

34

KRIHS
SPECIAL
REPORT
2016

Vol.

THE PROMISE OF BIG DATA
IN REAL ESTATE POLICY:
IDENTIFYING PROBLEMS AND
WORKABLE SOLUTIONS IN
REAL TIME



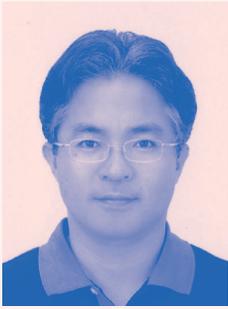
KRIHS Korea Research Institute for
Human Settlements

The Promise of Big Data in Real Estate Policy: Identifying Problems and Workable Solutions in Real Time

The study makes suggestions to actively utilize big data in the policymaking process. The target of real estate policy is spatial and the required information is mostly related to location. Many kinds of big data can also be transformed into geo-referenced data, and these geospatial big data allow effective and customized policymaking and implementation.

Policy effects on market revitalization was investigated using real-time data of housing sales and how the sales volume and prices changed over time at the various spatial units from parcel to administrative boundary. As a result, big data turned out to be extremely useful in the policymaking process because they provided a good opportunity to understand the trends or pattern changes of the real world quickly, and made possible timely effective policymaking and real-time evaluation of policy effects. Social media data turn out very useful to understand opinions of people on real estate policy.

THE PROMISE OF BIG DATA IN REAL ESTATE POLICY: IDENTIFYING PROBLEMS AND WORKABLE SOLUTIONS IN REAL TIME



Daejong Kim

Dr. Kim is the director of Geospatial Analytics Center at Geospatial Information Research Division of Korea Research Institute for Human Settlements.

He received his Ph.D in geography focused on spatial statistics at the State University of New York at Buffalo, USA. His major research area includes geospatial analytics, spatially explicit modelling and collaboration based geospatial platform.

His major research projects are as follows:

- Spatial knowledge platform development and utilization for collaborative policy research, 2016, KRIHS
- A study on building and utilization of spatial knowledge platform for scientific territorial planning, 2014-2015, KRIHS
- KOPSS(Korea Planning Support System) development, 2009-2015, Ministry of Land, Infrastructure and Transportation
- Big data utilization for monitoring territorial policy responses and predicting policy demand, 2014, KRIHS
- Developing a conceptual framework for urban simulation model to assess impacts of compact city policies, 2013, OECD
- Land use change prediction with spatiotemporal pattern analysis and strategies for urban policy, 2011, KRIHS
- Land use change prediction based on urban policy scenarios, 2011, KRIHS
- Land management information system development, 1998-2003, Ministry of Land, Infrastructure and Transportation
- Land use planning with GIS for Jeju Island, 1995-1998, Jeju Province
- The application of GIS to land use planning process, 1998, KRIHS

Contents

Summary	04
Chapter I. Introduction	
1. Background and Purpose	07
2. Scope and Methodology	08
Chapter II. The Concept of Big Data and Real Estate Big Data	
1. The Concept of Big Data	09
2. Types of Big Data	10
3. Big Data and Spatial Big Data	10
4. Major Big Data as Related to Real Estate	12
Chapter III. Using Big Data to Determine the Effectiveness of Real Estate Policy	
1. Methods of Assessing Effectiveness of Real Estate Policy	15
2. Establishment of Big Data	18
3. Big Data Analytics	23
Chapter IV. Analysis of Responses to and Demand for Real Estate Policy	
1. Analysis of Public Opinion on Real Estate Policy	25
2. How Effective Is Real Estate Policy?	35
3. Identifying Key Issues in Real Estate Policy	42
Chapter V. Conclusion	
1. Utilization of Big Data and Its Implications	51
2. Limitations of Big Data Utilization	52
3. Recommendations	52
References	54

Summary

Effective and transparent real estate policymaking requires scientific and objective analyses of various data. The real estate policymaking process is composed of setting up policy agenda from real-world problems, creation of alternatives, making decisions, implementation and evaluation. The process requires recent data for each step to formulate effective real estate policy but unfortunately, this is not easy due to circumstances. First, time is required to collect required data, and worse, most of the data were created in the past while issues were ongoing. Second, precise diagnosis of status to identify problems is impossible because there are not enough tools for evaluating how policy works and what the associated issues are. Third, real estate policy associated with complicated conflict tends to be settled in a political or top-down manner without data analysis.

On the other hand, information telling how people respond and behave to real estate policy are being produced in real time thanks to advancement in IT. A variety of information including from Twitter, blogs and internet cafes, travel, credit card payments and calling information are being produced every minute and the volume is huge. Such data are called big data and recognized as a core source for customized policymaking because they provide avenues to understanding how people think, make decisions and behave.

The goal of this study is to provide a method for how to use big data for transparent and scientific policymaking in real estate. First of all, the study defines the concept of big data, identifies why they are crucial and what kinds of big data are available for real estate policymaking as an example.

A case study on response monitoring in the real estate market from relevant policies was conducted to identify how helpful big data are and utilize them. A package of real estate policies in 2013 was announced twice to revitalize the stagnant real estate market on April 1 and October 28. At first, the public's positive and negative responses to the policy could be assessed using sensitivity and opinion analyses of crawled big data from social network services including Twitter, blogs, online cafes and news with the keywords "acquisition tax" and "transfer income tax." In addition, what caused the positive or negative responses was identified. Policy effects on market revitalization was investigated using real-time data of housing sales and how the sales volume and prices changed over time at the various spatial units from parcel to administrative boundary. As a result, big data turned out to be extremely useful in the policymaking process because they provided a good opportunity to understand the trends or pattern changes of the real world quickly, and made possible timely effective policymaking and real-time evaluation of policy effects.

Finally, the study makes suggestions to actively utilize big data in the policymaking process. The target of real estate policy is spatial and the required information is mostly related to location. Many kinds of big data can also be transformed into geo-referenced data, and these geospatial big data allow effective and customized policymaking and implementation. The Korean government is implementing a geospatial big data project, and to be successful, a concrete vision and strategies should be identified and public and private domains need to work together to promote the sharing of big data.

Introduction

I. Background and Purpose

For real estate policy to be effective, how the real estate market works must be precisely understood. The policymaking process requires consideration of current issues, identification of solutions and effective policy implementation. Each step of this process requires a huge volume of relevant data. Current issues must be identified rapidly and with a high degree of accuracy, the data must be analyzed in a comprehensive manner, and the performance of the resulting policy decisions must be evaluated with objectivity.

Timely data is crucial to this process, but not always easy to obtain. Obstacles include the time required to collect data and the rapidness with which data loses its relevance. The widespread distribution of devices such as computers and smartphones, however, has made a greater volume of timely data available. Traffic movements are captured on cameras and at tollgates. Credit card payments and phone calls generate vast amounts of data in real time. Twitter and other social networking services, blogs, internet forums and so on are rich sources of what is now known as big data. Big data is becoming increasingly important in the development of a customized policymaking process because it can be used to track people's opinions, decisions and behavior.

This article looks at how big data, one of the hottest topics today, can be used in the process of formulating real estate policy. For example, analyzing big data on real estate can show real estate policymakers how people are responding to policy measures and what issues might be arising. In addition, the analysis of real estate transactions and rent contract data can be used to assess the market impact of recent developments in the nation's real estate policy.

“

In addition, the analysis of real estate transactions and rent contract data can be used to assess the market impact of recent developments in the nation's real estate policy.

”

2. Scope and Methodology

This research utilizes big data to assess public opinion on policy measures the Korean government announced in 2013 to stimulate housing sales and measure their impact on the market. To determine where changes might be needed in real estate policy, we also analyzed public opinion about the real estate market.

01. A policy announced on April 1, 2013, to increase demand for housing sales and alleviate the rapid increase in *jeonse* prices as demand for *jeonse* exceeded demand for housing sales due to a sluggish real estate market. (*Jeonse* is a home rental system unique to Korea that refers to the way apartments are leased. Instead of paying monthly rent, a tenant gives the landlord a lump-sum deposit that is refunded upon the tenant moving out.) The policy exemption provided tax breaks and exemptions for those eligible in transfer and acquisition taxes, abolished a transfer surtax that affected select homeowners and extended financial support for housing purchases.
02. A follow-up policy to the April 1 measure that entailed a lower acquisition tax, income tax deductions, greater financial support for housing purchases and the expansion of long-term mortgages with low interest rates.

To these ends, this study used big data from both the public and private sectors. To measure the public's response to the April 1⁰¹ and August 28⁰² measures, we collected content through crawling for the period beginning March 2013 and ending September 2013. The content was drawn from social networking services, news reports, internet forums and blogs and contained keywords related to both measures. A sentiment analysis was then conducted to determine whether the response to a policy was positive or negative.

One aim of both measures was to encourage prospective tenants to purchase their homes rather than conclude rental contracts under the *jeonse* system (under which a tenant gives the landlord a large lump-sum deposit in lieu of monthly rent; the sum is given back to the tenant when he or she moves out). To determine whether the two measures had the intended impact on housing sales, housing sales figures were studied to see if transaction volumes had increased year on year. Accordingly, housing sales data was collected for the period between January 2011 and November 2013 and converted into spatial data using a cadastral map. Through visualization, the map allowed researchers to see which parts of the country experienced a higher volume of sales transactions. To analyze real estate market trends in greater detail, we also collected online documents that included keywords such as "housing," "redevelopment," "reconstruction" and "country(sub-urban) house."

The Concept of Big Data and Real Estate Big Data

I. The Concept of Big Data

To understand the significance of big data, the first step is to define data. Data is information about the real world stored in computer files or other media in various forms such as characters, numbers, images, plans, maps (spatial information) and videos. Because the production of data requires complex technology and considerable cost, the public sector has mainly had the resources to produce and distribute data. Large volumes of data have been produced at intervals, like on an annual basis.

With rapid technological development that enable the production and sharing of vast amounts of data, data has been proliferated. Smartphones and social networking services have made it easier for anyone to produce and share information, and sensor technology enables the automatic production of data (for example, traffic data) in real time. Moreover, the development of satellite technology has allowed collection of data about changes on the earth's surface with far greater precision and timeliness than in the past. This proliferation of data has created a need for new technologies to store, manage and process the huge volume of data at people's fingertips.

The three defining properties of big data, or “the three Vs” as Gartner (2011) referred to them, are volume, variety and velocity.⁰³ That means big data is characterized by the huge volume of information that it conveys, the rapid pace at which it is produced and processed, and the variety of forms that it takes. On Twitter, for example, an average of 5,700 tweets were generated per second in 2013, or about 500 million tweets per day. And in a single day—January 1, 2013—4.8 billion messages were sent via the messaging app Kakao Talk. And thanks to information technology, the public sector can update in real time daily population changes and land transaction data. These are also examples of the power of big data.

Sensors that collect data about real-time events—surveillance camera footage, satellite and aerial photos, and data on traffic conditions—are considered important generators of big data. Blogs, photos, videos and social networking content—all forms of unstructured data produced by individuals—are becoming increasingly important as sources of information unobtainable from structured data.

03. O'Reilly Radar Team. *Planning for Big Data*. 2012.

2. Types of Big Data

Big data can be classified as structured, unstructured or semi-structured. Structured data refers to any data stored in a fixed field. Most data stored in relational databases are structured data. Structured data is also accumulated when companies and public organizations use information systems in their business operations. For example, land transaction information includes fixed data fields on buyers and sellers, the addresses and characteristics of the properties being bought and sold, and other details.

Unstructured data is the opposite because it is produced through blogs, internet forums, social networking services, news media and others. Examples of unstructured data include text and voice messages, images and video clips. Unstructured data is expanding much more rapidly than other types of data thanks to the widespread use of smartphones and social networking services.

Semi-structured data lies somewhere in between the above categories. Photographs are classified as unstructured data, but details such as when and where photos are taken comprise structured data contained in file metadata. Other examples of semi-structured data are word-processing and presentation documents, email and PDF files.

3. Big Data and Spatial Big Data

About 80 percent of all information in the world is geospatially referenced (Burghardt Hahmann 2013). That means about 80 percent of all information can be linked to or expressed as geospatial data. Geospatial data refers to vector data (points, lines and polygons) and raster data (lattice data) that represents the location of objects in the world and objects with shapes.

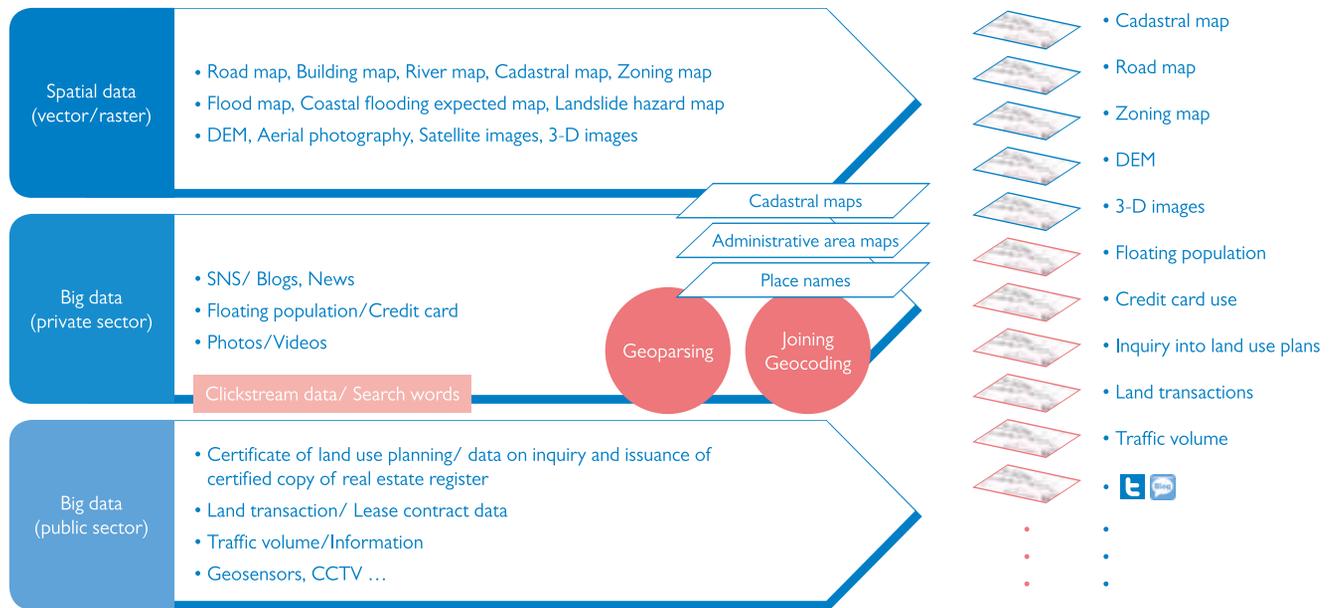
“

Thus spatial big data refers to structured, semi-structured or unstructured big data integrated with or expressed as geospatial data. Figure I shows the relationship between big data and spatial big data.

”

Big data takes various forms such as characters, numbers, images and video clips. Because big data usually has information on where it was produced, it can usually be translated into or expressed in the form of a map and incorporated into cadastral or administrative area maps and place name descriptions, all of which constitute spatial big data (Daejong Kim et al. 2014). For example, address data on real estate contracts can be incorporated into cadastral maps showing real estate transactions. In addition, by extracting location data from blogs and social networking content, it is possible to synthesize such address data with general location data. Floating population data produced by telecom companies represents the locations of smartphone users in “smart cells.” Thus spatial big data refers to structured, semi-structured or unstructured big data integrated with or expressed as geospatial data. Figure I shows the relationship between big data and spatial big data.

Figure 1. Relationship between Big Data & Spatial Big Data



Source: Daejong Kim et al. 2014, p.27.

Spatial big data is important because of the insights it offers into the interactions between people and their environments. Existing spatial data mainly concerns the physical and institutional environment, but spatial big data—for example, data on floating populations, credit card use and real estate transactions—relates to people’s thoughts and behavior. So analyzing spatial data alongside spatial big data can show how people think and behave in specific environments. In this way, the public sector can learn more effectively and in a timelier manner how to improve processes for policy formulation and implementation.

4. Major Big Data as Related to Real Estate

Real estate portals in Korea's private sector, such as Real Estate 114 and Naver Real Estate, have accumulated region-specific information on houses for sale and users' search logs and research data. The public sector, for its part, has collected other kinds of information: how many land-use plan certificates have been viewed and issued, searches for land-use regulations, real estate transaction data, requests for certified copies of real estate agencies' business registration certificates and real transaction prices, to name a few. Real estate portals listing homes for sale always record visitor searches and page views, and before executing transactions, the agencies must issue land-use planning certificates and certified copies of the real estate agencies' business registration certificates. All of these activities generate enormous amounts of data.

Such data constitutes the real estate market's leading indicators and the basis of an effective direction of real estate policy. In response to recent changes in the real estate market, certain public organizations plan to create a database to encompass inquiries about lots of land and searches for real estate portals about homes for sale. Figure 2 shows the results of a time-series analysis of real estate lease prices in each of Korea's *dong* (a neighborhood smaller than a *gu* in Korea's land administrative designation). The results reflect the impact of seasons, policies and regional factors on rental rates.

Social networking services like Facebook and Twitter allow people to use computers or smartphones to interact with each other and share information and opinions. Data generated through social networking services are mostly unstructured data such as photos, news reports and comments. As Facebook blocks web crawlers from accessing its services, data collection directly from Facebook is difficult, but Facebook sells data after processing. Twitter allows access by web crawlers and is a major source of big data from social networking services.

Blogs provide information on individual's interests and their daily lives. Internet forums also focus on users' specific areas of interest. The nature of such platforms renders them useful for the harvesting of data on certain issues only. In Korea, Naver blogs and online forums are major sources of big data. Figure 3 represents Twitter search results about the April 1 measure (announced on April 1, 2013, to revitalize the housing market).

Figure 2. Time-series Jeonse Prices in Seoul Neighborhoods

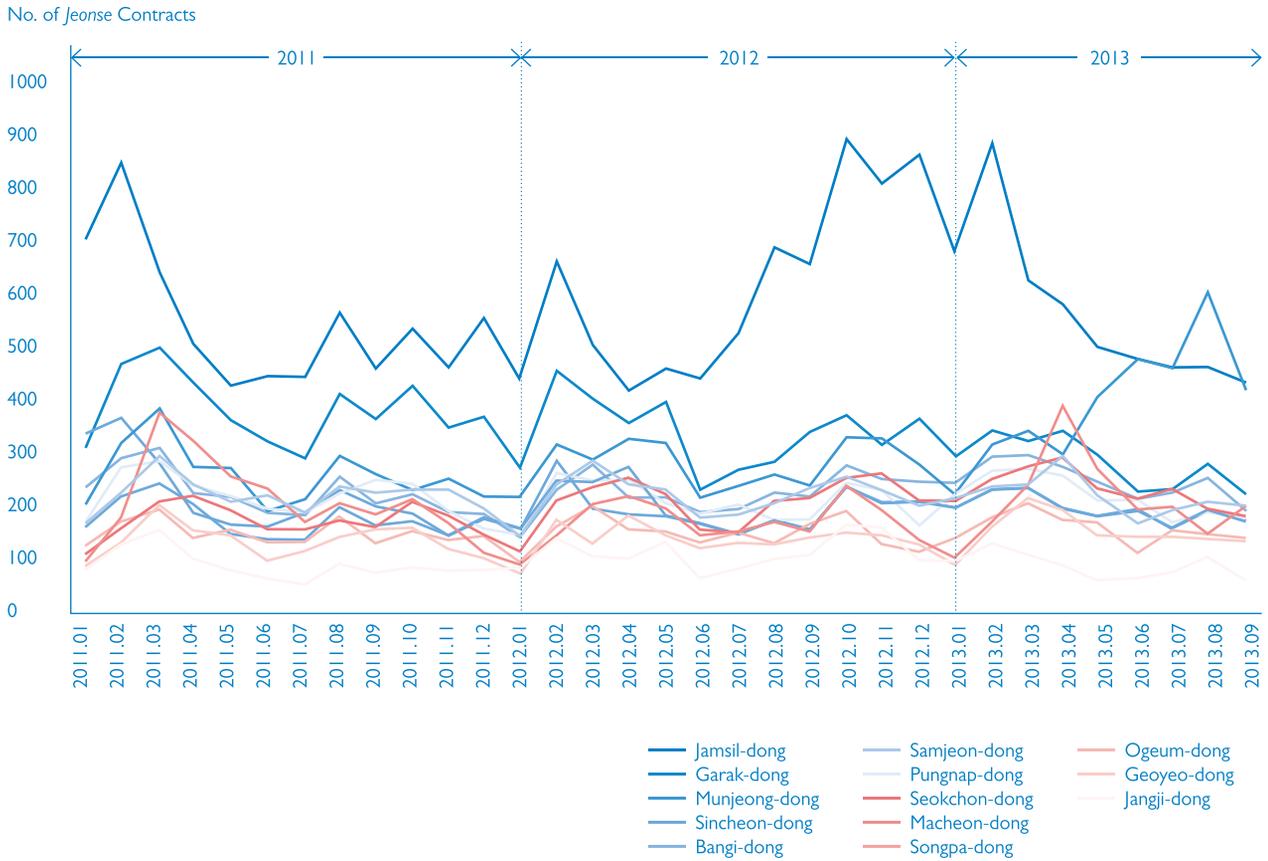


Figure 3. Twitter Search Results on April I Measure

 The front page of the Maeil Business Newspaper reported that demand for small apartments in northern Seoul was expected to rise “like wriggling bugs.” It’s funny that they always say real estate market prices are “wriggling” as if they’re bugs. Don’t trust these reports as they’ve been almost the same since the **April I measure** was announced. I think they’ll last for only two months at maximum.
2013. 09. 02

 The Korean government has become the boy who cried wolf as its real estate **measure** faces obstacles in the National Assembly.... **The April I** and August 28 **measures** failed to pass parliament because of political disputes.
2013. 10. 24

 The opposition party: “The benefits for the rich of the government’s real estate measure are 25 times higher than those for the working class.” Among the six new non-taxable and tax reduction items of the **April I measure**, three items provide much more benefits to the rich.
2013. 10. 18

 Jeonse loans for low-income people who do not own homes will be launched in August. Please visit I I0 Blog to check what support is provided for low-income people under the **April I measure**.
2013. 8. 13

Using Big Data to Determine the Effectiveness of Real Estate Policy

I. Methods of Assessing Effectiveness of Real Estate Policy

I.1 Overview of April 1 and August 28 Measures

As the real estate market remained sluggish in 2013 and home ownership grew less affordable, demand for *jeonse* increased, leading to greater difficulties with supply. This is why the Korean government announced the April 1 and August 28 measures. This article aims to evaluate the effects of the government's measures on the market, with a focus on housing sales transactions.

Prior to the April 1 measure, housing sales had been declining due to a sluggish market and falling prices in the face of growing fears over the *jeonse* market. A severe decline in the housing market can negatively affect the economy and undermine the stability of financial markets. Thus the government implemented the April 1 measure to take on fears over the economy, stabilize housing options and help low-income people. Major points of the measure included cuts and exemptions to transfer and acquisition taxes. Table I shows details on housing sales.

“

Thus the government implemented the April 1 measure to take on fears over the economy, stabilize housing options and help low-income people. Major points of the measure included cuts and exemptions to transfer and acquisition taxes. Table I shows details on housing sales.

”

Source: Ministry of Land, Infrastructure, and Transport. 2013. "Comprehensive Measure to Normalize the Housing Market for Stabilization of Housing for Low-Income People." Briefing materials.

Table 1. Major Points of April 1 Measure

	Applicable cases and conditions	Details of implementation
Exemption of transfer tax	<ul style="list-style-type: none"> • KRW 900 million or less, 85m² or smaller, people who own one home/ homes purchased by end of 2013 • If unsold new housing units are purchased / people who own a house purchase homes 	<ul style="list-style-type: none"> • Temporary exemption of transfer tax when homes are purchased by year's end
Abolition of heavy transfer tax	<ul style="list-style-type: none"> • If people who own multiple houses transfer houses. ※ The August 28 measure repealed imposition of a differential transfer tax on people who own multiple home. 	<ul style="list-style-type: none"> • 40% transfer tax if person transfers home after owning it for less than one year • Proportional tax rate of 6–38% applied if person transfers home after owning it for one year or more.
Exemption of acquisition tax	<ul style="list-style-type: none"> • First-time home purchase KRW 600 million or less in price, less than 85m², combined income of couple KRW 60 million or less • Houses acquired by end of 2013 	<ul style="list-style-type: none"> • Acquisition tax fully exempted.
Financial support for housing purchase	<ul style="list-style-type: none"> • Non-homeowner, KRW 600 million or less, 85m² or less, income of KRW 60 million or less • Financing can be provided up to 70% of loan-to-value ratio. 	<ul style="list-style-type: none"> • Annual rate of 3.5% or less • Criteria for single person is annual income of KRW 50 million. Support to be expanded by KRW 2.5 trillion

Despite the measure, the government determined that housing sales remained sluggish and *jeonse* prices continued to rise due to an imbalance between supply and demand in the *jeonse* and monthly rent markets. So a follow-up measure came on August 28 after consultations with relevant organizations. Highlights of the August 28 measure included the normalization of the housing market, expansion of the rental housing supply and the removal of burdens that had previously fallen on tenants. Details can be found in Table 2.

Table 2. Highlights of August 28 Measure

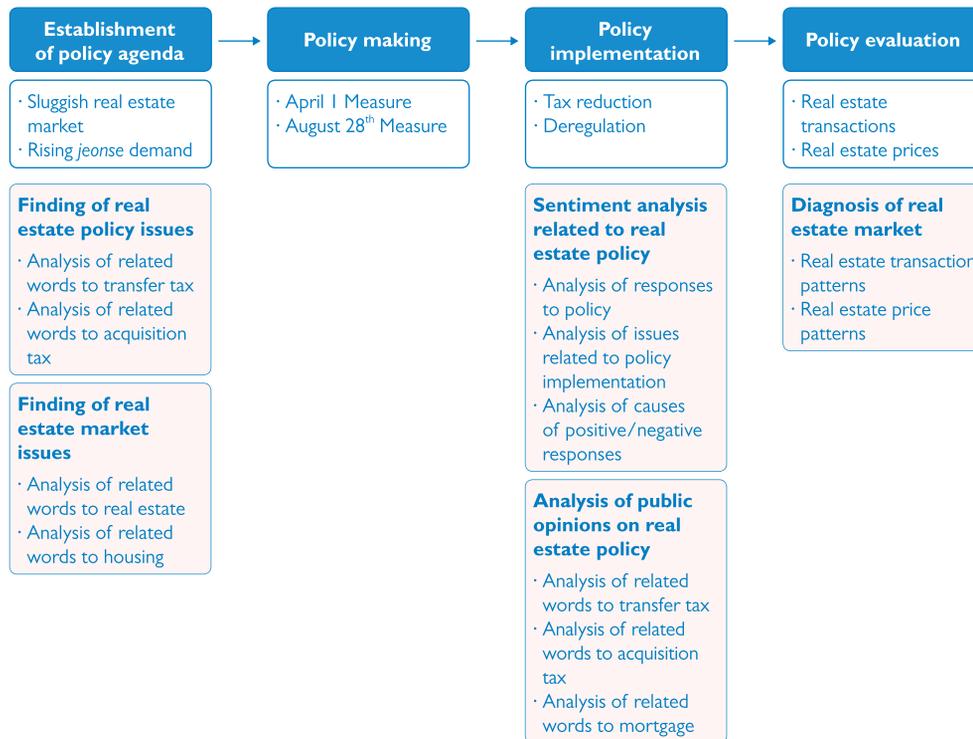
Source : Ministry of Land, Infrastructure, and Transport. 2013. "Comprehensive Measure to Normalize the Housing Market for Stabilization of Housing for Low-Income People." Briefing materials.

	Applicable cases & conditions
Reduction of acquisition tax	<ul style="list-style-type: none"> • Range of acquisition prices <ul style="list-style-type: none"> - Up to KRW 600 million : 2%→1% - KRW 600 million–900 million : 2%→2% - More than KRW 900 million : 4%→3%
Income tax deduction for mortgage with a long-term fixed-rate installment plan	<ul style="list-style-type: none"> • Income tax deduction for mortgage on home with standard market price of up to KRW 400 million • Income tax deduction for mortgage on house if homeowner purchases new house for replacement
Expansion of financial support for housing purchase	<ul style="list-style-type: none"> • Criteria on combined income of couple increased to KRW 60 million. • Loan amount of up to KRW 200 million • Criteria on applicable housing price increased to maximum of KRW 600 million (including studio apartments) • Applicable interest rates adjusted to 2.8–3.6%.
Introduction of profit-sharing/ profit-loss-sharing mortgage	<ul style="list-style-type: none"> • Profit-sharing mortgage: structure that offers low interest rates, but profits should be shared with government's national housing fund when housing prices rise • Profit-loss-sharing mortgage: structure where profit or loss is shared depending on changes in housing prices or ratio of loans • Pilot project for up to 3,000 households (implementation scheduled for around October)

I.2 Analytical Framework

So to what extent have the government’s measures succeeded in advancing its goals? How much was achieved in stabilizing the *jeonse* market and stimulating real estate transactions? Have the policy measures proven disappointing, and if so, why? This study aims to ascertain where and how big data can be utilized to devise policy for the real estate market, implement sound policy decisions and conduct effective follow-up assessments. Figure 4 represents a conceptual model showing the value of big data for such purposes.

Figure 4. Big Data Analysis Method to Assess Effectiveness of Real Estate Policy



In the initial stages, when policymakers work on devising a policy agenda, big data can help identify issues, for example, in the form of tweets that contain terms such as “transfer tax,” “acquisition tax,” “*jeonse*” and “lease.” Other issues can be uncovered by analyzing big data that includes terms such as “real estate,” “housing,” “apartments” and “villa” (a residential building in Korea usually two to five stories high).

In the implementation stage, sentiment analysis can help show how people respond to particular measures in real estate policy. Positive and negative responses can be analyzed to determine why people feel the way that they do, and related matters can also come to light in this process. Meanwhile, collection and analysis of big data using keywords such as “transfer tax,” “acquisition tax” and “mortgage” can help policymakers identify issues related to policy implementation.

When evaluating existing policies, prices and the volume of real estate transactions can be analyzed to determine how effective a policy is and how much progress it has made toward its original objectives.

As Figure 4 illustrates, the use of big data presents limitations in the policy making phase because experts usually decide whether to adopt a policy. But before such decisions are made, big data can be useful in gauging public opinion through seminars and other activities.

2. Establishment of Big Data

2.1 Data Collection

When real estate transactions are concluded and contracts executed, the information is recorded and stored in real time by way of the Real Estate Management System (RTMS). This research used data on real estate sales stored in the RTMS, with details supplied in Table 3. Contract information was analyzed to determine whether the government’s real estate policy measures were effective in reinvigorating the market. To arrive at a comprehensive analysis, we collected housing sales data over 35 months beginning in January 2011.

Table 3. Information Collected for Public Big Data

Type	Details
Housing sale information	<ul style="list-style-type: none"> • Content: address of houses, contract amount, date of contract, fixed date of contract, place of contract or types of buildings, area (306,140 cases) • Area: Seoul • Time range: January 2011 to November 2013

Big data in the private sector was collected from social networking sites, news media, online forums and blogs through web crawling. Social networking content was collected from Twitter, which releases data, and news media information was collected from a variety of Korean websites. Because Naver and Daum host over 90 percent of all online forums and blogs in Korea, all blog and forum data were collected from the two portals. Image content was excluded at the time of web crawling, as were comments on online postings. See Table 4 for more details.

Table 4. Information Collected for Private Big Data

Type	Details
Social networking service	<ul style="list-style-type: none"> • Only Twitter, data were collected due to accessibility
News media	<ul style="list-style-type: none"> • Examples include the Chosun Ilbo & JoongAng Daily
Online forums/ blogs	<ul style="list-style-type: none"> - Various contents in internet forums or blogs provided by search portals • Naver & Daum are main sources. • Information collected from Twitter, news media, internet forums & blogs through web crawling

For private big data, two datasets were collected separately: one for sentiment analysis and the other for related-word analysis. In the sentiment analysis, the objective of which was to measure the public's response to the real estate policy initiatives in question, the keywords were "transfer tax," "acquisition tax" and "mortgage," and a time range was set from March 2013 to September 2013. See Table 5 for details.

Unstructured big data comprised about 94,000 items collected through web crawling. Of these, blogs accounted for about half, the highest proportion, followed by online forums, news and tweets. This dataset for sentiment analysis was the basis of a case study for measuring public opinion and recommending a policy agenda.

On the dataset for analysis of related words, documents containing the word "housing" or related were collected through web crawling between January 2011 and September 2013. The dataset for analysis of related words, unlike that for sentiment analysis that included the full text of each document collected, only recorded the frequency with which words (nouns) appeared in the documents. For this reason, analysis of related words is used in case studies to develop a policy agenda.

Table 5. Details of Documents Collected

	Blogs	Online Forums	News	Twitter	Total
Volume (No. of items)	43,368	30,058	13,138	7,313	93,877
Ratio (%)	46.2	32.0	14.0	7.8	100

2.2 Cleansing of Data on Real Estate Transactions

From public big data used for the case study, that on real estate sales contracts was processed through the open-source software R. The data was then classified as that to calculate transaction volume or data to calculate the percentage of price change for data cleansing. To calculate transaction volume, data was extracted from contracts signed after January 2011 but excluded in cases of obvious data input errors. In calculating price changes, we excluded data showing errors such as incorrect area and price information, as well as overlapping data produced when the same person purchased multiple villas or apartments. Data to calculate transaction volume required a different process of data cleansing from data to calculate price changes because their expected results differed. Thus the number of data removals depended on a number of factors. Table 6 and 7 show the number of items removed during data cleansing in the category of real estate sales transactions, and the number of items that became part of the final data.

Table 6. Data Cleansing to Calculate Transaction Volume

	No. of remained records	No. of removed records
Total contracts	306,140	-
Removal of data before Jan. 2011	298,369	7,771
Removal of Dec. 2011 and 2012 data	283,509	14,860
Removal of data with incorrect fields	283,329	180

Table 7. Data Cleansing to Calculate Percentage of Price Change

	No. of remained records	No. of removed records
Total contracts	306,140	-
Removal of data with incorrect fields	305,917	223
Removal of data with area of 10m² or smaller	304,817	1,100
Removal of price input errors	297,640	7,177
Price area of the largest and smallest 0.05%	291,716	5,924
Removal of data before Jan. 2011	284,254	7,462
Removal of overlapping data on apartments & villas	266,964	17,290

2.3 Processing of Unstructured Big Data

Private big data consisted of documents extracted through web crawling and classified as positive or negative responses through sentiment analysis as indicated in Figure 5. The process for sentiment analysis largely consisted of the preprocessing of sentences, analysis of morphemes and extraction of semantic patterns. In the preprocessing of sentences, special characters used in sentences—periods, commas and brackets—are removed. Analysis of morphemes refers to a process of identifying the structure of a sentence by breaking it into nouns, adjectives and postpositions. Extraction of semantic patterns refers to a process to extract morphemes that can be used to determine whether responses are positive or negative, and that is the key to sentiment analysis. In the initial stage of classifying documents, a large number of words that mean positive or negative sentiment are collected through the three stages mentioned above. Then the sentiment (positive or negative) of the collected documents is determined by counting how often the positive and negative words appear.

Data processing for analysis of related words is similar to data processing for sentiment analysis. First, special characters are removed from sentences and morphemes are analyzed. Then sentences to be analyzed are divided into nouns, adjectives and postpositions, and only nouns are considered when researchers count to see how frequently nouns appear in documents.

“
Then the sentiment (positive or negative) of the collected documents is determined by counting how often the positive and negative words appear.
”

Figure 5. Data Classification for Sentiment Analysis

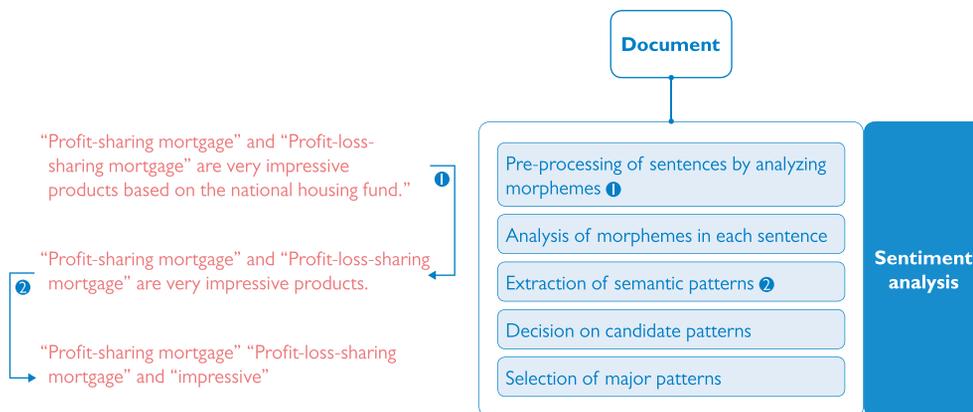
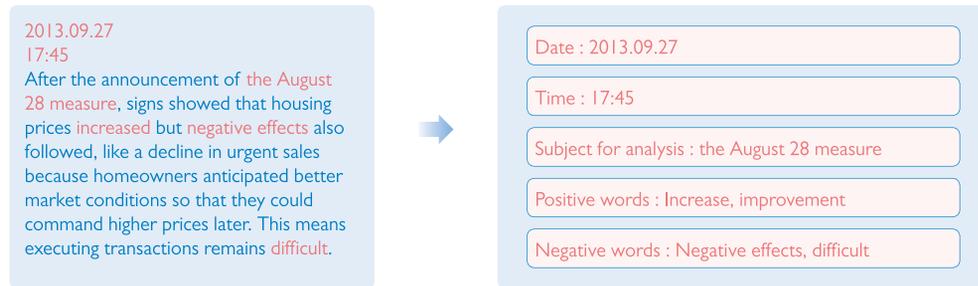


Figure 6 shows an example of processing based on text mining. When a smartphone user posted his or her opinions about real estate policy on a social networking site in the form of sentences, online documents containing keywords are collected through web crawling. Then classification and extraction begin through text mining for reprocessing for various analytical purposes such as date and time of extraction, subjects to be extracted, and positive and negative words.

Figure 6. Example of Data Extraction via Text Mining



2.4 Turning Big Data into Spatial Data

Structured, semi-structured and unstructured big data can contain various types of location information. Unstructured data can include place names, addresses or information about where photos were taken or where people used Twitter. Structured data including location information contains latitude and longitude coordinates. But in the case of unstructured data, location is indicated in words or sentences within the text; thus a method to convert them into spatial data is needed.

Big data that contains location information can be converted into spatial data through integration with spatial data, a process called geocoding. To carry out geocoding, certain information is needed to connect the two. For example, if the address information in big data and that in spatial data are same, they can be combined.

04. Big data produced in the public sector such as that of land transaction store address information mostly in the form of a 19-digit code: *si* (metropolitan or special city) /*do* (province) (2) + *si* (city) /*gun* (county) /*gu* (ward) (3) + *eup* (town) /*myeon* (township) /*dong* (neighborhood) (3) + *ri* (village) (2) + unit of a lot (1) + main number of a lot (4) + sub-number of a lot (4)

The level of detail in location information contained in big data can determine the level of detail obtainable for spatial data. Depending on the level of detail present in location information in big data, it can be combined with the central point of a lot of land or an administrative district such as an *eup* (town), *myeon* (township), *dong* (neighborhood), *si* (city including a metropolitan or special city), *gun* (county), *gu* (ward), or *do* (province). For example, addresses in real estate transactions or lease contract data⁰⁴ contain *beonji* (house address or street number), which can be connected to specific lots of land. Statistical data at the level of a *si* (city), *gun* (county), or *gu* (ward) can be combined with an administrative map at similar levels and converted into spatial data. In addition, online addresses in clickstream data, which are accumulated when internet users click on links, can be linked to some level of location.

3. Big Data Analytics

3.1 Big Data Analytics

Big data analytics is a process used to gain insightful and quality knowledge from a large volume of data with relatively low value (Manjai Lee 2012). That is, big data analytics collectively refers to various processing methods used to analyze and utilize big data. Big data analytics can be largely divided into techniques of analysis and visualization (Google, Wikipedia). Data mining, machine learning, natural language processing and pattern recognition are techniques or methodologies used to analyze big data. With the proliferation of unstructured data in such forms as social networking and news content, more attention has recently gone to text mining, opinion mining, cluster analysis and technology for analysis of social networking service.

Meanwhile, visualization techniques are also an important element of big data analytics because a new trend or patterns can be found depending on how data is expressed. In addition, communication with other people is important. Visualization techniques vary from simple graphs to 3-D maps, and sometimes require creativity on the part of analysts. Information graphics is a term similar to visualization.

3.2 Big Data Visualization Methods

Location information is extremely useful for data analysis because it enables analysis of spatiotemporal patterns as well as space, and, most importantly, because the data is easy for users to understand. Spatial big data can provide more precise and valid insights by using correlations between phenomena in a certain space as well as relationships among attributes for analysis. Figure 7 represents an example of spatial big data analytics. In this example, spatial big data was used to visualize on a map how many people moved to San Francisco and where they were originally from. This was to facilitate understanding of spatial distribution and volume.

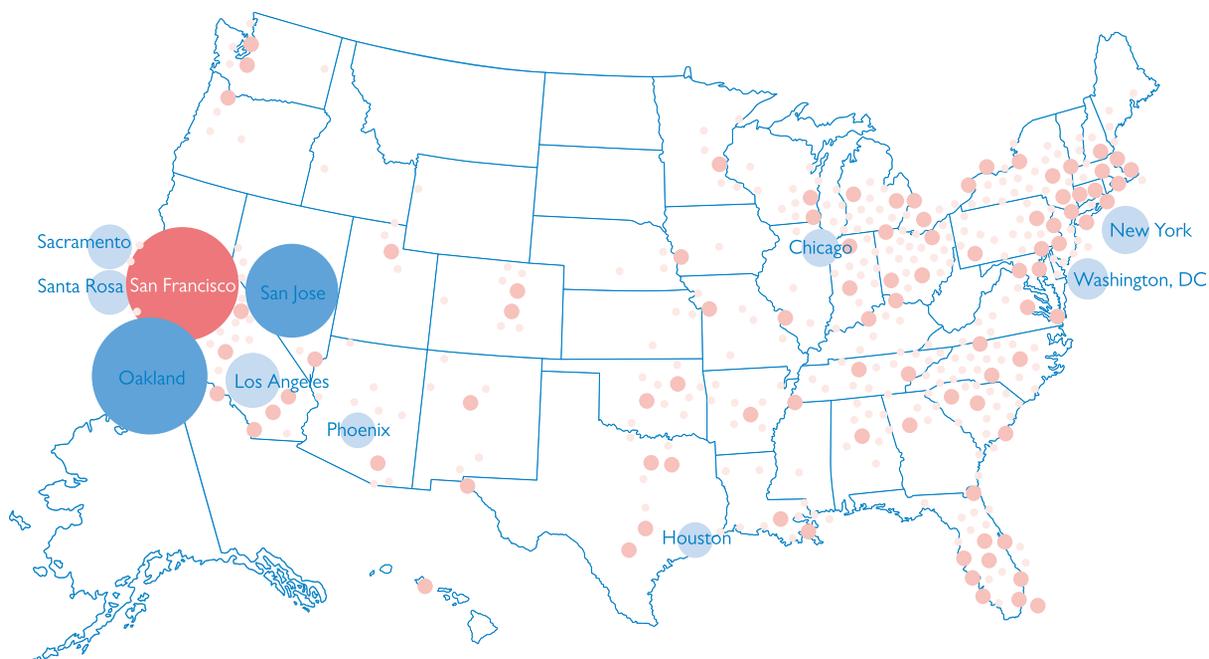
“

Spatial big data can provide more precise and valid insights by using correlations between phenomena in a certain space as well as relationships among attributes for analysis.

”

Figure 7. Visualization of House Hunters Moving to San Francisco

- Inbound searches
- Outbound searches



This map features home searched happening in the top 100 US metros. Click a metro to see the top ten metros where house hunters are looking to move in from (Inbound Searches) and move out to (Outbound Searches). The bigger the circle, the more home searches that are coming into or out of the selected metro area.

Source: Trulia trends (<http://trends.truliablog.com/vis/metro-movers>)

Analysis of Responses to and Demand for Real Estate Policy

I. Analysis of Public Opinion on Real Estate Policy

I.1 Analysis of Public Opinion on Transfer Tax

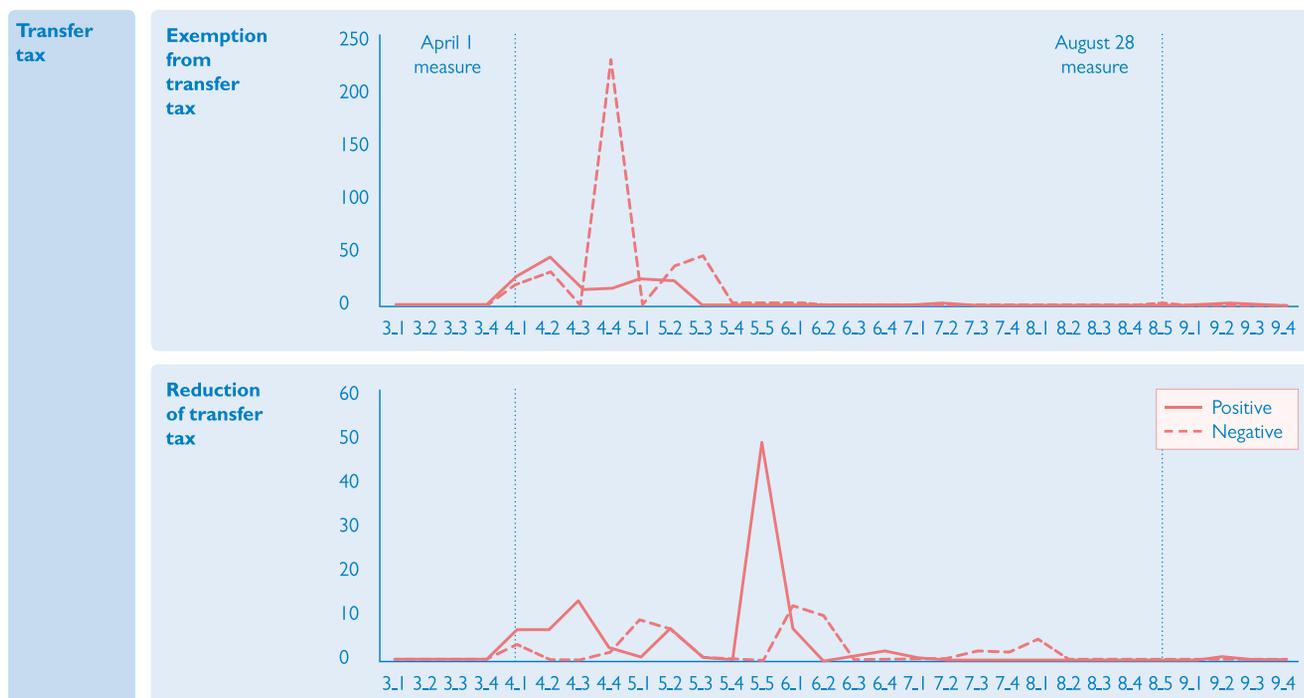
The April 1 measure temporarily suspended the transfer tax for all homeowners and reduced the tax for owners of multiple homes. The August 28 measure abolished the surtax for owners of multiple homes. Figure 8 shows the results of public opinion analyses of the decisions to cut and later abolish the transfer tax for multiple home owners. The horizontal axis represents weeks of the month and the vertical axis the number of positive/negative documents around dates when the two measures were announced. On the suspension of the transfer tax for first-time homebuyers, positive responses appeared after the April 1 measure was announced. But after mid-April, the negative responses far outnumbered the positive ones. On the reduction of the transfer tax for multiple home owners, positive responses outnumbered negative ones, but the frequency of the phrase “transfer tax reduction” decreased significantly.

“

But after mid-April, the negative responses far outnumbered the positive ones. On the reduction of the transfer tax for multiple home owners, positive responses outnumbered negative ones, but the frequency of the phrase “transfer tax reduction” decreased significantly.

”

Figure 8. Monitoring of Responses to Transfer Tax Breaks & Exemptions



In analyzing public opinion on the transfer tax, the researchers sought to determine more than just the prevalence of negative and positive responses. We also performed a word-cloud analysis to gain more insight into respondents’ thoughts. Word clouds are graphical representations of word frequency that give greater prominence to words appearing more frequently in a source text. The more frequently the word appears, the more important it is. Figures 9 through 12 show the word clouds that resulted from our questions about tax breaks and exemptions on the transfer tax. Table 8 shows the frequency of each word that appeared.

Positive comments on the transfer tax suspension consisted mainly of words expressing respondents’ expectations of recovery in the real estate market such as “reinvigoration” and “expectations.” But the volume of web documents extracted was smaller than that of negative comments. But in negative comments on transfer tax waiver, words such as “confusion,” “effectiveness,” “feeling of deprivation” and “inequality” appeared frequently. Moreover, the frequency of such negative words was higher than that of positive words. When the original text was analyzed, researchers found many negative opinions. For example, many respondents said the criteria for transfer tax exemptions were too limited, that only homeowners benefited and that they doubted the measure’s effectiveness.

Among the positive comments, many expressed optimism for the recovery of the real estate market, like comments on transfer tax exemption. Negative comments on the transfer tax cut expressed fears about and objections to the measure, and words such as “uncertainty,” “objection” and “unclear” were used to express doubt over the measure’s effectiveness.

By its very nature, the transfer tax measure benefited those who owned homes or who expected to own them later. Such a measure might be conducive to stimulating additional housing transactions, but is of no value for those who cannot afford to purchase homes. Such opinions also were reflected in the sentiment analysis. Thus we can conclude that the government’s real estate policy needs to benefit people who cannot afford to purchase homes, for example, by facilitating better terms and conditions on loans or vouchers.

Figure 9. Exemption from Transfer Tax – Positive Responses

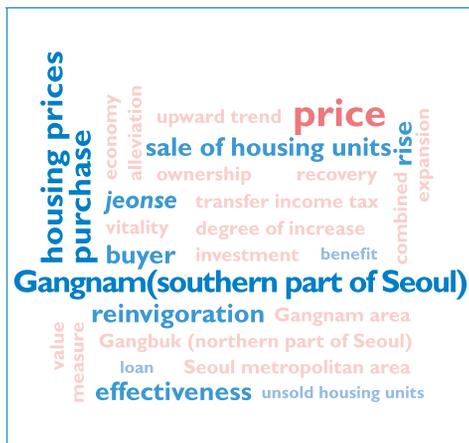


Figure 10. Exemption from Transfer Tax – Negative Responses

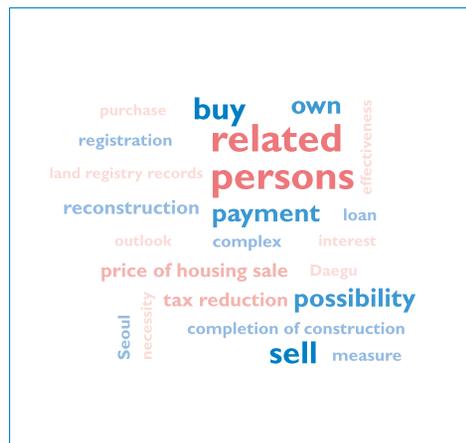


Figure 11. Reduction of Transfer Tax – Positive Responses



Figure 12. Reduction of Transfer Tax – Negative Responses



Table 8. Frequency of Words Appearing in Original Text of Responses to Sentiment Analysis on Transfer Tax

Exemption of Transfer Tax				Reduction of Transfer Tax			
Positive Words		Negative Words		Positive Words		Negative Words	
Reinvigoration	39	Confusion	112	Reinvigoration	96	Concern	106
Expectation	34	Loss	75	Recovery	71	Objection	81
Energy	19	Long Face	71	Benefit	65	Ember	51
Recovery	19	Blow	71	Boom	56	Effectiveness	44
Benefit	17	Burden	67	Feeling of Expectation	38	Dissatisfied	43
Fortunate	16	Confusion	52	Expansion	37	Voice	42
Possibility	14	Effectiveness	50	Sentiment	35	Isolation	42
Interest	14	Congestion	49	Positive	21	Confusion	41
Opportunity	10	Founder	48	Expectation	17	Cliff	36
		Perplexity	47	Interest	14	Congestion	35
		Feeling of Deprivation	46			Exclusion	28
		Shrink	44			Sensitive	26
		Panic	44			Uncertain	26
		Deterioration	42			Chaos	26
		Fear	35			Expected to Plummet	23
		Risk	34			Unclear	23
		Sluggish	33			Hostility	21
		Inequality	29			Excessive	19
		Doubt	26			Anxiety	19
						Pessimism	19
						Unclarity	14
						Unclear	12
						Criticism	12

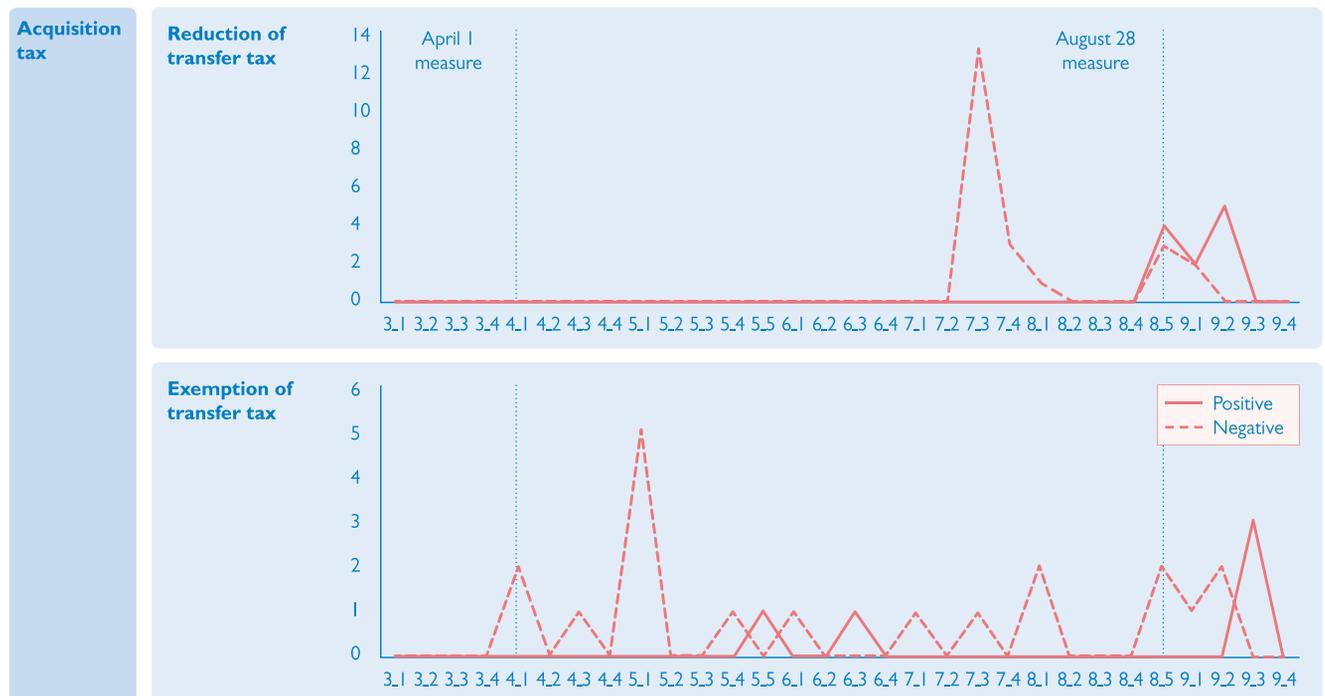
1.2 Analysis of Public Opinion on Acquisition Tax Provisions

The April 1 measure exempted first-time homebuyers from the acquisition tax. Also exempt were couples whose combined income did not exceed KRW 60 million, those purchasing homes worth KRW 600 million or less, and those buying smaller housing units (maximum 85 square meters). Cuts to the acquisition tax were a major part of the August 28 measure as well, but the government had earlier introduced a separate cut of the acquisition tax on a temporary basis, from January 2013 until the end of June the same year.

The sentiment analysis showed the appearance of many negative opinions about the April 1 measure even when it was first announced, and such opinions remained strong when the August 28 measure came out. This can be interpreted as evidence that the easing of the acquisition tax had an insignificant impact on the market.

On the cut in the acquisition tax, the ratio of positive to negative opinions was almost even immediately after the announcement. What is interesting about this is the presence of negative opinions about the reduction of the acquisition tax even before the August 28 measure was announced, dating as far back as July. These opinions were apparently related to the temporary reduction of the acquisition tax from January 2013 to the end of June that year, based both on the word-cloud analysis and our analysis of the original texts. Figure 13 shows the results of our monitoring of public opinions on both the reduction of the acquisition tax and exemptions. Table 9 shows the frequency with which words related to each measure appeared.

Figure 13. Monitoring of Opinions on Exemptions from / Reduction of Acquisition Tax



Our word-cloud analysis of the reduction of the acquisition tax found more comments on the tax cut, which benefited more people, than on the decision to exempt certain first-time homebuyers from the tax. The word cloud shown in Figure 14, representing positive responses, illustrates the level of anticipation that the measure would boost the real estate market. But in Figure 15, which represents negative responses, words such as “nonsense,” “temporary” and “shortage” frequently appeared. This can be interpreted as failure to meet public expectations on the part of the provision within the April 1 measure to reduce the acquisition tax, and the public perception of the provision merely

having a short-term effect. Furthermore, when the April 1 measure was announced, a separate measure existed to lower the acquisition tax and was set to remain in place through the end of June. Around the end of June that year, many pundits said the tax breaks should be extended. The prevalence of such comments can be interpreted as an indication that prospective homebuyers held a positive view of tax breaks for home purchases.

Figures 16 and 17 represent the results of our word-cloud analysis about the acquisition tax exemption, which had limited effects and applied to a limited number of people. Beneficiaries had to meet one of several requirements: they had to be first-time homebuyers, couples with combined annual incomes of up to KRW 60 million or purchasers of homes worth up to KRW 600 million. In particular, potential buyers who could afford to purchase homes were not exempt. Moreover, the acquisition tax exemption benefited those who wanted *jeonse* contracts but could also afford to buy homes, not those who wanted *jeonse* contracts but could not afford to buy. The acquisition tax cut suffered from the same limitation, and a number of related comments appeared in the original texts of collected documents.

The big data analysis found that the acquisition tax cut benefited a greater number of homebuyers than the exemption, and also had positive effects on the real estate market. So the market can be expected to respond in a positive way to the government's announcement on the real lower acquisition tax stipulated in the August 28 measure that it would retroactively apply to all home purchases after August 28, not January 2014 as originally planned. But considering the negative opinions about the acquisition tax rate, which was higher than expected, and the requirement that a newly purchased home cost no more than KRW 600 million, we recommend that the government reconsider both the acquisition tax rate and the eligibility criteria for tax breaks.

Figure 14. Reduction of Acquisition Tax – Positive Responses



Figure 15. Reduction of Acquisition Tax – Negative Responses

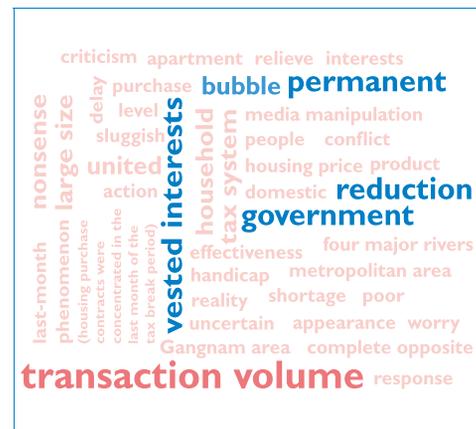


Figure 16. Exemptions from Acquisition Tax – Positive Responses

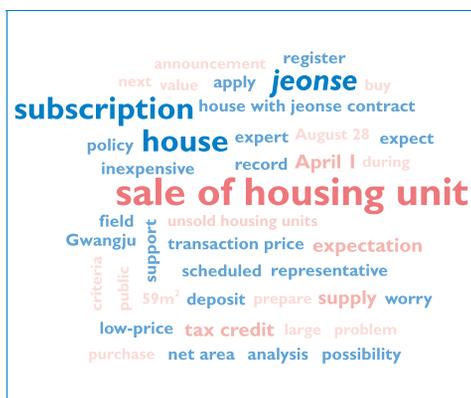


Figure 17. Exemptions from Acquisition Tax – Negative Responses



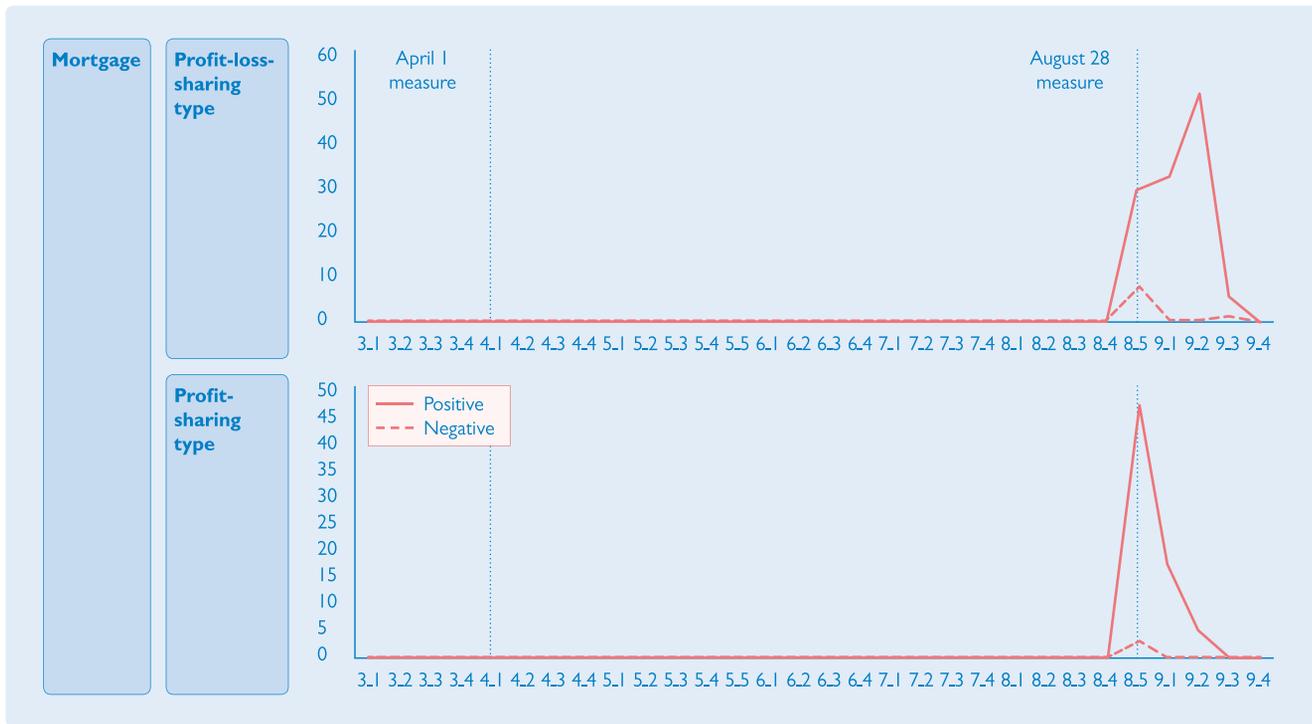
Table 9. Frequency of Words Appearing in Original Text of Responses to Sentiment Analysis on Acquisition Tax

Reduction of Acquisition Tax		Exemption of Acquisition Tax					
Positive Words	Negative Words	Positive Words	Negative Words	Positive Words	Negative Words		
Reinvigoration	27	Sham	45	Interest	4	Bubble	12
Effects	13	Temporary	32	Inexpensive	4	Shackle	12
Positive	9	Stupid	31	Effectiveness	4	Warning	9
Expectation	7	Uncertain	18	Expectation	3	Land Registry Records	9
Advantageous	7	Shortage	17	Success	2	Collapse	8
Benefits	7	Criticism	17			Delay	8
Recovery	7	Complete Opposite	17			Bubble	7
Attractive	6	Worry	16			Room	6
Impressive	6	Nonsense	16			Temporary	6
		Poor	16			Momentary	5
		Fluctuate Between Hopes & Fears	16			Slowdown	4
		Momentary	16			Burden	4
		Explicit	15			Weak	4
		Pussyfoot	15			Gross Real Estate Tax	4
		Loss	15			Shock	4
		Failure	15			Voice	3
		Some Influential Group	15			Complain	3
		Cliff	14			Instigation	3
		Transaction Cliff (Pend-up demand)	9			No Progress	3

1.3 Public Opinion on Mortgage Policy

Mortgage policy was part of the August 28 measure, and our study covered profit-sharing and profit-loss-sharing types separately. Figure 18 shows the results of the sentiment analysis of mortgage provisions in the August 28 measure. Positive public opinions about the policy appeared even before its implementation, and once it was announced, many people said it would greatly help first-time home buyers. Provisions on both types of mortgage, profit-loss-sharing and profit-sharing, were well received, but the former type received more positive responses for a longer period of time. Our review of the original texts found a number of comments suggesting that the August 28 measure was insufficient to handle rising demand for *jeonse* contracts, but that the mortgage policy would have a positive effect on the recovery of the housing market. Negative responses mostly concerned the overall content of the August 28 measure, rather than the mortgage provisions specifically.

Figure 18. Monitoring of Responses to Mortgage Provisions



Figures 19 and 20 represent positive and negative responses to the measure's provisions for profit-loss-sharing mortgages, while Figures 21, 22 show positive and negative responses for the term “profit-sharing mortgage.” Table 10 represents the frequency of words appearing in connection with each mortgage type. The word-cloud analysis found a great deal of interest in both mortgage types and more positive than negative responses to the mortgage policy. The higher level of interest in the profit-loss-sharing type suggests that fears over falling housing prices remained. Looking at the positive public opinions, we found many comments suggesting that the mortgage policy was unprecedented and expectations for recovery of the real estate market were high. The mortgage policy was recognized as a completely new policy and the number of positive responses far outnumbered those to existing policies. Under this policy, loans for home purchases were provided at low interest rates of 1-2 percent, and this was expected to provide real benefits to homebuyers through equity sharing. But a limited number of people could benefit from this policy, meaning the exclusion of numerous people because of conditions such as first-time homebuyer status, annual income requirement and the location and cost of the home being purchased. We believe the benefits should be expanded in light of the issues raised by respondents.

Figure 19. Profit-loss-sharing Mortgage Provisions – Positive Responses

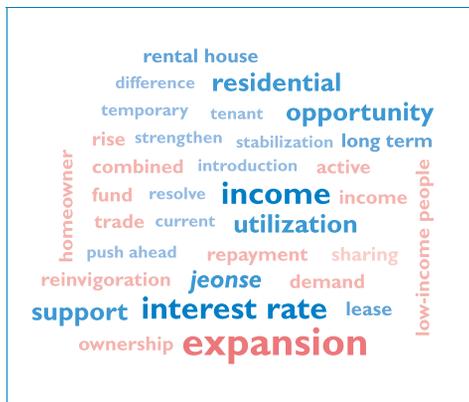


Figure 20. Profit-loss-sharing Mortgage Provisions – Negative Responses



Figure 21. Profit-sharing Mortgage Provisions – Positive Responses



Figure 22. Profit-sharing Mortgage Provisions – Negative Responses

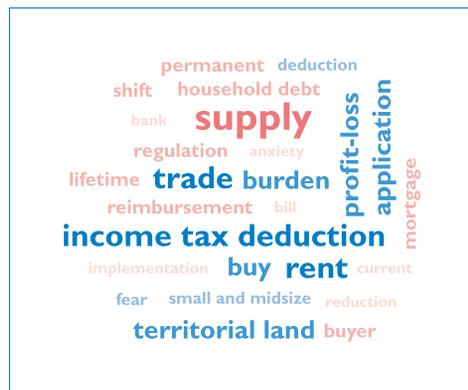


Table 10. Frequency of Words that Appeared in Original Texts on Mortgage Used for Sentiment Analysis

Profit-loss-sharing type				Profit-sharing type			
Positive words		Negative words		Positive words		Negative words	
Improvement	187	Burden	23	Resolve	156	Burden	23
Reinvigoration	144	Limit	17	Possibility	90	Room	16
Profit	141	Problem	13	Stable	87	Anxiety	10
Possibility	135	Loss	12	Possible	86	Fear	9
Stable	132	Damage	9	Effectiveness	76	Loss	8
Resolve	130	Dissatisfaction	8	Reinvigoration	51	Shortage	6
Strong Point	117	Blind Spot	8	Impressive	48	Momentary	5
Effectiveness	93	Fear	6	Isolation	47	Fear	4
Impressive	86	Fear	4	Feeling of Expectation	39	Deterioration	4
Success	59	Unsatisfactory	4	With Low Interest Rate	34	Fall behind	3
Adequate	53	Objection	4	Positive	33	Risk	3
Suitable	53	Crazy	2	Solve	28	Uncertain	3
Expect	45			Recovery	28	Inadequate	3
Feeling of Expectation	39			Strong Point	26	Limit	3
Satisfy	39			Expect	25		
Recovery	37			Help	21		
Inexpensive	23			Suitable	17		
Solve	22			Housing Stability	14		
Positive	21			Protection	9		

The mortgage policy was extremely popular; applications closed just 54 minutes after they opened. As of this writing, 2,275 applicants were selected for the program after undergoing credit and field reviews, and 76 percent of those selected signed loan contracts. Of the selected applicants who did not execute contracts, most gave up for reasons such as dissatisfaction with the terms and conditions or having changed their minds and chosen the loans for first-time home buyers instead. The profit-loss-sharing type accounted for 78.5 percent (1,785) of all contracts executed and the profit-sharing type 21.5 percent (490). The results of the sentiment analysis found greater interest in profit-loss-sharing mortgages, but the number of contracts concluded was higher for profit-sharing mortgages.

2. How Effective Is Real Estate Policy?

2.1 Changes in Housing Transaction Volume

The April 1 and August 28 measures aimed to reinvigorate the real estate market by shifting demand for *jeonse* to that for home purchases, and this study found a few signs of success. Such signs were uncovered through calculating and comparing the annual number of transactions; classifying transactions according to administrative unit (for example, cities, counties, city wards, towns, townships and neighborhoods—in Korean, *si*, *gu*, *gun*, *eup*, *myeon*, and *dong*); tracking changes in transactions and their regional distribution by using a map and marking transaction volumes on the locations of executed transactions; and other methods. This research analyzed data on housing sales transactions in Seoul from 2011 to 2013. As the data collection period in 2013 was from January to November, however, transaction volumes in December 2011 and 2012 were excluded so that researchers could compare the same periods for each year.

“
As the data collection period in 2013 was from January to November, however, transaction volumes in December 2011 and 2012 were excluded so that researchers could compare the same periods for each year.

”

Table 11 represents the volume of housing sales transactions in Seoul. The number of housing sales transactions was 105,117 in 2011, 77,793 in 2012 and 100,419 in 2013. The number of transactions decreased 27,324 (25 percent year on year) in 2012 from 2011. But in 2013, when the April 1 and August 28 measures were announced, the number of transactions increased 22,626 (29.0 percent) from 2012. This rise can be interpreted as evidence that the 2013 measures reinvigorated a sluggish housing market.

Table 11. Annual Volume of Housing Sales Transactions in Seoul

Year	2011	2012	2013
Transaction volume (No. of transactions)	105,117	77,793	100,419
Change (%)	-	-25.9	29

To arrive at a detailed analysis of transaction volume in Seoul, annual transaction volumes were recorded on a map of 25 districts (*gu* in Korean). This detailed graph (Figure 23) shows transaction volume changes for all of Seoul, broken down by district. Transaction volume in each of Seoul's districts declined in 2012 but nearly recovered to 2011 levels in 2013. The city wards of Nowon-gu, Seongbuk-gu and Gangseo-gu saw their transaction volumes recover to 2011 levels. Furthermore, transactions in Seocho-gu, Gangnam-gu, and Songpa-gu returned to 2011 levels.

Figure 23. No. of Housing Sales Transactions in Seoul Districts

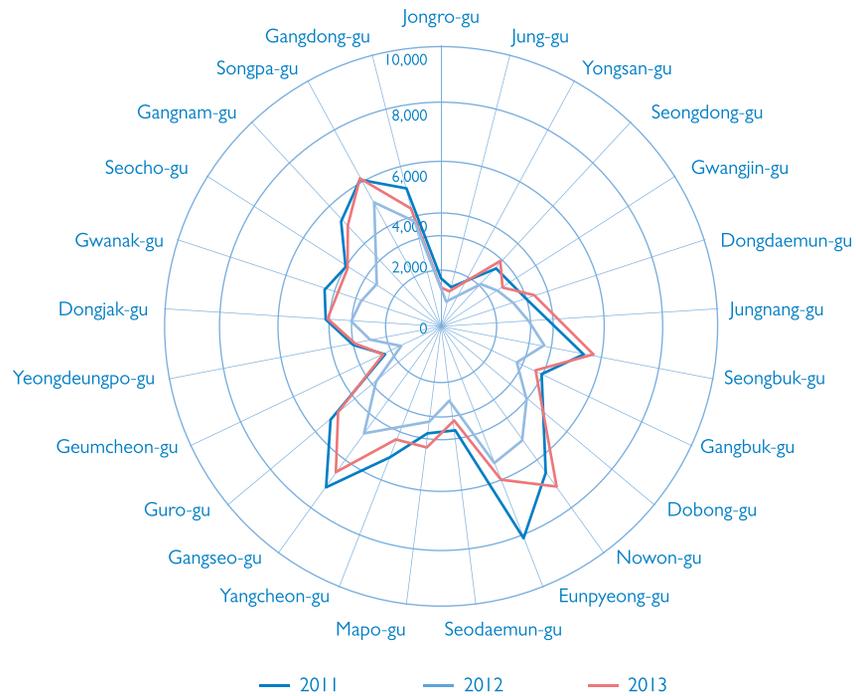
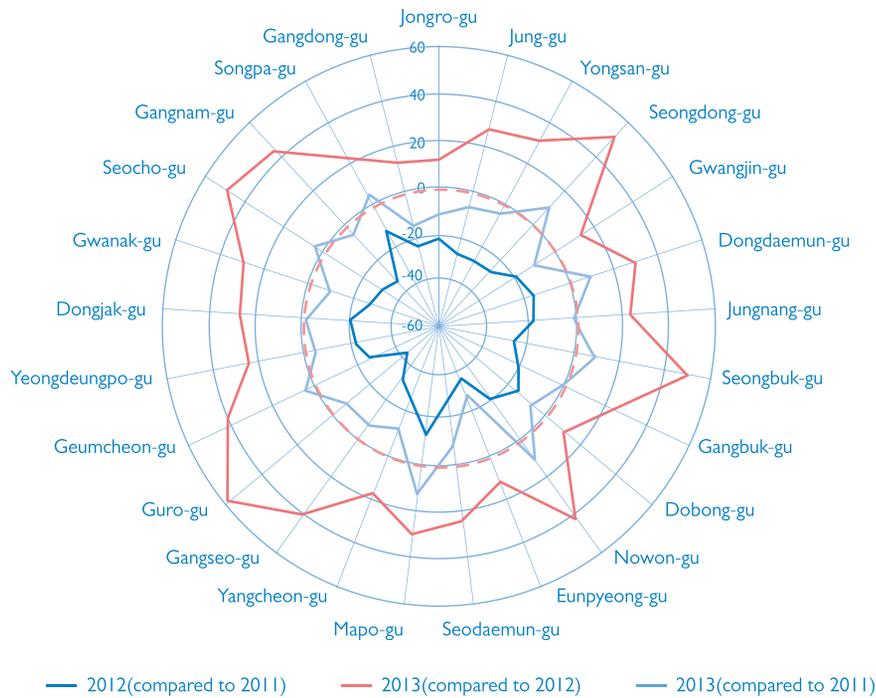


Figure 24 depicts transaction volume rates in each of Seoul’s districts in 2012 and 2013. The 2013 data reflects changes from 2011 and 2012. In 2012, the figures declined in all districts within the city. In particular, transaction volume was down almost 40 percent in Seocho-gu, Gangseo-gu and Eunpyeong-gu. But in 2013, transaction volumes increased year on year in the range of 20–50 percent. Transactions increased in Seongdong-gu, Seongbuk-gu, Nowon-gu, Gangseo-gu, Guro-gu and other parts of the city. Seocho-gu and Gangnam-gu saw especially noticeable increases after seeing significant declines in 2012. Most *gu* (wards) saw transaction volumes recover to 2011 levels.

Figure 24. Changes in Transaction Volume in Seoul Districts



As mentioned above, changes in transaction volume can be identified with relative changes in quantity after annual and regional data are collected, but such changes can be also presented on a map by using address information. The advantage of a map is that regional distribution and quantitative changes in transactions executed can be easily identified. In this research, data on transaction volume was classified per plot of land and year, and changes in 2012 and 2013 transaction volumes were calculated. For 2013, changes in transaction volumes were calculated compared with 2011 as well as 2012. Then kernel density estimation⁰⁵ was used to present changes in housing sales transaction volume on a map. For kernel density estimation, the radius was 1,000 meters and the size of the grid 100 meters.

05. How to make a kernel density map is as follows: First, make a grid for areas to be analyzed. Then to calculate a kernel density value of the first grid, calculate distances to all points within the grid's set radius. Then use a reciprocal number for each distance as a weighted value to multiply it by the attribute value of the related point. Then add all of the resulting values for the first kernel density value. Repeat this process for all the grids to make a kernel density map.

Figure 25 represents the results of a kernel density estimation of transaction volume change in 2012 compared with 2011. The redder the map is, the higher the growth in transaction volume. The bluer it is, the more transaction volume declined. In 2012, most of Seoul's districts experienced a decline in transaction volume. In particular, transactions plummeted in Yangcheon-gu, Eunpyeong-gu, Nowon-gu, Gangdong-gu and Dobong-gu.

A kernel density estimation of changes in 2013 compared with 2012 is depicted in Figure 26. Compared with 2012, the part of the distribution colored in red was larger in all of Seoul in 2013, which means housing sales transactions increased in most areas of the city from 2012. The northern part of the city reflected this trend the most, especially the wards of Seongbuk-gu, Nowon-gu and Dobong-gu. The level of recovery in 2013 can be estimated through comparison of the 2011 and 2013 transaction volumes. For this purpose, transaction volume changes in 2013 from 2011 are shown in Figure 27 using kernel density estimation. As Table 11 shows, the 2013 transaction volume was 95.5 percent of the 2011 level; the kernel density map in Figure 27 also shows that transaction volume increased or remained similar to 2011 levels in most districts. The exceptions were parts of Gangseo-gu and Eunpyeong-gu.

Increased housing sales transactions indicate recovery in the real estate market. So the figures support the conclusion that the April 1 and August 28 measures had positive effects on the housing market. We believe the government's policy intervention helped to revitalize housing sale transactions.

Figure 25. Changes in Transaction Volume in 2012 (vs. 2011)

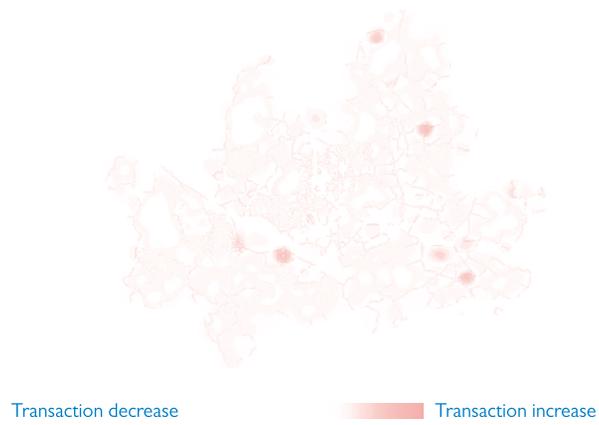


Figure 26. Change in Transaction Volume in 2013 (vs. 2012)

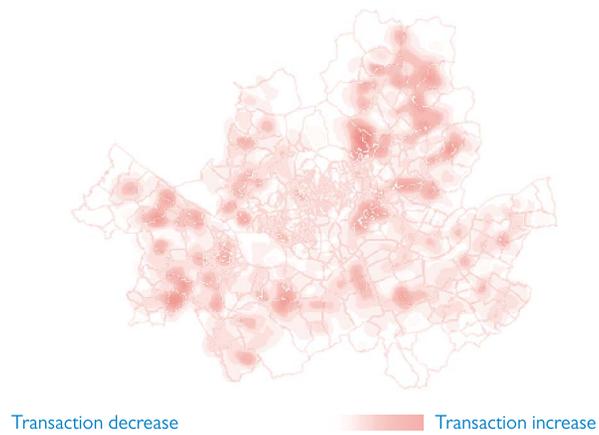
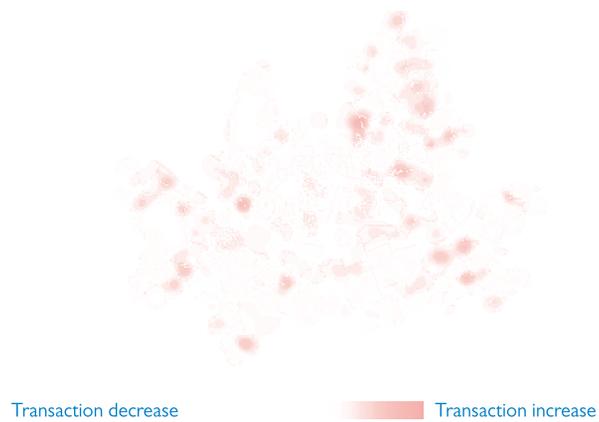


Figure 27. Change in Transaction Volume in 2013 (vs. 2011)



2.2 Changes in Housing Prices

06. The ratio of unit price in the current year to unit price in the previous year on the same house.

To evaluate the success of the policy measures, changes in housing prices should be considered along with the volume of sales transactions. Rising housing prices encourage potential homebuyers to consider buying homes rather than signing *jeonse* contracts, because this provides reassurance that the value of their investments will continue to grow. At the very least, rising prices assuage the fears of those reluctant to purchase homes. This research aimed to determine changes in housing prices as well as transaction volumes after the implementation of the policy. For this purpose, the percentage of annual price changes⁰⁶ for each unit of house was calculated by selecting houses on which contracts were executed annually from 2011 to 2013. Then the kernel density method was used on a map.

Figure 28 is a kernel density map showing percentages of price changes in 2012 from 2011. The redder the map, the higher the percentage of positive price change. The bluer it is, the smaller the percentage of negative price change. The map shows a rise in housing sale prices in select wards including Geumcheon-gu, Gwanak-gu, Gangdong-gu and Jungnang-gu, but prices declined in most other areas. This can be seen as the result of lower demand.

Then what is the percentage of price change in 2013 when transaction volume recovered to the 2011 level? Figure 29 represents the percentage of price change in 2013 from 2012. In 2012, prices decreased in most parts of Seoul, but in 2013, certain areas where prices decreased saw them rise again, albeit slightly. In particular, this occurred in Seocho-gu, Gangnam-gu, Songpa-gu and Gangdong-gu. Transaction volume increases in 2013 did not directly lead to higher prices, but counteracted price declines in most parts of Seoul in 2012.

The percentage of price changes between 2011 and 2013 can be seen in Figure 30, which shows a negative trend in all but a few districts such as Geumcheon-gu and Gangdong-gu. This means housing sale prices were down in 2013 from 2011, though the pace of price decline slowed and certain districts experienced price increases. So changes in transaction volume apparently did not affect housing prices immediately, but a time gap was to be expected.

After the government implemented real estate policy measures in 2013, the volume of housing transactions jumped year on year. And the pace of decline in housing prices also slowed, with a modest increase in certain districts. Though the pace of recovery in housing prices was slower than that of transaction volume, housing prices are expected to rise steadily if existing volume of housing transactions is maintained. In conclusion, the government policy had positive effects on the market in that it prevented housing prices from declining further, but did not spur a strong increase in housing prices.

Figure 28. Percentage of Price Changes in 2012 (vs. 2011)



Figure 29. Percentage of Price Changes in 2013 (vs. 2012)

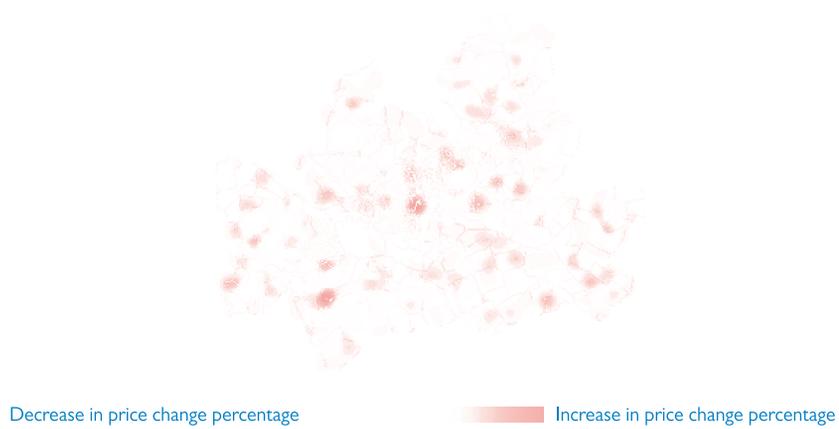
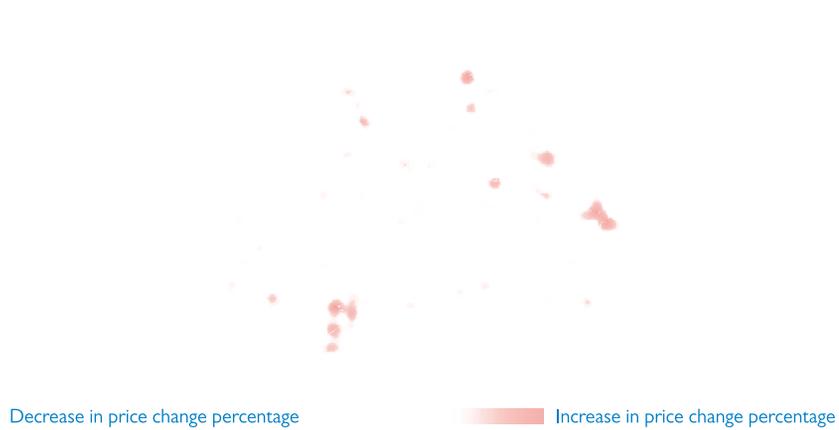


Figure 30. Percentage of Price Changes in 2013 (vs. 2011)



3. Identifying Key Issues in Real Estate Policy

For the development of a policy agenda, big data was divided into datasets for sentiment analysis (analysis of public opinion) and for analysis of related words (development of a policy agenda). The dataset for sentiment analysis consisted of only documents that were considered positive or negative responses. Thus they consisted of certain words closely related to a specified theme, which helped identify relevant issues quickly. But the dataset for the analysis of related words included a wide range of materials, not only documents with positive or negative responses, but also all documents linked to related words. Thus quick identification of issues is difficult due to the large volume of data acquired. But a variety of issues can be found thanks to the wide range of information included.

3.1 Developing a Policy Agenda Using Dataset for Sentiment Analysis

Analysis of “Transfer Tax”

The result of a sentiment analysis in which “transfer tax” was a keyword, conducted alongside an analysis of the original text of the comments, showed that the words “exemption” and “reduction of transfer tax” appeared the most frequently. The term “transfer tax exemption” appeared more frequently, possibly because more people stood to benefit from this than the tax cut, which applied only to owners of multiple homes.

Table 12 shows nouns extracted from the original text of comments used for the sentiment analysis of the transfer tax. The results show that on transfer tax, “imposition of surtax on transfer tax” and “abolition of surtax on transfer tax” appeared continuously from March, when data collection started, to before the August 28 measure was implemented. The abolition of the surtax previously imposed on owners of multiple homes was not included in the April 1 measure, but Koreans frequently mentioned the need for its abolition and the possibility of it being abolished. Politicians also held discussions of the matter.

“

While the real estate market was highly interested in the issue, the government announced the abolition of the surtax as part of the August 28 measure. Negative opinions about this provision said it would only benefit the rich. But the prevailing view within public opinion was that the transfer surtax, which had been introduced to deter real estate speculation, was aggravating the sluggishness of the real estate market. Thus the abolition of the surtax was expected help reinvigorate the real estate market.

”

While the real estate market was highly interested in the issue, the government announced the abolition of the surtax as part of the August 28 measure. Negative opinions about this provision said it would only benefit the rich. But the prevailing view within public opinion was that the transfer surtax, which had been introduced to deter real estate speculation, was aggravating the sluggishness of the real estate market. Thus the abolition of the surtax was expected help reinvigorate the real estate market.

Table 12. Analysis of Transfer Tax

	March		April		May	
Rank	Related words	Freq.	Related words	Freq.	Related words	Freq.
1	Transfer tax	5	Transfer tax	10,644	Transfer tax	4,473
2	Waiver of transfer tax	1	Waiver of transfer tax	200	Waiver of transfer tax	113
3	Abolition of heavy transfer tax	1	Reduction of transfer tax	77	Reduction of transfer tax	56
4	Reduction of transfer tax	1	Benefit from waiver of transfer tax	30	Benefit from transfer tax	13
5			Heavy transfer tax	21	Those who are subject to reduction of transfer tax	13
6			Benefit from transfer tax	17	Benefit from reduction of transfer tax	11
7			Abolition of heavy transfer tax	16	Amount of transfer tax reduction	8
8			Criteria for waiver of transfer tax	14	Waiver of transfer tax on studio apartment	7
9			Criteria for reduction of transfer tax	14	Exempt from transfer tax	5
10			Exempt from transfer tax	12	Those who are subject to waiver of transfer tax	4
11			Benefit from reduction of transfer tax	11	Benefit from waiver of transfer tax	4
12			Transfer tax rate	11	Amount of transfer tax reduction	4
13			Temporary waiver of transfer tax	7	Applicable rate of transfer tax	4
14			Time when Transfer tax is reduced	6	Reduction of transfer tax on studio apartment	2
15			Those who are subject to waiver of transfer tax	6	Temporary waiver of transfer tax	2
16			Questions about transfer tax	5	Transfer tax on equity	2
17			Abolition of transfer tax	4	Conditions to be exempt from transfer tax	2
18			Those who are subject to reduction of transfer tax	3	Imposition of heavy transfer tax	2
19			Areas to which waiver of transfer tax is applied	3	Calculation of transfer tax	1
20			Transfer tax on absentee landlord	2	Areas to which reduction of transfer tax is applied	1
21			Transfer tax on owners of multiple homes	2	Criteria for reduction of transfer tax	1
22			Application of transfer tax reduction	2	Transfer tax on new apartment unit	1
	June		July		August	
Rank	Related words	Freq.	Related words	Freq.	Related words	Freq.
1	Transfer tax	1,911	Transfer tax	968	Transfer tax	781
2	Waiver of transfer tax	44	Waiver of transfer tax	25	Imposition of heavy transfer tax	27
3	Reduction of transfer tax	29	Transfer tax on owners of multiple homes	7	Waiver of transfer tax	12
4	Abolition of heavy transfer tax	11	Benefit from reduction of transfer tax	6	Reduction of transfer tax	10
5	Transfer tax rate	11	Reduction of transfer tax	5	Transfer tax on owners of multiple homes	5
6	Imposition of heavy transfer tax	10	Those who are subject to reduction of transfer tax	4	Heavy transfer tax on owners of multiple homes	3
7	Exempt from transfer tax	6	Total of transfer tax amount	4	Homes subject to reduction of transfer tax	3
8	Benefit from transfer tax	5	Abolition of heavy transfer tax	4	Abolition of heavy transfer tax	6
9	Transfer tax on owners of multiple homes	4	Automatic calculation of transfer tax	4	Exempt from transfer tax	3

10	Benefit from waiver of transfer tax	3	Conditions to be exempt from transfer tax	4	Transfer tax rate	2
11	Benefit from reduction of transfer tax	3	Exempt from transfer tax	3	Heavy transfer tax on owners of multiple homes	1
12	Transfer tax on studio apartment	2	Imposition of heavy transfer tax	2	Consulting on transfer tax	1
13	Reduction of transfer tax, exempt from transfer tax	1	Benefit from transfer tax	2	Real estate subject to heavy transfer tax	1
14	Application of transfer tax reduction	1	Transfer tax rate	1		
15	Period for transfer tax waiver	1	Heavy transfer tax on owners of multiple homes	1		
16	Details of transfer tax reduction	1	Those who are subject to transfer tax	1		
17	Those who are subject to reduction of transfer tax	1	At the time of transfer tax return	1		
18	Related to transfer tax	1	Real rights to transfer tax waiver	1		
19			Period for transfer tax waiver	1		
20			Procedure of transfer tax return	1		
21			Regulations on reduction of transfer tax	1		
22						

September

Rank	Related words	Freq.
1	Transfer tax	738
2	Reduction of transfer tax	9
3	Abolition of heavy transfer tax	5
4	Transfer tax on owners of multiple homes	4
5	Imposition of heavy transfer tax	4
6	Transfer tax apartment	4
7	Imposition of heavy transfer tax	3
8	Abolition of transfer tax	2
9	Act on temporary abolition of heavy transfer tax	2
10	Waiver of transfer tax	2
11	Abolition of transfer tax on owners of multiple homes	1
12	Benefit from waiver of transfer tax	1
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

Analysis of Acquisition Tax

After exemptions to the acquisition tax were announced as part of the April 1 measure, the public showed great interest in the tax, especially in the tax breaks rather than exemptions. Table 13 shows nouns with positive or negative meanings appearing in online documents and classified into positive or negative responses in the sentiment analysis of the acquisition tax.

On the exemptions to the acquisition tax, people were interested in who was eligible for the exemption as well as in the conditions and its validity period. But the level of interest was not high due to the limited scope of eligibility. Furthermore, an acquisition tax cut had been in place since early January 2013 and lasting until the end of June. In May and June, when the lower tax rate was set to end, the words “period” and “extension” were newly mentioned and continued to appear by September. As shown in the result of sentiment analysis, a number of opinions demanding the extension of the tax reduction appeared in Table 13, as the end of the temporary tax reduction had a negative impact on the real estate market.

The government met market demand by announcing the permanent reduction of the acquisition tax as part of the August 28 measure, but this move was scheduled to take effect in January 2014. Because it would not be applied for six months, people purchasing homes between July and December could not benefit. That led to fears that the policy could have a negative impact on the market in the short term. In addition, words related to the acquisition tax such as “acquisition tax rate” and “retrospective application of acquisition tax” started to appear more frequently from July. That meant public opinion wanted the lower rate of acquisition tax to apply to homebuyers purchasing homes over those six months.

On November 4, 2013, the government announced that the reduced tax rate would apply retroactively to people who purchased homes after August 28, 2013. This can be interpreted as the government responding to plummeting transaction volume called a “transaction cliff” in July after the tax break ended at the end of June.⁰⁷ This is a good example of the government implementing a policy in a timely manner to meet market demand. But attention is required to a measure to help those who purchased homes from July 1 to August 28, during which time no tax breaks were available.

07. Yonhap News (July 30, 2013), “Last month, the transaction cliff” phenomena deteriorated whenever acquisition tax was reduced temporarily.”

Table 13. Analysis of Acquisition Tax

March			April		May	
Rank	Related words	Freq.	Related words	Freq.	Related words	Freq.
1	Acquisition Tax	6	Acquisition Tax	7,398	Acquisition Tax	3,654
2			Waiver of acquisition tax	148	Waiver of acquisition tax	75
3			Reduction of acquisition tax	90	Reduction of acquisition tax	64
4			Period for transfer tax reduction	33	Acquisition tax rate	36
5			Acquisition tax rate	17	Benefit from acquisition tax reduction	16
6			Acquisition tax amount	12	Conditions on acquisition tax waiver	7
7			Acquisition tax revenue	6	Burden of acquisition tax imposed	7
8			Benefit from acquisition tax reduction	5	Period for transfer tax reduction	6
9			Notice on acquisition tax reduction	4	Acquisition tax amount	4
10			Comparison of acquisition tax amounts	4	Benefit from acquisition tax	4
11			Acquisition tax on first-time homebuyers	4	Measure to reduce acquisition tax	4
12			Related to acquisition tax	4	Amount of acquisition tax waiver	4
13			Criteria for waiver of acquisition tax	4	Criteria for waiver of acquisition tax	3
14			Criteria for acquisition tax	3	Real estate acquisition tax	3
15			Acquisition tax on housing	3	Acquisition tax on first-time homebuyers	3
16			Areas subject to waiver of acquisition tax	3	Details on acquisition tax reduction	2
17			Benefit from acquisition tax waiver	2	Benefits from acquisition tax reduction	2
18			Acquisition tax table	2	Details on acquisition tax waiver	2
19			Temporary waiver of acquisition tax	2	Amount of reduced acquisition tax	1
20			Conditions on acquisition tax waiver	2	Reduction of acquisition tax on first-time home buyers	1
21			Benefit from acquisition tax	2	Transfer tax acquisition tax	1
22			Applicable acquisition tax	2		
June			July		August	
Rank	Related words	Freq.	Related words	Freq.	Related words	Freq.
1	Acquisition Tax	3,266	Acquisition Tax	3,323	Acquisition Tax	1,521
2	Acquisition tax rate	184	Acquisition tax rate	629	Acquisition tax rate	363
3	Reduction of acquisition tax	132	Waiver of acquisition tax on first-time home buyers	48	Waiver of acquisition tax on first-time homebuyers	15
4	Waiver of acquisition tax	34	Reduction of acquisition tax	38	Permanent reduction of acquisition tax	13
5	Benefit from acquisition tax reduction	18	Acquisition tax on housing	34	End of acquisition tax reduction	11
6	End of acquisition tax reduction	7	End of acquisition tax reduction	21	Reduction of acquisition tax	8
7	Those who are subject to acquisition tax waiver	6	Acquisition tax decrease	12	Permanent reduction of acquisition tax rate	3
8	Period for transfer tax reduction	5	Measure to waiver acquisition tax	9	Waiver of acquisition tax	3
9	Measure to reduce acquisition tax	5	Waiver of acquisition tax	9	Benefit from acquisition tax reduction	2
10	Acquisition tax on housing	4	Permanent reduction of acquisition tax	6	Benefit from acquisition tax waiver	2
11	Extension of acquisition tax reduction	4	Real estate acquisition tax	6	Policy to reduce acquisition tax	1
12	Acquisition tax on first-time homebuyers	3	Measure to reduce acquisition tax	4	Measure to reduce acquisition tax	1
13	Transfer tax acquisition tax	1	Benefit from acquisition tax reduction	3	Acquisition tax registration tax	1

14	Real estate acquisition tax		Regulations on reduction of acquisition tax	2	Change in Acquisition tax rate	
15	Acquisition tax registration tax		Acquisition tax shock	2	Measure to adjust acquisition tax rate	
16	Calculation of acquisition tax registration tax		Acquisition tax revenue	2	Acquisition tax decrease	
17	End of acquisition tax measure		Extension of acquisition tax reduction	2		
18	Those who are subject to acquisition tax reduction		Period for reduction of real estate acquisition tax			
19	Calculation of acquisition tax		Acquisition tax on first-time homebuyers			
20	Official in charge of acquisition tax		Waiver of acquisition tax on low-income people			
21			Transfer tax acquisition tax			
22						
	September					
Rank	Related words					Freq.
1	Acquisition Tax					1,975
2	Acquisition tax rate					377
3	Permanent reduction of acquisition tax					35
4	Reduction of acquisition tax					4
5	Waiver of acquisition tax					4
6	Acquisition tax return and payment					4
7	Acquisition tax decrease					3
8	Standard assessment of acquisition tax					2
9	Retrospective application of acquisition tax					
10	Period for transfer tax reduction					
11	Method to calculate acquisition tax					
12	Waiver of acquisition tax on first-time homebuyers					
13	Measure to decrease acquisition tax					
14	Retrospective application of acquisition tax reduction					
15	Decrease in acquisition tax rate					
16	End of acquisition tax reduction					
17	Acquisition tax retrospective					
18	Application of acquisition tax					
19	Acquisition tax act					
20	Acquisition tax exemption					
21						
22						

3.2 Developing Policy Agenda Using Dataset for Related-word Analysis

To develop a policy, recent issues and other related matters need to be identified. In this part, the frequency of nouns that appeared in online documents collected through keyword searches including “housing,” “reconstruction,” “re-development,” “unsold housing units” and “country house” was analyzed. Unlike the sentiment analysis, the analysis of related words targets a wide range of words, thus a huge volume of documents are collected and a variety of words extracted. The time range was between January 2011 and September 2013 and words were presented in descending order by monthly frequency.

Table 14 shows the frequency of words that appeared in documents collected with “housing” as a keyword. The word “apartment” ranked first, and “Seoul,” “government,” “monthly rent,” “jeonse” and “supply” continued to rank high.

Table 14. Analysis of Words Related to “Housing”

Jan. 2011		Feb. 2011		March 2011		April 2011		May 2011	
Related word	No. of frequency cases	Related word	No. of frequency cases	Related word	No. of frequency cases	Related word	No. of frequency cases	Related word	No. of frequency cases
Apartment	9,804	Apartment	9,375	Apartment	10,124	Apartment	10,107	Apartment	10,207
Housing	7,775	Housing	7,441	Housing	8,024	Housing	6,838	Housing	7,393
Real estate	6,364	Real estate	5,880	Seoul	6,880	Real estate	6,407	Real estate	6,433
Seoul	6,316	Seoul	5,354	Real estate	6,611	Seoul	6,189	Seoul	6,260
Household	5,543	Household	4,917	Household	5,186	Household	4,949	Household	6,008
Increase	3,331	Monthly rent	4,227	Government	4,051	Government	3,475	Region	4,512
Supply	3,199	jeonse	4,098	Monthly rent	3,706	Region	3,250	Government	4,096
Government	3,182	Supply	3,496	Region	3,104	Sale of housing units	3,009	Sale of housing units	3,288
Monthly rent	3,121	Government	3,484	Supply	3,097	Country house	2,827	Korea Land & Housing Corp.	3,254
Region	2,965	Region	2,866	Sale of housing units	3,049	Project	2,729	Seoul metropolitan area	2,792
jeonse	2,929	Land	2,591	Land	2,972	Supply	2,627	Supply	2,674

Figure 31 represents a time-series graph showing the level of interest in reconstruction, redevelopment and unsold housing units. Interest in reconstruction and redevelopment continued from 2011, and people were more interested in reconstruction (led by the private sector) than redevelopment (part of public urban planning). The level of interest peaked in the second half of 2011 but continued to decline after May 2012. In contrast, interest in unsold housing units jumped in September 2012, when the level of interest in reconstruction and redevelopment stopped plummeting. This can be interpreted as a weakening of the reconstruction and redevelopment market and a rise in the number of unsold housing units due to a slowdown in the real estate market. This result shows that the analysis of changes in the level of interest in the housing market can indirectly help estimate the business activities of the real estate market.

Meanwhile, the market for country(sub-urban) houses showed an upward trend, despite the slowdown in the real estate market. A country house refers to a home built in the countryside on farmland or green space with a rural atmosphere. The biggest advantage of country houses is that the land is more affordable than that in Seoul and Gyeonggi Province. In the past, major buyers of country homes were senior citizens in their 60s or 70s who wanted life in the countryside after retirement, but an increasing number of people in their 30s and 40s have recently sought homes in the countryside near Seoul. This is the result of rising real estate prices and difficulties in finding *jeonse* contracts.⁰⁸ As shown in Figure 32, this trend has been apparent since 2011.

08. Korea Business News (September 9, 2013), "For weekend vs. daily residence"...where can I find a country house suitable for me?

Country houses can serve as a solution to the shortage of *jeonse* housing now that the Seoul population has exceeded 10 million. But the lack of education and transportation facilities should be resolved first. Such issues should be taken on gradually in the form of a long-term national plan. Over the short term, subsidies and tax credits for new country houses can be considered and thus provide an alternative for people forced to move to the countryside due to rising *jeonse* demand.

“

Country houses can serve as a solution to the shortage of *jeonse* housing now that the Seoul population has exceeded 10 million. But the lack of education and transportation facilities should be resolved first. Such issues should be taken on gradually in the form of a long-term national plan. Over the short term, subsidies and tax credits for new country houses can be considered and thus provide an alternative for people forced to move to the countryside due to rising *jeonse* demand.

”

Figure 31. Level of Interest in Reconstruction, Redevelopment & Unsold Housing Units

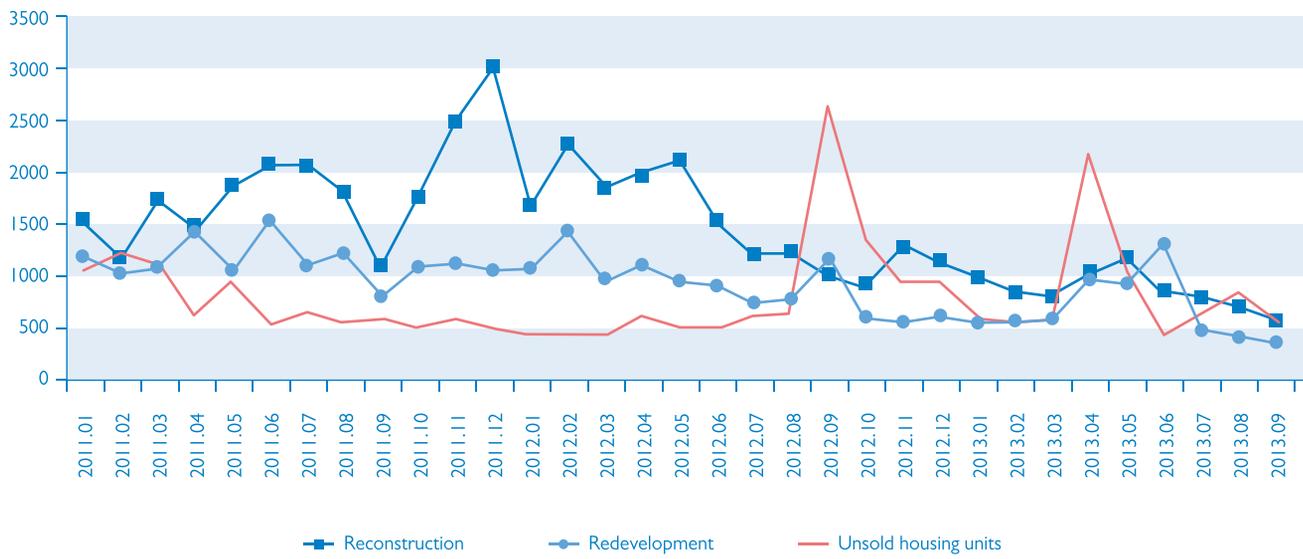
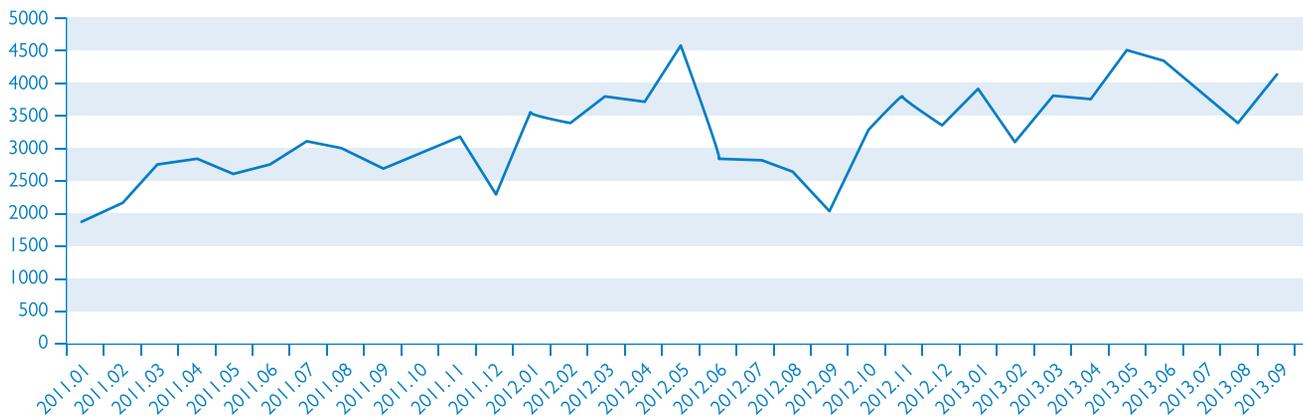


Figure 32. Level of Interest in Country Houses



Conclusion

I. Utilization of Big Data and Its Implications

To understand how big data can be used in the real estate policymaking process, this author conducted a case study on the April 1 and August 28 measures and examined the possibility of analyzing public opinion on policy through sentiment analysis of big data, evaluation of policy effectiveness by using real estate transaction data and development of policy agenda through related-word analysis.

Sentiment analysis based on big data can be helpful in monitoring how people respond to government policies and identify relevant issues. In particular, big data analysis can help identify positive or negative public opinions on the April 1 and August 28 measures in a timely fashion, and raise precise understanding of the reasons for such positive or negative opinions through detailed analysis of the full text of each comment.

To analyze policy effectiveness, the volume of real estate transactions and prices were analyzed through time-series statistical analysis visualized on geographical space. This process featured the cross-analysis of big data, a leading indicator, and real estate transaction data, a lagging indicator. Together, these forms of data shed light on the impact the two measures had on the market.

Lastly, the analysis of related words showed similar results to the sentiment analysis and helped to identify the extent of public interest in certain words or issues. In addition, it enabled time-series analysis of overlooked or newly emerging issues that can be used to inform policy.

This research showed the limited impact of the taxation policy of the April 1 and August 28 measures, such as the cut in the acquisition tax and exemptions to the transfer tax, which aimed to shift demand for *jeonse* to that for purchased homes to alleviate the supply-demand imbalance. Due to low interest rates, the number of *jeonse* contracts converted to monthly rent keeps rising, and shifting demand for *jeonse* to that for purchased homes is difficult considering the slowdown in the real estate market. The lack of stability in *jeonse* prices means that the government's policy suffered from limitations. But positive public opinion about the government's pilot mortgage project highlighted the necessity of new real estate policy, also showing that vouchers and expansion of rental homes are needed to ensure stable housing for low-income people.

2. Limitations of Big Data Utilization

Big data used for real estate policy basically consists of unstructured data collected through web crawling. But the quality of data collected through this method varies, depending on the crawling engine and the capabilities of engineers. The same applies to sentiment analysis. Interpretation by analysts affects how they determine positive or negative responses (database), and the results of the analysis depends on the database's quality. For example, big data collected for the case study showed that public opinion was positive about the supply of rental housing in May. Yet a review of the full text of responses classified as positive showed that they included news articles that had merely reported the details of the April 1 measure in a neutral manner. Furthermore, the same documents were collected several times, meaning the techniques used to collect and process big data can often determine the results.

Another problem with utilizing big data is that a sentiment analysis is not objective enough to shed significant light on a phenomenon. The sentiment analysis used in the case study showed that the public was more interested in and positive about the transfer tax than the acquisition tax. But an analysis of policy effectiveness based on real estate transaction data found that the acquisition tax played a more important role in home purchasing decisions than the transfer tax. This means the results of a big data analysis can differ from findings in the real world, depending on the criteria and methods used.

Big data provide a huge volume of data for the policy-making process, giving better insights. Meanwhile, the case study found potential issues related to the collection of big data and limitations when big data are the only source. Thus if big data are to be used in the policymaking process, such issues and limitations need to be considered in the process of data collection and analysis. In addition, when measures are prepared and determined, expertise and public participation are needed to resolve the limitations of big data.

3. Recommendations

To further utilize big data in the real estate policymaking process, several issues must be resolved. First, the big data and spatial big data platforms the government plans to set up should be completed as soon as possible. The private sector in Korea and developed nations utilizes big data in several ways, but the public sector still lays the foundation for the utilization of big data. This is why it takes so much time to provide scientific and transparent administration and customized policy. Thus a platform where various types of big data can be utilized easily is needed as soon as possible. And a big data platform should be connected to that of spatial big data.

Second, a system is needed to utilize big data. Big data produced in the public sector, in particular, should be made freely accessible. The private sector should be allowed access to structured big data produced by information systems of the government and log information and clickstream data accumulated when using the systems. The development of a variety of applications based on such free access will contribute to the transparency of policies. What should be considered for free access to big data is technology to process personal information that could exist in big data, so that sensitive information cannot be leaked and justified reasons for free access.

Third, a collaborative system is needed to use big data in the private sector in the real estate policymaking process. Portals in the private sector accumulate a variety of big data crucial to policy, such as search words and clickstream data. But companies are reluctant to share such data for proprietary reasons. The high prices makes the purchase and use of such data difficult. For the public purposes of policymaking and national research, a collaborative system that enables better utilization of big data in the private sector is needed.

Fourth, technologies to process big data should be developed. For this study, an outside service provider was hired to collect and process big data in the private sector such as Twitter. The problem was that the service provider feared the potential leak of confidential information, and provided no details on how it collects big data and performs sentiment analysis to determine positive and negative responses to a policy. In particular, government investment is required to develop technologies to identify slight differences between Korean words. Meanwhile, a long time is required to analyze the spatiotemporal patterns of a huge volume of spatial big data by using existing technologies, which could result in untimely analysis of big data collected in real time. Thus SpatialHadoop needs further development to rapidly analyze spatial big data.

Last but not least, data scientists who can extract useful information from big data should be nurtured. The advancement of ICT and personal ICT devices will lead to a hyperconnected society, and the volume of big data will continue to grow. Thus the success of policy depends on capabilities to analyze big data. For this, data scientists who can better utilize big data are needed so that the government can develop, decide on and implement effective policies in a timely manner. In particular, data scientists who can use spatial big data are urgently required for real estate policy, so the government needs to train experts who have the capabilities to find and interpret patterns in a spatial context and integrate big data and spatial data.

References

Gartner. *Hype Cycle for Emerging Technologies*. 2011.

Hahmann, Burghardt. "How Much Information Is Geospatially Referenced? Networks and Cognition." *International Journal of Geographical Information Science* 27.6 (2013).

Kim, Daejong and Hyungsu Gu. *Land Use Change Prediction with Spatiotemporal Pattern Analysis and Strategies for Urban Policy*. Korea Research Institute for Human Settlements, 2011. [in Korean]

Kim, Daejong, Myunghwa Hwang, Seoyeon Yoon, and Tae-sung Suh. *A Study on Building and Utilization of Spatial Knowledge Platform for Scientific Territorial Planning (I)*. Korea Research Institute for Human Settlements, 2014. [in Korean]

Lee, Manjai. "Big Data Analytics and Utilization of Public Data." *Communications of the Korean Institute of Information Scientists and Engineers* 30.6 (2012). [in Korean]

National Information Society Agency. *Big Data Era for New Future*. 2012. [in Korean]

National Information Society Agency. *Utilization of Big Data and Direction of Policy*. 2013. [in Korean]

O'Reilly Radar Team. *Planning for Big Data*. 2012

INDEX

Figures

Figure 1. Relationship between Big Data & Spatial Big Data	11
Figure 2. Time-series Jeonse Prices in Seoul Neighborhoods	13
Figure 3. Twitter Search Results on April 1 Measure	13
Figure 4. Big Data Analysis Method to Assess Effectiveness of Real Estate Policy	17
Figure 5. Data Classification for Sentiment Analysis	21
Figure 6. Example of Data Extraction via Text Mining	22
Figure 7. Visualization of House Hunters Moving to San Francisco	24
Figure 8. Monitoring of Responses to Transfer Tax Breaks & Exemptions	26
Figure 9. Exemption from Transfer Tax – Positive Responses	27
Figure 10. Exemption from Transfer Tax – Negative Responses	27
Figure 11. Reduction of Transfer Tax – Positive Responses	27
Figure 12. Reduction of Transfer Tax – Negative Responses	27
Figure 13. Monitoring of Opinions on Exemption from / Reduction of Acquisition Tax	29
Figure 14. Reduction of Acquisition Tax – Positive Responses	30
Figure 15. Reduction of Acquisition Tax – Negative Responses	30
Figure 16. Exemptions from Acquisition Tax – Positive Responses	31
Figure 17. Exemptions from Acquisition Tax – Negative Responses	31
Figure 18. Monitoring of Responses to Mortgage Provisions	32
Figure 19. Profit-loss-sharing Mortgage Provisions – Positive Responses	33
Figure 20. Profit-loss-sharing Mortgage Provisions – Negative Responses	33
Figure 21. Profit-Sharing Mortgage Provisions – Positive Responses	33
Figure 22. Profit-Sharing Mortgage Provisions – Negative Responses	33
Figure 23. No. of Housing Sales Transactions in Seoul Districts	36
Figure 24. Changes in Transaction Volume in Seoul Districts	37
Figure 25. Changes in Transaction Volume in 2012 (vs. 2011)	39
Figure 26. Change in Transaction Volume in 2013 (vs. 2012)	39
Figure 27. Change in Transaction Volume in 2013 (vs. 2011)	39
Figure 28. Percentage of Price Changes in 2012 (vs. 2011)	41
Figure 29. Percentage of Price Changes in 2013 (vs. 2012)	41
Figure 30. Percentage of Price Changes in 2013 (vs. 2011)	41
Figure 31. Level of Interest in Reconstruction, Redevelopment & Unsold Housing Units	50
Figure 32. Level of Interest in Country Houses	50

Tables

Table 1.	Major Points of April 1 Measure	16
Table 2.	Highlights of August 28 Measure	16
Table 3.	Information Collected for Public Big Data	18
Table 4.	Information Collected for Private Big Data	19
Table 5.	Details of Documents Collected	19
Table 6.	Data Cleansing to Calculate Transaction Volume	20
Table 7.	Data Cleansing to Calculate Percentage of Price Change	20
Table 8.	Frequency of Words Appearing in Original Text of Responses to Sentiment Analysis on Transfer Tax	28
Table 9.	Frequency of Words Appearing in Original Text of Responses to Sentiment Analysis on Acquisition Tax	31
Table 10.	Frequency of Words that Appeared in Original Texts on Mortgage Used for Sentiment Analysis	34
Table 11.	Annual Volume of Housing Sales Transactions in Seoul	35
Table 12.	Analysis of Transfer Tax	43
Table 13.	Analysis of Acquisition Tax	46
Table 14.	Analysis of Words Related to "Housing"	48

THE PROMISE OF BIG DATA IN REAL ESTATE POLICY: IDENTIFYING PROBLEMS AND WORKABLE SOLUTIONS IN REAL TIME

Edited by the Global Development Partnership Center (GDPC)
Publisher Ilho Chung
Publication Registration 25100-1994-2
Printed on December 6, 2016
ISBN 979-11-5898-153-2

Published by Korea Research Institute for Human Settlements (KRIHS)
254 Simin-daero, Dongan-gu, Anyang-si, Gyeonggi-do, 14067, Korea.
Tel. +82-31-380-0114 / Fax. +82-31-380-0470
<http://www.krihs.re.kr/>
<http://gdpc.kr/>

Designed and Produced by Designintro Co., Ltd.
<http://www.gointro.com>
Tel. +82-2-2285-0789

Copyright 2016 © Korea Research Institute for Human Settlements
All Rights Reserved.
Printed and Bound in the Republic of Korea
First Published in 2016

* The views expressed in this report are those of KRIHS
and do not necessarily reflect the government's official position.