

# 第一次作业你的报告题目

Code ▾

舒明

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 房屋名称 (property\_name) 的数值描述与图形
  - 4.2 区域 (property\_region) 的数值描述与图形
  - 4.3 房屋总价 (price\_ttl) 的数值描述与图形
  - 4.4 单价 (price\_sqm) 的数值描述与图形
  - 4.5 建筑面积 (building\_area) 的数值描述与图形
  - 4.6 关注人数 (followers) 的数值描述与图形
  - 4.7 楼栋总层数 (property\_t\_height) 的数值描述与图形
  - 4.8 房间数 (bedrooms) 的数值描述与图形
  - 4.9 客厅数 (livingrooms) 的数值描述与图形
  - 4.10 房屋主要朝向 (directions1) 的数值描述与图形
  - 4.11 房屋次要朝向 (directions2) 的数值描述与图形
  - 4.12 装修状况 (decoration) 的数值描述与图形
  - 4.13 房屋在所在楼栋所处位置 (property\_height) 的数值描述与图形
  - 4.14 建筑形式 (property\_style) 的数值描述与图形
  - 4.15 是否靠近地铁 (near\_subway) 的数值描述与图形
  - 4.16 产证是否满2年 (if\_2y) 的数值描述与图形
  - 4.17 中介是否有钥匙 (has\_key) 的数值描述与图形
  - 4.18 是否支持VR看房 (vr) 的数值描述与图形
  - 4.19 探索问题1: 房屋单价、装修状况和房屋在所在楼栋所处位置的关系
  - 4.20 探索问题2: 房屋单价、房间数和建筑形式的关系
  - 4.21 探索问题3: 房屋单价、关注人数和建筑形式的关系
- 5 发现总结

## 1 你的主要发现

- **房屋单价、装修状况和房屋在所在楼栋所处位置的关系：**
  - 装修状况对房价有显著影响，精装修房屋的单价通常高于简装房屋。
  - 楼层位置对房价也有影响，高层房屋单价通常高于低层房屋。
  - 装修状况和楼层位置的组合可能对房价有更大的影响。
- **房屋单价、房间数和建筑形式的关系：**
  - 建筑形式对单价有影响，板楼和板塔结合的单价相对稳定，而塔楼在大房间数时单价较低。
  - 房间数对单价有一定影响，平房的单价在不同房间数下变化不大。
- **房屋单价、关注人数和建筑形式的关系：**
  - 建筑形式对关注人数有影响，板楼明显获得更多关注。
  - 单价与关注人数之间存在关联，但不是唯一决定性因素，单价适中的房屋可能获得更多关注。

## 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(1)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	prop
---------------	-----------------	-----------	-----------	----------	-------------	---------------	-------------	-------------	------------	------

property\_name property\_region price\_ttl price\_sqm bedrooms livingrooms building\_area directions1 directions2 decoration prop

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", "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：

- 房屋总价 (price\_ttl) 的分布范围广泛，从10.6万元到1380万元不等。大多数房源的价格集中在95万到188万元之间，中位数为137万元，均值为155.9万元。这表明市场上多数房源的价格处于中等水平，但存在一些高价房源，使得均值略高于中位数。
- 单价 (price\_sqm) 的分布也较为广泛，从1771元/平方米到44656元/平方米。大多数房源的单价集中在10799元到18211元/平方米之间，中位数为14404元/平方米，均值为15148元/平方米。这表明市场上多数房源的单价处于中等水平，但同样存在一些高单价房源，使得均值略高于中位数。
- 房间数 (bedrooms) 主要集中在2到3个房间，中位数为3，均值为2.695。最小值为1，最大值为7，表明市场上既有单间也有较大的多居室户型，但主流是2室和3室户型。

- 客厅数 (livingrooms) 主要集中在1到2个，中位数为2，均值为1.709。最小值为0，最大值为4，表明市场上多数房源有1到2个客厅，但也存在一些没有客厅或拥有多个客厅的房源。
- 建筑面积 (building\_area) 主要集中在84.92平方米到117.68平方米，中位数为95.55，均值为100.87。最小值为22.77，最大值为588.66，表明市场上多数房源的建筑面积处于中等水平，但也存在一些大面积房源，使得均值略高于中位数。
- 楼栋总层数 (property\_t\_height) 的分布从2层到62层不等，中位数为27层，均值为24.22层。大多数楼栋的总层数在11到33层之间，表明市场上以中高层建筑为主，但也存在一些低层或多层建筑。
- 关注人数 (followers) 的分布范围从0到262人，中位数为3，均值为6.614。大多数房源的关注人数在1到6人之间，但少数房源受到大量关注，表明市场上部分房源非常受欢迎。
- 房屋名称 (property\_name)、区域 (property\_region)、主要朝向 (directions1)、次要朝向 (directions2)、装修状况 (decoration)、楼层高度 (property\_height)、楼栋风格 (property\_style)、是否靠近地铁 (near\_subway)、产证是否满两年 (if\_2y)、中介是否有钥匙 (has\_key)、是否支持VR看房 (vr) 这些变量都是字符型数据，没有数值统计信息，需要通过频数分布或词云图来进一步分析其具体分布情况。

## 4 探索性分析

### 4.1 房屋名称 (property\_name) 的数值描述与图形



发现：

- “花园”、“国际”、“城”、“二期”、“一期”、“三期”、“小区”、“时代”是房屋名称的高频词，可能反映了市场上主流房产项目的命名方式。其中“花园”一词最为高频，可能表明主流项目的大多数都会在名称中通过该词强调其绿化环境和居住的舒适度以增加卖点。
- 一些知名的房地产品牌如“保利”、“万科”、“金地”等在词云中占据显著位置，这可能反映了这些品牌在武汉二手房市场中的影响力和认可度。

### 4.2 区域 (property\_region) 的数值描述与图形

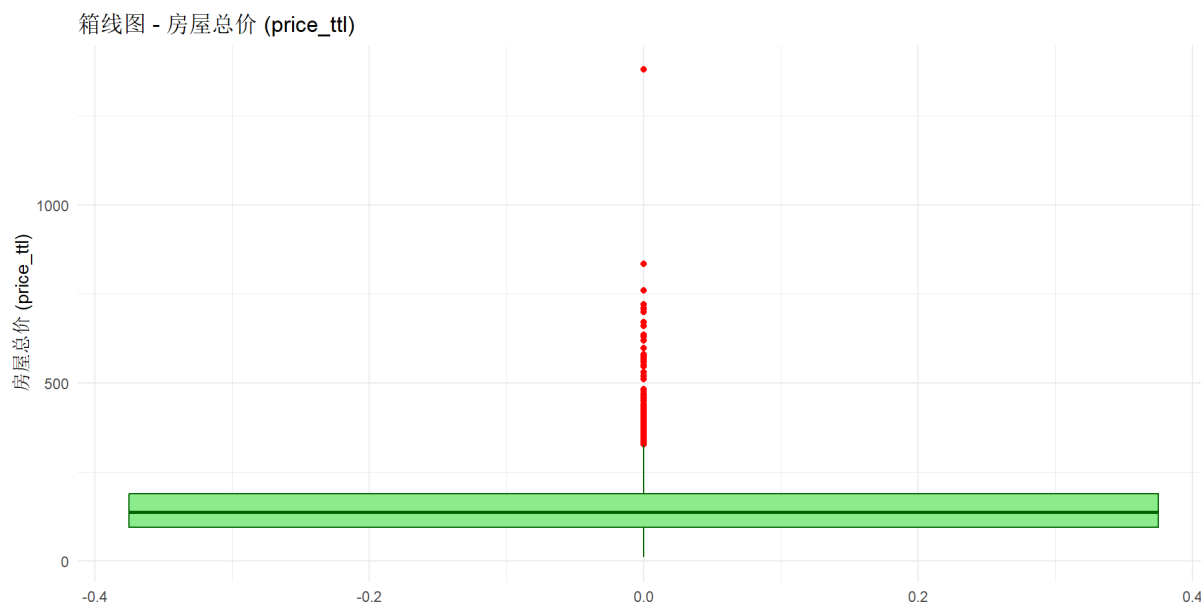
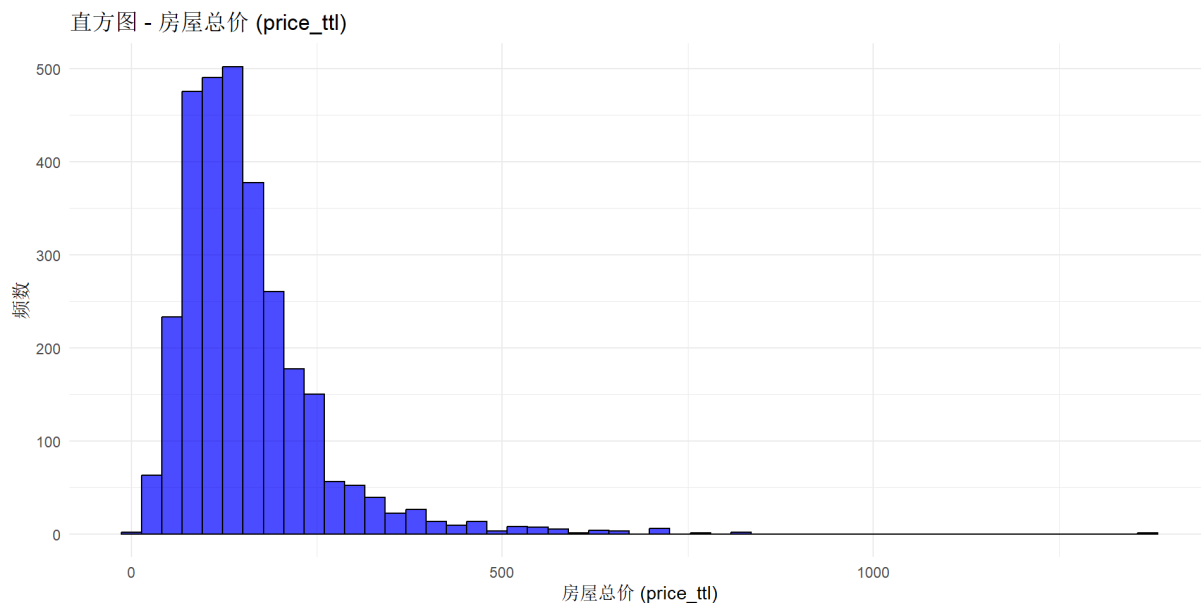


发现：

- “湖”、“大道”、“光谷”、“白沙洲”、“盘龙城”、“四新”是房屋名称的高频词，包含这些词的区域可能是热门的购房区域。其中“湖”一词最为高频，可能和武汉湖泊资源丰富、且常用于各类地理名称有关。
- 高频词基本和我所了解的年轻人购房需求区域重合。

## 4.3 房屋总价 (price\_ttl) 的数值描述与图形

```
##
## ----- 分析变量: 房屋总价 (price_ttl) -----
## 均值 (Mean): 155.8628
## 中位数 (Median): 137
## 众数 (Mode): 105
## 极差 (Range): 1369.4
## 四分位距 (IQR): 93
## 方差 (Variance): 9129.445
## 标准差 (Standard Deviation): 95.54813
## 偏度 (Skewness): 2.753223
## 峰度 (Kurtosis): 16.11672
```



发现:

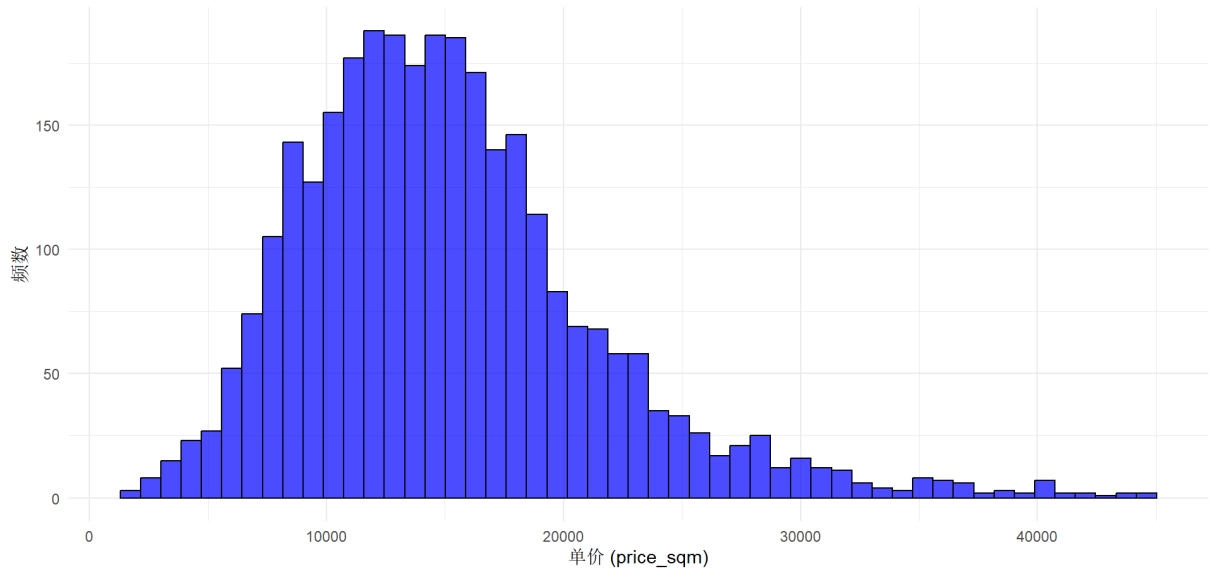
- 分布:
  - 直方图显示房屋总价的分布呈现右偏（正偏）分布，大多数房屋的总价集中在较低的区间，而随着价格的增加，房屋数量迅速减少。
  - 偏度 (Skewness) 为2.753223，这是一个非常高的正偏度值，表明数据分布有长尾向右延伸，存在一些极端的高值。
- 集中趋势:
  - 均值 (155.8628) 远高于中位数 (137)，这进一步证实了分布的右偏性，因为如果分布是对称的，均值和中位数应该是相等的。
  - 众数 (105) 表明最常见的房屋总价是105万，这可能反映了市场上小户型或低价房屋的普遍性。

- 离散程度和异常值：
  - 标准差 (95.54813) 和四分位距 (93) 表明房屋总价之间存在较大的差异，且数据分布较为分散。
  - 极差 (1369.4) 非常大，从最小值到最大值的范围很广，这表明市场上既有非常便宜的房屋，也有非常昂贵的房屋。
  - 箱线图显示存在一些异常值，这些点位于上须之外，表明有少数房屋的总价远高于大多数房屋。
- 多样性：
  - 房屋总价的广泛分布表明市场上供应的房屋价格差异很大，从低价到高价不等，这为不同预算的购房者提供了广泛的选择。
- 其他：
  - 整体分布情况同李汛老师上次课后作业里收入的分布极为类似——少数人拥有极高的收入，房屋总价上似乎也印证了这一点。

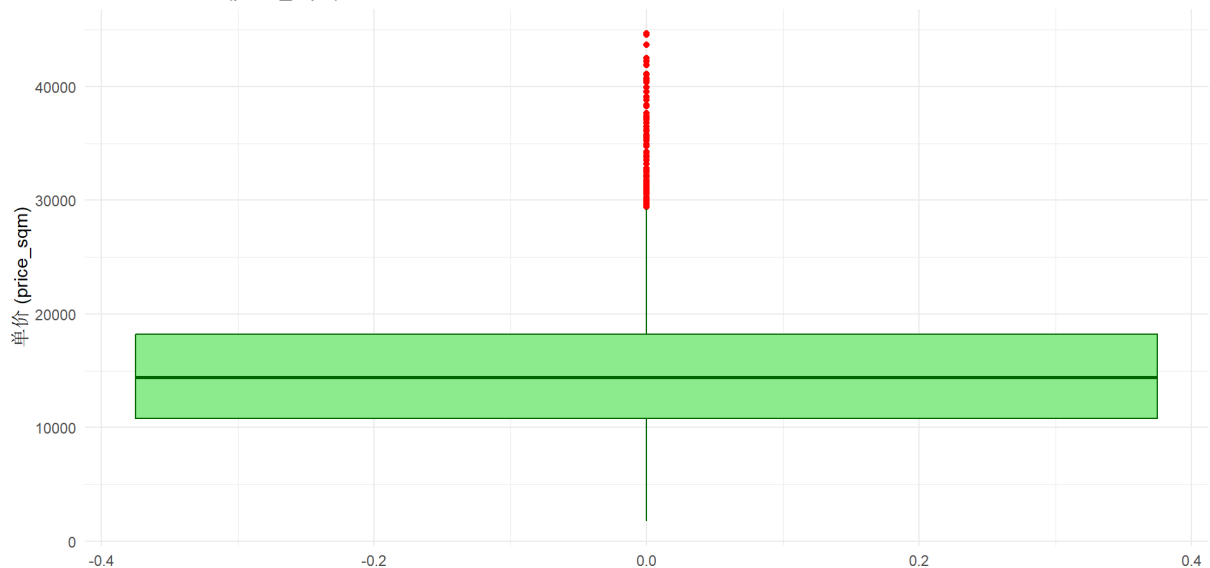
## 4.4 单价 (price\_sqm) 的数值描述与图形

```
##
## ----- 分析变量: 单价 (price_sqm) -----
## 均值 (Mean): 15148.49
## 中位数 (Median): 14404
## 众数 (Mode): 14394
## 极差 (Range): 42885
## 四分位距 (IQR): 7411.75
## 方差 (Variance): 39982547
## 标准差 (Standard Deviation): 6323.175
## 偏度 (Skewness): 1.079464
## 峰度 (Kurtosis): 2.025625
```

直方图 - 单价 (price\_sqm)



箱线图 - 单价 (price\_sqm)



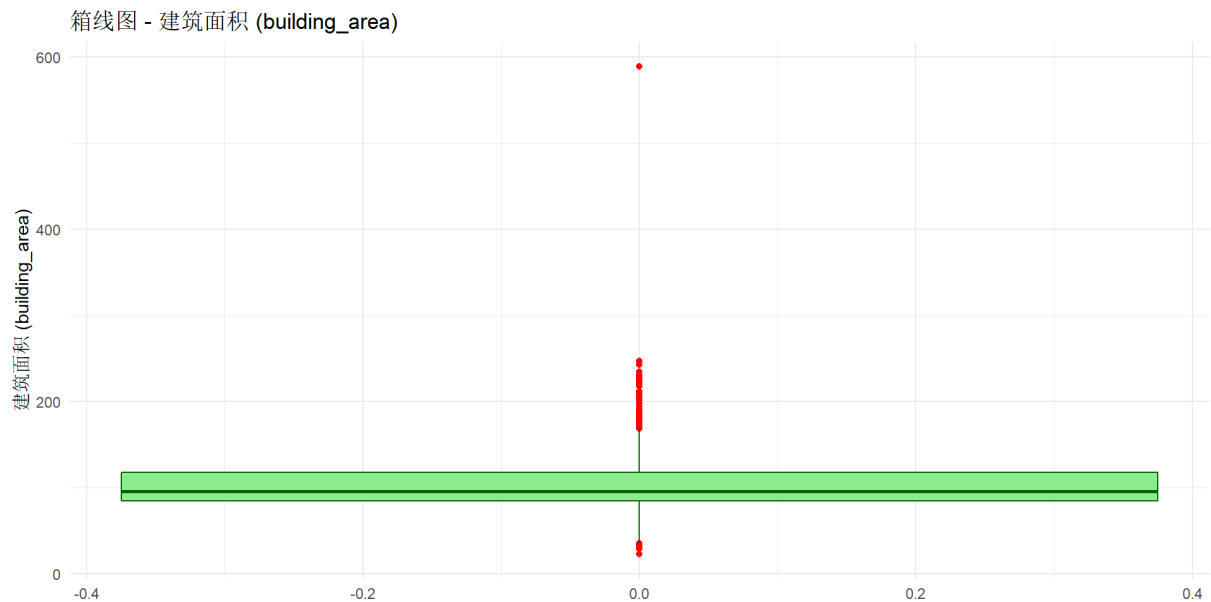
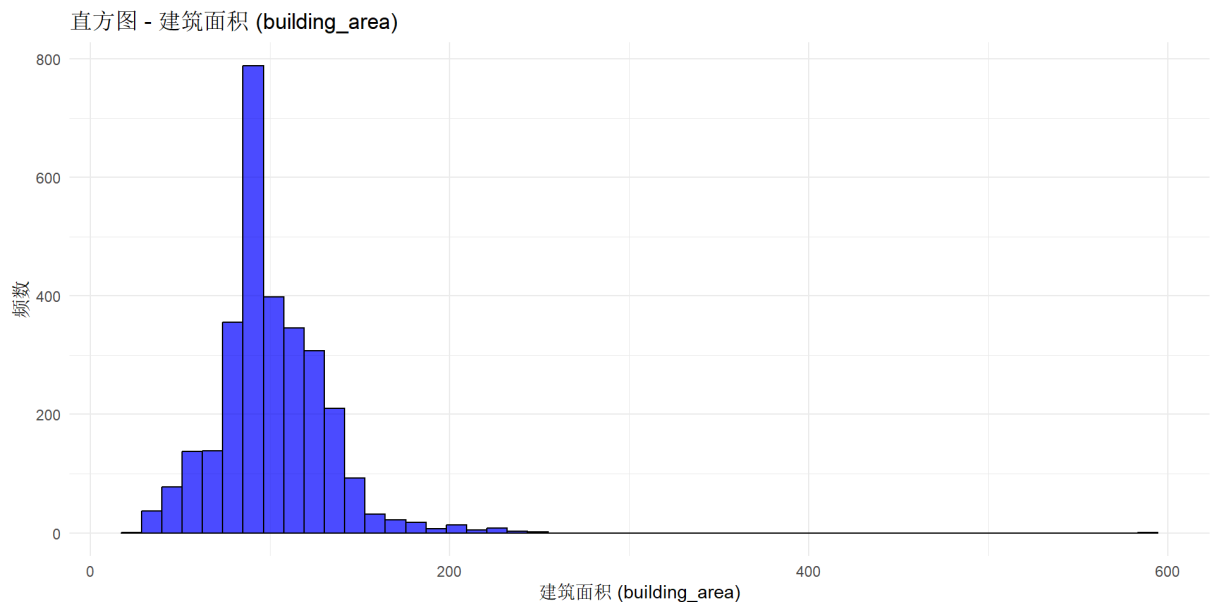
发现：

- 分布：
  - 直方图显示单价的分布呈现右偏（正偏）分布，大多数房屋的单价集中在10000到20000之间。

- 偏度 (Skewness) 为1.079464, 表明数据分布有一定的右偏, 存在一些极端的高值。
- 集中趋势:
  - 均值 (15148.49) 略高于中位数 (14404), 这进一步证实了分布的右偏性。
  - 众数 (14394) 表明最常见的单价是14394, 这可能反映了市场上某一特定价格区间的房屋较为普遍。
- 离散程度和异常值:
  - 标准差 (6323.175) 和四分位距 (7411.75) 表明单价之间存在较大的差异, 且数据分布较为分散。
  - 极差 (42885) 非常大, 从最小值到最大值的范围很广, 这表明市场上既有非常便宜的房屋, 也有非常昂贵的房屋。
  - 箱线图显示存在一些异常值, 这些点位于上须之外, 表明有少数房屋的单价远高于大多数房屋。
- 多样性:
  - 单价的广泛分布表明市场上供应的房屋价格差异很大, 从低价到高价不等, 这为不同预算的购房者提供了广泛的选择。

## 4.5 建筑面积 (building\_area) 的数值描述与图形

```
##
## ----- 分析变量: 建筑面积 (building_area) -----
## 均值 (Mean): 100.8669
## 中位数 (Median): 95.545
## 众数 (Mode): 60
## 极差 (Range): 565.89
## 四分位距 (IQR): 32.7625
## 方差 (Variance): 922.9442
## 标准差 (Standard Deviation): 30.38
## 偏度 (Skewness): 2.079785
## 峰度 (Kurtosis): 23.63585
```

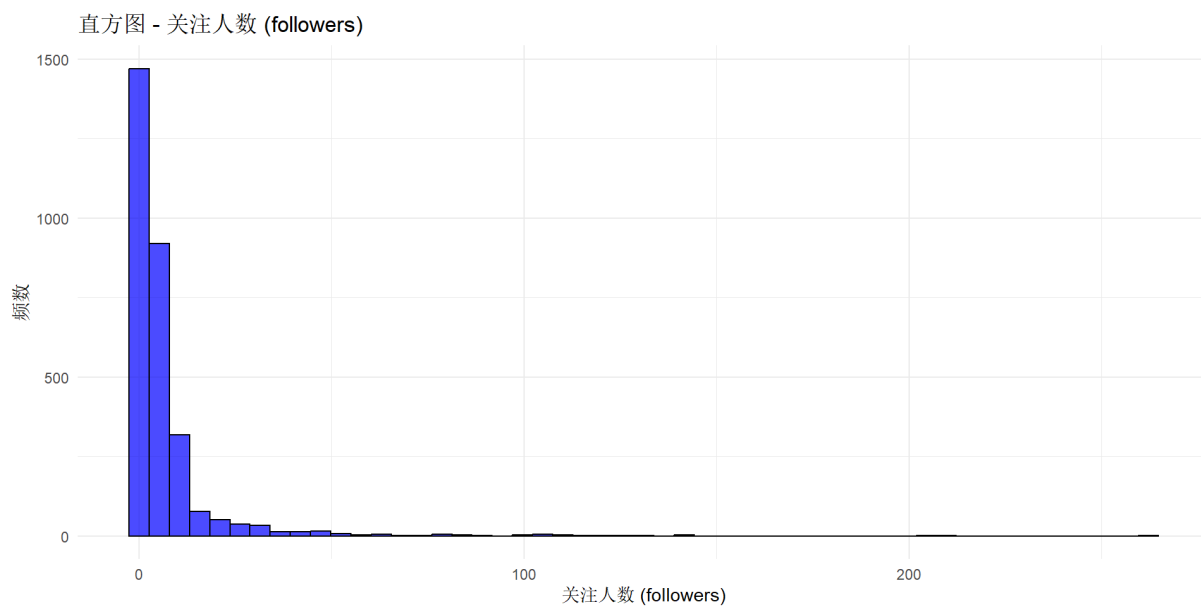


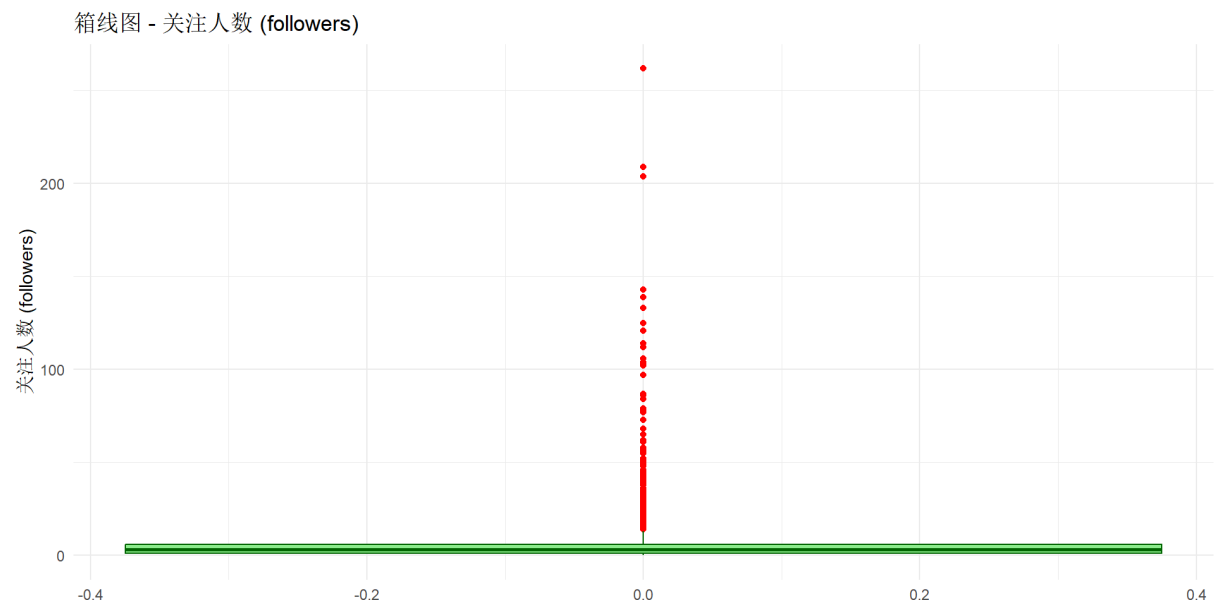
发现:

- 分布：
  - 直方图显示建筑面积的分布呈现右偏（正偏）分布，大多数房屋的建筑面积集中在较小的区间，随着建筑面积的增加，房屋数量迅速减少。
  - 偏度（Skewness）为2.079785，表明数据分布有显著的右偏，存在一些极端的高值。
- 集中趋势：
  - 均值（100.8669）略高于中位数（95.545），这进一步证实了分布的右偏性。
  - 众数（60）表明最常见的建筑面积是60平方米，这可能反映了市场上小户型房屋的普遍性。
- 离散程度和异常值：
  - 标准差（30.38）和四分位距（32.7625）表明建筑面积之间存在一定的差异，且数据分布较为分散。
  - 极差（565.89）非常大，从最小值到最大值的范围很广，这表明市场上既有非常小的房屋，也有非常大的房屋。
  - 箱线图显示存在一些异常值，这些点位于上须之外，表明有少数房屋的建筑面积远高于大多数房屋。
- 多样性：
  - 建筑面积的广泛分布表明市场上供应的房屋类型多样，从小户型到大户型不等，为不同需求的购房者提供了广泛的选择。

## 4.6 关注人数 (followers) 的数值描述与图形

```
##
## ----- 分析变量: 关注人数 (followers) -----
## 均值 (Mean): 6.613667
## 中位数 (Median): 3
## 众数 (Mode): 0
## 极差 (Range): 262
## 四分位距 (IQR): 5
## 方差 (Variance): 231.5576
## 标准差 (Standard Deviation): 15.21702
## 偏度 (Skewness): 6.900767
## 峰度 (Kurtosis): 68.17057
```





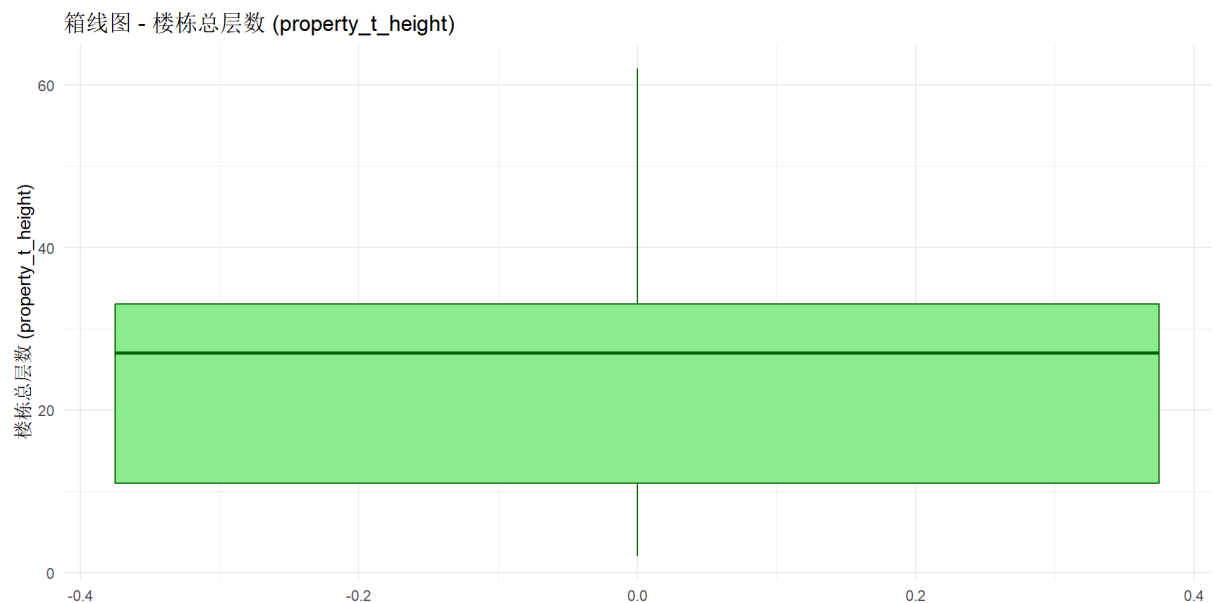
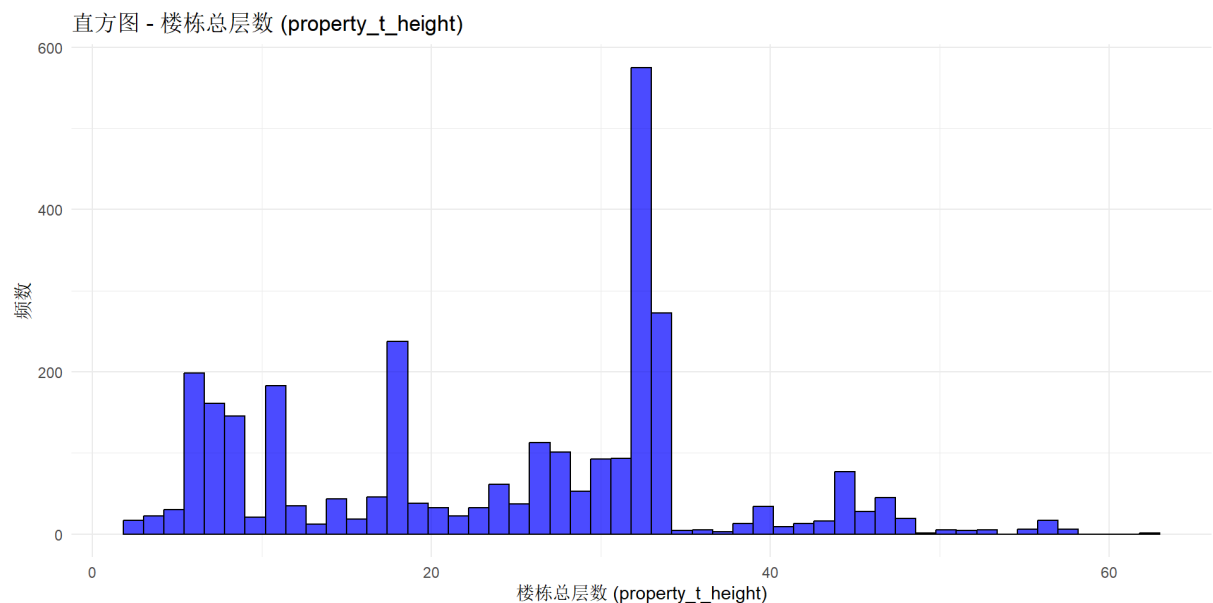
发现:

- 分布:
  - 直方图显示关注人数的分布呈现极端的右偏（正偏）分布，大多数数据点集中在较低的关注人数区间，而随着关注人数的增加，频数迅速下降。
  - 偏度（Skewness）为6.900767，这是一个非常高的正偏度值，表明数据分布有长尾向右延伸，存在一些极端的高值。
- 集中趋势:
  - 均值（6.613667）远高于中位数（3），这进一步证实了分布的右偏性，因为如果分布是对称的，均值和中位数应该是相等的。
  - 众数（0）表明最常见的关注人数是0，这可能意味着大多数二手房源没有获得关注或者关注人数非常少。
- 离散程度和异常值:
  - 标准差（15.21702）和四分位距（5）表明关注人数之间存在较大的差异，且数据分布较为分散。
  - 极差（262）非常大，从最小值到最大值的范围很广，这表明存在一些二手房源获得了非常高的关注。
  - 箱线图显示存在一些异常值，这些点位于上须之外，表明有少数房源的关注人数远高于大多数房源。
- 多样性:
  - 关注人数的广泛分布表明市场上房源的关注度差异很大，一些房源可能因为各种原因（如价格、位置、条件等）获得了较高的关注，而大多数房源则没有获得太多关注。

## 4.7 楼栋总层数 (property\_t\_height) 的数值描述与图形

```
##
## ----- 分析变量: 楼栋总层数 (property_t_height) -----
## 均值 (Mean): 24.218
## 中位数 (Median): 27
## 众数 (Mode): 33
## 极差 (Range): 60
## 四分位距 (IQR): 22
## 方差 (Variance): 154.9588
## 标准差 (Standard Deviation): 12.44824
## 偏度 (Skewness): 0.04850289
## 峰度 (Kurtosis): -0.797201
```





发现:

- 分布:
  - 直方图显示楼栋总层数的分布呈多峰形态，主要集中在若干特定的层数范围内，而非均匀分布。这表明市场上存在几种常见的楼栋高度类别。
  - 偏度 (Skewness) 接近于零 (0.04850289)，表明数据分布大致对称，没有明显偏向某一侧的趋势。
- 集中趋势:
  - 均值 (24.218) 和中位数 (27) 相近，但略有差异，这可能是由于数据分布中的多峰特性影响了均值的计算。
  - 众数 (33) 表明最常见的楼栋总层数是33层，这可能反映了某种建筑设计规范或市场需求的影响。
- 离散程度和异常值:
  - 标准差 (12.44824) 和四分位距 (22) 表明楼栋总层数之间存在一定的差异，但相对于整个分布范围而言，数据分布相对紧凑。
  - 极差 (60) 表明从最小到最大的楼栋总层数跨度很大，涵盖了各种高度的建筑物。
  - 箱线图显示存在一些异常值，这些点位于上下须之外，表明有少数楼栋的总层数远高于或低于大多数楼栋。
- 多样性:
  - 楼栋总层数的广泛分布表明市场上供应的楼栋类型多样，从小高层到超高层都有覆盖，为不同需求的购房者提供了广泛的选择。

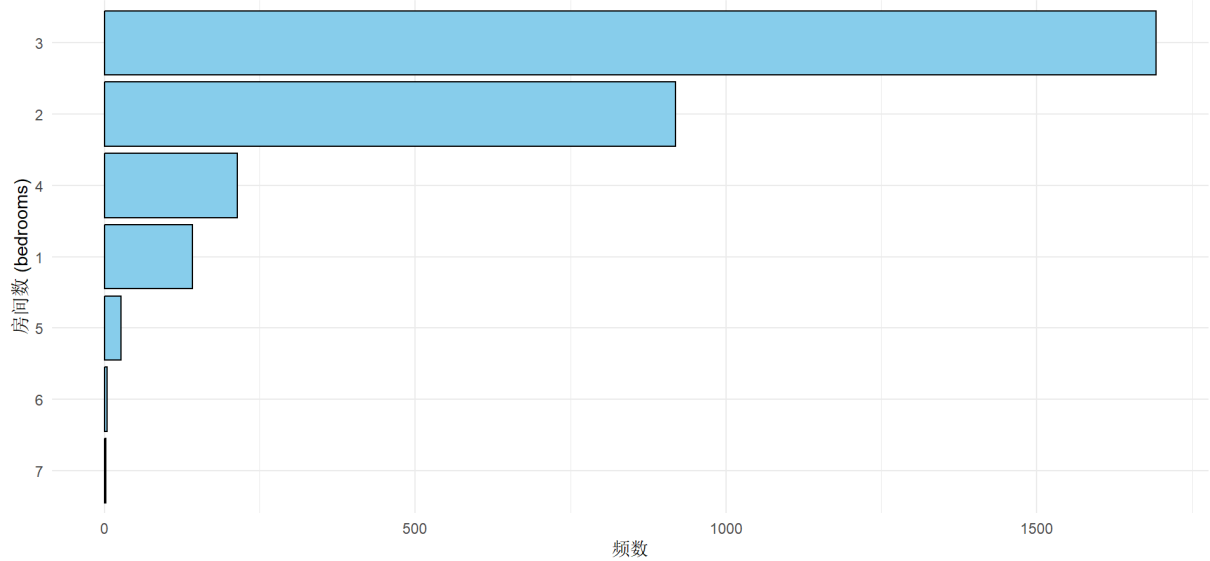
## 4.8 房间数 (bedrooms) 的数值描述与图形

```
## ----- 分析变量: 房间数 (bedrooms) -----
```

```
## 频数分布:
```

```
##
## 1 2 3 4 5 6 7
## 142 919 1692 214 27 4 2
```

条形图 - 房间数 (bedrooms)



发现：

- 三室的房屋数量最多，有1692套，这表明三室房屋在市场中非常普遍，可能是家庭居住的首选房型。二室的房屋数量也较多，有919套，这可能适合小家庭或单身人士。市场上供应的房型以中小面积为主，这可能与购房者的购买力和居住需求有关。
- 四室的房屋数量为214套，这表明四室房屋在市场中也占有一定比例，可能适合有较大家庭或需要额外空间的家庭，或者用于出租给需要更多房间的租户。一室的房屋数量为142套，这可能适合独居者或作为投资的出租物业。这两种实际需求都存在，但相对上一条会少很多，可能并非主流需求。
- 五室及以上的房屋数量显著减少，五室房屋仅有27套，六室和七室房屋的数量更是稀少，分别为4套和2套。这可能表明大户型房屋在市场中较为稀缺，或者这些房型的市场需求相对较小（购买门槛较高）。

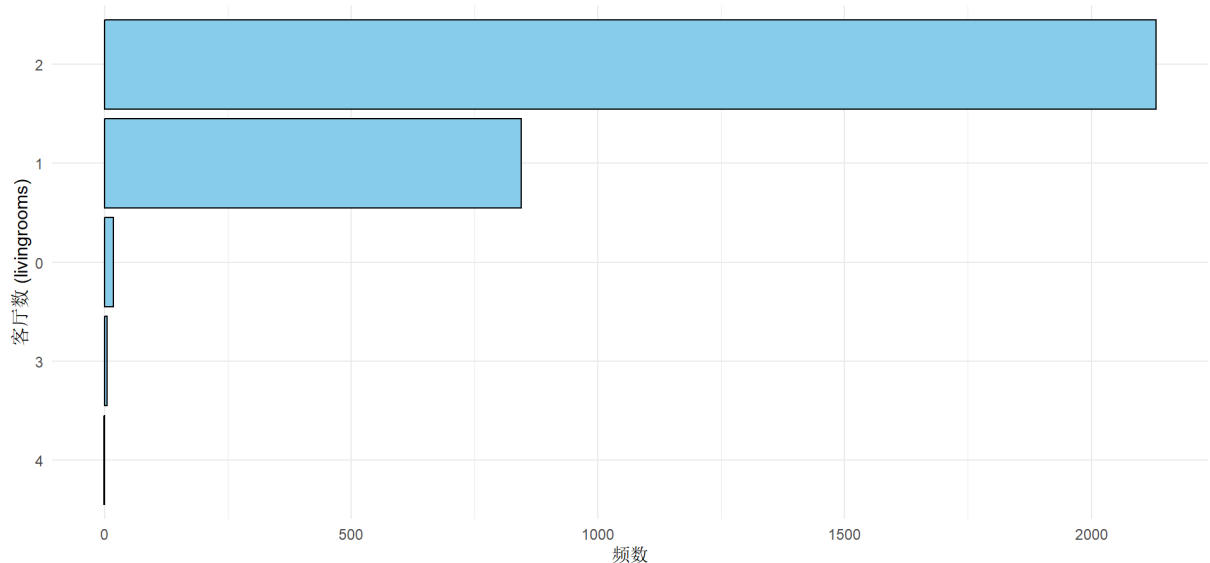
## 4.9 客厅数 (livingrooms) 的数值描述与图形

```
## ----- 分析变量：客厅数 (livingrooms) -----
```

```
## 频数分布：
```

```
##
## 0 1 2 3 4
## 18 845 2131 5 1
```

条形图 - 客厅数 (livingrooms)

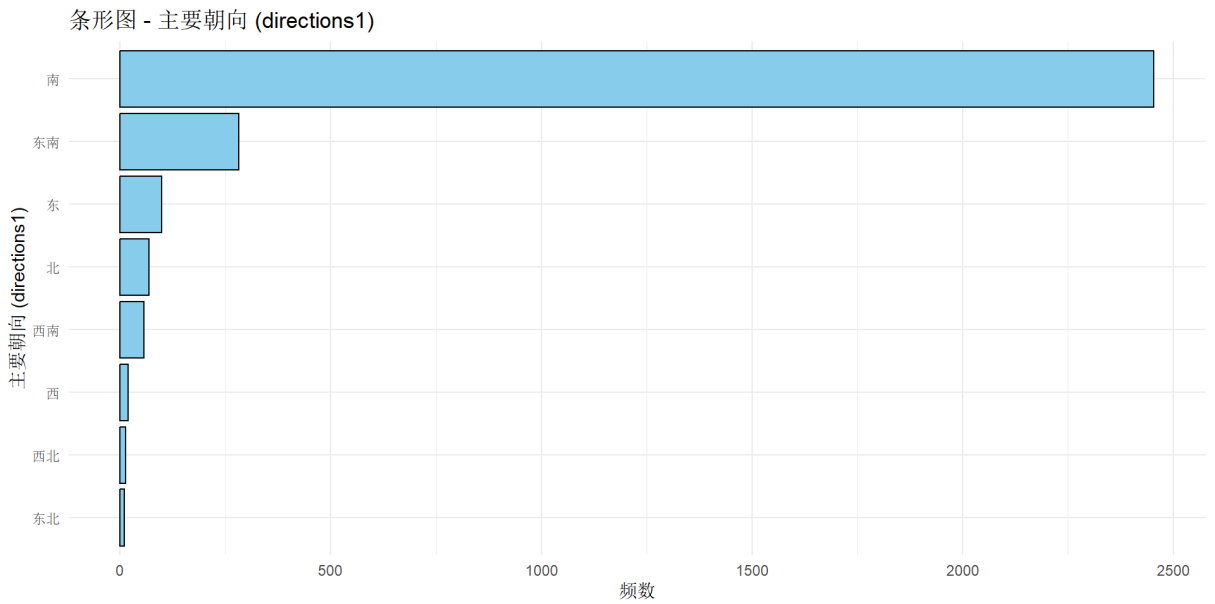


发现：

- 两厅的房屋数量最多，有2131套，这表明两厅房屋在市场中非常普遍，可能是家庭居住的首选房型。一厅的房屋数量也较多，有845套，这可能适合小家庭或单身人士。两厅和一厅房屋占据了绝大多数，这可能表明市场上的房屋设计倾向于提供至少一个客厅，以满足基本的居住和社交需求。
- 没有客厅的房屋数量为18套，这可能表明这种设计不太受欢迎，或者适用于特定的居住需求，如工作室或小型公寓。
- 三厅和四厅的房屋数量非常少，分别为5套和1套。这可能是因为这些房型的购买门槛较高，或者是因为这些房型的市场需求相对较小。

4.10 房屋主要朝向 (directions1) 的数值描述与图形

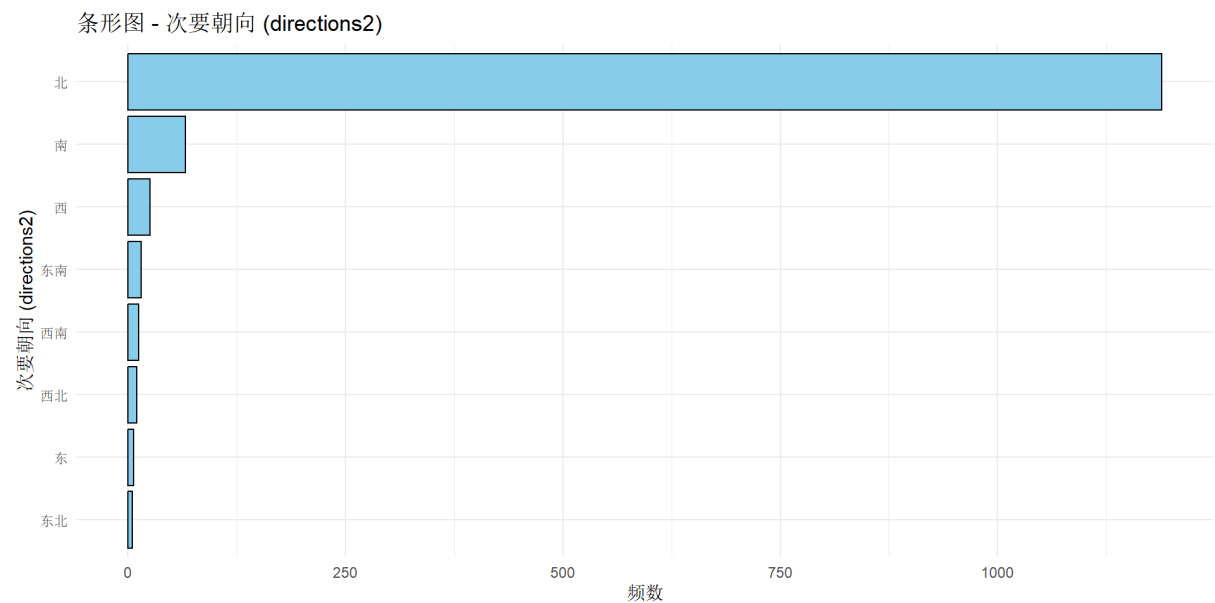
## ----- 分析变量：主要朝向（directions1） -----									
## 频数分布：									
##									
##	北	东	东北	东南	南	西	西北	西南	
##	68	98	10	281	2454	19	13	57	



- 发现：
- 南向（南）的房屋数量最多，有2454套，这表明南向房屋在市场中非常常见，可能是由于南向房屋通常具有更好的采光和通风条件。另外从风水的角度，“坐北朝南”被认为是吉祥的朝向，更有利于居住者的运势和健康——这点在国内算是一种共识，从数据上似乎也能体现。
  - 东南向（东南）的房屋数量也较多，有281套，这可能也是因为东南向房屋也具有相对良好的采光条件。
  - 东向（东）、北向（北）、东北（东北）、西北（西北）、西南（西南）和西向（西）的房屋数量非常少，分别为98套、68套、10套、13套、57套和19套。这可能表明这些朝向的房屋在市场中不太受欢迎，或者这些朝向的房屋数量本身就较少。

4.11 房屋次要朝向 (directions2) 的数值描述与图形

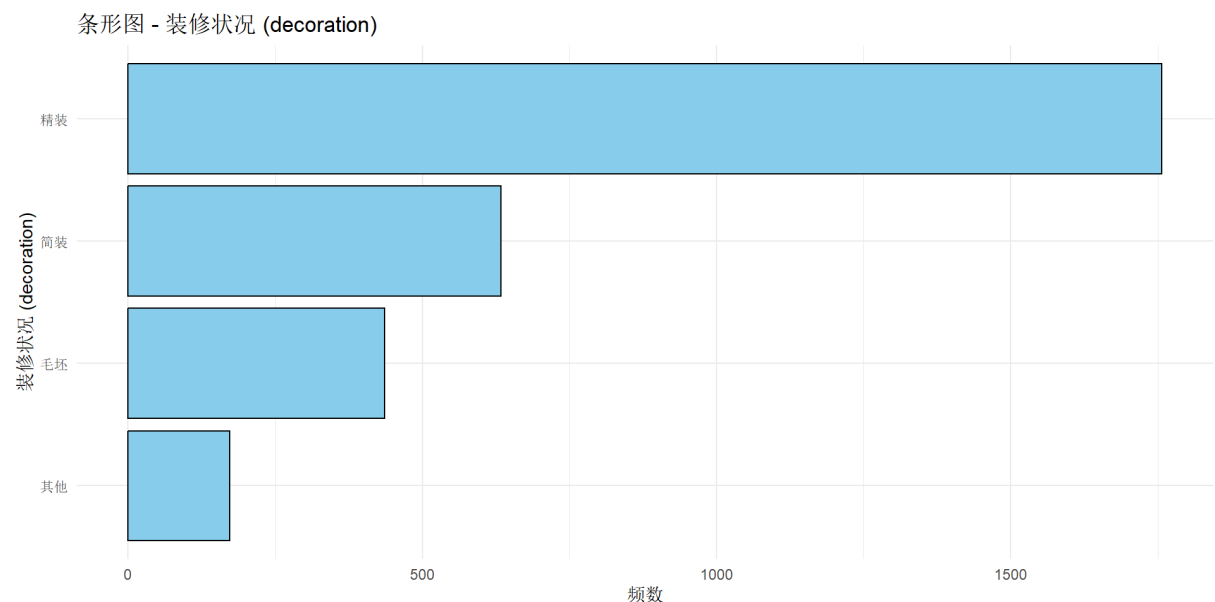
## ----- 分析变量：次要朝向（directions2） -----									
## 频数分布：									
##									
##	北	东	东北	东南	南	西	西北	西南	
##	1189	6	5	15	66	25	10	12	



发现：

- 北向（北）的房屋数量最多，有1189套。在考虑次要朝向时，北向房屋在市场中占据了很大的比例。南向（南）的房屋数量也较多，有66套，这可能表明南向房屋在次要朝向中也较为常见。次要朝向上主要以北向和南向为主，这可能与购房者对房屋采光和通风的偏好有关。
- 其他朝向的房屋数量相对较少，东向（东）和东北向（东北）各有6套和5套，东南向（东南）有15套，西向（西）有25套，西北向（西北）和西南向（西南）分别有10套和12套。可能是因为这些朝向的房屋需求相对较小，或者是因为建筑规划和设计上的限制。

## 4.12 装修状况 (decoration) 的数值描述与图形



发现：

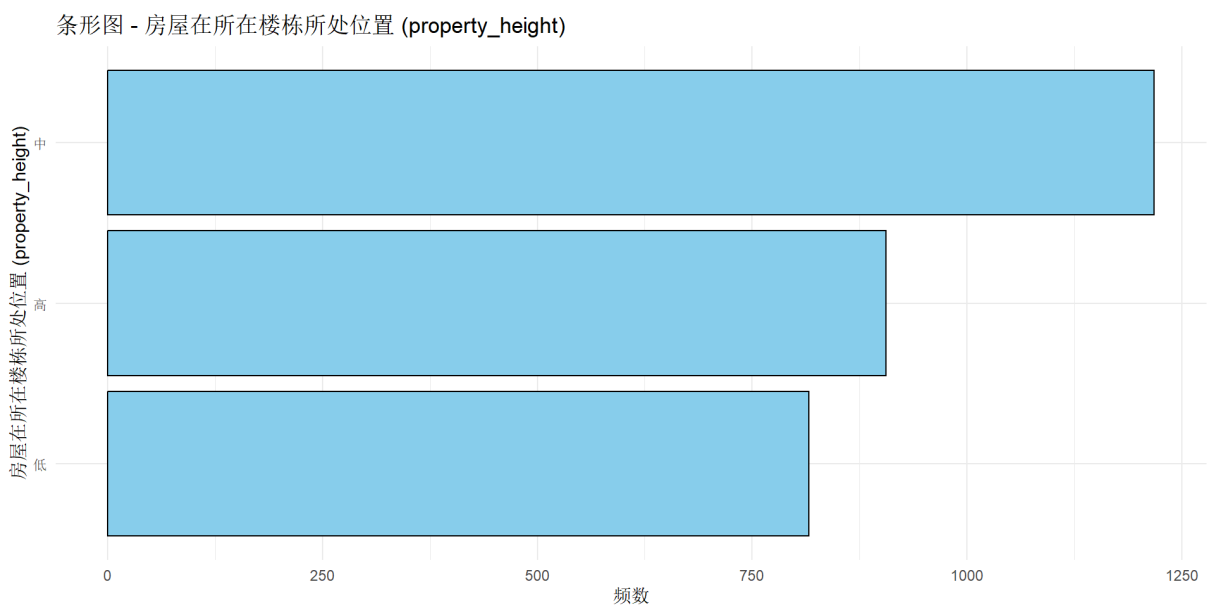
- 精装（精装）的房屋数量最多，有1757套，这表明精装房屋最为常见，可能是因为很多开发商交房时就会精装。简装（简装）的房屋数量也较多，有634套。精装和简装房屋占据了绝大多数，这表明市场上的房屋在二级市场出售时更多是以装修的形式，可能是因为基本都是装修后自住再出售。
- 毛坯（毛坯）的房屋数量为436套，供应量较少，这可能是因为开发商更倾向于提供装修好的房屋以吸引买家。
- 其他（其他）的房屋数量最少，有173套，这可能包括了一些特殊装修状况或不常见的装修类型。

4.13 房屋在所在楼栋所处位置 (property\_height) 的数值描述与图形

```
## ----- 分析变量：房屋在所在楼栋所处位置 (property_height) -----

## 频数分布：

##
## 低 高 中
## 816 906 1218
```



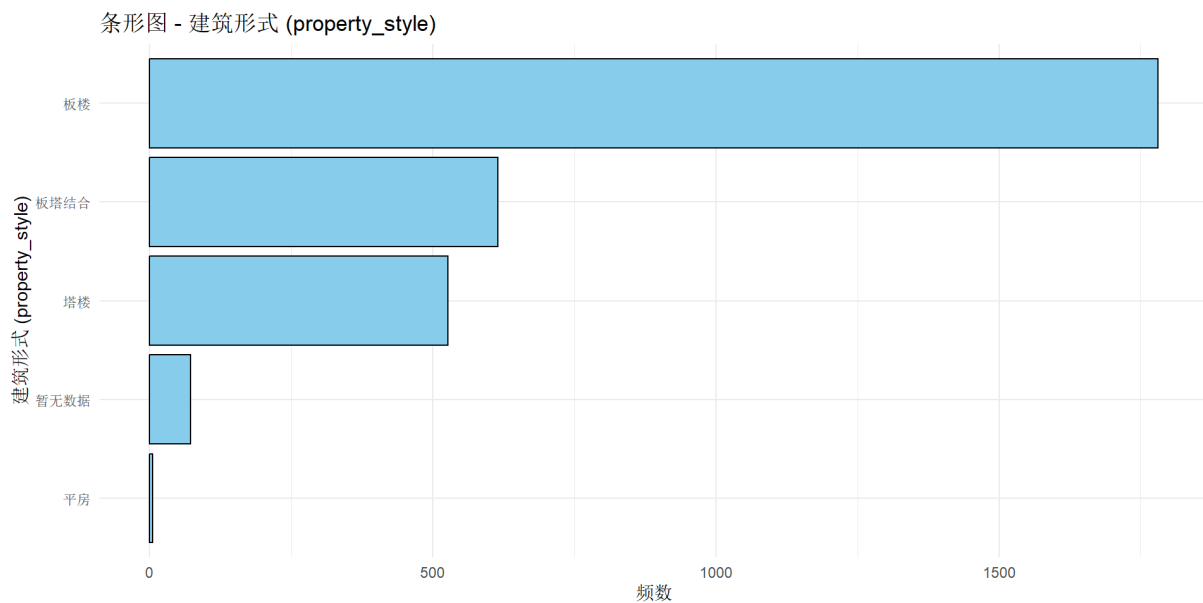
- 发现：
- 中层位置供应最多，有1218套，这表明市场上中层位置的房屋最为常见。可能是因为中层既能避免底层潮湿的问题，又能享受较好的视野和采光。
  - 高层和低层供应相当，高层的906套略高于低层的816套。两者相差不大，表明市场上对于高层和低层的位置需求相对均衡。高层可能受喜欢开阔视野的人群喜爱，而低层则可能因方便出行等因素受到另一部分人的青睐。
  - 尽管中层位置的房屋数量最多，但高层和低层的数量也不少，说明市场上对于不同楼层位置的房屋可能都有一定需求。

4.14 建筑形式 (property\_style) 的数值描述与图形

```
## ----- 分析变量：建筑形式 (property_style) -----

## 频数分布：

##
## 板楼 板塔结合 平房 塔楼 暂无数据
## 1781 615 5 527 72
```



发现：

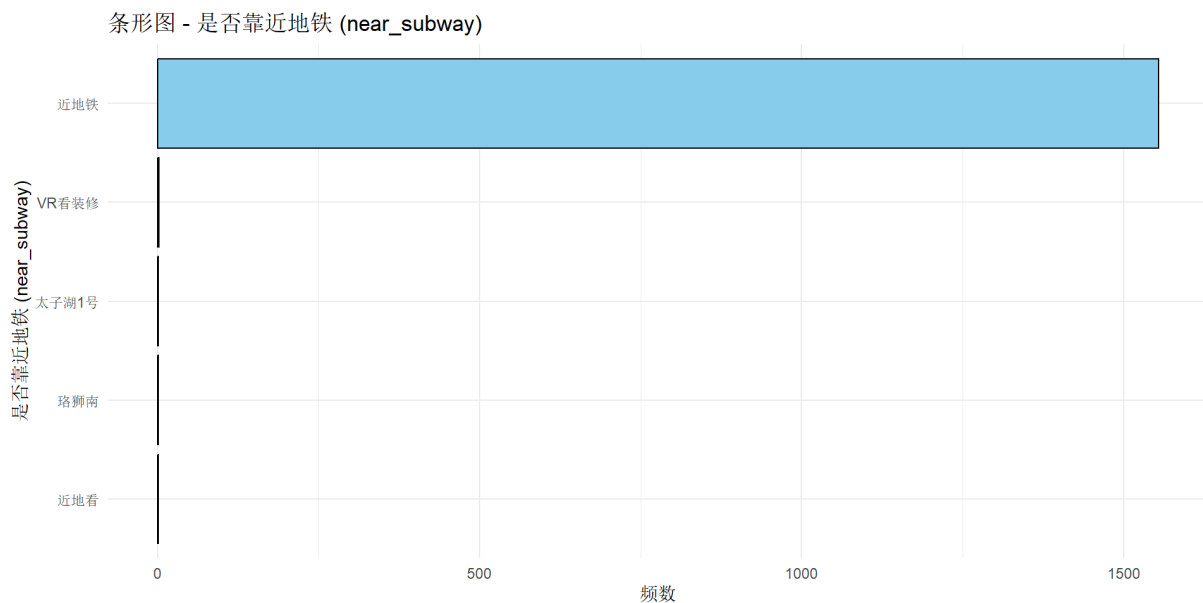
- 板楼是最常见的建筑形式，共有1781套。这表明市场上对于这种类型的建筑供应量最大，可能是因为其结构简单、成本较低，并且能够提供良好的自然采光和通风效果。
- 排名第二的是板塔结合的建筑形式，共有615套。这种设计结合了板式建筑的优点（如良好的采光和通风）和塔式建筑的优势（如更高的密度和更多的公共设施）。它适合那些希望在有限的土地资源内实现更高居住效率的项目；塔楼的数量为527套，介于板塔结合和平房之间。塔楼通常指高层或多层建筑，可以容纳更多居民并且提供更多公共服务设施，但在某些情况下可能会牺牲一部分自然光线和空气流通的质量。这两种建筑可能相对板楼都有一定局限性，导致并非主流。
- 平房的数量仅为5套，而暂无数据的记录也只有72套。这两种情况在样本中占比较小，可能是因为平房通常适用于农村地区或别墅区，而不是城市中心地带；而暂无数据的情况可能是由于数据收集过程中的遗漏或错误导致的。

## 4.15 是否靠近地铁 (near\_subway) 的数值描述与图形

```
## ----- 分析变量：是否靠近地铁 (near_subway) -----
```

```
## 频数分布：
```

```
##
## VR看装修    近地看    近地铁    珞狮南 太子湖1号
##          2          1      1554          1          1
```



发现：

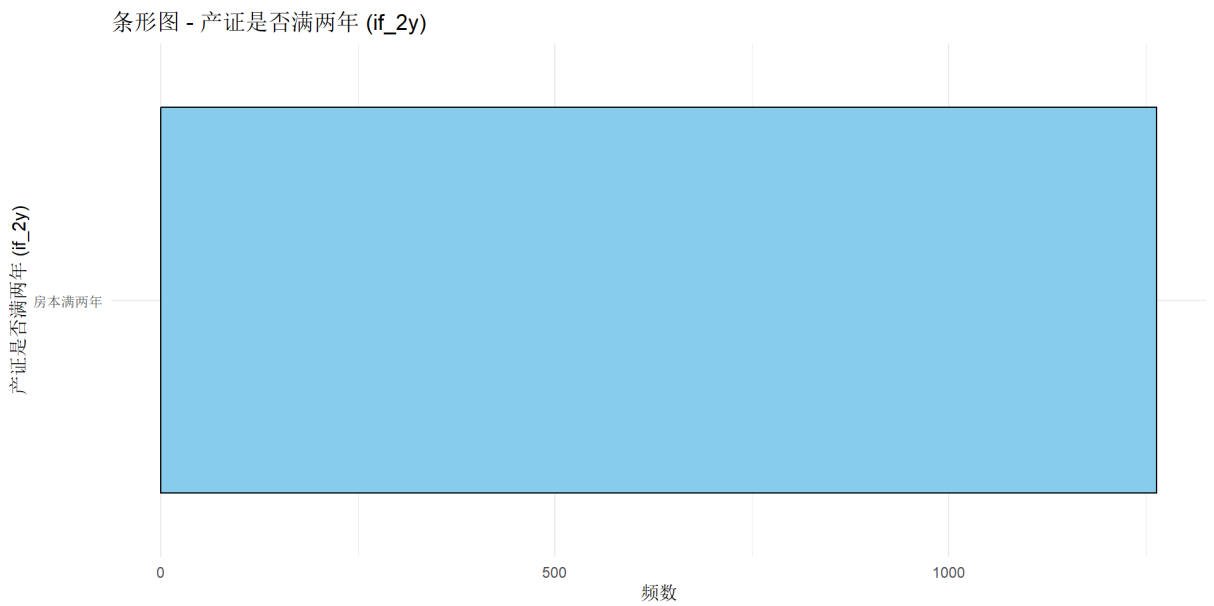
- 靠近地铁的房源数量最多，共有1554套。这表明市场上对于靠近地铁站的房源需求量较大，可能是因为交通便利性是许多购房者的重要考量因素之一，靠近地铁的房源具有较高的竞争力和吸引力。
- 其他选项如VR看装修、近地看、珞狮南太子湖1号等的房源数量都非常少，分别为2套、1套、1套和1套。这意味着这些选项在当前市场上的关注度不高，或者是由于数据采集过程中的一些特殊原因导致的结果。

4.16 产证是否满2年 (if\_2y) 的数值描述与图形

```
## ----- 分析变量：产证是否满两年 (if_2y) -----

## 频数分布：

##
## 房本满两年
##      1264
```



发现：

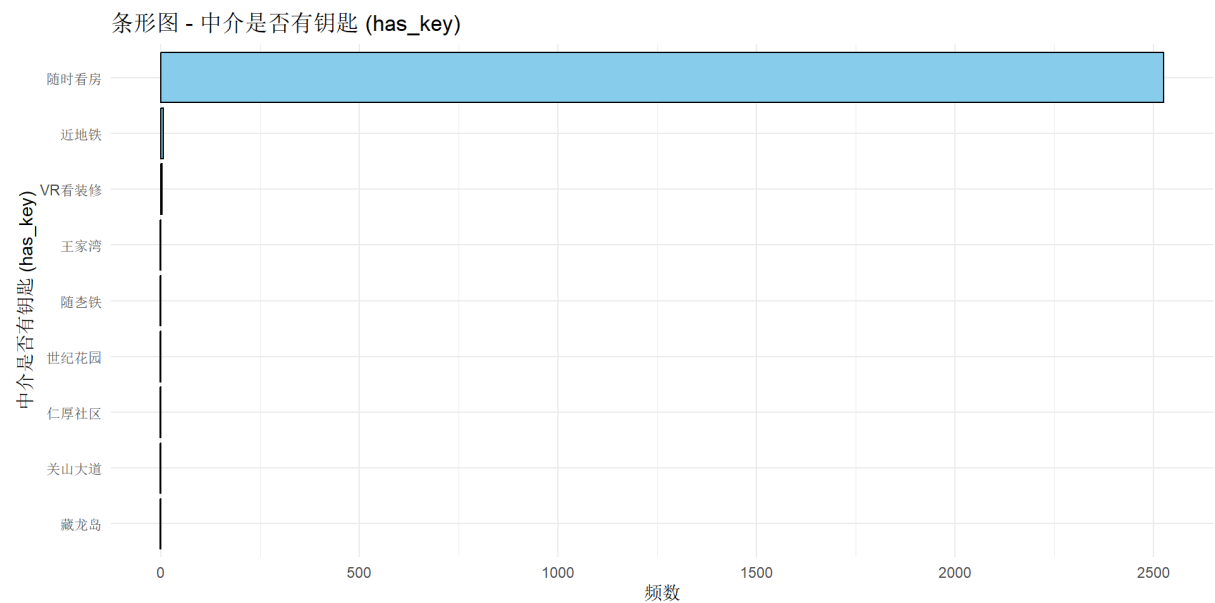
- 产权证书已满两年的房源数量为1264套。这表明市场上相当多的房源都已经达到了规定的持有年限要求，符合交易条件；剩余的房源此变量为缺失值，缺失值数量大于1264套，无法确定是何种情况。
- 目前只看到了一种状态——即产权证书已满两年的情况。这可能是因为数据集本身的特点所致，也可能暗示着当前市场环境下，新购入未满两年的房产较少进入市场流转。

4.17 中介是否有钥匙 (has\_key) 的数值描述与图形

```
## ----- 分析变量：中介是否有钥匙 (has_key) -----

## 频数分布：

##
## VR看装修  藏龙岛  关山大道  近地铁  仁厚社区  世纪花园  随时看房  随李铁
##      4      1      1      7      1      1      2525      1
##   王家湾
##      1
```



发现：

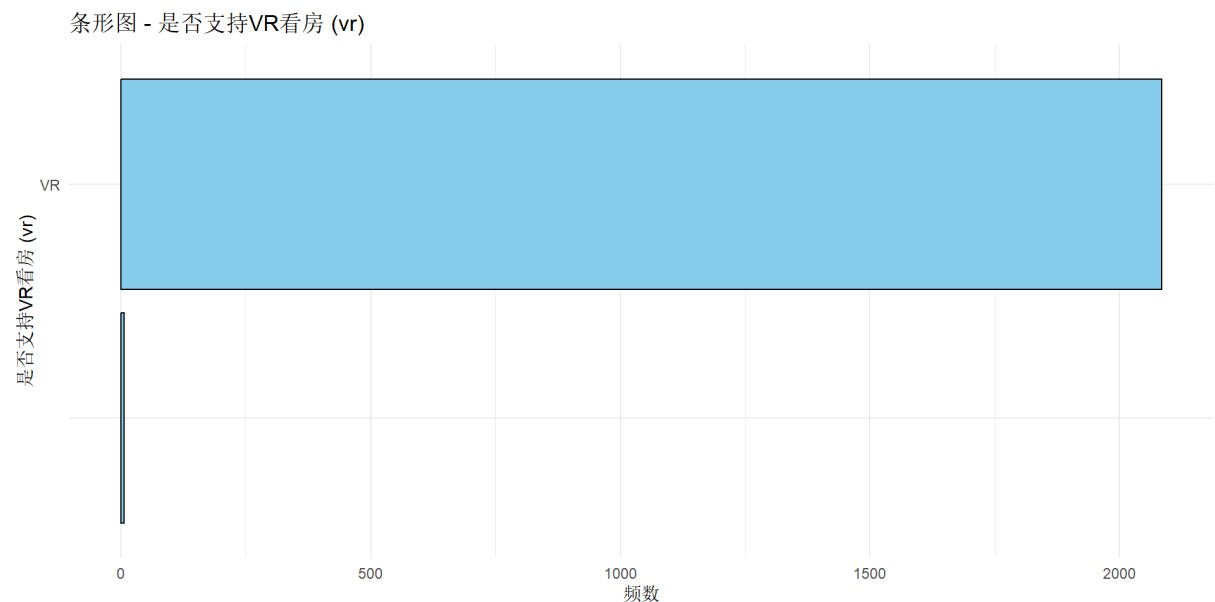
- “随时看房”的比例最高，共有2525个案例，是最常见的方式。这可能是因为市场上大多数房源都可以随时安排参观，无需预约等待，这种方式可大大提高了购房者的看房体验和效率，更受青睐。
- 其他选项如VR看装修、藏龙岛、关山大道、近地铁、仁厚社区、世纪花园、随空铁、王家湾等的占比都非常低，每个选项只有1至7个案例。这可能意味着这些地点或服务方式在当前市场中并不流行，或者是因为数据采集过程中的一些特殊原因导致的结果。

## 4.18 是否支持VR看房 (vr) 的数值描述与图形

```
## ----- 分析变量：是否支持VR看房 (vr) -----
```

```
## 频数分布：
```

```
##  
##      VR  
## 6 2084
```



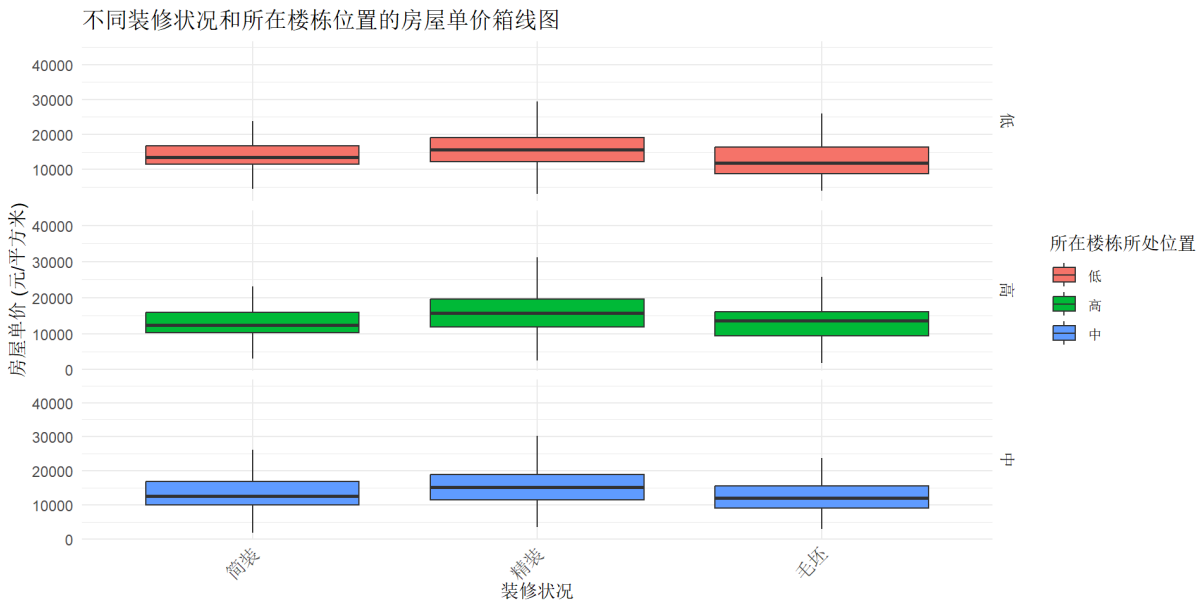
发现：

- 支持VR看房的房源数量为2084套。这表明市场上有一部分房源已经开始采用虚拟现实技术来进行展示，为购房者提供了更为直观和便捷的看房体验。几乎所有的房源都在利用VR技术吸引客户注意力。这种现象可能源于近年来科技发展带来的创新应用以及消费者对高科技产品的高度认可度。
- 相比之下，此列为空的房源有6套。这个数字非常之小，几乎可以忽略不计，可能是因为数据采集过程中的一些特殊原因导致的结果。

## 4.19 探索问题1：房屋单价、装修状况和房屋在所在楼栋所处位置的



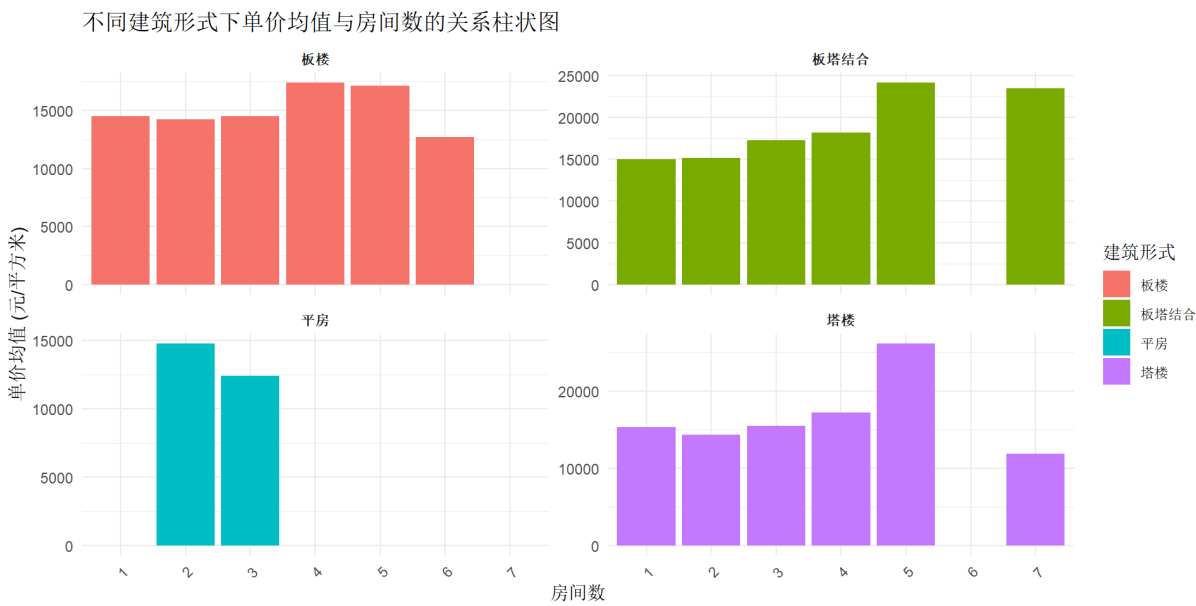
关系



发现：

- 装修状况对房价的影响：
  - 不同装修状况下的房屋单价存在显著差异。例如，精装修的房子单价明显高于简单装修的房子。这表明装修水平是影响房价的一个重要因素。
- 楼层位置对房价的影响：
  - 同样类型的装修下，位于较高楼层的房子单价往往高于较低楼层的房子。比如，在精装修类别中，高层房子的单价要高于中层和低层的房子。这反映了楼层位置也是决定房价的关键因素之一。
- 关联效应：
  - 装修状况和楼层位置之间的关系也很重要。例如，在某些情况下，即使装修水平相同，但由于楼层的不同，房价也会有很大差别。这暗示了两者结合在一起可能会产生更大的价格变化幅度。
- 价格分布情况：
  - 图表还展示了各个组别内房价的分布情况。可以看到，即使是同一类别的房子，其价格也可能有很大的波动性。这可能是由于其他未考虑的因素（如地理位置、周边环境等）造成的结果。

4.20 探索问题2：房屋单价、房间数和建筑形式的关系



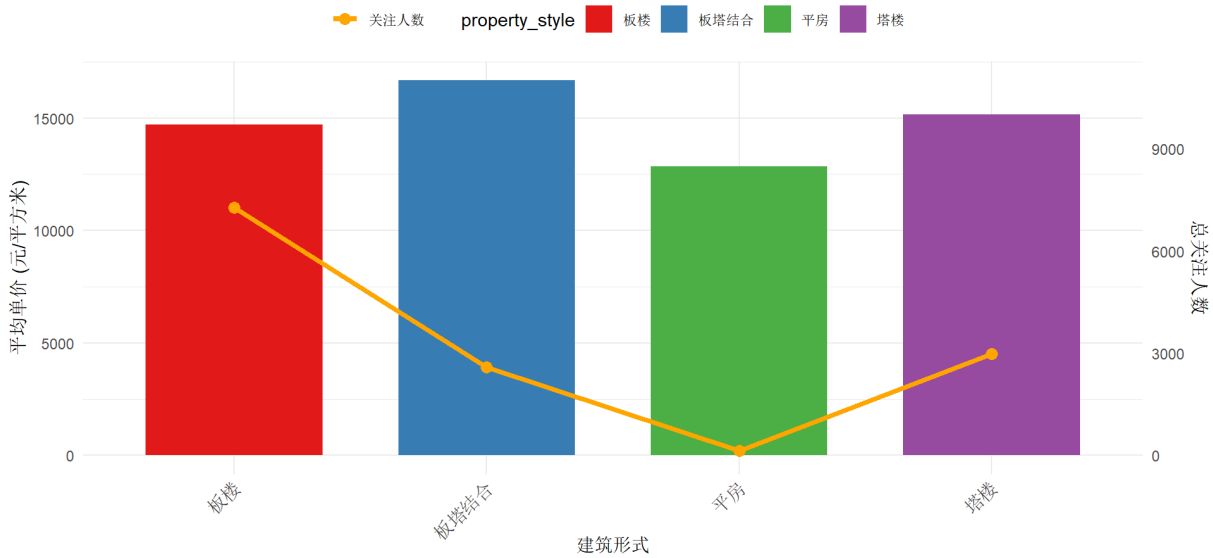
发现：

- 建筑形式对单价均值的影响：
  - 板楼的单价均值在不同房间数下相对稳定，这可能表明板楼的市场需求和价格稳定性较高。
  - 板塔结合和塔楼的单价均值在不同房间数下有所波动，这可能反映了这两类建筑的价格敏感性。
  - 平房的2室均价高于3室，但差别不大，这可能意味着平房的户型和价格相对稳定。
  - 塔楼的单价均值在7室有明显下降趋势，这可能表明塔楼在较大房间数时的单位面积价格较低。

- 建筑形式与房间数的关系：
  - 板楼、板塔结合和塔楼的房间数分布较为均匀，这可能表明这两种建筑形式在市场上的供应较为全面，能够满足不同的需求。
  - 平房仅有2室和3室，可能是受限于建筑设计或其他因素所致。

### 4.21 探索问题3：房屋单价、关注人数和建筑形式的关系

不同建筑形式下的房屋单价与关注人数



发现：

- 建筑形式与平均单价的关系：
  - 板塔结合的平均单价最高，接近15000元/平方米，这可能反映了板楼在市场上的受欢迎程度或其建筑质量较高。
  - 平房的平均单价最低，这可能意味着平房在市场上的需求较低或其建筑质量或设计可能不如其他建筑形式。
- 建筑形式与关注人数的关系：
  - 板楼和塔楼平均单价差不多，但平均关注人数要高得多，这可能与其有更高的性价比有关，或者有某些个别楼型影响了均价。
  - 平房的关注人数最低，这可能与其较低的平均单价有关，可能意味着平房在市场上的需求较低。

## 5 发现总结

通过对武汉二手房市场的探索性分析，我发现房屋的单价受到装修状况、楼层位置、建筑形式和房间数等因素的影响。精装修房屋和高层房屋通常单价较高，而塔楼在大房间数时单价较低。此外，房屋的关注度与单价和建筑形式有关，板楼更受市场欢迎。这些发现有助于购房者、投资者和市场分析师更好地理解市场动态，为决策提供数据支持。