Visualization of New York City Street Trees

Dr. Shilpa Balan

CIS 5250 – Visual Analytics

College of Business and Economics

California State University, Los Angeles

By:
Ninelia Talverdi

Fall 2020

Introduction

Trees are an important part of the urban environment, making our cities cleaner and more beautiful places to live and work. Trees can help to block unattractive views of our neighborhoods. Some recent studies also have found a link between presence of street trees and lower stress levels in urban citizens.

New York City is one of the greenest cities in the United States. There are over 5 million "forested natural areas" along with thousands street trees. So, I thought it would be interesting to see how populated the city is by greens by analyzing the street trees in New York City.

So, what makes a street tree different from any other tree? Mainly its location. A street tree is defined as any tree growing within the public right of way or within 15 feet of the curb; not in a park or on a private property. Although they reside in the public right of way (or within the jurisdiction of The Department of Transportation) they are the property of and cared for by the NYC Department of Parks and Recreation.[1]

In this project, I will try to explore and visualize the data for the following questions:

- How many street trees are in New York City? what is their distribution by location or zip code?
- What is the tree distribution by borough? And what is the condition of the street trees in different areas?
- What is the average diameter of trees in each borough?
- What are the most common problems? or How many problems they have?
- What is the relationship between number of problems and average size of tree? or what species have the most problem?

[1] *Kanpetch, Belinda. "Exploratory Visualization of Manhattan Street Trees." NYC Data Science Academy, 1 May 2016, nycdatascience.com/blog/student-works/exploratory-visualization-of-manhattan-street-trees/.*

Data Description

The dataset used for this visualization is “NYC_STREET_TREE_CENSUS_2015” dataset which is collected as part of TreeCount!2015, a street tree census conducted by volunteers and staff organized by NYC Department of Park and Recreation and partner organizations.

The Size of the dataset is 620.8MB. It has 42 columns and 683.8k rows.

The tree data includes information such as tree species, diameter, status, perception of health, specific locations of the trees, problems of trees within all 5 boroughs of New York City, and etc.

The below table describes the dataset in details:

Field Name	Type	Description
created_at	double	The date tree points were collected in the census software.
tree_id	double	Unique identification number for each tree point.
block_id	double	Identifier linking each tree to the block in the block face table/shapefile that it is mapped on.
the_geom	char	Indicate the geometry points.
tree_dbh	double	Diameter of the tree, measured at approximately 54" / 137cm above the ground. Data was collected for both living and dead trees; for stumps, use stump_diam.
stump_diam	double	Diameter of stump measured through the center, rounded to the nearest inch.
curb_loc	char	Location of tree bed in relationship to the curb; trees are either along the curb (OnCurb) or offset from the curb (OffsetFromCurb)
status	char	Indicates whether the tree is alive, standing dead, or a stump.
health	char	Indicates the user's perception of tree health. Good, Fair, Poor
spc_latin	char	Scientific name for species, e.g. "Acer rubrum"
spc_common	char	Common name for species, e.g. "red maple"
steward	char	Indicates the number of unique signs of stewardship observed for this tree. Not recorded for stumps or dead trees. 1or2, 3or4, 4orMore, None

guards	char	Indicates whether a guard is present, and if the user felt it was a helpful or harmful guard. Not recorded for dead trees and stumps.
sidewalk	char	Indicates whether one of the sidewalk flags immediately adjacent to the tree was damaged, cracked, or lifted. Not recorded for dead trees and stumps.
user_type	char	Indicates the category of user who collected this tree point's data.
problems	char	Indicates the trees problems.
root_stone	char	Indicates the presence of a root problem caused by paving stones in tree bed. Yes or No.
root_grate	char	Indicates the presence of a root problem caused by metal grates in tree bed. Yes or No.
root_other	char	Indicates the presence of other root problems. Yes or No.
trnk_wire	char	Indicates the presence of a trunk problem caused by wires or rope wrapped around the trunk. Yes or No.
trnk_light	char	Indicates the presence of a trunk problem caused by lighting installed on the tree. Yes or No.
trnk_other	char	Indicates the presence of other trunk problems. Yes or No.
brnch_light	char	Indicates the presence of a branch problem caused by lights (usually string lights) or wires in the branches. Yes or No.
brnch_shoe	char	Indicates the presence of a branch problem caused by sneakers in the branches. Yes or No.
brnch_other	char	Indicates the presence of other branch problems. Yes or No.
address	char	Nearest estimated address to tree.
zipcode	double	Five-digit zip code in which tree is located.
zip_city	char	City as derived from zip code. This is often (but not always) the same as borough.
cb_num	double	Community board in which tree point is located.
borocode	double	Code for borough in which tree point is located. 1 (Manhattan) 2 (Bronx) 3 (Brooklyn) 4 (Queens) 5 (Staten Island)
boroname	char	Name of borough in which tree point is located. Manhattan, Bronx, Brooklyn, Queens, Staten Island
cncldist	double	NYC Council District in which tree point is located.

st_assem	double	NYC State Assembly District in which tree point is located.
st_senate	double	NYC State Senate District in which tree point is located
nta	char	The NTA Code corresponding to the neighborhood tabulation area from the 2010 US Census that the tree point falls into.
nta_name	char	The NTA name corresponding to the neighborhood tabulation area from the 2010 US Census that the tree point falls into.
boro_ct	double	The boro_ct identifier for the census tract that the tree point falls into.
state	char	All features given value 'New York'.
latitude	double	Latitude of point, in decimal degrees.
longitude	double	Longitude of point, in decimal degrees.
x_sp	double	X coordinate, in state plane. Units are feet.
y_sp	double	Y coordinate, in state plane. Units are feet.

***The highlighted fields in green are used in this project.**

Analysis and Visualizations

Let's take a look at the overall distribution of street trees in New York City before I start the analysis.

A quick summary of the dataset revealed a total of 683,788 trees total in New York City. However, there were lots of missing values for different fields such as unidentifiable species type, health, problems, etc. So, I decided to use filtering options to exclude missing data in my analysis. Also, I noticed 174,263 data was gathered in 2016 so I created a data source filter to only include the data that was created in 2015.

1. What is the tree distribution by zip code?

The above map shows how street trees are located in different parts of New York city based on the zip code.

The highest percent of trees (4.11% - 20,939 in total) is distributed in an area with “10312” zip code which is located in “Staten Island” borough. However, this doesn’t mean that “Staten Island” has the highest number of tress among the 5 boroughs as we will see in the next graph.

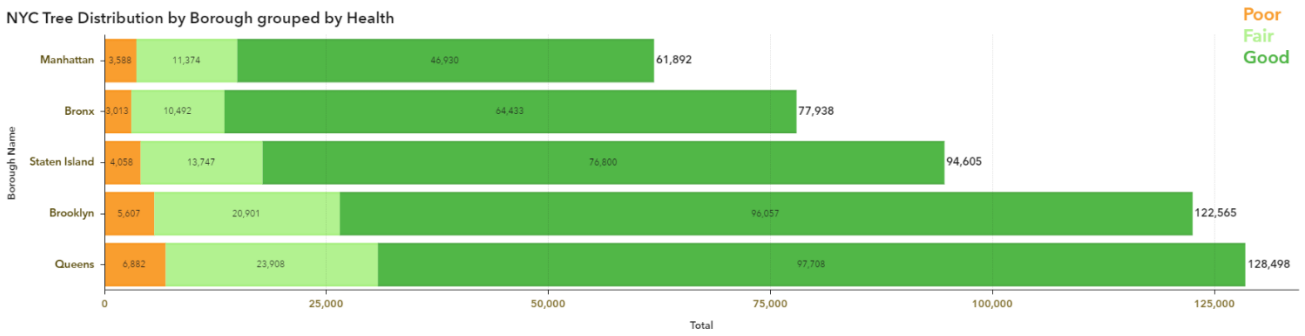
On the other hand, the lowest percent of trees (almost 0% - 7 in total) is distributed in an area with “10115” zip code which is located in Manhattan borough.

SAS features:

- Graph: **Geo Map**

- Since I wanted to use a Geo Map for this visualization, I created a **“Geography” data item** based on the “zipcode” field and map the values according to US zip codes.
- I also pinned the 2 areas to show which location has the highest percent of trees and lowest percent of trees on the map.

2. What is the tree distribution by borough? And what is the condition of the street trees in different areas?



The above chart shows the distribution of the trees with their health condition across the five boroughs. Based on the chart, Queens has the largest amount of street trees with 128,498 trees among the 5 boroughs. Brooklyn has the second greatest number of trees and Staten Island comes in at third place. One of the reasons that Queens has more trees is that it has a lot more landmass, so there's more room to plant trees.

On the other hand, Manhattan has the lowest number of trees with 61,892 trees which is almost half of the number of trees in Queens. I assume one of the reasons of having low number of tress can be because of the large buildings and narrow streets that exist in Manhattan.

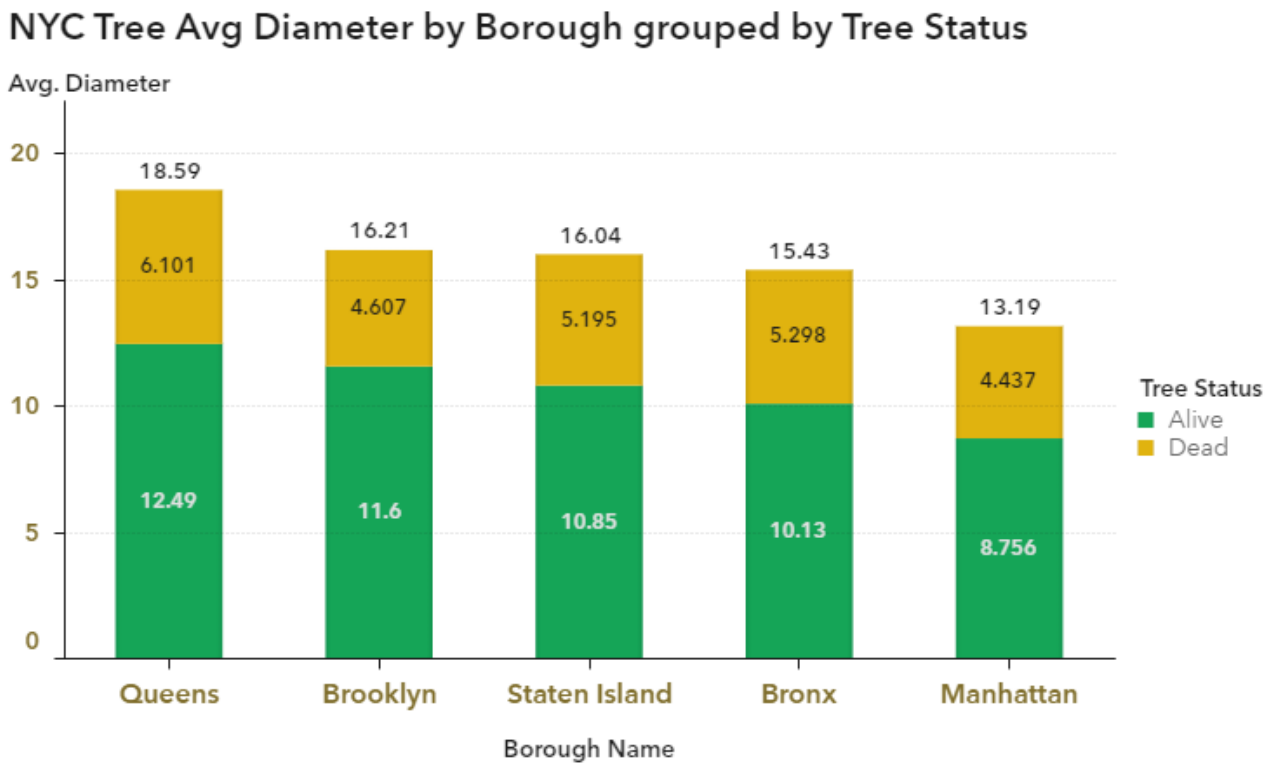
In general, just over 90% of the trees were rated in fair to good condition. As you can see, the total count of trees in Good condition is high in Queens and Brooklyn whereas Bronx and Manhattan have lower values. The same statement is true when making the comparisons for Fair and Poor conditions. The total count of trees in Fair or Poor condition is high in Queens and Brooklyn but it is low in Staten Island, Manhattan and Bronx.

SAS features:

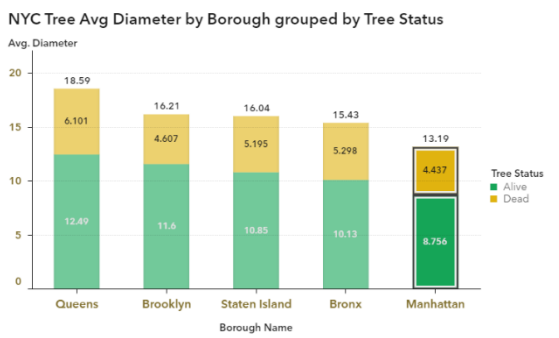
- Graph: **Bar Chart**
- **Horizontal Stacked Bar Chart** is used for this visualization since I wanted to show the **health of trees and the relationship of each part with the total amount.**

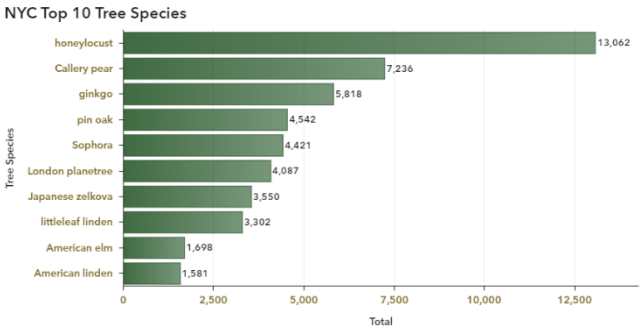
- The health condition of almost 24k of trees were missing from the data set so I used a **filter option** to **exclude** the **missing values**.

3. What is the average diameter of trees in each borough?



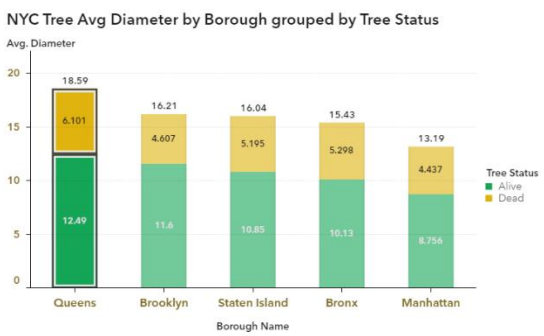
The above chart shows the average diameter of the trees by their status across the five boroughs. According to the average size of trees by borough, Manhattan has the smallest trees with an average size of 13.19 inch. The reason why the average tree size in Manhattan is small is because there are many *Honeylocusts*, which are relatively small trees.

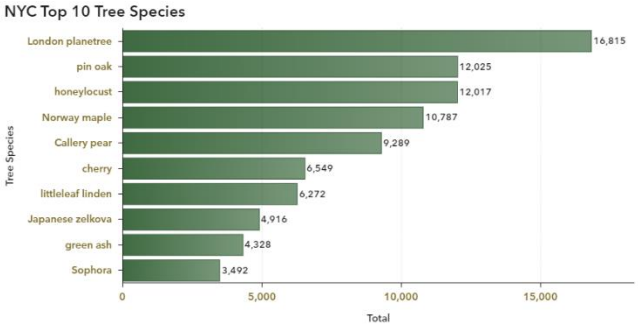




I think maybe they planted a lot of trees that weren't too big because Manhattan is considered downtown in New York City.

On the other hand, the Queens has a relatively large number of *London planetrees*, so the average size of the trees is large with an average size of 18.59 inch.

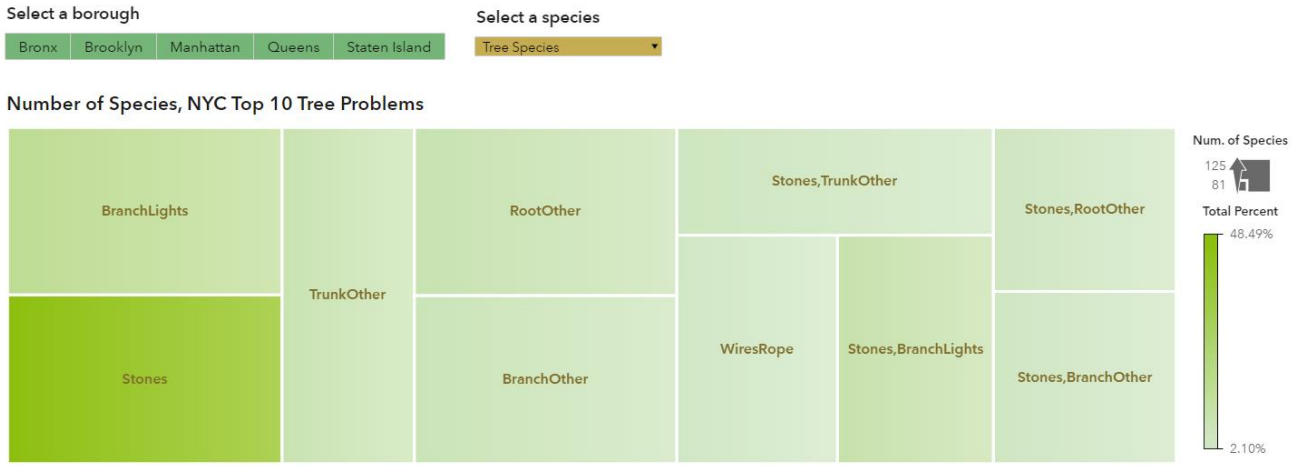




SAS features:

- Graph: **Bar Chart**
- **Vertical Stacked Bar Chart** is used for this visualization since I wanted to see if there is a difference in the average size if the tree is alive or dead.
- The original dataset has a field called "tree_dbh" that contains the diameter of the tree. **In order to get the average diameter, I created a new data item as "Avg. Diameter" in the "Measure" group and used "Average" as the aggregation.**
- **I grouped by "Tree Status" to field since I wanted to see if the dead trees affect the size of the trees.**
- The status of almost 12k of trees were missing from the data set so I used a **filter option** to exclude the **missing values**.
- For additional charts, I used the **rank option** to **show the top 10 tree species** and I used a **filter option** to **exclude** the **missing values for species.**
- I used an **"Image" object** to show the **picture of the top species**.

4. What are the most common problems?



The above chart shows the top 10 tree problem and the number of species that has the problem.

A quick review of the dataset reveals that there are 167,823 problems. As you can see, most of the top problems are related to "Stones". In fact, Census-takers identified "Stones" as the greatest problem (with 48.49%) that street trees face across all 5 boroughs and 125 number of tree species have this problem. The second highest problem is related to "BranchLights" with 14.49% that 124 number of tree species faces this issue. I guess, this issue is mostly because of different decorations that are done in New York City.

When you filter the map by borough, you can find which issue is the highest concern in that borough. As expected, "Stones" is the highest problem in each borough.

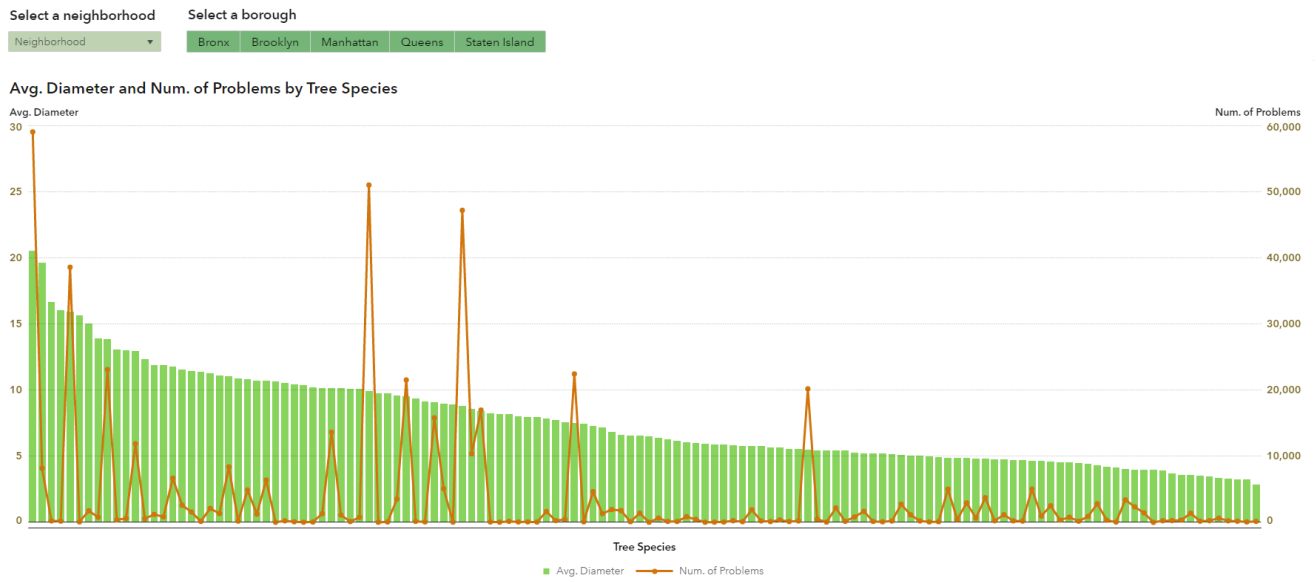
SAS features:

- Graph: **Treemap**
- Probably "Treemap" graph was not the best type to use to show the top 10 problems However, I wanted to explore a different graph, so I decided to use the "Treemap" graph.
- The original dataset has a field called "spc_common" that contains the name of tree species. **In order to get the number of species, I created a new data item as "Num. of Species" in the**

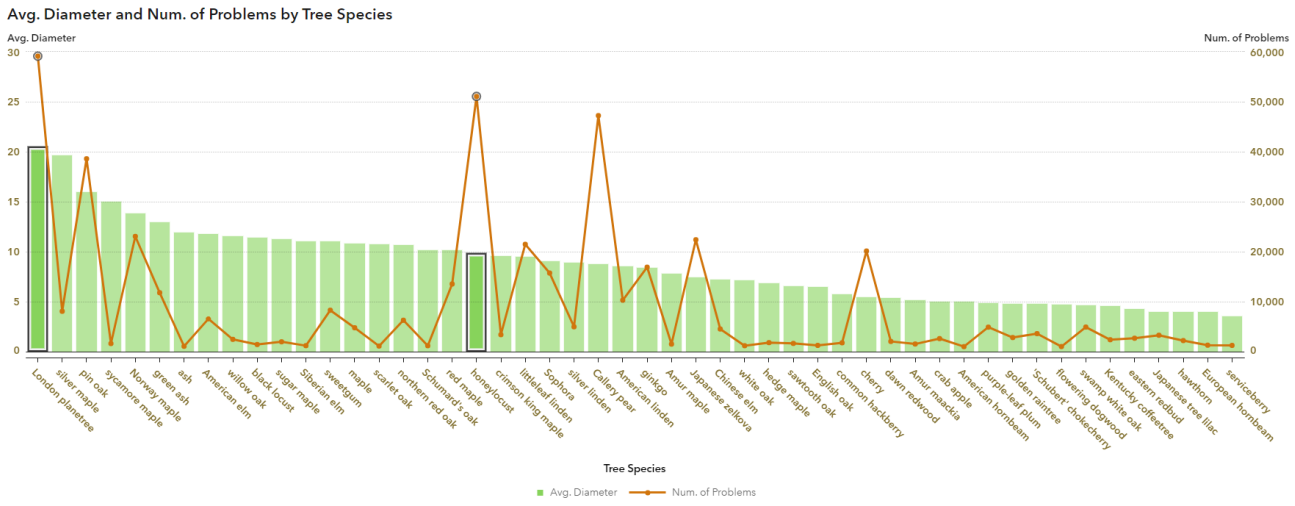
“Measure” group and used “Distinct count” as the aggregation. (*Distinct [_ByGroup_] ('Tree Species'n))*.

- I used the **rank option** to **show the top 10 problems** and I used a **filter option** to **exclude** the **“None” values for problems** since the tress with “Dead” status do not have any values for problems.
- I added a **“Button Bar” control** to be able to see the problems information for each borough.
- I added a **“Drop-Down List” control** to be able to see the problem details for each species.

5. What is the relationship between number of problems and average size of tree?



I wanted to see if there is any relationship between the number of problems and diameter of tree so I tried to compare the average diameter and number of problems in the above chart. Based on the graph, the tree with the largest average size (20.56 inch) which is *London planetree*, has the highest number of problems (59,123 in total). However, the second tree with the highest number of problems (51,081 in total) is “Honeylocust” which is a small tree with an average size of 9.95 inch. So, I figure out trees can have different number of problems regardless of their size.

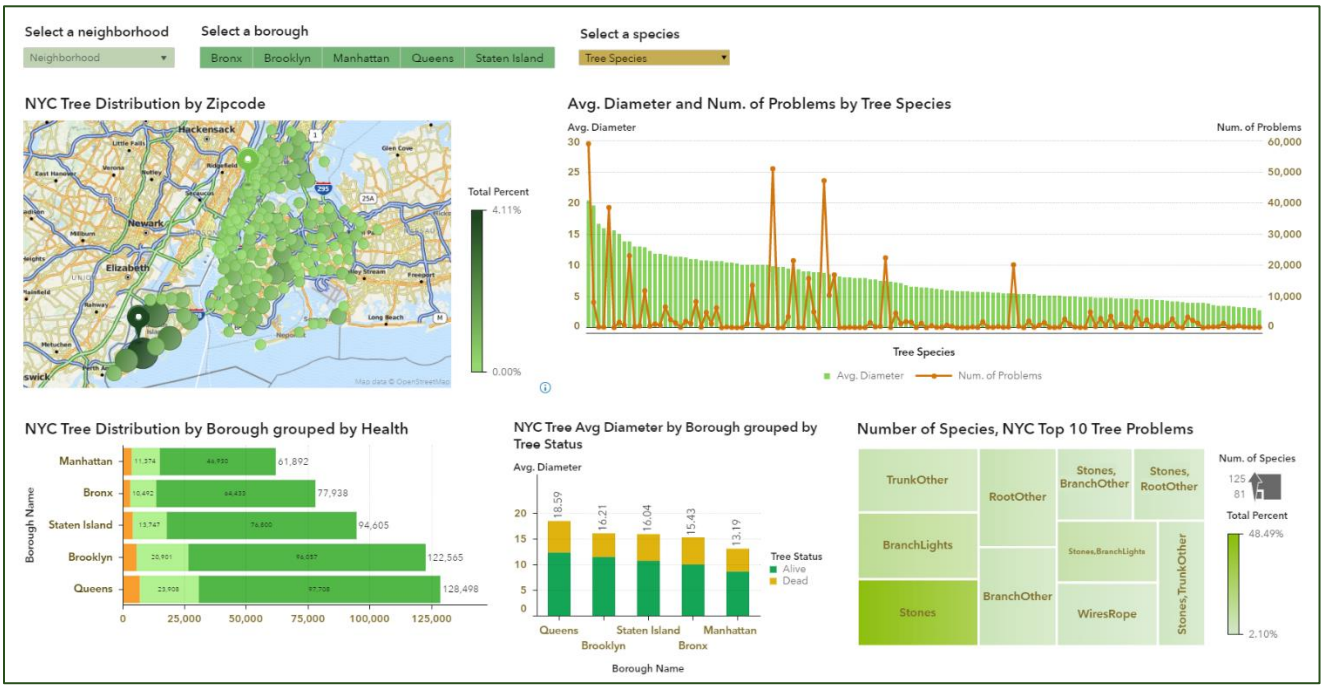


SAS features for this graph:

- Graph: **Dual Axis bar-line chart**
- I wanted to compare two data series ("Avg. of diameter" and "Num. of Problems") in one graph, so I used the "Dual Axis chart" type.
- The original dataset has a field called "tree_dbh" that contains the diameter of the tree. **In order to get the average diameter, I created a new data item as "Avg. Diameter" in the "Measure" group and used "Average" as the aggregation.**
- The original dataset has a field called "problems" that contains the name of tree species. **In order to get the number of problems, I created a new data item as "Num. of Problems" in the "Measure" group and used "Count" as the aggregation.** (*Count [_ByGroup_] ('problems'n[Raw])*).
- I used a **filter option** to **exclude** the **"None" values for problems** since the tress with "Dead" status do not have any values for problems.

Dashboard and Story Telling

One of the many names that New York City is famously known by is "Concrete Jungle". Since 1995, New Yorkers have joined together to count their trees by creating baseline inventory of the City's street trees. Every ten years, NYC Parks has worked with volunteers to record the location, size, species, and condition of all public curbside trees. TreesCount! 2015 is the third citizen participatory inventory of street trees in New York City[2] that I decided to analyze, to gain a more quantitative understanding of how populated the city is by greens. The below dashboard includes my findings. Let's take a look at it:



A glance at the geomap revealed that the area with the highest percentage of trees is located in Staten Island. However, when I analyzed the tree distribution by borough, I found out that Queens is the borough with the most and healthiest trees and Manhattan has the lowest number of trees. Next, I tried

[2] *nycgovparks. "2015 Street Tree Census Report." NYC Street Tree Census - 2015 Report, 2017, media.nycgovparks.org/images/web/TreesCount/Index.html.*

to analyze the size of trees in different boroughs. I got the similar results as the previous graph. The average size of the trees is large in Queens because it has a relatively large number of *London planetrees* and Manhattan had the smallest trees because there are many *Honeylocusts*, which are relatively small trees. During my research, I found out that the largest tree in New York City can also be found in Queens. It's an 87-inch diameter *pin oak* located on Douglaston Parkway and Barrows Court in Little Neck. It removes approximately 6 pounds of air pollutants each year, as well as reducing carbon dioxide levels by approximately 17,746 pounds.[3]

Since there were lots of different problems, I tried to explore the most common ones that trees face during their life. Based on my analysis, the most common issues were related to stones and branch lights. Moreover, I found out that trees can have different number of problems regardless of their size.

During my research, I found out that there is a tree service where people can report any issues related to street trees by submitting a request to NYC forestry services.[4]

Lastly, it's worth to mention that street trees provide multiple benefits that improve quality of life in New York City. They purify and cool the air, reduce stormwater runoff, and conserve energy. They also increase property values, beautify neighborhoods, and improve human health and well-being. Therefore, we must respect and protect them for the future.

---

[3] *Sperling, Jonathan. "Five Key Queens Tree Facts To Help You Breathe Easy." Queens Daily Eagle, Queens Daily Eagle, 1 Mar. 2019, queenseagle.com/all/2019/3/1/five-key-queens-tree-facts-to-help-you-breathe-easy.*

[4] *"Street Tree Planting." Street Tree Planting : NYC Parks, 0AD, www.nycgovparks.org/trees/street-tree-planting.*