

11: Crafting Reports

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

Spring 2021

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
- 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).>

- R Markdown is **FREE**
- R Markdown allows for your *code* and *results* to be documented and shared
- R Markdown converts your results to a **PDF**

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)

Data set Name	Relevant Information
Peter Paul	Temperature of Lakes
EPA Air	Air Pollutants

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

```
NTL_zero<-filter(NTL, depth == 0)
NTL_reduced<-select(NTL_zero, lakeid, depth_id, comments)
NTL_reduced2<-drop_na(NTL_reduced)
```

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.

```
##      lakename      tn_ug      sampleddate
## Length:5836      Min.   :  0.0      Min.   :1991-05-20
## Class :character  1st Qu.: 351.9      1st Qu.:1993-09-03
## Mode  :character  Median : 453.9      Median :1996-08-01
##                               Mean   : 593.3      Mean   :1999-06-04
##                               3rd Qu.: 664.9      3rd Qu.:2001-08-16
##                               Max.   :3497.7      Max.   :2016-08-17
##                               NA's   :2330

##      lakename      tp_ug      sampleddate
## Length:5836      Min.   : -6.349      Min.   :1991-05-20
## Class :character  1st Qu.: 11.000      1st Qu.:1993-09-03
## Mode  :character  Median : 18.663      Median :1996-08-01
##                               Mean   : 30.040      Mean   :1999-06-04
##                               3rd Qu.: 33.999      3rd Qu.:2001-08-16
##                               Max.   :352.056      Max.   :2016-08-17
##                               NA's   :317
```

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 Summary Stats

lakename	tn_ug	sampledate
Length:5836	Min. : 0.0	Min. :1991-05-20
Class :character	1st Qu.: 351.9	1st Qu.:1993-09-03
Mode :character	Median : 453.9	Median :1996-08-01
NA	Mean : 593.3	Mean :1999-06-04
NA	3rd Qu.: 664.9	3rd Qu.:2001-08-16
NA	Max. :3497.7	Max. :2016-08-17
NA	NA's :2330	NA

Table 3: Total P Summary Stats

lakename	tp_ug	sampledate
Length:5836	Min. : -6.349	Min. :1991-05-20
Class :character	1st Qu.: 11.000	1st Qu.:1993-09-03
Mode :character	Median : 18.663	Median :1996-08-01
NA	Mean : 30.040	Mean :1999-06-04
NA	3rd Qu.: 33.999	3rd Qu.:2001-08-16
NA	Max. :352.056	Max. :2016-08-17

lakename	tp_ug	sampleddate
NA	NA's :317	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.

```
## Warning: Removed 2330 rows containing missing values (geom_point).
```

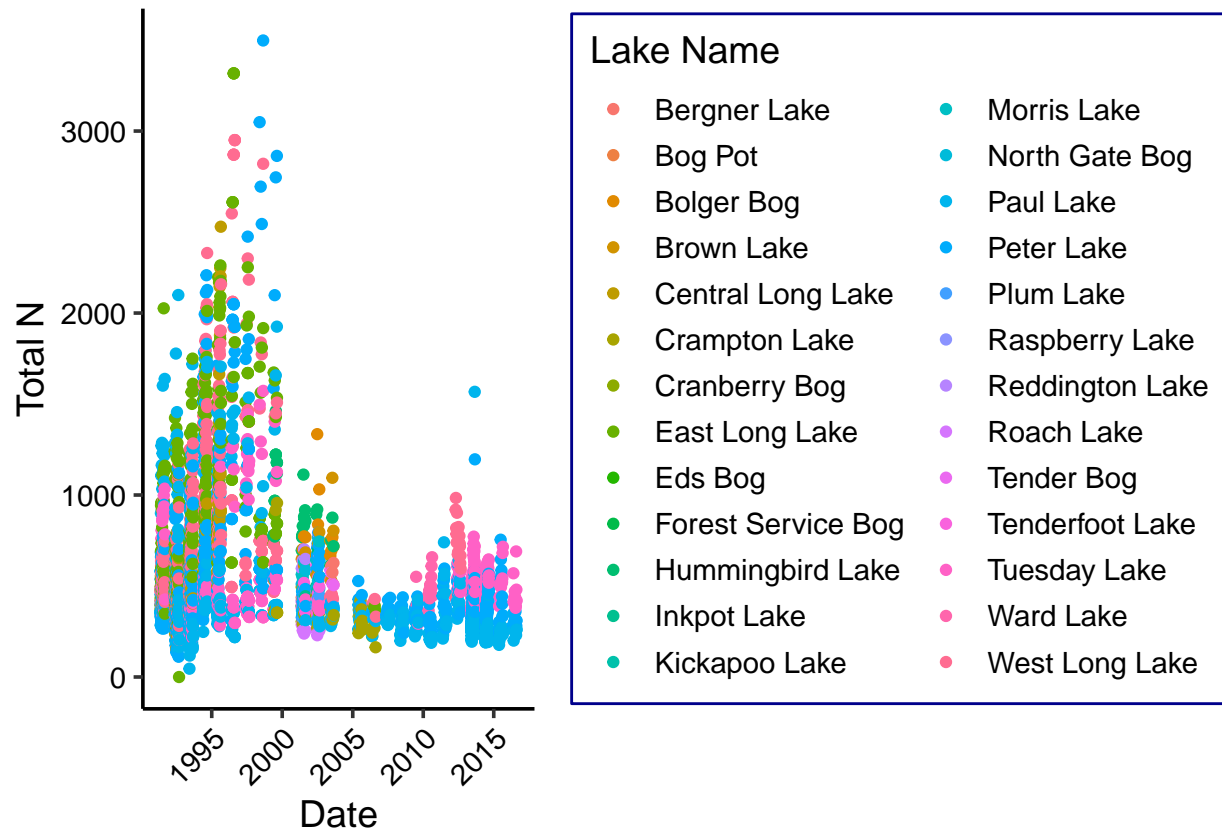


Figure 1: Total N per Lake

```
## Warning: Removed 317 rows containing missing values (geom_point).
```

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?

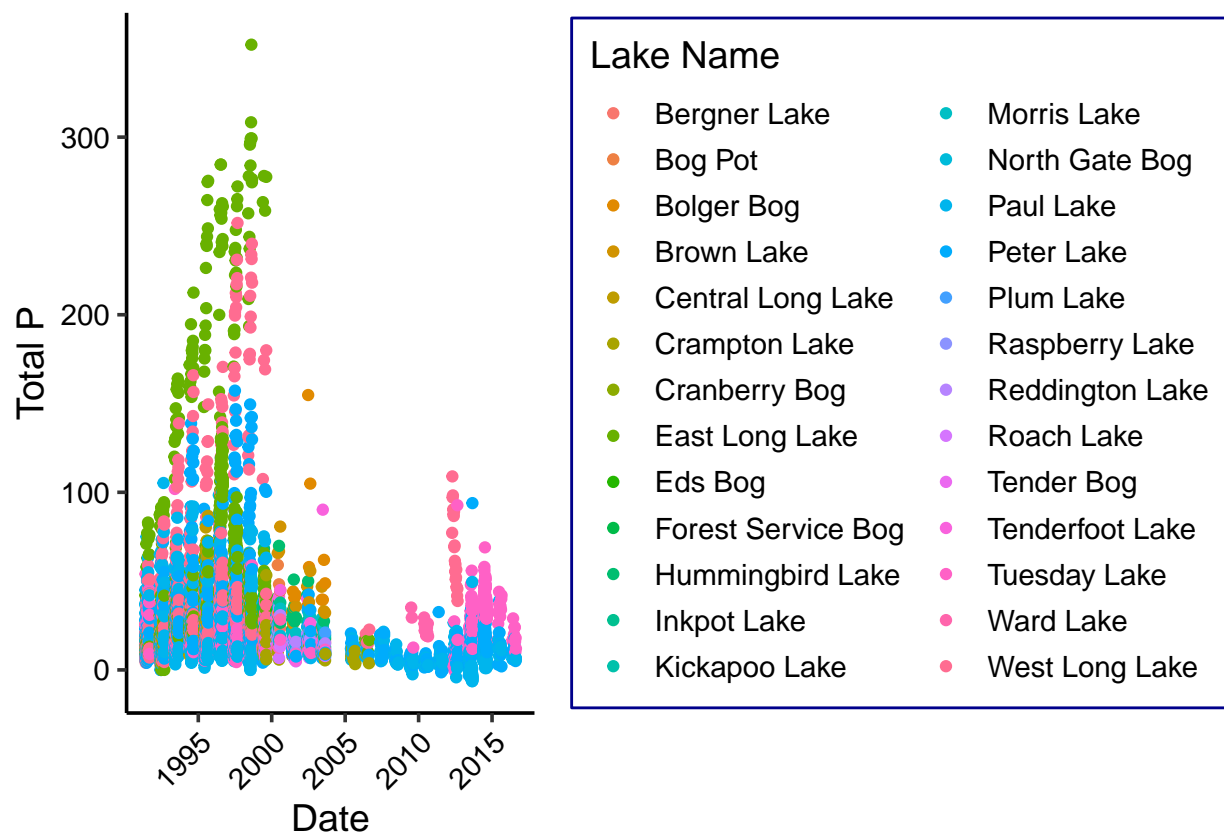


Figure 2: Total P per Lake

Results: The analysis done above show the the total nitrogen and phosphorous levels of each lake in our data set. Once again, the data set consisted of surface level nitrogen and phosphorous values found in lakes over a twenty year time span. From our graphs, we can tell that the total nitrogen and phosphorous levels have decreased in the past twenty years.

Future studeis: Moving forward, it would also be important to see how nitrogen and phosphorous levels changed over the past twenty years as we decreased in depth. This would be important for bottomw dwelling species found in lakes. Another question to look at would be, are there lakes that increase in nitrogen and phosphorous levels over time? Our graphs indicated that most levels decreased but if we look at lakes individually we might find lakes that increased over the past twenty years. If there are examples of this, more research could be done to hypthosize why this may be the case. Are these lakes close to agriculture or cities?

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:
```