

武汉二手房数据分析报告

张超颖

2023-10-19

目录

1	你的主要发现	1
2	数据介绍	2
3	数据概览	2
4	探索性分析	5
4.1	房屋总价的数值描述与图形	5
4.2	房屋单价的数值描述与图形	6
4.3	装修状况的数值描述与图形	7
4.4	武汉各区域房价是否具有明显的差别?	7
4.5	武汉二手房的建筑面积是否对房价有影响?	10
4.6	房屋所在地区和房屋建筑形式有关系吗?	13
5	发现总结	13

1 你的主要发现

1. 武汉市二手房各项基本特征及房源分布情况
2. 区域对二手房房价的影响
3. 房价与建筑面积的相关性
4. 建筑形式与区域的联系

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。

##	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"		

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, 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      <fct> 南湖名都A区, 万科紫悦湾, 东立国际, 新都汇, 保利城一~
## $ property_region    <fct> 南湖沃尔玛, 光谷东, 二七, 光谷广场, 团结大道, 庙山, ~
## $ 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~
```

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

```
## $ building_area      <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, ~
## $ directions1        <fct> 南, 南, 南, 北, 东南, 南, 南, 南, 南, 南, 南, 南, 东~
## $ directions2        <fct> 北, NA, NA, 东, NA, 北, NA, 北, 北, 北, 北, NA, 西南~
## $ decoration          <fct> 精装, 精装, 简装, 精装, 简装, 精装, 简装, 其他, 简装~
## $ property_t_height  <dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ~
## $ property_height     <fct> 中, 中, 低, 高, 中, 低, 低, 中, 低, 低, 高, 高, 中, ~
## $ property_style      <fct> 塔楼, 板楼, 塔楼, 塔楼, 板塔结合, 板楼, 板楼, 板塔结~
## $ followers           <dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ~
## $ near_subway         <fct> 近地铁, NA, 近地铁, 近地铁, NA, NA, 近地铁, 近地铁, ~
## $ if_2y               <fct> NA, 房本满两年, NA, 房本满两年, 房本满两年, 房本满两~
## $ has_key             <fct> 随时看房, 随时看房, 随时看房, 随时看房, 随时看房, 随~
## $ vr                  <fct> NA, VR看装修, NA, NA, VR看装修, NA, VR看装修, NA, NA~
```

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
## 3rd Qu.:33.00 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\学区房: 1
## 随时看: 1
## 保利拉: 1
## 育才花: 1
## (Other) : 6
## NA's : 906

```

可以看到:

- 武汉市二手房房源最多的区域为白沙洲有 167 套二手房, 其次为盘龙城 126 套和四新 116 套。
- 武汉市二手房平均房屋总价为 155.9 万元, 中位房屋总价为 137 万元; 二手房平均每平方米 1.51 万元, 中位数为 1.44 万元。
- 武汉市二手房的平均房屋面积为 100.87 平方米, 大多数房屋的主要朝向是南方, 精装房的数量最多为 1757 套。

- 武汉市二手房总层数平均数为 24.22 层，中位数为 27 层，但最多的建筑形式为板楼，板楼一般建筑层数不会超过 12 层

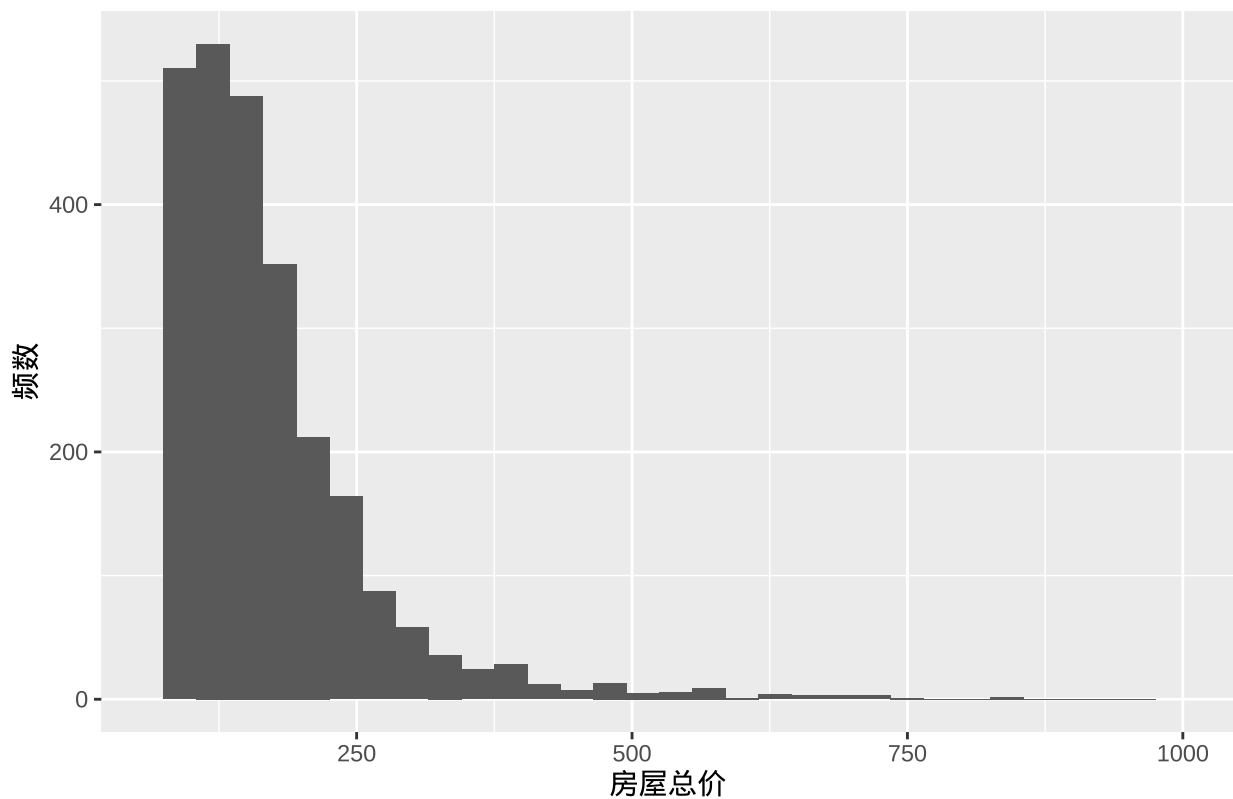
4 探索性分析

4.1 房屋总价的数值描述与图形

```
## # A tibble: 1 x 13
##   variable      n  min  max median    q1    q3   iqr  mad  mean   sd   se
##   <fct>      <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 price_ttl 3000  10.6 1380   137    95   188   93  66.7  156.  95.5  1.74
## # i 1 more variable: ci <dbl>

## [1] 105
```

武汉市二手房房屋总价直方图



发现：

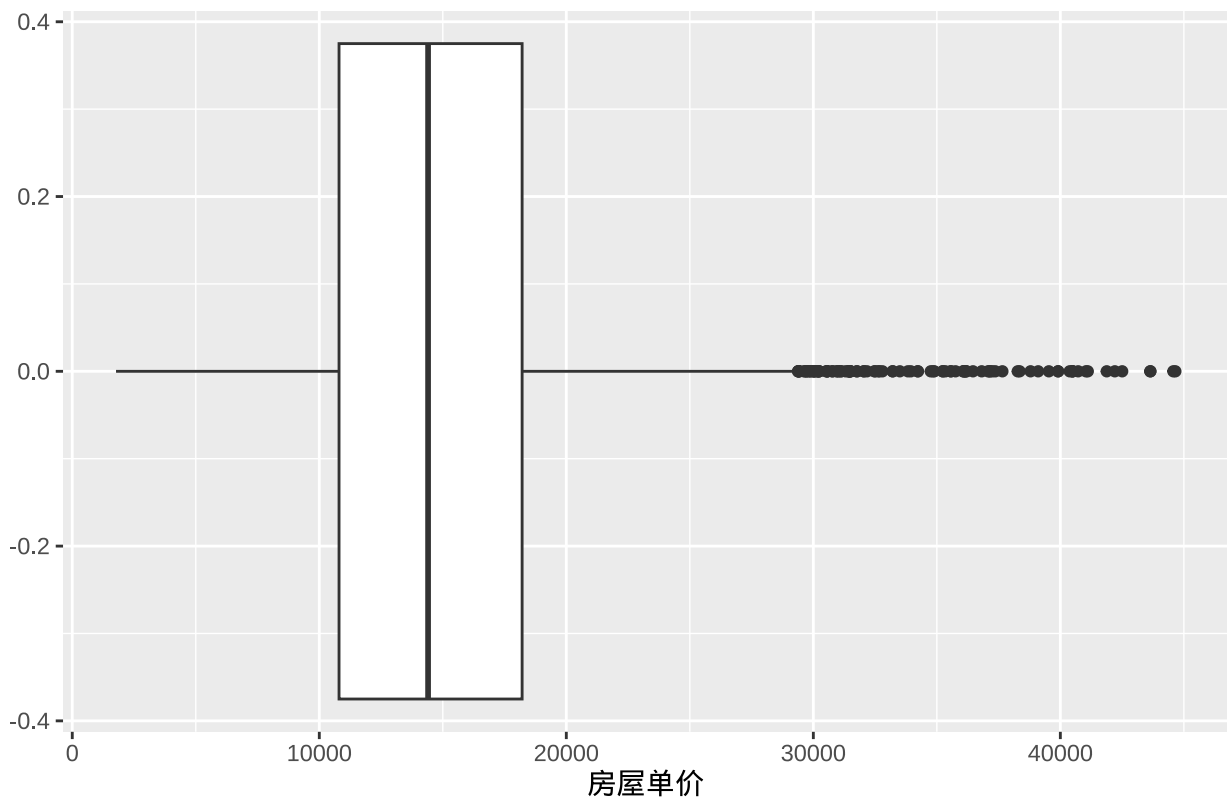
- 从直方图可以看出，数据分布整体右偏非常明显，武汉市二手房的价格主要集中在 100 万元到 250 万元区间，其中 100 万元到 175 万元占多数，500 万元以上占少数，最高达到 1380 万元。也符合我们对房价的认知，大部分人还是只能承担较低的房价。

4.2 房屋单价的数值描述与图形

```
## # A tibble: 1 x 13
##   variable      n  min  max median    q1    q3  iqr  mad  mean  sd  se
##   <fct>      <dbl> <dbl> <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 price_sqm 3000  1771 44656 14404 10799. 18211 7412. 5466. 15148. 6323. 115.
## # i 1 more variable: ci <dbl>

## [1] 14394 15123 15661
```

武汉市二手房房屋单价箱型图

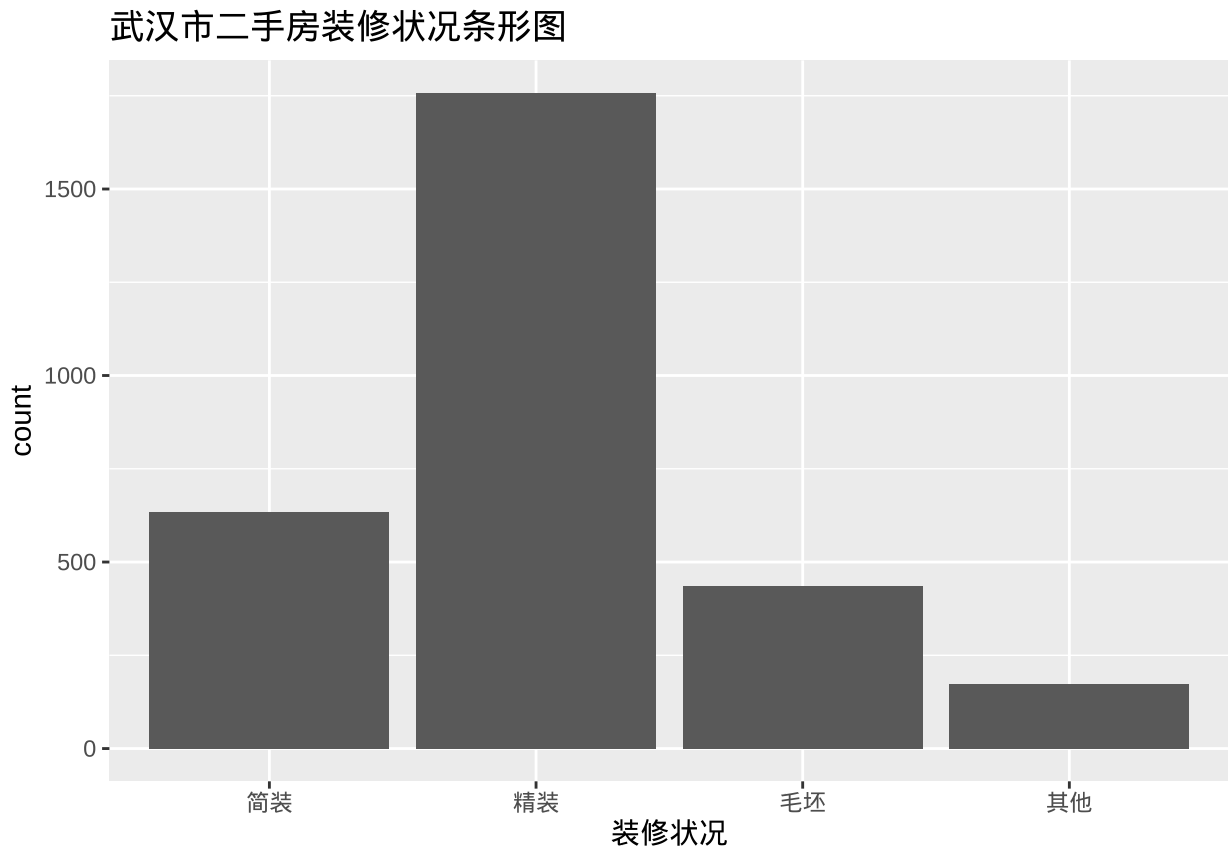


发现：

- 武汉市二手房房屋单价成典型的右偏态，异常值集中在较大值的一侧，尾部很长，这说明很多二手房的房屋单价严重偏离武汉市二手房平均水平，房价差异较大。
- 武汉市二手房房屋单价主要在 1.0-1.8 万元之间。

4.3 装修状况的数值描述与图形

```
##  
##      简装      精装      毛坯      其他  
## 21.133333 58.566667 14.533333  5.766667
```



发现:

- 武汉市二手房房源装修状况主要以精装修为主，占有二手房中的 58.57%，其次为简装房，占比为 21.13%。

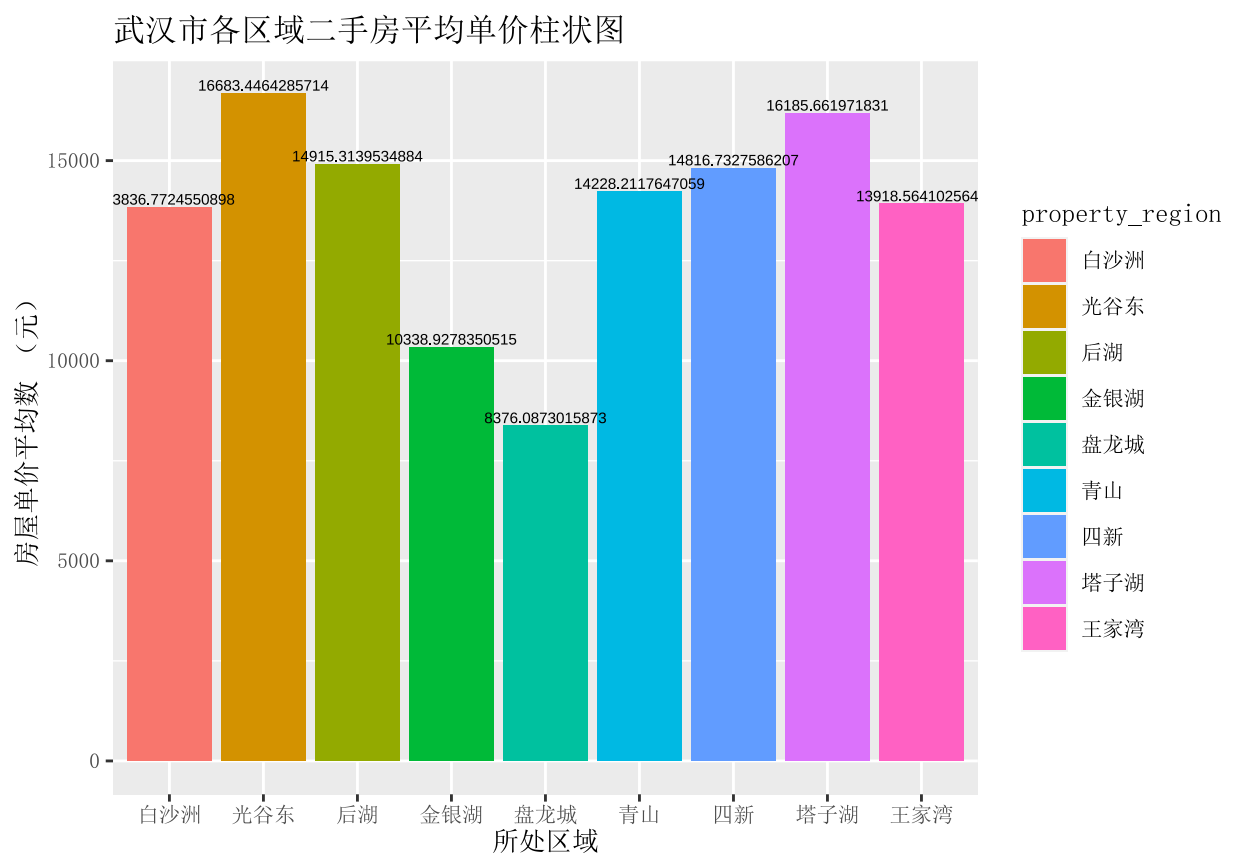
4.4 武汉各区域房价是否具有明显的差别？

武汉市各区域二手房数量超过 70 套的区域有：

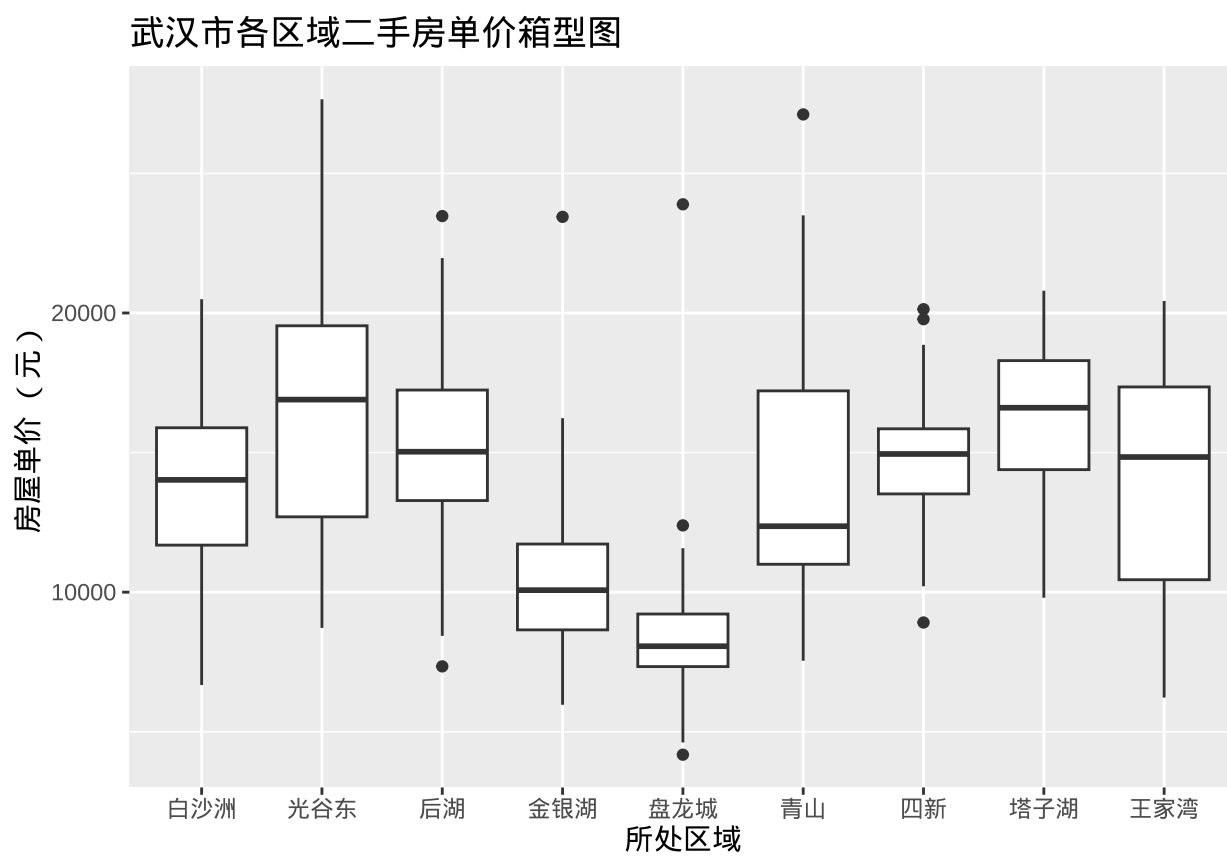
```
## # A tibble: 9 x 2  
##   property_region      n  
##   <fct>             <int>  
## 1 白沙洲             167
```

## 2	光谷东	112
## 3	后湖	86
## 4	金银湖	97
## 5	盘龙城	126
## 6	青山	85
## 7	四新	116
## 8	塔子湖	71
## 9	王家湾	78

武汉市各区域二手房平均单价柱状图:

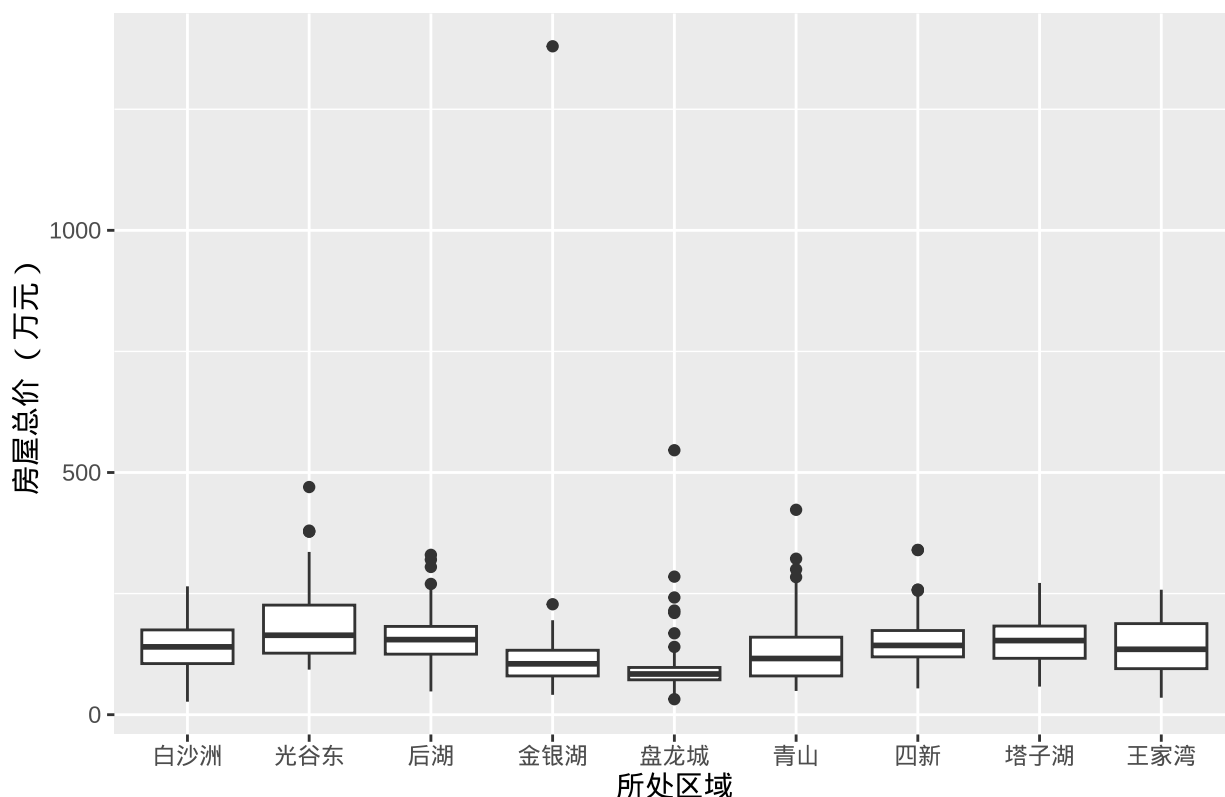


武汉市各区域二手房单价箱型图:



武汉市各区域二手房总价箱型图:

武汉市各区域二手房总价箱型图



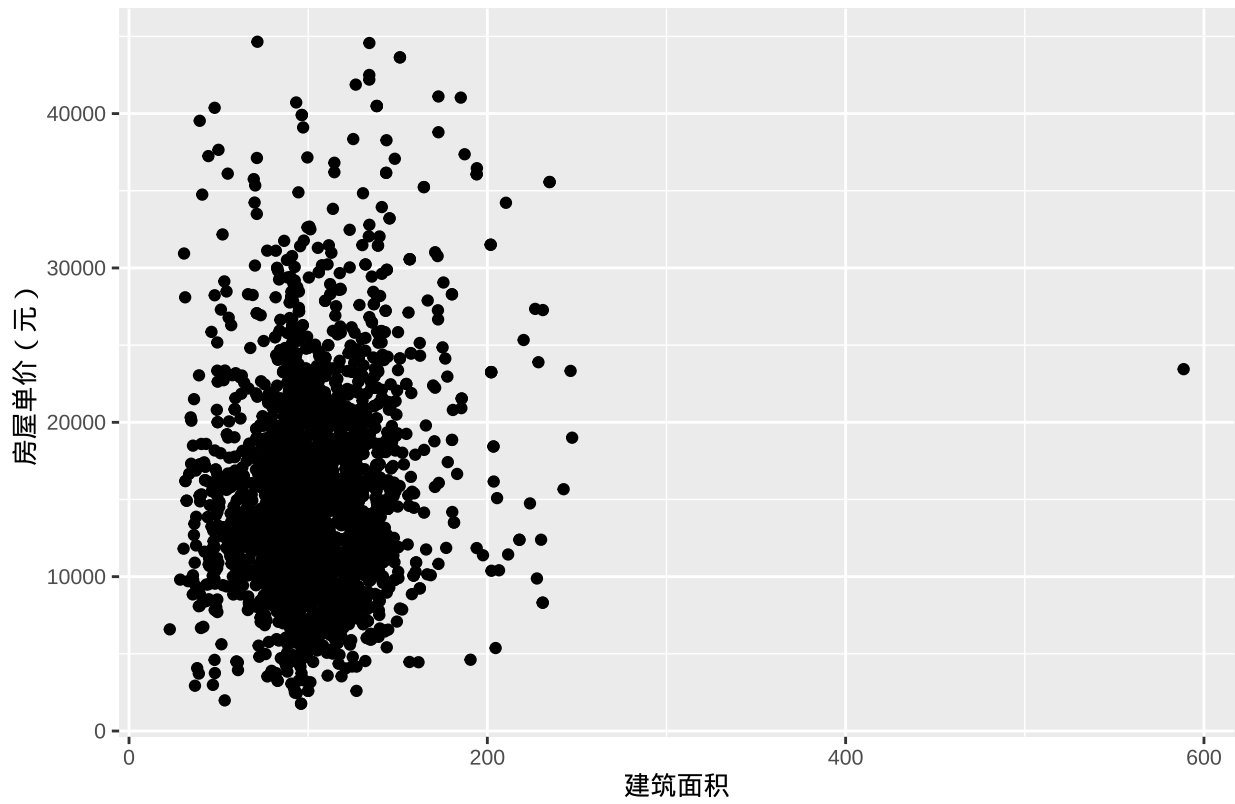
发现:

- 在二手房数量大于 70 的区域中,我们从柱状图中可以看出,光谷东和塔子湖平均单价最高,都在 16000 元/平方米左右。光谷东由于地处东湖高新区,近几年有很多的大型企业入驻,就业机会相对较多,因此刚需购房者较多,所以房价居高不下;而平均房价最低的盘龙城平均单价仅在 8376 元/平方米,也是因为盘龙城地处武汉远城区,离武汉市各中心有一定的距离,所以房价较为低廉。
- 二手房平均单价虽然是一个重要参考数据,但平均值不能有效的表示出数据整体上的分布情况,特别是数据中一些离散值的分布情况,这些信息的表现则需要借助箱型图。从武汉市各区域二手房单价箱型图来看,塔子湖,白沙洲和金银湖房屋单价在中位数两侧均匀分布,而青山的房屋单价多集中于高价区,平均数明显大于中位数,呈右偏分布。与之相反的是光谷东和王家湾的房屋单价集中在低价区。从异常值的角度上分析,盘龙城内平均单价差异最大,其次是金银湖;白沙洲,塔子湖,和王家湾的房屋总价相对较为集中,也可以看出武汉市大多数二手房都在 100 万元-200 万元之间。同时在二手房数量大于 70 的区域中,房屋总价最高的二手房在金银湖。房屋平均单价最高的光谷东,平均房屋总价也最高。盘龙城和青山这两个区域内房屋总价差异也比较大。

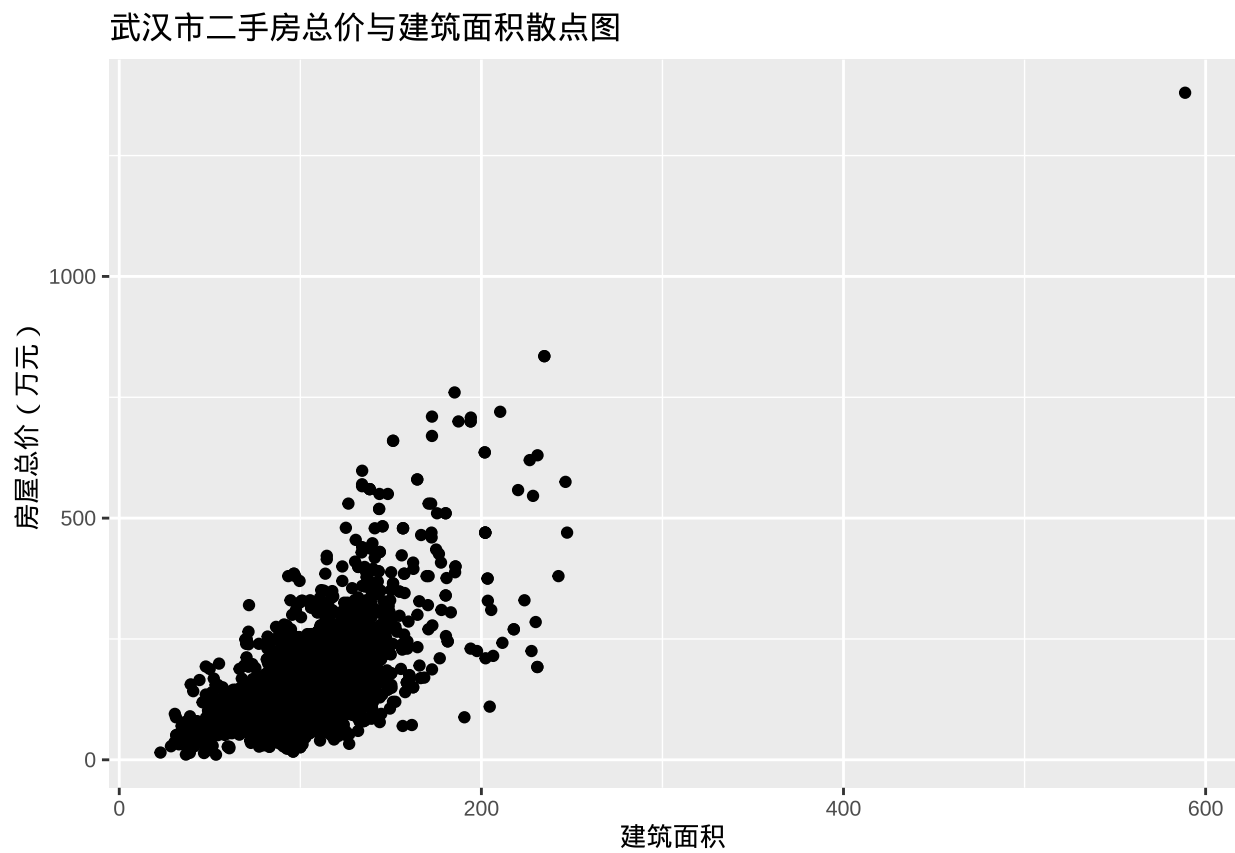
4.5 武汉二手房的建筑面积是否对房价有影响?

武汉市二手房单价与建筑面积散点图

武汉市二手房单价与建筑面积散点图



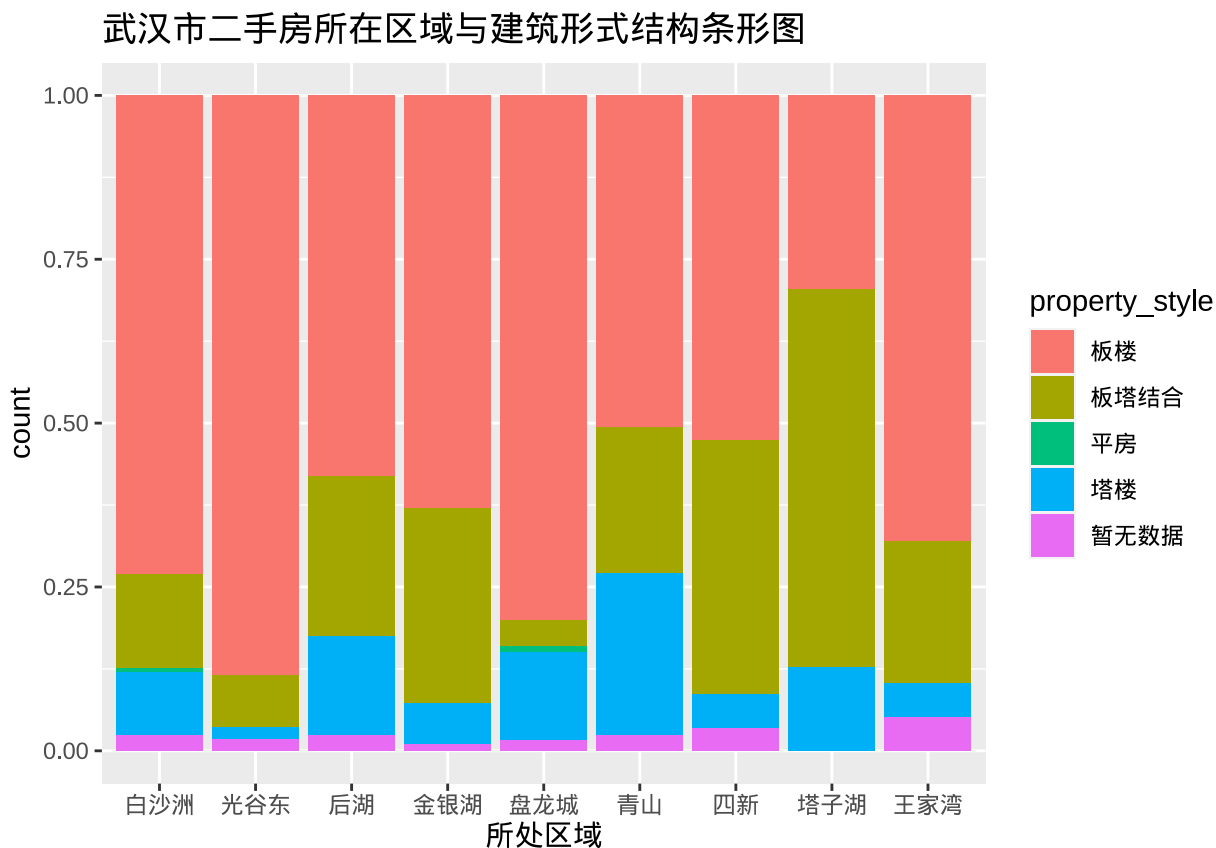
武汉市二手房总价与建筑面积散点图



发现：

- 房屋面积和房屋总价为正相关，数据点分布也比较集中在总价 0-500 万元之间，建筑面积 0-200 平米之间。
- 房屋面积和房屋单价并没有明显关系，样本分布也较为集中，离散值不多。

4.6 房屋所在地区和房屋建筑形式有关系吗？



发现：

- 武汉市不同区域的主要房屋建筑形式不同，可以发现大多数区域二手房的主要建筑形式为板楼，而塔子湖区域二手房的主要建筑形式为板塔结合。塔楼在青山所在的比例比其他都要高。

5 发现总结

根据上述分析，可以总结出以下几点结论：

- 武汉市二手房平均房屋总价为 155.9 万元，平均每平方米 15100 万元。房屋总价和单价都成右偏态，房价差异较大。总价集中在 100 万元-250 万元区间，单价主要在 1.0-1.8 万元之间。二手房所处的地区对于房价有较为明显的影响，其中光谷东总价和平均单价最高，单价多集中于低价区，而总价却集中于高价区，可以说明光谷东房屋的建筑面积一般都比较大。而受到地理位置的影响，地处远城区的盘龙城总价和平均单价最低，但单价却差异很大，最高可达 24000 万元/平方米，最低为 4000 元/平方米。

- 武汉市二手房的房屋面积和房屋总价为正相关，但房屋面积和房屋单价并没有明显关系。另外，武汉市不同区域的主要房屋建筑形式不同，总体主要建筑形式是板楼。大多数二手房的主要朝向是南方，精装房的数量最多，占有二手房中的 58.57%。