

链家武汉二手房网站数据描述性统计分析

朱波

2023-10-19

目录

1	你的主要发现	2
2	数据介绍	2
3	数据概览	2
4	探索性分析	6
4.1	房价的总体分布	6
4.2	“临近地铁” 的数值描述与图形 1	7
4.3	“临近地铁” 的数值描述与图形 2	8
4.4	“建筑形式” 的数值描述与图形	10
4.5	高层住宅房价是否高于低层住宅，平均房价最高的住宅一般 是多少层？	12
4.6	随着楼层升高，房价是否会显著升高？	13
4.7	装修等级是否会对房价产生影响？	15
4.8	建筑形式与楼层是否存在相关？	17
5	发现总结	18

1 你的主要发现

1. 二手房房屋单价区间差距非常大，大多数集中在 2 万平米以下。
2. 在中低房价区间，地铁、所属区域和装修等级是影响二手房单价的重要因素之一，临近地铁的二手房单价普遍高于非临近地铁，精装高于简装高于毛坯。
3. 在房型朝向选择上，南北通透的二手房存量最大，最受消费者欢迎。
4. 在高房价区间，装修等级、建筑类型和楼层等因素对二手房差异性有重要的影响。
5. 建筑形式与总楼层存在一定关联，板塔结合形式的建筑平均楼层最高，平房最低。

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

各变量的简短信息：

表 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68
万科紫悦湾	光谷东	127.0	14613	3	2	86.91
东立国际	二七	75.0	15968	1	1	46.97
新都汇	光谷广场	188.0	15702	3	2	119.73
保利城一期	团结大道	182.0	17509	3	2	103.95
加州橘郡	庙山	122.0	10376	3	2	117.59
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19
保利上城东区	白沙洲	193.8	16336	3	2	118.75
石化大院	中南丁字桥	325.0	32631	4	1	99.15
阳光花园	杨汊湖	192.0	17403	3	2	110.35

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", "~
## $ 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
```

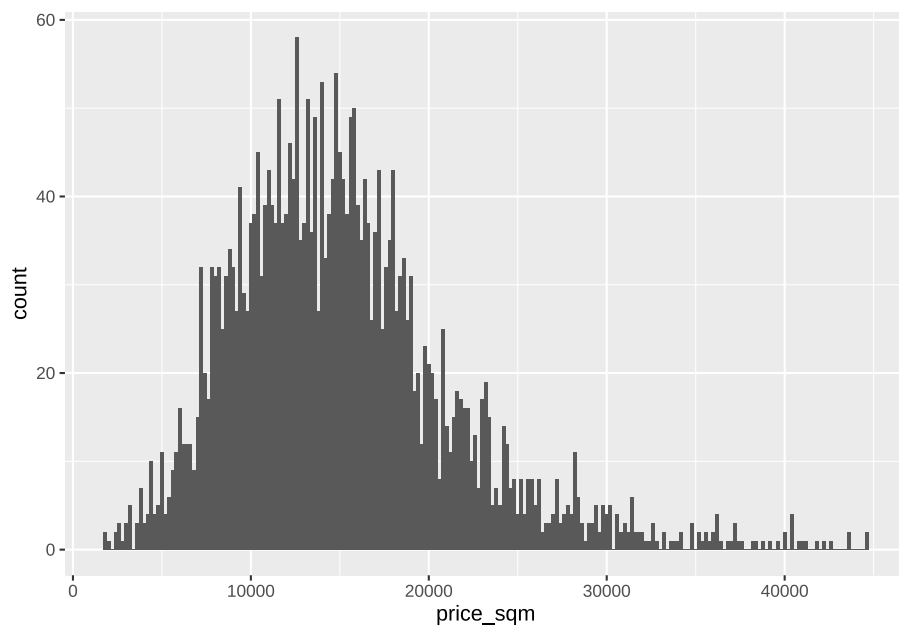
```
## Class :character   Class :character  
## Mode  :character   Mode  :character  
##  
##  
##
```

可以看到：

- 二手房房屋单价区间差距非常大，大多数集中在 2 万平方米以下，均价和中位数差别不大。
- 两室两厅和三室两厅的户型占大多数。
- 建筑面积集中在 100 平米左右。
- 大多数楼层集中在 20 至 30 层。

4 探索性分析

4.1 房价的总体分布

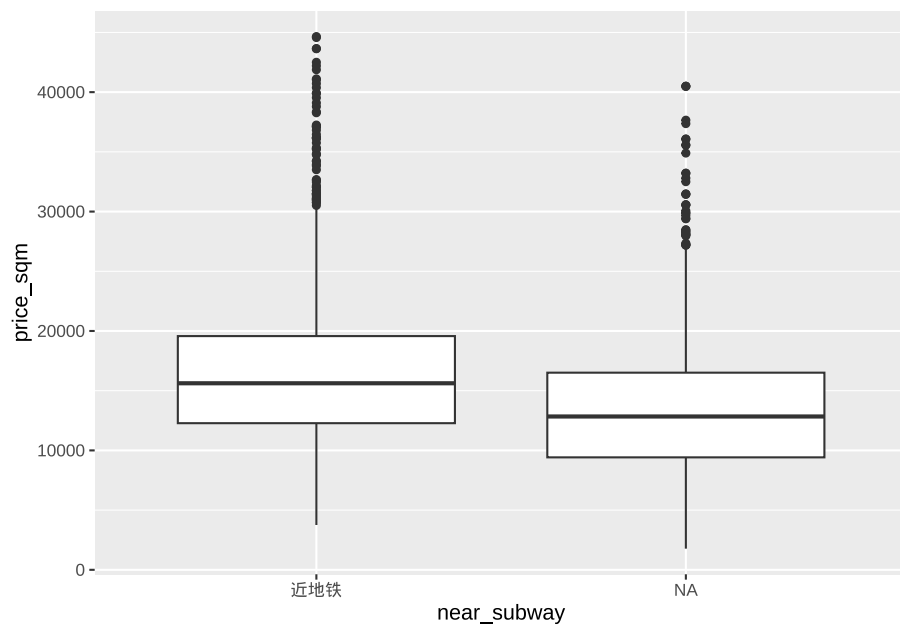


```
## # A tibble: 1 x 2
##   百分比  超高房价
##   <dbl>   <dbl>
## 1    0.95  27188.
```

发现:

- 房价服从正态分布，且集中分布在 10000 至 20000 之间，房价整体右偏，因此会针对右偏极值部分做针对性分析。
- 百分之九十五的房价低于 27187. 元/平，为简化描述，在以下的分析描述中，以 27000 元/平为超高房价的临界点。

4.2 “临近地铁”的数值描述与图形 1

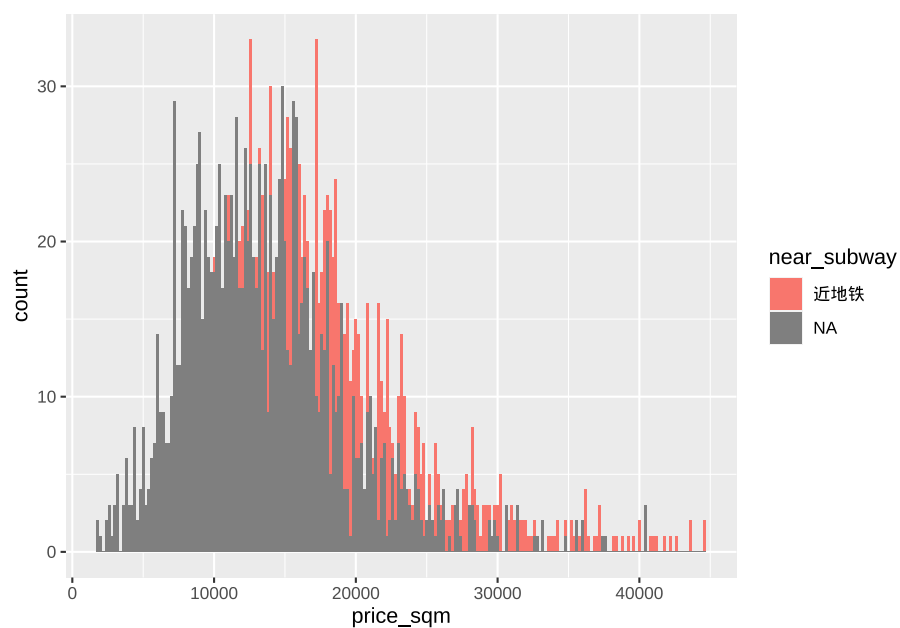


发现：

- 是否临近地铁作为影响房价的重要因素之一，临近地铁的房价普遍高于非临近地铁。

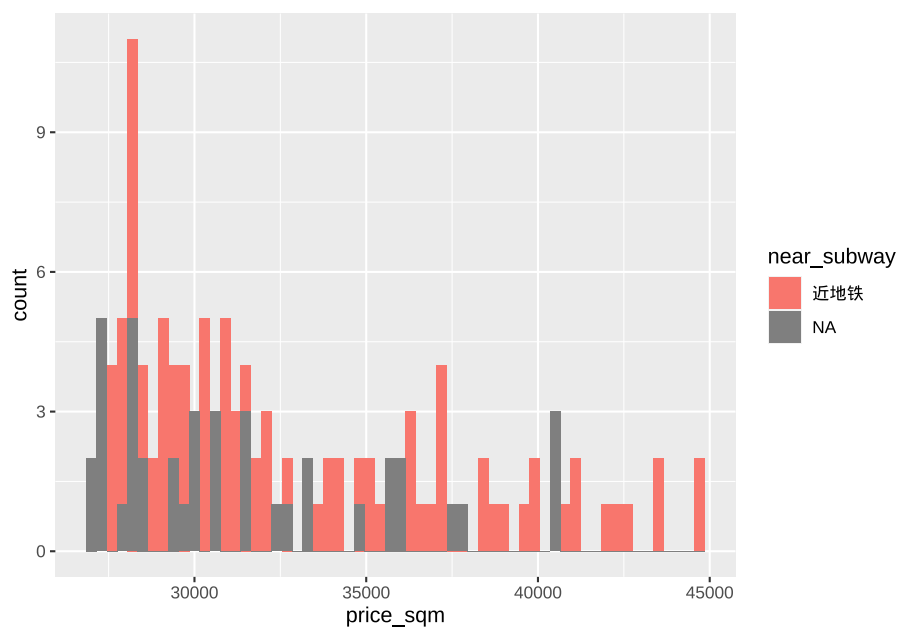
4.3 “临近地铁”的数值描述与图形 2

4.3.1 “临近地铁”的总体分布



4.3.2 超高房价区域“临近地铁”的数值统计与图形

```
## # A tibble: 2 x 2
##   near_subway count
##   <chr>         <int>
## 1 近地铁         113
## 2 <NA>          41
```

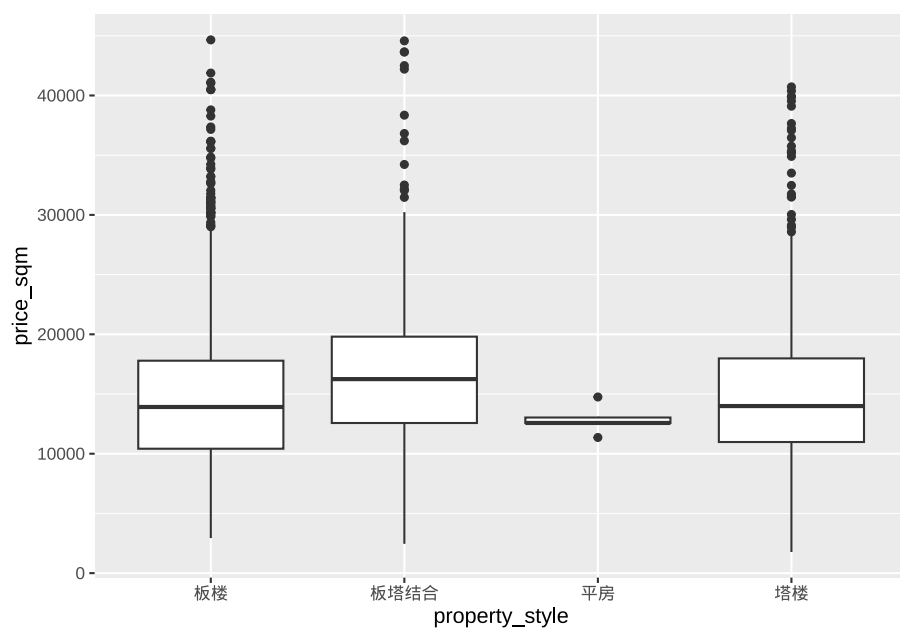



发现：

- 在高房价区间，临近地铁的二手房数量多于非临近地铁；在低房价区间，非临近地铁二手房数量多于临近地铁。
- 在均价 15000 元/平左右选择最为丰富，并且既可以选择临近地铁也可以选择非临近地铁。
- 在超高房价区间，临近地铁的二手房数量远多于非临近地铁，且主要集中在 30000 元/平附件。

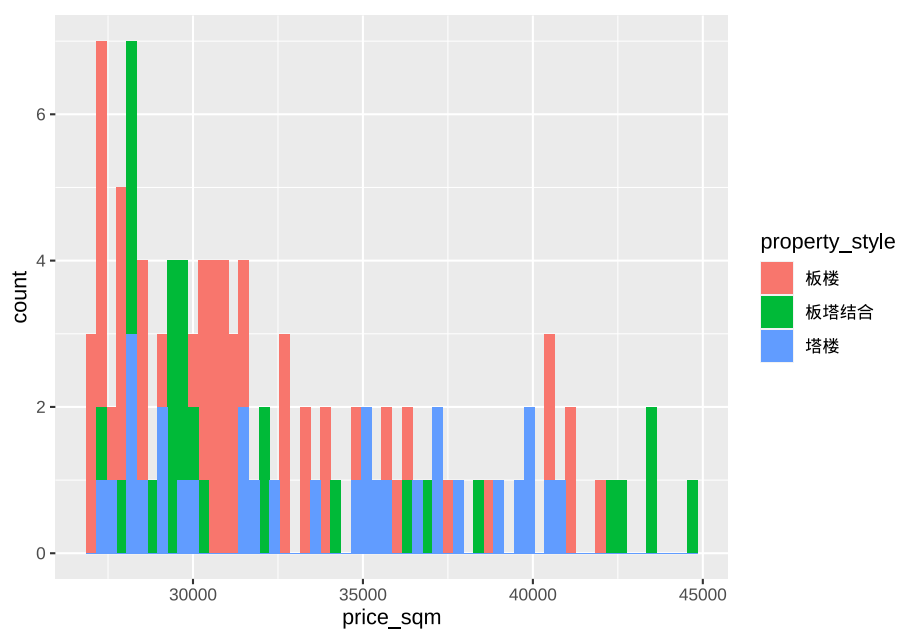
4.4 “建筑形式”的数值描述与图形

4.4.1 “建筑形式”的总体箱型图



4.4.2 超高房价区域，“建筑形式”的数量统计与分布

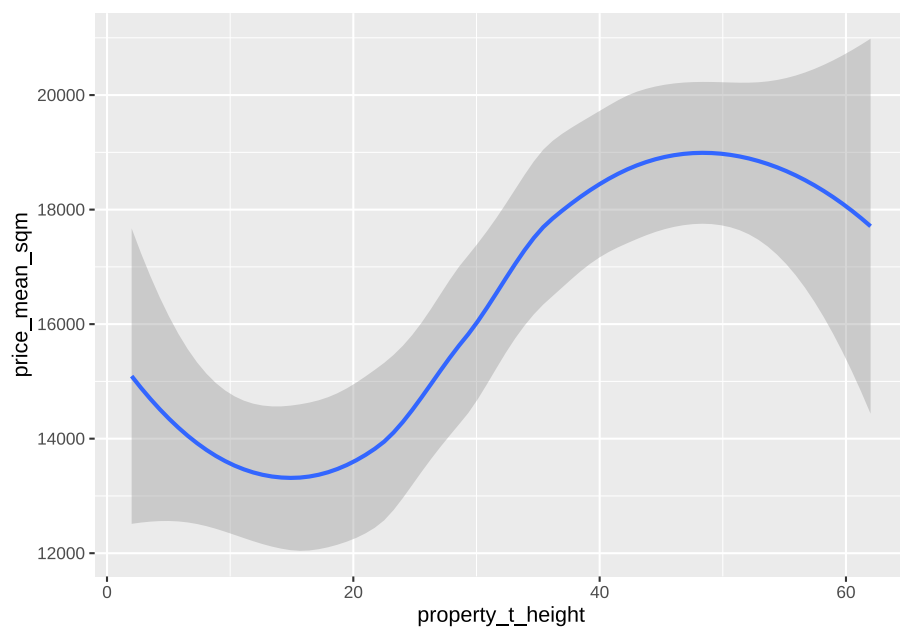
```
## # A tibble: 3 x 2
##   property_style count
##   <chr>           <int>
## 1 塔楼             30
## 2 板塔结合        37
## 3 板楼            82
```



发现:

- 平房作为一种落后的建筑形式，在二手房市场的存量很小且均价较低，被市场淘汰是趋势。
- 板塔结合作为最受欢迎的建筑形式，平均房价最高，塔楼和板楼价格区间近似。
- 在超高房价区间，板楼是最常见的建筑形式，远多于塔楼和板塔结合，塔楼和板塔结合形式数量相当，但在极高房价区间，尤其是超过 42000 元/平以后，板塔结合的建筑形式为主要房型。

4.5 高层住宅房价是否高于低层住宅，平均房价最高的住宅一般是多少层？

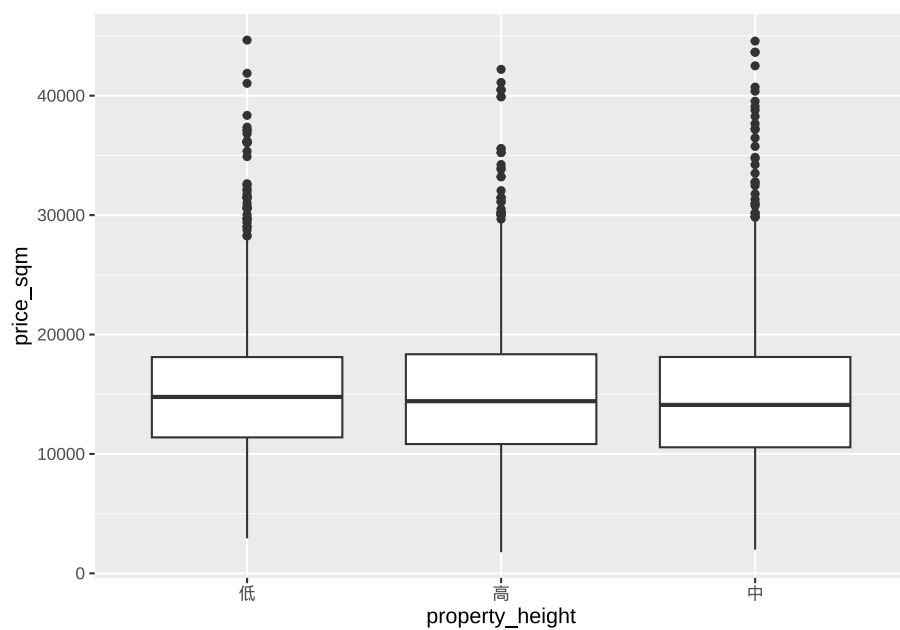


发现：

- 高层住宅的房价高于低层住宅的房价反应了整体趋势，但在各别区间内也有特殊情况
- 在 2 层至 20 层的区间，随着楼层的升高，平均房价呈现下降趋势；在 20 层至 50 层，层呈现单向递增；在 50 层达到峰值后，房价随着超高层的住宅出现开始回落。

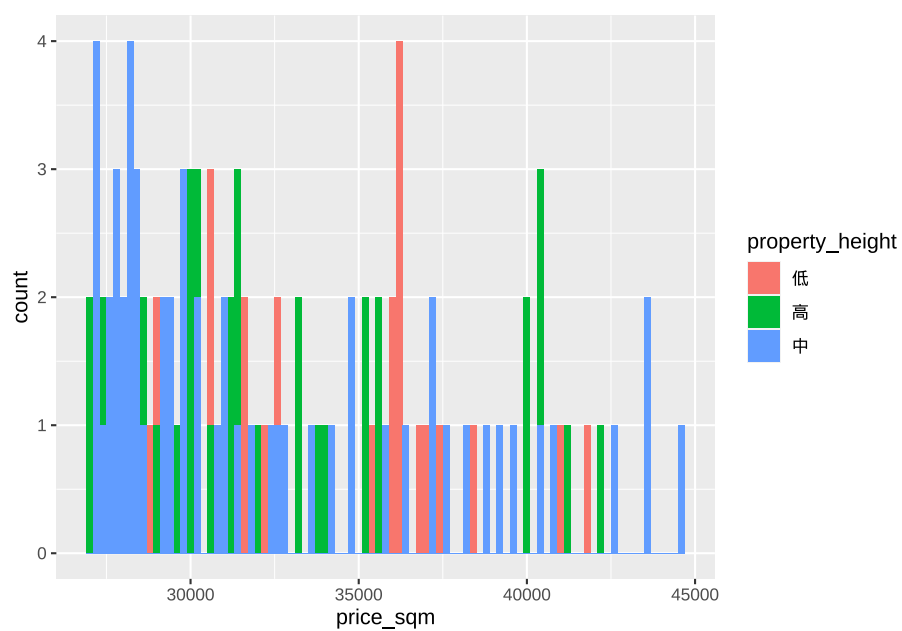
4.6 随着楼层升高，房价是否会显著升高？

4.6.1 楼层与房价的总体箱型图



4.6.2 超高房价区间楼层与房价的数据统计与分布图

```
## # A tibble: 3 x 2
##   property_height count
##   <chr>           <int>
## 1 中               56
## 2 低               43
## 3 高               52
```

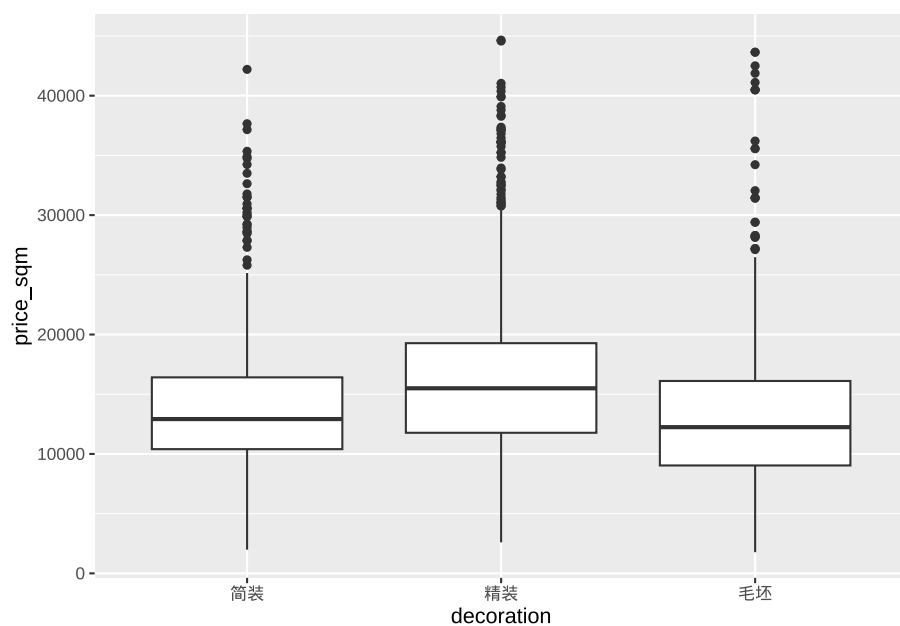


发现:

- 在普通房价区间，房价与所处楼层高低并无显著的关联。
- 在超高房价区间，低层房源相对中层和高层较少，其中中层最多。低层房源主要集中在 35000 元/平，高层分布相对均匀，中层主要集中在区间两端。

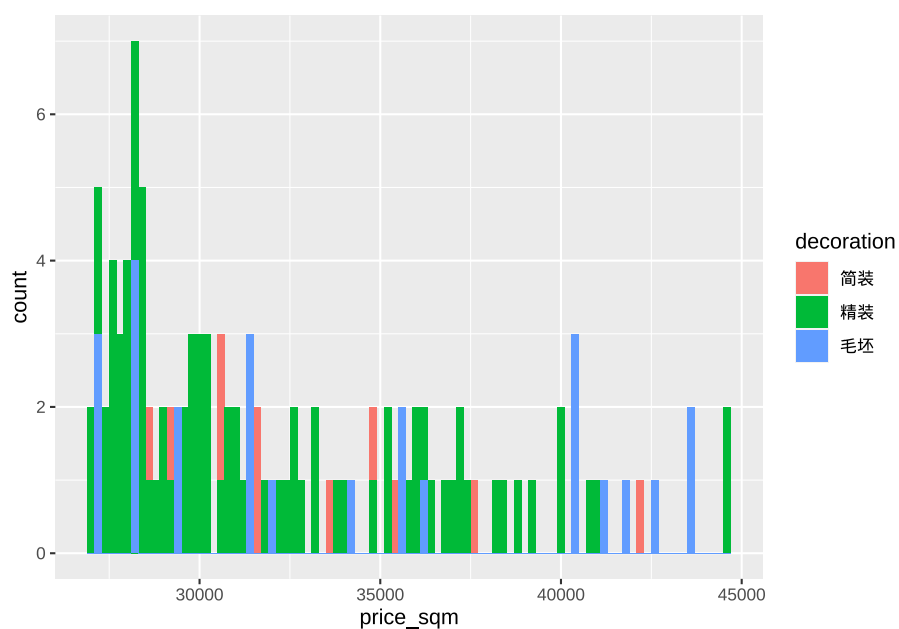
4.7 装修等级是否会对房价产生影响？

4.7.1 装修等级与房价的总体箱型图



4.7.2 超高房价区间，装修等级与房价的数量统计与图形分布

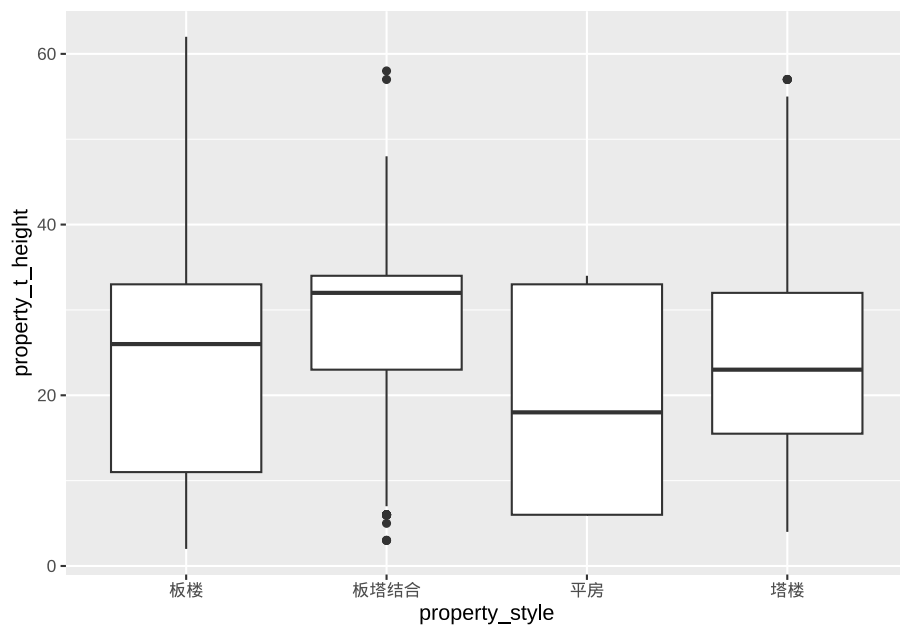
```
## # A tibble: 3 x 2
##   decoration count
##   <chr>      <int>
## 1 毛坯         25
## 2 简装         29
## 3 精装        94
```



发现:

- 在普通房价区间，装修等级对房价有着显著影响，精装的房价高于简装，简装高于毛坯。
- 在超高房价区间，精装二手房数远多于简装与毛坯，精装对房价的附加值提升影响显著。
- 在超高房价区间毛坯房数量少于精装和简装，超高房价区间住宅同时具有投资和消费双重属性。但在极高房价区间，在房价超过 40000 元/平后，毛坯房的数量分布相对增加，该区间二手房源开始呈现奢侈品属性，即无形价值相比有形价值显著提高，具有很强的纯投资行为。

4.8 建筑形式与楼层是否存在相关？



```
## # A tibble: 4 x 2
##   property_style property_t_height_mean
##   <chr>          <dbl>
## 1 塔楼           23.4
## 2 平房           19.4
## 3 板塔结合       29.1
## 4 板楼           23.1
```

发现:

- 建筑形式与总楼层存在一定的相关性，板塔结合形式的建筑平均楼层最高，平房最低，塔楼和板楼相近，但板楼的楼层分布的差异性最大。

5 发现总结

- 从总体数据来看，15000 元/平的房源是武汉二手市场交易的主流，在这个普通价格区间内，中位数和均值相差不大，可供选择的范围很多，可以综合考虑是否临近地铁，楼层建筑形式，楼层，装修等级等因素，但这个价格区间一般只可选择总楼层 20-30 层的楼栋。
- 在中低房价区间，临近地铁、所属区域与装修等级构成了二手房房价的主要影响因素，其他因素对房价没有特别明显的关联。
- 在超高房价区间，板楼作为最受欢迎的建筑形式，存量最多，同时，低层的二手房数量较少，高层和中层较多，从某种程度上来说，高房价消费中普遍青睐中层和高层，此外，精装二手房数远多于简装与毛坯，精装对房价的附加值提升影响显著，超高房价区间住宅同时具有投资和消费双重属性。
- 在极高房价区间，特别针对超过 40000 元/平的二手房源，板塔结合、中层和毛坯的存量较多，该区间二手房源开始呈现奢侈品属性，即无形价值相比有形价值显著提高，具有很强的纯投资行为。