

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



- 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).

- Allows you to combine R chunks and text and format them nicely into a ***PDF, HTML or Word doc.***
- Facilitates *groupwork*

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	Relevant Information
EPAair	Air quality monitoring of PM2.5 and ozone in North Carolina in 2017 and 2018
Litter	Litter monitoring at Niwot Ridge Long-Term Ecological Research (LTER) station from 2016-2019.
Nutrients	Data from studies on several lakes in the North Temperate Lakes District in Wisconsin, USA
Flow Data	Streamflow data from the USGS streamflow gage site 02085000 (Eno River at Hillsborough, NC)
NC County	Shapefile contains county level social vulnerability data for North Carolina
ECOTOX	Data from studies on several neonicotinoids and their effects insects

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`, `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.

```
#Check working directly
getwd()
```

```
## [1] "/Users/clara/Desktop/DATAN/Environmental_Data_Analytics_2022"
```

```
#Load packages
require(tidyverse)
require(lubridate)
require(knitr)

#Set ggplot theme
mytheme <- theme_classic(base_size = 14) +
  theme(axis.text = element_text(color = "black"),
        legend.position = "top")
theme_set(mytheme)
```

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.

```
##   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
```

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

```
#Create processed dataset that meets conditions
Lakenew <- Lake %>%
  drop_na() %>%
  select(lakename, year4, daynum, sampledate, depth, tn_ug, tp_ug, nh34, no23, po4) %>%
  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: Lake Surface Samples: Total Nitrogen

lakename	mean.nitrogen	minimum.nitrogen	maximum.nitrogen	sd.nitrogen
Central Long Lake	83.697481	1.832	169.274	54.499435
Crampton Lake	1.639667	0.631	2.651	1.010003
East Long Lake	107.040315	0.427	634.587	147.530437
Hummingbird Lake	7.504750	5.013	9.112	1.775485
Paul Lake	3.686810	0.000	18.060	3.294013
Peter Lake	41.468372	0.000	451.758	80.049647
Tuesday Lake	2.578067	0.841	5.554	1.089038
West Long Lake	113.402556	0.000	878.939	166.426599

Table 3: Lake Surface Samples: Total Phosphorus

lakename	mean.phosphorus	minimum.phosphorus	maximum.phosphorus	sd.phosphorus
Central Long Lake	2.124889	0.000	7.000	1.5182339
Crampton Lake	1.226667	0.983	1.608	0.3344702
East Long Lake	5.743630	0.000	62.535	8.4669011
Hummingbird Lake	2.335750	1.240	3.263	0.9744100
Paul Lake	2.276544	0.000	10.000	2.1335013
Peter Lake	3.175346	0.000	12.000	2.8039453
Tuesday Lake	1.524300	0.000	6.324	1.3136693
West Long Lake	3.331083	0.000	29.000	3.9450956

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.

```
## 'geom_smooth()' using formula 'y ~ x'
```

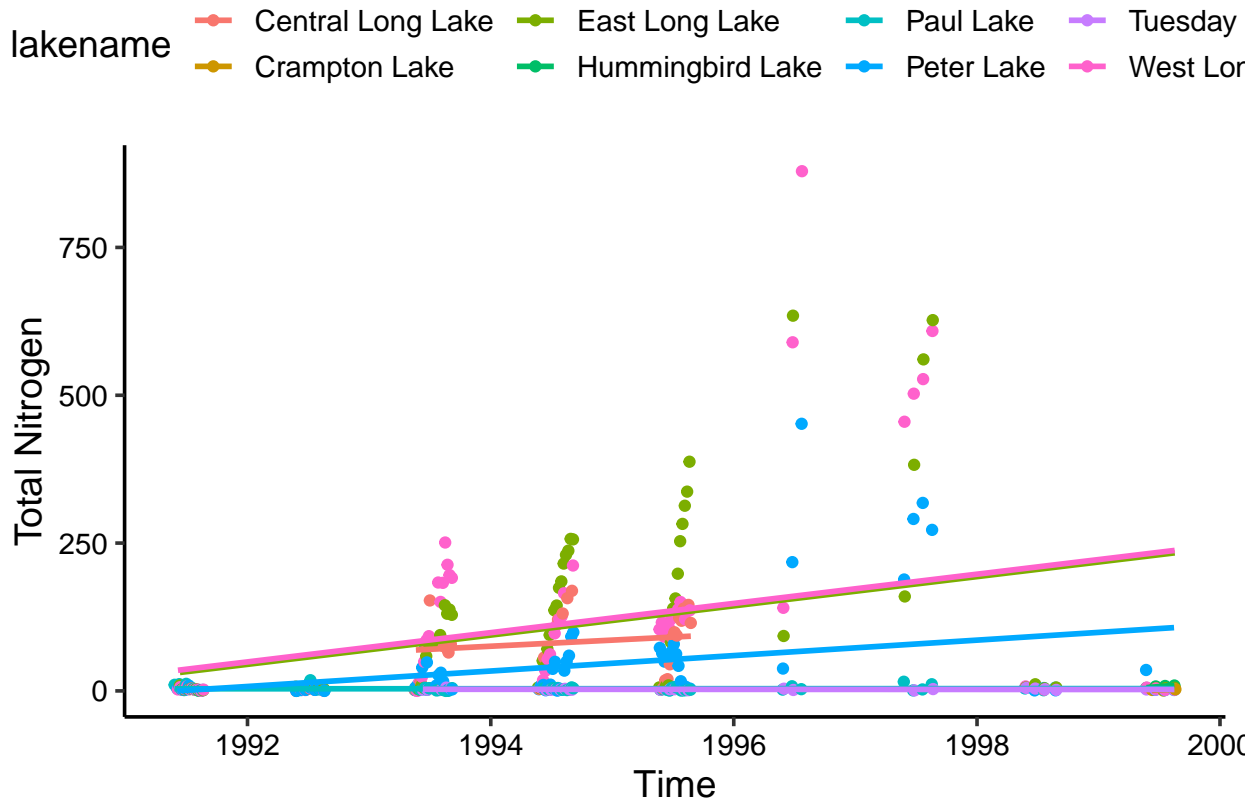


Figure 1: Nitrogen concentration over time

```
## 'geom_smooth()' using formula 'y ~ x'
```

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?

The coding challenge above dealt with a dataset looking at nutrient concentrations in several lakes between 1991-1999. Overall, it revealed that nitrogen and phosphorus concentrations have increased over time (1991-1999) for the majority of the sampled lakes. Some lakes did not show an increase over time.

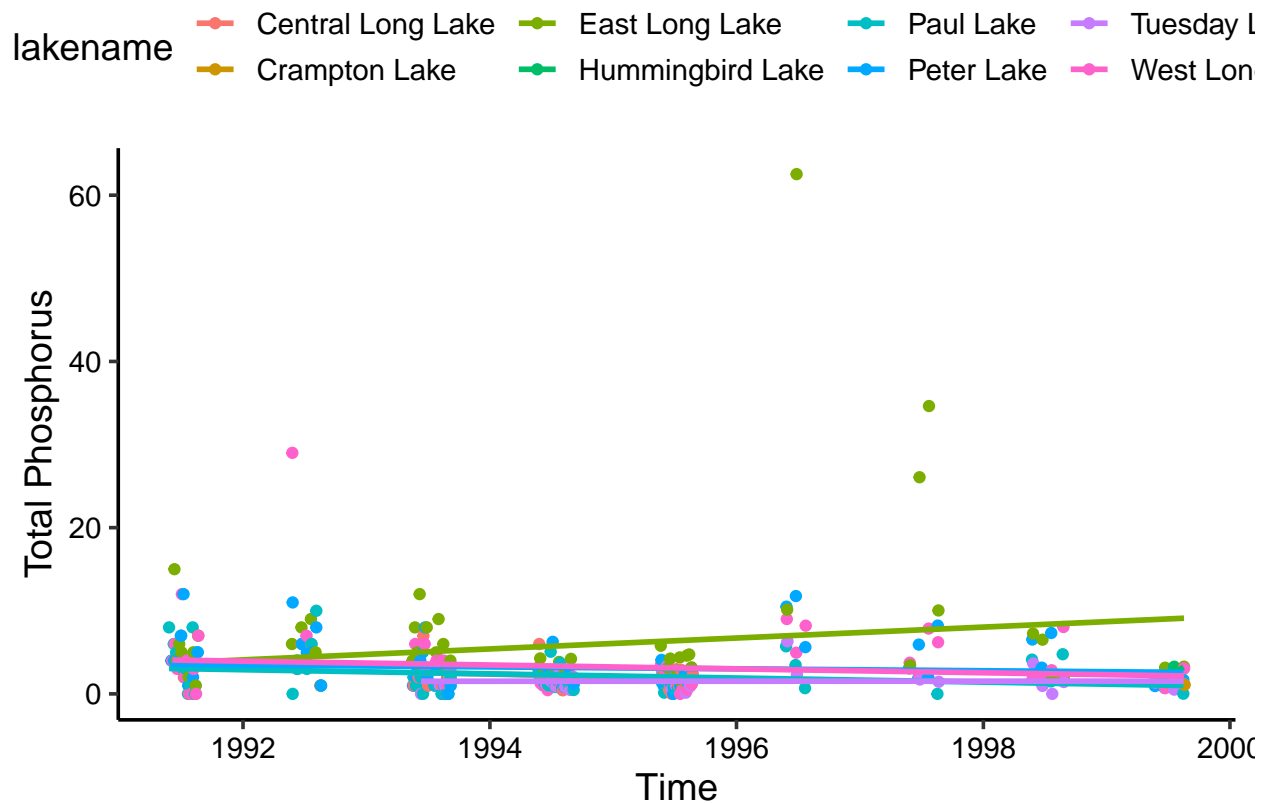


Figure 2: Phosphorus concentration over time

To further analyze this dataset, it would be of interest to determine whether seasonality plays a role in nutrient concentrations found in the sampled lakes. To do this, we would create a time series for the data and proceed to decompose the time series into several components to isolate seasonality. We could also investigate whether nutrient concentration is spatially correlated, by connecting this dataset to spatial data where the lakes are found, and then running a model on the variables of nutrient concentration (phosphorus or nitrogen) and latitude/longitude.

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:
```