

第一次作业你的报告题目

Code ▾

your name

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 房屋单价的数值描述与图形
 - 4.2 建筑面积的数值描述与图形
 - 4.3 房屋总价的数值描述与图形
 - 4.4 二手房建筑年限房屋数量分布
 - 4.5 房屋所处楼层对房屋单价的影响
 - 4.6 房屋所处区域对房屋单价的影响
 - 4.7 是否靠近地铁对房屋单价的影响
- 5 发现总结

1 你的主要发现

- 房屋单价、总价及面积分布均呈左偏趋势，偏度较大
- 房屋所属区域对房屋单价影响较大，所处楼层及受否靠近地铁影响较小
- 房屋建筑年限分布与国家楼市政策密切相关

2 数据介绍

本报告链家数据获取方式如下：

报告人在2024年10月27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_title, property_name, property_region, price_ttl, price_sqm, bedrooms, livingrooms, building_area, directions1, directions2, decoration, hml, building_height, building_year, building_style, followers, pub_days, near_subway, two_year, taxfree, haskey, vr, price_date, building_age等24个变量,共4700行。表的前10行示例如下：

武汉链家二手房

property_title	property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2	d
汉口湖畔不临										
马路，东南朝										
向，中间楼层										
汉口湖畔 盘										
龙城										

各变量的简短信息：

```
## Rows: 4,700
## Columns: 24
## $ property_title <chr> "汉口湖畔不临马路，东南朝向，中间楼层"，"双地铁口顶楼..."
## $ property_name <chr> "汉口湖畔"，"同鑫花园"，"东原启城一期"，"新华联青年城"..."
## $ property_region <chr> "盘龙城"，"后湖"，"白沙洲"，"纸坊"，"后湖"，"南湖沃尔..."
## $ price_ttl <dbl> 65.00, 160.00, 139.00, 130.00, 148.00, 229.00, 346.00, ...
## $ price_sqm <dbl> 7514, 14957, 14305, 9249, 11937, 18078, 26806, 10462, ...
## $ bedrooms <dbl> 2, 5, 3, 4, 3, 3, 3, 3, 3, 3, 3, 3, 2, 4, 4, 3, ...
## $ livingrooms <dbl> 2, 2, 1, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 1, 2, 2, 2, ...
## $ building_area <dbl> 86.51, 106.98, 97.17, 140.57, 123.99, 126.68, 129.08, ...
## $ directions1 <chr> "南"，"南"，"南"，"南"，"南"，"南"，"南"，"南"，"南"，"东南"..."
## $ directions2 <chr> NA, "北"，NA, NA, "北"，"北"，"北"，"北"，NA, NA, NA, ...
## $ decoration <chr> "毛坯"，"精装"，"精装"，"毛坯"，"精装"，"精装"，"精装"..."
## $ hml <chr> "中"，"高"，"中"，"中"，"低"，"中"，"低"，"高"，"中"，...
## $ building_height <dbl> 25, 6, 32, 18, 6, 17, 6, 34, 11, 33, 44, 24, 27, 20, 3...
## $ building_year <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ building_style <chr> "板楼"，"板楼"，"塔楼"，"板塔结合"，"板楼"，"塔楼"，"...
## $ followers <dbl> 4, 8, 4, 2, 15, 14, 45, 1, 8, 1, 10, 9, 0, 1, 6, 3, 9, ...
## $ pub_days <chr> NA, "2个月"，"2个月"，"2个月"，"2个月"，"2个月"，NA, "...
## $ near_subway <dbl> 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, ...
## $ two_year <dbl> 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, ...
## $ taxfree <dbl> 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, ...
## $ haskey <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
## $ vr <chr> NA, NA, NA, NA, "VR看装修"，NA, "VR看装修"，"VR看装修"..."
## $ price_date <date> 2024-01-31, 2024-01-31, 2024-01-31, 2024-01-31, 2024-...
## $ building_age <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
```

各变量的简短统计：

```
## property_title property_name property_region price_ttl
## Length:4700 Length:4700 Length:4700 Min. : 9.6
## Class :character Class :character Class :character 1st Qu.: 99.0
## Mode :character Mode :character Mode :character Median : 142.0
## Mean : 162.7
## 3rd Qu.: 196.0
## Max. :2300.0
##
## price_sqm bedrooms livingrooms building_area
## Min. : 2332 Min. :0.000 Min. :0.000 Min. : 12.25
## 1st Qu.:11424 1st Qu.:2.000 1st Qu.:1.000 1st Qu.: 82.52
## Median :15241 Median :3.000 Median :2.000 Median : 93.75
## Mean :16107 Mean :2.625 Mean :1.704 Mean : 98.98
## 3rd Qu.:19336 3rd Qu.:3.000 3rd Qu.:2.000 3rd Qu.:116.45
## Max. :66659 Max. :8.000 Max. :4.000 Max. :345.04
##
## directions1 directions2 decoration hml
## Length:4700 Length:4700 Length:4700 Length:4700
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
## building_height building_year building_style followers
## Min. : 1.00 Min. :1985 Length:4700 Min. : 0.000
## 1st Qu.:11.00 1st Qu.:2007 Class :character 1st Qu.: 1.000
## Median :27.00 Median :2012 Mode :character Median : 4.000
## Mean :24.29 Mean :2011 Mean : 7.365
## 3rd Qu.:33.00 3rd Qu.:2015 3rd Qu.: 8.000
## Max. :60.00 Max. :2021 Max. :260.000
## NA's :48 NA's :2612 NA's :2088
## pub_days near_subway two_year taxfree
## Length:4700 Min. :0.0000 Min. :0.0000 Min. :0.0000
## Class :character 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Mode :character Median :1.0000 Median :0.0000 Median :0.0000
## Mean :0.5302 Mean :0.3633 Mean :0.4349
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000
## NA's :2088 NA's :2088 NA's :2088
## haskey vr price_date building_age
## Min. :0.0000 Length:4700 Min. :2021-12-01 Min. : 0.00
## 1st Qu.:1.0000 Class :character 1st Qu.:2021-12-01 1st Qu.: 6.00
## Median :1.0000 Mode :character Median :2024-01-31 Median : 9.00
## Mean :0.8105 Mean :2023-02-13 Mean :10.35
## 3rd Qu.:1.0000 3rd Qu.:2024-01-31 3rd Qu.:14.00
## Max. :1.0000 Max. :2024-01-31 Max. :36.00
## NA's :2088 NA's :2612
```

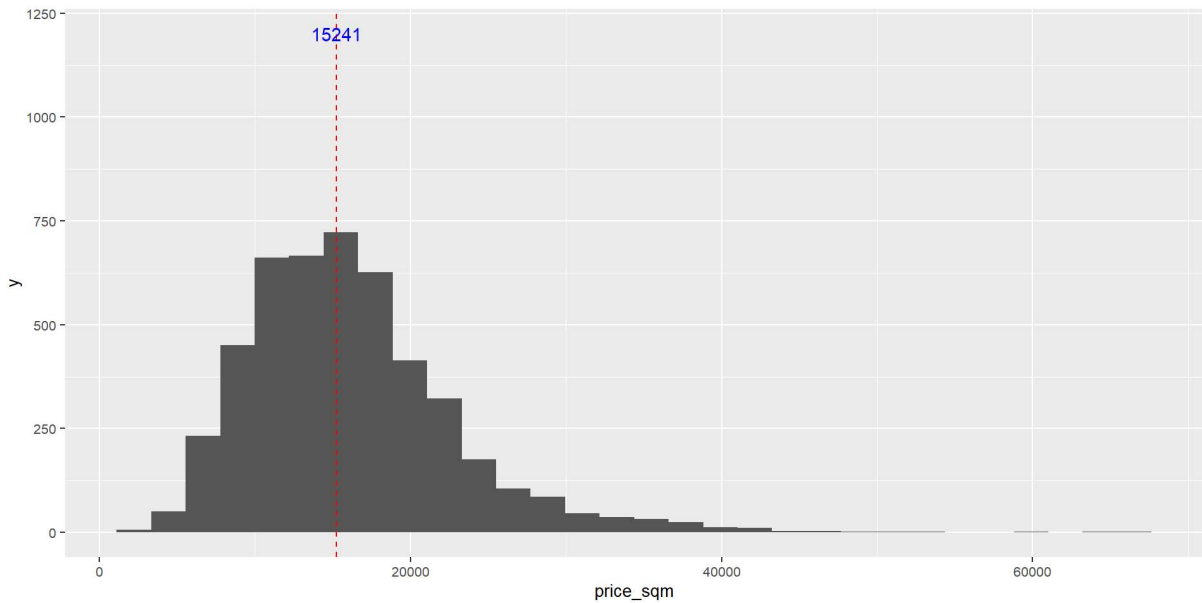
可以看到：

- 二手房房屋售价总价主要集中于100万至200万之间
- 二手房房屋售价均价1.1万/平米至1.9万/平米之间
- 二手房房屋年限平均为10年

4 探索性分析

4.1 房屋单价的数值描述与图形

```
##   vars   n   mean     sd median trimmed   mad min  max range skew
## X1    1 4700 16107.21 6663.87  15241 15477.36 5831.81 2332 66659 64327 1.31
##      kurtosis   se
## X1         3.72 97.2
```

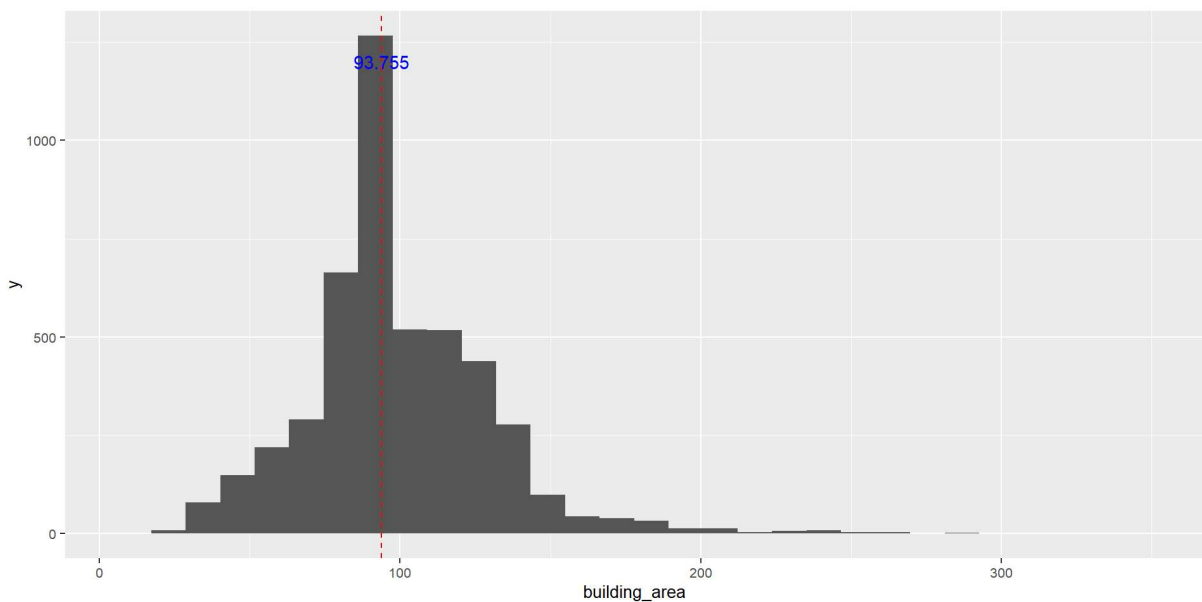


发现：

- 房屋单价分布呈左偏趋势
- 高价位房屋价格较为分散，低价位房屋价格较为集中

4.2 建筑面积的数值描述与图形

```
##   vars   n  mean    sd median trimmed   mad min  max range skew kurtosis
## X1    1 4700  98.98 30.68  93.75  97.59 24.03 12.25 345.04 332.79 1.12    4.44
##      se
## X1  0.45
```

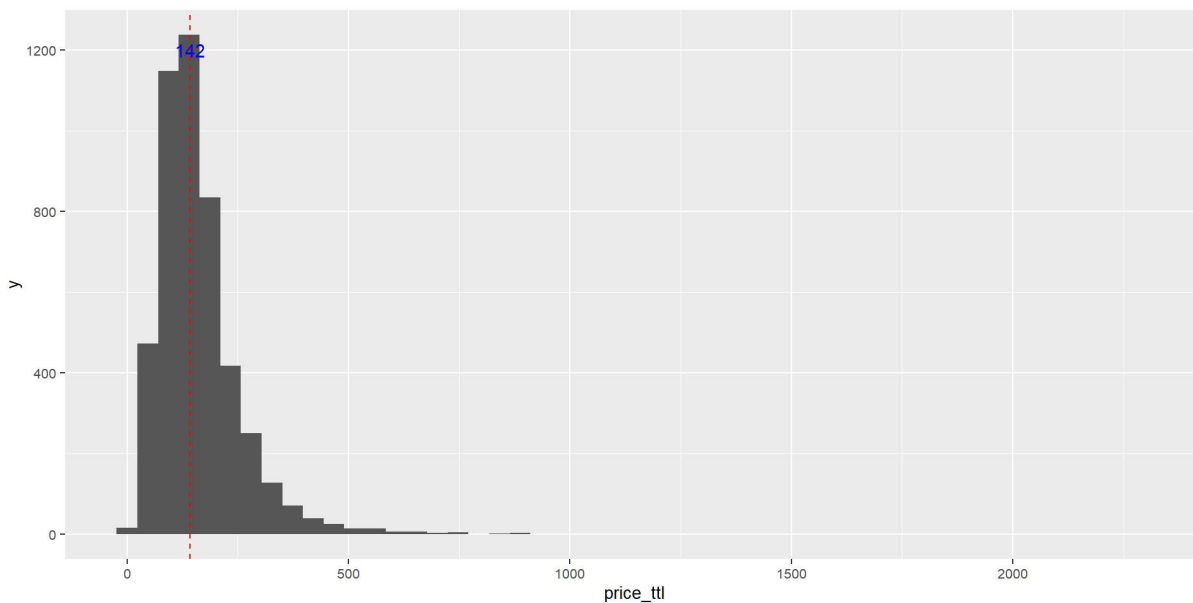


发现：

- 二手房建筑面积分布较为集中，呈左偏趋势
- 90-100平米面积的二手房屋数量最高，200平米以上房屋数量分散

4.3 房屋总价的数值描述与图形

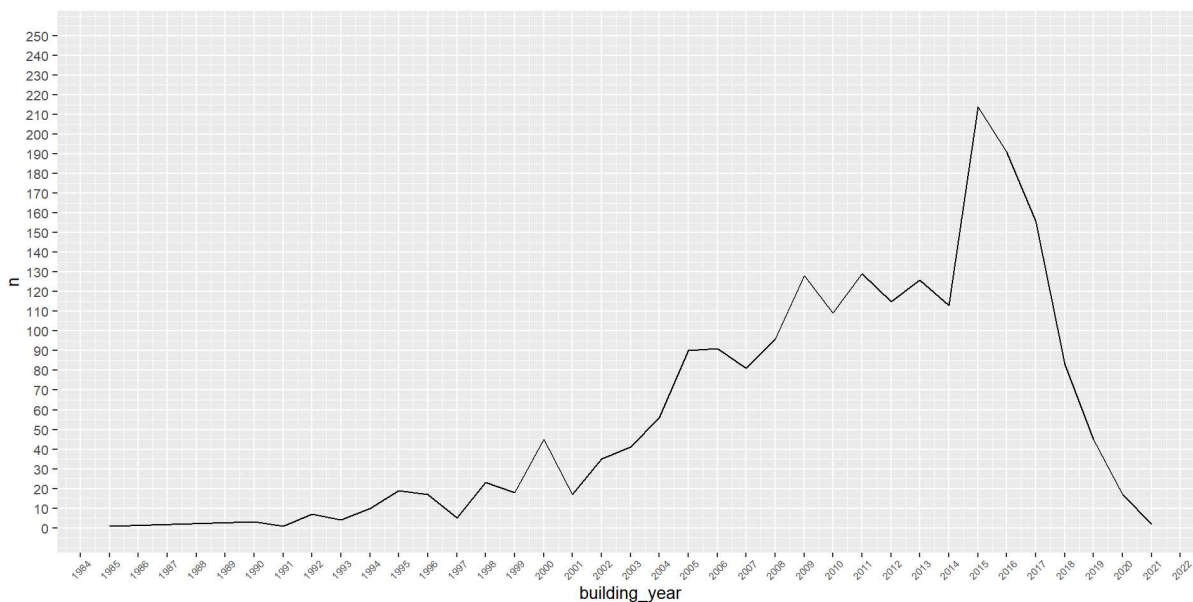
```
##      vars   n  mean    sd median trimmed  mad min  max range skew kurtosis
## X1      1 4700 162.66 103.72   142  148.87 69.68 9.6 2300 2290.4 4.45    55.55
##           se
## X1 1.51
```



发现：

- 房屋总价分布十分集中，mad切尾均值约为70万元，中位数为142万
- 高价房屋分布分散

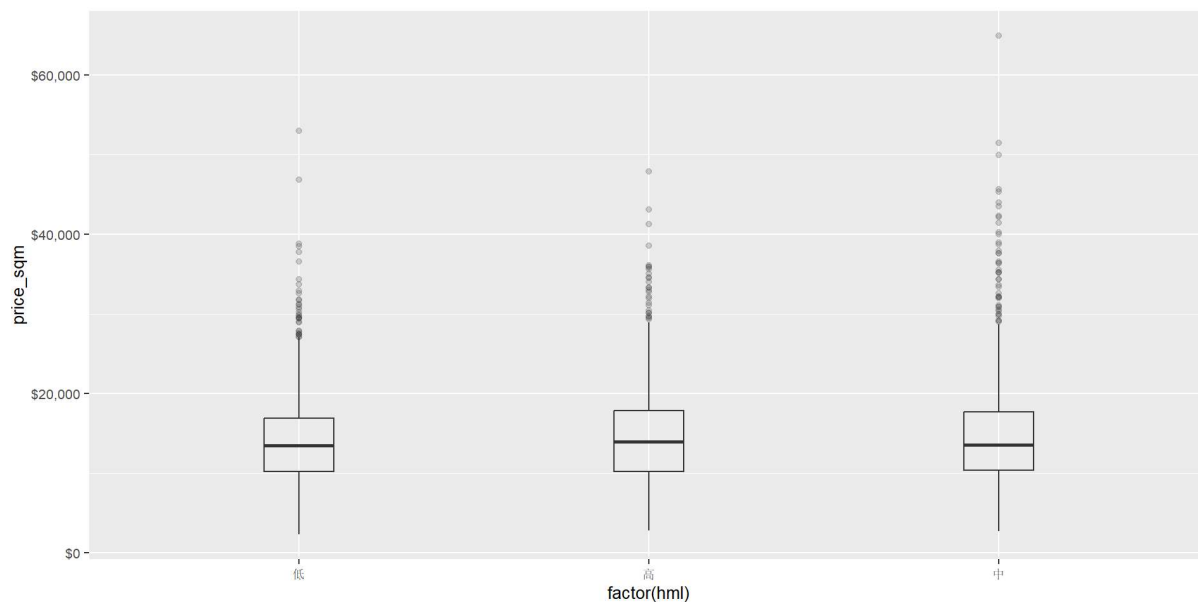
4.4 二手房建筑年限房屋数量分布



发现：

- 二手房供应建筑年限与国家楼市政策密切相关，例如2015年首套房首付款比例降低等促进售房政策导致房价上升，供房增加

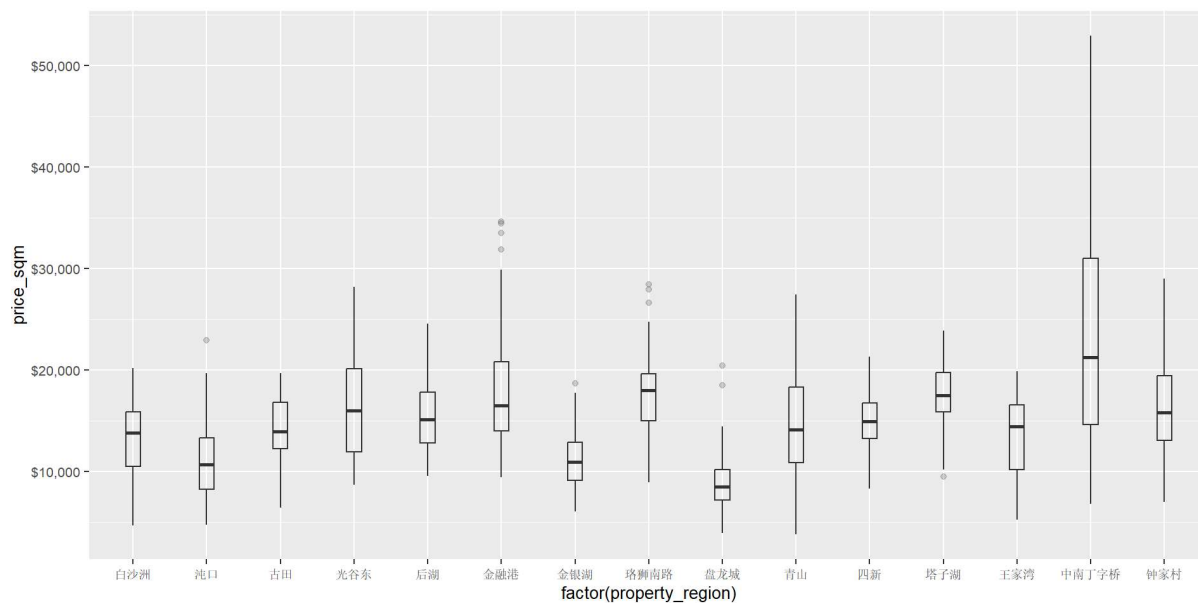
4.5 房屋所处楼层对房屋单价的影响



发现:

- 不同楼层房屋单价分布差异较小，房屋楼层对单价影响较小
- 高楼层房屋单价要稍高于其他类型

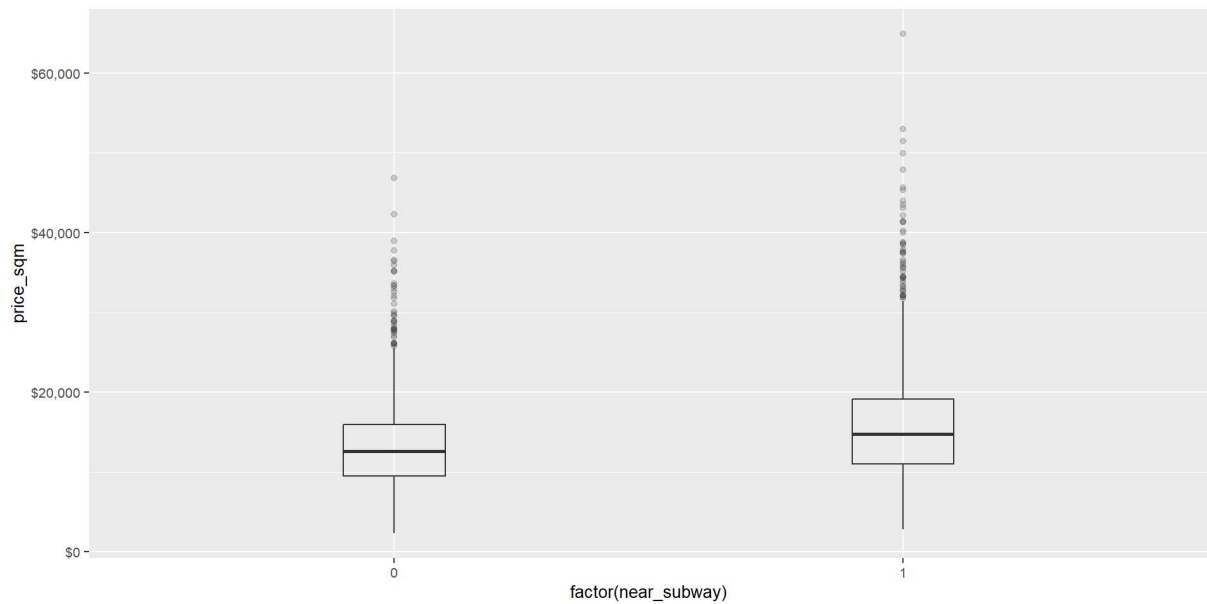
4.6 房屋所处区域对房屋单价的影响



发现: 选取房源数量大于100的区域进行房屋单价分析

- 不同区域房屋单价分布差异较大，房屋所属区域对房屋单价有较大影响
- 市内房屋价格普遍高于郊区房屋价格

4.7 是否靠近地铁对房屋单价的影响



发现:

- 否靠近地铁对房屋单价有所影响，但差异较小
- 靠近地铁的房屋单价分布要稍高于非靠近地铁的房屋

5 发现总结

1. 房屋单价、总价及面积分布均呈左偏趋势，偏度较大
2. 房屋所属区域对房屋单价影响较大，所处楼层及受否靠近地铁影响较小
3. 房屋建筑年限分布与国家楼市政策密切相关