

# 关于武汉房屋价格的描述性统计分析

祝晶

2023/10/18

## 你的主要发现

- 1. 房屋总价与房屋均价存在高度相关关系。
- 2. 房屋总价与房屋均价均成正偏态分布。
- 3. 装修与建筑结构对房屋价格影响不大，区域间房屋价格相差较大。

## 数据介绍

本报告链家数据获取方式如下：

报告人在 2023 年 9 月 12 日获取了链家武汉二手房网站数据。

- 链家二手房网站默认显示 100 页，每页 30 套房产，因此本数据包括 3000 套房产信息；
- 数据包括了页面可见部分的文本信息，具体字段及说明见作业说明。

说明：数据仅用于教学；由于不清楚链家数据的展示规则，因此数据可能并不是武汉二手房市场的随机抽样，结论很可能有很大的偏差，甚至可能是错误的。

## 数据概览

数据表 (lj) 共包括 property\_name, property\_region, price\_ttl, price\_sqm, bedrooms, livingrooms, building\_area, directions1, directions2, decoration, property\_t\_height, property\_height, property\_style, followers, near\_subway, if\_2y, has\_key, vr 等 18 个变量, 共 3000 行。表的前 10 行示例如下：

Table 1: 武汉链家二手房

| property_name | property_region | price_ttl | price_sqm | bedrooms | livingrooms | building_area | directions1 | directions2 | decoration | property_t_height | property_height | property_style | followers | near_subway | if_2y | has_key | vr     |
|---------------|-----------------|-----------|-----------|----------|-------------|---------------|-------------|-------------|------------|-------------------|-----------------|----------------|-----------|-------------|-------|---------|--------|
| 南湖名都 A 区      | 南湖沃尔玛           | 237.0     | 18709     | 3        | 1           | 126.68        | 南           | 北           | 精装         | 17                | 中               | 塔楼             | 3         | 近地铁         | NA    | 随时看房    | NA     |
| 万科紫悦湾         | 光谷东             | 127.0     | 14613     | 3        | 2           | 86.91         | 南           | NA          | 精装         | 28                | 中               | 板楼             | 1         | NA          | 房本满两年 | 随时看房    | VR 看装修 |
| 东立国际          | 二七              | 75.0      | 15968     | 1        | 1           | 46.97         | 南           | NA          | 简装         | 18                | 低               | 塔楼             | 3         | 近地铁         | NA    | 随时看房    | NA     |
| 新都汇           | 光谷广场            | 188.0     | 15702     | 3        | 2           | 119.73        | 北           | 东           | 精装         | 32                | 高               | 塔楼             | 2         | 近地铁         | 房本满两年 | 随时看房    | NA     |
| 保利城二期         | 团结大道            | 182.0     | 17509     | 3        | 2           | 103.95        | 东南          | NA          | 简装         | 34                | 中               | 板塔结合           | 3         | NA          | 房本满两年 | 随时看房    | VR 看装修 |
| 加州城郡          | 融山              | 122.0     | 10376     | 3        | 2           | 117.59        | 南           | 北           | 精装         | 34                | 低               | 板楼             | 1         | NA          | 房本满两年 | 随时看房    | NA     |
| 省建英立公司西区      | 光谷广场            | 99.0      | 12346     | 2        | 1           | 80.19         | 南           | NA          | 简装         | 7                 | 低               | 板楼             | 0         | 近地铁         | NA    | 随时看房    | VR 看装修 |
| 保利上城东区        | 白沙洲             | 193.8     | 16336     | 3        | 2           | 118.64        | 南           | 北           | 其他         | 34                | 中               | 板塔结合           | 0         | 近地铁         | 房本满两年 | 随时看房    | NA     |
| 石化大院          | 中南丁字桥           | 325.0     | 32631     | 4        | 1           | 99.60         | 南           | 北           | 简装         | 5                 | 低               | 板楼             | 2         | 近地铁         | NA    | 随时看房    | NA     |
| 阳光花园          | 杨汊湖             | 192.0     | 17403     | 3        | 2           | 110.33        | 南           | 北           | 其他         | 7                 | 低               | 板楼             | 0         | 近地铁         | 房本满两年 | 随时看房    | NA     |

各变量的简短信息：

```
## Rows: 3,000
## Columns: 18
## $ 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", "团结大~
## $ price_ttl <dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8~
## $ price_sqm <dbl> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163~
## $ bedrooms <dbl> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
```

```
## $ livingrooms      <dbl> 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2~
## $ building_area    <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19,~
## $ directions1      <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "南", "~
## $ directions2      <chr> "北", NA, NA, "东", NA, "北", NA, "北", "北", "北", ~
## $ decoration        <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简~
## $ property_t_height <dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ~
## $ property_height   <chr> "中", "中", "低", "高", "中", "低", "低", "中", "低",~
## $ property_style    <chr> "塔楼", "板楼", "塔楼", "塔楼", "板塔结合", "板楼", ~
## $ followers         <dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ~
## $ near_subway       <chr> "近地铁", NA, "近地铁", "近地铁", NA, NA, "近地铁", ~
## $ if_2y             <chr> NA, "房本满两年", NA, "房本满两年", "房本满两年", "~
## $ has_key           <chr> "随时看房", "随时看房", "随时看房", "随时看房", "随~
## $ vr               <chr> NA, "VR看装修", NA, NA, "VR看装修", NA, "VR看装修", ~
```

各变量的简短统计:

```
## 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_height
## Length:3000        Length:3000        Min.   : 2.00    Length:3000
## Class :character    Class :character    1st Qu.:11.00    Class :character
## Mode  :character    Mode  :character    Median :27.00    Mode  :character
##                                     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 :character
## Mode  :character    Median : 3.000     Mode  :character    Mode  :character
##                                     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
##
##
##
```

将房屋总价均价按所处区域分类排序

```
##      Group.1      x
## 1      阳逻    5015.833
```

|       |       |           |
|-------|-------|-----------|
| ## 2  | 汉南其它  | 5512.000  |
| ## 3  | 蔡甸城区  | 5735.649  |
| ## 4  | 黄陂其它  | 5764.167  |
| ## 5  | 江夏其它  | 6612.500  |
| ## 6  | 后官湖   | 7055.000  |
| ## 7  | 中法生态城 | 7065.000  |
| ## 8  | 蔡甸其它  | 7356.000  |
| ## 9  | 武湖    | 7626.133  |
| ## 10 | 汉口北   | 7974.000  |
| ## 11 | 盘龙城   | 8376.087  |
| ## 12 | 吴家山   | 8399.286  |
| ## 13 | 纸坊    | 9276.903  |
| ## 14 | 庙山    | 9456.812  |
| ## 15 | 东西湖其它 | 10149.143 |
| ## 16 | 沌口    | 10268.667 |
| ## 17 | 金银湖   | 10338.928 |
| ## 18 | 藏龙岛   | 10944.514 |
| ## 19 | 常青路   | 10945.700 |
| ## 20 | 将军路   | 10961.105 |
| ## 21 | 文化大道  | 11474.258 |
| ## 22 | 长丰常码头 | 11820.143 |
| ## 23 | 七里庙   | 11885.686 |
| ## 24 | 堤角    | 11968.838 |
| ## 25 | 黄家湖   | 12331.000 |
| ## 26 | 古田    | 12666.585 |
| ## 27 | 大智路   | 12706.143 |
| ## 28 | 白沙洲   | 13836.772 |
| ## 29 | 近地铁   | 13911.667 |
| ## 30 | 王家湾   | 13918.564 |
| ## 31 | 青山    | 14228.212 |
| ## 32 | 汉正街   | 14261.800 |
| ## 33 | 台北香港路 | 14619.458 |
| ## 34 | 首义    | 14703.292 |
| ## 35 | 四新    | 14816.733 |
| ## 36 | 后湖    | 14915.314 |
| ## 37 | 集贤    | 15045.200 |
| ## 38 | 徐东    | 15165.745 |
| ## 39 | 杨汊湖   | 15269.697 |
| ## 40 | 前进江汉  | 15425.684 |
| ## 41 | 武昌火车站 | 15463.762 |
| ## 42 | 钟家村   | 15483.585 |
| ## 43 | 国际百纳  | 15664.000 |
| ## 44 | 团结大道  | 15773.769 |
| ## 45 | 三环南   | 15827.556 |
| ## 46 | 珞狮南路  | 16163.881 |
| ## 47 | 塔子湖   | 16185.662 |
| ## 48 | 随时看   | 16259.500 |
| ## 49 | 洪山其它  | 16284.833 |
| ## 50 | 百步亭   | 16369.912 |
| ## 51 | 金融港   | 16408.288 |
| ## 52 | 常青花园  | 16428.472 |
| ## 53 | 光谷广场  | 16493.160 |
| ## 54 | 宗关    | 16620.853 |
| ## 55 | 新华路万达 | 16637.125 |

|       |        |           |
|-------|--------|-----------|
| ## 56 | 光谷东    | 16683.446 |
| ## 57 | 长港路    | 16758.048 |
| ## 58 | 宝丰崇仁   | 16991.727 |
| ## 59 | 关西长职   | 16994.391 |
| ## 60 | 杨园     | 17115.727 |
| ## 61 | 唐家墩    | 17119.556 |
| ## 62 | 光谷南    | 17205.778 |
| ## 63 | 育才花桥   | 17855.111 |
| ## 64 | 街道口    | 18016.923 |
| ## 65 | 华科大    | 18134.038 |
| ## 66 | 随时看房   | 18945.667 |
| ## 67 | 新南湖    | 19444.793 |
| ## 68 | 老南湖    | 19460.926 |
| ## 69 | 民族大道   | 19621.689 |
| ## 70 | 二七     | 19923.395 |
| ## 71 | 东湖东亭   | 20376.000 |
| ## 72 | 沙湖     | 20631.062 |
| ## 73 | 武广万松园  | 20755.889 |
| ## 74 | 卓刀泉    | 21285.733 |
| ## 75 | 街道可    | 21323.500 |
| ## 76 | 中南丁字桥  | 21373.451 |
| ## 77 | 积玉桥    | 21403.413 |
| ## 78 | 关山大道   | 21480.040 |
| ## 79 | 楚河汉街   | 21957.933 |
| ## 80 | CBD西北湖 | 22272.429 |
| ## 81 | 虎泉杨家湾  | 23902.095 |
| ## 82 | 南湖沃尔玛  | 24180.818 |
| ## 83 | 三阳路    | 24777.312 |
| ## 84 | 黄埔永清   | 24956.783 |
| ## 85 | 水果湖    | 28561.556 |
| ## 86 | 中北路    | 32727.778 |
| ## 87 | VR看装修  | 38351.000 |

可以看到：

- 直观结论 1 房屋总价的中位数为 137.0 万元，均值为 155.9 万元，房屋单价的中位数为 14404 元，均值为 15148 元，中位数均低于均值，说明成正偏态分布。大部分房屋价格都小于平均房屋价格。
- 直观结论 2 按照房屋所处区域房屋单价均值排序，均价最低的是阳逻区，为 5015.83 元/平方，均价最高的是中北路，为 32727.78 元/平方。

## 探索性分析

房屋总价的数值描述与图形

描述房屋总价

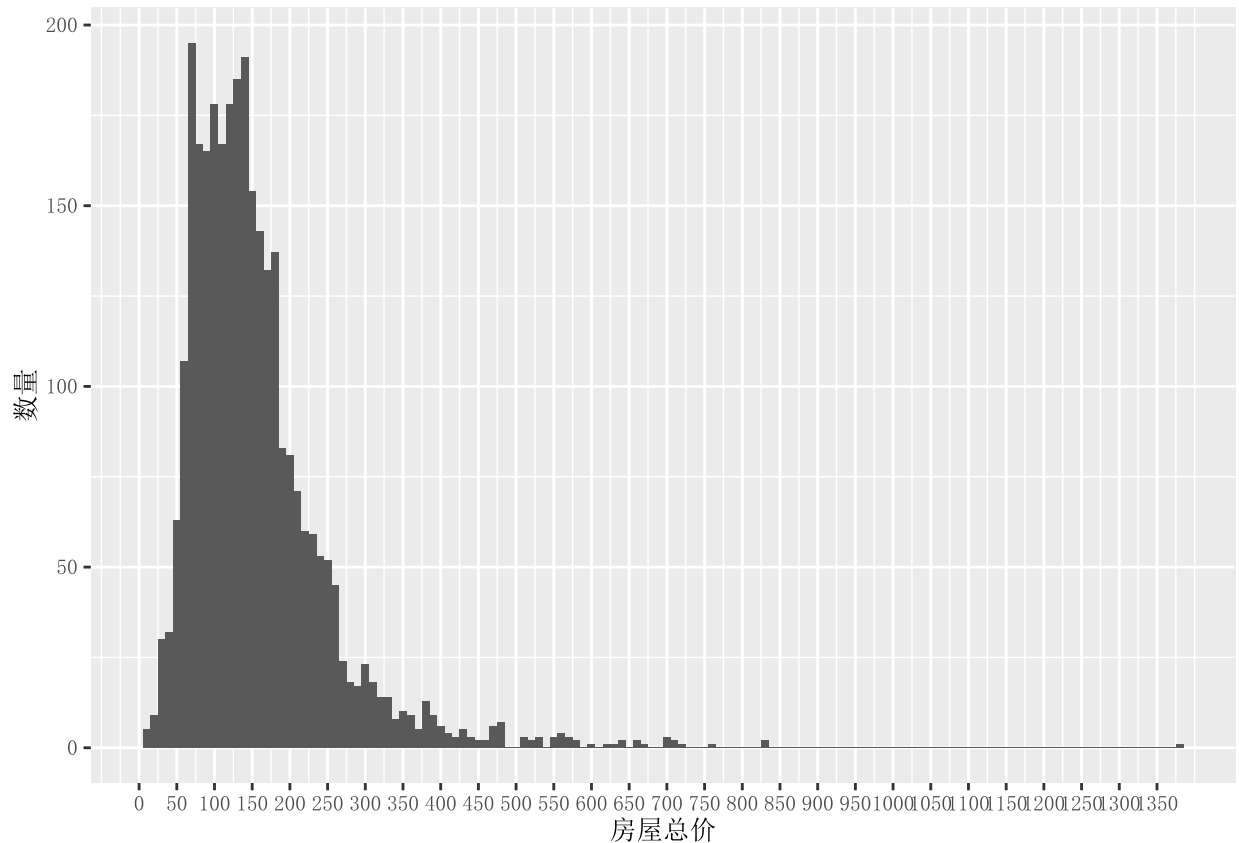
求房屋总价众数

```
## [1] "105"
```

Table 2: 房屋总价

|                   | vars | n    | mean         | sd           | median    | trimmed      | mad        | min     | max      | range    | skew        | kurtosis    | se          |
|-------------------|------|------|--------------|--------------|-----------|--------------|------------|---------|----------|----------|-------------|-------------|-------------|
| property_name*    | 1    | 3000 | 631.533000   | 391.9013470  | 622.500   | 625.136250   | 519.65130  | 1.00    | 1345.00  | 1344.00  | 0.0915593   | -1.2156518  | 7.1551069   |
| property_region*  | 2    | 3000 | 44.698333    | 25.4259347   | 48.000    | 45.032500    | 29.65200   | 1.00    | 87.00    | 86.00    | -0.1445066  | -1.1364377  | 0.4642119   |
| price_ttl         | 3    | 3000 | 155.862780   | 95.5481281   | 137.000   | 142.481767   | 66.71700   | 10.60   | 1380.00  | 1369.40  | 2.7532227   | 16.1167156  | 1.7444622   |
| price_sqm         | 4    | 3000 | 15148.493333 | 6323.1753620 | 14404.000 | 14579.246667 | 5465.60490 | 1771.00 | 44656.00 | 42885.00 | 1.0794641   | 2.0256251   | 115.4448594 |
| bedrooms          | 5    | 3000 | 2.695000     | 0.7299447    | 3.000     | 2.682917     | 0.00000    | 1.00    | 7.00     | 6.00     | 0.1356027   | 1.6357110   | 0.0133269   |
| livingrooms       | 6    | 3000 | 1.708667     | 0.4731450    | 2.000     | 1.765417     | 0.00000    | 0.00    | 4.00     | 4.00     | -0.9914220  | -0.1840918  | 0.0086384   |
| building_area     | 7    | 3000 | 100.866920   | 30.3799964   | 95.545    | 99.593079    | 23.26941   | 22.77   | 588.66   | 565.89   | 2.0797845   | 23.6358521  | 0.5546603   |
| directions1*      | 8    | 3000 | 4.783000     | 0.9547503    | 5.000     | 4.934583     | 0.00000    | 1.00    | 8.00     | 7.00     | -1.4886364  | 7.6220936   | 0.0174313   |
| directions2*      | 9    | 1328 | 1.447289     | 1.3792275    | 1.000     | 1.007519     | 0.00000    | 1.00    | 8.00     | 7.00     | 3.0336388   | 8.1537013   | 0.0378475   |
| decoration*       | 10   | 3000 | 2.049333     | 0.7649148    | 2.000     | 1.989583     | 0.00000    | 1.00    | 4.00     | 3.00     | 0.6896433   | 0.5341913   | 0.0139654   |
| property_t_height | 11   | 3000 | 24.218000    | 12.4482447   | 27.000    | 23.841667    | 11.86080   | 2.00    | 62.00    | 60.00    | 0.0485029   | -0.7972010  | 0.2272728   |
| property_height*  | 12   | 2940 | 2.136735     | 0.8205909    | 2.000     | 2.170918     | 1.48260    | 1.00    | 3.00     | 2.00     | -0.2568888  | -1.4702882  | 0.0151340   |
| property_style*   | 13   | 3000 | 1.831333     | 1.2190352    | 1.000     | 1.634167     | 0.00000    | 1.00    | 5.00     | 4.00     | 1.2242713   | -0.0149867  | 0.0222564   |
| followers         | 14   | 3000 | 6.613667     | 15.2170168   | 3.000     | 3.526250     | 2.96520    | 0.00    | 262.00   | 262.00   | 6.9007673   | 68.1705708  | 0.2778234   |
| near_subway*      | 15   | 1559 | 2.998717     | 0.0947852    | 3.000     | 3.000000     | 0.00000    | 1.00    | 5.00     | 4.00     | -5.9853266  | 394.0145339 | 0.0024006   |
| if_2y*            | 16   | 1264 | 1.000000     | 0.0000000    | 1.000     | 1.000000     | 0.00000    | 1.00    | 1.00     | 0.00     | NaN         | NaN         | 0.0000000   |
| has_key*          | 17   | 2542 | 6.978757     | 0.3179361    | 7.000     | 7.000000     | 0.00000    | 1.00    | 9.00     | 8.00     | -15.0022325 | 249.6071514 | 0.0063060   |
| vr*               | 18   | 2094 | 2.021012     | 0.3690586    | 2.000     | 2.000000     | 0.00000    | 1.00    | 11.00    | 10.00    | 19.0577035  | 387.3680211 | 0.0080650   |

房屋总价绘图



发现:

- 发现 1: 房屋总价大部分集中于 50 万—150 万之间, 与均值 155.86 万大致相等。
- 发现 2: 房屋总价成右偏分布, 总价最高值 1380 远远高于众数 105。

## 房屋单价的数值描述与图形

描述房屋单价

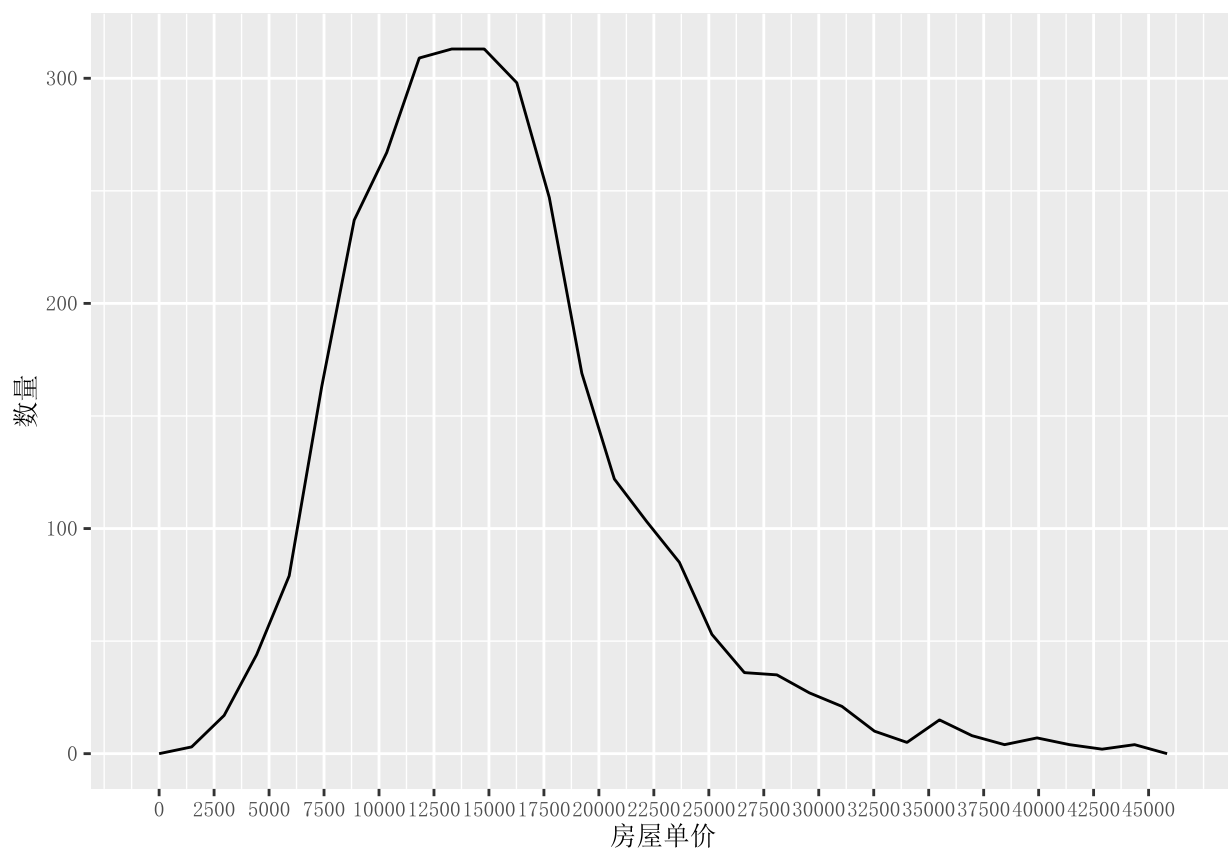
```
##      vars      n      mean      sd median trimmed      mad min      max range skew
```

```
## X1      1 3000 15148.49 6323.18 14404 14579.25 5465.6 1771 44656 42885 1.08
##      kurtosis      se
## X1      2.03 115.44
```

求房屋单价众数

```
## [1] "14394"
```

房屋单价绘图



发现：

- 发现 1: 房屋单价大部分集中于 10000 元/平方—16000 元/平方之间，与均值 15148 元/平方大致相等。
- 发现 2: 房屋总价成右偏分布，单价最高值 44656 远远高于众数 14394.

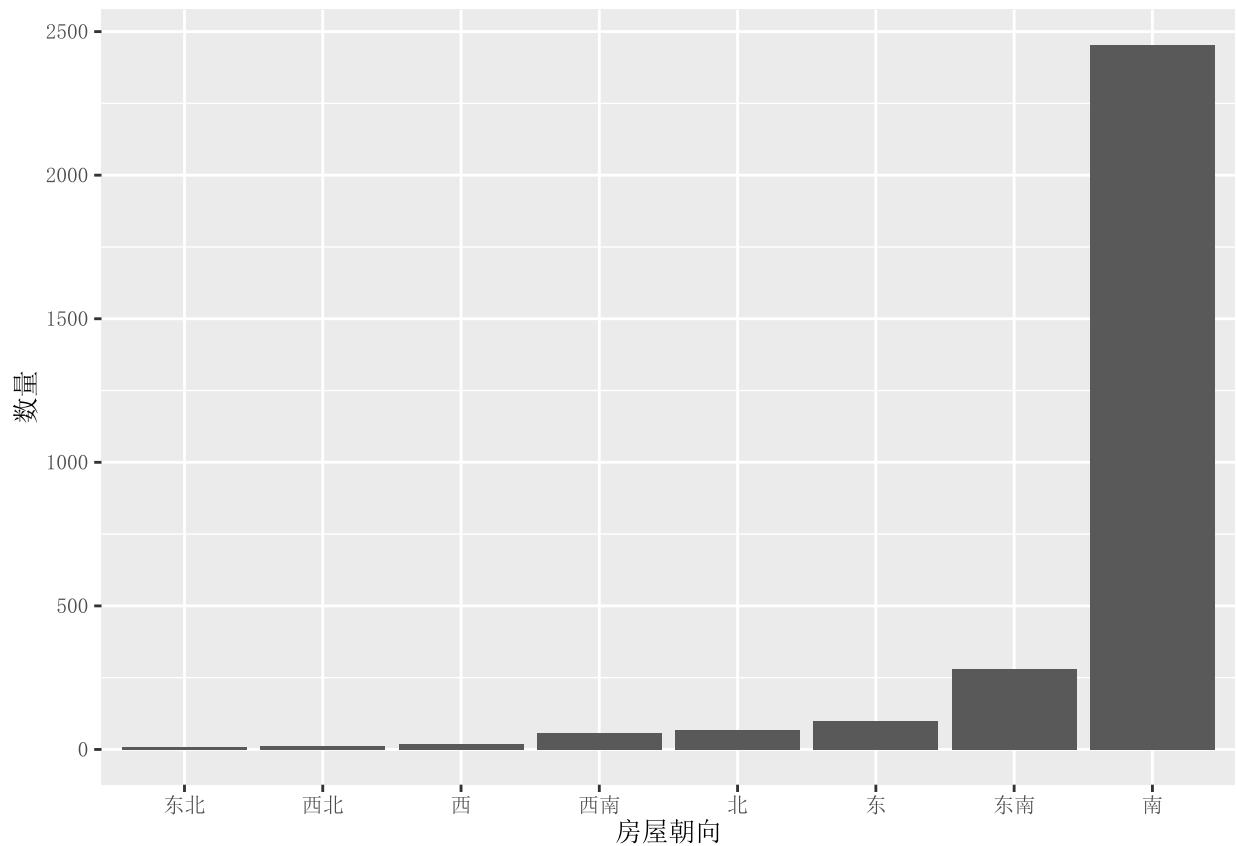
房屋主要朝向的数值描述与图形

房屋朝向计数

Table 3: 房屋朝向

| 北  | 东  | 东北 | 东南  | 南    | 西  | 西北 | 西南 |
|----|----|----|-----|------|----|----|----|
| 68 | 98 | 10 | 281 | 2454 | 19 | 13 | 57 |

房屋朝向绘图



发现:

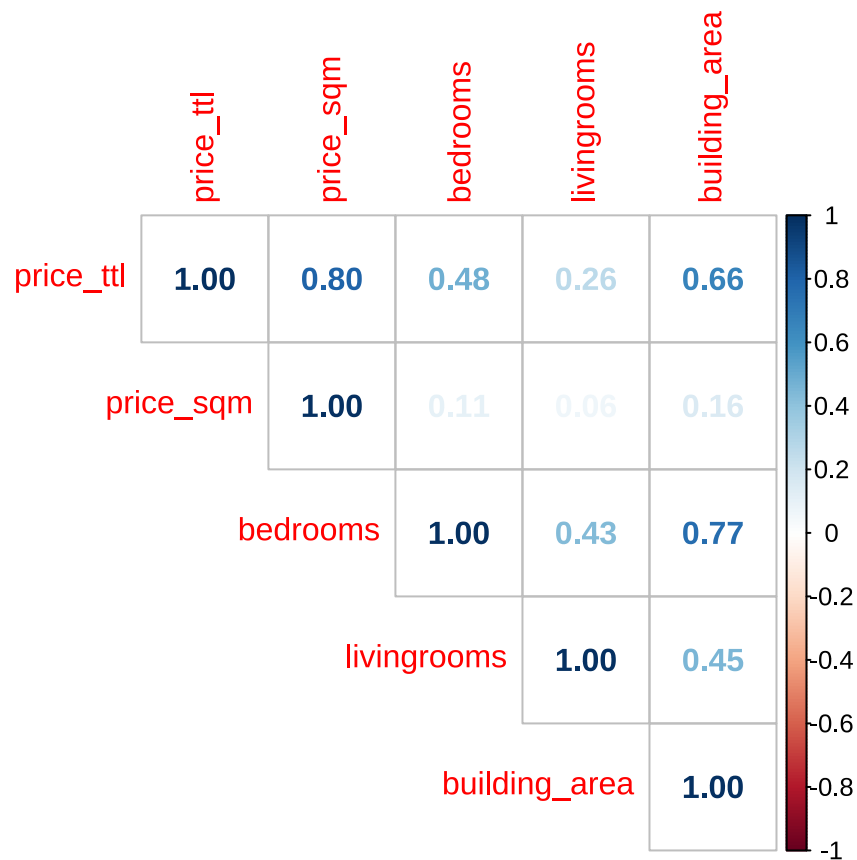
- 发现 1: 二手房网站出售的房屋主要朝向多数为南边。
- 发现 2: 其他所有朝向的房屋加起来占有率约为  $(3000-2454)/3000=18.2\%$ 。

在该二手房网站的房屋总价与什么因素相关

计算房屋总价相关性

```
##          price_ttl price_sqm bedrooms livingrooms building_area
## price_ttl          1.00      0.80      0.48      0.26          0.66
## price_sqm          0.80      1.00      0.11      0.06          0.16
## bedrooms          0.48      0.11      1.00      0.43          0.77
## livingrooms        0.26      0.06      0.43      1.00          0.45
## building_area       0.66      0.16      0.77      0.45          1.00
```

绘图



计算 P 值

```
##
## Pearson's product-moment correlation
##
## data:  lj$price_ttl and lj$price_sqm
## t = 74.125, df = 2998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7913449 0.8166323
## sample estimates:
##      cor
## 0.8043525

##
## Pearson's product-moment correlation
##
## data:  lj$price_ttl and lj$building_area
## t = 48.566, df = 2998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6430516 0.6831320
## sample estimates:
##      cor
## 0.6635677
```



发现:

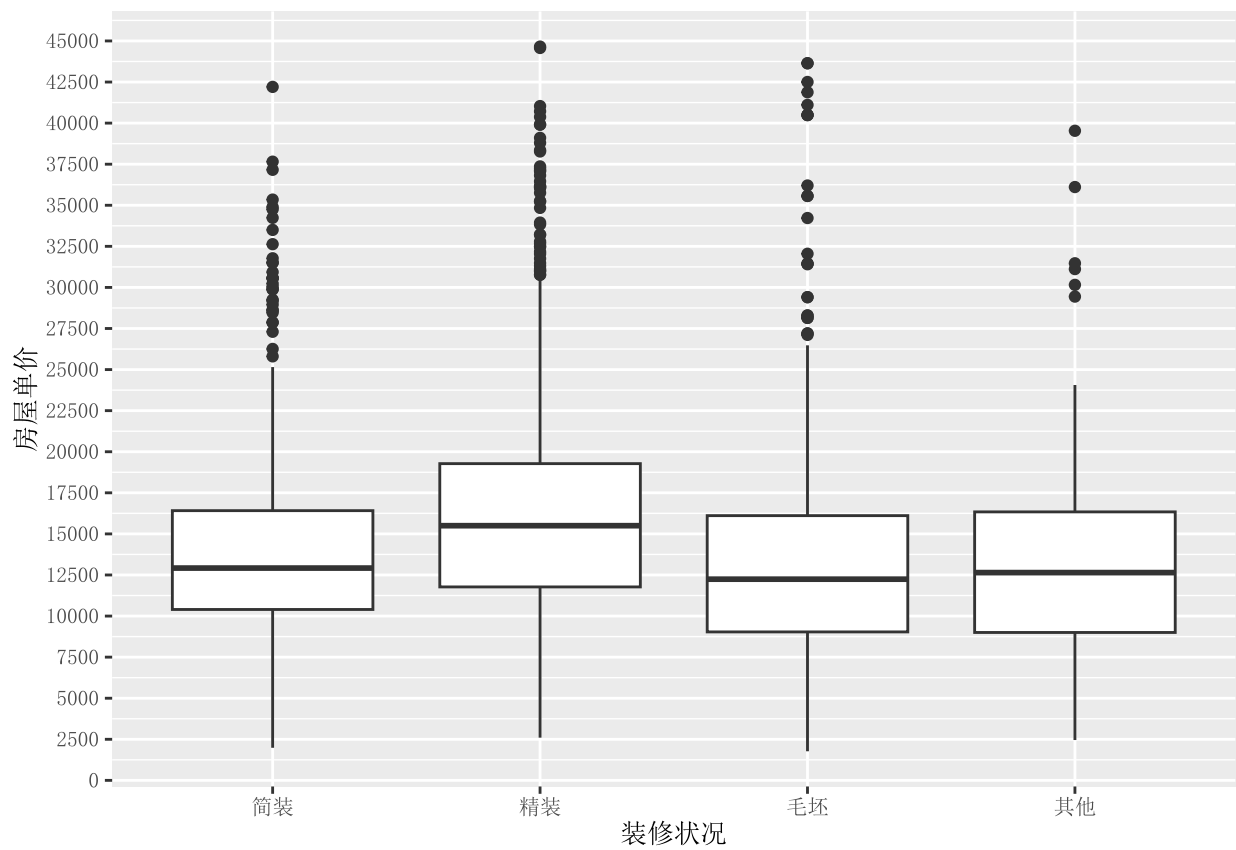
- 发现 1: 房屋总价与房屋均价的 p 值  $<0.001$ , 存在显著相关, 且相关系数为 0.8, 两者强相关。
- 发现 2: 房屋总价与建筑面积的 p 值  $<0.001$ , 存在显著相关, 且相关系数为 0.66, 两者中等相关。

## 装修对价格的影响

描述装修对价格的影响

```
##
## Descriptive statistics by group
## group: 简装
##   vars    n   mean    sd median trimmed   mad  min max range skew kurtosis   se
## X1      1 634 134.07 80.59    118   122.4 62.27 10.6 636 625.4 2.25    8.34 3.2
## -----
## group: 精装
##   vars    n   mean    sd median trimmed   mad  min max range skew kurtosis
## X1      1 1757 165.58 94.81    148   153.01 65.23 10.8 1380 1369.2 2.94    20.46
##       se
## X1 2.26
## -----
## group: 毛坯
##   vars    n   mean    sd median trimmed   mad  min max range skew kurtosis   se
## X1      1 436 161.02 117.51   129.5   140.93 68.94   14 835   821 2.53    8.64 5.63
## -----
## group: 其他
##   vars    n   mean    sd median trimmed   mad  min max range skew kurtosis   se
## X1      1 173 124.02 69.18    105   116.02 50.41 15.5 399 383.5 1.25    1.74 5.26
```

绘图



发现:

- 发现 1: 简装房与毛坯房价格基本相当，精装房价格高于简装和毛坯房。
- 发现 2: 离群点多集中于上半部分，上下四分位数都比较集中。

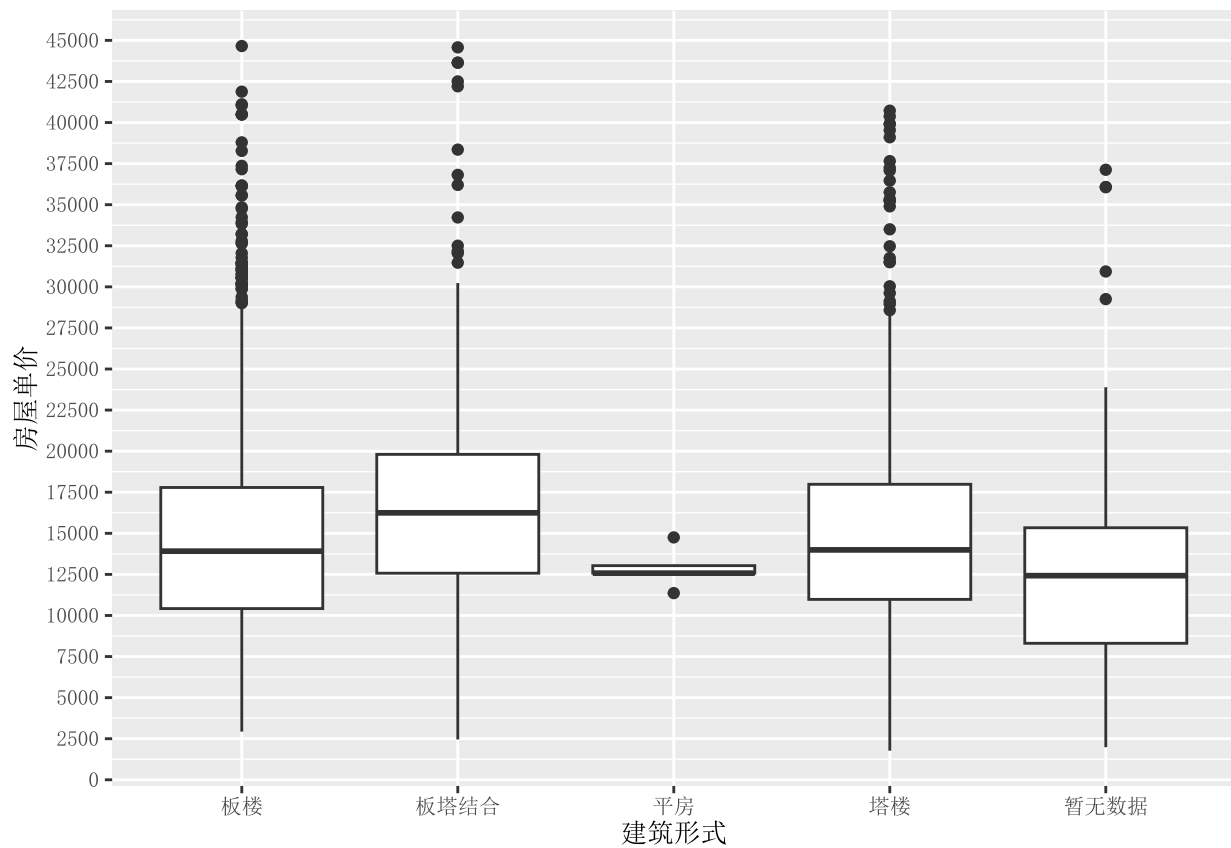
## 建筑形式对价格的影响

描述建筑形式对价格的影响

```
##
## Descriptive statistics by group
## group: 板楼
## vars n mean sd median trimmed mad min max range skew
## X1 1 1781 14698.78 6129.05 13914 14122.72 5343.29 2934 44656 41722 1.08
## kurtosis se
## X1 1.89 145.23
## -----
## group: 板塔结合
## vars n mean sd median trimmed mad min max range skew
## X1 1 615 16683.35 6295.36 16248 16293.33 5411.49 2456 44574 42118 0.88
## kurtosis se
## X1 2.11 253.85
## -----
## group: 平房
## vars n mean sd median trimmed mad min max range skew kurtosis
```

```
## X1      1 5 12859.8 1226.42  12577 12859.8 674.58 11361 14752  3391 0.37   -1.42
##          se
## X1 548.47
## -----
## group: 塔楼
##   vars  n    mean      sd median trimmed   mad  min  max range skew
## X1     1 527 15147.3 6594.23  13992 14450.4 5206.89 1771 40721 38950 1.31
##   kurtosis    se
## X1       2.55 287.25
## -----
## group: 暂无数据
##   vars  n    mean      sd median trimmed   mad  min  max range skew
## X1     1 72 13329.94 7273.16  12422 12343.9 5754.71 1984 37126 35142 1.5
##   kurtosis    se
## X1       2.51 857.15
```

绘图



发现:

- 发现 1: 平房的价格非常稳定, 差异非常小。
- 发现 2: 板塔结合的价格稍高, 板楼和塔楼价格相差无几。

## 发现总结

综合以上分析，我认为房屋的价格主要与所处区域及房屋面积有关。装修状况和建筑形式可以在一定程度上影响房价，但并不显著。武汉房屋价格趋势较为集中，波动幅度不大，存在远超出平均水平的价格。