Democratization and Linguistic Complexity. The Effect of Franchise Extension on Parliamentary Discourse, 1832–1915.

Priscila Stisman & Samuel Cohen

February 13, 2025

Index

Introduction

- **Research Question:** The Second Reform Act of 1867 roughly doubled the size of the electorate in the United Kingdom. How did the language of backbenchers and cabinet members in parliament change over time as a result of this expansion?
- **Data:** 675,997 speeches made by MPs from 1832 to 1915
- **Hypotheses:**
- (1) Language becomes less complex over time for cabinet MPs, as they need to appeal to the "common man."
- (2) Backbenchers do not need to adjust language as much, since they are relatively out of sight ("vote for the party, not the man")

Methods

- **What We Recreate:**
- Readability trends (by cabinet member, syllables per word over time)
- Hypothesis testing results (cabinet effect on readability)

- **What We Add:**
- Party affiliation
- Text Analysis: TFIDF, Cosine Similarity, Tokenization, etc.

Methods

- **Bigframe Data:**
- Word Count, Syllable per Word Count, FRE Scores
- Cabinet, Party, Competitiveness, Year
- Used for: Regressions, trends over time

- **Raw Speeches:**
- Slightly cleaned speech text (word consolidation but no stemming, stop word removal, etc.). Sample of 10,000 speeches used
- Used for: Text analysis

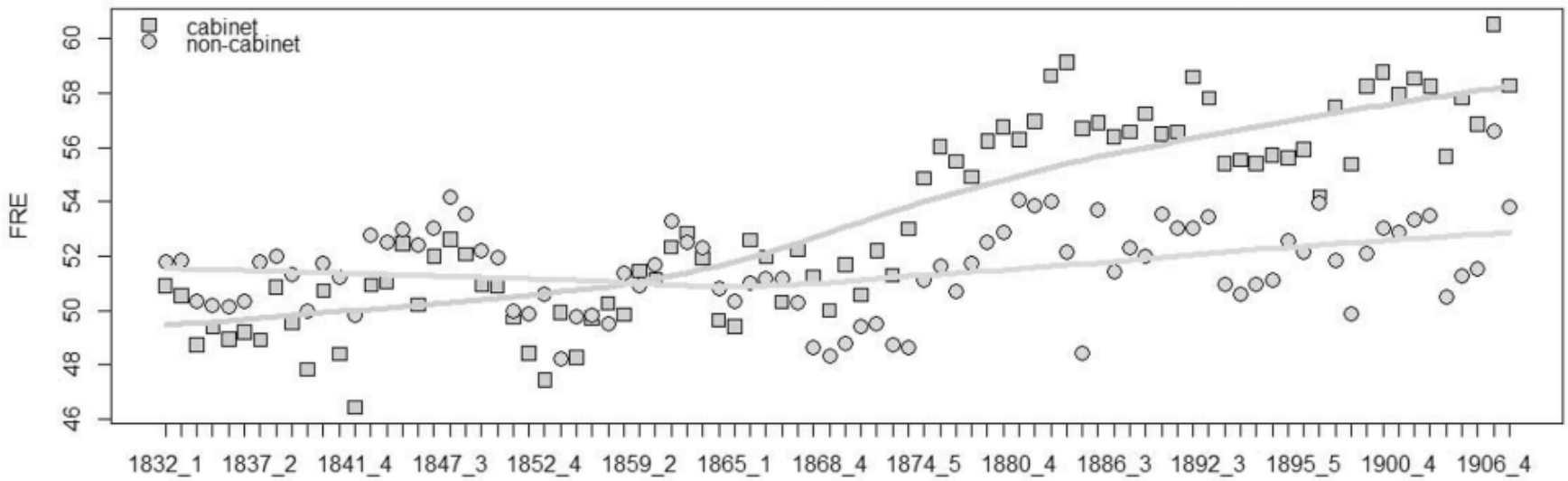
Results

FRE Statistics

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-301.78	41.81	54.75	51.16	66.53	121.22

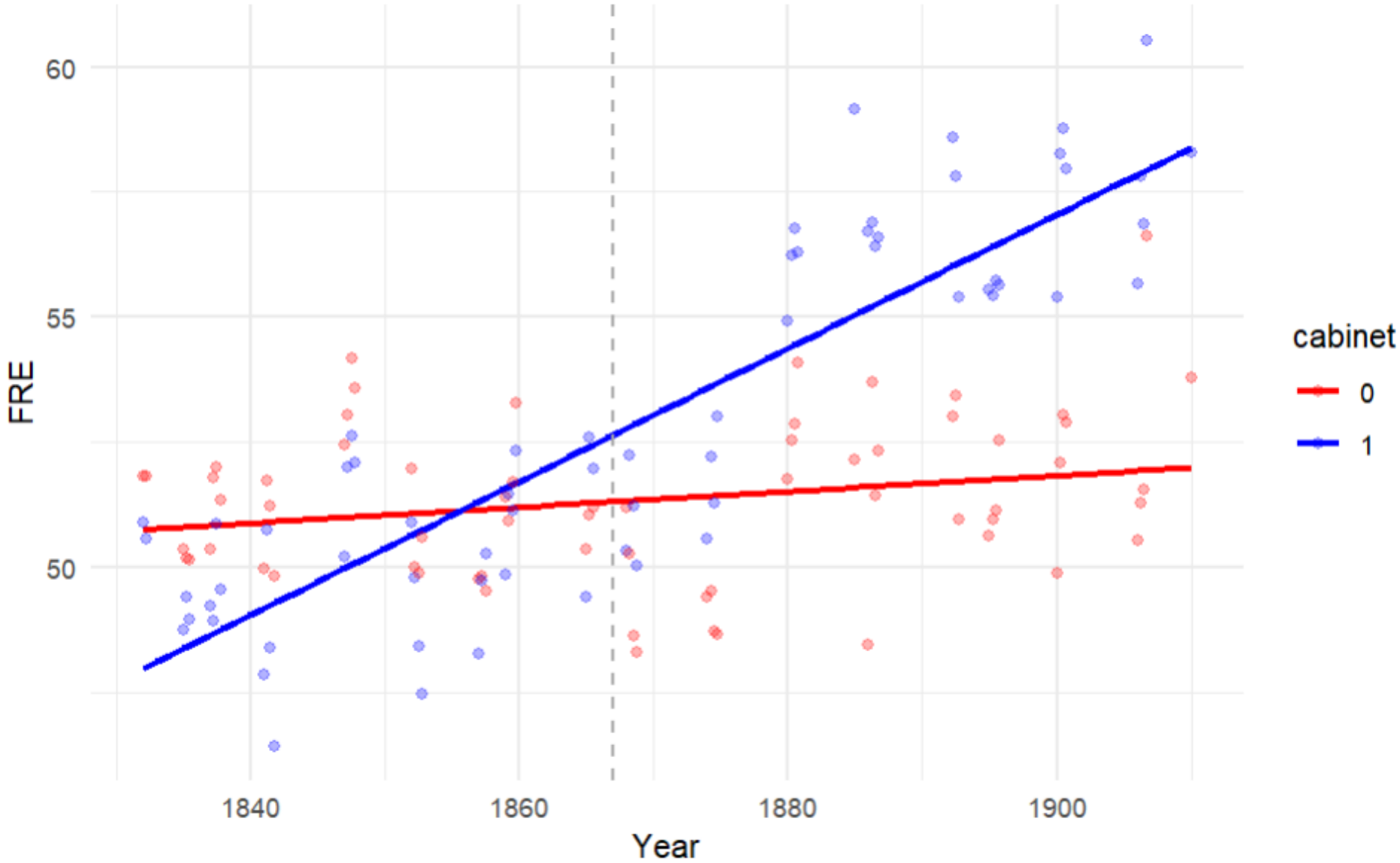
Results

Mean Readability Over Time by Cabinet Status



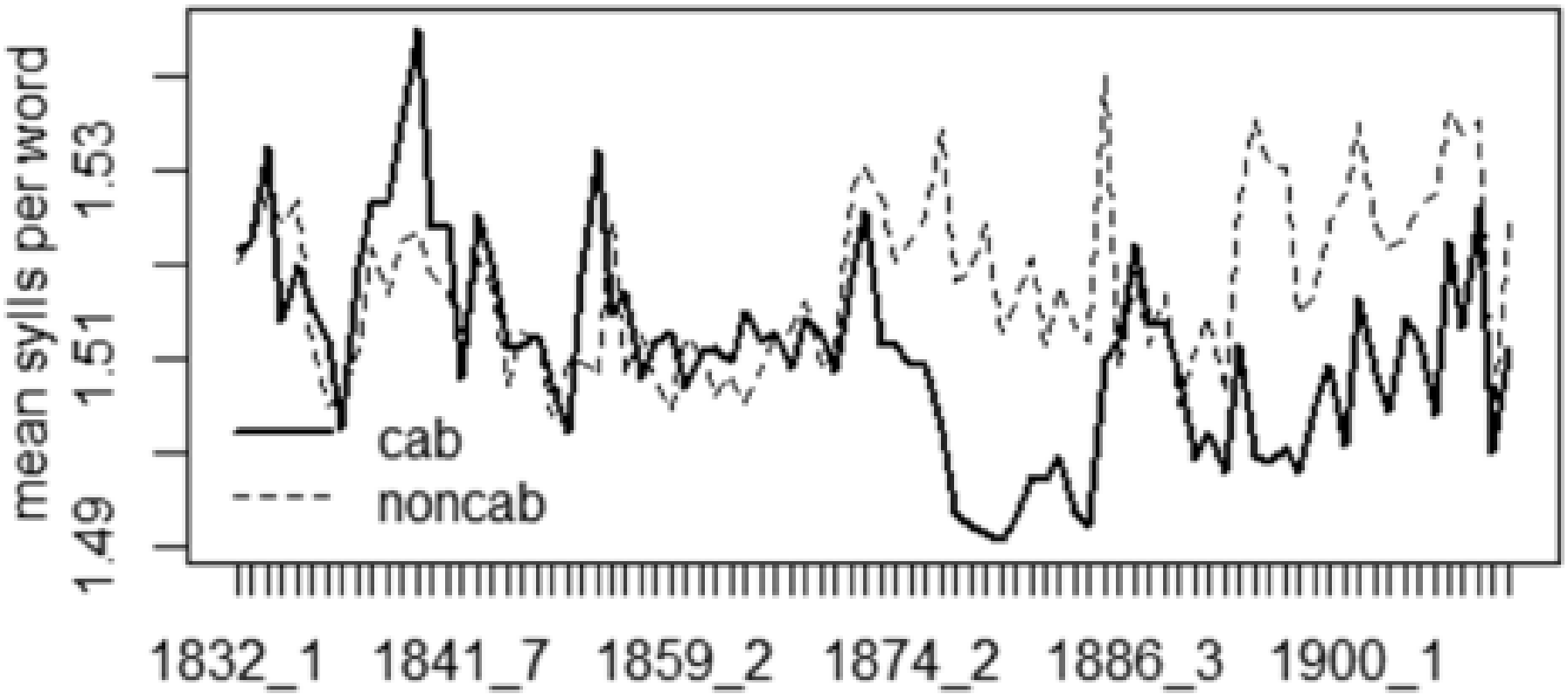
Results

Mean Readability Over Time by Cabinet Status (Replication)



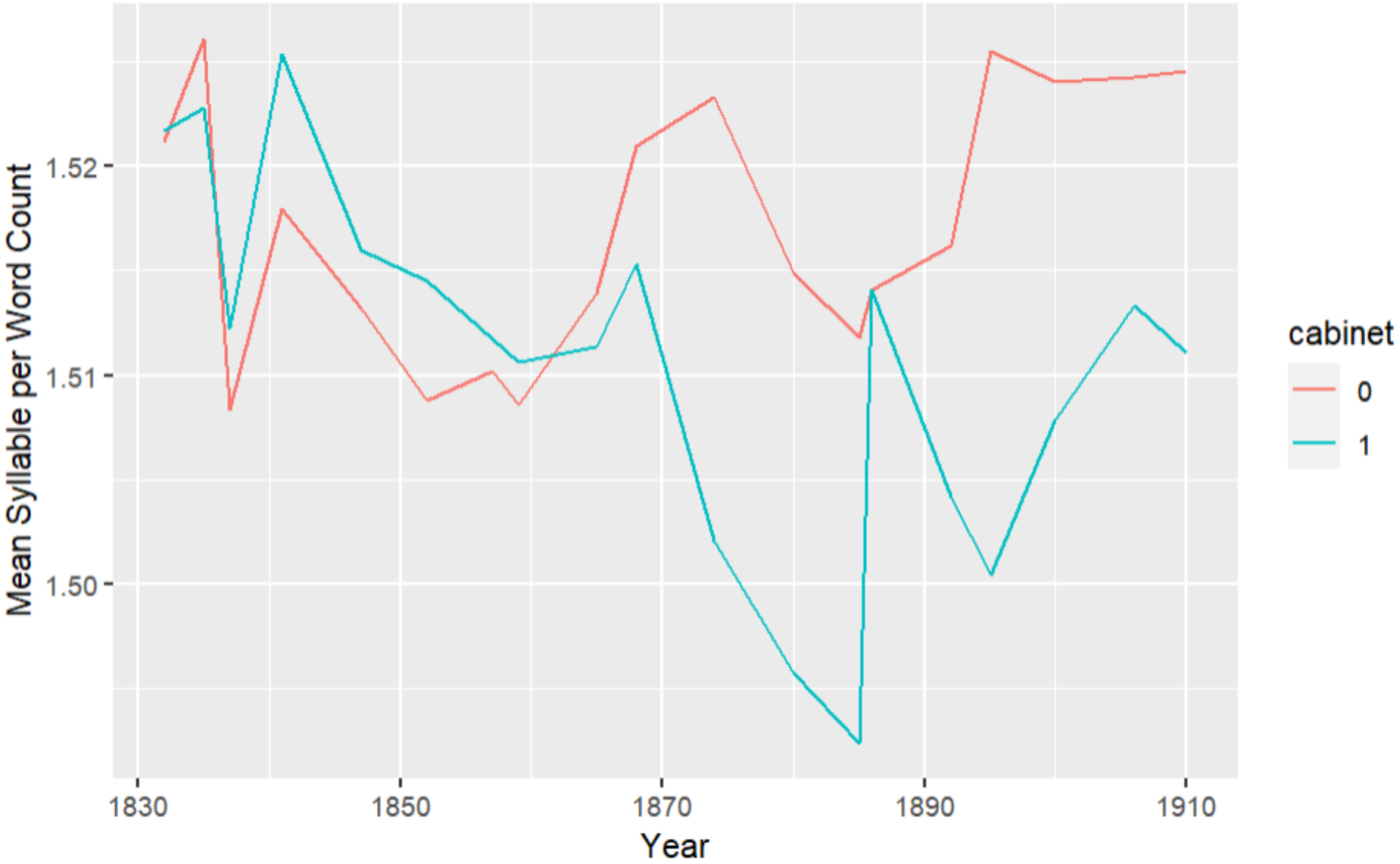
Results

Mean Syllable per Word Count by Cabinet Status



Results

Mean Syllable per Word Count by Cabinet Status (Replication)



Results

Multivariate Regression

	Reform Act Interaction	With Controls
(Intercept)	51.3976 ***	51.5920 ***
	(.2176)	(.5634)
Cabinet member	−.8189*	−.7803
	(.4176)	(.4271)
Reform Act dummy	.7172*	.5172
	(.3371)	(.3591)
Cabinet × Reform Act	5.3060 ***	5.2251 ***
	(.7195)	(.7172)
Liberal MP		.4511

Results

Multivariate Regression

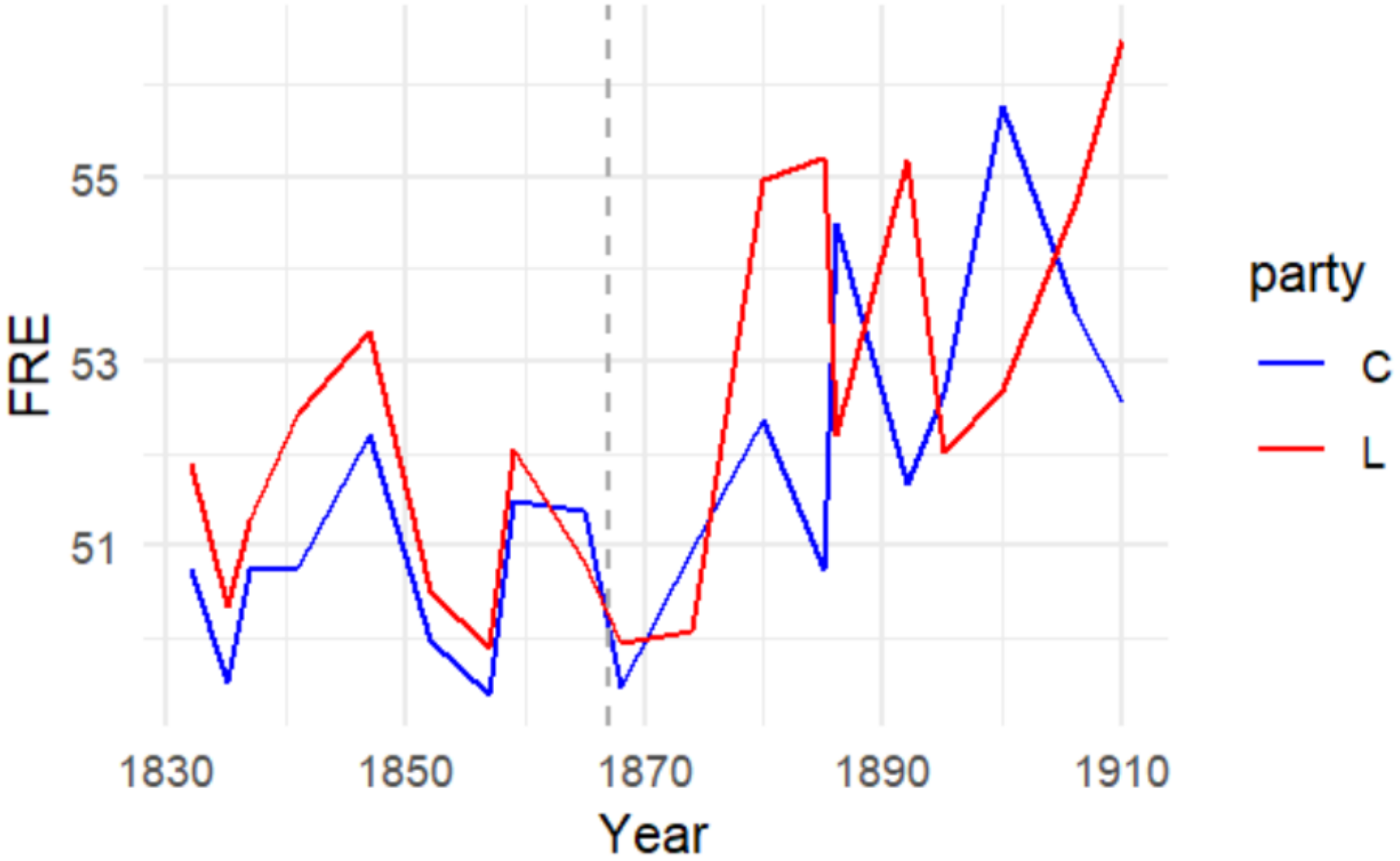
term <chr>	estimate <dbl>
(Intercept)	51.58964043
cabinet1	-0.78025119
post_reform	0.51715963
partyL	0.45105739
competitiveness	0.02301681
word.count	-0.00132767
cabinet1:post_reform	5.22513624

Extensions

- Party affiliation analysis
- TF-IDF Weighting for each token to find which words carried the most importance throughout the corpus
- Cosine Similarity to find the speeches that are most similar with each other

Extensions

Party Affiliation and Speech Readability



Extensions

TF-IDF and cosine similarity

	names <chr>	variable <fctr>	cosine <dbl>
1	text7214	text1779	0.9969217
2	text1779	text7214	0.9969217
3	text8866	text7214	0.9969217
4	text7214	text8866	0.9969217
5	text8998	text8393	0.9957456
6	text8393	text8998	0.9957456
7	text6084	text4254	0.9953603
8	text4254	text6084	0.9953603
9	text1062	text55	0.9917438
10	text55	text1062	0.9917438

1–10 of 10 rows

Extensions

Examples of high cosine similarity:

- "said, he would repeat the Question on Monday." $\rightarrow$ "I will repeat the question on Monday." (0.998)
- "said, he would withdraw the Amendment." $\rightarrow$ "Then I will withdraw my Amendment." (0.998)
- "objected to the Amendment." $\rightarrow$ "said, he had no objection to the Amendment." (0.998)

Differences and Similarities

- **FRE Statistics:** The minimum, first quartile, median, mean, and third quartile closely match those in the paper. Maximum differs (205 vs. 121).
- **FRE Trends:** The average readability score, in both the paper and our replication exercise, indicates that around the year 1860, the average cabinet speech becomes more comprehensible than the average non-cabinet speech
- **OLS Regression:** Very similar coefficients, slight differences in p-values.

Autopsy

- Data retrieval for raw data proved more difficult than for cleaned and preprocessed "bigframe" data
- Sampling of raw speech data may have led to some discrepancies in outcome
- Very high and very low FRE scores led to some confusion
- Satisfaction with multivariate regression replication results
- Overall, replicated trends reflected trends in the original paper
- Our additions were relatively easy to implement

Suggested Improvements

- Easier access to raw data
- More party affiliation analysis
- More detailed text analysis
- Comparing FRE to different readability scores (Dale Chall, SMOG)