UPSTAT-2015 DATA COMPETITION

# URBAN ANALYTICS CHALLENGE

The data was obtained from a “loop detector” underneath the pavement of the southbound lane 200 feet north of the intersection of East Main Street and Culver Road in Rochester, NY. It was provided by the Monroe County Department of Transportation under the Freedom of Information Law. This data set represents eight months of observations from Oct 13, 2013 to June 7, 2014.

Loop detectors are used to identify areas and episodes of traffic congestion. Here, we offer them as a challenge in urban analytics.

The variables are

* **Volume** is the number of vehicles per hour going by the location.  It is computed as twelve times the actual volume measured over five minutes (this converts it to an hourly flow rate).
* **Speed** (expressed in miles per hour) is calculated from the difference in time between the beginning and ending of vehicle detection.  It is the average over all estimated speeds of detected vehicles within a five-minute window.
* **Delay** (given in seconds) estimates how long vehicles wait between arrival at and departure from the intersection.
* **Stops** (given as numbers of vehicles per hour) count how many vehicles are approaching a red light. It approximates the length of the queue when the light is red.
* **DateTime** includes day, month, year and time.

The goal in this challenge is to discover the most compelling, appealing and practical patterns in the episodes of traffic congestion. The merits will be measured in terms of

* novelty of the patterns discovered,
* usability of the recommendations made, and
* originality of the scientific methods used.

Notes:

* This data set contains a few extreme outliers on Delays and Stops. You are strongly encouraged to provide your best guess of the possible origin of such spikes and/or extreme outliers. Central imputation or other imputation methods may be used to replace these outliers.
* The recommendations made by the winning teams will be given to the Monroe County Department of Transportation. In the interest of usability, the recommendation should use as little statistical jargon as possible.
* The format of the final report and submission rules can be found at up-stat.org.
* The data set is contained in the file *UrbanAnalytics2015.csv* and has a total of 68,437 rows with 5 columns.