Branching Out: Tree Health and Climate Equity Across DC Wards

Ashley Totten, Max Harris, Meg Norten

May 8, 2025

Introduction

Before conducting our analysis, we conducted background research to better understand the context of our issue. By understanding what causes changes in tree coverage and how this can have negative socioeconomic impact, we can have a different perspective to view our findings and how this is significant.

The first article we examined was "Disparities in Urban Neighborhood Conditions: Evidence from GIS Measures and Field Observation in New York City". This article was published in the Journal of Public Health Policy, and aimed to examine how different neighborhood conditions affected public health outcomes. The article used GIS, which is a tool that utilizes satellite imagery, as well as field work to reach their conclusions. The researchers realized that while many urban areas have high scores in traditional walk ability metrics, the health outcomes such as rate of diabetes and heart disease did not correlate. They found that the absence of street trees contributed heavily to this, reinforcing the importance of trees in tandem with other walk ability features to improve neighborhood health.

The second article we utilized was "Tree canopy change and neighborhood stability: A comparative analysis of Washington, D.C. and Baltimore, MD". This article compared the two mid Atlantic cities, ensuring that climate and region were not confounding variables, and evaluated how tree canopy change corresponds with neighborhood stability. Neighborhood stability is a metric that scores how much a neighborhood changes in terms of wealth, either positively or negatively. In the article, the researchers found that while wealth was not a strong predictor of tree canopy in Baltimore, it was in Washington, D.C. This shows that there are systems in place in D.C. causing the disparity that we aim to evaluate.

Finally, we analyzed "Benefit-Cost Analysis of Modesto's Municipal Urban Forest". This article published in the Journal of Arboriculture is a case study on the City of Modesto in California. It analyzes whether the spending by the city is justifiable from a purely economic perspective. The researchers tallied total expenditure by the city on its urban canopy and

then compared to the economic benefits achieved by this expenditure, particularly in home value. In total, Modesto spent $2.6 million and had a return of $4.95 million. This nearly 2:1 ratio demonstrates the utility of an urban canopy and how its benefits are key to ensuring economic growth.

Data Breakdown

```
suppressMessages(library(tidyverse))
suppressMessages(library(janitor))
suppressMessages(library(ggplot2))
suppressMessages(library(RColorBrewer))
suppressMessages(library(DescTools))
suppressMessages(library(patchwork))
suppressMessages(library(gmodels))
Tree_data <- read_csv("street_tree_archive.csv", show_col_types = FALSE)
tree_clean <- clean_names(Tree_data)
```

```
nrow(tree_clean)
```

```
[1] 1861805
```

```
ncol(tree_clean)
```

```
[1] 12
```

```
colnames(tree_clean)
```

```
 [1] "objectid"   "tbox_stat"  "dbh"        "condition"  "genus_name"
 [6] "year"       "ward_id"    "anc_id"     "smd_id"     "sci_nm"
[11] "cmmn_nm"    "globalid"
```

The dataset we utilized for our research was the Street Tree Archive Dataset, part of the Open Data DC library. This dataset contains information about trees across the city of Washington from 2014 to the present. Each row represents a tree, and each column is a different characteristic of the tree. A thing to note is that there is no ID variable for each tree, so the trees cannot be tracked over time. There are 1.8 million total observations, meaning the best guess for the total number of trees being measured is 180,000. Each tree still has a significant amount of information identified. Firstly, the common and scientific name of the

species of tree as well as the genus it falls in. Next, the status of the tree box was recorded, whether it is open, there are plants at the bottom, or there is conflict with another tree. The year of the observation is also recorded as a date variable. Additionally, the size of the tree, determined using the common measure diameter at breast height, is the only quantitative variable present. The location of each tree was determined using three different categorical variables, Ward, Advisory Neighborhood Commission, and Single Member District. Ward is the most general of the three, with only 8 wards in D.C., leading us to select ward as the location variable we utilized. We focused primarily on the condition of the tree, defined as "Dead", "Poor", "Fair", "Good", or "Excellent". The D.C. Forest Service division within the D.C Department of Transportation created and maintains the dataset, which has the primary goal of showing annual changes in species composition and tree health.

Research Goals and Questions

Our initial research goals were to investigate the changes in the proportion of "Good" and "Excellent" trees in DC from 2014 to 2024 and to pinpoint any differences in tree condition by ward in the year 2024. Going off of these goals, we formed the following research questions:

- To what extent are there differences in the condition of trees between wards? Are there greater proportions of "Good" and "Excellent" trees in certain wards compared to other?
- To what extent are there differences in the condition of trees between 2014 and 2024? Has the proportion of "Good" and "Excellent" trees increased from 2014 to 2024 under UFD management?

Initial Hypotheses

We hypothesized that there would be differences in the condition of trees from 2014 to 2024, and specifically, that the proportions of "Good" and "Excellent" trees would differ over this time period. We also predicted that with proper UFD management, the proportions of "Good" and "Excellent" trees in DC would increase over time. We also hypothesized that in the year 2024, there would be differences in the condition of trees and proportion of "Good" and "Excellent" trees between wards. We did not form a hypothesis about which wards would have larger proportions than others, although finding indicators that may predict these proportions by ward could be the basis for interesting research in the future.

Exploratory Data Analysis

Data Cleaning

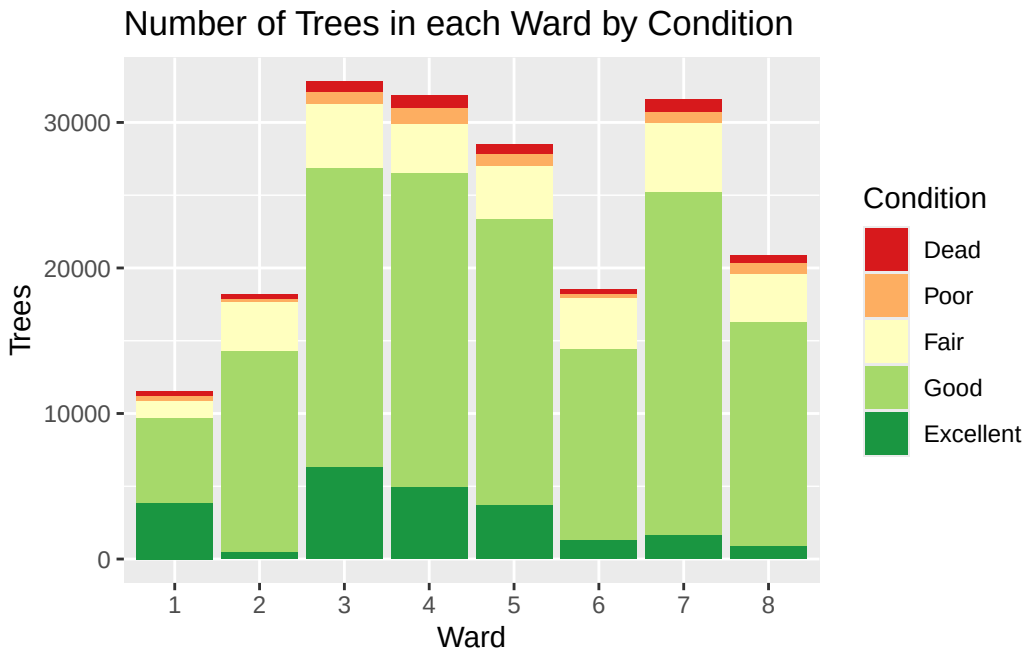
As part of our data cleaning, we removed unnecessary variables, including object id, single member district id, the genus name, and the global id. We also converted ward id, condition,

and year to factors so they were easier to work with and ordered the levels of condition, so it was easier to read and analyze in the graphs. In order to do the later tests, we also made a data set that was just for the years 2014 and 2024. To compare the proportion of trees by condition in each ward, we filtered a separate data set to only include the year 2024. Then, we counted the number of trees in the 2024 data set in each ward that were either "Good" or "Excellent", which were stored in one data set and "Dead" and "Poor", which were stored in another data set. Then, using the 2024 data set, we counted the number of total trees in each ward. These counts were then used to calculate the proportion of "Good" and "Excellent" and "Dead" and "Poor" trees in each ward by dividing the number of "Good" and "Excellent" trees in each ward by the total number of trees in each ward, repeating this process for "Dead" and "Poor" trees.

```
tree_clean2 <- tree_clean %>%
  mutate(objectid = NULL, smd_id = NULL, genus_name = NULL, globalid = NULL) %>%
  mutate(tbox_stat = as.factor(tbox_stat), ward_id = as.factor(ward_id),
         condition = as.factor(condition), anc_id = as.factor(anc_id),
         year = as.factor(year))
```

```
tree_clean3 <- tree_clean2 %>%
  filter(!is.na(condition)) %>%
  filter(!is.na(ward_id)) %>%
  filter(year == "2024") %>%
  mutate(condition = factor(condition, levels = c("Dead", "Poor", "Fair", "Good", "Excellent"

ggplot(aes(x = ward_id, fill = condition), data = tree_clean3) +
  geom_bar()+
  labs(x = "Ward", y = "Trees", title = "Number of Trees in each Ward by Condition", fill = "
  scale_fill_brewer(palette = "RdYlGn")
```



This bar graph displays the number of trees in each ward by condition. The condition variable on the side assigns "Dead" as red, "Poor" as orange, "Fair" as yellow, and "Good" and "Excellent" as different shades of green. On the x-axis, "Ward" is a factor variable, where each ward has its own category. The number of trees are on the y-axis. Wards 3,4, and 7 all have over 30,000 trees, while ward 1 has the least with a little over 10,000 trees. All of the wards have the most trees that are "Good" compared to any other condition. Wards 4 and 7 have the most "Dead" trees, and Ward 4 also has the most "Excellent" trees.

```
tree_sum_dead_poor <- tree_clean3 %>%
  filter(condition == "Dead"| condition == "Poor") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_sum_good_excel <- tree_clean3 %>%
  filter(condition == "Good"| condition == "Excellent") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_total_ward <- tree_clean3 %>%
```

```
  count(ward_id) %>%
  rename(total_trees = n)

tree_proportions <- tree_total_ward %>%
  mutate(dead_poor_trees = tree_sum_dead_poor$sum) %>%
  mutate(good_excel_trees = tree_sum_good_excel$sum) %>%
  mutate(dead_poor = dead_poor_trees / total_trees) %>%
  mutate(good_excel = good_excel_trees / total_trees)
```

```
tree_sum_dead_poor <- tree_clean3 %>%
  filter(condition == "Dead"| condition == "Poor") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_sum_dead_poor <- tree_clean3 %>%
  filter(condition == "Good"| condition == "Excellent") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_total_ward <- tree_clean3 %>%
  count(ward_id) %>%
  rename(total_trees = n)
```

```
tree_clean3 <- tree_clean3 |>
  mutate(
    good_excellent = condition %in% c("Good", "Excellent")
    ) |>
  group_by(ward_id) |>
  mutate(
    total_trees = n(),
    total_good_excellent = sum(good_excellent),
    prop_good_excellent = total_good_excellent/total_trees
  ) |>
  ungroup()
```

```
tree_clean4 <- tree_clean3 |>
  mutate(
```

```
    dead_poor = condition %in% c("Poor", "Dead")
    ) |>
  group_by(ward_id) |>
  mutate(
    total_trees = n(),
    total_dead_poor = sum(dead_poor),
    prop_dead_poor = total_dead_poor/total_trees
  ) |>
  ungroup()
```

```
tree_clean5 <- tree_clean2 %>%
  filter(!is.na(condition)) %>%
  filter(!is.na(ward_id)) %>%
  mutate(condition = factor(condition, levels = c("Dead", "Poor", "Fair", "Good", "Excellent"

tree_sum_dead_poor <- tree_clean5 %>%
  filter(condition == "Dead"| condition == "Poor") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_sum_dead_poor <- tree_clean5 %>%
  filter(condition == "Good"| condition == "Excellent") %>%
  count(condition, .by = ward_id) %>%
  mutate(sum = sum(n), .by = ".by") %>%
  distinct(.by, sum) %>%
  rename(ward_id = .by)

tree_total_ward <- tree_clean5 %>%
  count(ward_id) %>%
  rename(total_trees = n)
```

```
tree_clean5 <- tree_clean5 |>
  mutate(
    good_excellent = condition %in% c("Good", "Excellent")
    ) |>
  group_by(ward_id) |>
  mutate(
    total_trees = n(),
    total_good_excellent = sum(good_excellent),
    prop_good_excellent = total_good_excellent/total_trees
```

```
  ) |>
  ungroup()
```

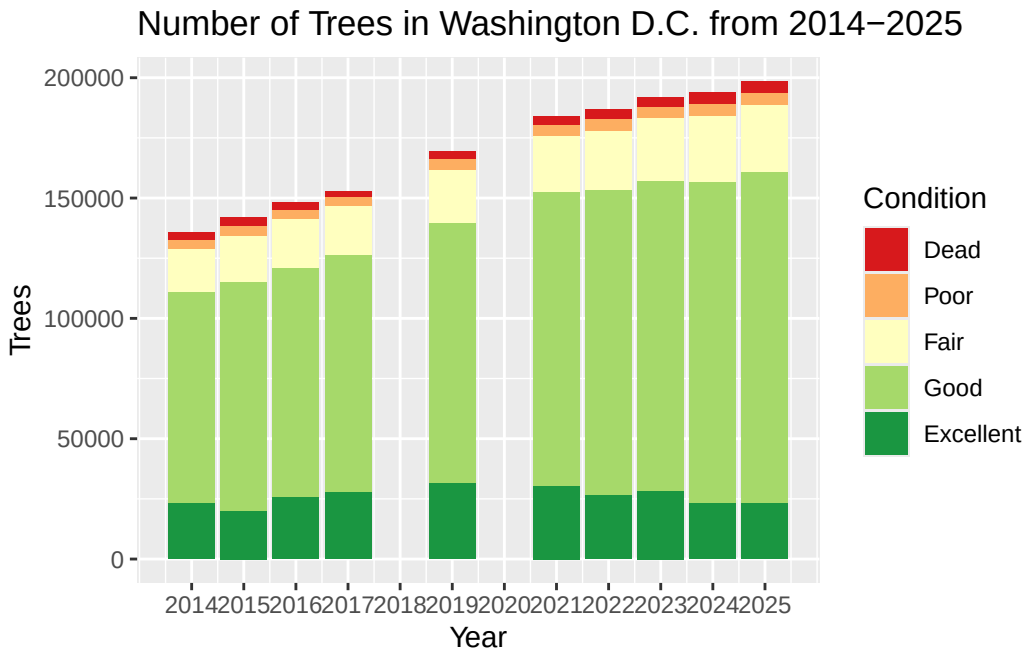
```
tree_clean5 <- tree_clean5 |>
  mutate(
    dead_poor = condition %in% c("Poor", "Dead")
    ) |>
  group_by(ward_id) |>
  mutate(
    total_trees = n(),
    total_dead_poor = sum(dead_poor),
    prop_dead_poor = total_dead_poor/total_trees
  ) |>
  ungroup()
```

```
tree_clean6 <- tree_clean5 %>%
  mutate(year = factor(year, levels=c('2024', '2014')))
```

```
tree_clean7 <- tree_clean %>%
  mutate(objectid = NULL, smd_id = NULL, genus_name = NULL, globalid = NULL) %>%
  mutate(tbox_stat = as.factor(tbox_stat), ward_id = as.factor(ward_id),
         condition = as.factor(condition))

tree_clean8 <- tree_clean7 %>%
  filter(!is.na(condition)) %>%
  filter(!is.na(ward_id)) %>%
  mutate(condition = factor(condition, levels = c("Dead", "Poor", "Fair", "Good", "Excellent
```

```
ggplot(aes(x = year, fill = condition), data = tree_clean8)+
  geom_bar()+
  scale_x_continuous(breaks = c(2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 
  scale_fill_brewer(palette = "RdYlGn")+
  labs(x = "Year", y = "Trees", fill = "Condition", title = "Number of Trees in Washington D
```

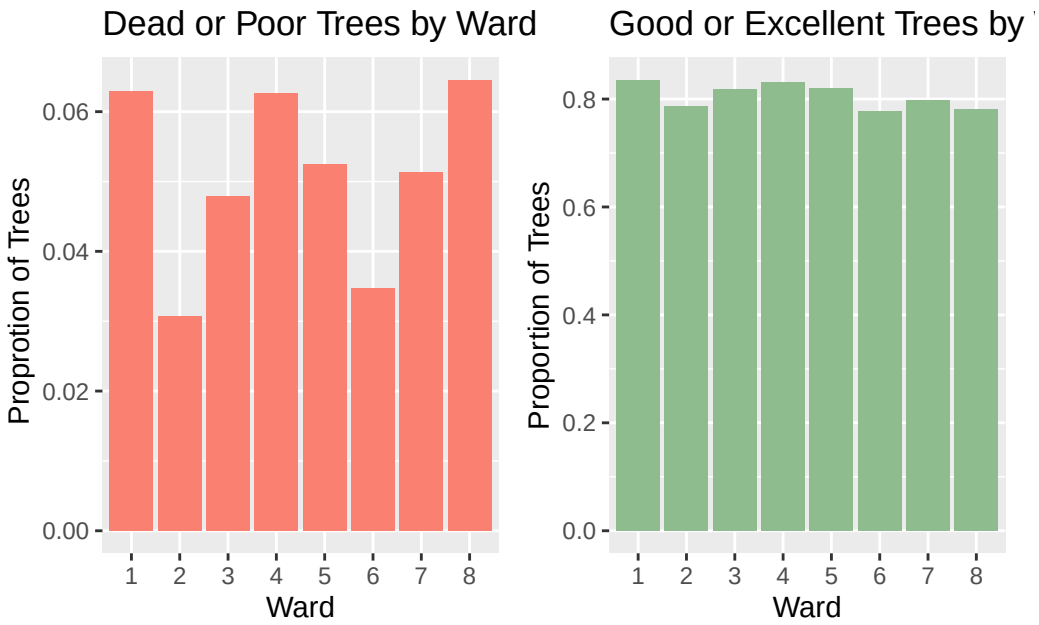


This bar growth shows the number of trees in Washington DC from 2014-2025. On the x-axis, "Year" is a continuous variable, so that the graph highlights the gaps in the data. There is no data collected in 2018 or 2020. The number of trees in DC increased steadily from 2014-2025. This could be because more trees were planted, or it could be because more trees were being surveyed by the data collectors. The number of "Dead" and "Poor" trees seems to be consistent from 2021-2025 if not a slight increase. There is an increase in "Good" trees from 2021-2025, but a decrease, overall, in "Excellent" trees.

```
p1<- ggplot(aes(x = ward_id, y = dead_poor), data = tree_proportions)+
  geom_col(fill = "salmon")+
  labs(x = "Ward", y = " Proprotion of Trees", title = "Dead or Poor Trees by Ward")

p2<- ggplot(aes(x = ward_id, y = good_excel), data = tree_proportions)+
  geom_col(fill = "darkseagreen")+
  labs(x = "Ward", y = "Proportion of Trees", title = "Good or Excellent Trees by Ward")

p1 + p2
```



This bar graph displays the proportion of trees that are "Dead" and "Poor" in each ward and the proportion of "Good" and "Excellent" trees in each ward. The two graphs have different scales. The "Dead" and "Poor" trees graph goes up to 7% proportion of those trees in each ward whereas the graph of the "Good" and "Excellent" trees goes up to 90% proportion of these trees in each ward. Wards 1,4, and 8 all have more than 6% of their trees as "Dead" or "Poor". Wards 2 and 6 have the least amount of "Dead" and "Poor" trees at around 3% and 3.5%. All of the wards have very similar proportions of "Good" and "Excellent" trees with all of them being around 80%. Ward 1 has the highest proportion, while Wards 6 and 8 have the lowest.

We also created a numerical summary of the number and proportions of "Good" and "Excellent" and "Dead" and "Poor" trees in each ward in 2024. This allowed us to understand the specific values associated with each proportion and to see how the proportions of "Good" and "Excellent" trees compared to the proportions for "Dead" and "Poor" trees.

```
tree_proportions |>
  group_by(ward_id) |>
  summarise(
    num_good_excel = good_excel_trees,
    num_dead_poor = dead_poor_trees,
    prop_good_excel = good_excel,
    prop_dead_poor = dead_poor
  )
```

```
# A tibble: 8 x 5
  ward_id num_good_excel num_dead_poor prop_good_excel prop_dead_poor
  <fct>            <int>         <int>           <dbl>          <dbl>
1 1                 9634           726           0.835         0.0630
2 2                14296           559           0.787         0.0308
3 3                26831          1573           0.818         0.0479
4 4                26496          1996           0.831         0.0626
5 5                23383          1496           0.821         0.0525
6 6                14419           645           0.778         0.0348
7 7                25204          1619           0.798         0.0513
8 8                16296          1348           0.780         0.0645
```

Upon creating our table, we noticed that Ward 1 had the largest proportion of "Good" and "Excellent" trees with 0.835 while Ward 6 had the smallest proportion with 0.778. We also observed that the ward with the smallest proportion of "Good" and "Excellent" trees—Ward 6, still far exceeded the largest proportion of "Dead" and "Poor" trees in Ward 8 with 0.0645. Overall, we found that there were noticeably larger proportions of "Good" and "Excellent" trees across all wards compared to the "Dead" and "Poor" trees.

We also conducted a simple Analysis of Variance F-test to see if there were significant differences in the diameter of trees between the tree conditions.

```
anova_diam <- aov(dbh~condition, data=tree_clean4)
anova(anova_diam)
```

```
Analysis of Variance Table

Response: dbh
              Df   Sum Sq Mean Sq F value    Pr(>F)
condition      4  1200145  300036  3464.1 < 2.2e-16 ***
Residuals 193909 16795154      87
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After running our ANOVA model, we found there was convincing evidence to suggest that at least one of the tree conditions' mean diameters significantly differed from the others (p-value < 0.0001).

```
PostHocTest(anova_diam, method = "lsd")
```

```
  Posthoc multiple comparisons of means : Fisher LSD
    95% family-wise confidence level

$condition
                      diff     lwr.ci      upr.ci   pval
Poor-Dead         4.045395   3.679533   4.4112563 <2e-16 ***
Fair-Dead         2.321300   2.034939   2.6076617 <2e-16 ***
Good-Dead        -1.267089  -1.536063  -0.9981154 <2e-16 ***
Excellent-Dead   -6.762202  -7.052633  -6.4717711 <2e-16 ***
Fair-Poor        -1.724095  -2.000038  -1.4481514 <2e-16 ***
Good-Poor        -5.312484  -5.570338  -5.0546301 <2e-16 ***
Excellent-Poor  -10.807596 -11.087760 -10.5274328 <2e-16 ***
Good-Fair        -3.588389  -3.709372  -3.4674066 <2e-16 ***
Excellent-Fair   -9.083502  -9.246719  -8.9202846 <2e-16 ***
Excellent-Good   -5.495112  -5.625435  -5.3647902 <2e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After running our Posthoc test to compare the mean diameter in each tree condition group to each other, we found that all of the condition-to-condition comparisons yielded significantly different results, indicating all of the mean diameters in each condition group were significantly different from the other groups (p-value < 0.0001). Interestingly, we noticed that "Poor" and "Fair" trees had significantly larger diameters than "Dead" trees with a positive difference between the two, while "Good" and "Excellent" trees had significantly smaller diameters than "Dead" trees with a negative difference. We believe this might be the case since new saplings that are planted might be in "Good" or "Excellent" condition and have smaller diameters at first, making these average diameters smaller than the "Dead" diameters. This could be a basis for further research in the future.

Statistical Methods, Modeling, and Results

To study our main research questions, we used chi-squared tests to explore associations between Tree Condition, Ward ID, and Year. We employed this technique since our main variables of interest were all categorical. Additionally, there were large expected values for the number of trees within each condition when grouped by both Ward ID and Year, no individual tree was counted in more than one cell in our contingency table, and our data is likely representative of the larger population of trees in DC (there is data for around 180,000 trees each year which is less than 10% of the total number of trees in DC). For the purposes of our study we will assume that the trees studied are independent of each other; however, this is a complicated assumption to make and might not hold true in practice when considering the biology of trees.

Since our variables are categorical, there are large expected values, and we have mutually exclusive and representative data, we conducted a chi-squared test with the following two sets of hypotheses.

Chi Square Test on Difference of Tree Condition by Ward:

- H0: There is no association between between tree condition and ward
- H1: There is an association between tree condition and ward

Chi Square Test on Difference of Tree Condition between 2014 and 2024:

- H0: There is no association between tree condition and year
- H1: There is an association between tree condition and year

Table of Counts Between Ward ID and Tree Condition

We began by creating a table of counts and expected values between Ward ID and Tree Condition. The table of expected counts displays the number of trees we would expect to find in each condition within each ward if there was no association between Tree Condition and Ward ID. Importantly, all of our expected counts are large (greater than 5) and we can see that in each ward, some of the observed counts of trees in each condition are above what was expected with no association while others are below the expected count.

```
# Create a data frame from the main data set.
tree_data_chi = data.frame(tree_clean4$ward_id,tree_clean4$condition)

# Create a contingency table with the needed variables.
tree_data_chi = table(tree_clean4$ward_id,tree_clean4$condition)

print(tree_data_chi)
```

```
   Dead  Poor  Fair  Good Excellent
1   338   388  1171  5782      3852
2   320   239  3320 13846       450
3   721   852  4402 20571      6260
4   956  1040  3408 21559      4937
5   691   805  3608 19711      3672
6   328   317  3478 13152      1267
7   865   754  4766 23559      1645
8   544   804  3240 15422       874
```

Table of Expected Values Between Ward ID and Tree Condition

```
chi.out <- chisq.test(tree_data_chi, correct=F)
chi.out$expected
```

```
       Dead     Poor     Fair      Good Excellent
1 283.2294 309.1560 1628.911  7944.577  1365.127
2 446.4223 487.2873 2567.467 12522.130  2151.693
3 805.7952 879.5569 4634.295 22602.531  3883.821
4 783.5417 855.2663 4506.311 21978.319  3776.562
5 699.7101 763.7608 4024.178 19626.846  3372.506
6 455.4367 497.1269 2619.311 12774.984  2195.142
7 775.9028 846.9281 4462.378 21764.048  3739.744
8 512.9619 559.9179 2950.150 14388.565  2472.405
```

We then used our table of counts to run a chi-squared test between Ward ID and Tree Condition in the year 2024.

```
chi.out
```

```
    Pearson's Chi-squared test

data:  tree_data_chi
X-squared = 12980, df = 28, p-value < 2.2e-16
```

Upon running our test, we found a statistically significant association between Ward ID and Tree Condition, that is, the distribution of tree conditions varies significantly across different wards (p-value < 0.0001).

Pairwise Proportion Test

We wanted to gain a deeper understanding of the extent to which the specific proportions of "Good" and "Excellent" trees varied across wards. To address this question, we conducted a pairwise proportion test which tests the significance of differences in proportions of "Good" and "Excellent" trees between each ward. We also used the Bonferroni correction to minimize the likelihood of a Type I error since we conducted multiple comparisons between wards.

```r
# Compare proportion of 'Good' and 'Excellent' trees in each ward
good_excel_trees <- tree_clean4[tree_clean4$condition %in% c("Good", "Excellent"),]
ward_table <- data.frame(table(good_excel_trees$ward_id))

# Compare this to all trees per ward to get proportions
total_trees <- data.frame(table(tree_clean4$ward_id))
```

```r
pairwise_prop <- pairwise.prop.test(x = ward_table$Freq, n = total_trees$Freq, p.adjust.metho
pairwise_prop
```

```
    Pairwise comparisons using Pairwise comparison of proportions

data:  ward_table$Freq out of total_trees$Freq

  1       2       3       4       5       6       7
2 < 2e-16 -       -       -       -       -       -
3 0.00061 3.2e-16 -       -       -       -       -
4 1.00000 < 2e-16 0.00062 -       -       -       -
5 0.01401 < 2e-16 1.00000 0.04588 -       -       -
6 < 2e-16 1.00000 < 2e-16 < 2e-16 < 2e-16 -       -
7 < 2e-16 0.07872 3.5e-09 < 2e-16 2.7e-11 2.3e-06 -
8 < 2e-16 1.00000 < 2e-16 < 2e-16 < 2e-16 1.00000 3.8e-05

P value adjustment method: bonferroni
```

Upon gathering our output, we discovered that most ward to ward comparisons yielded significantly different results, suggesting that the proportion of "Good" and "Excellent" trees significantly differs between most wards (p-value < 0.05). Specifically, five ward to ward comparisons did not show a significant difference in the proportion of "Good" and "Excellent" trees at the five percent significance level. When looking closer at these wards, Wards 1 and 4, 2 and 6, and 6 and 8 are all neighboring wards. We believe the proximity of these wards to each other could contribute to them having similar proportions of "Good" and "Excellent" trees if they have similar land, building usage, and proximity to water.

Proportion Table

```r
# Proportion of Good + Excellent per ward - include count
prop_table <- prop.table(tree_data_chi, margin = 1)
```

```
combined_props <- rowSums(prop_table[, c("Good", "Excellent")])
round(combined_props, 3)
```

```
    1     2     3     4     5     6     7     8
0.835 0.787 0.818 0.831 0.821 0.778 0.798 0.780
```

To verify the results of our pairwise proportion test, we created a proportion table to visually display the similarities in proportions across wards. This table verifies our results, as the five wards with insignificant differences are the wards with the most similar proportions of "Good" and "Excellent" trees in our proportion table.

Table of Counts Between Year and Tree Condition

Before running our chi-squared test between Year and Tree Condition, we created a table of counts and expected values between these variables. Once again, all of our expected counts were large (greater than 5), and within each year, some of the observed counts of trees in each condition fell below what was expected with no association while others were above the expected count of trees.

```
# Create a data frame from the main data set.
tree_data_chi_year = data.frame(tree_clean6$year,tree_clean6$condition)

# Create a contingency table with the needed variables.
tree_data_chi_year = table(tree_clean6$year,tree_clean6$condition)

print(tree_data_chi_year)
```

```
      Dead  Poor  Fair  Good Excellent
2024  4763  5199 27393 133602     22957
2014  3156  3852 17704  87794     23113
```

Table of Expected Values Between Year and Tree Condition

```
chi.out.year <- chisq.test(tree_data_chi_year, correct=F)
chi.out.year$expected
```

```
          Dead     Poor     Fair      Good Excellent
2024 4659.943 5326.069 26537.37 130280.68  27109.93
2014 3259.057 3724.931 18559.63  91115.32  18960.07
```

We repeated our previous methodology and used the table of counts to run a two-by-two chi-squared test between Ward ID and Year using only the years 2014 and 2024.

```
chi.out.year
```

```
    Pearson's Chi-squared test

data:  tree_data_chi_year
X-squared = 1831.5, df = 4, p-value < 2.2e-16
```

Upon gathering our output, we found a statistically significant association between year and tree condition. That is, the distribution of tree conditions varies significantly between 2014 and 2024 (p-value < 0.0001).

Pairwise Proportion Test Output

To further explore the extent to which the proportions of "Good" and "Excellent" trees varied between 2014 and 2024, we conducted another pairwise proportion test to assess the significance of differences in proportions of "Good" and "Excellent" trees between these years.

```
# Compare proportion of 'Good' and 'Excellent' trees in each ward
good_excel_trees_year <- tree_clean6[tree_clean6$condition %in% c("Good", "Excellent"),]
year_table <- data.frame(table(good_excel_trees_year$year))

# Compare this to all trees per ward to get proportions
total_trees_year <- data.frame(table(tree_clean6$year))
```

```
pairwise.prop.test(x = year_table$Freq, n = total_trees_year$Freq, p.adjust.method = "bonferr
```

```
    Pairwise comparisons using Pairwise comparison of proportions

data:  year_table$Freq out of total_trees_year$Freq
```

```
  1
2 5.3e-14

P value adjustment method: bonferroni
```

Upon gathering our output, there was evidence to suggest the proportion of "Good" and "Excellent" trees significantly differed between 2014 and 2024 where 1 represents 2024 and 2 represents 2014. We found these results unsurprising as over time, we would expect trees to age and potentially drop to worse conditions. On the other hand, the proportion of "Good" and "Excellent" trees could have significantly increased over time if UFD management properly maintained the trees.

Proportion Table

To further explore these potential differences in proportion, we created a proportion table to understand the direction of change over time.

```
# Proportion of Good + Excellent by year
prop_table_year <- prop.table(tree_data_chi_year, margin = 1)
combined_props_year <- rowSums(prop_table_year[, c("Good", "Excellent")])
round(combined_props_year, 3)
```

```
 2024  2014
0.807 0.818
```

Upon creating our proportion table, it appears that the proportion of "Good" and "Excellent" trees in 2024 is 0.011 lower than the proportion in 2014. This small, yet significant difference in the proportion of "Good" and "Excellent" trees indicates that over time, the proportion of these trees significantly decreased. As previously alluded to, this change could be attributed to trees aging, especially considering that the same trees were supposed to be studied over time in this data set. It is also important to recognize that our earlier bar graphs indicated that over time, the total number of trees being studied overall seemed to increase from 2014 to 2024. If more trees were planted each year, an increase in the total number of trees over time could also reduce the proportion of "Good" and "Excellent" trees even if the number of "Good" and "Excellent" trees remains constant.

Data-driven Questions

Some questions for future research are as follows: - Is there a relationship between the proportion of dead and poor or good and excellent trees in each ward and the income of the ward?

More specifically, do higher income ward have a higher proportion of good and excellent trees? Do lower income wards have a higher proportion of dead and poor trees? - Is there a relationship between the usage of the land (urban, suburban, or green space) and the condition of the trees in each ward? - Is there any relationship between population density and tree conditions in each ward?

These questions all seek to find explanations for the differences in tree conditions by ward that were indicated in our data analysis. Since we have shown that there are small, yet significant differences in the proportion of "Good" and "Excellent" trees between most wards and across time, we want to investigate the factors that contribute to these differences. While we are unable to find causal relationships with our observational studies, finding associations between different elements of DC and the tree conditions is of interest to our team.

Discussion

We were able to detect small, yet significant differences in the proportion of "Good" and "Excellent" trees between wards and years. While as a whole the proportion of "Good" and "Excellent" trees is similar across wards, there are still statistically significant differences across wards. The fact that only 5 ward pairings have similar values shows the discrepancy across different areas of the city, and reflects that there is a significant difference in the way that DCFS is planting and maintaining trees.

These findings are consistent to our takeaways from the literature. Our sources discussed how wealth can be correlated with tree canopy, and how poorer neighborhoods are more strongly affected. Additionally, the health consequences of this discrepancy are profound, and more awareness should be raised to the public health ramifications of unequal tree planting. Our research has shown that D.C. does have a discrepancy in the condition of trees between wards, and reinforces the importance of further researching the local implications of these differences.

Moving forward, something we discussed that would fortify our research is adding a wealth element to the analysis. While we discussed the impact of wealth on canopy and attempted to generate those takeaways anecdotally, incorporating another dataset that has wealth and income information for the wards of D.C. would be an impactful way to solidify our findings and ensure our research can be applied in the future.

References

Chuang, Wen-Ching; Boone, Christopher G.; Locke, Dexter H.; Grove, J. Morgan; Whitmer, Ali; Buckley, Geoffrey; Zhang, Sainan. 2017. Tree canopy change and neighborhood stability: A comparative analysis of Washington, D.C. and Baltimore, MD. *Urban Forestry & Urban Greening.* 27: 363-372. https://doi.org/10.1016/j.ufug.2017.03.030.

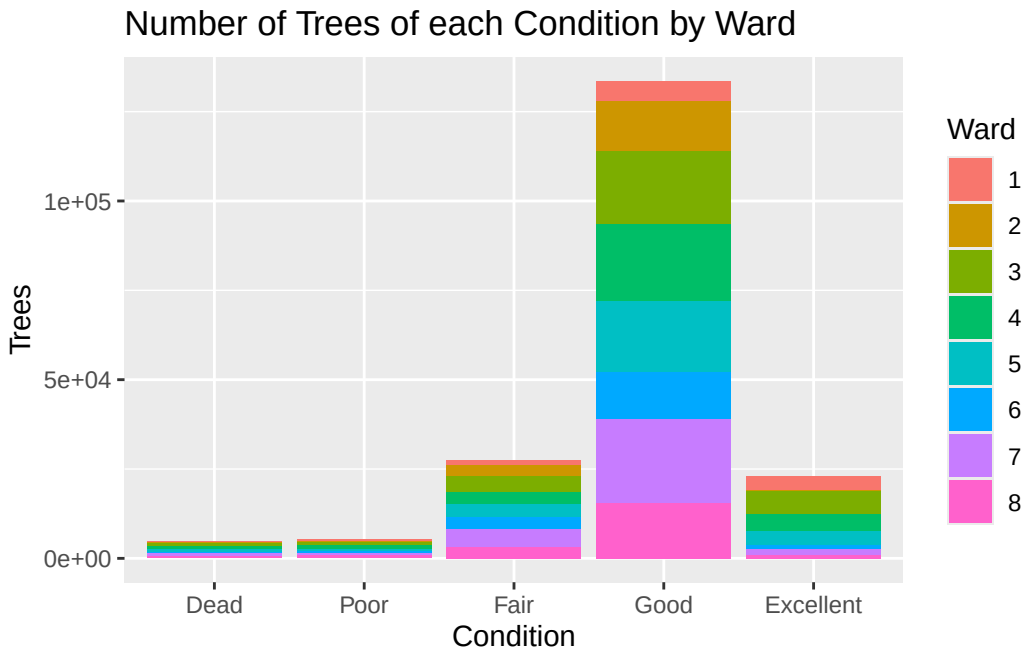
McPherson, G. E., Simpson, J. R., Peper, P. J., & Xiao, Q. 1999. Benefit-Cost Analysis of Modesto's Municipal Urban Forest. *Journal of Arboriculture*, *25*(5), 235–248. https://doi.org/10.48044/jauf.1999.033

Neckerman, K. M., Lovasi, G. S., Davies, S., Purciel, M., Quinn, J., Feder, E., Raghunath, N., Wasserman, B., & Rundle, A. 2009. Disparities in Urban Neighborhood Conditions: Evidence from GIS Measures and Field Observation in New York City. *Journal of Public Health Policy*, 30, S264–S285. http://www.jstor.org/stable/40207263

U.S. Environmental Protection Agency. 2008. Trees and Vegetation. In: Reducing Urban Heat Islands: Compendium of Strategies. Draft. https://www.epa.gov/heat-islands/heat-island-compendium.

Appendix

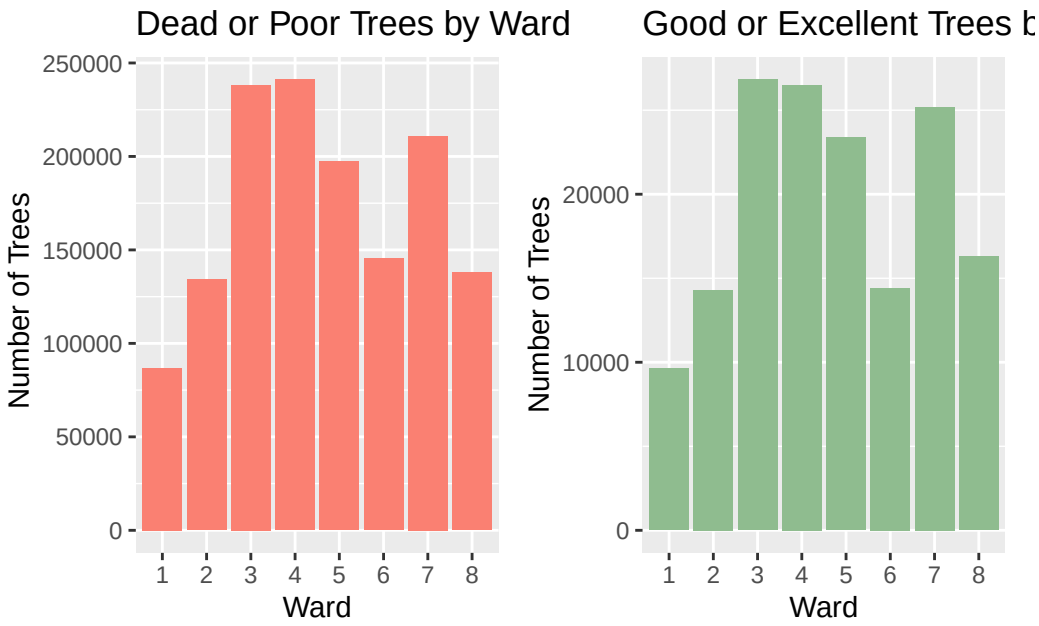
```
ggplot(aes(x = condition, fill = ward_id), data = tree_clean3)+
  geom_bar()+
  scale_x_discrete(limits = c("Dead", "Poor", "Fair", "Good", "Excellent"))+
  labs(x = "Condition", y = "Trees", title = "Number of Trees of each Condition by Ward", fi
```



```
p1<- ggplot(aes(x = ward_id, y = sum), data = tree_sum_dead_poor)+
  geom_col(fill = "salmon")+
  labs(x = "Ward", y = " Number of Trees", title = "Dead or Poor Trees by Ward")

p2<- ggplot(aes(x = ward_id, y = sum), data = tree_sum_good_excel)+
  geom_col(fill = "darkseagreen")+
  labs(x = "Ward", y = "Number of Trees", title = "Good or Excellent Trees by Ward")

p1 + p2
```



Data Used

The Street Tree Archive data was pulled from Open Data DC and collected by DC District Department of Transportation's (DDOT) Urban Forestry Division (UFD). It can be found by going to opendata.dc.gov, scrolling down to "Search here for more data and apps..." on the homepage, and typing "Street Tree Archive". From here, an overview of the data is displayed on the page. To download the data, click the "Download" box for ease of use in future research. The direct link to the webpage with the data to download on Open Data DC is provided below.

Data: Street Tree Archive