

Pre-Processing

Luke Wertis

2022-10-20

```
library(data.table)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

Read in the files

Here we are reading in the csv files for the three locations that we have already performed tree segmentation and collected the average R,G,B value within the individual tree polygons.

```
BR <- fread("C:\\Users\\wertisml\\Documents\\Hemlocks\\Scripts_and_Tools\\tree_crowns\\Final\\csv\\Blowing_Rock.1.csv")
Parkway_Bridge <- fread("C:\\Users\\wertisml\\Documents\\Hemlocks\\Scripts_and_Tools\\tree_crowns\\Final\\csv\\Parkway_Bridge.2.csv")
Valle_Crusis <- fread("C:\\Users\\wertisml\\Documents\\Hemlocks\\Scripts_and_Tools\\tree_crowns\\Final\\csv\\Valle_Crusis.3.csv")
```

Reformatting

Here we want to rename the R,G,B columns so that they all share the same column names to make it easier later on to combine dataframes.

```
BR <- BR %>%
  rename(R = Blowing_Rock.1,
         G = Blowing_Rock.2,
         B = Blowing_Rock.3)

Parkway_Bridge <- Parkway_Bridge %>%
```

```

  rename(R = Parkway_Bridge.1,
         G = Parkway_Bridge.2,
         B = Parkway_Bridge.3)

Valle_Crusis <- Valle_Crusis %>%
  rename(R = Valle_Crusis.1,
         G = Valle_Crusis.2,
         B = Valle_Crusis.3)

```

Seperating By Tree Type

At this point we want to create seperate dataframes for each tree type, hemlock, pine, deciduous. There might be more efficient methods of doing it but the approach that I took was once I was able to view the tree polygons, aerial imagery, and true point locations, I identified the best trees and noted their treeId and what type of tree it was. This treeID is how I am able to select the trees I want to extract from the dataframes.

```

#=====#
# Blowing Rock
#=====#

# Hemlock

Hemlock_Trees <- c(453, 791, 862, 5659, 5702, 6065, 6060, 5448, 5416, 5435)

BR_Hemlock <- BR %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Hemlock_Trees)) %>%
  mutate(Tree_Type = "Hemlock")

# Pine Tree

Pine_Trees <- c(5177, 3339, 3347)

BR_Pine <- BR %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Pine_Trees)) %>%
  mutate(Tree_Type = "Pine")

# Deci Tree

Deci_Trees <- c(558, 5757, 5720, 5785, 5852, 5744, 5815, 5956, 6003, 6127, 6246,
               6305, 6341, 6311, 6258, 6072, 6113, 6138, 5846, 5885, 5916, 6007,
               6181, 6293, 6154, 6112, 6180, 6174, 6126, 6200, 6235)

BR_Deci <- BR %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Deci_Trees)) %>%
  mutate(Tree_Type = "Deciduous")

#=====#
# Parkway Bridge
#=====#

# Hemlock

```

```

Hemlock_Trees <- c(7279, 6819, 5675, 5752, 5268, 5243, 5193, 5064, 5117, 5231,
                  6119, 6465, 7766, 5977, 5501, 5347, 4454)

Parkway_Bridge_Hemlock <- Parkway_Bridge %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Hemlock_Trees)) %>%
  mutate(Tree_Type = "Hemlock")

#####
# Vallie Crusis
#####

# Hemlock

Hemlock_Trees <- c(6385, 6358, 6329, 6278, 5746, 5495, 5424, 5727, 5495, 5811,
                  6251, 6279, 6318, 6632, 6642)

VC_Hemlock <- Valle_Crusis %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Hemlock_Trees)) %>%
  mutate(Tree_Type = "Hemlock")

# Pine Tree

Pine_Trees <- c(6823, 6843, 6889, 6878, 6868, 6718, 6312, 6362, 6495, 6541, 6525,
               6572, 6578, 6588, 6561, 6585, 6624, 6601, 6621, 6645, 6651, 6991,
               6977, 6928, 6787, 6752, 6699, 6683, 6649, 6638, 6619, 6612, 6579,
               6557, 6578, 6496, 6765)

VC_Pine <- Valle_Crusis %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Pine_Trees)) %>%
  mutate(Tree_Type = "Pine")

# Deci Tree

Deci_Trees <- c(6640, 6622, 6644, 6637, 6573, 6443, 6198, 6212, 6599, 5552, 5448,
               5435, 5409, 5375, 5441, 5644, 5484, 5671, 5553, 5624, 5696, 5768,
               5866, 5886, 5984, 6077, 6252, 6307, 6369, 6580, 6569, 6576, 6544,
               6623, 6614, 6602, 6613, 6639, 6627, 6646, 6663, 6676, 6687, 6704,
               6761, 6757, 6777, 6814, 6834, 6841, 6841, 6786, 6817, 6791, 6826,
               6842, 6897, 6920, 6941, 7031, 7020, 7131, 7141, 7150, 7019, 7043,
               7090, 7046, 7023, 7012, 7000)

VC_Deci <- Valle_Crusis %>%
  filter_at(vars(contains("treeID")), any_vars(. %in% Deci_Trees)) %>%
  mutate(Tree_Type = "Deciduous")

```

Create a New Dataframe

Now we want to combine all the hemlocks, pines, and deciduous trees into their individual dataframes and then combine into an overall dataframe. Again, there could be a more efficient method for doing this.

In the end you have one dataframe named Trees, it contains all the information that was available in the original dataframes at the start of this process with the addition on a column called Tree_Type, which indicates what type of tree it is. This dataframe is also noticeably smaller than the originals and smaller than

the total number of true point locations because this represents the most correctly segmented trees. This data will then be used in the classification methods to determine if we can correctly classify trees.

```
# Hemlocks

Hemlocks <- BR_Hemlock %>%
  rbind(Parkway_Bridge_Hemlock, fill = TRUE) %>%
  rbind(VC_Hemlock, fill = TRUE)

# Pine Trees
Pines <- BR_Pine %>%
  rbind(VC_Pine, fill = TRUE)

# Deciduous

Deciduous <- BR_Deci %>%
  rbind(VC_Deci, fill = TRUE)

Trees <- Hemlocks %>%
  rbind(Pines) %>%
  rbind(Deciduous)
```