This project is funded by the European Union.

TESEV BRIEFS

# Qualitative Synthesis Maps for Data-Driven Urban Governance: A Methodological Evaluation

CİNSİYET EŞİTLİĞİ İZLEME DERNEĞİ

TESEV

## INTRODUCTION

Since 2016, TESEV has been producing maps that support urban policy development and participation in the areas of Urban Governance and Sustainability, Gender Equality and Child Policy.

**Why does TESEV prepare these maps?**

Data supply in Turkey, especially at the neighborhood scale, comes with significant limitations. This is a fact acknowledged by all experts working in this area. Nevertheless, compiling the available data into legible maps helps to reveal urban patterns that would otherwise be unnoticed. These maps are invaluable as they depict similarities and differences that are hard to perceive and data on services provided by public institutions. At times, crowdsourced data may provide additional clues for policy priorities in the city. Over the last six years, these maps have attracted great interest from public institutions, civil society organizations, and academia, and they have been used to develop relevant urban policies targeting different social groups.

**What makes TESEV's maps distinctive?**

The maps are prepared in collaboration with the Istanbul Studies Center at of Kadir Has University. Their accuracy and legibility stems from a new and sophisticated methodology which was presented in detail in oral presentations. Given the increasing interest in the maps, time has come for a detailed methodological evaluation that highlights the distinctive aspects of these maps. Prof. Dr. Murat Güvenç, who prepared the maps series for TESEV since 2016, wrote the following methodological overview which we are pleased to share with you.

**Prof. Dr. Murat Güvenç**
Kadir Has University
İstanbul Studies Center

Murat Güvenç is born in Ankara in1953. Graduated from St. Joseph French High School Istanbul. Studied urban and Regional Planning in Middle East Technical University (METU). Received hi Ph. D from (METU) 1991. Worked as a Faculty at METU for about 28 years, joined the Graduate Program in Architectural Design of Istanbul Bilgi University in 2005. He worked on urban social geography, metropolitan industrial and electoral geography and institutional history . He co-curated the Istanbul 1910-2010 exposition. He is the co-author of Emlak Bankası 1926-1998: An Institutional history of Real Estate Bank and Electoral Atlas of Turkey 1950-2009. Professor Güvenç acted as a member of the executive board of TESEV. He joined Kadir Has University in 2014 and nominated as the Director of Istanbul Studies Center. He was a member of of the award winning Şerif Süveydan design team in Taksim Square Competition (2019). Since 2016, he is directing the data analysis and mapping teams for TESEV's urban data platforms.

Cities are complex social formations with a high level of internal differentiation, shaped by technological possibilities, socially shared values, dominant worldview, and practices of social production and reproduction. These characteristics make cities difficult to conceptualize and talk about. Their representation has been a research priority since early modernity when urbanization processes gathered momentum. There are different theoretical perspectives that define the city in different ways. Thus, cities are defined as (1) isolated **market centers** that shape the agricultural landscape in their immediate surroundings, (2) **centers** that provide services to settlements in their hinterland the size of which is defined by to their own position in settlement hierarchy, (3) as **areas of spatial interaction** shaped by the production of goods and services for export and the local population. (4) In social/cultural studies, cities are viewed as **ecological formations** shaped by their interacting communities. (5) Theories focusing on urban social movements consider cities as spatial aggregations shaped by **modalities of collective consumption**. Finally, in political economy cities are taken up as (6) agglomerations shaped by cyclical crises of **capital accumulation**.[1]

**The concise, reliable, and valid representation of qualitative data and the communication as well as visualization and mapping of findings require robust and replicable methods.**

While only a limited share of the world's population was living in cities at the beginning of the 20th century, more than half of the global population currently lives in urban areas. Today's cities are unprecedented spatial formations that have grown beyond administrative boundaries. As such, they are very difficult to perceive, monitor, and conceptualize. Meanwhile, good governance is vital to prevent these formations from becoming a threat to social peace, **social cohesion**, environmental and cultural values, architectural heritage, and ecological sustainability. The success of good governance depends upon constant monitoring and representation of urban processes. This (monitoring) requires up-to-date, high-resolution, accessible, and detailed datasets, as well as innovative approaches and participatory decision support systems **emphasizing the quality of urban life and the respect of ethical values**. However, it is almost impossible to generate such detailed, comprehensive **high-resolution** datasets through individual initiatives or in the context of small academic research projects. The provision of such detailed data at high resolution is one of the major issues of **urban governance today**.[2] Datasets produced in earlier periods are not sufficient to monitor urban formations if they are not updated. Nowadays, the continuity, accessibility, sustainability, updating, scope, and comparability of urban datasets are often more important than their content.

The concise, reliable, and valid representation of qualitative data and the communication as well as visualization and mapping of findings require robust and replicable methods that are resistant to the problems caused by disaggregation and high levels of resolution (that minimize spatial and classification errors) while also being compatible with our cognitive constraints.[3] In practice, this problem is mitigated with reference to general categories such as "residential area", "central business district (CBD)", "industrial zone", "small industrial area", "area of blue-white collar employment", etc. While such categories are useful for an overall assessment, they do not provide clues as to the internal differentiation. General categories may complicate policy design for target groups.[4] It is true that disaggregation can alleviate this problem. Yet, **disaggregation alone does not yield relevant clues about the context.** The representation of urban formations and social profiles, the identification of problems and opportunities, and a reliable and valid policy design all require categories at the everyday level that avoid the "pitfalls" of overgeneralization and overly high resolution. Gray proposes to solve this problem by disaggregating complex formations into sub-components.[5] The conceptual solution elaborated by Gray can be operationalized by a new approach developed by L. Lebart.

The methodological approach employed in TESEV's projects on urban, children and women issues since 2016 is an operationalized version of a pattern recognition model developed by French cartographer J. Bertin in the mid-1960's. The distinctive aspects of this approach are presented in **four paragraphs** of this research note. **The first section** discusses the problems related to the visualization of observation units defined by administrative boundaries. Large tables containing continuous variables can be easily reduced and mapped through factor analysis. The same does not hold true for categorical data sets however. The problem that one runs into at this point can be solved by the "**graphical information processing approach**" Bertin has developed in mid 60's.[6] Though no longer applied in the same way today, the features of this graphical approach to pattern recognition are described in detail in the **second section** of this presentation. In the mid-80s, it was shown that the Bertin's model could be operationalized through correspondence and cluster analyses.[7] Yet, by themselves, neither correspondence analysis nor cluster analysis are suitable for processing disaggregated and high-resolution datasets. We are going to see that the complementary use of these models stipulated by L. Lebart facilitates the processing of highly detailed qualitative data sets at high levels of resolution. This model is introduced in the **third section** of this piece.[8] This methodological evaluation concludes with a discussion regarding the evaluation of the findings and maps drawn with this procedure.

**Visualization issues related to observation units**

In Turkey, as in many other countries, comprehensive and detailed social data are obtained through censuses conducted on administrative units. In other words, administrative units are fixed references for interpreting the collected data. But while

**In such maps the periphery will have an overwhelming impact on city centers and boundaries determined by administrative concerns will unavoidably affect the perception and interpretation. This is often passed over in silence, but it constitutes an important problem.**

units at the same level of administrative hierarchy are often defined in the same way, their surface area is inversely proportional to the population density and the diversity of economic activities. The area of administrative units generally increases the further one moves away from central business districts. While local formations in and around central business districts tend to be **dense and differentiated**, those in the periphery are generally **less dense and specialized**. When mapping is based upon administrative boundaries, the central areas where most of the economic activity takes place are overshadowed by the colors assigned to the periphery. In such maps the periphery will have an overwhelming impact on city centers and boundaries determined by administrative concerns will unavoidably affect perception and interpretation. This is often passed over in silence, but it constitutes an important problem. Due to practical problems such as legal regulations, budget restraints, logistics, etc. it is difficult –if not impossible– to collect data in a scope and at a level of resolution that could compete with the census. One way to cope with these difficulties is using cartograms, which are part map, part graphic representations. Despite the distortions they cause, cartograms can be seen as more realistic than maps. For instance, a cartogram allowing settlements with more unemployed people than the general average to occupy more space than those with fewer unemployed people may be much closer to our mental maps. Meanwhile, the distortions caused by the cartogram representation can be reduced by choosing hexagon-based units. This mode of representation does not contain any empty spaces (i.e., any discontinuity). The Atlas of the United Kingdom created by Daniel Dorling and Bethan Thomas based on 2001 census data is one of the most striking examples. It should be noted, however, that in order to forestall interpretation errors, each map in this atlas is displayed in both cartogram and actual map format.[9]

Census units with fixed boundaries are likely to be the ultimate solution. However, even in this case, there is the problem of linking previously collected data to newly created units. This constitutes a huge work package that individual researchers or civil society organizations with their limited resources cannot fulfill. In this TESEV project (*Empowering Civil Society and Municipalities for Data-Driven Participatory Gender Equality Policies*) the problem regarding visualization arising from the difference in surface area was mitigated by **marking only areas of settlement** instead of relying on administrative boundaries. This approach (also known as asymmetric mapping) makes it possible to map findings related to large administrative units in the periphery

**without de-emphasizing central sectors**. A major drawback of this approach is the time-consuming process of manually digitizing settlement areas. That said, one can expect machine learning to help solve this problem in the very near future. Problems related to the legibility of maps can be mitigated through zoom-in. Maps produced in this way did not receive any criticism regarding legibility.

### Mapping Local Profiles Obtained from the Census

Census data stored on crosstabs show the distribution of selected characteristics across observation units. Cross-tabulation is a data storage technology enabling easy access to the connection between units of observation and their attributes. Yet, in themselves these cell-based readings are not practical at all. Knowing that there are x-number of, say, wage earners in settlement A is not very informative – except of course for those who for some reason are interested in settlement A or wage earners. For this reason, numbers are nowadays seldom communicated without reference to a fixed reference.[10] Compared to single cells, the row or column profiles offer much richer information. That said, only comparable crosstab profiles are meaningful. Normalization based on row and column sums can alleviate the problem but does not fully solve it. Thus, normalization based on row sums facilitates local profile comparisons while filtering off differences in scale. For example, if settlement A has a population of 200 and settlement B has a population of 2000 and the ratio of children in pre-school age is 5% in both settlements, these two settlements are not equivalent in terms of pre-school investment priorities. In such a scenario, decision-makers would be expected to prioritize the settlement with 100 children and come up with a special solution for settlement B. In a similar manner, this also applies to column totals. Row or column normalization can be misleading if there is no reference for comparison. This weakness is addressed by the signed chi indices, which are signed indicators depicting differences from an expected frequency value for each cell of the table. Developed by census geographers, this indicator is an interesting tool facilitating both the comparison and mapping of census data.[11]

#### 1. Bertin's Model for Mapping of Local Profiles

We start to see that evaluation of local profiles required sound references and comparisons. Notice however that the number of comparisons increases with the square of the number of observations. For example, in a city with 1,000 observation units (neighborhoods), this requires about half a million (n2/2-n) operations.[12] In large tables with hundreds of thousands of observations, the number of comparisons becomes too large even for computers to handle.

Representing local attribute profiles **through differences from the overall average profile prepares the ground for significantly reducing the number of comparisons**. However, this method still requires that one makes as many comparisons as the number of observations. Therefore, it has a limited scope of application. In the second half of the 1960s, J. Bertin proposes a partial remedy to the comparison problem with his **reduced and re-ordered**

**matrices**. A more advanced version of this model is applied in TESEV's *Empowering Civil Society and Municipalities for Data-Driven Participatory Gender Equality Policies* project. However the following example, showing the occupational geography of Paris in 1950 based on neighborhood-level data, provides some guidance for the interpretation of the maps created in the project. In this example, occupational differentiation is described through nine categories.

1. Domestic help,
2. Professionals,
3. Employees
4. Craftsmen,
5. Managerial personnel,
6. Unskilled laborers
7. Workers,
8. Manufacturers/big businessmen
9. Miscellaneous

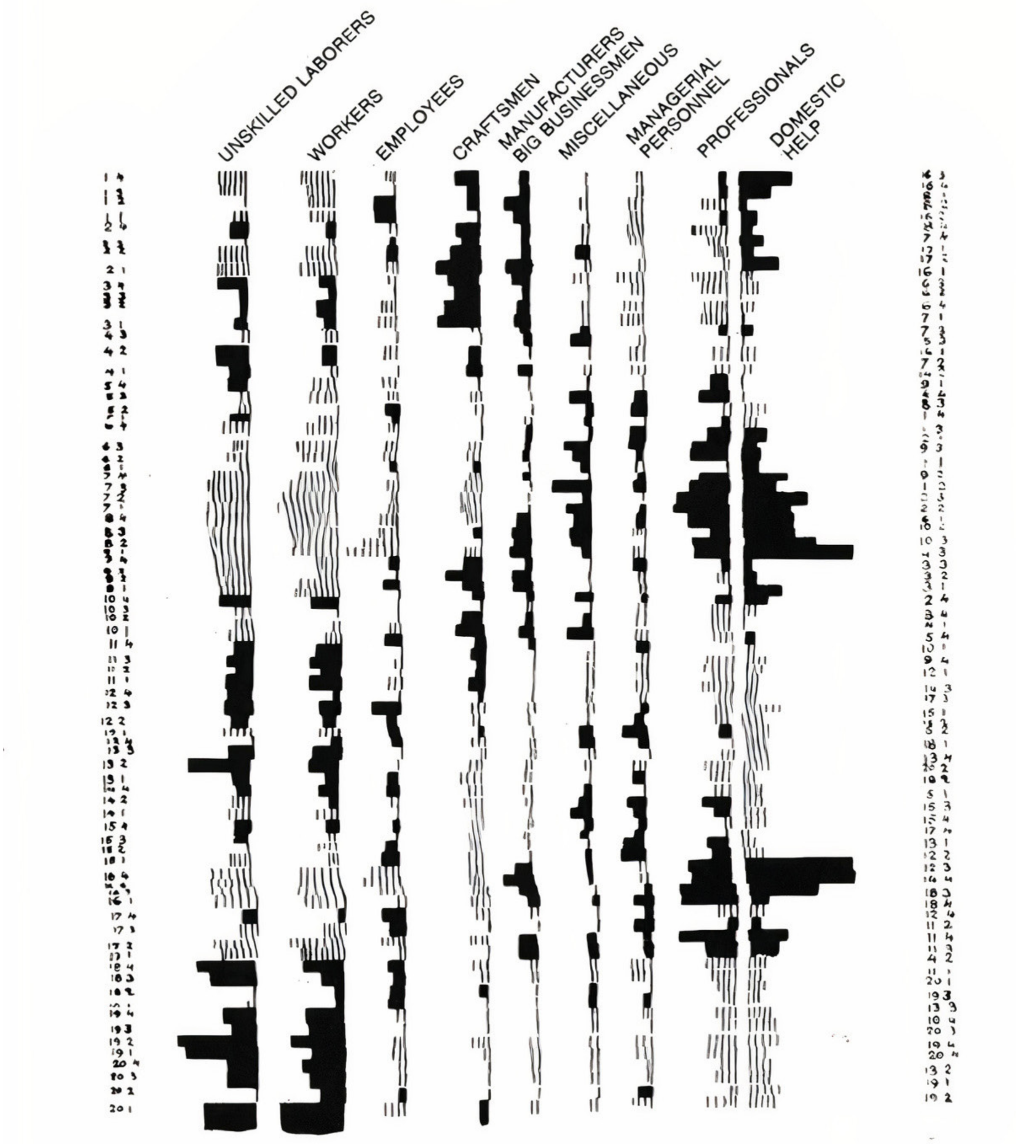



**Figure 1: Re-ordered Image File**
Source: Jacques Bertin. *Semiology of Graphics; Diagrams, Networks, Maps*, Redlands CA: ESRI.2011 p.230.

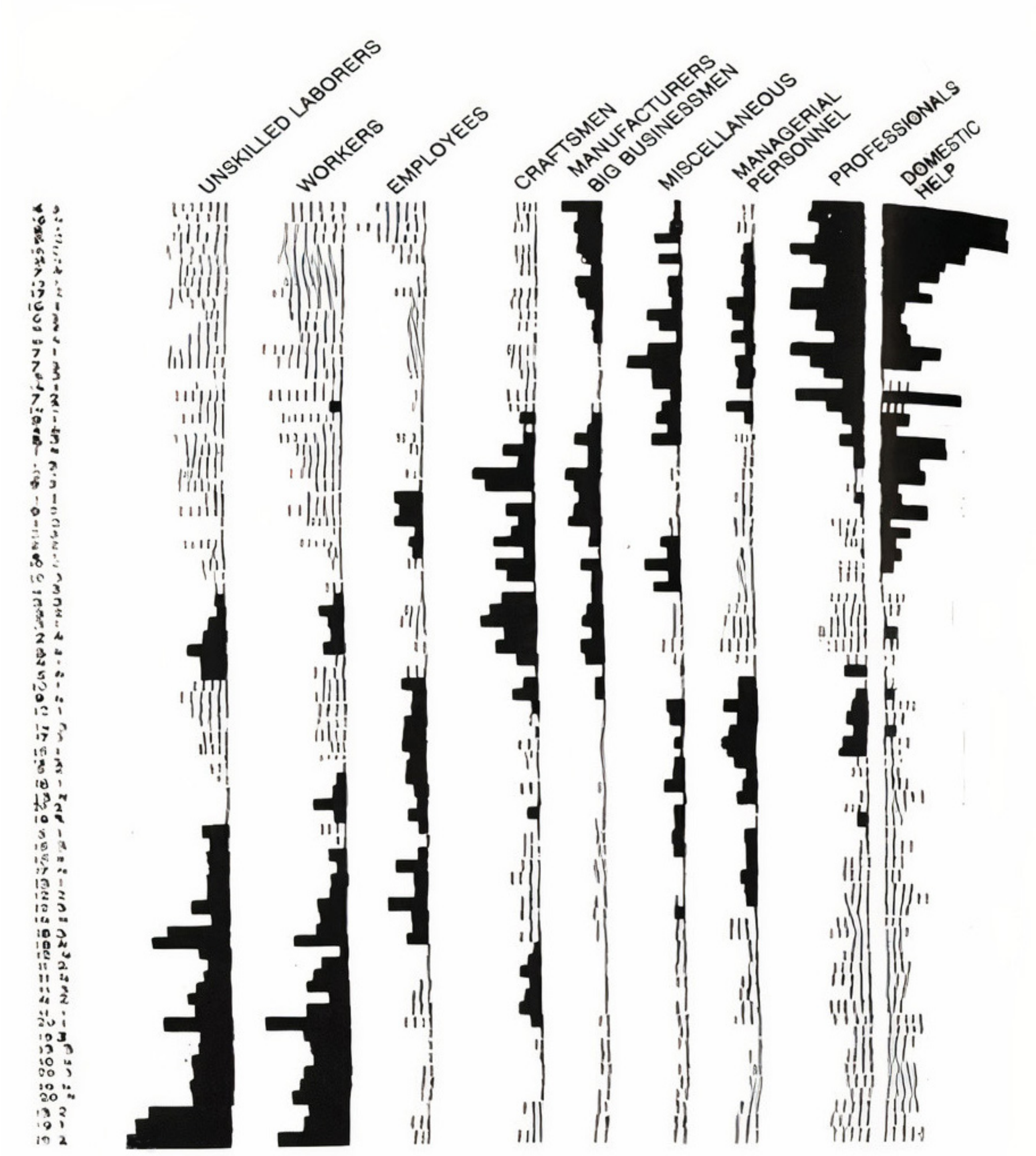



**Figure 2: Reduced and Reordered Image File**
Source: Jacques Bertin. *Semiology of Graphics; Diagrams, Networks, Maps*, Redlands CA: ESRI.2011 p.230.

Bertin's method involves **reordering and clustering**. In the first stage the differences between the normalized local profiles and the normalized overall profile are calculated.[13] The largest difference (deviation) in each column is determined. The columns of the table are re-ordered with respect to largest differences. The re-ordered table detects the characteristics that are closest to the average profile and those that are distant from it (i.e., distinctive).[14] The former are usually uninteresting while the latter are important and prepare the ground for further studies. Cell values above (and below) the overall average are displayed in different sizes and colors **to create an image file (Figure 1)**. Rows with a **consistent pattern in the image file** are combined[15] **to produce a diagonal matrix (Figure 2)**. The diagonal matrix reveals areal connections/associations between observations and features that are not visible to the naked eye. Notice the meaningful yet unexpected association between "domestic help". Finally, the patterns (identified by the naked eye) are cut at their breaking points to create legend categories that indicate (mutually exclusive) occupations that are concentrated together. This operation, which reduces the size of the table by clustering rows with compatible profiles, constitutes the reduction stage. Categories are named with reference to the over-represented characteristics (**Figure 3**). The categories obtained as a result of this operation are listed below.

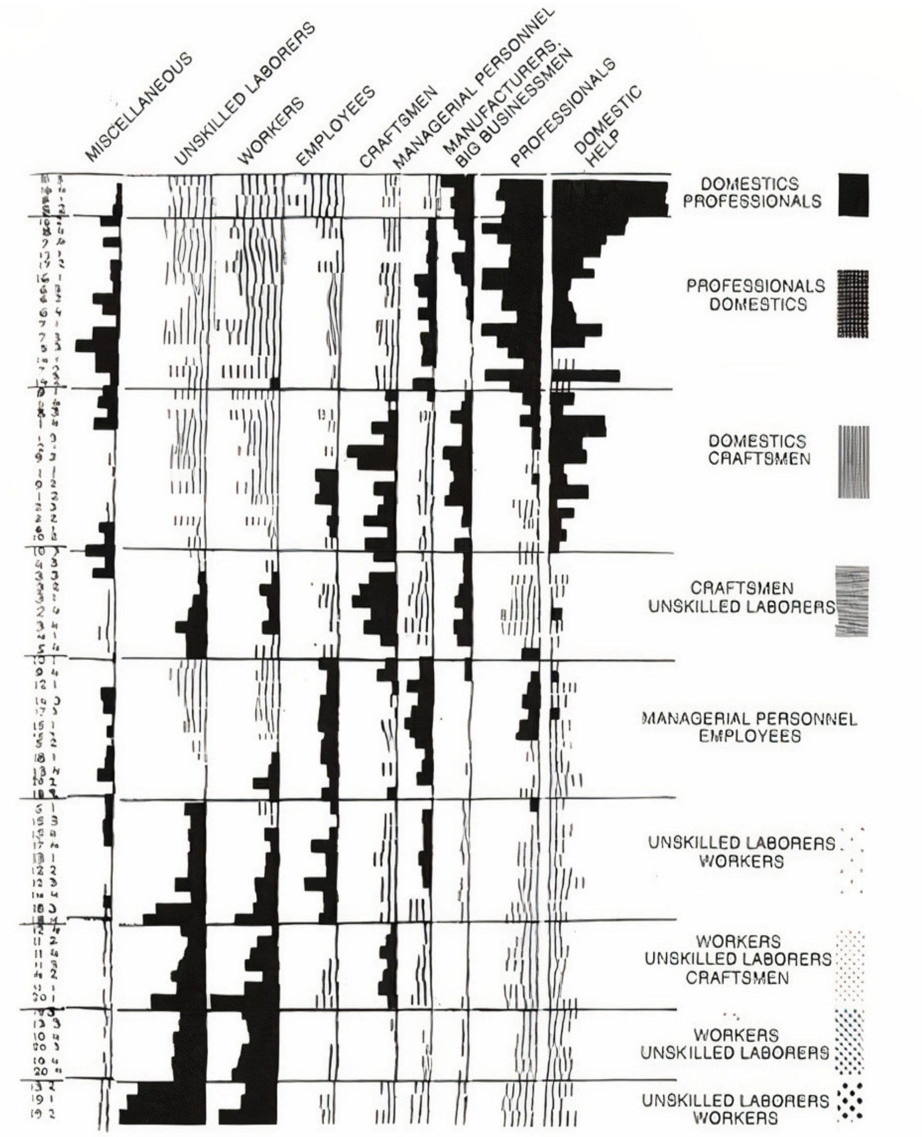
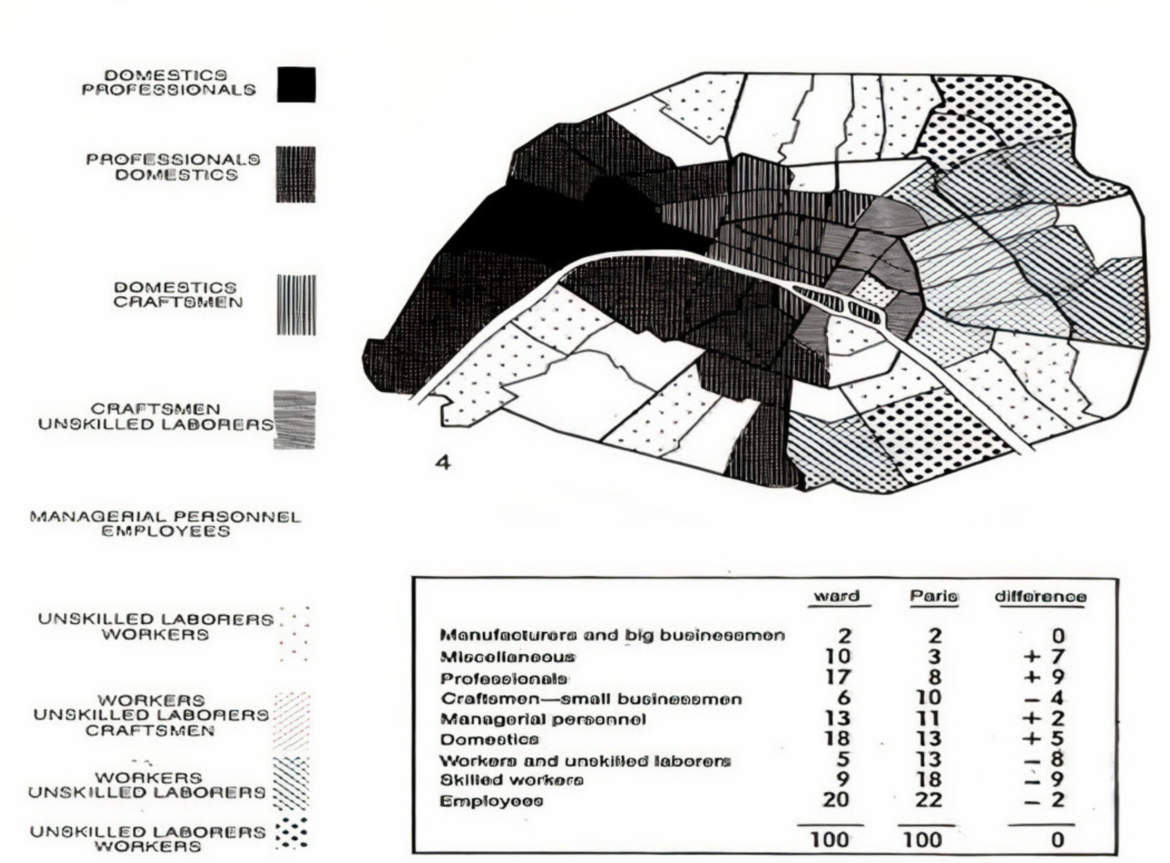



**Figure 3: Reduced and Re-ordered Image File with Legend Categories**
Kaynak: Jacques Bertin. *Semiology of Graphics; Diagrams, Networks, Maps*, Redlands CA: ESRI.2011 p.231.




| | ward | Paris | difference |
|---|---|---|---|
| Manufacturers and big businessmen | 2 | 2 | 0 |
| Miscellaneous | 10 | 3 | + 7 |
| Professionals | 17 | 8 | + 9 |
| Craftsmen—small businessmen | 6 | 10 | − 4 |
| Managerial personnel | 13 | 11 | + 2 |
| Domestics | 18 | 13 | + 5 |
| Workers and unskilled laborers | 5 | 13 | − 8 |
| Skilled workers | 9 | 18 | − 9 |
| Employees | 20 | 22 | − 2 |
| | 100 | 100 | 0 |

**Figure 4: Synthesis map of occupational differentiation in Paris neighborhoods**
Kaynak: Jacques Bertin. *Semiology of Graphics; Diagrams, Networks, Maps*, Redlands CA: ESRI.2011 p.231.

1. **Professionals**, domestic help

2. **Domestic help**, craftsmen,

3. **Craftsmen**, unskilled laborers

4. **Managerial personnel**, employees

5. **Unskilled laborers**, workers

6. **Workers**, unskilled laborers, craftsmen,

7. **Workers**, unskilled laborers

8. **Unskilled laborers**, workers

9. **Miscellaneous**

The process visualized in Figures 1-3 ultimately produces a original and legible synthesis map (**Figure 4**).

The map in Figure 4 shows a striking differentiation, enabling even those who do not know anything about Paris in the mid-1950s to discern that

- Professionals, managerial personnel and big businessmen,
- Craftsmen and employees in administrative positions,
- And blue-collar workers

are concentrated in different sectors of the city. The legends on the map are consistent with the image file: blue collar workers are rarely encountered in the segment with a high concentration of white-collar workers and vice versa; craftsmen and those working in small

**In the 1980s, it was shown that Bertin's model could be applied with correspondence analysis and clustering algorithms. This contribution significantly expanded the scope of application of the model, which would subsequently evolve into a general pattern recognition model.**

administrative positions choose a location independent of these two groups. This salient differentiation is reflected in the urban space. A more detailed analysis detects more subtle distinctions between the occupational profiles of the *left* and *right* banks of the Seine. Finally, it can be observed that, in the mid-1950s, the gendered division of labor is reflected in urban space through the predominantly female domestic workers, who live in the same neighborhoods as the upper-level managers and artisans, rather than blue-collar workers. How do domestic service workers managed to reside in these expensive neighborhoods? Do they live in their "workplaces"? Is the structure observed in these neighborhoods also valid in sectors where craftsmen depict high concentrations? If not, how can this differentiation be explained? Such research questions, otherwise difficult to raise, can be easily developed with reference to this map.

In the light of this summary evaluation, one may emphasize the following features of the Bertin model:

- The application of the model is limited by the **visual pattern recognition capacity** of the researcher. It is not suitable for mapping detailed qualitative datasets at high level of resolution. In practice, it enabled the evaluation of tables containing at most 300 units and up to 40 features.

- The model is sensitive to distinctive cases rather than to those that are close to the general trend. This is an advantage in some studies and a disadvantage in others. Categories that are not **particularly distinctive in an unusual manner** cannot be mapped as precisely as those that are. This may be an important disadvantage of the model.

- The re-ordered/reduced table approach is sensitive to first order contrasts. This implies that second and third order cannot be interpreted through this model. Notice that the category "*miscellaneous*" is not included in the map legend.

- This drawback can be overcome by using correspondence analysis, allowing all dimensions to be evaluated together.

Despite its above-cited disadvantages and limited scope of application, Bertin's is a pioneering model for pattern recognition. In fact, all maps prepared within the scope of TESEV's *Empowering Civil Society and Municipalities for Data-Driven Participatory Gender Equality Policies* project were created using a developed variant of the method developed by Bertin. They can be interpreted in the same way. The labels used in the legends of these maps, i.e., "High School, College, and Graduate School", "Child, Youth, Adult", "Married, Divorced", etc., depict statistically significant areal concentrations.

In the 1980s, it was shown that Bertin's model could be applied with correspondence analysis and clustering algorithms.[16] This contribution significantly expanded the scope of application of the model, which would subsequently evolve into a general pattern recognition model. However, it soon became clear that in themselves, neither correspondence analysis nor clustering algorithms are suitable **for data processing at high levels of resolution and disaggregation**. Therefore, these new methods did not address these weaknesses of the graphic data processing approach, and this for conventional three important reasons.

**1.** Clustering algorithms [with the exception of those developed in machine learning] are not capable of dimensionality reduction. It is therefore difficult to interpret the clustering results of detailed tables. However, minimization of classification errors requires processing qualitative data at a high level of detail. To meet this requirement, Gray proposes to decompose "complex" general concepts into components.[17] Disaggregated variables do not cause problems in classification, but they do lead to interpretation and communication problems. This limits the scope of application of clustering algorithms. According to Gray, decision-makers and bureaucrats prefer to use colloquial rather than technical terms.[18] Translating detailed categories into everyday language enhances ambiguity.

**2.** Secondly, hierarchical clustering algorithms compare all observations against each other and generate large (n x n) dissimilarity matrices. It follows that they are not particularly appropriate for data processing at high resolution. This problem can be overcome through genetic algorithms (such as k-means). However, genetic algorithms have methodological weaknesses. They require that the number of clusters to generate is determined in advance. The results depend on the starting points. It follows that replicability and the comparability of results constitute important methodological problems.

**3.** Finally, correspondence analysis is a representational tool with limited visual communication capabilities, vulnerable to the misleading effects of outliers. It is not suitable for analyzing high resolution and detailed datasets.[19]

## 2. Lebart's Model

The misleading effects of outliers affects all the dimensions of correspondence analysis, thus limits its scope of implementation. Its hypersensitivity to outliers makes it difficult to represent features (and observations) that are not particularly distinctive. The widening of the scope of application of correspondence analysis and the reliability and validity of findings depend on controlling the misleading effects of outliers. Referring to the connection between the nodes of dendrograms and the dimensions extracted by the correspondence analysis, L. Lebart emphasizes that **these models can be used complementarily**. Clustering algorithms are less affected by unusual observations in the early stages of the (clustering) process.[20] This advantage of the clustering model alleviates the over sensitivity of correspondence analysis to the presence of outliers in the dataset. In its turn, the correspondence analysis transforms

→ **In simple correspondence analysis, clusters are represented by aggregated observation profiles or weighted central objects. In Multiple Correspondence Analysis, clusters are necessarily represented through weighted central objects.**

the variables into a limited number of factors and reduces the size of the dataset. In Lebart's approach;

• The weakness of cluster analysis in handling high dimensional is resolved with the help of correspondence analysis[21] and

• The problems related to the high sensitivity of correspondence analysis to outliers is alleviated through two step clustering.

Application of the model takes place in two steps. The reduced table that is to form the basis for the final clustering is produced in **two stages. In the first stage**, correspondence analysis (or multiple correspondence analysis) is used to determine the number of dimensions over which will represent the attributes of the data set as well as the coordinates of the observations on these dimensions.[22] In correspondence analysis (or multiple correspondence analysis) tables with *m* features are represented with *m-1* dimensions. However, dimensions accounting for less than *1/m*100* of the total eigenvalue are marginal and are not worth considering. Observations are represented through their coordinate sets and attributes are represented by dimensions (determined by correspondence analysis).[23] Congruent profiles are defined by similar coordinate sets. Profiles that differ from the general trend are located at the antipodes of the dimensions. **In the second stage**, observations with respect to their scales and dimensions are weighted with respect to their eigenvalue shares and clustered.[24]

The method applied in the second step is the same as in the first step. Correspondence analysis (or multiple correspondence analysis) is used to determine the distinctive dimensions and coordinates of preliminary clusters. The composition and labels of the legend categories are determined through cluster analysis.

The stages of the above described data reduction process are summarized in Tables 1 and 2. Lebart's model can be applied to both cross tabulations and categorical datasets with only minor modifications. The most important difference between the two models concerns the representation of clusters in the second step. In simple correspondence analysis, clusters are represented by **aggregated** observation profiles or weighted central objects. In multiple correspondence analysis, clusters are necessarily represented through **weighted central objects**. Lebart's two-step representation model allows for the clustering and mapping of detailed datasets at high resolution with minimal and measurable loss of information. Different heuristics such as the "elbow method", minimization of within-cluster variances,"silhouette indices" assist students in determining the adequate number of clusters and the quality of classifications. The results can be interpreted with reference to their position on the final correspondence maps and with respect to statistically over-represented attributes. The model for multiple correspondence analysis is a version of correspondence analysis adapted to categorical data. Lebart's model has a wide range of applications. It can be used to process all kinds of statistically significant cross-tabulations or qualitative datasets.

| **Table 1: Mapping Detailed Crosstabs at High Resolution** |
|---|
| **First Step** |
| **Simple Correspondence Analysis of the Dataset** |
| • Determination of distinctive dimensions and<br>• The coordinates of observations on these dimensions,<br>**Pre-Clustering:** |
| o Weighting of "*observations*" according to their "*scales*" and<br>o Dimensions according to their eigenvalue shares,<br>Representation of cluster through aggregation or central objects.<br>o Generation of the interim reduced table for final analysis.<br>**Final Clustering** |
| **Simple Correspondence Analysis of the New Data Created in the First Stage** |
| • Determination of distinctive dimensions and their eigenvalue shares,<br>• Determination of the coordinates and weights of new clusters,<br>**Final Clustering with Ward's Method** |
| • Clustering of weighted central objects or clusters of objects objects or new cases by aggregating observations assigned to each cluster<br>• Final; clustering labeling, creation of a list of legends for a geographic information system |

| Table 2: Mapping of Categorical Datasets |
|---|
| **First-Step** |
| **Multiple Correspondence Analysis of the Dataset** |
| • Determination of distinctive factors (dimensions) and their eigenvalue shares<br>o Generation of the interim reduced table for final analysis |
| **Pre-Clustering:** |
| • Weighting observations with respect to their scales and dimensions according to their eigenvalue shares<br>• Generation of the interim reduced table for final analysis . |
| **Second-Step** |
| **Multiple Correspondence Analysis of the Central Object Dataset Created in the First Stage** |
| • Redefining the distinctive dimensions accounting for more than 1/n of total inertia<br>• Determination of the weights of central objects and factors<br>• Final clustering (via Ward's criterion)<br>o Final clustering; labeling, creation of a list of legends for a geographic information system |

## Evaluation and Conclusion

Lebart's model allows for the processing and mapping of disaggregated tables and/or qualitative datasets defined at high resolution. It detects the distinctive dimensions of social-spatial formations that are difficult to discern with the naked eye, without the need for expensive equipment or technical skills. All thematic and synthesis maps of this project and TESEV's previous projects such as Urban95 and Morharitam, were created using the model described above. Thematic maps were created based on the process described in Table 1, while synthesis maps were created based on the process described in Table 2.

The maps enable decision-makers, researchers, and stakeholders to perceive, comprehend, and monitor differentiation and formations at the scale of the metropolitan region and to develop testable hypotheses and come up with new research questions. The flexible nature of the model allows for the development of more comprehensive or detailed research. The quality and precision of the findings can be tested through statistical indicators, and their validity with participatory methods.

That said, we must not forget that no representation model is perfect.[25] Even though it works better than its alternatives, this also applies to Lebart's model. Despite its methodological advantages Lebart's model is only one of many representation models. The definition of individuals and the method of measuring differences (metric) adopted by this model to improve the representational quality of outliers are not without alternatives and may not be equally relevant in each research problem. Clustering algorithms that consider exclusively shared characteristics may be more successful in practice than those that depend upon external references (metrics).[26] Defining individual profiles in terms of differences from the overall average profile, Lebart's model generates **a top-down typology**, with reference to general structural features. However, clustering can also be realized with respect to locally shared features.

Recent algorithms based on the concept of affinity, generate clusters with respect to local references and produce **a bottom-up typology**. Which of these approaches will be better suited depends largely on the research question and priorities. Applied to the same dataset, different clustering models can yield very different results (**see Figure 5**). While top-down models may be employed for general assessments, bottom-up typologies based on shared characteristics may be useful in studies shedding light upon local specificities. However, this also depends upon the shape of the input data. Notice that the patterns in the **third**, **fifth** and to some extent **fourth** rows in Figure 5 can be identified by all models, while the results generated for input data in the **first**, **second** and **sixth** rows may vary depending on the model.

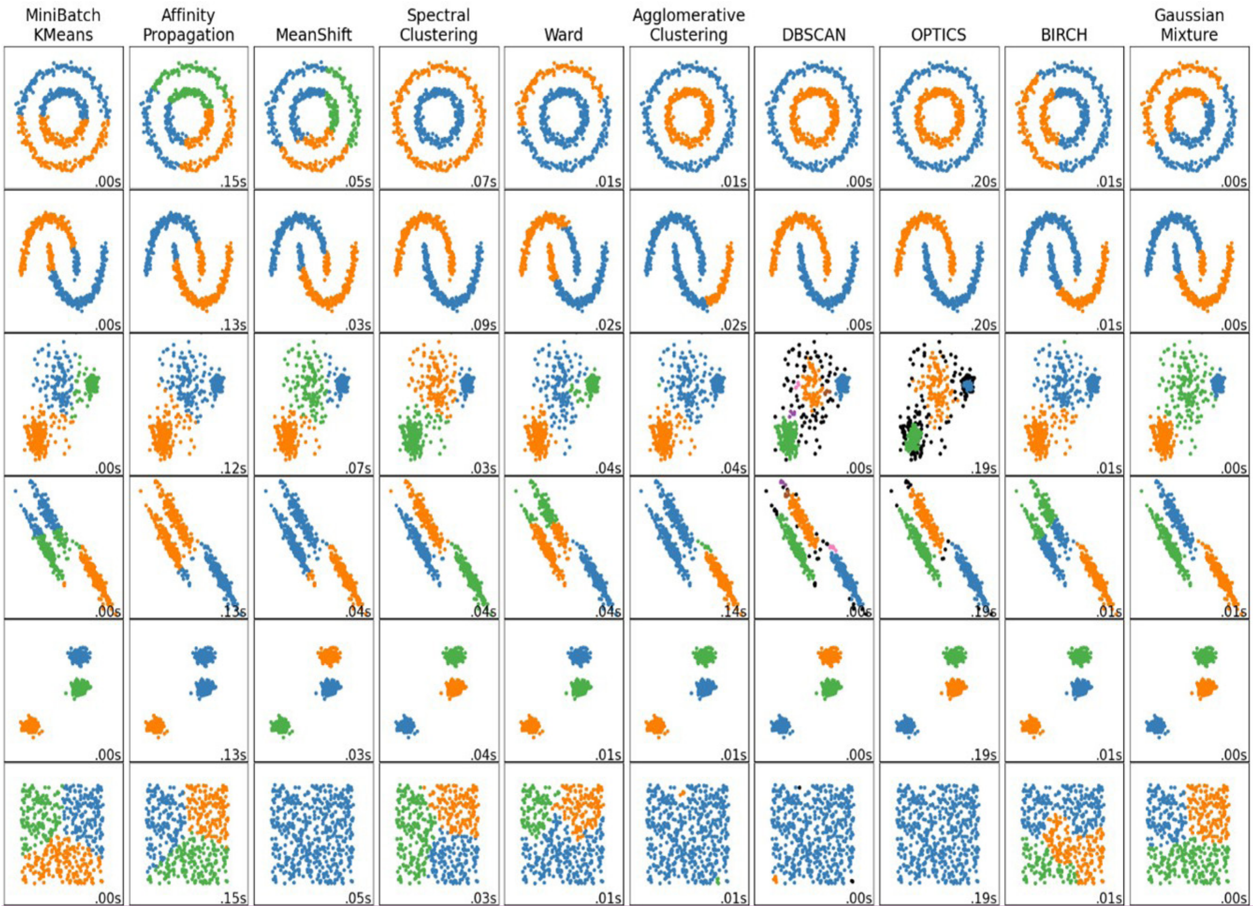



**Figure 5: Results obtained from different data sets using different clustering models.**
Source: scikit-learn, Comparing different clustering algorithms on toy datasets. Accessed: 25.10.2022, https://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_comparison.html

**The flexible structure of the model and the fact that the findings are published on the web allows larger public participation and eventually for periodic revisions and corrections.**

We may emphasize that the results of clustering exercises reflect both the characteristics of social-spatial differentiation reflected in census data as well as the characteristics of the taxonomic procedure.

• This exploratory exercise in the evaluation, visualization, and mapping of data aims to exhibit significant differences, general contrasts, and agglomerations, rather than comparing places to each other. Maps can provide decision-makers with new perspectives and stakeholders with a new and relevant understanding of urban issues, and relevant arguments of negotiation.

• This model makes it possible to monitor socio-spatial formations, which are often situated outside the city boundaries in sectors too large to be perceived by individual observers.

• One should make sure to take the maps obtained with Lebart's model as only one possible representation, as references that can be used until a better one is developed, i.e., as the accuracy of findings is tested through more detailed evaluations. Applying different models at a higher level of resolution and/or detail may yield different results.

• **It would be misleading and inappropriate from the point of view of research ethics to develop policy proposals** without scrutinizing the exploratory findings through participatory evaluation. Therefore, these maps should be considered as a starting point

• The flexible structure of the model and the fact that the findings are published on the web allows larger public participation and eventually for periodic revisions and corrections. In light of this assessment, it can be said that these new thematic and composite maps generate sound references for TESEV's research projects on urban governance, women and children.

# NOTES

1. On the characteristics of different theoretical approaches to the city, see: İlhan Tekeli: *Yerleşmeler için Temsil Sorunları ve Strateji Önerileri*, İdeal Kent, Ankara, 2016.

2. Violet, M. Gray, Classification as an Impediment to the Reliable and Valid Use of Spatial Information: A Disaggregate Approach. In Proceedings of the Spatial Information Theory A Theoretical Basis for GIS, Semmering, Austria, 21–23 September 1995; Hirtle, S.C., Frank, A.U., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; ss.. 137–149

3. Human cognitive capacity allows visual perception of up to seven categories and maps displaying seven to eleven colors. When these critical values are exceeded, verbal and visual communication of the content created becomes difficult. The maps lose their visual communication capabilities and turn into maps that need to be read. 1/25000 to 1/250000 scale military operation maps or maps for fire insurance purposes are examples of maps that are sensitive but lack visual communication capabilities and therefore need to be read. On the limits of human cognitive capacity see: Jacques Bertin J. Semiology of Graphics; Diagrams, Networks, Maps, ESRI, Redlands CA, 2011.

4. Violet, M. Gray, Classification as an Impediment to the Reliable and Valid Use of Spatial Information: A Disaggregate Approach. In Proceedings of the Spatial Information Theory A Theoretical Basis for GIS, Semmering, Austria, 21–23 September 1995; Hirtle, S.C., Frank, A.U., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; p. 143-144.

5. Ibid., p. 145

6. Jacques Bertin, Graphics and *Graphical Information Processing*, De Gruyter, 1981.

7. Jean-Hugues Chauchat, Alban Risson A. Bertin's Graphics and Multidimensional Data Analysis in J. Blasius and M. Greenacre (eds.) *Visualization of Categorical Data*, London: Academic Press, 1998, pp. 37-45.

8. Ludovic Lebart, 'Complementary Use of Correspondence Analysis and Cluster Analysis' in M. Greenacre and J.Blasius (eds.), *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, London: Academic Press, 1994, pp.162-178.

9. Daniel Dorling ve Bethan Thomas, P*eople and Places: A 2001 Census Atlas of the UK*, Policy Press: Bristol, 2004.

10. Jacques Bertin. *Semiology of Graphics; Diagrams, Networks, Maps*, Redlands CA: ESRI, 2011, p.16.

11. Anthony C. Gatrell, Any Space for Spatial Analysis. R. Johnson (der.) *The Future of Geography*, New York NY: Routledge, 2014, pp. 190-208.

12. It should be mentioned in passing that large settlements such as Istanbul Ankara Izmir are represented by 1,000-1,200 units.

13. Normalization is done over row sums.

14. AThe same operation can also be performed on expected values with a statistical approach.
See: Jacques Bertin, *Graphics and Graphical Information Processing*, De Gruyter, 1981, pp. 224-225.

15. Before the computer, this was done with a device called a **domino or permutator**. See: Jacques Bertin, *Graphics and Graphical Information Processing*, De Gruyter, p.168 & p.256.

# NOTES

16. Jean-Hugues Chauchat, Alban Risson A. Bertin's Graphics and Multidimensional Data Analysis in J. Blasius and M. Greeacre (eds.) *Visualization of Categorical Data*, London: Academic Press, 1998, pp. 37- 45

17. Violet, M. Gray, Classification as an Impediment to the Reliable and Valid Use of Spatial Information: A Disaggregate Approach. In Proceedings of the Spatial Information Theory A Theoretical Basis for GIS, Semmering, Austria, 21–23 September 1995; Hirtle, S.C., Frank, A.U., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; p. 145

18. Ibid., p. 143

19. For the weaknesses of correspondence maps regarding visual communication, see: Jean-Hughues Chauchat, Alban Risson A. Bertin's Graphics and Multidimensional Data Analysis in J. Blasius and M. Greeacre (eds.) *Visualization of Categorical Data*, London: Academic Press, 1998, pp. 37-45For the effects of unusual observations and characteristics on the results of correspondence analyses, see: Ludovic Lebart, 'Complementary Use of Correspondence Analysis and Cluster Analysis' in M. Greenacre and J.Blasius (eds.), *Correspondence Analysis in the Social Sciences: Recent Developments and Applications*, London: Academic Press, 1994, pp.162-178.

20. Ludovic Lebart, 'Complementary Use of Correspondence Analysis and Cluster Analysis' in M. Greenacre and J.Blasius (eds.), Correspondence Analysis in the Social Sciences: Recent Developments and Applications, London: Academic Press, 1994, pp.162-178.

21. Correspondence analysis represents observations with a single point on the map of **row points** and features with a single point on the map of **column points.** The fact that observations are represented with a single point, regardless of the number of indicators, facilitates communication. Congruent profiles are located close to each other, incongruent profiles are located far apart. Each position corresponds to a distinctive profile.

22. Correspondence analysis is used in applications based on crosstabs, and multiple correspondence analysis is used in tables consisting of qualitative indicators.

23. Correspondence analysis explains all the information in the dataset through m-1 dimension factors. Dimensions are listed from most explanatory to least explanatory. The excluded dimensions are usually related to single observations, this simplification and reduction of the number of dimensions of the table is important for filtering out unusual observations. It is impossible to introduce all the characteristics of correspondence analysis within the scope of this brief. For a comprehensive review see: Phillips D. and Phillips J. 'Visualising Types The Potential of Correspondence Analysis' in D. Byrne and C.Ragin (eds.) *Case Based Methods*. London: Sage, 2013, Chapter 8.

24. For large tables, Lebart recommends the use of k-means-type genetic algorithms in the first step and a hierarchical model in the second step. In all clustering applications, observations are weighted by their scale and dimensions by their eigenvalue share. If the number of observations is low, the Ward algorithm is used, which minimizes the variance within cluster.

25. Uprichard, E, 2013 'Introducing Cluster Analysis: What Can It Teach us about the Case? in D. Byrne and C.Ragin (eds.) *Case Based Methods*. London: Sage, Chapter 7

26. Jiawei Han, Micheline Kamber, ve Jian Pei, ***Data Mining: Concepts and Techniques***, 3rd edition, Haryana, India; Burlington, MA: Morgan Kaufmann, 2011, ss.519-22.

## How to cite:

Güvenç, Murat. 2022. "Qualitative Synthesis Maps for Data-Driven Urban Governance: A Methodological Evaluation" TESEV Briefs, 2022/12.

https://www.tesev.org.tr/en/research/qualitative-synthesis-maps-for-data-driven-urban-governance-a-methodological-evaluation/

This brief was translated from its Turkish original by Sebastian Heuer.



This publication has been prepared with the financial support of the European Union within the scope of the CEIDizler Grant Programme. The content is the sole responsibility ofTurkish Economic and Social Studies Foundation (TESEV) and does not necessarily reflect the views of the European Union and Association for Monitoring Gender Equality.



ceid CİNSİYET EŞİTLİĞİ İZLEME DERNEĞİ

TESEV