README-for-details-on-analysis-using-R

Raw Data

This folder contains a data file that can be loaded into R that includes the raw data downloaded from the Open Baltimore data portal on February 3, 2016. This R data file is named:

open-bpd-raw-data-downloaded-2016-02-03.RData

This file includes two data frames that contain the same information that is in crimes.original.csv and arrests.original.csv, both of which are offered in the top folder of this data distribution.

Analysis Data

This folder contains a data file that can be loaded into R that includes the data analyzed for Appendices 2 and 3. The R data file is named:

model-ready-data.RData

This file includes two data frames that contain the same information that is in crimes.type.week.csv and arrests.type.week.csv, both of which are offered in the top folder of this data distribution.

R Analysis Code (as a project in RStudio)

The open_bpd folder contains an RStudio project file named:

open_bpd.Rproj

Open this file with the "Open Project…" command from the dropdown File menu in RStudio. This command will also load a file, .Rprofile, which is a hidden file within the open_bpd folder. The .Rprofile file will install the pacman package, which will then download necessary R packages for the analysis. The list of required packages for the project is written on lines 8-11 of the .Rprofile. The user should verify that all of these packages have been properly installed.

The pdf files for Appendix 2 and Appendix 3 can be recreated using RStudio, which will call LaTeX using the knitr package. The package knitr will be installed by the .Rprofile file when the project is loaded (see above), but LaTeX needs to be installed separately. Information on LaTeX is available here: https://www.latex-project.org/. We used the MacTeX installation here: http://tug.org/mactex/. The open_bpd.Rproj file will set RStudio to use knitr and pdfLaTeX. To understand knitr, see (as of March 15, 2016): http://yihui.name/knitr/ (for knitr), http://yihui.name/knitr/demo/rstudio/ (for using knitr with RStudio).

.

To estimate the models and generate the pdf files that summarize them, open one of the two files below and then execute the "Compile PDF" command in RStudio:

1. analysis-crime-five-period-01.Rnw

   This file generates Appendix 2 (results for crime incidents)

2. analysis-arrests-five-period-01.Rnw

   This file generates Appendix 3 (results for arrests)

Each of these .Rnw files calls the following files, the latter three while looping over the outcome for each set of models:

A. functions-analysis.R

   This file contains all functions used to parameterize and estimate models, as well as all functions used to plot the raw and modeled data.

B. child-fit-and-comparison-table.Rnw

   For the specified outcome, this file generates results for several least squares models. It then takes raw coefficients from these models, scales them by the distribution of the outcome in the pre-Ferguson period, and writes the raw and scaled coefficients to an empty row in a .csv file in the "out" folder of the project. The .csv file has a name that ends with "intervention-effects.csv" and where the prefix is based on the data file specified for the intervention.models function (line 20).

C. child-least-squares-pre-fit-plot.Rnw

   This file estimates four least squares models and then generates a table for the estimated coefficients using the package stargazer. It then plots two time series graphs for the second model, extrapolated to the post-Ferguson period, against the raw data (with and without smoothing, using three-week moving averages). It then generates autocorrelation plots for the observed outcome and residuals for the second model, during the pre-Ferguson period.

D. child-poisson-pre-fit-plot.Rnw

As a companion to the least squares models, this file estimates four poisson regression models, offering the R output for the analog to the second least squares model. It then offers two analogous plots to of the second model along with the raw data for the outcome.

Advice for Estimating Alternative Models

Users of this data distribution who wish estimate alternative models can do so in a variety of ways in R. One strategy is to write additional functions for our script, functions-analysis.R, or modify the specifications already in that script. With this method, the child .Rnw files can be edited (or replaced by new child .Rnw files) to summarize these new or additional models in alternative ways.

Another strategy would be to build a new R script that first loads model-ready-data.Rdata and then estimates models of interest, ignoring our workflow that uses .Rnw files. Our models can be estimated by extracting R code from the gray boxes (also known as "chunks") in the the various .Rnw files and inserting the R code in the user's new R script. For our code to work, the user's new R script will need to source our file, functions-analysis.R.