RMarkdown Tutorial

Jill Riddell

2018-11-28

## Introduction and Purpose

The purpose of R Markdown is to generate easily readable and executable reports of code that make it simple and straightforward for others to reproduce your work. R Markdown files (.RMD) can be saved as html files, PDF’s, slideshows, and MS Word documents among others. These formats have different uses depending on the audience.

### Uses of Different File Types

1. html output and .Rmd The html output and .Rmd file is often most useful when sharing code with other scientists and researchers; it allows others to easily replicate, modify, and/or execute the code. Consider using this format for data reproducibility and accountability. This form can also act as your “digital lab notebook” and be a record of your data manipulation.
2. Word documents and PDF’s These output formats can be most useful for teaching or presenting results. For example, a scientist working for a consulting firm may export their coding results for long-term environmental monitoring data as a PDF because the client only cares about the results, not necessarily the code used to obtain the results. However, in the case of R workshops and classes, word documents may be more useful as it allows students to take notes more efficiently.

## Opening R Markdown

To open a new .Rmd file open R studio and click File -> New File -> R Markdown; an untitled .Rmd file will be opened and the following example and introduction appear:

## R Markdown

This is an R Markdown document. 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>.

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

## Including Plots

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.

### Code Chunking

In R markdown code is written in chunks. This allows the author or other user to run sections of code from the file. Chunks are usually named according to the output of the code and to make it easy to navigate and use different parts of a code. Chunks can be inserted into a .Rmd filein three ways:

1. By manually typing {r} and closing with
2. Using the keyboard shortcut cmd/ctrl + shift + enter
3. Or using the “insert” button in the editor toolbar.

Several options exist for code chunking and it’s output in the final file. Some examples include: eval = FALSE, which prevents the code from being run and is useful for displaying examples; message = FALSE, which does not allow error messages or warnings to appear; echo = FALSE, only allows the results to be printed in the final final; include = FALSE, this runs the code but doesn’t show it in the final output.

## An Example

The following code is to evaluate a statistical model for diel behavior of pH, temperature, and dissolved oxygen at a coal mine drainage stream.

First, install necessary packages and read in the data.

library(car) #install packages  
library(devtools)  
library(easyGgplot2)

## Loading required package: ggplot2

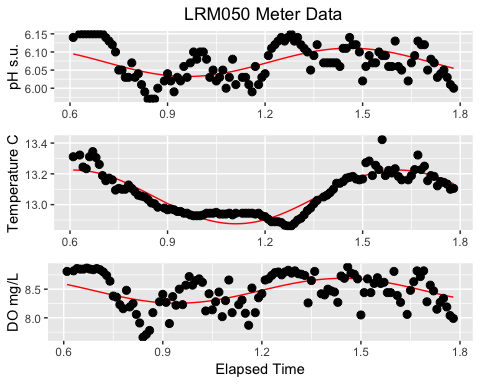
LRM050\_1 <- read.csv("LRM\_050\_1\_summer\_raw\_data.csv") #read in csv

Next, a fourier series will be added to the date to give a prediction of each value at time t and add these calculated values to the data frame

pH050 <- lm(pH ~ sin(2\*pi\*DateTimeFormatExt)+cos(2\*pi\*DateTimeFormatExt),data=LRM050\_1) #calculates a fitted value for pH at time t  
pHfitted <- pH050$fitted.values  
#creates a vector of the fitted values  
LRM050\_1$pHfitted <- pHfitted  
#adds the vector to the data  
  
Temp050 <- lm(Temp ~ sin(2\*pi\*DateTimeFormatExt)+cos(2\*pi\*DateTimeFormatExt),data=LRM050\_1)  
Tempfitted <- Temp050$fitted.values  
LRM050\_1$Tempfitted <-Tempfitted  
  
ODO050 <- lm(ODO ~ sin(2\*pi\*DateTimeFormatExt)+cos(2\*pi\*DateTimeFormatExt),data=LRM050\_1)  
ODOfitted <- ODO050$fitted.values  
LRM050\_1$ODOfitted <- ODOfitted

Now, create plots that compare our raw data to our cosine model.

pH050plot <- ggplot(LRM050\_1, aes(x=DateTimeFormatExt, y=pHfitted))+  
 geom\_line(col="red",lwd=0.5)+  
 geom\_point(aes(y=pH), size=2.5)+  
 ggtitle("LRM050 Meter Data")+  
 theme(plot.title = element\_text(hjust = 0.5))+  
 labs(x=NULL, y="pH s.u.")  
  
Temp050plot <- ggplot(LRM050\_1, aes(x=DateTimeFormatExt, y=Tempfitted))+  
 geom\_line(col="red",lwd=0.5)+  
 geom\_point(aes(y=Temp), size=2.5)+  
 labs(x=NULL, y="Temperature C")  
  
ODO050plot <- ggplot(LRM050\_1, aes(x=DateTimeFormatExt, y=ODOfitted))+  
 geom\_line(col="red",lwd=0.5)+  
 geom\_point(aes(y=ODO), size=2.5)+  
 labs(x="Elapsed Time", y= "DO mg/L")  
  
ggplot2.multiplot(pH050plot,Temp050plot,ODO050plot, cols=1)



Finally, evaluate the statistical significance of the model to the data. Is this model a good predictor for our data?

summary(pH050)

##   
## Call:  
## lm(formula = pH ~ sin(2 \* pi \* DateTimeFormatExt) + cos(2 \* pi \*   
## DateTimeFormatExt), data = LRM050\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.089289 -0.036325 -0.008658 0.037541 0.075907   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.071429 0.004018 1511.198 < 2e-16 \*\*\*  
## sin(2 \* pi \* DateTimeFormatExt) 0.009500 0.005422 1.752 0.0825 .   
## cos(2 \* pi \* DateTimeFormatExt) -0.037860 0.005899 -6.418 3.6e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.04192 on 110 degrees of freedom  
## Multiple R-squared: 0.2812, Adjusted R-squared: 0.2681   
## F-statistic: 21.52 on 2 and 110 DF, p-value: 1.297e-08

outlierTest(pH050)

##   
## No Studentized residuals with Bonferonni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferonni p  
## 86 -2.197224 0.030121 NA

summary(Temp050)

##   
## Call:  
## lm(formula = Temp ~ sin(2 \* pi \* DateTimeFormatExt) + cos(2 \*   
## pi \* DateTimeFormatExt), data = LRM050\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.10491 -0.03826 -0.01069 0.03061 0.20590   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.050268 0.005157 2530.69 <2e-16 \*\*\*  
## sin(2 \* pi \* DateTimeFormatExt) -0.111416 0.006960 -16.01 <2e-16 \*\*\*  
## cos(2 \* pi \* DateTimeFormatExt) -0.134487 0.007572 -17.76 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0538 on 110 degrees of freedom  
## Multiple R-squared: 0.847, Adjusted R-squared: 0.8442   
## F-statistic: 304.5 on 2 and 110 DF, p-value: < 2.2e-16

outlierTest(Temp050)

## rstudent unadjusted p-value Bonferonni p  
## 92 4.151142 6.5868e-05 0.0074431

summary(ODO050)

##   
## Call:  
## lm(formula = ODO ~ sin(2 \* pi \* DateTimeFormatExt) + cos(2 \*   
## pi \* DateTimeFormatExt), data = LRM050\_1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.63944 -0.18784 0.00129 0.21903 0.45940   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.47194 0.02533 334.508 < 2e-16 \*\*\*  
## sin(2 \* pi \* DateTimeFormatExt) 0.07514 0.03418 2.198 0.03 \*   
## cos(2 \* pi \* DateTimeFormatExt) -0.20352 0.03719 -5.473 2.82e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2642 on 110 degrees of freedom  
## Multiple R-squared: 0.2331, Adjusted R-squared: 0.2191   
## F-statistic: 16.71 on 2 and 110 DF, p-value: 4.588e-07

outlierTest(ODO050)

##   
## No Studentized residuals with Bonferonni p < 0.05  
## Largest |rstudent|:  
## rstudent unadjusted p-value Bonferonni p  
## 23 -2.508789 0.013586 NA

### Knitting

Knitting is the file step in producing the final report. This sends the R markdown file to knitr which completes all executables in the markdown and then the selected output is processed by pandoc into the file type of the users choosing.

### Some notes

If this above code were going to be presented to a client or at a conference, it may be pertinent to include echol = FALSE and to knit the product as a PDF. For teaching purposes an html or word document may be more appropriate.

## References and important links

1. <https://r3ds.had.co.nz/r-markdown.html>
2. <https://rmarkdown.rstudio.com>
3. <https://bookdown.org/yihui/rmarkdown/>