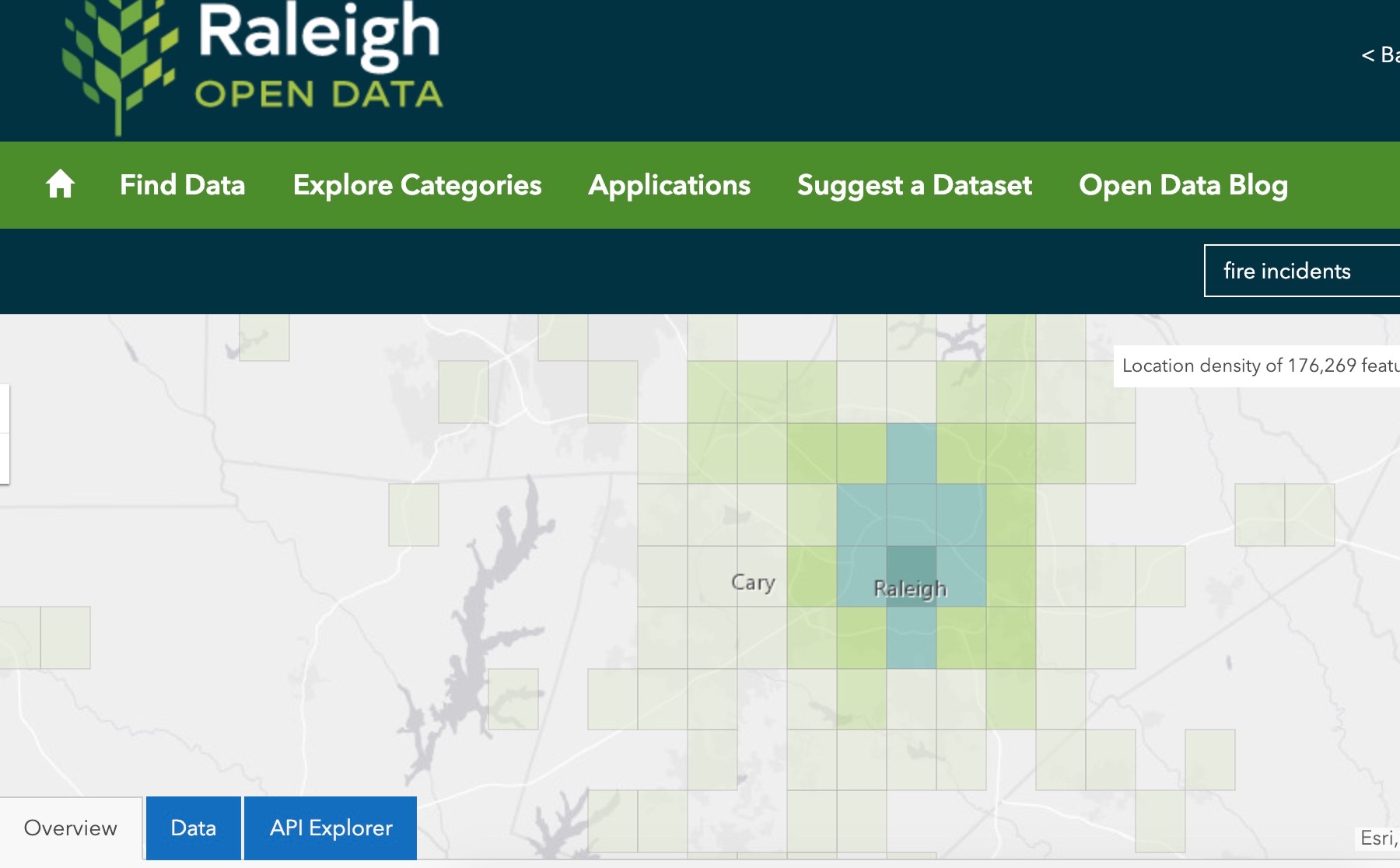
Raleigh Fire Incident Analysis

Data Analytics Bootcamp Project 1 - Team 7

January 16, 2019



# Introduction

Raleigh launched its open data project in 2012 when the City Council adopted the Open Source Resolution. Since then it has made available numerous datasets using the ArcGIS platform reflecting the functioning of a growing metropolitan center. We chose to focus on the fire incident data, which uses a nationally adopted system for coding emergency service calls into categories, and tracks dispatch and arrival times at incidents. The data allows us to view metrics in overall number of incidents over time, response speed by stations, and compare the number of incidents with population growth, weather, and more.

From the Raleigh Open Data website, we know there are 177,738 rows available. The API limits calls to 2000 records per request. We created a numpy array to make a list of start record number in increments of 2000, up to 176,000. We used that list to set the parameter for the starting record number in a loop, running 89 API calls to get all the incident and coordinate data.

Our try-except loop caught one call that repeatedly failed, but we were able to pull that set down from the web interface manually. We then parsed the two JSON files separately (one was a list, one was a single JSON set), created Pandas dataframes from the parsed data and appended the single set to the larger one.

There were 175738 records in the main set, but 1425 lacked coordinates, 1626 records lacked incident\_type, 1436 were missing dispatch date/time and two did not have cleared date. The ‘missing’ set had 2000 records, 24 missing coordinates, 49 missing type.

For data analysis, we had to reformat the POSIX date time columns to datetime objects using the Pandas to\_datetime function. This was tricky!

Fire incident analysis:

<Add per-type analyses>

* Added a column (timedelta) ‘Seconds’ that was arrive time minus dispatch time to track overall and individual station performance.
* Removed rows with null stations.
* Grouped this dataset by Year, counted the IDs to get tally of incidents per year.
* Calculated the mean, median standard deviation of response times.
* Raleigh FD had a mean incident rate of 11793.73 per year, median 11529.0, with a standard deviation of 1009.29 and error 304.31
* The most incidents came in 2017 when there were 14423, the minimum in 2009 when there were 10915.
* <graphics/total\_fi\_year.png>
* Fire incidents have been on the rise in Raleigh over the period of study

Station response time analysis:

* A list of stations was obtained from the Wake County website.
* There was one record with station number 13, which does not exist.
* There were 30034 rows without arrival times.
* These nulls fell into the following categories:
  + Dispatched & cancelled en route
  + False alarm or false call Other
  + Cover assignment standby moveup
  + Alarm system sounded due to malfunction
  + Service Call other
  + Dispatched & cancelled en route
* Of the remaining, there were 59 incidents with response times more than 3 standard deviations from the mean.
* Before removing them, the RFD stations had a 307 second response time, with standard deviation 1046, error 3.0.
* After removing outliers, the RFD stations had a 292.35 second response time, with standard deviation 145.54.
* The station with the fastest response times averaged 232.76 seconds while the slowest averaged 348.36.
* There were 18 stations with response times above average: [2, 4, 6, 8, 11, 15, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]
* <‘./graphics/station\_response\_time\_summary.png'>