



Studying Events: Creating Event Data

GRAEME ROBERTSON

UNC AT CHAPEL HILL

Outline

- ▶ Why do we care about protest?
- ▶ Why do we need two theories to study events?
- ▶ Different data generating processes
 - ▶ Official sources
 - ▶ Artisanal data
 - ▶ State Archives
 - ▶ Media
 - ▶ Activists
 - ▶ Surveys
 - ▶ Generic Issues with artisanal data
- ▶ A practical application – The MVD Dataset – Contention in Russia 1997-2000

Why do we care about protest events?

- ▶ Is there something underlying dimension, some feeling, some force that we are trying to capture? e.g relative deprivation
- ▶ Are protests a proxy for some deeper social forces? e.g. class consciousness
- ▶ Or do we just care about protest?
- ▶ Or about what protest tells us about context – how do events relate to each other (waves) and to other elements of political context (POS)?

Tilly's 2 related questions

- ▶ What is the theory of the underlying phenomena (Type 1 Theory)?
 - ▶ What causes events?
 - ▶ What do events cause?
- ▶ What is the theory of the data generating process (Type 2 Theory)?
 - ▶ What is the chain of events by which I receive a trace/some evidence of this event?
- ▶ And what can the latter tell me about the former (and often the inverse too).

Data Generating Processes: Official Sources

- ▶ Most obvious subject of official data on contention are strikes.
- ▶ Consequently (partly) these are the most studied -- Organization; Bargaining Power; Information.
 - ▶ Roberto Franzosi's amazing 1995 book, *The Puzzle of Strikes*. Also his article, "One hundred years of strike statistics: Methodological and theoretical issues in quantitative strike research." *ILR Review* 42.3 (1989): 348-362.
 - ▶ See also Haimson, Leopold H., and Charles Tilly, eds. *Strikes, wars, and revolutions in an international perspective: Strike waves in the late nineteenth and early twentieth centuries*. Cambridge University Press, 2002.
- ▶ But, like any official reporting then we have to think about the bureaucratic data generating process. e.g Russian strike data.

Artisanal Data Collection: Archives

- ▶ What is the data generating process for archives?
- ▶ Depends a lot on the archival source
- ▶ Finkel, Evgeny, Scott Gehlbach, and Tricia D. Olsen. "Does reform prevent rebellion? Evidence from Russia's emancipation of the serfs." *Comparative Political Studies* 48.8 (2015): 984-1019.
 - ▶ Tsarist era administration – varying intensity of reporting over time – why? Who wants to report what? Police agencies versus political appointees?
 - ▶ Soviet era political agendas.

Artisanal Data Collection: Media

- ▶ Probably the most commonly used approach to coding data on protest and contention.
- ▶ Extraordinary variety of sources:
 - ▶ Broad collection versus targeted data collection
 - ▶ (e.g. Sullivan Mexico protest collection vs Teitelbaum HPS Dataset).
- ▶ Problems with newspaper accounts
 - ▶ Theoretical: Media are not concerned with creating a valid account of all the incidents of a certain kind of event.
 - ▶ Empirical: Cross-national comparisons are very difficult indeed, perhaps especially for cross-sectional analyses, but also for different kinds of events and different places within countries
 - ▶ Herkenrath, Mark and Alex Knoll. 2011. "Protest Events in International Press Coverage: An Empirical Critique of Cross-national Conflict Databases," *International Journal of Comparative Sociology* 52, 3: 163-180.
 - ▶ Some commonly used data sources – Banks Protest Event Data – are very poor indeed.

Artisanal Data Collection: Activists

- ▶ Often a good source where official sources or media attention is (deliberately) sparse.
 - ▶ e.g. Peasants are generally very difficult to study using archival or media sources.
- ▶ Particularly useful in cases of international or indigenous intellectual involvement.
 - ▶ Local newspaper collections compiled by activists
 - ▶ Guillermo Trejo's work on the Zapatista uprisings and peasant disputes in Chiapas and southern Mexico.
- ▶ Online sources:
 - ▶ Reuter and Robertson (IKD.ru; kprf.ru)
 - ▶ Lankina namarsh.ru

General Problems with Artisanal Data

- ▶ Depends often on luck and serendipity
 - ▶ Are there activists and how reliable are they?
 - ▶ Thank goodness for The Wayback Machine!
- ▶ Absence of standardization makes cross-case comparisons difficult.
- ▶ Data sources come ... and go...

Alternatives to event counts -- surveys

- ▶ Population surveys
 - ▶ Can allow comparison between protesters and non-protesters (Javeline 2003, Pop-Eleches, Robertson and Rosenfeld 2018).
 - ▶ But the small numbers problem
 - ▶ though Rosenfeld case control method (APSR 2017)
 - ▶ Also survey methods are sometimes highly misleading (Biggs 2015)
- ▶ Protester surveys
 - ▶ Allow you to capture many more protesters (Smyth and Oates 2015, Onuch 2014)
 - ▶ Selection on the dependent variable
 - ▶ Substantial problems in comparisons across issues (Walgrave et al. 2016)

Practical Application: MVD Dataset on Contention in Russia 1997-2000

- ▶ Serendipity and the data collection process.

The Raw Data

МВД-ОПЕРАТИВНЫЕ_СВОДКИ от 20.10.2000

ИНФОРМАЦИЯ ОБ ОБСТАНОВКЕ " 19 " октября 2000 год

З прошедшие сутки в Российской Федерации общественнополитическая обстановка существенно не изменялась.

В городе Ростове-на-Дону на площади Советов у здания полномочного представителя Президента РФ в Южном федеральном округе прошли несанкционированные пикетирования 40 бывших военнослужащих контрактной службы, принимавших участие в боевых действиях на территории Чеченской Республики, с требованием выплаты задолженности по денежному довольствию.

Тем же у здания Министерств социальной защиты и селения по Ростовской области 56 участников ликвидации в районе Чернобыльской АЭС провели несанкционированные пикетирования с требованием перерасчета выделяемых денежных средств и возмещение причиненного вреда здоровью.

С требованием возмещения задолженности по заработной плате продолжают забастовки - учителя 15 школ (433 человек) в Кош-Агачском районе Республики Алтай; работники муниципального предприятия объединенных котельных и тепловых сетей (170 человек) в городе Лесозаводске Приморского края;

голодовку - врачи детской поликлиники (10 человек) в городе Кулебики Нижегородской области.

С неадекватными требованиями возобновили забастовку члены экипажа судна "Берег мечты" ОАО "Востоктрансфлот" (29 человек) в городе Владивостоке.

С требованием выплаты задолженности по пособиям из причиненный ущерб здоровью продолжают голодовки: участники ликвидации в районе Чернобыльской АЭС в офисе Ворошиловского отделения общественной организации

"Союз чернобыльцев" в городе Ростове-на-Дону (4 человек); в городе Константиновске (5 человек) и Шхты (10 человек) Ростовской области. Прекратили забастовку работники ЖКХ Киясовского района (70 человек) Удмуртской Республики.

Converting text to numbers: generating a codebook

- ▶ Convert text to data so that you could convert data back into text.

Principles of aggregation

- ▶ Aggregation is always theory dependent (Type 1 Theory).
- ▶ Tilly's orienting questions:
 - ▶ What are they doing (repertoire)?
 - ▶ Who is protesting (collective identity)?
 - ▶ What would it take to satisfy demands (state)?
 - ▶ Who would need to act (public authorities)?
- ▶ Problems of aggregation:
 - ▶ Aggregation necessarily involves information loss
 - ▶ How to measure protest: Events, participants or intensity (Biggs 2016)?