

High throughput data extraction from primary literature: The metaDigitise R package

Joel L. Pick ?...? & Daniel W.A. Noble

October 19, 2017

Prepared for: Journal of Statistical Software?

Abstract

Keywords: meta-analysis, data extraction, R

Meta-analysis is becoming increasingly common in many fields. It relies foremost on data extracted from primary literature. This data is often presented in figures and so needs to be manually extracted. Although there are several existing tools to perform tasks like this, these tools are either not made for this specific purpose (i.e. meta-analysis) and so require a large amount of downstream data manipulation or do not allow easy import into statistical software (such as R). Furthermore these tools are generally not designed for a high throughput of images. Here, we present an interactive R package designed for large scale data extraction from figures, with meta-analysis in mind. To this end, we provide tools specific to data extraction from common plot types (mean and error plots, boxplots, scatterplots and histograms) and functions that allow for the summary statistics needed for downstream analysis to be calculated (i.e. mean sd and n and r). The package

1 Functionality

1.1 Figure Rotation

-for extraction from older, copied papers. also to flip sideways plots

1.2 Calibration

1.3 Bulk Digitisation

2 Plot Types

2.1 Mean and error plots

2.2 Boxplots

2.3 Scatterplots

2.4 Histograms

3 Summary Data

4 Reproducibility