Software Development Principles for Statistical Modelling
Basic Data Manipulation in R

Mick Cooney
mickcooney@gmail.com

January 31, 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 to `dplyr`

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

Exercise 1.1 Load the `mpg` data into memory.

Exercise 1.2 Calculate the average mileage for both city and highway driving across the dataset.

Exercise 1.3 Calculate the mileage by transmission type.

Exercise 1.4 Calculate the average mileage by car year.

Exercise 1.5 Which car manufacture has the most fuel economy overall?

Exercise 1.6 Append two new columns to the tibble, giving the mileage in kilometres.

Exercise 1.7 Append two new columns to show the mileage in kilometres per litre.

Exercise 1.8 Create a new version of the tibble that contains the model, the year and km/l fuel efficiency.

2. Reading and Manipulating Data Files

Exercise 2.1 Load the data from the file `SPY_data_2007.csv` into a tibble.

Exercise 2.2 Reload the data so that the year column is read in as a string.

Exercise 2.3 Load all the SPY data across all the years into a single file.

Exercise 2.4 Write the data out as a single file, but only with the columns `symbol`, `date` and `adj_close`.

3. Data Visualisation with `ggplot2`

Exercise 3.1 Create a line plot of the SPY adjusted close price.

Exercise 3.2 Create a line plot of the SPY data adjusted close price from 2015 onwards.

Exercise 3.3 Create boxplots of the city fuel efficiency of the `mpg` cars by manufacturer.

Exercise 3.4 Facet the above boxplots by cylinder count

Exercise 3.5 Create a density plot of the overall highway fuel efficiency.

Exercise 3.6 Create multiple fuel efficiency plots facetted by manufacturer. The city and highway efficiencies should be on the same plot with different colours.

Exercise 3.7 Redo all the above plots, but add labels to the $x-$ and $y-$ axes and add a title to each plot.

4. Introduction to `rmarkdown`

Exercise 4.1 Create an rmarkdown document for all the work in this module.

Exercise 4.2 Give the document an appropriate title.

Exercise 4.3 Add your name and the current date to the heading of the document.

Exercise 4.4 Render this document as both a HTML and a PDF document.