Introduction to the TAF package

1 Background

Arni Magnusson and Colin Millar

Overview

Objectives

The overarching goal of the Transparent Assessment Framework (TAF) is to support open and reproducible research. To achieve this goal, the following objectives have guided the design of TAF

Objectives

1. Provide a standard workflow structure that is general enough for any analysis that can be run from R.

Objectives

2. Introduce minimal constraints or learning curve, making it easy for a beginner
to create a new workflow or convert an existing workflow to TAF format.

Objectives

3. Enable reviewers to browse the data, model settings, and results, without being experts in R or the specific methods used.

Objectives

4. Enable anyone to rerun the analysis on another computer and get the same results.

Objectives

5. Require the scientist to briefly describe the data that are used in the analysis and where they came from.

Objectives

6. Invite the scientist to document with scripts how they processed the data before feeding them to the model.

Objectives

7. Invite the scientist to specify which versions of software are used, so the original analysis can be rerun at a later time.

Objectives

1. Provide a standard workflow structure that is general enough for any analysis that can be run from R.

2. Introduce minimal constraints or learning curve, making it easy for a beginner to create a new workflow or convert an existing workflow to TAF format.

3. Enable reviewers to browse the data, model settings, and results, without being experts in R or the specific methods used.

4. Enable anyone to rerun the analysis on another computer and get the same results.

5. Require the scientist to briefly describe the data that are used in the analysis and where they came from.

6. Invite the scientist to document with scripts how they processed the data before feeding them to the model.

7. Invite the scientist to specify which versions of software are used, so the original analysis can be rerun at a later time.
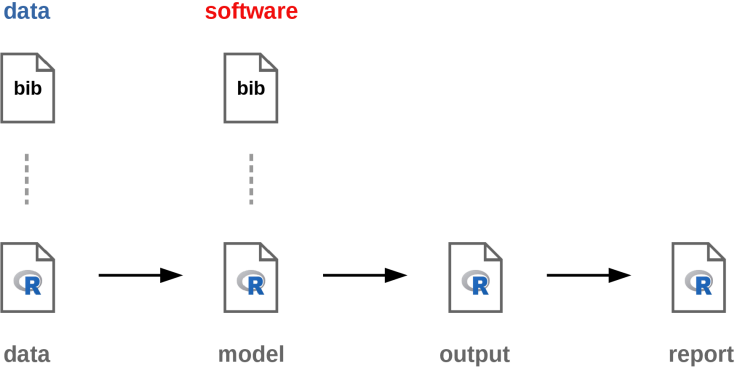
Design

TAF divides a workflow into four steps:

Script	Purpose
data.R	Prepare data, write CSV tables
model.R	Run model
output.R	Extract results, write CSV tables
report.R	Plots and tables for report

These scripts all share the same general structure,
starting with loading packages and reading in files,
then performing computations and writing out files.

They are run sequentially in alphabetical order, where
each script reads from files created in a previous step.

Design

Design

The initial data that are used in the analysis are declared in a file called **DATA.bib**, which is processed by the `taf.boot()` function.

During this boot procedure, each data entry is processed and the TAF system then makes the data available in the `boot/data` subfolder, where the `data.R` script will read it.

The **SOFTWARE.bib** file is optional. It is not used in the following linear regression example, but its functionality is covered in the slides on the boot procedure.

Summary