Blokes in Blue: Bobby Activity in the UK

ANLY 502 Spring 2016 Final Project Proposal

TEAM

Tim Ahn <ta437@georgetown.edu>
John Hotchkiss <jsh104@georgetown.edu>
Josh Kaplan <jk1805@georgetown.edu>

ABSTRACT

Around the world police are receiving greater scrutiny in all aspects of duty. In the interests of increased transparency of police operations, the United Kingdom has compiled and released detailed data covering street-level crime, outcomes, and police response. Using past applications of this data for inspiration, we intend to geolocate police activity in the hopes of teasing out interesting relationships between police departments and their "stop and search" and crime response activities. In addition, we will evaluate how these practices change depending on Socioeconomic status (SES) indicators of populations in the vicinity of those police activities.

DATA SETS

- Street-level crime, outcome, and stop and search information, broken down by police force and 2011 lower layer super output area (LSOA) from https://data.police.uk/data/ (government open data). 19 million rows of structured data from December 2010 to January 2016 for England, Wales, and Northern Ireland. Data is downloadable in CSV format, with three tables per month for each of the 45 police departments. 1.9GB.
- Middle Layer Super Output Area (MSOA) economic and social statistics, from http://www.nomisweb.co.uk. Government open data from the 2011 census. Data is structured, and downloadable in various formats, including CSV. 7201 rows per MSOA-level table, at the moment we are planning on using religion and unemployment measures, but we will certainly add more data. ~1MB per table.
- Lookup file to map LSOA-level crime data to MSOAs, from https://data.gov.uk/dataset/output-areas-2011-to-local-authority-districts-2015-to-lower-layer-super-output-areas-2011-to-m. Structured government open data, 171,373 observations, downloadable as CSV. 22MB.
- 2011 MSOA GIS shapefiles (.shp) from https://census.edina.ac.uk/easy_download_data.html?data=England_msoa_2011. 150mb in 10 files. Structured government open data. If we can figure it out, these will be used to map the crime and census data to MSOAs.

WORKING HYPOTHESIS

Socioeconomic status (SES) is correlated with crime rates: Areas containing populations of, on average, lower SES have higher crime rates than areas containing populations of higher average SES.

ANALYSIS TOOLS

- We need to figure out what tools can make the kinds of visualizations we want, something like this: http://londondatastore-upload.s3.amazonaws.com/instant-atlas/msoa-atlas/atlas.html
    - Try the 30-day free trial of http://www.instantatlas.com/
    - Google fusion tables tables.googlelabs.com/
    - Tableau
    - Bokeh
- Spark, on AWS

- Hive or Pig to combine datasets
- Skflow

VISUALIZATIONS

This will actually be one of the core pieces of our project. As alluded to above, we hope to plot the different police activities on various maps along with SES indicators in order to organize the massive quantities of data that we are working with in a meaningful way. We imagine that gaining an understanding of shapefiles and developing these maps and visualizations will be a sizable effort.

TEAM DUTIES

We believe that given our relative inexperience with the various tasks associated with our final project, it is not possible to directly divide tasks among the group members in an equitable manner at this time. Having worked together on group projects in the past, we find that treating all tasks as a group effort, succeeding and failing as a team, will lead to the best outcome. Tasks include:

-Data Joining/Merging/Aggregating
-Perform Analysis
-Create map visualizations
-Prepare Presentation
-Write Report

TIMELINE AND MILESTONES

April 4-11: Once the proposal has been accepted, make sure that we have successfully mapped the the SES and police data to the same level. Figured out how to create at least one map, or portion of a map, successfully. More or less, this is a proof of concept week.

April 11-18: Based on difficulties encountered proofing the concept, we will decide on what, more exactly, we hope to get into the final deliverables. This week should be spent performing the analysis and finalizing data combinations for visualizations.

April 18-25: Finalize analyses and finish at least enough of the visualizations for the presentation.

April 25-May 2: Finish presentation and presentation prep.

May 2-13: Write and submit report.

FINAL DELIVERABLES

- The report will detail our data sources, methodology, and findings.
- For the presentation, we will briefly summarize the report, and hopefully present some interactive maps (At the very least, static ones).
- Our deliverables will be visualizations and the final report.