

# 关于链家二手房商业分析报告

瞿芬芳

## 目录

0.1	.....	2
1	一. 主要发现:	2
2	二. 数据介绍及整体结构	3
2.1	1. 本报告链家数据获取方式如下: .....	3
2.2	2. 数据概览 .....	3
2.3	3. 数据表前 10 行示例如下: .....	4
2.4	4. 数据整体结构 .....	4
3	三. 数据清洗	5
3.1	1. 查看数据缺失情况并去重 .....	5
3.2	2. 查看各数据简短信息 .....	6
4	四. 探索性研究	10
4.1	1. 二手房主要类型特征 .....	10
4.2	4. 探索问题 .....	13
4.3	问题 2: 区域对房屋均价的影响? .....	15
4.4	问题 4: 区域房屋单价分布情况 .....	17
5	五. 决策建议:	17

## 0.1

## 1 一. 主要发现:

发现 1: 武汉二手房数据有如下特征 (1) 房屋总价: 最低价 10.60 万元, 最高价 1,380.00 万元, 中位数 136.00 万元, 平均数 154.80 万元, 数据分布呈右偏分布。(2) 房屋单价: 最低价 1,771 元/ $m^2$ , 最高价 44,656 元/ $m^2$ , 中位数 14,309 元/ $m^2$ , 平均数 15,110 元/ $m^2$ , 数据分布呈现右偏分布, 因为房屋面积不同, 整体单价的右偏幅度较房屋总价较小。(3) 建筑面积: 最小面积 22.77 $m^2$  最大面积 588.66 $m^2$ , 中位数 95.46 $m^2$ , 平均数 100.67 $m^2$ , 呈现右偏分布。(4) 其他二手房数据特征: -房间数: 以 2~3 居室居多 -客厅数: 以 1~2 客厅居多 -楼栋整层数: 平均楼层为 24 层的中层建筑

发现 2: 二手房面积集中在 100 $m^2$ ; 以三室两厅房型为主。

发现 3: 装修过的二手房占比近 80%, 其中精装数占比 50% 以上。

发现 4: 建筑类型为板楼类型最多 1781 套, 其次是板塔结合 615 套。

发现 5: 在售二手房数排名前 10 区域显示, 远城区二手房数量相对较多。而排名后 5 区域显示, 吴家山/沙湖/CBD 西北湖/沌口/团结大道二手房数量相对少。

发现 6: 武汉在售二手房地区房屋均价在 15000~20000 元间; -均价最高值: 中北路约 30000 元; -均价最低值: 阳逻约 5000 元。

发现 7: 面积对单价有如下影响 -面积约在 120 $m^2$  内的房屋, 面积的增长对单价影响较平稳; -面积大于 350 $m^2$  的房屋, 单价随面积增加略有下降; -面积在 120 $m^2$  至 350 $m^2$  区间, 单价随面积上升幅度较大; -分析极值点的离散情况, 高房价存在个别极值, 说明离散度较高。

发现 8: 区域房屋单价分布情况, 其中光谷、青山、王家湾等区域房屋单价离散度较高; 盘龙城、四新等区域房屋单价集中性度较高。

## 2 二. 数据介绍及整体结构

### 2.1 1. 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。

### 2.2 2. 数据概览

变量	解释
property_name	小区名字
property_region	所处区域
price_ttl	房屋总价，单位万元
price_sqm	房屋单价，单位元
bedrooms	房间数
livingrooms	客厅数
building_area	建筑面积
directions1	房屋主要朝向
directions2	房屋次要朝向
decoration	装修状况
property_t_height	楼栋总层数
property_height	房屋在所在楼栋所处位置，取值为高中低
property_style	建筑形式，如板楼、塔楼等
followers	在该二手房网站的关注人数
near_subway	是否靠近地铁

表 2: 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

变量	解释
if_2y	产证是否满 2 年
has_key	中介是否有钥匙，标注”随时看房”表示有钥匙
vr	是否支持 VR 看房

2.3 3. 数据表前 10 行示例如下：

数据表 (lj\_wuhan) 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。

2.4 4. 数据整体结构

```
## 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", "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", ~
```

3 三. 数据清洗

3.1 1. 查看数据缺失情况并去重

样本数据存在重复，去重后由原先样本数 3000，得到实际可用样本数 2515。  
大多数分类的数值缺失度较低，对数值分析影响较小。

表 3: Table continues below

property_name	property_region	price_ttl	price_sqm
Mode :logical	Mode :logical	Mode :logical	Mode :logical
FALSE:2515	FALSE:2515	FALSE:2515	FALSE:2515
NA	NA	NA	NA

表 4: Table continues below

bedrooms	livingrooms	building_area	directions1	directions2
Mode	Mode	Mode	Mode	Mode
:logical	:logical	:logical	:logical	:logical
FALSE:2515	FALSE:2515	FALSE:2515	FALSE:2515	FALSE:2515
NA	NA	NA	NA	NA

表 5: Table continues below

decoration	property_t_height	property_height	property_style
Mode :logical	Mode :logical	Mode :logical	Mode :logical
FALSE:2515	FALSE:2515	FALSE:2462	FALSE:2515
NA	NA	TRUE :53	NA

followers	near_subway	if_2y	has_key	vr
Mode	Mode	Mode	Mode	Mode
:logical	:logical	:logical	:logical	:logical
FALSE:2515	FALSE:1310	FALSE:1050	FALSE:2092	FALSE:1754
NA	TRUE :1205	TRUE :1465	TRUE :423	TRUE :761

### 3.2 2. 查看各数据简短信息

表 7: Table continues below

property_name	property_region	price_ttl	price_sqm
Length:2515	Length:2515	Min. : 10.6	Min. : 1771
Class :character	Class :character	1st Qu.: 95.0	1st Qu.:10765
Mode :character	Mode :character	Median : 136.0	Median :14309
NA	NA	Mean : 154.8	Mean :15110

property_name	property_region	price_ttl	price_sqm
NA	NA	3rd Qu.: 188.0	3rd Qu.:18213
NA	NA	Max. :1380.0	Max. :44656

表 8: Table continues below

bedrooms	livingrooms	building_area	directions1
Min. :1.000	Min. :0.000	Min. : 22.77	Length:2515
1st Qu.:2.000	1st Qu.:1.000	1st Qu.: 84.45	Class :character
Median :3.000	Median :2.000	Median : 95.46	Mode :character
Mean :2.689	Mean :1.706	Mean :100.67	NA
3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:118.03	NA
Max. :7.000	Max. :4.000	Max. :588.66	NA

表 9: Table continues below

directions2	decoration	property_t_height	property_height
Length:2515	Length:2515	Min. : 2.00	Length:2515
Class :character	Class :character	1st Qu.:11.00	Class :character
Mode :character	Mode :character	Median :27.00	Mode :character
NA	NA	Mean :24.05	NA
NA	NA	3rd Qu.:33.00	NA
NA	NA	Max. :62.00	NA

表 10: Table continues below

property_style	followers	near_subway	if_2y
Length:2515	Min. : 0.000	Length:2515	Length:2515
Class :character	1st Qu.: 1.000	Class :character	Class :character
Mode :character	Median : 2.000	Mode :character	Mode :character

property_style	followers	near_subway	if_2y
NA	Mean : 6.326	NA	NA
NA	3rd Qu.: 6.000	NA	NA
NA	Max. :262.000	NA	NA

has_key	vr
Length:2515	Length:2515
Class :character	Class :character
Mode :character	Mode :character
NA	NA
NA	NA
NA	NA

总结数据特征：2023 年 9 月 12 号链家武汉二手房数据特征：##（1）房屋总价：## 最低价 10.60 万元。

```
##  property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1      农垦街      汉南其它      10.6      1984      1      1
##  building_area directions1 directions2 decoration property_t_height
## 1      53.43      南      简装      5
##  property_height property_style followers near_subway if_2y has_key vr
## 1      中      暂无数据      3      <NA> <NA> 随时看房 VR看装修
```

## 最高价 1,380.00 万元。

```
##  property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1 万科高尔夫城市花园      金银湖      1380      23444      7      3
##  building_area directions1 directions2 decoration property_t_height
## 1      588.66      南      北      精装      3
##  property_height property_style followers near_subway if_2y has_key vr
## 1      <NA>      板塔结合      32      <NA> <NA> 随时看房 <NA>
```

## 中位数 136.00 万元，平均数 154.80 万元，数据分布呈右偏分布。



## (2) 房屋单价 ## 最低价 1,771 元/ $m^2$

```
## property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1 天益华庭 阳逻 17 1771 2 2
## building_area directions1 directions2 decoration property_t_height
## 1 96.03 东南 毛坯 7
## property_height property_style followers near_subway if_2y has_key vr
## 1 高 塔楼 11 <NA> <NA> 随时看房 VR看装修
```

## 最高价 44,656 元/ $m^2$

```
## property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1 中商宿舍 中南丁字桥 320 44656 2 1
## building_area directions1 directions2 decoration property_t_height
## 1 71.66 南 北 精装 7
## property_height property_style followers near_subway if_2y has_key vr
## 1 低 板楼 1 近地铁 <NA> 随时看房 VR看装修
```

## 中位数 14,309 元/ $m^2$ ，平均数元 15,110/ $m^2$ ，数据分布呈现右偏分布，  
因为房屋面积不同，整体单价的右偏幅度较房屋总价较小。

## (3) 建筑面积 ## 最小面积 22.77 $m^2$

```
## property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1 金色港湾MIMI空间 沌口 15 6588 1 1
## building_area directions1 directions2 decoration property_t_height
## 1 22.77 南 精装 31
## property_height property_style followers near_subway if_2y has_key vr
## 1 高 板楼 2 近地铁 <NA> 随时看房 <NA>
```

## 最大面积 588.66 $m^2$

```
## property_name property_region price_ttl price_sqm bedrooms livingrooms
## 1 万科高尔夫城市花园 金银湖 1380 23444 7 3
## building_area directions1 directions2 decoration property_t_height
## 1 588.66 南 北 精装 3
## property_height property_style followers near_subway if_2y has_key vr
```

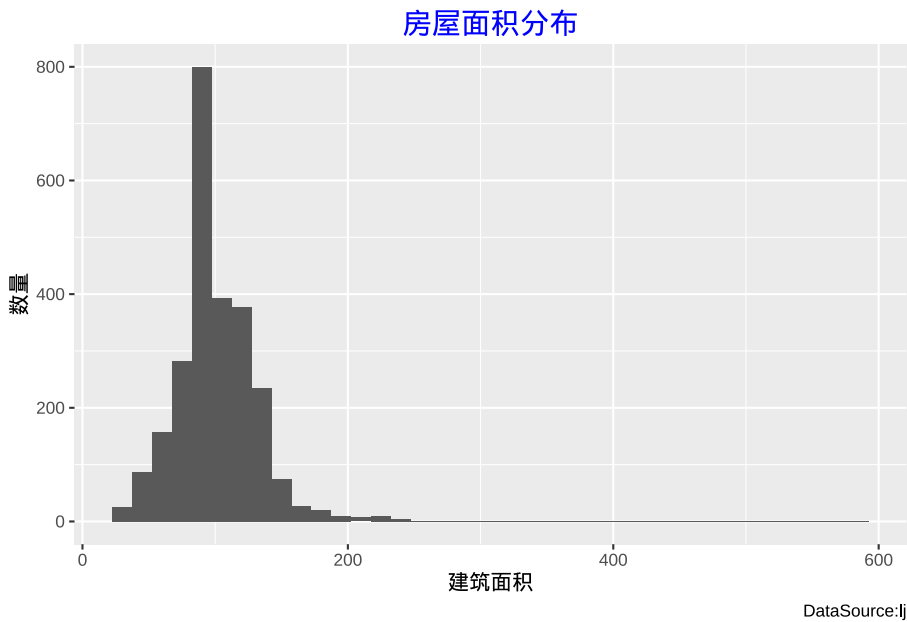
## 1 <NA> 板塔结合 32 <NA> <NA> 随时看房 <NA>

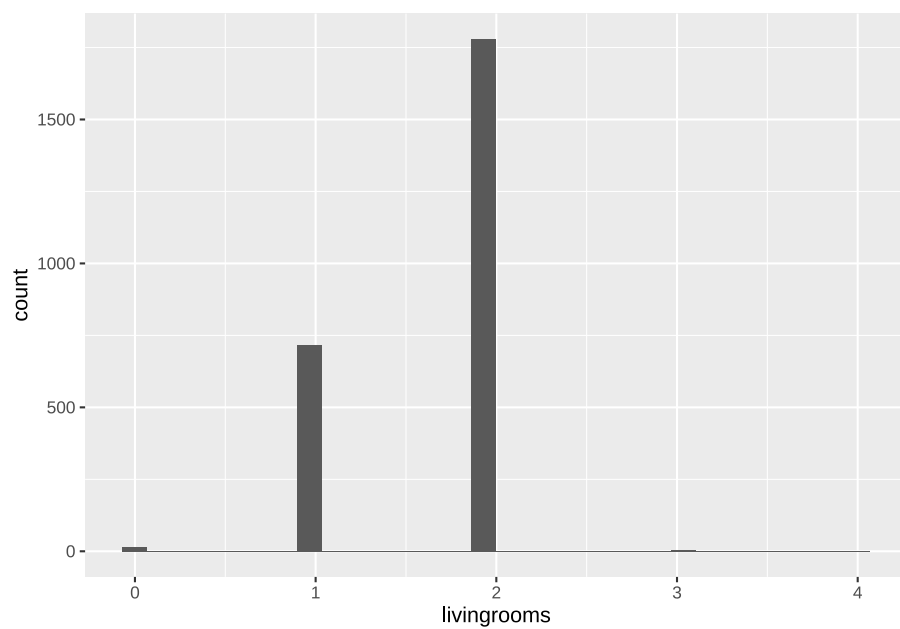
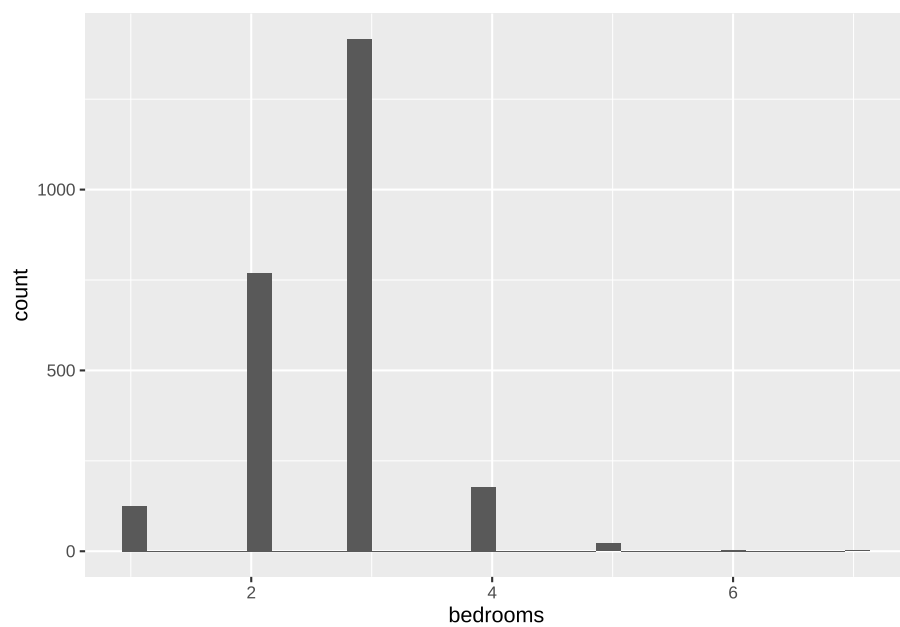
## 中位数  $95.46m^2$ ，平均数  $100.67m^2$ ，呈现右偏分布。

## (4) 其他二手房数据特征：房间数：以 2~3 居室居多客厅数：以 1~2 客厅居多楼栋整层数：平均楼层为 24 层的中层建筑

4 四. 探索性研究

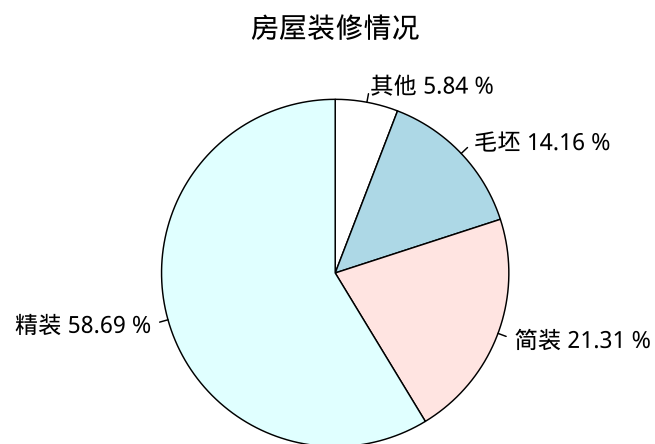
4.1 1. 二手房主要类型特征





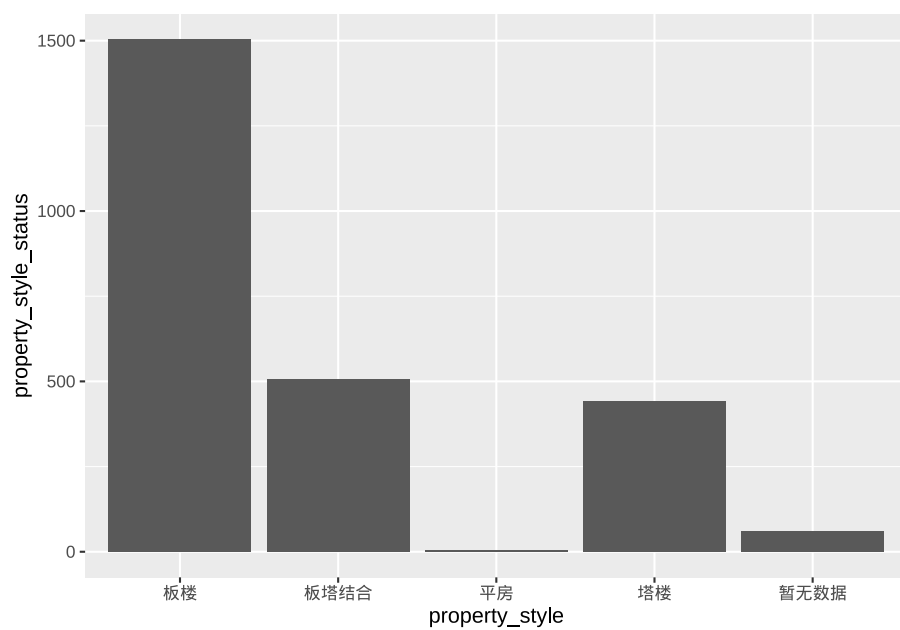
结论：-二手房面积集中在  $100m^2$ ；-以三室两厅房型为主。

##2. 房屋装修情况



结论：装修过的二手房占比近 80%，其中精装数占比 50% 以上。

### ##3. 楼房类型分布

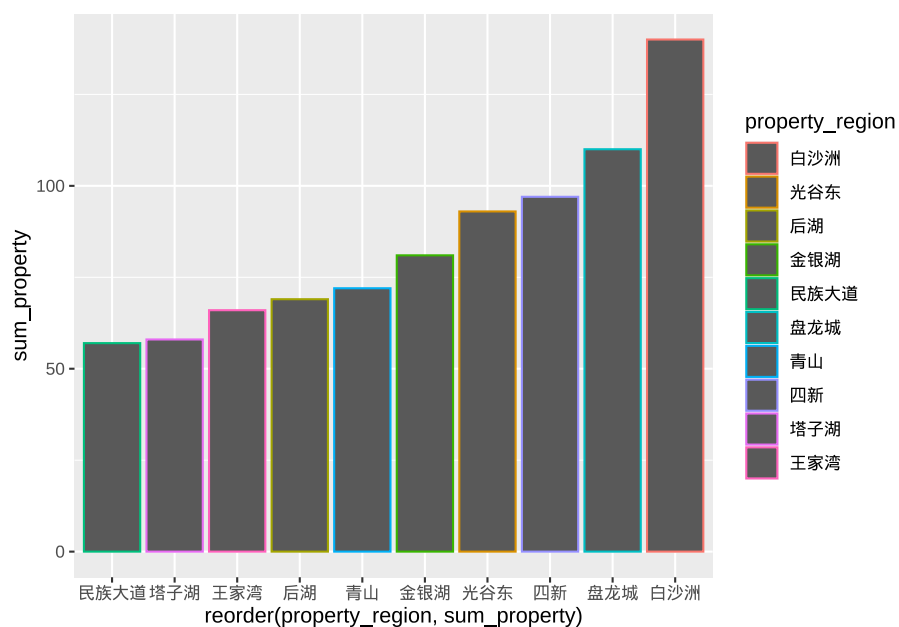


```
## # A tibble: 5 x 2
##   property_style property_style_status
##   <chr>           <int>
## 1 塔楼             441
## 2 平房             4
## 3 暂无数据         61
## 4 板塔结合        506
## 5 板楼           1503
```

结论：一板楼类型最多 1781 套，其次是板塔结合 615 套。

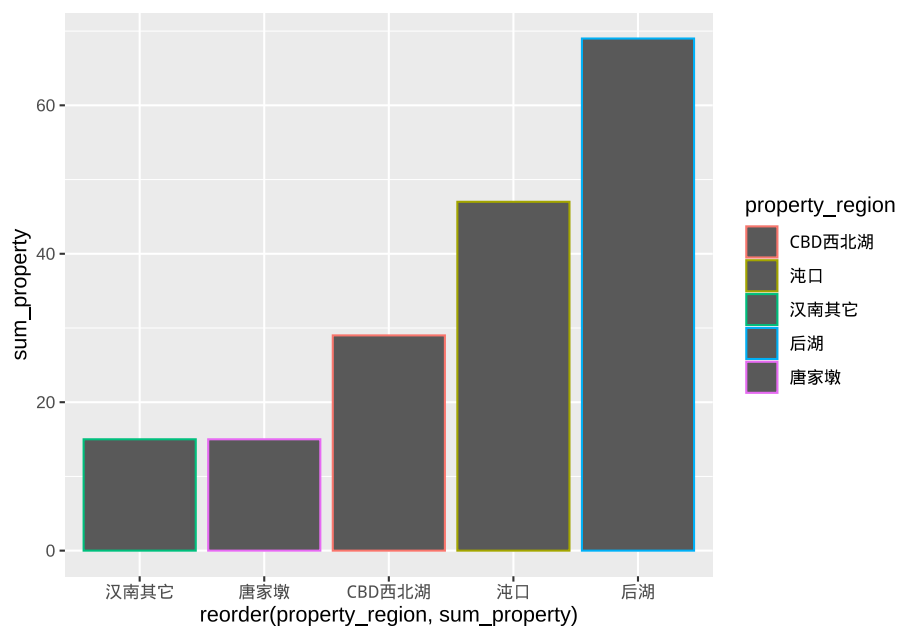
## 4.2 4. 探索问题

### 问题 1：二手房有何区域分布特征？



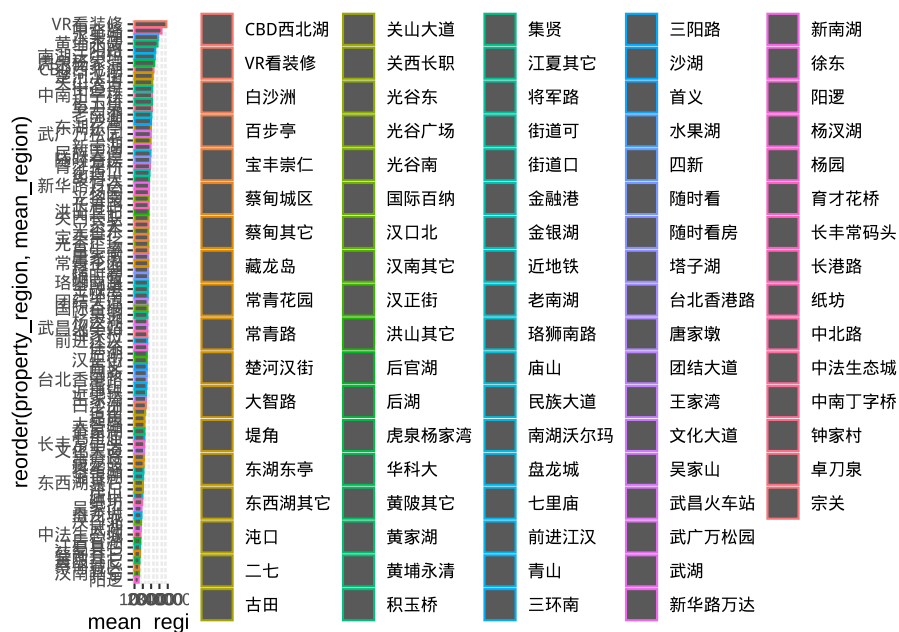
```
## # A tibble: 10 x 2
##   property_region sum_property
##   <chr>           <int>
## 1 白沙洲           140
```

```
## 2 盘龙城      110
## 3 四新        97
## 4 光谷东      93
## 5 金银湖      81
## 6 青山        72
## 7 后湖        69
## 8 王家湾      66
## 9 塔子湖      58
## 10 民族大道   57
```



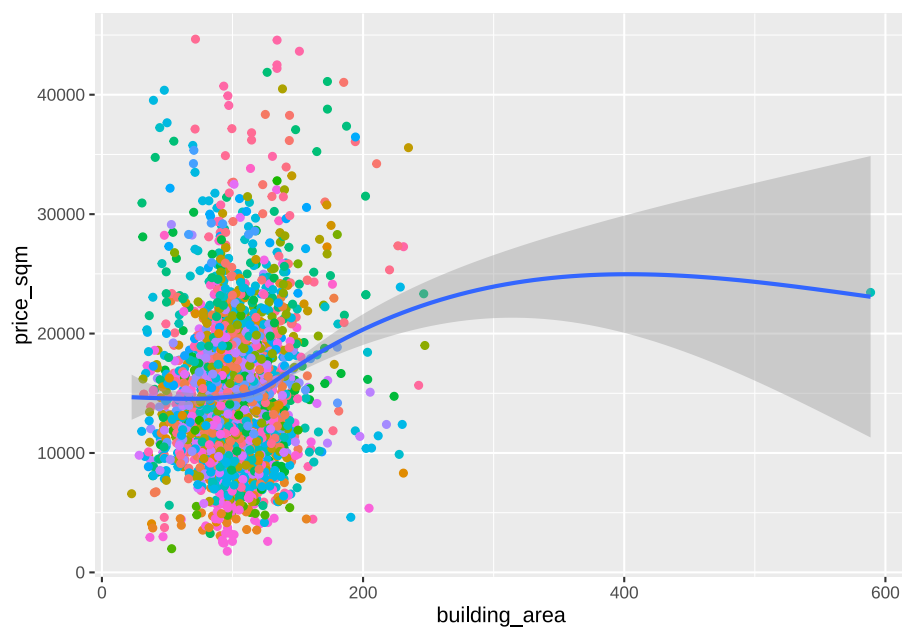
```
## # A tibble: 5 x 2
##   property_region sum_property
##   <chr>           <int>
## 1 后湖             69
## 2 唐家墩          15
## 3 汉南其它        15
## 4 CBD西北湖      29
## 5 沌口            47
```

## 4.3 问题 2：区域对房屋均价的影响？



结论：-武汉在售二手房地区房屋均价在 15000~20000 元间；-均价最高值：中北路约 30000 元；-均价最低值：阳逻约 5000 元。

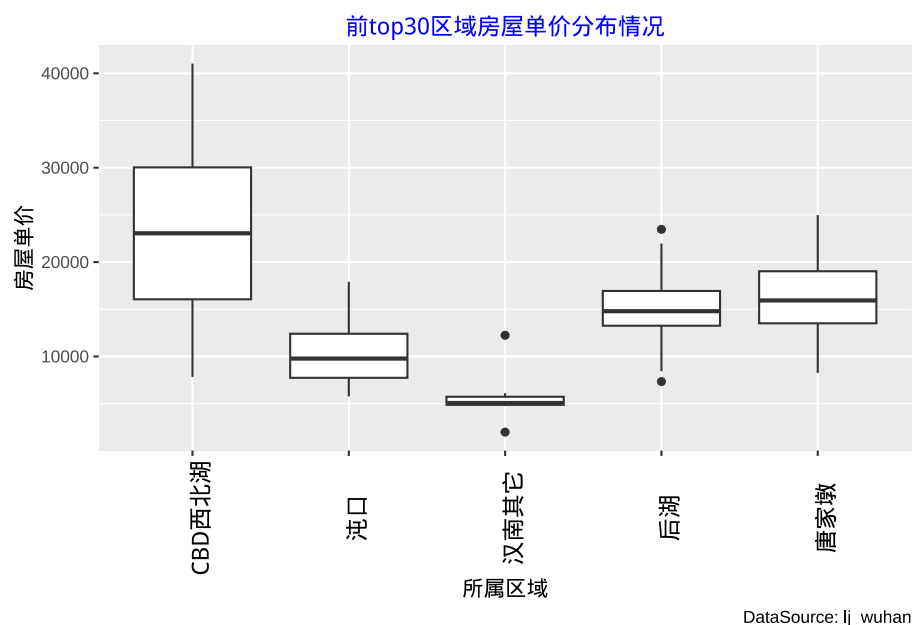
## 问题 3：房屋面积对单价的影响



**结论:** -面积约在  $120m^2$  内的房屋, 面积的增长对单价影响较平稳; -面积大于  $350m^2$  的房屋, 单价随面积增加略有下降; -面积在  $120m^2$  至  $350m^2$  区间, 单价随面积上升幅度较大; -分析极值点的离散情况, 高房价存在个别极值, 说明离散度较高。



#### 4.4 问题 4: 区域房屋单价分布情况



**结论:** 其中光谷、青山、王家湾等区域房屋单价离散度较高; 盘龙城、四新等区域房屋单价集中性度较高。

## 5 五. 决策建议:

##1, 房地产商: 从二手房面积方面来看, 集中在  $100m^2$ , 以三室两厅房型为主, 此类房型因为交易度最活跃, 相应可以投入更多建设。从建筑类型来看, 二手房板楼类型更多, 因为其南北通透户型方正, 交易度更活跃, 相应可以投入更多建设。

##2, 卖家: 面积大于  $350m^2$  的房屋, 单价随面积增加略有下降, 因此该面积范围内的住房不适宜在二手房市场流通。

##3, 买家: 面积约在  $120m^2$  内的房屋, 面积的增长对单价影响较平稳, 而面积在  $120m^2$  至  $350m^2$  区间, 单价随面积上升幅度较大, 因此买家在购买面积超过  $120m^2$  的二手房, 尽量按实际需求购买。针对区域房屋单价分布

情况, 如有在光谷、青山、王家湾等区域购房需求, 考虑房屋单价离散度较高, 应多多比价选择性价比更高的房屋。