Temporal and Spatial Analyses of Street Crime in Inner London

by Blae Quayle

Springboard Foundations of Data Science
April-June 2017

Problem Definition

- Metropolitan Police operate across 32 boroughs and aim to make London the ‘safest global city’
- No accessible method allowing residents and tourists to check for crime hotspots
- Allow for informed decisions and encourage preventative action

Project Aims

- Investigate temporal and geographical patterns of street crime occurrences
  - Annual variations
  - Seasonal variations
  - Identify crime hot spots
- Create visualisation allowing identification of risk level for specific crimes
- Can a model be built to predict the expected number of crimes in a borough?

Data Extraction

- Street level crime for London is available from data.police.uk
- Reported monthly and available from Dec 2010 to Mar 2017
- Associated API implemented as standard JSON web service:

  https://data.police.uk/api/crimes-street/all-crime?poly=52.268,0.543:52.794,0.238:52.130,0.478&date=2013-01

Geographic and temporal extents

Data Exploration

- Data are stored at a crime incident level i.e. one observation for each incident with unique identifier
  - Almost 500,000 observations for central London polygon for full time period
- 11 variables returned by API – 4 are of primary interest: category, month, longitude, latitude

Data Limitations

- Anonymisation – no day/time provided and no specific location to protect victim confidentiality
- Integration with other datasets – using initial polygon returned dataset over arbitrary area, not at borough level
- City of London not included in dataset as not policed by the Met

Data Wrangling 1

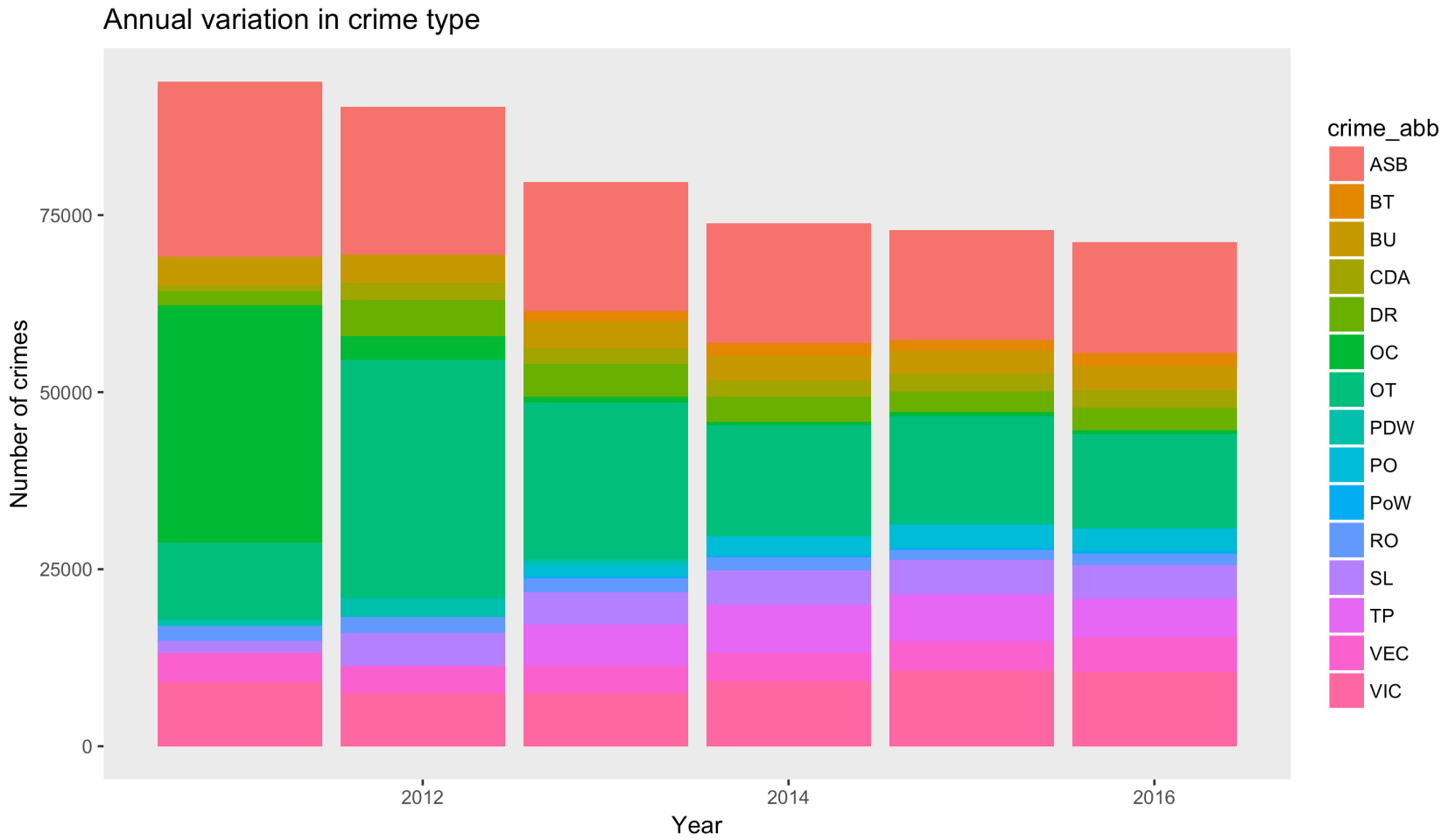
- Loaded **dplyr, tidyr** and **readr**
- Removed variables with no useful civilian data using *select()*
- *Rename()* category as crime_type and location_type as service
- Set category and location_type to factor variables using *as.factor()*
- Check for missing values using *is.na()*
- **Lubridate** package used to set *month()* and *year()* as separate variables

Data Wrangling 2

- Create additional data frames for analysis using *mutate(), case_when(), group_by(), summarize()*
  - Monthly crime
  - Annual crime
  - Abbreviation of crime_type for clearer visualisation e.g Anti-social behaviour = ASB
  - Seasonal crime – addition of 'season' variable
- Load **rgdal, rgeos** and **sp**. Filter borough shapefile using *subset() and* convert 2016 data to spatial points dataframe, remove all crimes outside of 12 boroughs

Data Visualisation 1: geom_col()

2011-2017
"Central"

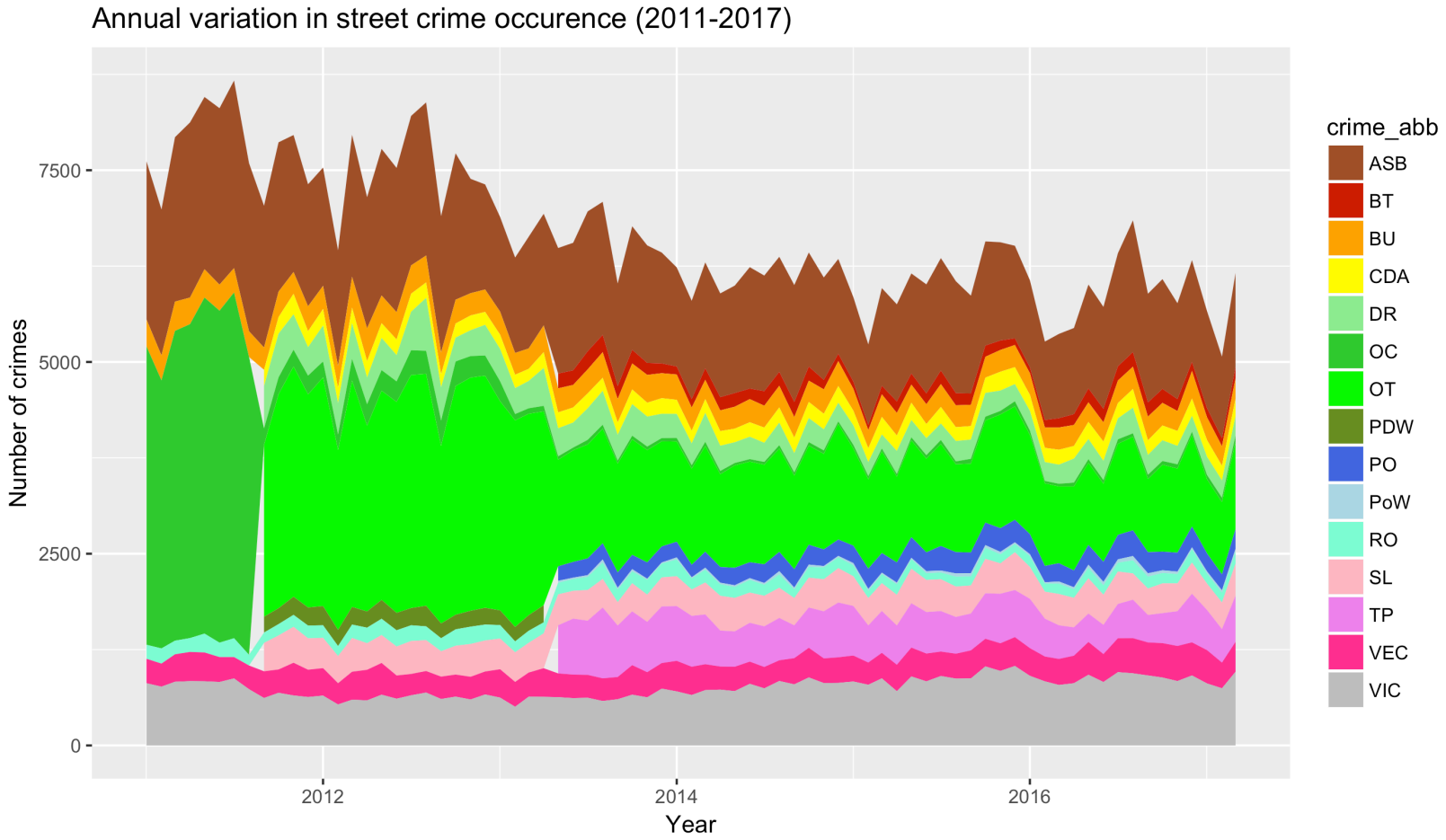


Decreasing since 2011, largest drop of 12% between 2012 and 2013

Other crime, other theft and anti-social behaviour largest decrease

Data Visualisation 2: geom_area()

2011-2017
"Central"

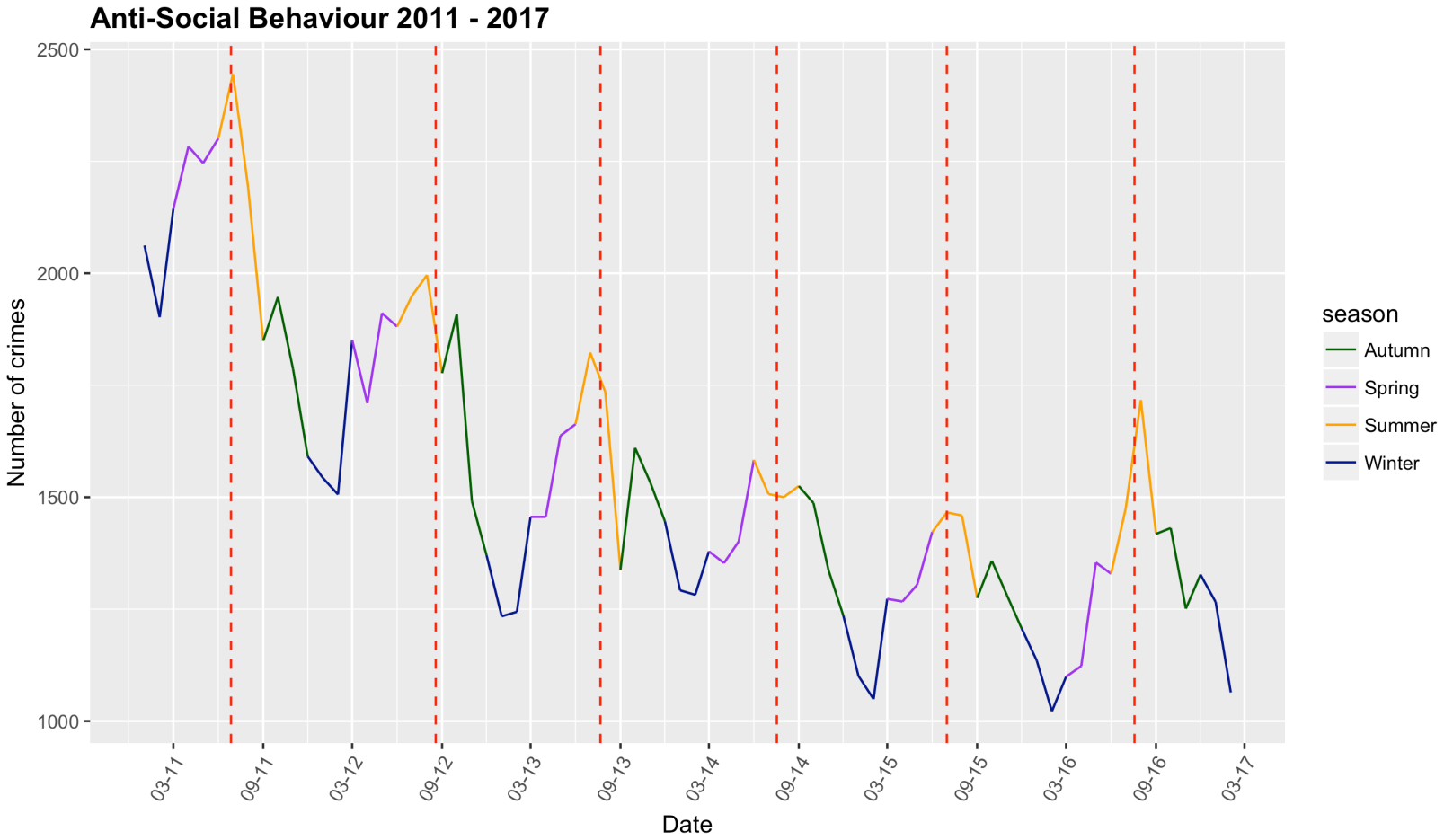


Identifying seasonal trends is difficult when all crimes are combined

Anti-social behaviour appears to show peaks most summers

Data Visualisation 3: geom_line()

2011-2017
"Central"

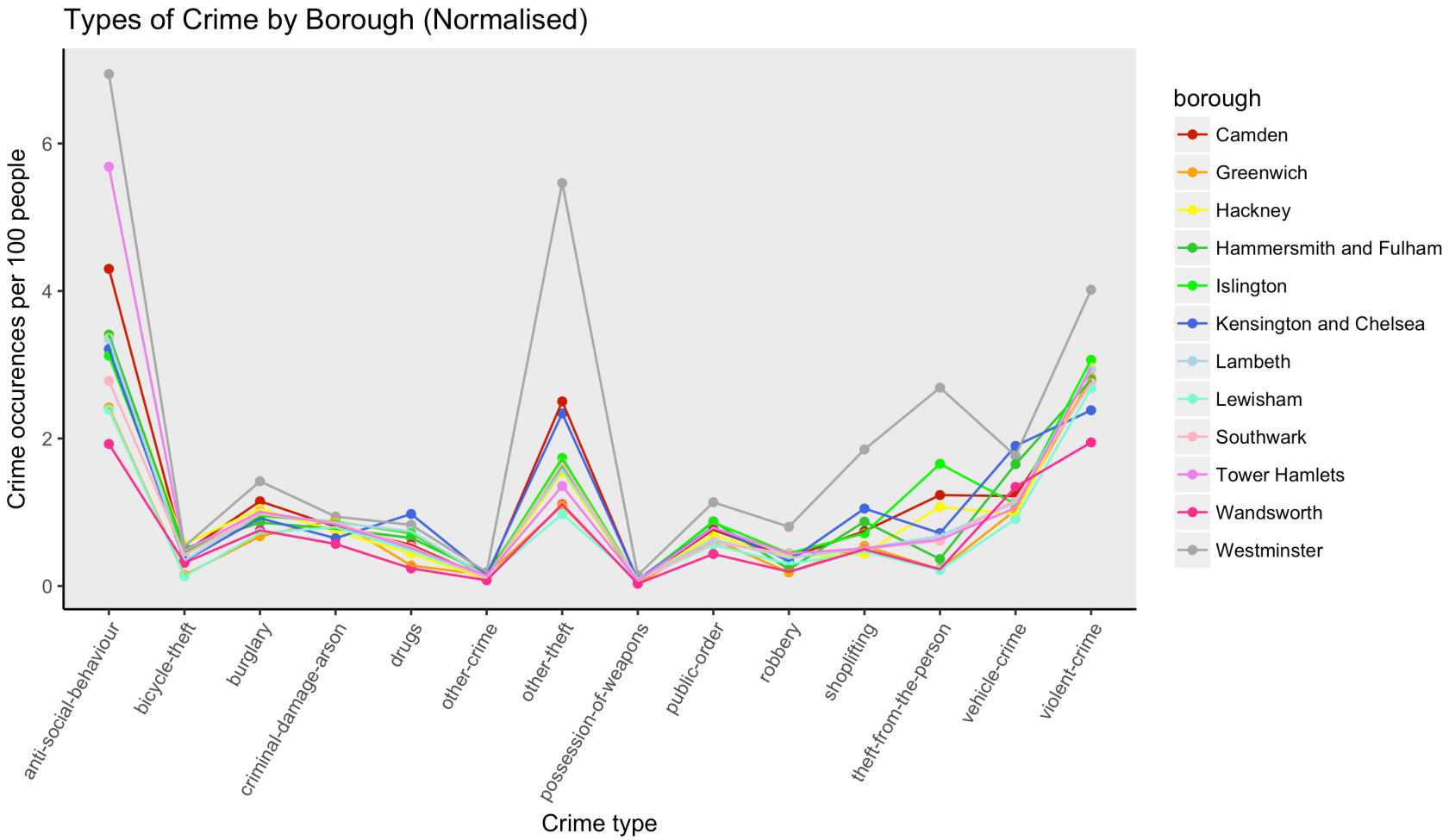


Decrease in anti-social behaviour from 2011 until 2016

Clear peaks visible usually coinciding with school summer holidays

Data Visualisation 4

2016
"Inner Boroughs"

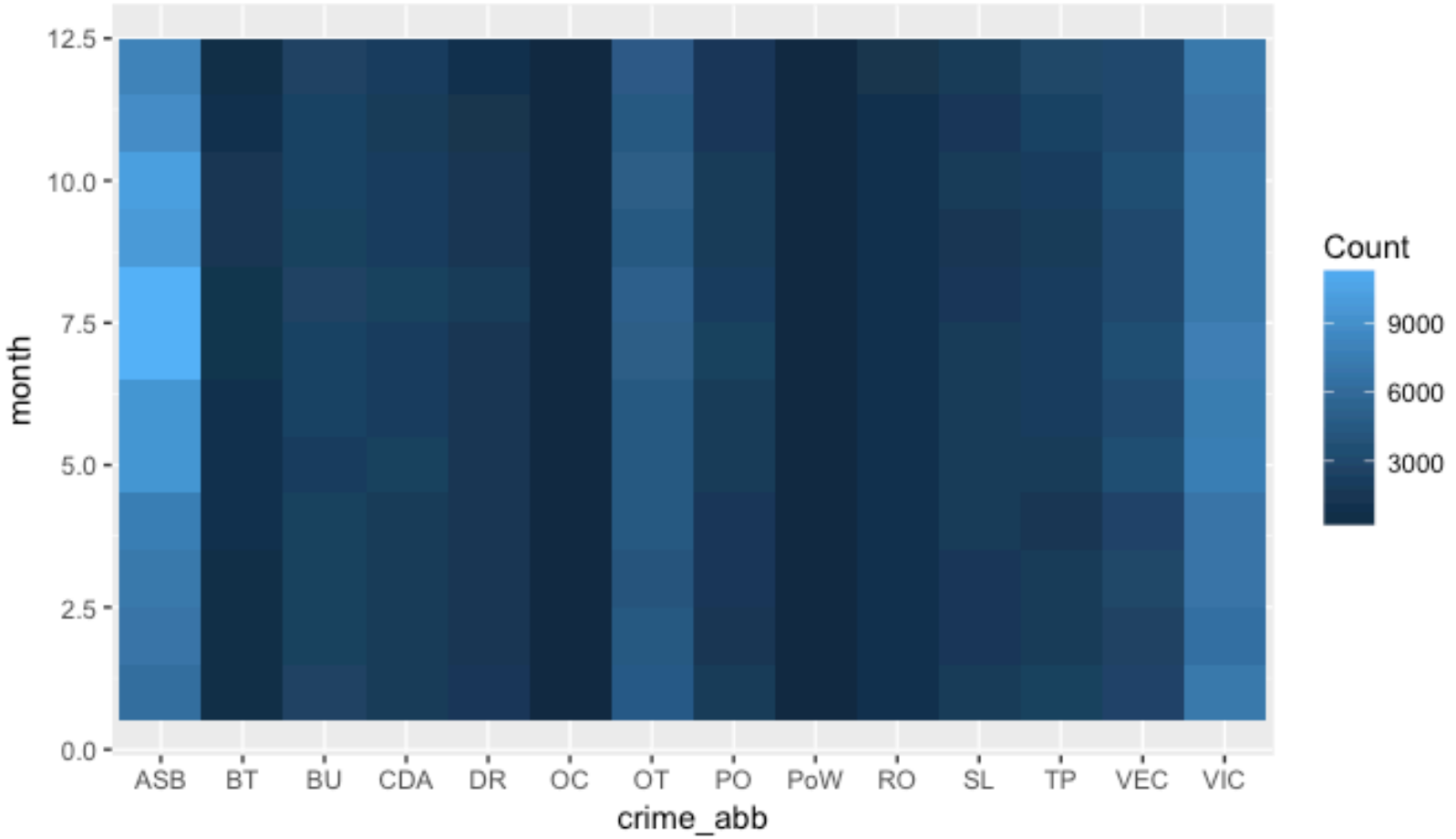


Westminster highest rate 28.7 crimes/100 people, however also borough with most tourists

Wandsworth lowest rate 9.7 crimes/100 people

Data Visualisation 4

2016
"Inner Boroughs"

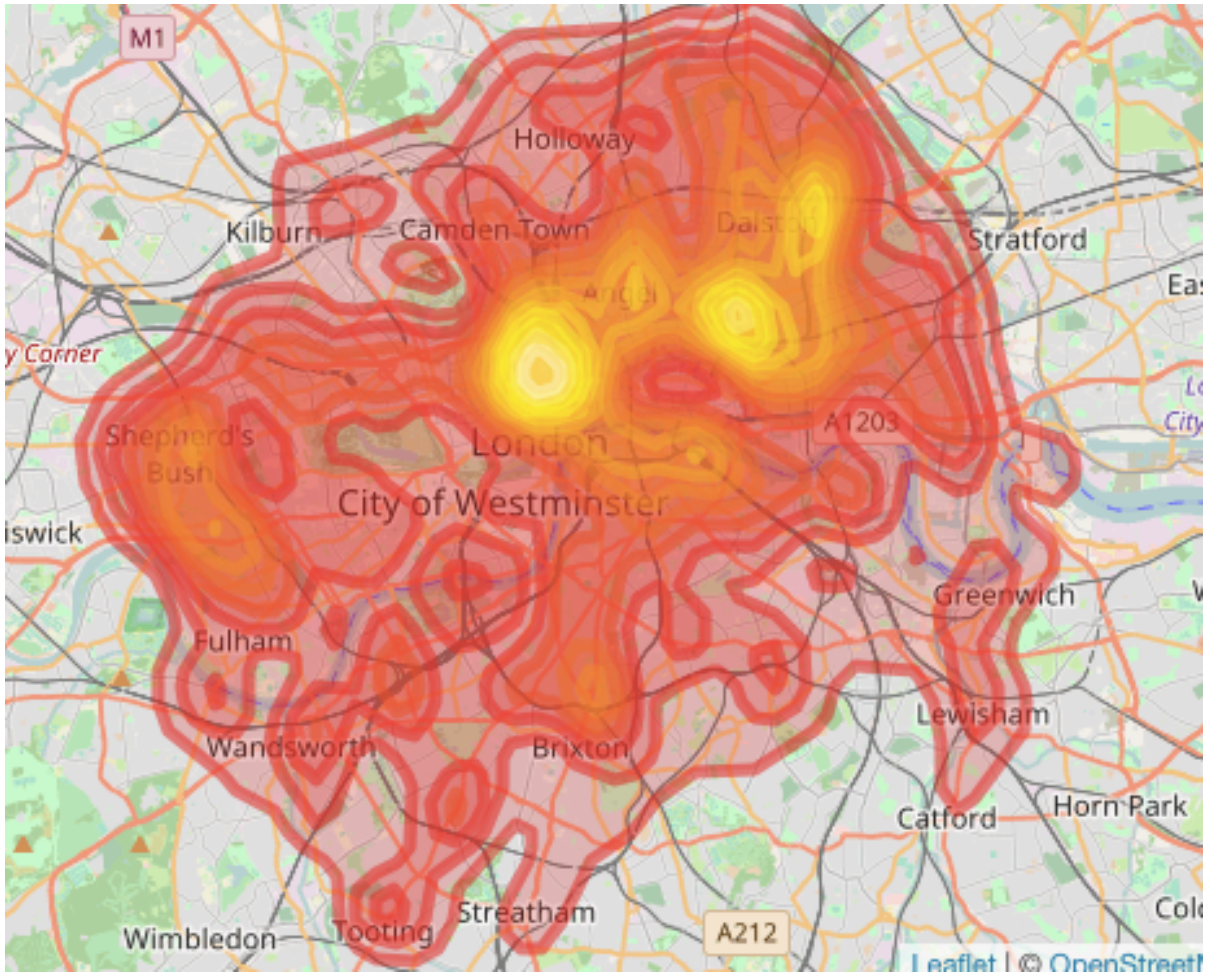


Heatmaps can be used to get an overview of seasonal variation across all crimes. Summer peaks in anti-social behaviour clear, but also violent crime and other theft more prevalent

Crime hotspots 1

2016
"Inner Boroughs"

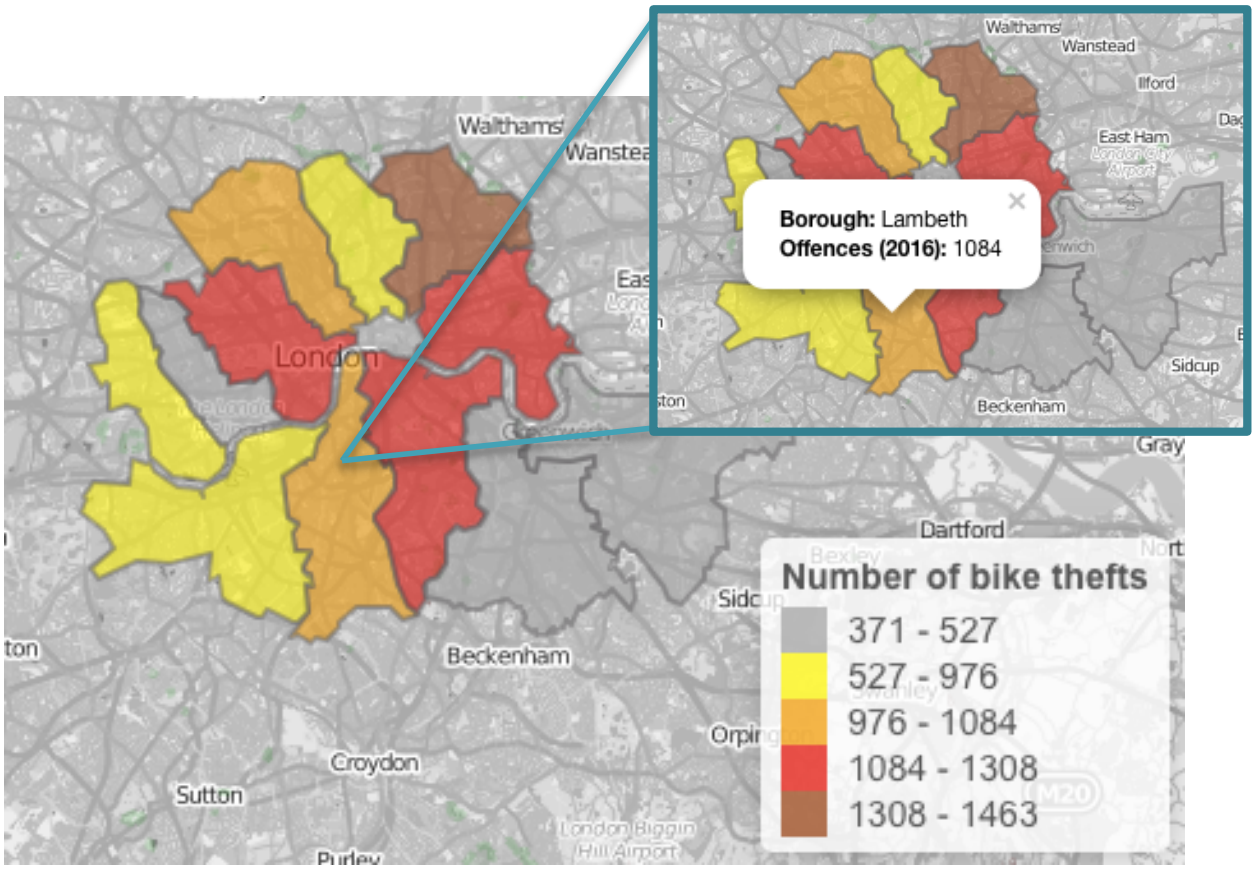
Using **KernSmooth, sp** and **Leaflet** to create Kernel Density Estimation maps for bicycle thefts - high density around Russel Square (SOAS, UCL and Birkbeck Universities), Shoreditch and Dalston



Crime hotspots 2

2016
"Inner Boroughs"

Using **maptools, classInt** and **Leaflet** to create interactive chloropleth maps for bicycle thefts – highlights Westminster, Southwark and Tower Hamlets as boroughs with highest number of bike thefts



Point pattern analysis

2016
"Inner Boroughs"
Burglary
November

- Most isolated occurrences of burglary tend to occur on the periphery of the inner London boroughs or within the parks.
- G and F functions indicate clustering, confirmed by the VMR.
- Key constraint of the dataset used is the lack of unique geographic location resulting in more than half the records being discounted from the analysis as they are non-unique.
- The P value found using the Kolmogorov-Smirnov test is very small (2.2e-16) - the null hypothesis of a continuous distribution can be rejected.

Conclusions 1

- Crime rates have been dropping consistently from 2011, with a high decrease between 2012 and 2013 noted. However, certain crimes have shown an increase over that time e.g. violent crime.

- Seasonal patterns in crime present
  - Peaks of anti-social behaviour correlating to the highest temperature peaks in the summer
  - Bicycle theft cyclical with peaks in the summer and troughs in the winter, against a background of fairly constant thefts
  - Burglary broadly decreasing since 2012 but peak often visible in the run up to Christmas followed by a sharp New Year drop.

Conclusions 2

- KDE is a valuable method for visualising the density of crimes when displaying a very large dataset
- Particular due to the location anonymisation meaning many crime observations overlie
- Russel Square and Shoreditch were identified as bicycle theft hotspots, with Soho and Fitzrovia as burglary hotspots, and Covent Garden and Soho for theft from the person.
- Westminster, with one of the lowest populations, displays the largest number of crimes at 28.7 per 100 people.

Recommendations

1. Making the Shiny App publicly available could encourage both residents and tourists to check crime hotspots. This would support the Met's aim of making London the safest global city and puts data directly into the hands of the general public, allowing them to take informed decisions.

2. For tourists, always be on high alert for pickpockets when visiting tourist attractions in Westminster, especially the areas around Soho and Covent Garden, and when visiting Shoreditch.

3. For local residents, be extra cautious and vigilant if locking your bike up in the area around the universities in Russell Square or Shoreditch and for the 'safest' inner London borough move to Wandsworth

Further Work

- A potential feature which could be added to the KDE maps would be location services so user could identify current position automatically and return a risk rating from 1 to 5 for example.

- Chloropleth maps should be adjusted to display data normalised by population, so the effect of population is removed from the records.

- Ideally, the dataset would include the day and time of the month, plus the specific location, as made available within crime datasets in other cities. Additional granularity would widen the scope for predictive machine learning methods to be used.