**Scott Kleinman's Topic Model (25 Topics) of *LA Times*, *New York Times*, & *The Guardian* from 2013-2014** (1/12/16)

**Source Corpus:**

**Deduplication:**

**Scrubbing:**

**Topic Model Settings** (Mallet options):

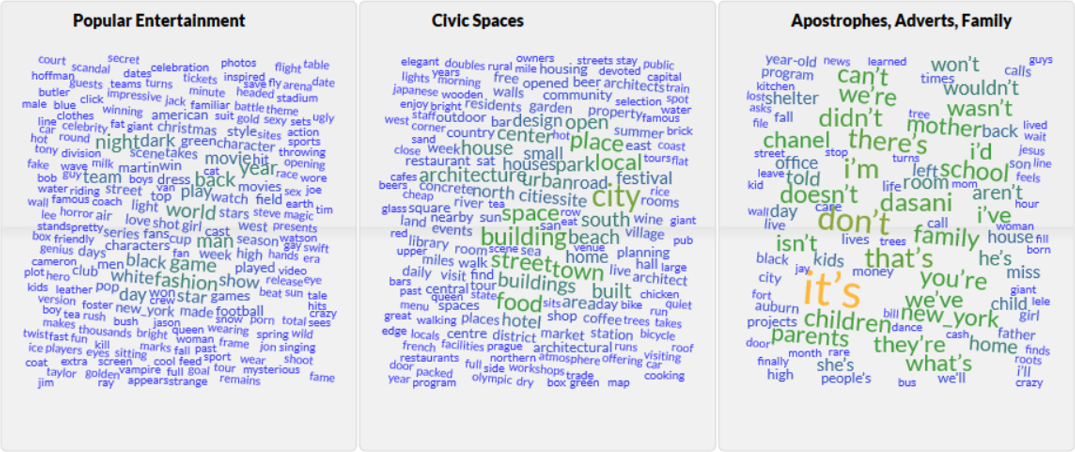
* Number of Topics: 25
* Mallet optimize-interval: 20

**Other Information:**

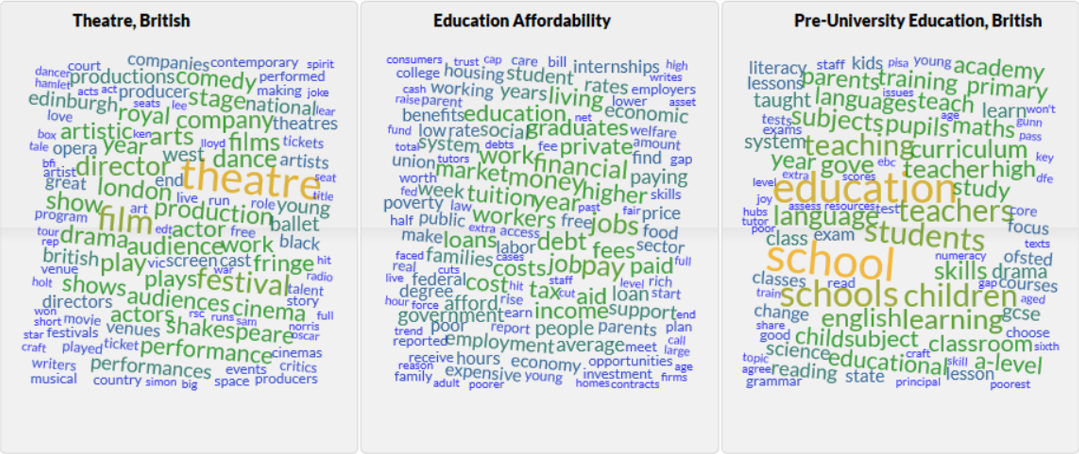
After topics were converted to documents using Lexos, Topic Clouds were generated, and document names were changed to interpretive labels. Clustering was performed, and the resulting dendrogram was exported to Newick format using a development version of Lexos. A second dendrogram was produced in EvolView by from the Newick export.

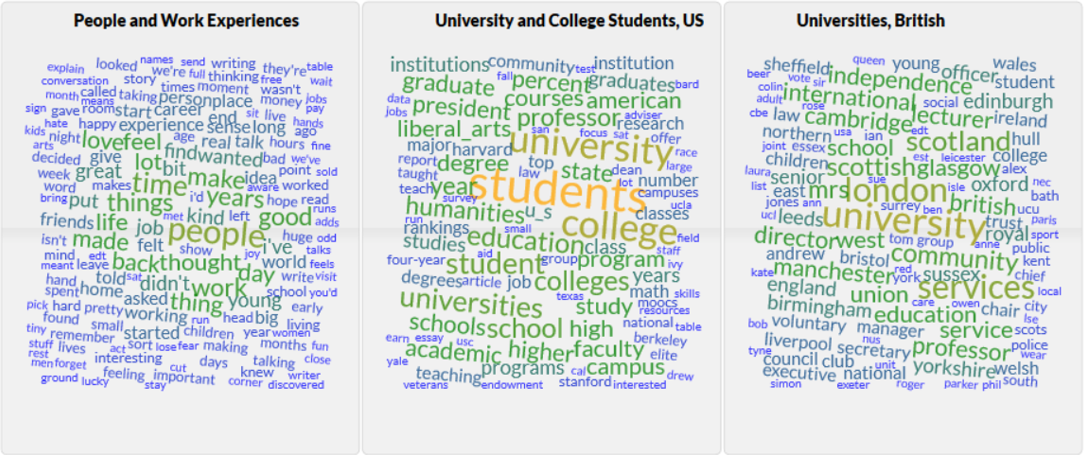
**Topic Clouds**

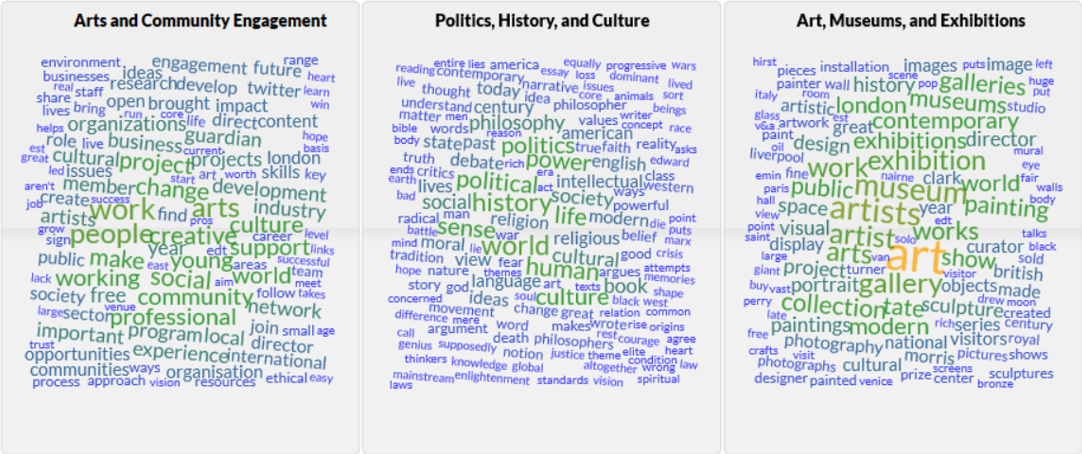








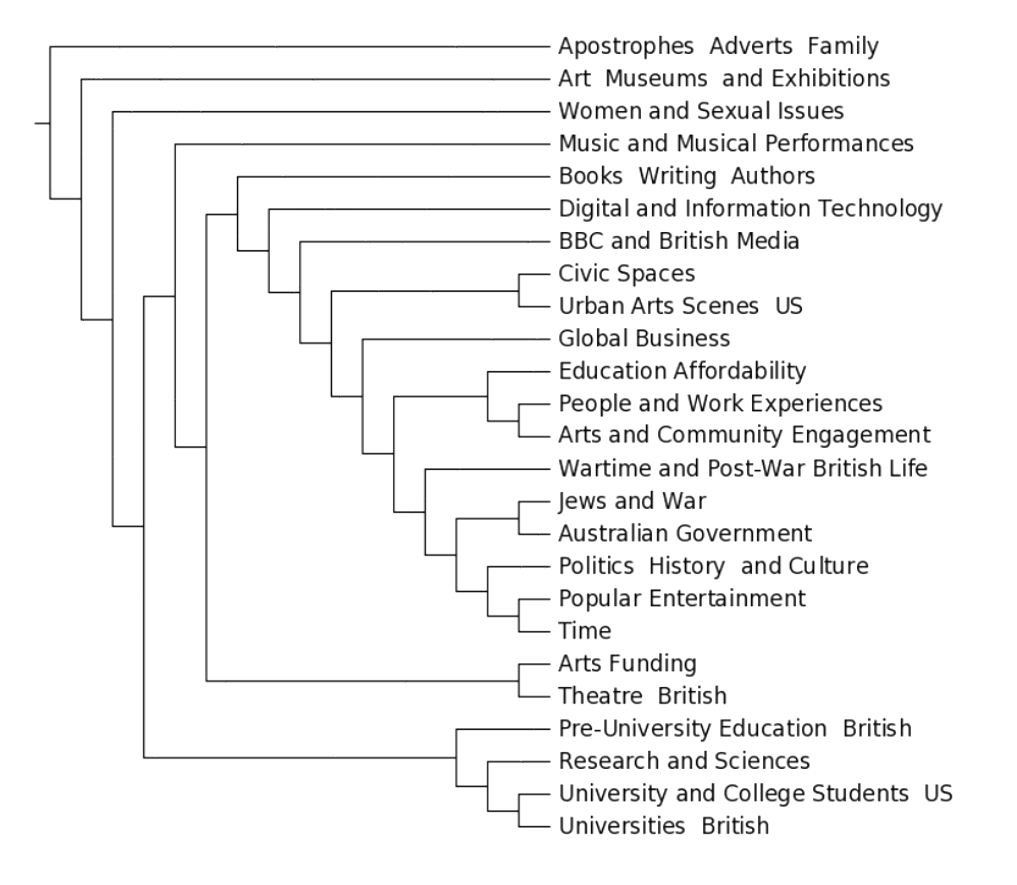








**Hierarchical Cluster Analysis** (Distance Metric=Euclidean, Linkage Method=Weighted)



**Preliminary Analysis**

Apart from the junk topic “Apostrophes, Adverts, Family”, the model clearly distinguishes different families of topics. Museums and Art exhibitions surprisingly have little in common with topics related to the civic arts scene and (public) funding. But funding for “the arts” is closely tied to theatre. Another closely-related set of topics seems to revolve around the history and political scene of the last century or so. “Education affordability” is surprisingly not linked to the topics most directly related to education, but rather to work experience (perhaps unsurprisingly) and community engagement. The close linkage between civic spaces and the urban arts scene is not apparent from the topic clouds, but it might suggest that the civic and more Bohemian expressions of public “humanities/arts” experience. “Liberal arts” seems prominent only in the sphere of US colleges.

Finally, it is remarkable how prominent specifically British terms and topics are when *The Guardian* is added to the Corpus. Very few of the topics seem to reflect as distinctly a US experience, with the exception of “University and College Students, US”. It will be interesting to see if that remains the case with a larger number of topics.