

# URBAN SOCIAL DISORDER CODEBOOK

KARIM BAHGAT (KARIM.BAHGAT.NORWAY@GMAIL.COM)  
HENRY THOMSON (HENRY.THOMSON@ASU.EDU)  
HALVARD BUHAUG (HALVARD@PRIO.ORG)  
HENRIK URDAL (URDAL@PRIO.ORG)

VERSION: 3.0  
DATE: DAY MONTH 2022

When using this dataset please cite:

Thomson, Henry, Karim Bahgat, Henrik Urdal, and Halvard Buhaug, 2022.  
“Urban Social Disorder 3.0: A global, city-level event dataset of political mobilization and disorder”,  
*Journal of Peace Research* X(Y): AA–BB.

This codebook should be cited as:

Bahgat, Karim, Henry Thomson, Halvard Buhaug, and Henrik Urdal. 2022.  
“Urban Social Disorder Codebook, version 3.0”, Peace Research Institute Oslo.  
<https://www.prio.org/Data/Armed-Conflict/Urban-Social-Disorder/>

## Acknowledgements:

Work on the USD 3.0 dataset has been funded by grants from the... Previous versions of the dataset have been supported by grants from the European Research Council (#648291), the Norwegian Ministry of Foreign Affairs (#QZA-13/0365), the research Council of Norway (#240315), and the World Bank (#PO7145929). The information contained in the USD dataset has been collected and coded by: Abhimanyu Arora, Karim Bahgat, Stephanie Frommer, Sydnee Hertle, Elyse Leonard, Oddbjørn Midttømme, Elisabeth L. Rosvold, Kiran Sagoo, Friederike Schwebler, Mia Senechal, Nathaniel Smith, Henry Thompson, Esther Trappeniers, Tarjei Vaa, and Nan Wu.

## Contents

Introduction.....	1
List of Variables.....	2
The Events Dataset .....	2
The Reports Dataset.....	6
The Cities Dataset .....	7
On the data collection .....	10
The selection of cities .....	10
Source material .....	10
The Search .....	11
Screening the Reports .....	12
Coding the Events.....	13
Inter-coder reliability .....	14
References.....	16
Appendix 1: Specific coding rules .....	18

## Introduction

For the first time, more people now live in cities than in rural areas. By 2050, rural-urban migration in the developing world will take the global urbanization rate to over seventy percent (Saunders, 2010; United Nations, 2018). While urbanization is generally linked with economic and social progress and increasing living standards (World Bank, 2016), it is no longer synonymous with rapid economic growth, leading to increasing concerns about premature deindustrialization, poverty, and inequality in cities of the global South (Gollin, Jedwab, and Vollrath, 2016; Rodrik, 2016). Scholars are also beginning to address the potential consequences of contemporary urban growth for democracy in the developing world (Glaeser and Steinberg, 2017; Post, 2018).

As the world's population becomes increasingly concentrated in urban areas, so too do political mobilization and violence. Unlike conventional forms of armed conflict, which – despite an uptick in recent years – have been on a long-term decline (Goldstein 2011; Pettersson and Öberg, 2020), urban social disorder is becoming more prevalent, and there is an upward trend in lethal urban violence (Bahgat, Buhaug, and Urdal, 2018). This trend threatens to continue in future decades (Goldstone, 2010; Beall, Goodfellow, and Rodgers, 2013; Moncada, 2013) and has attracted the attention of policymakers: The United Nations included urban security in both the 2015 Sustainable Development Goals and the 2016 Habitat New Urban Agenda.

However, empirical research on urban mobilization and violence has been hampered by a lack of data with a broad geographic and temporal scope. Existing efforts to collect detailed information on conflict events, such as the Armed Conflict Location and Event Dataset (ACLED) (Raleigh et al. 2010); the Social Conflict Analysis Database, (SCAD) (Salehyan et al. 2012); the Uppsala Conflict Data Program's Georeferenced Event Dataset (UCDP GED) (Sundberg and Melander 2013); and the Global Database on Events, Language, and Tone project (GDELT; [www.gdeltproject.org](http://www.gdeltproject.org)) primarily cover the post-Cold War period and therefore cannot detect long-run changes in patterns of urban political mobilization and violence. Datasets with broader temporal scope, such as the Cross-National Time-Series Data Archive (Banks and Wilson, 2020) and the Nonviolent and Violent Campaigns and Outcomes Data Project (Chenoweth and Lewis, 2013), do not provide detailed information on the location of individual events within states and capture only specific types of collective mobilization and violence.

As an alternative, we provide the third iteration of the Urban Social Disorder (USD) event dataset that facilitates empirical research on political mobilization, contention, and violence in urban areas across the globe. The scope of the USD dataset is to collect events that occur in select individual cities, not to represent the situation of entire countries. Whereas previous versions of the USD data (Urdal and Hoelscher 2012; see also Urdal 2008) was limited to particular countries or continents, the updated version is the first release with global coverage of cities that match the city selection criteria, containing 186 national capitals and major urban centers in the period 1960-2014. The latest version also includes additional metadata about the capital status of cities,<sup>1</sup> and minor updates to the data, e.g. changes or removals of events included in previous versions to ensure that all events conform to the coding rules.

---

<sup>1</sup> Additionally, some city names were updated to reflect the most recently accepted variant: Calcutta was changed to Kolkata; New Delhi to Delhi; Rangoon to Yangon; Saigon to Ho Chi Minh City. However, the city ID continues remain the same across all versions.

## List of Variables

### The Events Dataset

The events dataset is the main dataset used for analysis, and contains the following variables:

<b>EVENTID</b>	<b>Event ID</b> (text; ID) A globally unique ID for the event. It consists of a unique city ID and a unique number for each event coded separated by a dash, to make it globally unique, i.e. “[City]-[Event]”.
<b>CITY_ID</b>	<b>ID of City</b> (ID) ID of the city in which the event occurred. This city ID can for instance be used to join with the “cities” table that contains additional metadata about the city.
<b>CITY</b>	<b>Name of City</b> (text) Name of the city in which the event occurred.
<b>COUNTRY</b>	<b>Name of Country</b> (text) Current official name of the country in which the city is located (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>ISO3</b>	<b>ISO3 Alphanumeric Country Code</b> (3 character text; ID) Alpha-3 country code according to the ISO 3166-1 standard (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>GWNO</b>	<b>Gleditsch-Ward Country Code</b> (1-3 digit numeric; ID) Numeric country code according to the Gleditsch and Ward (1999) list of independent states. This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>COUNTRY_HIST</b>	<b>Name of Historical Country</b> (text) Contemporaneous name of the country in which the city was located at the time of the event, according to the Cshapes dataset of historic country borders (Weidmann, Kuse, and Gleditsch 2010). For cities in countries prior to independence this is set to missing.
<b>GWNO_HIST</b>	<b>Historical Gleditsch-Ward Country Code</b> (1-3 digit numeric; ID) Contemporaneous numeric country code according to the Gleditsch and Ward (1999) list of independent states, according to the Cshapes dataset of historic country borders (Weidmann, Kuse, and Gleditsch 2010). For cities in countries prior to independence this is set to missing.
<b>REGION</b>	<b>Name of Region</b> (text) Name of the continent/geographical region to which the city belongs.  Asia Latin America Middle East and North Africa Sub-Saharan Africa

<b>LONG</b>	<b>Longitude</b> (decimal) Longitude degree of the city's location.
<b>LAT</b>	<b>Latitude</b> (decimal) Latitude degree of the city's location.
<b>PTYPE</b>	<b>Problem Type</b> (2 digit numeric; nominal) Identifies the general level and type of event according to the following list of values. While the PTYPE variable is considered to be a nominal variable, it is structured with some sense of ordinal ranking of societal conflict magnitude (based on degree of institutionalization of political activity).
10	<b>General Warfare:</b> Distinct event related to a protracted, interactive, and violent conflict involving at least one, organized, non-state actor group fighting with government authorities. Can be either over ethnic, political or economic issues.
20	<b>Inter-communal Warfare:</b> Distinct event related to a protracted, interactive, and violent conflict involving at least two militant organizations representing a singular ethnic or religious identity group (i.e., non-state actor) fighting each other without (explicit) involvement of a government.
30	<b>Armed Battle/Clash:</b> Distinct, continuous, and coordinated interaction involving opposing, organized armed forces representing government and/or group interests.
31	<b>Armed Attack:</b> Distinct, continuous, and coordinated action staged by a singular, militant political or identity group against government authorities or institutions representing an "other" group.
40	<b>Pro-Government Terrorism (Repression):</b> Distinct event related to a persistent, directed-violence campaign waged primarily by government authorities, or by groups acting in explicit support of government authority, targeting individual, or "collective individual," members of an alleged opposition group or movement. Limited to public acts of violence such as assassinations or massacres, but not systematic violent practices like torture or non-violent acts like mass-arrests or crackdowns.
41	<b>Anti-Government Terrorism:</b> Distinct event related to a persistent, directed-violence campaign waged primarily by a non-state group against government authorities or symbols of government authorities (e.g., transportation or other infrastructures).
42	<b>Communal Terrorism:</b> Distinct event related to a persistent, directed-violence campaign waged primarily by a non-state group targeting individual, or "collective individual," members of an alleged oppositional group or movement.
50	<b>Organized Violent Riot:</b> Distinct, continuous, and coordinated action staged by members of a singular political or identity group and directed toward members of a distinct "other" group or government authorities.
51	<b>Spontaneous Violent Riot:</b> Distinct, continuous, and uncoordinated action resulting from an originally non-violent protest and directed toward members of a distinct "other" group or government authorities.
60	<b>Organized Demonstration:</b> Distinct, continuous, and coordinated largely peaceful action directed toward members of a distinct "other" group or government authorities.
61	<b>Pro-Government Demonstration:</b> Distinct, continuous, and largely peaceful action in support of government. May be coordinated or uncoordinated.
62	<b>Spontaneous Demonstration:</b> Distinct, continuous, and uncoordinated largely

		peaceful action directed toward members of a distinct “other” group or government authorities.
70		<b>Other:</b> Reserved for rare events that don’t fit into any of the above categories.
<b>BDAY</b>		<b>Begin Day</b> (2 digit numeric; ordinal) Records the day the event begins.
	1-31	Day event begins
	99	Not known
<b>BMONTH</b>		<b>Begin Month</b> (2 digit numeric; ordinal) Records the month during which the event begins.
	1-12	Month event begins
	99	Not known
<b>BYEAR</b>		<b>Begin Year</b> (4 digit numeric; ordinal) Records the four-digit year during which the event begins. Always present.
<b>EDAY</b>		<b>End Day</b> (2 digit numeric; ordinal) Records the day the event ends.
	1-31	Day event ends
	99	Not known
<b>EMONTH</b>		<b>End Month</b> (2 digit numeric; ordinal) Records the month during which the event ends.
	1-12	Month event ends
	99	Not known
<b>EYEAR</b>		<b>End Year</b> (4 digit numeric; ordinal) Records the four-digit year during which the event ends. Always present.
<b>ACTOR1</b>		<b>Actor #1</b> (text) Records the general political or identity group (i.e., actor) directly involved in the fighting, violence or protest that defines the event.
<b>ACTOR2</b>		<b>Actor #2</b> (text) Records the general political or identity group (i.e., actor) directly involved in the fighting, violence or protest that defines the event.
<b>ACTOR3</b>		<b>Actor #3</b> (text) Records the general political or identity group (i.e., actor) directly involved in the fighting, violence or protest that defines the event.
<b>TARGET1</b>		<b>Target Group #1</b> (text) Records the general political or identity group (i.e., target) directly targeted by the fighting, violence or protest that defines the event.
<b>TARGET2</b>		<b>Target Group #2</b> (text) Records the general political or identity group (i.e., target) directly targeted by the

fighting, violence or protest that defines the event.

**NPART Total Number of Active Participants and People Directly Affected** (2 digit numeric; ordinal)

Records estimated total number of participants and people directly affected by the event according to the following scale. This is the sum of actors (participants of all sides) and targets (including the deaths listed in NDEATH as well as those wounded or otherwise affected):

- 1 less than 10
- 2 10-100
- 3 101-1,000
- 4 1,001-10,000
- 5 10,001-100,000
- 6 100,001-1,000,000
- 7 over 1,000,000
- 11 Unknown, but probably relatively small number (less than 1,000)
- 12 Unknown, but probably large number (1,000-100,000)
- 13 Unknown, but probably very large number (over 100,000)
- 99 Unknown

**NDEATH Total Number of Deaths Reported** (9 digit numeric; cardinal)

Records the best estimate of the number of persons killed as a direct result of the event. Estimates in ranges are separated by a dash in increasing order, e.g. "100-800". If the reported numbers are only a smaller sample of the total, this becomes the lowest possible estimate and is preceded with a "greater than" symbol (">"). If the highest possible estimate is given, such as a combined estimate for multiple cities, precede with a "lower than" symbol ("<").

- 99 Unknown; not reported (note the minus here)

**ELOCAL Event Locality** (text)

Identifies by name the locality in which the event mainly takes place if other/more specific locality than city name is provided.

- 99 Unknown/not provided

**REPORTID1 Report ID #1** (text; ID)

Unique identifier of the Keesing's report used to code the event, and can be linked with the REPORTID in the reports table. This ID is the primary report used for the event. No other event can use the same report as their primary report. Consists of a city ID and a report ID separated by a dash to make it globally unique, i.e. "[City]-[Report]".

**REPORTID2 Report ID #2** (text; ID)

Unique identifier of a second Keesing's report mentioning the event. Note that any number of events can point to the same report.

**REPORTID3 Report ID #3** (text; ID)

Unique identifier of a third Keesing's report mentioning the event. Note that any number of events can point to the same report.

<b>REPORTID4</b>	<b>Report ID #4</b> (text; ID) Unique identifier of a fourth Keesing’s report mentioning the event. Note that any number of events can point to the same report.
<b>REPORTID5</b>	<b>Report ID #5</b> (text; ID) Unique identifier of a fifth Keesing’s report mentioning the event. Note that any number of events can point to the same report.
<b>REPORTID6</b>	<b>Report ID #6</b> (text; ID) Unique identifier of a sixth Keesing’s report mentioning the event. Note that any number of events can point to the same report.
<b>SUMMARY</b>	<b>Description of Event</b> (text) A summary description of the event. Usually direct quotes from the news article, with three stars “***” indicating a break in the original text, and “...” indicating mid-sentence cutoffs. Combines the relevant text from the reports in REPORTID1-6.

### **The Reports Dataset**

The reports dataset contains information about the reports and source text on which events are based, and contains the following variables:

<b>REPORTID</b>	<b>Report ID</b> (text; ID) Unique identifier of an event described in a Keesing’s report. Consists of the city ID to make it globally unique, a number that identifies the report (1, 2, 3, ...), and an identifying letter (A, B, C, ...) for the event within that report, i.e. “[City]-[Report][ReportEvent]”. Can be linked with REPORTID1-6 in the events table.
<b>TITLE</b>	<b>TITLE</b> (text) The original title header of the Keesing’s report. Typically also includes the month and year of the report.
<b>CITY_ID</b>	<b>ID of City</b> (ID) ID of the city for the event which this part of the report describes. This city ID can for instance be used to join with the “cities” table that contains additional metadata about the city.
<b>CITY</b>	<b>Name of City</b> (text) Name of the city for the event which this part of the report describes.
<b>COUNTRY</b>	<b>Name of Country</b> (text) Current official name of the country in which the city is located (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>ISO3</b>	<b>ISO3 Alphanumeric Country Code</b> (3 character text; ID) Alpha-3 country code according to the ISO 3166-1 standard (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.

<b>GWNO</b>	<b>Gleditsch-Ward Country Code</b> (1-3 digit numeric; ID) Numeric country code according to the Gleditsch and Ward (1999) list of independent states. This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>FULLTEXT</b>	<b>All extracts from the report relevant to the event</b> (text) Contains the original text from the sections of the report that describe the event or provide relevant context. Three stars “***” indicates a break in the original text, and “...” indicates mid-sentence cutoffs.
<b>SUMMARY</b>	<b>Summary description of the reported information</b> (text) A summary description of the parts of the report that describe the event. Usually direct quotes from the news article, with three stars “***” indicating a break in the original text, and “...” indicating mid-sentence cutoffs.

### The Cities Dataset

The cities dataset contains additional information and metadata for each city, such as the capital city status or link identifiers to match with other datasets:

<b>CITY_ID</b>	<b>ID of City</b> (ID) ID of the city. This city ID can for instance be used to join with the main “events” table.
<b>CITY</b>	<b>Name of City</b> (text) Name of the city.
<b>CITY_ALT</b>	<b>Alternate Names of City</b> (text) Alternate names of the city, including any historical names that have been in use during the time period covered by the dataset (1960-2014).
<b>COUNTRY</b>	<b>Name of Country</b> (text) Current official name of the country in which the city is located (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>ISO3</b>	<b>ISO3 Alphanumeric Country Code</b> (3 character text; ID) Alpha-3 country code according to the ISO 3166-1 standard (as of June 2017). This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>GWNO</b>	<b>Gleditsch-Ward Country Code</b> (1-3 digit numeric; ID) Numeric country code according to the Gleditsch and Ward (1999) list of independent states. This is based on the situation at the time that the data is released and does <u>not</u> change to reflect changing ownership over time.
<b>REGION</b>	<b>Name of Region</b> (text) Name of the continent/geographical region to which the city belongs.  Asia

Europe  
Latin America  
Middle East and North Africa  
North America  
Oceania  
Sub-Saharan Africa

<b>LONG</b>	<b>Longitude</b> (decimal) Longitude degree of the city's location.
<b>LAT</b>	<b>Latitude</b> (decimal) Latitude degree of the city's location.
<b>CAPITAL</b>	<b>Capital</b> (text) Indicates if the city is or has been a capital during the time period 1960-2014.  Yes No
<b>CAPSTART</b>	<b>Capital start date</b> (date) If the city became the capital city between 1960 and 2014, indicates the date on which the city became the capital (YYYY-MM-DD).
<b>CAPEND</b>	<b>Capital end date</b> (date) If the city ceased being the capital city between 1960 and 2014, indicates the date on which the city ceased to be the capital (YYYY-MM-DD).
<b>ID_UNWUP</b>	<b>ID of City in the UN World Urbanization Prospects dataset</b> (numeric; ID) This lists the city ID used in the 2018 revision of the UN World Urbanization Prospects dataset, which contains yearly population estimates for urban agglomerations. <a href="https://population.un.org/wup/">https://population.un.org/wup/</a>
<b>ID_UNDAT</b>	<b>ID of City on the UN Data platform</b> (date) This lists the city ID used in the city datasets available on the UN Data platform. <a href="http://data.un.org/">http://data.un.org/</a>
<b>NOTE</b>	<b>Note</b> (text) Any special notes regarding the city and its inclusion in the dataset.
<b>NEVENTS</b>	<b>Number of events</b> (text) Total number of events occurring in the city between 1960 and 2014.
<b>DEATHEVENTS</b>	<b>Number of deadly events</b> (text) Total number of events resulting in at least one death occurring in the city between 1960 and 2014.
<b>NODEATHEVENTS</b>	<b>Number of non-deadly events</b> (text) Total number of events resulting in zero deaths occurring in the city between 1960 and 2014.

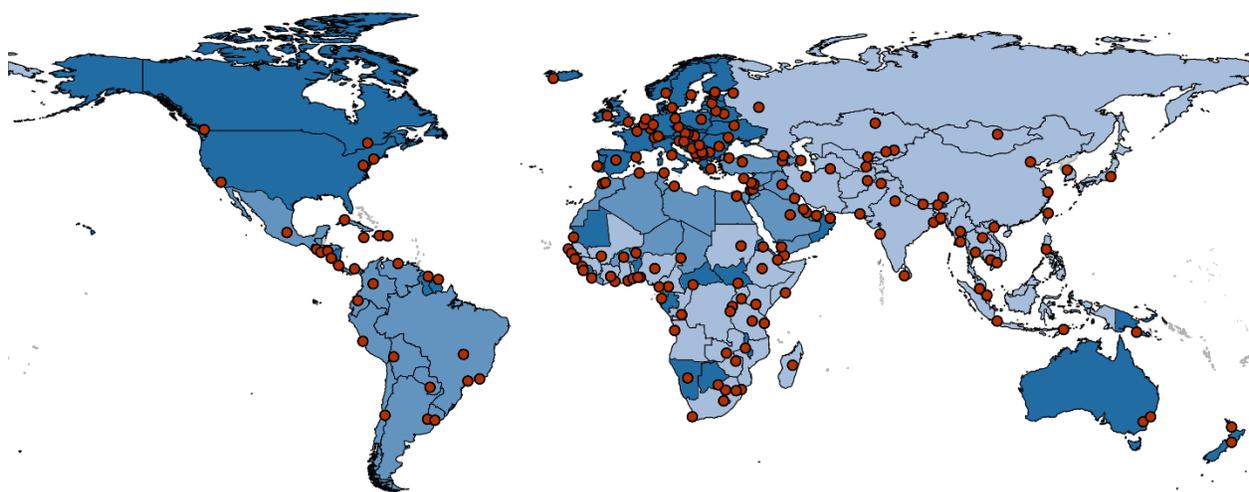


## On the data collection

### The selection of cities

The main selection criteria is that we code the capital city of each country, limited to capital cities of above 100,000 inhabitants in the last year of the dataset (2014), including any city that was once a capital in the period since the start of the dataset (1960).<sup>2</sup> In addition, selected major non-capital cities were also included. The non-capital cities fall into two categories. The first include cases where the capital city is considerably smaller (defined as less than 50% of the population size) than the largest city in the country. An example would be Nigeria, where the capital, Abuja, is considerably smaller than Nigeria's by far largest city, Lagos. For some major countries (China, India, Brazil, USA) with multiple mega-cities we have coded additional, select major cities.<sup>3</sup>

**Figure 1. Geographic coverage of the Urban Social Disorder datasets**



*Note:* The map shows the geographic coverage of the Urban Social Disorder dataset. Red symbols denote the location of the cities included in the complete dataset. Light blue countries reflect the coverage of the first original (USD v. 1.0), medium blue countries reflect the coverage of the second version (USD v. 2.0), whereas dark blue are the countries that have been added in the USD 3.0 update. Light gray countries are not covered, which are mostly smaller island-states.

### Source material

Consistent with the previous versions of the Urban Social Disorder dataset (Urdal and Hoelscher 2012), version 3.0 contains records of unrest events coded from electronic news reports in the online version of the *Keesing's Record of World Events*. Keesing's is a highly regarded and widely used resource of information on political events. As a news aggregator, they publish yearly, event-specific, and monthly summaries of news of political, economic and social

<sup>2</sup> Capitals of smaller island-states or non-sovereign countries like the State of Palestine or Western Sahara were not included. For countries with multiple capitals, only those considered most important were included. The city population criteria were determined using the population of the broader urban agglomeration or region, as reported by the UN and other sources.

<sup>3</sup> Cities included in this final category include Shanghai (China), Mumbai and Kolkata (India), Sao Paulo and Rio de Janeiro (Brazil), New York and Los Angeles (USA), to name a few.

significance from across the world, based on a variety of news sources.

One limitation with Keesing's is that they do not generally extract information directly from local language sources, but depend on events being reported in news media in major languages. This brings us to the more general challenges that most event datasets using news sources have to face. While news reports are likely to cover all events of *major* political significance, there are potentially important biases in event data that need to be carefully addressed, particularly when analyzing cross-sectional (between geographical areas) and time-series trends.

First, strong and autocratic regimes may to some degree succeed in censoring information about events that are considered unwanted, preventing events from entering into news sources such as those used by Keesing's. At the same time, such regimes are probably also relatively successful in preventing undesirable political events from happening. However, distinguishing between bias and regime effect is inherently difficult.

Second, the consumers of news sources such as Keesing's, which are primarily institutions and individuals based in the developed world, may take a stronger interest in certain geographic areas than others, influencing the news providers' priorities of what areas to cover in greater detail. Hence, events happening in countries that are low on the international agenda are possibly less likely to be reported than similar events in countries of high political and economic strategic importance.

Third, it is possible that improvements in communications technology and greater international presence in more places generally means that more events are being reported by international media. Hence, it is difficult to assess whether a general increase in the number of events reflects more events, or just better reporting. Fourth, reporting biases may differ over time for different areas as certain regions wane in economic or geopolitical importance over time, and others wax. While these biases are not easily corrected, students of event data based on news sources need to be alerted to these potential shortcomings.

### **The Search**

The first step of the coding process involved doing an electronic search in the Keesing's database, and was done manually by human coders. All searches included a search term for the city name, including multiple searches for cities formerly known by a different name or that go under multiple names or spellings. We did not make use of any of the filters such as 'Region' or 'Country' in order to facilitate the return of relevant reports that were that were not country-specific. As the data collection effort has been undergoing for some years, the exact nature of this step has undergone several changes over time.

When the data collection first started in 2006, electronic searches were initially conducted in the online version of the database (<http://www.keesings.com/>), and later on a CD-ROM version of the database. The reason for abandoning online searches was that in the spring of 2007, a deficiency occurred in the online search facility after Keesing's changed the online search interface. According to Keesing's, the CD-ROM version should contain exactly the same records up to the end of 2006 as the online database. Our own searches in the CD-ROM version, of cities previously coded on the basis of online searches, did indeed confirm that the content was identical. We switched back to using the online search engine in 2010.

Another difference is that in the first round of data collection, in 2006-2007, we not only searched for the city name, but also added specific search terms to filter only those articles associated with political violence and disorder. The search query was as such:

terroris\* or riot\* or war\* or armed\* or death\* or dead\* or protest\* or communal\* or demonstra\* or insurgen\* or conflict\* or clash\* or violen\* or hostage\* or assassin\* or strike\* or bomb\* or kill\* or suicid\* and [cityname]\* or [cityname]\*

One possible consequence of this is that certain types of events not covered by the keywords might not have been coded, thus underreporting the overall levels of disorder. For all cities and updates coded after 2007, we opted for a broader and more extensive search by searching only for the city name and going through all articles related to that city.

There have also been some minor changes and errors on the Keesing's website that may have affected our search results and data coding. Up until around 2011 we used wildcard characters ('\*') to replace diacritics and other variable parts of city names, but by 2015 the website no longer accepted such requests. Instead we would make separate searches for the various ways of spelling the name. One final issue we came across during 2015-2016 was that for searches returning very large numbers of articles, some articles would be hidden by instead returning a duplicate of a previous article. In the cases where we encountered such duplicates, we would therefore just temporarily limit the time-span of the search in order to uncover the hidden article, and then reset the time-span again. This was a relatively rare occurrence.

### **Screening the Reports**

In the next stage, the coder would go through the search results starting with the oldest reports and work their way forward. Each returned report was screened manually to determine relevance. We used two main criteria for coding an event. First, as this is a city-specific dataset, only events that took place within the strict perimeters and immediate suburbs of the selected cities were included in the dataset. The second main criteria being the event would have to fall under the umbrella of urban social disorder, and fit into one of the event type categories as laid out in the codebook. The term 'urban social disorder' is here understood to encompass social actions directed against a political target (broadly understood as including also different social or identity groups) and/or challenging political authority. Actors may vary considerably in terms of organization, number of participants, use of (non-)violence, and type of political target. Relevant events include public strikes, demonstrations, rioting, terrorism, assassinations, massacres, and military battles (as listed in the "Variables" section). USD events are separable from crime in that they are politically motivated, and although that distinction sometimes is blurred. Crucially, an event is considered relevant, and thus included in the dataset, if the nature of the target can be considered political.<sup>4</sup> The full list of coding criteria and rules can be found in Appendix 1.

For the most part, reports were screened by skimming through the text and doing browser searches for the city name and in some cases for relevant keywords indicating urban disorder. However, from 2016 and out, the switch was made to using an alternative browsing software that

---

<sup>4</sup> An example of a type of event that is not included is clashes between the police and gang members since this is arguably part of an ordinary process to uphold order and security, unless there is evidence to the contrary. For similar reasons, prison riots were not coded, even if the rioters were mainly political prisoners aiming to make a political statement. On the other hand, we code clashes between police and political groups, or politically motivated attacks on prisons from the outside.

interfaced with the website, automatically highlighting an extensive set of predefined keywords in different colors.<sup>5</sup> This may have increased the chances that the coder picked up on relevant events, but all in all probably only to a small degree. Prior to the switch, coders were likely to have been much more diligent at skimming the report text, and the decision to use the software was seen more as a way of increasing the efficiency and speed of the coding process.

### **Coding the Events**

Once an event was found in a report, the relevant text was copied into a Word document for that city, so we could keep track of the original source material. Each report and each event extract was given a unique ID for later reference. Following this, the coder would enter into the Excel sheet information on location, type of event, start- and end date, actor(s) and target(s), reported number of participants, and reported number of deaths. The adoption of the new software in 2016 did not make any significant changes to this workflow, except saving time and making it easier to copy the relevant parts of the reports along with correctly formatted IDs for the reports and events.

The first important factor to determine relevance was whether an event happened within the bounds of the city in question. Events can be described as happening in, on the outskirts, or near the city. And sometimes, reports would mention events taking place in very specific locations such as squares or neighborhoods, or close to certain buildings or monuments. Events were generally included if the location was reported as having happened:

- in the ‘suburbs’ or ‘outskirts’ of the city in question<sup>6</sup>
- near or in a central government building when coding a national capital
- at a location that the coders were able to locate as being within the city (for instance a specific palace or monument)

An event was not coded if the location was reported as being ‘near’ or a certain distance ‘outside’ or ‘north/east/south/west of’ the city.

The second important factor is that the disorder event had to be political, in a broad sense of the word. Events that according to the report appeared to have a purely criminal, non-political motive were excluded, although in some cases it was not possible to verify whether an event was of political or criminal character. A particularly common category was the assassination of public figures, where the assassin was unknown. Generally, we have coded assassinations of government officials and political and military leaders as political events unless specific information is provided in the reports suggesting that the assassination was motivated by personal, economic or other non-political factors.<sup>7</sup>

A particular problem with coding social disorder event data relates to how to separate between discrete events that are parts of complex processes of social disorder. There are no simple criteria to employ, and news reports on complex situations usually contain less information on each discrete event than reports dealing with one discrete event only. For this project, we developed

---

<sup>5</sup> This software, called the “USD Coder”, is available for download from the official USD download page, and its use is described in a separate “Coding Procedures” document.

<sup>6</sup> For instance, in the case of Johannesburg we included the major township of Soweto.

<sup>7</sup> Yet another bias may arise from the likely underreporting of events where the target belongs to more marginal or remote segments of society, particularly in situations of complex disturbances.

three broader criteria to guide the coding of complex events to determine whether it was possible to identify several discrete events. Events were coded as discrete events if (in decreasing order of importance):

- It was possible distinguish between different actors and targets
- Events took place in different locations
- Reported motives for the events were different

As a rule of thumb, at least two of these criteria had to be met before an event was coded as a separate event. However, if a report clearly identified different actors and/or targets, events were coded separately.

Furthermore, the time period within which a series of events happened determined whether events were coded as discrete. If a series of events involving the same actors and targets happened within a short period of time, this would normally be coded as one event (e.g. several bombs against government targets happening within few days). If events involving the same actors and targets are spaced by at least one week (seven days), they would normally be coded as different events. Other times, a report will summarize a collection of events happening over a long time-period, such as a period of continuous demonstrations or battle-clashes lasting for several months. In the absence of information about each individual event these aggregated events are recorded as one long event.

The relatively low share of events with missing information on actor and target reflects to some degree the use of generic terms for these two categories. While individuals and groups are frequently mentioned in reports, the actors behind many events, and sometimes also the targets, are described only in generic terms such as ‘guerilla group’, ‘attackers’, ‘gunmen’, ‘government’, ‘demonstrators’, and ‘rioters’. In these cases we used the generic term as it was used in the original report. When no comment was made as to the identity of the actor or target, and these could not be inferred from the context, as in the case of a bomb attack with no suspected actor or target in a country with several ongoing conflicts, we used the missing code 99.

Finally, it should be noted that the 12 different ‘problem types’ (see Table 1) that all events are categorized according to, are by no means mutually exclusive categories. Demonstrations that are initially peaceful may develop into riots, and the activities of armed opposition groups may in one context (and one time period) be labeled rebellion, in another terrorism. In cases where events escalated as such, we coded it at the highest level of severity. While we have tried to be consistent in the coding of such events, one should be careful in treating the categories as clearly distinguishable phenomena. The provision of the text extracts allow researchers to overrule the current coding of types by going into individual cases carefully.

### **Inter-coder reliability**

The issue of inter-coder reliability is an important one when it comes to event datasets. Not only in terms of different coders finding the same events, but also in terms of how similarly they code the information at hand. We chose to approach this in several ways. First, new coders were trained and assessed by assigning them a previously coded city, and then reviewing the coding with the person who previously coded that city. This level of continuity and dialogue between previous and new coders contributed towards more consistency across coders. Furthermore,

coders would make note of difficult cases that they would then bring up and discuss at regular meetings with the project leader and other coders. As such, difficult coding decisions were typically made as a collective decision and in a way to enhance coding consistency.

## References

Bahgat, Karim, Halvard Buhaug and Henrik Urdal. 2018. "Urban Social Disorder: An Update." PRIO Papers. (<https://www.prio.org/Publications/Publication/?x=10888>).

Banks, Arthur S. and Kenneth A. Wilson. 2020. Cross-National Time-Series Data Archive. Jerusalem: Databanks International.

Beall, Jo, Tom Goodfellow and Dennis Rodgers. 2013. "Cities and Conflict in Fragile States in the Developing World." *Urban Studies* 50(15): 3065-3083.

Chenoweth, Erica and Orion A. Lewis. 2013. "Unpacking Violent Campaigns: Introducing the NAVCO 2.0 dataset." *Journal of Peace Research* 50(3): 415-423.

Glaeser, Edward L. and Bryce Millett Steinberg. 2017. "Transforming Cities: Does Urbanization Promote Democratic Change?" *Regional Studies* 51(1): 58-68.

Gleditsch, Kristian S. & Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25: 393-413.

Goldstein, Joshua S. 2011. *Winning the War on War: The Decline of Armed Conflict Worldwide*. New York: Plume.

Goldstone, Jack A. 2010. "The New Population Bomb: The Four Megatrends that will Change the World." *Foreign Affairs* 89(1): 31-43.

Gollin, Douglas, Remi Jedwab and Dietrich Vollrath. 2016. "Urbanization With and Without Industrialization." *Journal of Economic Growth* 21(1): 35-70.

Moncada, Eduardo. 2013. "The Politics of Urban Violence: Challenges for Development in the Global South." *Studies in Comparative International Development* 48(3): 217-239.

Pettersson, Therese and Magnus Öberg. 2020. "Organized Violence, 1989-2019." *Journal of Peace Research* 57(4): 597-613.

Post, Alison E. 2018. "Cities and Politics in the Developing World." *Annual Review of Political Science* 21: 115-133.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47(5): 651-660.

Rodrik, Dani. 2016. "Premature Deindustrialization." *Journal of Economic Growth* 21(1): 1-33.

Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull and Jennifer Williams. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38(4): 503-511.

Saunders, Doug. 2010. *Arrival City: How the Largest Migration in History is Reshaping Our*

World. London: William Heinemann.

Sundberg, Ralph and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50(4): 523-532.

United Nations. 2018. *World Urbanization Prospects: 2018 Revision*. New York: United Nations.

Weidmann, Nils B., Doreen Kuse, and Kristian Skrede Gleditsch. 2010. *The Geography of the International System: The CShapes Dataset*. *International Interactions* 36 (1).

World Bank. 2016. "Measuring Living Standards within Cities". Washington, DC: World Bank. <https://microdata.worldbank.org/index.php/catalog/3399/download/45899>

## Appendix 1: Specific coding rules

### A. Location

An event should be coded if location is reported as being:

- i. in the ‘suburbs’ or ‘outskirts’ of the city in question
- ii. if it happens ‘near’ or in a central government building when coding a national capital
- iii. if it happens at a location that we are able to locate (for instance a specific palace or monument)

An event should not be coded if location is reported as being:

- iv. a certain distance ‘outside’ or ‘north/east/south/west of’ the city

### B. One event or more?

In some cases it may be difficult to determine whether a report contains one or more events. If there are two or more events happening simultaneously or over a short time interval, we should code them as different if (in decreasing order of importance):

- i. We can distinguish between different actors and targets
- ii. Events take place in different locations
- iii. Reported rationales (motives) for the events are different

A coding judgment should be based on these criteria, and as a rule of thumb at least two of them should be met before we code events as separate. However, if we are able to clearly distinguish between actors/targets we may code them as different events if there is no (contradictory) information provided on location and rationale.

If a series of events involving the same actors and targets are happening within a short period of time, this should be coded as one event (e.g. several bombs against government targets happening within few days). If events involving the same actors are spaced by at least one week (seven days), they should be coded as different events. Other times, a report will summarize a collection of events happening over a long time-period, such as a period of continuous demonstrations or battle-clashes lasting for several months. In the absence of information about each individual event these aggregated events are recorded as one long event.

Unclear cases will be discussed in the project group to enhance coding consistency.

### C. Casualties

If casualties are reported as a range or if different sources report different casualty numbers, the whole range is to be provided. If the report mentions casualties, but not how many, enter ‘>0’. If it is clear that many deaths occurred but only a few are mentioned specifically, those few should be coded as the minimum possible number with ‘>X’. If the report mentions an aggregate casualty number for several cities including the target city, that aggregate should be listed as the maximum possible with an ‘<’ (e.g. if 30 people total were killed in New Delhi, Mumbai, Calcutta, and Chennai, then list <30 for New Delhi’s casualties). If no mentioning of casualties, enter ‘0’ if it is fairly certain none occurred (e.g. peaceful demonstration), enter ‘>0’ if it is almost certain that it did occur (e.g. a major battle/warfare), otherwise enter -99.

- D. If no specific information is provided for actors/target, use the terminology being used in the original report, such as: guerilla group, attackers, gunmen, government, demonstrators, rioters, etc.
- E. All items where no information is provided should be coded 99, including ‘target’ when there is no obvious target (e.g. many pro-government demonstrations). No supplementary sources shall be consulted in attempts to gather more details.
- F. If uncertain of how to code, highlight and add a comment to the report text, which will then be discussed at the next meeting.
- G. Instructions aimed at coder coherence:
- a. Mass events that are not political should not be coded (i.e. funeral attended by several hundred thousand people, audience of 2 million at President Nasser’s resignation speech, large crowd at airport to greet Henry Kissinger)
  - b. For airplane hijackings, code if:
    - i. The violent event took place in the target city (i.e. the plane was blown up on a runway at Cairo airport)
    - ii. The plane was hijacked while leaving the target city’s airport.
    - iii. Do not code if the hijacked airplane simply refueled in the target city, or peacefully landed in the target city.
  - c. Systematic repressory practices like censorship, banning of organizations, seizing of property, mass arrests, police raids, executions, or torture are to be considered outside the scope of USD dataset.
  - d. Coups d’etat should almost always be coded as Armed Attack. Even if the coup is bloodless, the assumption is that the attacking party would have used force if necessary. Sometimes, General Warfare can be used if the target country is in a continual state of war and the coup is related to this warfare.
  - e. Nationwide events should usually be coded for capital cities even if specific action is not reported for the capital city. The assumption is that nationwide events would almost certainly occur in the capital, as it is an important and often largest city.
  - f. Attacks do not have to be directed at people, e.g. covert sabotage of infrastructure, electricity, dams, etc. should be counted anti-government terrorism. However, only if the actual sabotage happened within the city, so that a city experiencing blackout from an attack on an electrical facility outside the city would not be coded. Cyber attacks or hacks should in no case be coded.
  - g. Events that escalate should be coded at their highest level of intensity, so that demonstrations that result in government repression is coded as the latter.

- h. A foreign power using clandestine dirty-tactics like bombing or assassination in a country is not to be understood as pro-government activity, since “government” refers to the ruling power in the country being coded.
- i. Demonstrators who occupy/barricade some building is coded as organized riot, even if not involving any actual violence, since it implies the threat of violence.