17: Crafting Reports

Environmental Data Analytics | Kateri Salk

Spring 2019

LESSON OBJECTIVES

1. Describe the purpose of using R Markdown as a communication and workflow tool
2. Incorporate Markdown syntax into documents
3. Communicate the process and findings of an analysis session in the style of a report

BASIC R MARKDOWN DOCUMENT STRUCTURE

1. **YAML Header** surrounded by — on top and bottom
    - YAML templates include options for html, pdf, word, markdown, and interactive
    - More information on formatting the YAML header can be found in the cheat sheet
2. **R Code Chunks** surrounded by "`on top and bottom + Create using`Cmd/Ctrl+Alt+I'
    - Can be named {r name} to facilitate navigation and autoreferencing
    - Chunk options allow for flexibility when the code runs and when the document is knitted
3. **Text** with formatting options for readability in knitted document

A handy cheat sheet for R markdown can be found here. Another one can be found here.

WHY R MARKDOWN?

<Fill in our discussion below with bullet points. Use italics and bold for emphasis (hint: use the cheat sheets to figure out how to make bold and italic text).>

- *italicize* using asterisks surrounding a word
- **bold** using two asterisks on each side of a word or phrase
- Documentation of what you are doing
- Allows for text, graphics, and code all in one clean(ish) document
- Control over what is displayed
- Easy to share/distribute
- Allows for more organization

TEXT EDITING CHALLENGE

Create a table below that details the example datasets we have been using in class. The first column should contain the names of the datasets and the second column should include some relevant information about the datasets. (Hint: use the cheat sheets to figure out how to make a table in Rmd)

Dataset Name	Description
EPA Air	Air quality data
Leaf Litter	Litter collection data
LTER Lake Data	Lake data from NTL

R CHUNK EDITING CHALLENGE

Installing packages

Create an R chunk below that installs the package `knitr`. Instead of commenting out the code, customize the chunk options such that the code is not evaluated (i.e., not run).

```
install.packages("knitr")
```

Setup

Create an R chunk below called "setup" that checks your working directory, loads the packages `tidyverse` and `knitr`, and sets a ggplot theme. Remember that you need to disable R throwing a message, which containts a check mark that cannot be knitted.

Load the NTL-LTER_Lake_Nutrients_Raw dataset, display the head of the dataset, and set the date column to a date format.

Customize the chunk options such that the code is run but is not displayed in the final document.

```
## [1] "C:/Users/thoma/Thomas/2018 Grad School/Duke MEM/ENV 872/Environmental_Data_Analytics_2020"
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
## Warning: package 'tidyr' was built under R version 3.6.3
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

```
## Warning: package 'forcats' was built under R version 3.6.3
```

```
## Warning: package 'knitr' was built under R version 3.6.3
```

```
##   lakeid  lakename year4 daynum sampledate depth_id depth tn_ug tp_ug nh34 no23
## 1      L Paul Lake  1991    140    5/20/91        1  0.00   538    25   NA   NA
## 2      L Paul Lake  1991    140    5/20/91        2  0.85   285    14   NA   NA
## 3      L Paul Lake  1991    140    5/20/91        3  1.75   399    14   NA   NA
## 4      L Paul Lake  1991    140    5/20/91        4  3.00   453    14   NA   NA
## 5      L Paul Lake  1991    140    5/20/91        5  4.00   363    13   NA   NA
## 6      L Paul Lake  1991    140    5/20/91        6  6.00   583    37   NA   NA
##   po4 comments
## 1  NA
## 2  NA
## 3  NA
## 4  NA
## 5  NA
## 6  NA
```

```
## [1] "factor"
```

```
## [1] "Date"
```

Data Exploration, Wrangling, and Visualization

Create an R chunk below to create a processed dataset do the following operations:

- Include all columns except lakeid, depth_id, and comments
- Include only surface samples (depth = 0 m)

```
NTL.Nutrients.Processed <- NTL.Nutrients %>%
  select(-c("lakeid", "depth_id", "comments")) %>%
  filter(depth == 0)
```

Create a second R chunk to create a summary dataset with the mean, minimum, maximum, and standard deviation of total nitrogen concentrations for each lake. Create a second summary dataset that is identical except that it evaluates total phosphorus. Customize the chunk options such that the code is run but not displayed in the final document.

Create a third R chunk that uses the function `kable` in the knitr package to display two tables: one for the summary dataframe for total N and one for the summary dataframe of total P. Use the `caption = " "` code within that function to title your tables. Customize the chunk options such that the final table is displayed but not the code used to generate the table.

Table 2: Nitrogen Summary

lakename	meanN	minN	maxN	stdevN
Central Long Lake	675.8338	343.020	953.063	203.25838
Crampton Lake	362.6813	353.380	376.304	12.05748
East Long Lake	796.0141	299.310	3316.892	413.16241
Hummingbird Lake	1036.6695	779.053	1221.960	204.36889
Paul Lake	365.0679	45.670	628.625	107.31194
Peter Lake	548.2733	131.830	2048.151	320.83105
Tuesday Lake	410.2571	237.363	554.418	71.53021
West Long Lake	737.5649	303.170	2950.343	437.18914

Table 3: Nitrogen Summary

lakename	meanP	minP	maxP	stdevP
Central Long Lake	21.59478	8.190	37.270	7.067785
Crampton Lake	11.16033	5.803	15.555	4.946759
East Long Lake	39.16260	7.160	136.671	29.814346
Hummingbird Lake	36.21925	32.765	42.119	4.146717
Paul Lake	10.27233	0.110	36.070	4.507096
Peter Lake	21.33466	0.000	66.893	14.110973
Tuesday Lake	11.37472	4.413	18.663	3.076367
West Long Lake	22.34388	2.690	63.243	12.622100

Create a fourth and fifth R chunk that generates two plots (one in each chunk): one for total N over time with different colors for each lake, and one with the same setup but for total P. Decide which geom option will be appropriate for your purpose, and select a color palette that is visually pleasing and accessible. Customize the chunk options such that the final figures are displayed but not the code used to generate the figures. In addition, customize the chunk options such that the figures are aligned on the left side of the page. Lastly, add a fig.cap chunk option to add a caption (title) to your plot that will display underneath the figure.

```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```

Other options

What are the chunk options that will suppress the display of errors, warnings, and messages in the final document?

ANSWER: Respectively: error, warning, message

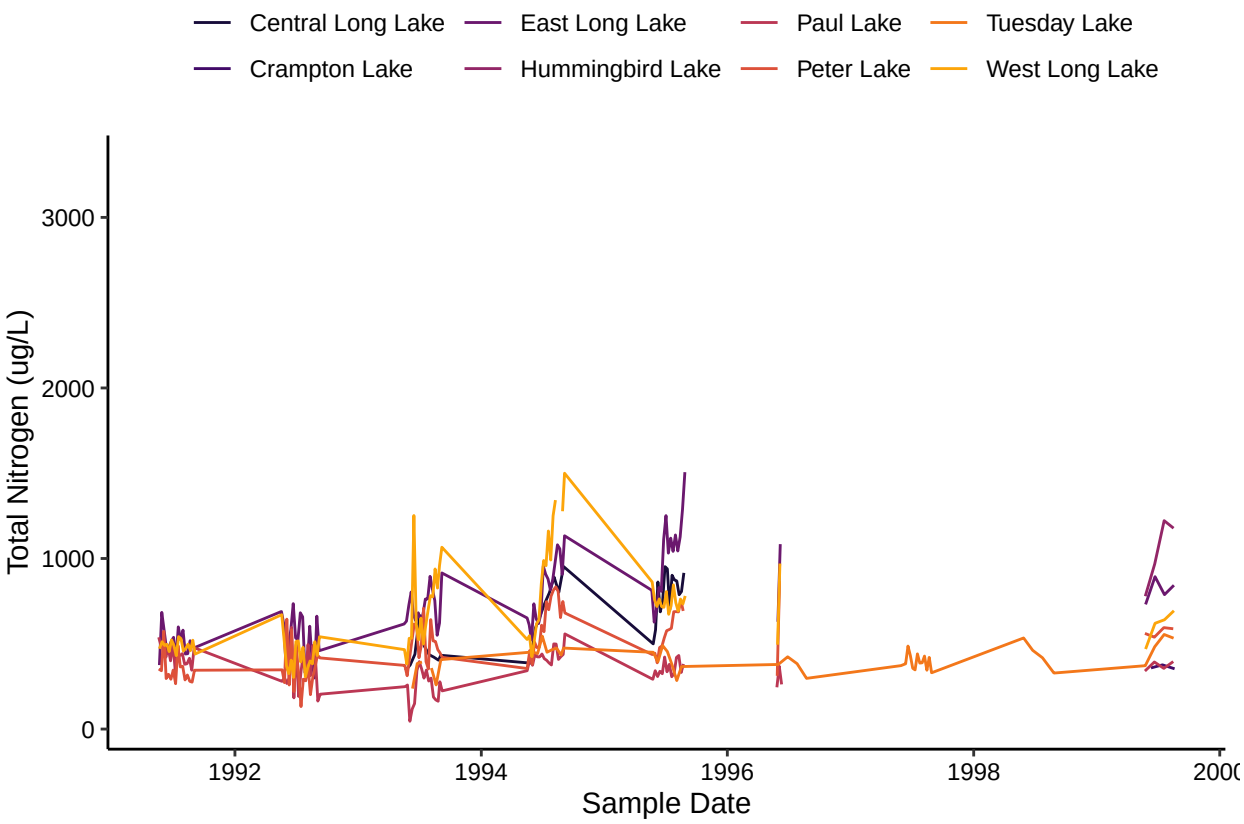

Figure 1: Total Nitrogen Over Time

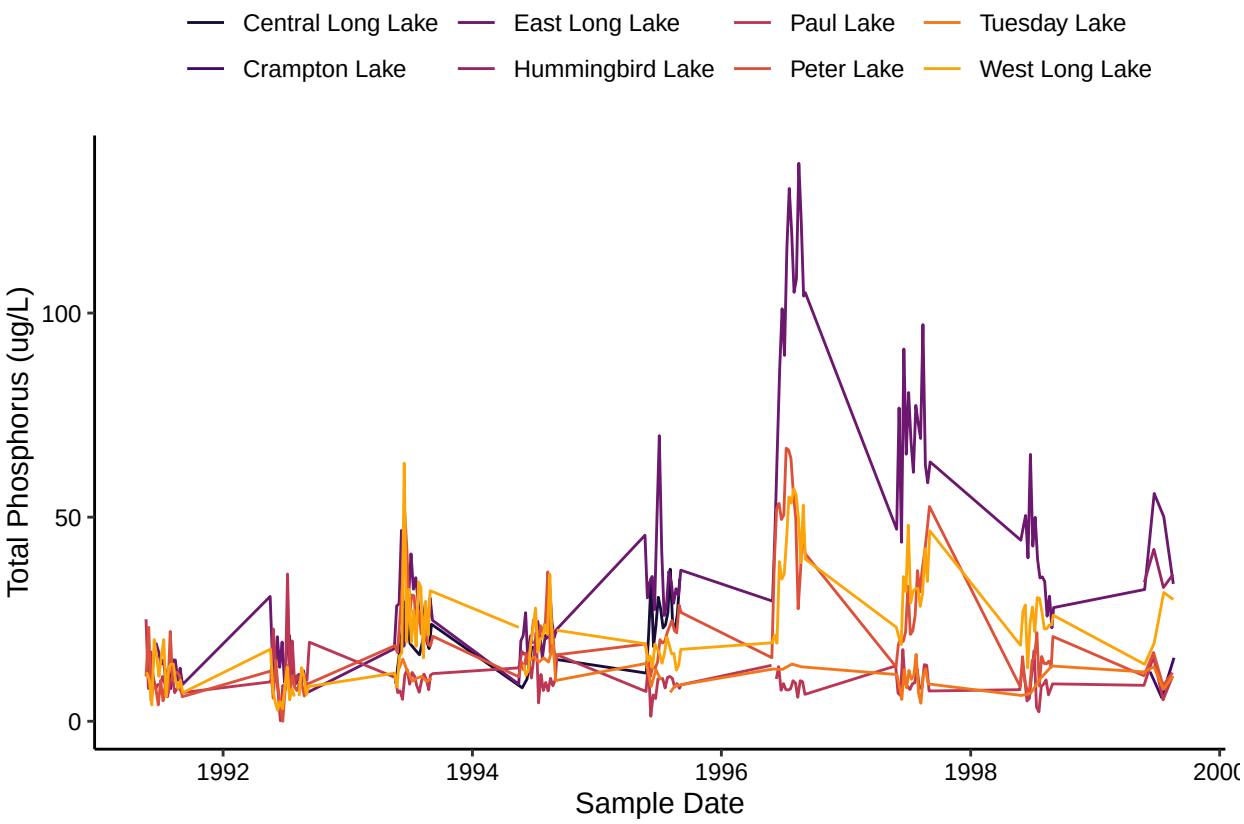

Figure 2: Total Phosphorus Over Time

Communicating results

Write a paragraph describing your findings from the R coding challenge above. This should be geared toward an educated audience but one that is not necessarily familiar with the dataset. Then insert a horizontal rule below the paragraph. Below the horizontal rule, write another paragraph describing the next steps you might take in analyzing this dataset. What questions might you be able to answer, and what analyses would you conduct to answer those questions?

This is paragraph one, describing the data set. The data has various data about lake conditions such as total nitrogen, total phosphorous, and species of each chemical. Other data including temperature and sampling conditions are also included.

This is the second paragraph, describing the next steps I would take to analyze this data. It would be useful to have a specific research question, so that is probably where I would go next. Doing some statisitical tests across the lakes or across time seem like good starting places for more exploration as well.

KNIT YOUR PDF

When you have completed the above steps, try knitting your PDF to see if all of the formatting options you specified turned out as planned. This may take some troubleshooting.

OTHER R MARKDOWN CUSTOMIZATION OPTIONS

We have covered the basics in class today, but R Markdown offers many customization options. A word of caution: customizing templates will often require more interaction with LaTeX and installations on your computer, so be ready to troubleshoot issues.

Customization options for pdf output include:

- Table of contents
- Number sections
- Control default size of figures
- Citations
- Template (more info here)

pdf_document:
toc: true
number_sections: true
fig_height: 3
fig_width: 4
citation_package: natbib
template: