**Syracuse University**

**IST-736 Assignment 8**

ThulasiRam RuppaKrishnan

IST 736

Professor Ami Gates

Contents

[Introduction 3](#_Toc18014617)

[Analysis and Models 6](#_Toc18014618)

[**About the data** 6](#_Toc18014619)

[**Models** 7](#_Toc18014620)

[Results 12](#_Toc18014621)

[Conclusion 16](#_Toc18014622)

## 

## **Introduction**

**Overview of the Legislative Process**

Article I of the U.S. Constitution grants all legislative powers to a bicameral Congress: A House of Representatives and a Senate that are the result of a “Great Compromise” seeking to balance the effects of popular majorities with the interests of the states. It currently provides for a two-year term of office for House members from the 435 population-based districts. In the Senate, voters of each state elect two Senators, who serve 6-year terms that overlap (such that only one-third of the chamber is up for election in any given election cycle).

The two chambers are fundamentally equal in their legislative roles and functions. Only the House can originate revenue legislation, and only the Senate confirms presidential nominations and approves treaties, but the enactment of law always requires both chambers to separately agree to the same bill in the same form before presenting it to the President.

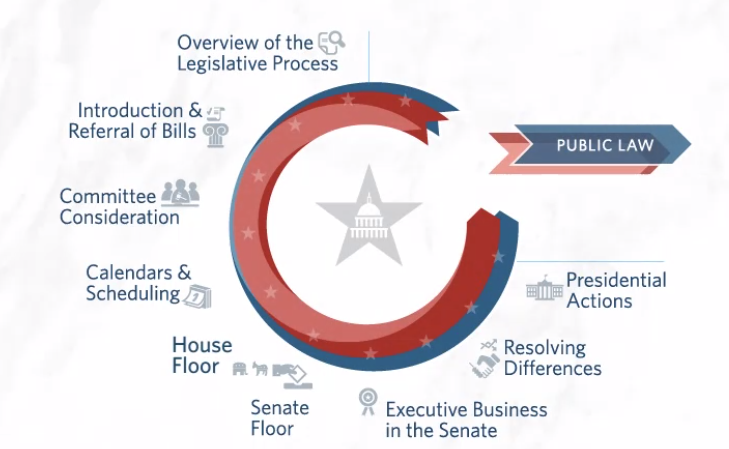
Because each chamber has the constitutional authority to make its own rules, the House and Senate have developed some very different ways of processing legislation, perhaps partially flowing from their constitutional differences. In general, House rules and practices allow a numerical majority to process legislation relatively quickly. Senate rules and procedures, on the other hand, favor deliberation over quick action, as they provide significant procedural leverage to individual Senators.

Congressional action is typically planned and coordinated by party leaders in each chamber, who have been chosen by members of their own caucus or conference – that is, the group of members in a chamber who share a party affiliation. Majority party leaders in the House have important powers and prerogatives to effectively set the policy agenda and decide which proposals will receive floor consideration. In the Senate, the leader of the majority party is generally expected to propose items for consideration, but formal tools that allow a numerical majority to act are few. Instead, majority party leadership typically must negotiate with minority party leaders (and often all Senators) to effectively conduct Senate floor action.

In both chambers, much of the policy expertise resides in the standing committees – panels of members from both parties that typically take the lead in developing and accessing legislation. Members typically serve on a small number of committees, often for many years, allowing them to become highly knowledgeable in certain policy areas. All committees are chaired by a member of the majority party, though chairs often work closely with the committee’s ranking member, the most senior member of the minority party on the committee. In almost all cases, the ratio of majority party to minority party members on a committee roughly reflects the overall partisan ratio in the congressional chamber.

Committee members and staff focus much of their time on drafting and considering legislative proposals, but committees engage in other activities, as well. Once law is enacted, Congress has the prerogative and responsibility to provide oversight of policy implementation, and its committees take the lead in this effort. Both chambers provide their committees with significant powers and latitude for oversight and investigations into questions of public policy and its effects. While the engine of legislative ideas and action is Congress itself, the President has influence in the legislative process, as well. The President recommends an annual budget for federal agencies and often suggests legislation. Perhaps more significantly, the power to veto legislation can affect the content of bills passed by Congress. Since it is quite unusual for law to be enacted over a presidential veto, Congress typically must accommodate the president’s position on proposed policies.

The process by which a bill becomes law is rarely predictable and can vary significantly from bill to bill. In fact, for many bills, the process will not follow the sequence of congressional stages that are often understood to make up the legislative process. The presentations on specific topics that follow present a more detailed look at each of the common stages through which a bill may move, but keep in mind that complications and variations abound in practice.



**Figure 1.1 Legislative process workflow**

**House Floor**

The House considers bills under a variety of procedures, each of which differs in the amount of time allotted for debate and the opportunities given to members to propose amendments. Most bills are considered under the suspension of the rules procedure, which limits debate to 40 minutes and does not allow amendments to be offered by members on the floor. However, for the House to pass a bill under suspension of the rules requires two-thirds of members voting to agree, so this method is not designed for bills that do not have supermajority support in the House.

Bills not considered on the House floor under suspension of the rules are typically considered instead under terms tailored for each bill. The House establishes these parameters on a case-by-case basis through the adoption of a simple House resolution called a special rule. Special rules are reported by the House Rules Committee. This committee, which is often referred to as the traffic cop of the House, is heavily dominated by the majority party, and works closely with House majority party leadership on the main elements of each special rule. Common provisions found in a special rule include selection of the text to be considered, limitations on debate, and limits on the amendments that can be offered on the floor. For instance, sometimes the committee reports a rule that places few restrictions at all on amending, which can result in dozens of amendments being offered on the floor during consideration. In other cases, the special rule will allow only specific pre-determined amendments to be offered, or even preclude floor amendments all together.

Note that House procedures place certain other limitations on the content of amendments, unless the special rule waives these restrictions. For instance, amendments must typically meet certain germaneness standards, meaning that they must be on the precise subject of the legislation being considered. After the Rules Committee reports a rule for consideration of a bill, the House first considers that special rule itself on the House floor, for approximately one hour. After debate, the House votes on adopting the special rule. Only after its adoption will the House proceed to consider the bill itself, under the terms specified by the special rule. In this situation, the House typically will consider the bill in a procedural setting called the Committee of the Whole, which allows members an efficient way to consider and vote on amendments. After any amendments are offered and debated, members vote on approval, and each amendment requires a simple majority to be agreed to.

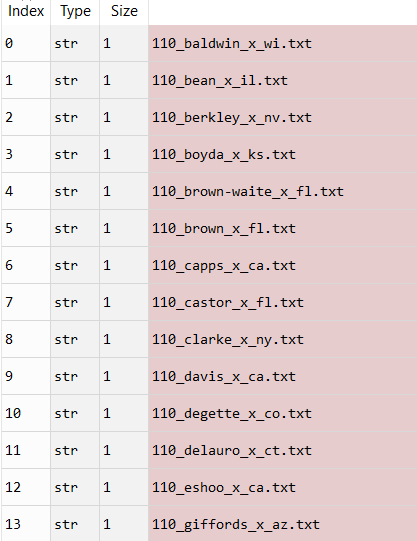
After the amendment process is complete, the Committee of the Whole rises and reports to the full House any recommended amendments, which are then usually approved by the House by voice vote. Just prior to voting on final passage, members typically will briefly debate and then vote on a motion to recommit, which allows the minority party to effectively propose its own amendment. In the House, some votes are taken by voice, but many votes are taken by electronic device, a method that records the individual position of each member who voted.

Reference: <https://www.congress.gov/legislative-process/house-floor>

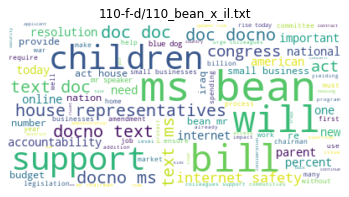
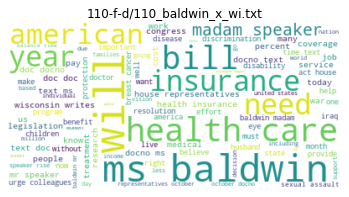
## **Analysis and Models**

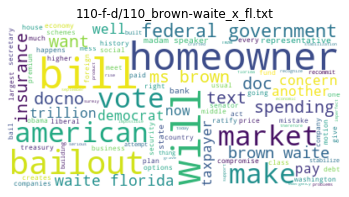
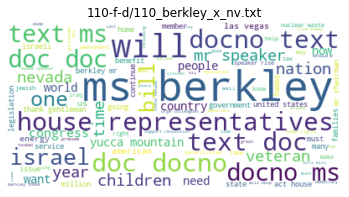
### **About the data**

The dataset used are from floor debate of the 110th Congress (House only). According to political scientists, there are usually 40-50 common topics going on in each Congress. In this exercise, 10 topics are modelled from these documents. Figure 2.1 and Figure 2.2 shows the overview of this sample dataset.



**Figure 2.1 Sample debate file from 110th Congress**





**Figure 2.2 Sample word cloud from 110th Congress debate**

### **Models**

Topic modeling is the process of identifying topics in a set of documents. This can be useful for search engines, customer service automation, and any other instance where knowing the topics of documents is important. There are multiple methods of going about doing this and one of them is **Latent Dirichlet Allocation** (LDA).

**The Algorithm**

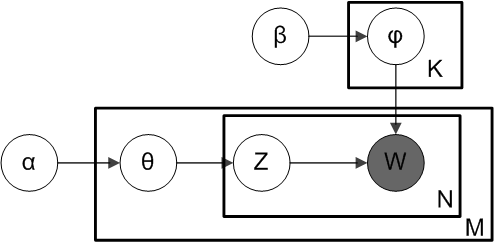
LDA is a form of unsupervised learning that views documents as bags of words (i.e. order does not matter). LDA works by first making a key assumption: the way a document was generated was by picking a set of topics and then for each topic picking a set of words and then reverse engineers this process.

To do this it does the following for each document m:

* Assume there are k topics across all the documents
* Distribute these k topics across document m (this distribution is known as α and can be symmetric or asymmetric, more on this later) by assigning each word a topic.
* For each word w in document m, assume its topic is wrong but every other word is assigned the correct topic.
* Probabilistically assign word w a topic based on two things:
  + what topics are in document m
  + how many times word w has been assigned a topic across all of the documents (this distribution is called β, more on this later)
* Repeat this process several times for each document until it is done!

**The Model**

Smoothed LDA from <https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation>



**Figure 3.1 Plate diagram of an LDA Model**

Above is what is known as a plate diagram of an LDA model where:

* α is the per-document topic distributions,
* β is the per-topic word distribution,
* θ is the topic distribution for document m,
* φ is the word distribution for topic k,
* z is the topic for the n-th word in document m, and
* w is the specific word

**Tweaking the Model**

In the plate model diagram above, w is grayed out. This is because it is the only observable variable in the system while the others are latent. Because of this, to tweak the model there are a few things which are given below

α is a matrix where each row is a document and each column represent a topic. A value in row i and column j represents how likely document i contains topic j. A symmetric distribution would mean that each topic is evenly distributed throughout the document while an asymmetric distribution favors certain topics over others. This affects the starting point of the model and can be used when you have a rough idea of how the topics are distributed to improve results.

β is a matrix where each row represents a topic and each column represents a word. A value in row i and column j represents how likely that topic i contains word j. Usually each word is distributed evenly throughout the topic such that no topic is biased towards certain words. This can be exploited though in order to bias certain topics to favor certain words. For example, if there is a topic about Apple products it can be helpful to bias words like “iphone” and “ipad” for one of the topics in order to push the model towards finding that topic.

**Model 1.1: gensim.models.LdaMulticore on frequency**

This model is created by using Genism package and with the help of LDA multicore function applied on word frequencies and the below are the outcome of 10 topics that this model identified

**Model parameters are as follows**

* num\_topics=10,
* passes=2,
* random\_state=100,
* chunksize=10,
* alpha='symmetric',
* iterations=100,
* per\_word\_topics=True

**Topics are as follows**

Topic: 0

Words: 0.004\*"obey" + 0.003\*"physician" + 0.003\*"mckeon" + 0.003\*"hoyer" + 0.003\*"wilson" + 0.003\*"schip" + 0.003\*"jackson" + 0.002\*"scott" + 0.002\*"dingel" + 0.002\*"medicaid"

Topic: 1

Words: 0.005\*"coal" + 0.004\*"barrel" + 0.003\*"colombia" + 0.003\*"ethanol" + 0.003\*"flake" + 0.003\*"welch" + 0.003\*"schip" + 0.003\*"hensarl" + 0.002\*"crude" + 0.002\*"chart"

Topic: 2

Words: 0.003\*"oberstar" + 0.003\*"israel" + 0.003\*"arcuri" + 0.003\*"smith" + 0.002\*"schip" + 0.002\*"coal" + 0.002\*"blackburn" + 0.002\*"kagen" + 0.002\*"campbel" + 0.002\*"forest"

Topic: 3

Words: 0.005\*"agent" + 0.005\*"hoyer" + 0.005\*"coal" + 0.004\*"chart" + 0.003\*"barrel" + 0.003\*"westmoreland" + 0.003\*"physician" + 0.003\*"leas" + 0.003\*"heat" + 0.003\*"patrol"

Topic: 4

Words: 0.005\*"hast" + 0.003\*"foxx" + 0.003\*"bishop" + 0.003\*"israel" + 0.003\*"rahal" + 0.003\*"lake" + 0.003\*"wilson" + 0.002\*"leas" + 0.002\*"forest" + 0.002\*"kaptur"

Topic: 5

Words: 0.008\*"blunt" + 0.004\*"obey" + 0.003\*"gingrey" + 0.003\*"stem" + 0.003\*"barrel" + 0.003\*"ellison" + 0.003\*"chart" + 0.003\*"blackburn" + 0.003\*"schip" + 0.002\*"ryan"

Topic: 6

Words: 0.005\*"ryan" + 0.004\*"lewi" + 0.003\*"meek" + 0.003\*"grijalva" + 0.003\*"eddi" + 0.003\*"bernic" + 0.003\*"obey" + 0.002\*"smith" + 0.002\*"capp" + 0.002\*"darfur"

Topic: 7

Words: 0.006\*"jackson" + 0.005\*"murphi" + 0.004\*"conyer" + 0.003\*"aid" + 0.002\*"schip" + 0.002\*"meek" + 0.002\*"altmir" + 0.002\*"maloney" + 0.002\*"foreclosur" + 0.002\*"stem"

Topic: 8

Words: 0.005\*"dreier" + 0.004\*"stem" + 0.004\*"occup" + 0.003\*"woolsey" + 0.003\*"pallon" + 0.003\*"oberstar" + 0.002\*"barrel" + 0.002\*"contractor" + 0.002\*"rail" + 0.002\*"thompson"

Topic: 9

Words: 0.005\*"israel" + 0.004\*"hunter" + 0.003\*"skelton" + 0.003\*"rail" + 0.002\*"stem" + 0.002\*"lake" + 0.002\*"abort" + 0.002\*"mccarthi" + 0.002\*"hare" + 0.002\*"missil"

**Model 1.2: gensim.models.LdaMulticore on TFIDF**

This model is created by using Genism package and with the help of LDA multicore function applied on word TFIDF and the below are the outcome of 10 topics that this model identified

**Model parameters are as follows**

* num\_topics=10,
* passes=2,
* random\_state=100,
* chunksize=10,
* alpha='symmetric',
* iterations=100,
* per\_word\_topics=True

**Topics are as follows**

Topic: 0 Word: 0.002\*"skelton" + 0.001\*"rangel" + 0.001\*"hoyer" + 0.001\*"tanner" + 0.001\*"rey" + 0.001\*"jackson" + 0.001\*"stark" + 0.001\*"walsh" + 0.001\*"cardoza" + 0.001\*"blumenau"

Topic: 1 Word: 0.001\*"gingrey" + 0.001\*"ellison" + 0.001\*"arcuri" + 0.001\*"dick" + 0.001\*"doggett" + 0.001\*"hare" + 0.001\*"putnam" + 0.001\*"tribe" + 0.001\*"cuellar" + 0.001\*"neal"

Topic: 2 Word: 0.002\*"filner" + 0.002\*"boozman" + 0.001\*"clay" + 0.001\*"lowey" + 0.001\*"knollenberg" + 0.001\*"sutton" + 0.001\*"pascrel" + 0.001\*"issa" + 0.001\*"defazio" + 0.001\*"gifford"

Topic: 3 Word: 0.002\*"ryan" + 0.002\*"patent" + 0.002\*"schip" + 0.001\*"meek" + 0.001\*"murphi" + 0.001\*"colombia" + 0.001\*"hast" + 0.001\*"jone" + 0.001\*"smith" + 0.001\*"barton"

Topic: 4 Word: 0.002\*"hawaiian" + 0.001\*"meth" + 0.001\*"blunt" + 0.001\*"latourett" + 0.001\*"boustani" + 0.001\*"cantor" + 0.001\*"blackburn" + 0.001\*"andrew" + 0.001\*"eshoo" + 0.001\*"klein"

Topic: 5 Word: 0.002\*"roger" + 0.002\*"bishop" + 0.001\*"cramer" + 0.001\*"everett" + 0.001\*"hawaiian" + 0.001\*"lamborn" + 0.001\*"mccaul" + 0.001\*"emanuel" + 0.001\*"shadegg" + 0.001\*"biggert"

Topic: 6 Word: 0.002\*"obey" + 0.001\*"peterson" + 0.001\*"wamp" + 0.001\*"edward" + 0.001\*"flake" + 0.001\*"hensarl" + 0.001\*"welch" + 0.001\*"ross" + 0.001\*"boyda" + 0.001\*"mccreri"

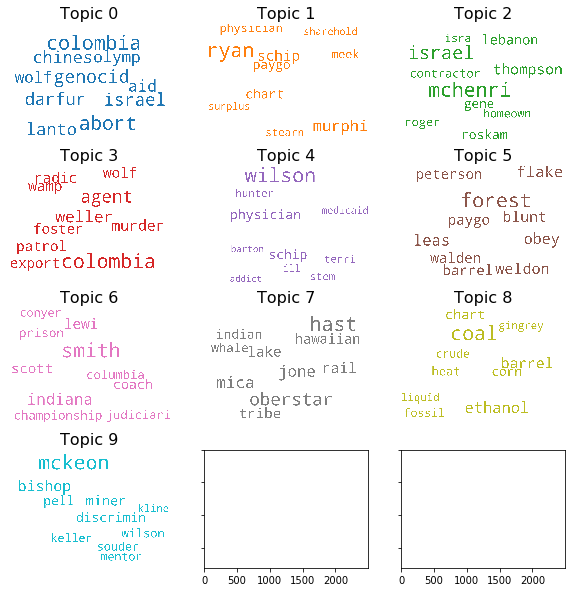
Topic: 7 Word: 0.002\*"rollcal" + 0.001\*"orlean" + 0.001\*"mccarthi" + 0.001\*"stem" + 0.001\*"platt" + 0.001\*"mcdonald" + 0.001\*"millend" + 0.001\*"boehner" + 0.001\*"ehler" + 0.001\*"mcgovern"

Topic: 8 Word: 0.001\*"israel" + 0.001\*"moran" + 0.001\*"campbel" + 0.001\*"wilson" + 0.001\*"saxton" + 0.001\*"lebanon" + 0.001\*"jordan" + 0.001\*"mckeon" + 0.001\*"chabot" + 0.001\*"kucinich"

Topic: 9 Word: 0.001\*"neugebau" + 0.001\*"baird" + 0.001\*"pearc" + 0.001\*"sire" + 0.001\*"grijalva" + 0.001\*"levin" + 0.001\*"frelinghuysen" + 0.001\*"tiahrt" + 0.001\*"mchugh" + 0.001\*"viscloski"

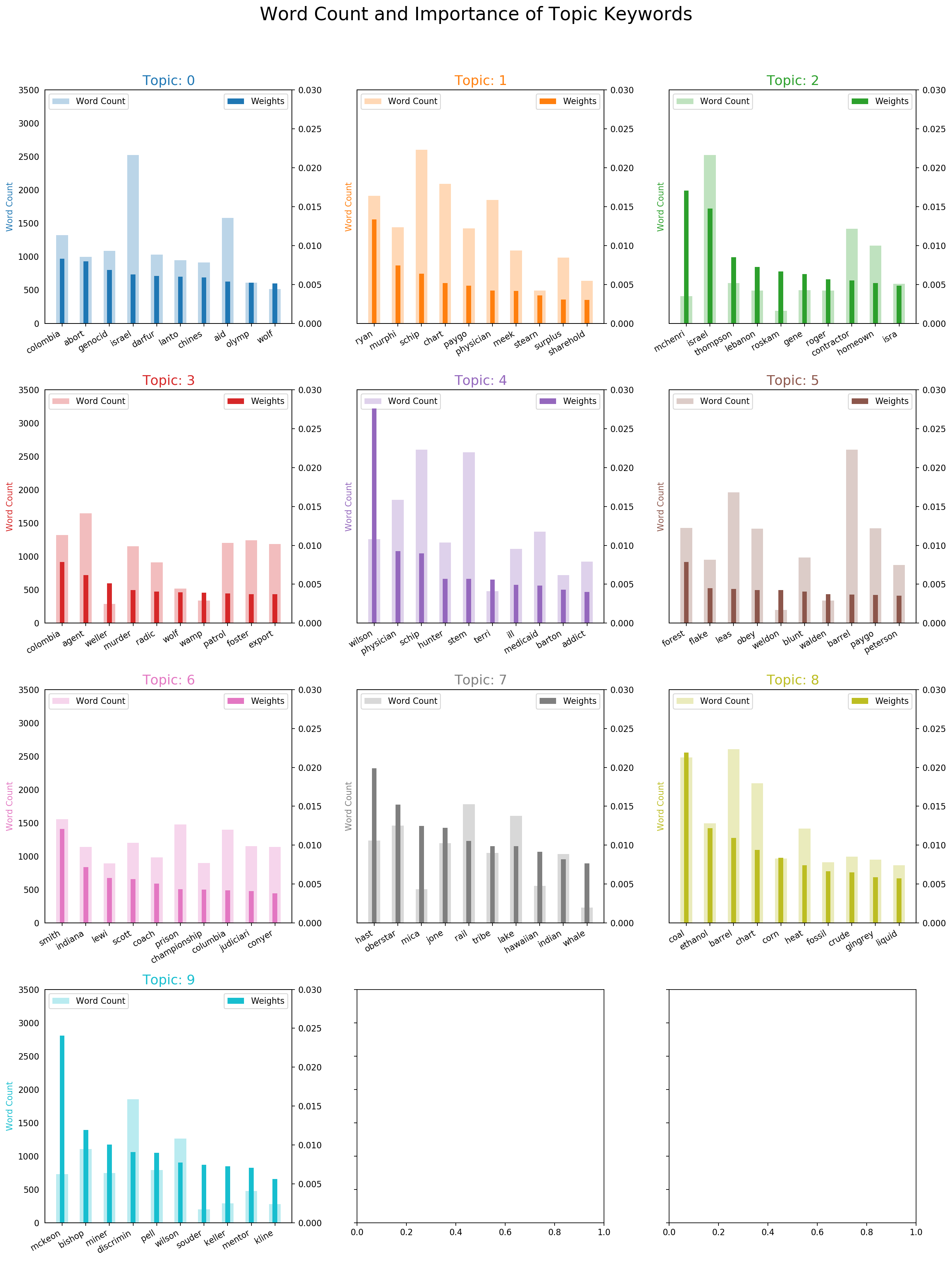
## **Results**

Word Clouds of top N keywords in each Topic are presented in the below **Figure 3.1**



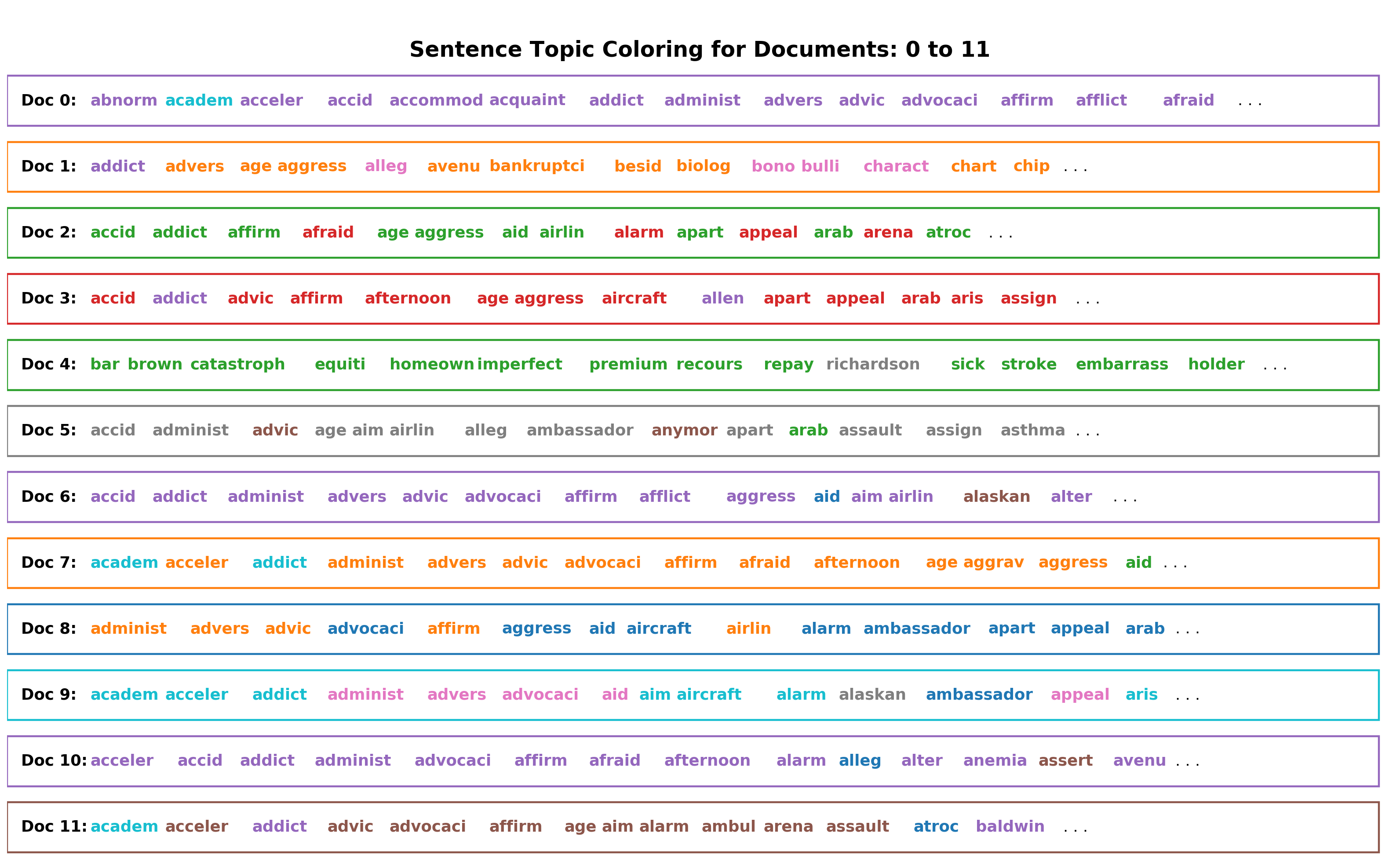
**Figure 3.1 Word Clouds of top N keywords in each Topic**

Word count of top N keywords in each Topic are presented in the below **Figure 3.2**



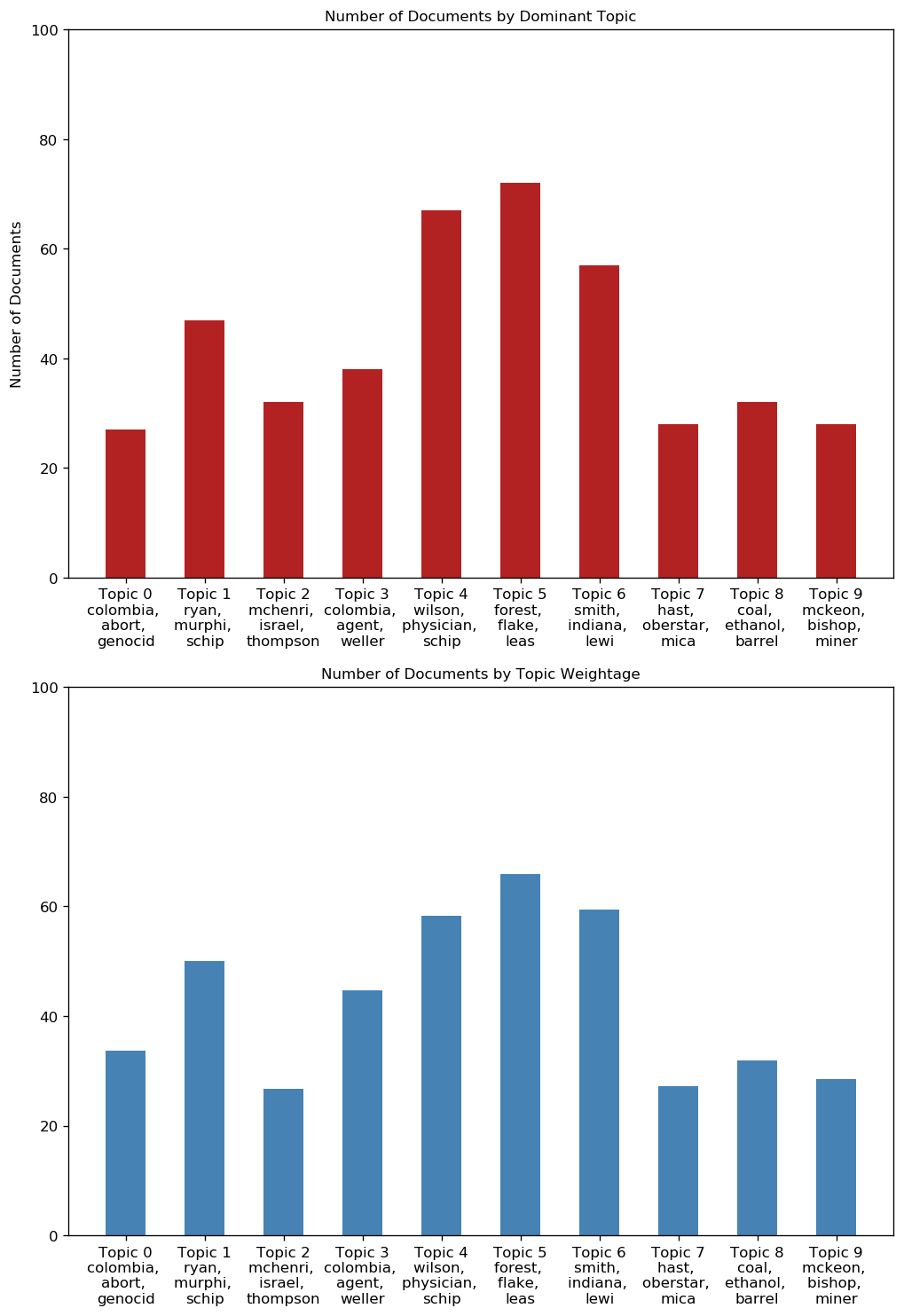
**Figure 3.2 Word count of top N keywords in each Topic**

Sample of 11 document are presented here and the tokens are colored relative to the topic in which each of these tokens holds high weightage



**Figure 3.3 Sentence topic coloring**

In the Figure 3.4 given below, it shows the number of documents by dominant topics and the number of documents by topic weightage

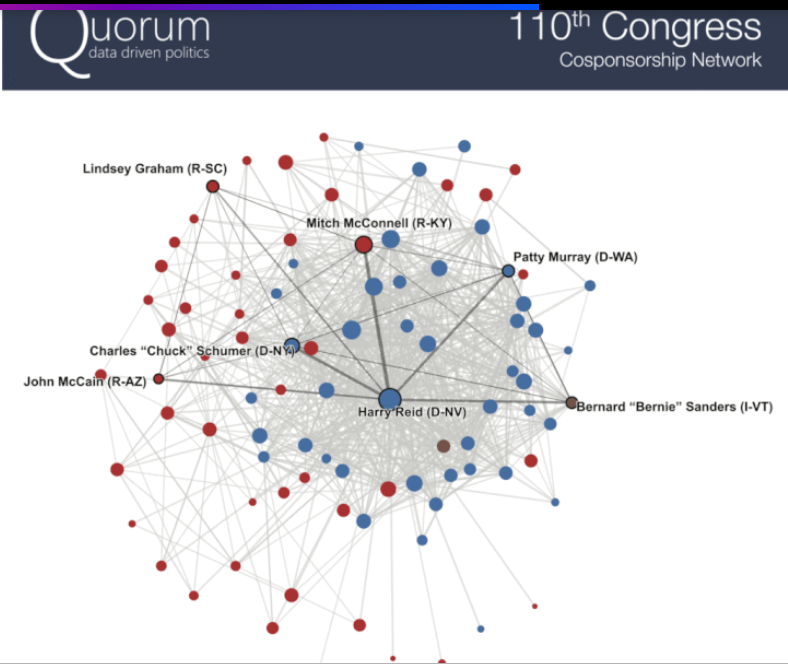


**Figure 3.4 documents by topic and its weightage**

## **Conclusion**

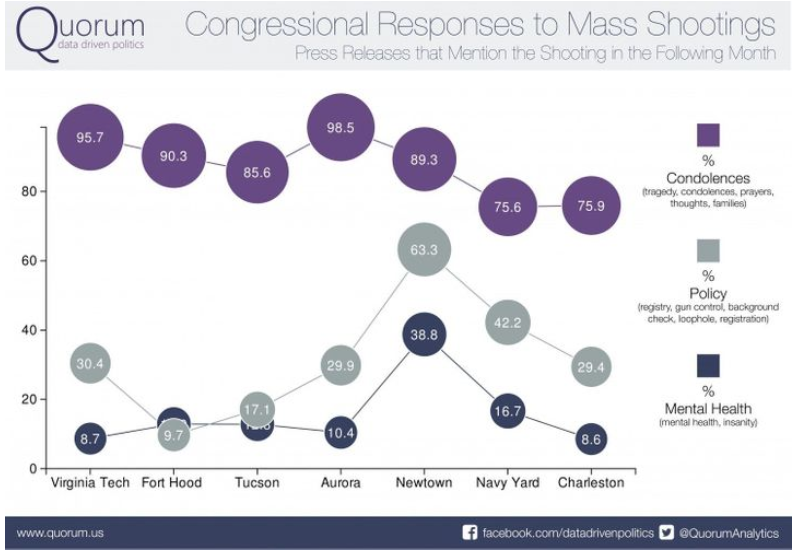
As the House of Representatives grew in number, however, revisions to the House rules limited debate. In the smaller Senate, unlimited debate continued the grounds that any senator should have the right to speak if necessary, on any issue.

In the last two years that Congress was in session, 11,059 bills were introduced, and only 4% of those became law. Most bills died in the committee of the chamber where it was introduced. 3724 were introduced in the Senate, and 7335 were introduced in the House. Another difference is who they represent. Senators represent their entire states, but members of the House represent individual districts. The number of districts in each state is determined by a state's population. Debate is nearly unlimited in the Senate and all members have an opportunity to influence legislation.



One of the interesting things that can be done by aggregating the debate data that’s there. “The Library of Congress goes through and tags every single bill related to a specific issue area. Finding topics based on the above model and determining which members are most active on a specific area, in terms of bill sponsored, bills cosponsored, or amendments adopted. Data analytics can also show how the power dynamics in Congress change from year to year This kind of data analysis can help lawmakers understand how efficiently their chamber is working and the metrics showing how influential or effective Congressmen are for the public to see

Quorum releases some of the insights from its data to the public, like this graphic on [how Congress reacts to mass shootings](https://www.quorum.us/blog/congress-reaction-mass-shootings):



Certainly, with someone like the House Majority Leader putting a bigger focus on data, people are beginning to talk about it differently and use the numbers to reference what’s going on in Congress,” Wirth said. “Obviously, our hope is that this becomes part of the ‘must-haves’ with communicating about how this Congress is different, measuring not just talking points but using data to prove it.

“The dream is that eventually we’d like to be able to bring [campaign finance data](https://www.huffpost.com/entry/federal-election-commission-data_559d8776e4b0967291556791) into the platform and put that together side by side with the legislative data, so that you could actually see and start to show how money has an influence on politics,” said Wirth.