

Intro to RMarkdown

Andrew Mullins

November 2, 2015

Motivation

New Workflow

Other Tools

Learning More

Motivation

Old Workflow - R to L^AT_EX

1. Do some cool stuff in R:

```
mod1 <- lm(mpg~hp, data = mtcars)
with(mtcars, plot(mpg~hp))
abline(mod1)
summary(mod1)
```

2. Copy/save the output/figure.
3. Paste into Latex document:

```
\begin{verbatim}
mod1 <- lm(mpg~hp, data = mtcars)
...
\includegraphics{figures/mpg.pdf}
```

Old Workflow - Problems

- ▶ What if we need to make a change?
- ▶ Latex is hard and requires a lot of boilerplate.
- ▶ Distracting: doesn't facilitate higher-order thinking.

Reproducibility

- ▶ If I win the lottery and drop out of graduate school, could someone else pick up where I left off?
- ▶ How long would it take them to do so?

Interactivity

- ▶ We live in the future. Why should our documents be static?
- ▶ Interactive documents are useful to everyone: ourselves, our students, executives, etc.

New Workflow

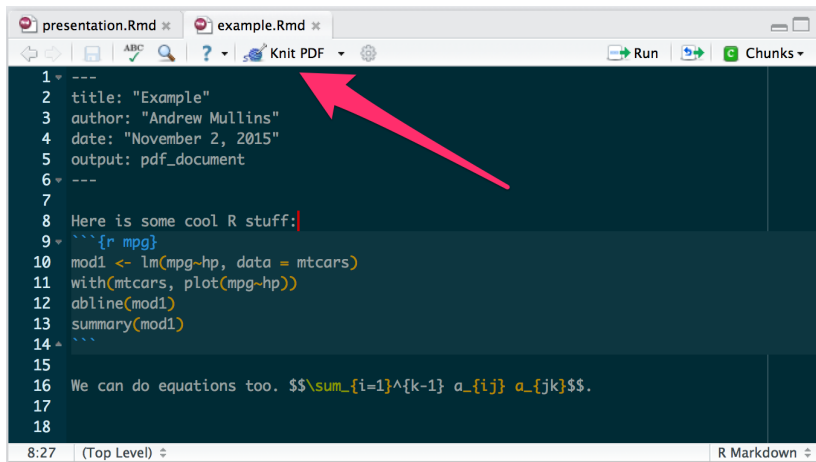
New Workflow - Unified Document

1. Do some cool stuff in RMarkdown.
2. Push button.
3. Get PDF.

New Workflow - Write Code

```
1 ---
2 title: "Example"
3 author: "Andrew Mullins"
4 date: "November 2, 2015"
5 output: pdf_document
6 ---
7
8 Here is some cool R stuff:
9 ```{r mpg}
10 mod1 <- lm(mpg~hp, data = mtcars)
11 with(mtcars, plot(mpg~hp))
12 abline(mod1)
13 summary(mod1)
14 ```
15
16 We can do equations too.  $\sum_{i=1}^{k-1} a_{ij} a_{jk}$ .
17
18
```

New Workflow - Push Button



The screenshot shows the RStudio interface with two tabs: 'presentation.Rmd' and 'example.Rmd'. The 'presentation.Rmd' tab is active, displaying a document with a title, author, date, output format, and a block of R code. A red arrow points to the 'Knit PDF' button in the toolbar. The status bar at the bottom shows '8:27', '(Top Level)', and 'R Markdown'.

```
1 ---
2 title: "Example"
3 author: "Andrew Mullins"
4 date: "November 2, 2015"
5 output: pdf_document
6 ---
7
8 Here is some cool R stuff:
9 ```{r mpg}
10 mod1 <- lm(mpg~hp, data = mtcars)
11 with(mtcars, plot(mpg~hp))
12 abline(mod1)
13 summary(mod1)
14 ```
15
16 We can do equations too.  $\sum_{i=1}^{k-1} a_{ij} a_{jk}$ .
17
18
```

New Workflow - Get PDF 1

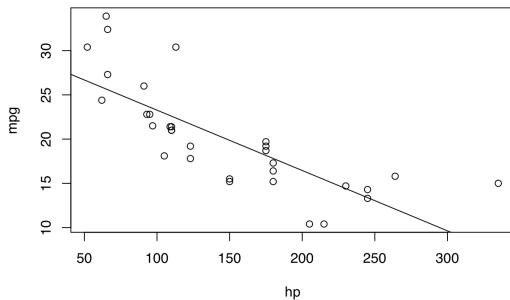
Example

Andrew Mullins

November 2, 2015

Here is some cool R stuff:

```
mod1 <- lm(mpg~hp, data = mtcars)
with(mtcars, plot(mpg~hp))
abline(mod1)
```



New Workflow - Get PDF 2

```
summary(mod1)
```

```
##  
## Call:  
## lm(formula = mpg ~ hp, data = mtcars)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -5.7121 -2.1122 -0.8854  1.5819  8.2360   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)
```

1

```
## (Intercept) 30.09886    1.63392   18.421 < 2e-16 ***  
## hp          -0.06823    0.01012   -6.742 1.79e-07 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.863 on 30 degrees of freedom  
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892   
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

New Workflow - Get PDF 3

We can do equations too:

$$\sum_{i=1}^{k-1} a_{ij} a_{jk}$$

.

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

New Workflow - Make a change

Just change it in one place and press button.

Other Tools

Sweave

Problem: Sometimes we need the fine-tuned control you can get with Latex.

Solution: Sweave let's use get both the power of Latex while still being able to use in-line R code.

Sweave Example Code

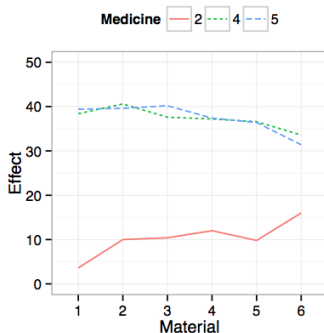
```
202 In the ANOVA table above, there is evidence that the effect of the medicine-material
    interaction is different between the two species. Here are two plots visualizing
    this relationship:
203
204 \begin{figure}[H]
205 \caption{\textbf{Interaction Plots}}
206 <<intplot1, echo=FALSE, fig.height = 4>>=
207 # Plot second-order interaction for each species
208 species1clean <- species1 %>%
209   filter(!(Medicine %in% c(1,3)))
210
211 species2clean <- species2 %>%
212   filter(!(Medicine %in% c(1,3)))
213 intp1 <- ggplot(species1clean, aes(x=Material, y = Effect,
214                                   group = Medicine,
215                                   color = Medicine,
216                                   lty = Medicine)) +
217   stat_summary(fun.y = mean, geom = "line") +
218   ggtitle("Material-Medicine (Species 1)") +
219   ylim(0,50) + theme_bw() + theme(legend.position = "top")
```

Sweave Example Output

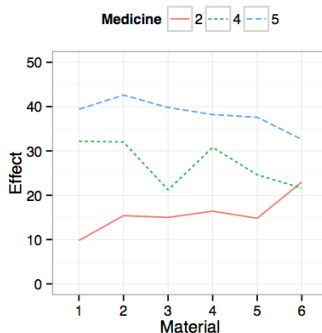
In the ANOVA table above, there is evidence that the effect of the medicine-material interaction is different between the two species. Here are two plots visualizing this relationship:

Interaction Plots

Material–Medicine (Species 1)



Material–Medicine (Species 2)



xtable/stargazer

Problem: R output is ugly.

Solution: xtable/stargazer creates Markdown/Latex code that is inserted into your document as-is.

xtable Example Code

```
188 \section*{Analysis}
189 Having prepared the data for analysis as described above, I fit a three-way ANOVA
    model with interactions. Here is the resulting ANOVA table:
190 <<anova1, echo=FALSE, results='asis'>>=
191 # Three-way ANOVA
192 infecmod <- lm(Effect~Medicine*Material*Species, data = infectedClean)
193 print(xtable(summary(aov(infecmod)),
194           caption = c("Three-way ANOVA")), table.placement = "H",
195           caption.placement = "top")
196 @
```

xtable Example Output

Three-way ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medicine	2	20373.03	10186.52	1580.67	0.0000
Material	5	291.80	58.36	9.06	0.0000
Species	1	74.76	74.76	11.60	0.0009
Medicine:Material	10	1686.17	168.62	26.16	0.0000
Medicine:Species	2	1963.14	981.57	152.31	0.0000
Material:Species	5	90.58	18.12	2.81	0.0187
Medicine:Material:Species	10	134.32	13.43	2.08	0.0293
Residuals	144	928.00	6.44		

stargazer Example Code

```
\section*{Appendix - Tables and Charts}
\label{effectstats}
<<effectstats, echo=FALSE, results='asis'>>=
# Summary statistics for response
stargazer(infected, keep = c(5), title = c("Summary of Effect Variable"),
          table.placement = "H")
@
```

stargazer Example Output

Appendix - Tables and Charts

Summary of Effect Variable

Statistic	N	Mean	St. Dev.	Min	Max
Effect	300	17.743	16.046	0	44

Problem: A lot of common R objects have bizarre inner structure.

broom - Example: lm

Call:

```
lm(formula = mpg ~ wt + qsec, data = mtcars)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.3962	-2.1431	-0.2129	1.4915	5.7486

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	19.7462	5.2521	3.760	0.000765	***
wt	-5.0480	0.4840	-10.430	2.52e-11	***
qsec	0.9292	0.2650	3.506	0.001500	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.596 on 29 degrees of freedom

Multiple R-squared: 0.8264, Adjusted R-squared: 0.8144

F-statistic: 69.03 on 2 and 29 DF, p-value: 9.395e-12

broom - Example: lm

```
> str(mod2)
List of 12
 $ coefficients : Named num [1:3] 19.746 -5.048 0.929
   ..- attr(*, "names")= chr [1:3] "(Intercept)" "wt" "qsec"
 $ residuals    : Named num [1:32] -0.8151 -0.0482 -2.5273 -0.1806 0.5039 ...
   ..- attr(*, "names")= chr [1:32] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710" "Hornet 4 Drive"
 ...
 $ effects      : Named num [1:32] -113.65 -29.116 -9.103 0.357 0.503 ...
   ..- attr(*, "names")= chr [1:32] "(Intercept)" "wt" "qsec" "" ...
 $ rank         : int 3
 $ fitted.values: Named num [1:32] 21.8 21 25.3 21.6 18.2 ...
   ..- attr(*, "names")= chr [1:32] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710" "Hornet 4 Drive"
 ...
 $ assign       : int [1:3] 0 1 2
 $ qr           :List of 5
   ..$ qr       : num [1:32, 1:3] -5.657 0.177 0.177 0.177 0.177 ...
   .. ..- attr(*, "dimnames")=List of 2
   .. .. ..$ : chr [1:32] "Mazda RX4" "Mazda RX4 Wag" "Datsun 710" "Hornet 4 Drive" ...
   .. .. ..$ : chr [1:3] "(Intercept)" "wt" "qsec"
   .. ..- attr(*, "assign")= int [1:3] 0 1 2
```

broom - Example: lm

```
library(broom)
tidy(mod2)[,1:4]
```

##	term	estimate	std.error	statistic
## 1	(Intercept)	19.746223	5.2520617	3.759709
## 2	wt	-5.047982	0.4839974	-10.429771
## 3	qsec	0.929198	0.2650173	3.506179

broom - Supports

- ▶ `lm`, `glm`, `htest`, `anova`, `nls`, `kmeans`, `manova`
- ▶ `TukeyHSD`, `arima`, `lme4`, `glmnet`, `boot`, `gam`
- ▶ `survival`, `lfe`, `zoo`, `multcomp`, `sp`, `maps`, ...

Learning More

Examples

- ▶ These slides were created in RMarkdown
- ▶ Sweave file used for Spring 2015 QEM Submission
- ▶ <https://github.com/andrewjlm/jrcolloquium>