Sharing Your Work

There are various ways to share your work. In this session we talk about several approaches. The first is simply by sharing your .RData file with others. Remember that when you quit an R session it asks you if you wish to save it. If you answer "yes" then it creates a file called .RData in that folder/directory. If you already have one in your current directory then do:

```
system("ls .Rdata",intern=T)
[1] ".Rdata"
```

You can also save specific environment items to a given .RData file:

```
ls()
[1] "a"    "hold"
save(hold,file="xfer.RData")
```

You can now share xfer.RData with someone else. They can load it like:

```
load("xfer.RData")
ls()
[1] "hold"
```

Sharing Your Work

This works over the Internet also:

```
load(url("http://www.bimcore.emory.edu/BIOS560R/SUPP.DIR/xy.Rdata"))

ls()
[1] "hold"   "mtcars" "x"      "y"
```

Sharing Your Work - .RData

The advantages of this approach is that you can rapidly share things with others. The disadvantage is that if your environment file is too big then you can't email it. You have to share it across DropBox which , of course, isn't really a problem - you just have to have an account.

If you just want to share R code you can always paste in the contents of a function into an email message. The problem there is that you then have to supply some test data unless it is implied or you provide instructions on how to generate it.

Reproducible Research

Data deposition: Missing data mean holes in tree of life

Bryan T. Drew

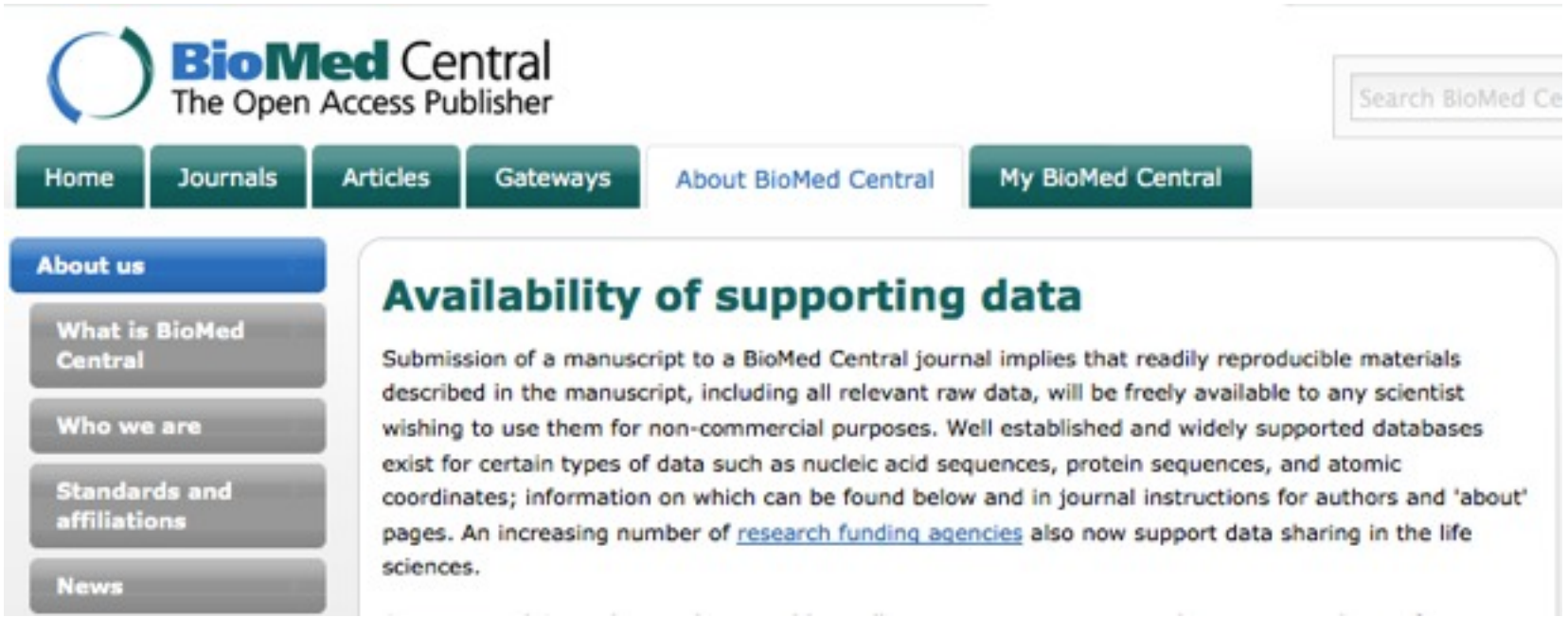
Nature **493**, 305 (17 January 2013) | doi:10.1038/493305f
Published online 16 January 2013

Subject terms: Databases · Genomics · Evolution

As part of the Open Tree of Life project (http://opentreeoflife.org), we surveyed publications covering all domains of life and found that most phylogenetic trees and nucleotide alignments from the past two decades have been irrevocably lost.

Of 6,193 papers we surveyed in more than 100 peer-reviewed journals, only 17% present accessible trees and alignments (used to infer relatedness). Contacting lead authors to procure data sets was only 19% successful. DNA sequences were deposited in GenBank for almost all these studies, but it is the actual character alignments that are pivotal for reproducing phylogenetic analyses. We estimate that more than 64% of existing alignments or trees are permanently lost.

Reproducible Research



Reproducible Research

Nature [edit]

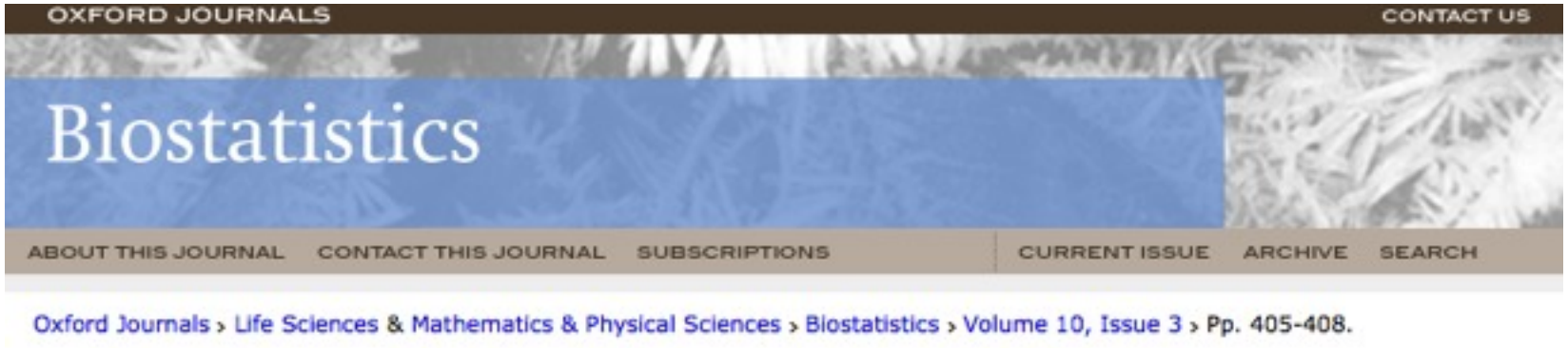
"Such material must be hosted on an accredited independent site (URL and accession numbers to be provided by the author), or sent to the *Nature* journal at submission, either uploaded via the journal's online submission service, or if the files are too large or in an unsuitable format for this purpose, on CD/DVD (five copies). Such material cannot solely be hosted on an author's personal or institutional web site.[7]

Nature requires the reviewer to determine if all of the supplementary data and methods have been archived. The policy advises reviewers to consider several questions, including: "Should the authors be asked to provide supplementary methods or data to accompany the paper online? (Such data might include source code for modelling studies, detailed experimental protocols or mathematical derivations.)"[8]

Science [edit]

Science supports the efforts of databases that aggregate published data for the use of the scientific community. Therefore, before publication, large data sets (including microarray data, protein or DNA sequences, and atomic coordinates or electron microscopy maps for macromolecular structures) must be deposited in an approved database and an accession number provided for inclusion in the published paper.[9]

Reproducible Research



Reproducible research and *Biostatistics*

1. INTRODUCTION AND MOTIVATION

The replication of scientific findings using independent investigators, methods, data, equipment, and protocols has long been, and will continue to be, the standard by which scientific claims are evaluated. However, in many fields of study there are examples of scientific investigations that cannot be fully replicated because of a lack of time or resources. In such a situation, there is a need for a minimum standard that can fill the void between full replication and nothing. One candidate for this minimum standard is "reproducible research", which requires that data sets and computer code be made available to others for verifying published results and conducting alternative analyses.

Reproducible Research

2.1 Dimensions of reproducibility

The Associate Editor for reproducibility (AER) will handle submissions of reproducible articles. Currently, the AER's involvement with a submission begins only when an article has been accepted for publication. The AER will consider three different criteria when evaluating the reproducibility of an article.

1. Data: The analytic data from which the principal results were derived are made available on the journal's Web site. The authors are responsible for ensuring that necessary permissions are obtained before the data are distributed.
2. Code: Any computer code, software, or other computer instructions that were used to compute published results are provided. For software that is widely available from central repositories (e.g. CRAN, Statlib), a reference to where they can be obtained will suffice.
3. Reproducible: An article is designated as reproducible if the AER succeeds in executing the code on the data provided and produces results matching those that the authors claim are reproducible. In reproducing these results, reasonable bounds for numerical tolerance will be considered.

Authors can choose to meet a subset of these criteria if they wish. For example, an author may choose to release code showing how a

Reproducible Research

* Research Should be reproducible. Anything in a scientific publication should be reproducible by the reader

* This means that the "entire analysis" of any data in the paper should be reproducible

* Many authors make "supplemental materials" available via web pages many times "forever" which makes it an ideal "place" to provide additional materials.

Reproducible Research

* Research Should be reproducible. Anything in a scientific publication should be reproducible by the reader

* This means that the "entire analysis" of any data in the paper should be reproducible

* Many authors make "supplemental materials" available via web pages many times "forever" which makes it an ideal "place" to provide additional materials.

* With R one reads in data from some place, runs some functions, and then presents visualizations for consideration. While it is easy to put up the data source and R code what about the "narrative" that includes the analysis ?

* Is there a way to include data, code, and narrative in one document ?

Literate Programming

The idea is that you can provide a description of what your analysis and code is doing with the added benefit of being able to embed actual code (in the form of macros) within the description.

Then you can "compile" you document and it will run the code and insert the results into your document alongside with the narrative in a, (hopefully), seamless way.

```
Sweave Document Layout

Section 1

This is section one. Blah, blah, etc.

<<>>
my.lm = lm(mpg~wt, data=mtcars)
@

Section 2
Here we do some regression

<<>>
plot(my.lm)
@

Section 3
```

Literate Programming

Sweave Example 1

Steve

November 19, 2013

1 Section 1

This is section one. Blah, blah, etc.

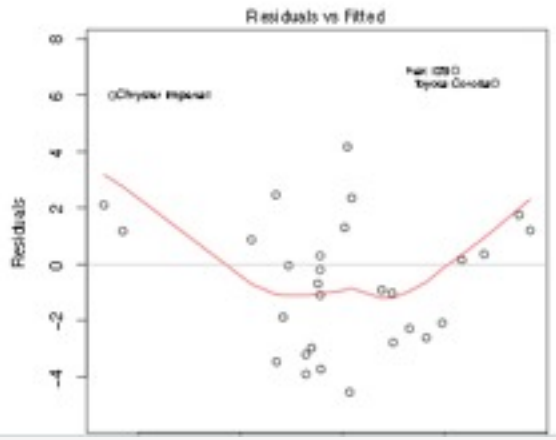
```
> my.lm = lm(mpg~wt,data=mtcars)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.2851	1.8776	19.86	0.0000
wt	-5.3445	0.5591	-9.56	0.0000

2 Section 2

Here we do some regression

```
> plot(my.lm)
```



Sweave - Reproducible Research

Yes - Sweave to the rescue. http://www.stat.uni-muenchen.de/~leisch/Sweave/

What Is It ?

"Sweave is a tool that allows to embed the R code for complete data analyses in LaTeX documents. The purpose is to create dynamic reports, which can be updated automatically if data or analysis change.

Instead of inserting a prefabricated graph or table into the report, the master document contains the R code necessary to obtain it.

When run through R, all data analysis output (tables, graphs, etc.) is created on the fly and inserted into a final latex document.

The report can be automatically updated if data or analysis change, which allows for truly reproducible research"

Sweave - Reproducible Research

Yes - Sweave to the rescue. http://www.stat.uni-muenchen.de/~leisch/Sweave/

What Is It ? A More Detailed Answer

Sweave is a utility function that lets one insert/embed R code into LaTex documents.

Your Sweave document will contain the text of your narrative / discussion in LaTex markup and will have the "chunks" of R code that will be replaced by results when the document is "compiled".

You can also keep the R statements that generated the results.

You can always go back to the document and update it if and when your data and subsequent analysis changes.

Sweave generates LaTex code - Think of Sweave as an add-on to LaTex

Sweave - Reproducible Research

Yes - Sweave to the rescue. http://www.stat.uni-muenchen.de/~leisch/Sweave/

Where Can I Get It ?

The Sweave software itself is part of every R installation, see

```
help("Sweave", package="utils")
```

to get started. The home page for Sweave features additional material that does not ship with standard R, like papers and additional examples.

Also see

```
vignette("Sweave")
```

Since Sweave is contingent upon LaTex then we need to consider it also. At this point we need to take a detour. Don't worry, we will come back to Sweave momentarily.

Sweave - Reproducible Research

Sweave User Manual

Friedrich Leisch and R-core

May 16, 2013

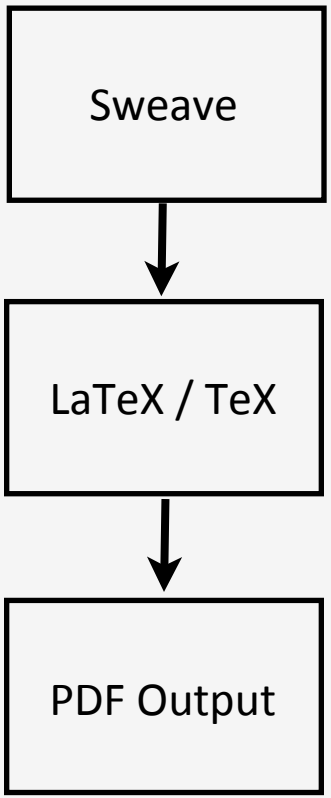
1 Introduction

Sweave provides a flexible framework for mixing text and R code for automatic document generation. A single source file contains both documentation text and R code, which are then *woven* into a final document containing

- the documentation text together with
- the R code and/or
- the output of the code (text, graphs)

This allows the re-generation of a report if the input data change and documents the code to reproduce the analysis in the same file that contains the report. The R code of the complete analysis is embedded into a LaTeX document[1] using the `noweb` syntax (Ramsey, 1998) which is usually used for literate programming Knuth (1984). Hence, the full power of LaTeX (for high-quality typesetting) and R (for data analysis) can be used simultaneously. See Leisch (2002) and references therein for more general thoughts on dynamic report generation and pointers to other systems.

Detour - LaTex



Sweave documents are a mixture of LaTeX statements and R code "weaved" together

The LaTeX and TeX systems process the Sweave file including the R code

The output is a PDF file although users can freely share the original Sweave document

Detour - LaTex

LaTex - http://www.latex-project.org/

So What Is LaTex ?

LaTeX is a high-quality typesetting system; it includes features designed for the production of technical and scientific documentation. LaTeX is the *de facto* standard for the communication and publication of scientific documents. LaTeX is available as free software for all major operating systems.

Detour - LaTex - For Apple

LaTex - http://www.latex-project.org/

How Do I Get It For My Computer ?

For Apple Computers check out http://www.tug.org/mactex/ and download the MacTex.pkg - http://mirror.ctan.org/systems/mac/mactex/MacTeX.pkg

The current distribution is MacTeX-2013 - This distribution requires Mac OS 10.5 Leopard or higher; see other links on page for Mac OS 10.3 or 10.4.

Also download the MacTexExtras.pkg which will provide you want a LaTex Editor such as TexMaker http://mirror.ctan.org/systems/mac/mactex/MacTeXtras.zip

Note that RStudio can also serve as a LaTex editor.

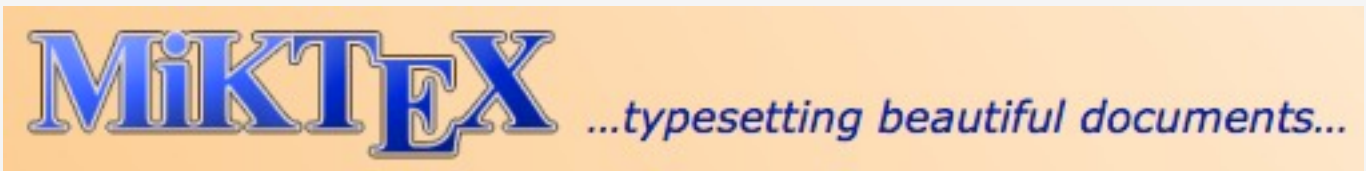
Detour - LaTex - For MS-Windows

LaTex - http://www.latex-project.org/

How Do I Get It For My Computer ?

For Windows Computers check out MicTex package @ http://miktex.org/
The download can be found at http://mirrors.ctan.org/systems/win32/miktex/setup/basic-miktex-2.9.4813.exe

This distribution supports most modern versions of MS Windows (Vista, 7, and even XP with service pack 3). Older versions of Windows (Win 95, NT, ME, 2000) are not supported.Follow the instructions in the Installer to get the package setup.



you can also download a LaTex editor called WinEdit. It is free for 31 days
www.winedt.com/

Detour - LaTex

LaTeX is based on TeX which whose goals are

1) to allow anybody to produce high-quality books using a reasonably minimal amount of effort, and

2) to provide a system that would give exactly the same results on all computers, now and in the future.[

TeX was written by Donald Knuth in 1978 !

LaTex comes with many templates depending upon your goal. Article, Slides, Brochures, Abstracts, Legal, and many more. It also has superior mathematical and scientific symbol support.

Detour - LaTex

http://www.youtube.com/watch?v=75Ju0eM5T2c

Detour - LaTex

You can create a document using RStudio's editor. Create a LaTeX document called coolpkg.Rnew with the following contents:

```
\documentclass{article}
\usepackage{amsmath}
\title{Nobel Prize Winning Article}
\author{Ziggy Stardust}
\begin{document}
\SweaveOpts{concordance=TRUE}
\maketitle
\section{Introduction}
Simplictity is everything. This is why I like being simple. It's too difficult being
difficult.
\section{Results}
Some data analysis results. If you don't understand this then there is no hope for
you.
 \begin{align*}
    \sin A \cos B &= \frac{1}{2}\left[ \sin(A-B)+\sin(A+B) \right] \\
 \end{align*}

\begin{equation*}
\int_{\partial \Omega} \omega = \int_{\Omega} d\omega.
\end{equation*}

\section{Conclusion}
Conclusion texts are here.
\end{document}
```

Detour - LaTex

Nobel Prize Winning Article

Ziggy Stardust

November 16, 2013

1 Introduction

Simplicity is everything. That is why I like being simple. It's too difficult to be difficult.

2 Results

Some data analysis results. If you don't understand this then there is just no hope for you:

$$\sin A \cos B = \frac{1}{2}\left[\sin(A-B) + \sin(A+B)\right]$$

$$\int_{\partial\Omega} \omega = \int_{\Omega} d\omega.$$

3 Conclusion

Conclusion texts are here.

Detour - LaTex - Advantages

* Free

* More professional looking. Preferred for Scientific Documents.

* Finer grained control of formatting.

* Faster because it's a simple text file.

* Most importantly, can support any type of mathematical symbols and equations.

* Note that you can now type LaTex commands within a Word document to generate equations.

* Great Bibliographic Tools

Detour - LaTex - Disadvantages

* Can be tedious.

* No "What You See Is What You Get" editing.

* All formatting and typesetting needs to be accomplished manually. This can be quite tedious at times.

* Lack of useful features such as "track changes" and collaborative tools.

Detour - LaTex - Learning

Some resources for learning LaTeX

Practical LaTeX for Public Health and Medicine

http://www1.maths.leeds.ac.uk/latex/epitex.pdf

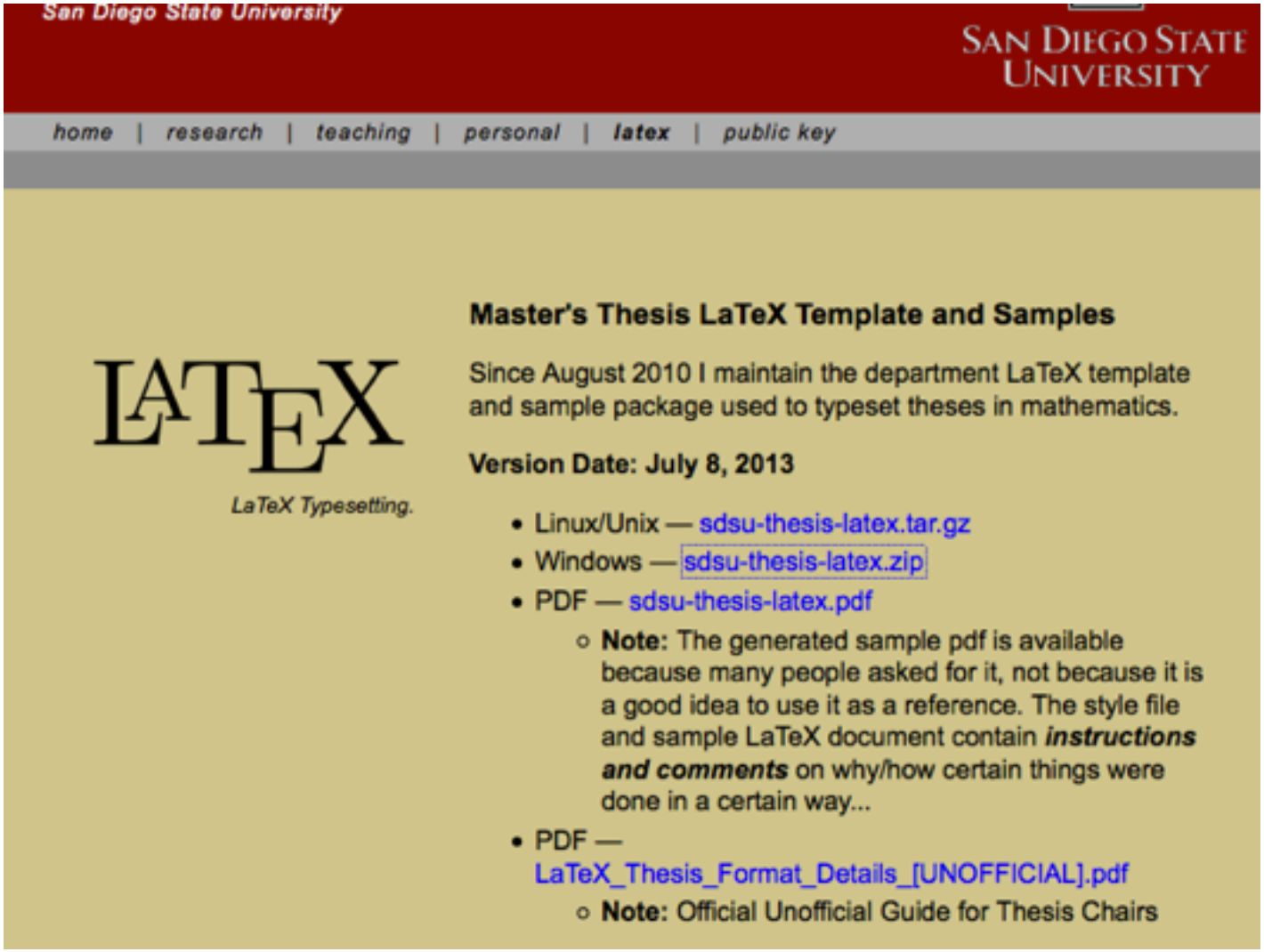
The Not so Short Introduction to LaTeX

http://www1.maths.leeds.ac.uk/school/students/docs/latex-short.pdf

LaTeX WikiBooks

http://en.wikibooks.org/wiki/LaTeX

Detour - LaTex - Disadvantages



Detour - LaTex

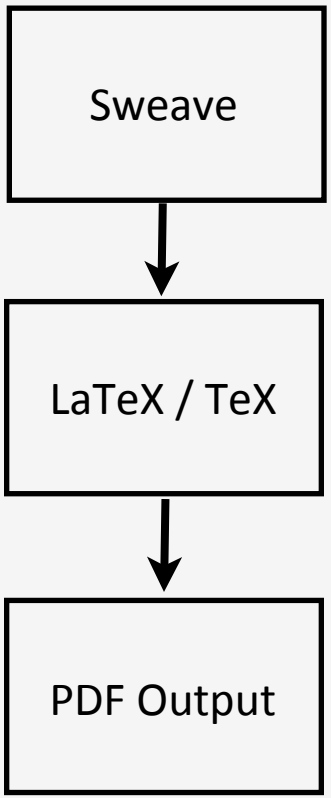
In addition to LaTex you might want to download dedicated editors if you plan on doing serious work with LaTex. For example TexMaker is a good one for OSX and WinEdit is a good one for MS Windows.

However, if all you want to do is generate nice documentation then you might be able to get by just fine by using RStudio as both your LaTex and Sweave editor.

If you plan on doing lots of LaTex editing and you are working on a big document then you might lean towards the dedicated editor approach since they tend to offer more "bells and whistles".

In this session I will describe two ways to generate nice documents. One using the traditional method from within R. The second will use RStudio.

Detour - LaTex



Sweave documents are a mixture of LaTeX statements and R code "weaved" together

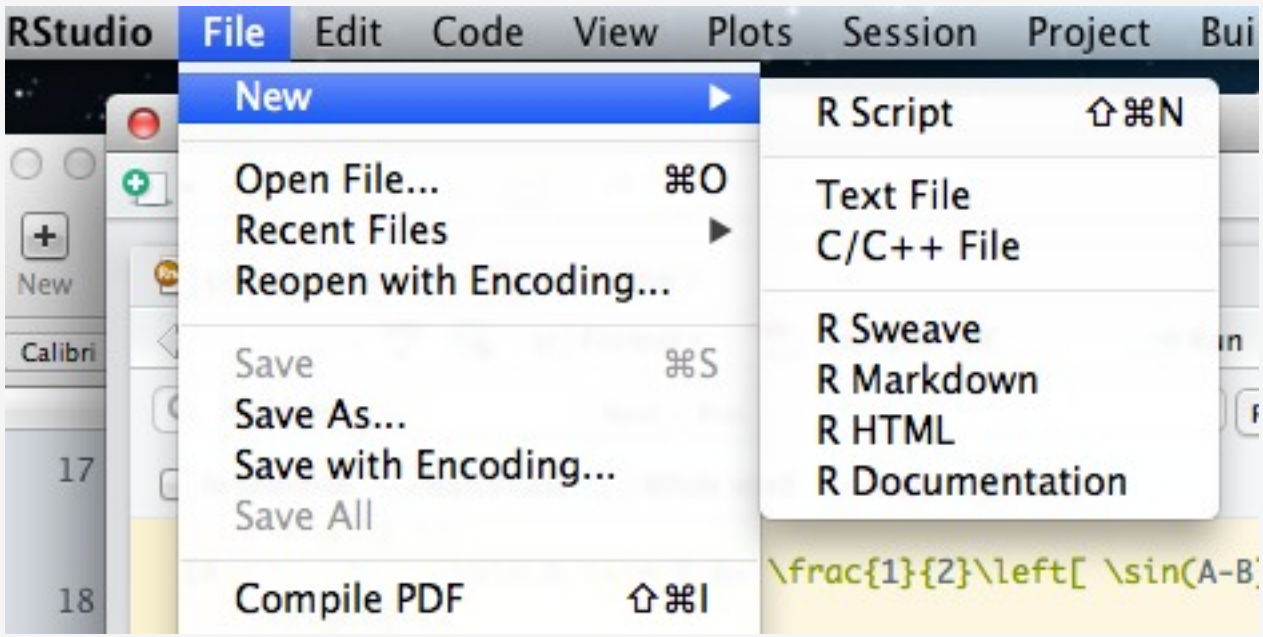
The LaTeX and TeX systems process the Sweave file including the R code

The output is a PDF file although users can freely share the original Sweave document

Detour - LaTex

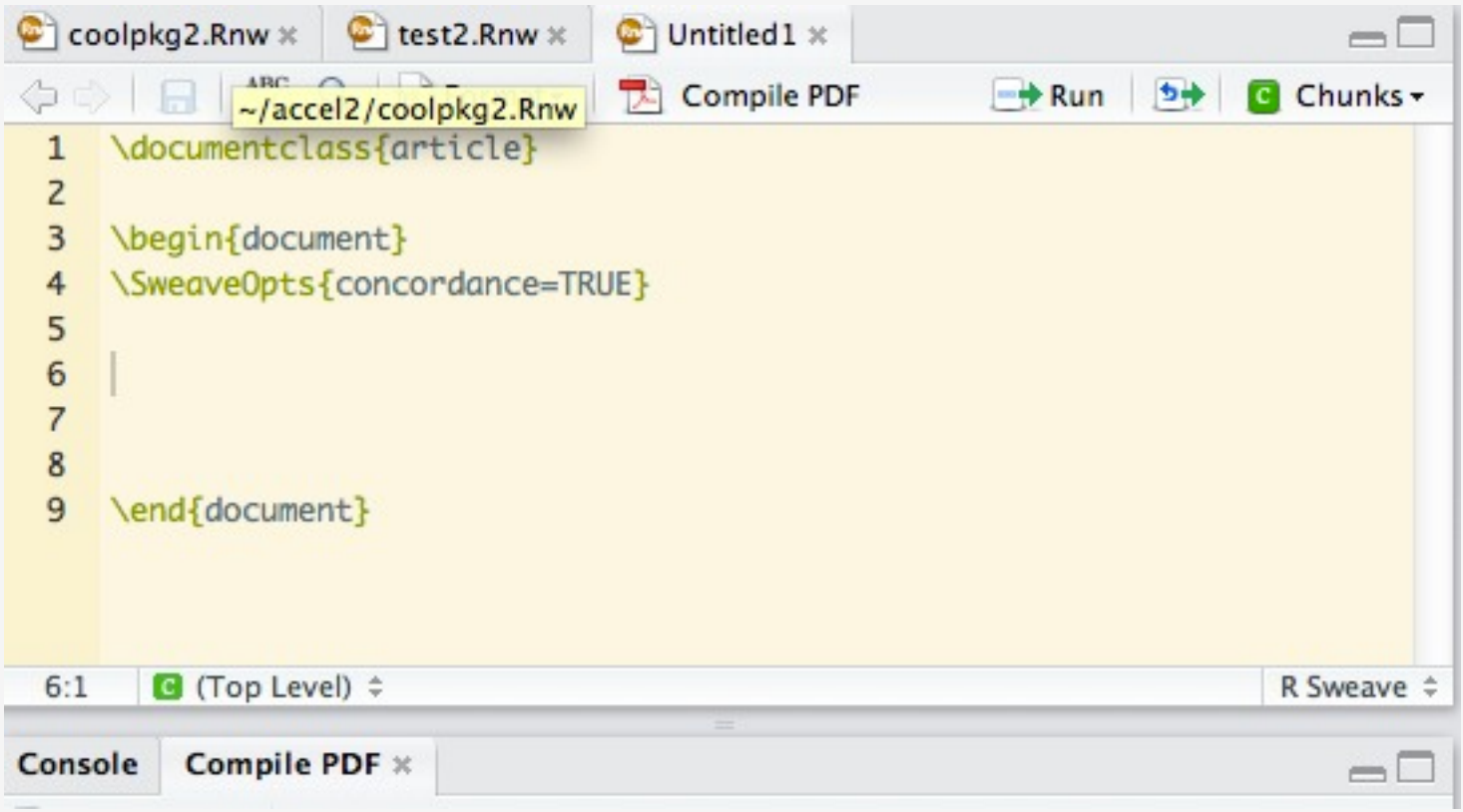
You can use one of the editors to prepare a LaTeX document. Or you can use RStudio to manage a LaTeX document for you. This is probably the simplest thing to do. As an example let's do:

From the File menu select File > New > R Sweave



Detour - LaTex

From the File menu select File > New > R Sweave. In the file pane you get an Untitled document that looks like. Save the file to something like "example1".

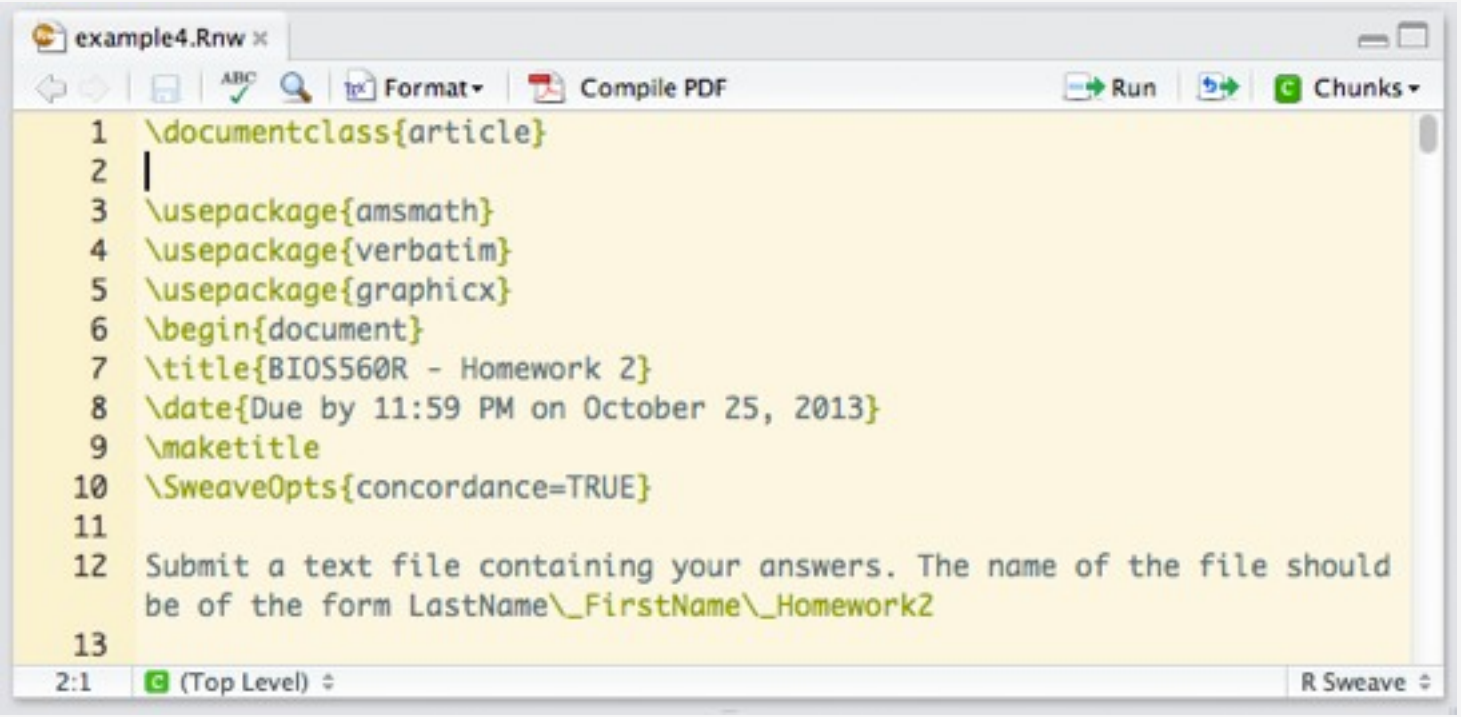


Detour - LaTex

Clear all the existing contents from the current window.

Next, load up http://www.bimcore.emory.edu/BIOS560R/SUPP.DIR/latexhwork2.txt select all the contents and paste into the empty RStudio edit window.

Then resave the document. Then click the "Compile PDF" button.



Detour - LaTex

BIOS560R - Homework 2

Due by 11:59 PM on October 25, 2013

Submit a text file containing your answers. The name of the file should be of the form LastName_FirstName_Homework2

1 Problem 1 - Data Frame Questions - 30 points

Please read in the following data set as follows. This data frame contains information on several important characteristics of diamonds. You might want to familiarize yourself with the layout and column names using some of the functions we've discussed in class. In all questions please present the R statements used to arrive at the answer.

```
nyd = read.table("http://www.bimcore.emory.edu/BIOS560R/DATA.DIR/my.diamonds.csv",header=T,sep=",")

names(nyd)
 [1] "carat"   "cut"     "color"   "clarity" "depth"   "table"   "price"
 [8] "x"       "y"       "z"

head(nyd)
  carat       cut color clarity depth table price    x    y    z
1  1.83      Fair     H      I1  68.0    57  6887 7.47 7.44 5.07
2  0.36     Ideal     F     VS1  60.1    57  1013 4.67 4.62 2.79
3  0.99 Very Good     G     SI1  62.8    56  4863 6.34 6.36 3.99
4  0.30     Ideal     H     VS1  62.2    57   545 4.28 4.31 2.67
5  0.30     Ideal     D    VVS2  61.6    57   814 4.31 4.32 2.66
6  0.53     Ideal     F     VS2  61.1    56  1753 5.26 5.22 3.20
```

Detour - LaTex

R Journal http://journal.r-project.org/archive/2013-1/eugster-schlesinger.pdf

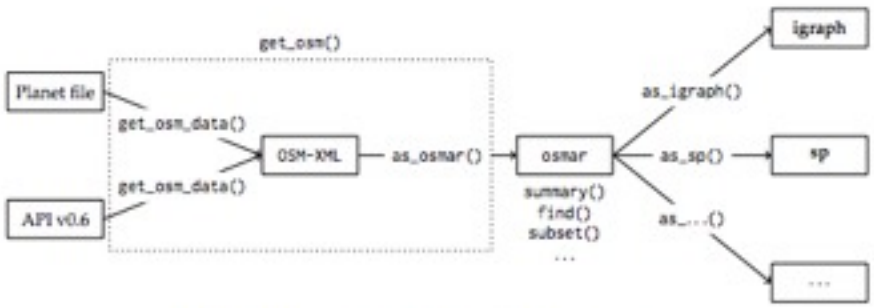


Figure 1: Schematic description of the **osmar** concept.

Getting the data

We begin with defining the data source. Currently two sources, HTTP-API and planet files, are supported. In this section we use the API of the OSM project with the default URL `url = http://api.openstreetmap.org/api/0.6/`:

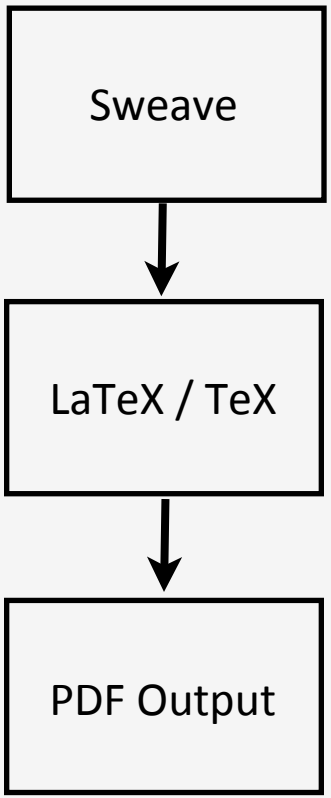
```
> library("osmar")

Loading required package: XML
Loading required package: RCurl
Loading required package: bitops
Loading required package: gtools
Loading required package: geosphere
Loading required package: sp

Attaching package: 'osmar'

The following object(s) are masked
from 'package:utils':
```

Detour - LaTex



Sweave documents are a mixture of LaTeX statements and R code "weaved" together

The LaTeX and TeX systems process the Sweave file including the R code

The output is a PDF file although users can freely share the original Sweave document

Back to Sweave

* The Sweave file has a file extension of .Rnw and is a mix of LaTex and R code

* The R code is contained within special tags that start with <<>>= and end with @ This is where it can be tedious - keeping up with these without making a mistake.

* Control commands can be specified within <<>>

* For example consider the following. In our coolpkg.Rnw make the "Results" section look like this:

```
\section{Results}
Some data analysis results. Below is example of table function.
<<>>=
table(mtcars$cyl)
@
```

Back to Sweave

The typical steps of Sweave involve the following work flow:

1) Using an editor (e.g. RStudio) create a file with the extension .Rnw This text file is a mixture of LaTex and R code.

2) Process the file using the Sweave function that is built-in to R. Here we use the command line console but RStudio would be easier as you will soon see.

```
> Sweave("coolpkg.Rnw")
Writing to file
coolpkg.tex
Processing code chunks with options ...
You can now run (pdf)latex on ‘coolpkg.tex'
```

3) Process the resulting .tex file using your LaTex compiler (latex, pdflatex, etc)

-OR-

Let RStudio handle all of this for you

Back to Sweave

This is what the document coolpkg.Rnw looks like:

```
\documentclass{article}
\usepackage{amsmath}

\title{Manual of coolpkg}
\author{John Doe}

\begin{document}
\maketitle

\section{Introduction}
Some introduction to my coolpkg.

\section{Results}
Some data analysis results. Below is an example of the table function.
<<>>=
table(mtcars$cyl)
@
\section{Conclusion}
Conclusion texts are here.

\end{document}
```

Back to Sweave

Manual of coolpkg

John Doe

August 6, 2013

1 Introduction

Some introduction to my coolpkg.

2 Results

Some data analysis results. Below is example of table function.

```
> table(mtcars$cyl)

 4  6  8
11  7 14
```

3 Conclusion

Conclusion texts are here.

Back to Sweave

So let me remind you why this is worth the trouble....

Well for starters you can embed any type of R code you want in the document and it will be evaluated to its full result so you don't have to format the output. So if you update your document over time you are in effect creating a document that reproduces the original research as opposed to you just pasting in some hard coded results.

Second you have at your disposal all the cool things in LaTex which is a professional typesetting system for preparation of publication ready documents. It is especially useful for any type of work that require mathematical and scientific notation of any type. There really is no other system that offers similar flexibility.

Third, you might have to use a tool like this to prepare your thesis, a publication, or an abstract.

Back to Sweave

```
\documentclass{article}
\usepackage{amsmath}
\title{Manual of coolpkg}
\author{John Doe}
\begin{document}
\maketitle
\section{Introduction}
Some introduction to my coolpkg.
\section{Results}
Some data analysis results. Below is an example of the lm function.
<<>>=
my.lm = lm(mpg~wt, data=mtcars)
print(my.lm)
@
<<fig=TRUE>>=
par(mfrow=c(2,2))
plot(my.lm)
@
\section{Conclusion}
Conclusion texts are here.
\end{document}
```

Back to Sweave

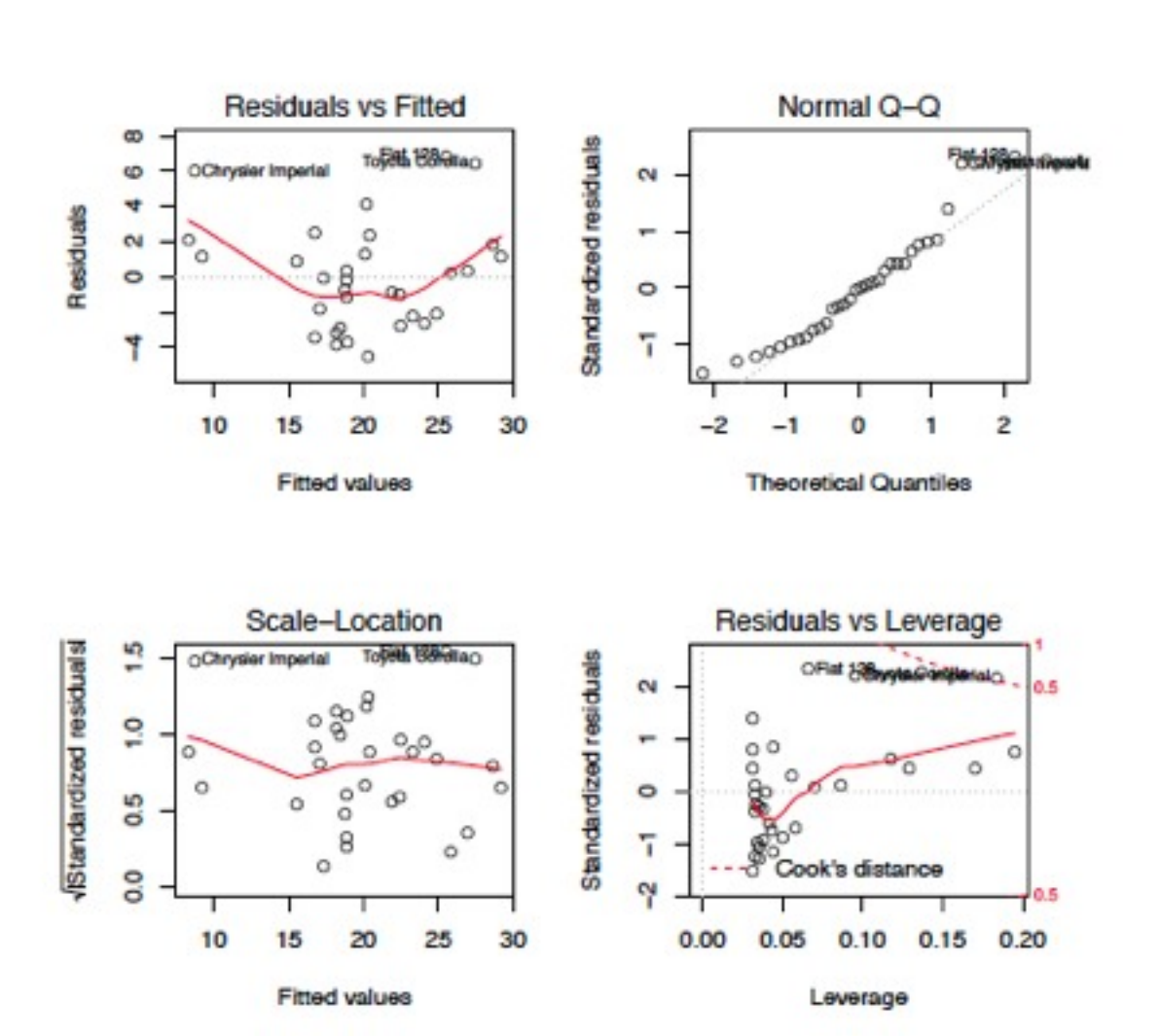
```
> my.lm = lm(mpg~wt, data=mtcars)
> print(my.lm)

Call:
lm(formula = mpg ~ wt, data = mtcars)

Coefficients:
(Intercept)           wt
      37.29        -5.34

> par(mfrow=c(2,2))
> plot(my.lm)
```

Back to Sweave



Back to Sweave

```
\documentclass[a4paper]{article}
\title{Sweave Example 1}
\author{Friedrich Leisch}
\usepackage{Sweave}
\begin{document}
\SweaveOpts{concordance=TRUE}
\maketitle
In this example we embed parts of the examples from the
\texttt{kruskal.test} help page into a \LaTeX{} document:

<<>>=
data(airquality , package="datasets")
library ("stats")
kruskal.test(Ozone ~ Month, data = airquality)
@

which shows that the location parameter of the Ozone distribution varies
significantly from month to month. Finally we include a boxplot of the data:
\begin{center}
<<fig=TRUE>>=
library ("graphics")
boxplot(Ozone ~ Month, data = airquality)
@
\end{center}
\end{document}
```

Back to Sweave

Sweave Example 1

Friedrich Leisch

November 19, 2013

In this example we embed parts of the examples from the `kruskal.test` help page into a LaTeX document:

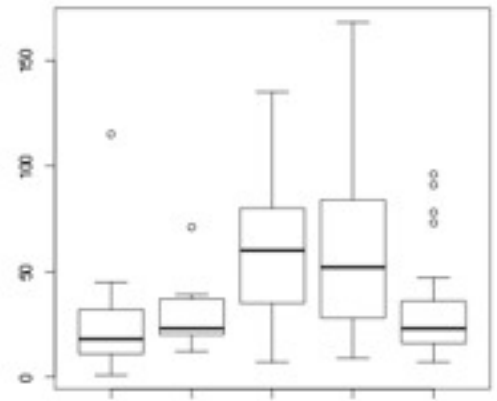
```
> data(airquality , package="datasets")
> library ("stats")
> kruskal.test(Ozone ~ Month, data = airquality)

        Kruskal-Wallis rank sum test

data:  Ozone by Month
Kruskal-Wallis chi-squared = 29.3, df = 4, p-value = 6.901e-06
```

which shows that the location parameter of the Ozone distribution varies significantly from month to month. Finally we include a boxplot of the data:

```
> library ("graphics")
> boxplot(Ozone ~ Month, data = airquality)
```



Back to Sweave

```
<<eval=TRUE,echo=TRUE>>=
data(airquality , package="datasets")
library ("stats")
kruskal.test(Ozone ~ Month, data = airquality)
@
```

What happens if we play around with the eval and echo ?

```
eval=TRUE, echo=TRUE  (Prints R statements and their results)

eval=FALSE,echo=TRUE  (Prints just the R statements with no results)

eval=TRUE,echo=FALSE  (Prints only the result of the statements)

eval=FALSE,echo=FALSE (Pointless - nothing gets printed)
```

Back to Sweave

So when processing Sweave documents you might sometimes see a message like this:

```
This is pdfTeX, Version 3.1415926-2.4-1.40.13 (TeX Live 2012)
 restricted \write18 enabled.
..
..
! LaTeX Error: File `Sweave.sty' not found.

Type X to quit or <RETURN> to proceed,
or enter new name. (Default extension: sty)
```

Back to Sweave

This is the first of several possible errors that you might encounter when using Latex, Sweave, and building R packages. This doesn't always happen and the fix for this is quick.

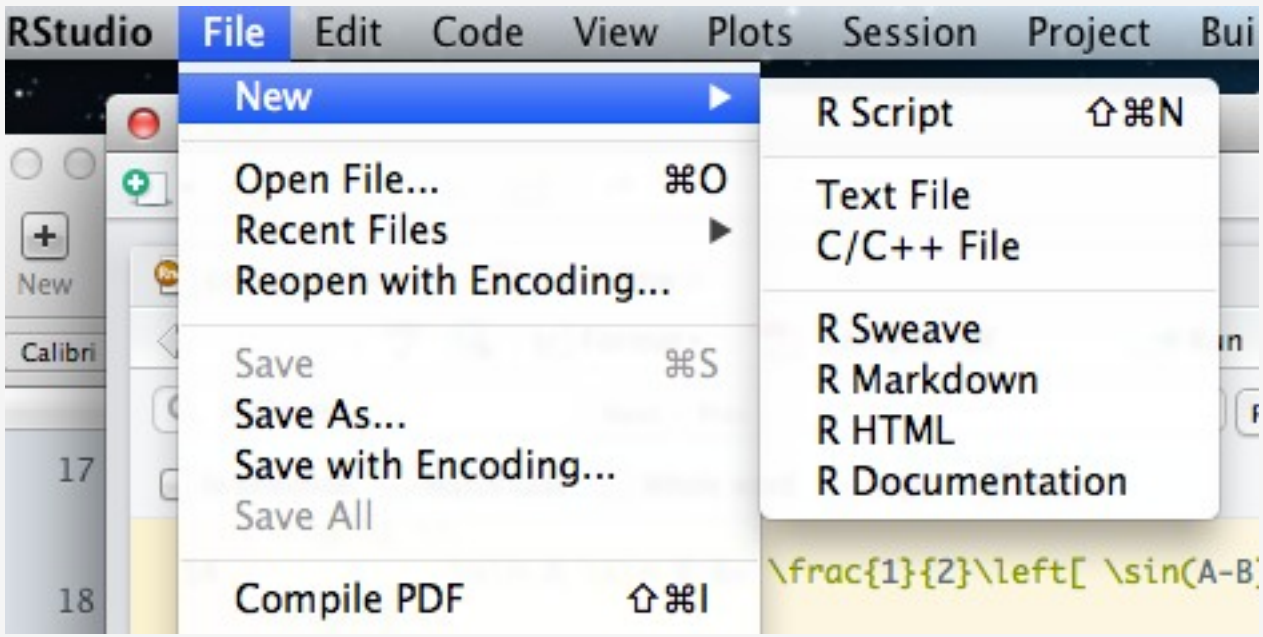
Go to http://www.bimcore.emory.edu/BIOS560R/DATA.DIR/Sweave.sty and do a "Save As" to whatever Folder or directory you are currently working in. The one that contains the cookpkg.Rnw file in this case.

Then repeat the above steps and all should be well.

Back to Sweave

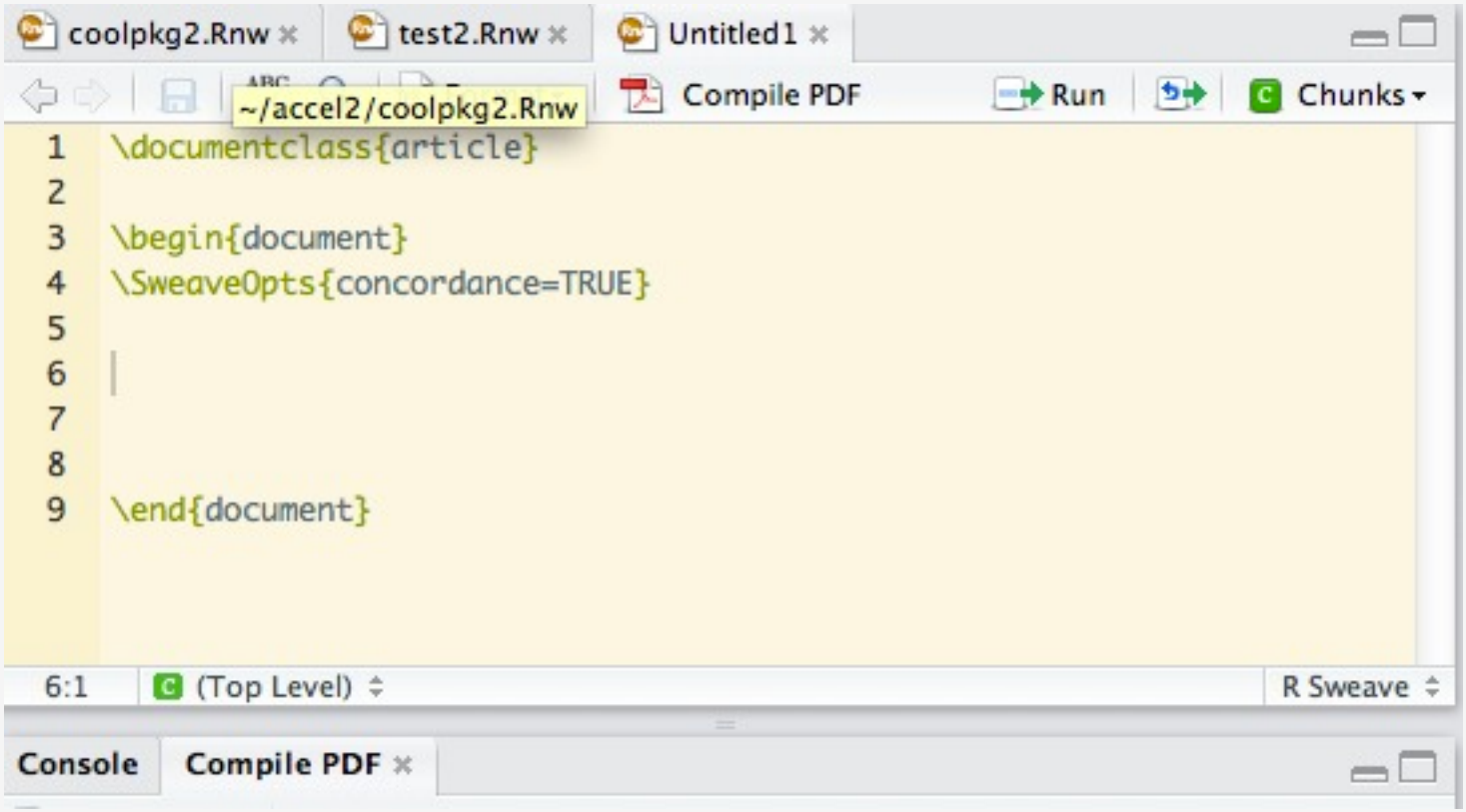
As I said earlier this can also be done with RStudio assuming that it has been set up correctly. I took you through the "hard way" first since it is the most general method of working with Sweave. With that knowledge you could generate your PDFs in most environments. To see what Rstudio has on offer just crank it up.

From the File menu select File > New > R Sweave



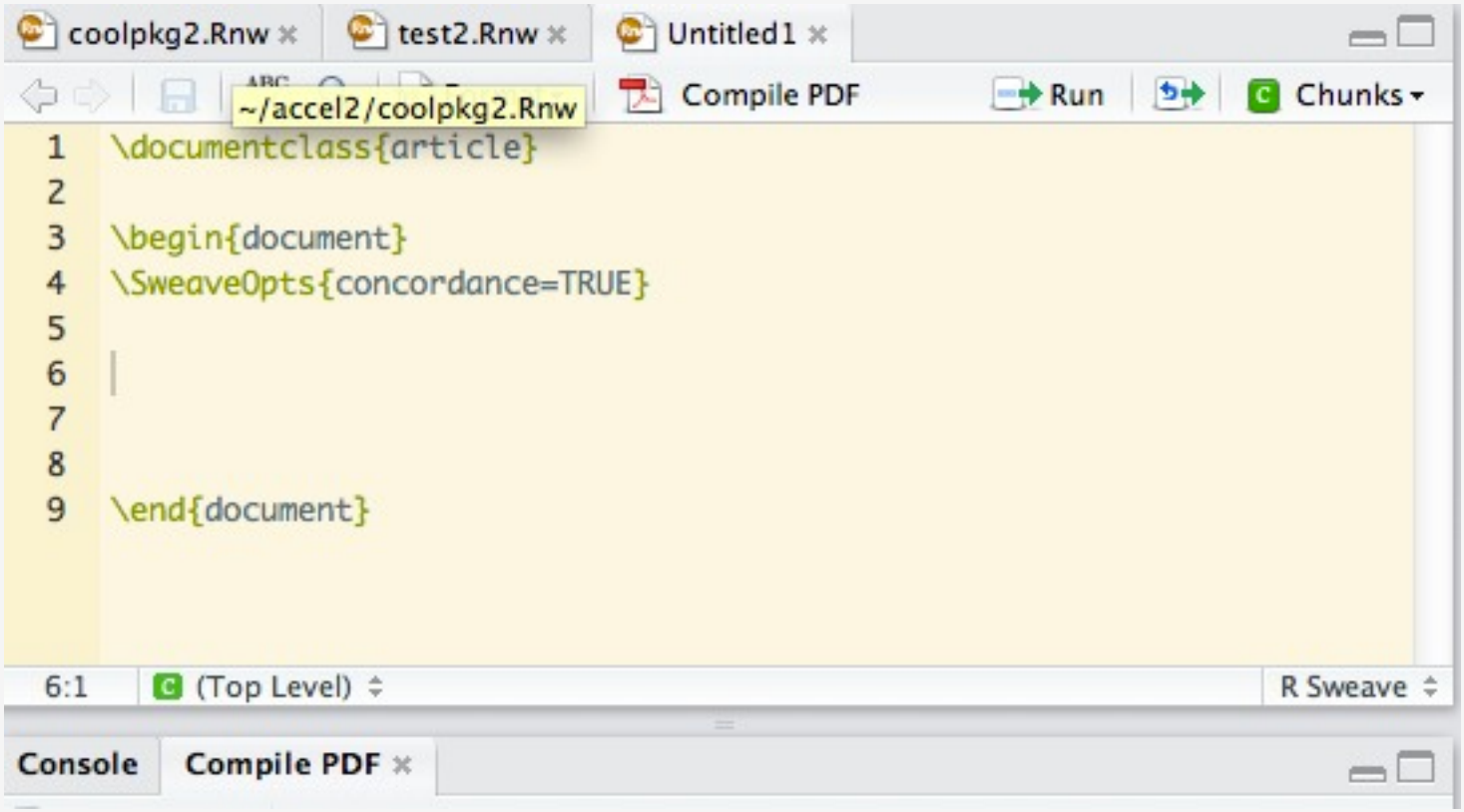
Back to Sweave

From the File menu select File > New > R Sweave. In the file pane you get an Untitled document that looks like:



Back to Sweave

You may now enter in LaTex commands and/or Sweave chunks. RStudio even has a "Chunks" menu to add things in for you.

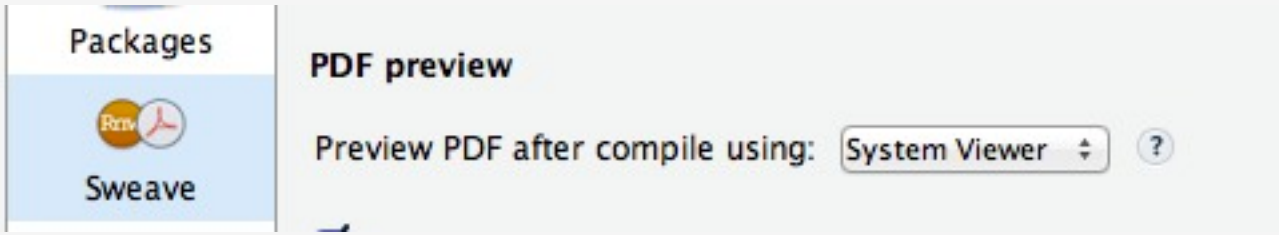


Back to Sweave

Or you can just paste in the chunks you want. Save your document to whatever name you want. Then hit the "Compile PDF" button. It should process your file and then launch a viewer with the results.

Back to Sweave

To get the viewer to work I had to go to Preferences -> Sweave and tell RStudio that I wanted to use the "System Viewer" as my PDF viewer.



The default was "RStudio Viewer" which I could not get to work. I'm not saying that you will have to do this but just keep it in mind if you encounter trouble.

Template

If you wanted to write your own Sweave document then you can start with this template:

www.bimcore.emory.edu/BIOS560R/SUPP.DIR/template.Rnw

Just load it up and copy/paste it into a blank Sweave document.