

第一次作业：武汉链家二手房简要分析

杨玉瑾

2023-10-19

你的主要发现

1. 发现 1 根据武汉市链家二手房 3000 套房源信息，可以初步发现，武汉市挂牌二手房房源单价均价主要集中在 14000-15000 元居多，总价在 140 万-150 万元居多，面积在 100 平居多，户型多为两室一厅或三室两厅，精装房源为主，板楼多，大多数近地铁。
2. 发现 2 房源地段、户型、楼层、交通便利程度均在不同程度上影响房源价格。
3. 发现 3 中北路、水果湖的房源价格最高，存在一些面积大、房间多的房源和面积小于 20 平的房源，差异化大。

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。

##	property_name	property_region	price_ttl	price_sqm
##	"character"	"character"	"numeric"	"numeric"
##	bedrooms	livingrooms	building_area	directions1
##	"numeric"	"numeric"	"numeric"	"character"
##	directions2	decoration	property_t_height	property_height
##	"character"	"character"	"numeric"	"character"
##	property_style	followers	near_subway	if_2y
##	"character"	"numeric"	"character"	"character"
##	has_key	vr		
##	"character"	"character"		
##	property_name	property_region	price_ttl	price_sqm
##	"factor"	"factor"	"numeric"	"numeric"
##	bedrooms	livingrooms	building_area	directions1
##	"numeric"	"numeric"	"numeric"	"factor"
##	directions2	decoration	property_t_height	property_height
##	"factor"	"factor"	"numeric"	"factor"
##	property_style	followers	near_subway	if_2y
##	"factor"	"numeric"	"factor"	"factor"
##	has_key	vr		
##	"factor"	"factor"		

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", "~
## $ 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:

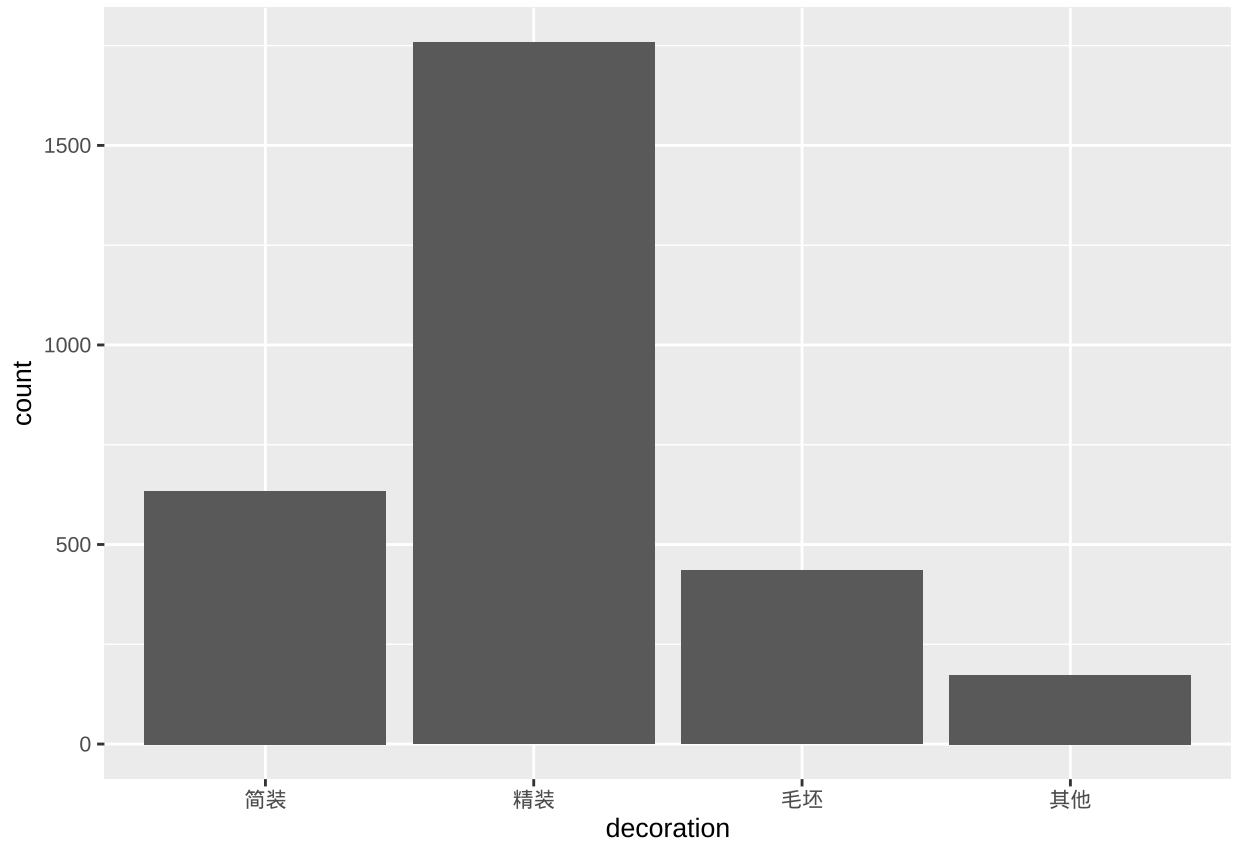
```

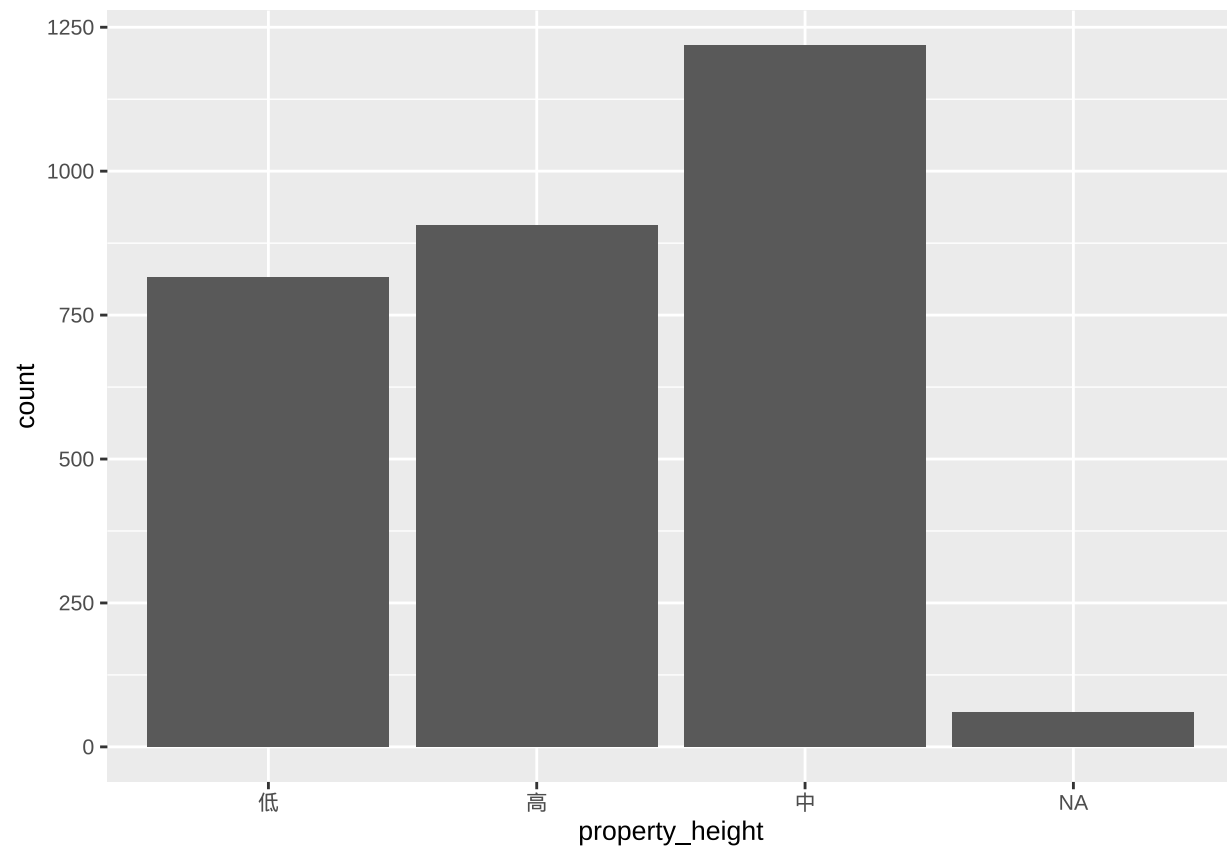
##      property_name  property_region  price_ttl      price_sqm
## 东立国际      : 22 白沙洲 : 167    Min.      : 10.6    Min.      : 1771
## 保利中央公馆 : 16 盘龙城 : 126    1st Qu.: 95.0    1st Qu.:10799
## 朗诗里程      : 16 四新   : 116    Median   : 137.0  Median :14404
## 恒大名都      : 15 光谷东 : 112    Mean     : 155.9  Mean   :15148
## 阳光100大湖第: 15 金银湖 : 97     3rd Qu.: 188.0  3rd Qu.:18211
## 保利城一期    : 13 后湖   : 86     Max.     :1380.0  Max.   :44656
## (Other)       :2903 (Other):2296
##      bedrooms      livingrooms  building_area  directions1  directions2
## Min.      :1.000    Min.      :0.000  Min.      : 22.77  南          :2454  北          :1189
## 1st Qu.:2.000    1st Qu.:1.000    1st Qu.: 84.92  东南        : 281  南          : 66
## Median :3.000    Median :2.000    Median : 95.55  东          : 98  西          : 25
## Mean   :2.695    Mean   :1.709    Mean  :100.87  北          : 68  东南        : 15
## 3rd Qu.:3.000    3rd Qu.:2.000    3rd Qu.:117.68  西南        : 57  西南        : 12
## Max.   :7.000    Max.   :4.000    Max.   :588.66  西          : 19  (Other): 21
##                                     (Other): 23  NA's       :1672
##      decoration  property_t_height  property_height  property_style
## 简装: 634    Min.      : 2.00    低      : 816    板楼      :1781
## 精装:1757    1st Qu.:11.00    高      : 906    板塔结合: 615
## 毛坯: 436    Median :27.00    中      :1218    平房      : 5
## 其他: 173    Mean   :24.22    NA's: 60      塔楼      : 527
##                                     暂无数据: 72
##                                     Max.     :62.00
##
##      followers      near_subway      if_2y      has_key
## Min.      : 0.000  VR看装修 : 2    房本满两年:1264  随时看房:2525
## 1st Qu.: 1.000    近地看   : 1    NA's       :1736  近地铁   : 7
## Median : 3.000    近地铁   :1554  VR看装修: 4
## Mean   : 6.614    珞狮南   : 1    藏龙岛   : 1
## 3rd Qu.: 6.000    太子湖1号: 1    关山大道: 1
## Max.   :262.000  NA's      :1441  (Other)   : 4
##                                     NA's      : 458

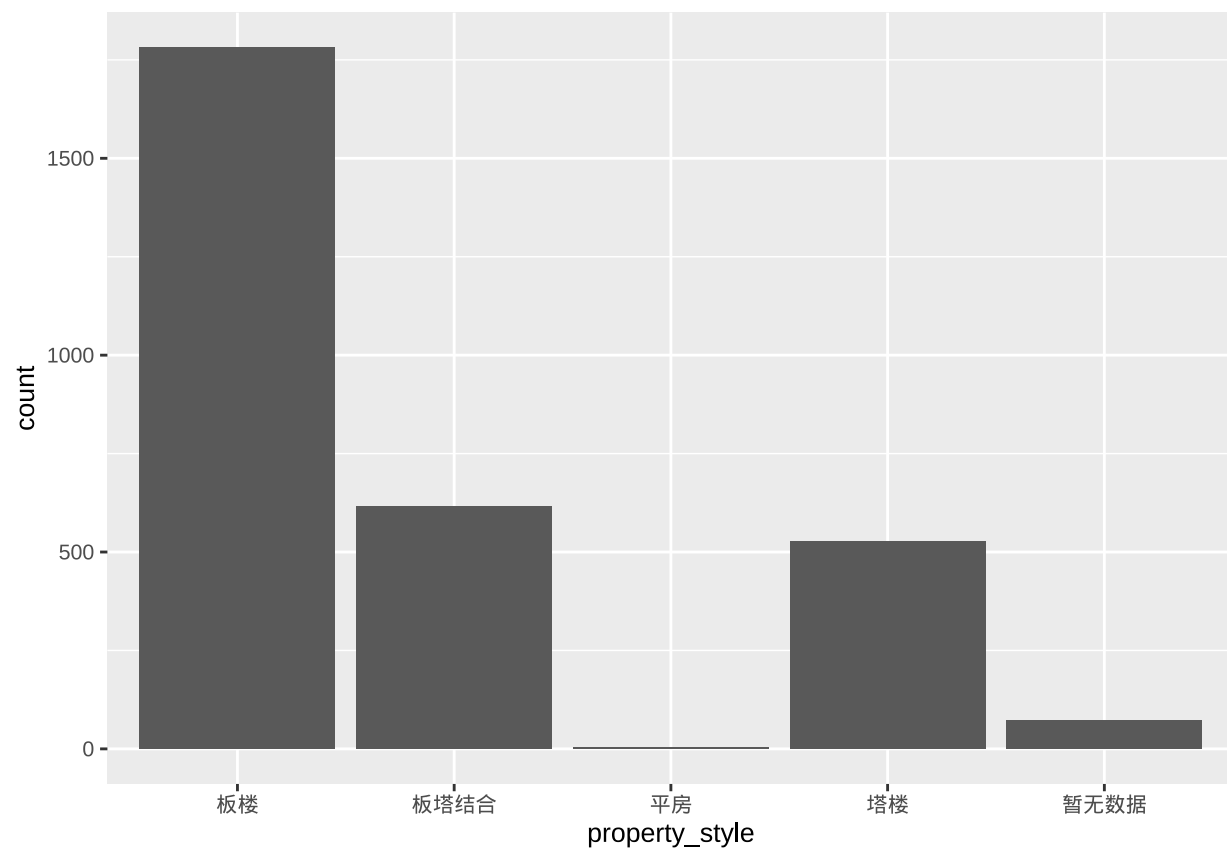
```

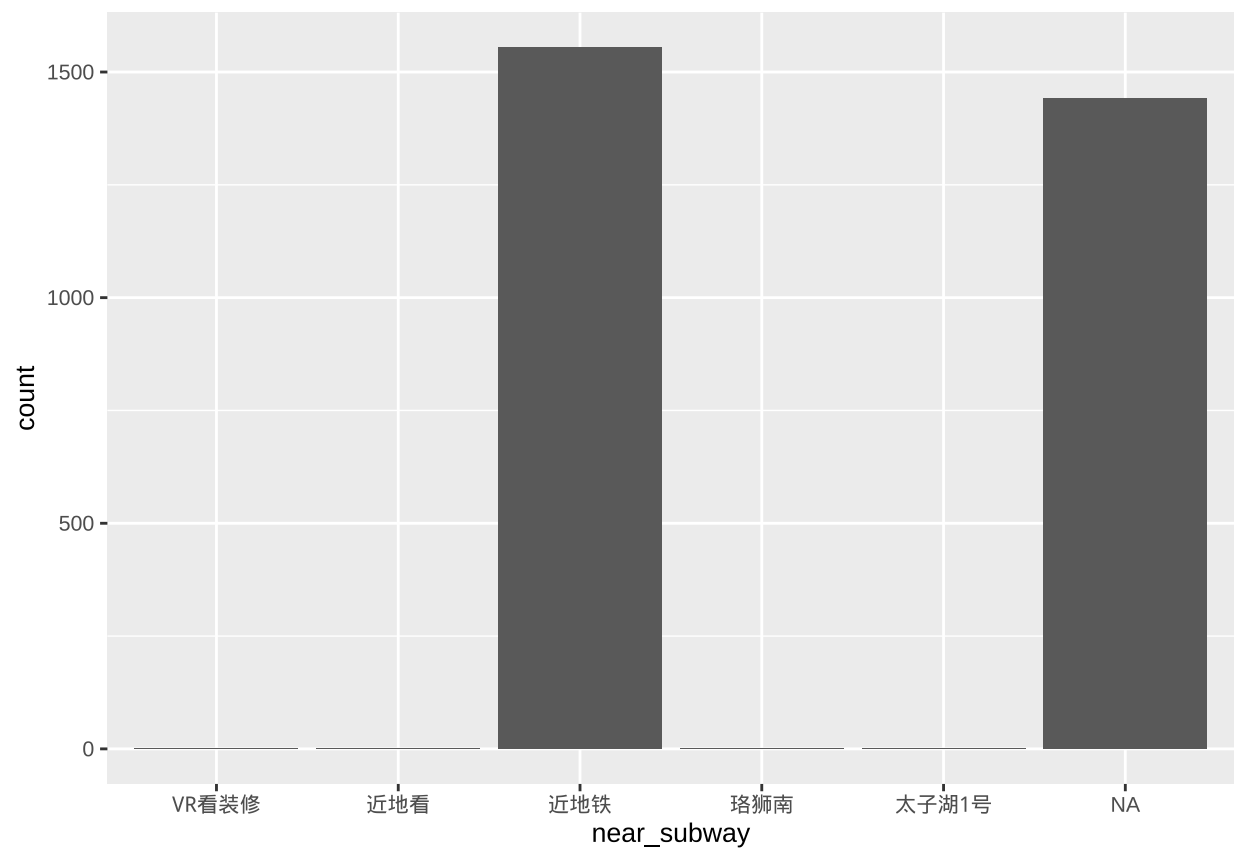
```
##          vr
## VR看装修:2084
## VR 甸城  :  1
## 保利拉   :  1
## 泓悦府   :  1
## 江景湾   :  1
## (Other)  :  6
## NA's     : 906

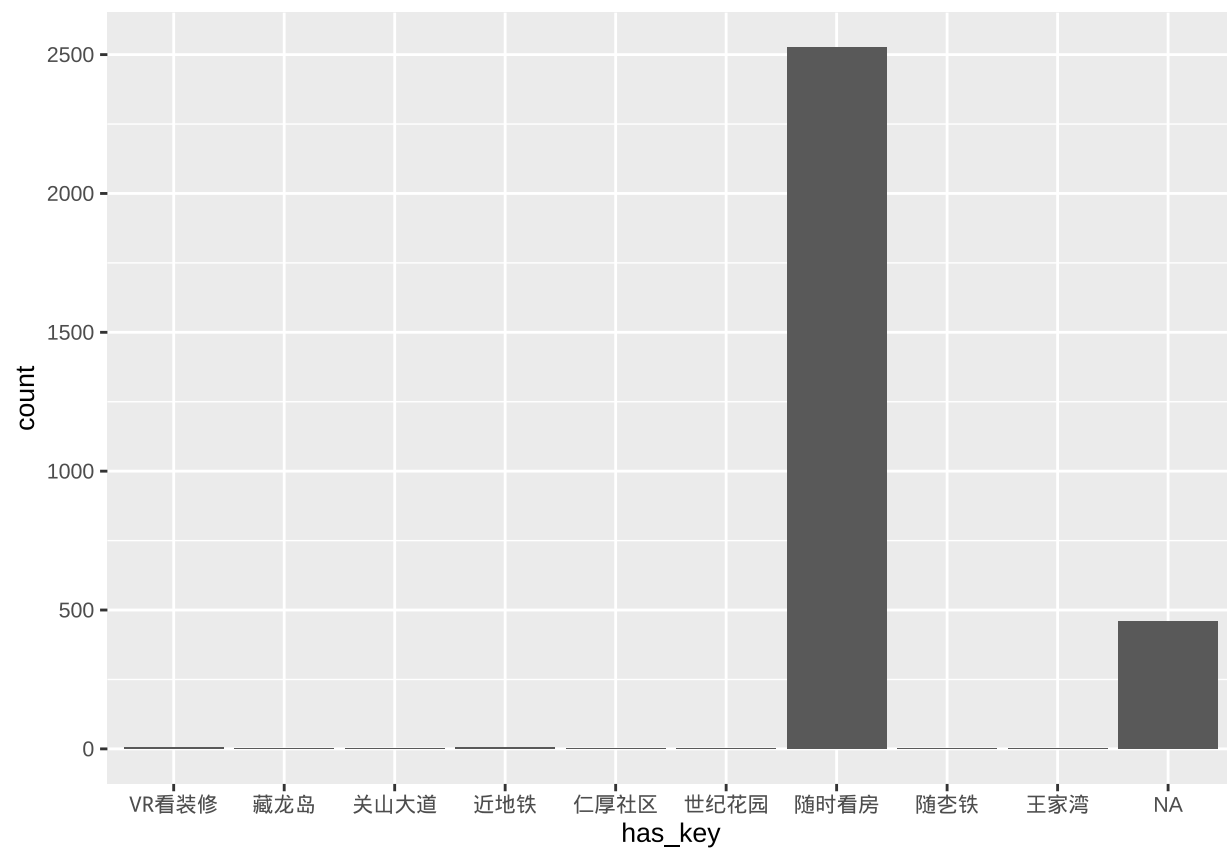
## [1] 9129.445
## [1] 39982547
```

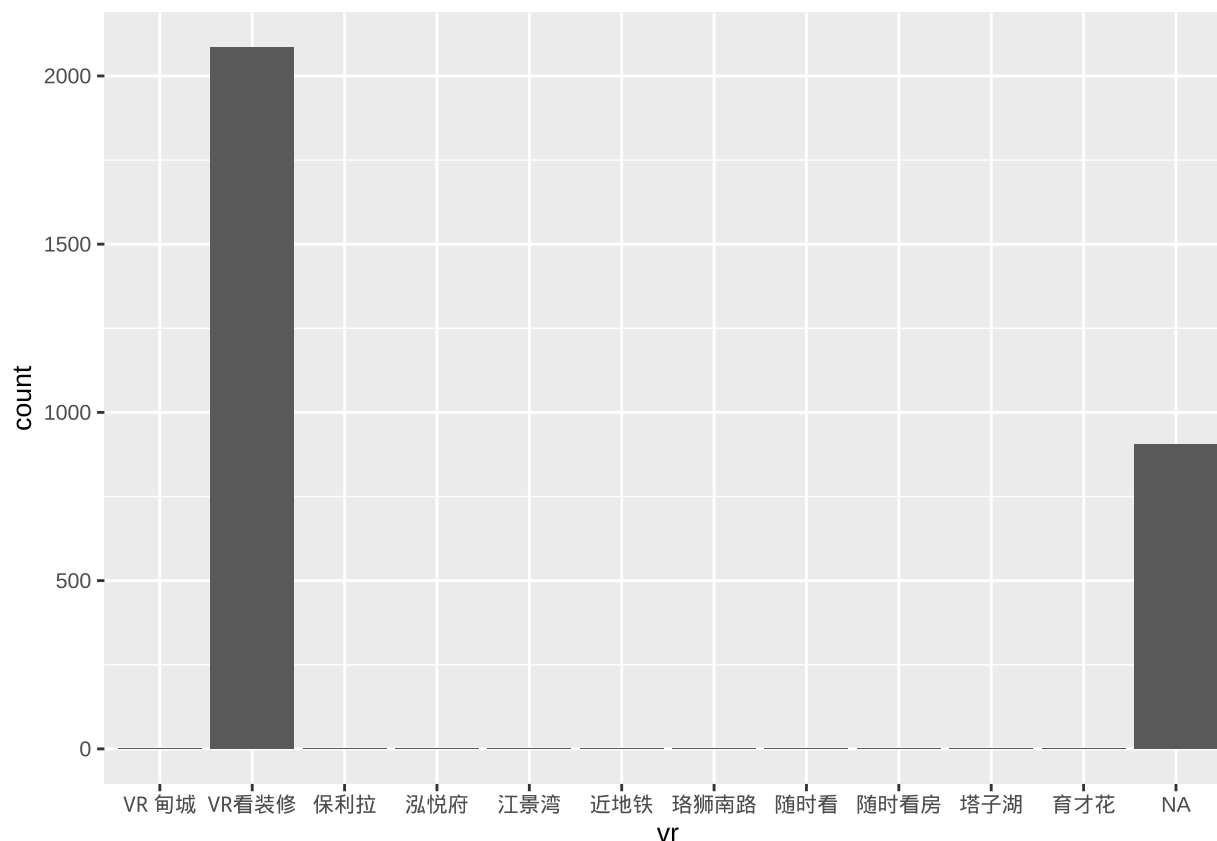












可以看到：

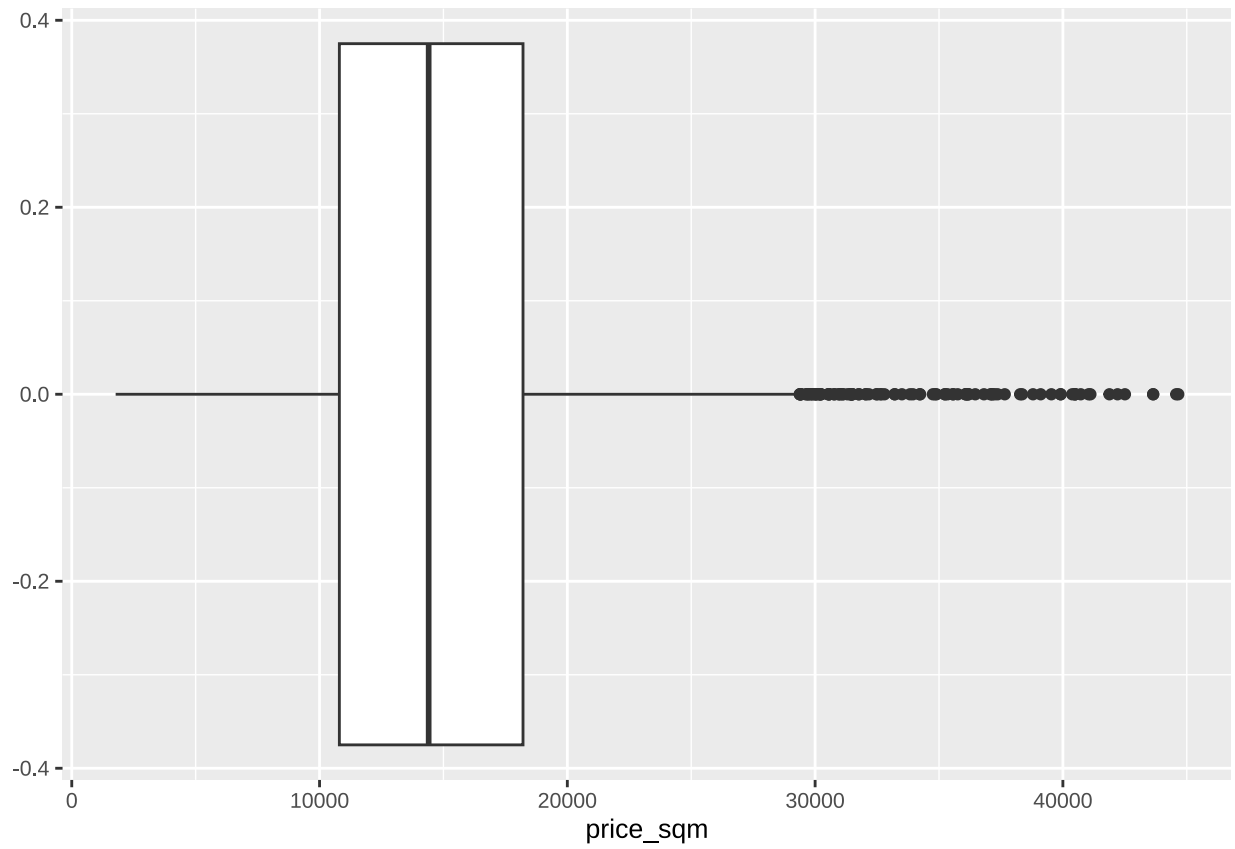
- 直观结论 1 武汉市链家挂牌二手房房源每平方米价格平均数为 15148 元，中位数为 14404 元，最高值为 44656 元，最低值为 1771 元，极差为 42885 元，方差为 39982547，数据波动大，离散程度较低，可能存在一些异常值。
- 直观结论 2 武汉市链家挂牌二手房房源总价平均数为 155.9 万，中位数为 137 万，最高值为 1380 万，最低值为 10.6 万，极差为 1369.4 万，离散程度较低，可能存在一些异常值。
- 直观结论 3 武汉市链家挂牌二手房房源面积平均数为 100.87 平，中位数为 95.55 平，最高值为 588.66 平，最低值为 22.77 平，极差为 565.89 平，可能存在一些异常值。
- 直观结论 4 武汉市链家挂牌二手房房源至少有 1 间卧室，大部分有 2 至 3 间卧室，最多有 7 间卧室；大部分有 1 至 2 间客厅，最多有 4 间客厅，最少的没有客厅。
- 直观结论 5 武汉市链家挂牌二手房房源以精装为主，标注精装的占 58.57%（1757 套），标注简装的占 21.13%（634 套）。房源楼层分布较为均匀，中楼层房源相对较多，40.6% 的房源（1218 套）位于中楼层，30.2% 的房源（906 套）位于高楼层，27.2% 的房源（816 套）位于低楼层。房源为板楼的居多，占 59.37%（1781 套），其次是板塔结合和塔楼，分别占 20.5%（615 套）、17.57%（527 套）。约一半房源（51.8%，1554 套）标注”近地铁”，交通便利；大部分房源（84.17%，2525 套）可以”随时看房”，房屋中介有钥匙更方便带看；69.47% 的房源（2084 套）可以”VR 看装修”，利于有意向的买家足不出户初步了解房屋状况。

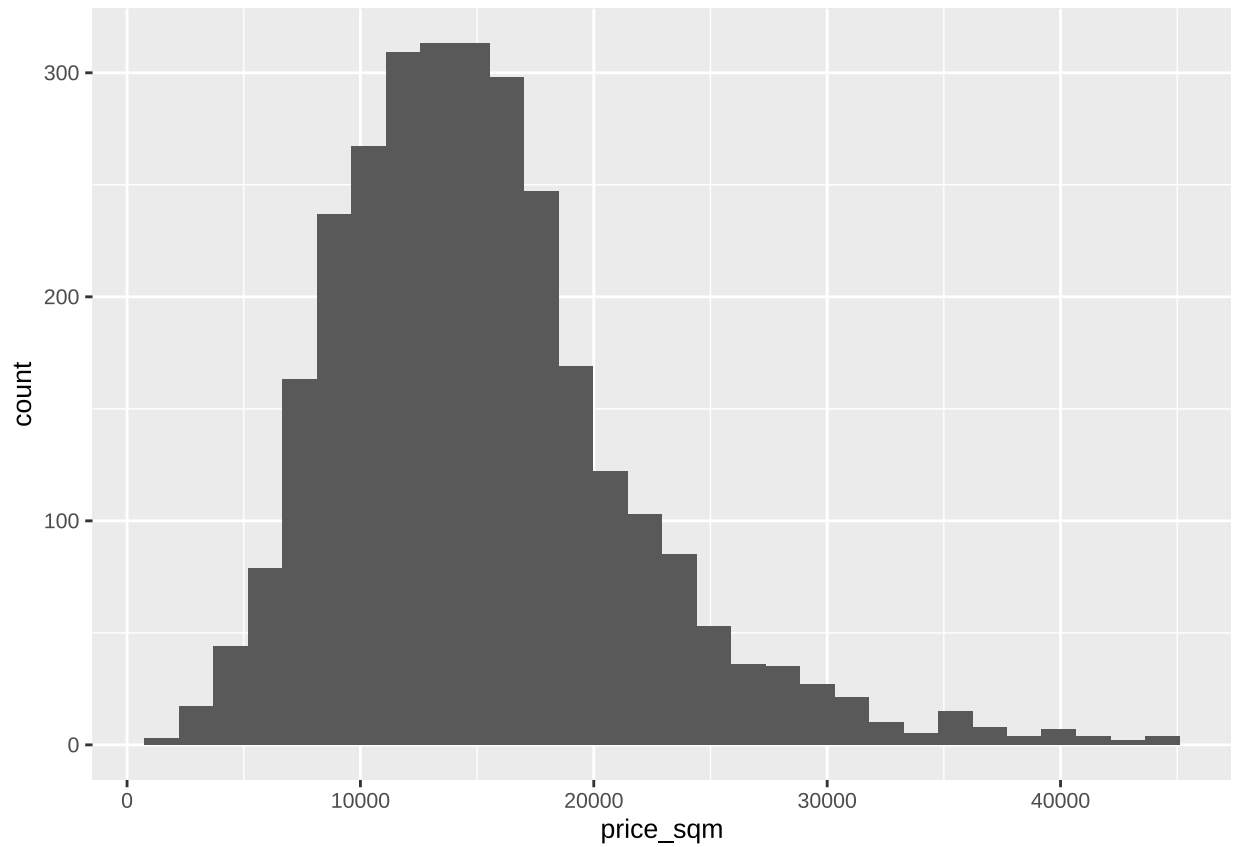
探索性分析

每平方米价格的数值描述与图形

```
## [1] 1.080004
```

```
## [1] 5.028977
```



```
## [1] 27050.35
```

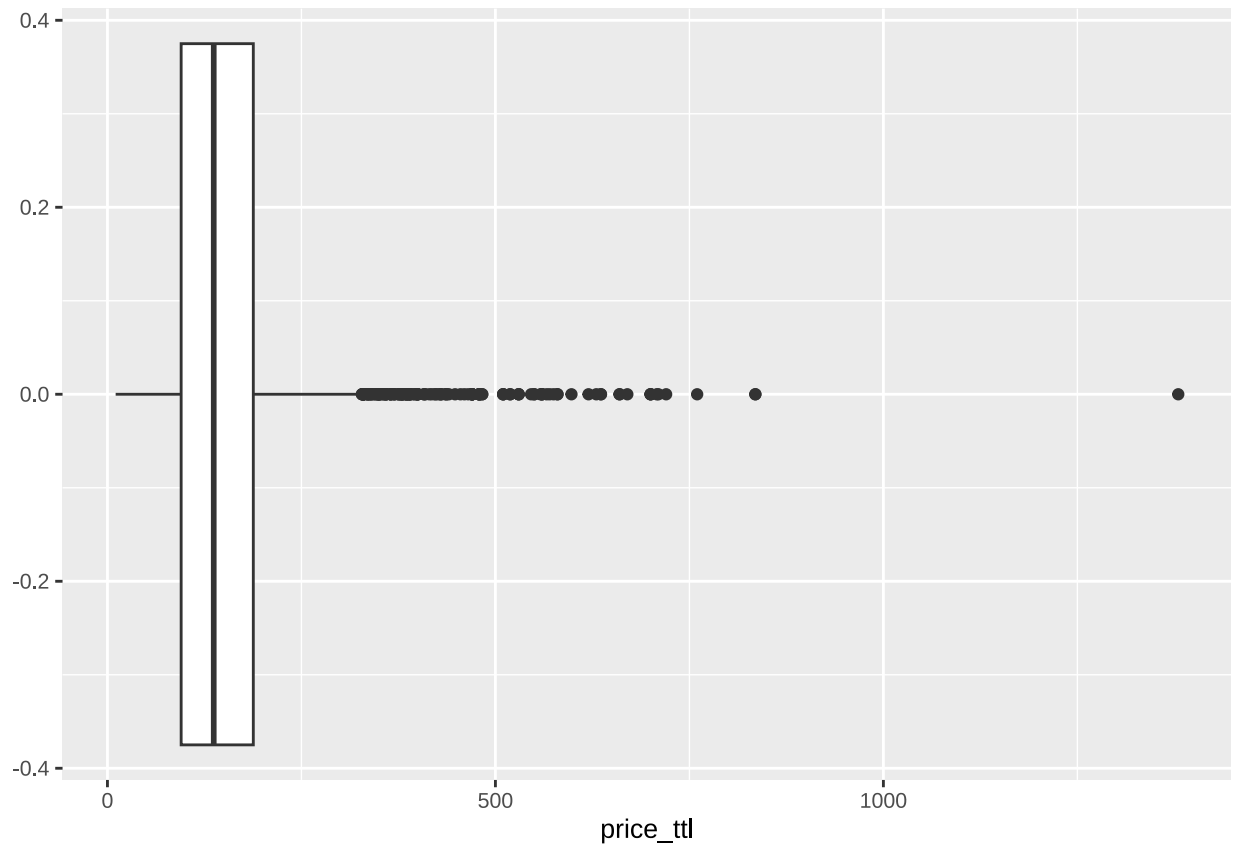
发现：

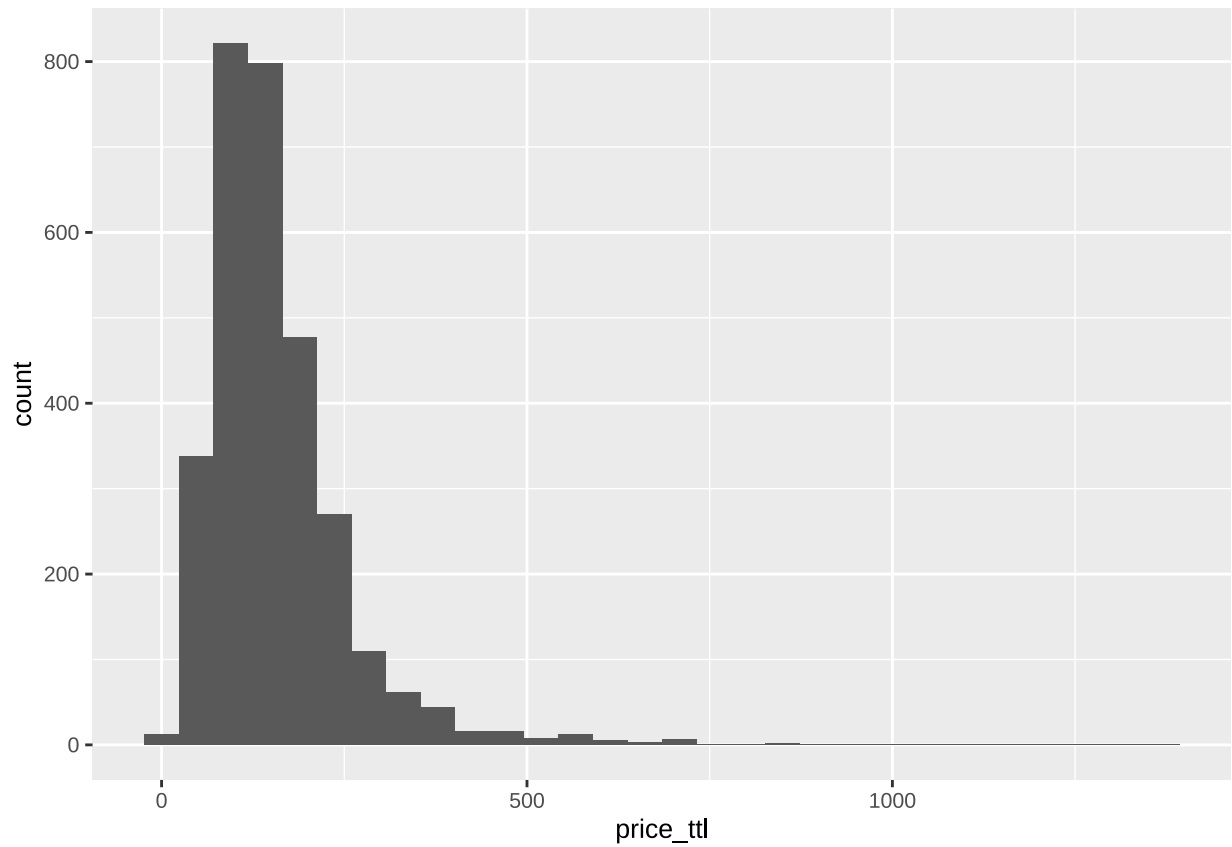
- 发现 1 每平方米价格主要集中在 1 万元至 2 万元之间，假设高于中位数 2 个标准差为异常值，则存在 27050 元以上的异常值。
- 发现 2 每平方米价格分布的尾部在右侧，具有比正态分布更尖锐的峰，不满足正态分布。

房源总价的数值描述与图形

```
## [1] 2.7546
```

```
## [1] 19.12947
```





```
## [1] 328.0963
```

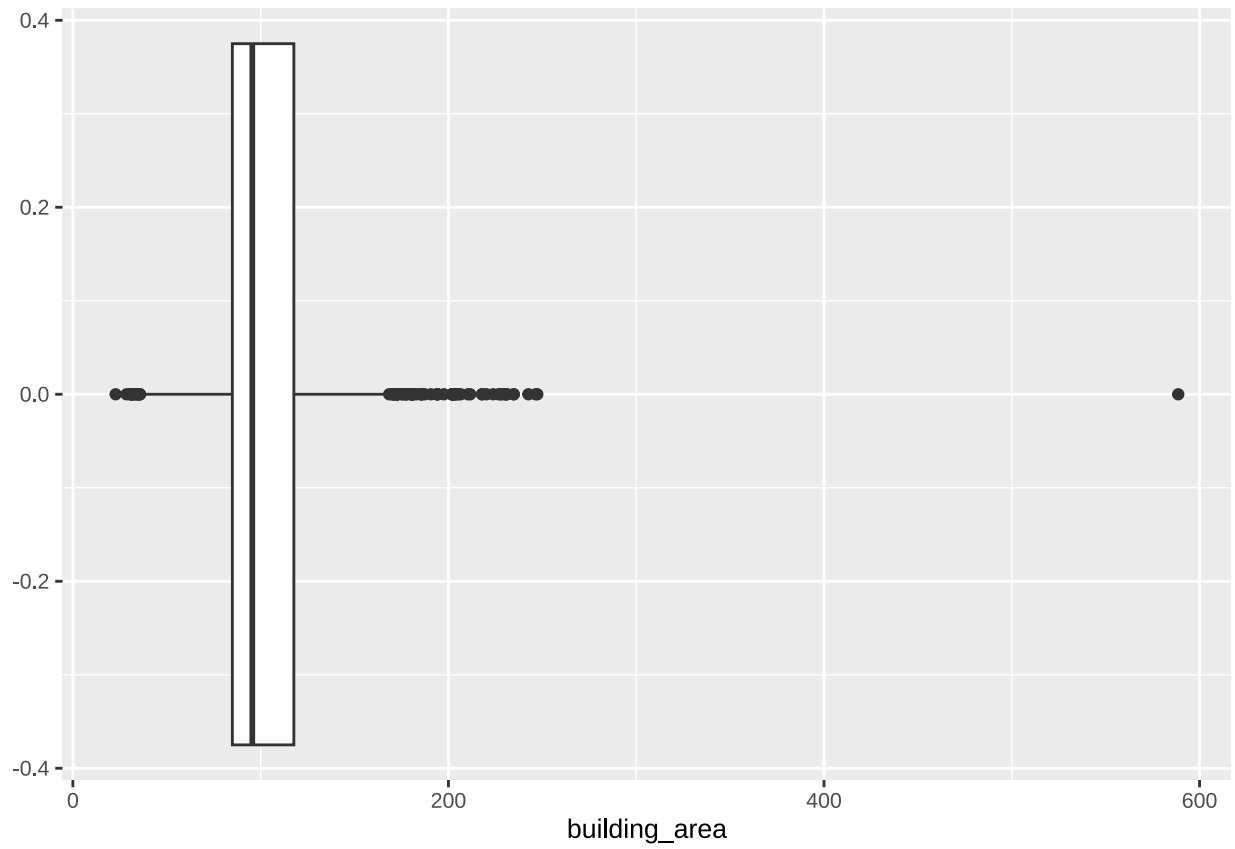
发现：

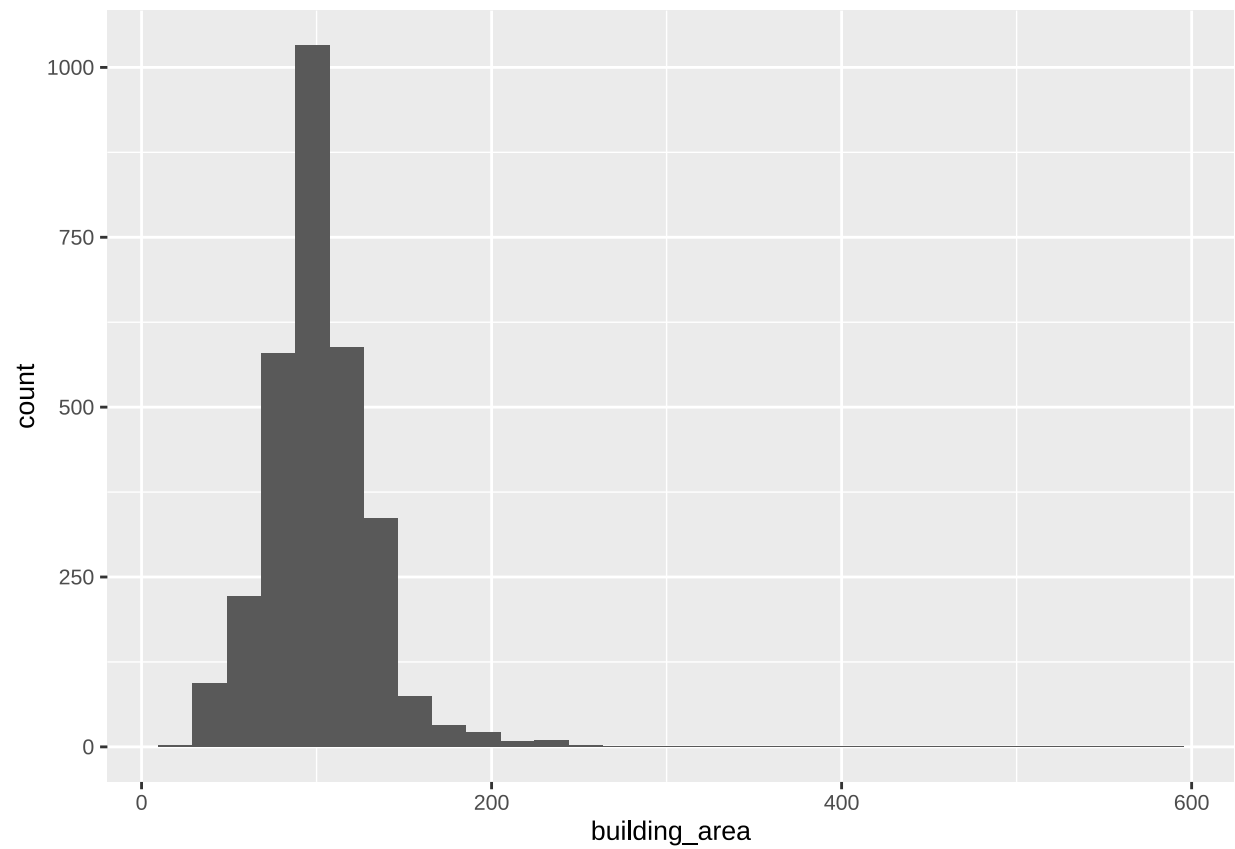
- 发现 1 房源总价主要集中在 30 万元至 250 万元之间，假设高于中位数 2 个标准差为异常值，则存在 328 万元以上的异常值。
- 发现 2 房源总价分布的尾部在右侧，具有比正态分布更尖锐的峰，不满足正态分布。

房源面积的数值描述与图形

```
## [1] 2.080825
```

```
## [1] 26.65362
```



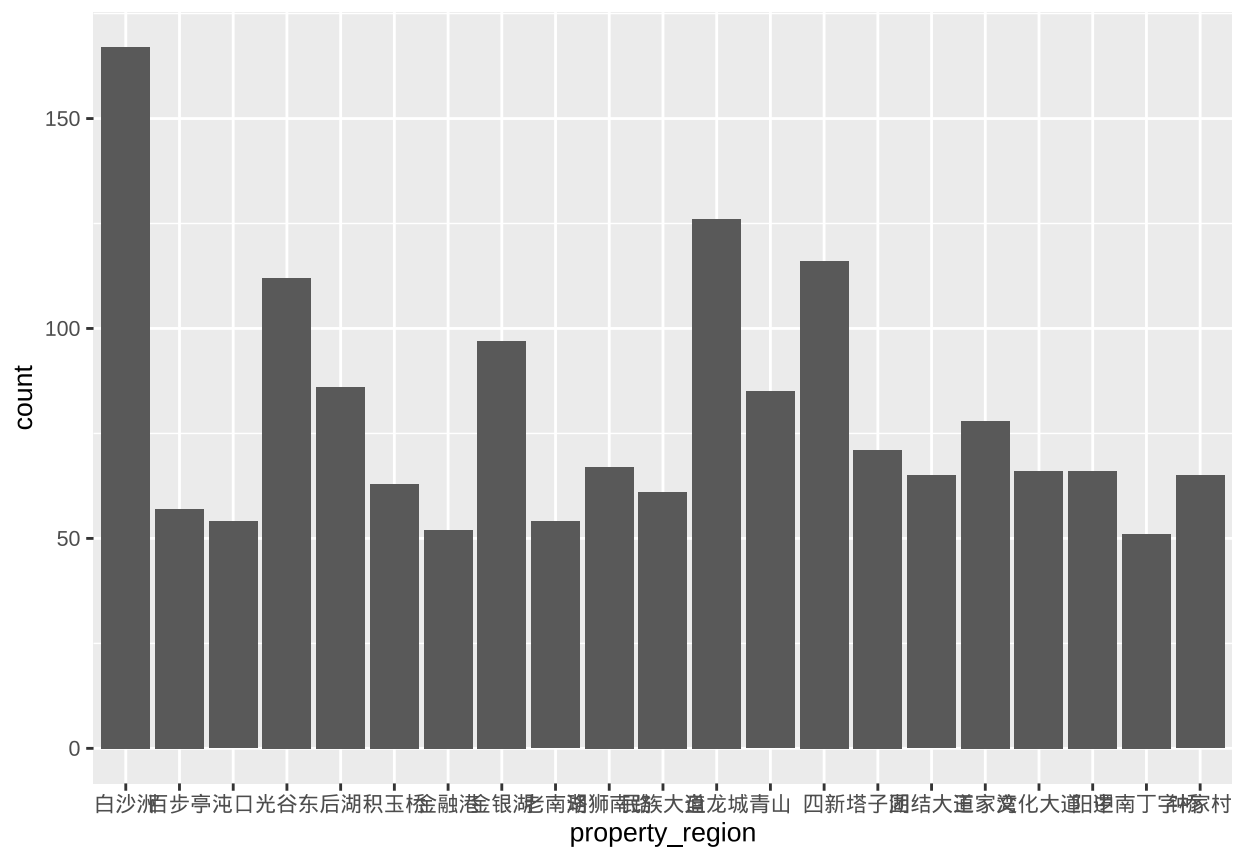


[1] 156.305

发现:

- 发现 1 房源面积主要集中在 60 平至 150 平之间，假设高于中位数 2 个标准差为异常值，则存在 156 平以上的异常值。
- 发现 2 房源面积分布的尾部在右侧，具有比正态分布更尖锐的峰，不满足正态分布。

各区域房源数量的数值描述与图形

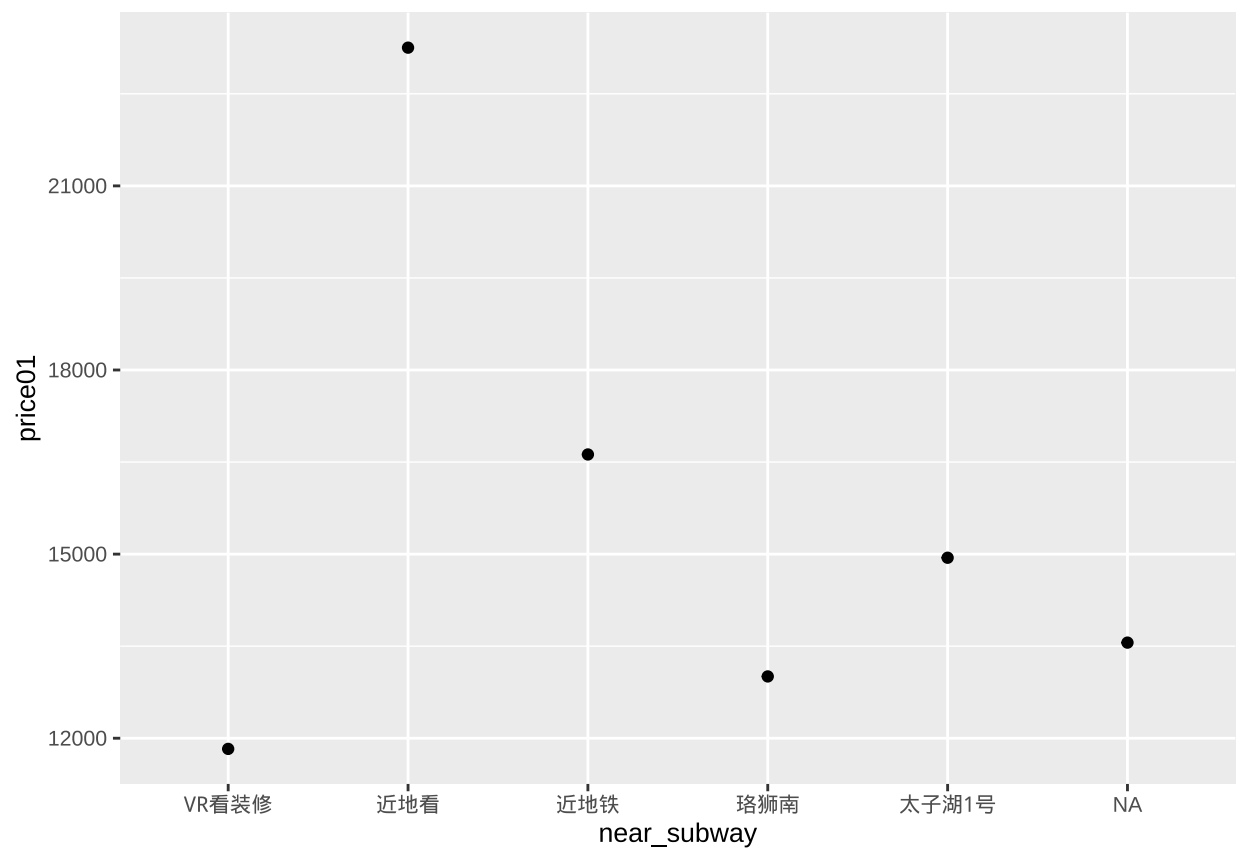


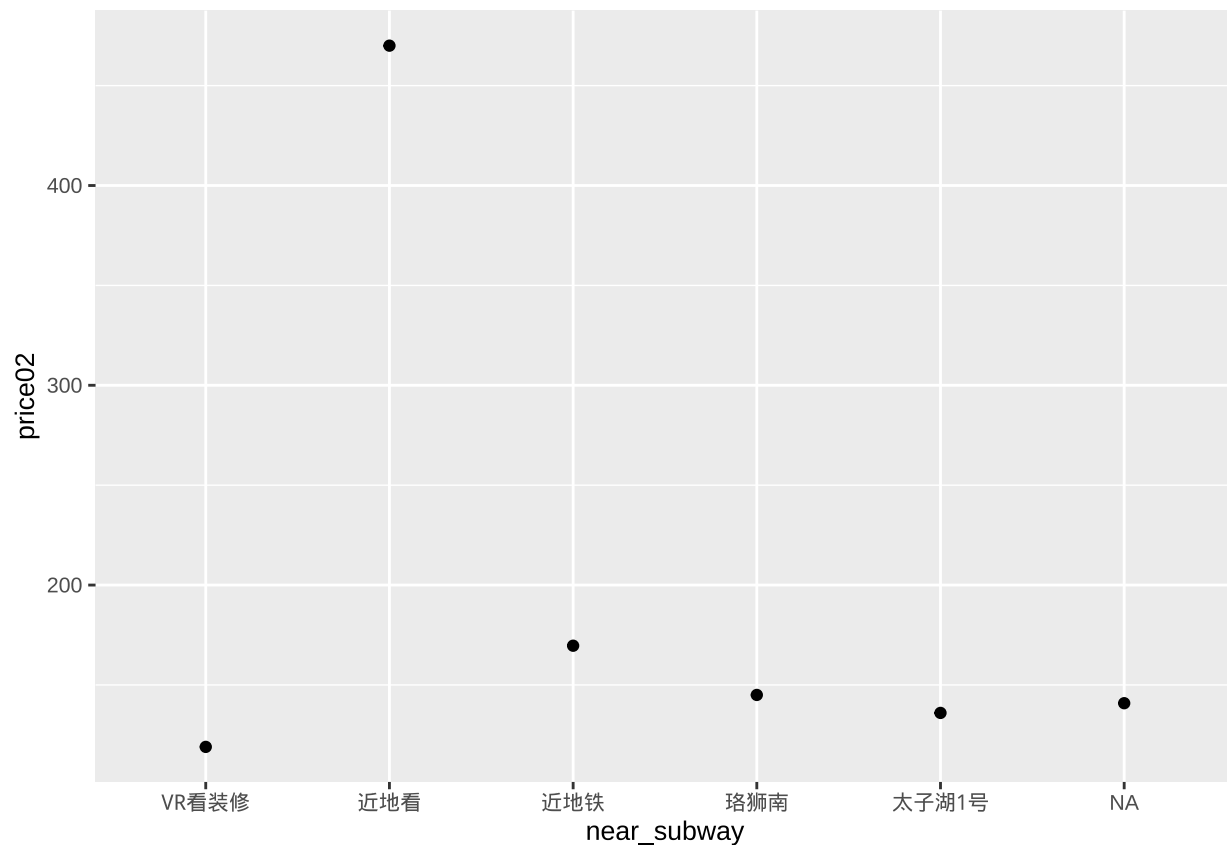
发现:

- 发现 白沙洲区域的房源最多，其次是盘龙城、四新、光谷东。

探索问题 1

```
## # A tibble: 21 x 2
##   property_region price00
##   <fct>           <dbl>
## 1 积玉桥           21403.
## 2 中南丁字桥       21373.
## 3 民族大道         19622.
## 4 老南湖           19461.
## 5 光谷东           16683.
## 6 金融港           16408.
## 7 百步亭           16370.
## 8 塔子湖           16186.
## 9 珞狮南路         16164.
## 10 团结大道        15774.
## # i 11 more rows
## [1] 23681.59
```

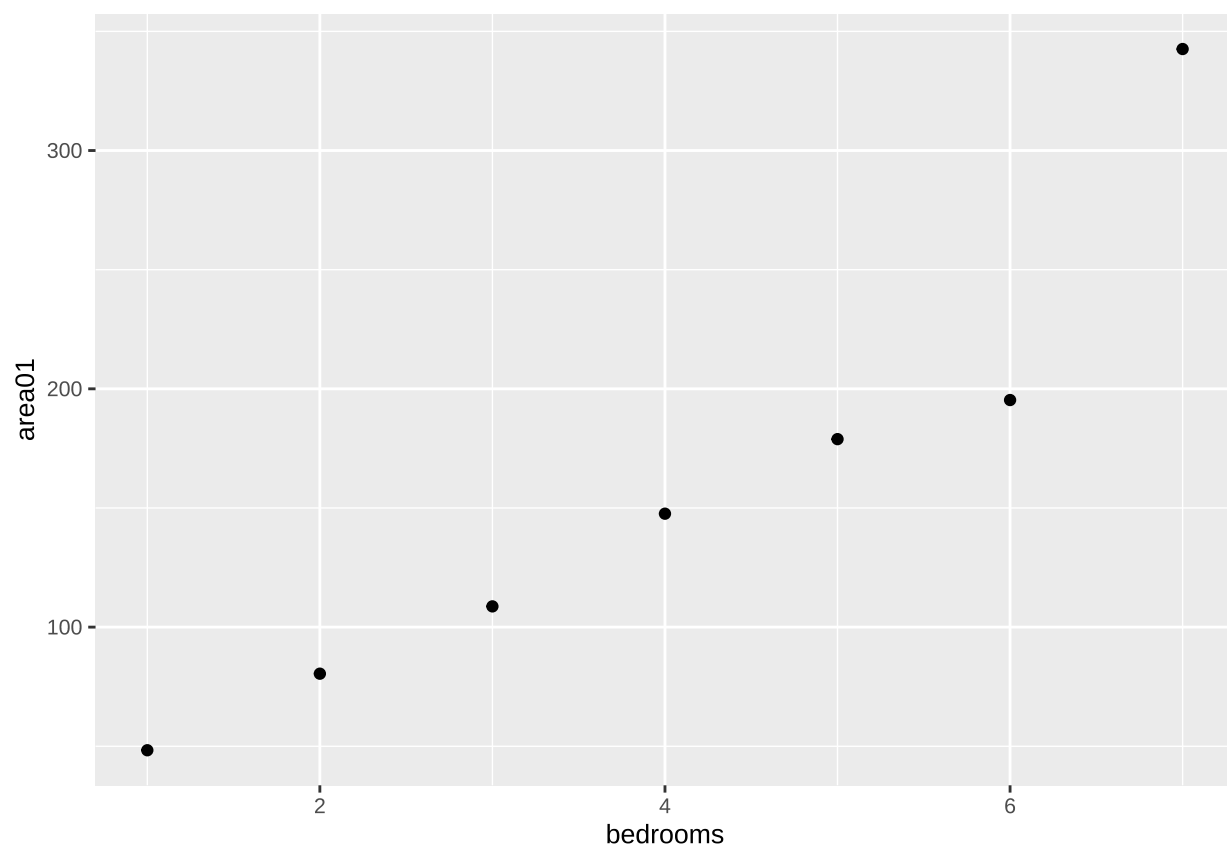


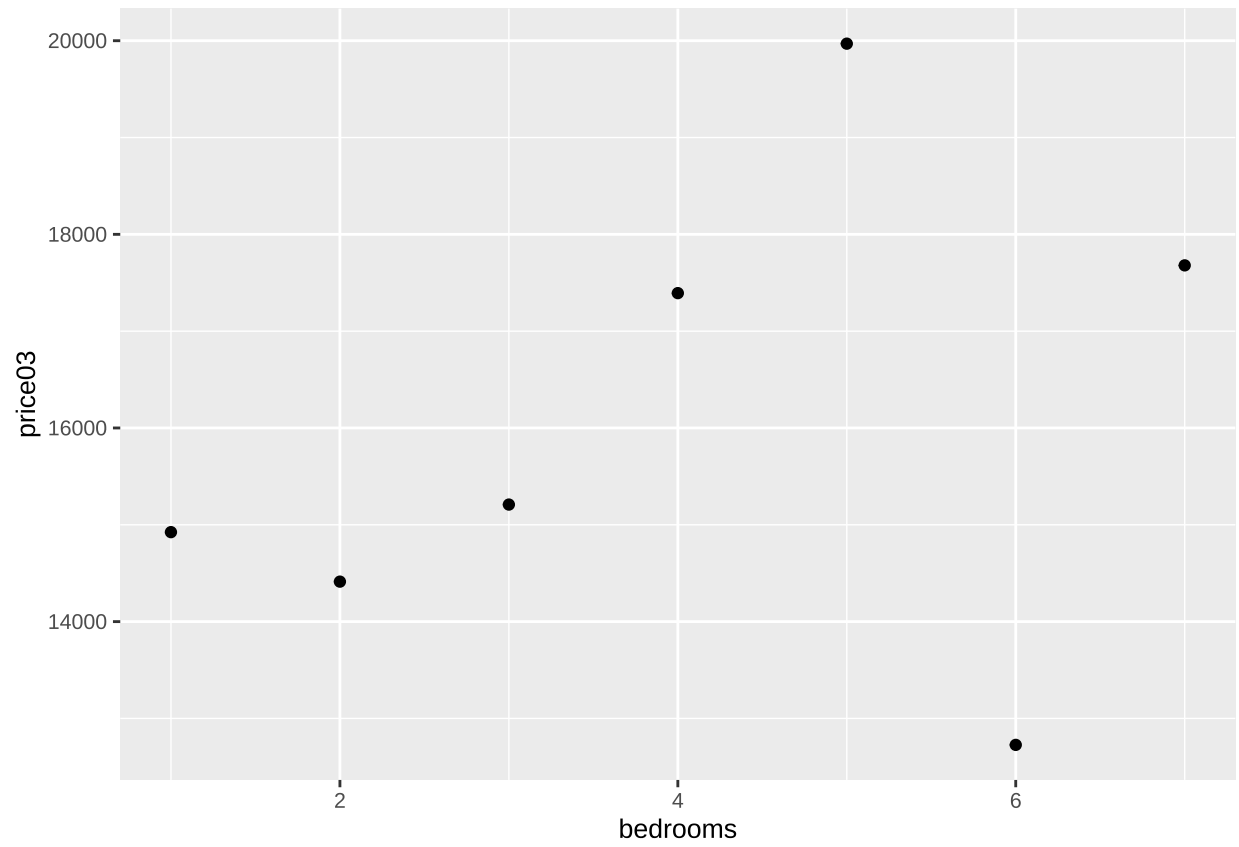


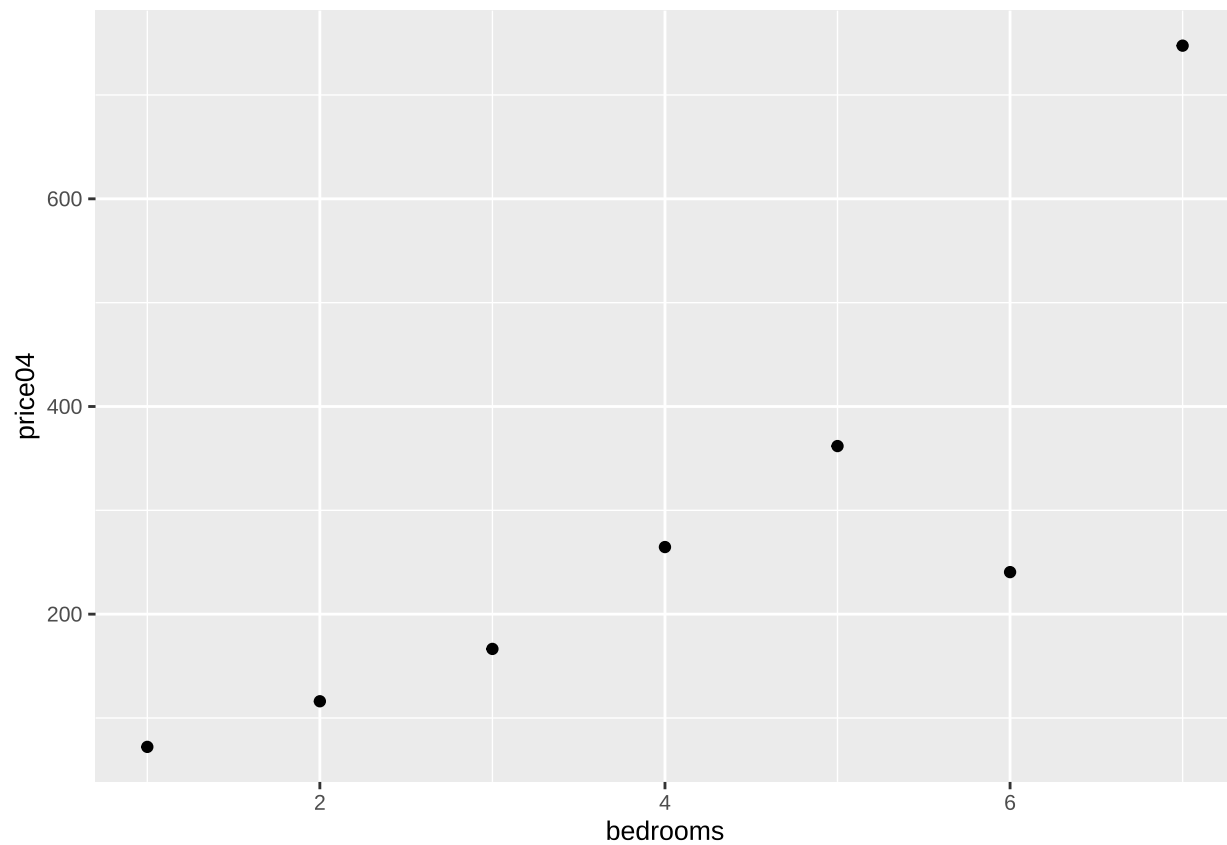
发现：

- 发现 1 剔除掉错误数据（如“VR 看装修”）后，中北路、水果湖、黄埔永清、三阳、南湖沃尔玛、虎泉杨家湾、CBD 西北湖、楚河汉街、关山大道区域的房源总价最高，基本集中在武昌区、江汉区、洪山区等中心城区，与武汉市热门商圈地段相符。假设高于中位数 2 个标准差为异常值，则中北路、水果湖单价为异常值，房源价格偏高。
- 发现 2 剔除掉错误数据（如“VR 看装修”“近地看”“珞狮南”“太子湖 1 号”）后，标注“近地铁”的房源单价为 16624 元，总价为 169.64 万元，均高于未标注“近地铁”的房源单价 13558 元，以及总价 140.86 万元，说明交通便利程度对房价有影响，近地铁的房源可能有更高的价格。

探索问题 2



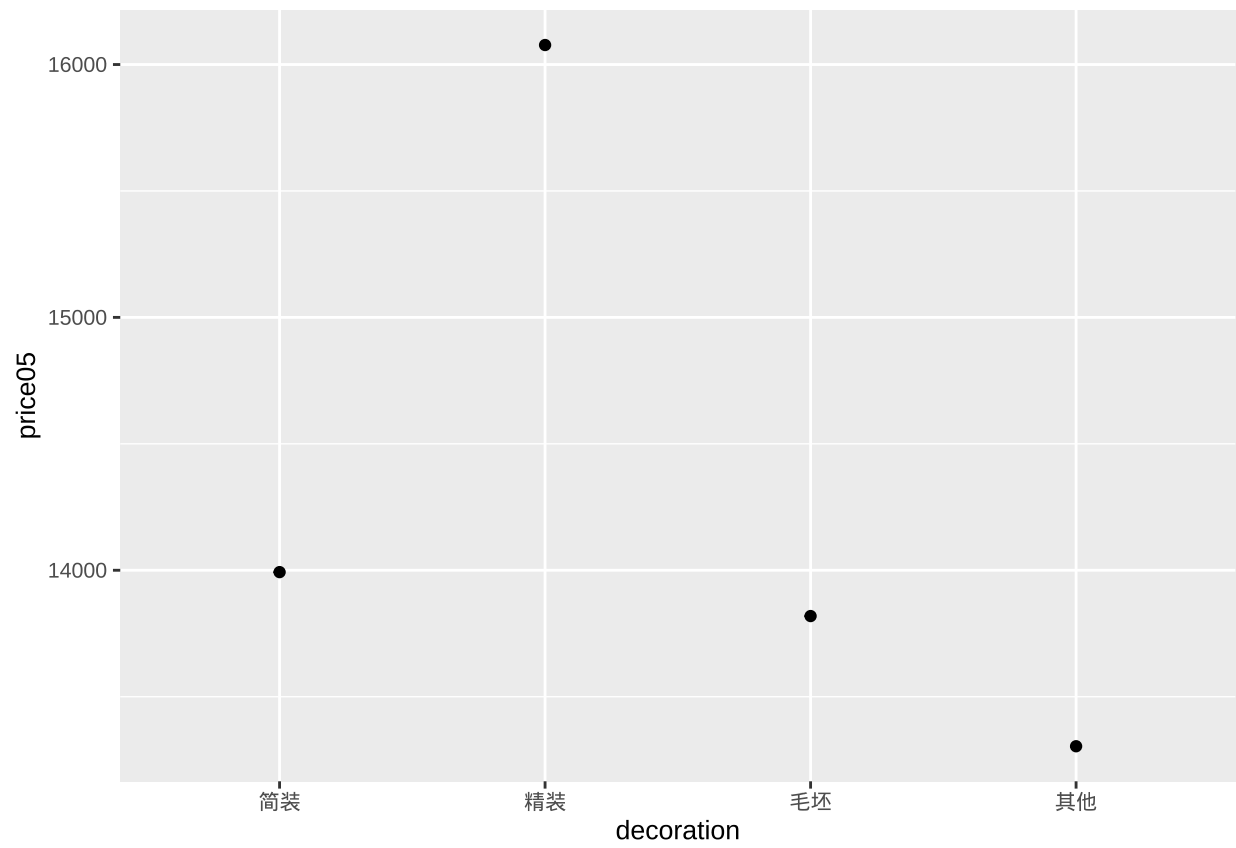


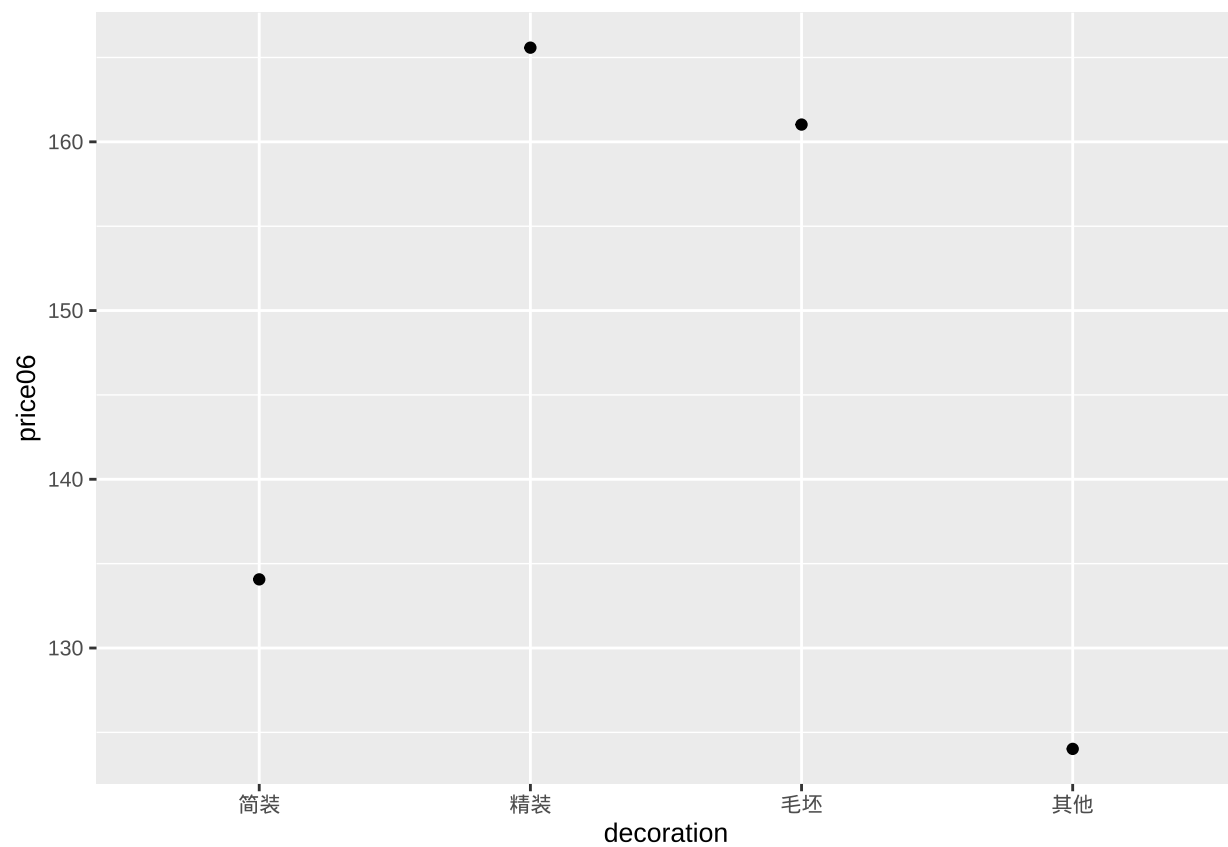


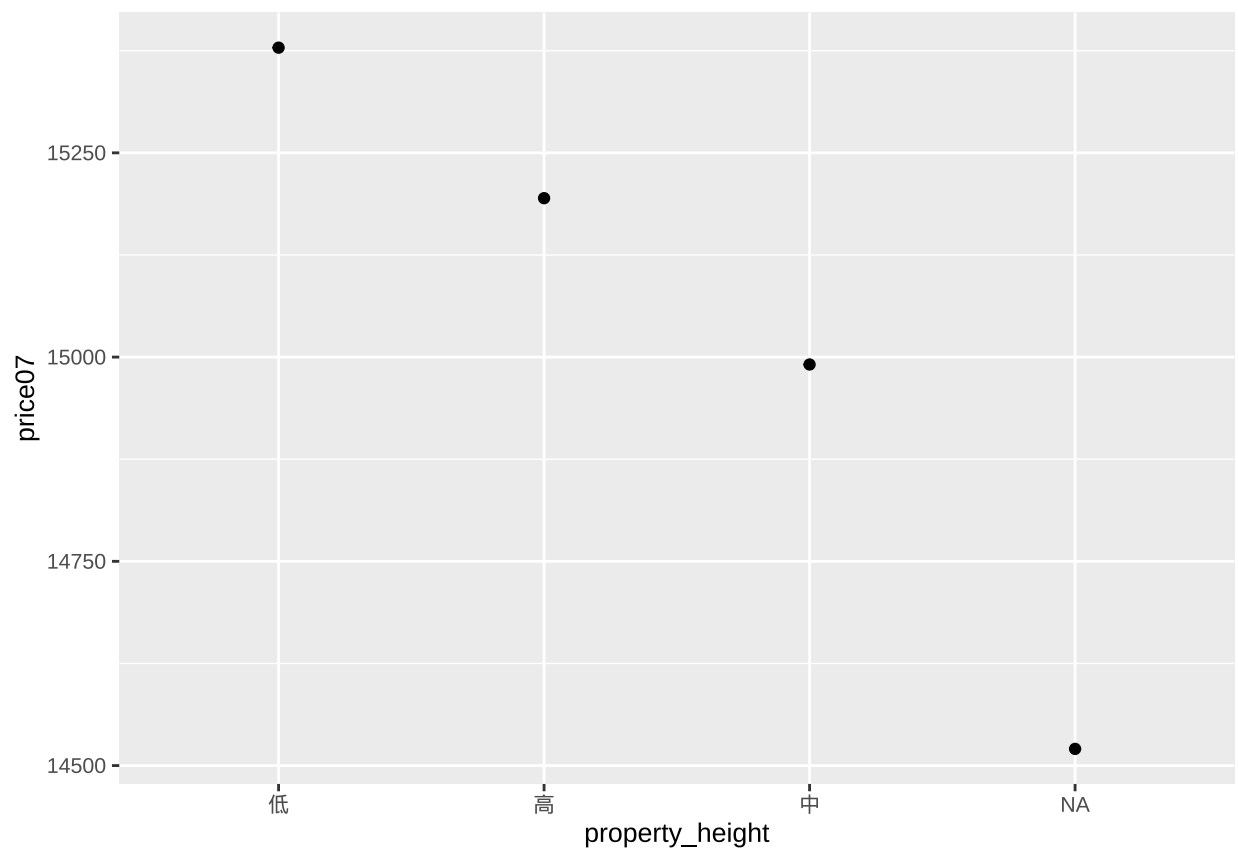
发现：

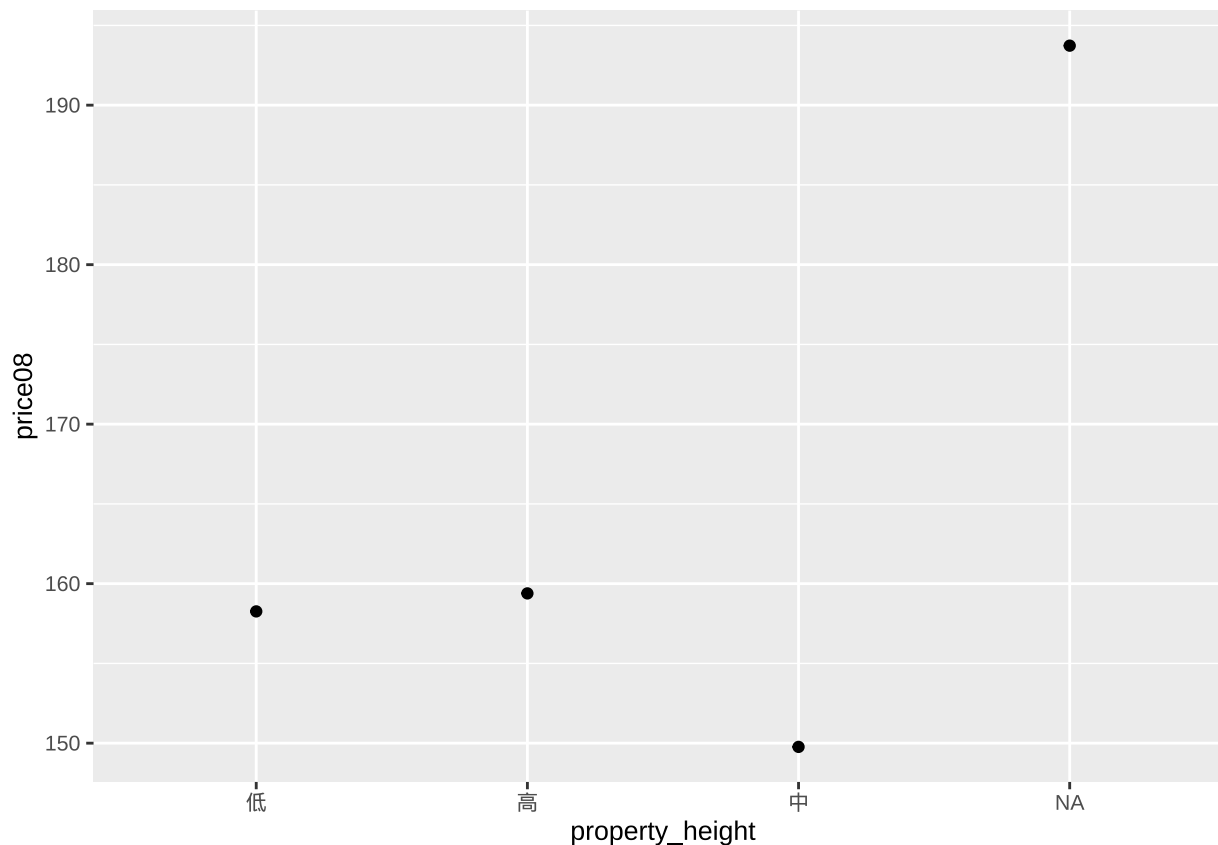
- 发现 1 随着房间数的增加，房源的面积也越大，二者正相关。
- 发现 2 随着房间数的增加，房源的单价和总价均持续增加，在 5 室时达到最高，此后开始回落，说明房间数越多，价格越高。

探索问题 3









发现：

- 发现 1 房源的装修状况对房价有一定的影响，精装房源的挂牌总价和单价均高于简装和毛坯房，精装房平均总价高于简装房近 30 万，精装房平均单价高于简装和毛坯房超过 2000 元，差距不小，说明卖家对房屋的前期投入也同样体现在了价格上，精装房方便买家拎包入住，因此普遍有更高的价格。但毛坯房单价虽略低于简装房，总价却高于简装房，与精装房总价差距较小，可能是因为毛坯房相比较简装房有更大的装修自由度，让买家可以将房屋装修成心仪的样子，同时毛坯房可能意味着建筑年代较新，所以卖家还没来得及装修，至少是从未住过人，这也在一定程度上满足了一些买家的心理。
- 发现 2 房屋在所在楼栋所处位置对房价也有一定的影响，总体差距不大。如果房源属于中楼层，那么总价与单价都处于相对较低的位置。高楼层平均总价在 160 万以内，高于中楼层平均总价近 10 万元，价格更高基本符合开发商对一手房的定价，也就是卖方购房的成本更高，高楼层的采光、隔音效果都更优于其他楼层，导致高楼层的房价也较高。低楼层的高价格可能更多源于上下楼梯便利度的优势，或是受异常值如别墅等超高价格的影响。

发现总结

根据武汉市链家二手房 3000 套房源信息，可以初步发现，武汉市挂牌二手房房源单价均价主要集中在在 14000-15000 元居多，总价在 140 万-150 万元居多，面积在 100 平居多，户型多为两室一厅或三室两厅，精装房源为主，高、中、低楼层都有，板楼多，大多数近地铁，能够随时看房或 VR 看装修。房源地段、户型、楼层、交通便利程度均在不同程度上影响房源价格。存在一些异常值，如中北路、水果湖的房源，以及面积大、房间多的高端别墅或面积小的车库，使得房源差异化大。