

商务统计分析 1st_assignment

鄢圣东 2023281051003

2023-10-19

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。

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：

```
## # A tibble: 10 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 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>
```

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", "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:

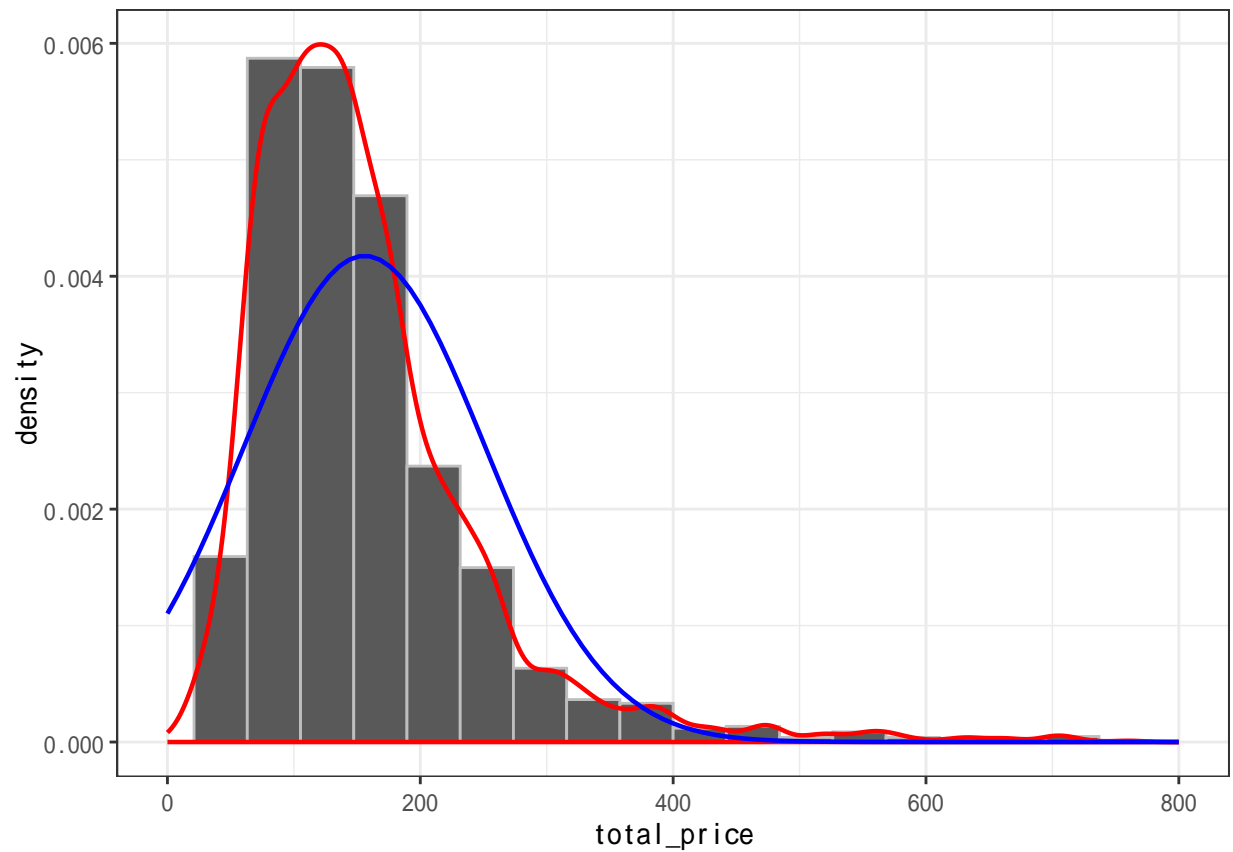
- 直观结论 1

数据中有 3000 套住房信息。价格特征: 1、房屋总价平均价格为 155.9 万/套, 价格主要分布在 95.0~188.0 万/套, 最贵一套为 1380.0 万, 最便宜一套为 10.6 万 2、房屋每平米价格均价为 14404 元/平方米, 价格主要分布在 10799~18211 元/平方米, 最贵一套为 44656 元/平方米, 最便宜一套 1771 元/平方米 - 直观结论 2 2 房型特征: 1、3000 套房屋中卧室平均个数约为 2.7 个, 住房中卧室个数主要分布在 2~3 个, 最小卧室数量为 1 个, 最大卧室数量为 7 个。2、3000 套房屋中卫生间平均个数为 1.7 个, 住房中卫生间个数主要分布在 1~2 个, 最小卫生间数量为 1 个, 最大数量/为 4 个。3、3000 套房屋平均面积为 100.87 平方米, 面积大小主要分布在 84.92~117.68 平方米, 最小面积为 22.77, 最大面积为 588.66 平方米 - 直观结论 3 3 楼型特征: 1、3000 套房屋所在楼层平均数约在 24.22 层, 所在楼层主要分布 11~33 层之间, 最小楼层为 2, 最大楼层为 62 层 2、3000 套房屋所在楼层电梯数量平均数为 6.6 个, 电梯数量主要分布在 1~6 个之间, 有的住宅没有电梯, 最大电梯数量为 262 个。

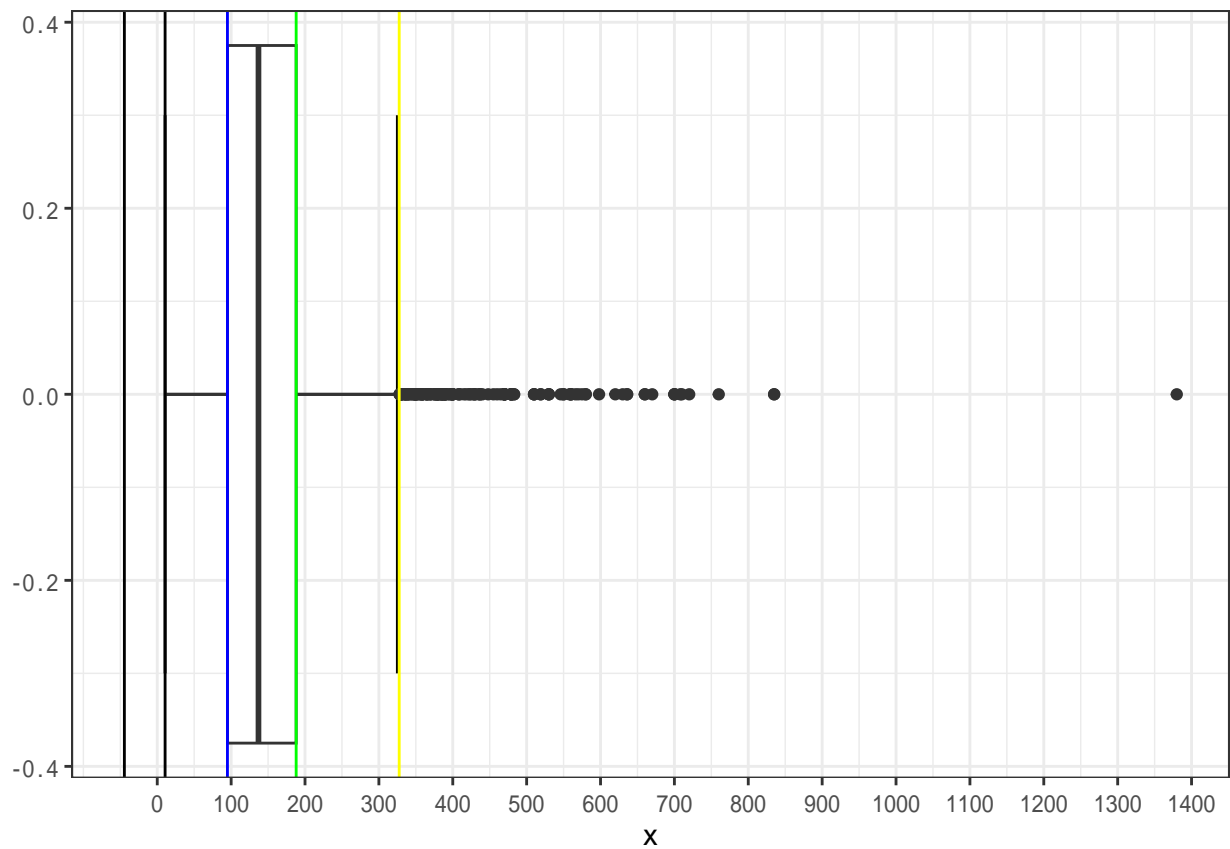
探索性分析

变量 price_ttl 的数值描述与图形

price_ttl 的数据特征 Min. 1st Qu. Median Mean 3rd Qu. Max. 10.6 95.0 137.0 155.9 188.0 1380.0 -
直方图发现：根据直方图图形显示数据分布情况，概率密度曲线基本满足正态分布趋势，曲线更高更窄，数据更加集中，并且向右偏移。

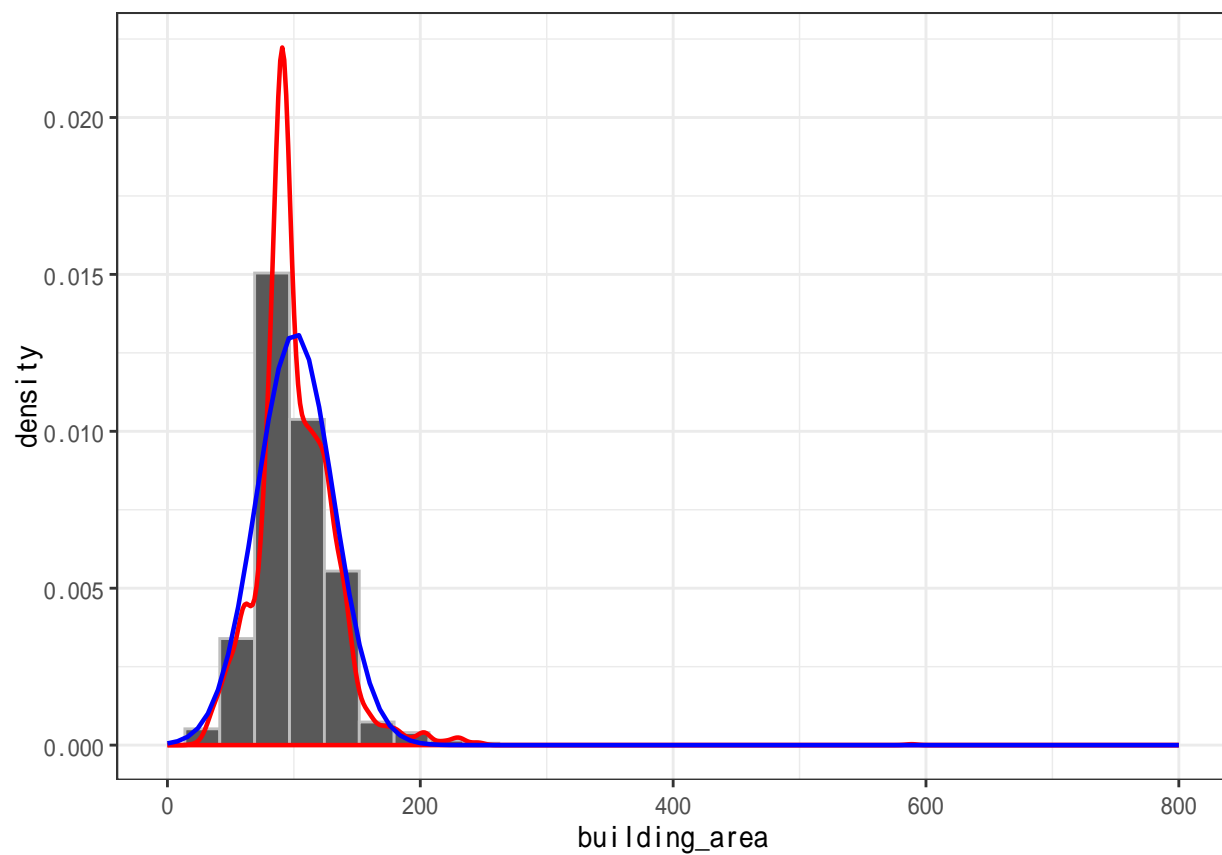


- 箱型图发现：箱线图展示了变量的 4 分位以及上下分界，价格集中区域。

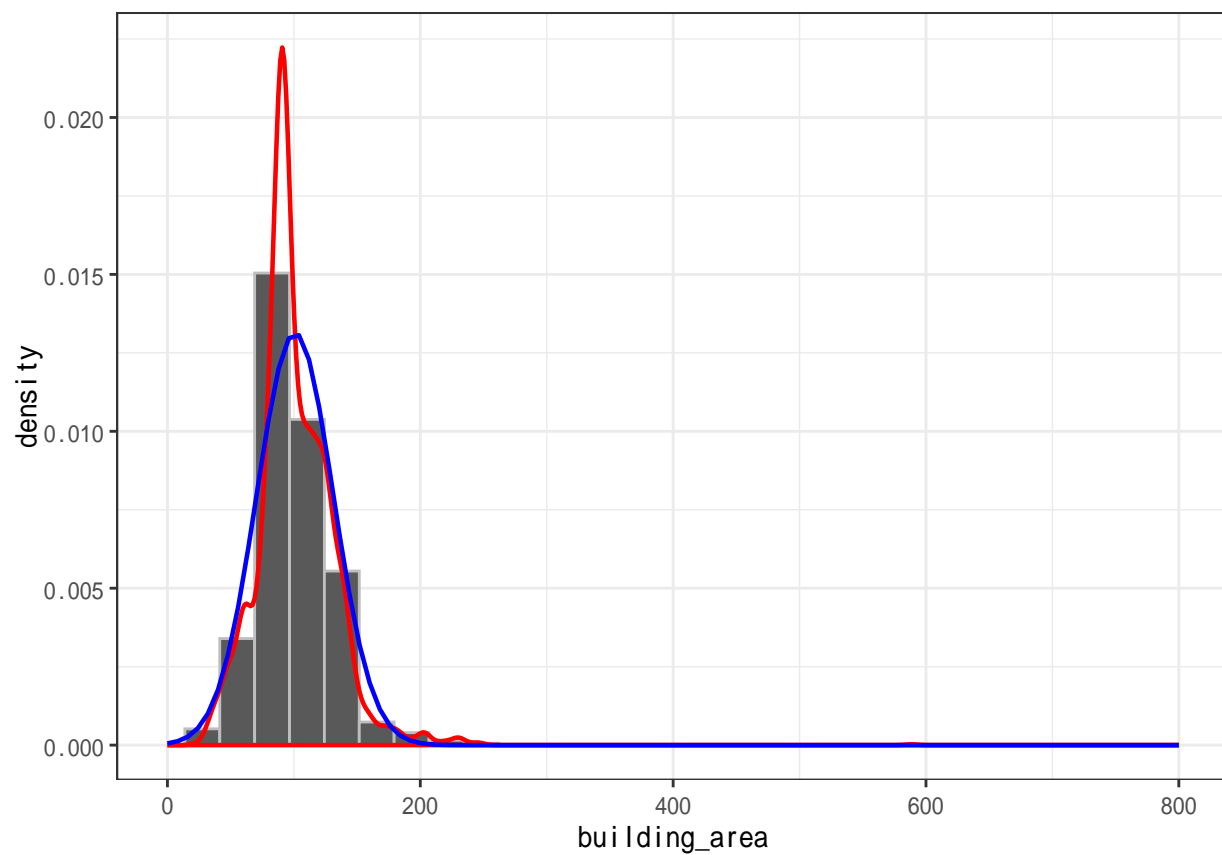


变量 `building_area` 的数值描述与图形

building_area 数值特征 Min. 1st Qu. Median Mean 3rd Qu. Max. 22.77 84.92 95.55 100.87 117.68 588.66 - 箱型图发现：住房面积主要集中与 80 万元与 120 万元之间, 其中集中区域非常集中。

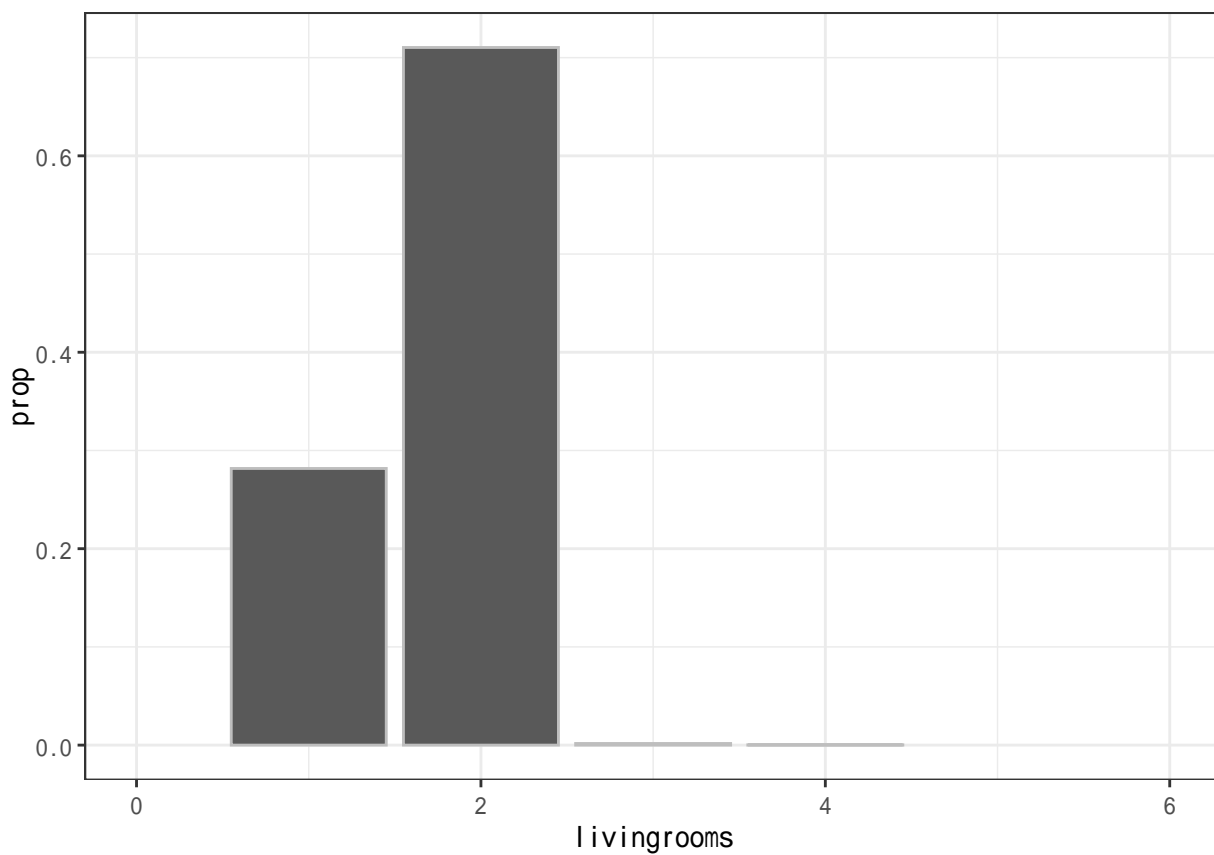


- 直方图发现：图形中密度曲线满足正态分布趋势



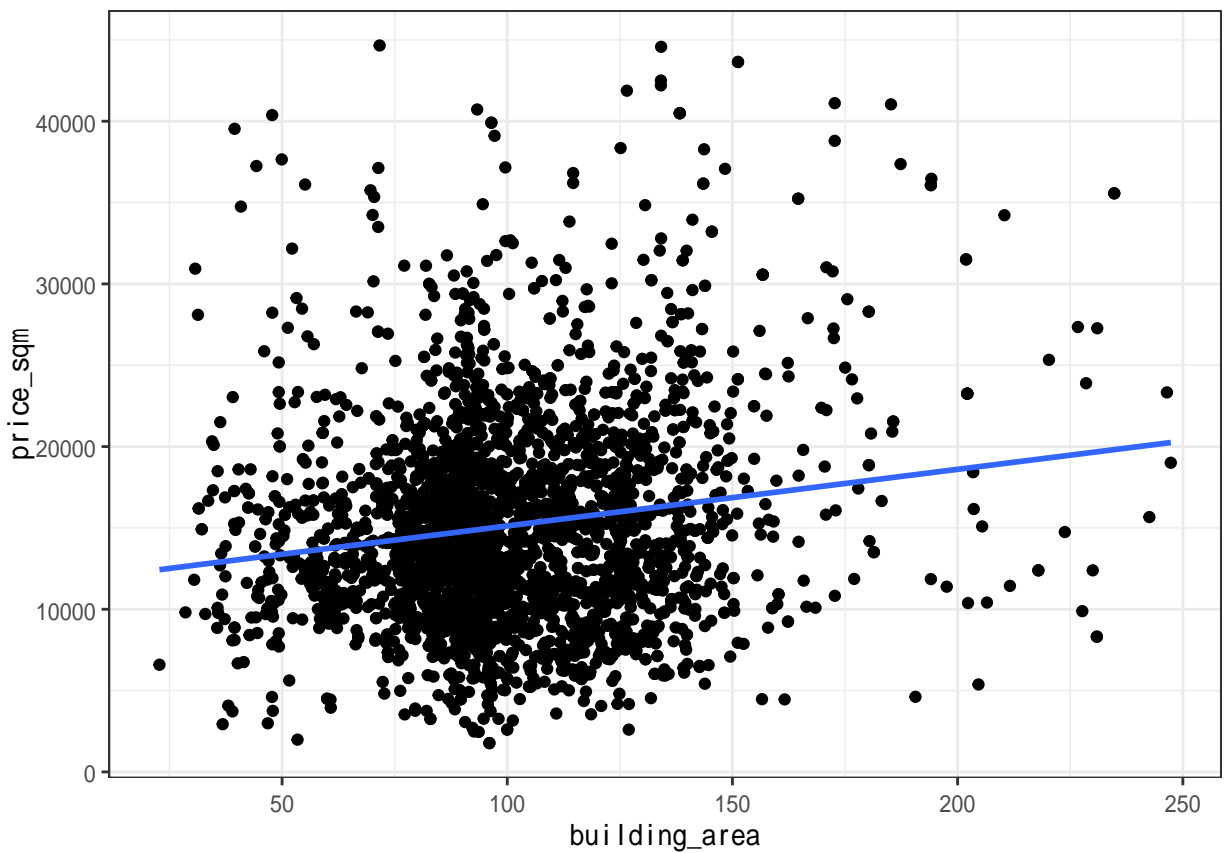
变量 livingrooms 的数值描述与图形

变量 livingrooms 数值特征 summary(lj\$livingrooms) Min. 1st Qu. Median Mean 3rd Qu. Max. 0.000 1.000 2.000 1.709 2.000 4.000 — 条形图发现：房屋中卧室数量为 2 的房屋数量最多，超过样本数量的 70%。



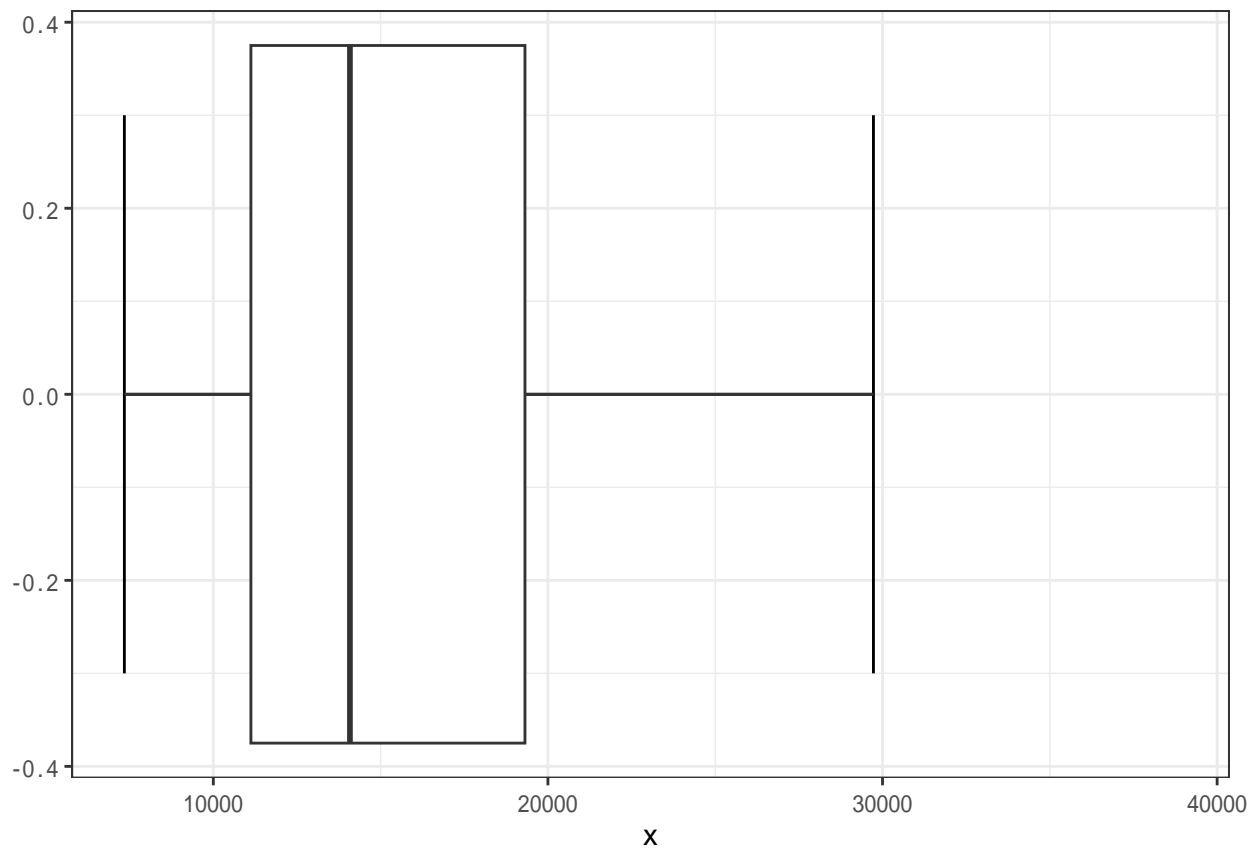
住房单价与住房面积存在什么关系

发现一：住房面积越大，房屋的单价就越高，具有正相关性。所以所购买的房子越大，其单价越贵，其总价也就越高，对于购房者也许要考虑房子面积大小所带来的房子单价的上涨趋势。



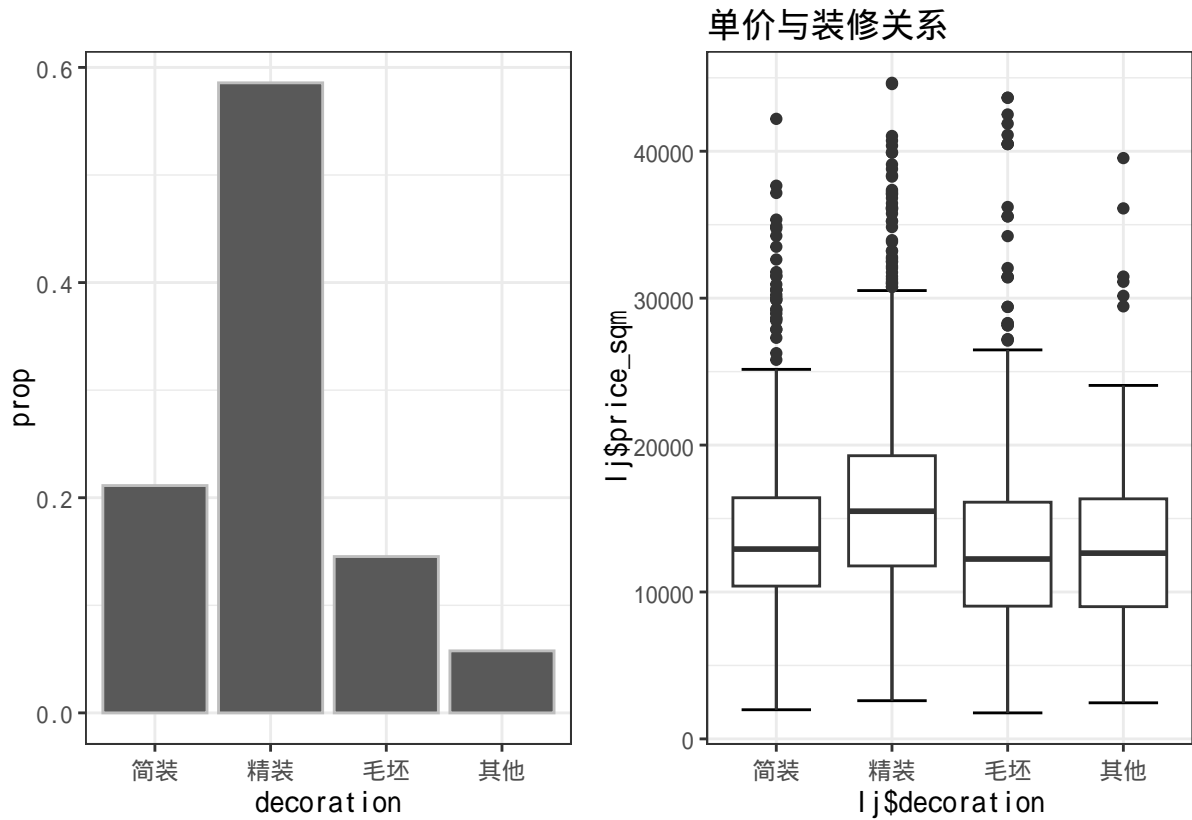
在武汉人们主要考虑的购房单价是多少

发现：1. 看房人数前 50 的房子的单价基本集中于 1.1 万元到 2 万元之间，这个区间是人们的刚性需求。



装修风格对住房价格的影响

发现：1、二手房市场绝大多数是精装房，这表明大多数出售的房屋为自住房，而不是用来炒作的出租房或者炒房。2、简装房和毛坯房单价集中区域基本相同，说明这部分房产具有共性，而精装房价格明显高于其他。



发现总结

通过对数据的分析,我发现了如下几点:1、二手房市场房源的总价格区间主要集中在 80₁₃₀ 万元区, 其中热门房源的单价区间在 1.1² 万元之间, 说明武汉当地第一套房的刚需住房条件, 后续为改善生活出售第一套房产情况最多。2、房屋价值的影响: 房子面积越大, 房子单价会被抬高。精装相比与其他装修方式, 房子单价也会高于其他。3、通过市场的装修风格的分析, 大部分的二手房源为自住房, 因为绝大多数是自己居住, 才会在装修投入更多。