Big-data Analysis on Energy Consumption of Office Buildings in Seoul, Korea

Ki Uhn Ahn[1], Han Sol Shin[1], Cheol Soo Park[1], Kwang Woo Kim[2]
[1]School of Civil, Architectural Engineering and Landscape Architecture, Sungkyunkwan University, Suwon, South Korea
[2]Department of Architecture and Architectural Eng., College of Eng., Seoul National University, Seoul, South Korea

Abstract

It is important to understand the characteristics of actual energy consumption of existing buildings, e.g. energy use intensity (EUI, kwh/m2.yr) vs. building age. This paper reports a big-data analysis on the actual energy consumption of 4,625 office buildings in Seoul, South Korea. The data, a total of 226,625 data points, were obtained from the national building energy database released by Korean government in 2015 and include the following information of each individual building: location, age, usage, total floor area, the number of floors, the number of elevators, and monthly energy consumption. In order to deliver an intuitive understanding of the thermal characteristics on the large building stock, this paper presents a big-data analysis as follows: (1) energy consumption of buildings vs. their location, (2) energy consumption of buildings vs. building age, and (3) correlations between energy consumption of buildings and other building data. The analyses were visualized with scatterplot matrix and density plot. In the paper, it is shown that building energy consumption, represented as EUI (kwh.m2.yr), was irrelevant to the building age as well as strict prescriptive building energy codes.

Introduction

The energy consumed by existing buildings accounts for more than 30% of global energy use (Wu et al. 2016). With the prospect that the building sector is and will continue to be a major energy consumer in the years ahead (Liu et al. 2015), the replacement rate of existing buildings by the new buildings is only about 1.0-3.0% per annum (Wu et al. 2016). Therefore, there has been awareness that retrofitting existing buildings is important to reduce the building energy consumption.

During past several decades, a number of energy saving policies and certifications have been developed to improve energy efficiency in existing buildings: LEED and Energy Star in U.S.; BREEAM in U.K.; Energy Performance Certificate in E.U.; HK-BEAM in H.K. The International Energy Agency (IEA) has also conducted several projects to improve energy efficiency of existing buildings: Annex 46, Annex 55, and Annex 56. In addition, a number of studies have been conducted to develop technical measures, e.g. simulation for high-performance windows or insulations, model predictive controls, measurement and verification of energy efficiency, etc. in order to reduce and optimize energy consumption of existing buildings.

In 2015, the government of South Korea released the national building energy dataset including monthly energy consumption and other building data to the public. This paper aims to investigate the relevance of actual monthly/yearly energy consumption of office buildings (4,625 buildings) to location, building age, use, total floor area, and the number of floors and elevators. A total of data points used in this study is 226,625.

Building energy database & data-mining

Big-data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization (Beyer and Laney, 2012). There has been millions of databases used in business management, government administration, scientific and engineering academia, and may other applications (Chen et al., 1996). In building domain, building automation systems (BASs) and building energy management and control systems (EMCSs) can be a kind of big data since they can store a huge amount of building energy data in real-time (Zhun et al., 2016).

A data mining, which is also referred to as knowledge discovery in database (KDD), means a process of nontrivial extraction of implicitly, previously unknown and potentially useful information (such as knowledge, constraints, regularities) from database (Fayyad et al., 1996). The data mining studies in Building Simulation (BS) domain have been focused on prediction and/or pattern extraction of energy consumption and occupant behavior in a specific target building. Recently, large-scale dataset collected from national building stock have been highlighted since it can provide statistics of the building energy information. Several datasets including DOE Building Performance Database (BPD), Commercial Building Energy Consumption Survey (CBECS), Residential Energy Consumption Survey (RECS), and California Energy End-Use Survey (CEUS) have been maintained by federal or state government.

Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases and visualization to address the issue of information extraction from large databases (Cabena et al. 1997). Several kinds of techniques can be used for data mining tasks such as classification, clustering, association rules, regression, generation and summarization, pattern based similarities (Han et al. 2012). It is worth noting that visual data exploration, which is to present the data in a visual form, allows the human to get intuitive insight or understanding in the data (Keim 2002). Therefore, the visualization for the big-data of existing buildings can be an effective way to capture the relationships between energy consumption and the relevant building information.

The Korean government released the a database (https://open.eais.go.kr, http://open.greentogether.go.kr/) which include information on building's address, monthly electric and/or gas energy consumption, building name, building usage, total floor area, height, and the number of households, the number of floors, the number of elevators, the number of parking lots, and building age. In this study, the 4,625 office buildings in the database located in 25 different districts, Seoul were used.

Question #1: Do old buildings consume more energy than new buildings? The answer is no.

Figure 1 shows the location of buildings in the database, their year built, yearly electric energy consumption per unit area, yearly gas energy consumption per unit area, and yearly primary energy consumption per unit area

Green dots in (Figure 1 (a)) indicate the buildings built in 1970 or even earlier and are concentrated in the district named Junggu. The electric, gas and primary energy consumption of buildings do not seem correlated to building age (Figure 1 (a) vs. Figure 1 (b), (c), (d)).

As shown in Table 1, the average of year built in Junggu district is 1986, which is 12 years older than the average year built (1998) of the entire buildings located in Seoul. According to Figure 1 and Table 1, it is difficult to identify a notable relationship between the building age and building energy consumption. For example, the 50th quantiles of year built of the buildings in the Junggu

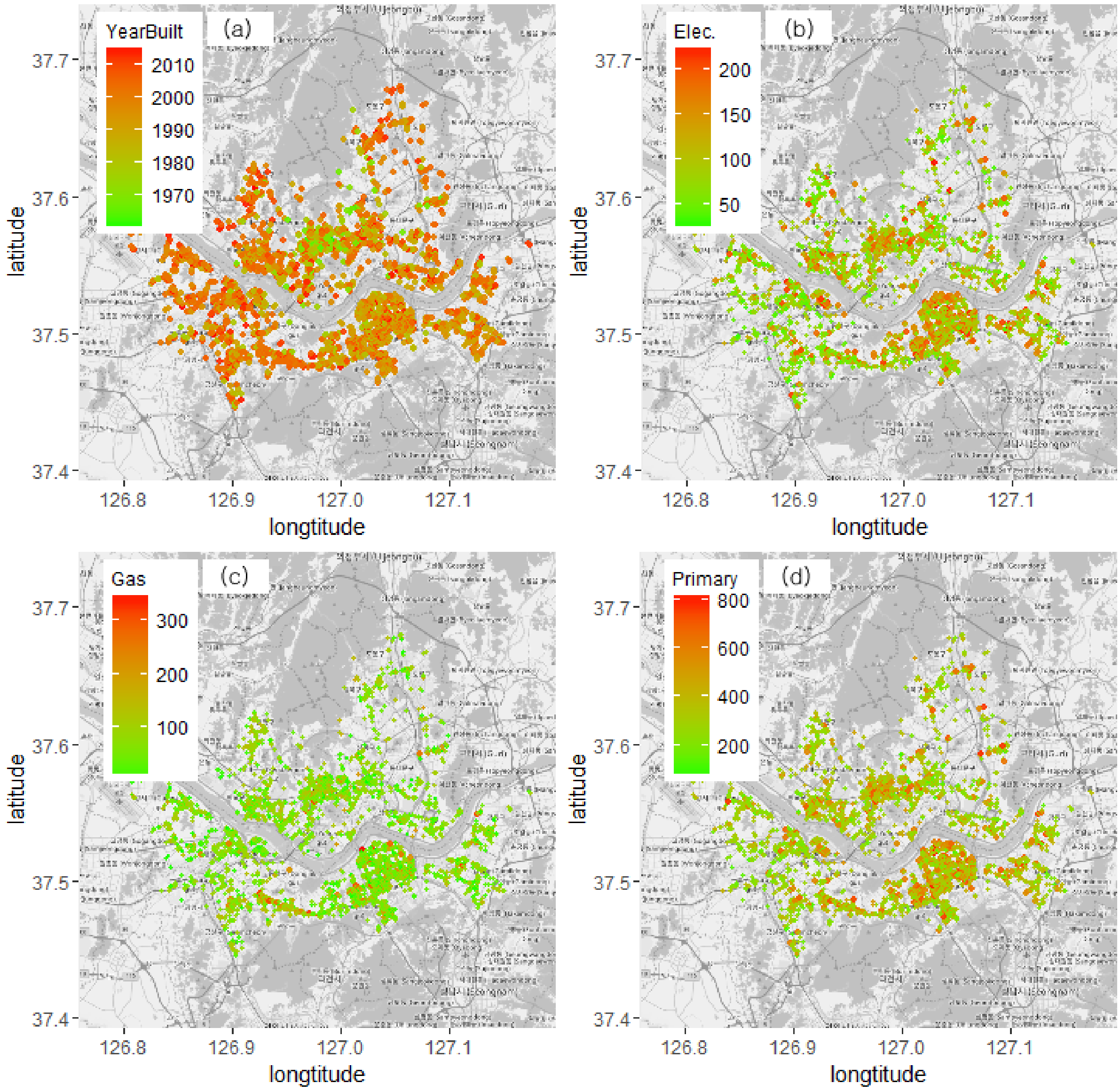


Figure 1: building energy consumption and year built: (a) year built, (b) electric energy consumption (kWh/m²·year), (c) gas energy consumption (kWh/m²·year), (d) primary energy consumption (kWh/m²·year)

Table 1: Statistic results of primary energy consumption and year built

Name of district	Year built					Yearly primary (elec. + gas) energy consumption [kWh/m2·year]				
	quantile			avg.	std.	quantile			avg.	std.
	25th	50th	75th			25th	50th	75th		
Dobong	1996	2003	2005	2001	7	242	278	335	295	90
Dongdaemun	1987	1994	2004	1996	11	206	277	342	287	107
Dongjak	1990	1997	2003	1997	10	264	325	458	368	136
Eunpyeong	2002	2003	2004	2002	8	232	276	327	297	99
Gangbook	1990	1997	2003	1996	10	266	328	377	333	106
Gangdong	1992	1998	2004	1998	8	251	302	387	338	135
Gangseo	1997	2002	2004	2001	7	226	274	325	288	91
Gangnam	**1991**	**1995**	**2004**	**1997**	**9**	**291**	**361**	**445**	**375**	**123**
Geumcheon	1996	2004	2007	2001	10	219	274	361	291	110
Guro	1991	2003	2008	2000	10	218	270	340	292	107
Gwanak	2002	2003	2004	2002	5	261	297	360	320	94
Gwangjin	1992	2000	2004	1999	9	209	272	348	300	117
Jongro	1983	1991	2002	1991	12	280	331	424	352	111
Junggu	**1971**	**1986**	**1999**	**1986**	**15**	**286**	**364**	**447**	**374**	**136**
Jungnang	1996	2003	2004	2000	7	257	284	343	313	100
Mapo	1991	2002	2004	1998	9	254	307	386	330	113
Nowon	1990	2000	2004	1998	7	269	345	414	367	135
Seocho	1991	1994	2003	1996	8	271	340	417	356	118
Seodaemun	1992	2003	2005	1999	10	261	310	369	325	111
Seongbuk	1988	2000	2004	1996	13	245	317	369	326	118
Seongdong	1990	1998	2006	1998	11	225	289	386	316	115
Songpa	1991	1994	2003	1996	7	260	332	402	347	121
Yangcheon	1996	2002	2004	2001	7	206	265	305	271	96
Yeongdeungpo	1991	2001	2005	1998	10	232	285	372	315	115
Yongsan	1988	1994	2004	1995	11	247	298	395	330	114
Total (Seoul)	**1991**	**1998**	**2004**	**1997**	**10**	**257**	**318**	**400**	**340**	**119**

district and in the Gangnam district are 1,986 and 1,995 respectively, and yearly primary energy consumptions of the buildings in the Junggu and Gangnam districts are similar to each other (364 kWh/m^2 vs. 361 kWh/m^2) (Table 1).

Figure 2 shows the scatterplots visualizing the relationship between year built and energy consumption. As indicated in the values of R^2 (coefficient of determination) in Figure 2, the building age and the level of building energy consumption seem to be irrelevant to each other.

It is genrally agreeable that energy performance of new buildings (e.g. insulation, u-values of building envelopes, effcient cooling and heating systems/plants, controls) is better than that of old buildings. But, it is noteworthy that energy use intensity of new buildings is not generally less than that of old buildings. Even, the relevance of building age to EUI is almost negligible (the ranges of R^2 is from 0.00068 to 0.22). Further investigation is required with regard to this findings, e.g. taking several buildings and conducting more detailed studies, or sensitivy analysis of thermal performance of buildings with regard to the building energy use.

Question #2: Have strict prescriptive specification contributed to reduction of energy consumption for past decades? The answer is not necessarily.

Since the 1980s, a number of building energy regulations have been enacted in South Korea, including energy conservation design criteria, energy performance index (EPI) evaluation, and energy consumption calculation for Energy Performance Certificate (EPC) (Park et al., 2015). The energy conservation design criteria and EPI evaluation are a prescriptive method with regard to building envelopes and systems. The energy conservation design criteria have a total of nineteen mandatory rules in areas of architectural design (e.g. u-values for walls, roofs and windows, airtightness of building envelopes), mechanical system design (e.g. insulation for ducts and pipes, installation of certified energy efficient pumps/fans) and electrical system design (e.g. installation of efficient luminaires). The energy conservation design criteria encourage architectural designers and engineers to adopt energy-efficient technologies.

The EPI is calculated based on fifty evaluation items with respect to architectural, mechanical, electrical, and

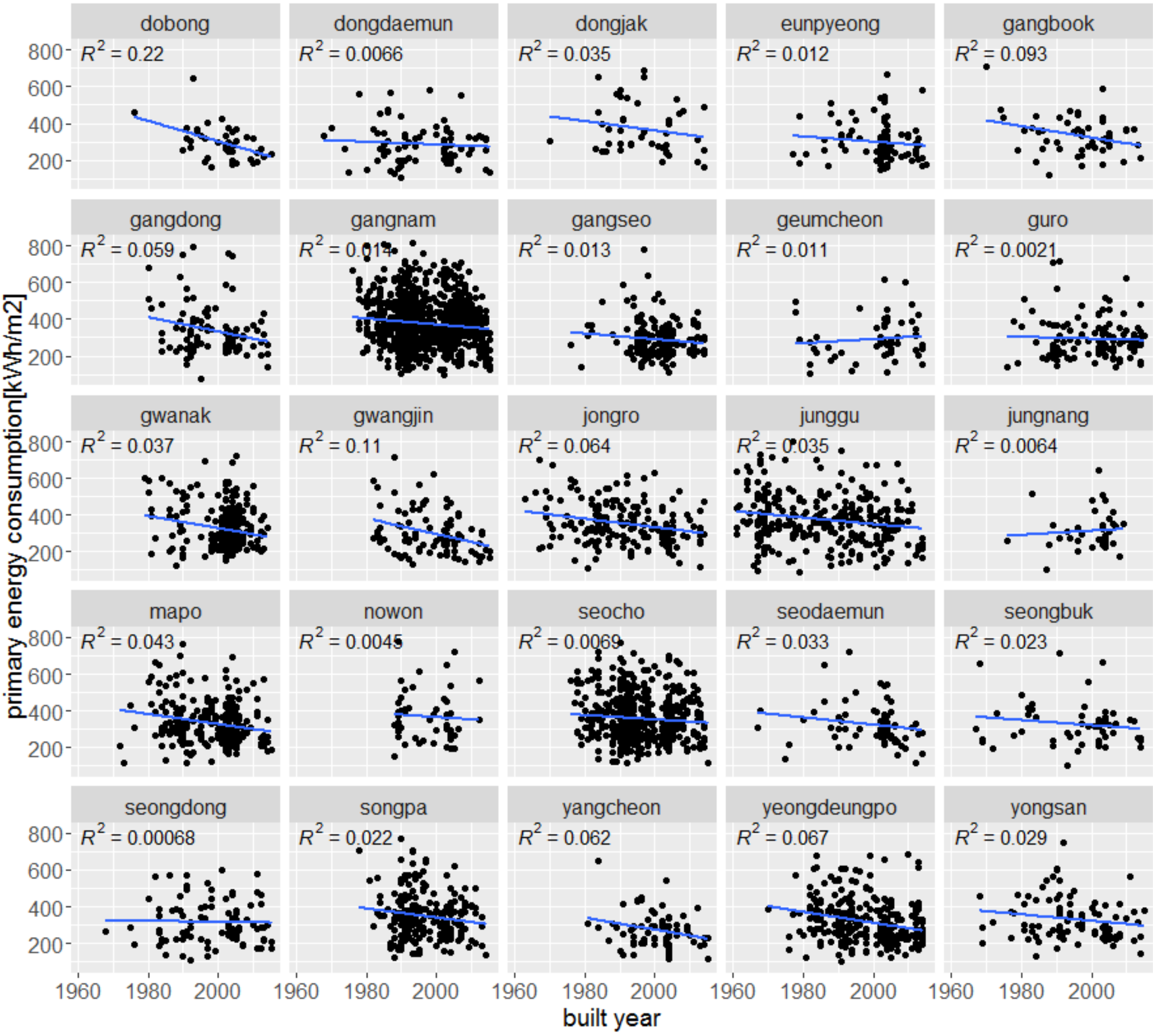


Figure 2: Scatterplots for year built and yearly energy consumption of the districts

renewable energy aspects. The EPI score, which is a sum of the credits obtained by each evaluation item, must be greater than a reference value designated by Korean government. The level of energy conservation design criteria as well as EPI have been raised periodically. Table 2 and Table 3 show the revisions of energy conservation design criteria for external walls and windows, respectively.

In the aforementioned revisions, there has been an implicit intention to save more energy with the use of stricter building energy codes (Tables 2-3). However, it is difficult to say that stricter prescriptive energy codes have contributed to reduction of building energy consumption, as shown in Figures 1-3 and Table 1.

Table 2: Revisions of energy conservation design criteria for u-value of external walls in South Korea

Year	Central zone	Southern zone	Jeju Island
1992	0.58 W/m²·K	0.76 W/m²·K	1.16 W/m²·K
2001	0.47 W/m²·K	0.58 W/m²·K	0.76 W/m²·K
2010	0.36 W/m²·K	0.45 W/m²·K	0.58 W/m²·K
2013	0.27 W/m²·K	0.34 W/m²·K	0.44 W/m²·K
2016	0.26 W/m²·K	0.32 W/m²·K	0.43 W/m²·K

Table 3: Revisions of energy conservation design criteria for u-value of windows in South Korea

Year	Central zone	Southern zone	Jeju Island
2001	3.84 W/m²·K	4.19 W/m²·K	5.23 W/m²·K
2008	3.40 W/m²·K	3.80 W/m²·K	4.40 W/m²·K
2010	2.40 W/m²·K	2.70 W/m²·K	3.40 W/m²·K
2013	2.10 W/m²·K	2.40 W/m²·K	3.00 W/m²·K
2016	1.50 W/m²·K	1.80 W/m²·K	2.40 W/m²·K

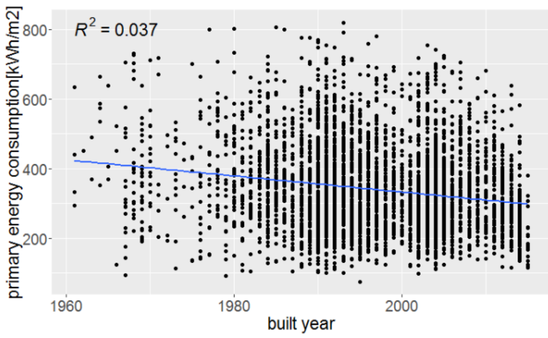


Figure 3: Scatterplots for year built and energy consumption of the Seoul

Question #3: By which factor among the data in the national building energy database is building energy consumption most influenced? It is hard to find.

The authors investigated the correlations between the building energy consumption and the following building information released from Korean national building energy database:

- District: 25 districts are located in Seoul and defined as 1 to 25
- Elec(Gas)_M1-Elec(Gas)_M12: monthly electric(gas) energy consumption per unit area in Jan. to Dec. 2015 [kWh/m^2]
- Elec(Gas)_Y: yearly electric(gas) energy consumption per unit area in 2015 [kWh/m^2]
- Primary_Y: yearly primary energy consumption per unit area in 2015 [kWh/m^2]
- YearBuilt: year built
- NumofFlr: the number of floors
- NumofElev: the number of elevator
- FloorArea: total floor area [m^2]

Figure 4 shows the scatter plots with regard to yearly primary energy consumption vs. the aforementioned building information. The yearly primary energy consumption was classified into four levels (0-250

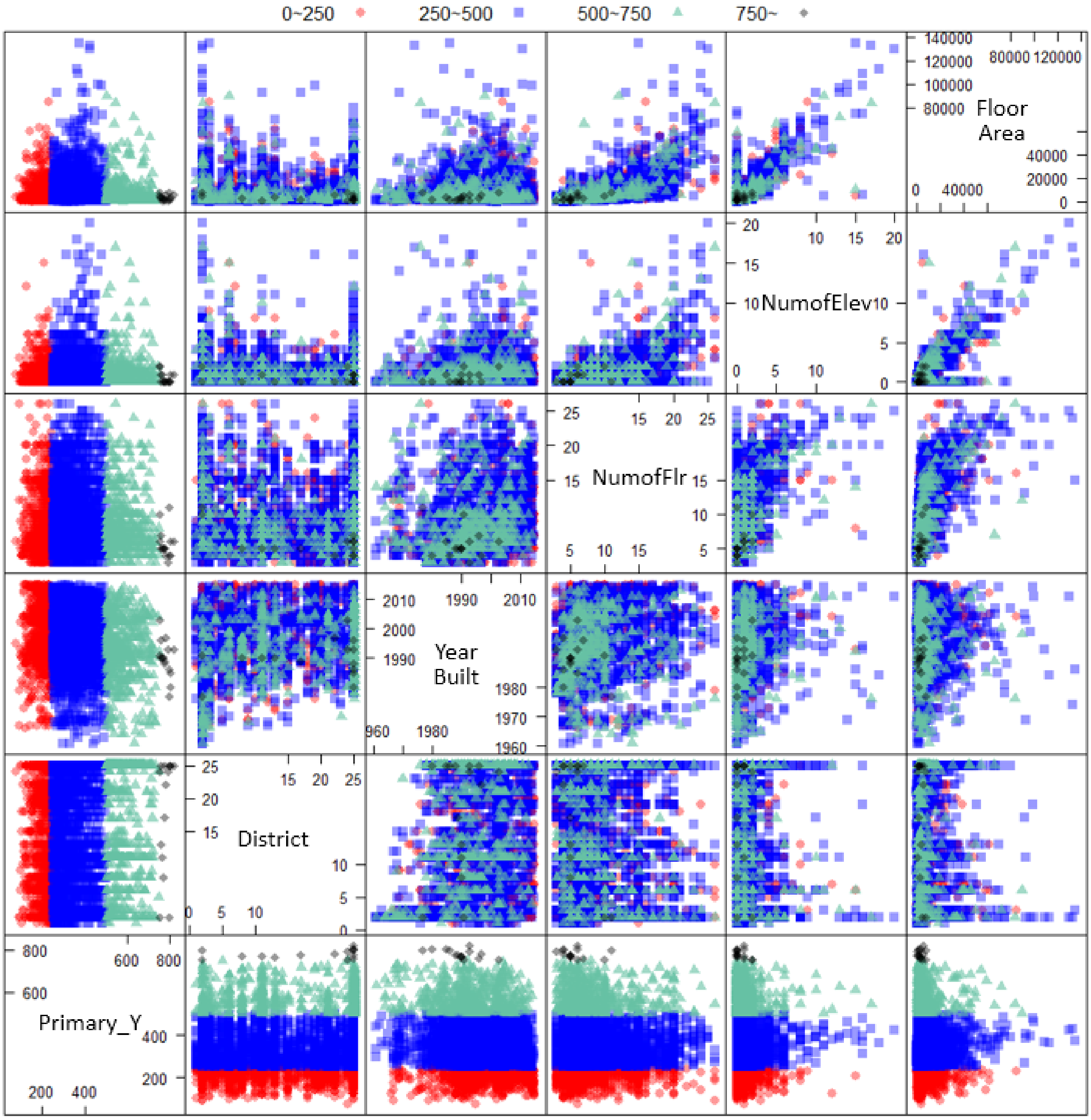


Figure 4: Scatterplot matrix for yearly primary energy consumption and building information (Primary_Y: yearly primary energy consumption per unit area in 2015 [kWh/m2], District: index of 25 districts, YearBuilt: year built, NumofFlr: the number of floors, NumofElev: the number of elevator, FloorArea: total floor area [m2])

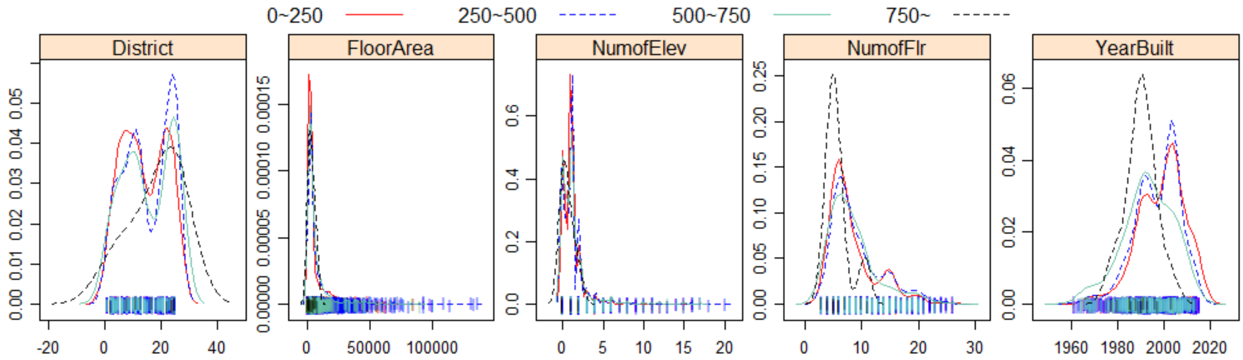


Figure 5: Probability density plots of primary energy consumption with respect to building information

[represented by red circle], 250-500[blue rectangular], 500-750[green triangle], greater than 750 kWh/m^2 [grey diamond]) and mapped to the points shown in the each plane of scatterplots. Four levels of energy consumption are uniformly scattered, to some extent, over in the scatterplot planes. In other words, it is difficult to find explicit relationships between energy consumption vs. location, year built, the number of floors, the number of elevators, and the floor area.

Figure 5 shows four probability density graphs in each plane, and each graph represents the probability density of the aforementioned building information with four energy consumption levels. Table 4 is the statistic results of each probability densities for the levels of energy consumption. As shown in Figure 5 and Table 4, the degree of energy consumption is not relevant to the building information.

Table 4: Statistic results of probability density plots

Building Information	Statistic	energy consumption levels [kWh/m^2]			
		0-250	250-500	500-750	750-
District	Avg.	13.9	15.1	15.1	18.2
	Std.	7.6	8.1	8.6	8.7
Total Floor Area [m^2]	Avg.	5,857	7,116	6,950	3,412
	Std.	9,003	11,605	11,070	2,440
The num. of Elev.	Avg.	1.0	1.2	1.2	0.5
	Std.	1.3	1.7	1.9	0.7
The num. of floors	Avg.	8.3	9.1	8.4	5.7
	Std.	4.2	4.4	4.2	2.2
Year Built	Avg.	1998	1997	1992	1989
	Std.	9.9	9.6	10.7	6.5

Figure 6 is the correlation plot matrix where the relationship between two attributes, represented with a scale from -1.0 to 1.0. The blue color scaled to 1.0 means positive correlation, while the red color scaled to -1.0 means negative correlation. The yearly primary energy consumption is correlated to the monthly/yearly electric energy consumption (blue color), but the other attributes including monthly/yearly gas energy consumption have a week correlation with the yearly primary energy consumption. The monthly and yearly gas energy consumption are correlated with each other except the cooling seasons (from July to September).

The probability densities for each energy consumption are shown in Figure 7. The amount of energy was normalized from 0.0 to 1.0 on X-axis. Monthly probability densities for electric and primary energy consumption are similar to each other over 12 months. However, the probabilities of monthly gas energy consumption seasonably vary. In addition, the electric and primary energy consumption can be regarded as a normal distribution, while the distributions of monthly gas energy consumption are severely right-skewed and heavy-tailed. Therefore, the correlation between electric and primary energy consumption is stronger rather than that between gas and primary energy consumption. Also, the monthly gas energy consumption during winter months (Jan, Feb. Mar vs. Nov. Dec.) are cross-correlated (Figures 6 & 7).

Discussion & Conclusion

This paper presented a big data analysis for Korean national building energy database. For this study, the energy database for 4,625 office buildings located in Seoul were used. This study addresses the noteworthy findings as follows: (1) building age is not relevant to the energy consumption level (kwh/m2.year), (2) stricter prescriptive building energy codes have not contributed to reduction of the energy consumption level and (3) there is no explicit relationship between energy consumption vs. location, year built, the number of floors, the number of elevators, and the floor area.

The detailed reasons for the aforementioned three findings will be investigated as a further study. For example, the integration of public database with the stored data from the BASs (e.g. type of HVAC and plant system, and their control records, occupancy, occupant behavior) and EMCs (e.g. set-point temperature for indoor air and plant systems, energy saving strategy records, retrofit and/or commissioning records) of selected buildings would be used for further big-data analysis.

Acknowledgement

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20152020105550).

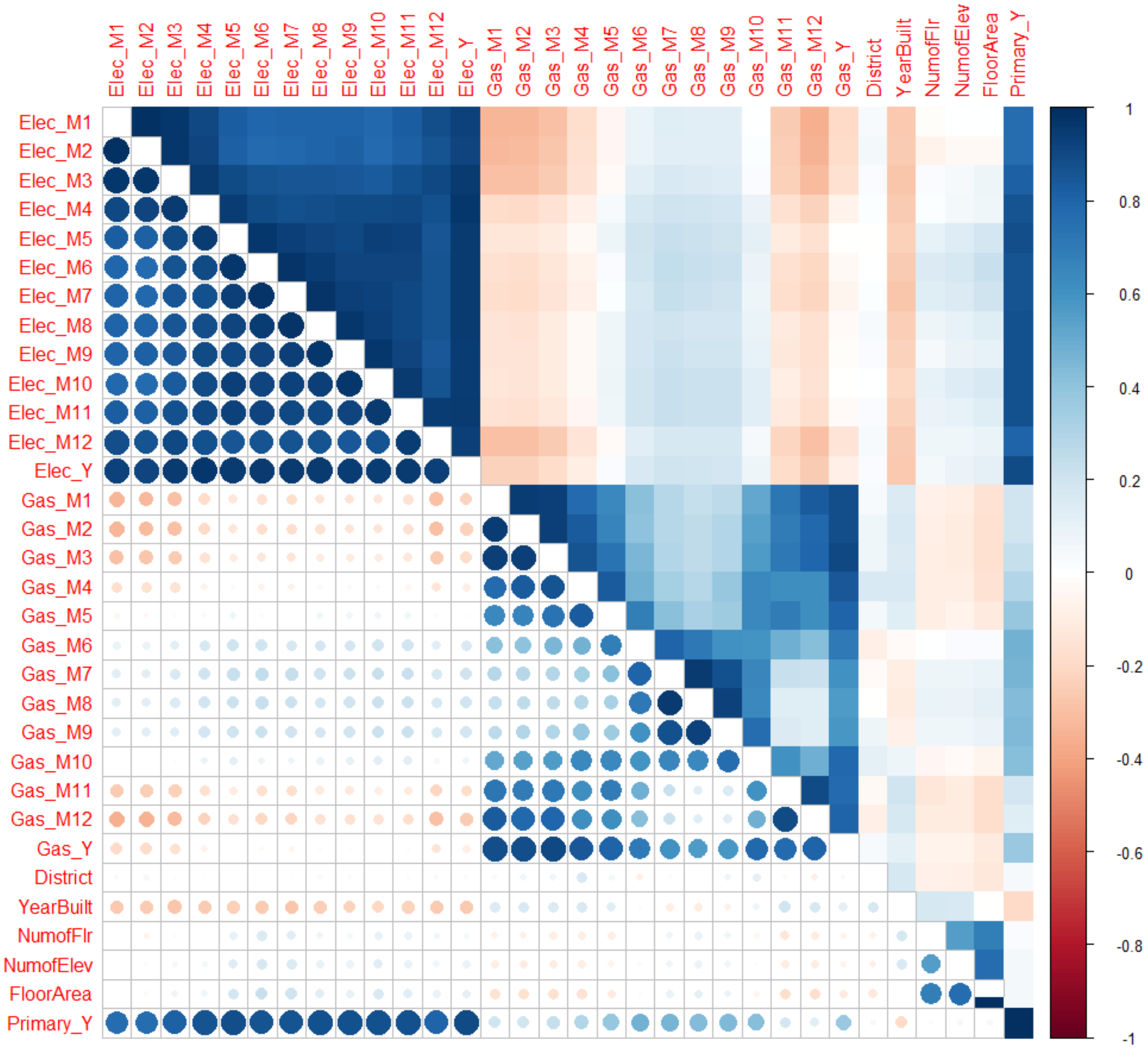
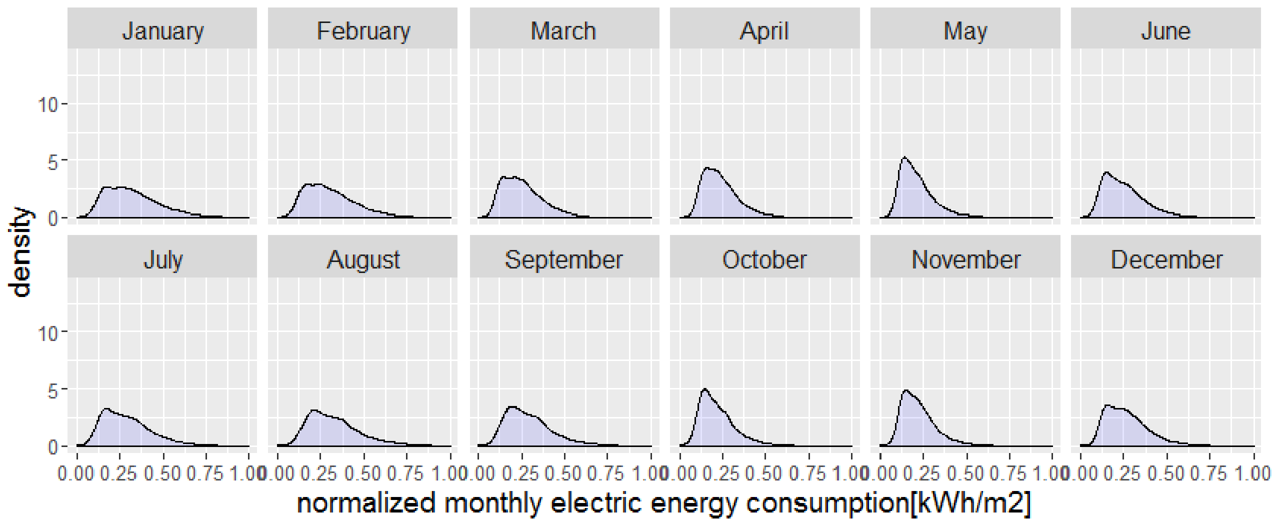
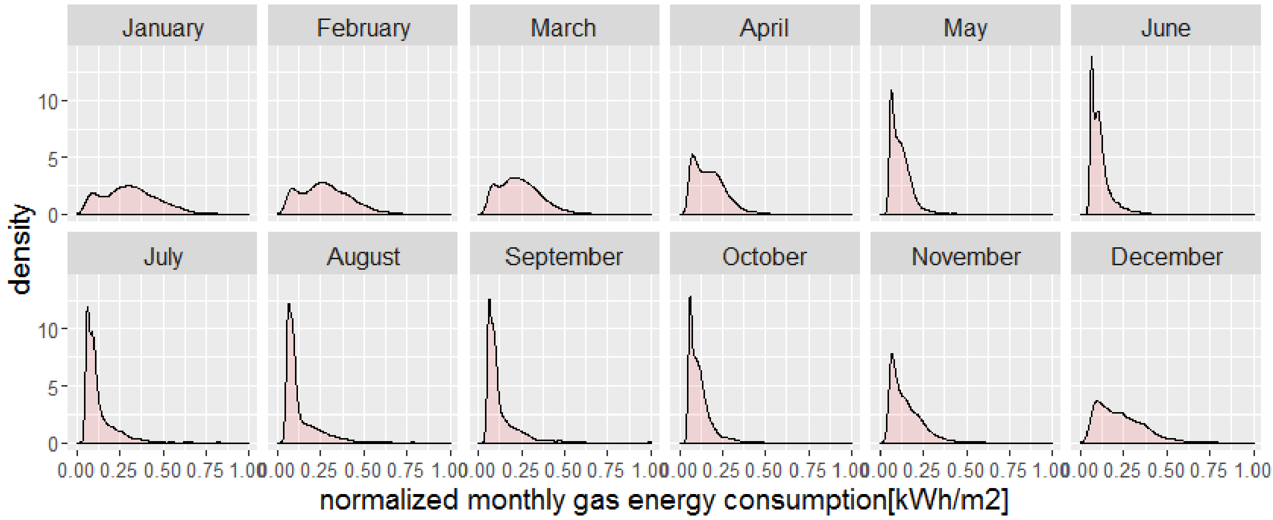


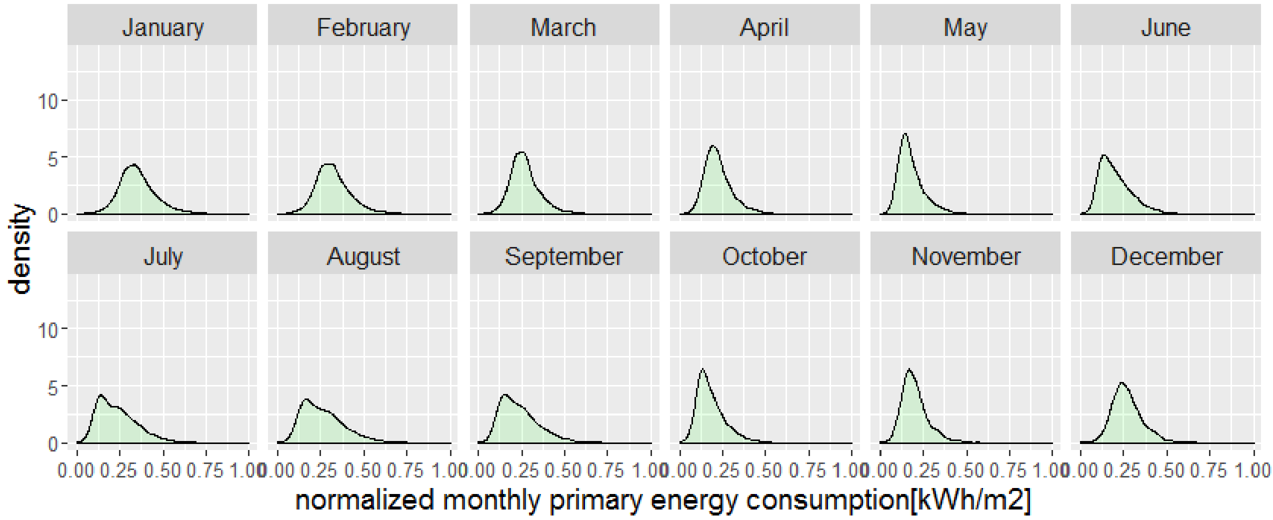
Figure 6: Correlation matrix for yearly primary energy consumption and building information



(a) Electric energy consumption



(b) Gas energy consumption



(c) Primary energy consumption

Figure 7: Probability density plots of monthly energy consumption

References

Beyer, M. and D. Laney (2012). The importance of 'big data': A definition, Gartner, Tech. Rep.

Cabena, P., P.O. Hadjinaian, R. Stadler, J. Verhees, A. Zanasi (1997). Discovering data mining: From concept to implementation. Upper Saddle River, NJ: Prentice Hall.

Chen, M.S., J.W. Han, P.S. Yu (1996). Data mining: An overview from a database perspective, IEEE Transactions on Knowledge and Data Engineering, 8 (6), 866-883.

Fayyad, U., G. PiatetskyShapiro, P. Smyth (1996). From data mining to knowledge discovery in databases, Ai Mag, 17 (3), 37-54.

Han, J., M. Kamber, J. Pei (2012). Data mining concepts and techniques (3rd ed.). Waltham, US: Morgan Kaufmann.

Keim D.A. (2002). Information visualization and visual data mining, IEEE Transactions on Visualization and Computer Graphics, 8 (1), 1-8.

Liu, W.L., J.Y. Zhang, B. Bluemling, A.P.J. Mol, C. Wang (2015). Public participation in energy saving retrofitting of residential buildings in China, Applied Energy, 147, 287-296.

Park, D.J., K.H. Yu, Y.S. Yoon, K.H. Kim and S.S. Kim (2015). Analysis of a building energy efficiency certification system in Korea, Sustainability, (7).

Wu, Z., B. Wang, X. Xiac (2016). Large-scale building energy efficiency retrofit: Concept, model and control, Energy, 109 (15), 456-465.

Zhun F.H., Yu, Benjamin C.M. Fung (2016). Advances and challenges in building engineering and data mining applications for energy-efficient communities, Sustainable Cities and Society, 25, 33-38.