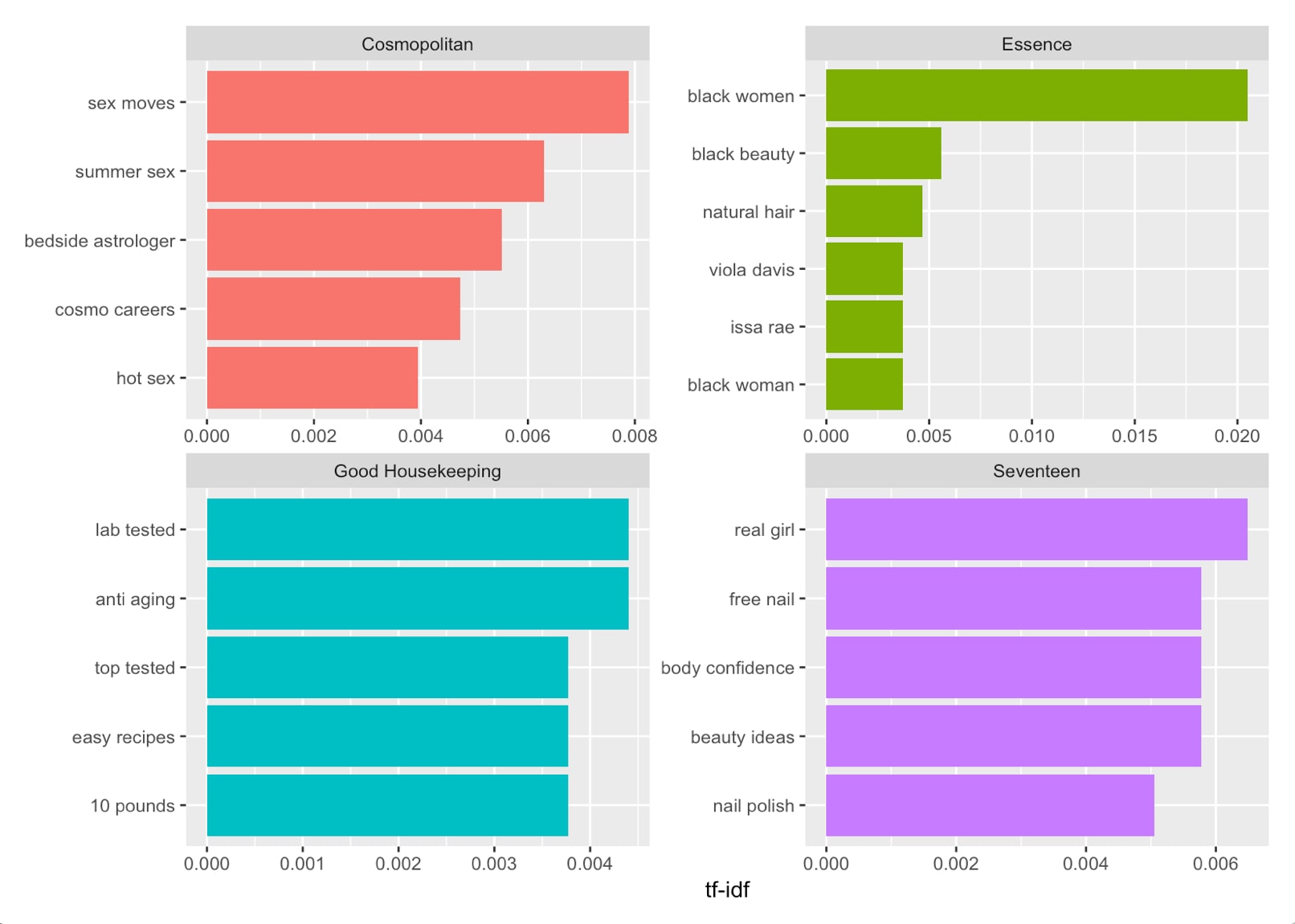
Tf-idf (Term Frequency - Inverse Document Frequency)

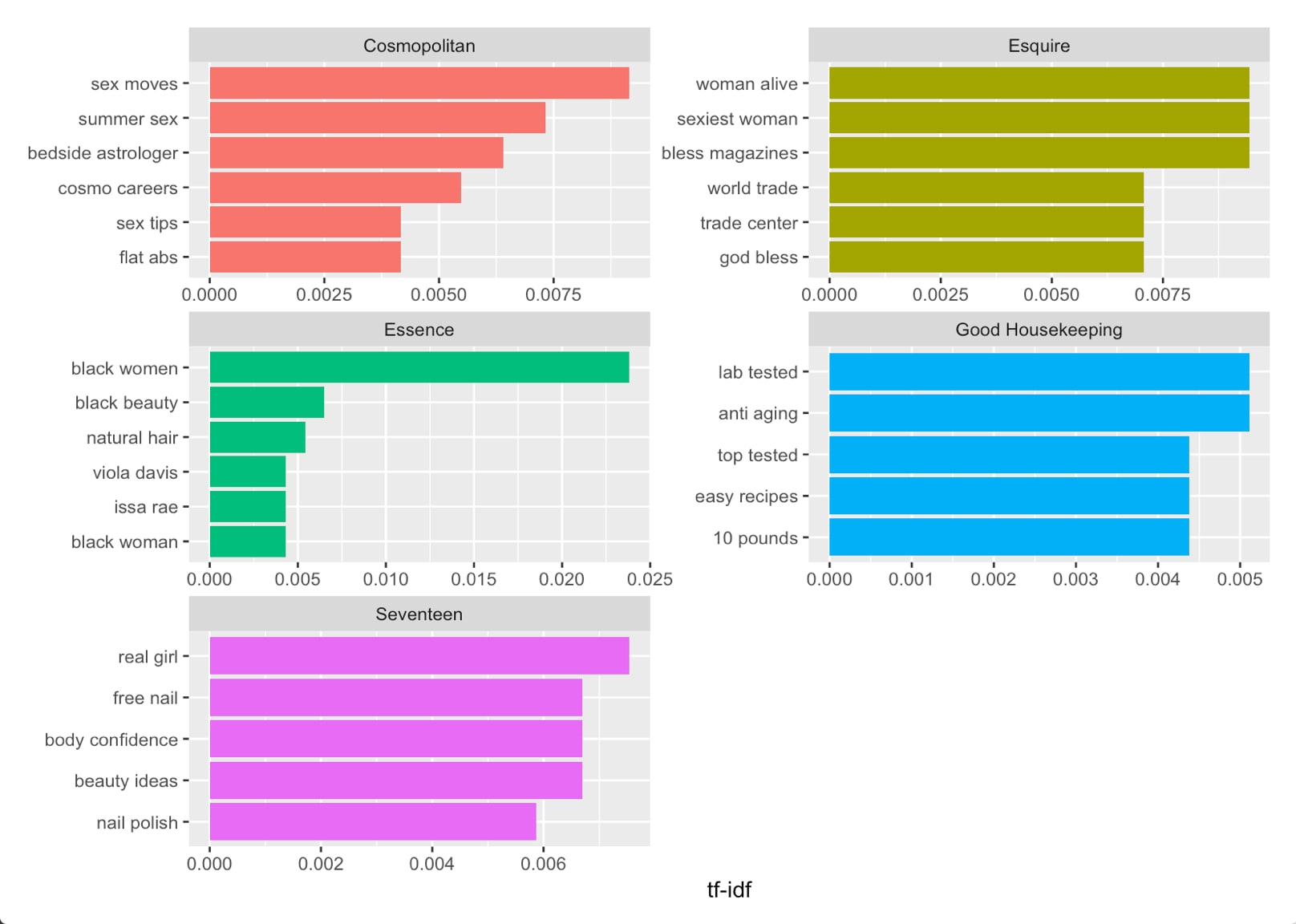
* Term frequency measures how frequently a word occurs in the document
* Inverse document frequency decreases the weight of words commonly used in the document and increases the weight of words not used very often in the document
* Tf-idf is the multiplication of these two quantities for each word (frequency of each word adjusted for how often it occurs)
* Document in this case is a magazine
* The charts below are the 10 items (single words, bigrams, trigrams) with the highest tf-idf

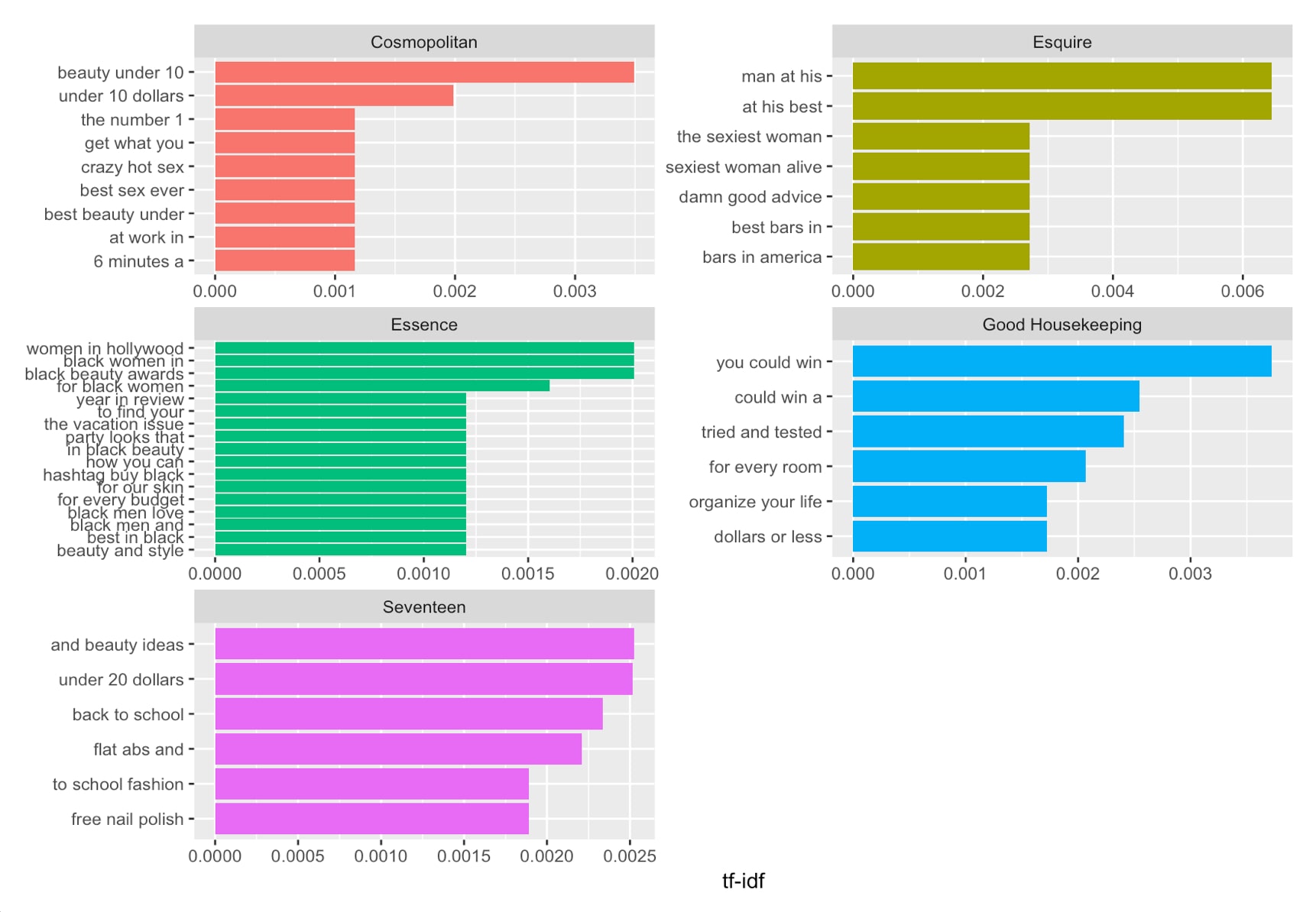
**Single Word tf-idf**



**Bigram tf-idf**

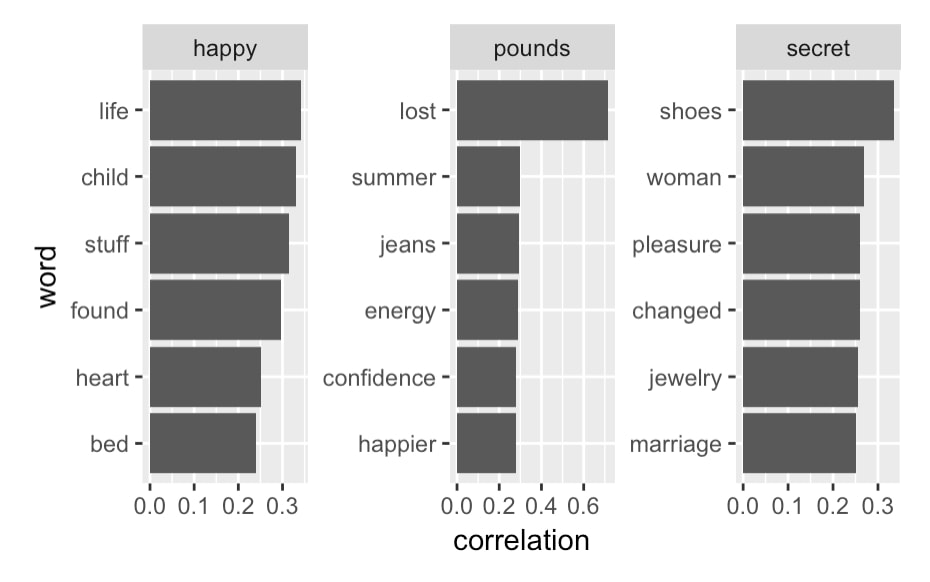




**Trigram tf-idf**

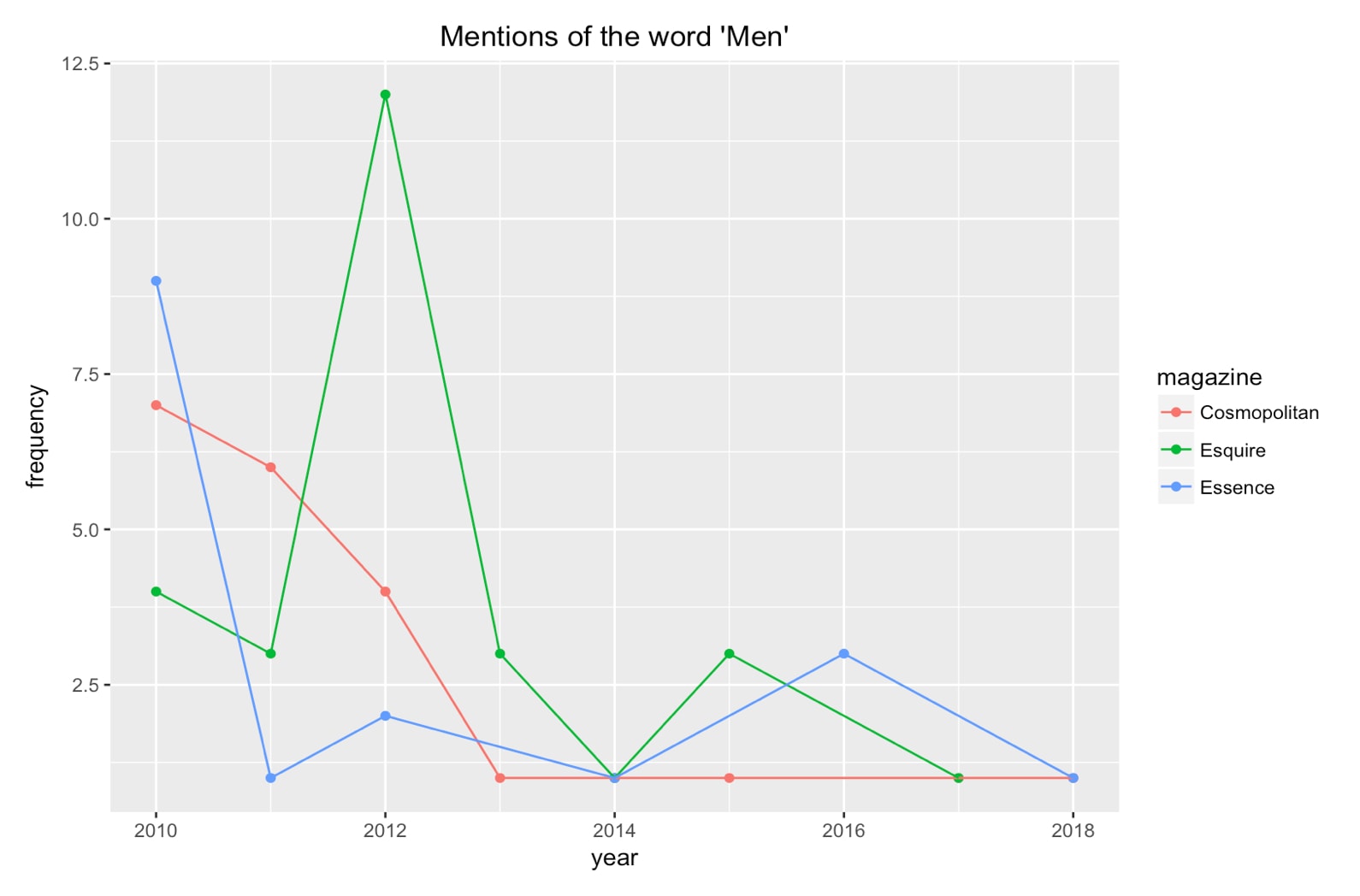
**Other preliminary visualizations**

Pairwise correlation

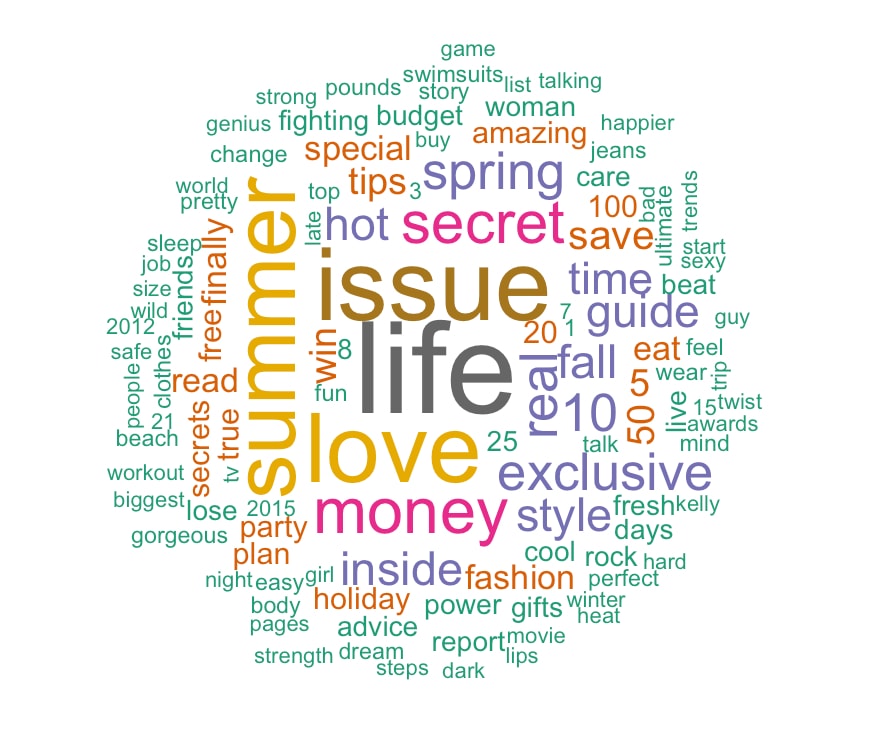
* Measure of how frequently words appear together on a magazine cover

**Word Clouds!**

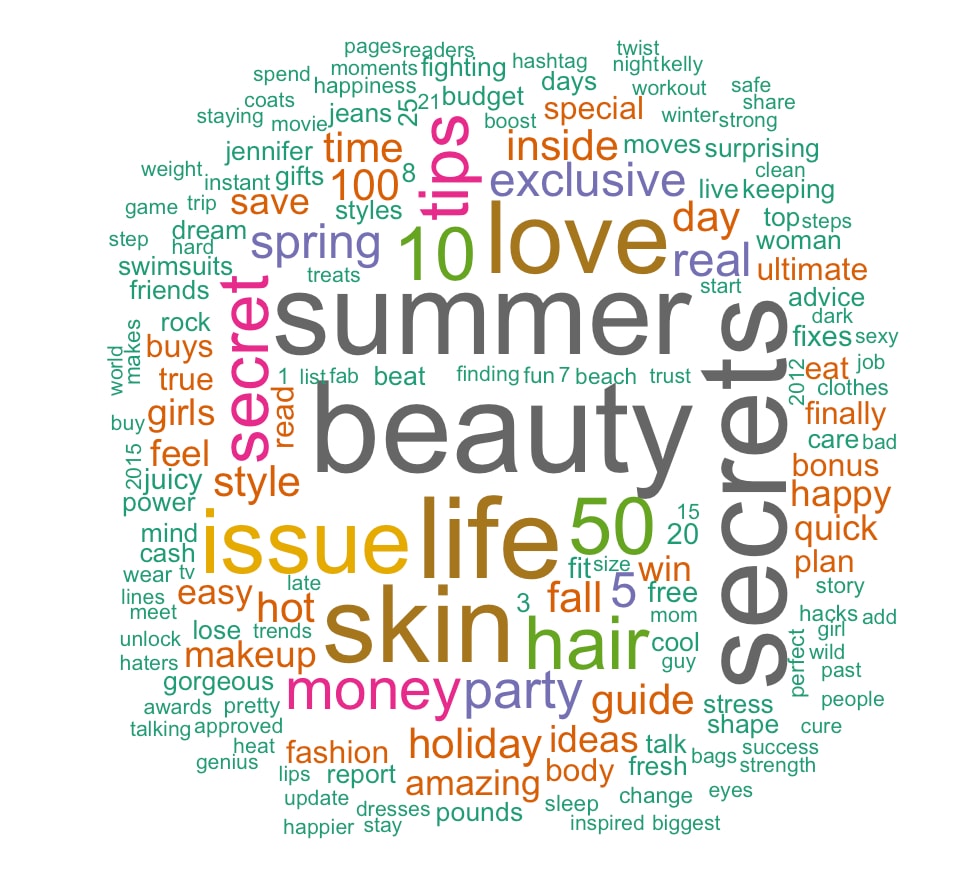
Simple cloud (word frequencies) Comparison cloud (comparing word frequencies across magazines

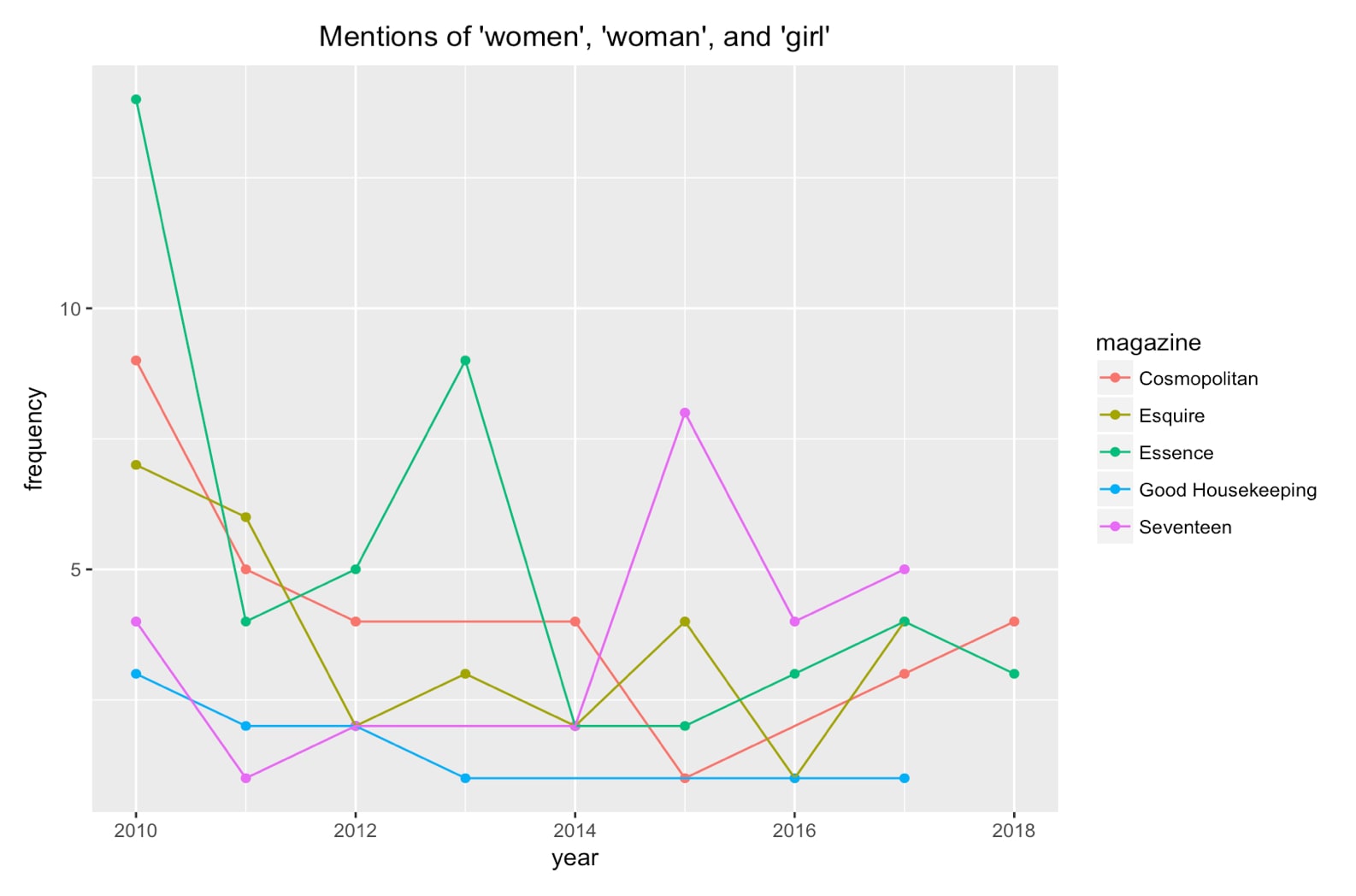


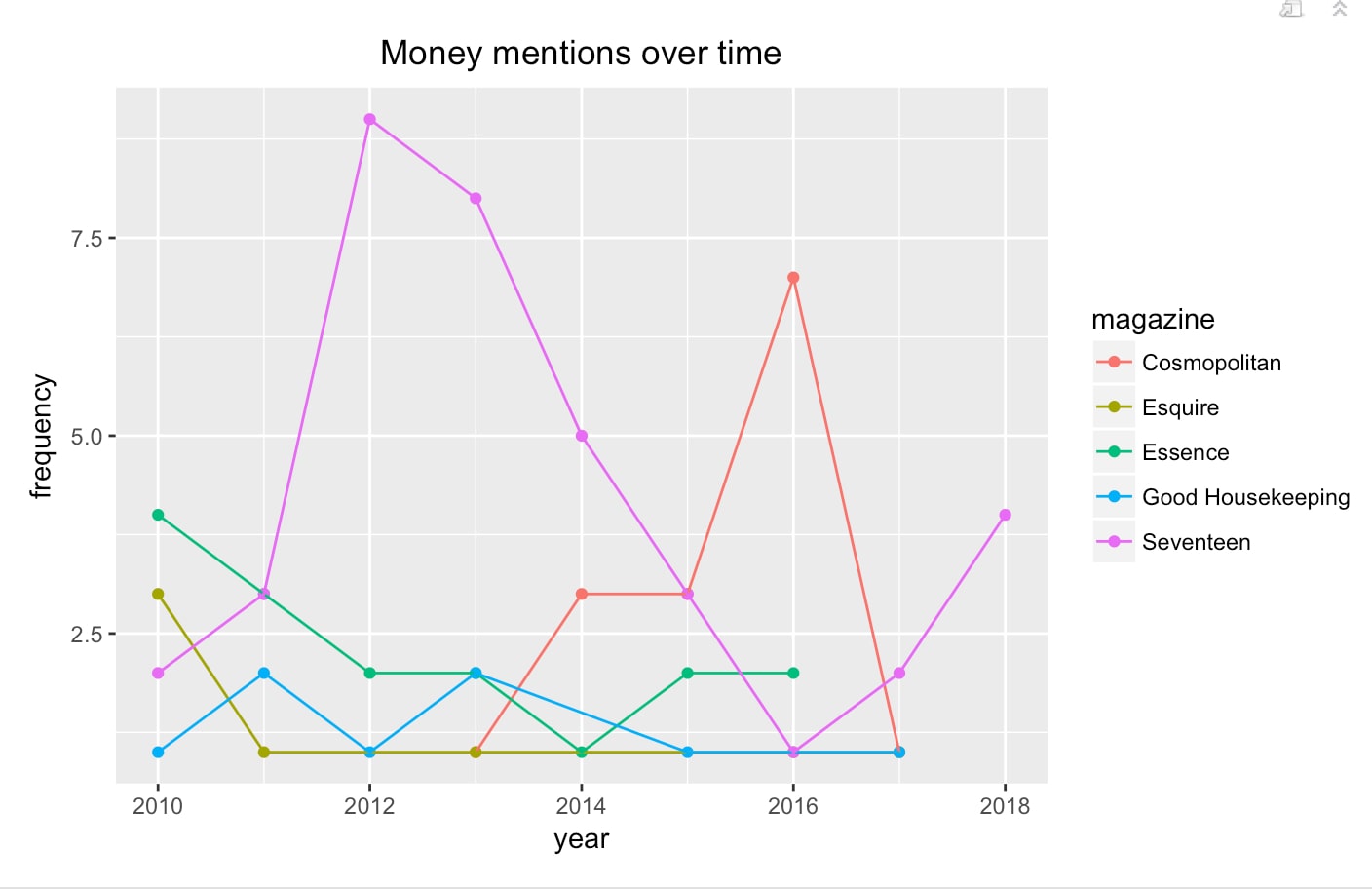
**Commonality cloud (words that appear in all magazines, size proportional to frequency)**

All magazines

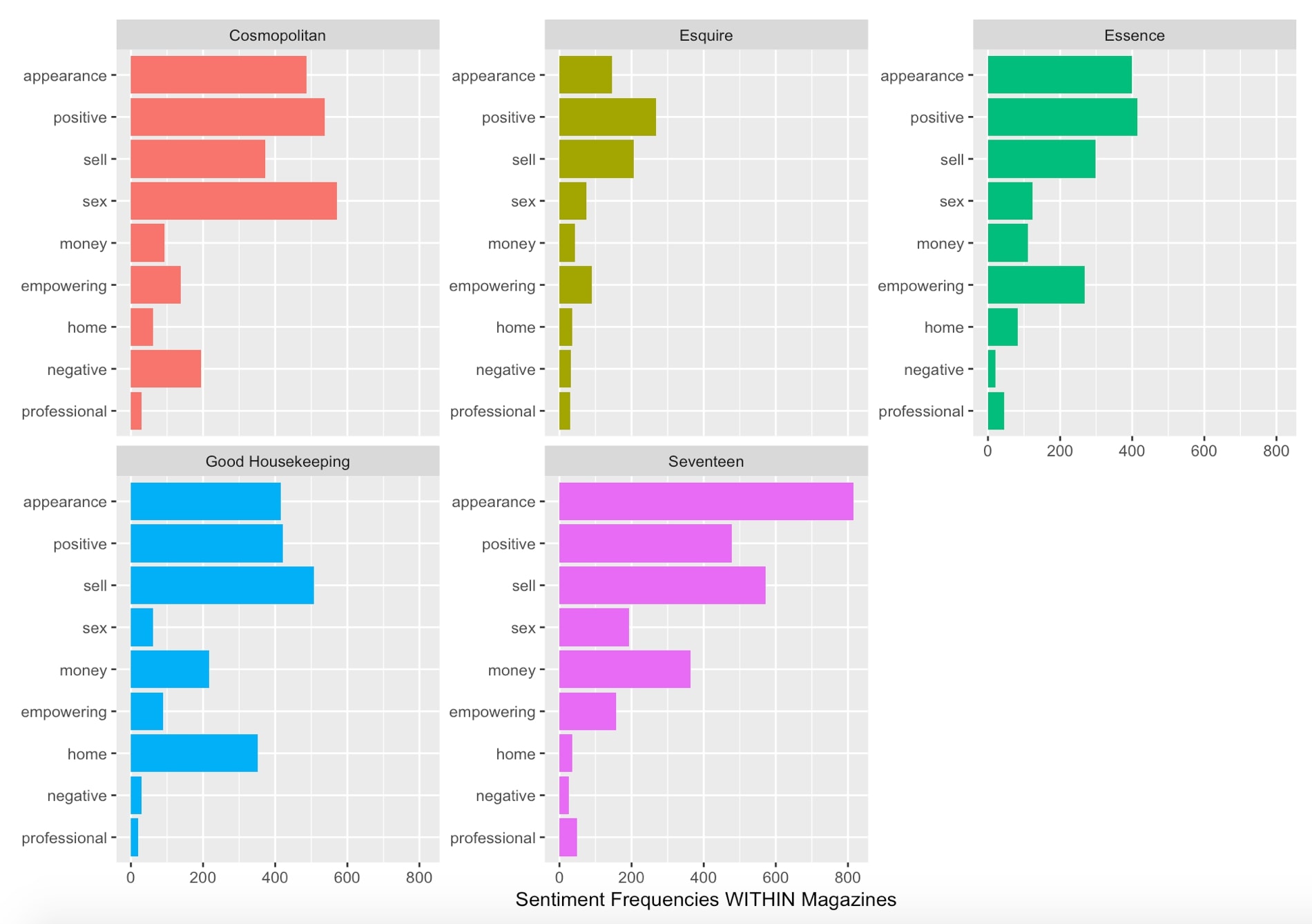
Women’s Magazines



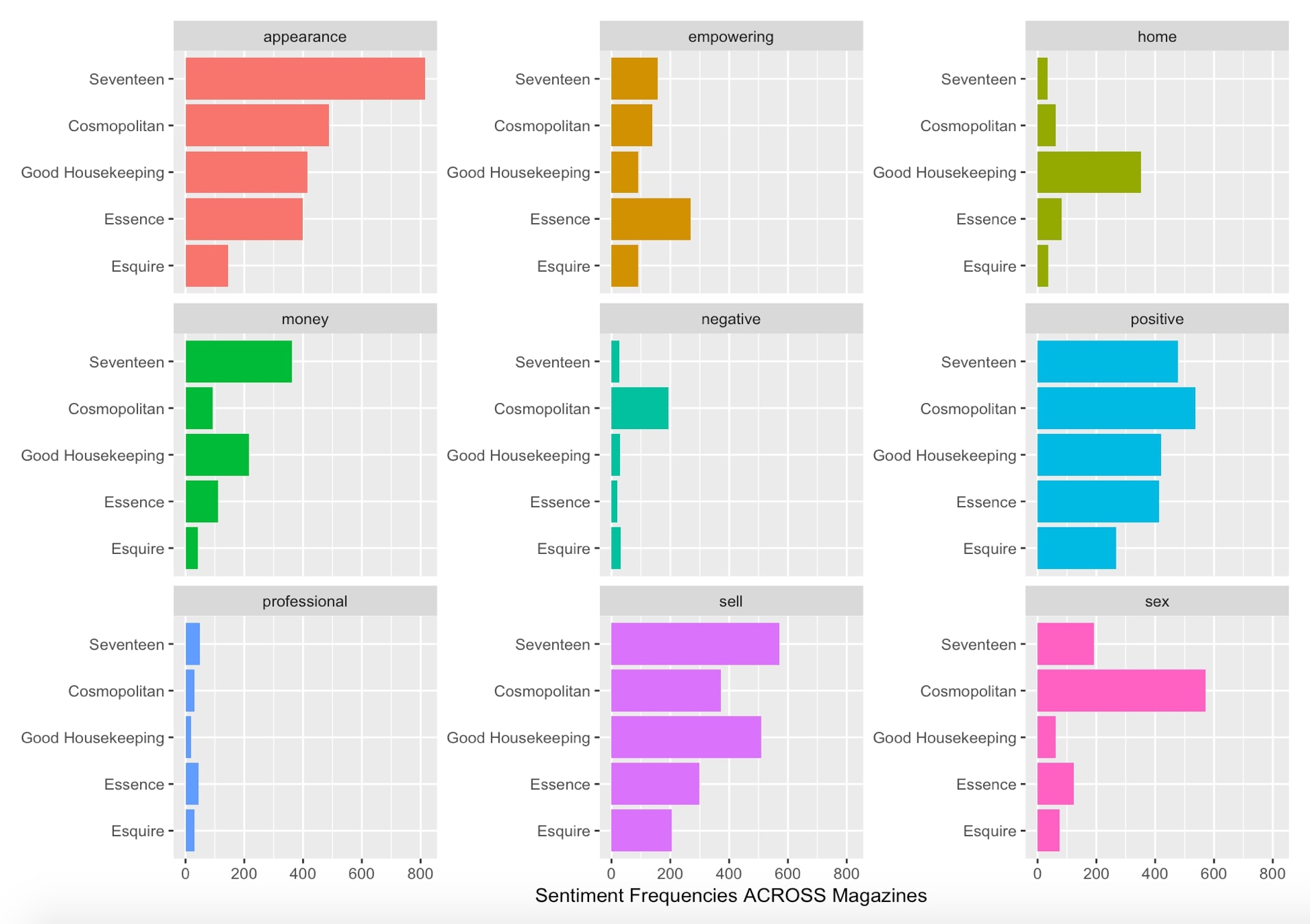
**Some example graphs of word frequency over time**

****

**Sentiments Analysis**



Frequency of different sentiments in each magazine

****