Correspondence Analysis

Foundation Entries

SAGE Research Methods Foundations

By: Omar Lizardo & Marshall A. Taylor
Published:2019
Length: 10,000 Words
DOI: http://dx.doi.org/10.4135/9781526421036
Methods: Correspondence Analysis
Online ISBN: 9781526421036
Disciplines: Anthropology, Marketing, Sociology, Computer Science
Access Date: February 6, 2020
Publishing Company: SAGE Publications Ltd
City: London

Abstract

Correspondence analysis (CA) is a data analytic method appropriate for data matrices consisting of categorical variables. These can be cross-tabulations of two or more categorical variables or can be entity-by-entity (e.g., person-by-variable) data sets of more general interest to social and behavioral scientists, as long as the variables are categorical and homogenous. Although CA is at its core a dimension reduction technique, the method is tightly coupled to a geometric interpretation and a visual mode of presentation: the *biplot*, where the row and column elements are projected into a low-dimensional space. In this sense, CA is also a data visualization method. This entry proceeds in the following steps. First, it presents more details behind what exactly CA is and how it has been characterized in different kinds of literature. Second, it describes the basic concepts behind CA: profiles, masses, χ^2 distances, inertias, and computational details. Third, it discusses an important generalization of CA, known as MCA. Fourth, it details some other extensions of CA, including how the method can be used to analyze one- and two-mode network data and text data. It concludes with an overview of currently available software implementing CA and related methods.

Introduction

The anglophone literature on correspondence analysis (CA) has developed rapidly since the publication of the first edition of Michael Greenacre's (1984) classic introductory monograph *Correspondence Analysis*, along with his 1993 overview, *Correspondence Analysis in Practice*, whose third edition was published in 2017 (Greenacre, 2017). Along the way, various characterizations of CA have cropped up in the literature. This entry distinguishes two broad conceptions of CA as a data-analytic method. One based on the types of data most appropriate for the application of CA, the other based on particular ideas as to what data-analytic task CA is best capable of handling.

In the first group, some see CA as a tool for the analysis of *cross-classified* or *cross-tabular* data, similar to that presented in Table 1. From this perspective, CA belongs in the same class as log-linear modeling and canonical correlation analysis (Goodman, 1996). Others see CA, and especially its *multiple correspondence analysis* (MCA) extension, as a general method for the analysis of rectangular data matrices such as those obtained from survey questionnaires administered to samples of individuals. The idea is that typically the analyst is faced with analyzing a structured table with individuals and the rows and a series of categorical variables associated with a given topic. Here CA is seen as a *data reduction* method, allowing the analyst to group variables (and individuals) into meaningful underlying dimensions, which can then later be used as an input for more traditional statistical techniques. From this perspective, CA and MCA belong to the class of data reduction and variable scaling techniques such as factor analysis, principal component analysis (PCA), and multidimensional scaling.

Finally, some see CA as a general method for the analysis of two-mode data. As Stephen P. Borgatti and

Martin G. Everett (1997) note, in a two-way data matrix, a "mode" refers to a type of entity. The "adjacency matrices" popular in social network analysis, are "one-mode" because the type of entity (e.g., people) is found in both the rows and the columns. By extension, the traditional (individuals by variables) data matrix has two-modes (individuals in the rows and variables in the columns). However, so does an *affiliation network* in which individuals report their membership to a set of organizations, or countries are assigned to the international treaties they have ratified.

In this last respect, traditional social science questionnaire data are a subset of the broader category of two-mode data, with the last including some "relational" (network) data types that have sometimes been contrasted to "variable-" based data (Borgatti & Everett, 1997). Another increasingly prominent two-mode data type is the document-term matrices from natural language processing and statistical analysis of textual data in computational linguistics. From this perspective, CA belongs to the broader field of *matrix factorization techniques* allowing for the discovery of structure in two-mode data; Ludovic Lebart, A. Salem, and L. Berry (1997) introduce CA and MCA as a text analysis method.

There are other characterizations of CA in the literature that will not be discussed here, mostly because some scholars see them as less useful a way to consider the utility of the method. For instance, a popular way to characterize CA in the literature is as an *exploratory* technique, presumably distinct from *confirmatory* techniques oriented to hypothesis testing. Related to this distinction, some see CA as mainly a *descriptive* technique distinct from those having a more solid basis in statistical theory or probability.

In terms of the last issue, as Brigitte Le Roux and Henry Rouanet (2004) point out, all data-analytic techniques have elements of description, and therefore, there is no such thing as a "nondescriptive" statistical technique. Furthermore, CA can be given meaningful mathematical characterizations in terms of linear algebra and geometry. Thus, the method has as much of a mathematical basis as other data analysis techniques that are based on mathematical probability theory. Because of this, as Zvi Gilula and Shelby J. Haberman (1986), Leo A. Goodman (1996), and Albert Gifi (1990) have argued from different perspectives, CA can be adapted to the traditional framework of "confirmatory" data analysis as a probabilistic model, which involves the setting of parameter restrictions and statistical criteria for the choice of a number of reduced dimensions to account for the data (discarding the rest as noise). While these "statistical" variants are important developments in the CA literature, they will not be discussed further in this entry.

Basic Concepts of CA

The prototypical data source for CA is the contingency table cross-classifying two categorical variables with each cell containing the relative frequencies with which two categories of each variable co-occur among individuals in the sample. This type of data is ubiquitous in social science research. An example is shown in Table 1. However, frequency data are not the only type of data that can be handled by CA, as the method can be extended to the analysis of two-way tables featuring any type of meaningful non-negative integer entries—including data on percentages or rates, where some quantity is expressed as a fraction of some total that sums to 100, with the only restriction being that the data be expressed on the same scale (Greenacre, 2017).

One of the main goals of CA is to gain insight into how the categories of the cross-classified variables relate to one another. The analyst may be interested in the row categories or the column categories exclusively (*asymmetric* analysis) or the simultaneous relationship between row and column categories (*symmetric* analysis). Proceeding from CA to the visualization and interpretation stage differs depending on what approach is taken.

Table 1. Cross-classification of education by taste for country music.

	Country Music Taste			**Row Total**
	Dislike	**Mixed Feelings**	**Like**	
Education				
<High school	16	18	45	79
High school	102	88	231	421
Some college	109	133	282	524
Associate’s	42	60	107	209
Bachelor’s	171	161	258	590
Masters	69	58	79	206
PhD/professional	25	20	26	71
Column total	534	538	1,028	2,100

There are three fundamental concepts in CA: row and column *profiles,* row and column *masses*, and the χ^2 *distance.* Other concepts, such as *inertia*, are defined in terms of these more basic ones.

Profiles

We can transform the rows of Table 1 by taking the observed frequency in each cell and dividing it by the corresponding row total. This would yield the data shown in Table 2, telling us that while about 55% of people who have a high school degree like country music, only 37% of people with a PhD or professional degree feel the same. Thus, differences between educational groups concerning country music taste can be characterized by their *row profiles*—contrasting, say, those with some college education (0.21, 0.25, 0.54) against those with a master's degree (0.34, 0.28, 0.38). The same could be done with each modality of country music taste concerning education if we calculated the column profiles (see Table 3). In this respect, row or column profiles can be thought of as a vector specifying the position of a given row or column category in an *n*-dimensional space, where *n* is the length of the row or column profile vector.

Table 2. Row profiles of Table 1 data.

	Country Music Taste		
	Dislike	**Mixed feelings**	**Like**
Education			
<High school	0.20	0.23	0.57
High school	0.24	0.21	0.55
Some college	0.21	0.25	0.54
Associate's	0.20	0.29	0.51
Bachelor's	0.29	0.27	0.44
Master's	0.34	0.28	0.38
PhD/professional	0.35	0.28	0.37
Average row profile	0.25	0.26	0.49

In addition to making relative comparisons of row and column profiles against one another, we may be interested in the relative contrast between a given category profile and the average profile (also referred to as the *centroid*). For instance, in Table 2, the average row profile is (0.25, 0.26, 0.49). From this, we can surmise that people with high levels of education (e.g., PhD/professional) are more likely to dislike country than expected. This can be done by comparing the proportion of people who have a PhD who claim to dislike country with the corresponding entry in the average profile and finding that the ratio exceeds 1.0 (0.35/0.25 = 1.4).

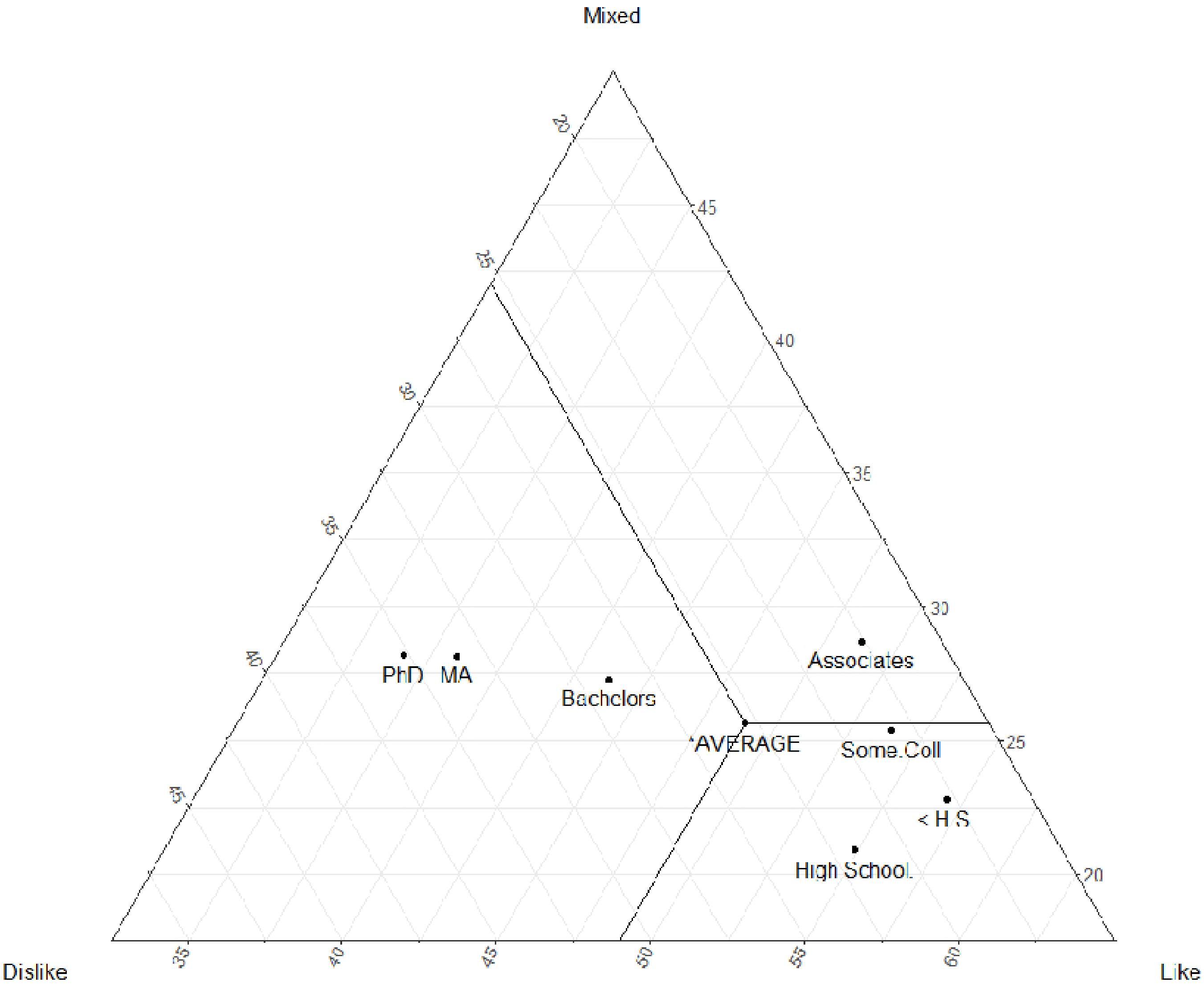
Table 3. Column profiles of Table 1 data.

	Country Music Taste			**Average Column Profile**
	Dislike	**Mixed Feelings**	**Like**	
Education				
<High school	0.03	0.03	0.04	0.04
High school	0.19	0.16	0.22	0.20
Some college	0.20	0.25	0.27	0.25
Associate’s	0.08	0.11	0.10	0.10
Bachelor’s	0.32	0.30	0.25	0.28
Master’s	0.13	0.11	0.08	0.10
PhD/professional	0.05	0.04	0.03	0.03

Masses

The average (row and column) profiles play an important role in the geometric interpretation of CA. Of particular importance is the number of individuals falling in any given row or column category as a proportion of the total sample. These are referred to as (row or column) masses and are indicative of which row and column categories contain a larger (or smaller) proportion of individuals in the sample. The full vector of row and column masses for all categories constitutes the *average row and column profile*, respectively. Thus, we would say (looking at Table 3) that the row mass corresponding to the education category “High School Degree” is 0.20, which is about 6 times as big as that corresponding to PhD and professional degree holders (0.03).

Figure 1. Ternary plot of the row profiles shown in Table 2.



One way to visualize the relationships between profiles in Table 2 is via the use of a *ternary plot*, such as the one shown in Figure 1. Here, each education group is a point in a three-dimensional space defined by their respective profiles. The average profile is highlighted using lines connecting each of the profile values and the associated point in the plot. The plot makes clear what we already suspected from looking at the table: high education groups are closer to the dislike pole, while low education groups are closer to the like pole. Finally, we can also see that some education groups (such as PhD and professional degree holders) are relatively distant because their profiles deviate from the average, while others, such as people with a bachelor’s degree are closer to the average profile for the sample.

As noted by Greenacre (2017), the position of each of the education categories in the ternary plot space can be interpreted as a *weighted average* of their “attraction” to each one of the three vertices in the triangle (Dislike, Mixed Feelings, and Likes), where the corresponding row profile gives the weights. Thus, the row profile for people with a high school degree (0.24, 0.21, 0.55) is “tilted” towards the Like pole (hence the positioning of the point corresponding to that category closer to this vertex), and in this way this category contrasts to higher education groups such as master’s degree holders who are less tilted towards Liking and closer to Mixed Feelings and Disliking.

In this way, the traditional conceptions of (bivariate) “association” and deviation from independence (from the associational analysis contingency tables) are defined geometrically in CA, in terms of row and column category profile distances from the corresponding row or column average. Thus, if it were the case that all education categories fell near the point corresponding to the average profile in Figure 1, we would say that education is not associated with a taste for country music.

The fact that some of the profiles are not at the same position as the average (and that they arrange themselves into an intuitive pattern) reveals that there exists an association between the two categorical variables. Also, we can guess which category profiles contribute more to this association, and which ones contribute less. As a rule, the more distant a given category profile is from the average profile, the more it contributes to the association, while the closer a given category profile is from the corresponding average, the less it contributes to the association.

χ2 Distances

Formally, in CA, distances between each row and column profile and the corresponding average profile are measured using *the χ^2 distance*, not the raw Euclidean distances as displayed in Figure 1. The χ^2 distance can be thought of as a *weighted* version of the Euclidean distance between row and column profiles and the corresponding average profile in which the weights are the corresponding row and column masses.

In the case of a particular row profile, the χ^2 distance is given by:

(1)

$$\mathrm{X}^2_{\mathrm{dist}_i} = \sqrt{\sum_j \frac{\left(a_{ij} - r_j\right)^2}{r_j}}$$

where a_{ij} is the j^{th} element of the i^{th} row profile, and r_j is the corresponding element of the average row profile vector. Thus, the distance between the PhD/professional row profile and the average row profile would be:

$$\mathrm{X}^2_{\mathrm{dist}_{(\mathrm{PhD})}} = \sqrt{\frac{(0.35 - 0.25)^2}{0.25} + \frac{(0.28 - 0.26)^2}{0.26} + \frac{(0.37 - 0.49)^2}{0.49}}$$

which equals 0.26. The larger this number, the bigger the distance between each row profile and the average.

Note that we can also compute the weighted χ^2 distance between any pair of row (or column) profiles. This is given by:

(2)

$$\mathrm{X}^2_{\mathrm{dist}_{ii'}} = \sqrt{\sum_j \frac{\left(a_{ij} - a_{i'j}\right)^2}{r_j}}$$

where everything is as before. This quantity lies at the center of CA as a geometrically motivated data visualization method, in that plots based on CA are designed to represent the χ^2 distances between row (and column) profiles as well as between row and column profiles and their respective average profile. Thus, distances between points representing the row or column categories are formally defined, but distances between a particular row and a particular column category are not.

In this respect, CA can be thought of as a way to link the formal geometry of χ^2 distances to a data visualization and dimension-reduction approach, such that we can preserve the structure of associations—encoded in the interprofile distances—between the row and column categories. CA also preserves and visualizes the association between each row and column category and the corresponding average profiles. Categories can then be visualized in this space, with distances between rows and/or between column categories having a clear interpretation in terms of profile (dis)similarity and difference from the average and one another.

A critical property of χ^2 distances in CA is that of *distributional equivalence*. This means that if two rows (or columns) with identical profiles were to be merged (e.g., by adding up their respective frequencies), then this would not affect the interprofile distances of the other row (or column) points.

Inertia

We can sum each of the (squared) row profile distances from the average profile (weighted by the corresponding mass of each row category) to get an overall sense of the total amount of association in the table. This is called the *inertia*, and it is given, in the case of row profiles, by:

(3)

$$I = \sum_{i} \omega_i \left(\sqrt{ \sum_{j} \frac{\left(a_{ij} - r_j\right)^2}{r_j} } \right)^2$$

where everything else is as in Equation 1, and w_i is the mass corresponding to the ith row (education) category (shown in the last column of Table 3). The inertia is equivalent to the “mean square contingency” coefficient (or its square usually referred to as the ϕ^2 coefficient) in log-linear modeling and summarizes how highly associated the row and column variables are.

From Equation 3, it is evident that the inertia will be higher when more row profiles deviate from the average (as given by the χ^2 distances). In this last sense, the inertia gives us a sense of how dispersed the row (or column) points are around their expected values. The total inertia in the table is fully decomposable into the share attributable to each row. Finally, an essential property of CA is the duality (in terms of all quantities of interest) between analyses performed on the rows and the columns. Thus, the inertia in the table would be the same if we calculated it using the column profiles.

Computation

Computation of CA is relatively straightforward as it relies on the same matrix factorization (sometimes referred to as spectral or eigen decomposition) techniques behind principal components and canonical correlation analysis. The factorization is performed on a transformed version of the original table of counts (or non-negative integers).

Let **N** be such a table with I rows and J columns and entries n_{ij}. The first step consists of creating a new matrix, referred to as the *correspondence matrix* (**P**), in which each entry is equal to the original data divided by the sum of all of the cells, or the table total (n).

(4)

$$\boldsymbol{P} = \boldsymbol{N}\frac{1}{n}$$

The row ($\mathbf{r}_i$) and column ($\mathbf{c}_j$) sums of the correspondence matrix are vectors containing the row and column masses. If every row and column profile was equal to the average profile (implying that there is no association between the row and column variable), then each cell in the correspondence matrix would be equal to the product corresponding of the row and column masses ($\mathbf{r}_i\mathbf{c}_j$). Therefore, any association between the variables above and beyond independence is encoded in the *deviations* from this expected value. These can be stored in a new matrix (**S**) where each cell contains the *standardized residuals*. These residuals are computed by finding the deviation of the observed cells in the correspondence matrix ($\mathbf{P}_{ij}$) from the expected value under independence, and then weighting each residual by the square root of the product of the row and column masses ($\mathbf{S}_{ij} = (\mathbf{P}_{ij} - \mathbf{r}_i\mathbf{c}_j)/\mathbf{r}_i\mathbf{c}_j^{1/2}$). The standardized residuals are equivalent to the square root of that cell's contribution to the table's overall χ^2 statistic. The singular value decomposition (SVD) of S gives the CA scores for each row and column categories across all possible dimensions:

(5)

$$\boldsymbol{S} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$$

Where **Σ** is an $s \times s$ diagonal matrix of singular values arranged in descending order ($\sigma_1 \geq \sigma_2 \geq \ldots \sigma_s$) and s is the rank of the original **S** matrix, which is given by $\min(I, J) - 1$ and is also the maximum possible value of the inertia in the table; **U** is a matrix of *right singular vectors* of dimensions $I \times s$ and **V** is a matrix of *left singular vectors* of dimensions $J \times s$. Both **U** and **V** are orthonormal, implying that $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$ (the identity matrix). A revealing property in CA is that the sum of the eigenvalues (e.g., the squared singular values) across all dimensions s is equivalent to the total inertia of the Table 1 defined earlier. This means that the first dimension accounts for the bulk of the variation in the table, followed by the second one, and so on.

All of the quantities required to project the original data into a low-dimensional space allowing for geometric interpretation can be recovered from the SVD of **S.** First, there are the *standard row coordinates* for each dimension s, given by $\mathbf{a}_{is}= \mathbf{u}_{is}/\mathbf{r}_i^{1/2}$ along with the *standard column coordinates* given by $\mathbf{b}_{js} = \mathbf{v}_{js}/\mathbf{c}_j^{1/2}$. Second, there are the *principal row coordinates* for each dimension s, given by $\mathbf{f}_{is} = \mathbf{u}_{is}\boldsymbol{\sigma}_s/\mathbf{r}_i^{1/2}$, and the *principal column coordinates* given by $\mathbf{g}_{js} = \mathbf{v}_{js}\boldsymbol{\sigma}_s/\mathbf{c}_j^{1/2}$. Note that the main difference between standard and principal coordinates is that the latter is multiplied by the eigenvalue corresponding to that dimension. Both are weighted by the square root of the corresponding (row or column) masses $\mathbf{r}_i$ and $\mathbf{c}_j$.

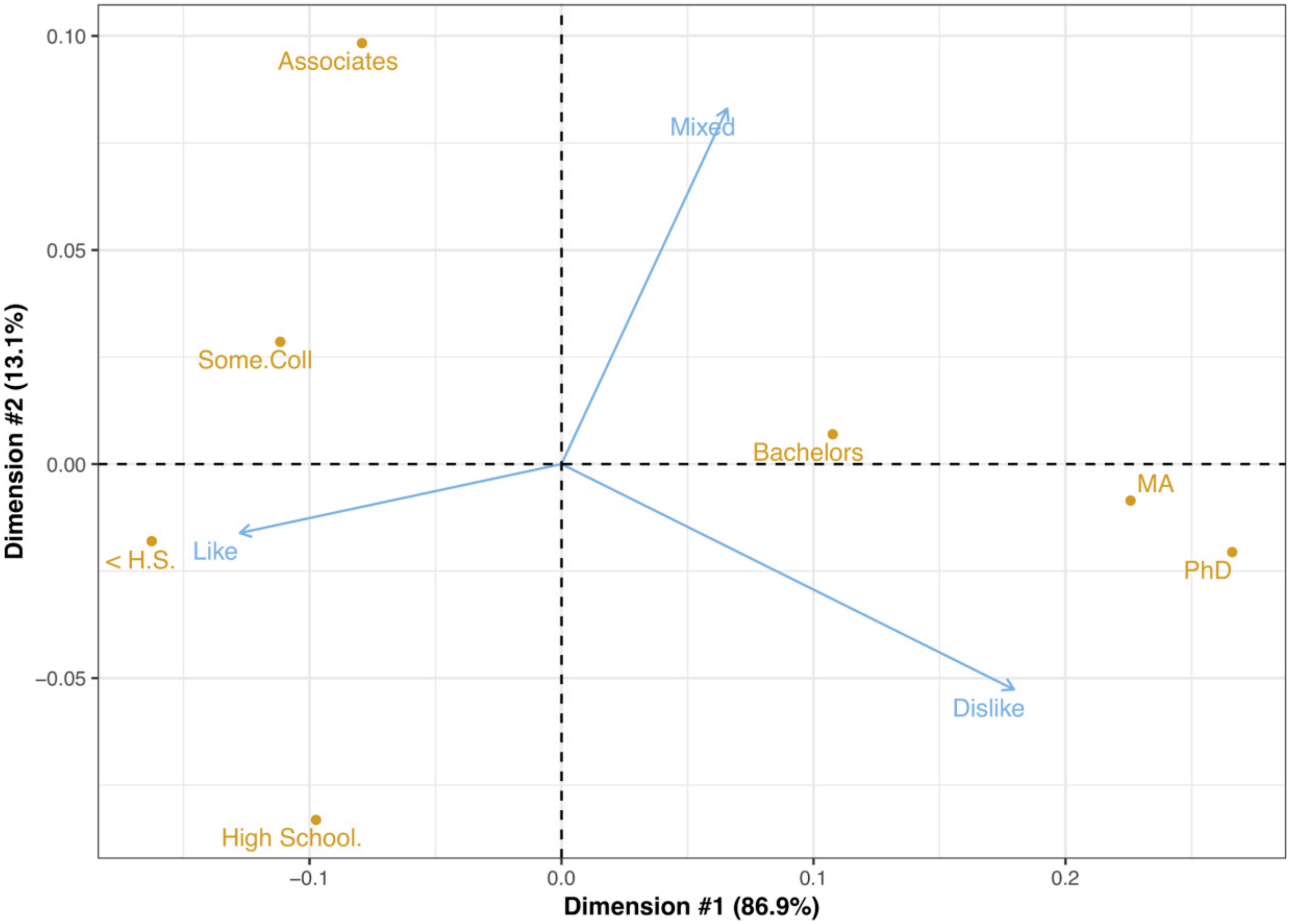
In the *symmetric plot*, row and column categories are plotted in the same space using their respective principal coordinates. In the *asymmetric plot*, as will be seen later, one of the variables is plotted in principal coordinates, while the other is plotted in standard coordinates. Table 4 shows the principal coordinates obtained from a CA of the data in Table 1. Since the rank of the matrix is min(7, 3) − 1 = 2, the data can be described entirely in a two-dimensional space. The eigenvalues corresponding to each dimension are shown in the bottom row of the table. The resulting CA plot is shown in Figure 2. Traditionally, the first (and most important) dimension is plotted on the horizontal axis and the second most important dimension on the vertical axis. For matrices of larger rank (resulting in more dimensions), the same approach can be followed to plot higher order dimensions (e.g., plotting the second dimension on the horizontal axis and the third one on the vertical).

Table 4. Correspondence analysis principal coordinates from the factorization of Table 1.

Row Cats	**Dim 1**	**Dim 2**	**Column cats**	**Dim 1**	**Dim 2**
<High school	−0.163	−0.018	Dislike	0.180	−0.053
High school	−0.097	−0.083	Mixed	0.066	0.083
Some college	−0.112	0.029	Like	−0.128	−0.016
Associates	−0.079	0.098			
Bachelors	0.108	0.007			
MA	0.226	−0.009			
PhD	0.266	−0.021			
Eigenvalues	0.017	0.003		0.017	0.003

Figure 2 shows the symmetric plot of the CA coordinates, as defined earlier. In this plot, the distances between row and column categories correspond (in this plot exactly, in higher-dimensional tables approximately) to the χ^2 distances between the row (and column) points. Distances between a given row and a column point are not interpretable in the same terms (since profile distances are not defined across rows and columns), although one can use one set of points (e.g., in this case, the column categories) as a substantive reference to interpret the meaning of the location of the other set (in this case the row categories) along each axis. Another key feature of the symmetric plot is that the origin (point 0,0) corresponds to where the average row or column profile would be placed. Thus, distances of each row and column category are interpretable in terms of (approximations to) χ^2 distances of each category from the average profile in each dimension.

Figure 2. Symmetric plot along the two principal axes obtained from a CA of the data shown in Table 1.



In this application, we may be interested in how the row (education) categories are differentiated based on their taste for country music (with the taste categories highlighted with arrows pointing from the origin to their location in the Figure 2 plot). We may have an *ex-ante* hypothesis of “homology” indicating that high education groups should be similar in their tastes, as given by similarities in their row profiles, but different from low education groups, who are also expected to be similar to one another.

As Figure 2 shows, we do find such a differentiation, with high education categories appearing closer to another in the two-dimensional space, and distant from low education categories. As given by the placement of the column categories (and as we already suspected from inspected raw profile differences in the ternary plot), this is because low education groups are substantially more likely to like country music. Hence, the placement of low education categories on the left side of the horizontal dimension (corresponding to “Like”) and the placement of the high education categories on the right side (corresponding to “Dislike”). This Like/Dislike axis accounts for about 87% of the total inertia in the table, indicating that this is the primary source of differentiation across education groups based on their country music taste. The remaining 13% (second axis) appears to separate those educational groups more likely to express “Mixed Feelings” (such as people with an associate’s degree) from the rest.

MCA

MCA is an extension of CA useful for the analysis of tables in which the rows represent individuals and the columns represent variables indexing a set of characteristics of the individuals. The key restriction is that the variables classify individuals into a set of *disjunctive* (mutually exclusive) categories and that the variables be "homogeneous" (e.g., all in the same broad domain, such as demographics, cultural practices, or attitudes in a given domain of social life). In traditional statistical parlance, the first criterion restricts variables considered in MCA to be "categorical" (irrespective of whether they are "nominal" or "ordered").

This is why in the literature MCA is sometimes portrayed as a categorical counterpart to PCA for the statistical analysis of tables of individuals by variables (Hill, 1974; A. Lebart, Morineau, & Warwick, 1984; Le Roux & Rouanet, 2004). This is not entirely inaccurate, as the similarities between MCA and PCA are more than superficial; both methods rely on the SVD of a transformed version of the individuals by variables table and yield a low-dimensional representation of the data in the form of ordered principal components which can then be used for the visualization of their constituent categories as well as the placement of each individual in the same reduced multidimensional space as briefly noted below.

Basic Concepts

We denote the total number of questions in the columns of the original data in Table 2 by Q and a single question as q. Each of the response categories in a question is referred to as a *modality*. We use p_q to denote the qth modality in the pth question, the total number of modalities considered would then be given by:

(6)

$$p = \sum_{q=1}^{Q} p_q$$

Take for instance a data set in which, instead of having a cross-classification of education groups by the taste for a single musical genre as in the CA example, we have reports for each respondent for their taste for *four* musical genres (Classical, Country, Hip-Hop, and Rock). A sample (the first 10 of a total of 2,100 cases) of the data are shown in Table 5. In these data, there are four questions (corresponding to the taste reports for each genre), with a total of 3 modalities (1 = *like*, 2 = *dislike*, and 3 = *mixed feelings*) per question. Thus, $p = 4 \times 3 = 12$. We also have information on the highest degree completed by the respondent, which will enter the analysis as a *supplementary* or *passive* variable (to be defined later).

Table 5. First 10 rows of a data set of the musical tastes of a representative sample of Americans for four musical genres (SSI, 2012, $N = 2{,}100$).

Id	Degree	Classical	Country	Hip-Hop	Rock
2	College	3	1	1	3
5	College	2	2	2	2
9	Master’s	3	1	1	3
13	Some college	3	3	2	3
14	Master’s	1	3	1	2
17	High school	3	3	1	3
18	College	1	1	3	1
19	College	3	3	3	2
22	Master’s	3	3	2	2
23	Some college	3	3	3	1

Note. 1 = *like*; 2 = *dislike*; 3 = *mixed feelings*.

The first step in MCA is to transform the original data matrix (also referred to as the *condensed coding matrix*) into an *indicator matrix* **Z**. Part of the indicator matrix corresponding to Table 3 is shown in Table 6. This matrix has the same rows as those in the original (the number of individuals, in this case, N = 2,100), but has a column for each possible modality; thus, **Z** is 2,100 × 12. For each of the three columns corresponding to each question, 1 is placed in the column corresponding to the modality chosen by the respondent and 0 in the modalities not chosen. Thus, the row corresponding to each individual in **Z** is a vector of zeros and ones. The row sum for each individual ($\mathbf{Z}_{i+}$) is fixed and equal the number of questions (Q). The column sums ($\mathbf{Z}_{+j}$) correspond to the number of individuals who chose that modality for a particular question. The row masses are also fixed and are equal to Q/p (the number of questions divided by the total number of modalities), while the column masses are equivalent to the proportion of individuals who choose modality p_q (e.g., the proportion of people who like hip-hop).

The **Z** matrix can thus be thought of as a concatenation of Q submatrices, each corresponding to a question:

(7)

$$\boldsymbol{Z} = \left[\boldsymbol{Z}_1, \boldsymbol{Z}_2, \quad \dots, \boldsymbol{Z}_Q\right]$$

Table 6. First 10 rows of the indicator matrix, corresponding to the data shown in Table 3.

Classical			Country			Hip-Hop			Rock		
1	2	3	1	2	3	1	2	3	1	2	3
0	0	1	1	0	0	1	0	0	0	0	1
0	1	0	0	1	0	0	1	0	0	1	0
0	0	1	1	0	0	1	0	0	0	0	1
0	0	1	0	0	1	0	1	0	0	0	1
1	0	0	0	0	1	1	0	0	0	1	0
0	0	1	0	0	1	1	0	0	0	0	1
1	0	0	1	0	0	0	0	1	1	0	0
0	0	1	0	0	1	0	0	1	0	1	0
0	0	1	0	0	1	0	1	0	0	1	0
0	0	1	0	0	1	0	0	1	1	0	0

Note. 1 = "*like*"; 2 = "*dislike*"; 3 = "*mixed fneelings*."

Each of the $\mathbf{Z}_q$ submatrices is such that the i^{th} row contains $p_q - 1$ zero entries and exactly one cell with a value of 1, corresponding to the chosen modality for that question for the i^{th} individual.

An important feature of the **Z** matrix is that we can derive a square matrix containing all possible cross-tabulations of every variable included in the analysis by multiplying the transpose of **Z** times itself:

(8)

$$\boldsymbol{B} = \left[Z^T \times Z\right]$$

Matrix **B**, of dimensions $p \times p$ is called the *Burt table* (after Burt, 1950) and is shown as Table 7. The Burt table is a "block" of subtables containing all the two-way cross-classifications (contingency tables) of each of the variables (q, q') in the original data. The block diagonals of the Burt matrix contain the cross-classification of each variable against itself, including diagonal entries. Note that the trace of **B** (e.g., the vector containing only diagonal entries) is thus equivalent to the column sums of **Z**.

Table 7. Burt table of musical taste for four musical genres (SSI, 2012, N = 2,100).

	Classical_Like	**Classical_Dislike**	**Classical_Mixed**	**Country_Like**	**Country_Dislike**	**Country_Mixed**	**Hip-Hop_Li**
Classical_Like	292	0	0	110	61	121	121
Classical_Dislike	0	638	0	123	184	331	242
Classical_Mixed	0	0	1,170	301	293	576	639
Country_Like	110	123	301	534	0	0	267
Country_Dislike	61	184	293	0	538	0	241
Country_Mixed	121	331	576	0	0	1,028	494
Hip-Hop_Like	121	242	639	267	241	494	1,002
Hip-Hop_Dislike	51	177	243	81	147	243	0
Hip-Hop_Mixed	120	219	288	186	150	291	0
Rock_Like	82	93	228	152	96	155	261
Rock_Dislike	75	217	324	114	229	273	271
Rock_Mixed	135	328	618	268	213	600	470

The simplest way to think about MCA is as a simple CA of either the indicator (**Z**) or Burt matrices (**B**). SVD of the residualized **Z** or **B** matrices will return the same singular vectors. The standard coordinates of the columns of **Z** will be the same as the ones corresponding to either the row or columns of **B**, resulting in equivalent geometric interpretations of the relationship between variable modalities, the eigenvalues corresponding to the CA of **B** being the squares of those corresponding to **Z.**

In the case of two variables, a CA of either **Z** or **B** is equivalent to a simple CA of the corresponding contingency table (A. Lebart, Morineau, & Warwick, 1984). Simple CA of either **Z** or **B** also produces scores for each individual (rows in the original data). In the CA of Z, these are obtained directly from the SVD, while in the CA or **B**, these are obtained indirectly, by assigning each individual the average of the standard coordinates corresponding to each of the modalities they chose for each question in each dimension. This is equivalent to projecting each individual as a “passive” variable in the space of modalities obtained from the CA of the Burt table.

The rank of the **Z** matrix determines the maximum number of dimensions (and corresponding eigenvalues) required to describe the inertia in the **Z** or **B** matrices. This is given by:

(9)

$p - (Q - 1)$
Which in our example is 12 − (4 −1) = 9.

As before, the geometric interpretation of basic concepts remains in MCA, with some modifications. In the indicator matrix, all rows masses are fixed to a single value be equivalent to p/n, which in our example is 4/2,100 = 0.002. Column masses, on the other hand, correspond to the proportion of respondents who chose modality p_q. Because the rows of the indicator matrix are binary 0/1 vectors, the corresponding profiles can contain one of two values: either 0 or $1/n$. Thus, χ^2 distances between row profiles computed as in Equation 2, which essentially counts the (squared) differences in the nonzero ($1/n$) entries for each row vector for pairs of individuals weighted by the column mass corresponding to each modality. This makes mismatches corresponding to rarely chosen modalities count more in determining interindividual distances than mismatches corresponding to “common” categories. Geometric interpretation of χ^2 distances between columns is less straightforward because each nonzero entry in the column of the indicator matrix is weighted by a different factor (the corresponding column masses).

However, this issue can be resolved when we note the equivalence between the indicator and Burt matrices. In **B**, χ^2 distances between pairs of columns (or rows) track the co-occurrence distribution concerning other modalities. Thus, pairs of modalities with distributions similar to other modalities will be closer to one another, with “matches” to rarer modalities having greater weight in determining the distance.

Figure 3. Eigenvalue scree plot corresponding to MCA Shown in Figure 4.

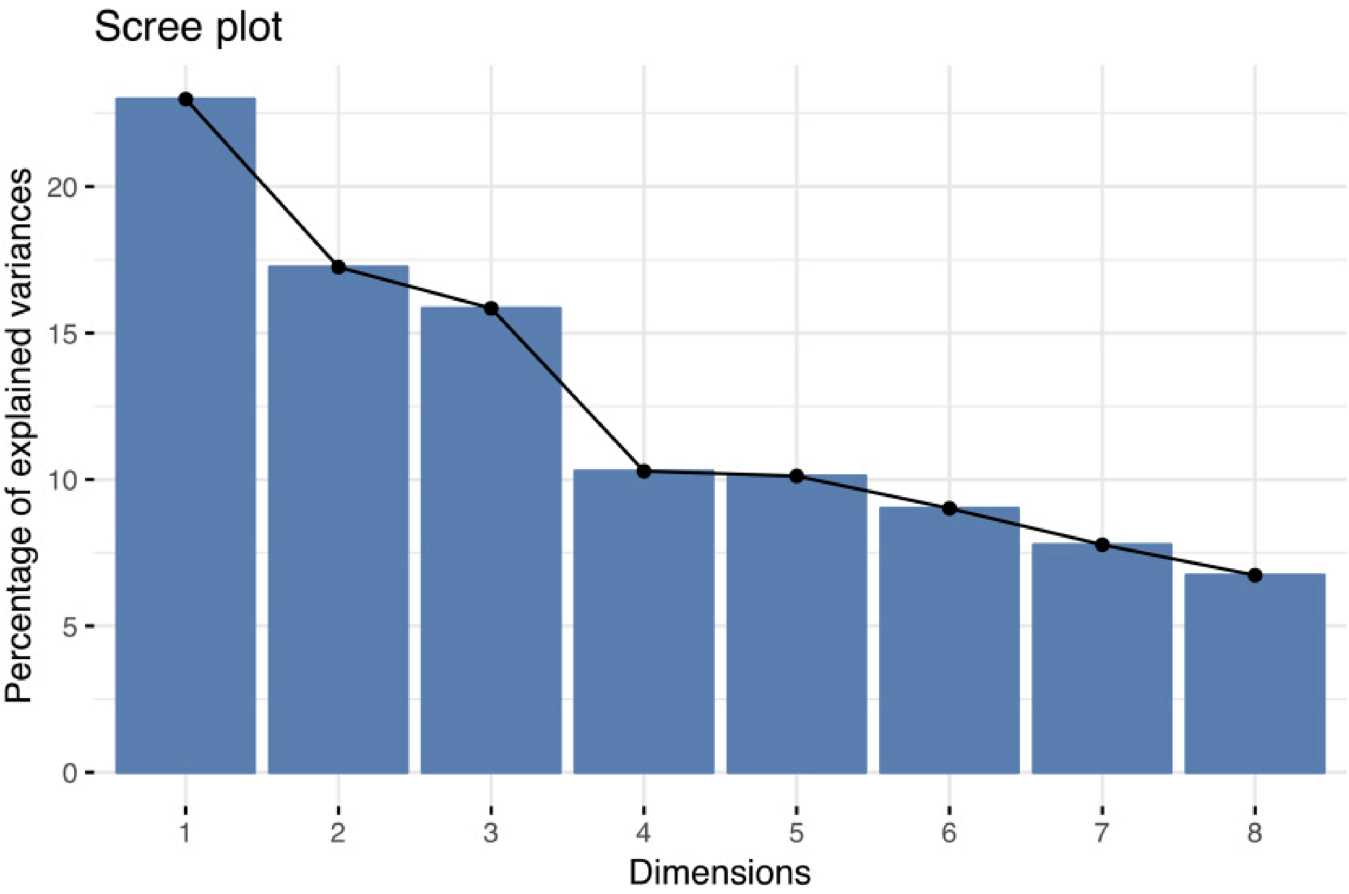
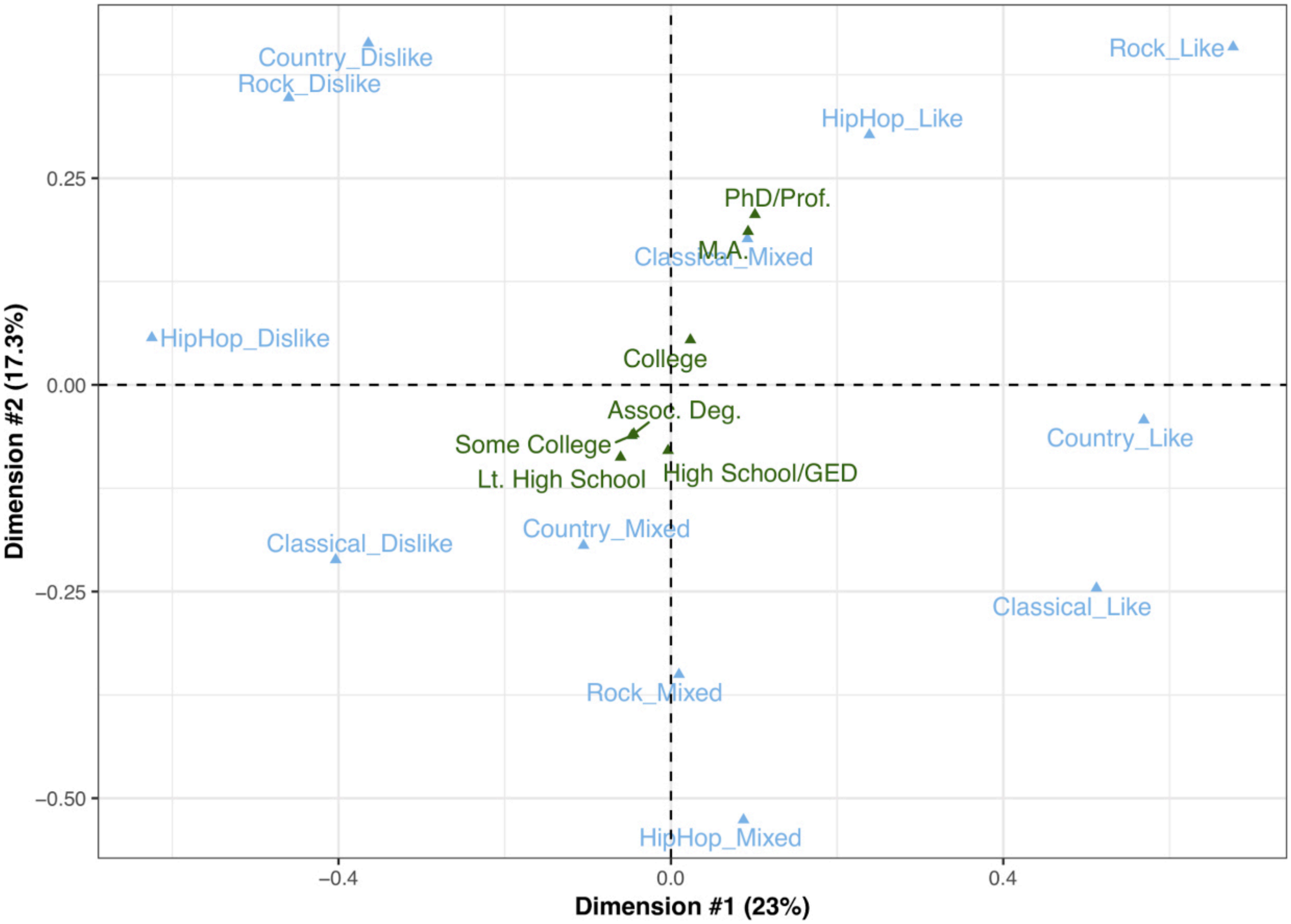


Figure 4. Symmetric plot along the two principal axes obtained from an MCA of the data shown in Table 6.



Performing a CA of the Burt table reveals that eight dimensions are sufficient to explain the inertia in the original table. The 12 taste modalities included in the analysis are plotted in Figure 3 along the first two dimensions. Like in CA, dimensions are ordered in terms of importance (the proportion of inertia they account for). Figure 3 shows a *scree plot* summarizing the proportion of the inertia in the indicator matrix accounted for by each of the eight dimensions. We can see that, jointly, the first two dimensions account for about 40% of the inertia in the original data and the first four dimensions for about 66%. This means that the bulk of the associations in the original data can be reduced to a relatively low dimensional space.

Figure 4 shows the plot of all 12 modalities in principal coordinates, as well as the education categories projected onto the same space as *passive* or *supplementary* variables. Supplementary variables do not directly contribute to the estimation of the row or column scores. Instead, the modalities corresponding to supplementary variables are projected into the space induced by the other *active* variables, by computing the average score on each dimension for each group of individuals that falls in a given passive modality. For instance, in Figure 4, people with a PhD/professional degree have positive average scores in both the first and second dimensions and thus appear in the upper right-hand quadrant of the biplot.

The biplot of active variables is interpreted as before. The first (horizontal) axis is a like/dislike dimension, opposing positive dispositions towards the four musical genres (on the right) against negative dispositions. The second (vertical) dimensions encode an opposition between those who have “mixed feelings” for each musical genre (especially Rock and Hip-Hop), at the bottom against those who make an assured judgment of taste one way or the other (at the top).

From MCA to Structured Data Analysis

As noted earlier, MCA also produces scores for each row of the indicator matrix which, in the case of standard social science data sets, corresponding to the individuals included in the survey. This means that individuals, just like the variable modalities, can also be projected into a low dimensional space, and their relations in this "social space" can be examined in a systematic way. Extensions of MCA and other techniques such as analysis of variance and hierarchical clustering to the analysis of this "cloud individuals" leads to a general approach to the statistical analysis of multivariate data that has been referred to as *structured data analysis* (SDA).

The SDA approach has roots going all the way back to the seminal work of the original founder of the "French" approach to CA and MCA, Jean-Paul Benzécri, and has been developed subsequently by his students (Le Roux & Rouanet, 2004). SDA merges the geometric emphasis on visualization reviewed here with techniques aimed at classification (assigning individuals into clusters based on their scores on a subset of the MCA dimensions) and statistical inference (e.g., ascertaining whether individuals assigned to clusters differ systematically on certain characteristics), both classical and nonparametric. It is beyond the scope of this entry to cover the SDA approach in detail; Le Roux & Rouanet (2004) provide a detailed technical introduction from a classic "French" perspective based on the Benzécri approach, and Johs Hjellbrekke (2018) provides a relatively nontechnical introduction to SDA aimed at practicing social scientists.

Extensions of CA and MCA to Other Data Types

CA of Relational Data

Relational data are typically presented in graph form, where vertices represent actors and edges (or "ties") between any *i* and *j* elements represent a type of relation between actors (e.g., friendship). Simple graph representations, however, do not always scale up well with large and/or very "dense" networks. Further, without a specified spacing algorithm, the placement of vertices and edges in the plotting space is arbitrary. CA, then, can be a valuable tool for visualizing relational data in a way that preserves the main axes of variation in the distribution of ties while reducing the complexity of the network (L. Lebart, 1998; Noma & Smith, 1985; Roberts, 1996).

Relational matrices are often square: that is, the same number of elements are arrayed across the rows and columns. When the elements across the rows and columns are the same entity, a square matrix is called an *adjacency matrix*. When the cells represent the presence or absence of an edge, it is a binary adjacency matrix, otherwise known as a *one-mode network*.

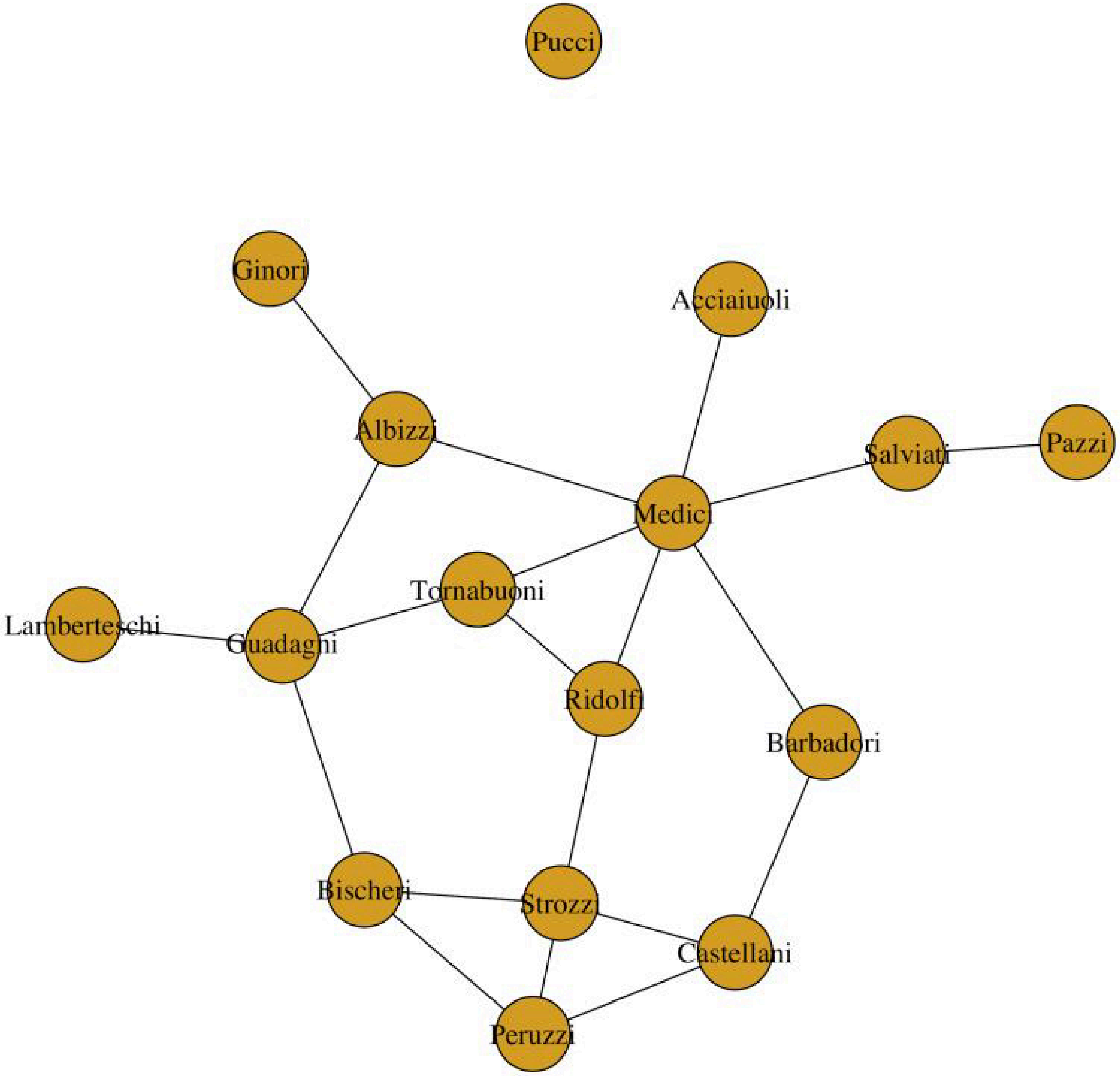
Consider a popular example in the social networks literature: the Florentine Families data, a network of 16 families (vertices), and their interfamily marriages in Renaissance-era Florence, Italy (Padgett & Ansell, 1993). The first six rows and six columns of the adjacency matrix **M** is given in Table 8, where 1 indicates marriage, and 0 indicates a lack of a marriage. The matrix is symmetric: that is, $\boldsymbol{M}^T = \boldsymbol{M}$, so there is no “direction” the marriage.

Table 8. Snippet of Florentine families marriage adjacency matrix.

	Acciaiuoli	**Albizzi**	**Barbadori**	**Bischeri**	**Castellani**	**Ginori**
Acciaiuoli	0	0	0	0	0	0
Albizzi	0	0	0	0	0	1
Barbadori	0	0	0	0	1	0
Bischeri	0	0	0	0	0	0
Castellani	0	0	1	0	0	0
Ginori	0	1	0	0	0	0

This matrix can be represented as a *node-link plot*, where an edge for each cell containing a nonzero entry in the matrix. This is shown in Figure 5. This plot shows useful information: For example, the Medici family has the most interfamily marriages, and the Pucci family is the only isolate with no interfamily marriages. MCA, however, can be used to bring less apparent properties of this social structure to light.

Figure 5. Graph representation of Florentine families marriage data.

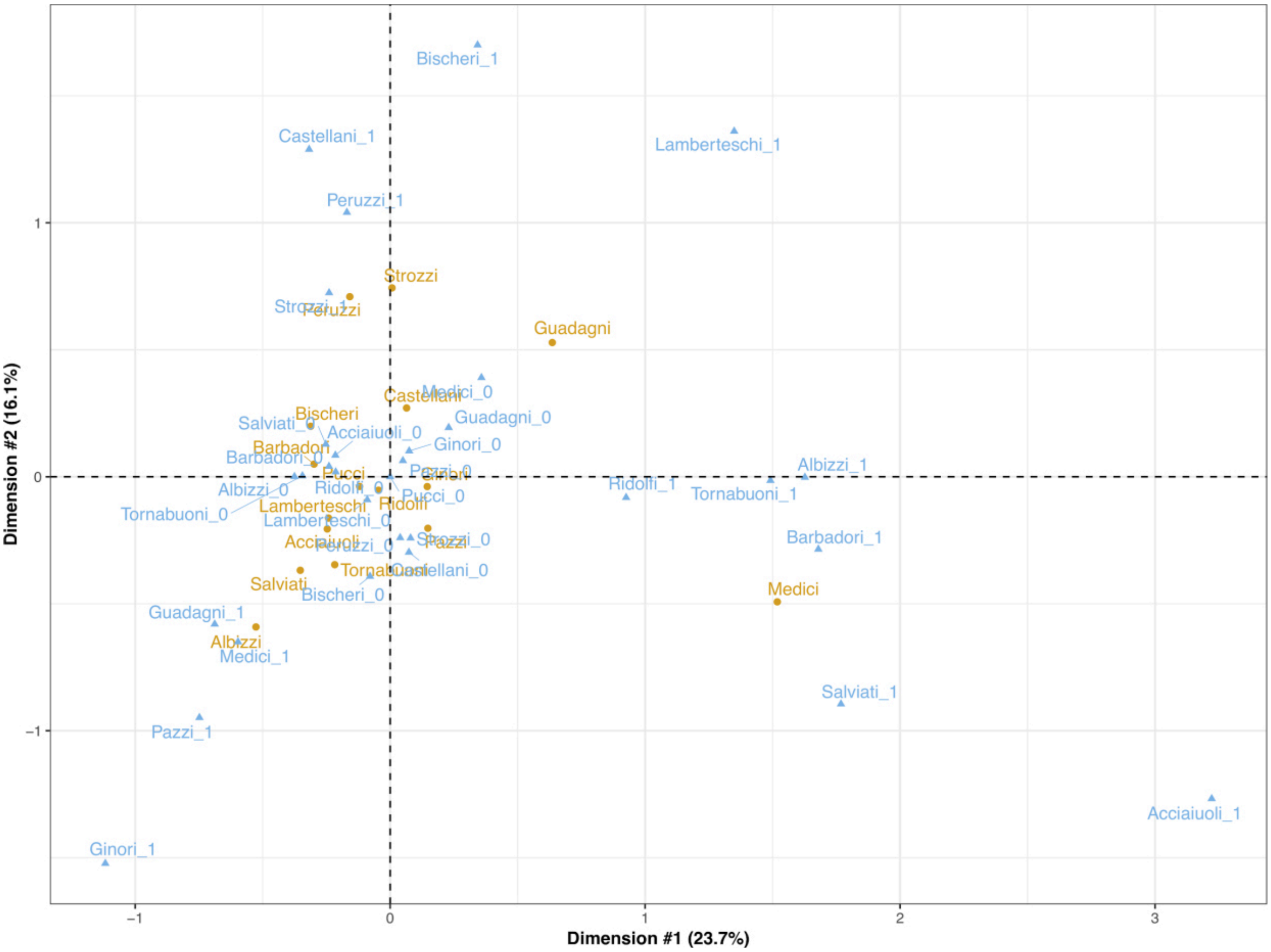


The first step is to turn the adjacency matrix into an indicator matrix, like the one in Table 8. MCA of the indicator matrix yields (symmetrical) row and column scores, which can be represented as a biplot, like in Figure 6. The first dimension of the biplot shows that the associations with the Medici family account for most of the deviations from independence in the matrix. All of the families (in blue) which are connected to the Medicis through marriage are positioned closely to the Medici row (in orange), with the Acciaiuoli family occupying the upper limit of the axis because of its lack of ties to any *other* family.

Table 9. Snippet of the indicator matrix of Florentine families data.

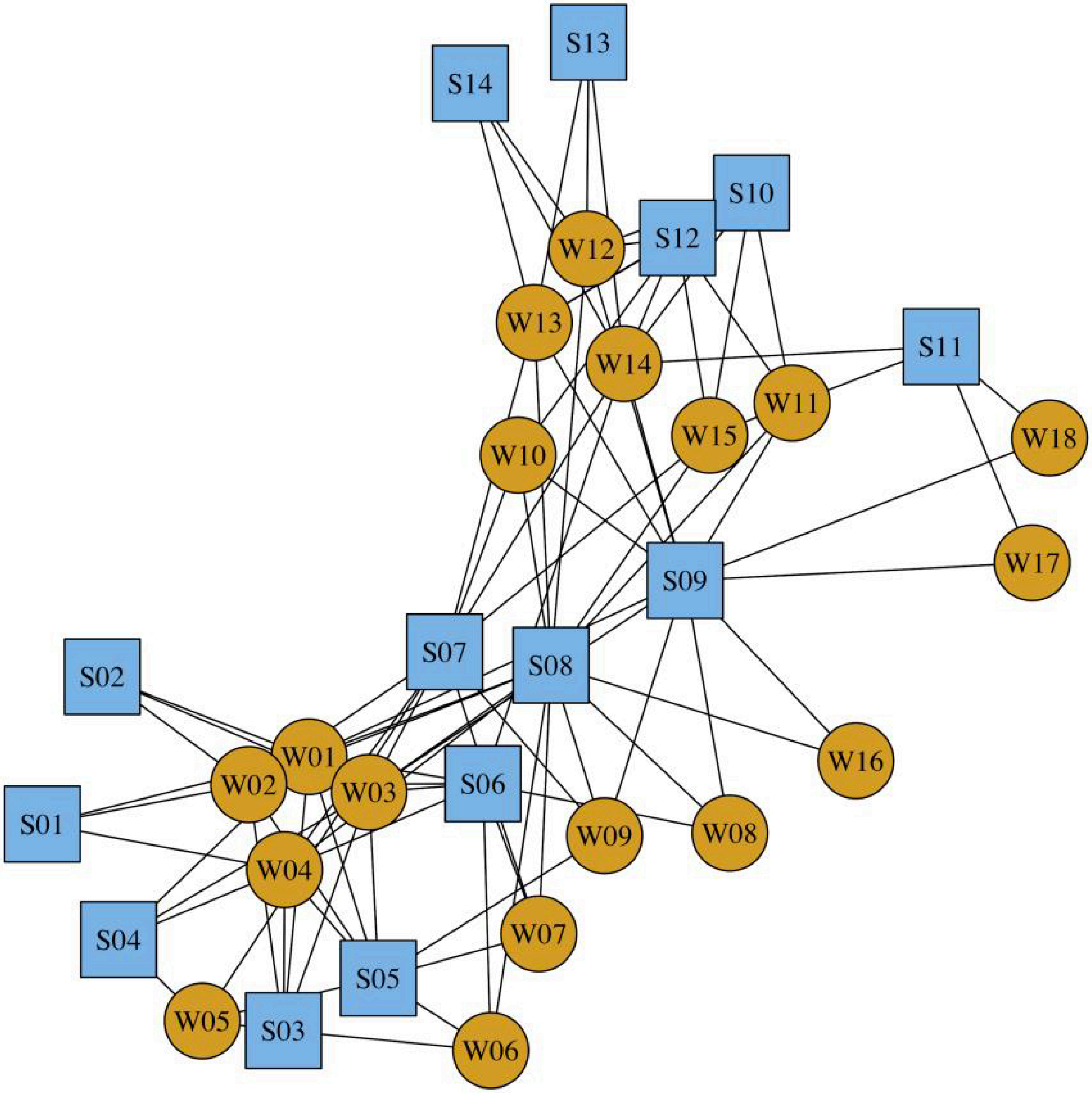
	Acciaiuoli		**Albizzi**		**Barbadori**		**Bischeri**		**Castellani**		**Ginori**	
No	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	
Acciaiuoli	1	0	1	0	1	0	1	0	1	0	1	0
Albizzi	1	0	1	0	1	0	1	0	1	0	0	1
Barbadori	1	0	1	0	1	0	1	0	0	1	1	0
Bischeri	1	0	1	0	1	0	1	0	1	0	1	0
Castellani	1	0	1	0	0	1	1	0	1	0	1	0
Ginori	1	0	0	1	1	0	1	0	1	0	1	0

Figure 6. Symmetric plot along the two principal axes obtained from Florentine families data



MCA has also proven useful for uncovering structure in two-mode networks (Borgatti & Everett, 1997; D’Esposito, De Stefano, & Ragozini, 2014). In a *two-mode network*, the rows and columns represent different entities—for example, the rows might be individuals and the column groups to which they belong (Breiger, 1974). Consider, for example, the Southern Women data—a two-mode relational data set that includes 18 women in a Mississippi town and the 14 social events they attended for nine months in the 1930s (Davis, Gardner, & Gardner, 2009; Freeman, 2003). This is often called an *incidence matrix*, and, in this case, the rows correspond to each of the 18 women and the columns are the 14 social events although one-mode projections are often used to analyze two-mode data—where the row projection of incidence matrix **N** is equal to $\boldsymbol{N}\boldsymbol{N}^{T}$ and the column projection to $\boldsymbol{N}^{T}\boldsymbol{N}$—the matrix can also be represented as a *bipartite graph*, such as the one shown in Figure 7.

Figure 7. Bipartite graph representation of Southern women data.



Note. Blue boxes indicate events (columns); orange circles indicate women (rows).

The two-mode nature of the network and the larger number of vertices and edges make this graph more difficult to eyeball than the Florence family graph in Figure 5. MCA can again be used to decompose the incidence matrix (after first converting it to an indicator matrix) to find the underlying dimensions that maximize the amount of total inertia explained. Figure 8 is a biplot of the first two dimensions from an MCA of the Southern women data.

Figure 8. Symmetric plot along the two principal axes obtained from Southern women data.

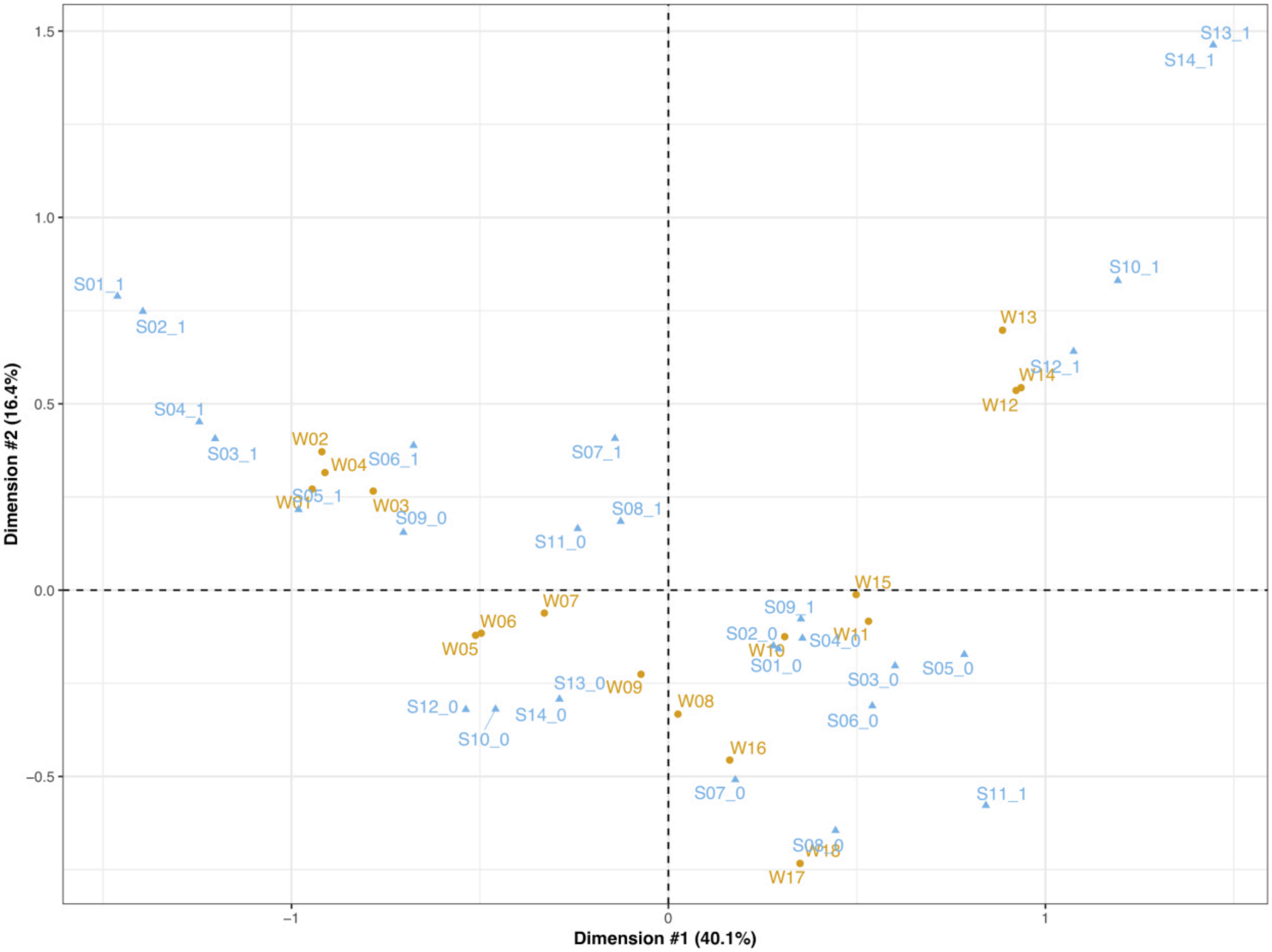


Figure 8 underscores structural patterns in the network not readily discernible in the bipartite graph representation. For instance, since variable categories (the absence or presence of an event) are mapped in the space and not the variable itself, individuals can be described simultaneously by their presence at some events *and absence at others*. W01, W02, W03, and W04 all have similar profiles (meaning that they each have similar distributions of edges across the other 17 women), and we see that these profiles are defined in large part by coattending SO1, SO2, and SO4, but *not* attending S09 and S011.

The fact that MCA positions elements of the same mode closer based on how similar their distributional profiles are across the other array highlights an important feature of CA for analyzing relational data. Social network analysts are often interested in identifying “roles” in a network, where two actors occupy a similar role to the extent that they have the same ties to others (or in a two-mode network, to elements of the other column array). This property is known as *structural equivalence* (White, Boorman, & Breiger, 1976). Unlike a graph-based method such as blockmodeling, however, MCA can be used to locate structurally equivalent actors on more of a “fuzzy” gradient, and along different (and, importantly, orthogonal) axes of variation (D’Esposito, De Stefano, & Ragozini, 2014).

CA of Text Data

Extending the logic of CA to two-mode relational data highlights another domain for CA application: *computational text analysis* (CTA). Scholars from a wide range of disciplines have begun to recognize CA as a useful dimension reduction technique for textual data (Francesco & Roberta, 2019). CA is appropriate since the raw data for CTA are frequently represented in the form of two-mode *document-term matrices* and are therefore amenable to most matrix factorization algorithms (including PCA and *t*-distributed stochastic neighbor embedding—two other popular dimension reduction techniques for document-term matrices; Maaten & Hinton, 2008).

Consider a set of documents (a "corpus") which we wish to represent as a document-term matrix (DTM), **D**. We begin by vectorizing the corpus so that each document, *d*, is represented as a distribution of counts over a vocabulary of words. Consider the following document, d_i (composed of only two sentences): "See spot run. He runs really fast." Now consider another document in the same corpus, $d_{i'}$: "Spot can swim really fast." After standard preprocessing (wherein, e.g., capitalization, punctuation, numbers, and select words are removed, and words are stemmed so that words like "run" and "runs" are recognized as the same token), d_i and $d_{i'}$ could be represented as the row vectors in a DTM as shown in Table 9.

Table 10. Two-document corpus as a document-term matrix.

	Can	Fast	He	Really	Run	See	Spot	Swim
d_i	0	1	1	1	2	1	1	0
$d_{i'}$	1	1	0	1	0	0	1	1

Once the documents are represented in such a way, CA can be applied.

The U.S. State of the Unions (SOTUs) corpus provides a useful example. The data subset used here contains all SOTUs from George W. Bush's inaugural address in 2001 to Donald Trump's speech in 2019 ("The American Presidency Project," 2019). After document preprocessing (including removing terms that fail to show up in at least 90% of the speeches), we get a DTM with 19 speeches and 55,865 words across 3,092 unique terms. Table 10 shows a subset of this DTM across 10 terms.

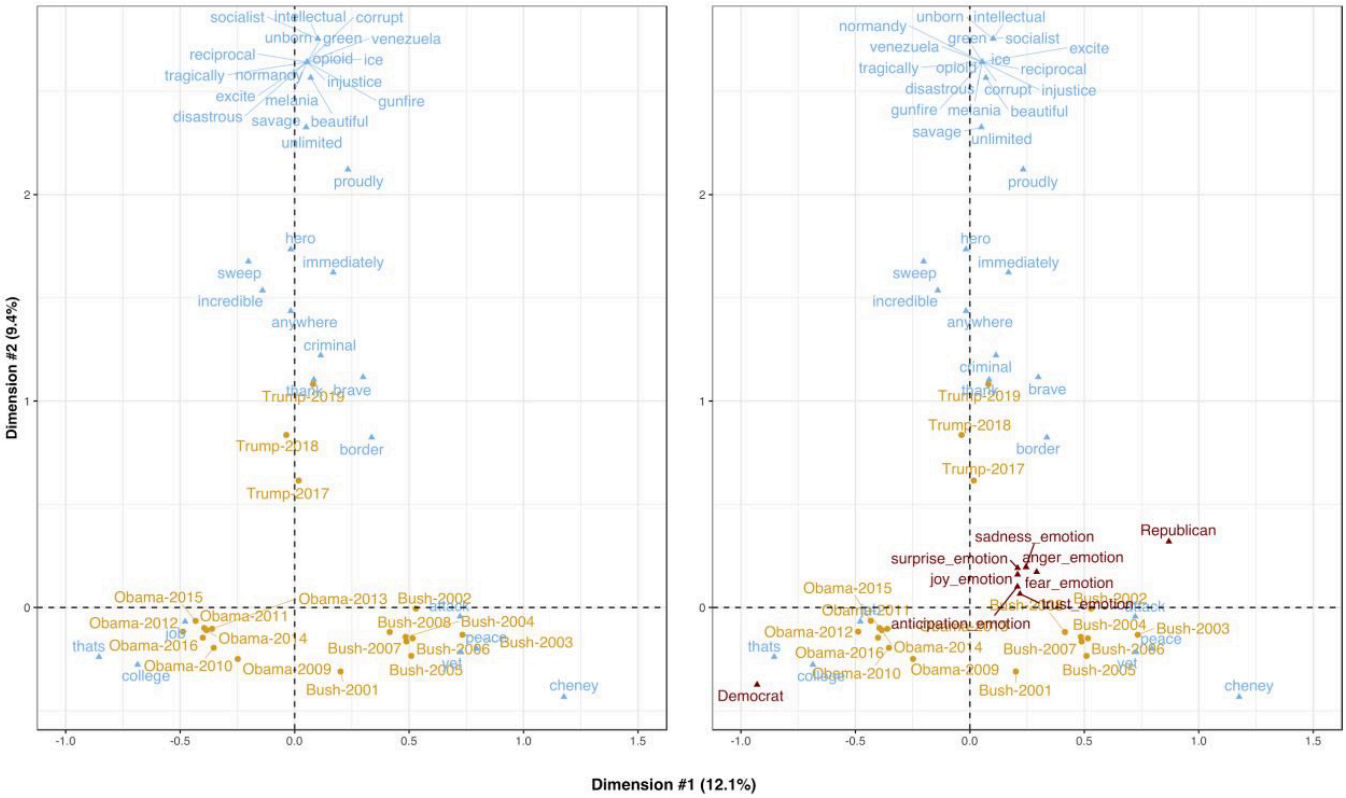
Table 11. Subset of State of the Unions document-term matrix.

	America	American	Can	Good	Make	New	People	Will	Work	Year
Bush-08	30	31	20	22	14	25	32	53	12	35
Obama-09	18	40	30	14	24	20	21	86	14	25

	America	American	Can	Good	Make	New	People	Will	Work	Year
Obama-10	18	46	24	17	30	20	33	62	30	47
Obama-11	18	30	36	35	34	36	33	67	27	39
Obama-12	30	54	26	22	33	27	23	58	27	37
Obama-13	23	28	27	36	33	24	20	58	34	40
Obama-14	28	48	35	30	29	29	24	43	43	37
Obama-15	35	37	24	33	45	42	23	35	34	40
Obama-16	28	37	19	33	33	19	27	27	30	35
Trump-18	22	55	20	12	12	21	33	38	14	30

We can then perform CA on this DTM. As before, we apply the SVD to the DTM—where each cell is transformed to a standardized residual or the square root of that cell's contribution to the DTM's χ^2 statistic—and find the document/term scores that preserve the document–document/term–term χ^2 distances and their respective distances from the average document/term profiles. The underlying dimension can then be projected as a biplot, as in Figure 9. Given the large number of terms (columns) in most DTMs, it is often useful to only show those terms above some threshold: In this case, we show only terms with a squared cosine (a measure of association) above 0.7 across the two dimensions.

Figure 9. Symmetric plot along the two principal axes obtained from a correspondence analysis (CA) of State of the Unions document-term matrix with (left) and without (right) supplementary variables.



In the left panel of Figure 9, speeches (in orange, located in the bottom half of the figure) with similar distributions across the words (i.e., similar profiles) are projected into similar positions on the map. Columns (in blue) with similar distributions across the speeches are grouped in the same manner. The further away a speech of term is from a given axis, the distinctive (i.e., the higher its "loading") it is on that dimension.

We can see, for example, that Dimension #1 is defined by an opposition between Obama and Bush speeches (with Trump not deviating much from the average speech on this dimension). Dimension #2, however, is defined more by a tension between Trump, on the one hand, and Obama and Bush, on the other, with the latter two being close the average speech on this dimension. Similarly, for terms, we see that Dimension #1 opposes talk of "college" and "job" on one end and "attack," "peace," and "Cheney" on the other. Dimension #2 mainly accentuates the distinctiveness of Trump's language, with terms such as "socialist," "unborn," "disastrous," and "Melania." Importantly, the CA decomposition does not formally define the association between the row and column elements in the symmetric biplot; as such, to interpret such row–column associations one should create an asymmetric biplot where the row (document) coordinates are either projected into the column coordinates or the other way around (Kassambara, 2017).

One last point of interest in using CA with textual data is the use of *supplementary variables*. As before, these variables are not part of the matrix factorization, do not contribute to the matrix inertia being decomposed (Sourial et al., 2010), and are instead projected based on their associations with the two latent dimensions plotted. In the case of textual data, these variables can be textual or nontextual and can be continuous or categorical. For instance, perhaps an analyst wants to examine how political party (a categorical variable) and expressed sentiment (continuous variables, generated from a sentiment analysis; Jockers, 2017) are associated with Dimensions #1 and #2 in the left panel of Figure 9. The resulting biplot is shown in the right panel of Figure 9, with the sentiment and party categories are projected (in maroon, toward the center of the plot) onto the same mapping in the left panel.

The party categories are positioned as one might expect: “Democrat” is to the left with Obama speeches, and “Republican” is in between Bush and Trump speeches (although closer to Bush, given Bush’s more traditional conservative discourse). All of the sentiment categories are placed close to the origin, suggesting that emotion distinctions do not radically differentiate the speeches along the first two dimensions. To the extent that they are correlated with the dimensions, though, they are closest to the Bush speeches—not surprising, given the tragedy of September 11, 2001, and ensuing events that occurred during this administration.

Software

CA and MCA routines are readily available across major statistical software platforms. In R, there is ca package (Nenadic & Greenacre, 2007), MASS (Venables & Ripley, 2013), ade4 (Dray, Dufour, and Others, 2007), and the FactoMineR and factoextra packages (Lê et al., 2008; Kassambara, 2017). The FactoMineR and factoextra packages were used to carry out the CA/MCA example analyses in this chapter. Python also has a number of options, including the mca (Safak, 2018) and prince (Halford, 2019) libraries. Stata comes equipped with a ca command for simple CA and a mca command for MCA, SAS with PROC CORRESP, and SPSS with CORRESPONDENCE. There is also a package in MATLAB called CAR that was designed for CA (Lorenzo-Seva, van de Velden, & Kiers, 2009). This list is of course not exhaustive, and the relevant packages are likely to change as new computational tools are developed.

References

The American Presidency Project. (2019). *The American Presidency Project*. Retrieved from https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/annual-messages-congress-the-state-the-union

Borgatti, S. P., & **Everett, M. G.** (1997). Network analysis of 2-mode data. *Social Networks*. doi:10.1016/s0378-8733(96)00301-2

Breiger, R. L. (1974). The duality of persons and groups. *Social Forces*, 53, 181–190.

Burt, C. (1950). The factorial analysis of qualitative data. *The British Journal of Mathematical and Statistical Psychology*, 3, 166–185.

Davis, A., **Gardner, B. B.**, & **Gardner, M. R.** (2009). *Deep South: A social anthropological study of caste and class*. Columbia: University of South Carolina Press. (Original work published 1941)

D’Esposito, M. R., **De Stefano, D.**, & **Ragozini, G.** (2014). On the use of multiple correspondence analysis to visually explore affiliation networks. *Social Networks*, 38, 28–40.

Dray, S., & **Dufour, A. B.** (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20.

Francesco, G., & **Roberta, G.** (2019). Cross-country analysis of perception and emphasis of hotel attributes. *Tourism Management*, 74, 24–42.

Freeman, L. C. (2003). Finding social groups: A meta-analysis of the southern women data. In **R. L. Breiger**, **K. Carley**, & **P. Pattison** (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers* (pp. 39–77). Washington, DC: The National Academies Press.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York, NY: Wiley.

Gilula, Z., & **Haberman, S. J.** (1986). Canonical analysis of contingency tables by maximum likelihood. *Journal of the American Statistical Association*, 81, 780–788.

Goodman, L. A. (1996). A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of pearson, yule, and fisher, and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association*, 91, 408–428.

Greenacre, M. (1984). *Correspondence analysis*. London, England: Academic Press.

Greenacre, M. (2017). *Correspondence analysis in practice* (3rd ed.). New York, NY: Chapman and Hall/CRC.

Halford, M. (2019). *Prince: Statistical factor analysis in python* (Version 0.6.2). Retrieved from https://pypi.org/project/prince/

Hill, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Applied Statistics*, 23, 340.

Hjellbrekke, J. (2018). *Multiple correspondence analysis for the social sciences*. New York, NY: Routledge.

Jockers, M. (2017). Syuzhet: Extracts sentiment and sentiment-derived plot arcs from text. *R Package Version*, 1.

Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M(CA), FAMD, MFA, HCPC, Factoextra*. STHDA (http://www.sthda.com).

Lê, S., **Josse, J.**, & **Husson, F.** (2008). FactoMineR: An R Package for multivariate analysis. *Journal of Statistical Software*, 25, 1–18.

Lebart, A., **Morineau, A.**, & **Warwick, K.** (1984). *Multivariate statistical descriptive analysis*. New York, NY: Wiley.

Lebart, L. (1998). Correspondence analysis, discrimination, and neural networks. In **C. Hayashi**, **K. Yajima**, **H. H. Bock**, **N. Ohsumi**, **Y. Tanaka**, & **Y. Baba** (Eds.), *Data science, classification, and related methods* (pp. 423–430). Tokyo, Japan: Springer.

Lebart, L., **Salem, A.**, & **Berry, L.** (1997). *Exploring textual data*. Norwell, MA: Kluwer Academic Publishers.

Le Roux, B., & **Rouanet, H.** (2004). *Geometric data analysis: From correspondence analysis to structured data analysis*. Dordrecht, the Netherlands: Springer Science & Business Media.

Lorenzo-Seva, U., **van de Velden, M.**, & **Kiers, H. A. L.** (2009). CAR: A MATLAB Package to compute correspondence analysis with rotations. *Journal of Statistical Software*, 31, 1–14.

Nenadic, O., & **Greenacre, M.** (2007). Correspondence analysis in R, with two-and three-dimensional graphics: The ca Package. *Journal of Statistical Software*, 20. Retrieved from https://goedoc.uni-goettingen.de/handle/1/5892

Noma, E., & **Smith, D. R.** (1985). Scaling sociomatrices by optimizing an explicit function: Correspondence analysis of binary single response sociomatrices. *Multivariate Behavioral Research*, 20, 179–197.

Padgett, J. F., & **Ansell, C. K.** (1993). Robust action and the rise of the medici, 1400–1434. *The American Journal of Sociology*, 98, 1259–1319.

Roberts, J. M., Jr. (1996). Alternative approaches to correspondence analysis of sociomatrices. *The Journal of Mathematical Sociology*, 21, 359–368.

Safak, E. (2018). *MCA: Multiple Correspondence Analysis with Pandas* (Version 1.0.3). Retrieved from https://pypi.org/project/mca/

Sourial, N., **Wolfson, C.**, **Zhu, B.**, **Quail, J.**, **Fletcher, J.**, **Karunananthan, S.**, … **Bergman, H.** (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *Journal of Clinical Epidemiology*, 63, 638–646.

van der Maaten, L., & **Hinton, G.** (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.

Venables, W. N., & **Ripley, B. D.** (2013). *Modern applied statistics with S-PLUS*. New York, NY: Springer Science & Business Media.

White, H. C., **Boorman, S. A.**, & **Breiger, R. L.** (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *The American Journal of Sociology*, 81, 730–780.