

Let's get some more practice with regular expressions, file paths, and downloading and unzipping files via R. Please turn in an RMarkdown document that completes the following tasks. Explain the main code chunks in a few sentences.

1. Finding incomplete links on a webpage and completing them.

- (a) Using the following pseudo code, find all Daily AQI by County links to zip files from the following URL, https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI.

```
library(httr)
library(stringr)
a_variable_name <- GET(URL)
html <- content(a_variable_name, "text")
incomplete_links <- str_match_all(html, pattern="")
```

- (b) Using string manipulations, make all of the `incomplete_links` complete, by prepending (adding to the front of the string) the necessary "http://STUFF." In the end, you should have a vector of complete URLs.

2. Figuring out where to store downloaded data.

- (a) Create (with your mouse) a subdirectory of your MATH 385 directory that might store downloaded data.
- (b) Store the file path of the directory into a variable named `base_dir`.
- (c) Write a `for` loop that postpends appropriate file names to `base_dir` using the function `file.path`. The variable `incomplete_links` should help here. Hint: before diving into a `for` loop, make sure you know how `file.path` works.

3. Working with downloaded data.

- (a) Download the following URL: https://aqs.epa.gov/aqsweb/airdata/daily_aqi_by_county_2016.zip. Information about these data is found at https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI.
- (b) Unzip the file 2016 data. Try to use R's function `unzip` instead of your mouse.
- (c) Read in both 2017 and 2016 air quality data into R.
- (d) Perform a simple analysis on these data sets, so as to tell me something interesting about them.