Visualization of Correlation Tables by Positive/Negative Threshold for Coefficients Significance

Mullat, Joseph

9 March 2025

Visualization of Correlation Tables by Positive/Negative Threshold for Coefficients Significance

Abstract

This work introduces a novel approach to the visualization of correlation tables, guided by positive and negative significance thresholds of correlation coefficients. Traditional methods for visualizing correlation matrices often rely on heuristic color schemes, which lack a robust analytical foundation. To address this limitation, we propose a method that constructs and analyzes an ordered sequence of so called *momentums*, separating positive and negative correlations based on their significance. By leveraging mathematical principles of an optimal solutions, this approach enhances the clarity and interpretability of correlation patterns.

Keywords:correlation, classification, coefficient, defining, sequence

1. Intoduction

We establish a visualization algorithm that identifies significance levels for positive and negative correlations, enabling clear differentiation of meaningful patterns within the matrix. The key feature of the method is the ***defining sequence***, constructed by pairing ordered correlation coefficients with their statistical momentums and analyzing its single-peaked property. Theoretical justification for this property is provided via the Basic Visualization Theorem, which demonstrates that the local maximum of the defining sequence corresponds to the global maximum of a specific function defined over all subsets of the correlation matrix.

The proposed method draws inspiration from Mullats (1971) foundational work on single-peaked sequences but adapts the defining sequence in reverse order for the purposes of visualization. The approach offers a systematic, mathematically supported framework for representing correlation tables, with mirrored extensions for negative correlations to ensure comprehensive analysis.

The visualization of correlation tables is a fundamental yet highly nuanced task in data analysis and statistical exploration. A correlation table represents a matrix of correlation coefficients, each ranging from -1 to $+1$, capturing the linear relationships between pairs of variables in a dataset. The conventional approach to visualizing these tables involves mapping numerical values onto a color palette. For instance, negative correlations are often depicted with shades of lighgt blue that gradually grow to strong blue as the magnitude of the negative correlation decreases, while positive correlations might be represented by a gradient transitioning from light pink to deep red or pink for higher positive correlations. The strength of the correlation is emphasized through the intensity of the color, creating a vivid heatmap-style representation.

While such visualizations are visually appealing and heuristic, they often lack analytical depth and rigor. These methods provide a quick and intuitive way to identify patterns but fall short of delivering a comprehensive or objective understanding of the data. For researchers aiming to uncover the intricacies of a system, such heuristic approaches can be unsatisfactory. This is because they prioritize subjective interpretations and aesthetic considerations over the statistical robustness and objectivity required to address complex research questions.

1.1. The Need for a Rigorous Approach

For those aiming to combine the intuitive appeal of visualization with strict statistical analysis, more advanced methods are required. Such methods must balance clarity with analytical depth, allowing researchers to derive meaningful insights from complex datasets. Achieving this balance often necessitates employing the entire arsenal of statistical tools, including both classical and modern approaches to data analysis. However, this endeavor can be resource-intensive, demanding significant effort and time, especially when decisions need to be made swiftly based on the analysis.

Statistical methods are traditionally designed for scenarios where the underlying numerical distributions are well-understood. However, this assumption is not always realistic. In many practical applications, the distribution of the data is unknown, adding layers of complexity to the analysis. This uncertainty can compromise the reliability of the results and, in some cases, lead to conclusions that contradict intuitive understandings of the system under study.

1.2. Simplicity as a Strength

In situations where the complexity of statistical methods outweighs their utility, it may be prudent to adopt simpler visualization techniques. The simplicity of these methods can sometimes offset the disadvantages of forgoing more sophisticated approaches. By providing a straightforward and transparent representation of the data, simple visualizations can serve as a practical alternative, particularly when the goal is to communicate insights effectively rather than conduct an exhaustive analysis.

The simplicity of data visualization does not necessarily equate to a lack of rigor. By grounding visualization techniques in a solid mathematical framework, it is possible to maximize the quality of the visualization while retaining its interpretability. For example, optimization techniques can be used to enhance the clarity and informativeness of visualizations, ensuring that they are both accessible and analytically meaningful.

1.3. A Proposed Method for Visualizing Correlation Tables

In this study, we propose a novel method for visualizing correlation tables that balances simplicity with mathematical rigor. Our approach is rooted in the optimization of visualization quality metrics, ensuring that the resulting visualizations are not only aesthetically pleasing but also analytically robust. This method addresses the limitations of traditional heuristic approaches, providing researchers with a tool that is both intuitive and grounded in a strict mathematical foundation.

By focusing on the maximization of visualization quality functions, our method ensures that the patterns and relationships within the data are represented as clearly as possible. This approach bridges the gap between intuitive visualization and rigorous statistical analysis, offering a practical solution for researchers navigating the challenges of modern data analysis.

1.4. Supporting Literature and References

Several studies and tools have explored the visualization of correlation matrices, emphasizing both heuristic and mathematically rigorous approaches:

- **Heatmap Visualization Techniques**: The use of heatmaps for correlation matrices is discussed in depth in [Davila et al., 2023; Pleil et al., 2011; Henderson and Velleman, 1981.], where the effectiveness of color gradients in representing relationships is evaluated.
- **Challenges of Unknown Distributions**: Research on the implications of unknown data distributions can be found in [Wendy & Liu, 2018; Johnstone and Titterington (2009], highlighting the limitations of traditional statistical methods in such scenarios.
- **Optimization in Data Visualization**: The mathematical foundations of optimizing visualization quality are explored in [Midway, 2020; Sun et al., 2023], providing a basis for the proposed method.

1.5. Practical Applications of Simple Visualization

Case studies demonstrating the utility of simple visualization techniques in various fields are presented in [Schwabish, 2021; Mullat, 2023; Jöreskog, 1978], reinforcing the value of simplicity in data communication.

By integrating insights from these studies, we aim to contribute a practical and theoretically sound method for visualizing correlation tables, catering to the needs of both intuitive understanding and rigorous analysis.

2. Explanation of Our Simple Method for Visualizing the Correlation Matrix

In presenting the core principles of our simple method for visualizing correlation matrices, we aim to minimize the use of complex mathematical notations and instead articulate the method in a more accessible verbal form. This approach is intended to make the mathematical underpinnings of the visualization algorithm comprehensible without compromising the rigor necessary for proving its main theorem, which is provided in a clear verbal explanation.

Let us begin by defining the structure of a correlation matrix. A correlation table is represented as a matrix $\|r_{i,j}\|$, where $-1 \le r_{i,j} \le 1$, with $i = \overline{1,n}$ and $j = \overline{1,n}$. The matrix is a square in shape, consisting of $n \times n$ elements, where each cell is identified by the row index i and the column index j. The diagonal elements of the correlation matrix are equal to 1, representing the perfect self-correlation of each variable. To simplify indexing, we introduce a single cell number k, calculated as $k = (i-1) \times n + j$.

For our purposes, we propose a different perspective on the correlation matrix: instead of treating it as a block structure, we transform it into a linear list of values r_k. Additionally, we separate positive correlations from negative correlations, enabling independent visualization for each. This separation allows us to treat positive and negative correlations distinctly, ensuring clarity and precision in the visualization process.

2.1. Algorithm for Finding Correlation Significance Levels

To effectively visualize the correlation matrix, we introduce an algorithm for determining significance levels for both positive and negative correlations. These significance levels serve as thresholds that filter which correlations are included in the final visualization.

Defining Positive and Negative Significance Levels

The algorithm distinguishes between two significance levels:

- Positive significance level (sp): A positive value in the range $0 \le sp \le 1$.
- Negative significance level (sn): A negative value in the range $-1 \le sn \le 0$.

These thresholds allow us to separately assess the significance of positive and negative correlations.

Visualization of Positive Correlations

Positive correlations that meet or exceed the positive significance level $(r_k \ge sp)$ are visualized using a designated color scheme. Cells that fall below this threshold are excluded from the visualization pattern, ensuring that only significant positive correlations are highlighted.

Visualization of Negative Correlations

Negative correlations are handled similarly but with a focus on values that are equal to or more negative than the negative significance level $(r_k \le sn)$. These correlations are visualized using a different color scheme, distinct from the one used for positive correlations.

Exclusion of Non-Significant Cells

Cells that do not meet either the positive or negative significance criteria are excluded from the visualization. This exclusion reduces visual clutter and ensures that the resulting matrix highlights only the most meaningful correlations.

Color Coding for Enhanced Clarity

To facilitate intuitive interpretation, the visualization employs two separate color palettes: one for positive correlations and another for negative correlations. For instance, positive correlations might be represented with warm tones (e.g., shades of red or pink), while negative correlations could use cool tones (e.g., shades of blue). The intensity of the color reflects the strength of the correlation, with higher absolute values corresponding to deeper or more saturated colors.

2.2. The Value of Significance-Based Visualization

The introduction of significance levels addresses a common challenge in visualizing correlation matrices: the overwhelming amount of information that can obscure meaningful patterns. By focusing on significant correlations, the visualization becomes more interpretable, enabling researchers to quickly identify key relationships. Furthermore, the separation of positive and negative correlations into distinct visual channels ensures that contrasting patterns are not conflated, providing a clearer overall picture.

Our approach not only simplifies the visualization process but also provides a mathematical basis for determining what is deemed significant. This balance of simplicity and rigor makes the method accessible to a broad audience, from researchers conducting exploratory data analysis to practitioners seeking to communicate findings effectively.

In future work, we aim to extend this framework by incorporating additional metrics for correlation strength and significance, as well as exploring alternative visualization techniques that may enhance interpretability in specific applications.

This extended version emphasizes clarity, expands on key concepts, and provides a structured explanation of the algorithm and its value. If youd like, I can further refine it or help you locate relevant references for supporting claims. Finally, having assembled all the necessary tools and concepts for explaining our algorithm, we will now outline its main provisions.

2.3. Main Procedure

Let us begin by considering the set of all positive correlation coefficients in our correlation matrix, denoted as r_k. Arrange these coefficients in descending order, which we will denote as $\vec{r}_k$. Without loss of generality, we assume that the ordered sequence satisfies $\vec{r}_1 \geq \vec{r}_2 \geq \vec{r}_3 \geq \ldots$

The core idea of our algorithm is to use this descending sequence, $\vec{r}_1 \geq \vec{r}_2 \geq \vec{r}_3 \geq \ldots$, together with the sequence $k = 1, 2, 3, \ldots$, to construct a new sequence. Specifically, we pair each coefficient $\vec{r}_k$ with its corresponding index k and compute the products $\vec{r}_1 \times 1, \vec{r}_2 \times 2, \vec{r}_3 \times 3 \ldots$. This results in a sequence of values $\vec{r}_k \times k$, which we refer to as the ***defining sequence***. The indicators $\vec{r}_k \times k$ are also called in statistic as moments.

The defining sequence exhibits a characteristic behavior: its values either increase or decrease as one progresses along the sequence. Such sequences are common in numerical datasets and exhibit the properties of $\cap$-**single-peak sequences**, which are characterized by having exactly one local maximum when traversed from left to right.

2.4. The Basic Visualization Theorem

As the title of this section suggests, the reader might expect long and intricate mathematical constructions, culminating in the formulation of a theorem or proposition. However, this is not the case. We aim to convince the reader of the validity of our statements without resorting to exhaustive mathematical analysis. Instead, we will present the theorem on the defining sequence in the form of a proposition that "*Local Maximum in the Defining Sequence also Corresponds to the Global Maximum of a Particular Function Defined Over the System of All Subsets of the Defining Sequence.*"

Preliminary Construction

Let us begin with the following setup. Consider a set W consisting of cells in the correlation matrix that contain positive correlation coefficients. From this set W, we can derive the collection of all possible subsets H, where $H \subseteq W$.

For any subset H, we restrict our attention to the correlation coefficients r_k corresponding to the cells within H. We then scale these coefficients proportionally to the size of the subset H, also referred to as the cardinality of H, denoted $|H|$. Specifically, if a cell $r_k = r_{i,j}$ belongs to H, we define the scaled value as: $\pi(k,H) = r_k \times |H|$ (also known as statistcal moments).

Next, based on this scaled set of correlation moments $\pi(k,H)$, we define a function $F(H)$ as the minimum value $\pi(k,H)$ of over all k in H: $F(H) = \min_k \pi(k,H)$.

This function $F(H)$ evaluates the scaled minimum of correlation moments over any subset H of W. When considering the function $F(H)$, the scaling by $|H|$ ensures that the global maximum of $F(H)$ aligns with the subset of W that corresponds to this local maximum in the defining sequence. This is because the defining sequence captures the interplay between the correlation coefficients and their relative positions, ensuring that the subset maximizing $F(H)$ is precisely the one associated with the peak of the defining sequence.

This reasoning confirms that the local maximum of the defining sequence is not merely a numerical artifact but has a direct and meaningful correspondence to the global maximum of the function $F(H)$.

Theorem

The local maximum (or peak) of the defining sequence constructed from positive correlation coefficients coincides with the global maximum of the function $F(H)$ over all possible subsets H of the set W.

Main Outline of the Proof

Contrary to the claim that a global maximum of function $F(H)$ is achieved at the defining sequence peak for some index k^*, assume there exists a subset H' of correlation cells in the correlation matrix such that the function $F(H')$ attains a value greater than the local maximum $F(H^*)$, $F(H') > F(H^*)$, where $H^* = \{\vec{r}_1, \vec{r}_2, ..., \vec{r}_{k^*}\}$.

If we combine the set H^*, which corresponds to the peak correlation coefficient at k^*, with the cells in H', the resulting combined set $H^* \cup H'$ can be extended to some subset $\widetilde{H} = \{\vec{r}_1, \vec{r}_2, ..., \vec{r}_{k'}\}$, where $k' > k^*$. The set $H^* \cup H'$ lies within the subset $\widetilde{H}$, $H^* \cup H' \subseteq \widetilde{H}$.

On this extended subset $\widetilde{H}$, the function $F(\widetilde{H})$ would allegedly satisfy the inequality $F(\widetilde{H}) > F(H^*)$, owing to the monotonicity of our statistical moments' definition. This inequality implies that the defining sequence would have at least one additional local maximum greater than $F(H^*)$.

Such a scenario violates the single-peakedness property of the defining sequence. Thus, the initial assumption is invalid, and the theorem holds.

This reasoning builds on the ideas originally presented by Mullat (1971) in his work "On a Maximum Principle for Certain Functions of Sets", where he established the fundamental properties of single-peaked sequences. However, it is important to note that the defining sequence discussed here is in reverse order compared to the sequence analyzed in Mullat's proof. For further details, see Mullat's original article from 1971.

Note on Negative Correlations

The same reasoning applies to negative correlations with appropriate modifications. Specifically, the signs of the inequalities are reversed, and the minimum used to define $F(H)$ is replaced by a maximum. Therefore, all conclusions regarding the positive significance levels of correlations are mirrored for the negative significance levels in our correlation matrix.

3. Visualization Aspects and Illustration

We have now reached the stage where we introduce the thoughtful (and technically inclined) reader to the mechanics of the proposed method for visualizing correlation coefficients—or any other interrelated elements that may need to be analyzed to support decision-making. While we assume the reader has some familiarity with data analysis in Microsoft Excel, which is highly convenient for both computation and visualization, the proposed method is accessible through macro from http://www.datalaundering.com/download/joreskog.xls (Excel spreadsheet) even to those who lack deep knowledge of Excel. A layman-friendly procedure is provided to simplify understanding without requiring immersion in technical complexities.

3.1. Layman's Procedure for Visualizing a Set of Numerical Values

Consider a given set of numerical values representing parameters or similar quantities for analysis. For example, take the set of 10 numbers:

$$\langle 1, 0.81, 0.64, 0.49, 0.36, 0.25, 0.16, 0.09, 0.04, 0.01 \rangle .$$

Step 1: Arrange the data

Enter these numbers as a row in Microsoft Excel. Beneath this row, enter another row containing sequential indices: $\langle 1, 2, 3, 4{,}5, 6, 7, 8, 9, 10 \rangle$.

Step 2: Compute the products or moments

In a third row, compute the product of these values in the first row with their corresponding indices in the second row:

(1×1, 0.81×2, 0.64×3, 0.49×4, 0.36×5, 0.25×6, 0.16×7, 0.09×8, 0.04×9, 0.01×10).

This results in the following sequence of products (often referred to as statistical moments): **(1, 1.62, 1.92, 1.96, 1.80, 1.50, 1.12, 0.72, 0.36, 010).**

Step 3: Identify the local maximum

In this sequence, the local maximum occurs at the 4th number, 1.96, as you progress from left to right. This value represents the level of significance (**sp**) that is of primary interest.

Step 4: Visualize the results

To incorporate this significance level into a visualization, color the first four numbers of the sequence **(1, 1.62, 1.92, 1.96, 1.80, 1.50, 1.12, 0.72, 0.36, 010),** in pink, leaving the remaining numbers uncolored.

Table 1

Index (k)	1	2	3	4	5	6	7	8	9	10
Indicators	1	0,81	0,64	0,49	0,36	0,25	0,16	0,09	0,04	0,01
Moments	1	1,62	1,92	1,96	1,8	1,5	1,12	0,72	0,36	0,1

This table illustrates the fundamental steps without requiring detailed familiarity with Excel functions. The remaining task is purely technical—how to automate the procedure within Microsoft Excel.

3.2. Illustration of the Visualization Procedure on Real Data

For the thoughtful reader, this section addresses the applicability and efficiency of the proposed method in real-world scenarios involving correlation coefficient analysis. To this end, we identified a noteworthy article in the public domain (Jöreskog, 1978) titled "***Structural Analysis of Covariance and Correlation Matrices***", which contains a correlation table well-suited for our purposes. It is not necessary to delve into the detailed analysis of the table presented in Jöreskogs article. Instead, we will focus on visualizing the data using our method, supplemented by a few explanatory comments.

It is evident that visualizing the matrix using these significance thresholds partially aligns with the diagonalization of the correlation matrix. This diagonalization, in turn, reflects the classification of the 24 parameters under analysis based on their linear dependence on one another. Through this visualization method, the correlation table is effectively divided into four parameter groups, as follows (with group numbers borrowed from Jöreskogs article):

- Group No.1: 01, 02, 03, 04, 23; • Group No.2: 05, 06, 07, 08, 09, 22;
- Group No.3: 10, 11, 12, 13, 24; • Group No.4: 14, 15, 16, 17, 18, 19, 20, 21

This diagonalization technique for analyzing relationships between parameters is a distinct aspect of data multivariate analysis and classification (Ishii, et al., 2021; Mirkin, 2011; Frey and Võhandu, 1966), or as demonstrated in Braverman Readings, 2017.

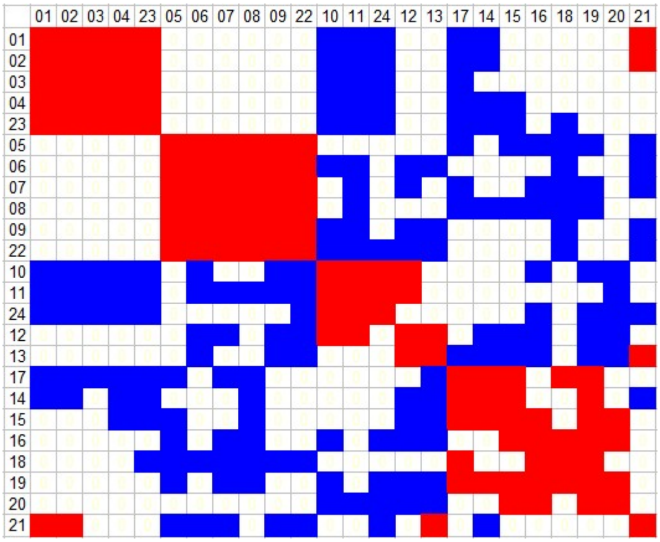
Tables visualization using macro from:
http://www.datalaundering.com/download/joreskog.xls

Table 2

Negative significance level →		-0,412																							
Positive significance level →		0,671																							
		01	02	03	04	23	05	06	07	08	09	22	10	11	24	12	13	17	14	15	16	18	19	20	21
Visual-Perception	01	1,000	0,979	0,929	0,915	0,831	-0,329	-0,337	-0,256	0,052	-0,328	0,077	-0,428	-0,592	-0,795	0,045	0,540	-0,537	-0,574	-0,360	0,328	0,097	0,114	0,076	0,842
Cubes	02	0,979	1,000	0,983	0,959	0,899	-0,234	-0,183	-0,136	0,131	-0,172	0,259	-0,600	-0,731	-0,890	-0,158	0,373	-0,521	-0,479	-0,284	0,399	0,015	0,190	0,254	0,720
Paper-Form-Board	03	0,929	0,983	1,000	0,947	0,901	-0,201	-0,087	-0,084	0,131	-0,075	0,375	-0,733	-0,801	-0,956	-0,327	0,199	-0,434	-0,336	-0,153	0,502	0,007	0,308	0,424	0,603
Flags	04	0,915	0,959	0,947	1,000	0,985	0,049	0,071	0,142	0,404	0,081	0,460	-0,633	-0,853	-0,833	-0,252	0,340	-0,698	-0,576	-0,444	0,201	-0,267	-0,007	0,230	0,571
Series-Completion	23	0,831	0,899	0,901	0,985	1,000	0,215	0,242	0,310	0,543	0,252	0,595	-0,675	-0,917	-0,796	-0,353	0,251	-0,743	-0,563	-0,474	0,119	-0,418	-0,077	0,267	0,430
General-Information	05	-0,329	-0,234	-0,201	0,049	0,215	1,000	0,929	0,988	0,923	0,926	0,719	-0,125	-0,389	0,239	-0,389	-0,232	-0,517	-0,222	-0,460	-0,661	-0,969	-0,633	-0,021	-0,627
Paragraph-Comprehension	06	-0,337	-0,183	-0,087	0,071	0,242	0,929	1,000	0,965	0,815	1,000	0,882	-0,416	-0,526	0,023	-0,697	-0,542	-0,302	0,073	-0,149	-0,344	-0,863	-0,301	0,346	-0,750
Sentence-Completion	07	-0,256	-0,136	-0,084	0,142	0,310	0,988	0,965	1,000	0,929	0,964	0,817	-0,275	-0,510	0,098	-0,515	-0,306	-0,510	-0,176	-0,399	-0,544	-0,964	-0,529	0,119	-0,623
Word-Classification	08	0,052	0,131	0,131	0,404	0,543	0,923	0,815	0,929	1,000	0,814	0,740	-0,233	-0,605	-0,017	-0,318	0,039	-0,793	-0,519	-0,685	-0,625	-0,989	-0,683	-0,072	-0,293
Word-Meaning	09	-0,328	-0,172	-0,075	0,081	0,252	0,926	1,000	0,964	0,814	1,000	0,888	-0,427	-0,536	0,011	-0,704	-0,543	-0,303	0,073	-0,147	-0,335	-0,861	-0,294	0,355	-0,745
Problem-Reasoning	22	0,077	0,259	0,375	0,460	0,595	0,719	0,882	0,817	0,740	0,888	1,000	-0,779	-0,849	-0,450	-0,869	-0,516	-0,367	0,036	-0,080	0,017	-0,727	-0,019	0,618	-0,466
Addition	10	-0,428	-0,600	-0,733	-0,633	-0,675	-0,125	-0,416	-0,275	-0,233	-0,427	-0,779	1,000	0,871	0,868	0,880	0,506	0,080	-0,227	-0,278	-0,613	0,170	-0,531	-0,890	0,079
Code	11	-0,592	-0,731	-0,801	-0,853	-0,917	-0,389	-0,526	-0,510	-0,605	-0,536	-0,849	0,871	1,000	0,799	0,692	0,149	0,558	0,250	0,228	-0,216	0,516	-0,075	-0,563	-0,068
Arithmetic-Problems	24	-0,795	-0,890	-0,956	-0,833	-0,796	0,239	0,023	0,098	-0,017	0,011	-0,450	0,868	0,799	1,000	0,529	0,080	0,194	0,047	-0,126	-0,696	-0,100	-0,541	-0,665	-0,424
Counting-Dots	12	0,045	-0,158	-0,327	-0,252	-0,353	-0,389	-0,697	-0,515	-0,318	-0,704	-0,869	0,880	0,692	0,529	1,000	0,817	-0,113	-0,489	-0,419	-0,433	0,326	-0,449	-0,913	0,543
Straight-Curved-Capitals	13	0,540	0,373	0,199	0,340	0,251	-0,232	-0,542	-0,306	0,039	-0,543	-0,516	0,506	0,149	0,080	0,817	1,000	-0,597	-0,868	-0,751	-0,412	0,043	-0,547	-0,796	0,804
object-Number	17	-0,537	-0,521	-0,434	-0,698	-0,743	-0,517	-0,302	-0,510	-0,793	-0,303	-0,367	0,080	0,558	0,194	-0,113	-0,597	1,000	0,916	0,935	0,561	0,710	0,720	0,369	-0,342
Word-Recognition	14	-0,574	-0,479	-0,336	-0,576	-0,563	-0,222	0,073	-0,176	-0,519	0,073	0,036	-0,227	0,250	0,047	-0,489	-0,868	0,916	1,000	0,961	0,581	0,431	0,741	0,646	-0,593
Number-Recognition	15	-0,360	-0,284	-0,153	-0,444	-0,474	-0,460	-0,149	-0,399	-0,685	-0,147	-0,080	-0,278	0,228	-0,126	-0,419	-0,751	0,935	0,961	1,000	0,762	0,628	0,885	0,672	-0,351
Figure-Recognition	16	0,328	0,399	0,502	0,201	0,119	-0,661	-0,344	-0,544	-0,625	-0,335	0,017	-0,613	-0,216	-0,696	-0,433	-0,412	0,561	0,581	0,762	1,000	0,671	0,976	0,762	0,203
Number-Figure	18	0,097	0,015	0,007	-0,267	-0,418	-0,969	-0,863	-0,964	-0,989	-0,861	-0,727	0,170	0,516	-0,100	0,326	0,043	0,710	0,431	0,628	0,671	1,000	0,697	0,081	0,418
Figuree-Word	19	0,114	0,190	0,308	-0,007	-0,077	-0,633	-0,301	-0,529	-0,683	-0,294	-0,019	-0,531	-0,075	-0,541	-0,449	-0,547	0,720	0,741	0,885	0,976	0,697	1,000	0,772	0,030
Deduction	20	0,076	0,254	0,424	0,230	0,267	-0,021	0,346	0,119	-0,072	0,355	0,618	-0,890	-0,563	-0,665	-0,913	-0,796	0,369	0,646	0,672	0,762	0,081	0,772	1,000	-0,327
Numerical-Puzzles	21	0,842	0,720	0,603	0,571	0,430	-0,627	-0,750	-0,623	-0,293	-0,745	-0,466	0,079	-0,068	-0,424	0,543	0,804	-0,342	-0,593	-0,351	0,203	0,418	0,030	-0,327	1,000

Table 3

The table above and the one to the right both correspond precisely to the **24×24** correlation matrix derived from the correlation coefficients in Jöreskogs article. The cells are color-coded as follows: pink for positive correlation coefficients ≥ the positive significance threshold (**sp=+0.671**), blue for negative correlation coefficients ≤ the negative significance threshold (**sn=0.412**), and uncolored for insignificant values.



4. Concuding Remarks

In conclusion, it is important to emphasize that the application of the described method in fields such as data analysis and data visualization should not be narrowly confined to correlation matrices. While correlation matrices provide a convenient and illustrative example, the method is versatile and can be applied to any numerical data reflecting systems of measurements, statistical indicators, or data sets presented in various forms, such as time series, sequences, or other arrangements.

Moreover, the method can extend to analyzing deviations from the significance levels of indicators in any direction. This includes exploring variations, moments, and other types of statistical deviations, limited only by the creativity and expertise of the researcher. Such flexibility allows for an expansive range of applications, as the method is not constrained by predefined structures or formats.

The key takeaway is that, although the method is demonstrated here using a correlation table as an example, it is fundamentally general in nature. This broader applicability was the primary motivation behind presenting this methodological approach, offering a valuable tool for thoughtful readers seeking robust and adaptable techniques for data analysis.

Appendix

Excel Macro (Ctrl-s) – Short Description

http://www.datalaundering.com/download/joreskog.xls
(Visualization Implementation in Microsoft Excel 2000)

This macro combines several functionalities, primarily for analyzing stock data, various data tables, correlation matrices, and more. **Two rows at the top of the spreadsheet must remain available** for macro messages and significance thresholds. Below is an analysis of its structure and operation.

How to Download and Implement the Macro

1. In any spreadsheet (new or existing), go to the **Functions** section and then the **Macros** subsection.
2. Create a dummy macro and copy the source code into it.
3. Edit the dummy macro so that the first line reads: ***vba (copy) Sub Significant()***
4. Assign the letter "**s**" to the macro properties, enabling the macro to work with the **Ctrl-s** shortcut.
5. Ensure the macro is applied to the spreadsheet range containing the indicators to be visualized.

Overview of Key Features

1. **Selection Processing**
   - Operates on a user-selected range of cells.
   - Differentiates between positive and negative numbers, processing them separately.
2. **Shell Sort Implementation**
   - Uses the Shell Sort algorithm to sort numbers in the selected range.
   - Sorts positive numbers in descending order and negative numbers in ascending order.
3. **Significance Analysis (Ctrl-s)**
   - Identifies "significant levels" for positive and negative numbers based on their sorted values.
   - Applies conditional formatting to indicate significance.
4. **Conditional Formatting**
   - Positive and negative cells are formatted differently (e.g., bold fonts, colored backgrounds).
   - Provides visual cues to distinguish significant positive and negative indicators.
5. **Error Handling**
   - Includes error checks for various conditions, such as:
   - Selection of only one cell.
   - Fewer than two positive or negative numbers.
   - Presence of non-numeric values in the selection.

6. **Output**
   - Outputs significant positive and negative levels to specific cells.
   - Transfers the formatting of the first "significant" positive and negative cells to designated positions.

Strengths

1. **Robust Sorting Algorithm**
   - Efficiently handles moderate-sized data ranges with a gap-based Shell Sort implementation.
2. **Dynamic Range Handling**
   - Adapts dynamically to the user-selected range.
3. **Enhanced Usability**
   - Provides clear, user-friendly outputs, including visual formatting and indication of significant values.
4. **Error Handling**
   - Prevents crashes due to common issues, such as non-numeric inputs or insufficient data.

Literature Cited

Davila F., Paz F., Moquillaza A. (2023). Usage and Application of Heatmap Visualizations on Usability User Testing: A Systematic Literature Review. In: Marcus, A., Rosenzweig, E., Soares, M.M. (eds) Design, User Experience, and Usability. HCII 2023. Lecture Notes in Computer Science, vol 14032. Springer, Cham. https://doi.org/10.1007/978-3-031-35702-2_1.

Frey, T. and Võhandu, L.K. (1966). Uus Meetod Klassifikatsiooniühikute Püstitatimiseks, Eesti NSV Akadeemia Toimetised, xV Kõide, Bioloogiline Seeria, Nr. 4, 565-576 (Estonian), http://www.datalaundering.com/download/Frey-Vyhandu-RG.pdf (English version)

Henderson, H.V. and Velleman, P.F. (1981). Building multiple regression models interactively for statistical analysis and data visualization. Biometrics, 37, 391_411.

Ishii Aki, Yata Kazuyoshi, Aoshima Makoto (2021). Geometric classifiers for high-dimensional noisy data, Journal of Multivariate Analysis 188 (2022) 104850, https://doi.org/10.1016/j.jmva.2021.104850.

Johnstone Iain M. and Titterington D. Michael (2009). Statistical challenges of high-dimensional data, Published: https://doi.org/10.1098/rsta.2009.0159.

Jöreskog Karl G. (1978). Structural Analysis of Covariance and Correlation Matrices, Psychometrica, vol. 43, No.4, pp.443-477.

Midway Stephen R. (2020). Principles of Effective Data Visualization, Patterns (N Y). Nov 11;1(9):100141, https://pmc.ncbi.nlm.nih.gov/articles/PMC7733875/ .

Midway Stephen R. (2020). Principles of Effective Data Visualization, ScienceDirect, Patterns, Volume 1, Issue 9, 11 December 2020, 100141. Principles of Effective Data Visualization - ScienceDirect.

Mirkin B.G. (2011). Core Concepts in Data Analysis: Summarization, Correlation, Visualization, Department of Computer Science and Information Systems, Birkbeck, University of London, Malet Street, London WC1E 7Hx UK, Textbook, 426pp, https://link.springer.com/book/10.1007/978-0-85729-287-2 .

Mullat J.E. a) (2022). Case Study of Fuel Consumption by Vehicles Utilizing the Postulates of Bounded Rationality, Private Publishing Platform, Monotone Phenomerna of Issues Behind Baragaining Games and Data Analysis, pp.172-191, http://www.datalaundering.com/download/RG-Case-Vehicle-Study.pdf; b) (1971). On a Maximum Principle for Certain Functions of Sets, in "Notes on Data Processing and Functional Analysis", Proceedings of the Tallinn Polytechnic Institute Series A, No. 313, Tallinn Polytechnic Institute, pp. 37–44, (in Russian), http://www.datalaundering.com/download/RG-193-201-Modular.pdf, (English).

Pleil Joachim D., Stiegel Matthew A., Madden Michael C., Sobus Jon R. (2011). Heat map visualization of complex environmental and biomarker measurements. Chemosphere, Volume 84, Issue 5, Pages 716-723. https://doi.org/10.1016/j.chemosphere.2011.03.017.

Rozonoer Lev, Mirkin Boris, Muchnik Ilya (2017). Key Ideas from Inception to Current State: International Conference Commemorating the 40th Anniversary of Emmanuil Bravermans Decease, Boston, MA, USA, April 28-30, 2017, Invited Talks

Schwabish Jonathan (2021). The Practice of Visual Data Communication: What Works, SageJournals, Volume 22, Issue 3, https://doi.org/10.1177/15291006211057899.

Sun Guodao, Zhu Zihao, Zhang Gefei, xu Chaoqing, Wang Yunchao, Zhu Sujia (2023)."Application of Mathematical Optimization in Data Visualization and Visual Analytics: A Survey," in IEEE Transactions on Big Data, vol. 9, no. 4, pp. 1018-1037, https://ieeexplore.ieee.org/document/10081451.

Wendy K., Cho Tam, Liu Yan Y. (2018). Sampling from complicated and unknown distributions: Monte Carlo and Markov Chain Monte Carlo methods for redistricting, Physica A: Statistical Mechanics and its Applications, Volume 506, 15 September 2018, Pages 170-178. https://doi.org/10.1016/j.physa.2018.03.096 .