Mining economic policy issues in the agendas of presidential candidates of the US

by
Alibek Korganbekov

Submitted to
Central European University
Department of Economics

In Partial Fulfillment of the Requirements for the
degree of
Master of Arts

Supervisor: Professor Rosario N. Mantegna

Budapest, Hungary

2016

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Rosario Mantegna and Professor Luca Marotta for their support.

ABSTRACT

This paper looks into economic policy issues in the agendas of presidential candidates of the US. By applying text mining tools author tries to define differences between competitors. Paper provides insights into the word structures, readability and sentiment types in texts of the candidates.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Barack Obama, current president of the United States of America, is famous for his rhetorical skills. His abilities are often compared to those of his predecessors, great master orators like Lincoln, Roosevelt, Kennedy and others. Donald Trump, a candidate for 2016 presidential elections, on the other hand, is considered, at least, different. A number of analyses on his speech patterns and usage of words noted his incoherent style and, at the same time, high level of readability. Though very different, both of them are highly popular in the US.

Obviously, personal charisma plays a great role in nominee's popularity. Though there is a number of papers on measuring one's charisma, it's beyond the scope of this work. Language, on the other hand, is arguably the only tool of any presidential candidate to communicate with electorate. During current and previous two elections Americans named economic problems as the most important factor to their vote (Casselman, 2015). It's debatable, whether those with right policy choices and ways of addressing issues in text win the elections. Thus, analysis of these texts can be helpful to derive differences between successful candidates and runners-up.

Defined as a process of processing text-based content to get valuable insights from it, text mining is being mostly developed for business applications. The biggest examples are companies like Facebook and Twitter, where users create text content on everyday basis. Social networks try to utilize this information with goals such as understanding user's posts, deriving patterns in her behaviour and personalizing her content. Since this data is in open access it is widely used for deriving trends, checking correlations with stock market and measuring public opinion (O'Connor et al., 2010).

Grimer and Stewart (Grimmer and Stewart, 2013) define two methods of political text analysis basing on researcher's aims: ideological scaling and classification. Though authors argue, that goals of scaling are not properly articulated, one of the possible results of it is placing political actors in space basing on their texts. Researchers apply these algorithms to predict political positions

and party affiliation of different texts (Laver, Benoit, and Garry, 2003). Various data sources are used for such researches, to note most common: political news, electoral programs ('Manifesto Corpus') or speeches of presidents (for instance, see, 'The American Presidency Project').

This work offers an analysis of the most important text of a presidential candidate - agenda. As already mentioned, I look at those parts, which are related to economics. I apply classification methods, namely: simple statistical analysis of the text, its categorization, clustering, topic modelling and sentiment analysis. The time period of the research covers elections of 2004, 2008, 2012 and 2016. With an a posteriori knowledge of 3 of them, I try to see whether economic agendas of candidates were on point.

The paper is organised as follows. I start with a short description of presidential elections in the US and data. Third chapter reviews text mining tools used for this paper. In the fourth chapter I present the results, which are discussed in conclusion.

Chapter 2

AGENDAS OF PRESIDENTIAL CANDIDATES AS DATA SOURCE

Presidential elections occur every 4 years in the United States. Typically, candidates announce themselves 1.5 years prior to the elections, which marks the start of their campaign. First of all, candidates have to win the nomination of their party in the series of primary elections and caucuses. Primaries are held in each state and territory of the US. Main competitors for presidency has always been nominees of Democratic and Republican parties.

As a rule, each candidate publishes a program on different issues with policy proposals he or she aims to realize once being elected. A number of issues that a candidate might address varies. Nevertheless, economic problems are mainly the same and usually cover economy itself and topics like budget, income, labour market, taxes, education, healthcare, trade, immigration and Wall Street.

It was elections of 1996, when first ever websites for candidates' (Bill Clinton, Bob Dole) campaigns were launched[1]. Since then each nominee had a website with listed issues and plans of the candidate.

Though there are websites, which aggregate information on views of candidates basing on their publications, interviews, speeches and debates[2], they usually shorten it to bullet points. For this research, I use data from official campaign websites of nominees.

Agendas of elections of 2004, 2008 and 2012 were collected manually through 'Internet Archive' project[3]. Data on issues of candidates of 2016 was gathered from respective active sites of Hilary Clinton, Bernie Sanders and Donald Trump.

[1]Shields, Mike (February 18, 2016). "An Oral History of The First Presidential Campaign Websites in 1996", Wall Street Journal, accessed on May 25, 2016, http://www.wsj.com/articles/an-oral-history-of-the-first-presidential-campaign-websites-in-1996-1455831487

[2]For instance, see OnTheIssues.org, accessed on May 25, 2016, http://www.ontheissues.org

[3]Internet Archive Wayback Machine, accessed on May 12, 2016, http://archive.org/web/

It should be mentioned, that some candidates attached files with more details on their plans (for instance, John Kerry in 2004). Due to the big size of these files and assumption that most visitors read only the main pages of 'Issues', I did not include them in data. Furthermore, Donald Trump was the first one to use videos on his 'Issues' page and publish the details in text on a page called 'Positions'. Due to the small size of the transcripts of the videos related to economy (541 words), it was decided to keep texts from both pages.

In the Table 2.1 below one can find names of candidates and links to their campaign websites, which were used to collect data. Since it's not clear who will be the nominee of the Democratic party, both Clinton and Sanders were included.

Year of election	Democratic party	Republican party
2004	John Kerry (www.johnkerry.com/issues/)	George Bush (www.georgewbush.com/Agenda/)
2008	Barack Obama (www.barackobama.com/issues/)	John McCain (www.johnmccain.com/Informing/Issues/)
2012	Barack Obama (www.barackobama.com/record/)	Mitt Romney (www.mittromney.com/issues)
2016	Hilary Clinton (www.hillaryclinton.com/issues/), Bernie Sanders (berniesanders.com/issues/)	Donald Trump (www.donaldjtrump.com/issues) (www.donaldjtrump.com/positions)

Table 2.1: Agendas data and sources

Next chapter is dedicated to methods, which are applied to this data.

Chapter 3

CLASSIFICATION METHODS FOR MINING TEXTS OF AGENDAS

One of the first serious attempts of political texts mining began with the Manifesto Project in 1980s. The database includes information on political manifestos of more than 1000 parties from more than 50 countries starting from 1945[1]. All texts are coded sentence by sentence by political experts into one of 56 categories (Werner, Lacewell, and Volkens, 2011). Examples of these categories in economics domain are 'Free market economy', 'Market regulation', 'Protectionism: positive/negative', 'Nationalisation' and so on[2]. The project is praised for its volume and openness[3], and it's widely used in political science. At the same time, it was noted by both project and its critics, that human nature of encoding lacks reliability and can contain significant misclassification[4].

A significant number of empirical research on this dataset mainly uses idealogical scaling approach(Grimmer and Stewart, 2013). For this paper's data I apply so called classification methods for political texts. Analysis is performed in R with using special packages for data manipulation and text mining.

First of all, I start with preprocessing texts for analysis, which includes steps like removing special symbols, numbers, stopwords (articles, common words), stemming the text and getting stripping extra whitespace.

Analysis starts with basic statistics, like overall word count in the text, most frequent terms and measuring its readability. Note, that since the amount of total words in agendas differ, I use relative frequency measure, i.e. frequency of a term in agenda to total number of terms in agenda. Furthermore, readabil-

[1]The Manifesto Project, accessed on May 25, 2016, https://manifesto-project.wzb.eu/information/documents/information

[2]See "Manifesto Project Dataset" for complete list, accessed on May 25, 2016, https://manifestoproject.wzb.eu/down/documentation/codebook_MPDataset_MPDS2015a.pdf

[3]For example, it was recognized at "Lijphart/Przeworski/Verba Data Set Award" http://community.apsanet.org/comparativepolitics/awards/lijphartprzeworskiverbadatasetaward1

[4]See, for instance, Mikhaylov, Slava, Michael Laver, and Kenneth R. Benoit. "Coder reliability and misclassification in the human coding of party manifestos." Political Analysis 20, no. 1 (2012): 78-91.

ity is measured by commonly used Flesch–Kincaid reading ease score[5]. The formula of the score is represented below:

$$206.835 - 1.015 * (\frac{totalwords}{totalsentences}) - 84.6 * (\frac{totalsyllables}{totalwords}).$$

Scores range from 0 to 100, with those above 60 considered to be easy to read. The score is also assigned to a grade a reader has to get to fully understand the text, thus a grade above 12 assumes reader's completion of high school.

Second, I perform hierarchical clustering. It should be mentioned, that due to the large amount of words, first clusters had a lot of noise. Pre-installed in R methods are argued to be inappropriate for text data, thus Jaccard index was from 'Proxy' package was taken as measure of the distance (Berry and Castellanos, 2004). Clustering follows divisive approach and operates on dissimilarity matrix. To get more meaningful results, sparse terms were removed and bounds to filter low frequency words were applied to term document matrix.

Third step is to allocate topics to chosen documents, which is achieved by using topic modelling algorithm. In particular, I use currently most common and available Latent Dirichlet allocation approach. LDA is "three-level hierarchical Bayesian model, in which each item of a collection is modelled as a finite mixture over an underlying set of topics" (Blei, Ng, and Jordan, 2003). LDA uses iterative approach of recreating documents in the corpus and then adjusting relative importance of topics, which are "characterized by a distribution over words". The R package 'topicmodels' was used to produce the results.

The final step is sentiment analysis, which is used to get subjective information from texts. Usually, such algorithms try to identify the *polarity* of a text, its negative or positive tone. To achieve it, words are allocated to dictionaries and marked with the level of sentiments and emotions. There's a number of algorithms and dictionaries to perform such task, for instance NRC Word-Emotion Association Lexicon[6] (EmoLex). EmoLex covers 2 sentiments (positive, negative) with association scores of 0 or 1 and eight emotions

[5]Rudolf Flesch, "How to Write Plain English", accessed on May 20, 2016, http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml

[6]NRC Word-Emotion Association Lexicon, accessed June 1, 2016, http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

(anger, anticipation, disgust, fear, joy, sadness, surprise, trust), which has 4 association scores (not associated, weakly, moderately, or strongly associated) (Mohammad and Turney, 2010).

Though the above method provides a lot of opportunities, such programs are dedicated for specific texts like reviews of product. To perform sentiment analysis on a political text I apply a more conservative way and use Harvard's General Inquirer dictionary ((Wilson, Wiebe, and Hoffmann, 2005)).

This approach allows to measure positivity, negativity of the text, its subjectivity and polarity of the opinion. To be specific, positivity/negativity of the text is measured as total number of positive (p)/negative (n) sentiment references to total number of references. Polarity is calculated next way:

$$Polarity = \frac{p - n}{p + n}.$$

While subjectivity is a sum of positive and negative references to total number of references.

In the next chapter I show the results of the analysis.

Chapter 4

RESULTS OF TEXT MINING ANALYSIS

4.1 Statistics and readability

Statistics of agendas is represented in Table 4.1 below. Note that for Donald Trump both 'Positions' and 'Issues' were calculated, with latter shown in brackets.

Candidate	Total words number	Most frequent word	Readability score	Grade level
George Bush	11797	'president'	35.4	13.6
John Kerry	1662	'John'	46.9	11.1
Barack Obama	8757	'Obama'	36.8	12.8
John McCain	7360	'John'	44.7	11.2
Barack Obama	3046	'president'	46.1	11.3
Mitt Romney	5977	'federal'	40.5	11.7
Hilary Clinton	11291	'Hillary'	45	11.3
Bernie Sanders	12628	'tax'	50.8	10.3
Donald Trump	5864 (521)	'tax' ('jobs')	46.6 (77.9)	10.8 (5.2)

Table 4.1: Statistics of agendas by candidate

Average number of words of agendas is 7598 and standard deviation is 3866 (Donald Trump issues not included), the largest agenda text belongs to Bernie Sanders and smallest to John Kerry.

Both Bush and Obama used their position of 'president' very frequently, when running for the second term. Moreover, Obama's second agenda was much shorter than his first one. At the same time, John Kerry and John McCain used their first names the most often, while Barack Obama used his last name a lot while running for the first time. It should be noted, that John Kerry was running with John Edwards as vice-presidential candidate, which led to higher frequency. 'Federal' was the most frequent word in Mitt Romney's agenda, while 'tax' was mentioned the most by Bernie Sanders and Donald Trump. One can find most frequent words by candidate on Figure .1 of the Appendix.

Average readability score of agendas is 43.64, which is described as difficult to read document and requires at least ongoing college level of education. Except

for Donald Trump's videos, the most comprehensible text belongs to Bernie Sanders (50.8). The hardest one to understand was written for George Bush (35.4).

4.2 Hierarchical clustering

Results of hierarchical clustering for elections of 2004 are shown on Figure 4.1 and Figure 4.2.

Basing on this graph, one can see again, that "President Bush" often goes as a phrase in the document and, unsurprisingly, the same happens to "health" and "care". Taxes, on the other hand seem to be connected to "funding" and federal budget issues overall and, also, "american families".

Agenda of John Kerry, on the other hand, seems to focus on partnership of Kerry and Edwards and their plan. Once again, we see healthcare with close adjective "affordable", there is a criticism of "president", who is on the same branch with "cut" and "waste". Obviously, with a slogan "A Stronger America", Kerry was targeting national feelings, especially those of middle-class.

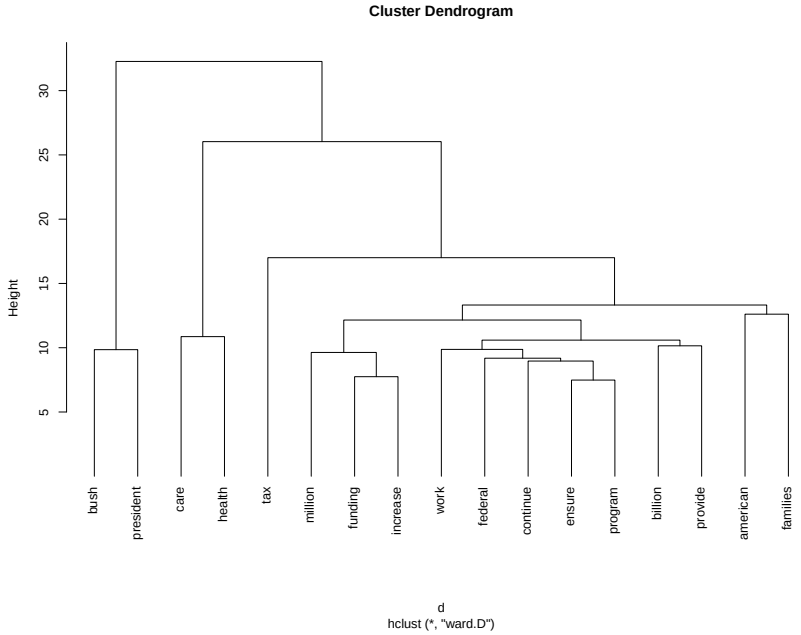


Figure 4.1: George Bush agenda dendrogram

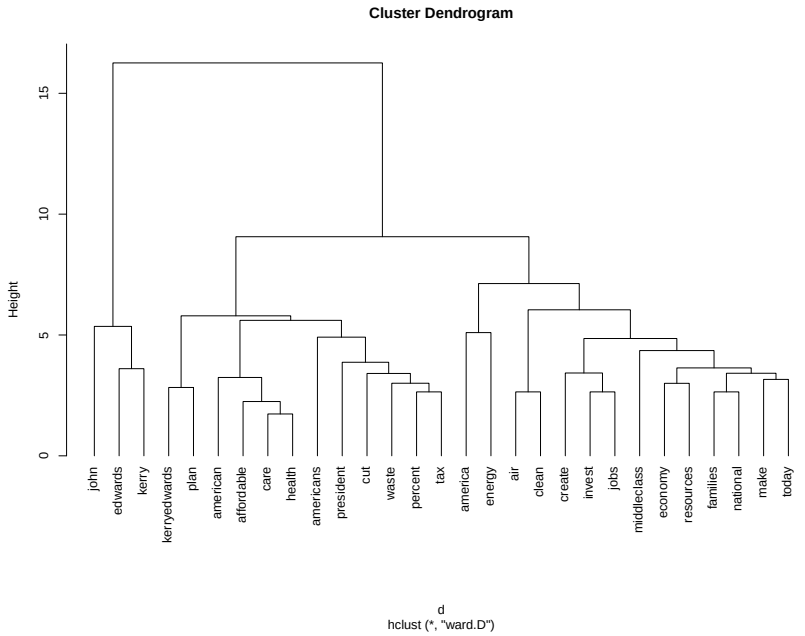


Figure 4.2: John Kerry agenda dendrogram

Let's look now at Obama's agendas of 2008 and 2012. As already mentioned, his second agenda had more than twice less words than his first one. I applied same options of filtering from noise to both trees, which can be found below on Figure 4.3 and Figure 4.4. As one can see, the second time Obama was much more laconic and focused on his "acts" or "actions" in healthcare (Obamacare reform of 2010) and jobs. Though Great Recession increased unemployment up to 10% (highest rate since 1983[1]), by 2012 elections he managed to decrease it to 8%.

In his 2008 agenda, Obama's name as we know was the most frequent and from the dendrogram one can see that it was often paired with his healthcare plans. Other than that, his agenda included points on energy, job programs, education and regulation of corporates.

Dendrograms of agendas of his respective competitors in 2008 and 2012 can be found in Appendix as Figure **??**. John McCain wasn't as wide on topics as Obama in 2008, he also focused on healthcare and proposed reforms of tax system. Mitt Romney, did not use his name as much as his party predecessor,

[1]Bureau of Labour Statistics, Databases, Tables & Calculators by Subject, accessed on June 2, 2016 http://data.bls.gov/pdq/SurveyOutputServlet

yet he targeted failures of Obama and was highly concerned about economic governance and federal spending.

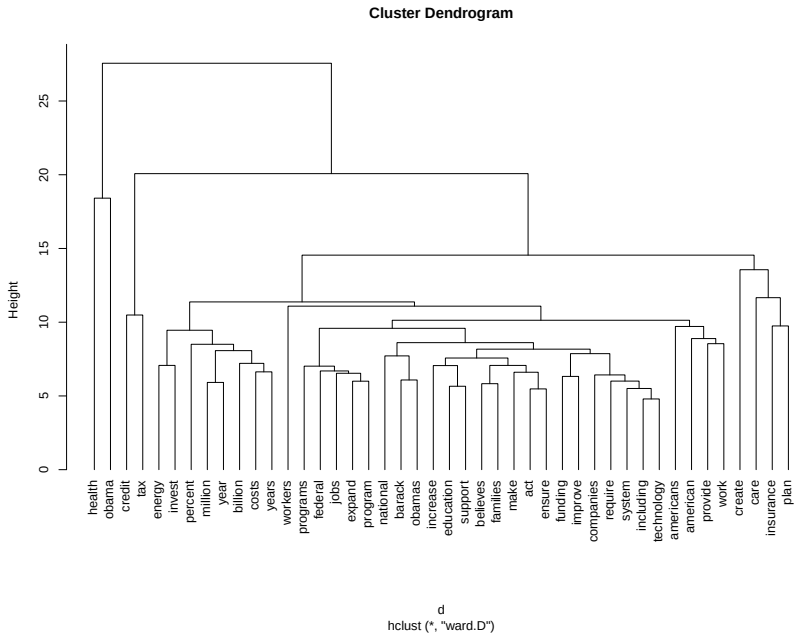


Figure 4.3: Barack Obama 2008 agenda dendrogram

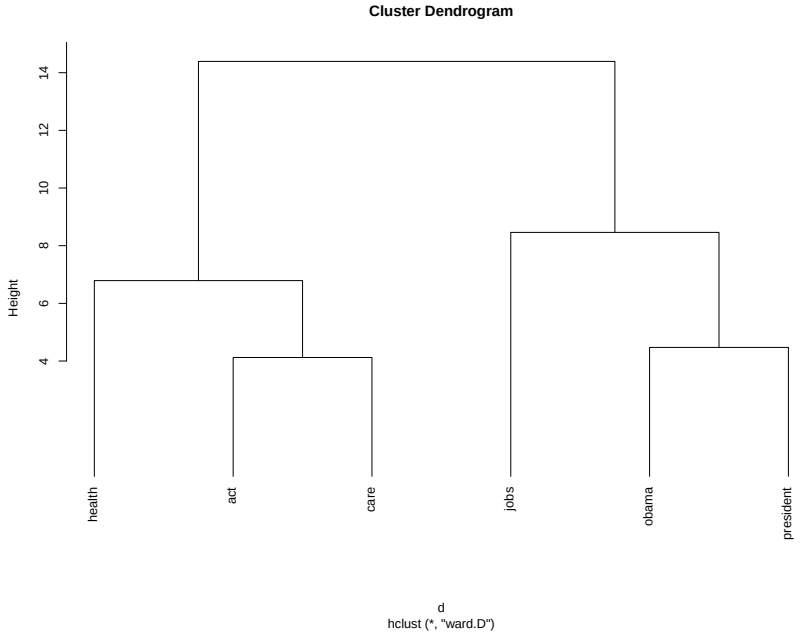


Figure 4.4: Barack Obama 2012 agenda dendrogram

In the elections of 2016 main competitors for nominations from Democratic

party are Hilary Clinton and Bernie Sanders[2]. In the Republican party, after withdrawal of senator Ted Cruz in the beginning of May, Donald Trump got the nomination.

The cluster tree of agenda of Bernie Sanders is represented on Figure 4.5. Healthcare is a big issue for him, while targeting the inequality, which he is the most famous for, is also visible on the dendrogram ('tax', 'million', 'income', 'corporation'). Obviously, 'Wall Street' is a big theme for him, which is also seen on the tree. Another big topic of his concerns is labour market, which is a big cluster on the right side.

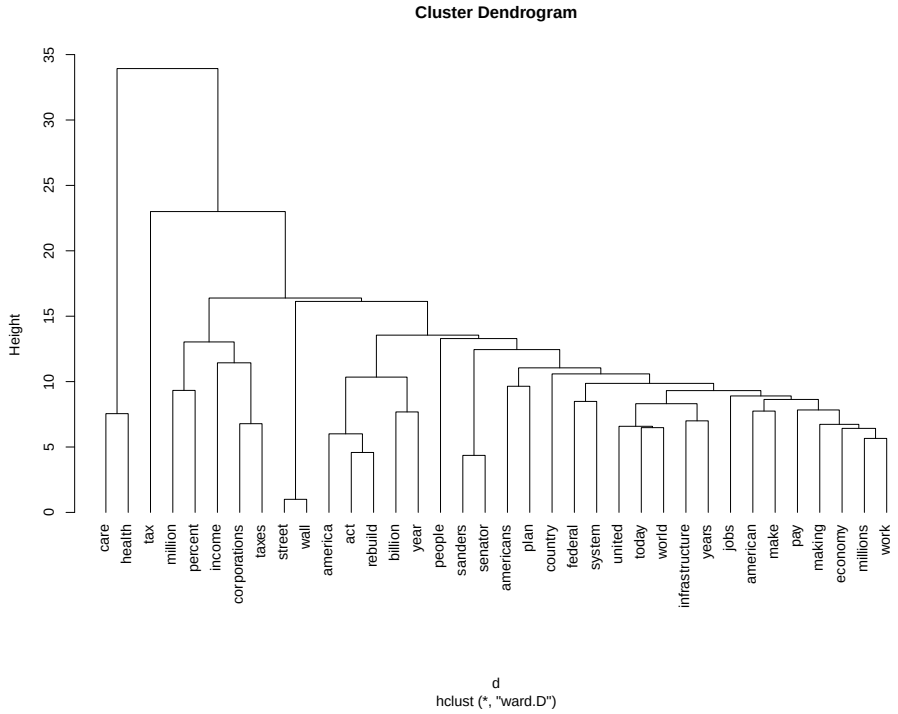


Figure 4.5: Bernie Sanders

The most interesting feature of Hilary Clinton's dendrogram (Figure .2) is the high correlation of words 'fight' and 'president'. In this scope of candidates, aside from Bush and Obama, who were rerunning for the Cabinet, she's the only one to frequently use the name of a position. If one takes a look at data, the phrase 'As a president' occurs very often in her text.

Donald Trump (Figure .2), unlike any of his competitors or predecessors, is the first one who doesn't significantly focus on healthcare. Trade, on the other

[2]As of June 1, 2016

hand, is a big topic for him. Furthermore, he emphasizes on China, which is highly correlated in his agenda with 'trade' and 'jobs'.

4.3 Topic modelling

Defining number of topics for a corpus of text has a number of different empirical approaches (for instance, (Greene, O'Callaghan, and Cunningham, 2014)). Since this algorithms are usually oriented to large datasets or even Big Data and my sample is comparably small, I apply 5 topics for each candidate, which was defined experimentally. Though, one can notice an over-clustering in some cases even with 5 topics.

As one can see, in the first topic of George Bush he concentrated on his 'program' and its continuing character, while not forgetting to add something 'new' (Table 4.2).

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	bush	bush	president	president	president
2	program	president	health	families	bush
3	taxes	billion	bush	care	american
4	continue	care	million	health	federal
5	funding	children	taxes	federal	help
6	new	continue	provide	increase	care

Table 4.2: Topics of agenda of George Bush

John Kerry, on the other hand, focused mostly energy issues and his partnership with John Edwards (Table 4.3).

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	americans	john	plan	john	john
2	can	kerry	america	kerry	edwards
3	energy	energy	john	america	health
4	cut	america	kerry	energy	middleclass
5	economy	new	edwards	cut	new
6	health	percent	clean	plan	clean

Table 4.3: Topics of agenda of John Kerry

In his first run Barack Obama's main topic was his plan on reforming healthcare and insurance. Throughout all five topics one can see this theme to be robust (Table 4.4).

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	plan	health	health	obama	obama
2	health	care	new	plan	health
3	taxes	barack	create	create	create
4	barack	taxes	plan	invest	new
5	credit	new	american	care	work
6	work	federal	barack	expand	billion

Table 4.4: Topics of agenda of Barack Obama, 2012

During his second campaign, his presidential status and achievements in affordable healthcare were underlined. Moreover, this was also supported by mentioning creation of new jobs and small businesses after the Great Recession. One can also notice the switch between his usage of his first name and surname to only surname.

Obama's first competitor John McCain (Table .1) also used his name a lot, while his policy was oriented on taxes, healthcare. Interestingly, a term 'believes' occurred quite often in his agenda. Mitt Romney in 2012 (Table .2) as was already seen, targeted Obama's mistakes. In his topics 'president', 'government' and 'Obama' are represented consistently.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	care	obama	act	president	president
2	health	care	obama	care	obama
3	obama	jobs	reform	new	jobs
4	affordable	taxes	president	small	insurance
5	insurance	affordable	care	businesses	american
6	president	million	insurance	job	health

Table 4.5: Topics of agenda of Barack Obama, 2008

In the elections of 2016, Hillary Clinton, as we saw in clustering part, used the word 'president' very frequently. Topic modelling showed, that her usage of her first name is also significant. As already mentioned, Barack Obama used his name hugely prior to her in 2008 elections. This strategy seems valid, considering that her current surname might be associated with her husband, 42nd president of the USA, Bill Clinton. Her topics seem to focus on family issues, with attention to themes like manufacturing, affordable healthcare and security.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	care	hillary	care	hillary	hillary
2	families	families	hillary	health	health
3	access	plan	plan	affordable	plan
4	manufacturing	every	health	taxes	support
5	family	leave	access	families	security
6	make	new	security	law	work

Table 4.6: Topic of agenda of Hillary Clinton

Her main competitor for the party nomination Bernie Sanders concentrates on inequality. His plans include taxis corporations and just distribution of income (Table .3).

The republican Donald Trump, on the other hand, is concerned with China's role in the economy of the US. He assigns it both trade and employment (Table .4).

4.4 Sentiment analysis

The result of sentiment analysis is shown below (Table 4.7).

	polarity	subjectivity	pos_refs_per_ref	neg_refs_per_ref	senti_diffs_per_ref
George Bush	0.2488	0.07856	0.05944	0.01911	0.04033
John Kerry	0.08331	0.09219	0.04990	0.04229	0.007607
Barack Obama	0.2604	0.14056	0.08460	0.05597	0.02863
John McCain	0.1206	0.08278	0.05238	0.03039	0.02199
Barack Obama	0.1149	0.06038	0.04234	0.01804	0.02429
Mitt Romney	0.0633	0.09797	0.05051	0.04746	0.003055
Hilary Clinton	0.2205	0.11258	0.07753	0.03505	0.04248
Bernie Sanders	0.07811	0.1349	0.07297	0.06192	0.01105
Donald Trump	-0.06152 (-0.05899)	0.0904 (0.1607)	0.03937 (0.08806)	0.05103 (0.0726)	0.01166 (0.01546)

Table 4.7: Sentiment analysis, means of values

The most noticeable moment in this table is Donald Trump's negative polarity score. Both his 'positions' and video recorded 'issues' seem to be on average more negative than positive. Moreover, his issues scored the maximum in subjectivity, comparing to all other candidates.

The closest other candidate to Trump's polarity is Mitt Romney. This is not surprising, since, as mentioned before, he mainly targeted mistakes of Obama in his agenda.

Interestingly, Barack Obama in 2008 has the highest polarity score of 0.26 with quite high subjectivity of his agenda. This might be due to his overall campaign idea, which was based on positive slogans like 'Yes we can' and 'Hope'. Nevertheless, in 2012 elections he was more than twice less positive and subjective.

George Bush's second run for the Cabinet also shows positivity. This can be connected to overall well situation in the economy at that time.

Hilary Clinton comes third, with a score of polarity equal to 0.2205 and high subjectivity of her text. Her closest competitor Bernie Sanders is more subjective and less positive, due to targeting an inequality issue.

Chapter 5

CONCLUSION

It's hard to define a pattern in a given dataset of texts. For instance, 'taxes' seem to be consistent throughout most candidates, yet the context they are used in differs.

Nevertheless, this paper offers insights into situational issues.

Donald Trump's videos on issues supported the idea of their easiness in readability ((Schumacher and Eskenazi, 2016)). While Sanders, surprisingly, scored the maximum, Trump's positions were also very understandable.

Basing on hierarchical clustering, one can notice that most candidates try to use their name a lot. The record is held by Barack Obama, who in 2012 though switched to underlying his presidential position and highly focused on achievements in Obamacare.

Unlike topics of male candidates, Hilary Clinton's results in LDA topic modelling showed concentration on her first name. This is obvious, due to desire to be different from her husband and ex-president. Moreover, her focus on American families is also noticeable.

As I mentioned in introduction, both Obama and Trump are very popular within electorate, yet their popularity differs very much. Sentiment analysis has shown, that they are not just different, they might be opposite.

Trump is the only one to get negative polarity score, while Obama in his first run had very positive attitude.

It seems, that Obama's success in 2008 was largely due to this kind of attitude and charisma, which was targeted to give people hope. Moreover, unlike his competitor McCain, who's text seems to be indecisive ('believes') about his policies, Obama had at least one significant goal in healthcare.

Predicting who is going to win elections of 2016 is out of scope of this work, yet basing on positivity of the Hilary Clinton's text and her usage of word 'president' I can say that most probably she will be the nominee of the Democratic party. Her competition with Trump will mainly depend on his behaviour.

BIBLIOGRAPHY

Berry, Michael W and Malu Castellanos (2004). "Survey of text mining". In: *Computing Reviews* 45.9, p. 548.

Blei, David M, Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *the Journal of machine Learning research* 3, pp. 993–1022.

Casselman, Ben (2015). *The Big Issues Of The 2016 Campaign*. URL: `http://fivethirtyeight.com/features/year-ahead-project/#part1`.

Feinerer, Ingo (2015). *Introduction to the tm Package Text Mining in R*.

Greene, Derek, Derek O'Callaghan, and Pádraig Cunningham (2014). "How many topics? stability analysis for topic models". In: *Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 498–513.

Grimmer, Justin and Brandon M Stewart (2013). "Text as data: The promise and pitfalls of automatic content analysis methods for political texts". In: *Political Analysis* 21.3, pp. 267–297.

Hornik, Kurt and Bettina Grün (2011). "topicmodels: An R package for fitting topic models". In: *Journal of Statistical Software* 40.13, pp. 1–30.

Laver, Michael, Kenneth Benoit, and John Garry (2003). "Extracting policy positions from political texts using words as data". In: *American Political Science Review* 97.02, pp. 311–331.

Mohammad, Saif M and Peter D Turney (2010). "Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon". In: *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. Association for Computational Linguistics, pp. 26–34.

O'Connor, Brendan et al. (2010). "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." In: *ICWSM* 11.122-129, pp. 1–2.

Pang, Bo and Lillian Lee (2008). "Opinion mining and sentiment analysis". In: *Foundations and trends in information retrieval* 2.1-2, pp. 1–135.

Schumacher, Elliot and Maxine Eskenazi (2016). "A Readability Analysis of Campaign Speeches from the 2016 US Presidential Campaign". In: *arXiv preprint arXiv:1603.05739*.

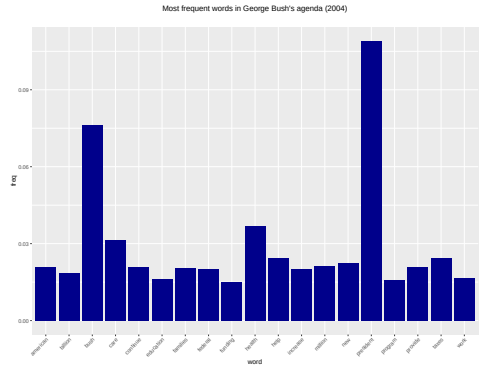
Werner, Annika, Onawa Lacewell, and Andrea Volkens (2011). "Manifesto Coding Instructions (4th fully revised edition)". In: *Wissenschaftszentrum Berlin für Sozialforschung (WZB)*. URL: `https://manifesto-project.wzb.eu/information/documents`.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann (2005). "Recognizing contextual polarity in phrase-level sentiment analysis". In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, pp. 347–354.

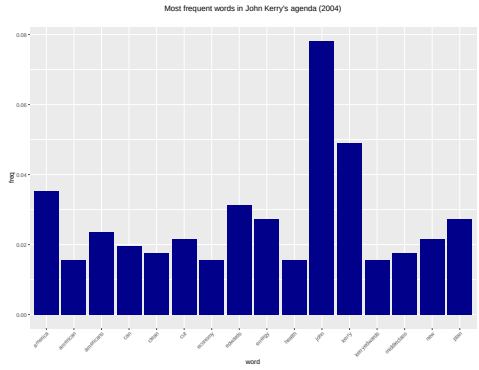
APPENDIX

Most frequent words by candidate

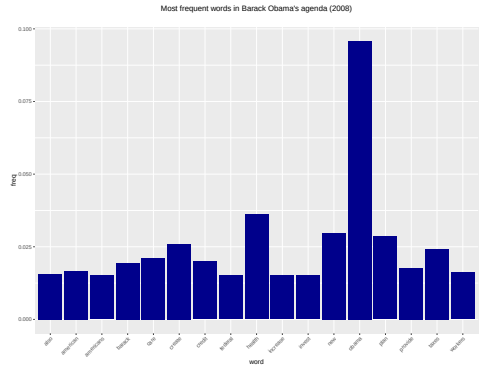
All terms with relative frequency above 0.15 are represented below.



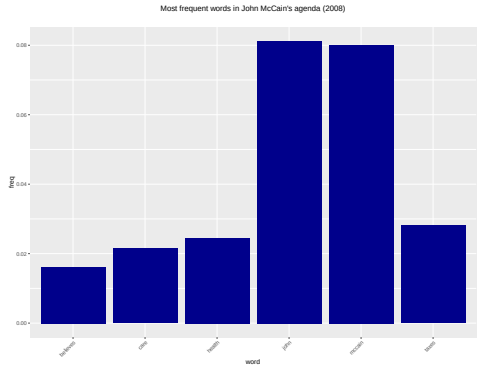
(a) George Bush



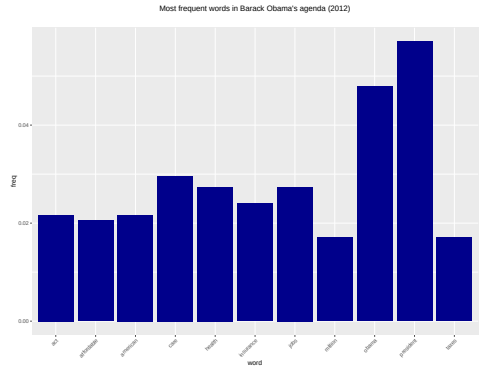
(b) John Kerry



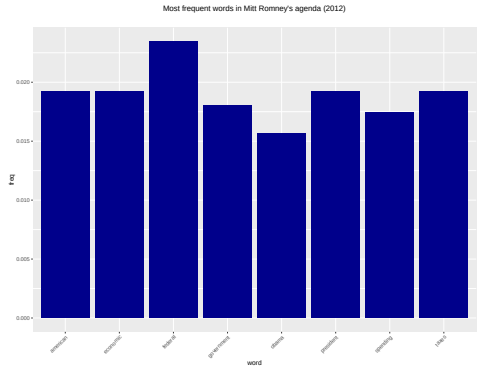
(c) Barack Obama (2008)



(d) John McCain

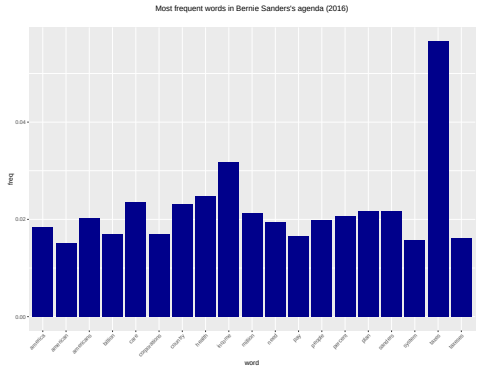


(e) Barack Obama (2012)

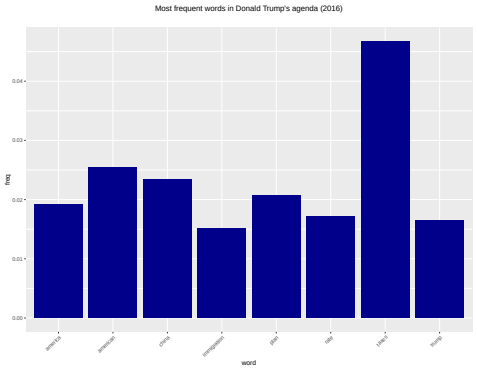


(f) Mitt Romney

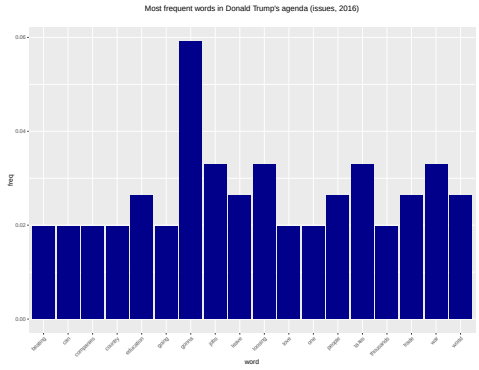
Figure .1: Most frequent words in agendas by candidate



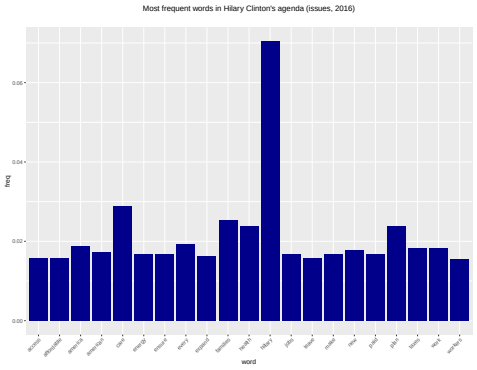
(g) Bernie Sanders



(h) Donald Trump



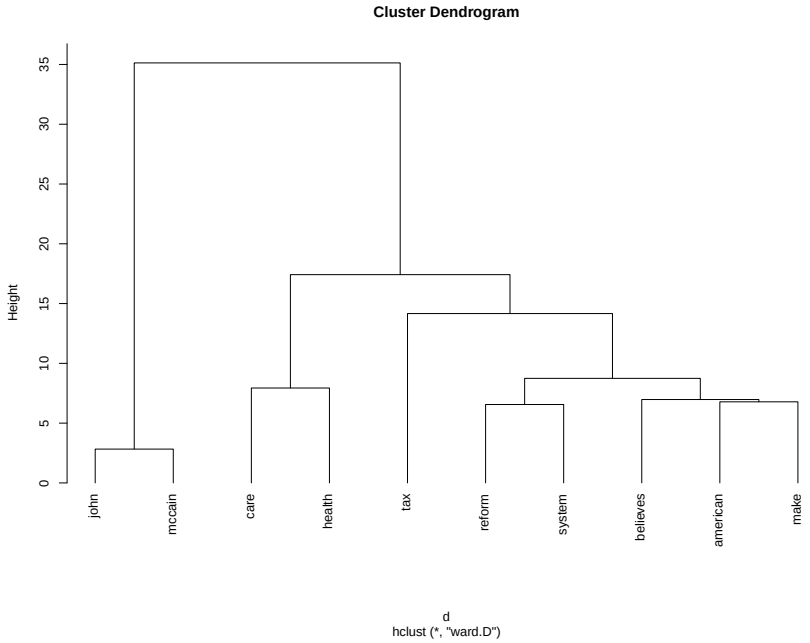
(i) Donald Trump (issues)



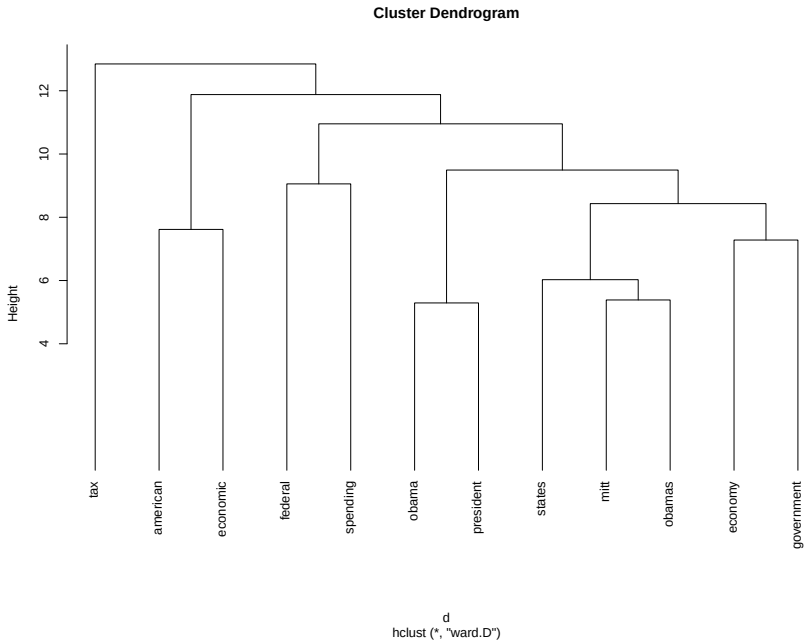
(j) Hilary Clinton

Figure .1: Most frequent words in agendas by candidate (continued)

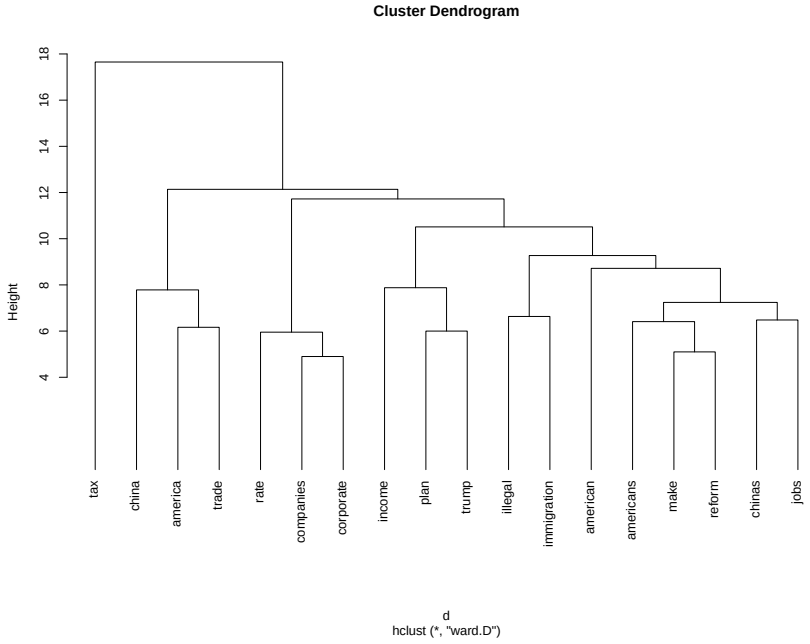
Dendrograms of agendas of presidential candidates



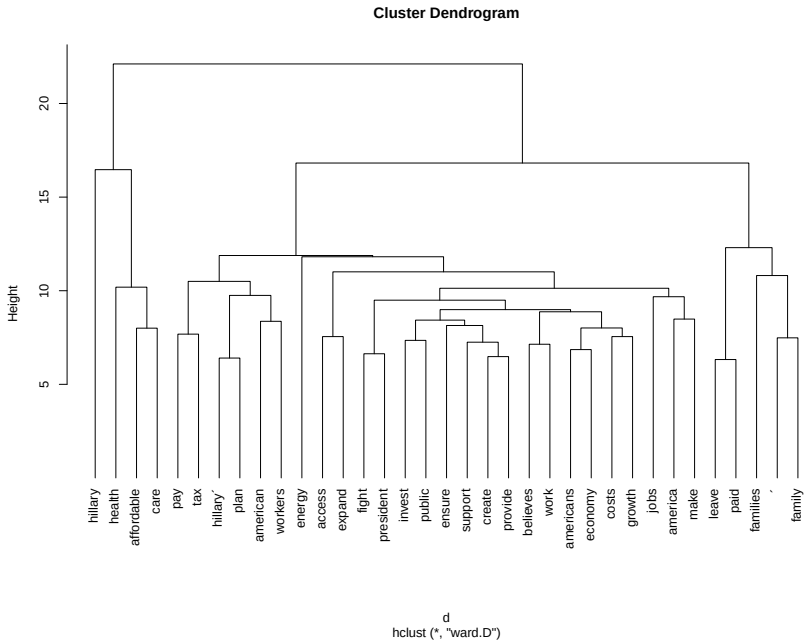
(a) John McCain



(b) Mitt Romney



(c) Donald Trump



(d) Hilary Clinton

Figure .2: Cluster trees of agendas (continued)

Topic models of presidential candidates

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	john	mccain	health	mccain	mccain
2	taxes	john	care	john	john
3	mccain	health	mccain	can	care
4	care	make	program	new	health
5	reform	believes	taxes	taxes	percent
6	costs	system	american	must	believes

Table .1: Topic of agenda of John McCain

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	economic	taxes	american	president	jobs
2	federal	spending	obama	obama	federal
3	government	federal	government	percent	spending
4	investment	president	federal	american	taxes
5	spending	economic	economic	energy	government
6	jobs	american	labor	growth	must

Table .2: Topic of agenda of Mitt Romney

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	taxes	percent	taxes	taxes	taxes
2	sanders	country	percent	income	million
3	plan	income	care	americans	care
4	corporations	billion	american	need	people
5	income	taxes	health	people	energy
6	country	health	plan	just	america

Table .3: Topic of agenda of Bernie Sanders

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	taxes	taxes	plan	american	taxes
2	states	plan	china	china	rate
3	trump	must	america	taxes	chinese
4	chinas	jobs	american	america	income
5	corporate	trump	workers	must	deductions
6	immigration	healthcare	trump	corporate	jobs

Table .4: Topic of agenda of Donald Trump