

链家二手房数据统计分析报告

田赐惠

目录

1 数据介绍	1
2 数据概览	2
3 探索性分析	5
3.1 变量 1 的数值描述与图形	5
3.2 变量 2 的数值描述与图形	9

1 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。

表 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68
万科紫悦湾	光谷东	127.0	14613	3	2	86.91
东立国际	二七	75.0	15968	1	1	46.97
新都汇	光谷广场	188.0	15702	3	2	119.73
保利城一期	团结大道	182.0	17509	3	2	103.95
加州橘郡	庙山	122.0	10376	3	2	117.59
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19
保利上城东区	白沙洲	193.8	16336	3	2	118.75
石化大院	中南丁字桥	325.0	32631	4	1	99.15
阳光花园	杨汊湖	192.0	17403	3	2	110.75

2 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, 2, 1, 2, 2, 2~
## $ 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       <lgl> TRUE, FALSE, TRUE, TRUE, FALSE, FALSE, TRUE, TRUE, T~
## $ if_2y             <lgl> FALSE, TRUE, FALSE, TRUE, TRUE, TRUE, FALSE, TRUE, F~
## $ has_key           <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE~
## $ vr               <lgl> FALSE, TRUE, FALSE, FALSE, TRUE, FALSE, TRUE, FALSE,~
```

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
## 精装:1757    Min.      : 2.00    低      : 816    板楼      :1781
## 简装: 634    1st Qu.:11.00    中      :1218    板塔结合: 615
## 毛坯: 436    Median :27.00    高      : 906    平房      : 5
```

```
## 其他: 173   Mean   :24.22   NA's: 60   塔楼   : 527
##           3rd Qu.:33.00           NA's   : 72
##           Max.   :62.00
##
## followers      near_subway      if_2y      has_key
## Min.   : 0.000   Mode :logical   Mode :logical   Mode :logical
## 1st Qu.: 1.000   FALSE:1446     FALSE:1736     FALSE:475
## Median : 3.000   TRUE :1554     TRUE :1264     TRUE :2525
## Mean    : 6.614
## 3rd Qu.: 6.000
## Max.    :262.000
##
## vr
## Mode :logical
## FALSE:916
## TRUE :2084
##
##
##
##
```

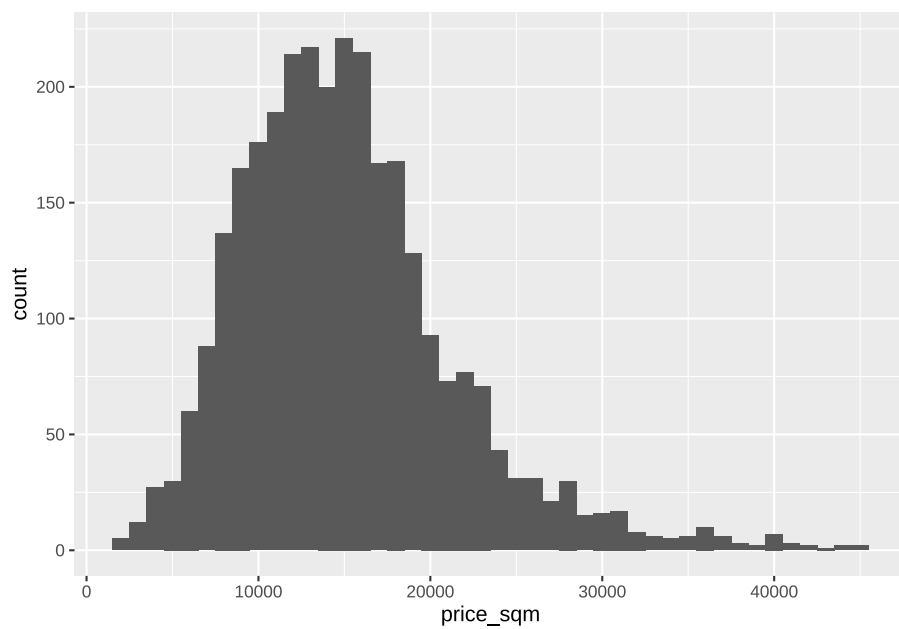
可以看到:

- 直观结论 1 从数量上来看, 白沙洲、盘龙城、四新等三环以及三环外区域的二手房套数最多
- 直观结论 2 每平方米价格极差非常大, 超过 40000 元, 但平均数与中位数都在 15000 元附近, 且集中在 10799 元至 18211 元范围内。这个价位是最为普遍的。
- 精装修房屋占大多数 -地铁覆盖了超过一半的二手房

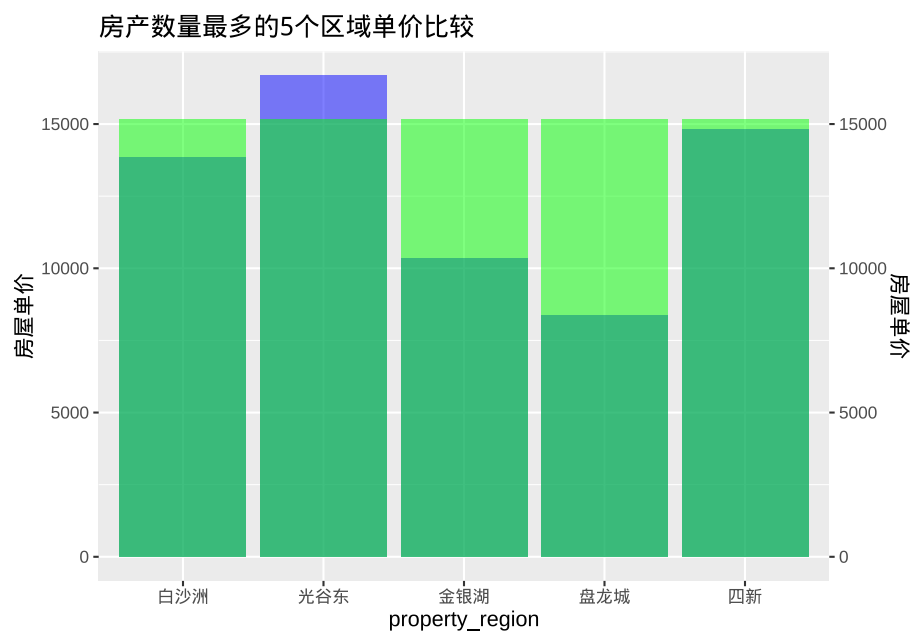
3 探索性分析

3.1 变量 1 的数值描述与图形

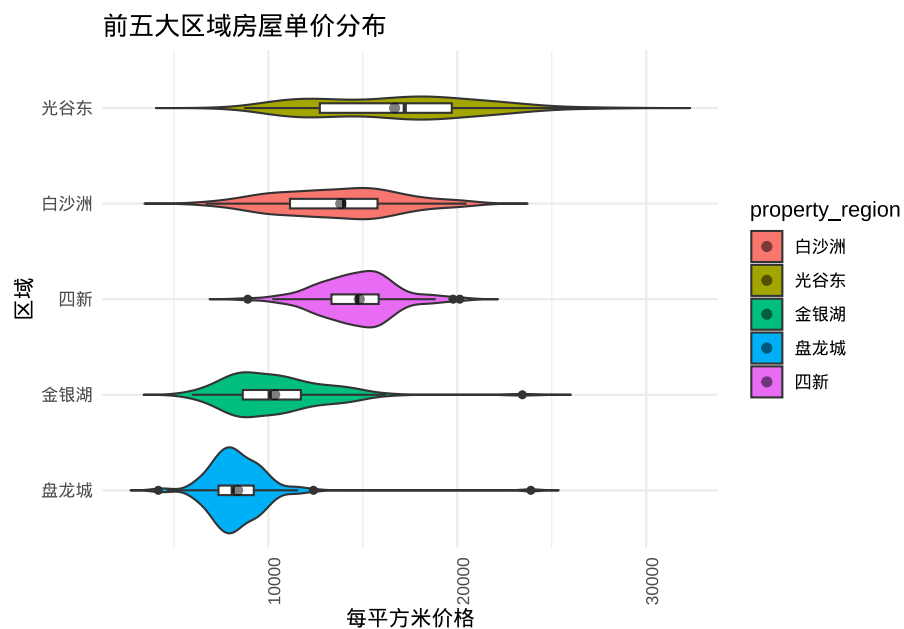
房屋单价整体呈现右偏的形态，一半房屋单价在 14404 元以下，一半房屋单价处于 14404 元至 44656 元区间



房产数量最多的 5 个区域与总体房价的比较，除了光谷东房价高于均值外，其他热门区域房价均低于全市均价，特别是盘龙城和金银湖，均价低于 11000 元。



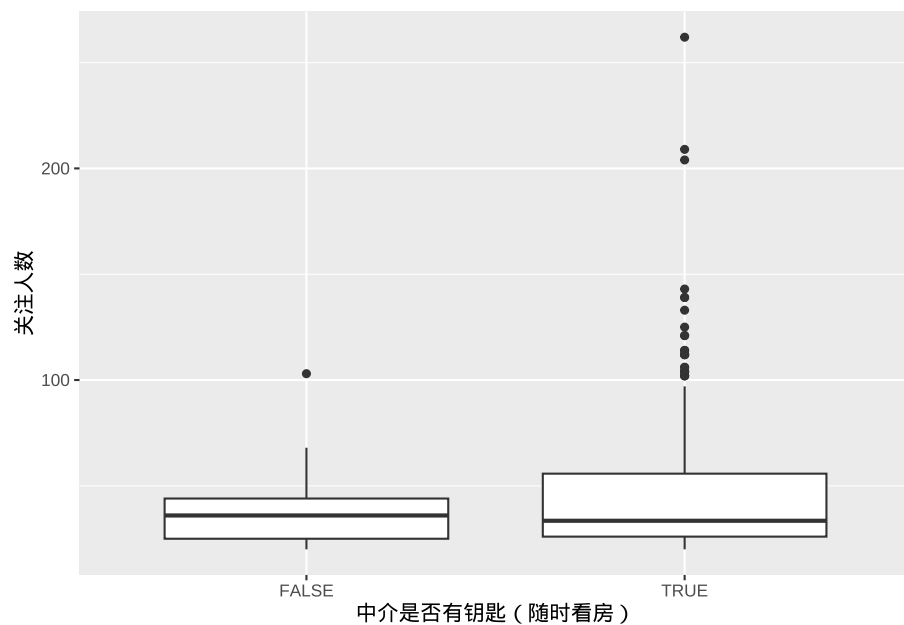
根据区域单价分布，发现光谷东与白沙洲房价分布非常分散，价格高低相差很大。而盘龙城和四新价格则较为集中。

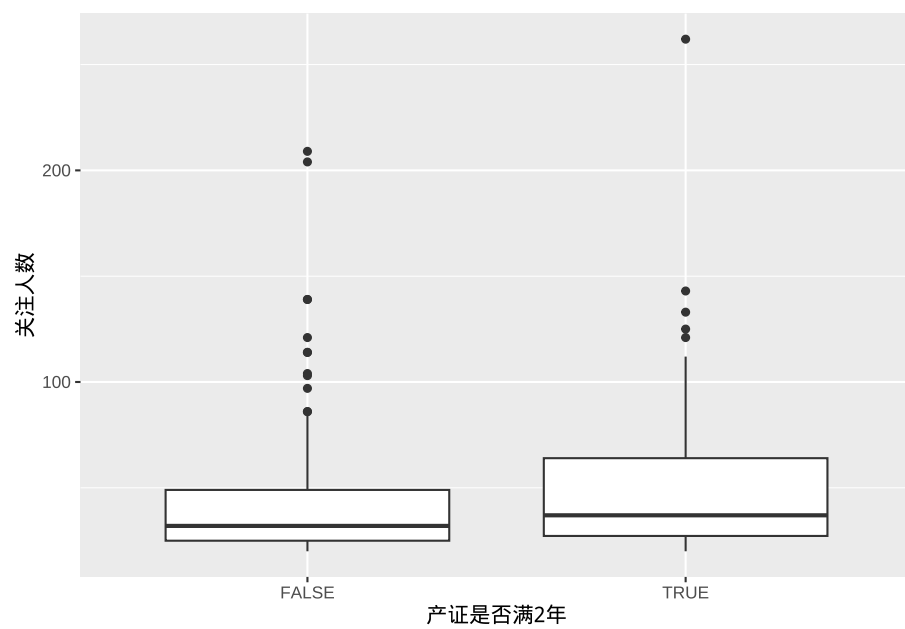
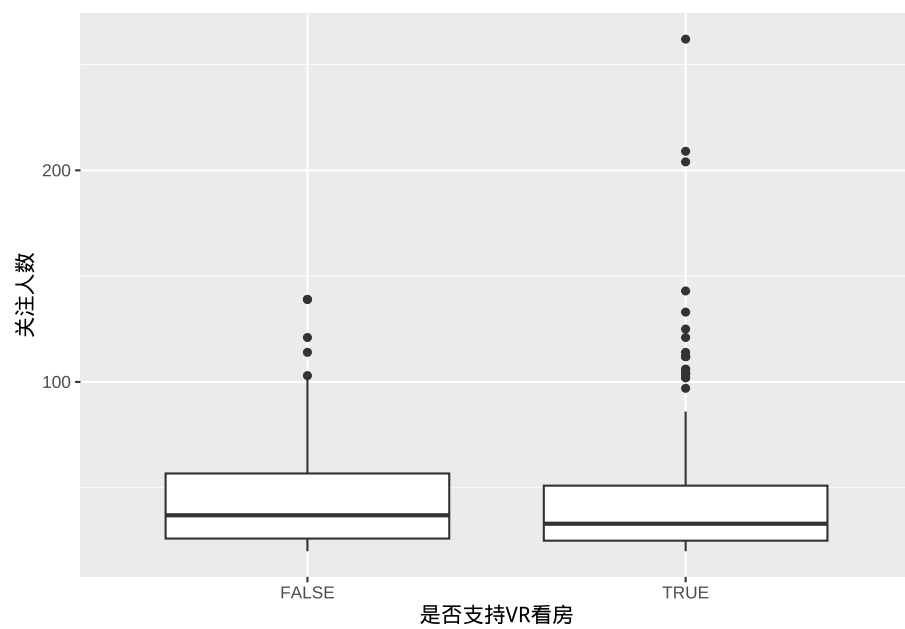


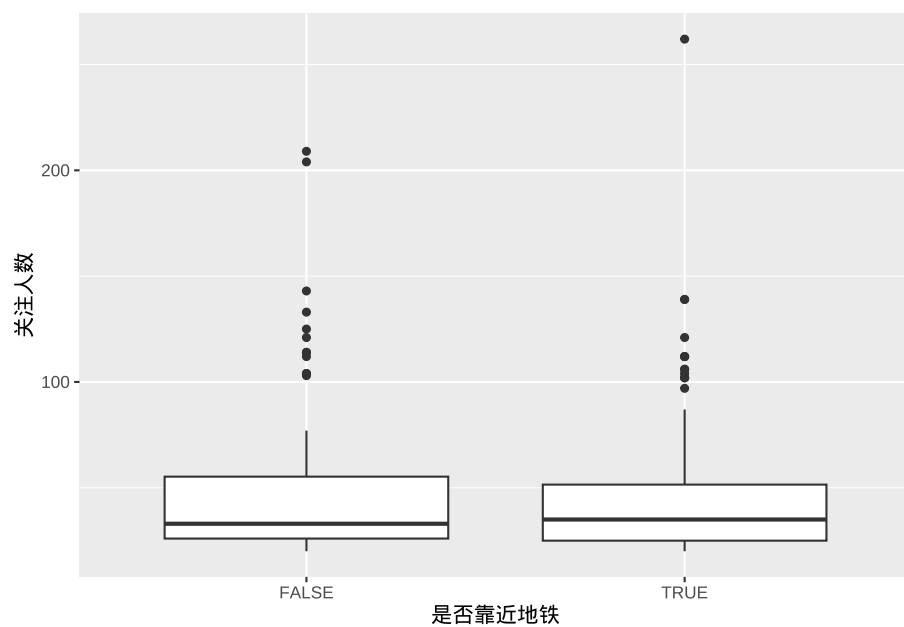
统计关注人数排名前 500 的房屋，光谷东的二手房受到最多的关注，这可能也是该区域房价较高的原因之一。

```
## # A tibble: 5 x 2
##   property_region count
##   <fct>           <int>
## 1 白沙洲             19
## 2 光谷东             33
## 3 积玉桥             22
## 4 金银湖             18
## 5 团结大道          21
```

尝试研究关注人数是否与临近地铁、产证满 2 年、随时看房、VR 看房这几个要素有关。发现很难找到明显的影响。



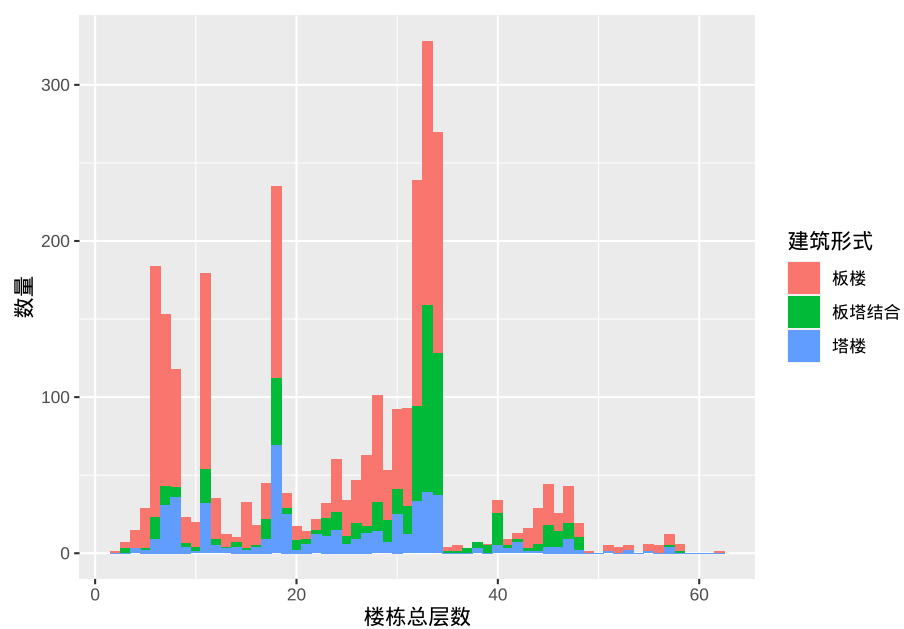




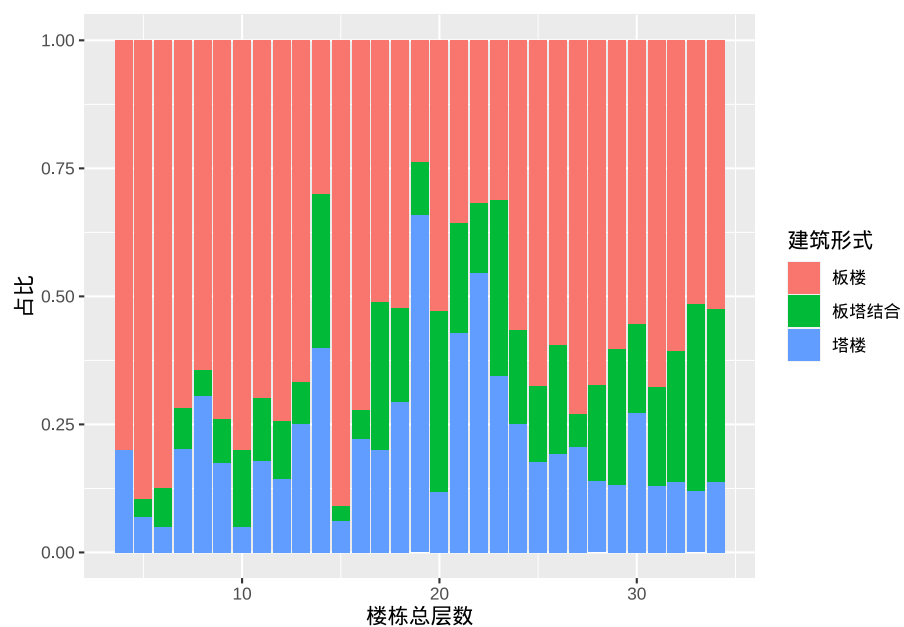
- 发现 1
- 发现 2

3.2 变量 2 的数值描述与图形

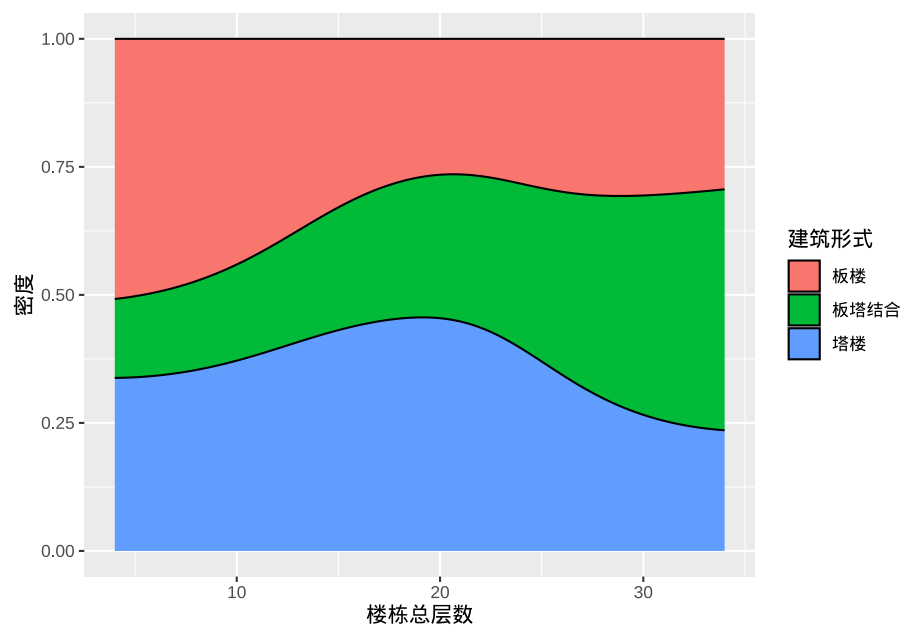
通过直方图观察楼栋总层数的分布，发现很多住宅楼高度为 32 至 34 层，18 层，11 层，6 至 8 层。



在板楼，塔楼，板塔结合三种建筑形式中，占比最高的是板楼。这里为了避免干扰，选取了总层数为 4 至 34 层的数据。



观察建筑形式在不同层高的变化趋势，发现随着楼层增高，板楼这种形式的占比减小直至平稳，板塔结合占比增大。



查看二手房的装修状况，分为精装，简装，毛坯，其他四类。精装房的平均单价明显高于另外三类，简装与毛坯均价差异不大。

```
## # A tibble: 4 x 3
##   decoration  计数  平均单价
##   <fct>      <int>    <dbl>
## 1 精装      1757    16077.
## 2 简装       634    13993.
## 3 毛坯       436    13819.
## 4 其他       173    13304.
```

分别画出价格分布图，观察到毛坯房较简装房，价格分布更加分散。

