

Adding a corporate identity to reproducible research

Thierry Onkelinx, Ivy Jansen & Paul Quataert



Flanders
State of the Art

Reproducible research

- ▶ Markup languages separate content and style
- ▶ Straightforward to apply different styles to a document
- ▶ **knitr** facilitates the combination of R code with markup languages
 - ▶ \LaTeX
 - ▶ HTML
 - ▶ Markdown
- ▶ Benefit: R code and accompanying text in the same document
- ▶ **rmarkdown** facilitates the conversion of Markdown to different output formats
 - ▶ PDF
 - ▶ HTML
 - ▶ MS Word
 - ▶ ...

Our solution

Two R packages

- ▶ **INBOtheme**
 - ▶ Define custom themes for **ggplot2** graphics
 - ▶ Corporate identity theme
 - ▶ Themes as required by scientific journals
- ▶ **INBOMd**
 - ▶ Currently focus on PDF output
 - ▶ \LaTeX styles
 - ▶ Pandoc templates
 - ▶ **rmarkdown** functions

Requirements corporate identity

- ▶ Consistent
 - ▶ Markup language with appropriate output format gives consistent corporate identity
- ▶ Easy to use
 - ▶ Markdown is relatively easy to learn and use
 - ▶ \LaTeX is available for more fine-grained control
- ▶ Easy to maintain
 - ▶ Packages are maintained in a private git repository
 - ▶ Allows for different versions and branches
- ▶ Easy to distribute
 - ▶ `devtools::install_git` makes it easy to install from a git repository
 - ▶ Avoids the need to pack the code for different R versions on different architectures
 - ▶ Users are recommended to install packages at a default location
 - ▶ The manual assumes it is installed at that location
- ▶ Documented
 - ▶ R has nice infrastructure for the documentation of code in a package
 - ▶ Vignettes allow to write more elaborate manuals and examples
- ▶ Different types of output
 - ▶ Report
 - ▶ Presentation
 - ▶ Handouts
 - ▶ Poster
- ▶ Create matching graphics

INBOtheme

- ▶ **ggplot2** has a convenient system to define themes
- ▶ The default theme is easy to set and change
- ▶ The user can use the standard **ggplot2** code and apply the theme afterwards
- ▶ Easy to reuse code for different purposes (report, scientific journal)
- ▶ Each available theme is applied to all example graphics in a vignette
 - ▶ Provides examples for the user
 - ▶ Allows for visual quality control of the themes

INBOtheme example

```
library(ggplot2)
library(INBOtheme)
theme_set(theme_inbo2015(base_size = 25))
test <- data.frame(
  Letters = LETTERS[1:9],
  Y = runif(9, 0.25, 1)
)
ggplot(
  test,
  aes(x = Letters, y = Y, fill = Letters)
) +
  geom_bar(stat = "identity")
```

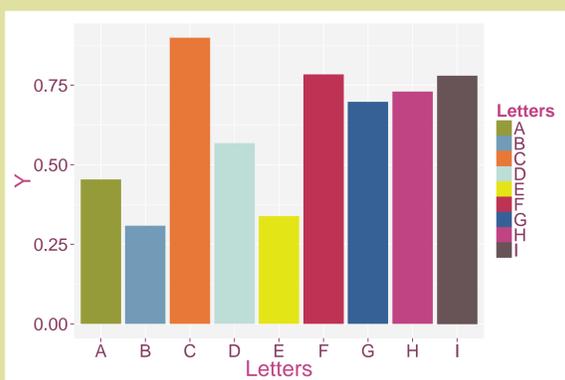


Figure: Example **ggplot2** graphic with corporate identity

INBOMd package structure

- ▶ R contains an **rmarkdown** function for each output style to map the variables of the YAML block to the correct pandoc template
- ▶ `inst/pandoc` contains the pandoc templates
- ▶ `inst/local_tex` contains a \TeX Directory Structure (TDS) with the \LaTeX styles
- ▶ `inst/inbo.cls` is a bibliography style
- ▶ `vignettes`
 - ▶ Manuals on using the corporate identity
 - ▶ Dummy documents illustrating the various components of the corporate identity

INBOMd \LaTeX usage

- ▶ Add `local_tex` from the package to the \LaTeX roots
- ▶ Refresh the FNDB (File Name Data Base)
 - ▶ Required each time files are added to `local_tex`
- ▶ Add the appropriate `\usepackage` to the preamble of the Sweave file

INBOMd Markdown usage

- ▶ Prepare the \LaTeX styles
- ▶ Add the relevant YAML block at the top of the Rmd file
 - ▶ Indicates the required render function
 - ▶ Defines optional variables to pass to the Pandoc template
- ▶ Use `rmarkdown::render` to render the Rmd file to the required output
 - ▶ A small script allows to render the same document to different output formats
 - ▶ E.g. presentation, handouts and a report version of the presentation (useful when teaching R code)
 - ▶ RStudio users can use the "Knit" button, which renders the first format in the YAML block

YAML block

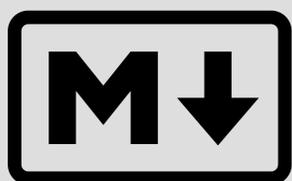
- ▶ Contains rendering metadata
- ▶ Is placed at the top of the Rmd file
- ▶ `variable: value` structure
- ▶ Hierarchy is indicated by indentation
- ▶ `output: lists` the defined output formats
- ▶ `package::function:` tells **rmarkdown** which function to use for rendering
- ▶ Other variables are passed to the function
 - ▶ Variables at the highest hierarchical level are passed to all output formats
 - ▶ Variables at the level below an output format are only passed to that output format

A YAML block example

```
title: "Markdown, R and corporate identity"
author: "Thierry Onkelinx"
bibliography: "INBOMd.bib"
keep_tex: true
output:
  INBOMd::inbo_slides:
    location: "Anderlecht, 2014/12/4"
    institute: "INBO"
  INBOMd::inbo_handouts:
    location: "Anderlecht, 2014/12/4"
    institute: "INBO"
  INBOMd::inbo_rapport:
    cover: "useR-large.png"
    subtitle: "An introduction"
    reportnr: "2014/12/4"
    year: 2014
    email: thierry.onkelinx@inbo.be
    cover_offset: 150mm
    cover_text: "The R logo"
```

Markdown to MS Word?

- ▶ Pandoc can do the conversion
- ▶ MS Word templates are limited to style definitions
- ▶ Variables are not available in MS Word templates
- ▶ Conclusions:
 - ▶ A lot of postprocessing would be required
 - ▶ Useful when collaborating with co-authors requiring MS Word for revisions



```
> sessionInfo()
[1] "June 30 - July 3, 2015"
[2] "Aalborg, Denmark"
```