Effects of a Defaunation Gradient on Tropical Forest Structure in Ivindo National Park, Gabon

https://github.com/israelgolden/GoldenGriffithsKnierMalinowski_ENV872_EDA_FinalProject

Tashsa Griffiths, Israel Golden, Aubrey Knier, and Mishka Malinowski

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

2 Dataset Information

This dataset if provided by the Pouslen Tropical Ecology Lab here at Duke. yadda yadda...

...briefly explain what the project is / what the data are...

The dimensions and variable information of the raw dataset are below:

```
## [1] 45681    21
```

Column name	Description	Unit	Range (if applicable)
E	Field expedition season	Season-Year	Winter - Summer 2021
Data_entry	Name of individual inputting data to Excel	Name	
Date..dd.mm.yyyy.	Date of Excel data entry	Date/Month/Year	March 2021 - November 2022
File_name	Photo file name of field data sheet	.JPG	
Date..dd.mm.yyyy..1	Data of field data collection	Date/Month/Year	June - January 2021
Note_taker	Name of individual recording field data	Name	
Project	Defaunated forest (DF) or intact forest (IF) plot	Category	DF or IF
Plot	Unique plot identification	Category	1A, 1B, 2A, 2B, 3A, 3B, 4A, 4B, 5A, 5B, 6A, 6B
Grid	Within-plot grid where data were collected	Category	
TAG_SUM	The most unique identifier, using plot grid and plant tag	Category	
Plant_tag	Identifer assigned to each sample	Letter-Number Combination	
X_coord	X coordinate of sample location	Degrees	0.00 - 9.80
Y_coord	Y coordinate of sample location	Degrees	0.00 - 8.75
Tool	Tool used to measure diameter (DBH or caliper (CP))	Category	DBH or CP
POM	Point of measurement for diameter	Meters	0.00 - 11.00
DBH.mm	Diameter at breast height (DBH)	Millimeters	0.00 - 173.00
Height..meters.	Height of plant	Meters	0.07 - 70.00

Column name	Description	Unit	Range (if applicable)
Type_Field	Vegetation type or size class of plant	Category	Seedling, Sapling, Liana, Tree
Note_Field	Miscellaneous field notes	Phrase	
ID	Latin species identification	Name	
Treatment	Future plot treatments (fungicide/insecticide)	Category	LMC, LME, MME, MMC

With such a large dataset, data cleaning and wrangling was an essential process for creating a manageable dataset that was relevant for answering our research questions. First, we subset our selected six plots for our analysis:

```
## [1] "DF_3B" "IF_2A" "DF_5A" "DF_5B" "DF_6A" "DF_6B"
```

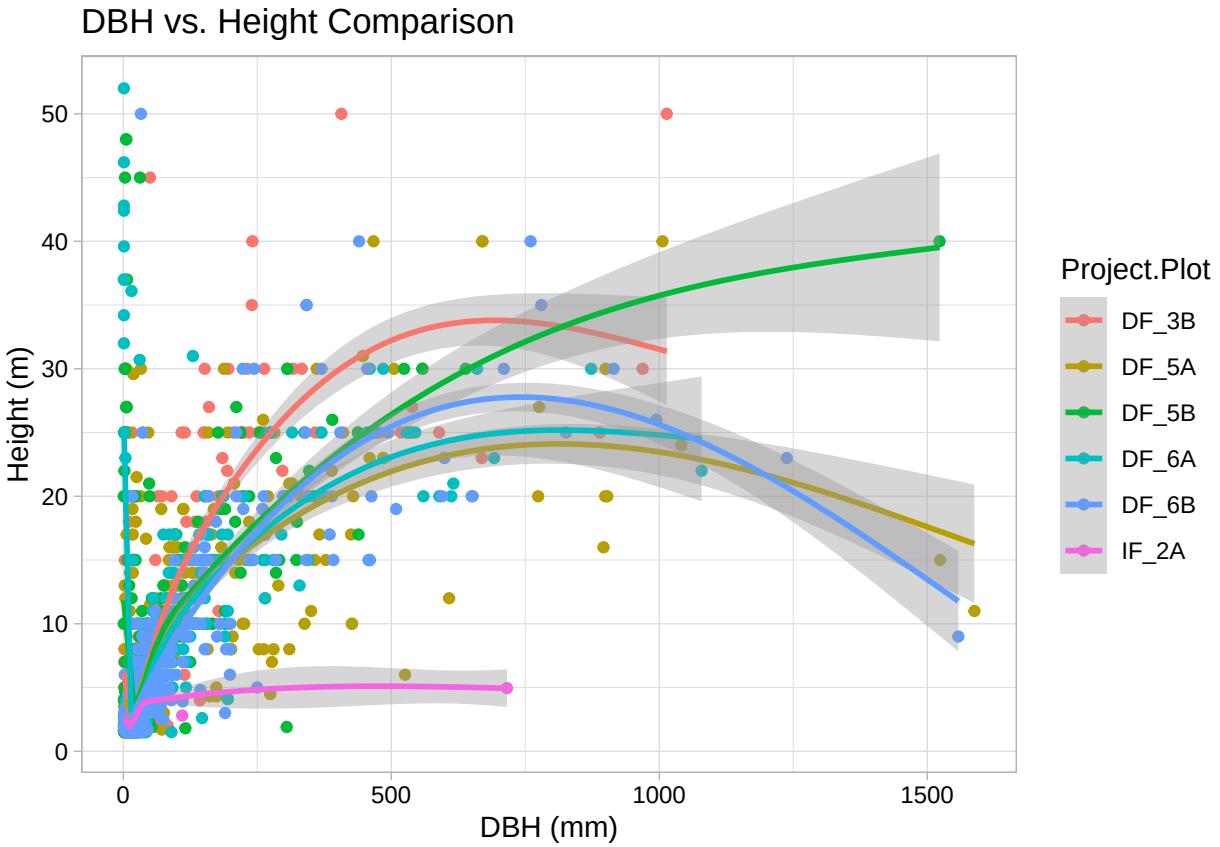
These plots were chosen out of the total 20 plots because they were the only ones that had species identifications attached to samples, which was needed for our investigation of how defaunation affects species composition.

Next, we only selected columns that contained variables of interest:

```
## [1] "Project.Plot" "Plant_tag"    "DBH_mm"       "Height_m"     "Veg_Type"
## [6] "ID"
```

We removed absent or unreasonable values from the dataset. This involved simply removing blank cells or "NAs", as well as measurements that were likely incorrect, probably as a result of improper unit conversions. Additionally, we improved uniformity in the dataset by removing samples that had a height less than 1.5m and lianas. This was because not all plots measured individuals smaller than 1.5m, and height measurements for lianas are less reliable, so we decided to only analyze trees. We also found strange instances in the data: some samples were relatively tall yet had very low DBHs (Figure ___); this is probably due to some error in units. Therefore, we removed any samples that had a DBH less than 1mm and a height above 1.5m to improve accuracy.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



We added in variables to support our research questions and analyses. We created two new columns: "Status" and "Distance_km"

Column name	Description	Unit	Range
Status	Indicates whether each plot is defaunated or intact forest	Category	Defaunated - Intact
Distance_km	Distance of each plot from Mokokou	Kilometers	8.177 - 40.224

The categorical variable, "Status", will help with data visualization, and the "Distance_km" variable will be used as a proxy from the defaunation gradient in our analyses.

The new dimensions of this dataset are:

```
## [1] 6479    8
```

Our cleaned dataset looks as follows:

```
##   Project.Plot Plant_tag DBH_mm Height_m Veg_Type                    ID
## 1        DF_3B      1554  558.8     30.0     Tree   Heisteria parvifolia
## 2        DF_3B        69   15.8      2.4     Tree   Dialium pachyphyllum
```

```
## 3        DF_3B      4371   11.7      1.8  Sapling Scorodophloeus zenkeri
## 4        DF_3B       607   19.4      2.5     Tree    Odjendja gabonensis
## 5        DF_3B      7150   21.5      3.7     Tree Scorodophloeus zenkeri
## 6        DF_3B      7110   65.0      7.5     Tree Centroplacus glaucinus
##       Status Distance_km
## 1 Defaunated      20.195
## 2 Defaunated      20.195
## 3 Defaunated      20.195
## 4 Defaunated      20.195
## 5 Defaunated      20.195
## 6 Defaunated      20.195
```

Accidental misspellings are common in datasets such as this with thousands of manual entries of complex Latin species names. This is a concern because two samples that are supposed to be the same species, but have different spellings, will not be identified as the same species in our analyses. By looking at a list of the unique species names in the dataset, we found this to be the case in several instances. Identifying these errors and correcting them was very labor intensive as this can only really be done with the human eye using personal judgment as to what names were meant to be the same. Before cleaning the species names, there were 349 "species"; after correcting for spelling mistakes, there were only 323 species. This means that 26 "species" were falsely identified prior to data cleaning.

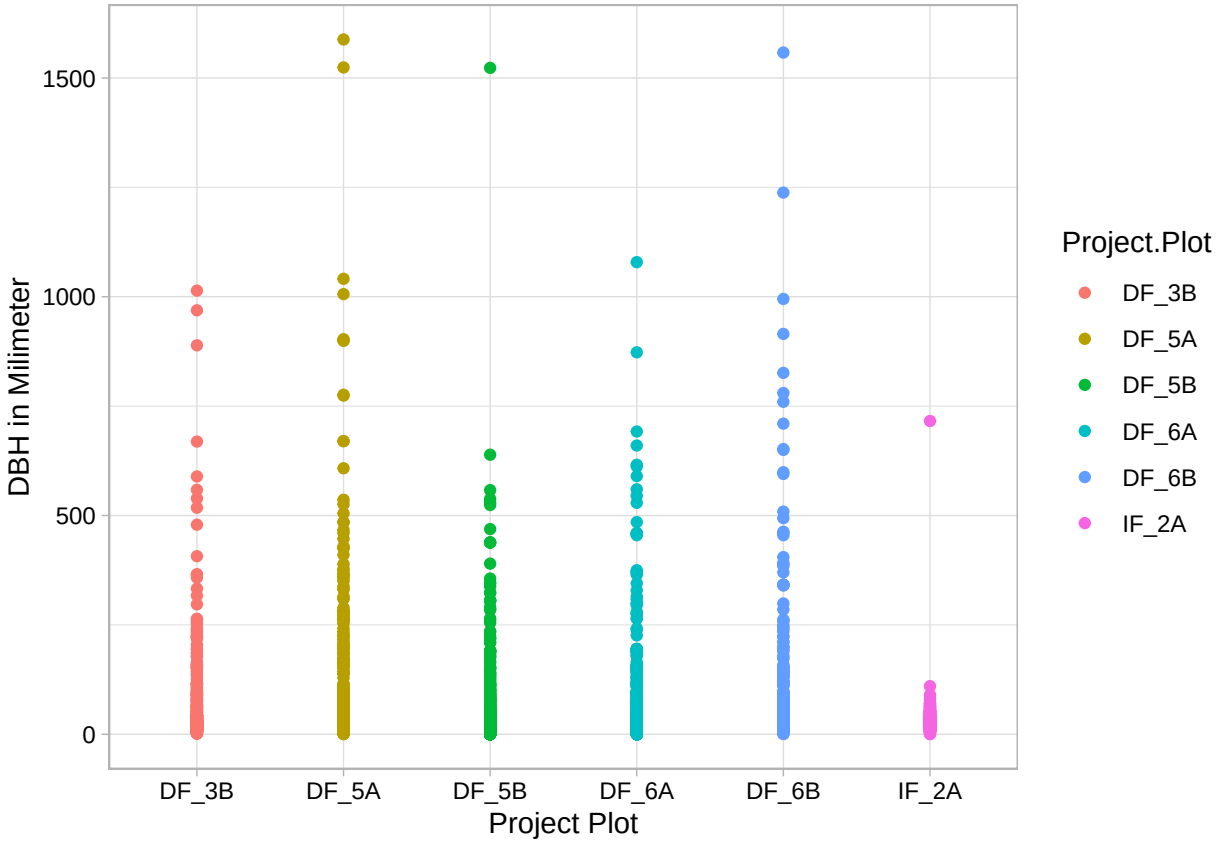
The dimensions of the processed, clean dataset are as follows:

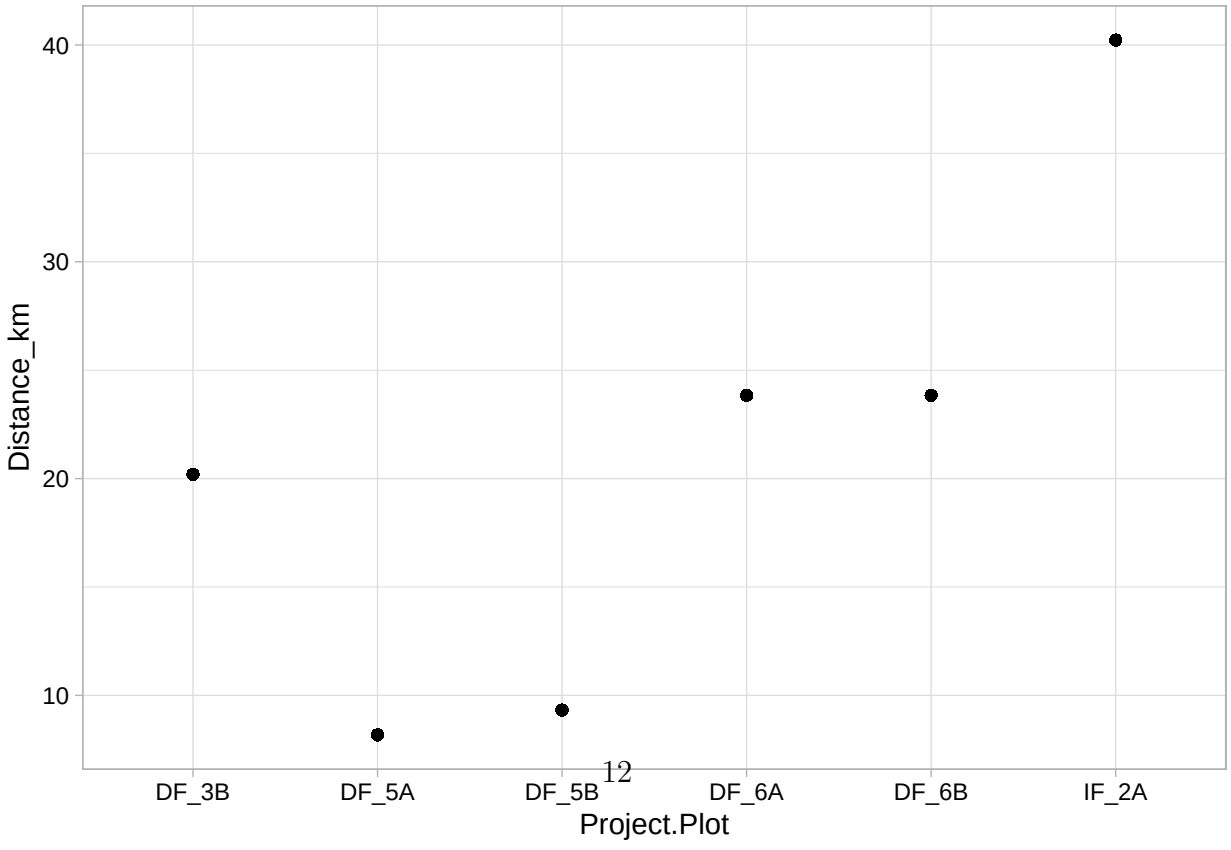
```
## [1] 6340    7
```

3 Exploratory Analysis

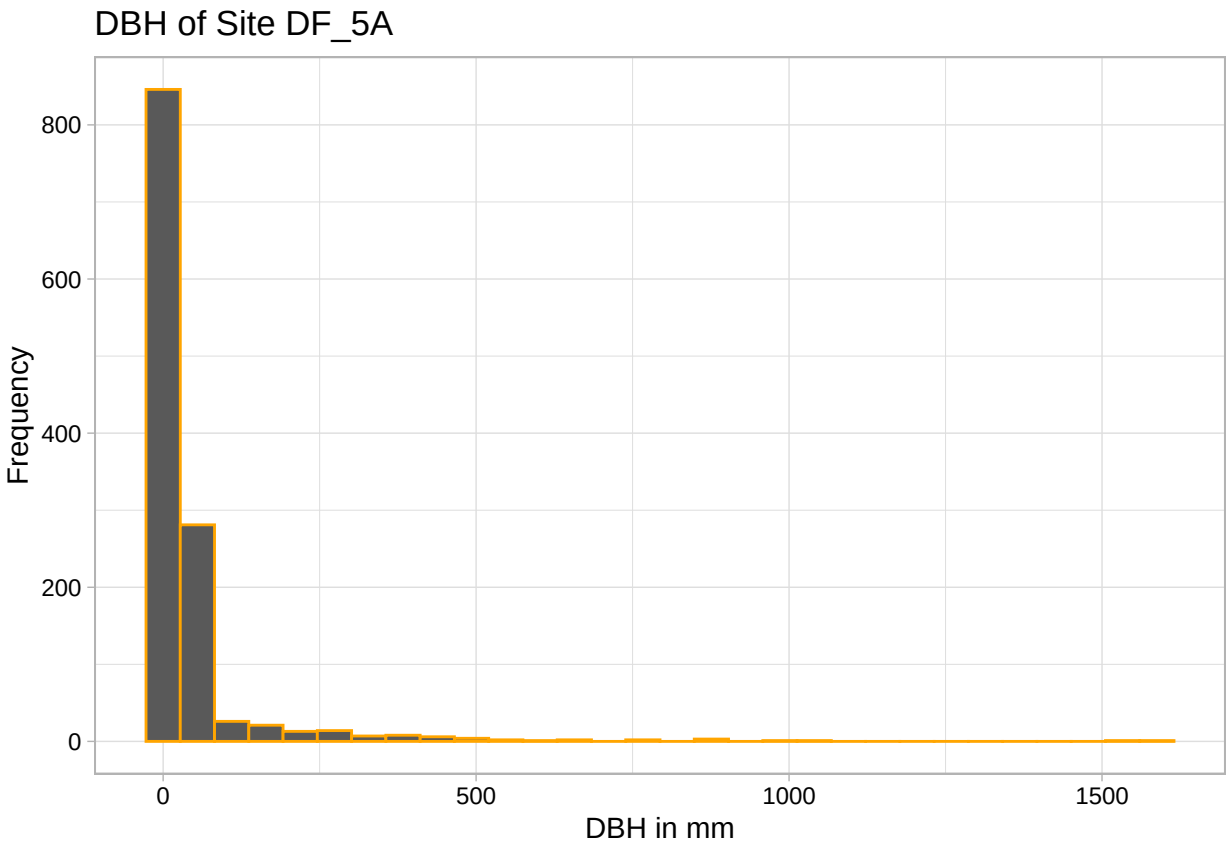
3.1 Spatial Analysis

3.2 DBH





```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

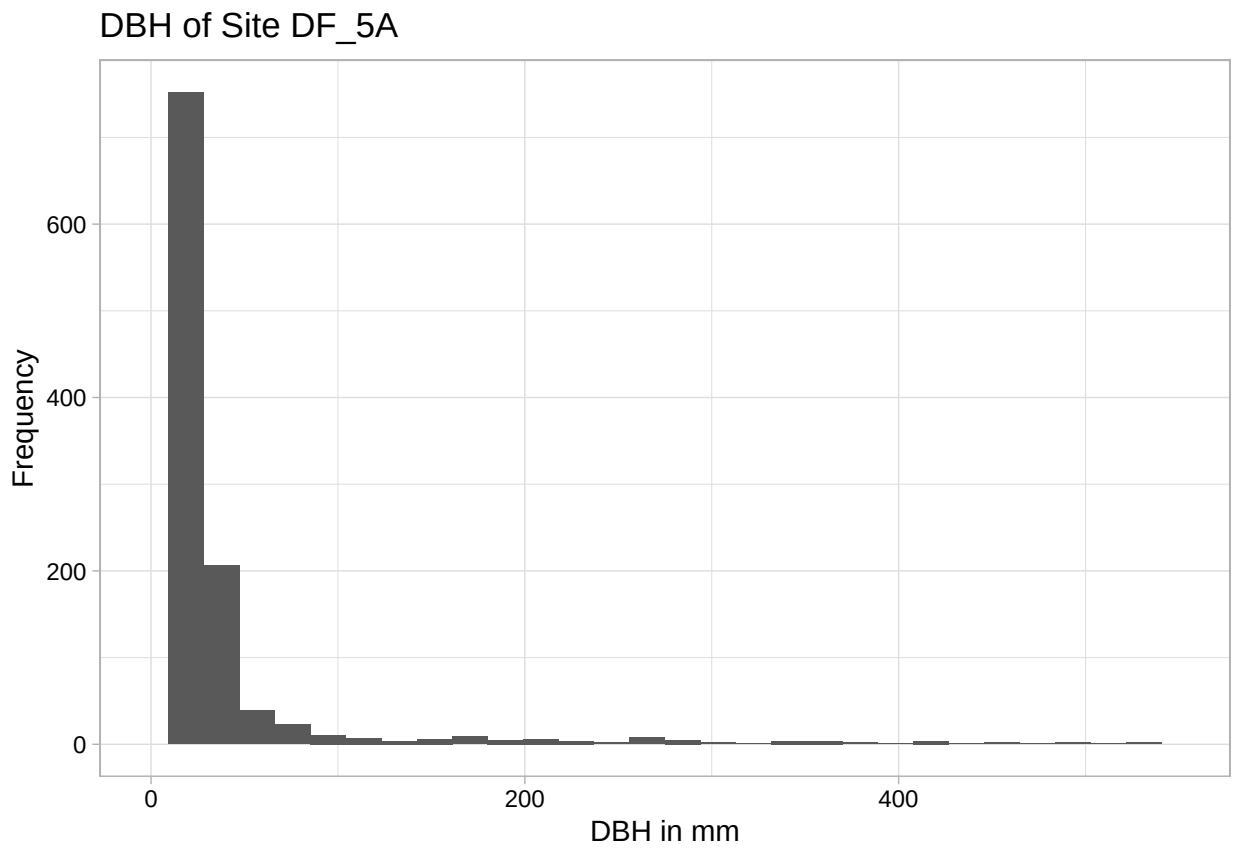



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

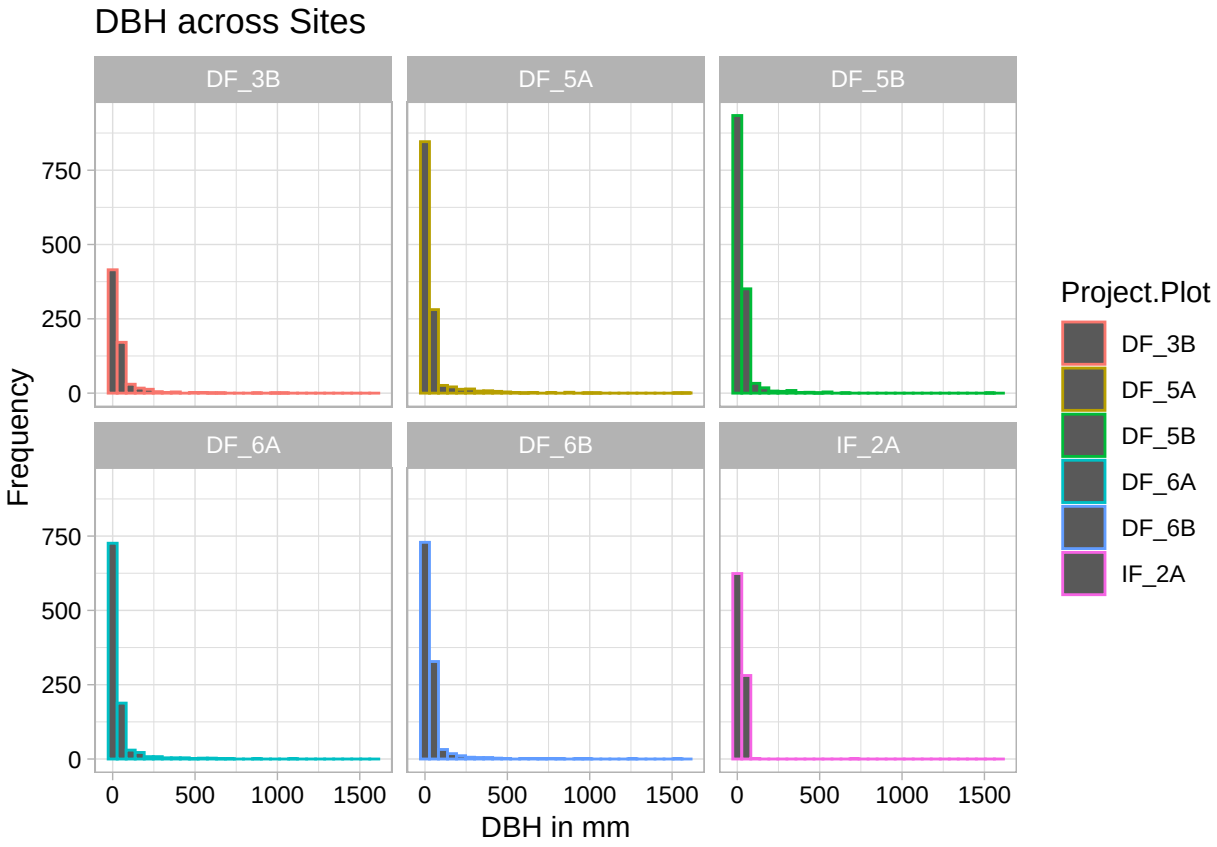

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 12 rows containing non-finite values (stat_bin).

## Warning: Removed 2 rows containing missing values (geom_bar).
```



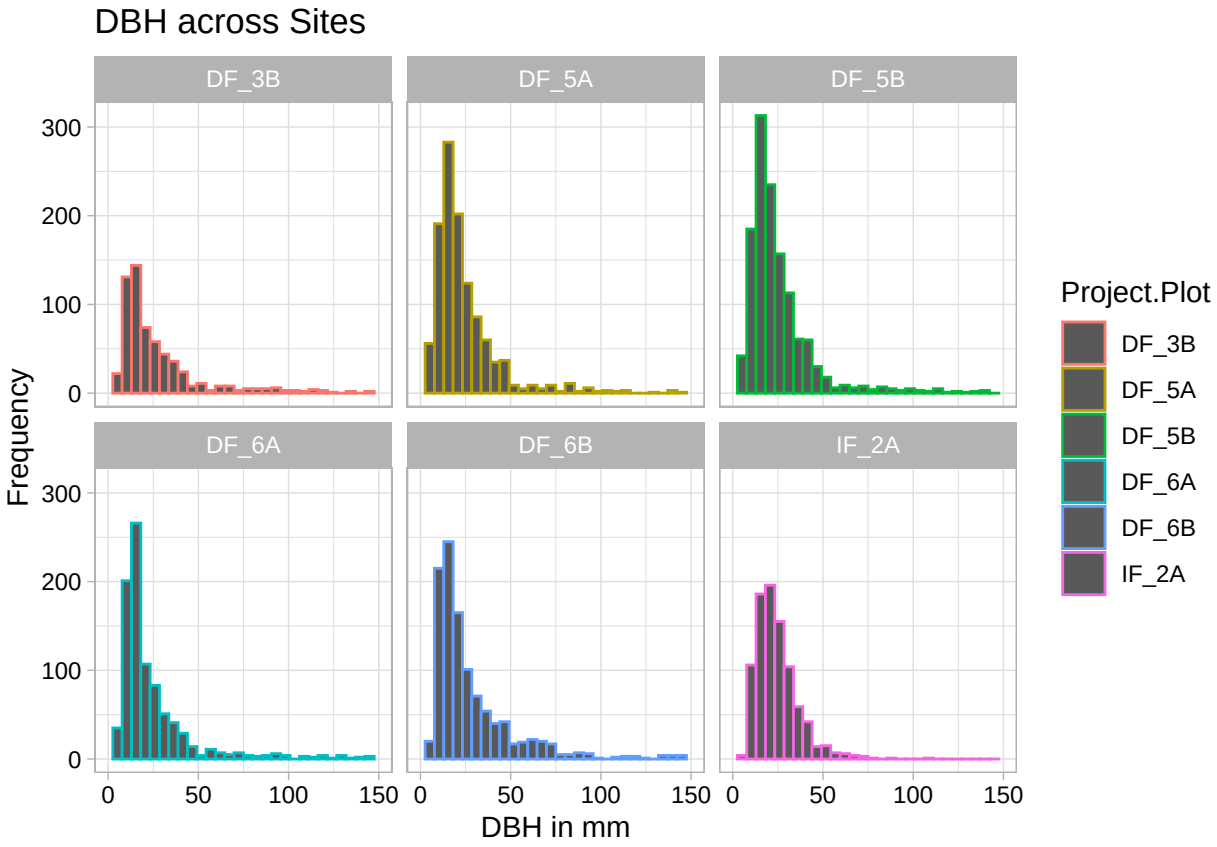

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



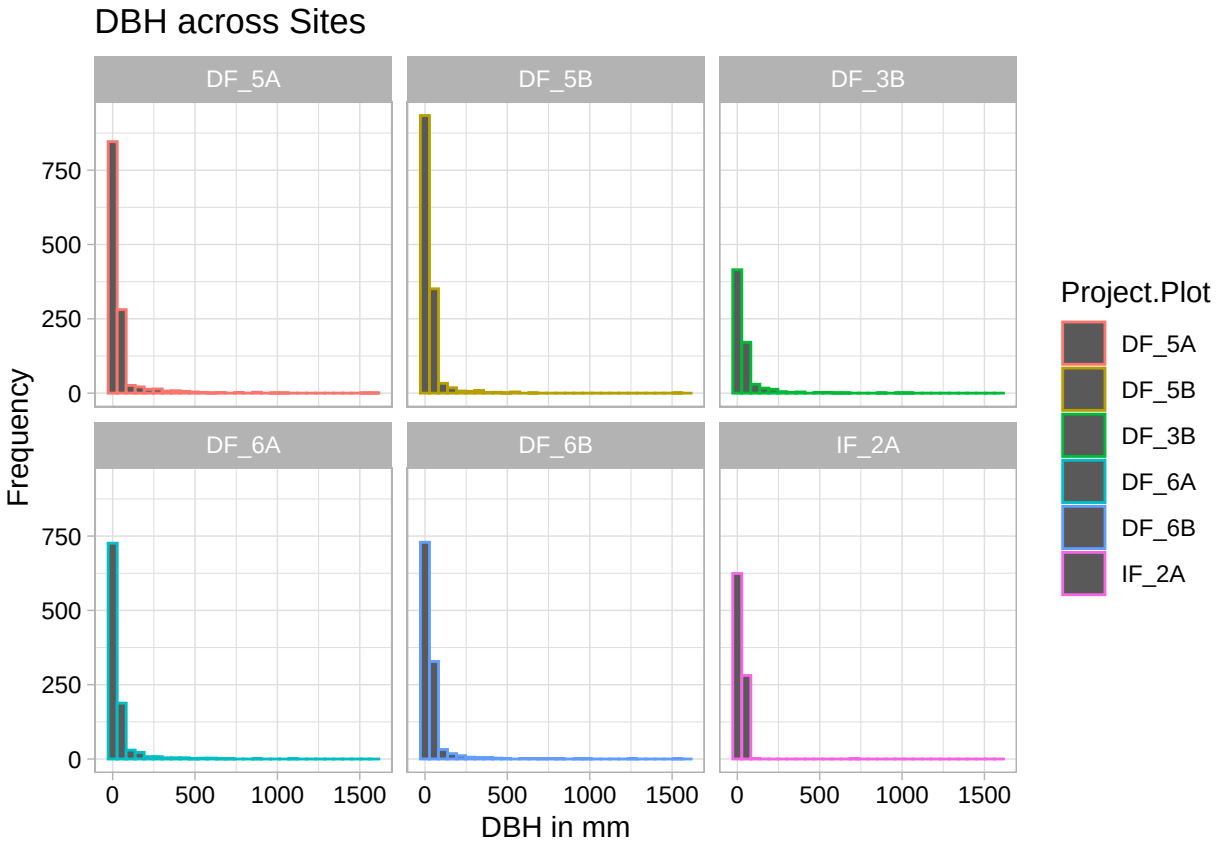

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 284 rows containing non-finite values (stat_bin).

## Warning: Removed 12 rows containing missing values (geom_bar).
```



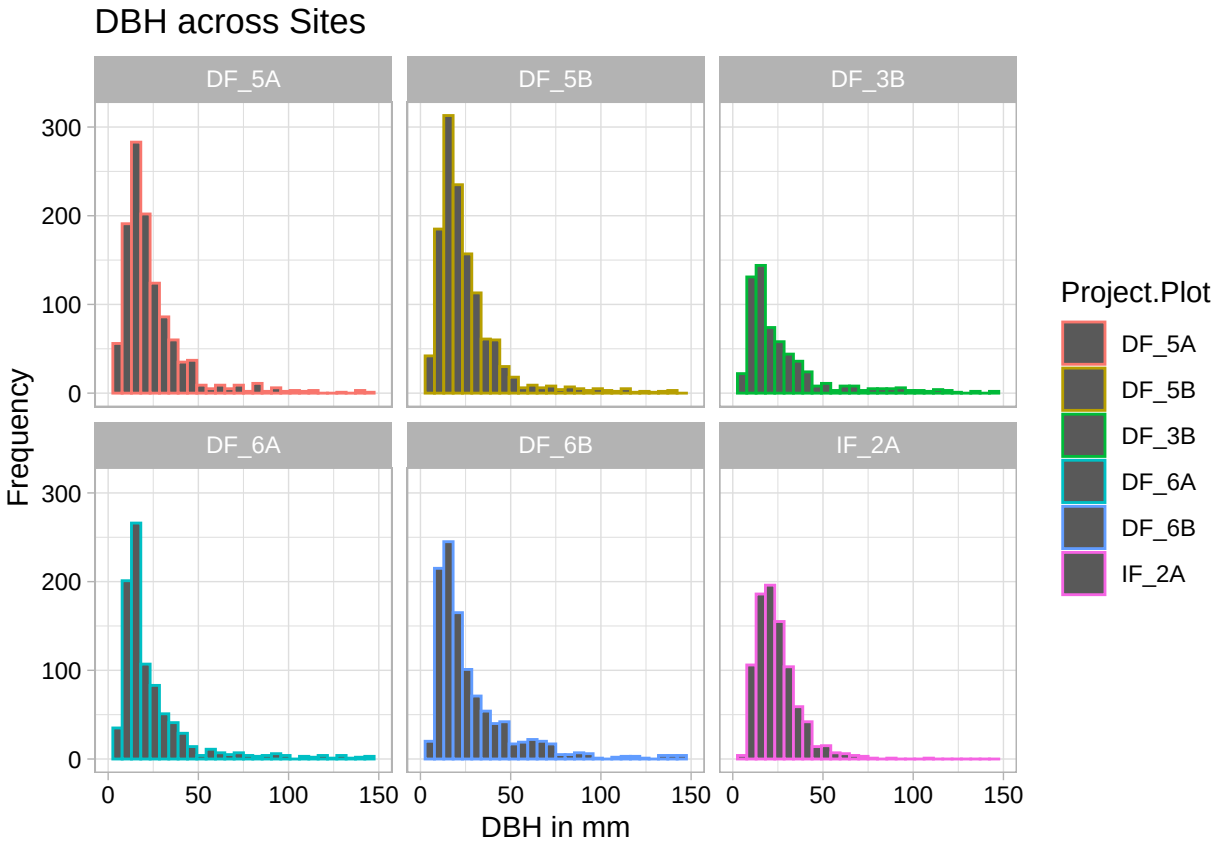

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 284 rows containing non-finite values (stat_bin).

## Warning: Removed 12 rows containing missing values (geom_bar).
```

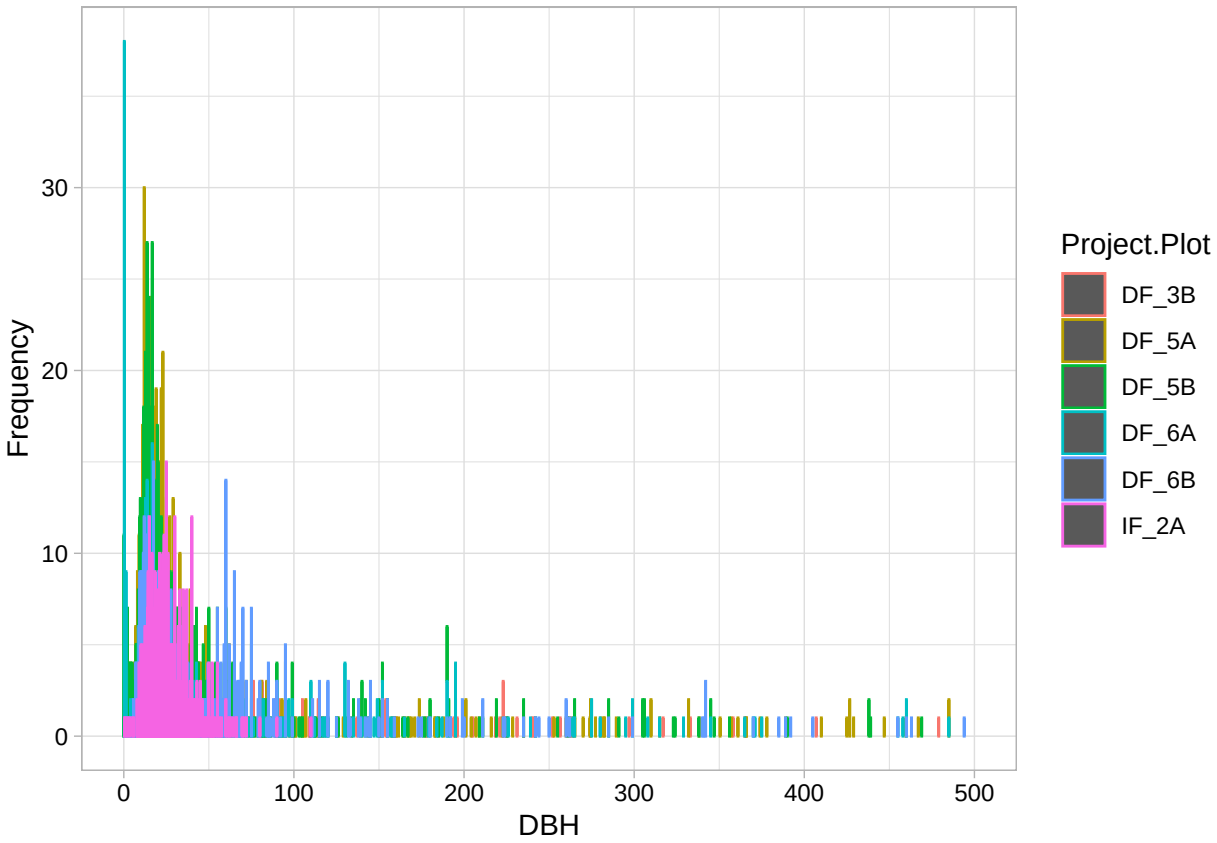



```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.39   14.00   20.15   48.09   32.02 1588.00

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.40   16.00   22.20   25.09   29.80  716.00

## Warning: Removed 53 rows containing non-finite values (stat_count).

## Warning: position_stack requires non-overlapping x intervals
```



4 Analysis

##GLMs

So it seems like most sites are pretty not uniform, and small trees are over-represented at each location. Let's compare sites to see which site has the most species.

So from this we can see that there are no great disparities between sites when it comes to species. At 146, DF_6B has the most species and at 96 DF_5B has the fewest species. Most sites have right around 100. We can also see how each site compares in terms of basal area - i.e., how densely forested each site is. IF_2A has the lowest basal area where DF_5A has the greatest at around 15 square meters per hectare. Given IF_2A's high species richness but low basal area, these data seem to suggest that IF_2A has many small trees but few large ones. Let's continue our exploration by seeing which genuses contribute the most to each site's basal area.

From these graphs we can see which genera make up the majority of basal area at each site. So how similar are these sites to one another? I wonder. ANOVA?

Let's begin to see if there is any relationship between basal area, species richness, and distance to towns (i.e., along the defaunation gradient)?.

Looking at these graphs, it doesn't appear that there's much of an obvious relationship between basal area, species richness, and distance along the defaunation gradient. Still, looks can be deceiving. Let's feed these data into a linear model to see what relationships can be statistically proved. We begin by checking for correlations among variables with a corrplot.

from this it appears that basal area per hectare is negatively correlated with mean distance at the same time, there appears to be a weak positive correlation between total species and mean distance. Let's build a model and see how well these correlations predict basal area and species richness.

In this linear model we use species richness and distance from developed area to predict the basal area per hectare of a given site. The null hypothesis is that there is no relationship between basal area per hectare, species richness, and distance to town. The alternative hypothesis is that either both species richness and/or distance to town will influence basal area per hectare. The results of the model ($p = 0.16$) indicate that no such significant ($p<0.05$) relationships exist in this subset of data. However, based on these observations ($n = 6$), there does appear to be a weak negative relationship ($p = 0.09$) between distance to town and basal area per hectare. According to the model, with every additional kilometer of distance there is an associated 1.3 square meter decline in basal area of forested land. Once again, these relationships are not considered significant, which means the explanatory variables included in this model are not sufficient to explain observed patterns in stand density along the defaunation gradient. There is no relationship between distance to town and species richness ($p = 0.74$).

4.1 Question 1: <insert specific question here and add additional subsections for additional questions below, if needed>

4.2 Question 2:

5 Summary and Conclusions

6 References

<add references here if relevant, otherwise delete this section>