

武汉市二手房浅析

李游游

2023-10-16

你的主要发现

1. 靠近地铁的房屋数量与远离地铁的房屋数量几乎一致，但靠近地铁房屋均价比远离地铁大约高 3000 元，中位数大约高 2700 元
2. 白沙洲二手房市场相较于其他区域比较活跃
3. 有 95% 的信心武汉市二手房均价在 14922.13 元至 15374.85 元之间

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：

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, 2, 1, 2, 2, 2~
## $ building_area       <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, ~
```

Table 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68	南	北
万科紫悦湾	光谷东	127.0	14613	3	2	86.91	南	NA
东立国际	二七	75.0	15968	1	1	46.97	南	NA
新都汇	光谷广场	188.0	15702	3	2	119.73	北	东
保利城一期	团结大道	182.0	17509	3	2	103.95	东南	NA
加州橘郡	庙山	122.0	10376	3	2	117.59	南	北
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19	南	NA
保利上城东区	白沙洲	193.8	16336	3	2	118.64	南	北
石化大院	中南丁字桥	325.0	32631	4	1	99.60	南	北
阳光花园	杨汊湖	192.0	17403	3	2	110.33	南	北

```
## $ directions1      <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "~
## $ directions2      <chr> "北", NA, NA, "东", NA, "北", NA, "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
##
##
##
```

可以看到：

- 房屋单价的 最小值：1771 元，最大值为：44656 元，均值为：15148 元，中位数：14404 元
- 房屋总价的 最小值：10.6 万元，最大值为：1380 万元，均值为：155.9 万元，中位数：137 万元
- 房屋价格（单价、总价）跨度非常之大，例如单价从 1771 元至 44656 元
- 随着科技的发展，VR 看装修已经非常普及，这对买方、卖房来说都节省很多成本

探索性分析

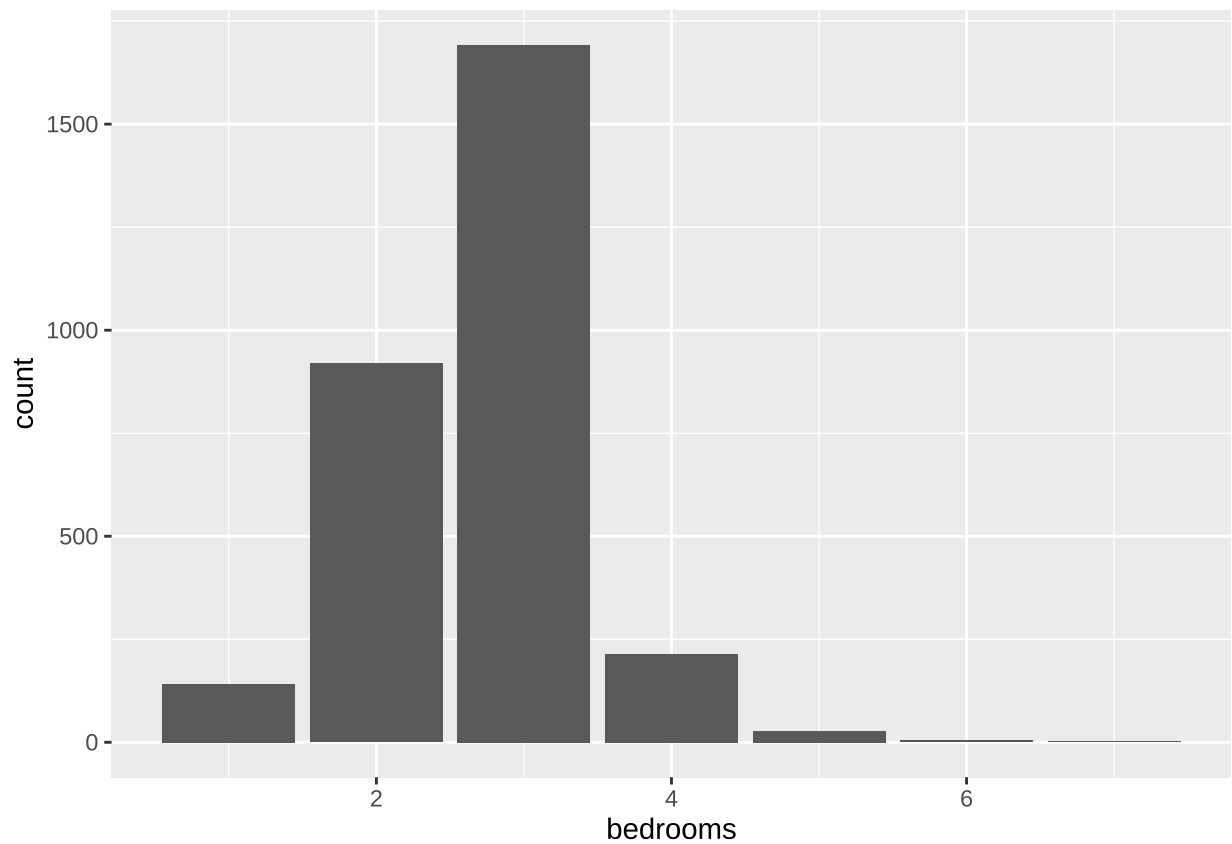
变量 near_subway（是否靠近地铁）对房价的影响

```
## # A tibble: 2 x 4
## near_subway count mean_sqm mid_sqm
## <chr> <int> <dbl> <dbl>
## 1 近地铁 1554 16624. 15622
## 2 <NA> 1441 13558. 12840
```

发现：

- 样本数据近地铁与不近地铁数据几乎持平
- 靠近地铁房屋均价比远离地铁大约高 3000 元，中位数大约高 2700 元

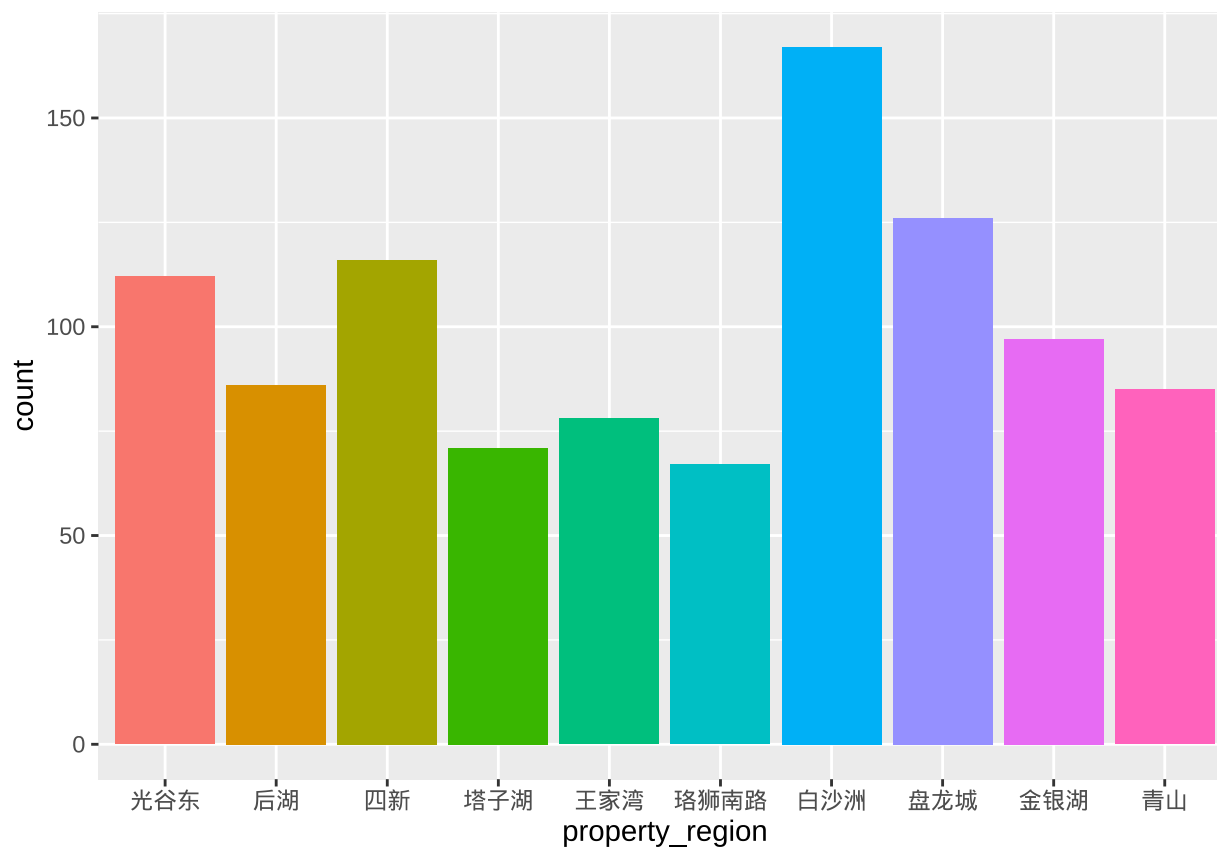
变量 bedrooms (房间数) 的数值描述与图形



发现:

- 两室、三室还是房型的主流
- 相较于两室，三室数量还是高出不少，更符合现代人的需求

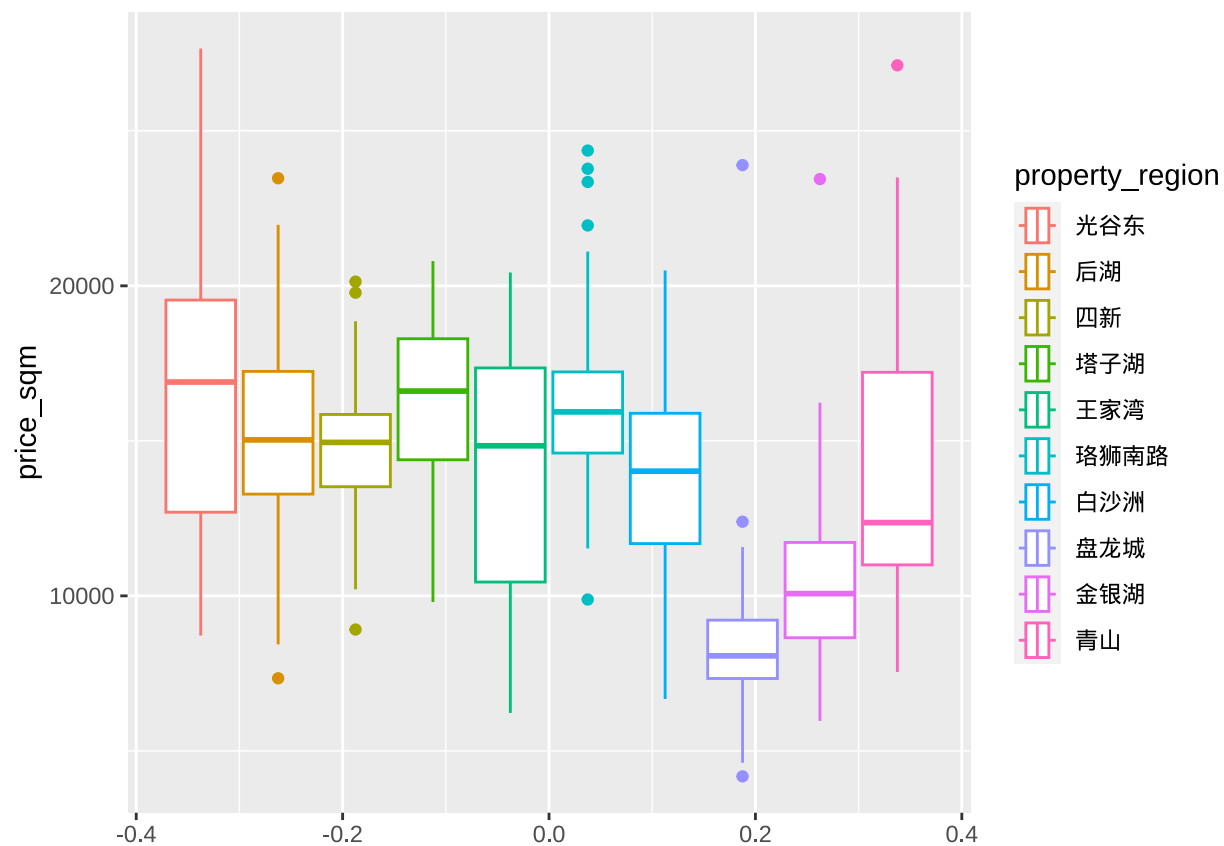
变量 property_region (区域) 的数值描述与图形



发现:

- 相较于市中心，三环周边二手房市更为活跃
- 白沙洲整个区块算是三环内比较新且比较大的居民区，房市非常活跃
- 光谷东为东湖高新发展重中之重，房市也非常活跃

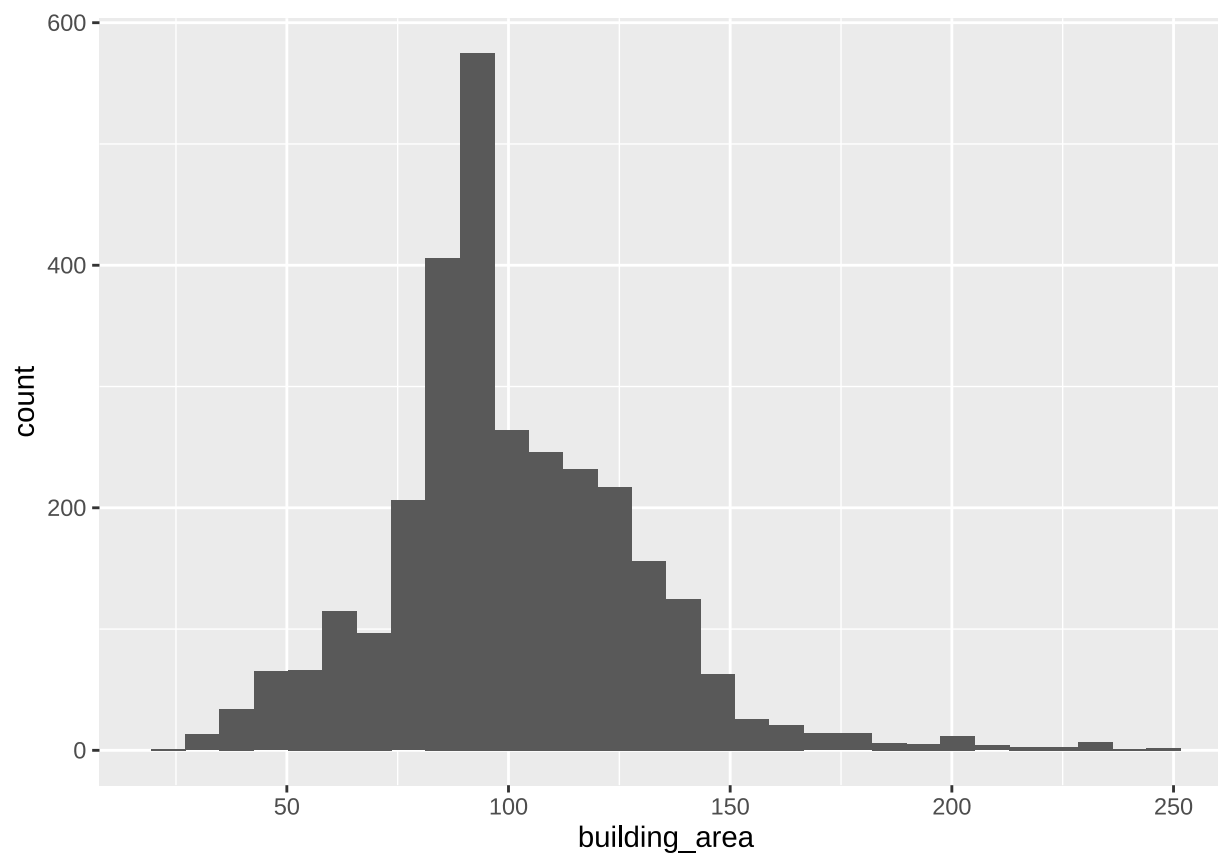
探索问题 1 数量前十区域房价分布



发现:

- 光谷东作为东湖高新发展新动力，价格活动范围更大，为新发展局域房屋较新，房价也相对较高
- 盘龙城为三环外靠近机场方向，虽然数量可观，但房价整体偏低且价格更为集中
- 青山范围较广，覆盖老城、火车站及部分长江沿岸区域，价格范围也广，但低价位房子更多

探索问题 2 房屋面积市场占比

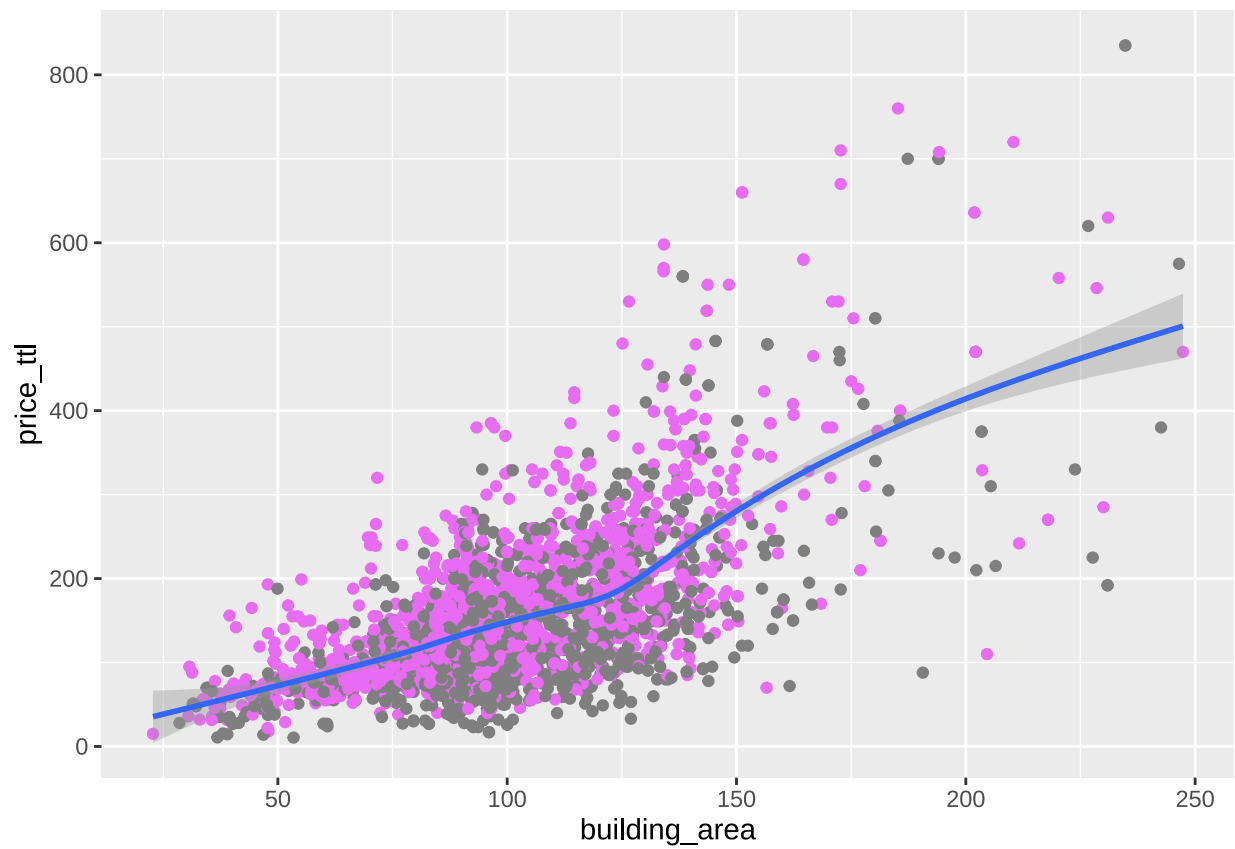


发现：

- 建筑面积在 100 平米左右数量居多，居民对小三室的市场需求

探索问题 3 房价与面积的趋势

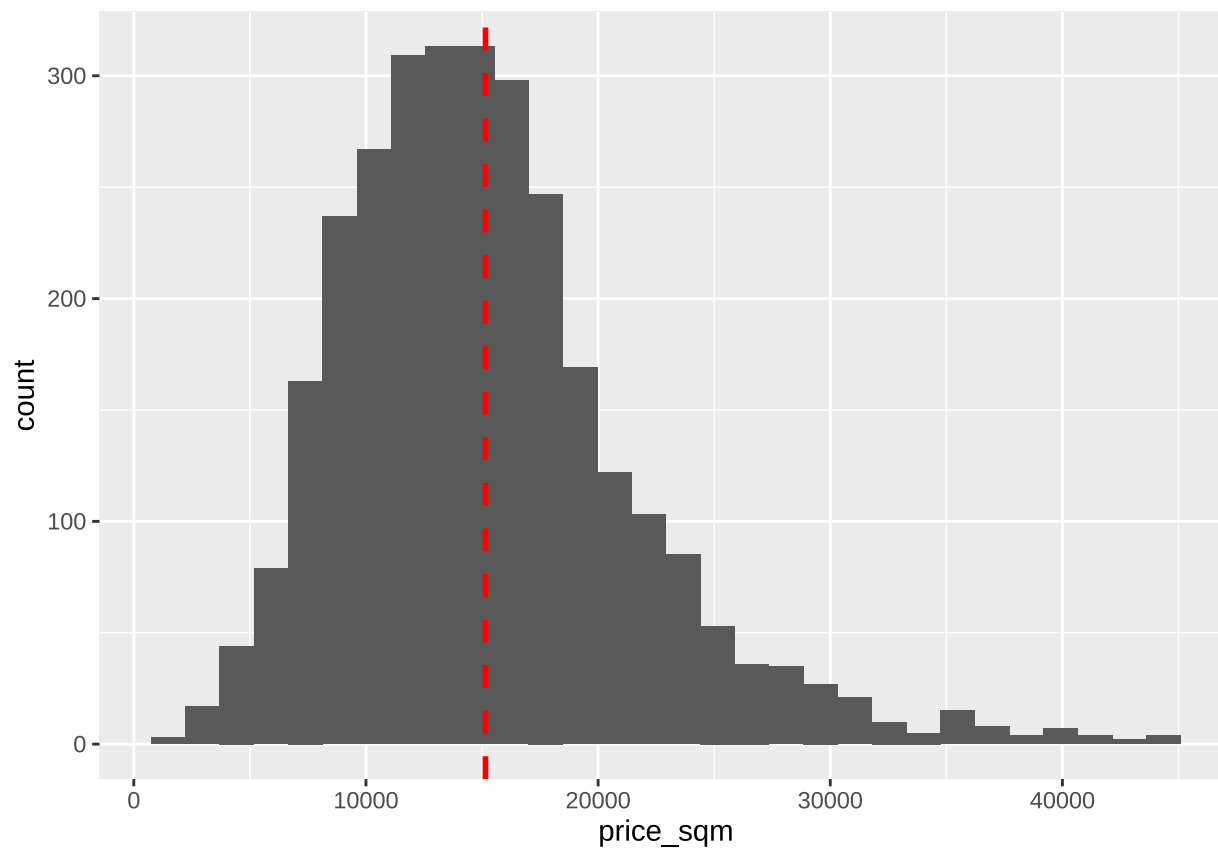
去掉数据集中 300 平米以上的数据



发现:

- 房价与面积称正相关性
- 同一个面积靠近地铁房价会更高一些

探索问题 4 整体房价预估



```
##
## One Sample t-test
##
## data:  lj$price_sqm
## t = 131.22, df = 2999, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  14922.13 15374.85
## sample estimates:
## mean of x
##  15148.49
```

发现:

- 房屋单价整体分布符合正态分布
- 通过构建置信区间, 有 95% 的信心武汉市二手房均价在 14922.13 元至 15374.85 元之间

发现总结

样本整体还算比较均匀，比较重要的一些地区都有数据出现且极少出现无效数据，根据简要的分析结果来看样本数据可以大致体现武汉总体二手房价的情况。

本分析报告还是比较浮于表面，没有更深层次的分析，比如同一地区不同因素对房价的影响，也没有做回归、聚类等相关分析。另外如果有时间因素可使报告更加丰富，目前也缺失房价随着时间变化的趋势。