DISASTERS IN THE LITERATURE OF HUMAN FACTORS: AN ANALYSIS OF SUBJECT AND APPLICATION TERMS

Nestor L. Osorio[1] and Gabriel E. Osorio

Abstract – In this preliminary study, we analyzed the literature of disasters as covered by the field of human factors. The purpose is to reach a better understanding of the structure of human factors information and its major components as they relate to disaster technological information. Classification terms from Ergonomics Abstracts were used for a co-occurrence analysis. Six clusters have been identified. Correspondence analysis to determine associations between classification terms and applications terms produced a description of their relationship.

Keywords: Co-occurrence analysis, cluster analysis, correspondence analysis, human factors, disasters.

Introduction

A basic definition of Ergonomics is presented in the *International Encyclopedia of Ergonomics and Human Factors* [1, p.102]: "Ergonomics (or Human Factors) is the scientific discipline concerned with fundamental understanding of interactions among human and other elements of a system, and the application of appropriate methods, theory and data to improve human well-being and overall system performance." It continues by indicating that there are three major areas in ergonomics: physical ergonomics; cognitive ergonomics; and social or organizational ergonomics. In the same publication [2] it is indicated that while ergonomics is related more to physical aspects human factors stresses the cognitive perspective. However, "there is a growing consensus that human factors and ergonomics refer essentially to a common body of knowledge." *Grolier's Encyclopedia* [3] defines disasters as: "a major misfortune, an occurrence that causes great suffering or loss. Its distinctive feature is the sudden and unexpected loss of human life."

In general terms two types of disasters have been defined [4]: Natural disasters (i.e., earthquakes and volcanoes) and man-made or technological disasters (i.e., nuclear accidents and aircraft crashes). More recently, terrorism is also considered a man-made disaster. Disasters belong to the occurrences characterized by complex systems and many of the research components on disasters are related to human factors. Relationships between disasters and human factors are well documented in the literature. For example, O'Hare's work [5] on accident analysis pointed out that human factors are one of the most significant reasons for the failure of a complex system. In this study, we analyzed the literature of disasters as covered by the field of human factors. The purpose is to reach a better understanding of the structure of human factors information and its major components as they relate to disaster technological information by providing some visualization analyses (cluster analysis and correspondence analysis). This is the first study that has been done about finding relationships and associations between the subject terms in these two fields. This study intends to be a starting point for further examination of the human factors literature.

Methodology

Ergonomics Abstracts (EA), a publication produced by the Ergonomics Information Analysis Center of the University of Birmingham, UK was used to collect data. This publication provides reviews of hundreds of research items from journals, conference proceedings, technical reports and books. The hierarchical classification system of EA contains 638 terms divided into twelve major categories. Each EA record contains complete bibliographic information; an abstract; classification terms; classification codes; and application terms. It is also considered to be the leading index for the literature of human factors.

The data collection process started by doing preliminary keyword searching. The initial result produced 199 entries on the entire database (1985-2003). Abstracts of these entries were analyzed producing a larger number of significant keywords that were used for a second round of searches. This method is similar to the one used by Schaible [6] in his data mining technique in order to identify trends in experimental and theoretical materials researched using *INSPEC*.

[1] Nestor L. Osorio, Northern Illinois University, Founders Library 303, DeKalb, IL 60115 nosorio@niu.edu. Gabriel E. Osorio, Northern Illinois University, DeKalb, IL 60115 gosorio75@hotmail.com.

Twelve hundred documents covering the period of 1990-2002 were obtained. After further evaluation of the abstracts five hundred and fifty documents were selected as documents related to both disasters and human factors. EA assigns a principal Classification Term and several secondary Classification Terms to each entry. Principal Classification Terms assigned to each document were tabulated. EA gives to its documents "applications" terms to describe the applicability of the work presented. For example, the "subject term" assigned to a document could be evacuation and the "application" term could be nuclear plants. Principal Applications Terms assigned by EA were also collected. This is the second part of the study of the literature of disasters and human factors. In the first part [7], using the same data mining technique, bibliographic sources (journals, books, conference proceedings, etc.), and some other characteristics of these publications were examined.

Cluster Analysis.

A total of 133 principal Classification Terms (CT) were identified; and the top 34 CT's representing nearly 75% of the total number of documents were selected. These 34 CT's became the basis for a co-occurrence analysis, see Table 1.

Table 1. Classification terms

#	CLASSIFICATION TERM	%
1	Decision making & risk management	10.55
2	Evacuation procedures	21.09
3	Human reliability & system reliability	28.00
4	Work design & organization for H&S	32.18
5	Information & communication design	35.82
6	Rescue	39.09
7	Training	42.00
8	Errors, accuracy & reliability	44.55
9	Education, training & safety programs	46.55
10	Emergency services	48.55
11	Surveys, statistics & analysis	50.55
12	Data collection & recording	52.18
13	General health & safety	53.82
14	Stress	55.27
15	Information systems & communication	56.55
16	Respiratory equipment	57.82
17	Workload demands	59.09
18	Behavioral & social processes	60.18
19	Low temperature	61.27
20	Modeling human characteristics	62.36
21	Signs	63.45
22	Clothing ensembles	64.36
23	Conspicuity aids	65.27
24	Fear, anxiety, mood & emotion	66.18
25	Movement through working areas	67.09
26	Selection & screening	68.00
27	Standards, codes of practice, guidelines	68.91
28	Teamwork	69.82
29	Causation models	70.55
30	Emergency & warning devices	71.27
31	General workplace design & buildings	72.00
32	Performance strategies	72.73
33	Physical fitness	73.45
34	Psychological disorders	74.18

Co-occurrence analysis is a bibliometric technique used to systematically find relationships between keywords of subject terms. It is used in this study in preparation for determining closed relationships (clusters) between the CT's in the field of disasters and human factors.

Data for a co-occurrence analysis is obtained by searching in the database (EA) each term of the list of CT's against each other. In this case, 561 searches were performed through the entire database (1985-2003). To limit EA documents to only those related to disasters and human factors a filter was created composed of keyword terms representing the field of disasters. The total number of filtered documents obtained was 19,000 records.

Table 2. Raw Data: Term Co-occurrence

	CT 1	CT 2	CT 3	CT 4	CT 5	CT 6	CT 7	CT 8	CT 9	CT 10
CT1	0	18	29	118	320	21	256	241	87	84
CT2	18	0	2	14	49	25	24	15	21	32
CT3	29	2	0	37	16	0	14	121	4	3
CT4	118	14	37	0	114	14	76	166	288	72
CT5	320	49	16	114	0	13	43	87	82	53
CT6	21	25	0	14	13	0	20	3	14	15
CT7	256	24	14	76	43	20	0	129	119	50
CT8	241	15	121	166	87	3	129	0	48	18
CT9	87	21	4	288	82	14	119	48	0	31
CT10	84	32	3	72	53	16	50	18	31	0

The raw co-occurrence data creates a matrix of 34 variables and 34 by 34 elements with a diagonal of zeros. Table (2) shows the first 10 elements of this matrix. This data set was presented to a common factor analysis using the SAS procedure Proc Factor. The procedure followed here has been successfully used by Morris [8] in a study about visualizing structural relationships in medical informatics. The zero diagonal in the row data set was transformed by calculating the mean value of each column. This was done because it has been reported in a similar bibliometric analysis by Morris [9] that Pearson's *r* ,calculated by Proc Corr, is sensitive to zero values and can affect the outcome of the analysis, This transformation was done as a safety device.

The Proc Corr procedure of SAS converted the modified row data set into similarity/distance values (Pearson's *r* correlations). The new standardized data values are correlation coefficients that reflect similarities and differences (dissimilarities) between the 34 variables.

SAS Proc Factor was used with the raw co-occurrence matrix to run a principal components analysis, the criteria selected was for eigenvalues greater than or equal to 1. This analysis will be used to support the number of clusters chosen in the cluster analysis.

The standardized matrix was presented to the SAS Proc Cluster procedure. The basic purpose of cluster analysis is to identify groups of variables based on a defined quantitative methodology. Since the number of clusters in this project is not known a hierarchical cluster method was selected. The method chosen measured the similarity of the variables. Ward's cluster method is a hierarchical method that measures the similarity of the between-cluster sum of squares to determine homogeneity. The resulting minimized within-cluster sum of squares is known as the error sums of squares ESS [10]. In the dendrogram the vertical axis shows the ESS values.

The selection of a method in cluster analysis is based on the kind of problem being studied. For this type of bibliometric study the complete linkage method and Ward's has been reported as appropriate methods [9]. After preliminary explorations of both methods, Ward's method was selected because it gave the best interpretable solution.

Correspondence Analysis

The second part of this study analyzes relationships between Application Terms (AT) and Classification Terms (CT). Based on the data mining technique mentioned above a total of 75 ATs were identified. In this section, the top 25 CTs and ATs were chosen. They represent 67% and 60% respectively of documents found. These top terms - having at least 4 hits -can be considered the main components of the literature under examination. Term by term searches between CT's and AT's were performed producing a matrix of 25 CT by 25 AT variables with 625 elements. Table (3) shows the first 10 values of this matrix, and Table (4) are the Applications Terms.

Table 3. Cross-tabulation data

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
CT1	58	24	18	6	249	126	290	16	25	69
CT2	61	69	6	39	39	11	117	7	17	63
CT3	1	6	2	1	5	53	19	2	2	13
CT4	26	58	6	4	15	176	56	33	39	63
CT5	30	30	19	16	23	164	114	14	5	121
CT6	48	8	2	1	11	2	11	17	6	2
CT7	33	55	4	7	133	168	286	9	8	165
CT8	7	121	4	9	39	351	250	20	24	195
CT9	25	22	7	4	17	29	31	27	9	28
CT10	73	10	16	8	18	26	33	6	11	10

Table 4. Application terms

#	APPLICATION TERM	%
1	Fire services	12.36
2	Air safety	16.18
3	Natural resources	19.82
4	Air passengers	23.27
5	Armed forces	26.55
6	Nuclear power plants	29.64
7	Aircrew	32.55
8	Mining	35.45
9	Offshore industries	38.36
10	Aircraft	40.55
11	Ambulance services	42.55
12	Chemical industry	44.55
13	Helicopters	46.55
14	Decision support systems	48.36
15	Railways	50.18
16	Astronautics	51.45
17	Fault diagnosis	52.73
18	Gas industry	54.00
19	Ships	55.09
20	Coast guards	56.00
21	Diving	56.91
22	Offices	57.82
23	Air traffic control	58.55
24	Sea navigation	59.27
25	Meteorology	60.00

CT terms against AT terms identify the number of documents that are common to each pair. Correspondence analysis is a visualization technique that represents multiple associations (25 variables) into a low-dimensional coordinate space to establish relationships between the row and column variables [11]. The SAS Proc Corresp produced a two dimensional representation of these variables.

Results

Cluster Analysis

The factors analysis suggested a solution of 8 factors, Table 7, which can be interpreted as eight clusters [8]; another way to estimate the number of clusters is by using the screed plot of the eigenvalue provided in the factor analysis. The screed plot suggested that at point 6 the eigenvalues start becoming flat, Figure 3. Therefore, six clusters might be a good "stopping point" for a solution of the number of clusters found. As suggested by Sharman [10] another way to estimate the number of clusters is by plotting the SPRSQ. SPRSQ (Semipartial R- Squared) measures the loss of homogeneity when a new cluster is formed, a drop in the value of SPRSQ indicates a solution for numbers of clusters. Ward's method provides SPRSQ statistics; Figure 4 shows a significant drop in point 7, suggesting a solution of seven clusters.

The selection of numbers of clusters is a judgmental decision; the statistics provided are heuristic in nature [4]. Confronted with an eight, seven or six cluster solutions and after detailed analysis done of the clusters presented in the dendrograms we have concluded that the six-cluster model is the best interpretable solution in this case. In the dendrogram this occurs at ESS = 0.150.

Table 5. Six-Clusters Model

Clusters	Definition
CL1	COGNITION - MODELLING - PREVENTION Members: CT1; CT20; CT9
CL2	WORKPLACE & WORK DESIGN - SOCIAL & PSYCHOPHYSIOLOGICAL VARIABLES Members: CT4; CT15; CT25; CT13; CT24; CT12; CT16; CT21; CT28
CL3	EMERGENCY SERVICES - HUMAN PERFORMANCE - INFORMATION & COMMUNICATION Members: CT2; CT6; CT11; CT31; CT32; CT8; CT33; CT5; CT17; CT30
CL4	PREVENTION - ILLNESSES - CAUSATION MODELS Members: CT10; CT34; CT29
CL5	TRAINING - STANDARDS - ENVIRONMENT - PERFORMANCE Members: CT7; CT19; CT14; CT27
CL6	METHODS - SAFETY - ORGANIZATION - BEHAVIORAL & SOCIAL ASPECTS Members: CT3; CT23; CT22; CT26; CT18

The detailed analysis of the clusters in the dendrogram, Figure 2, was performed by looking at each Classification term within a potential cluster and finding the Classification Code of each of its members (a six numerical code divided into three hierarchical elements). In this hierarchical system (EA), we looked at the main level, second level, and when possible the third level of classification terms. As mentioned above EA is divided into twelve main categories (main level): General ergonomics, Human characteristics, Performance related factors, Information presentation and communication, Display and control design, Workplace and equipment design, Environment; System Characteristics, Work design and organization, Health and Safety, Social and economic impact of the system, and Methods and techniques. As in a multidisciplinary science, such as ergonomics, these major categories are not isolated islands of knowledge and there is a great level of interrelation among them. Efforts were made in producing a qualitative image of the field with the information provided by the dendrogram from the Ward's cluster analysis and by further searching the literature for a strong indicator of supporting arguments as suggested by a group of documents representing these clusters. Table 5 shows the six-cluster model.

According to the analysis done in the literature these six clusters can be described as:

Cluster 1 covers the role of psychological aspects in dealing with the prevention of disasters. For examples, the modeling of organizational cognition of risk management; the use of integrated system simulation for risk and reliability assessment; and the process of designing emergency services.

Cluster 2 covers studies of how the design of the workplace and work design play a role in the social and psychophysiological aspects of the job performed and their relation to, for example, risk, human failure/errors, hazardous situations and the like.

Cluster 3 covers human performances in emergency situations and the role of proper or improper information and communication among the parties involved.

Cluster 4 covers prevention management systems and devices to minimize causes of accidents (disasters) due to physical or emotional illnesses in the work place.

Cluster 5 covers the role of training about safety standards and other industrial codes and regulations in situations of extreme environments. For example, proper methods to deal with the stress produced by very low temperatures in an accident location.

Cluster 6 covers the benefits of prevention, response, and safety management programs and their application within an organization model of a specific workplace. The behavioral and social processes involved in implementing them for a successful work operation.

Correspondence Analysis

As mentioned before correspondence analysis is a visualization technique [11], based on the plot, Figure 1, produced by the Proc Corresp procedure, seven groups of association between AT's and CT's were identified and are shown in Table 6.

In this table applications terms are first followed by a dash, the classification terms are then listed. In the diagram, the boundaries of each group are not always clearly defined but "pockets" of CT's and AT's can be recognized.

Table 6. Correspondence Analysis Groups

	Quadrant(s)	Description
Group 1	I	Fire services; Ambulance services; Coast guards - Rescue; Emergency services; Information systems & comm.; Movement through working areas
Group 2	I	Air passengers; Offices - Evacuation procedures; Behavioral & social aspects
Group 3	I & IV	Gas industry; Mining; Railways; Offshore industries; Chemicals industry - Education training & safety; Signs
Group 4	I & II	Astronautics; Naval resources; Helicopters; Ships; Sea navigation - General health & safety; Workload & demands; Fear, anxiety & mood; Respiratory equipment
Group 5	II	Decision support systems; Armed forces - Decision making & risks
Group 6	III & IV	Aircrew; Aircraft; Fault diagnosis - Training; Stress; Low temperature; Information comm.. & design for health/safety
Group 7	IV	Nuclear power plants - Emergency & accuracy; Human reliability & systems

An interpretation of this correspondence analysis would be that the literature of human factors shows seven areas of active research involving these AT's and CT's. From the applications point of view the groups in the diagram could be interpreted as follows: Group 1 deals with fire and ambulance services; Group 2 is about people in closed environments such as in airplanes and offices; Group 3 includes several major industries (gas mining, chemical, etc); Group 4 is related to navigational devices; Group 5 is about decision support systems; Group 6 focuses on the aircrew and aircrafts; and Group 7 is about nuclear power plants. The classification terms attached to each group are according to the diagram those subject terms more associated to them. Also, in the same diagram there are two examples of very close associations: Diving - Conspicuity aids (AT21-CT23), and Nuclear power plants - Errors & accuracy (AT6-CT8). A very close association might indicate that often those two subjects together are part of an investigation. Another possible way to explore the diagram is by looking at the distances between specific AT's and CT's. A large distance between two terms might imply that little research on those combined topics have been done. For example, separated by a long distance are Rescue (CT6) and Mining (AP8), therefore, it is possible to say that at least in the last 18 years not much research has been done in this area. Further, the questions would be: why? Is this due to new mining and safety techniques? In the same fashion we could find areas of research being neglected.

Conclusions

This is an exploratory study about finding common relationships (similarities or associations) between the major subject areas of disasters as studied from the human factors perspective. Since there has been no previous analysis of this type in the literature of ergonomics, comparisons with similar studies could not be brought out. Results obtained in both the cluster analysis and correspondence analysis shows the complexity of the topic being investigated. A possible way to further investigate this topic would be by searching its literature in *Compendex* which is a more comprehensive index to the literature of engineering and engineering technology.

References

[3] Bahr, Lauren S. ed. (1996) "Collier's Members: Encyclopedia with Bibliography and Index", New York: Colliers 8, p. 256.

[4] Hy, Ronald J. and Waugh, William L. (1990) "The Function of Emergency Management", in: "Handbook of Emergency Management. Program and Policies Dealing with Major Hazards and Disasters", Edited by William L. Vaugh and Ronald, J, Hy, Greenwood Press, New York, p. 11-26.

[11] Johnson, R. A., (1998) "Applied Multivariate Statistical Analysis", 4th ed., Upper Saddle River, NJ: Prentice Hall.

[2] Karwowski, Walemar, ed. (2001) "IEA Definitions of Ergonomics", in: "International Encyclopedia of Ergonomics and Human Factors", Francis & Taylor, London, p.102.

[9] Morris, Theodore, (2001) "Structural Relationship Within Medical Informatics. A Classification/ Indexing Co-occurrence analysis". Doctoral Dissertation, Philadelphia, PA: Drexel University.

[8] Morris, Theodore, (2001) "Visualizing the Structure of Medical Informatics Using Term Co-occurrence Analysis: II. INSPEC perspective". *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology*, 38 p. 489-497.

[5] O'Hare, D. (2000) "The 'Wheel of Misfortune': a Taxonomic Approach to Human Factors an Accident Investigation and Analysis in Aviation and Other

Complex Systems". *Ergonomics*, 43(12) p. 2001-20019.

[9] Osorio, Nestor L. and Hurych, Jitka, (2003) "Literature of Disasters from the Human Factors Point of View: A Descriptive Analysis", *Collection Building*, 14 p., (Accepted 12/1/2003).

[6] Schaible, Max (2001), "Searching Scientific Databases for Guide to Experiment and Theory". *Computing in Science and Engineering* 5(4) p. 30-39.

[10] Sharma, Subhash, (1996) "Applied Multivariate Techniques", Chapter 7: Cluster Analysis, New York: Wiley & Sons.

[1] Wogalter, M.S., P.G. Dempsey, and P.A. Hancock, (2001) "Defining Ergonomics/Human Factors", in: *International Encyclopedia of Ergonomics and Human Factors*, Francis & Taylor, London, p.35-37.

```
                         Plot of Dim2*Dim1.  Symbol is value of ctype.

       „ƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒ†
  Dim2 ‚                                                                                   6  ‚
 1.100 ˆ              II                                   ‚      I                        c  ˆ
       ‚                                                   ‚                              c   ‚
       ‚                                                   ‚                             25   ‚
       ‚                                                                                      ‚
       ‚                                                   ‚                    1    20       ‚
       ‚                                                                        a     a       ‚
 0.825 ˆ                                                   ‚                                  ˆ
       ‚                                                                           11         ‚
       ‚                                                   ‚                       a          ‚
       ‚                                                     22              c   c      ca    ‚
       ‚                                                   ‚  c             10  15    23  21  ‚
       ‚                                                                              20c     ‚
 0.550 ˆ                                                   ‚                                  ˆ
       ‚                                                                            a 4       ‚
       ‚                                                   ‚       3               18 c     a ‚
       ‚                                                           a  a                    22 ‚
       ‚                                  a   5            ‚          16              c 2     ‚
       ‚a 14         c 1                                                   13 a     a 18      ‚
 0.275 ˆ                       a                           ‚         13 c                     ˆ
       ‚                      25                                                      c 9     ‚
       ‚                                                   ‚      17 c  c 24                  ‚
       ‚                                                                a                     ‚
       ‚                                         24 a      ‚    c       19              a 8   ‚
       ‚                                                       16                             ‚
 0.000 ˆ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ˆ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ƒ ˆ
       ‚                                            7 a                 15 a    a             ‚
       ‚                               14  19              ‚                    9  c 12       ‚
       ‚                   a     a       cc         c 7                      c                ‚
       ‚                  17    23             c           ‚              a  21               ‚
       ‚                                       5                         12                   ‚
-0.275 ˆ                                                   ‚                      a 2         ˆ
       ‚                                                      a               4 c             ‚
       ‚                                                   ‚  10                c 11          ‚
       ‚                                                   ‚6                                 ‚
       ‚                                                    ac                                ‚
-0.550 ˆ       III                                    c  ‚ 8                            IV    ˆ
       ‚                                              3                                       ‚
       ŠƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒŒ
       -1.00       -0.75       -0.50       -0.25       0.00        0.25        0.50        0.75

                                                   Dim1
```

Figure 1. Correspondence Analysis Diagram. (a = AT, column coordinates; c = CT, row coordinates)

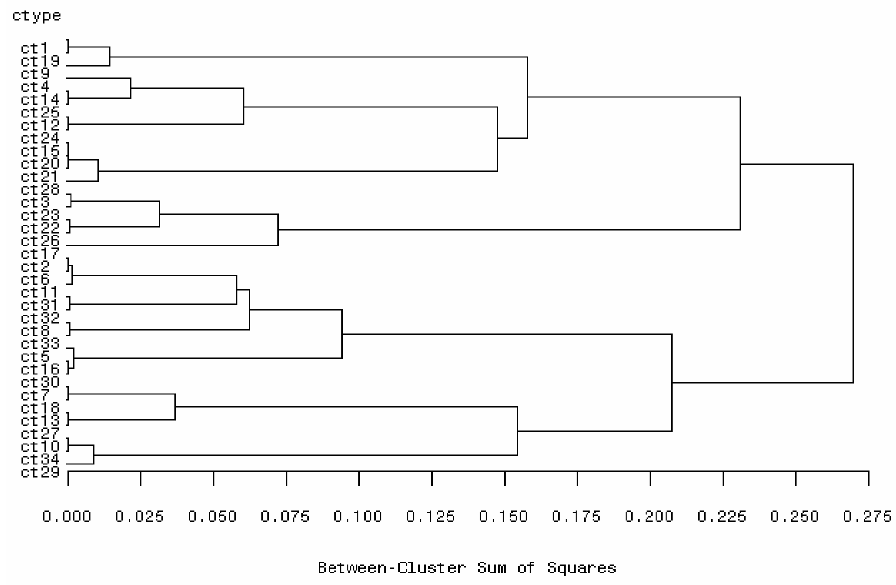

Figure 2. Cluster Analysis Dendrogram. Ward's method. X axis = ESS; Y axis = clusters

Table 7. Eight factors based on raw data

Factor 1	Factor 2	Factor 3	Factor 4
9.7722765	4.8772304	3.9955248	3.6670924
Factor 5	Factor 6	Factor 7	Factor 8
1.7230677	1.5510403	1.3474139	1.0735070

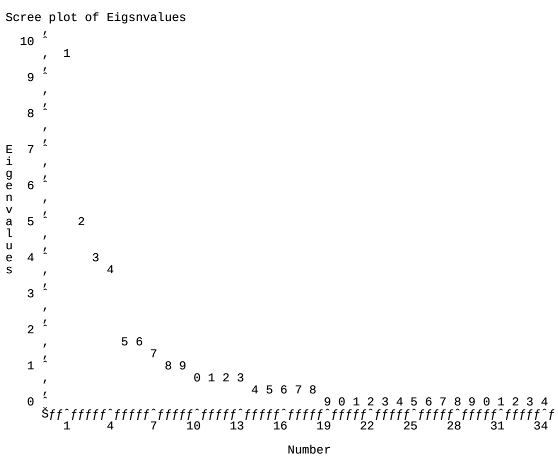

Figure 3. Plot of eigenvalues. X axis = number of factors.

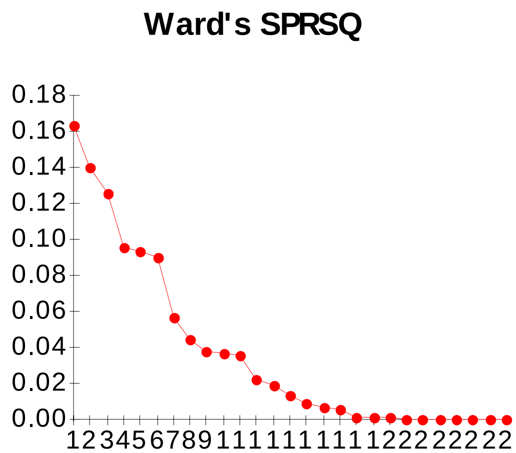

Figure 4. Ward's SPRSQ. X axis = number of clusters