

目录

1	Introduce, Opinions on Data Science	1
1.1	What is data science?	1
1.2	Why data science?	2
1.3	What do data scientists do?	3
1.4	What can data science do?	3
1.5	New technologies for data science	4
1.6	Future development of data science	6
2	Brief Introduction to Two Data Sets and Views on Public Data	7
2.1	Chicago criminal record from 2001 to the present	7
2.2	Leading Causes of Death in United States	9
2.3	The advantages of open data publicly	10
2.4	The disadvantages of open data publicly	11
2.5	Suggestion to government on open data policies	11
3	Research on second-hand housing in Beijing from 2011 to 2017	12
3.1	Foreword	12
3.2	Introduction to the dataset used	12
3.3	Price overview	14
3.4	Building Structure	16
3.5	Predict to house price	18

1 Introduce, Opinions on Data Science

1.1 What is data science?

Many definitions have been made to data science since it's such a young subject of science. Simply from its name "Data Science", data science is a subject that about data's science or research data, tying to research data as a "Natural Object". So, data science is defined as the theory, method and technique to research data.

- Explore and research data itself, research its various types, status, attributes, change forms and changing patterns.
- Offer a new method called data method for scientific research to the research of natural science and social science, aimed to reveal and to explain nature and human behavior phenomena and its pattern.

There's a common misunderstanding here: that is, many people think that data science is only used to study data itself, not as a research method or a tool to discover new knowledge. The premise of data science to become a science is that the relationship and patterns found in a certain field have universal applicability that can be extended to other fields, not just a tool for some specific fields. Data science is all about how we take data, use it to get knowledge, and then use the knowledge to do the following things [2]:

- Make decisions
- Predict the future
- Understand the past / future
- Create new industry / products

So, this is my favorite definition to data science: Data science is the art and science of acquiring new knowledge through data.

1.2 Why data science?

At present time, there's no doubt that we have an amazingly huge amount of data. And here comes the "Big Data". Though its definition is vague, but there are four widely accepted features: Volume, Velocity, Variety and Value [3]. In addition to the value, the remaining three characteristics determined that big data is difficult to work with [4].:

- Data volume can be up to PB, EB or even ZB sized, for example, dozens of data can be made by a midsized city's surveillance cameras.
- The reality puts forward higher requirements for the processing speed of data, that is, the real-time requirement of collecting data, and the

need to integrate data into business processes and decision-making processes, all this reflected the difference from traditional data processing.

- The type and structure of data is completed. in the past, the data we made or process is mostly structured, but now more and more new technology is producing many unstructured data such as XML, blog and instant messages [5].

Because of these, big data is usually difficult to be used directly and efficiently, and the very main goal of data science is to develop clear practices and procedures for faster and more efficient use of big data.

1.3 What do data scientists do?

It's widely said that data scientists are the sexiest job of the 21st century. As a possible future career, it is necessary for us to understand what exactly data scientists do. A complete data science project requires many people to work together, and data scientists are an important part of it. They are responsible for performing the steps necessary to succeed in the project, including designing project steps, selecting data sources and tools used, and be responsible for project planning. On the technical side, they need to be good in statistics and machine learning, to examine data, perform statistical test and processing; use machine learning models, and evaluate results. These are also the daily tasks of a data scientist.

1.4 What can data science do?

As we mentioned earlier, data science can describe, generalize, abstract data, and make predictions for the real world to help people making decisions. But specifically, what exactly can it do? There are some examples I found.

1.4.1 Public opinion forecast for Obama campaign

The Obama administration is an active promoter of the big data and data science boom. In November 2012, Obama defeated rival Romney and

successfully re-elected the president. One of the secrets of his successful re-election is the data analysis team he calls “core code.” The key to win is to get support from more voters. His team has stored a large amount of voter information, and they use this information to dig out propaganda strategies for different voter groups to make promotion more efficient. Different from the past, the Obama team abandoned the strategy of deciding the advertising based on the opinions of media consultants, but built his advertising strategy on a large database they own. The result of advertising in Chicago showed that TV advertising efficiency increased by 14% [6]. Obviously, Obama’s winning reflects the advance of data analysis, because the storage, analysis and processing power of big data is the basis of macro decision-making and the key to gaining information in advance.

1.4.2 Intelligent Transportation in Los Angeles, USA

With big data and data analysis, transportation system can become more efficient and smarter. By the analysis of real-time monitoring data, an intelligent transportation system can achieve real-time analysis and prediction of traffic conditions, traffic path planning and many other aspects. In Los Angeles, for example, since the 1989 Olympics, the government has built an automatic traffic monitor and control center. By using data analysis, they developed an adaptive traffic control system that automatically adjusts the signal time based on the traffic volume, reducing vehicle waiting time. Other than this, combined with a large amount of data, the system can increase the average traffic speed by 12% for the intersections in Los Angeles and reduce the waiting time by 30% [7]. Through these examples, we can clearly see that big data combined with data analysis can affect many aspects of society, make government activities more efficient, make society more convenient, etc. Data science is making our lives easier. Better.

1.5 New technologies for data science

Big data brings opportunities and challenges to data science. Traditional data processing methods have been unable to meet the needs of big data. In recent years, the rapid development of a new generation of tech-

nologies, such as bath towel computing, Internet of Things, mobile Internet, etc., has provided greater opportunities and development conditions for data collection, storage and processing.

1.5.1 Cloud Computing

Today, data scientists can use many open source machine learning frameworks such as R, SciKit Learn, Spark MLlib, TensorFlow, MXnet, and CNTK. However, managing the configuration and environment of these frameworks on a laptop or local server can be bothering and can take a lot of unnecessary time. In the age of cloud computing, that unnecessary time consume will be greatly reduced. The usage-based pricing model for cloud computing is very effective for machine learning. Cloud computing also makes it easier to explore different machine learning frameworks, with cloud computing sellers offering model hosting and deployment options. In addition, public cloud computing service providers offer smart capabilities as a service, and this reduces the barriers to integrating these features into new products or applications.

1.5.2 Machine Learning

The core of timely and automated processing of data is often algorithms. With the breakthrough development of deep learning technology, machine learning algorithms can be developed at a high speed. In short, the intelligence of big data processing must be machine learning and deep learning. In recent years, machine learning algorithms are having a rapid development, and various networks and algorithms have emerged one after another. For example, BERT, vid2vid, graph_nets. The BERT paper published by Google AI in October has aroused strong concern in the deep learning community. This paper presents a deep two-way Transformer model. BERT has refreshed the best performance of 11 NLP tasks, including the Stanford Q&A data set (SQuAD). And NVIDIA's Video-to-Video (vid2vid) synthesis paper introduces an amazingly productive model that has been one of the most popular deep learning areas of the past few years.

1.6 Future development of data science

In the age of big data, the size, flexibility, and the ability to collect and use information that a company have will determine their core competencies. Data will become the key asset of the company. Owning the data is owning the market, which means a huge return on investment. Companies with data and data science thinking mode are creating new business models. The following are some of my personal view on the development of data science in the age of big data:

1.6.1 The application of data science will move from the Internet to traditional industry applications

With the increase of big data and the spread of data thinking, more and more industries and companies will realize that data and information will become assets and resources, and data analysis and processing capabilities will bring bigger benefits to companies. And with the advancement of enterprise information and the popularize of Internet service models, traditional industries will become more and more accessible to and apply big data and data analysis [8].

1.6.2 Data services will be more industrialized

With the popularize of big data and the Internet, the demand for data analysis in various industries is getting bigger, and there will be higher requirements for data analysis [9]. As a result, more companies will turn to data services for the industry. Third-party data services will become a direct way of profit, such as selling raw data, selling analyzed results, or collecting, processing, analyzing, and delivering value to form the data chain to help cooperative company to form core competitiveness. More and more companies are turning to such data services.

2 Brief Introduction to Two Data Sets and Views on Public Data

2.1 Chicago criminal record from 2001 to the present

This dataset reflects reported incidents of crime that occurred in the City of Chicago from 2001 to present. There are samples from data:

```
library(knitr)
data_crime <- read.csv("../crimes.csv")
kable(data_crime[1:5,1:3],caption="Chicago criminal record")
```

表 1: Chicago criminal record

ID	Case.Number	Date
11690881	JC267643	05/17/2019 11:50:00 PM
11690822	JC267651	05/17/2019 11:45:00 PM
11690835	JC267648	05/17/2019 11:44:00 PM
11690776	JC267644	05/17/2019 11:41:00 PM
11690957	JC267647	05/17/2019 11:40:00 PM

```
kable(data_crime[1:5,4:7])
```

Block	IUCR	Primary.Type	Description
064XX S HONORE ST	460	BATTERY	SIMPLE
081XX S CENTRAL PARK AVE	486	BATTERY	DOMESTIC BATTERY SIMPLE
016XX S CENTRAL PARK AVE	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN
047XX W JACKSON BLVD	5111	OTHER OFFENSE	GUN OFFENDER: ANNUAL
073XX S STEWART AVE	1310	CRIMINAL DAMAGE	TO PROPERTY

2 BRIEF INTRODUCTION TO TWO DATA SETS AND VIEWS ON PUBLIC DATA8

```
kable(data_crime[1:5,8:14])
```

Location.Description	Arrest	Domestic	Beat	District	Ward	Community.Area
STREET	FALSE	FALSE	726	7	15	67
VEHICLE NON-COMMERCIAL	TRUE	TRUE	834	8	18	70
ALLEY	TRUE	FALSE	1021	10	24	29
STREET	TRUE	FALSE	1131	11	28	25
APARTMENT	FALSE	FALSE	732	7	6	68

```
kable(data_crime[1:5,15:20])
```

FBI.Code	X.Coordinate	Y.Coordinate	Year	Updated.On	Latitude
08B	1165117	1862070	2019	05/24/2019 04:15:50 PM	41.77712
08B	1153877	1850223	2019	05/24/2019 04:15:50 PM	41.74484
15	1152642	1891572	2019	05/24/2019 04:15:50 PM	41.85833
26	1144751	1898269	2019	05/24/2019 04:15:50 PM	41.87686
14	1174898	1856180	2019	05/24/2019 04:15:50 PM	41.76074

```
kable(data_crime[1:5,21:24])
```

Longitude	Location	Historical.Wards.2003.2015	Zip.Codes
-87.67023	(41.777118516, -87.670233634)	44	22257
-87.71175	(41.744838886, -87.711753447)	6	4300
-87.71519	(41.858330876, -87.715189131)	36	21569
-87.74399	(41.876860534, -87.743985641)	11	22216
-87.63455	(41.760743124, -87.63455219)	17	21559

```
kable(data_crime[1:5,25:29])
```

Community.Areas	Census.Tracts	Wards	Boundaries...ZIP.Codes
65	281	3	23
69	199	30	8
30	800	14	57
26	69	23	32
66	511	32	11

This data set was collected and published by the government of city. The Chicago police conducted extensive analysis of these data, for example, to explore patterns.

2.2 Leading Causes of Death in United States

This dataset presents the age-adjusted death rates for the 10 leading causes of death in the United States beginning in 1999. Data are based on information from all resident death certificates filed in the 50 states and the District of Columbia using demographic and medical characteristics.

```
data_death <- read.csv("../death.csv")
kable(data_death[1:5,1:2],caption="Leading Causes of Death in United States")
```

表 7: Leading Causes of Death in United States

Year	X113.Cause.Name
2012	Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)
2016	Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)
2013	Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)
2000	Intentional self-harm (suicide) (*U03,X60-X84,Y87.0)
2014	Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)

```
kable(data_death[1:5,3:6])
```

Cause.Name	State	Deaths	Age.adjusted.Death.Rate
Kidney disease	Vermont	21	2.6

Cause.Name	State	Deaths	Age.adjusted.Death.Rate
Kidney disease	Vermont	30	3.7
Kidney disease	Vermont	30	3.8
Suicide	District of Columbia	23	3.8
Kidney disease	Arizona	325	4.1

2.3 The advantages of open data publicly

Open data is valuable information that is free and easy to access without any restrictions. Many governments have recognized that the benefits of more open and transparent data. Open government data has the potential to accelerate major changes in urban operations and residents' lives. Here are the benefits of open data.

2.3.1 Increase the transparency of government's work

Open data means that the public can keep track of and monitor the operations of the local government. The public nature of open data is responsible for the government's consequences. Public can know exactly what the government has done and what has not been done. Unfinished work or unfulfilled goals will be also published and the public will be able to monitor these tasks. Conversely, meeting or exceeding goals will help build greater, more trusting relationships with local residents.

2.3.2 Promote progress and innovation

Open data provides new opportunities for business applications and build the foundation for new technological innovation and economic growth. A third party that does not have the resources to collect these data for itself will be able to use that information to develop new applications and services. The information provided in this way is also important for academic, public and industry-based research communities. Open data greatly increases the value of information so that it can spread and reach its full potential.

2.3.3 Better store information for a long time

Public data can be accessed in a single, fixed location, and the public will have the ability to spread and store data, which will greatly assist in the long-term storage of data.

2.4 The disadvantages of open data publicly

2.4.1 Inaccurate data can cause big problems

Inaccurate public data can cause big problems because the application that uses them may give false information. Some media found that an application in Australia issued a forest fire alarm that did not exist based on big data [10]. The data came from public emergency calls but was not verified before it was released. This means that application users will receive all possible fire alarms, but it can also cause unnecessary panic.

2.4.2 Public data violation of personal privacy

Most of the public data is desensitized, which means that citizens' private information will be erased. For example, Netflix's film rental dataset for algorithmic competitions desensitizes the user's identity, but there are still users who find their personal information in the data. In addition, the handling of sensitive information is not always taken into account, and the mistakes of government departments may also result in the disclosure of personal information without being processed.

2.5 Suggestion to government on open data policies

- Government departments should improve the level of care of their work to protect personal information.
- Relevant laws and regulatory systems should be detailed and improved to make sure data can be stored safely.
- Hold a positive and optimistic attitude towards public data, reduce the phenomenon of "data islands" and make the data truly show their value.

3 Research on second-hand housing in Beijing from 2011 to 2017

3.1 Foreword

In recent years, housing prices have become an important issue for Chinese people. How does Beijing's housing prices change? What factors will affect house prices? How can I predict the price of second-hand housing? Next we'll try to present the information behind the data through graphs and data analysis.

3.2 Introduction to the dataset used

The dataset includes URL, ID, Lng, Lat, CommunityID, TradeTime, DOM(days on market), Followers, Total price, Price, Square, Living Room, number of Drawing room, Kitchen and Bathroom, Building Type, Construction time, renovation condition, building structure, Ladder ratio(which is the proportion between number of residents on the same floor and number of elevator of ladder. It describes how many ladders a resident have on average), elevator, Property rights for five years (It's related to China restricted purchase of houses policy), Subway, District, Community average price.

```
library(knitr)
data_house <- read.csv("../house.csv", fileEncoding="GBK")
data_house <- cbind(data_house, seq(1:nrow(data_house)))
kable(data_house[1:5,2:9], caption="Samples from whole table")
```

表 9: Samples from whole table

id	Lng	Lat	Cid	tradeTime	DOM	followers	totalPrice
1.0109E+11	116.4211	40.06830	1.11103e+12	2016/8/21	435	33	109.0
1.01091E+11	116.4211	40.06830	1.11103e+12	2016/7/20	286	32	132.5
1.01091E+11	116.4810	39.94474	1.11103e+12	2017/2/26	467	96	1600.0
1.01092E+11	116.3153	39.98667	1.11103e+12	2016/8/27	209	23	363.0

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201713

id	Lng	Lat	Cid	tradeTime	DOM	followers	totalPrice
1.01092E+11	116.3866	40.08650	1.11103e+12	2016/6/27	123	60	530.0

```
kable(data_house[1:5,10:17])
```

price	square	livingRoom	drawingRoom	kitchen	bathRoom	floor	buildingType
19758	55.17	1	0	0	1	低 10	3
21722	61.00	1	0	1	1	低 10	3
61539	260.00	4	2	1	2	低 10	3
84695	42.86	1	1	1	2	低 10	3
35100	151.00	4	2	1	2	低 10	4

```
kable(data_house[1:5,18:22])
```

constructionTime	renovationCondition	buildingStructure	ladderRatio	elevator
2008	4	6	0.167	1
2008	4	6	0.174	1
2008	4	6	1.000	1
未知	4	6	0.087	1
2009	4	6	0.500	1

```
kable(data_house[1:5,23:26])
```

fiveYearsProperty	subway	district	communityAverage
0	1	6	24224
0	1	6	24224
0	1	7	63185
1	1	8	96798
1	0	6	47574

Most data is traded in 2011-2017, some of them is traded in Jan,2018, and some is even earlier(2010,2009)

All the data was fetching from <https://bj.lianjia.com/chengjiao>. Full dataset can be downloaded from <https://www.kaggle.com/ruiqurm/lianjia/downloads/lianjia.zip/2>

3.3 Price overview

First, let's look at the overall change in second-hand housing prices in Beijing. To show the main part, we exclude the transaction information before 2011.

```
library("magrittr")
library("tidyverse")
#Load pipeline operation package magrittr and drawing package ggplot

data_house$tradeTime <- data_house$tradeTime %>%
  as.character() %>%
  as.Date()
#Convert to date format

data_house$district <- as.factor(data_house$district)

data_house_f <- data_house[data_house$tradeTime>as.Date("2010-01-01"),]

ggplot(data_house_f)+ 
  geom_smooth(aes(tradeTime,price),color="steelblue")+
  labs(x = "Trade time", y = "Price")+
  geom_smooth(aes(tradeTime,price,color=data_house_f$district))+ 
  scale_color_discrete(name="District",labels=c("Dong cheng","Feng tai","Yi zhuang","Da xing","Fang shan","Chang ping","Cai qing"))

dst_name <- c("Dong cheng","Feng tai","Yi zhuang","Da xing","Fang shan","Chang ping","Cai qing")
```

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201715

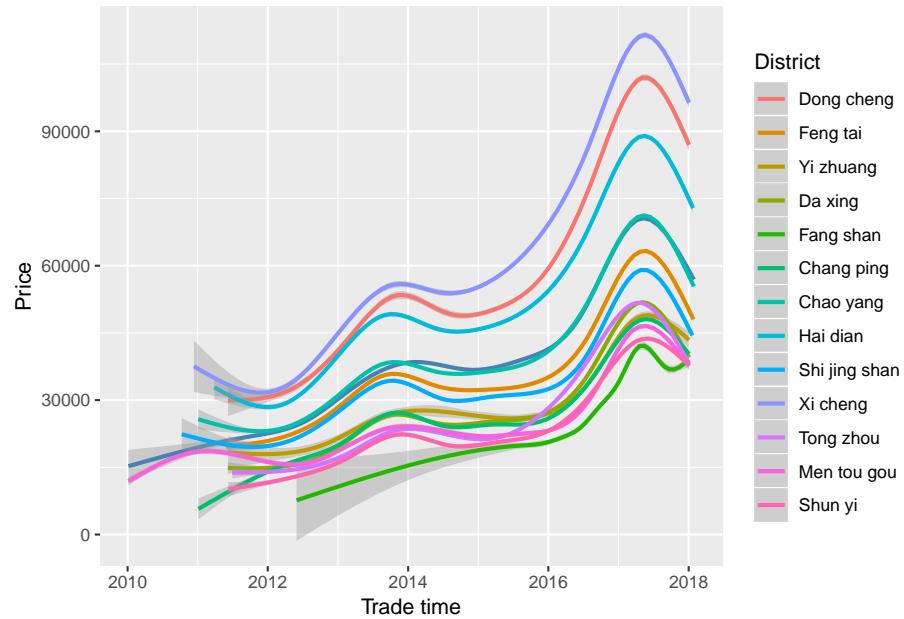
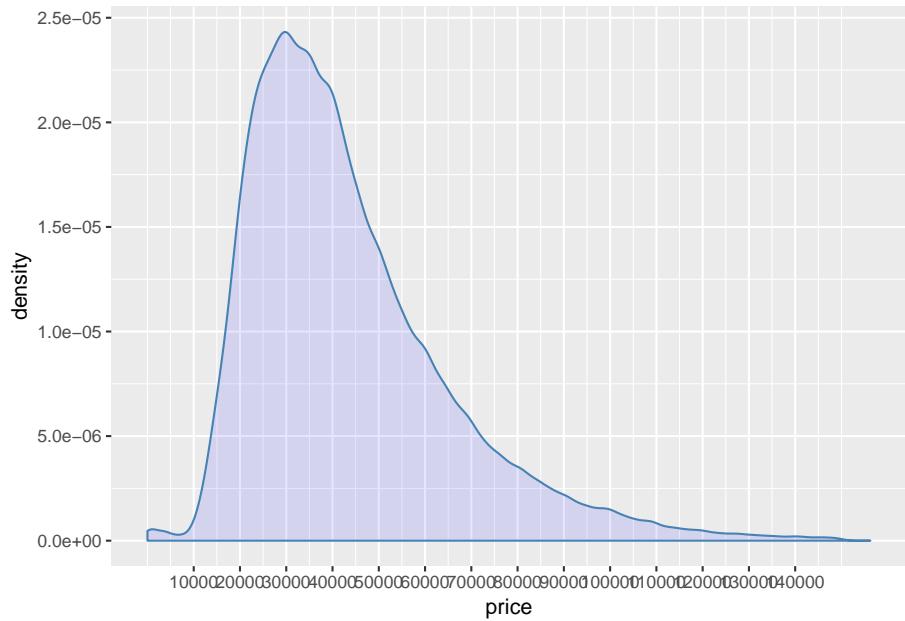


图 1: Beijing second-hand housing prices with time curve

```
ggplot(data_house_f,aes(price))+geom_density(color="steelblue",fill="blue",alpha=0.1)+
```



We can see:

- The price of Chaoyang District is not much different from the average house price in Beijing, which can represent the price of Beijing to a certain extent.
- Among them, the core Haidian District, Dongcheng District and Xicheng District have far higher housing prices and the average house price in Beijing.

The house prices will be displayed in the form of a heat map, and we will see these features more clearly. Here we use the remapH in the **REmap** package to draw the heat map.

```
library("REmap")
points_map <- data.frame(data_house$Lng,data_house$Lat,data_house$communityAverage)
plot(remapH(points_map, maptype="北京", blurSize=5))
```

3.4 Building Structure

Next, let's take a look at what kind of house is the most

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201717

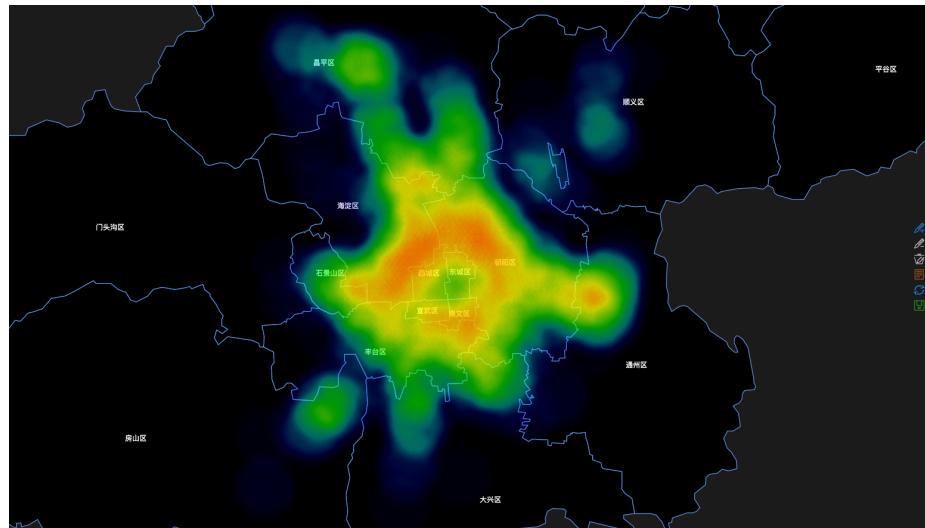
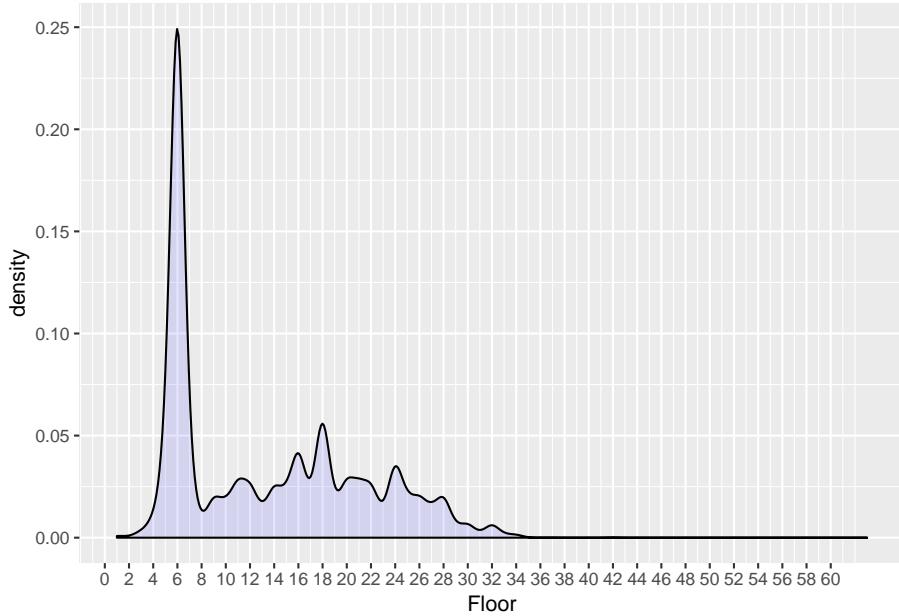


图 2: “Heat map of house price of Beijing”

```
library("stringr")

data_house$floor <- as.character(data_house$floor)
Floor <- substr(data_house$floor,3,nchar(data_house$floor))
Floor <- as.numeric(Floor)
Floor <- data.frame(Floor)
data_house$Floor <- Floor
ggplot(Floor)+  
  geom_density(aes(Floor),fill="blue",alpha=0.1)+  
  scale_x_continuous(breaks=seq(0,60,2))
```



We can see that there are more than 6 floors in the second-hand housing. This may be because the state has stipulated that more than 7 floors need to be equipped with elevators.

3.5 Predict to house price

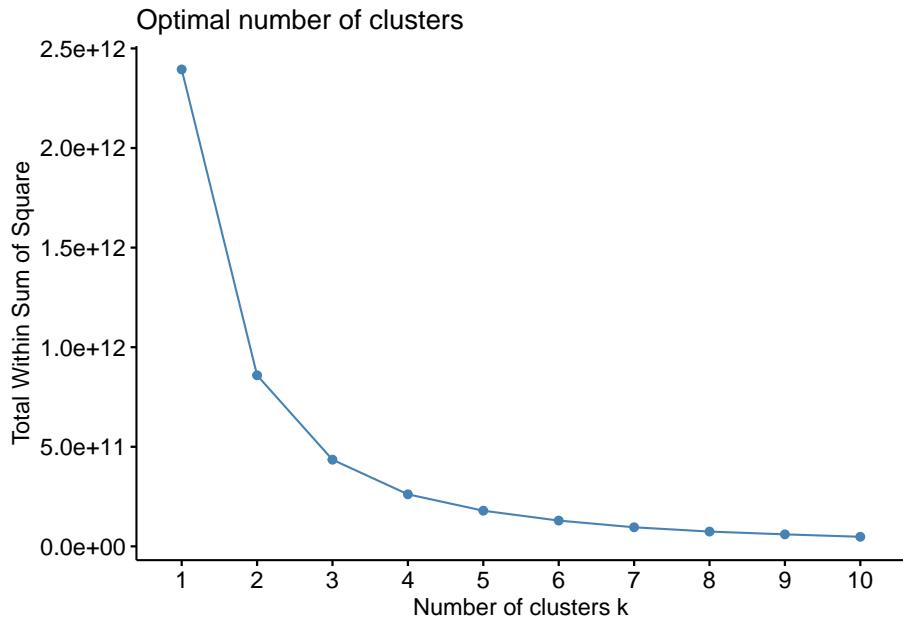
Next we try to model the house price First cluster the data.

```
library(factoextra)

## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl

library(sampling)
d <- data_house[sample(nrow(data_house), 5000), ]
d_sam <- data.frame(d$totalPrice,d$price,d$square)
fviz_nbclust(d_sam, kmeans, method = "wss") + geom_line()

## geom_path: Each group consists of only one observation. Do you need to
## adjust the group aesthetic?
```



```
data_house$type <- paste(data_house$livingRoom,data_house$drawingRoom,sep="室")
```

It can be seen that it is more appropriate to classify all houses into 4 categories.

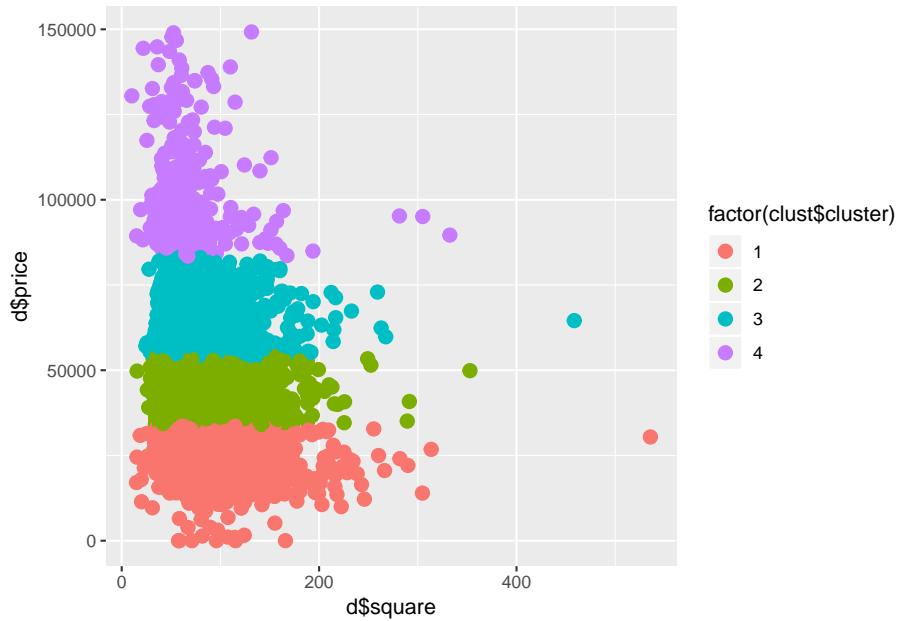
Then we use k-means to classify these houses.

```
set.seed(1111)
clust <- kmeans(x = d_sam, centers = 4, iter.max = 100)

#Classify houses by category

ggplot(d,aes(x = d$square,y = d$price, color = factor(clust$cluster)))+ geom_point(size=
```

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201720



```
table(d$district,clust$cluster)
```

```
##  
##      1   2   3   4  
## 1  22 109  96  43  
## 2 220 185  64   2  
## 3  24  12   0   0  
## 4 166  49  12   0  
## 5  25  13   0   0  
## 6 418 155  17   0  
## 7 531 766 345  37  
## 8  79 260 199  67  
## 9  95  51  14   0  
## 10 45 138 187 159  
## 11 182  41  10   1  
## 12 17   5   0   0  
## 13 117  18   4   0
```

Next we try to establish a linear regression equation. First, let's take

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201721

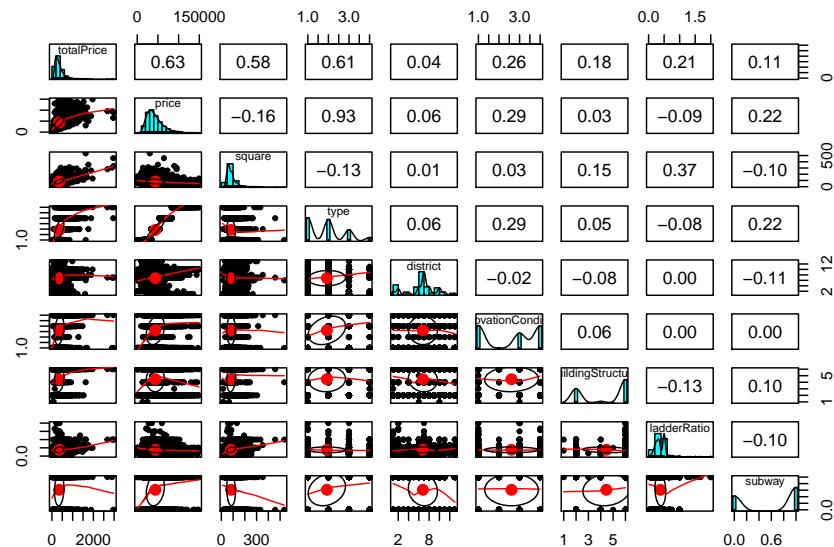
a look at what factors have a greater impact on house prices.

```
library(psych)

## 
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
## 
##     %+%, alpha

d$type <- as.factor(clust$cluster)
pairs.panels(d[c("totalPrice", "price", "square", "type", "district", "renovationCondition",
```



We can see that the house area, the ladder ratio, the building structure, the renovation condition, the number of subways and the type of housing previously clustered on the house have a greater impact on the price. So we choose these factors to fit a multivariate linear model:

3 RESEARCH ON SECOND-HAND HOUSING IN BEIJING FROM 2011 TO 201722

```
model <- lm(d$price~d$square+d$type+d$ladderRatio+d$buildingStructure+d$renovationCondition
summary(model)

## 
## Call:
## lm(formula = d$price ~ d$square + d$type + d$ladderRatio + d$buildingStructure +
##      d$renovationCondition + d$subway)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -28083 -5098   -169   4765  48262 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            24358.416   443.281  54.950 < 2e-16 ***
## d$square                 -18.880    3.031  -6.230 5.05e-10 ***
## d$type2                  16843.993   244.152  68.990 < 2e-16 ***
## d$type3                  38788.650   300.359 129.141 < 2e-16 ***
## d$type4                  74566.432   452.733 164.703 < 2e-16 ***
## d$ladderRatio             -671.417   640.014  -1.049  0.2942    
## d$buildingStructure       106.644    55.480   1.922  0.0546 .  
## d$renovationCondition    732.421    81.480   8.989 < 2e-16 ***
## d$subway                  1412.401   214.225   6.593 4.75e-11 ***
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 7111 on 4991 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.8944 
## F-statistic:  5295 on 8 and 4991 DF,  p-value: < 2.2e-16
```

From the results of the fitting, this model can fit and predict the house price very well.

参考文献

- [1] 赵勇, 林辉, 沈寓实等. 大数据革命-理论、模式与技术创新 [M]. 北京: 电子工业出版社, 2014: 9
- [2] Sinan Ozdemir. Principles of Data Science[M]. 南京: 东南大学出版社, 2017: 4
- [3] 赵勇, 林辉, 沈寓实等. 大数据革命-理论、模式与技术创新 [M]. 北京: 电子工业出版社, 2014: 3
- [4] 赵勇, 林辉, 沈寓实等. 大数据革命-理论、模式与技术创新 [M]. 北京: 电子工业出版社, 2014: 34
- [5] Sinan Ozdemir. Principles of Data Science[M]. 南京: 东南大学出版社, 2017: 5
- [6] 汤姆. 奥巴马连任背后: 政界大数据时代即将到来 [OL].
<http://tech.qq.com/a/20121108/000182.htm> 2012.11.8
- [7] 流雨. 洛杉矶智慧交通 [OL]. <https://www.scitycase.com/>
- [8] 黄鑫. 大数据如何影响传统行业 [OL].
<https://cj.sina.com.cn/article/detail/6176827668/312244.html> 2017.7.8
- [9] 爱分析 ifenxi. 数据服务产业链出现, 数据应用机会最大 [OL].
<https://36kr.com/p/5061834.html> 2017.1.2
- [10] Stefaan Verhulst. The downside of Open Data[OL].
<https://thelivinglib.org/the-downside-of-open-data/> 2015.1.22