Crowd Counting Consortium  
Coding Guidelines 3.0

January 2025–

# Introduction

The [Crowd Counting Consortium](https://sites.google.com/view/crowdcountingconsortium/home) is a longtime collaboration between Profs. Erica Chenoweth (Harvard University) and Jeremy Pressman (University of Connecticut). Since January 2017, CCC has collected data on political protest events in the United States, including but not limited to rallies, marches, demonstrations, and acts of civil disobedience.

More specifically, CCC follows the [Dynamics of Collective Action](https://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/) project in defining its events of interest as “any type of activity that...is carried out with the explicit purpose of articulating a grievance against a [political] target, or expressing support of a [political] target.” That is a wide net to cast, so we use a few additional criteria to narrow the scope further.

* Following Soule and Davenport (2009: 8), we limit our collection to events that are **open to the public and free of charge**.
* We also limit our work to events that are **nonviolent** in the sense that they are not primarily organized to cause direct harm to any persons. Per Chenoweth (2021: 35–37), in the study of contentious collective action, “nonviolent” does not necessarily mean free of conflict, coercion, or property damage. What we intend to exclude with this criterion are military or paramilitary actions in which attacking or otherwise harming other people is the focal tactic.
* Finally, we only encode data on events that **occur in the United States**, including the U.S. territories of Puerto Rico, Guam, and the U.S. Virgin Islands. The project has occasionally encoded data on events in other countries, but its core mission is limited to making structured data about contentious political events in the United States, and events outside that country are not included in the [compiled and augmented version](https://github.com/nonviolent-action-lab/crowd-counting-consortium) of the CCC data.

Together, these criteria include all of the aforementioned forms of action while excluding things like paid fundraisers, workshops, trainings, teach-ins, panel discussions, distributions of mutual aid, many commemorative or celebratory festivals or parades (but not all—see the [FAQ](#_1tl2zonkhwxw) for more on this), concerts, cookouts, court support, block parties, town hall meetings, political campaign events that are closed to the general public, and signature-collecting or canvassing activities. Claims voiced during comment sessions at meetings of government bodies (e.g., legislative sessions, city council meetings, school board meetings, court proceedings) are also excluded unless the people making those claims deliberately and successfully disrupted or derailed the larger meeting. Rallies associated with campaigns for public office are included as long as they are open to the public free of charge (i.e., not paid fundraisers). Vigils, memorials, and commemorative gatherings are only included if participants also articulate grievances around broader political issues (e.g., community violence, police brutality, domestic violence, or gun control).

We also choose to define “gathering” broadly and include online events in our collection. Virtual gatherings have become an increasingly common and important form of collective action, and that trend only accelerated in 2020–2021 as a result of the COVID-19 pandemic. In light of this fact, we encode data about virtual versions of events that also meet the criteria enumerated above.

Finally, there is no minimum crowd size for inclusion in the CCC dataset. We record events involving just one or a few participants, as long as they satisfy the criteria enumerated above.

# Sources

CCC uses a broad and always-evolving set of sources to find public information about relevant events, including online print and broadcast media sources, social media, and public submissions via an [online form](https://sites.google.com/view/crowdcountingconsortium/submit-a-record). On a day-to-day basis, the research process proceeds along four main channels.

* **Automated web crawl of online print and broadcast media sources**. From 2017 through early 2021, the [Count Love](https://countlove.org/) project shared the output of its web crawler with CCC via daily emails. Since early 2021, CCC has run its own instance of that crawler using Count Love’s open-source code. The daily output from the crawler is now stored in an Amazon Web Services bucket and written to a Google Sheet that refreshes daily.
* **Regular review of selected social media accounts**. CCC uses a constantly evolving set of social-media platforms, and lists of accounts on those platforms, to surface public information about protest events, many of which are covered incompletely or not at all by commercial news media. At the start of 2025, CCC was primarily using Instagram, Telegram, and, to a lesser extent, Twitter/X, with a focus on open accounts maintained on those platforms by organizers, activists, and independent journalists.
* **Google News searches on relevant keywords**. On an ad hoc basis, CCC coders also use Google News searches on relevant keywords to uncover stories the web crawler may have missed.
* **Irregular review of organization websites**. On an ad hoc basis, CCC collects information directly from the websites of organizations hosting large, multi-location protests or maintaining public online calendars or lists of past or upcoming events.
* **Public submissions**. CCC also solicits submissions and corrections from the public via an [online form](https://sites.google.com/view/crowdcountingconsortium/submit-a-record). Responses are reviewed and encoded on an occasional basis. Events identified through this process comprise a tiny fraction of our data set, and we only create records from online submissions when we are able to confirm that the reported event actually took place.

# Data-Making Process

CCCC coders record information about reported events in [Google Sheets](https://sites.google.com/view/crowdcountingconsortium/view-download-the-data) that cover individual calendar months (e.g., January 2025).

For CCC, protest events, not news reports, are the units of observation. This means that, in the dataset, each row represents a single event, defined by the gathering of a particular set of people to articulate a specific set of grievances in a particular place on a certain date. When multiple news reports or social-media posts describe the same event, the relevant information is amalgamated into a single record (row). Rows are ordered by date, then alphabetically by locality (city or town) and then state (postal code). When entering data, coders should assume that each record (row) needs to stand on its own and therefore should never reference other rows (e.g., “see above”) in their entries.

Collection is an ongoing endeavor, and coders regularly make additions or revisions to existing records as new information surfaces. This process places twin burdens on coders.

* Coders are asked to take care to avoid duplicating events that are already represented in the data. This means carefully checking existing records for events that occurred in the same locality on the same date before starting a new one.
* When multiple reports exist for the same event, coders have to decide how to synthesize that information. We generally resolve that tension by recording the relevant bits from all of the accounts, using semicolons within fields to make it easy for other researchers to distinguish those bits and make their own choices about how to synthesize them or otherwise handle that uncertainty.
* CCC’s instance of the Count Love web crawler sometimes surfaces multiple news stories about the same event. In these cases, coders are instructed only to add a link to the relevant record if they add information to the record from that particular story. In other words, we do not necessarily store all links we see about a particular event, so the count of links associated with a record is not a reliable indicator of that event’s supposed newsworthiness.

# Data Fields

For each event, coders should fill in as many of the following fields as possible, using information from all relevant sources. When in doubt about what to record from a particular source or how to record it, please consult with the project’s supervisor(s).

## date

Record the date on which the event occurred, in YYYY-MM-DD format.

In the Google Sheets, this column uses data validation to reduce the risk of invalid responses. If you make an invalid entry, the sheet will notify you. When this happens, please take a moment to correct the entry before moving on. NOTE: data validation does not correct for typos in elements of the date (e.g., 2201-07-21 instead of 2021-07-21, or 2020-07-21 for 2021-07-21), so please take care when entering this element.

Some events span more than one day. For events that recur daily or that span multiple days for other reasons (e.g., many strikes and occupations), create a separate record for each day on which the event occurred or continued. The exception to this rule are events that begin one day and continue into the early-morning hours of the next day before wrapping up. In these cases, just create a single record and associate it with the date on which the event began.

In situations where the date of the event is not specified, try to use other contextual information to suss it out. For example, if you know that the protest targeted a city council meeting, you can search for the city council’s meeting schedule and see if that clears it up.

If the timing is only given vaguely (e.g., “last week”, “over the weekend”), make an educated guess about which day and give some indication in the *notes* field that you did that (e.g., “post only said ‘last weekend’, so arbitrarily assigning to Saturday”).

If none of the above apply, date the event to the day before the publication of the news article (for news media) or post (for social media) about it and make a note of that in the *notes* field (e.g., “date not specified, so dating to day before post”).

## locality

Record the name of the city, town, or township in which the event occurred.

* Take time to check for correct spelling. When in doubt, confirm that the spelling you use is recognized by Google Maps.
* Avoid abbreviations. For example, enter “Saint Louis” rather than “St. Louis”.
* In cases where a single event began in one locality and traveled to one or more others (e.g., a march or caravan), do not create a separate record for each locality. Instead, choose one locality—usually the start or end point—and enter it here, then enter the names of the other towns in the ‘location’ field (see below).
* For events occurring in any of the five boroughs of New York City (Manhattan, Brooklyn, the Bronx, Queens, and Staten Island), enter “New York” as the locality, not the borough name. The borough is captured in the ‘location’ field (see below).
* For named neighborhoods of other cities, use the city name as the locality and, if it is given, capture the neighborhood name in the ‘location’ field. For example, an event described as occurring in “Roxbury’s Nubian Square” would get “Boston” as its locality, and the location might be recorded as “Nubian Square, Roxbury”. When in doubt about whether something is a neighborhood or a city, use Google Maps to check.
* In rare cases where the locality is unclear (e.g., “on the Merrimack River” with no other information about event location), leave this field blank.
* For online events, see the guidance specific to that type of event below.

## state

Record the two-character postal code for the state or territory in which the locality named above is located.

In the Google Sheets, this column uses data validation to ensure that only valid responses may be entered. If you enter an invalid character pair, the sheet will show you a notification, and nothing will be recorded. Please try again.

If you are not certain of the correct postal code, please look it up (see the USPS column in [this table](https://en.wikipedia.org/wiki/List_of_U.S._state_and_territory_abbreviations)). Some of them are easy to confuse (e.g., AR/AK for Arkansas/Alaska, MA/ME for Massachusetts/Maine).

## location

Record the specific location or locations at which the event occurred within the aforenamed locality.

* If a **landmark** is specified (e.g., “Massachusetts State House”, “Hermann Park”, “Hilton Midtown Hotel”), record its name, ideally in the format recognized or returned by Google Maps (e.g., “on the capitol steps” in an article about a protest in Boise, Idaho, should be recorded as “Idaho State Capitol”, or “outside city hall” in Akron should be recorded as “Akron City Hall”).
* If the protest occurred at a **school**, record only the name of the school (e.g., “Harvard University”) and *not* the specific location on that school’s campus (e.g., Widener Library). Feel free to capture that more specific information in the *partipant\_measures* field (e.g., “study-in at Widener Library”), however.
* If an **address** is given and there is no associated landmark at that spot, record the address.
* If there is no clear landmark or address but an **intersection** is given (e.g., “Hollywood and Vine”), record it, ideally in the format recognized or returned by Google Maps (e.g., “Hollywood Blvd and Vine St” or “Hollywood Blvd & Vine St”, either is fine).
* For **events occurring in New York City**, always include the name of the borough as well as the landmark or address within it (e.g., “633 3rd St, Manhattan” or “Grand Army Plaza, Brooklyn”). When in doubt about the borough, please look it up.
* **When events move from location to location**, record as many as you can recognize, separated by semicolons. For example, a march that traveled from Foley Square in Manhattan over the Brooklyn Bridge to Barclays Center in Brooklyn would get “Foley Square, Manhattan; Brooklyn Bridge; Barclays Center, Brooklyn” as its location. Meanwhile, an event that started in Silver Spring, MD, and traveled to the White House in Washington, DC, might get “Washington” as its locality and then “Silver Spring, MD; The White House” as its location.
* For [**online events**](#_m72z7od9q1h), enter “online” in this field and nothing else.

In any case where the location is not clear, you are encouraged to use Google Maps to help clarify or confirm the location and how to record it. As often as possible, we want the entries in this field to be machine readable, meaning they could be fed as is into Google Maps and we could expect to get a specific set of geocoordinates back. So, when in doubt, record the location as Google Maps gives it

## title

Record the name of the event, if one is given in materials announcing the event in advance.

On flyers and Facebook event pages, the title is usually easy to distinguish. When it isn’t and several phrasings appear, make an educated guess about which part is the title, and err on the side of including too much information rather than too little. With Facebook events and many other calendar listings, the title can often be copied and pasted. To avoid carrying undesired formatting into the Google Sheet, you can copy the text, then a) paste it into a simple text editor (e.g., Notepad or TextEdit), copy it again, and then paste it into the Sheet; or b) use the Paste Special -> Values Only command (ctrl+shift+V) when pasting it into Google Sheets.

## event\_type

Identify the form or forms the event took from the list below, using semicolons to separate entries when there is more than one type (e.g., “rally; march”).

NOTE: In cases where the action includes a specific subtype of possible interest (including but not limited to the ones enumerated under several types below), please include that subtype in the *participant\_measures* field.

* **Demonstration**. A crowd gathering to demonstrate their support for a set of political demands or claims to an external audience, typically with things like signs and banners and flags displayed to passers-by in a public space. This is probably the most common type of action we see. Some common subtypes include:
  + picket
  + sit-in (assuming it does not qualify as civil disobedience; see below)
  + die-in (assuming it does not qualify as civil disobedience; see below)
  + vigil
  + banner drop
  + art installation
  + mobile billboard (e.g., parked truck with LED signs)
* **Rally**. A crowd gathering in public to demonstrate and reinforce their mutual support for a set of political demands or claims, typically by making speeches directed at each other. These include but are not limited to campaign rallies, and these events sometimes include or support press conferences. Festivals that meet our criteria (e.g., LGBTQ+ pride festivals) should also be labeled as rallies.
* **March**. A crowd traveling on a public route in a coordinated fashion as a means to express their support for a set of political demands or claims, either to each other or to bystanders or both. Some common subtypes include:
  + walk
  + run
  + parade
  + procession/posada
  + car caravan
  + bicycle ride
  + boat parade or flotilla
* **Protest**. A crowd gathering in public to express disagreement with, or disapproval or anger or frustration toward, a specific individual or organization that is at or near the crowd’s gathering point (e.g., a politician giving a speech, a corporate headquarters, a bank branch, a construction site, a city hall), or in negative reaction to a recent or current event (e.g., the killing of George Floyd, the reversal of Roe v. Wade).
* **Civil Disobedience**. For purposes of this project, we understand civil disobedience as deliberate, public, nonviolent law-breaking acts intended to express a political point. In recent years, most acts of civil disobedience have involved demonstrators engaging in some kind of disruptive action, deliberately ignoring police instructions to stop or disperse, and then cooperating with police during the arrests that follow. Common forms involve occupations of space that impede others from doing things to which the protesters object—e.g., occupying a lawmaker’s office, forming a line to block the entrance to a corporate office building, or blocking a public roadway.  
    
  For purposes of this project, we consider many acts that anarchists would label as “direct action” to be forms of civil disobedience as well, as long as they do not directly threaten to harm other humans (cf. Berglund 2023). Common forms include disruptions at work sites (e.g., protesters locking themselves to construction equipment), vandalism (e.g., graffiti on a targeted office building or bank branch), sabotage (e.g., concrete powder poured into toilets in a targeted hotel), and looting when it is clearly connected to an act of political protest (e.g., during a march in response to a police killing).
* **Counter-protest**. See the [dedicated section below](#_5ysd5a9anpi6) for more on what these are.
* **Other**. If you encounter an action that does not fit into any of the above categories, label it “other” in this field and then include relevant description or details in the *participant\_measures* field.

## organizations

Record the names of all organizations mentioned in your sources as having participated in the event in any way, separated by semicolons if there’s more than one.

Organizational participation can take many forms, from organizing and leading the event to sponsoring or co-sponsoring it to providing one or more speakers for it to just showing up to the event as a recognizable presence (e.g., with a banner or in a uniform). Err on the side of inclusion. If a reporter quotes a protester and associates them with a named organization, include that organization’s name. The goal is to facilitate research into connections and interactions between organizations, so we want to capture as much of this information as possible.

When no organizations are mentioned as organizing or participating in the event, leave this field blank.

## claims\_summary

Summarize the main political demands or grievances expressed by event leaders and organizers.

Think of this field as a space to describe what the protest is about, or *why* participants are protesting. Try to paint as vivid and complete a picture as you can in as efficient a manner as possible. This helps casual users of the data understand what the event was about, and it gives more advanced users of the data richer material to draw connections or contrasts across events. Recognizing that this is a bit of an art form, here are some general pointers.

* **Start by summarizing the participants’ stated or implied purpose of the event in one or a few short phrases** that begin with “for” or “against” or, where relevant, “in remembrance of”, “in celebration of”, “in commemoration of”, or “in solidarity with”. For example, instead of writing “Trump rally”, write something like “for Donald Trump for U.S. president”. Or, instead of writing “stop Line 3”, write something like “against Line 3 oil pipeline”.
* **When crafting this summary, assume that none of the other fields are visible to the person reading i**t. So, for example, if the event involves teachers picketing for higher pay at Long Beach School District, don’t write something like “for higher pay” and assume that data users could infer the “teachers” and “Long Beach High School” parts from the ‘participants’ and ‘location’ fields, respectively. Instead, write something like “for higher pay for public school teachers in Long Beach Unified School District”.
* **When relevant, make connections to broader issues explicit in your summary**. For example, protests against the Line 3 project often involve concerns about water pollution and the sovereignty of Indigenous peoples whose territories the project crosses. So, instead of just writing “against Line 3 oil pipeline” for events where that’s the case, write something like “against Line 3 oil pipeline, for environmental protection, for Indigenous sovereignty” if those other issues are evident in protesters’ claims as well.
* CCC applies regular expressions to the claims summaries you write to try to attach tags representing one or more of a few dozen major political issues (e.g., racism, environment, education, LGBTQ+ rights) with each event. Those regular expressions rely on keywords or phrases. **If there are specific issues you think an event should be associated with, be sure to include at least one keyword or key phrase from the relevant regular expression(s) in your summary**. For example, if you’re describing a protest against mask mandates in schools, you would want to write something like “against coronavirus mask mandates in schools'' instead of the shorter “against mask mandates” so the keyword “coronavirus” would trigger the ‘covid’ tag and the term “schools” would trigger the ‘education’ tag. For the actual regular expressions, see [this R script](https://github.com/nonviolent-action-lab/crowd-counting-consortium/blob/master/data_compilation/ccc_issue_regex_list.R).

## claims\_verbatim

Record verbatim as many unique claims made by organizers and participants during the event as you can, including phrases from signs, banners, and chants evident in photos or videos. **Separate claims phrases with semicolons**.

* By “unique”, we mean that you only need to record a single instance of a phrase you see or hear more than once at the same event. So, for example, if dozens of people are carrying signs that say “Black lives matter”, you should still only record that particular phrase once in this field. If, however, one person has a “Black lives matter” sign and another has a “Black trans lives matter” sign, you would record both. In cases where the exact same phrase is both seen on signs and heard in chants, only capture it once.
* When recording **chants**, end the phrase with an exclamation point (e.g., “stop fossil fuels!”). To avoid confusing slogans from placards and banners with chants, do not include exclamation points at the end of verbatim captures of those (e.g., a sign reading “climate action now!” should be recorded as “climate action now”).
* In cases where **images** are an important part of a sign or banner, put a word or phrase describing the image in brackets (e.g., “never again [wire hanger]” or “we need to talk about the elephant in the womb [GOP]” at a rally for abortion access).
* When protesters carry **flags**, list them in the summary as well, using the structure “\_\_\_\_ flag”, where the blank is filled with the name or descriptor of the flag, e.g., “Gadsden flag”, “Progress Pride flag”, “An Appeal to Heaven flag”, or “Trump flag”. See [here](https://docs.google.com/spreadsheets/d/1PWYLtJawr72oQA1UjK64SBpggE1_139v_uqP8tOWWKw/edit?usp=sharing) for a project flag glossary.
* For **national flags**, use the name of the country rather than its adjectival form (e.g., write “USA flag” instead of “American flag”, or “Mexico flag” rather than “Mexican flag”).
* Where verbatim claims include **line breaks**, use “ / ” to represent the break (e.g., “arms embargo now / free Palestine”). And please note the spaces on either side of the slash.
* When signs, banners, or chants use **languages other than English**, record the verbatim version in the original language if you are able to do so and the language uses the Latin alphabet. If the language uses something other than the Latin alphabet and you are able to understand that language, you may capture either a) a transliterated version in the original language or b) an English translation. If you do either one of those, please start the captured phrase with a bracketed clarification of how it was transformed. For example, if a sign reads “мир миру” and you read Russian, you could write either “[transliterated Russian] mir miru” or “[translated Russian] peace to the world”. And if you don’t understand that language or aren’t sure what’s written or said, just ignore it.

## valence

Use this field to capture the left-right political orientation of the participants’ claims as follows.

2 = right wing

1 = left wing

0 = other/neither

This is a judgment call, and we recognize that it will often be unclear. Here are some general pointers.

* Events organized by or in support of Republican Party organizations or candidates generally warrant a 2, while events organized by or in support of Democratic Party candidates or organizations generally warrant a 1.
* Events expressing claims traditionally associated with one or the other major U.S. political party should be coded accordingly. For example, the GOP has long advocated for restricting abortion rights, so “pro-life” events warrant a 2. Ditto for events advocating for Second Amendment rights. By contrast, events advocating for abortion rights or for gun control would generally get a 1 here, as those positions are traditionally associated with the Democratic Party.
* Events focused on hyper-local issues (e.g., opposition to a new residential development in the neighborhood) or essentially non-partisan issues (e.g., domestic violence, suicide awareness, drug addiction, animal cruelty) should generally be coded as 0.
* Events focused on foreign affairs generally warrant a 0 as well. The rare exceptions are cases where claims about foreign affairs are mixed with explicitly partisan claims about domestic issues (e.g., “stand with Israel, vote for Trump” or “solidarity with Palestine, against racism, defund police“).
* For claims that are not associated with either major political party but are broad and politicized, you will need to use your judgment. For example, while the core claims of many actions associated with the 2020 George Floyd uprising were not endorsed by leaders of the Democratic Party, we chose to give them a valence of 1 because they were clearly left wing and were firmly rejected by President Trump and other Republican leaders.
* When using your judgment about fuzzier cases, aim to be consistent with past practice on similar issues. When in doubt, consult project leaders.

## conf

This field is used to identify events that we know have actually occurred because we have seen reports from or about them as or after they happened. The point is to distinguish these confirmed events from events about which we only have evidence that they were planned or announced in advance.

Enter a 1 in this field when you see information confirming an event’s occurrence and the field was previously blank. If another coder has already entered a 1 here, leave it alone.

## macroevent

When relevant, this field is used to record a unique ID that links a counter-protest (or counter-protests) to the event it (or they) target. (For more on what constitutes a counter-protest, see [Counter-protests](#_5ysd5a9anpi6) below.) The goal is to facilitate analysis of protester/counter-protester interactions by making it easy to filter and group events where these interactions occurred.

Macroevent IDs are composed of three parts: a date, a location, and some brief descriptor of the central event’s content or theme. For example a Black Lives Matter protest and counter-protest in Boston on June 10, 2021, might be given matching macroevent IDs that read “20210610-boston-blm”.

The length of the ID doesn’t matter much, and the exact construction of the IDs (including typos) is not really important, either. The main concerns are the following.

* IDs must be unique to a given protest/counter-protest pair or set (hence the use of the date-location-descriptor construction, which virtually ensures this)
* IDs must match exactly within protest/counter-protest pairs or sets (best practice is to use copy/paste to ensure this)
* IDs should not contain any spaces or special punctuation marks, just numbers and letters and hyphens (e.g., “20211001-nyc-abortion”, not “20211001 New York abortion”)

## participants

Record words or phrases describing the participants in the event, separated by semicolons (e.g., “employees” or “parents; teachers; students”).

* As with the ‘organizations’ field, err on the side of inclusion. The goal is to capture as much information as we can about the kinds of people who participated, as distinct from any organizations they represent or to which they belong.
* When sources do not offer descriptions of participants but you can glean some from images or videos (e.g., “children”, “workers”), feel free to record those.
* When elected officials of any kind participate in the action in any way, please include “elected officials” in this field.
* When no descriptions are given or the only ones given are uninformative (e.g., “protesters”, “activists”) and nothing stands out from images or video of the action, leave this field blank.

## targets

Use this field to indicate the institutions or individuals at whom protesters’ demands or other claims are directed. If more than one, separate with semicolons.

* In many cases, these will be specific government bodies (e.g., “city government”, “state government”), specific elected officials (e.g., “President Donald Trump”, “Mayor Eric Adams”, “U.S. Rep. Judy Chu”), or specific organizations or corporations (e.g., “Chevron”, “Citibank”, “Planned Parenthood”).
* When the target is vague (e.g., the general public, passers-by) or unclear, leave this field blank.

## notables

Use this field to capture the names of notable individuals who participated in the protest action, using semicolons to separate names when there are more than one. If the person has an official title that’s relevant to their notability, please include that title (e.g., “U.S. Rep. Cori Bush” rather than “Cori Bush”). Err on the side of inclusion.

## size\_text

Record the bit(s) of text describing the number of people who participated in the event.

Note that we do not record “expected” or “predicted” turnout here or in the ‘size\_low’ and ‘size\_high’ fields that follow. If the event has not yet happened, leave this cell blank. Instead, we wait to see how many people actually show up according to public reports.

* If an exact number is reported (e.g., “37”, “two”), record it.
* When the crowd size is described in vague words—e.g., "a few dozen" or "hundreds"—record those words here. Likewise for combinations of words and numbers (e.g., “more than 20” or “nearly 1,000”). If the source uses words that can be converted to numbers in these phrases, please do the conversion (e.g., “more than a hundred” should be recorded as “more than 100”, but “hundreds” should be recorded as a word).
* If a range is given, record it using “xx-xx” format (e.g., “30-40” and “30 to 40” should both be recorded here as 30-40).
* When different sources provide different estimates, include them all, separating them with semicolons (e.g., “dozens; a few dozen; more than 50”). The order does not matter.
* When a single source gives multiple estimates (e.g., “a few dozen to start...crowd grew to more than 100”), use only the largest one.
* When no sources describe the crowd size but there are photos or videos available, use that imagery to try to estimate the crowd size. If you do this, record either “count pic” or “count video” in this field, depending on which one you (primarily) used. Do not try to estimate crowd size from a video or a picture of a very large event (roughly thousands or higher).
* When no sources describe the crowd size and pictures and video are unavailable or uninformative but the event has a Facebook Event page, check that page after the event for a count of people who “went” to it. If there is one, record it here as FB: Number. For example, if the Facebook event page says 37 people went, put FB: 37 in this column.
* If a coder witnesses an event or receives information about event size from a credible source and no other credible information about crowd size is available, record “eyewitness” in this field.

## size\_low

Record the lowest reported estimate of the crowd’s size, using the following rules to convert words into numbers. If only a single, specific number is reported, record it here and in ‘size\_high’. Do not include commas in the numbers (so, 1000 rather than 1,000).

* With fuzzy plural terms, assume the lowest possible multiple. So, for example, “dozens” becomes 24, “hundreds” becomes 200, and “thousands” becomes 2000.
* Interpret “several” and “a few” as 3. So, for example, “a few hundred” and “several hundred” would both become 300.
* Interpret “handful” as 5.
* If a range is reported, record the low end of it here.
* If you must estimate the crowd size from photos or video, you are encouraged to recognize that your estimate will be imprecise and guess a range rather than a single number. If you do this, record the low end of that range here.
* When an exact count is given or a Facebook Event page count is used, record it here.

## size\_high

Record the highest reported estimate of the crowd’s size, using the following rules to convert words into numbers. If only a single, specific number is reported, record it here and in ‘size\_low’. Do not include commas in the numbers (so, 1000 rather than 1,000).

* With fuzzy plural terms, assume the lowest possible multiple. So, for example, “dozens” becomes 24, “hundreds” becomes 200, and “thousands” becomes 2000.
* Interpret “several” and “a few” as 3. So, for example, “a few hundred” and “several hundred” would both become 300.
* Interpret “handful” as 5.
* If a range is reported, record the high end of it here.
* If you must estimate the crowd size from photos or video, you are encouraged to recognize that your estimate will be imprecise and guess a range rather than a single number. If you do this, record the high end of that range here.
* When an exact count is given or a Facebook Event page count is used, record it here.

## participant\_measures

Briefly describe what the protesters planned to do or, if information is available after the fact, what they did (e.g., “stood at intersection with signs and banners”).

This field can be especially useful for describing the nature of civil disobedience or direct actions in cases where those were involved (e.g., “roadblock”, “sleeping dragon”, “protesters locked themselves to construction equipment”, “vandalized building facade”), or for describing interactions between protesters and police (e.g., “heckled police; threw water bottles”) or protesters and counter-protesters (e.g., “verbal confrontation and scuffles with counter-protesters”, “physical fighting with protesters”).

## police\_measures

Indicate if police were on the scene at any point during the action, and if so, briefly describe what they did. If no information is provided about police presence or actions, leave this field blank.

In many cases, this field will be blank or something as simple as “on scene” or “followed marchers”. In cases where police took more aggressive action, you might see a summary like “riot gear; formed skirmish line; declared unlawful assembly”. In cases where police used additional crowd-control measures, try to be as specific as possible (e.g., “sprayed pepper spray; tossed flash-bangs; carried less-lethal munitions”). If information is available about the size or scope of the law enforcement presence, it is helpful to capture that as well (e.g., “dozens of officers” or “National Guard and state troopers deployed”).

## participant\_injuries

Record the number of participants in the action who were reportedly injured during the action.

* If a single number is reported, enter the number (e.g., 3).
* If a phrase is used to describe an ambiguous number of injuries (e.g., “more than 3”), record the phrase.
* If protesters appear to have been injured but neither of the above applies, enter “unspecified”.
* If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., an organizer’s after-action report). If none appears authoritative or all accounts are ambiguous, record “unspecified”.
* If no participants were reportedly injured, leave blank.

Note that injuries of counter-protesters should not be recorded or included here. Instead, they should be recorded in the row representing the counter-protest.

## police\_injuries

Record the number of police or law enforcement officers on the scene who were reportedly injured during the action.

* If a single number is reported, enter the number (e.g., 3).
* If a phrase is used to describe an ambiguous number of injuries (e.g., “more than 3”), record the phrase.
* If officers were reportedly injured and neither of the above applies, enter “unspecified”.
* If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., the latest report from a law enforcement agency). If none appears authoritative, record “unspecified”.
* If no officers were reportedly injured, leave blank.

## arrests

Record the number of participants in the action who were reportedly arrested or detained during the action.

* If a single number is reported, enter the number (e.g., 3).
* If a phrase is used to describe an ambiguous number of arrests (e.g., “more than 3”), record the phrase.
* If protesters appear to have been arrested but neither of the above applies, enter “unspecified”.
* If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., a law enforcement agency’s after-action report vs. an eyewitness’s statement). If none appears particularly authoritative or all accounts are ambiguous, record “unspecified”.
* If no participants were reportedly arrested, leave blank.

Note that a citation or even a detention alone is not enough to count toward this tally. To count here, participants must be arrested. The distinction between detention and arrest [is not always clear](https://www.nolo.com/legal-encyclopedia/arrest-vs-detention-how-tell-whether-you-ve-been-arrested-simply-detained.html), however, so err on the side of inclusion.

Note also that arrests of counter-protesters should not be counted here. Instead, they should be recorded in the row representing the counter-protest. If it is unclear from reports whether arrested individuals were protesters or counter-protesters or both, put “unspecified” in the ‘arrests’ field for both events.

## property\_damage

In words or short phrases, summarize any property damage caused by participants in the action (e.g., “graffiti; broken windows; dumpster fire”). If no property damage is reported, leave blank.

## police\_deaths

Record the number of police or law enforcement officers who reportedly died during the action.

* If a single number is reported, enter the number (e.g., 3).
* If a phrase is used to describe an ambiguous number of deaths (e.g., “more than 3”), record the phrase.
* If officers reportedly died and neither of the above applies, enter “unspecified”.
* If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., the latest report from a law enforcement agency). If none appears authoritative, record “unspecified”.
* If no officers reportedly died, leave blank.

## participant\_deaths

Record the number of participants in the action who reportedly died.

* If a single number is reported, enter the number (e.g., 3).
* If a phrase is used to describe an ambiguous number of deaths (e.g., “more than 3”), record the phrase.
* If protesters reportedly died but neither of the above applies, enter “unspecified”.
* If multiple sources provide conflicting accounts, use your judgment about whether one is more authoritative (e.g., an organizer’s after-action report). If none appears authoritative or all accounts are ambiguous, record “unspecified”.
* If no participants reportedly died, leave blank.

Note also that deaths of counter-protesters should not be counted here. Instead, they should be recorded in the row representing the counter-protest.

## notes

Use this field to record any information or observations about the event that you think are interesting or important but are not already captured in the preceding fields. For example, do you want to clarify something about how or why the event was coded? Is there relevant context that isn’t captured in the other fields? This is your space to use or not, but aim to keep your observations civil and non-judgmental.

## coder

Coders should enter their first and last initials (e.g., “JU”). If you add information to an event already coded, add your initials after a semicolon and a space (e.g., “JU; JP”).

## source[n]

Paste the URL of a source from which information was added. Or, for sources that do not have URLs, enter a short description (e.g., “public submission” or “eyewitness”).

Please note: if a news source does not provide any new information that you add to the dataset, you should not paste the link in the spreadsheet. That can happen because it’s an original story that just didn’t offer any new information, or it can happen because the story is identical to one already included (e.g., an AP wire story that appears in numerous outlets).

# 

# Special Cases

A few types of events commonly arise that require some additional explanation on how to encode them.

## Future Events

If you come across information about an event that is planned but has not yet happened, add the event to the spreadsheet but leave blank the columns relating to things like attendance, arrests, property damage, and injuries. Sometimes you will later come across an after-event report that will allow you to fill in the missing details (and enter a 1 in the *event\_confirmed* column).

If you later learn that the event was canceled, delete the row. If you learn that the event was rescheduled, edit the date and move it to the correct place in the sheet.

## Counter-protests

A counter-protest is an action that is organized or occurs in direct response to another action and usually (but not always) engages directly with that original action. We treat actions as counter-protests whenever a) an event appears to have been organized primarily for purposes of confronting or challenging another event or b) when organizers describe their action as a counter-protest to another CCC-coded event and hold it on the same date, even if the counter does not take place in the same location as the action it is targeting.

This category includes planned protests, rallies, demonstrations, and the like that intend to confront other actions (e.g., demonstrators outside a venue in which a political candidate is holding a rally, or antifascist activists who mobilize to confront a right-wing march). CCC also uses this category to capture more spontaneous confrontational actions with apparent political intent, including individuals or small groups who engage in counter-messaging protesters in a sustained way (e.g., following and heckling them) or who violently attack protesters (e.g., ram protesters with their vehicle). Note, however, that an action described as a response or counter to another action is not coded as a counter-protest if it occurs on a different date.

A protest and a counter-protest each have their own row in the sheet. Whenever you record a counter-protest, be sure to do the following three things.

* Label it a “counter-protest” in the ‘type’ field.
* Assign a common macroevent ID to it and the event it is countering, using copy/paste to ensure that the IDs match.
* Use the ‘participant\_measures’ fields in the two rows to capture available information about the nature of interactions between protesters and counter-protesters (e.g., “verbally confronted and scuffled with protesters”).

Everything else should be handled just as you would any other event.

NOTE: If an event that is initially planned as a counter-protest occurs but the event it is meant to counter does not occur (i.e., is canceled), the event that does occur should be labeled something else in the ‘type’ field (e.g., “rally”, “demonstration”), and no ‘macroevent’ ID should be assigned.

## Online Events

As noted in the Introduction, CCC covers online events that otherwise meet its criteria. This means that we capture information about things like virtual rallies, protests, vigils, and walkouts (e.g., student or teacher Zoom-outs during the COVID-19 pandemic) but not online petitions, paid fundraisers, celebrations, and so on.

Online events are treated just like other events with three exceptions.

* The locality in which an online event occurs is not always specified or clear. Here is how we handle that ambiguity.
  + If an online event is associated with a specific town, record it as you usually would.
  + If the event is associated with a specific state but not a town, enter the name of the state capital in the ‘locality’ field. For example, an online get-out-the-vote rally for Wisconsin Democrats would get “Madison” as its locality and “WI” as its state.
  + If the event is intentionally national in scope, use “Washington” and “DC” as the locality and state, respectively.
  + If the event is intentionally international in scope, do not record it.
* In the ‘location’ field, enter “online” and nothing else.

## Repeating Events

Many activists across the U.S. organize demonstrations or actions that repeat on a regular schedule—monthly, weekly, daily, or otherwise—for weeks, months, years, or even decades. Many of these repetitive actions meet all of our criteria for inclusion, but they are often poorly covered by the news media, so information about them can be hard to find.

To facilitate record-keeping on these actions, we maintain a “repeaters” tab in our monthly data-collection sheet that groups repetitive events by their time cycles: daily, weekly, or otherwise. When a new repeating event is discovered, we add a row for it to this tab. And, of course, we remove rows from this tab as we learn of the end of repeating actions.

In some cases, organizers don’t announce the end of the action, and we realize that this can lead to errors of commission as we record numerous events that did not happen. Consistent with the general ethos of this project, however, we choose to err on the side of inclusion and prefer these errors of commission to the large number of errors of omission we would make if we did not allow for repeating actions in this way. To reduce those errors of commission, we periodically review the list of repeating events to see if we can update our information on their conclusion or continuing occurrence.

## Riots

CCC does not apply the label “riot” to any events. As Chenoweth (2021: 55) notes, that term has become pejorative and is often used by authorities or opponents to delegitimize protesters’ actions. In the light of the political weight of this term, we choose to apply one or more of the terms listed above—including “civil disobedience” in cases where protesters deliberately destroy or seize property but do not deliberately harm people—instead of labeling them as riots.

# FAQ

#### Do I enter an event even if sources don’t provide any crowd counts?

Yes, just leave empty the columns pertaining to crowd size (size\_text, size\_low, size\_high).

#### Do I enter an event that has not happened yet?

Yes, but leave blank the columns relating to attendance, arrests, damage, injuries, and the like. Sometimes you will later come across an after-event report that will allow you to fill in those missing details. If you later learn that the event was canceled, delete the row. If you learn that the event was rescheduled, edit the date and move it to the correct place in the sheet.

#### For location, do I need to include every reported detail?

No. Focus instead on trying to capture major addresses or landmarks that Google Maps recognizes (and you are encouraged to use Google Maps to sort out what that means). In cases where an event moves from location to location, record as many of those major points as possible, separating them with semicolons.

#### What do I enter for locality if the event spanned more than one town or traveled between towns?

Pick one notable locality (often the start or end point) and enter it in the ‘locality’ field, and then list the other towns in the ‘location’ field. For example, if protesters caravanned from Palo Alto to Sacramento and then rallied at the state capitol there, you might enter “Sacramento” as the locality, “CA” as the state, and then record “Palo Alto, CA; California State Capitol” in the ‘location’ field.

#### What if protesters from two locations or organizations join together? Are they listed as a single event or two separate ones?

If participants organize separately and begin their events separately, then their actions are listed as two separate events, even if they eventually merge their actions ([example](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.baynews9.com_fl_tampa_news_2021_10_02_hundreds-2Dtake-2Dpart-2Din-2Dreproductive-2Drights-2Drally-2Din-2Ddowntown-2Dtampa&d=DwMGaQ&c=WO-RGvefibhHBZq3fL85hQ&r=5VODqbn-CqmBWkKUzoQm8tbI0cNqXeAvJT9IJUx9asI&m=BNXzLB4fwlByLVOCmjARRxRxu0lLkcC5fIuhu9kDfSs&s=m6jIPlUMjT30iI7bkG9LSOKH2zz7uZ193vCIWkFEVME&e=)). However, if the source or sources list a meeting point so people can travel together to an event, the meeting point—e.g., a train station or parking lot—is not listed as its own event. Also, if multiple organizations each advertise on their own that they are bringing a group to the same action, that action is still listed as one event, not one event per organization.

#### Should I list an event that only appears on a Facebook event page or on an organizational calendar or website?

Yes. Keep the event listed even if you cannot find a news report. We assume it is more likely that an event happened but did not get media coverage than that the listed event didn’t happen.

#### Do we count online protests?

Yes, we also track crowds that gather online. See the Online Events section under the Special Cases header above for guidance on how to encode them.

#### How do we handle events that have both an online and in-person component?

Some events combine in-person and online elements, often but not always in the form of a livestream of the in-person action. In these cases, create separate records (rows) for the in-person and online parts, and use a common macroevent ID to link the two. This approach is similar to how we handle coordinated actions across multiple in-person locations. It also gives maximum flexibility to researchers who may want to limit analysis to in-person actions (so they can filter out the online gathering without losing the in-person action), and it allows us to record distinct information for the two (or more) parts when relevant (e.g., about the number of participants). (NOTE: This does not extend to participants or observers who independently livestream an event. We are talking specifically about cases where organizers of an in-person action also organize a parallel online component or gathering.)

#### What if an event recurs daily, weekly, or monthly?

Each day or iteration counts as a unique event and gets its own row.

#### What if a single event lasts for multiple days in a row?

Same story: each day or iteration counts as a unique event and gets its own row. The partial exceptions are events that start one day and last into the early-morning hours of the next; in these situations, only record a single event on the date the action started.

#### In the arrest column, do we count protesters who are issued citations?

No. To count as arrests, participants must be arrested or detained. A citation alone is not enough to be listed.

#### Do festivals and parades count as protest events?

The answer depends on the nature of the festival or parade and who organizes it. We generally do include festivals and parades connected to salient political themes, like LGBTQ+ pride or racial justice or Indigenous Peoples’ Day, but not ones associated with less overtly political and more “traditional” themes like July 4th/Independence Day or Veterans Day. Also, CCC generally doesn’t cover events that are solely organized by governments or government agencies, and that extends to parades as well. Ditto for only including events that are open to the public and free of charge. So, for example, a Pride flag-raising event organized by a city council would not be included, but a Pride festival and parade organized by a non-governmental organization would be, unless you have to pay to attend or watch. And, of course, when people demonstrate or protest at festivals and parades, we do include those demonstrations or protests, even if the festival or parade is not within our scope.

# Works Cited

Berglund, Oscar. 2023. “Disruptive Protest, Civil Disobedience & Direct Action.” *Politics* (online).

Chenoweth, Erica. 2021. *Civil Resistance: What Everyone Needs to Know*. New York: Oxford University Press.

Soule, Sarah and Christian Davenport. 2009. "Velvet Glove, Iron Fist, or Even Hand? Protest Policing in the United States, 1960-1990." *Mobilization: An International Quarterly* (Vol. 14, Issue 1): 1-22.