

第一次作业-探索武汉二手房市场

赵炜

2024-11-03

目录

主要发现	1
数据介绍	2
数据概览	2
探索性分析	4
房子总价的数值描述与图形	4
房价的数值描述与图形	5
探索房价与区域的关系	6
房子类型与房价的关系	7
交通便利度与房价的关系	9
发现总结	11

主要发现

- 1. 发现 1：武汉二手房总价的平均值在 140 万元，在 70 万元到 170 万元之间的房子最多；房价均价在 1.4 万元每平方米，大部分集中在 1 万到 1.8 万之间。
- 2. 发现 2：武汉不同区域差异房价明显，水果湖区域房子总价和每平方的均价都很高，尤其是板楼价格很坚挺；华科大与关山大道板块的板塔结合的楼挂牌较多，而且总价与每平方的均价都很贵，不愧是“武汉宇宙中心”。
- 3. 发现 3：房子周边交通的便利性也影响着房价，靠近地铁的房子明显要比不靠近的贵，平均贵约 2000 元/平方米。

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。

```
#> # 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>
```

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, 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", "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:

- 直观结论 1: 50% 多的房子的总价在 95 万到 188 万之间, 房子均价在 1 万每平方到 1.8 万每平方, 建筑面积在 95 平方到 117 平方;
- 直观结论 2: 精装房最多, 大部分是 3 房, 约 60% 是板楼;
- 近一半的房子没有填写是否满 2 年; 近一半的房子不靠近地铁。

探索性分析

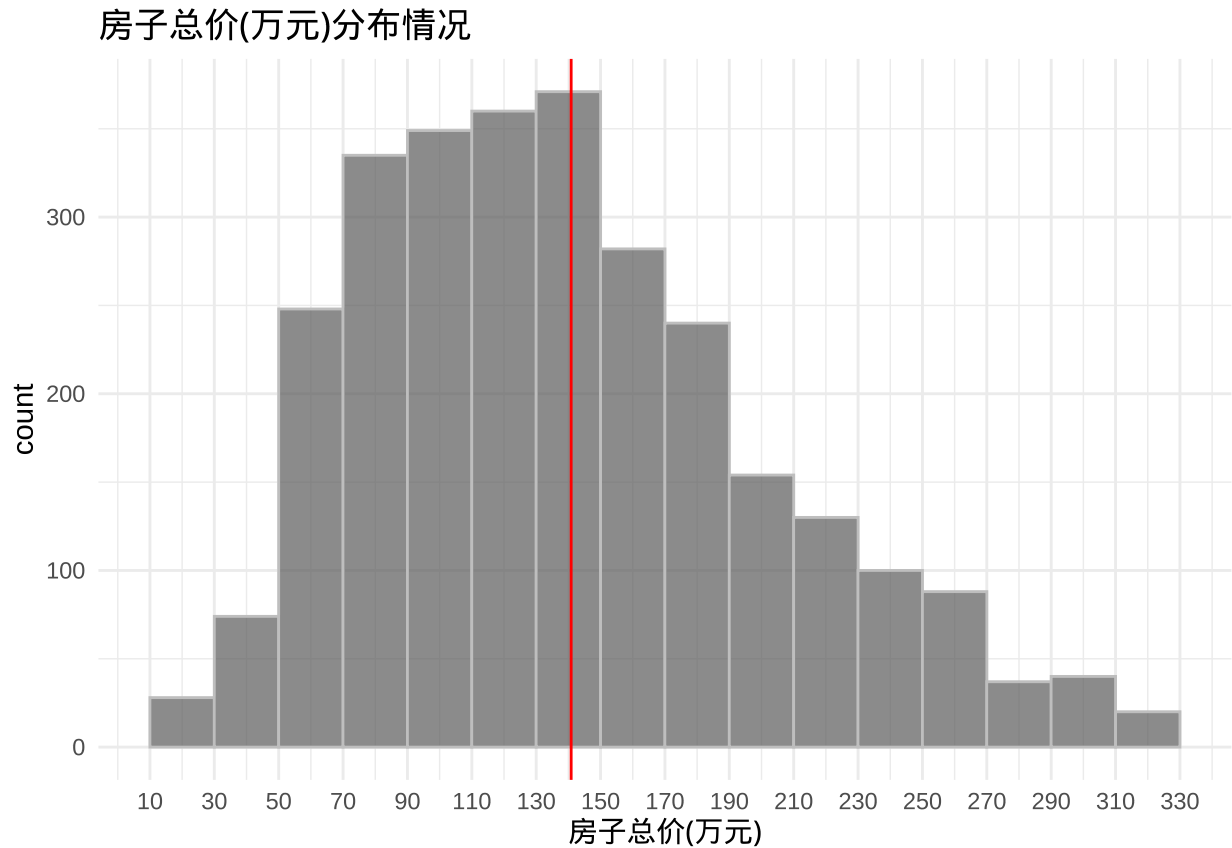
房子总价的数值描述与图形

房子总价统计指标:

```
#>   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>   10.6   95.0   137.0   155.9   188.0   1380.0
```

使用 IQR 的方法，剔除极值后，房价的分布情况如下图所示：

房子总价分布图：



剔除异常值后发现：

- 发现 1：房子总价的平均值在 140 万元，在 70 万元到 170 万元之间的房子最多；
- 发现 2：房子总价的分布呈现右偏，长尾；

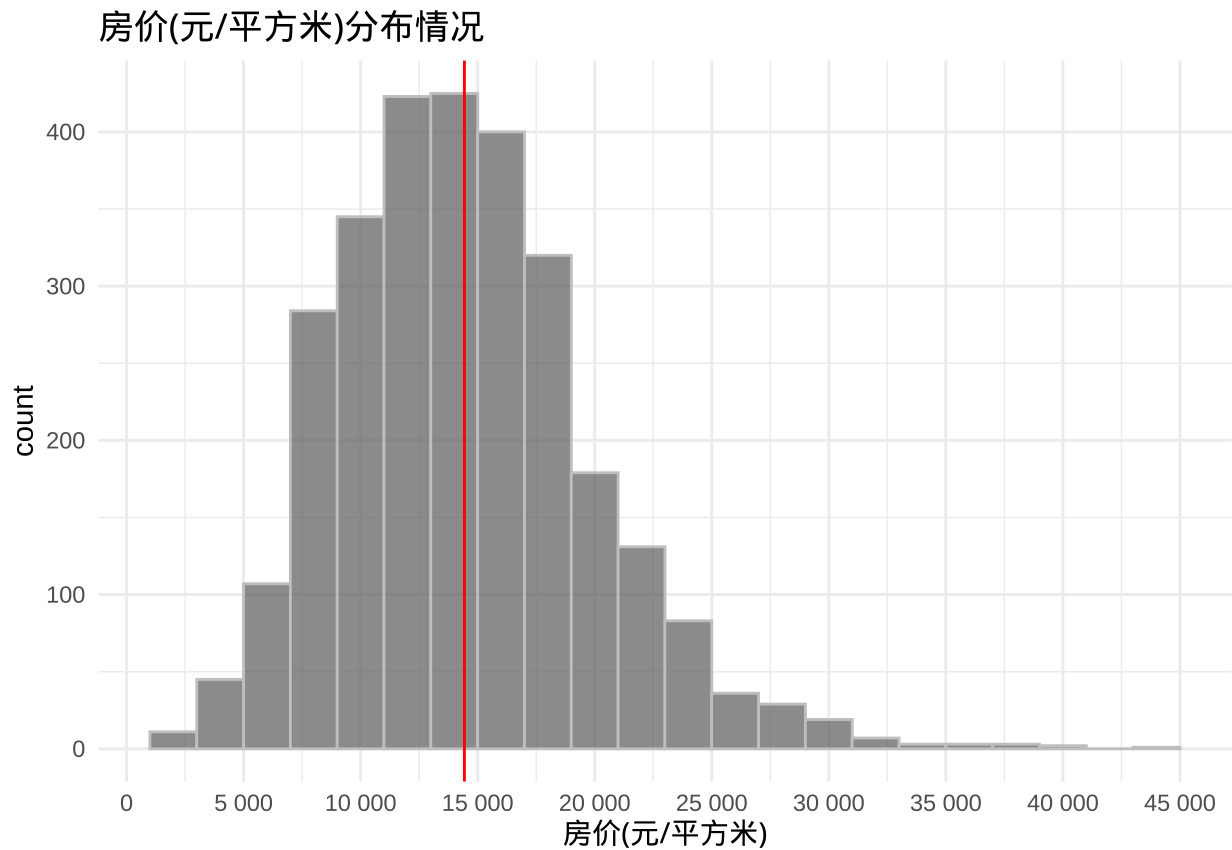
房价的数值描述与图形

房价（元/平方米）统计指标：

```
#>   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#>  1771  10564  14022  14433  17650  44656
```

可以看到房价的差异还是比较大的，下面通过直方图可以看看房价的分布情况。

房价（元/平方米）分布直方图：

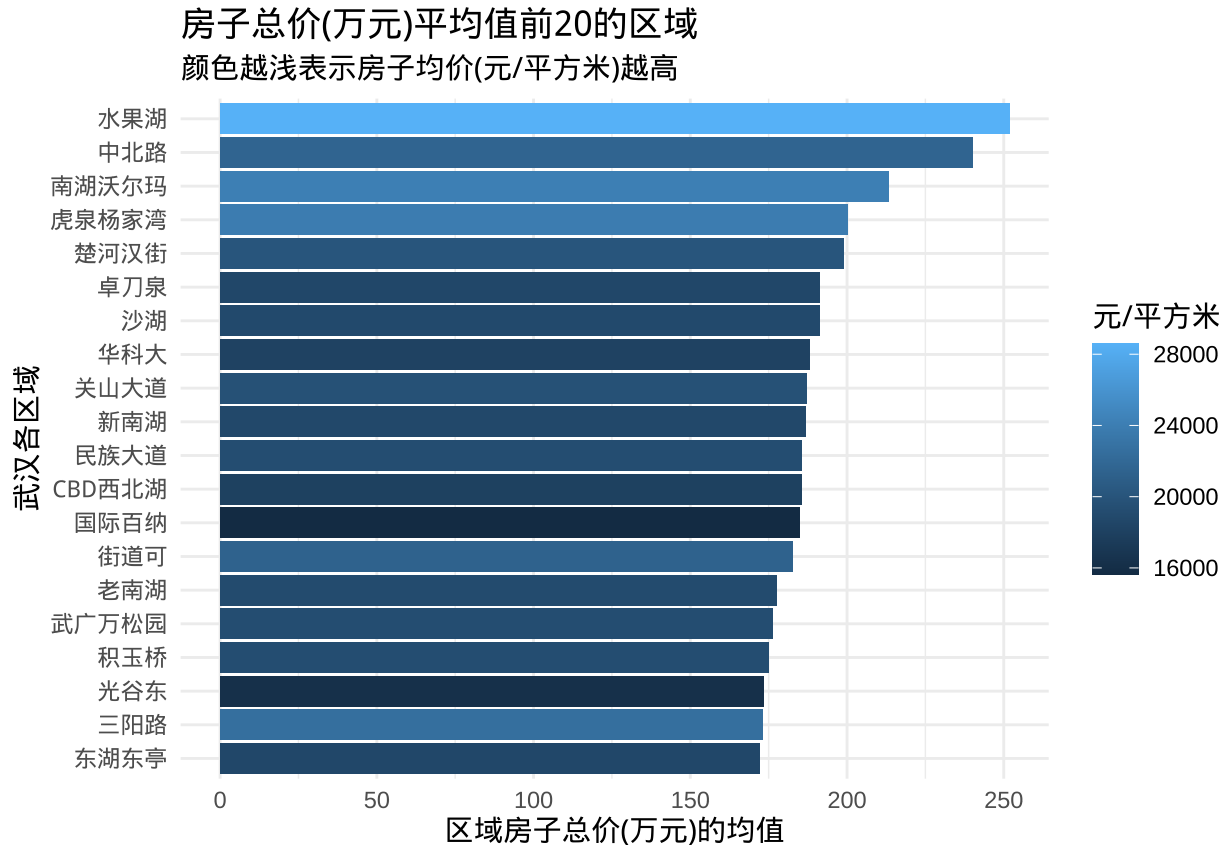


从图中可以发现：

- 1. 房价均价在 1.4 万元每平方米，大部分集中在 1 万到 1.8 万之间
- 2. 房价（元/平方米）的分布接近正态分布，略右偏，呈现长尾趋势，即存在相对较少的豪宅出售。

探索房价与区域的关系

房子位置对房价的影响很大，特别是繁华地带的，下面将分析武汉不同的区域的房价差异。



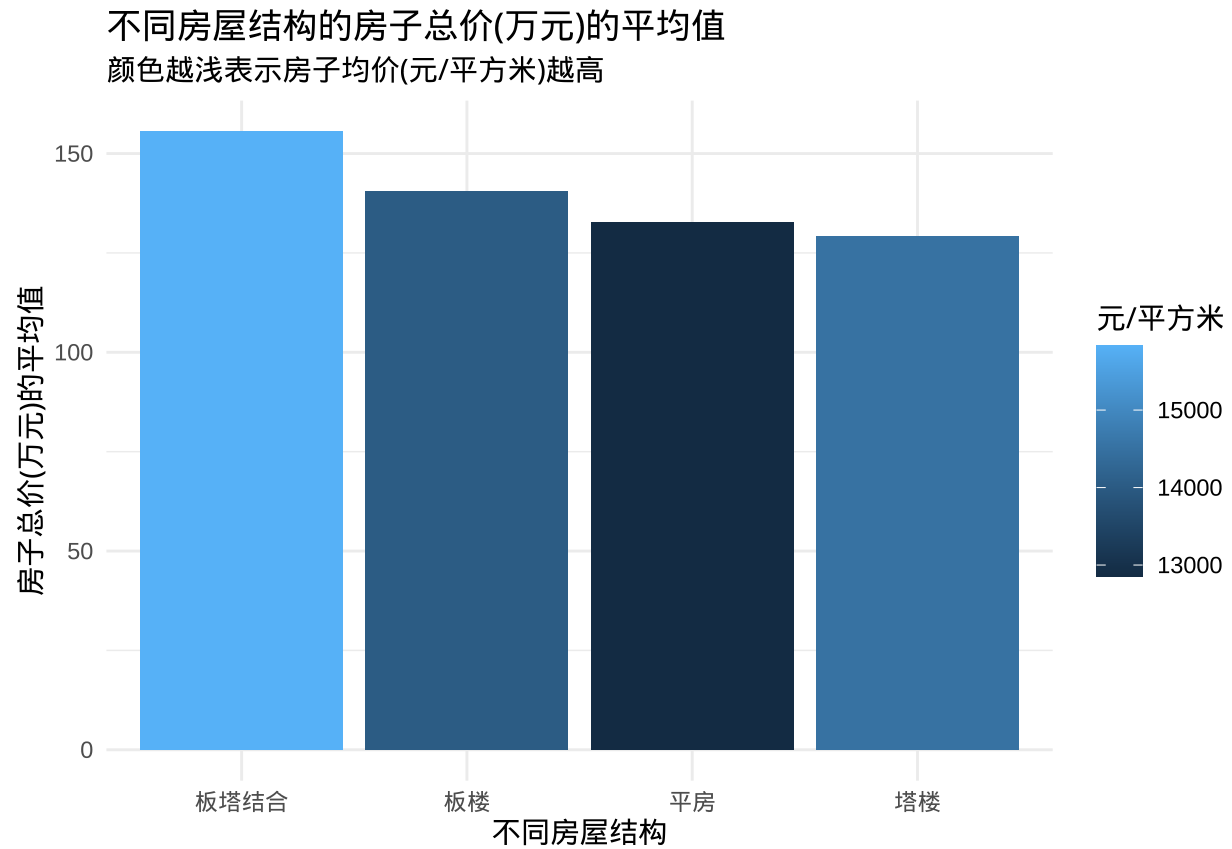
发现：

- 水果湖区域房子总价高，并且均价也高；中北路房子总价高，但是均价相对较低，说明该区域二手房的面积相对较大。
- 街道可（口）和三阳路的两个区域的房子均价要相对较高，相同总价情况下，房子的建筑面积可能要小一些

房子类型与房价的关系

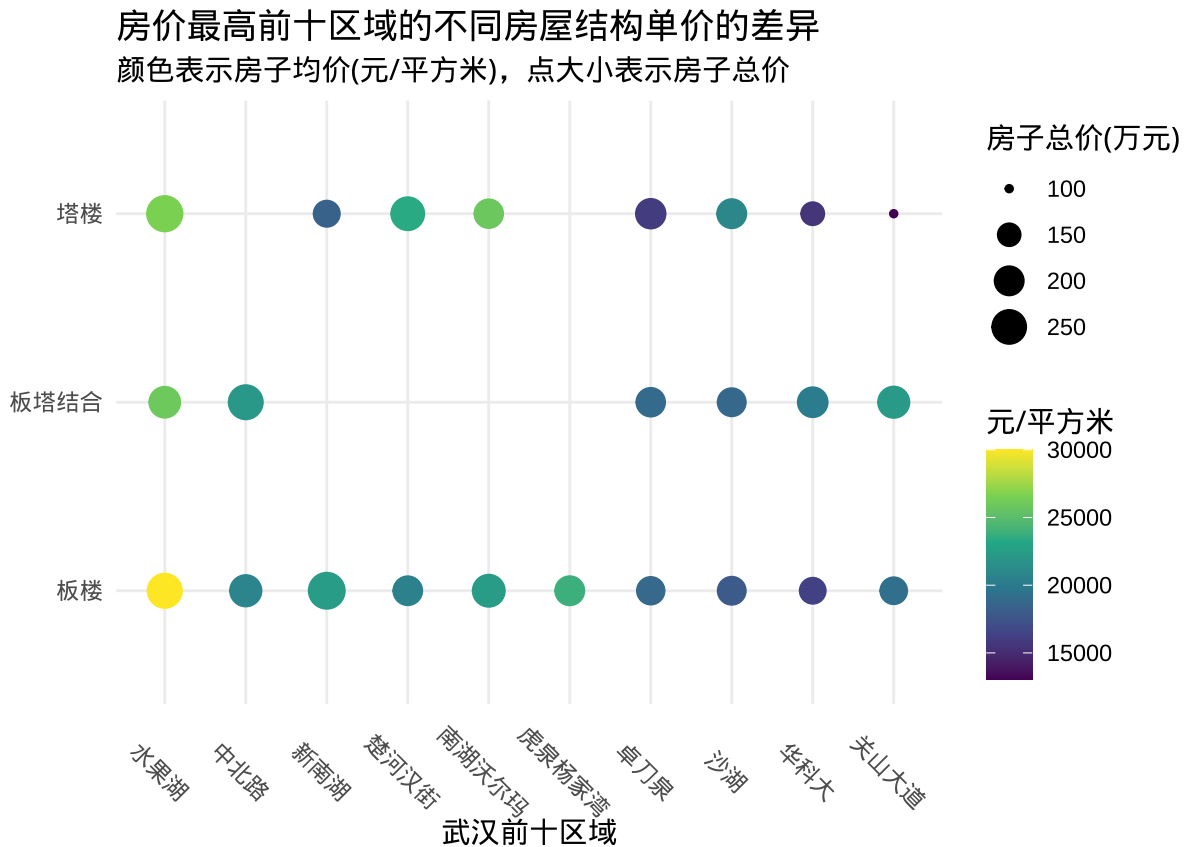
房屋的结构（板楼、塔楼）与是否精装修势必会影响到房价，下面分析一下这两个因素与房价的关系。

房屋结构与房价关系：



下面，进一步探索房子总价最高的前是个区域的不同房屋结构的在价格上的差异：

前十个地区为：水果湖，中北路，南湖沃尔玛，虎泉杨家湾，楚河汉街，卓刀泉，沙湖，华科大，关山大道，新南湖。



发现：

- 1. 水果湖区域，板楼的单价非常高，超过 3 万元每平方米；
- 2. 新南湖区域，板楼总价与单价均高于塔楼；
- 3. 板塔结合的房屋结构，在华科大与关山大道区域总价及单价均高于其他类型的房屋结构

详细看一下关山大道的房子情况：

可以看出关山大道板塔结合的房子，所在的楼栋总楼层较高，单价远高于平均值 1.4W。这与该地域繁华程度较高，土地资源稀缺的事实相呼应。

交通便利度与房价的关系

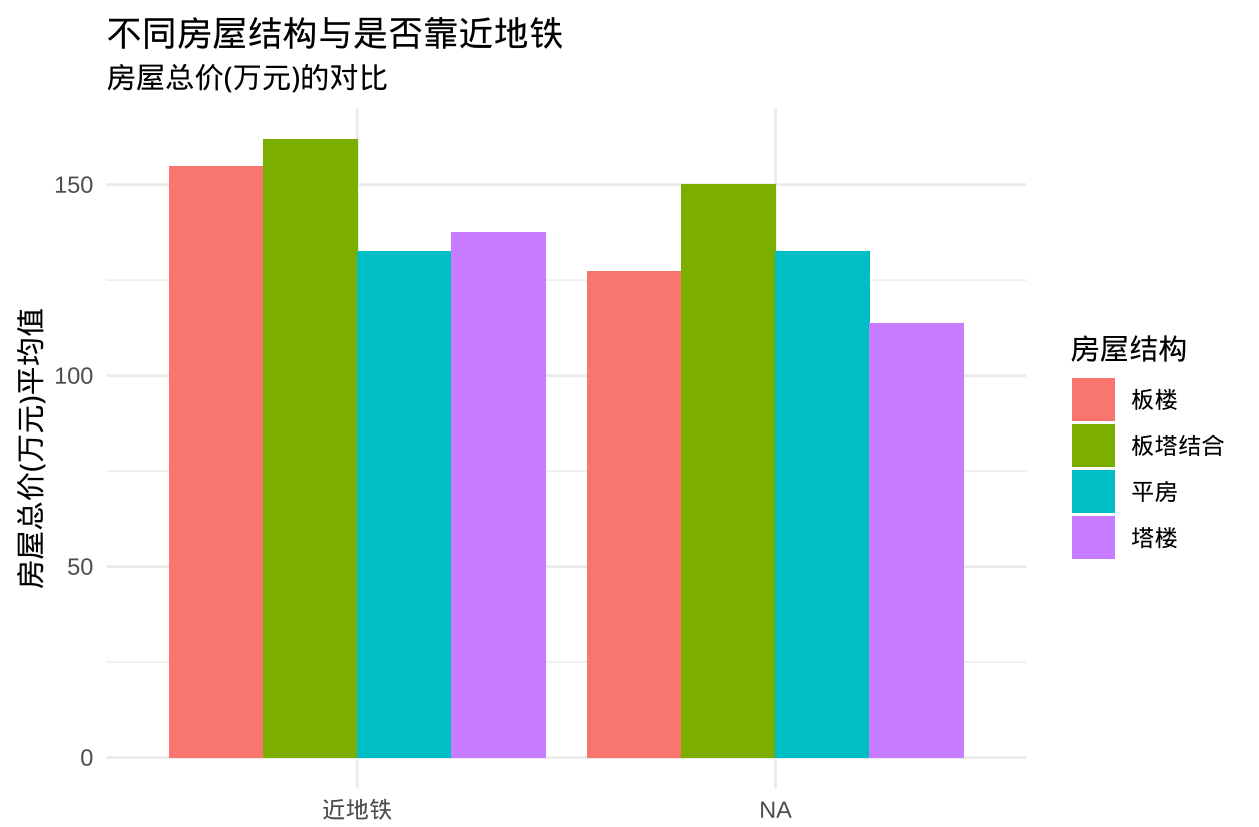
一般认为离地铁近的房子，交通便利度高，房价会比没有地铁的高，下面基于这份数据探索一下。

先看一下离地铁的房子情况：

near_subway	avg_price_ttl	avg_price_sqm	n
近地铁	151.2154	15723.66	1454
NA	130.1230	13094.43	1402

有近一半的房子靠近地铁（去掉了房价过高数据），近地铁房子总价及均价都要高于非近地铁的房子。

下面结合房屋类型看一下



发现：是否靠近地铁，只有对于平房没有影响，仔细看一下平房数据：

property_name	property_region	price_ttl	price_sqm	building_area	decoration	property_type	height	property_style	near_subway
万科金色城市	白沙洲	125	13032	95.92	精装	34	平房	NA	
23 地块									
天汇龙城龙华苑	盘龙城	140	11361	123.23	精装	33	平房	NA	
洪山教委宿舍	中南丁字桥	135	12577	107.34	简装	6	平房	近地铁	
洪山教委宿舍	中南丁字桥	135	12577	107.34	简装	6	平房	近地铁	

property_name	property_region	price_sqm	price_total	building_area	decoration	property_type	height	property_style	near_subway
长安丽都	CBD 西北湖	128	14752	86.77	简装	18	平房		近地铁

发现数据可能存在填写问题，平房应该没有 33 高，6 层高的教委宿舍是有可能，所以该类型的房子没有分析的必要性。

总体上看，靠近地铁的房子，不论是总价还是单价均高于不靠近地铁的房子。

发现总结

1. 武汉二手房总价的平均值在 140 万元，在 70 万元到 170 万元之间的房子最多；房价均价在 1.4 万元每平方米，大部分集中在 1 万到 1.8 万之间。相比其他一线新一线城市，武汉整体的房价水平是偏低的，普通人努努力还是可以够得着的。
2. 武汉不同区域差异房价明显，水果湖区域房子总价和每平方的均价都很高，尤其是板楼价格很坚挺；华科大与关山大道板块的板塔结合的楼挂牌较多，而且总价与每平方的均价都很贵，不愧是“武汉宇宙中心”。自住需求的，买房还是尽量避开这些区域，溢价过高。当然如果是投资，还是需要在热门区域投资，成交量大，容易出手。
3. 买房靠近地铁，将来交易的时候，平均要贵约 2000 元/平方米，所以尽量购买能靠近地铁房子，保值一些。