

武汉二手房数据分析

郑红娟 2023281051034

2023-10-19

你的主要发现

1. 武汉二手房房屋总价平均值为 155.9 万元，总价中位数为 137.0 万元，总价偏态明显，少数总价高的房源拉高了总体平均水平。单价平均值为 1.51 万元/m²，单价中位数为 1.44 万元/m²，平均值与中位数基本相等，均价分布比较均匀。
2. 在所给的房屋相关字段中，户型中“2 室 3 厅”，朝“南”、“精装”的二手房都超过半数，是二手房中普遍属性。
3. 二手房所在行政区对房价影响比较大，其次是户型。武昌区的单价和均价均为最高且极值较多，东湖高新区作为近年新兴区域，价格紧随其后。汉口老城区 CBD 区域，价格稳居前三，符合普遍认知。

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。

数据概览

各变量的简短信息：

```
## Rows: 3,000
## Columns: 20
## $ 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> 1.8709, 1.4613, 1.5968, 1.5702, 1.7509, 1.0376, 1.23~
## $ 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, 1, 2, 2, 2, 2~
## $ building_area    <dbl> 126.68, 86.91, 46.97, 119.73, 103.95, 117.59, 80.19,~
## $ 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~
## $ roomtype          <fct> 1室3厅, 2室3厅, 1室1厅, 2室3厅, 2室3厅, 2室3厅, 1室2~
## $ matched_regions   <fct> 洪山, 东湖高新, 汉口, 洪山, 洪山, 江夏, 洪山, 洪山, ~
```

各变量的简短统计:

```
##      property_name  property_region  price_ttl      price_sqm
## 东立国际      : 22 白沙洲 : 167  Min.    : 10.6  Min.    :0.1771
## 保利中央公馆 : 16 盘龙城 : 126  1st Qu.: 95.0  1st Qu.:1.0799
## 朗诗里程      : 16 四新   : 116  Median : 137.0 Median :1.4404
## 恒大名都      : 15 光谷东 : 112  Mean   : 155.9 Mean   :1.5148
## 阳光100大湖第: 15 金银湖 : 97   3rd Qu.: 188.0 3rd Qu.:1.8211
## 保利城一期    : 13 后湖   : 86   Max.    :1380.0 Max.    :4.4656
## (Other)       :2903 (Other):2296

##      bedrooms      livingrooms      building_area      directions1      directions2
## Min.    :1.000  Min.    :0.000  Min.    : 22.77  南      :2454  北      :1189
## 1st Qu.:2.000  1st Qu.:1.000  1st Qu.: 84.92  东南    : 281  南      : 66
## Median :3.000  Median :2.000  Median : 95.55  东      : 98   西      : 25
## Mean   :2.695  Mean   :1.709  Mean   :100.87  北      : 68   东南    : 15
## 3rd Qu.:3.000  3rd Qu.:2.000  3rd Qu.:117.68  西南    : 57   西南    : 12
## Max.    :7.000  Max.    :4.000  Max.    :588.66  西      : 19   (Other): 21
##                                     (Other): 23  NA's    :1672

##      decoration  property_t_height  property_height  property_style
## 简装: 634  Min.    : 2.00  低 : 816  板楼 :1781
## 精装:1757  1st Qu.:11.00  高 : 906  板塔结合: 615
## 毛坯: 436  Median :27.00  中 :1218  平房   : 5
## 其他: 173  Mean   :24.22  NA's: 60  塔楼   : 527
##                                     暂无数据: 72
##                                     Max.    :62.00
```

```
##
##      followers      near_subway      if_2y      has_key
##  Min.   : 0.000    近地铁:1559    房本满两年:1264    随时看房:2525
##  1st Qu.: 1.000    NA's   :1441    NA's       :1736    近地铁   :   7
##  Median : 3.000                                VR看装修:   4
##  Mean   : 6.614                                藏龙岛   :   1
##  3rd Qu.: 6.000                                关山大道:   1
##  Max.   :262.000                                (Other)  :   4
##                                              NA's     : 458
##
##      vr      roomtype      matched_regions
##  VR看装修      :2084    2室3厅 :1416    洪山      :581
##  VR\ue7甸城\ue5\x8c:   1    2室2厅 : 502    汉口      :401
##  随时看\ue6\x88   :   1    1室2厅 : 417    武昌      :340
##  保利拉\ue8\x8f   :   1    1室3厅 : 273    汉阳      :294
##  育才花\ue6\xa1   :   1    2室4厅 : 186    东湖高新:283
##  (Other)        :   6    1室1厅 : 124    江汉      :199
##  NA's          : 906    (Other):  82    (Other) :902
```

可以看到:

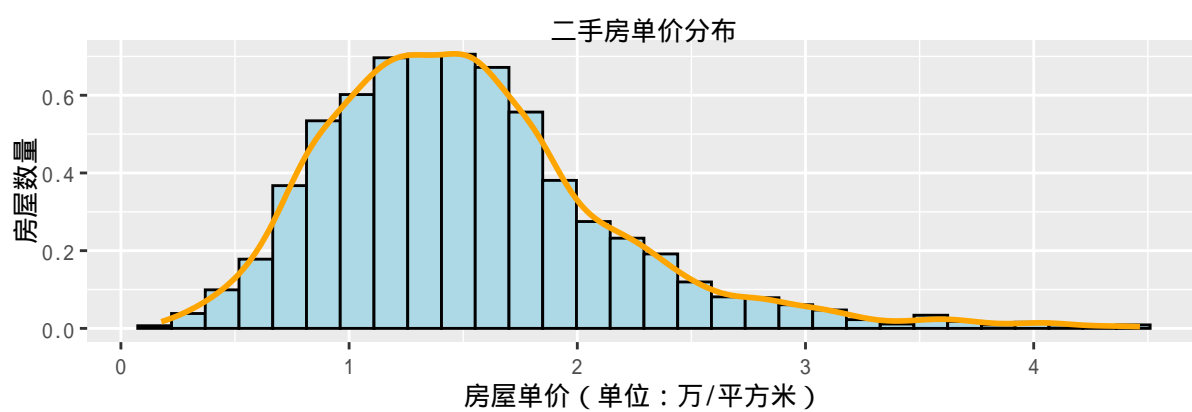
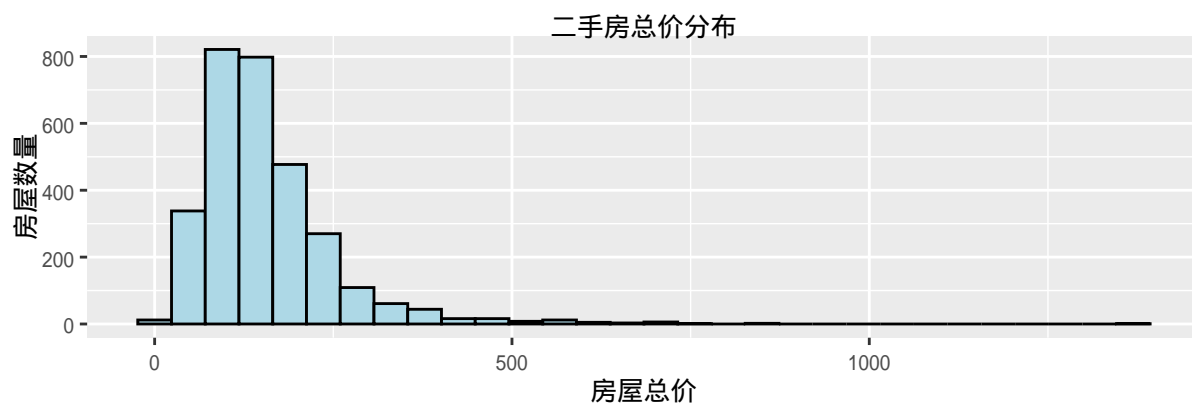
- 直观结论 1

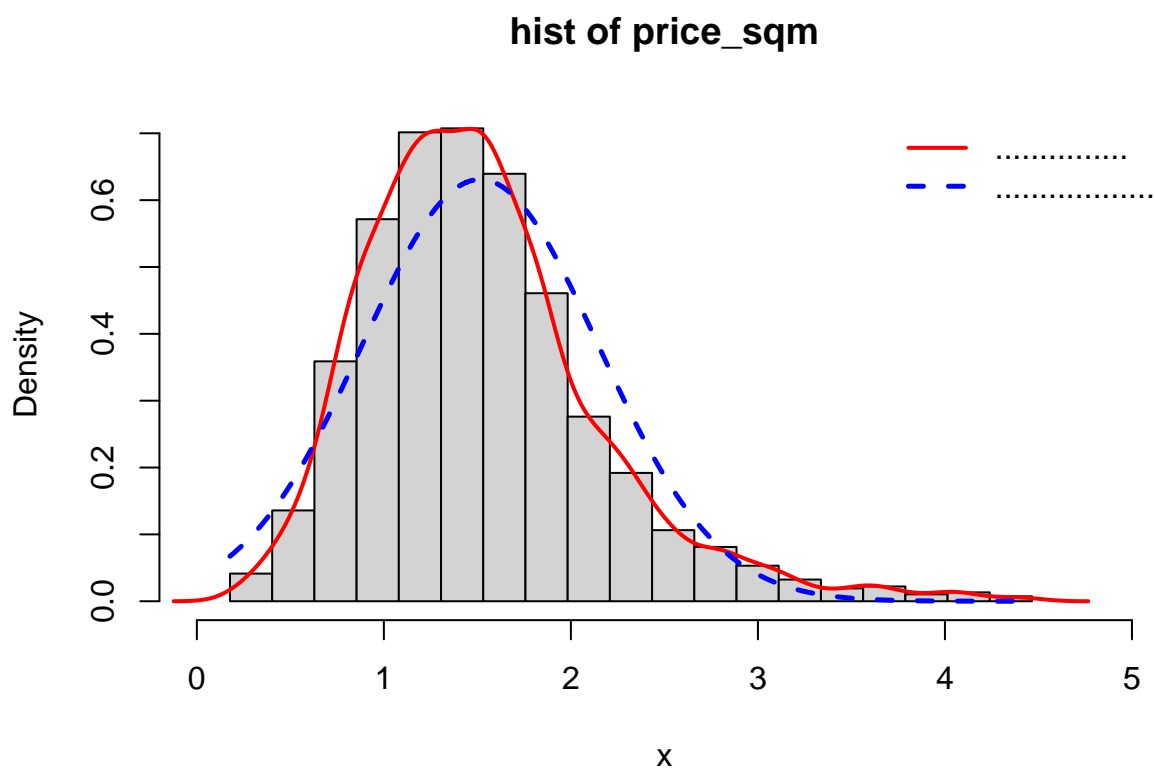
数据集中包含武汉的 3000 套二手房的信息, 包括“小区名字”、所处区域、房屋总价、房屋单价等 13 个字段信息。武汉二手房房屋总价平均值为 155.9 万元, 总价中位数为 137.0 万元, 说明少数总价高的房源拉高了总体平均水平。单价平均值为 1.51 万元/m², 单价中位数为 1.44 万元/m², 平均值与中位数基本相等, 说明房价升高趋势比较线性。

- 直观结论 2

1. 最关注的小区: 十里和府, 位于七里庙, 其次是阜华领秀中南-中南路, 光谷广场的米兰印象。
2. 小区总价最高的: 万科高尔夫城市花园, 1380 万, 是万科在武汉唯一的别墅区。
3. 小区均价最高的: 中商宿舍, 复地东湖国际小区, 因对口武昌武汉小学, 意料之中。

探索性分析

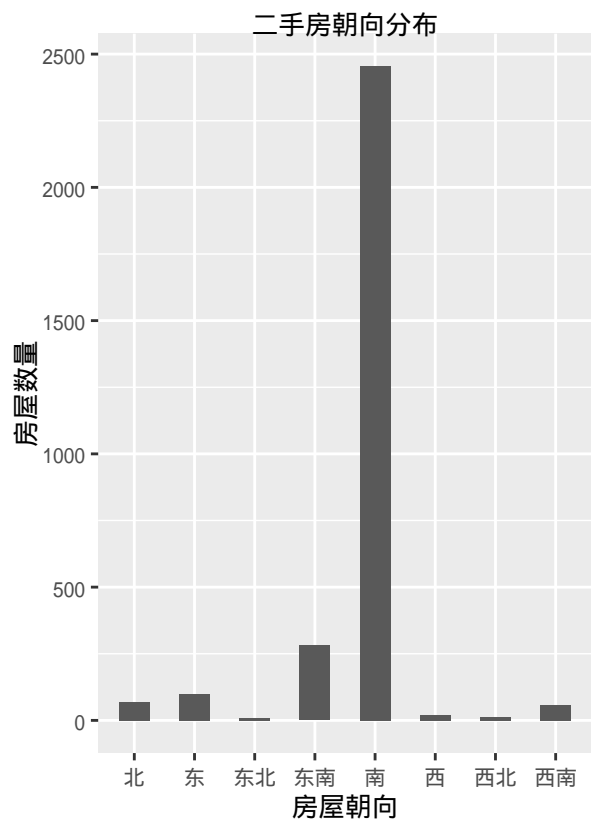
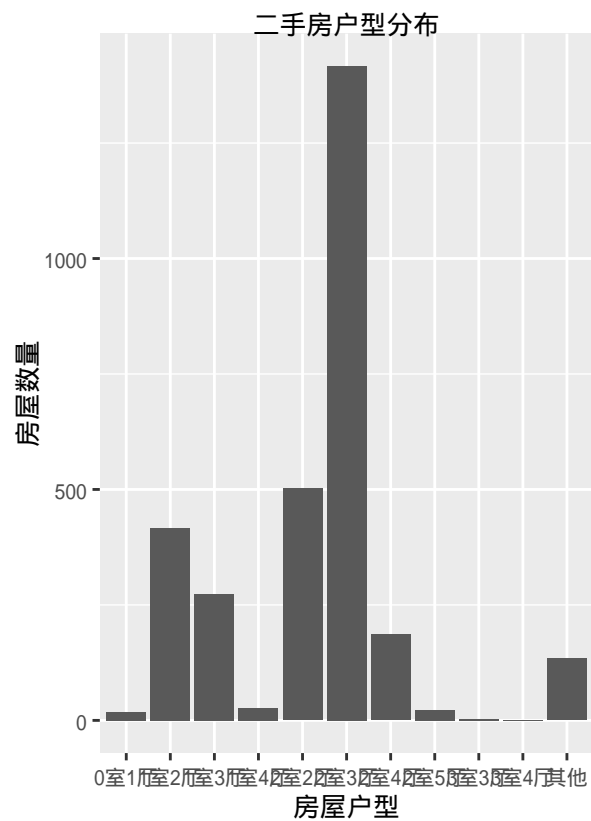


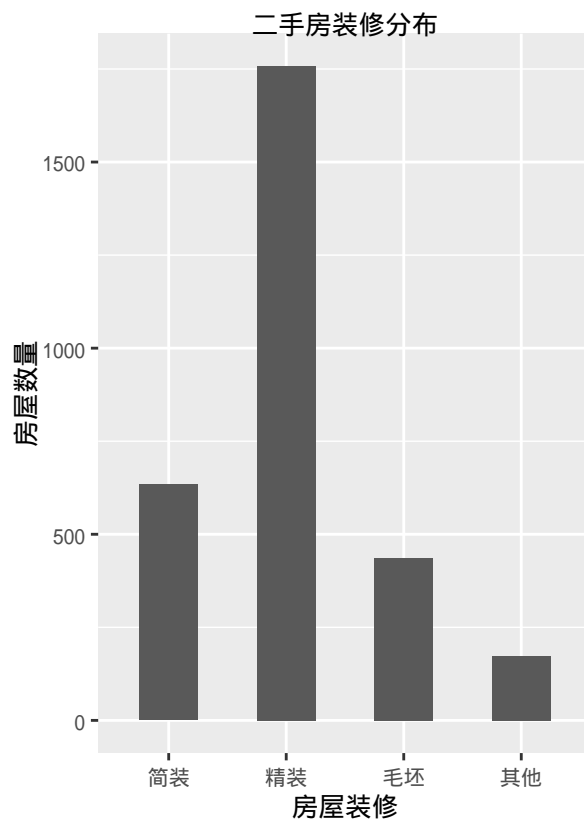
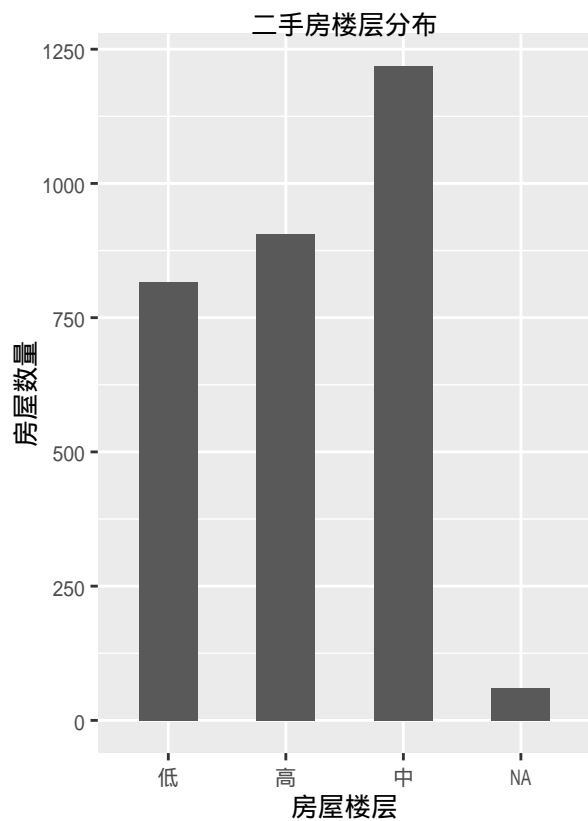


“房价”的数值描述与图形

发现：

对二手房房屋总价做直方图，整体呈拖尾型，数据分布整体右偏非常明显，集中分布在 100 万 ~200 万之间，符合大众对房价的认知。二手房房屋均价偏态比总价好，观测数据的分布与正态分布有一定差距，进一步进行正态检验分析，发现两者密度曲线比较相似。

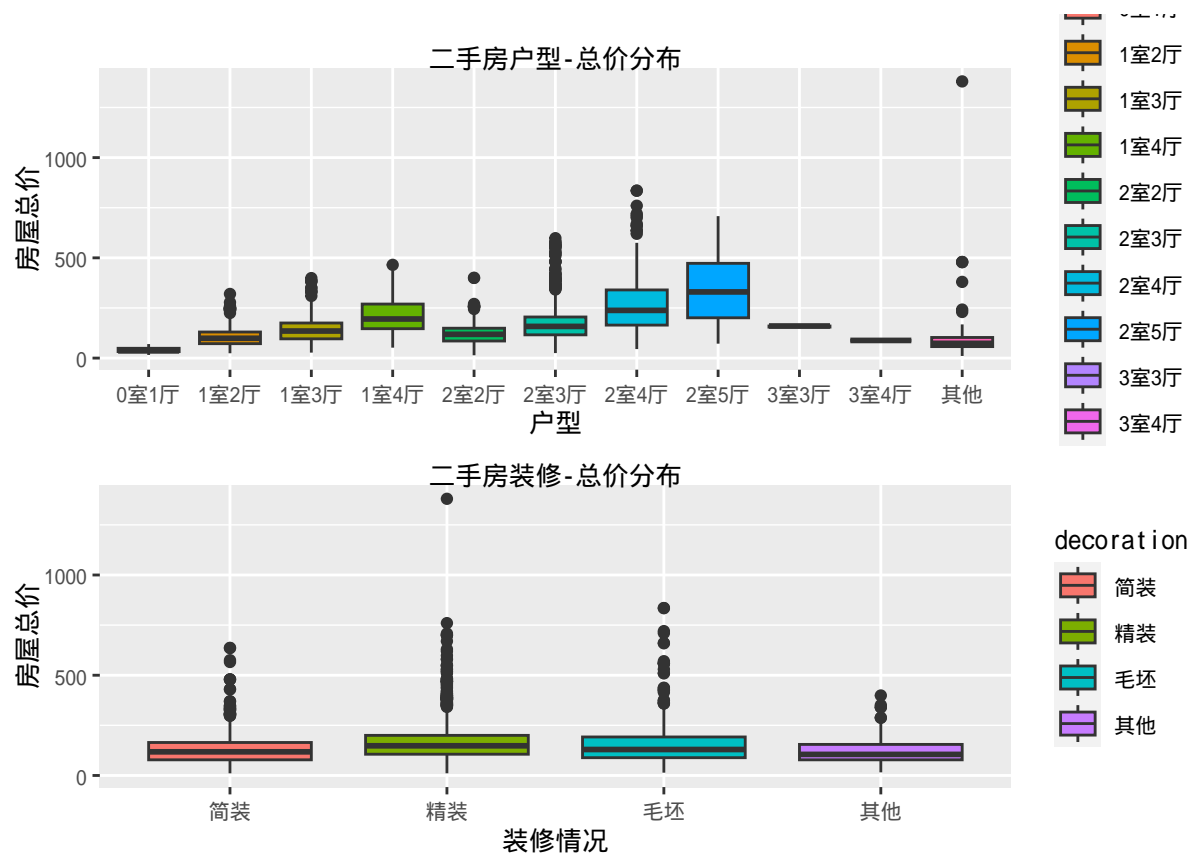


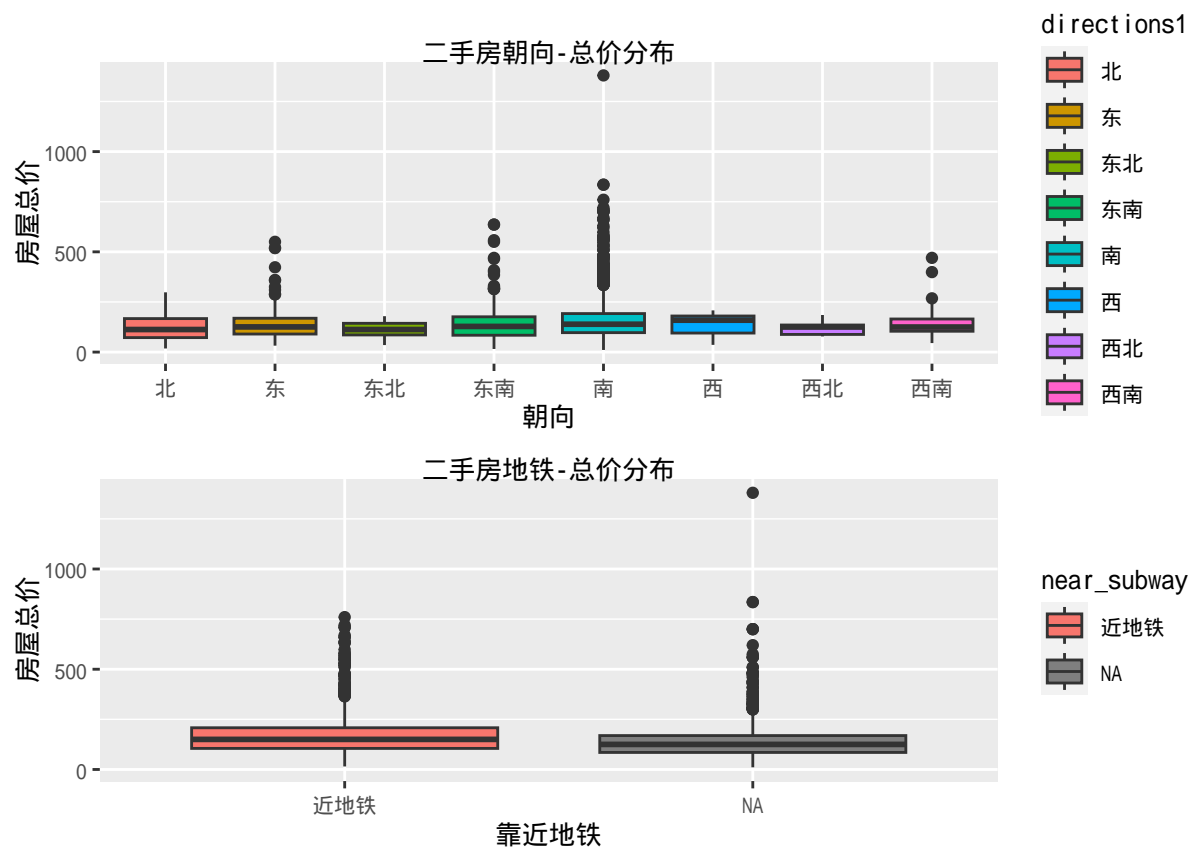


“房屋属性”的数值描述与图形

发现：

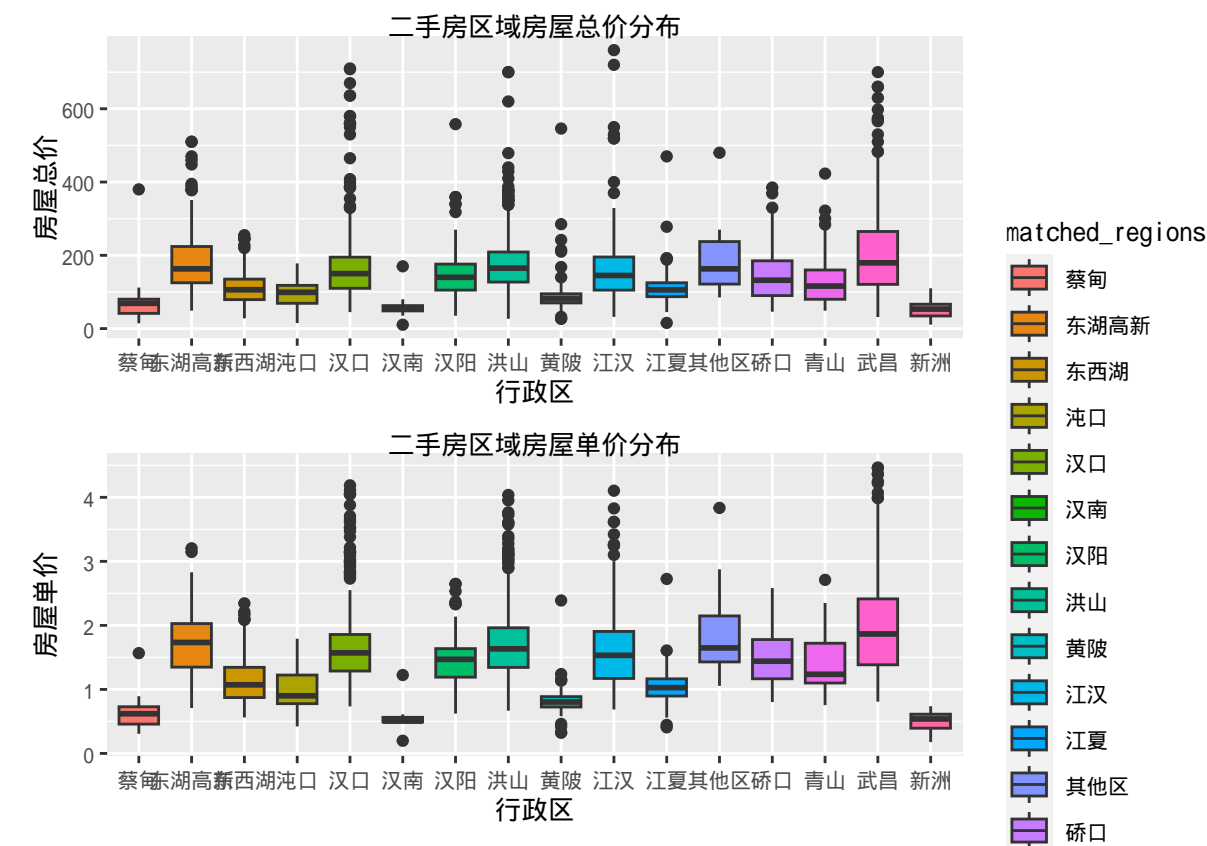
- 房屋属性中户型、朝向、装修都呈严重的偏态分布，户型中“2 室 3 厅”，朝“南”、“精装”的二手房都超过半数，符合普遍认知。





-对房屋属性和房屋总价做箱线图分析，四个变量中发现户型对房屋总价的影响最大。

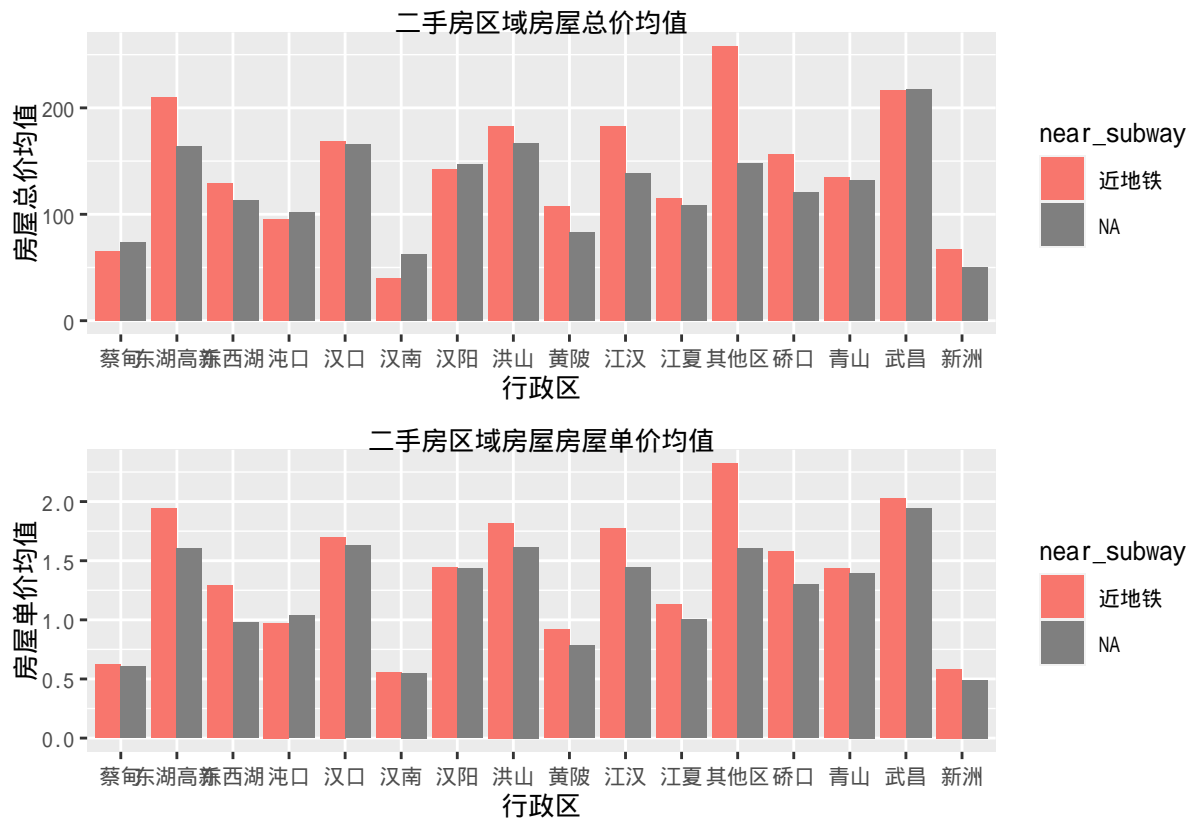
探索问题 1: 房屋所属行政区域的房价分布:



发现:

对房价所属片区进行行政区映射后，发现房屋在不同行政区，房屋的总价和均价的中心值都有较大差别，远城区的房屋价格都比较稳定。所属行政区的房屋单价中，武昌区的单价和均价均为最高且极值较多，东湖高新区作为近年新兴区域，价格紧随其后。汉口老城区 CBD 区域，价格稳居前三，符合普遍认知。

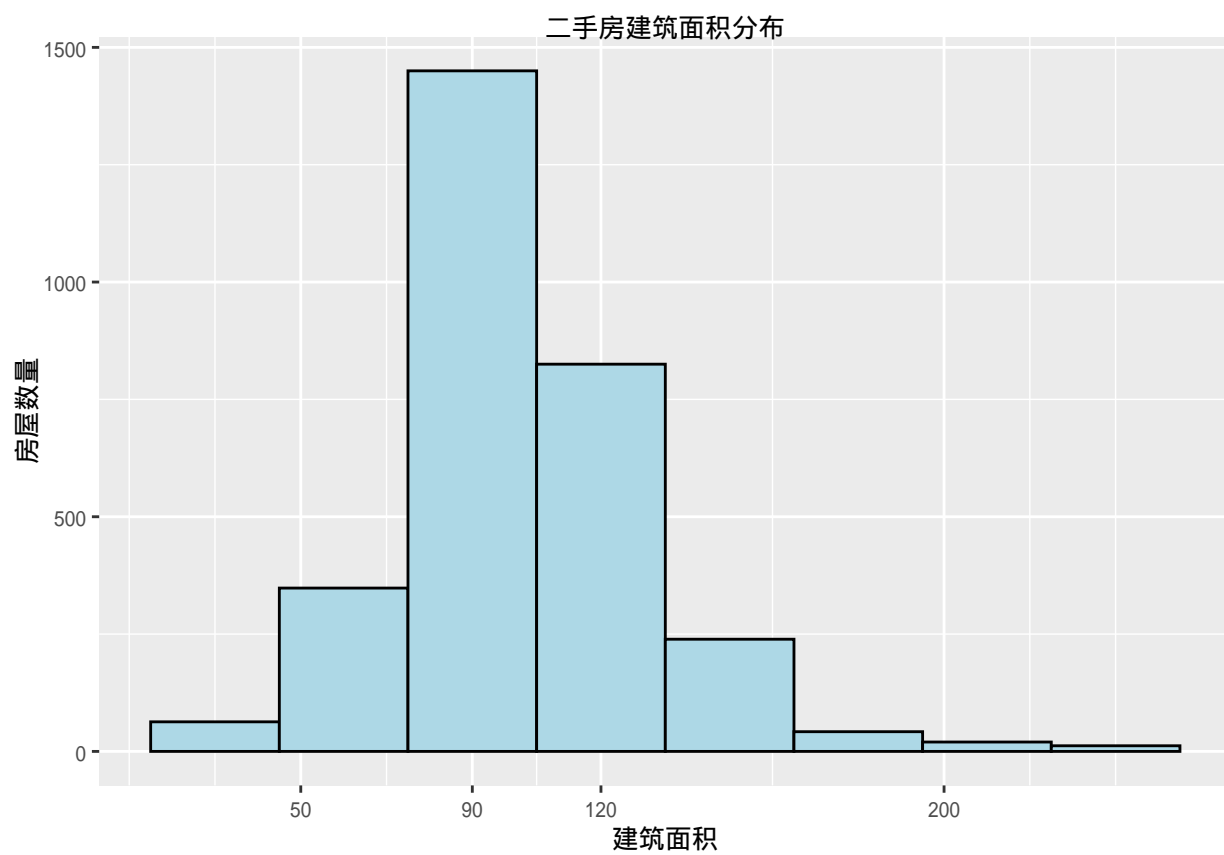
探索问题 2: 靠近地铁是否对房价会产生影响:

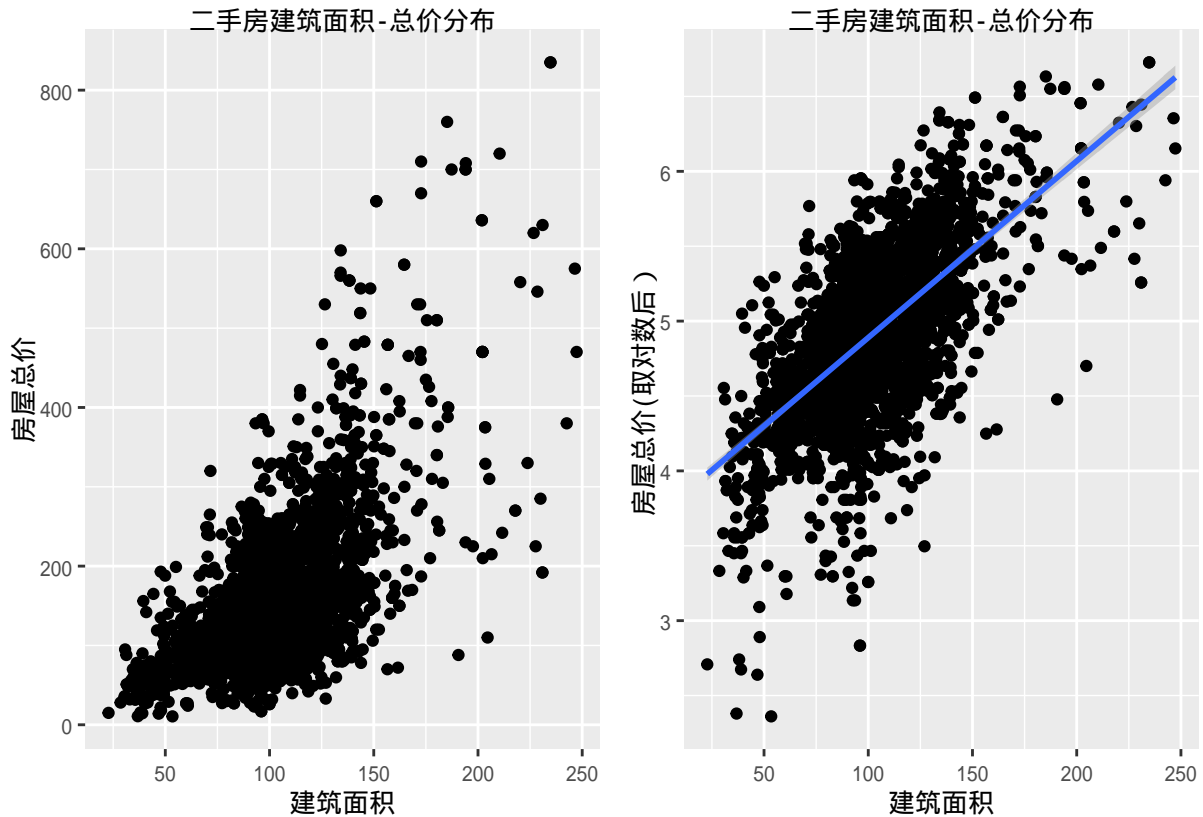


发现:

在 1.2 中对房屋属性中靠近地铁和未靠近地铁两组数据没有明显区别, 猜想是受黄陂、蔡甸等远城区实际中没有地铁到达的影响。对房屋靠近地铁对房价进一步进行分析, 按照行政区进行分组后, 虽然新洲、蔡甸等地区近地铁和未靠近地铁两组房价数据没有区别, 但在武昌、汉口区域两组数据差值也没有相等差距, 故无法得出靠近地铁房价比较高的结论。

探索问题 3: 对房屋面积对房价分布进行检验:





发现：

- 房屋面积为连续性变量，对房屋面积进行分组统计，发现在 90-120 平方米的房屋最多。
- 做房屋面积和总价的散点图，散点图呈现左密集右疏散，对总价取对数后，设置 95% 的置信区间，进行线性拟合，也是集中在 90-120 平方米区间中。

发现总结

1. 本次作业中对房屋属性的相关性进行检验时，发现数值型变量较少，只有建筑面积一个连续性变量进行分布分析，而且两次取对数后相关性系数并没有太大变化（0.68，0.63，0.68），而且分布也没有太大改变，有点黔驴技穷。
2. 其他字符型变量在后面工作中可以考虑进行数值标准化进行进一步分析。
3. 对于二手房关注房价的相关性影响，符合大众普遍认知是区域、面积、学区房这些因素，但是做出来的图不一定符合自我常识，所以结论根据观测值做出定性的结果有一点怀疑。