

对于武汉市二手房房价的一些探讨

李梓青

目录

1 你的主要发现	3
2 数据介绍	3
3 数据概览	4
4 探索性分析	7
4.1 房屋单价和房屋总价与房屋面积的关系	7
4.2 房屋数量对房屋单价的影响	9
4.3 层高对房屋单价的影响	10
4.4 地铁对房屋单价的影响	12
4.5 装修对房屋单价的影响	13
4.6 建筑形式对房屋单价的影响	14
4.7 房本是否满两年对房价的影响	15
4.8 高价房屋是否有其特殊的地方呢	17
4.9 探索各因素对于高价房屋的影响	18
5 发现总结	31

```
knitr::opts_chunk$set(echo = TRUE)
#load library
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.3      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr      1.0.2

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts

library(lubridate)
library(scales)

##
## 载入程辑包: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor

library(patchwork)
library(ggrepel)
library(ggplot2)
library(gridExtra)

##
## 载入程辑包: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(gplots)

##
## 载入程辑包: 'gplots'
##
## The following object is masked from 'package:stats':
##
##      lowess

library(showtext)

## 载入需要的程辑包: sysfonts
## 载入需要的程辑包: showtextdb

showtext_auto(enable = TRUE)
windowsFonts("MyFont" = windowsFont("Songti SC"))
```

1 你的主要发现

1. 武汉市二手房的普遍价格在 2 万/平以内且中等户型更多
2. 2-3 房的二手房反而能卖出更高价格
3. 各种变量因素对高价二手房的价格影响更多

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。

```
data <- read_csv("C:/Users/Shirl/Documents/2023-09-12_cleaned.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details.
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 3000 Columns: 18
## -- Column specification -----
## Delimiter: ","
## chr (11): property_name, property_region, directions1, directions2, decorati...
## dbl (7): price_ttl, price_sqm, bedrooms, livingrooms, building_area, proper...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

3 数据概览

表的前 10 行示例如下：

```
head(data, 10)
```

```
## # A tibble: 10 x 18
##   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 12 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>
```

各变量的简短信息:

```
glimpse(data)
```

```
## 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          <dbl> 18709, 14613, 15968, 15702, 17509, 10376, 12346, 163~
## $ 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, 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", "~
## $ 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, 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          <dbl> 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", ~
```

各变量的简短统计:

```
summary(data)
```

```
## property_name      property_region      price_ttl      price_sqm
## Length:3000      Length:3000      Min.   : 10.6      Min.   : 1771
## Class :character      Class :character      1st Qu.: 95.0      1st Qu.:10799
## Mode  :character      Mode  :character      Median : 137.0      Median :14404
##                                     Mean  : 155.9      Mean  :15148
##                                     3rd Qu.: 188.0      3rd Qu.:18211
##                                     Max.   :1380.0      Max.   :44656
## bedrooms      livingrooms      building_area      directions1
## Min.   :1.000      Min.   :0.000      Min.   : 22.77      Length:3000
## 1st Qu.:2.000      1st Qu.:1.000      1st Qu.: 84.92      Class :character
## Median :3.000      Median :2.000      Median : 95.55      Mode  :character
## Mean   :2.695      Mean   :1.709      Mean   :100.87
## 3rd Qu.:3.000      3rd Qu.:2.000      3rd Qu.:117.68
## Max.   :7.000      Max.   :4.000      Max.   :588.66
## directions2      decoration      property_t_height      property_height
## Length:3000      Length:3000      Min.   : 2.00      Length:3000
## Class :character      Class :character      1st Qu.:11.00      Class :character
## Mode  :character      Mode  :character      Median :27.00      Mode  :character
##                                     Mean   :24.22
##                                     3rd Qu.:33.00
##                                     Max.   :62.00
## property_style      followers      near_subway      if_2y
## Length:3000      Min.   : 0.000      Length:3000      Length:3000
## Class :character      1st Qu.: 1.000      Class :character      Class :character
## Mode  :character      Median : 3.000      Mode  :character      Mode  :character
##                                     Mean   : 6.614
##                                     3rd Qu.: 6.000
##                                     Max.   :262.000
## has_key      vr
## Length:3000      Length:3000
## Class :character      Class :character
```

```
## Mode :character Mode :character
##
##
##
```

可以看到：

- 1. 武汉市二手房房屋总价相差极大，最低只需要 10 万，最高 1380 万，但是绝大多数总价都不到两百万。房屋总价均值为 155.9 万，中位数为 137 万。
 - 2. 武汉市二手房房屋单价相差极大，最低 1771 元/平，最高 44656 元/平，但是绝大多数都在 2 万/平以内。
 - 3. 武汉市二手房的房间数主要还是以 2 房或者 3 房居多。大多数都有 1 个或者 2 个客厅。
 - 4. 武汉市二手房的建筑面积最小只有 22.77 平方米，最大有 588.66 平方米，中位数是 95.55 平方米，均值 100.87 平方米。
- 也就是说小户型卖的并不多，100 平左右，甚至 100 平以上的房子卖的挺多的。
- 5. 武汉市二手房房屋在所在楼栋所处位置最低在 2 楼，最高在 62 楼，中位数为 27 楼。绝大多数房屋都在 10 楼到 33 楼之间，也就是说房屋在中间楼层的卖的更多。
 - 6. 在该二手房网站的关注人数均值为 6.6 人，中位数是 3 个人。最火爆的房屋有 262 个人关注，但是有的房屋无人关注。

4 探索性分析

4.1 房屋单价和房屋总价与房屋面积的关系

```
p1 <- ggplot(data, aes(price_ttl)) + geom_histogram(binwidth=0.5, fill="blue")

p2 <- ggplot(data, aes(building_area)) + geom_histogram(binwidth=0.5, fill="blue")
```

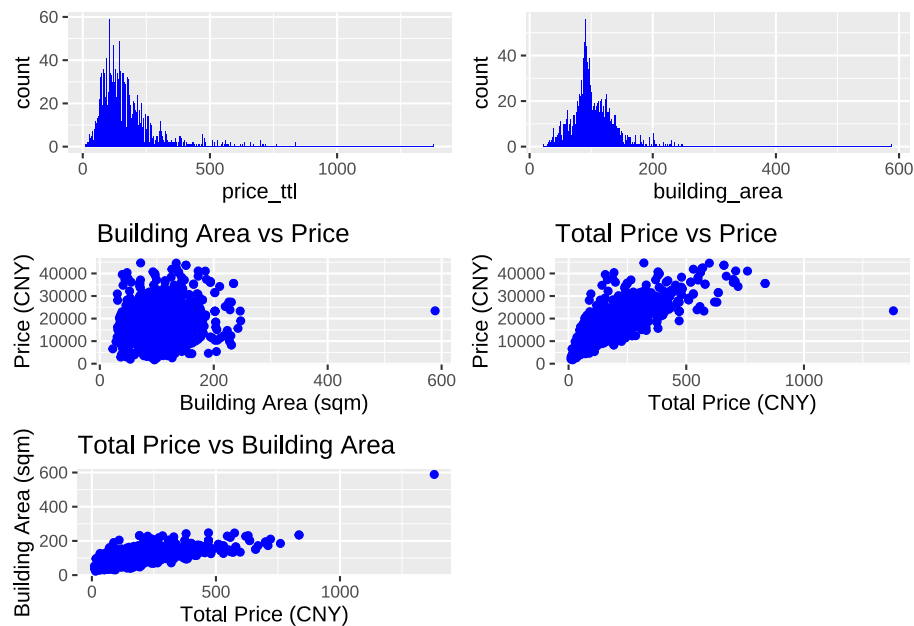
```

p3 <- ggplot(data, aes(x = building_area, y = price_sqm)) +
  geom_point(color = "blue") +
  labs(x = "Building Area (sqm)", y = "Price (CNY)") +
  ggtitle("Building Area vs Price")

p4 <- ggplot(data, aes(x = price_ttl, y = price_sqm)) +
  geom_point(color = "blue") +
  labs(x = "Total Price (CNY)", y = "Price (CNY)") +
  ggtitle("Total Price vs Price")

p5 <- ggplot(data, aes(x = price_ttl, y = building_area)) +
  geom_point(color = "blue") +
  labs(x = "Total Price (CNY)", y = "Building Area (sqm)") +
  ggtitle("Total Price vs Building Area")
grid.arrange(p1, p2, p3, p4, p5, nrow=3)

```



可以看到:

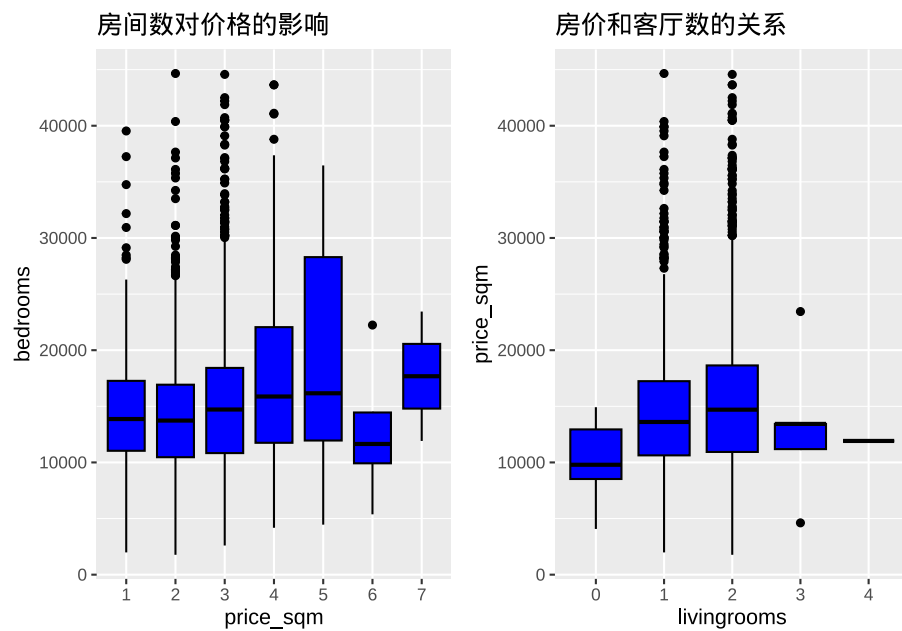
-从直方图中可以看出房屋总价和单价曲线呈现右偏的趋势，但是分部趋势不完全相同，里面一定有一些其他因素扰动。

600 万以上总价和 300 万以上建筑面积在分析的时候可以作为异常值剔除掉。

-从散点图可以看出，单价和总价，建筑面积和总价正相关。均价高的地方和建筑面积相关性没那么强，其中一定有其他因素的影响。

4.2 房屋数量对房屋单价的影响

```
p1 <- ggplot(data, aes(x = factor(bedrooms), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "price_sqm", y = "bedrooms") +  
  ggtitle(" 房间数对价格的影响")  
  
p2 <- ggplot(data, aes(x = factor(livingrooms), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "livingrooms", y = "price_sqm") +  
  ggtitle(" 房价和客厅数的关系")  
  
grid.arrange(p1, p2, nrow=1)
```

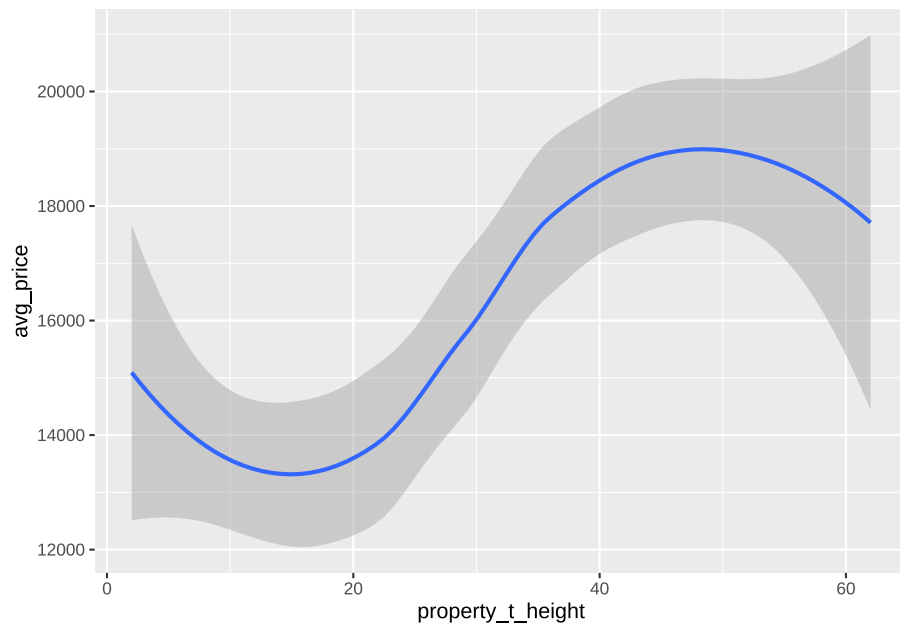


发现：

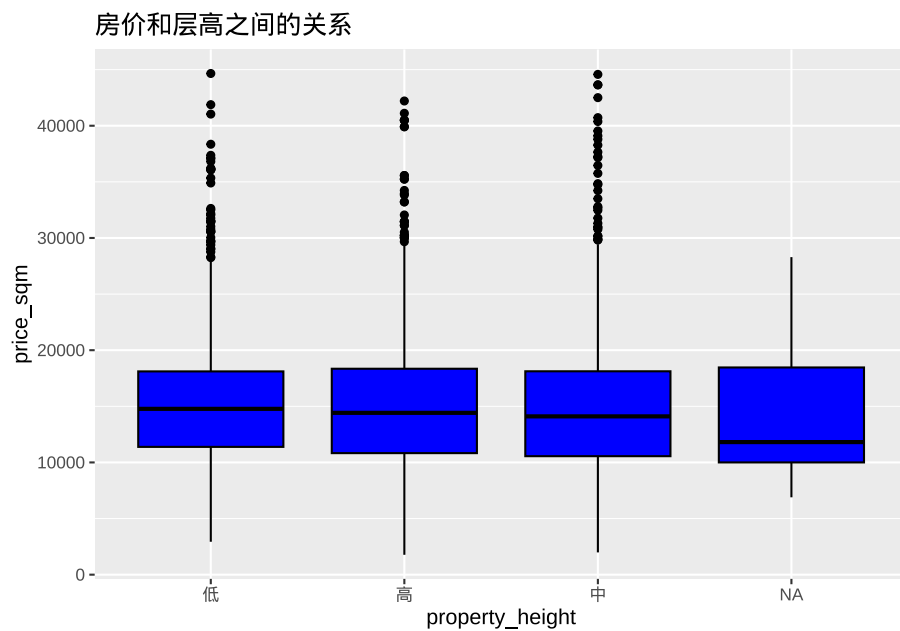
- 房屋价格的中位数随着房间数的增加会略微上升，但不是非常显著。
- 对于有五间房屋的房子，房屋单价相差特别大
- 有两个厅的房子房价是相对更高的。

4.3 层高对房屋单价的影响

```
average_prices <- data %>%  
  group_by(property_t_height) %>%  
  summarise(avg_price = mean(price_sqm))  
  
ggplot(average_prices, aes(x = property_t_height, y = avg_price)) +  
  geom_smooth()  
  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



```
ggplot(data, aes(x = factor(property_height), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "property_height", y = "price_sqm") +  
  ggtitle(" 房价和层高之间的关系")
```

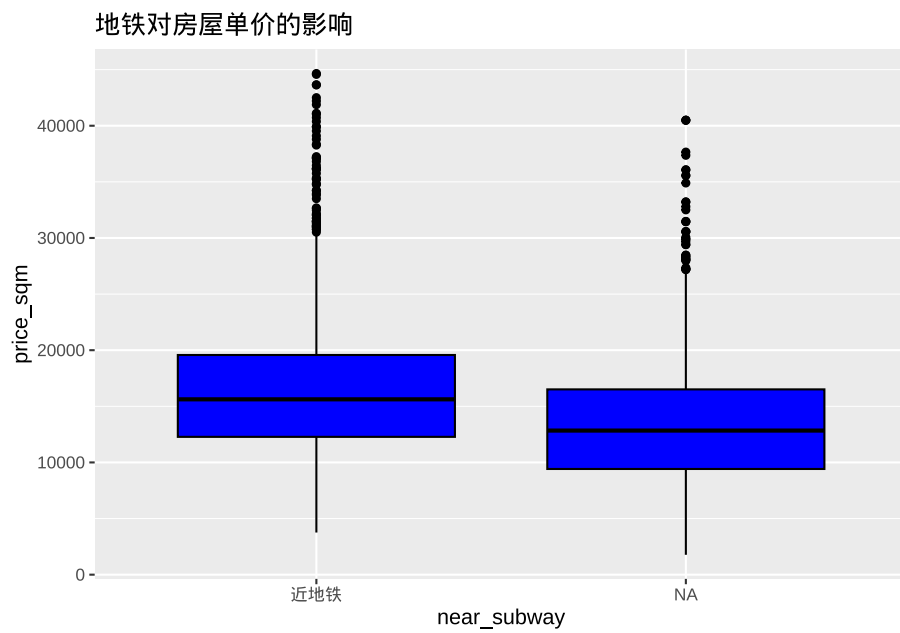


发现：

- 15 到 20 楼的房价最低；40-55 的房价最高。
- 对于房屋总体而言，低中高层对房价的影响不是很显著，低层房价比高层和中层房价更高一点。

4.4 地铁对房屋单价的影响

```
data_1 <- data %>%  
  filter(is.na(near_subway) | near_subway == "近地铁")  
  
ggplot(data_1, aes(x = factor(near_subway), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "near_subway", y = "price_sqm") +  
  ggtitle("地铁对房屋单价的影响")
```

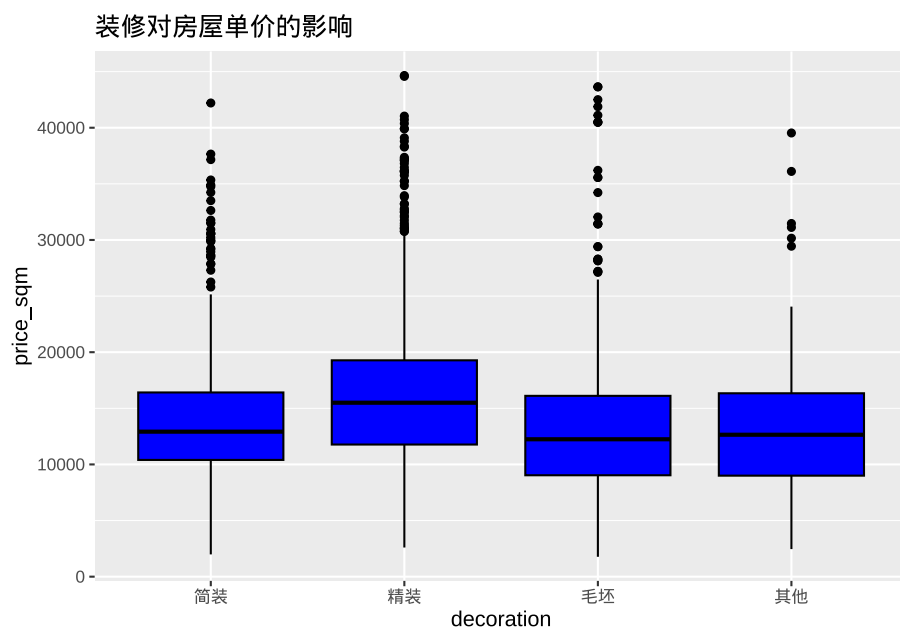


发现:

-靠近地铁的房价会更高一些

4.5 装修对房屋单价的影响

```
ggplot(data, aes(x = factor(decoration ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "decoration", y = "price_sqm") +  
  ggtitle(" 装修对房屋单价的影响")
```

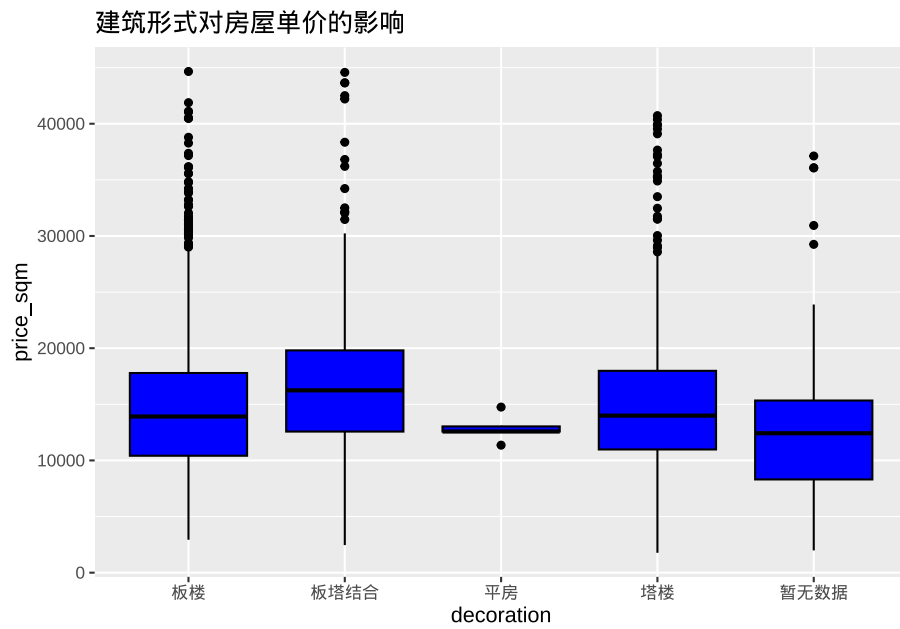


发现：

-装修程度对二手房影响挺大的。精装的价格普遍高于简装高于毛坯

4.6 建筑形式对房屋单价的影响

```
ggplot(data, aes(x = factor(property_style), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "decoration", y = "price_sqm") +  
  ggtitle(" 建筑形式对房屋单价的影响")
```

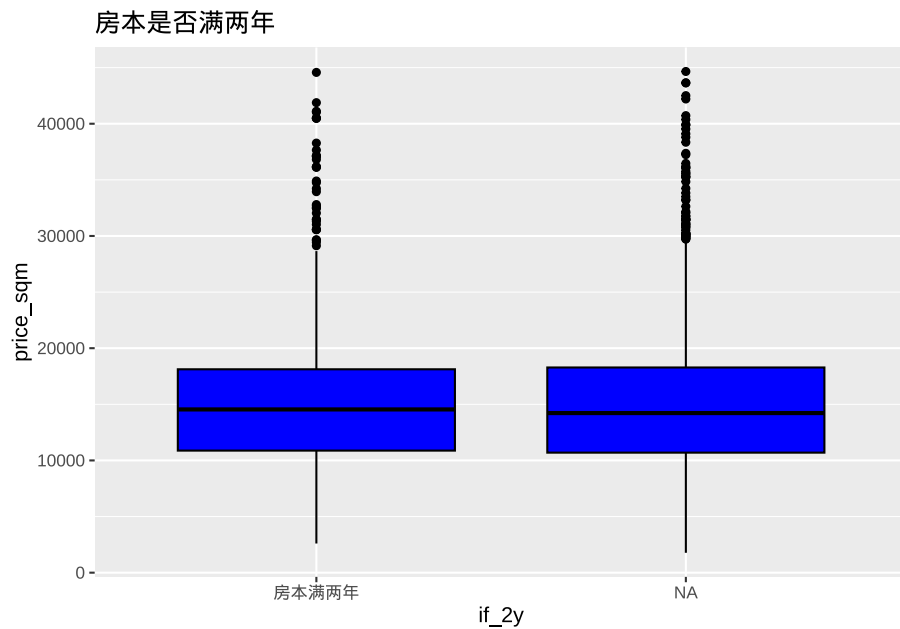


发现:

- 塔楼的房价略高于板楼
- 板塔结合是大家最喜欢的建筑形式，价格更高

4.7 房本是否满两年对房价的影响

```
ggplot(data, aes(x = factor(if_2y ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "if_2y", y = "price_sqm") +  
  ggtitle(" 房本是否满两年")
```

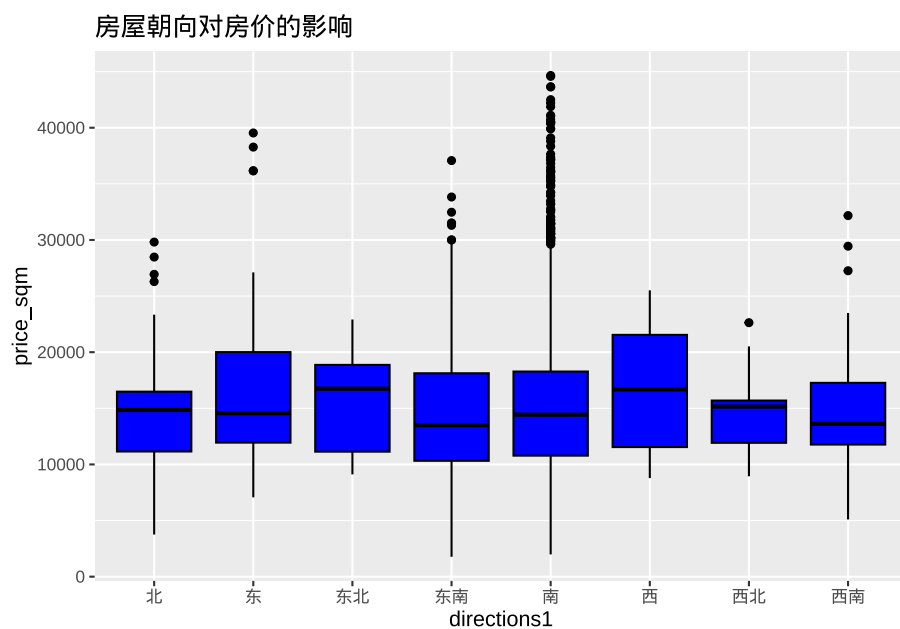


发现：

房本是否满两年对房价的影响并不是特别大。房本满两年的房子会略高于未满足两年的房子，这应该是由税的影响。但是从极值看，对高价房屋来说，没什么影响，这也许是因为有钱人对税费没那么敏感。

房屋主要朝向对房价的影响

```
ggplot(data, aes(x = factor(directions1 ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "directions1", y = "price_sqm") +  
  ggtitle(" 房屋朝向对房价的影响")
```

发现：房屋朝向对房价影响不是特别显著

4.8 高价房屋是否有其特殊的地方呢

```
top_n <- 20 # 想要找出的最高房价的地方数量
top_places <- data[order(-data$price_sqm), ][1:top_n, ]

# 输出最高房价的地方
print(top_places)
```

```
## # A tibble: 20 x 18
```

	property_name	property_region	price_ttl	price_sqm	bedrooms	livingrooms
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	中商宿舍	中南丁字桥	320	44656	2	1
## 2	复地东湖国际五六期	中北路	598	44574	3	2
## 3	复地东湖国际一期	中北路	660	43643	4	2
## 4	复地东湖国际一期	中北路	660	43643	4	2
## 5	复地东湖国际五六期	中北路	570	42503	3	2

##	6	复地东湖国际五六期	中北路	566	42205	3	2
##	7	华发外滩首府	黄埔永清	530	41878	3	2
##	8	华发外滩首府	黄埔永清	710	41110	4	2
##	9	华发中城荟	CBD西北湖	760	41037	4	2
##	10	复地东湖国际二期	中北路	380	40721	3	2
##	11	中信泰富滨江金融城~	二七	560	40492	3	2
##	12	中信泰富滨江金融城~	二七	560	40492	3	2
##	13	中信泰富滨江金融城~	二七	560	40492	3	2
##	14	水域天际	南湖沃尔玛	193	40377	2	1
##	15	复地东湖国际二期	中北路	385	39909	3	1
##	16	复地东湖国际二期	中北路	385	39909	3	1
##	17	水域天际	南湖沃尔玛	156	39534	1	1
##	18	复地东湖国际一期	中北路	380	39103	3	1
##	19	华发外滩首府	黄埔永清	670	38794	4	2
##	20	华清园	VR看装修	480	38351	3	2

```
## # i 12 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>
```

发现:

房屋单价最高的地方几乎很集中, 中北路, 黄埔永清, 二七, 南湖沃尔玛。

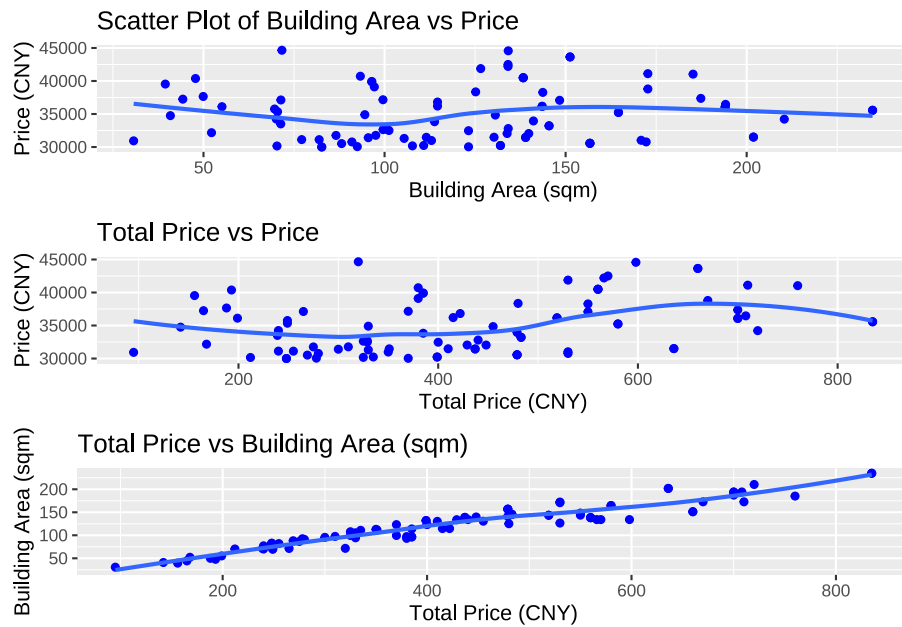
大胆猜测一些, 这些可能是学区房。复地东湖国际对应武昌实验小学(中北路), 华发外滩首府(黄埔永清)和中信泰富(二七)对应七一中学。

可惜这里没有具体的学区信息。和地图进行进一步对比太复杂了。

4.9 探索各因素对于高价房屋的影响

```
high_price_houses <- data[data$price_sqm >= 30000, ]
```

```
p1 <- ggplot(high_price_houses, aes(x = building_area, y = price_sqm)) +  
  geom_point(color = "blue") +  
  geom_smooth(se = FALSE)+  
  labs(x = "Building Area (sqm)", y = "Price (CNY)") +  
  ggtitle("Scatter Plot of Building Area vs Price")  
  
p2 <- ggplot(high_price_houses, aes(x = price_ttl, y = price_sqm)) +  
  geom_point(color = "blue") +  
  geom_smooth(se = FALSE)+  
  labs(x = "Total Price (CNY)", y = "Price (CNY)") +  
  ggtitle("Total Price vs Price")  
  
p3 <- ggplot(high_price_houses, aes(x = price_ttl, y = building_area)) +  
  geom_point(color = "blue") +  
  geom_smooth(se = FALSE)+  
  labs(x = "Total Price (CNY)", y = "Building Area (sqm)") +  
  ggtitle("Total Price vs Building Area (sqm)")  
  
grid.arrange(p1, p2,p3, nrow=3)  
  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

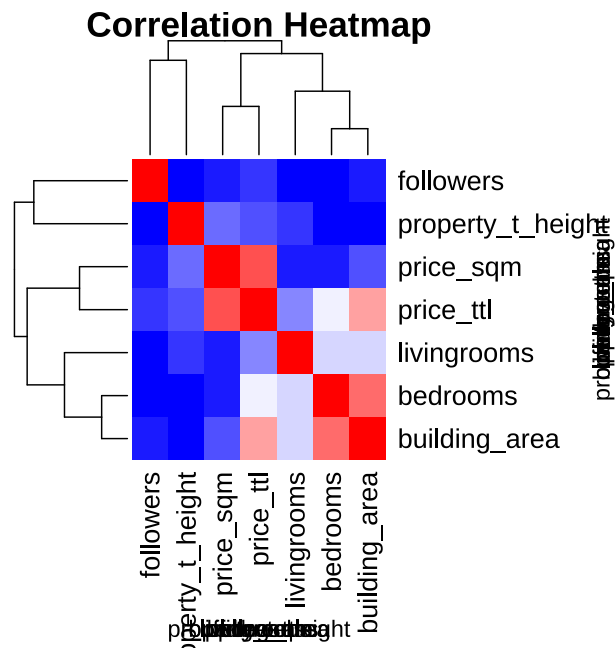


发现:

-和总体比，房屋的建筑面积和房屋总价的相关性变高了

总体热力图与高价房屋热力图进行对比

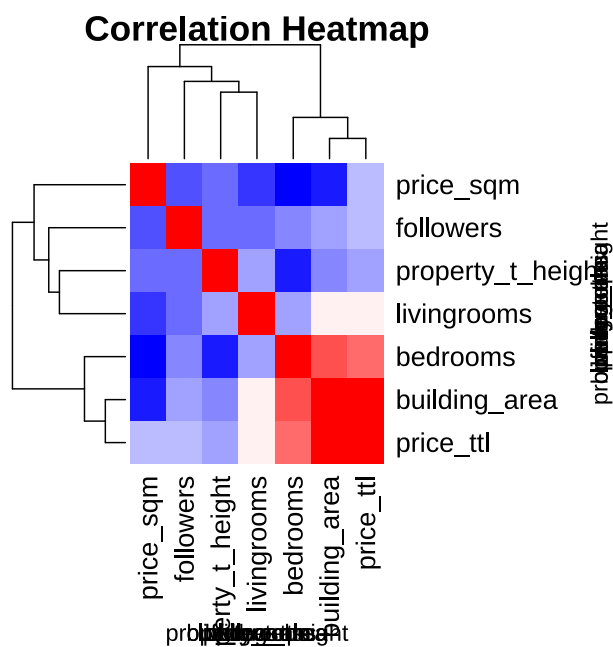
```
selected_data <- data[c("price_ttl", "price_sqm", "livingrooms", "bedrooms", "building_a",
                        "followers", "property_t_height")]
correlation_matrix <- cor(selected_data)
heatmap(correlation_matrix,
        col = colorRampPalette(c("blue", "white", "red"))(20), # 定义颜色范围
        main = "Correlation Heatmap", # 图表标题
        xlab = colnames(correlation_matrix), ylab = colnames(correlation_matrix), # X
        margins = c(8, 10), # 设置边距
        cex.lab = 1.5, cex.axis = 1.5, # 设置标签文本大小
        key.title = NULL, # 不显示颜色键标题
        symm = TRUE # 对称显示
)
```



```

selected_data <- high_price_houses[c("price_ttl", "price_sqm", "livingrooms", "bedrooms",
                                     "followers", "property_t_height")]
correlation_matrix <- cor(selected_data)
heatmap(correlation_matrix,
        col = colorRampPalette(c("blue", "white", "red"))(20), # 定义颜色范围
        main = "Correlation Heatmap", # 图表标题
        xlab = colnames(correlation_matrix), ylab = colnames(correlation_matrix), # X
        margins = c(8, 10), # 设置边距
        cex.lab = 1.5, cex.axis = 1.5, # 设置标签文本大小
        key.title = NULL, # 不显示颜色键标题
        symm = TRUE # 对称显示
)

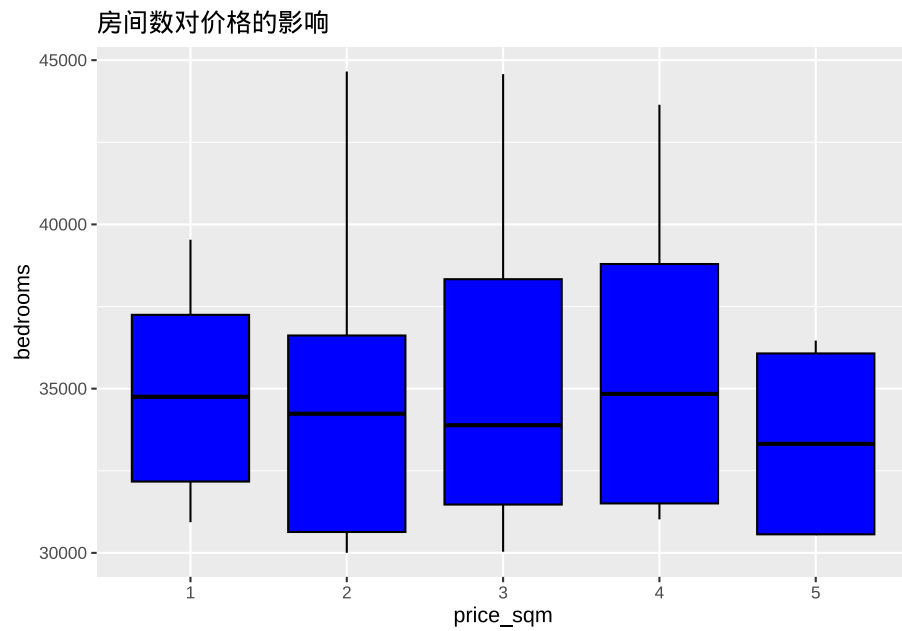
```



发现：

- 和总体热力图相比，高价房屋的热力图红色增加，深蓝色减少。各个变量对高价二手房的影响更深
- 对于高价房屋来说，房屋的面积和房间的数量和房价相关性很高。这个结论和前面的线型图是对应的。

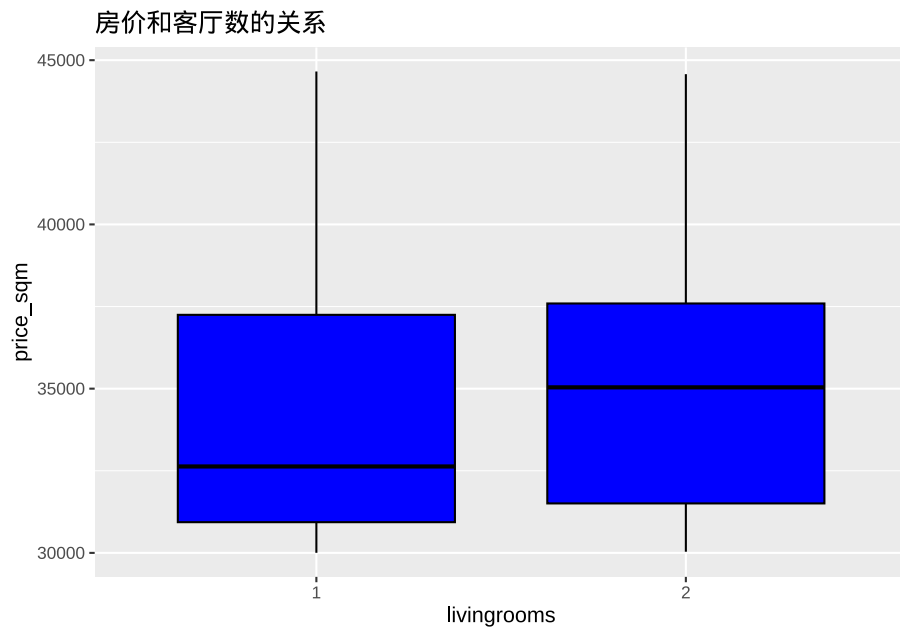
```
ggplot(high_price_houses, aes(x = factor(bedrooms), y = price_sqm)) +
  geom_boxplot(fill = "blue", color = "black") +
  labs(x = "price_sqm", y = "bedrooms") +
  ggtitle(" 房间数对价格的影响")
```



发现：

- 对于高价房屋来说，房屋单价范围比较接近
- 有三四间房屋的房子房价偏高一些
- 但是最高单价的房屋是只有 2 间房的

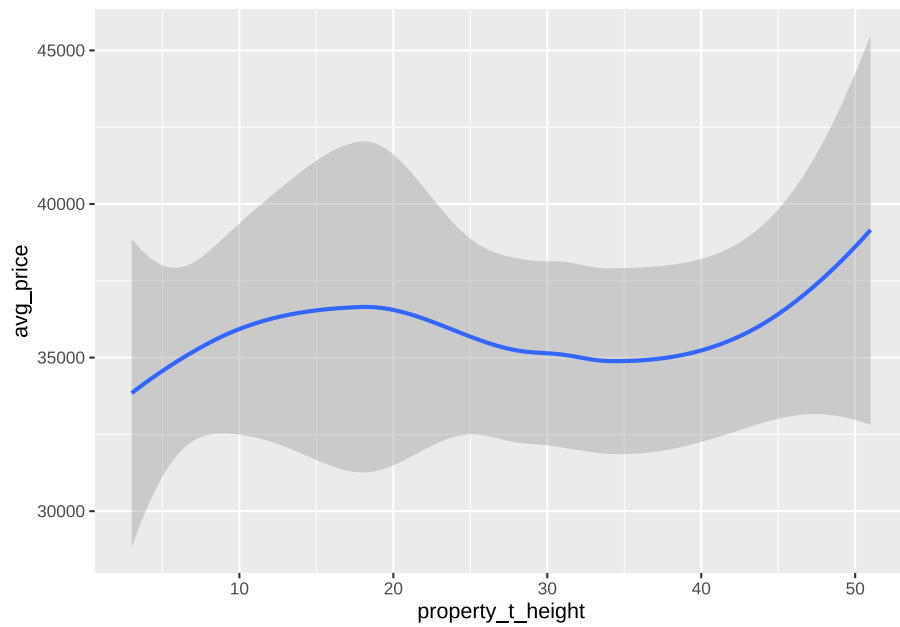
```
ggplot(high_price_houses, aes(x = factor(livingrooms), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "livingrooms", y = "price_sqm") +  
  ggtitle(" 房价和客厅数的关系")
```



发现:

-对于高价房屋来说，有两个厅的房子普遍更贵一些

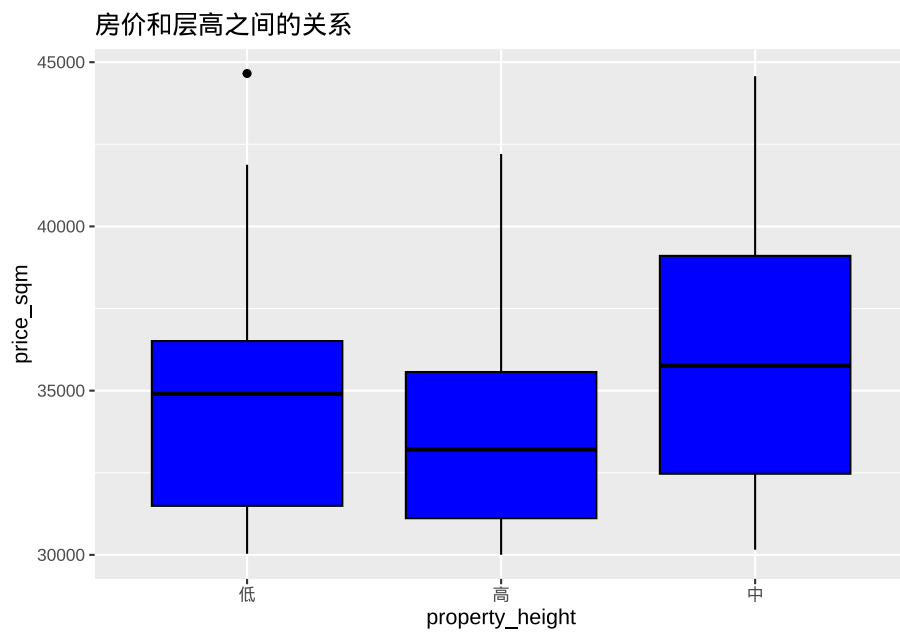
```
average_prices1 <- high_price_houses %>%  
  group_by(property_t_height) %>%  
  summarise(avg_price = mean(price_sqm))  
  
ggplot(average_prices1, aes(x = property_t_height, y = avg_price)) +  
  geom_smooth()  
  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

发现:

-在高价房屋中，18 楼左右或者 50 楼以上的房价会更高一些。和前面相比，层高对价格的影响没那么大了。

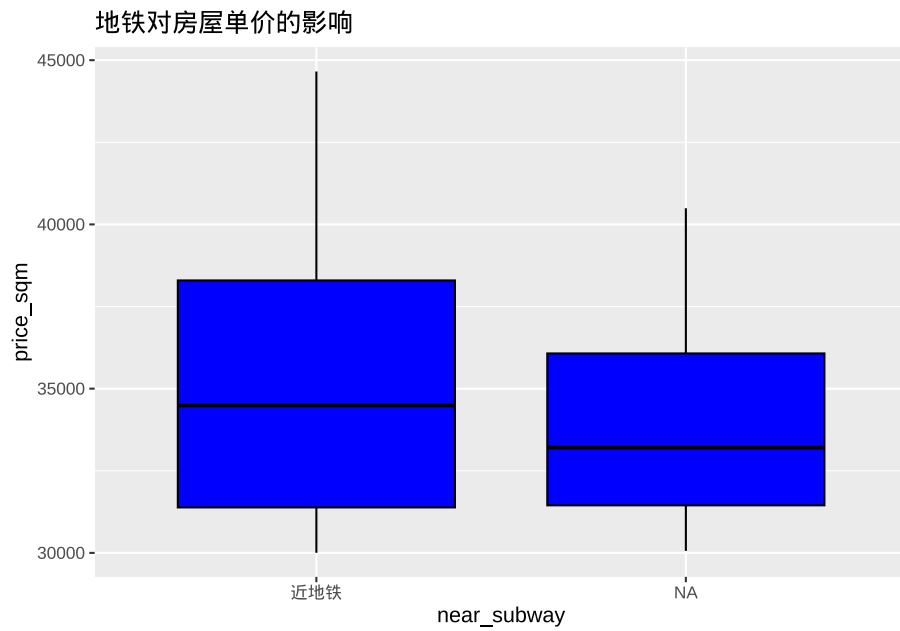
```
ggplot(high_price_houses, aes(x = factor(property_height), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "property_height", y = "price_sqm") +  
  ggtitle(" 房价和层高之间的关系")
```



发现:

-在高价房屋中，中层房屋的价格会更贵一些

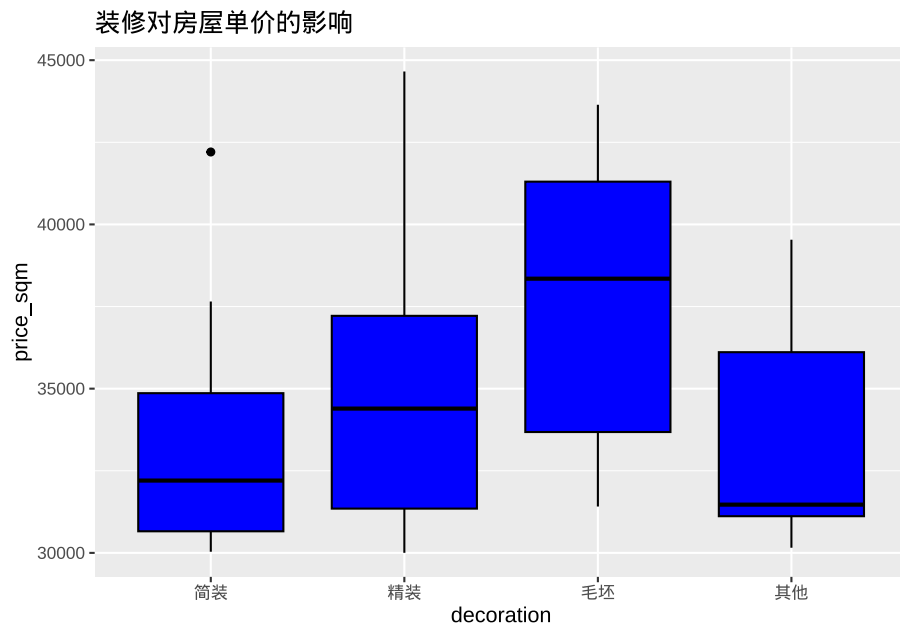
```
data_2 <- high_price_houses %>%  
  filter(is.na(near_subway) | near_subway == "近地铁")  
  
ggplot(data_2, aes(x = factor(near_subway), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "near_subway", y = "price_sqm") +  
  ggtitle("地铁对房屋单价的影响")
```



发现:

-有无地铁对高价房屋来说没有总体那么显著，可能是因为有钱人一般都有车。

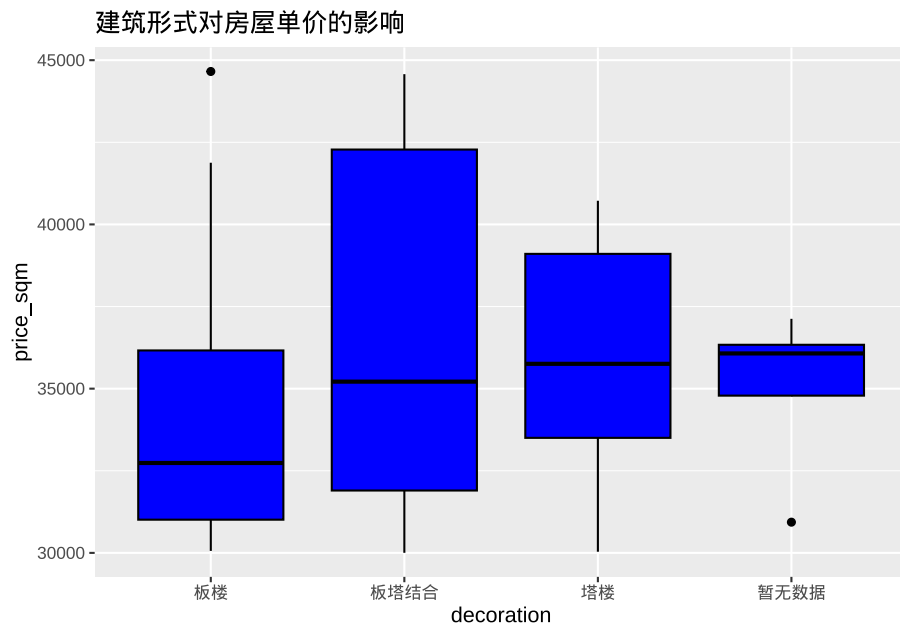
```
ggplot(high_price_houses, aes(x = factor(decoration ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "decoration", y = "price_sqm") +  
  ggtitle(" 装修对房屋单价的影响")
```



发现:

-对于高价房屋来说，毛坯房的价格反而更高一些。

```
ggplot(high_price_houses, aes(x = factor(property_style), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "decoration", y = "price_sqm") +  
  ggtitle(" 建筑形式对房屋单价的影响")
```

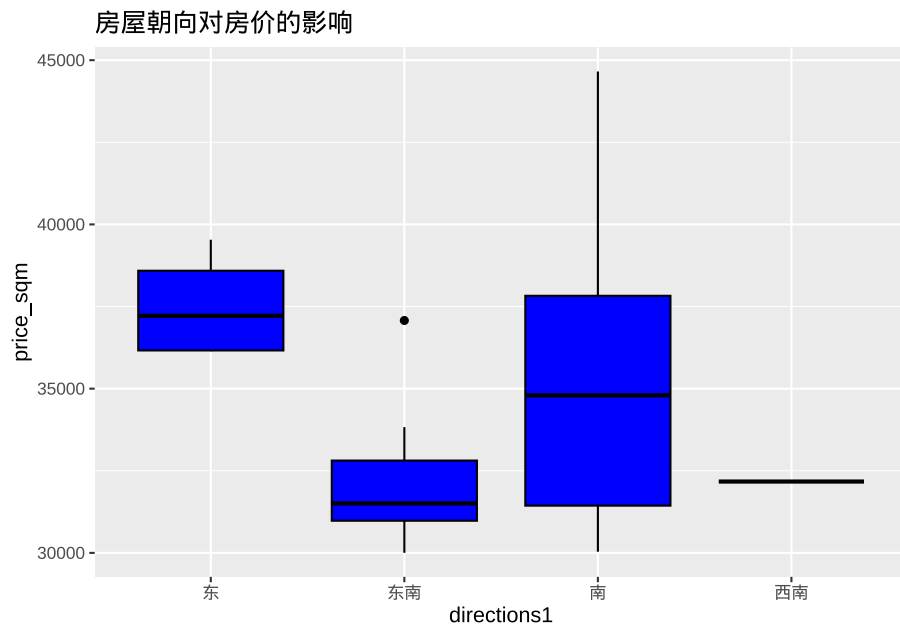


发现:

-在高价房屋里面，塔楼的价格显著要与板楼

-板塔结合的房屋价格相差极大

```
ggplot(high_price_houses, aes(x = factor(directions1 ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "directions1", y = "price_sqm") +  
  ggtitle(" 房屋朝向对房价的影响")
```

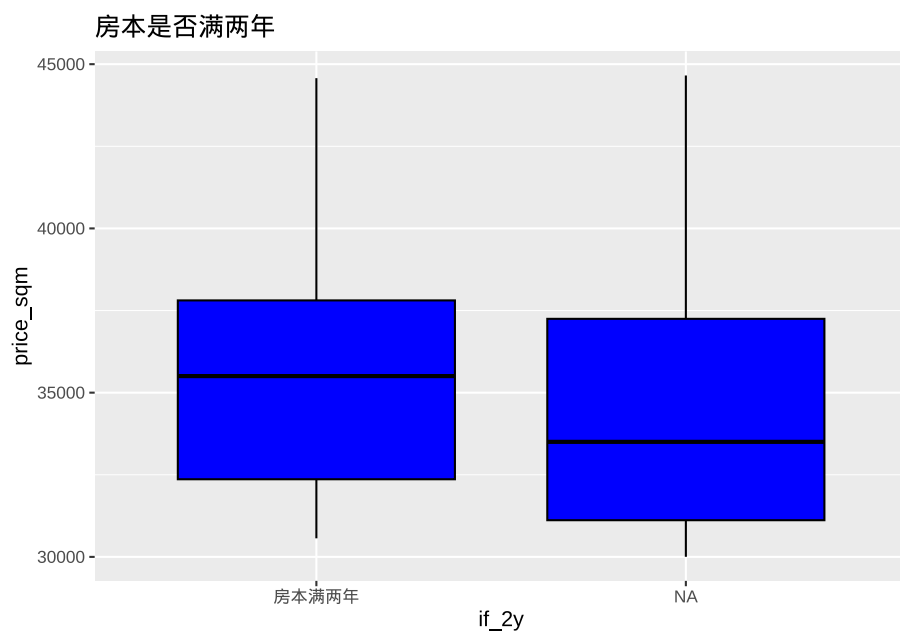


发现:

-在高价房屋中，主要房屋朝东价格更高

-主要房屋朝南的价格相差极大

```
ggplot(high_price_houses, aes(x = factor(if_2y ), y = price_sqm)) +  
  geom_boxplot(fill = "blue", color = "black") +  
  labs(x = "if_2y", y = "price_sqm") +  
  ggtitle(" 房本是否满两年")
```



发现:

-在高价房屋中，满两年的房子价格更高

5 发现总结

从总体数据来看，1.5 元/平的房源是武汉二手市场交易的主流，在这个价格区间内，中位数和均值相差不大，可供选择的范围比较多，可以考虑是否临近地铁，楼层建筑形式，楼层，装修等级等因素。

对于 3 万以上高价格的房子来说，地理位置、朝向、大小、楼层建筑形式对房价的影响更高一些。可能是因为买家不再是以价格优先，会更多的考虑舒适度和使用价值。与普通房源相比，毛坯高价房反而能卖出更高价格。高价房的买家也会关注地铁、房龄等因素，但是相对而言重视度没那么高，他们不是高价房屋价值的决定性因素。