

关于武汉链家二手房的数据规律分析

李嘉欣

2023-10-18

目录

| | | |
|-----|---------------------------------|----|
| 1 | 你的主要发现 | 2 |
| 2 | 数据介绍 | 2 |
| 3 | 数据概览 | 2 |
| 4 | 数据清洗与处理 | 7 |
| 5 | 探索性分析 | 11 |
| 5.1 | 变量 1 住宅的价格与总价的数值描述与图形 | 11 |
| 5.2 | 变量 2（建筑形式）的数值描述与图形 | 12 |
| 5.3 | 变量 3（建筑总层数）的数值描述与图形 | 14 |
| 5.4 | 变量 4（朝向）的数值描述与图形 | 16 |
| 5.5 | 探索问题 1 房间个数与房间面积关系 | 18 |
| 5.6 | 探索问题 2：特定小区（东立国际）楼层位置与平均价格关系 | 18 |
| 6 | 发现总结 | 21 |

1 你的主要发现

1. 发现 1：整体发现整体价格：单位面积均价和总面积均价都高于中位数；单位面积均价较高的小区 and 区域较多在武昌区；整体面积：武汉二手房平均面积超 100 m²，为中等户型；两室一厅和三室两厅为较常见的格局；
2. 发现 2：公共配套设施（如地铁）对房价有较明显影响；当前武汉链家中的二手房里，近地铁与不近地铁比例接近 1:1，近地铁的单位面积均价（16623.93 元/m²）要高于近地铁的单位面积均价（13562.86 元/m²）
3. 发现 3：建筑自身格局影响：南向贵，板塔结合贵，总高高的贵，住宅位置高的贵；板楼多，多层 6-8 层，多层中 11、18 层，32-35 层多，主朝向为南向的多，房间个数 2、3、4 个的多；

2 数据介绍

本报告链家网数据获取方式如下：

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 (lj) 共包括 property_name, property_region, price_ttl, price_sqm, bedrooms, livingrooms, building_area, directions1, directions2, deco-

ration, 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, 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 |
|----|------------------|------------------|----------------|---------------|
| ## | 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
##
##
## # A tibble: 87 x 4
##   property_region count_n avg_building_area avg_price_sqm
##   <chr>          <int>          <dbl>          <dbl>
## 1 白沙洲          167          101.          13837.

```

```
## 2 盘龙城          126          108.          8376.
## 3 四新            116          102.          14817.
## 4 光谷东          112          110.          16683.
## 5 金银湖          97          112.          10339.
## 6 后湖            86          109.          14915.
## 7 青山            85           90.3          14228.
## 8 王家湾          78          98.9          13919.
## 9 塔子湖          71          93.0          16186.
## 10 珞狮南路       67          108.          16164.
## # i 77 more rows
```

```
## # A tibble: 1,345 x 4
```

```
##   property_name count_n avg_building_area avg_price_sqm
##   <chr>          <int>          <dbl>          <dbl>
## 1 东立国际         22           85.1          18303.
## 2 保利中央公馆     16           92.6          18921.
## 3 朗诗里程         16           89.6          16994.
## 4 恒大名都         15          119.           8786.
## 5 阳光100大湖第     15          123.           9320.
## 6 保利城一期        13           95.2          16885.
## 7 锦绣龙城         13           97.8          14454
## 8 保利华都         12           89.7          27141.
## 9 统建同安家园      12          106.          14693.
## 10 金地自在城K2     12           86.9          18484
## # i 1,335 more rows
```

```
## # A tibble: 1,345 x 5
```

```
##   property_name count_n avg_followers avg_building_area avg_price_sqm
##   <chr>          <int>          <dbl>          <dbl>          <dbl>
## 1 十里和府         1           262           83.6          10765
## 2 阜华领秀中南     1           209           58.1          11711
## 3 万达公馆         2           139           186.          21542
## 4 中环星干线       3           112           79.9           9512
## 5 汉港大厦         1           112           134.          9704
```

```

## 6 中南新苑西区          1          102          145.          15147
## 7 洪山教委宿舍          2           86          107.          12577
## 8 吉祥谷                1           84          95.0          13259
## 9 越秀星汇君泊越天地    1           79          112.          10298
## 10 越秀国际金融汇二期   2           78          106.          29729
## # i 1,335 more rows

## # A tibble: 218 x 5
##   property_name      count_n avg_followers avg_building_area avg_price_sqm
##   <chr>            <int>      <dbl>          <dbl>          <dbl>
## 1 联投花山郡一期（香颂）      4         59          118.          12843
## 2 保利心语六期                4         53           87.0          18162
## 3 中建东湖锦城                4         52          126.          18384
## 4 米兰映象                    5        48.2          112.          18996.
## 5 兴和嘉园                    4        36.5          106.           8558.
## 6 当代国际城                  4        31.2          112.          19780.
## 7 融科花满庭                  4        25.2          107.          13581.
## 8 中建大公馆一期              6        25.2          123.          25081.
## 9 世茂锦绣长江五期            7        24.7          114.          22322.
## 10 华发中城荟                 4         22          154.          37910.
## # i 208 more rows

## # A tibble: 3,000 x 18
##   property_name property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>          <chr>          <dbl>   <dbl>   <dbl>   <dbl>
## 1 十里和府      七里庙             90    10765      2      2
## 2 阜华领秀中南 中南丁字桥          68    11711      2      1
## 3 米兰映象      光谷广场         175    16948      3      2
## 4 兴和嘉园      后湖              64     7342      2      2
## 5 万达公馆      积玉桥          400    21542      2      2
## 6 万达公馆      积玉桥          400    21542      2      2
## 7 中建大公馆一期 民族大道         258    23841      3      2
## 8 联投时代中心  积玉桥          142    22915      1      1
## 9 枫丹白鹭      黄埔永清         179    15424      3      1

```

```
## 10 新力帝泊湾      盘龙城      70      7740      3      1
## # i 2,990 more rows
## # i 12 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>
```

可以看到：

- 直观结论 1 当前武汉链家中的二手房里：价格：单位面积均价和总面积均价都高于中位数，单位面积均价为 15148 元/m²，总面积均价为 155.9 万元；单位面积均价和总面积均价的最大值和最小值差距较大；单位面积均价较高的小区 and 区域较多在武昌区；面积：武汉二手房平均面积超 100 m²，为中等户型；根据中位数推断，两室一厅和三室两厅为较常见的格局；
- 直观结论 2 当前武汉链家中的二手房里，白沙洲、盘龙城、四新、光谷东的二手房数量较多，市场供给量较大，属于近些年新开发较多楼盘，平均面积在 100-110 m²波动；

4 数据清洗与处理

将有缺失值和错乱值字段的数据进行内容的填充；

```
## # A tibble: 3,000 x 18
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>   <dbl>   <dbl>   <dbl>
## 1 南湖名都A区     南湖沃尔玛     237    18709     3       1
## 2 万科紫悦湾     光谷东         127    14613     3       2
## 3 东立国际       二七           75     15968     1       1
## 4 新都汇         光谷广场      188    15702     3       2
## 5 保利城一期     团结大道      182    17509     3       2
## 6 加州橘郡       庙山          122    10376     3       2
## 7 省建筑五公司西区 光谷广场      99     12346     2       1
```

```
## 8 保利上城东区    白沙洲    194.    16336    3    2
## 9 石化大院        中南丁字桥    325    32631    4    1
## 10 阳光花园       杨汊湖    192    17403    3    2
## # i 2,990 more rows
## # i 12 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>
```

将近地铁字段的数据进行处理为仅有近地铁和不近地铁:

```
## # A tibble: 1,554 x 18
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>    <dbl>    <dbl>    <dbl>
## 1 南湖名都A区    南湖沃尔玛    237     18709     3        1
## 2 东立国际      二七          75     15968     1        1
## 3 新都汇        光谷广场    188     15702     3        2
## 4 省建筑五公司西区 光谷广场    99     12346     2        1
## 5 保利上城东区    白沙洲    194.    16336     3        2
## 6 石化大院      中南丁字桥    325     32631     4        1
## 7 阳光花园      杨汊湖    192     17403     3        2
## 8 东湖1号      中北路    630     27273     5        2
## 9 光谷创新天地1期 光谷东    470     19003     4        2
## 10 佰昌公馆B区   堤角    118     13119     3        1
## # i 1,544 more rows
## # i 12 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>
```

```
## # A tibble: 1,446 x 18
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>    <dbl>    <dbl>    <dbl>
## 1 万科紫悦湾    光谷东    127     14613     3        2
## 2 保利城一期    团结大道    182     17509     3        2
```



```

## 3 加州橘郡      庙山      122      10376      3      2
## 4 银泰御华园    百步亭      245      20604      3      2
## 5 锦绣龙城      民族大道    127      14642      2      2
## 6 三环花园      阳逻      36       3739      3      2
## 7 长港路303号    杨汊湖    130      11453      3      2
## 8 汉飞洋房印象  盘龙城      60       6266      3      1
## 9 爱家名校华城  白沙洲     83.8     9279      3      2
## 10 翡翠城1期     长丰常码头 123      11348      3      2
## # i 1,436 more rows
## # i 12 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>

## # A tibble: 76 x 4
##   property_region count_n avg_building_area avg_price_sqm
##   <chr>           <int>           <dbl>           <dbl>
## 1 白沙洲           96           104.           15268.
## 2 青山             56           89.6           14363.
## 3 塔子湖           50           94.9           16494.
## 4 珞狮南路         47           104.           16650.
## 5 积玉桥           46           119.           22650.
## 6 王家湾           43           96.5           14164.
## 7 光谷东           42           118.           20513.
## 8 钟家村           41           102.           15910.
## 9 后湖             39           110.           14620.
## 10 老南湖          39           102.           20742.
## # i 66 more rows

## [1] 16623.93

## # A tibble: 76 x 4
##   property_region count_n avg_building_area avg_price_sqm
##   <chr>           <int>           <dbl>           <dbl>
## 1 四新             95           104.           14818.

```

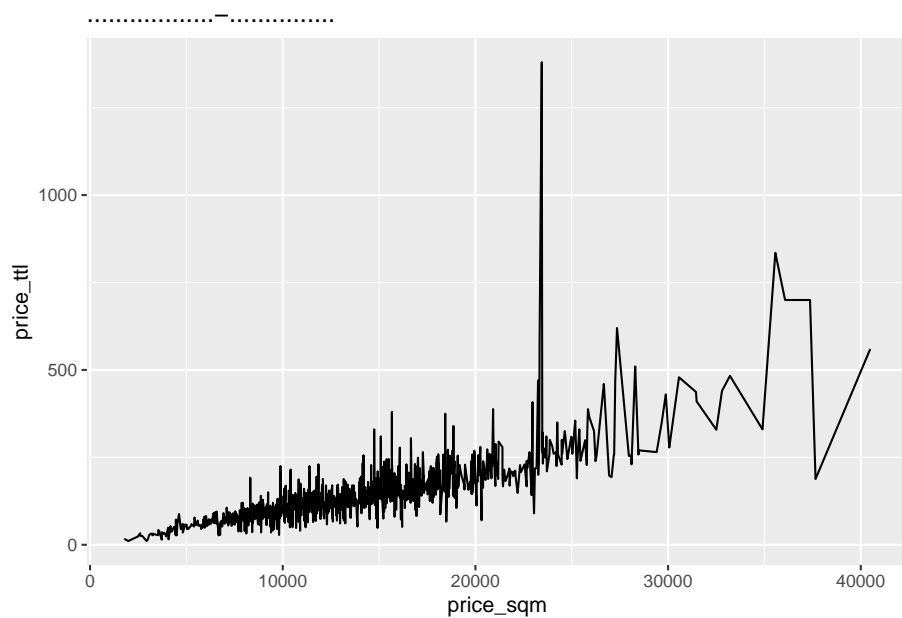
```
## 2 盘龙城          91          106.          8025.
## 3 白沙洲          71          97.7          11901.
## 4 光谷东          70          105.          14386
## 5 金银湖          65          112.          9451.
## 6 民族大道        61          98.4          19622.
## 7 阳逻            59          101.          4918.
## 8 后湖            47          108.          15161.
## 9 百步亭          40          101.          16084.
## 10 王家湾         35          102.          13617.
## # i 66 more rows

## [1] 13562.86
```

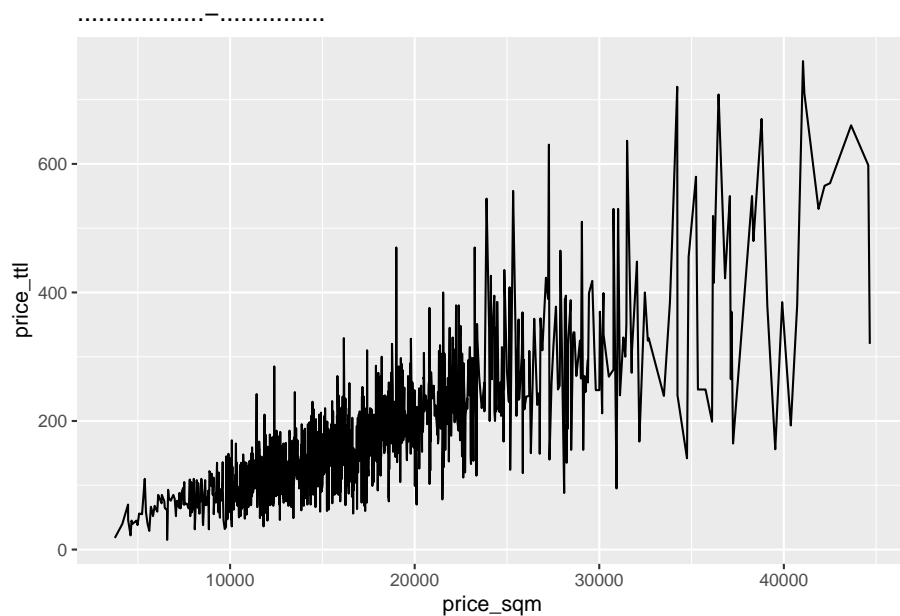
- 直观结论 3 当前武汉链家中的二手房里，近地铁与不近地铁比例接近 1:1，近地铁的单位面积均价（16623.93 元/m²）要高于不近地铁的单位面积均价（13562.86 元/m²）

5 探索性分析

5.1 变量 1 住宅的价格与总价的数值描述与图形



- 发现 1 不近地铁的小区整体来说，小区整体价格在 23000 元/m²前和单位面积价格呈正相关；



- 发现 2 近地铁的小区整体来说，小区整体价格和单位面积价格呈正相关，但是波动幅度较大，受其他因素影响较多；

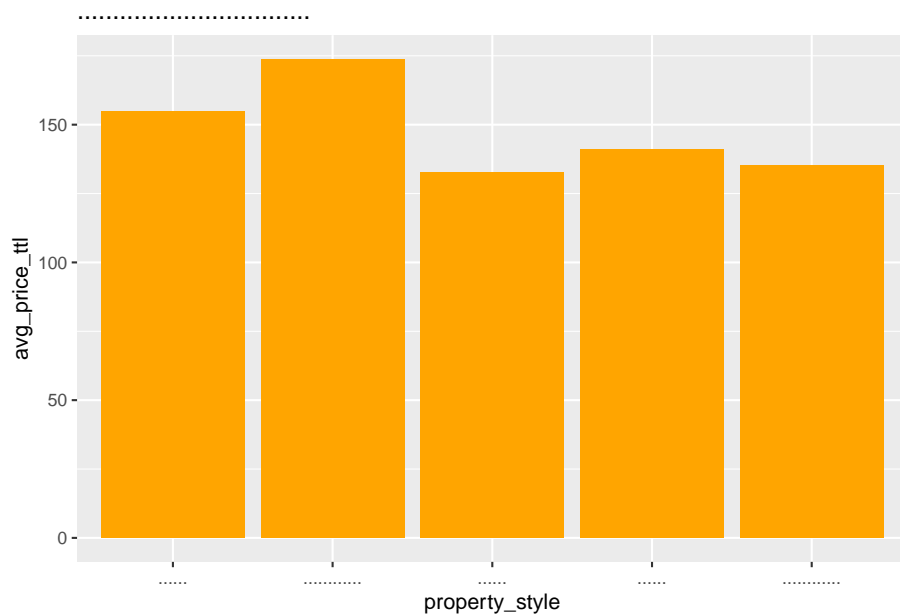
5.2 变量 2（建筑形式）的数值描述与图形

```
## [1] 152.0895
```

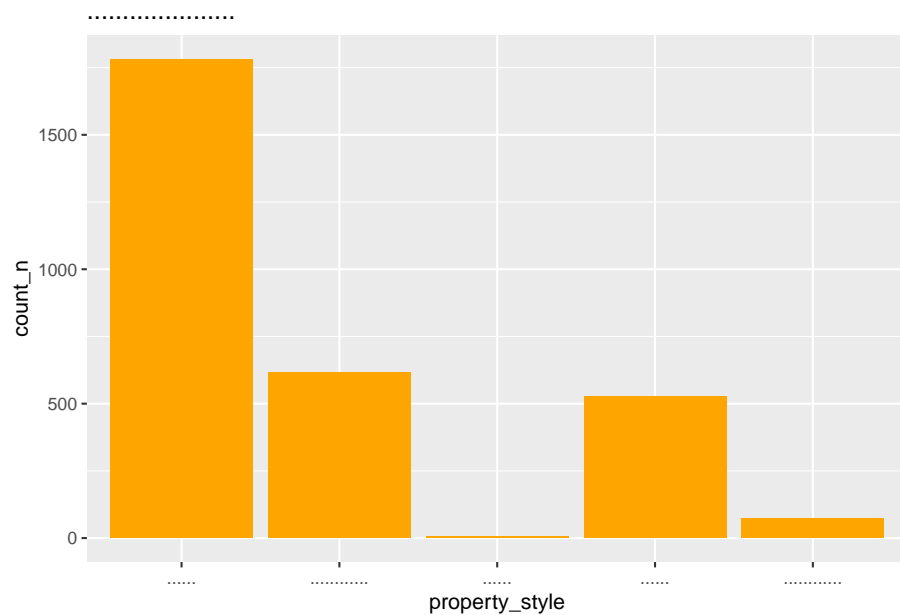
```
## [1] 158.6101
```

```
## # A tibble: 5 x 3
```

| ## | property_style | count_n | avg_price_ttl |
|------|----------------|---------|---------------|
| ## | <chr> | <int> | <dbl> |
| ## 1 | 板楼 | 1781 | 155. |
| ## 2 | 板塔结合 | 615 | 174. |
| ## 3 | 塔楼 | 527 | 141. |
| ## 4 | 暂无数据 | 72 | 135. |
| ## 5 | 平房 | 5 | 133. |



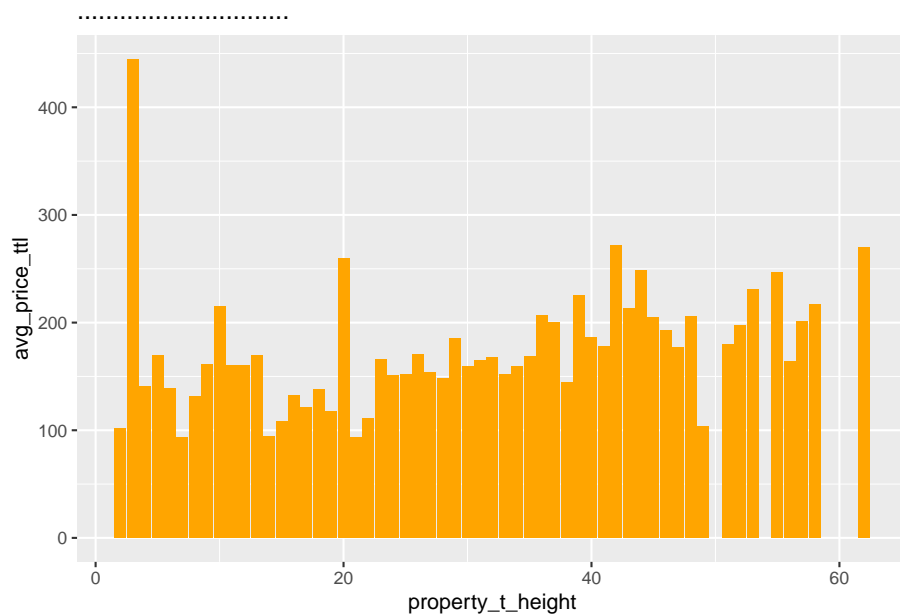
- 发现 1 建筑形式均价：相对来说板塔结合的平均价格高于塔楼和板楼的房子，板塔结合的建筑的性能各方面较好；



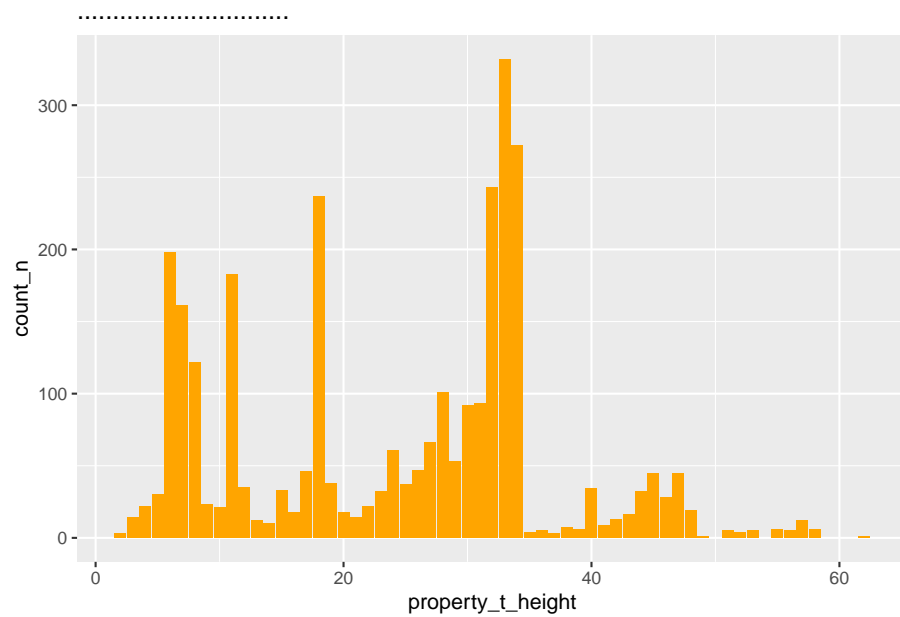
- 发现 2 建筑形式数量：相对来说，武汉二手房是板楼属性的数量远远高于其他类型的建筑形式；

5.3 变量 3（建筑总层数）的数值描述与图形

```
## # A tibble: 56 x 3
##   property_t_height count_n avg_price_ttl
##           <dbl>    <int>         <dbl>
## 1             33      332          152.
## 2             34      272          160.
## 3             32      243          168.
## 4             18      237          138.
## 5              6      198          139.
## 6             11      183          161.
## 7              7      161           93.5
## 8              8      122          131.
## 9             28      101          148.
## 10            31       93          165.
## # i 46 more rows
```



- 发现 1 建筑总层数：相对来说，总层高高于 100m 的超高层，均价高于高层，高于多层；2 层，商品房作为商铺，均价最高；20 层和 42 层价格相对来说较高；

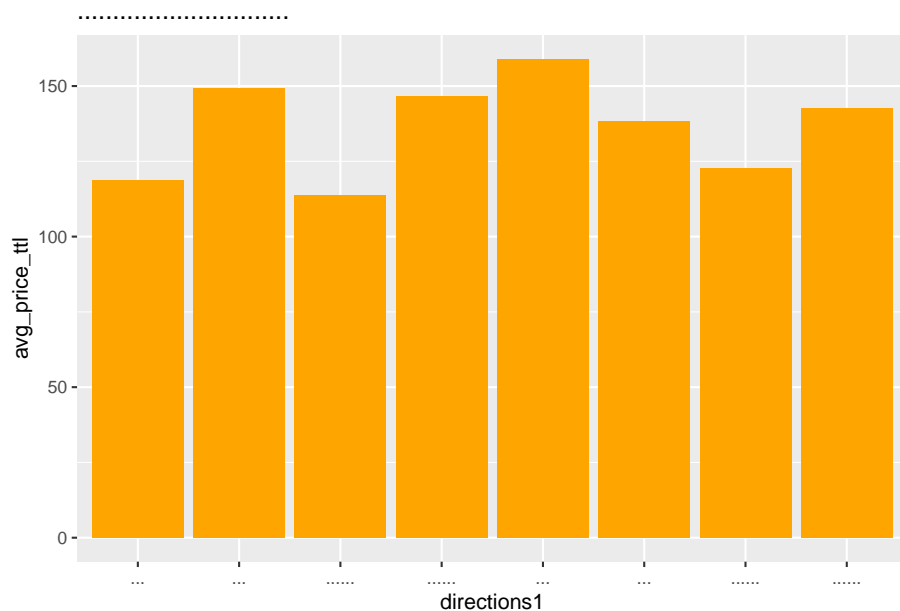


- 发现 2

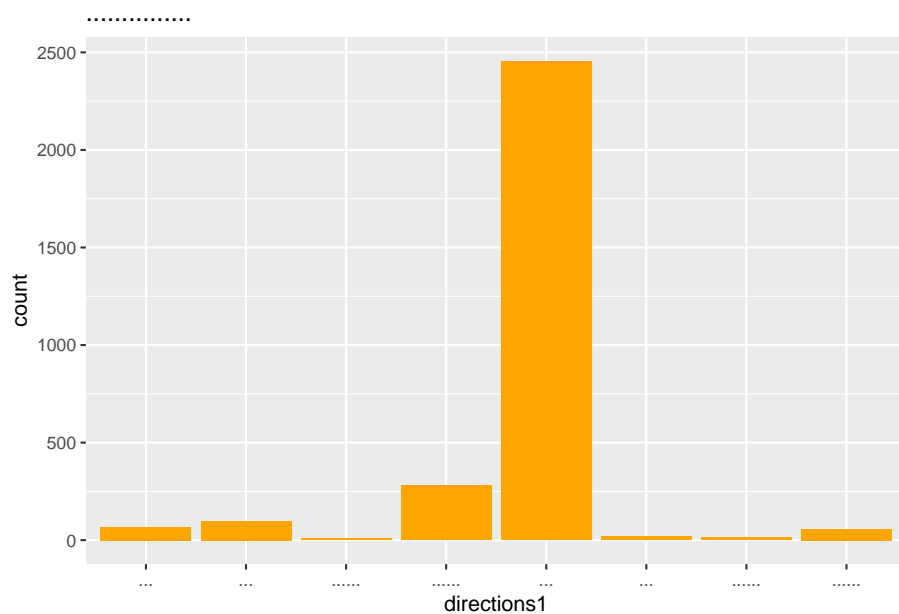
建筑总层数：相对来说，多层 6-8 层，多层中 11、18 层；超高层中，32-35 层为开发商较为喜欢的层高，可以达到各项最优，效率最高；

5.4 变量 4（朝向）的数值描述与图形

```
## # A tibble: 8 x 3
##   directions1 count_n avg_price_ttl
##   <chr>         <int>     <dbl>
## 1 南             2454      159.
## 2 东南           281      147.
## 3 东             98      149.
## 4 北             68      119.
## 5 西南           57      143.
## 6 西             19      138.
## 7 西北           13      123.
## 8 东北           10      114.
```

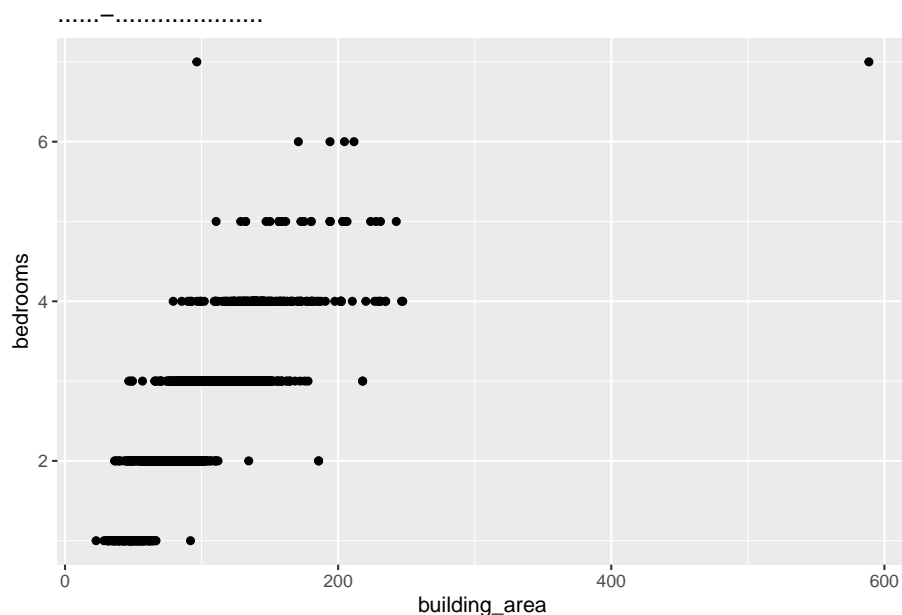


- 发现 1 建筑朝向：相对来说，南向、东、东南向为日照采光最好的朝向，均价会受朝向较大影响；东北和西北向为建筑朝向较差，故均价最低；



- 发现 2 建筑朝向：相对来说，南向、东、东南向为日照采光最好的朝向，开发商更多会设计南向作为主要朝向，因此南向、东向、东南向数量较多；

5.5 探索问题 1 房间个数与房间面积关系

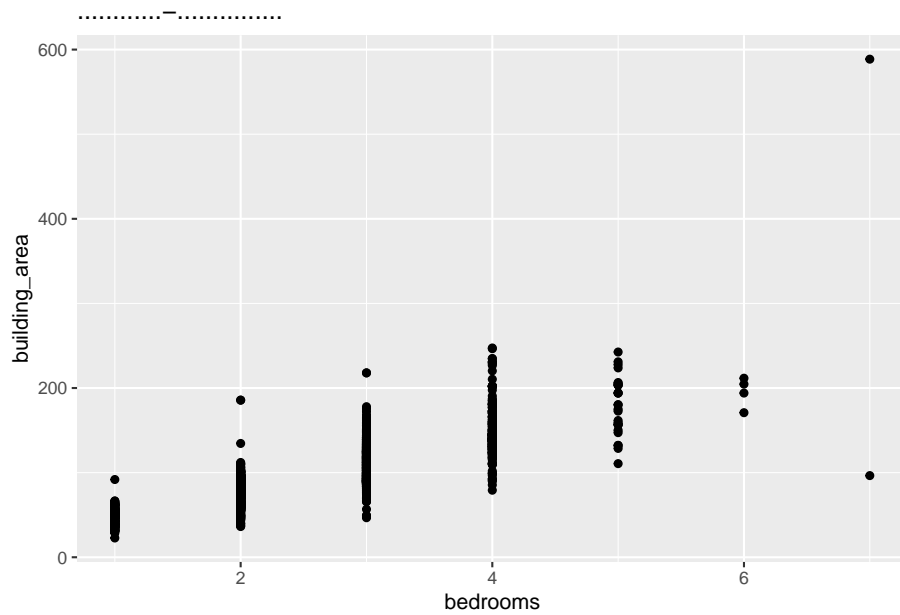
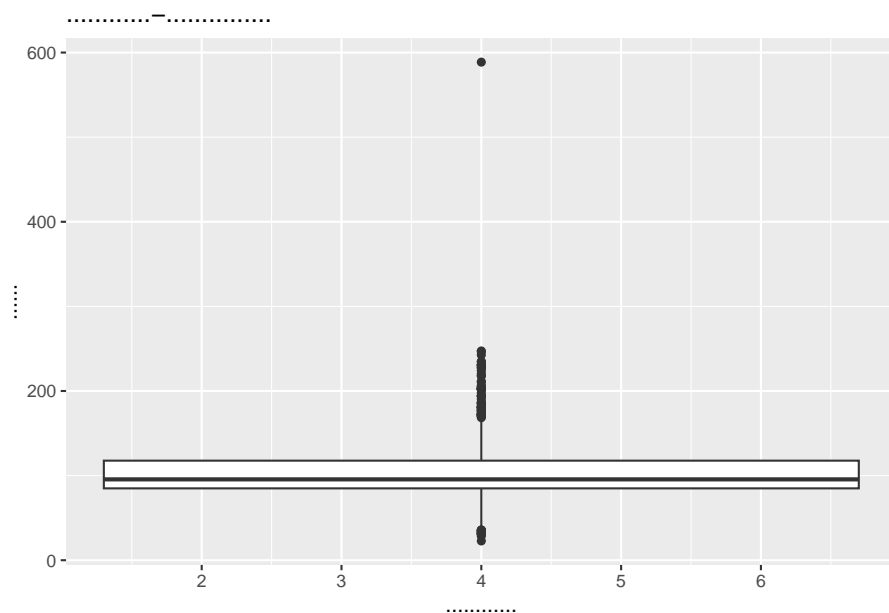


- 发现 1 房间个数 2、3、4 较多；2 个房间的较少超过 120 m²，3 个房间较少超过 180 m²

5.6 探索问题 2：特定小区（东立国际）楼层位置与平均价格关系

```
## # A tibble: 8 x 3
##   directions1 count_n avg_price_ttl
##   <chr>         <int>         <dbl>
## 1 南             2454           159.
## 2 东南            281           147.
## 3 东              98           149.
## 4 北              68           119.
## 5 西南            57           143.
## 6 西              19           138.
## 7 西北            13           123.
```

```
## 8 东北      10      114.
```

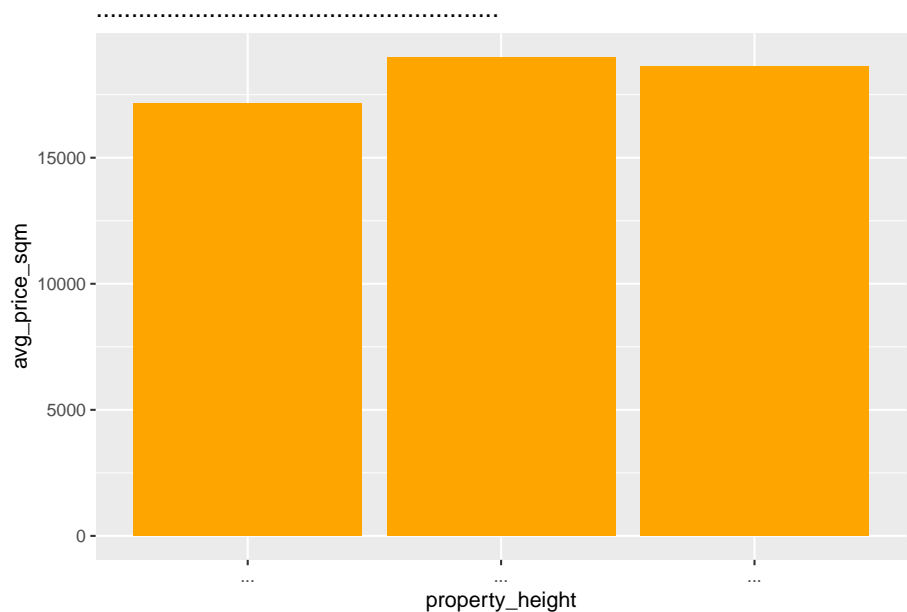


```
## # A tibble: 22 x 18
```

```
##   property_name property_region price_ttl price_sqm bedrooms livingrooms
```

```
##      <chr>          <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 东立国际      二七              75     15968        1        1
## 2 东立国际      二七             208     20853        3        2
## 3 东立国际      二七             153     17373        2        2
## 4 东立国际      二七             128     17463        2        1
## 5 东立国际      二七              80     18601        1        1
## 6 东立国际      二七             180     18064        3        2
## 7 东立国际      二七             180     18064        3        2
## 8 东立国际      二七             180     18064        3        2
## 9 东立国际      二七             214     24593        3        2
## 10 东立国际     二七             176     17709        3        2
## # i 12 more rows
## # i 12 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>

## # A tibble: 3 x 3
##   property_height count_n avg_price_sqm
##   <chr>          <int>    <dbl>
## 1 高              9     18991.
## 2 低              7     17157.
## 3 中              6     18607.
```



- 发现 1 东立小区住宅位置较低的单位面积单价要低位置高和中的；同一小区里，位置高单位面积的房间较高，可能与通风采光条件更好有关；

6 发现总结

1. 整体表现：当前武汉链家中的二手房里：价格：单位面积均价和总面积均价都高于中位数，单位面积均价为 15148 元/m²，总面积均价为 155.9 万元；单位面积均价和总面积均价的最大值和最小值差距较大；单位面积均价较高的小区 and 区域较多在武昌区；面积：武汉二手房平均面积超 100 m²，为中等户型；根据中位数推断，两室一厅和三室两厅为较常见的格局；当前武汉链家中的二手房里，白沙洲、盘龙城、四新、光谷东的二手房数量较多，市场供给量较大，属于近些年新开发较多楼盘，平均面积在 100-110 m²波动；当前武汉链家中的二手房里，近地铁与不近地铁比例接近 1:1，近地铁的单位面积均价（16623.93 元/m²）要高于近地铁的单位面积均价（13562.86 元/m²）

2. 配套：武汉二手房符合一般规律，近地铁仍然单位面积均价会有较大差异；不近地铁的小区整体来说，小区整体价格在 23000 元/m²前和单位面积价格呈正相关；近地铁的小区整体来说，小区整体价格和单位面积价格呈正相关，但是波动幅度较大；

3. 建筑自身格局影响：南向贵，板塔结合贵，总高高的贵，住宅位置高的贵；板楼多，多层 6-8 层，多层中 11、18 层，32-35 层多，主朝向为南向的多，房间个数 2、3、4 个的多；建筑形式均价：相对来说板塔结合的平均价格高于塔楼和板楼的房子，板塔结合的建筑的性能各方面较好；建筑形式数量：相对来说，武汉二手房是板楼属性的数量远远高于其他类型的建筑形式；建筑总层数：相对来说，总层高高于 100m 的超高层，均价高于高层，高于多层；2 层，商品房作为商铺，均价最高；20 层和 42 层价格相对来说较高；建筑总层数：相对来说，多层 6-8 层，多层中 11、18 层；超高层中，32-35 层为开发商较为喜欢的层高，可以达到各项最优，效率最高；建筑朝向：相对来说，南向、东、东南向为日照采光最好的朝向，均价会受朝向较大影响；东北和西北向为建筑朝向较差，故均价最低；建筑朝向：相对来说，南向、东、东南向为日照采光最好的朝向，开发商更多会设计南向作为主要朝向；房间个数 2、3、4 较多；2 个房间的较少超过 120 m²，3 个房间较少超过 180 m² 特定小区（东立小区住宅个数最多）住宅位置较低的单位面积单价要低位置高和中的；同一小区里，位置高单位面积的房间较高，可能与通风采光条件更好有关；