



jumping rivers

Next Steps in the Tidyverse

The *tidyverse* is essential for any statistician or data scientist who deals with data on a day-to-day basis. By focusing on small key tasks, the tidyverse suite of packages removes the pain of data manipulation. This course takes the next steps in using the tidyverse and examines how and where to use packages such as broom and purrr in an analysis.

Course Outline

The course will cover

- forcats: factors for the tidyverse
- broom: Tidying statistical output
- purrr: A functional programming toolkit
- stringr: Strings, the band of a data scientist's life

Learning Outcomes

By the end of the day participants will understand...

- what tidy data means for statistical modelling
- the challenges and solutions when working with strings
- the basics of functional programming and how it relates to the tidyverse
- when and where to use factors
- the types of problems regular expressions can help with

Course Structure

Factors are used to work with categorical variables, variables that have a fixed and known set of possible values. Used correctly, factors are incredibly useful. However, base R's obsession with converting everything to a factor is annoying.

- The what, why and where of factors
- Manipulating factors with the forcats package

broom is an attempt to bridge the gap from untidy outputs of predictions and estimations to the tidy data we want to work with. The package contains a variety of functions that enable output from common R functions, into a tidy format.

- An overview of the available tidiers
- The tidying functions: `tidy()`, `augment()` and `glance()`

R is a functional programming language, e.g. the `apply` family. However, due to the evolution of the language, the interface has some idiosyncrasies. The `purrr` package provides a complete and consistent set of tools for working with functions and vectors.

- The `map()` functions and formula notation
- Using `nest()` and `unnest()`

Strings aren't glamorous, and while base R can handle all tasks, it isn't always clear how to approach each task. The `stringr` package provides a cohesive set of functions designed to make working with strings straightforward.

- Getting to grips with strings
- The fundamentals of regular expressions

Prior Knowledge

This course assumes basic familiarity with R and the tidyverse, e.g. the *Mastering the tidyverse*.