

Datenvisualisierung in R mit ggplot2

Jens Müller

2025-04-14

```
library(readxl) # for reading Excel files
library(ggplot2) # for creating charts and plots
library(RColorBrewer) # a popular color palette
library(forcats) # for sorting factor variables
library(sf) # for reading shape files

## Linking to GEOS 3.13.1, GDAL 3.10.3, PROJ 9.6.0; sf_use_s2() is TRUE
library(dplyr) # for data manipulation

##
## Attaching package: 'dplyr'

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

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

library(tidyr) # for drop_na
```

Datenvisualisierung

Dates des Marteloskops Möllergrab laden

```
df <- read_excel('data/Marteloscope_Möllergrab_2021_DE.xlsx', sheet = '2021')

## New names:
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`

df

## # A tibble: 895 x 15
##   Zahl `Baum Nr` Baumart `d 1.3 [cm]` `h [m]` QMP Zustand `G [m²]` `V [m³]`
##   <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1 1 1 RBU 9.6 8.9 1 1 0.00724 0.0285
## 2 2 2 2 RBU 16.8 14 1 1 0.0222 0.0714
## 3 3 3 3 RBU 13.6 14.4 1 1 0.0145 0.0634
## 4 4 4 4 RBU 11.5 8.7 1 1 0.0104 0.0304
```

```
## 5      5      5 GKI      39.5    30.6    1      1  0.123    1.63
## 6      6      6 RBU      8        8.2    1      1  0.00503  0.0247
## 7      7      7 GKI      40.6    30.4    1      1  0.129    1.71
## 8      8      8 GKI      32.3    28      1      1  0.0819   1.01
## 9      9      9 GFI      7.6     7.6    1      1  0.00454  0.0183
## 10     10     10 RBU     17.6    17.6    1      1  0.0243   0.108
## # i 885 more rows
## # i 6 more variables: `Nr. TReMs` <dbl>, ...11 <lgl>, ...12 <chr>, ...13 <chr>,
## #   ...14 <chr>, ...15 <chr>
```

Beschreibende Statistiken

```
summary(df)
```

```
##      Zahl      Baum Nr      Baumart      d 1.3 [cm]
## Min.   : 1.0    Min.   : 1.0    Length:895    Min.   : 6.80
## 1st Qu.:224.2  1st Qu.:235.2  Class :character 1st Qu.:12.10
## Median :447.5  Median :464.5  Mode  :character Median :17.10
## Mean   :447.5  Mean   :461.9                Mean   :21.76
## 3rd Qu.:670.8  3rd Qu.:691.5                3rd Qu.:32.80
## Max.   :894.0  Max.   :917.0                Max.   :62.80
## NA's   :1      NA's   :1                NA's   :1
##      h [m]      QMP      Zustand      G [m²]
## Min.   : 5.60   Min.   :1.000   Min.   :0.0000   Min.   : 0.00363
## 1st Qu.:20.43   1st Qu.:1.000   1st Qu.:1.0000   1st Qu.: 0.01150
## Median :28.40   Median :2.000   Median :1.0000   Median : 0.02297
## Mean   :25.29   Mean   :2.404   Mean   :0.9899   Mean   : 0.09658
## 3rd Qu.:30.00   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.: 0.08488
## Max.   :37.20   Max.   :4.000   Max.   :1.0000   Max.   :43.17020
## NA's   :495     NA's   :1      NA's   :1      NA's   :1
##      V [m³]      Nr. TReMs      ...11      ...12
## Min.   :0.00000   Min.   :0.0000   Mode:logical   Length:895
## 1st Qu.:0.01559   1st Qu.:0.0000   NA's:895       Class :character
## Median :0.01559   Median :0.0000                Mode  :character
## Mean   :0.47054   Mean   :0.6437
## 3rd Qu.:1.05605   3rd Qu.:1.0000
## Max.   :3.77631   Max.   :7.0000
## NA's   :3         NA's   :8
##      ...13      ...14      ...15
## Length:895     Length:895     Length:895
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##
```

```
# we ignore the missing data for this notebook
```

```
df_clean <- df |>
  drop_na(`h [m]`)
df_clean
```

```
## # A tibble: 400 x 15
##      Zahl `Baum Nr` Baumart `d 1.3 [cm]` `h [m]` QMP Zustand `G [m²]` `V [m³]`
##      <dbl>      <dbl> <chr>      <dbl>      <dbl> <dbl> <dbl>      <dbl>      <dbl>
```

```
## 1      1      1 RBU      9.6      8.9      1      1 0.00724 0.0285
## 2      2      2 RBU     16.8     14      1      1 0.0222 0.0714
## 3      3      3 RBU     13.6     14.4    1      1 0.0145 0.0634
## 4      4      4 RBU     11.5      8.7    1      1 0.0104 0.0304
## 5      5      5 GKI     39.5     30.6    1      1 0.123  1.63
## 6      6      6 RBU      8       8.2    1      1 0.00503 0.0247
## 7      7      7 GKI     40.6     30.4    1      1 0.129  1.71
## 8      8      8 GKI     32.3     28     1      1 0.0819 1.01
## 9      9      9 GFI      7.6      7.6    1      1 0.00454 0.0183
## 10     10     10 RBU     17.6     17.6    1      1 0.0243 0.108
## # i 390 more rows
## # i 6 more variables: `Nr. TReMs` <dbl>, ...11 <lgl>, ...12 <chr>, ...13 <chr>,
## #   ...14 <chr>, ...15 <chr>
```

```
summary(df_clean)
```

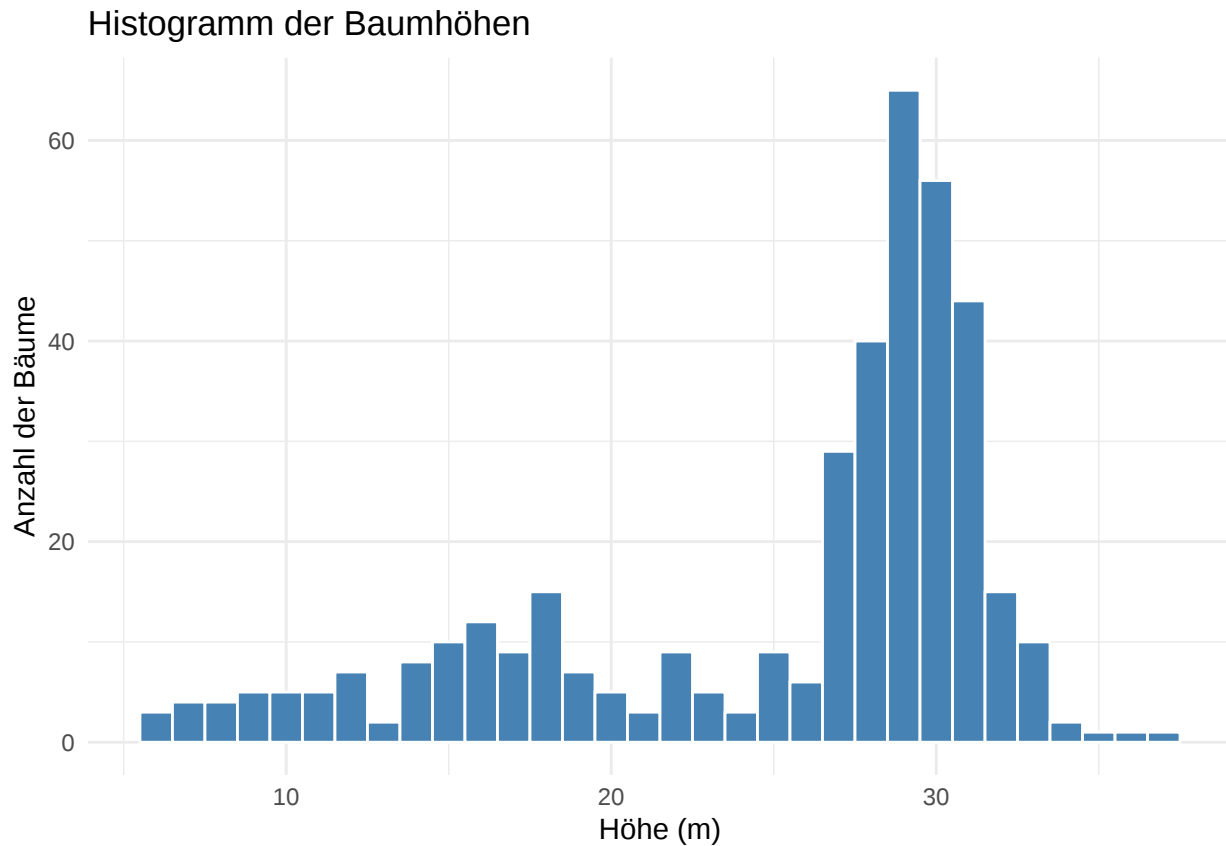
```
##      Zahl      Baum Nr      Baumart      d 1.3 [cm]
## Min.   : 1.0   Min.   : 1.0   Length:400   Min.   : 7.20
## 1st Qu.:121.8  1st Qu.:123.8  Class :character 1st Qu.:16.70
## Median :255.0  Median :272.5  Mode  :character  Median :33.85
## Mean   :337.4  Mean   :346.7           Mean   :29.83
## 3rd Qu.:543.8  3rd Qu.:562.8           3rd Qu.:39.02
## Max.   :893.0  Max.   :885.0           Max.   :62.80
##
##      h [m]      QMP      Zustand      G [m²]
## Min.   : 5.60   Min.   :1.000   Min.   :0.0000   Min.   :0.004071
## 1st Qu.:20.43   1st Qu.:1.000   1st Qu.:1.0000   1st Qu.:0.021904
## Median :28.40   Median :1.000   Median :1.0000   Median :0.089993
## Mean   :25.29   Mean   :1.942   Mean   :0.9875   Mean   :0.082172
## 3rd Qu.:30.00   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:0.119612
## Max.   :37.20   Max.   :4.000   Max.   :1.0000   Max.   :0.309748
##
##      V [m³]      Nr. TReMs      ...11      ...12
## Min.   :0.01835   Min.   :0.000   Mode:logical   Length:400
## 1st Qu.:0.13604   1st Qu.:0.000   NA's:400       Class :character
## Median :1.14315   Median :1.000           Mode  :character
## Mean   :1.03240   Mean   :1.144
## 3rd Qu.:1.55537   3rd Qu.:2.000
## Max.   :3.77631   Max.   :6.000
##      NA's :5
##      ...13      ...14      ...15
## Length:400      Length:400      Length:400
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
##
##
##
##
```

Visualisierungen

```
# consistent colors for tree species
colors_tree_species <- brewer.pal(8, "Set2")
```

Histogramm

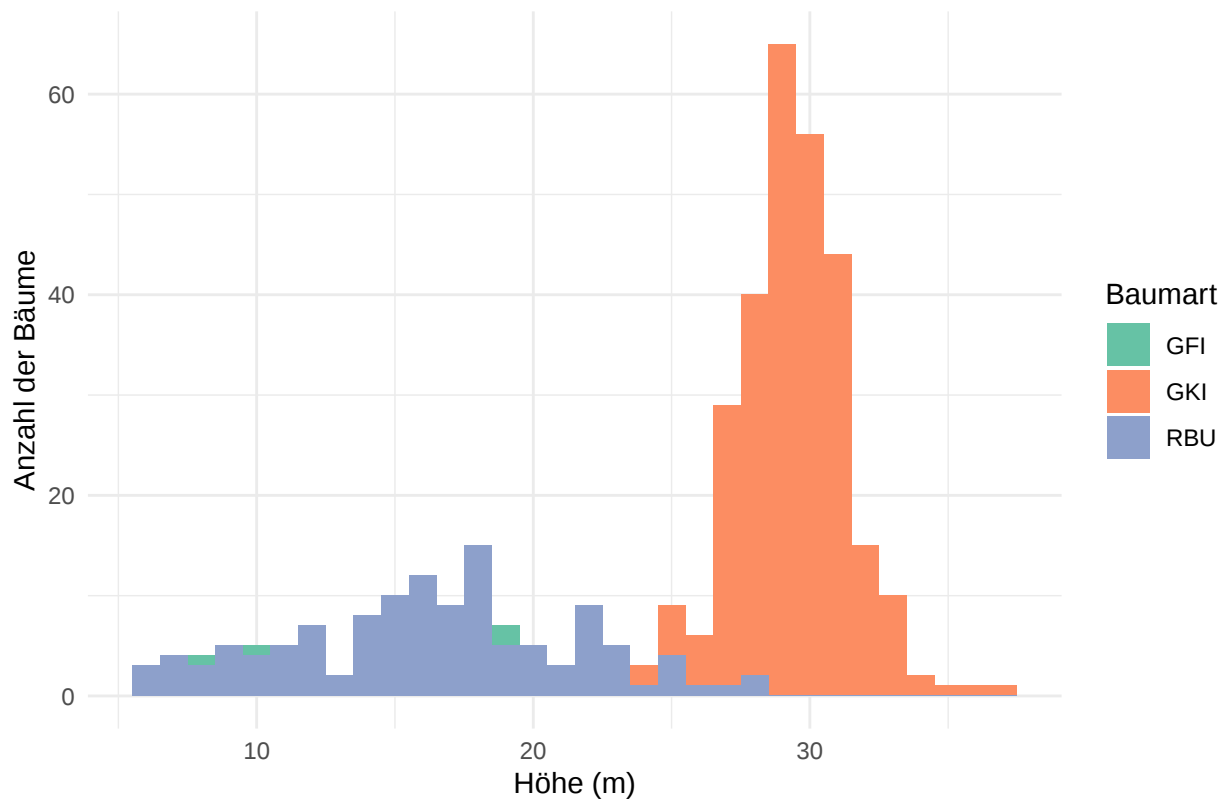
```
ggplot(data = df_clean, aes(x = `h [m]`)) +  
  geom_histogram(binwidth=1, fill = 'steelblue', color = 'white') +  
  labs(title = 'Histogramm der Baumhöhen',  
        x = 'Höhe (m)',  
        y = 'Anzahl der Bäume') +  
  theme_minimal()
```



Gestapeltes Säulendiagramm (Stacked Bar Chart)

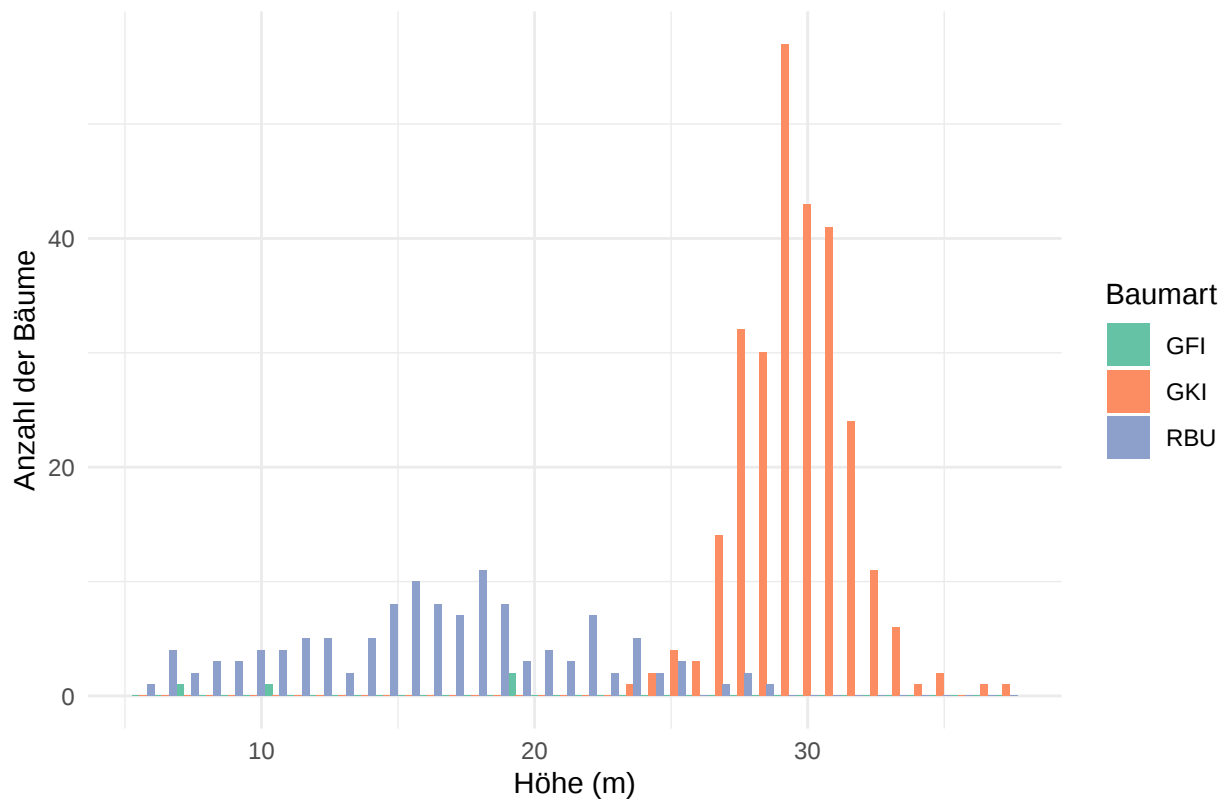
```
ggplot(data = df_clean, aes(x = `h [m]`, fill = factor(`Baumart`))) +  
  geom_histogram(binwidth=1) +  
  labs(title = 'Histogramm der Baumhöhen gestapelt nach Baumart',  
        x = 'Höhe (m)',  
        y = 'Anzahl der Bäume',  
        fill = 'Baumart') +  
  scale_fill_manual(values = colors_tree_species) +  
  theme_minimal()
```

Histogramm der Baumhöhen gestapelt nach Baumart



```
ggplot(data = df_clean, aes(x = `h [m]`, fill = factor(`Baumart`))) +
  geom_histogram(position = 'dodge', bins = 40) +
  labs(title = 'Histogramm der Baumhöhen nach Baumart',
       x = 'Höhe (m)',
       y = 'Anzahl der Bäume',
       fill = 'Baumart') +
  scale_fill_manual(values = colors_tree_species) +
  theme_minimal()
```

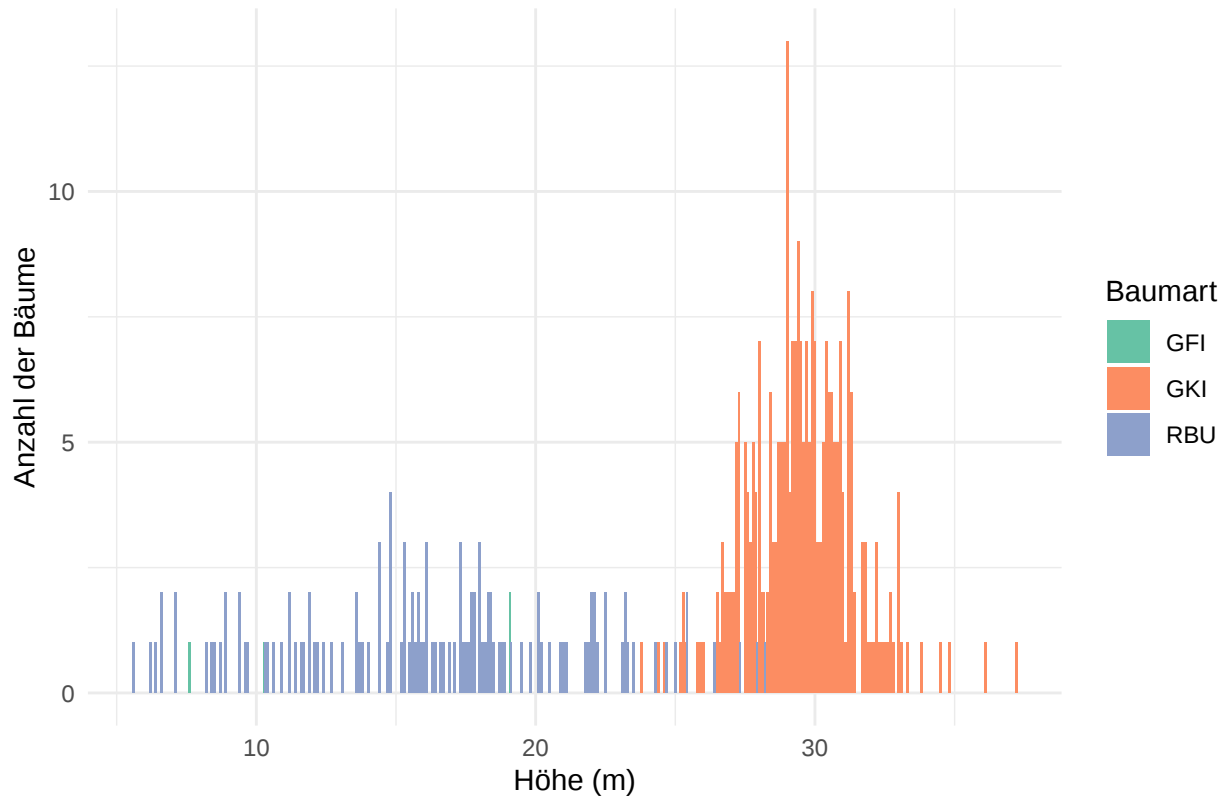
Histogramm der Baumhöhen nach Baumart



Säulendiagramm

```
ggplot(data = df_clean, aes(x = `h [m]`, fill = factor(`Baumart`))) +
  geom_bar(position = 'dodge') +
  labs(title = 'Säulendiagramm der Baumhöhen mit Baumarten',
        x = 'Höhe (m)',
        y = 'Anzahl der Bäume',
        fill = 'Baumart') +
  scale_fill_manual(values = colors_tree_species) +
  theme_minimal()
```

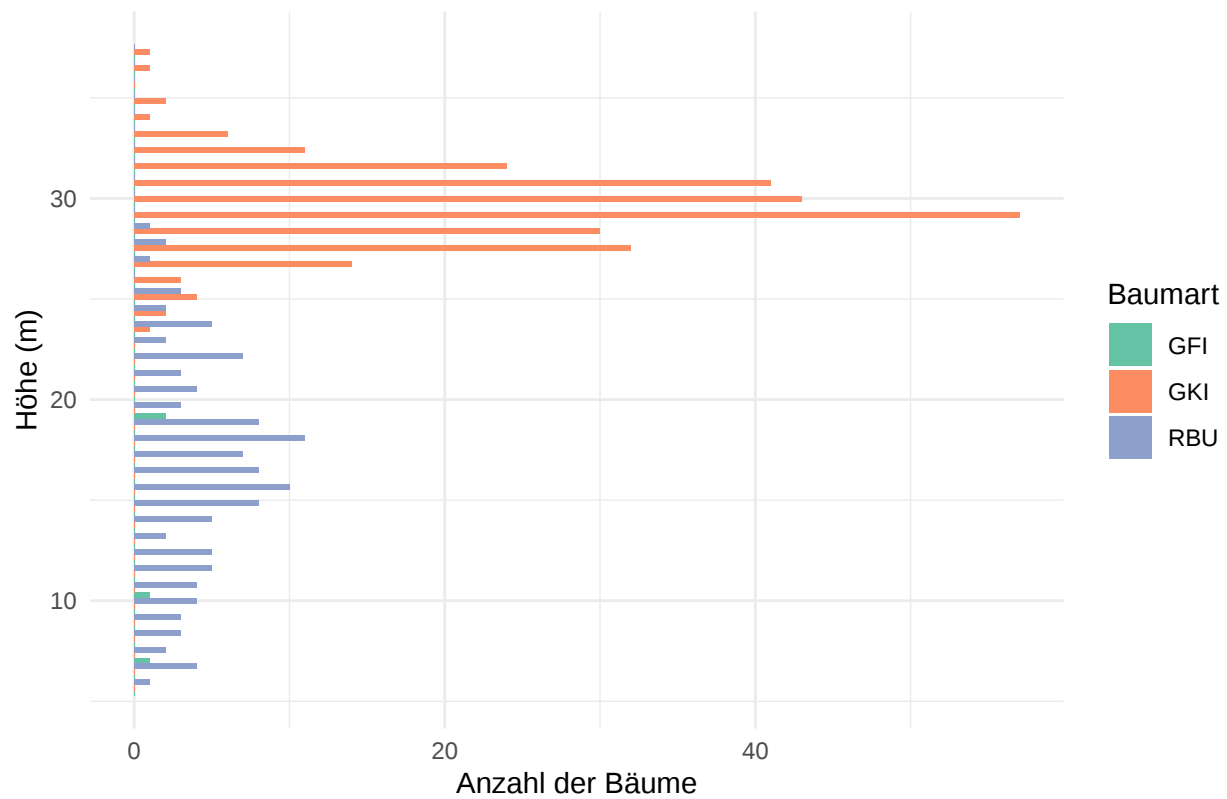
Säulendiagramm der Baumhöhen mit Baumarten



Balkendiagramm

```
ggplot(data = df_clean, aes(x = `h [m]`, fill = factor(`Baumart`))) +
  geom_histogram(position = 'dodge', bins = 40) +
  coord_flip() + # for a horizontal bar chart
  labs(title = 'Balkendiagramm der Baumhöhen mit Baumarten',
        x = 'Höhe (m)',
        y = 'Anzahl der Bäume',
        fill = 'Baumart') +
  scale_fill_manual(values = colors_tree_species) +
  theme_minimal()
```

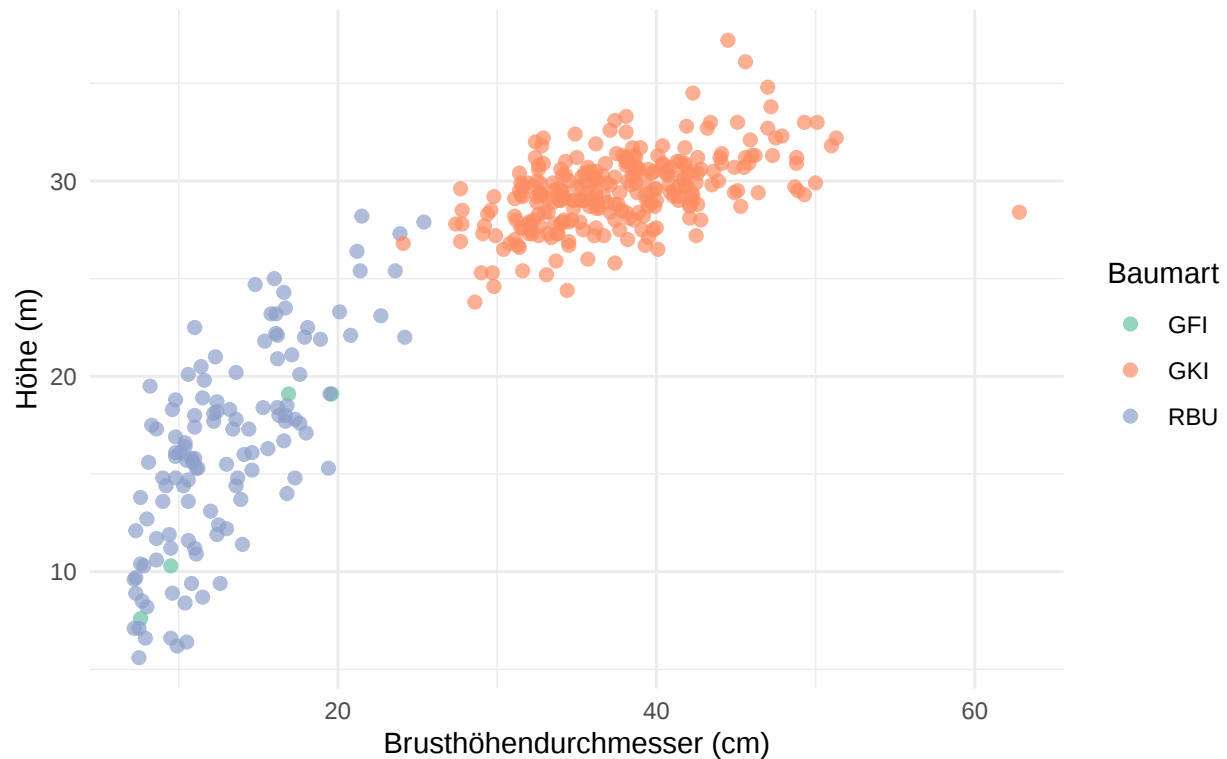
Balkendiagramm der Baumhöhen mit Baumarten



Streudiagramm (Scatter Plot)

```
ggplot(data = df_clean, aes(x = `d 1.3 [cm]`, y = `h [m]`, color = `Baumart`)) +
  geom_point(size = 2, alpha = 0.7) +
  labs(title = 'Brusthöhendurchmesser und Baumhöhe für unterschiedliche Baumarten\nim Marteloskop Mölle',
        x = 'Brusthöhendurchmesser (cm)',
        y = 'Höhe (m)') +
  scale_color_manual(values = colors_tree_species) +
  theme_minimal()
```

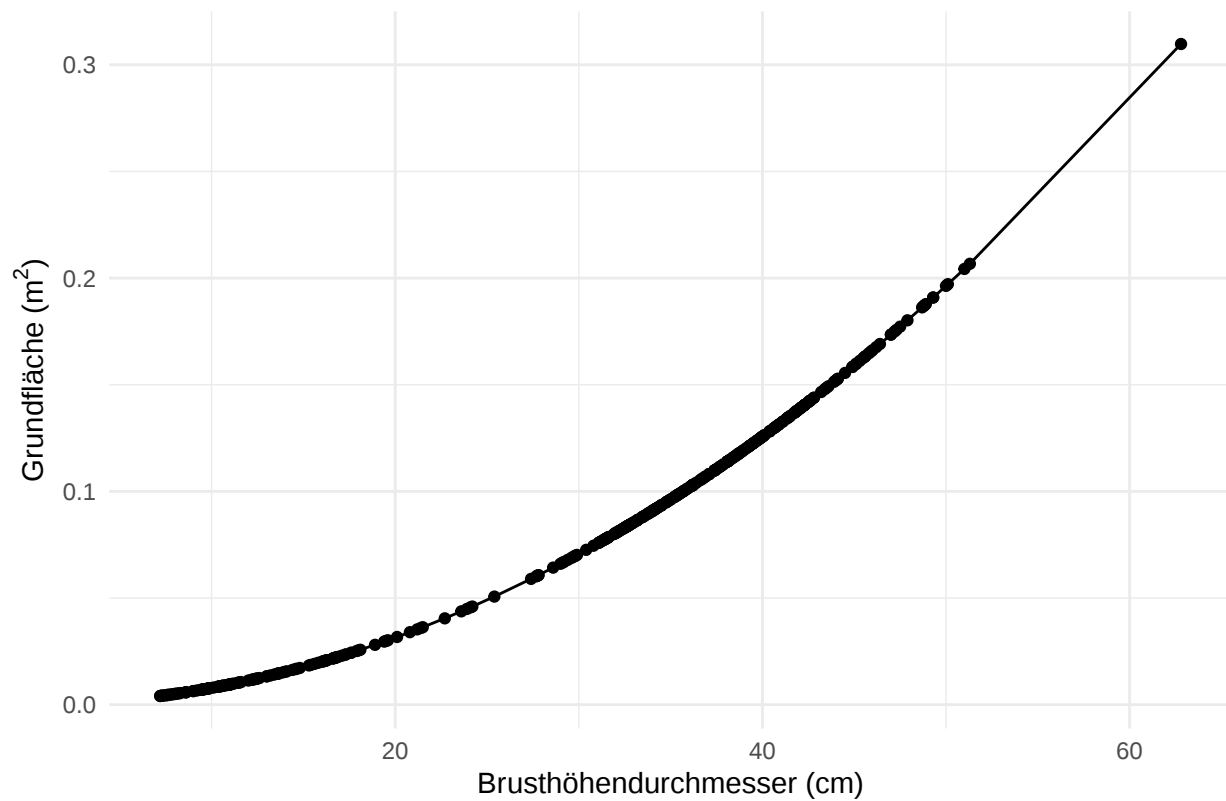
Brusthöhendurchmesser und Baumhöhe für unterschiedliche Baumarten im Marteloskop Möllergrab in 2021



Liniendiagramm (Line Chart)

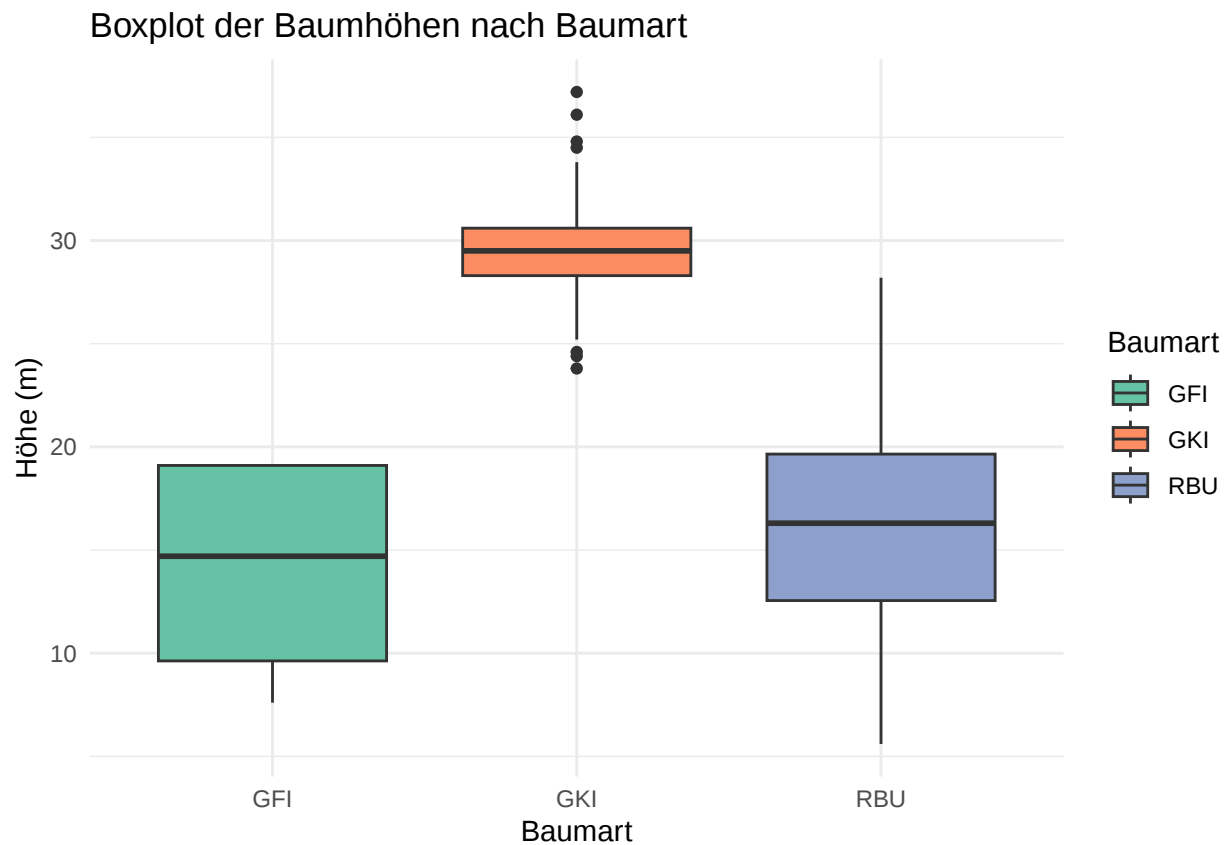
```
ggplot(data = df_clean |> arrange(`d 1.3 [cm]`), aes(x = `d 1.3 [cm]`, y = `G [m²]`)) +
  geom_point() +
  geom_line() +
  labs(title = 'Grundflächen in Abhängigkeit zum BHD im Marteloskop Möllergrab in 2021',
        x = 'Brusthöhendurchmesser (cm)',
        y = expression('Grundfläche (' * m^2 * ')')) +
  theme_minimal()
```

Grundflächen in Abhängigkeit zum BHD im Marteloskop Möllergrab in 202



Boxplot

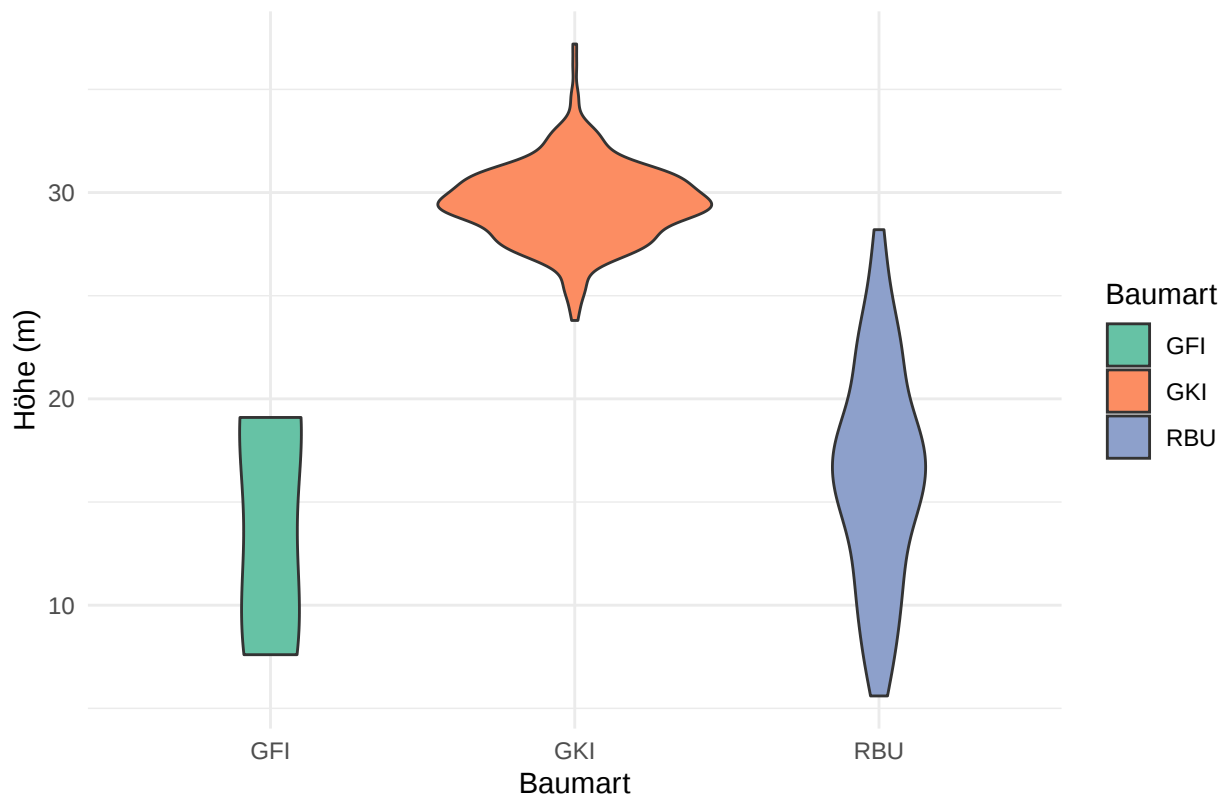
```
ggplot(data = df_clean, aes(x = factor(`Baumart`), y = `h [m]`, fill = factor(`Baumart`))) +  
  geom_boxplot() +  
  labs(title = 'Boxplot der Baumhöhen nach Baumart',  
        x = 'Baumart',  
        y = 'Höhe (m)',  
        fill = 'Baumart') +  
  scale_fill_manual(values = colors_tree_species) +  
  theme_minimal()
```



Violinplot

```
ggplot(data = df_clean, aes(x = factor(`Baumart`), y = `h [m]`, fill = factor(`Baumart`))) +
  geom_violin() +
  labs(title = 'Boxplot der Baumhöhen nach Baumart',
        x = 'Baumart',
        y = 'Höhe (m)',
        fill = 'Baumart') +
  scale_fill_manual(values = colors_tree_species) +
  theme_minimal()
```

Boxplot der Baumhöhen nach Baumart



Karten

```
quadrants <- st_read('data/Geodaten_Moellergrab/EW_Marteloscop_A0Ipolygons_32633.shp')
```

```
## Reading layer `EW_Marteloscop_A0Ipolygons_32633' from data source
##   `/Users/jkm/sync/projects/it_im_wald/data/Geodaten_Moellergrab/EW_Marteloscop_A0Ipolygons_32633.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 4 features and 7 fields
## Geometry type: POLYGON
## Dimension:      XY
## Bounding box:   xmin: 419227.8 ymin: 5851552 xmax: 419352.6 ymax: 5851677
## Projected CRS: WGS 84 / UTM zone 33N
```

quadrants

```
## Simple feature collection with 4 features and 7 fields
## Geometry type: POLYGON
## Dimension:      XY
## Bounding box:   xmin: 419227.8 ymin: 5851552 xmax: 419352.6 ymax: 5851677
## Projected CRS: WGS 84 / UTM zone 33N
##
```

	Source	AREA	PERIMETER	ANGLE	WIDTH
## 1	EW_Marteloscope_subA0I4_polygon_32633	2500.000	200	-73.00000	50.00000
## 2	EW_Marteloscope_subA0I3_polygon_32633	2500.000	200	107.00000	50.00000
## 3	EW_Marteloscope_subA0I2_polygon_32633	2500.000	200	-73.00001	50.00000
## 4	EW_Marteloscope_subA0I1_polygon_32633	2500.001	200	-73.00001	50.00001

```
## HEIGHT subID geometry
## 1 50 4 POLYGON ((419290.2 5851614,...
## 2 50 3 POLYGON ((419304.8 5851567,...
## 3 50 2 POLYGON ((419227.8 5851648,...
```

```
## 4      50      1 POLYGON ((419275.6 5851662,...
```

```
st_crs(quadrants)$epsg
```

```
## [1] 32633
```

```
ggplot(data = quadrants) +  
  geom_sf() +  
  geom_sf_text(aes(label = subID), size = 5) +  
  theme_minimal()
```



```
trees <- st_read('data/Geodaten_Moellergrab/EW_Marteloscop_TreeXY_32633.shp')
```

```
## Reading layer `EW_Marteloscop_TreeXY_32633' from data source  
##   `/Users/jkm/sync/projects/it_im_wald/data/Geodaten_Moellergrab/EW_Marteloscop_TreeXY_32633.shp'  
##   using driver `ESRI Shapefile'  
## Simple feature collection with 857 features and 7 fields  
## Geometry type: POINT  
## Dimension:      XY  
## Bounding box:   xmin: 419229.8 ymin: 5851556 xmax: 419351.5 ymax: 5851675  
## Projected CRS: WGS 84 / UTM zone 33N
```

```
trees
```

```
## Simple feature collection with 857 features and 7 fields  
## Geometry type: POINT  
## Dimension:      XY  
## Bounding box:   xmin: 419229.8 ymin: 5851556 xmax: 419351.5 ymax: 5851675  
## Projected CRS: WGS 84 / UTM zone 33N  
## First 10 features:
```

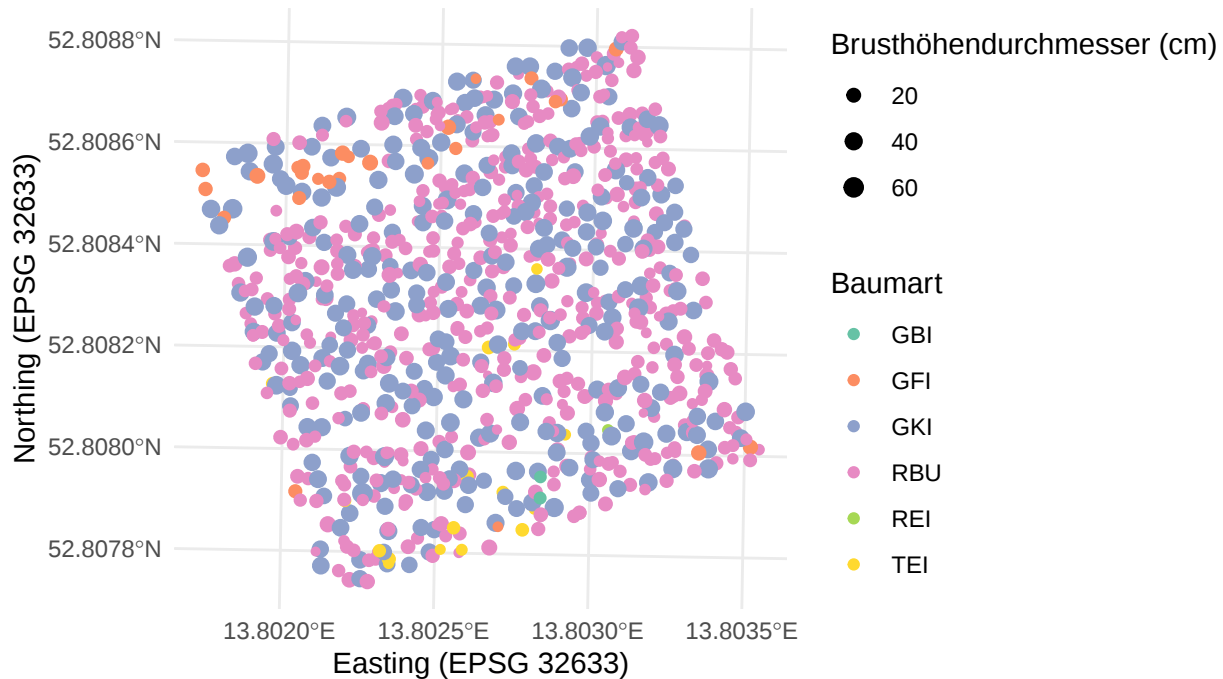
```
##           X           Y treeID spec_short dbh_cm height_m subID
## 1  419276.4 5851659         1      RBU    9.5    11.6     2
## 2  419277.9 5851660         2      RBU   15.8    19.6     1
## 3  419278.0 5851660         3      RBU   12.9    17.5     1
## 4  419277.4 5851662         4      RBU   11.0    13.4     1
## 5  419280.1 5851661         5      GKI   38.8    31.9     1
## 6  419281.1 5851661         6      RBU    7.9     9.1     1
## 7  419285.4 5851665         7      GKI   39.6    31.9     1
## 8  419289.0 5851666         8      GKI   32.1    28.7     1
## 9  419289.6 5851666         9      GFI    7.5     7.2     1
## 10 419294.2 5851665        10      RBU   16.8    17.4     1
##
##           geometry
## 1  POINT (419276.4 5851659)
## 2  POINT (419277.9 5851660)
## 3  POINT (419278 5851660)
## 4  POINT (419277.4 5851662)
## 5  POINT (419280.1 5851661)
## 6  POINT (419281.1 5851661)
## 7  POINT (419285.4 5851665)
## 8  POINT (419289 5851666)
## 9  POINT (419289.6 5851666)
## 10 POINT (419294.2 5851665)
```

```
st_crs(trees)$epsg
```

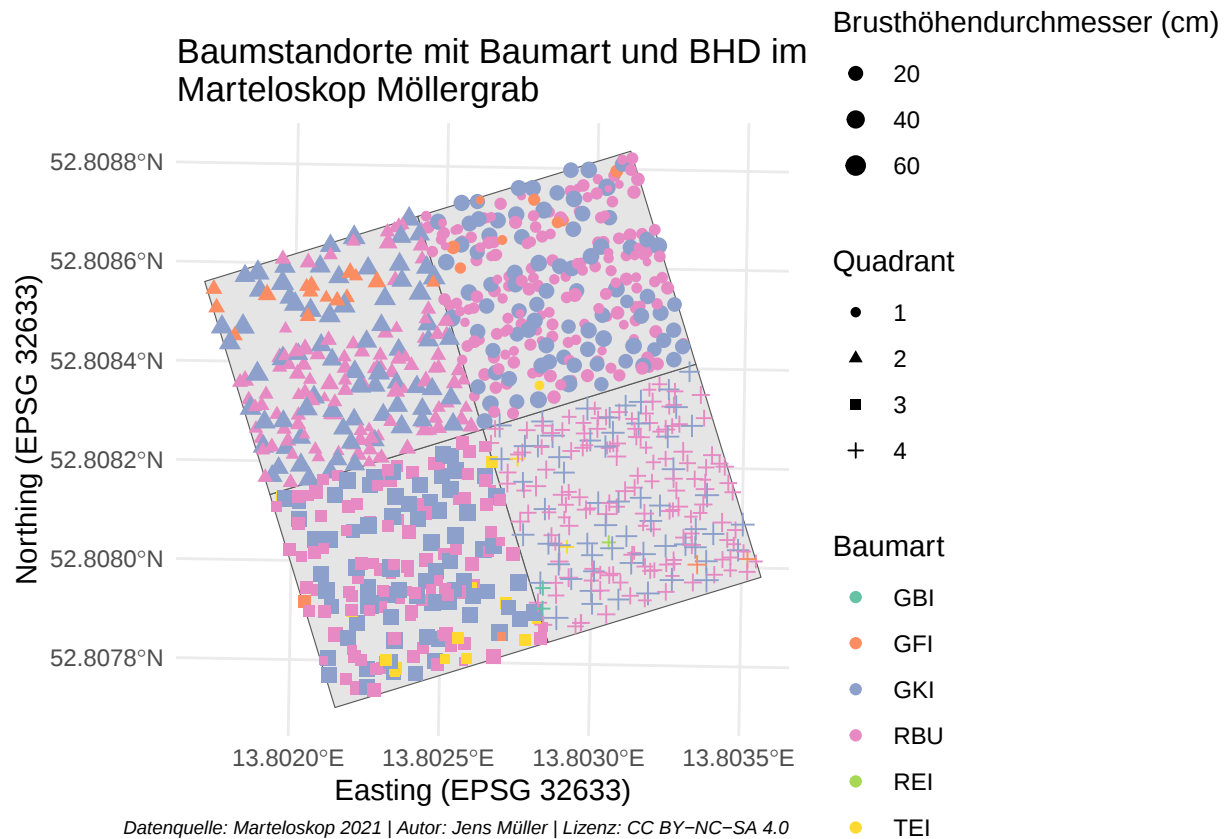
```
## [1] 32633
```

```
ggplot(data = trees) +
  geom_sf(aes(color = spec_short, size = dbh_cm)) +
  scale_size(range = c(1, 3)) +
  labs(title = 'Baumstandorte mit Baumart und Brusthöhendurchmesser\nim Marteloskop Möllergrab',
        x = 'Easting (EPSG 32633)',
        y = 'Northing (EPSG 32633)',
        color = 'Baumart',
        size = "Brusthöhendurchmesser (cm)") +
  scale_color_manual(values = colors_tree_species) +
  theme_minimal()
```

Baumstandorte mit Baumart und Brusthöhendurchmesser im Marteloskop Möllergrab



```
ggplot(data = quadrants) +
  geom_sf() +
  geom_sf(data = trees, aes(color = spec_short, size = dbh_cm, shape = factor(subID))) +
  scale_size(range = c(1, 3)) +
  labs(title = 'Baumstandorte mit Baumart und BHD im\nMarteloskop Möllergrab',
       x = 'Easting (EPSG 32633)',
       y = 'Northing (EPSG 32633)',
       color = 'Baumart',
       size = 'Brusthöhendurchmesser (cm)',
       shape = 'Quadrant',
       caption = "Datenquelle: Marteloskop 2021 | Autor: Jens Müller | Lizenz: CC BY-NC-SA 4.0") +
  scale_color_manual(values = colors_tree_species) +
  theme_minimal() +
  theme(plot.caption = element_text(hjust = 1, size = 7, face = "italic"))
```



```
# save plot in PDF file
ggsave("plot.pdf")
```

```
## Saving 6.5 x 4.5 in image
```