

# 基于 R 语言对链家武汉二手房交易价格进行统计分析

孙雪菲

## 目录

1 你的主要发现	1
2 数据介绍	2
3 探索性分析	12
3.1 变量 1【建筑面积】的数值描述与图形	12
3.2 变量 2【房源数量】的数值描述与图形	14
3.3 变量 3【单位均价】的数值描述与图形	15
3.4 变量 4【卧室个数】的数值描述与图形	17
4 发现总结	21

## 1 你的主要发现

武汉二手房交易均价的影响因素多元，主要可以分为以下三类：

1、供需： 2、区位： 3、产品力。

其中供需影响分析主要考虑从房源数量与交易均价的波动分析引入；区分影响分析主要考虑从各小区所在地区对应行政区划分比较；产品力分析受限于篇幅及 r 工具使用熟练度，主要考虑从业态、户型两方面进行分析。【发现 1】供需：房源较多的区域，单位均价偏低，呈现出负相关。【发现 2】区

位：不同区域二手房的单位均价差别很大，武昌区房价远高于其他区域。二手房的区域位置十分重要，地处经济发达城区的二手房具备地理优势，单位均价更高。可以推测区域是影响房价的极其重要的因素之一。【发现 3】产品力：不同业态二手房的单位均价差别很大，各区域超高层价格普遍高于其他业态；不同户型二手房的单位均价差别很大，普遍大户型均价高于其他户型。推测原因可能为大户型通常定位为改善型客户，因此无论是品质或者成本均通常高于刚需型。

## 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。

```
## [1] TRUE

## # A tibble: 3,000 x 19
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>   <dbl>   <dbl>     <dbl>
## 1 南湖名都A区     南湖沃尔玛      237    18709     3         1
## 2 万科紫悦湾      光谷东          127    14613     3         2
## 3 东立国际        二七             75    15968     1         1
## 4 新都汇          光谷广场        188    15702     3         2
## 5 保利城一期      团结大道        182    17509     3         2
## 6 加州橘郡        庙山            122    10376     3         2
## 7 省建筑五公司西区 光谷广场         99    12346     2         1
```

```
## 8 保利上城东区    白沙洲    194.    16336    3    2
## 9 石化大院        中南丁字桥    325    32631    4    1
## 10 阳光花园        杨汊湖    192    17403    3    2
## # i 2,990 more rows
## # i 13 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>
```

# 整理数据-按总楼层区分新增变量【业态】

```
## # A tibble: 3,000 x 20
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>            <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 南湖名都A区      南湖沃尔玛      237    18709      3      1
## 2 万科紫悦湾      光谷东          127    14613      3      2
## 3 东立国际        二七             75    15968      1      1
## 4 新都汇          光谷广场        188    15702      3      2
## 5 保利城一期      团结大道        182    17509      3      2
## 6 加州橘郡        庙山            122    10376      3      2
## 7 省建筑五公司西区 光谷广场         99    12346      2      1
## 8 保利上城东区    白沙洲          194.    16336      3      2
## 9 石化大院        中南丁字桥      325    32631      4      1
## 10 阳光花园       杨汊湖          192    17403      3      2
## # i 2,990 more rows
## # i 14 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>
```

# 整理数据-按总建筑面积区分新增变量【户型】

```
## # A tibble: 3,000 x 21
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
```

```

##      <chr>          <chr>          <dbl>      <dbl>      <dbl>      <dbl>
##  1 南湖名都A区      南湖沃尔玛      237      18709        3        1
##  2 万科紫悦湾      光谷东      127      14613        3        2
##  3 东立国际        二七      75      15968        1        1
##  4 新都汇          光谷广场      188      15702        3        2
##  5 保利城一期      团结大道      182      17509        3        2
##  6 加州橘郡        庙山      122      10376        3        2
##  7 省建筑五公司西区 光谷广场      99      12346        2        1
##  8 保利上城东区      白沙洲      194.      16336        3        2
##  9 石化大院        中南丁字桥      325      32631        4        1
## 10 阳光花园        杨汊湖      192      17403        3        2
## # i 2,990 more rows
## # i 15 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>

# 整理数据-计算平均每间面积并新增变量, 利用 IQR 识别建筑面积异常值

## # A tibble: 3,000 x 22
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>            <chr>          <dbl>    <dbl>    <dbl>    <dbl>
##  1 南湖名都A区      南湖沃尔玛      237      18709        3        1
##  2 万科紫悦湾      光谷东      127      14613        3        2
##  3 东立国际        二七      75      15968        1        1
##  4 新都汇          光谷广场      188      15702        3        2
##  5 保利城一期      团结大道      182      17509        3        2
##  6 加州橘郡        庙山      122      10376        3        2
##  7 省建筑五公司西区 光谷广场      99      12346        2        1
##  8 保利上城东区      白沙洲      194.      16336        3        2
##  9 石化大院        中南丁字桥      325      32631        4        1
## 10 阳光花园        杨汊湖      192      17403        3        2
## # i 2,990 more rows

```

```
## # i 16 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>
```

```
## # A tibble: 3,000 x 23
```

	property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	南湖名都A区	南湖沃尔玛	237	18709	3	1
## 2	万科紫悦湾	光谷东	127	14613	3	2
## 3	东立国际	二七	75	15968	1	1
## 4	新都汇	光谷广场	188	15702	3	2
## 5	保利城一期	团结大道	182	17509	3	2
## 6	加州橘郡	庙山	122	10376	3	2
## 7	省建筑五公司西区	光谷广场	99	12346	2	1
## 8	保利上城东区	白沙洲	194.	16336	3	2
## 9	石化大院	中南丁字桥	325	32631	4	1
## 10	阳光花园	杨汊湖	192	17403	3	2

```
## # i 2,990 more rows
```

```
## # i 17 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>,
## #   outlier <chr>
```

# 整理数据-利用 IQR 识别单位均价异常值

```
## # A tibble: 3,000 x 24
```

	property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	南湖名都A区	南湖沃尔玛	237	18709	3	1
## 2	万科紫悦湾	光谷东	127	14613	3	2
## 3	东立国际	二七	75	15968	1	1

```
## 4 新都汇          光谷广场          188      15702      3          2
## 5 保利城一期      团结大道          182      17509      3          2
## 6 加州橘郡        庙山              122      10376      3          2
## 7 省建筑五公司西区 光谷广场          99       12346      2          1
## 8 保利上城东区    白沙洲            194.     16336      3          2
## 9 石化大院        中南丁字桥        325      32631      4          1
## 10 阳光花园       杨汊湖            192      17403      3          2
## # i 2,990 more rows
## # i 18 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>,
## #   outlier <chr>, outlier2 <chr>
```

# 整理数据-查找为空单元格所在列名称, 该 6 个字段信息不全不建议用作分析

```
## [1] "directions2"      "property_height" "near_subway"     "if_2y"
## [5] "has_key"          "vr"
```

# 整理数据-查找所在位置字段信息有误

```
## [1] "南湖沃尔玛" "光谷东"      "二七"        "光谷广场"    "团结大道"
## [6] "庙山"        "白沙洲"      "中南丁字桥" "杨汊湖"      "中北路"
## [11] "百步亭"      "堤角"        "民族大道"    "阳逻"        "盘龙城"
## [16] "长丰常码头" "青山"        "四新"        "藏龙岛"      "汉南其它"
## [21] "后湖"        "金银湖"      "王家湾"      "三环南"      "长港路"
## [26] "沌口"        "七里庙"      "虎泉杨家湾" "珞狮南路"    "街道可"
## [31] "古田"        "东西湖其它" "武昌火车站" "黄埔永清"    "蔡甸城区"
## [36] "徐东"        "钟家村"      "常青花园"    "文化大道"    "关山大道"
## [41] "台北香港路" "光谷南"      "金融港"      "新南湖"      "杨园"
## [46] "武湖"        "纸坊"        "塔子湖"      "中法生态城" "卓刀泉"
## [51] "关西长职"    "宝丰崇仁"    "东湖东亭"    "唐家墩"      "黄家湖"
## [56] "首义"        "随时看房"    "老南湖"      "常青路"      "三阳路"
```

```
## [61] "积玉桥"      "武广万松园" "楚河汉街"   "育才花桥"   "华科大"
## [66] "新华路万达" "沙湖"        "近地铁"     "洪山其它"   "后官湖"
## [71] "蔡甸其它"    "国际百纳"    "吴家山"     "CBD西北湖"  "前进江汉"
## [76] "街道口"      "汉正街"      "集贤"        "宗关"        "江夏其它"
## [81] "随时看"      "大智路"      "黄陂其它"   "将军路"     "水果湖"
## [86] "汉口北"      "VR看装修"
```

```
## # A tibble: 3,000 x 24
```

```
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>    <dbl>    <dbl>    <dbl>
## 1 南湖名都A区      南湖沃尔玛        237      18709      3         1
## 2 万科紫悦湾      光谷东            127      14613      3         2
## 3 东立国际        二七              75       15968      1         1
## 4 新都汇          光谷广场          188      15702      3         2
## 5 保利城一期      团结大道          182      17509      3         2
## 6 加州橘郡        庙山              122      10376      3         2
## 7 省建筑五公司西区 光谷广场          99       12346      2         1
## 8 保利上城东区    白沙洲            194      16336      3         2
## 9 石化大院        中南丁字桥        325      32631      4         1
## 10 阳光花园       杨汉湖            192      17403      3         2
```

```
## # i 2,990 more rows
```

```
## # i 18 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>,
## #   outlier <chr>, outlier2 <chr>
```

```
# 筛选数据-非异常非字段信息有误数据另存为数据源
```

```
## # A tibble: 2,828 x 24
```

```
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>    <dbl>    <dbl>    <dbl>
## 1 南湖名都A区      南湖沃尔玛        237      18709      3         1
## 2 万科紫悦湾      光谷东            127      14613      3         2
```

```
## 3 东立国际      二七      75      15968      1      1
## 4 新都汇      光谷广场      188      15702      3      2
## 5 保利城一期      团结大道      182      17509      3      2
## 6 加州橘郡      庙山      122      10376      3      2
## 7 省建筑五公司西区 光谷广场      99      12346      2      1
## 8 保利上城东区      白沙洲      194.      16336      3      2
## 9 阳光花园      杨汊湖      192      17403      3      2
## 10 东湖1号      中北路      630      27273      5      2
## # i 2,818 more rows
## # i 18 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>,
## #   outlier <chr>, outlier2 <chr>
```

# 整理数据-按所在地区对照“武汉三镇及远城区”新增变量【区域】

```
## # A tibble: 2,828 x 25
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <chr>           <chr>           <dbl>    <dbl>    <dbl>    <dbl>
## 1 南湖名都A区      南湖沃尔玛      237      18709      3      1
## 2 万科紫悦湾      光谷东      127      14613      3      2
## 3 东立国际      二七      75      15968      1      1
## 4 新都汇      光谷广场      188      15702      3      2
## 5 保利城一期      团结大道      182      17509      3      2
## 6 加州橘郡      庙山      122      10376      3      2
## 7 省建筑五公司西区 光谷广场      99      12346      2      1
## 8 保利上城东区      白沙洲      194.      16336      3      2
## 9 阳光花园      杨汊湖      192      17403      3      2
## 10 东湖1号      中北路      630      27273      5      2
## # i 2,818 more rows
## # i 19 more variables: building_area <dbl>, directions1 <chr>,
## #   directions2 <chr>, decoration <chr>, property_t_height <dbl>,
```



```
## #   property_height <chr>, property_style <chr>, followers <dbl>,
## #   near_subway <chr>, if_2y <chr>, has_key <chr>, vr <chr>, House_type <chr>,
## #   Building_type <chr>, Property_type <chr>, Area_per_room <dbl>,
## #   outlier <chr>, outlier2 <chr>, Area <chr>
```

## 各变量的简短信息:

在经过异常值处理后数据集 (lj2023\_1) 一共有 2828 条, 每条代表一个房源信息。数据集 (lj2023\_1) 集一共涉及 25 个变量 (含新增变量区域、户型、建筑类别等), 其中文本类数据 17 列、数值类数据 8 列。本次数据抓取 1278 个小区分布于 82 个地区, 考虑分组分析对比故将地区整合为武汉三镇及远城区。本次分析对象是武汉市链家平台二手房交易价格, 所以将单位均价 (price\_sqm) 视为因变量。区域、户型、建筑类别等作为自变量, (directions2、near\_subway、if\_2y、vr 列变量缺失值较多所以本次价格影响因素分析重点不放在这几个变量上)

```
## Rows: 2,828
## Columns: 25
## $ 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, 3, 5, 3, 3, 3, 2, 3, 4, 3, 3~
## $ livingrooms        <dbl> 1, 2, 1, 2, 2, 2, 1, 2, 2, 2, 2, 1, 2, 2, 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", NA, NA, "东", NA, "北", NA, "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, 7, 25, 32, 31, 44, 19~
## $ 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, 0, 0, 0, 0, 0, 1, 0, 1, 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", ~
## $ House_type  <chr> "三室一厅", "三室二厅", "一室一厅", "三室二厅", "三~
## $ Building_type <chr> "小高层", "高层", "小高层", "高层", "超高层", "超高~
## $ Property_type <chr> "中户型", "小户型", "小户型", "中户型", "中户型", "~
## $ Area_per_room <dbl> 31.67000, 17.38200, 23.48500, 23.94600, 20.79000, 23~
## $ outlier     <chr> "正常值", "正常值", "正常值", "正常值", "正常值", "~
## $ outlier2    <chr> "正常值", "正常值", "正常值", "正常值", "正常值", "~
## $ Area        <chr> "武昌区", "远城区", "汉口区", "远城区", "远城区", "~

## 各变量的简短统计:
```

表 1: Data summary

Name	lj2023_1
Number of rows	2828
Number of columns	25
Column type frequency:	
character	17
numeric	8
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
property_name	0	1.00	2	16	0	1278	0
property_region	0	1.00	2	6	0	82	0
directions1	0	1.00	1	2	0	8	0
directions2	1579	0.44	1	2	0	8	0
decoration	0	1.00	2	2	0	4	0
property_height	55	0.98	1	1	0	3	0
property_style	0	1.00	2	4	0	5	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
near_subway	1389	0.51	3	5	0	3	0
if_2y	1613	0.43	5	5	0	1	0
has_key	423	0.85	3	5	0	9	0
vr	824	0.71	3	5	0	11	0
House_type	0	1.00	2	4	0	13	0
Building_type	0	1.00	2	3	0	5	0
Property_type	0	1.00	3	3	0	3	0
outlier	0	1.00	3	3	0	1	0
outlier2	0	1.00	3	3	0	1	0
Area	0	1.00	3	3	0	4	0

### Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
price_ttl	0	1	145.08	70.32	10.60	94.75	134.00	180.00	630.00	
price_sqm	0	1	14462.32	111.21	1771.00	689.00	1107.00	780.25	254.00	
bedrooms	0	1	2.68	0.70	1.00	2.00	3.00	3.00	6.00	
livingrooms	0	1	1.72	0.46	0.00	1.00	2.00	2.00	3.00	
building_area0	0	1	99.56	26.09	28.55	85.02	95.02	116.51	231.00	
property_t_height	0	1	24.26	12.41	2.00	11.00	27.50	33.00	62.00	
followers	0	1	6.31	14.90	0.00	1.00	2.00	6.00	262.00	
Area_per_room	0	1	22.77	3.75	12.30	19.98	22.43	25.03	33.34	

通过数据概览可以直观看到：

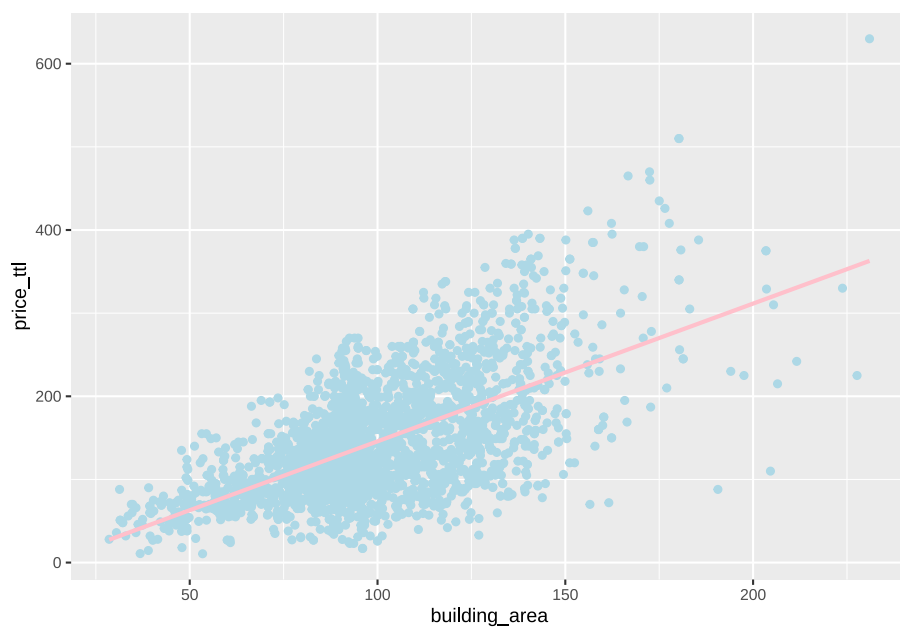
- **【直观结论 1】：**房屋总价（price\_ttl）平均值 145 万，最低价 11 万，中位数 134 万，最高价 630 万，超过 75% 的房源在 180 万以下。
- **【直观结论 2】：**房屋均价（price\_sqm）平均值 1.4 万元/平，中位数 1.4 万元/平，最高 2.9 万元/平，超过 75% 的房源在 1.8 万元/平以下。

- **【直观结论 3】**: 建筑面积 (building\_area) 平均值 100 平米, 中位数 95 平米, 最大 231 平米, 超过 75% 的房源在 117 平米以下。

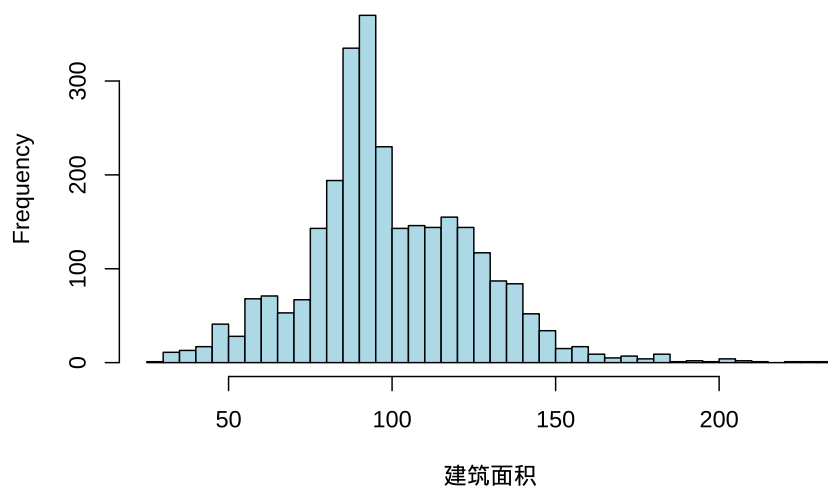
## 3 探索性分析

### 3.1 变量 1【建筑面积】的数值描述与图形

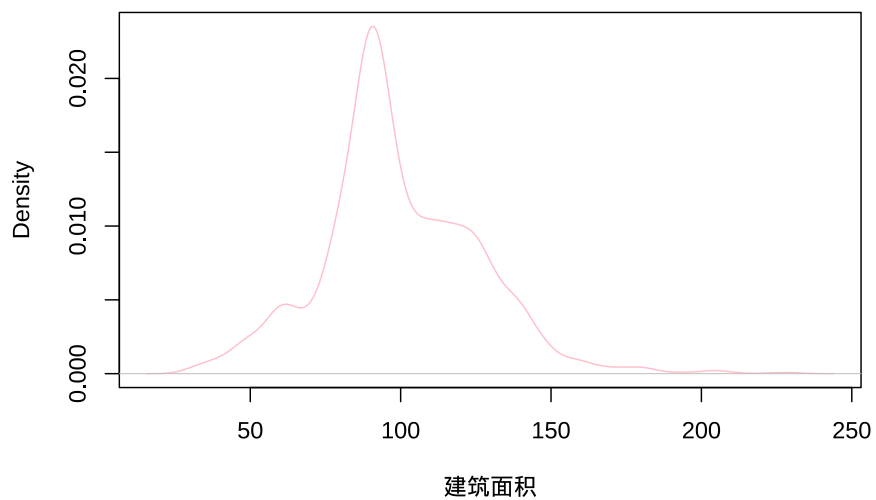
发现: 根据**【建筑面积与房屋总价图】**整体来看, 建筑面积越大, 房屋总价越高, 斜率代表单位均价。根据**【建筑面积直方图与概率密度曲线图】**, 建筑面积的分布为右偏分布, 相比于正态分布更加陡峭。建筑面积主要集中在 80-100 平米。**【建筑面积与单位均价图】**, 建筑面积与单位均价分布呈现散射状, 有一定正向关系, 但不呈现线性关系, 因此房价的影响因素应该是多元的。

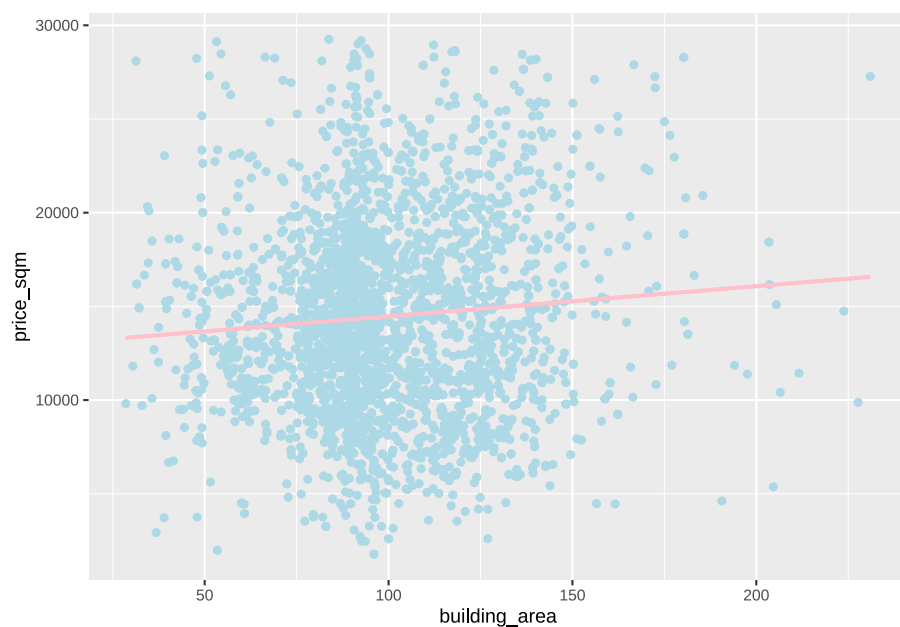


建筑面积直方图



建筑面积概率密度曲线

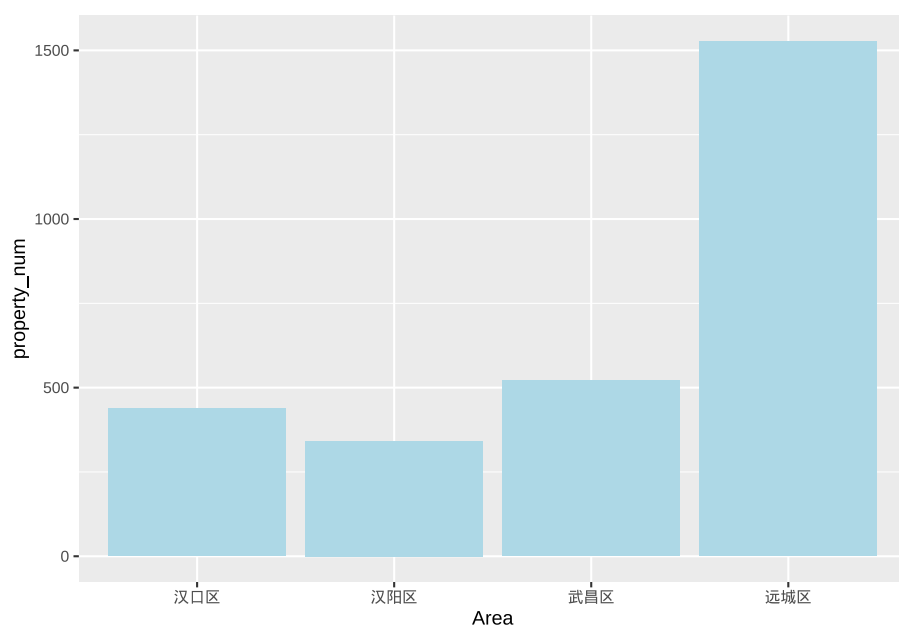




### 3.2 变量 2【房源数量】的数值描述与图形

发现：远城区房源数量远高于武汉三镇中心城区。经查询相关政策可知，早期中心城区全部限购，远城区都不限购。远城区想迁入中心城区必须参照外地人落户的政策执行；中心城区迁远城区，可以直接按武汉市内迁转的政策执行；远城区的落户主城区需要在主城区购买 100 平方以上的住宅房产。因此推断远城区二手房房源数量远高于武汉三镇原因除了辖区范围大，户口及购房政策因素也存在部分影响。

```
## # A tibble: 4 x 3
##   Area      property_num propertyprice_sqm
##   <chr>          <int>          <dbl>
## 1 武昌区             521           17303.
## 2 汉口区             438           15534.
## 3 汉阳区             342           13922.
## 4 远城区            1527           13491.
```

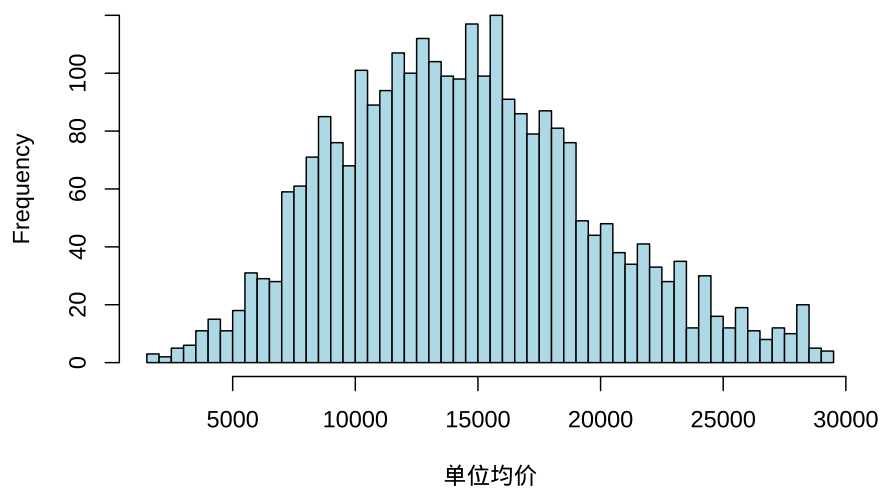


```
## $x
## [1] "区域"
##
## $y
## [1] "房源数量"
##
## $title
## [1] "房源数量分布"
##
## attr(,"class")
## [1] "labels"
```

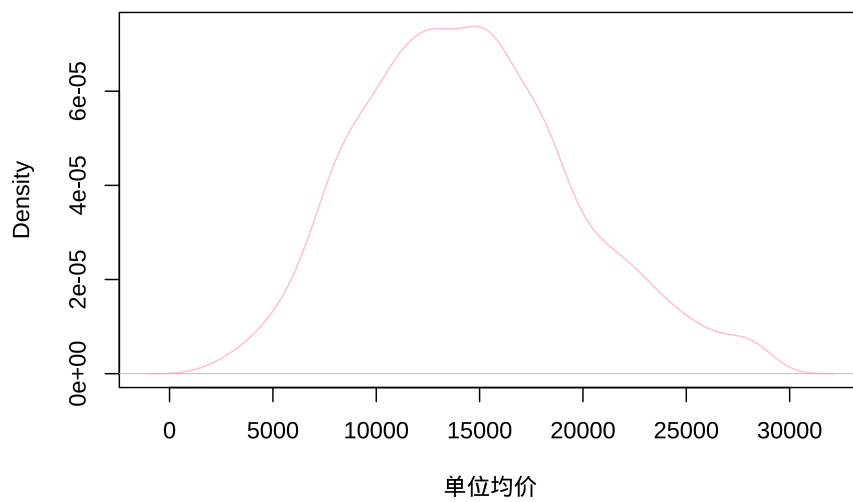
### 3.3 变量 3【单位均价】的数值描述与图形

发现：根据下图可以发现，单位均价的分布为右偏分布，相比于正态分布更加陡峭。单位均价主要集中在 1 万-2 万元/平米。整体均价分布从 0.2 万-2.9 万，跨度较大。相比较而言，高价位的房源数量要远低于中等价位。

单位均价直方图



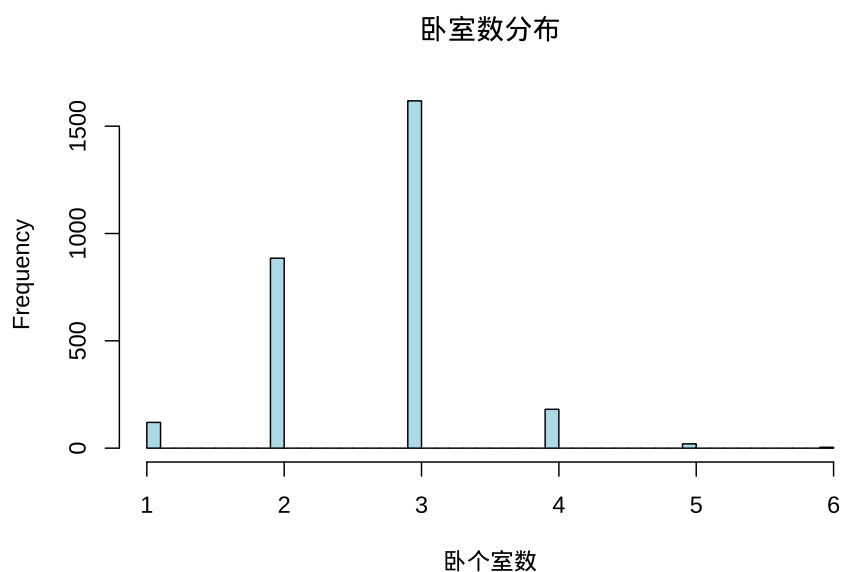
单位均价概率密度曲线



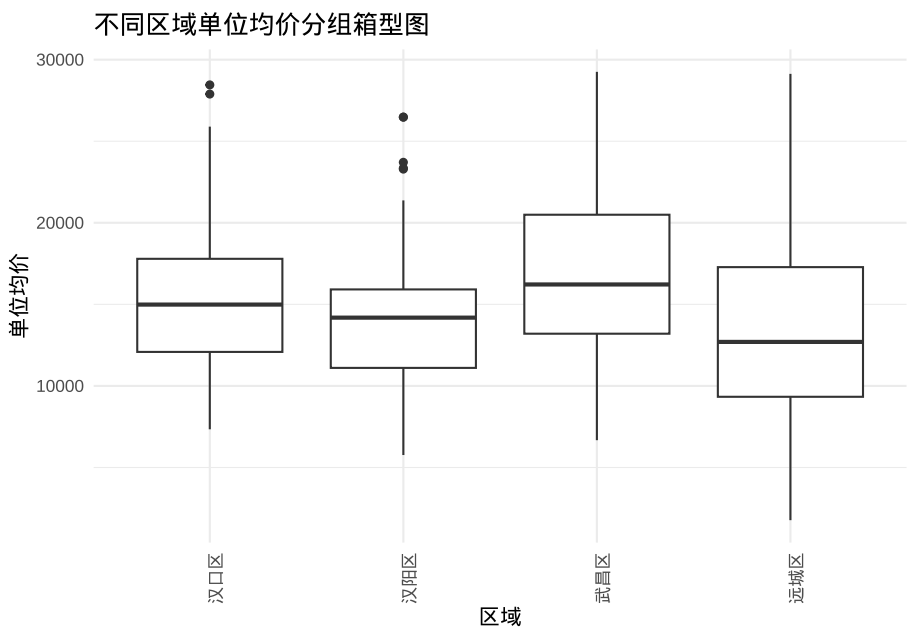


### 3.4 变量 4【卧室个数】的数值描述与图形

发现：三室的房源较多，符合常规家庭成员结构。



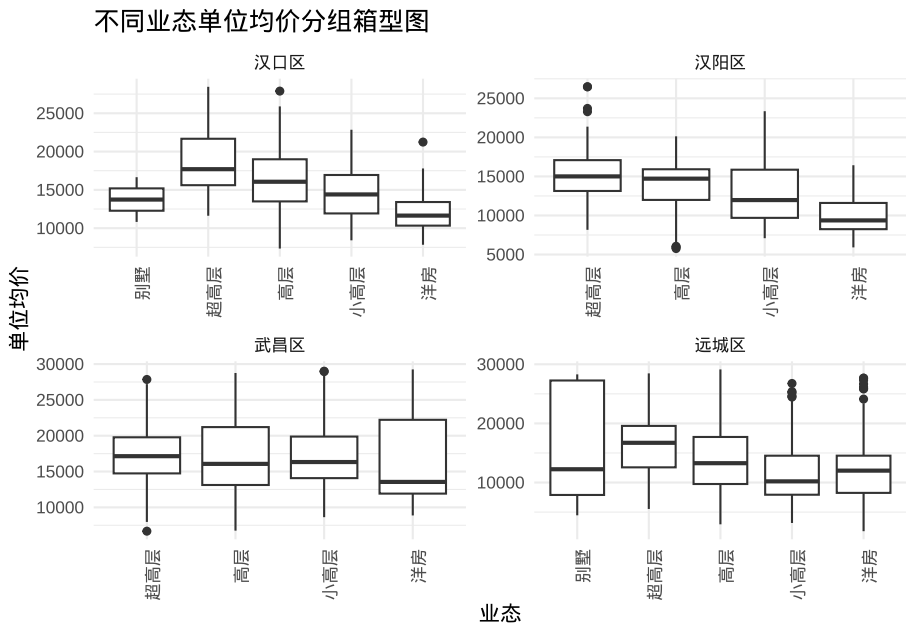
## 探索问题 1：单位均价是否受所在区域影响【不同区域的单位均价分组箱线图】发现：不同区域二手房的单位均价差别很大，武昌区房价远高于其他地区。二手房的区域位置十分重要，地处经济发达城区的人二手房具备地理优势，单位均价更高。可以推测区域是影响房价的极其重要的因素之一



```
##
## Call:
## lm(formula = price_sqm ~ Area, data = lj2023_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11685.0  -3763.9  -497.7   3188.6  15674.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15324.7      239.6   63.971  < 2e-16 ***
## Area汉阳区   -1561.0      361.8   -4.315 1.65e-05 ***
## Area武昌区    1820.8      325.0    5.602 2.32e-08 ***
## Area远城区   -1868.7      271.7   -6.877 7.52e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5014 on 2824 degrees of freedom
```

```
## Multiple R-squared:  0.07542,    Adjusted R-squared:  0.07443
## F-statistic: 76.78 on 3 and 2824 DF,  p-value: < 2.2e-16
```

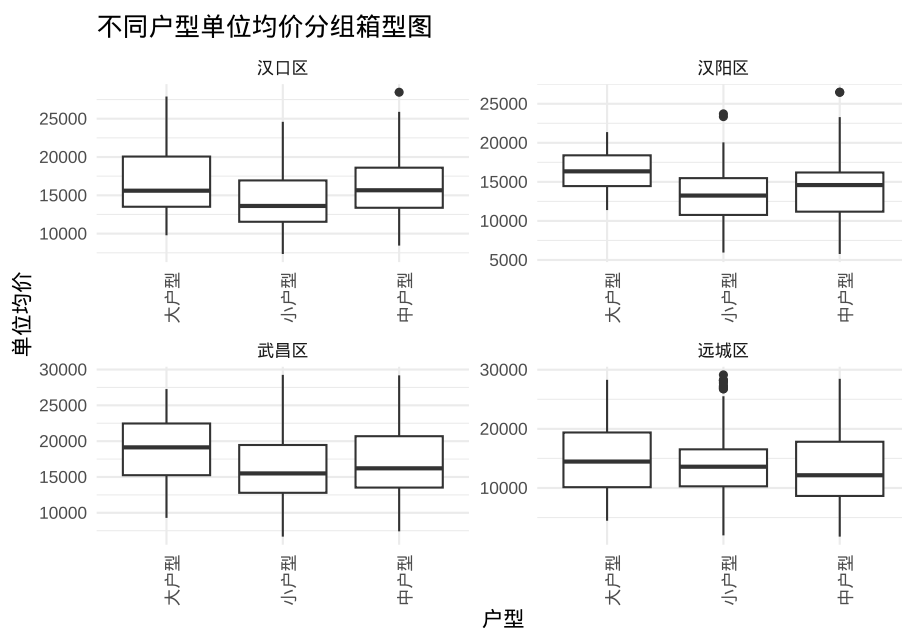
## 探索问题 2：单位均价是否受业态影响【不同业态单位均价分组箱型图】发现：不同业态二手房的单位均价差别很大，各区域超高层价格普遍高于其他业态。推测超高层可能建设年份偏后，地价高于早点楼盘。



```
##
## Call:
## lm(formula = price_sqm ~ Building_type, data = lj2023_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11902.0  -3516.6   -342.9   2974.7  16830.3
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      15066      1506  10.004  <2e-16 ***
## Building_type超高层      1779      1521   1.170   0.2422
```

```
## Building_type高层      -230      1513  -0.152   0.8792
## Building_type小高层    -1830     1521  -1.203   0.2290
## Building_type洋房      -2642     1522  -1.736   0.0826 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4995 on 2823 degrees of freedom
## Multiple R-squared:  0.08261,    Adjusted R-squared:  0.08131
## F-statistic: 63.55 on 4 and 2823 DF,  p-value: < 2.2e-16
```

# 探索问题 3：单位均价是否受户型影响【不同户型单位均价分组箱型图】  
发现：不同户型二手房的单位均价差别很大，各区域大户型价格普遍高于其他业态。推测原因可能为大户型通常定位为改善型客户，因此无论是品质或者成本均通常高于刚需型，所以价格略高。



```
##
## Call:
## lm(formula = price_sqm ~ Property_type, data = lj2023_1)
##
```

```
## Residuals:
##      Min      1Q   Median      3Q      Max
## -12749.3 -3748.5  -303.3   3296.2 15115.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      16438.4      470.1  34.964 < 2e-16 ***
## Property_type小户型 -2299.9      496.9  -4.628 3.85e-06 ***
## Property_type中户型 -1918.1      487.1  -3.938 8.42e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5193 on 2825 degrees of freedom
## Multiple R-squared:  0.007701,   Adjusted R-squared:  0.006999
## F-statistic: 10.96 on 2 and 2825 DF,  p-value: 1.809e-05
```

## 4 发现总结

武汉二手房交易均价的影响因素多元，主要可以分为以下三类：1、供需；2、区位；3、产品力。

其中供需影响分析主要考虑从房源数量与交易均价的波动分析引入；区位影响分析主要考虑从各小区所在地区对应行政区划分比较；产品力分析受限于篇幅及 r 工具使用熟练度，主要考虑从业态、户型两方面进行分析。

1、供需：房源较多的区域，单位均价偏低，呈现出负相关。

2、区位：不同区域二手房的单位均价差别很大，武昌区房价远高于其他区域。二手房的区域位置十分重要，地处经济发达城区的二手房具备地理优势，单位均价更高。可以推测区域是影响房价的极其重要的因素之一。

3、产品力：不同业态二手房的单位均价差别很大，各区域超高层价格普遍高于其他业态；不同户型二手房的单位均价差别很大，普遍大户型均价高于其他户型。推测原因可能为大户型通常定位为改善型客户，因此无论是品质

或者成本均通常高于刚需型。