**Lab 1 – Cleaning 911 Calls – Part 1**

The overall goal of the first project is to use visualizations to illustrate interesting facts about the frequency, timing, and location of 911 calls. The attached data contains information on all 911 calls made in Charleston, North Carolina during 2010 and 2011.

**Overview of Part 1**

Our first tasks involve cleaning up the raw data and creating new columns that we will use in our visualization. You will be working in groups of 2-3 students on the following tasks.

1. Open up that file name **Charleston\_Calls\_for\_Service\_and\_Officer\_Initiated\_Calls\_2010\_2011.csv** in OpenRefine.
2. Clean up all of the textual columns using clustering and/or text facets. In particular, you will want to get the *nature* column as clean as possible.
3. Make a new column based on the *nature* column, but with fewer and more broad categories. For example, you might combine all accidents (bicycle, automobile, etc.) into one category.
4. Create a **length of call** variable using *calltime* and *timeclose*. You will want to convert both of these columns to a date column, then use an expression that uses the diff function. Documentation on diff and other OpenRefine date functions can be found [here](https://github.com/OpenRefine/OpenRefine/wiki/GREL-Date-Functions):
5. Extract various dateParts (another GREL function describe in the documentation linked above).
   1. Month
   2. Day of week
   3. Time of day
   4. Etc.
6. Compute **time\_since\_jan\_1st** (elapsed time since 1/1/2010) for various granularities. You can use the toDate function on a string to specify 1/1/2010.
   1. Days
   2. Hours
   3. Weeks
   4. Months

**Other Details**

1. Keep a detailed log of the steps/transformations along with your reasoning behind these changes. You will submit this in a separated attached word document.
2. Part of your grade will be based on an evaluation by your peers. **Make sure you do your part**!