

R Notebook

Yue Jin

Step 1 - Data Preparation

- Set environment
- Import data
- Remove extra white space, convert all letters to the lower case, remove stop words, removed empty words due to formatting errors, and remove punctuation.

Topic Modeling

- LDA

Extract the most likely terms for each topic or the most likely topics for each document.

```
##
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
## 1 1 1 2 18 1 1 1 1 1 3 2 2 1 1

##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] "feder"    "emerg"    "listen"  "effect"  "america" "found"
## [2,] "hungri"    "enact"    "substitut" "econom"  "live"    "higher"
## [3,] "tax"       "split"    "quiet"    "north"   "god"     "path"
## [4,] "spend"     "submit"   "heard"    "benefit" "promis"  "collaps"
## [5,] "administr" "employ"   "pain"     "save"    "dream"   "girl"
## [6,] "arsenal"   "communic" "white"    "trade"   "build"   "inhabit"
## [7,] "west"      "1789"     "middl"    "inalien" "last"    "settl"
## [8,] "monument" "afraid"   "alarm"    "develop" "courag"  "claim"
## [9,] "slow"      "ancestor" "reap"     "major"   "say"     "fascism"
## [10,] "commonplac" "clean"    "allegi"   "desir"   "share"   "lend"
##      [,7]      [,8]      [,9]      [,10]     [,11]     [,12]
## [1,] "easi"     "sorrow"   "mind"    "ensur"   "pledg"   "permit"
## [2,] "show"     "coven"    "liber"    "respons" "far"     "fix"
## [3,] "founder"  "mine"     "charact"  "brother" "war"     "militari"
## [4,] "ambit"    "short"    "creed"    "spring"  "side"    "europ"
## [5,] "break"    "stranger" "fire"     "role"    "light"   "prison"
## [6,] "servic"   "lifetim"  "meant"    "abroad"  "mankind" "trial"
## [7,] "imper"    "second"   "run"      "choos"   "lie"     "defens"
## [8,] "grand"    "battalion" "institut" "adversari" "globe"   "confer"
## [9,] "habit"    "deepen"   "oppress"  "poor"    "seek"    "spoken"
## [10,] "ride"    "candor"   "defeat"   "endless" "perfect" "symbol"
##      [,13]     [,14]     [,15]
## [1,] "reserv"   "throughout" "read"
## [2,] "easili"   "father"     "potenti"
## [3,] "wise"     "bow"        "heaven"
## [4,] "adequ"    "door"       "saw"
## [5,] "incom"    "crime"      "class"
## [6,] "loss"     "breez"      "driven"
## [7,] "stricken" "birthright" "enlarg"
```

```

## [8,] "tide"      "vow"      "21st"
## [9,] "urg"       "magic"    "20th"
## [10,] "kind"     "teacher"  "flag"

##      Topic 1  Topic 2  Topic 3  Topic 4  Topic 5  Topic 6
## [1,] "believ"  "vision"  "earth"  "econom"  "will"  "found"
## [2,] "weapon"  "given"   "moment"  "state"  "nation"  "women"
## [3,] "let"     "remain"  "good"    "men"    "america"  "oath"
## [4,] "system"  "congress"  "protect"  "democraci"  "peopl"  "may"
## [5,] "tax"     "face"    "job"     "unit"   "world"  "see"
## [6,] "begin"   "must"    "whether"   "believ"  "new"    "declar"
## [7,] "limit"   "emerg"   "oath"    "creat"  "american"  "deni"
## [8,] "much"    "honor"   "voic"    "nation"  "time"   "happi"
## [9,] "well"    "arm"     "wealth"  "secur"  "govern"  "hour"
## [10,] "feder"  "threaten"  "capit"  "program"  "must"  "prosper"
## [11,] "reduc"  "uniti"    "child"  "save"    "great"  "right"
## [12,] "struggl"  "common"  "crisi"  "right"  "live"  "victori"
## [13,] "prayer"  "cost"    "quiet"  "trade"  "one"  "equal"
## [14,] "support"  "hatr"    "remain"  "intern"  "work"  "health"
## [15,] "economi"  "histor"  "across"  "realiz"  "everi"  "mere"

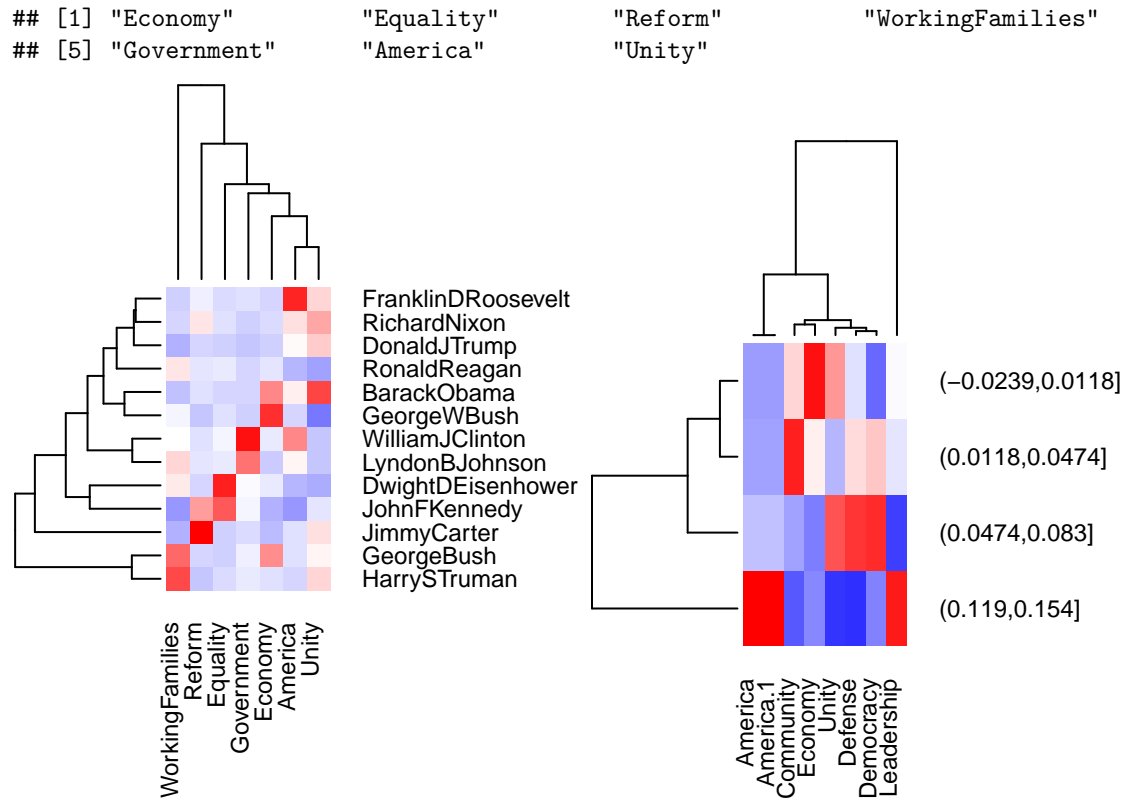
##      Topic 7  Topic 8  Topic 9  Topic 10  Topic 11  Topic 12
## [1,] "servic"  "man"    "freedom"  "let"    "can"    "free"
## [2,] "less"    "without"  "liberti"  "respons"  "peac"  "freedom"
## [3,] "civil"   "believ"  "know"    "better"  "world"  "secur"
## [4,] "forc"    "possibl"  "human"   "peac"    "help"  "men"
## [5,] "never"   "chang"   "ideal"   "spirit"  "nation"  "must"
## [6,] "across"  "union"   "duti"    "home"    "peopl"  "upon"
## [7,] "care"    "uniti"   "societi"  "celebr"  "shall"  "hope"
## [8,] "polit"   "find"    "word"    "commit"  "know"  "life"
## [9,] "common"  "mani"    "caus"    "respect"  "war"  "may"
## [10,] "among"  "rest"    "carri"   "era"     "seek"  "equal"
## [11,] "call"   "becom"   "choic"   "high"    "need"  "common"
## [12,] "show"   "first"   "mind"    "other"   "good"  "law"
## [13,] "knowledg"  "learn"  "stori"   "proud"   "ask"  "princip"
## [14,] "beyond"  "destini"  "accept"  "role"    "meet"  "rule"
## [15,] "founder"  "fear"    "advanc"  "abroad"  "power"  "strength"

##      Topic 13  Topic 14  Topic 15
## [1,] "upon"     "thank"  "centuri"
## [2,] "measur"    "thing"  "polit"
## [3,] "action"    "mean"   "famili"
## [4,] "spirit"    "friend"  "give"
## [5,] "constitut"  "father"  "clear"
## [6,] "task"      "like"   "class"
## [7,] "leadership"  "take"   "stand"
## [8,] "democraci"  "hard"   "take"
## [9,] "order"     "part"   "terror"
## [10,] "part"     "fact"   "assum"
## [11,] "purpos"   "love"   "done"
## [12,] "mere"     "need"   "enough"
## [13,] "restor"   "word"   "hear"
## [14,] "kind"     "hand"   "power"
## [15,] "line"     "vice"   "read"

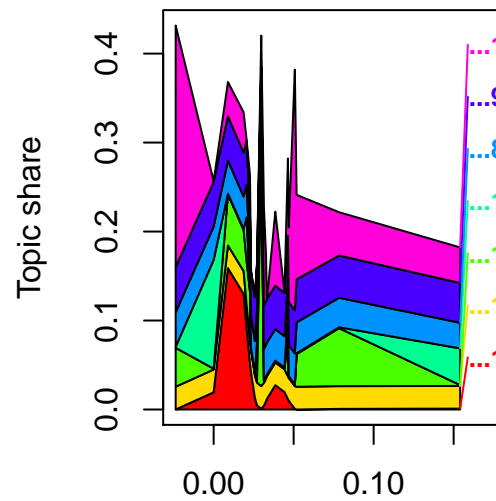
```

- Summarize Each Topic with One Word

- Clustering of topics



Topic Shifts

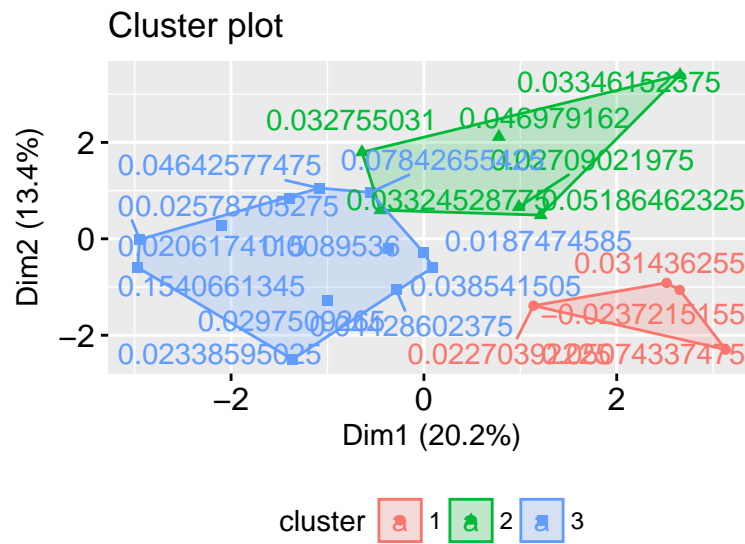


- How topics shift as economic condition changes

[1] 0.05857093 0.11714186 0.17571279 0.23428373 0.29285466 0.35142559

[7] 0.40999652

How much topics differ given different economic conditions



Reference

- Text mining with `tidytext`.
- Basic Text Mining in R