

Data Carpentry in R

Fabrice Rossi

December 18, 2017

1 In memory data manipulations

The following exercises are based on the [HM Land Registry Price Paid Data](#). Extract from the original data are available on the [course web site](#)¹.

Data should be loaded from those files in the `data.table` format.

1.1 Simple calculations

1. Determine the County for which the median Price is the largest (over the whole time period).
2. Plot the evolution of the yearly median Price of the least expensive County (overall based on the median Price).
3. Determine the month of the year that sees the largest number of transactions over the whole time period.
4. Compute the number of distinct cities in the data set.

1.2 Advanced calculations

1. Plot the evolution of the monthly median Price over the period, together with the first and third monthly quartiles of the Price.
2. Create a new data table containing for each County the median Price per year. More precisely, the data table should contain as many columns as there are years in the data set plus one column for the County.
3. Create a new data table listing for each County the Town/City with the largest number of transactions over the full period.

¹<http://apiacoa.org/teaching/big-data/index.en.html>