

第一次作业你的报告题目

Code ▾

徐颖

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...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

- 1. 发现1
- 2. 发现2
- 3. 发现3

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	prop
南湖名都A区 南湖沃尔		237.								

各变量的简短信息：###也可以用str(lj)

```
## 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 不同的区域，房价不一样
- 直观结论2 同一个区域，楼层对价格的影响 热门区域
- ...

4 探索性分析

4.1 变量1的数值描述与图形

property_region的数值描述和图形

发现：

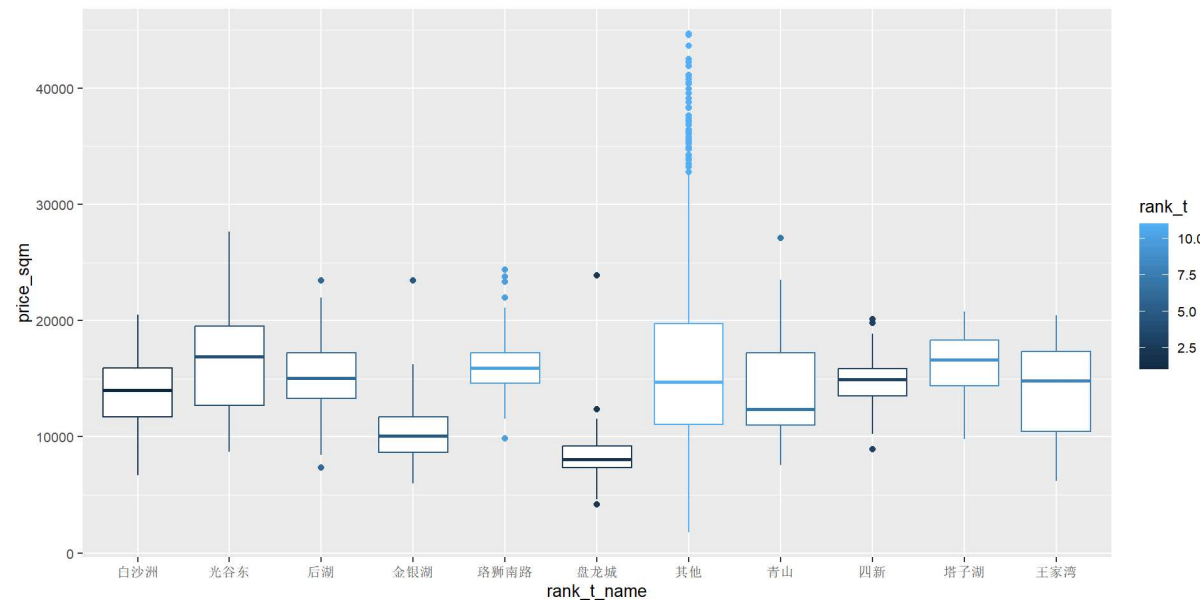
-发现1：

白沙洲是当前二手房源最多的区域，其次是四新/光谷东/盘龙城/金银湖，说明这些区域的二手房市场较活跃。

```
## [1] "白沙洲"
```



- 发现2：房源较多的区域如白沙洲/金银湖/盘龙城/四新的价格会比其他地区的低，且价格波动更小；



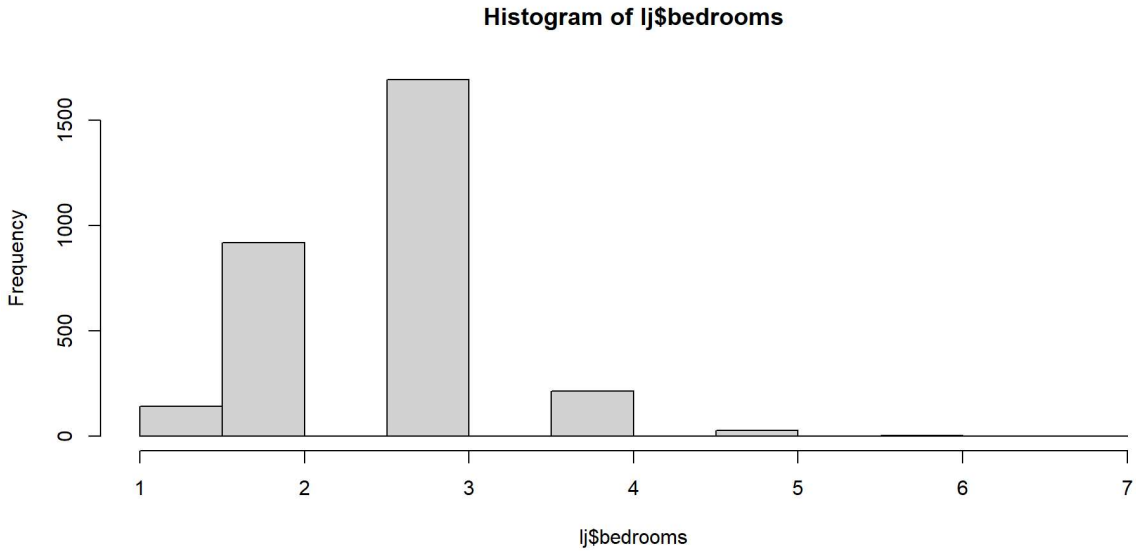
4.2 变量2的数值描述与图形

bedrooms的数值描述与图形

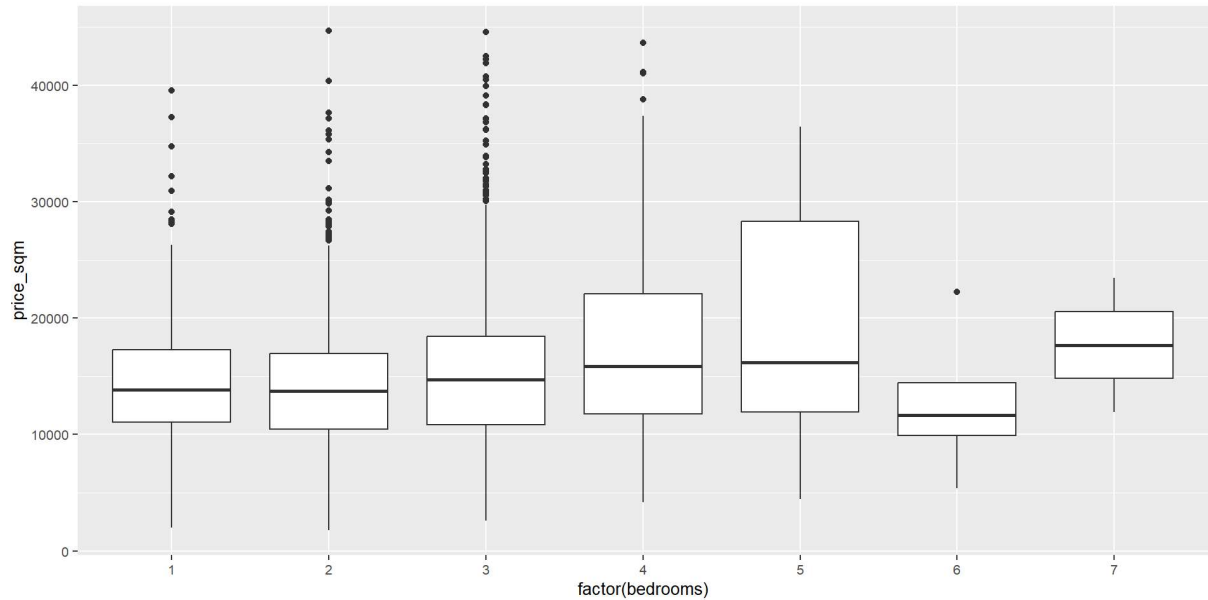
发现：

-发现1：

市面上流通的房子以2-3个房间数为主，3个房间的房子最多，其次是2个房间的，其他房间数量的房子比较少。



- 发现2:
房间数对价格的影响不大，但是房间数越多，房价的分散程度越大，比如说5个房间的房间波动是最大的，5个房间里面可能包含部分门高档住宅或别墅之类的，使价格波动较大。



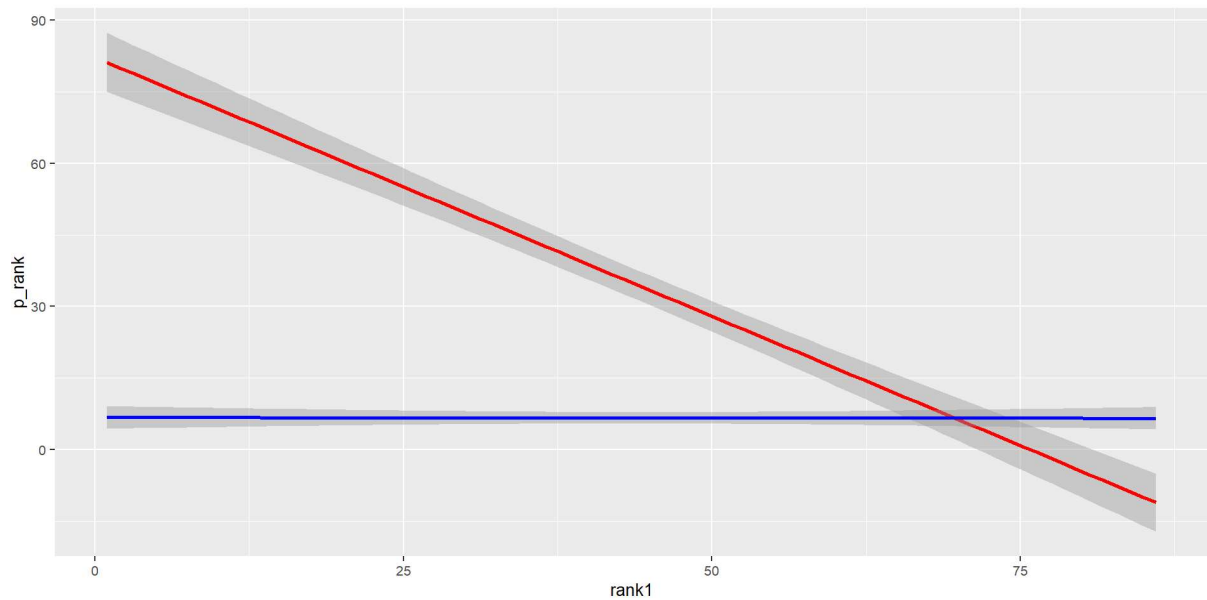
4.3 变量...的数值描述与图形

property_region和followers的数值描述和图形

发现:

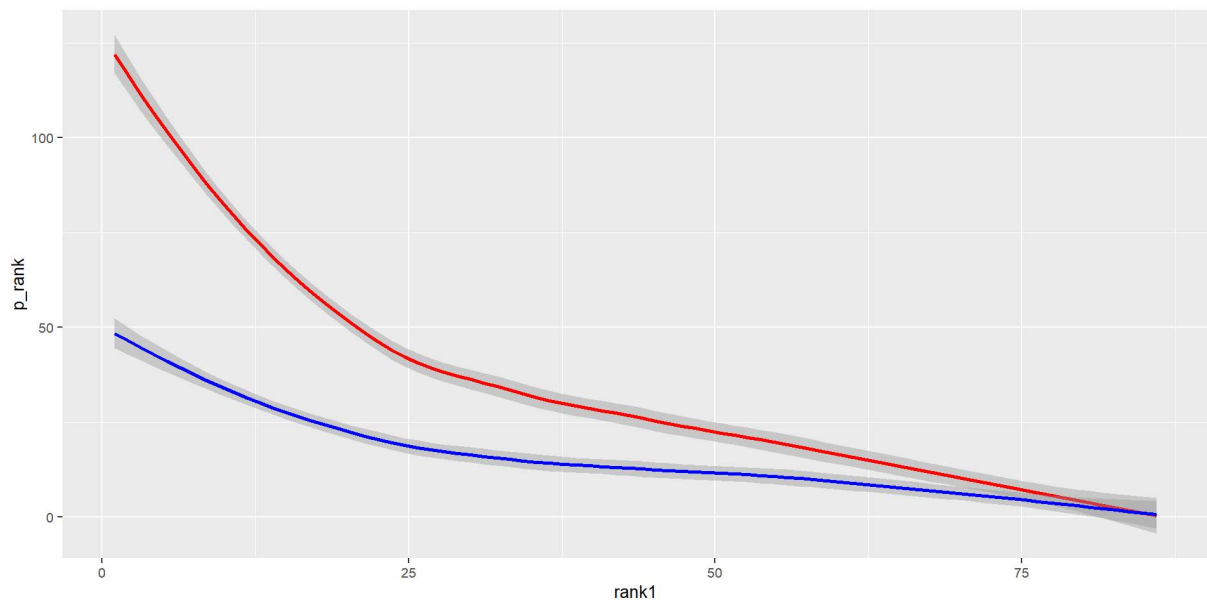
-发现1:

各区域的关注人数较稳定，房源越多并没有更多的关注人数，尤其是白沙洲/金银湖/盘龙城/四新等房源较多的区域，可能会出现房源积压的情况；而房源较少的区域，可适当增投房源数。



- 发现2:

上架的房源数与区域的小区数呈现正相关，比如白沙洲的小区数量就是最多的，但是随着小区数量的减少，上架的房源数急剧下降，房源数的降幅高于小区数的降幅。

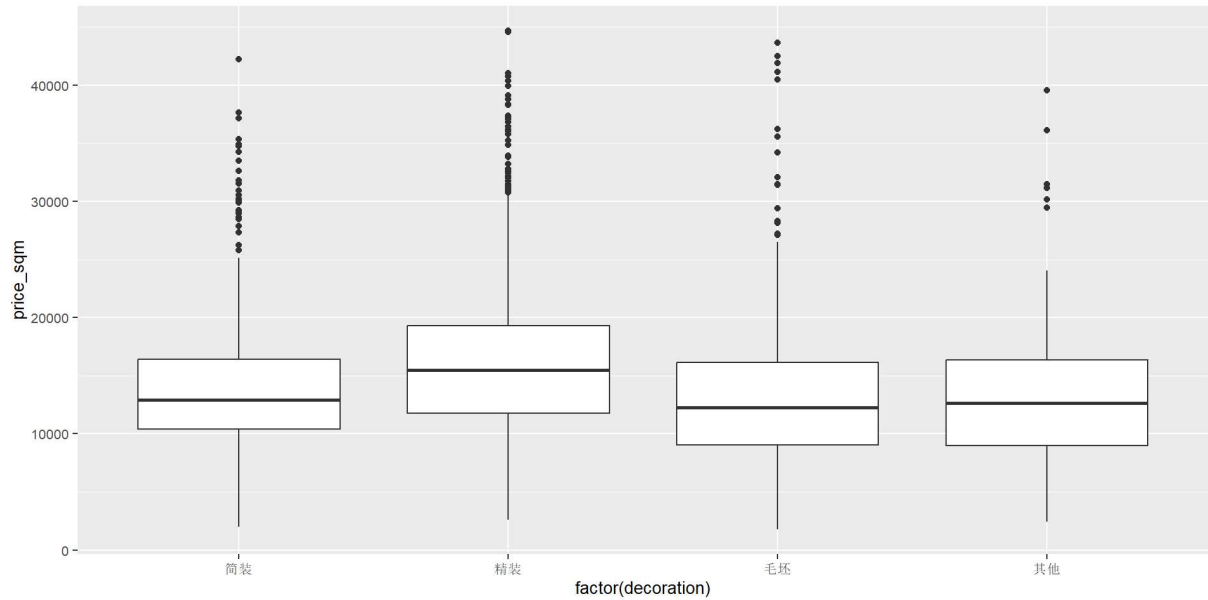


4.4 探索问题1

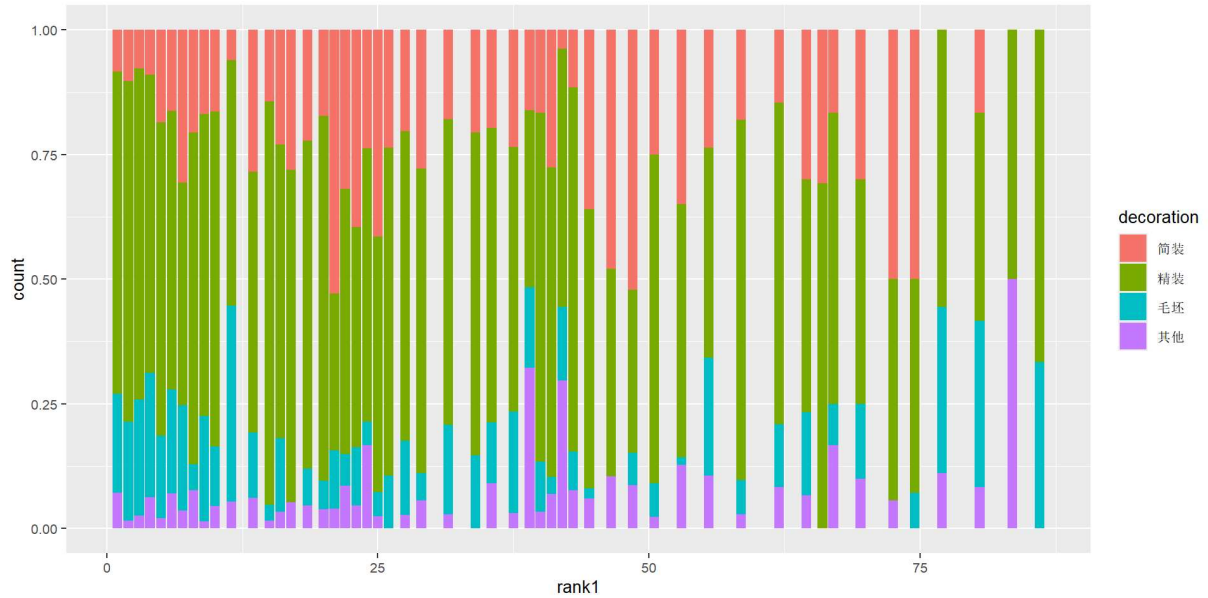
装修情况对房价的影响？

发现1:

二手房中，精装房的房价会比其他几种装修的房价高，且精装房的房价波动较大，说明装修好坏是影响房价的一个因素。



- 发现2:
从各区域的二手房装修情况来看，精装房占比最高，其次是简装房，毛坯房较少。



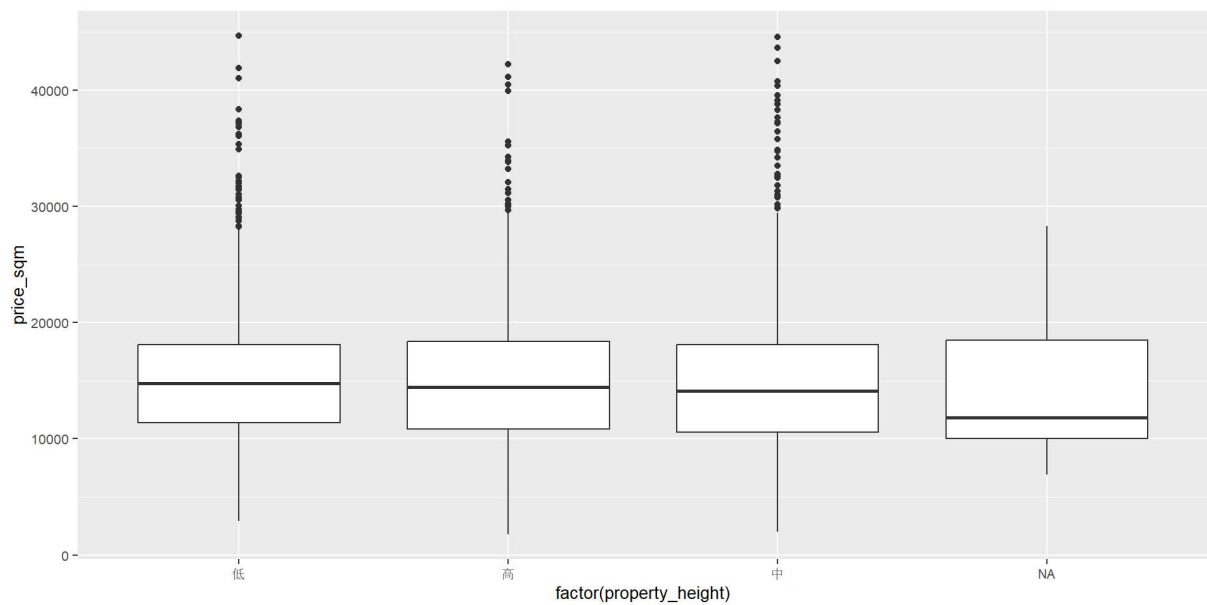
4.5 探索问题2

房屋在所在楼栋所处的位置对房价的影响？

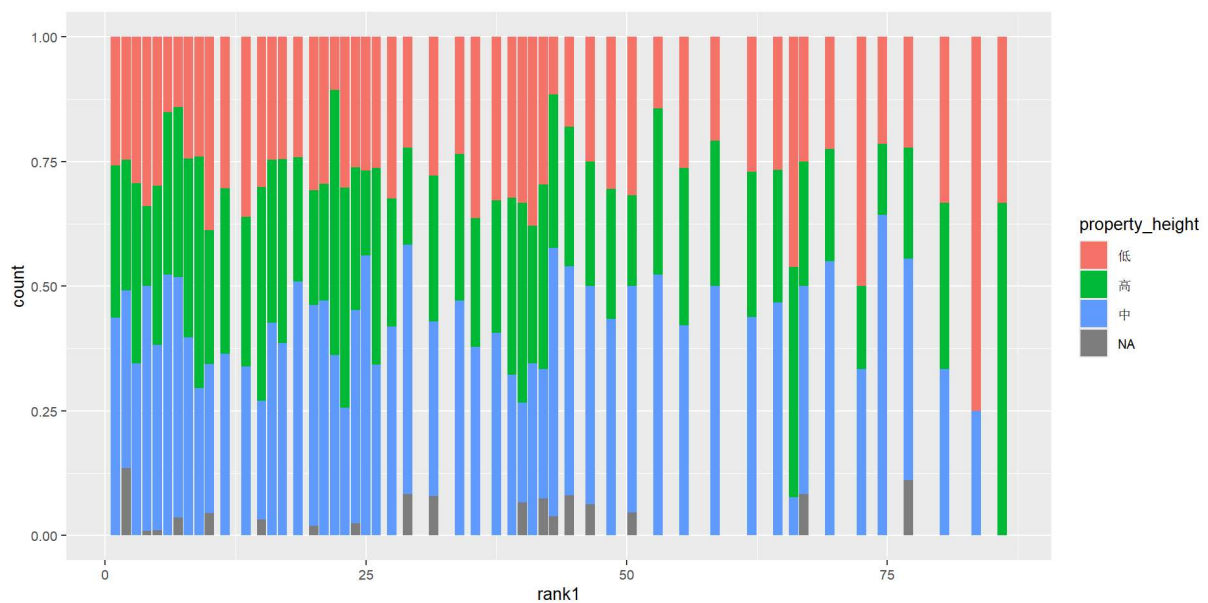
发现：

- 发现1：

房屋在所在楼栋中所处的位置对房价的影响较小，几乎不影响。



- 发现2:
各楼层占比分布相对较均匀, 中层房屋占比略多。



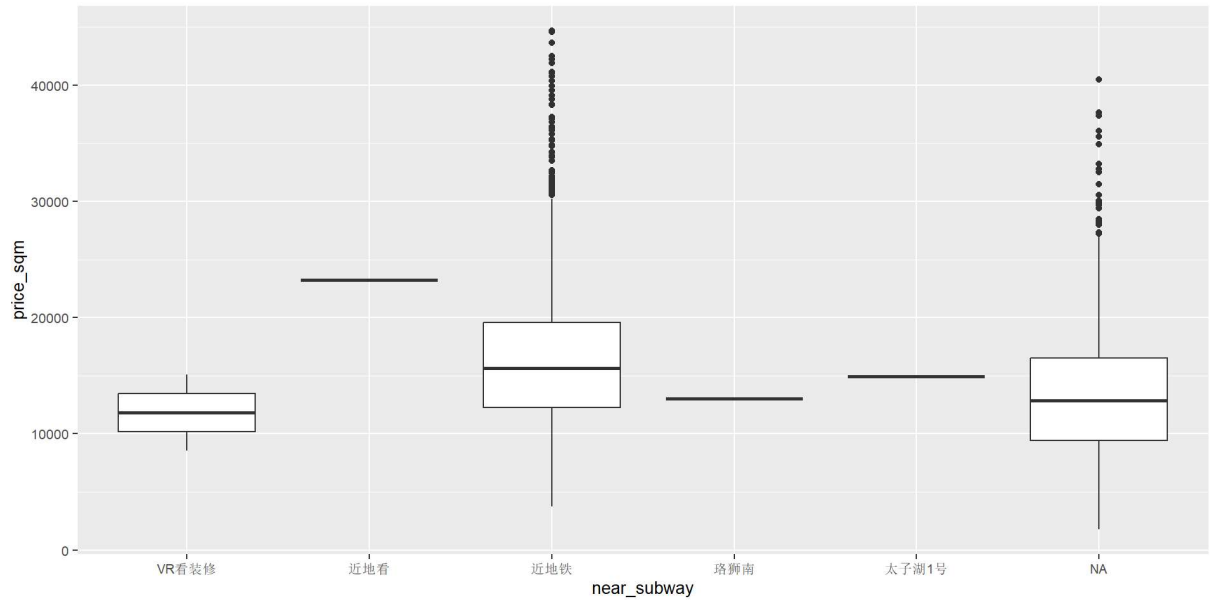
4.6 探索问题3

地铁对房价的影响

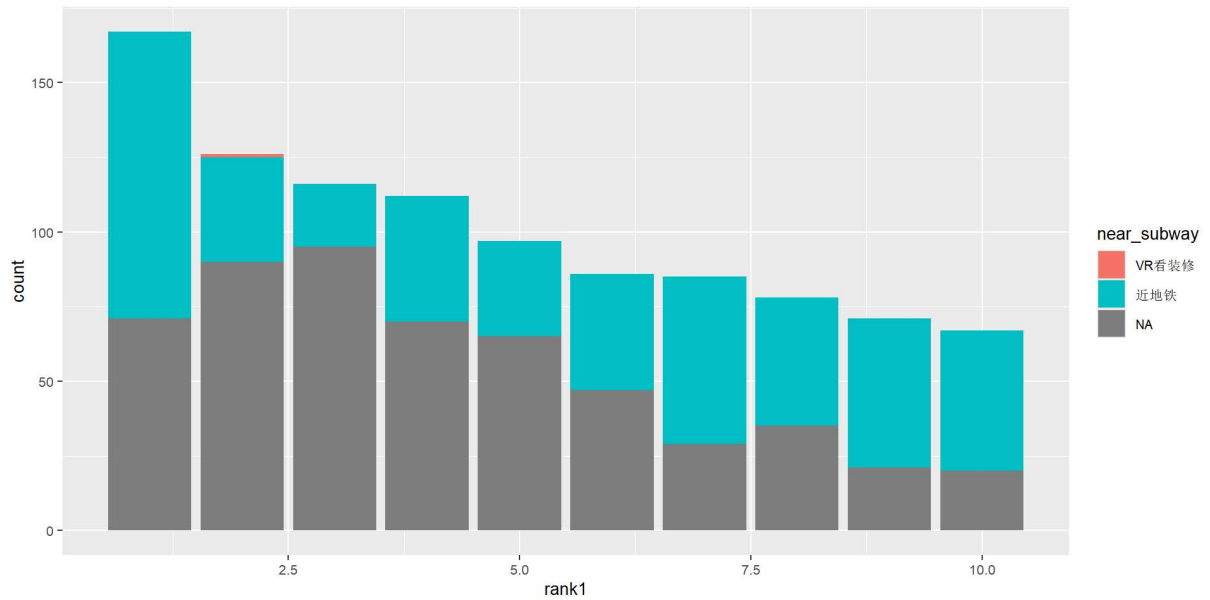
发现:

- 发现1:

靠近地铁的房价比不靠近地铁的房价高, 说明地铁是利好房价的一个因素



- 发现2:
二手房源数量top10的区域中，不靠近地铁的房屋会偏多，因此也验证了前面房源数量top的区域房价会偏低一些。



5 发现总结

用1-3段话总结你的发现。1.房屋数量多的区域，覆盖的小区也偏多，这些地方的地铁偏少，相对没有那么便利，价格波动也较小；2.影响房价的因素有：区域、装修情况、是否靠近地铁等因素；而楼层、房间数量等因素对房价的影响较小；