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** (Yet Another Markdown Lesson) 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/Option+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; determine what shows in our document, what runs, what doesn't run
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?

- Code, output, and test/notes together in one document
- Knit to useful formats (pdf, html, docx)
- Legible code and output
- Git friendly - version control!
- Reproducible
- Updating capabilities
- Focus on output and conclusions, not code (flexible formatting)
- Simple syntax and autoreferencing (will need to autoreference figures and tables in R for final project)

TEXT EDITING CHALLENGE

Create a table below that details the example datasets we have been using in class. The first column should contain the name of the dataset and the second column should include some relevant information about the dataset.

Dataset Name	Detail about dataset
NTLER	Nutrients in lakes
Ecotox	Neonicotinoids
EPA Air	Air quality
USGS	Stream flow data
NTLER	Carbon in lakes
NTLER	Lake chemistry and physical data

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")
install.packages("viridis")
```

Setup

Create an R chunk below called "setup" that checks your working directory, loads the packages `tidyverse` and `knitr`, and sets a ggplot theme.

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.

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)

```
nutrientsprocessed <- nutrients %>%
  select(-lakeid, -depth_id, -comments) %>%  #for filter use &, for select use commas to indidcate and
  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: Total N in Lakes

lakename	mean_tn	min_tn	max_tn	std_tn
Central Long Lake	NA	NA	NA	NA
Crampton Lake	362.6813	353.380	376.304	12.05748
East Long Lake	NA	NA	NA	NA
Hummingbird Lake	1036.6695	779.053	1221.960	204.36889
Paul Lake	NA	NA	NA	NA
Peter Lake	NA	NA	NA	NA
Tuesday Lake	NA	NA	NA	NA
West Long Lake	NA	NA	NA	NA

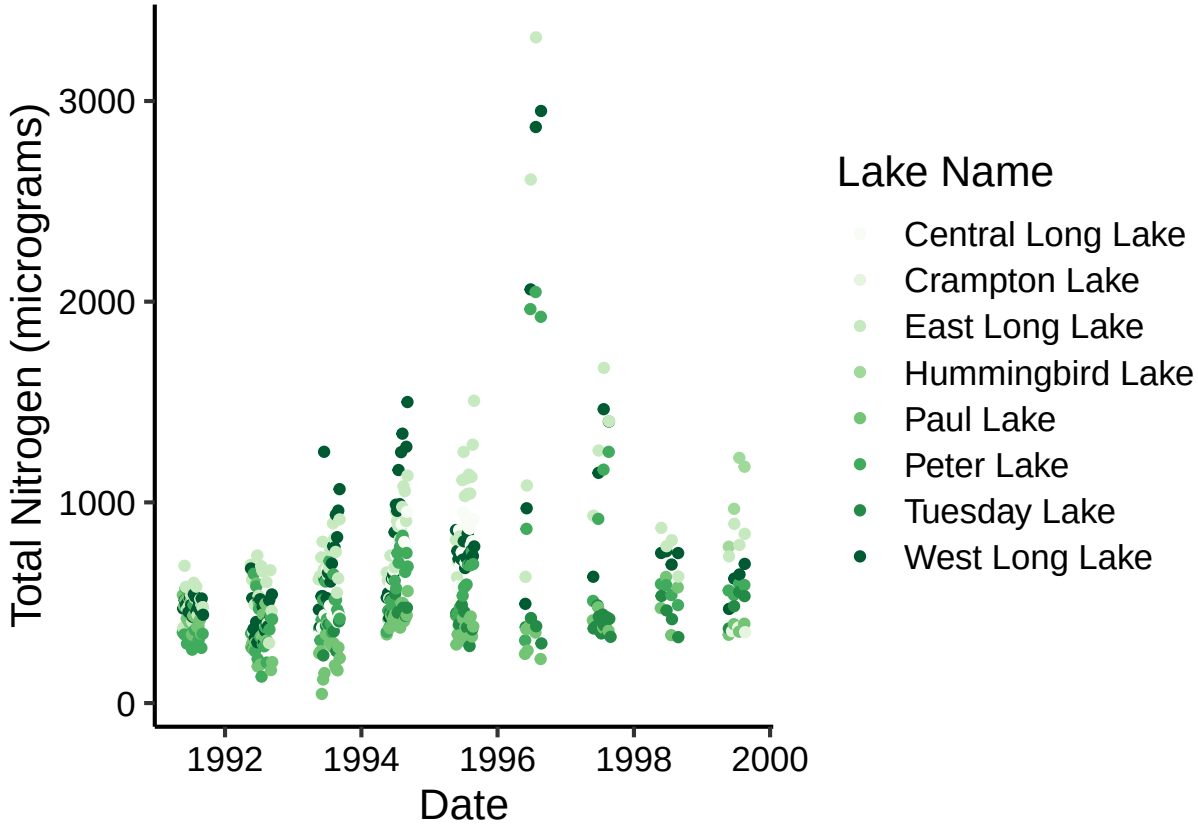

Figure 1: Total Nitrogen in Lakes

Table 3: Total P in Lakes

lakename	mean_tp	min_tp	max_tp	std_tp
Central Long Lake	21.59478	8.190	37.270	7.067785
Crampton Lake	11.16033	5.803	15.555	4.946759
East Long Lake	NA	NA	NA	NA
Hummingbird Lake	36.21925	32.765	42.119	4.146717
Paul Lake	NA	NA	NA	NA
Peter Lake	21.33466	0.000	66.893	14.110973
Tuesday Lake	NA	NA	NA	NA
West Long Lake	NA	NA	NA	NA

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.

Other options

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

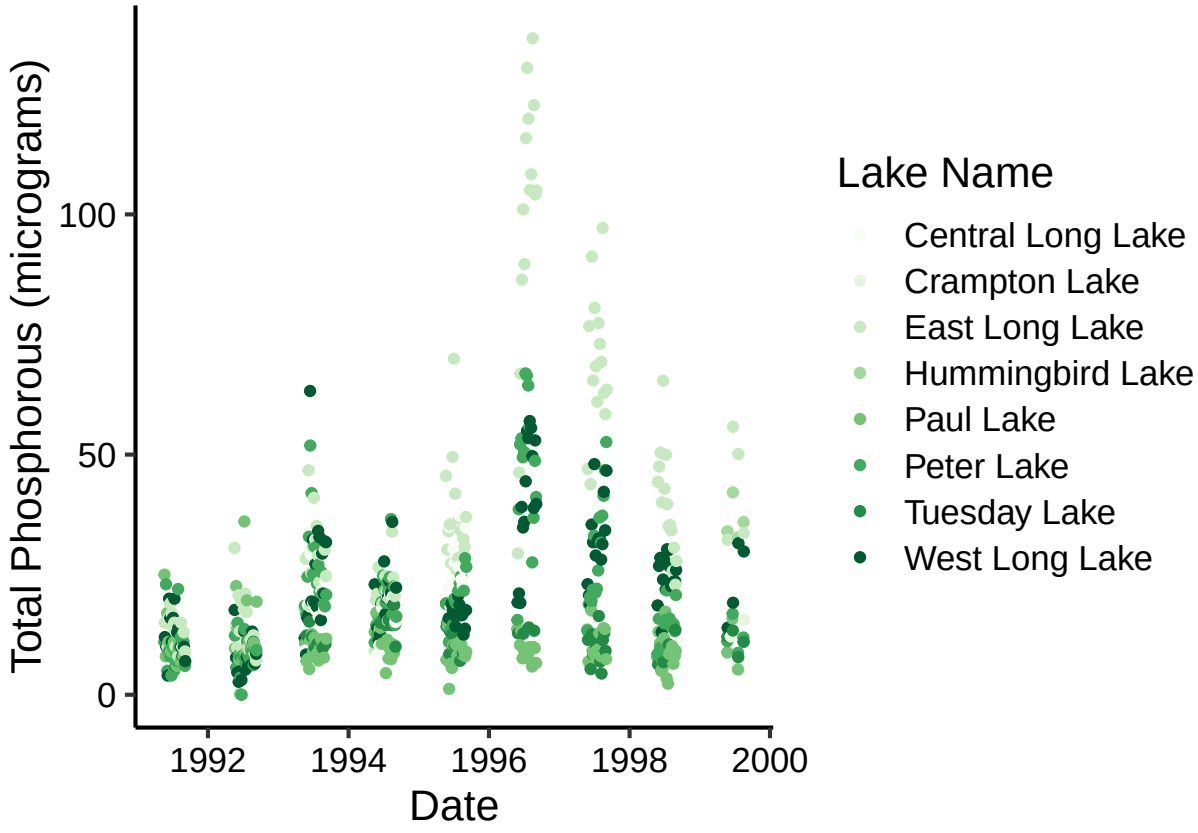

Figure 2: Total Phosphorous in Lakes

ANSWER:

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?

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: