# Introduction

NYC forestry department is responsible for maintaining tree points in City Of New York. Tree maintenance not only includes planting, periodic inspection and maintenance but also preventing and responding hazards caused dues to trees. Variety of work related to maintenance make it complex to plan like some service request like fallen tree or root impacting sewer line require urgent attention while requests like pruning at traffic signal or utility line are not urgent but required to prevent accidents.

Using data from NYC forestry department, we will explore type of service request received related to maintenance of Tree points, priority, action taken to predict response time in relationship with location, weather, time of the year as well as tree geometry.

## Dataset

NYC Open Data has dataset available on Forestry Service request, Forestry inspection and Forestry Work order from 2015 till Sep 2017.

Primary Data set is Forestry Service request

<https://data.cityofnewyork.us/Environment/Forestry-Service-Requests/mu46-p9is>

This Data set contain total 162,386 observations with each request has attributes on location, creation date, close date, service request category, complaint type, resolution , resolution status, Borough, community board, StateSenate, Congress , council districts as well as longitude and latitude.

There are other data set related to forestry as well as demographic information used to analyze impact demographic date on service request resolution.

Forestry WO: Total 192K with 59 Columns

<https://data.cityofnewyork.us/Environment/Forestry-Work-Orders/bdjm-n7q4>

Forestry Inspection: Total 239 K with 28 Columns

<https://data.cityofnewyork.us/Environment/Forestry-Inspections/4pt5-3vv4>

Community District geographical area: Area per Community District

<https://data.cityofnewyork.us/City-Government/Community-Districts/yfnk-k7r4>

Community District Breakdowns: Demographic Breakdown of community district

<https://data.cityofnewyork.us/City-Government/Community-District-Breakdowns/w3c6-35wg/data>

Council District geographical area: Area per council district

<https://data.cityofnewyork.us/City-Government/City-Council-Districts/yusd-j4xi>

## Data Wrangling

All above datasets are available as CSV and are directly imported in R.

In the imported Forestry Service request Dataset all dates Created Date, Updated Date and Closed Date are in Char class. Closed Date is not populated properly while updated date has more values available showing some action take but not closed. For our analysis, we will use updated date to calculate Response time from created date.

1. Both Created date and Updated date are in Char Class. First step is to convert both Created date and Updated date to date class using as.date function.
2. Separate Month and Year from created date and updated date in separate columns to get Month and Year values for each observation.
3. Calculate response time using mutate function as Updated Date-Created Date
4. In some observations, Response time is negative which is manual error in updating date. Use filter function to filter these observations.
5. Create New dataset by Joining Service Request, Inspection and Work Order Dataset.
   1. Join Inspection Dataset and Service request using Global id in Forestry service request and ServiceRequest Global id.
   2. Join with Work order dataset using Global Id in inspection with Inspection Global id in work order dataset.
   3. This gives a new data set of 43551 observation where Service request has been raised and it has been inspected and work order is initiated to close service request. Other service request might be open or closed with inspection only.
6. Have looked for outliers where response time is more than 400 days. Has not removed any of these as more analysis is required.
7. We are not removing any NA in any of the columns as during the further analysis it will be removed.