

Topic Modeling Experiment Instructions:

Topic Identification

Attached are 4 sets of topics: **15A – 15B – 16A – 16B**

Step 1: Identify a label the topic based on the text below.

For example:

Topic 1 Top Words:

Highest Prob: american, spend, america, interest, obamacar, best, wrong, congress, welfar, budget

FREX: american, spend, america, interest, obamacar, special, balanc, dirti, phone, best

Lift: apolig, bailout, brief, bui, dean, demo, envolv, foolish, furious, ship

Score: spend, american, interest, obamacar, america, best, congress, welfar, budget, wrong

- ☐ One could determine this topic as: **Policy – Social** or **Government Spending**
- ☐ There are no correct answers just your interpretation of what this topic should be labeled.
- ☐ Note that the partial words are stems to a word for example: **apolig** = likely means apologize or could be apologizing

Step 2: Once completed assigning labels, then select below which model in your mind showed preformed best at centering around the topics in general.

- 15 A
- 15 B
- 16 A
- 16 B

Additional Definitions

Highest Prob = the words within each topic with the highest probability (inferred directly from topic-word distribution parameter)

FREX = the words that are both frequent and exclusive, identifying words that distinguish topics. This is calculated by taking the harmonic mean of rank by probability within the topic (frequency) and rank by distribution of topic given word (exclusivity)

Lift = Specifically, lift calculates how much more likely a word occurs in a topic compared to its frequency in the entire corpus of text

Score = refers to a metric that evaluates the quality of the identified topics.