

Example of Knitr

John Muschelli

July 28, 2015

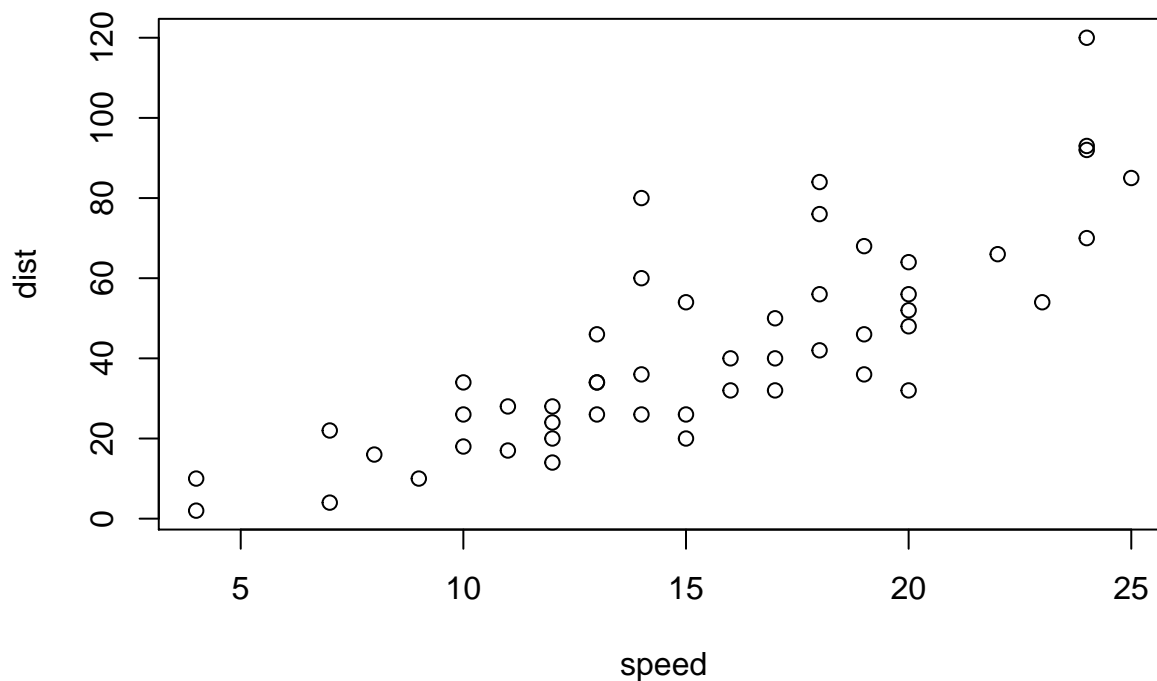
A generic `knitr` document is like any other document. The output can be HTML/PDF/Word:

The generic setup from RStudio is this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median:15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

You can also embed plots, for example:



The mean car speed is 15.4. Sometimes you want to use `sprintf` and standard number formatting in the command, such as `sprintf('%03.2f', mean(cars$speed))` to get the mean car speed is 15.40. A lot of times, this has to do with p-values and formatting.

Reproducible Research

Have you ever run something once? I haven't. For any project. Ever. Now it's only been about 8 years for my biostats/analysis experience, but I'm sure I never will do anything (plot/table/test/final figure) once.

This brings me to the real purpose of `knitr` and weaving/knitting. Reproducible research.

Wouldn't it be nice if you could drop one patient out that someone forgot to tell you didn't really meet the inclusion criteria and run your whole paper's analysis? Well, if you write a markdown document, you can!

Think of 2 scenarios. You have a `data.frame`/matrix/csv, whatever that stores your subject ids.

Scenario 1

You wrote "I had 30 people in my analysis" (that's a terrible sentence). Oh no! Now you have 29, better go back through and adjust the p-values/tables/figures/etc.

Scenario 2

You wrote "I had `nrow(df)` people in my analysis" (still a terrible sentence). Oh no! Now you have 29, but R only sees `nrow(df)` and now `df` has 29 rows. As long as your other files are updated accordingly, all works

P-value Bonanza

We all use p-values, don't lie. Now you updated to 29 people, and the $p < 0.01$ is not $p = 0.02$. Oh snap! Reviewer 1 is about to have something to write ON and ON about.

What if you calculated a p-value in your R code and then said something like `stopifnot(pvalue < 0.01)`. When you run your new data, the code errors and you know something's wrong.

"Meta-Writing": A Dangerous/Strange Road

I see that people could do some crazy stuff in the future like this:

```
if (pvalue < 0.05){
  mystr = "The result was significant (p < 0.05) blah blah"
} else {
  mystr = paste0("The result was not significant (p = ", pvalue, ") blah blah")
}
```

And then embedding the sentence `mystr` in the text. I do see some value in this, but maybe it could be good. Maybe your collaborator which asks you to do things 8000 different ways switches their mind every day. Then having a flag to day `do_fun_analysis = TRUE` and scenarios where this is `FALSE`. I've done this, but it's a bit too far in my opinion.

Startup cost

At the end of the day, the startup cost is non-trivial, as with all changes. The changes can save you significant time down the road. Even if collaborators are not finicky, reviewers sure can be, and ask you to add things that do not make sense.