Factors Influencing Above Ground Biomass of Forest Plots in the Congo

https://github.com/nikki-egna/ENV872_Final_Project_Congo_Carbon

Nikki Egna

Abstract

Add the abstract here

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Above ground biomass can be a predictor of overall health of a forest ecosystem. It is also important for looking at carbon storage potential of forests, which has significant implications for climate change mitigation. (Ducanson et al., 2019) This study aims to look at various environmental factors, and how they might contribute to differences in above ground biomass. The environmental factors included in this study are Human Footprint Index (HFI), GlobCover, annual precipitation, soil type, distance from the nearest road, distance from the nearest village, distance from the nearest river, distance from the nearest saw mill, and distance from the nearest protected area.

Question 1:

How has above ground biomass changed over time at the plot locations?

Question 2:

Which environmental factors are signicant predictors of above ground biomass?

2 Dataset Information

This repository uses data from the Poulsen Tropical Ecology Lab, and contains information on tree diameter and species from plots in the Congo. This data is being collected in order to model the amount of above ground biomass (AGB) within each plot across the study sites. Utilizing remote sensing techniques, the Poulsen Lab performed analysis to determine several covariate values at each plot, including distance to nearest road, distance to nearest village, distance to nearest river, distance to nearest protected area, Human Footprint Index, and Globcover value. This project will look at which covariates have a significant impact on AGB at the study sites.

The CongoCarbon_Raw_Data.csv was collected by the Poulsen Tropical Ecology Lab, and their on-ground team based in the Congo. This data is not public facing, and access was granted through Dr. John Poulsen. The CongoCarbon_AGB_by_Plot.csv was created by John Poulsen, Anna Nordseth, and Nikki Egna based on the raw data. The covariate data (CongoCarbon_Plot_Covariates.csv) was created by Nikki Egna in March of 2020, and the data it utilizes was obtained from the following sources: Human Footprint Index was derived from Wildlife Conservation Society (WCS), Center for International Earth Science Information Network (CIESIN), and Columbia University and was accessed in December of 2019. The GlobCover raster was downloaded from the ESA Globcover 2005 Project, led by MEDIAS-France/POSTEL, in December of 2019. Precipitation data was obtained from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) in December of 2019. The roads, river, soil, saw mill, and villages data comes from the Ministère de l'Économie Forestière CNIAF_MEFDD (https://cog.forest-atlas.org), accessed in April, 2020.

2.1 Metadata

2.1.1 CongoCarbon_Raw_data.csv

Column Name	Description
Tree ID	Identification number assigned to the unique tree (Numeric)
Tag No	Secondary tree identification number (Numeric)
Plot	Study plot number (Numeric)
Subplot	Subplot number within each plot (Numeric)
Family	Tree family name (Character)
Species	Tree species name (Character)
WD	Wood density (Numeric)
DBH0_05	First diameter measurement in 2005 (Numeric; Centimeters)
DBH1_05	Second diameter measurement in 2005 (Numeric; Centimeters)
DBH2_05	Third diameter measurement in 2005 (Numeric; Centimeters)
DBH3_05	Fourth diameter measurement in 2005 (Numeric; Centimeters)
DBH4_05	Fifth diameter measurement in 2005 (Numeric; Centimeters)
DBH0_09	First diameter measurement in 2009 (Numeric; Centimeters)

Column Name	Description
DBH1_09	Second diameter measurement in 2009 (Numeric; Centimeters)
DBH2_09	Third diameter measurement in 2009 (Numeric; Centimeters)
DBH3_09	Fourth diameter measurement in 2009 (Numeric; Centimeters)
DBH4_09	Fifth diameter measurement in 2009 (Numeric; Centimeters)
DBH0_13	First diameter measurement in 2013 (Numeric; Centimeters)
DBH1_13	Second diameter measurement in 2013 (Numeric; Centimeters)
DBH2_13	Third diameter measurement in 2013 (Numeric; Centimeters)
DBH3_13	Fourth diameter measurement in 2013 (Numeric; Centimeters)
DBH4_13	Fifth diameter measurement in 2013 (Numeric; Centimeters)
AGB05.MgE	AGB estimate 2005 (Numeric; milligrams)
AGB09.MgE	AGB estimate 2009 (Numeric; milligrams)
AGB13.MgE	AGB estimate 2013 (Numeric; milligrams)

2.1.2 CongoCarbon_Plot_Covariates.csv

Column Name	Description
Plot	Tree plot ID number (Factor)
Latitude	Latitude of the plot (Numeric)
Longitude	Longitude of the plot (Numeric)
HFI	Human Footprint Index value at the plot location (Factor)
GlobCover	GlobCover vegetation index value at the plot location (Factor)
Precip_sum_2013	Annual precipitation sum for 2013 at plot location (Numeric; millimeters)
Soil	Soil type index at plot location (Factor)
Dist_Road_m	Distance from plot to the nearest road (Numeric; Meters)
Dist_Village_m	Distance from plot to the nearest village (Numeric; Meters)
Dist_River_m	Distance from plot to the nearest river (Numeric; Meters)
Dist_PA_m	Distance from plot to the nearest protected area (Numeric; Meters)
Dist_Saw_Mills_m	Distance from plot to the nearest saw mill (Numeric; Meters)

3 Exploratory Analysis

3.1 Raw Data

```r
#Check first few rows of data and the structure of the data
head(Raw_CongoCarbon_Data)
str(Raw_CongoCarbon_Data)

#Check class of necessary columns
lapply(Raw_CongoCarbon_Data, class)

#Change necessary classes
Raw_CongoCarbon_Data$Plot <- as.factor(Raw_CongoCarbon_Data$Plot)

#Check for NAs
anyNA(Raw_CongoCarbon_Data$AGB05.MgE)
anyNA(Raw_CongoCarbon_Data$AGB09.MgE)
anyNA(Raw_CongoCarbon_Data$AGB13.MgE)

#Look at number of NA values
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB05.MgE)))
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB09.MgE)))
nrow(subset(Raw_CongoCarbon_Data, is.na(AGB13.MgE)))
```

3.2 Covariate Data

```r
head(CongoCarbon_Covariates)
str(CongoCarbon_Covariates)

#Check the class for all columns in CongoCarbon_Covariates
lapply(CongoCarbon_Covariates, class)
```

4 Analysis

Wrangle raw datasets to create a final dataframe with covariate values, sum of AGB for each year, and the change in AGB by year for each plot location

```
#Create columns with the sum AGB by plot for each year '05, '09, '13
AGB_by_Plot <- Raw_CongoCarbon_Data %>%
  group_by(Plot) %>%
  summarize(sum_AGB05 = sum(AGB05.MgE, na.rm = TRUE),
            sum_AGB09 = sum(AGB09.MgE, na.rm = TRUE),
            sum_AGB13 = sum(AGB13.MgE, na.rm = TRUE))

#Create columns for change in AGB from '05-'09, '09-'13, & '05-'13
AGB_by_Plot$change0509 <- with(AGB_by_Plot, sum_AGB09 - sum_AGB05)
AGB_by_Plot$change0913 <- with(AGB_by_Plot, sum_AGB13 - sum_AGB09)
AGB_by_Plot$change0513 <- with(AGB_by_Plot, sum_AGB13 - sum_AGB05)

#Combine AGB_by_Plot and CongoCarbon_Covariates by plot
AGB_and_Covariates_by_Plot <-
  merge(CongoCarbon_Covariates, AGB_by_Plot, by="Plot")
```

Prepare data in correct format for analysis and graphing

```
#Change dataframe from horizontal to vertical
melted_AGB_by_Plot <- melt(AGB_by_Plot, id=c("Plot"))
#Create a column for year
melted_AGB_by_Plot$Year <- melted_AGB_by_Plot$variable
melted_AGB_by_Plot$Year <- gsub("sum_AGB05","2005",
                                melted_AGB_by_Plot$Year, fixed=T)
melted_AGB_by_Plot$Year <- gsub("sum_AGB09","2009",
                                melted_AGB_by_Plot$Year, fixed=T)
melted_AGB_by_Plot$Year <- gsub("sum_AGB13","2013",
                                melted_AGB_by_Plot$Year, fixed=T)

#Create dataframe for the sum of AGB by plot
sum_AGB_by_plot <- melted_AGB_by_Plot[melted_AGB_by_Plot$variable
                                      %like% "sum", ]
#Create dataframe for the change in AGB by plot
change_AGB_by_plot <- melted_AGB_by_Plot[melted_AGB_by_Plot$variable
                                         %like% "change", ]
#Create dataframe for the change in AGB by year
AGB_change_by_year <- change_AGB_by_plot %>%
  group_by(variable) %>%
  summarise(sum=sum(value))
```

```r
#Raw data
#Add column for percent change
Raw_CongoCarbon_Data$Percent_Change_0513 <-
  ((Raw_CongoCarbon_Data$AGB13-Raw_CongoCarbon_Data$AGB05)/
     Raw_CongoCarbon_Data$AGB05.MgE)*100

#Subset necessary data
Raw_data_subset <- Raw_CongoCarbon_Data %>%
  dplyr::select(Tree.ID, Plot, AGB05.MgE, AGB09.MgE, AGB13.MgE)

#Melt data
melted_raw_data <- melt(Raw_data_subset, id=c("Tree.ID","Plot"))

#Create a column for year
melted_raw_data$Year <- melted_raw_data$variable
melted_raw_data$Year <- gsub("AGB05.MgE","2005",
                             melted_raw_data$Year, fixed=T)
melted_raw_data$Year <- gsub("AGB09.MgE","2009",
                             melted_raw_data$Year, fixed=T)
melted_raw_data$Year <- gsub("AGB13.MgE","2013",
                             melted_raw_data$Year, fixed=T)
```

4.1 Data visualization

4.2 Question 1: Is there significant difference in AGB between each of the years of data collection (2005, 2009, and 2013)?

```r
AGB.by.year.model <- lm(data=melted_raw_data, value ~ as.factor(Year))
summary(AGB.by.year.model)
```

```
##
## Call:
## lm(formula = value ~ as.factor(Year), data = melted_raw_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -0.91  -0.64  -0.56  -0.30 106.06
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.6136     0.0237   25.94   <2e-16 ***
```

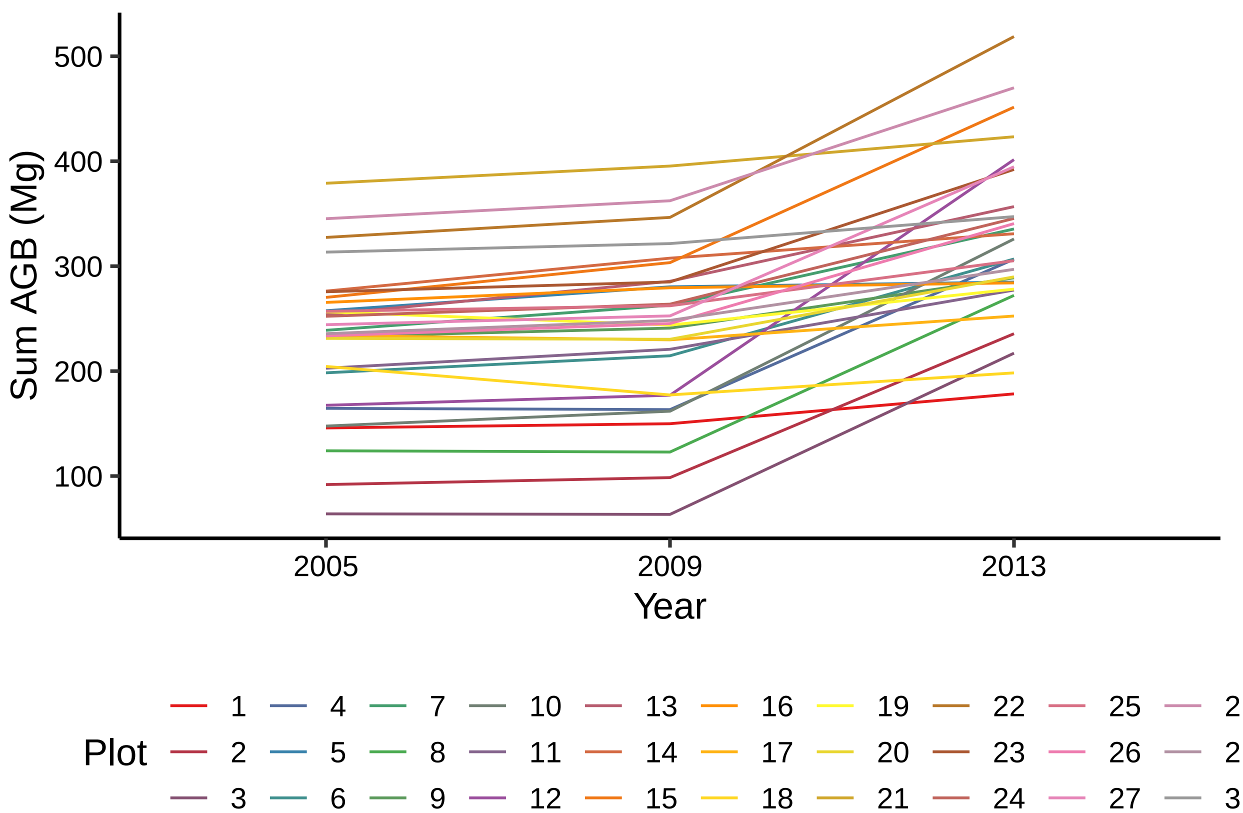


Figure 1: Relationship of the changes in total above ground biomass at each plot between 2005, 2009, and 2013.

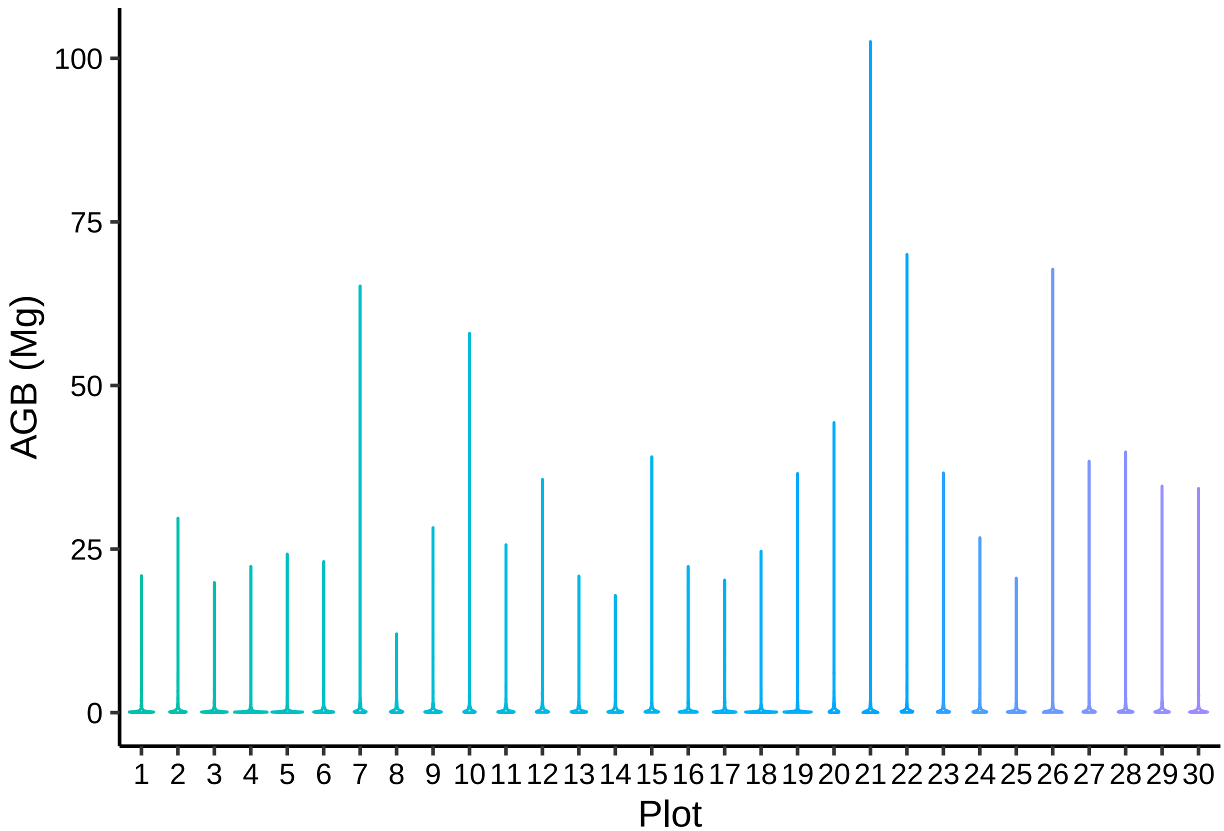


Figure 2: Relationship between above ground biomass at each plot in 2013.

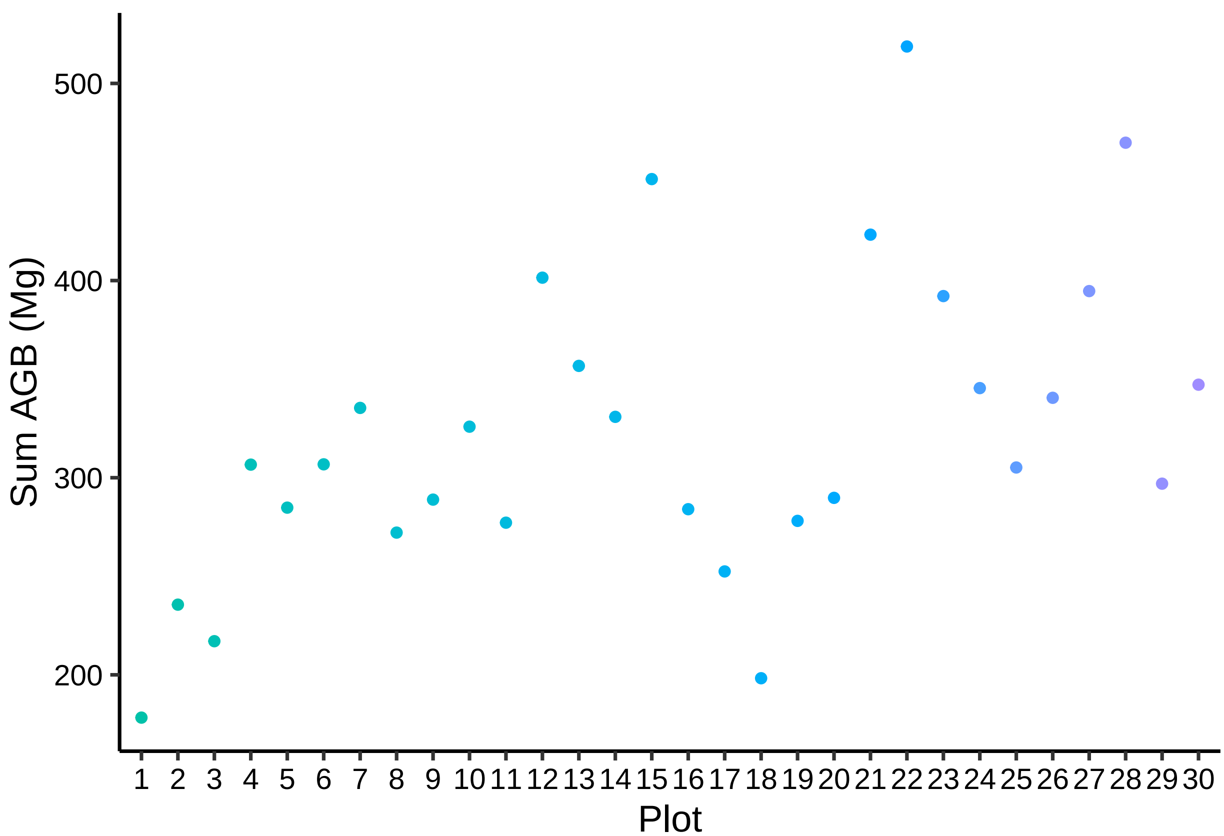


Figure 3: Relationship between total above ground biomass at each plot in 2013.

```
## as.factor(Year)2009   0.0509     0.0338    1.51     0.13    
## as.factor(Year)2013   0.2939     0.0339    8.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.51 on 32750 degrees of freedom
##   (4930 observations deleted due to missingness)
## Multiple R-squared:  0.0026, Adjusted R-squared:  0.00253 
## F-statistic: 42.6 on 2 and 32750 DF,  p-value: <2e-16
```

```r
TukeyHSD(aov(data=melted_raw_data, value ~ as.factor(Year)))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
## 
## Fit: aov(formula = value ~ as.factor(Year), data = melted_raw_data)
## 
## $`as.factor(Year)`
##             diff     lwr   upr p adj
## 2009-2005 0.0509 -0.0282 0.130 0.287
## 2013-2005 0.2939  0.2146 0.373 0.000
## 2013-2009 0.2430  0.1629 0.323 0.000
```

4.3 Question 2: Which environmental factors are significant predictors of AGB?

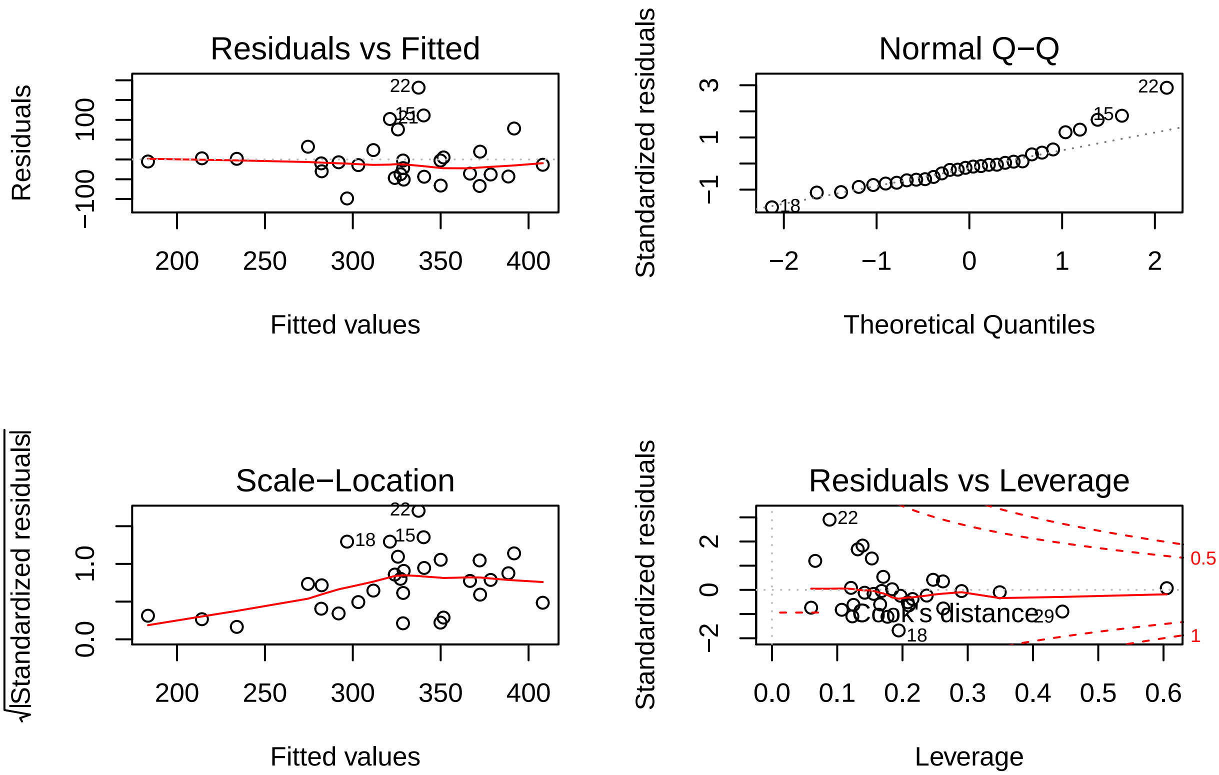
```r
#Main effects sum '13
AGB.main.13 <- lm(data = AGB_and_Covariates_by_Plot, sum_AGB13 ~
                    as.factor(HFI) + as.factor(GlobCover) + as.factor(Soil) +
                    Precip_sum_2013 + Dist_Road_m + Dist_Village_m + Dist_PA_m +
                    Dist_Saw_Mills_m)
#summary(AGB.main.13)

#step(AGB.main.13)

#Best model
AGB.main.13 <- lm(formula = sum_AGB13 ~ Precip_sum_2013 +
                    Dist_Road_m + Dist_Village_m + Dist_PA_m +
                    Dist_Saw_Mills_m, data = AGB_and_Covariates_by_Plot)
AGB.main.13
```

```
##
```

```
## Call:
## lm(formula = sum_AGB13 ~ Precip_sum_2013 + Dist_Road_m + Dist_Village_m +
##     Dist_PA_m + Dist_Saw_Mills_m, data = AGB_and_Covariates_by_Plot)
##
## Coefficients:
##      (Intercept)   Precip_sum_2013       Dist_Road_m    Dist_Village_m
##        -2.20e+02          3.84e+00          9.83e-03          5.78e-03
##        Dist_PA_m  Dist_Saw_Mills_m
##        -7.54e-03         -1.03e-02
```

```r
summary(AGB.main.13)
```

```
##
## Call:
## lm(formula = sum_AGB13 ~ Precip_sum_2013 + Dist_Road_m + Dist_Village_m +
##     Dist_PA_m + Dist_Saw_Mills_m, data = AGB_and_Covariates_by_Plot)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -98.5  -37.8   -8.5   16.1  181.3
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.20e+02   4.13e+02   -0.53    0.599
## Precip_sum_2013   3.84e+00   2.23e+00    1.72    0.099 .
## Dist_Road_m       9.83e-03   5.96e-03    1.65    0.112
## Dist_Village_m    5.78e-03   3.33e-03    1.73    0.096 .
## Dist_PA_m        -7.54e-03   3.71e-03   -2.03    0.053 .
## Dist_Saw_Mills_m -1.03e-02   5.26e-03   -1.95    0.062 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 65.4 on 24 degrees of freedom
## Multiple R-squared:  0.427,  Adjusted R-squared:  0.307
## F-statistic: 3.57 on 5 and 24 DF,  p-value: 0.0148
```


Residuals vs Fitted
Residuals
Fitted values
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Cook's distance
Leverage

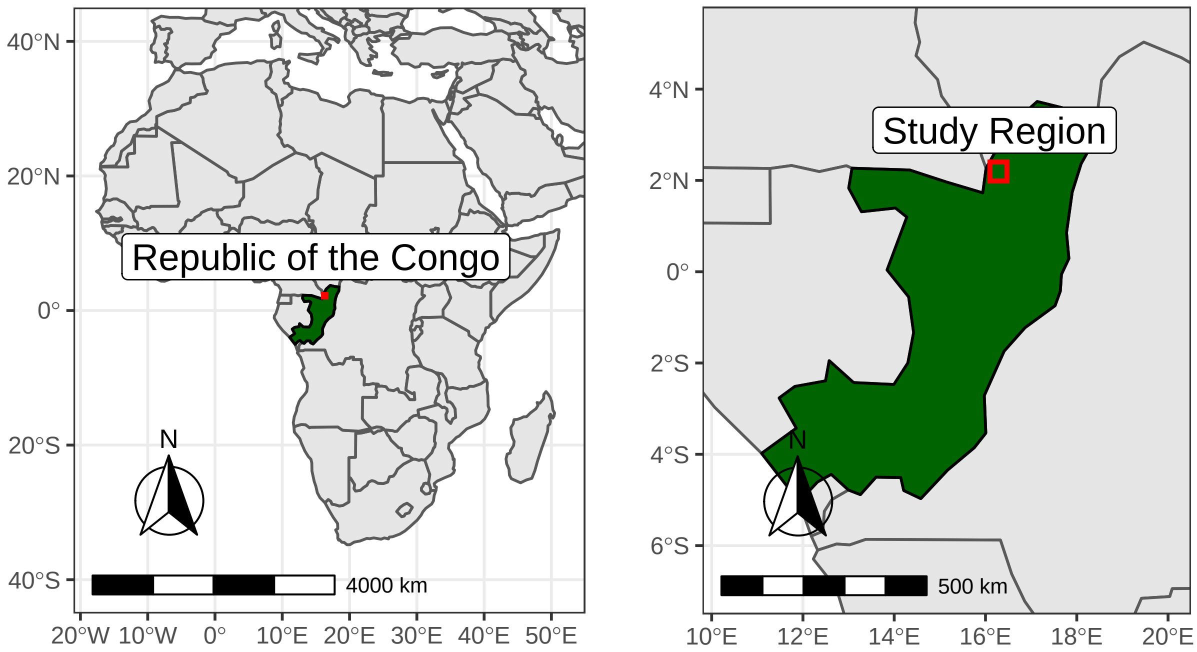


Figure 4: Map displaying the location of the forest plots within the Republic of Congo, Africa.

```
##        Plot    Latitude   Longitude
## "character"   "numeric"   "numeric"
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=2.195,16.28&zoom=11&si
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=2.25,16.25&zoom=10&siz
```

```
## Scale on map varies by more than 10%, scale bar may be inaccurate
## Scale on map varies by more than 10%, scale bar may be inaccurate
```

```
## Coordinate system already present. Adding new coordinate system, which will replace t
```

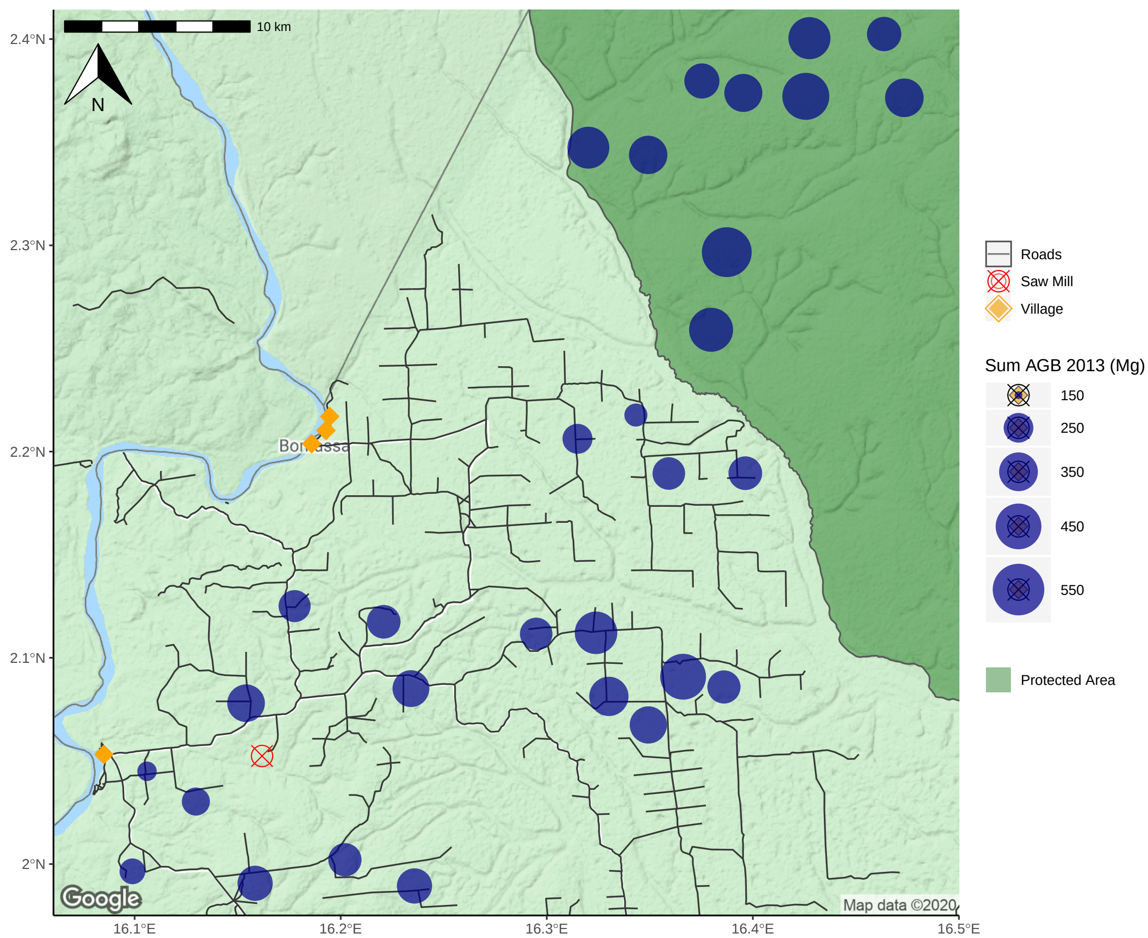


Figure 5: Sum of above ground biomass (Mg) at each plot location in 2013. Nearby road, village, saw mill, and protected area locations are plotted to visualize the relationship between these variables and the amount of above ground biomass.

5 Summary and Conclusions

6 References

Duncanson, L., Armston, J., Disney, M. et al. The Importance of Consistent Global Forest Aboveground Biomass Product Validation. Surv Geophys 40, 979–999 (2019). https://doi.org/10.1007/s10712-019-09538-8 <add references here if relevant, otherwise delete this section>