Recognition of Risk Information --- Adaptation of J. Bertin's Ordinal Matrix for social communication

Keiichi Ishida [a]
[a] Hosei University / Zurich Insurance company, Tokyo, Japan

Abstract: This paper aims to show capability of the Ordinal Matrix of Jacques Bertin which is a visualization method of data analysis and/or a method to recognize data. This matrix shows data by replacing numbers with visual elements. As an example, using a data set of natural hazard rankings for certain metropolitan cities in the world, this paper describes how the Ordinal Matrix handles the data set and shows characteristic factors of this data to understand it. Not only showing a kind of risk ranking of cities, the Ordinal Matrix shows how different dangers impact different cities. Furthermore, we will see that the visualized data by Ordinal Matrices allows us to see the characteristics of the data set comprehensively and instantaneously.

Keywords: Visual analytics, Data mining, Decision making support, Social communication, Ordinal Matrix, Jacques Bertin.

1. INTRODUCTION

Risk information should be assessed, for risk management purpose, based on the perceptions of risk stakeholders to control it. There are certain impact levels of risk causing unexpected results individually and/or socially. In the case of risks relevant to social benefit, specialist's assessments are significantly more important than ordinary peoples' in terms of quality and quantity. However, looking at current circumstance such as nuclear power plant security, product liability law to protect consumers, informed-consent processes in hospitals, etc., people tend to think that one-way explanations from specialists are not sufficient to determine what measures should be taken in the next steps against concerned risks. The specialists' advice is sometimes too simplified (i.e. yes or no) or too complicated (i.e. many specialized terms) for other stakeholders. There must be an establishment of decision making based on mutual understanding of risk information among all stakeholders. Only then, mutual understanding of risk information will be integrated into improvements in risk management.

Considering the gap of knowledge between specialists and ordinally people (*qua* non-specialists), it is not realistic to have the totally same level of understanding for all concerned peoples. Therefore, a mutual understanding between specialists and non-specialists is the key point in this paper. It is not an issue in this paper how ordinary people can look thoroughly at all detailed documents, therefore becoming "virtual"specialists. The issue in this paper is about communication to avoid omissions of potentially understandable information as much as possible regarding concerned matters rather than to provide just one definitive result. For example, in the insurance industry there is aa metric called Probable Maximum Loss (PML) to measure the likelihood of vulnerability of real estate properties against natural hazards in a given return period per year, i.e. a PML for a 500 year return period. This is normally calculated by probabilistic simulation software based on multiple factors: location of earthquake epicenter, soil quality, ground shaking intensity, building foundation techniques, construction type, number of stories, shape of building, etc. One property owner who spent a lot of money to build a solid structure could have a similar PML result with another owner who did not spend the money to build a solid structure. A reason could be due to location of property: how far from earthquake epicenters, soil quality, or due to many other reasons. There are many different factors of PML calculations for different property owners. A specialist might think that it is too complicated to explain PML results in an understandable way to a non-specialist property owner. On the other hand, it is sometimes difficult for the non-specialist to believe a result which is simply provided as a percentage of damage without any explanation for the non-specialist. Or the non-specialist may wonder if all necessary considerations were appropriately considered in the result. Certainly, it is very hard to follow how the specialist establishes each of the calculations, and, furthermore, difficult to understand which factors would make different results between a PML of 5.1% and a PML of 4.6%, for example. But ordinary people can grasp which elements are essential factors without deeply comprehending how each factor is used in complicated calculations.

One solution is to have an intermediate level of communication for a situation such as above, and another solution is a communication tool which can easily show the key points. Text or language is commonly used for communication between people. However, since letters and words only represent phenomena or events, understanding of between experts and non-experts are sometimes different. For example, the word "risk" has several meanings: pure risk, which causes purely negative impacts; speculative risk, which includes a possibility of benefit and loss; and other meanings might be used in certain situations. For accurate information transmission in communication, the time spent in communicating is an important issue to share correct definitions of words and contents used in discussions. Therefore, not only text explanation, but illustrations are commonly used to describe certain complex things and can also explain them to different language users.

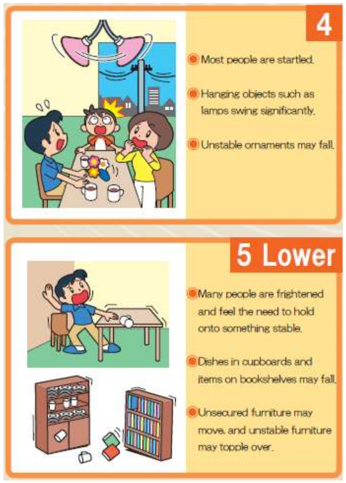

Fig. 1. Examples of JMA seismic intensity scale (source: JMA)

A brochure of the seismic intensity scale issued by Japan Meteorology Agency (JMA) is an example (Fig.1). By using the illustration, the difference between strong ground shaking on the JMA scale of intensity 4 and intensity 5-Lower (the scale limit is 7) is easily seen. Rather than taking time to read detailed explanations by text, it is easy to understand the circumstances by illustrations. Also, a comparison of these 2 illustrations gives a certain idea of the difference between earthquake intensity 4 and 5-Lower. The visualizations are used because of their excellent characteristics of information transmission: almost instantaneous understanding.

2. FORMATTING AND UNITS

Visualizations are useful for sharing an idea between people, however, this visual communication must be used with consideration as to how visual materials can be accurately applied. Jacques Bertin's Ordinal Matrix, which is referred to as an analysis tool in this paper, describes numerical data sets by visual images. This means a replacing ordinal numbers (1^{st}, 2^{nd}, 3^{rd} …) with visual elements. The visual elements used must be consistently relative to the original data figures and have a good visibility, as such, replaced visual element. We refer again J. Bertin's work which established graphic semiology and contributed to graphic and cartographic language to provide good graphical communication between people without causing visual confusion. He developed a notion of visual variables based on perception effects. As Fig.2 is shows, there are 6 visual variables on two dimensions: Size, Value, Texture, Color, Orientation and Shape.

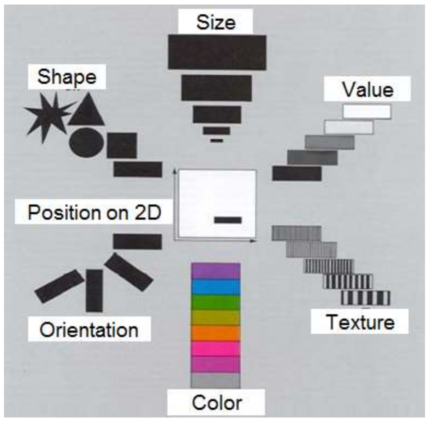

Fig. 2. Visual variables (source: J. Bertin)

Two variables, Size and Value, give a clear selectivity regarding difference of numerical figures, whereas the other variables give same visibility. Fig.3 shows a comparison of how each variable is identified and its visibility and selectivity in a sample matrix. The Size of large or small scale and/or the Value of deep or light density give different perceptions in terms of numerical data: Large numbers are represented by large Size or dense Value, and small numbers are represented by small Size or light Value. This rule of representation is generally understandable for anyone who observes visualized data.

In the case of the Texture, Color, Orientation and Shape, we can see a variable in a cell giving different “appearances”, but there is no different intensity of visibility. The Texture, Color, Orientation and Shape does not have “inherently” any *a priori* meaning of view strength. So, there is no natural, cross-cultural meaning to help evaluate visual data between red and green color, triangle and round shape, horizontally and vertically oriented bar..., etc. Furthermore, if we would like to use the Color or other variables to show numerical differences, additional definitions are required: Which color represents which number. However, no one has, *a priori*, any definitive combination of colors with numbers (no common sense that red is larger than green, for example). Therefore, such created definition requires observers’ unnecessary tasks: a conversion in mind from colors to relative data figures complying with the definition.

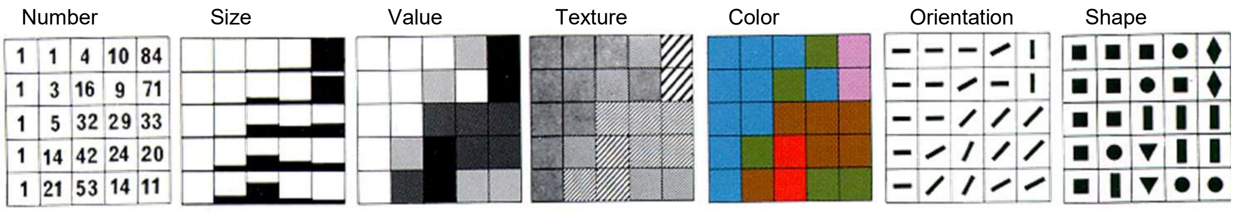

Fig. 3. Comparison of visual variables in matrix (source: J. Bertin)

J. Bertin’s Ordinal Matrix is a visualization of data which is a replacement of numeric data table by visual variables (Size and/or Value are the main points in this paper). And another characteristic is re-ordering of rows and columns in matrix. A matrix with permutations of rows and columns improves visibility of the table and makes it more readable, in view of the concerned data set. Then, the Ordinal Matrix lets observers grasp insights of the data as a whole. An example is in Fig.4. Objective data comes from a survey with respect to urban equipment for several areas where people live. And this tried to discover a categorization: city, suburban or country, depending on different equipment in each place in question. An Ordinal Matrix, just after replacing numerical figures to visual element, is shown at (1) in Fig.4; and (2) is after re-ordering of columns and rows. By looking at the Ordinal Matrix of (2), the observer can easily and quickly see the categories of area: city, suburban or country by identification of which equipment makes such categorization (highlighted by red frame). Areas categorized as city are C, H and K where we see properties: “high school”, “station” and “police station”. On the other hand, the county is represented by “only 1 class in school”, “no medicine” and “no water conveyance”. So, the Ordinal Matrix makes observers relatively easily to see correspondences between township categories and urban equipment.

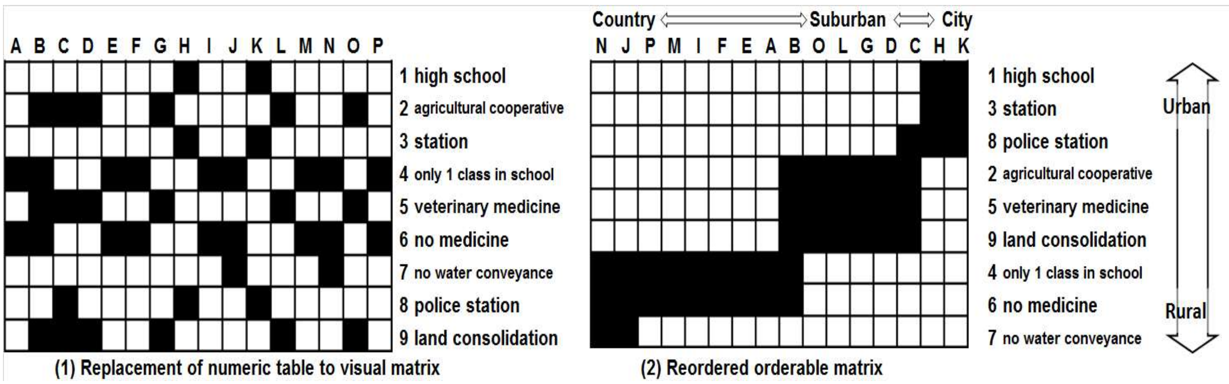

Fig. 4. Example of data handling in the Ordinal Matrix (source: J. Bertin)

2.1. Application to risk information

A report published by a reinsurance company provides a global ranking for metropolitan cities potentially affected by five natural catastrophes: Earthquake, Tsunami, Wind storm, (or Typhoon), River flood and Storm surge. The ranking indicates Tokyo-Yokohama region as the first place as shown in Tab.1. About 57.1 million people are potentially affected. Some journals published articles, saying that

Tokyo is the most vulnerable city or that Tokyo-Yokohama metropolitan area is the riskiest place in the world to do business.

This statistic was made by a simple accumulation of potentially affected peoples by the five aforementioned natural catastrophes. At the same time, this report provides top 10 rankings for each of the five events respectively in their annex as shown in Tab. 2. Undoubtedly, one summarized statistic table as Tab.1 can show data more simply than several tables as shown in Tab.2. Here is a dilemma for the presenter in how to introduce the data: a brief idea with less figures of data, or a deep understanding with much more data. Naturally, it would be better to handle all available data and to see the data more accurately if we can see these data in a convenient manner. The ranking tables for respective natural catastrophe are very interesting, but it will take much time to get the same idea of the summarized table as Tab.1, indeed. In other words, it is almost impossible to bring such summary to mind. We need to take a mental note as to which city is concerned which natural events, how serious damages are…, etc. One solution is to use the Ordinal Matrix to provide us an observational approach to see all necessary data without compromise. So, for an comprehensive understanding of a risk portfolio, the detailed ranking data sets are going to be analyzed by an Ordinal Matrix.

Tab. 1. Number of people potentially affected by 5 natural catastrophes (in million) (Source: Mind the risk)

Tokyo-Yokohama (JPN)	57.1
Manila (PHL)	34.6
Pearl-River Delta (CHN)	34.5
Osaka-Kobe (JPN)	32.1
Jakarta (IND)	27.7
Nagoya (JPN)	22.9
Kolkata (IND)	17.9
Shanghai (CHN)	16.7
Los Angeles (USA)	16.4

Tab. 2. Number of people potentially affected for five perils (in million) (Source: Mind the risk)

Earthquake ranking			
	Metro area	Area (km²); population (mn)	People potentially affected(mn)
1	Tokyo-Yokohama (JPN)	16 300 37.1	29.4
2	Jakarta (IDN)	11 600 33.1	17.7
3	Manila (PHL)	2 900 20.9	16.8
4	Los Angeles (USA)	14 400 15.4	14.7
5	Osaka-Kobe (JPN)	13 600 18.6	14.6
6	Tehran (IRN)	11 000 15.1	13.6
7	Nagoya (JPN)	15 600 11.6	9.4
8	Lima (PER)	2 600 8.9	8.9
9	Taipei (TWN)	2 100 8.1	8.0
10	Istanbul (TUR)	4 100 11.5	6.4

Storm ranking			
	Metro area	Area (km²); population (mn)	People potentially affected(mn)
1	Pearl River Delta (CHN)	20 600 42.4	17.2
2	Tokyo-Yokohama (JPN)	16 300 37.1	14.1
3	Manila (PHL)	2 900 20.9	12.6
4	Osaka-Kobe (JPN)	13 600 18.6	7.8
5	Taipei (TWN)	2 100 8.1	5.4
6	Shantou (CHN)	5 200 10.0	5.1
7	Nagoya (JPN)	15 600 11.6	4.3
8	Mumbai (IND)	2 600 20.6	4.3
9	Chennai (IND)	1 600 8.5	4.0
10	Tainan-Kaohsiung (TWN)	3 100 5.1	4.0

River flood ranking			
	Metro area	Area (km²); population (mn)	People potentially affected(mn)
1	Pearl River Delta (CHN)	20 600 42.4	12.0
2	Shanghai (CHN)	8 000 17.6	11.7
3	Kolkata (IND)	3 200 19.1	10.5
4	Jakarta (IDN)	10 600 33.1	10.0
5	Delhi (IND)	5 700 21.9	8.9
6	Tokyo-Yokohama (JPN)	16 300 37.1	8.9
7	Bangkok (THL)	3 500 9.5	7.1
8	Mexico City (MEX)	4 500 19.6	6.1
9	Cairo (EGY)	3 500 17.7	5.5
10	Tianjin (CHN)	2 600 5.8	5.5

Storm surge ranking			
	Metro area	Area (km²); population (mn)	People potentially affected(mn)
1	Pearl River Delta (CHN)	20 600 42.4	5.3
2	Osaka-Kobe (JPN)	13 500 18.6	3.0
3	Mumbai (IND)	2 600 20.6	2.6
4	Tokyo-Yokohama (JPN)	16 300 37.1	2.3
5	Amsterdam-Rotterdam (NDL)	10 800 5.4	1.8
6	Nagoya (JPN)	15 600 11.6	1.7
7	Shanghai (CHN)	8 000 17.6	1.4
8	Kolkata (IND)	3 200 19.1	1.4
9	Ho Chi Minh (VNM)	2 000 9.1	1.3
10	New York-Newark (USA)	11 900 16.5	1.1

Tsunami ranking			
	Metro area	Area (km²); population (mn)	People potentially affected(mn)
1	Tokyo-Yokohama (JPN)	16 300 37.1	2.4
2	Nagoya (JPN)	15 600 11.6	2.4
3	Osaka-Kobe (JPN)	13 500 18.6	1.8
4	Shantou (CHN)	5 200 10.0	0.7
5	Kolkata (IND)	3 200 19.1	0.6
6	Dhaka (BGD)	1 800 12.9	0.4
7	Izmir (TUR)	1 400 2.7	0.4
8	Jiaojing (CHN)	2 700 3.1	0.3
9	Guayaquil (ECU)	900 2.4	0.3
10	Chennai (IND)	1 600 8.5	0.2

2.2. Overlooking of data without omission

The data previously shown in Tab. 2 contains not only the number of people potentially affected by the different natural catastrophes, but also the number of people living in concerned agglomeration areas. Thus, this data can lead observers to see damage impacts by a percentage of people affected by the five different natural catastrophes to total people living in cities.

Now, all data is integrated into one table as shown in Tab.3. Then, this numerical data table is replaced by an Ordinal Matrix as Fig.5. The width of columns represents the number of people living in each listed city areas: areas with wider columns represent that number of people living in that area are represented much people than ones with narrow width. By this arrangement, the observers can see a potential social exposure against natural catastrophes. The damageability (number of people affected by natural event divided by a total number of people living in concerned city) is presented by different sizes of blocks in each matrix cell. So, Size of black blocks in cells show importance of damage due to each event: bigger blocks represent higher damageability. Finally, based on the perception of such importance of cells, rows and columns are permuted to have a better view how different risks are in each area as shown in Fig.6. The arrangement of permutation is done by a principal rule: similar views of intensity of cells shall be gathered. Then, certain groups of cells emege in the Ordinal Matrix. As shown in Fig.6, we see 4 groups which are highlighted in a red frame and some exceptions shown in blue. So, the variable of black blocks, being initially scattered separate parts of the matrix, are now re-ordered to be readily seen, therefore showing what are the main drivers producing the characteristic of this data set and what correspondences are observed between natural catastrophes and cities.

Tab. 3. Potential cities affected by 5 natural catastrophes (in million) (based on the original data from Mind the risk)

City	Tokyo-Yokohama	Jakarta	Manila	Los Angeles	Osaka-Kobe	Tehran	Nagoya	Lima	Taipei	Istanbul	Pearl River Delta	Shantou	Mumbai	Shanghai	Kolkata	Delhi	Bangkok	Mexico City	Tianjin	Cairo
Country	JPN	IDN	PHL	USA	JPN	IRN	JPN	PER	TWN	TUR	CHN	CHN	IND	CHN	IND	IND	THL	MEX	CHN	EGY
Population	37.1	33.1	20.9	15.4	18.6	15.1	11.6	8.9	8.1	11.5	42.4	10.0	20.6	17.6	19.1	21.9	9.5	19.6	5.8	17.7
Earthquake	29.4	17.7	16.8	14.7	14.6	13.6	9.4	8.9	8.0	6.4										
Wind storm	14.1		12.6		7.8		4.3		5.4		17.2	5.1	4.3							
River flood	8.9	10.0									12.0			11.7	10.5	8.9	7.1	6.1	5.5	5.5
Storm surge	2.3				3.0		1.7				5.3		2.6	1.4	1.4					
Tsunami	2.4				1.8		2.4					0.7			0.6					

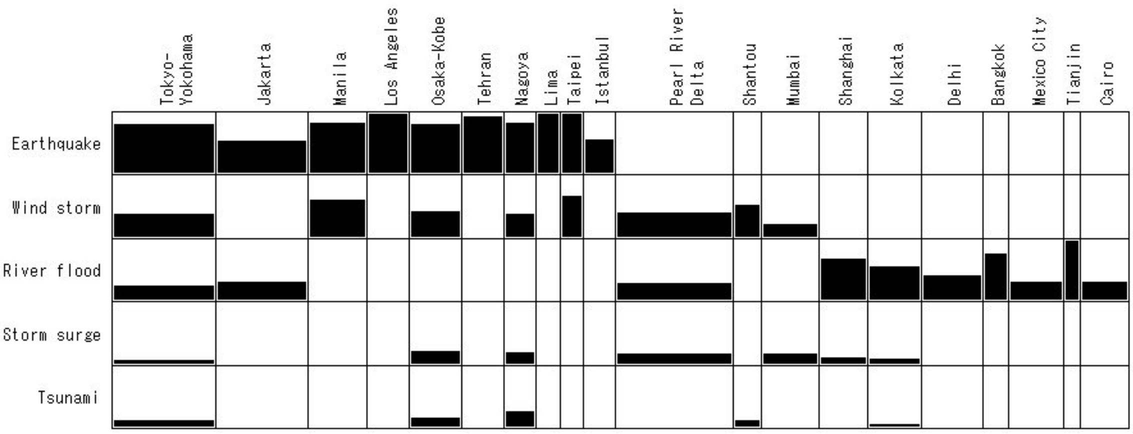

Fig. 5. Ordinal Matrix

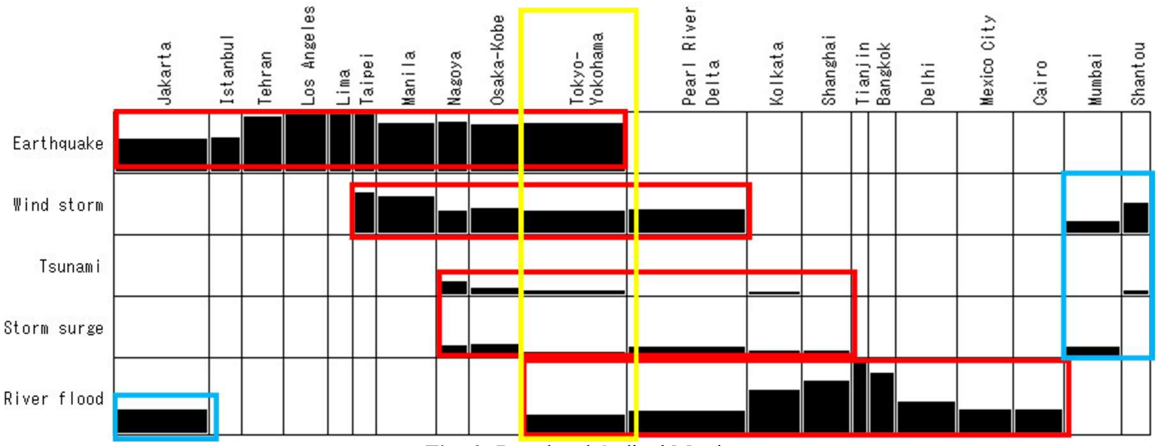

Fig. 6. Reordered Ordinal Matrix

Firstly, we should understand that each different natural catastrophe does not cause the same level of damageability to the cities. In a holistic level of view, it is understood that the Earthquake, Wind storm and River flood give stronger impact to the cities than Tsunami and Storm surge. And we see one combination of risk on Tsunami and Storm surge by grouping of cells. Both risks are shown in adjacent areas in the matrix. It is easily understood that areas near shorelines are exposed to marine related risks.

The index of 57.1 million potential victims in the Tokyo-Yokohama area shown in Tab.1, which may be the most hazardous city in the literature, has been decomposed by each natural catastrophe, and is compared with other city areas. The dangers of Tokyo area clarifies how different dangers among different cities. Tokyo is an area exposed by the all catastrophes. Osaka and Nagoya, then, follow Tokyo. This demonstrates how Japan is the most natural hazard prone area. This relative comparison in partial level view of analysis in the matrix could not be observed by Tab.1 which is a simple accumulation of number of potentially affected peoples for all catastrophes.

When we look at the matrix in detailed level, high hazard of Tokyo does not mean that the other areas are relatively secured from the catastrophes. Some of cities would have destructive damage due to certain perils. In case of that black blocks fully occupy cells means almost a 100% loss. For example, Los Angeles, Lima and Taipei should be concerned cities of total loss against earthquake. And river flood causes serious damage to other cities: Kolkata, Shanghai, Jakarta, Delhi, Mexico City, Cairo Bangkok and Tianjin. The Ordinal Matrix comprehensively shows how serious each natural catastrophe is in different cities.

The awareness through the analysis above might lead us to make a hypothesis for the next step to prevent damages due to concerned perils, not necessary to be afraid of the risk ranking. Further, it would be interesting to look for a complement data to be in blanks of table: Tab.3 which was left out from the original top 10 data set described in Tab.2. And this could allow us to see exactly the natural catastrophe exposures of each city.

2.3. Study on visualization

The data visualization by Ordinal Matrix lets us observe risk exposure against natural catastrophes in some metropolitan cities as described above. However, since there are other visual variables to use and other ways of re-ordering rows and columns in the Ordinal Matrix as suggested by J. Bertin, other ways of visual establishment are available. A matter for study here is to see different nuances resulting from other versions of the Ordinal Matrix, even if observed and/or analyzed information is not changed.

Fig.7 is one variation which shows the same data as in Fig.6. A difference of Fig.7 compared to Fig.6 is to introduce only 5 levels of different size of black block, rather totally matching visual variable size and data figures. This would allow to see a data set more simply than before.

Fig.8 is an Ordinal Matrix using a variable of Value (a gray scale gradation from black to white). And this is a simpler version than the original because of the exclusion of the notion of population size; then the all widths of columns are same. We are purely focusing on damageability of cities.

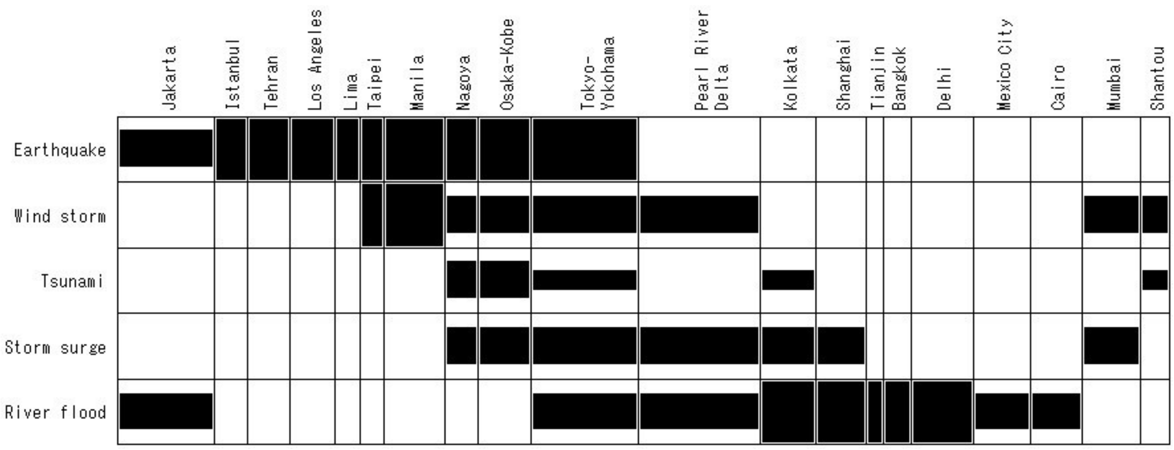

Fig. 7. Variation on Ordinal Matrix, in case of introducing 5 level of visual variable of Size

With this version, another ordering of columns could be suggested. For example, Jakarta at left side of the matrix could be at right side after Cairo as Fig.9. But again, there is no major influence on the analysis that we made before.

Any data set coming from reality contains exceptions. Therefore, it is possible to have differences in matrix views. However, one objective of using the Ordinal Matrix is not to have one definitive re-ordering method, but to be a valuable tool for social communication.

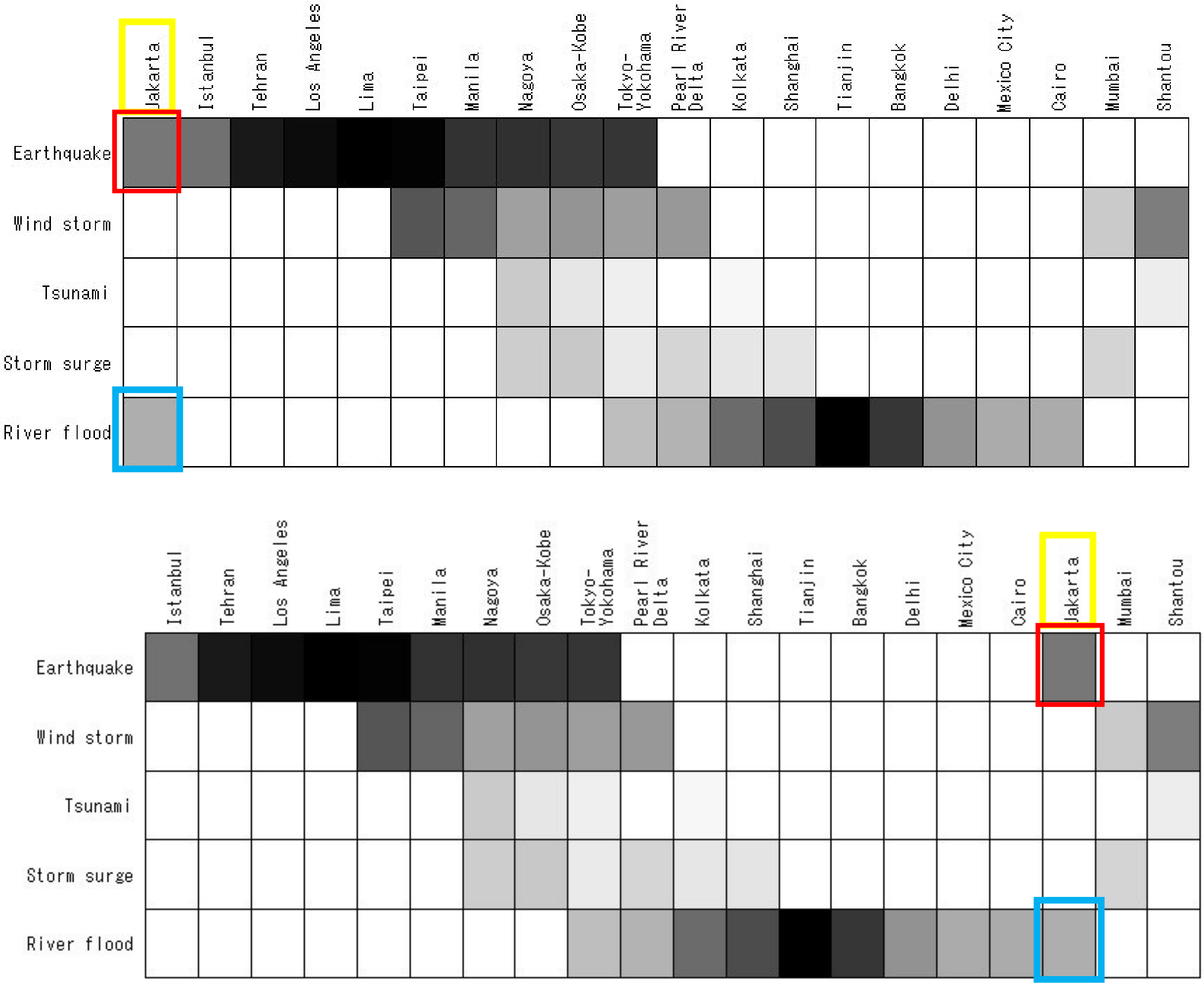

Fig. 9. Variation on Ordinal Matrix: re-postponement of item

3. Conclusions

This paper has highlighted the usability of the Ordinal Matrix of Jacques Bertin, which allows us to analyze visually a data set regarding natural catastrophes. The Ordinal Matrix has been able to show how danger concerns cities. This could not be found out from the top 10 indices of five different natural catastrophes which are presented separately by a numerical data table. The panoramic view of all data sets arranged by Ordinal Matrix allow us to see visually a portfolio of natural hazards instantaneously and comprehensively: There are no individual interventions of observers, whereas a reading of five different numerical data tables requires certain tactics to read and to understand integrally all data, totally depending on observation methods with the personal capability of data reading. So, the Ordinal Matrix itself would be considered as a recognition tool to allow each observer to have the same quality of understanding. This will be, in other words, the same level of risk perception for all stakeholders. Then, from this common view of risk, we can expect a decision commonly shared with everyone for the next step of risk management.

Nowadays, large data can be handled easily with currently developed IT infrastructure; however, analysis methods of such data is a major issue. The Ordinal Matrix could solve some aspects of this issue. Regarding natural hazard, there is other information such as historical disasters, recovery period from disasters, economical damage... etc., which can be added to the Ordinal Matrix that we used in this paper. And the large Ordinal Matrix might give principal elements to describe characteristics of large data sets based on a broader visualization of data. To demonstrate this, we need to examine the possibility of the Ordinal Matrix with more complex and large data, and then, the usability of the Ordinal Matrix will be reinforced again for social communication where we find always certain conflicts between specialists and non-specialists. This is the radical question on the Ordinal Matrix: how essential information shall be selected and how that information is arranged to show according to requirements of people who observe it.

References

[1] Bertin, J. *"La graphique et le traitement graphique de l'information"*. Flammarion. 1977, Paris.

[2] Bertin, J. *"Sémiologie graphique"*, Mouton, 1966, Paris.

[3] Fukazawa, N. "*Risk perception and human behaviours"*, Kobundo-shuppansha, 2005, Tokyo.

[4] Fekete, JD., Dragicevic, P. and Perin, C. *"Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations"*, IEEE. p2082-2091, 2014.

[5] Hirota, S., Masuda, S. and Sakagami, T. *"Physiologically described world of risk."*, Keikodaigaku-shuppankai, 2002, Tokyo.

[6] Ishida K. *"Analyses of Images of streets in France and Japan by cartography"*, The city planning institute of Japan, 2006

[7] Ishida K. *"Image of Streets in Japan and France, Image Map Analysis and Visualized Data Analysis"*, Journal of the Japan Cartographers Association, p12-26 Vol.47 No.1, 2009

[8] Ishida K. *"Geographical portfolio management for enterprise risk management"*, International Cartographic Association, 25th ICC, 2011

[9] Kamei. K. *"Management strategies of French company and risk management"*, Horitsubunka-sha, 2001, Tokyo.

[10] Morita, T. *"Zu no Kigougaku"*, Heibonsha, 1982, Tokyo.

[11] Neisser, U [translation by Kozaki, T]. *"Cognition and reality"*, Sceince-sha, 1978, Tokyo.

[12] Omi, M., Yosizawa, T., Takao, A., Amari, K. and Kubo, E. *"Elements of risk and insurance"*, Dobunkan, 2012, Tokyo.

[13] de Saussure, F. *"Cours de linguistique générale"*, Payot, 1992, Paris.

[14] Japan Meteorological Agency. http://www.jma.go.jp/jma/indexe.html

[15] SwissRe. *"Mind the risk"*, http://www.swissre.com/library/expertisepublication/Mind_the_ risk_a_global_ranking_of_cities_under_threat_from_natural_disasters.html. Accessed 26October 2016.

[16] Reuters, http://www.reuters.com/article/swissre-risks-idUSL5N0HD2RT20130918. Accessed 26October 2016.

[17] USA Today, http://www.usatoday.com/story/weather/2013/09/18/worlds-most-disaster-prone-cities-tokyo/2833667/. Ac-cessed 26October 2016.

[18] PML simulation software providers: AIR Worldwide, CoreLogic and RMS are commonly used. (AIR Worldwide. http://www.air-worldwide.com/, CoreLogic. http://www.corelogic.com/default.aspx, RMS. http://www.rms.com/)