Overview
The members of congress make decisions everyday that impact the lives of Americans yet there remains a disconnect between voters and those who represent them. The bills legislators write in congress are often long and complex. IBM Watson’s Natural Language Understanding component can quickly parse bill text and extract the important concepts, subjects and sentiment. This project uses keyword features extracted by Watson to analyze U.S. Senator behavior and present it in a simplified summary to aid in voter understanding of their representatives.

Congressional Data
All modern records that belong to the U.S. Congress are publicly available through official U.S. government websites. There is open source software that enables scraping and formatting of the data. This project scraped and stores legislator, full bill details and sponsorship records from the 113th and 114th U.S. Senate. Data is stored in a local MongoDB for fast retrieval.

Feature Extraction
IBM Watson services are easily available through IBM’s Bluemix service. This project use the Natural Language Understanding Service to parse full bill text and extract the key features that represent bills.

Keyword Vectors
The issues that are most important to senators are encoded in the keywords of bills that they sponsor. For each senator, a single weighted keyword vector is constructed from all of the bills they have previously supported, where the weight is the number of bills a keyword has appeared in.

Visualization:
By leveraging and comparing keyword vectors we can visualize and present simplified information about bills and legislators as seen in figures 2 and 3.

Comparing Keyword Vectors
A benefit of encoding bills and senator behavior into keyword vectors is that we can measure the distance between them to determine how similar/different they are. This allows us to compare bills to bills, legislators to legislators and legislators to bills. The comparison between keyword vectors can be achieved with any vector distance metric. For this system we use a weighted distance score that is the sum of all common key words minus the weighted sum of all disjoint keywords. Disjoint keywords are weighted 0.02 times less than matches. The higher the score, the more similar the bill/legislator. Making these comparisons enables us to estimate which legislators are most similar or different to each other.

Conclusion
The data derived from bill text and legislator sponsorship of these bills can reveal information about the main issues senators support. Through Watson NLU keywords can be extracted from bills and placed into keyword vectors to simplify the representation of issues import to bills and legislators in general. This new representation can be represented in a graphical form that makes the data more accessible at a glance.