

“Simultaneous and Systematic Abolition”?

Automated Content Analysis and the 1789 *Cahiers de Doléances*

Jeff Jacobs

jjp2122@columbia.edu

Columbia University

June 11, 2018

The French Revolution

- France, 24 January 1789: King Louis XVI summons the Estates-General, requesting “a concourse of our faithful subjects, to assist us in surmounting all the difficulties we find relative to the state of our finances.”
- Why is this interesting?

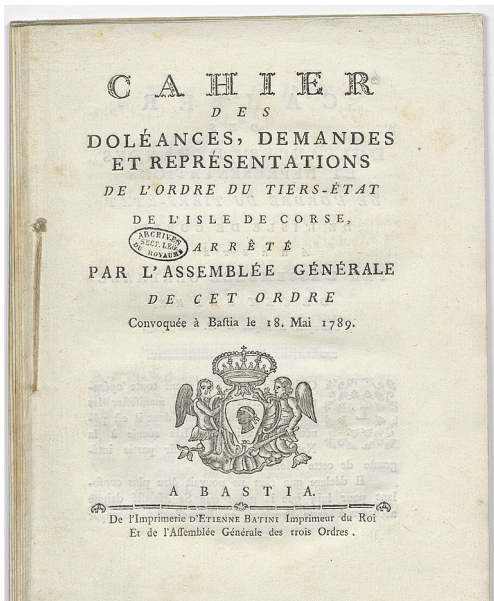
The French Revolution

- France, 24 January 1789: King Louis XVI summons the Estates-General, requesting “a concourse of our faithful subjects, to assist us in surmounting all the difficulties we find relative to the state of our finances.”
- Why is this interesting?

sée. Et pour cet effet vous mandons et très-expressément enjoignons qu'incontinent, la présente reçue, vous avez à convoquer et assembler en notre ville de dans le plus bref temps que faire se pourra tous ceux des trois États du bailliage (ou sénéchaussée de...) pour conférer et pour commaniquer ensemble, tant des remontrances, plaintes et doléances, que des moyens et avis qu'ils auront à proposer en l'assemblée générale de nosdits États;

tl;dr: Each of the three Estates within every Bailliage should confer and draft a book of “remontrances, complaints, and grievances” to be brought by their representatives to the Estates-General.

The Cahiers de Doléances



The Significance of the *Cahiers*

“The only [revolution] at the beginning of which the different classes of society were enabled to present an authentic account of the ideas they had conceived, and express the feelings which animated them, before the revolution had distorted or modified those ideas and feelings.” (Tocqueville 1856, 262)

Interpretations

- **Tocqueville**: “In short, reading the grievance books of the nobles leads to the conclusion that they might have made the Revolution themselves had they only been commoners, for that was the only ingredient they lacked.” (Tocqueville 1856, 244)
 - ⇒ Content of Noble *cahiers* \approx content of Third Estate *cahiers*
- **Lefebvre’s** (1939) Marxist interpretation: Social revolution, a class struggle of the Bourgeoisie (“upper crust” of the Third Estate) against the entrenched aristocracy (Nobility/Clergy)
 - ⇒ Content of Noble *cahiers* \neq content of Third Estate *cahiers*
- **Cobban’s** (1956) revisionism: Not a social but a political revolution. Feudalism had disappeared long before. Peasantry, not bourgeoisie, were central actors. “A revolution not for, but against, capitalism.” (Cobban 1964, 172)
 - ⇒ Content of Parish *cahiers* \neq content of Third Estate *cahiers*

Topic Modeling



Figure 1: Results of a Latent Dirichlet Allocation (LDA) topic model run on the New York Times corpus with the number-of-topics parameter $K = 10$. Figure from Blei (2012).

The Data

- Extracted from the *Archives Parlementaires* (French Revolution Digital Archive), Vols. 1–6
 - Invalid XML – *beware!*
- *Cahiers* for 280 Bailliage × Estate pairs
 - 138 Noble *cahiers* and 142 Third Estate *cahiers* from 167 Bailliages
- 10,006 extracted grievances
 - Regular Expressions; could potentially do better with ML
 - 35.7 grievances per *cahier*
- Preprocessing: stopwords removed (*des, il, si*), French Snowball Stemmer (*malade, malades* → *malad*), terms appearing in less than 3 grievances removed

Most Indicative Words

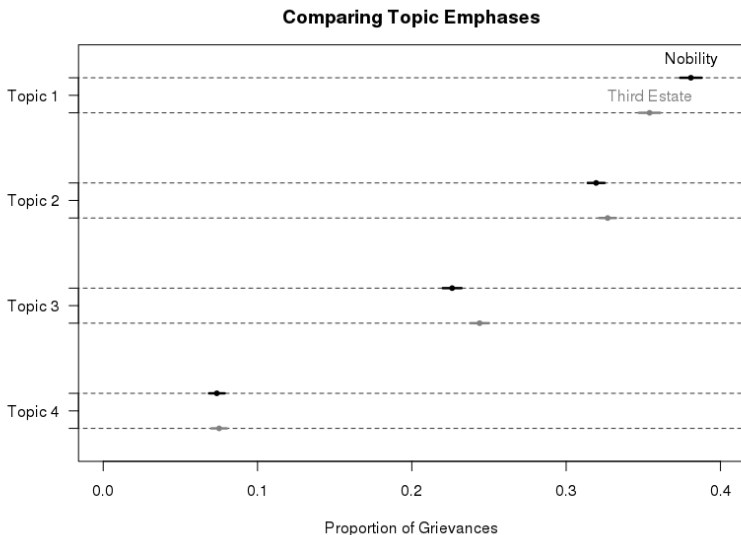
Topic 1	Topic 2	Topic 3	Topic 4
Assembly Procedures	Judicial Reform	Rights	Church Reform
etat general provinc nation assembl provincial consent bailliag ordre mesur	vill criminel civil jug ministr libert municipal proc justic abus	droit seigneur commerc libert privileg offic citoyen religion lettr tarif	ecclesiast vicair dioces paroiss conseil revenus abbay communaut avril ordon

Table 1: 10 Most Indicative Words for Each Topic

Inferred Topics

- (1) *Assembly Procedures*: The structure and decision-making processes that will be utilized in the upcoming Estates-General. Most likely concerns the issue of voting by order or by head, and how to represent the Third Estate.
- (2) *Judicial Reform*: Grievances/recommendations regarding criminal and civil law. Often mentioning specific judges or abuses observed in the particular bailliage.
- (3) *Rights: Seignorial, Commercial, and Civil*: Grievances regarding feudal structure, freedom of commerce, and (whether implicitly or explicitly) the Rights of Man and the Citizen.
- (4) *Ecclesiastical Reform*: Grievances concerning the structure and revenue-collection of the Church. Sometimes specifically corrupt dioceses or parish priests are mentioned.

Topic Emphasis By Estate

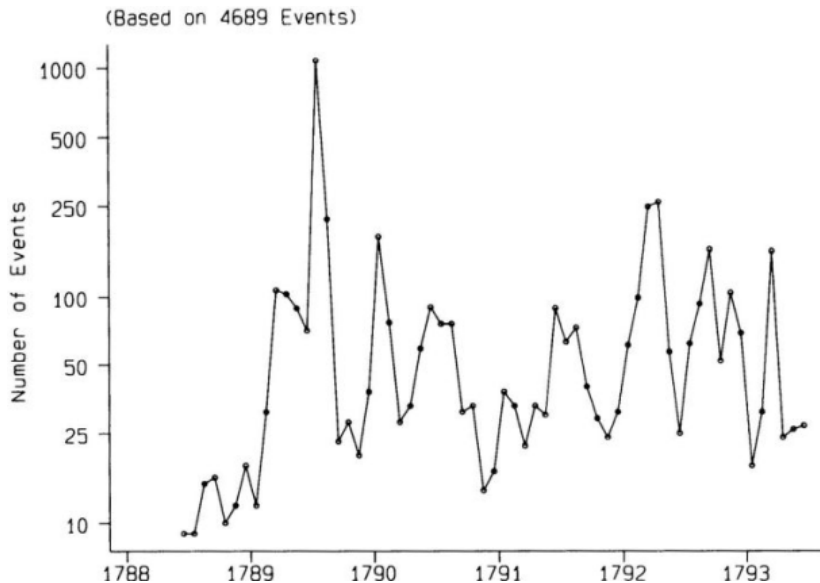


Why Is This Cool?

Code0	Code1	Topic	Agreement	Predicted	Residual
C	EG	* Estates-General	0.343	0.412	-0.070
C	PL	Political liberties	0.284	0.312	-0.028
E	0	* Economy (misc)	0.086	0.172	-0.086
J	1	Justice (general)	0.198	0.193	0.005
J	CM	Criminal prosecution	0.247	0.252	-0.006
J	CO	Courts	0.184	0.207	-0.023
J	CV	Civil law	0.173	0.171	0.002
J	DP	Due process	0.359	0.378	-0.019
J	EA	Court enforcement	0.274	0.263	0.011
J	LP	Legal professions	0.150	0.143	0.007
R	0	Religion (misc)	0.053	0.040	0.013
R	1	Religion (general)	0.111	0.011	0.100
R	CF	Church finances	0.240	0.289	-0.049
R	CL	Clergy	0.224	0.230	-0.006
R	CO	Church organization	0.141	0.123	0.018
R	CS	Church-state relations	0.188	0.198	-0.010
R	DI	Tithe	0.170	0.186	-0.016
R	MO	Morality	0.255	0.267	-0.012
S	CM	* Criteria of mobility	0.212	0.300	-0.087
S	SE	* Seigneurial regime	0.148	0.203	-0.055

Table 2: Table 15-4 from Shapiro and Markoff (1998)

Covariates: Revolt Data from Markoff (1996)



Findings: Aggregate

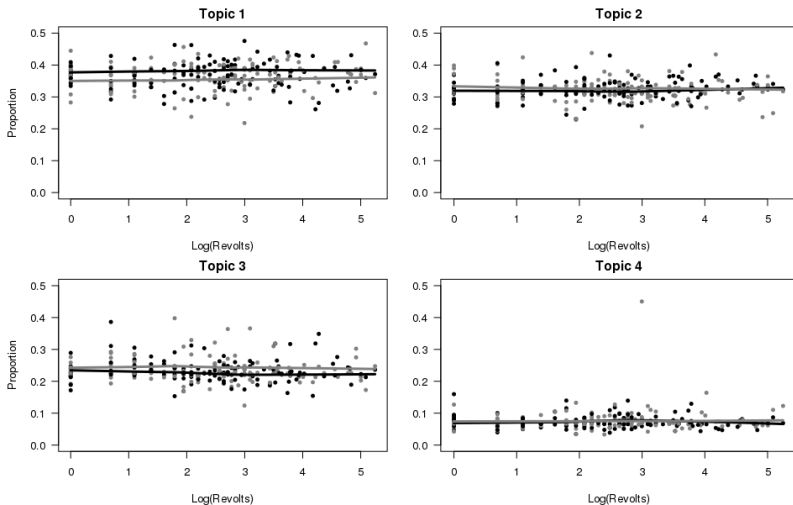


Figure 2: Topic emphasis vs. Frequency of peasant revolt. Black indicates Nobility, grey indicates Third Estate data.

Findings: Pre-Cahier Drafting

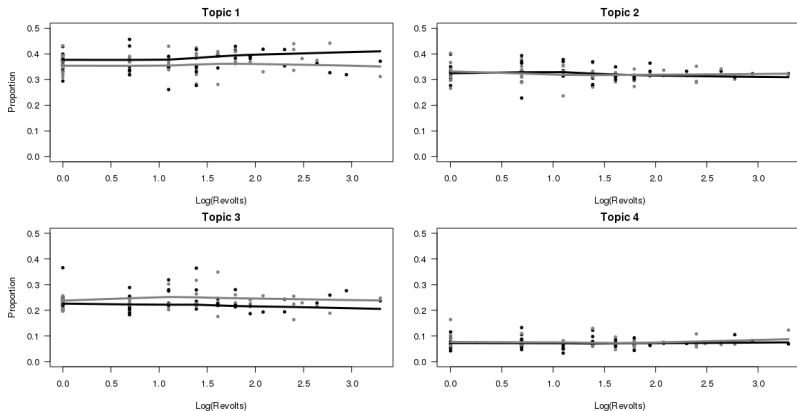


Figure 3: Topic emphasis vs. Frequency of peasant revolts *before* drafting. Black indicates Nobility, grey indicates Third Estate data.

Findings: Post-*Cahier* Drafting

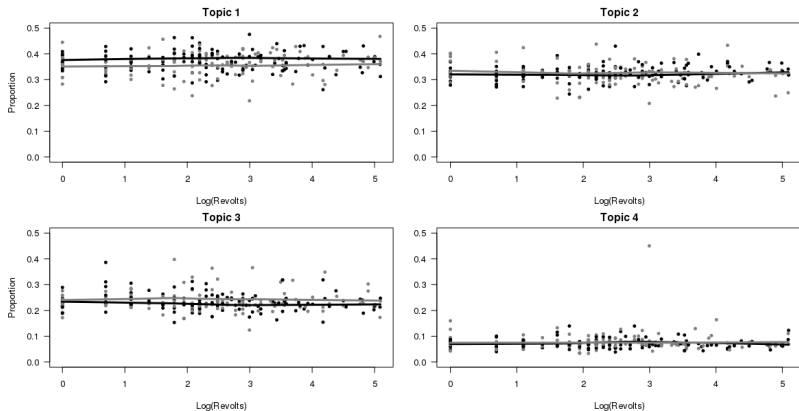


Figure 4: Topic emphasis vs. Frequency of peasant revolts *after* drafting. Black indicates Nobility, grey indicates Third Estate data.

Comparing Estate Coefficients

	Topic 1		Topic 2	
	Nob Emphasis	TE Emphasis	Nob Emphasis	TE Emphasis
Any Revolts	0.009 (0.012)	0.008 (0.016)	-0.001 (0.008)	0.006 (0.014)
Constant	0.374*** (0.011)	0.347*** (0.016)	0.320*** (0.008)	0.321*** (0.013)
Observations	135	139	135	139
R^2	0.004	0.002	0.0001	0.001
Adjusted R^2	-0.003	-0.006	-0.007	-0.006
Coeff. Δ z-score		0.048		-0.424
Coeff. Δ p-value		0.962		0.672

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3: Topic emphasis on any-revolts dummy. Coefficient-comparison test (Clogg et al. 1995) results indicate no significant differences between the relationship for Nobility and Third Estate *cahiers*.

Comparing Estate Coefficients (cont'd)

	Topic 3		Topic 4	
	Nob Emphasis	TE Emphasis	Nob Emphasis	TE Emphasis
Any Revolts	-0.012 (0.009)	-0.013 (0.016)	0.005 (0.011)	-0.0003 (0.008)
Constant	0.237*** (0.008)	0.257*** (0.015)	0.069*** (0.011)	0.075*** (0.008)
Observations	135	139	135	139
R ²	0.015	0.005	0.001	0.00001
Adjusted R ²	0.007	-0.002	-0.006	-0.007
Coeff. Δ z-score		0.037		0.367
Coeff. Δ p-value		0.970		0.714

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 4: Topic emphasis on any-revolts dummy. Coefficient-comparison test (Clogg et al. 1995) results indicate no significant differences between the relationship for Nobility and Third Estate *cahiers*.

Takeaways

- Main point was to show feasibility of automated content analysis with covariate analysis
- All code+data is on GitHub:
<https://github.com/jjacobs3/cahiers-de-doleances>
- Change hyperparameters → Discover new topics
- Now, alternative covariates can be “dropped in”:
 - Degree of Urbanization (Shapiro et al. 1973)
 - Rate of Emigration (Greer 1951)
 - Incidence of the Terror (Greer 1935)

Extensions

- Supervised approach
 - If you *do* know precisely what you're looking for, “steer” the topic discovery via training examples
 - How few training examples (what proportion p of the N total grievances) do we need to get within ϵ of the “true” results from a manual content analysis?

$$\begin{array}{ll} \text{minimize} & p \\ \text{subject to} & \text{MSE}(p, N) = \sum_{i \in S}^N (t_i(R_p) - a_i(R_p))^2 < \epsilon \end{array}$$

- Mostly done
- Diachronic corpora
 - Code is done, but covariate collection still in progress (next slide)

Corpora

Cross-Sectional

- Literature of the “Hundred Flowers”: Nieh (1981)
 - Digitized, OCRred, and simple (non-comparative) topic model + word clouds produced
- Reports on “contentious performances” from 18th- and 19th-century Britain: Tilly (2008)
 - TBD

Diachronic

- Rap Lyrics, 1979–Present: RapGenius
 - >2 million songs scraped, w/metadata
 - Covariates: Metropolitan areas, militarization of police, murder rate, poverty rate, incarceration rate, introduction of Stop & Frisk, police shootings, Black Lives Matter
 - Working on this w/an undergrad

The End

Thank You!