










# Analyzing The Correlations Between Tree Characteristics And Their Surroundings

This manuscript ([permalink](#)) was automatically generated from [uiceds/cee-492-term-project-fall-2022-her@df91ac1](#) on September 22, 2022.

## Authors

---

- **Hadil Helaly**  
 [55555-55-5555](#) ·  [hadilhelaly](#) ·  [johndoe](#)  
Department of Civil and Environmental Engineering · Funded by Grant XXXXXXXX
- **Emma Golub**  
 [XXXX-XXXX-XXXX-XXXX](#) ·  [emmaagolub](#)  
Department of Civil and Environmental Engineering
- **Riley Blasiak**  
 [XXXX-XXXX-XXXX-XXXX](#) ·  [blasiak2](#)  
Department of Civil and Environmental Engineering
- **Rupesh Rokade**  
 [XXXX-XXXX-XXXX-XXXX](#) ·  [RupeshRokade16](#)  
Department of Civil and Environmental Engineering

## Introduction

---

The Urban tree database, which was collected by the US Forest Service Research Archive of the US Department of Agriculture, includes data about tree growth in urban areas across 17 cities and 13 states over the span of 14-years (from 1998-2012). The states included in the study are: Arizona, California, Colorado, Florida, Hawaii, Idaho, Indiana, Minnesota, New Mexico, New York, North Carolina, Oregon, and South Carolina. The data come from measurements taken to over 14,000 street and urban park trees, and the data can be obtained by downloading the 1.08 MB compressed “data publication” file from [Link](#). Some measurements of interest include tree age, location, height, crown diameter, leaf area, foliar biomass, and utility line interference. Tree age, for example, was determined from interviews with residents, street construction dates, aerial and historical photos, the city’s urban forester, and laboratory cores developed by the Lamont-Doherty Earth Observatory’s Tree Ring Laboratory.

The downloaded folder includes 9 data sheets in CSV format. The most interesting data files are 1. TS1\_Regional\_information.csv, 2. TS2\_Regional\_species\_and\_counts.csv, and 3. TS3\_Raw\_tree\_data.csv. First, the “TS1\_Regional\_information.csv” file contains information about region code, city, state, airport codes, and collection year. Second, the “TS2\_Regional\_species\_and\_counts.csv” file contains information (columns) regarding region, scientific and common names of trees, tree type, and 9 columns of dbh\_class, which represent a species diameter at breast height and are used to predict tree height, crown diameter, crown height, and leaf area. The file contains a total of 347 rows. Finally, the “TS3\_Raw\_tree\_data.csv” file includes 14487 observations (rows) of raw tree data. For each observation, 41 different variables were collected (columns). A detailed description of each of these 41 variables is as followed:

- DbaseID = Unique id number for each tree.

- Region = 16 U.S. climate regions, abbreviations are used.
- City = City/state names where data collected.
- Source = Original \*.xls filename (not available in this data publication).
- TreeID = Number assigned to each tree in inventory by city.
- Zone = Number/ID/name of the management area or zone that the tree is located in within a city; or nursery if young tree data collected there.
- Park/Street = Data listed as Park, Street, Regional Big Tree, or Nursery (for young tree measurements).
- SpCode = 4 to 6 letter code consisting of the first two letters of the genus name and the first two letters of the species name followed by two optional letters to distinguish two species with the same four-letter code (See \_Regional\_species\_and\_counts.csv for a list of the SpCodes and corresponding scientific names.)
- ScientificName = Botanical name of species.
- CommonName = Common name of species.
- Tree Type = 3 letter code where first two letters refer to life form (BD=broadleaf deciduous, BE=broadleaf evergreen, CE=coniferous evergreen, PE=palm evergreen) and the third letter is mature height (S=small which is < 8 meters, M=medium which is 8-15 meters, and L=large which is > 15 meters).
- Address = From inventory, street number of building where tree is located.
- Street = From inventory, the name of the street the tree is located on. (NOTE: zero values denote data were not recorded in that city. These values were left unchanged because they originated from city inventories.)
- Side = From inventory, side of building or lot tree is located on (F=front, M=median, S=side, P=park). (NOTE: zero values denote data were not recorded in that city. These values were left unchanged because they originated from city inventories.)
- Cell = From inventory, the cell number (i.e., 1, 2, 3, ...), where protocol determines the order trees at same address are numbered (e.g., driving direction or as street number increases).
- OnStreet = From inventory (omitted if not a field in city's inventory), for trees at corner addresses when tree is on cross street rather than addressed street. FromStreet = From inventory, the name of the first cross street that forms a boundary for trees lining un-addressed boulevards. Trees are typically numbered in order (1, 2, 3 ...) on boulevards that have no development adjacent to them, no obvious parcel addresses.
- ToStreet = From inventory, the name of the last cross street that forms a boundary for trees lining un-addressed boulevards.
- Age = Number of years since planted. (NOTE: zero values represent newly planted trees, < 1 year old.)
- DBH (cm) = Diameter at breast height (1.37 meters [m]) measured to nearest 0.1 centimeters (tape). For multi-stemmed trees forking below 1.37 m measured above the butt flare and below the point where the stem begins forking, as per protocol.
- TreeHt (m) = From ground level to tree top to nearest 0.5 m (omitting erratic leader).
- CrnBase (m) = Average distance between ground and lowest foliage layer to nearest 0.5 m (omitting erratic branch).
- CrnHt (m) = Calculated as TreeHT minus Crnbase to nearest 0.5 m. (NOTE: zero values indicate no live crown was present, hence no other tree dimension data were available.)
- CdiaPar (m) = Crown diameter measurement taken to the nearest 0.5 m parallel to the street (omitting erratic branch).
- CDiaPerp (m) = Crown diameter measurement taken to the nearest 0.5 m perpendicular to the street (omitting erratic branch).
- AvgCdia (m) = The average of crown diameter measured parallel and perpendicular to the street.
- Leaf (m<sup>2</sup>) = Estimated using digital imaging method to nearest 0.1 squared meter (m<sup>2</sup>).
- Setback = Distance from tree to nearest air-conditioned/heated space (may not be same address as tree location): 1=0-8 m, 2=8.1-12 m, 3=12.1-18 m, 4=> 18 m.
- TreeOr = Taken with compass, the coordinate of tree taken from imaginary lines extending from walls of the nearest conditioned space (may not be same address as tree location).

- CarShade = Number of parked automotive vehicles with some part under the tree's drip line. Car must be present (0=no autos, 1=1 auto, etc.).
- LandUse = Predominant land use type where tree is growing (1=single family residential, 2=multi-family residential [duplex, apartments, condos], 3=industrial/institutional/large commercial [schools, gov't, hospitals], 4=park/vacant/other [agric., unmanaged riparian areas of greenbelts], 5=small commercial [minimart, retail boutiques, etc.], 6=transportation corridor).
- Shape = Visual estimate of crown shape verified from each side with actual measured dimensions of crown height and average crown diameter (1=cylinder [maintains same crown diameter in top and bottom thirds of tree], 2=ellipsoid, the tree's center [whether vertical or horizontal is the widest, includes spherical], 3=paraboloid [widest in bottom third of crown], 4=upside down paraboloid [widest in top third of crown]).
- WireConf = Utility lines that interfere with or appear above tree (0=no lines, 1=present and no potential conflict, 2=present and conflicting, 3=present and potential for conflicting). (NOTE: -1 denotes data were not collected.)
- dbh1 = Dbh (centimeters [cm]) for multi-stemmed trees; for non-multi-stemmed trees, dbh1 is same as Dbh (cm).
- dbh2 = Dbh (cm) for second stem of multi-stemmed trees.
- dbh3 = Dbh (cm) for third stem of multi-stemmed trees.
- dbh4 = Dbh (cm) for fourth stem of multi-stemmed trees.
- dbh5 = Dbh (cm) for fifth stem of multi-stemmed trees.
- dbh6 = Dbh (cm) for sixth stem of multi-stemmed trees.
- dbh7 = Dbh (cm) for seventh stem of multi-stemmed trees.
- dbh8 = Dbh (cm) for eighth stem of multi-stemmed trees.

Additionally, a fourth data set may be of later interest for estimating leaf area, species dominance at a spatial scale, and carbon storage estimates. The TS5\_Foliar\_biomass\_leaf\_samples.csv contains urban foliar samples data by species for 17 U.S. cities. A total of 261 rows are provided.

The breadth of this dataset allows for a myriad of problems to be explored. The primary data that will be utilized for this project is the "TS3\_Raw\_tree\_data.csv" file, as this contains the most columns which will result in more feasible predictions during the machine learning portion of the project. This data can be used to analyze correlations between tree characteristics and their surroundings. One potential research question using the "TS3\_Raw\_tree\_data.csv" file is: how does utility line interference affect the growth of a certain type of tree in one state versus a different state. the preliminary 14 variables that can be used in the proposed analysis include "Address", "Age", "Shape", "WireConf", "Setback", "CarShade", "DBH", "TreeHt", "CrnBas", "CrnHt", "CdiaPar", "CDiaPerp", "AvgCdia", "Leaf".

After tidying the dataset, we can compare the effect of the WireConf, Setback, CarShade on the remaining variables of similar trees. Since we also contain the addresses of the trees, along with visualizing graphs from results of the comparisons, we can create maps to understand the variance of these effects across different cities. Further, a machine learning model can be created to possibly target and predict The above results for a city that is not mentioned in the dataset Predict the missing values in the dataset.

## References

---