**TBLAND 510 A Business Analytics**

**Fall Quarter**

**Enrique Otanez**

**12/09/20**

**510 Final Project**

**City Maintenance**

**The purchasing and planting of trees in NYC**



I certify that I have completed this assignment within the Academic Integrity guidelines presented in the UW General Catalog. Further, I certify that I do not have any knowledge of any other individual(s) violating these guidelines.

Contents

[Executive Summary 3](#_Toc58674831)

[Background 4](#_Toc58674832)

[Learning objectives 4](#_Toc58674833)

[Methods & Description of the Data 4](#_Toc58674834)

[Data Transformation 5](#_Toc58674835)

[The Dashboard 6](#_Toc58674836)

[The Storyboard 19](#_Toc58674837)

[Results/Insights/Conclusions 34](#_Toc58674838)

[Limitations 35](#_Toc58674839)

[References 36](#_Toc58674840)

# Executive Summary

The background of this project is one of altruism, the betterment of a single neighborhood in New York City (NYC). A single report spurred the governor of New York to act on one small neighborhood, Hunt’s Point. Located in in the south Bronx borough the report claimed that Hunt’s point was the worst neighborhood in NYC for 2020 (Sparks, Sam), and this spurred the governor into action, by having others do the work. That is where we come in, the newly formed data analytics team of the NYC parks and recreation governmental department. Waiting for an opportunity to fall in our laps to show the public that their taxpayer dollars are not going to us to buy video games, booze, and other slightly questionable things for nothing, we eagerly accepted the governor’s request to try and improve the neighborhood of Hunt’s point through data analytics. We saw a narrative, learning objectives and opportunities with the dataset that we chose, the problem and solution we saw through this dataset, trees. The NYC parks and recreation department has been collecting data on all the trees in NYC for many years updated annually via community volunteers, our own staff, and partnered organization, this was the perfect opportunity to show the metal that NYC parks was made of.

The mode of collecting the data was already in place and the data could not be more inclusive. Many variables show the health of the tree, what was damaging it, location of the tree and other important variables like mapping data longitude and latitude. It was perfect, almost to perfect. After analyzing it we realized something we should have from the start, but it was far too late, many of the important variables that could end up being our chosen KPI’s were string values. This set us back a few weeks as we had to do ETL practices on the data, creating calculated fields and creating meaningful measures like percentages for visualization. It was an endeavor, but we were finally able to run analysis on our business problem: What tree or trees is best suitable for replacement use for all the stumps and dead trees in Hunt’s Point? So glorious, this opportunity was, to show our skills at replacing something that symbolizes death with brand new life. Needless to say, the parks and recreation department were thrilled. But what can be learned from this analysis?

The reader can learn how we created the calculated fields, prescriptive analytics via improvements of trees via KPI’s, make inferences on the results of the project and other meaningful insights through the analysis process. Tutorials are provided to get the user started and learn some of what Tableau has to offer via sheets for analysis, dashboards for painting a quick picture of the business problem, and storyboards for understanding how the project was conducted decision to decision. The Tableau tutorials also cover many features to customize your visualizations and to better the analysis like filters, shading, coloring and more.

The results based off this analysis could be better, as further analysis should take place before implementation of the suggested trees for replacement use. The overall feeling inside the office, wonderful. The sophora, ginkgo, eastern redbud, and callery pear were chosen via the analysis as the best trees for replacement use so far, and it is a wonderful thing the ginkgo made the list, its an office favorite. These were based off improvements we noticed of trees of each variable comparing the trees in its sub-variables and how they improved by percentage and list movement. The results did show some more abundant trees as good candidates but that is why we want to investigate the analysis further with other KPI’s and testing existing KPI’s against each other in the future.

# Background

We, the New York City (NYC) parks and recreation government department have updated our staff and processes to include the most recent developments in the business world; data analytics. This was due to legislation change and the priorities of New York’s existing governor. After reading an article called “The 10 Worst Neighborhoods in New York City for 2020” by Sam Sparkes of RoadSnacks back in January(Sparks, Sam), the governor decided to take action. Thus, the data analytics division of the parks and recreation of NYC was born.

# Learning objectives

* Learning Objective 1: Make string values more meaningful and workable by changing them into percentages.
* Learning Objective 2: Understand what it means to compare improvement, by percentage and by moving up in the list.
* Learning Objective 3: Learn how to determine the best tree or trees for replacement use.

# Methods & Description of the Data

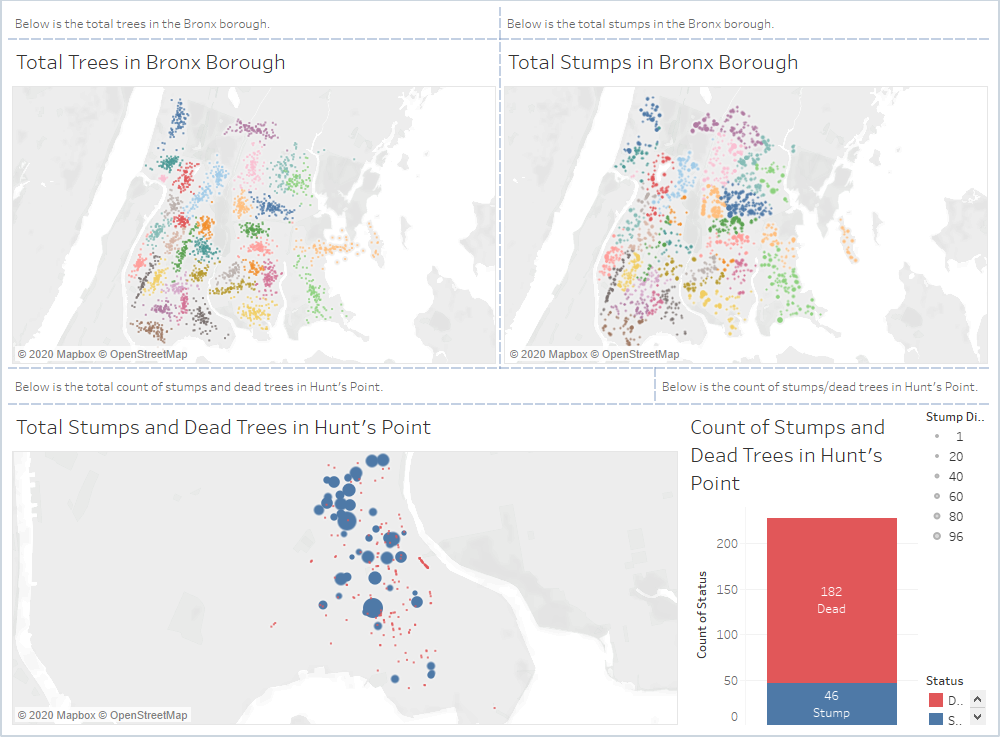
The dataset is about all the trees in NYC, but specifically all the trees on the streets and sidewalks of NYC, not the trees in the established parks. It is a very accurate representation of the city trees of NYC as there are three ways NYC parks and recreation were able to receive the data, community volunteer, NYC parks and recreation staff, and partner organizations, yet there are also over 600,000 records inside this dataset. In references citation number 2 is a link to the dataset found on Kaggle.com. Downloading the data and unzipping it to reveal and CSV that then needed to be converted to a XLSX file. There are noticeable categories of variables inside the dataset, general information about the trees, general information about the data recorded, specific generalized information about the trees, Boolean variables of problems to the trees, general location data, and specific mapping data. How these categories can be used, general information shows variables like tree diameter and stump diameter, which can help identify large stumps or trees to either be taken down or left standing. General information about the data recorded can help show the block id as well as finding out accuracy based on what user (volunteer, NYC parks staff, or partner organization) recorded the record. Specific information about the tree’s category show a lot of the status of the tree, health and alive vs dead and stump, mapping this could show popularity of trees in different areas of the neighborhood of Hunt’s Point. Depending on analysis, the Boolean variables of problems could be a main point of learning objectives, as finding out the most consistently damaged trees could be insightful. General location data can help narrow down inside boroughs and neighborhoods where they can be broken down further. Specific mapping data is very important as to no matter how the data is to be analyzed, with the mapping data provided we can map all the trees anywhere in NYC, down to the street address. We used a KPI from almost every one of those categories, or at least used a variable from each category as filters and more. The KPI’s we used for this first iteration of analysis are steward, curb\_loc, guards, health, and sidewalk. Steward is the amount of recorded stewardship that was shown to a tree. That can be things like helpful tree guards, mulch or woodchips, intentionally planted flowers, tree bed maintenance and more (see data dictionary provided with dataset). As said before, this data is collected via volunteer, NYC parks and recreation staff, and partner organization but all this information is gathered from the tree census and put onto Kaggle.com. “The city has an open data platform, and they update their information according to the amount of data that is brought in” (City of New York, Data). This data is updated to 2019 and the data is updated annually.

# Data Transformation

The dataset used was incredibly clean. There were no missing values among the 600,000 records or if there were any, they would be completely negligible. What nulls that did exist was when a tree was dead or a stump, the data gatherers did not record anything else about the tree. So dead trees and stumps would not have data on health status, stewardship, even the name of the tree for dead trees are not included in the dataset. The biggest problem when the data was first put into Tableau was that because most of the variables were string values, many meaningful visualizations could not be generated, and the best you could get was comparing values against itself for a count of trees pertaining to that variable. Mapping data did work, and some of the first visuals were made via plotting trees in NYC. Even the Boolean values of yes and no had to be converted to true or false for Tableau. So, with that in mind the first thing done was divide each applicable variable into numbers. Health for example has three statuses used to describe the condition of the tree which are good, fair, and poor. To split health between its statuses calculated fields were made using this code in Tableau; if(([Health])='Poor') then 1 else 0 END, for each of the health statuses all with a common nomenclature of PoorHealth for example. At first, all the variables underwent this transformation into a numeric data type because the KPI’s were not decided upon. Then each one underwent another calculated field to find the percentage of that variable and this was useful because it could then be found for instance, what percent of good health trees can be found on the streets in NYC. For this, each of the new fields was used for another calculated fields with the common nomenclature of %GoodHealth for example with this Tableau code: sum([GoodHealth]) / total(count([Health])). This was believed to be good enough until the percentages did not represent the total amount of trees displayed in the visuals. At first it was thought that the percentages still represented the entirety of NYC even though a filter was put in place to find the percentage of trees via variable Goodhealth in the Bronx borough. But was discovered was that the percentage was still including all the other health statuses and according to the first equation, it was counting 1 and 0. Thus another calculated field was then made with the common nomenclature of JustFairHealth which used the Tableau code { EXCLUDE [FairHealth]: count(0)}. This way Tableau recognizes it as a single measure and now the final percentages can be made from this most recent calculated field. Using JustFairHealth as an example, the common nomenclature for the final calculated fields would look like %JustFairHealth using this Tableau code: sum([JustFairHealth]) / total(count([Health])). With these final percentages recognizing only the number of trees that truly do have fair health in whatever area you are filtering for, for instance the Hunt’s Point neighborhood in south Bronx, accurate and eye pleasing visualizations could be made. Below are the steps done to get the data to work in Tableau for clarity.

1. Divide the subcategories health using this code if(([Health])='Poor') then 1 else 0 END.
2. Find the percentage of each new fields using this code: sum([GoodHealth]) / total(count([Health])) for each new field.
3. Excluding the number of trees in the field that was not desired with this code: { EXCLUDE [FairHealth]: count(0)}.
4. Finding the percentages of the new and final fields using this code: sum([JustFairHealth]) / total(count([Health])).

# The Dashboard



<https://public.tableau.com/profile/enrique.otanez#!/vizhome/510FinalAssignmentTableauProjectFile/Initial?publish=yes>

A. Simple Business Scenario

Sam Sparks of RoadSnacks claimed that of the overall “snackability” (how different variables compare to each other) via census data like income, crime, home prices, etc.; Sam was able to figure out the worst neighborhood in 2020, which is Hunt’s Point in the south Bronx. The governor with this information tasked me, the head of the NYC parks and recreation government department, and my team of newly hired data analytics employees with how to improve that area. Luckily for my team, eager to get their hands dirty with data, we have a community volunteered, NYC parks and recreation, and partnered organization gathered dataset of all the trees in NYC planted among the streets. After deliberating with the team, we have come to a chosen method of how to improve the area via our resources and expertise, using the tree dataset to find the best tree or trees to use for replacing all the stumps and dead trees existing in Hunt’s Point. Hopefully, this will be a good start to increasing property value, attract new and retain existing businesses, and to help citizens feel safter and prouder of the Hunt’s Point neighborhood, and more. Hunt’s Point is only the first area we can look at with this dataset, but first we need to get our findings and implement them, as to see the success of this upcoming change.

A. Purpose/Questions/Applications

The purpose and overarching goal of this project is to better what is the worst neighborhood in NYC as of 2020 (Sparks, Sam). To better understand how we can improve the neighborhood with something like tree data, the problem that can be answered needed to be found first within the dataset, and that most pertinent problem looks to be stumps and dead trees. The dataset contains much more information other than the KPI’s chosen for this project, stewardship, tree curb location, guards on trees, health of trees and sidewalk damage caused by trees. What was immediately at our attention as our team deliberated was that although there could be damage shown on tree trunks, branches, and roots, choosing a tree that was familiar and liked by the citizens of Hunt’s Point was pertinent. The goal to better the neighborhood also aligned with what people want to see as they leave their doorstep every morning for work, and a better liked tree can satisfy the human emotion of feeling safer and prouder of their neighborhood. Although further analysis of damage to the tree itself is important, the KPI’s selected were much more pertinent in our deliberations because they would consider the overall conditions of the trees as well as the consequences of us putting the trees into the ground on city sidewalks.

A. A sample list of decisions to be made is provided below to understand how we tackled this project.

* Show analysis via percentage of trees in each variable in Hunt’s Point.
  + The reason percentage is used as the main basis of understanding is because via percentage, we can get a single more manageable measure of the number of trees in a specific variable in Hunt’s Point. If to say we used count, then a large portion of trees may seem significant when there is large variation but using percentage, we can understand how a single tree type is compared to another variable or sub variable.
* Exclusion of the top five trees in Hunt’s Point.
  + The top five trees are so numerous that many of those trees would control every visualization. This would also cause our selected trees for replacement use as one of the top five trees in Hunt’s Point as well as all of NYC. It does not give a good representation of what trees could be used for replacement use just because the top five would be the highest in consistency of each variable and sub-variable.
* After removal of the top five trees in Hunt’s Point, then select the top ten trees for each variable.
  + The reason why after the top five trees were removed to select then the next top ten trees in each variable was a simple answer. Although some of the top ten shared the same problem with the top five, appearing in each visualization, the next top then was still representative of the types of trees the citizens liked via stewardship. Because of this, more variations of trees appeared across the visualizations, but also popular trees can be selected without being underrepresented by the top five trees previously removed.

A. A sample list of questions we could answer and why throughout the analysis is listed below.

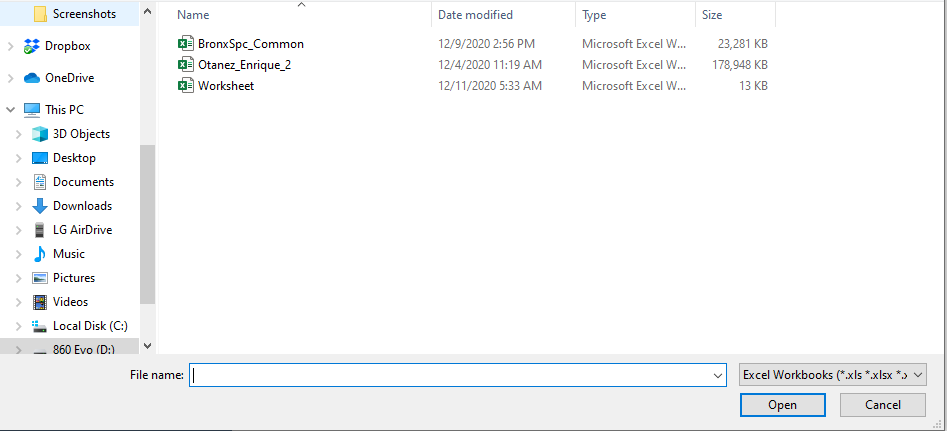
* What trees show the most consistent stewardship and the highest amount of stewardship?
  + The reason for stewardship is to try and understand how the citizens of Hunt’s Point neighborhood likes their trees and more importantly, what trees they like the most. It is a good base measure as well. Although it seems more practical to find the best tree for replacement use objectively, keeping the peoples minds in the analysis keeps that human touch, and again, can boost the overall improvement of the neighborhood. What we are looking for is between the signs of stewardship sub-variables, 4orMore, 3or4, or 1or2, we can see if there is any improvement of stewardship among the top ten trees of stewardship. Although we are looking for improvement, the top ten trees of the 4orMore sub-variable are considered as the basis of other variable improvement. If one of the chosen trees at the end of the analysis is one of the trees that shown improvement in stewardship, then it is a bonus reason to select that tree for replacement use.
* What trees are consistently on or off the curb, and what trees show improvement from on to off the curb?
  + The reason why this is an important question to ask is not as clear as say the damage to sidewalks in the last question of this list, on or off the curb involves the tree bed and roots of the tree. It is safe to assume that in real world application outside of this dataset trees with their roots and bed going off the curb can be harmful to the tree, and there is no use in planting a tree if their roots and tree beds consistently go off the curb. This can also affect the health of the tree and things like traffic and parking and etc.
* What trees show any harm or help from guards placed on them as well as improvements from none to helpful and harmful to helpful guards?
  + Guards for trees can be the metal grates the NYC parks and recreation department can put on the tree beds, but any other type of guard that can be put up by citizens. There is a list of examples included in the data dictionary. Unbeknownst to the citizens and even to our department, sometimes these guards can cause damage to the tree or even result in poor or fair health for the tree. It is important to understand that sometimes the metal grate is needed, and that the citizens will want to contribute to the tree via guards, and understanding what trees are less affected by guards and what trees improve with guards or not guards is an important question to ask when selecting the final tree for replacement use.
* What trees show good, fair, and poor health and what trees show improvement in health?
  + Like the others there is a trend here. By selecting the trees that show improvement in the respective KPI’s we can try to determine the tree that is best used for replacement. It is very important that we select a tree that is of consistent good health and that shows signs of improvement in health. Planting a tree that would soon die after fifty years, only when the tree is starting to show its true value to the neighborhood would defeat the whole purpose of the governor’s task.
* What trees show sidewalk damage and what trees show improvement from sidewalk damage to no sidewalk damage?
  + Why we ask this is because we do not want our efforts to show any negative affects after implementation. Most important is to find trees that are most consistently not doing damage to sidewalks and quite possibly the road, but to also include trees that show improvement to from damaged sidewalks to no damage sidewalks as well.
  + Find the tree type that are most consistently off the curb and assume that this may lead to further damage.

A. Procedure – Opening and getting the data into Tableau

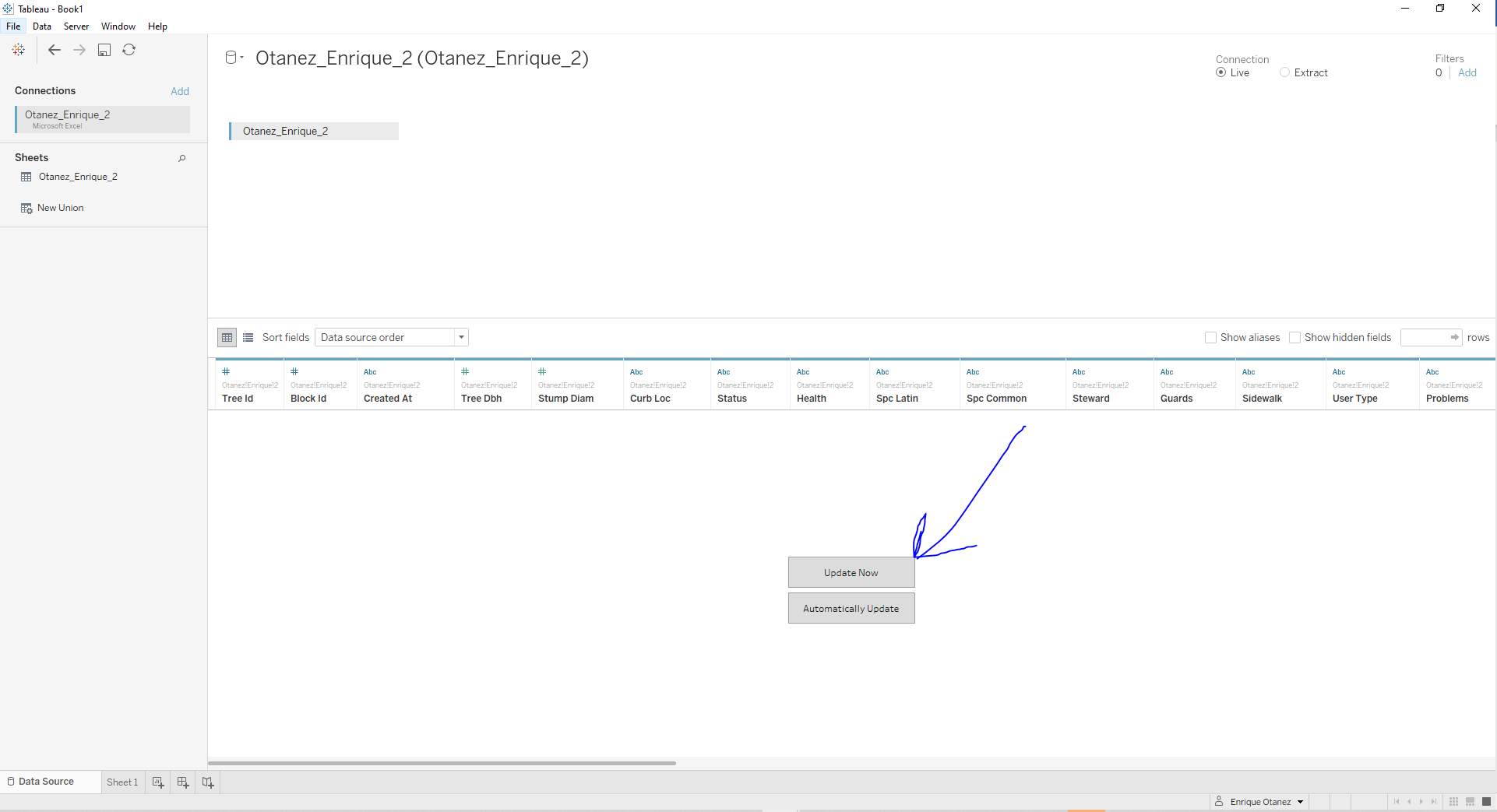
1.



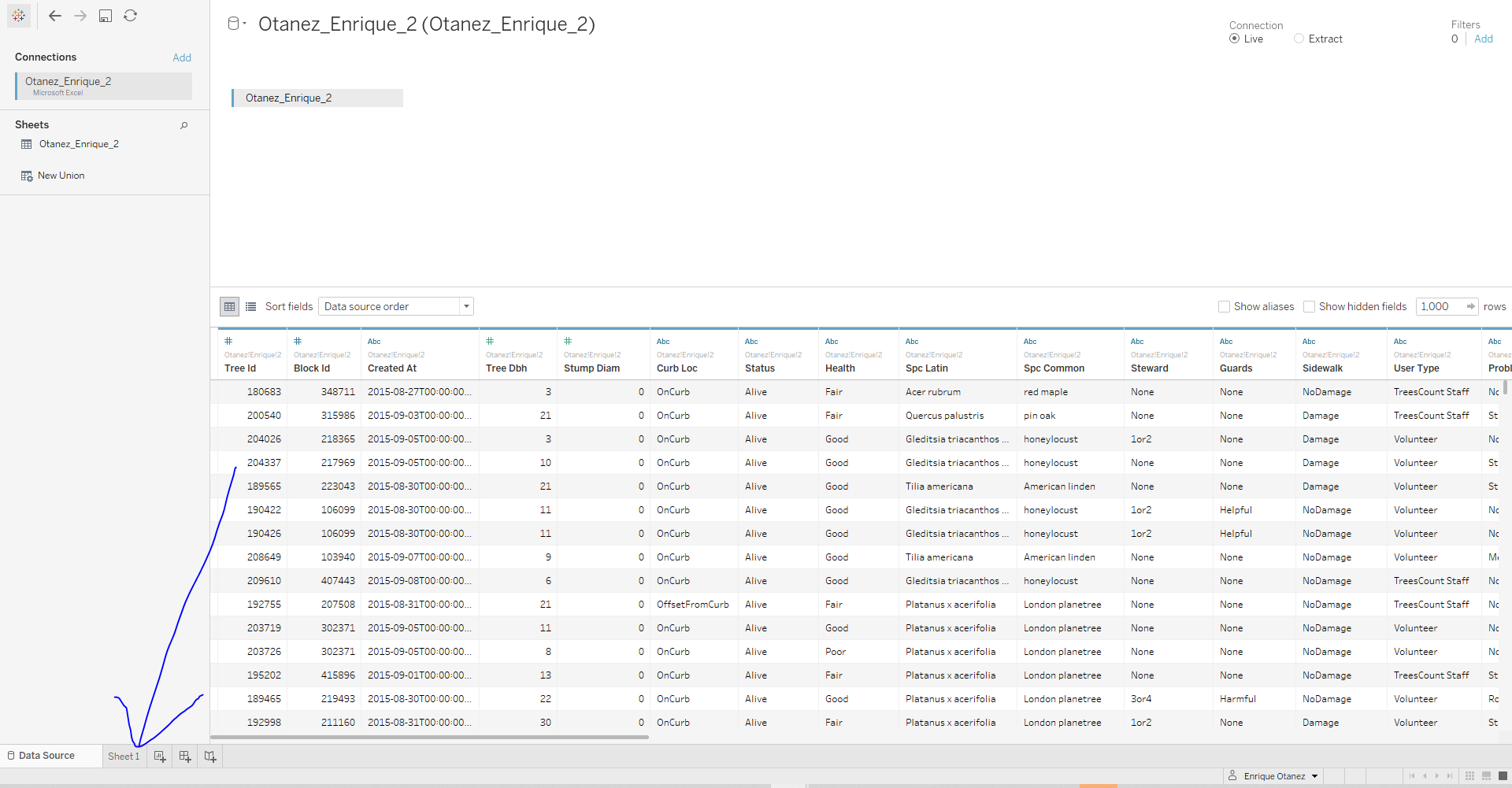
2.



3.



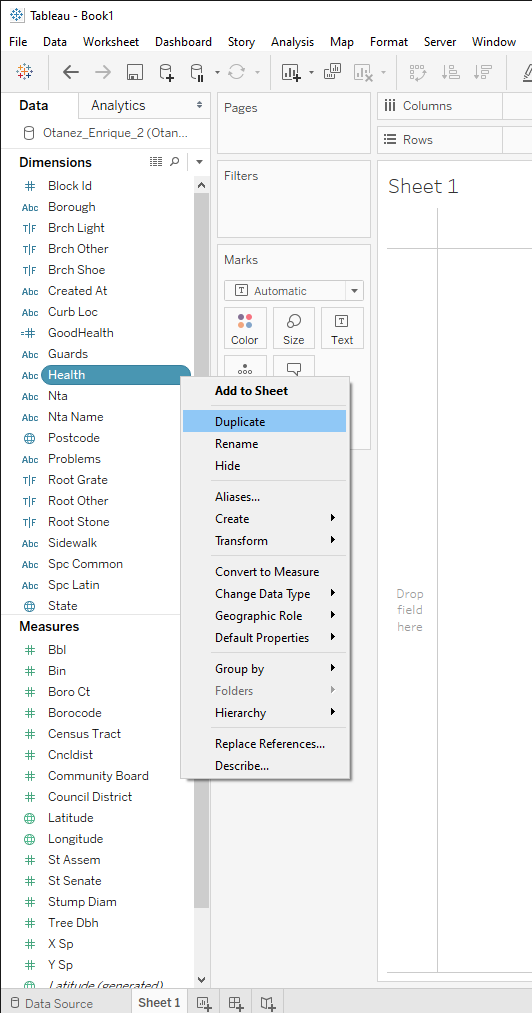
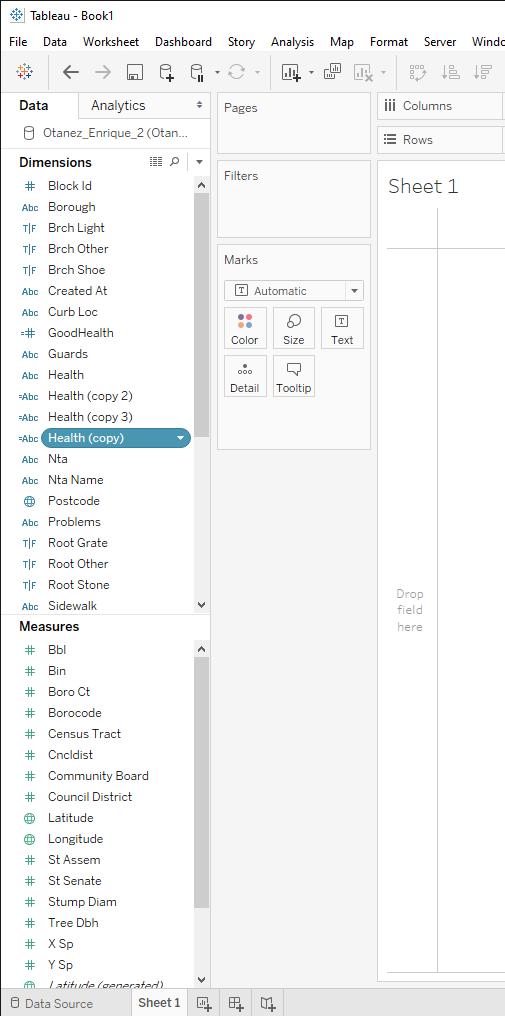
4.



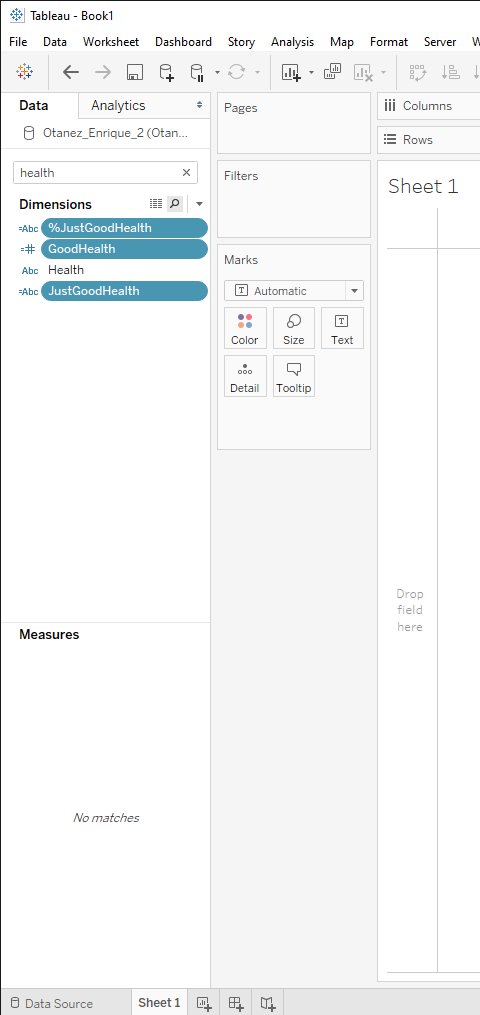
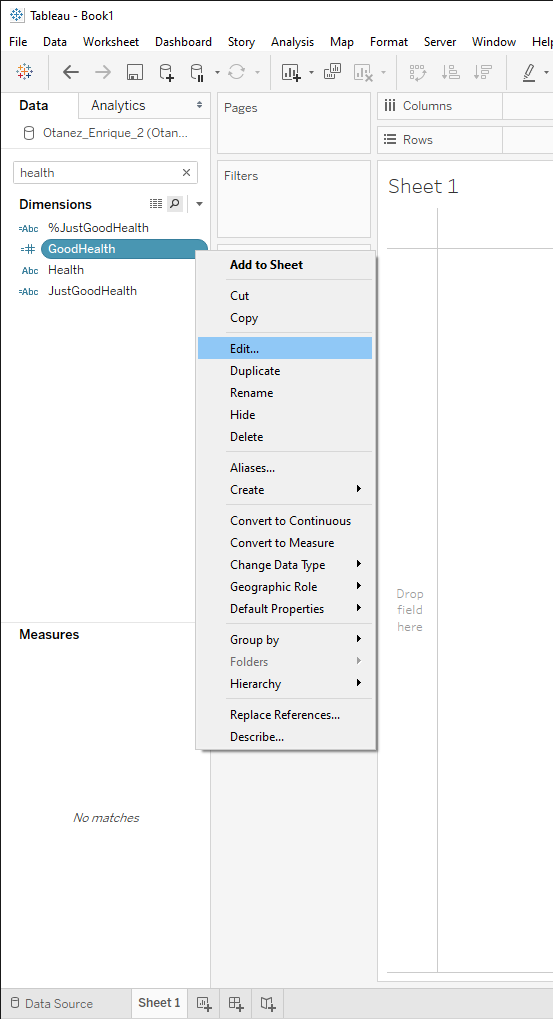
The above four images show how you need to get started with Tableau. First is opening the desktop application like in picture 1. Once the application is open then go and click on the Microsoft Excel under “To a File”. Once selected proceed to navigate and select the datafile you want to manage, in this case it is Otanez\_Enrique\_2, or the tree dataset in the second image like in picture 2. The third image is what appears after you select the data from your file directory. Click on the Update Now button and wait for your data to load (this sometimes can take a while) like in picture 3. Once uploaded you will see the data displayed at the bottom like in picture 4. Now onto the first sheet to start making the calculated fields necessary for this analysis.

A. Procedure – Calculated Fields

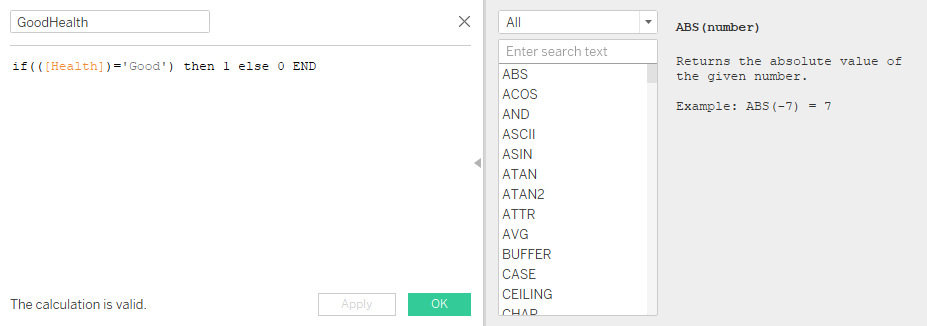
1. 2.

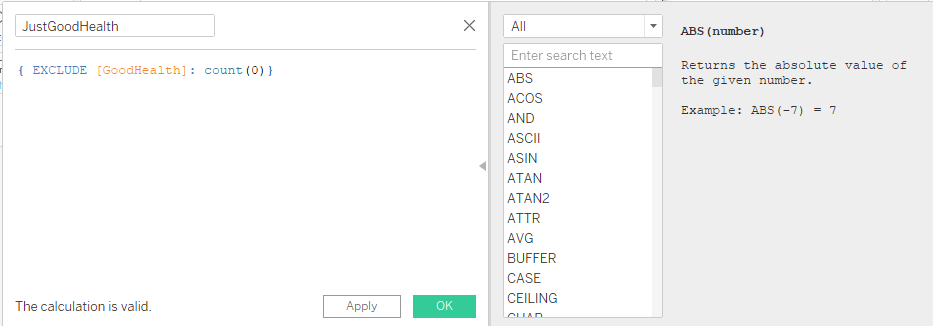
3. 4.

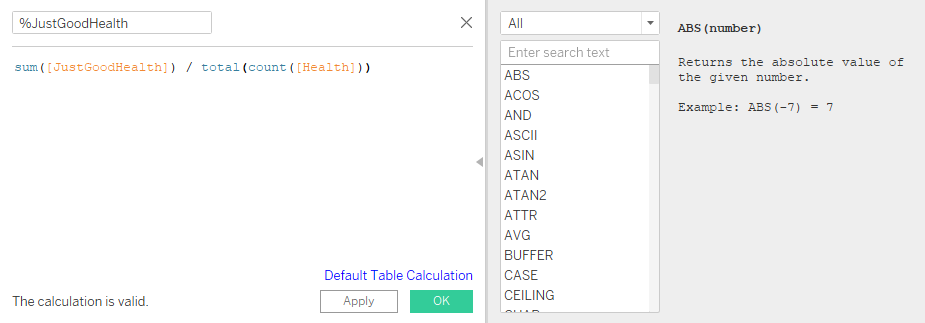
5.



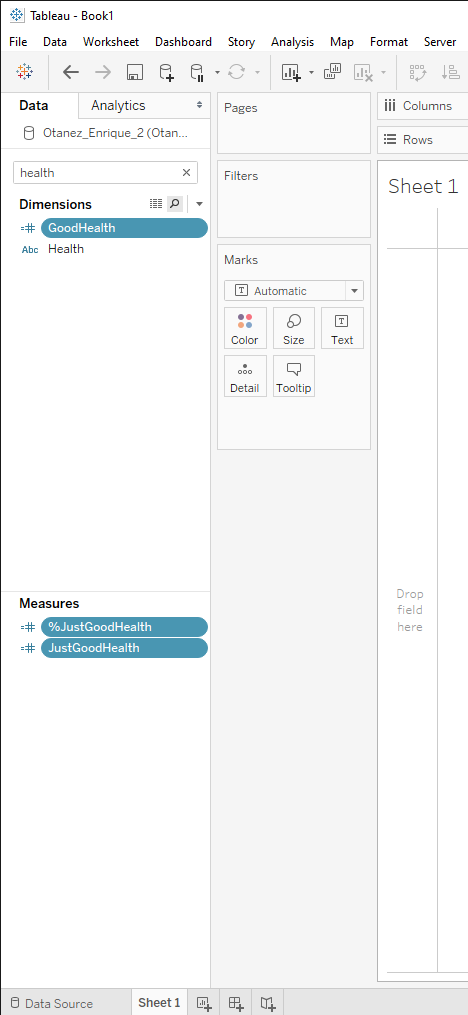
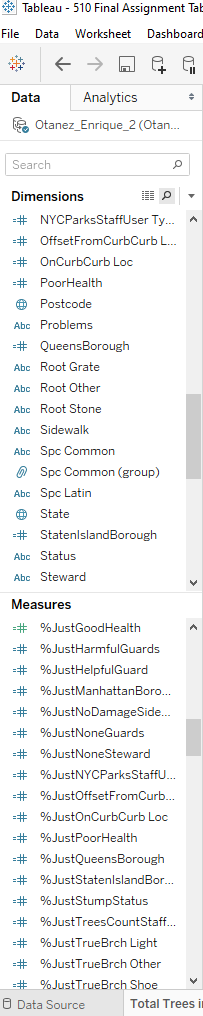
6.



7.



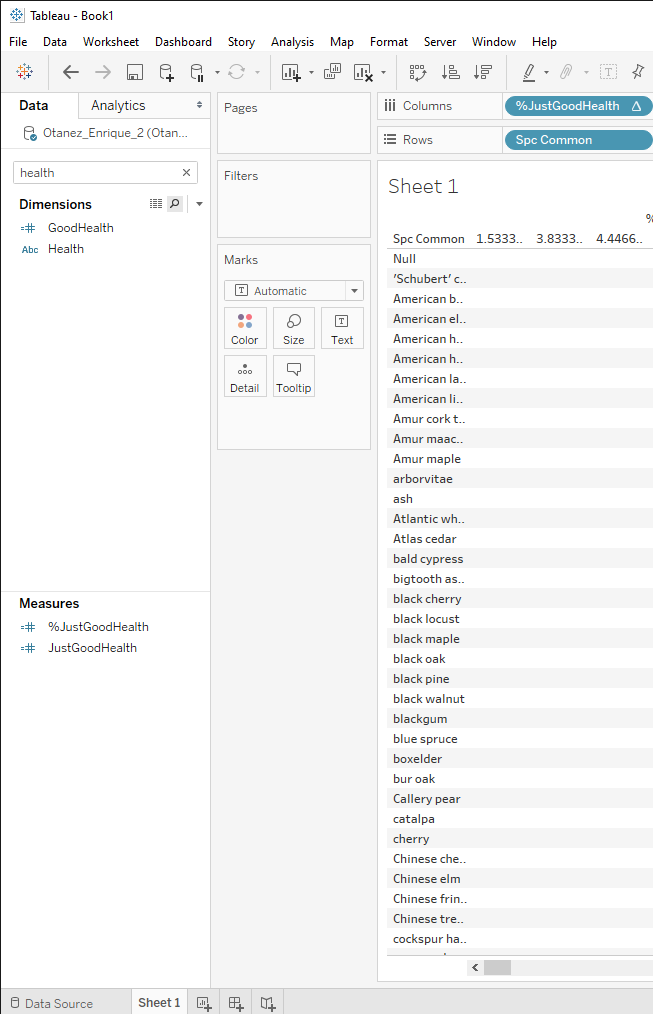
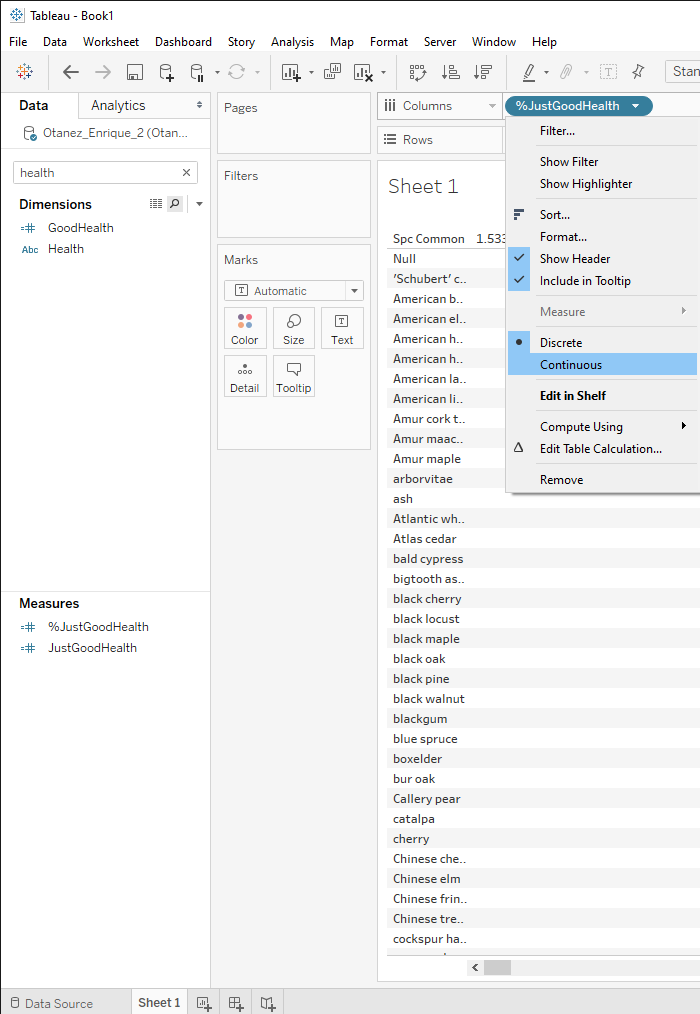
8. 9.

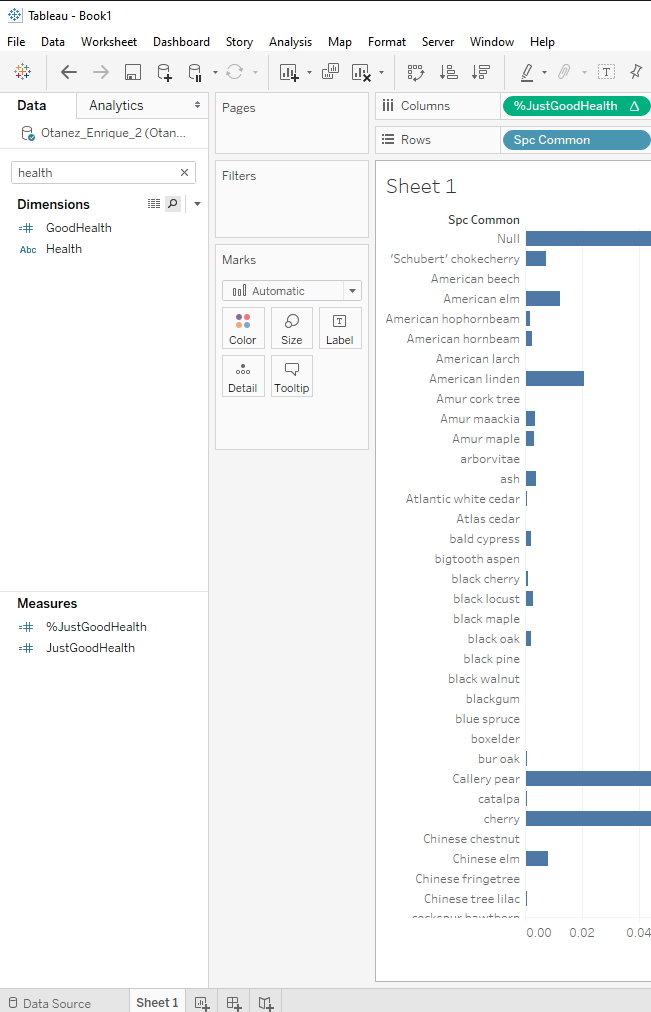
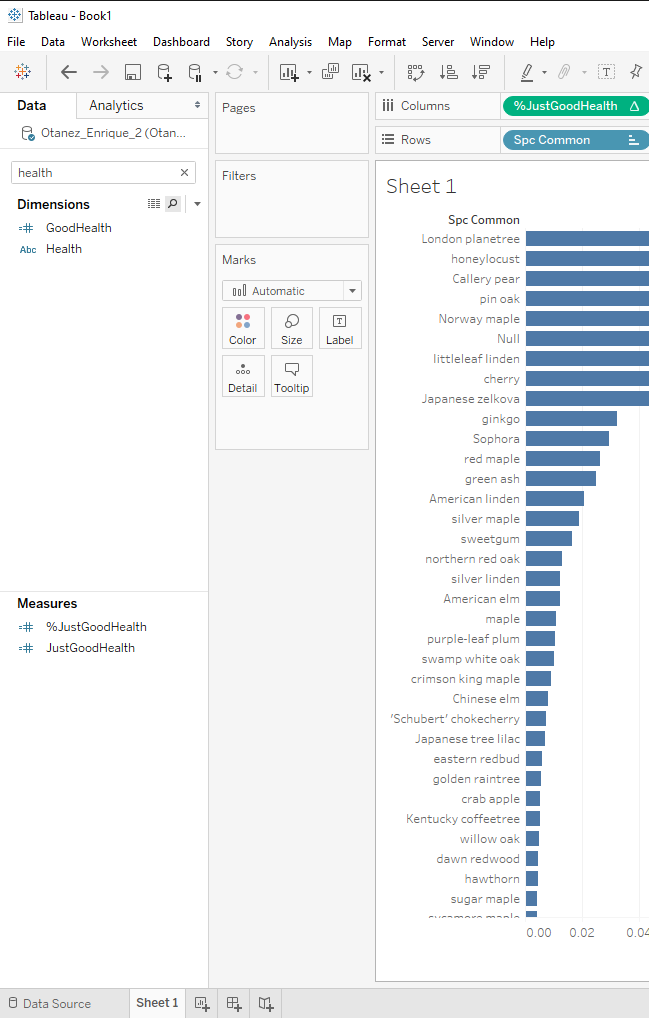
These next eight pictures show how you need to create all the necessary calculated fields for the analysis. At the minimum you will be needed to do all the KPI’s; steward, curb\_loc, guards, health, and sidewalk. For this set of pictures, the KPI health will be used as an example. Picture 1 shows that you need to duplicate the variable into a new pill three separate times like it is shown in picture 2. Next in picture 3 you need to name the new fields or “pills” as the nomenclature shows in the picture. This is because you need to maintain the same naming convention so that they are easier to find in what may be many generated fields you need to do. It also helps to search by health at this point as when putting in code in later pictures they may change from dimensions to measures. Picture 4 shows in the drop down or by right clicking the pill you are then able to see the drop-down menu where you can edit the pill. This results in the first pill you need to edit in picture 5. Disclaimer: you need to do these pills in order of instruction or in order of the pictures. Enter the code into the pill and hit ok and do that for the next two pictures 6 and 7. Finally at picture 8 we can see all the variables have been edited with their respective codes and now have even moved places from dimensions to measures. As said before you may be needed to do more than just the KPI’s depending on how far in the analysis you want to go. Picture 9 is a good example of all the possible variables you can create calculated fields for and all of them went under the same procedure.

A. Procedure – Make the first analysis: Health

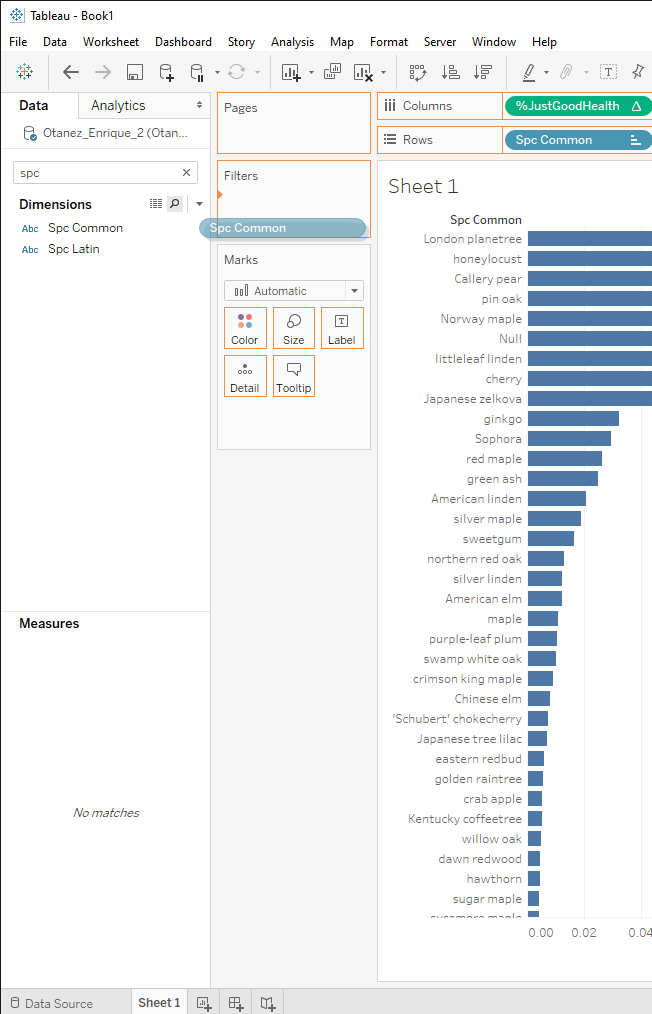
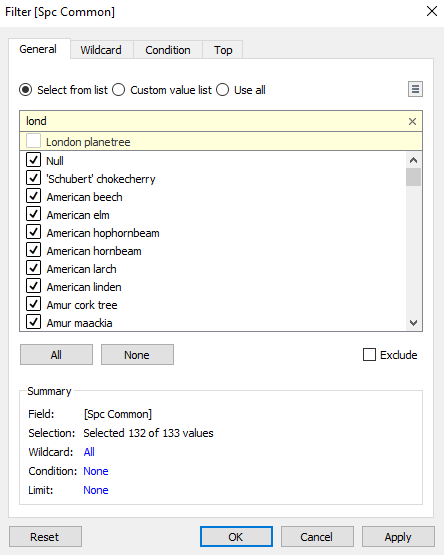
1. 2.

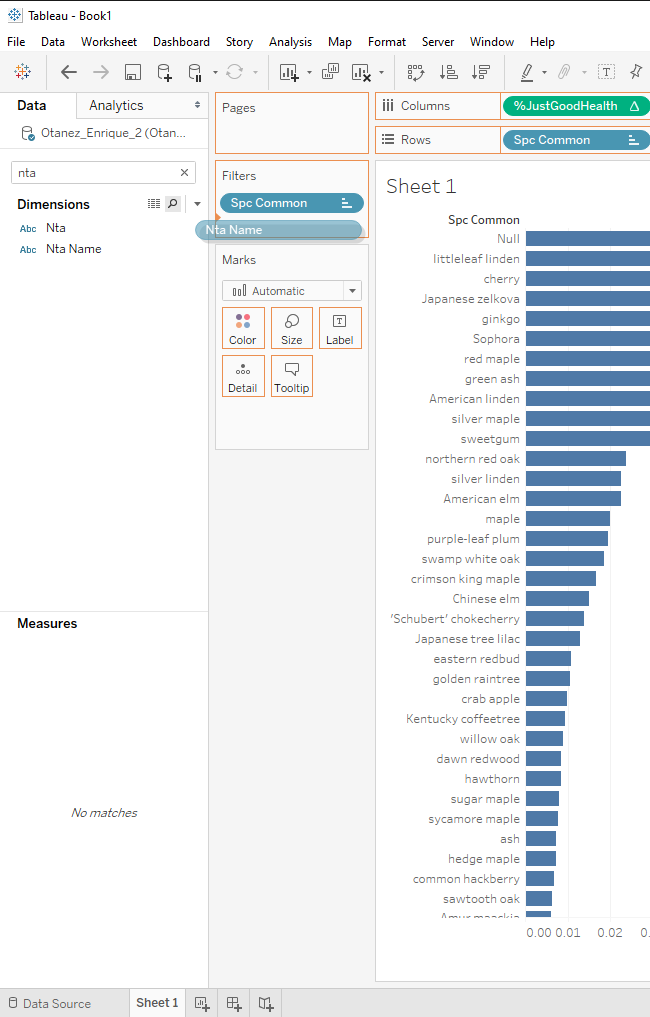
3. 4.

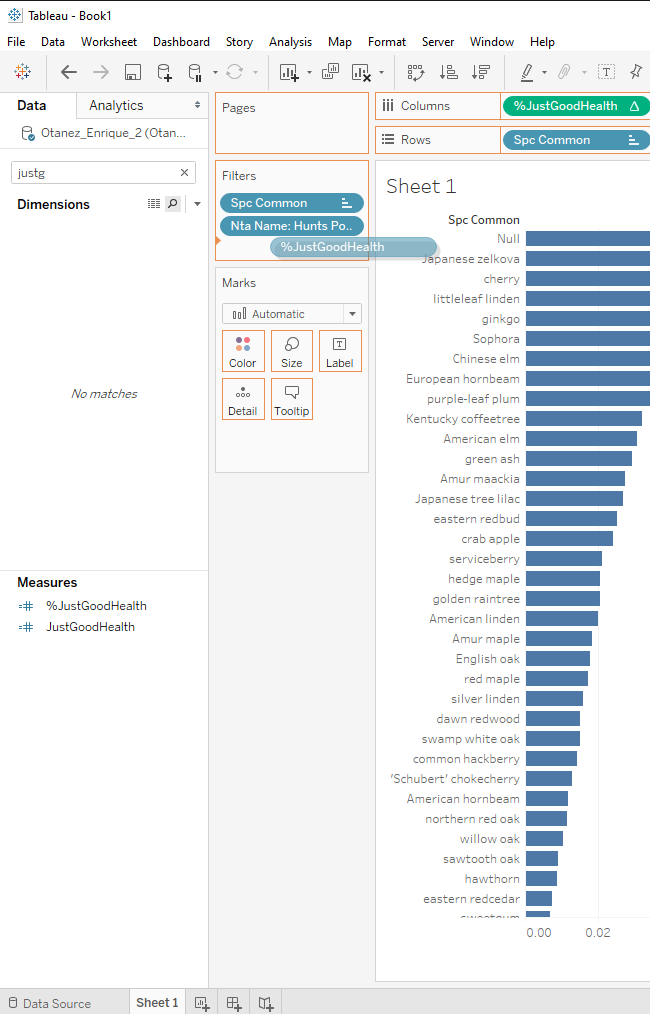
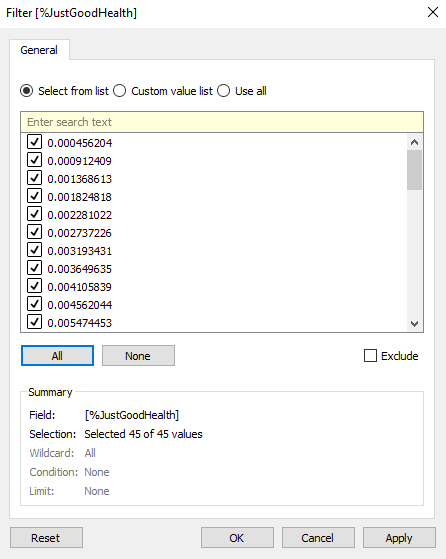
5. 6.

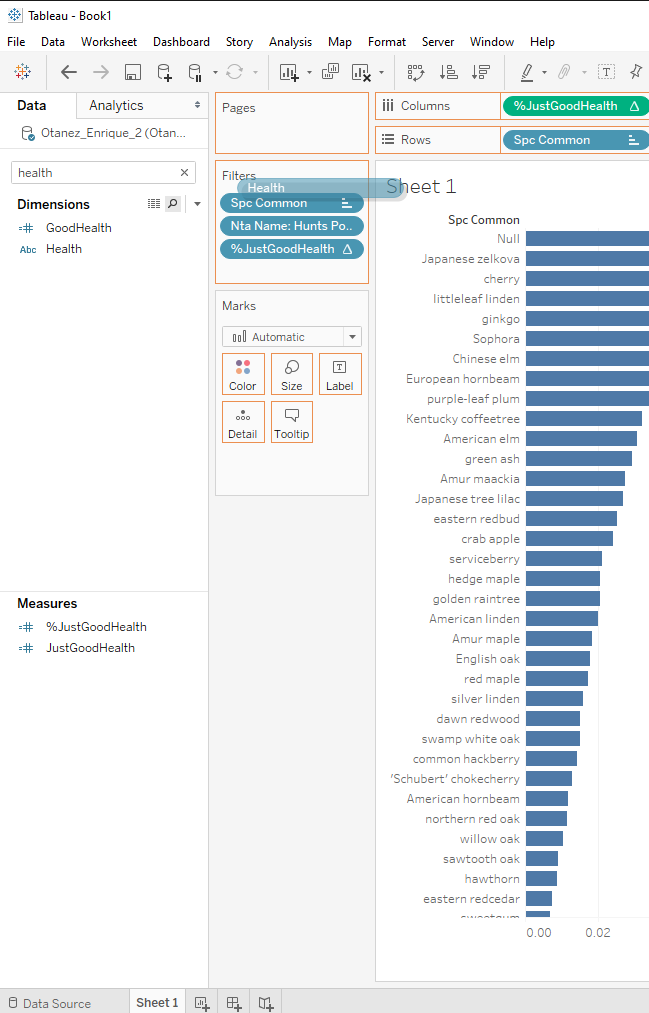
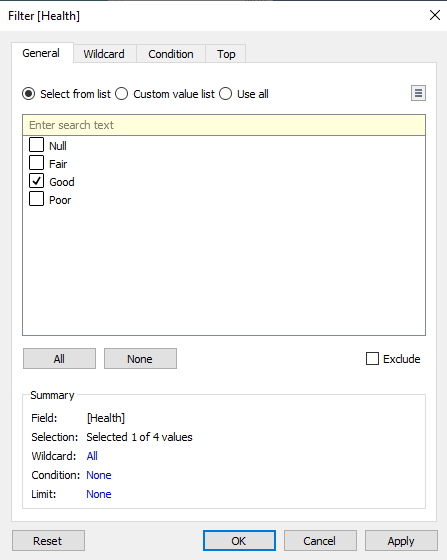
7. 8.

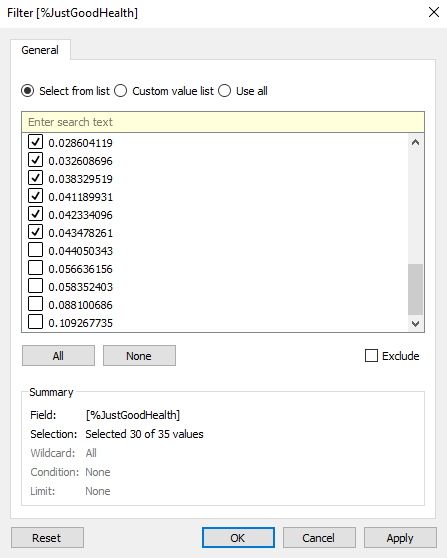
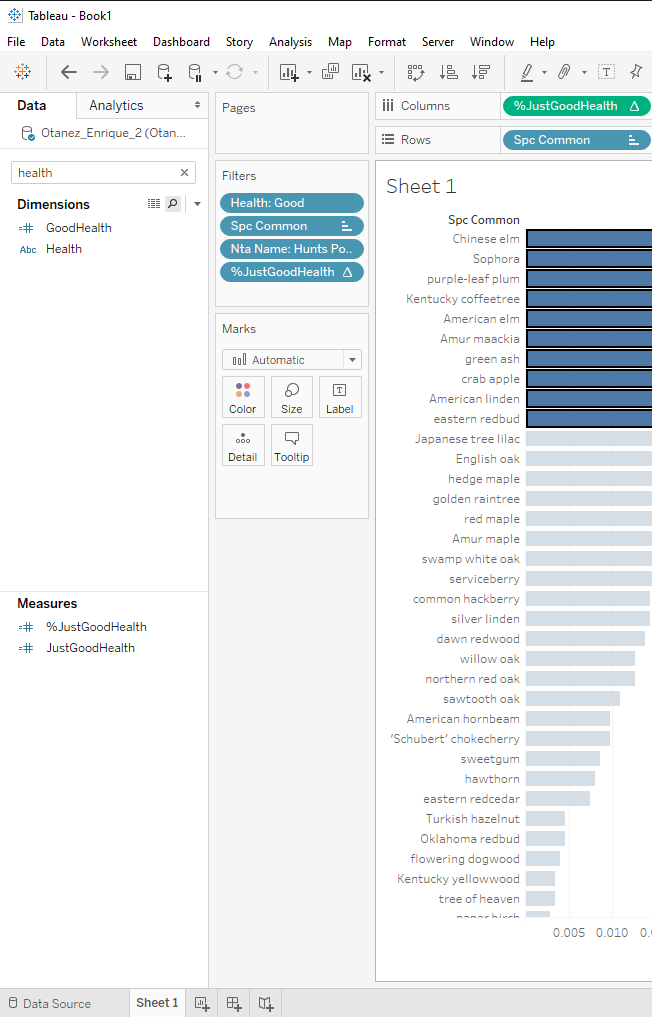
9. 10.

11. 12.

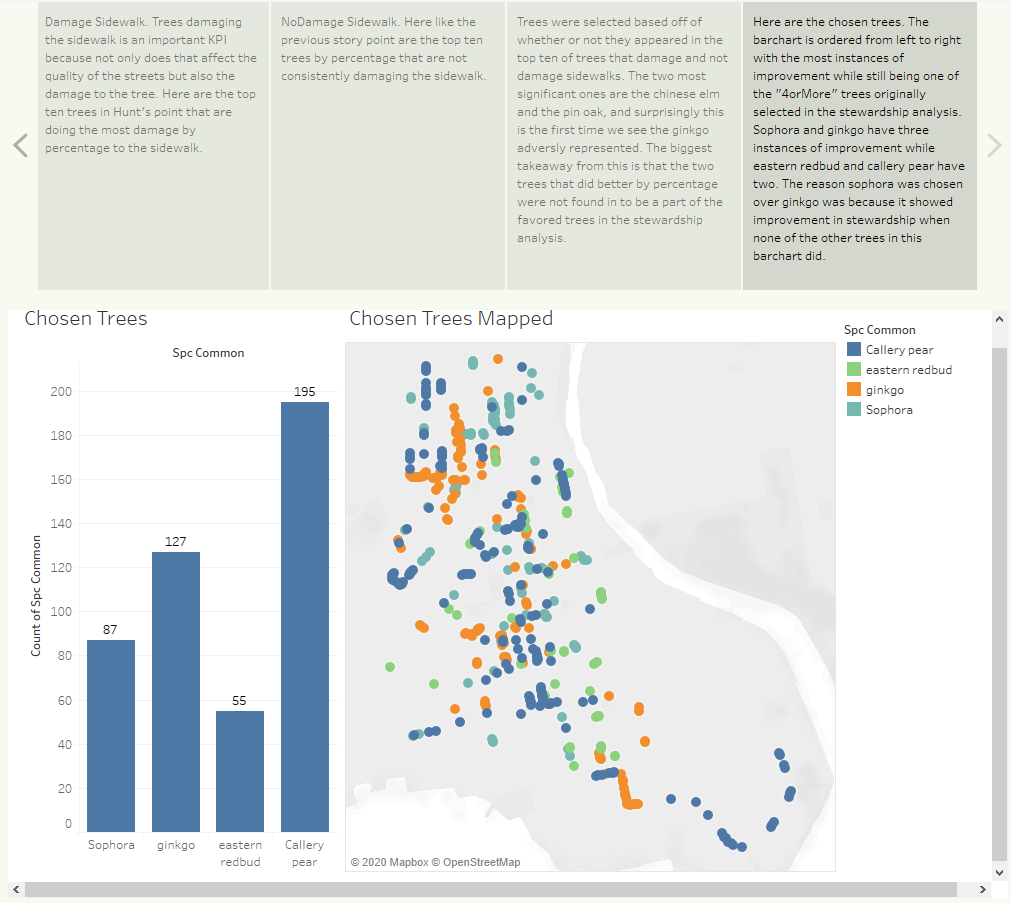
 

13. 14.

The first thing that should be done is to create a new sheet and then do what is in picture 1 and bring over spc common and %JustGoodHealth in rows and columns section. Next is to open the drop-down box in the %JustGoodHealth that was placed into the column bar and select continuous like in picture 2, and picture 3 should be the outcome you will see. Picture 4 is to show some ordering found by selecting the buttons right below server in the ribbon on the top. Next for picture 6 and many of the following picture, drag over spc common over to the filter section and in picture 6, remove the top 5 trees in the data. In picture 7 move over nta\_name, this is so that in picture 8 we can filter and only select the neighborhood Hunt’s Point. Again, in picture 9 we bring over %JustGoodHealth and make sure that all of them are selected, we will be coming back here in picture 13. In picture 10 we see that all are selected, and we can move onto picture 11 where we bring in Health to the filter section as well and this time only selecting Good like in picture 12. Now onto picture 13. Open the filter again and uncheck the five most abundant trees by scrolling to the bottom and unchecking the boxes. Finally, in picture 14 we see the final visual with the top ten trees selected. This is where the rest of the analysis can take start from. This is only one of thee that will be made to try and find some improvement in health to the top ten trees of this variable. Like the other KPI’s to follow, this will be done twice more for health to show for fair health and poor health.

# The Storyboard



<https://public.tableau.com/profile/enrique.otanez#!/vizhome/510FinalAssignmentTableauProjectFile/Story1?publish=yes>

B. Simple Business Scenario – Prescriptive Analysis

First things first, we plotted all of the trees and in the Bronx area to understand what it would look like. After that we moved to finding all the stumps in the Bronx as that is the crux of the issue at hand. Doing so while color categorizing and showing stump diameter showed an impressive map of many stumps that can be replaced in the Bronx borough of NYC. Wanting Hunt’s Point, a map of all the stumps and dead trees were mapped for the neighborhood and there was a total of 228 stumps and dead trees that can be replaced. As said before in the first procedure the top five trees in Hunt’s Point were removed as the abundance of them often did not provide insightful visuals when included in the analysis. After this the most favored trees based off stewardship in Hunt’s point was established and as well as the improvements of stewardship of those threes. On and off the curb of the stewardship trees were then found showing many trees were more favored when considering that on percentage, more of those trees found to be favored higher in the “4orMore” sub-variable rather than the “1or2” sub-variable. The rest of the KPI’s were looked upon objectively meaning that objectively finding the top ten trees of the new variables rather than testing the top trees of stewardship against it like what was done with the curb analysis. This process is then done with all the remaining KPI’s finding more and more improvements until a conclusion is reached based off which tree or tree has the most consistent improvements of all the KPI’s.

B. Purposes or Questions or Application

A more in-depth purpose is to understand what this analysis can provide outside of the goal of replacing stumps and dead trees in Hunt’s Pont. Choosing the top ten trees of each variable’s influence on Hunt’s Point was based off the assumption that these threes are abundant enough to show that they are planted and chosen in hindsight because they are more successful trees and that they do well in the climate of NYC. Yet the top ten does seem somewhat arbitrary as it underrepresents the trees that are not as common based solely for other reasons like space availability after most of the spots in city blocks and streets were taken. The insight here is that more consideration should have been placed on what criterium was chosen for trees selected when variables were tested against them. But a further purpose that our analysis did provide was that there were trees shown to have improvement that were not a part of the standard trees chosen in the stewardship analysis. Because of this there are more candidates to choose from that are not based off of stewardship but more objectively to consider.

B. A sample list of decisions to be made is provided below to understand how we tackled this project.

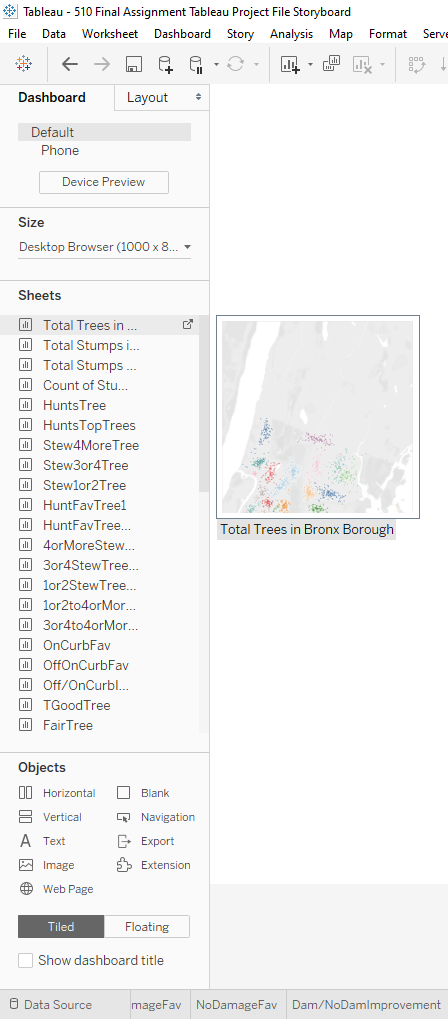
* There will be one test of KPI’s tested against each other.
  + This was to see the capabilities of our analysis and dataset with future analysis in mind, but only one instance was made, and the results were conclusive to not much discernible information. Regardless of that, the analysis should be taken towards the future and steps to compare KPI’s should be taken.
* The criterium of final chosen trees will be that of the most improvements of trees selected in the primary trees found in the stewardship analysis.
  + We wanted the focus to be on what trees would be most liked. Because of this we were looking for the trees with the most improvements that were also in the trees picked from out stewardship analysis. The reason was good enough to go forward with the overall analysis of the project thanks to many of the trees having a higher abundance but it often left out many other trees as good candidates as well.
* All but curb\_loc analysis will have trees look objectively rather than trees based off trees selected in the stewardship analysis.
  + We wanted to look objectively at the top ten trees of each variable even though we were looking at the trees that showed improvement that would then show on the streets picked in the stewardship analysis. We wanted the trees to be uninfluenced in the improvements and had hopes that the trees would be among the trees selected for the stewardship test.

B. A sample list of questions we could answer and why throughout the analysis is listed below.

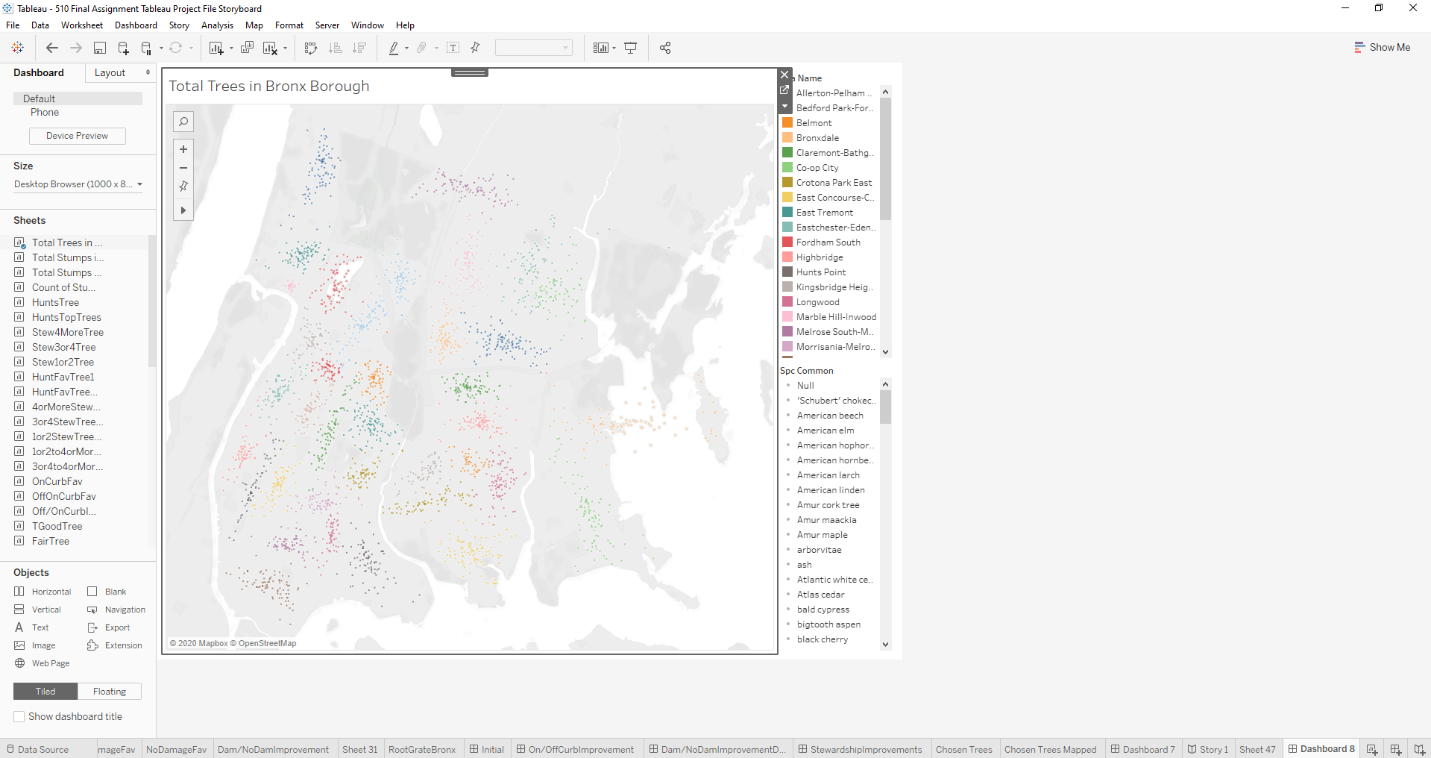
* What trees are showing improvement amongst the KPI’s?
  + The trees that shown improvements in KPI’s are the callery pear, littleleaf linden, ginkgo, sophora, American elm, the eastern redbud, silver linden, cherry, and Kentucky coffeetree trees. One thing that is interesting to note is what trees improved in what variable analysis. For stewardship it was the callery pear, littleleaf linden, sophora, American elm, and the eastern redbud. For damage/ no damage it was the ginkgo and sophora. For on or off the curb it was the ginkgo, American linden, and the eastern redbud. For health it was the ginkgo and sophora. Finally for guard it was the silver linden, cherry, and the Kentucky coffeetree. For guard, those three where not a part of the selected trees for the stewardship test. The instances of improvement for each tree are as follows. Callery pear – 2, litteleaf linden – 1, Norway maple – 0, ginkgo – 3, sophora – 3, American linden – 1, American elm – 1, eastern redbud – 2, silver linden – 1, cherry – 1, Kentucky coffeetree – 1. Why is this important information? Why it is important is this is how we selected the best trees for replacement use. The trees with three instances were the first choice, the ginkgo and the sophora, with the callery pear and eastern redbud following second.
* Why are some trees more consistent among datasets?
  + The first obvious reason is the abundancy of trees. Earlier it was a decision to remove the top five most abundant trees in the entirety of the dataset because the as they were so abundant that by choosing the top ten of each variable, the top five would most likely show up in every visualization regardless. But it is still hard to escape the fact that other trees though less abundant by comparison, is still quite abundant compared to many of the lesser used trees, and there is such a big disparity between many of the trees compared to the most abundant trees still existing in the analysis. Other reasons is up to assumption. One that was previously stated was that for reasons being that the abundant trees were planted more because they could handle the climate, are hardier than other trees and can survive the light pollution, city smog, and hardness of the ground underneath all the concrete and asphalt. Other reasons could be that the abundant trees were planted so many times because at the time of planting, they were simply popular at the time.

B. Procedure – Creating a Dashboard, the basics

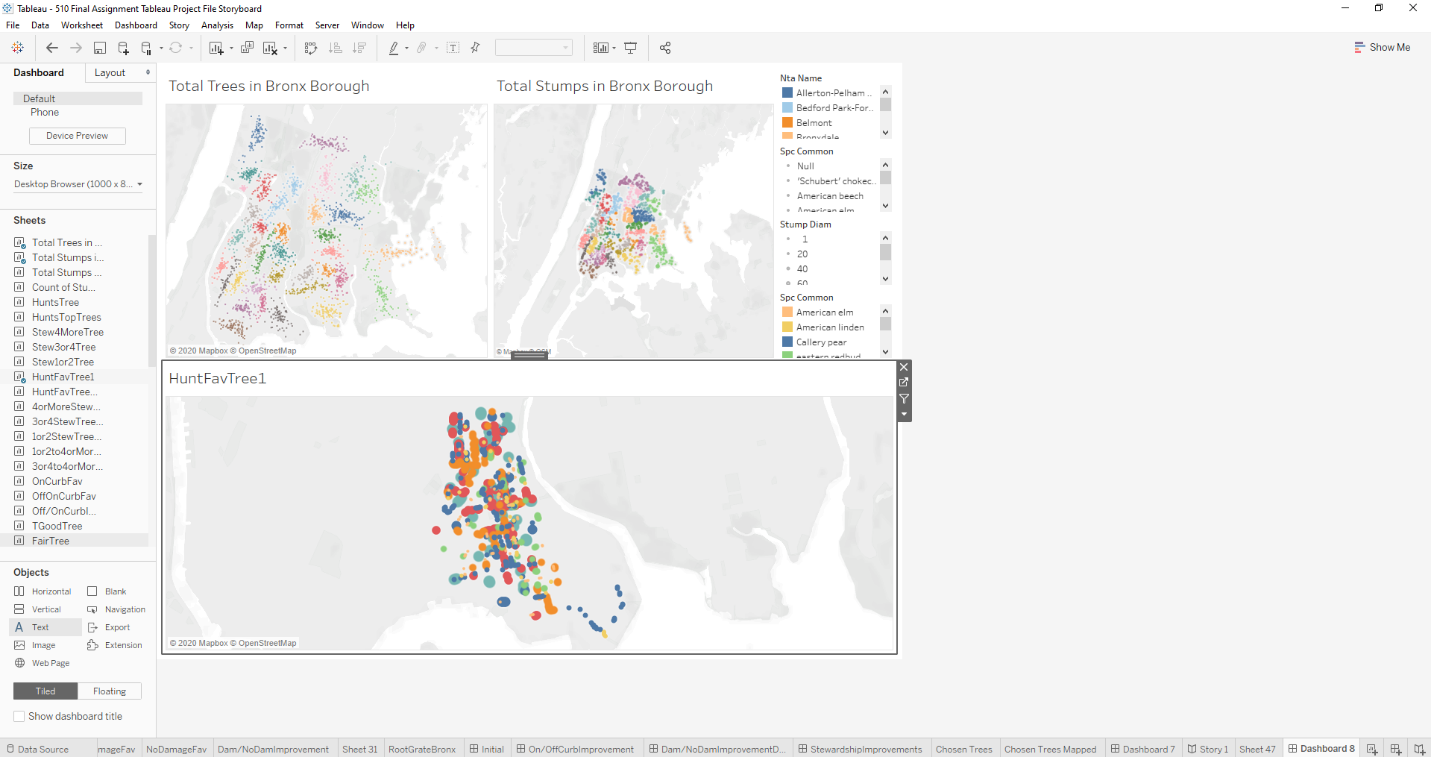
1. 2.

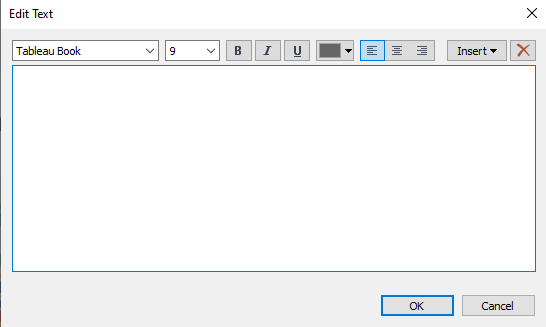
3.



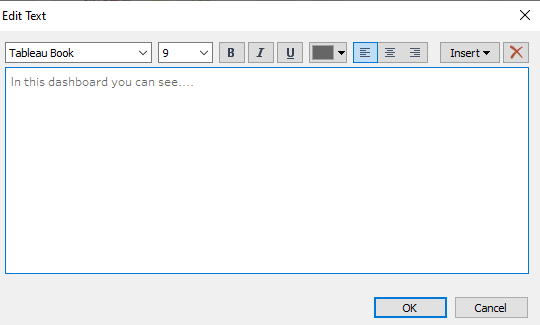
4.



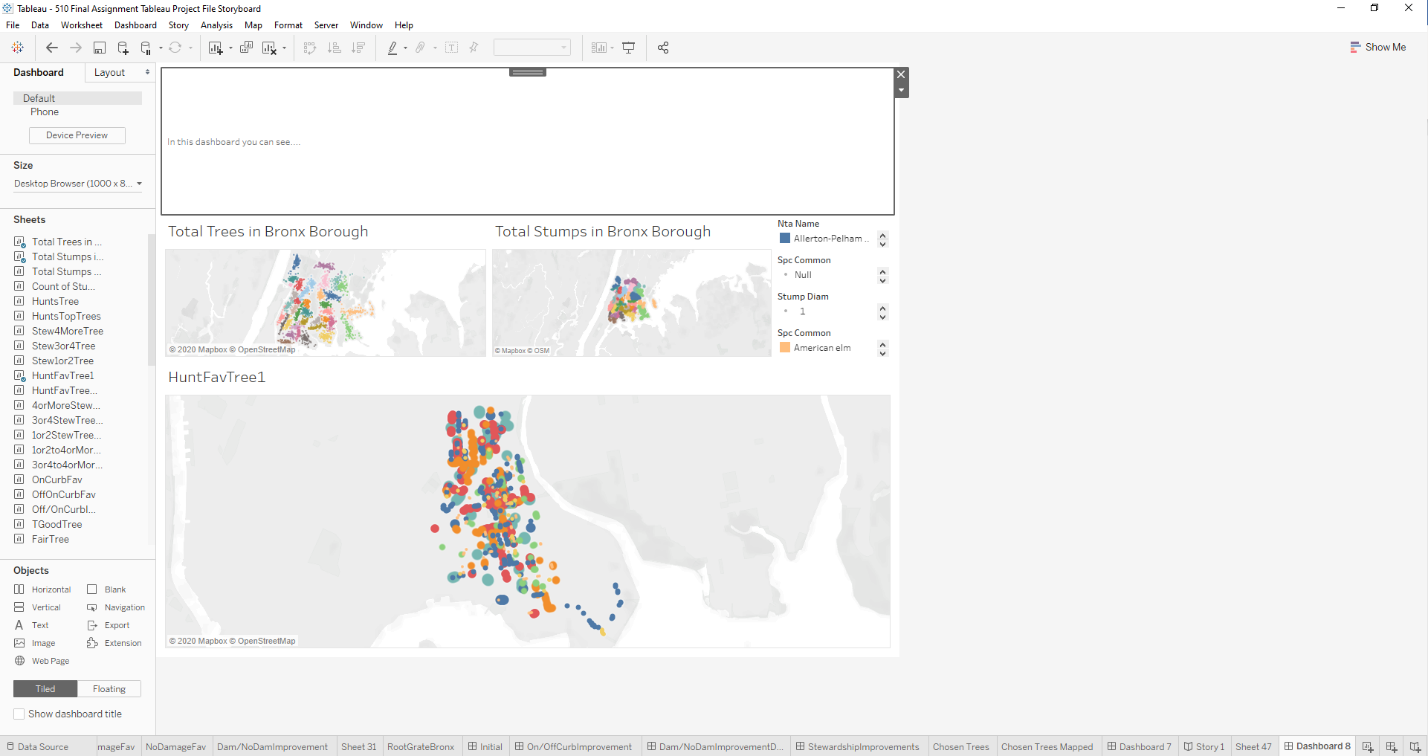
5.



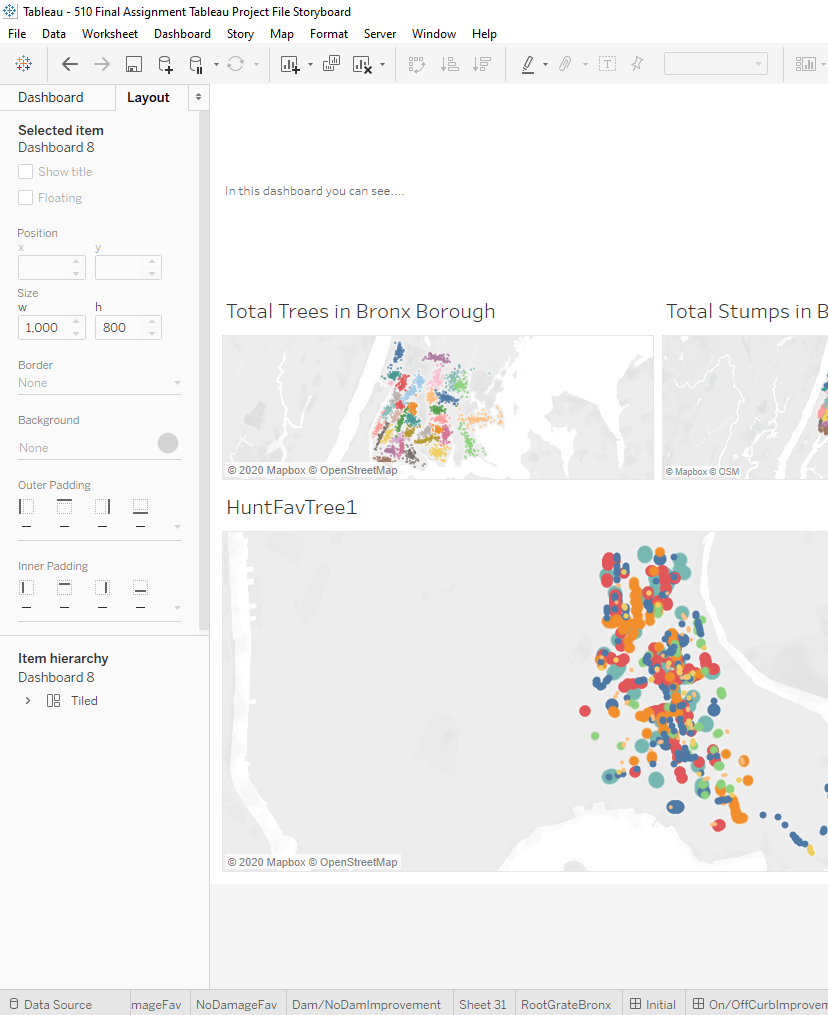
6.



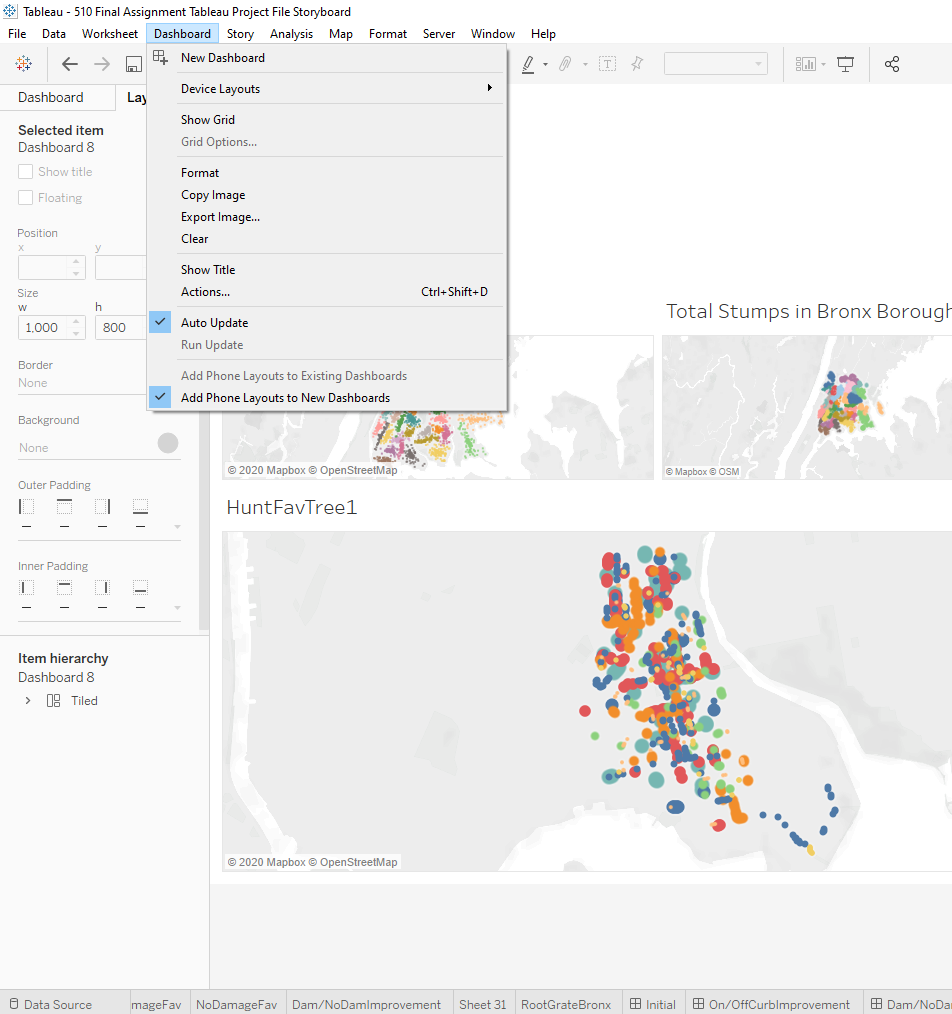
7.



8.



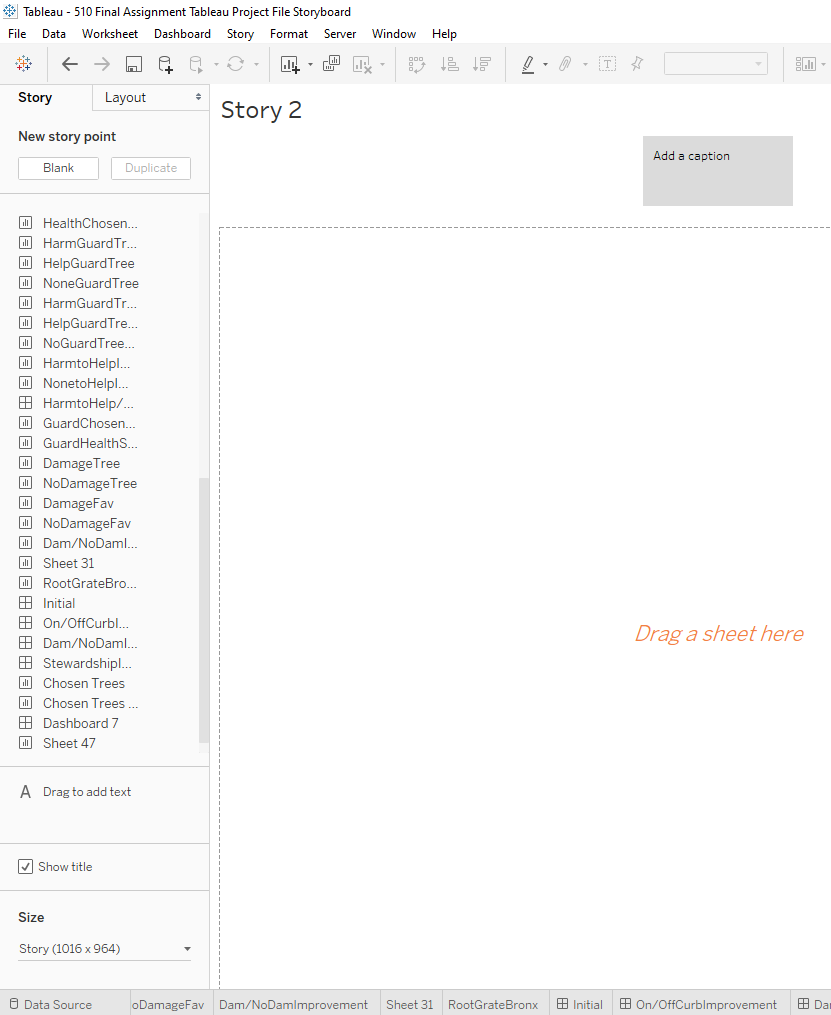
9.



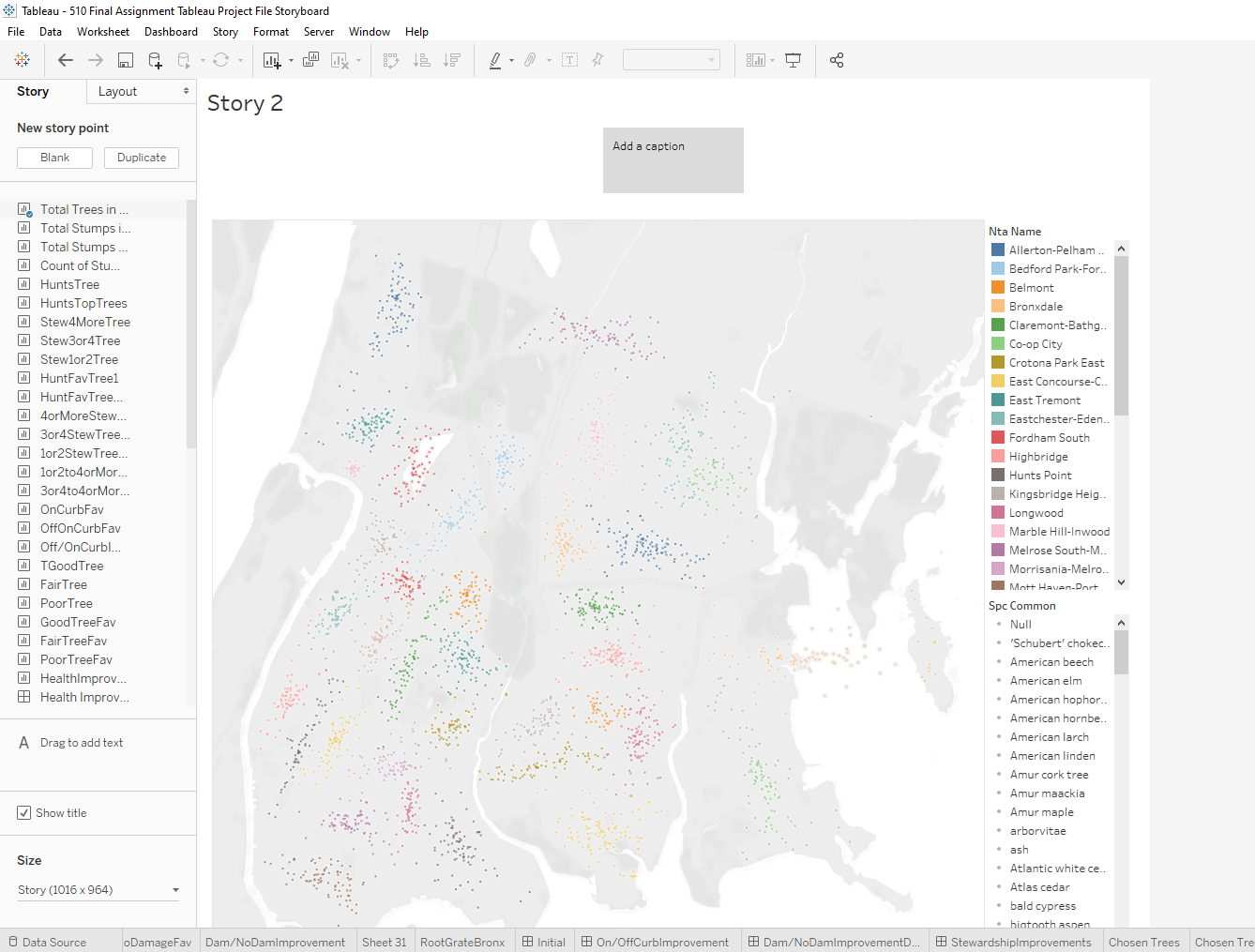
Now that you have has some time to familiarize yourself with Tableau and the layout created a dashboard should be easy. Like creating a new sheet for analysis, you open a new dashboard shown in picture 1 and in picture 2 is the result. Picture 2 also shows on the right you can drag and drop the analysis you have made via sheets and drag them onto the dashboard and picture 3 is the result of that. Picture 3 shows how more than one sheet can be moved and manipulated on the board for more eye pleasing visualization layout. On picture 4 you can see in the lower left-hand corner that the text box is lightly highlighted. By double clicking on that you then see picture 5 a text entry box to give some description to the dashboard, so that viewers have some context as to what they are looking at. Picture 6 shows some text and picture 7 shows the textbox placed within the dashboard. Picture 8 shows that if selecting the layout in the top right corner of the dashboard screen you have more options to customize the board. Picture 9 also shows more options you can place upon the board for more individualization.

B. Procedure – Creating a Storyboard, the basics

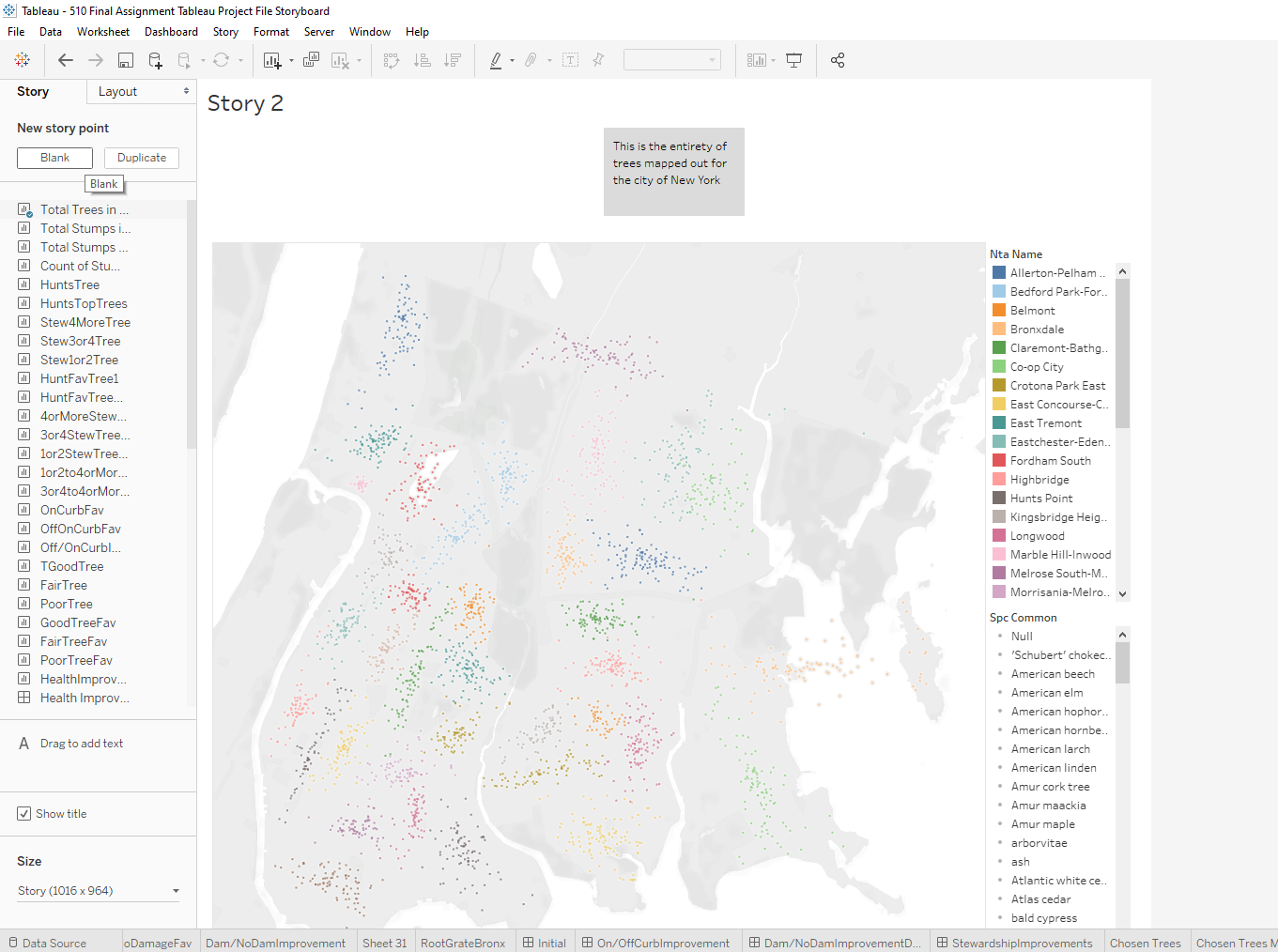
1. 2.

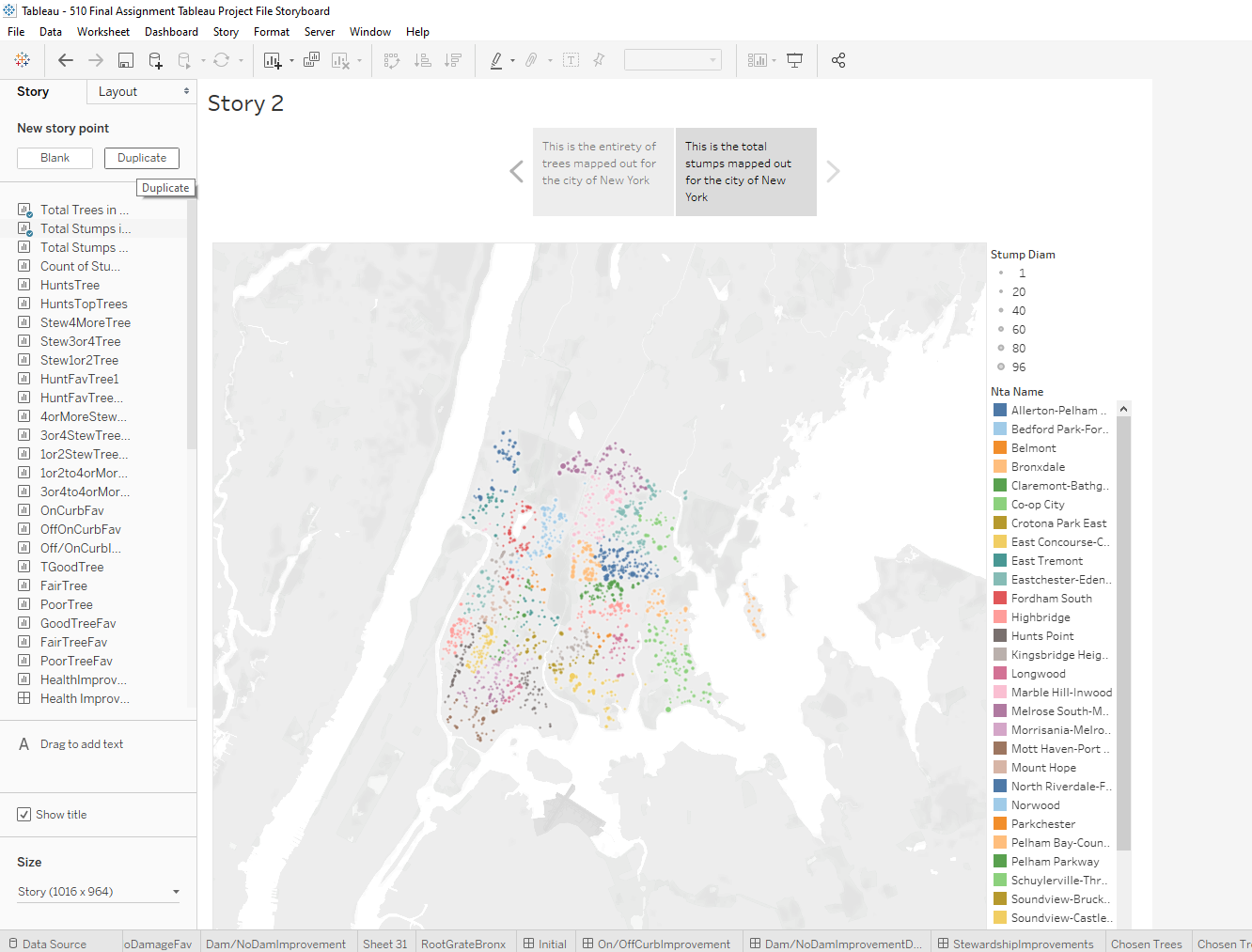
3.



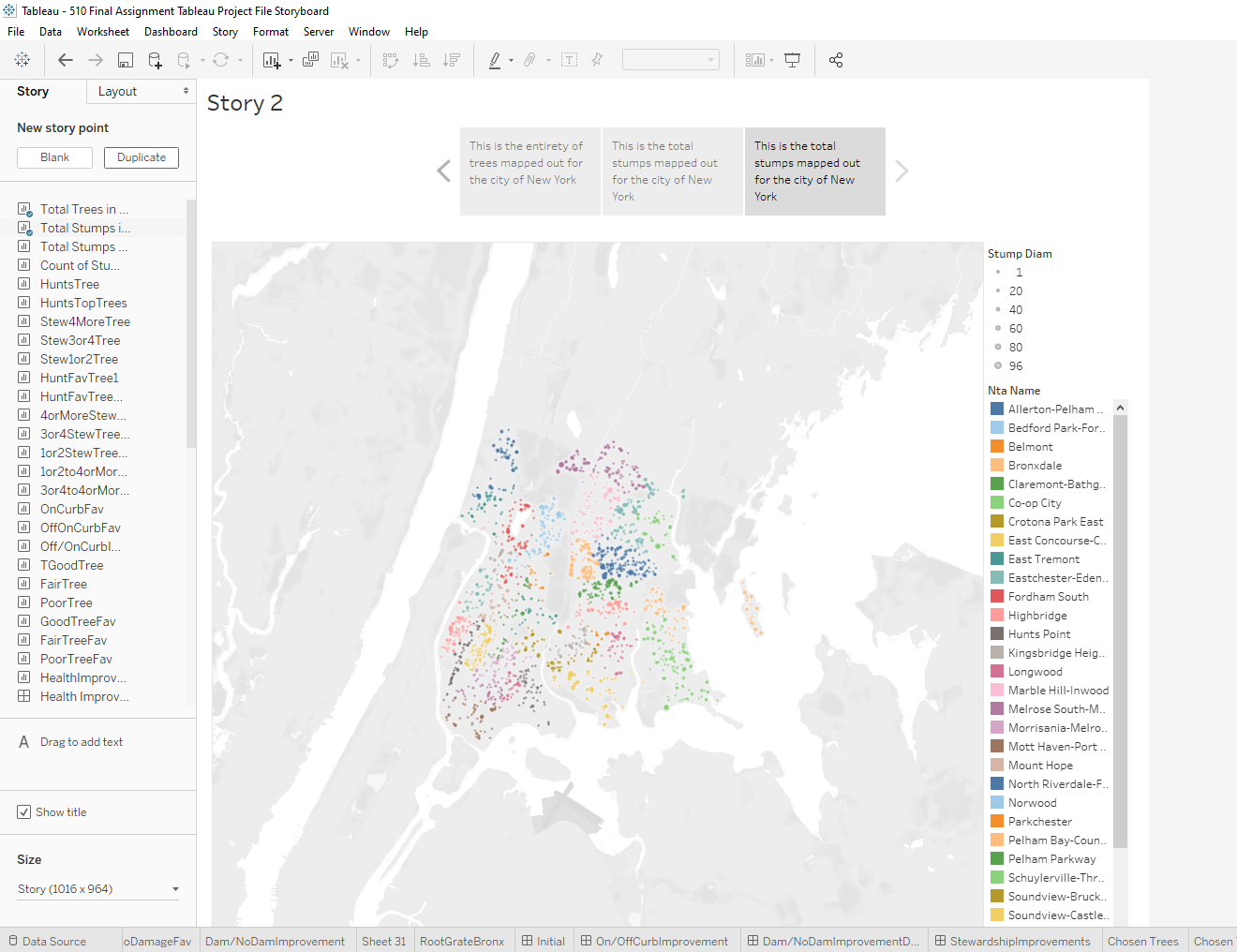
4.



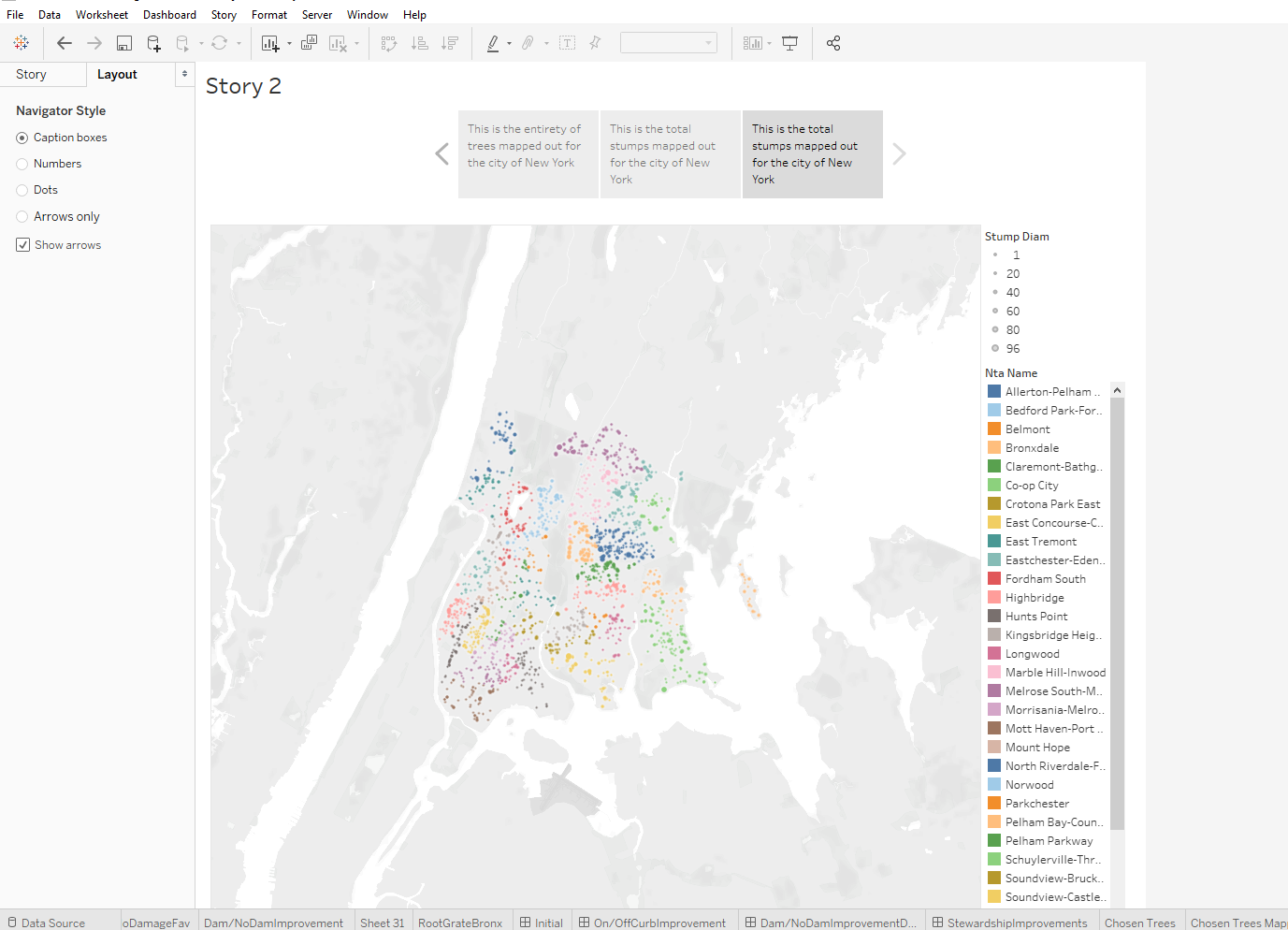
5.



6.



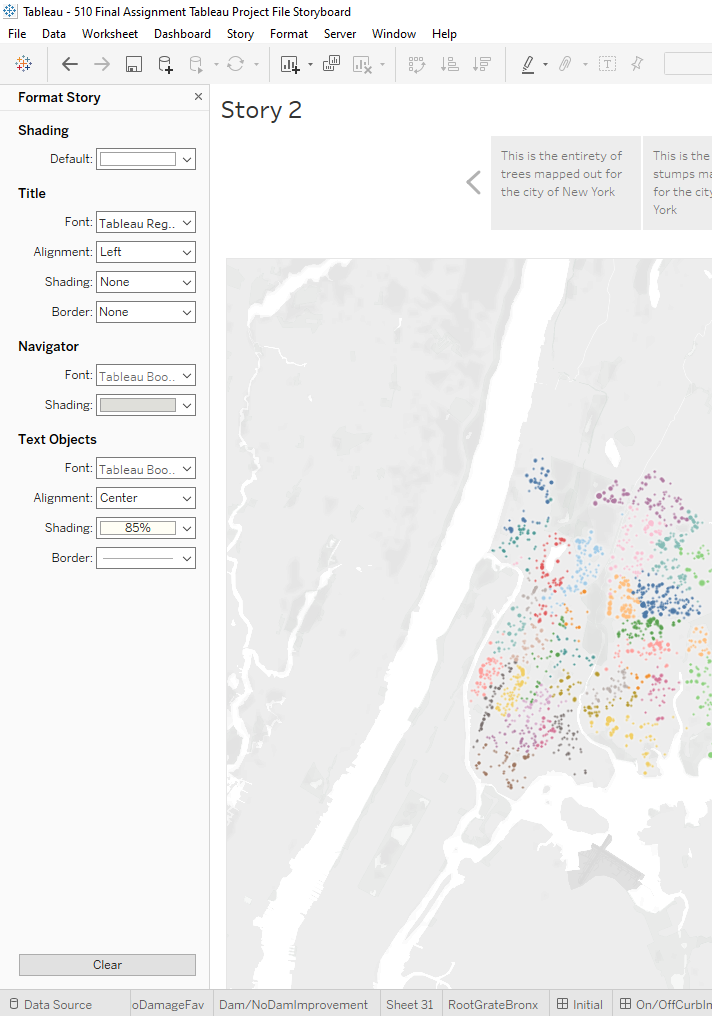
7.



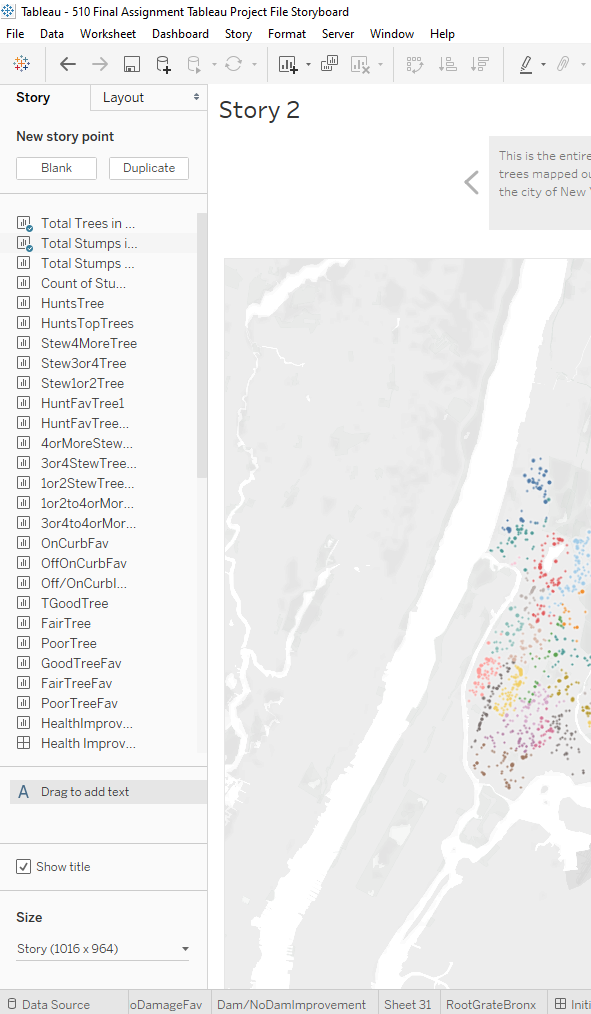
8.



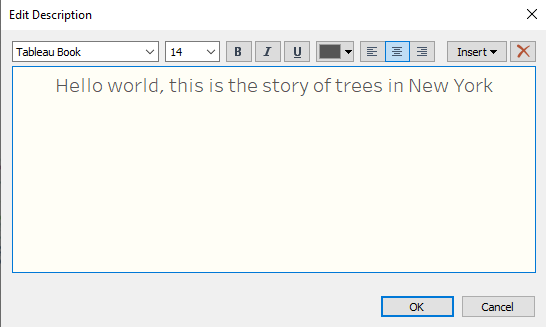
9.



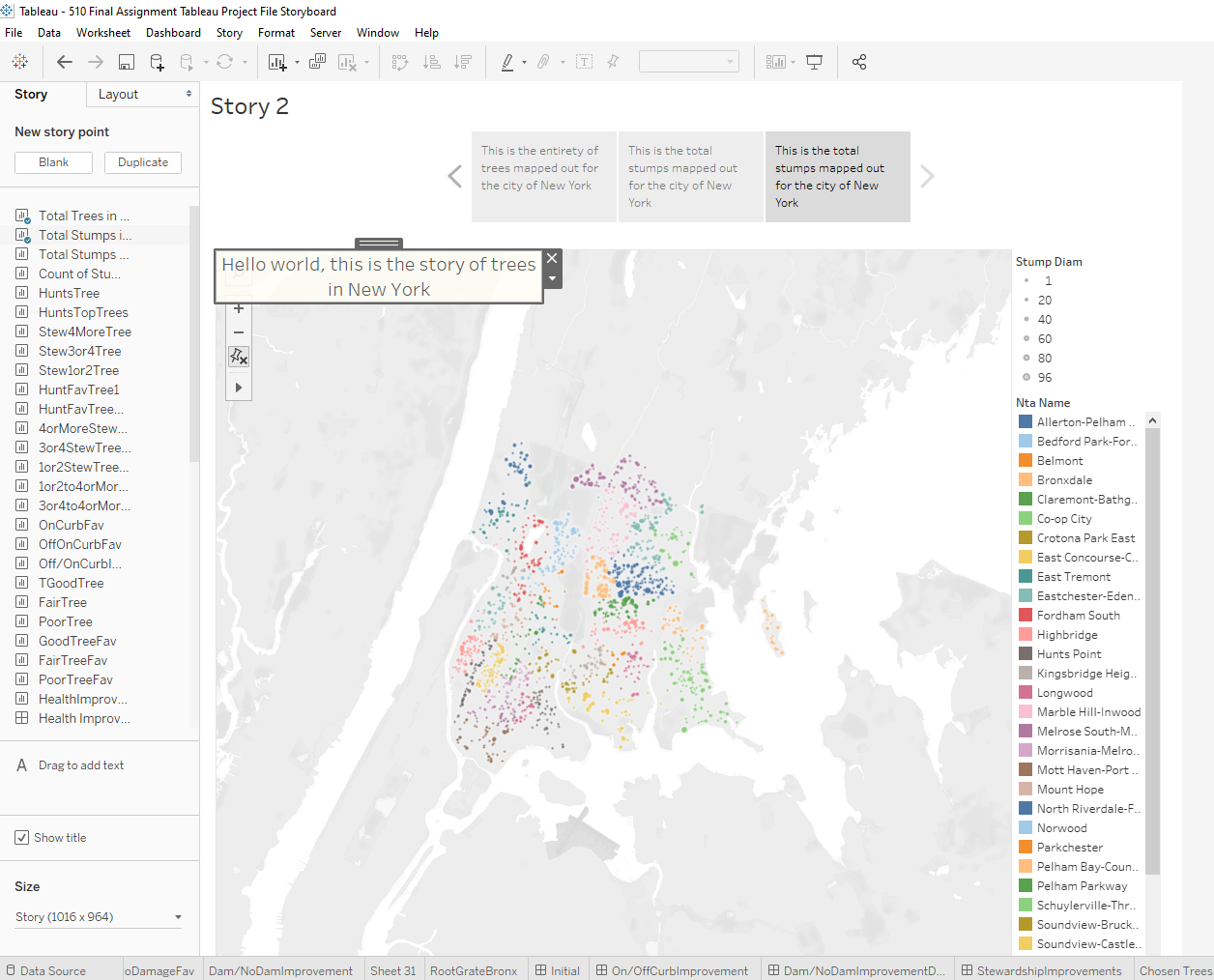
10.



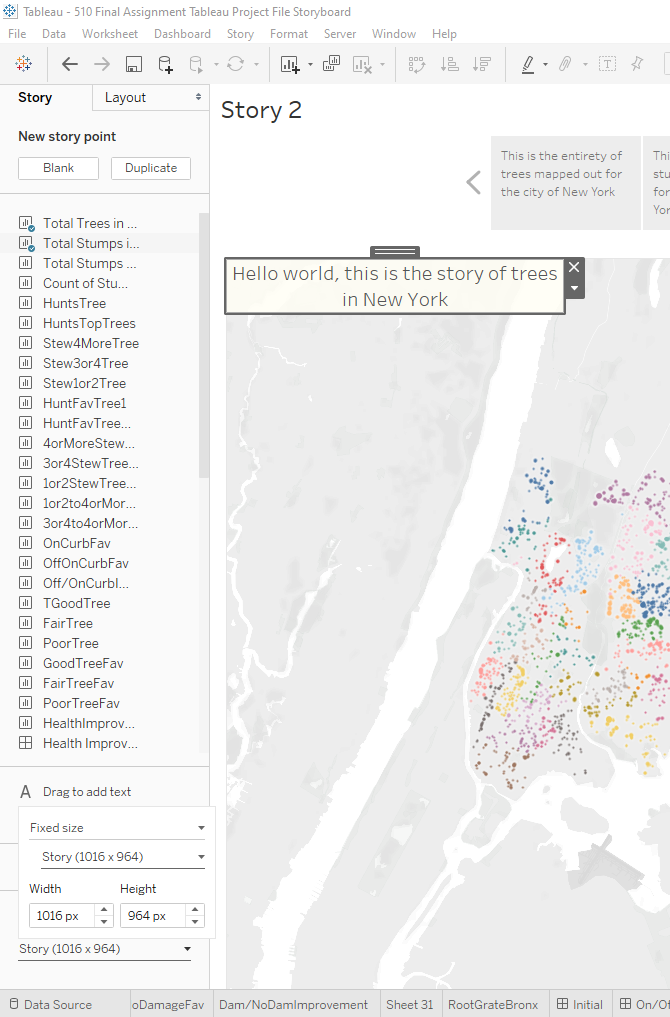
11.



12.



13.



Picture 1 like dashboard before it starts with selecting story at the bottom right of Tableau desktop and picture 2 is the result of that. In picture 3 we can now see that from the left column labeled story you can drag and drop the sheets and now dashboards over to the storyboard pane, but only once sheet can go on to the storyboard at a time. To fix this in picture 4 we see that in the top right corner in story you can select blank to then create a new story point where a new sheet or dashboard can be placed within. Picture 5 shows the new sheet inside the storyboard with the caption filled out at the top. You can make duplicates of your story point like in picture 6 with the button duplicate right next to the blank button where we made the new story point if there is any more clarity you need to say without clouding up a story point with too much information. In picture 7 you can see that switching over from story to layout in the upper left of the storyboard pane there are more options to customize the storyboard. Picture 8 shows more options for your storyboard by hovering over story in the ribbon up on top of Tableau Desktop and by clicking on format you can see the outcome which is Picture 9, more options to customize the storyboard. Picture 10 shoes a very useful option which is to add text via drag and drop from the bottom left corner of the storyboard pane. Once that is done you will see the result of that in picture 11. Picture 12 is the text box fitted in a nice position on the storyboard. Lastly, picture 13 shows that another option that is available to you is on the bottom left hand corner of the storyboard pane which is to the size of the board to fit any type of screen you would need be it laptop or projector or monitor or tv screen.

# Results/Insights/Conclusions

The results are based off what trees consistently showed improvements across all KPI’s used for the overall business problem analysis. We were looking for improvements from good to bad in all of our KPI’s and noticed that many times trees would improve but sometimes the percentage of say good health ginkgo trees would be less overall than poor health gingko trees. But a useful insight was to understand that the tree itself made a jump improvement from poor to fair health with a distinct increase of percentage for example. In that case we would still consider the tree to be improved for both cases because of the two instances of the ginkgo improving. Over all improvements between KPI’s, the trees with the most improvements will be selected, and these trees are the sophora at number one, the ginkgo for number two, the eastern redbud for number three, and the final spot belongs to the callery pear. The sophora and ginkgo both had three instances of improvement but the sophora was chosen for the number one spot because of its improvements was in stewardship, our main KPI. Currently, there should be no actions taken based off this analysis alone. More analysis must be done with the remaining possible KPI’s not used in this analysis. There are also real-world applications to planting trees in neighborhoods that should be considered as well and possibly even added to this dataset for future analysis. The next steps would then be to enact the above. Firstly, to give the current finding to the governor and update the governor on the analysis and get feedback. Secondly is to consider the consequences, good or bad, to implementing the findings now rather than what should be done before implementation which is testing the current KPI’s against each other for more insight and testing the remaining KPI’s. The innovative solution currently is to do surveys about this current tree, and as to what tree they like independently of the analysis, and then of the tree they like most of the recommended trees thus far.

|  |  |  |
| --- | --- | --- |
| Learning objectives | Activities/Tasks | How would you measure the reader’s learning |
| Changing variables to percentages for meaningful measures. | Create calculated fields that divides variables into its subcategories called sub-variables and then create percentages of those new fields. | The reader successfully created calculated fields with the correct naming conventions that then created percentages that truly represented the variable status. |
| Understand what improvement means via sub-variable statuses. | Through analysis create visualizations of each sub-variable to trees in the Hunt’s Point neighborhood and see how they improve via percentage and list movement. | Successfully creating analyses for each sub-variable KPI and visually determining what tree improved via that KPI. |
| Learn how to determine the tree or trees for replacement use. | Compare the trees in each analysis that improved in the list and count the instances that they improved via percentage or list movement. | Counting the amount of instances of improvement comparatively to the trees picked in the stewardship analysis and determining trees best suitable for replacement use. |

# Limitations

The limitations of this project are scattered throughout this tutorial as insights were discovered as analysis were made, but the main limitations that were consistent throughout the analysis are discussed below.

Firstly, the results are inconclusive, mostly because of accuracy issue from the dataset and how the analysis was conducted. Further implementations are required but because of that there may be some decisions that were made based off assumption that were not sound. For instance, why stewardship for the trees that improved to compare too? Sure, the reasons are there, but real-world application shows that instead of the tree the citizens like the most, its going to be the tree that will look the nicest last the longest and is the cheapest to get. The other big thing is the percentages. They often changed between visualizations to represent either the total percentage of only the trees selected, or the entirety of the trees in Hunt’s Point. Even if it was consistent (which we assume) it was not apparent and interpreting improvement was a challenge based off percentage. A better measure could be determined, or percentage could be altered or visually displayed better. There must have been a better way to deal with the overabundance of trees. Possibly normalizing the data or found a different decision to select trees. It seems that they could have been potential, and the opportunity cost was created when they could have just been analyzed like the rest of the trees if it were somehow made to work with the analyses better.

# References

1. Sparkes, Sam. “These Are The 10 Worst Neighborhoods In New York For 2019.” *RoadSnacks*, 12 Jan. 2020, [www.roadsnacks.net/worst-neighborhoods-in-new-york-city/](http://www.roadsnacks.net/worst-neighborhoods-in-new-york-city/) (Sparks, Sam)
2. York, City of New. “NY 2015 Street Tree Census - Tree Data.” Kaggle, 6 Dec. 2019, [www.kaggle.com/new-york-city/ny-2015-street-tree-census-tree-data](http://www.kaggle.com/new-york-city/ny-2015-street-tree-census-tree-data). (City of New York, Data)