Homework 2

Data science for high shcool students, Summer 2018

13 July, 2018

1. R Markdown example

Create an R Markdown file for the code in HW1 Problem 3. The aim of this excercise to create a simple one page report which has a few lines of text about AQI values in zipcode 20871 (or another of of your choice), a table showing the AQI values by date and a plot (same as in HW1 Problem 3).

1. Set the 'echo=F' in the R section so that code is not displayed in the rendered pdf.

2. BBC News API

The BBC news API can be used (amongst many things) to get the latest news articles and then we can do some analytics on those. This is what we would be attempting in this problem.

1. Get an API key from https://newsapi.org/ (signup process is very simple and ofcourse free).
2. The following URL (the correct term to use in this context is **API end point**) is an example of the API to get the most recent 20 articles for news related to the U.S.

https://newsapi.org/v2/top-headlines?country=us&apiKey=eed8134a37b33bac8a3416f94a2e64a8

(*I have changed the API key so this URL will not work as is, replace the API key with your own API key before trying.*)

As you can see we are passing parameters **country** and **apiKey** as part of the URL. See an example of this API in action at https://newsapi.org/docs/get-started. For this part of the problem just try getting the API to work by copy pasting the URL in a web browser (with your own API key).

3. Now try to run this API from an R script. The code will be very similar to the AirNow API code.

3.1 Make sure you print and check the status code from the response.

3.2 The JSON returns by this API is more complex than what we saw in the AirNow example. Convert the JSON response to a data frame using the familiar fromJSON function. To understand the structure of the response (and this is true for any data structure) use the **str** function. Once you understand the structure you will be able to print the **articles** dataframe as a table, try it. You will see that this is too much information, usually we will extract out fields of interest.

3.3 Extract out the **id** field from the articles dataframe (it is not directly under articles but under articles and then source), plot a bar chart where each bar respresents the number of articles from each source. See example code here https://www.r-bloggers.com/make-a-bar-plot-with-ggplot/.

So at this point you have accomplished writing R code to programatically get data from the BBC News API and make a bar chart of the various sources of news articiles provided by the API.