

Plotrr: Functions for making visual exploratory data analysis with nested data easier.

Charles Crabtree¹ and Michael J. Nelson²

¹University of Michigan

²Pennsylvania State University

12 February 2017

Paper DOI: <http://dx.doi.org/10.21105/joss.00190>

Software Repository: <https://github.com/cdcrabtree/plotrr>

Software Archive: <http://dx.doi.org/10.5281/zenodo.344900>

Summary

Social scientists can improve their research by conducting exploratory data analysis (Tukey 1977). One reason why they might not, however, is because it takes additional time. This is particularly true for researchers who use nested data. The issue here is that few existing functions help scholars *easily* visualize relationships within groups/units.

plotrr helps address this issue by providing several functions that make visual EDA easier to conduct. The focus of many of the package's functions is to create plots that can help researchers explore relationships within nested data. Among other things, these functions can help scholars assess the extent to which expected relationships between variables occur in specific cases. **bivarplots** creates a bivariate plot for every group/unit in the data, **dotplots** creates a dot plot for every group/unit, and **violinplots** creates a violin plot for every group/unit.

As demonstrated in Charles Crabtree and Nelson (2017), creating and interpreting plots like this this can help scholars find initial support for their theoretical expectations prior to conducting analysis with pooled data. The intuition here is that researchers can check their initial priors about relationships within cases. When the data support those priors, scholars have some additional evidence that the processes they theorize actually occur in the real world.

Another set of functions helps the researcher evaluate correlations between measures within groups/units. **groupcor** returns a tibble data frame with group/unit identifiers and the correlation coefficient between two measures for each group/unit in a data frame. **groupcorplot** performs a similar function and returns a figure that plots the correlation coefficient between measures within groups/units. These functions can help researchers identify important patterns across groups/units.

The visual EDA functions can also help scholars check the correlation between competing measures - such as different measures of state human rights practices - within units. This is one way to check the construct validity of the measures (Trochim and Donnelly 2008). This also might be useful to generate new empirical puzzles; if measures disagree in important cases, why is this the case?

In addition to these functions, the package also includes **histplots**, which creates histograms of a measure for each group/unit, and **bivarplot**, which returns a plot of the bivariate relationship between two measures alongside a rugplot of each measure.

Finally, the package also contains several “helper,” or convenience, functions.

References

- Charles Crabtree, and Michael J. Nelson. 2017. “New Evidence for a Positive Relationship Between de Facto Judicial Independence and State Respect for Empowerment Rights.” *International Studies Quarterly*.
- Trochim, William M. K., and James P. Donnelly. 2008. *Research Methods Knowledge Base*. Cengage Learning.
- Tukey, John W. 1977. *Exploratory Data Analysis*. Pearson.