Title	ReportMD: Writing complex scientific reports in R
Author	*Peter Humburg*
Affiliation	Garvan Institute of Medical Research
Contact	p.humburg@garvan.org.au
URL	https://github.com/humburg/reportmd
License	MIT

R is commonly used for complex analyses of large datasets. To improve the reproducibility of such analyses and provide the necessary documentation it is increasingly common to use literate programming techniques. Although R has had the ability to embed R code in long-form documents describing the analysis and presenting the results for many years via *Sweave*, it has been the introduction of *RMarkdown* that has driven the increasing popularity of this approach. The relative simplicity of authoring *Markdown* documents and the ability to convert these into a large variety of output formats via *pandoc* combined with the tight integration with *RStudio* make this approach easy to use and powerful at the same time.

While this is an essential technique for reproducible data analysis and much progress has been made over recent years to facilitate this literate programming approach, some challenges remain. One such issue is that complex analyses are best split into smaller units to retain some modularity but at the same time are likely to depend on the results of earlier stages of the analysis, i.e. modules may have complex interdependencies.

The solution to this problem provided by *knitr* is the use of child documents. These can be referenced in a main document and will then be evaluated separately and concatenated into one, potentially very large, document. While this has the advantage that multiple analysis modules can be combined into a larger report and all parts of the analysis are collated in a single document, ensuring that they are not inadvertently separated, it also has its downsides. In particular, it does not distinguish well between the data analysis and the presentation of the results obtained through this analysis. While both of these are intimately related and should not be entirely separated the natural flow of an analysis and the order imposed by internal dependencies is not always well suited to the effective presentation of results.

At the other extreme a commonly employed alternative largely separates the presentation layer from the analysis documents. Although this approach incorporates selected outputs from the analysis, such as figures and tables, into the presentation layer these connections are often fragile and a disconnect between the presentation and analysis layers can easily occur.

An approach that falls between these two extremes may be preferable. The aim is to produce a document that is designed to facilitate the presentation of results without being constraint by the structure of the analysis documents, forming a presentation layer on top of the analysis layer, while maintaining tight links between the two. To accomplish this such a document should

- have direct access to the R objects produced during the analysis;
- respond to changes in the analysis documents with corresponding changes in the presentation layer; and
- facilitate cross-references between individual documents.

The R package *ReportMD* presented here aims to provide the infrastructure required to achieve this. Build on top of *Rmarkdown* it is easy to use in existing work flows. *ReportMD* extends *knitr*'s chunk dependencies to extend across document boundaries and provides facilities to automatically load cached results from chunks in other documents. This makes it straightforward to include figures, tables and other summaries in the presentation layer while the required computations are documented in detail in an analysis document.