Readme

Chartreader James Kurtak

May 8, 2017

The data management and analysis for the paper were conducted using R version 3.3.2 "Sincere Pumpkin Patch" using RStudio version 1.0.136. It has been written to run on a Windows operating system. In addition to the base R package, several packages are used. These can be seen at the top of the Rmd document loaded in the first code chunk. The paper was knitted to pdf using the elsevier article format contained in the rticles package. Knitting to pdf requires a LATEX installation on your computer, though you will not need to write any LATEX code directly.

The documentation is organized into three main directories (listed in order of importance for reproduction of results):

- **Source code:**

Here you will find the Rmd file that reproduces the entire paper, along with all auxiliary files created in process of knitting the pdf. Appended to the paper, you will find the Data Appendix which describes all variables used in greater detail.

- **Processing and analysis:**

contains all data called by the Rmd file in analysis.The importable data folder contains the original data files modified to the minimal extent necessary to import them into R. All further processing is done in the code.

- **Original Data and Metadata:**

contains the data we collected in the most original form possible. For the IRS and Open PV data, these files as exactly as downloaded from the websites. The census data was inputed manually into an excel sheet and was gathered from Factfinder.com. Should you want to verify or replicate the data, simply go to https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml and input one of the zip codes found in the SolMerge data frame within the RMD file found in the Source code folder. After searching for a given zip code, the data used in this analysis can be found in the link titled "General Population and Housing Characteristics (Population, Age, Sex, Race, Households and Housing, . . .)" on the main results page, and the link Educational Attainment (High School, Bachelor's, Advanced Degree, . . .) found through the Education tab. (Note, although the data link is under the "2015 American Community Survey" subheader, it also contains data found through the 2010 census).

To reproduce all results reported in paper:

First, copy the Documentation folder to your computer. Make sure to leave the folder structurally intact. The Documentation folder can be moved or renamed as necessary, but everything else must remain unchanged.

Next, open the Rmd file in the Source Code folder using R Studio. Make sure all required packages are installed. Run all code in the file. If the output format is changed to an R notebooks format, the chunks can be run one at a time to see individual results.

Finally, knit the file to elsevier article. (Knitting to other pdf formats should change only the formatting. You must have a LATEX installation on your computer to do this.