A simple workflow for using R with Microsoft products

Marcus W. Beck

USEPA NHEERL Gulf Ecology Division, Gulf Breeze, FL
Email: beck.marcus@epa.gov, Phone: 850 934 2480

May 21, 2014
The problem...

- R is great and has an increasing user base
- RStudio is integrated with multiple document preparation systems
- Output documents are not in a format that facilitates collaboration with non R users, e.g., pdf, html
- Data coming to you may be in a proprietary format, e.g., xls spreadsheet
The solution?

- Solution one - Make liberal use of ‘projects’ within RStudio
- Solution two - Use gdata package to import excel data
- Solution three - Get pandoc to convert document formats - http://johnmacfarlane.net/pandoc/
The solution?

- Solution one - Make liberal use of ‘projects’ within RStudio
- Solution two - Use gdata package to import excel data
- Solution three - Get pandoc to convert document formats - [http://johnmacfarlane.net/pandoc/](http://johnmacfarlane.net/pandoc/)

*Not recommended for simple tasks unless you really, really love R*
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users

- Idea is to never leave the RStudio environment - dynamic documents!

- General workflow…
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users

- Idea is to never leave the RStudio environment - dynamic documents!

- General workflow…

1. Install necessary software and packages
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users.

- Idea is to never leave the RStudio environment - dynamic documents!

- General workflow...

1. Install necessary software and packages
2. Create project in RStudio
3. Setup supporting docs/functions
4. Import with gdata, summarize
5. Create HTML document using knitr Markdown
6. Convert HTML doc to Word with Pandoc
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users

- Idea is to never leave the RStudio environment - dynamic documents!

- General workflow...

1. Install necessary software and packages
2. Create project in RStudio
3. Setup supporting docs/functions
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users
- Idea is to never leave the RStudio environment - dynamic documents!
- General workflow...

1. Install necessary software and packages
2. Create project in RStudio
3. Setup supporting docs/functions
4. Import with gdata, summarize
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users

- Idea is to never leave the RStudio environment - dynamic documents!

- General workflow...

1. Install necessary software and packages
2. Create project in RStudio
3. Setup supporting docs/functions
4. Import with gdata, summarize
5. Create HTML document using knitr Markdown
6. Convert HTML doc to Word with Pandoc
An example workflow

- I will present a workflow for integrating Microsoft products within RStudio as an approach to working with non R users
- Idea is to never leave the RStudio environment - dynamic documents!
- General workflow...

1. Install necessary software and packages
2. Create project in RStudio
3. Setup supporting docs/functions
4. Import with gdata, summarize
5. Create HTML document using knitr Markdown
6. Convert HTML doc to Word with Pandoc
The example
You are sent an Excel file of data to summarize and report but you love R and want to do everything in RStudio...

<table>
<thead>
<tr>
<th>SiteName</th>
<th>Year</th>
<th>Restoration</th>
<th>Reference</th>
<th>Observer.Names</th>
<th>Precipitation</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGH</td>
<td>2005</td>
<td>3</td>
<td>3</td>
<td>Tyler_Amanda</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Kelly</td>
<td>2005</td>
<td>4</td>
<td>2</td>
<td>Patrick_Chelsea</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Carlton</td>
<td>2005</td>
<td>2</td>
<td>3</td>
<td>David_Megan</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>IGH</td>
<td>2006</td>
<td>9</td>
<td>6</td>
<td>Tyler_Amanda</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>Kelly</td>
<td>2006</td>
<td>9</td>
<td>1</td>
<td>David_Megan</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>Carlton</td>
<td>2006</td>
<td>7</td>
<td>3</td>
<td>Patrick_Chelsea</td>
<td>0</td>
<td>52</td>
</tr>
<tr>
<td>IGH</td>
<td>2007</td>
<td>12</td>
<td>7</td>
<td>David_Megan</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>Kelly</td>
<td>2007</td>
<td>2</td>
<td>18</td>
<td>Jeremy_Lucy</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>Carlton</td>
<td>2007</td>
<td>11</td>
<td>2</td>
<td>Patrick_Chelsea</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>IGH</td>
<td>2008</td>
<td>9</td>
<td>4</td>
<td>Tyler_Amanda</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>Kelly</td>
<td>2008</td>
<td>14</td>
<td>5</td>
<td>David_Megan</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>Carlton</td>
<td>2008</td>
<td>13</td>
<td>3</td>
<td>Patrick_Chelsea</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>IGH</td>
<td>2009</td>
<td>18</td>
<td>7</td>
<td>Patrick_Chelsea</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Kelly</td>
<td>2009</td>
<td>16</td>
<td>5</td>
<td>David_Megan</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Carlton</td>
<td>2009</td>
<td>20</td>
<td>1</td>
<td>Tyler_Amanda</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>IGH</td>
<td>2010</td>
<td>12</td>
<td>2</td>
<td>David_Megan</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>Kelly</td>
<td>2010</td>
<td>15</td>
<td>3</td>
<td>Patrick_Chelsea</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>Carlton</td>
<td>2010</td>
<td>24</td>
<td>4</td>
<td>Tyler_Amanda</td>
<td>0</td>
<td>61</td>
</tr>
</tbody>
</table>
Step 1

Install necessary software and Packages
Step 1

Install necessary software and Packages

- R and RStudio (can do with other R editors)
- Microsoft Office
Step 1

Install necessary software and Packages

- R and RStudio (can do with other R editors)
- Microsoft Office
- Strawberry Perl for using gdata package
- Pandoc
Step 1

Install necessary software and Packages

- R and RStudio (can do with other R editors)
- Microsoft Office
- Strawberry Perl for using gdata package
- Pandoc
- Packages: gdata, knitr, utils, xtable, others as needed...
Step 2
Create a project in RStudio

- Create a folder or use existing on local machine
- Add .Rprofile file to the folder for custom startup
- Move all data you are working with to the folder
- Literally create project in RStudio
- Set options within RStudio
Step 3
Setup supporting docs/functions, i.e., .Rprofile, functions, report, master

.Rprofile

# library path
.libPaths("C:\\Users\\mbeck\\R\\library")

# startup message
cat("My project...

# packages to use
library(utils)  # for system commands
library(knitr)  # for markdown
library(gdata)  # for import xls
library(reshape2)  # data format conversion
library(xtable)  # easy tables
library(ggplot2)  # plotting

# perl path for gdata
prl_pth <- "C:/strawberry/perl/bin/perl.exe"

# functions to use
source("my_funcs.r")
Step 3
Setup supporting docs/functions, i.e., .Rprofile, functions, report, master

my_funcs.r

```r
### functions for creating report, created May 2014, M. Beck

### processes data for creating output in report, 'dat_in' is input data as
### data frame, output is data frame with converted variables
proc_fun <- function(dat_in) {

    # convert temp to C
    dat_in$Temperature <- round((dat_in$Temperature - 32) * 5/9)

    # convert data to long format
    dat_in <- melt(dat_in, measure.vars = c("Restoration", "Reference"))

    return(dat_in)
}

### creates linear model for data, 'proc_dat' is processed data returned from
### 'proc_fun', output is linear model object
mod_fun <- function(proc_in) lm(value ~ variable + Year, dat = proc_in)
```

M. Beck (USEPA NHEERL)  R with Microsoft  May 21, 2014
Step 3
Setup supporting docs/functions, i.e., .Rprofile, functions, report, master

---

Here's a report I made for `r gsub('/|\.xlsx','',name)`

---

```r
```{r echo=F, include=F}

# import data
url <- paste0('http://beckmw.files.wordpress.com/2014/05', name)
dat <- read.xls(xls = url, sheet = 'Sheet1', perl = prl_pth)

# process data for tables/figs
dat <- proc_fun(dat)

# model of data
mod <- mod_fun(dat)
```{r reg_fig, echo=F, fig.width = 5, fig.height = 3, dpi=200}

ggplot(dat, aes(x = Year, y = value, colour = variable)) +
  geom_point() +
  stat_smooth(method = 'lm')
```{r results='asis', echo=F}

### Model summary
```{r echo=F}
print.xtable(xtable(mod, digits = 2), type = 'html')
```

### Figure of restoration and reference by year
```{r echo=F}
ggplot(dat, aes(x = Year, y = value, colour = variable)) +
  geom_point() +
  stat_smooth(method = 'lm')
```
Step 3

Setup supporting docs/functions, i.e., .Rprofile, functions, report, master

---

### master.r

```r
# file to process
name <- "/my_data.xlsx"

# rmd to html
knit2html("report.Rmd")

# pandoc conversion of html to word doc
system(paste0("pandoc -o report.docx report.html"))
```

---
Steps 4 - 6

After creating supporting documents in Project directory, final steps are completed by running ‘master.r’

- Step 4 - xlsx file imported using gdata package, implemented in ‘report.Rmd’
- Step 5 - HTML document created by converting ‘report.Rmd’ with knit2html in ‘master.r’
- Step 6 - HTML document converted to Word with Pandoc by invoking system command

```r
# file to process
name <- "~/my_data.xlsx"

# rmd to html
knit2html("report.Rmd")

# pandoc conversion of html to word doc
system(paste0("pandoc -o report.docx report.html"))
```