DATA 602 Project: Exploring the City of Calgary's Public Trees Dataset

Harneet Cheema, Michael Ellsworth

Introduction

- Calgary has ~7 million trees
- Calgary should not have ~7 million trees
- Maintaining Public Trees is a costly endeavour
- How could Calgary's tree population improve?
- What insights can we gain from public tree data?

Dataset

- Calgary has data on ~500,000 trees
- 20 different features
- Analysis on tree condition
    - Common Name / Genus / Size
- Analysis on planting trends
    - Proportions in Developing / Developed Communities

Packages

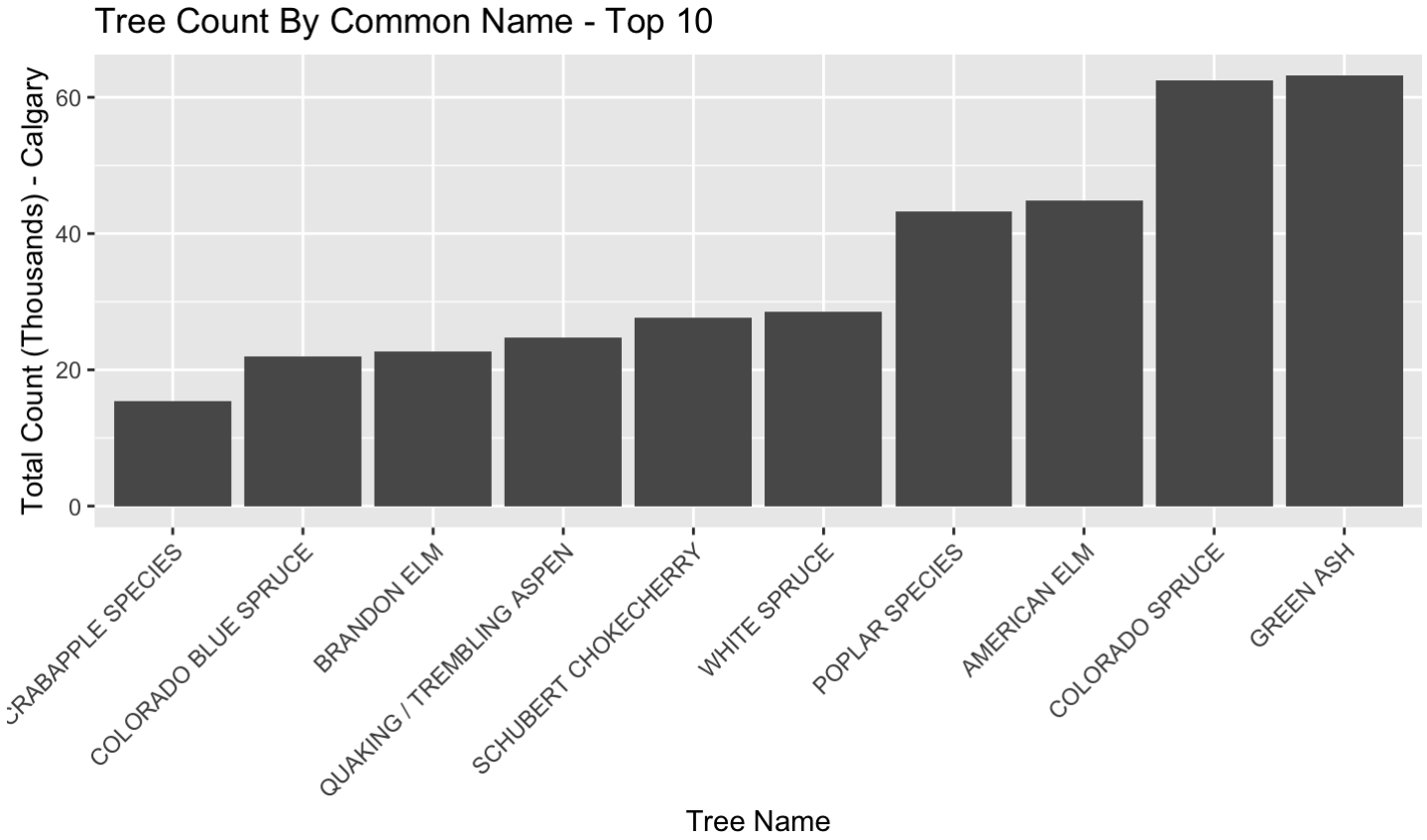
- dplyr
- ggplot2
- readr
- stringr
- mosaic

Explore the data

- Many different features to look at
- Determine the most popular trees by name / genus
    - Isolate the analysis on these types of trees
- Tree size

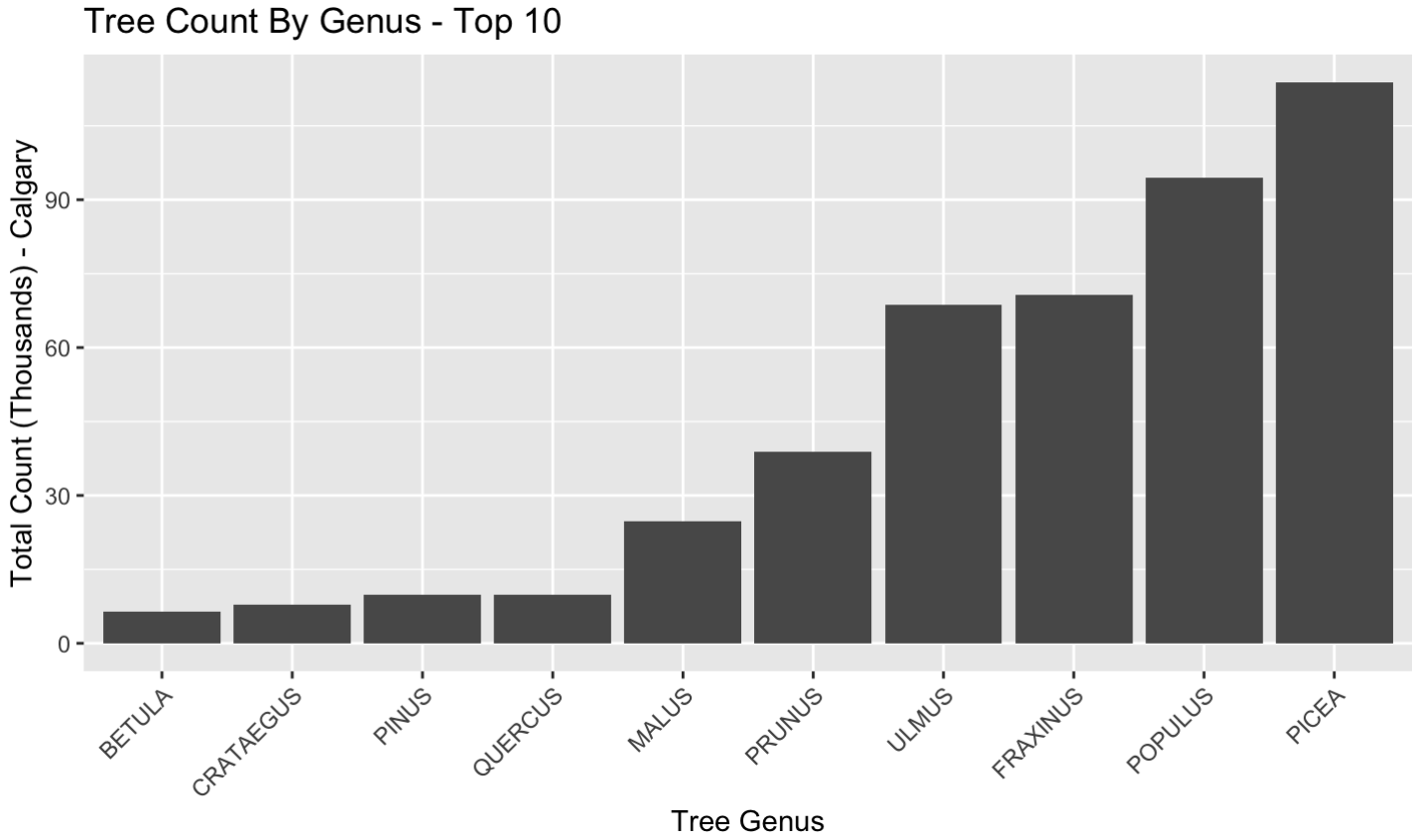
Explore - Common Tree Names

- Top 10 Trees by Common Name in Calgary
- Select "Green Ash"



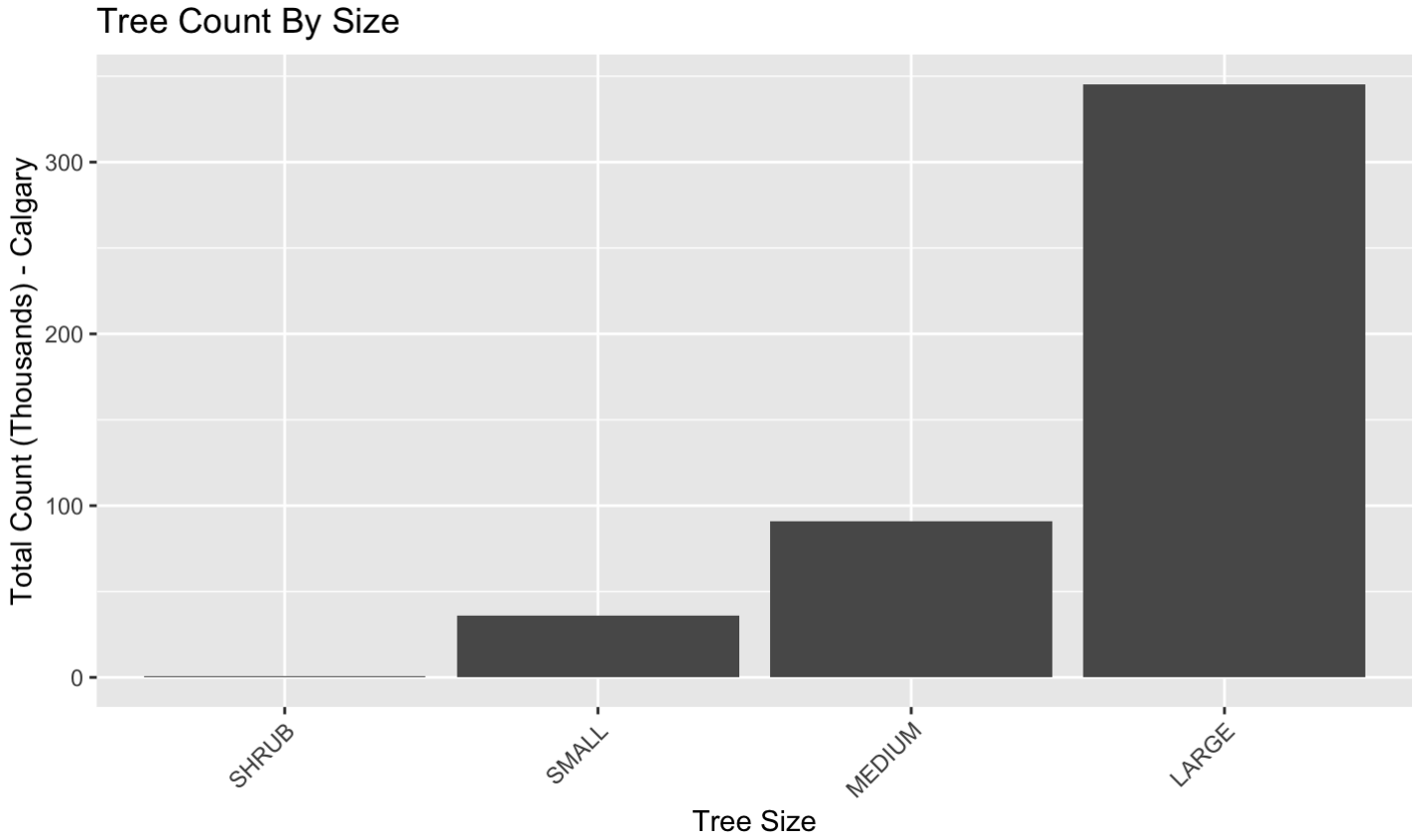
Explore - Genus

- Top 10 Trees by Genus in Calgary
- Select "Picea"



Explore - Size

- Large trees are most common, small are least common
- For comparison, Large and Small trees are selected



Difference of Tree Condition Means - Green Ash and Other

- Test how average Green Ash condition compares against all other trees
- Is there a reason why this is the most popular tree in Calgary?
- Should it be the most popular tree in Calgary?

Difference of Tree Condition Means - Green Ash and Other

Stastical Hypothesis

$$H_0 : \quad \mu_{\text{Tree Condition}(GreenAsh)} \leq \mu_{\text{Tree Condition}(Other)}$$

$$H_A : \quad \mu_{\text{Tree Condition}(GreenAsh)} > \mu_{\text{Tree Condition}(Other)}$$

- Statistical Hypothesis is based on assumption that the most popular tree would have a better overall condition than other trees

Difference of Tree Condition Means - Green Ash and Other

- Convential confidence interval estimation of the difference between two population means
- Conditions can be relaxed since n = 450,632

```
##
##  Welch Two Sample t-test
##
## data:  TREE_CONDITION_RATING_PERC by COMMON_NAME
## t = -5.7177, df = 94963, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.3120178        Inf
## sample estimates:
## mean in group GREEN ASH     mean in group OTHER
##                60.65806                60.90037
```

Difference of Tree Condition Means - Green Ash and Other

- The P-Value in the t-test is 1
- Cannot reject our null hypothesis
- Average tree condition of other trees is greater than average condition of Green Ash
- Should Green Ash be the most popular tree if the condition doesn't out perform the others?

Difference of Tree Condition Means - Picea and Other

- Test how average Picea condition compares against all other trees
- Is there a reason why this is the most popular tree Genus in Calgary?
- Should it be the most popular tree Genus in Calgary?

Difference of Tree Condition Means - Picea and Other

Stastical Hypothesis

$$H_0 : \quad \mu_{\text{Tree Condition}(Picea)} \leq \mu_{\text{Tree Condition}(Other)}$$

$$H_A : \quad \mu_{\text{Tree Condition}(Picea)} > \mu_{\text{Tree Condition}(Other)}$$

- Statistical Hypothesis is based on assumption that the most popular tree would have a better overall condition than other trees

Difference of Tree Condition Means - Picea and Other

- Convential confidence interval estimation of the difference between two population means
- Conditions can be relaxed since n = 450,632

```
##
##  Welch Two Sample t-test
##
## data:  TREE_CONDITION_RATING_PERC by GENUS
## t = -41.57, df = 197400, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -1.521122
## sample estimates:
## mean in group OTHER mean in group PICEA
##            60.48765             62.07144
```

Difference of Tree Condition Means - Picea and Other

- The P-Value in the t-test is less than 0.05
- Reject our null hypothesis
- Average tree condition of Picea is greater than average condition of other trees
- Picea appear to be a good pick for the most popular tree Genus in Calgary!

Difference of Tree Condition Means - Large and Small trees

- Test how average tree condition compares between Large and Small trees
- Should Calgary be planting more Small trees?

Difference of Tree Condition Means - Large and Small trees

Stastical Hypothesis

$$H_0 : \quad \mu_{\text{Tree Condition}(Large)} \geq \mu_{\text{Tree Condition}(Small)}$$

$$H_A : \quad \mu_{\text{Tree Condition}(Large)} < \mu_{\text{Tree Condition}(Small)}$$

- Statistical Hypothesis is based on our assumption that large trees would have lower tree condition because Calgary isn't a great climate for trees

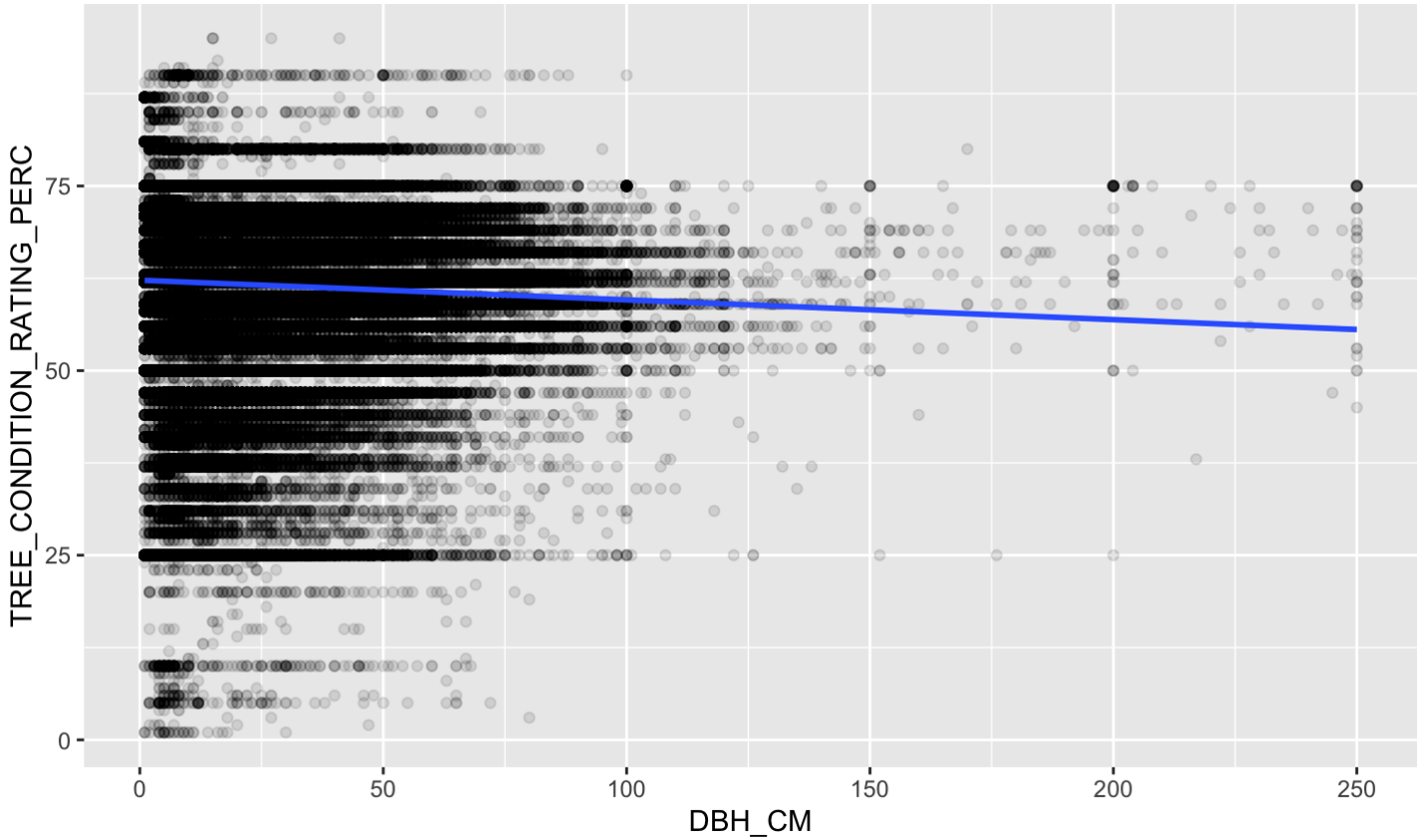
Difference of Tree Condition Means - Large and Small trees

```
##
##  Welch Two Sample t-test
##
## data:  TREE_CONDITION_RATING_PERC by MATURE_SIZE
## t = -1.8853, df = 40574, p-value = 0.0297
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##         -Inf -0.01564818
## sample estimates:
## mean in group LARGE mean in group SMALL
##            60.94202            61.06476
```

Difference of Tree Condition Means - Large and Small trees

- The P-Value in the t-test is 0.0297
- Reject our null hypothesis
- Average tree condition of small trees is greater than average condition of large trees
- Should small trees be planted more often than large trees?
- The mean values are very close - is it negligible?

Tree Condition vs Size (Diameter of Trunk)



Tree Condition vs Size (Diameter of Trunk)

- Coefficient of determination is very low
- Only 0.2% of the data can be described by this linear model

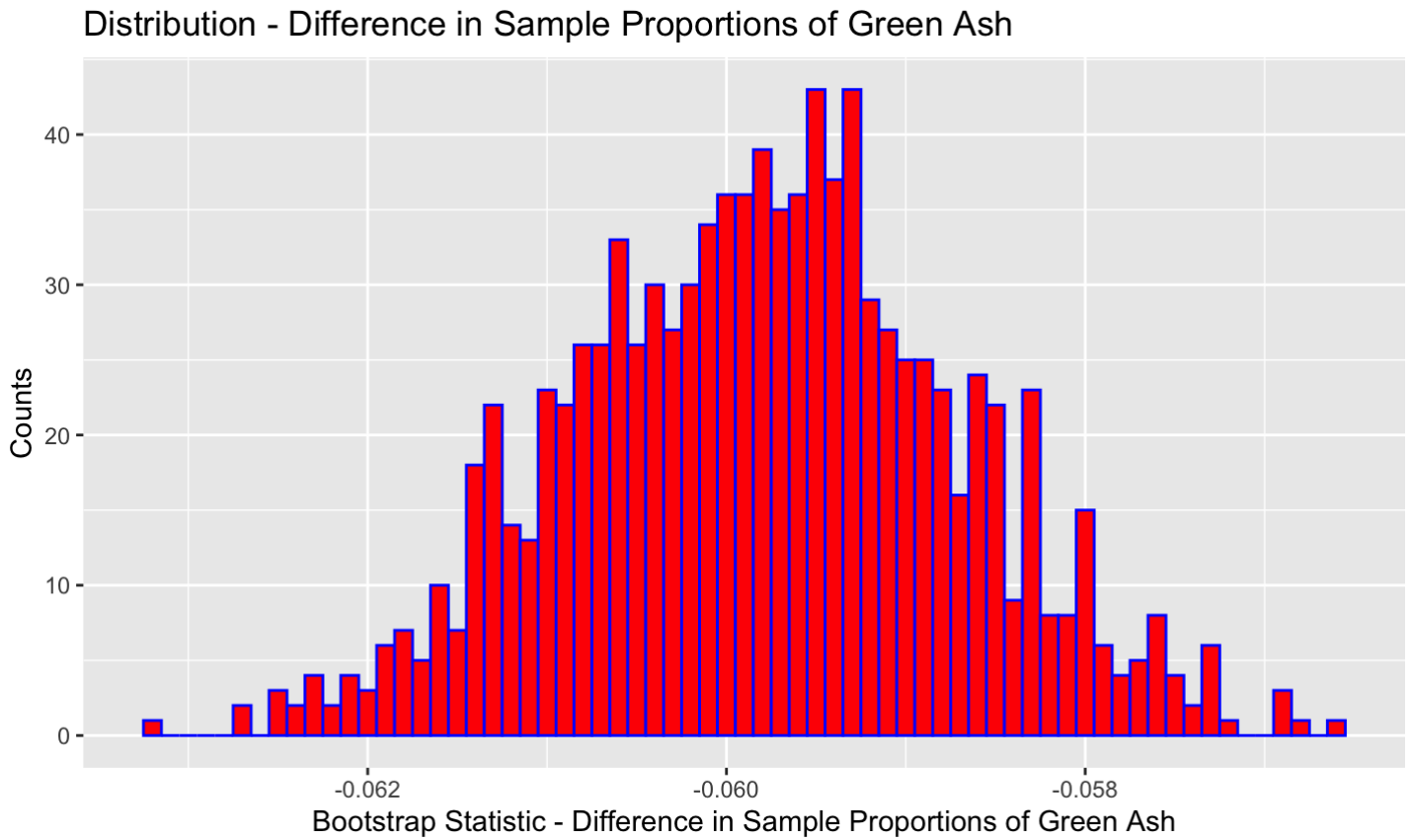
```
## (Intercept)      DBH_CM
## 62.25473402 -0.02672251
```

```
## [1] 0.002649334
```

Green Ash Proportion

- Determine if there is a difference between the proportions of Green Ash in developing and developed communities
- Does the city know that they should be planting fewer Green Ash?
- Bootstrapping to create confidence intervals

Green Ash Proportion Differences



Green Ash Proportion Differences

```
##         min          Q1      median          Q3         max        mean
##  -0.0632317 -0.06056134 -0.05978292 -0.05909098 -0.05662287 -0.05979703
##          sd    n missing
##  0.00107439 1000       0
```

```
##          quantile     p
## 2.5%  -0.06188652 0.025
## 97.5% -0.05764818 0.975
```

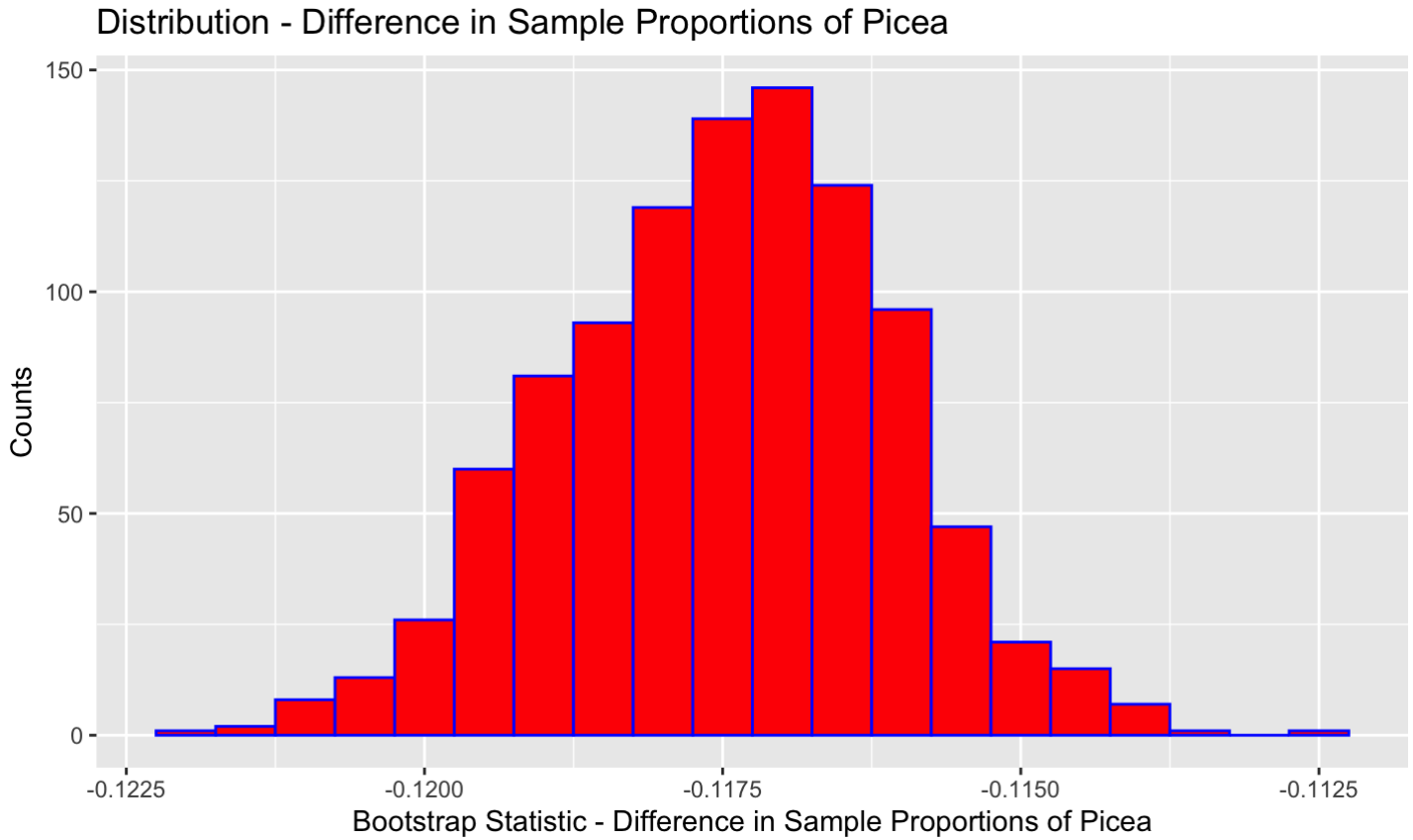
Green Ash Proportion

- There is a clear difference between the Green Ash proportion in developing and developed communities
- Fewer Green Ash are being planted in developing communities
- Is it because Green Ash condition is on average less than the average overall condition of other trees?

Picea Proportion

- Determine if there is a difference between the proportions of Picea in developing and developed communities
- Does the city know that they should be planting more Picea?
- Bootstrapping to create confidence intervals
- Fewer Genus categories (fewer naming inconsistencies than common names)

Picea Proportion Differences



Picea Proportion Differences

```
##         min         Q1     median         Q3        max       mean
##  -0.1222125 -0.1184362 -0.1174318 -0.1165313 -0.1125122 -0.1174781
##          sd    n missing
##  0.00138449 1000       0
```

```
##         quantile     p
## 2.5%  -0.1201641 0.025
## 97.5% -0.1148295 0.975
```

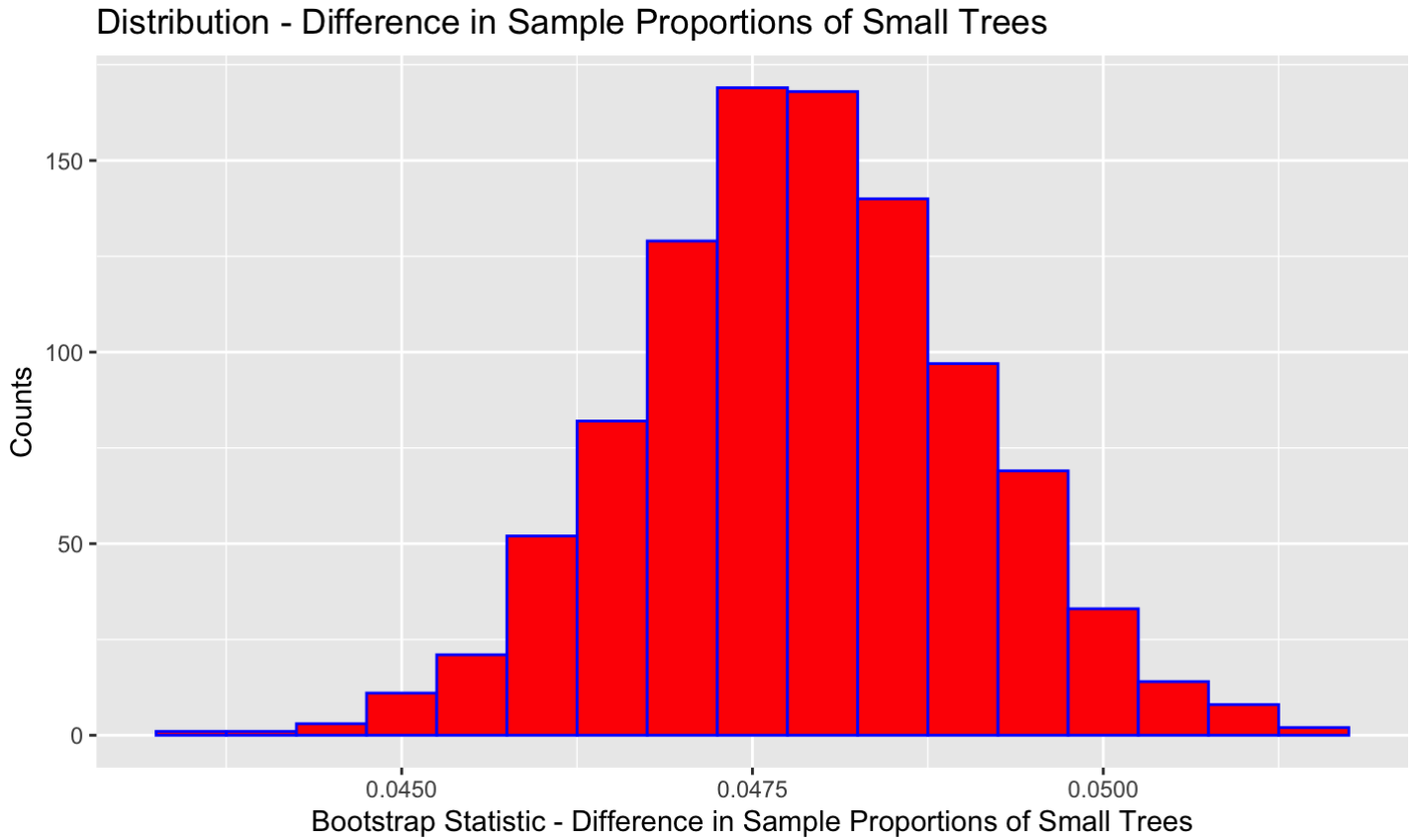
Picea Proportion

- There is a clear difference between the Picea proportion in developing and developed communities
- Fewer Picea are being planted in developing communities
- Why is this the case? Shouldn't the city be planting more?
- Are there other reasons for the decline in proportion of Picea?

Mature Size Proportion

- Determine if there is a difference between the proportions of Small trees in developing and developed communities
- Does the city know that they should be planting more small trees?
- Bootstrapping to create confidence intervals

Small Proportion Differences



Small Proportion Differences

```
##         min         Q1     median         Q3        max      mean
##  0.04369133 0.04706572 0.04785151 0.04862192 0.05147679 0.0478573
##           sd    n missing
##  0.001185796 1000       0
```

```
##         quantile     p
## 2.5%  0.04556293 0.025
## 97.5% 0.05020030 0.975
```

Small Proportion

- There is a clear difference between the proportions of Small trees in developing and developed communities
- More Small trees are being planted in developing communities
- Is it because the condition of Small trees is on average greater than the condition of Large trees?
- We may need to check the proportions of Medium sized trees as well to make an accurate conclusion

Conclusions

- The City of Calgary Public Trees dataset gives us an understanding of tree condition over a variety of variables
- The most popular tree NAME appear to have a lower average tree condition than the average tree condition of all other trees
- The most popular tree GENUS appear to have a higher average tree condition than the average tree condition of all other trees
- Small trees on average have a higher tree condition than large trees

Conclusions

- The proportion of Green Ash appears to be reducing in developing communities
    - Good! Based on average tree condition of Green Ash, this is a good trend.
- The proportion of Picea appears to be reducing in developing communities
    - Bad! Based on average tree condition of Picea, this is not a good trend.
- The proportion of Small trees appears to be increasing in developing communities
    - Good! Based on average tree condition of Small trees, this is a good trend.

References

- Public Trees
  - City of Calgary (2019) Public Trees [Online]. Available at: https://data.calgary.ca/Environment/Public-Trees/tfs4-3wwa (Accessed: 28 September 2019) Community Points
  - City of Calgary (2019) Community Points [Online]. Available at: https://data.calgary.ca/Base-Maps/Community-Points/j9ps-fyst (Accessed: 28 September 2019)