Incorporating time into studying crime and place

Matthew P J Ashby1

1 Jill Dando Institute of Security and Crime Science, University College London

# 

# 1 Introduction

Understanding how crime varies over time is just as important as understanding how it varies between places. Very few places are crime hotspots all the time – business districts might be hotspots of pickpocketing in the daytime but deserted at night, while a nearby entertainment district may be quiet in the daytime but a violence hotspot at night.

Temporal variation in crime frequency is widespread. There are often long-term trends, such as the crime drop that occurred in many countries from the 1990s onward. Trends also exist at smaller spatial scales: micro places such as street segments have different trajectories of crime, which drive crime trends at larger spatial scales (Andresen, Curman, & Linning, 2017).

[Figure 1 about here]

Non-trend temporal variations can be divided into two categories: cyclical variations (including seasonal patterns) and short-term fluctuations (including random ‘noise’). Figure 1 shows how weekly counts of various crimes in Washington, DC, differ from the mean weekly count. Some types (such as assault with a dangerous weapon) exhibit strong seasonal patterns, while others (such as robbery) do not. There are also substantial short-term variations.

Focusing only on where crime happens, rather than analyzing when and where together, leads to crime-prevention failures. Many police (the author included) will have been sent to patrol a hotspot at a time when no crimes had ever occurred there, because a crime analyst had considered where crimes concentrated, but not when. Equally, a theory explaining why crime concentrates at places but not at times is likely to lack predictive power.

# 2 Why studying temporal data can be challenging

Temporal data can be challenging for several reasons. Dates can be stored in many formats: 17 January 1981 can be 1/17/1981 in the United States, 17.01.81 in Germany and 17/01/1981 in France. Many formats are ambiguous – does 10/02 represent 10 February or 2 October, or even October 2002? Is 3.45 in the early hours of the morning or mid-afternoon? Dates and times should be stored in an unambiguous format such as 1981-01-17 13:43, with hours represented using the 24-hour clock. Date-format problems can be hard to detect: even if an analyst knows what a value represents, collaborators or software may make different assumptions. For example, Microsoft Excel guesses (often incorrectly) whether numbers in a cell represent a date.

## 2.1 Imprecision in temporal data

[Figure 2 about here]

Temporal data is often imprecise. Some values (such as when a call to 911 was made) are known to the second, but others (such as the time of death of a decomposed body) might only be known to the nearest day or week. A common example is the times of crimes reported by victims. A person who has been robbed on the street might know it happened ‘about a half-hour ago’, leading either victims or officers to round-off the time at which they record a crime as occurring. Figure 2 shows that robberies in Washington, DC are much more likely to be reported as occurring on the hour than would be expected by chance. An extreme version of this problem is when a victim doesn’t know when a crime occurred, only the earliest and latest possible times – dealing with this using aoristic analysis is the subject of Chapter 14.2.

## 2.2 Choosing a temporal scale

One reason temporal data are challenging is the conceptual nature of time. Time is linear: the instant in time at 14:19:34 on Wednesday 4 November 2020 will never be repeated in the Gregorian Calendar. But time also has overlapping cycles: a 4th of November every year, a 4th day every month, Wednesday every week, 14:19 every day and 19 minutes past every hour. There are also exceptions to these rhythms: February in some years has 29 days (and the year 366 days); some days are 23-hours long and others 25-hours long due to daylight savings changes.

[Figure 3 about here]

These complications make it important to choose an appropriate temporal scale for analysis. Figure 3 shows the frequency of thefts from automobiles in Washington, DC, at different temporal scales. Panel A shows daily frequency. Panel B shows the same data, but aggregated to monthly counts. This is a common way to present trends, but monthly counts should almost always be avoided. Not all months are the same length (March is 11% longer than February), so it’s likely Panel B would show month-by-month variations even if crime occurred at a constant rate. Months also don’t include equal numbers of each weekday, which is important when (as shown in Panel C) crime concentrates on certain days. The only reason to use monthly counts of crime is if data are only available in that format (e.g. administrative data published monthly). Otherwise, use weeks or multiples of weeks as a unit of analysis.

Panel C shows thefts on each day of the week. There is a weekly pattern, with thefts peaking on Fridays and Saturdays. This is obscured by noise in Panel A, and hidden within monthly counts in Panel B – it is often useful to visualize patterns at multiple temporal scales.

Panel D shows thefts by hour of the day. The horizontal axis begins at 05:00, following Felson and Poulsen (2003), since human activity patterns rarely begin or end neatly at midnight. Thefts have two peaks, one around midday and one around 18:00, with thefts gradually tailing off either side. With a plot axis beginning at midnight, the increase before midday would be obvious but the gradual decrease after midnight could have been missed, being split between the two sides of the plot.

Panel E also shows hourly crime counts, but for hours of the week rather than the day. Combining units of analysis in this way can be useful to identify how (for example) hourly patterns vary between days. This additional detail, however, comes at the cost of statistical noise in the data being more apparent, making some trends more difficult to identify. Choosing the appropriate trade off between noise and temporal detail requires judgment, which can often be helped by experimenting with multiple temporal scales to find the best balance between factors. It is also possible to use smoothing techniques (as in Panel A) such as moving averages or LOESS smoothing (Buskirk, Willoughby, & Tomazic, 2013) to identify patterns amid noise.

The appropriate temporal scale also depends on the question being asked. Identifying changes in crime over several years has little need for hourly data, but might need weekly counts or counts for 28-day periods (to avoid the problems of monthly periods). It is often worthwhile considering a smaller temporal unit than initial instinct suggests, since (as can be seen by comparing panels C and E a scale that is too large masks important variations. This does not mean always plotting data at the smallest available scale – the non-nested cycles seen in temporal data mean choosing a scale is more complicated than moving up and down the ‘cone of resolution’ used for spatial data (Brantingham, Dyreson, & Brantingham, 1976). For example, Panel A shows the data in its rawest form, but arguably conveys less information that the other panels that show aggregated data.

## 2.3 Handling temporal ordering and autocorrelation

One feature of temporal data is that each data point has an inherent serial order relative to other points. So whereas (for example) survey respondents are typically independent of one another, repeated measures of a quantity at different time points are not. One consequence of this is temporal (or serial) autocorrelation: observations of data taken close together (e.g. the count of crime last week and the count this week) are usually more similar than observations taken further apart (e.g. the count last week compared to the count six months ago). As Panel A of Figure 4 shows, high weekly counts of vehicle theft are more likely to be followed by other high counts, and low counts more likely to be followed by other low counts (in this case, because of seasonality).

[Figure 4 about here]

Temporal autocorrelation is challenging because many statistical techniques produce misleading results if the assumption that observations are independent is violated. In particular, -tests and linear regressions may be inaccurate (De Boef, 2004). Autocorrelation can be identified using tests such as the Ljung-Box test (Ljung & Box, 1978), which shows that there is significant autocorrelation in the weekly counts of vehicle theft in Panel A of Figure 4 (, ). Temporal autocorrelation can also be analyzed graphically using an autocorrelation function (ACF) plot (Box & Jenkins, 1970). In Panel B of Figure 4 there is significant autocorrelation present (i.e. ACF values outside the confidence interval) for the first several lags.

Autocorrelation can be handled by using statistical methods appropriate for temporal data. Autoregressive integrated moving average (ARIMA) models are widely implemented in statistical software and commonly used in studying temporal crime patterns. These models can handle many common features of temporal data, including seasonal variation.

# 3 Incorporating time into spatial analysis

[Figure 5 about here]

Time is often relevant to spatial analysis because spatial crime patterns frequently vary over time. Figure 5 shows the frequency of assaults with a dangerous weapon throughout the week in two districts of Washington, DC. The First District includes Capitol Hill, half the National Mall and the Southwest Federal Center government buildings, and some residential areas. The Seventh District is almost entirely residential except for a large military base. Figure 5 shows similarities between weekly patterns of serious assaults in both districts, but with notable differences. Dangerous-weapon assaults in the First District are clustered on Friday and Saturday evenings, whereas in the Seventh District offenses peak the in late evening on every weeknight, with some additional peaks during afternoons.

Variations in temporal patterns also exists at smaller geographies. Figure 6 shows hotspots of assaults with dangerous weapons in the area around the White House at different times[[1]](#footnote-1). Farragut Square (marked A on Figure 6) is a significant hotspot only during the daytime, unsurprisingly since it hosts many large businesses that generate daytime foot traffic. The Penn Quarter/Chinatown neighborhood (marked B), which hosts many restaurants and entertainment venues, is a hotspot both in the daytime and the evening. Meanwhile, Dupont Circle (marked C) is a hotspot only overnight.

[Figure 6 about here]

Ignoring the temporal dimension of crime and place risks committing an ecological fallacy. This can be avoided by using an appropriate theoretical perspective to consider the role of time in a given situation. The routine activities approach can identify how the aggregate activity patterns of different actors in a crime event might influence the spatio-temporal distribution of crime (Eck & Madensen, 2015). For example, crime in a city park may require analysis that distinguishes crimes during daylight and at night, or crimes on weekdays and at weekends. The multi-faceted nature of temporal crime variation (apparent in Figure 3) means it is often fruitful to explore empirically multiple ways of ‘slicing’ data to understand how time influences spatial crime patterns.

# 4 References

Andresen, M. A., Curman, A. S., & Linning, S. J. (2017). The trajectories of crime at places: Understanding the patterns of disaggregated crime types. *Journal of Quantitative Criminology*, *33*, 427–449. <https://doi.org/10.1007/s10940-016-9301-1>

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Brantingham, P. J., Dyreson, D. A., & Brantingham, P. L. (1976). Crime seen through a cone of resolution. *American Behavioral Scientist*, *20*(2), 261–273. <https://doi.org/10.1177/000276427602000207>

Buskirk, T. D., Willoughby, L. M., & Tomazic, T. T. (2013). Nonparametric statistical techniques. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology: Vol. 2: Statistical analysis*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199934898.013.0006>

De Boef, S. (2004). Serial correlation. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The Sage encyclopedia of social science research methods*. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781412950589.n915>

Eck, J. E., & Madensen, T. D. (2015). Meaningfully and artfully reinterpreting crime for useful science: An essay on the value of building with simple theory. In M. A. Andresen & G. Farrell (Eds.), *The criminal act: The role and influence of routine activity theory* (pp. 5–18). Basingstoke, Hampshire: Palgrave Macmillan. <https://doi.org/10.1057/9781137391322.0006>

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

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. <https://doi.org/10.1093/biomet/65.2.297>

*Figure* *1:*. Difference from mean frequency crimes in Washington, DC, 2016

*Figure* *2:*. Burglaries reported by minute of the day in Washington, DC, 2016

*Figure* *3:*. Thefts from automobiles in Washington, DC, 2016, using different temporal units of analysis

*Figure* *4:*. Motor vehicle thefts by week in Washington, DC, 2016

*Figure* *5:*. Assaults with dangerous weapons by hour of the week in Washington, DC, 2016

*Figure* *6:*. Hotspots of assault with a dangerous weapon at different times of day in Washington, DC, 2016

1. Hotspots were identified by counting assaults in each 500-metre-wide cell in a hexagonal grid, then calculating Getis–Ord (Getis & Ord, 1992) for each cell. scores greater than (equivalent to ) are treated as significant. [↑](#footnote-ref-1)