# 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.

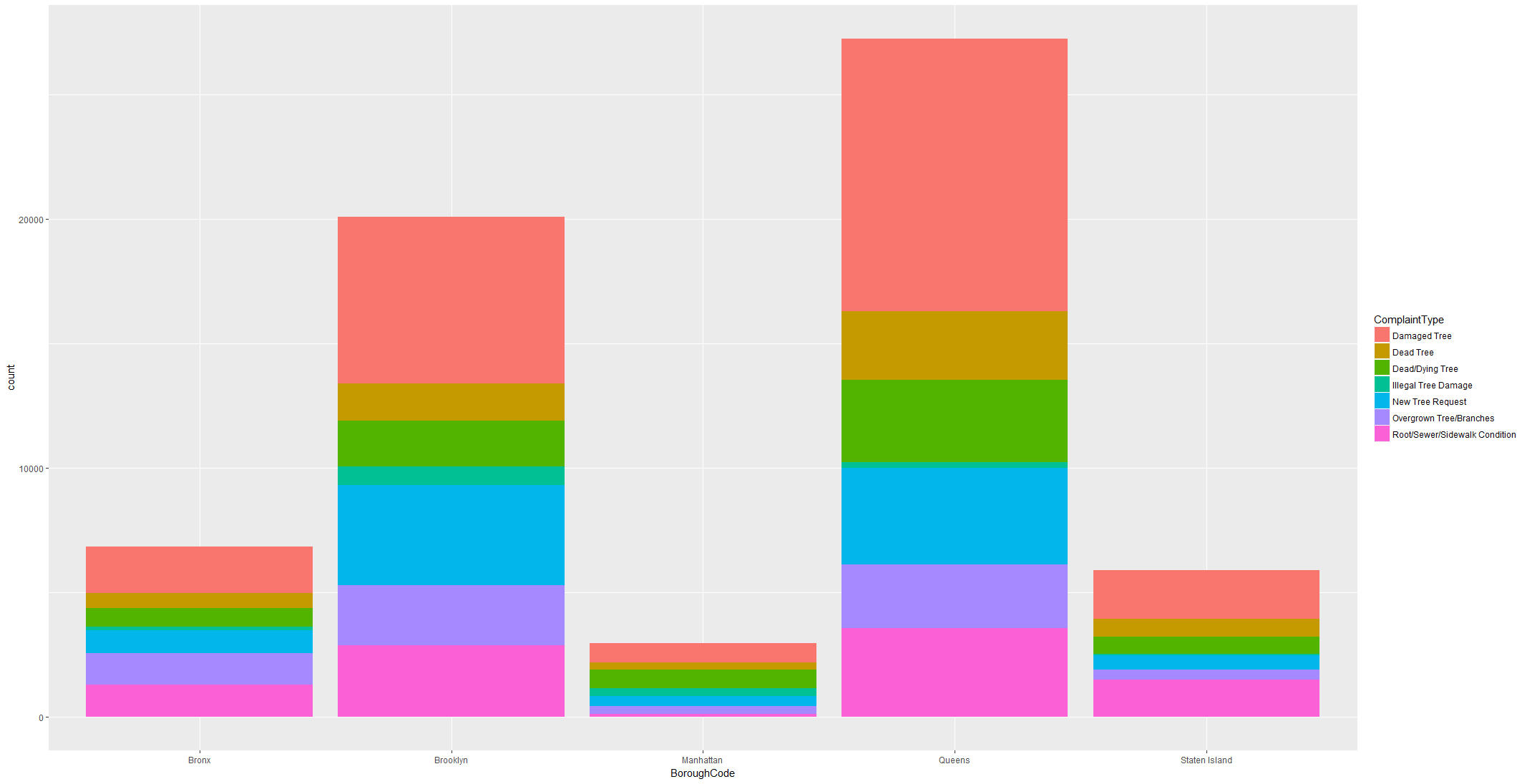
In the imported Forestry Service Request Dataset all dates Created Date, Updated Date and Closed Date are in Char class. Closed Date is not populated properly while updated date has more values available showing some action take but not closed. For our analysis, we will use updated date to calculate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

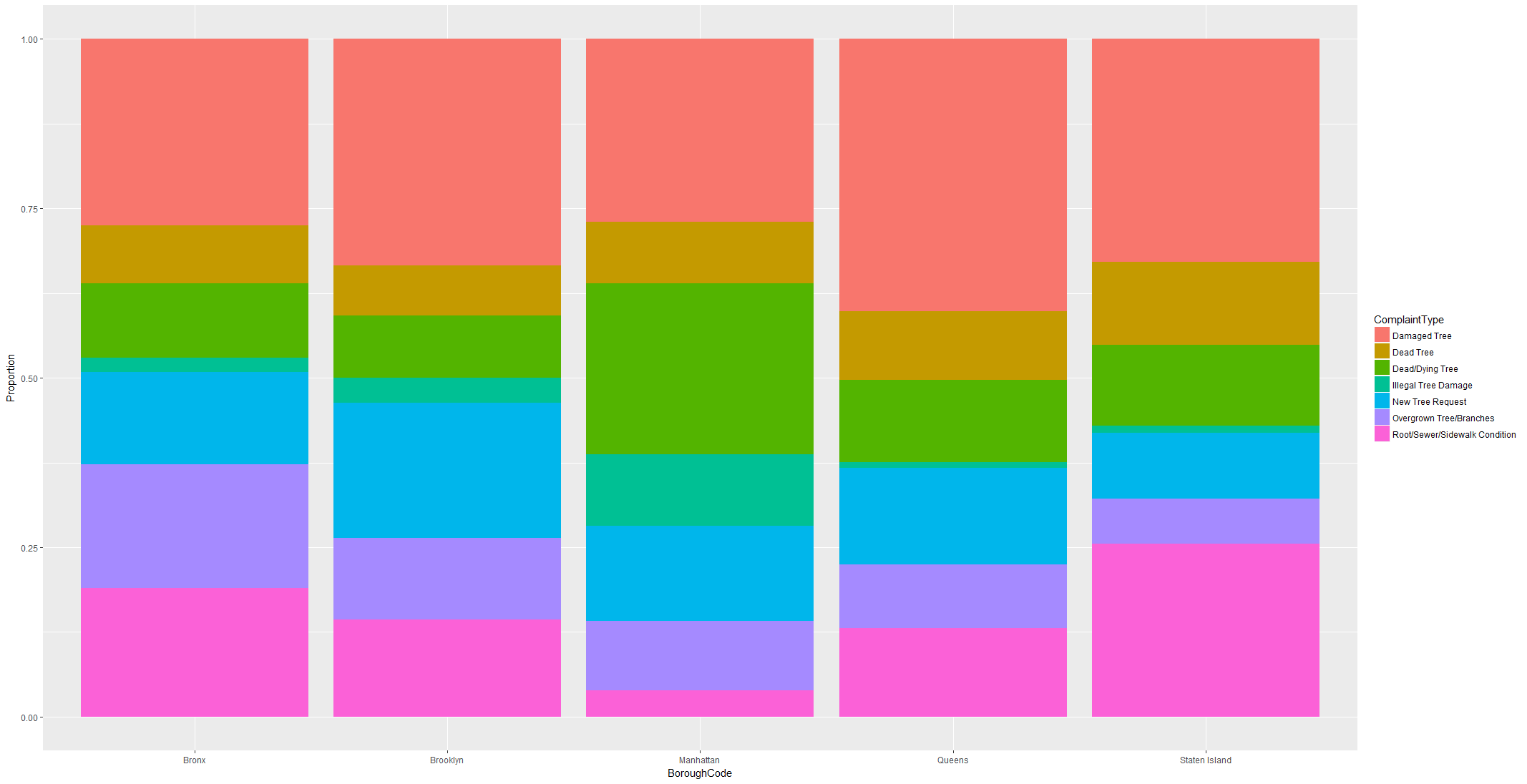
1. Both Created date and Updated date are in Char Class. First step is to convert both Created date and Updated date to date class using as.date function.
2. Separate Month and Year from created date and updated date in separate columns to get Month and Year values for each observation.
3. Calculate 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.

# Statistical Analysis of Data-set:

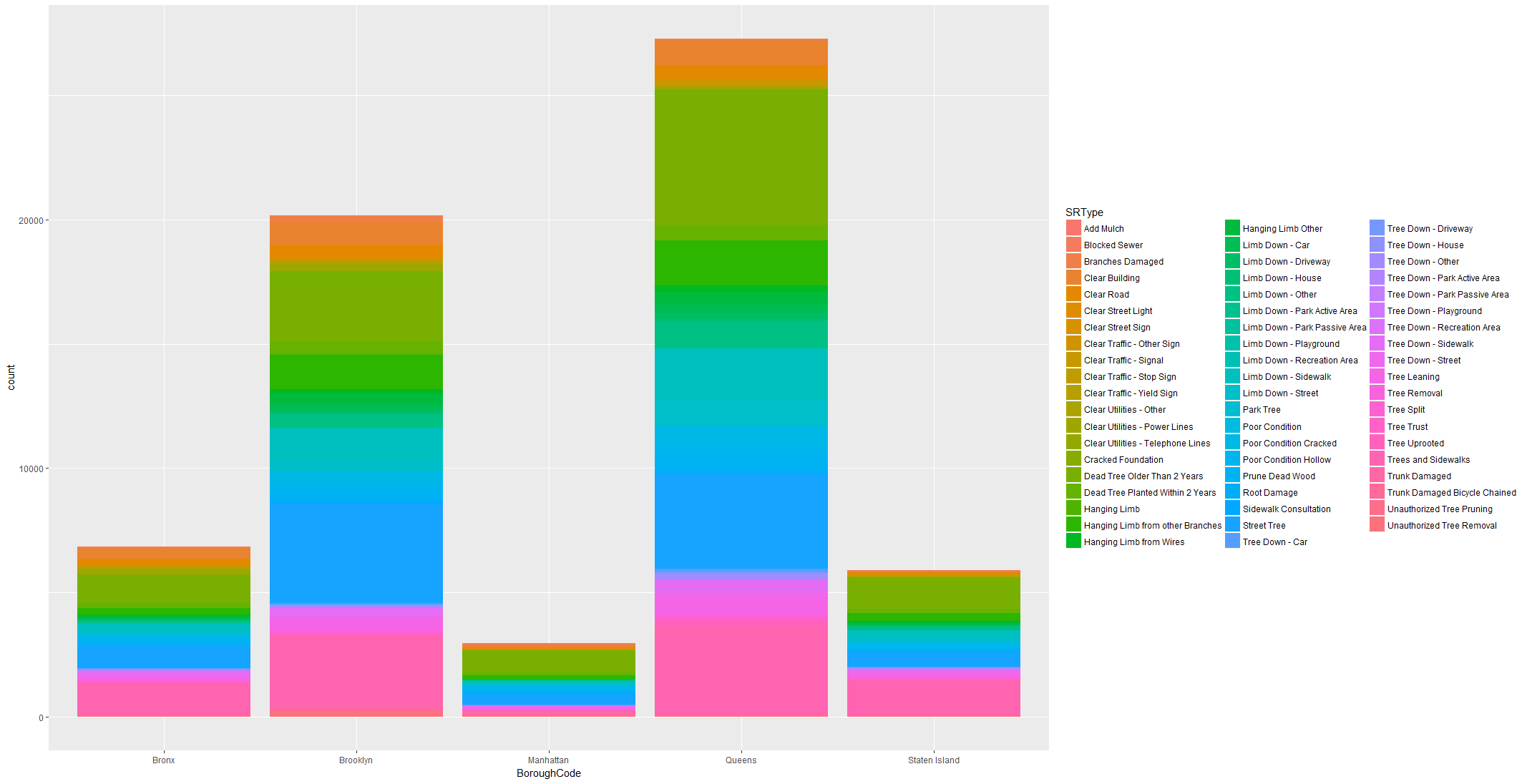
## Distribution of Service request by Borough code

Queens and Brooklyn lead in most number of service request raised in the dataset. Checking proportion of service request, it shows type of request vary by Borough Code. Out of total request from Manhattan around 45% are about New Tree Requests while for other Boroughs 30% of requests are around Damaged Tree and Root/Sewar/SideWalk complaints.



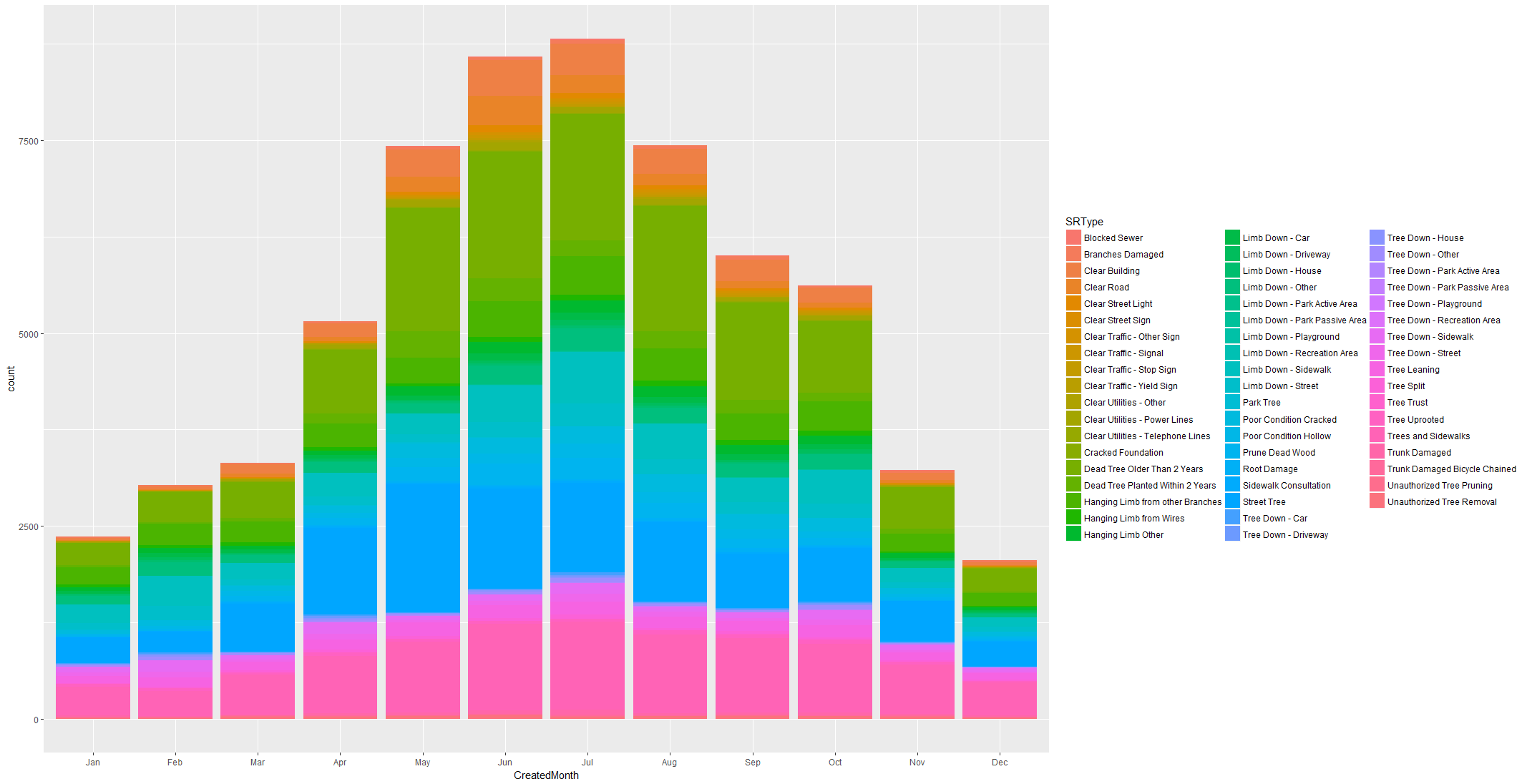


Complaints Type are further broken down into SRType which shows details of each complaint type which gives more detail on problem faced like Tree Down on Car or Tree Down on house etc.



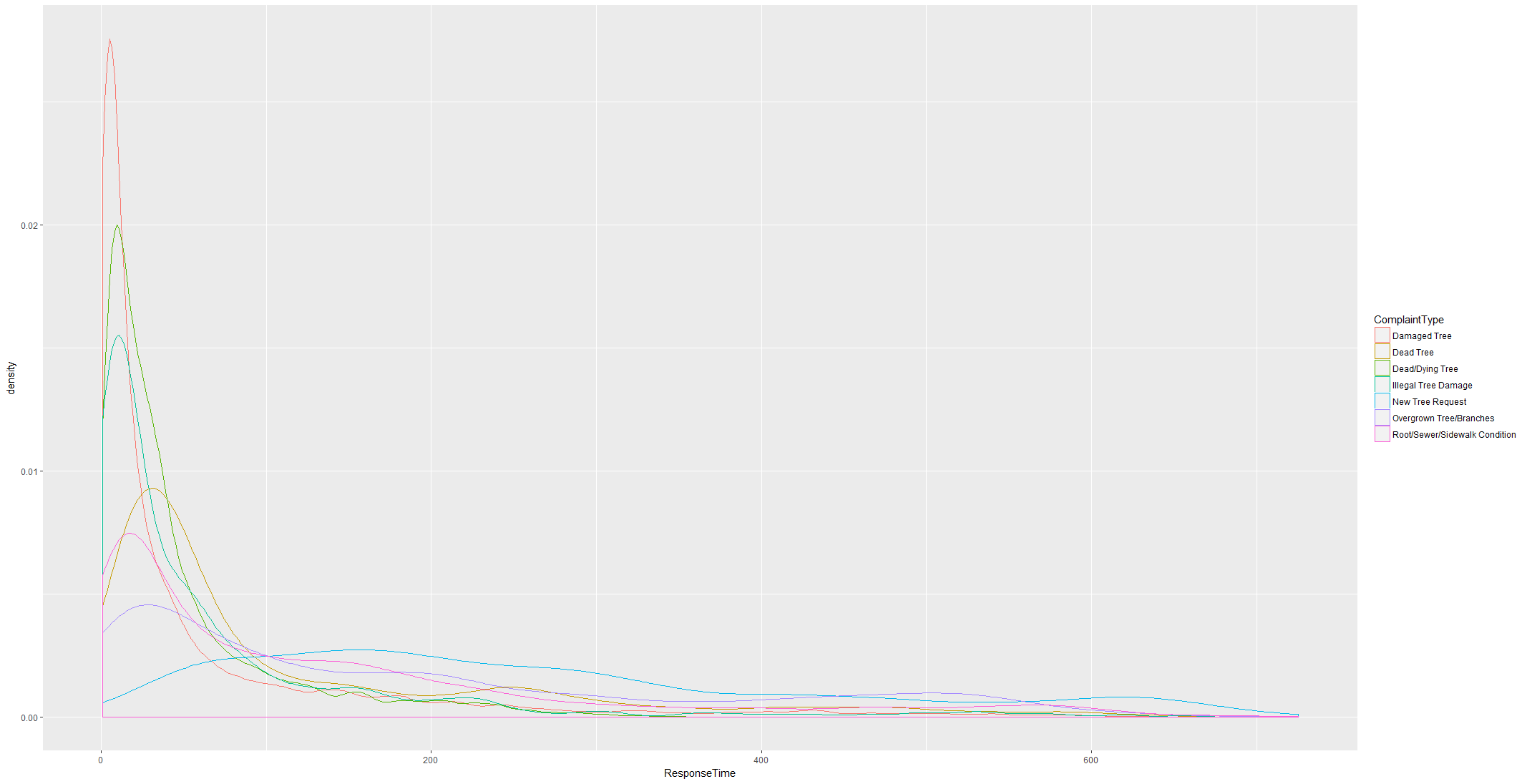
## Distribution of Service Request over Time

Mapping SR creation month, data shows most of the requests are created in Summer from May to Aug and almost normal distribution. It seems post spring during April to Sep most of the complaints increased.



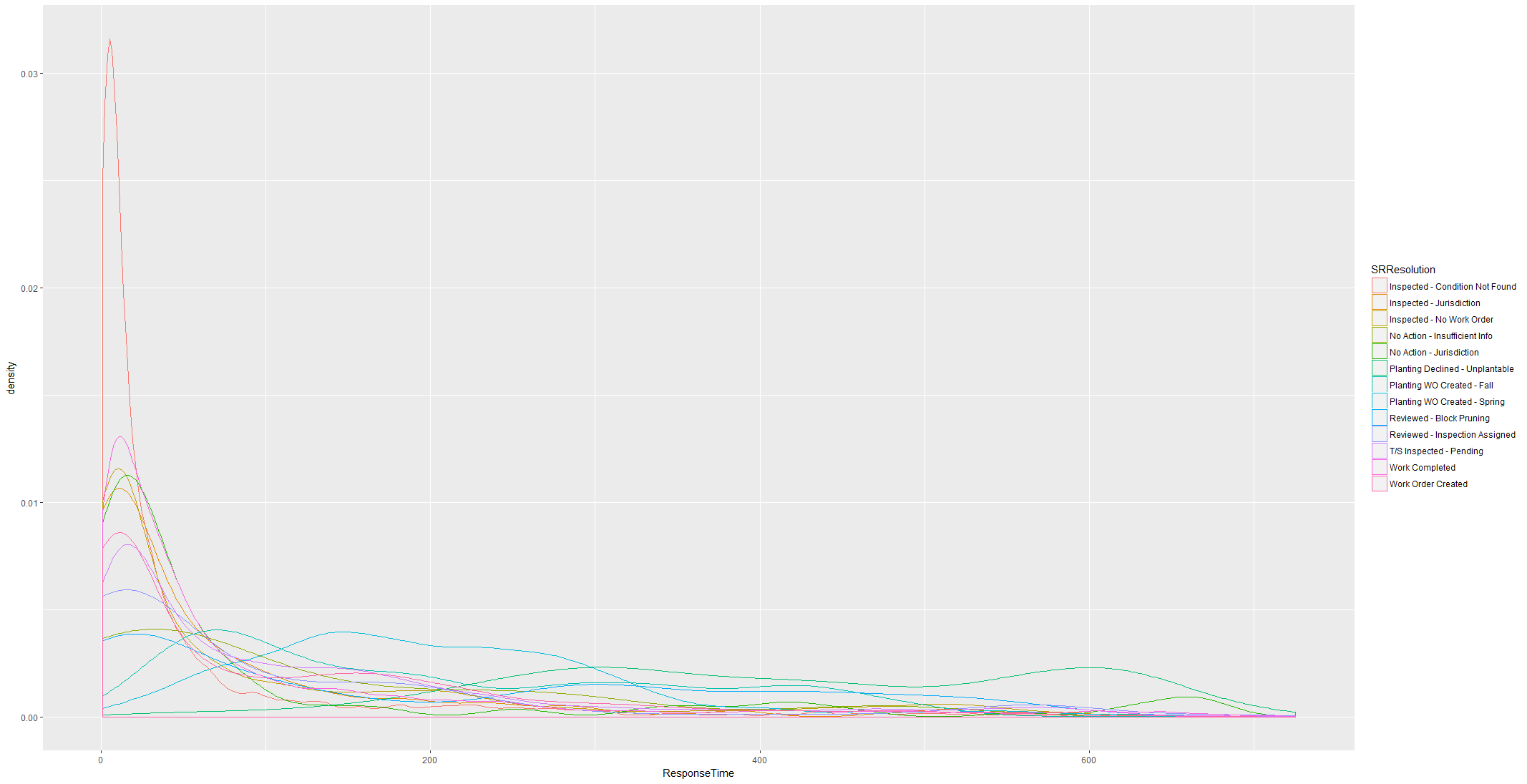
## Response to Complaints

Dataset contains Creation date, closure date and updated date for most of the service request. Also, SR resolution which indicates how each of the request or complaints closed. Response time is calculated using Updated Date-Created Date. Updated Date is the date some action is taken on the request either for closure or any other action. It is more consistently updated than closure date.



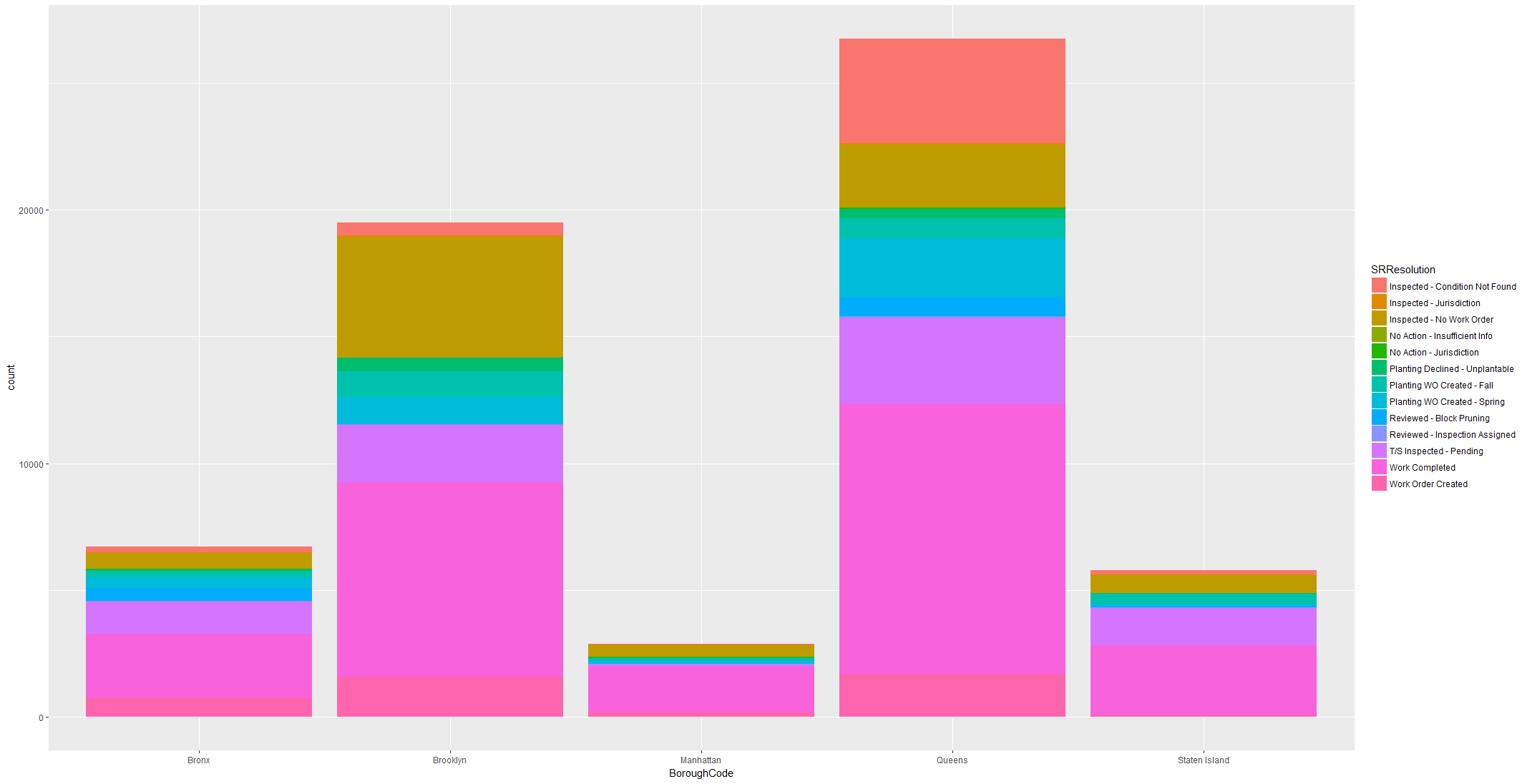
Response time varies with Complaint type. While Damaged Tree and Dead Tree are responded within 100 days, Root/Sidewalk conditions and overgrown branches have response time in more towards 200 days.

Plotting SR Resolution and Response time, data shows that most of the cases for which inspection is completed and condition not found are closed early. Complaints related to new tree request where solution is planting WO created response time is spread over 200 days. For SR resolution WO Created or completed Response time is spread over 100 days because update date is updated once WO is created or work is completed. Cases where information is not sufficient to are having longer update time as it might be waiting for more information before closure.

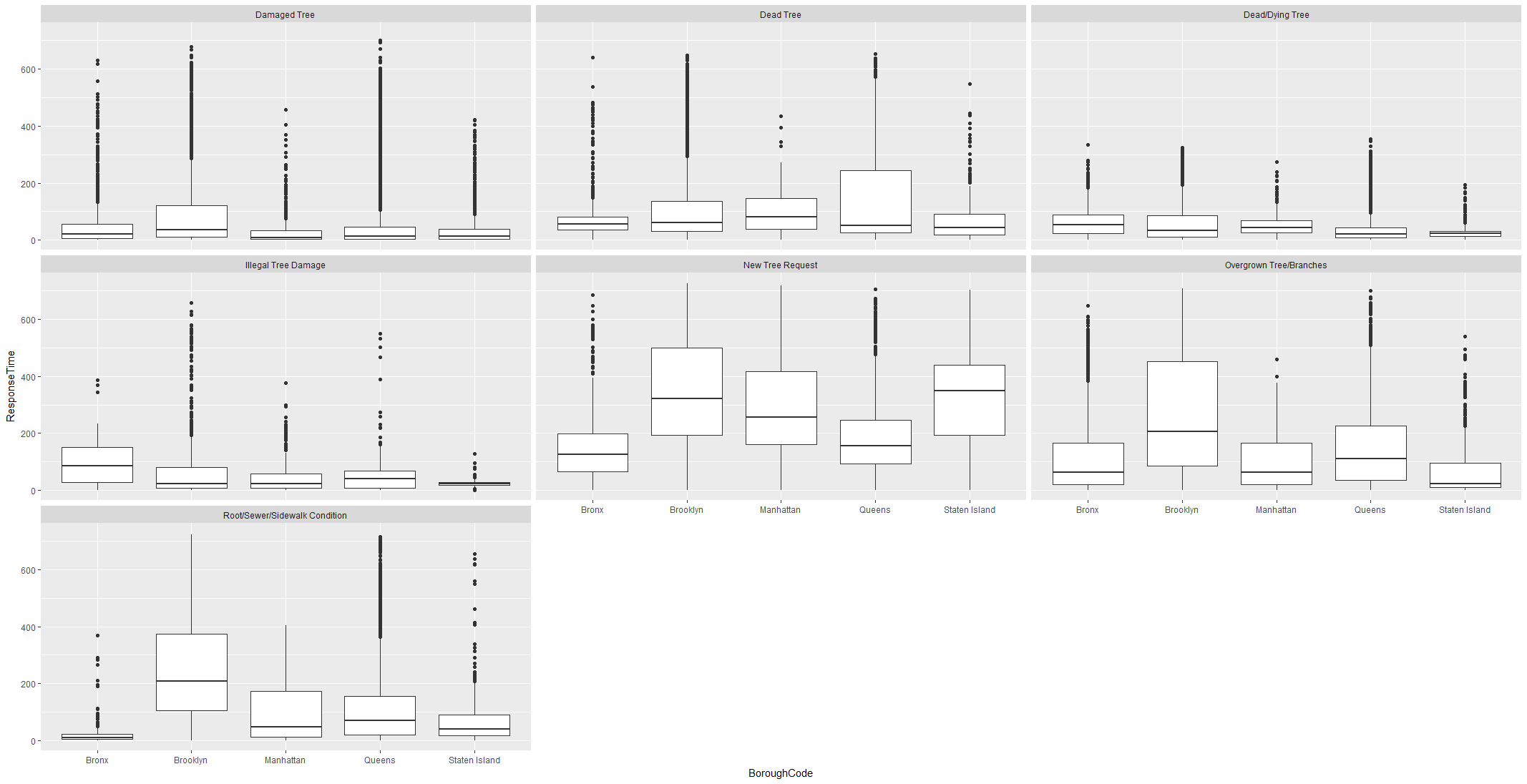


### Distribution of Resolution by location

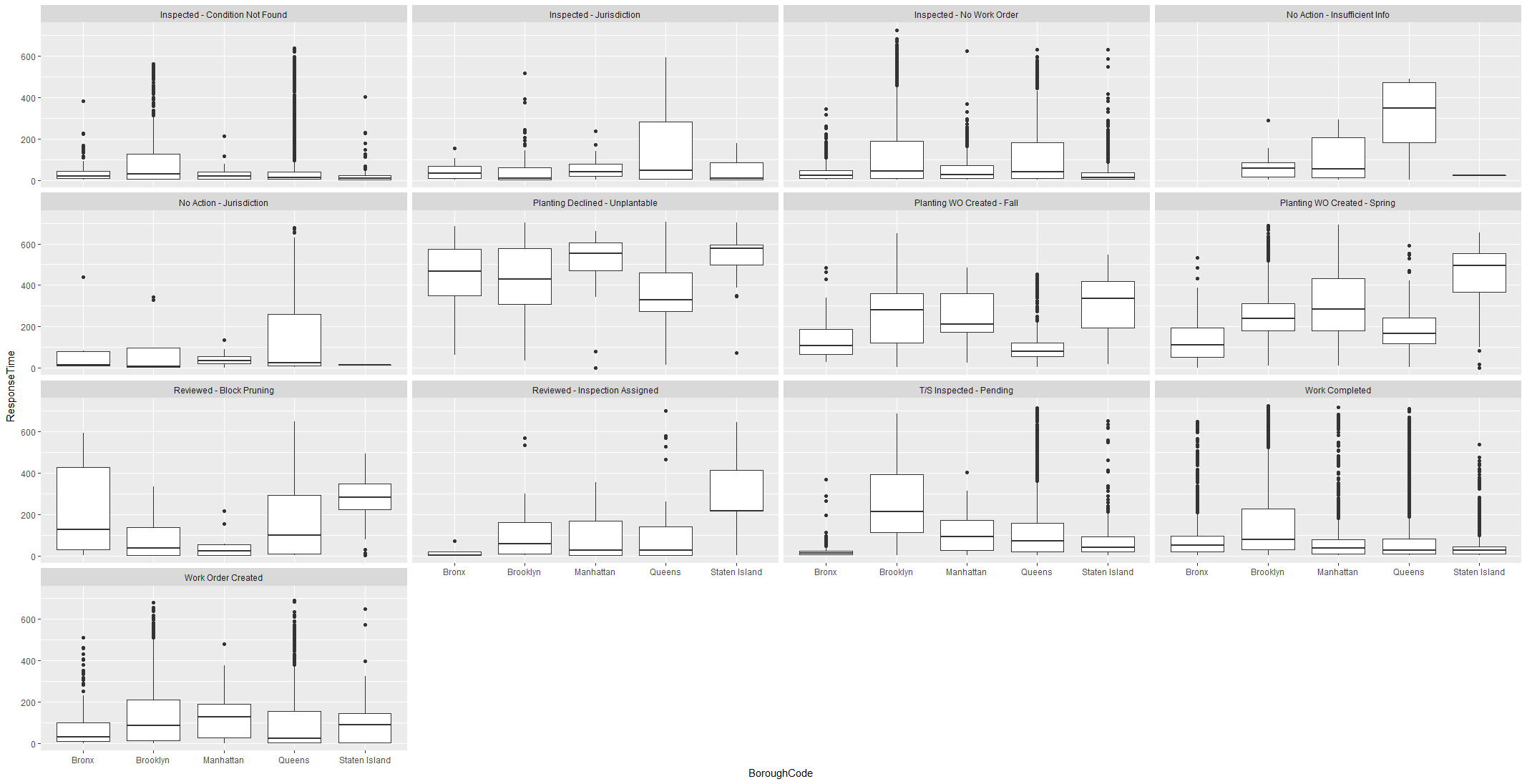
By plotting SR Resolution by Borough Code, it case be seen that Queens has most Inspected Condition Not Found as Resolution Type while Brooklyn has most Inspected-No Work Order resolution.



Plotting Box Plot facet by Borough Code on Response time , Response time for each complaint type can be compared side by side. It can be Seen that for same complaint type Response time Median and IQR varies from BoroughCode. Brooklyn has high IQR of Response Time for complaint Type Root/Sewar , New Tree Request, Overgrown Trees.



Lets plot same box plot on SR Resolution Type to see how response time varies per BoroughCode for same SR Resolution.



For Queens , No action – Insufficient Info has median around 350 days while for others it is less than 100 days. It means either inspection is delayed or it most of the cases pending for information and closed late. Similar to that inspection-Jurisdiction and No action Jurisdiction has Median time is less than 50 days but for some request it is around 300 days.

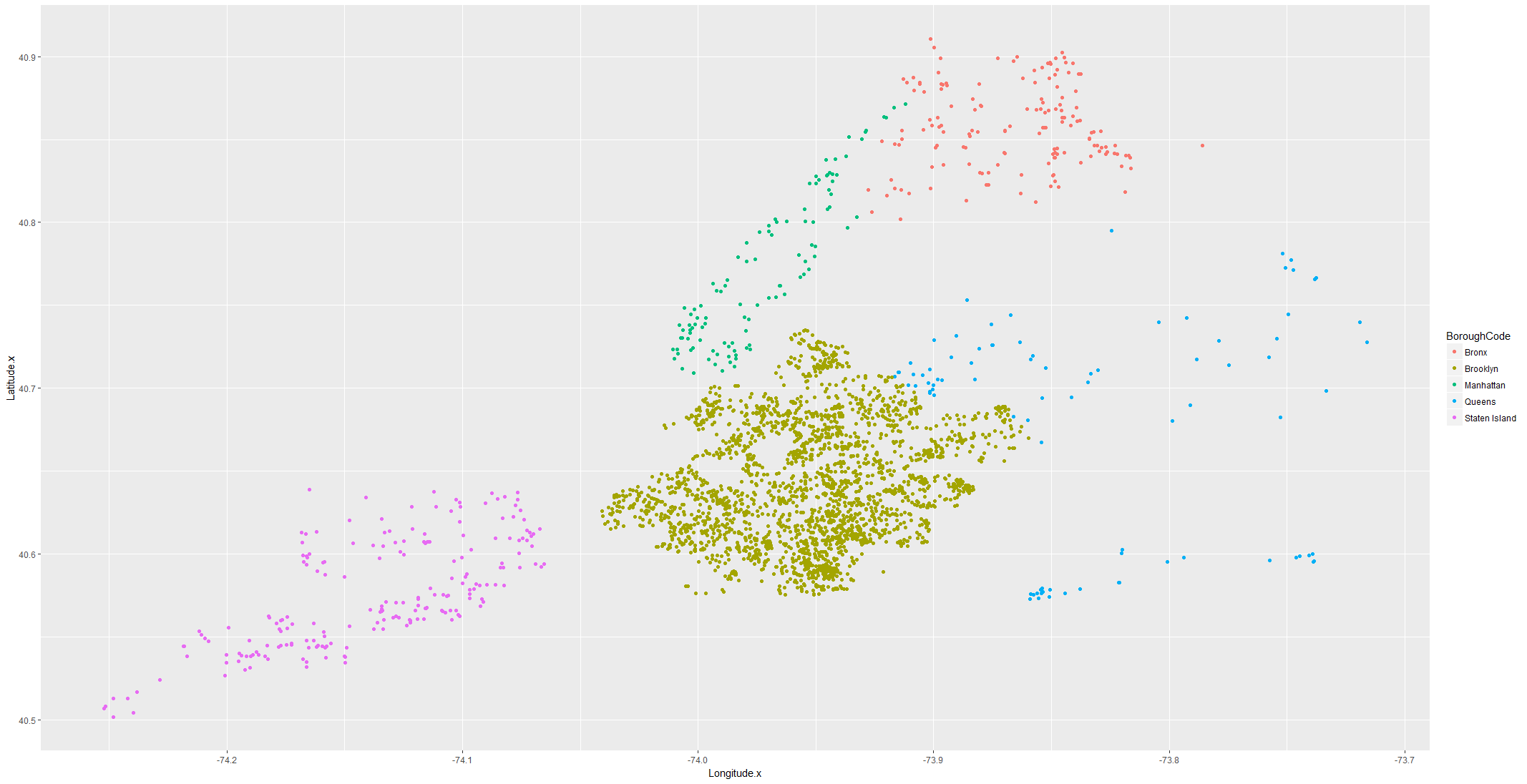
For Block Pruning , IQR for Bronx is around 400 days.

For Brooklyn, Inspected Pending , Inspection Assigned, Work Order Created and Work Order completed IQR are more than other Boroughs. Even though number of request similar to Queens but IQR is more than Queens. It can be inferred that for complaints to be resolved in Brooklyn has chance of taking more time than in other Boroughs.

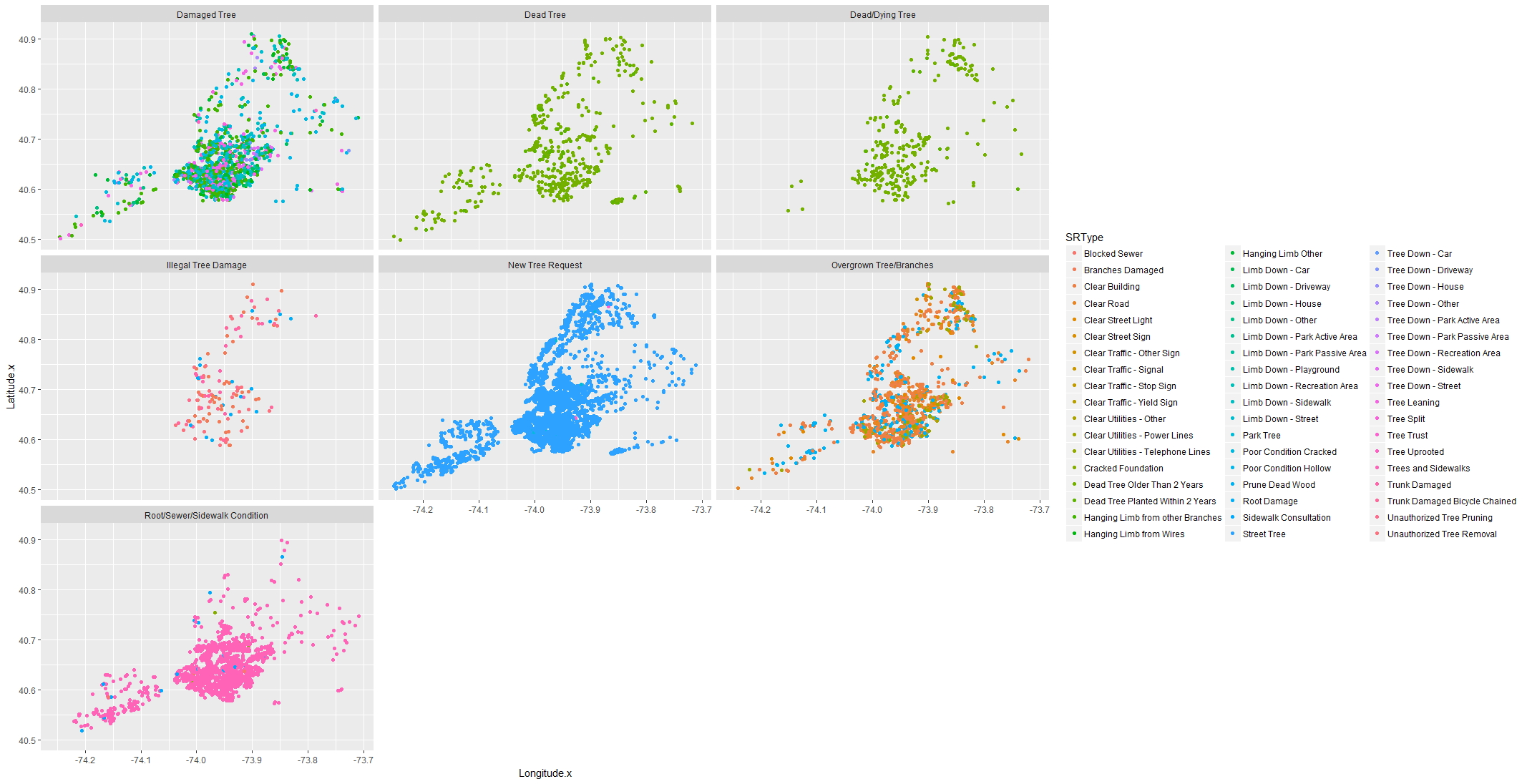
#### Curious Case of Brooklyn

Plotting Borough Code on longitude and latitude with filter on response Time more than 400 days for all complaint types Brooklyn has more presence than any other boroughs even though total number of complaints are less than that of Queens.

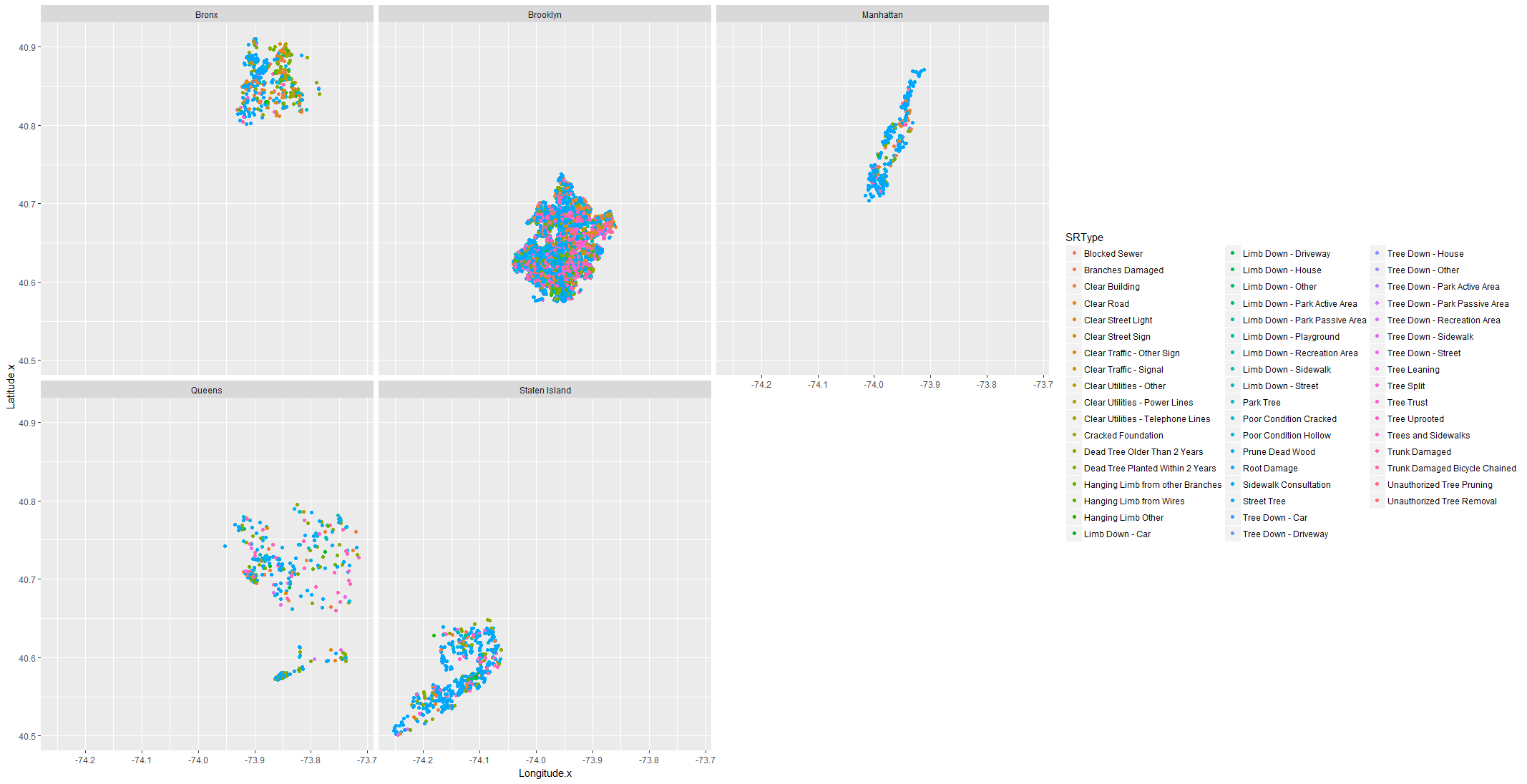
Brooklyn has more Root /Sidewalk complaints than other boroughs.



Plotting SRType over Longitude and Latitude with Response Time more than 150 Days colored by complaint type , we can see that Brooklyn and adjacent Queen communities 405 and 402 has more presence than any other borough.



Plotting SRTYPE on Longitude and Latitude faceted by Borough for Complaints having more than 200 Days.



With above information we can hypothesize that

1. Service request raised from Brooklyn and Adjacent Queens neighborhood chances of update or closure is more than 200 Days and can be 400 days+ as well.
2. Brooklyn has more cases related to Sidewalk repair.
3. In Bronx, Block pruning type of request take more time than other request.
4. Queen has more Damaged Tree Request and most of the cases has not sufficient information to process request which leads to longer closure time.

We need to explore further on

1. Is there any specific Trees under complaints?
2. Are there any repeated request at same location and same Tree Point?
3. How many requests went for inspection and time taken for inspection?
4. Are there any delays in request to inspection and inspection completion to Work order which might be contributing in closure time?
5. What type of work required to resolve complaints

# Data Story

Primary Analysis on service request data provide information on distribution of complaint type by Borough, Complaints creation month, resolution of complaints and response time for each complaint at each borough.

To explore further, we need to understand relate service request data with inspection and work order to get insight on what work is required to close service request.

To get Information on Inspection , this dataset need to joined with Inspection Dataset which provides information on Inspection Type,Inspection status, Inspection initiation date ,Inspection closure time, Inspection Tree Point and Tree Point Diameter.

Then Inspection Dataset is joined with Tree point Dataset which provides additional information on Species of tree and tree Diameter for each request.

Then this dataset is joined with Workorder dataset which provides additional information on Workorder Category, Workorder equipment, Workorder entity, Workorder created date and Workorder closure date.

All information from Inspection, Tree point and Workorder are merged back into service Request dataset to relate service request with Inspection, Tree and Workorder.

In the resultant dataset containing 162k service requests

1. There are 63108 requests where Inspection has been initiated.
2. There are 40300 requests where Wo has been initiated.

We need to further explore that these cases unique cases or repeating cases means

1. Same location repeated complaints have been initiated
2. Same tree points repeated complaints have been initiated

We are trying to find out

1. Is there any trend in Repeated calls?
2. Is there any cause of repeating calls like work is taking longer time and multiple request has been initiated?
3. Is there