Focus Article

Biplots: the joy of singular value decomposition

Michael J. Greenacre*

The biplot is a generalization of a scatterplot for two variables to the case of many variables. Instead of having samples represented as points with respect to two perpendicular axes, as in a bivariate scatterplot, there are as many axes as variables pointing in different directions. Samples are then perpendicularly projected onto axes to obtain approximate values of the data. The word 'approximate' is important, because it is not possible to represent data on many variables exactly by this procedure, but the biplot arranges the axes to display the data as accurately as possible, usually by least-squares fitting. The 'bi' in biplot refers to the rows and columns of a multivariate data matrix, where the rows are usually cases and the columns are variables. Biplots are almost always displayed in a two-dimensional plot but can just as well be displayed in three-dimensions, with more accurate data representation, using suitable graphical software, for example dynamic rotation or conditioned plots. The usual linear biplot, using least-squares approximation, relies analytically on the singular value decomposition, which in turn can be thought of as a two-sided regression problem. Biplot geometry underlies many classical multivariate procedures, such as principal component analysis, simple and multiple correspondence analysis, discriminant analysis, and other variants of dimension reduction methods such as log-ratio analysis.

How to cite this article:
WIREs Comput Stat 2012, 4:399–406. doi: 10.1002/wics.1200

Keywords: multivariate analysis; visualization; singular value decomposition; principal component analysis; correspondence analysis; discriminant analysis

INTRODUCTION

The term 'biplot' originates in a paper by Gabriel,[1] who showed that if a matrix $\mathbf{Y}$ ($n \times m$) is written as the product of two matrices, $\mathbf{A}$ ($n \times p$) and the transpose of $\mathbf{B}$ ($m \times p$):

$$\mathbf{Y} = \mathbf{A}\mathbf{B}^{\mathrm{T}}, \tag{1}$$

then the rows and columns of $\mathbf{Y}$ can be displayed in a p-dimensional Euclidean space using coordinates in the rows of $\mathbf{A}$ and $\mathbf{B}$, respectively (algebraically, p is the rank of $\mathbf{Y}$). Each element y_{ij} of $\mathbf{Y}$ is then equal to the scalar product between the ith row vector of $\mathbf{A}$ and the jth row vector of $\mathbf{B}$, a result which has many interesting geometric interpretations, as we will show later. $\mathbf{Y}$ is called the *target matrix* and the matrices $\mathbf{A}$ and $\mathbf{B}$ the *left* and *right matrices* of the biplot. The interest of the result for practical use is when p is small, say 2 or 3, which gives a two- or three-dimensional representation of the matrix $\mathbf{Y}$. For general matrices $\mathbf{Y}$ of high rank it is thus necessary to find a target matrix of low rank that approximates $\mathbf{Y}$—most often an approximation of rank 2 is sought to obtain a planar display. Matrix approximations can be obtained most conveniently by least-squares using the singular value decomposition (SVD), which is the generalization of the eigenvalue–eigenvector decomposition to rectangular matrices. The beauty of the SVD is that it not only provides the target matrix approximation but also the left and right matrices $\mathbf{A}$ and $\mathbf{B}$ needed for the biplot. Before giving the general definition of the biplot, we consider so-called regression biplots, which provide a natural introduction to the geometric concepts needed later.

*Correspondence to: michael.greenacre@upf.edu

Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain

TABLE 1 | Six Economic Indicators for the 27 European Union Countries in 2011[1]

Countries	(abbr.)	*CPI*	*UNE*	*INP*	*BOP*	*PRC*	*UN%*
Belgium	BE	116.03	4.77	125.59	908.6	6716.5	−1.6
Bulgaria	BG	141.20	7.31	102.39	27.8	1094.7	3.5
Czech Rep.	CZ	116.20	4.88	119.01	−277.9	2616.4	−0.6
Denmark	DK	114.20	6.03	88.20	1156.4	7992.4	0.5
Germany	DE	111.60	4.63	111.30	499.4	6774.6	−1.3
Estonia	EE	135.08	9.71	111.50	153.4	2194.1	−7.7
Ireland	IE	106.80	10.20	111.20	−166.5	6525.1	2.0
Greece	EL	122.83	11.30	78.22	−764.1	5620.1	6.4
Spain	ES	116.97	15.79	83.44	−280.8	4955.8	0.7
France	FR	111.55	6.77	92.60	−337.1	6828.5	−0.9
Italy	IT	115.00	5.05	87.80	−366.2	5996.6	−0.5
Cyprus	CY	116.44	5.14	86.91	−1090.6	5310.3	−0.4
Latvia	LV	144.47	12.11	110.39	42.3	1968.3	−3.6
Lithuania	LT	135.08	11.47	114.50	−77.4	2130.6	−4.3
Luxembourg	LU	118.19	3.14	85.51	2016.5	10051.6	−3.0
Hungary	HU	134.66	6.77	115.10	156.2	1954.8	−0.1
Malta	MT	117.65	4.15	101.65	359.4	3378.3	−0.6
Netherlands	NL	111.17	3.23	103.80	1156.6	6046.0	−0.4
Austria	AT	114.10	2.99	116.80	87.8	7045.5	−1.5
Poland	PL	119.90	6.28	146.70	−74.8	2124.2	−1.0
Portugal	PT	113.06	9.68	89.30	−613.4	4073.6	0.8
Romania	RO	142.34	4.76	131.80	−128.7	1302.2	3.2
Slovenia	SI	118.33	5.56	105.40	39.4	3528.3	1.8
Slovakia	SK	117.17	9.19	156.30	16.0	2515.3	−2.1
Finland	FI	114.60	5.92	101.00	−503.7	7198.8	−1.3
Sweden	SE	112.71	6.10	100.50	1079.1	7476.7	−2.3
United Kingdom	UK	120.90	6.11	90.36	−24.3	6843.9	−0.8

CPI, consumer price index (index = 100 in 2005); UNE, unemployment rate in 15–64 age group; INP, industrial production (index = 100 in 2005); BOP, balance of payments (€/capita[2]); PRC, private final consumption expenditure (€/capita[2]); UN%, annual change in unemployment rate.
[1] *CPI* November 2011, *UNE* June 2011, *INP* September 2011, *BOP* second quarter 2011, *PRC* first quarter 2011, *UN%* second quarter 2011 compared to second quarter 2010.
[2] Per capita computed with respect to population in 15–64 age group.

REGRESSION BIPLOTS

Consider the data in Table 1, six economic indicators gleaned from the latest statistics for the 27 European Union countries.[2] These variables are on different scales so we first standardize them all to have mean zero and variance 1, then use two of them, industrial production (*INP*) and balance of payments (*BOP*), to represent the 27 countries in a scatterplot (see Figure 1). Added to this plot are the other four variables, using their regression coefficients on *INP* and *BOP* to define vectors. For example, *CPI* has a linear regression function $CPI = 0.216INP - 0.118BOP$ and is thus represented by the vector with coordinates 0.216 and −0.118. *PRC* has a longer vector because it is better predicted by *INP* and *BOP*, and its regression coefficients, −0.529 and 0.438, are higher in absolute value. The R^2 values for the four regressions are 5.9% (*CPI*), 17.1% (*UNE*), 45.3% (*PRC*), and 14.9% (*UN%*), and even though the regression of *PRC* turns out to be the only statistically significant one, all relationships are still worth visualizing.

A pair of regression coefficients constitutes the *gradient*, defined as the vector of partial derivatives of the response variable with respect to the predictors. Each gradient vector defines a *biplot axis* that points in the direction of maximum slope of the respective

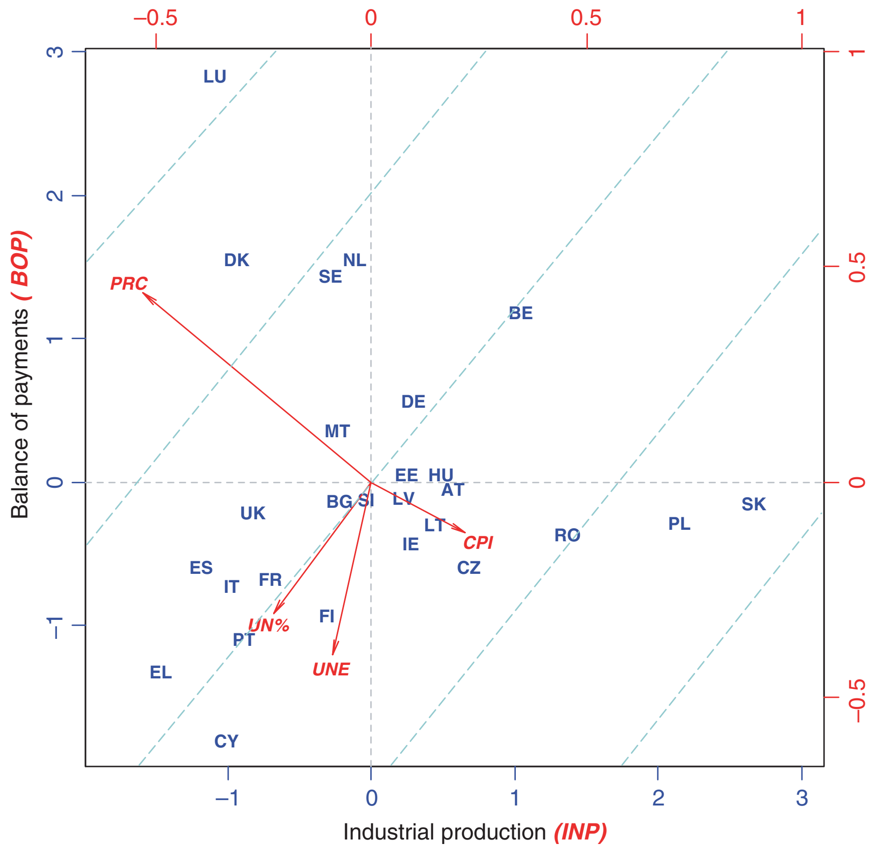

FIGURE 1 | Regression biplot of four economic variables *CPI*, *UNE*, *PRC*, and *UN%*, as linear functions of two other variables, *INP* and *BOP*. All variables are prestandardized. The 27 countries are shown in a scatterplot with respect to *INP*, *BOP*. The four response variables are depicted by their gradient vectors, that is, vectors with coefficients equal to the respective regression coefficients. The dashed lines perpendicular to the vector PRC indicate contours of its regression plane corresponding to estimated values of −2, −1, 0, 1, and 2 standard deviations of PRC, increasing from bottom right to top left in the direction of the gradient vector of PRC. Notice separate scales for countries and variables.

regression plane, and contours of equal height (i.e., equal regression estimates) are perpendicular to the biplot axis. For example, in Figure 1 the contours of the estimated values of *PRC* are shown, at intervals of 1 unit of standard deviation, and thus the perpendicular projections of the country points onto this biplot axis give *PRC*'s estimated values from the regression. The relationship between all the estimated values from the four regressions and row and column coordinates in Figure 1 can be summarized by the matrix product, as in Eq. (1):

$$\hat{\mathbf{Y}} = \mathbf{A}\mathbf{B}^{\mathsf{T}} = \begin{bmatrix} 1.026 & 1.189 \\ -0.201 & -0.124 \\ \vdots & \vdots \\ -0.838 & -0.201 \end{bmatrix} \times \begin{bmatrix} 0.216 & -0.088 & -0.529 & -0.226 \\ -0.118 & -0.400 & 0.438 & -0.304 \end{bmatrix}$$

where $\hat{\mathbf{Y}}$ is the 27 × 4 target matrix of predicted values from the regressions, **A** is the 27 × 2 left matrix of standardized values of the two predictor variables *INP* and *BOP*, which were used as the *support* to plot the country points in Figure 1, and **B** is the 4 × 2 right matrix of regression coefficients, used to draw the vectors. Notice that the scale of the variable vectors is different in Figure 1, to allow expansion of otherwise very short vectors.

Greenacre[3] (chapter 3) shows how a set of generalized linear models with responses modeled on the same predictors can lead to similar biplots, the only difference being that the underlying scales on the biplot axes are linear in the transformed mean (e.g., logarithm of mean for Poisson regression, logit function of probability for logistic regressions), and thus nonlinear if back-transformed to the original scale.

BIPLOTS AND THE WONDERFUL SVD

The difference between a regression biplot and a regular biplot is the definition of the space that acts as support for the cases. In a regression biplot

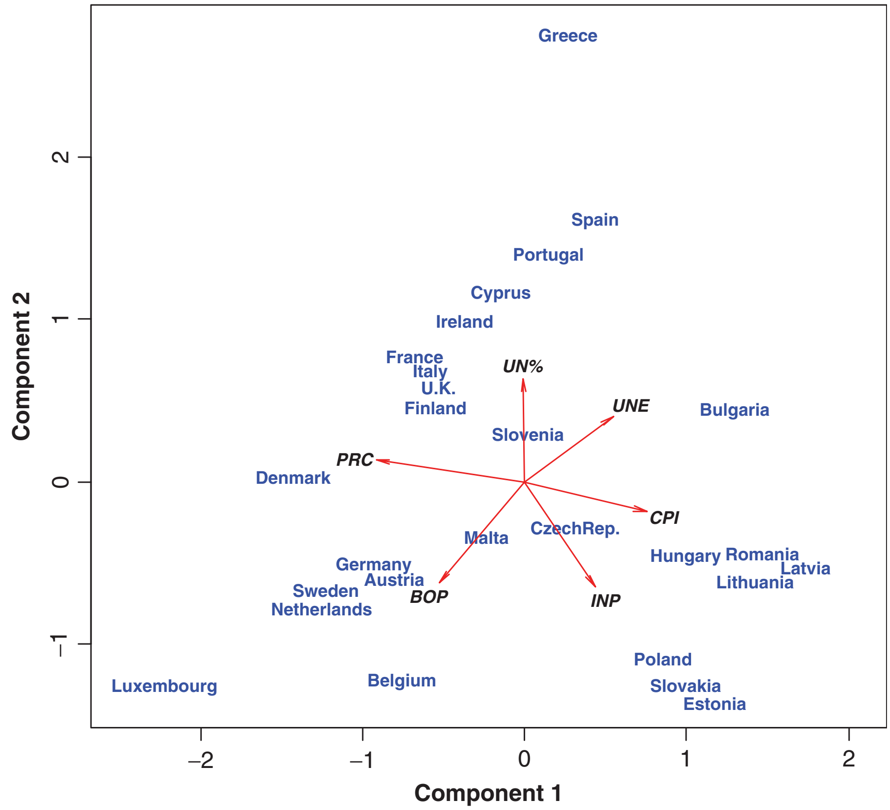

FIGURE 2 | Correlation biplot of data in Table 1, with respect to the first two principal component axes. The percentage of the variables' variance explained is 63.3%.

the supporting axes are defined by a few chosen observed variables, considered as predictors, and which are correlated, whereas in a regular biplot the axes are defined by (linear) combinations of all the variables, and uncorrelated. For example, the principal component analysis (PCA) biplot would compute such combinations that maximize the variance explained, called the principal components. The first two principal components of the standardized data in Table 1 are:

$$x_1 = 0.510CPI + 0.372UNE + 0.290INP - 0.363BOP - 0.620PRC + 0.021UN\%,$$

$$x_2 = -0.170CPI + 0.336UNE - 0.534INP - 0.493BOP + 0.120PRC + 0.562UN\%,$$

and have variances 2.26 and 1.49, respectively. As the principal components are by definition uncorrelated and the total variance of six standardized variables is equal to 6, the first two components account for a $(2.26 + 1.54)/6$ part of the variance, or 63.3%.

To mimic the regression biplot, we plot the countries according to the first two principal components, standardized, and then perform regression analyses of the six variables on the principal components to obtain vectors for each variable (see Figure 2). As the variables themselves define the supporting axes, we expect the regressions of the variables on the axes to be much more successful: indeed, the R^2 for the six variables are all highly significant: 0.630, 0.482, 0.638, 0.692, 0.892, and 0.461, respectively, with an average of 0.633, the overall proportion of variance explained. Notice that the scale of the variables in Figure 2 does not have to be expanded, as was necessary in Figure 1, because of the much higher standardized regression coefficients, hence longer arrows.

Rather than a two-step process (i.e., plotting the cases, then adding the variables by regression), the results of a biplot can be obtained in their entirety thanks to the singular value decomposition (SVD) of a rectangular matrix. A general definition of a biplot, incorporating weights on the rows and columns, is as follows:

- Suppose that matrix $\mathbf{Y}$ $(n \times m)$ (usually pre-centered, optionally normalized) is to be biplotted, and suppose that $\mathbf{r} = [r_1\ r_2 \cdots r_n]^{\mathrm{T}}$ and $\mathbf{c} = [c_1\ c_2 \cdots c_m]^{\mathrm{T}}$ are prescribed row and column weights. The objective is to find a low-rank weighted least-squares matrix approximation $\hat{\mathbf{Y}}$ such that $\sum_{i=1}^{n}\sum_{j=1}^{m} r_i c_j (y_{ij} - \hat{y}_{ij})^2$ is minimized and $\hat{\mathbf{Y}} = \mathbf{A}\mathbf{B}^{\mathrm{T}}$.

- $\mathbf{Y}$ is first transformed: $\mathbf{D}_r^{1/2}\mathbf{Y}\mathbf{D}_c^{1/2}$, and then decomposed by the SVD:

$$\mathbf{D}_r^{1/2}\mathbf{Y}\mathbf{D}_c^{1/2} = \mathbf{U}\mathbf{D}_\alpha\mathbf{V}^{\mathrm{T}}, \qquad (2)$$

where the columns of $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vectors: $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V}$ (i.e., orthonormal) and the singular values in the diagonal of $\mathbf{D}_\alpha$ are positive and in descending order: $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$.

- A target matrix approximation of rank (dimension) p and corresponding left and right biplot coordinate matrices can be obtained from

$$\hat{\mathbf{Y}} = \mathbf{D}_r^{-1/2}\mathbf{U}_{[p]}\mathbf{D}_{\alpha[p]}(\mathbf{D}_c^{-1/2}\mathbf{V}_{[p]})^{\mathrm{T}} = \mathbf{A}\mathbf{B}^{\mathrm{T}}, \qquad (3)$$

where the subscript $[p]$ denotes the first p columns of matrices $\mathbf{U}$ and $\mathbf{V}$ and corresponding first $p \times p$ diagonal part of $\mathbf{D}_\alpha$. Different choices of $\mathbf{A}$ and $\mathbf{B}$ are possible, depending on how the singular values are allocated—the coordinates which are scaled by the singular values are called *principal coordinates*, otherwise if they are unscaled they are called *standard coordinates*. The two most common choices are:

$$\mathbf{A} = \mathbf{D}_r^{-1/2}\mathbf{U}_{[p]}\mathbf{D}_{\alpha[p]} \text{ and } \mathbf{B} = \mathbf{D}_c^{-1/2}\mathbf{V}_{[p]}$$
$$\text{(rows principal, columns standard)}, \qquad (4)$$
$$\mathbf{A} = \mathbf{D}_r^{-1/2}\mathbf{U}_{[p]} \text{ and } \mathbf{B} = \mathbf{D}_c^{-1/2}\mathbf{V}_{[p]}\mathbf{D}_{\alpha[p]}$$
$$\text{(rows standard, columns principal)}. \qquad (5)$$

- The total variance in the data is the sum of squared elements of $\mathbf{D}_r^{1/2}\mathbf{Y}\mathbf{D}_c^{1/2}$, which is computed equivalently as the sum of squared singular values $\sum_k \alpha_k^2$. The variance explained in a p-dimensional biplot is the sum of squares of the first p singular values, usually expressed as a percentage of the total.

Figure 2 is the biplot of Table 1, after standardization, with equal row and column weights $r_i = 1/27$, $i = 1, \ldots, 27$ and $c_j = 1/6$, $j = 1, \ldots, 6$, $p = 2$, with rows displayed in standard coordinates and columns in principal coordinates, as in Eq. (5). If the original data are centered but not normalized this is often called the *covariance biplot*[4]—in the present example, where variables are normalized to have variances equal to 1, it is called the *correlation biplot*. In Figure 2 the cosines of the angles between the variables approximate the pairwise correlation coefficients, and the nearer the length of a variable vector is to 1, the more of its variance is being explained in the biplot (hence, the unit circle in some outputs of computer packages). The country points in Figure 2 have variance 1 on both axes.

Figure 3 is the biplot with the alternative scaling in Eq. (4), that is singular values assigned to the left matrix. Now the country points are a projection of their projections in six dimensions onto the optimal planar subspace, and their interpoint distances are approximations of Euclidean distances based on their standardized values. The variable points now have variance 1 on the respective axes.

In both Figures 2 and 3 the approximation of the data by row–column scalar products is valid. This means that in either biplot we can project country points onto biplots axes defined by the variable vectors and the projections will be approximations of the data, with the approximation being only as good as the percentage of variance explained, which is 63.3%. Each biplot axis can be calibrated in the original scale of the variable if there is interest in actually reading off approximate values, but generally the analyst is interested only in the positions of the countries relative to the direction vectors, knowing that the center represents the mean of each variable.

Once a biplot is established, an additional variable $\mathbf{y}$ such as GDP per capita, which should also be standardized in this example, can be added by regression onto the axes: the coordinates of $\mathbf{y}$ ($n \times 1$) would be $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}$. Similarly, additional countries $\mathbf{x}$ ($m \times 1$) can be added as $(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{x}$. These added variables or cases are called *supplementary* (or *passive*) points and can be thought of as points in the original data set but with zero weight and thus playing no part in determining the solution space. Coordinates of the *active* rows (countries) and columns (variables) satisfy similar regression relationships, hence the SVD is often called "two-sided regression", which can be solved by "alternating least squares".

CONTRIBUTIONS TO VARIANCE

The measure of variance explained by the biplot solution (e.g., 63.3% in Figures 2 and 3) is a global figure for all row or all column points, but one can calculate such a measure for each point separately. Suppose $\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}_\alpha$ is the full set of row principal coordinates (for a total of K, say, dimensions, the rank of the matrix being analysed) and consider the

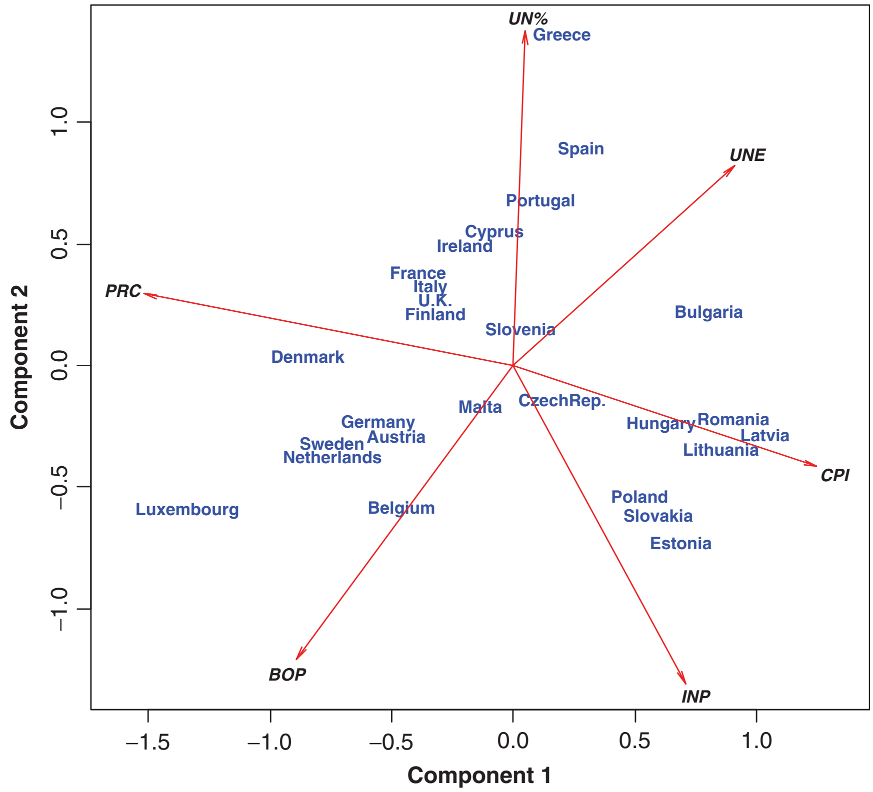

FIGURE 3 | Biplot of Table 1 using scaling Eq. (4). The percentage of the countries' variance explained is 63.3%.

following table:

$$\begin{array}{cccc} r_1 f_{11}^2 & r_1 f_{12}^2 & \cdots & r_1 f_{1K}^2 \\ r_2 f_{21}^2 & r_2 f_{22}^2 & \cdots & r_2 f_{2K}^2 \\ \vdots & \vdots & & \vdots \\ r_n f_{n1}^2 & r_n f_{n2}^2 & \cdots & r_n f_{nK}^2 \end{array} \tag{6}$$

The sum of this table is equal to the total variance, the columns sums are equal to α_k^2, $k = 1, \ldots, K$, and the row sums are equal to the variances of each row point i (i.e., usually a case) contributing to the total. Expressing the column sums (α_k^2) relative to their total are the parts of variance for the complete data set explained on each dimension, while expressing each row of Eq. (6) relative to its row sum gives parts of variance explained for each case point. Similarly, expressing each column relative to its sum α_k^2 quantifies how much each point contributes to the variance on a dimension. The above explanation applies in a symmetric way to the columns of the data matrix, using the column principal coordinates $\mathbf{G} = \mathbf{D}_c^{-1/2}\mathbf{V}\mathbf{D}_\alpha$. When the variance of the column points is decomposed as in (6) (i.e., usually the m variables), then the contributions of each variable to the dimension variances (α_k^2) give useful numerical diagnostics for interpretation, which can be incorporated graphically into what is called a *contribution biplot*.[5]

DIFFERENT BIPLOTS FOR DIFFERENT DATA TYPES

Biplot methodology extends to all data types as long as the matrix to be approximated by weighted least-squares can be appropriately set up according to the nature of the data, with special attention to the measurement scale(s) of the variables and their normalization.

Multidimensional Scaling Biplots

Figure 3 is equivalent to performing a classical multidimensional scaling of the Euclidean distances between countries based on the (standardized) economic variables, onto which the same (standardized) variables are regressed. Given any data type and an appropriate distance function between the cases, an MDS (classical or otherwise) can be performed to represent the cases in a map and then the variables, suitably normalized, can be regressed onto the dimensions of this display to be represented by their regression coefficients.

Log-Ratio Biplots

Log-ratio analysis (LRA) applies to a table of positive ratio-scale data, which is logarithmically transformed, double-centered and then biplotted. This is the method of choice for compositional data (data vectors that sum to 1), because the biplot displays all pairwise ratios of the components.[6] Weighting the rows and columns can improve the properties of this approach.[7]

Correspondence Analysis Biplots

Competing with the log-ratio biplot, correspondence analysis analyses mainly count data, but in general any table of nonnegative data for which the row and column margins are strictly positive. Instead of the log-ratio approach of visualizing pairwise ratios, the rows and the columns relative to their respective sums, called *profiles*, are biplotted. Row and column profiles are normalized using the marginal sums of the table as estimates of variance, and the resulting inter-profile distance function is called the *chi-square distance*. The total variance is then the Pearson chi-square statistic divided by the grand total of the table.[8] This method, which has no problem with data zeros, is flexible and has been applied to contingency tables, abundance and presence/absence tables in ecology, large sparse data matrices in archaeology and linguistics, and product-attribute tables in marketing.

Multiple Correspondence Analysis Biplots

An extension of CA, multiple correspondence analysis (MCA) biplots cases, often respondents in a sample survey, and a set of categorical variables, for example questions in a survey.

Discriminant Analysis Biplots

If samples are grouped, a biplot of the group mean vectors maximizes between-group dispersion rather than between-sample dispersion. Each mean vector is weighted proportional to the group size and the metric for the space is often the Mahalanobis metric, to account for correlations between variables within each group, leading to linear discriminant analysis, or canonical variate analysis. The interpretation is the same as PCA, except that it is the group mean vectors and their relationship to the variables that is focused on.

Nonlinear Biplots

There are two ways to generalize the biplot to be nonlinear. First, one can allow nonlinear transformations of the variables, which effectively means that the biplots axes are calibrated with nonlinear scales. Alternatively, the biplot axes themselves can be allowed to be curved, and projection of a case onto a variable trajectory is defined as the case's closest point to the curve.[9]

REFERENCES

1. Gabriel KR. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 1971; 58:453–467.
2. Eurostat Statistics Data Base. Economic indicators. Available at: http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database. (Accessed March 8, 2012).
3. Greenacre M. *Biplots in Practice*. BBVA Foundation, Madrid; 2010. Available at: http://www.multivariatestatistics.org. (Accessed March 8, 2012).
4. Greenacre M, Underhill LG. Scaling a data matrix in low-dimensional Euclidean space. In: Hawkins DM (ed), *Topics in Applied Multivariate Analysis*. Cambridge: Cambridge University Press; 1982, 183–268.
5. Greenacre M. Contribution biplots. *J Comput Graph Stat* 2012. In press.
6. Aitchison J, Greenacre M. Biplots of compositional data. *Appl Stat* 2002; 51:375–392.
7. Greenacre M, Lewi PJ. Distributional equivalence and subcompositional coherence in the analysis of compositional data, contingency tables and ratio scale measurements. *J Classif* 2009; 26:29–54.
8. Greenacre M. Biplots in correspondence analysis. *J Appl Stat* 1993; 20:251–269.
9. Gower JC, Harding SA. Nonlinear biplots. *Biometrika* 1988; 75:445–455.

FURTHER READING

Apart from the book by Greenacre, cited above,[3] which gives a practical introduction to all aspects of biplots, other texts on the subject include:

Bradu D, Gabriel KR. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 1978, 20:47–68.

Constantine AG, Gower JC. Graphical representation of asymmetry. *Appl Stat* 1978, 27:297–304.

Gabriel KR. Analysis of meteorological data by means of canonical decomposition and biplots. *J Appl Meteorol* 1972, 11:1071–1077.

Gabriel KR. Goodness of fit biplots and correspondence analysis. *Biometrika* 2002, 89:423–436.

Gabriel KR, Odoroff CL. Biplots in biomedical research. *Stat Med* 1990, 9:469–485.

Gower J, Hand D. *Biplots*. London: Chapman & Hall; 1996.

Gower J, Lubbe S, le Roux N. *Understanding Biplots*. UK: John Wiley & Sons; 2011.

Gower JC. Generalized biplots. *Biometrika* 1992, 79:475–493.

Gower JC. The geometry of biplot scaling. *Biometrika* 2004, 91:705–714.

Gower JC, Meulman JJ, Arnold GM. Non-linear metric biplots. *J Classif* 1999, 16:181–196.

Krzanowski WJ. Biplots for multifactorial analysis of distance. *Biometrics* 2004, 60:517–524.

Ter Braak CJF. Interpreting canonical correlation analysis through biplots of structural correlations and weights. *Psychometrika* 1990, 55:519–531.

Ter Braak CJF, Looman CWN. Biplots in reduced-rank regression. *Biomet J* 1994, 36:983–1003.

Underhill LG. The coefficient of variation biplot. *J Classif* 1990, 7:41–56.

Software for computing biplots in R is described in:

De Leeuw J, Mair P. Gifi methods for optimal scaling in R : the package homals. *J Stat Softw* 2009. Available at http://www.jstatsoft.org/v31/i04/paper. (Accessed March 8, 2012).

La Grange A, le Roux N, Gardner-Lubbe S. BiplotGUI: Interactive biplots in R. *J Stat Softw* 2008. Available at http://www.jstatsoft.org/v30/i12/paper. (Accessed March 8, 2012).

Nenadić O, Greenacre M. Correspondence analysis in R, with two- and three-dimensional graphics. *J Stat Softw* 2007. Available at http://www.jstatsoft.org/v20/i03/paper. (Accessed March 8, 2012).

Thiolouse J, Dray S. Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. *J Stat Softw* 2007. Available at http://www.jstatsoft.org/v22/i05/paper. (Accessed March 8, 2012).

Finally, for an educational and entertaining introduction to the singular value decomposition (SVD) in song, watch and listen to “It had to be U”, written by Michael Greenacre, and sung by Gurdeep Stephens, at http://www.youtube.com/StatisticalSongs. The reference and direct web link is:

Greenacre M, Stephens G. It had to be U: the SVD song. *Youtube* 2011. Available at http://www.youtube.com/watch?v=JEYLfIVvR9I. (Accessed March 8, 2012).

