

SPATIAL PROPERTY SALES STATISTICS OF HUNGARIAN HOT-SPOT SETTLEMENTS

Data Visualization Challenge – BI Forum, Budapest, 2016

Zsuzsanna Szabó

Environmental physicist-geochemist, PhD, Data Science and R enthusiast

Contact: +36-30-286-0879, zsszabo86@gmail.com, <https://www.linkedin.com/in/zsuzsanna-szab%C3%B3-64002430>

Introduction and goals

Ingtatlan.com has provided a dataset for the purpose of a data visualization challenge at BI Forum, Budapest, 2016. The dataset contains Hungarian property posts, either for sale or rent, which were created after 09 February 2015 and were not active on 01 September 2016. All properties are below 70 m². In this project an app is developed for the visualization of the spatial distribution of a few statistics of properties for sale.

Methods

All the work has been done by R in RStudio. The visualization is focused on “hot-spot” settlements which are the ones with a minimum of a hundred records in the dataset (package *dplyr*). The names of locations had to be manually edited and their coordinates were also added by the help of *geocode* function.

In an exploratory data analysis the price/sqm (m²) values have been grouped by date and their median values were taken. I observed that this daily median price close to linearly increased with time in the available period. Therefore, a linear trend line was used to calculate the price increase (trend) in %/month units for the different locations. By the same equation, I also calculated a corrected median price for each day if it were at the last day of the dataset.

After the preparations of the dataset, the app has been coded by the package *shiny* where the main output is an interactive map made by the package *leaflet*. The settlements and districts of Budapest are marked by different circles where the radius is related to the corrected median price and the color to the trend. If the user clicks a circle a popup provides a few, more detailed information. The user may also select a city to zoom at on the map, then the app also plots by package *ggplot2* the trend line for the given location (giving some idea of the uncertainty of the numbers presented on the map), and two histograms of add views per day and active days which are also useful variables when considering **investment**.

Product

Spatial Property Sales Statistics of Hungarian Hot-spot Settlements

