Safer Traffic Recovery from the Pandemic in London – Spatiotemporal Data Mining of Car Crashes

Kejiang Qian[1] · Yijing Li[1]

Abstract

In the aim to provide evidence for deployment policies towards post-pandemic safer recovery from COVID-19, this study investigated the spatiotemporal patterns of age-involved car crashes and affecting factors, upon answering two main research questions: (1) "What are spatiotemporal patterns of car crashes and any observed changes in two years, 2019 and 2020, in London, and waht were the influential factors for these crashes?"; (2) "What are spatiotemporal patterns of casualty by age, and how do people's daily activities affect the patterns pre- and during the pandemic"? Three approaches, spatial analysis (network Kernel Density Estimation, NetKDE), factor analysis, and spatiotemporal data mining (tensor decomposition), had been implemented to identify the temporal patterns of car crashes, detect hot spots, and to understand the effect on citizens' daily activity on crash patterns pre- and during the pandemic. It had been found from the study that car crashes mainly clustered in the central part of London, especially busier areas around denser hubs of point-of-interest (POIs); the POIs, as an indicator for citizens' daily activities and travel behaviours, can be of help to analyze their relationships with crash patterns, upon further assessment on interactions through the geographical detector; the casualty patterns varied by age group, with distinctive relationships between POIs and crash pattern for corresponding age group categorised. In all, the paper introduced new approaches for an in-depth analysis of car crashes and their casualty patterns in London to support London's safer recovery from the pandemic by improving road safety.

Keywords Car crash · Casualty · Point-of-interest · Spatiotemporal pattern · Tensor decomposition

✉ Yijing Li
yijing.li@kcl.ac.uk

Kejiang Qian
kejiang.qian@kcl.ac.uk

[1] Centre for Urban Science and Progress London, Department of Informatics, King's College London, London WC2B 4BW, UK

🙂 Springer

Introduction

Car crashes have long been recognised as exerting tremendous effects on economic development and human well-beings among societies. Most people involved in road traffic crashes will get injured or even die from the accidents, where the number of crash deaths peaked at 1.35 million around the world in 2016, with more than half of fatalities aged between 15 and 44 (World Health Organization (WHO, 2018). London also engenders such challenges and has around 80,000 reported collisions during 2017–2019 before the pandemic in 2020, when people aged between 25 and 59 accounting for the most casualties, whilst those over 60 have a higher fatality rate. Such an issue was predicted to get worse alongside the rapid population growth, as the Greater London Authority (GLA) projected a potential population of around 11 million in 2050 based on 8.17 million citizens censused in 2011 (GLA, 2020). On the other hand, the COVID-19 pandemic exerted significant influences on road traffic activities with the figure that, UK Office for National Statistics (ONS, 2020) reported up to 40% of workers work from home during the pandemic, dramatically reducing traffic demand and expected to have a long-term effect on people's daily travel behaviour (Katrakazas et al., 2020) and transport mode preference (Batty, 2020). For example, Almansouri & Cetin, (2021) had found that commuters' frequency of visits, and their preference of distances to shopping malls had changed significantly upon the pandemic. It had been further testified by Yasin et al., (2021) and Amberber et al., (2021)'s work that, there was a significant reduction in traffic collisions during the pandemic due to changes in traffic volumes, driver behaviours, and travel behaviours. To investigate such changes in depth, spatial analysis is recognised as one of the most efficient methods (Kaya et al., 2019). Therefore, studies on road safety by varied daily patterns during the target period were considered to be in a better position formulating car crash prevention deployment strategies.

Empirical studies on crash distributions identified the spatial and temporal variations that, crashes tend to occur more around working places rather than residential places (Levine et al., 1995), and had been thought to be affected by multiple factors, e.g., daily human activities and build environment factors (Ziakopoulos & Yannis, 2020). Land use types play an important role in estimating car crashes (Pulugurtha et al., 2013) and teen crash frequency (Mathew et al., 2022) at aggregated levels, which can provide valuable insights for design intervention measures, in the aim to reduce the risk of crashes among vulnerable population. However, land use characteristics are of less use in analyzing car crashes occurred in highly urbanized and cosmopolitan cities or regions, where a high-degreed mosaic of land uses and spatial heterogeneity are expected. To address such issue at finer spatial scale in urbanised regions, POIs had been selected as the proxy for land use types, in order analyse the relationships between urban forms and collision patterns (Worachairungreung et al., 2021). Besides, POIs could be more powerful in explaining the varied adversity from collisions by age group. Because collision-vulnerable population were normally defined by

age groups, considering the concentration of certain age groups in proximity to POIs, the differences in mobility patterns among age groups, and the varied travel behaviour and mode preferences by age group (Pour et al., 2017). However, there were still few research examining the interactive effect between multiple POIs and traffic crashes by age groups, making the topic worthy of further exploration.

Being motivated by the incentive to improve road safety intelligently and support traffic recovery, this study aims to identify spatiotemporal patterns of age-categorised traffic crashes, and further to investigate the influential factors towards severer crashes pre- and during pandemic. Besides, to gain richer knowledge of traffic crashes' variations among age groups in London pre- and during pandemic, two main objectives are expected to be arrived at:

The spatiotemporal pattern and varied recognition of car crashes in two different representative years, and the underlying influences from POI factors.
The spatiotemporal casualty patterns by age group, as well as the effects from people's varied daily activities pre- and during the pandemic.

To realise such objectives, this study will (1) analyse the spatiotemporal patterns of car crashes considering injury severity; (2)quantify their relationships with human activities presented by POI categoriesthrough the geographical detector; (3) extract the spatial and temporal crashes patterns of victims by age groups; and (4) measure the relationship between their daily activities and corresponding patterns of car crashes based on the road safety data in 2019 and 2020, respectively.

Literature Review

The COVID-19 pandemic had a significant impact on road safety and vehicle collision patterns, such as changes in their frequency and severity. Specifically, there is a notable decline in the number of vehicle collisions and associated injuries due to reduced traffic volume and remote replacements for social activities (Cappellari & Weber, 2022; Sutherland et al., 2020). On the contrary, the severity of injuries and deaths occurred from those collisions has relatively increased (Yasin et al., 2021). However, the nexus among different factors, especially those related to casualties need to be further analyzed to fully understand these changes. This understanding is crucial for policymakers and experts to develop effective strategies for road safety.

Their pattern changes and factors can be investigated through various methods. Two potential approaches are spatial analysis and factor analysis.. The spatial analysis techniques include examining car crashes' first-order effect, i.e., the hot spot of car crashes detection, and the second-order effect, i.e., spatial clusters detection, with Kernel Density Estimation (KDE) as a widely adopted approach. In order to map traffic crashes, Okabe et al.'s proposes network-based KDE (NetKDE) estimating crashes along the roadway and the development of the SANET toolbox, followed by an increasing number of studies on car crashes using the NetKDE approach (Xie & Yan, 2008; Zahran et al., 2021). Besides, Koloushani et al., (2022) found that hot spots of young-driver-involved crashes data supported the justification that NetKDE

has a better performance than planar KDE when receiving localized focuses. The second-order effect had traditionally been measured by widely recognised Moran's *I* index, in the aim to estimate the strength of spatial autocorrelation and discover statistically significant risk clusters. For example, Iyanda, (2019)'s work explores the distribution of accident severity, and Shabanikiya et al., (2020)'s work identifies high-risk areas of pedestrian crashes with children's involvement.

To investigate the influences on traffic crashes from environmental factors, factor analysis had been utilised as the mainstream with car crashes being treated as point events, and the count data models as the basic methods embracing spatial heterogeneity existence. For example, the spatial zero-inflated negative binomial model (Champahom et al., 2020), and the fixed bandwidth geographically and temporally weighted ordered logistic regression model (Chen et al., 2022). However, these models failed to capture the dynamic interactions between selected factors and traffic crashes. To fulfill such a gap, geographical detector proposed by Wang et al., (2016) had been considered in light of its being excellent at measuring the influence of factors on geographical phenomena based on spatial stratified heterogeneity analysis, and received extensive applications in varied research fields, like housing price (Wang et al., 2017) and air pollution (Ding et al., 2019), as well as car crash analysis (Zhang et al., 2020) assessing injury factors' influences on casualties considering their mutual interaction, but still leaving the research on interactions of factors for road safety less explored.

In parallel, spatiotemporal data mining technique has been taken as a more efficient method exploring patterns of spatiotemporal data (Han et al., 2012), in comparison with traditional spatial analysis methods. This data mining rooted method on spatiotemporal data mining and urban computing (Zheng et al., 2014a) was represented by the most favourite approach, tensor decomposition (Kolda & Bader, 2009). Tensor decomposition has been advantaged by its modelling multidimensional data and dealing with data sparsity issues, with example applications in analysing human mobility from OD-matrix data and decomposing the spatiotemporal patterns of human mobility through measuring the variability of multidimensional mobility patterns (Sun & Axhausen, 2016; Yao et al., 2015); Zheng et al. (2014b) also applied it onto New York City's sparse noise data analysis, and tried to recover the noise level by combining with POIs, road network, and social media data. However, there was still very limited literature applying it in traffic crash research.

In a nutshell, the limitations of empirical studies (summarized in Table 1) rendered this study the opportunity for improvements in: combining spatiotemporal data mining with spatial analysis and factor analysis techniques, onto age-categorised traffic crashes analysis from three dimensions which are time, space, and casualty severity by age group, to highlight the age-bounded effects from traffic crashes.

Data and Methodology

London, as the capital of the United Kingdom, has been chosen as the study area with its 4,835 Lower Super Output Area (LSOA) units (GLA, 2014), where each has an average population of 1500 (Green et al., 2011).

Table 1 Limitations of current studies

Authors & Year	Regions	Methodologies	Limitations
Okabe et al., (2006); Blazquez and Celis, (2013); Koloushani et al., (2022)	Tokyo, Japan Santiago, Chile Florida, USA	Planar KDE & NetKDE	This method has worse performance than NetKDE when there is a localized focus
Wang et al., (2016); Ding et al., (2019); Zhang et al., (2020)	China China Shenzhen, China	Geographical detector	There are rare studies of crash factor analysis by considering specific risky indicators by age breakdown, the relationship of the POIs, and the interactive effects of factors
Zheng et al., (2014b); Yao et al., (2015); Sun & Axhausen, (2016)	New York, USA; Massachusetts, USA; Singapore	Tensor decomposition	This method is very useful, but there was only very limited literature applying it in traffic crash research

Data Acquisition

Road safety data of London had been collected from the GOV.UK in point format at LSOA scale from 01 January to 30 December 2019 and from 01 January to 29 December 2020, with detailed information on recorded car crashes at 29,023 in 2019 and 23,551 in 2020 respectively (Table 2). Among the records, 84.47% in 2019 and 85.3% in 2020 were slight, but 0.57% in 2019 and 0.53% in 2020 were fatal. In reference to Department for Transport (DfT)'s 11 breakdown categories of the casualties by age group, this study restructured the age groups into seven research conveniences as listed in Tables 2 and 3.

It could be read from Table 2 that, the car crashes and causality in 2020 had decreased from the level in 2019 during the pandemic, but with a similar distribution pattern across age group breakdowns; the frequency of all crashes or by varied severity in 2020 had been decreased proportionally by age group against 2019. Since this study is aiming to explore the pattern and mechanism of safer traffic recovery from pandemics, 2019 and 2020 as the representative normal and abnormal years, respectively, would be sufficient to suggest the patterns, hence shed light on the traffic recovery strategies upon recovery from 2020 back to normal. So the following analysis will focus on the case study in 2019 and 2020 to explore the patterns in detail.

POI datasets in this study had been collected from the Ordnance Survey, consisting of geospatial information on all London business, entertainment, education, etc. in 2019, with ten categories: accommodation, eating and drinking, attractions, commercial services, sport and entertainment, education and health, public infrastructure, transport, manufacturing and production, and retail, as summarised and represented as relevant icons in Table 3.

Table 2 Car crashes and causalities by age group (2020 vs. 2019)

Age groups	Frequency of slight crashes		Frequency of serious crashes		Frequency of fatal crashes		Frequency of all crashes	
	2020	2019	2020	2019	2020	2019	2020	2019
0–18	2108	2858	376	498	5	14	2489	3370
19–25	3525	3886	523	730	17	29	4065	4645
26–35	6087	6846	845	1048	34	34	6966	7928
36–45	3845	4786	618	770	22	21	4485	5577
46–55	2544	3331	489	588	17	16	3050	3935
56–65	1294	1679	275	331	7	14	1576	2024
Over 65	686	1131	211	375	23	38	920	1544
Total							23,551	29,023

Fatal accident refers to an accident in which at least one person is killed; other casualties (if any) may have serious or slightly injuries; serious accident means one in which at least one person is seriously injured but no person (other than a confirmed suicide) is killed; slight accident represents one in which at least one person is slightly injured but no person is killed or seriously injured. Definitions retrieved from official document at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/259012/rrcgb-quality-statement.pdf

Table 3 Statistical summary of POI categories in London

POI category	Count	Mean	Std	Min	Max
(WORK)	142,005	29.369	4685.475	0	3218
(ATTRACTION)	11,269	2.329	41.347	0	207
(EDUCATION)	11,969	2.475	8.373	0	8
(SHOP)	49,990	10.339	578.338	0	759
(RESTAURANT)	29,708	6.144	18.271	0	807
(STATION)	21,209	4.386	25.098	0	140
(ENTERTAINMENT)	12,691	2.624	4.432	0	118
(ACCOMMODATON)	1,749	0.362	2.525	0	53
(PARKING)	1,533	0.317	0.872	0	27

Severity-weighted Index

Severity-weighted index (SWI) had been calculated to indicate the seriousness of crashes rather than purely the volume of crashes, based on the combined 5–3-1 weighting system (Geurts et al., 2004). The system was developed and deployed by Flanders government of Belgium to locate dangerous accident locations. Based on sensitivity analysis, Geurts et al., (2004) proved that it is a moderate approach to highlight the importance of deadly accidents, whilst capture the overall severity of injuries resulting from collisions, with each fatal injury weighted at 5, serious injury weighted at 3, slight injury weighted at 1 in Eq. (1):

$$SWI = 5 \times x_{fatal} + 3 \times x_{serious} + x_{slight} \tag{1}$$

where x_{fatal}, $x_{serious}$ and x_{slight} respectively presents for the total number of fatal, serious, and slight traffic crashes.

Spatial Analysis

Spatial analysis of crash distributions on basis of road networks and planar maps at LSOA level, to discover the first-order and second-order effects through NetKDE and Moran's *I* statistics respectively.

Hot spots Detection Based on the NetKDE

NetKDE targeting at point events in a network space is an extension of planar KDE with its two main specialties (Xie & Yan, 2008): (1) network distance QUOTE d_{is}

from crash point QUOTE i to location QUOTE s is calculated as the shortest-path distance along the road network; and (2) the density estimator is computed per linear unit as defined below in Eq. (2):

$$\lambda(s) = \frac{1}{r} \sum_{i=1}^{n} w_i \cdot K(\frac{d_{is}}{r}) \tag{2}$$

where $\lambda(s)$ presents the density estimator at location ; is the bandwidth of the KDE, and is the weight of crash point i. K is the kernel modeled as the kernel function of distance d_{is}. The method assumes that crashes occur alongside roadways. Crashes data need to be snapped on the network within a 10-m road network distance, because previous crashes may not be recorded on road networks due to issues such as the low precision of the measurement system or car moving right after a crash.

Xie & Yan, (2008) also pointed out that the network needs to be divided into an equal-length linear unit, which may have an impact on the local variation details. So this study selects the 200-m road segment as the basic unit considering London's vast road network, in order to decrease the number of linear units so for a good balance between computational efficiency and accuracy. On the other hand, the choice of bandwidth also affects performance of the network KDE, As suggested by (Okabe & Sugihara, 2012) an optimal bandwidth should be 100–300 m would be optimal, so this study designed a bandwidth of 200 m for network KDE and Gaussian function as the kernel function. It is hypothesised that network KDE can improve the drawback of planar KDE (Xie & Yan, 2008) on over-detecting clustered point events, hence enhancing the accuracy of density estimation in intersections.

Spatial Clustering Identification using Moran's I Statistics

Spatial autocorrelation can reflect the strength of the spatial dependence between factors, so Moran's *I* statistics could be utilised to identify clusters with similar SWIs or the dispersion of those with dissimilar SWIs. The global Moran's *I* statistic, denoted as QUOTE , can be computed as in Eq. (3):

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} d_i d_j}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \sum_{i=1}^{n} d_i^2} \tag{3}$$

where is the total number of factors, and present the deviations of the -th and -th LSOA's SWI to their means: and , and is the spatial weight between factor and , which represents the relationship between research units under Queen criterion. Down to the local indicators of spatial association (LISA) measuring the degree of local spatial autocorrelation of SWIs could be expressed in Eq. (4) as:

$$LISA_i = z_i \sum_j w_{ij} z_j z_i = \frac{SWI_i - \overline{SWI}}{\sqrt{\frac{1}{n} \sum (SWI_i - \overline{SWI})^2}} \tag{4}$$

where LISA clusters denoted as High-High (HH) illustrate the high-valued target clustered with high-valued neighbours; the low-low (LL) cluster indicates for low values are clustered with each other; the high-low (HL) or low–high (LH) clusters are high values surrounded by low values and vice versa.

Geographical Detector for Factor Analysis

Upon exploring the spatial clustering of crashes, it is necessary to quantify the influence of POI factors on traffic crashes through geographical detectors, which can reflect daily human activities for our in-depth interpretation. In this study, it is assumed that if road traffic crashes were affected by POI categories, they may follow a similar spatial distribution pattern. Among the four geographical detectors, which are factor detector, interactive detector, ecological detector, and risk detector, the former two are deployed in this study to measure the nonlinear or linear relationship between factors. The factor detector is used to find the dominant factor of the geographical phenomena, while the interactive detector is applied to measure the interconnected effects between pairs of factors on crashes' distributions.

The factor detector computes the value of the power determinant (PD) through q statistics, denoting the percentage of explainability for target variable by another variable, which is similar to R-square statistics of regression models, as in Eq. (5):

$$\mathrm{PD} = 1 - \frac{1}{N\sigma^2} \sum_{h=1}^{1} N_h \sigma_h^2 \tag{5}$$

where denotes the number of LSOA units, and represent the variance of the dependent variable of the whole study area and that in the units respectively; the PD value indicates the strength of the spatially stratified heterogeneity of crash distributions and the contribution of the POIs to their spatial pattern.

The interaction detector is good at estimating the interactions between two individual factors. If the PD value of interaction is greater than the accumulation of the PD values from individual effect, a significant enhancement influence from the interaction is present. This study will build up the model with the number of POI categories in each LSOA as an explanatory variable, and numeric variables need to be stratified for detectors; whilst Jenks's natural breaks (JNB) method is applied to determine the optimised categorisation of each POI, in the criteria of minimising the variances within classes but maximising them between classes (Chen et al., 2013).

Spatiotemporal Pattern Detection using Tensor Decomposition

Tensor decomposition excels at analysing the multidimensional road safety data in time and space patterns, hence being able to decide the various crash scenarios for geographical detector methods to fit in. For example, if a crash pattern involving minors during rush hours has been identified with a high PD value between POIs of education, that indicates a scenario about students & rush hours.

Tensor Construction and Decomposition

An Nth-order tensor is a multidimensional matrix of N vector spaces, with each vector as a first-order tensor ($N=1$) and the matrix as a second-order tensor ($N=2$), but in vision to present three- or higher-dimensional data hereby with a higher order of the tensor. For example, three-dimensional data can be written in a tensor as $A \in R^{I_1 \times I_2 \times I_3}$ with three dimensions denoting I_1, I_2, and I_3. To extract and discover latent patterns from tensors, approaches such as singular vector decomposition (SVD), latent factor modeling (LFM), and principal component analysis (PCA) are required. However, such methods may cause information loss due to corrupted data, and make tensor decomposition outstanding with much better performances (Kolda & Bader, 2009). The most efficient tensor decomposition method is tucker tensor decomposition (Tucker, 1966), to decompose the tensor $A \in R^{I_1 \times I_2 \times \cdots \times I_N}$ into a core tensor $G \in R^{J_1 \times J_2 \times \cdots \times J_N}$ and a set of nonsingular matrices $A^{(1)} \in R^{I_1 \times J_1}$, $A^{(2)} \in R^{I_2 \times J_2}$, QUOTE, and $A^{(n)} \in R^{I_n \times J_n}$. It should be noted that the size of the core tensor must be defined manually, which may affect the performance of the model in Eq. (6).

$$A \cong \left[G; A^{(1)}, A^{(2)}, \cdots, A^{(n)} \right] = G \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_n A^{(n)} \tag{6}$$

where is the n-mode tensor-matrix product of a tensor with a matrix , which means multiplying a tensor by an n-mode matrix and is defined in Eq. (7):

$$\left[G \times_n A \right]_{j_1, \cdots, j_{n-1}, i, j_{n+1}, \cdots j_N} = \sum_{j_n=1}^{J_n} G_{j_1, \cdots, j_{n-1}, i, j_{n+1}, \cdots j_N} A_{i, j_n} \in R^{J_1, \cdots, J_{n-1}, I, J_{n+1} \cdots J_N} \tag{7}$$

Nonnegative Tucker Decomposition

The Tucker tensor decomposition also has a drawback on possible negative elements in its results, which deviates from the fact that no negative SWI of crashes data, hereby calls up the utilisation of nonnegative Tucker decomposition (NTD) (Shashua & Hazan, 2005) to find a nonnegative core tensor QUOTE and nonnegative matrices $A^{(1)}$, $A^{(2)}$, QUOTE $\cdots$, QUOTE $A^{(n)}$, and be modelled in Eq. (8) for optimisation:

$$\text{minimize} \frac{1}{2} \| G \times_1 A^{(1)} \times_2 A^{(2)} \cdots \times_n A^{(n)} - A \|_F^2 \tag{8}$$

subject to $G \in R^{J_1 \times J_2 \times \cdots \times J_N} \geq 0$, $A^{(n)} \in R^{I \times J_n} \geq 0 \forall n = 1, 2, 3 \cdots, N$ where $\| \bullet \|_F$ presents the Frobenius norm of a tensor $\chi \in R^{I_1 \times I_2 \times \cdots \times I_N}$ defined as:

$$\| \chi \|_F = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \cdots i_N}^2} \tag{9}$$

Since the size of the core tensor influences the capability of the NTD model, it is crucial to find an appropriate size on basis of Kullbask-Leibler (KL) divergence,

which is also known as the relative entropy and is widely used to measure the similarity between two discrete density distributions M and N (Hershey & Olsen, 2007).

$$D_{KL}(M \| N) = - \sum m(x) \log n(x) + \sum m(x) \log m(x) = H(M, N) - H(M) \quad (10)$$

where the entropy of distribution M is referred to as $H(M)$, and the cross-entropy of distributions M and N is presented as $H(M, N)$. Additionally, the following equation illustrates the measurement of the original tensor $A \in R^{i \times j \times k}$ and decomposed tensor $\hat{A} \in R^{i \times j \times k}$ based on the KL divergence: the smaller value is, the closer between two distributions; and the optimal size will be indicated by the converging point of the minimum value.

Results

Temporal Pattern of Road Traffic Crashes

The SWI index of traffic crashes (accumulative) and their temporal patterns in 2019 and 2020 are summarized in Fig. 1 (a, b, c), illustrating similar distribution patterns. The daily patterns in either time-of-day or day-of-week could be presented in Fig. 1(a), where the *y-axis* indicates day-of-week, the *x-axis* stands for time-of-day by every two hours, and the legend on the right scaled values of SWI. It is obvious that crashes occur more frequently on weekdays than on weekends; the peak hours on weekdays are routinely 8 am to 9 am and 5 pm to 7 pm; comparatively, weekends have a higher risk of traffic crashes at midnight in these two years. An exceptional observation on Wednesday afternoon with a minor peak time 3 pm to 4 pm might owe to its routinely being the sports day after 2 pm. However, there was a small difference that the maximums of the SWI during the pandemic were less than those in 2019. This phenomenon is also illustrated in Fig. 1b and c on the characteristics of crashes' breakdown by seven age groups on weekdays and weekends, respectively. Besides, Fig. 1(b) clearly shows the higher causality for 26–35 years old during peak hours on weekdays, followed by the relatively higher causality for children and youngsters (0–18 and 19–25) after school. Weekends' car crashes rarely happen at midnight, as reflected in Fig. 1(c), but more tend to accumulate among young and middle-aged groups (19–25, 26–35 and 36–45) from rush hours in the afternoon throughout to midnight.

Spatial Analysis of Road Traffic Crashes

High-risk locations of traffic crash in London pre- and during the pandemic are detected by NetKDE, as shown in Fig. 2, and colored in red. It illustrates that crashes are mainly distributed in the central part of London and extended in all directions alongside the road network, with most being located at road intersections and complex road structures such as roundabouts. Besides, there are smaller number of high-risk locations of crashes in 2020 than those in 2019, which shows a hollow distribution.
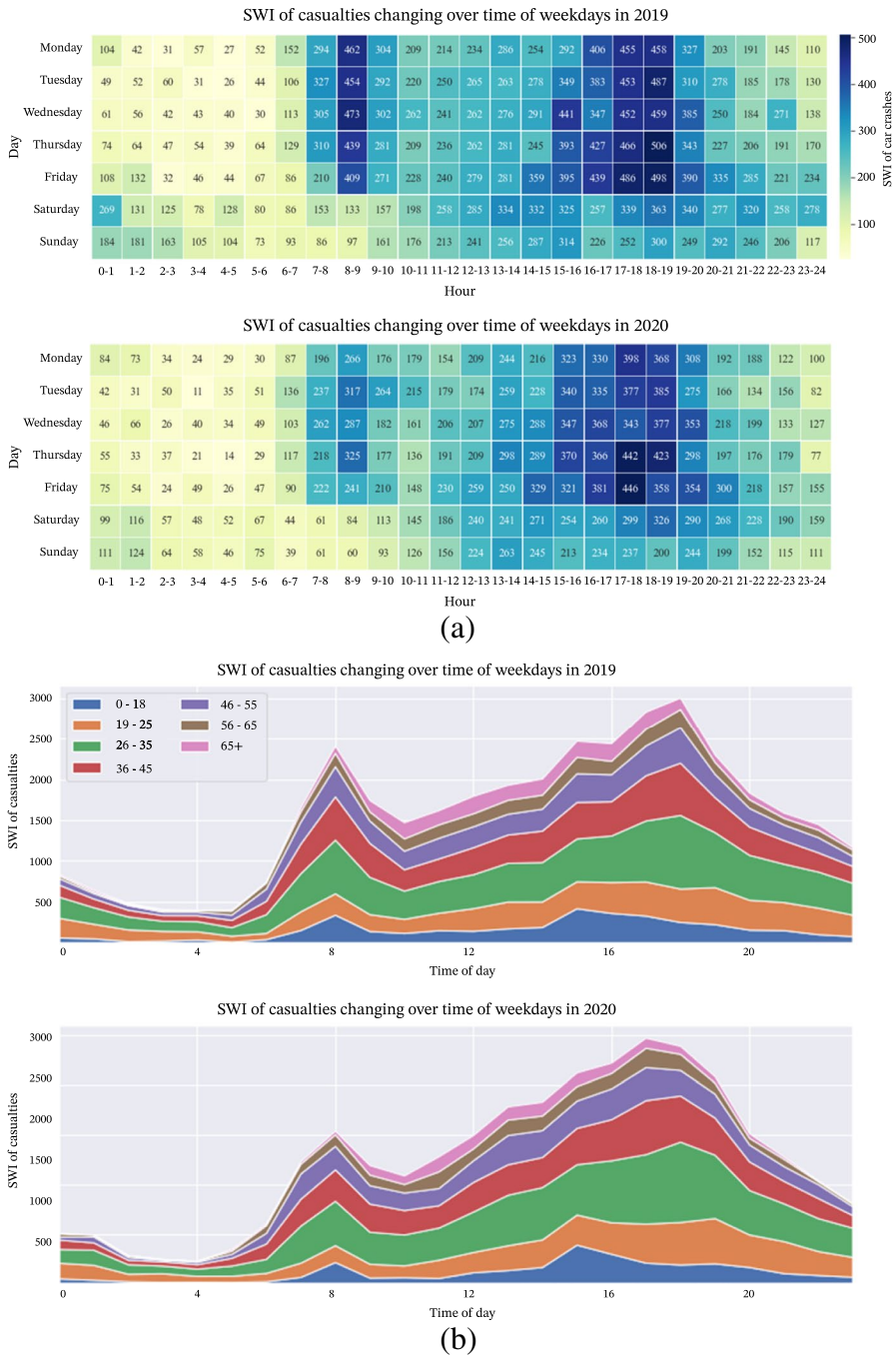
Fig. 1 Temporal patterns of car crashes. (**a**) SWI of traffic crashes in each time of a week in 2019 & 2020; (**b**) SWI index of casualties changing over time of weekdays in 2019 & 2020; (**c**) SWI index of casualties changing over time of weekends in 2019 & 2020

SWI of casualties changing over time of weekdays in 2019

SWI of casualties changing over time of weekdays in 2020

(c)

Fig. 1 (continued)


(a) 2019

(b) 2020
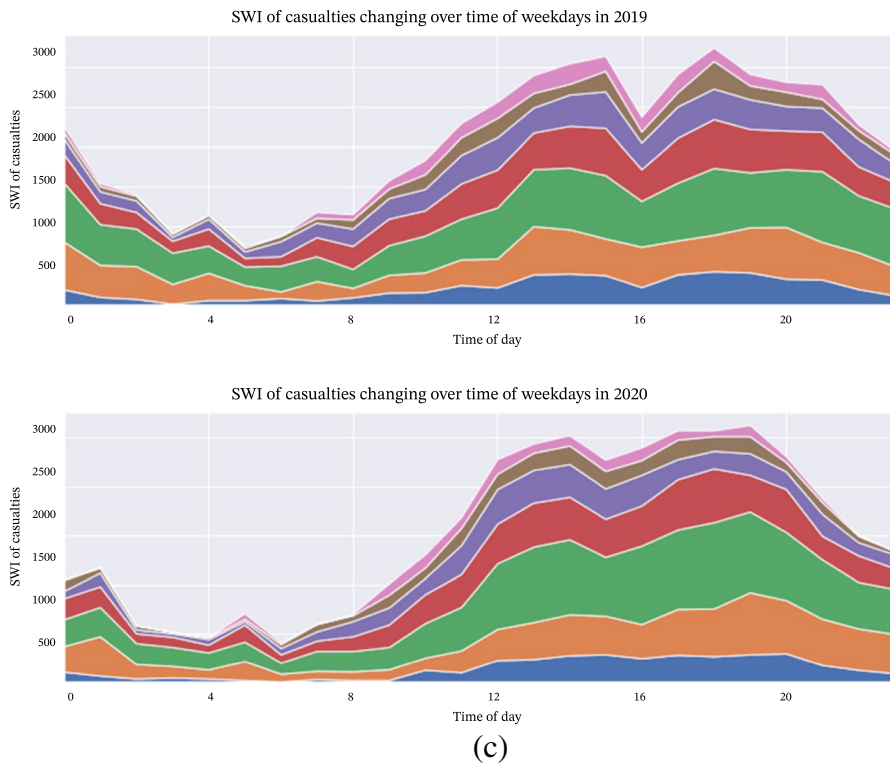
Fig. 2 Network KDE of crashes in 2019 & 2020

To further explore the riskiest roads, the top four and first densest road segments in 2019 and 2020 respectively, which are in the same density interval, had been selected in North London or nearer to central London (Fig. 3a and b). In 2019, the first road segment is Seven Sister Roads in the borough of Haringey

(a) 2019

(b) 2020
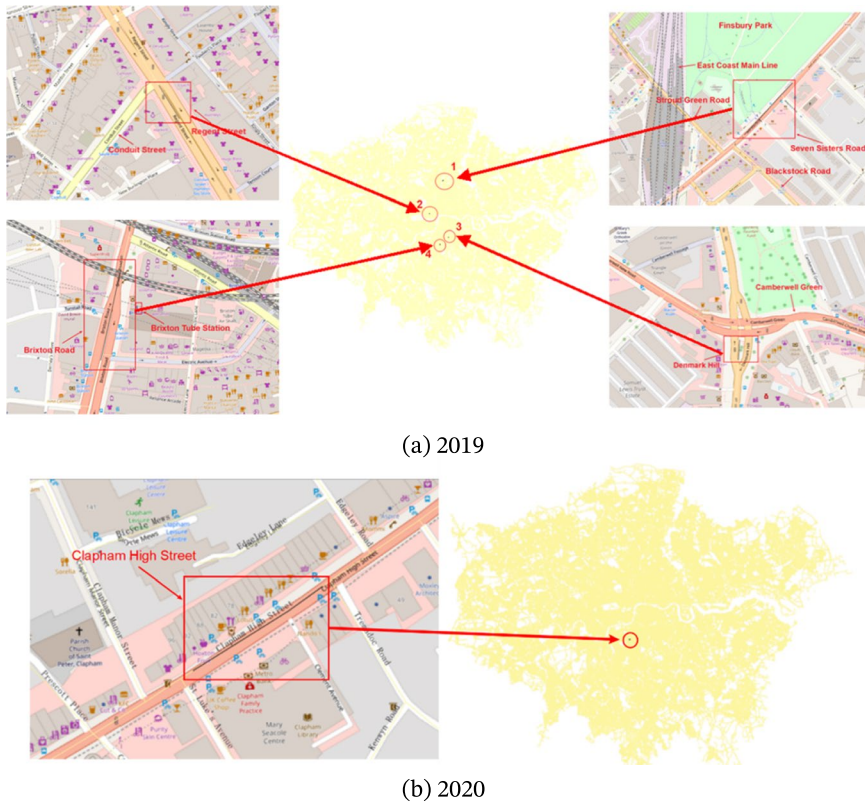
Fig. 3 The riskiest roads of Network KDE in 2019 & 2020

of north London, in a busy multicultural neighbourhood near Finsbury Park and close to the major transport hub, the Finsbury Park Station featured busy local amenities. The second road is part of Regent Street in the borough of Westminster, hosting the most famous shopping centre and many flagship retail stores, shops, and restaurants. The third road locates at the intersection of Denmark Hill and Camberwell Green in Southwark, as the access road to one of the busiest stations in South London annoyed by congestion and overcrowding. The last road is Brixton Road in the borough of Lambeth, characterised by its being overcrowded but narrow with a pair of Bus Lines, and embracing several markets as well as the Brixton Tube Station. In 2020, there is only one riskiest road, Clapham High Street, which is near Brixton Road and has a similar characteristic. In terms of the hotspot analysis, it is clear that the pandemic and its prevention policies like lockdown have a great influence on crash hotspots.

It is obvious that all these roads are bustling with POIs, i.e., shops, transport stations, and restaurants, evidencing the necessity to explore traffic crashes by varied POI categories.

The spatial cluster characteristics of two years' traffic crashes had been measured by Moran's *I* statistics, arriving at the significant global Moran's *I* at 0.25 and 0.18 (*p*-value of 0.001), respectively, indicating positive spatial autocorrelation hence clustering tendency at LSOA level of severity-weighted crashes. To further locate where are the clusters, LISA cluster map had been drawn using local Moran's *I* in Fig. 4a and b. It could be found that the four most significant high-high (HH) clusters are in northern London, the city centre (high density of POIs), and on the west and east (the Heathrow airport location) wings of the city. In addition, Fig. 4c and d highlighted similarities and differences pre- and during the pandemic, where the area in a black circle indicates those unchanged clusters, and areas in green or red circle are clusters experiencing an increase or decrease, respectively. For example, there are similar clusters of the Heathrow airport, Ealing borough (major metropolitan centre), and the City of London in 2019 and 2020, indicating the modest effect of the pandemic on such transport junctions and city centre. Majority of workplace clusters located in eastern London alongside the River Thames had shrunk in 2020, due to the work-from-home policy. On the contrary, parks and entertainments such as, Richmond Park, Hampton Court Park, Walthamstow Wetlands, etc., circled in green had expanded during the pandemics, showcasing citizens' rising leisure activities.


(a) 2019	(b) 2020
(c) 2019	(d) 2020
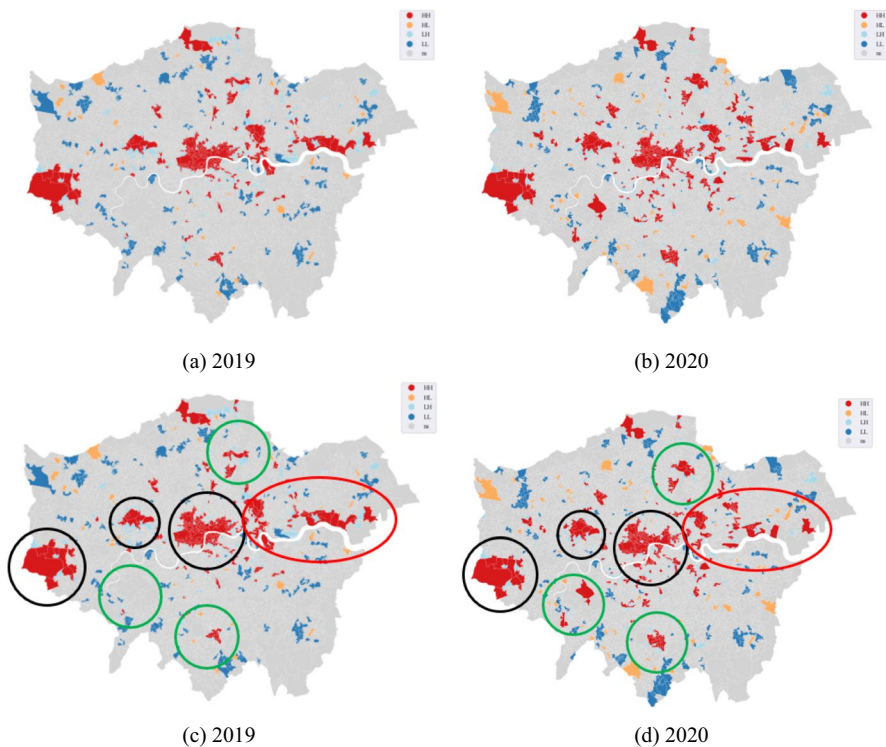
Fig. 4 LISA cluster map for crashes in 2019 & 2020

Factor Analysis of Road Traffic Crashes

A geographical detector measures the individual and interactive effects of POIs on crash distribution and realise the POIs' categorisation through the JNB division method. A further factor detector implied the categorisation's statistical significance pre- and during the pandemic, as shown in Table 4. In general, cliff drops of PD values had been witnessed but with unchanged rankings throughout the observation years, except for the rank for SHOP (⊞) increased from 5th to 3rd in 2020. It shows that the impact of POIs distribution on car crashes remained relatively stable over time by interpreting the PD value. For instance, the PD values for STATION (⊖) were the highest, with 43.2% and 29.6% for corresponding years respectively, followed by the second highest PD values for WORK (💼) at 0.397 and 0.229 in these two years, respectively. The locations of these two factors could then signify car crash spatial distributions.

An interaction detector had been utilised to further quantify the interconnected effect of paired factors on crashes in 2019 and 2020. In theory, it would be 36 pairs of interactions among nine factors, but only listed the top five in Table 4 for

Table 4 PD values of individual and interactive effects of POIs in 2019 & 2020

Effect type	POI categories	PD value(ranking)		p-value
		2019	2020	
Individual effect	⊖	0.432(1)	0.296(1)	<0.05
	💼	0.397(2)	0.229(2)	<0.05
	🍴	0.366(3)	0.200(4)	<0.05
	🎬	0.313(4)	0.177(5)	<0.05
	🛒	0.296(5)	0.207(3)	<0.05
	🎡	0.289(6)	0.141(6)	<0.05
	🏛	0.220(7)	0.121(7)	<0.05
	🏠	0.125(8)	0.076(9)	<0.05
	Ⓟ	0.088(9)	0.089(8)	<0.05
Interactive effect	🍴 ∩ ⊖	0.562(1)	0.365(2)	<0.05
	💼 ∩ ⊖	0.554(2)	0.363(3)	<0.05
	🛒 ∩ ⊖	0.540(3)	0.376(1)	<0.05
	🎬 ∩ ⊖	0.510(4)	0.344(4)	<0.05
	🎡 ∩ ⊖	0.503(5)	-	<0.05
	🏛 ∩ ⊖	-	0.323(5)	<0.05

information, with STATION as the compulsory. There are also declines of PD values for the consistent top four pairs of interactive factors, where the SHOP interacting with STATION that ranks from 3rd to 1st pre- and during the pandemic, by replacing the interaction between RESTAURANT (✂🍴) and STATION. Such results implied a possible high risk of car crashes in areas with a higher density of top five POIs near transport stations. With restrictions on eating indoors and the rise of online shopping, people may have visited shops near transport stations more often than restaurants, leading to an increased risk of car crashes in these areas. Another difference is that the 5th interactive factor has been changed from STATION interacting with ENTERTAINMENT (▷) to its being interacting with EDUCATION (🏛). Londoners may use public transport more often to travel to educational institutions, instead of visiting entertainment venues which experienced shut down or being less accessible due to pandemic restrictions.

Spatiotemporal Data Mining of Road Traffic Crashes

To locate the locations and periods with higher car crashes rates by age group, this section models the crashes in each LSOA using two tensors, $A_{weekday}, A_{weekend} \in R^{I \times J \times K}$, or weekdays and weekends in 2019 and 2020, respectively. The three dimensions in the tensors are I for regions, J for age groups, and for time slots in a single day. Since there are 4,835 LSOAs in London, *I* will be presented by $r = [r_1, r_2, \cdots, r_{4835}]$; *J* for age groups are $a = [a_1, a_2, \cdots, a_7]$ and *K* for time slots by an hour as $t = [t_1, t_2, \cdots, t_{24}]$. The SWI index, therefore, is stored in entry (i,j,k), which refers to a cumulative value of the SWI index of age groups a_j in LSOA r_i and time slot t_k throughout the study period.

Tensor Size Selection

The sizes of core tensors had been calculated respectively for 2019-weekday and 2020-weekday as well as 2019-weekend and 2020-weekend scenarios, with three decomposed time patterns for weekdays (morning peak hours, afternoon peak hours, and other times) and two distinctive patterns for weekends (evening peak hours and other times). After multiple attempts and integration of three similar age groups, the KL divergence parameter suggests a core tensor size at $[13, 5, 3]$ for weekdays and $[8, 4, 2]$ for weekends pre- and during the pandemic (Fig. 5). In Fig. 5a and c, the x-axis indicates different numbers of spatial patterns, and the y-axis lists the corresponding value of KL divergence, illustrating that the iteration of the parameter with initial value $KL = 3.1 \& 3.2$ converges to nearly 2.4 & 2.5 where $x = 13$ as the most appropriate size; the 2019-weekend's and 2020-weekend's selection in Fig. 5b and d exhibits four
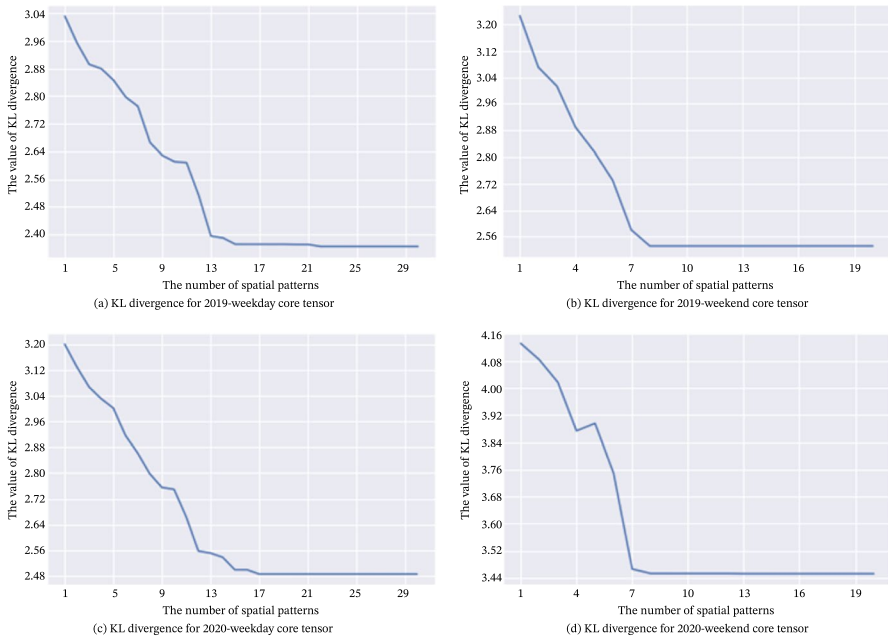
Fig. 5 Size selections of four core tensors

main patterns by age group, with 8 optimal spatial patterns being suggested by the KL divergence.

Tensor Decomposition

These two tensors for car crashes in 2019 and 2020, respectively, QUOTE $A_{weekday}$ and QUOTE $A_{weekend}$, are decomposed separately into a core tensor and three matrices as presented: QUOTE $G_{weekday} \in R^{13 \times 5 \times 3}$ and QUOTE $G_{weekend} \in R^{8 \times 4 \times 2}$, Fig. 6 shows the spatial patterns, age patterns, and temporal patterns on weekdays are identifiable as matrices of (a) QUOTE $A_{weekday} \in R^{4835 \times 13}$, QUOTE $B_{weekday} \in A^{7 \times 5}$, QUOTE $C_{weekday} \in T^{24 \times 3}$, and (b) QUOTE $A_{weekend} \in R^{4835 \times 8}$, QUOTE $B_{weekend} \in A^{7 \times 4}$, QUOTE $C_{weekend} \in T^{24 \times 2}$. It should be noted that the value of the SWI index storing each entry of tensors is normalised to QUOTE $[0, 1]$ for nonnegative Tucker decomposition; the degree of importance for study subjects has been referred to as the "degree of importance", where a high value means a vital role of a target object in its pattern and vice versa.

Pattern analysis of road traffic crashes during weekdays

Upon tensor decomposition, Fig. 7 depicted daily patterns, age patterns and a core tensor of traffic crashes on weekdays in 2019 and 2020, respectively. Figure 7(a) exhibited the obvious daily pattern pre- the pandemic in the green line, with a peak at 8:00 am, when car crashes occurred in the morning peak hours;
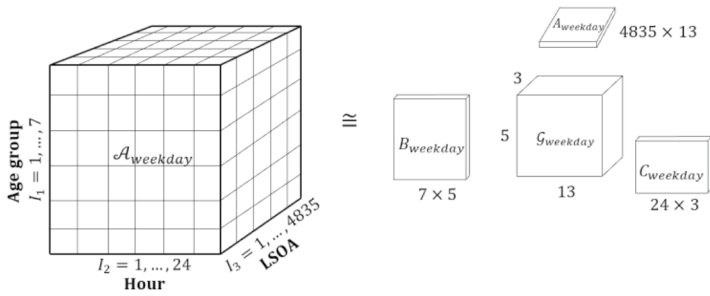
(a) 2019-weekday & 2020-weekday tensor
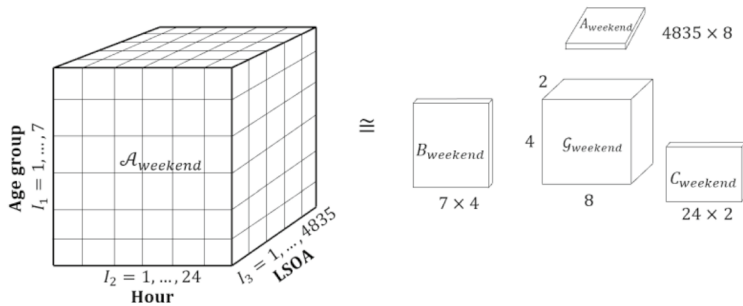
(b) 2019-weekend & 2020-weekend tensor

Fig. 6 Tensor decomposition of tensors

this morning peak hour is not obvious in 2020 but the dinner peak hours with a peak at 5:00 pm. The orange line peaked at 5:00 pm—6:00 pm in 2019 and 4:00 pm in 2020 are approximately the afternoon rush hours, implying the afternoon peaks of car crashes in 2020 appear earlier than those in 2019; the blue line illustrated the pattern during off-peak hours.

In light of the varied age groups presented in Fig. 7c and d, the purple colour with higher proportions indicated the main subject of the age-group pattern, especially the patterns for 26–35-year-olds, 36–45-year-olds, 46–55-year-olds, 18–25-year-olds, and 0–18-year-olds. Other age group patterns were difficult to detect due to their possible high similarities. In total, 13 spatial patterns have been detected with complicated information about car crashes in these two years, contributing to the later geographical detector. To match with the core tensor reflected nexus of such patterns at three dimensions, four further sliced matrices from five age dimensions had been drawn in Fig. 7e and f, where the light purple indicates a high degree of closeness. For example, the first matrix reflected first age patterns for 26–35 age group, and it has 1st spatial pattern & off-peak hours daily pattern, 3rd spatial pattern & afternoon peak hour daily pattern, and
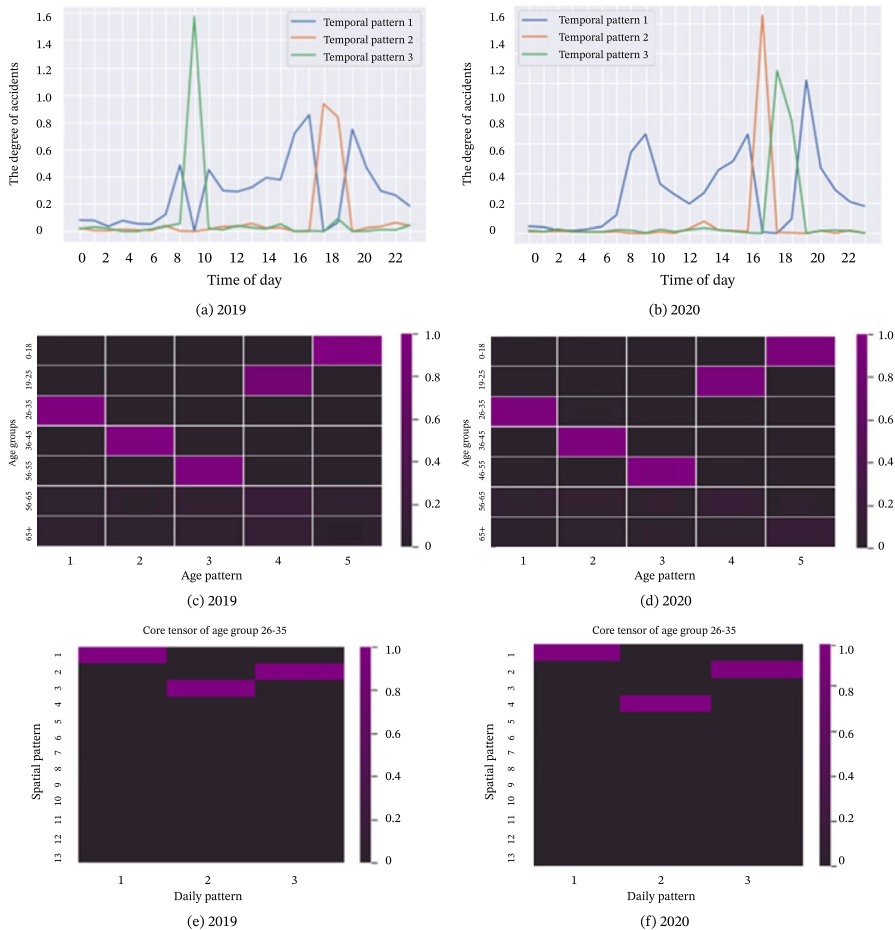
Fig. 7 Decomposed patterns of 2019-weekday & 2020-weekday tensor. (**a**) and (**b**) Temporal pattern; (**c**) and (**d**) Age pattern; (**e**) and (**f**) Examples of matrices of Core tensor. The matrices of core tensor are only for illustrating our workflow but not for comparative analysis, so one example of this matrix is shown for two years respective

2nd spatial pattern & morning rush hour daily pattern. It should be noted that the third and fourth age patterns have a major pattern and minor pattern, rendering their potential similar patterns with other age groups.

Pattern Analysis of Road Traffic Crashes During Weekends

In a similar design, decomposed tensor results on weekends for two years were presented in Fig. 8. As described in Sect. 4.1, most car crashes happened during the weekend evenings. In Fig. 8a, the orange line depicted that age-involved crashes occurred at 20:00 were distinctive, hence being reorganised as the pattern of evening
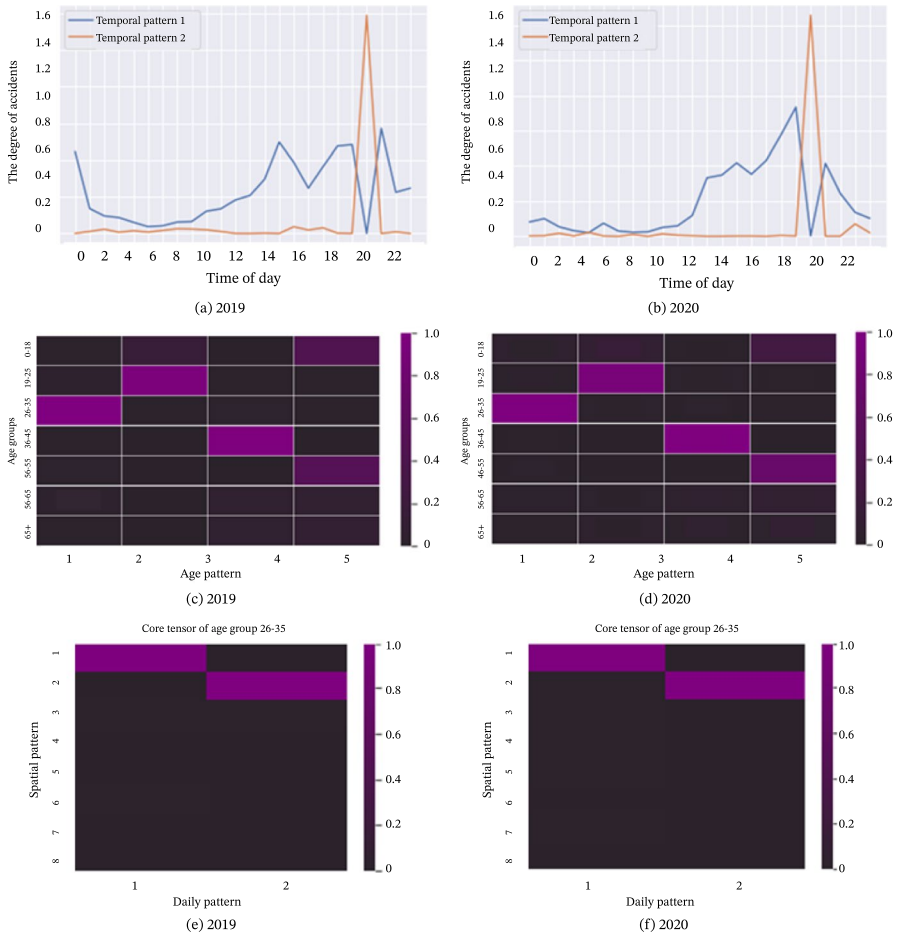
Fig. 8 Decomposed patterns of weekday tensor. (**a**) Temporal pattern; (**b**) Age pattern; (**c**) Matrices of Core tensor

peak hours; the blue line indicated car crashes distribution at other times on the weekends.

Figure 8b and d clearly exhibited four age patterns for people aged between 26 and 35, 18 to 25, 36 to 45, and 0–18 & 46–55. The patterns for people over 55 are not revealing due to their being similar as on weekdays. It could be observed that people aged 0–18 and 46–55 have the same crash patterns. There are eight spatial patterns and a core tensor for weekends being sliced by age dimensions but in identified metrics, with the last matrix representing the age patterns of 0–18 and 46–55.

Table 5 Dominant influence and interaction of factors for age groups during weekdays in 2019 & 2020

Age group	Year	Off-peak hours		Morning peak hours		Afternoon peak hours		Dinner peak hours	
		Individual factor	Interactive factors	Individual factor	Interactive factors	Individual factor	Interactive factors	Individual factor	Interactive factors
0–18	19	[STATION]	[STATION] ∩ [WORK]	[STATION]	[STATION] ∩	-	-	[STATION]	[STATION] ∩ [WORK]
	20	-	-	-	-	-	-		∩ [PARK]
19-25	19	[STATION]	[STATION] ∩ [WORK]	[STATION]	[STATION] ∩ [WORK]	-	-	[STATION]	[STATION] ∩
	20	[STATION]	[STATION] ∩ [WORK]	-	-	[SHOP] or [STATION]	[SHOP] ∩ [PARK]	[WORK]	[SHOP] ∩
26–35	19	[STATION]	[STATION] ∩ [FOOD]	[WORK]	[STATION] ∩ [WORK]	-	-	[SHOP]	[STATION] ∩ [SHOP]
	20	[WORK]	[STATION] ∩ [WORK]	-	-	[FOOD]	[STATION] ∩ [FOOD]	[FOOD]	[FOOD] ∩ or [STATION]
36–45	19	[STATION]	[STATION] ∩ [SHOP]	[STATION]	[STATION] ∩	-	-	[STATION]	[STATION] ∩
	20	[STATION]	[STATION] ∩ [SHOP]	-	-	[FOOD] or [WORK]	[STATION] ∩ [WORK]	[STATION]	[STATION] ∩ [WORK]
46–55	19	[STATION]	[STATION] ∩ [WORK]	[STATION]	[STATION] ∩ [FOOD]	-	-	[STATION] or [FOOD]	[STATION] ∩ [FOOD]
	20	[STATION]	[STATION] ∩ [PARK]	-	-	[STATION] or [SHOP]	[SHOP] ∩ [WORK]	[STATION]	[STATION] ∩ [FOOD]

Spatiotemporal Pattern Analysis for each Age Group

On basis of such spatial and temporal patterns in 2019 and 2020 by age group, both dominant POI factors and interactive effects of severity-weighted crashes could be explored through geographical detector, with results presented in Tables 5 and 6.

According to Table 5, STATION and WORK factors have the same significant effects on car crashes involving every age group on weekdays pre- and during the pandemic, except for victims aged above 55 years old. Nevertheless, it is obvious that every age group has different dominant POI factors. In terms of 0–18 age group

Table 6 Dominant influence and interaction of factors for age groups during weekends in 2019 & 2020

Age group	Year	Off-peak hours Individual factor	Off-peak hours Interactive factors	Evening peak hours Individual factor	Evening peak hours Interactive factors
0–18	19	[TRANSPORT]	[TRANSPORT] ∩ [SHOP]	[TRANSPORT]	[TRANSPORT] ∩ [PARKING]
0–18	20	[TRANSPORT]	[TRANSPORT] ∩ [EDUCATION]	[RESTAURANT]	[TRANSPORT] ∩ [ENTERTAINMENT]
19-25	19	[TRANSPORT]	[TRANSPORT] ∩ [RESTAURANT]	[ENTERTAINMENT]	[TRANSPORT] ∩ [ENTERTAINMENT]
19-25	20	[TRANSPORT]	[TRANSPORT] ∩ [RESTAURANT]	[PARKING]	[TRANSPORT] ∩ [PARKING]
26–35	19	[RESTAURANT]	[TRANSPORT] ∩ [SHOP]	[RESTAURANT]	[TRANSPORT] ∩ [RESTAURANT]
26–35	20	[TRANSPORT]	[TRANSPORT] ∩ [RESTAURANT]	[WORK]	[SHOP] ∩ [ATTRACTION]
36–45	19	[TRANSPORT]	[TRANSPORT] ∩ [RESTAURANT]	[WORK] or [RESTAURANT]	[WORK] ∩ [RESTAURANT]
36–45	20	[WORK]	[TRANSPORT] ∩ [WORK]	[WORK] or [RESTAURANT]	[TRANSPORT] ∩ [RESTAURANT]
46–55	19	[TRANSPORT]	[TRANSPORT] ∩ [SHOP]	[TRANSPORT]	[TRANSPORT] ∩ [PARKING]
46–55	20	[TRANSPORT]	[TRANSPORT] ∩ [EDUCATION]	[RESTAURANT]	[TRANSPORT] ∩ [ENTERTAINMENT]

specifically, POIs factors do not have significant PD-values in 2020, except for the EDUCATION and its interaction with ENTERTAINMENT during dinner peak hour. It suggested that residents had changed their routine behaviour throughtout the pandemic, i.e., spending more time at home rather than visiting POIs as frequently as they would have pre-pandemic, and areas with a high density of schools should be considered to protect them from crashes. Adults (age groups 19–25, 46–55), had similar changes of dominant factor, the SHOP imposing its greater influence on

patterns in 2020 than 2019. With more people working from home and avoiding crowded places, the number of people paying visit to shops near transport stations had increased. Besides, people may take shopping as a replacement of entertainment due to the shutdown of entertainment venues. These behaviour changes were likely to be found in younger adults, who were more inclined to engage in leisure and shopping activities in commercial areas; as well as those middle-aged adults, who had more disposable incomes and greater possibility to work from home, or flexibility in adjusting their working hours. It illustrates more car crashes near the areas with a high density of shops, especially shops near attractions, entertainment venues and workplaces, during afternoon and dinner peak hours. For age group of 26–35 and 36–4, the RESTAURANT and WORK as well as their interactions with STATION become the dominant factors in 2020, respectively, identifying they may drive more frequently to restaurants were reopen or offering take-aways, and those risky areas with a high density of workplaces and restaurants around transport stations.

In comparison to weekdays' factors, the WORK on weekends has lower effects on crash distribution and led out the dominant effects to ENTERTAINMENT and RESTAURANT since people tend to have more leisure time hence be able to engage in these recreational activities. There are also differences of distinctive factors and interactions at the weekend by age group pre- and during the pandemic, especially during evening peak hours. For example, RESTAURANT becomes the dominant factor for age groups 0–18 and 46–55, during the evening peak hours, although STATION and its interactions with EDUCATION and ENTERTAINMENT have the dominant effect in 2020 because restaurants tend to be located in areas that are popular for entertainment and may also be located near schools. Besides, during evening peak hours, the individual factors and interactive factors for 19–25 and 26–35 age groups have been changed from ENTERTAINMENT, RESTAURANT, and their interactions with STATION to PARKING, STATION, WORK and SHOP interacting with ENTERTAINMENT during the pandemic. This could be because people were more likely to engage in outdoor activities and visit places with more parking, such as parks or shopping centres.

In summary, in the observing years, the distributions of transport stations, workplaces, and restaurants significantly affected casualty patterns of all age groups, while transport stations nearer to restaurants and workplaces have the largest impact on crashes patterns, calling for extra attention to deploy corresponding countermeasures in such areas to reduce traffic crashes risk behaviours, and to apply certain technique support in consolidating motorway safety engineering. Additionally, these results can be applied to identify the riskier areas requiring extra attention towards better traffic management and control strategies, tailoring for featured age groups through their featured activities pre- and during the pandemic.

Conclusion

This paper provides an in-depth analysis of the changes in factors influencing car crashes during the COVID-19 pandemic and how these changes differ by age group, day of the week and time of day, and suggests potential interventions that can be

taken to reduce the risk of crashes in specific areas. The findings of this paper can be used by policymakers, urban planners and transport engineers to improve road safety after the pandemic and future pandemic.

Upon recognising the spatial and temporal variations of traffic crashes in London in two representative years towards recovery from pandemic, it had been found that weekdays and weekends exhibited distinct patterns with daily trends against the time of day. Specifically, the riskiest roads and the main high-risk locations were spatially distributed in central London but spread out along the road, with a higher likelihood to happen at busier places and highly correlated to their neighbouring areas. Especially, areas with many transport stations and surrounding shops require particular attention to local transportation planning and traffic management. Such findings could be used to target specific areas with collision-alleviation intervention measures, such as traffic calming measures during rush hours or improved pedestrian crossings; it could also be informative for urban planners and policymakers to make evidence-based decisions on future design of POIs, taking into consideration of local safety. Hereby the study contributed to the literature on age-group involved crashes' casualty and daily human activity analysis through its utilisation of geographical detector, in the results to discover their varied latent mechanism of spatiotemporal patterns with defined POI categories. In addition, it highlights the importance of data-driven approach for age-group involved collision patterns and their relationship to the built environment. Policymakers could develop evidence-based targeted intervention measures on basis of such findings, and further evaluate their effectiveness to improve road safety for different age groups over time.

Despite its contribution, some limitations could get improved in future extended work. For example, traffic demand for public transport has been taken as a temporary phenomenon with a speedy recovery rise after a short-term decline. Analysis of their long-term changes should be further undertaken to support traffic management decision-making in the long run.Besides, the spatiotemporal patterns of casualties aged over 55 have been unclear, which may owe to its relatively lower proportion of casualties hindered the representative pattern emerging; some more POI categories, such as hospitals and petrol stations, deserve to be added for comprehensive studies in the future towards a more in-depth study.

In the future, retrospective research on deploying up-to-date data will be conducted once with official datasets in place, to test the methodologies proposed in this study onto the real data and try to capture the full picture on London's road safety upon recovery from the pandemic.

Declarations

Conflict of Interest None.

References

Almansouri, H. M. S. (2021). The Impact of Covid on Shopping Centers in Libya. *Scientific* Research *Communications, 1*.

Amberber, N., Howard, A., Winters, M., Harris, M. A., Pike, I., Machperson, A., Cloutier, M. S., Richmond, S. A., Hagel, B., Fuselli, P. (2021). Road traffic injury during the covid-19 pandemic: Cured or a continued threat?. *University of Toronto Journal of Public Health, 2*.

Batty, M. (2020). The Coronavirus crisis: What will the post-pandemic city look like? : SAGE Publications Sage UK: London, England.

Blazquez, C. A., & Celis, M. S. (2013). A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accident Analysis & Prevention, 50*, 304–311.

Cappellari, P., & Weber, B. S. (2022). An analysis of the New York City traffic volume, vehicle collisions, and safety under COVID-19. *Journal of Safety Research, 83*, 57–65.

Champahom, T., Jomnonkwao, S., Karoonsoontawong, A., Ratanavaraha, V. (2020). Spatial zero-inflated negative binomial regression models: Application for estimating frequencies of rear-end crashes on Thai highways. *Journal of Transportation Safety & Security*, 1–18.

Chen, J., Yang, S., LI, H., Zhang, B., & Lv, J. (2013). Research on geographical environment unit division based on the method of natural breaks (Jenks). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3*, 47–50.

Chen, Y., Luo, R., King, M., Shi, Q., He, J., & Hu, Z. (2022). Spatiotemporal analysis of crash severity on rural highway: A case study in Anhui. *China. Accident Analysis & Prevention, 165*, 106538.

Ding, Y., Zhang, M., Qian, X., Li, C., Chen, S., & Wang, W. (2019). Using the geographical detector technique to explore the impact of socioeconomic factors on PM2. 5 concentrations in China. *Journal of Cleaner Production, 211*, 1480–1490.

Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2004). Identification and Ranking of Black Spots: Sensitivity Analysis. *Transportation Research Record, 1897*, 34–42.

GLA. (2014). Statistical GIS Boundary Files for London. London.

GLA. (2020). *2019-based Trend Projections* [Online]. London: GLA. Available: https://data.london.gov.uk/dataset/trend-based-population-projections [Accessed].

Green, J., Muir, H., & Maher, M. (2011). Child pedestrian casualties and deprivation. *Accident Analysis & Prevention, 43*, 714–723.

Han, J., Kamber, M., Pei, J. (2012). 13 - Data Mining Trends and Research Frontiers. *In:* HAN, J., KAMBER, M. & PEI, J. (eds.) *Data Mining (Third Edition).* Boston: Morgan Kaufmann.

Hershey, J. R., Olsen, P. A. (2007). Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models. IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, 15–20 April 2007 2007. IV-317-IV-320.

Iyanda, A. E. (2019). Geographic analysis of road accident severity index in Nigeria. *International Journal of Injury Control and Safety Promotion, 26*, 72–81.

Katrakazas, C., Michelaraki, E., Sekadakis, M., & Yannis, G. (2020). A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. *Transportation Research Interdisciplinary Perspectives, 7*, 100186.

Kaya, E., Agca, M., Adiguzel, F., & Cetin, M. (2019). Spatial data analysis with R programming for environment. *Human and Ecological Risk Assessment: An International Journal, 25*, 1521–1530.

Kolda, T. G., & Bader, B. W. (2009). Tensor Decompositions and Applications. *SIAM Review, 51*, 455–500.

Koloushani, M., Ghorbanzadeh, M., Ulak, M. B., Ozguven, E. E., Horner, M. W., Vanli, O. A. (2022). The Analysis of Spatial Patterns and Significant Factors Associated with Young-Driver-Involved Crashes in Florida. *Sustainability, 14*.

Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis & Prevention, 27*, 663–674.

Mathew, S., Pulugurtha, S. S., & Duvvuri, S. (2022). Exploring the effect of road network, demographic, and land use characteristics on teen crash frequency using geographically weighted negative binomial regression. *Accident Analysis & Prevention, 168*, 106615.

Okabe, A., Okunuki, K.-I., & Shiode, S. (2006). SANET: A Toolbox for Spatial Analysis on a Network. *Geographical Analysis, 38*, 57–66.

Okabe, A., & Sugihara, K. (2012). *Spatial Analysis Along Networks: Statistical and Computational Methods*. Wiley.

ONS. (2020). *Coronavirus and Homeworking in the UK Labour Market: 2019*. Office of National Statistics.

Pour, A. T., Moridpour, S., Rajabifard, A., & Tay, R. (2017). Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area. *Road & Transport Research, 26*, 4–20.

Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis & Prevention, 50*, 678–687.

Shabanikiya, H., Hashtarkhani, S., Bergquist, R., Bagheri, N., Vafaeinejad, R., Amiri-Gholanlou, M., Akbari, T., & Kiani, B. (2020). Multiple-scale spatial analysis of paediatric, pedestrian road traffic injuries in a major city in North-Eastern Iran 2015–2019. *BMC public health, 20*, 1–11.

Shashua, A., Hazan, T. (2005). Non-negative tensor factorization with applications to statistics and computer vision. Proceedings of the 22nd international conference on Machine learning, 792–799.

Sun, L., & Axhausen, K. W. (2016). Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part b: Methodological, 91*, 511–524.

Sutherland, M., Mckenney, M., & Elkbuli, A. (2020). Vehicle related injury patterns during the COVID-19 pandemic: What has changed? *The American Journal of Emergency Medicine, 38*, 1710–1714.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279–311.

Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators, 67*, 250–256.

Wang, Y., Wang, S., Li, G., Zhang, H., Jin, L., Su, Y., & Wu, K. (2017). Identifying the determinants of housing prices in China using spatial regression and the geographical detector technique. *Applied Geography, 79*, 26–36.

WHO. (2018). Global *Status Report on Road Safety* [Online]. Geneva: World Health Organization. Available: https://www.who.int/publications-detail/global-statusreport-on-road-safety-2018 [Accessed].

Worachairungreung, M., Ninsawat, S., Witayangkurn, A., & Dailey, M. N. (2021). Identification of Road Traffic Injury Risk Prone Area Using Environmental Factors by Machine Learning Classification in Nonthaburi. *Thailand. Sustainability, 13*, 3907.

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems, 32*, 396–406.

Yamada, I., & Thill, J. C. (2007). Local indicators of network-constrained clusters in spatial point patterns. *Geographical Analysis, 39*, 268–292.

Yao, D., Yu, C., Jin, H., & Ding, Q. (2015). Human mobility synthesis using matrix and tensor factorizations. *Information Fusion, 23*, 25–32.

Yasin, Y. J., Grivna, M., & Abu-Zidan, F. M. (2021). Global impact of COVID-19 pandemic on road traffic collisions. *World Journal of Emergency Surgery, 16*, 51.

Zahran, E.-S.M.M., Tan, S. J., Tan, E. H. A., Mohamad'asri Putra, N. A. A. B., Yap, Y. H., & AbdulRahman, E. K. (2021). Spatial analysis of road traffic accident hotspots: evaluation and validation of recent approaches using road safety audit. *Journal of Transportation Safety & Security, 13*, 575–604.

Zhang, Y., Lu, H., Qu, W. (2020). Geographical Detection of Traffic Accidents Spatial Stratified Heterogeneity and Influence Factors. *International Journal of Environmental Research and Public Health,* 17.

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*, 1–55.

Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., & Chang, E. (2014b). Diagnosing New York city's noises with ubiquitous data. *Proceedings of the 2014b ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 715–725.

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis & Prevention, 135*, 105323.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.