

---

# TOPIC MODELING WITH GENSIM

Parul Sethi



@parulsethi



@parul1sethi

---



---

Millions of Documents have been leaked!



What we do we look for? About what the leaks are? If elections, then which country

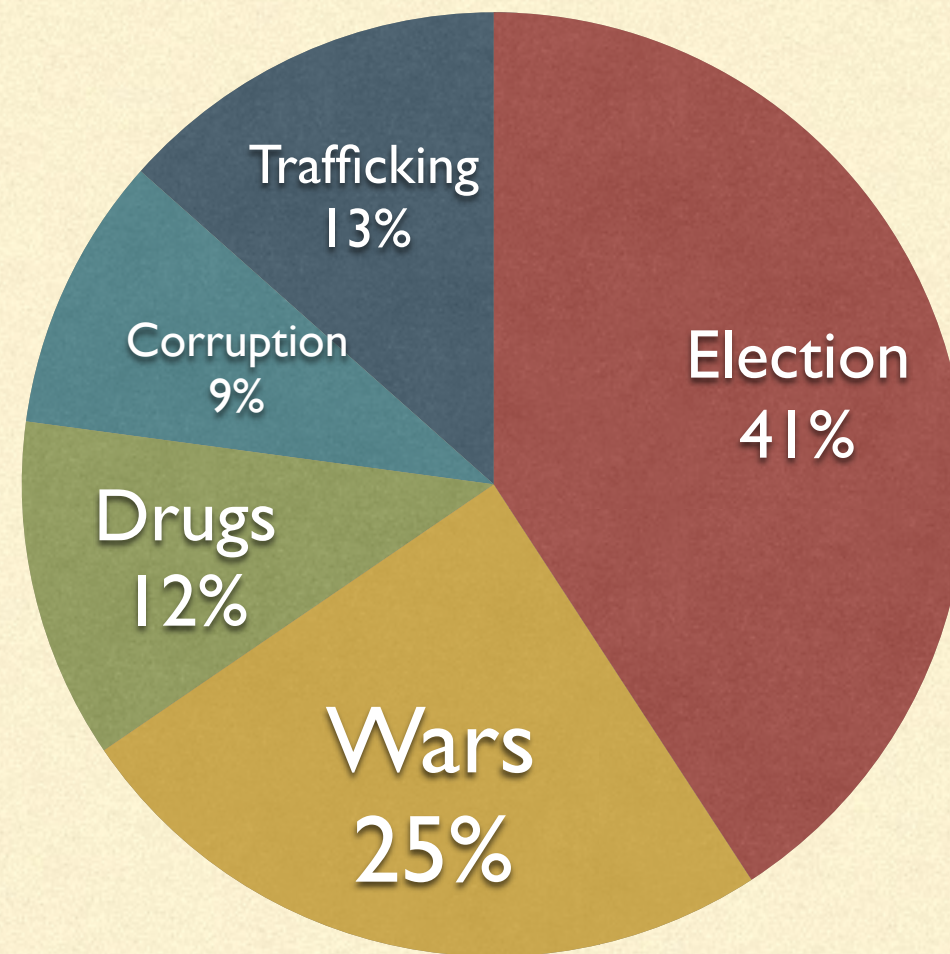
---



---

# BIRD'S EYE VIEW OF DOCUMENTS

---





---

# HOW ARE THE LABELS INFERRED

---

Topic model gives us a set of words for each topic that it finds from the collection of documents.



It looks through a corpus for these clusters of words and groups them together. In a good topic model, the words in topic make sense, for ex. “navy, ship, captain” and “tobacco, farm, crops.”

---



# WHAT ABOUT DOCUMENTS

- Which topics a document belong to
- Which documents belong to this topic

DOCUMENT		
Conservative	Liberal	
Rain	President	Sun Snow
Minister		
Olympics	Football	

Recipe		
Politics	Weather	Sports
50%	30%	20%

Politics	Weather	Sports
Conservative	Rain	Olympics
Trump	Sunny	Football
Clinton	Beach	Championship
....	....	....



---

# ENOUGH OF THEORY! LET'S GET PRACTICAL NOW

---

- [https://github.com/parulsethi/PydataDelhi2017/blob/master/topic\\_modelling.ipynb](https://github.com/parulsethi/PydataDelhi2017/blob/master/topic_modelling.ipynb)



---

THANK YOU!

---