

# 武汉二手房数据观察（第一次作业）

杨雪枫 2024281050971

2024-10-31

## 目录

1	你的主要发现	2
2	数据介绍	3
3	数据概览	3
4	探索性分析	7
4.1	单价的数值描述与图形 . . . . .	7
4.2	总价的数值描述与图形 . . . . .	9
4.3	房屋面积的数值描述与图形 . . . . .	11
4.4	关注人数的数值描述与图形 . . . . .	12
4.5	装修类型分布趋势 . . . . .	12
4.6	装修与单价关系 . . . . .	14
4.7	楼层位置与单价关系 . . . . .	15
4.8	房屋主要朝向与单价关系 . . . . .	16
4.9	建筑形式与单价关系 . . . . .	17
4.10	卧室数量与单价关系 . . . . .	18

1 你的主要发现	2
4.11 是否近地铁与单价关系	19
4.12 户型与单价关系	20
4.13 进一步分析装修、是否近地铁与价格关系	21
4.14 尝试建模寻找显著性	22
4.15 区域分布集中度	23
5 发现总结	23

## 1 你的主要发现

- 发现 1

武汉二手房均价为 15148，中位数为 14404；总价均值为 155.9 万，中位数为 137.0 万；面积均值为 100 平，中位数为 95 平；三者均呈现正偏分布，均值大于中位数，总价格与面积非常集中，总价集中于 90-160 万区间，面积集中于 100 平左右，以中户型居多，占比 50%，大户型与小户型占比接近；卧室主要三室居多，其次为二室；大厅主要为二厅，其次为一厅；市场供应以精装修为主，占比近 60%，其次为简装，占比超 20%，再次为毛坯，占比近 15%；市场整体关注人数不多，大部分房屋关注人数为 6 人左右，部分热门房屋，关注人数 200 人以上。

- 发现 2

武汉二手房市场不同楼层位置、卧室数量、房屋主要朝向对二手房单价影响不明显；房屋建筑形式、装修类型、是否近地铁、户型等对二手房单价有一定影响，其中：（1）精装修单价比其他装修类型单价高，简装、毛坯单价相近，高价房受装修类型的影响不明显；（2）板塔结合房屋单价较高，板楼与塔楼房屋单价相似；（3）近地铁房单价均价高于不近地铁房单价均价，进一步探究发现，是否近地铁对毛坯房影响非常大；（4）大中户型二手房单价接近，无明显差别，但是小户型均价较低，上三分位单价较低。

- 发现 3

二手房主要位于盘龙城、四新、后湖、青山、王家湾、积玉桥、百步亭，这些区域可能房屋存量较多，或人们挂牌需求较大，建议关注后续价格走势。

## 2 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。

## 3 数据概览

数据表（wh）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：

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2
南湖名都 A 区   南湖沃尔	237.	1870	3	1	1	6.68	南	北
万科紫悦湾   光谷东	12	.0  14	13 3	2		86.91	南	
东立国际   二七		75.0	5968 1	1		46.97	南	
新都汇   光谷	场   1	8.0  1	702 3	2		119.73	北	东
保利城一期   团结大道	182	0  175	9 3	2		03.95	东南	

加州橘郡   庙山		22.0	0376 3	2		117.59	南	北
省建筑五公司西区   光谷广场	99.0	12346	1		80	19	南	北
保利上城东区   白沙洲	193	8  163	6 3	2		18.64	南	北
石化大院   中南丁	桥   325	0  326	1 4	1		99.60	南	北
阳光花园   杨汊湖	1	2.0  1	403 3	2		110.33	南	北

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          <int> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163~
## $ bedrooms           <fct> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
## $ livingrooms        <fct> 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", "~
## $ directions2        <chr> "北", "", "", "东", "", "北", "", "北", "北", "北", ~
## $ decoration          <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简~
## $ property_t_height  <int> 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          <int> 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 directions2
## 1: 142 0: 18 Min. : 22.77 Length:3000 Length:3000
## 2: 919 1: 845 1st Qu.: 84.92 Class :character Class :character
## 3:1692 2:2131 Median : 95.55 Mode :character Mode :character
## 4: 214 3: 5 Mean :100.87
## 5: 27 4: 1 3rd Qu.:117.68
## 6: 4 Max. :588.66
## 7: 2
## decoration property_t_height property_height property_style
## Length:3000 Min. : 2.00 Length:3000 Length:3000
## Class :character 1st Qu.:11.00 Class :character Class :character
## Mode :character Median :27.00 Mode :character Mode :character
## Mean :24.22
## 3rd Qu.:33.00
## Max. :62.00
##
## followers near_subway if_2y has_key
## Min. : 0.000 Length:3000 Length:3000 Length:3000
## 1st Qu.: 1.000 Class :character Class :character Class :character
## Median : 3.000 Mode :character Mode :character Mode :character
## Mean : 6.614
## 3rd Qu.: 6.000
## Max. :262.000
##
## vr
## Length:3000
## Class :character
## Mode :character

```

```
##  
##  
##  
##
```

可以看到：

- 直观结论 1

数据共包含 3000 行数据，18 个变量

- 直观结论 2

武汉二手房单价均值为 15148，中位数为 14404，最高价为 44656；武汉二手房总价均值为 155.9 万，中位数为 137.0 万，最高价为 1380.0 万；

- 直观结论 3

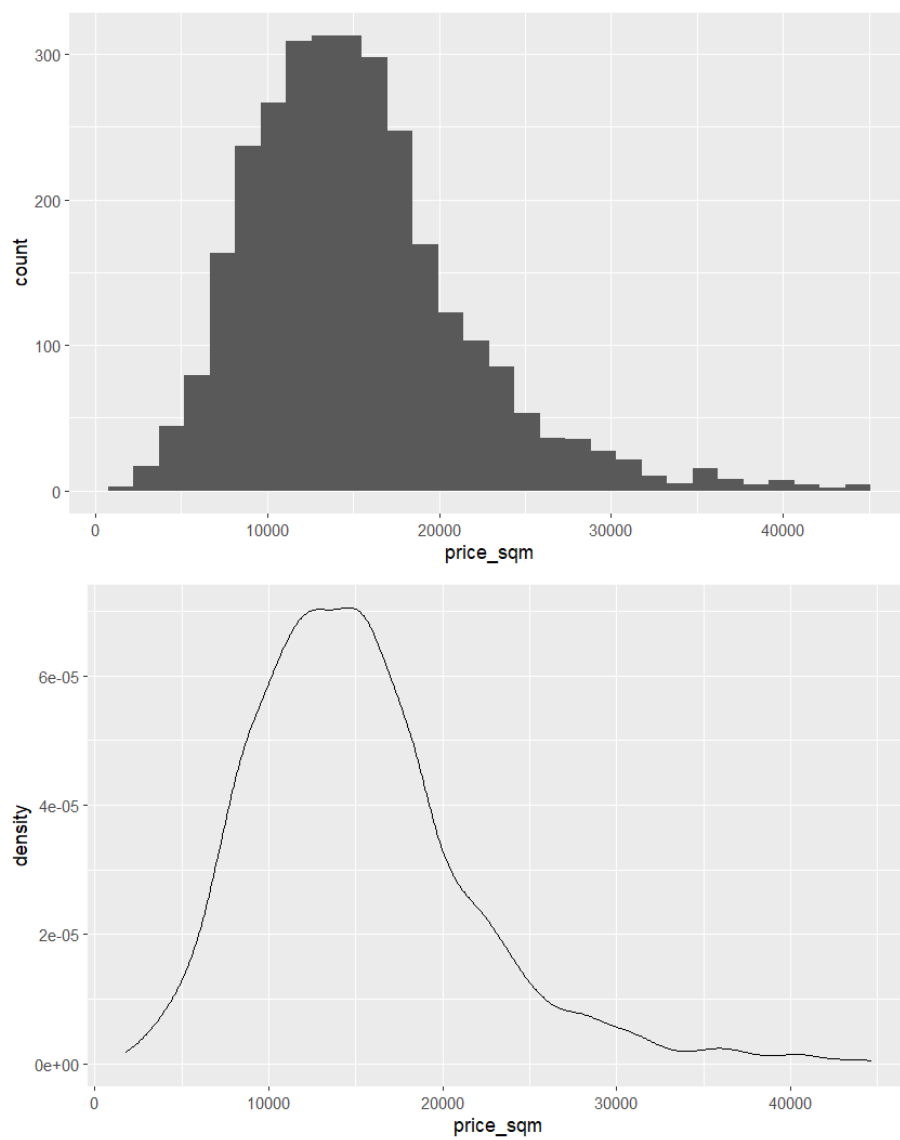
武汉二手房主要为三室，其次二室居多；武汉二手房主要为二厅，其次为一厅；武汉二手房面积均值为 100 平，中位数为 95 平，上三分位为 117 平；

表 2: 武汉二手房数据

price_ttl	price_sqm	building_area	property_t_height
Min. : 10.6	Min. : 1771	Min. : 22.77	Min. : 2.00
1st Qu.: 95.0	1st Qu.:10799	1st Qu.: 84.92	1st Qu.:11.00
Median : 137.0	Median :14404	Median : 95.55	Median :27.00
Mean : 155.9	Mean :15148	Mean :100.87	Mean :24.22
3rd Qu.: 188.0	3rd Qu.:18211	3rd Qu.:117.68	3rd Qu.:33.00
Max. :1380.0	Max. :44656	Max. :588.66	Max. :62.00

## 4 探索性分析

### 4.1 单价的数值描述与图形



```
## [1] 1.079464
```

```
## [1] 2.025625
```

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

```
## [1] 7411.75
```

```
## [1] 39982547
```

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

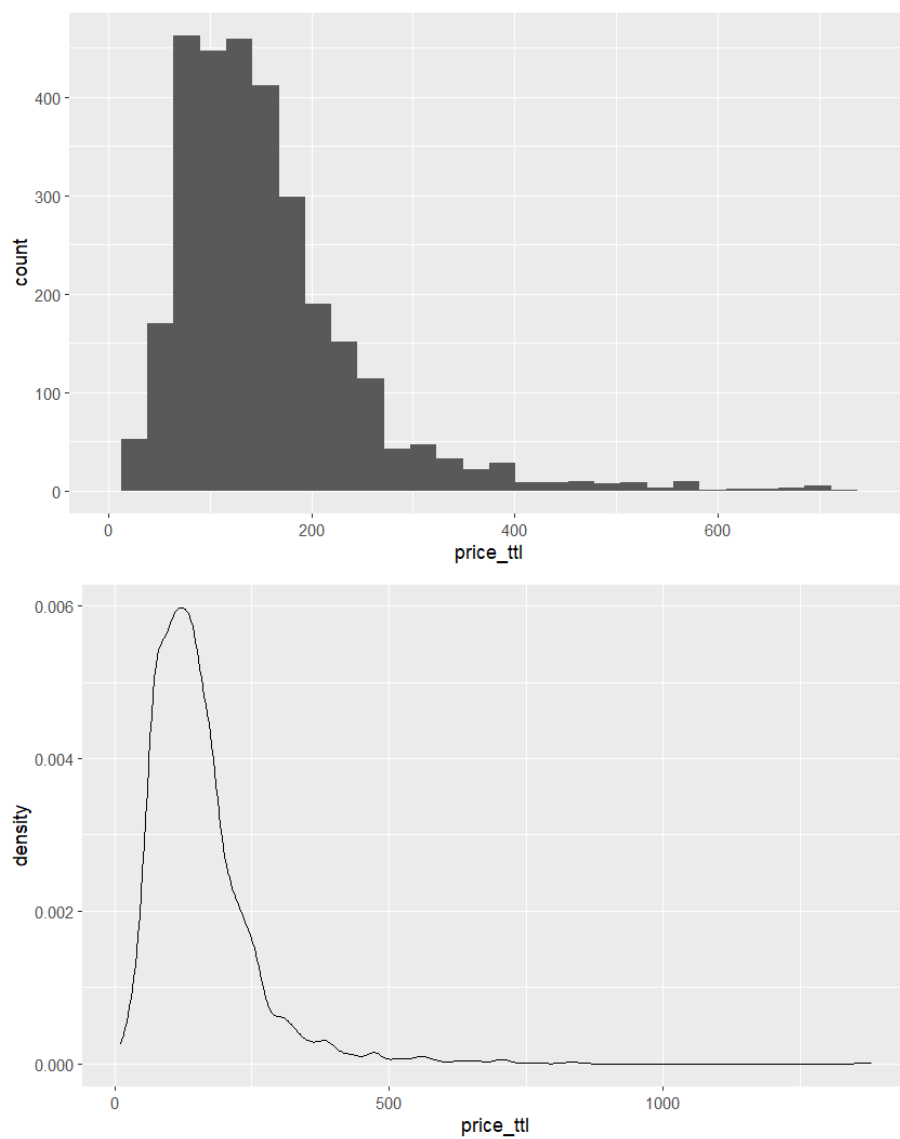
```
## [1] 5465.605
```

发现:

- 发现 1  
武汉二手房单价整体呈正偏分布，偏态系数为 1.079
- 发现 2  
武汉二手房最低单价与最高单价极差较大，标准差为 6323 元



## 4.2 总价的数值描述与图形



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

```
## [1] 93
```

```
## [1] 9129.445
```

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

```
## [1] 66.717
```

```
## [1] 2.753223
```

```
## [1] 16.11672
```

发现:

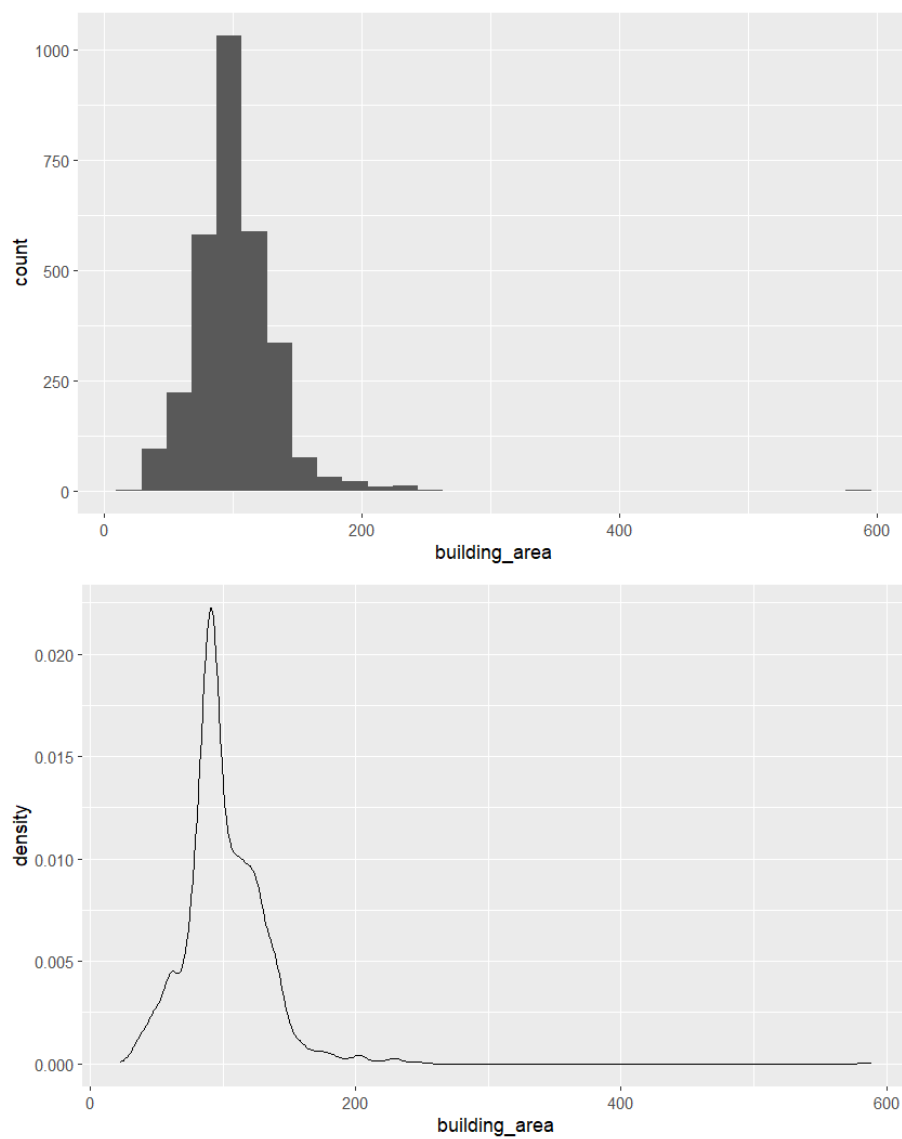
- 发现 1

武汉二手房总价整体呈正偏分布，偏态系数为 2.753，总价非常集中，峰度系数为 16.11

- 发现 2

武汉二手房最低单价与最高单价极差较大，存在异常值影响，四分位距为 93 万，标准差为 95 元

### 4.3 房屋面积的数值描述与图形



```
## [1] 22.77 588.66
```

```
## [1] 32.7625
```

```
## [1] 922.9442
```

```
## [1] 30.38
```

```
## [1] 23.26941
```

```
## [1] 2.079785
```

```
## [1] 23.63585
```

发现:

- 发现 1  
武汉二手房房屋面积整体呈正偏分布, 偏态系数为 2.080; 面积非常集中, 峰度系数为 23.63
- 发现 2  
武汉二手房房屋面积积极差较大, 存在偏离值影响, 四分位距为 32 平, 标准差为 30 平

#### 4.4 关注人数的数值描述与图形

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.000   1.000   3.000   6.614   6.000 262.000
```

发现:

- 发现 1  
武汉二手房每栋房关注人数不多, 中位数为 3 人, 均值为 6 人; 但部分房子关注人数达 200 人以上, 存在部分热门房屋。

#### 4.5 装修类型分布趋势

发现:

- 发现 1  
武汉二手房以精装修为主, 占比近 60%, 其次为简装, 占比超 20%, 再次为毛坯, 占比近 15%

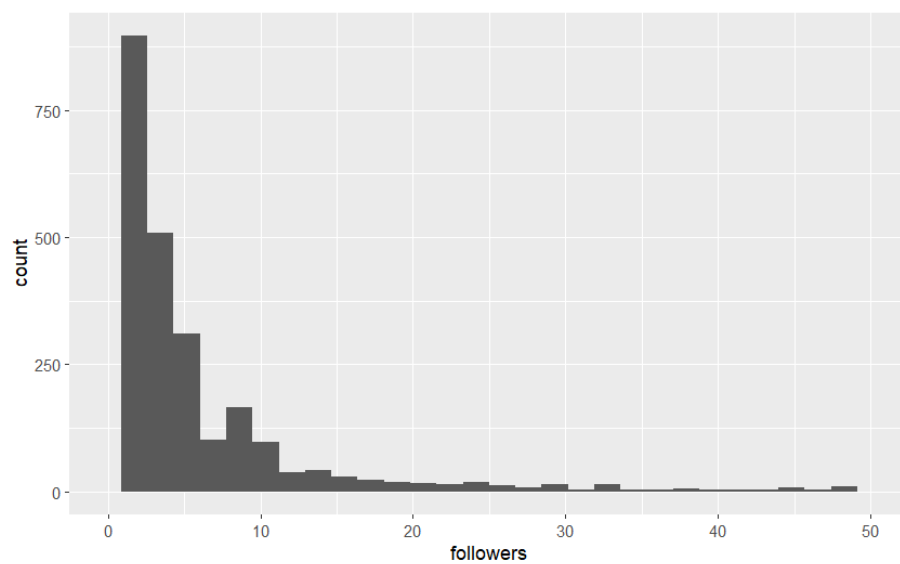


图 1: 关注人数直方图

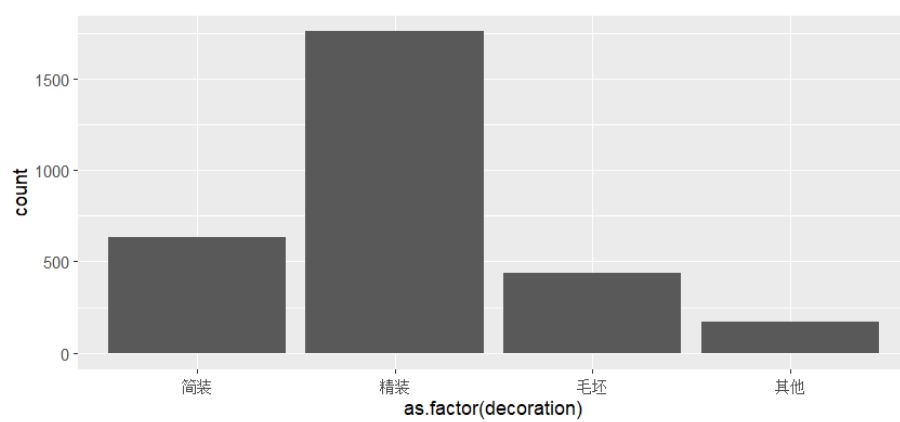
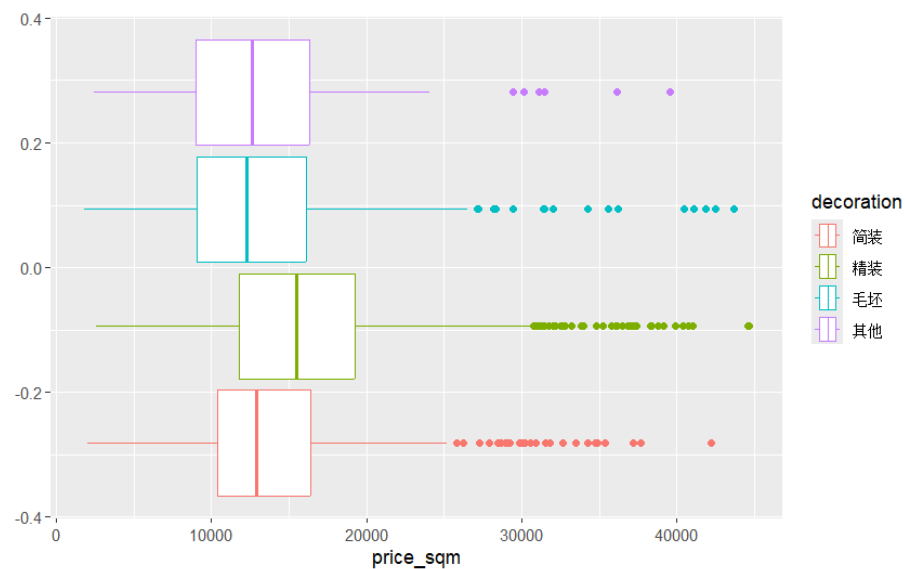


图 2: 装修类型频次分布

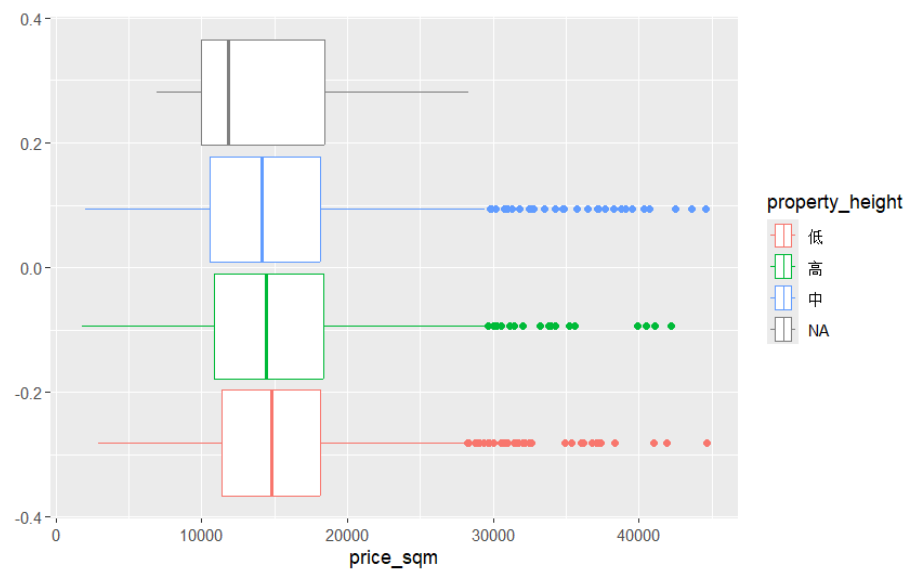
### 4.6 装修与单价关系



发现:

- 发现 1  
装修对单价有一定影响，其中精装单价较高，简装、毛坯及其他单价相近，影响不明显
- 发现 2  
不论什么装修类型，均出现一定的高价偏离值，说明高价房受装修类型的影响不明显

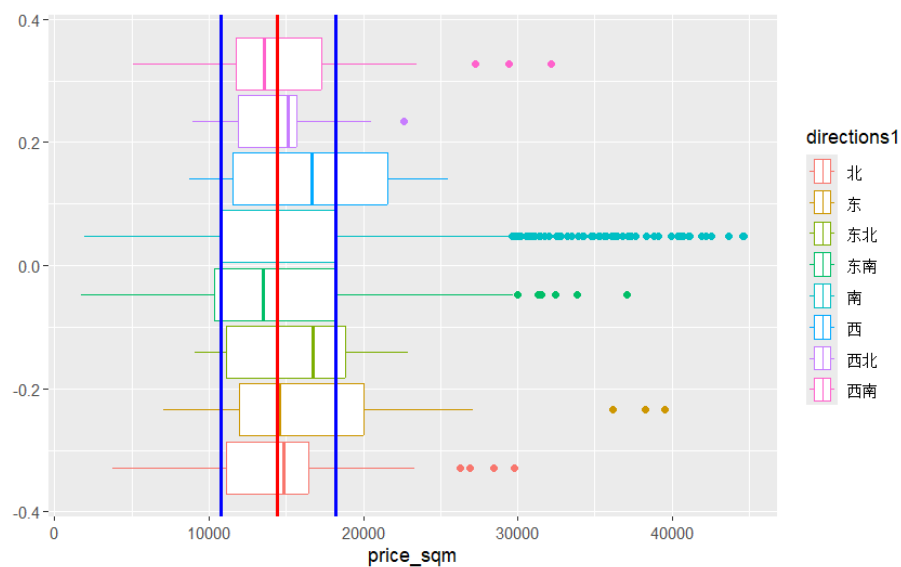
### 4.7 楼层位置与单价关系



发现:

- 发现 1  
不同楼层位置均值及分位数单价接近，对二手房单价影响不明显

### 4.8 房屋主要朝向与单价关系



发现:

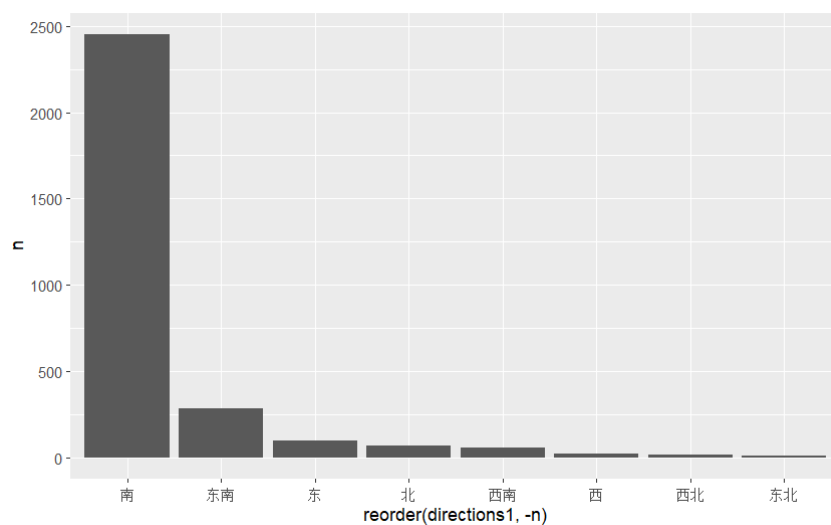
- 发现 1

房屋主要朝向对二手房单价影响不明显，大部分处于上下三分位数内，均价分布中位数两侧

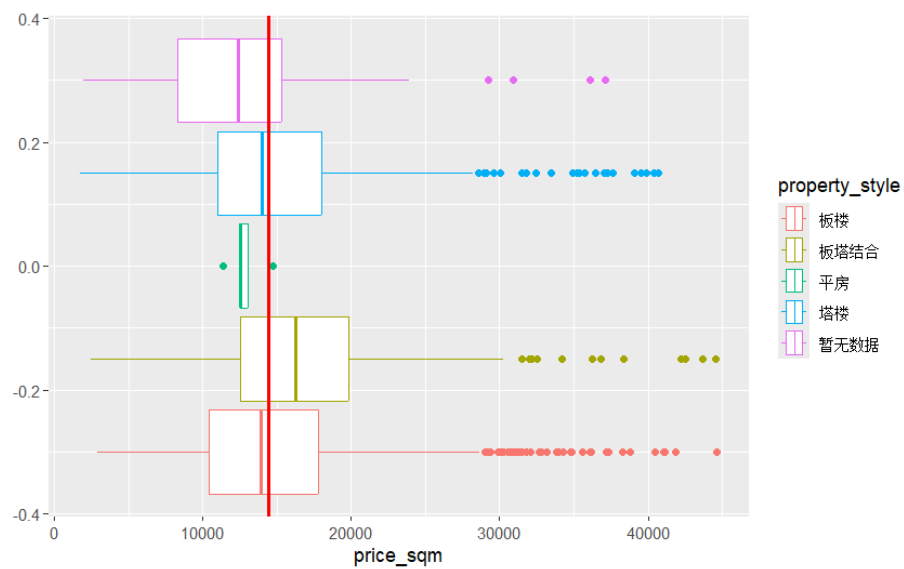
- 发现 2

部分朝向显示对单价有一定影响。如：西向均价及上三分位数单价均较高；东南向均价偏低；北向及西南向价格较集中，上三分位数单价偏低。这些可能受样本值数据或其他因素影响。





#### 4.9 建筑形式与单价关系

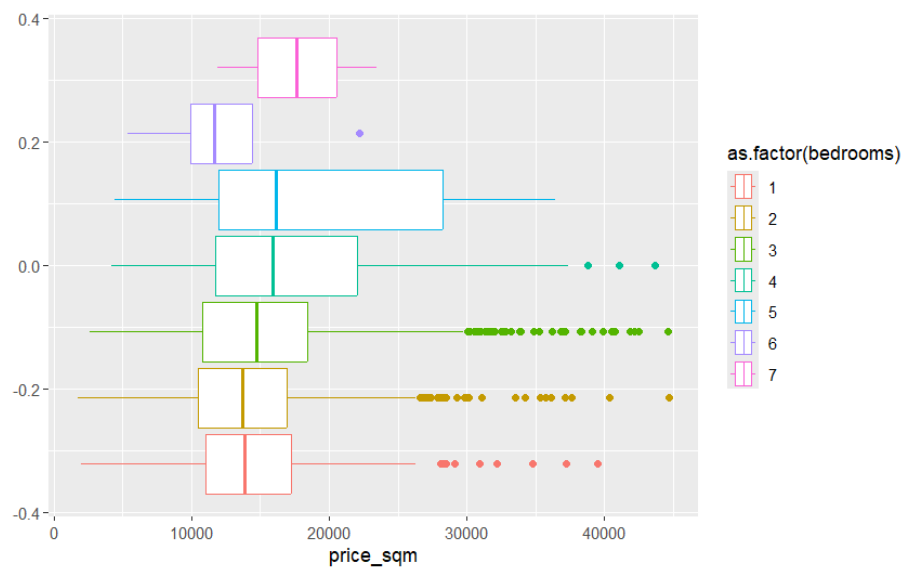


发现:

- 发现 1

板塔结合房屋单价较高, 板楼与塔楼房屋单价相似

### 4.10 卧室数量与单价关系



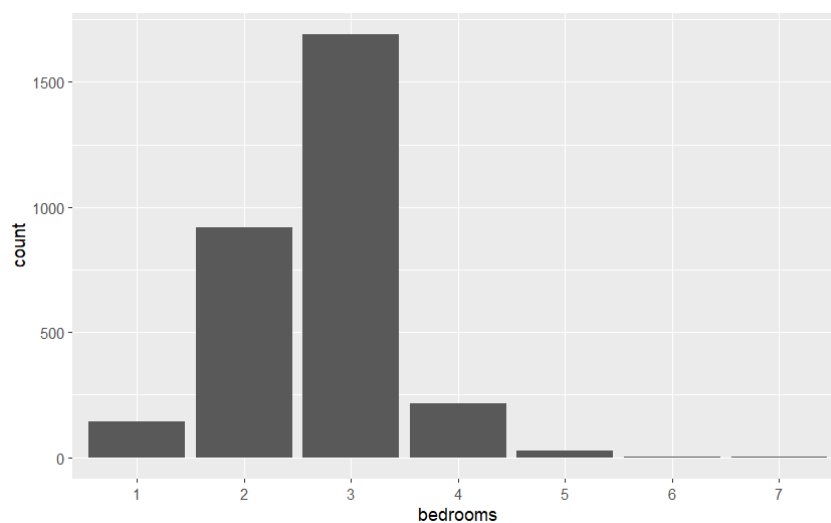
发现:

- 发现 1

卧室数量为 1、2、3 的单价接近，数量为 4 的卧室单价均价及上三分位单价较高

- 发现 2

卧室数量为 5、6、7 的单价偏离较大,尤其是数量为 5 的单价上三分位单价很高,认为可能是受样本值数据或其他因素影响,卧室数量为 5、6、7



的样板较少。

#### 4.11 是否近地铁与单价关系

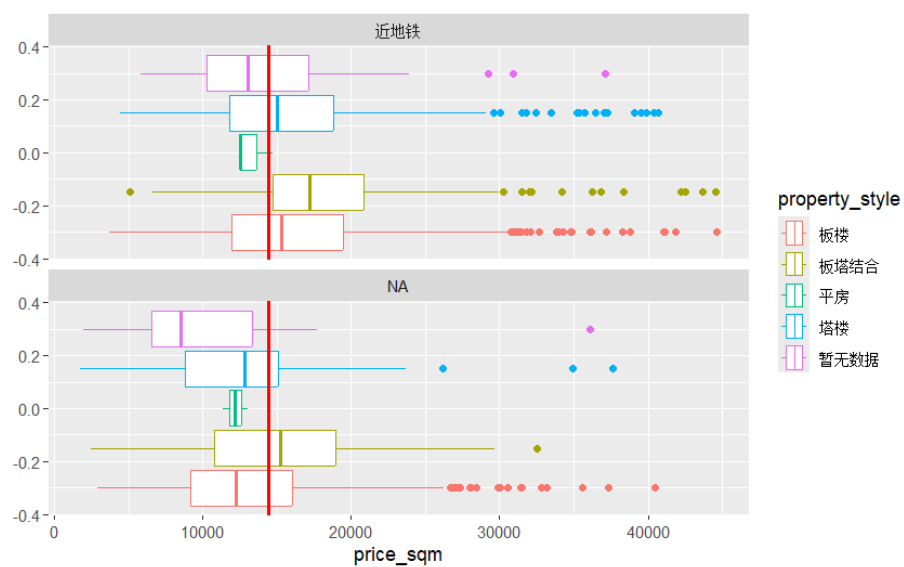


图 3: 是否近地铁与单价关系

发现:

- 发现 1

同等建筑形式下，是否近地铁对单价有明显影响，近地铁单价均价明显高于不近地铁单价均价

#### 4.12 户型与单价关系

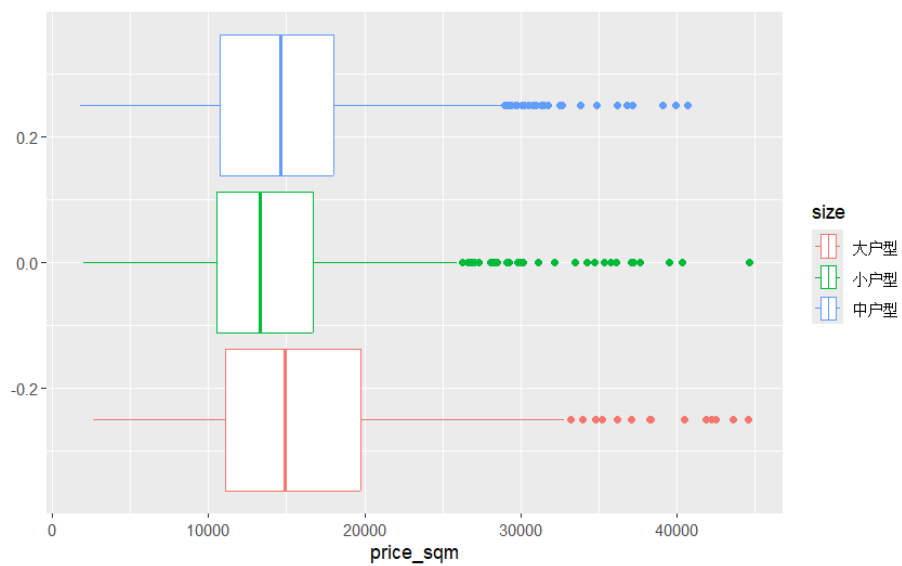
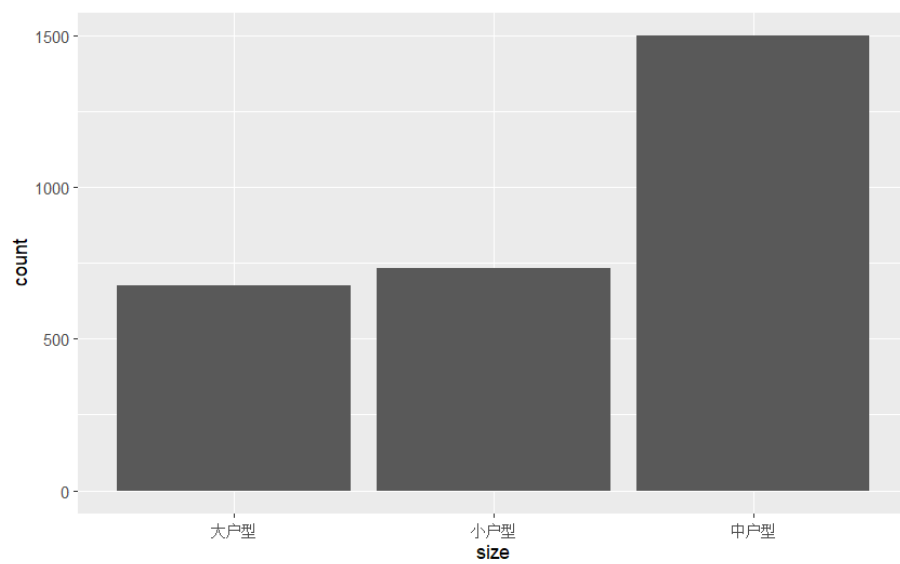


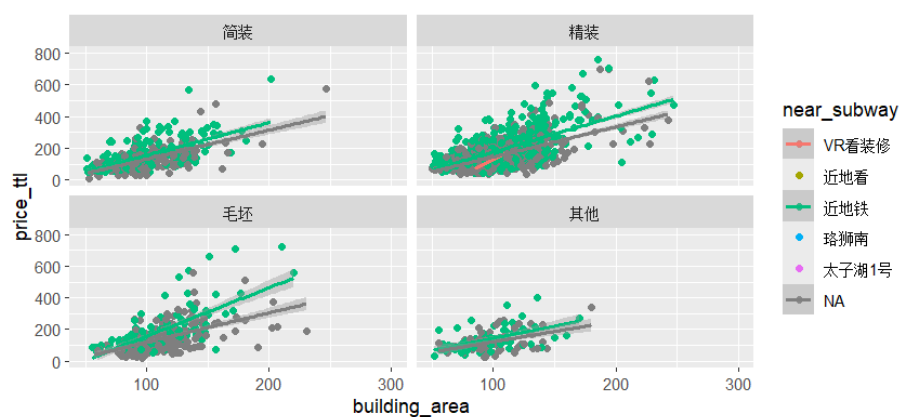
图 4: 户型与单价关系



发现:

- 发现 1  
大中户型单价接近，无明显差别
- 发现 2  
小户型均价较低，上三分位单价较低
- 发现 3  
武汉二手房以中户型居多，占比 50%，大户型与小户型相当，占比接近

#### 4.13 进一步分析装修、是否近地铁与价格关系



发现:

- 发现 1  
与前面结论一致，精装修基础单价比其他装修类型单价高
- 发现 2  
是否近地铁对毛坯房影响大，近地铁毛坯房斜率每平方单价更高，其他装修类型近地铁斜率每平方单价相近

## 4.14 尝试建模寻找显著性

```
Call:
lm(formula = price_sqm ~ directions1 + decoration + property_style +
    near_subway + size, data = wh_area_size)

Residuals:
    Min       1Q   Median       3Q      Max
-13544  -3936   -904    2790   28662

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      9619.2     4487.3   2.144  0.03222 *
directions1东      1984.4     1250.7   1.587  0.11281
directions1东北     260.0     2287.7   0.114  0.90951
directions1东南    1516.3     1133.3   1.338  0.18112
directions1南     1103.9     1027.9   1.074  0.28304
directions1西     1228.6     2111.2   0.582  0.56070
directions1西北     412.9     2391.7   0.173  0.86294
directions1西南    1671.8     1543.0   1.083  0.27878
decoration精装     2181.0       391.3   5.574 2.95e-08 ***
decoration毛坯      534.5       606.1   0.882  0.37799
decoration其他    -298.7       762.5  -0.392  0.69526
property_style板塔结合 1935.1       421.0   4.596 4.67e-06 ***
property_style平房 -1490.3     3565.6  -0.418  0.67603
property_style塔楼    442.2       403.5   1.096  0.27326
property_style暂无数据 -1404.5       970.1  -1.448  0.14788
near_subway近地铁    5074.4     4341.2   1.169  0.24263
near_subway珞狮南    1401.6     7503.8   0.187  0.85185
near_subway太子湖1号 3335.6     7503.8   0.445  0.65673
size小户型     -1983.9       448.2  -4.426 1.03e-05 ***
size中户型     -1298.6       394.6  -3.291  0.00102 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6125 on 1491 degrees of freedom
(因为不存在, 1401个观察量被删除了)
Multiple R-squared:  0.06706,    Adjusted R-squared:  0.05517
F-statistic:  5.64 on 19 and 1491 DF,  p-value: 9.719e-14
```

发现:

- 发现 1

这个线性回归模型  $R^2$  很低, 解释能力有限, 但也可以观察到某些自变量在统计上系数是显著的, 需要进一步分析改进模型。

- 发现 2

观察显著性指标, 精装修、板塔结合、小户型等显著性水平较高, 与我们前面结论有相似性。

#### 4.15 区域分布集中度



发现:

- 发现 1

二手房屋主要位于盘龙城、四新、后湖、青山、王家湾、积玉桥、百步亭，这些区域可能房屋存量较多，或人们挂牌需求较大，建议关注后续价格走势

## 5 发现总结

1. 武汉二手房均价为 15148，中位数为 14404；总价均值为 155.9 万，中位数为 137.0 万；面积均值为 100 平，中位数为 95 平；三者均呈现正偏分布，均值大于中位数，总价格与面积非常集中，总价集中于 90-160 万区间，面积集中于 100 平左右，以中户型居多，占比 50%，大户型与小户型占比接近；卧室主要三室居多，其次为二室；大厅主要为二厅，其次为一厅；市场供应以精装修为主，占比近 60%，其次为简装，

占比超 20%，再次为毛坯，占比近 15%；市场整体关注人数不多，大部分房屋关注人数为 6 人左右，部分热门房屋，关注人数 200 人以上。

2. 经过探究发现，武汉二手房市场不同楼层位置、卧室数量、房屋主要朝向对二手房单价影响不明显；房屋建筑形式、装修类型、是否近地铁、户型等对二手房单价有一定影响，其中：（1）精装修单价比其他装修类型单价高，简装、毛坯单价相近，高价房受装修类型的影响不明显；（2）板塔结合房屋单价较高，板楼与塔楼房屋单价相似；（3）近地铁房单价均价高于不近地铁房单价均价，进一步探究发现，是否近地铁对毛坯房影响非常大；（4）大中户型二手房单价接近，无明显差别，但是小户型均价较低，上三分位单价较低。
3. 进一步分析，二手房主要位于盘龙城、四新、后湖、青山、王家湾、积玉桥、百步亭，这些区域可能房屋存量较多，或人们挂牌需求较大，建议关注后续价格走势。