

第一次作业-武汉二手房信息分析报告

Code

屈意伟

2024-10-31

- 1 你的主要发现
- 2 数据介绍
- 3 数据概览
- 4 探索性分析
 - 4.1 变量1的数值描述与图形
 - 4.2 变量2的数值描述与图形
 - 4.3 变量3的数值描述与图形
 - 4.4 探索问题1
 - 4.5 探索问题2
 - 4.6 探索问题3
- 5 发现总结

1 你的主要发现

发现1: 武汉房屋均价差别较大, 最高价为44656元吗, 最低价为1771元。但平均价格在1.5万元左右, 平均面积在100平米左右, 以三室两厅或两室一厅为主要房型。

发现2: 武汉二手房源层高主要集中在20-30层之间, 相对较高。但随着总层数增加, 单价相应增加。

发现3: 武汉市二手房装修类型以精修为主, 其次简装, 毛坯第三, 坐北朝南的房屋仍然为中国人最喜欢的房屋朝向。房间数对价格的影响程度较小, 但房间数越多的房子房价的上限更高。

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 (https://wh.lianjia.com/ershoufang/)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, 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:

武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2	decoration	pro
南湖名都A区	237.									
南湖沃尔										

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
##
##
##
```

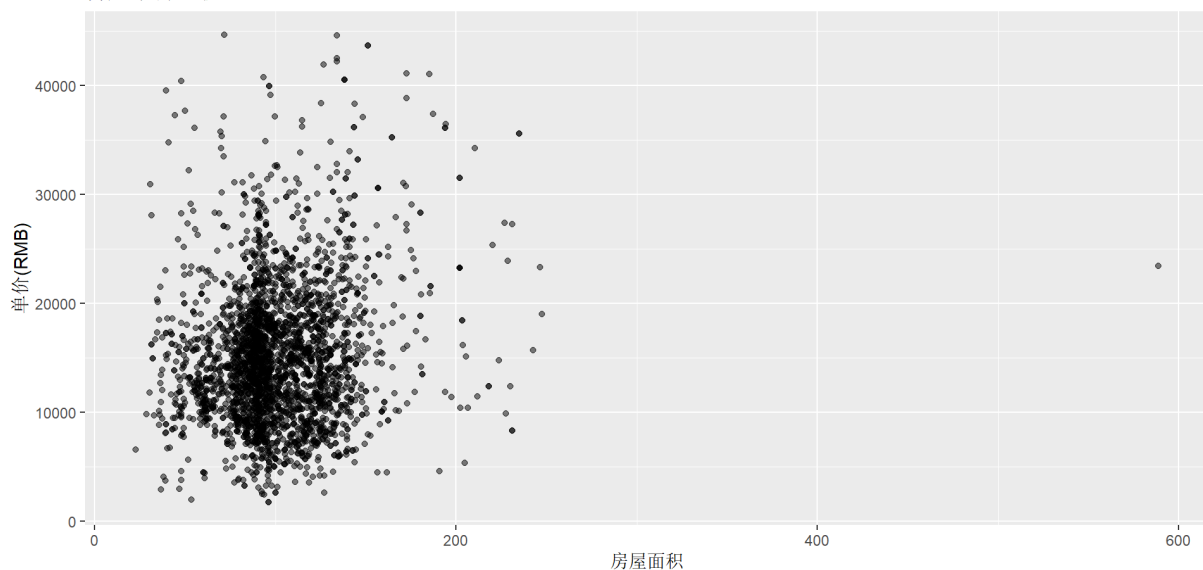
可以看到：

- 直观结论1：数据共包含3000行数据，18个变量
- 直观结论2：房源以1-2个客厅，3个卧室为主。房源平均面积为100平米左右。
- 直观结论3：房屋单价从1771元至44656元不等，均价为15148元。总价平均为155.9元。

4 探索性分析

4.1 变量1的数值描述与图形

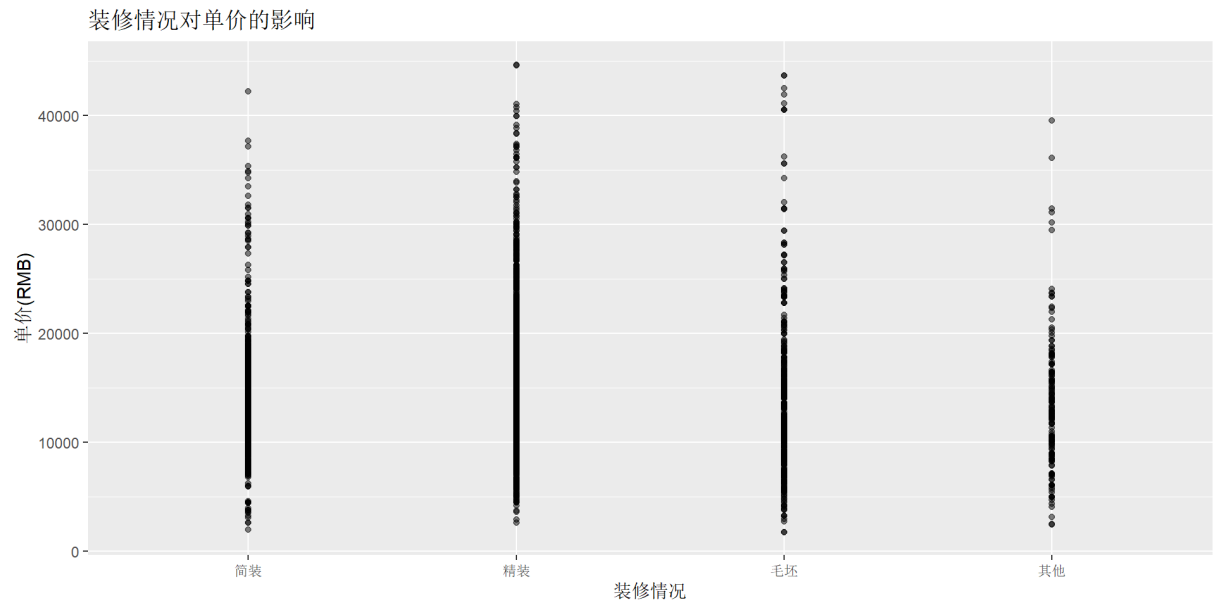
房屋单价面积对比表



发现：

- 发现1：武汉二手房市场价格主要维持在10000—20000元之间，房屋面积主要在200平米以下。
- 发现2：仅有一处小区面积在600平米左右，该小区均价在25000元左右。

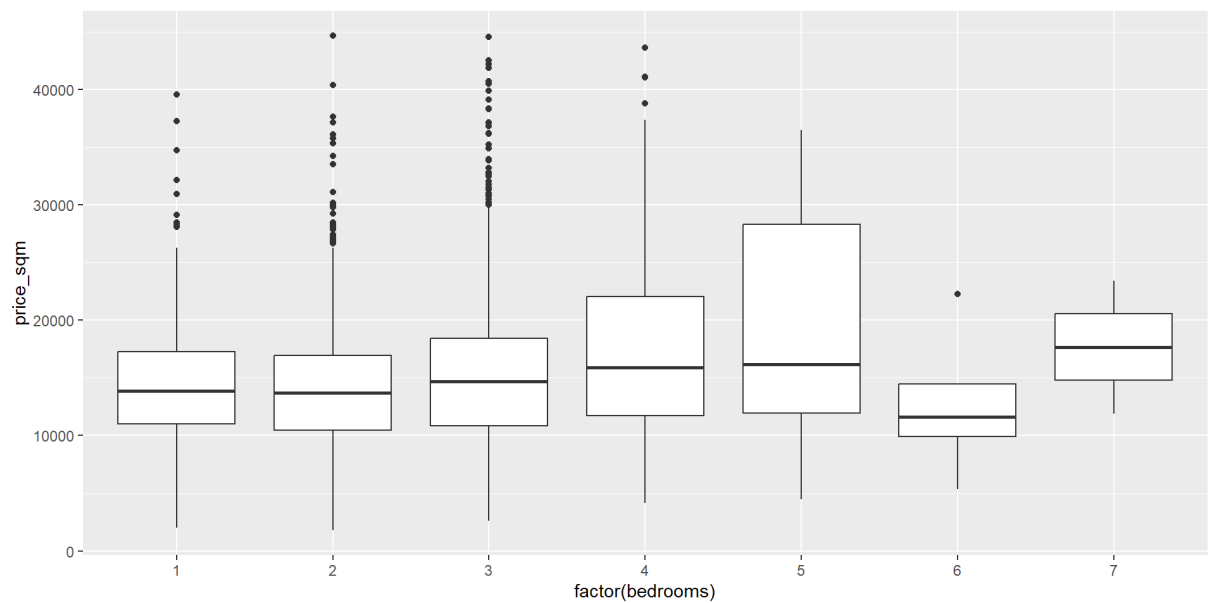
4.2 变量2的数值描述与图形



发现:

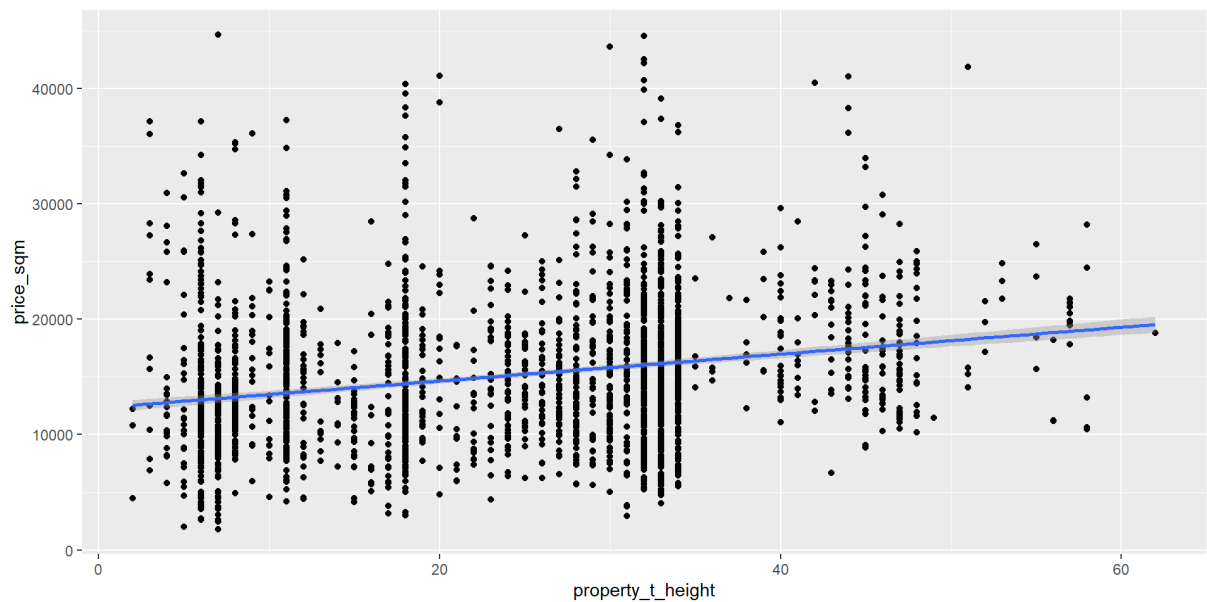
- 发现1: 装修情况越好, 单价相对越高。毛坯均价显著低于精装房。

4.3 变量3的数值描述与图形



发现: 房间数对价格的影响程度较小。但是一定程度上房间数越多, 房价的分散程度越大。表明房间数越多的房子房价的上限更高。

4.4 探索问题1



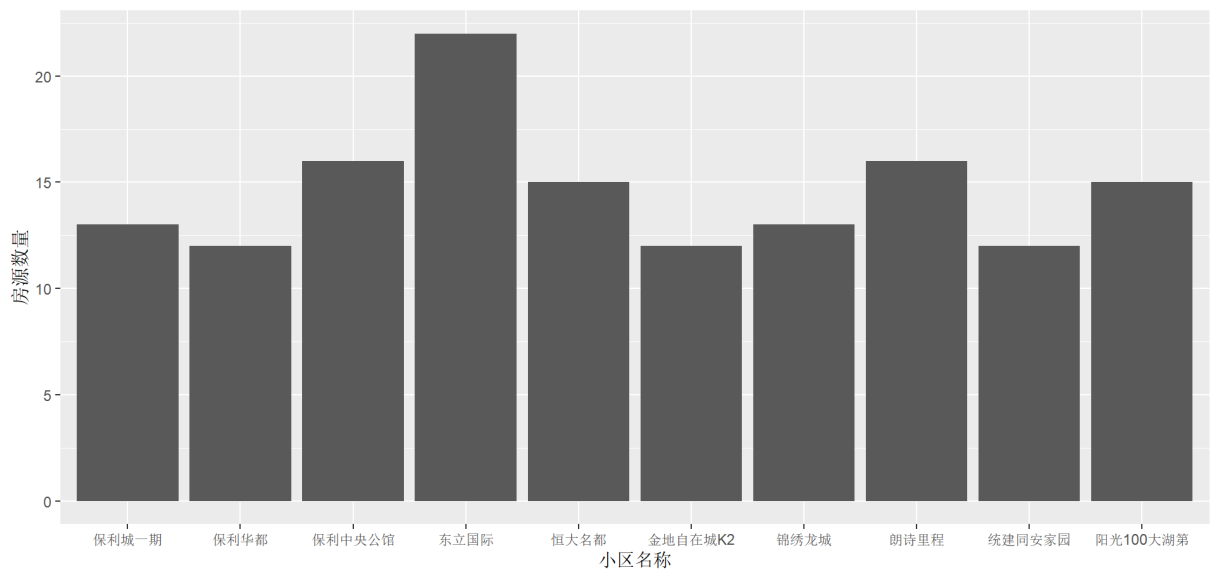
发现：

- 发现1：小区的总层数主要维护在30层级及以下。
- 发现2：小区单价与总层数呈正相关。小区楼层越高，单价相应升高。

4.5 探索问题2

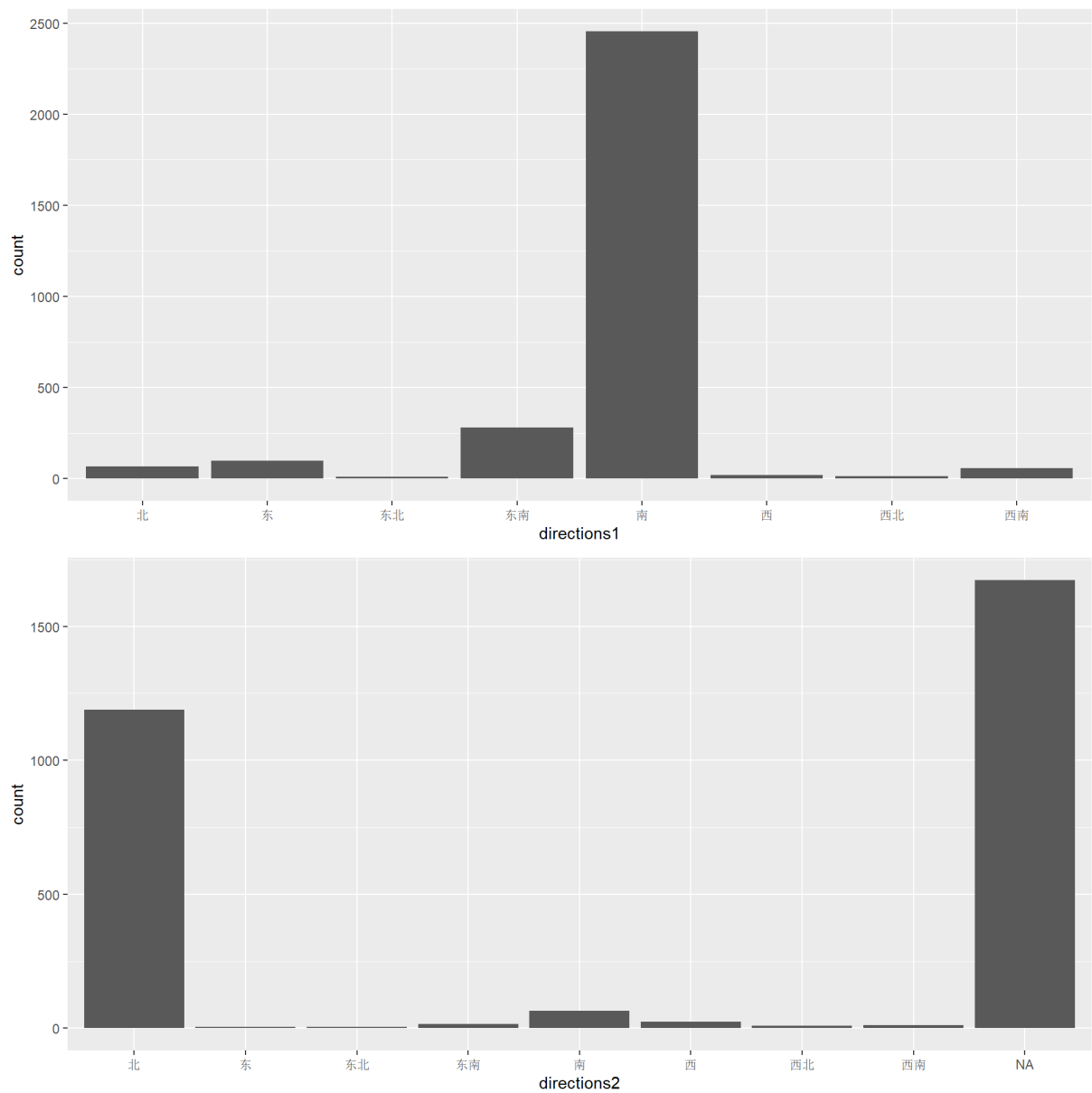
```
## # A tibble: 10 × 2
##   property_name count
##   <chr>         <int>
## 1 东立国际         22
## 2 保利中央公馆    16
## 3 朗诗里程         16
## 4 恒大名都         15
## 5 阳光100大湖第    15
## 6 保利城一期       13
## 7 锦绣龙城         13
## 8 保利华都         12
## 9 统建同安家园     12
## 10 金地自在城K2    12
```

房源最多的小区



发现：东立国际、保利中央公馆、朗诗里程为武汉房源最多的三大小区。

4.6 探索问题3



发现：武汉二手房源主要以坐北朝南为主，表明该朝向房屋容易接收阳光，更为中国人所喜爱。

5 发现总结

发现1：武汉房屋均价差别较大，最高价为44656元吗，最低价为1771元。但平均价格在1.5万元左右，平均面积在100平米左右，以三室两厅或两室一厅为主要房型。

发现2：武汉二手房源层高主要集中在20-30层之间，相对较高。但随着总层数增加，单价相应增加。

发现3：武汉市二手房装修类型以精修为主，其次简装，毛坯第三，坐北朝南的房屋仍然为中国人最喜欢的房屋朝向。房间数对价格的影响程度较小，但房间数越多的房子房价的上限更高。