Reproducible Analytical Pipelines in Offender Management Statistics

Christopher Fairbanks, Prison Statistician, Ministry of Justice,

Government Statistical Service (GSS)

London, England

Abstract

The production of official statistics across sectors of government often presents a bespoke set of problems relating to data collection, methodology and evaluation. However, the largest drain on time, common to all official statistics, is often updating and styling publications in an attractive and engaging way. Through my three years on working with official statistics, I have become frustrated with the repetitive, tedious task of making official statistics 'pretty'.

Whilst it is important to present official statistics in a clean and interpretive manner, I find that "dotting the i's and crossing the t's" erodes limited time that could be spent on understanding and drawing out key messages from data. Put simply, **is it possible to produce official statistics automatically from data we can trust?**

A Reproducible Analytical Pipeline (RAP) is a system for producing reproducible 'styled and formatted' official statistics documents from 'raw' administration data. RAP borrows ideas from automated processes often found in software engineering, but can be achieved through tools such as R, which are commonly used within the statistical community. The aim of RAP is to replace monotonous processes; to improve accuracy and timeliness of the official statistics, and to provide a 'digital' third eye on the content of statistical publications