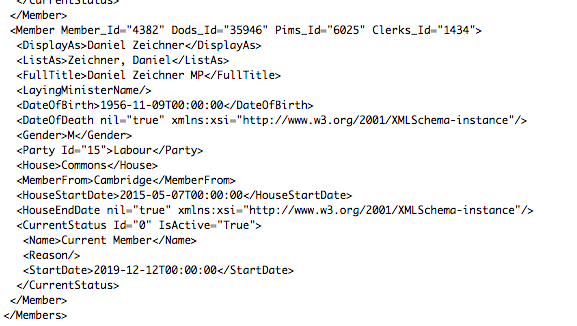
Data cleaning issues



The fact that there were four ID values (important identifiers) embedded in the child node <Member> meant that this XML file could not be converted correctly into a data frame in R.

It was possible to read the file into R using the XML package and the xmlParse or xmlTreeParse function. When the xmlToDataFrame function is used, the values in the <Member> node are truncated and excluded from the data frame.

This was critical problem as the ID value are needed as the unique values for the member IDs.

After a lot of googling, I discovered I could import XML to Excel using the Developer > Import function. This would splendidly and after some spot testing, I could confirm the data has been converted correctly from XML into a table in Excel.

After this I was able to import the excel file as a dataframe in R using the Openxlsx package.

Next….data exploration! Finally!