LaTeX and R Integration

Mark Westcott

July 31, 2015

1 Idea

The idea is to integrate your LaTeX and R code into a single file, which will have an .Rnw extension. You can then compile a PDF direct from RStudio.

2 RStudio configuration

There are two systems that can do the work of producing a TeX file, Sweave and knitr. Tell RStudio to use 'knitr', by going to Tools, Global Options, Sweave and seting 'Weave Rnw file using' to 'knitr'.

3 Starting a project

Create a new project inside RStudio. Projects are neat - you get your environment saved between sessions as well as all your history. Create one by selecting File, New Project and clicking through the wizard.

4 Creating a Rnw file

Your combined LaTeX/R file should have the Rnw extension. Create one with File, New File, R Sweave. This creates a LaTeX document that you can compile by pressing 'Compile PDF'

5 Adding R code

You include your R code by inserting a named 'Chunk' into your .Rnw file like this:

<<myFirstChunk>>=
myRcode()
@

Hide the code by setting echo to FALSE or hide results by setting results to hide

<<hiddenChunk, echo=F, results='hide' >>=
myRcode()
@

You can also use \Sexpr{} in your LaTeX code to subsitute R output into LaTeX.

<<, echo=F, results='hide' >>=
l <- 2.2
@

l has been assigned the value \Sexpr{l} yields:

l has been assigned the value 2.2

6 Tables with stargazer

We'll be using stargazer to produce LaTeX tables direct out of our R regressions. There are lots of packages for doing this. If stargazer doesn't do everything you want (for example, it won't produce booktables), then look at texreg instead, it is more flexible, but slightly harder to figure out the documentation.

Set the results option of the chunk to 'asis'. This will tell knitr to copy the LaTeX output of stargazer straight into your document, without trying to reformat it.

<<, echo=F, results='asis' >>=
call to stargazer goes here
@

6.1 Minimal example

```
#install.packages("stargazer")
library(stargazer)

df <- data.frame(y=rnorm(10),x1=1:10,x2=sample(10))
lm1 <- lm(y ~ x1 + x2, df)
```

```
stargazer(lm1)
```

Table 1:

	Dependent variable:
	y
x1	0.051
	(0.090)
x2	−0.247**
	(0.090)
Constant	1.076
	(0.809)
Observations	10
R^2	0.561
Adjusted R^2	0.436
Residual Std. Error	0.797 (df = 7)
F Statistic	4.473* (df = 2; 7)
Note:	*p<0.1; **p<0.05; ***p<0.01

6.2 A regression on the 'PublicSchools' dataset

```
library("AER")
data(PublicSchools)
public_schools_r <- lm(Expenditure ~
                       Income + I(Income^2), data=PublicSchools)
```

```
stargazer(public_schools_r)
```

Table 2:

	Dependent variable:
	Expenditure
Income	−0.183**
	(0.083)
I(Income^2)	0.00002***
	(0.00001)
Constant	832.914**
	(327.292)
Observations	50
R^2	0.655
Adjusted R^2	0.641
Residual Std. Error	56.679 (df = 47)
F Statistic	44.684*** (df = 2; 47)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can provide extra options to stargazer to customize the output:

```
stargazer(public_schools_r,
          covariate.labels = c("Income","Income$^2$"),
          omit.stat = c("f", "adj.rsq", "ser"),
          style="default",
          title="Regression Outputs")
```

```
coefficient_on_income = sprintf("%.2f",public_schools_r$coefficients["Income"])
```

We have 51 observations in the dataset but only 50 in the regression output, due to a missing observation. The coefficient on income is -0.18.

Table 3: Regression Outputs

	Dependent variable:
	Expenditure
Income	-0.183^{**}
	(0.083)
Income^2	0.00002^{***}
	(0.00001)
Constant	832.914^{**}
	(327.292)
Observations	50
R^2	0.655
Note:	*p<0.1; **p<0.05; ***p<0.01

6.3 Different SEs across models

```
vc_homo <- vcov(public_schools_r)  #homosk
vc_hc3 <- vcovHC(public_schools_r) #robust, default is HC3
vc_hc1 <- vcovHC(public_schools_r, type = "HC1") #robust, Stata

ses_homo <- sqrt(diag(vc_homo)) # homosk
ses_hc3 <- sqrt(diag(vc_hc3))   # hdc3
ses_hc1 <- sqrt(diag(vc_hc1))   # hc1
```

```
stargazer(public_schools_r, public_schools_r, public_schools_r,
          se = list(ses_homo, ses_hc3, ses_hc1),
          keep = c("Income"),
          covariate.labels = c("Income","Income$^2$"),
          omit.stat = c("f", "adj.rsq", "ser"),
          style="default",
          intercept.bottom = F,
          intercept.top = F,
          title="Regression Outputs")
```

Table 4: Regression Outputs

	Dependent variable:		
	Expenditure		
	(1)	(2)	(3)
Income	−0.183**	−0.183	−0.183
	(0.083)	(0.298)	(0.128)
Income^2	0.00002***	0.00002	0.00002*
	(0.00001)	(0.00002)	(0.00001)
Observations	50	50	50
R^2	0.655	0.655	0.655
Note:		*p<0.1; **p<0.05; ***p<0.01	

6.4 Different models

```
admissions <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")

myols <- lm(admit ~ gre + gpa + rank, data = admissions)

myprobit <- glm(admit ~ gre + gpa + rank, binomial("probit"),
                data = admissions)

mylogit <- glm(admit ~ gre + gpa, binomial("logit"),
                data = admissions)
```

```
stargazer(myols, myprobit, mylogit,
          keep = setdiff(names(myols$coefficients),"(Intercept)"),
          covariate.labels = c("GRE score","GPA score","Rank in class"),
          omit.stat = c("f", "adj.rsq", "ser","aic","ll"),
          dep.var.labels = c("Admissions decision"),
          add.lines = list(c("Some control","Y","Y","Y")),
          style="default",
          intercept.bottom = F,
          intercept.top = F,
          title="Regression Outputs")
```

Table 5: Regression Outputs

	Dependent variable:		
	Admissions decision		
	OLS	*probit*	*logistic*
	(1)	(2)	(3)
GRE score	0.0004**	0.001**	0.003**
	(0.0002)	(0.001)	(0.001)
GPA score	0.151**	0.464**	0.755**
	(0.063)	(0.195)	(0.320)
Rank in class	−0.110***	−0.332***	
	(0.024)	(0.075)	
Some control	Y	Y	Y
Observations	400	400	400
R^2	0.096		
Note:		*p<0.1; **p<0.05; ***p<0.01	

7 Summary tables and custom tables

Summary tables:

```
nice_labels = c("Miles/(US) gallon",
                "Cylinders",
                "Displacement",
                "Gross horsepower",
                "Drat",
                "Weight (lb/1000)",
                "1/4 mile time",
                "V/S",
                "Is automatic",
                "Gears",
                "Carburetors")

stargazer(mtcars, summary.stat = c("mean","sd"),
                  covariate.labels=nice_labels)
```

You an also output any data frame you want by passing the 'summary=FALSE' argument.

Table 6:

Statistic	Mean	St. Dev.
Miles/(US) gallon	20.091	6.027
Cylinders	6.188	1.786
Displacement	230.722	123.939
Gross horsepower	146.688	68.563
Drat	3.597	0.535
Weight (lb/1000)	3.217	0.978
1/4 mile time	17.849	1.787
V/S	0.438	0.504
Is automatic	0.406	0.499
Gears	3.688	0.738
Carburetors	2.812	1.615

```
stargazer(Duncan[1:15,], summary=FALSE)
```

Table 7:

	type	income	education	prestige
accountant	prof	62	86	82
pilot	prof	72	76	83
architect	prof	75	92	90
author	prof	55	90	76
chemist	prof	64	86	90
minister	prof	21	84	87
professor	prof	64	93	93
dentist	prof	80	100	90
reporter	wc	67	87	52
engineer	prof	72	86	88
undertaker	prof	42	74	57
lawyer	prof	76	98	89
physician	prof	76	97	97
welfare.worker	prof	41	84	59
teacher	prof	48	91	73

8 Short texreg example

	Model 1
Income	0.05
	(0.09)
Income^2	-0.25^{**}
	(0.09)
R^2	0.56
Num. obs.	10
RMSE	0.80

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 8: Statistical models

```
library(texreg)
```

```
texreg(lm1,
       dcolumn = TRUE, booktabs = TRUE, use.packages=F,
       stars = c(0.1,0.05,0.01),
       include.adjrs = FALSE,
       omit.coef = "Interc",
       custom.coef.names = c("Intercept","Income","Income$^2$")
       )
```

9 Including graphics

This really easy – any graphical output automatically gets put in a \figure environment.

<<, fig.cap="Fuel Efficienty" >>=

```
library(ggplot2)
ggplot(data = mtcars, aes(x = mpg, y = wt)) + geom_point()
```

@

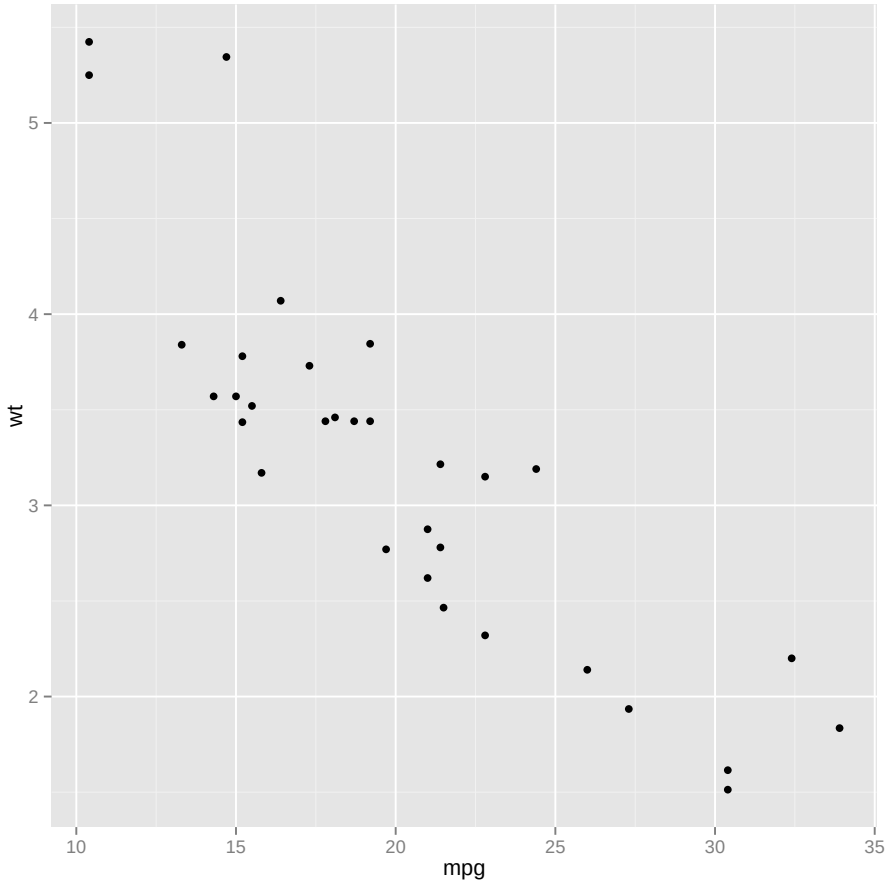

Figure 1: Fuel Efficiency