MATH550/SCC461 Programming for Data Science
Lab 5
Assessment

Tom Palmer

Deadline for coursework submission is **9am Monday** 6th **November 2017**.

Assessment

In the last assessment you looked at the Australian severe storm events data set recording all the Australian storm events in a 40 year period from 28th September 1975 to 28th September 2015. The data consists of a row for each severe weather event and 14 columns. The last 6 columns consist of Comments (titled `Comments, X, X.1, X.2, X.3, X.4`) were ignored in last weeks assessment but we will make use of this narrative text to extract additional information. Reminder that the first 8 columns of the data set are:

Variable(s)	Description
Event.ID	Unique event ID
Database	Type of severe storm event.
ID	A secondary ID variable.
Date.Time	Date and time in the format dd/mm/yyyy hh:mm.
Nearest.town	Nearest town to the event.
State	Australian state in which the event was recorded.
Latitude	Latitude of the event.
Longitude	Longitude of the event.

Tasks

Download the data set to your working direction, and produce a script that performs the following:

1. Load the data into R.

   Combine the comments from columns (`Comments, X, X.1, X.2, X.3, X.4`) into a single column containing comments. Call the column `All.comments`.

   Select the following columns to keep for further analysis, `Event.ID`, `Database, Date.Time, State, All.comments` and make sure all variables of the appropriate type.

   After which you should add the following command to your script:

```
print(sapply(DF, class))
```

where 'DF' is the name of the data frame.

[3 marks]

2. Create an indicator variable which states whether or not a storm event has resulted in a flash flood.
   Hint: Make sure you sort out all terms relating to flash floods.

   Print a plot of the number of flash floods per year from 1975-2015.
   Hint: Create a vector to contain the number of flash floods per year.

[3 marks]

3. For severe wind events often the wind speed is given. The wind speed is given in knots or km/h.

   Extract all wind speeds both those in knots and km/h.
   Hint: Knots can be abbreviated by kts or kt. Also note that wind speed can be a single, double or triple digit number.

   Convert km/h wind speeds to knots (1 knot = 1.852 km/h) rounding the wind/speed to the nearest knot.
   Hint: It is helpful to work with a reduced data frame which includes only those observations with a wind speed recorded.

   Print a boxplot of the wind speeds recorded per state.

[4 marks]