Data Science Workshop

British Society for Proteomic Research Meeting 2018

Alistair Bailey

June 27 2018

Contents

Overview

This book covers:

1. An introduction to R and RStudio
2. An introduction to tidyverse and base R
3. Importing and transforming proteomics data
4. Visualisation of proteomics analysis

The analysis is of an example data set of observations for 7702 proteins from cells in three control experiments and three treatment experiments. The observations are signal intensity measurements from the mass spectrometer. These intensities relate the concentration of protein observed in each experiment and under each condition. The analysis transforms the data to examine the effect of treatment on the cellular proteome and visualise the output using a volcano plot and a heatmap. Click here to download the csv file.

Requirements

An up to date version of R (R Core Team, 2018) and RStudio (RStudio Team, 2018). If you are new to R, then the first thing to know is that R is a programming language and RStudio is a program for working with R called an integrated development environment (IDE). Further details in Chapter @ref((#r-rstudio).

Download R here and Download RStudio Desktop here.

These materials were generated using R version 3.5.0.

The following R packages:

```
install.packages(c("tidyverse","gplots","pheatmap"))
```

Chapter 1

Introduction

Placeholder

1.1 What are R and RStudio?

1.1.1 Environments

1.2 Why learn R, or any language ?

1.3 Finding your way around RStudio

1.3.1 What is real?

1.4 Where am I?

1.5 R projects

1.6 Naming things

1.7 Seeking help

1.7.1 Asking for help

Chapter 2

Getting started in R and the tidyverse

Placeholder

2.1 The tidyverse and tidy data

2.2 Data visualisation

2.3 Workflow basics

2.3.1 Assigning objects

2.3.2 Function anatomy

2.3.3 Atomic vectors

2.3.4 Attributes

2.3.5 Factors

2.3.6 Lists

2.3.7 Matrices and arrays

2.3.8 Data frames

2.4 Learning more R

Chapter 3

Creating scripts and importing data

Placeholder

3.1 Some definitions

3.2 Using scripts

3.3 Running code

3.4 Creating a R script

3.5 Setting up our environment

3.5.1 Bioconductor

3.6 Importing data

3.7 Exploring the data

Chapter 4

Transformation and visualisation

Placeholder

4.1 Fold change and log-fold change

4.2 Dealing with missing values

4.3 Data normalization

4.3.1 T-test

4.4 Visualising data

4.5 Creating a volcano plot

4.6 Creating a heatmap plot

Chapter 5

Going further

5.1 Learning `dplyr` verbs

5.2 Getting help and joining the R community

5.3 Communication: creating reports, presentations and websites

References

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2018). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.