5
Applications and Visualisations

This chapter serves as a gallery of possible applications that can arise from the data generated from the Smart street sensor project. As we saw in the chapter 2 availability of such granular, longitudinal data on movement and distribution of people at such spatial extent has numerous uses in various fields of study. In this chapter we first start by looking at the use of the data in understanding the broad footfall landscape of the United Kingdom deriving sample insights on how retail footfall in UK has been performing for the past couple of years. The sample analysis were done on various levels from National index to individual locations.

5.1 *Footfall Landscape of United Kingdom*

Figure 5.1 shows the weekly footfall index of UK through 2017 to 2018 which in addition to showing larger trends also shows sudden short term changes such as week during the storm in Feb 2018. Figure 5.2 shows the spatial distribution of the footfall change in UK between these two years across cities showing the growth and decline of retail in the cities of West Bromwich and Ipswich. These changes could be further looked at in detail temporally as shown in figure 5.4. The difference in even smaller intra day patterns showing the nature of the economies of cities can also be inferred by their data profiles as shown in figure 5.4.
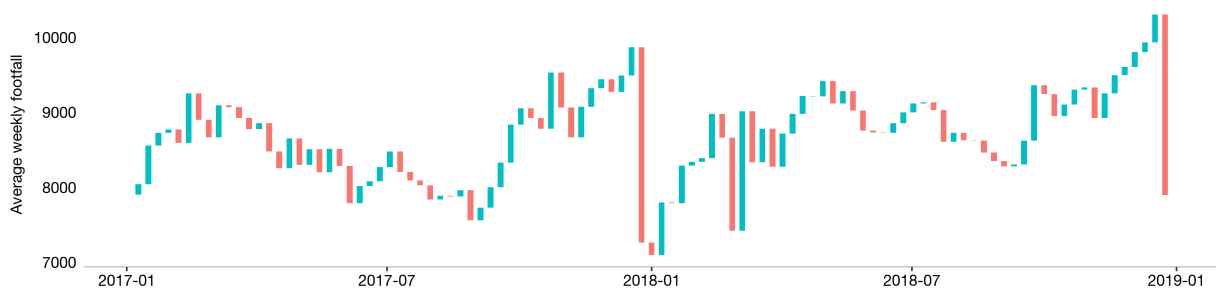

Figure 5.1: A weekly footfall index for United Kingdom showing the change in footfall from 2017 to 18
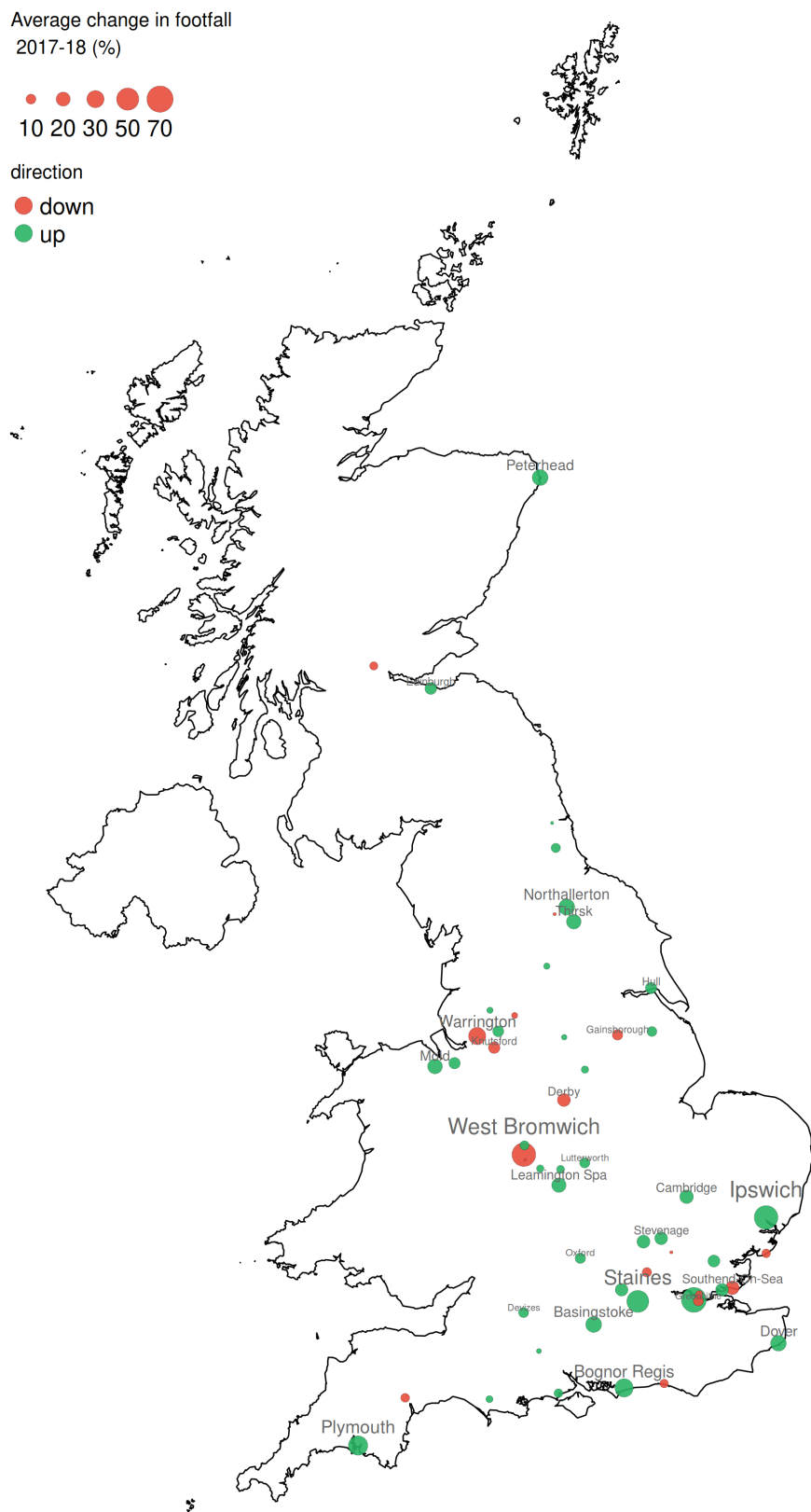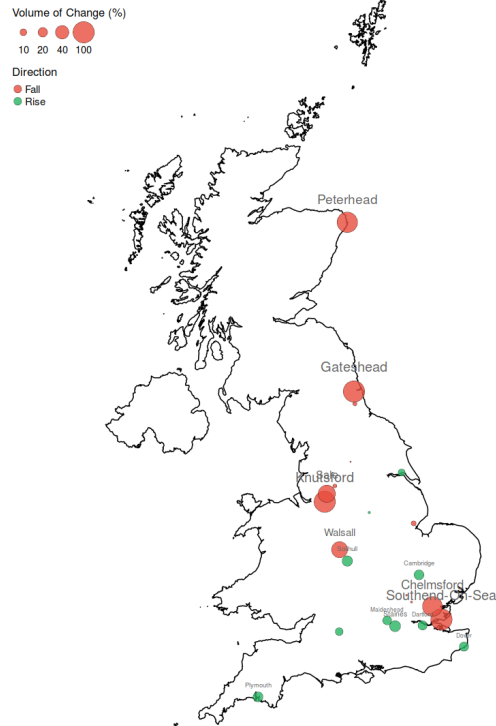
Figure 5.2: The change (%) in average weekly footfall of towns across UK in 2018 compared to 2017.
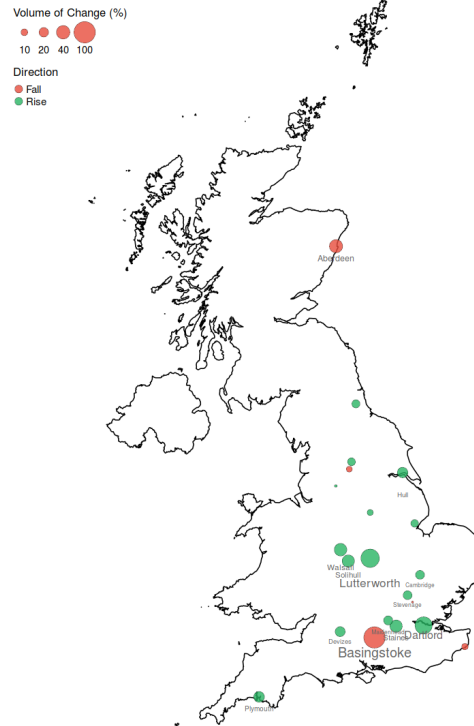
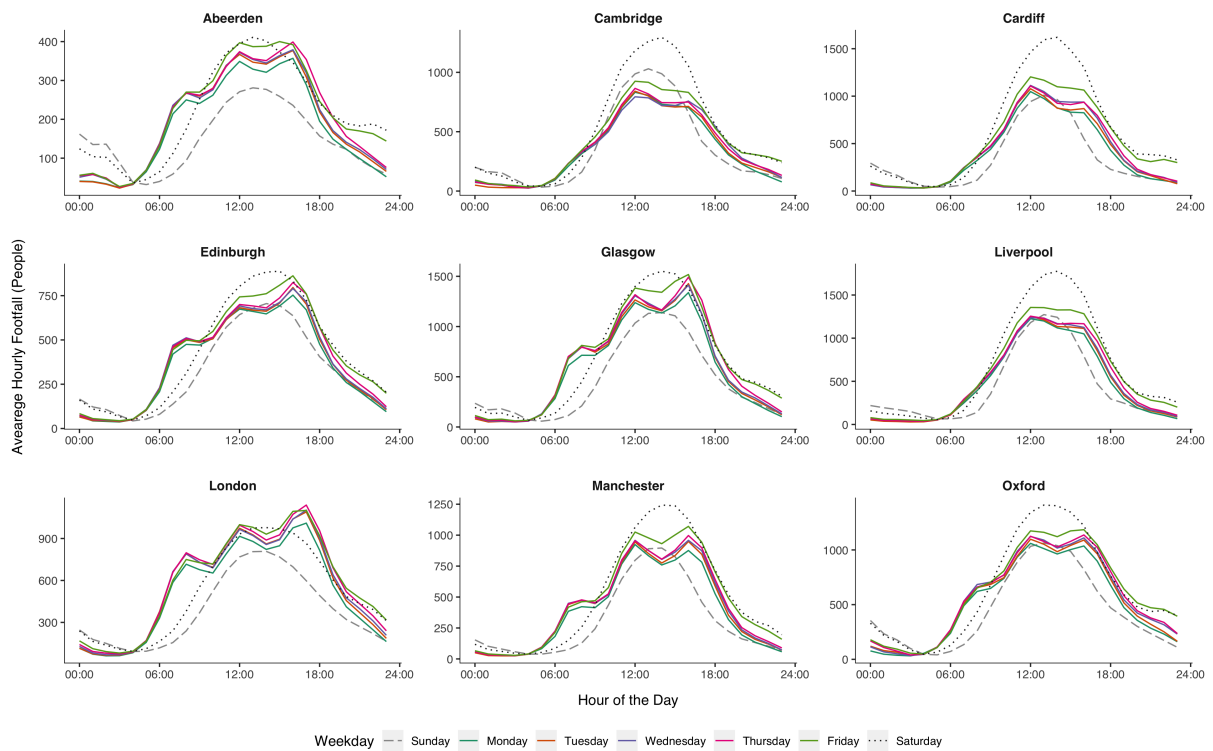Figure 5.3: The change (%) in monthly average footfall in towns across UK in April and May 2009.

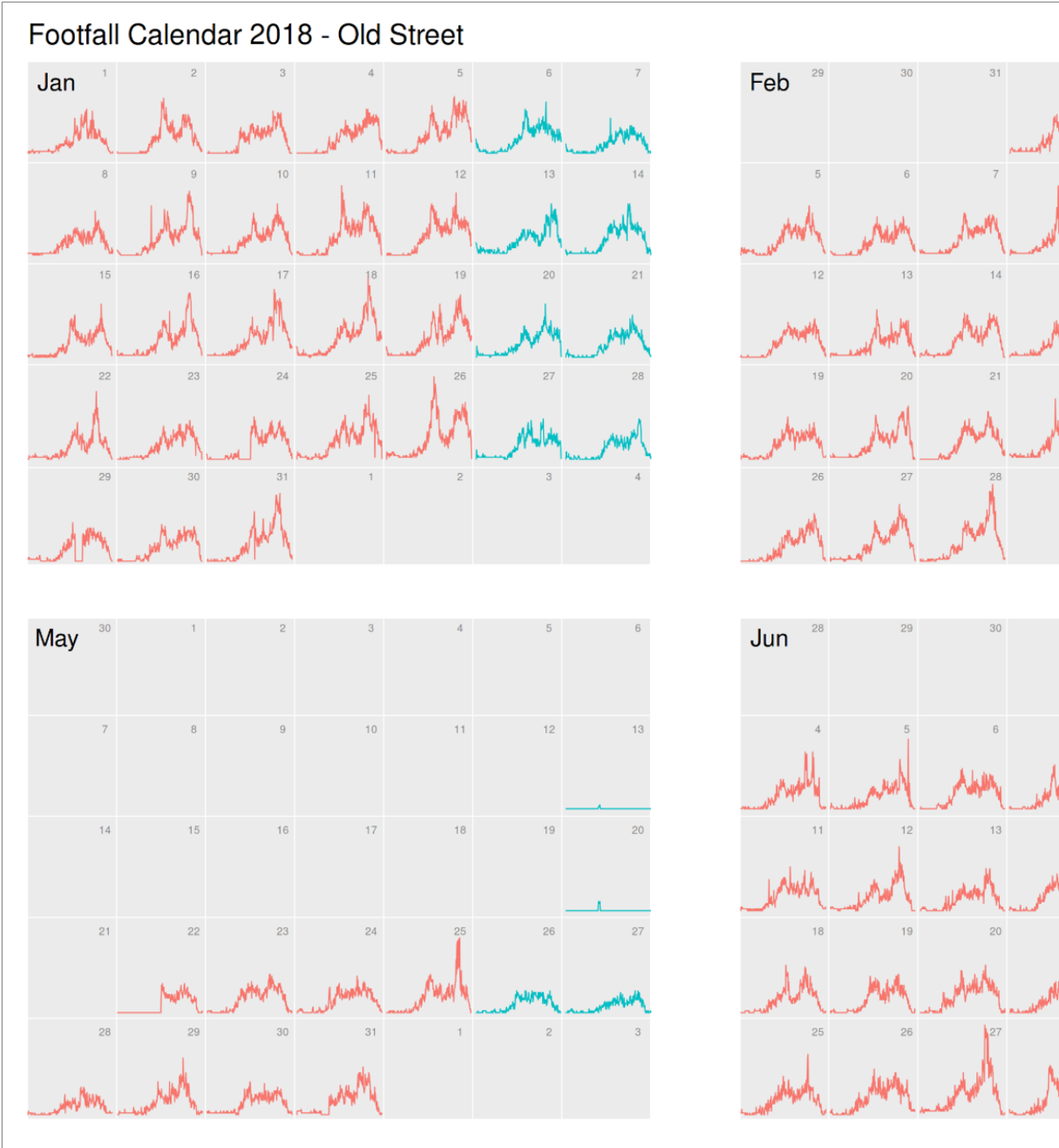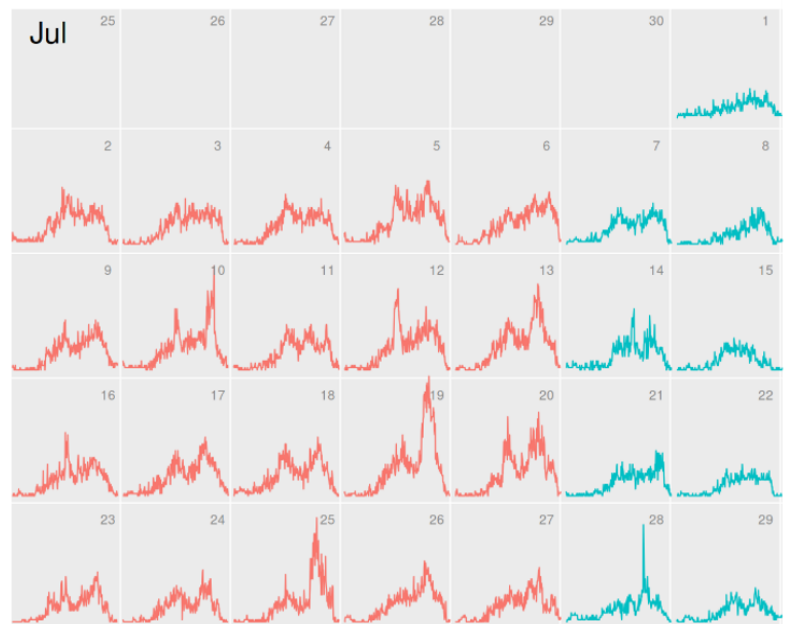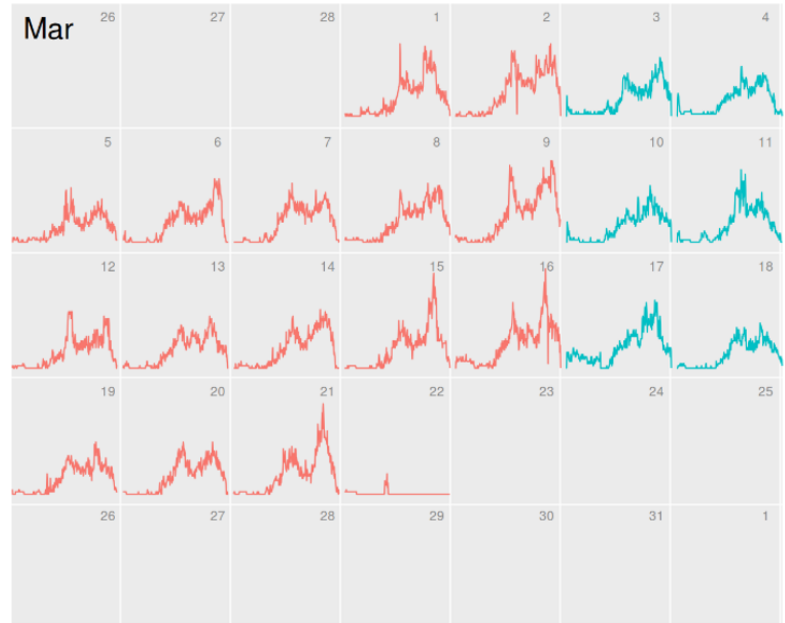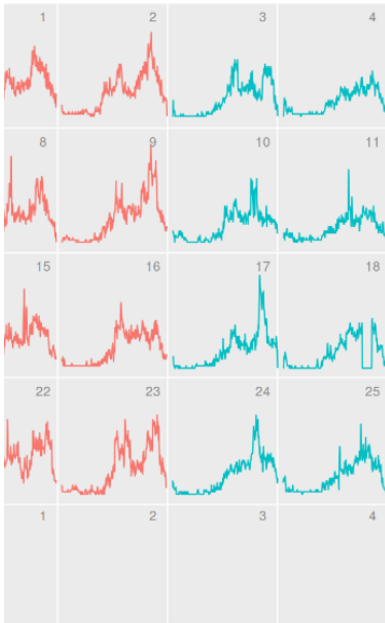Figure 5.4: Intra-day footfall profile of major cities in United Kingdom

Figure 5.5: Footfall calendar showing the profiles of daily volume of footfall at Old Street, London.
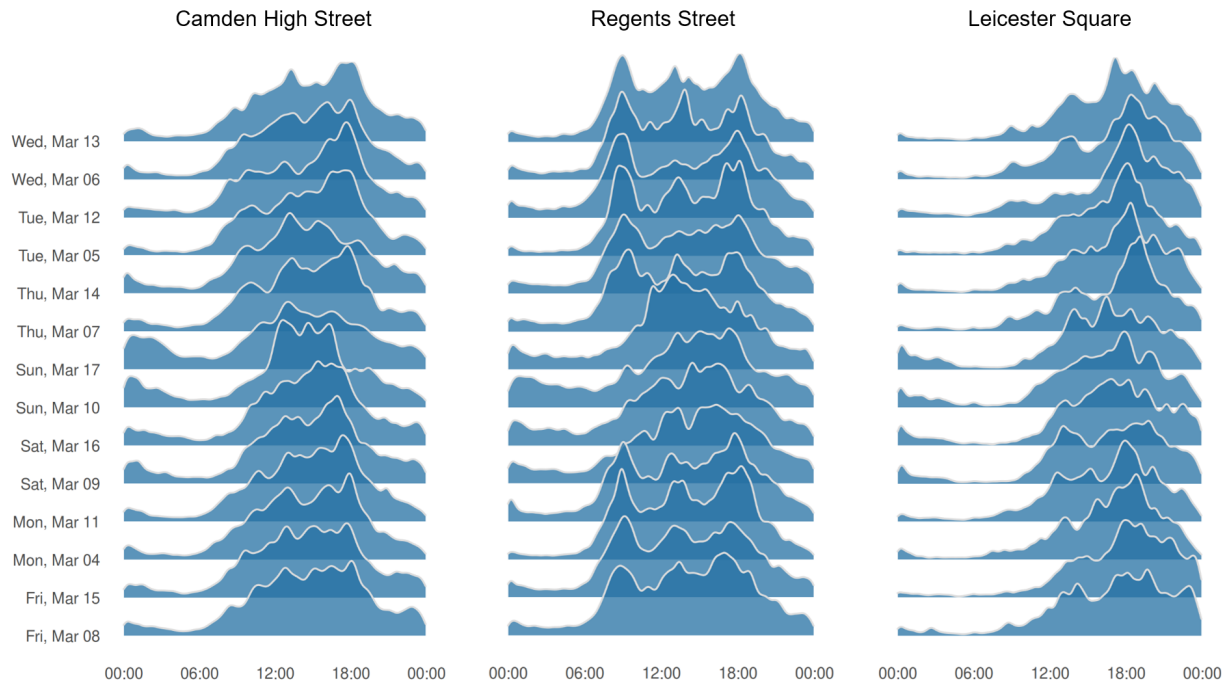
Figure 5.6: The profiles can be tracked longitudinally to reveal nature and change

5.2   Event Detection

When we look at the footfall longitudinally across locations we can uncover wide range of information about the context which resulted in the patterns in the footfall. Figure 5.7 shows the normalised weekly footfall of 10 different locations across Cardiff for two years 2017 and 2018. The patterns in the footfall clearly show numerous events that were happening at Cardiff as unusually high or low footfall at the corresponding week. The most significant one was in the Feb 2018, when all the sensor report the lowest numbers they have ever recorded. This coincides with the cold wave in UK named 'Beast from the East' which brought adverse conditions to all over UK leading to significant reduction in footfall. The other events we can identify are the bank holiday weekends which result in higher than normal footfall and the holiday shopping season when the footfall is the highest. Finally it is interesting to see the difference in the footfall in the summer weeks between 2017 and 2018 which could be explained by the FIFA world cup that took place in these weeks. This example shows the usefulness of the footfall data to detect events that happen in real life from the data in near real time. It can also be used to measure the effect of events on the footfall and hence understand the impact of these events for retail and economy in general.

5.2.1   Football world cup

In addition to the long term changes and events, the footfall data can be used to identify the smaller effects of these events at an area scales.

Figure 5.8 shows the footfall on two days in Leicester square, London when the quarter final and semi final matches of the FIFA world cup took place. Both matches happened in the evening and lead to increase in footfall around the match time. The more interesting observation is the effect the outcome of the game had on the footfall. On the day of the quarterfinal the winning of the English team led to a post match celebration after the match pushing the footfall to the highest during the day unlike the semifinal when the English team lost. This not only shows the usefulness of the data in understanding effect events have on local footfall, but also can be used by these retailers to predict the effect of the results of these sports events might affect them.
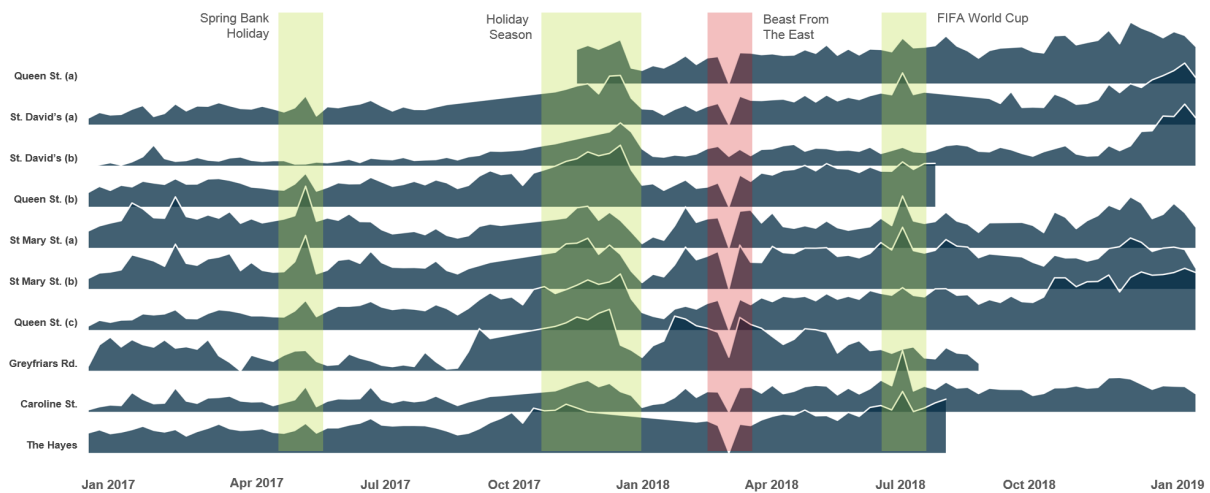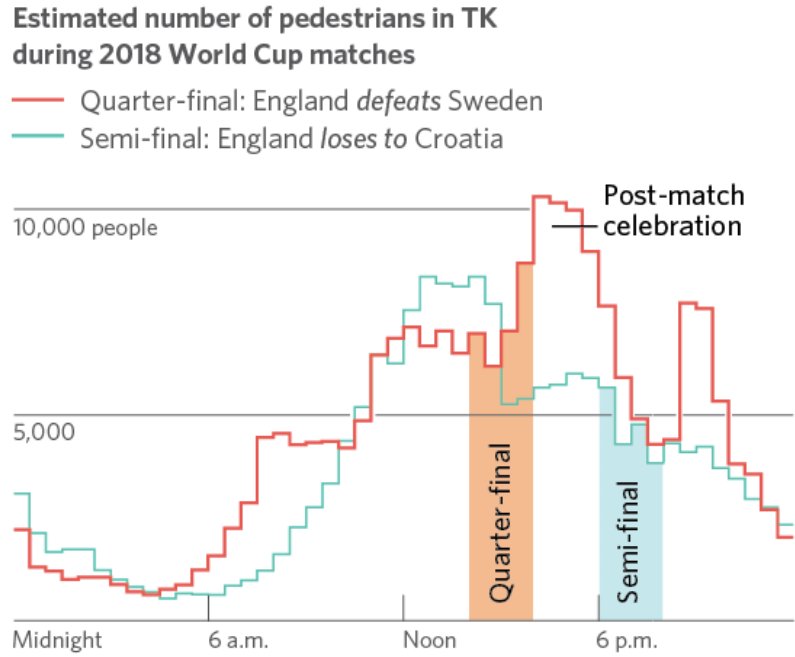

Figure 5.7: Normalised weekly footfall index at locations across Cardiff from 2017 to 2018

These examples show the importance of footfall data in detecting events. Even a simple visual analytics of the dataset reveal interesting information on events. This would be much more useful when used in tandem with advanced machine learning/data mining techniques and can predict much better results as we collect more data.

5.3 Pedestrian Flows

Detecting general trends in flow of people between spatial locations is neither obvious nor an trivial task due the high cost of capturing these movements without compromising the privacy of the users since the primary way to collect such detailed data involves handling data of their precise locations. This project specifically removes any such information because of MAC randomisation and hashing and hence seems like it might not be suitable for such studies on human mobility. Alternatively, this problem can be solved by examining the movement of people in a Smart Street Sensors network at a fine spatial and temporal resolution using a novel methodology to the field of Big Data using mathematical models from information theory - Transfer Entropy. This section serves as a small pilot study using an area in central London as a case study to

Figure 5.8: The difference in footfall distribution at Leicester square, London after FIFA world cup quarter-final and semi-final matches. Source: Oliver Uberti and James Cheshire (Sample needs to be replaced with original graphic)

Estimated number of pedestrians in TK during 2018 World Cup matches

— Quarter-final: England *defeats* Sweden
— Semi-final: England *loses to* Croatia

Post-match celebration

10,000 people

5,000

Quarter-final

Semi-final

Midnight   6 a.m.   Noon   6 p.m.

demonstrate the usefulness of TE as a measure of flow of pedestrians. [1]

Consider the array of sensors shown at Figure 5.9. Assuming that we have a flow of people walking past the location 116 and then diffusing towards the remaining sensors. Counts generated by the sensor are aggregated per five minute intervals, so if, for example, it takes one minute to walk from the location 116 to the location 117, the number of people detected at 117 from minutes 2 to 5, would correspond to the percentage of people detected at 116 from minutes 1 to 4. In other words, the similarity between the time series of counts at the locations under consideration are correlated. Hence the aim here is to, without actually tracking people, provide a measure for the size of the flow between each pair of sensors. One way to accomplish this, is to think of this motion of people as flows of information among distinctive sources, so we can relate the number of people reaching one sensor from another by measuring the uncertainty between two interacting random variables. For this, we used an information theory concept known as Transfer Entropy TE defined by:

$$TE(X, Y) = \sum_{t=1} p(y_{t+1}, y_t, x_t) \times log \frac{p(y_{t+1} \mid y_t, x_t)}{p(y_{t+1} \mid y_t)} \qquad (5.1)$$

Where t indicates a given point in time. 5.1 measures the reduction in uncertainty at y_t, given x_t and y_{t-1}. In comparison with the case when only y_{t-1} is known. This measure is applied directly to our people's movement problem and X = location i, Y = location j and t runs for a whole day, the TE would represent an indicator for the direction of the flow, as the counts at y_{t+1} are more accurately estimated using the information of x_t.
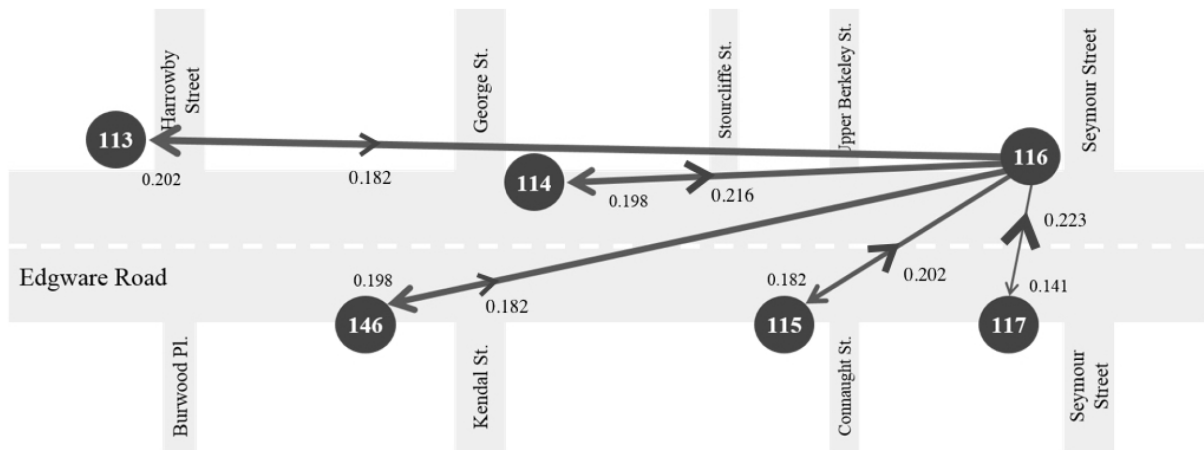
Figure 5.9: Illustration of transfer entropies between set of locations along Edgware Road, London.

Taking again Figure 5.9 as a reference, we measured the *TE* between sensor 116 and the rest of the sensors. The walking time is not constant and each sensor has counts at all times *i*, *j*. There are people passing by these sensors that came from locations outside the network. The numbers at each line represent the *TE* measured between each pair of sensor locations. The largest *TE* value found was between 117 and 115. The asymmetry of the TE is clear here, as the value in the opposite direction (115 to 117) is considerably lower. Another interesting value is the pair 116-117, where TE(116,117) « TE(117,116). This demonstrates that in this four-way crossing the predominant direction of flow is from location 117 to location 116 (from the bottom of the figure upwards, or from west to east in reality). These results suggest that, in general there is a larger flow of people from West side to East side of Edgware road and larger flow of people from South to North. The results are consistent with our intuition that there is a larger flow of people from South to North along this road towards the Edgware road underground station.

There is still a series of uncertainties yet to be addressed by this model, such as the decay of probabilities with distance and the number of interventions of opportunity encountered by people while walking from one sensor to another. However, this first initial set of results is encouraging in measuring flow between spatial points without actually tracking these users.

6
Discussion and Conclusions

Sentient technologies make measurement of the human activities that are the life blood of the smart city possible. Yet the data that they harvest are frequently relevant only to the sub-groups within society that avail themselves of particular goods and services – such as social media applications, transport modes or retail offers. In each of these cases, it is necessary to remember that the resulting data are by-products of consumer transactions, and will as a consequence, only pertain to users of the relevant goods or services. If the smart city is to be socially inclusive, it therefore follows that sentient data must represent entire populations, whether by design or by triangulation with external, population wide, sources. This is a non-trivial task, since the ebbs and flows of smart device-enabled citizens rarely pertain to any clearly defined population in either administrative or functional terms [Massam, 1975].

Our objective here has been to collect, rather than re-use, data on smart city functioning, by recording Wi-Fi probes and ultimately reconciling them with manual counts in order to infer ambient populations. The internal validation methodology set out in the technical sections of this paper, allied to external validation from pedestrian counts, renders the method inclusive and robust when recording activity levels in retail centres in real time. We have described the collection and processing of a novel consumer Big Dataset that enables valid measures of levels of footfall activity which has been scaled across a wide network of sensors [Longley et al., 2018]. In both conceptual and technical terms, it illustrates the ways in which passively collected consumer data can be 'hardened' to render them robust and reliable by using related procedures of internal and external validation.

Internal validation addresses the issues of screening out device probes that do not indicate footfall, and the further screening of device probes to 'fingerprint' the effects of MAC randomization. It is important to note that the filtering process work based solely on the information present in the probe requests and their temporal distribution. This ensures that although the mobile devices were uniquely identified, there was no further personal data generated by linking the probe requests to the users of the mobile devices. This method essentially gave us a way to estimate the footfall in real-time without identifying or tracking the mobile devices themselves. External validation then entailed reconciling adjusted counts with the footfall observed at sample locations. This procedure makes it

possible to generalise from locations at which manual footfall surveys are conducted to all others in the system, and to develop a classification of device locations that are more or less susceptible to noise generation.

This Wi-Fi based footfall counting methodology offers a large number of applications and benefits for real time spatial analysis. Since Wi-Fi based sensors are inexpensive and the data model is scalable, it is possible to use this methodology for a large network of sensors to gather granular data on pedestrian footfall. A snapshot showing a week's worth of precise footfall in area around Charring cross, London is shown in Figure **??** in order to demonstrate the potential for such a dataset. Projects such as SmartStreetSensors [Longley et al., 2018], may utilise this methodology to overcome the challenges introduced by the implementation of MAC address randomisation.

The vicissitudes of MAC randomisation, and the provisions of privacy legislation such as EU General Data Protection Regulations mitigate against tracking individuals across the smart city using this approach. This can be modelled using agent-based methods [Heppenstall et al., 2011], however. In our own research we have also begun to link store time-lagged till receipts to footfall, and have used such data to better understand the dwell times that characterise such different retail uses as stores with window displays and fast food restaurants. Such analysis not only provides a more nuanced picture of movement through retail areas, but also enables valorisation of micro sites within retail centres. In the UK, for example, this is of immediate practical importance in evaluating business rates on properties, and has still wider implications for the setting of retail unit rental values. There are obvious extensions to understanding the ebbs and flows of activities in the 24-hour smart city such as understanding urban mobility [Gariazzo et al., 2019] and conceptualising them with a people dimension [Nam and Pardo, 2011].

More broadly still, extensions to this strand of smart city research are likely to seek to differentiate the quality of different elements within footfall according to mission e.. travel to adjacent workplace zones, leisure, etc. and personal characteristics such as spending power. In this respect, future research may not only simulate linkage of harmonised footfall counts between sensor locations, but also link these in turn to disaggregate origin-destination matrices for bikeshare and other public transport modes. Our own investigations will consider these and other challenges to understanding the functioning of the sentient city.