

A Visual Framework for Longitudinal and Panel Studies (with Examples in R)

Andrey Davydenko^{1*} and K. Charith²

¹Independent Researcher, Antalya, Turkey

²Department of Computer Science and Informatics, Uva Wellassa University, Badulla, Sri Lanka

Visual analysis is an essential part of modeling as it helps identify potential data issues and select appropriate methods for further analysis. We focus on simple yet effective visual tools applicable to panel and longitudinal data. Our objective was to find suitable tools for the following sequence of tasks: i) detect data anomalies/outliers, ii) describe patterns for missing and zero values, iii) identify data patterns and patches requiring special attention or modeling approaches, iv) assess the properties of distributions for the variables of interest, v) choose most suitable transformations, vi) assess the evidence for trends. This study demonstrates that existing software is not always suitable for the above tasks and there are areas for improvements and also found that many journal papers lack the visual analysis part for panel studies providing only summary tables, which does not give a clear picture of data features. Thus, the study proposes a framework aiming to solve the above tasks using a methodology based on a set of principles for effective statistical graphics. One well-known difficulty when plotting panels occurs due to the problem of overlapping: when a lot of points belonging to different panels are shown on one plot, the plot becomes hard to read. The framework proposed consists of a set of visual techniques that help solve this problem and increase the readability of plots. In particular, the study use colors and symbols with high visual discrimination and several plots to describe different data features keeping consistent colors/symbols for all plots. Th study describes several tools for the improved analysis, provide our guidelines and present R codes for the implementation of the tools. The study demonstrates the flexibility and the ease of use of R to plot, summarize, slide-and-dice panels, and to transform or impute variables. And the study used real data (relating to financial variables of Sri Lanka's companies) to show how the framework works.

Keywords: Exploratory data analysis, Visual tools, Panel data, Longitudinal data, R programming language