JW_Acacia_v_Trees

James Waterford

2023-03-02

##Fresh Start

Previously, we looked at the UHURU data set, and got a lot of information from there. Maybe too much. Let's start again and only consider columns that are relevant to us.

First, we must extract and prepare the files.

```
library(readr)
trees <- read_tsv("../197-raw_storage/TREE_SURVEYS.txt",
                  col_types = list(HEIGHT = col_double(),
                                   AXIS_2 = col_double()))
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

Calculations

I want to know the area under the canopy. This is not provided to us in our data set, but we can calculate and create a designated column.

To create a column, we write it into existence. `df$[New_Col] <- c([Val])`

Use `c([function])` to fill row values with data.

```
trees$CANOPY <- (trees$AXIS_1 * trees$AXIS_2)

head(trees$CANOPY)
```

```
## [1] 30.5000 69.7125 79.6500 39.0500 40.7500  6.1600
```

Awesome. So we now have a new column added, but we still need to subset the data. There are a few methods.

```
subset([DATA.FRAME], select = c([ columns-to-keep ]))
```

```
new_trees <- subset(trees, select = c(SURVEY, YEAR, SITE, CANOPY, HEIGHT, TREATMENT, SPECIES))

t1 <- data.frame(trees$CANOPY, trees$SURVEY, trees$YEAR, trees$SITE)

## This first comma means "Leave the row values alone" ##
t2 <- trees[,c("YEAR", "SITE", "CANOPY", "HEIGHT")]
```

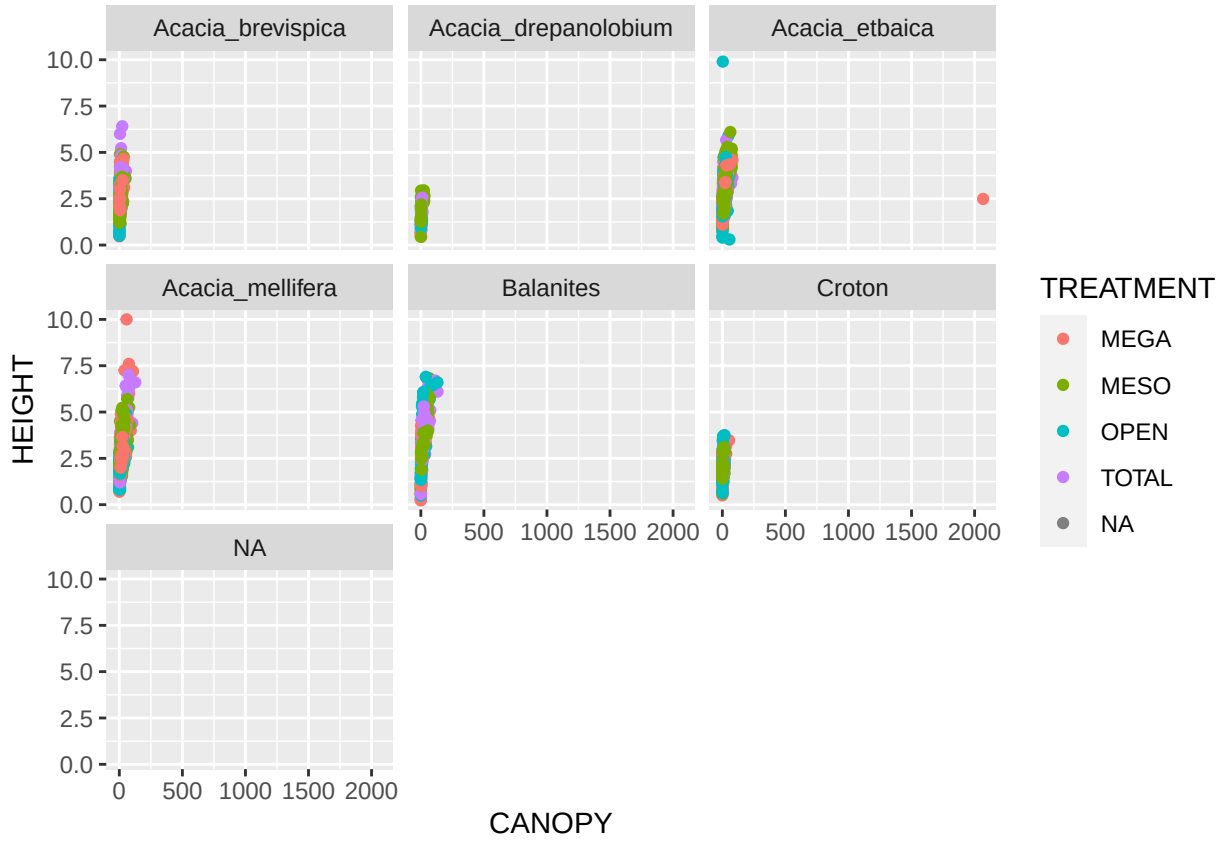
Graphical Analysis

Let's try creating a *scatterplot* and see how it looks.

We'll call `ggplot` and use `facet_wrap` to separate the data by each **species**

```
library(ggplot2)
ggplot(new_trees, aes(x = CANOPY,y= HEIGHT, color = TREATMENT)) +
  geom_point() +
  facet_wrap(~SPECIES)
```

```
## Warning: Removed 215 rows containing missing values (`geom_point()`).
```



This is good! But there is a reason that is looks so messy.

The outlier in this data set is really throwing things off..

Let's remove via the `subset()` function.

```
new_trees <- subset(new_trees, CANOPY < 2000)
```

And then let's graph it again?

```
library(ggplot2)
ggplot(new_trees, aes(x = CANOPY,y= HEIGHT, color = TREATMENT)) +
  geom_point() +
  facet_wrap(~SPECIES)
```

There we go! Now we have some clean(ish) data to look at. As we can see, Height is quite variable, and might be correlated in some species.

```
Acacia_sub <- subset(trees, SPECIES == "Acacia_etbaica" | SPECIES == "Acacia_drepanolobium" | SPECIES ==


other_sub <- subset(trees, SPECIES != "Acacia_etbaica" & SPECIES != "Acacia_drepanolobium" & SPECIES !=



ggplot() +
  geom_point(data = Acacia_sub, aes(x= CIRC, y= HEIGHT), color = "slategray") +
  geom_smooth(data = Acacia_sub,  method = "lm", mapping = aes(x= CIRC, y= HEIGHT), color = "grey23") +
  geom_point(data = other_sub, aes(x= CIRC, y= HEIGHT), color = "red") +
  geom_smooth(data = other_sub,  method = "lm", mapping = aes(x= CIRC, y= HEIGHT), color = "tomato2") +
  scale_y_sqrt() +
  scale_x_sqrt()
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 292 rows containing non-finite values (`stat_smooth()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 115 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 292 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 115 rows containing missing values (`geom_point()`).
```