

Report on Open Research Hackathon Event: Exploring the Relationship between Crowds, Safety and Crime

Reka Solymosi, Samuel Langton, Taku Fujiyama

Background

We know that crowded environments can be unsafe. For instance, crime levels increase in and around sports stadiums when attendees flock to attend events (Kurland et al., 2014). Facilities which attract high ambient populations, such transit nodes (e.g. bus stations), are known to produce criminogenic hotspots (Ceccato et al., 2015). Crowding on street corners has been found to facilitate pickpocketing (Levine et al., 1986), and in the reverse causal direction, *restrictions* on crowding and population mobility have been identified as the leading cause of the crime drop witnessed during the COVID-19 pandemic lockdown (Halford et al., 2020).

That said, the relationship is not clear-cut. While transit stations might be criminogenic places on aggregate, there is considerable temporal variation to these patterns (Irvin-Erickson & La Vigne, 2015). Passengers often feel most safe, and at a minimal risk of being victimized, when public transport is busy (Newton, 2004). Out on the street, pedestrians passing-by potential targets can create ‘eyes on the street’ which deter victimization and reduce criminal opportunity (Cozens, 2011). The density of pedestrians in public spaces has been found to decrease fear of crime, but only to a threshold, at which point pedestrian density can be detrimental to feelings of safety (Tchinda & Kim, 2020).

The notion that crowding impacts on human behavior is not new, and it is certainly not restricted to geography or criminology disciplines. The impact of crowding on human behavior has been examined in psychology for some time (Freedman, 1975; Stokols, 1972). Here, findings point towards a potential relationship with crime. For instance, the invasion of private space (e.g. high crowd density, crossing someone’s path) has been found to prompt negative emotional responses, such as aggression (Engelniederhammer et al., 2019). This might insight criminal acts such as assault. In policing and event management, unsuitable interior architecture (e.g. poorly placed exit and entry doors) in nightclubs has been found to produce sub-optimal pedestrian flows, which in turn facilitate overcrowding and aggressive behavior (Macintyre & Homel, 1997).

To contribute to this growing (but incomplete) evidence-base, we organized an online event, bringing together students and researchers from a diverse array of backgrounds and academic disciplines. Exploiting novel sources of open data, we organized a hackathon event ‘Exploring the Relationship between Crowds, Safety and Crime’ in collaboration with the University of Manchester and the International Association of Traffic Safety Sciences. In this paper, we summarize the design, format and key findings from this event to (1) actively contribute to scientific advancement in the field, and (2) provide a ‘proof of concept’ for wider usage in crime and safety research.

Hackathon

Design

The hackathon was held over two days in December, 2020. Due to COVID-19 social distancing measures the event was hosted online using the Zoom conferencing platform. We advertised the hackathon through mailing lists and social media in the weeks prior. No costs were charged to attendees and we explicitly stated that the event was open to a diverse array of professions (i.e. academic and non-academic), disciplines and experience in data-driven research. Between 25 and 30 participants attended the event throughout the two days. Most attendees were from higher education institutes in the United Kingdom, although the online

platform did encourage some international participants from Malaysia and observers from International Association of Traffic Safety Sciences (IATSS) in Japan. Participants' experience range from a high school pupil to senior academic staff, and the disciplines represented included geography, psychology, environmental design, criminology and data science.

We designed the programme so that the hackathon served as both a substantive research activity and an educational tool for participants. The educational component included a range of parallel workshops which people could choose from, including introduction to Application Programming Interfaces (API), introduction to Python, introduction to R, access to open police-recorded crime data, and an introduction to GitHub and version control. Two guest speakers provided lectures on topics on their respective expertise, namely, Dr Taku Fujiyama on crowding, traffic density and safety, and Dr Andrew Newton on theoretical and empirical research on 'busyness' from a crime and place perspective. These activities served to extend attendees' existing expertise, but also to ensure that less experienced participants could actively contribute to the hackathon.

For the research component, attendees were divided into teams according to their research background. Prior to the hackathon, those who registered were sent a questionnaire to gauge their research interests and experience with software (e.g. GIS, R). Three teams were manually selected to ensure diversity and balance based on these answers. Each team was tasked with exploring the topic of 'crowds, safety and crime'. We let groups pick their own research questions, drawing upon their own research backgrounds and inspired by the educational workshops and guest lectures during the hackathon. Teams could communicate with one another and the organisers via a dedicated Slack through the two days.

Data

To assist the teams in answering their research questions we provided a series of datasets via an open Dropbox link. These data were collated and processed specifically for usage during the hackathon, and offered a unique opportunity for participants to produce novel findings. The datasets are summarized in Table 1. When appropriate, we also provided participants with access to the R script used to pre-process the data, so that attendees could check or amend the raw data for their own purposes.

Crowding measures A key challenge was to ensure that participants had access open measures of 'crowding' which could then be matched with open data on crime and safety. We sourced our measures for crowding from two datasets: Retrosheet and Safegraph. Retrosheet is an open online database containing game logs on historical Major League Baseball games. These logs include relevant information including: date of the game, time of the game, the away team, the home team, the host stadium, the winner, the length of the game, and perhaps most importantly, the attendance figure for the game. The game attendance figure thus provides a measure of crowding around a specific stadium on a specific day. Although game logs go back to 1871, we subsetting the 2018 simplicity and to ensure a match with the open crime described in the next section.

Safegraph is a company which provides geospatial data on Points of Interest and measures of footfall, primarily in the United States and Canada. Subsets of their *neighbourhood patterns* dataset are openly available, containing information on footfall counts at census block group level in June, 2020. A number of measures of footfall are available based on individual 'pings' from mobile phone devices, including the number of pings by day and by hour for the whole month (i.e. 720 measurements) in each census block group. We subsetting this data to Los Angeles for computational ease, and to ensure a match with the crime data.

Crime measures To facilitate a match between the measures of 'crowds' and crime and safety, we provided attendees with access to open police-recorded crime data. Open police data in the United States is unique in terms of its spatial and temporal accuracy. Two sources were used. First, geocoded police-recorded crime data in Los Angeles for the month of June, 2020. This could be aggregated to the census block group level and merged with the Safegraph data on footfall. Second, a subset of police-recorded crime data for 2018 obtained using the `crimedata` R package (Ashby, 2018, 2019). This dataset included crime data on 16 major cities in the United States, and thus provided reasonable coverage of Major League Baseball match host cities in the Retrosheet data.

Table 1: Summary of data provided to participants

Description	Source
Police-recorded crime incidents during June 2020 in Los Angeles	Los Angeles open data archives.
Police-recorded crime incidents in 16 major US cities during 2018.	R package crimedata (Ashby, 2018).
Major League Baseball game logs for 2018.	Retrosheet website.
Stadium coordinates for Major League Baseball stadiums in US crime data cities.	Open Street Map API.
Shapefile of census block groups in California.	Government open data portal.
List of census block groups nested within Los Angeles.	Government open data portal.
Footfall counts in California during June 2020 at census block group level.	Safegraph website.

Supplementary data A number of small supplementary data files were provided to participants to provide support. These included (1) a file containing the latitude-longitude coordinates of the host baseball stadiums in the Retrosheet data; (2) a shapefile of census block groups in California; (3) a list of census block group identification codes for the city of Los Angeles.

Research outputs

Each team approached the research component of the hackathon uniquely. This demonstrated the benefits of choosing teams based on the diversity of research backgrounds, and highlighted the depth and complexities of the topic. We showcase the outputs of each team in turn.

Team 1

Need to get slides.

Team 2

The second team focused on crime during match days for the LA Dodgers. They hypothesised that, not only would attendance correlate with crime, but that the *result* of the game (i.e. whether the LA Dodgers won or not) would factor into this relationship. The team also expected this relationship to vary by crime type, reflecting the unique opportunity structures of different offences. For instance, a an LA Dodgers loss may result in higher rates of disorderly conduct or assault, whereas crimes like motor vehicle theft may be unaffected by the game result.

Findings suggested that more crime occurred in and around the LA Dodgers stadium on match days compared to non-match days, particularly for assault and disorder. There was some evidence to suggest that crime was *higher* on game days, although this relationship was not clear. Overall, higher game attendance was associated with less crime, both in and around the stadium, but also in Los Angeles as a whole. This suggests that, although games appear to elevate crime in the area, the crowding brought about by high attendance may act as a deterrent.

Team 3

Team 3 used the Safegraph data to explore the spatial and temporal relationship between vehicle crime and footfall in Los Angeles. The team hypothesised that higher footfall (i.e. crowding) would be associated with elevated vehicle crime due to the anonymity of crowds. Specifically, the team expected a lagged relationship: that an increase in crowds during time point t would be associated with an increase in vehicle crime during $t+1$.

The team found a temporal variation in vehicles crimes, and there was some evidence to suggest an association with crowding. In a typical day during June, vehicle crime counts peaked during the afternoon rush hour and remained high into the evening (see Figure 1). The team found a positive, statistically significant relationship between footfall counts and crowding, providing support for their hypothesis. However, marginal effects models suggested that the nature of this relationship (strength and direction) varied depending on the time of day.

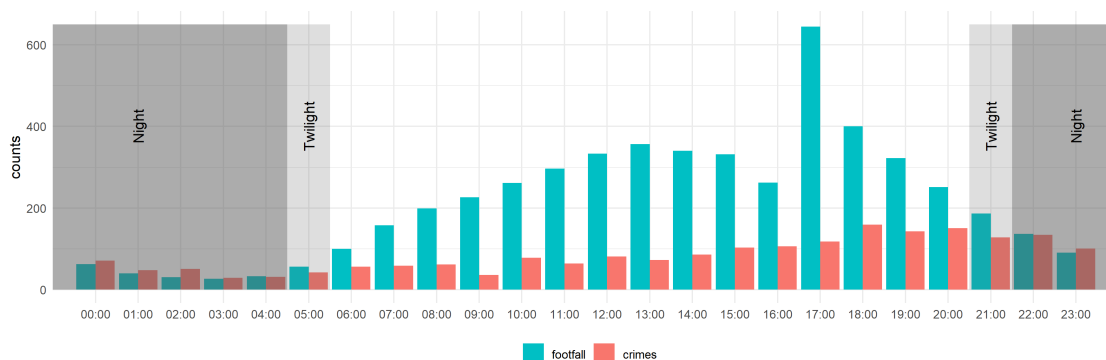


Figure 1: Example visual produced by Team 3 during the hackathon using Safegraph footfall data and open police-recorded crime data on vehicle crimes.

Discussion

- A key outcome from the hackathon was the discussion and ideas that came from these preliminary analyses.
- One thing that is yet to be exploited is the crowd *density* measure rather than simply a measure of people counts in an area.
- There is important theoretical justification for this.

Conclusions

In this paper we have reported on the design and outputs from a community-orientated open research format to tackle research questions on crowds, safety and crime. Contributions from the event are twofold. First, participants actively contributed to the field’s scientific evidence-base. Using novel data sources from Retrosheet and Safegraph, teams produced theoretically-driven findings which described and explained the temporal and spatial relationships between crowding and police-recorded crime. Teams intend to advance these analyses for publication in crime science and geography journals. Secondly, the hackathon-style event showcases the benefits and importance of a community-driven, open research format. In reporting the event design here, we hope to provide a template and inspire future events to examine topics in criminology, policing and urban analytics.

References

- Ashby, M. P. J. (2018). *Studying crime and place with the crime open database*. SocArXiv pre-print. <https://doi.org/10.31235/osf.io/9y7qz>
- Ashby, M. P. J. (2019). Studying crime and place with the crime open database: Social and behavioural sciences. *Research Data Journal for the Humanities and Social Sciences*, 4(1), 65–80.
- Ceccato, V., Cats, O., & Wang, Q. (2015). The geography of pickpocketing at bus stops: An analysis of grid cells. In *Safety and security in transit environments* (pp. 76–98). Springer.

- Cozens, P. M. (2011). Urban planning and environmental criminology: Towards a new perspective for safer cities. *Planning Practice and Research*, 26(4), 481–508.
- Engelniederhammer, A., Papastefanou, G., & Xiang, L. (2019). Crowding density in urban environment and its effects on emotional responding of pedestrians: Using wearable device technology with sensors capturing proximity and psychophysiological emotion responses while walking in the street. *Journal of Human Behavior in the Social Environment*, 29(5), 630–646.
- Freedman, J. L. (1975). *Crowding and behavior*. WH Freedman.
- Halford, E., Dixon, A., Farrell, G., Malleson, N., & Tilley, N. (2020). Crime and coronavirus: Social distancing, lockdown, and the mobility elasticity of crime. *Crime Science*, 9(1), 1–12.
- Irvin-Erickson, Y., & La Vigne, N. (2015). A spatio-temporal analysis of crime at washington, dc metro rail: Stations’ crime-generating and crime-attracting characteristics as transportation nodes and places. *Crime Science*, 4(1), 1–13.
- Kurland, J., Johnson, S. D., & Tilley, N. (2014). Offenses around stadiums: A natural experiment on crime attraction and generation. *Journal of Research in Crime and Delinquency*, 51(1), 5–28.
- Levine, N., Wachs, M., & Shirazi, E. (1986). Crime at bus stops: A study of environmental factors. *Journal of Architectural and Planning Research*, 339–361.
- Macintyre, S., & Homel, R. (1997). Danger on the dance floor: A study of interior design, crowding and aggression in nightclubs. *Policing for Prevention: Reducing Crime, Public Intoxication and Injury*, 7(1), 91–113.
- Newton, A. D. (2004). Crime on public transport: ‘Static’ and ‘Non-static’ (Moving) crime events. *Western Criminology Review*, 5(3).
- Stokols, D. (1972). On the distinction between density and crowding: Some implications for future research. *Psychological Review*, 79(3), 275.
- Tchinda, P. E., & Kim, S.-N. (2020). The paradox of “eyes on the street”: Pedestrian density and fear of crime in yaoundé, cameroon. *Sustainability*, 12(13), 5300.