Creating statistical reports in the past, present and future
On reproducible research and literate programming

Gergely Daróczi
daroczig@rapporter.net

Easystats Ltd, UK

27 March 2014
"The further back you can look, the further forward you are likely to see."
– Churchill

Source: Significance magazine
"This apparatus [Hollerith 1890 Census Tabulator] works unerringly as the mills of the gods, but beats them hollow as to speed."

– The Electrical Engineer, 11 Nov 1891.
Programming Environment

Assembly, Fortran, C(++), S(+), Java, R, Python, Julia, JavaScript

Creating statistical reports then and now

R, MATLAB, SPSS, STATA, Mathematica

Google Trends reference value

User Interface

Programming in console, Graphical User Interface, IDE, Software-as-a-Service

Distance Between Brownian Motions

Gergely Darócz (rapporter.net)
Reporting tools
A somewhat skewed cross table

<table>
<thead>
<tr>
<th></th>
<th>Past</th>
<th>Present</th>
<th>Future (?)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hardware</strong></td>
<td>Mainframes</td>
<td>PC, servers</td>
<td>Cloud, mobile</td>
</tr>
<tr>
<td><strong>Software</strong></td>
<td>FORTRAN</td>
<td>R, Python, D3</td>
<td>Julia, JS</td>
</tr>
<tr>
<td><strong>Users</strong></td>
<td>Statisticians</td>
<td>Data scientists</td>
<td>Everyone</td>
</tr>
<tr>
<td><strong>Stats/IT skills</strong></td>
<td>High</td>
<td>Medium to high</td>
<td>Low to high</td>
</tr>
<tr>
<td><strong>User interface</strong></td>
<td>console</td>
<td>IDE, GUI</td>
<td>browser</td>
</tr>
<tr>
<td><strong>Main content</strong></td>
<td>tables</td>
<td>interactive plots</td>
<td>text</td>
</tr>
<tr>
<td><strong>Reporting lang.</strong></td>
<td>LaTeX</td>
<td>HTML</td>
<td>markdown</td>
</tr>
<tr>
<td><strong>Main tool</strong></td>
<td>program</td>
<td>function</td>
<td>template</td>
</tr>
<tr>
<td><strong>Workflow</strong></td>
<td>writing</td>
<td>reproducible research</td>
<td>literate</td>
</tr>
<tr>
<td></td>
<td>reports</td>
<td></td>
<td>programming</td>
</tr>
</tbody>
</table>
What does “pander” do inside of “rapport”? A tool for literate programming that automatically transforms R objects into markdown

# A quick analysis on mtcars

```r
<% for (v in names(mtcars)) { %>
The mean of <%= v %> is <%= mean(mtcars[, v]) %>, and the standard deviation is <%= sd(mtcars[, v]) %>. Let us also check the frequency table:

<%= table(mtcars[, v]) %>

```r
```r
## Tables are boring!

<%=
set.caption(paste("Histogram of", v))
hist(mtcars[, v], xlab = v, col = sample(colors(), 1), main = "")
%
```r
```r
<% } %>
```
# A quick analysis on mtcars

The mean of am is _0.4062_ and the standard deviation is _0.499_. Let us also check the frequency table:

```
-----
   0  1
--- ---
19  13
-----
```

## Tables are boring!

![Histogram of am](/tmp/RtmpL0K2Q/plots/f2457fb575.png)

...
What does “pander” do inside of “rapport”? Pandoc.brew: calling Pandoc to convert the results to HTML

A quick analysis on mtcars

The mean of am is 0.4062 and the standard deviation is 0.499. Let us also check the frequency table:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

Tables are boring!
What does “pander” do inside of “rapport”? Pandoc.brew: calling Pandoc to convert the results to MS docx

A quick analysis on mtcars

The mean of \(am\) is 0.4062 and the standard deviation is 0.499. Let us also check the frequency table:

\[
\begin{array}{cc}
0 & 1 \\
19 & 13 \\
\end{array}
\]

Tables are boring!

Histogram of \(am\)
“rapport” overview
The overall structure of a reusable report template

<!--head
meta:
  title: ...
  author: ... 
  description: ...
  packages: ...
inputs:
  - name: ...
    class: ...
head-->

<% for (...) { %>

## Subtitle with <%= inline code chunk %>

<%= table(...) %>

<% } %>

http://rapport-package.info/
Gergely Daróczi (rapporter.net)
Creating statistical reports then and now 27/3/2014 11 / 17
“rapport” demo

Header: meta information on the reporting template and input(s) specification

<!--head
meta:
  title: Rapport demo
  author: daroczig
  description: This is POC demo on the usage of rapport templates
  packages:
    - ggplot2
    - pander

inputs:
 - name: v
   label: Variable to analyse
   required: yes
   class: numeric
   length:
     min: 1.0
     max: 1.0
 - name: color
   label: Color of the histogram
   standalone: yes
   value: red
   class: character
head-->
# A quick analysis on `<%= v.name %>`

The mean of `<%= v.name %>` is `<%= mean(v) %>` and the standard deviation is `<%= sd(v) %>`. Let us also check the frequency table:

`<%= table(v) %>`

## Tables are boring!

`<%=
set.caption(paste('Histogram of', v.name))
hist(v, xlab = v, col = color, main = '')
%>`
```r
> rapport('rapport-demo', data = mtcars, v = 'am')
```

# A quick analysis on `am`

The mean of am is _0.4062_ and the standard deviation is _0.499_. Let us also check the frequency table:

```
-------
 0  1
--- ---
19 13
-------
```

## Tables are boring!

![Histogram of am](/tmp/Rtmpf30K2Q/plots/f2457fb575.png)
Online demo
Analysing the results of a dialect survey with Google Maps and ordinary English language

http://blog.rapporter.net/2013/07/uk-dialect-maps.html
Summary

The most popular category in the United Kingdom was <<pop>> for <<Pop or soda?>> chosen by four tenth of the respondents.

And the most important differences between the countries can be summarised as:

- it seems, that two tenth of British people disagree with <<other>> that is low comparing to e.g. Scottish people
- eventually, less then one tenth of British people tends to dislike the answer <<soft drink>> that is low compared to let's say Northern Irish people
- it seems, that one half of people living in Northern Ireland tends to like the answer <<soft drink>> that is high comparing to e.g. Welsh citizens
- it seems, that two tenth of Scottish people tends to dislike the answer <<pop>> that is low compared to the average
- it seems, that five tenth of Scottish people love the answer <<other>> that is high compared to let's say British people

http://blog.rapporter.net/2013/07/uk-dialect-maps.html
Rapporter packages

References: integrating the reporting framework in web applications

- **pander**: A Pandoc’s markdown writer in R
- **rapport**: A report templating system with dynamic inputs
- **rapportools**: Helpers functions
- **sandboxR**: Filtering ”malicious” R calls

Further documentation:

- [http://hackme.rapporter.net](http://hackme.rapporter.net)

Q & A: daroczig@rapporter.net