

Exercise: Xpath

Peter Meißner

29 February 2016

1) xml2 package

a) getting started

- load the xml2 package
- use `ls(package:xml2)` to get a list of available functions
- use `read_xml()` to read in the following resource: <http://www.r-datacollection.com/materials/ch-4-xpath/potus/potus.xml> and save it myxml
- use your browser to look at the XML

b) convenience functions

- how many entries has the XML, which function did you use
- extract all the text of the XML and use `cat()` to print the text to the screen
- use a function of the xml2 package to have a look at the structure of the XML

c) extraction

- use XPath to get all religion nodes
- use CSS-Selectors to get all religion nodes
- extract the text from the nodes and use `table()` and `sort()` to produce a sorted table
- use `xml_node()` in combination with XPath to extract the node for the first president
- extract the name node for all presidents after the 40th
- using XPath get the nodes for Pierce, Garfield and Taylor
- extract the occupation
- transform the XML into a data.frame