Advancing in R

Module 2: Data manipulation

Supplementary exercises

Tidying and reshaping time series data

Locate and then load the 'OBrienKaiser' data:

```
OBK <- read.csv("OBrienKaiser.csv")
```

This data set contains data for individuals split into different treatment groups, and measured multiple times in each of 3 separate periods (5 repeat times in each of pre-treatment, post-treatment, and follow-up). Can you restructure this data into a tidy format?

Make sure you name your variables properly, and also think about whether you also need to mutate them into different data types.

Creating a 'Quarter' variable for Weather data

We may want to inspect the weather data by quarter, rather than individual months (i.e., months 1-3 would be quarter 1, months 4-6 quarter 2, etc). How would you add such a variable to the WEATHER data frame?

Hint: you could set up nested 'ifelse' calls within a mutate function, but perhaps it would be less confusing to create a supplementary data frame that you could join with your main one…?

Asking further questions with dplyr

Try installing the 'nycflights13' library and loading the data:

```
install.packages("nycflights13")
library(nycflights13)
data(flights)
?flights
```

Look at the description file (from the '?flights' command): there is a huge amount of information in this data set! Brainstorm questions with the person next to you, then approach these with dplyr.