

定量分析第一次作业

刘杨

2023-10-20

你的主要发现

1. 发现 1：建筑形式影响房屋单价。
2. 发现 2：不同区域间房屋单价相差较大。
3. 发现 3：人们似乎更关注靠近地铁的房子。

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：

各变量的简短信息：

Rows: 3,000

Columns: 18

\$ 小区名字 <chr> "南湖名都A区", "万科紫悦湾", "东立国际", "新都汇", "保利城~

\$ 所处区域 <chr> "南湖沃尔玛", "光谷东", "二七", "光谷广场", "团结大道", "庙~

\$ 房屋总价 <dbl> 237.0, 127.0, 75.0, 188.0, 182.0, 122.0, 99.0, 193.8, 325.0~

\$ 房屋单价 <dbl> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 16336, 326~

表 1: 武汉链家二手房

小区名字	所处区域	房屋总价	房屋单价	房间数	客厅数	建筑面积	主要朝向	次要朝向	装修状况	总层数	所在楼层	建筑形式	关注人数	靠近地铁
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68	南	北	精装	17	中	塔楼	3	近地铁
万科紫悦湾	光谷东	127.0	14613	3	2	86.91	南	NA	精装	28	中	板楼	1	NA
东立国际	二七	75.0	15968	1	1	46.97	南	NA	简装	18	低	塔楼	3	近地铁
新都汇	光谷广场	188.0	15702	3	2	119.73	北	东	精装	32	高	塔楼	2	近地铁
保利城一期	团结大道	182.0	17509	3	2	103.95	东南	NA	简装	34	中	板塔结合	3	NA
加州橘郡	庙山	122.0	10376	3	2	117.59	南	北	精装	34	低	板楼	1	NA
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19	南	NA	简装	7	低	板楼	0	近地铁
保利上城东区	白沙洲	193.8	16336	3	2	118.64	南	北	其他	34	中	板塔结合	0	近地铁
石化大院	中南丁字桥	325.0	32631	4	1	99.60	南	北	简装	5	低	板楼	2	近地铁
阳光花园	杨汊湖	192.0	17403	3	2	110.33	南	北	其他	7	低	板楼	0	近地铁

```
## $ 房间数      <dbl> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4, 3, 3, ~
## $ 客厅数      <dbl> 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, ~
## $ 建筑面积    <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, 118.64~
## $ 主要朝向    <chr> "南", "南", "南", "北", "东南", "南", "南", "南", "南", "南", "南~
## $ 次要朝向    <chr> "北", NA, NA, "东", NA, "北", NA, "北", "北", "北", "北", N~
## $ 装修状况    <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简装", "其~
## $ 总层数      <dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, 44, 19, ~
## $ 所在楼层    <chr> "中", "中", "低", "高", "中", "低", "低", "中", "低", "低", "~
## $ 建筑形式    <chr> "塔楼", "板楼", "塔楼", "塔楼", "板塔结合", "板楼", "板楼", ~
## $ 关注人数    <dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, 1, 1, 0~
## $ 靠近地铁    <chr> "近地铁", NA, "近地铁", "近地铁", NA, NA, "近地铁", "近地铁~
## $ 满2年       <chr> NA, "房本满两年", NA, "房本满两年", "房本满两年", "房本满两~
## $ 中介有钥匙  <chr> "随时看房", "随时看房", "随时看房", "随时看房", "随时看房", ~
## $ VR看房      <chr> NA, "VR看装修", NA, NA, "VR看装修", NA, "VR看装修", NA, NA, ~
```

各变量的简短统计:

```
##      小区名字      所处区域      房屋总价      房屋单价
## 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
##      房间数      客厅数      建筑面积      主要朝向
## 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
```

##	次要朝向	装修状况	总层数	所在楼层
##	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	
##	建筑形式	关注人数	靠近地铁	满2年
##	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		
##	中介有钥匙	VR看房		
##	Length:3000	Length:3000		
##	Class :character	Class :character		
##	Mode :character	Mode :character		
##				
##				
##				

可以看到:

- 直观结论 1

通过查看房屋单价的分布情况,最小值为 1771 元,最大值为 44656 元,说明房价波动较大;平均房价为 15148 元,中位数为 14404 元,平均值大于中位数,说明房价整体偏高,同时从分布形态来看整个数据可能呈右偏分布。此外,第三个四分位数为 18211 元,说明约 75% 的房屋单价在 18211 元以下。

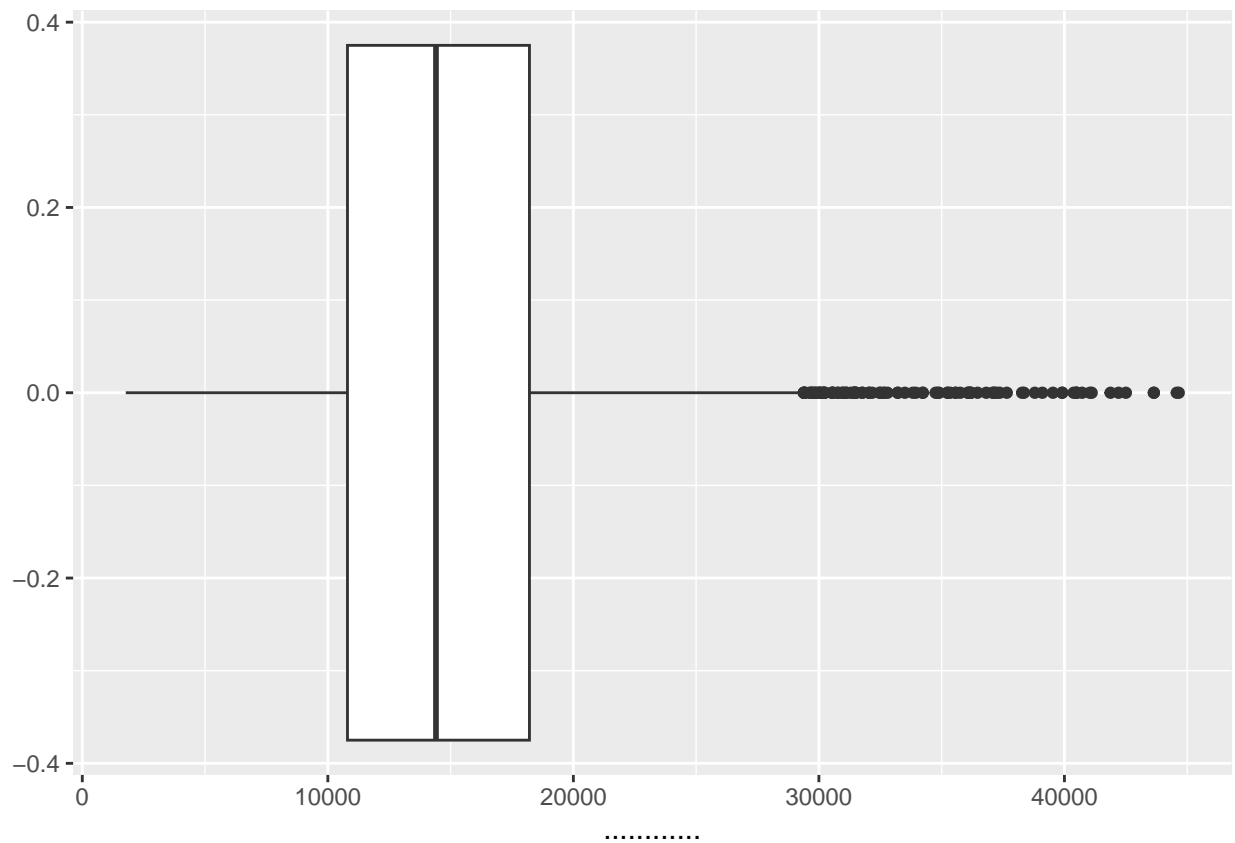
- 直观结论 2

大部分房间的数量为 2-3 间,客厅数为 1 间左右;大部分房屋的建筑面积在 95.55 至 100.87 平米之间;楼栋的总层数偏低。

探索性分析

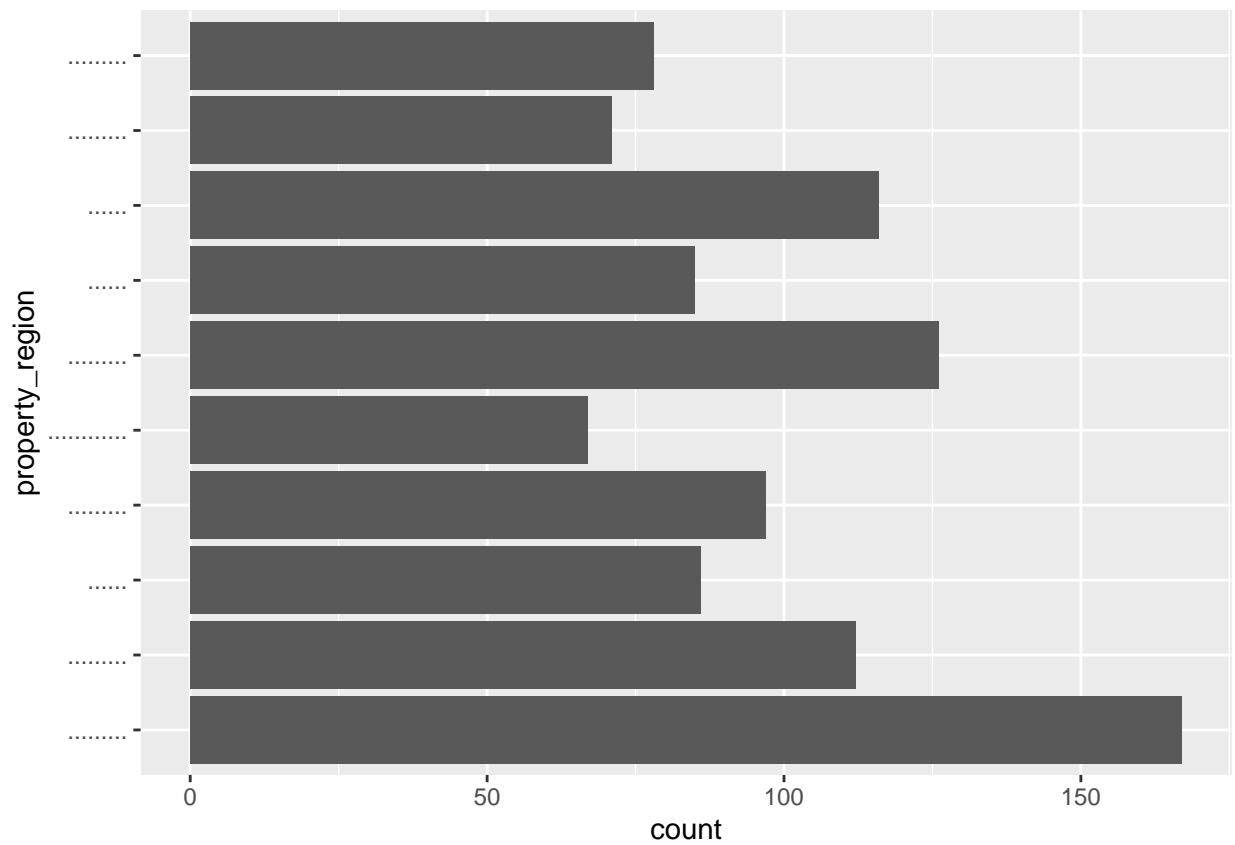
房屋单价的数值描述与图形

房屋单价最小值为 1771 元,最大值为 44656 元,平均房价为 15148 元,中位数是 14404 元,第一个四分位数为 10799 元,第三个四分位数为 18211 元,方差 6323。



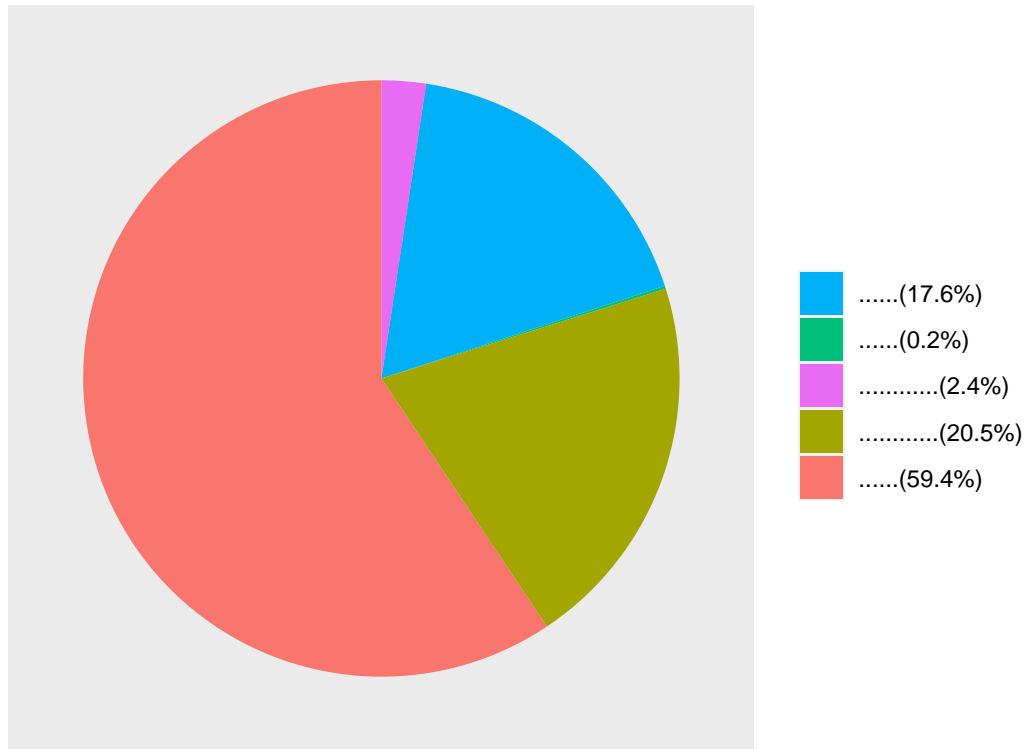
各区域的房屋数量统计描述与图形

我们筛选各区域房屋数量发现，房屋数量最多的区域在白沙洲附近，如下图展示排名靠前的十大区域。



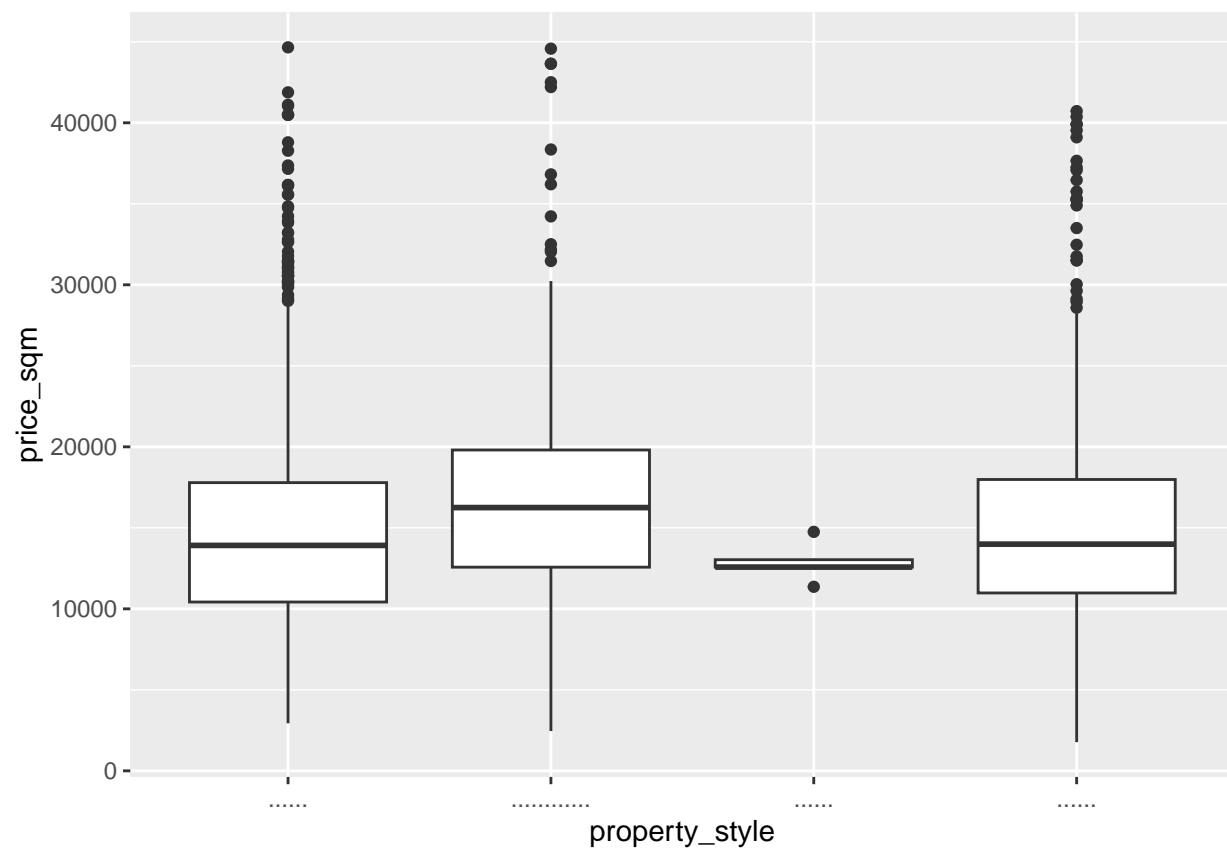
不同建筑形式的数值描述与图形

在 3000 个数据中，除 72 个房屋暂无相关数据外，我们发现数量最少的类型是平房只有 5 间，数量最多的类型是板房，共有 1781 个。各类占比如下图。



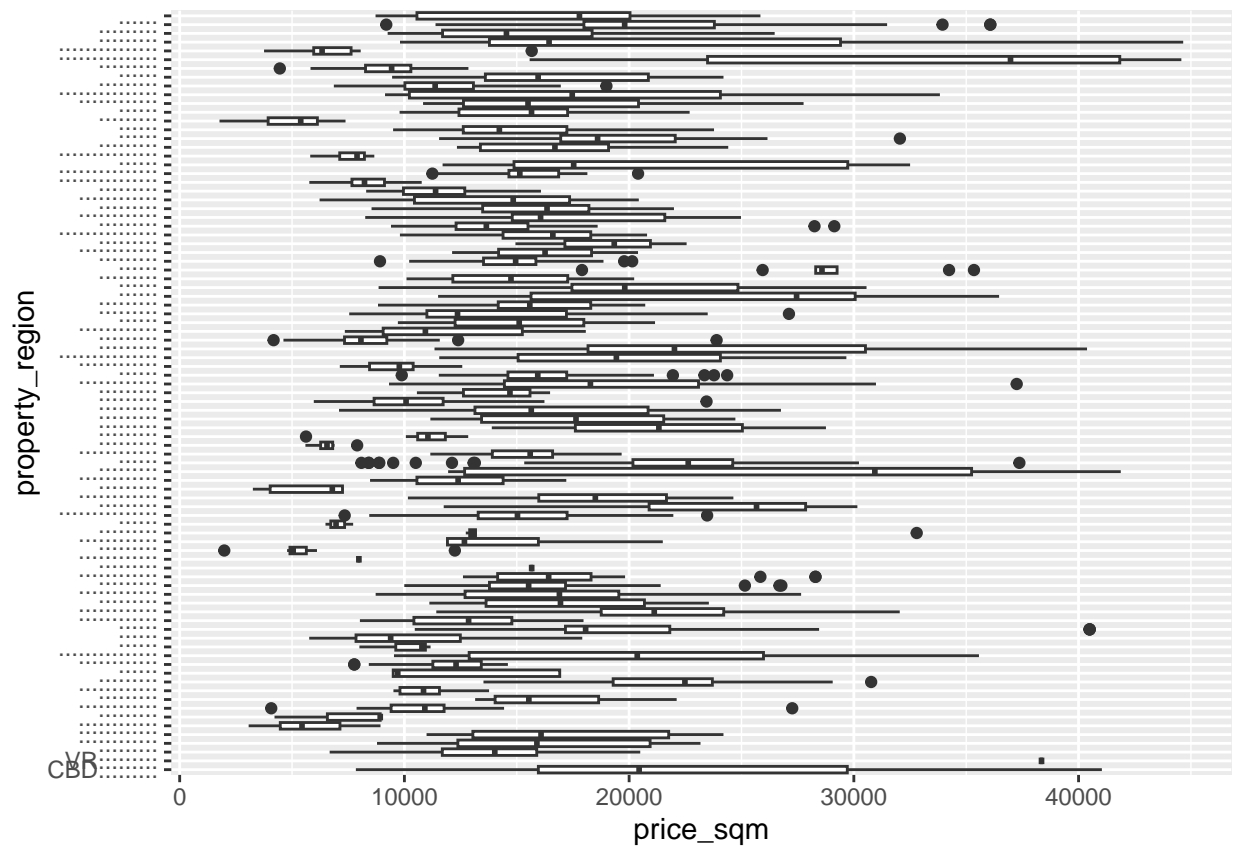
探索问题 1：房屋单价与建筑形式存在什么关系？

剔除房屋类型是“暂无数据”的行形成新表 `lj_02`，以该表为基础，画出建筑形式和房屋单价关系的箱线图。从大部分的数据来看，平房的单价最低，板塔结合的房屋单价最高，板楼和塔楼居于二者之间。

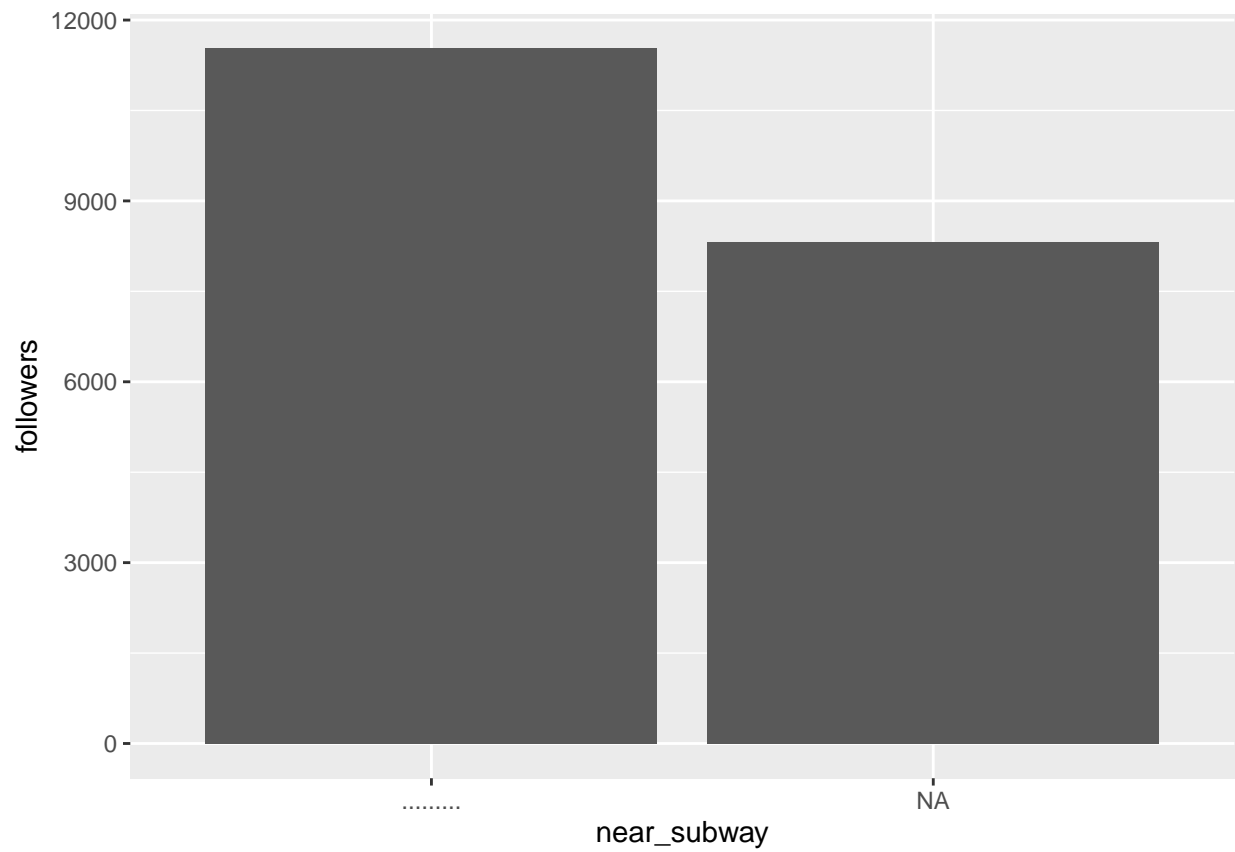


探索问题 2：房屋单价与区域存在什么关系？

不同区域间房屋单价相差较大。



探索问题 3：人们是否更关注房屋附近有地铁？



人们似乎更关注靠近地铁的房子。