

武汉市二手房数据分析

魏勤伟

目录

1	你的主要发现	1
2	数据介绍	2
3	发现	3
4	数据概览	3
5	探索性分析	5
5.1	变量 1price_sqm(房屋单价) 的数值描述与图形	5
5.2	变量 2bedrooms(房间数) 的数值描述与图形	6
5.3	变量 3 directions1(主要朝向) 的数值描述与图形	7
5.4	探索问题 1: 精装修价格与其他类型均价差异分析:	8
5.5	探索问题 3: directions1(主要朝向) 与 price_sqm(平均房价) 之间的关系	11
6	发现总结	12

1 你的主要发现

- 发现 1: 并不是房间数量越多，房价就越高
- 发现 2: 武汉市房屋单价最大值比最小值高数十倍

- 发现 3：房屋主要朝向虽然朝南最多，但朝西方向的单价最高

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。

##	key	value	key	value
##	"property_name"	"0"	"property_region"	"0"
##	key	value	key	value
##	"price_ttl"	"0"	"price_sqm"	"0"
##	key	value	key	value
##	"bedrooms"	"0"	"livingrooms"	"0"
##	key	value	key	value
##	"building_area"	"0"	"directions1"	"0"
##	key	value	key	value
##	"directions2"	"1672"	"decoration"	"0"
##	key	value	key	value
##	"property_t_height"	"0"	"property_height"	"60"
##	key	value	key	value
##	"property_style"	"0"	"followers"	"0"
##	key	value	key	value
##	"near_subway"	"1441"	"if_2y"	"1736"
##	key	value	key	value
##	"has_key"	"458"	"vr"	"906"

表 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126
万科紫悦湾	光谷东	127.0	14613	3	2	86
东立国际	二七	75.0	15968	1	1	46
新都汇	光谷广场	188.0	15702	3	2	119
保利城一期	团结大道	182.0	17509	3	2	103
加州橘郡	庙山	122.0	10376	3	2	117
省建筑五公司西区	光谷广场	99.0	12346	2	1	80
保利上城东区	白沙洲	193.8	16336	3	2	118
石化大院	中南丁字桥	325.0	32631	4	1	99
阳光花园	杨汊湖	192.0	17403	3	2	110

3 发现

- 发现 1: directions2、property_height、property_style、near_subway、if_2y、has_key、vr 这些字段均存在 NA 值，需要在后续分析中注意处理

4 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
```

```
## $ 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", "南", "~
## $ directions2        <chr> "北", NA, NA, "东", NA, "北", NA, "北", "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
##
##
##

```

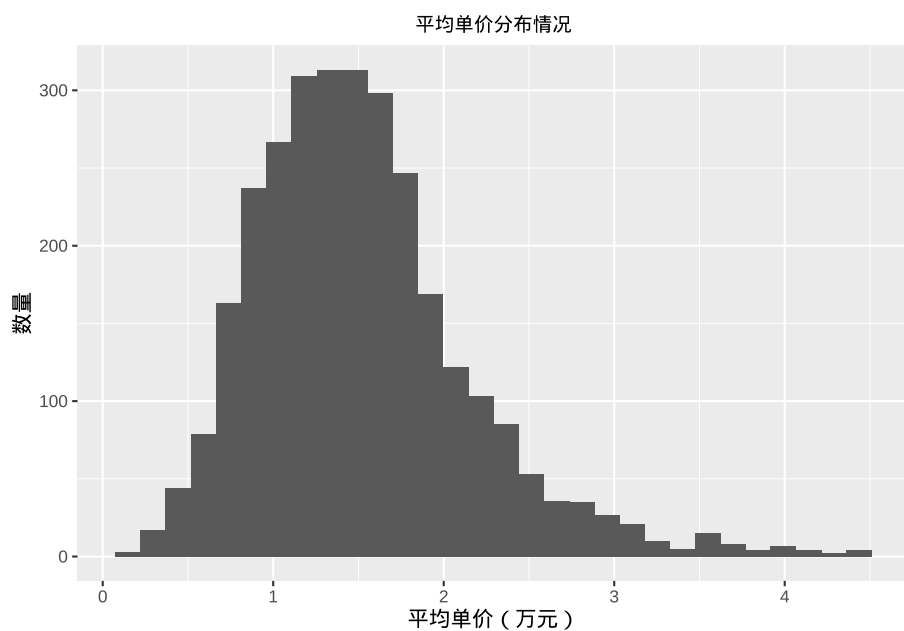
可以看到:

- 直观结论 1: 房屋主要朝向大部分为南或东南
- 直观结论 2: 大部分房子都是标称靠近地铁

5 探索性分析

5.1 变量 1price_sqm(房屋单价) 的数值描述与图形

```
## [1] "武汉地区房屋单价情况分析, 最高单价: 4.4656 最低单价: 0.1771 单价中位数: 1.44
```

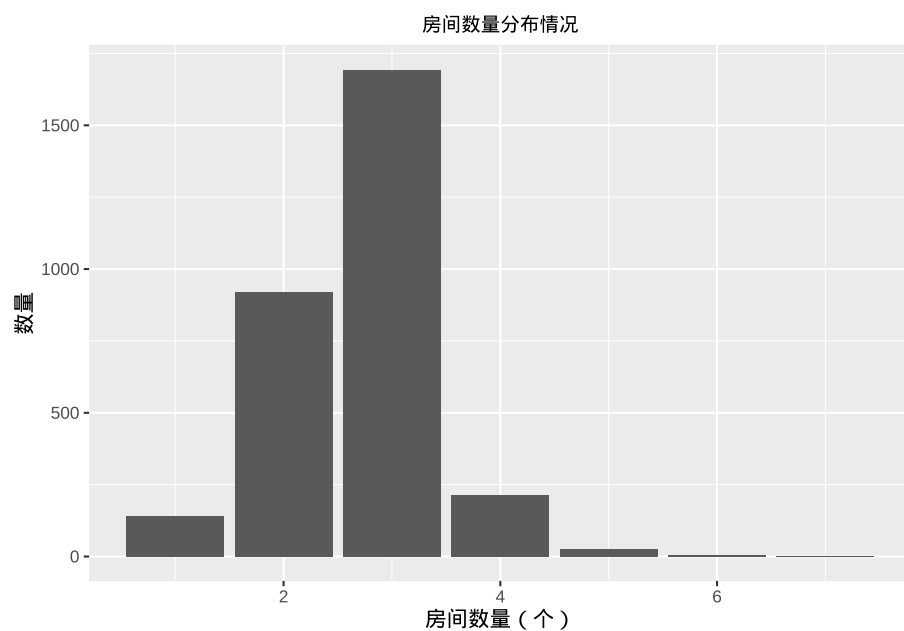


发现

- 发现 1：二手房最高单价与最低单价差距巨大，达到数十倍。 — 发现 2：价格中位数小于均值，说明高价房屋数量大于低价房屋

5.2 变量 2bedrooms(房间数) 的数值描述与图形

```
## [1] "武汉地区房屋房间数量情况分析，最大房间数： 7 最小房间数： 1 房间数中位数： 3 房
```

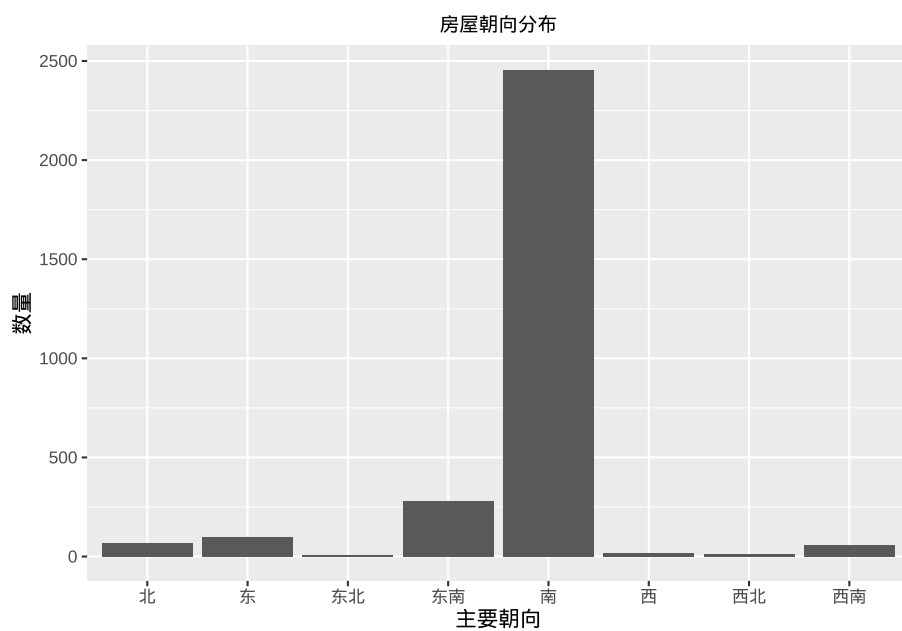


发现:

- 发现 1: 大部分房子房间数在 2-3 个 — 发现 2: 房屋数量大于 5 的房屋占比非常小

5.3 变量 3 directions1(主要朝向) 的数值描述与图形

```
## # A tibble: 8 x 2
##   directions1 count_directions1
##   <chr>         <table[1d]>
## 1 东             98
## 2 东北          10
## 3 东南         281
## 4 北            68
## 5 南          2454
## 6 西           19
## 7 西北          13
## 8 西南          57
```

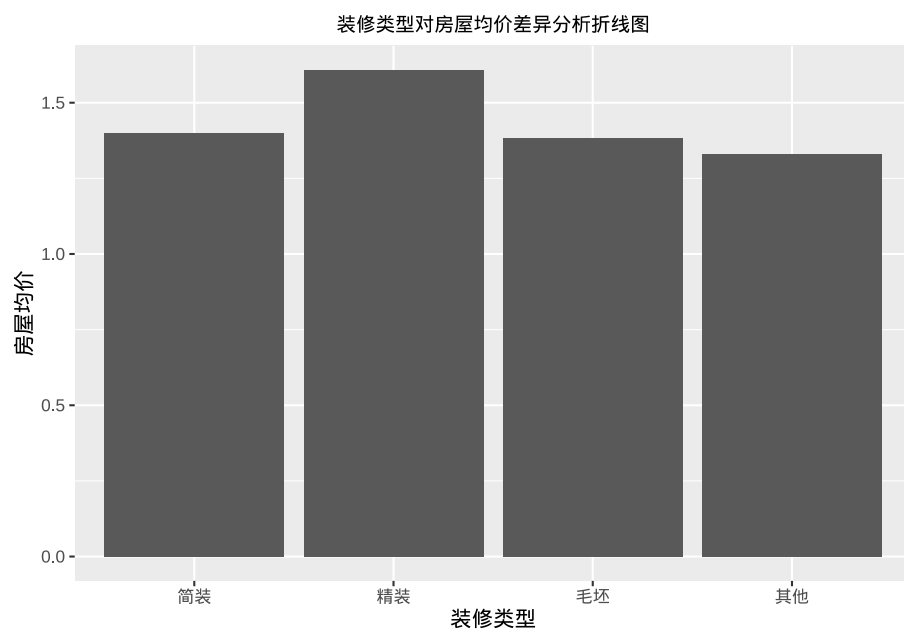


发现:

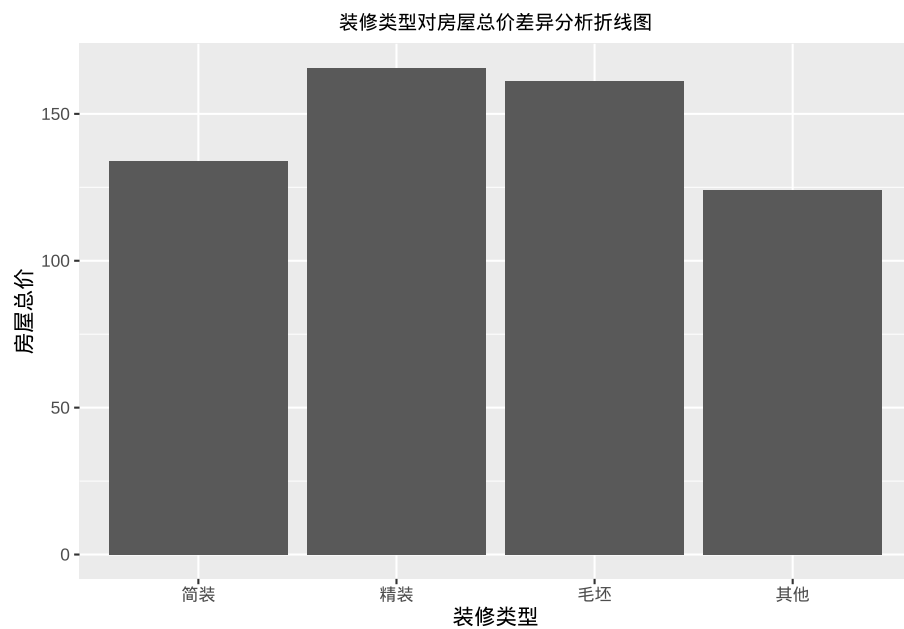
- 发现 1: 绝大部分房型主要朝向为东南方向
- 发现 2: 各种朝向的房屋均价中位数差距不大, 但极端值 (均价偏高) 中南朝向的较多, 且离中位数差异最大

5.4 探索问题 1: 精装修价格与其他类型均价差异分析:

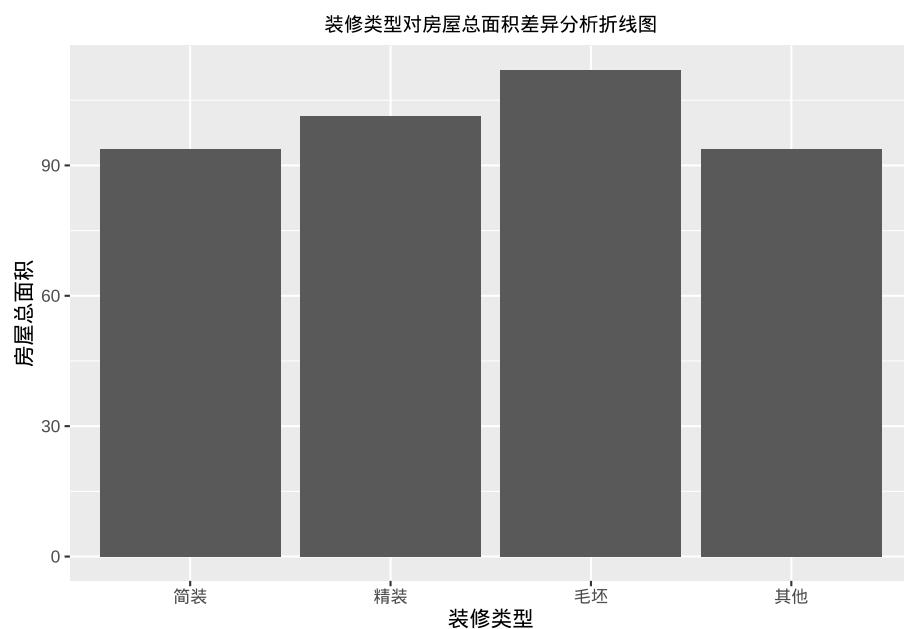
```
## # A tibble: 4 x 4
##   decoration avg_ttl avg_sqm avg_building_area
##   <chr>      <dbl>   <dbl>         <dbl>
## 1 其他        124.    1.33          93.8
## 2 毛坯        161.    1.38         112.
## 3 简装        134.    1.40          93.8
## 4 精装        166.    1.61         101.
```

根据总价值绘制条形图



根据总平均面积绘制条形图

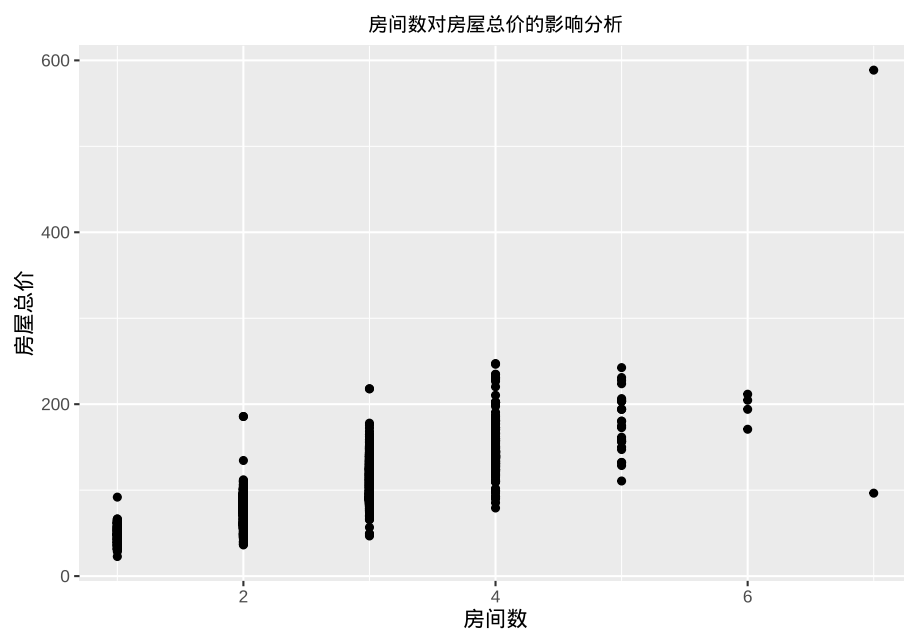


发现:

- 发现 1: 通过上图我们可以显著发现精装修的房屋均价、总价均高于其他装修形式的房屋
- 发现 2: 毛坯房的平均建筑面积是所有类型中最高的

探索问题 2: 分析 bedrooms(房间数) 与 price_ttl (房屋总价) 之间的关系

```
## # A tibble: 7 x 2
##   bedrooms price_ttl
##   <dbl>     <dbl>
## 1       1      72.2
## 2       2     116.
## 3       3     167.
## 4       4     265.
## 5       5     362.
## 6       6     240.
## 7       7     748.
```



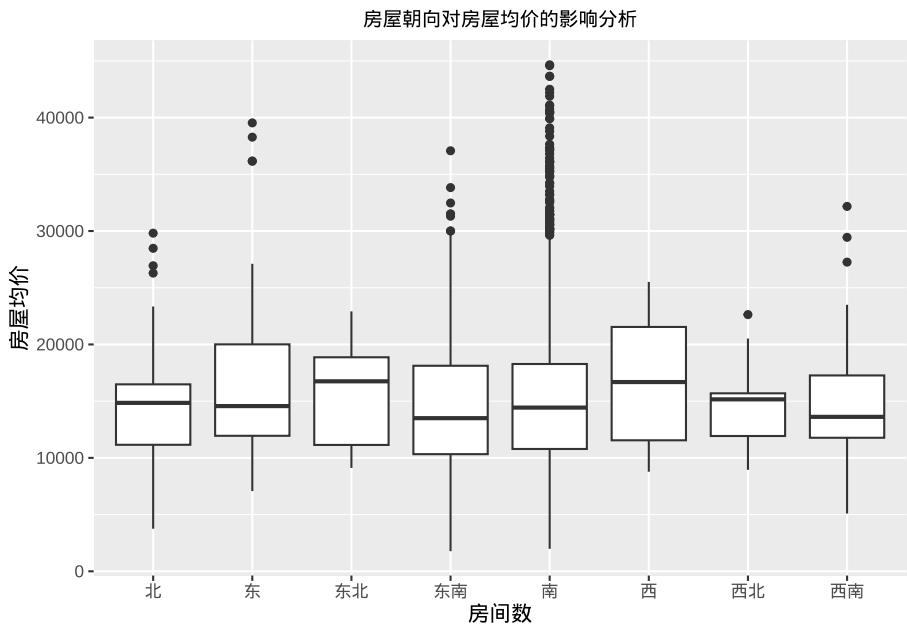
发现:

- 发现 1: 在 1-4 个房间的房型中, 平均房屋总价与房间数量整体呈正比, 但 6 个以上的房间的房型总价未发现明显规律, 可能是因为在数据集中相关样本数量太少导致
- 发现 2: 绝大部分时候房间数量越大, 房屋总价越高

5.5 探索问题 3: directions1(主要朝向) 与 price_sqm(平均房价) 之间的关系

```
## # A tibble: 8 x 2
##   directions1 price_sqm
##   <chr>         <dbl>
## 1 东             1.62
## 2 东北          1.58
## 3 东南          1.48
## 4 北            1.45
```

## 5 南	1.52
## 6 西	1.72
## 7 西北	1.46
## 8 西南	1.44



- 发现：
- 发现 1：朝西方向的房屋平均的单价最高
 - 发现 2：朝西北方向的房屋单价上四分位值与中位数较为接近，说明该方向上房间单价分布较为接近。

6 发现总结

用 1-3 段话总结你的发现。通过运用数据分析工具对 3000 个二手房数据的分析，打破了我的一些固有印象，主要发现如下：发现 1：并不是房间数量越多，房价就越高发现 2：武汉市房屋单价最大值比最小值高数十倍，说明

贫富差距还是很大的发现 3: 房屋主要朝向虽然朝南最多, 但朝西方向的单价最高, 这个比较反直觉, 需要进一步细化对朝西方向的房屋的抽样调查, 了解为什么会这样