

BIPLOT

Authored by
mohammad looti

November 12, 2025

RECOMMENDED CITATION

mohammad looti (2025). *BIPLOT*. PSYCHOLOGICAL SCALES. Retrieved from
<https://scales.arabpsychology.com/?p=68253>

Biplot

Primary Disciplinary Field(s): Multivariate Statistics, Data Visualization, Psychometrics, Ecology

1. Core Definition

The **biplot** is a fundamental graphical tool within the field of **multivariate statistics**, designed specifically for the simultaneous visualization of both the rows (observations or samples) and the columns (variables or features) of a high-dimensional data matrix. The term "biplot" itself highlights this duality, providing a single, unified two-dimensional graphical representation that summarizes the underlying structure of the data. Essentially, it serves as a method for displaying the results of a dimensionality reduction technique, most commonly **Principal Component Analysis (PCA)** or Factor Analysis, where the original data matrix has been approximated or decomposed into a lower-rank structure, typically rank two.

Unlike traditional scatter plots that display only observations or bar plots that summarize only variable distributions, the biplot integrates these two perspectives into a single coordinate system. The observations are typically represented as points, while the variables are represented as vectors or axes originating from the plot's center. This combination allows researchers to visually assess relationships not only among the samples themselves (clustering or dispersion) and among the variables themselves (correlation or orthogonality) but, crucially, also the relationship between specific samples and specific variables (which samples score highly or lowly on which variables). The effectiveness of the biplot lies in its ability to condense a complex, multidimensional dataset into an easily interpretable visual summary, making latent patterns, interactions, and potential outliers immediately apparent.

Mathematically, the biplot is intrinsically linked to the factorization of the data matrix, often achieved through techniques like Singular Value Decomposition (SVD). If the original data matrix X is approximated by the product of two matrices, $X \approx AB^T$, where A represents the coordinates for the rows (samples) and B represents the coordinates for the columns (variables), the biplot graphically renders the entries of A and B in a shared coordinate plane. The quality of the biplot representation is therefore directly related to how well the rank-two approximation captures the total variance present in the original data matrix, a critical metric that must be assessed when interpreting the derived visualization.

2. Etymology and Historical Development

While graphical techniques for displaying multivariate data have roots extending back to the early 20th century--particularly in the context of factor analysis and multidimensional scaling--the formal concept and coinage of the term "**biplot**" are attributed primarily to statistician K. Ruben Gabriel.

In his seminal 1971 paper, "The Biplot Graphical Display of Matrices with Application to Principal Component Analysis," Gabriel formalized the mathematical framework necessary to simultaneously plot the row and column markers derived from a matrix decomposition. This work provided a rigorous foundation, connecting the visual display directly to the algebraic properties of the data matrix and establishing the biplot as a statistically justifiable visualization tool rather than merely an exploratory graph.

Prior to Gabriel's formalization, similar ideas were explored in specialized fields. For instance, the techniques inherent in **Correspondence Analysis (CA)**, which plots categories of rows and columns derived from contingency tables, exhibit the properties of a biplot, although the underlying scaling and decomposition methods differ slightly from those used in PCA biplots. However, Gabriel's contribution was to unify these approaches under a single, generalized framework applicable to any data matrix X that could be approximated via a low-rank SVD, thereby expanding its utility far beyond specialized categorical data analysis.

Following its introduction, the biplot rapidly gained traction, particularly in applied statistical fields such as psychometrics, agriculture, and ecology, where the interpretation of complex, high-dimensional datasets is commonplace. Subsequent methodological work focused on refining the various scaling options available within the biplot framework (e.g., focusing on maximizing the representation of distances between samples versus the angles between variables) and extending its application to other multivariate methods, including canonical correlation analysis and discriminant analysis. These developments solidified the biplot's status as a critical interpretive tool for complex data reduction techniques throughout the late 20th and early 21st centuries.

3. Key Components and Mathematical Basis

The construction and interpretation of a biplot rely on understanding the distinct roles of its two primary components: the **row markers** (observations/samples) and the **column markers** (variables). The mathematical foundation invariably involves the Singular Value Decomposition (SVD) of the centered data matrix X . The SVD decomposes X into three matrices: $X = U D V^T$, where U and V are orthonormal matrices, and D is a diagonal matrix containing the singular values. A rank-two approximation of X is then used to generate the 2D biplot, utilizing the first two columns of $U D$ for row coordinates and the first two columns of V for column coordinates, or variations thereof, depending on the chosen scaling.

The configuration of the biplot is highly dependent on the choice of scaling, often referred to as α -scaling. Common variants include the standard biplot ($\alpha=1$ or $\alpha=0$) and the correlation biplot ($\alpha=0.5$). In a typical **Principal Component Biplot**, the observation points are plotted such that the Euclidean distance between any two points in the biplot approximates the original distance between those two observations in the high-dimensional space. Conversely, the

variable vectors (arrows) are interpreted by their length and direction: the length of a vector approximates the variance of that variable captured by the plane, and the angle between any two vectors approximates the correlation between those two variables (a small angle indicates high positive correlation, a 90-degree angle indicates near-zero correlation, and a 180-degree angle indicates high negative correlation).

A particularly powerful interpretive feature is the ability to project observation points orthogonally onto a variable vector. The position of this projection indicates the approximate value of that observation for that specific variable. For instance, an observation point projected far along the positive direction of a variable vector suggests that the sample has a high value (score) for that variable, relative to the mean. This simultaneous representation of sample values relative to variable profiles is what distinguishes the biplot from simpler scatter plots of principal components, providing a richer context for interpreting the structure of the latent dimensions.

4. Applications and Examples

The biplot is an exceptionally versatile tool, finding critical application across diverse scientific and commercial disciplines wherever complex datasets require visual condensation and interpretability. In **Ecology** and **Environmental Science**, biplots are frequently used to visualize species-environment relationships. For example, researchers might plot sampling sites (observations) and environmental factors (variables like temperature, pH, or nutrient availability) simultaneously to determine which environmental factors drive the grouping of certain habitats or which sites are most heavily influenced by a specific pollutant.

In **Psychometrics** and **Social Science**, biplots are invaluable for interpreting the results of complex surveys and psychological testing. If a PCA is performed on responses to a battery of attitude questions, the biplot can display individual respondents (points) and the survey questions (vectors). This allows analysts to visually confirm whether certain questions cluster together (suggesting they measure the same underlying construct or factor) and to identify which groups of respondents align most closely with particular attitudinal profiles. This application is crucial for validating construct validity and identifying distinct subgroups within a population.

Furthermore, in **Genetics** and **Agronomy**, the biplot serves as the standard visualization method for analyzing Genotype-by-Environment (GxE) interaction data, often referred to as the **GGE Biplot** (Genotype plus Genotype-by-Environment). Here, genotypes (rows) and environments (columns) are plotted to visualize which genotypes perform optimally in which environments, allowing breeders to select stable and high-performing crop varieties efficiently. The clarity provided by the biplot in immediately identifying mega-environments and superior genotypes makes it indispensable for agricultural research and development worldwide, demonstrating its practical impact far beyond theoretical statistics.

5. Significance and Impact

The primary significance of the biplot lies in its ability to bridge the gap between abstract mathematical results derived from multivariate analysis and intuitive, concrete interpretation. Prior to widespread use of biplots, the results of PCA were often communicated through complex tables of factor loadings and component scores, requiring significant expertise to synthesize the relationships between variables and observations. The biplot collapses this complexity into a single visual plane, democratizing access to the interpretation of high-dimensional data structures. This immediate visual feedback aids dramatically in the exploratory data analysis phase, allowing researchers to quickly identify patterns, check for outliers, and formulate new hypotheses.

Moreover, the biplot reinforces the understanding of key statistical concepts, such as variance and correlation, by representing them geometrically. For instance, the closer a variable vector is to the edge of the biplot space, the higher the percentage of its variance is explained by the first two principal components, providing a measure of the variable's importance in defining the primary structure. Similarly, the angle-based representation of correlation offers a tangible interpretation of linearity between variables, which is often difficult to grasp purely from correlation coefficients alone.

In pedagogical contexts, the biplot is crucial for teaching multivariate statistics. It serves as a visual aid to demonstrate how high-dimensional data is projected and approximated onto lower dimensions, helping students grasp the meaning of concepts like eigenvalues, eigenvectors, and component axes. Its profound impact is evident in its ubiquitous presence in statistical software packages (such as R, SAS, and specialized statistical environments), where it remains the default visualization method for presenting the output of principal component and factor analyses across the scientific literature.

6. Debates and Criticisms

Despite its utility, the biplot is subject to several methodological debates and practical limitations that users must recognize. The most fundamental criticism revolves around the inherent loss of information associated with the **dimensionality reduction** process. Since the biplot usually displays only the first two principal components, it effectively ignores the variance captured by all subsequent components. If the first two components capture a low percentage of the total variance (e.g., less than 60-70%), the biplot representation may be misleading, as the distances and relationships shown are poor approximations of the true high-dimensional relationships. Researchers must always report the explained variance alongside the biplot to contextualize its reliability.

Another source of contention involves the choice of **scaling method** (α -scaling). Different scaling methods prioritize different relationships. For instance, one scaling might accurately

represent the distances between samples (points), while sacrificing the accuracy of the correlation representation (angles between vectors), and vice versa. Choosing the wrong scaling for the research question can lead to incorrect conclusions about the data structure. Analysts must explicitly justify their scaling choice based on whether they are primarily interested in comparing samples or assessing variable correlations.

Finally, interpretation complexity can be a challenge, particularly when the variable vectors are densely clustered or when projections are difficult to visualize accurately due to overlapping points. While the biplot simplifies the output of multivariate analysis, its effective interpretation still requires a solid understanding of multivariate geometry and the specific decomposition method employed. Misinterpretation--especially confusing the length of the vector with the magnitude of the underlying variable value--remains a common pitfall for inexperienced users.

Further Reading

[Biplot \(Wikipedia\)](#)

[Singular Value Decomposition \(SVD\)](#)

Gabriel, K. R. (1971). The Biplot Graphical Display of Matrices with Application to Principal Component Analysis. *Biometrika*, 58(3), 453-467.

Gower, J. C., & Hand, D. J. (1996). *Biplots*. Chapman & Hall.