Assignment 05

Bus disruptions

Consider the Twitter data in the file `translink.RData` provided in the archive for this homework (you can load this using the R command `load`). Write an R function `translink` that takes 4 numerical arguments: a year (a numeric value such as 2020), a month (a numeric value between 1 and 12, inclusive with 1 indicating January), and a day of the month, and an hour of the day (in 24 hour time). The R function should return a list with two elements: 1) an element with the name `start` and with value specifying a character vector enumerating all bus routes that started to have disruptions during the hour indicated by the date and time provided to the function, 2) an element with the name `stop` and with value specifying a character vector enumerating all bus routes that stopped having disruptions during the hour indicated by the dat and time provided to the function. You don't have to care about timezones for this question: you can assume that the time specified by the parameters to the function are in the same time zone as the data in `translink.RData`. Disruptions are defined as starting or stopping only if a Tweet indicating such is present in `translink.RData` (*i.e.*, you don't have to consider times that fall outside of the data provided in `translink.RData`). Note that some of the tweets in `translink.RData` may be truncated: you may ignore the truncated portions (this is approximate). Note that there are some corner cases among the tweets (not all of the disruptions are indicated with the same format). For full marks, consider some of the corner cases. Example usage is as follows:

```
> disruptions = translink(2020, 1, 26, 3)
> disruptions$start
[ 1 ] "401" "406"
> disruptions$stop
[1] "23"
```

For bonus points, have your function use the Twitter API to download and process tweets for the provided hour (instead of using the static `translink.RData` file).

a) Provide a *pdf* containing all of the code that you wrote for this project. If your code spans multiple files, put all of the code into a single *pdf* and use comments to indicate the filenames. You may use code from outside public domain sources provided that you follow the guidelines for academic honesty, and you cite the sources clearly in the comments.

(10 points)

b) Write a short report arguing that the code you've provided works as required, and provide a *pdf* of the report. For example, show the `translink` function being called with different arguments, and show that the corresponding elements of the data file prove that the correct bus routes are returned. You could also show that for a given element of the data file, the appropriate parameters provided to the `translink` function can be provided to return that specific route.

(5 points)