Allison Theobold
Montana State University
Bozeman, MT 59717

allisontheobold@montana.edu

http://www.math.montana.edu/graduate/pages/atheobold

This project aims to promote data science literacy among environmental science graduate students, both at Montana State University and in the broader academic community. This goal will be achieved through a suite of four 3-hour hands-on workshops, taught in-person every semester at Montana State University.

This project was funded by National Network of Libraries of Medicine Pacific Northwest Region under Award 11582. The content is solely the responsibility of the author and does not necessarily represent the official views of the NNLM.

Statistical Computing Workshops for Data-Intensive Scientific Research

Allison Theobold
Montana State University
Bozeman, Montana

Introduction to R

This workshop teaches the basics of R, one of the most-used languages of data analytics. The workshop begins with an understanding objects, the building blocks of programming in R. Next, we load data into R and learn to use basic R functions to inspect, extract, manipulate, summarize, and visualize the data. We also discuss how packages developed for R can be accessed and used and how R documentation can be used to help us debug our code.

Workshop Structure

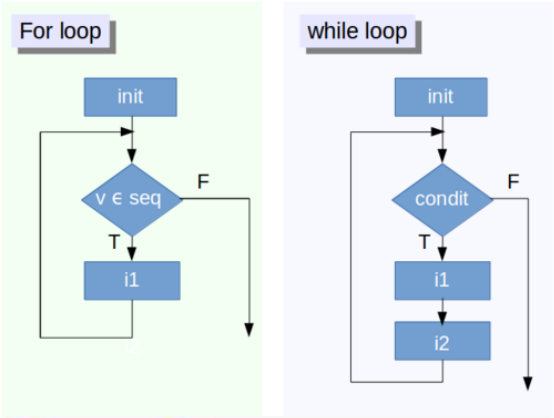
Every workshop is presented using the RStudio Cloud, an easy way to share projects. Each workshop's materials are publicly available and require no software installation, data transfer, or package installation.

RStudio Cloud

RStudio Cloud Project:
https://rstudio.cloud/project/46123

Intermediate R

This workshop builds off the Introduction to R workshop without any expectations of additional knowledge or skills. The workshop begins with introducing relational statements and logical operators as the building blocks of data manipulation. We then use relational statements to introduce and explore conditional statements, building from `if` to `if else` to `else if`. The workshop then provides a motivating example of why looping is good coding practice and how to create for-loops in R. We connect these concepts with the final topic of user-defined functions. A discussion is had about the difference between functions and loops and the ways they each can help to make our code more efficient.

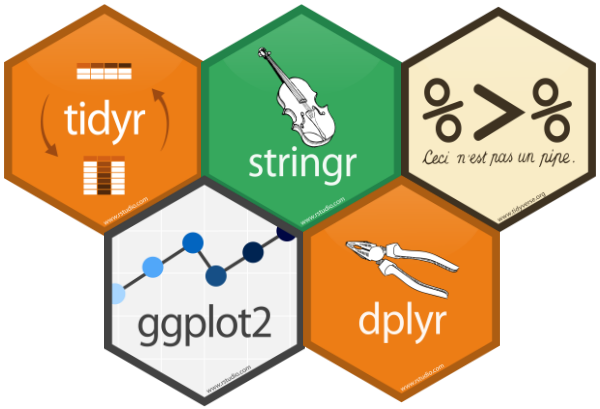



```
if(y < 0){
  "y is a negative number"
}else if(y == 0){
    "y is zero"
}else{
    "y is positive"}
```

RStudio Cloud Project:
https://rstudio.cloud/project/46127

Data Wrangling in R

This workshop covers the foundations of manipulating data in R, using the `dplyr` and `tidyr` packages. We use dplyr to introduce the verbs of data manipulation which allow for users to select, filter, mutate, arrange, and summarise their data. While exploring these verbs, the pipe operator is introduced as a way to chain a set of operations together. Once we have nicely formatted the data, we discuss the concepts of wide and long data formats, for which purposes each of these formats are useful, and how to use `tidyr` to reshape data frames from one orientation to the other.



Data Visualization in R

This workshop targets the ability to create meaningful and attractive data visualizations, a key skill for most researchers. Using the `ggplot2` package, this workshop provides an overview of the grammar of graphics plotting concepts. We explore `ggplot`'s three plotting layers, data, aesthetics, and geometries, for many univariate and multivariate graphics.

RStudio Cloud Projects:
https://rstudio.cloud/project/46135
https://rstudio.cloud/project/46136