Advanced R

Exercises for session 6: XML.

1. (a) Write a function to read in the airport locations from the file `airportlocation.csv` used in session 2, and write out a KML file for a single airport (specified by name or abbreviation) or a set of airports. If you don't have Google Earth installed but do have internet access, you can put your KML file on a website and then supply its URL to Google Maps to see the results.

(b). Using the Seattle flight data, plot mean arrival delay and mean departure delay for each airport. Use `identify()` to identify when a point is clicked, and write a KML file for that point.

2. Look at the URL http://ip-api.com/docs/api:json, which provides a geolocator service returning JSON. Use it to find where you are.

3. Read in the XML file `phiSITE767857.xml`, which describes promoter sites for a set of bacteriophage viruses (from phisite.org).
(a) Use `xpathApply()` to extract the organism name for each site (`/phisite/site/organism/name`) and the sequence (`/phisite/site/sequence`)

4. Try the PUBMED query service from slides 5 and 6.

5. Using the `phiSITE767857.xml` file, extract the sequence for promoter sites that have experimental evidence using `xpathApply()`.
Notes:
- `xpathApply()` takes a function as its third argument, which is passed each of the XML elements returned by the xpath.
- If `site` is an element of type "site" then the element `site[["evidence"]][["type"]]` is the type of evidence (experimental or predicted) for the site
- `xmlValue()` extracts the actual content from an XML element (eg a string)
- The sequence element for a site is `site[["sequence"]]`