

Codebook for New York Times Index Dataset

Overview and Sampling

This database is a random sampling of the New York Times Index from 1946 to 2003 (over 37,000 records). The Index itself includes thousands of entries per year. Each entry includes a short abstract of the story, the date when the story appeared, and some additional information such as the page on which it appeared; whether the story was short, medium-length, or long; etc. Entries are grouped together under subject headings. Subject headings also include many references to other related subjects where additional entries are located.

The unit of analysis for this database is an entry. Subject headings, "see also" references, and everything else are skipped. The sample is derived by coding the first entry on every odd-numbered page of the index. Entries are generally identifiable because they have a date, section number, and a page number. Subject headings and other references do not have full citations. Exceptions to the first-entry-on-odd-pages rule occurred infrequently when the first entry on the odd page consisted of only the date, section number, and page number, with no accompanying summary. (In such a case we coded the next entry.)

Tracking Issues Across Time and Across Datasets

The topics variable allows you to track an issue across time as it is treated in the media, in congressional activity and in budgetary activity.

Final Note

There are two complications to the New York Times and its Index that users should keep in mind. First is that occasional format changes have led to different numbers of articles on each page. Second is that the New York Times and its Index have varied in size over the decades. The accompanying Excel spreadsheet, NYT-Pages.xls gives detailed information on the number of pages in the Index and on an estimate of the number of articles per page for each of the years included in our dataset.

We recommend that users adjust our NYTimes data by dividing by the total number of entries per year. Analyzing the data in this way will control for any effects of changing size and composition of the annual indices.

Coding Scheme

1. Page number

Actual page number of the NY Times Index where the entry is found (for identification purposes)

2. First 3 words

The first three words exactly as they appear, for identification purposes.

3. Date

Date of the story in this format: mm/dd/yy

The New York Times Index uses the following abbreviations to identify the month in which the story appears:

- Ja January
- F February
- Mr March
- Ap April
- My May
- Je June
- Jl July
- Ag August
- S September
- O October
- N November
- D December

4. Length:

Each entry has a note that indicates whether the story was short, medium, or long.

- 1. Short
- 2. Medium
- 3. Long
- 9. Not indicated (before 1974, this information was generally not included)

5. Page One

Each entry mentions the page number where the story begins. The New York Times Index uses various abbreviations to indicate the pagination of the paper. These are described in the beginning section of each annual volume (the first few pages include a section on "How to Use this Volume" or something similar). In some years, a page-one story would be indicated by something like My 5, A1:3, which would mean that the article appeared on May 5th, Section A (the main section), page 1, column 3. In other years, that article might be listed as My 5, 1:3. Also, sometimes it differs from the

Sunday edition to the daily editions. Sometimes it is listed as My 5, I 1:3, using roman numerals to indicate the section.

- 1. Front page story.
- 0. Not a front page story.

6. Summary of story topic:

We include a short summary the topic of the story. This entry should be complete enough to allow anyone to read through the database and understand what the article was about. Some topics can have a quick description ("movie review" or "home improvement idea" but other topics should have more complete descriptions: "President Nixon refused to divulge information to Watergate committee in spite of court order to do so" or "President Clinton is 15 points ahead in polls; Dole campaign considers shake-up") Whenever space permits, we include specific mentions of public figures, courts, interest groups, etc.

7. Topic code

We have 27 topic codes. Each article is coded by topic area and checked by at least 3 coders to ensure consistency and correspondence with the congressional hearings, statutes, and CQ stories datasets of the Policy Agendas Project. The topic codes are as follows. For more detail, including examples of each topic, see the detailed topic codebook.

- 1. Macroeconomics
- 2. Civil Rights
- 3. Health
- 4. Agriculture
- 5. Labor, Immigration, and Employment
- 6. Education
- 7. Environment
- 8. Energy
- 10. Transportation
- 12. Law, Crime, and Family Issues
- 13. Social welfare
- 14. Community Development and Housing
- 15. Banking, Finance and Domestic Commerce
- 16. Defense
- 17. Space, Science, Technology, and Communications
- 18. Foreign Trade
- 19. International Affairs
- 20. Federal Government Operations
- 21. Public Lands and Water Management
- 24. State and Local Government Administration
- 26. Weather and Natural Disasters

- 27. Fires
- 28. Arts and Entertainment
- 29. Sports and Recreation
- 30. Death Notices
- 31. Churches and Religion
- 99. Other, Miscellaneous, and Human Interest

8. NY State or Metro news

Many stories in the New York Times will focus specifically on things that happen within city or in its region, or report on the actions of the State government in Albany. For instance, many of the crime stories will be reports of recent events in New York. Other stories are part of its reporting of national affairs. Since the local news may be driven by different concerns than the national news, we distinguish between these.

- 1. Exclusively NY City news.
NY City is comprised of five boroughs: Manhattan, Brooklyn, Queens, Staten Island, and the Bronx. Also includes stories about Long Island. This code is designed to identify stories of purely local interest.
- 2. NY State or regional news
(actions in the legislature, by the Governor, things that happen in small towns in the New York region, and regional things without national impact). Also includes CT or NJ news.
- 3. Anything else.
(National events that occur in New York City, like meetings at the United Nations, are included here. Activities of federal government officials and activities that have national or international importance but happen to occur in New York City, are included here.)

9. U.S. or foreign

In distinguishing between domestic and foreign stories, we add one complication. A story about crime in Detroit is clearly "domestic." One about crime in Calcutta is clearly "foreign." However, a story about how crime overseas is causing the American government to send advisors, or American tourists to alter their plans, or that mentions any other US angle on the story, should be coded as "US angle on foreign story." Discussions of foreign wars, for example, may or may not discuss the probability that the US may get involved or the effect of the fighting on US businesses or citizens abroad.

- 1. Domestic
(includes US territories abroad like Puerto Rico)
- 2. Discussion of US angle on foreign story
(includes US actions abroad, such as war). Anytime the US is mentioned in the same entry with a foreign country or with terms like "international," "worldwide," etc. Also includes wars in which the US is a party, even if the US is not

specifically mentioned. So, for example, all stories concerning the Vietnam or the Persian Gulf wars are coded 2, even if they do not specifically mention the US.

- 3. Foreign, with no discussion of US involvement or concern.
This includes UN actions that do not specifically mention the US.

10. U.S. Government Action

Some entries are about events, problems, and social concerns having nothing to do with the government. Such things as new movies, sports events, crimes, and the like often mention nothing about public policy. Others mention something about government activities or discuss action the government should take. This variable is meant to distinguish between those articles mentioning anything about government activities or government officials and those that do not.

- 1. Mentions U.S. government official, government agency, government action or discussion of action the government has taken (or should take), etc. Includes any level of US government, including state and local. Does not include foreign governments.
- 0. Unrelated to questions of American public policy.

11. Federal agency

- 1. Mentions a federal (national-level) agency.
Includes references to the heads of Departments, such as the Secretary of State, US Trade Negotiator, as well as references to any federal agency, such as the Army, Post Office, EPA, etc. Also includes any mention of national parks, military activity, diplomatic activity, astronauts, postmaster, etc., even if the name of the specific agency is not mentioned. Includes executive branch advisory bodies, commissions, etc. unless these are clearly under the auspices of the President or Congress (see below).
- 0. does not.

12. The President or the Vice President

- 1. mentions the current President or his staff (including "White House," "Clinton Administration," "Presidential spokesman"). Also includes historical references to past presidents concerning actions while in office. Also includes the Vice President. Also includes presidential advisory bodies, commissions, etc.
- 0. does not

13. Congress

- 1. mentions the Congress, the House or Senate, Members of Congress, staff members, or the legislative process. Also includes congressional advisory bodies, Library of Congress, other legislative branch activities of the federal government. Also includes discussion of legislative debates where they clearly took place in

Congress, even if Congress is not specifically mentioned. Does not include discussion of previously passed legislation unless Congress is specifically mentioned.

- 0. does not.

14. The courts

- 1. mentions court or judicial activities (may be federal, state, or courts at any level; also includes grand jury activities, as well as mentions of "filing suit" and "standing trial," "lawsuit," "contract suit," "trial news," "litigation," "jury," "indicted," "pleaded guilty," "acquittal," "witness testimony," "file motion," "sentenced," "tort system," "legal system," "US penal experts," and "warrant issue").
- 0. does not.

15. State or local government actions

- 1. Mentions state or local government officials or agencies other than courts. Includes activities of state and local government officials only, not federal officials who happen to represent particular states or localities: Members of Congress, Senators, US Attorneys, and federal prosecutors are all federal employees, not state or local officials. State and local officials include: district attorneys, sheriffs, county prosecutors, state agency officials (e.g., "health officials," "environmental officials"), members of the state senate or legislature, and other state and local workers ("NY City transit workers," "NY City road repair crew," "NY City bus driver," and public school administrators, teachers and school boards acting in an official capacity). Also includes "city fathers," "civic leaders," etc. Includes mentions of public libraries, public schools, municipal parking garages, etc. Also includes police activities, state prison activities, etc. In police activities, includes any mention of people being "arrested," "charged," or "jailed." In campaign coverage, includes mentions of current state or local officials as either the incumbent or challenger, or as the winner of the election (e.g., governor-elect, state representative-elect).
- 0. does not.

16. Candidates and campaigns

- 1. discusses candidates for office or elections, including incumbents in election campaigns. Includes mentions of political party conventions, news coverage of the campaigns/elections, redistricting, FEC activities, campaign contributions, party platforms, and candidate debates.
- 0. does not.

17. Interest groups

- 1. mentions a specific group (e.g., "NAACP"). Includes political parties and individuals who hold positions within the party organization. Does not include the lobbying activities of specific businesses. Includes unions, foundations (e.g., Geraldine Dodge Foundation), institutes (e.g., International Food Policy Research Institute), and non-profit organizations (e.g., Girl Scouts of America). Includes discussion of publications and newsletters put out by interest groups (e.g., "ABA Journal," "JAMA." The Roman Catholic Church and the pope are considered interest groups when either takes a policy stand through an editorial, etc. Does not include activities of foreign interest groups abroad.
- 0. does not

Written by:

Frank R. Baumgartner and Bryan D. Jones

(Revised February 1999)