Private Label Demand: Zillow Data Construction

Günter J. Hitsch

1/13/2017

```
library(bit64)
library(data.table)
library(stringr)          # For the str_sub function
```

Get the Zillow data (in csv format) from the Zillow website:

```
zillow_DT = fread("http://files.zillowstatic.com/research/public/Zip/Zip_Zhvi_AllHomes.csv")
names(zillow_DT)[1:20]
```

```
 [1] "RegionID"   "RegionName" "City"       "State"      "Metro"
 [6] "CountyName" "SizeRank"   "1996-04"    "1996-05"    "1996-06"
[11] "1996-07"    "1996-08"    "1996-09"    "1996-10"    "1996-11"
[16] "1996-12"    "1997-01"    "1997-02"    "1997-03"    "1997-04"
```

Note that all Zillow home value indices are in columns!

`melt` the table from wide to long format:

```
zillow_DT = melt(zillow_DT, id.vars       = names(zillow_DT)[1:7],
                            measure.vars  = names(zillow_DT)[8:ncol(zillow_DT)],
                            variable.name = "date",
                            value.name    = "zillow_index")
head(zillow_DT)
```

```
   RegionID RegionName     City State     Metro CountyName SizeRank    date
1:    61639      10025 New York    NY New York   New York        1 1996-04
2:    84654      60657  Chicago    IL  Chicago       Cook        2 1996-04
3:    84616      60614  Chicago    IL  Chicago       Cook        3 1996-04
4:    93144      79936  El Paso    TX  El Paso    El Paso        4 1996-04
5:    61616      10002 New York    NY New York   New York        5 1996-04
6:    84640      60640  Chicago    IL  Chicago       Cook        6 1996-04
   zillow_index
1:           NA
2:       146700
3:       198000
4:        70800
5:           NA
6:       102300
```

Clean the data:

1. Convert `date` to year and month
2. `RegionName` is the ZIP code variable
3. Remove some variables, set preferred column order and key

```
zillow_DT[, date := as.character(date)]
zillow_DT[, `:=`(year     = as.integer(as.numeric(str_sub(date, 1L, 4L))),
                 month    = as.integer(as.numeric(str_sub(date, -2L))),
                 zip_code = as.integer(RegionName))]
zillow_DT[, c("RegionID", "RegionName") := NULL]
```

```r
setcolorder(zillow_DT, c("zip_code", "year", "month", "zillow_index", "date",
                         "City", "State", "Metro", "CountyName", "SizeRank"))
setkey(zillow_DT, zip_code, year, month)
```

Save the data.

```r
save(zillow_DT, file = "./Data/Zillow-Data.RData")
```