# Software Development Principles for Statistical Modelling
# Basic Data Manipulation in R

Mick Cooney
mickcooney@gmail.com

January 30, 2017

`https://www.github.com/kaybenleroll/training_courses`.

Code is available in the `sdpsm_intro` directory.

Content in this workshop is based on the 'Software Carpentry' course: `http://software-carpentry.org/`

# 1. Introduction

## 2. Introduction to `dplyr`

Do the following exercises by manipulating the tibbles. Your output should be in the form of a tibble.

**Exercise 2.1** Load the `mpg` data into memory.

**Exercise 2.2** Calculate the average mileage for both city and highway driving across the dataset.

**Exercise 2.3** Calculate the mileage by transmission type.

**Exercise 2.4** Calculate the average mileage by car year.

**Exercise 2.5** Which car manufacture has the most fuel economy overall?

**Exercise 2.6** Append two new columns to the tibble, giving the mileage in kilometres.

**Exercise 2.7** Append two new columns to show the mileage in kilometres per litre.

**Exercise 2.8** Create a new version of the tibble that contains the model, the year and km/l fuel efficiency.

## 3. Reading and Manipulating Data Files

**Exercise 3.1** Load the data from the file `SPY_data_2007.csv` into a tibble.

**Exercise 3.2** Reload the data so that the year column is read in as a string.

**Exercise 3.3** Load all the SPY data across all the years into a single file.

**Exercise 3.4** Write the data out as a single file, but only with the columns `symbol`, `date` and `adj_close`.