

title: “第一次作业” CJKmainfont: Songti SC author: “刘琪” date: “2024-11-01” output: pdf_document: latex_engine: xelatex word_document: toc: yes
html_document: code_folding: show fig_caption: yes fig_width: 10 highlight: tango number_sections: yes theme: cosmo toc: yes —

你的主要发现

- 1. 发现1 中商宿舍房价最高。
- 2. 发现2 阳逻地区平均房价最低。
- 3. 发现3 绝大部分的房子朝南。

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。

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	prop
南湖名都A区 南湖沃尔	237.									

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, 包含小区名称、小区位置、房产总价、房子每平方单价等18个变量。
- 直观结论2 武汉二手房单价均值为15148, 中位数为14404, 最高价为44656; 武汉二手房总价均值为155.9万, 中位数为137.0万, 最高价为1380.0万;

探索性分析

#1、分析各区域的房产均价

```
## # A tibble: 87 × 2
##   property_region avg_price_sqm
##   <chr>          <dbl>
## 1 VR看装修      38351
## 2 中北路        32728.
## 3 水果湖        28562.
## 4 黄埔永清      24957.
## 5 三阳路        24777.
## 6 南湖沃尔玛    24181.
## 7 虎泉杨家湾    23902.
## 8 CBD西北湖     22272.
## 9 楚河汉街      21958.
## 10 关山大道      21480.
## #   77 more rows
```

发现:

-中北路房子的平均每平米房价最高。-阳逻房子的平均每平米房价最低。

#2、分析各区域的房产房产平均面积

```
## # A tibble: 87 × 2
##   property_region avg_building_area
##   <chr>          <dbl>
## 1 汉口北        125.
## 2 中北路        125.
## 3 VR看装修      125.
## 4 中法生态城    122.
## 5 楚河汉街      121.
## 6 卓刀泉        120.
## 7 国际百纳      118.
## 8 蔡甸其它      117.
## 9 CBD西北湖     117.
## 10 黄家湖        117.
## #   77 more rows
```

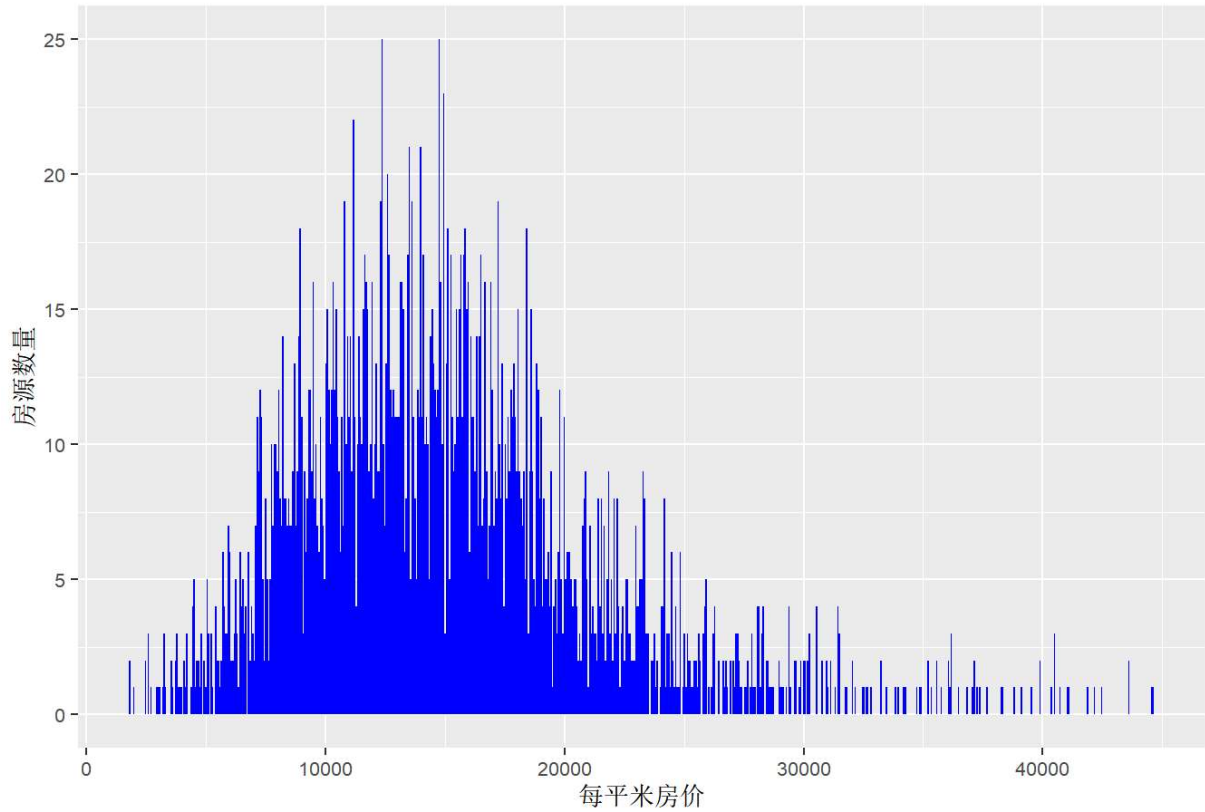
发现:

-汉口北房子的平均面积最大。

-古田房子的平均面积最小。

##3、每平米房价分析

每平米房价直方图

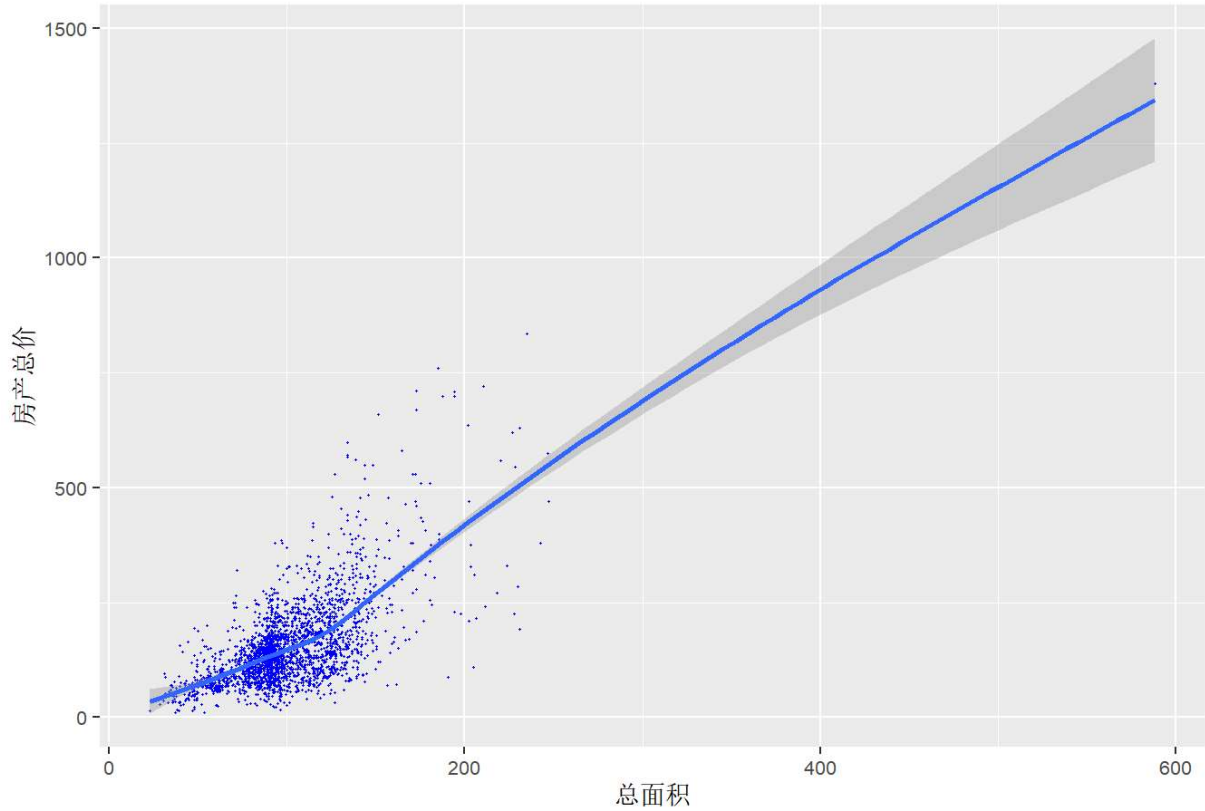


发现:

- 每平米房价直方图，房价的分布不均匀，存在许多离散的尖峰，这可能表明数据中有许多异常值或极端值。
- 大部分的房价集中在较低的区间，随着价格的增加，房源数量逐渐减少。这种分布可能表明市场上大多数房产的价格相对较低，而高价位的房产数量相对较少。

##4、房产总价与房产总面积直接的关系

房产总价与总面积的关系

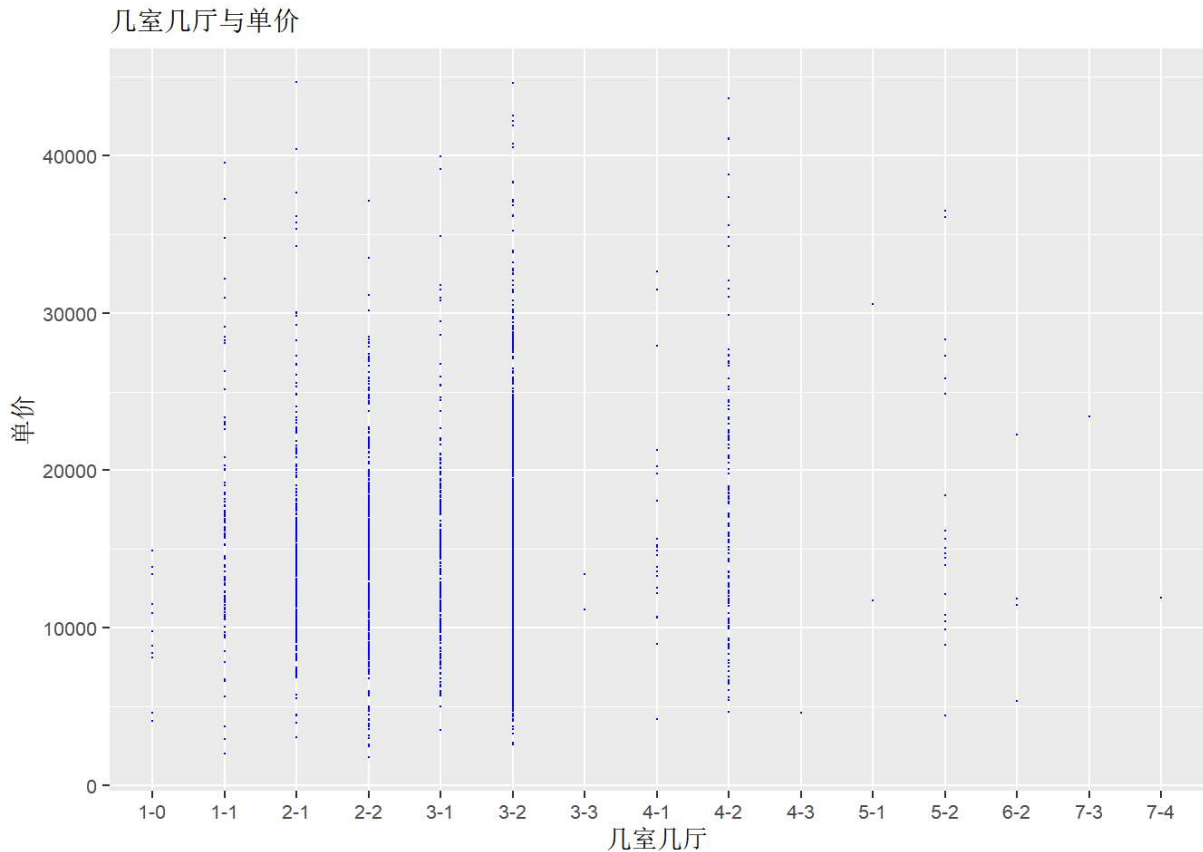
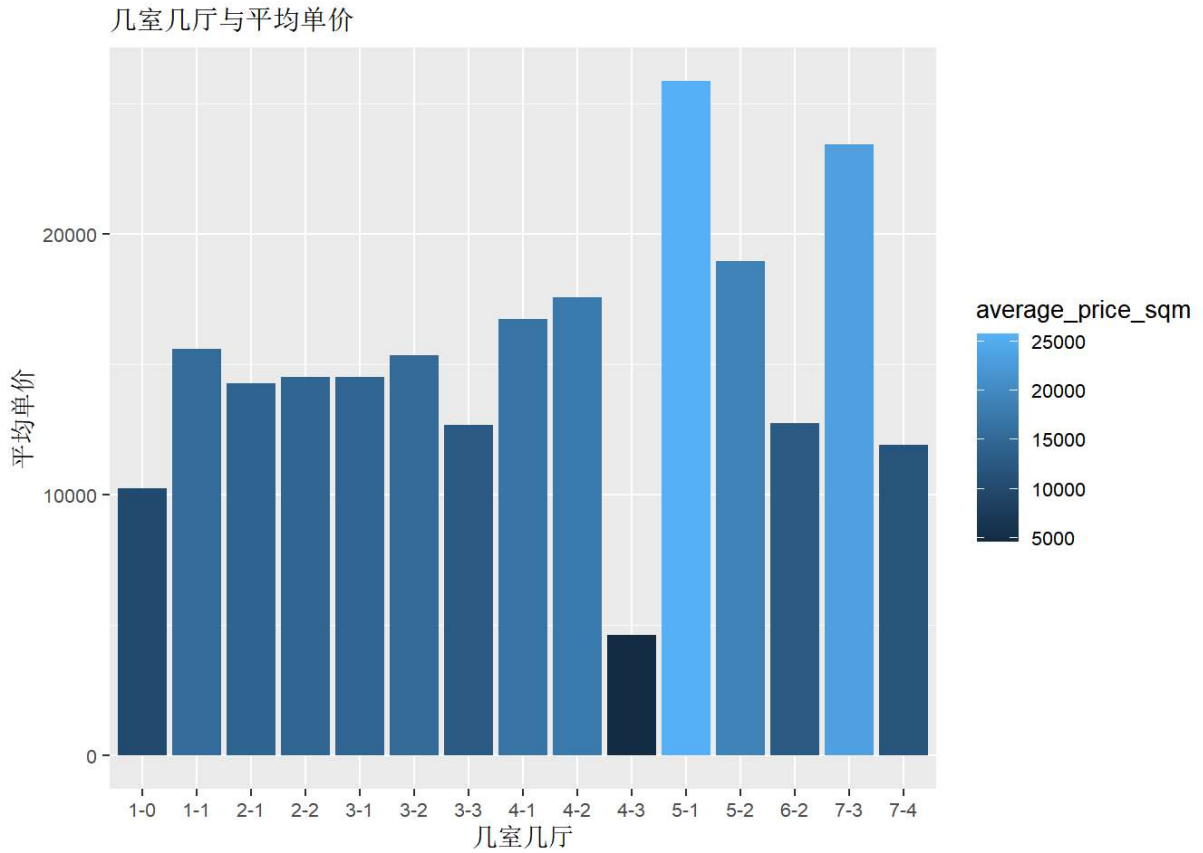


发现:

- 从图中可以看出，房产总价随着总面积的增加而增加，这表明两者之间存在正相关关系。图中还绘制了一条蓝色的回归线，表示房产总价与总面积之间的线性趋势。回归线周围的灰色区域表示置信区间，反映了回归线的不确定性。

变量...的数值描述与图形

探索问题1、卧室数量与起居室数量（几室几厅）与房产平均单价之间的关系？

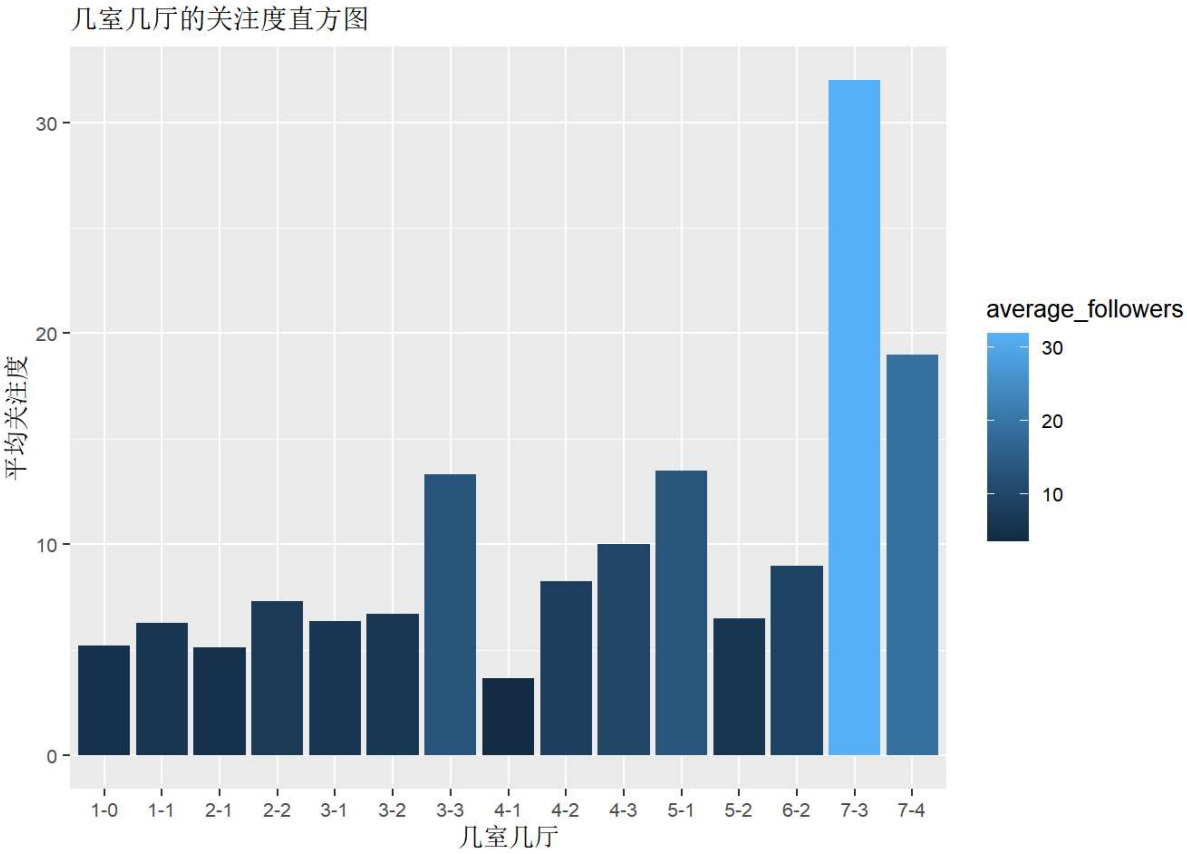
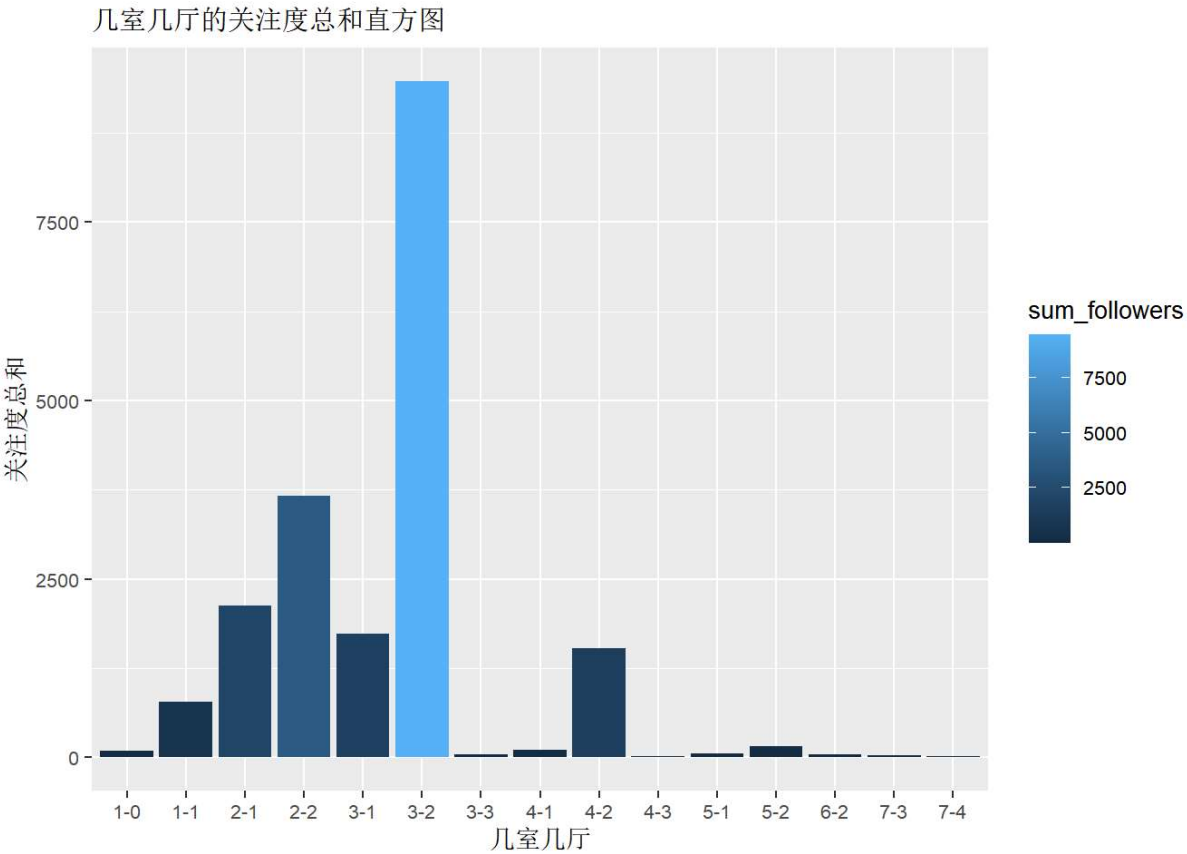


发现:

- 从直方图可以看出，不同户型组合的平均单价有显著差异。其中五室一厅户型的平均单价最高，而四室三厅户型的平均单价最低。这表明在市场上，五室一厅户型更昂贵或更稀缺，因此单价更高。

- 从散点可以看出，不同户型组合的单价分布范围很广。其中三室两厅户型的单价分布最为集中，而其他户型三室三厅及以上的单价分布则较为分散。这表明在市场上，某些特定户型的单价可能更加标准化或统一，而其他户型的单价则可能因位置、条件或其他因素而有较大差异。

探索问题2、卧室数量与起居室数量（几室几厅）哪种组合更受关注？



发现：

- 不同户型的关注度有显著差异。例如，七室三厅户型的平均关注度最高，超过了30，而四室一厅户型的关注度最低。这表明在市场上，七室三厅户型可能更受欢迎或更稀缺，因此吸引了更多的关注。
- 三室两厅户型的关注度总和最高，接近8000，而其他户型的关注度总和相对较低。这表明在市场上，三室两厅户型可能更受欢迎或更常见，因此吸引了更多的关注。

发现总结

用1-3段话总结你的发现。 1、三室两厅户型的单价分布最为集中，且三室两厅户型的关注度总和最高，说明三室两厅为市面上的主流户型，价格也相对统一。 2、大部分买房者会选择三室两厅等主流户型，说明了三室两厅的性价比较高，其拥有适中的卧室数量和客厅数量，能满足正常的居住需求且价格偏差小，合适的面积也保证了总价不会太高，所以这类户型受到欢迎。 2、市场房源众多，价格受地段、户型等多方面影响。买房群体同样会考虑自身经济状况与家庭结构。整个市场里对于中等规模的户型有较高的需求，表明买房群体更倾向于购买舒适实用的房子。