

武汉市二手房数据分析

龚天纳

- 数据介绍
- 查看数据总体样貌
- 探索性分析
 - 房屋单价的数值描述与图形
 - 房屋朝向的数值描述与图形
- 发现总结

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。

##	[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"

查看数据总体样貌

Data summary	
Name	lj
Number of rows	2515
Number of columns	18
Column type frequency:	
character	11
numeric	7
Group variables	
None	

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
property_name	0	1.00	2	17	0	1345	0
property_region	0	1.00	2	6	0	87	0
directions1	0	1.00	1	2	0	8	0
directions2	1397	0.44	1	2	0	8	0
decoration	0	1.00	2	2	0	4	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
property_height	53	0.98	1	1	0	3	0
property_style	0	1.00	2	4	0	5	0
near_subway	1205	0.52	3	5	0	5	0
if_2y	1465	0.42	5	5	0	1	0
has_key	423	0.83	3	5	0	9	0
vr	761	0.70	3	5	0	11	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
price_ttl	0	1	154.82	94.42	10.60	95.00	136.00	188.00	1380.00	
price_sqm	0	1	15110.42	6347.03	1771.00	10765.00	14309.00	18212.50	44656.00	
bedrooms	0	1	2.69	0.73	1.00	2.00	3.00	3.00	7.00	
livingrooms	0	1	1.71	0.47	0.00	1.00	2.00	2.00	4.00	
building_area	0	1	100.67	30.71	22.77	84.45	95.46	118.03	588.66	
property_t_height	0	1	24.05	12.46	2.00	11.00	27.00	33.00	62.00	
followers	0	1	6.33	15.01	0.00	1.00	2.00	6.00	262.00	

##	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
##			Mean : 154.8	Mean :15110
##			3rd Qu.: 188.0	3rd Qu.:18213
##			Max. :1380.0	Max. :44656
##	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	
##	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:118.03	
##	Max. :7.000	Max. :4.000	Max. :588.66	
##	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
##			Mean :24.05	
##			3rd Qu.:33.00	
##			Max. :62.00	
##	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
##		Mean : 6.326		
##		3rd Qu.: 6.000		
##		Max. :262.000		
##	has_key	vr		
##	Length:2515	Length:2515		
##	Class :character	Class :character		
##	Mode :character	Mode :character		
##				
##				
##				

```
## property_region 的去重值为:
## [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" "黄浦永清" "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"
##
## directions1 的去重值为:
## [1] "南" "北" "东南" "西南" "西北" "东" "东北" "西"
##
## directions2 的去重值为:
## [1] "北" NA "东" "西南" "东南" "南" "西" "西北" "东北"
##
## decoration 的去重值为:
## [1] "精装" "简装" "其他" "毛坯"
##
## property_t_height 的去重值为:
## [1] "17" "28" "18" "32" "34" " 7" " 5" "25" " 8" "31" "44" "19" "47" "39" "27"
## [16] "11" "22" " 6" " 3" "33" "24" "48" "15" "43" "30" "29" "21" " 9" "57" "23"
## [31] "45" "12" "26" "46" "40" "42" "14" "10" " 4" "56" "13" "20" "16" "55" "35"
## [46] "51" "38" "41" "52" "58" "53" "37" " 2" "49" "36" "62"
##
## property_height 的去重值为:
## [1] "中" "低" "高" NA
##
## property_style 的去重值为:
## [1] "塔楼" "板楼" "板塔结合" "暂无数据" "平房"
##
## followers 的去重值为:
## [1] " 3" " 1" " 2" " 0" " 10" " 4" " 8" " 11" " 6" " 5" "102" " 50"
## [13] " 9" " 7" " 12" " 17" " 18" " 25" " 24" " 48" " 16" "104" " 13" " 23"
## [25] " 15" " 30" "204" " 14" " 26" " 31" " 22" " 41" " 19" " 33" " 78" " 34"
## [37] " 29" " 84" " 38" "112" " 73" " 36" " 45" " 44" " 28" " 55" "121" " 58"
## [49] " 32" " 20" " 35" " 46" " 21" " 51" "103" " 62" " 77" " 79" "106" " 86"
## [61] " 27" " 43" "114" "143" " 61" " 39" " 42" "262" " 97" " 56" " 40" "133"
## [73] " 57" "209" "139" " 65" " 87" " 68" " 49" " 52" "125"
##
## near_subway 的去重值为:
## [1] "近地铁" NA "太子湖1号" "VR看装修" "珞狮南" "近地看"
##
## if_2y 的去重值为:
## [1] NA "房本满两年"
##
## has_key 的去重值为:
## [1] "随时看房" "王家湾" "近地铁" NA "关山大道" "仁厚社区"
## [7] "世纪花园" "藏龙岛" "VR看装修" "随李铁"
##
## vr 的去重值为:
## [1] NA "VR看装修" "江景湾"
## [4] "VR\xe7甸城\xe5\xe8c" "珞狮南路" "保利拉\xe8\xe8f"
```

##	[7]	"随时看\xe6\x88"	"泓悦府"	"随时看房"
##	[10]	"近地铁"	"塔子湖"	"育才花\xe6\xa1"

property_name
<chr>
常青花园五小区
城市之光
红光小区(江汉)
金色雅园金沙苑
华清园
中北春天
6 rows

property_name	property_region
<chr>	<chr>
常青花园五小区	常青花园
红光小区(江汉)	长港路
城市之光	光谷东
华清园	三阳路
4 rows	

可以看到：

- 直观结论1 武汉市二手房每平米均价1.5万元，过半房屋的单价在1.1-1.8万之间
- 直观结论2 武汉市二手房2房和3房的户型居多，大部分房屋面积在85-120平之间
- 直观结论3 武汉市二手房中高楼层居多，过半房屋楼层在10-33层之间

Data summary








Name	lj
Number of rows	2514
Number of columns	19
Column type frequency:	
character	12
numeric	7
Group variables	
None	

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
property_name	0	1.00	2	17	0	1345	0
directions1	0	1.00	1	2	0	8	0
directions2	1396	0.44	1	2	0	8	0
decoration	0	1.00	2	2	0	4	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
property_height	53	0.98	1	1	0	3	0
property_style	0	1.00	2	4	0	5	0
near_subway	1210	0.52	3	3	0	1	0
if_2y	1464	0.42	5	5	0	1	0
has_key	440	0.82	4	4	0	1	0
vr	770	0.69	5	5	0	1	0
property_region.y	2502	0.00	3	4	0	4	0
region	0	1.00	2	6	0	84	0

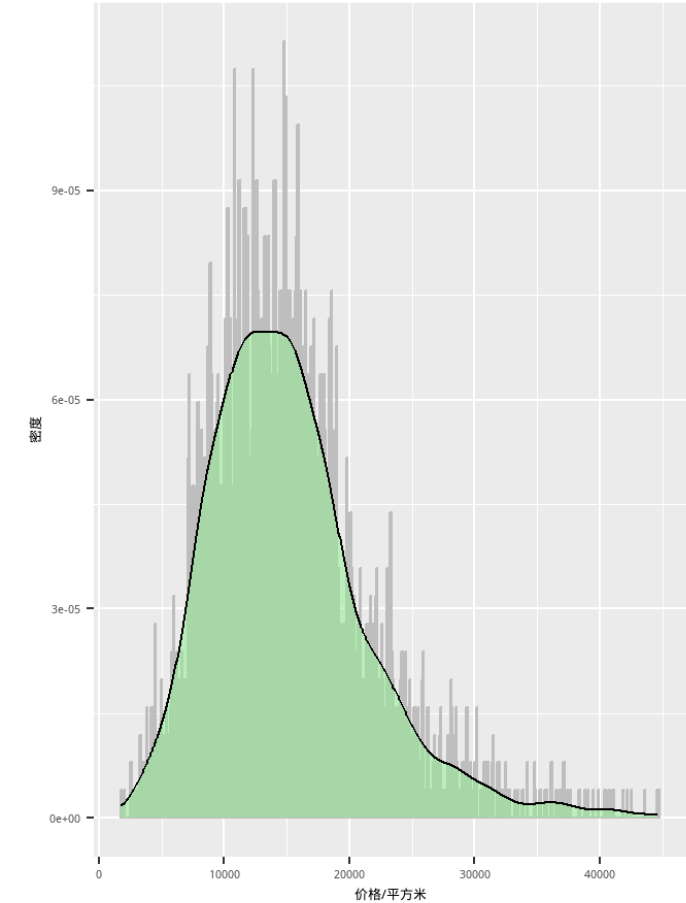
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
price_ttl	0	1	154.79	94.43	10.60	95.00	136.00	188.00	1380.00	
price_sqm	0	1	15108.73	6347.73	1771.00	10765.00	14307.00	18209.25	44656.00	
bedrooms	0	1	2.69	0.73	1.00	2.00	3.00	3.00	7.00	
livingrooms	0	1	1.71	0.47	0.00	1.00	2.00	2.00	4.00	
building_area	0	1	100.66	30.71	22.77	84.44	95.42	117.98	588.66	
property_t_height	0	1	24.05	12.46	2.00	11.00	27.00	33.00	62.00	
followers	0	1	6.33	15.01	0.00	1.00	2.00	6.00	262.00	

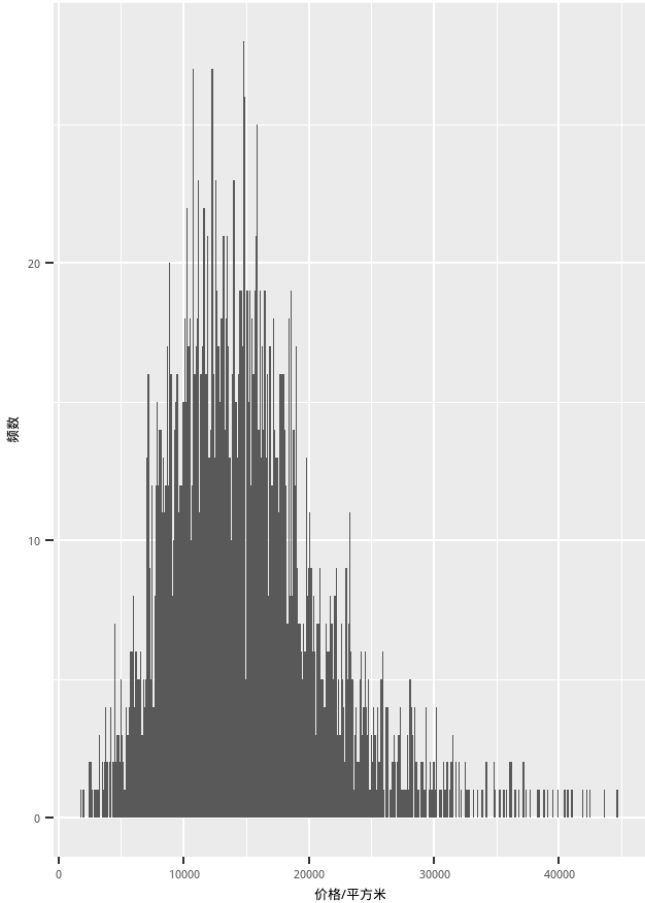
探索性分析

房屋单价的数值描述与图形

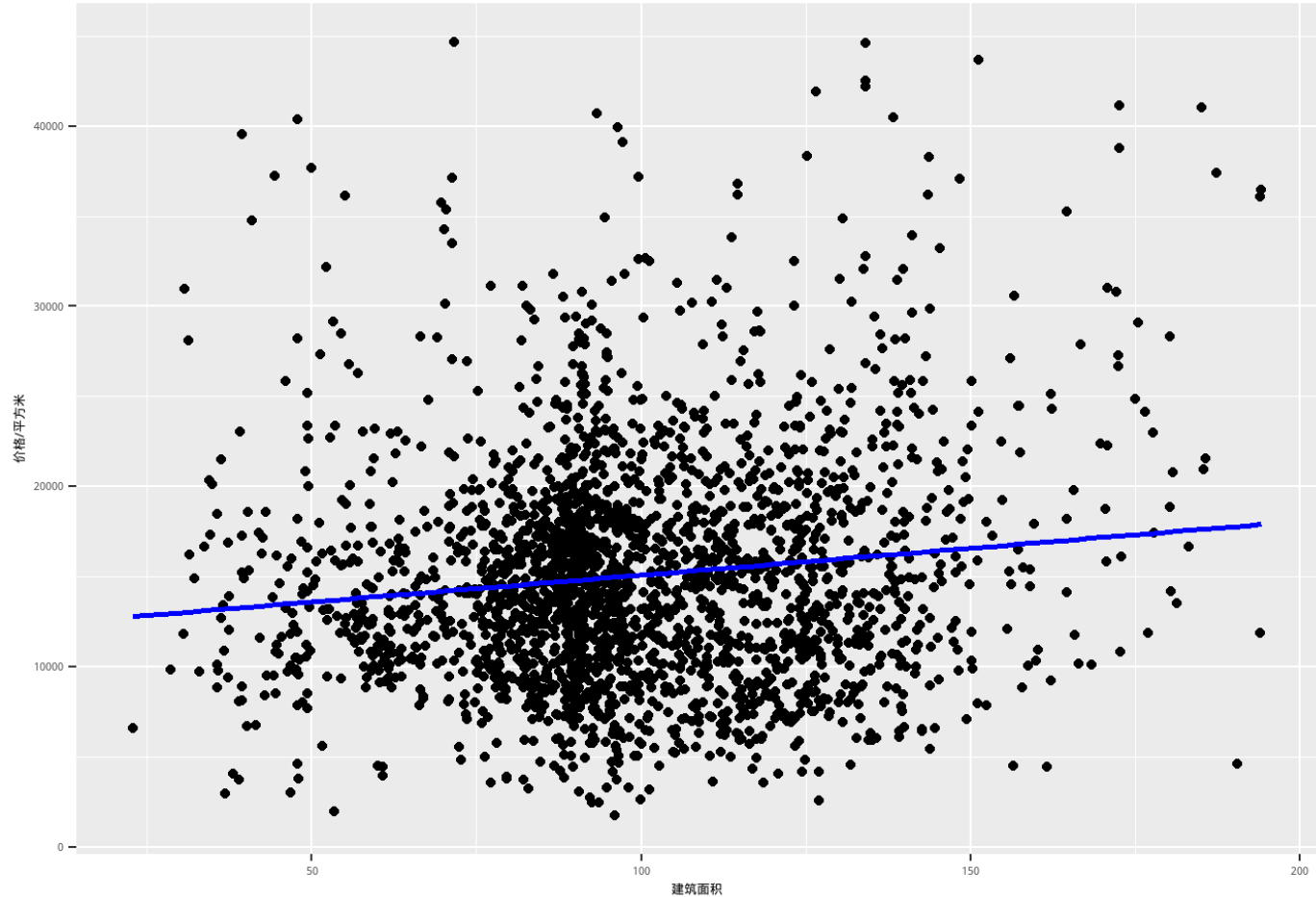
price_sqm直方图与概率密度曲线

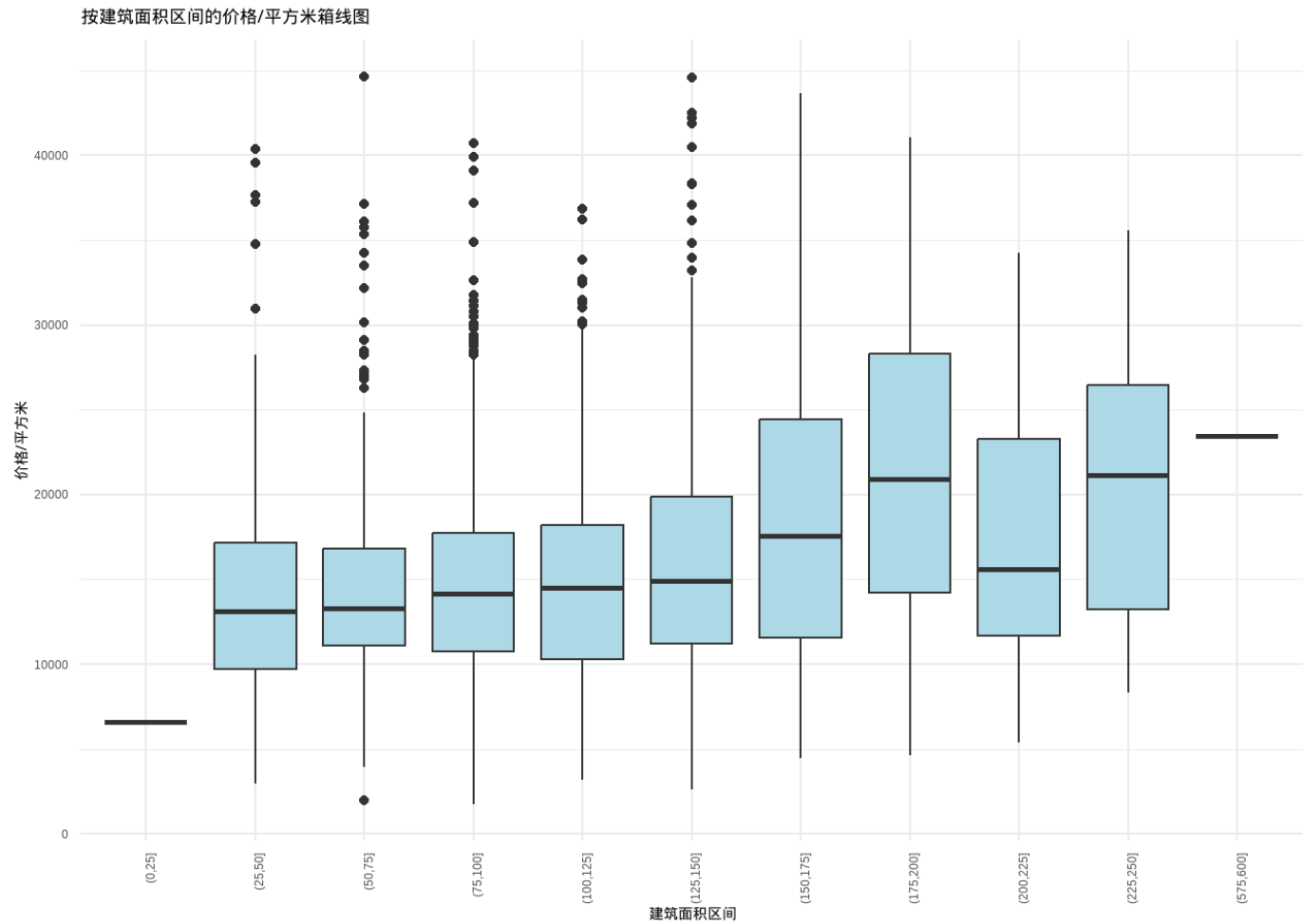


price_sqm直方图



price_sqm 和 building_area 的关系 (去除极值)

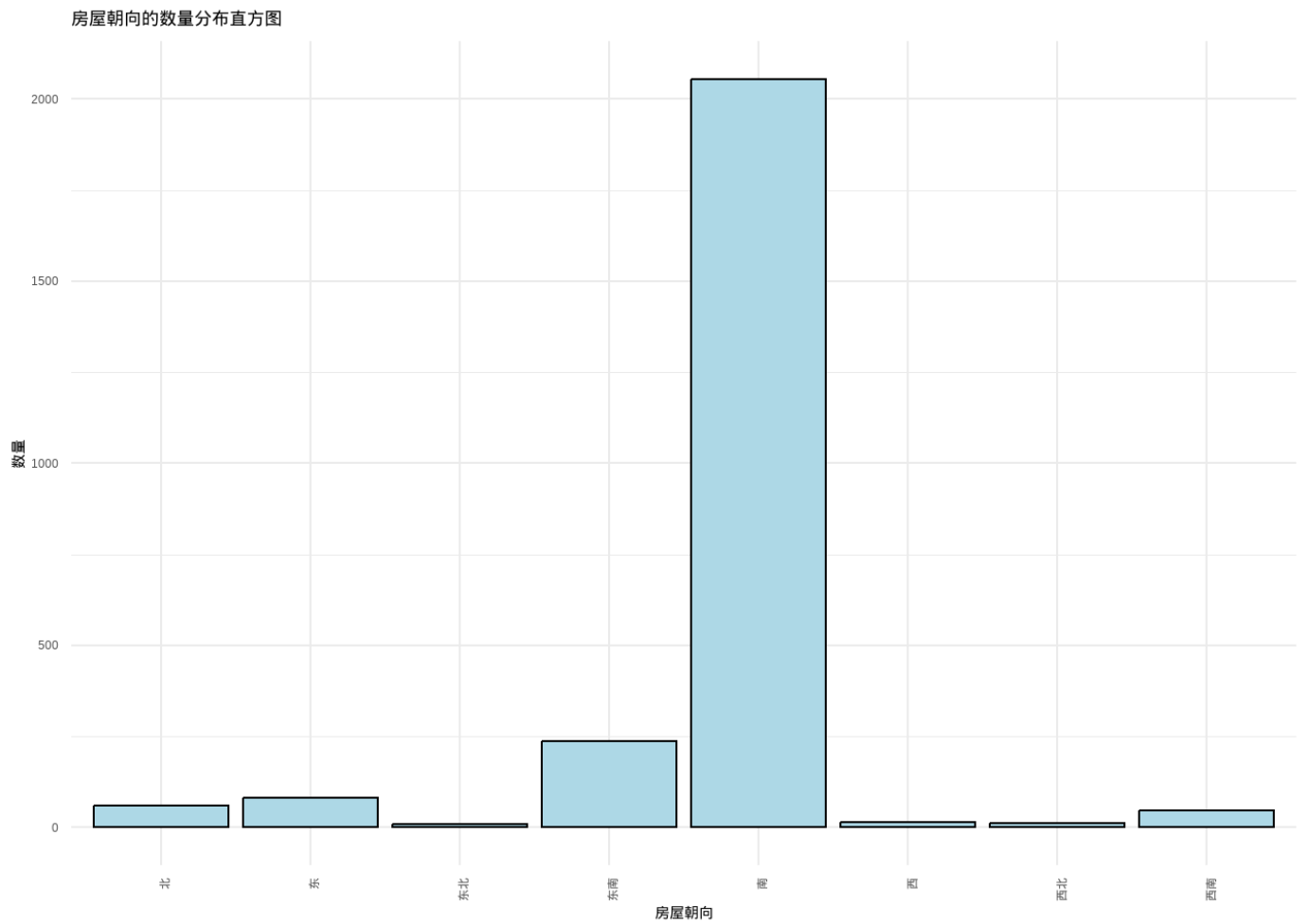




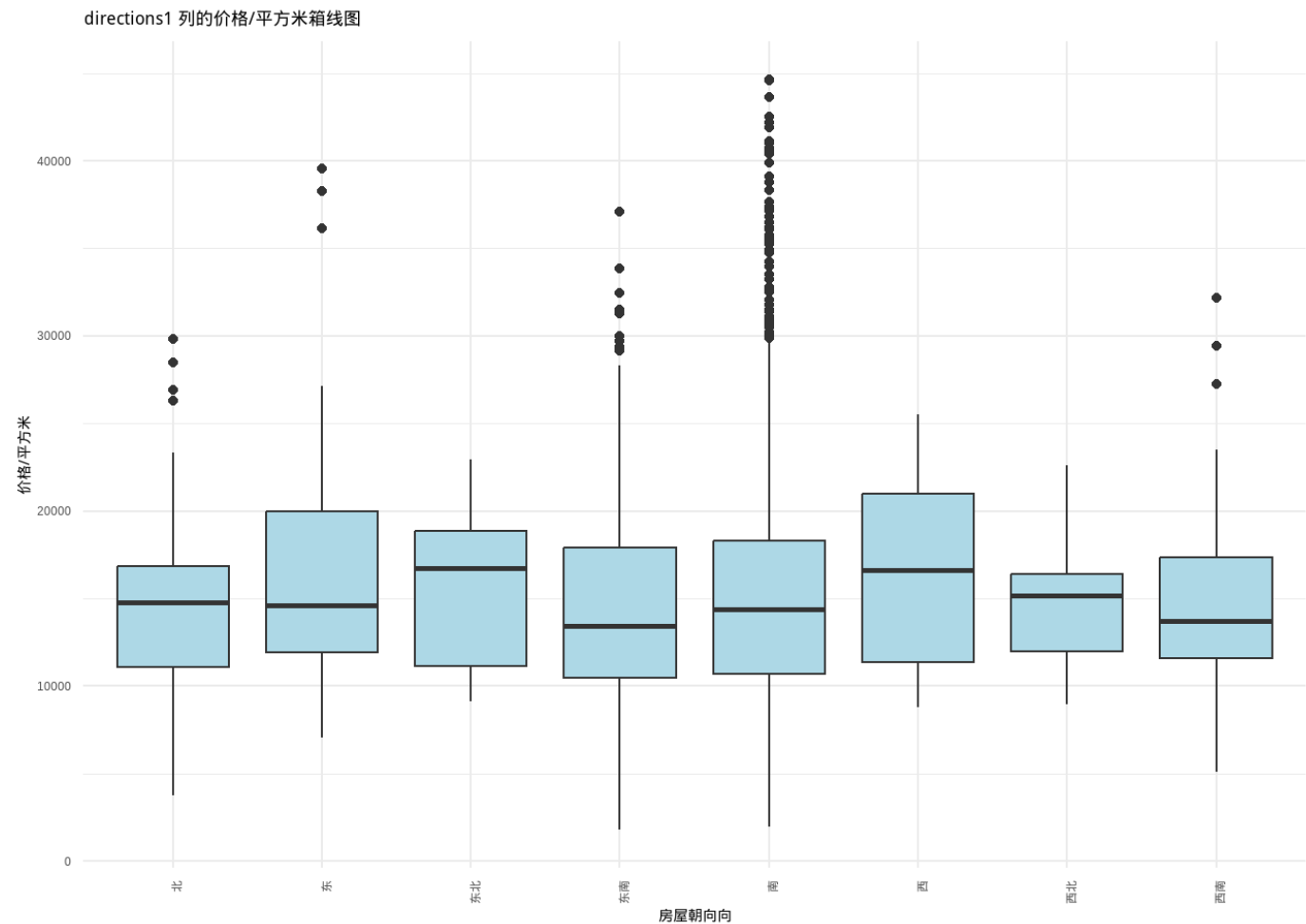
发现:

- 发现1 房屋单价分布呈右偏，大部分房屋单价在3万以内
- 发现2 房屋单价与面积成正比，面积越大的房子，单价可能更高
- 发现3 房屋单价与面积成正比，面积越大的房子，单价可能更高，175-200平以及225-250平的房子单价比其他面积更贵

房屋朝向的数值描述与图形



```
## # A tibble: 8 × 3
##   directionsl count proportion
##   <chr>      <int>      <dbl>
## 1 东           81    0.0322
## 2 东北          10    0.00398
## 3 东南         238    0.0947
## 4 北           59    0.0235
## 5 南        2055    0.817
## 6 西           14    0.00557
## 7 西北          11    0.00438
## 8 西南          46    0.0183
```

发现:

- 发现1 大约有82%的房子房屋朝向为南
- 发现2 单价在3万以上的房子，房屋朝向为南和东南可能是个很好的卖点

发现总结

武汉市二手房均价1.5万，过半房屋的单价在1.1-1.8万之间，房屋单价与面积成正比，面积越大的房子，单价可能更高。2房和3房的户型居多，大部分房屋面积在85-120平之间，房中高楼层居多，过半房屋楼层在10-33层之间，大约有82%的房子房屋朝向为南。