Protest in the 6ix

What can *new* and *traditional* data sources tell us about Toronto protest between 1995 and 2020?

Abstract

News media has often been used for the identification and analysis of protest event, the analysis of which has provided many important revelations. Twitter data reaffirms some of these conclusions, such as those about protest themes and occurrence. it also challenges some of the previous findings, indicating important avenues for future research. Used in conjunction, both sources can provide a rich illustration of activist movement in Toronto.

Methods

There were two methods of data collection used to identify protest events. The first, which aligns with a more traditional means of event identification, was the use of newspapers articles to find events. To maintain consistency with previously collected datasets, the Toronto Star was the primary source. A series of key words such as 'protest', 'rallied', and 'demonstrated' were used to identify articles. Once An article was found it was read, and if relevant, added to the dataset via a preconstructed coding methodology.

The second method sought to take advantage of the changing forms of media that have become available. Of specific interest in this case, is twitter. Using a python based application that works through Twitters API, we scraped public posts with strings containing 'protest', 'rallied', and 'demonstrated'. These key words were selected to maintain as much comparability with the news based searches as possible.

Due to limitations in the date range accessible for scraping, the Twitter dataset started July 15th and ended August 15th. To compare datasets collected from each method, the same one month date range was used from each year between 1995 and 2020). For trends suspected to extend beyond the one month period, full sets from selected and relevant years were analysed.

Findings: News Media

In general, previous observations about Toronto protest are reflected in the trends found in news sourced datasets. These observations include Toronto's focus on local or provincial issues, relative low volume of events in July-August, and common protest locations. More novel are observations of protest subject. Though anti-poverty and affordable housing have been the subject of continuous protest since 1995, there appears to be a decrease in the volume of poverty focused events and an increase in police brutality and racism focused events.

Beyond just subject matter, there are also fluctuations in event volume. The attached chart indicates the change in number of events overtime for the selected period. Major changes tend to come around world events. Post 9/11 there is a steep drop of events, likely related to the changes in security infrastructure seen in North America. Spikes in protest during this time come during periods of right leaning provincial governance ('95-03', 18'-20'). This is an avenue of research worth further exploration.

number of events by year for July 15-August 15
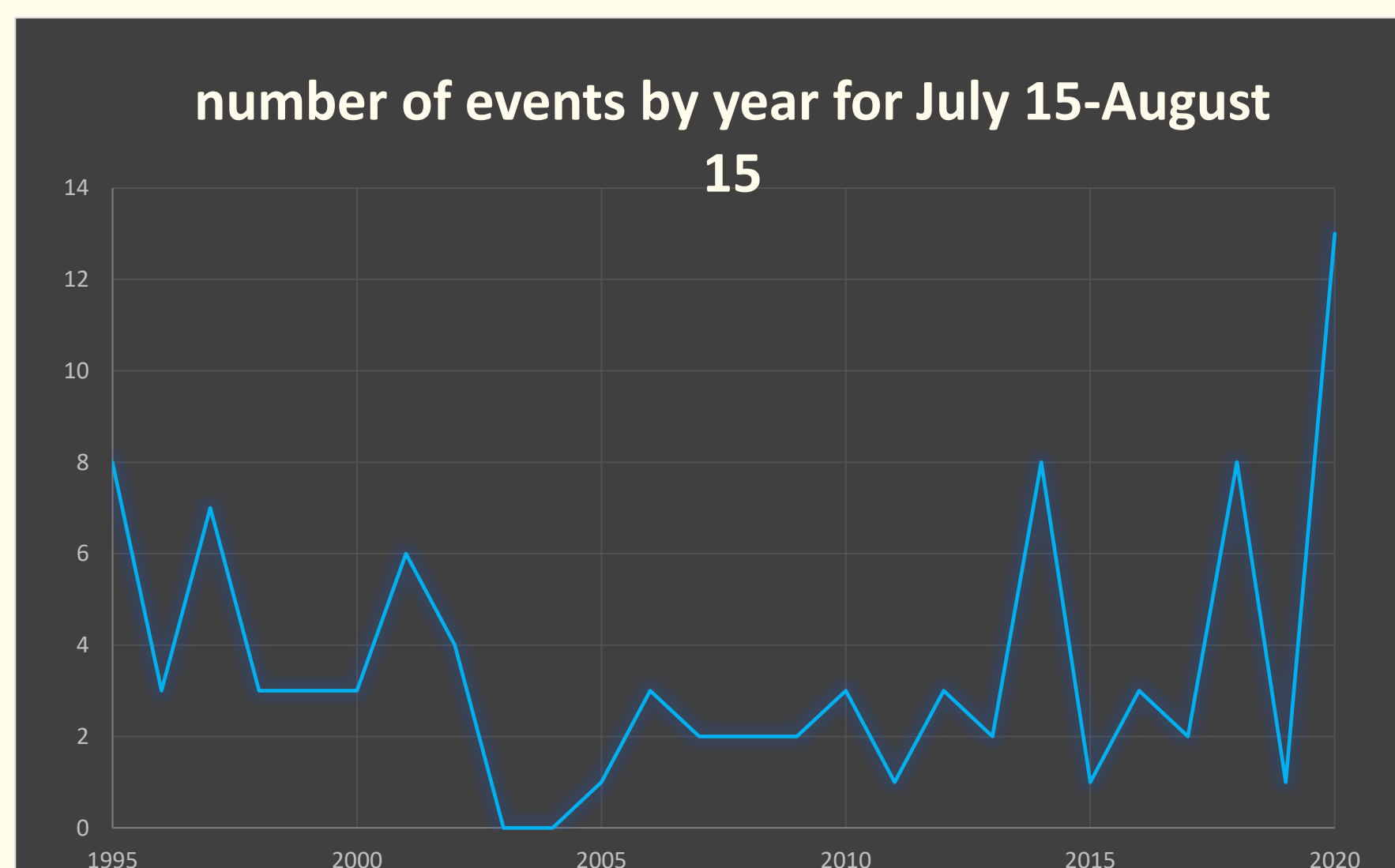
For the years with complete data sets, similar trends can be observed for the entirety of the year.

Findings: Social Media

Although limited in its range, using twitter for event identification has proved fruitful. It affirms several key findings from news media. First it affirms the key issues of the current activist climate. Issues surrounding racism and policing take precedence within the time frame observed. It also supports news medias reports about what kind of events are occurring (ex. rally, march, creative action, etc.).

Several new findings arise from this data set as well. Most obvious is event volume. Using news media, 11 events were identified. Via Twitter, more than 45 events were found. Whether this is proportionally reflected in the findings from news media remains to be seen.

Twitter also identified a greater proportion of events focused on non-domestic issues than was found vie news media. The groups involved with a given protest were found to be much more diverse in nature when observed through social media.

In comparison...

Using the two datasets together appears to be the optimal methodology. A finding first taken from the news dataset, but given powerful support by the twitter dataset, is the overall reduction in time between a catalytic event and the related protest. For example, a week or more elapsed between the police killing of Edmond Yu, Otto Vass, and Hugh Dawson, and the first protests centered around each murder. After the rise and proliferation of social media (2004 onwards), it appears to take 3 days or less between an occurrence or police brutality and the ensuing protests (ex. Sammy Yatim – 2 days, G20 – immediate).

A greater understanding of the kinds of people protesting is also gained by combining the methods. The News reports tend to focus on left leaning protest, giving the appearance of a general lack of protest from conservative/right leaning groups. Twitter data shows a significantly greater number of right leaning protests. This could be due to the generally smaller conservative protests being outshined by the much larger protests on police brutality and racism. Despite a multitude of possible explanations, the discrepancy between the two sources of data may highlight a bias in the sorts of events identifiable through news media. If this is the case, further research is needed to understand how this bias has influenced our understanding of protest in Toronto.

Conclusion

News media has provided strong indications of the key trends in Toronto protest. The addition of Twitter data has strengthened some of these conclusions and called others into question. Both data collection methods have gaps: the news standardizes information and tends to homogenize activist groups, Twitter on the other hand offers a more diverse set of information and views but rarely indicates event size or police activity. The research application used here also has its limitations, predominantly created by time limitations. The Twitter sample was small, it is more useful as an indicator for future research than a source for conclusions about protest itself. The month selected from the news dataset may also not be representative of the whole year. A full analysis of the data would likely provide a multitude of insights.