Sweave:
Integrating R into your TEX documents

Lizzie Wolkovich

10 October 2012

1 Important note: Sweave is dead easy

Sweave is just like LATEX, except with a dash of extra in the preamble. If you know R and LATEX and you know the following:

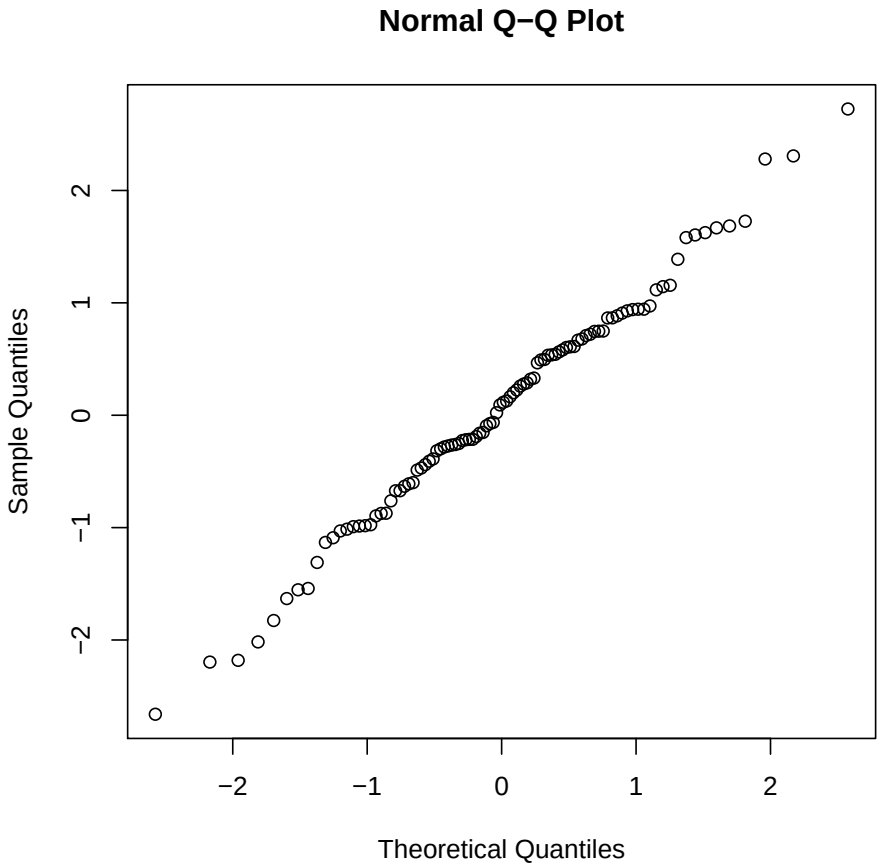
<<>>=
@

. . . then you can Sweave! Congrats. Now we can break out of here and go grab a drink.

No, no, just kidding. These things are often a little better in practice.

Let's practice.

2 Embed R code in your TEX document!



Okay, let's break down what just happened (behind the scenes for those just looking at the pdf):

1. The code <<>>= tells TEX that you are about to feed it R code.
2. The code @ tells TEX the R code has ended.
3. Within the <<>>= and @ you just type R code.
4. Within the <<>>= is the big place to get crazy! Here I told it a couple things:
   (a) `label=nqpugly` is just a handy label, when your Sweave file is compiling and breaks it's nice to have this, among other handy things you can use labels for. I called mine 'nqpugly' for normal quantile plot (NQP) that is ugly. I could just have easily called it 'sweaverocks', but I didn't.
   (b) `echo=FALSE` means I don't want TEX to show the R code, note that if you are in the `.Rnw` file you see `qqnorm (rnorm(100))`, but in the output you don't. That's no echoing (pardon my English). The default is to echo.
   (c) `results=hide` means don't show the results either. The default is to show all your results.
   (d) `fig=TRUE` means 'yes! show the figure.' This, in my mind, is a hack figure. Let's fix it up a touch now (we'll center it and add a caption), since we're here.

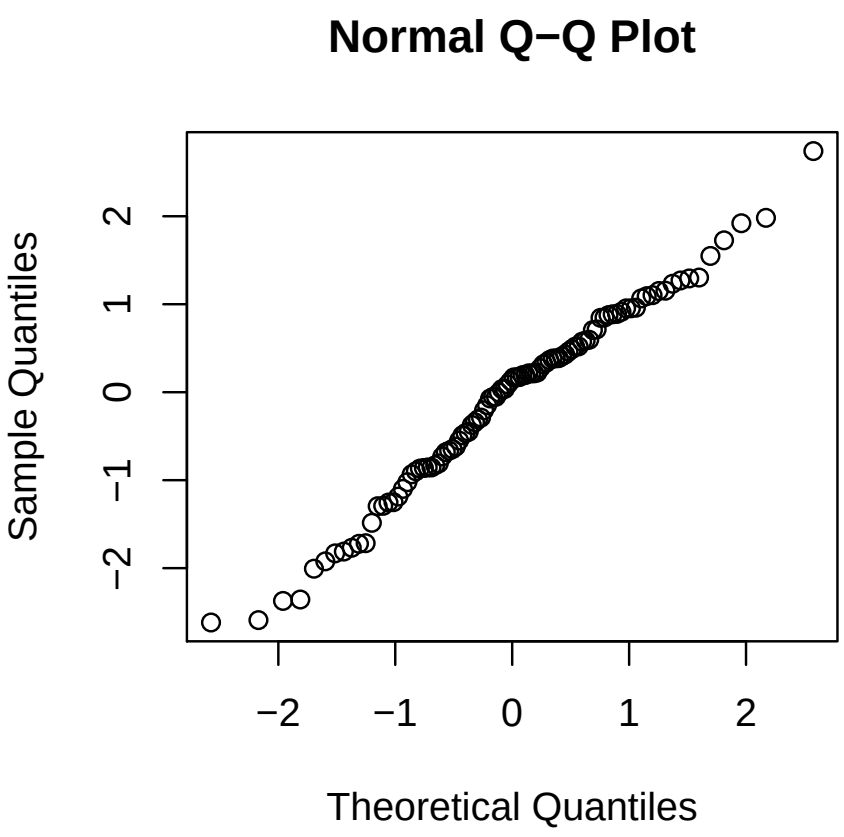


Figure 1: It's a normal quantile plot, built from random data.

Excellent, moving on let's use some real data. We'll grab some data and ask R about it.

```
  year harvestdate
1 1370        27.0
2 1371        25.0
3 1372        28.1
4 1373        20.7
5 1374        28.2
6 1375        20.2
```

```
[1] 1370 2003
```

Since I did not write `results=hide` you see the output.

Let's contrast the above if I let echoing happen.

```
> setwd("/Users/Lizzie/Documents/git/demoSweave")
> pinot <- read.delim("data/pinotnoir.txt", header = TRUE)
> head(pinot)
```

```
  year harvestdate
1 1370        27.0
2 1371        25.0
3 1372        28.1
4 1373        20.7
5 1374        28.2
6 1375        20.2
```

```
> range(pinot$year)
```

```
[1] 1370 2003
```

By the way, I took these data from the National Climatic Data Center in the USA (see: `http://www.ncdc.noaa.gov/paleo/pubs/chuine2004/chuine2004.html` if you want the metadata for them). They show the grape harvest records (harvest date is given as days after September 1) from Burgundy for the past 600 or so years (the dominant variety in Burgundy is Pinot Noir, hence my naming of the dataframe). Grape harvest records, like these, have been used to reconstruct past climate. They also offer evidence of recent changes in climate, at least partially associated with increasing greenhouse gas emissions (whee—that last clause is how all my climatologists always tell me to say it).

Next, let's make some tables. For this we use the `xtable` package from R. According to the Intergovernmental Panel on Climate Change (IPCC) significant warming began in 1970, so we'll subset the data to after then and look for a trend.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
year	1	560.96	560.96	7.68	0.0095
Residuals	30	2190.99	73.03		

Table 1: An ANOVA table, so little effort for so much prettiness

Note the `results=tex` command, this makes the output in TEX formatting.

ANOVA tables are pretty dull though, so let's do a more real-life example. Let's contrast this ~ 30 year time period with three others I haphazardly selected. We'll show the degrees of freedom, F statistic and p-value, as well as the mean value (which is change per year).

Table 2: Comparisons of regression of harvest dates for Pinot Noir in Burgundy region of France across different centuries

	days/year	DF	F	p-value
1970-2003	-0.44	30	7.68	0.009
1770-1803	0.03	30	0.04	0.836
1570-1603	0.15	30	0.74	0.396
1370-1403	-0.07	30	0.23	0.633

Let's break down a little of what happened here (again, you'll enjoy this more looking at the `Rnw` file, instead of the `pdf`):

1. First you see a bunch of ugly R code to pull together what I want in each row, depending on what you're doing, you can automate this and make it much prettier!

2. Next I bind the rows into a dataframe where I set the row names to be useful.
3. I make the column headers (`names`) useful as well.
4. Then I call `xtable`, I give it the dataframe to make into a table, I tell it a caption, and I tell it the number of significant digits for each row. You don't have to specify the `digits` command (as we didn't in the ANOVA table example), but here it's nice to adjust it for the DFs versus other values.

Okay, how about one last trick? We saw that since 1970 harvest dates of the delicious Pinot Noir grapes from Burgundy have been advancing each year. Let's say we want to discuss the exact value in the text. We want to say that harvest dates of the delicious Pinot Noir grapes from Burgundy have been shifting about -0.44 days each year, or -14 days since 1970.

So, all we did was use `Sexpr{}`, and stuck in some R code in line. I also tossed in that `digits` command again with `format`, if I didn't it would look like this: ... delicious Pinot Noir grapes from Burgundy have been shifting about -0.439171170396767 days each year. And, let's be honest, the latter is just not going to impress your friends or reviewers that much.

3 Now you!

Now would be a great time for you to muck with this document. First, get it to run on your machine[1], next try turning on and off things like `echo=FALSE` or `results=hide`.

Then, may I suggest you try plotting the Pinot Noir data yourself? Maybe add a fit line? I specifically avoided such fun so you could see it. Try plotting all of it, then 1970 onward. You could also see if things were a little weird around 1883 when Krakatoa blew and mucked up the global climate a little (I haven't even done that myself so I would be interested to see it).

4 A bunch of random notes

Where to next, fellow Sweaver?

Well, as you may have guessed, we didn't cover everything here. I keep learning more every time I build a giant new Sweave file and suspect you will too. I have some resources on my webpage (`http://www.zoology.ubc.ca/~wolkovich/emwresource.htm#latex`) and you'll find many others online.

Before we part though some random notes I wanted to share:

- To pull just your R code out of a Sweave doc you can use the following in R: `Stangle("filename.Rnw",output="filenameout")`.
- One easy error is to forget to close your R code. To be safe and avoid this, every time I write the opening `<<>>=` I also write the close @.
- You can use escape slashes inside R chunks of code to get LaTeX formatting you may want when it's written (for example, in a caption).
- The file extension for Sweave files with R is `.Rnw`, for Sweave files with S it's `.Snw`. This doesn't actually matter in practice, I just thought I would share it.

[1] Important note for TeXshop users! Be sure you set your drop-down menu from LaTeXto Sweave before you try to compile.