

# “Automatic” interpretation of multiple correspondence analysis (MCA) results for nonexpert users, using R programming

“Automatic”  
interpretation  
of MCA results

Stratos Moschidis

*Hellenic Statistical Authority, Piraeus, Greece and  
Department of Accounting and Finance, University of Macedonia,  
Thessaloniki, Greece*

Angelos Markos

*Department of Primary Education, Democritus University of Thrace,  
Alexandroupoli, Greece, and*

Athanasios C. Thanopoulos

*Hellenic Statistical Authority, Piraeus, Greece*

Received 21 July 2022  
Revised 2 September 2022  
Accepted 17 September 2022

## Abstract

**Purpose** – The purpose of this paper is to create an automatic interpretation of the results of the method of multiple correspondence analysis (MCA) for categorical variables, so that the nonexpert user can immediately and safely interpret the results, which concern, as the authors know, the categories of variables that strongly interact and determine the trends of the subject under investigation.

**Design/methodology/approach** – This study is a novel theoretical approach to interpreting the results of the MCA method. The classical interpretation of MCA results is based on three indicators: the projection (F) of the category points of the variables in factorial axes, the point contribution to axis creation (CTR) and the correlation (COR) of a point with an axis. The synthetic use of the aforementioned indicators is arduous, particularly for nonexpert users, and frequently results in misinterpretations. The current study has achieved a synthesis of the aforementioned indicators, so that the interpretation of the results is based on a new indicator, as correspondingly on an index, the well-known method principal component analysis (PCA) for continuous variables is based.

**Findings** – Two (2) concepts were proposed in the new theoretical approach. The interpretative axis corresponding to the classical factorial axis and the interpretative plane corresponding to the factorial plane that as it will be seen offer clear and safe interpretative results in MCA.

**Research limitations/implications** – It is obvious that in the development of the proposed automatic interpretation of the MCA results, the authors do not have in the interpretative axes the actual projections of the points as is the case in the original factorial axes, but this is not of interest to the simple user who is only interested in being able to distinguish the categories of variables that determine the interpretation of the most pronounced trends of the phenomenon being examined.

**Practical implications** – The results of this research can have positive implications for the dissemination of MCA as a method and its use as an integrated exploratory data analysis approach.

**Originality/value** – Interpreting the MCA results presents difficulties for the nonexpert user and sometimes lead to misinterpretations. The interpretative difficulty persists in the MCA’s other interpretative proposals. The proposed method of interpreting the MCA results clearly and accurately allows for the interpretation of its



---

results and thus contributes to the dissemination of the MCA as an integrated method of categorical data analysis and exploration.

**Keywords** Data mining, Data visualization, Machine learning, Multiple correspondence analysis, Dimension reduction, Categorical data

**Paper type** Research paper

---

## 1. Introduction

Dimension reduction methods seek to reduce the number of dimensions [1, 2] in the variable space whilst also preserving the most important structure or relationships between the variables, i.e. without significant loss of information (capturing the essential information) [3]. They have also the advantage of handling and visualizing the results of complex and massive amounts of data [4–7]. Principal Component Analysis (PCA) is a popular method for performing dimension reduction [8] of a set of continuous variables, an effective approach to capture characteristics [9], with the aim of identifying those variables that contribute most to the creation of new, composite variables unlike in feature selection [10], known as principal components or dominant factorial axes. This is achieved via the diagonalization of a symmetric correlation or covariance matrix [11]. To identify which of the original variables contribute most to the creation of each principal component, the coordinates of the projection [1] of the variable points on each factorial axis can only be used, that express the correlation coefficients of each variable with factorial axis. These coordinates express the coefficients of correlation of each variable with the axis.

In this paper we focus on Multiple Correspondence Analysis (MCA), a generalization of PCA for categorical data and a generalization of simple correspondence analysis [12], that is widely used in various scientific fields such as marketing, psychology, health, economics, management and others [13]. The goal of MCA is to describe the associations between the categories of two or more nominal variables in a low-dimensional space containing these categories. Whilst interpreting the results of MCA, users have to identify which column categories have a major contribution to the definition of the factorial axes. MCA has been used as a first step for reducing predictors in classification problems [2], or for receiving new coordinates for performing Hierarchical Clustering on Principal Components [14] or even as a method for meta analyzing literature findings in marketing [15, 16]. Proper interpretation of the MCA results by nonexpert users can often become a difficult task, and consequently it might lead to misinterpretations [17]. The purpose of this paper is to provide nonexpert users with an “automatic” and clear interpretation of the most important points of MCA’s results, via an alternative visualization scheme, based on the construction of the so-called interpretive axes and the corresponding interpretive factorial planes. These proposals come to shrink the existing research gap providing originality to the study. This eliminates the requirement for the user to examine and evaluate the tabular MCA output, as well as looking for numbers and statistics for which additional calculations are frequently required. The proposed scheme is similar to the one used in the context of PCA and users familiar with PCA can easily comprehend this one as well. The novelty of this study is the discovery of a geometrical locus of points on the so-called interpretive plane that improves current alternative approaches on interpreting the most important points of a factorial plane. The paper is organized as follows. [Section 2](#) presents the basic concepts for MCA. [Section 3](#) reviews corresponding literature, discuss about research gap and alternatives addressing the problem. The interpretive axis and the interpretive plane for visualizing MCA results are introduced in [Section 4](#). [Section 5](#) demonstrates the proposed approach on a real data set and compares results. [Section 6](#) includes conclusions, discussion, limitations and implications regarding this research.

## 2. Basic MCA concepts

Let  $\mathbf{X}$ , a  $n \times s$  data matrix, where  $n$  objects or individuals are characterized by  $s$  nominal variables,  $X_k$ . Given that each variable,  $X_k$ , in the original data matrix has  $r_k$  categories, the data matrix is transformed into the so-called indicator or 0-1 matrix,  $\mathbf{Z}$ . Each object in  $\mathbf{Z}$  receives a value of 1 in a single category of each variable (presence) and the value of 0 (absence) in the other categories of the variable. The indicator matrix is of size  $n \times p$ , where  $p = \sum_{k=1}^s r_k$  [2] is the total number of categories of all  $s$  nominal variables. The sum of each row in the indicator matrix is  $s$ , i.e. the total number of variables, while the sum of each column (variable category) is generally different and determines the number of object instances in each variable category. The total sum of rows or columns is  $n \cdot s$ . The values (0 or 1) of each column (variable category) divided by their sum are known as *column profiles* [12]. The set of column profiles is the cloud  $N(J)$  of category points [18]. The category profiles are points of the vector space  $\mathbb{R}^n$  ( $n$  the number of objects). Each sum of a column  $j$  in  $\mathbf{Z}$  is denoted as  $K_{.j}$  and each sum of a row  $i$  as  $K_{i.}$ , respectively. The sum of all variable categories is thus equal to  $n \cdot s$ . If we divide  $K_{.j}$  by the total sum  $n \cdot s$ , the result is known as the mass or weight of category  $j$  and is denoted as  $m_j$ . The column profile points together with their mass constitute the cloud of column profiles and are the ones that carry the information of the data. The center of gravity of the column points is also easily shown to be the point  $n$ -tuple  $g$ , where  $g = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ . Another central concept of MCA is the concept of inertia [12]. Let  $j$  be a category profile point of a variable with sum  $K_{.j}$ . The square of the distance  $d_{\chi^2}^2(j, g)$  of the profile point  $j$  from the center of gravity  $g = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ , is defined with the chi-square distance,  $\chi^2(d_{\chi^2}^2)$  and is given by:  $d_{\chi^2}^2(j, g) = \frac{m_j}{K_{.j}} - 1$ . The inertia  $I_j$  of the point  $j$  with respect to the center of gravity  $g$  is denoted as  $I_j = (\text{mass of } j) \times d_{\chi^2}^2(j, g)$  and it can be easily shown that  $I_j = \frac{1}{s} \cdot (1 - \frac{K_{.j}}{n})$ . The total inertia,  $I$ , of all  $p$  points of the category profiles is then  $I = \frac{p}{s} - 1$  [12]. That is, the inertia of the points depends both on the total number of variable categories,  $p$ , and the total number of categorical variables,  $s$ , in contrast to PCA, where the corresponding inertia is  $s$ , the number of continuous variables. At the core of MCA is the investigation of the structure (form) of the cloud of points  $N(J)$ , that is, the determination of the main directions of inertia. These directions are perpendicular to one another and they are passing through the center of the cloud  $g$ , and optimally describe the cloud  $N(J)$ . These axes, known as factorial axes are obtained via the decomposition of a special variance-covariance matrix and are automatically sorted in descending order, according to their associated eigenvalues,  $\lambda_1 > \lambda_2 > \lambda_3 \dots$ . The inertia of a point  $j$  along a factorial axis,  $a$ , is defined as the square of the Euclidean distance  $d^2(F_a, g)$ , where  $F_a$  is the projection of coefficient of point  $j$  on the factorial axis  $a$ . The sum of the inertias of all points along the factorial axis  $a$ , also known as inertia along the axis or inertia interpreted by the axis, is equal to the eigenvalue  $\lambda_a$  [19]. For example, the inertia along the first factorial axis or the inertia interpreted by this axis, equals to the largest eigenvalue,  $\lambda_1$ . In practice, we usually interpret the first few and most important factorial axes, resulting in a loss of information. The number of important factorial axes can be determined using criteria such as the scree plot and the percentage of explained inertia (variance). The points of interpretive interest on each axis are those which contribute most to the total inertia of the axis. The standard MCA output includes: the standard visualization obtained from the projection of all the points on the factorial axes and/or the factorial planes (e.g. the map created by the first and second factorial axes). These plots, however, do not allow for a direct interpretation of the most important points. Determining the most important points requires considering their inertia along each axis. Based on the (M)CA (Correspondence Analysis) literature, one can

rely on the following aids to interpretation [20, 21] in the form of numerical diagnostics: (a) the coordinate (projection) of point  $j$  on a factorial axis  $a$ ,  $F_a(j)$ . (b) the COR or squared correlation of point  $j$  on axis  $a$ ,  $COR_a(j)$ .

The COR index expresses the amount of inertia of point  $j$  explained by axis  $a$ , that is:

$$COR_a(j) = \frac{m \cdot d^2(F_a, g)}{m \cdot d^2(j, g)} = \left( \frac{d(F_a, g)}{d(j, g)} \right)^2 = \cos^2 \theta [22],$$

(c) the CTR index. The total inertia along an axis or equivalently the inertia interpreted by axis  $a$  is denoted by  $\lambda_a$ . This total inertia is the sum of the inertias of all points in the direction of the axis  $a$ . The ratio of the inertia of the point  $j$  in the direction of the axis  $a$  to the total inertia of the axis  $a$  is called contribution of the point  $j$  and is denoted as  $CTR_a(j)$  and is given by  $CTR_a(j) = \frac{m \cdot F_a^2(j)}{\lambda_a}$  [22]. Obviously, those points with the highest values of the CTR indicator contribute more to the construction of the axis. The contribution indicator shows the points that contribute to the construction of the axis, so it is on these points that the possible interpretation of the axis is based. The combined use of the aforementioned indicators is sufficient for interpretation but a nonexpert user without this knowledge can easily be misled.

### 3. Literature review – alternative approaches of visualizing important points of the MCA maps

The problem of interpreting the results of MCA with visualizations has been a subject of interest and this can be highlighted from a number of published studies [23–26]. Special interest presents the effort of interpreting the most important points on a factorial plane [2, 24–26]. Approaches addressing the issue of correct interpretation (importance [24] or proximity of points [2]) of factorial maps vary from the usage of symbolic means and points' colorization to mathematical transformations. Even though MCA has been implemented in different software and programming languages (e.g. SPSS, SPAD, Python etc.) in this study we focus on  $R$ . In  $R$  there are packages that are developed to perform the MCA and visualize its results and others that produce visualizations aiming to help users interpreting the results. FactoMiner [14] is a package that can compute MCA's results and offers the ability of producing the classical factorial maps. Users can also colorize points according to their contributions that shift the problem of distinguishing the most important points to another variable information. CainterprTools [23] is an  $R$  package that uses additional optical means (dotplots, scatter plots) in order to help users. However, the user has to resort in additional graphical and numerical information that is not a good solution for nonexpert users. CA package [27] includes research of [24, 25] regarding the asymmetrical biplots [26] and through its functions user can pass parameters that transform the points by multiplying points' standard coordinates with their corresponding masses. Consequently, points on visualizations receive information of masses which in case of the factorial axis is adequate for immediate extraction of the most important points. While in the factorial plane the points retain and generalize their informative character, here surface one issue. From the moment that the printed arrows (the ones which the ca package is using and connect a point with the beginning of the axes) are not placed on the same axis their lengths must be measured in order to be safe in interpreting the most important points of the plane. This is a drawback, and it is being resolved through our proposed interpretive plane. We also like to underline the findings of [2] where scholars have presented a solution to the problem of interpreting proximity of points on the factorial maps through the defining of "tolerance distance".

### 4. Methodology – proposal of interpretive axes and interpretive planes

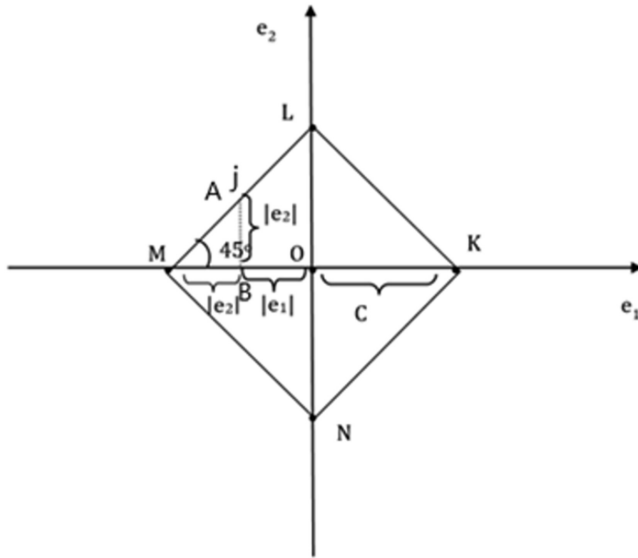
In this section we introduce the notions of interpretive axis and interpretive plane. We consider that each factorial axis corresponds to an interpretive axis that incorporates or combines all the interpretive information of the coordinates,  $F$ , the COR index and the CTR

index. This interpretive axis is defined as follows: The coordinate,  $e_a(j)$ , of a point  $j$  on the interpretive axis  $a$ , is given by  $e_a(j) = \text{sign}(F_a(j)) \cdot \lambda_a \cdot \text{CTR}_a(j)$ , where  $\text{sign}(F_a(j))$  denotes its sign on factorial axis  $a$ . Here the product,  $\lambda_a \cdot \text{CTR}_a(j)$  is the inertia of the point  $j$  in the direction of the factorial axis  $a$ . Since  $e_a(j)$  has the same sign as  $F_a(j)$ , we have information about the direction of each point on the axis. The next step is to define which points of the interpretive axis are the most important which at the same time they will be the most important for interpretation of the corresponding factorial axis. Important interpretive points  $j$  should satisfy the following conditions:

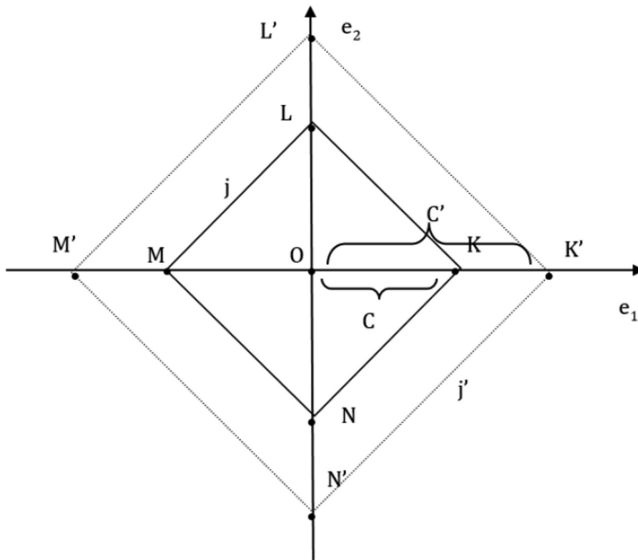
*1st condition:*  $|e_a(j)| > \frac{\lambda_a}{p} \cdot 100$ , where  $\lambda_a$  is the inertia of axis  $a$  and  $p$  the total number of categories, that is important points on the axis are initially considered to have an inertia contribution above the average inertia contribution in the direction of the axis. We multiply by 100 since most software packages report the contributions in %. *2nd condition:* Since we consider that, for a nonexpert user, it is better for an important point to be included on a single factorial axis, in a number of first factorial axes selected by the user ( $k$ ) (e.g first  $k = 5$  factorial axes with the largest percentage of variance), we additionally require that:  $\text{COR}_a(j) = \max(\text{COR}_k(j))$ , where  $a$  is the candidate preferred axis and  $k$  denotes all the selected axes in which point  $j$  satisfies the 1st condition. Conditions 1 and 2 are applied to the 1st factorial axis at first, then the 2nd etc. Second condition addresses cases such as for example the case of a point that satisfies the 1st condition for both the 1st and 2nd factorial axis and at the same time gives the largest value of its inertia on the 2nd factorial axis. At this case this point should then be interpreted as important in the second factorial axis. Consequently, points that satisfy both conditions with the explained sequence are considered as the most important for interpretation of the factorial axis. Therefore, the interpretive axis  $a$ , allows as to evaluate the points with the largest interpretive weight for a factorial axis, based on the value of a single index, the interpretive coordinate  $e_a(j)$ . We can now generalize the notion of the interpretive axis to the interpretive plane and specifically at the first plane which is usually of the greatest interest. The first interpretive plane is created by the first and the second interpretive axes. Important interpretive points  $j$  on the first interpretive plane, which are the most important points of the first factorial plane, should satisfy the following conditions. *1st Condition:*  $|e_1(j)| + |e_2(j)| > \frac{(\lambda_1 + \lambda_2)}{p} \cdot 100$  where  $\lambda_1, \lambda_2$  are the inertias of axis 1 and 2, respectively. First condition is satisfied at least by all the points that have already satisfied the corresponding condition 1 for both the first and second interpretive axes. *2nd Condition:*  $\text{COR}_1(j) + \text{COR}_2(j) = \max(\text{COR}_k(j) + \text{COR}_l(j))$ , where  $k, l$  are the axes that create a plane  $k \times l$ , created by a combination of axes  $k, l$  from a ( $q$ )number first factorial axes selected by the user, in which point  $j$  satisfies 1<sup>st</sup> condition. Points that satisfy 1<sup>st</sup> and 2<sup>nd</sup> conditions are considered as the most important of the first plane. Now, notice that all the points  $j$  on an interpretive plane with  $|e_1(j)| + |e_2(j)| = c$ , with  $c$  a constant, lie at the perimeter of a square KLMN centered at the origin, O, with its diagonals being on the interpretive axes and the semi-diagonal (OK) equal to  $(OK) = c$ . This is shown in [Figure 1](#).

Proof: consider the point  $j$  in position A which has coordinates  $(e_1(j), e_2(j))$ . So its projection (position B) on the first interpretive axis is the distance from the beginning O which is equal to  $|e_1(j)|$ . We then take the point M on the first interpretive axis ([Figure 1](#)) so that its distance from the beginning O is  $c$ . It is now apparent that the length of the line segment BM is equal to  $|e_2(j)|$ . Therefore, the triangle MBA is an isosceles right triangle. Therefore, point  $j$  is generally on the perimeter of the square KLMN. It follows that point  $j$  that is on the square K'L'M'N' with  $(OK') = c' > c$ , is more interpretive than point  $j$  on the first factorial plane. This is shown in [Figures 1](#) and [2](#). The squares in [Figures 1](#) and [2](#) are important for the interpretation. More specifically, the squares on the interpretive plane allow the user to directly compare the contribution of the points. For example, points that belong to the same square have the same contribution regardless of their coordinates. Consequently, points that

ACI



**Figure 1.**  
Defining most important points of an interpretive plane



**Figure 2.**  
Most important points of interpretive plane with "the square" view visualization

are important for the interpretation of the first factorial plane are closer to the most distant squares which are formed from the points of the first plane. This visualization of the interpretive plane eliminates the need to resort to the values of COR and CTR or asymmetric biplot with vector lengths.

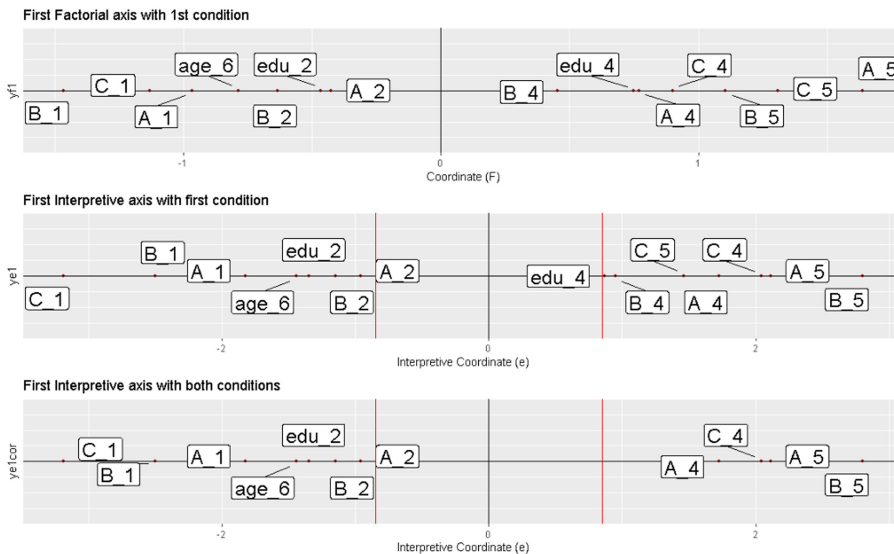
## 5. Applications

In this section we illustrate the visualizations presented in previous section and compare them with existing alternatives using the "wg93" dataset that can be found on the ca package

[27]. More information on dataset can be found on chapter 2 of [12]. MCA was performed with FactoMineR package [14] while libraries “tidyverse” [28], “ggplot” [29–31], “ggrepel” [32], “plotly” [33], “cainertools” [23], “factoextra” [34], “shiny” [35], “DT” [36], “this.path” [37], “soc.ca” [38], “egg” [39] were used as well to produce content that can be found either on the manuscript or on the [supplementary material](#). At the end of the manuscript there is a link that connects with the [supplementary material](#), which can reproduce all visualizations being discussed at this paper and also data tables with numerical evidence for verification. At this section we compare the proposed visualizations with the classical factorial maps while also we compare proposed first interpretive plane with what we consider as the best alternative among approaches that have been discussed on literature section. We encourage readers to download and explore the [supplementary material](#) because it’s important for the completeness of the paper.

In [Figure 3](#) we observe that on first factorial axis, point A\_5 is the most distant point from the right side of the axis, with maximum F value but without the maximum CTR value among points of the right side (A\_5, F: 1.64, CTR: 7.32, e: 2.11) while point B\_5 that has the third biggest F value from the right side is the one with the maximum CTR value (B\_5, F: 1.10, CTR: 9.71, e: 2.80). The same occurs between points B\_1 and C\_1 on the left side of the axis. This can mislead nonexpert user resulting to an erroneous interpretation of the most contributing points of this axis and an erroneous overall characterization of the factorial axis. Our proposed first interpretive axis, on the contrary, guarantees a successful interpretation of the most important points of the first factorial axis.

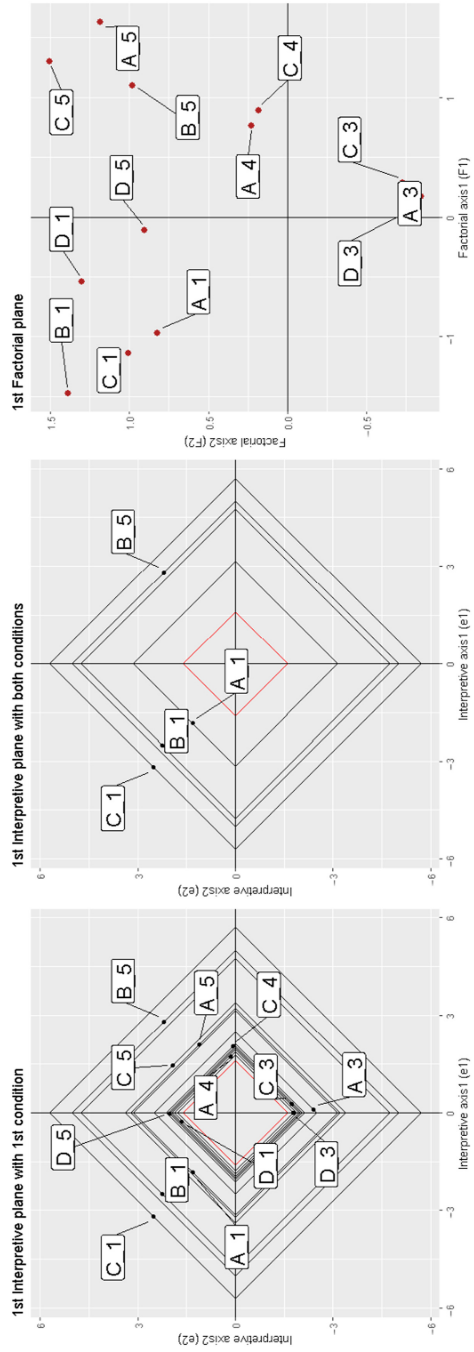
In [Figure 4](#) we observe the comparison between the classic first factorial plane and the proposed first interpretive plane that implements the discovery of the geometrical locus of points (squares). On the first factorial plane we observe that some of the most distant points are the C\_5 and B\_1. Here, a user in order to make a successful interpretation must manipulate MCA’s output and perform additional calculations in order to extract inertias of each point for each of the two axes for the comparison of different points. Therefore, a nonexpert user could easily be led to erroneous interpretations considering C\_5 and B\_1 as



**Note(s):** Only points with contribution greater than the average

**Figure 3.**  
Comparison of first  
factorial axis with first  
interpretive axis





**Figure 4.**  
Comparison of first factorial plane with first interpretive plane

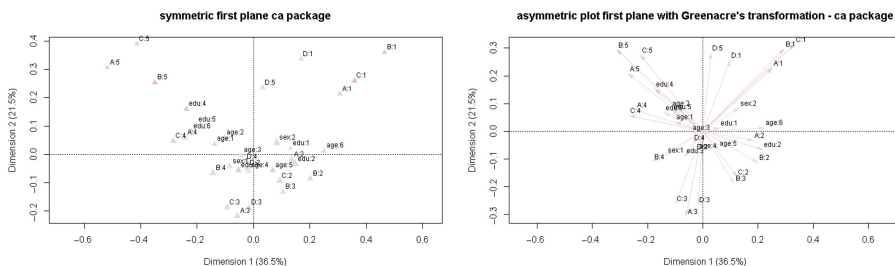


the most important points of the factorial plane. On the other hand our proposed visualization enables a fast and accurate evaluation of the most important points of the factorial plane. As can be seen on first interpretive plane (Figure 4), points C\_5(total inertia on first factorial plane: 3.39) and B\_1(total inertia on first factorial plane: 4.75) are less important for example from points C\_1(total inertia on first factorial plane: 5.71) and B\_5(total inertia on first factorial plane: 5.00) that are in fact the two most important points of the first plane. In our visualization point C\_1 is located on the most distant square while also point B\_5 is located on the second most distant square and this observation “automatically” gives correct interpretation. Similar comparisons can be observed by reader in other points as well.

In Figure 5 we observe the first factorial planes and the one on the right incorporates the asymmetrical biplot theory with the Greenacre’s transformation (contribution biplots). In this transformation points in standard coordinates are multiplied by the square root of the corresponding masses. However, this is a different transformation to ours but we acknowledge that provides improvement over the basic factorial plane’s visualization and the others that are discussed in literature section. This visualization comparing to ours, lacks of the extremely important observation of the geometrical locus of points which eliminates the need for any further calculation to reach to decision about the importance of a point. In Figure 5 (right plot) now notice that points with similar arrows’ lengths need to be measured (with some calculation) and then compared. For example points B\_5 and B\_1 has visually similar arrows’ lengths, therefore is hard to tell which point is more contributing than the other without the evidence from a numerical calculation about their lengths; on the contrary our proposed interpretive plane that incorporates the squares as a geometrical locus of points is a superior and improving approach than the one that is depicted in Figure 5. On our proposed visualization (Figure 4 middle plot) the points B\_5 and B\_1 are easily compared to each other since the B\_5 stands in a more distant square than the B\_1 so it’s more important than B\_1 in the first factorial plane. For a correct interpretation, our proposed visualization requires no additional calculations but only observation.

## 6. Conclusions

This paper proposes a new visualization scheme through the introduction of the interpretive coordinate, the interpretive axis and the interpretive plane which address the problem of finding and interpreting the most important points on the MCA’s factorial maps by nonexpert users. Several scholars through their work have indicated this research gap and provided corresponding solutions. We presented and compared them with our proposal, and we concluded that our interpretive plane with the squares is a quicker and overall better way to find and interpret the most important points of a factorial plane. The originality/novelty of our work is the discovery of the geometrical locus of points which provide immediate optical



**Figure 5.**  
First factorial plane  
(symmetric plot) and  
first asymmetric biplot  
with Greenacre’s  
transformation using  
the ca package

---

identification of most important points. In short, the further a point is from the beginning of the interpretive axis, the more important it becomes for the factorial axis, and also when a point on the interpretive plane is at the perimeter of a more distant square, the more important that point is to the factorial plane. This work can have practical implications through disseminating the use of MCA in a wider audience while it opens a new window for theoretical research on the geometrical relations of the points in factorial maps. Interpretive coordinates as a transformation outcome cannot be used for other analysis methods (e.g. hierarchical clustering) so users must use original MCA coordinates and that can be considered as a limitation. Future research involves more theoretical investigation on geometrical relationships of the points on the factorial maps and providing Information Technology (IT) tools which will help to educate users about this new visualization scheme.

### References

1. Nguyen S, Golas E, Zywiak W, Kennedy K. Dimension reduction in bankruptcy prediction: a case study of North American companies. *Adv Bus Manag Forecast*. 2019; 13: 83-92. Emerald Publishing. doi: [10.1108/S1477-407020190000013010](https://doi.org/10.1108/S1477-407020190000013010).
2. Almeida R, Infantosi A, Suassuna J, Costa J. Multiple correspondence analysis in predictive logistic modelling: application to a living-donor kidney transplantation data. *Comput Methods Programs Biomed*. 2009; 95(2): 116-28. doi: [10.1016/j.cmpb.2009.02.003](https://doi.org/10.1016/j.cmpb.2009.02.003).
3. Ghosh D. Sufficient dimension reduction: an information-theoretic viewpoint. *Entropy*. 2022; 24(2), Art. no. 2. doi: [10.3390/e24020167](https://doi.org/10.3390/e24020167).
4. Diday E. Principal component analysis for bar charts and metabins tables. *Stat Anal Data Mining: ASA Data Sci J*. 2013; 6(5): 403-30. doi: [10.1002/sam.11188](https://doi.org/10.1002/sam.11188).
5. Fernstad SJ, Shaw J, Johansson J. Quality-based guidance for exploratory dimensionality reduction. *Inf Visualization*. 2013; 12(1): 44-64. doi: [10.1177/1473871612460526](https://doi.org/10.1177/1473871612460526).
6. Gardner-Lubbe S, Le Roux NJ, Maunders H, Shah V, Patwardhan S. Biplot methodology in exploratory analysis of microarray data. *Stat Anal Data Mining: ASA Data Sci J*. 2009; 2(2): 135-45. doi: [10.1002/sam.10038](https://doi.org/10.1002/sam.10038).
7. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Phil Trans R Soc A: Math Phys Eng Sci*. 2016; 374(2065): 20150202. doi: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202).
8. Kurita T. Principal component analysis (PCA). In: *Computer vision: a reference guide*. Cham: Springer International Publishing; 2019. p. 1-4. doi: [10.1007/978-3-030-03243-2\\_649-1](https://doi.org/10.1007/978-3-030-03243-2_649-1).
9. Alsaqre F, Almathkour O. Moving objects classification via category-wise two-dimensional principal component analysis. *Appl Comput Inform*. 2020; 18(1/2): 136-50. doi: [10.1016/j.aci.2019.02.001](https://doi.org/10.1016/j.aci.2019.02.001).
10. AlNuaimi N, Masud MM, Serhani MA, Zaki N. Streaming feature selection algorithms for big data: a survey. *Appl Comput Inform*. 2020; 18(1/2): 113-35. doi: [10.1016/j.aci.2019.01.001](https://doi.org/10.1016/j.aci.2019.01.001).
11. Salem N, Hussein S. Data dimensional reduction and principal components analysis. *Proced Comput Sci*. 2019; 163: 292-9. doi: [10.1016/j.procs.2019.12.111](https://doi.org/10.1016/j.procs.2019.12.111).
12. Greenacre M, Blasius J, editors. *Multiple correspondence analysis and related methods*. New York: Chapman and Hall/CRC; 2006. doi: [10.1201/9781420011319](https://doi.org/10.1201/9781420011319).
13. Blasius J, Greenacre M. *Visualization and verbalization of data*. New York: Chapman and Hall/CRC; 2014.
14. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008; 25: 1-18. doi: [10.18637/jss.v025.i01](https://doi.org/10.18637/jss.v025.i01).
15. Saura J, Ribeiro-Soriano D, Palacios-Marqués D. Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research. *Ind Marketing Manag*. 2021; 98: 161-78. doi: [10.1016/j.indmarman.2021.08.006](https://doi.org/10.1016/j.indmarman.2021.08.006).

16. Saura J, Palacios-Marqués D, Ribeiro-Soriano D. Digital marketing in SMEs via data-driven strategies: reviewing the current state of research. *J Small Business Manag.* 2021; 1-36. doi: [10.1080/00472778.2021.1955127](https://doi.org/10.1080/00472778.2021.1955127).
17. Greenacre MJ. Interpreting multiple correspondence analysis. *Appl Stochastic Models Data Anal.* 1991; 7(2): 195-210. doi: [10.1002/asm.3150070208](https://doi.org/10.1002/asm.3150070208).
18. le Roux B, Rouanet H. Multiple correspondence analysis. SAGE Publications; 2010. 163.
19. Moschidis OE. A different approach to multiple correspondence analysis (MCA) than that of specific MCA. *Mathématiques Sciences Humaines Mathematics Soc Sci.* 2009; 186. Art. no. 186. doi: [10.4000/msh.11091](https://doi.org/10.4000/msh.11091).
20. Kaciak E, Louviere J. Multiple correspondence analysis of multiple choice experiment data. *J Marketing Res.* 1990; 27(4): 455-65. doi: [10.1177/002224379002700407](https://doi.org/10.1177/002224379002700407).
21. Lebart L. Techniques de la description statistique: Méthodes et logiciels pour l'analyse des grands tableaux. Paris: Dunod; 1977.
22. Abdi H, Valentin D. Multiple correspondence analysis. *Encyclopedia Meas Stat.* 2007; 2(4): 651-7.
23. Alberti G. CAInterprTools: an R package to help interpreting Correspondence Analysis' results. *SoftwareX.* 2015; 1(2): 26-31. doi: [10.1016/j.softx.2015.07.001](https://doi.org/10.1016/j.softx.2015.07.001).
24. Greenacre M. Contribution biplots. *J Comput Graphical Stat.* 2013; 22(1): 107-22. doi: [10.1080/10618600.2012.702494](https://doi.org/10.1080/10618600.2012.702494).
25. Gabriel KR, Odoroff CL. Biplots in biomedical research, *Statistics in medicine.* 1990; 9(5), 469-85. doi: [10.1002/sim.4780090502](https://doi.org/10.1002/sim.4780090502).
26. Greenacre MJ. Biplots in correspondence analysis. *J Appl Stat.* 1993; 20(2): 251-69.
27. Nenadić O, Greenacre M. Correspondence analysis in R, with two- and three-dimensional graphics: the ca package. *J Stat Softw.* 2007; 20: 1-13. doi: [10.18637/jss.v020.i03](https://doi.org/10.18637/jss.v020.i03).
28. Wickham H, *et al.* Welcome to the tidyverse. *J open source Softw.* 2019; 4(43): 1686.
29. Wickham H. Data analysis. In: *ggplot2*. Springer; 2016: 189-201.
30. Wickham H. *ggplot2: elegant graphics for data analysis*. In: 2009. Corr. 3rd printing 2010 edition. 1st ed. New York: Springer; 2010.
31. Wickham H, Grolemund G. R for data science: import, tidy, transform, visualize, and model data. Sebastopol, CA: O'Reilly Media; 2016.
32. Slowikowski K. Ggrepel: automatically position non-overlapping text labels with 'ggplot2'. 2021. R Package Version 0.9, CRAN repository. 2019; 1.
33. Sievert C. Interactive web-based data visualization with R, plotly, and shiny. *Plotly-r.com*; 2022. Available from: <https://plotly-r.com/>
34. Kassambara A. Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, Factoextra. 2017; 2. Sthda.
35. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. Package 'shiny,'; 2015. Available from: <http://citereerx.ist.psu.edu/viewdoc/download>
36. Xie Y, Cheng J, Tan X. DT: a wrapper of the JavaScript library 'DataTables,'. R Package Version 0.4, CRAN repository. 2018.
37. Simmons A. This.path: get executing script's path, from 'RStudio', 'Rgui', 'rscript' (shells including windows command-line/unix terminal), and 'source.'. 2022; 11. Available from: <https://CRAN.R-project.org/package=this.path>
38. Larsen AG, Andrade S. Package 'soc. ca,'; 2016. Available from: <https://cran.r-project.org/web/packages/soc.ca/soc.ca.pdf>
39. Auguie B. Extensions for 'ggplot2': custom geom, custom themes, plot alignment, labelled panels, symmetric scales, and fixed panel size [R package egg version 0.4.5]. *Cran.r-project.org.* 2022. Available from: <https://cran.r-project.org/web/packages/egg/index.html>

---

ACI

**Supplementary material**

Supplementary material for this article can be found online at: <https://drive.google.com/drive/folders/1Xxz4RxSaltYcwJLi2jiCbR4a-WjNklr4?usp=sharing>

**Corresponding author**

Stratos Moschidis can be contacted at: [smos@statistics.gr](mailto:smos@statistics.gr)

---

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)