# ***Analysis of Response Time to Handle Service Request related to Forestry Dept of NYC.***

# ***Abhay Mourya***

# 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.

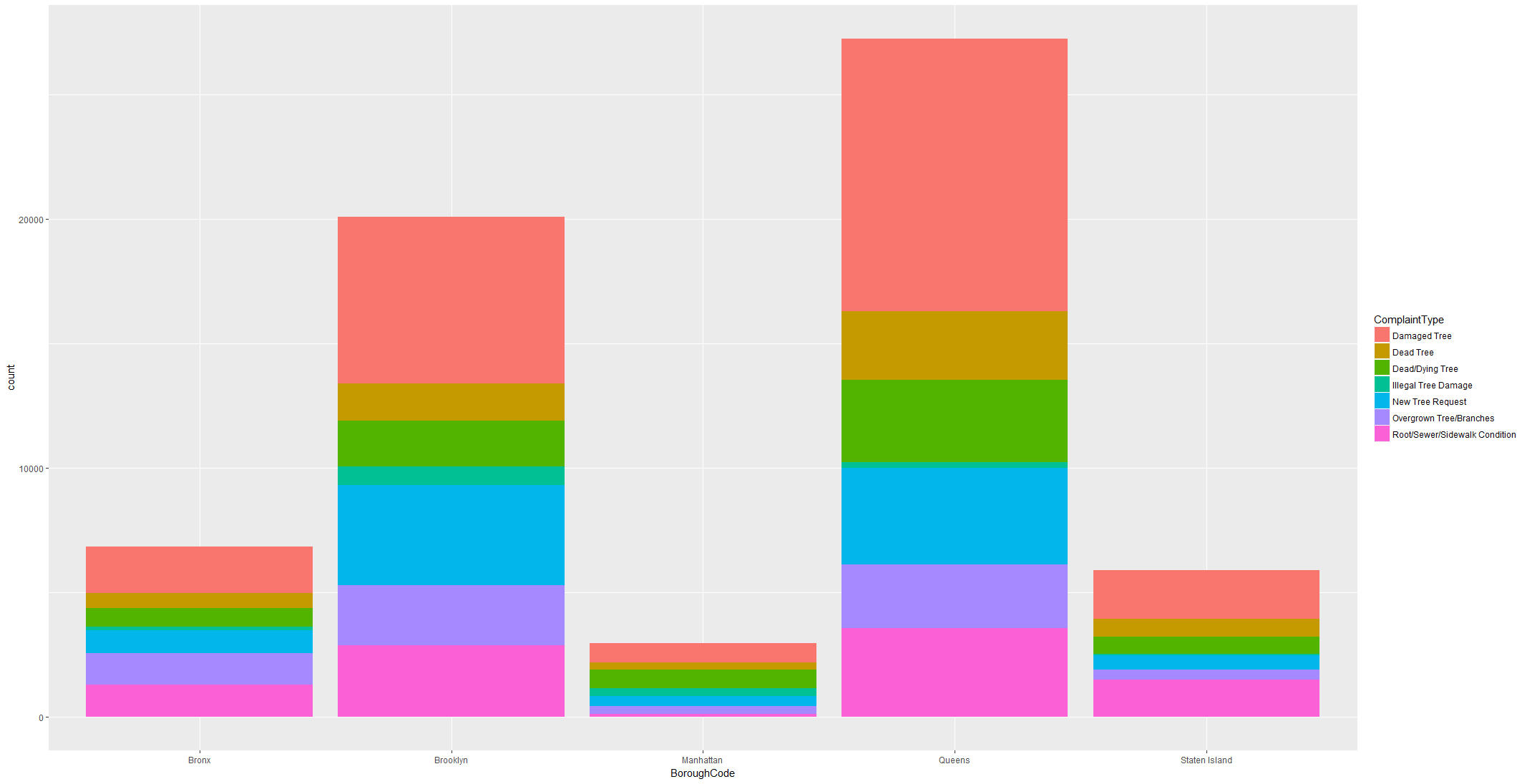
In the imported Forestry Service Request Dataset all dates Created Date, Updated Date and Closed Date are in Char class. Closed Date is not populated properly while updated date has more values available showing some action take but not closed. For our analysis, we will use updated date to calculate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

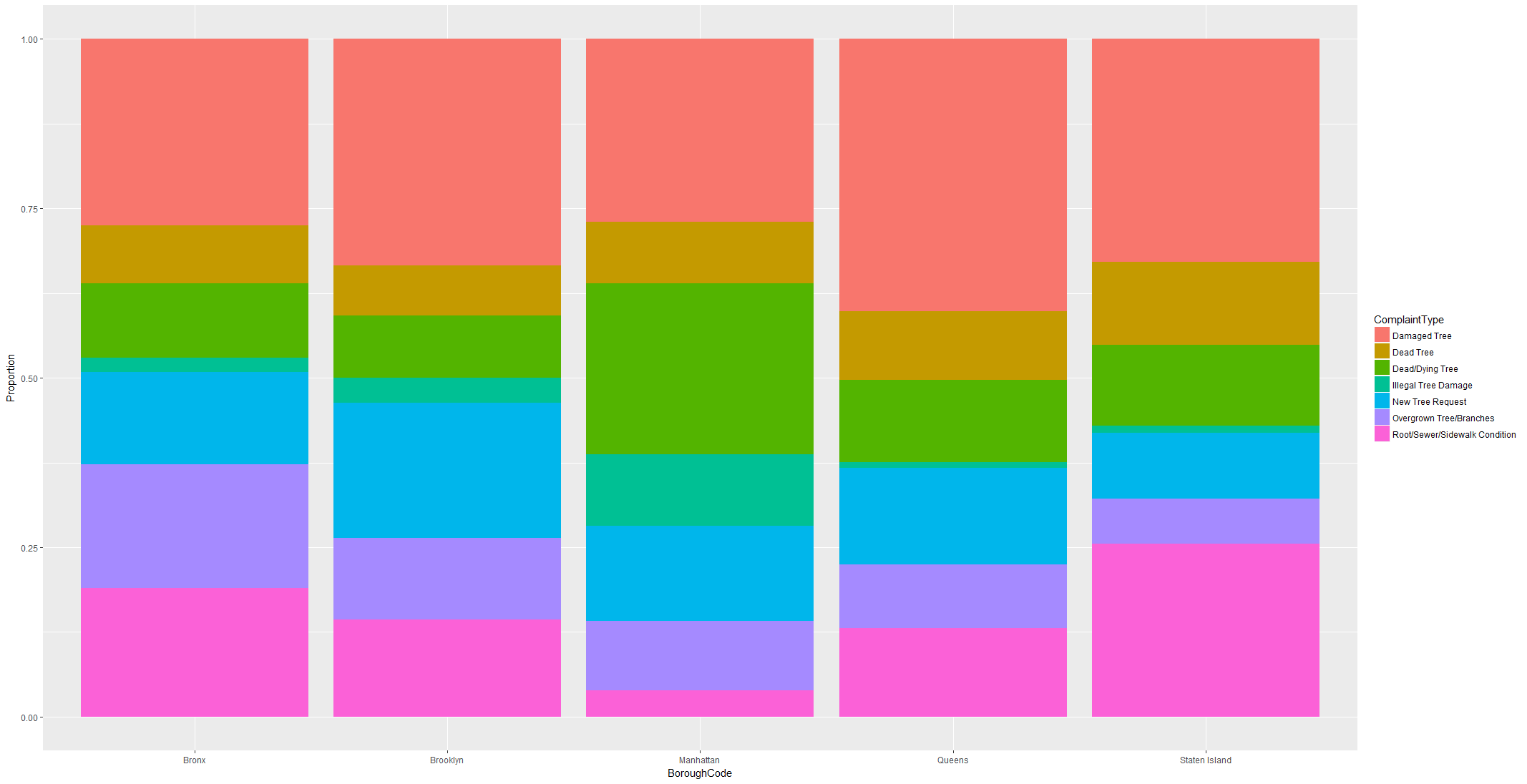
1. Both Created date and Updated date are in Char Class. First step is to convert both Created date and Updated date to date class using as.date function.
2. Separate Month and Year from created date and updated date in separate columns to get Month and Year values for each observation.
3. Calculate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

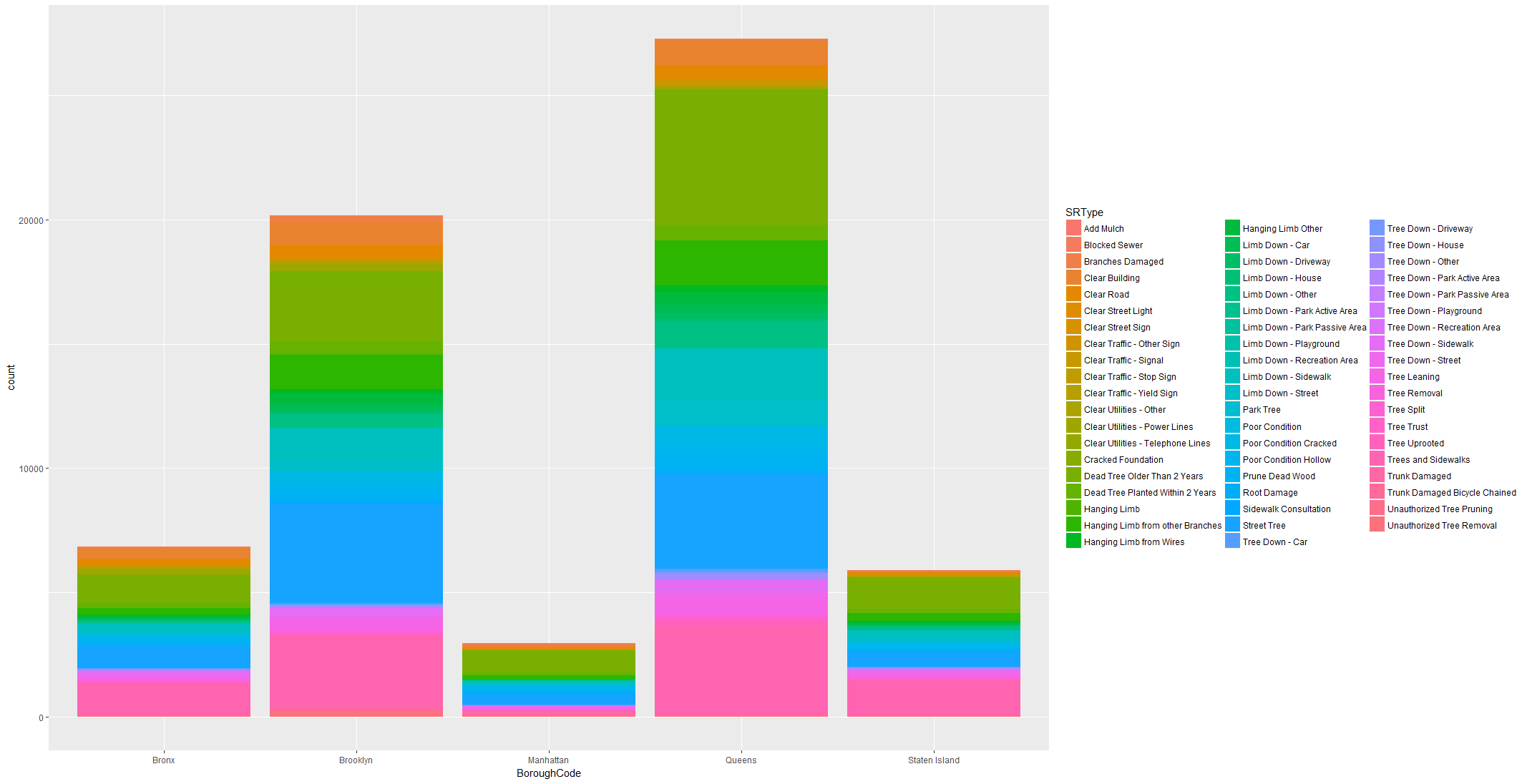
## Distribution of Service request by Borough code

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.



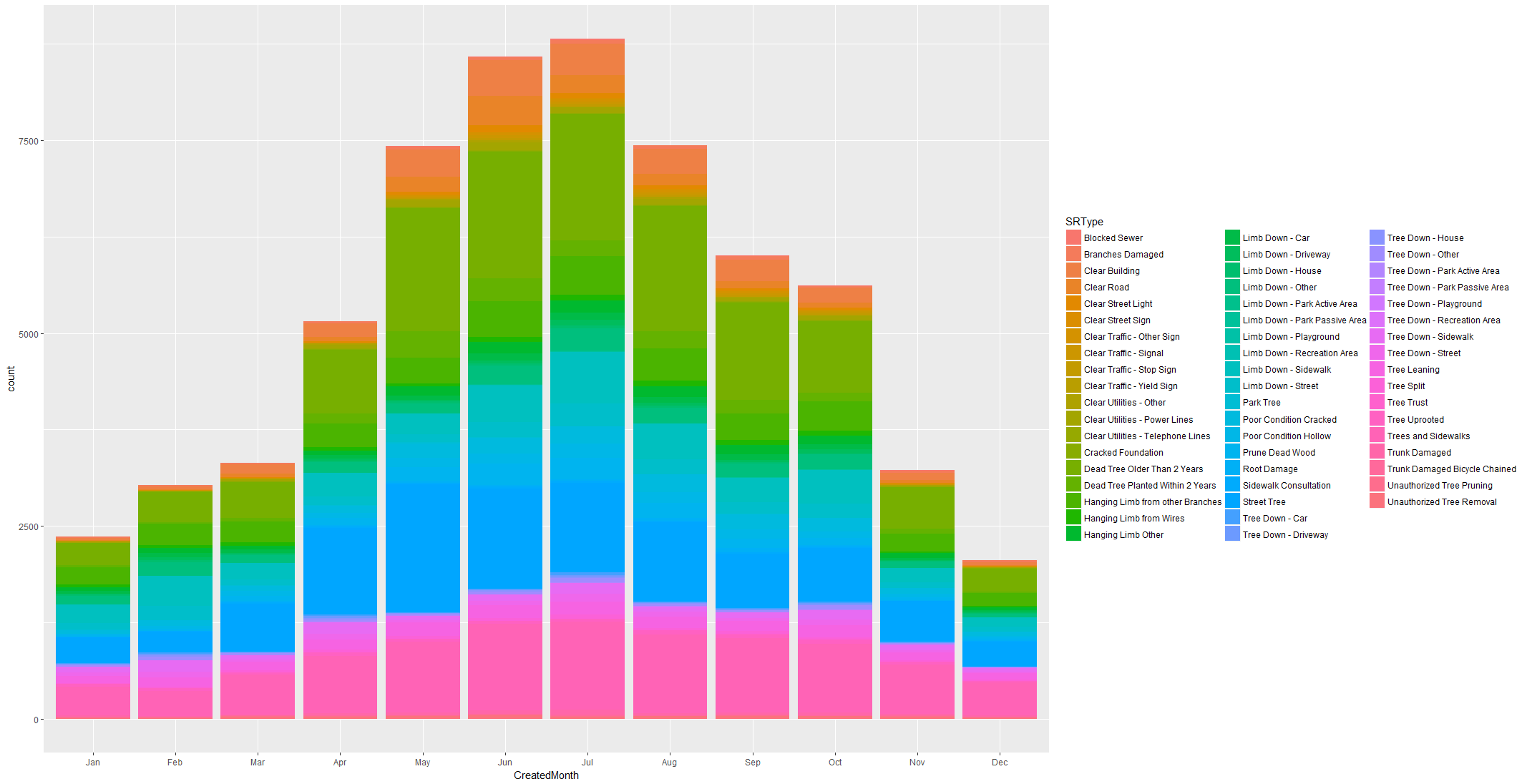


Complaints Type are further broken down into SRType which shows details of each complaint type which gives more detail on problem faced like Tree Down on Car or Tree Down on house etc.



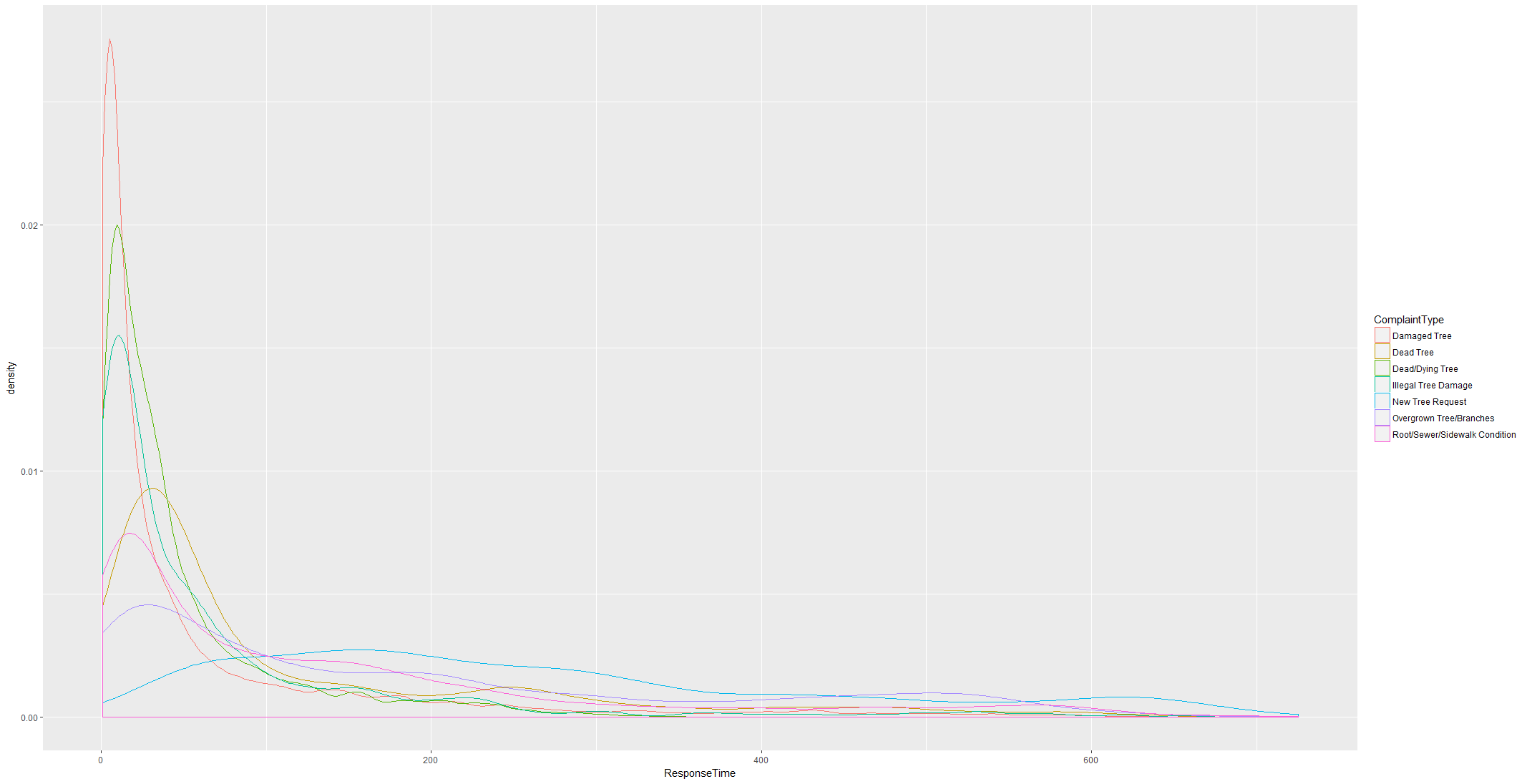
## Distribution of Service Request over Time

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.



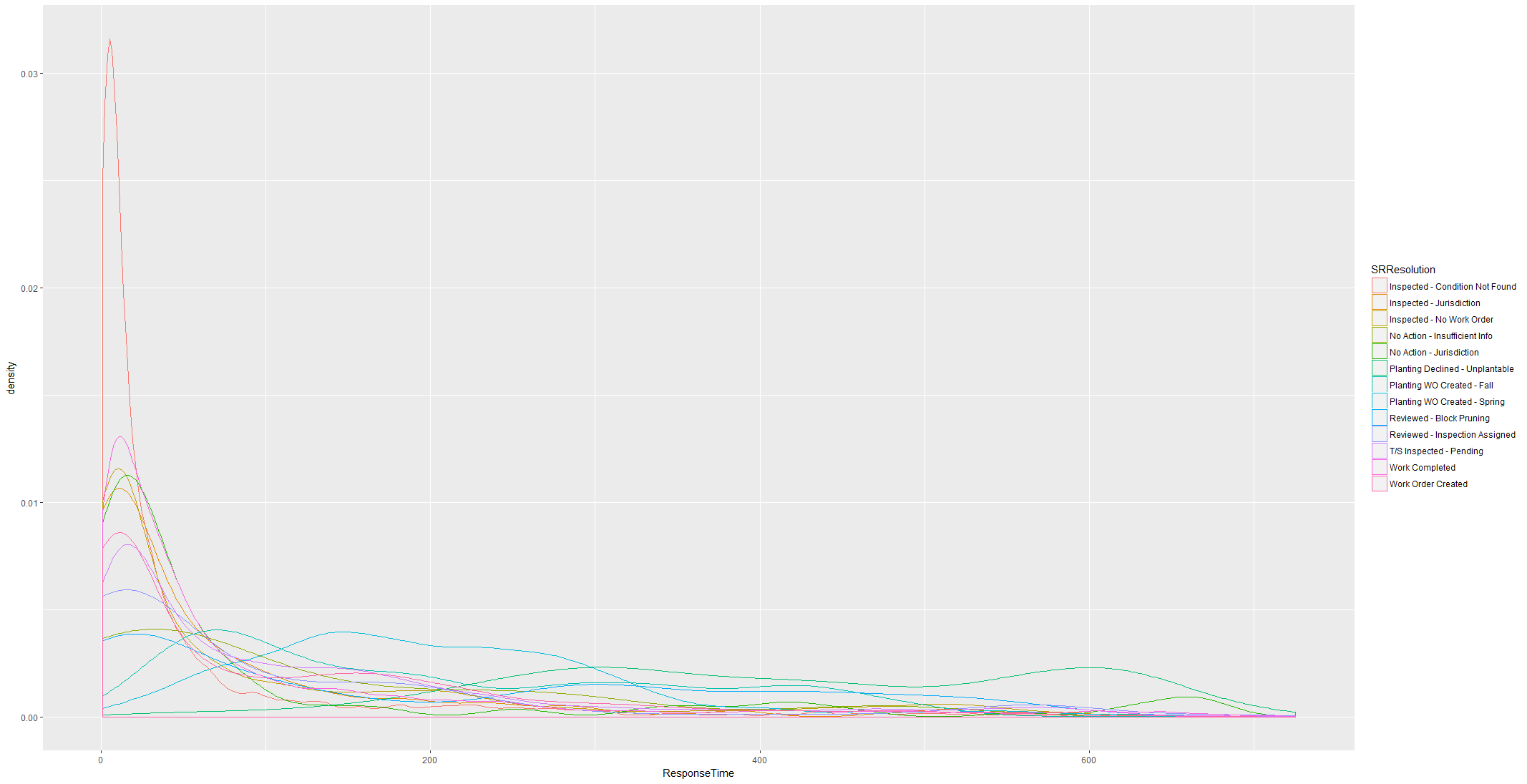
## Response to Complaints

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.



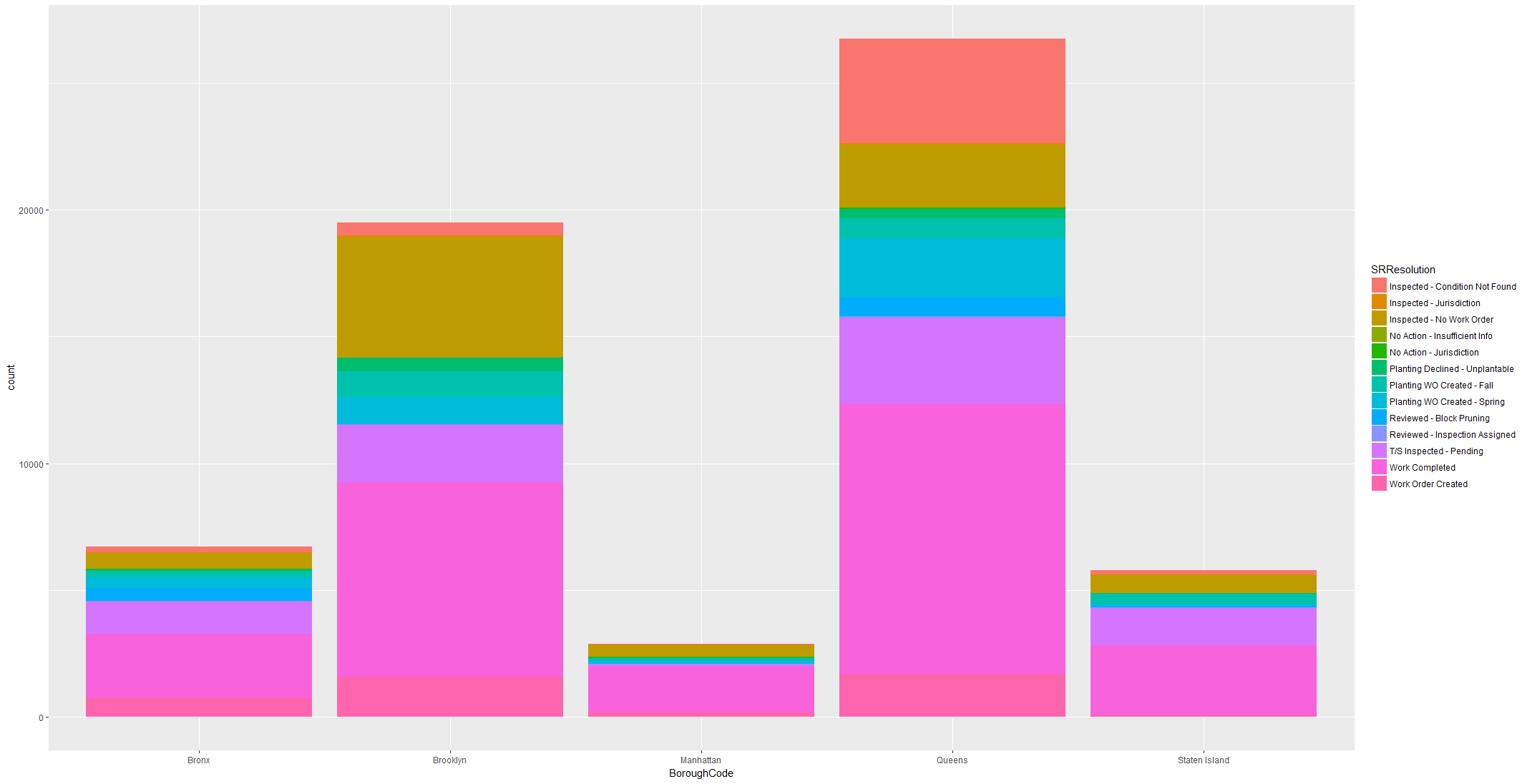
Response time varies with Complaint type. While Damaged Tree and Dead Tree are responded within 100 days, Root/Sidewalk conditions and overgrown branches have response time in more towards 200 days.

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.

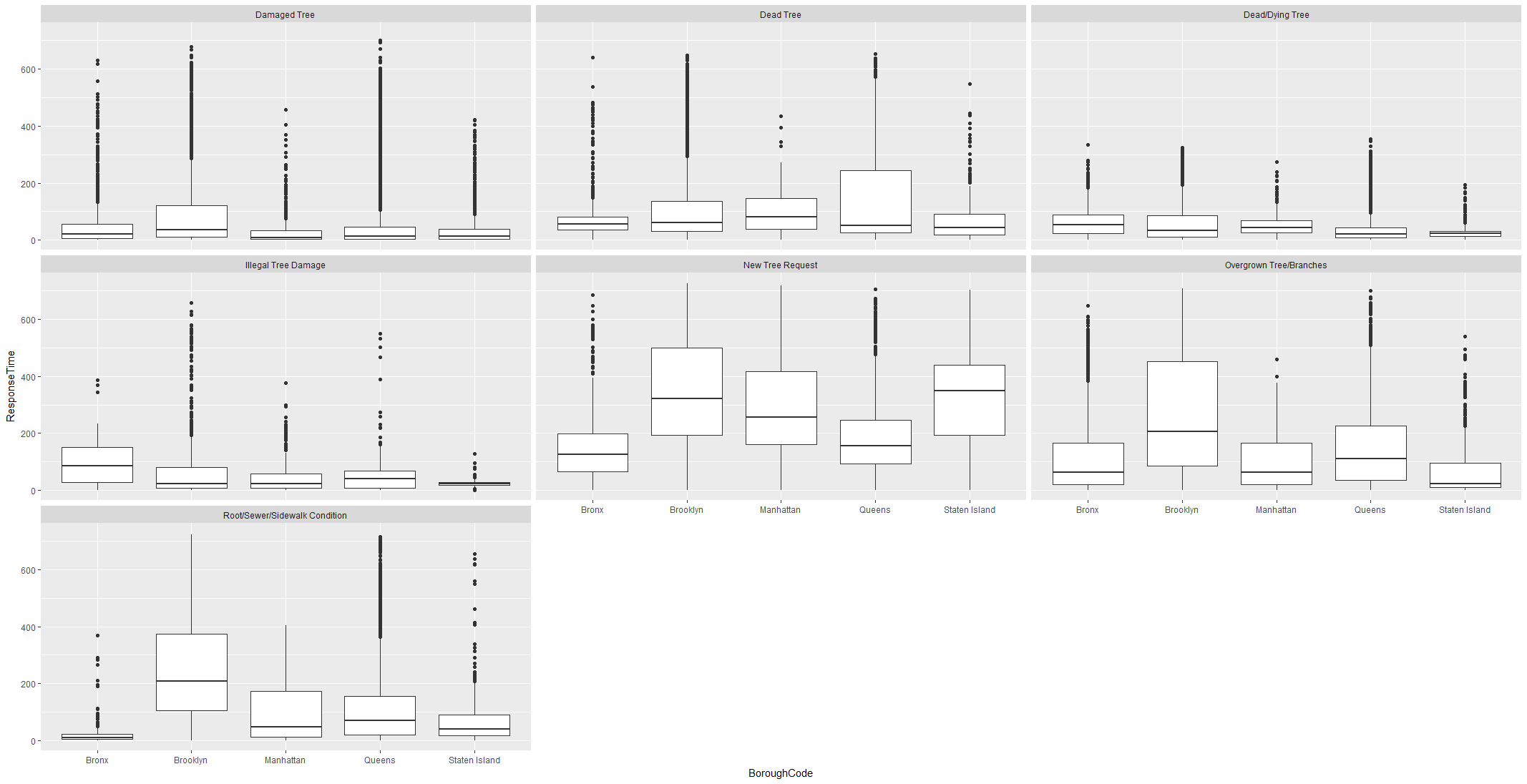


### Distribution of Resolution by location

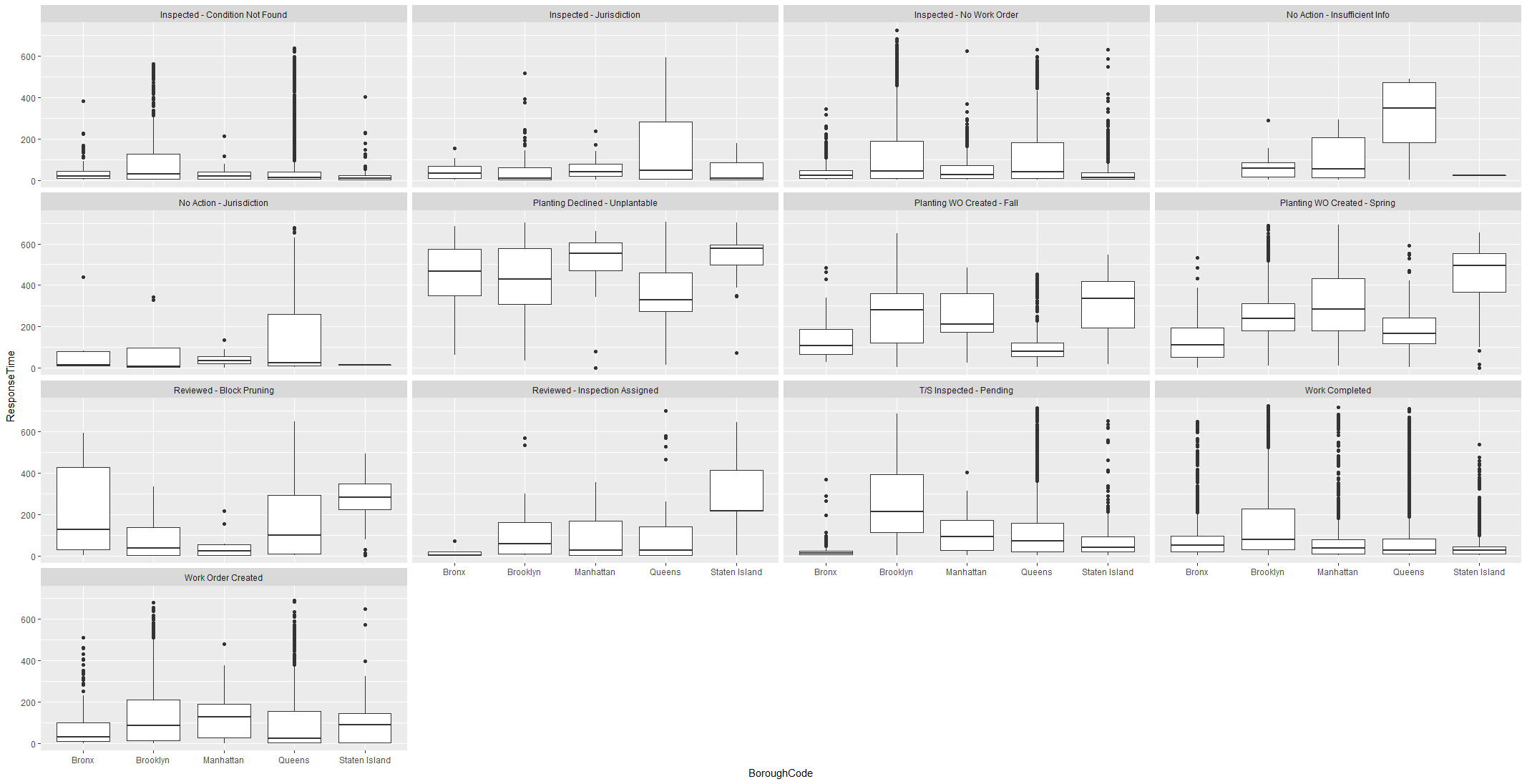
By plotting SR Resolution by Borough Code, it case be seen that Queens has most Inspected Condition Not Found as Resolution Type while Brooklyn has most Inspected-No Work Order resolution.



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.

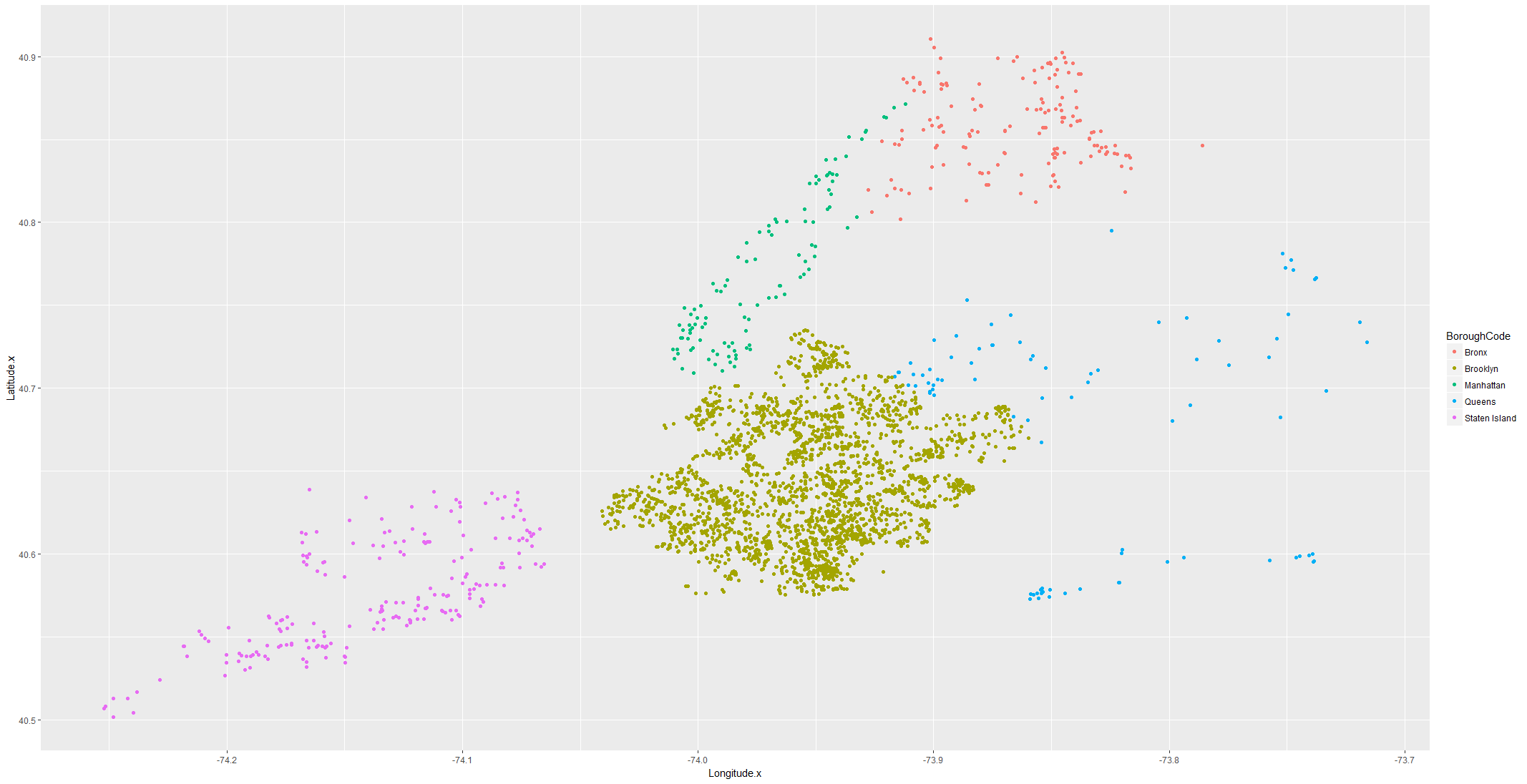
For Block Pruning , IQR for Bronx is around 400 days.

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.

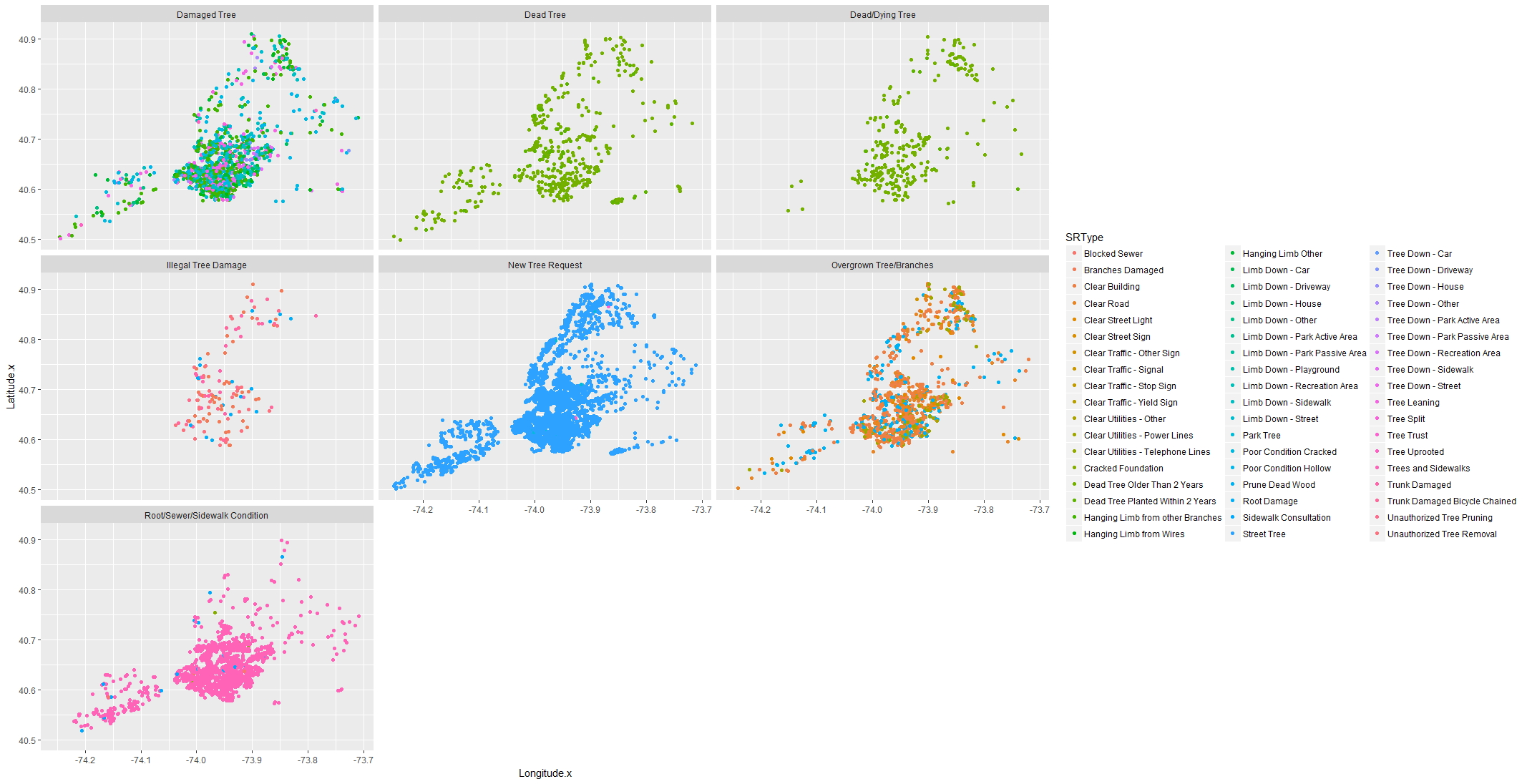
#### Curious Case of Brooklyn

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.

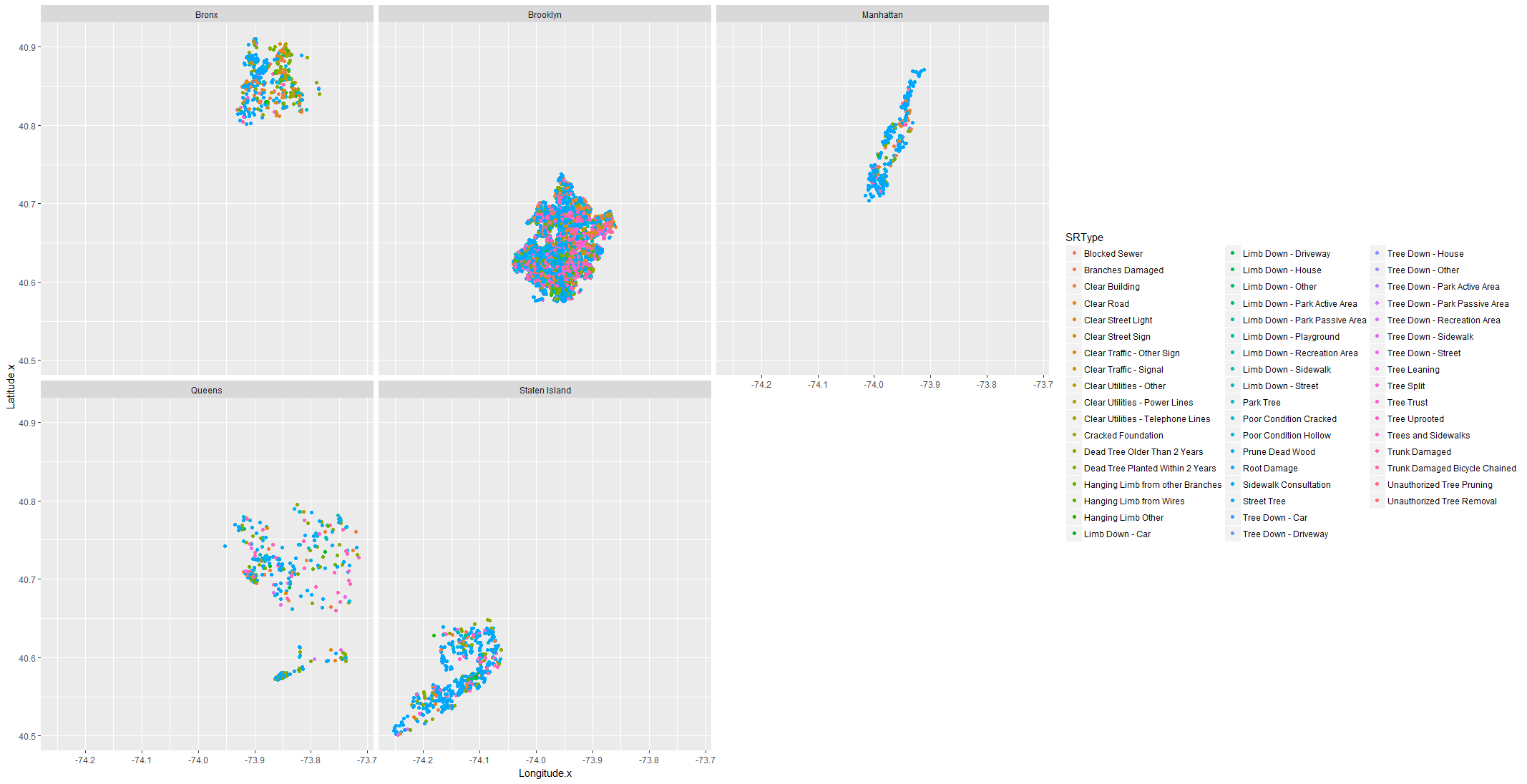
Brooklyn has more Root /Sidewalk complaints than other boroughs.



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.

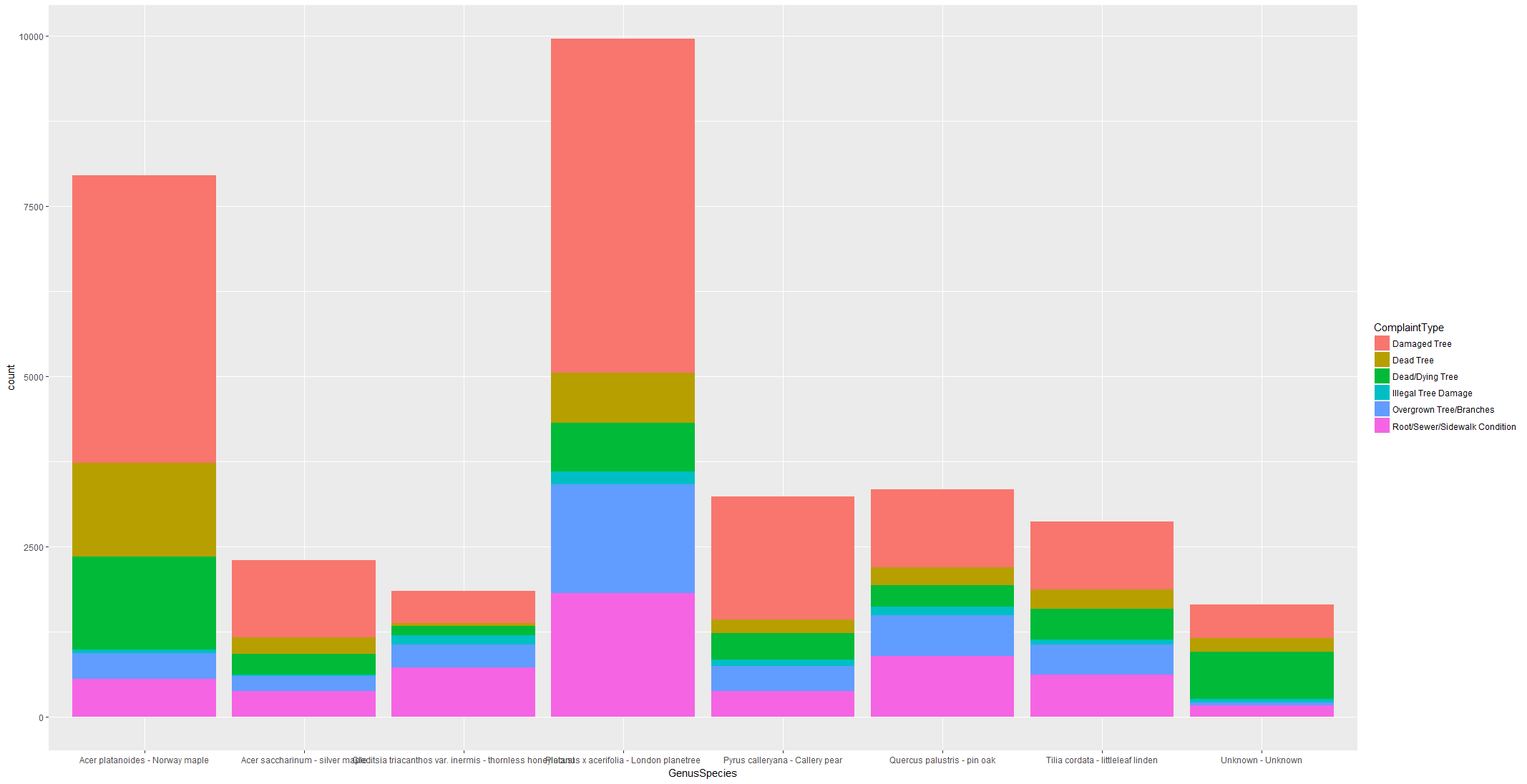


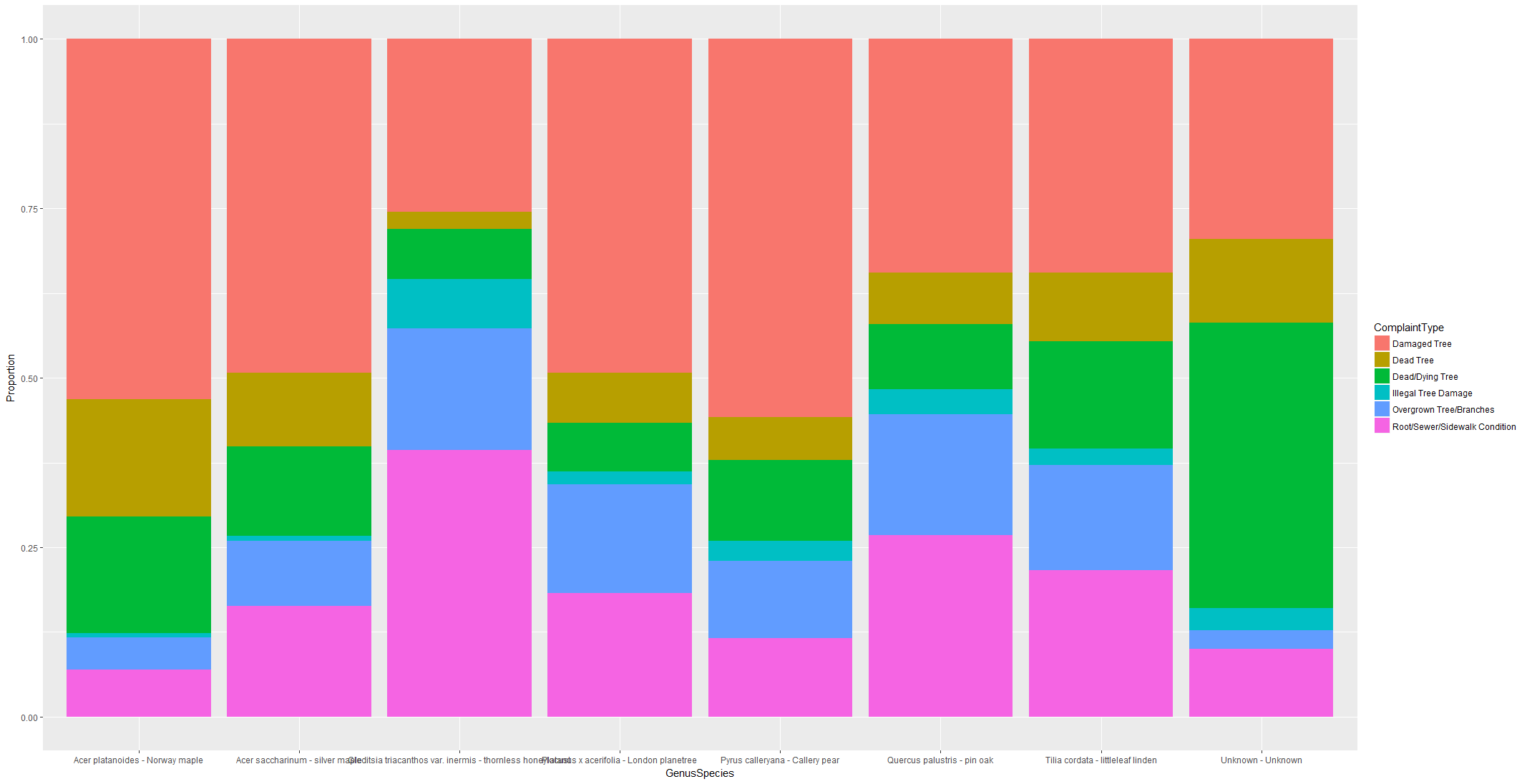
Plotting SRTYPE on Longitude and Latitude faceted by Borough for Complaints having more than 200 Days.



## Tree Species under request

Joining with tree point provides additional variable in terms of Species and Tree structure. Summarizing Species of tree under a service request we can find which species are contributing most in complaint types. This is plotted with species which have more than 1500 complaints.





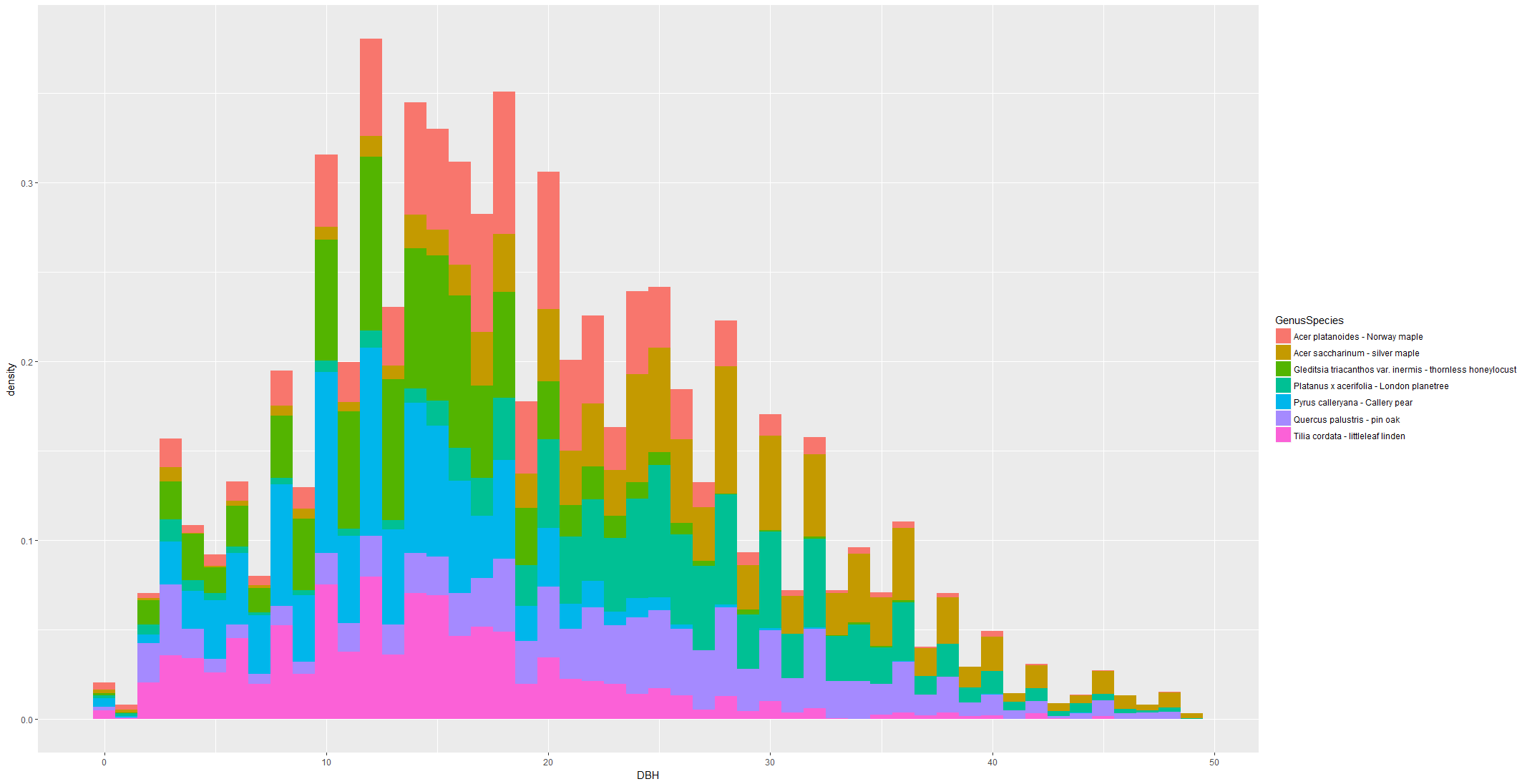
We can see that

1. London PlaneTree has most cases around 10000 and 50 % of request more of type Damaged tree, Overgrown Branches, Root/Sidewallk problem and has less count on Dead or Dying Tree.
2. Norway Maple has more than 50% cases on Damaged Tree and almost 40 % cases in Dead/Dying Tree.
3. Pin Oak has second most cases of Root/Sidewalk problem and Overgrown Branches and Thornless Honeylocust has similar proportion of requests.
4. Callery Pear has thirdmost Damaged Tree cases after London Planetree and Norway Maple.

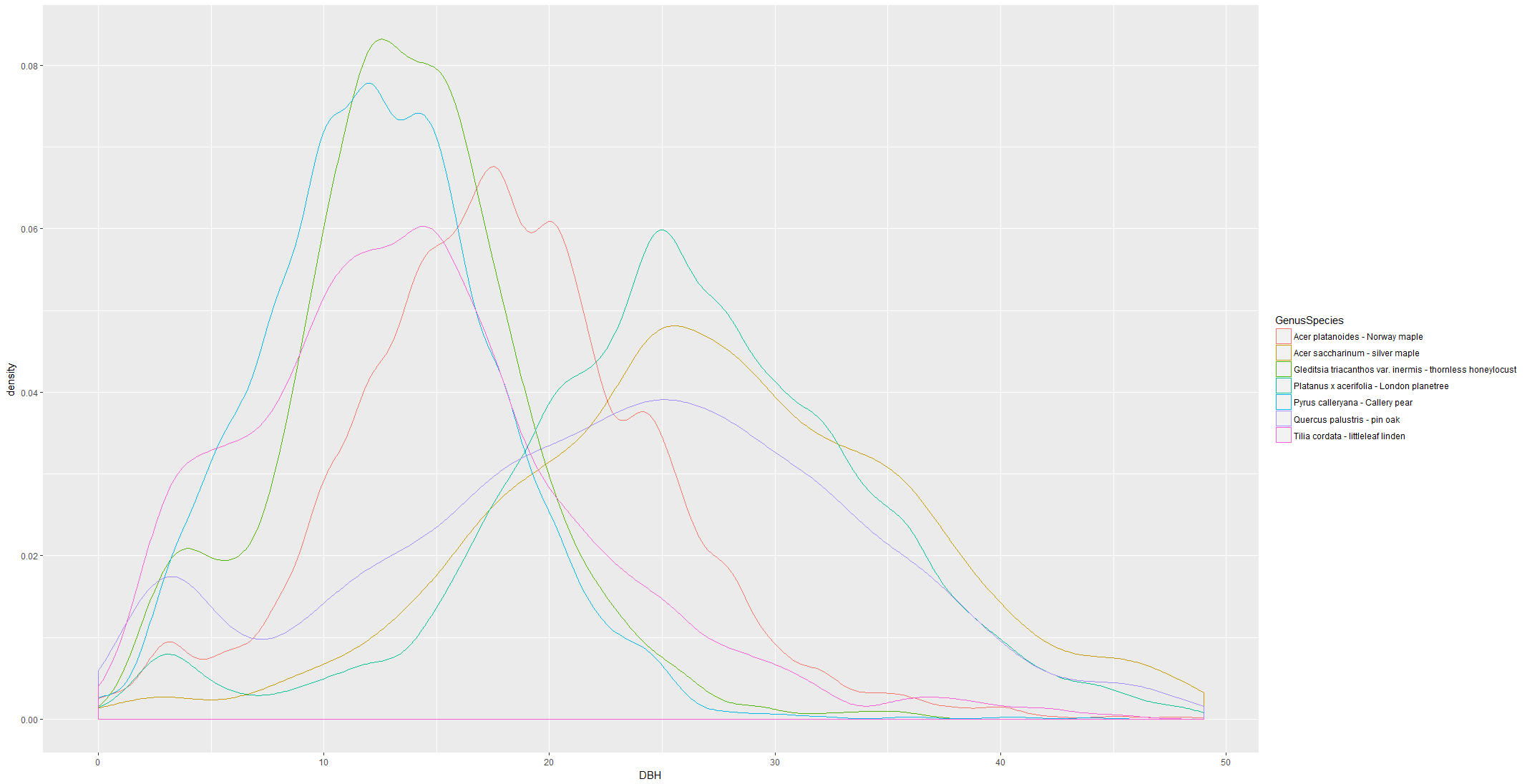
Comparing these counts with overall species count from FTP, We can see that out of total Norway Maple it percentage in complaints are more than others.

|  |  |
| --- | --- |
| Platanus x acerifolia - London planetree | 93856 |
| Gleditsia triacanthos var. inermis - thornless honeylocust | 63667 |
| Pyrus calleryana - Callery pear | 60861 |
| Quercus palustris - pin oak | 56527 |
| Acer platanoides - Norway maple | 38675 |
| Tilia cordata - littleleaf linden | 32171 |
| Zelkova serrata - Japanese zelkova | 31587 |
| Prunus serrulata 'Green leaf' - 'Green leaf' Japanese flowering cherry | 29059 |
| Ginkgo biloba - maidenhair tree | 22157 |

Exploring Diameter of these Species under complaint to see if specific tree diameter is related to complaint. Each species has specific Tree Diameter Range in the complaints.



Plotting Diameter Density for these species will provide cleared picture.

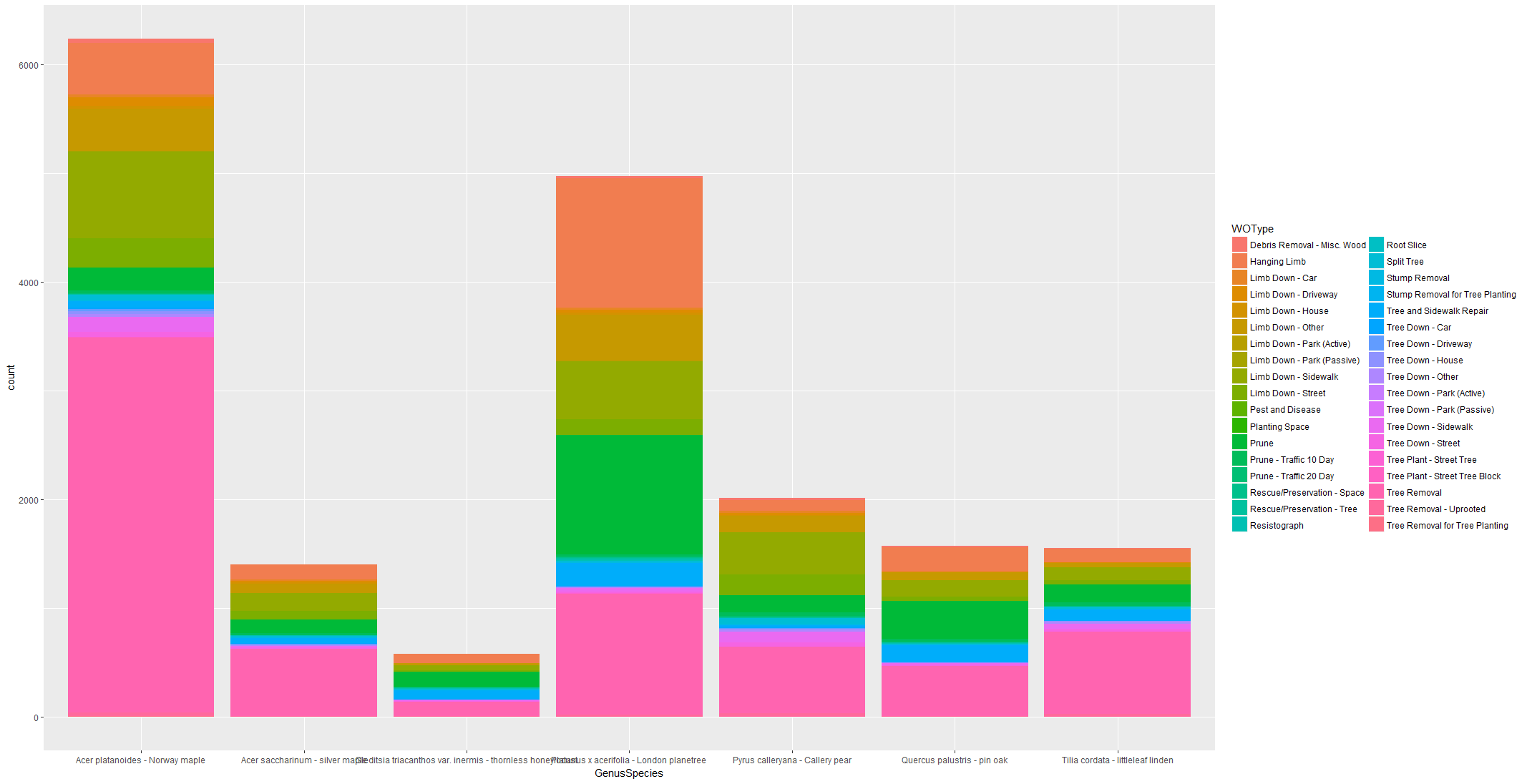


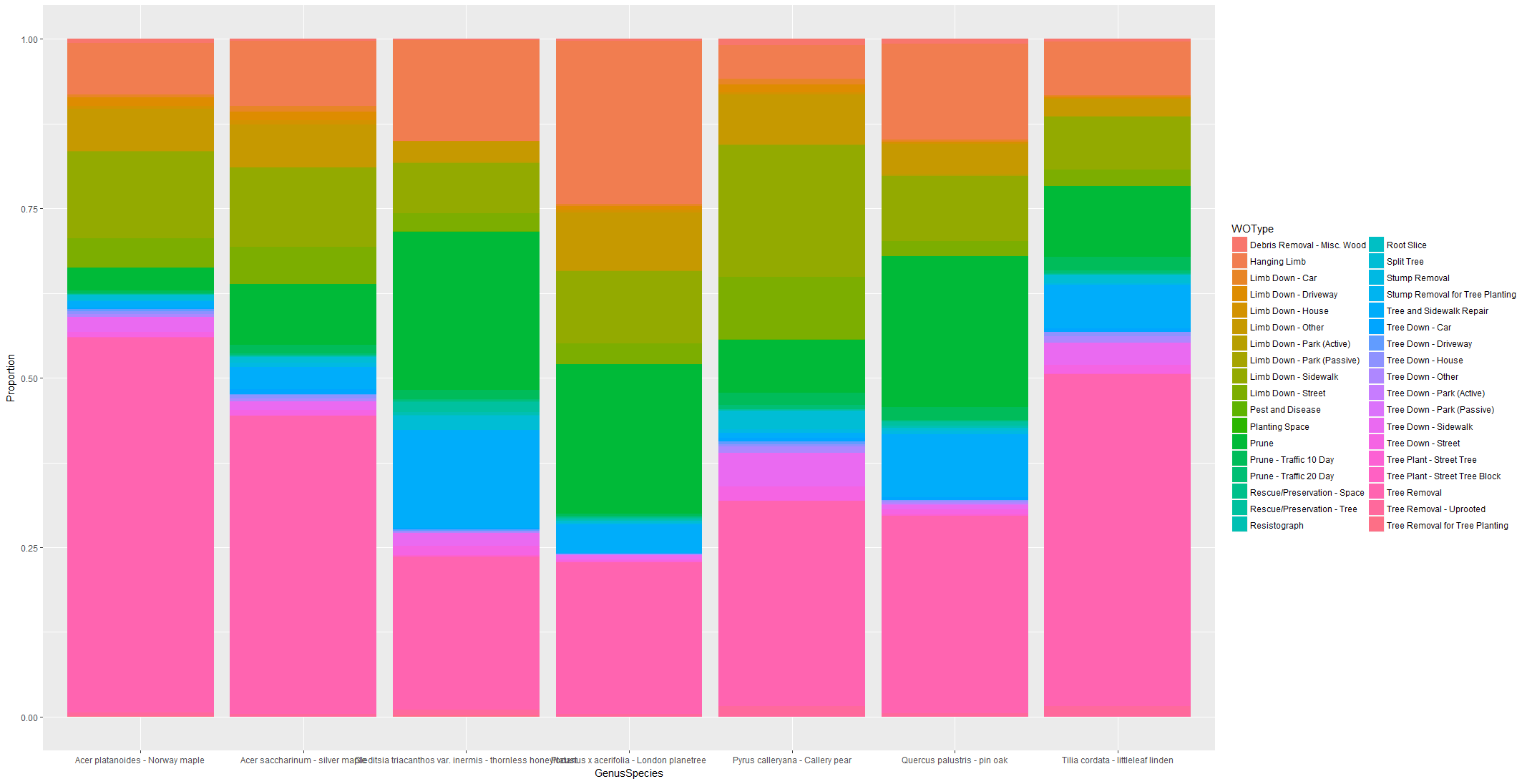
For each species Diameter is falling under almost normal curve.

1. Most of request for Norway Maple is between DBH 10 to 27 While London plantree most cases when DBH between 15 to 40.
2. Pin Oak has more cases at DBH more than 20 while Pyrus and Tilia as well as Callery Pear has most cases at less than 20 DBH
3. Thornless Honeylocust has more cases in 10-20 DBH

#### Work Order category on frequent Species

We are exploring if specific species has specific Work order that is done as part of complaint. For each species some of the work type is more prominent.



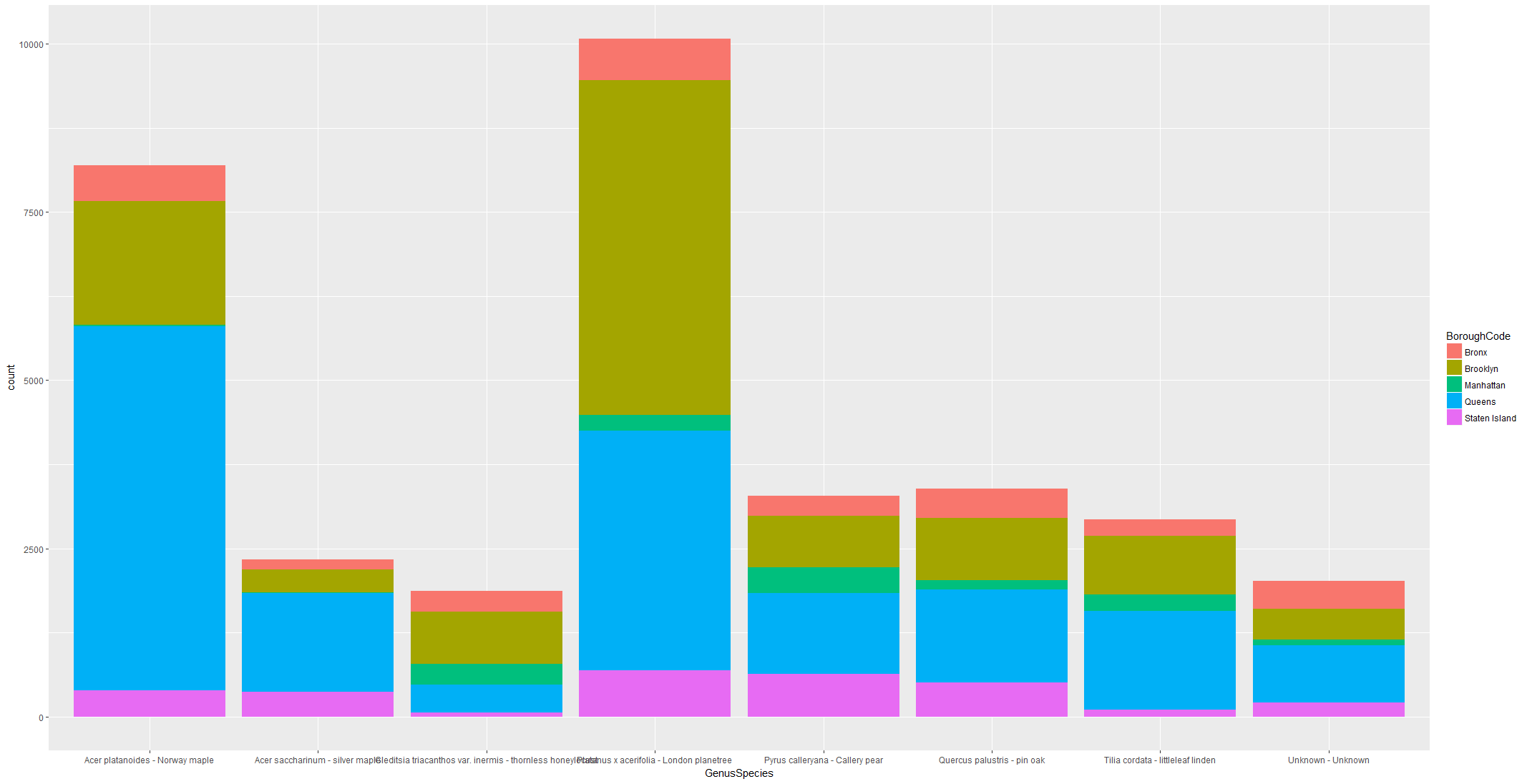


Relating Work order completed on these species, we can see that

1. Norway Maple has more work done even though it is second in complaints and it as more work type related to Tree Down to Tree Removal than Debris removal or pruning.
2. London plane tree has second most work type completion, with most of the work is related to Limb down, Pruning an Debris Removal.
3. Tree Removal in Till Cordata is almost as high as London Planetree while complaint raised on these species are 1/10 of London Planetree.
4. Thornless Honelocust and Pin oak has percentage of work work type related to Stump and Sidewalk repair after Pruning.
5. Cellary Pear has more percentage of work related to Limb down which relates to complaint type damaged tree initiated on these trees.

### Location of frequent Species:

Relating species under complaints and relating with location might provide some information on if any species at specific location is cause of request.



We can see that

1. Queens has most Norway Maple cases which are second most species under complaint and it can inferred that most of these cases are concentrated in Queens.
2. Brooklyn has most or second most presence in all species and comparing with Complaint type they have almost more complaints in each category only second in Tree Removal. Most complaints related to London Planetree are concentrated in Brooklyn which has almost siliar proportion of all complaint types.

With above information we can hypothesize that

1. Queens has more number of cases, these are more related to one Species Norway maple and Mostly Tree removal which are promptly responded.
2. Brooklyn has more variety of complaints on multiple species which is causing might reason for higher response time.
3. 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.

We need to explore further on

1. Are there any repeated request at same location and same Tree Point?
2. How many requests went for inspection and time taken for inspection?
3. Are there any delays in request to inspection and inspection completion to Work order which might be contributing in closure time?
4. 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 relate service request data with inspection and work order to get insight on what work is required to close a 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 which is causing duplicate inspections?
3. Is there trend in Inspection completion time, Work order completion time similar to service request response time?
4. Is there any species of tree which are contributing more towards complaints which can help to take prevention measures?

## Time intervals

Each service request goes through cycle of creation, inspection, Work order to closure. We are interested to find out how time taken from creation to closure as well as interval between life stages impact overall closure time.

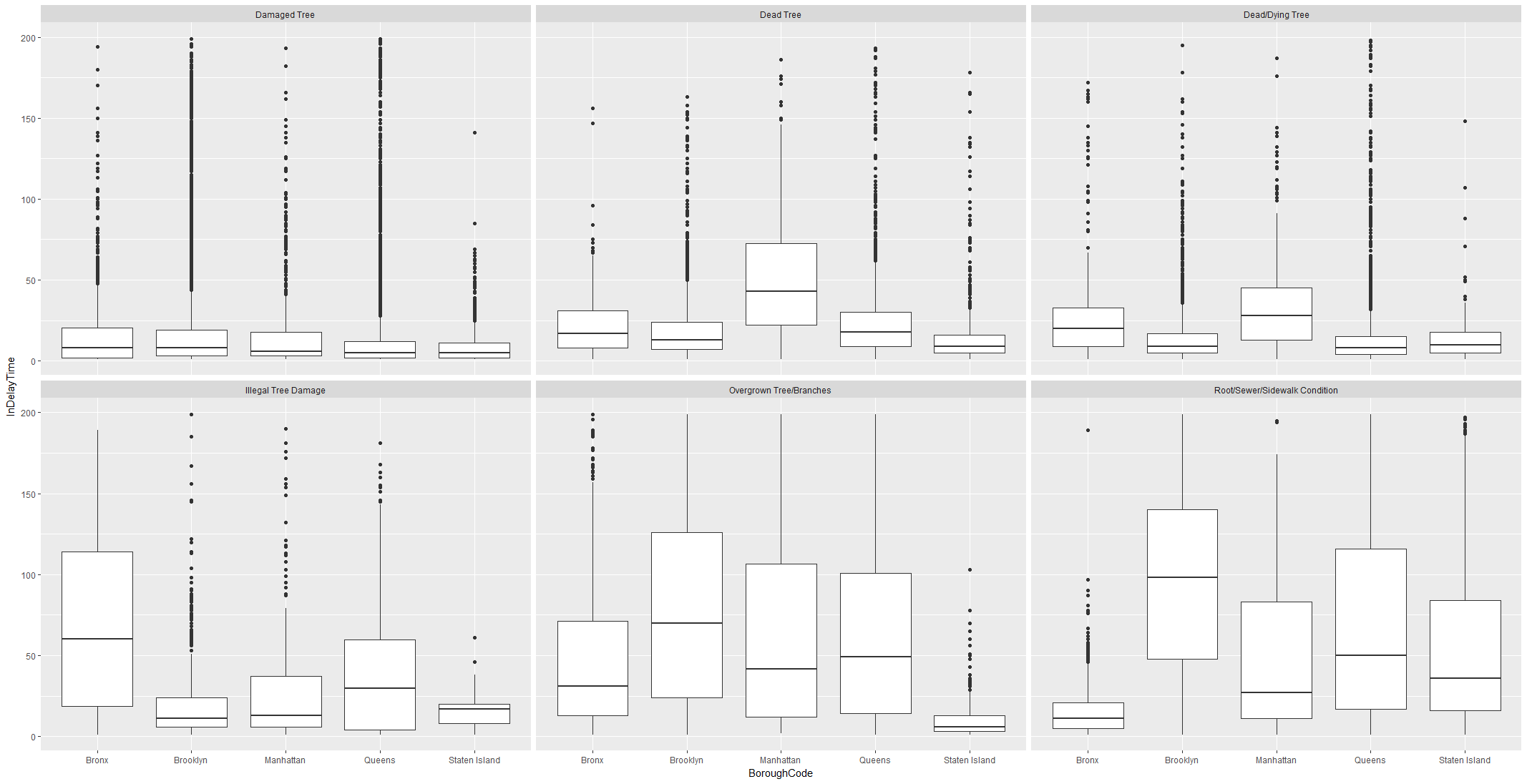
1. Service creation time
2. Inspection creation time
3. Inspection update time
4. Work order Creation time
5. Work order update time

## Time Intervals:

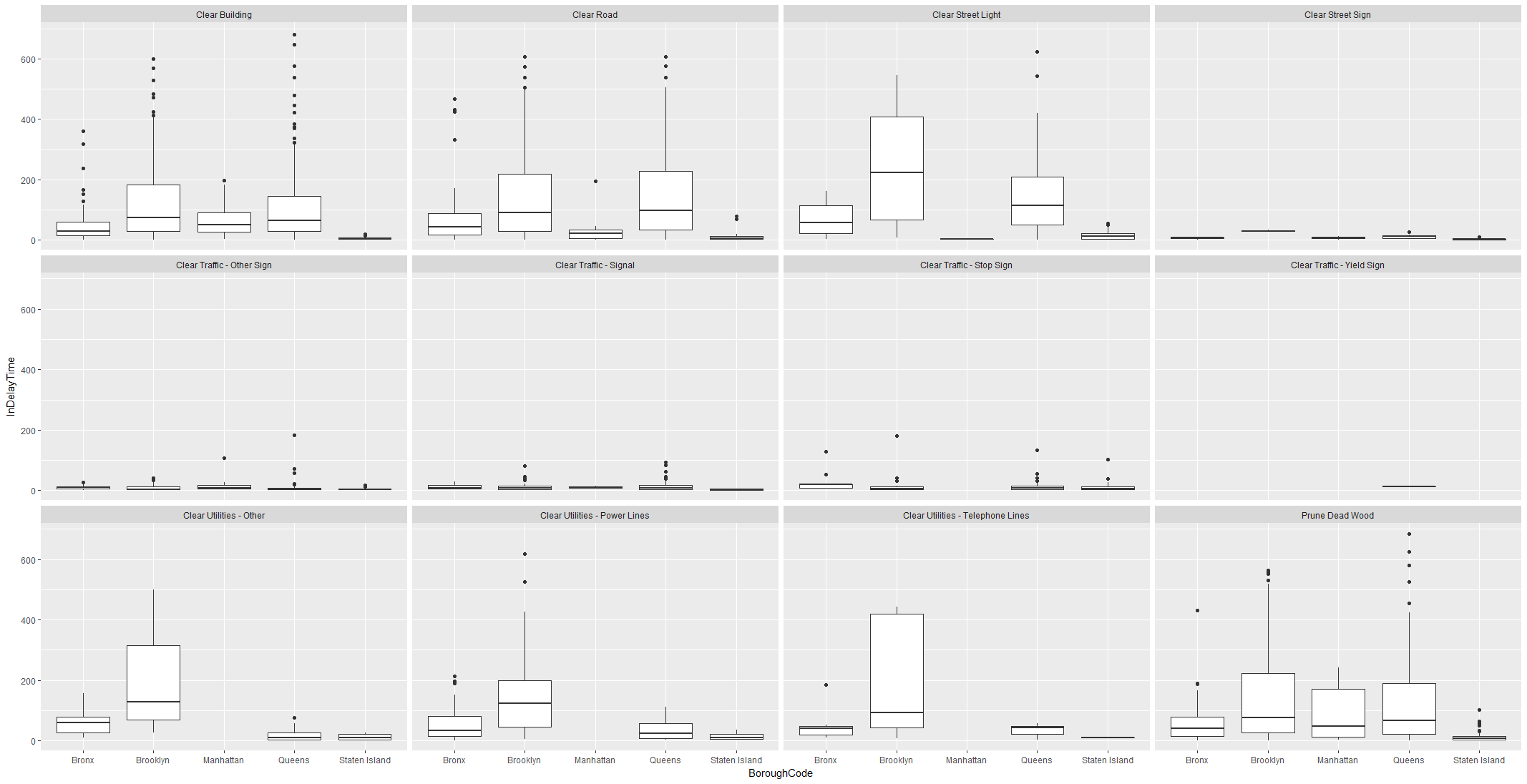
1. Response time: Service update date-Service creation date
2. Inspection delay time: Inspection creation date-Service creation date
3. Inspection closure time: Inspection update date-Inspection creation date
4. Workorder delay time: Workorder creation-Inspection update date
5. Workorder completion time: Work order update date-Work order creation date
6. SRWOclosure time: Workorder update date - service creation date

### Inspection Delay interval

Time interval from service creation to inspection initiation date. We can see that for specific complaints at specific Borough Median and IQR time varies. Plotting SR type gives much more detail on priority for inspection.



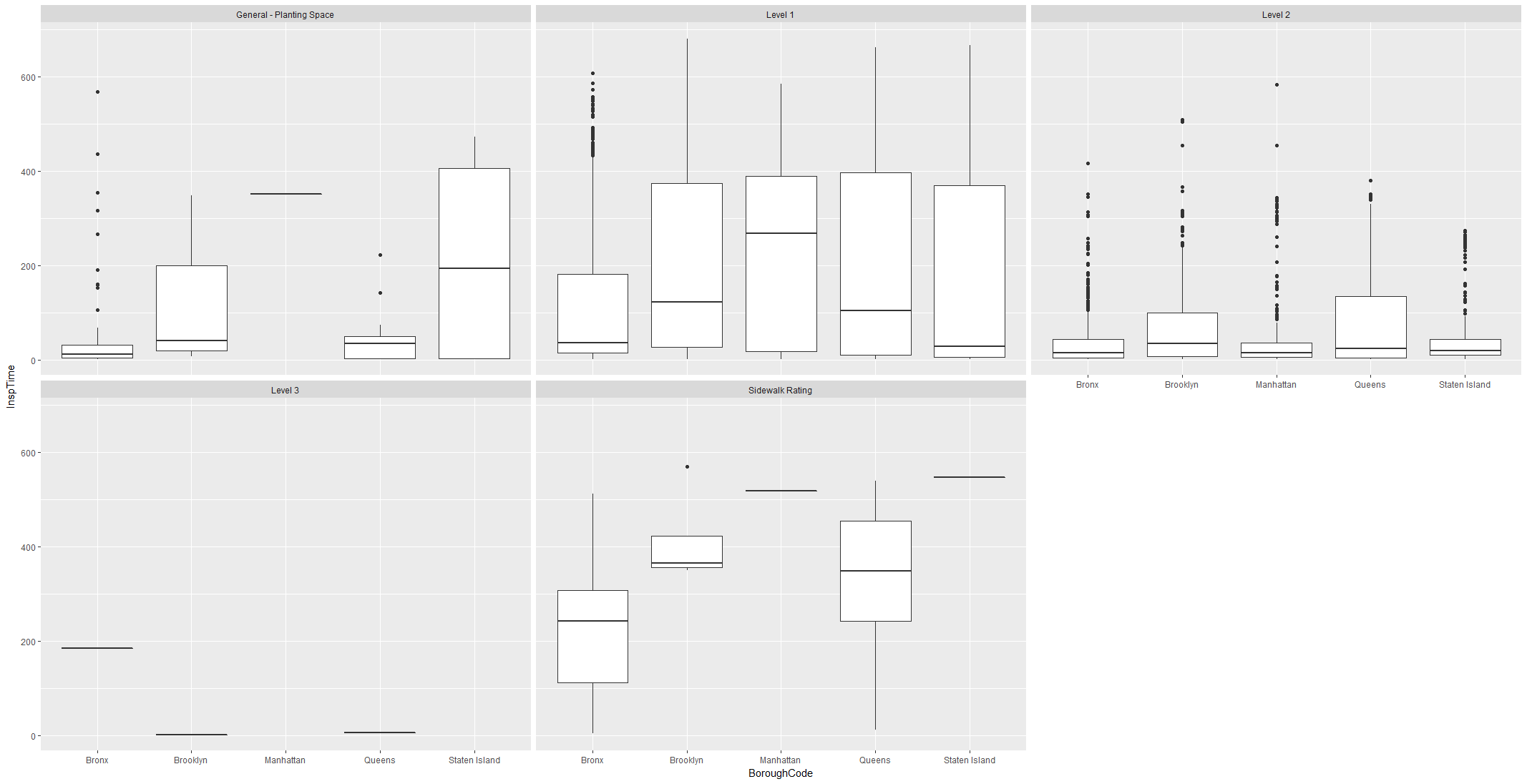
Digging deeper into one of the complaint type and plotting SRType will provide more detail relationship between inspection initiation delay per SRtype at specific Borough Code. For example lets take Complaint Type as Overgrown Branches & inspection as well as WO has been initiated and plot SRType and time taken to initiate inspection:



We can see that for a complaint type , inspection delay actually depend on actual service request type. For example Clearing Traffic Signal has been prioritized over other SRtypes. We can still see for each SRTYpe Brooklyn IQR is more than other boroughs.

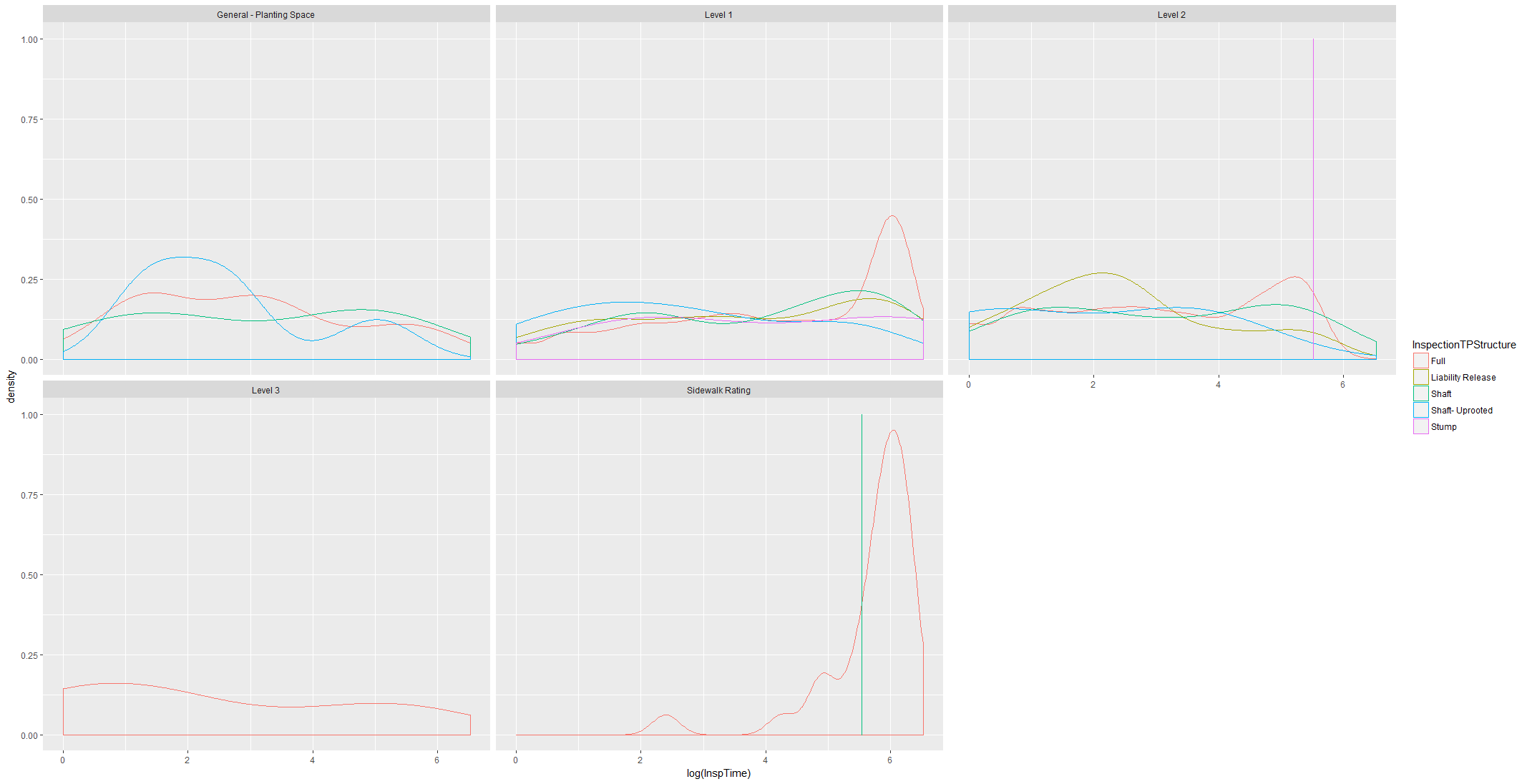
## Inspection completion time

Inspection completion time is difference between Inspection close or update date and Inspection creation date. There is different type of inspections designated by level of inspection. Plotting Inspection closure time should provide on relationship between inspection closure time and level of inspection for complaint type as ‘Damaged Tree”



Ideally level 2 inspection should take more time to close as it is detailed risk assessment but IQR is more for level 1 inspection type.

Is it inspection time is dependent on some other variable like Inspection Structure which is structure of Tree at point of inspection. Plotting log of Inspection time for various TP structure can provide some insight.

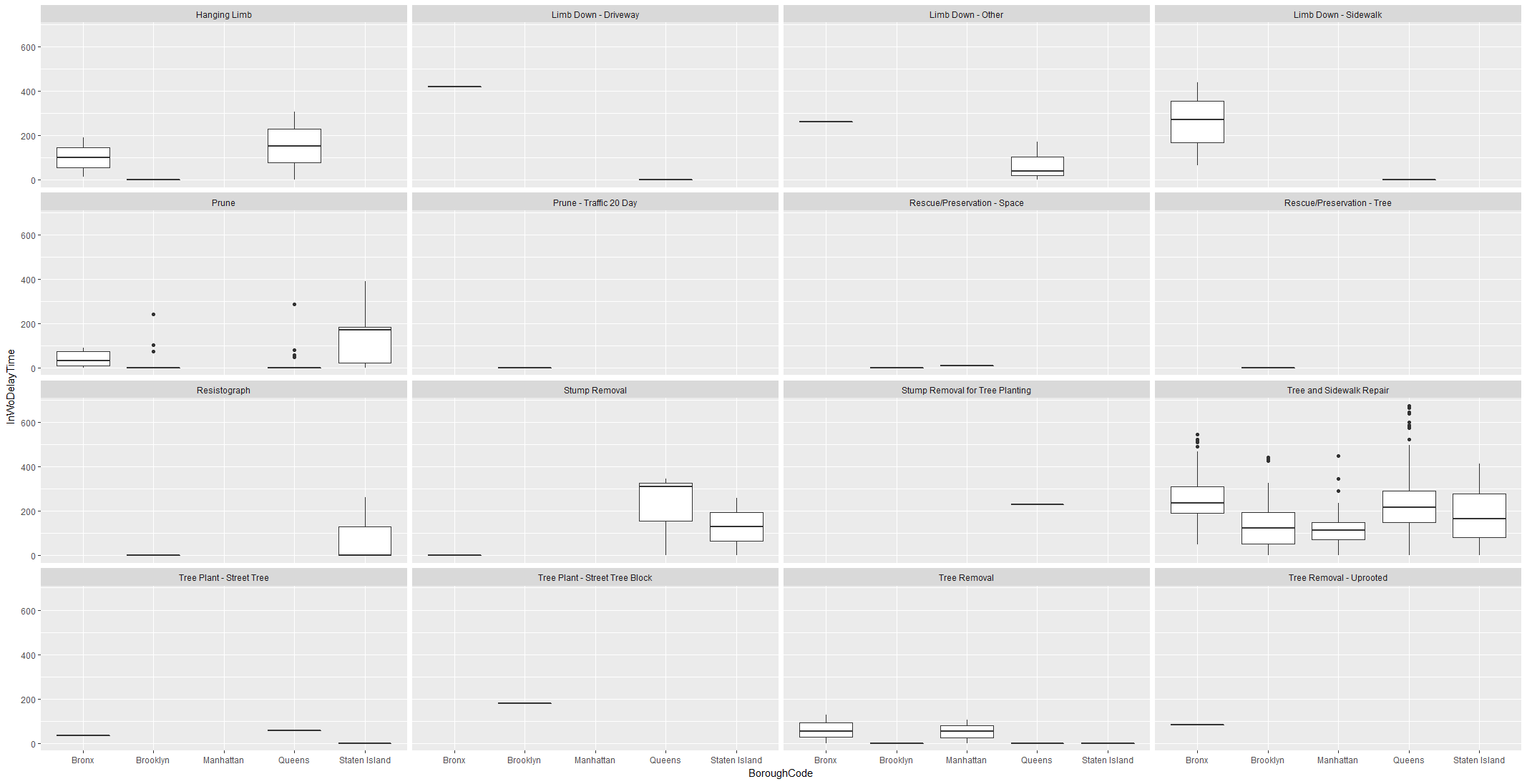


We can that for each level if inspection, probability of inspection time longer is more for full TP structure. It shows that relationship of inspection completion time on Tree Point structure instead of level of inspection.

## Work order delay time

Work order delay is time interval between inspection completion to Work order start date. Some of the values are negative it means that Work order is creation even when inspection is not closed. Given that we can see delay between Work order creation and Inspection creation date and plot it with WO type to see if WO Type has any relationship with Delay in Work order creation.

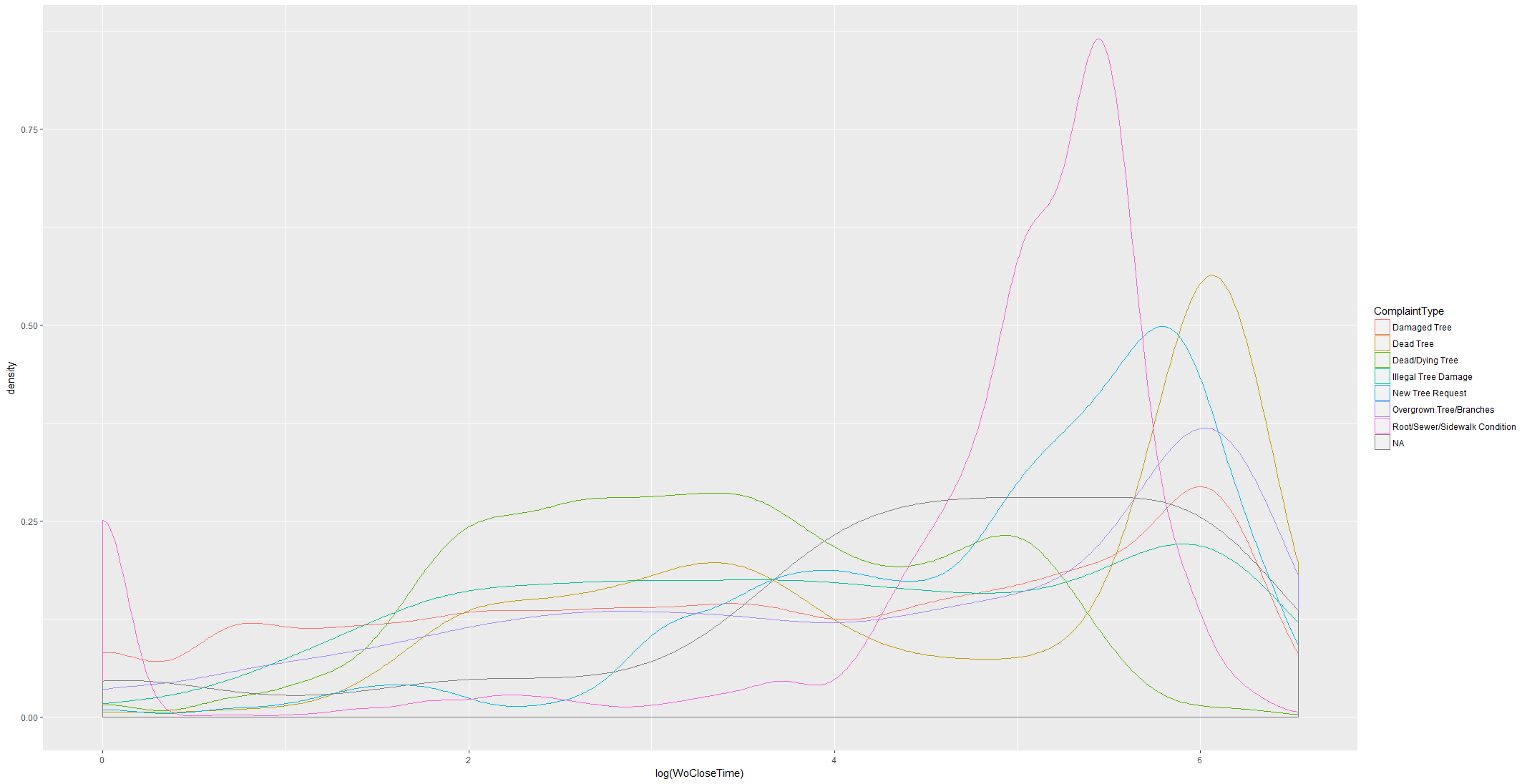
Following Plot showing box plot of Work Order delay time for complaint Type “ Root Side Walk Condition” per borough Code faceted by Work order type.



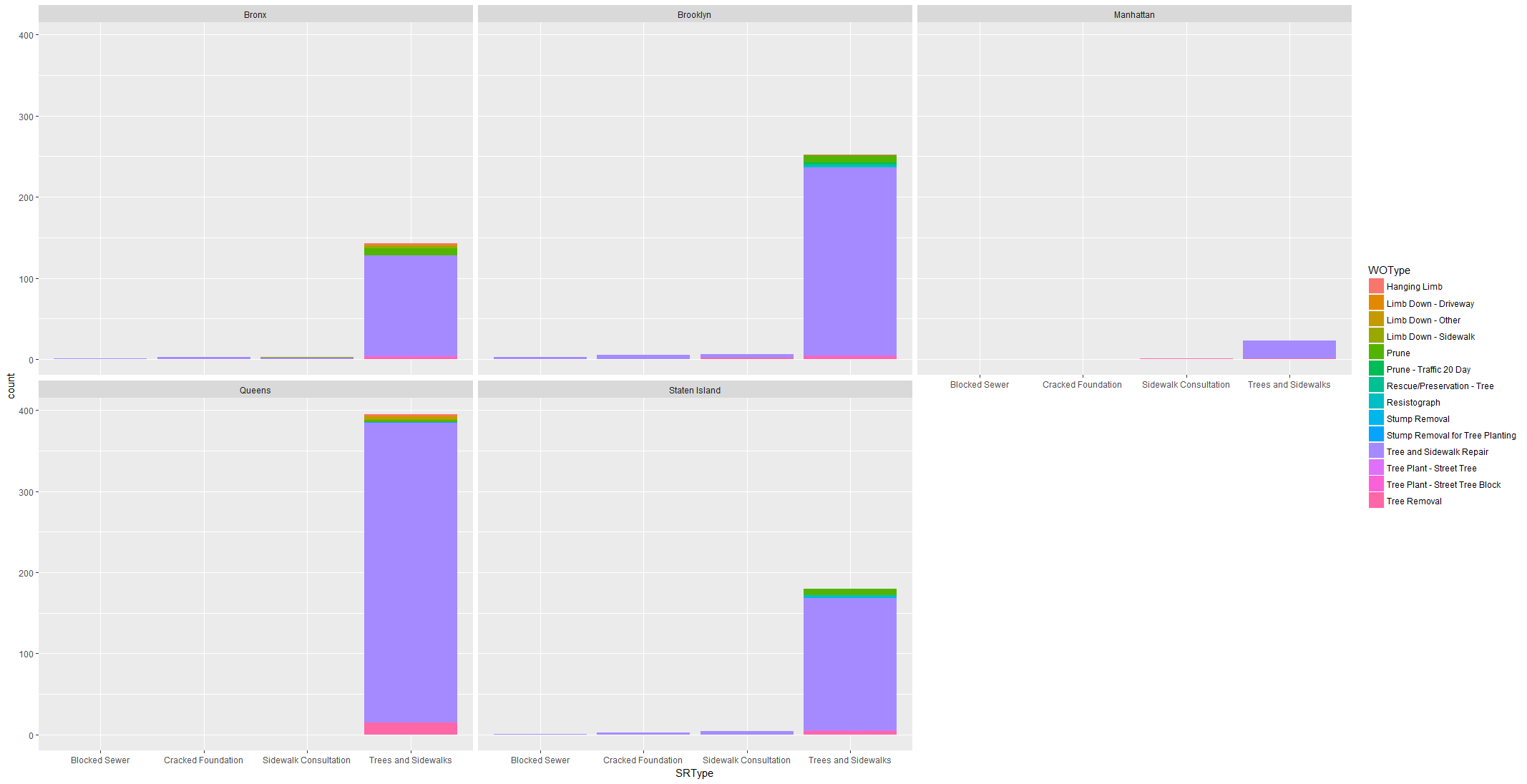
We can see that for Work Order Type Side Walk repair, Medians are more than any other work order type. We can infer that if work order type is Sidewalk repair a delay in work order creation can be expected.

## Work Order Completion time

Work order completion time is actual completion of work from Work order creation date to Work order completion time or update time. Plotting Workorder close time per complaint type we can see that some complaints have more Work order Close time.

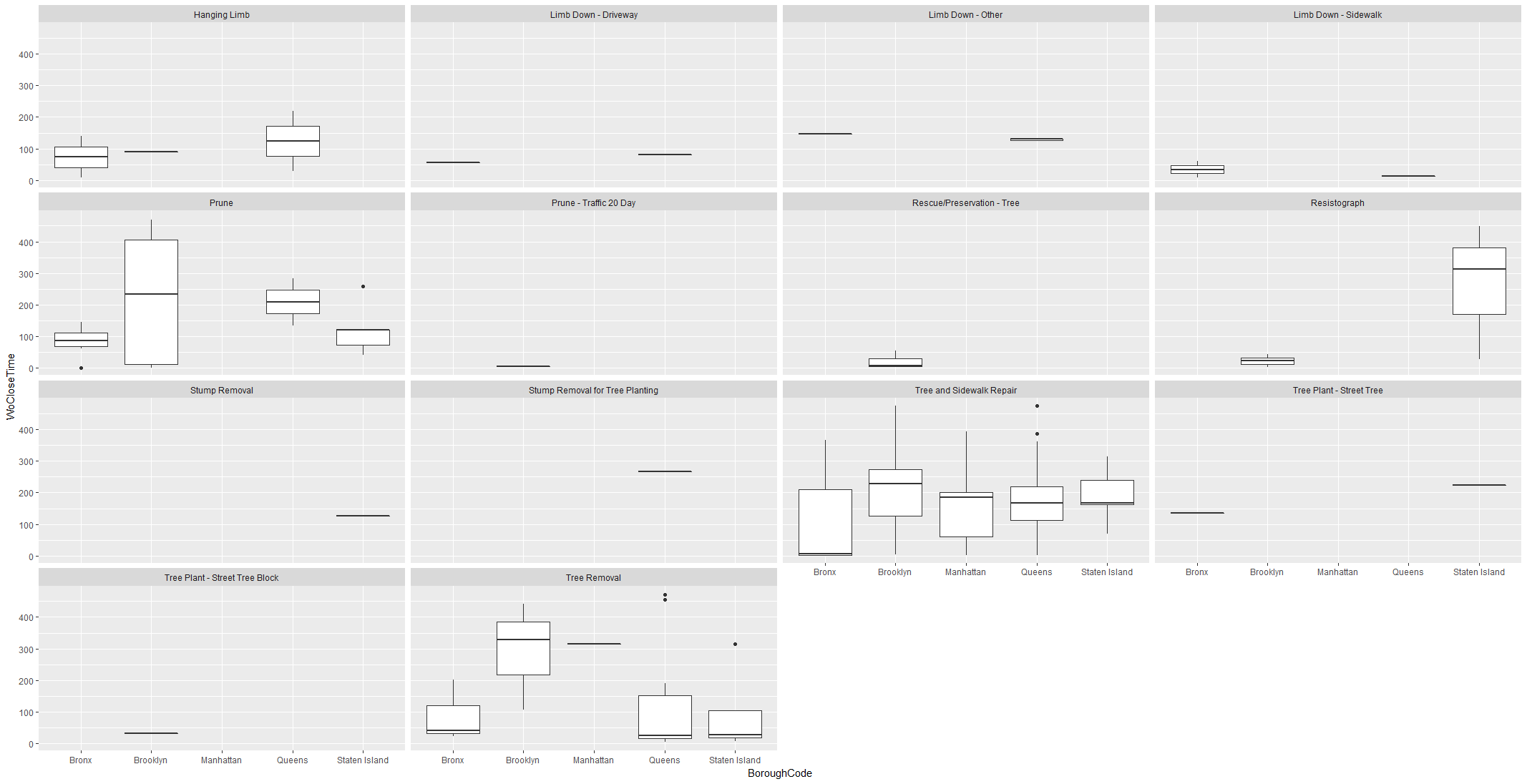


Complaint Type Root/Sewar/Sidewalk and Dead Tree has high probability of high closure time. Getting detail into Root/Sewar/SideWalk complaint , we need to explore further type of Work order completed for these complaints.



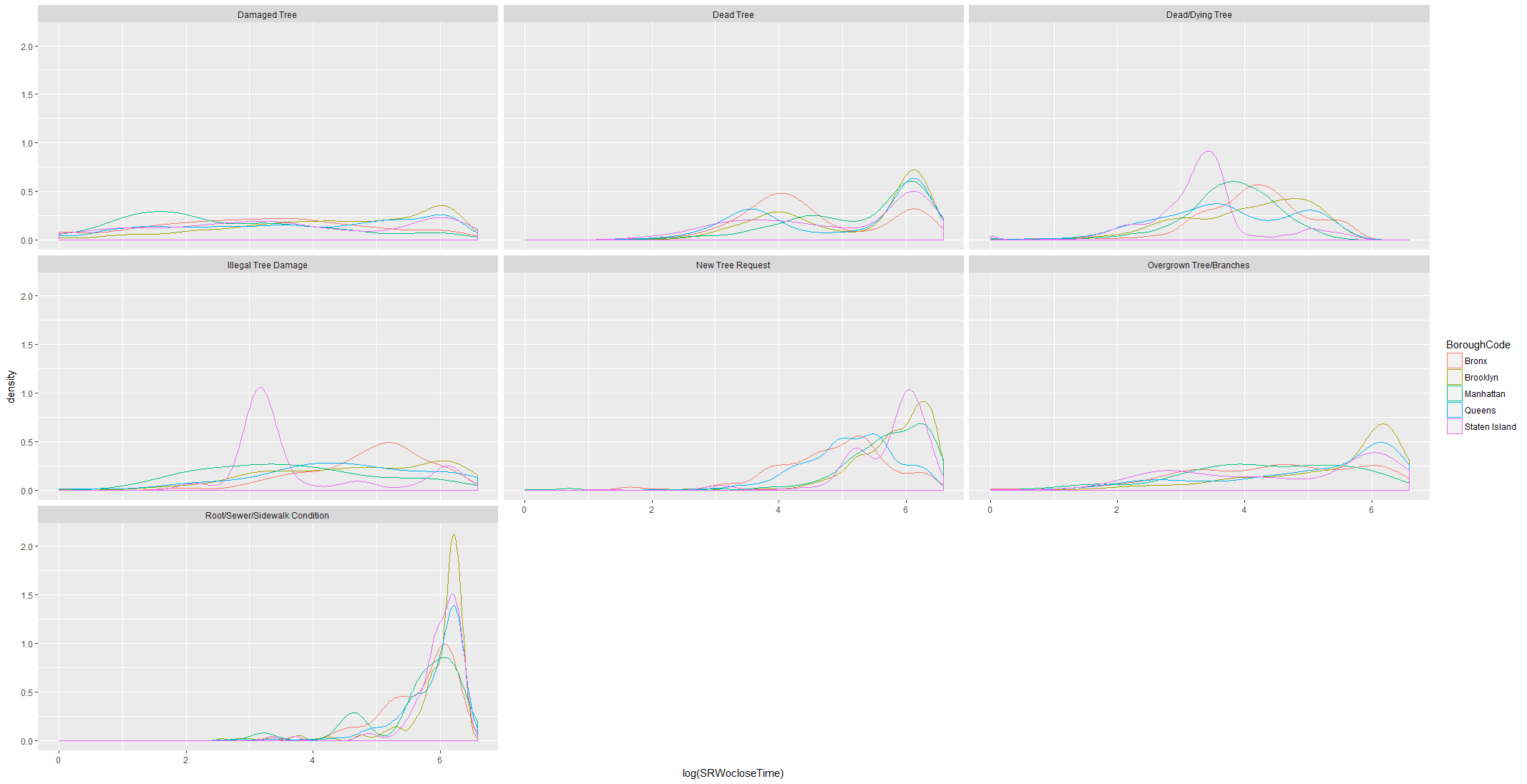
We can see that for SRtype Tree and Sidewalk, most of the work is Tree and Sidewalk repair followed by Tree Removal and Prune.

We also observed that count of SRType in Queens is more than that of Brooklyn. Let’s compare if WO close time has any relationship on Borough or SRtype for this complaint type. We can see that even though count is less for Brooklyn than Queens but IQR is range is more than other for Sidewalk Repair and Tree Removal.



## Overall Closure Time

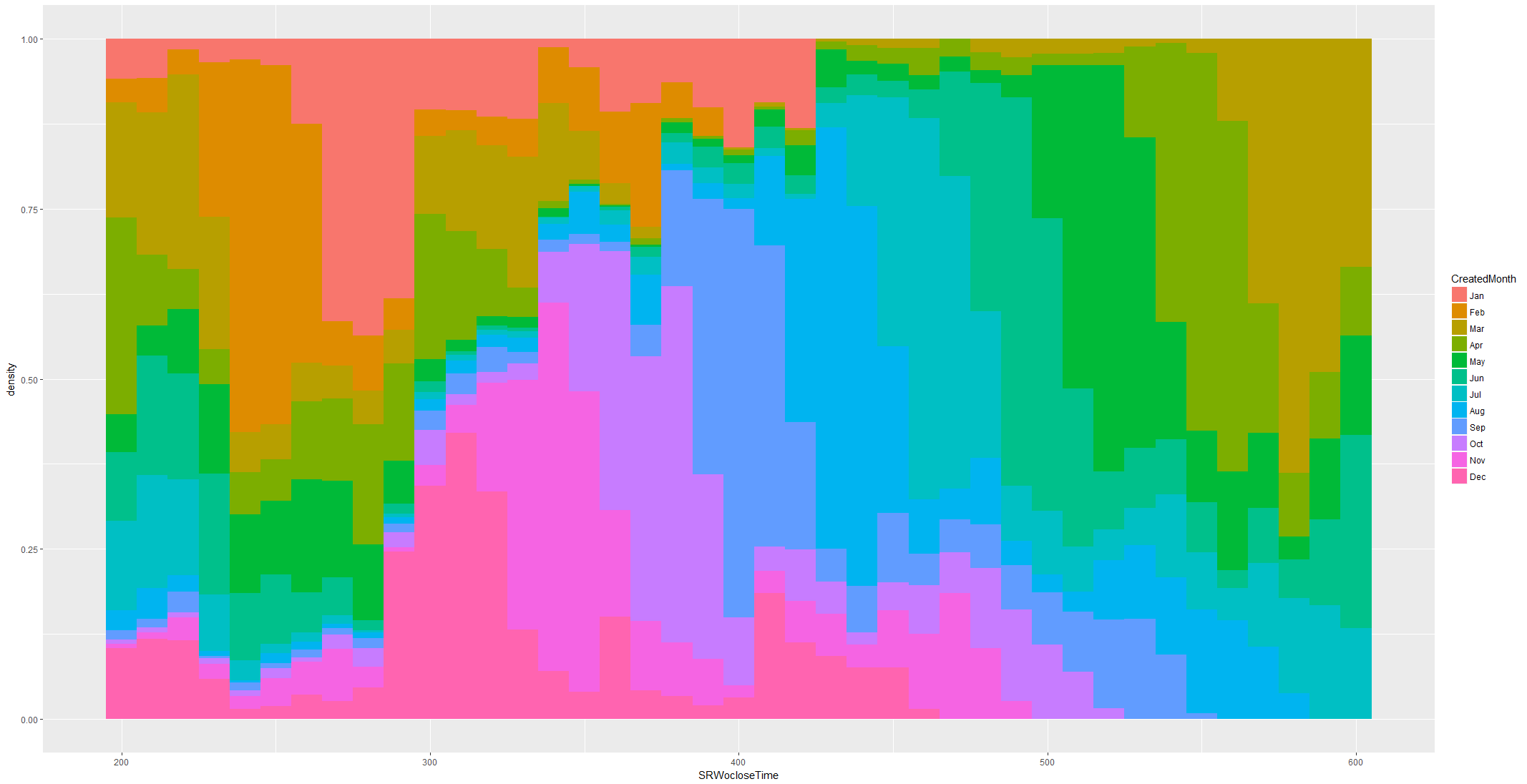
Overall closure time is time interval between Service Creation to Work Order Closure. Let plot for each Borough facet by complaint type.



With above plot we can see that

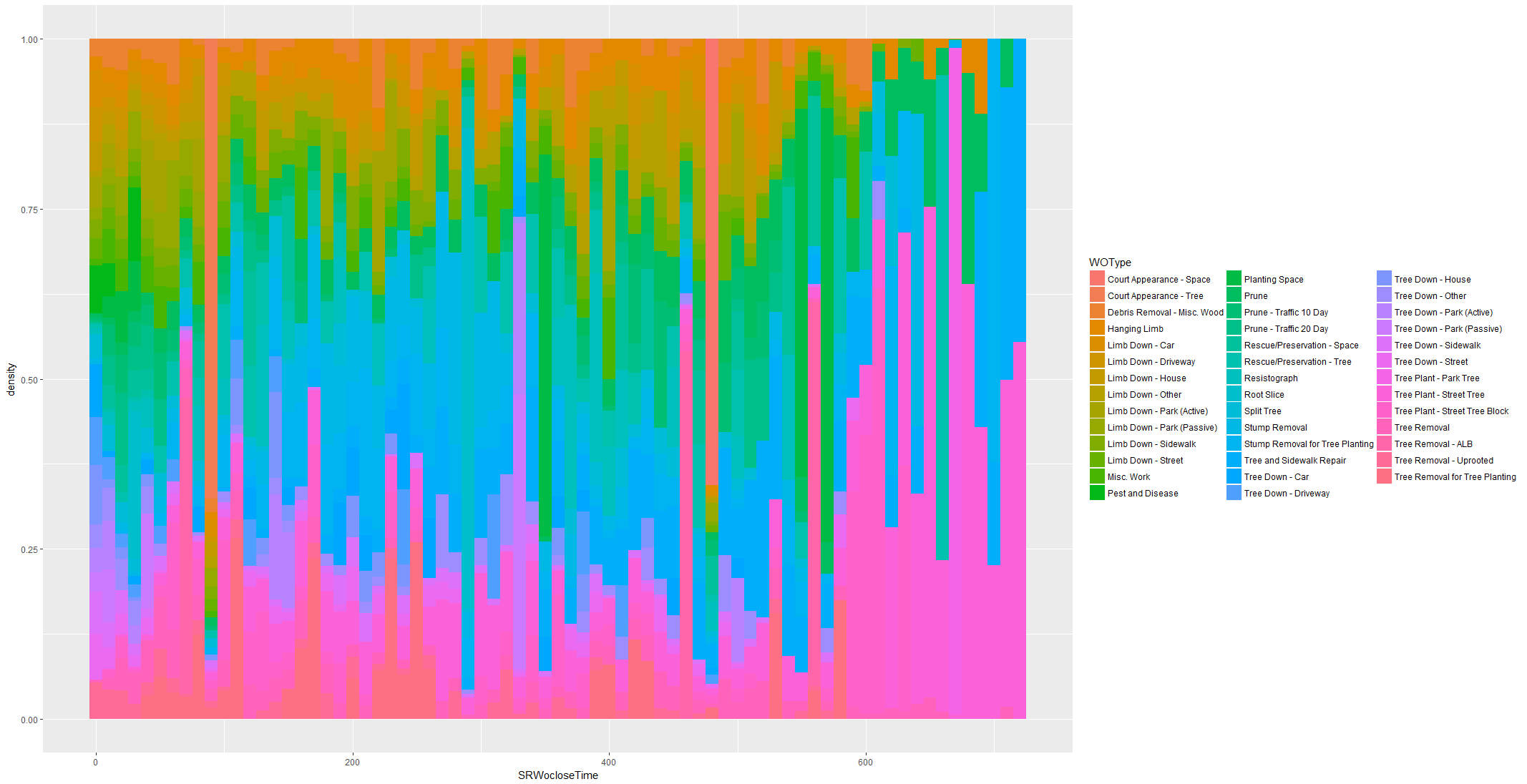
1. Root/Sewar/Sidewalk has more cases in higher time range with Brooklyn has higher density than others. Sidewalk Complaint in Brooklyn there are more chances that it will take more time to close than other Borough.
2. In all complaint Type , if complaint is from Brooklyn it is highly likely that closure time will be higher than other Boroughs.

Lets check if Closure time is related to service request creation month.



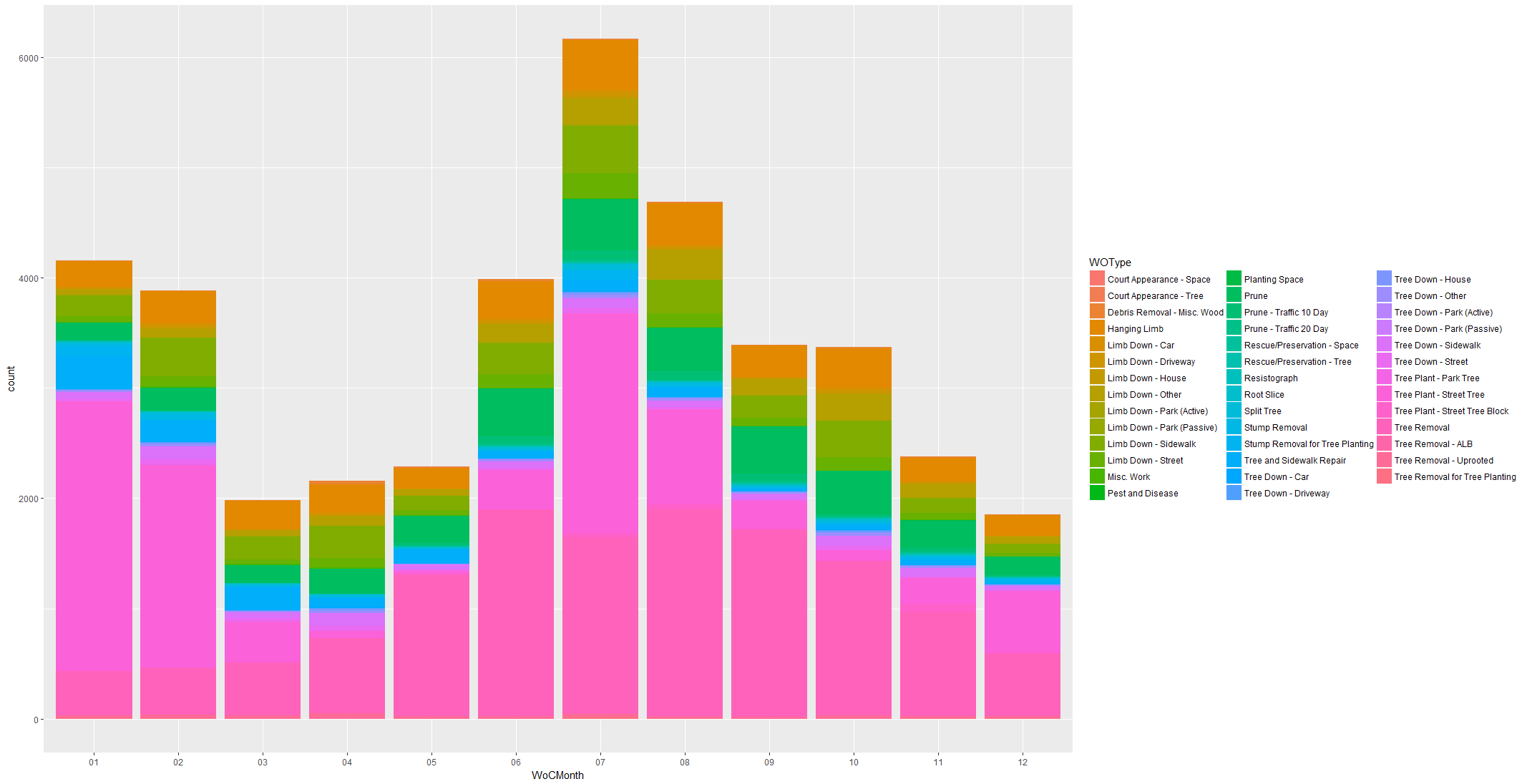
We can see that Case which have more like to have more than 400 days are created in March to Sep.

By plotting Closure time with WOType , which relates type of work completed and time taken to close the request we can see that Tree Down Sidewalk WOtype is more likely to be more than 400 Days. Tree removal, Tree Down car, Stump removal are more likely to be completed around 150 to 400 days. Tree Down house is completed within 50 days.

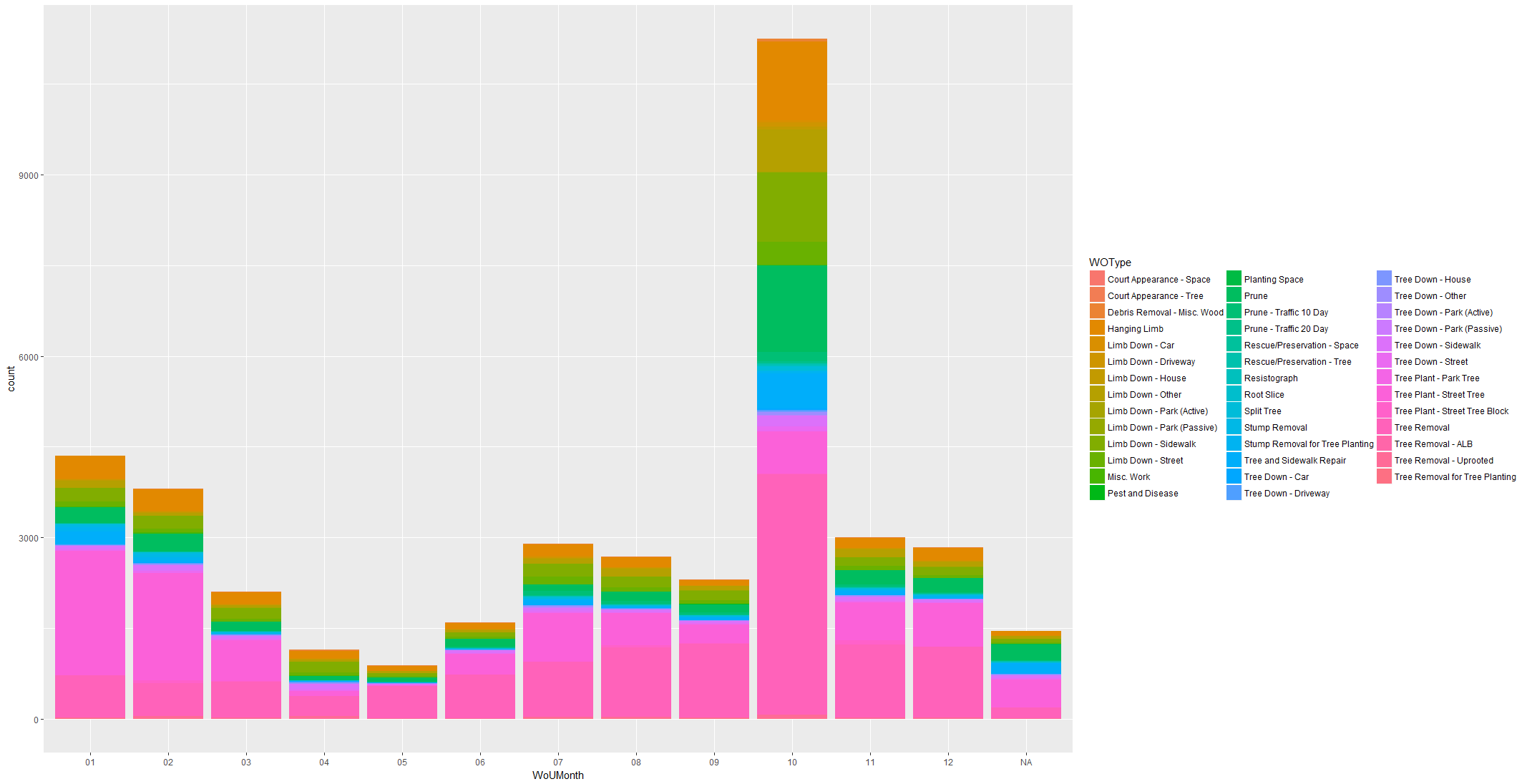


#### Exploring work order creation and Update time

By plotting Workorder creation and workorder update month we can see distribution of work order completion. We can see that most of the work order is created in Month of July as most of the request received from May to July.

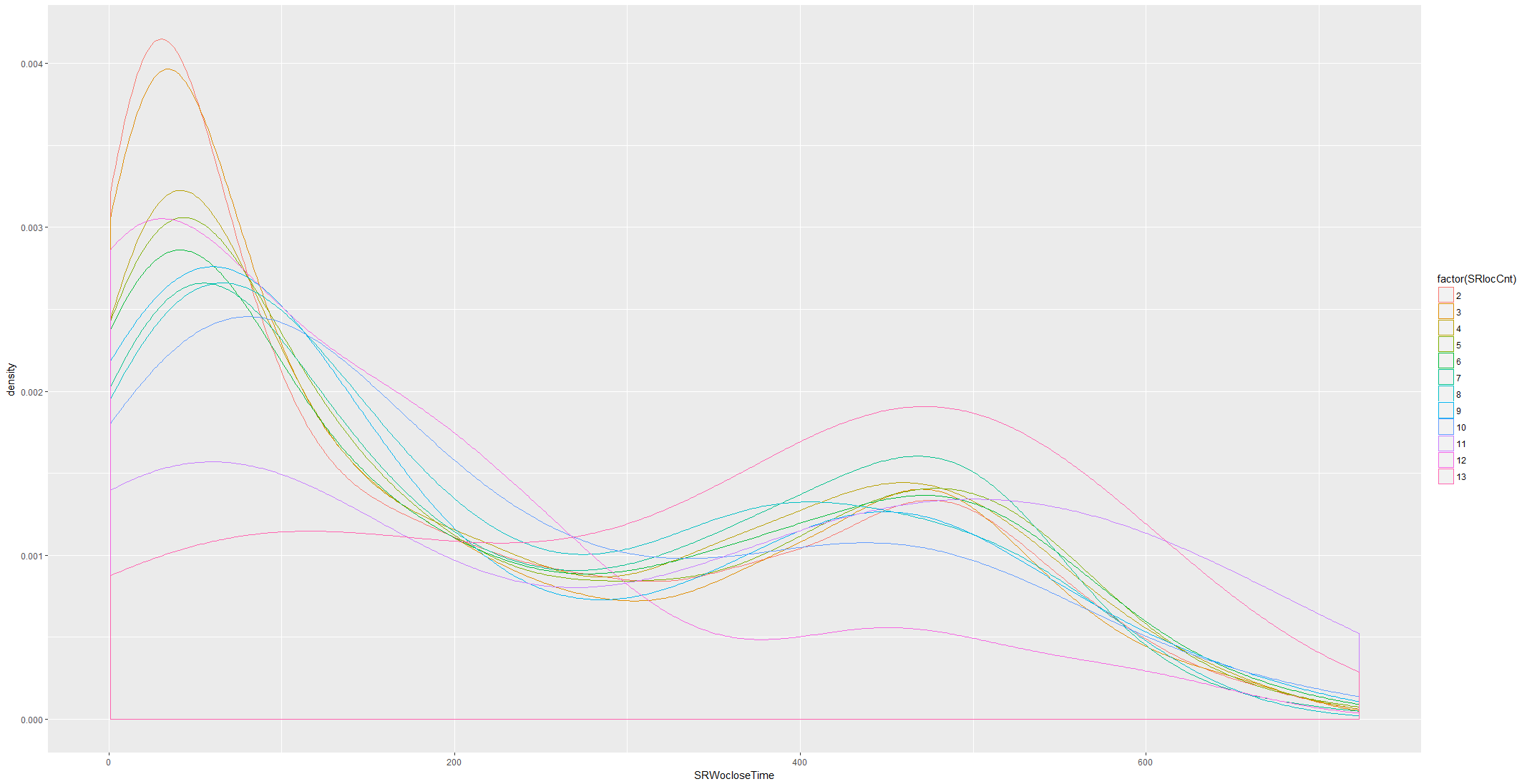


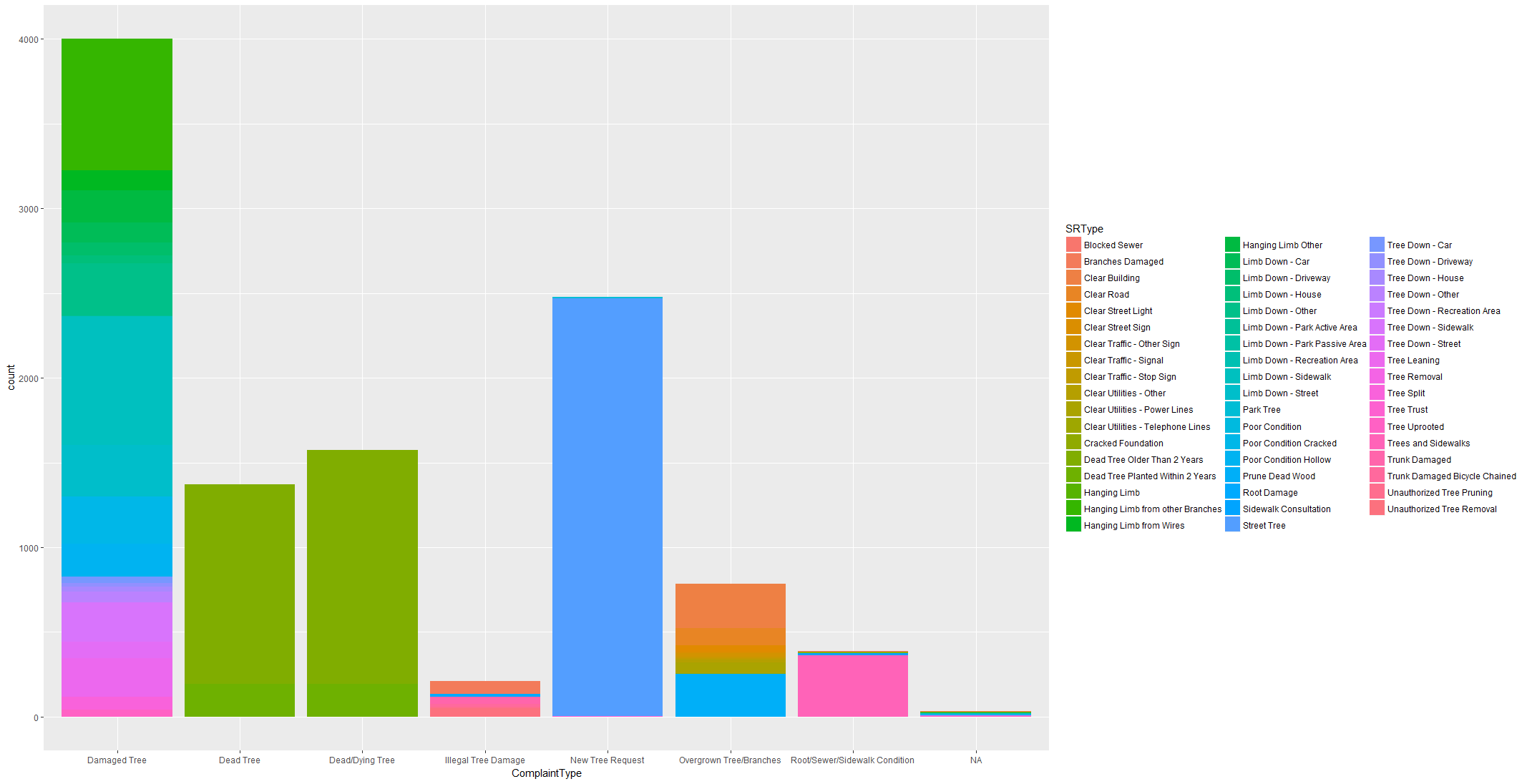
Looking at distribution of workorder we can see that most of the work order is updated in month of October. As we have seen overall completing time is is peeking around 100 days this support the hypothesis that most of the request created before July are getting updated in Oct.



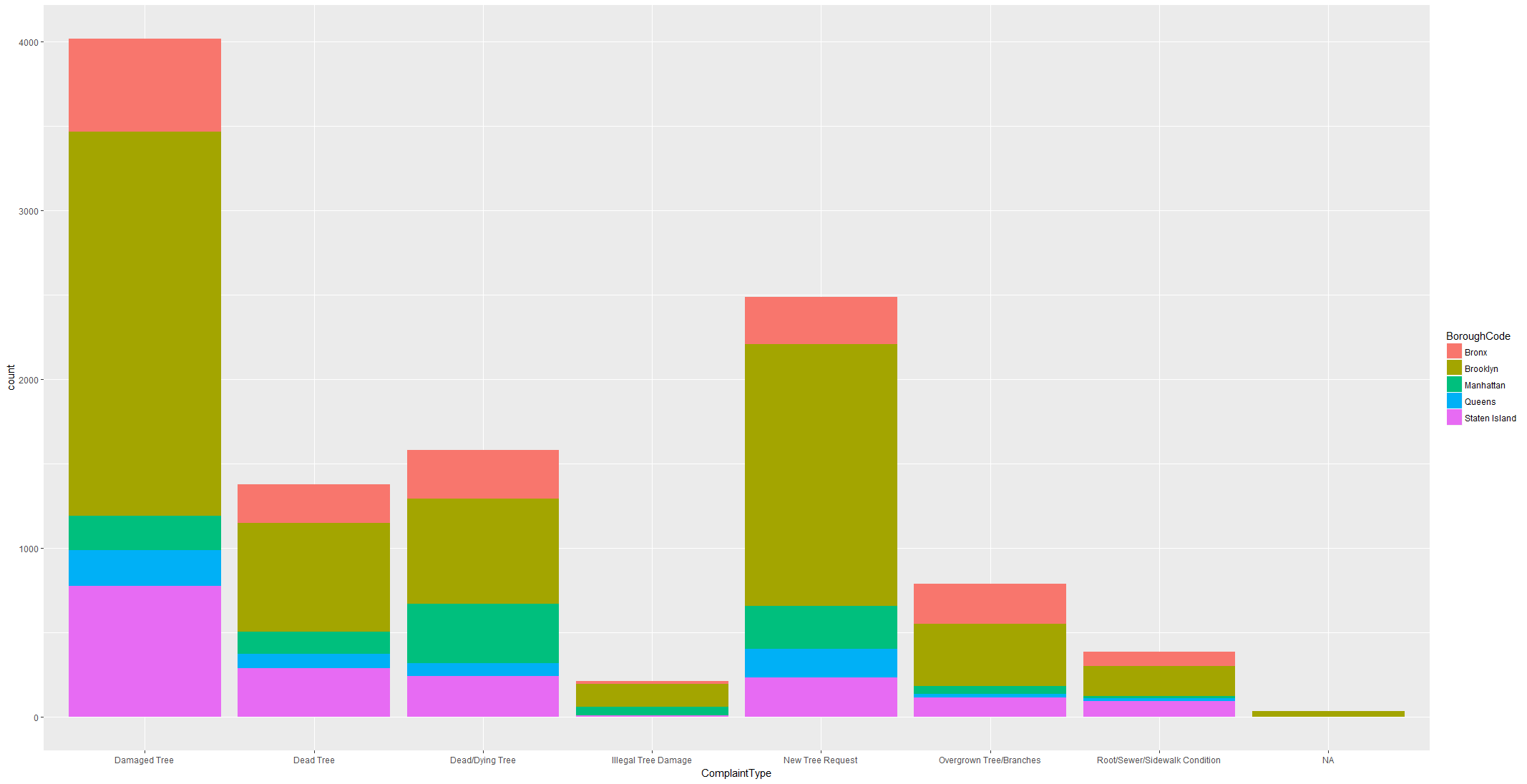
### Impact of Repeating Requests

Repeating Requests are request in same location. Reason might be work not completed on time or it might be urgent which makes individual to raise request again. Plotting SR to WO close time on same location with multiple request we can see that higher repeated calls have higher closure time. It might since service request is taking more time , repeated calls has been made.

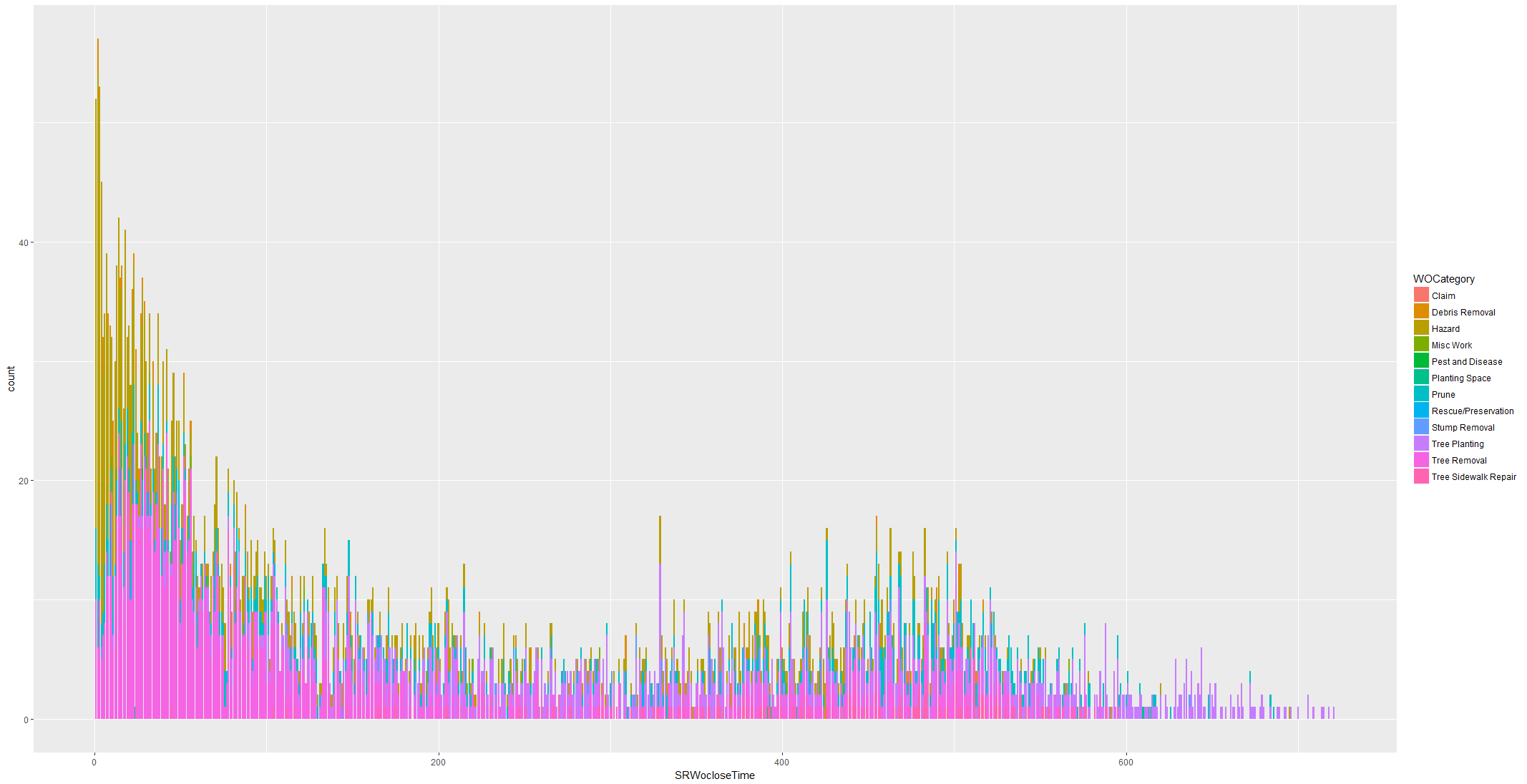


Let’s see what type of complaints, repeated calls are made. We can see most of the repeated calls are made on Damaged Tree with SRtype related to tree removal, limb down,poor condition etc.

Plotting repeated calls by Boroughs we can see that Brooklyn has most repeated calls for all complaint types.

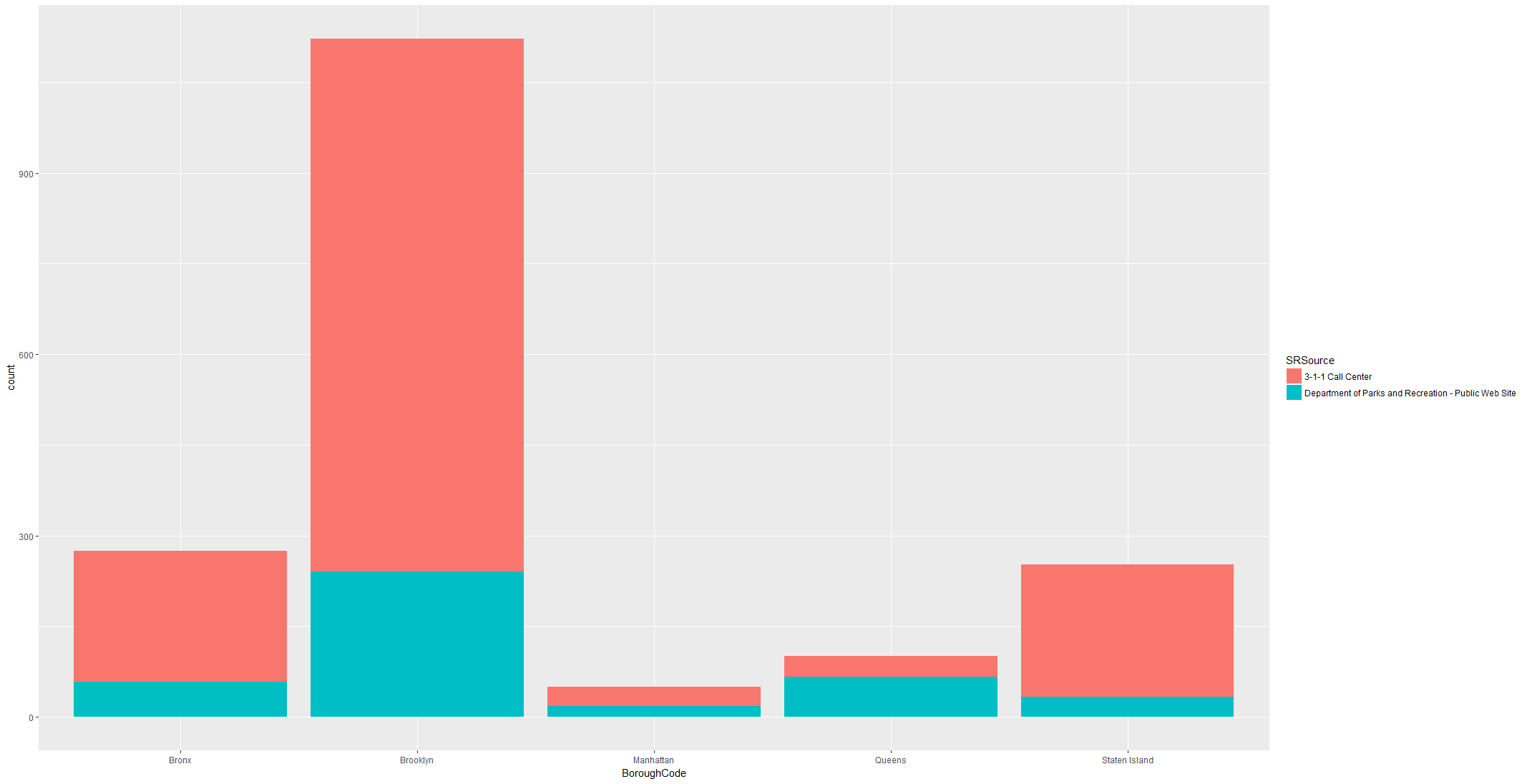


This support overall hypothesis that Brooklyn repeated calls are made and has higher closure time with most of complaints.



Plotting Work Order category on these repeated complaints, it can be inferred that Tree Removal and Tree SideWalk repair as Work Category is likely to be more than 400 days in alignment with overall trend.

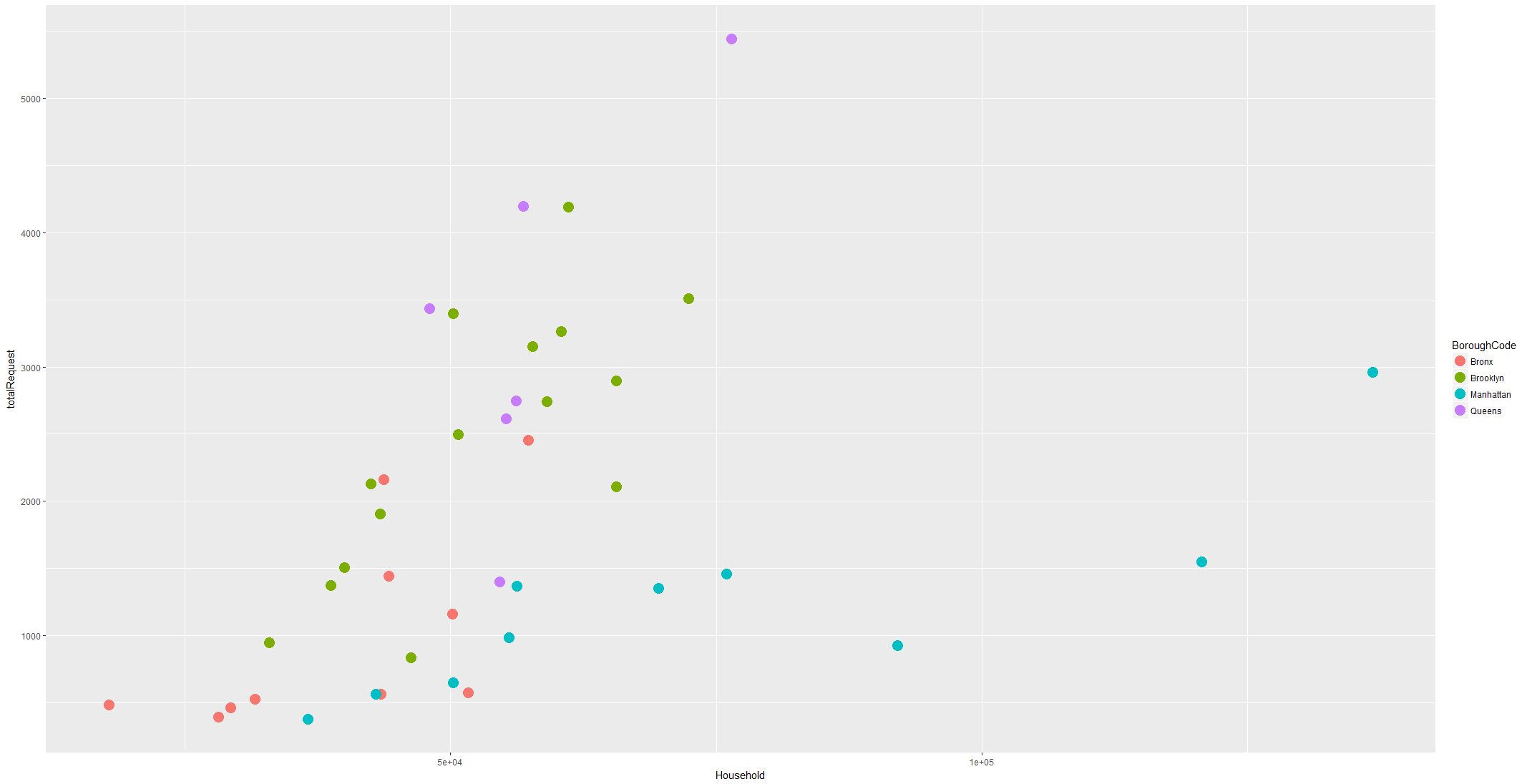
Let check source of repeated calls. We can see that most of these calls from Brooklyn using both main sources. In fact, repeated calls using website is much more than other boroughs.

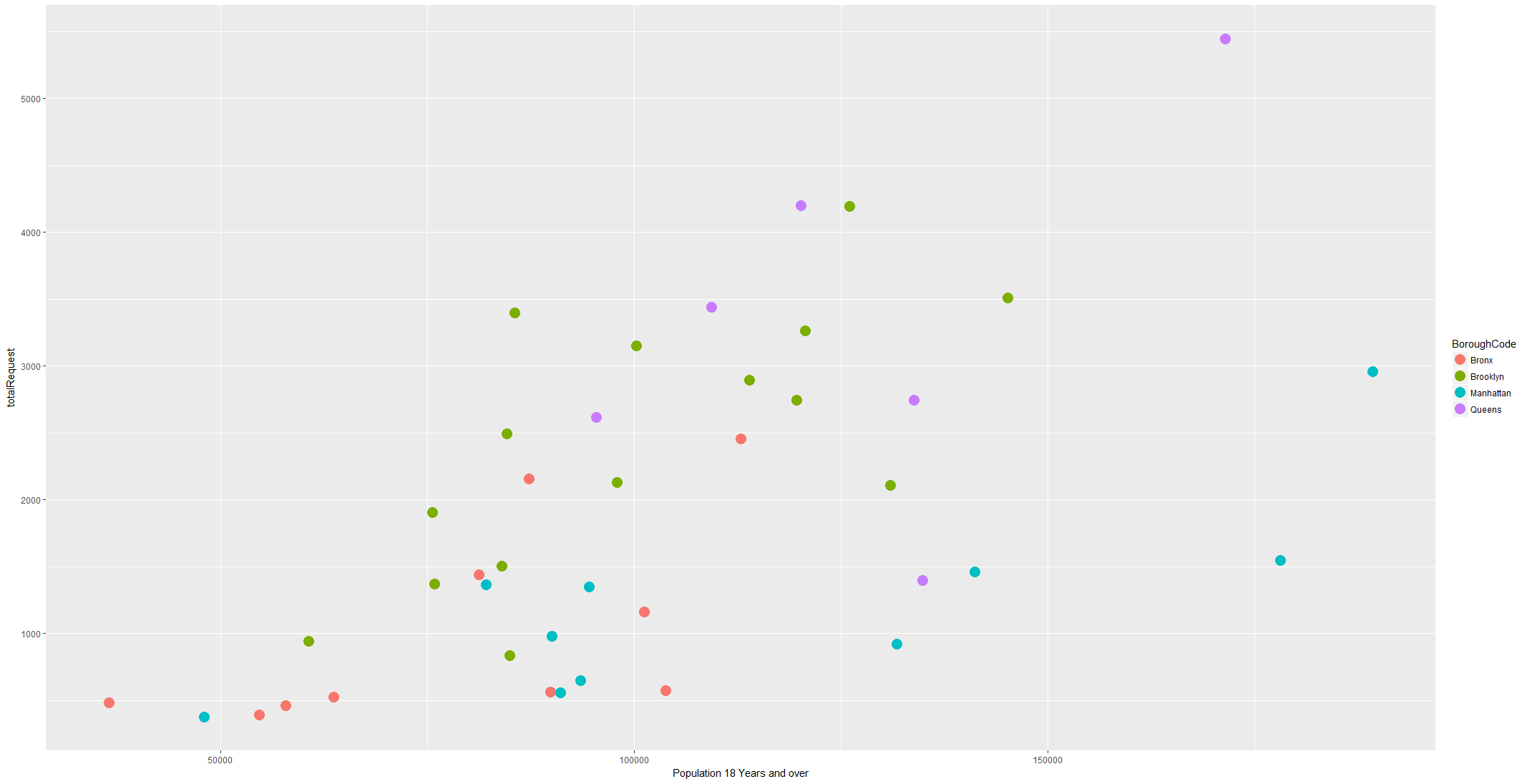


Brooklyn has most repeated call through all sources.

## Community Demographic Analysis

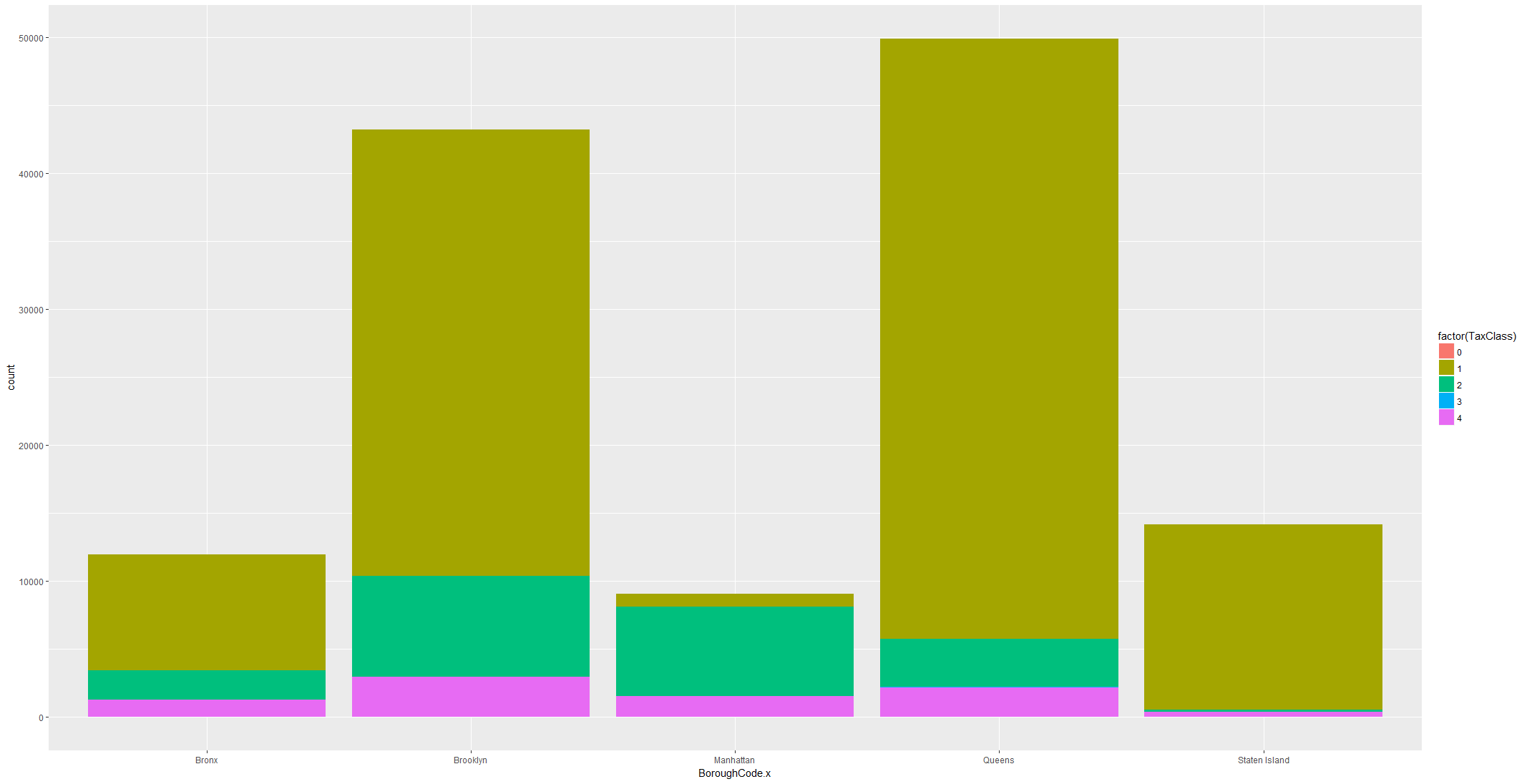
We have demographic data on community district like population, Household, Race , Gender etc. While plotting Response time and Service request to wo closure time , it is observed that it is not related to composition of population rather it depend more on number of household and population.





Though both seems related, it also shows that for each Borough relationship is not same. Like Manhattan population increase doesn’t increase total request while for Brooklyn rate of increase is high. One reason being Individual ownership of houses and spread of population.

This can be validated by Tax class of the area.



Tax Class 1 and Tax Class 2 are residential property while tax class 4 is office rental space.

## Findings

Till now we are able to find out that

Response time on service request is dependent on

* Inspection delay time which is dependent on
* Type of request like tree removal is prioritized over others
* Inspection time is dependent on Inspection structure like full inspection take more time than partial even for level 1 inspection
* Post inspection SR resolution is decided which resolve the request or Work order is created
* For SR resolution information not provided, court appearance and refer to other dept takes more time to close
* If Work order is created WOType , which relates type of work completed and time taken to close the request we can see that Tree Down Sidewalk WOtype is more likely to be more than 400 Days. Tree removal, Tree Down car, Stump removal are more likely to be completed around 150 to 400 days. Tree Down house is completed within 50 days.
* Most probably request is created in May to July while most of the Work order is created in July to September and most of the work order updates are in October.

Even number of request are higher in Queens, Brooklyn has most repeated calls which leads to hypothesis that response time must be high in Brooklyn and validated by high response time in Brooklyn.

High response time in Brooklyn can be attributed to variety of complaints, specifies of trees and variety of work order to be completed.

Brooklyn and Queens has more Tax class 1 and 2 properties

1. Tax class 1 : Three-unit residential properties
2. Tax class 2 : Residential property with more than 3 units including cooperatives & condominium

## Feature engineering

Based on the analysis we already added repeated count on location and repeated count on species which might be impacting response time. We already have other factors like inspection status and work order category which can help in identifying cases which went full cycle of request to work order completion.

Some variables like Inspection Delay Time and Service Request work order Close time has been not considered as independent in model as they are one of the output similar to response time.

Created Year has also not been considered to ensure Year doesn’t impact model.

We will try to use all remaining independent variables to predict response variable and check which variables has higher impact on Response Time.

# Applying Machine learning

The objective of model is to predict Response Time of service requests related to forestry dept. of NYC based on dataset from 2015 till 2017. We already have set of variables which might be statistically significant in predicting Response Time. We are going to use supervised learning on set of variables in training data set build a regression model.

## Building model

Response variable is continuous variable while most of the independent variable are categorical variable. To build regression model we need to dummify categorical variable into continuous variable. Using “Caret” Package all the independent categorical variables are converted into continuous variable at the same time dataset has become wide with 173 variables.

Once all independent variables are dummified, data set is split into 70% training dataset and 30 % test dataset.

Dataset prepared for modeling: fsrforMOD\_df.new

All factor variables have been removed as those are already dummyfied.

### Fitting the model

While fitting the model using linear regression, system is taking lot of time due to wide data set. We are using H2o.ai to run Generalized Regression Model, Random Forest and Gradient Boosting method model.

#### Generalized Linear Modeling using H20.ai

The Generalized Linear model was used to fit the model with family as “Gaussian”. The dataset has been already split into a training (70%) and a test (30%) set, and the model was fit on the training data after normalizing the features. Summary of performance of model on Training and Test Dataset a below:

dataset RMSE MAE R2

1 Training 121.4544 85.53995 0.2852869

2 Test 121.6521 85.76100 0.2813671

R squared value for GLM model is 0.28 which shows only 28% of variability in the data can explained using linear model. Test and Training dataset has similar coefficient showing model is fitting well on test data set as well.

#### Random forest algorithm using H2o.ai

Using Random forest, we found that R-squared value from the fitted random forest regression on the response data is much lower than the GLM—0.09 and also RMSE as well as MAE is also lower than GLM.

dataset RMSE MAE R2

1 Training 136.5609 102.6122 0.09643792

2 Test 136.4425 102.5933 0.09600275

However, the feature importance’s are shown below to give a sense of which variables the model found to have the largest impact on mean decrease node impurity during the training.

variable relative\_importance scaled\_importance percentage

1 SRCategory.Plant.Tree 14940109824 1.0000000 0.10783034

2 SRType.Street.Tree 13009389568 0.8707693 0.09389536

3 WOCategory.Tree.Planting 12017478656 0.8043769 0.08673623

4 BoroughCode.Brooklyn 8447867392 0.5654488 0.06097254

5 SRResolution.Planting.Dec 5510671872 0.3688508 0.03977331

6 Longitude.x 3422561792 0.2290855 0.02470236

Variable importance shows that SR category as Plant Tree, SRtype as Street Tree, WOCategory as Tree pl-anting has most impact on response time. It seems service request Plant Tree has highest impact on res-ponse time.

#### Gradient boosting algorithm using H20.ai

Using GBM algorithm , we found that R-squared value is much higher at 0.60 on training and 0.59 on test dataset.

Dataset RMSE MAE R2

1 Training 108.0567 73.21972 0.4342712

2 Test 108.6959 73.73865 0.4262872

Take a look at variable importance in model using GBM,

variable relative\_importance scaled\_importance percentage

1 SRCategory.Plant.Tree 11124196352 1.0000000 0.22275542

2 SRCategory.Prune 5271790080 0.4739030 0.10556446

3 BoroughCode.Brooklyn 3749020928 0.3370150 0.07507191

4 SRResolution.Work.Completed 3221082624 0.2895564 0.0645002

5 SRResolution.Reviewed...Inspection.Assigned 3106930176 0.2792948 0.06221443

6 BoroughCode.Staten.Island 2283439360 0.2052678 0.04572452

Analyzing GBM model variable importance , similar to Random forest Plant Tree has highest importance While SRcat as Prune and Brooklyn as borough code are next important variable in GBM model. Also, GBM model has vastly improved model R squared and RMSE than GLM or Random Forest.

#### Comparing predicted values with actual test set

actual predict predict.1 predict.2

Min. : 0 Min. :-96.25 Min. : 87.75 Min. :-65.76

1st Qu.: 9 1st Qu.: 79.13 1st Qu.:104.51 1st Qu.: 67.42

Median : 48 Median :112.02 Median :118.45 Median :134.71

Mean :112 Mean :119.81 Mean :137.71 Mean :130.80

3rd Qu.:167 3rd Qu.:164.46 3rd Qu.:182.30 3rd Qu.:162.48

Max. :723 Max. :414.29 Max. :206.41 Max. :630.79

NA's :20296

# Model analysis

We have analyzed NYC service request dataset joined with Inspection, WO and Tree dataset to predict response time based on number of independent variables. While analyzing that we hypothesised that Brooklyn request takes longer time than other Boroughs and delay in inspection time is important factor in determining response or closure of service request. We wanted to understand relationship of other variables like SRTYpe, Complaing type, SR resolution, Inspection type , WO category , Tree parameters as well as location parameters has any impact on Response time or not.

Using GBM model variable importance we can see that

* SR Category Plant tree and Prune tree has more impact on response time. It is also consistent with Random forest model where Plant Tree has most impact on response time. Even though these two categories are lower than Damaged Tree , impact is highest. It might be data has been updated most in these categories.
* BoroughCode Brooklyn has also have more impact on response time than other Borough Code and as we have seen in analysis Brooklyn has most cases with high IQR response time on most of the SR Categories.
* SR resolution Work completed and inspection assigned has more impact than others as these are together higher in count than others.
* We had hypothesis that species and Tree parameters might impact response time. Analyzing variable importance Tree Diameter have little impact “scaled importance” of 0.04 on response time while Tree species doesn’t have much impact at all.

# Conclusion and Future Work

Our analysis of request shows that response to service request related to tree vary a lot from 1 days to 600 days as there are various categories for request. Most of the cases where time period is high due to unavailability of proper information or dependencies on other dept like Sidewalk repair etc.

Recommendations for Forestry department:

* This model can used to set up SLA for type of request and estimated timeline based on dependency.
* Request accurate information to be sufficient to estimate SLA for service request as we have seen Inspection Delay time and Non-availability of information are main cause of delay.
* Learn from the model and provide immediate possibility of solution based in learnings instead of delaying for inspection and then denying the request. While analyzing SR resolution lot of cases are denied after inspection and time.
* Complaint type Street tree and Sidewalk repairs are cases where coordination with other dept. takes lot of time and are mostly delayed. Setup a communication process for joint inspection to reduce delay in work estimation.

For future work:

* Analyze relationship between Tree species and type or count of request raised. While analysis we found that specific species are contributing more on specific service request.
* Request related to Street Tree, Pruning and Sidewalk repair can be shared with Traffic dept to prevent accidents related to these tree requests. One can study traffic accidents related to tree service request.
* Refine this model with accurate data and removing some categories which are not contributing much in model. Some cases update time is lower than created time which is not logically possible.
* While analysis we found that Brooklyn cases has higher range of response time but unable to find any correlation with any parameters. Most probable reason we assumed is variation of cases. We can study further to find root cause of variation in response time.
* Also, we have seen there are repeated calls from same location as well as on same tree point which means either response is delayed or solution is not working. Forestry dept can look into cause of repeated calls to prevent expense of effort on these cases.