

11: Crafting Reports

Environmental Data Analytics | John Fay & Luana Lima | Developed by Kateri Salk

Spring 2022

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

USE OF R STUDIO & R MARKDOWN SO FAR...

1. Write code
2. Document that code
3. Generate PDFs of code and its outputs
4. Integrate with Git/GitHub for version control

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

RESOURCES

Handy cheat sheets for R markdown can be found: [here](#), and [here](#).

There's also a quick reference available via the **Help**→**Markdown Quick Reference** menu.

Lastly, this website give a great & thorough overview.

THE KNITTING PROCESS

- The knitting sequence



Figure 1: knitting

- Knitting commands in code chunks:
 - `include = FALSE` - code is run, but neither code nor results appear in knitted file
 - `echo = FALSE` - code not included in knitted file, but results are
 - `eval = FALSE` - code is not run in the knitted file
 - `message = FALSE` - messages do not appear in knitted file
 - `warning = FALSE` - warnings do not appear...
 - `fig.cap = "..."` - adds a caption to graphical results

WHAT ELSE CAN R MARKDOWN DO?

See: <https://rmarkdown.rstudio.com> and class recording. * Languages other than R... * Various outputs...

WHY R MARKDOWN?

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

- enables organization of rcode
- different coding languages
- can hide/show code- managing how products are is displayed
- add **formatted** documentation along with code
- everything is just code/text, so *reproducibility* is easy

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 title	Summary
ECOTOX	data from U.S. EPA ECOTOX Knowledgebase website
EPAair	data from U.S. EPA air quality monitoring of PM2.5 and ozone in North Carolina in 2017 and 2018
NEON_NIWO_Litter	litterfall and woody debris data (needles, leaves, twigs, lichen, etc)
NTL-LTR_Lake_Nutrients	nutrient data from studies on several lakes in the North Temperate Lakes District in Wisconsin, USA. Data were collected as part of the Long Term Ecological Research station established by the National Science Foundation.
USGS_Site02085000_Flow	U.S. geological survey streamflow data from the gage site 02085000 (Eno River at Hillsborough, NC)

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).

Setup

Create an R chunk below called “setup” that checks your working directory, loads the packages `tidyverse`, `lubridate`, and `knitr`, and sets a ggplot theme. Remember that you need to disable R throwing a message, which contains 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.

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)
- Drop rows with missing data

```
Lake_Nutrients_processed <- Lake_Nutrients_Raw %>%  
  select(lakename:sampleddate, depth:po4, Date) %>%  
  filter(depth==0) %>%  
  drop_na()
```

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 Table

lakename	meanNitrogen	stdevNitrogen	minNitrogen	maxNitrogen
Central Long Lake	610.9982	333.8124	45.67	2870.302
Crampton Lake	610.9982	333.8124	45.67	2870.302
East Long Lake	610.9982	333.8124	45.67	2870.302
Hummingbird Lake	610.9982	333.8124	45.67	2870.302
Paul Lake	610.9982	333.8124	45.67	2870.302
Peter Lake	610.9982	333.8124	45.67	2870.302
Tuesday Lake	610.9982	333.8124	45.67	2870.302
West Long Lake	610.9982	333.8124	45.67	2870.302

Table 3: Phosphorus Summary Table

lakename	meanPhos	stdevPhos	minPhos	maxPhos
Central Long Lake	18.96872	12.72575	0	101.05
Crampton Lake	18.96872	12.72575	0	101.05

lakename	meanPhos	stdevPhos	minPhos	maxPhos
East Long Lake	18.96872	12.72575	0	101.05
Hummingbird Lake	18.96872	12.72575	0	101.05
Paul Lake	18.96872	12.72575	0	101.05
Peter Lake	18.96872	12.72575	0	101.05
Tuesday Lake	18.96872	12.72575	0	101.05
West Long Lake	18.96872	12.72575	0	101.05

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.

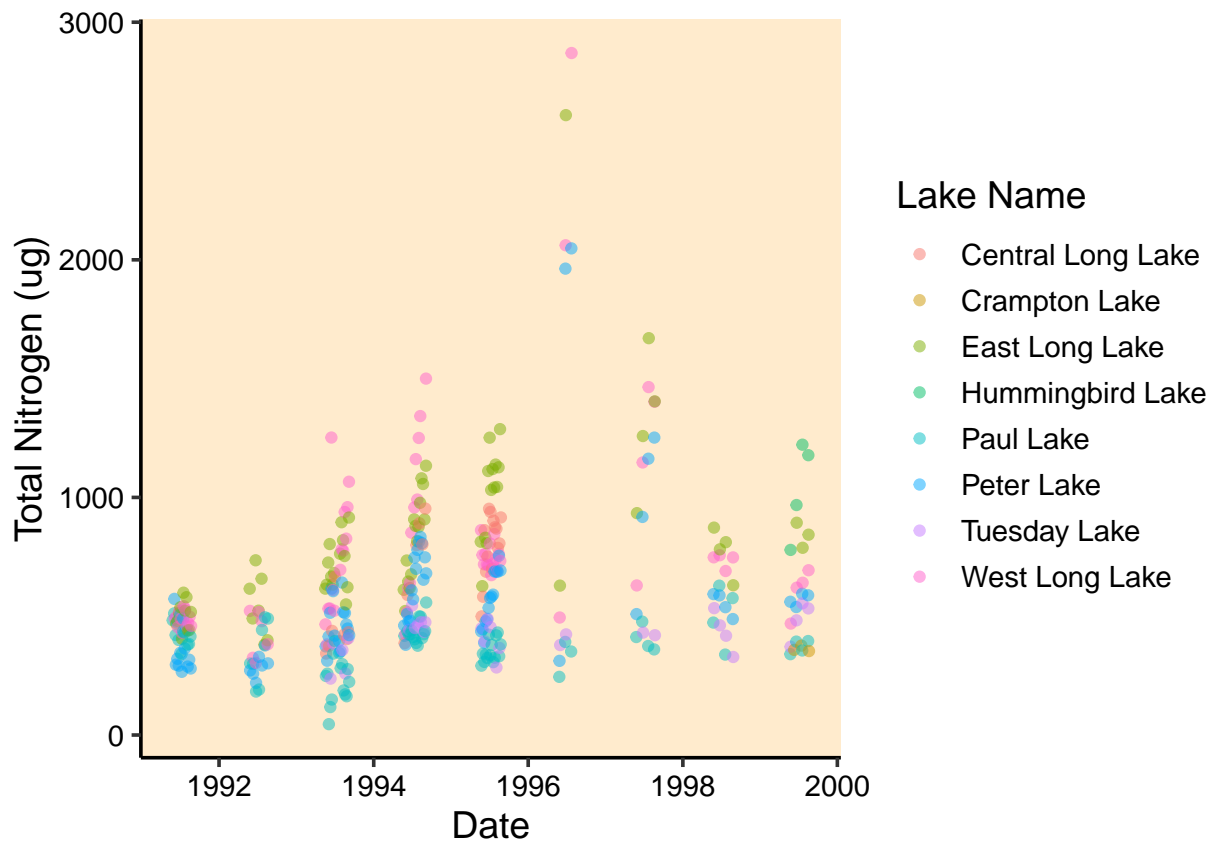


Figure 2: Total Nitrogen Over Time by Lake

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?

Answer: This data includes measurements of nutrients in a number of lakes across Wisconsin.

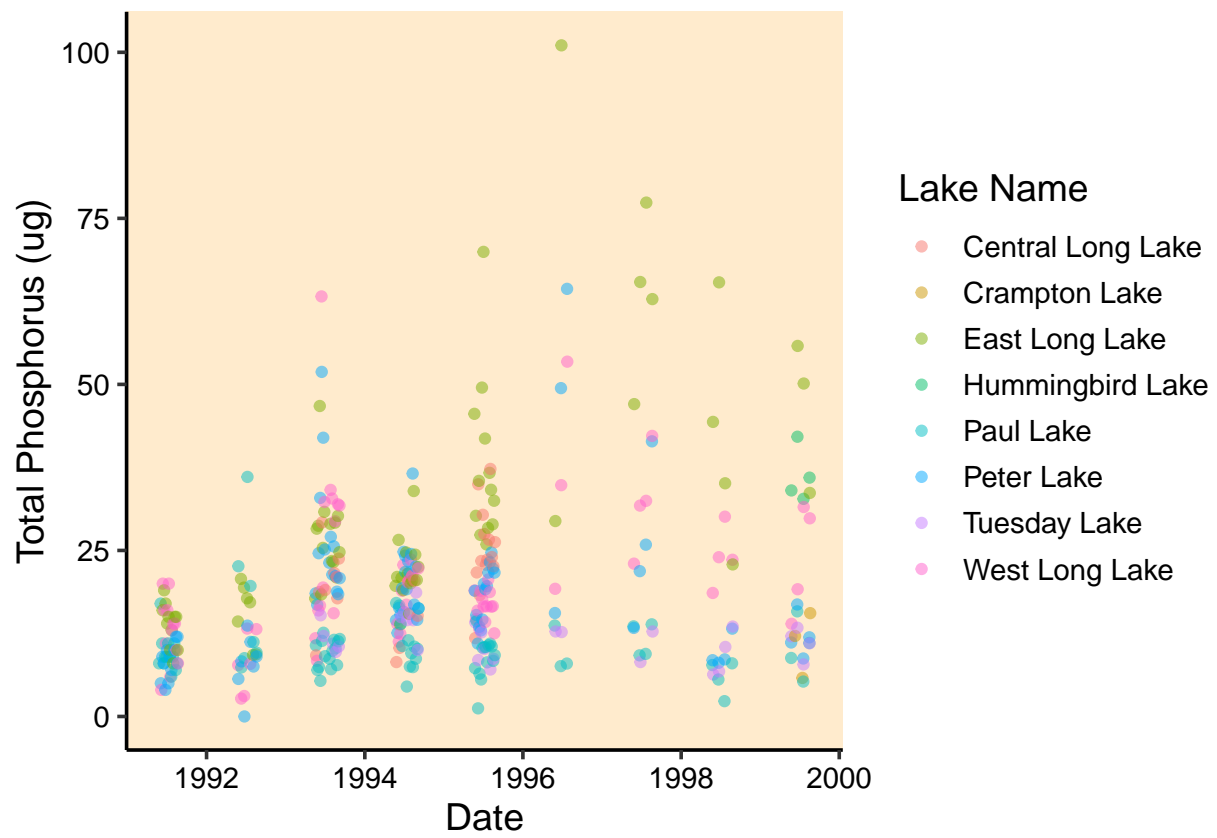


Figure 3: Total Phosphorus Over Time by Lake

The graphs show us that overall, total phosphorus is much lower than total nitrogen across all lakes (look at the scale of the y-axis). I see a bit of a hump around 1996 in both nutrients, and wonder what may have caused that (extra runoff for some reason?). Especially at East Long Lake, it appears to be an outlier in both nitrogen and phosphorus.

My next steps would include testing lines of best fit, to see what kind of model might fit the data. I might also run an ANOVA to see if there are statistically significant differences between nitrogen and phosphorus levels between lakes.

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: