

# 11: Crafting Reports

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

Spring 2022

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

## LESSON OBJECTIVES

### 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

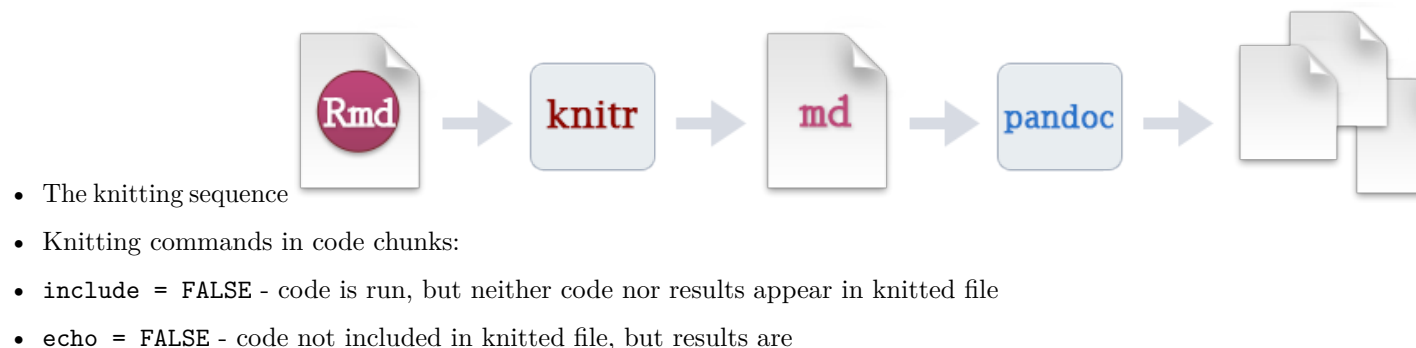
## RESOURCES

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



- `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).>

## 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)

## 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.

```
## [1] "C:/Users/curtx/Desktop/Environmental Data Analytic/Environmental_Data_Analytics_2022/Lessons"
## Warning: package 'pacman' was built under R version 4.1.3
## Rows: 5836 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (4): lakeid, lakenname, sampleddate, comments
## dbl (9): year4, daynum, depth_id, depth, tn_ug, tp_ug, nh34, no23, po4
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## # A tibble: 6 x 13
##   lakeid lakenname year4 daynum sampleddate depth_id depth tn_ug tp_ug nh34 no23
##   <chr>   <chr>    <dbl> <dbl> <date>         <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1 L      Paul La~ 1991    140 1991-05-20      1 0      538    25    NA    NA
## 2 L      Paul La~ 1991    140 1991-05-20      2 0.85    285    14    NA    NA
## 3 L      Paul La~ 1991    140 1991-05-20      3 1.75    399    14    NA    NA
## 4 L      Paul La~ 1991    140 1991-05-20      4 3      453    14    NA    NA
## 5 L      Paul La~ 1991    140 1991-05-20      5 4      363    13    NA    NA
## 6 L      Paul La~ 1991    140 1991-05-20      6 6      583    37    NA    NA
## # ... with 2 more variables: po4 <dbl>, comments <chr>
```

## 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

```
data_clean <- NTL_LTER_Lake_Nutrients_Raw %>%
  select(-lakeid, -depth_id, -comments) %>%
  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 1: Summary Statistics for Nitrogen Concentrations

Mean	Min	Max	SD
610.9982	45.67	2870.302	333.8124

Table 2: Summary Statistics for Phosphorous Concentrations

Mean	Min	Max	SD
18.96872	0	101.05	12.72575

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.

## 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

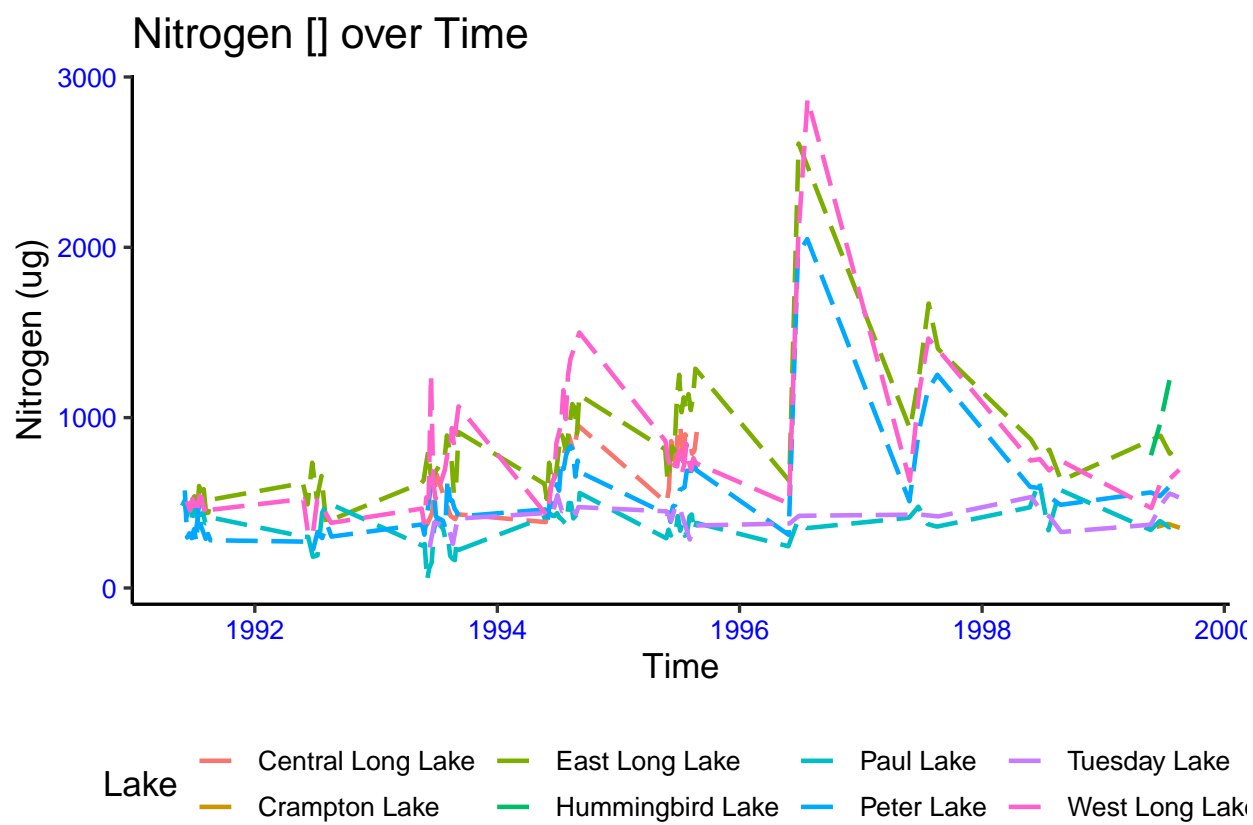


Figure 1: N (ug) over Time

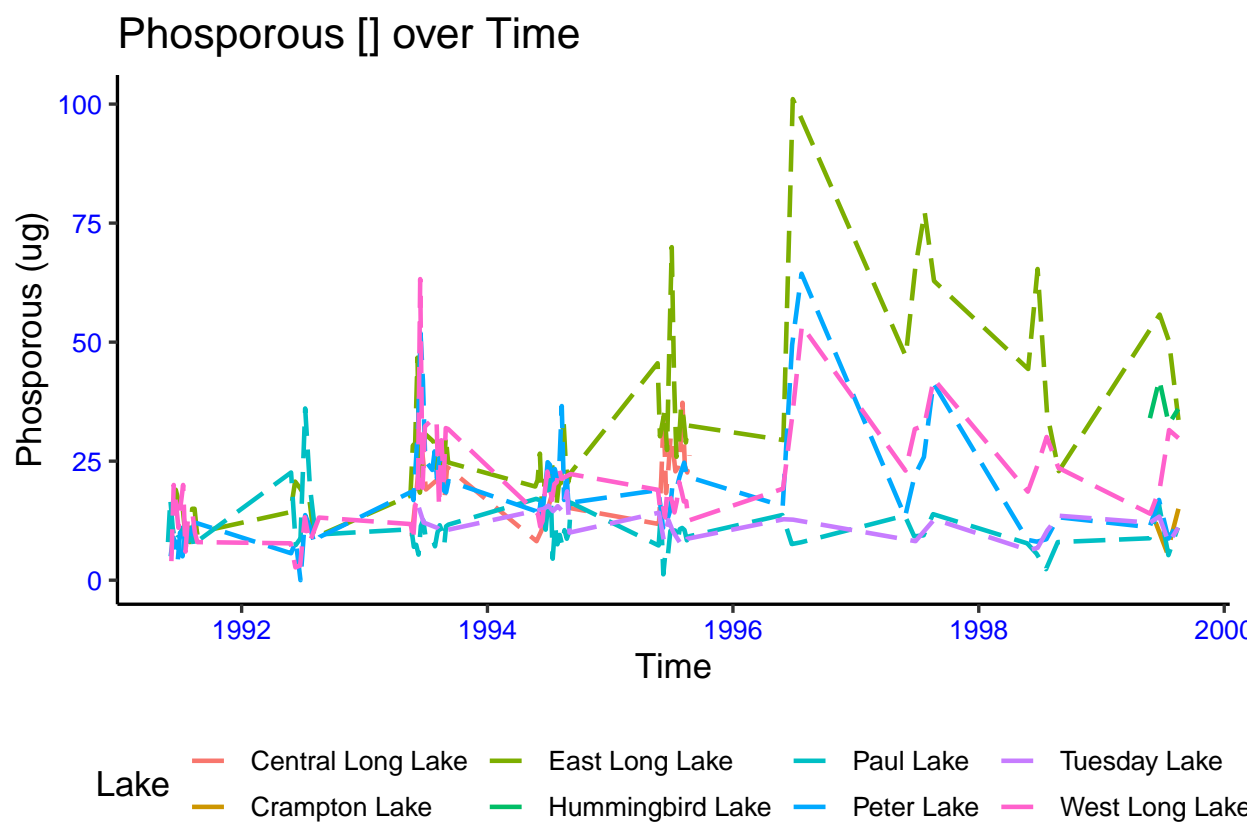


Figure 2: P (ug) over Time

take in analyzing this dataset. What questions might you be able to answer, and what analyses would you conduct to answer those questions?

The data shows that both Nitrogen and Phosphorous have similar trends in concentrations over time. Nitrogen is found in significantly more abundance than Phosphorous according to the tables above. The mean concentration for Nitrogen is 610.9 ug while that of Phosphorous is 18.9 ug. Seasonally, both Nitrogen and Phosphorous drastically increase in middle of the year (Summer/Fall) for all lakes. However, for the year 1996, both Nitrogen and Phosphorous concentrations peaked for all lakes except Paul Lake and Tuesday Lake (concentrations of 2870.3 ug and 101.1 ug, respectively).

---

Moving forward, I would conduct a time-series analysis to determine if there is an overall increase in Nitrogen and Phosphorous across all lakes and for each lake in this study. Since Paul Lake and Tuesday Lake do not share the same pattern as other lakes, a generalized linear model would also be another option for future analysis. If the lake name has a significant effect in determining Nitrogen and Phosphorous, then it's possible that the physical, geographical, or social uses of the lake impact their Nitrogen and Phosphorous concentrations.

## 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:
```