|  |
| --- |
| IST707 text MINING |
|  |

|  |  |
| --- | --- |
| Date Submitted  8/25/2020 | WK8: LDA |
| Professor | Dr. Ami Gates |
| Submitted by: | Beverlyn Tucker |

**Introduction**

The mainstream media is a traditional reference for news entities capable of shaping and molding opinions of the general public. The dissemination of knowledge covers a wide variety of platforms, including television, online, and print. Vast capabilities of communication allow news agencies to maximize information delivery. Moreover, benefits potentially include a better-informed and engaged general public. However, the underlying facilities often relate to the ideals of a network and general interest factions, which can be limited by social and technological constructs. MSNBC is principally known to favor a left-wing perspective, while Fox News indulges more a right-wing viewpoint. CNN attempts to report an unbiased perspective. However, in the recent wave of political events, CNN has been perceived as more leftist than in prior years.

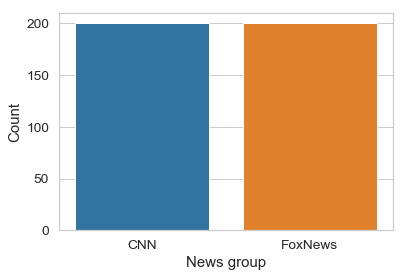
People often gravitate to those with similar beliefs, which may further polarize an already established belief structure. Furthermore, in 2015 Microsoft released a study indicating that the average human attention span in 2000 shortened from 12 seconds to just 8 seconds in 2013. A well-known idiom states, "Road to hell is paved with good intentions." With the advent of technology, individuals must carefully weigh information from all possible sources. The 21st century's social interactions have allowed social media to be a focal point of social exchange theory. Concepts of maximized rewards include receiving positive remarks from like-minded individuals while minimizing cost entail the least effort of information retrieval..

**Data prepation**

Data tweets in CNN and fox news from Jan to March of 2020. Removing special characters, utilizing stop word English. With the size 199 count each.

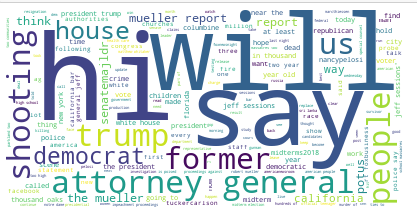
**Results**

A word cloud for the CNN media indicates a dispersion of words heavily related to the Mueller Report and President Trump.



**World Cloud**

**CNN and Fox News CNN News**

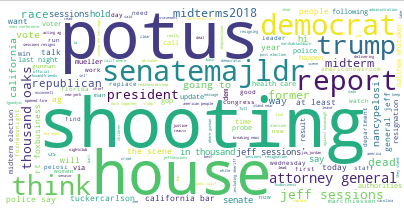


The combine word most common

words is hi will say, attorney general and shooting

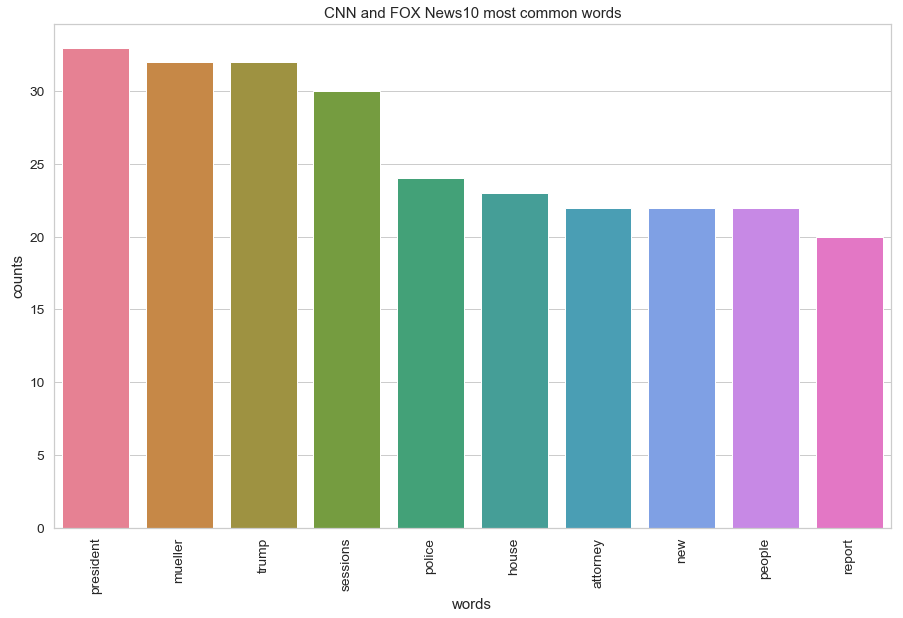
The CNN news talk over and over about muller report

**FOX News**

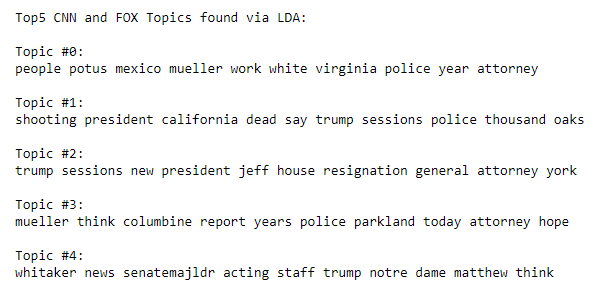


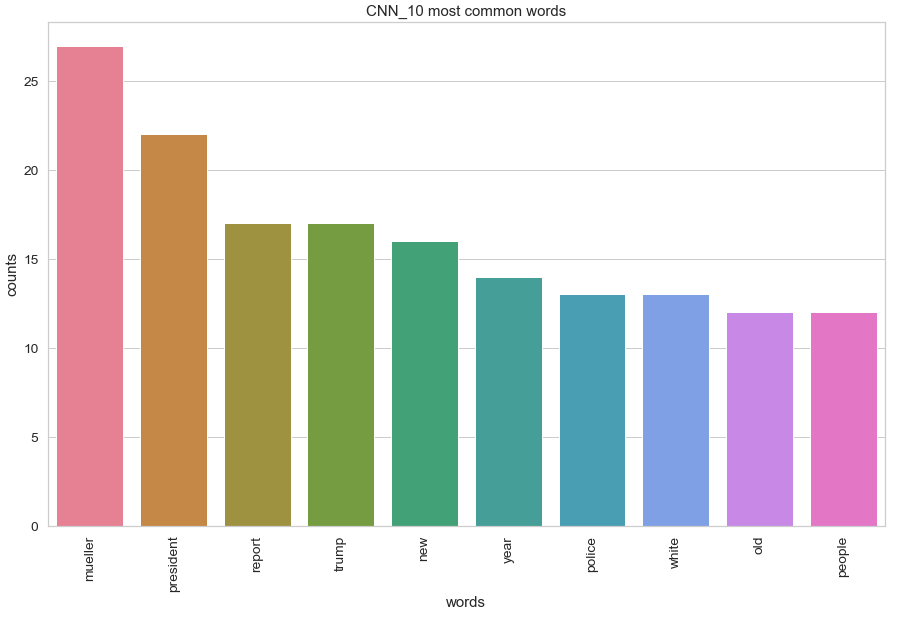
FOX talks about the potus, shooting that is happening and the house

**Top 10 Words**

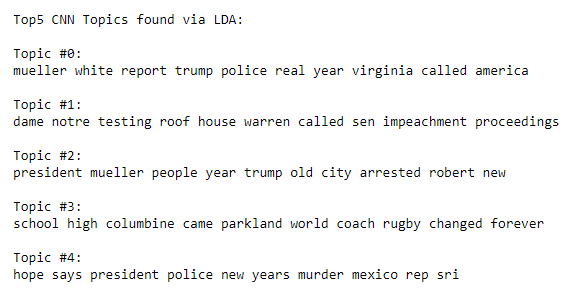


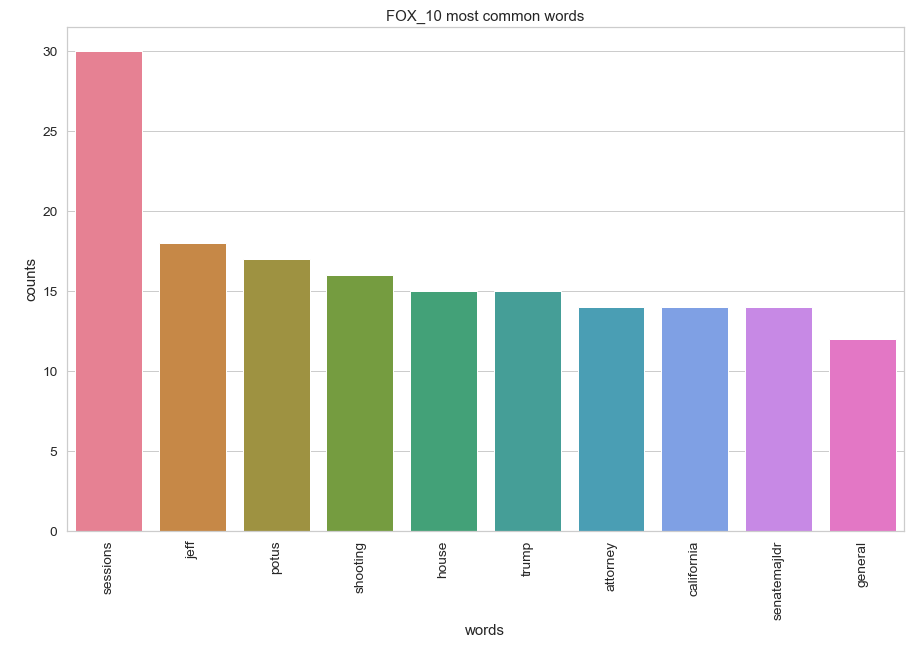
Presedent was mention about 28 times, muller and trump mentioned about the same then the sessions



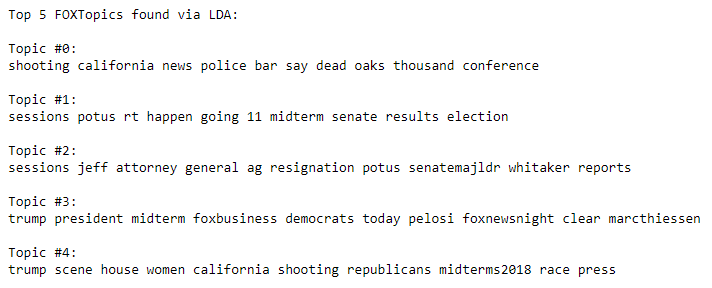


CNN talks more about muller 28 times, president report must be muller report then about trump.



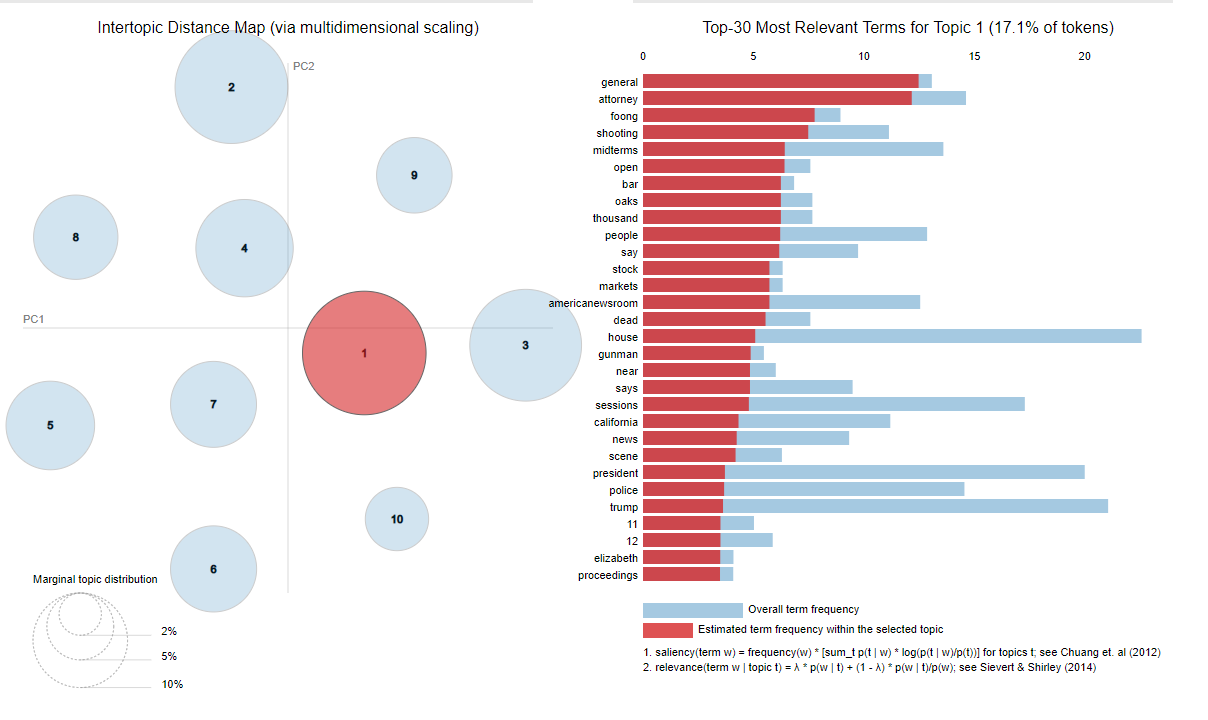


The most topic for FOX news is the sessions, then jeff, putus and shooting.



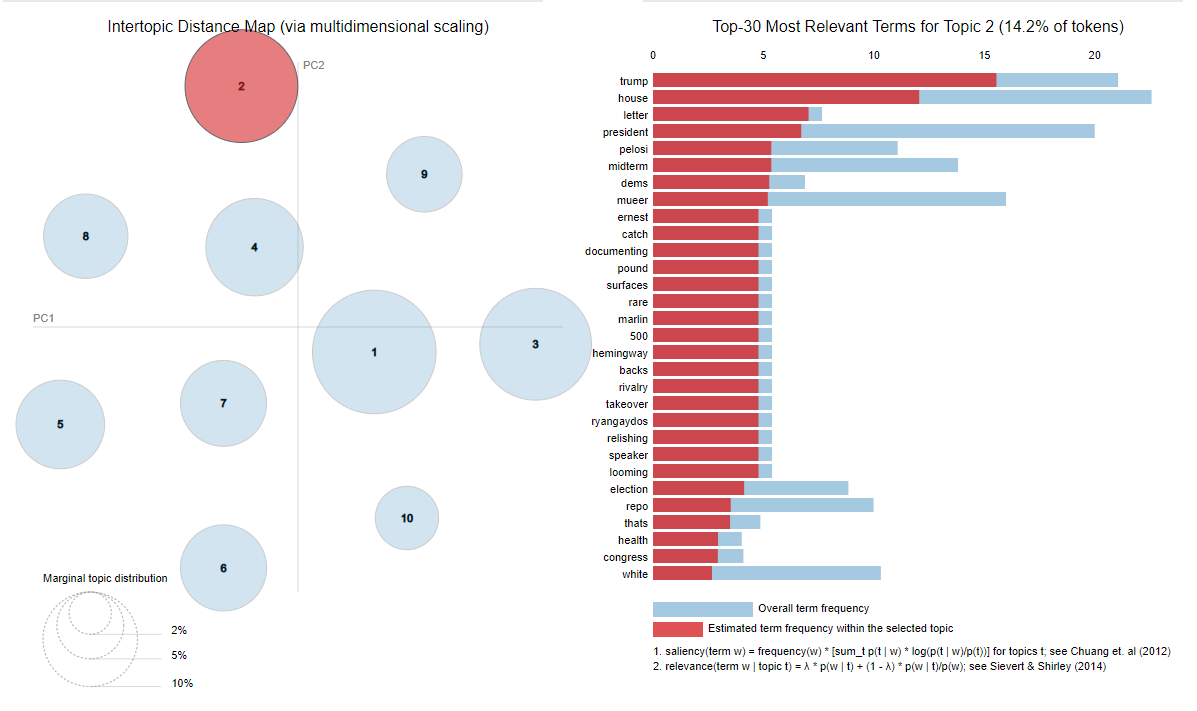
Each bubble represents a topic. The larger the bubble, the more prevalent is that topic.

**ALL Data CNN and FOx News**

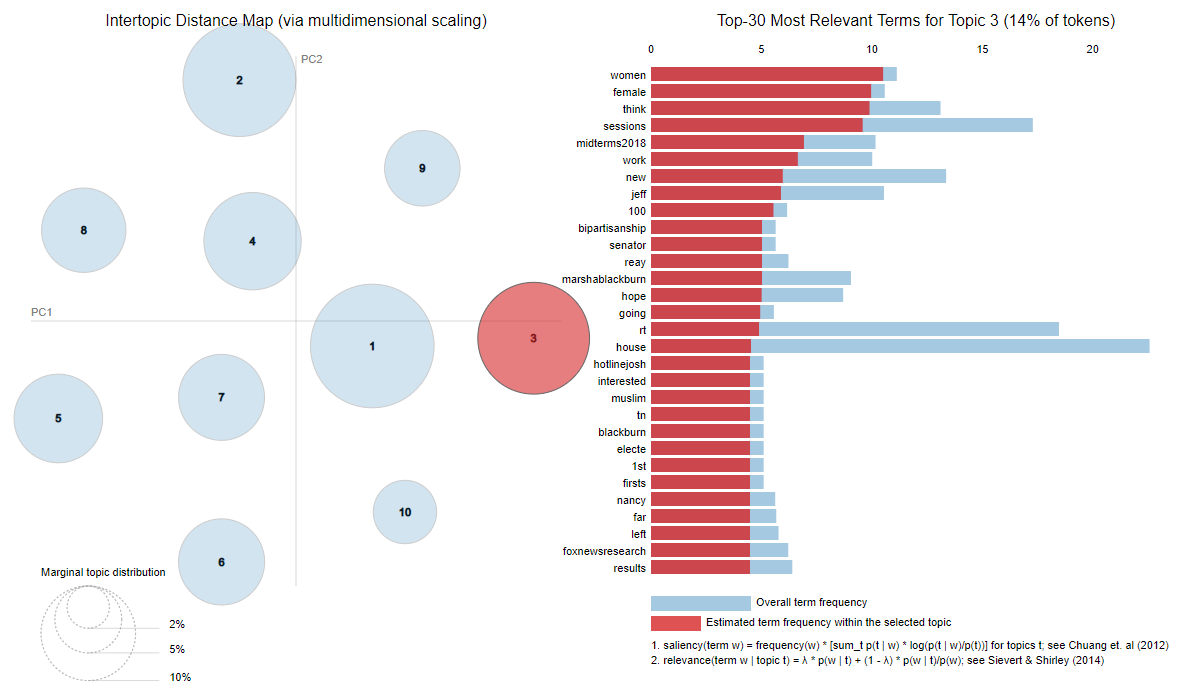


The topic about the attorney general and shooting, midterm, open which is open border concern

one with 17.1%



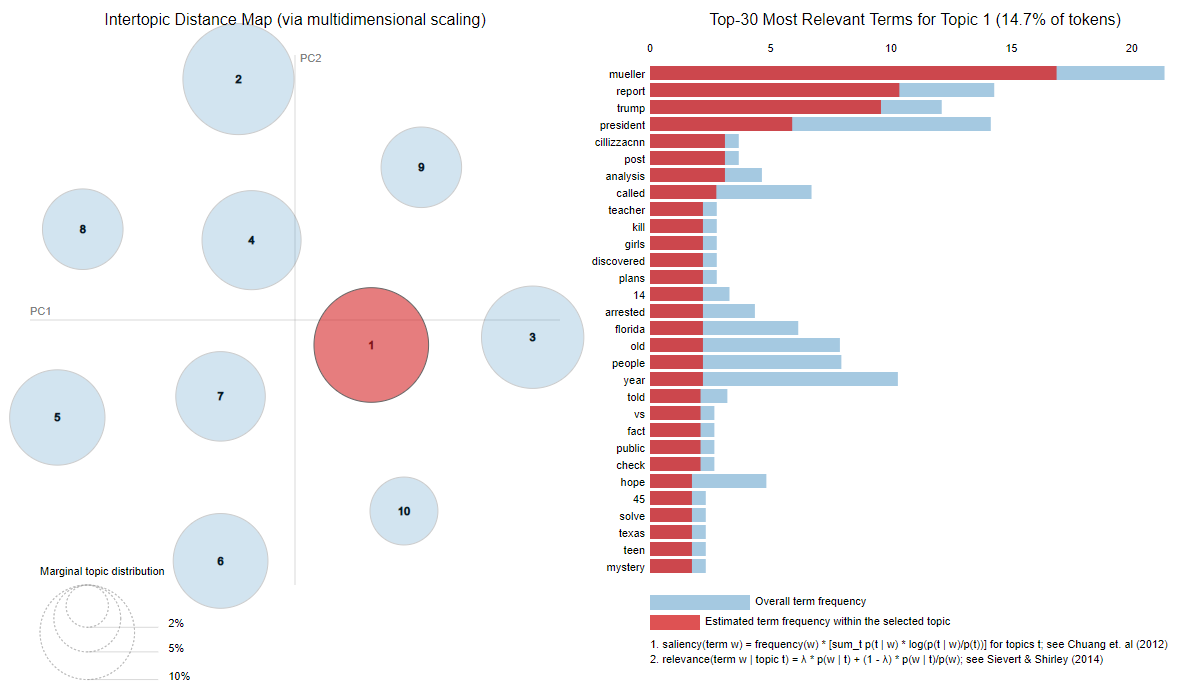
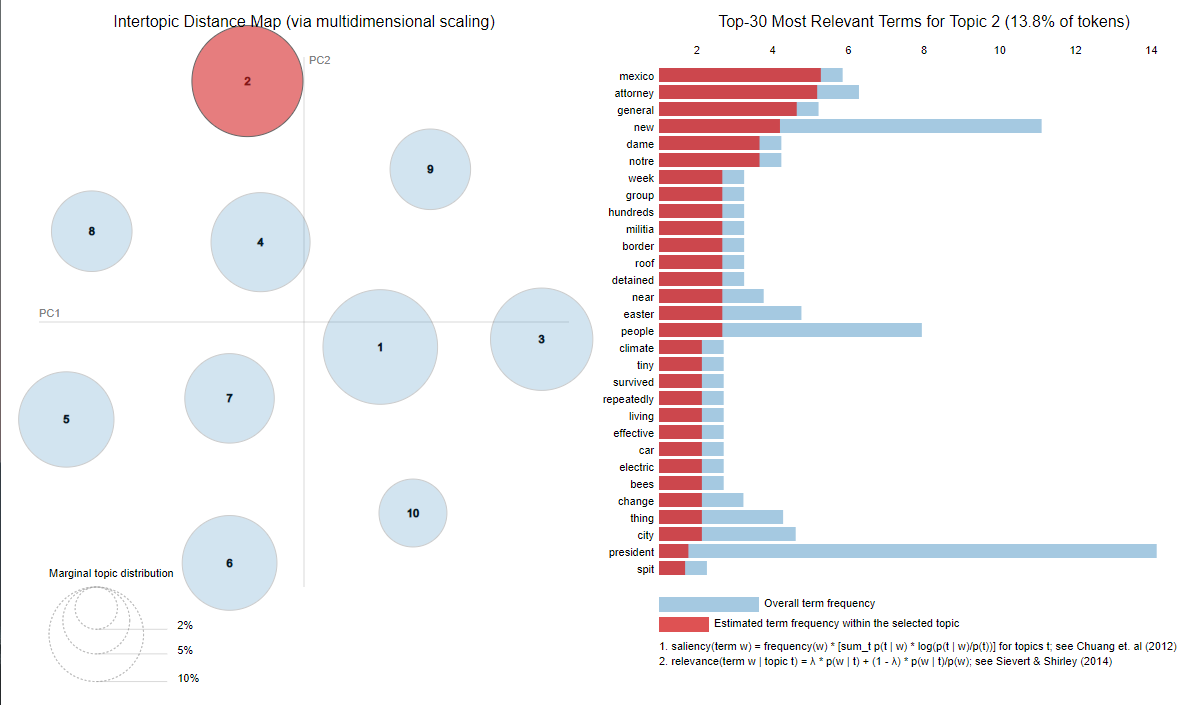
Talks about president trump, the house, letter, president midterm and more, with 14.2% most relevant term of the topic.

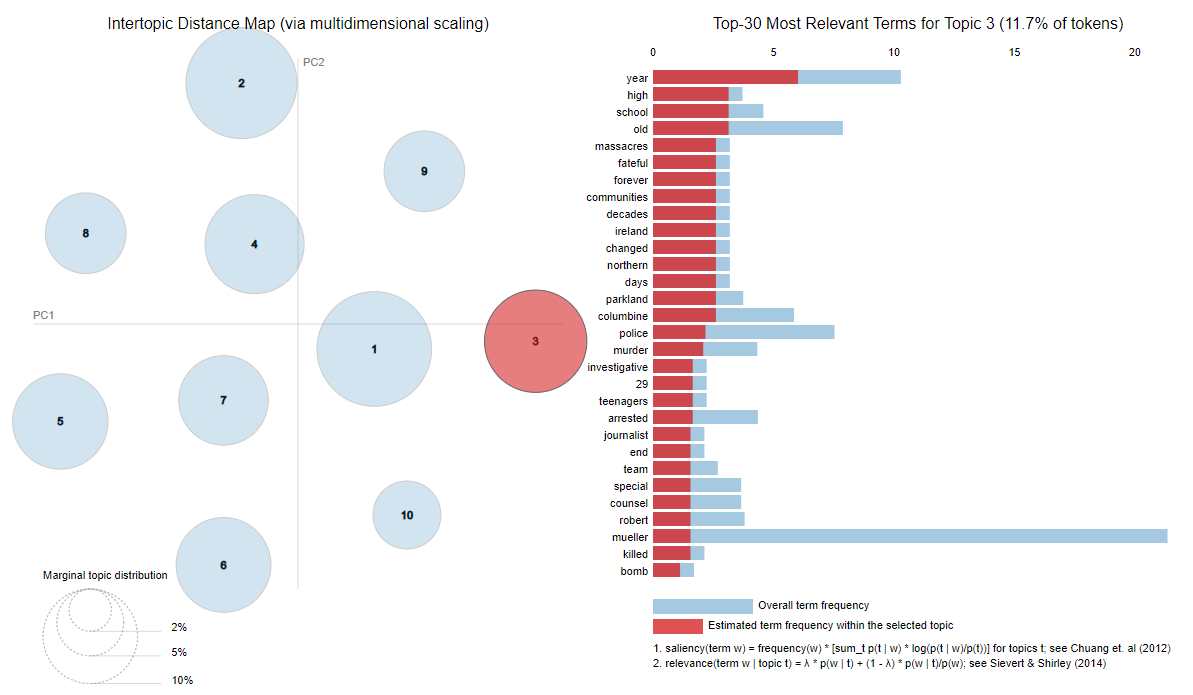


Third topic talks about the women, female, sessions, midterm with 14% of most relevant term.

Overall talks about the general attorney shooting, midterm with topic one 17% most relevant term. The topic 1 is more closer to number 3 with 14% compare to number 2 with 14.2% of most relevant topic

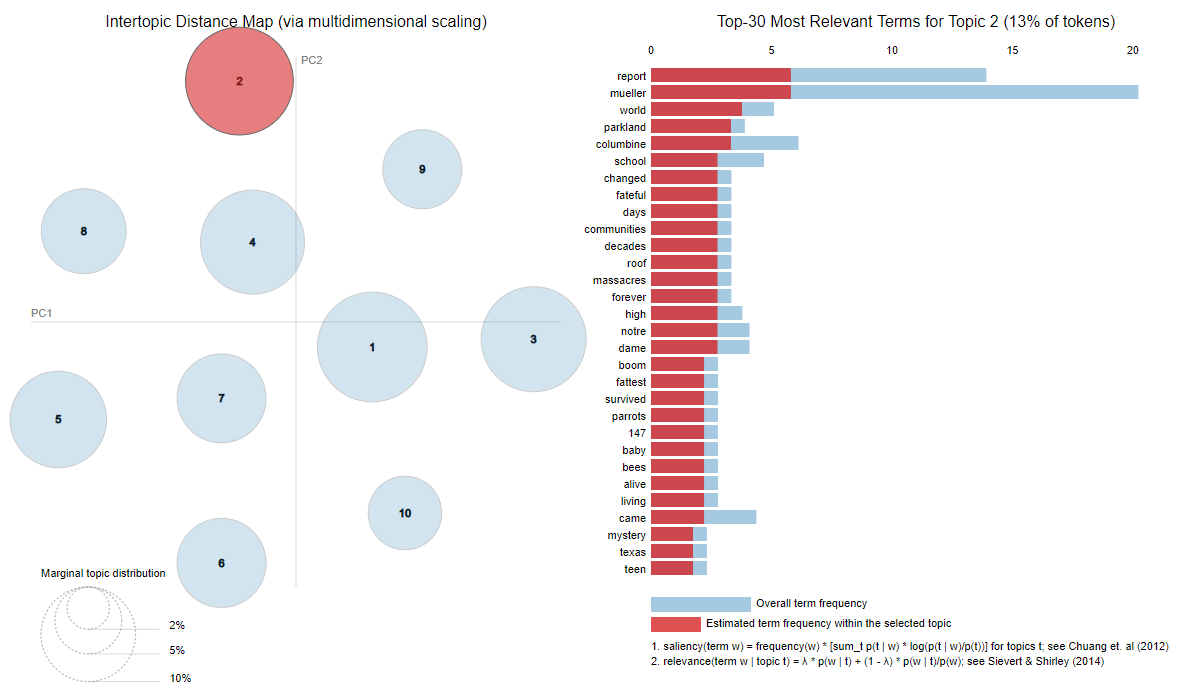
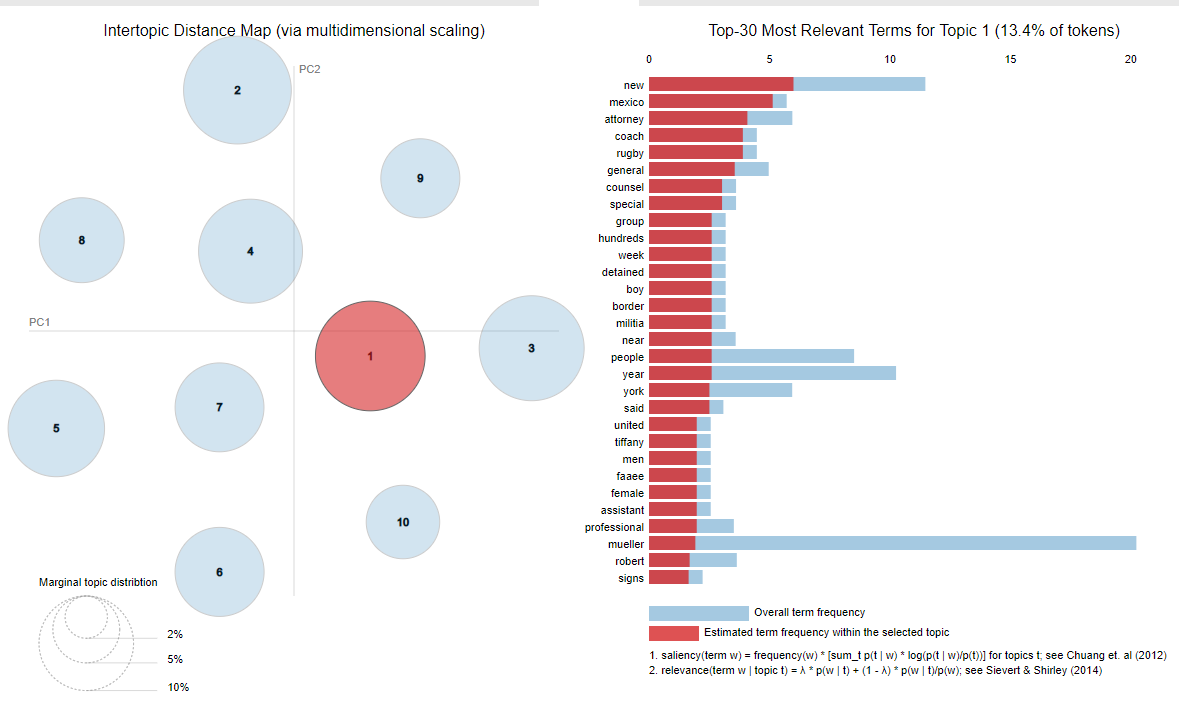
**CNN News**



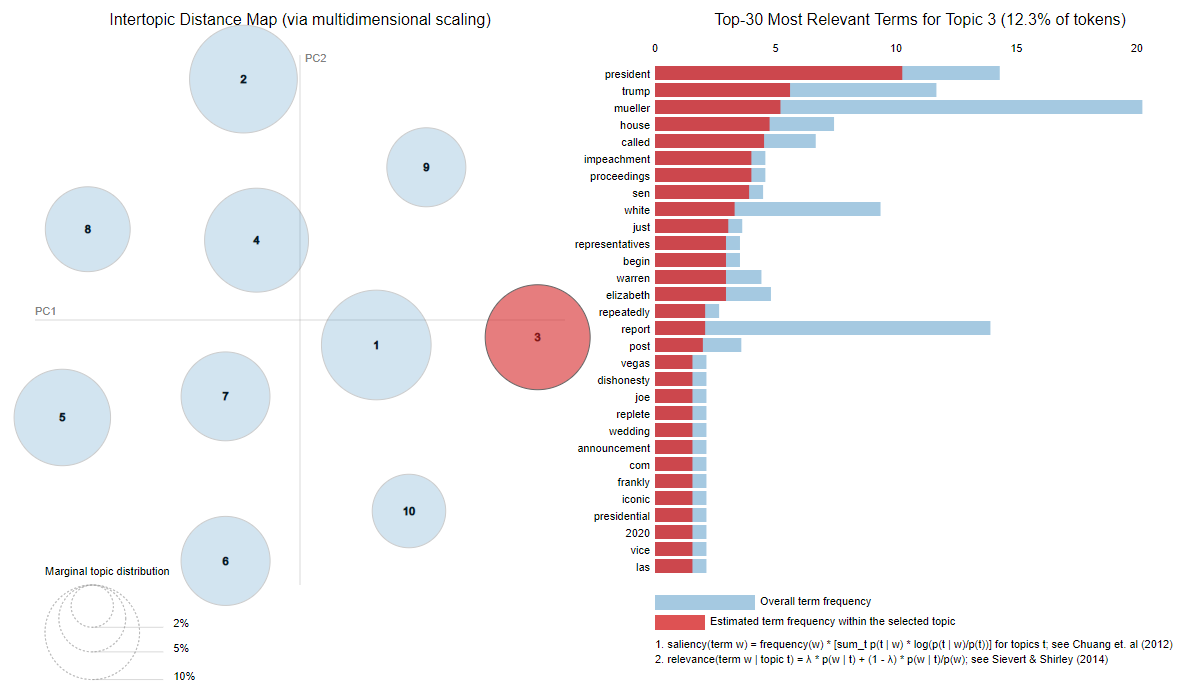
CNN news talks more about Muller report and the president with 14.7% on topic1 then topic 2 with 13.8% of the token. Next is topic 3 with 11.7 %

**Fox news**



FOX news topic one and topic two has almost the same relevance 13%, topic 3 have 12% one that is noticeable on the fox news has a positive notion compare to CNN news.

On first topic talks about the new Mexico attorney general second muller report and about the world, third pertaining to president trump, muller report and he house.



**Result Summary**

Data showed is a good topic modeling non-overlapping bubbles scattered throughout the chart instead of being clustered in one quadrant. There is no overlap.

The distance between the relevant term is the same spot. But have a different weight each topic of relevance into it. The different news media sources as agents with unique personalities. The latent Dirichlet allocation (LDA) a probabilistic topic modeling the different agents’ personality trait with respect to various topics or concepts. An agent’s personality traits affect its inclination to word a certain event in a specific way.

**Conclusion**

In a digitized world, information travels through many different mediums. The same information may reach the end-user from various sources, possibly causing information loss and even distortion. The same concept borrowed from network packets, fundamental to internet protocols, could apply to social theory. Multiple dimensions, including social media, the internet that connects users to social media, attention span, and interests of individuals and groups, and group theory vs. individualism, all play a similar pattern. Precisely, the potential bottlenecks in the social arena should portray the same stochastic tendencies as network packets.

Information theory contains a rich field of varying perspectives. Bottlenecks of information can throttle and impact the outlook of the mass. A short attention span may further limit an individual’s interest in acquiring information from credible sources. These shortcomings may encourage users to rely on information loss, obtained from general conversations of different social media. When enough individuals with a degree of these limitations share an interest, information loss and distortion is more likely to occur.

Furthermore, media outlets like CNN and Fox News often provide differing viewpoints. Each tends to have a different fundamental ideology. To be mindful of the current event is important to know the limitations and degree of perspective from various sources. Having an open interpretation (at least initially), and not marrying one ideology is sometimes the right approach. A curious stance, especially to lower degree information, is essential. What exactly is lower degree information? It is information that has traversed one or many nodes.

Furthermore, the concept is similarly genuine with social sciences. Each time information traverses an individual or group of individuals, and the original message is subject to greater distortion. This traversal between individuals or groups of individuals decreases the degree of unique information.