Advanced Data and Coding

You are being watched

CCTV cameras watch our every move.

What for?

—

By Nicu Calcea

Interactive story: https://nicucalcea.github.io/cctv/

Code: https://github.com/nicucalcea/nicucalcea.github.io/tree/master/cctv

For this project, I decided to map all the CCTV cameras in London on top of a heatmap of crimes in the city.

After initial research, I found that each London borough operates its own security cameras. Most of these boroughs provide no information on their official website. Older FOIA requests revealed that most boroughs publish data that's difficult to map (just street names, proximity to various buildings, JPG maps, etc.). The rest of them argued in FOIA replies that this information is exempt from publishing. Only Islington, Camden, Hillingdon and Waltham Forest provide data with coordinates in some form.

Because I wanted to publish my story, I focused on Islington and Camden. I pitched the story to Islington Tribune and their sister paper, Camden New Journal and they accepted my pitch.

The first step was to download the data from the Islington and Camden council websites. Both provide online maps of CCTV cameras in JSON and XML respectively. Camden was a bit more challenging, as it only returned the 10 closest cameras for any particular address. To ensure I get all data, I downloaded the XML responses for each postcode query in Camden and extracted the unique locations.

Next, I imported data about cameras operated by TfL. TfL provides an "example feed" with about 70 of their cameras in XML. For the full information, I decided not to go through their official API as I couldn't find how to specifically request the information I needed. Instead, I used Table to Excel for Firefox to download an HTML table from a third-party website. While not as elegant, this was much easier. Since this data is provided for the entire London, I downloaded the boroughs outlines shapefile and filtered out the cameras outside Islington and Camden. This was one of the more difficult tasks, as I had to transform the types of data I have, mess with coordinates (geometrical to geographical), etc. I realised working with spatial data takes a bit more tinkering.

Next, I gathered data about crime in my boroughs. I didn't use the police API since it only gives out data about specific postcodes, so I just downloaded the entire archive for 2018 and filtered it.

I started to map my data with static images in ggmap. I still kept these maps, plus data from OpenStreetMap, for a scrollytelling experiment on the website. But the boroughs are quite big and I wanted to let people zoom in to see the area they live or work in more closely, so I decided to go with Leaflet instead.

For the website, I used a Bootstrap theme that I modified to suit my needs. I like adding gimmicks to my stories and projects, so I used some code I found on JSFiddle to add an icon of a CCTV camera that follows the cursor in the header of the page. It doesn't do much on mobile (you have to tap to change the icon orientation), but I thought it adds a bit of character to the page.

Another gimmick was a totally unnecessary scrollytelling element on the page. I used the ggmap static plots I initially wanted to use for my story as background for some text to pass by. It doesn't work at all on mobile, but it doesn't significantly break the site so I left it in. I also added a live feed of a TfL CCTV camera from Islington in the story to make it seem more "alive" and to hone in the idea that these cameras are actually out there.

There were a lot of challenges along the way. One issue I stumbled upon again and again is trying to work with data in different formats (XML, JSON, CSV, shapefiles) and making it all work. I had to do a lot of conversions from and to dataframes, matrixes, lists, XML lists, etc. I had to figure out I need to tell R to treat some data as numeric when it didn't. A lot of work went into trying to filter out values outside my shapefiles, extracting values from XML files and even simple things such as merging two dataframes. Thankfully, Stack Overflow and R-bloggers were there to help.

If I had more time for this project, I would take the project to a city-wide level (or another city). Right now, data is inconsistent (41 council cameras in Camden and 156 in Islington) and it doesn't really tell a story. That's why, as time passed on, my story shifted from trying to figure out if the locations of surveillance cameras correspond to where crime happens to showing that CCTV is very ineffective at what authorities claim it does.

Nonetheless, I believe the story is interesting for local publishers even if just shows a map of cameras in their borough and where crimes happen.