

# Topic 6.2: Open Data

Samuel Langton (MMU) & Reka Solymosi (University of Manchester)

June 2020

## Introduction

Access to data in crime and place research, and even criminology more broadly, has traditionally been reserved to those who have the means to collect fresh data themselves, pay for access, or obtain data through formal data sharing agreements. Even when access is granted, the usage of these data often comes with conditions that circumscribe how the data can be used through licensing or policy (Kitchin 2014). The public dissemination of findings which emerge from analysis might even be subject to restrictions. This can lead to unequal access, controlled usage and curb the diffusion of findings, severely limiting the insight that can be attained from data.

Open data are a response to these shortcomings, broadening access and participation in research, removing the requirement for permissions, formal agreements and negotiations (Manovich 2011). Not only this, but open data can also lead to all sorts of novel insight within criminology and crime and place research, tapping into constructs and processes which are difficult to capture through surveys, interviews and other traditional measures (Solymosi and Bowers 2018). As such, it is important that social scientists, researchers, crime analysts, and others interested in making sense of the social world around them have the skills and know-how to access, accurately interpret, assess and analyse open data sets.

This chapter aims to lay the groundwork for developing such skills by providing a framework to approach and meaningfully interpret open data, in terms of understanding its origin and the implications this might have on the data itself. The chapter also offers a practical hands-on guide to accessing, wrangling, and analysing open data in order to draw conclusions about crime and place. We achieve this by first giving a background on open data, discussing the key types of data out there, critically assessing the strengths and limitations, and building a how-to guide checklist for researchers to follow when approaching an open data source. Finally, we work through an example that shows how to access, wrangle, link, and interpret open data from various sources, providing a template which can be applied in other research areas.

## Background

### What is open data?

First, it is useful to clarify what we mean when we refer to ‘open data’. The *Open Data Handbook*, compiled by the Open Knowledge Foundation, states that open data are “data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike” (Dietrich et al. 2009). Specifically, open data can be defined along three key domains (Foundation 2016).

- **Availability and Access:** the data must be accessible via a public domain, at no more than a reasonable reproduction cost, and through a convenient medium, such as the internet. Ideally, it should be machine-readable and modifiable. For example, it should be downloadable from a website as an unlocked spreadsheet, or json file, rather than presented as a summary table in a PDF document.
- **Re-use and Redistribution:** the data must be provided under conditions that allow free use, such as modification, separation, or compilation of the data and permit re-use and redistribution including the intermixing (merging) with other datasets.

- **Universal Participation:** everyone should be allowed to use, re-use and redistribute the data and its derivatives without restriction. For instance, restrictions that only allow non-commercial or educational usage would not constitute open data.

In short, ‘open data’ should be data that are readily and publicly available, in a usable format, and without restrictions on usage, modification and re-distribution.

Open data practices, and the extent to which it is a reality, vary both within and between countries. For example, the United States have a history of making public sector data sets openly available. The United Kingdom can be more strict, releasing data under a licence (‘Open Government Licence’), with some data requiring a fee (Kitchin 2014). The *Global Open Data Index* (Foundation 2020) is one tool for tracking the annual global benchmark for publication of open government data by countries. It providing a rank which indicates how well governments across the world follow open data practices on aggregate, but also according to specific domains. Many of these have direct relevance to crime and place research, such as administrative boundaries and geographic locations (e.g. coordinates).

## What are types of open data?

Now we have a reasonable idea about what open data is in a broad sense, we can consider the different types of open data which you might come across. Here, we formulate a typology of open data sets based on their origin to help provide an overview.

### Public sector (including publicly funded research).

Much of the open data movement has focused on information generated by local and national state agencies, or publicly funded research. Open data has become a key component of transparency and accountability in government, given that much of the data collected has been funded by the public purse (Kitchin 2014). In some countries, the national statistics body is independent from government, which serves to improve accountability and increase public trust in the data being collated and published. The *Global Open Data Index* introduced earlier focuses on identifying the gaps in governmental organisations, encouraging them think about how public sector information can become more usable, and ultimately, more impactful (Foundation 2020).

As researchers studying crime and place, public sector data is particularly important. Many variables thought to explain the non-random distribution of crime, offenders and victims in space are collated and published by governmental bodies. For instance, demographic characteristics about resident populations can be collected via national censuses, with data usually released at aggregations which constitute ‘neighbourhoods’ - a theoretically important scale in criminology. Data describing urban structure, such as street networks, are often generated and distributed by public bodies, and can be used in street syntax analysis to assess whether the configuration of streets dictates things like violent crime victimisation (Summers and Johnson 2017). In many countries, victimisation surveys are openly available for download. This is not only useful for substantive research (e.g. quantifying long-term crime trends), but also scrutiny over the reliability and accuracy in other forms of open data, such as police recorded crime records. Both of these examples represent an interesting challenge to open data principles, since data tends to be anonymised to some degree before being made available to the public. Access to the raw records is dependent on a number of requirements (e.g. background checks, contractual agreements, ICT security measures) which might exclude many individuals due to cost and practicality.

The drive towards transparency and scrutiny of public information is further enhanced by an increasing need for replication - a hallmark for open science research practice. Opening up data sets used in criminology publications and wider social sciences fosters and facilitates a culture of replication (Pridemore, Makel, and Plucker 2018). As such, open data fits into the wider discourse around transparency, open review and open scrutiny, which (in time) may even become policy requirements irrespective of the research field or the source of funding (Vuong, n.d.). Increasingly, researchers are making use of online repositories for their papers (e.g. [www.osf.io](http://www.osf.io)), data (e.g. [www.ukdataservice.ac.uk/deposit-data](http://www.ukdataservice.ac.uk/deposit-data)) and code (e.g. [www.github.com](http://www.github.com)), all in the name of transparency and replication. These are guidelines and tools worth considering, not just

when conducting analysis on secondary data sources, but also collecting and sharing your own data for research projects.

### **Private sector**

Opening up data generated by the private sector represents a significant challenge, largely due to the proprietary value to its creators (Kitchin 2014). The aims and objectives of private companies, who have a duty to shareholders and operate in a competitive market environment, differ considerably from local and central government. That said, some datasets are released openly by private sector organisations, albeit often only a subset of what would be available as a paying customer. The website for ArcGIS, a piece of software maintained by the Environmental Systems Research Institute (ESRI), a private company, host a number of open geographic datasets on their online cloud platform ([www.hub.arcgis.com/search](http://www.hub.arcgis.com/search)). Indeed, many papers have made use of data collected or distributed by private organisations to explore crime and place, such as Google (Langton and Steenbeek 2017) and Twitter (Malleon and Andresen 2015). In this book you will learn to use the free Twitter API for such purposes in criminology (see Section 3, Topic 7).

### **Open crowdsourced data**

Finally, there is a specific group of open data sources that collate data collectively, generated by large groups of individuals who do not specifically belong to any organisation, but instead work together on a collaborative project. These are crowdsourced data sets. Examples include Wikipedia ([www.wikipedia.org](http://www.wikipedia.org)), an online encyclopedia where anyone can contribute, and Flickr ([www.flickr.com](http://www.flickr.com)) an online photo gallery where people can upload and tag their photos. Often, the organisations who maintain and monitor these data collection activities are charities or non-governmental organisations that do not operate for profit, but instead provide some form of social good. An example is the online reporting platform FixMyStreet ([www.fixmystreet.com](http://www.fixmystreet.com)), where people can report problems such as instances of graffiti, vandalism or environment issues. The data has been used to explore signal crimes theory, for instance, offering insight into people's experiences with incivilities (Solymosi, Bowers, and Fujiyama 2018). Another chapter in this book (see Section 3, Topic 6) discusses the merits and pitfalls of crowdsourced data, and what to watch out for when analysing open data of this type.

### **Strengths and limitations**

Each one of the above categories might have its own strengths and limitations, and it is very important to consider the source of the data you are working with, and what might mean for what and who is represented or excluded from these data. Here we will provide a general overview, and some guidance on how to make sure you take all necessary precautions when analysing open data of any kind.

Open data is hailed for its ease of use afforded by its availability and access; there is no need to ask for permissions or negotiate data access in order to use it for research (Manovich 2011). In this sense, the great advantage of this data is that it is already out there. To collect it, all that is needed is a way to be able to interpret what this data means (for example by applying a framework for its analysis), and some skill in data scraping, wrangling and cleaning, in order to be able to transform it into a usable format for research (Boyd and Crawford 2012).

But besides being easy to access, there are further advantages to using such data. One specific motivation for using open data comes from its potential to address many limitations associated with traditional surveying methods, such as social desirability bias, or issues associated with memory and recall (Mayer-Schönberger and Cukier 2013). By generating data often as a byproduct of other activities, we might gain more honest insight into people's everyday lives and associated social processes (Solymosi and Bowers 2018). At the same time, there are new threats to validity that may emerge in these sorts of data, and researchers should be careful in interpreting people's communication online as authentic (Manovich 2011).

A related, and key concern about these data relates to issues about sampling, and ultimately the generalisability of any findings that can be drawn from them. In the case of crowdsourced data, where the sample of the population who contributes to such data is self-selected, giving way for people more motivated to speak

about the issue, which means the data are likely to contain inherent bias (Longley 2012). Specifically it tends to be men, between ages of 20-50, with a college or university degree who are most likely contributors (Budhathoki 2010; Haklay 2010).

Looking into what contextual factors influence participation in Open Street Map, Mashhadi, Quattrone, and Capra (2013) found that, socio-economic factors such as population density, dynamic population, distance from the centre and poverty all play an important role. These are important to keep in mind when reporting findings based on analysis of such data.

Despite these limitation, the social sciences must embrace these new forms of data that, although messy, biased and noisy, have the potential to describe social phenomena better than well-organized small surveys or even national censuses (Savage and Burrows 2007). The inclusion of crowdsourced data which are both up-to-date and specific to the problem at hand can still provide new insight in addition to the knowledge from established sources of data collected in traditional methods (Birkin et al. 2011). All of this helps us to understand the relations between the events and the occurrences that come together in unique places and provides a framework for understanding how places change over different time periods (Longley 2012).

---

In sum, researchers and analysts using open data must first ask critical questions about:

- **where** does this data come from?
- **why** was the data collected in the first place?
- **who** is represented in the data, and who is excluded?
- **what** concepts and constructs can and cannot be operationalised with this data?
- **when** did data collection take place, and how might have that influenced results?

And meaningfully engage with the answers to these questions during conceptualisation and operationalisation phases of any research project.

---

## Practical exercise

In this exercise we will demonstrate:

- how to access open data using Application Programming Interfaces (APIs) and through downloading from official sites
- an example of how to clean and wrangle data for analysis
- how to link distinct data sets
- how open data from different sources may tell us different things, and why engaging with the critical questions outline above is crucial to properly use open data in research and analysis.

## Our aim

We are interested in finding out how bus stops might affect crime.

## Accessing data

We will be using two different types of open data in this exercise, public sector data (crime data and transport authority data) and crowdsourced data (from Open Street Map). To access them we will use three different methods: 1) Direct download, 2) Direct request to an API, 3) Request to an API using a wrapper. This will give you few different ideas about how you might go about accessing other open data sets relevant to your research.

## Direct download

The simplest way that open data is often made available is through direct download from a website. This is the case for open police data made available in the United Kingdom through the `data.police.uk` webportal. To access this data, simply type the url into any web browser ( `https://data.police.uk/` ), and follow the guidance to acquire data. In the case of `police.uk` you select the tab which says “Data”, choose your “Date range:” (ie for what months you would like to download crime data for), choose the police forces you want crime data from, choose some options (whether you want to include crime data, outcomes data, and/or stop and search data) and click on ‘Generate File’ to download.

Once you download this data you will have to save this locally on your computer. We have saved ours in a subfolder called “data” and will be loading the data into R directly from there.

Some data downloads may come bundled with a data dictionary, explaining important information about your data set, which you will need to answer the critical questions outlined in previous sections. In other cases this might be found online. This is the case with the `police.uk` dataset, where the data dictionary can be found on the “About” page. Make sure you always read through this correctly so you can answer your critical questions and fully understand your data set.

## Direct request to an API

An application programming interface (API) is a tool which defines an interface for a programme to interact with a software component or system, for example, what sort of requests or calls can be made to it, and how these can be made. We are specifically using the term APIs to denote tools created by an open data provider to provide access to different subsets of their content. APIs facilitate scripted and programmatic extraction of content, as permitted by the API provider (Olmedilla, Martínez-Torres, and Toral 2016). APIs can take many different forms and be of varying quality and usefulness (Foster et al. 2016). For the purposes of accessing open data from the web, we are specifically talking about **RESTful APIs**. REST stands for Representational State Transfer. These APIs work directly over the web, meaning that we as users can play with the API with relative ease to understand how it works (Foster et al. 2016).

Here we will use the example of the British transit authority Transport for London (TfL) to practice constructing HTTP requests via their API, and then parsing the data that are returned. Much like how downloading data directly requires access to a data dictionary to make sense of the data and answer our critical questions, we must find a similar document to understand how the API we plan on using works. In the case of the TfL API this is provided via the TfL Open Data page for their unified API and the documentation page. We will engage more with this in the exercises.

## Request to an API using a wrapper

Often developers who work with APIs will share their code for doing so, and release them in the form of a package or module. This is called a **wrapper** because it’s code that wraps the API in a nice package, making it easier for future users to access. The wrapper can take many forms, it could be a Python module, or an R package, or it can even be a web interface that provides a graphical user interface (GUI) for accessing the API in question.

To demonstrate this, we will be accessing data from Open Street Map (OSM). OpenStreetMap is a crowd-sourced open data set of spatial information built by a community of mappers that contribute and maintain data about all sorts of environmental features, such as roads, trails, cafes, railway stations, and much more, all over the world. The result is a map rich in local knowledge, which provides valuable information about places and their features.

OSM has two types of wrappers available for its API, a web-based GUI called Overpass Turbo, and an R package called `osmdata`. We will explore using both.

## Walk-through

Let's get started! Getting back to our problem at hand, crime at bus stops, we want to know how crimes and bus stops interact...

First, ensure that you have the relevant packages installed in R. Although the main packages used in this demonstration are `rjson` for parsing the TfL API results and `osmdata` for querying the Open Street Map API, we use a number of additional packages for data handling and visualisation. If you don't have these packages installed, use the `install.packages()` function prior to loading each one with `library()`.

```
library(jsonlite)
library(osmdata)
library(ggplot2)
library(sf)
library(readr)
library(dplyr)
library(tidyr)
```

Now let's focus on acquiring our different data sets

### Direct Download of crime data

#### Direct request to TfL API

Our first step is to read through the . The tells us what sort of data is possible to acquire. We know that currently we want to find the location of bus stops, and any information that is available from TfL about these bus stops. In the documentation is where we find specifically *how* we can request this.

Specifically, we can scroll down to where we find ". Next to this we see the example URL: <https://api.tfl.gov.uk/line/24/route/sequence/outbound> This example demonstrates how we can request this information, and gives us a template to alter, but also allows us to take a peak into the result this call will return.

To see, try to take the above URL, and copy and paste it into your browser window. What do you see? It should be something like this:

```
knitr::include_graphics("img/tfl_api_browser.png")
```

This contains the information we requested (bus route 24 stops outbound) in a format called **JSON**, which stands for JavaScript Object Notation. JSON is an open standard way of storing and exchanging data, and will most likely be the format in which data are returned from most API calls.

Great, but how do we get this into R? Well this is where the `rjson` package is handy, and specifically we can use the `fromJSON()` function to parse this JSON p=object, which we can read directly from the URL into R using the `readLines()` function (base R). So, if we wanted the results of the bus route 24, we would type:

```
api_call <- fromJSON(readLines("https://api.tfl.gov.uk/line/24/route/sequence/outbound"))
```

This gives us an object (`api_call`) which contains all the information returned by the TfL API. Now JSON is slightly different to dataframes which we are more used to dealing with, in that there is nested information

```
json_data_frame <- as.data.frame(api_call)
```

### Using wrappers to access Open Street Map data

All queries begin with a bounding box specification to define the study region. This can be obtained manually, which requires some existing knowledge about an area using the latitude and longitude coordinates, but it is generally easier to use a search term. Here, we select Greater London in the United Kingdom using the `getbb()` function, specifying that we want the content as a simple features (sf) polygon.

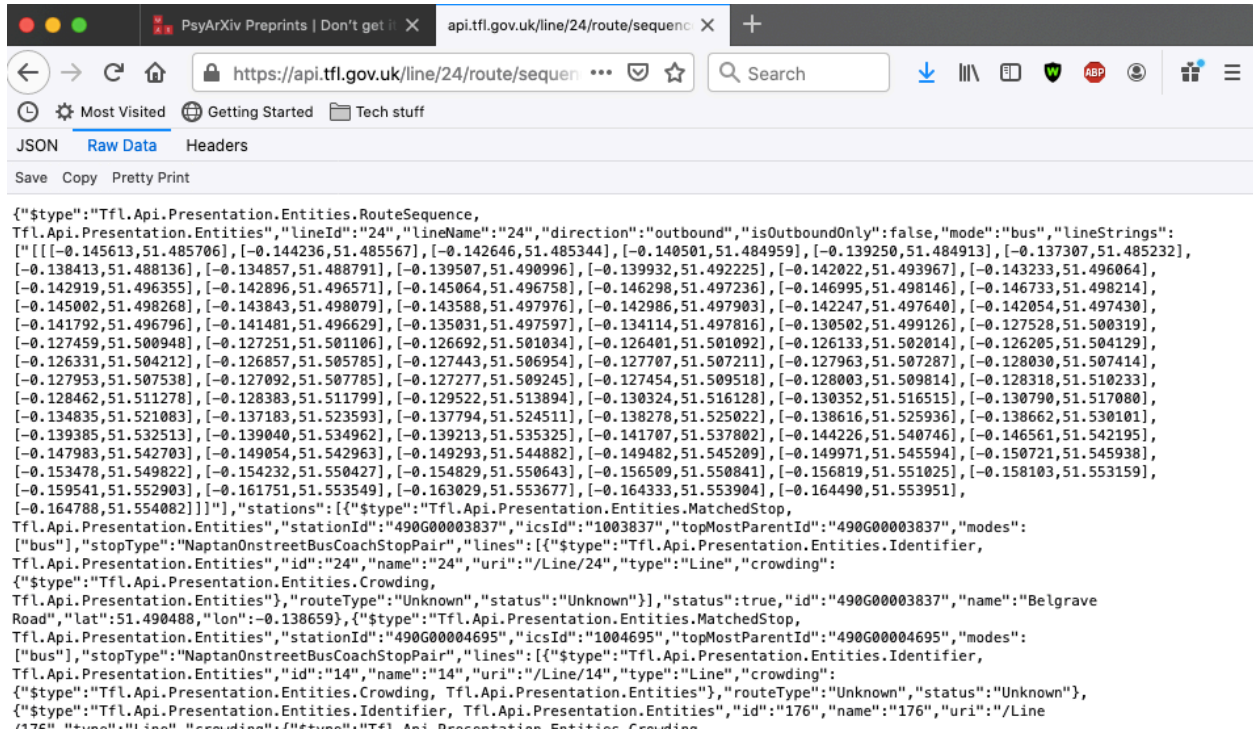
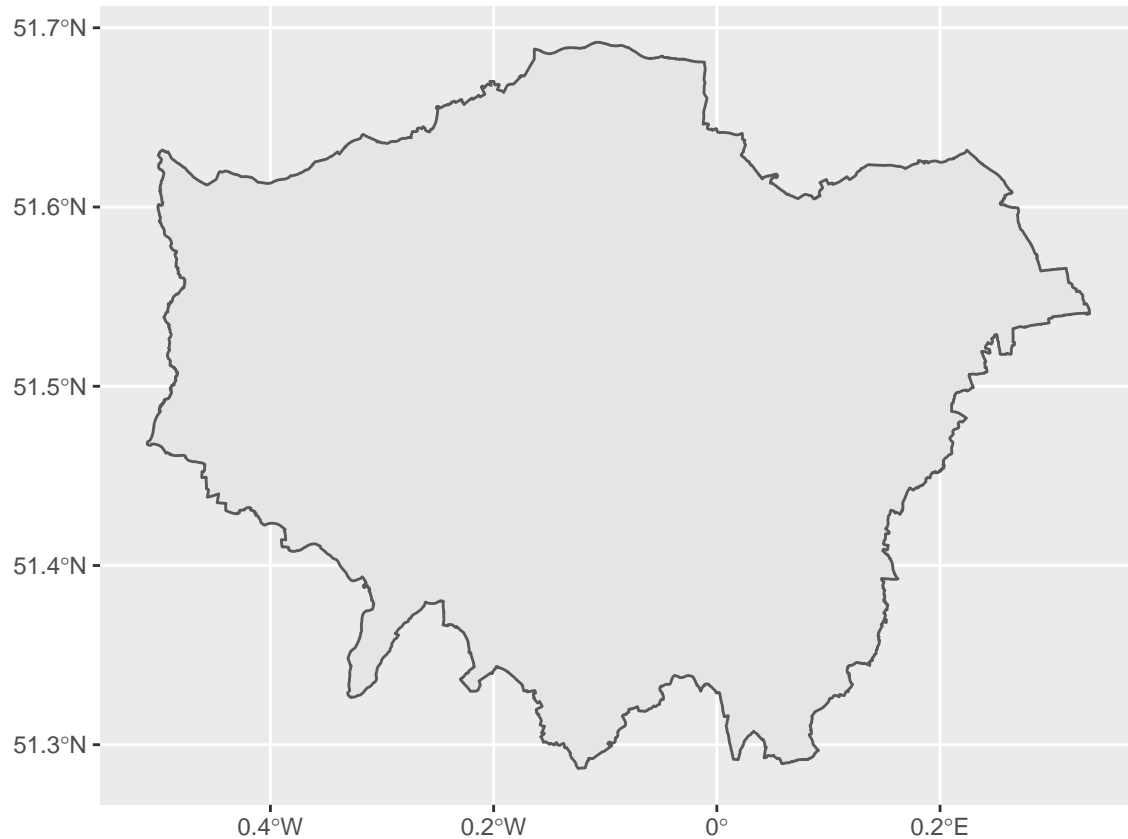


Figure 1: Data from TfL

```
bb_sf <- getbb(place_name = "greater london united kingdom", format_out = "sf_polygon")
```

We now have our study region defined as the administrative boundaries of Greater London. This can be visualised using `ggplot2` and the *simple features* geometry `geom_sf()` available with `sf`. Note that by default, the Coordinate Reference System (CRS) is the World Geodetic System 84.

```
ggplot(data = bb_sf) +  
  geom_sf()
```



Now we have our study region, we can scrape data from the OSM API using the `opq()` function, which is short for ‘Overpass query’. This allows you to build an Overpass query, outlined in the previous section, from within the R environment. We specify the bounding box object which is our study area, and pass this through using a pipe `%>%` to `add_osm_feature()` in which we define what we want to pull from the API. As we noted earlier, features in OSM have are defined through keys and values. Here, we specify that we want amenities (the key) defined as bicycle parking (the value). This query is then piped through to `osmdata_sf()` which ensures that the resulting object is a *simple features* class for easy plotting with `ggplot2`. We trim the features pulled from the API using `trim_osmdata()` to ensure that everything stays within the boundaries of our study region.

```
bikes_sf <- opq(bbox = bb_sf) %>%                                # select bounding box
  add_osm_feature(key = 'amenity', value = 'bicycle_parking') %>% # select features
  osmdata_sf() %>%                                              # specify class
  trim_osmdata(bb_poly = bb_sf)                                # trim to region
```

The resulting object `bikes_sf` contains lots of information. We can view the contents of the object by simply executing the object name into the Console.

```
bikes_sf
```

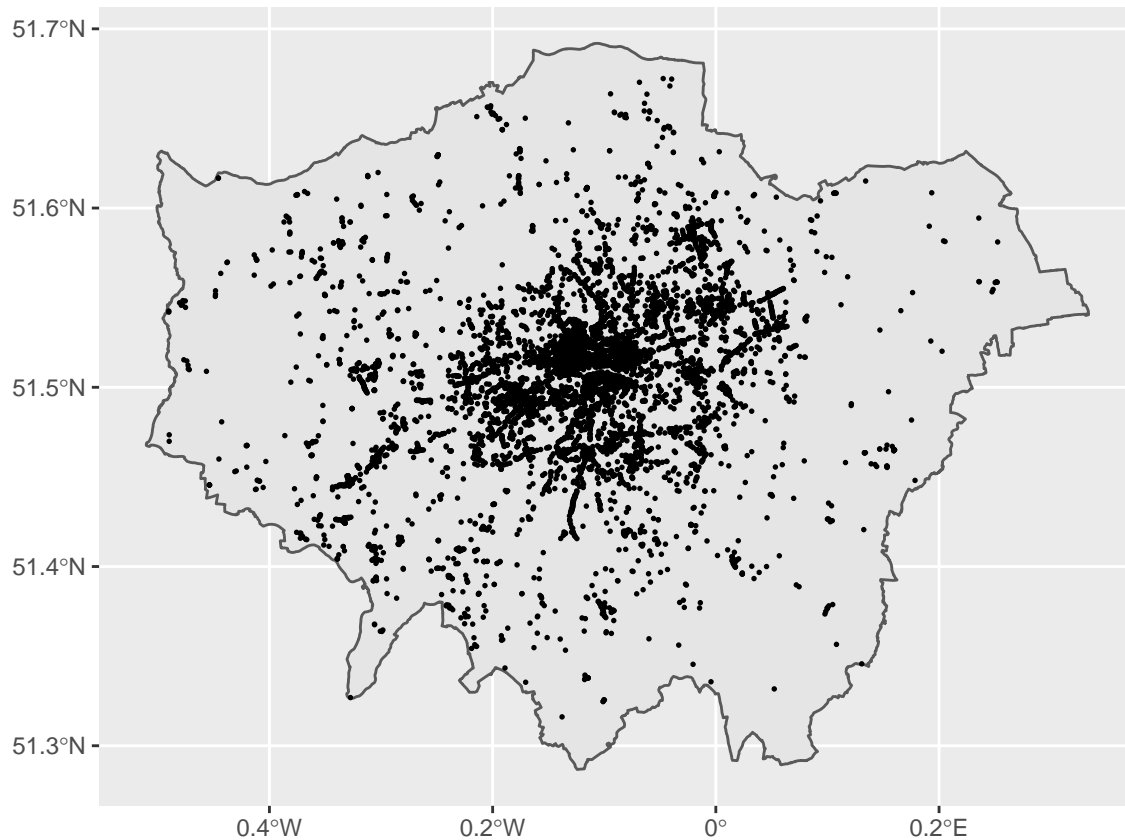
This confirms details like the bounding box, but also provides information on the simple features collected from the query. As one might expect, most information relating to bicycle parking has been recorded using points (i.e. two-dimensional vertices, coordinates) of which we have over seven thousand at the time of writing. We also have around one hundred polygons. For now, let’s extract the point information only and then transform the CRS to the BNG.

```
bikes_points_sf <- bikes_sf$osm_points
```



We can then plot these points over our original boundaries of Greater London. We reduce the default size of the point to ensure that we avoid too much overlap between bicycle parking locations.

```
ggplot() +
  geom_sf(data = bb_sf) +
  geom_sf(data = bikes_points_sf, size = 0.3)
```



As we can see, most bicycle parking spaces are clustered around the city centre, especially just north of the river Thames. It is also possible to make out key roads flowing in and out of the city centre, which contain bicycle parking all along the street.

Using open police recorded crime data we can then plot actual incidences of bicycle theft to explore whether there is a spatial relationship between bike theft and parking spots in Greater London. For this example, we just use crime recorded as occurring in January 2020. First, let's load in the data as it downloaded raw from <https://data.police.uk/data/>.

```
crime.df <- read_csv("data/2020-01-metropolitan-street.csv")
```

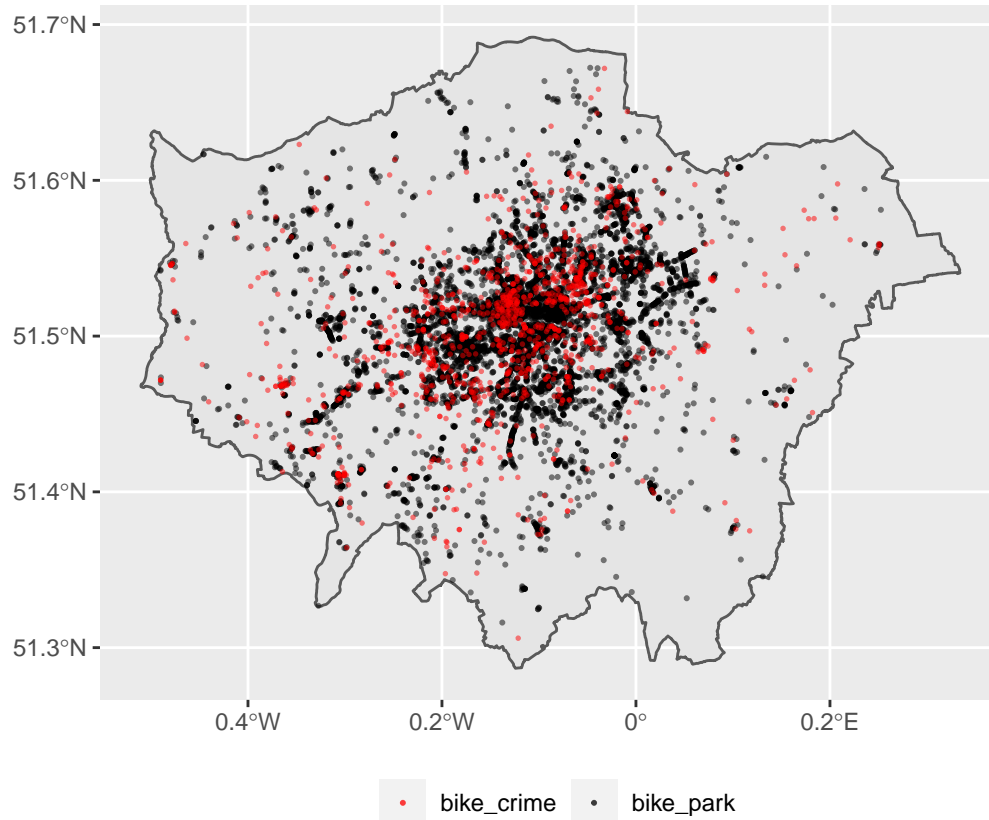
We then need to conduct a bit of preliminary data handling: filter crimes which were tagged as bicycle theft, convert the latitude and longitude columns to coordinates with *simple features*, state the WGS 84 CRS and then clip the points by our study region.

```
bike.crime.sf <- crime.df %>%
  filter(`Crime type` == "Bicycle theft") %>%
  drop_na(Longitude, Latitude) %>%
  st_as_sf(coords = c(x = "Longitude", y = "Latitude"), crs = 4326) %>%
  st_intersection(bb_sf)
```

To demonstrate the data in its entirety, we can plot the Greater London boundaries, overlaid with the

bicycle parking space locations, and open crime data about bicycle thefts. It is worth clarifying that open police recorded crime data in England and Wales is spatially anonymised by a process of snapping points to a pre-defined grid (see Tompson et al. (2015)). For that reason, many of these points overlap, and thus a degree of transparency is used for the points.

```
ggplot() +
  geom_sf(data = bb_sf) +
  geom_sf(data = bikes_points_sf, aes(colour = "bike_park"), size = 0.4, alpha = 0.5) +
  geom_sf(data = bike_crime_sf, aes(colour = "bike_crime"), size = 0.3, alpha = 0.5) +
  scale_colour_manual(name = NULL, values = c(bike_park = "black", bike_crime = "red")) +
  theme(legend.position = "bottom")
```



### Bus stop example

```
bus_sf <- opq(bbox = bb_sf) %>% # select bounding box
  add_osm_feature(key = 'highway', value = 'bus_stop') %>% # select features
  osmdata_sf() %>% # specify class
  trim_osmdata(bb_poly = bb_sf)
```

### Future of open data

- Threats to its sustainability (e.g. licence expiry).
- Prospects in crime of place research.
- Examples of cool projects (e.g. Colouring London).
- Suggestions for new avenues which can expand the field.

## Conclusion

- Re-cap on what we’ve covered.
- Wrap-up the key points.

## References

- Birkin, Mark, Nick Malleson, Andy Hudson-Smith, Steven Gray, and Richard Milton. 2011. “Calibration of a Spatial Simulation Model with Volunteered Geographical Information.” *International Journal of Geographical Information Science* 25 (8): 1221–39.
- Boyd, Danah, and Kate Crawford. 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” *Information, Communication & Society* 15 (5): 662–79.
- Budhathoki, Nama R. 2010. “Participants’ Motivations to Contribute Geographic Information in an Online Community.” PhD thesis, University of Illinois at Urbana-Champaign.
- Dietrich, Daniel, Jonathan Gray, Tim McNamara, Antti Poikola, P Pollock, Julian Tait, Ton Zijlstra, and others. 2009. “Open Data Handbook.” *Open Knowledge International*.
- Foster, Ian, Rayid Ghani, Ron S Jarmin, Frauke Kreuter, and Julia Lane. 2016. *Big Data and Social Science: A Practical Guide to Methods and Tools*. crc Press.
- Foundation, Open Knowledge. 2016. “Open Definition 2.1.” <https://opendefinition.org/od/2.1/en/>.
- . 2020. “Global Open Data Index.” <https://index.okfn.org/>.
- Haklay, Mordechai. 2010. “How Good Is Volunteered Geographical Information? A Comparative Study of Openstreetmap and Ordnance Survey Datasets.” *Environment and Planning B: Planning and Design* 37 (4): 682–703.
- Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.
- Langton, Samuel H, and Wouter Steenbeek. 2017. “Residential Burglary Target Selection: An Analysis at the Property-Level Using Google Street View.” *Applied Geography* 86: 292–99.
- Longley, Paul A. 2012. “Geodemographics and the Practices of Geographic Information Science.” *International Journal of Geographical Information Science* 26 (12): 2227–37.
- Malleson, Nick, and Martin A Andresen. 2015. “The Impact of Using Social Media Data in Crime Rate Calculations: Shifting Hot Spots and Changing Spatial Patterns.” *Cartography and Geographic Information Science* 42 (2): 112–21.
- Manovich, Lev. 2011. “Trending: The Promises and the Challenges of Big Social Data.” *Debates in the Digital Humanities* 2: 460–75.
- Mashhadi, Afra, Giovanni Quattrone, and Licia Capra. 2013. “Putting Ubiquitous Crowd-Sourcing into Context.” In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 611–22.
- Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.
- Olmedilla, M, M Rocío Martínez-Torres, and SL Toral. 2016. “Harvesting Big Data in Social Science: A Methodological Approach for Collecting Online User-Generated Content.” *Computer Standards & Interfaces* 46: 79–87.
- Pridemore, William Alex, Matthew C Makel, and Jonathan A Plucker. 2018. “Replication in Criminology and the Social Sciences.” *Annual Review of Criminology* 1: 19–38.
- Savage, Mike, and Roger Burrows. 2007. “The Coming Crisis of Empirical Sociology.” *Sociology* 41 (5): 885–99.

- Solymosi, Reka, and Kate Bowers. 2018. "The Role of Innovative Data Collection Methods in Advancing Criminological Understanding." *The Oxford Handbook of Environmental Criminology* 210: 210–37.
- Solymosi, Reka, Kate J Bowers, and Taku Fujiyama. 2018. "Crowdsourcing Subjective Perceptions of Neighbourhood Disorder: Interpreting Bias in Open Data." *The British Journal of Criminology* 58 (4): 944–67.
- Summers, Lucia, and Shane D Johnson. 2017. "Does the Configuration of the Street Network Influence Where Outdoor Serious Violence Takes Place? Using Space Syntax to Test Crime Pattern Theory." *Journal of Quantitative Criminology* 33 (2): 397–420.
- Tompson, Lisa, Shane Johnson, Matthew Ashby, Chloe Perkins, and Phillip Edwards. 2015. "UK Open Source Crime Data: Accuracy and Possibilities for Research." *Cartography and Geographic Information Science* 42 (2): 97–111.
- Vuong, Quan Hoang. n.d. "Open Data, Open Review and Open Dialogue in Making Social Sciences Plausible." *Nature: Scientific Data Updates*. Available%20at%20SSRN:%20<https://ssrn.com/abstract=3086667>.