Short example of R Markdown

Garret Christensen

Wednesday, June 10, 2015

This is an R Markdown document, opened in RStudio. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

R Studio is a free user-friendly implementation of R that brings a lot of tools (LaTeX and version control, most notably) to help with reproducibility into the same environment. Feel free to use these tools on their own for more flexibility, but it's a nice option to have them under one roof.

R Markdown is pretty similar to regular Markdown, which is called mark*down* because it's simpler than a full mark*up* language like LaTeX or HTML. Find a quick syntax reference under the '?' menu above.

The main thing we'll do is use R Markdown and R Studio to write a **dynamic document**. That's a document that has the data, analysis code, and output all in one place, and is updated automatically. Basically, you reduce the likelihood of error by eliminating copy and pasting, and your work is more easily reproduced by others.

There are two big dynamic document packages. Sweave, and the newer Knitr. We'll mostly talk about Knitr, but I have a simple Sweave example as well.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

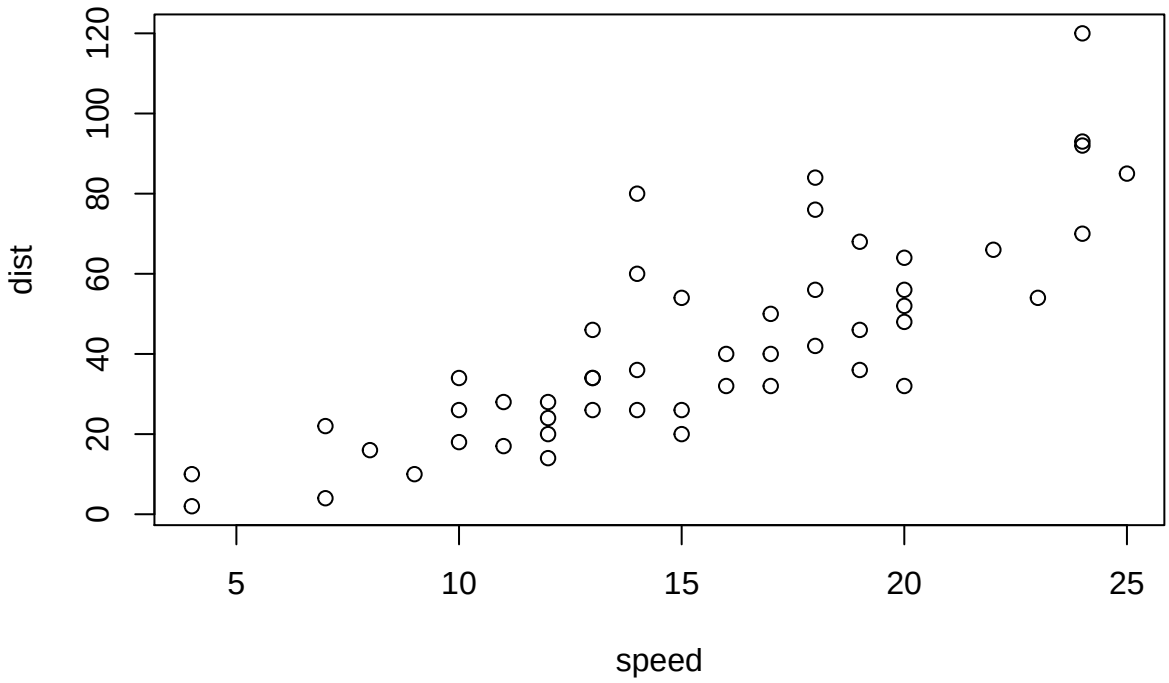
```
summary(cars)
```

```
##      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
```

```
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

You can also embed plots. For example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Loading Data

You can bring in Stata data (up to version 12) directly with the 'foreign' package. You only have to install a package once, but you have to load the library every time.

```
#install.packages("foreign")
library(foreign)
WASHB<-read.dta("C:/Users/garret/Documents/Research/BITSS/WorkshopSlides/Riverside/WASHBpublic_mock.dta
```

```
## Warning in `levels<-`(`*tmp*`, value = if (nl == nL) as.character(labels)
## else paste0(labels, : duplicated levels in factors are deprecated
```

But let's bring in Clara's states data files and merge them together.

```
setwd("C:/Users/garret/Documents/Research/BITSS/WorkshopSlides/BITSS2015Summer")
ratings<-read.csv("ratings.csv")
```

Running Analysis

```r
#Model 1=simple model
mod1 <- lm(meanSizeRating ~ Frequency + Class + meanWeightRating + DerivEntropy, data=ratings) #Without
mod2 <- lm(meanSizeRating ~ Frequency * Class + meanWeightRating + DerivEntropy, data=ratings) #With in

#Model 3=with hetero-robust se's--a bit more complicated than Stata
#install.packages("sandwich")
library(sandwich)
cov <- vcovHC(mod2, type = "HC")
robust.se <- sqrt(diag(cov))
```

Equations

Equations written with LaTeX syntax works, so you can write short reports all in one file.

$$\frac{dN}{dt} = r * N * (1 - \frac{N}{K})$$

Refer to Values

You can refer to values calculated in R by just surrounding "r" and the code with single accent marks. For example, the mean frequency is 4.7611931.

Simple Output

You can just use built in R functionality.

```r
summary(mod1)
```

```
##
## Call:
## lm(formula = meanSizeRating ~ Frequency + Class + meanWeightRating +
##     DerivEntropy, data = ratings)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.016662 -0.006126 -0.000257  0.005461  0.022904
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.745607   0.008563  87.073   <2e-16 ***
## Frequency        -0.021633   0.001170 -18.497   <2e-16 ***
## Classplant       -0.093553   0.006029 -15.517   <2e-16 ***
## meanWeightRating  0.898063   0.003020 297.389   <2e-16 ***
## DerivEntropy     -0.005552   0.002903  -1.912   0.0596 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008464 on 76 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 2.602e+05 on 4 and 76 DF,  p-value: < 2.2e-16
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = meanSizeRating ~ Frequency * Class + meanWeightRating +
##     DerivEntropy, data = ratings)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0154317 -0.0055852  0.0003123  0.0053354  0.0229305
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.742681   0.008039  92.389  < 2e-16 ***
## Frequency             -0.017815   0.001546 -11.527  < 2e-16 ***
## Classplant            -0.074998   0.007742  -9.687 7.43e-15 ***
## meanWeightRating       0.893774   0.003076 290.599  < 2e-16 ***
## DerivEntropy          -0.009617   0.002950  -3.260 0.001676 **
## Frequency:Classplant -0.005776   0.001655  -3.491 0.000811 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.007902 on 75 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 2.388e+05 on 5 and 75 DF,  p-value: < 2.2e-16
```

Fancier Output

R Markdown is mostly for simple stuff. Like I said, mark*down*, not mark*up*. But you can still get really nicely formatted regression output with a couple of R packages, xtable or stargazer. (Very similar to estout or outreg2 in Stata.)

Stargazer has three types of output (text, html, and LaTeX).

HTML->HTML

First, I'll do output as html. We can knit right to html here, and it looks great.

```
library(stargazer)
```

```
##
## Please cite as:
##
##  Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression and summary
##  R package version 5.1. http://CRAN.R-project.org/package=stargazer
```

```
stargazer(mod1, mod2, mod2, se=list(NULL, NULL, robust.se), type="html", out="outputR.html", title="Mad
```

Made Automatically in R

Dependent variable:

meanSizeRating

default

robust

controls

(1)

(2)

(3)

Frequency

-0.022***

-0.018***

-0.018***

(0.001)

(0.002)

(0.001)

Classplant

-0.094***

-0.075***

-0.075***

(0.006)

(0.008)

(0.009)

meanWeightRating

0.898***

0.894***

0.894***

(0.003)

(0.003)

(0.003)

DerivEntropy

-0.006*

-0.010***

-0.010***

(0.003)

(0.003)

(0.003)

Frequency:Classplant		
-0.006***		
-0.006***		
(0.002)		
(0.001)		
Constant		
0.746***		
0.743***		
0.743***		
(0.009)		
(0.008)		
(0.011)		
Observations		
81		
81		
81		
R2		
1.000		
1.000		
1.000		
Adjusted R2		
1.000		
1.000		
1.000		
Residual Std. Error		
0.008 (df = 76)		
0.008 (df = 75)		
0.008 (df = 75)		
F Statistic		
260,222.300*** (df = 4; 76)		
238,817.200*** (df = 5; 75)		
238,817.200*** (df = 5; 75)		
Note:		
p<0.1; ***p<0.05;*** p<0.01		

TeX->PDF

When we Knit a Markdown as a PDF, it actually makes that PDF using LaTeX. (See here.) So you can use the .tex output option from stargazer and get nice PDF documentation.

```
stargazer(mod1, mod2, mod2, se=list(NULL, NULL, robust.se), title="Made Automatically in R", out="outpu
```

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Fri, Jun 05, 2015 - 4:12:27 PM

Table 1: Made Automatically in R

	Dependent variable:		
	meanSizeRating		
	(1)	(2)	(3)
Frequency	−0.022***	−0.018***	−0.018***
	(0.001)	(0.002)	(0.001)
Classplant	−0.094***	−0.075***	−0.075***
	(0.006)	(0.008)	(0.009)
meanWeightRating	0.898***	0.894***	0.894***
	(0.003)	(0.003)	(0.003)
DerivEntropy	−0.006*	−0.010***	−0.010***
	(0.003)	(0.003)	(0.003)
Frequency:Classplant		−0.006***	−0.006***
		(0.002)	(0.001)
Constant	0.746***	0.743***	0.743***
	(0.009)	(0.008)	(0.011)
Observations	81	81	81
R^2	1.000	1.000	1.000
Adjusted R^2	1.000	1.000	1.000
Residual Std. Error	0.008 (df = 76)	0.008 (df = 75)	0.008 (df = 75)
F Statistic	260,222.300*** (df = 4; 76)	238,817.200*** (df = 5; 75)	238,817.200*** (df = 5; 75)
Note:			*p<0.1; **p<0.05; ***p<0.01

Everything All in One Place?

You can do citations. Plots, graphs, and citations, what else do you need for a research paper?

Complicated, time consuming for very long articles.

Maybe try Sweave (direct combo of LaTeX and R).

Send your output to .tex files, include those in your master paper file.