



Depends on how you count them: the value of general propensity choropleth maps for visualising databases of protest incidents

Martin Bekker

To cite this article: Martin Bekker (2023) Depends on how you count them: the value of general propensity choropleth maps for visualising databases of protest incidents, Journal of Maps, 19:1, 2064778, DOI: [10.1080/17445647.2022.2064778](https://doi.org/10.1080/17445647.2022.2064778)

To link to this article: <https://doi.org/10.1080/17445647.2022.2064778>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of Journal of Maps.



[View supplementary material](#)



Published online: 24 Apr 2022.



[Submit your article to this journal](#)



Article views: 1698



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)



Depends on how you count them: the value of general propensity choropleth maps for visualising databases of protest incidents

Martin Bekker ^{a,b}

^aSchool of Electronic and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa; ^bCentre for Social Change, University of Johannesburg, Johannesburg, South Africa

ABSTRACT

Public protest represents an important sanction on rulers and institutions. Protest is a quotidian phenomenon in South Africa; perhaps the defining element of post-apartheid political life. Geographic representations of protest abound – typically dot distribution maps – but these merely confirm that more protests occur where there are more people. Visualisations of protest per capita and protestors per capita (or ‘general propensity’), which are best rendered as choropleth maps, are well-placed to overcome this limitation. The South African Police Services’ database of protest is the largest publicly-available single-country protest event database. Having used machine learning to classify 89,000 protest events, I locate each within one of the country’s 234 municipalities, and depict these events using counts, count per capita, and the general propensity. This reveals a proportionally high number of rural protests, and that municipalities hosting major industries, along with provincial seats of government, present the highest propensity for protest.

ARTICLE HISTORY

Received 4 August 2021
Revised 29 March 2022
Accepted 29 March 2022

KEYWORDS

Protest; unrest; social movements; general propensity; South Africa

1. Introduction

The occurrence of public protest, especially if widespread, presents an important popular censure on rulers, and the institutions they represent. Though often depicted as a politically subversive undertaking, protest itself is not so much a malady as it is a symptom of severe social distress. From BLM protests in North America and anti-neoliberalism protests in South America, to anticorruption protests in Lebanon, and pro-democracy protest in Hong-Kong – the recent past has seen a rise in people taking to the streets to demonstrate against their political leadership’s actions, or failure to act (Ortiz et al., 2013).

Protest is a large and quotidian phenomenon in South Africa, and perhaps the defining element of post-apartheid political life in the country. Between 1997 and 2013 more than 14 protests occurred daily, on average, across the country (Bekker, 2021a), with scholars projecting increases since that time, is in other parts of the world, such that the number of protests in South Africa has nearly doubled over 20 years (Runciman et al., 2016, p. 5). As such, South Africa has been dubbed ‘the world’s protest capital’ (Odendaal, 2016, p. 287).

Scholarship generally tends to depict protests as an urban phenomenon (Alexander & Pfaffe, 2014; Biggs, 2018; Booysen, 2007; Dawson, 2014; Noble & Wright, 2013). Cities and metropolises are home to denser

arrangements of people, and most countries, including South Africa, have seen urbanisation rates rise above 50%. A visual representation of frequency counts of protests on a map of any country, using dot maps (the practice by leading social movement scholars) might intuitively place the majority of dots on the country’s largest cities. After all, urban population size and the number of large cities in a country are positively associated with protest incidence (Fox & Bell, 2016). For South Africa, several such maps exist (Figure 1). However, such depiction might turn out to be the manifestation of a less-than-ideal cartographic practice, which can obfuscate, instead of reveal, the view of what is happening on the ground.

In this paper, I offer two alternatives to maps depicting counts of protest – proportional counts (i.e. counts of protest per capita) and general propensity (i.e. the number of protestors per capita). The former might be expected to deemphasise urban localities, owing to high urban populations, while the latter will shift the focus to places where, when people protest, they tend to do so in large crowds (such as seats of governments). The ability to utilise such alternatives is, however, dependent on access to appropriate data.

Students of social movements and protest event analysis (PEA) have used various approaches to study protests, ranging from small-n case studies to

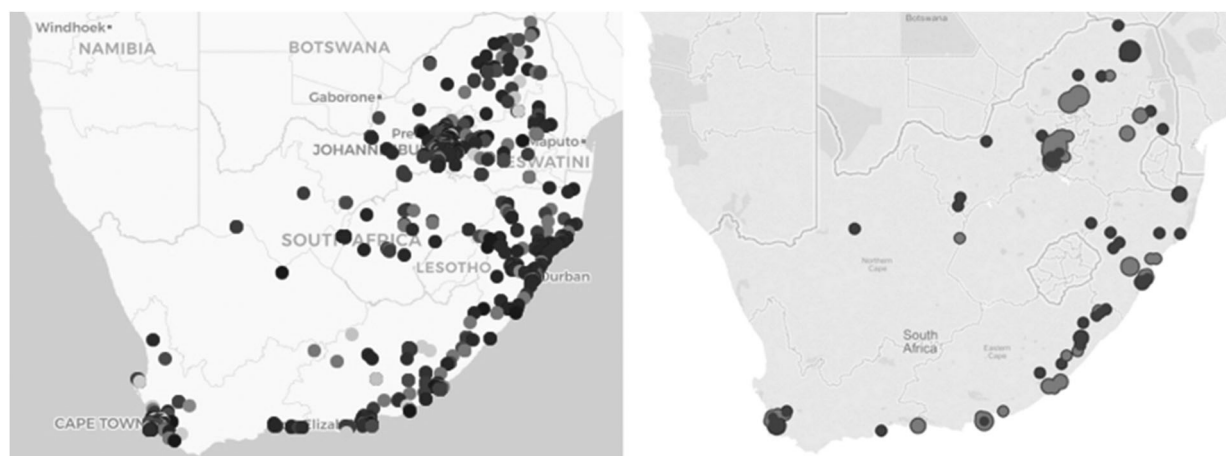


Figure 1. Depictions of protests in South Africa by the Institute of Security Studies (Institute for Security Studies, n.d.) on the left, and from the Armed Conflict Location & Event Data Project (Wigmore-Shepherd & Moody, 2016), on the right.

large-n quantitative studies. Quantitative studies generally make use of media-generated datasets of protests or subsets drawn from administered data. Media-derived datasets are well-understood to amplify the urban gaze (Duncan, 2016), their inclusion in news reports being contingent on newsworthiness to the reading public (meaning large, urban, violent events are much more likely to be reported than small, rural, peaceful ones). Owing to the size of administrative databases, generally, sample-based methods have been used when dealing with large datasets. Recently, natural language processing techniques have been adopted in PEA, showing the way for future protest analysis (Hanna, 2014; Kriesi et al., 2020; Weidmann & Rød, 2019).

This article shows how the world's largest publicly accessible dataset can be used to depict the local protest propensities in South Africa's 234 local municipalities. The next section provides an overview of the data, and a reflection on its scope and relative merits. Thereafter, I chronicle in brief the machine learning techniques and data processing steps followed. The resultant data, fit for mapping not only as counts, but also as protests per capita, and ultimately, depicting protestors per capita (drawing on aggregate crowd sizes), permits refined visualisations that grant insights into local pressure factors, with scholarly and policy implications alike. I end by discussing these, along with the merits of approaches to mapping that go beyond dot distribution maps.

2. Materials

The Incident Registration Information System (IRIS) is regarded as the world's best source of protest data (Alexander et al., 2015; Lancaster, 2018, p. 31). It is collected by the South African Police Service (SAPS), and captures, at the risk of oversimplification, all known 'assemblies,

demonstrations and all gatherings ... whether recreational, peaceful, or of an unrest nature' (SAPS 2014). I obtained these data via the Centre for Social Change at the University of Johannesburg (see discussion of the legal process in Alexander et al., 2015).

IRIS provides an unvarnished and complete record of crowd incidents attended by the SAPS. While it is in the SAPS' interest not to ignore or underreport incidents (for reasons of intra-police funding competition) reporting does, however, vary according to the resources available to public order policing (Runciman et al., 2016). This means that while IRIS does not *deliberatively* underrepresent protest counts, it may do so systematically.

Another salient consideration is how protests are captured in IRIS. A protest event, regardless of when it begins, is captured as an individual record (which includes date, place, motive, and extensive police notes, among other variables). But if the protest continues past midnight on the day it started, it is classified as two events. Using a 'day' to frame protest events implies protests that last several days – with crescendos and diminendos over time – are seen as separate events, with implications for calculations based on protest counts.

These imperfections notwithstanding, owing to its value, several scholars have engaged with IRIS, including, among others Alexander et al. (2015), who showed that only 10% of incidents appear to be violent; Runciman et al. (2016), who using a sample extrapolation, found that there might be as many as 70,000 cases (in the 17 years spanning 1997–2013); Duncan (2016), who revealed patterns of protest repression; and De Juan and Wegner (2019), who found correlations between levels of services and counts of protests (also using a relatively small subsample).

3. Methods

Below follows a summation of the data processing techniques, which is discussed in elaborate detail in Bekker (2021b). The data, stored in 32 files in csv format, were combined and the records cleaned (primarily removing duplicates or missing records). The variable of interest was the ‘notes’ feature, wherein police provided a narrative account of the protest event’s unfolding and denouement over time – ranging from a dozen to hundreds of words per record.

In order to extract the reported number of people protesting per event (where this was recorded), I used a Visual Basic script that identified the largest integer in the ‘notes’ feature. Thus, in the case of the following fictitious note, ‘a group of 10 men gathered, it grew to 15 people, eventually 25 people marched, later only 20 remained before dispersing’ – the number 25 would be returned. However, this also returned many unlikely crowd numbers: telephone numbers, identification (social security) and passport numbers, car registrations, etc. For ethical and practical reasons, all non-probable crowd numbers were eliminated.

Using an ensemble machine learning algorithm – gradient boosted decision trees – I classified each record according to a four-way taxonomy: non-protest, orderly protest, disruptive protest, or violent protest. This involved first hand-classifying a training and testing sample of a thousand cases, and considering the performance of several candidate algorithms and hyperparameters. Having classified the unlabelled data, I identified and removed non-protest incidents from the set.

Postprocessing of the data involved, among others, attributing each of the resultant 89,000 records of protest incidents to one of SA’s ‘local’ municipalities (including ‘metropolitan’ municipalities, but not ‘district’ municipalities – the latter made up of groups of local municipalities). Here, I used a combination of police station data, the ‘notes’ variable, online maps (OpenStreetViewer and Google Maps), the South African Independent Electoral Commission’s list of voting stations, and Statistics South Africa’s *mainplace* and *subplace* name lists. With a dataset containing protest incidents by type, level of tumult, and location in hand, the next step was to map these.

4. Results

Maps of protest frequencies are commonly rendered as symbol-based (dot maps) or heatmaps. In the case of South Africa, there are many examples of such maps. Figure 1 illustrates two such instances. Regarding both of these maps, one immediately appreciates the arguments advocating protests’ perceived urban predilections, with protest clusters at the extreme

southwest (Cape Town), along the east coast (Buffalo City to the south, eThekweni/Durban to the north), and the northern interior (Gauteng Province, where Johannesburg and Tshwane/Pretoria are located).

These maps show the incidents by location, with the number of dots (on the left), or the size of the dots (on the right) indicating counts. Visualising the occurrence of protest incidents helps to display the demographic realities of urbanisation and the uneven human density distributions between urban and rural geographic units in South Africa. Figure 1 provides cursory confirmation that four municipalities – the City of Tshwane, the City of Johannesburg, the City of Cape Town, and Nelson Mandela Bay – host the bulk of protests in the country, and thus contribute disproportionately to count-based analyses.

Instead of counts, one might choose to visualise the count of protest divided by the population size of each municipality, which one might call the ‘proportional occurrence’. Such a per capita model represents an improvement¹ on occurrence models; however, a drawback to its use is that in order to maintain a fair measure, the protests counted should be limited to years around the population headcount utilised. Following this, I only considered protest incidents from 2009 to 2013, owing to the South African national census having been taken in 2011.

Perhaps the closest one can get to the true measure of protest is to consider the number of participants in all protests in an area, per capita, or the ‘general propensity’ for protest, as I call it. That is, the product of the count of protests and the average crowd size for each municipality, divided by the municipal population, as illustrated in the following equation:

$$\frac{\text{municipal protest count} \times \text{average size of protests in municipality}}{\text{municipal population}} \quad (1)$$

Just as the per capita calculation brings a measure of proportionality to protest counts (which would otherwise overestimate the influence of areas with high populations), a general propensity count takes crowd sizes into consideration, thus largely correcting the biases introduced when small and large protests are treated equally.² The general propensity is thus sensitive to the fact that, while urban areas might proportionally host fewer protests than some rural areas, they tend to host *larger* protests; it also supports visualising the effect of this consideration. The inclusion of population size means that, as in the case of proportional occurrence visualisations, best-practice would be to confine protest incidents to those around the time at which population estimates are made.

The maps below show the outcomes of the steps developed above, starting with a depiction of South Africa’s municipalities by population, and moving to show the occurrence of protest (counts), proportional

occurrence (counts per capita), and general propensity (protestors per capita). Note that the maps are produced (online) in great detail under ‘supplementary material’.

Figure 2 is a choropleth map illustrating the population of each of the country’s municipalities, using their 2011 demarcation, with grading running from light (fewer) to dark (more people). The 15 municipalities with populations above 500,000 are labelled; three coastal cities and three municipalities in the heartland of Gauteng are home to more than a million residents, respectively, while in the arid and sparsely populated Northern Cape, no municipality, except for Sol Plaatje (seat of the town of Kimberley, the provincial capital) has more than 200,000 residents.

Figure 3 shows the frequency and distribution of protests; that is, protest occurrence. This confirms the rudimentary assumption that protests occur in the major population centres: one notes the high counts in the population centres in Cape Town (and surrounds) in the southwest, Johannesburg and Tshwane/Pretoria in the northern interior, eThek-wini/Durban on the east coast, and Nelson Mandela Bay (Gqberha/Port Elizabeth) in the Eastern Cape.

Figure 4 presents protest per capita. Viewed this way, the urban centres and provincial capitals tend not to be the sites of high proportional occurrence of protest. Instead, the rural landscape is suddenly

alive with protest. Areas with specific types of economic activities, or other local characteristics (discussed below) appear with higher protest per capita profiles. It is noted that none of SA’s major urban centres – Cape Town, eThek-wini/Durban, Johannesburg, and Tshwane/Pretoria – are listed among the places in the highest category of protests per capita. Instead, the map intimates that, on the one hand, the iron, platinum, chrome, and diamond mining belt in the North West Province (from Rustenburg to Naledi), and, on the other, the manufacturing sites at the southern coastal ports (plus some agricultural areas in the Free State and Southern Cape), see the most protests, speaking proportionally. Notably, this approach reveals that protest in South Africa is more rural than previously appreciated, with some ‘surprise’ municipalities that, notwithstanding little press coverage of their public protests, seem to host proportionally high levels of protest.

Figure 5 depicts South African municipalities by their general propensity to protest. Displaying protestors per capita, turns the focus somewhat back to the large population centres, where crowd sizes are bigger, and where focal points of large protest gatherings – such as provincial legislatures and company headquarters – tend to be located. Hence provincial capitals (Cape Town, Buffalo City, The Msunduzi, Mangaung/Bloemfontein, Mbombela, Tshwane/Pretoria,

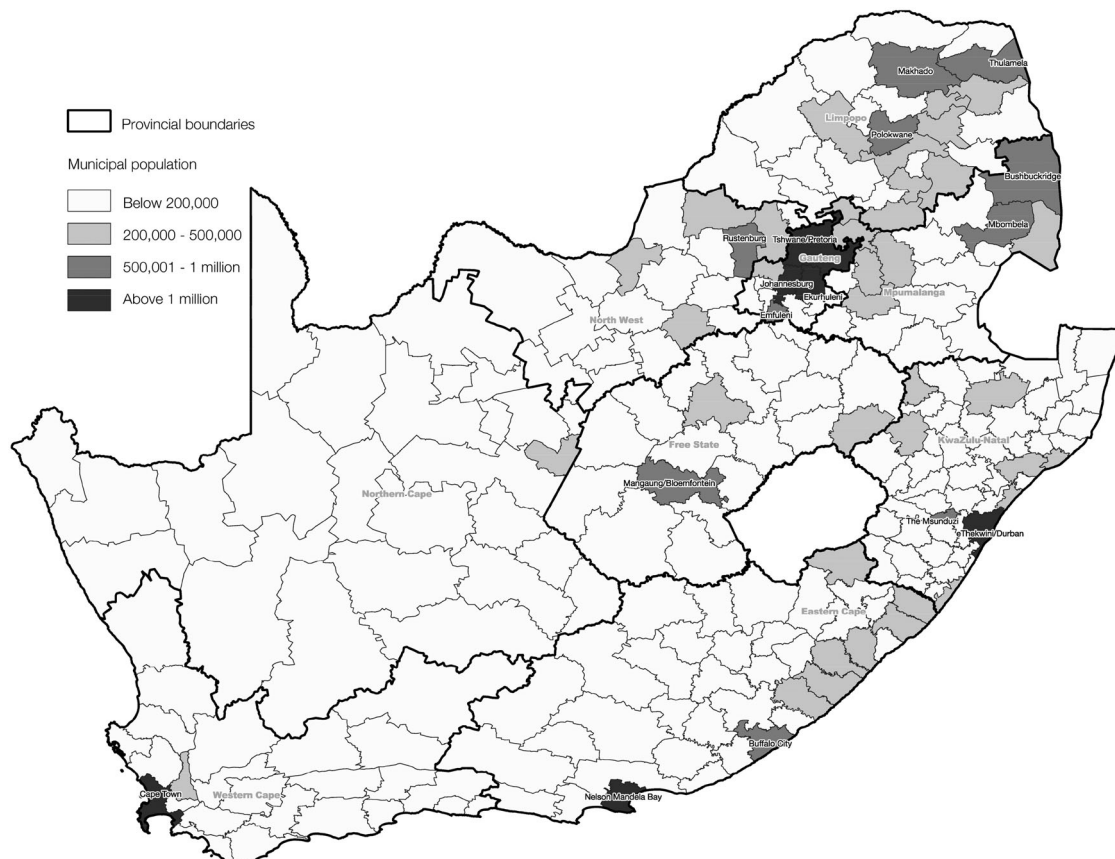


Figure 2. South African municipalities by population size (2011).

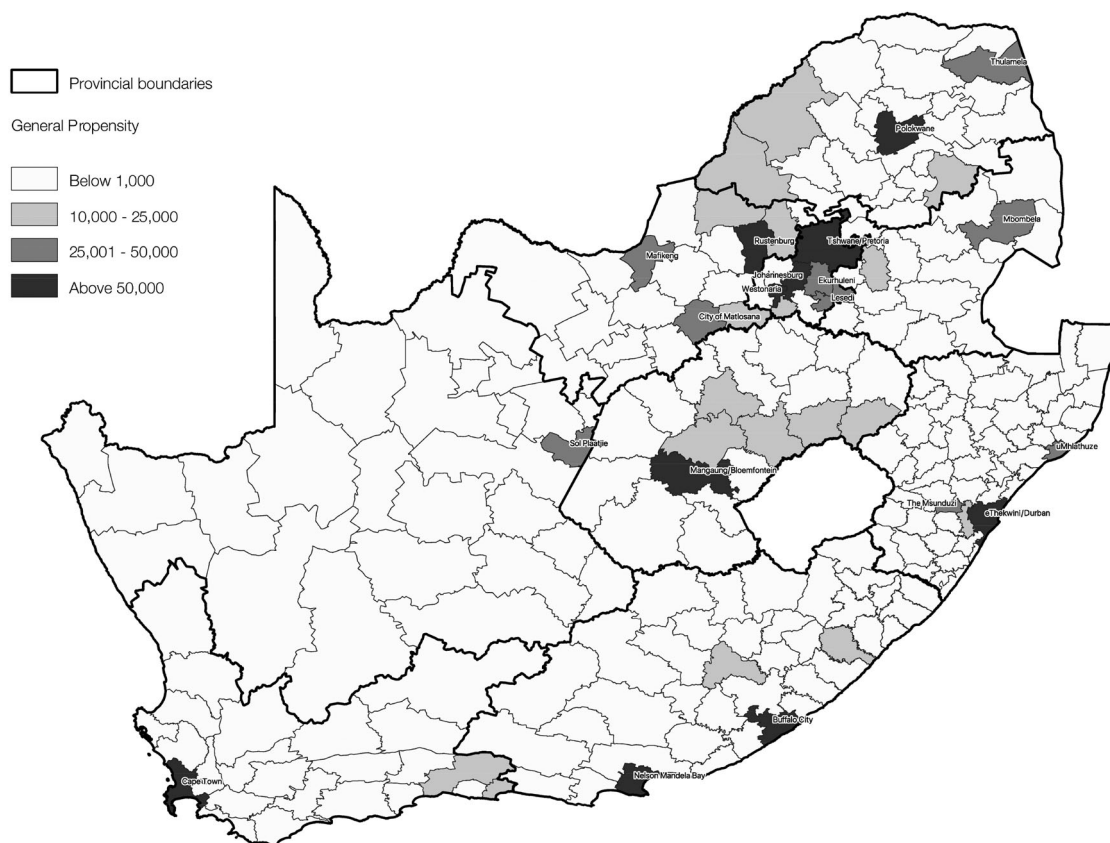


Figure 5. South African municipalities by number of protestors per capita (2009–2013).

Polokwane, Mafikeng, and Sol Plaatjie) are given prominence. Other notable sites are municipalities hosting sizeable primary or secondary industries with high trade-union representation, such as Rustenburg and Westonaria (mining), Nelson Mandela Bay (automotive manufacturing), and uMhlathuze/Richards Bay (dock work).

5. Discussion

Protest is often seen as a means of last resort (Nyir & Wray, 2012, p. 30) inasmuch as it implies that other, official channels of addressing collective grievances are not available or have been exhausted. As such, protest communicates ordinary citizens' grievances, a matter that ought to concern policy-makers and social scientists alike. Despite South Africa's transition to democracy, the country continues to experience numerous community and labour-related protests daily; by some claims, more than any other country.

Following the data analysis presented above, I found that South Africa saw more than 89,000 police-recorded protest events over the 17-year period from 1997 to 2013, implying an average of more than 14 protests per day (Bekker, 2021a). Moreover, each of South Africa's 234 local municipalities hosted at least one protest event between 1997 and 2013 (the range is two to 6 564 events).

While the number of protests increased over time, the size of protests did not. The estimated average crowd size was 435 people per protest in 1997 and 452 in 2013. Crowd size calculations allow one to consider the concept of ‘person-days of protest’ (or possibly ‘workdays lost’) over the period of consideration, calculated as the product of average crowd size and the number of protests. Assuming a non-significant number of people joining multiple protests on the same day, South Africa had over 22.4 million person-days of protest over a 17-year period, at an average of 1.3 million person-days per year.

This article presents an innovation with regard to mapping protest in South Africa. Crowd sizes of protests in South Africa have hitherto not been used to inform calculations of propensities to protest, most likely owing to most scholars relying on media reports by journalists, which unlike police reports, generally do not include crowd sizes, or, if they do, use conveniently rounded figures. Moreover, IRIS's 89,000 protest events have not before been differentiated, or represented cartographically by count, proportion, or general propensity.

Choropleth maps trade accuracy for locating individual protest events (at which dot maps excel) for a more accessible rendering of the aggregate picture of protests. The maps reveal there is quite a difference in the geographic patterns and findings in general, depending on how one counts protest; this, in turn,

should affect how one conceives of patterns of protest. Shifting from counting the number of protests per municipality to counting protest per capita (per municipality), alters the national profile of protest from predominantly urban to rural – a reality best illustrated visually. In light hereof, to select one among many advocates conceiving of protest as all but exclusively urban, Booysen is not wrong that protests tend to be ‘concentrated in the urban and metropolitan areas’ (2007, p. 23); however, such an analysis offers a limited improvement on our understanding of protest propensities and the locations of protest events. In addition to maps of protest proportions, propensity maps, which represent yet another perspective on protest, evince the prominence of political centres and trade-union bases in mobilising large demonstrations.

Data availability

Source data: Owing to the confidential and sensitive information contained in the South African Police Service’s IRIS, it is not available for public dissemination. To prevent misuse, it may be obtained via court order, citing the South African Promotion of Access to Information Act, Section 15, as set out by the South African Human Rights Commission.

Software

The maps and layout were created and edited using QGIS 2018, version 3.16 ‘Hannover’.

Notes

1. Consider a fictionalised example: if the City of Johannesburg and Mier Local Municipality (in the Northern Cape) both host 1000 protests in a year, it cannot be reasonably reported that they have the same levels of protest. Mier’s population is 7000 (thus 14.3 protests per capita), while the Johannesburg population is around 4000,000 (thus 0.025 protest per capita). Per capita reporting not only is more informative, but also more truthful.
2. For example, one might consider municipality A with ten protests, each attended by 20 people, contrasted to municipality B, with five protests of 100 people each. Common approaches would see this example as a case of ten protests against five (suggesting A has more protests), instead of 200 protesters (or participants) versus 500 (which shows that B has more protesters). In addition, the per capita calculation ensures that municipal population size is also factored in. However, crowd sizes are notoriously hard to capture accurately (Ranjan et al., 2018); my method of crowd size estimation is detailed in Bekker (2021b).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the South African Research Chair in Social Change, administered by the South African National Research Foundation and funded by the Department of Science and Technology. [Grant Number 71193].

ORCID

Martin Bekker  <http://orcid.org/0000-0001-8766-937X>

References

- Alexander, P., & Pfaffe, P. (2014). Social relationships to the means and ends of protest in South Africa’s ongoing rebellion of the poor: The Balfour insurrections. *Social Movement Studies*, 13(2), 204–221. <https://doi.org/10.1080/14742837.2013.820904>
- Alexander, P., Runciman, C., & Maruping, B. (2015). *South African Police Service (SAPS) data on crowd incidents: A preliminary analysis*. Social Change Unit, University of Johannesburg.
- Bekker, M. (2021a). Language of the unheard: Police-recorded protests in South Africa 1997–2013. *Review of African Political Economy*. <https://doi.org/10.1080/03056244.2021.1953976>
- Bekker, M. (2021b). Better, faster, stronger: Using machine learning to analyse South African police-recorded protest data. *South African Review of Sociology*. <https://doi.org/10.1080/21528586.2021.1982762>
- Biggs, M. (2018). Size matters: Quantifying protest by counting participants. *Sociological Methods & Research*, 47(3), 351–383. <https://doi.org/10.1177/0049124116629166>
- Booyesen, S. (2007). With the ballot and the brick: The politics of attaining service delivery. *Progress in Development Studies*, 7(1), 21–32. <https://doi.org/10.1177/146499340600700103>
- Dawson, H. H. J. (2014). Patronage from below: Political unrest in an informal settlement in South Africa. *African Affairs*, 113(453), 518–539. <https://doi.org/10.1093/afraf/adu056>
- De Juan, A., & Wegner, E. (2019). Social inequality, state-centered grievances, and protest: Evidence from South Africa. *Journal of Conflict Resolution*, 63(1), 31–58. <https://doi.org/10.1177/0022002717723136>
- Duncan, J. (2016). *Protest nation: The right to protest in South Africa*. University of KwaZulu Natal Press.
- Fox, S., & Bell, A. (2016). Urban geography and protest mobilization in Africa. *Political Geography*, 53(July), 54–64. <https://doi.org/10.1016/J.POLGEO.2016.02.004>
- Hanna, A. (2014, April 15). Developing a system for the automated coding of protest event data. SSRN Electronic Journal, Entry. <https://doi.org/10.2139/ssrn.2425232>
- Institute for Security Studies. (n.d.). *Public protest and violence map*. <https://issafrica.org/crimehub/maps/public-violence>
- Kriesi, H., Lorenzini, J., Wüest, B., & Hausermann, S. (Eds.). (2020). *Contention in times of crisis: Recession and political protest in thirty European countries*. Cambridge University Press.
- Lancaster, L. (2018). Unpacking discontent: Where and why protest happens in South Africa. *South African Crime Quarterly*, 64(64), 29–44. <https://doi.org/10.17159/2413-3108/2018/v0n64a3031>

- Noble, M., & Wright, G. (2013). Ring of fire: Socially perceived necessities in informal urban settlements in South Africa. *Policy and Politics*, 41(2), 259–277. <https://doi.org/10.1332/030557312X655440>
- Nyar, A., & Wray, C. (2012). Understanding protest action: Some data collection challenges for South Africa. *Transformation: Critical Perspectives on Southern Africa*, 80(1), 22–43. <https://doi.org/10.1353/trn.2012.0050>
- Odendaal, A. (2016). South Africa's incomplete peace. In O. Richmond, S. Pogodda, & J. Ramovic (Eds.), *The Palgrave handbook of disciplinary and regional approaches to peace* (pp. 287–298). Palgrave Macmillan.
- Ortiz, I., Burke, S., Berrada, M., & Cortés, H. (2013). World protests 2006–2013. Initiative for Policy Dialogue and Friedrich-Ebert-Stiftung New York Working Paper.
- Ranjan, V., Le, H., & Hoai, M. (2018). *Iterative crowd counting*. Proceedings of the European Conference on Computer Vision (ECCV), 8–14 September. Munich, Germany. (pp. 270–285).
- Runciman, C., Alexander, P., Rampedi, M., Moloto, B., Maruping, B., Khumalo, E., & Sibanda, S. (2016). *Counting police-recorded protests: Based on South African police service data*. Centre for Social Change.
- South African Police Service. (2014). *Annual Report 2013/14 Financial Year*. Pretoria: South African Police Service.
- Weidmann, N., & Rød, E. (2019). *The internet and political protest in autocracies*. Oxford University Press.
- Wigmore-Shepherd, D., & Moody, D. (2016). *South Africa May 2016 Update*. Armed Conflict Location and Event Data Project. <https://acleddata.com/2016/05/06/south-africa-may-2016-update/>