Biplots in Practice

MICHAEL GREENACRE

Professor of Statistics at the Pompeu Fabra University

Chapter 1 Offprint

Biplots–the Basic Idea

Biplots—the Basic Idea

The basic idea of the biplot is very simple, and like all simple solutions to complex problems it is both powerful and very useful. The biplot makes information in a table of data become transparent, revealing the main structures in the data in a methodical way, for example patterns of correlations between variables or similarities between the observations. Rectangular data matrices are the "raw material" in many research areas, existing in spreadsheets or databases. The rows of a data matrix are usually observed sampling units such as individuals, countries, demographic groups, locations, cases,…, and the columns are variables describing the rows, such as responses in a questionnaire, economic indicators, products purchased, environmental parameters, genetic markers, … Throughout this book several data matrices from different areas of research will be used as illustrations of the power of the biplot to reveal the inherent structure in the data. In this initial chapter, we describe the basic geometric concepts that form both the foundation of the biplot's definition as well as its practical interpretation.

Contents

A biplot is the generalization of the well-known *scatterplot* of observations on two variables. Exhibit 1.1 shows data for 12 European countries in 2008 on the following three variables: X_1 = purchasing power per capita (expressed in euros), X_2 = gross domestic product (GDP) per capita (indexed at 100 for all 27 countries in the European Union for 2008) and X_3 = inflation rate (percentage). Exhibit 1.2 shows two scatterplots, of X_1 versus X_2 and X_3 versus X_2 respectively. At

Scatterplots

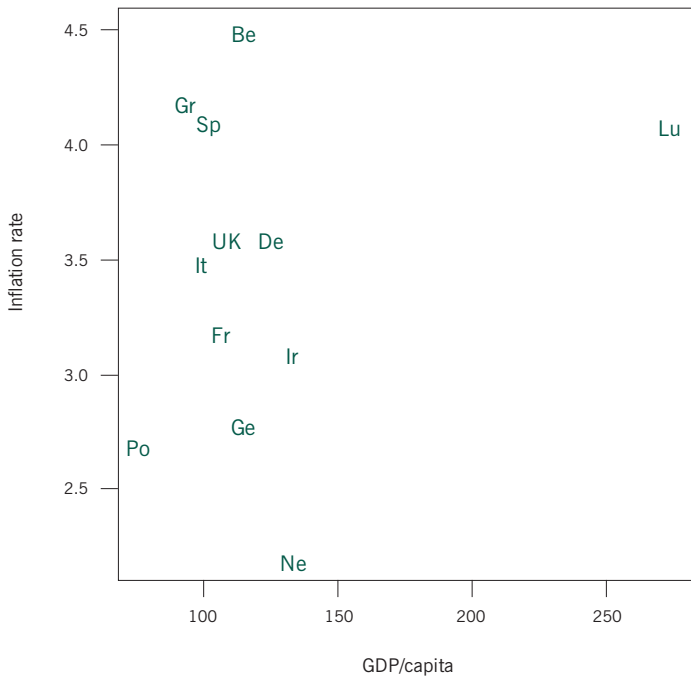|	COUNTRY	X_1	X_2	X_3
Be	Belgium	19,200	115.2	4.5
De	Denmark	20,400	120.1	3.6
Ge	Germany	19,500	115.6	2.8
Gr	Greece	18,800	94.3	4.2
Sp	Spain	17,600	102.6	4.1
Fr	France	19,600	108.0	3.2
Ir	Ireland	20,800	135.4	3.1
It	Italy	18,200	101.8	3.5
Lu	Luxembourg	28,800	276.4	4.1
Ne	Netherlands	20,400	134.0	2.2
Po	Portugal	15,000	76.0	2.7
UK	United Kingdom	22,600	116.2	3.6

a glance we can see in the first scatterplot that X_1 and X_2 are strongly correlated, as one would expect for these variables, whereas the second scatterplot shows the correlation to be weaker between X_3 and X_2. In this book we will be interested in visualizing the relationships between many variables, but let us consider for a start how we could see the inter-relationships between all three of these variables. One way is to use three-dimensional graphics, but we still have to resort to a planar representation of the points, as shown inside a cube in Exhibit 1.3. Two different views of the points are shown and the first view shows the countries lining up more or less in a linear spread, as indicated by the dashed line. We can capitalize on this by turning the cloud of points around so we look at it from a viewpoint perpendicular to this line of spread, as shown in the second view. The countries are now spread out more, and all that we miss in the second view is the small deviation of each country from the dashed line. In the second view of Exhibit 1.3 directions have been added, starting more or less at the middle of the cloud of country points and pointing in the directions of the variables as given by the sides of the cube. It is no coincidence that the two variables which were seen to be correlated turn out pointing in similar directions, since they were seen in the left hand scatterplot of Exhibit 1.2 to be lining up the countries in more or less the same order. What we have done in this informal example is reduce the three-dimensional example to a flat two-dimensional display, trying to lose as little of the spread of the countries in their original three-dimensional space as possible. The rest of this book formalizes this idea and extends it to showing clouds of points in high-dimensional spaces in a subspace of reduced dimensionality.

A simple example: a matrix expressed as the product of two other matrices

To introduce the biplot in a very simple way, consider the following equality between a 5 × 4 matrix on the left-hand side and the product of two matrices, 5 × 2 and 2 × 4 respectively:
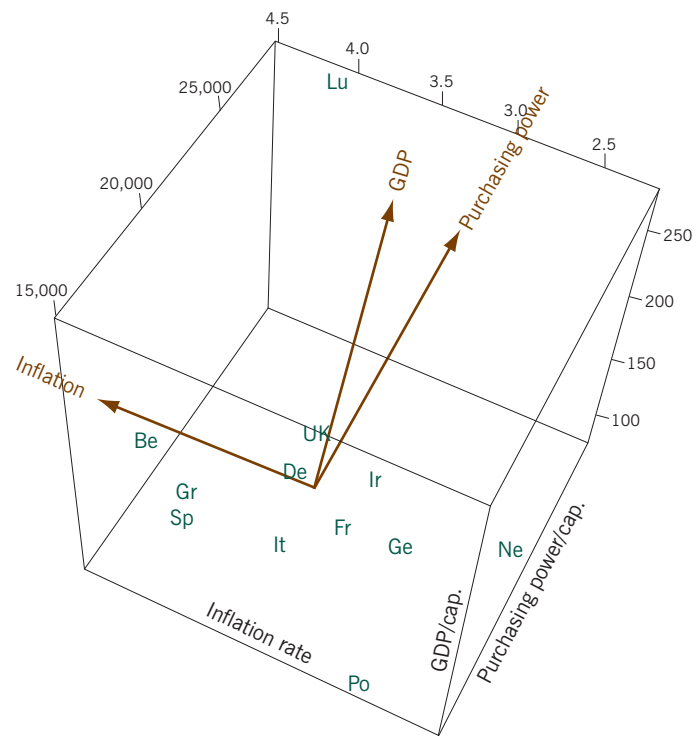
17
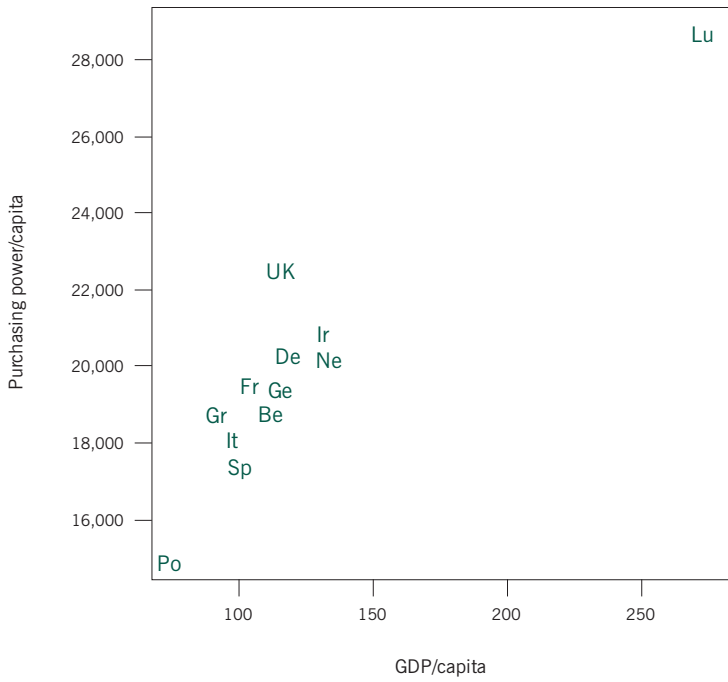
Exhibit 1.3:

A three-dimensional scatterplot of all three variables in Exhibit 1.1, seen from two points of view. The second one is more informative about the relative positions of the countries in the three-dimensional space and axes are shown parallel to their respective sides of the cube

Fundación **BBVA**
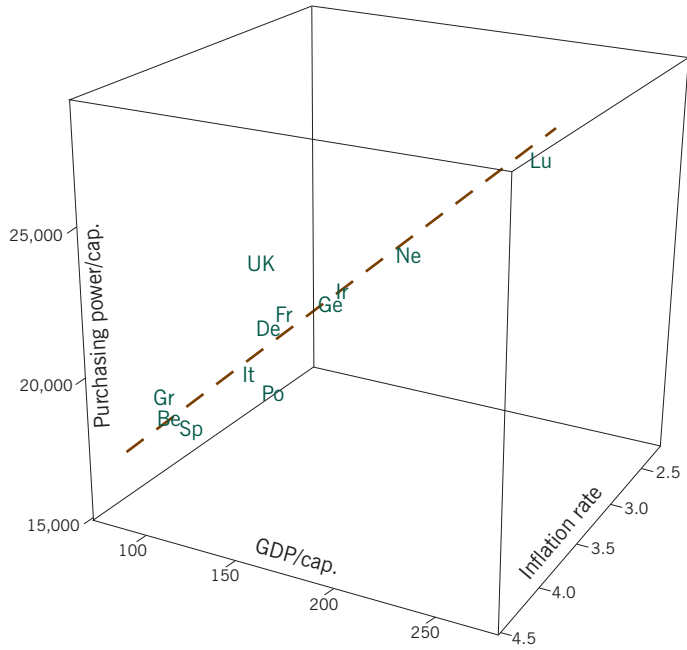
$$
\begin{pmatrix}
8 & 2 & 2 & -6 \\
5 & 0 & 3 & -4 \\
-2 & -3 & 3 & 1 \\
2 & 3 & -3 & -1 \\
4 & 6 & -6 & -2
\end{pmatrix}
=
\begin{pmatrix}
2 & 2 \\
1 & 2 \\
-1 & 1 \\
1 & -1 \\
2 & -2
\end{pmatrix}
\begin{pmatrix}
3 & 2 & -1 & -2 \\
1 & -1 & 2 & -1
\end{pmatrix}
\tag{1.1}
$$

To give some general names to these matrices, let us call the first matrix the *target matrix*, and the two matrices of the product the *left matrix* and the *right matrix* respectively, so that expressions of the form (1.1) can be written as:

$$
\text{target matrix} = \text{left matrix} \cdot \text{right matrix} \tag{1.2}
$$

Notice that for the matrix product to be valid, the number of columns of the left matrix must be equal to the number of rows of the right matrix. The rules of matrix multiplication are that the elements of a row of the left matrix are multiplied by the corresponding elements of a column of the right matrix, and then added, producing that particular row–column element of the target matrix. For example, the first value of the target matrix, 8, is equal to $2 \times 3 + 2 \times 1$; and the value −6 in row 1, column 4, is calculated from the first row of the left matrix and the fourth column of the right matrix, as $2 \times (-2) + 2 \times (-1)$. Such "sums of cross-products" are called *scalar products*, and are the basis of the geometry of the biplot.

For any two vectors $\mathbf{a}^\mathsf{T} = [a_1 \; a_2 \; \ldots \; a_m]$ and $\mathbf{b}^\mathsf{T} = [b_1 \; b_2 \; \ldots \; b_m]$, with m elements each, the *scalar product* between $\mathbf{a}$ and $\mathbf{b}$ is

$$
\mathbf{a}^\mathsf{T}\mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_m b_m \tag{1.3}
$$

The notation $^\mathsf{T}$ stands for "transpose of", turning rows into columns and vice versa. Notice that the convention is to write vectors as columns, so that a row vector is defined as the transpose of a column vector.

In the case of (1.1) each row vector of the left matrix has two elements and each column vector of the right matrix has two elements, so each of these row–column pairs has a scalar product equal to the corresponding element of the target matrix. We write (1.1) as

$$
\mathbf{S} = \mathbf{X}\mathbf{Y}^\mathsf{T} =
\begin{pmatrix}
\mathbf{x}_1^\mathsf{T} \\
\mathbf{x}_2^\mathsf{T} \\
\mathbf{x}_3^\mathsf{T} \\
\mathbf{x}_4^\mathsf{T} \\
\mathbf{x}_5^\mathsf{T}
\end{pmatrix}
\begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & \mathbf{y}_4 \end{pmatrix}
=
\begin{pmatrix}
\mathbf{x}_1^\mathsf{T}\mathbf{y}_1 & \mathbf{x}_1^\mathsf{T}\mathbf{y}_2 & \mathbf{x}_1^\mathsf{T}\mathbf{y}_3 & \mathbf{x}_1^\mathsf{T}\mathbf{y}_4 \\
\mathbf{x}_2^\mathsf{T}\mathbf{y}_1 & \mathbf{x}_2^\mathsf{T}\mathbf{y}_2 & \mathbf{x}_2^\mathsf{T}\mathbf{y}_3 & \mathbf{x}_2^\mathsf{T}\mathbf{y}_4 \\
\mathbf{x}_3^\mathsf{T}\mathbf{y}_1 & \mathbf{x}_3^\mathsf{T}\mathbf{y}_2 & \mathbf{x}_3^\mathsf{T}\mathbf{y}_3 & \mathbf{x}_3^\mathsf{T}\mathbf{y}_4 \\
\mathbf{x}_4^\mathsf{T}\mathbf{y}_1 & \mathbf{x}_4^\mathsf{T}\mathbf{y}_2 & \mathbf{x}_4^\mathsf{T}\mathbf{y}_3 & \mathbf{x}_4^\mathsf{T}\mathbf{y}_4 \\
\mathbf{x}_5^\mathsf{T}\mathbf{y}_1 & \mathbf{x}_5^\mathsf{T}\mathbf{y}_2 & \mathbf{x}_5^\mathsf{T}\mathbf{y}_3 & \mathbf{x}_5^\mathsf{T}\mathbf{y}_4
\end{pmatrix}
\tag{1.4}
$$

where $\mathbf{X}$ consists of a set of row vectors $\mathbf{x}_i^\mathsf{T}$, $i = 1,\ldots,5$ (each with two elements), and $\mathbf{Y}^\mathsf{T}$ consists of a set of column vectors $\mathbf{y}_j$, $j = 1,\ldots,4$ (also with two elements each). The right matrix is written as a transpose so that we have two matrices, $\mathbf{X}$ and $\mathbf{Y}$, whose *rows* both contain the vectors used in the scalar product calculations:
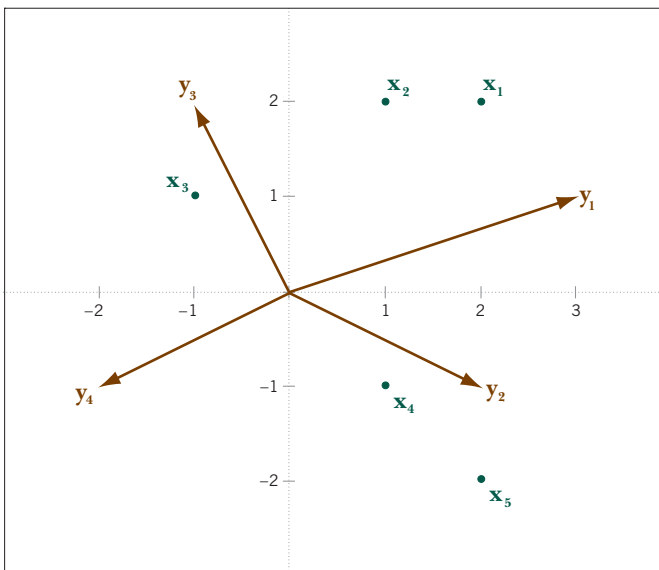
$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^\mathsf{T} \\ \mathbf{x}_2^\mathsf{T} \\ \mathbf{x}_3^\mathsf{T} \\ \mathbf{x}_4^\mathsf{T} \\ \mathbf{x}_5^\mathsf{T} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} \mathbf{y}_1^\mathsf{T} \\ \mathbf{y}_2^\mathsf{T} \\ \mathbf{y}_3^\mathsf{T} \\ \mathbf{y}_4^\mathsf{T} \end{pmatrix} \tag{1.5}$$

Geometric interpretation of scalar product

There is a specific geometric interpretation of the scalar product which gives its usefulness as a method of data visualization. Using the simple example above, let us plot the rows of $\mathbf{X}$ and the rows of $\mathbf{Y}$ as points in a Euclidean space, which is two-dimensional in this case:

Exhibit 1.4:

The five points $\mathbf{x}_i$ of the left matrix and four points $\mathbf{y}_j$ of the right matrix in decomposition (1.1) (the latter points are shown as vectors connected to the origin). The scalar product between the i-th row point and the j-th column point gives the (i,j)-th value s_{ij} of the target matrix in (1.1)

$$\mathbf{x}_1 = [\,2 \quad 2\,]^\mathsf{T}$$
$$\mathbf{x}_2 = [\,1 \quad 2\,]^\mathsf{T}$$
$$\mathbf{x}_3 = [-1 \quad 1\,]^\mathsf{T}$$
$$\mathbf{x}_4 = [\,1 \quad -1\,]^\mathsf{T}$$
$$\mathbf{x}_5 = [\,2 \quad -2\,]^\mathsf{T}$$

$$\mathbf{y}_1 = [\,3 \quad 1\,]^\mathsf{T}$$
$$\mathbf{y}_2 = [\,2 \quad -1\,]^\mathsf{T}$$
$$\mathbf{y}_3 = [-1 \quad 2\,]^\mathsf{T}$$
$$\mathbf{y}_4 = [-2 \quad -1\,]^\mathsf{T}$$

In Exhibit 1.4 one of the sets of points, in this case the four points of the right matrix, are drawn as vectors connected to the origin. Each of these vectors defines a *biplot axis* onto which the other set of points can be projected. By projection of a point onto a vector we mean dropping the point perpendicularly onto the vector as shown in Exhibit 1.5. For example, the projection of $\mathbf{x}_1$ onto the vector $\mathbf{y}_1$ can be calculated as having length 2.530 (using simple trigonometry, the length of $\mathbf{x}_1$

20

is 2.828, the square root of 8, and the angle between $\mathbf{x}_1$ and $\mathbf{y}_1$ is 26.57°, so the projection of $\mathbf{x}_1$ onto $\mathbf{y}_1$ has length $2.828 \times \cos(26.57°) = 2.828 \times 0.8944 = 2.530)$. Now if this projected length is multiplied by the length of $\mathbf{y}_1$ (equal to 3.162, the square root of 10), the result is $2.530 \times 3.162 = 8.00$. This result is nothing else but the scalar product between $\mathbf{x}_1$ and $\mathbf{y}_1$: $(2 \times 3) + (2 \times 1) = 8$. So we have illustrated the result, shown in Exhibit 1.5, that the scalar product between two vectors is the length of the projection of the first vector onto the second multiplied by the length of the second one (or vice versa):

$$\mathbf{x}^{\mathsf{T}}\mathbf{y} = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos(\theta) \qquad (1.6)$$

All biplots have the above property, in fact the property is inherent to the definition of the biplot. Two sets of points are graphed in a joint plot, according to coordinates provided by the left and right matrices of the decomposition of a target matrix. Either set can be chosen as a set of biplot axes, and often this set is connected by vectors to the origin of the biplot to indicate the choice. Let us call these the *biplot vectors* as opposed to the other set of *biplot points*. The biplot points can then be projected onto the biplot axes, and their projections, multiplied by the length of the corresponding biplot vectors, give the scalar products which are equal to the elements of the target matrix.

The practical value of the biplot property of representing data by scalar products is only truly apparent if one thinks of the biplot axes as being *calibrated*. This is the

Calibration of biplot axes

placing of tic marks on the biplot axes indicating a scale for reading off the values in the target matrix by simply projecting the biplot points onto the biplot axes. For example, consider again the biplot axis in Exhibit 1.4 defined by the biplot vector $\mathbf{y}_1$, and now consider the projections of all the biplot points $\mathbf{x}_1$, …, $\mathbf{x}_5$ onto this axis. These five projections will all be multiplied by the same value, the length of $\mathbf{y}_1$, to obtain the five values in the first column of the target matrix. This means that the five projected values are directly proportional to the target values, so we can calibrate the biplot axis with a scale that depends on the length of $\mathbf{y}_1$, after which we no longer have to perform the multiplication by the length of $\mathbf{y}_1$. To know what the length of one unit is along the biplot axis, consider again the way we obtain a particular target value in row i from the corresponding scalar product:
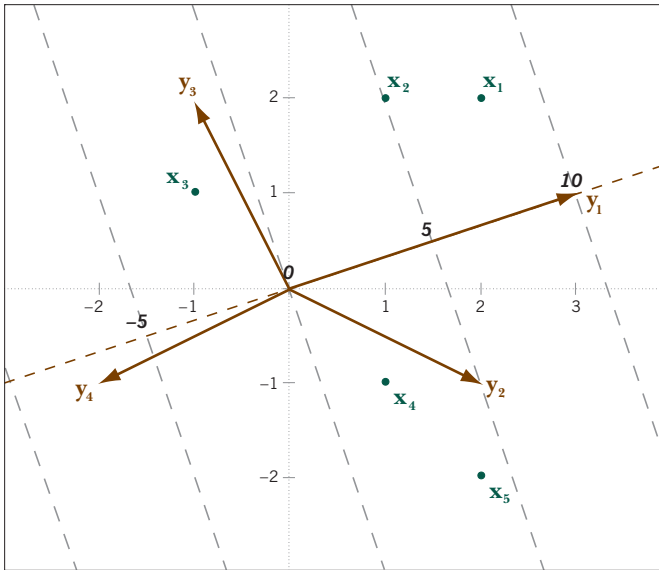
$$\text{target value in row } i = \left(\begin{array}{c} \text{length of projection} \\ \text{of } i\text{-th biplot point} \end{array} \right) \times \left(\begin{array}{c} \text{length of biplot} \\ \text{vector} \end{array} \right)$$

In order to know what one unit is on the biplot axis, we need to invert this formula for a target value of 1:

$$\text{length of projection of one unit} = 1 \ / \ \text{length of biplot vector} \qquad (1.7)$$

This means that the inverse of the lengths of the biplot vectors give us the lengths of one unit on the biplot axis (see Exhibit 1.6)—if a biplot vector is short, the intervals between units on that biplot axis are large (so values change slowly), and if the biplot vector is long, the intervals between units on that biplot axis are short (so values change fast). For example, the length of $\mathbf{y}_1$ is the square root of 10, 3.162. Hence, unit tic marks on the biplot axis through $\mathbf{y}_1$ are a distance of $1/3.162 = 0.3162$ apart.

Each vector $\mathbf{y}_j$ defines a biplot axis that can be calibrated in the same way. Although we will generally not show calibrations along the axes, the fact that axes can be calibrated gives the *raison d'être* of the biplot's interpretation. If the columns of the target matrix are variables and the rows are cases, the biplot representation will mean that variables can be depicted as axes pointing in a direction such that the values for the cases on that variable are obtained by projecting the cases onto the variables. The actual values of that variable are not as important as being able to see how the cases line up along that variable. And if two biplot axes lie in the same orientation we shall be able to deduce that the cases have the same relative positions along both variables, which translates into high inter-variable correlation.

In this introductory example of how scalar products and projections give values of a target matrix, the left and right matrices of the decomposition provide two-dimensional coordinates, easily visualized in the planar display of Exhibits 1.4 and 1.6. If these matrices were three-dimensional ($\mathbf{X}$ and $\mathbf{Y}$ with three columns), we could still see what was happening if we could view the points and vectors in three-dimensional space. Calibration and projection are performed in exactly the same way. Of course, for a given data matrix in practice, we will not be able to find a decomposition into a product of matrices with only two or three dimensions: in fact, the number of dimensions needed is equal to the *rank* of the target matrix. For these higher-dimensional situations we shall need methods for reducing the dimensionality of the target matrix—this theme will be introduced in Chapter 5 and used extensively in the rest of the book.

Moving into higher-dimensional spaces

1. *Scatterplots* typically plot observations on two variables with respect to rectangular coordinate axes. Three-dimensional scatterplots showing observations on three variables are possible using special software and a two-dimensional view of these three variables can be optimized to show a maximum amount of the variance (i.e., dispersion) of the plotted points. The biplot generalizes this idea to many variables being observed and viewed simultaneously in an optimal fashion.

2. *Biplots* are defined as the decomposition of a *target matrix* into the product of two matrices, called *left and right matrices*: $\mathbf{S} = \mathbf{X}\mathbf{Y}^{\mathsf{T}}$. Elements in the target ma-

SUMMARY:
Biplots—the Basic Idea

trix **S** are equal to *scalar products* between corresponding pairs of vectors in the rows of **X** and **Y** respectively.

3. The geometry of scalar products provides the following rule for interpreting the biplot graphically. The vectors in the left and right matrices provide two sets of points, one of which can be considered as a set of *biplot vectors* defining *biplot axes*, and the other as a set of *biplot points*. Points can be projected perpendicularly onto biplot axes to recover the values in the target matrix, since the lengths of these projections multiplied by the lengths of the corresponding biplot vectors are equal to the scalar products, and thus in turn equal to the target values.

4. *Calibration* of the biplot axes is possible, which means that values of the target matrix can be read off directly from the projections, just as in scatterplots where points are projected onto the axes to read their values.

5. The "bi" in biplot refers to the fact that two sets of points (i.e., the rows and columns of the target matrix) are visualized by scalar products, not the fact that the display is usually two-dimensional. The biplot and its geometry hold for spaces of any dimensionality, but we shall need *dimension-reducing techniques* in practice when data matrices have high inherent dimensionality and a representation is required with respect to a low number of dimensions, usually two or three.