Hotels in Brussels Analysis

Maeva Braeckevelt

Introduction

This analysis aimed at analyzing the pattern of association between the price of the hotels in Brussels and the distance from the center and then compare to the Vienna result. The data used was gathered in a csv files : The hotelbookingdata.csv. It was download from a comparison website and it was anonymized and slightly altered to ensure confidentiality.The main variables that I used were: the price in dollars (y) and the distance to the city center in miles (x). My sample is the price of the hotels in Brussels whithin 5 miles, in the weekday of November 2017 of a maximum of 400 dollars.

Histogram and summary statitics

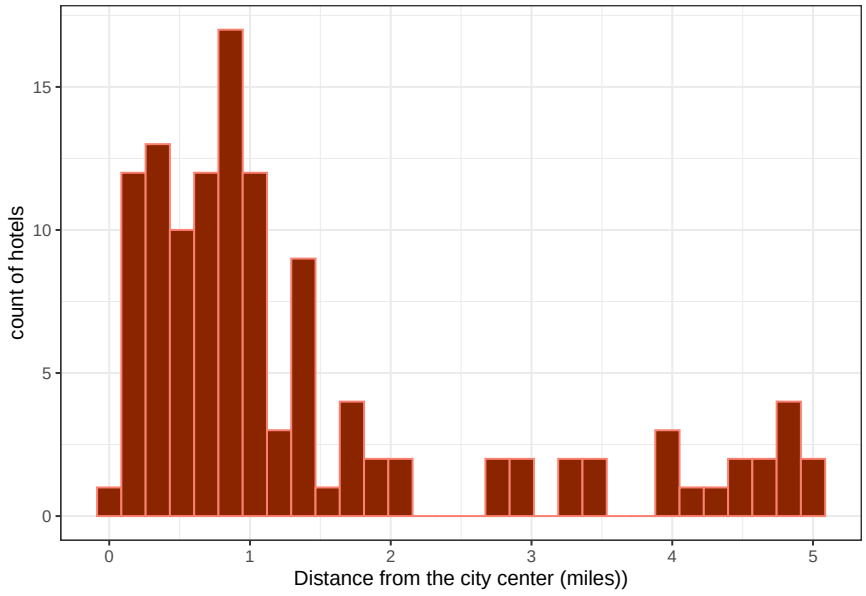


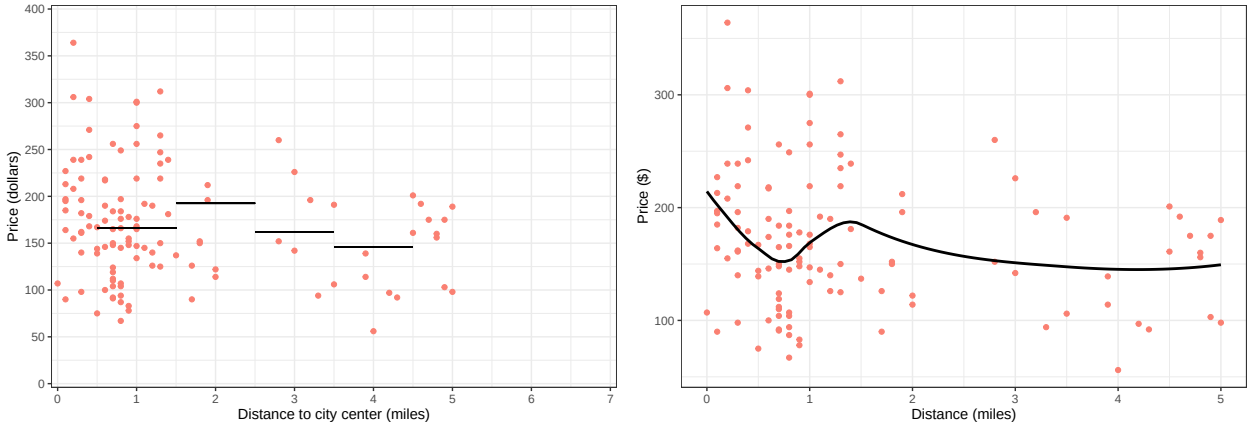
Table 1: Summary statistic of the Price

variable	mean	median	min	max	sd	skew
price	169.314	164	56	364	60.11642	0.6070223

The minimun price for a night is 56 dollars and the maximum 364 dollars. The mean is 169,31 dollars. I observed that the distribution of the distance is skewed with a right tail and some extreme values.

Non-parametric regression

Bin scatter with four bins & Lowess regression



I observed that there is a negative slope in general. Further away from the city center, cheaper are the hotels. However, I observed that around 2 miles, the price gets higher. This spike was not present where we were comparing hotels in Vienna. One explanation could be that around 2 miles from the center, there is a expensive neighborhood. With these two non-parametric regressions I can not answer quantitative answers, I can only evaluate by the graphs, a difference on average of 20$ between close and far hotels from the city center. So to have more quantitative data I will do a linear regression.

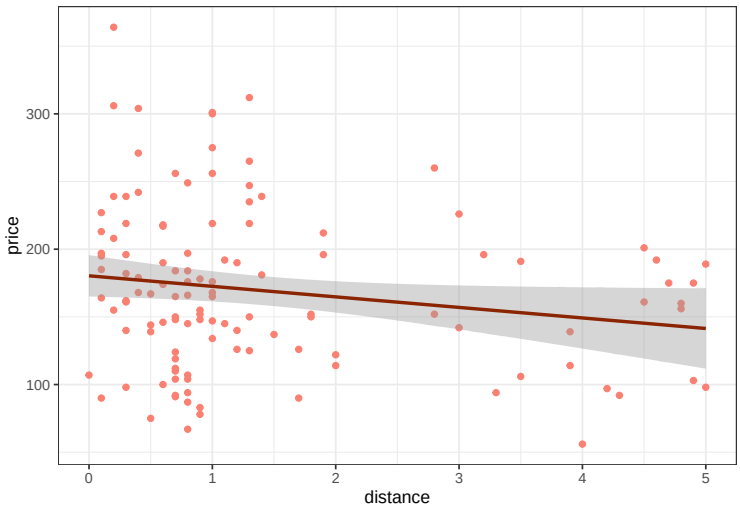
Simple Linear regression (A5)

Call: lm_robust(formula = price ~ distance, data = hotels_brussels, se_type = "HC2")

Standard error type: HC2

Coefficients: Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF (Intercept) 180.339 7.907 22.807 2.945e-45 164.68 195.996 119 distance -7.783 3.304 -2.356 2.013e-02 -14.33 -1.241 119

Multiple R-squared: 0.03239 , Adjusted R-squared: 0.02426 F-statistic: 5.549 on 1 and 119 DF, p-value: 0.02013



This graph represents the simple linear regression between the price of hotels in Brussels and the distance to the city center.

Formula : Price = 180,34 - 7,78 * distance

Alpha : 180,34 is the average price of the hotel when the distance is equal to 0.

Beta : the hotels that are 1 miles further away from the city center are, on average, 7,78$ cheaper

However my R square is 3%, so only 3% of variation of the price is captured in this regression. I observed as well that the line doesn't fit very well the graphs. There are some very high residuals.

	Linear
(Intercept)	180.34 ***
	(7.91)
distance	-7.78 *
	(3.30)
nobs	121
r.squared	0.03
adj.r.squared	0.02
statistic	5.55
p.value	0.02
df.residual	119.00
nobs.1	121.00
se_type	HC2.00

*** p < 0.001; ** p < 0.01; * p < 0.05.

Conclusion

I can see that the linear regression give a more quantitative answer but we could already capture the pattern with the non-parametric regression. Compare to Vienna (slope of 14), the price of hotels in Brussels don't get as much cheaper when you go further away from the city center.