

Big Data Analysis of Public Library Operations and Services by Using the Chernoff Face Method

Abstract

Purpose – The purpose of this paper is to conduct a big data analysis of public library operations and services of two cities in two countries by using the Chernoff face method.

Design/methodology/approach – The study is designed to evaluate library services by analysing the Chernoff face. Big data on public libraries in London and Seoul were collected respectively from CIPFA and the Korean government's website for drawing a Chernoff face. The association of variables and human facial features was decided by survey. Although limited in its capacity to handle a large number of variables (eight were analysed in this study) the Chernoff face method does readily allow for the comparison of a large number of instances of analysis. 58 Chernoff faces were drawn from the formatted data by using the R programming language.

Findings – The study reveals that most of the local governments in London perform better than those of Seoul. This consequence is due to the fact that local governments in London operate more libraries, invest more budgets, allocate more staff and hold more collections than local governments in Seoul. This administration resulted in more use of libraries in London than Seoul. The study validates the benefit of using the Chernoff face method for big data analysis of library services.

Practical implications – Chernoff face method for big data analysis offers a new evaluation technique for library services and provides insights that may not be as readily apparent and discernible using more traditional analytical methods.

Originality/value – This study is the first to use the Chernoff face method for big data analysis of library services in library and information research.

Keywords Public libraries, Chernoff face, Big data analysis, Performance evaluation, Data visualisation.

Paper type Research paper

Introduction

Since the world's first modern public library was established in 1852 in Manchester in the U.K., several thousand public libraries have been operating for more than 160 years all over the world. Although public library operations have been shrinking in many ways in

1
2
3
4
5
6
7
8 some countries, e.g. Finland, Germany, Norway, the UK, the USA, for the last 10~20 years,
9 in some other countries, e.g. Australia, Canada, Korea, Japan, public library services have
10 been steadily extending and developing. Surprisingly, in the U.K. 443 libraries (excluding
11 mobile libraries) were closed, opening hours were reduced and many local libraries are
12 operating by community volunteers between 1995 and 2014 (CIPFA, 1996; 2015). In stark
13 contrast, in Japan the number of libraries increased from 2,951 to 3,261 between 2005
14 and 2015 (JLA, 2015) and in Korea 438 new libraries have been built during the period
15 between 2004 and 2014 (NLSSK, 2015). It seems that the development of public library
16 infrastructure in Korea is well supported by the result of the research on the cost-
17 effectiveness of public library services carried out in Korea (Ko et al., 2012). The research
18 reveals that the ROI (Return On Investment) is 3.66 : 1. This means that there may be a
19 return of economic benefit of Korean Won ₩3,660 (about US\$ 3.18 based on currency
20 exchange rate of 1,152 on June 21, 2016) for every ₩1,000 spent on public libraries.

21 Although 'public libraries are no longer the taken-for-granted institutions they were
22 some decades ago' (Vakkari et al., 2014, p.927) in some developed countries, public
23 library services are still very important and necessary to local people. As Goulding (2006,
24 p.3) claims, 'the public library service is a thriving and dynamic system, successfully
25 supporting the literary, leisure, learning and cultural pastimes of the population' and also
26 providing various useful forms of information. Thus, in order to develop the operations,
27 and to improve services of public libraries further, a vast amount of research into their
28 effectiveness has been carried out throughout the world.

29 Since library operations and services are being weakened in some countries but in
30 contrast, developed in some other countries, the researchers considered that the analysis,
31 and comparison of the library operations and services in Korea and the UK will be
32 beneficial for developing and improving library services in both countries. This is because
33 the research involves comparison of two countries in which the library operations are
34 in stark contrast. Furthermore, it is recognized that examining which local government
35 performs better than others in one city is crucial to discover reasons for the decline of
36 library operations and services at a local and a national level. There are also practical
37 benefits to be had from visualization of the results of the analysis and comparison of all
38 of the local governments in a manner that can be easily understood by the community.
39 Nowadays, there is high interest in using big data analysis for academic research and
40 business. This trend leads to consideration of the importance of visualization of the
41 results of data analysis and recognition of the use of the Chernoff face as a useful
42 method for data visualization. The Chernoff face method is a very impressive tool, which
43
44
45
46
47
48
49
50
51
52
53
54

1
2
3
4
5
6
7
8 presents multivariate data as human facial features. Therefore, each Chernoff face can
9 represent the data for one person or one institution or one company or one country, etc.
10 That is to say, one Chernoff face can represent various data relating to public library
11 operations and services. As human faces are easily recognized individually, Chernoff faces
12 which individually represent each public library or each local government can be easily
13 identified. The Chernoff face is an effective method for comparing the performance
14 results of library operations and services in one city or one country.
15

16
17 The idea of using the Chernoff face method for this study also comes from the fact
18 that since in general, people in London and Seoul have different lifestyles and
19 expectations, there might be a different Chernoff face for library operations and uses in
20 the two cities. Therefore, the study aims to examine how the Chernoff face of the two
21 cities is different. As the first author is a Korean and has extensive knowledge about
22 public libraries of both Korea and the UK, he has conducted a comparison of the public
23 libraries of both countries. This comparison aims to examine what is different in the
24 library operations and services in Korea and the UK, whilst simultaneously testing the
25 benefit of using the Chernoff face method in library research.
26
27
28
29

30 The literature

31
32 'Chernoff face was first proposed by Herman Chernoff in 1973, as a way to present
33 multivariate data in a manner that is easily discernible by the human viewers' (Morris,
34 Ebert, and Rheingans, 2000). Lee et al. (2013, p.961) claim that 'Chernoff face analysis is
35 an interesting statistical method that people who have no special knowledge on it can
36 use in many areas'. The Chernoff face method has been used in many areas of research
37 such as social science (Turner, 1979; Ki, 2016), medical science (Rahu, 1989; Lott &
38 Durbridge, 1990), consumer research (Golden & Sirdesal, 1992), business administration
39 (Nel, Pitt, and Webb, 1994; Lee et al., 2013), sports analysis (Yau, 2011; Yang, 2012), and
40 political science (Spinelli & Zhou, 2004; Ki, 2015).
41

42
43 'The point of Chernoff face is to display multiple variables at once by positioning parts
44 of the human face, such as ears, hair, eyes, and nose, based on numbers in a dataset.
45 The assumption is that you can read people's faces easily in real life, so you should
46 recognize small differences when they represent data' (Yau, 2011). Yau confirms that the
47 Chernoff face method is one of many computer-supported data visualization tools. Chen,
48 Floridi and Borgo (2014) claim that computer-supported data visualization can save the
49 time of researchers in creating visualization and, of readers in understanding visualized
50 messages. As they argue, the method particularly enabled the authors of this study to
51
52
53
54

55
56
57
58
59
60

Comment [u1]: Because the reviewer says that the sentence "since ... in the two cities" is meaningless and not clear, thus, "difference faces" in the sentence have been changed into "different lifestyles and expectations".

1
2
3
4
5
6
7
8 accomplish more easily a visualization task and, will help the readers of this paper
9 understand multivariate data more quickly.

10 The Chernoff face method is used in sports analysis, e.g. the performance of NBA
11 basketball players (Yau, 2011), Korean baseball players (Yang, 2012). Lee et al. (2013)
12 used the method in evaluating 15 banks with four major bank economic indicators. The
13 research presented four different groups of Chernoff faces with 11 variables. The
14 research reveals that the method is a very useful for data analysis, but the method has
15 some disadvantages. Firstly, since people have different preferences for human faces,
16 people may interpret Chernoff faces based on their own feelings. Secondly, the method
17 can handle big data in the sense of many instances, for which there are only on a
18 relatively small number of variables. The method in general cannot handle more than 15
19 variables, for this study the restriction was limited to 8. Thirdly, different results (faces)
20 can be produced according to the changes of the association of facial features and
21 statistical variables. Ki (2016) also used the method to show Korean metropolitan
22 governments' local community health index with 8 variables. With regard to the use of
23 the method, Ki (2016, p.357) claims that 'Chernoff face can be a better research tool or
24 marketing method for local governments and elections, (but) it has some demerits such
25 as lack of clear standard for matching the facial elements and statistical variables'. The
26 literature review reveals that the method has advantages and disadvantages and that
27 since there has been a growing interest in big data analysis, the method is recently re-
28 acknowledged by many researchers as a very useful tool for data visualization of big
29 data analysis. This is because as data are getting bigger, data visualization as the way of
30 presenting the result of big data analysis is considered very important. Therefore, it
31 seems that the method is likely to be used by many researchers in various areas.

32 One of the interesting things of using the method for presenting the results of data
33 analysis of library operations and services is that we can explore changes in the library
34 operations and services over time. As people's faces change as time goes on, so the
35 Chernoff face of the public library in a certain area also will be changed. This means that
36 the differences of Chernoff faces of libraries in London and Seoul would be seen if the
37 same study as this is carried out every five or ten years. Therefore, this study can be
38 viewed as the baseline for future comparison. The literature review reveals that there is
39 no research using the method in library-related research. It seems that this study is the
40 first to use the method for big data analysis of public library operations and services in
41 library and information research.
42
43
44
45
46
47
48
49
50
51
52

Comment [u2]: Because the reviewer says that "The role of the faces remains unclear, and this is the weakest point of the paper; it needs to be expanded upon, so that the reader can be convinced that this is a genuinely useful technique", the new sentence was added.

Comment [u3]: Because it was suggested that "It might be better to admit at the outset that this method can handle big data in the sense of many instances, but only on a relatively small number of variable", thus the new sentence is added.

Methodology

The study aims to explore a new method of evaluating public library operations and services. Chernoff faces, drawn by using big data and representing the public library operations and services of each local government in London and Seoul, are analysed and compared with each other.

<Table 1> Summary of the two cities and their public library operations and uses
 *registered users, **including Seoul metropolitan library and 4 private libraries
 total library visits including study area uses, *active borrowers only

In terms of research methods, there were four main work procedures in conducting the study. Firstly, collecting big data as empirical data. This involved three different stages,

	London (Mar. 2014)	Seoul (Dec. 2014)
Population	8,416,700	10,103,200 (4,507,995*)
Number of local governments	33	25
Number of public libraries	366	132**
Total stock (items)	12,599,124	11,990,254
(per 1,000 people)	(1,497)	(1,187)
Total library purpose visits	49,154,112	46,143,554
(per 1,000 people)	(5,840)	(4,567) (71,043,565***)
Total borrowers	1,348,256****	3,559,517
(per 1,000 people)	(160)	(352)
Total issues	34,021,358	26,599,198
(per 1,000 people)	(4,042)	(2,633)

e.g. selection of research sites, deciding types of data to be gathered and actual data gathering work. Secondly, associating human facial features and variables to draw the Chernoff face. Thirdly, formatting collected data based on the association of human facial features and variables and drawing the Chernoff face by using the R programming language. Fourthly, analysing and examining the operations and services of the public library in London and Seoul by using the Chernoff face. The detailed procedures are described in <Table 1>.

The first step of carrying out the study is gathering raw data, which is essential to draw the Chernoff face representing library operations and services of each local

1
2
3
4
5
6
7
8 government in London and Seoul. This study is designed to compare two cities to
9 examine which city performs better in library operations and services. London and Seoul
10 are selected for research sites because these are the capital city of each country and they
11 might best represent the operations and services of public libraries in the whole country.
12 As seen in <Table 1>, there are more libraries and stock and more people actively use
13 library services in London than Seoul.
14

15
16 The study was designed such that raw data collection was limited to eight elements
17 relating to public library operations, services and uses. The eight elements are library
18 floor space, (total and professional) staff, collections, budgets, library number, materials
19 issues, library visits. Statistical data of libraries as raw data is normally gathered by a
20 library authority or by a local government.
21

22
23 In order to evaluate the library operations and services of the two cities, eight variables
24 were selected based on eight elements for the following reasons. In general, three main
25 elements required for operating a library are a building (and facilities), staff and
26 collections. The study selects variables of library floor space and number of libraries for
27 the element of a library building and variables of number of total staff and professional
28 staff for the element of staff. Another very important element for operating a library is
29 collections. Thus, collections are considered as another variable. Library budgets are one
30 of the five core management resources, thus, these are also defined as a variable. If the
31 above three main elements and budgets are well prepared and managed, then many
32 people will be satisfied with visiting the library, using library facilities and borrowing
33 library materials. Library visits and issues of library materials are selected as two other
34 variables because these are very often considered as vital performance indicators of
35 library operations and services. Although there is some argument whether this method is
36 appropriate or not, measuring number of visitors and issues seems to be one of the
37 widely used methods for performance evaluation of library services. McMenemy (2009,
38 p.147), claims that 'measuring number of visitors to a library, and number of issues, gives
39 an indication of how widespread the use of the services is, but it does not give any
40 sense of what patrons do with the services the public library provides.'
41

42
43 According to the list of performance indicators suggested by the International
44 Federation of Library Associations (IFLA, 2000), the variables used for the evaluation of
45 library services in this study overlap with the list in five areas, e.g. number of issues,
46 library visits, book stock and staff and amounts of budgets. It is reported that two
47 variables used in this study, e.g. number and square meters of libraries are also used in
48 the USA for measuring library services (McMenemy, 2009). The literature confirms that
49
50
51
52
53
54

1
2
3
4
5
6
7
8 the selection of variables for the evaluation of library services in this study is reasonable.

9 There are two main reasons why only eight variables were selected although Chernoff
10 face-making functions in R program enable the display of a maximum of 15 variables.
11 Firstly, if too many variables are displayed in one Chernoff face then, this may cause
12 confusion in understanding of how the Chernoff face represents variables. Secondly, as
13 this study compares London and Seoul, comparable variables are limited to major
14 elements of library operations and services. This is because, since there are differences in
15 the culture, history and policy of public libraries in Korea and the UK, data produced on
16 library operations and services are slightly different. For example, since March 2012 data
17 on library floor space of each library authority have not been produced any more in the
18 UK, whereas it is still produced in Seoul and as seen in <Table 1>, data on 'registered
19 users' and 'total library visits including study area use' are gathered in Seoul.
20

21
22
23
24 Raw data excluding the data on the number of professional staff and libraries were
25 edited for fair comparison between the two cities. Since the population of the two cities
26 is different, in order to ensure a fair comparison the data was recalculated into 'per 1,000
27 people'. The data on the number of professional staff and libraries were not edited
28 because the figure is not high. Data on library budgets were slightly differently edited in
29 order to make a clear comparison between the two countries. The data on the budget of
30 Korean libraries are presented as Korean won and of British libraries are presented as UK
31 pound. Thus, Korean won and UK pound were converted into US dollars. The whole
32 process of actual data gathering is described in detail in the '**Data collection**' section.
33
34
35

36 The second step of the study was to match variables (raw data) to face features. It is
37 very important in the research method of this study to match human facial features, e.g.
38 eyes, nose, mouth, etc. to each variable of a multivariate dataset. Yau (2007, p.2) claims
39 that 'unfortunately, the 'faces() function' (of a computer program 'R') doesn't let us
40 choose what face part to associate with each metric, so we need to find a workaround'.
41 Lee et al. (2013, p.974) also mention that 'one of the difficult problems using a Chernoff
42 face method is that depending on how face parts and variables are associated with each
43 other, a different result can be produced'.
44
45

46 In order to redeem the disadvantage of using the method and to help people more
47 easily understand what part of a Chernoff face is associated with each variable, the
48 researcher surveyed 20 Korean people as to which part of the face they would like to
49 display for four elements of a library operation such as number of libraries, staff and
50 collections, and budgets. The literature review on the Chernoff face method reveals that
51 no survey had been carried out before and that researchers designed association of
52
53
54

facial features and variables by themselves. This study is the first attempt to survey people's preferences of association of face part and variables. However, since the first author of this study who was mainly in charge of collecting data and surveying people is a Korean, he had some difficulties of carrying out such survey in the UK. This is because he planned to adopt a Korean mobile instant messaging application of smart phones for the survey, but he was only able to use the application with his own network of Korean librarians. This is a recognised limitation of the current study.

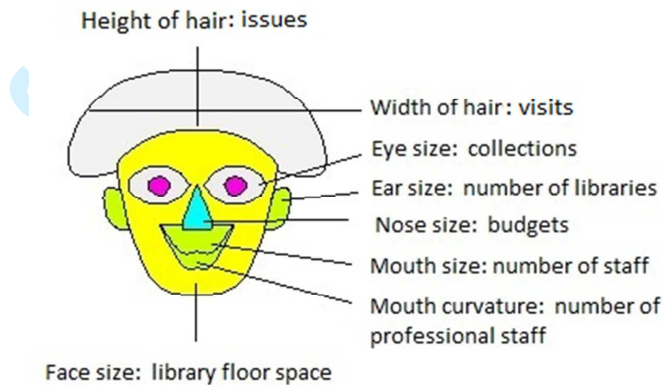
<Table 2> Participants' preference of association of face parts and variables

	Eyes	Nose	Mouth	Ears
Collections	<u>15</u>	1	2	2
Budgets	1	<u>8</u>	7	4
Staff	2	6	<u>8</u>	4
Number of libraries	2	5	3	<u>10</u>

The 20 people who are mostly librarians answered the question. As seen in <Table 2>, 15 among 20 participants prefer that eyes represent collections. This is because they considered that, when users visit libraries, collections can be seen by the users' eyes. They view that library budgets are very important for library operations. Budgets should be represented by a nose, which exists in the centre of a face. Eight out of 20 participants claim that a nose represents library budgets. They like a mouth to represent library staff. This is because a member of staff is a human being who can speak to library users. They suggest that ears can represent the number of libraries. This is because a library is the institution which can hear people's voice or their demands. Based on people's opinion, the researcher finally designed the association of facial features and variables as <Figure 1>.

As Chernoff (1973) explains, with Chernoff faces each data point 'is represented by a cartoon of a face whose features, such as length of nose and curvature of mouth, correspond to components of the point. Thus, every multivariate observation is visualized as a computer-drawn face. This presentation makes it easy for the human mind to grasp many of the essential regularities and irregularities present in the data.' Information and data are represented in the shape of a human face as seen in <Figure 1>. Thus, the individual parts, such as eyes, nose, mouth and ears represent values of the variables by their shape, size, placement and orientation.

Comment [u4]: Because there was the reviewer's comment "why was the test of public views carried out only in Korea; and why was no test done with different assignments to see if it made any difference?", thus, the new sentence was added.



<Figure 1> Association of facial features and eight variables in the Chernoff Face

<Table 3> Association of facial features and eight variables

1. Height of Face: **library floor space(m²) per 1,000 people***
2. Width of Face: library floor space(m²) per 1,000 people
3. Shape of Face: not allocated
4. Height of Mouth: **no. of staff per 1,000 people**
5. Width of Mouth: no. of staff per 1,000 people
6. Shape of Mouth: **no. of professional staff**
7. Height of Eyes: **no. of collection per 1,000 people**
8. Width of Eyes: no. of collection per 1,000 people
9. Height of Hair: **no of materials issued per 1,000 people**
10. Width of Hair: **no. of library visits per 1,000 people**
11. Shape of Hair: not allocated
12. Height of Nose: **library budget(\$) per 1,000 people**
13. Width of Nose: library budget(\$) per 1,000 people
14. Height of Ear: **no. of libraries**
15. Width of Ear: no. of libraries

* **Bold letters** are 8 variables

The third step of the study is formatting the data gathered. The data were edited and formatted by using the Microsoft Excel program as a CSV (comma separated values) file for analysis based on the association of facial features and eight variables as <Table 3>. Formatted data were processed by using the R programming language and resulted in the drawing of Chernoff faces. Initially 33 Chernoff faces and 25 Chernoff faces

1
2
3
4
5
6
7
8 representing each local government in London and Seoul were drawn for analysis of
9 library operations and services.

10 11 **Data collection**

12 In this study, data collection means two things: one is gathering raw data and the other
13 is drawing the Chernoff face by using the R programming language with the data
14 gathered for analysis of library operations and services. There might be argument as to
15 whether data like the raw data gathered in this study such as statistics on library
16 operations, services and uses are big data or not. Big data is defined as 'a massive
17 volume of both structured and unstructured data that is too large to be processed with
18 traditional database and software techniques' (Beal, 2016, n.p.). However, Arthur (2013,
19 n.p.) claims that 'big data is a collection of data from traditional and digital sources
20 inside and outside the company that represent a source for ongoing discovery and
21 analysis. ... Some people like to constrain big data to digital inputs like web behavior and
22 social network interactions; however, the COMs and ICOs ... agree that we can't exclude
23 traditional data derived from product transaction information, financial records and
24 interaction channels...'

25
26
27
28
29
30 This study needed to gather big data on the operations and uses of public libraries in
31 London and Seoul. As seen in <Table 1>, there are 366 public libraries and these are
32 operated by 33 local governments in London and 127 libraries are operated by 25 local
33 governments in Seoul. The problem of collecting big data on library operations, services
34 and uses is that in general, the data cannot be collected through the Internet as big data
35 in other areas normally does. The data on public libraries can only be collected when
36 library authorities or local governments release the data. In particular, this study aiming
37 to analyse and compare the data of 470 public libraries in 58 local governments in two
38 countries has huge difficulty in terms of gathering data. However, luckily Korea and the
39 UK release detailed statistical data on public libraries which were initially gathered by
40 each library, each library authority and finally by one organization in each country.

41
42
43
44
45 In the study, raw data on public library services in London were gathered through
46 CIPFA (2015) and the same data on Seoul were gathered through the Korean
47 government web site (NLSSK, 2015). IMTA & SCT have been collecting and releasing
48 statistical data on the public libraries in the UK since 1963. Each library authority in
49 London gathers the data and the authority provides the data to CIPFA. Raw data on
50 public libraries in Seoul were gathered through the Korean government website. Yau
51 (2011, p.xv) mentions that "a significant shift in release of government data came in mid-
52
53
54

1
2
3
4
5
6
7
8 2009, with the United States' launch of Data.gov. ... Soon after, many countries and cities
9 began to release government data". As Yau claims, the Korean government releases the
10 statistical data of the public library on the government website. These data were initially
11 gathered by each library and the library provided the data to the government institution.
12

13 Using these data and the R programming language, it was possible to draw a Chernoff
14 face for each local government of the two cities representing their government. London
15 consists of 33 boroughs (including the City of London) and Seoul consists of 25 local
16 governments. This means that 33 and 25 Chernoff faces were drawn respectively for
17 London and Seoul. Each of the 58 Chernoff faces represents one local government's
18 library operations, services and uses. The Chernoff faces of London and Seoul were
19 compared to examine which city and local government performs better in library
20 operations and services; to discover how and what library services are different; and
21 whether there are any similarities.
22
23
24
25

26 Findings

27 On the whole, as seen in <Figure 2> and <Figure 3>, the face of S-Jongno which is the
28 largest is well recognizable(S represents Seoul, and Jongno is the name of one of the 25
29 local governments in Seoul). S-Jongno, just like the City of London, is located in the
30 centre of Seoul and its population is low in comparison with other local government
31 areas. S-Jongno is the second after S-Jungu in population among 25 local governments
32 in Seoul. However, there is the biggest public library (13,266m²) in Seoul in that area and
33 the library's operations and services are very good: the library has excellent collections
34 and enough reading and study space. Thus, many people from other local government
35 areas come and use and borrow library materials. As a result, the value of this library's
36 variables is very high and this indicator affected the value of the variables of the whole
37 library operations and services in the S-Jongno area. Finally, this resulted in a good
38 shape for the S-Jongno face.
39
40
41
42
43

44 Face size represents floor space of the library. On the whole, the size of the faces of
45 the two cities looks almost the same except for a few faces, e.g. S-Jongno, L-Cof Lon (L
46 and 'Cof Lon' represents respectively London and City of London). However, in fact the
47 real size of the public library in the two cities is completely different. As the average size
48 of libraries in Seoul is 2,104m² (NLSSK, 2015) and of London is 594m² (CIPFA, 2015), the
49 size of the faces of Seoul should be bigger than the faces of London. However, face size
50 is calculated per 1,000 people. That is to say, in general, there are more people
51 (population) who should be served by the library in Seoul than the library in London.
52
53
54

1
2
3
4
5
6
7
8 Thus, the size of the faces of London and Seoul looks almost the same.

9 Eyes represent library collections. On the whole, except for a few faces, e.g. S-Jongno,
10 S-Jonggu, S-Yongsan, the eyes of most of the faces in Seoul look small. In contrast, the
11 eyes of the faces in London generally look bigger than those of Seoul. This means that
12 the libraries in London have more extensive collections than the libraries in Seoul.

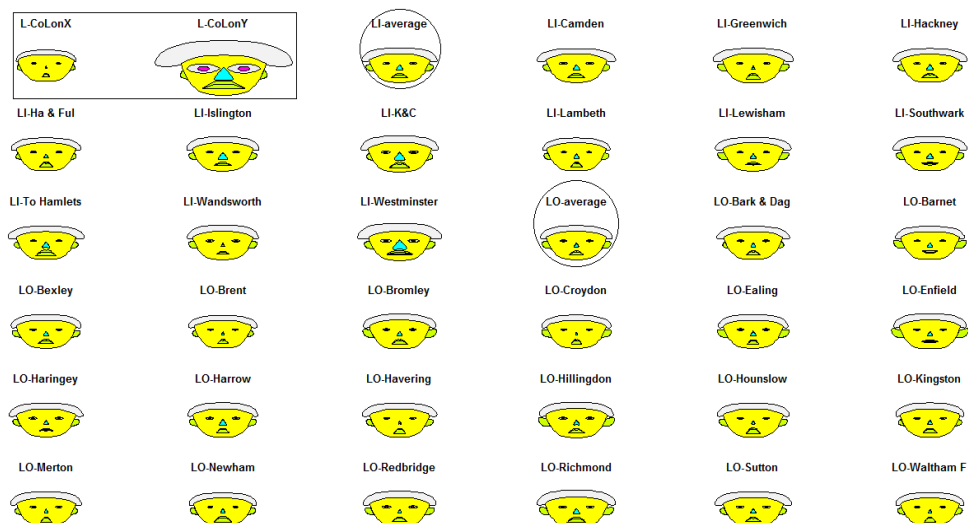
13 A nose represents budgets. The noses of London faces are more prominent than those
14 of Seoul. This means that in general, libraries in London have higher budgets than
15 libraries in Seoul.

16 As a mouth represents staff, the faces of London have a more defined mouth than the
17 faces of Seoul. This means that libraries in London have more library staff than libraries
18 in Seoul. There are only a few smiling faces among the 58 faces. A smiling face means
19 that there is a high proportion of professional staff among the total staff. The result of
20 the study reveals that only a few library authorities have many professional staff. The
21 most well defined smiling faces are S-Jongno and S-Yongsan.

22 Ears represent the number of public libraries a local government operates. The ears of
23 the London faces are differentiated from those of Seoul. The ears of the London faces
24 look bigger than those of Seoul. This means that there are more libraries in London than
25 Seoul. On average, each local government operates 11.1 libraries in London but only 5.1
26 libraries in Seoul.

27 The hair of the Chernoff face represents two variables. The width of hair displays the
28 number of 'library purpose visits' and the height of hair shows the number of issues of
29 library materials. Excepting several faces, e.g. S-Geumcheon, S-Jongno, S-Seongdong, S-
30 Yongsan, the width of hair of the faces in Seoul is narrow. The study reveals that the hair
31 of the faces in London is slightly wider than that of Seoul. This means that more people
32 in London visit the library than do the people in Seoul. As seen in <Table 1>, 1,000
33 people in London visit the library 5,840 times but 1,000 people in Seoul visit 4,567 times
34 per year. The height of the hair of the Chernoff face represents the number of issues.
35 Although not many faces of London have thick hair, in general, the hair of London faces
36 looks higher than the hair height of Seoul. This means that in general, more library
37 materials (4,042 items per 1,000 people) are borrowed in London than in Seoul (2,633
38 items per 1,000 people).
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



<Figure 2> Chernoff faces of public library operations and Uses of London

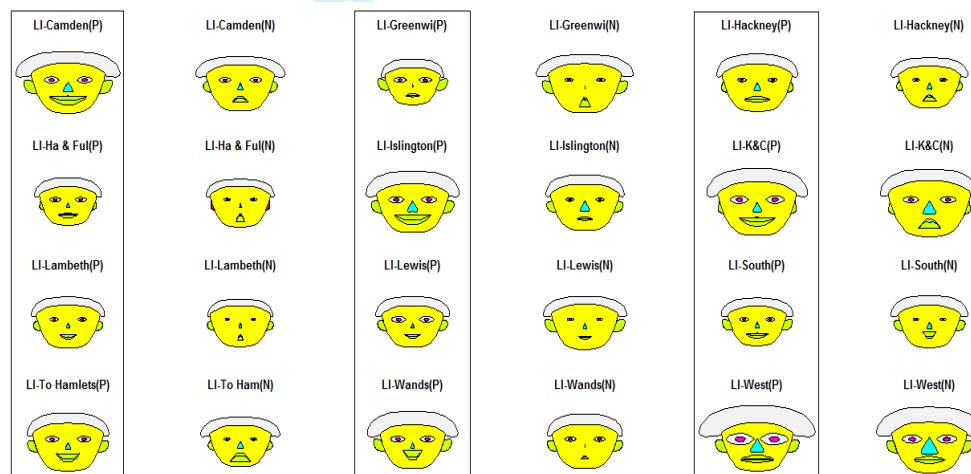


<Figure 3> Chernoff faces of public library operations and Uses of Seoul

Since the City of London is in an awkward location geographically, the City has a very small registered population. However, the library in the City serves as many people as other councils' libraries do. Because of this complex situation, two faces are drawn as seen in <Figure 2>. One (L-CoLonX), which is the smallest face among 58 faces, is drawn with the number of potential users (around 392,000 people) who are employed in the City of London. The other face (L-CoLonY), which is the largest, is drawn based on the

number of registered population (around 7,600 people). This face is even bigger than the face of S-Jongno.

The study discovers that one of the notable advantages of using the Chernoff face for visualizing research data is the capacity for comparison. For example, two Chernoff faces that each represents the past and the present situation can be compared. As seen in <Figure 4>, the study compares the operations and uses of public libraries in 12 Inner London boroughs over a period of 10 years (2004-2014).

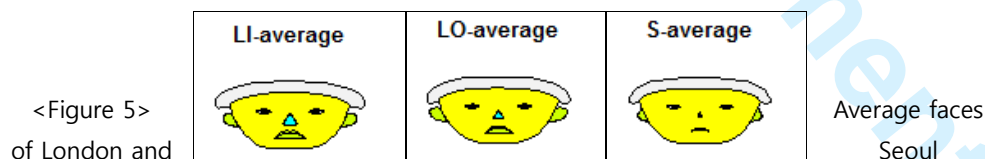


<Figure 4>Chernoff faces of past and present public library operations and uses of Inner London boroughs * (P): Past (2004), (N): Now (2014)

The study reveals that in general Chernoff faces representing the data of the past situation are slightly bigger than the faces representing the data of the present situation. This means that the size of public libraries has lessened during 2004-2014. Some differences between the two groups of faces identified are that, except in a few boroughs, in general, eye size is diminished and as seen in Wandsworth, Lambeth, Hammersmith & Fulham, the nose is less apparent. It can also be seen that there is a change of hair. The height of hair has lowered in the last ten years. This means that fewer people borrow library materials than was the case 10 years earlier. The most important finding of the study is surely the fact that the 2004 faces are smiling (representing a high proportion of professional staff) whereas by the 2014 faces are no longer smiling. This means that there are fewer professional library staff than 10 years ago.

1
2
3
4
5
6
7
8 If a standard face is available, then this face can be used as a tool for absolute
9 comparison. However, since there is no standard for library operations and services, a
10 standard face cannot be drawn. Instead of a standard face, an average face was drawn
11 and this face can be used as a tool for absolute comparison. As seen in <Figure 5>,
12 three average faces were drawn in order to reveal, firstly, which local government in each
13 city performs better than other governments in terms of library operations and uses and
14 secondly, on the whole, which city between London and Seoul performs better. In the
15 case of London, except the City of London, two average faces that each represents
16 respectively inner boroughs and outer boroughs were drawn.

17
18
19
20 The results of the study reveal that on the whole, apart from mouth curvature, every
21 face feature of London is bigger than the Seoul face: the average face of London has a
22 bigger (narrow long) eyes, nose, mouth and ears. This means that there are more
23 libraries in each local government area in London than Seoul and that London boroughs
24 have more library collections, budgets and (total) staff than local governments in Seoul.
25 The width and height of the hair of London face is slightly wider and higher than that of
26 Seoul. This means that more people in London use library services (visiting libraries and
27 borrowing library materials) than the people in Seoul. However, the average face of Seoul
28 has a slightly more curved (smiling) mouth than that of London. This means that there
29 are more professional library staff in Seoul than London.
30
31
32
33
34
35
36



In comparison between inner and outer boroughs of London on the whole, LI (London Inner Borough) has bigger (narrow long) eyes and nose than LO (London Outer Borough). This means that LI councils allocate more library collections and budgets than LO. The mouth of LI face is slightly bigger and the width of the hair of LI is slightly wider than that of LO. This means that there are slightly more library staff and slightly more people visit the library in LI than LO. In contrast, LO has bigger ears than LI. This means that LO has more libraries than LI. In terms of the height of the hair, LI and LO are almost the same. This means that almost the same number of people in both boroughs borrow

1
2
3
4
5
6
7
8 library materials.

10 Discussion

11 The results of the evaluation of the public library operations and services of the two
12 cities using the Chernoff face method reveal that on the whole, local governments in
13 London perform better than local governments in Seoul with regard to the provision and
14 operations of public libraries. The local governments in London operate more libraries,
15 invest more library budgets, allocate more staff and hold more collections than the local
16 governments in Seoul. This administration of public libraries resulted in more borrowing
17 of library materials and more visits to libraries than Seoul.

18 In terms of library size, the two cities operate almost the same size of libraries. With
19 regard to professional library staff, it was discovered that there are more professional
20 library staff in libraries in Seoul than London. This fact should be seriously considered in
21 terms of the future of public library services in London, because the role of professional
22 library staff is very important in maintaining professional standards of library operations
23 and services (IFLA, 2001).

24 It can be claimed from the results of the study that the overall performance of public
25 libraries in London is better than Seoul. However, the results of the analyses of the
26 changes of library operations and uses in inner London boroughs for the last ten years
27 (2004-2014) reveal that the library operations and services have diminished and library
28 uses have been decreasing. Therefore, it could be anticipated that the leading
29 performance of library operations and services, and uses in London will be overtaken by
30 Seoul. This is because the library operations and services have been shrinking in the UK,
31 whereas these have gradually been developing in Korea during 2004-2014. The latest
32 available data support this claim. According to the Taking Part study (Department for
33 Culture, Media and Sport, 2016), the number of British adults visiting libraries fell by 14.3%
34 between March 2006 (48.2%) and September 2015 (33.9%). The data from CIPFA also
35 reveal that the number of active borrowers fell by 23.5% between 2010/2011 and
36 2014/2015 (The Guardian, 2016).

37 Regarding the advantages and disadvantages of using the Chernoff face method in the
38 evaluation of library operations and services, the study discovered the following findings.
39 Firstly, the method can be very useful for use in the comparison of the performance of
40 library operations and services over a period of time. For example, as per this research, a
41 study on the examination of the changes of library operations and services for ten or
42 twenty years by using Chernoff faces will be very useful. Secondly, as described in the

1
2
3
4
5
6
7
8 **"Findings"** section, if a standard face is made available, then the Chernoff face could be
9 very useful to analyse the performance of library operations and services. That is to say,
10 the standard face can act as a scale for measuring the performance of library operations
11 and services.
12

13 It is considered that the Chernoff face method is not useful on the following occasions.
14 Firstly, if there are too many variables, this will make a Chernoff face difficult to
15 understand. The 'faces()' function' of the R programming language enables only a
16 maximum of 15 variables: this study used 8 variables, Yau (2011) used 7 and 8 variables
17 respectively for crime research and an analysis of NBA basketball players, Lee et al (2013)
18 used 11 variables for bank research and Ki (2016) used 8 variables for social research.
19 The results of this study suggest around 10 variables as a reasonable number of variables
20
21

22 Secondly, if there are too many faces, it will be difficult to compare and distinguish
23 faces. The current study has 58 faces and this number could be considered to be too
24 many. Yau (2011) drew 50 and 52 faces respectively for an analysis of NBA basketball
25 players and crime research; Yang (2012) drew only 8 faces; Lee et al. drew 15 faces; and
26 Ki (2014) drew 26 and 48 faces respectively for local government research and political
27 research. The study suggests a maximum of around 30~40 faces as a reasonable number
28 of faces.
29
30

31 Thirdly, it can be difficult to use the Chernoff face method in the comparison of
32 countries. This is because it is not easy to gather the same kinds of data from many
33 countries. The method is more suited for use in the comparison of library authorities and
34 other public sectors in one country or of libraries in one local or metropolitan
35 government.
36
37

38 The method is more commonly used for quantitative research than qualitative research.
39 In future research using the method, in order to overcome the limitation of quantitative
40 research, the method could be used alongside other qualitative research. Another way of
41 improving the limitation of the quantitative research is that data of people's satisfaction
42 on library services can be presented by using the method. As IFLA (2000) describes, data
43 on user satisfaction is a qualitative indicator for measurement of library services.
44
45
46

47 **Conclusion**

48 This is the first study to use the Chernoff face method for big data analysis of public
49 library operations and services of two cities in two countries. The study reveals that on
50 the whole, at the moment local governments in London perform better than Seoul: more
51 people in London use library services than Seoul. This finding suggests that Seoul still
52
53
54

1
2
3
4
5
6
7
8 needs more libraries, collections, budgets and staff for providing better library services
9 and to increase library use.

10 However, the results of this study including the analysis of changes of library
11 operations and uses in inner London boroughs, and other available data and reports
12 (Department for Culture, Media and Sport, 2016; The Guardian, 2016), demonstrate that
13 in the last five to ten years the use of library services has gradually decreased in the UK.
14 This consequence seems to be due to the fact that many libraries were closed, or are
15 operated by community groups, and opening hours and numbers of professional library
16 staff were reduced. If this situation lasts for some time, the level of library services and
17 uses in the UK might fall behind some OECD countries like Korea, in which many local
18 governments are keen to improve public library operations and services e.g. building
19 new libraries and increasing opening hours and new services, etc.

20 'As social media has become widespread, emoticons have played a significant role in
21 communication through technology' (Wikipedia, 2016). Looking like emoticons, the
22 Chernoff face which represents the library operations and services in a local government,
23 might attract people's interest in library services provided by their local government.
24 Thus, a Chernoff face will be effective in communicating with ordinary people about how
25 their local government operates public libraries and how other people use library
26 services.
27
28
29
30
31
32
33
34
35

36 · Note for data collection and related.

- 37 1. Raw data on public libraries in London is mainly collected from CIPFA's public
38 library statistics and the same data in Seoul is mainly collected from the National
39 Library Statistic System of Korea.
- 40 2. Most of the raw data collected on public libraries in London were produced in
41 March 2014 and in Seoul produced in December 2014. However, some raw data
42 on public libraries in London were produced between March 2013 and March
43 2011.
- 44 3. Since March 2011, most of the London boroughs do not provide statistics on
45 library floor space. Thus, raw data on library floor space in London boroughs was
46 collected in the way that the average floor space (682.6m² and 550m²) of the
47 public libraries in inner and outer London boroughs is respectively applied to
48 each borough. For example, Camden in inner London has 9 libraries, thus the
49 total floor space of the libraries is 6,143m². Barking & Dagenham in outer London
50
51
52
53
54

- 1
2
3
4
5
6
7
8 has 6 libraries, thus the total floor space of the libraries is 3,300m².
- 9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
4. Population statistics for London and Seoul are collected respectively by March 2014 and by December 2014.
 5. "Camden (Borough) have made the decision not to recognize professional qualifications", (CIPFA, 2015) so the number of professional staff is revealed as zero. However, in practice there will be some professional staff. There were 18.2 professional staff until March 2011. Therefore, the number of professional staff in Camden is calculated in the way that the average rate (15.8%) of the number of the professional staff in inner London boroughs is applied to Camden. The total staff number of Camden is 98. Thus, the number of professional staff in Camden is assumed as 14.1 staff.
 6. There is a Seoul metropolitan library which serves the whole Seoul area. This study does not include the data of this library.
 7. To ensure a fair comparison between London and Seoul:
 - mobile libraries are excluded from the number of total libraries.
 - part-time staff are excluded from the number of staff because their total working hours are unknown in Seoul.
 - library budgets presented respectively as British Pound and Korean Won were converted into US Dollar by using XE's currency converter system in March 2016.
 - E-books are excluded from collections because the same E-books are used by Library users in many local governments in Seoul. Thus, E-books and E-audios are excluded from the collections in both London and Seoul.
 - issues of E-books and E-audios are excluded from the total number of issues.
 - visits only for library purposes are considered in the study. Visits for business and only study area use are excluded from the total number of visits.
 8. IMTA & SCT. The Institute of Municipal Treasurers and Accountants later changed their name into the Chartered Institute of Public Finance and Accountancy (CIPFA) began to publish 'Public Library Statistics' together with the Society of County Treasurers from 1963. Since 1978, CIPFA alone has been publishing 'Public Library Statistics'.

References

- Arthur, L. (2013), "What is Big Data?", available at:
<http://www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-ata/#a1ff2d348723>
(accessed 15 March 2016).

- 1
2
3
4
5
6
7
8 Beal, V. (2016), "Big data", available at: http://www.webopedia.com/TERM/B/big_data.html
9 (accessed 28 January 2016).
- 10 Chen, M., Floridi, L., and Borgo, R. (2014) "What is visualization really for?" In: Floridi,
11 L. and Phyllis, I. ed. *The Philosophy of Information Quality, Synthese Library*, Vol.
12 358, Springer International Publishing Switzerland, pp. 75–93.
- 13 Chernoff, H. (1973), "The Use of Faces to Represent in K-Dimensional Space Graphically",
14 *Journal of the American Statistical Association*, Vol.68 No.342, pp. 361-368.
- 15 CIPFA. (2015), *Public Library Statistics 2014-15 Estimates and 2013-14 Actuals*,
16 CIPFA, London.
- 17 Department for Culture, Media & Sport. (2016), "Taking Part focus on: libraries", available
18 at: [https://www.gov.uk/government/uploads/system/uploads/attachment_data/
19 file/519675/Libraries_short_story_-_FINAL.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/519675/Libraries_short_story_-_FINAL.pdf) (accessed 31 May 2016).
- 20 The Guardian. (2016), "Library use falling sharply, study shows", available at:
21 [https://www.theguardian.com/books/2016/may/03/library-use-falling-sharply-
22 study-shows](https://www.theguardian.com/books/2016/may/03/library-use-falling-sharply-study-shows) (accessed 31 May 2016).
- 23 Golden, L. L. and Sirdesai, M. (1992), "Chernoff Faces: a Useful Technique for Comparative
24 Image Analysis and Representation", *Advances in Consumer Research*, Vol.19, pp.
25 123-128.
- 26 Goulding, A. (2006), *Public Libraries in the 21st Century: Defining Services and Debating
27 the Future*, Ashgate, Aldershot.
- 28 IFLA. (2000), "Revision of IFLA's Guidelines for Public Libraries", available at:
29 <http://archive.ifla.org/VII/s8/proj/gpl.htm> (accessed 28 September 2015).
- 30 IFLA. (2001), *The public library service: IFLA/UNESCO guidelines for development*, (IFLA
31 *Publications 97*), Saue, Munchen.
- 32 JLA (Japan Library Association). (2015), "Japan Library Statistics", available at:
33 <http://www.jla.or.jp/library/statistics/tabid/94/Default.aspx> (accessed 5 June 2016).
- 34 Ki, J. (2016), "A Big Data Analysis of Urban Statistics Expression -Chernoff Face- Based
35 Expression of Local Community Health Index in Korea", *Space and Society*, Vol.26
36 No.1, pp. 336-358.
- 37 Kim, Y. (2004), "Public Libraries and Friends of the Library Groups: the influence of Friends
38 groups on public library policy", Ph.D. Thesis. The University of Sheffield.
- 39 Kim, Y. (2011), "A Study on the Comparison of the Operation of Public Libraries Among
40 Local Governments in Seoul", *Journal of Library and Information Science Society*,
41 Vol.42 No.4, pp. 311-330.
- 42 Ko, Y. M., Shim, W., Pyo, S. H., Chang, J. S. and Chung, H. K. (2012), "An economic
43
44
45
46
47
48
49
50
51
52
53
54

- 1
2
3
4
5
6
7
8 valuation study of public libraries in Korea", *Library & Information Science*
9 *Research*, Vol.34 No.2, pp. 117-124.
- 10 Lee, J., Jeong, H., Kim, M., Kim, J. and Son, Y. S. (2013), "Good Bank Evaluation by
11 Chernoff Face Analysis using SAS macro faces", *The Korean Journal of Applied*
12 *Statistics*, Vol.26 No.6, pp. 959-975.
- 13
14 Lott, J. A. and Durbridge, T. C. (1990), "Use of Chernoff faces to follow trends in
15 laboratory data", *Journal of Clinical Laboratory*, Vol.4 No.1, pp. 59-63, available
16 at: <http://www.ncbi.nlm.nih.gov/pubmed/2107290> (accessed 28 September 2015).
- 17
18 McMenemy, D. (2009), *The Public Library*, Facet Publishing, London.
- 19
20 The Ministry of the Interior. (2016), "Population statistics", available at:
21 http://rcps.egov.go.kr:8081/jsp/stat/ppl_stat_jf.jsp (accessed 5 March 2016).
- 22
23 Morris, C. J., Ebert, D. S. and Rheingans, P. (2000), "An Experimental Analysis
24 of the Effectiveness of Features in Chernoff Faces", available at:
25 http://www.research.ibm.com/people/c/cjmorris/publications/Chernoff_990402.pdf
26 (accessed 28 September 2015).
- 27
28 NLSSK (National Library Statistics System of Korea). (2015), "Library Statistics", available at:
29 <http://www.libsta.go.kr/index.do> (accessed 28 September 2015).
- 30
31 Nel, D., Pitt, L. and Webb, T. (1994), "Using Chernoff faces to portray service quality data",
32 *Journal of Marketing Management*, Vol.10, pp. 247-255.
- 33
34 R Core Team (2016), "R: A language and environment for statistical computing. R
35 Foundation for Statistical Computing, Vienna, Austria", available at:
36 <https://www.R-project.org/> (accessed 28 September 2015).
- 37
38 Rahu, M. (1989), "Graphical representation of cancer incidence data: Chernoff faces",
39 *International Journal of Epidemiology*, Vol.18 No.4, pp. 763-767, available at:
40 <http://www.ncbi.nlm.nih.gov/pubmed/2621011#> (accessed 28 September 2015).
- 41
42 The World Trusted Currency Authority. (2016), "XE Currency Converter", available at:
43 <http://www.xe.com/> (accessed 28 September 2015).
- 44
45 Yau, N. (2007), "How to visualize data with cartoonish faces ala Chernoff", available at:
46 <http://flowingdata.com/2010/08/31/how-to-visualize-data-with-cartoonish-faces/>
47 (accessed 29 September 2015).
- 48
49 Yau, N. (2011), *Visualize this: the flowing data guide to design, visualization, and*
50 *statistics*, Wiley Publishing, Indianapolis.
- 51
52 Yang, W. (2012), "Looking at the face of Korean baseball teams in 2012", available at:
53 <http://wsyang.com/2012/05/2012-professional-baseball-korea/> (accessed 29 March
54 2016).

1
2
3
4
5
6
7
8 Wikipedia. (2016), "Emoticon", available at: <https://en.wikipedia.org/wiki/Emoticon>
9 (accessed 6 June 2016).
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Big Data Analysis of Public Library Operations and Services by Using the Chernoff Face Method

- Tables & Figures

<Table 1> Summary of the two cities and their public library operations and uses

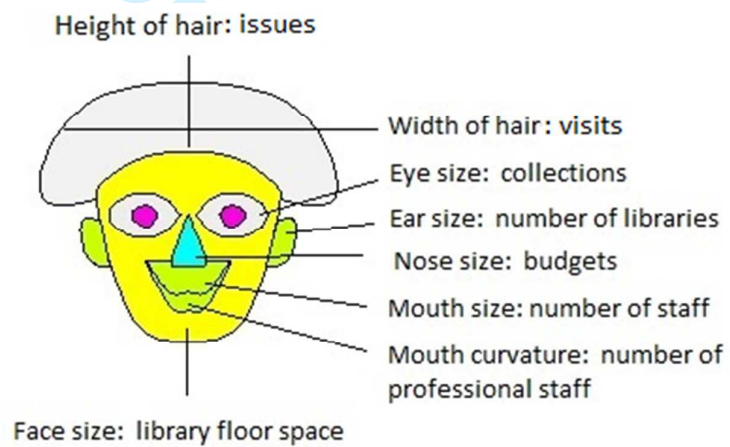
*registered users, **including Seoul metropolitan library and 4 private libraries

total library visits including study area uses, *active borrowers only

	London(Mar. 2014)	Seoul(Dec. 2014)
Population	8,416,700	10,103,200 (4,507,995*)
Number of local governments	33	25
Number of public libraries	366	132**
Total stock (items)	12,599,124	11,990,254
(per 1,000 people)	(1,497)	(1,187)
Total library purpose visits	49,154,112	46,143,554
(per 1,000 people)	(5,840)	(4,567) (71,043,565***)
Total borrowers	1,348,256****	3,559,517
(per 1,000 people)	(160)	(352)
Total issues	34,021,358	26,599,198
(per 1,000 people)	(4,042)	(2,633)

<Table 2> Participants' preference of association of face parts and variables

	Eyes	Nose	Mouth	Ears
Collections	<u>15</u>	1	2	2
Budgets	1	<u>8</u>	7	4
Staff	2	6	<u>8</u>	4
Number of libraries	2	5	3	<u>10</u>

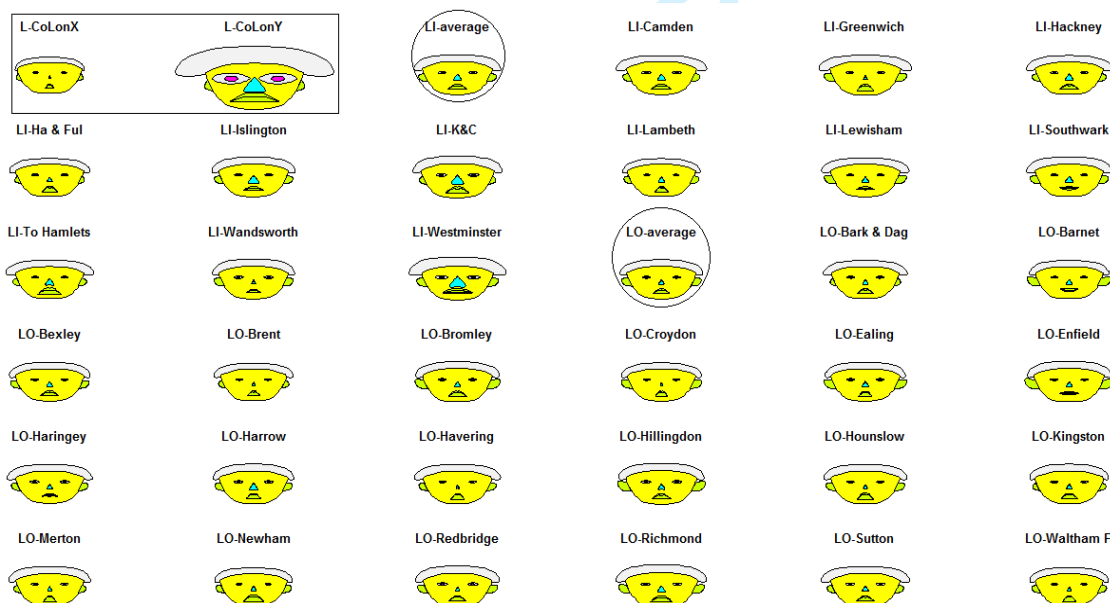


<Figure 1> Association of facial features and eight variables in the Chernoff Face

<Table 3> Association of facial features and eight variables

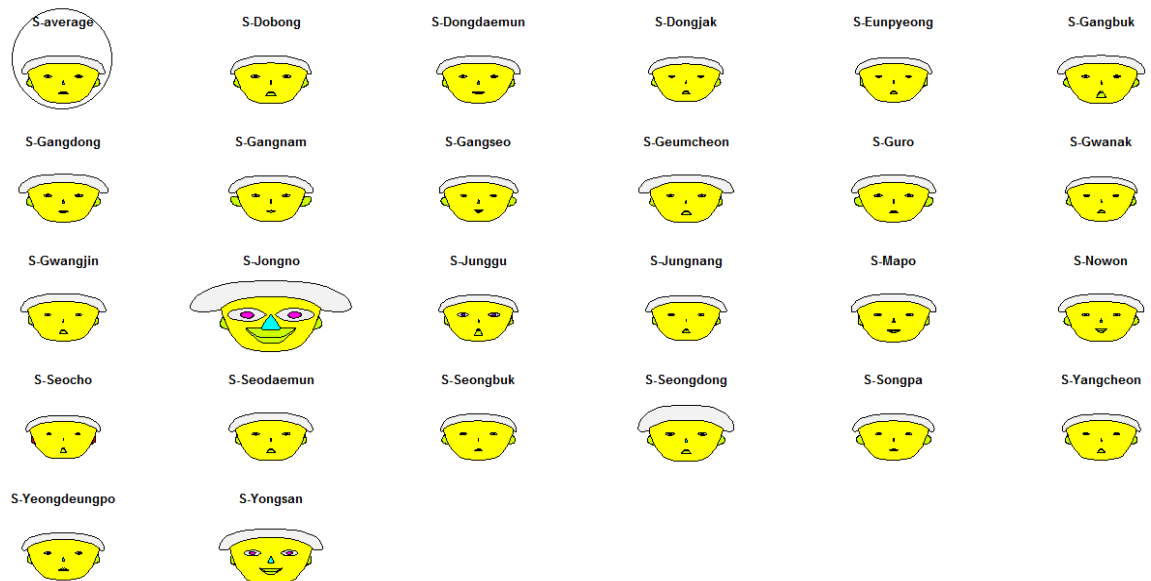
1. Height of Face: **library floor space(m²) per 1,000 people***
2. Width of Face: library floor space(m²) per 1,000 people
3. Shape of Face: not allocated
4. Height of Mouth: **no. of staff per 1,000 people**
5. Width of Mouth: no. of staff per 1,000 people
6. Shape of Mouth: **no. of professional staff**
7. Height of Eyes: **no. of collection per 1,000 people**
8. Width of Eyes: no. of collection per 1,000 people
9. Height of Hair: **no of materials issued per 1,000 people**
10. Width of Hair: **no. of library visits per 1,000 people**
11. Shape of Hair: not allocated
12. Height of Nose: **library budget(\$)** per 1,000 people
13. Width of Nose: library budget(\$) per 1,000 people
14. Height of Ear: **no. of libraries**
15. Width of Ear: no. of libraries

* **Bold letters** are 8 variables

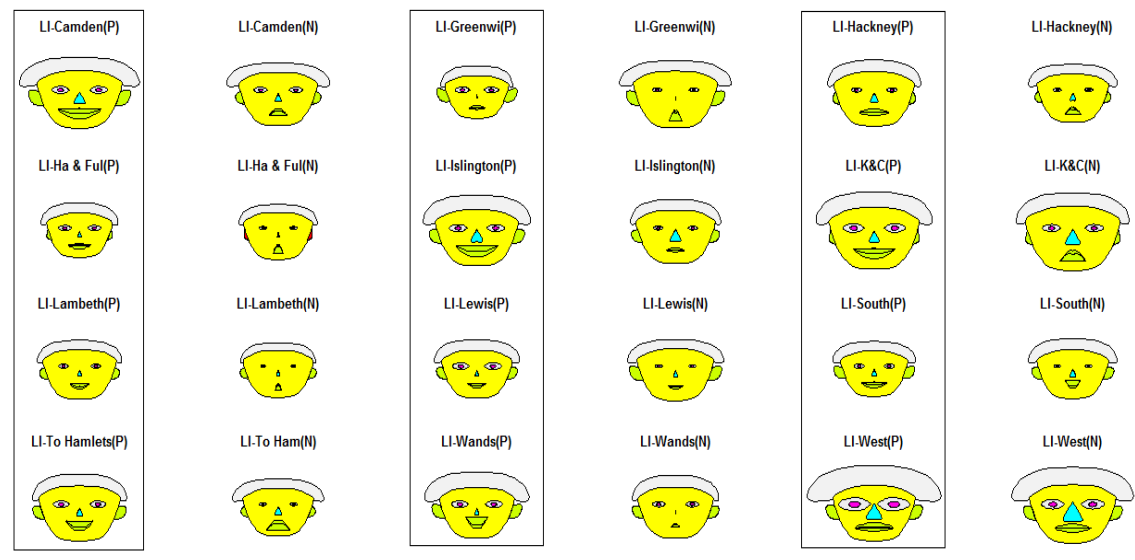


<Figure 2> Chernoff faces of public library operations and Uses of London

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



<Figure 3> Chernoff faces of public library operations and Uses of Seoul



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

<Figure 4> Chernoff faces of past and present public library operations and uses of inner London boroughs * (P): Past (2004), (N): Now (2014)

<Figure 5> Average faces of London and Seoul

