1 *Introduction*

Biplots have been with us at least since Descartes, if not from the time of Ptolemy who had a method for fixing the map positions of cities in the ancient world. The essential ingredients are coordinate axes that give the positions of points. From the very beginning, the concept of distance was central to the Cartesian system, a point being fixed according to its distance from two orthogonal axes; distance remains central to much of what follows. Descartes was concerned with how the points moved in a smooth way as parameters changed, so describing straight lines, conics and so on. In statistics, we are interested also in isolated points presented in the form of a scatter diagram where, typically, the coordinate axes represent variables and the points represent samples or cases. Cartesian geometry soon developed three-dimensional and then multidimensional forms in which there are many coordinate axes. Although two-dimensional scatter diagrams are invaluable for showing data, multidimensional scatter diagrams are not. Therefore, statisticians have developed methods for approximating multidimensional scatter in two, or perhaps three, dimensions. It turns out that the original coordinate axes can also be displayed as part of the approximation, although inevitably they lose their orthogonality. The essential property of all biplots is the two modes, such as variables and samples. For obvious reasons, we shall be concerned mainly with two-dimensional approximations but should stress at the outset that the *bi-* of biplots refers to the two modes and not the usual two dimensions used for display.

Biplots, not necessarily referred to by name, have been used in one form or another for many years, especially since computer graphics have become readily available. The term 'biplot' is due to Gabriel (1971) who popularized versions in which the variables are represented by directed vectors. Gower and Hand (1996) particularly stressed the advantages of presenting biplots with calibrated axes, in much the same way as for conventional coordinate representations. A feature of this book is the wealth of examples of different kinds of biplots. Although there are many novel ideas in this book, we acknowledge our debts to many others whose work is cited either in the current text or in the bibliography of Gower and Hand (1996).

Understanding Biplots John Gower, Sugnet Lubbe and Niël le Roux

1.1 Types of biplots

We may distinguish two main types of biplot:

- *asymmetric* (biplots giving information on sample units and variables of a data matrix);
- *symmetric* (biplots giving information on rows and columns of a two-way table).

In symmetric biplots, rows and columns may be interchanged without loss of information, while in asymmetric biplots variables and sample units are different kinds of object that may not be interchanged.

Consider the data on four variables measured on 21 aircraft in Table 1.1. The corresponding biplot in Figure 1.1 represents the 21 aircraft as sample points and the four variables as biplot axes. It will not be sensible to exchange the two sets, representing the aircraft as continuous axes and the variables as points. Next, consider the two-way table in Table 1.2. Exchanging the rows and columns of this table will have no effect on the information contained therein. For such a symmetric data set, both the rows and columns are represented as points as shown in Figure 1.2. Details on the construction of these biplots are deferred to later chapters.

Table 1.1 Values of four variables, ***SPR*** (specific power, proportional to power per unit weight), ***RGF*** (flight range factor), ***PLF*** (payload as a fraction of gross weight of aircraft) and ***SLF*** (sustained load factor), for 21 aircraft labelled in column 2. From Cook and Weisberg (1982, Table 2.3.1), derived from 1979 RAND Corporation report.

	Aircraft	SPR	RGF	PLF	SLF
A	FH-1	1.468	3.30	0.166	0.10
B	FJ-1	1.605	3.64	0.154	0.10
C	F-86A	2.168	4.87	0.177	2.90
D	F9F-2	2.054	4.72	0.275	1.10
E	F-94A	2.467	4.11	0.298	1.00
F	F3D-1	1.294	3.75	0.150	0.90
G	F-89A	2.183	3.97	0.000	2.40
H	XF10F-1	2.426	4.65	0.117	1.80
I	F9F-6	2.607	3.84	0.155	2.30
J	F100-A	4.567	4.92	0.138	3.20
K	F4D-1	4.588	3.82	0.249	3.50
M	F11F-1	3.618	4.32	0.143	2.80
N	F-101A	5.855	4.53	0.172	2.50
P	F3H-2	2.898	4.48	0.178	3.00
Q	F102-A	3.880	5.39	0.101	3.00
R	F-8A	0.455	4.99	0.008	2.64
S	F-104A	8.088	4.50	0.251	2.70
T	F-105B	6.502	5.20	0.366	2.90
U	YF-107A	6.081	5.65	0.106	2.90
V	F-106A	7.105	5.40	0.089	3.20
W	F-4B	8.548	4.20	0.222	2.90

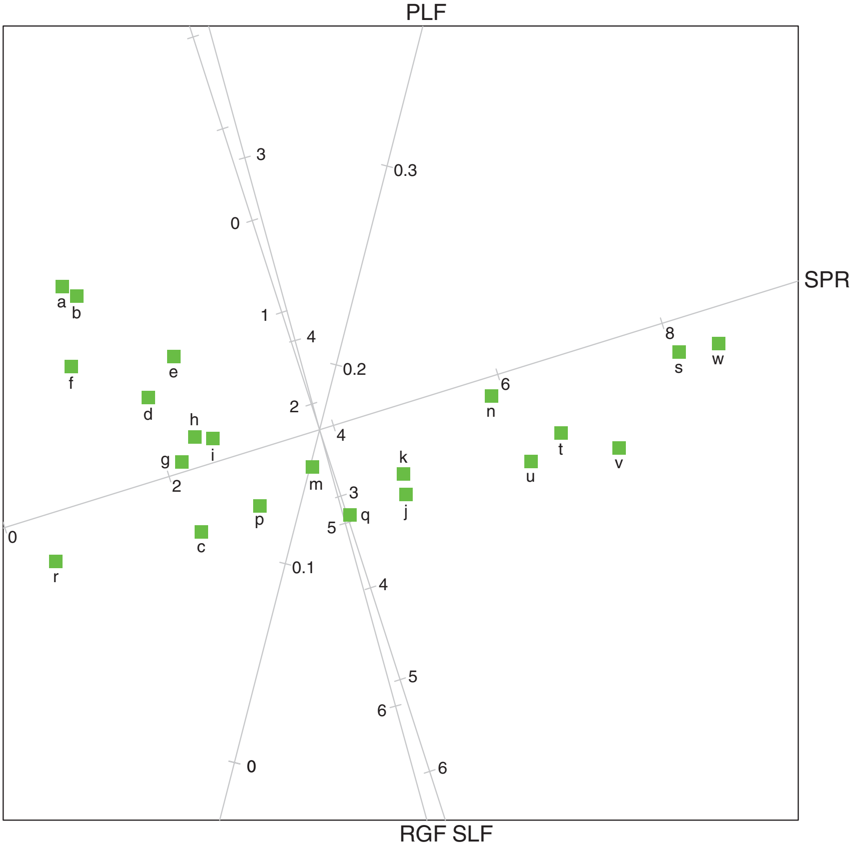

Figure 1.1 Principal component analysis biplot according to the Gower and Hand (1996) representation.

Table 1.2 Species × Temperature two-way table of percentage cellulose measured in wood pulp from four species after a hot water wash.

Temperature (°C)	Species			
	Amea	Edun	Egran	Emac
90	47.12	40.61	46.36	45.15
130	48.59	46.57	45.96	45.76
140	59.49	49.73	55.71	49.95
150	63.59	68.18	70.94	56.32
160	71.18	69.50	65.13	71.18
170	67.12	65.30	69.85	67.58

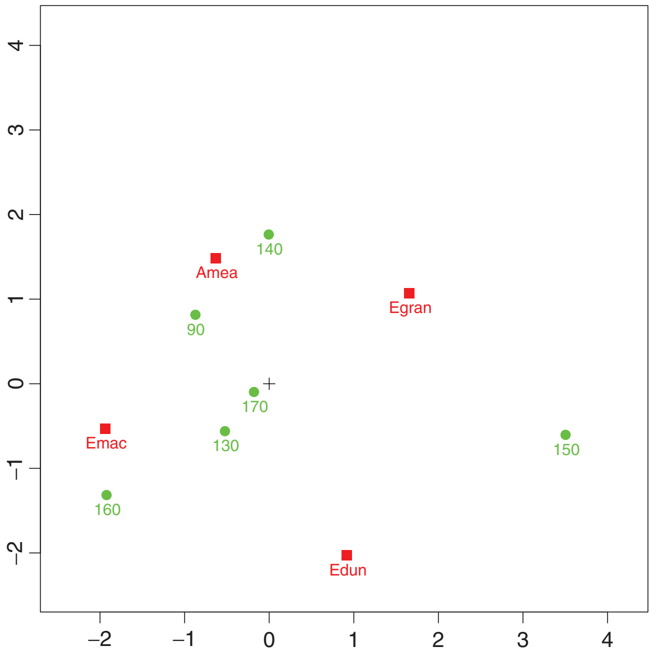

Figure 1.2 Biplot for a two-way table representing Species × Temperature.

We shall see that this distinction between symmetric and asymmetric biplots affects what is permissible in the construction of a biplot. Within this broad classification, other major considerations are:

- the types of variable (quantitative, qualitative, ordinal, etc.);
- the method used for displaying samples (multidimensional scaling and related methods);
- what the biplot display is to be used for (especially for prediction or for interpolation).

The following can be represented in an asymmetric biplot:

- distances between samples;
- relationships between variables;
- inner products between samples and variables.

However, only two of these characteristics can be optimally represented in a single biplot. In the simple biplot in Figure 1.1 all the calibration scales are linear with evenly spaced calibration points. Other types of scale are possible and we shall meet them later in other types of biplots. Figure 1.3 shows the main possibilities.

Figure 1.3(a) is the familiar equally spaced calibration of a linear axis that we have already met in Figure 1.1. Figure 1.3(b) shows logarithmic calibration of a linear axis;

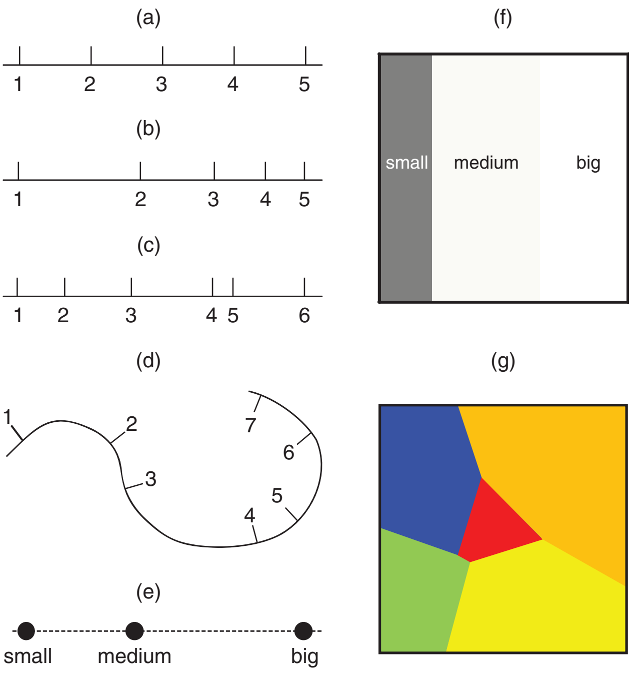

Figure 1.3 Different types of scale. (a) A linear scale with equally spaced calibration as used in principal component analysis. (b) A linear scale with logarithmic calibration. (c) A linear scale with irregular calibration. (d) A curvilinear scale with irregular calibration. (e) A linear scale for an ordered categorical variable. (f) Linear regions for ordered categorical variables (g) A categorical variable, colour, defined over convex regions.

this is an example of regular but unequally spaced calibration. In Figure 1.3(c) the axis remains linear but the calibrations are irregularly spaced. In Figure 1.3(d) the axis is nonlinear and calibrations are irregularly spaced; in principle, nonlinear axes could have equally spaced calibrations or regularly space calibrations, but in practice such combinations are unlikely. Figure 1.3(e) shows an ordered categorical variable, ***size***, not recorded numerically but only as *small*, *medium* and *big*. The calibration is indicated as a set of correctly ordered markers on a linear axis, but this is shown as a dotted line to indicate that intermediate markers are undefined (i.e. interpolation is not permitted). In Figure 1.3(f) the ordered categorical variable ***size*** is represented by linear regions; all samples in a region are associated with that level of ***size***. Figure 1.3(g) shows an unordered categorical variable, ***colour***, with five levels: *blue*, *green*, *yellow*, *orange* and *red*. These levels label convex regions. In general, the levels of unordered categorical variables may be represented by convex regions in many dimensions. Examples of these calibrations occur throughout the book.

1.2 Overview of the book

The basic steps for constructing many asymmetric biplots are summarized in Figure 1.4. Starting from a data matrix **X**, first we calculate a distance matrix **D**: $n \times n$. The essence

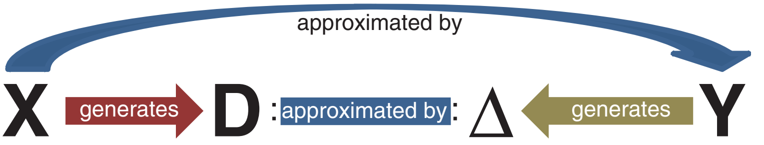

Figure 1.4 Construction of an asymmetric biplot.

of the methodology is approximating the distance matrix **D** by a matrix of Pythagorean distances $\mathbf{\Delta}$: $n \times n$. Operationally, this is achieved iteratively by updating r-dimensional coordinates **Y**, that generate $\mathbf{\Delta}$, to improve the approximation to **D**. It is hoped that a small choice of r (hopefully 2) will give a good approximation. Finally, the curved arrow represents two ideas: (i) in principal component analysis (PCA) **Y** approximates **X;** and (ii) more generally, information on **X** can be represented in the map of **Y** (the essence of biplots). These are the basic steps of multidimensional scaling (see Cox and Cox, 2001).

In general, the points given by **Y** generate distances in $\mathbf{\Delta}$ that approximate the values in **D**. In addition, and this is the special contribution of biplots, approximations to the true values **X** may be deduced from **Y**. In the simplest case, the PCA biplot, this approximation is made by projecting the orthogonal axes of **X** onto a subspace occupied by **Y**. In the subsequent chapters, we will discuss more general forms of asymmetric biplots. The most general of these, appropriately named the generalized biplot, has as special case the PCA biplot when all variables in **X** are continuous and the matrix **D** consists of Pythagorean distances. When restricting the variables in **X** to be continuous only, the rows of **X** represent the samples as points in p-dimensional space with an associated coordinate system. In the biplot, we represent the samples as points whose coordinates are given by the rows of **Y** and the coordinate system of **X** by appropriately defined biplot axes. These axes become nonlinear biplot trajectories when the definition of distance in the matrix **D** necessitates a nonlinear transformation from **X** to **Y**. The methodology outlined by Figure 1.4 allows us to also include categorical variables. Even though a categorical variable cannot be represented in the space of **X** by a linear coordinate axis, we can calculate the matrix **D** and proceed from there.

Thus, a biplot adds to **Y** information on the variables given in **X**. In multidimensional scaling, **D** may be observed directly and not derived from **X**, and then biplots cannot be constructed. The different types of asymmetric biplots discussed above depend on the properties of the variables in the matrix **X** and the distance metric producing the matrix **D**. Many special cases of importance fall within this general framework and are illustrated by applications in the following chapters. Several definitions of distance used in constructing **D** occur using both quantitative and qualitative variables (or mixtures of the two). For symmetric biplots, the position is simpler as we have only two main possibilities: (i) a quantitative variable classified in a two-way table and (ii) a two-way table of counts.

In Figure 1.5 the biplots to be discussed in the designated chapters are represented diagrammatically. The distances associated with the matrix **D** in Figure 1.4 is divided into subsets for the different types of biplots. The matrix $\mathbf{\Delta}$ always consists of Pythagorean distances to allow intuitive interpretation of the rows of **Y**.

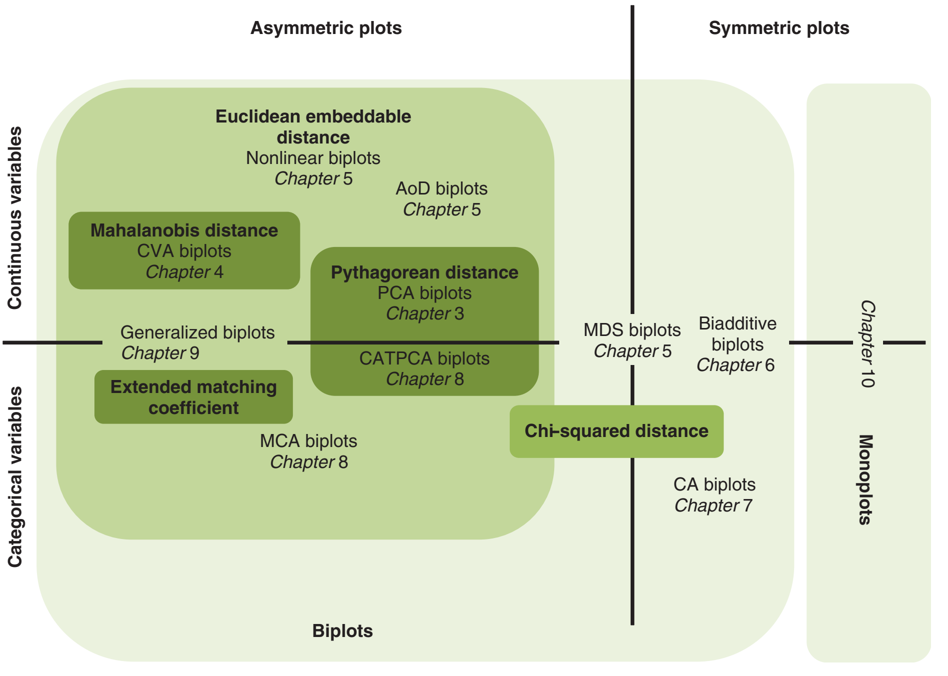

Figure 1.5 Summary of the different types of biplots discussed in subsequent chapters.

In a symmetric biplot, rows and columns have equal status and we aim to find two sets of coordinates $\mathbf{A}$ and $\mathbf{B}$, one for the rows and one for the columns respectively. Now, the main interest is in the inner product $\mathbf{AB}'$ and there is less interest in distance interpretations. A popular version of correspondence analysis (CA) approximates chi-squared distance, treating either the rows or columns as if they were 'variables' and thus giving two asymmetric biplots, not linked by a useful inner product. This form of CA is not a biplot and is sometimes referred to as a joint plot (see also Figure 10.4); other forms of CA do treat $\mathbf{X}$ symmetrically.

1.3 Software

A library of functions has been developed in the R language (R Development Core Team, 2009) and is available on the website www.wiley.com/go/biplots. Throughout this book reference will be made to the functions associated with the biplots being discussed. Examples of the commands to reproduce the figures in this book are given in the text. Sections are also included with specific information about the core functions needed for the different types of biplots.

1.4 Notation

Matrices are used extensively to enable the mathematically inclined reader to understand the algebra behind the different biplots. Bold upper-case letters indicate matrices and bold lower-case letters indicate vectors. Any column vector $\mathbf{x}$: $p \times 1$ when presented as a row vector will be denoted by $\mathbf{x}' : 1 \times p$. The following symbols are used extensively throughout the text:

n	number of samples
p	number of variables
K	number of groups or classes into which the samples are divided
m	$\min(p, K-1)$
$\mathbf{X} : n \times p$	a data matrix with n samples measured on p variables. Unless stated otherwise, the matrix $\mathbf{X}$ is assumed to be centred to have column means equal to zero.
$\mathbf{G}$	an indicator matrix, usually with n rows, where each row consists of zeros except for a one in the column associated with that particular sample
$\mathbf{N}$	diagonal matrix of the group sizes, $\mathbf{N} = (\mathbf{G}'\mathbf{G})^{-1}$
$\mathbf{n}$	diag($\mathbf{N}$)
$\overline{\mathbf{X}} : K \times p$	matrix of group means, $\overline{\mathbf{X}} = \mathbf{N}^{-1}\mathbf{G}'\mathbf{X}$
$\mathbf{I}$	identity matrix, size determined by context
$\mathbf{J} : p \times p$	$\begin{bmatrix} \mathbf{I}_r & \mathbf{0} : r \times (p-r) \\ \mathbf{0} : (p-r) \times r & \mathbf{0} : (p-r) \times (p-r) \end{bmatrix}$
$\mathbf{1}$	column vector of ones, size determined by context
d_{ij}	the distance between sample i and sample j
δ_{ij}	the fitted distance between sample i and sample j
$\mathbf{D} : n \times n$	a matrix derived from the pairwise distances of all n samples with ijth element $-\frac{1}{2}d_{ij}^2$. The latter quantities are termed *ddistances*.
diag($\mathbf{A} : p \times p$)	the $p \times p$ diagonal matrix formed by replacing all the off-diagonal elements of $\mathbf{A}$ with zeros; or, depending on the context, the p-vector consisting of the diagonal elements of $\mathbf{A}$
diag($\mathbf{a}$)	a diagonal matrix with the elements of the vector $\mathbf{a}$ on the diagonal
$\mathbf{R}$	diagonal matrix of row totals
$\mathbf{C}$	diagonal matrix of column totals
$\mathbf{E}$	$\mathbf{R11}'\mathbf{C}/n$
$\|\mathbf{A}\|^2$	tr($\mathbf{AA}'$)
$\mathbf{A}^*\mathbf{B}$	elementwise multiplication
$\mathbf{A}/\mathbf{B}$	elementwise division

The notion of *distance* is discussed in Chapter 5. Here we mention two concepts which the reader will need throughout the book. *Pythagorean distance* is the ordinary Euclidean distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ with

$$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2.$$

Any distance metric that can be embedded in a Euclidean space is termed *Euclidean embeddable*.

1.4.1 Acronyms

AoD	analysis of distance
CA	correspondence analysis
CVA	canonical variate analysis
EMC	extended matching coefficient
JCA	joint correspondence analysis
MCA	multiple correspondence analysis
MDS	multidimensional scaling
PCA	principal component analysis