

# 陈魏娟-2024281050965-第一次作业

Code

陈魏娟

2024-10-31

- 1 你的主要发现
- 2 数据介绍
- 3 探索性分析
  - 3.1 变量1的数值描述与图形
  - 3.2 变量2的数值描述与图形
  - 3.3 变量...的数值描述与图形
  - 3.4 探索问题1
- 4 房屋的位置是否影响房屋的单价，靠近地铁的房屋单价高于不靠近地铁的房屋
  - 4.1 探索问题2
- 5 房屋单价是否受房屋面积的影响
  - 5.1 探索问题3
- 6 房屋单价是否受房屋精装修或毛坯的影响
- 7 发现总结

## 1 你的主要发现

- 发现1 白沙洲的房子供应最多。
- 发现2 中北路附近的房屋单价最高。
- 发现3 大多数房屋的朝向都向南。

## 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。

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		237.								
南湖沃尔										

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          <int> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163…
## $ bedrooms          <int> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4…
## $ livingrooms        <int> 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", "", "", "东", "", "北", "", "北", "北", "北", …
## $ decoration         <chr> "精装", "精装", "简装", "精装", "简装", "精装", "简…
## $ property_t_height  <int> 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          <int> 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", …
```

```
## 'data.frame': 3000 obs. of 18 variables:
## $ 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" ...
## $ price_ttl : num 237 127 75 188 182 ...
## $ price_sqm : int 18709 14613 15968 15702 17509 10376 12346 16336 32631 17403 ...
## $ bedrooms : int 3 3 1 3 3 3 2 3 4 3 ...
## $ livingrooms : int 1 2 1 2 2 2 1 2 1 2 ...
## $ building_area : num 126.7 86.9 47 119.7 104 ...
## $ directions1 : chr "南" "南" "南" "北" ...
## $ directions2 : chr "北" "" "" "东" ...
## $ decoration : chr "精装" "精装" "简装" "精装" ...
## $ property_t_height: int 17 28 18 32 34 34 7 34 5 7 ...
## $ property_height : chr "中" "中" "低" "高" ...
## $ property_style : chr "塔楼" "板楼" "塔楼" "塔楼" ...
## $ followers : int 3 1 3 2 3 1 0 0 2 0 ...
## $ near_subway : chr "近地铁" NA "近地铁" "近地铁" ...
## $ if_2y : chr NA "房本满两年" NA "房本满两年" ...
## $ has_key : chr "随时看房" "随时看房" "随时看房" "随时看房" ...
## $ vr : chr NA "VR看装修" NA NA ...
```

```
## [1] "property_name" "property_region" "price_ttl"
## [4] "price_sqm" "bedrooms" "livingrooms"
## [7] "building_area" "directions1" "directions2"
## [10] "decoration" "property_t_height" "property_height"
## [13] "property_style" "followers" "near_subway"
## [16] "if_2y" "has_key" "vr"
```

```
## 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
##
##
##
```

```
## [1] 15148.49
```

```
## [1] 14404
```

```
## 0% 25% 50% 75% 100%
## 1771.00 10799.25 14404.00 18211.00 44656.00
```

```
## [1] 1771 44656
```

```
## [1] 7411.75
```

```
## [1] 39982547
```

```
## [1] 6323.175
```

```
## [1] 5465.605
```

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
##
##
##
```

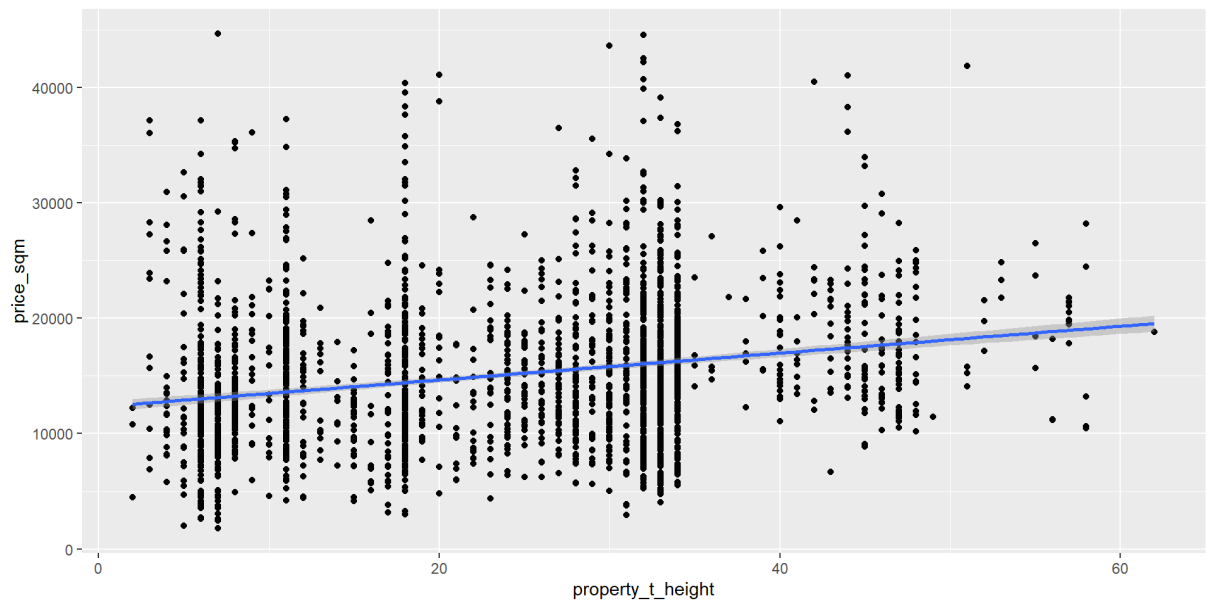


可以看到：

- 直观结论1 #白沙洲，盘龙城，四新，光谷东，王家湾的房源最多。
- 直观结论2 #总体来看武昌的房源比汉口的房源多。

### 3 探索性分析

3.1 变量1的数值描述与图形



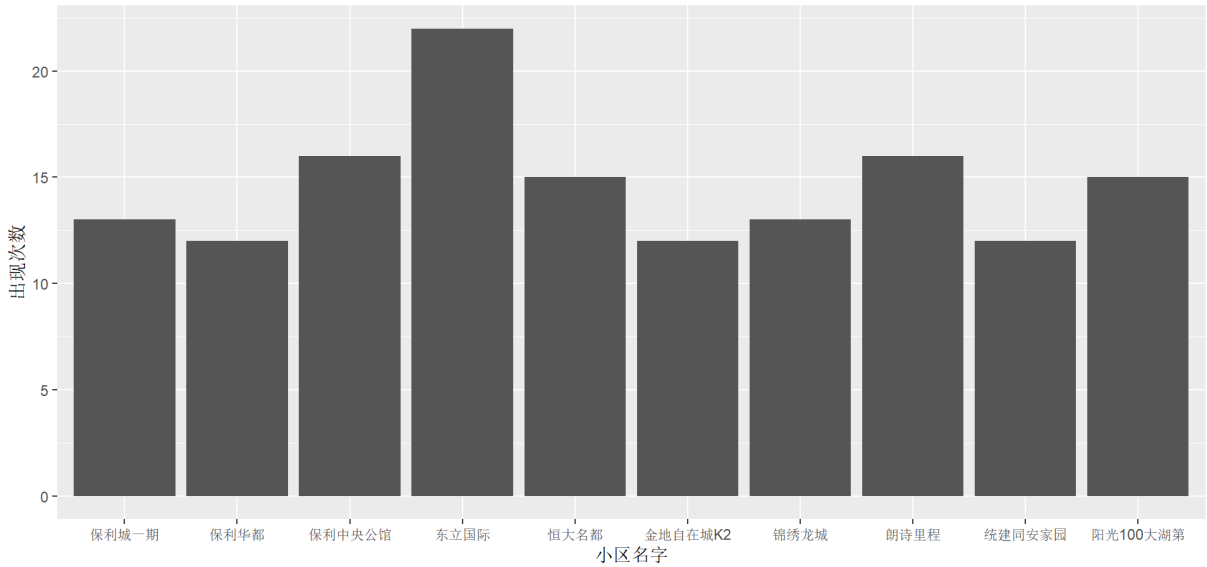
发现：楼层高低对房屋单价的影响

- 发现1 # 大多数房屋的总层数在20层左右。
- 发现2 # 单价均值随着总层数的增加而增加

3.2 变量2的数值描述与图形

```
## # A tibble: 10 × 2
##   property_name count
##   <chr>          <int>
## 1 东立国际         22
## 2 保利中央公馆    16
## 3 朗诗里程        16
## 4 恒大名都        15
## 5 阳光100大湖第   15
## 6 保利城一期      13
## 7 锦绣龙城        13
## 8 保利华都        12
## 9 统建同安家园    12
## 10 金地自在城K2   12
```

出现最多次数的小区名字



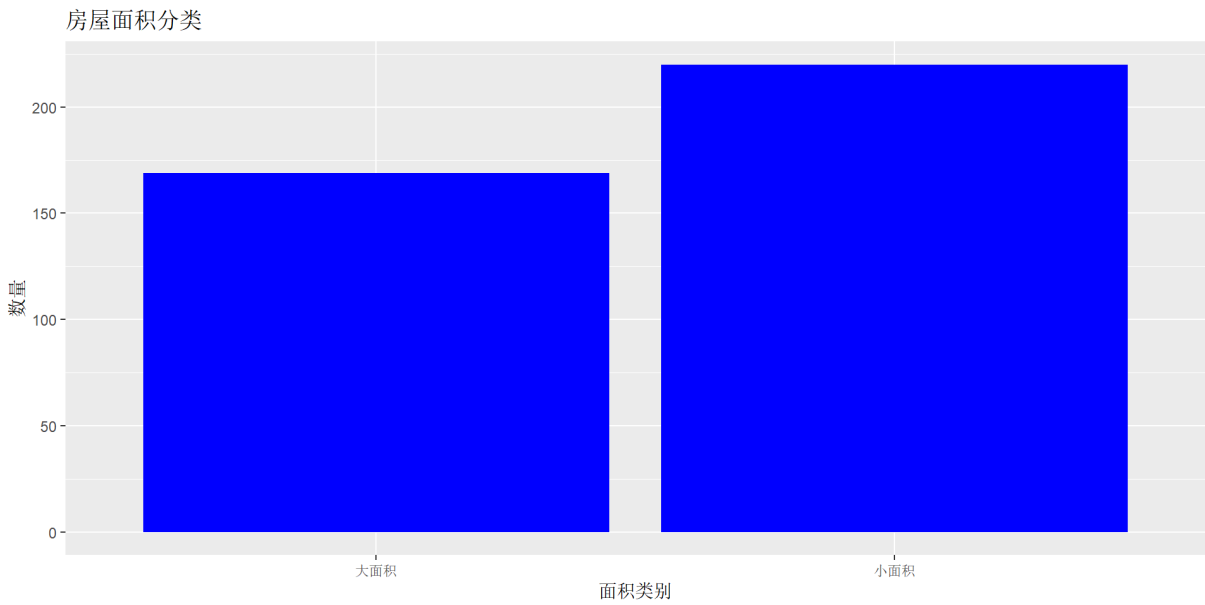
发现：2023年小区出现次数前十的名字

- 发现1 # 东立国际的房源最多。
- 发现2 # 前十个小区的房源都在十个以上。

3.3 变量...的数值描述与图形

```
##
## 大面积 小面积
##    169    220

##
##    大面积    小面积
## 0.4344473 0.5655527
```



发现：房源供应中大面积的房源多还是小面积的房源多。

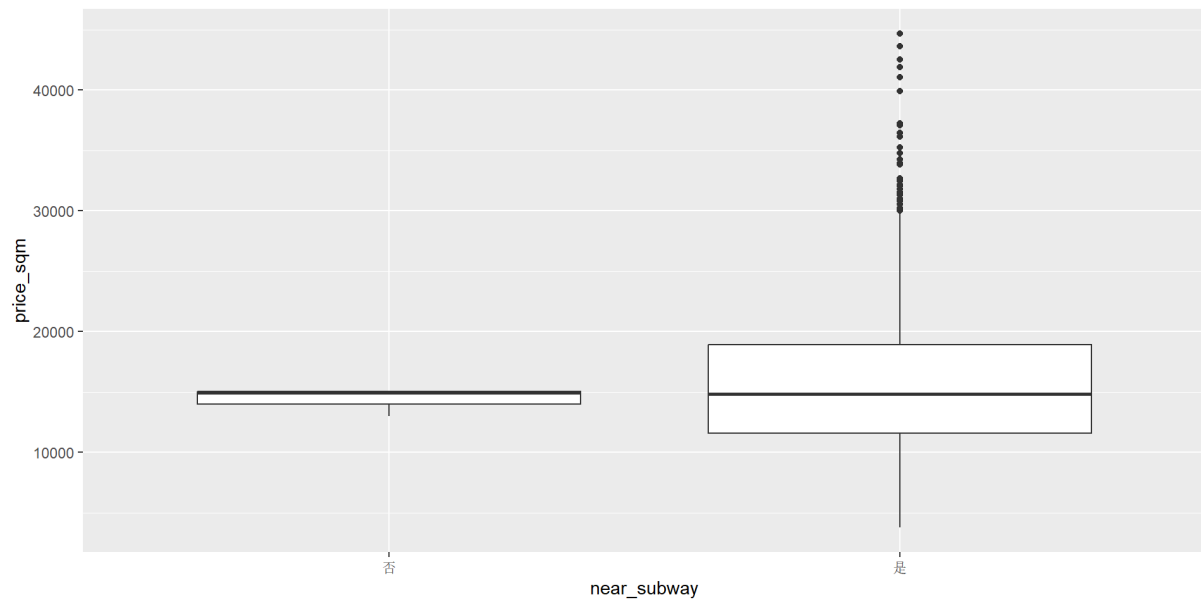
- 发现1 总体来看小面积的房源比大面积的房源多。
- 发现2 小面积的房源和大面积的房源数量差距不大。

3.4 探索问题1

4 房屋的位置是否影响房屋的单价，靠近地铁的房屋单价高于不靠近地铁的房屋

发现：

```
## # A tibble: 73 × 2
##   property_region num_properties
##   <chr>           <int>
## 1 白沙洲           36
## 2 青山             28
## 3 钟家村           25
## 4 中南丁字桥       23
## 5 塔子湖           21
## 6 老南湖           21
## 7 后湖             19
## 8 王家湾           19
## 9 台北香港路       18
## 10 CBD西北湖       17
## # 63 more rows
```



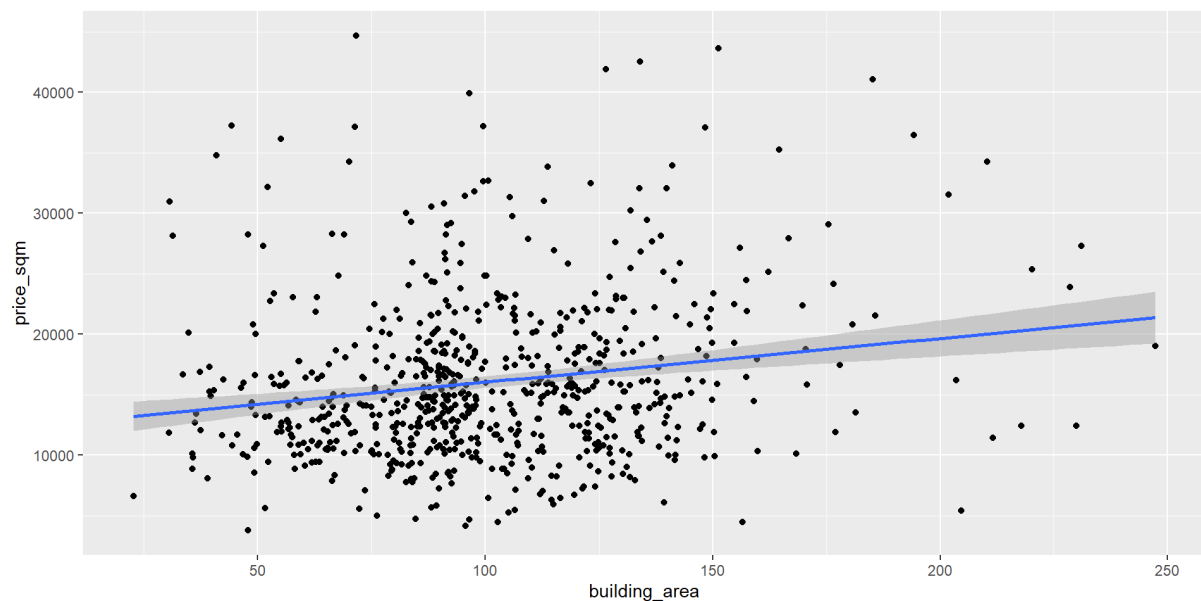
```
##
## Welch Two Sample t-test
##
## data: price_sqm by near_subway
## t = -2.3659, df = 2.5725, p-value = 0.1133
## alternative hypothesis: true difference in means between group 否 and group 是 is not equal to 0
## 95 percent confidence interval:
## -4214.4889 816.8688
## sample estimates:
## mean in group 否 mean in group 是
## 14352.00 16050.81
```

- 发现1 # 房屋的位置影响房屋的单价，靠近地铁的房屋单价高于不靠近地铁的房屋
- 发现2 # 白沙洲近地铁的房源最多。

#### 4.1 探索问题2

### 5 房屋单价是否受房屋面积的影响

发现:

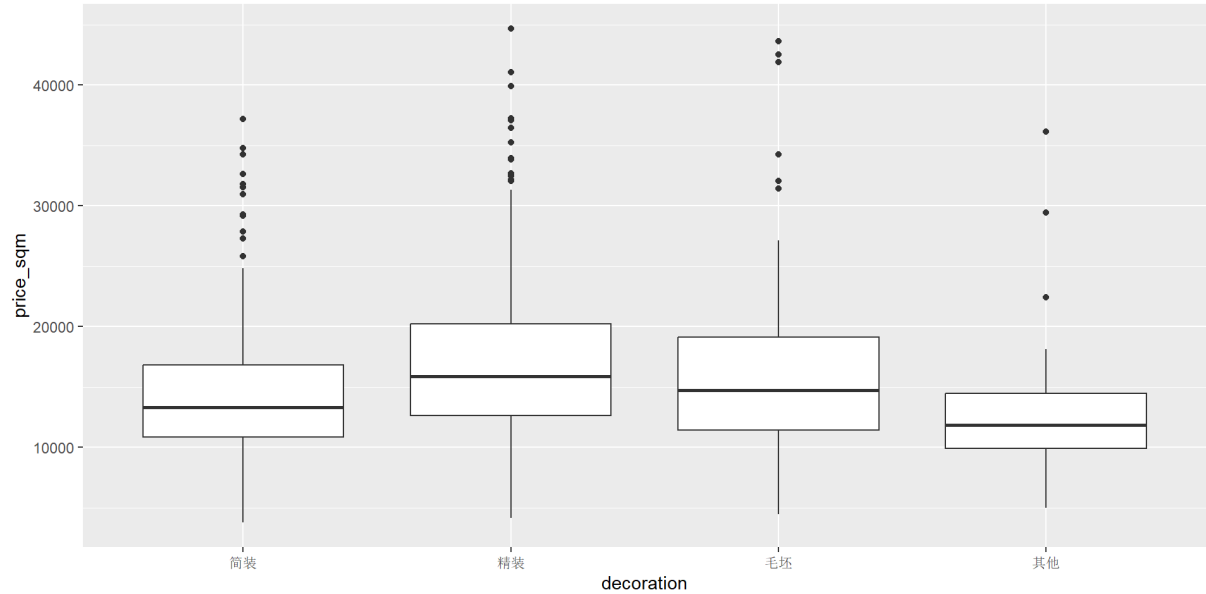


- 发现1 # 大致上来看房屋面积越大，房屋单价越高
- 发现2 # 大多数房源房屋的面积都在100平米左右

#### 5.1 探索问题3

### 6 房屋单价是否受房屋精装修或毛坯的影响

发现：



-发现1 # 目前的房源中，精装房的数量最多

- 发现2 # 房屋单价受是否精装的影响，精装房的房屋单价均值相对而言最高。

## 7 发现总结

用1-3段话总结你的发现。近几年武汉的房源供应情况以白沙洲，盘龙城等较为偏远的地方居多，市中心房源供应的情况较少，由此可见市区土地有限，人口在逐渐往郊区外扩。房屋的单价与房屋所在的地段，附近是否有地铁，房屋是否为精装修，房屋的层高以及房屋的面积均相关。一般在其他条件相同的情况下地段越中心，附近有地铁，房屋为精装修，楼高偏高，面积大的房屋单价越高。虽然房屋的单价受房屋面积的影响，面积打的房屋相对单价越高，但是小面积和大面积的市场供需来看，小面积房屋似乎更多。