Article

Sparse HJ Biplot: A New Methodology via Elastic Net

Mitzi Cubilla-Montilla [1,2,*], Ana Belén Nieto-Librero [3,4], M. Purificación Galindo-Villardón [3,4] and Carlos A. Torres-Cubilla [5]

1   Departamento de Estadística, Facultad de Ciencias Naturales, Exactas y Tecnología, Universidad de Panamá, Panama City 0824, Panama
2   Sistema Nacional de Investigación, Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT), Panama City 0824, Panama
3   Department of Statistics, University of Salamanca, 37008 Salamanca, Spain; ananieto@usal.es (A.B.N.-L.); pgalindo@usal.es (M.P.G.-V.)
4   Institute of Biomedical Research of Salamanca, 37008 Salamanca, Spain
5   Department of Data Analytics, Banco General, Panama City 07096, Panama; ctorres@bgeneral.com
*   Correspondence: mitzi.cubilla@up.ac.pa

Abstract: The HJ biplot is a multivariate analysis technique that allows us to represent both individuals and variables in a space of reduced dimensions. To adapt this approach to massive datasets, it is necessary to implement new techniques that are capable of reducing the dimensionality of the data and improving interpretation. Because of this, we propose a modern approach to obtaining the HJ biplot called the elastic net HJ biplot, which applies the elastic net penalty to improve the interpretation of the results. It is a novel algorithm in the sense that it is the first attempt within the biplot family in which regularisation methods are used to obtain modified loadings to optimise the results. As a complement to the proposed method, and to give practical support to it, a package has been developed in the R language called SparseBiplots. This package fills a gap that exists in the context of the HJ biplot through penalized techniques since in addition to the elastic net, it also includes the ridge and lasso to obtain the HJ biplot. To complete the study, a practical comparison is made with the standard HJ biplot and the disjoint biplot, and some results common to these methods are analysed.

Keywords: biplot; sparse; PCA; regularization; elastic net; multivariate analysis; R Software; TCGA; breast cancer

1. Introduction

Recently, the variety and the rapid growth of datasets have led to an increase in the amount of information in many disciplines and fields of study. Due to this increase in the volume of data, a statistical approach based on dimension reduction is an essential tool to project the original data onto a subspace of lower dimensions, in such a way that it is possible to capture most of the variability. This representation can be approximated by applying multivariate techniques including principal component analysis (PCA).

PCA has its origins in works developed by [1], but more substantial development was carried out by [1,2]. A more current reference can be found in [3]. PCA is traditionally undertaken through the singular value decomposition (SVD) technique of [4].

In PCA, data are projected on orthogonal axes of maximum variability in a space of reduced dimensions, usually a plane. Thus, each principal component is a linear combination of the initial variables and their contribution to each component is established. The coefficients of the combinations, which are called loadings and are usually different from zero, generate the main drawback of the PCA: its interpretation.

Several alternatives have been proposed to improve the interpretation of the results, ranging from rotation techniques to the imposition of restrictions on factor loadings. Initially, Hausman [5] proposed restricting the value that can be assigned to loads of the

principal components (PCs) to a set of integers $\{-1, 0, 1\}$ to build simplified components. Subsequently, Vines [6] welcomed the idea of [5] and suggested the use of arbitrary integers. A different method was proposed by McCabe [7] that consisted of selecting a subset of variables, identified as main variables, based on an optimisation criterion, without having to go through PCA. To solve this problem, Cadima and Jolliffe [8] presented a method called a simple threshold, which consisted of converting factorial loadings with absolute values less than a certain threshold to zero loadings.

Traditionally, rotation techniques have been used to simplify the structure of principal components (PCs) and facilitate their interpretation [9]. However, the reduction in the dimensionality of the data is not always sufficient to facilitate the interpretation of the PCs. An ad hoc approach is to use regularisation techniques, although these require a restriction parameter to induce projection vectors with modified loadings (null or near zero), the interpretation of the results improves significantly. Thus, Tibshirani [10] introduced the least absolute shrinkage and selection operator (lasso) method in which he combined a regression model with a procedure for the contraction of some parameters towards zero, imposing a penalty on the regression coefficients. A few years later, Jolliffe and Jolliffe and Uddin [11] presented a solution that modified the traditional approach to PCs using two stages, PCA and rotation. They proposed the simplified component technique (ScoT), in which the original PCs and the VARIMAX rotation are combined to shrink loadings towards zero, thus maintaining the decrease in the proportion of explained variance. Since the loadings obtained by ScoT achieve small values, but not null, Jolliffe et al. [12], proposed the SCoTLASS (simplified component technique subject to lasso) algorithm, which imposes a restriction in such a way that some loadings are completely null while sacrificing the variance. In the same sense, Zou et al. [13] proposed a penalized algorithm called sparse PCA, which applies the elastic net technique in addition to the lasso penalty [14] and which efficiently solved the problem using minimum angle regression [15]. Subject to cardinality restrictions (the number of zero loadings per component), Moghaddam et al. [16] built an algorithm for sparse components. Next, D'Aspremont et al. [17] explained the cardinality restriction based on semi-definite programming. Taking advantage of some of the previous ideas, Ref. [18] connected the PCA with the SVD of the data and obtained sparse PCs through regularisation penalties (sPCA-rSVD). Subsequently, Ref. [19] unified the low-range approach of [18] with the criterion of maximum variance of [12] and the sparse PCA method of [13], to give a general solution to the problem of sparse PCA. In the same context, Ref. [20] suggested maximising the explained variance by the cardinality of sparse PCs. Through an extension of the classical PCA, Ref. [21] built sparse components by replacing the l_2 norm in problems of traditional eigenvalues with a new norm consisting of a combination of the l_1 and l_2 norms. A modification to the PCA was presented by [22] which allowed identification of components of maximum variance, guaranteeing that each variable contributes to only to one of the factor axes obtained. Reference [23] formulated the CUR matrix decomposition, expressed in a small number of rows and/or columns in a low-range approximation of the original matrix. This is a different technique in the sense that it does not aim to obtain factorial axes as in the SVD.

A thorough review of the principal components analysis, from the traditional approach to the modern point of view through the sparse PCA, was conducted by [24]. Similarly, Ref. [25] provided an overview of sparse representation algorithms from the theory of mathematical optimisation.

Although PCA is probably the most popular technique in multivariate statistics [3,26], the structure of the observations is represented as points in the plane, but is not described jointly with the variables, which causes further inconvenience for the interpretation of the data. In this scenario, the biplot methods [27,28] have the advantage that they allow both variables and samples to be represented in a low-dimensional subspace. Hence, the present paper has been developed to take an important step in contributing to the process of analysing large-scale data, and at the same time to make a novel contribution in favour of the biplot methods.

The main objective is to propose a new biplot methodology called the Sparse HJ biplot as an alternative method to improve the interpretation of information provided by high-dimensionality data. The suggested algorithm adapts the restrictions to penalise the loadings and produce sparse components; that is, each component is a combination of only the relevant variables. In addition to the proposed method, a package was implemented in the R language to give practical support to the new algorithm.

The paper is organised into the following sections. In Section 2.1, descriptions of the approach proposed by [27] and HJ biplot [28] are presented and their properties are synthesised. Next, Section 2.2 describes the disjoint biplot (DBiplot) [29], a recent adaptation to the study of disjoint axes in the HJ biplot. Section 3 presents our main contribution, a method called Sparse HJ biplot, in which the elastic net penalty is applied to obtain zero loadings. Finally, in Section 4 this method is applied to a dataset and compared with results obtained by other methods.

2. Materials and Methods

In multivariate analysis when you have the interest of representing variables and individuals in the same coordinate system, biplots methods allow obtaining visually interpretable results. Although biplot methods and classical methods of multivariate analysis work with two-way matrices, the need to integrate techniques for manipulating large volumes of data is necessary for many scientific disciplines. The advance of science requires the design of advanced techniques to simplify the analysis of the data. In the case of Biplot methods, the literature only reports the disjoint biplot technique (described in Section 2.2), to produce axes with zero factorial loadings that facilitate the interpretation of massive data. Therefore, to analyze this type of data, in this paper an innovative multivariate technique, named "sparse HJ-Biplot", was developed. This is a new alternative of HJ-Biplot adapted to the analysis of large data sets, that raises the interpretation of the results. The mathematical structure of the algorithm is described in Section 2.3.

To illustrate the techniques used in this paper, we take a sample from data published by The Cancer Genome Atlas [30]. The data subset used is "breast.TCGA" and it was downloaded from the mixOmics R language package [31]. It has been selected the proteomics information which includes 150 breast cancer samples and the expression or abundance of omics data in 142 proteins. These samples are classified into three groups: Basal, Her2, Luminal A.

2.1. Biplot and HJ-Biplot

The biplot methods [27] are a form of low-dimensional graphic representation of a multivariate data matrix (individuals x variables). The two most important biplot factorisations presented by [27], were the GH biplot and the JK biplot. The GH biplot achieves a high-quality representation of the variables, while the JK biplot achieves this high quality in the ranks of individuals. Consequently, the biplot representation is not simultaneous for either of the two methods.

An alternative to optimise the biplot methods described by [27,28] proposed a multivariate technique called the HJ biplot. This contribution maximises the representation quality of both rows and columns simultaneously [32] for the same coordinate system. In this way, it is possible to interpret the relationships between individuals and variables at the same time.

From the algebraic point of view, the biplot methods are based on the same principles of PCA and SVD of a data matrix. In a biplot, the data are projected onto orthogonal axes of maximum variability in a space of reduced dimension. The HJ biplot facilitates the interpretation of the positions of rows, columns and row-column relations through the axes, in the same way as correspondence analysis [33–36].

In the HJ biplot, the algorithm starts with the decomposition into eigenvalues and eigenvectors of the matrix X_{nxp} defined previously:

$$X \cong UDV^T$$

where,

X: is the data matrix
U: is the matrix of data whose columns contain the eigenvectors of XX^T
V: is the matrix of data whose columns contain the eigenvectors of X^TX
D: is the diagonal matrix containing the eigenvalues of X
U and V must be orthonormal, that is, $U^TU = I$ and $V^TV = I$, to guarantee the uniqueness of the factorisation.

The best approximation to X is, therefore:

$$X \cong UDV^T \cong \sum_{k=1}^{r} d_k u_k v_k^T$$

Considering that de X matrix is centered, the markers for the columns in the HJ biplot are matched with the marked columns of the GH biplot; in turn, the markers for the rows are made to agree with the marker's rows of the JK biplot. Thus,

$$\mathrm{E} = UD \text{ and } \mathrm{G} = VD$$

In the same manner, the row markers in the HJ biplot correspond with the row markers of JK biplot, and the markers for the columns coincide with the markers for the columns in a GH biplot, concerning the factorial axes. Figure 1 shows the relationships between U and V.
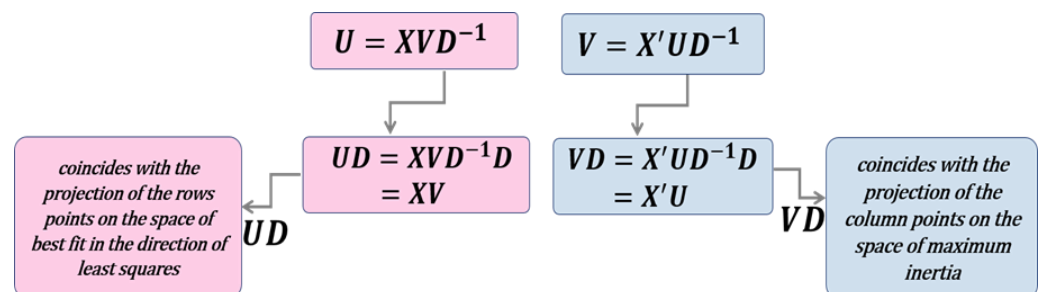
Figure 1. Markers in the HJ biplot.

The rules for the interpretation of the HJ biplot are:

- The proximity between the points that represent the row markers is interpreted as the similarity between them. Consequently, nearby points allow the identification of clusters of individuals with similar profiles.
- The standard deviation of a variable can be estimated by the module of the vector which represents it.
- Correlations between variables can be captured from the angles between vectors. If two variables are correlated, they will have an acute angle; if the angle they form is obtuse the variables will present a negative correlation; and, if the angle is a right angle it indicates that the variables are not correlated.
- The points orthogonally projected onto a variable approximates the position of the sample values in that variable.

2.2. Disjoint HJ Biplot

An algorithm for HJ biplot representation was proposed by [29,37]. The DBiplot allows for better interpretation of the extracted factorial axes and is a method that constructs disjoint factor axes guaranteeing that each variable of the original data matrix contributes only to one dimension. The algorithm starts with a random classification of the variables in the principal components, and the optimal classification is sought through an iterative procedure leading to maximisation of the explained variability.

The graphic representation in the subspace of reduced dimension is carried out through the HJ biplot. For this, a function called CDBiplot is hosted within the graphic interface biplotbootGUI [37]. The interface has three main functions. The CDBiplot function executes the interface to build the DBiplot. Using this interface, it is also possible to perform the representation of clusters, using the clustering biplot function (CBiplot) and a third alternative called clustering disjoint biplot, which allows the simultaneous representation of two ways, where the latter is characterised by disjoint axes.

Concerning the biplot, we have not found any other evidence for the formulation of alternative algorithms that penalise and contract the loadings of factorial axes to enhance the interpretation of the results. Because of this, we propose a new methodology for the HJ biplot that adapts the restrictions on the principal components applying the elastic net penalty. This new approach is called the Sparse HJ biplot.

2.3. Sparse HJ Biplot

The elastic net method for regression analysis, which combines the ridge and lasso regularisation techniques was presented by [14]. This method penalises the size of the regression coefficients based on the l_1 and l_2 norms, as follows:

$$l_1 = ||\beta||_1 = \sum_{j=1}^{p} |\beta_j| \ \ and \ \ l_2 = (||\beta||_2)^2 = \sum_{j=1}^{p} \beta_j^2$$

The Sparse HJ biplot proposes a solution to the problem of obtaining a linear combination of variables determined by a vector of sparse loadings that maximises data variability or minimises construction error. We use the concept of minimisation of the reconstruction error (*E*):

$$E = [||X - \hat{X}||^2] = Trace\ (E[(X - \hat{X})(X - \hat{X})^T])$$

Based on this approach, the ability to interpret the axes (sparse) obtained is greatly improved.

In this work, the elastic net regularisation method is implemented in the HJ biplot, combining the lasso and ridge techniques. The formulation of the HJ biplot as a regression problem imposes restrictions on factorial loadings to produce modified axes.

Since the HJ biplot does not reproduce the starting data, a factor is introduced to make this recovery possible. The following model is obtained:

$$\hat{X} = AD^{-1}B^T + E$$

From the elastic net regularisation method, modified loadings for the biplot are derived as follows:

$$V_{elasticnet} = \text{argmin} \left|\left| X - AD^{-1}B^T \right|\right|^2 + \lambda_2 \sum_{j=1}^{p} V_j^2 + \lambda_1 \sum_{j=1}^{p} |V_j|$$

Using the first *k* factorial axes, the matrices are defined:

$$A_{pxk} = [\alpha_1,\ \alpha_2, \ldots,\ \alpha_k] \text{ and } B_{pxk} = [\beta_1,\ \beta_2, \ldots,\ \beta_k].$$

For any $\lambda_2 > 0$, we have:

$$(\hat{A}, \hat{B}) = argmin \sum_{i=1}^{n} \left|\left| x_i - AB^T x_i \right|\right|^2 + \lambda_2 \sum_{j=1}^{k} ||\beta_j||^2 + \lambda_{1,j} \sum_{j=1}^{k} ||\beta_j||_1$$

subject to $A^T A = I_{KxK}$, where $\lambda_{1,j}$ is the lasso penalty parameter to induce sparsity (Sparsity: the condition of the penalty that refers to the automatic selection of variables, setting sufficiently small coefficients to null).

λ_2 is the regularization parameter to contract the loadings.

$||\cdot||^2$ denotes the l_2 norm and $||\cdot||_1$ denotes the l_1 norm.

This problem can be solved by alternating the optimisation between A and B, using the LARS-EN algorithm [14].

For fixed A, B is obtained by solving the following problem:

$$\hat{\beta}_j = argmin \, ||X\alpha_j - X\beta_j||^2 + \lambda_2||\beta_j||^2 + \lambda_{1,j}||\beta_j||_1$$
$$= (\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) + \lambda_2 \left|\left|\beta_j\right|\right|^2 + \lambda_{1,j} \left|\left|\beta_j\right|\right|_1$$

where each $\hat{\beta}_j$ is an elastic net estimator.

For fixed B, the penalty part is ignored and minimized

$$argmin \sum_{i=1}^{n} ||x_i - AB^T x_i||^2 = ||X - XBA^T||^2$$

subject to $A^T A = I_{KxK}$.

This a Procrustes problem, and the solution is provided by DVS, $(X^T X)B = UDV^T$ with $\hat{A} = UV^T$.

Recently [38] proposed a solution to the optimisation problem using the variable projection method. The main characteristic of this method is to partially decrease the orthogonally constrained variables.

The steps for the implementation of the elastic net regularisation method in the HJ biplot are detailed in the following algorithm (see Algorithm 1).

Algorithm 1 Sparse HJ biplot algorithm using elastic net regularisation.

---

1. Consider a nxp data matrix.
2. A tolerance value is set (1×10^{-5}).
3. The data is transformed (centred or standardised).
4. Decomposition of the original data matrix is performed via SVD.
5. A is taken as the loadings of the first k components $V[, 1{:}k]$.
6. β_j is calculated by:

$$\beta_j = (\alpha_j - \beta_j)^T X^T X \left(\alpha_j - \beta_j\right) + \lambda_2 \left|\left|\beta_j\right|\right|^2 + \lambda_{1,j}\left|\left|\beta_j\right|\right|_1$$

7. A is updated via SVD of $X^T X \beta$:

$$X^T X \beta = UDV^T \rightarrow A = UV^T$$

8. The difference between A and B is updated:

$$dif_{AB} = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|\beta_i|^2 |\alpha_i|^2} \sum_{j=1}^{m} \beta_{ij} - \alpha_{ij}$$

9. Steps 4, 5 and 6 are repeated until $dif_{AB} <$ tolerance.
10. The columns are normalized using $\hat{V}_j^{EN} = \frac{\beta_j}{||\beta_j||}$, $j = 1, \dots, k$
11. We then calculate the row markers and column markers.
12. The elastic net HJ biplot obtained by the previous steps is plotted.

The following scheme (Figure 2) presents the steps describing the application of the regularisation methods in the HJ biplot, which leads to modified axes being obtained.
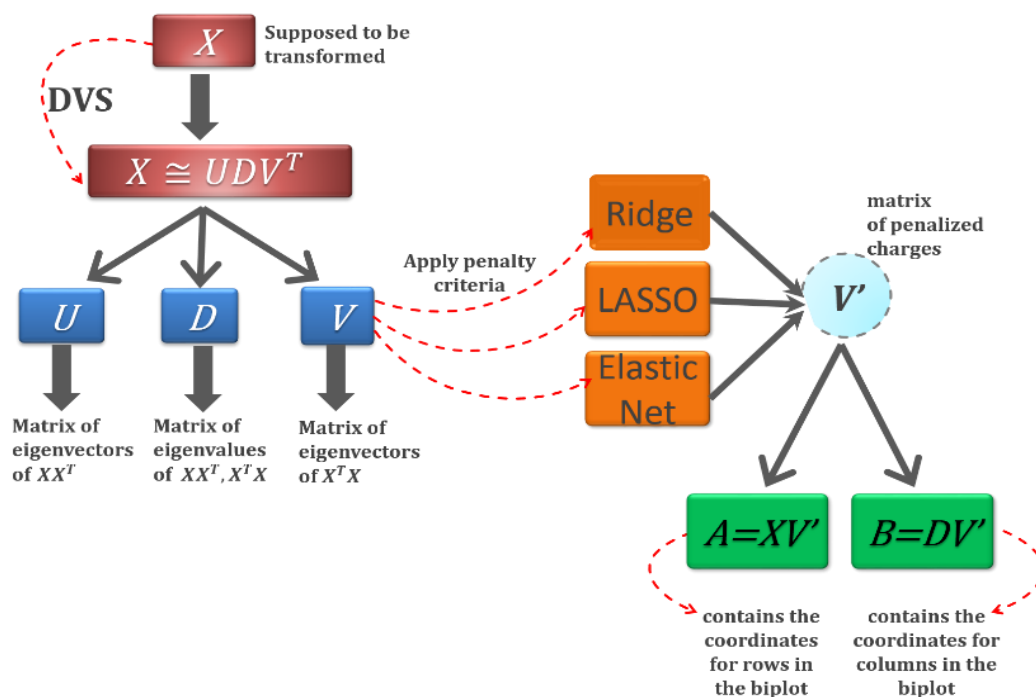
Figure 2. Sheme of the Sparse HJ biplot.

Explained variance by the Sparse HJ biplot.

In the HJ biplot, orthogonal loadings and uncorrelated axes are obtained from the transformation of the original variables. Both conditions are met, since for a covariance matrix $\hat{D} = X^T X$ then $V^T V = I$ and $V^T D V$ is a diagonal matrix. Taking $\hat{Z}$ as the estimated sparse PCs, the total explained variance is determined by $\text{tr}(\hat{Z}^T \hat{Z})$.

Although sparse PCs are capable of producing orthogonal loadings, their components are correlated [12]. Under these conditions, it is not appropriate to calculate the explained variance in the same way as in an ordinary biplot, since the true variance would be overestimated.

In this analysis, the aim is that each component should be independent of the previous ones; therefore, if linear dependence exists, it must be eliminated. From the alternatives that are available to obtain the adjusted total variance and give a solution to this problem in sparse components, [13] suggest using projection vectors to remove linear dependence. These authors denote $\hat{Z}_{j \cdot 1, \dots, j-1}$ as the residual of fitting $\hat{Z}$ for $\hat{Z}_1, \dots, \hat{Z}_{j-1}$, as follows:

$$\hat{Z}_{j \cdot 1, \dots, j-1} = \hat{Z}_j - H_{1, \dots, j-1} \hat{Z}_j$$

where $H_{1, \dots, j-1}$ is the projection matrix over $\{\hat{Z}_i\}_1^{j-1}$.

Therefore, the adjusted variance of $\hat{Z}_j$ is $\left\| \hat{Z}_{j \cdot 1, \dots, j-1} \right\|^2$ and the total explained variance is determined as $\sum_{j=1}^{k} \left\| \hat{Z}_{j \cdot 1, \dots, j-1} \right\|^2$. In the case of the estimated sparse components $\hat{Z}$ are uncorrelated, this formula matches $\text{tr}(\hat{Z}^T \hat{Z})$.

The QR decomposition, in which Q is an orthonormal matrix and R an upper triangular matrix, is a simpler way to estimate the adjusted variance. Taking $\hat{Z} = QR$ we have $\left\| \hat{Z}_{j \cdot 1, \dots, j-1} \right\|^2 = R_{jj}^2$.

The total explained variance is then $\sum_{j=1}^{k} R_{jj}^2$.

2.4. Software

To give practical support to the new algorithm and to provide a new tool that enables the use of the proposed method, a package was developed in the R programming language (R Core Team, 2021), called SparseBiplots [39].

This package includes a collection of functions that allow multivariate data to be represented on a low dimension subspace, using the HJ biplot methodology. The package implements the HJ biplot and three new techniques to reduce the size of the data, select variables, and identify patterns. It performs the HJ biplot adapting restrictions to decrease and/or produce zero loadings, using the methods of regularization ridge, lasso, and elastic net into the loadings matrix.

In each of the techniques implemented it is possible to obtain the eigenvalues, the explained variance, the loadings matrix, the coordinates for the individuals and the coordinates of variables. The package also allows the graphic representation into two selected dimensions to be obtained using the ggplot2 grammar [40].

This package combines an advanced methodology, the power of the R programming language and the elegance of ggplot2 to help models clearly explain the results and improve the informative capacity of the data. It is a new alternative to analysing data from different sources such as medicine, chromatography, spectrometry, psychology, or others.

3. Illustrative Example

In order to illustrate the new algorithm and to compare the results with the standard HJ biplot and the DBiplot, we consider the subset of data explained in Section 2.

The analysis of the data begins with the calculation of the loadings matrix associated with the factorial axes of each of the three methods, to compare the contributions of the proteins to the new axes (see Table 1). In this way a technical meaning is given to each axis, facilitating the interpretation of the results. No sparsity is induced in the HJ Biplot loadings matrix. Each factorial axis is obtained as a linear combination of all the proteins, making the characterisation of each axis difficult. DBiplot, on the other hand, generates disjoint factorial axes, where each protein contributes to a single factorial axis; this facilitates the grouping of proteins with similar performance making the axes mostly characterisable. However, the fact that each protein contributes its information entirely to a single axis severely limits the variability explained by the model. In this new proposal, the sparse biplot decreases the importance of proteins that contribute the smallest information to each axis, producing sparse factor axes with some loadings as zeros, by the penalty imposed. In contrast to the DBiplot, variables can contribute information to more than one axis. It does not limit the grouping of variables in different axes, conversely produces axes that can hold some similarities, but keeps the difference in their characterisation.

Table 1. Loadings matrix for the first three principal components obtained from the HJ biplot, DBiplot and the Sparse HJ biplot algorithms. To enhance the sparsity in each method, different fonts colors are used.

Proteins	HJ Biplot			Disjoint Biplot			Elastic Net HJ Biplot		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
14-3-3_epsilon	9.835	−0.791	0.698	1	0	0	6.330	0	0
4E-BP1	−1.127	3.408	−0.752	0	0	1	0	0	0
4E-BP1_pS65	−2.074	6.317	−2.116	0	0	1	0	1.633	0
4E-BP1_pT37	−1.862	2.997	−5.079	0	0	1	0	0	0
4E-BP1_pT70	0.486	5.227	−1.753	0	0	1	0	0.832	0
53BP1	−6.654	−3.875	1.235	0	0	1	−3.015	0	0
A-Raf_pS299	−4.47	2.681	−1.221	0	1	0	0	0	0
ACC1	−4.007	−3.042	−0.894	0	0	1	0	0	0
ACC_pS79	−4.094	−2.386	−2.147	0	0	1	0	0	0
AMPK_alpha	−1.103	−5.287	1.69	1	0	0	0	−0.917	0

Table 1. *Cont.*

Proteins	HJ Biplot			Disjoint Biplot			Elastic Net HJ Biplot		
	D1	**D2**	**D3**	**D1**	**D2**	**D3**	**D1**	**D2**	**D3**
AMPK_pT172	−0.86	−6.362	1.486	1	0	0	0	−1.075	0
ANLN	0.877	6.213	5.348	1	0	0	0	0	2.111
AR	−0.729	−6.68	4.084	1	0	0	0	−4.209	0
ARID1A	−3.643	0.852	1.409	1	0	0	0	0	0
ASNS	−4.067	8.449	−1.908	0	0	1	0	4.819	0
ATM	−5.234	−1.222	−0.396	0	1	0	−1.118	0	0
Akt	−5.513	−4.694	−0.288	0	0	1	−1.664	0	0
Akt_pS473	−1.049	1.326	−7.186	0	0	1	0	0	0
Akt_pT308	−1.782	3.054	−5.25	0	0	1	0	0	0
Annexin_I	6.102	−0.652	−4.703	1	0	0	1.919	0	0
B-Raf	−7.829	0.995	2.763	1	0	0	−4.104	0	0
Bak	9.633	−1.78	−1.534	1	0	0	6.042	0	0
Bax	4.12	−1.972	−2.539	1	0	0	0	0	0
Bcl-2	1.021	−6.875	4.678	1	0	0	0	−3.623	0
Bcl-xL	4.824	−0.207	1.384	1	0	0	0.189	0	0
Beclin	−3.283	4.485	6.71	1	0	0	0	0	2.459
Bid	9.885	1.076	1.13	1	0	0	6.612	0	0
Bim	0.715	−2.899	3.656	1	0	0	0	0	0
C-Raf	−7.355	−2.215	−0.384	1	0	0	−3.880	0	0
C-Raf_pS338	5.312	6.464	3.593	1	0	0	1.686	0	2.060
CD31	−2.116	8.398	7.088	1	0	0	0	0	4.871
CD49b	3.632	2.32	3.552	0	0	1	0	0	0
CDK1	0.488	9.008	−0.729	1	0	0	0	3.351	0.892
Caspase-7_cleavedD198	1.906	6.466	−1.048	1	0	0	0	2.174	0
Caveolin-1	7.827	−6.415	−1.311	1	0	0	3.643	−0.742	−0.702
Chk1	6.792	7.605	1.289	1	0	0	3.358	0.985	0.563
Chk1_pS345	2.418	8.966	5.551	1	0	0	0	0	4.071
Chk2	−6	3.914	−2.176	1	0	0	−1.815	1.190	0
Chk2_pT68	−2.247	10.009	4.991	1	0	0	0	0.914	4.810
Claudin-7	−4.187	0.804	4.021	1	0	0	0	0	0
Collagen_VI	8.506	−2.722	−0.097	1	0	0	4.628	0	0
Cyclin_B1	−4.571	7.465	−2.745	1	0	0	0	4.414	0
Cyclin_D1	8.872	−2.678	2.491	1	0	0	5.170	0	0
Cyclin_E1	−1.927	6.257	−3.637	1	0	0	0	3.837	0
DJ-1	3.216	−5.246	2.356	1	0	0	0	−1.053	0
Dvl3	−7.369	−0.063	−0.467	1	0	0	−3.462	0	0
E-Cadherin	−4.731	1.337	3.855	1	0	0	−0.142	0	0
EGFR	2.315	4.231	−2.325	0	0	1	0	0.244	0
EGFR_pY1068	−0.772	1.862	−2.4	0	0	1	0	0	0
EGFR_pY1173	8.702	1.484	1.23	0	1	0	5.321	0	0
ER-alpha	−0.686	−8.918	5.329	0	1	0	0	−6.509	0
ER-alpha_pS118	−3.542	−4.816	6.177	1	0	0	0	−2.834	0
ERK2	−4.911	−4.53	−1.404	0	1	0	−0.903	0	0
FOXO3a	7.666	3.783	1.127	1	0	0	4.027	0	0
Fibronectin	1.852	−2.596	−0.897	1	0	0	0	0	0
GAB2	−3.465	1.083	−0.098	1	0	0	0	0	0
GATA3	−1.972	−8.738	5.216	0	1	0	0	−6.058	0
GSK3-alpha-beta	−9.243	0.501	−0.787	1	0	0	−5.924	0	0
GSK3-alpha-beta_pS21_S9	−6.718	−0.029	−4.263	0	0	1	−2.273	0	0
HER2	−3.51	−0.115	0.883	0	0	1	0	0	0
HER2_pY1248	−0.973	1.28	−0.442	0	0	1	0	0	0
HER3	3.995	−3.859	0.899	0	1	0	0	0	0
HER3_pY1289	5.381	0.692	−1.768	0	0	1	1.365	0	0
HSP70	8.525	3.238	0.181	0	1	0	5.127	0	0

Table 1. *Cont.*

Proteins	HJ Biplot			Disjoint Biplot			Elastic Net HJ Biplot		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
IGFBP2	1.102	−3.158	1.768	1	0	0	0	0	0
INPP4B	−2.524	−6.657	6.597	1	0	0	0	−5.098	0
IRS1	4.045	−2.6	5.377	1	0	0	0	−0.741	0
JNK2	−0.584	−9.036	1.984	1	0	0	0	−4.625	−0.041
JNK_pT183_pT185	2.476	−2.463	−0.011	1	0	0	0	0	0
K-Ras	10.304	0.949	0.456	1	0	0	7.040	0	0
Ku80	−8.303	0.743	0.447	1	0	0	−4.768	0	0
LBK1	−2.52	1.968	7.864	1	0	0	0	0	1.173
Lck	4.528	3.158	−3.052	0	1	0	0.055	0	0
MAPK_pT202_Y204	−0.304	−2.731	−3.607	1	0	0	0	0	0
MEK1	2.993	−2.122	−2.706	1	0	0	0	0	0
MEK1_pS217_S221	−5.209	−1.263	−2.992	0	0	1	−0.514	0	0
MIG-6	4.206	2.429	1.513	0	1	0	0	0	0
Mre11	2.55	7.729	7.37	1	0	0	0	0	4.444
N-Cadherin	10.669	1.608	0.799	1	0	0	7.616	0	0
NF-kB-p65_pS536	−4.992	0.915	−2.066	0	0	1	−0.330	0	0
NF2	−4.468	−1.197	1.559	0	0	1	−0.284	0	0
Notch1	4.22	5.154	0.049	0	1	0	0	0.221	0
P-Cadherin	0.692	5.044	−4.532	1	0	0	0	2.793	0
PAI-1	2.668	0.836	−1.01	1	0	0	0	0	0
PCNA	5.345	2.069	−2.278	1	0	0	0.893	0	0
PDCD4	−7.3	4.39	2.29	1	0	0	−2.851	0	0.807
PDK1_pS241	−4.328	−7.468	1.103	0	0	1	−0.406	−2.068	0
PI3K-p110-alpha	−2.045	−1.893	−0.058	0	1	0	0	0	0
PKC-alpha	4.107	−2.008	−2.34	1	0	0	0	0	0
PKC-alpha_pS657	2.787	−0.643	−0.587	1	0	0	0	0	0
PKC-delta_pS664	−1.622	3.93	6.058	1	0	0	0	0	1.686
PR	−0.264	−6.858	5.072	0	0	1	0	−4.465	0
PRAS40_pT246	−4.816	6.243	−2.556	0	1	0	0	0.549	0
PRDX1	1.244	2.609	−0.483	1	0	0	0	0	0
PTEN	−3.443	−4.129	0.798	1	0	0	0	0	0
Paxillin	−3.448	−4.577	−2.127	1	0	0	0	0	0
Pea-15	3.573	−5.791	0.144	1	0	0	0	−0.581	0
RBM3	−3.695	−3.794	0.379	1	0	0	0	0	0
Rad50	−0.172	−5.001	2.657	0	1	0	0	−0.962	0
Rb_pS807_S811	−6.574	1.905	−2.987	1	0	0	−2.009	0	0
S6	−7.718	4.048	−0.745	1	0	0	−3.914	0.543	0
S6_pS235_S236	−2.383	2.566	−6.953	1	0	0	0	0.644	0
S6_pS240_S244	−2.97	2.075	−7.177	1	0	0	0	0.350	0
SCD1	−4.54	6.624	4.302	1	0	0	0	0	2.683
STAT3_pY705	4.967	−2.48	−2.268	1	0	0	0.564	0	0
STAT5-alpha	−4.779	−3.47	−2.199	0	0	1	−0.608	0	0
Shc_pY317	−4.083	3.86	2.936	0	1	0	0	0	0.740
Smad1	0.205	−0.579	2.104	1	0	0	0	0	0
Smad3	4.219	−5.98	2.297	1	0	0	0	−1.414	0
Smad4	9.908	−0.979	−1.163	1	0	0	6.333	0	0
Src	5.433	1.57	−0.883	1	0	0	1.023	0	0
Src_pY416	−2.997	8.115	2	1	0	0	0	0.303	2.653
Src_pY527	4.385	−0.67	−5.883	1	0	0	0.282	0	0
Stathmin	7.149	7.29	4.099	1	0	0	3.795	0	1.895
Syk	−4.569	2.55	−4.518	0	0	1	−0.140	0.699	0
Transglutaminase	−2.552	3.844	−0.091	0	1	0	0	0	0
Tuberin	−6.672	−6.243	−0.289	1	0	0	−3.259	−0.161	0
VEGFR2	−4.498	−4.125	2.757	0	1	0	−0.236	−0.168	0
XBP1	1.892	1.07	4.482	1	0	0	0	0	0
XRCC1	1.111	−0.622	2.963	0	0	1	0	0	0

Table 1. *Cont.*

Proteins	HJ Biplot			Disjoint Biplot			Elastic Net HJ Biplot		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
YAP_pS127	0.955	−1.745	−1.806	0	1	0	0	0	0
YB-1	−3.686	−0.954	1.81	0	1	0	0	0	0
YB-1_pS102	−2.047	2.807	−6.06	1	0	0	0	0	0
alpha-Catenin	−4.342	4.573	7.605	1	0	0	0	0	3.159
beta-Catenin	−6.877	−0.545	0.895	1	0	0	−2.921	0	0
c-Kit	2.082	3.856	−2.432	1	0	0	0	0.042	0
c-Met_pY1235	1.579	8.612	6.828	1	0	0	0	0	4.546
c-Myc	1.778	2.82	2.362	1	0	0	0	0	0
eEF2	−5.963	3.474	−1.39	1	0	0	−1.832	0.245	0
eEF2K	−5.193	−5.415	2.727	0	1	0	−1.232	−1.218	0
eIF4E	1.789	−0.64	−1.43	1	0	0	0	0	0
mTOR	−9.332	−2.094	2.38	1	0	0	−6.073	0	0
mTOR_pS2448	−4.804	0.348	−1.92	1	0	0	0	0	0
p27	3.678	−0.86	2.963	1	0	0	0	0	0
p27_pT157	−1.764	7.218	4.386	1	0	0	0	0	2.536
p27_pT198	−2.494	5.968	−1.246	1	0	0	0	1.908	0
p38_MAPK	0.767	−5.788	−0.57	1	0	0	0	0	−0.006
p38_pT180_Y182	0.215	−1.303	−3.117	1	0	0	0	0	0
p53	−2.966	9.064	4.314	1	0	0	0	0.846	3.663
p70S6K	−4.922	−1.661	1.338	1	0	0	−0.838	0	0
p70S6K_pT389	5.575	1.371	−1.5	1	0	0	1.619	0	0
p90RSK_pT359_S363	−6.319	1.839	−1.653	1	0	0	−1.687	0	0

The three techniques were then plotted using the type of cancer as a grouping factor for the samples.

In the HJ biplot (Figure 3) a slight pattern can be observed between the three types of cancer and the proteins, but this is not entirely clear. However, due to the high number of proteins, it is difficult to identify the contribution in each axis.
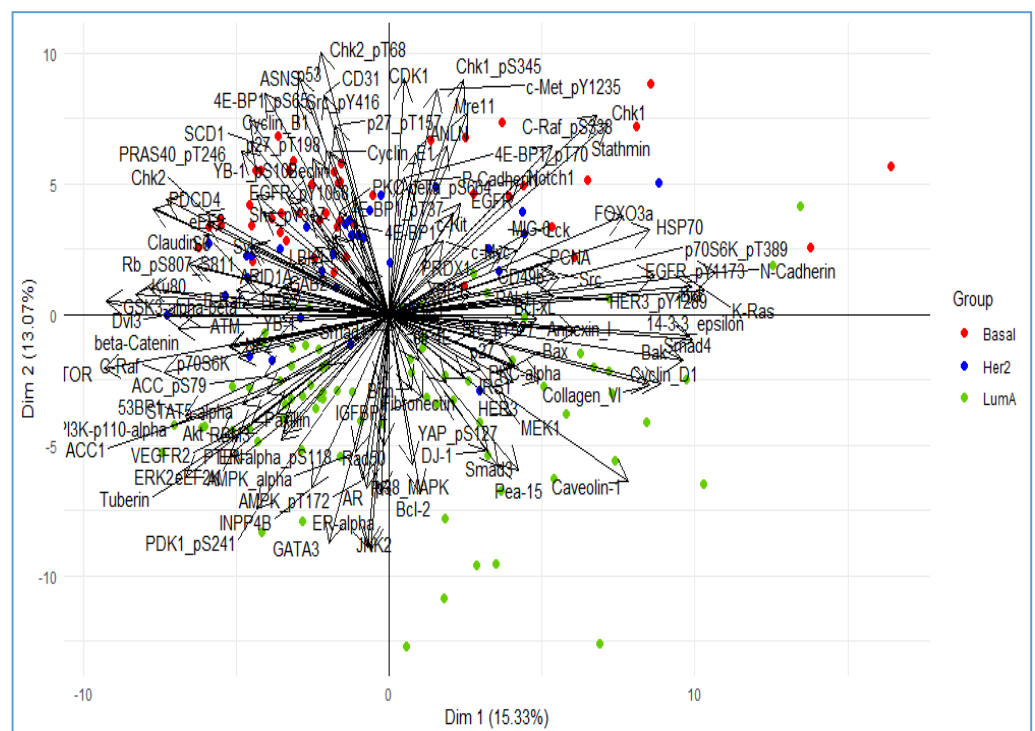

Figure 3. HJ biplot representation.

The disjoint biplot was realized with the help of the biplotbootGUI package, the data were analyzed by performing 1000 iterations of the algorithm obtaining the graph shown (Figure 4). The DBiplot shows a better structure in terms of interpretation, although the cancer–protein interaction is quite poor. A common characteristic of this technique is that most of the variables contribute mainly to the first axis.
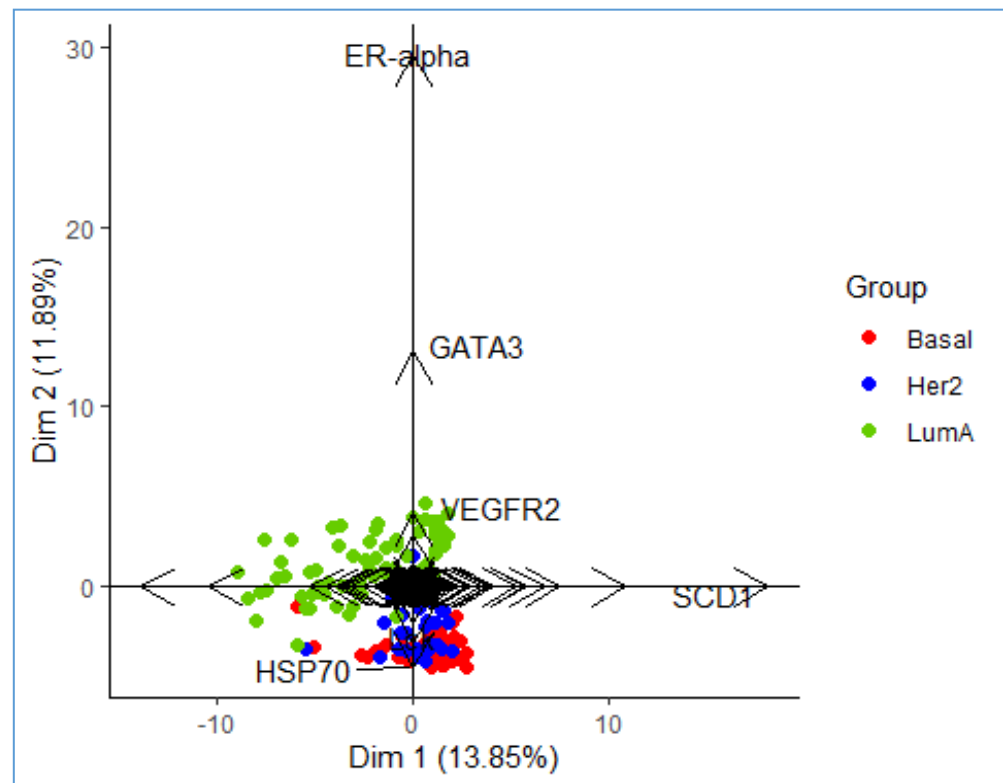

Figure 4. Disjoint biplot representation.

Finally, the SparseBiplot (Figure 5) shows an interaction between proteins and cancer types, that is clearer and easier to interpret. It shows that the variables that contribute negatively to the second axis have higher values for cancer type Luminal A. On the contrary, the proteins that contribute positively to the second axis have a higher value for cancer type Basal, and average values in Her2 type cancers. Axis 1 is a highly informative gradient (45.8%); and collectively with axis 2 captures approximately 75% of the variability of the data. The interpretation of the SparseBiplot makes possible the recognition of a proteomic characterisation of each of the groups, starting from a set of original proteins since the rest have obtained null loadings.
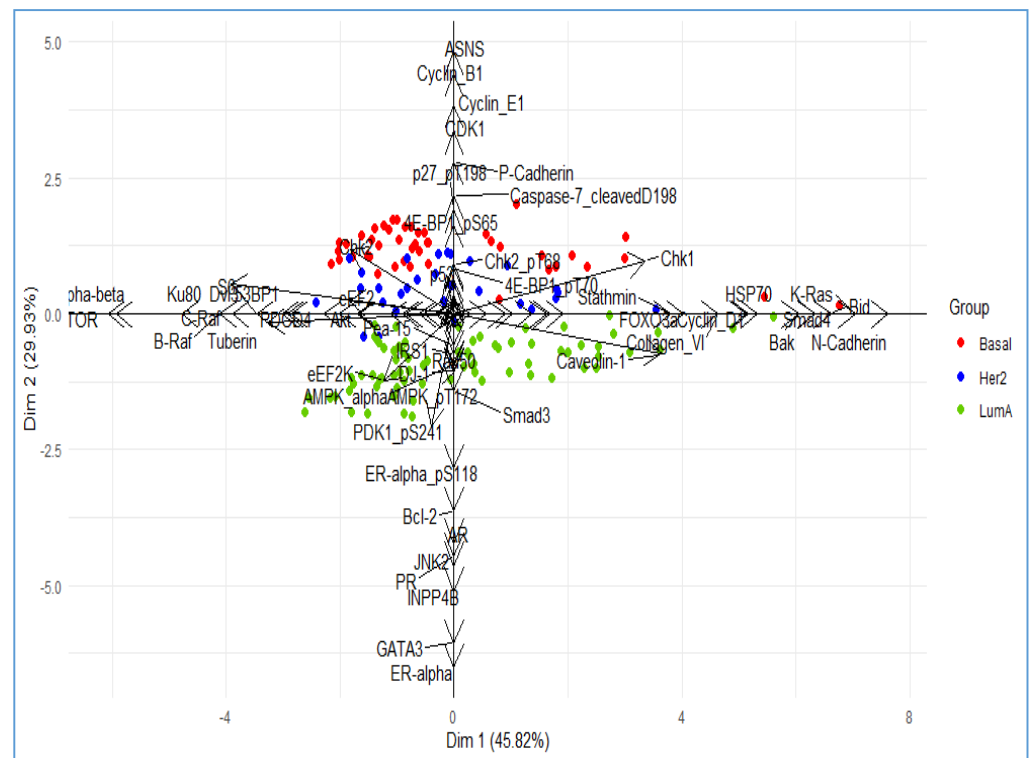
Figure 5. Elastic net HJ biplot representations.

4. Conclusions and Discussion

The theoretical contribution of this research means excellent progress for Big Data analysis and multivariate statistical methods. In this paper, we proposed an innovative algorithm to describe samples and variables in a low-dimensional space keeping the most relevant variables. Additionally, the developed software is a valuable tool that enables applying the theoretical contribution to the data analysis in every scientific field, as suggested by [41,42].

In contrast to the current developmental dynamic of the biplot method, no reference was found to applied regularisation techniques to the biplot. Therefore, the main contribution of this paper is to propose a new biplot version with penalized loadings, called the elastic net HJ biplot, using the ridge (based on the standard l_2 norm) and lasso (based on the standard l_1 norm) regularisation methods [43].

The advantage of applying penalization to induce sparsity in the HJ biplot produces the loadings matrix of the resulting axis to be sparse, allowing us to omit the slightly important variables into the biplot. Consequently, the interpretability of the biplot's results becomes clear by knowing the variables that contribute the most information to each axis in the biplot, providing efficient solutions to problems arising from the high dimension of the data [44].

A package, called SparseBiplots [39], was developed in the R programming language to perform the proposed elastic net HJ biplot algorithm. As mentioned by [45], in some study designs the sample size is smaller than the number of variables. In particular, our package has the advantage of being able to solve this type of problem in two-way tables.

The presented methodology opens a new road to future research in multivariate statistics, as it serves as the basis for it to be extended to three-way data analysis techniques, such as Statis and Statis Dual [46], partial triadic analysis [47], Tucker [48], Tucker3 [49], Parallel Factor Analysis (PARAFAC) [50], among others.

References

1. Pearson, K.F.R.S. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
2. Hotelling, H. Analysis of a Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417. [CrossRef]
3. Jolliffe, I. *Principal Component Analysis*; Wiley Online Library: Hoboken, NJ, USA, 2002.
4. Eckart, C.; Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1936**, *1*, 211–218. [CrossRef]
5. Hausman, R.E. Constrained Multivariate Analysis. In *Optimisation in Statistics*; Zanakis, S.H., Rustagi, J.S., Eds.; North-Holland Publishing Company: Amsterdam, The Netherlands, 1982; pp. 137–151.
6. Vines, S.K. Simple principal components. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2000**, *49*, 441–451. [CrossRef]
7. McCabe, G.P. Principal Variables. *Technometrics* **1984**, *26*, 137–144. [CrossRef]
8. Cadima, J.; Jolliffe, I.T. Department of Mathematical Sciences Loading and correlations in the interpretation of principle compenents. *J. Appl. Stat.* **1995**, *22*, 203–214. [CrossRef]
9. Jolliffe, I.T. Rotation of principal components: Choice of normalization constraints. *J. Appl. Stat.* **1995**, *22*, 29–35. [CrossRef]
10. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
11. Jolliffe, I.T.; Uddin, M. The Simplified Component Technique: An Alternative to Rotated Principal Components. *J. Comput. Graph. Stat.* **2000**, *9*, 689–710. [CrossRef]
12. Jolliffe, I.T.; Trendafilov, N.; Uddin, M. A Modified Principal Component Technique Based on the LASSO. *J. Comput. Graph. Stat.* **2003**, *12*, 531–547. [CrossRef]
13. Zou, H.; Hastie, T.; Tibshirani, R. Sparse Principal Component Analysis. *J. Comput. Graph. Stat.* **2006**, *15*, 265–286. [CrossRef]
14. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
15. Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least Angle Regression. *Ann. Stat.* **2004**, *32*, 407–499. [CrossRef]
16. Moghaddam, B.; Weiss, Y.; Avidan, S. Spectral Bounds for Sparse PCA: Exact and Greedy Algorithms. *Adv. Neural Inf. Process. Syst.* **2006**, *18*, 915.
17. D'Aspremont, A.; El Ghaoui, L.; Jordan, M.; Lanckriet, G.R.G. A Direct Formulation for Sparse PCA Using Semidefinite Programming. *SIAM Rev.* **2007**, *49*, 434–448. [CrossRef]
18. Shen, H.; Huang, J.Z. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* **2008**, *99*, 1015–1034. [CrossRef]
19. Witten, D.M.; Tibshirani, R.; Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **2009**, *10*, 515–534. [CrossRef]
20. Farcomeni, A. An exact approach to sparse principal component analysis. *Comput. Stat.* **2009**, *24*, 583–604. [CrossRef]
21. Qi, X.; Luo, R.; Zhao, H. Sparse principal component analysis by choice of norm. *J. Multivar. Anal.* **2013**, *114*, 127–160. [CrossRef]
22. Vichi, M.; Saporta, G. Clustering and disjoint principal component analysis. *Comput. Stat. Data Anal.* **2009**, *53*, 3194–3208. [CrossRef]
23. Mahoney, M.W.; Drineas, P. CUR matrix decompositions for improved data analysis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 697–702. [CrossRef]
24. Trendafilov, N.T. From simple structure to sparse components: A review. *Comput. Stat.* **2014**, *29*, 431–454. [CrossRef]
25. Zhang, Z.; Xuelong, L.; Yang, J.; Li, X.; Zhang, D. A Survey of Sparse Representation: Algorithms and Applications. *IEEE Access* **2015**, *3*, 490–530. [CrossRef]
26. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]

27. Gabriel, K.R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika* **1971**, *58*, 453–467. [CrossRef]

28. Galindo-Villardón, P. Una Alternativa de Representacion Simultanea: HJ-Biplot. *Qüestiió Quad. D'Estad. I Investig. Oper.* **1986**, *10*, 13–23.

29. Nieto-Librero, A.B.; Sierra, C.; Vicente-Galindo, M.; Ruíz-Barzola, O.; Galindo-Villardón, M.P. Clustering Disjoint HJ-Biplot: A new tool for identifying pollution patterns in geochemical studies. *Chemosphere* **2017**, *176*, 389–396. [CrossRef] [PubMed]

30. Cancer Genome Atlas Network. Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* **2012**, *490*, 61. [CrossRef] [PubMed]

31. Rohart, F.; Gautier, B.; Singh, A.; Cao, K.-A.L. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **2017**, *13*, e1005752. [CrossRef]

32. Galindo-Villardón, P.; Cuadras, C.M. Una Extensión del Método Biplot y su relación con otras técnicas. *Publ. Bioestad. Biomatemática* **1986**, *17*, 13–23.

33. Greenacre, M.J. Correspondence analysis. *Wiley Interdisc. Rev. Comput. Stat.* **2010**, *2*, 613–619. [CrossRef]

34. Cubilla-Montilla, M.; Nieto-Librero, A.-B.; Galindo-Villardón, M.P.; Galindo, M.P.V.; Garcia-Sanchez, I.-M. Are cultural values sufficient to improve stakeholder engagement human and labour rights issues? *Corp. Soc. Responsib. Environ. Manag.* **2019**, *26*, 938–955. [CrossRef]

35. Cubilla-Montilla, M.I.; Galindo-Villardón, P.; Nieto-Librero, A.B.; Galindo, M.P.V.; García-Sánchez, I.M. What companies do not disclose about their environmental policy and what institutional pressures may do to respect. *Corp. Soc. Responsib. Environ. Manag.* **2020**, *27*, 1181–1197. [CrossRef]

36. Murillo-Avalos, C.L.; Cubilla-Montilla, M.; Sánchez, M.; Ángel, C.; Vicente-Galindo, P. What environmental social responsibility practices do large companies manage for sustainable development? *Corp. Soc. Responsib. Environ. Manag.* **2021**, *28*, 153–168. [CrossRef]

37. Nieto-Librero, A.B.; Galindo-Villardón, P.; Freitas, A. Package biplotbootGUI: Bootstrap on Classical Biplots and Clustering Disjoint Biplot. Available online: https://CRAN.R-project.org/package=biplotbootGUI (accessed on 4 April 2021).

38. Erichson, N.B.; Zheng, P.; Manohar, K.; Brunton, S.L.; Kutz, J.N.; Aravkin, A.Y. Sparse Principal Component Analysis via Variable Projection. *SIAM J. Appl. Math.* **2020**, *80*, 977–1002. [CrossRef]

39. Cubilla-Montilla, M.; Torres-Cubilla, C.A.; Galindo-Villardón, P.; Nieto-Librero, A.B. Package SparseBiplots. Available online: https://CRAN.R-project.org/package=SparseBiplots (accessed on 4 April 2021).

40. Wickham, H. Ggplot2. *Wiley Interdisc. Rev. Comput. Stat.* **2011**, *3*, 180–185. [CrossRef]

41. Gligorijević, V.; Malod-Dognin, N.; Pržulj, N. Integrative methods for analyzing big data in precision medicine. *Proteomics* **2016**, *16*, 741–758. [CrossRef]

42. McCue, M.E.; McCoy, A.M. The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. *Front. Veter Sci.* **2017**, *4*, 194. [CrossRef]

43. Montilla, M.I.C. Contribuciones al Análisis Biplot Basadas en Soluciones Factoriales Disjuntas Y en Soluciones Sparse. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2019.

44. González García, N. Análisis Sparse de Tensores Multidimensionales. Ph.D. Thesis, Universidad de Salamanca, Salamanca, Spain, 2019.

45. Hernández-Sánchez, J.C.; Vicente-Villardón, J.L. Logistic biplot for nominal data. *Adv. Data Anal. Classif.* **2016**, *11*, 307–326. [CrossRef]

46. Lavit, C.; Escoufier, Y.; Sabatier, R.; Traissac, P. The Act (Statis Method). *Comput. Stat. Data Anal.* **1994**, *18*, 97–119. [CrossRef]

47. Jaffrenou, P.-A. Sur l'analyse Des Familles Finies de Variables Vectorielles: Bases Algébriques et Application à La Description Statistique. Ph.D. Thesis, Thèse de Troisième Cycle, Université de Lyon, Lyon, France, 1978.

48. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311. [CrossRef] [PubMed]

49. Kroonenberg, P.M.; De Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **1980**, *45*, 69–97. [CrossRef]

50. Harshman, R.A. Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multi-Modal Factor Analysis. *Work. Pap. Phon.* **1970**, *16*, 1–84.