The effects of transitions of power on the contents of municipal government websites

Markus Neumann
Bruce Desmarais
Hanna Wallach
Fridolin Linder

The Pennsylvania State University

January 6, 2018

Government Websites

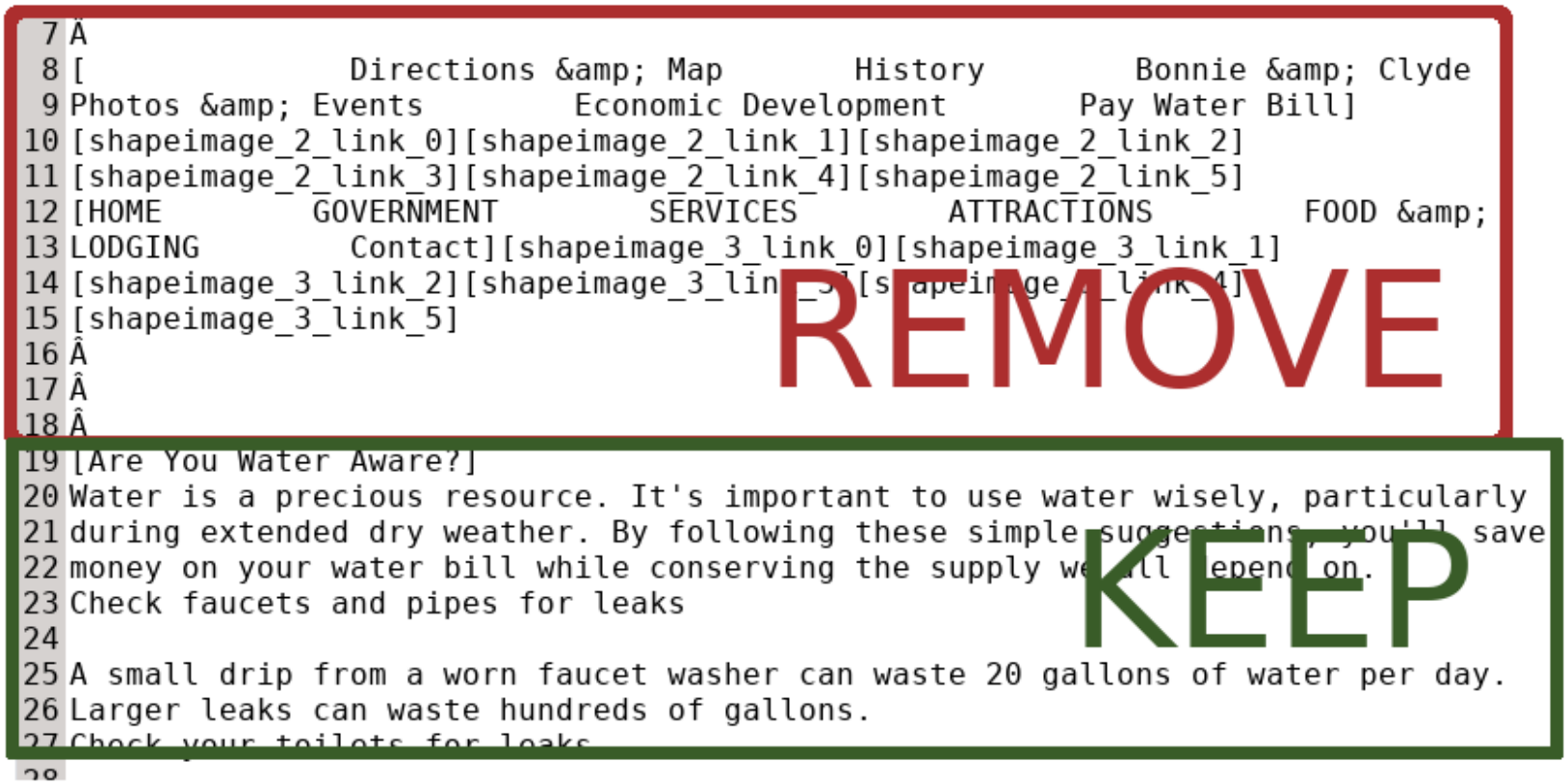
- Transparency

wget

Duplicate Line Removal

```
 7 Â
 8 [              Directions &amp;amp; Map       History        Bonnie &amp;amp; Clyde
 9 Photos &amp;amp; Events        Economic Development       Pay Water Bill]
10 [shapeimage_2_link_0][shapeimage_2_link_1][shapeimage_2_link_2]
11 [shapeimage_2_link_3][shapeimage_2_link_4][shapeimage_2_link_5]
12 [HOME        GOVERNMENT        SERVICES        ATTRACTIONS        FOOD &amp;amp;
13 LODGING        Contact][shapeimage_3_link_0][shapeimage_3_link_1]
14 [shapeimage_3_link_2][shapeimage_3_link_3][shapeimage_3_link_4]
15 [shapeimage_3_link_5]
16 Â
17 Â
18 Â
19 [Are You Water Aware?]
20 Water is a precious resource. It's important to use water wisely, particularly
21 during extended dry weather. By following these simple suggestions, you'll save
22 money on your water bill while conserving the supply we all depend on.
23 Check faucets and pipes for leaks
24
25 A small drip from a worn faucet washer can waste 20 gallons of water per day.
26 Larger leaks can waste hundreds of gallons.
27 Check your toilets for leaks
```

Duplicate Line Removal



Duplicate Line Removal

- Within each city, there is a lot of shared text
  - Boilerplate
  - Website elements
- If not removed, the text clusters into cities

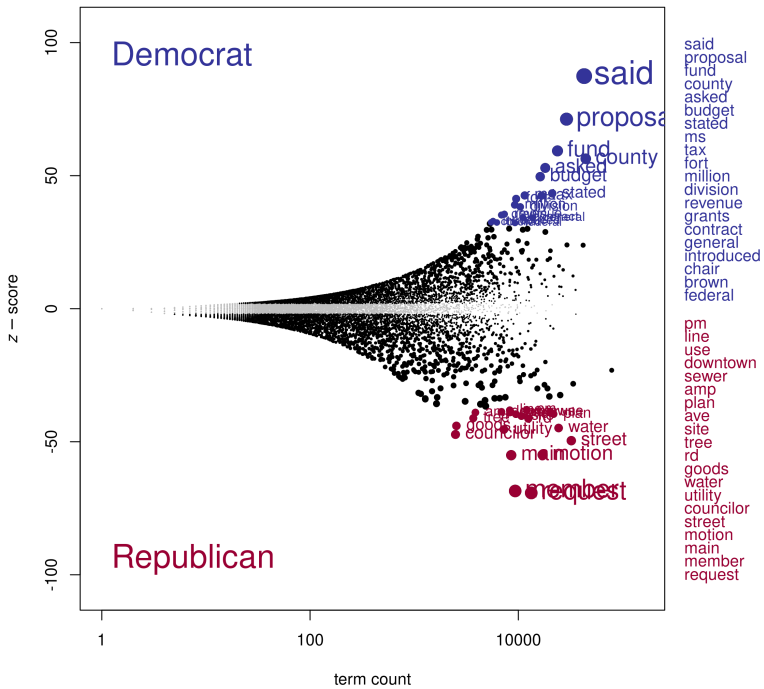
- Solution: Compare each line in each document to every other line in every document of that city
- Count duplicates
- Remove a line if it is duplicated within a city above some threshold
- Hash tables for computational efficiency

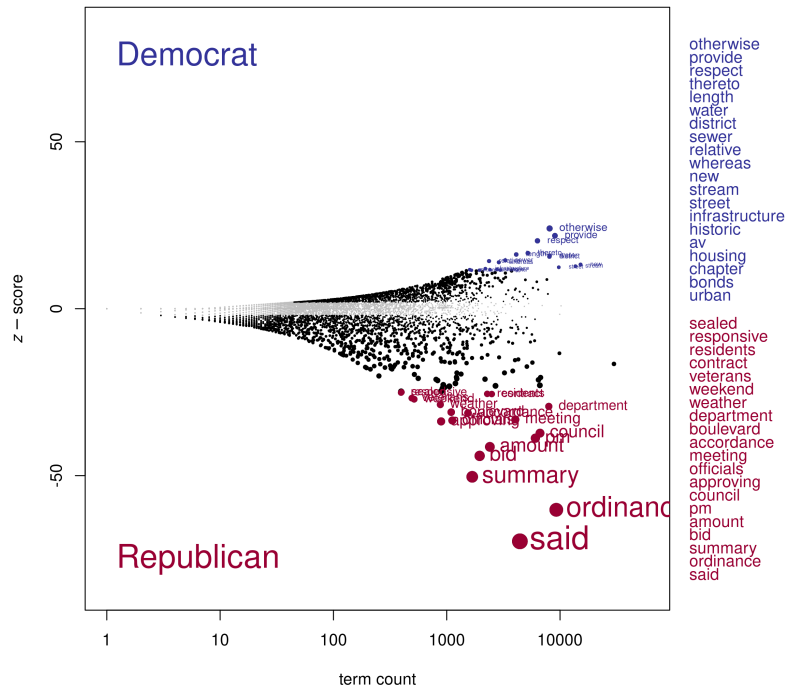
Other preprocessing

- Remove:
  - Dates
  - Numbers
  - Punctuation
  - Words not recognized by an English dictionary
  - Stopwords
  - Documents with too few unique words
- Set to lowercase
- Lemmatization

Hierarchical Clustering

Fightin' Words





Latent Dirichlet Allocation

Conclusion