

# Crowding and crime: a case study of Major League Baseball games

Samuel Langton (University of Leeds)

Reka Solymosi (University of Manchester)

## 1 Background

The relationship between crowding and crime in urban environments has been acknowledged for some time (Freedman, 1975). Although the association can vary by crime type, existing evidence suggests that the more ‘strangers’ present in an area the more crime tends to occur (Jarrell & Howsen, 1990). Overcrowded areas, such as those around public transport stops, are thought to facilitate deviant behavior (Loukaitou-Sideris, 1999) by providing opportunities for crime masked by a sea of anonymity. Studies have also found that cities hosting sports events such as baseball games can experience crime spikes (Mares & Blackburn, 2019). This is consistent with the idea that ambient populations are a fundamental part of explaining temporal variations in crime (Hipp et al., 2019). Here, we demonstrate a way of examining the relationship between crowds and crime by examining crowd attendance at Major League Baseball (MLB) match days using open data sources.

## 2 Data

### 2.1 Crime data

The crime data for this study is retrieved from an open database on police-recorded crime in the United States. Specifically, we make use of the `crimedata` package (Ashby, 2018, 2019) in R (R Core Team, 2020) which compiles geocoded crime records from recent years across 16 major cities.<sup>1</sup> For the purposes of this study, we first collated data on all crime types recorded during the year 2018.

### 2.2 Game logs

To obtain information about attendance at MLB games, we make use of data hosted on the *Retrosheet* website (<https://www.retrosheet.org/>) which compiles statistics about MLB games in Game Logs on an annual basis. From these we could obtain the date of matches, the names of home and away teams and a unique identifier for the stadium. This identifier could be matched with information (‘parkcodes’) from *Retrosheet* to retrieve the name of the stadium which hosted the match. A total of 11 host stadiums were identified for 2018 which had corresponding crime data for that city.

### 2.3 Open Street Map API

The geocoded location of each stadium was obtained computationally using the `tmtools` package (Tennekes, 2020) to query the Open Street Map API. We obtained data on building footprints within a 1-mile radius of each MLB stadium using the `osmdata` (Padgham et al., 2017) to query the API. Using this, we could adjust the match attendance figures for the built environment to create a measure of crowd density (persons per square meter), on the assumption that people entering the 1-mile buffer to attend the match would be channeled through areas that had not been built on. As a demonstration, Figure 1 visualizes the data on building footprints around the LA Dodger stadium, along with the approximate location of police-recorded crimes occurring within the 1-mile buffer zone.

---

<sup>1</sup>Austin, Boston, Chicago, Detroit, Fort Worth, Kansas City, Los Angeles, Louisville, Mesa, Nashville, New York, San Francisco, Seattle, St Louis, Tucson and Virginia Beach.

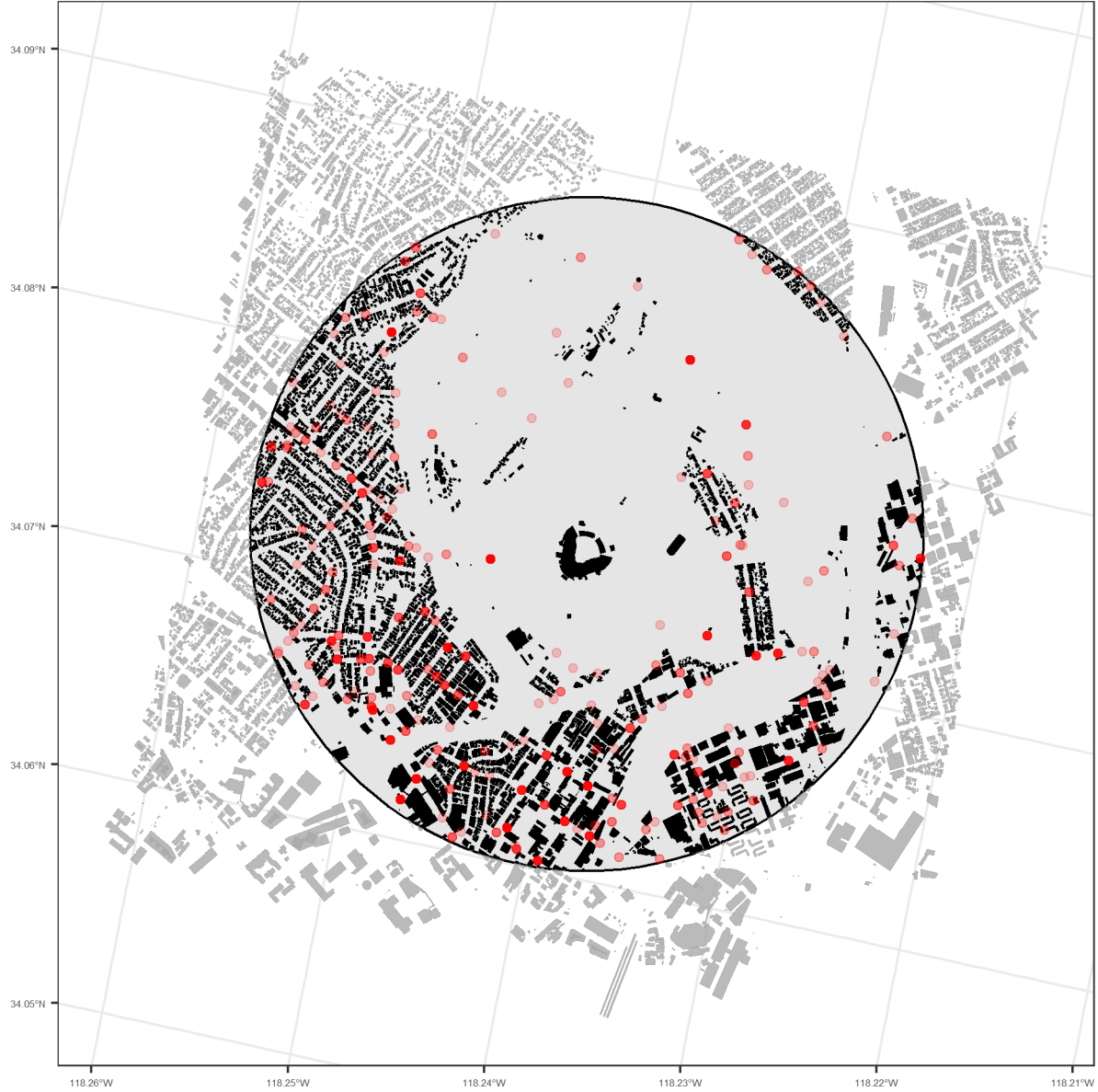


Figure 1: Building footprints around the LA Dodger stadium including 1-mile buffer and geocoded crime locations on match days during 2018. Transparency added to deonate crime locations that overlap due to police geocoding.

### 3 Preliminary findings

Overall findings suggest that there is indeed a relationship between crowd density and crime (see Figure 2). In other words, as the number of people per square meter increases, the volume of crime goes up. That said, this association appears to vary by stadium (see Figure 3). Moreover, for some stadiums (e.g. Safeco Field) the relationship flips when using crime rates, calculated by adjusting the crime count for the number of attendees (see Figure 3). In which case, as attendance increases (and in turn, crowd density) the number of crimes per person actually falls. Here, it is worth noting that both the crowd density and crime rate measures are computed using the attendance figure.

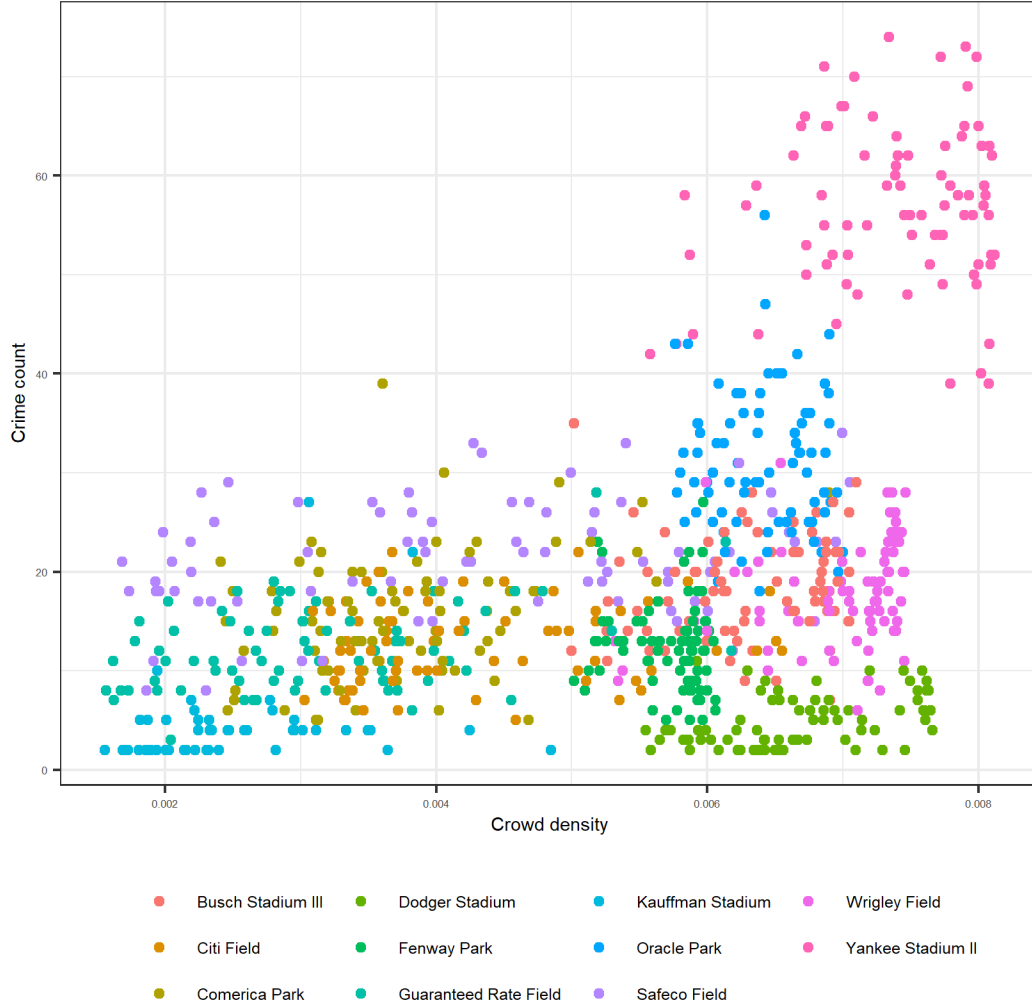


Figure 2: Relationship between crowd density and crime counts for Major League Baseball match days (Spearman's rank 0.42  $p < 0.001$ ).

## 4 Discussion

This short study has provided a demonstration of how to use open data to examine the relationship between crowding and crime at sports events. Using a measure of crowd density based on attendance per square meter of free space, we found preliminary evidence to suggest that there is a relationship between crowding and crime. The variation by stadium indicates that other factors in the urban environment might be determining criminal behavior, or mediating the relationship between crowding and crime (e.g. crowd control, alcohol sales, time of game). Future research might consider comparing games at different stadiums (thus different building footprints) that had a similar attendance figure. This would help unpick the relationship between crowd density and crime. All code to replicate this study can be found online ([https://github.com/langtonhugh/crowds\\_crime](https://github.com/langtonhugh/crowds_crime)).

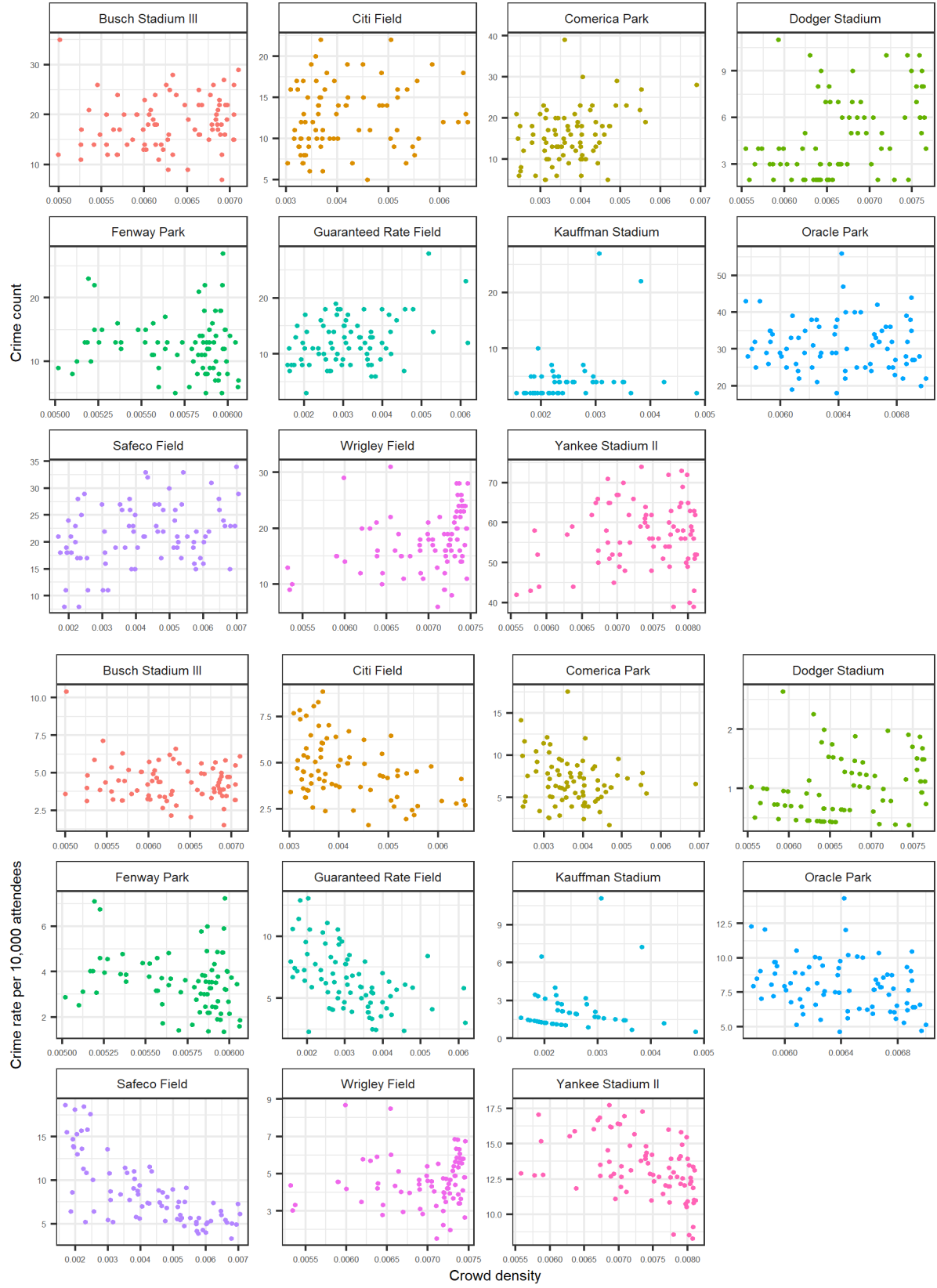


Figure 3: Relationship between crowd density and crime counts (top pane) and rates (bottom pane) by stadium for Major League Baseball match days.

## 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.
- Freedman, J. L. (1975). *Crowding and behavior*. WH Freedman.
- Hipp, J. R., Bates, C., Lichman, M., & Smyth, P. (2019). Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Quarterly*, 36(4), 718–748.
- Jarrell, S., & Howsen, R. M. (1990). Transient crowding and crime: The more “strangers” in an area, the more crime except for murder, assault and rape. *American Journal of Economics and Sociology*, 49(4), 483–494.
- Loukaitou-Sideris, A. (1999). Hot spots of bus stop crime: The importance of environmental attributes. *Journal of the American Planning Association*, 65(4), 395–411.
- Mares, D., & Blackburn, E. (2019). Major league baseball and crime: Opportunity, spatial patterns, and team rivalry at st. Louis cardinal games. *Journal of Sports Economics*, 20(7), 875–902.
- Padgham, M., Rudis, B., Lovelace, R., & Salmon, M. (2017). Osmdata. *The Journal of Open Source Software*, 2(14). <https://doi.org/10.21105/joss.00305>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Tennekes, M. (2020). *Tmaptools: Thematic map tools*. <https://CRAN.R-project.org/package=tmaptools>