

# 链家武汉二手房的数据分析报告

2023281051044-莫雪美-MEM

## 目录

你的主要发现	1
二、数据准备	1
二、数据概览	2
三、探索性分析	4
1. 价格特征分析 . . . . .	4
2. 小区所属区位特征分析 . . . . .	6
3. 房屋属性分析 . . . . .	7
4. 在售房屋数量与区位的关系 . . . . .	9
5. 房屋收藏人数的影响因素 . . . . .	9
6. 房屋单价和房屋特性的关系 . . . . .	14
四、发现总结	17
# 一、环境准备	

## 你的主要发现

1. 在售的二手房主要处于远离主城区的偏远位置；
2. 房屋出售价格比较集中，大部分房源的房屋总价主要与房屋面积呈正相关；
3. 在售房源关注度受多个因素影响，主要体现在房屋所在区域为近主城区，地理位置好，价格处于均价附近的房源。

## 二、数据准备

本报告链家数据获取方式如下：

表 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	direction
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68	南
万科紫悦湾	光谷东	127.0	14613	3	2	86.91	南
东立国际	二七	75.0	15968	1	1	46.97	南
新都汇	光谷广场	188.0	15702	3	2	119.73	北
保利城一期	团结大道	182.0	17509	3	2	103.95	东南
加州橘郡	庙山	122.0	10376	3	2	117.59	南
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19	南
保利上城东区	白沙洲	193.8	16336	3	2	118.64	南
石化大院	中南丁字桥	325.0	32631	4	1	99.60	南
阳光花园	杨汊湖	192.0	17403	3	2	110.33	南

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。

二、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 2515 行。表的前 10 行示例如下：

各变量的简短信息：

```
## Rows: 2,515
## 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           <int> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
## $ livingrooms        <int> 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", "随~
## $ 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:2515      Length:2515      Min.   : 10.6      Min.   : 1771
## Class :character  Class :character  1st Qu.: 95.0      1st Qu.:10765
## Mode  :character  Mode  :character  Median : 136.0     Median :14309
##                                     Mean   : 154.8     Mean   :15110
##                                     3rd Qu.: 188.0     3rd Qu.:18213
##                                     Max.   :1380.0     Max.   :44656
## bedrooms          livingrooms      building_area      directions1
## Min.   :1.000      Min.   :0.000      Min.   : 22.77      Length:2515
## 1st Qu.:2.000      1st Qu.:1.000      1st Qu.: 84.45      Class :character
## Median :3.000      Median :2.000      Median : 95.46      Mode  :character
## Mean   :2.689      Mean   :1.706      Mean   :100.67
## 3rd Qu.:3.000      3rd Qu.:2.000      3rd Qu.:118.03
## Max.   :7.000      Max.   :4.000      Max.   :588.66
## directions2        decoration          property_t_height  property_height
## Length:2515        Length:2515          Min.   : 2.00      Length:2515
## Class :character    Class :character     1st Qu.:11.00      Class :character
## Mode  :character    Mode  :character     Median :27.00      Mode  :character
##                                     Mean   :24.05
##                                     3rd Qu.:33.00
##                                     Max.   :62.00
## property_style      followers          near_subway          if_2y
## Length:2515        Min.   : 0.000      Length:2515        Length:2515
## Class :character    1st Qu.: 1.000      Class :character    Class :character
## Mode  :character    Median : 2.000      Mode  :character    Mode  :character
##                                     Mean   : 6.326
##                                     3rd Qu.: 6.000
##                                     Max.   :262.000
## has_key              vr
## Length:2515        Length:2515
```

```
## Class :character    Class :character
## Mode  :character    Mode  :character
##
##
##
```

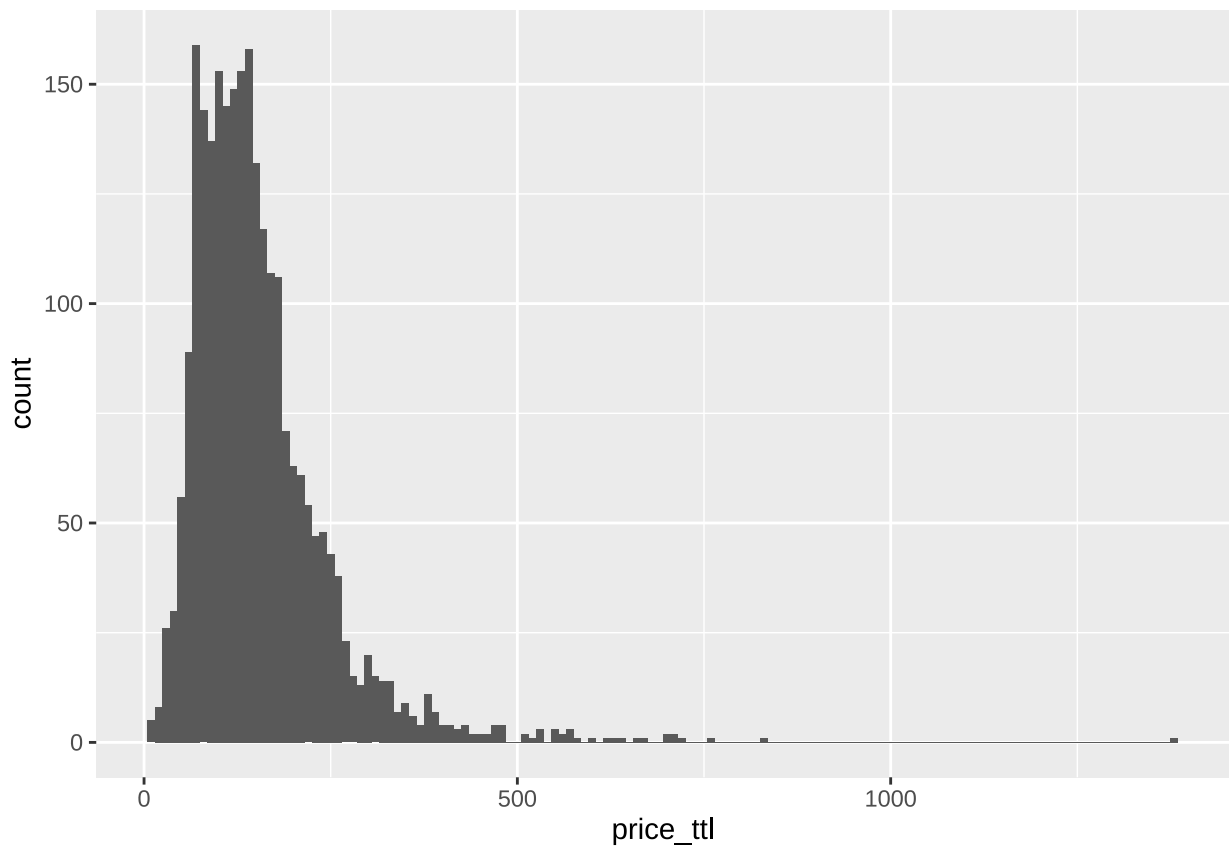
可以看到：

- 房屋最低总价为 10.6 万元，最高房屋总价为 1380 万元，严重偏离均值；房屋单价最低 1771 元/平方米，最高 44656 元/平方米，均价 15148 元/平方米与中值 14404 元/平方米较为接近
- 3000 套房平均有 2.695 个房间，平均 1.709 个客厅，房屋面积平均 100.87 平方米
- 房屋建筑均高 24.22 层，最低 2 层，最高 62 层，房屋风格有塔楼、板楼、塔板结合，房屋装修风格有：精装、简装、毛坯、其他
- 在该二手房网站的关注人数平均为 6.614，最多关注人数为 262 人，远大于平均值
- 统计了是否近地铁、房本是否满两年、中介是否有钥匙、是否支持 VR 看房 4 种外部变量因素

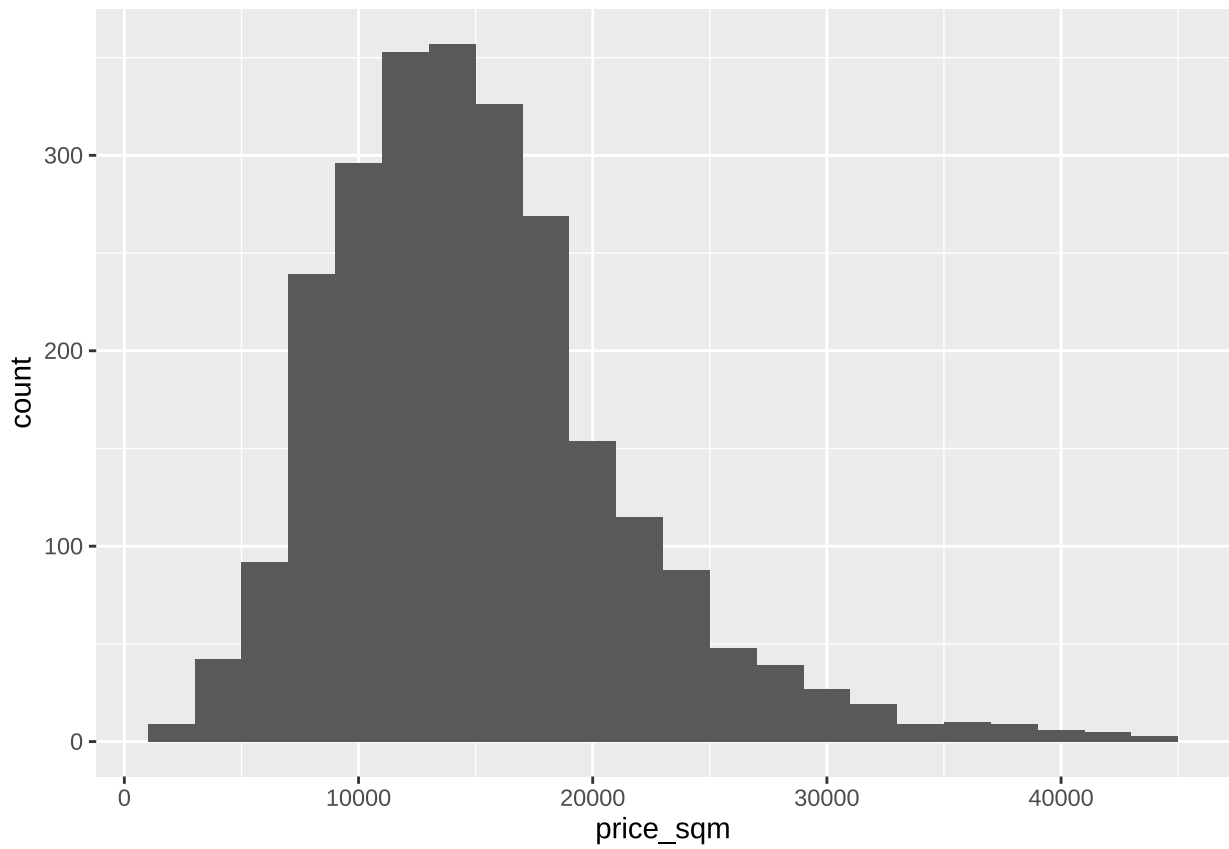
### 三、探索性分析

#### 1. 价格特征分析

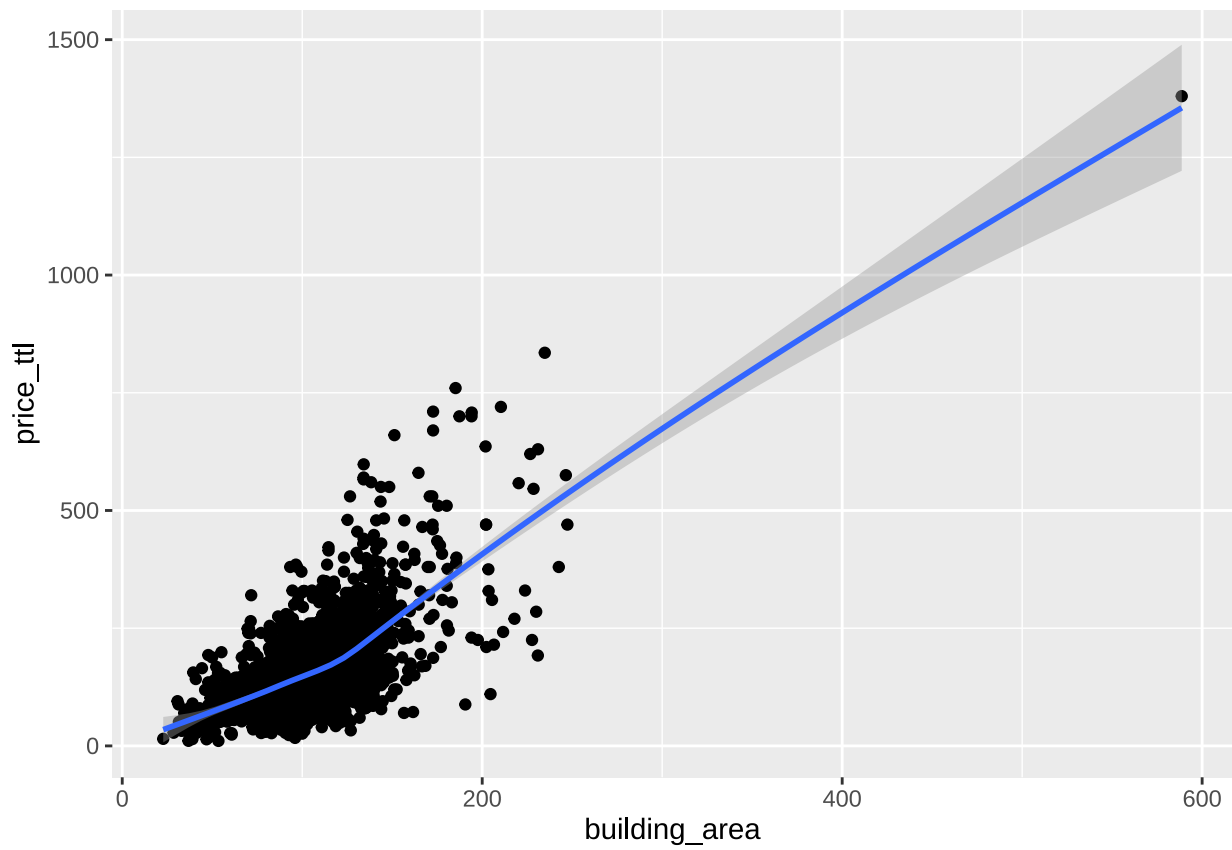
###(1) 房屋总价分布直方图



###(2) 房屋均价分布直方图



###(3) 房屋面积和房屋总价关系散点图



发现：

- 房屋总价主要集中在 50 万 ~250 万元价格区间，呈现右偏
- 房屋单价主要集中在 7000 元 ~20000 元价格区间，呈现右偏
- 房屋面积与房屋总价整体呈现正相关，房屋面积约小于 150 平方米时，二者相关性强，大于约 150 平方米时相关性减弱。

2. 小区所属区位特征分析

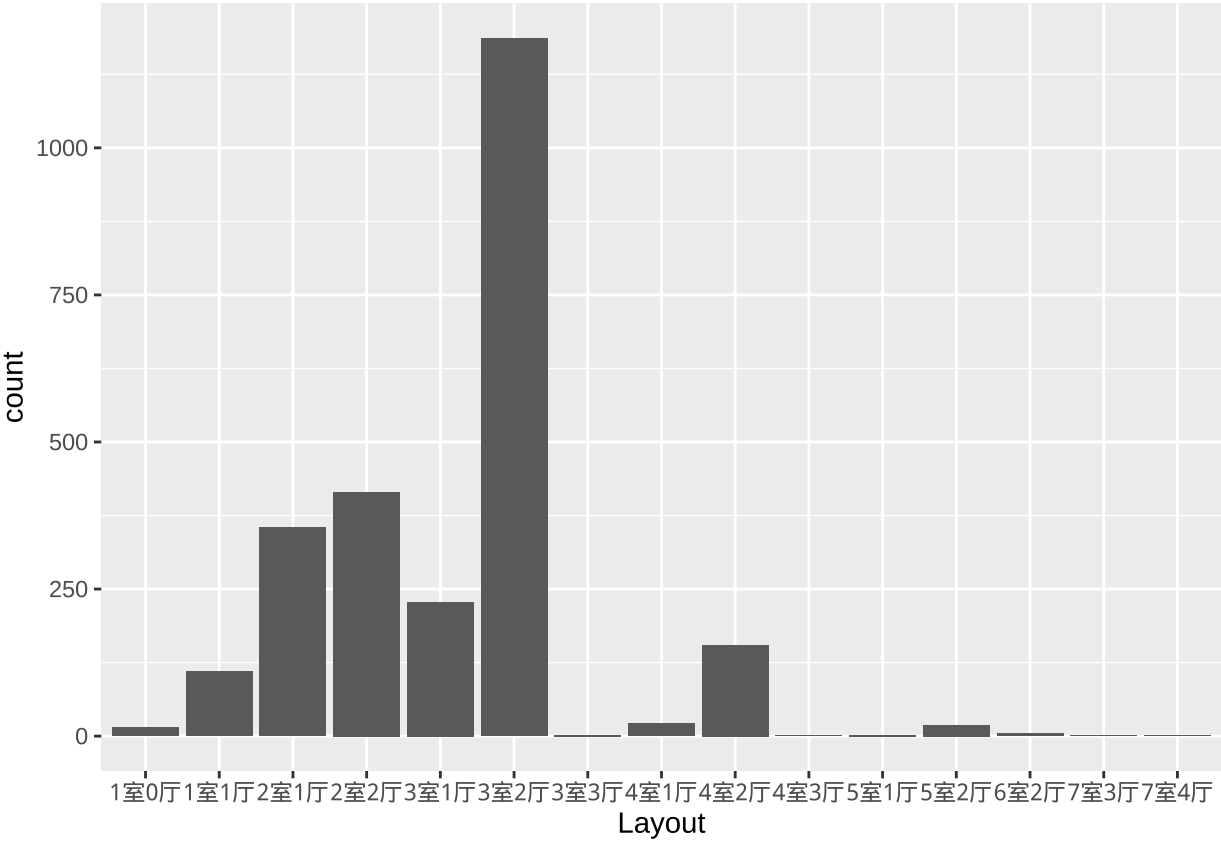
小区名称	在售房屋数	所在区域
东立国际	16	江岸区
恒大名都	13	洪山区
锦绣龙城	13	江夏区
朗诗里程	13	黄陂区
保利中央公馆	12	江夏区
阳光 100 大湖第	12	洪山区
保利华都	11	江夏区
水域天际	11	洪山区
保利城一期	10	洪山区
保利新武昌 K4 地块	10	江岸区

发现：

- 统计在售的二手房，在售数量前 10 的小区均大于等于 12 套房在售
- 在售数量前十的小区大部分位于洪山区、江夏区，地处武汉远城区

3. 房屋属性分析

###(1) 户型数据分布条形图



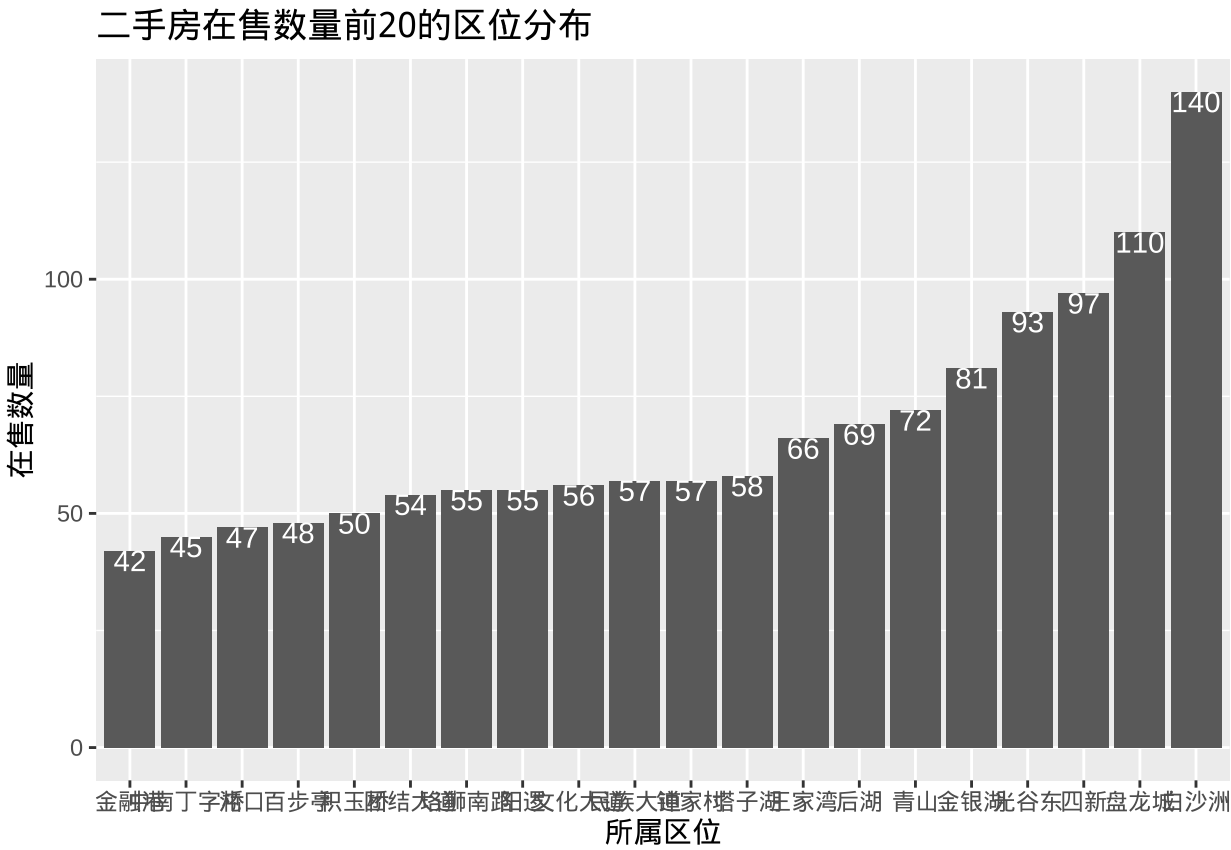
###(2) 朝向及装修分布条形图





- 在售房屋户型中，3 室 2 厅在售共 1186 套远超其他户型，其次是 2 室 2 厅、2 室 1 厅
- 在售房屋以朝南、南北朝, 共占比 81.1%；精装修为主，占比 58.7%
- 出售房屋类型以板楼为主，板楼中等楼层出售最多，共计 643 套，占比 25.6%

4. 在售房屋数量与区位的关系



发现：

- 统计在售前 20 名的区位，白沙洲、盘龙城、新四、光谷东、金银湖排名前五均属于远离主城区位置，与前面分析的远离主城区小区出售数量较大相呼应。

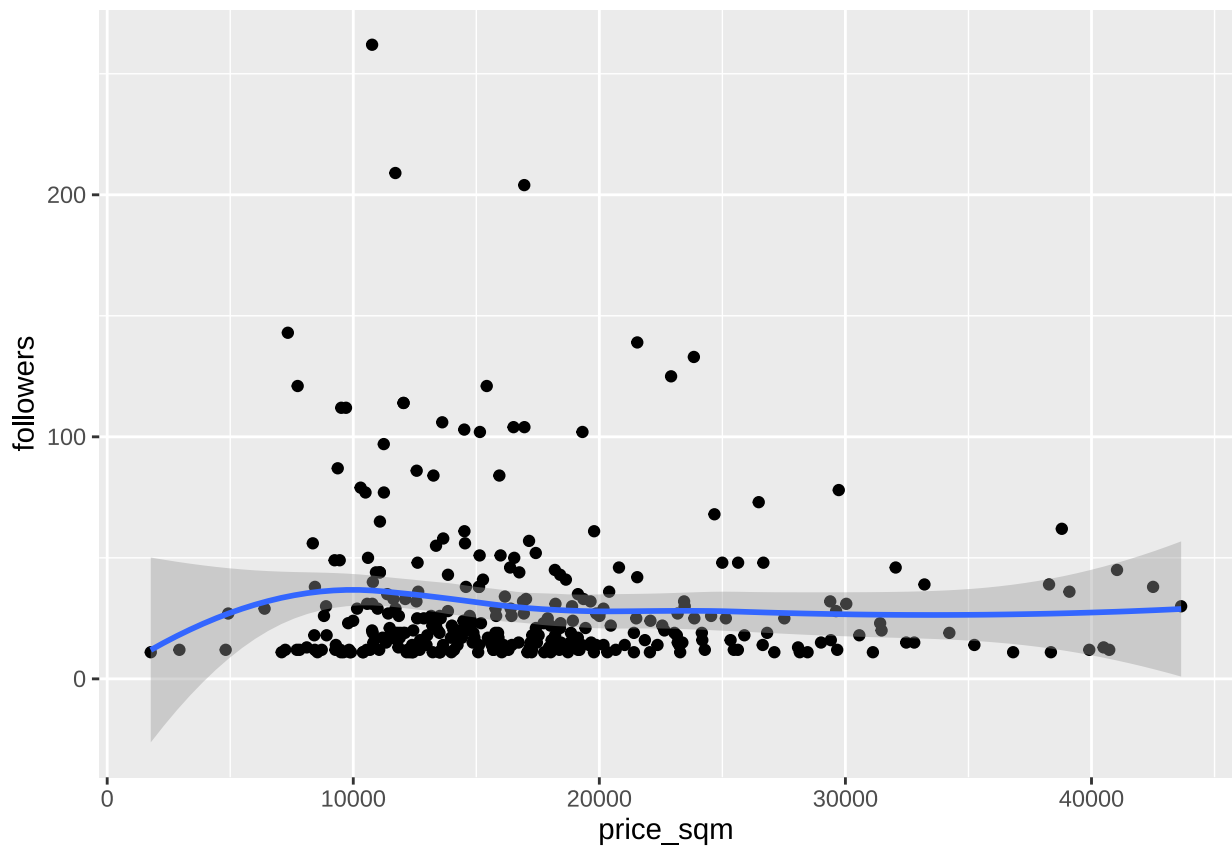
5. 房屋收藏人数的影响因素

###(1) 房屋单价与收藏人数的关系

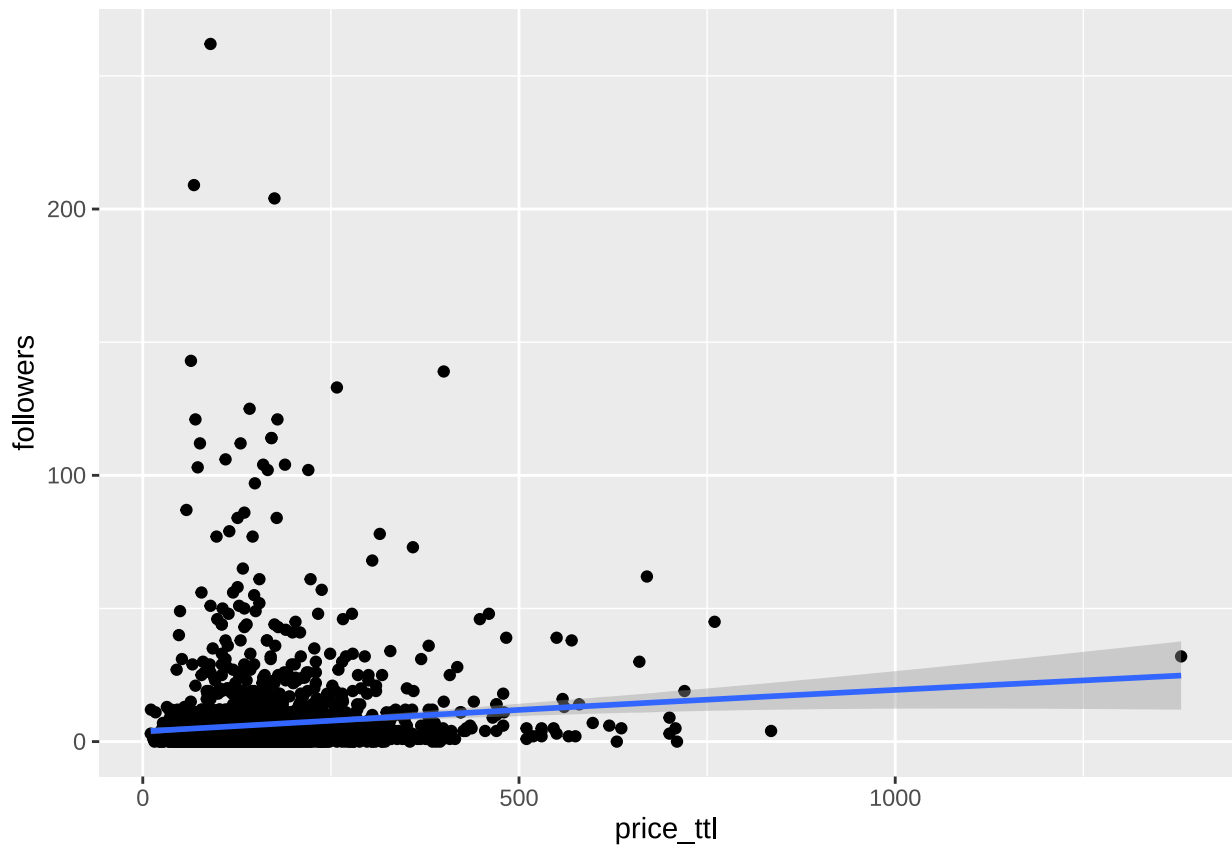
筛选出收藏人数大于 10 人的房屋数据并求其房屋单价均值

## [1] 17360.59

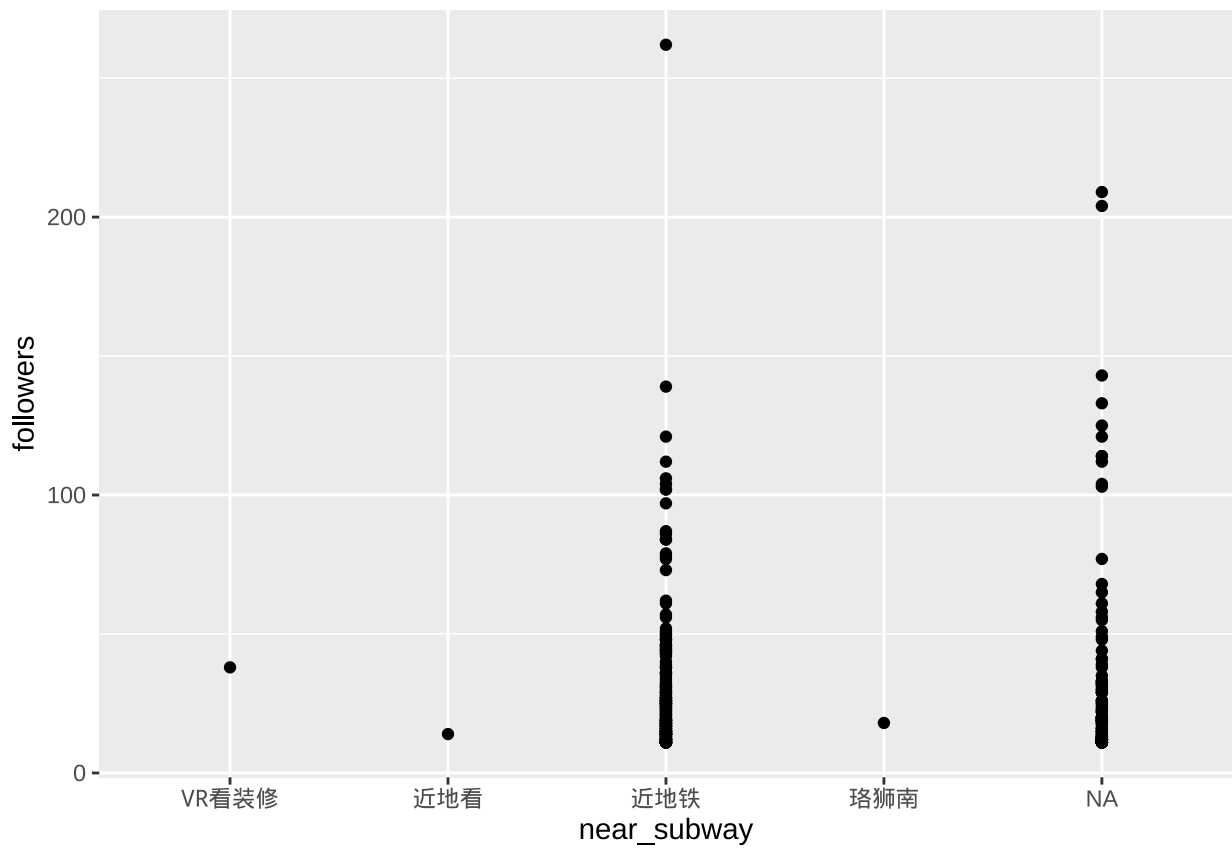
###(2) 房屋单价与关注人数的关系分布



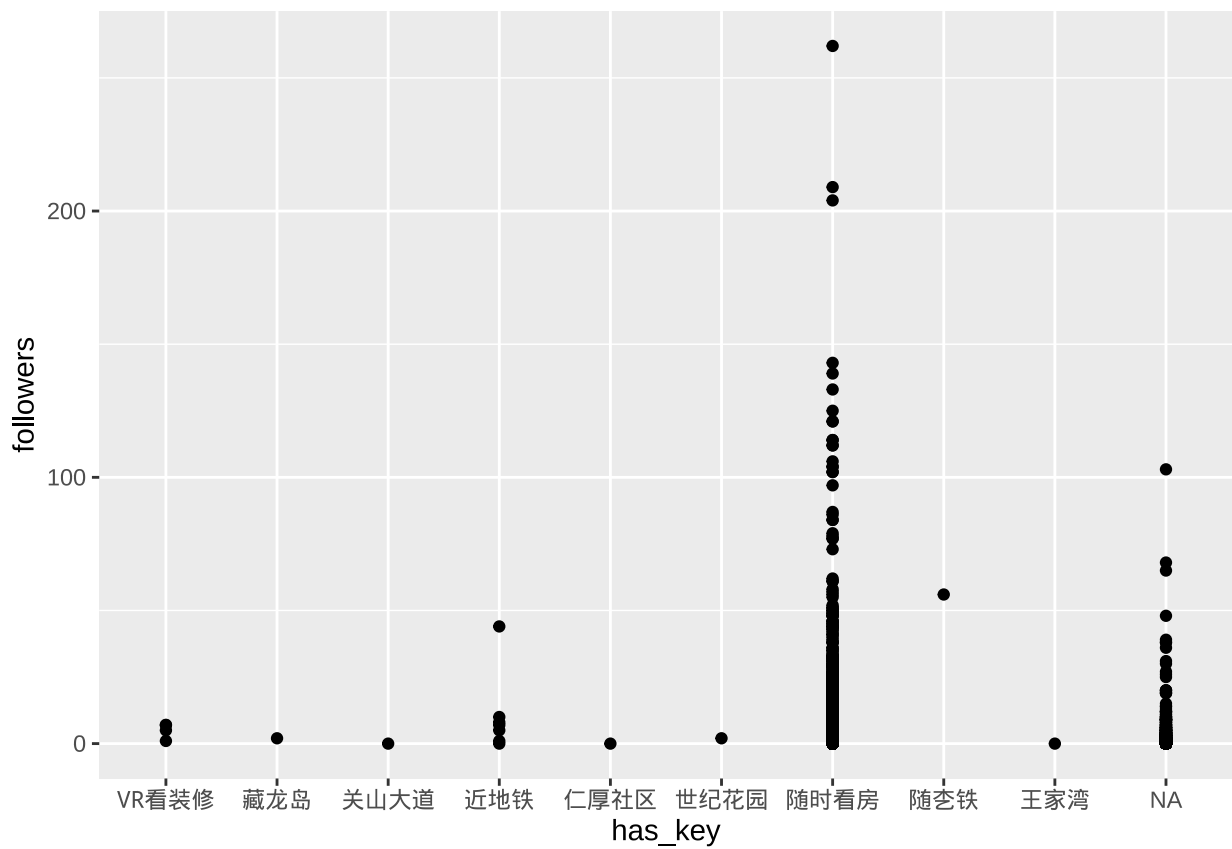
###(2) 房屋总价与关注人数的关系分布



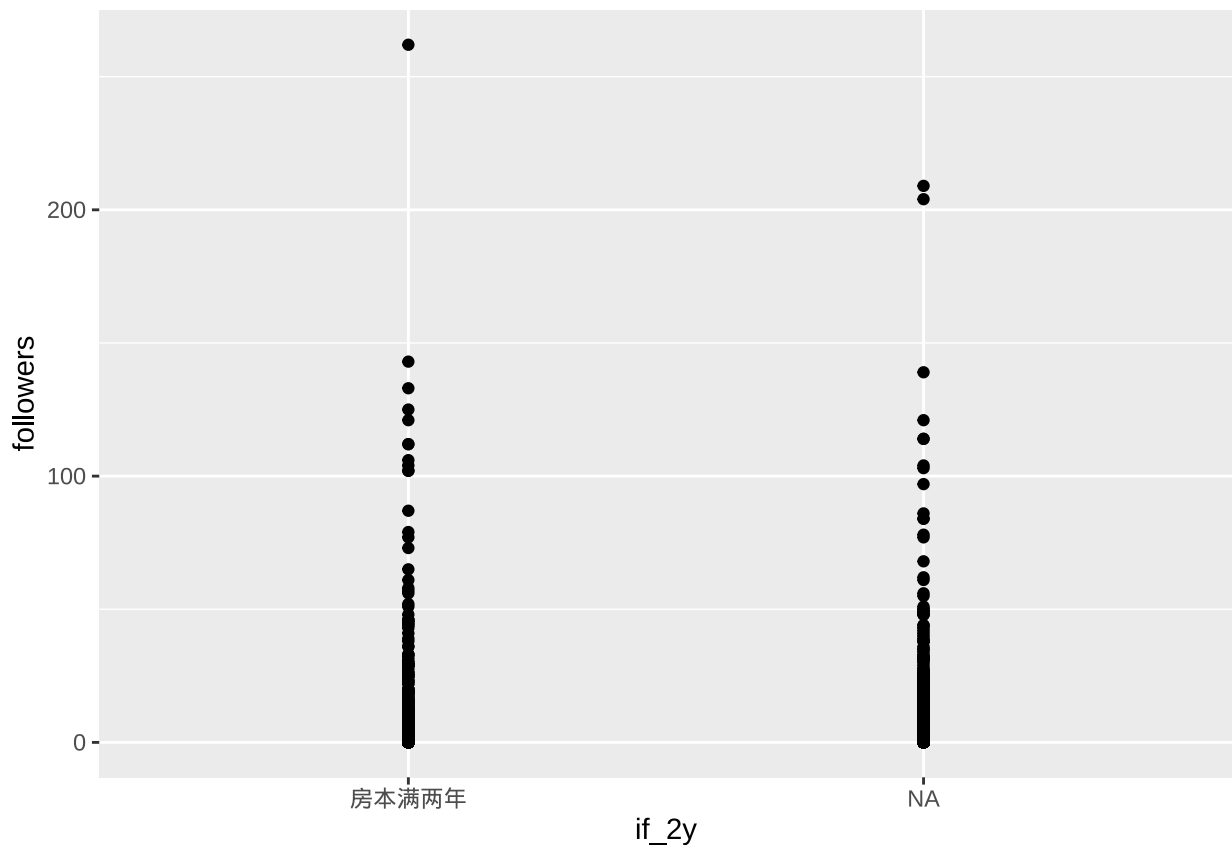
###(3) 是否近地铁与关注人数的关系分布



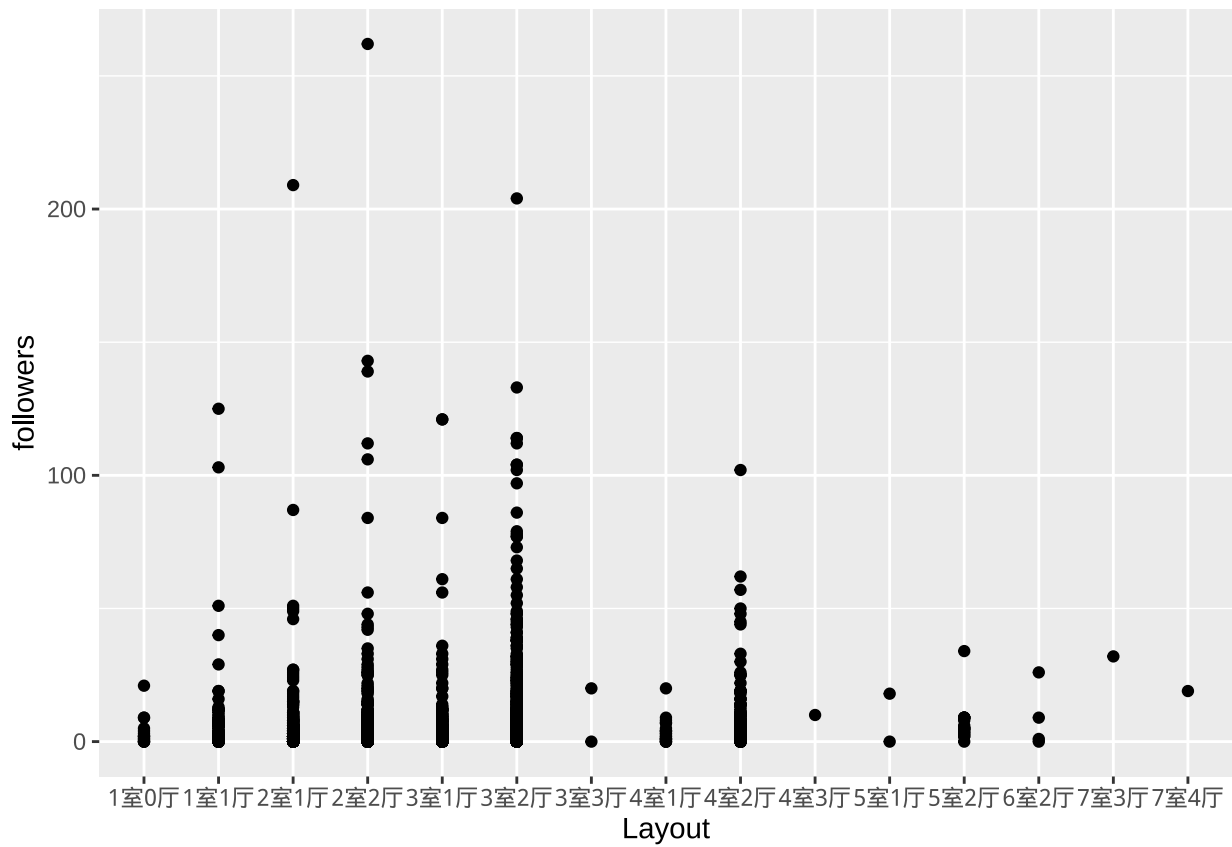
### ###(4) 是否支持随时看房与关注人数的关系分布



###(5) 是否房本满 2 年与关注人数的关系分布



###(6) 户型与关注人数的关系分布



###(7) 装修与关注人数的关系分布

decoration	followers_decoration_sum
精装	9052
简装	3650
毛坯	2786
其他	423

###(8) 装修与关注人数的关系分布

Layout	followers_sum
3 室 2 厅	7683
2 室 2 厅	2677
2 室 1 厅	1712
3 室 1 厅	1449
4 室 2 厅	1257
1 室 1 厅	726
5 室 2 厅	131
4 室 1 厅	82
1 室 0 厅	59
6 室 2 厅	36
7 室 3 厅	32
3 室 3 厅	20
7 室 4 厅	19
5 室 1 厅	18
4 室 3 厅	10

###(9) 所处区域关注度

收藏数量前 10 的房源数据分析

	property_name	followers	所处行政区	price_sqm	price_ttl
1788	十里和府	262	汉阳区	10765	90
1902	阜华领秀中南	209	武昌区	11711	68
431	米兰映象	204	洪山区	16948	175
1675	兴和嘉园	143	江汉区	7342	64
1923	万达公馆	139	武昌区	21542	400
1900	中建大公馆一期	133	洪山区	23841	258
2510	联投时代中心	125	武昌区	22915	142
836	枫丹白鹭	121	江岸区	15424	179
1551	新力帝泊湾	121	黄陂区	7740	70
1660	联投花山郡一期（香颂）	114	洪山区	12043	171

计算关注度前 10 的房源平均单价和平均总价

```
## [1] 15027.1
```

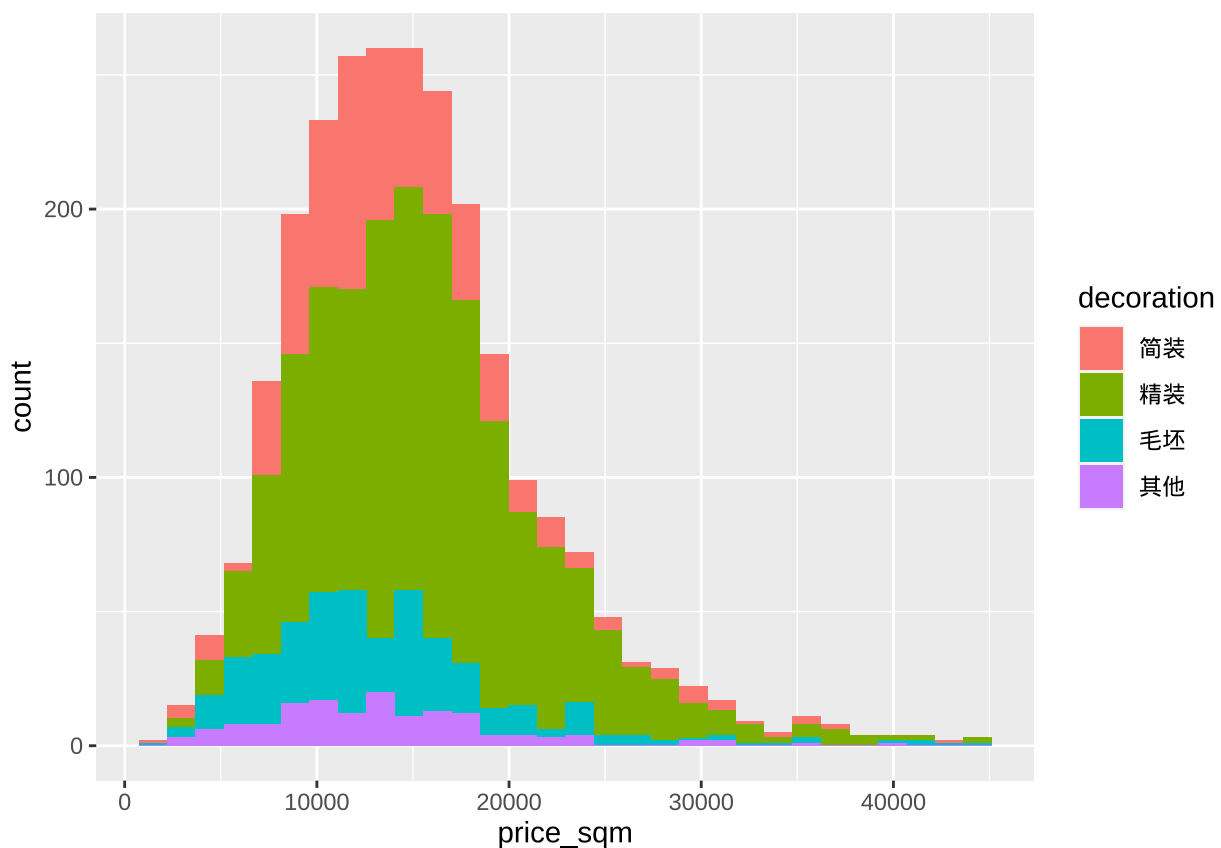
```
## [1] 161.7
```

发现：

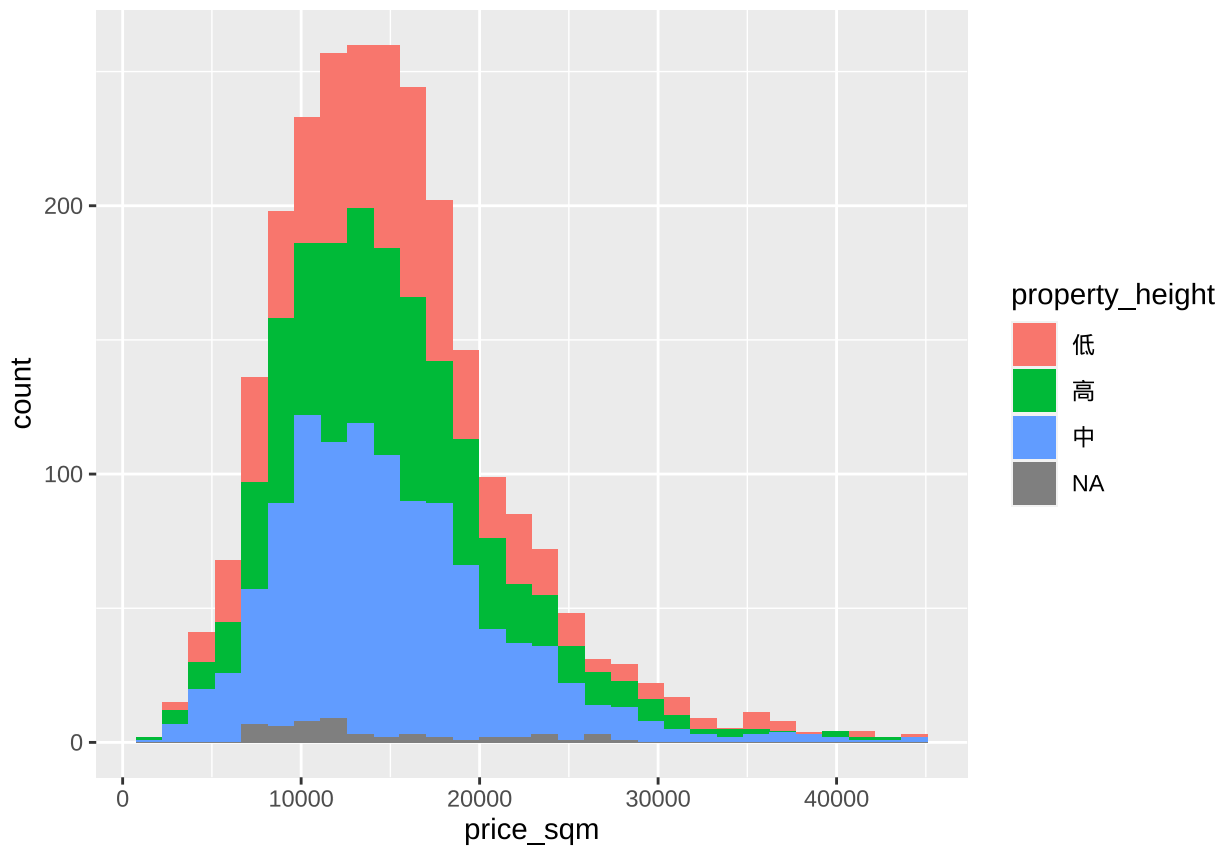
- 价格特征：被收藏较多（定义为收藏人数  $> 10$  人）的房屋单价均值为 17360.59 元/平方米，大部分人关注房价为 7000 元 ~15000 元每平方米的单价区间；大部分收藏房屋总价集中约在 400 万元以下，收藏人数与房屋总价无明显关联关系；
- 数据集中绝大部分房屋均近地铁、房本满两年，故均非关注人数的影响因素；支持随时看房的收藏数会更高；
- 收藏数量最多的装修状况是精装修类；
- 关注度最高的二手房户型依次为 3 室 2 厅 9475 人；2 室 2 厅 3660 人；2 室 1 厅 2127 人
- 分析收藏人数最高的前 10 套房源所处区域多为近主城区的区域，

## 6. 房屋单价和房屋特性的关系

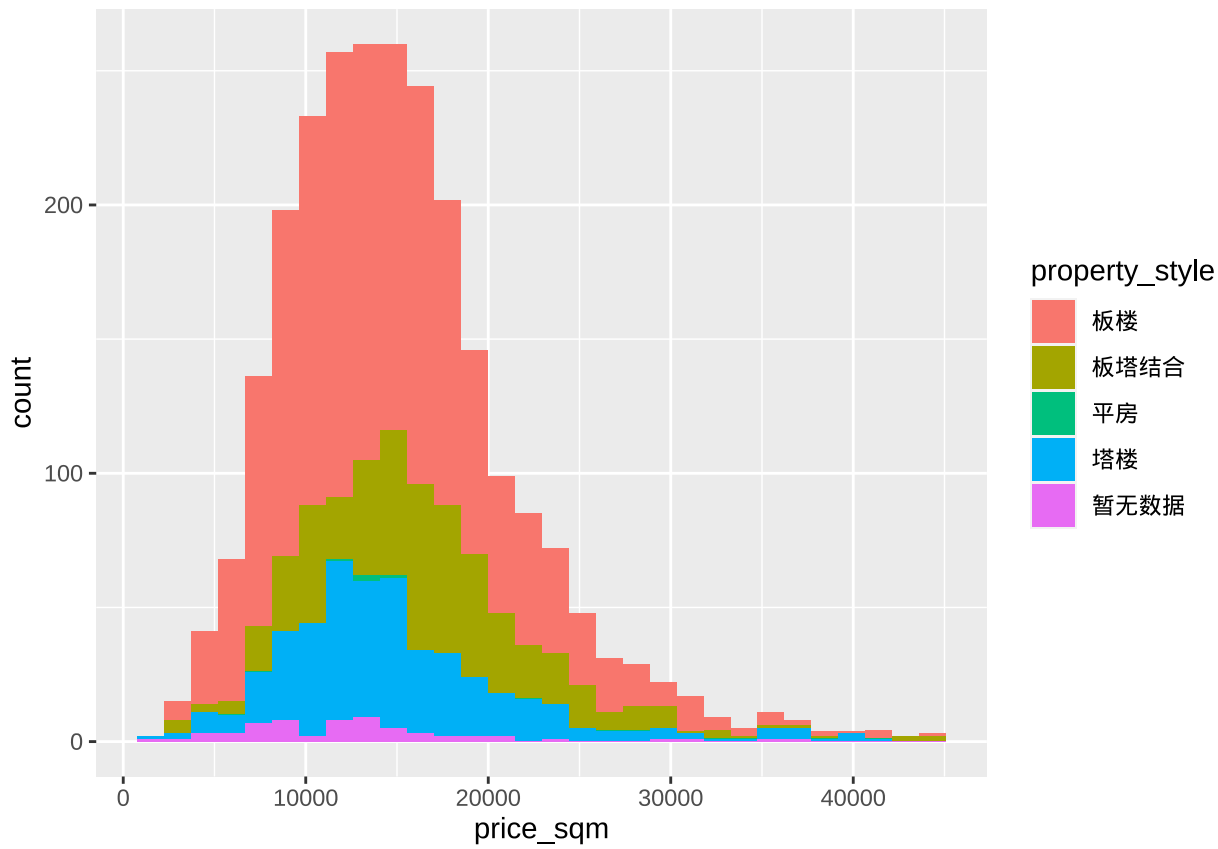
###(1) 房屋单价及对应装修状况分布直方图



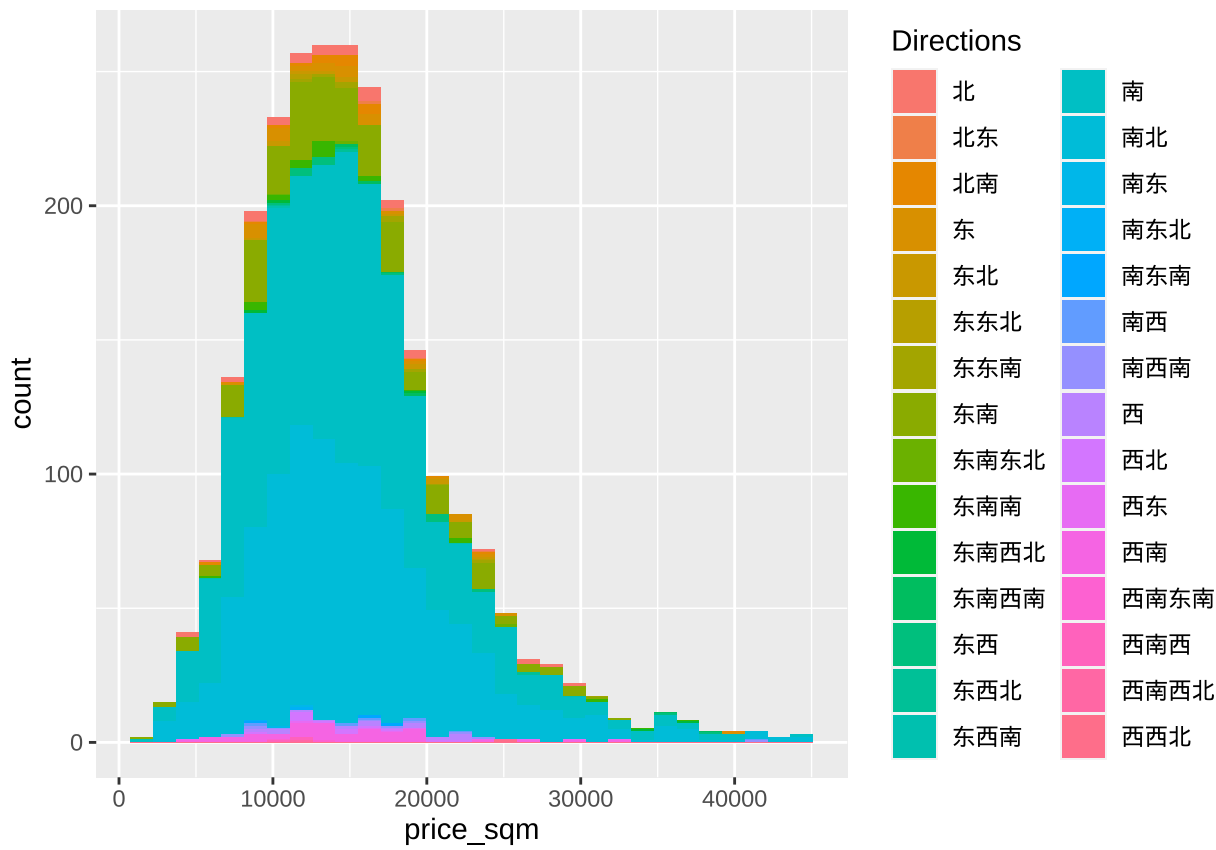
###(2) 房屋单价及对应楼层分布直方图



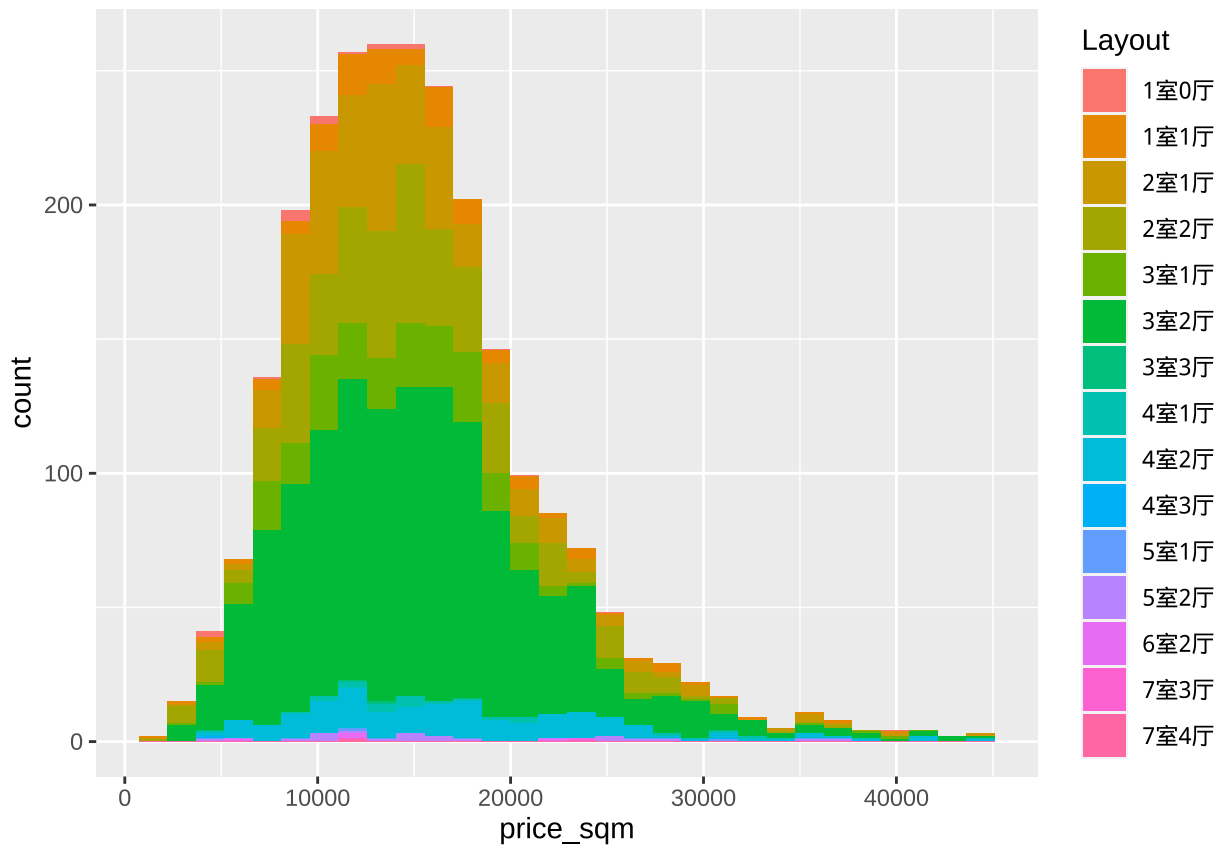
###(3) 房屋单价及对应建筑形式分布直方图



###(4) 房屋单价及对应朝向分布直方图



###(5) 房屋单价及对应户型分布直方图



###(6) 同时满足单价条件最优时与整体数据对比



```
##   Min_price Max_price Ave_price
## 1      4720     32797  15757.82
```

发现：

- 由直方图颜色分类所占面积大小和右偏程度可知，装修越好房屋单价越高；中楼层的房价比其他楼层高；板楼的单价比其他类型高；朝南、南北朝向房价更高；3室2厅户型单价较高；
- 综合分析多个变量时，计算其房屋单价平均值与总数据集偏离不大

---

## 四、发现总结

1. 在售房源显著特征 - 房屋价格呈现右偏，有较大占比在售房源价格较高；- 多数房源远离主城区位置，在售数量前 10 的小区均大于等于 12 套房在售，大部分位于洪山区、江夏区，地处武汉远城区位置；- 3 室 2 厅在售最多共 1186 套，远超其他户型，其次是 2 室 2 厅、2 室 1 厅；- 朝向以朝南、南北朝，共占比 81.1%；- 装修状况以精装修为主，占比 58.7%；- 板楼中等楼层出售最多，共计 643 套，占比 25.6%；
2. 价格影响因素 - 房屋面积与房屋总价整体呈现正相关，房屋面积约小于 150 平方米时，二者相关性强，大于约 150 平方米时相关性减弱 - 装修越好房屋单价越高；中楼层的房价比其他楼层高；板楼的单价比其他类型高；朝南、南北朝向房价比其他朝向更高；
3. 房源关注度影响因素通过分析在售房屋收藏数据可得以下房屋受欢迎因素：- 近主城区，地理位置较好；- 精装修；- 房屋单价与总体均价相近；- 支持随时看房；- 户型为 3 室 2 厅或 2 室 2 厅
4. 购房建议：- 远离主城区的房源价格偏低 - 户型房间数多的房源房屋单价会相对更低，在意性价比可关注房屋面积更大的房源；
5. 售房建议：- 近主城区、精装修、朝南或南北朝向、中等楼层、板楼、3 室 2 厅的房屋售价可以高于平均单价 - 发布房源信息时支持随时看房可以提高关注度。