Data manipulation with dplyr

About dplyr

Data manipulation is a key part of any data analysis, crucial for interactive exploration and as a precursor for visualisation and modelling. The dplyr package is the next iteration of the plyr package, designed specifically for the needs of data analysis. It features:

- a carefully crafted set of functions that make it easy to solve the majority of data manipulation problems
- blazing fast performance for in-memory datasets (100-10,000x faster than plyr)
- flexible backends that allow you to work with your data wherever it lives, whether that be in data frames, data tables, or databases.

What you'll learn

In this tutorial, you'll learn how to use dplyr with the `nycflights13` package which contains five interlinked data frames ranging from 16 to over 300,000 observations. You'll learn:

- The five most important data manipulation verbs: `select()`, `filter()`, `mutate()`, `arrange()` and `summarise()`.
- How to use the pipe operator, `%>%`, to build up rich workflows from simple primitives.
- How to combine `summarise()` with `group_by()` to compute rich grouped summaries.
- How to work with multiple data sets with the two-table verbs: matching joins (`left_join()`, `right_join()`, `inner_join()` and `outer_join()`), selecting joins (`semi_join()` and `anti_join()`), and set operations (`intersect()`, `union()` and `setdiff()`).

We'll focus on using dplyr with in-memory data frames, but I'll also give you a taste of how you can use dplyr with your existing data stored in relational databases like PostgreSQL, MySQL and SQLite.

Who should attend

You should attend this tutorial if you've struggled with manipulating datasets in R. You will get the most out of it if you already know how to get your data into R, and are familiar with the basics of tidy data. If you have used plyr before, you'll find the ideas of dplyr immediately familiar, and you'll appreciate the faster speed and more thoughtful syntax.

The tutorial will be a mixture of lecture and hands-on activities, so make sure to bring your laptop so that you can practice what you learn.