* Citations should be in Harvard format.
  + Changed
* Commas and other punctuation should go inside quotation marks.
  + I can't find any cases where this occurs (searching for all quotations in the document)
* Inconsistent use of "data set" vs. "dataset" throughout.
  + Change made: corrected to use 'dataset' throughout
* Abstract: missing word in second sentence: "We the use of one of these datasets–GDELT–for the micro-level study of political violence by comparing it to two hand-coded conflict event data sets."
  + Change made: "We **assess** the use of one of these datasets–GDELT–for the micro-level study of political violence by comparing it to two hand-coded conflict event data sets."
* pg. 2, first full paragraph: the semicolon in the following sentence should be a colon: "The results confirm a geographic bias; whereas GDELT over-reports violence close to the capital, the opposite applies in remote locations."
  + Change made
* Last sentence on pg.2: I would rephrase "Examples date a couple of decades back," to "Examples date back a few decades,"
  + Change made
* pg. 3, first paragraph: omitted space between KEDS and (Gerner et. al, 1994).
  + Change made
* Last full paragraph on pg.3: "All these datasets rely largely on human coding of news reports, and thus require significantly effort and time" should read "All these datasets rely largely on human coding of news reports, and thus require significant effort and time."
  + Change made
* Footnote no.1: the following sentence should include the comma indicated in brackets: "Some new data initiatives still in development, such as the Social, Political, and Economic Event Data Project (SPEED, http://www.clinecenter.illinois.edu/data/speed/) and the Event Data on Armed Conflict and Security Project (EDACS, <http://www.conflict-data.org/edacs/index.html>)[,]
  + attempt to utilize the speed of machine coding with direct human oversight."
* pg.4, second to last paragraph: suggested revision to the following sentence--replace "Last" with "Finally": "Last, even with identical source data and coding techniques, locational discrepancies could exist due to issues with the geo-referencing process."
  + Change made
* pg.5: "Our analysis compares GDELT to two other established data sets on political violence, the ACLED (Raleigh et al., 2010) and GED (Sundberg and Melander, 2013), for Africa from 1997 through 2008." It's unclear from the phrasing whether " for Africa from 1997 through 2008" applies to GED only or to both datasets.
  + Change made by splitting sentences: now reads “Our analysis compares GDELT to two other established data sets on political violence, the ACLED (Raleigh et al., 2010) and GED (Sundberg and Melander, 2013). To ensure overlap between these datasets, we use focus on African violence between 1997 and 2008.”
* Bottom of pg. 5: I would amend "Specifically, a GDELT event is kept if it meets the following criteria for information:" to read "Specifically, a GDELT event is kept in the dataset if it meets the following criteria for information"
  + Change made
* pg. 5: This is a bit confusing: "2. The event location is geolocated at least at the administrative district or city level." Would rephrasing this to "The event location can be geolocated to at least the administrative district or city level" preserve the intended meaning?
  + Change made
* pg. 6: Suggested revision: "The GED coding method attempts to create one record ‘per event’, meaning that events lasting more than one day only count as one observation, and differs from the other two data sets in that it only codes events where at least one fatality is recorded, leading to a lower overall event count" to "The GED coding method attempts to create one record ‘per event’, meaning that events lasting more than one day only count as one observation, and IT differs from the other two data sets in that it only codes events where at least one fatality is recorded, leading to a lower overall event count. "
  + Change made
* pg. 12: "Of particular interest here is the coefficient linking geographic remoteness, as measured by distance to the capital, to the presence of conflict in a given grid-month. " Is the grid-month different than the cell-month?
  + Change made: corrected to “cell-month” throughout
* pg. 12: Suggested revision: "Machine coding of event datasets can process large amounts of reports in little time, and thus have some advantages over their human-coded counterparts" to "Machine coding of event datasets can process A large NUMBER of reports in little time, and thus have some advantages over their human-coded counterparts."
  + Change made
* pg.14: "This will enable more thorough validation studies of automatic geo-coding, but also going beyond this." The "also going beyond this" strikes me a bit unclear.
  + Change made: rephrased sentence to read “This will enable more thorough validation studies not only of automatic geo-coding, but of event content coding as well.”
* pg.14: Two suggested revisions to the following sentence: "GDELT’s high level of noise in the data, coupled with geographic accuracy issues suggests that using GDELT instead of a more detailed hand- coded data set would likely be counterproductive. " First, I would suggest rewording the first two clauses as follows: "The high level of noise in the GDELT data, coupled with the geographic accuracy issues we identify, suggests..." Second, I would suggest revising the last clause of the sentence to restate for which goals the use of GDELT is likely to be counterproductive--this would make more of an impact as the second-to-last sentence in the piece.
  + Change made: rephrased sentence to read “The high level of noise in the GDELT data, coupled with the geographic accuracy issues we identify, suggests that using GDELT instead of a more detailed hand-coded dataset may lead to distorted or incorrect inferences when explaining spatial dynamics in civil war violence.”