

第一次作业 链家二手房数据分析

吴晗

2023-10-19

主要发现

1. 从小区上看，东立国际小区在售二手房最多。从地区上看，白沙洲地区 在售二手房最多。地区二手房数量高，其各个小区二手房数量不一定高。
2. 从房屋总价和房屋面积上看，房屋总价和面积总体是正比例增长，但也有短暂下降，例如白沙洲。
3. 从地区与房屋单价上看，光谷东的房屋单价跨度最广，普遍比其他要高，青山有高异常房屋单价。

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。

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    <fct> 南湖名都A区, 万科紫悦湾, 东立国际, 新都汇, 保利城一~
## $ property_region  <fct> 南湖沃尔玛, 光谷东, 二七, 光谷广场, 团结大道, 庙山, ~
## $ 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        <fct> 3, 3, 1, 3, 3, 3, 2, 3, 4, 3, 5, 3, 4, 3, 3, 2, 3, 4~
## $ livingrooms      <fct> 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2~
## $ building_area    <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19, ~
```

Table 1: 武汉链家二手房

property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms	building_area	directions1	directions2
南湖名都 A 区	南湖沃尔玛	237.0	18709	3	1	126.68	南	南
万科紫悦湾	光谷东	127.0	14613	3	2	86.91	南	南
东立国际	二七	75.0	15968	1	1	46.97	南	南
新都汇	光谷广场	188.0	15702	3	2	119.73	北	北
保利城一期	团结大道	182.0	17509	3	2	103.95	东南	东南
加州橘郡	庙山	122.0	10376	3	2	117.59	南	南
省建筑五公司西区	光谷广场	99.0	12346	2	1	80.19	南	南
保利上城东区	白沙洲	193.8	16336	3	2	118.64	南	南
石化大院	中南丁字桥	325.0	32631	4	1	99.60	南	南
阳光花园	杨汊湖	192.0	17403	3	2	110.33	南	南

```

## $ directions1      <fct> 南, 南, 南, 北, 东南, 南, 南, 南, 南, 南, 南, 南, 东~
## $ directions2      <fct> 北, NA, NA, 东, NA, 北, NA, 北, 北, 北, 北, NA, 西南~
## $ decoration        <fct> 精装, 精装, 简装, 精装, 简装, 精装, 简装, 其他, 简装~
## $ property_t_height <dbl> 17, 28, 18, 32, 34, 34, 7, 34, 5, 7, 25, 32, 8, 31, ~
## $ property_height   <fct> 中, 中, 低, 高, 中, 低, 低, 中, 低, 低, 高, 高, 中, ~
## $ property_style     <fct> 塔楼, 板楼, 塔楼, 塔楼, 板塔结合, 板楼, 板楼, 板塔结~
## $ followers         <dbl> 3, 1, 3, 2, 3, 1, 0, 0, 2, 0, 0, 0, 10, 0, 0, 1, 0, ~
## $ near_subway        <fct> 近地铁, NA, 近地铁, 近地铁, NA, NA, 近地铁, 近地铁, ~
## $ if_2y              <fct> NA, 房本满两年, NA, 房本满两年, 房本满两年, 房本满两~
## $ has_key            <fct> 随时看房, 随时看房, 随时看房, 随时看房, 随时看房, 随~
## $ vr                <fct> NA, VR看装修, NA, NA, VR看装修, NA, VR看装修, NA, NA~

```

各变量的简短统计:

```

##           property_name  property_region  price_ttl      price_sqm      bedrooms
## 东立国际      : 22  白沙洲 : 167    Min.      : 10.6    Min.      : 1771    1: 142
## 保利中央公馆 : 16  盘龙城 : 126    1st Qu.: 95.0    1st Qu.:10799    2: 919
## 朗诗里程      : 16  四新   : 116    Median : 137.0    Median :14404    3:1692
## 恒大名都      : 15  光谷东 : 112    Mean    : 155.9    Mean    :15148    4: 214
## 阳光100大湖第: 15  金银湖 : 97    3rd Qu.: 188.0    3rd Qu.:18211    5: 27
## 保利城一期    : 13  后湖   : 86    Max.    :1380.0    Max.    :44656    6: 4
## (Other)       :2903 (Other):2296                                7: 2
## livingrooms building_area  directions1  directions2  decoration
## 0: 18      Min.      : 22.77  南       :2454  北       :1189  简装: 634
## 1: 845     1st Qu.: 84.92  东南    : 281  南       : 66  精装:1757
## 2:2131     Median : 95.55  东       : 98  西       : 25  毛坯: 436
## 3: 5       Mean    :100.87  北       : 68  东南    : 15  其他: 173
## 4: 1       3rd Qu.:117.68  西南    : 57  西南    : 12
##          Max.    :588.66  西       : 19  (Other): 21
##          (Other): 23  NA's    :1672
## property_t_height property_height property_style  followers
## Min.      : 2.00  低      : 816  板楼      :1781  Min.      : 0.000
## 1st Qu.:11.00  高      : 906  板塔结合: 615  1st Qu.: 1.000
## Median :27.00  中     :1218  平房      : 5  Median : 3.000
## Mean    :24.22  NA's: 60  塔楼      : 527  Mean    : 6.614
## 3rd Qu.:33.00  暂无数据: 72  3rd Qu.: 6.000
## Max.    :62.00  Max.    :262.000
##

```

```
##      near_subway      if_2y      has_key      vr
## VR看装修 : 2 房本满两年:1264 随时看房:2525 VR看装修 :2084
## 近地看 : 1 NA's :1736 近地铁 : 7 VR\7甸城\5\8c: 1
## 近地铁 :1554 VR看装修: 4 随时看\6\88 : 1
## 珞狮南 : 1 藏龙岛 : 1 保利拉\8\8f : 1
## 太子湖1号: 1 关山大道: 1 育才花\6\1a : 1
## NA's :1441 (Other) : 4 (Other) : 6
## NA's : 458 NA's : 906
```

可以看到:

- 房屋总价最小 10.6 万元, 最大 1380 万元, 中间值 137 万元, 平方米价格最小 1771 元/平方米, 最大 44656 元/平方米。
- 房间卧室个数分布, 3 室最多, 7 室最少。客厅个数, 2 客厅最多, 4 客厅最小, 总体面积: 最小 22.77 最大 588.66。
- 房间装修精准最多, 板楼最多, 主要朝向南最多, 次要朝向北最多。

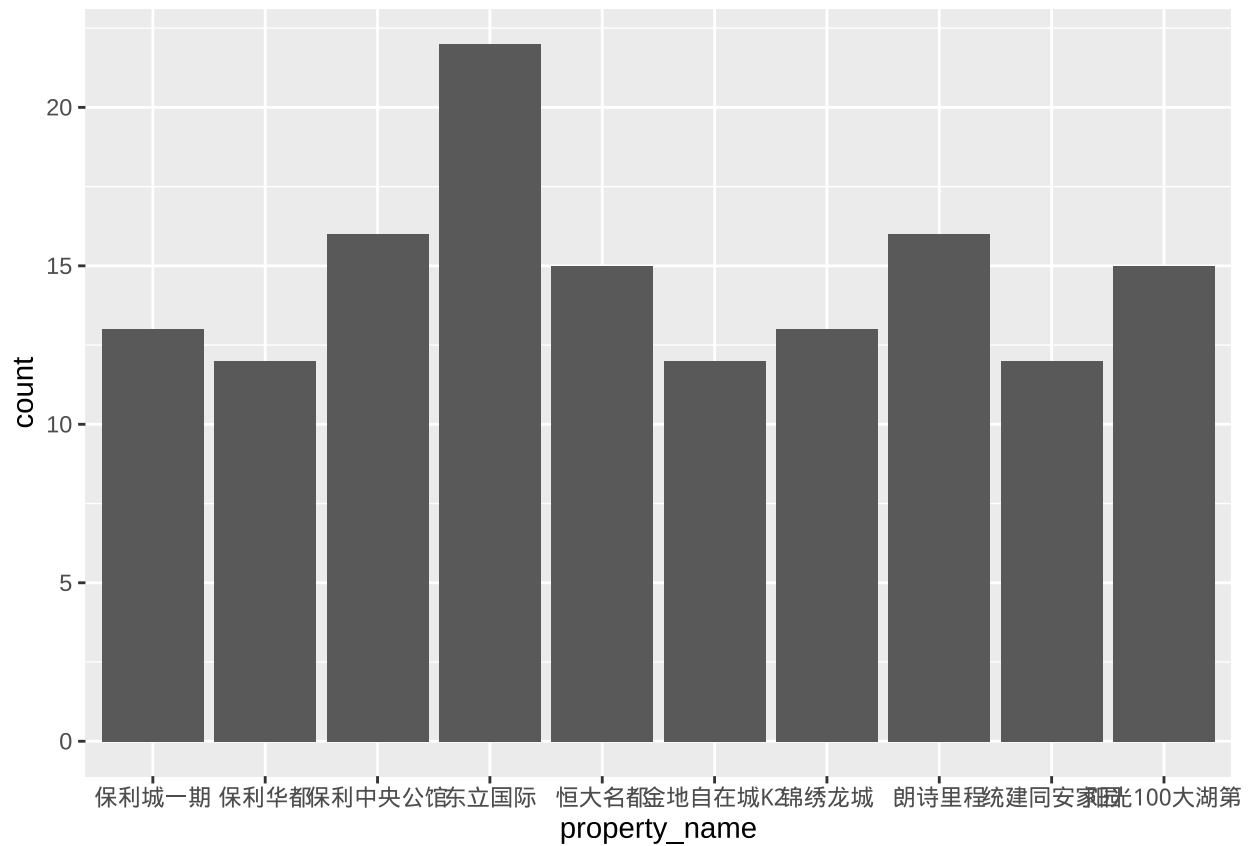
探索性分析

变量 `property_name` 的数值描述与图形

```
## # A tibble: 1,345 x 2
##   property_name count
##   <fct>          <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 金地自在城K2    12
## 10 统建同安家园   12
## # i 1,335 more rows
```

```
## # A tibble: 10 x 2
##   property_name count
##   <fct>          <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 金地自在城K2    12
## 10 统建同安家园   12
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00    1.00    2.00    2.23    3.00   22.00
```



发现:

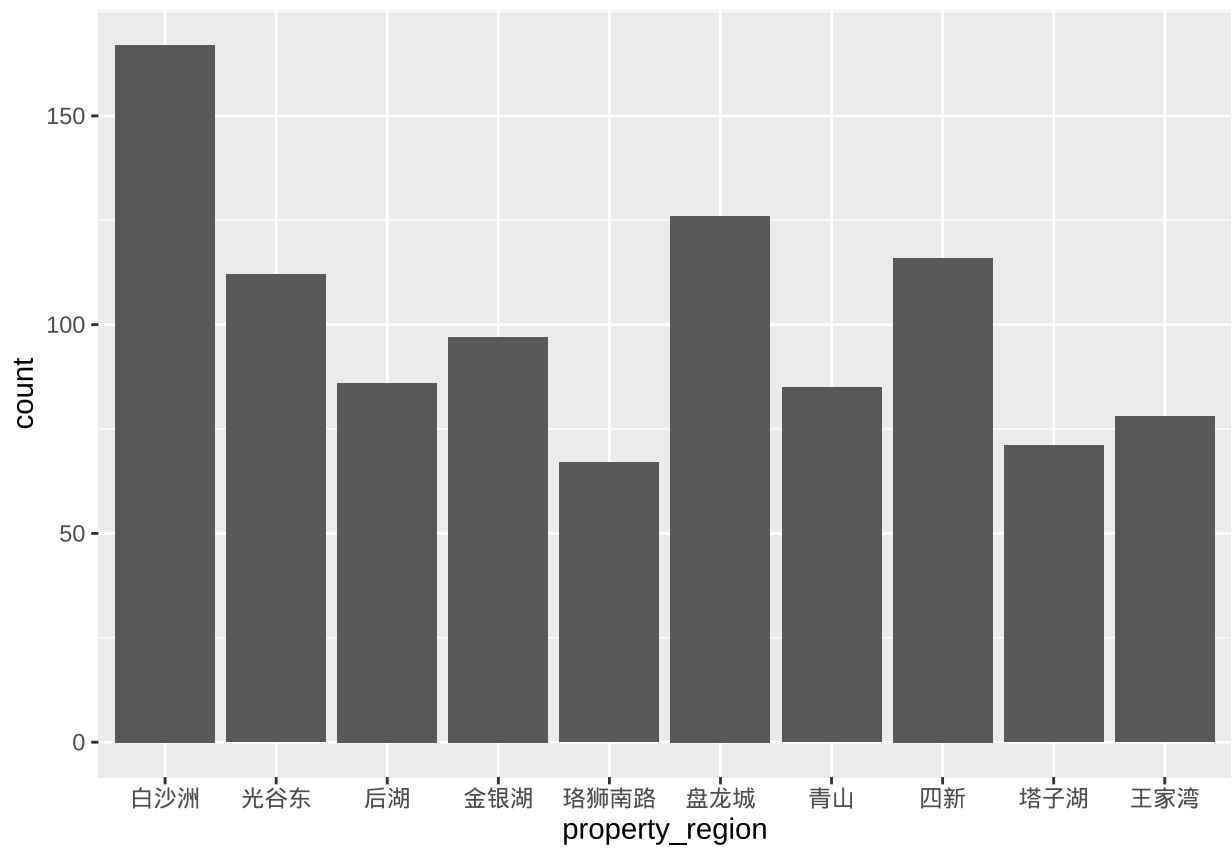
- 共有 1345 个小区在售二手房，其中东立国际小区在售二手房最多，有 22 间，最少小区仅一间二手房 大部分小区有 1-3 间二手房出售 平均每个小区有 2.23 间。
- 前 10 排名的小区至少有 12 间二手房出售。

变量 `property_region` 的数值描述与图形

```
## # A tibble: 87 x 2
##   property_region count
##   <fct>          <int>
## 1 白沙洲          167
## 2 盘龙城          126
## 3 四新            116
## 4 光谷东          112
## 5 金银湖           97
## 6 后湖             86
## 7 青山             85
## 8 王家湾           78
## 9 塔子湖           71
## 10 珞狮南路        67
## # i 77 more rows
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   14.00   25.00   34.48   45.00   167.00
```

```
## # A tibble: 10 x 2
##   property_region count
##   <fct>          <int>
## 1 白沙洲          167
## 2 盘龙城          126
## 3 四新            116
## 4 光谷东          112
## 5 金银湖           97
## 6 后湖            86
## 7 青山            85
## 8 王家湾           78
## 9 塔子湖           71
## 10 珞狮南路        67
```

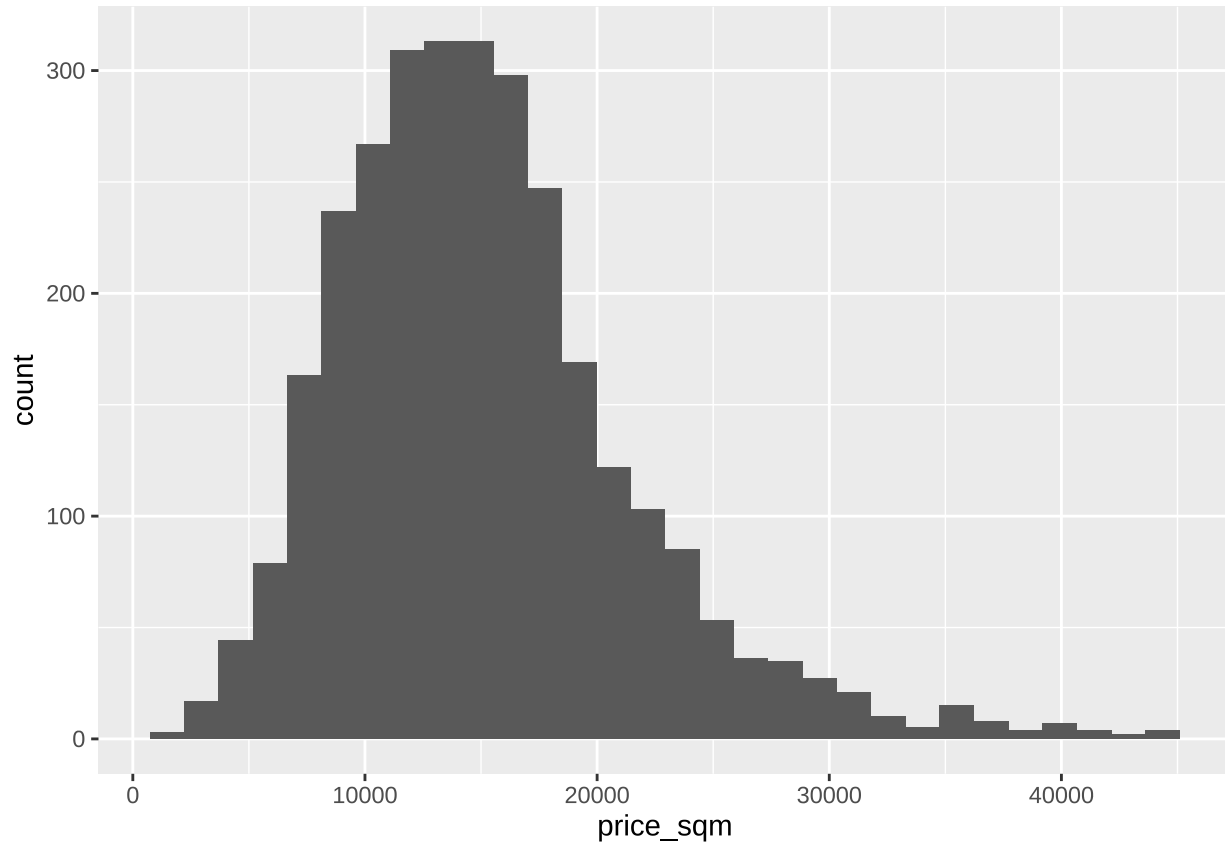


发现:

- 共有 87 个地区在售二手房，白沙洲地区在售二手房最多，有 167 间。最少地区仅一间二手房，平均每个地区有 34 间。
- 排名前 10 的地区最少也有 67 间在售二手房。

变量 `price_sqm` 的数值描述与图形

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1771  10799   14404   15148   18211   44656
```



发现:

- 房屋单价最小 1771, 最大 44656, 平均值 15148。
- 房屋单价集中在 10799 到 18211 之间。

探索问题 地区, 小区在售二手分布趋势

```
## # A tibble: 1,345 x 2
##   property_name count
##   <fct>         <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 金地自在城K2     12
```

```
## 10 统建同安家园      12
## # i 1,335 more rows

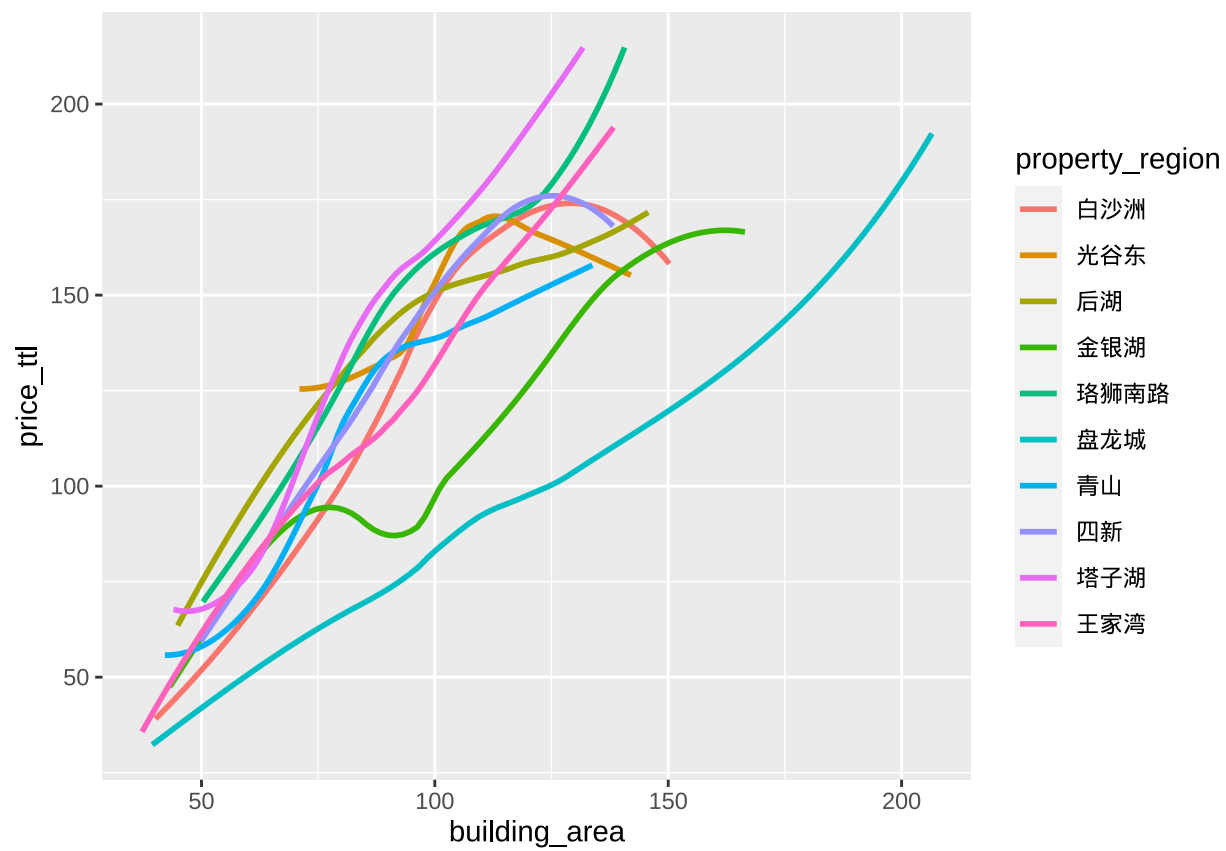
## # A tibble: 1,357 x 3
## # Groups:   property_region [87]
##   property_region property_name count
##   <fct>          <fct>      <int>
## 1 二七            东立国际        22
## 2 光谷东          朗诗里程        16
## 3 新南湖          保利中央公馆      16
## 4 藏龙岛          阳光100大湖第      15
## 5 盘龙城          恒大名都          15
## 6 民族大道        锦绣龙城          13
## 7 团结大道        保利城一期        13
## 8 后湖            统建同安家园      12
## 9 虎泉杨家湾      保利华都          12
## 10 团结大道       金地自在城K2      12
## # i 1,347 more rows
```

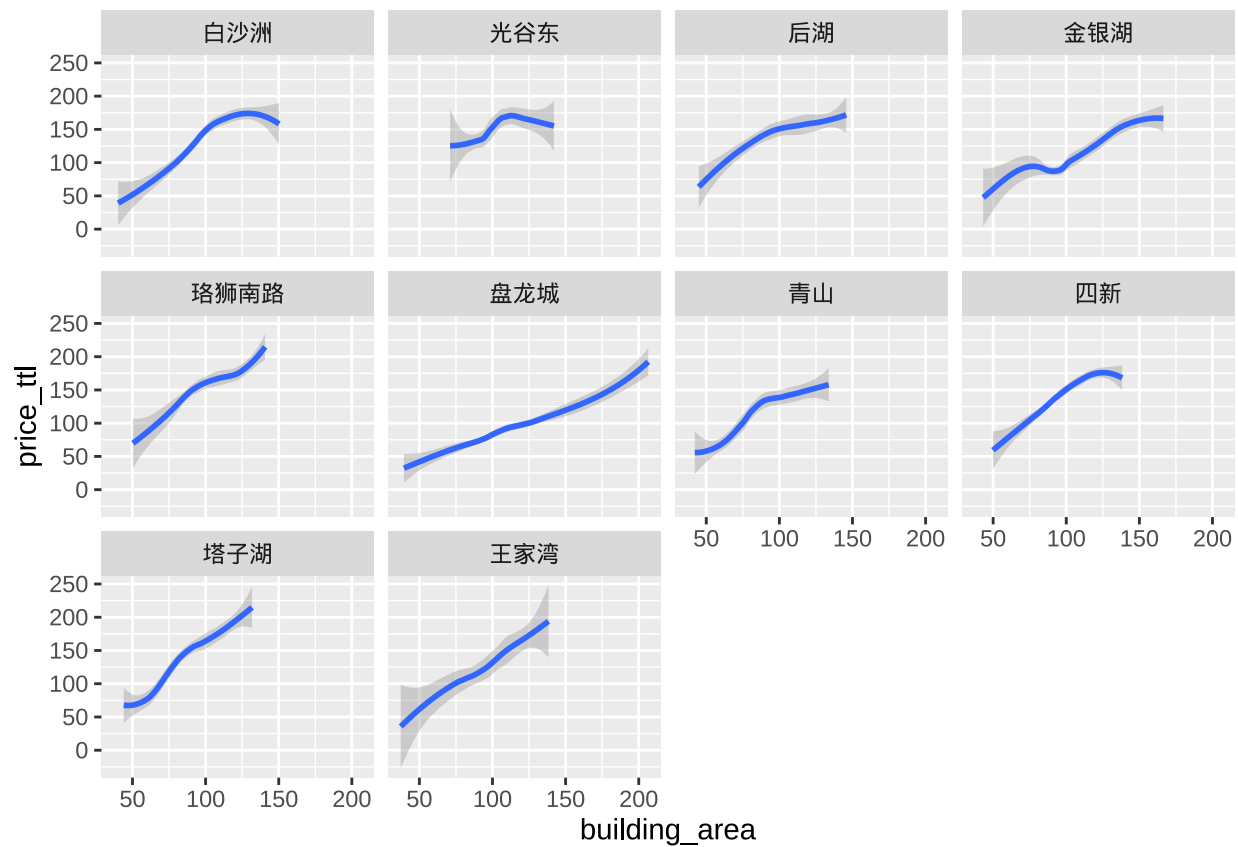
发现:

- 二七地区的东立国际小区在售二手房最高达 22，二七地区不在地区二手房前 10。
- 地区二手房数量高，其各个小区不一定高例如白沙洲地区数量最高，其小区数量甚至没有前 10。

探索问题 在售数量前 10 不同地区 相同面积下 地区与房屋总价之间的关系

```
## # A tibble: 1,005 x 18
##   property_name    property_region price_ttl price_sqm bedrooms livingrooms
##   <fct>          <fct>      <dbl>   <dbl> <fct>    <fct>
## 1 万科紫悦湾      光谷东        127     14613 3         2
## 2 保利上城东区    白沙洲        194.     16336 3         2
## 3 光谷创新天地1期 光谷东        470     19003 4         2
## 4 保利新武昌K4地块 白沙洲        215     20128 3         2
## 5 融科花满庭      白沙洲        259     16463 4         2
## 6 朗诗里程        光谷东        203     18816 3         2
## 7 汉飞洋房印象    盘龙城         60      6266 3         1
## 8 爱家名校华城    白沙洲        83.8     9279 3         2
## 9 东方玉龙居      青山          155     11590 3         2
## 10 广电兰亭盛荟1期 四新          135     15593 3         2
## # i 995 more rows
## # i 12 more variables: building_area <dbl>, directions1 <fct>,
## #   directions2 <fct>, decoration <fct>, property_t_height <dbl>,
## #   property_height <fct>, property_style <fct>, followers <dbl>,
## #   near_subway <fct>, if_2y <fct>, has_key <fct>, vr <fct>
```

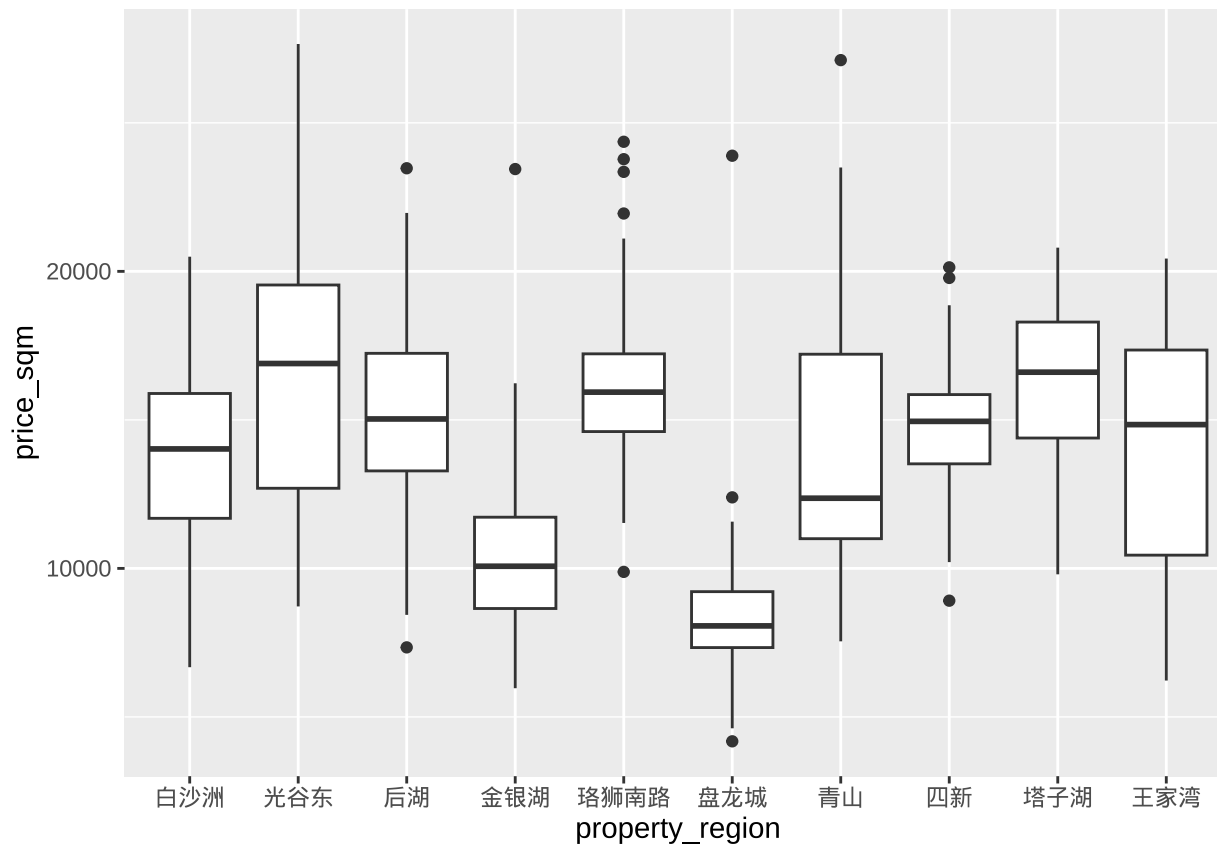




发现:

- 房屋总价和面积总体是正比例增长，但也有短暂下降，并不总是面积越大房价越高，例如白沙洲。
- 75 平方米内相同面积房屋，后湖房屋总价最高。75 往后相同面积房屋，塔子湖的房屋总价最高。

探索问题 在售数量前 10 不同地区 各自房屋单价走向



发现:

- 盘龙城房屋单价最低，中位数也是最低，金银湖第二低。光谷东的中位数最高，其次是塔子湖。
- 光谷东的房屋单价跨度最多，普遍价格比其他要高，青山有高异常房屋单价。

发现总结

二手房主要分布在白沙洲, 盘龙城, 四新, 光谷东, 金银湖, 后湖, 青山等地区, 地区二手房最多不代表各个小区的二手房也最多。房屋总价和面积总体是正比例增长, 但也有短暂下降, 说明还有其他因素影响房屋总价。不同地区, 房屋单价分布不同, 光谷东的房屋单价跨度最多, 从总体来看盘龙城房屋单价和房屋总价都比较低, 适合对地区要求不高的人购买。