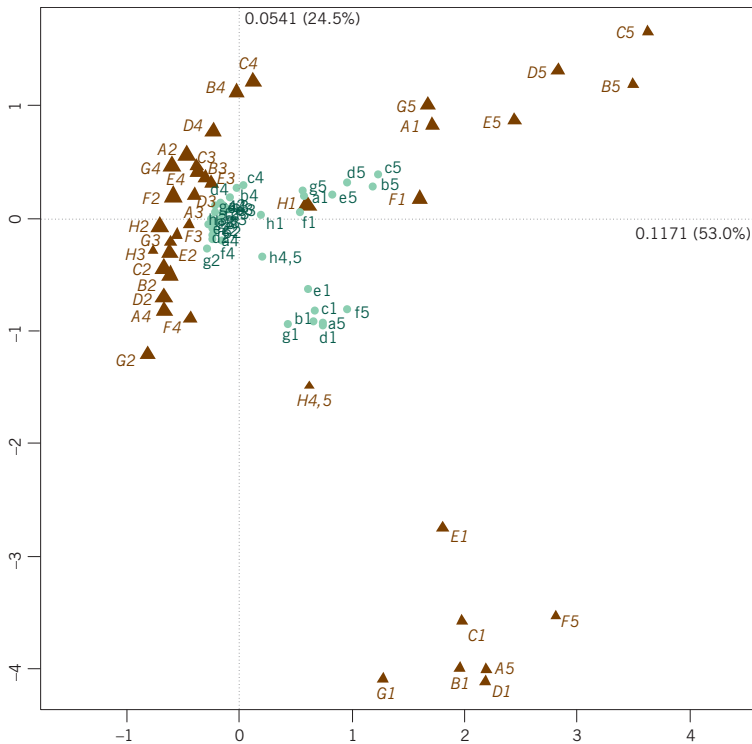


Exhibit 10.4:
Asymmetric map/biplot of the Burt matrix: columns in standard coordinates and rows in principal coordinates using adjusted principal inertias. Percentages of inertia on the two axes are 53.0% and 24.5% respectively

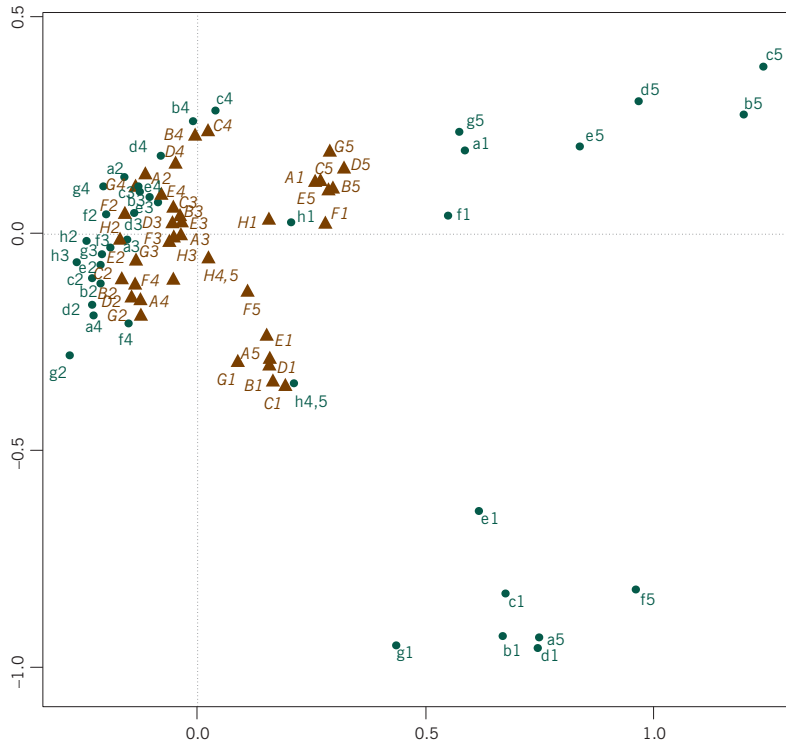
figuration to show which variables and which categories are contributing the most to the solution. The column points in standard coordinates are thus multiplied by the square roots of their masses to obtain the standard MCA biplot of Exhibit 10.5. Notice that the contribution coordinates of the middle categories (3's) are close to the centre, so these categories play a minor role in this biplot. However, they all contribute strongly as a group to the third dimension (not shown here). The separation of the middle response categories is a common phenomenon in survey data and is the theme of the case study in Chapter 14.

As we saw in Exhibit 10.3, every respondent (a row of the indicator matrix) has a position in the biplot depending on that respondent's particular choice of responses. Groups of respondents can be displayed by their mean positions and, optionally, some type of confidence region for each mean. Geometrically, this is simply finding the average of all respondents in the lowest education group, for example, in the display and this gives a point e0, or finding the average of all females in the age group up to 25 years, and this gives a point fa1. Analytically, this is achieved by adding an extra row to the data matrix which accumulates all the frequencies for males across the variables, or all females in the first age group, in

[Supplementary points](#)

Exhibit 10.5:

MCA contribution biplot. The row points (principal coordinates, in green—same as in Exhibit 10.4) show chi-square distances between the categories, while the column points (contribution coordinates, in brown) serve as directions for biplot axes as well as quantifying the contributions of the categories to each dimension



other words, exactly the rows of the concatenated matrix in Exhibit 9.1. The profiles of these rows define points in the MCA space which are the means of the corresponding demographic group.

Exhibit 10.6 shows the demographic group means added to the display of Exhibit 10.5 (categories shown in contribution coordinates only).

Because the upper right represents the most liberal attitude towards working women and the lower right the most conservative (and lower left the more moderate conservative attitude), one can see that the oldest male group and lowest education group are the most conservative while at the top end it is the two highest education groups and the younger female groups that are the most liberal. Not only have the combinations of gender and age been added but also the points representing all males (m) and all females (f) and each age been group (a1 to a6). A similar result to that observed in Exhibit 9.6 can be seen at the top where the young females (e.g., fa1) tend to be strongly liberal, whereas the corresponding male group (ma1) tends to the more moderate liberal side (upper left). The point a1 representing the age group as a whole is between these two points. At

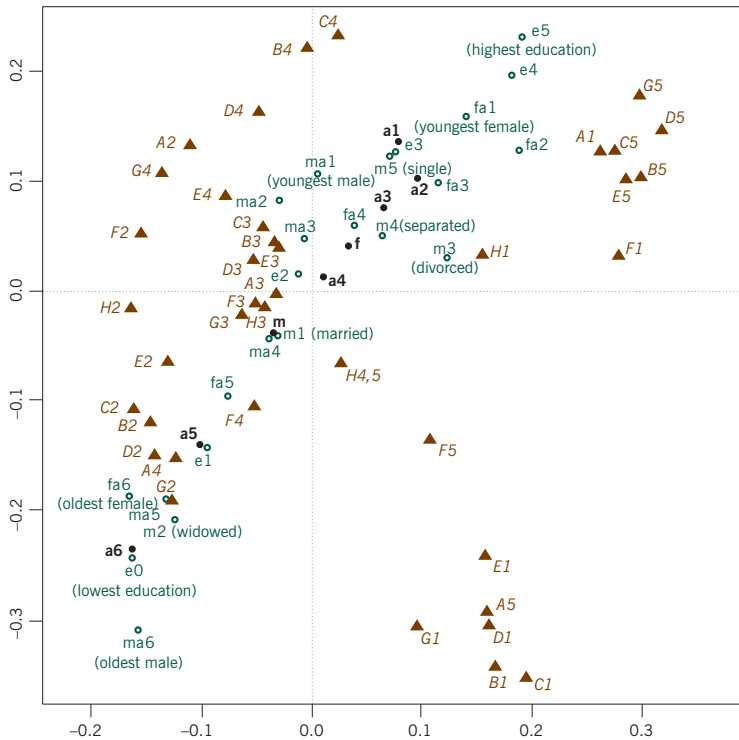


Exhibit 10.6:
MCA contribution biplot,
showing variables in their
contribution positions and
supplementary points added
for the demographic groups

bottom left we have a6 lying between fa6 and ma6 (but closer to fa6 since there are more females in this age group) but here the difference is simply that males are more conservative than females and there is not so much the strong versus moderate attitude observed in the youngest groups. In fact, there is no specific demographic group tending to the extreme conservative attitudes at bottom right.

In summary, Exhibit 10.6 is displaying many aspects of the original 2107×12 data matrix (8 questions and 4 demographic variables (see Chapter 9)). In the first instance it shows the principal axes of the question categories (green points in Exhibit 10.5) in such a way as to maximize the inertia accounted for in all the pair-wise relationships between the questions. Second, it shows the question categories (brown points) with lengths related to their contributions to the solution axes and with directions that define biplot axes, onto which the green points (in Exhibit 10.5) can be projected (excluding the green points for the same question, because projecting a question onto itself has been purposely avoided in finding the solution). Third, demographic groups are displayed as supplementary mean points of the respondents in each respective group.

The MCA in Exhibit 10.6 is primarily focused on displaying the associations among the eight questions, and secondarily on showing how the demographic groups relate to these dimensions. Exhibit 9.6, on the other hand, is focused on the associations between the demographic variables and the questions. From the biplot point of view the difference can be explained as follows. In Exhibit 10.6 if a biplot axis is drawn through *G5*, for example, then the projections of all the question categories *a1* to *a5*, *b1* to *b5*, etc. (but not *g1* to *g5*), will be approximating the profile values of these categories on *G5*—the overall quality of display of all these profiles with respect to the biplot axes is 77.5%, the inertia explained by the first two axes. The demographic categories can also be projected onto directions such as *G5*, but their displayed profile values have not been specifically optimized—the overall quality of display of these supplementary profiles is only 58.1% in Exhibit 10.6. In Exhibit 9.6, on the other hand, the projections of the demographic categories onto the biplot axes defined by the response categories (for example, onto *G5*), were optimal and there the overall quality of display was 86.5%.

SUMMARY:
Multiple Correspondence
Analysis Biplots II

1. One of the ways of defining and thinking about MCA is as the CA of the concatenated table of cross-tabulations of a set of categorical variables with themselves. This square symmetric block matrix is called the *Burt matrix*.
2. The Burt matrix includes down its diagonal blocks the cross-tabulations of each variable with itself, and these tables inflate the total inertia of the problem, leading to low percentages of inertia explained on the principal axes if the Burt matrix is displayed.
3. A simple adjustment of the principal inertias and the total inertia optimizes the solution to the off-diagonal tables that cross-tabulate distinct pairs of variables.
4. Because the rows and columns of the Burt matrix are identical, the contribution biplot is particularly useful: one of the sets, for example the rows, shows the category points in principal coordinates and so displays inter-profile distances, while the other set can display the categories both as biplot axes and with lengths related to their contributions to the solution.
5. In all MCA biplots the respondent points can also be displayed, but it is usually more interesting to show various average positions of groups of respondents in terms of their demographic characteristics. These are added as supplementary points.