

# Topic Modelling

$N$  documents  $\leadsto M$  unique words

Step 1: Derive document-term matrix

	Word 1	...	Word M
Document 1	2	...	0
...	...	...	...
Document N	0	...	1

---

Let's flip it around and produce a new document:

Step 1: Sample the topics

For example, we may have 60% Topic 1,  
40% Topic 2,  
0% Topic 3

Step 2: Sample the words

(a) Sample the topic from the which the word comes

(b) Sample the word using the distribution for sampled topic