

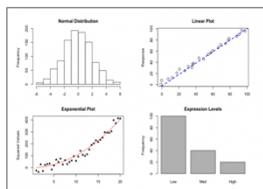
CAPS Bioinformatic Workshop Descriptions

Getting Started: Introduction to R



R is open-source software for statistical computing that has become widely utilized within the field of bioinformatics, and more generally, data analytics. This workshop provides a general introduction to R, and addresses specific topics such as installing and using R packages, understanding and using functions, importing and exporting data, and working with different types of R data structures.

Data Visualization with Base R



One especially powerful aspect of R is its ability to produce high quality, customizable, publication-ready figures. This workshop focuses on using packages and functionality associated with a basic installation of R (base R) to generate and customize several types of example plots, including histograms, barplots, scatterplots, and boxplots. Topics such as plot layout, resolution, and exporting plots from R are also addressed.

Data Visualization with ggplot



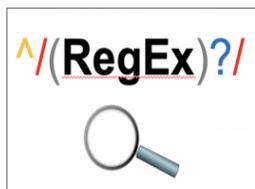
While Base R can be very useful for creating figures, it is limited in the types of plots it can produce. ggplot is an R package that uses the “grammar of graphics” for data visualization, allowing for much more flexibility. As such, ggplot has been adopted by many researchers as a go-to approach for visualizing their data. In this workshop, we create the same plots as in the Base R workshop (histograms, barplots, scatterplots, and boxplots), but do so with ggplot approaches. This allows for a comparison of the two primary methods for plotting in R, and provides participants with sufficient background and templates to subsequently create a wide variety of figures using ggplot functionality.

Getting Started: Terminal/Command Line Computing



This workshop provides a basic introduction to command line computing within a Terminal window. Topics include navigating within Terminal, manipulating and organizing files and directories, downloading data files, and using basic wildcard characters (globbing). Skills developed in this workshop are necessary for a wide range of bioinformatic tasks, including searching and manipulating large text-based data files and utilizing HPC resources such as the Ohio Supercomputer Center or Amazon Web Services.

Regular Expressions for Searching and Editing Large Data Files



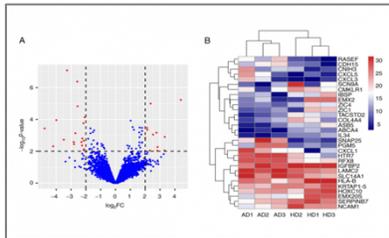
Regular expressions provide powerful methods for searching and editing text in large data files. In this workshop, we work with regular expressions in both GUI-based text editors and command-line tools such as ‘grep’ and ‘sed’. Participants will be introduced to some of the different types (flavors) of regular expressions, their associated syntax, and also some common pitfalls associated with their use. Basic familiarity working in Terminal (command line) is recommended.

Utilizing High Performance Compute Resources



Analyses of large datasets, including bioinformatic datasets, often demand substantial computational resources. In many cases, these are available as high-performance compute (HPC) clusters that are accessed remotely, either through a Terminal window, or a web browser. In this workshop, participants will learn some of the basics of how to interact with a HPC resource, using the Ohio Supercomputer Center as an example. Basic familiarity working in Terminal (command line) is recommended.

Data Visualization for ‘Omics Data in R



The scale of information produced in ‘omics studies can present challenges when it comes to creating figures that effectively communicate results. In this workshop, we’ll work with example ‘omics datasets to create some widely-used types of plots including PCA, volcano plots, heat maps, and Manhattan plots. A basic understanding of R syntax is recommended for participating in this workshop.

Documenting and Sharing Code With Rmarkdown



Developing reproducible code and sharing results are two important parts of any bioinformatics task. Rmarkdown provides a convenient framework for both writing and documenting code, and also executing the code to produce summary reports (e.g. pdf, html, etc). This workshop provides an introduction to creating an Rmarkdown document, and then using that document to generate a report that includes the code itself, notes about the analysis, and results figures and tables generated by the code. We also explore how code from programming languages outside of R (i.e. Perl, Python, Bash, etc) can be incorporated into these documents.