

1st-assignment

郑艾轩

2024-10-29

#读文件

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
```

```
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
```

```
## ✓ forcats   1.0.0      ✓ stringr    1.5.1
```

```
## ✓ ggplot2    3.5.1      ✓ tibble     3.2.1
```

```
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.1
```

```
## ✓ purrr     1.0.2
```

```
## — Conflicts — tidyverse_conflicts() —
```

```
## ✖ dplyr::filter() masks stats::filter()
```

```
## ✖ dplyr::lag() masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
shuju <- read.csv("C:/Users/DELL/Desktop/2023_09_12_cleaned(1).csv")
```

```
str(shuju)
```

```
## 'data.frame': 3000 obs. of 18 variables:
```

```
## $ property_name : chr "南湖名都A区" "万科紫悦湾" "东立国际" "新都汇" ...
```

```
## $ property_region : chr "南湖沃尔玛" "光谷东" "二七" "光谷广场" ...
```

```
## $ price_ttl : num 237 127 75 188 182 ...
```

```
## $ price_sqm : int 18709 14613 15968 15702 17509 10376 12346 16336 32631 17403 ...
```

```
## $ bedrooms : int 3 3 1 3 3 3 2 3 4 3 ...
```

```
## $ livingrooms : int 1 2 1 2 2 2 1 2 1 2 ...
```

```
## $ building_area : num 126.7 86.9 47 119.7 104 ...
```

```
## $ directions1 : chr "南" "南" "南" "北" ...
```

```
## $ directions2 : chr "北" "" "" "东" ...
```

```
## $ decoration : chr "精装" "精装" "简装" "精装" ...
```

```
## $ property_t_height: int 17 28 18 32 34 34 7 34 5 7 ...
```

```
## $ property_height : chr "中" "中" "低" "高" ...
```

```
## $ property_style : chr "塔楼" "板楼" "塔楼" "塔楼" ...
```

```
## $ followers : int 3 1 3 2 3 1 0 0 2 0 ...
```

```
## $ near_subway : chr "近地铁" NA "近地铁" "近地铁" ...
```

```
## $ if_2y : chr NA "房本满两年" NA "房本满两年" ...
```

```
## $ has_key      : chr "随时看房" "随时看房" "随时看房" "随时看房" ...
## $ vr           : chr NA "VR 看装修" NA NA ...
```

```
names(shuju)
```

```
## [1] "property_name"      "property_region"    "price_ttl"
## [4] "price_sqm"          "bedrooms"           "livingrooms"
## [7] "building_area"      "directions1"        "directions2"
## [10] "decoration"         "property_t_height"  "property_height"
## [13] "property_style"     "followers"          "near_subway"
## [16] "if_2y"              "has_key"            "vr"
```

```
summary(shuju)
```

```
## property_name      property_region      price_ttl      price_sqm
## Length:3000      Length:3000      Min.   : 10.6  Min.   : 1771
## Class :character  Class :character  1st Qu.: 95.0  1st Qu.:10799
## Mode  :character  Mode  :character  Median : 137.0 Median :14404
##
##                      Mean   : 155.9  Mean   :15148
##                      3rd Qu.: 188.0  3rd Qu.:18211
##                      Max.    :1380.0  Max.    :44656
## bedrooms      livingrooms  building_area  directions1
## Min.   :1.000  Min.   :0.000  Min.   : 22.77  Length:3000
## 1st Qu.:2.000  1st Qu.:1.000  1st Qu.: 84.92  Class :character
## Median :3.000  Median :2.000  Median : 95.55  Mode  :character
## Mean   :2.695  Mean   :1.709  Mean   :100.87
## 3rd Qu.:3.000  3rd Qu.:2.000  3rd Qu.:117.68
## Max.   :7.000  Max.   :4.000  Max.   :588.66
## directions2      decoration      property_t_height property_heig
ht
## Length:3000      Length:3000      Min.   : 2.00  Length:3000
## Class :character  Class :character  1st Qu.:11.00  Class :chara
cter
## Mode  :character  Mode  :character  Median :27.00  Mode  :charac
ter
##                      Mean   :24.22
##
##                      3rd Qu.:33.00
##
##                      Max.    :62.00
## property_style      followers      near_subway      if_2y
## Length:3000      Min.   : 0.000  Length:3000      Length:3000
## Class :character  1st Qu.: 1.000  Class :character  Class :chara
```

```

cter
## Mode :character Median : 3.000 Mode :character Mode :charac
ter
## Mean : 6.614

## 3rd Qu.: 6.000

## Max. :262.000

## has_key vr
## Length:3000 Length:3000
## Class :character Class :character
## Mode :character Mode :character
##
##
##

```

主要发现

#1.该网站大部分二手房都可以随时看房

#2.大部分二手房房产证都满了 2 年

#3.大部分二手房都有 1-2 个客厅

#数据分析 1:

#价格特点: 分析 price_sqm

```
mean(shuju$price_sqm)
```

```
## [1] 15148.49
```

```
median(shuju$price_sqm)
```

```
## [1] 14404
```

```
range(shuju$price_sqm)
```

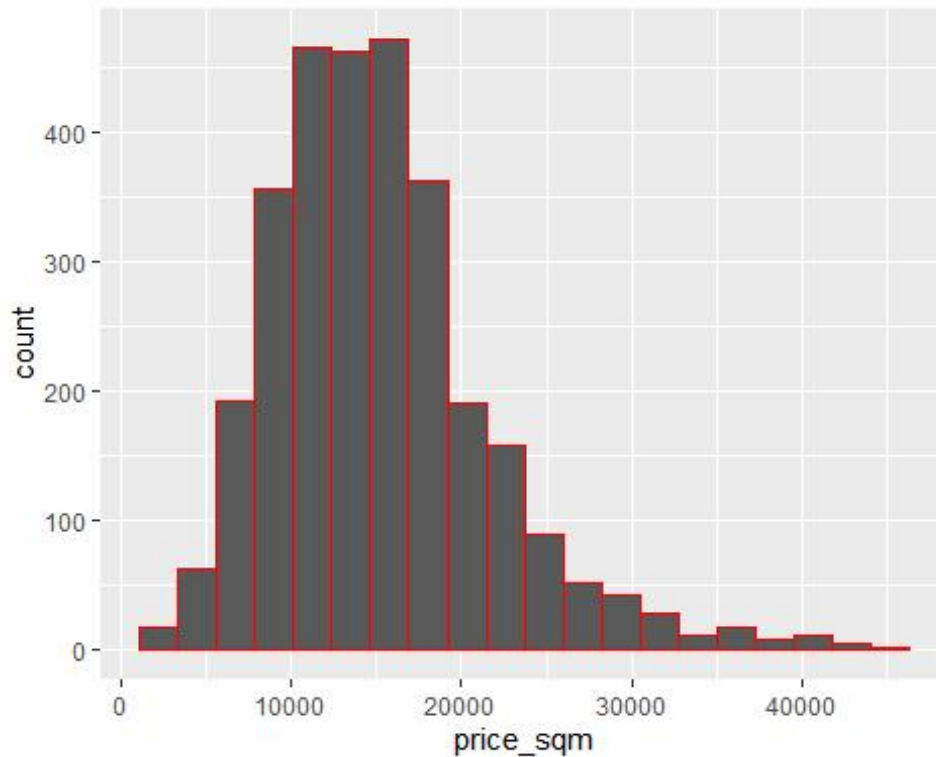
```
## [1] 1771 44656
```

```
sd(shuju$price_sqm)
```

```
## [1] 6323.175
```

#price_sqm 直方图

```
ggplot(shuju) +
  geom_histogram(aes(price_sqm),bins = 20,color="red")
```



#每平米价格 (*price_sqm*) 的均值为15148.49，中位数为14404，标准差为6323.18，离散程度较大。最小值为 1771.00，最大值为 44656.00，说明房价的波动范围较大

#数据分析2

#价格特点: 分析*price_ttl*

```
mean(shuju$price_ttl)
```

```
## [1] 155.8628
```

```
median(shuju$price_ttl)
```

```
## [1] 137
```

```
range(shuju$price_ttl)
```

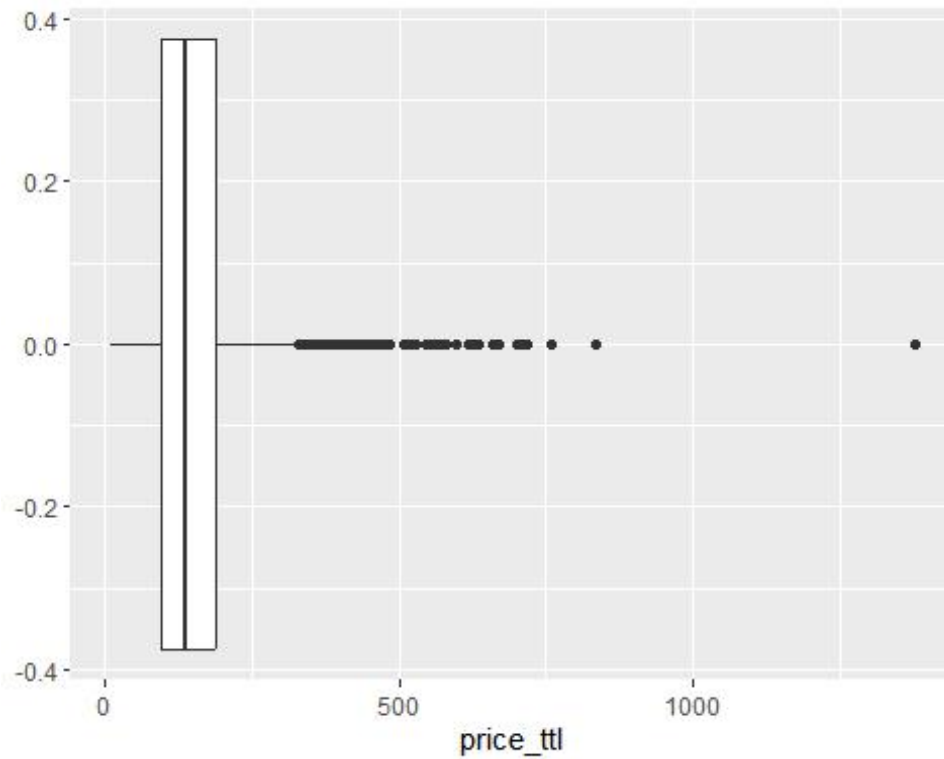
```
## [1] 10.6 1380.0
```

```
sd(shuju$price_ttl)
```

```
## [1] 95.54813
```

#*price_sqm* 箱线图

```
ggplot(shuju) +  
  geom_boxplot(aes(price_ttl))
```



#总价 (price_ttl) 的均值为 155.86, 中位数为 137, 标准差为 95.55, 说明总价分布较为分散。最小值为 10.60, 最大值为 1380.00, 表明房价存在较大的差异。

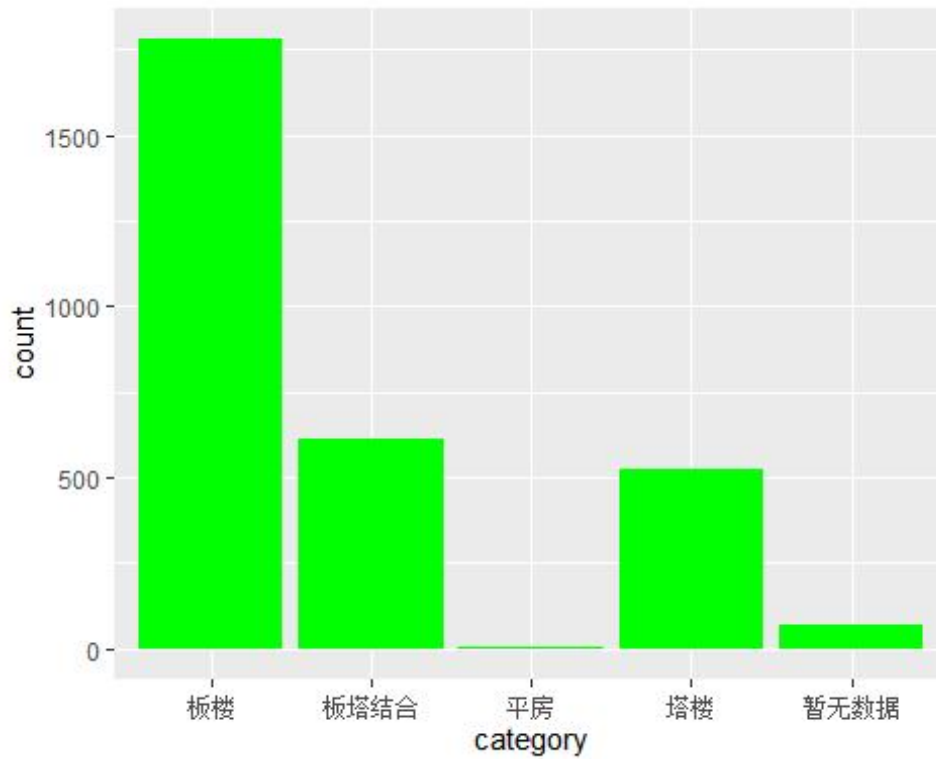
#数据分析3

#建筑风格进行分析

```
jianzhu_style<-shuju %>%
  group_by(property_style) %>%
  summarise(count = n()) %>%
  arrange(desc(count))
```

#建筑风格直方图

```
data <- data.frame(category = c('板楼', '板塔结合', '塔楼', '暂无数据', '平房'),
                    count = c(1781, 615, 527, 72, 5))
ggplot(data, aes(x = category, y = count)) +
  geom_bar(stat = "identity", fill = "green")
```

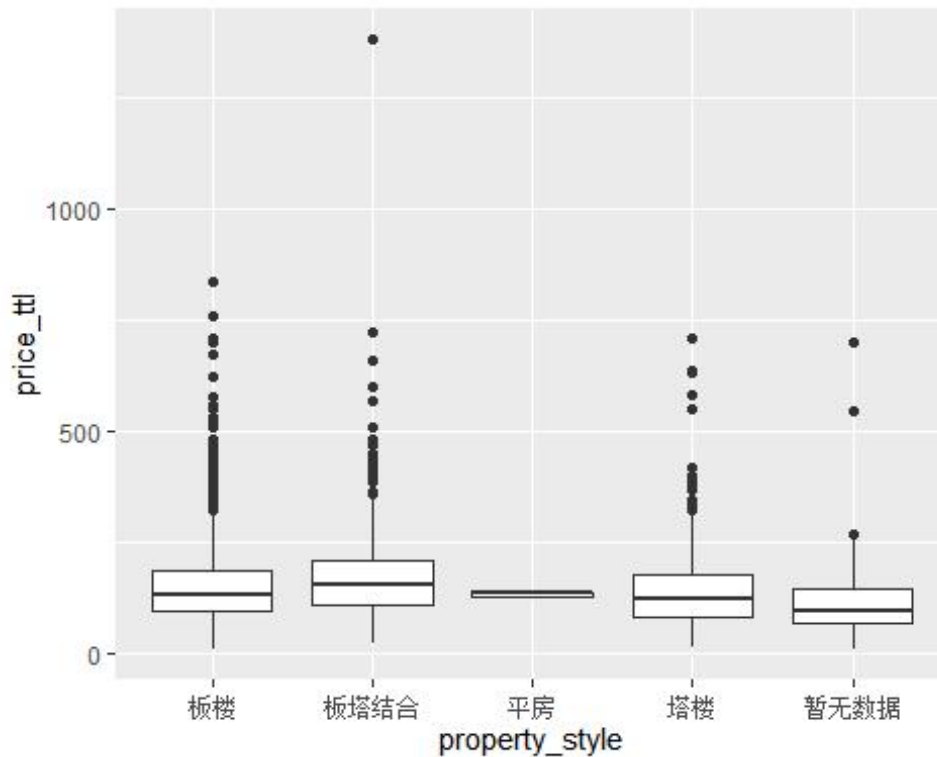


#结论：板楼的数量最多，为 1781 套；其次是板塔结合，有 615 套；塔楼有 527 套；暂无数据的有 72 套；平房最少，只有 5 套

探索问题1

#总价和建筑风格的关系：

```
ggplot(shuju, aes(x = property_style, y = price_ttl)) +  
geom_boxplot()
```



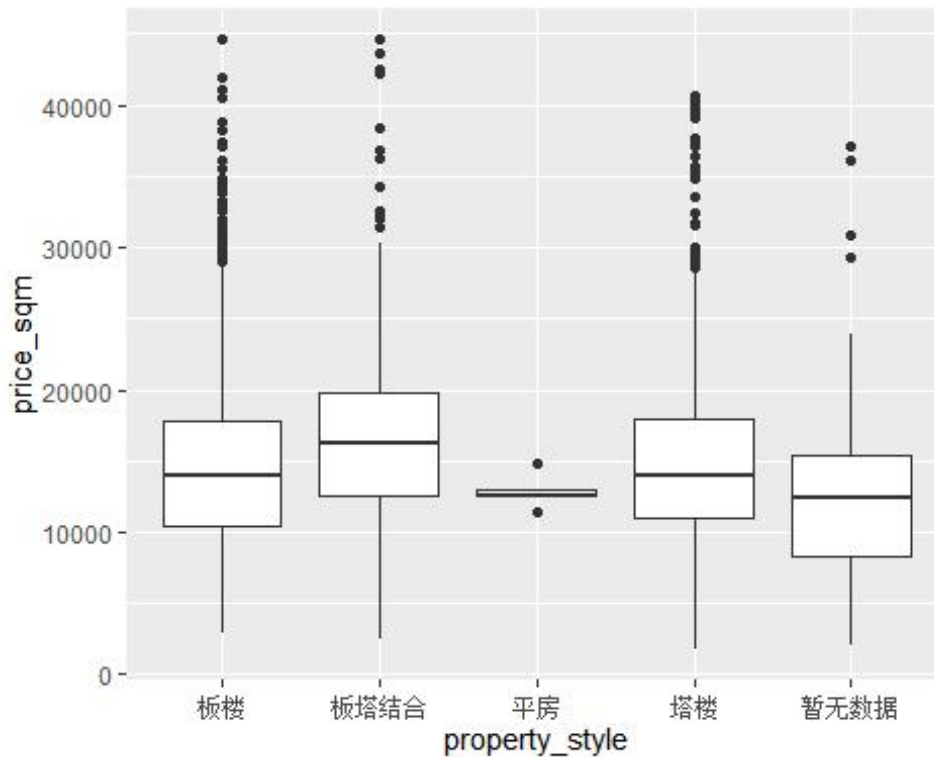
```
fenxi1<-shuju %>%
  group_by(property_style) %>%
  summarise(mean_price = mean(price_ttl),
            var_price = var(price_ttl),
            median_price=median(price_ttl),
            sd_price = sd(price_ttl))
```

#板塔结合型建筑的平均总价最高为174.8, 中位数最高为155, 但数据分布不均, 离散程度较大。平房平均总价最低, 数据较集中, 平均总价为132.6, 中位数为135

探索问题2

#每平米价格和建筑风格的关系:

```
ggplot(shuju, aes(x = property_style, y = price_sqm)) +
  geom_boxplot()
```



```
fenxi2<-shuju %>%
  group_by(property_style) %>%
  summarise(mean_price = mean(price_sqm),
            var_price = var(price_sqm),
            median_price=median(price_sqm),
            sd_price = sd(price_sqm))
```

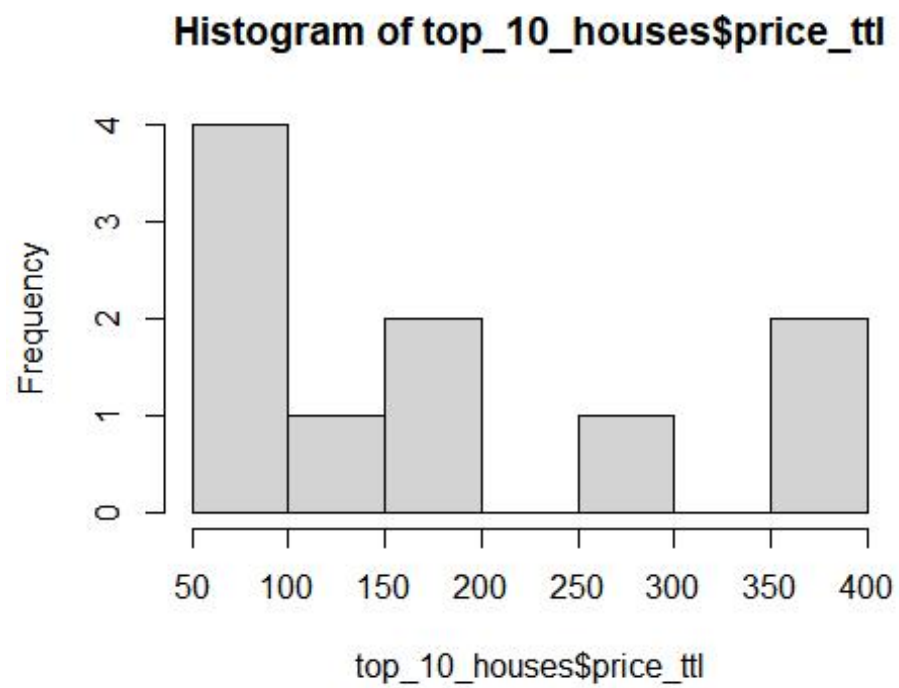
#结论：板塔结合型建筑的平均每平米单价最高为**16683.3**，中位数为**16248**，但数据分布不均。平房型数据较集中，平均单价是**12859.8**，中位数是**12577**

#结论：不同建筑形式的平均总价和每平米价格也有所不同。板塔结合的平均总价和每平米价格最高；平房的平均总价和每平米价格最低。

探索问题3

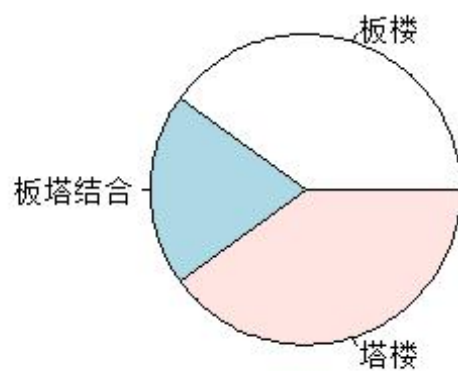
##关注人数前十名的房子有什么特点？

```
top_10_houses <- shuju %>%
  arrange(desc(followers)) %>%
  head(10)
# 价格特点，价格统计量
mean_price <- mean(top_10_houses$price_ttl)
median_price <- median(top_10_houses$price_ttl)
sd_price <- sd(top_10_houses$price_ttl)
# 绘制价格直方图
hist(top_10_houses$price_ttl)
```

```
# 建筑风格特点， 建筑风格频率
style_frequency <- table(top_10_houses$property_style)
# 绘制建筑风格饼图
pie(style_frequency, main = "关注人数前十名房子建筑风格分布")
```

关注人数前十名房子建筑风格分布



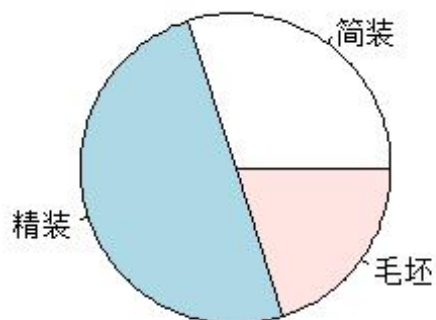
建筑装饰特点，建筑装饰频率

```
decoration_frequency <- table(top_10_houses$decoration)
```

绘制装修饼图

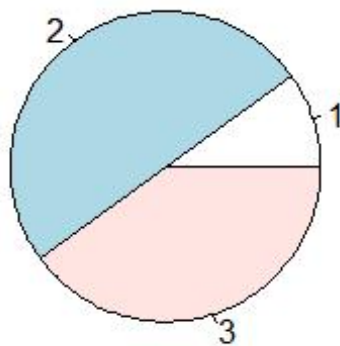
```
pie(decoration_frequency, main = "关注人数前十名房子装修风格分布")
```

关注人数前十名房子装修风格分布



```
# 房间特点
room_frequency <- table(top_10_houses$bedrooms)
# 绘制房间饼图
pie(room_frequency, main = "关注人数前十名房间分布")
```

关注人数前十名房间分布



#结论：在二手网站最受关注的房子一般都是精装房，板楼结构，有两个卧室，均价在 **184.6 万元一套**

发现总结

卖家在二手网站会关注的房子特征：精装房，板楼结构，有两个卧室，均价在 **184.6 万元一套**

板塔结合型建筑总价是所有建筑中最贵的