Studying Crime and Place with the Crime Open Database

Social and Behavioural Scienes

M. P. J. Ashby
Lecturer in Crime Science, Department of Security and Crime Science,
University College London, UK
matthew.ashby@ucl.ac.uk

Abstract

The study of spatial and temporal crime patterns is important for both academic understanding of crime-generating processes and for policies aimed at reducing crime. However, studying crime and place is often made more difficult by restrictions on access to appropriate crime data. This means understanding of many spatio-temporal crime patterns are limited to data from a single geographic setting, and there are few attempts at replication. This article introduces the Crime Open Database (CODE), a database of 16 million offenses from 10 of the largest United States cities over 11 years and more than 60 offense types. Open crime data were obtained from each city, having been published in multiple incompatible formats. The data were processed to harmonize geographic co-ordinates, dates and times, offense categories and location types, as well as adding census and other geographic identifiers. The resulting database allows the wider study of spatio-temporal patterns of crime across multiple US cities, allowing greater understanding of variations in the relationships between crime and place across different settings, as well as facilitating replication of research.

Keywords

open data – crime – crime data – police data – geography of crime – crime and place

- Related data set "Crime Open Database (CODE)" with URL https://www.osf.io/zyaqn/files in repository "Open Science Framework"

1.   Introduction

Research on crime and place has become a substantial field within the broader interdisciplinary study of crime and its effects. Numerous studies have demonstrated crime is concentrated in a few places while no crime at all occurs in most others (Johnson, 2010) and that patterns of land use can generate crime hotspots (Wilcox & Eck, 2011). These findings have implications for policy and practice, be it in decisions about where to deploy police officers (Ratcliffe, Taniguchi, Groff, & Wood, 2011) or the management of potentially criminogenic land uses (Eck & Wartell, 1999). Temporal crime patterns, and particularly interactions between spatial and temporal patterns, are also important (Tompson & Coupe, 2018). For example, many types of crime exhibit seasonal patterns (Andresen & Malleson, 2013), while a daily 'wave' of street crime hits cities in the late afternoon (Wheeler & Haberman, 2018).

It is important to research individual crime types, rather than considering crime as an undifferentiated whole (Cornish & Smith, 2012). Many spatial and temporal patterns differ across crime types, with (for example) residential burglary typically peaking in the afternoon while commercial burglaries most often happen overnight (Butler, 1994).

Also important is the choice of units of analysis. Although there is no single best analytical scale (Hipp, 2007), recent scholars have suggested that in spatial analysis "smaller is better" (Oberwittler & Wikström, 2009) because many environmental influences on crime operate at small spatial scales (Steenbeek & Weisburd, 2016). The same is true in temporal analysis, since some of the largest fluctuations in crime happen over short timescales such as hours and days (Felson & Poulsen, 2003).

Conducting analysis at these important micro scales requires micro-level data. For example, knowing on which street a crime occurred is more useful than only knowing the crime occurred in a certain city. Similarly, knowing the date and time of an offence is more valuable than knowing only the month or year. This makes many sources of official crime data – such as the Federal Bureau of Investigation (FBI) Uniform Crime Reports (UCR) – of only limited value in studying crime and place.

Many researchers have obtained micro-scale crime data by forming partnerships with a single local police agency, typically that covering the city in which the research team is based. The historically secretive nature of many police organizations and the importance of maintaining privacy for crime victims means such relationships often require substantial investment of time to establish trust between researchers and police leaders. When police do share

data, it is typically on condition that it be stored securely and not shared outside the original research team (Kane, 2007).

Single-city studies based on confidential data have generated valuable results on crime and place, but have at least two limitations. Firstly, the importance of environmental characteristics in generating spatio-temporal crime patterns calls into question the generalizability of studies conducted in a single setting. For example, Los Angeles is 54% larger in area than New York City but has less than half the population, creating a very different urban fabric. Social context differs, too: the median income of San Francisco, for example, is three times that of Detroit (US Census Bureau, 2018).

Secondly, studies using confidential data can be difficult to replicate, since other research teams typically cannot access the data or other materials. Replication is important both to identify errors (e.g. of method) in existing studies and to understand the extent and nature of outcome variation in different environments (Lösel, 2017; Weisburd & Taxman, 2000). However, in practice very few criminological studies are subject to replication (McNeeley & Warner, 2015).

Some phenomena in crime and place have been studied across multiple settings, such as the elevated short-term 'near repeat' risk of burglary victimization experienced by households surrounding a previous offence location (Grove, Farrell, Farrington, & Johnson, 2012). Nevertheless, the extensive data-access negotiations typically preceding such research mean that making use of results from different cities is typically only possible after several years. This may be too late for policy makers, who must typically make decisions over shorter time scales.

Given these limitations, a source of crime data that allows both simultaneous study of multiple geographic settings and sharing data with other researchers, would potentially benefit research into crime and place. In recent years, police agencies have begun to release crime data as 'open data', typically as part of efforts to increase transparency (Caplan, Rosenblat, & Boyd, 2015). Open data is information that is freely available for anyone to use and share with others (Open Knowledge International, 2018). Open crime data are popular both with citizens and journalists (Stoneman, 2015), but also provide a potential source of information for a multi-city crime dataset.

Since they are released for non-research purposes, open crime data are not necessarily immediately useful for criminological studies across cities. For example, crime types are typically published using categories that are bespoke to a particular city or state. More prosaically, elements such as dates or geographic co-ordinates are often provided in different, incompatible formats. As such, further processing is necessary to maximize the usefulness of these datasets.

The remainder of this article introduces the Crime Open Database (CODE), a dataset combining harmonized open crime data from large cities in the United States that can be used for multi-city studies of crime and place.

2.   Methods

2.1.   *Data Sources*
The municipal websites of the 50 largest cities in the United States by population were checked for a source of open crime data for inclusion in CODE. A city was included in the database if data were available:
1.   for at least four consecutive years (to allow longitudinal analysis),
2.   at the offense or incident level (rather than, for example, aggregated offense counts),
3.   including either geographic co-ordinates or offense addresses (to allow analysis at micro scales),
4.   including the date and (optionally) the time at which the offense occurred,
5.   with sufficient information about the offense type to map to a harmonized set of crime categories.

The United States was chosen because its laws (e.g. on personal data) and well-established open-data movement (Ubaldi, 2013) mean US cities may be more likely to release open data with sufficient detail in comparison to other countries. For example, although the UK Home Office releases a national open crime dataset at www.police.uk, it lacks detail in that it does not specify the date on which an offense occurred and uses very broad offense types.

Ten cities met these criteria and are shown in Figure 1. Links to the original data sources and brief explanations of why other cities could not be included are provided on the project website. It is hoped that more cities will be included in future as they begin to release suitable open crime data.

2.2.   *Data Processing*
Data from each city were processed to allow their use in multi-site research. Dates were converted to a consistent format and local co-ordinate reference systems were converted to latitude/longitude pairs using the WGS 84 co-ordinate system.

Geocoded offense locations were included in the open data published by all-but-three cities. For Fort Worth, Kansas City and Louisville, addresses were geocoded using the US Census Geocoder accessed via the MapChi package in R (Welgus, 2018), with residual ungeocoded locations resolved using the commercial Geocod.io geocoding API.

FIGURE 1   Cities included in the Crime Open Database

To ensure the privacy of crime victims, cities releasing open crime data typi-
cally introduce small inaccuracies into the addresses or co-ordinates reported
for each crime. This process differs between cities, but CODE offense locations
can be considered to be accurate to the nearest hundred block on a particular
street.[1] This spatial inaccuracy prevents analysis at the individual address level
using CODE data. However, many studies do not require address-level analy-
sis, particularly since the inaccuracy of locations reported by victims means
apparent address-level data accuracy is sometimes spurious (Ratcliffe, 2002).
Each CODE offense record is supplemented with the state, county, census tract,
block group and census block for the geocoded location.

    To produce a harmonized typology of offense types, offense categories in
each city were manually mapped to the 52 offense types included in the FBI
National Incident-Based Reporting System (NIBRS) using the offense defi-
nitions published in Federal Bureau of Investigation (2018). Since the NIBRS
categories do not distinguish between commercial and personal robbery, or
between residential and non-residential burglary, these offenses were further
distinguished where city crime categories allowed. The final categorization is
shown in Table 1.

---

1 Depending on the configuration of each city, a hundred block (i.e. the length of a street be-
   tween buildings with numbers 100 apart, e.g. between 101 and 201 Main Street) may not al-
   ways be the same as a physical block.

TABLE 1　Crime-type data available for each CODE city

Category	Type	CH	DT	FW	KC	LA	LO	NY	SF	TS	VB
Crimes against people											
homicide offenses	murder and nonnegligent manslaughter	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
	negligent manslaughter	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
kidnapping/abduction		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sex offenses	rape (except statutory rape)	×	✓	×	✓	✓	✓	✓	✓	✓	✓
	sodomy	×	×	×	✓	✓	✓	✓	✓	×	✓
	sexual assault with an object	×	✓	×	✓	✓	×	✓	✓	×	✓
	fondling	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
	sexual assault by penetration (including rape)[a]	✓	×	×	×	×	×	×	×	×	×
assault offenses	aggravated assault	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	simple assault	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	intimidation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sex offenses, nonforcible	incest	✓	×	×	✓	✓	✓	×	✓	×	✓
	statutory rape	×	✓	×	✓	✓	✓	×	✓	✓	✓
	other sex offenses[a]	×	×	×	×	✓	×	×	×	✓	×
human trafficking	human trafficking, commercial sex acts	✓	×	✓	×	✓	×	×	✓	✓	×
	human trafficking, involuntary servitude	✓	×	✓	×	✓	×	×	×	✓	×

Category	Type	CH	DT	FW	KC	LA	LO	NY	SF	TS	VB
robbery	personal robbery	✓	✓	✓	×	✓	✓	✓	✓	✓	×
	commercial robbery	✓	✓	✓	×	✓	✓	✓	✓	✓	×
	other robbery[a]	✓	✓	✓	×	✓	✓	×	×	✓	✓
Crimes against property											
arson		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
extortion/blackmail		✓	✓	✓	✓	✓	✓	✓	✓	×	✓
burglary/breaking & entering	residential burglary / breaking & entering	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	non-residential burglary / breaking & entering	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	other burglary / breaking & entering[a]	×	✓	✓	✓	×	×	×	×	✓	×
larceny/theft offenses	pocket-picking	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	purse-snatching	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	shoplifting	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	theft from building	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	theft from coin-operated machine or device	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	theft from motor vehicle (except theft of motor vehicle parts or accessories)	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
	theft of motor vehicle parts or accessories	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	all other larceny[a]	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
	other theft[a]	✓	×	×	×	×	×	×	×	×	×

TABLE 1   Crime-type data available for each CODE city (*Cont.*)

Category	Type	CH	DT	FW	KC	LA	LO	NY	SF	TS	VB
motor vehicle theft		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
counterfeiting / forgery		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
fraud offenses (except counterfeiting/forgery and bad checks)	false pretenses / swindle / confidence game	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	credit card / automated teller machine fraud	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	impersonation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	welfare fraud	×	✓	✓	✓	×	✓	×	✓	×	✓
	wire fraud	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	other fraud[a]	×	×	×	×	×	×	✓	×	✓	×
embezzlement		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
stolen property offenses		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
destruction / damage / vandalism of property (except arson)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
bribery		✓	✓	✓	×	✓	✓	✓	✓	×	✓
bad checks (except counterfeit checks or forged checks)		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Crimes against society											
drug/narcotic offenses	drug / narcotic violations	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	drug equipment violations	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
pornography / obscene material		✓	✓	✓	✓	✓	✓	✓	✓	×	✓

Category	Type	CH	DT	FW	KC	LA	LO	NY	SF	TS	VB
gambling offenses	betting / wagering	×	✓	✓	×	×	×	✓	✓	×	✓
	operating / promoting / assisting gambling	✓	✓	✓	✓	×	✓	✓	✓	×	✓
	gambling equipment violations	×	×	✓	×	×	✓	✓	✓	×	×
	sports tampering	×	×	×	✓	×	×	×	×	×	×
prostitution offenses	prostitution	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	assisting or promoting prostitution	✓	✓	✓	✓	✓	✓	✓	✓	×	✓
	purchasing prostitution	✓	×	✓	✓	✓	×	✓	✓	×	×
	weapon law violations	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	curfew / loitering / vagrancy violations	×	×	✓	✓	✓	✓	✓	✓	✓	✓
	disorderly conduct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	driving under the influence	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	drunkenness (except driving under the influence)	×	✓	✓	✓	×	✓	×	✓	✓	✓
	family offenses, nonviolent	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	liquor law violations (except driving under the influence and drunkenness)	✓	✓	✓	✓	×	✓	×	✓	✓	✓
	peeping tom	×	×	✓	✓	✓	✓	×	✓	✓	✓
	trespass of real property	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
All other offenses	all other offenses[a]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[a] This crime type may contain crimes that could not be correctly allocated to more-specific types within the same category. Crime counts in this category should not be compared across cities.

TABLE 2   CODE location categories.

Location category	Location types
commercial	construction, factory / warehouse, finance, gas station, office, storage, other commercial
education	child care, college, school
government	government
healthcare	healthcare
hotel	hotel
leisure	bar / club, entertainment, restaurant
open space	green space, vehicle parking, other open space
other	abandoned, religious, other
residence	residence
retail	auto dealer, convenience store, gas station, mall, supermarket, other retail
street	street
transportation	airport, bus, rapid transit, other transportation
utility	utility
vehicle	commercial vehicle, private vehicle

TABLE 3   Years of data available for each CODE city.

City	2007	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17
Chicago, IL	√	√	√	√	√	√	√	√	√	√	√
Detroit, MI			√	√	√	√	√	√	√	√	√
Fort Worth, TX	√	√	√	√	√	√	√	√	√	√	√
Kansas City, MO			√	√	√	√	√	√	√	√	√
Los Angeles, CA				√	√	√	√	√	√	√	√
Louisville, KY			√	√	√	√	√	√	√	√	√
New York, NY	√	√	√	√	√	√	√	√	√	√	√
San Francisco, CA	√	√	√	√	√	√	√	√	√	√	√
Tucson, AZ			√	√	√	√	√	√	√	√	√
Virginia Beach, VA							√	√	√	√	√

The inclusion of offense types varies between cities. While some offenses (such as aggravated assault) are recorded in every CODE city, others (such as gambling offenses) are included only by some. Cities may exclude offense types for several reasons, perhaps because an activity is not criminal in a particular state or to preserve victim privacy (particularly for sex offenses). While these exclusions limit the analysis possible using CODE data, this is an inevitable consequence of using open data. Nevertheless, multi-site studies remain possible for almost all crime types, albeit using fewer cities for some offenses.

It can be useful to disaggregate crimes according to the type of place in which they occur. For example, spatio-temporal patterns of assaults committed in bars may vary substantially from those in private homes. To facilitate such analysis, CODE includes fields for harmonized location category and location type (Table 2) for the five cities – Chicago, Fort Worth, Los Angeles, Louisville and New York – that include location descriptions or categories in their open crime data.

The project website at www.osf.io/zyaqn contains technical documentation for the database, lookup tables for offense and location types, summaries of quality-assurance tests undertaken on the data and links to the R code used to process it.

3.   Data

– **Crime Open Database (CODE) deposited at Open Science Framework –**
  URL: https://www.osf.io/zyaqn
– **Crime Open Database**
  – Data files – URL: https://www.osf.io/zyaqn/files
– **Temporal coverage:** 2007-2017

After being processed into a consistent format, data from each city were filtered to allow analysis across cities. Incidents occurring before 1 January 2007 (or the first complete year for which each city has published data – see Table 3), located outside the city boundary, or which were non-criminal in nature (e.g. traffic collisions) were excluded. Comparing the final dataset to the data published by each city demonstrates the importance of this processing and filtering. In Tucson, for example, 45% of records in the original data were for non-criminal matters such as traffic incidents. Meanwhile, some cities' data included a small number of historical offenses that occurred decades ago but have only been recently recorded.

The resulting dataset includes a single record for each criminal offense recorded by police in CODE cities. CODE data are released under a Creative Commons (CC) Attribution 4.0 International license in compressed comma-separated values files. Two files are provided for each year. The first contains only those variables that have been harmonized across cities, including geographic co-ordinates, census identifiers, dates and offense types. These 'core' files are likely to be sufficient for almost-all research purposes. If researchers wish to make use of unharmonized fields in the data published by specific cities, these can be found (together with the core fields) in the second, 'extended' data files.

CODE data can also be accessed using the crimedata package for the R statistical programming language, which can be downloaded from the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/package=crimedata

4. Research Potential

CODE contains records for more than 16 million offenses across 11 years and more than 60 offense types. The size, detail and spatial variety of this dataset opens up new opportunities for research into crime and place.

CODE is likely to be useful in three ways. Firstly, the large sample size available in this dataset will allow more-detailed study of some aspects of crime and place. For example, some environmental characteristics may be relatively rare in a single city, potentially leading to high uncertainty about relationships between those characteristics and crime that can be reduced by including data from multiple cities. Secondly, using data from different cities allows understanding of a wider range of potentially criminogenic environments that may not be present in every city. For example, inland cities typically do not have beaches, while many cities are not adjacent to military bases, making it impossible for single-city studies to explore the influences of the full range of settings that might be encountered elsewhere. Thirdly, it is possible to study differences in the effects of similar environments on crime in different cities. For example, is the relationship between high schools and crime in the surrounding area the same in different cities? CODE is likely to be particularly useful in improving the generalizability of crime-and-place research.

While CODE offers new opportunities for studying crime and place, care must be taken to understand its limitations. The database is primarily intended for the study of spatio-temporal patterns of crime at micro scales, and in particular how those patterns differ across geographic settings. It is not likely

to be useful for research that uses the city as the unit of analysis, such as studies analyzing associations between crimes and various city-level social policies. Official sources of crime data such as the UCR may be more appropriate in such cases. Due to the variations between laws and practices between cities, CODE is also unlikely to be useful for studies that attempt to study `crime' as an undifferentiated whole. While there are overlaps between spatio-temporal patterns for different types of crime (Newton, 2015), combining for analysis such crimes as (for example) welfare fraud and driving under the influence is unlikely to be productive.

Since CODE is based on police crime records, it inherits the limitations of those data. The principal limitation is the "dark figure" of crime (Grünhut, 1951, p. 149), the difference between the number of crimes that occur and those that are reported to police (for a discussion, see Maguire & McVie, 2017). The proportion of crimes appearing in police data varies by type, for example almost all vehicle thefts are reported while most assaults without injury are not (Gove, Hughes, & Geerken, 1985; Tarling & Morris, 2010). A particular problem is those crimes (such as drugs possession) that usually only become known to police if officers catch the offender in the act. For these "intangible" offenses (Chappell & Walsh, 1974, p. 494), police records reflect patterns of police activity more than the underlying distribution of offenses.

Although the dark figure is an acknowledged problem in criminological research, official figures are often the only available source of micro-scale spatio-temporal crime data (Brantingham & Brantingham, 1975). Attempts have been made to use other sources of data to capture unreported offenses (Solymosi & Bowers, 2018), but these have their own limitations. In many cases, police records may be the best available source of crime data.

Open crime data have further limitations. While a researcher with an ongoing relationship with a particular police agency may be able to request access to particular data fields, researchers using open data (particularly from multiple cities) are typically limited to whatever information the agency as decided to release. In addition, not having a direct relationship with those who generate and process crime data may make it harder for open-data researchers to understand any specific issues relating to data from a particular city.

Balancing the impact of these limitations versus the benefits of CODE data outlined above is a decision best taken in consideration of the individual needs of particular research studies. While all data sources have limitations, the emerging crime-and-place research that has used open data (Solymosi, Ashby, Cohen, & Sidebottom, 2017; Tompson, Johnson, Ashby, Perkins, & Edwards, 2014) suggests open crime datasets can complement existing alternatives. This

is likely to be increasingly the case as more cities begin to release open crime data, particularly if that allows for comparison of places in different countries.

References

Andresen, M. A., & Malleson, N. (2013). Crime seasonality and its variations across space. *Applied Geography*, 43, 25–35. www.doi.org/10.1016/j.apgeog.2013.06.007.

Brantingham, P. J., & Brantingham, P. L. (1975). The spatial patterning of burglary. *The Howard Journal of Criminal Justice*, 14(2), 11–23. www.doi.org/10.1111/j.1468-2311.1975.tb00297.x.

Butler, G. (1994). Commercial burglary: What offenders say. In M. Gill (Ed.), *Crime at work: Studies in security and crime prevention* (Vol. 1, pp. 29–41). Leicester: Perpetuity.

Caplan, R., Rosenblat, A., & Boyd, D. (2015). Open data, the criminal justice system, and the police data initiative. In *Data and civil rights: A new era of policing and justice*. Washington, DC: Data; Society. Retrieved from www.datacivilrights.org/pubs/2015-1027/Open_Data_Police_Data_Initiative.pdf.

Chappell, D., & Walsh, M. (1974). Receiving stolen property: The need for systematic inquiry into the fencing process. *Criminology*, 11(4), 484–497. www.doi.org/10.1111/j.1745-9125.1974.tb00609.x.

Cornish, D. B., & Smith, M. J. (2012). On being crime specific: Observations on the career of R V G Clarke. In N. Tilley & G. Farrell (Eds.), *The reasoning criminologist: Essays in honour of Ronald V Clarke* (pp. 30–45). Abingdon: Routledge.

Eck, J. E., & Wartell, J. (1999). *Reducing crime and drug dealing by improving place management: A randomized experiment*. Washington, DC: National Institute of Justice.

Federal Bureau of Investigation. (2018). *2019 National Incident-Based Reporting System user manual*. Washington, DC: US Department of Justice. Retrieved from https://ucr.fbi.gov/nibrs/nibrs-user-manual.

Felson, M., & Poulsen, E. (2003). Simple indicators of crime by time of day. *International Journal of Forecasting*, 19(4), 595–601. www.doi.org/10.1016/S0169-2070(03)00093-1.

Gove, W. R., Hughes, M., & Geerken, M. (1985). Are Uniform Crime Reports a valid indicator of the index crimes? An affirmative answer with minor qualifications. *Criminology*, 23(3), 451–501.

Grove, L. E., Farrell, G., Farrington, D. P., & Johnson, S. D. (2012). *Preventing repeat victimization: A systematic review*. Stockholm: Swedish National Council for Crime Prevention.

Grünhut, M. (1951). Statistics in criminology. *Journal of the Royal Statistical Society Series A (General)*, 114(2), 139–162.

Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review*, 72(5), 659–680. www.doi.org/10.1177/000312240707200501.

Johnson, S. D. (2010). A brief history of the analysis of crime concentration. *European Journal of Applied Mathematics*, 21(4/5), 349–370. www.doi.org/10.1017/S0956792510000082.

Kane, R. J. (2007). Collect and release data on coercive police actions. *Criminology and Public Policy*, 6(4), 773–780. www.doi.org/10.1111/j.1745-9133.2007.00485.x.

Lösel, F. (2017). Evidence comes by replication, but needs differentiation: the reproducibility issue in science and its relevance for criminology. *Journal of Experimental Criminology*, 1–22. www.doi.org/10.1007/s11292-017-9297-z.

Maguire, M., & McVie, S. (2017). Crime data and criminal statistics: A critical reflection. In A. Liebling, S. Maruna, & L. McAra (Eds.), *The Oxford handbook of criminology* (6th ed., pp. 163–189). Oxford: Oxford University Press.

McNeeley, S., & Warner, J. J. (2015). Replication in criminology: A necessary practice. *European Journal of Criminology*, 12(5), 581–597. www.doi.org/10.1177/1477370815578197.

Newton, A. (2015). Crime and the NTE: Multi-classification crime (MCC) hot spots in time and space. *Crime Science*, 4(1), 30. www.doi.org/10.1186/s40163-015-0040-7.

Oberwittler, D., & Wikström, P. O. H. (2009). Why small is better: Advancing the study of the role of behavioral contexts in crime causation. In D. Weisburd, W. Bernasco, & G. J. N. Bruinsma (Eds.), *Putting crime in its place: Units of analysis in geographic criminology* (pp. 35–59). New York: Springer.

Open Knowledge International. (2018). What is open data? Retrieved from www.opendatahandbook.org/guide/en/what-is-open-data/.

Ratcliffe, J. H. (2002). Damned if you don't, damned if you do: Crime mapping and its implications in the real world. *Policing and Society*, 12(3), 211–225. www.doi.org/10.1080/1043946290018463.

Ratcliffe, J. H., Taniguchi, T., Groff, E. R., & Wood, J. D. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology*, 49(3), 795–831. www.doi.org/10.1111/j.1745-9125.2011.00240.x.

Solymosi, R., & Bowers, K. J. (2018). The role of innovative data collection methods in advancing criminological understanding. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford handbook of environmental criminology* (pp. 210–237). Oxford: Oxford University Press. www.doi.org/10.1093/oxfordhb/9780190279707.013.35.

Solymosi, R., Ashby, M. P. J., Cohen, T., & Sidebottom, A. (2017). Alternative denominators in transport crime rates. Open Science Framework preprint. www.doi.org/10.17605/OSF.IO/5QV38.

Steenbeek, W., & Weisburd, D. (2016). Where the action is in crime? An examination of variability of crime across different spatial units in the Hague, 2001–2009. *Journal of Quantitative Criminology*, 32(3), 449–469. www.doi.org/10.1007/s10940-015-9276-3.

Stoneman, J. (2015). *Does open data need journalism?* Oxford: Reuters Institute for the Study of Journalism. Retrieved from http://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism.

Tarling, R., & Morris, K. (2010). Reporting crime to the police. *British Journal of Criminology*, 50(March), 474–490. www.doi.org/10.1093/bjc/azq011.

Tompson, L., & Coupe, T. (2018). Time and opportunity. In G. J. N. Bruinsma & S. D. Johnson (Eds.), *The Oxford handbook of environmental criminology* (pp. 695–719). Oxford: Oxford University Press. www.doi.org/10.1093/oxfordhb/9780190279707.013.19.

Tompson, L., Johnson, S. D., Ashby, M. P. J., Perkins, C., & Edwards, P. (2014). UK open source crime data: Accuracy and possibilities for research. *Cartography and Geographic Information Science*, 42(2), 97–111. www.doi.org/10.1080/15230406.2014.972456.

Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives*. Paris: OECD Publishing. www.doi.org/10.1787/19934351.

US Census Bureau. (2018). American community survey 5-year estimates. Retrieved from www.census.gov/programs-surveys/acs/.

Weisburd, D., & Taxman, F. S. (2000). Developing a multicenter randomized trial in criminology: The case of HIDTA. *Journal of Quantitative Criminology*, 16(3), 315–340. www.doi.org/10.1023/A:1007574906103.

Welgus, D. (2018). *MapChi: Tools for making maps of Chicago*. Retrieved from www.github.com/dmwelgus/MapChi.

Wheeler, A. P., & Haberman, C. P. (2018). *Modeling the spatial patterns of intra-day crime trends*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3136030.

Wilcox, P., & Eck, J. E. (2011). Criminology of the unpopular: Implications for policy aimed at payday lending facilities. *Criminology and Public Policy*, 10(2), 473–482. www.doi.org/10.1111/j.1745-9133.2011.00721.x.