

Johannes Kepler Universität Linz

***Abteilung für Datenanalyse und Ökonometrie***

Replikation des Artikels:

**Data Descriptor: A dataset of forest**

**biomass structure for Eurasia**

Ausgearbeitet von

**k0355435 | Rudolf Gruber**

**LVA-Leitung:**

[**Assoz. Univ.-Prof.in Dipl.-Ing.in Dr.in Bettina Grün**](http://www.jku.at/ifas/content/e98868/index_html?team_view=section&emp=e98868/employee_groups_wiss98983/employees103449)

Table of Content

[Vorwort 1](#_Toc491179760)

[1. Aufgabe 2](#_Toc491179761)

[2. Aufgabe 1](#_Toc491179762)

[3. Aufgabe 2](#_Toc491179763)

[4. Aufgabe 1](#_Toc491179764)

[5. Aufgabe 3](#_Toc491179765)

[6. Aufgabe 4](#_Toc491179766)

[7. Aufgabe 6](#_Toc491179767)

[8. Aufgabe 7](#_Toc491179768)

[9. Aufgabe 8](#_Toc491179769)

[Session Info 9](#_Toc491179770)

# Vorwort

In diesem Dokument repliziere ich für die Lehrveranstaltung „Programmieren mit statistischer Software – Sommersemester 2017“ den Artikel „A Dataset of Forest Biomass Structure for Eurasia“ von Sepaschenko et. al (2017).

Die Daten liegen als xlsx Dateiformat vor und sind von mir unter <https://doi.pangaea.de/10.1594/PANGAEA.871465> als auch unter <https://doi.pangaea.de/10.1594/PANGAEA.871491> heruntergeladen worden.

Der Datensatz von Sepaschenko et. al (2017) beinhaltet experimentelle Daten der Autoren als auch Daten vergangener wissenschaftlichen Publikationen und ist meines Wissens entsprechend, der umfangreichste veröffentlichte Datensatz über Biomass Structure.

Für die Replikation der Aufgabenstellungen nutze ich neben den Lehrinhalten des Kurses zu R die Pakete ggplot, dplyr und tidyr. Funktionen, welche für die Replikation zu implementieren waren befinden sich in einem Hilfsskript mit dem Titel util.R. Ein main.R Skript führt die Replikation als gesamtes durch. Skripte der Aufgabenstellungen sind entsprechend der Nummerierung der Aufgabe benannt (z.B: Aufgabe1.R für die 1. Aufgabe).

# Aufgabe

Plot enthält 44 Variablen und 10351 Observationen

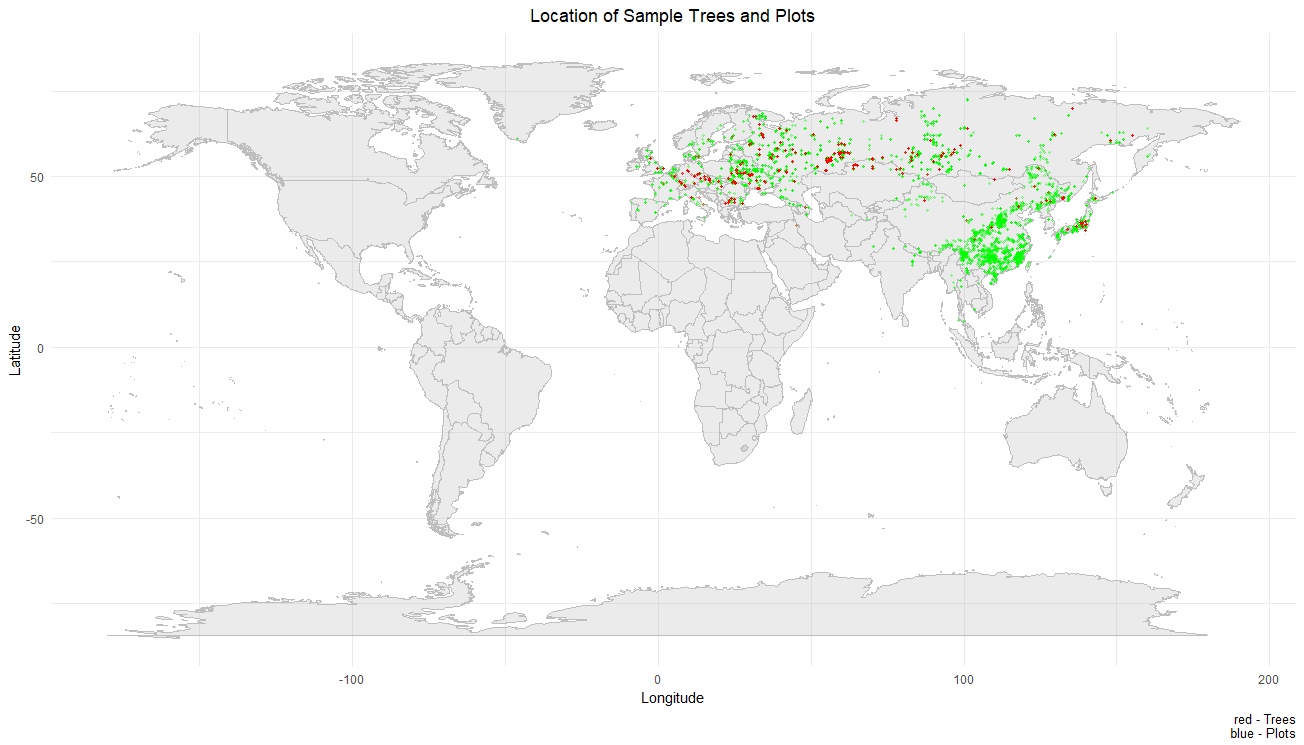
Tree enthält 27 Variablen und 9613 Observationen

> dim.plot

[1] 10351 44

> dim.tree

[1] 9613 27



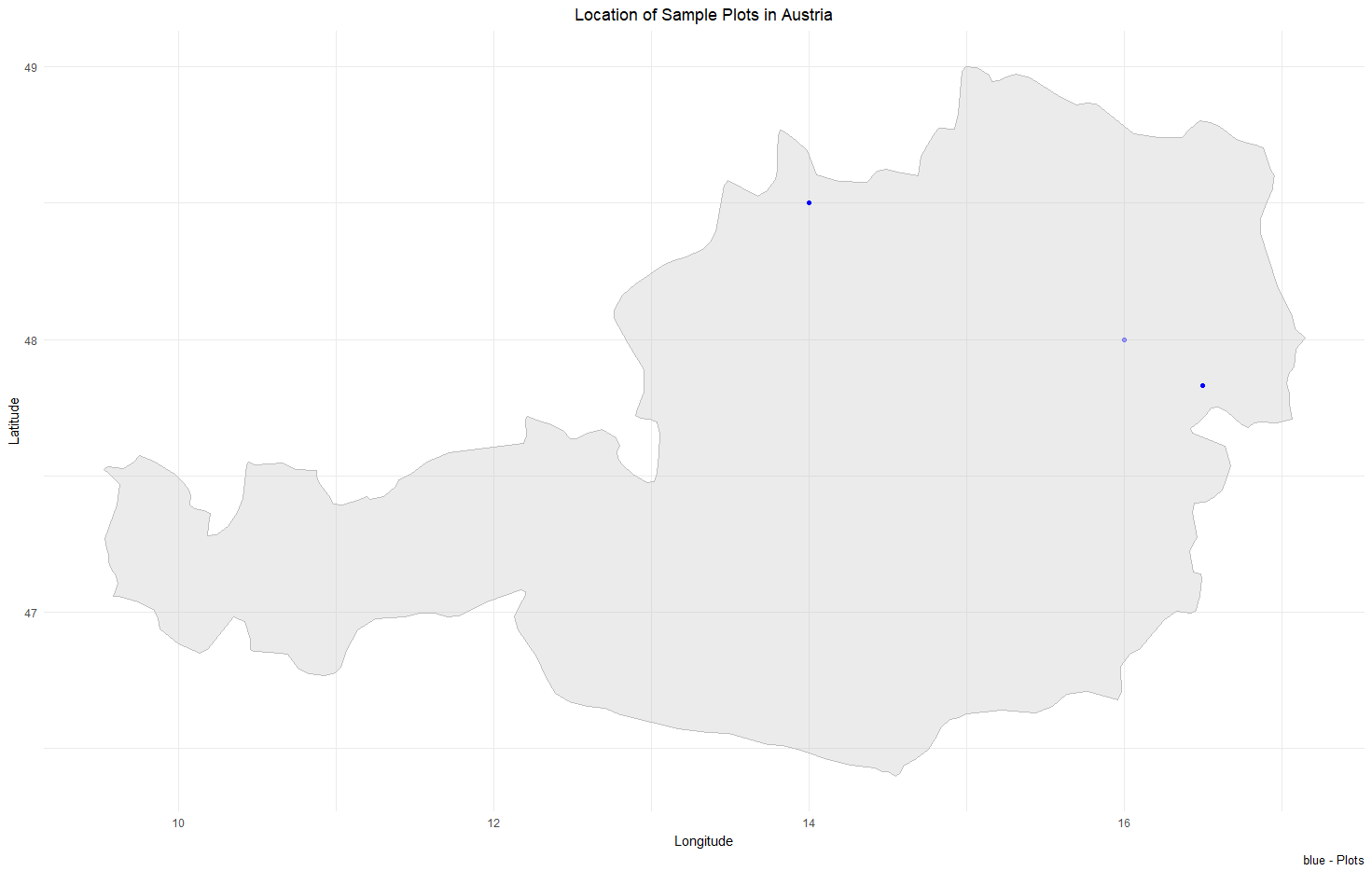
In Österreich sind 28 Plots vermerkt aber keine Trees.

> nrow(plot.aut)

[1] 28

> nrow(tree.aut) # 0 observations for Aut

[1] 0



Einer der Plots in Österreich befindet sich in Altenberg bei Linz.

# Aufgabe

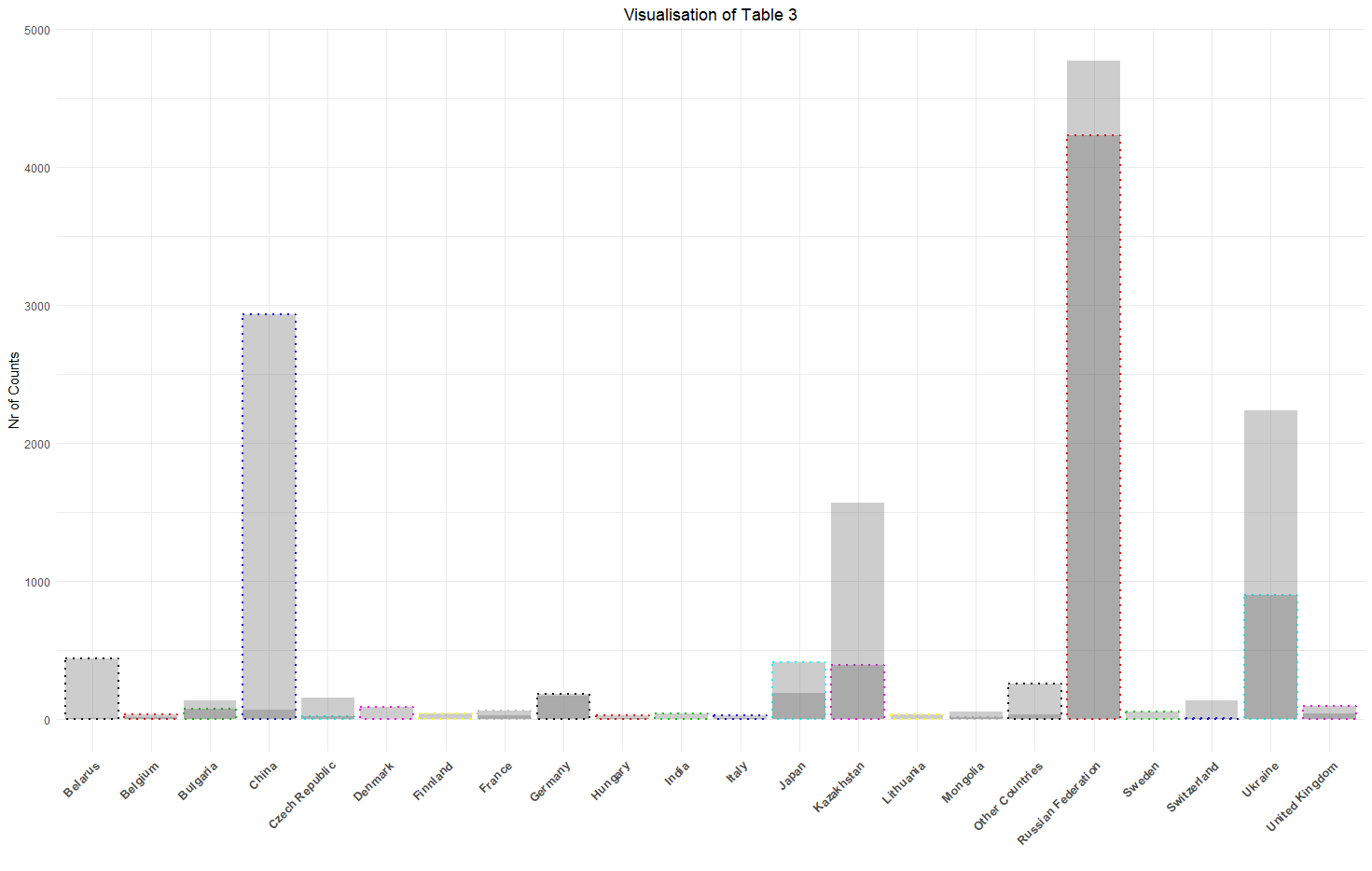
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field** | **Number** | **Unique** | **Median** | **Min** | **Max** |
| **Tree species (Latin me)** | 9613 |  | 4807 | 1 | 9613 |
| **Age (years)** | 9613 | 90 |  |  |  |
| **Diameter at breast height—DBH (cm)** | 9575 |  | 36 | 3 | 430 |
| **Height of the tree (m)** | 9518 |  | 12 | 0 | 98 |
| **Height to crown base (m)** | 8625 |  | 42897 | 0.13 | 44.2 |
| **Diameter (maximal) of the crown (m)** | 5774 |  | 42799 | 0 | 42972 |
| **Stem over bark volume (dm3)** | 4091 |  | 42401 | 0.02 | 42808 |
| **Stem bark volume (dm3)** | 7169 |  | 68.4 | 0.02 | 6984 |
| **Origin (tural or planted)** | 5404 |  | 11 | 0.01 | 678 |
| **Live biomass (kg oven dry matter1)** | 9530 | 4 |  |  |  |
| **Stem over bark** | 7466 |  | 29.43 | 0 | 4122 |
| **Bark of stem** | 4799 |  | 42982 | 0 | 280 |
| **Crown wood over bark** | 8862 |  | 31472 | 0 | 1091.8 |
| **Leaves & needles** | 8896 |  | 42827 | 0 | 305 |
| **Above ground** | 7474 |  | 36.59 | 0 | 5089 |
| **Stump & roots of trees** | 1746 |  | 34790 | 0 | 901 |
| **Total tree** | 1712 |  | 28.78 | 0 | 5134.8 |
| **Location and reference** | 7370 | 198 |  |  |  |
| **Country code (ISO ALPHA-3)** | 9613 | 21 |  |  |  |
| **Latitude** | 9613 |  | 54.67 | 42886 | 69.88 |
| **Longitude** | 9613 |  | 55.82 | -2.48 | 155 |
| **Altitude (m a.s.l.)** | 310 | 34 |  |  |  |
| **Number of trees per hectare** | 8592 |  | 2003 | 69 | 900000 |
| **Reference** | 9613 | 159 |  |  |  |
| **Notes** | 430 | 9 |  |  |  |
| **Ecoregion** | 9613 |  | 80608 | 80402 | 81211 |
| **Link to the sample plot (ID of the sample plot dataset)** | 6280 |  | 5760.5 | 192 | 12263 |

# Aufgabe

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field** | **Number** | **Unique** | **Median** | **Min** | **Max** |
| **Domint tree species (Latin me)** | 10351 | 465 |  |  |  |
| **Domint tree species code** | 10351 |  | 104 | 101 | 1012 |
| **Tree species composition** | 9819 | 1016 |  |  |  |
| **Age structure of stands (even or uneven aged)** | 700 | 5 |  |  |  |
| **Origin (tural or planted)** | 9860 | 6 |  |  |  |
| **Site index** | 7324 | 18 |  |  |  |
| **Mean stand age (year)** | 10232 | 298 |  |  |  |
| **Average height of the stand height of a tree with average DBH on the plot** | 10351 |  | 42926 | 0 | 65.1 |
| **Average DBH diameter at breast height (1.3 m above ground)** | 10351 |  | 42988 | 0 | 114 |
| **Number of trees per hectare** | 10351 |  | 1290 | 0 | 10530000 |
| **Relative stocking ratio of basal area of a plot to basal area of the normal stand** | 10351 |  | 0.46 | 0 | 2 |
| **Basal area total cross sectiol area of live trees at breast height in a plot Growing stock volume of stems of all living trees** | 10351 |  | 0 | 0 | 57.7 |
| **Stem over bark** | 10351 |  | 155 | 0 | 3831 |
| **Bark of stem** | 10351 |  | 68.9 | 0 | 1280.28 |
| **Crown wood over bark** | 10351 |  | 0 | 0 | 74.7 |
| **Leaves & needles** | 10351 |  | 42777 | 0 | 214 |
| **Stump & roots of trees** | 10351 |  | 42982 | 0 | 103.5 |
| **Fine roots with several thresholds** | 4035 | 2341 |  |  |  |
| **Undergrowth & shrubs above ground** | 10351 |  | 0 | 0 | 16.68 |
| **Undergrowth & shrubs including roots** | 10351 |  | 0 | 0 | 73 |
| **Green forest above ground** | 10351 |  | 0 | 0 | 78.7 |
| **Green forest including roots** | 10351 |  | 0 | 0 | 65.1 |
| **Dead organic matter** | 10351 |  | 0 | 0 | 35.23 |
| **Sgs** | 10351 |  | 0 | 0 | 74.6 |
| **Logs** | 10351 |  | 0 | 0 | 74.4 |
| **Dead branches of living trees** | 10351 |  | 0 | 0 | 79.03 |
| **Dead roots** | 10351 |  | 0 | 0 | 35.25 |
| **Litter** | 10351 |  | 0 | 0 | 193.07 |
| **Country code (ISO ALPHA 3)** | 10351 | 43 |  |  |  |
| **Latitude (8 N to 72 N)** | 10351 |  | 51.33 | 21367 | 72.47 |
| **Longitude (8 W to 160 E)** | 10351 |  | 70.17 | -45.97 | 160.67 |
| **Altitude (m a.s.l.)** | 3880 | 915 |  |  |  |
| **Year of measurement** | 10351 |  | 1982 | 0 | 2014 |
| **Number of trees selected for destructive sampling** | 1705 | 37 |  |  |  |
| **Leaf area index of trees** | 10351 |  | 0 | 0 | 41.78 |
| **Ecoregion** | 10351 |  | 80436 | 40106 | 81330 |
| **Reference** | 10351 | 1482 |  |  |  |

# Aufgabe

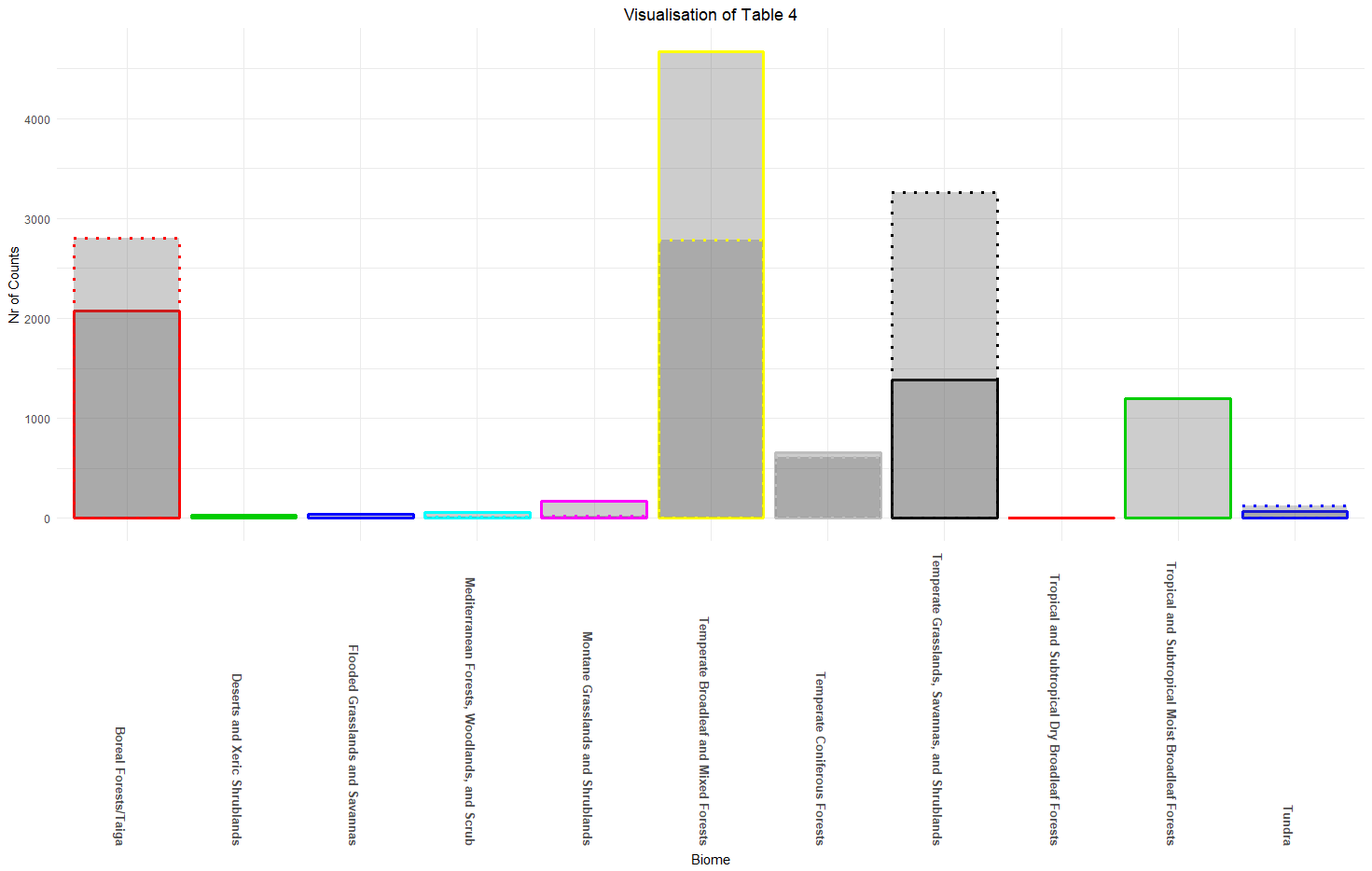
|  |  |  |
| --- | --- | --- |
| **Country** | **Plots** | **Trees** |
| **Belarus** | 439 | 8 |
| **Belgium** | 34 | 14 |
| **Bulgaria** | 75 | 137 |
| **Chi** | 2933 | 65 |
| **Czech Republic** | 20 | 153 |
| **Denmark** | 85 | 1 |
| **Finnland** | 38 | 0 |
| **France** | 63 | 24 |
| **Germany** | 184 | 167 |
| **Hungary** | 28 | 9 |
| **India** | 37 | 0 |
| **Italy** | 29 | 9 |
| **Japan** | 412 | 186 |
| **Kazakhstan** | 393 | 1564 |
| **Lithuania** | 33 | 0 |
| **Mongolia** | 10 | 57 |
| **Other Countries** | 254 | 32 |
| **Russian Federation** | 4228 | 4771 |
| **Sweden** | 56 | 0 |
| **Switzerland** | 6 | 136 |
| **Ukraine** | 897 | 2238 |
| **United Kingdom** | 97 | 42 |
| **Total** | **10351** | **9613** |



Visualisierung von Tabelle 4. Plots sind punktiert umrandet.

# Aufgabe

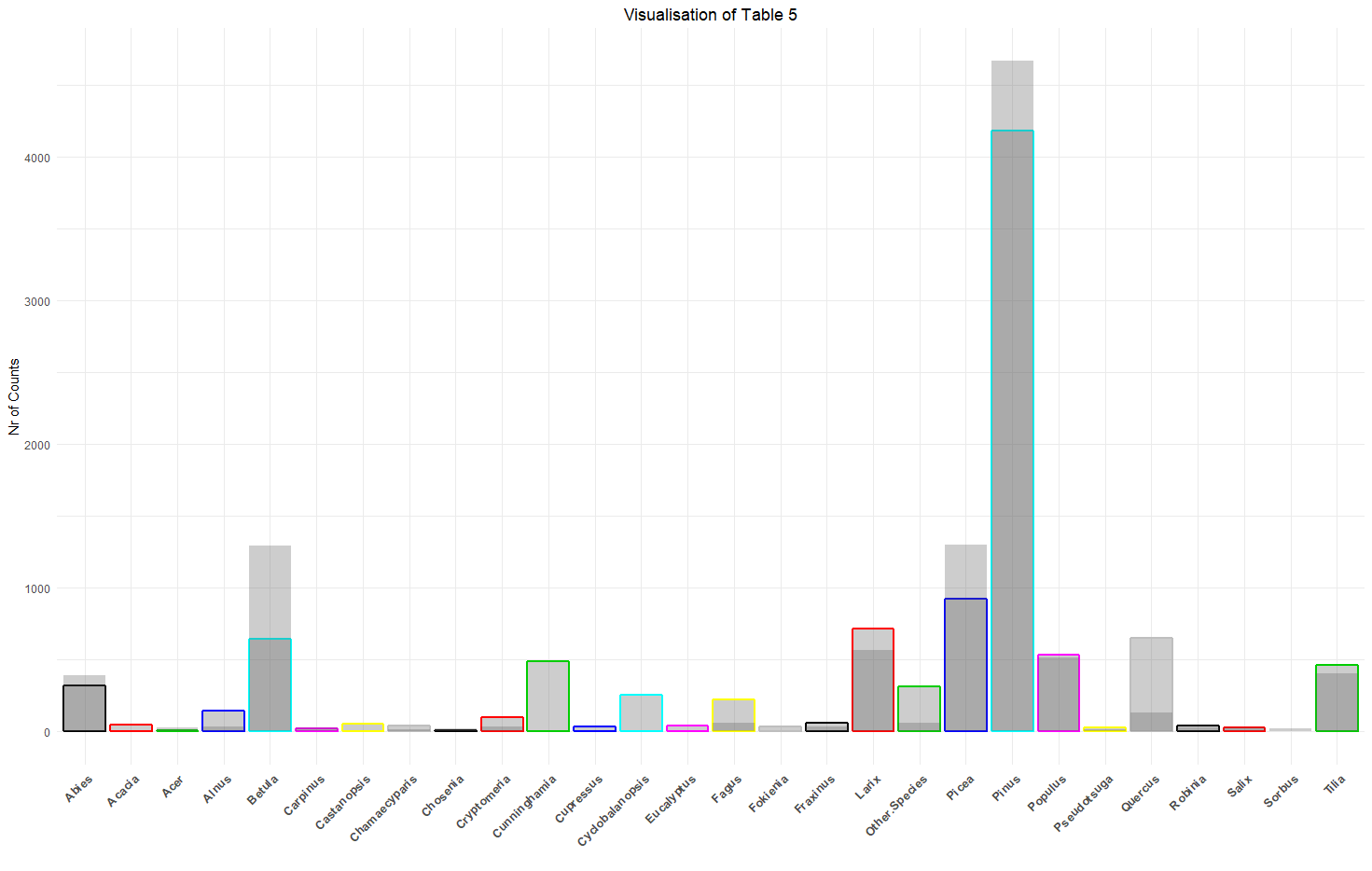
|  |  |  |
| --- | --- | --- |
| **Biome** | **Plots** | **Trees** |
| **Boreal Forests/Taiga** | 2072 | 2805 |
| **Deserts and Xeric Shrublands** | 31 | - |
| **Flooded Grasslands and Savans** | 39 | 1 |
| **Mediterranean Forests, Woodlands, and Scrub** | 55 | 8 |
| **Montane Grasslands and Shrublands** | 171 | 18 |
| **Temperate Broadleaf and Mixed Forests** | 4673 | 2788 |
| **Temperate Coniferous Forests** | 655 | 606 |
| **Temperate Grasslands, Savans, and Shrublands** | 1385 | 3264 |
| **Tropical and Subtropical Dry Broadleaf Forests** | 6 | - |
| **Tropical and Subtropical Moist Broadleaf Forests** | 1197 | - |
| **Tundra** | 67 | 123 |



Darstellung von Tabelle 4. Die Verteilung der Trees ist durch die punktierte Umrandung erkennbar. Die Verteilung der Plots ist mit vollen Rahmenlinien ersichtlich.

# Aufgabe

|  |  |  |
| --- | --- | --- |
| **Tree genus** | **Sample plots** | **Sample trees** |
| **Abies** | 320 | 386 |
| **Acacia** | 44 |  |
| **Acer** | 7 | 27 |
| **Alnus** | 142 | 31 |
| **Betula** | 643 | 1291 |
| **Carpinus** | 17 | 22 |
| **Castanopsis** | 53 |  |
| **Chamaecyparis** | 38 | 10 |
| **Chosenia** | 6 | 17 |
| **Cryptomeria** | 97 | 29 |
| **Cunninghamia** | 489 |  |
| **Cupressus** | 32 |  |
| **Cyclobalanopsis** | 254 |  |
| **Eucalyptus** | 37 |  |
| **Fagus** | 221 | 56 |
| **Fokienia** | 34 |  |
| **Fraxinus** | 55 | 31 |
| **Larix** | 711 | 566 |
| **Picea** | 921 | 1298 |
| **Pinus** | 4186 | 4668 |
| **Populus** | 533 | 513 |
| **Pseudotsuga** | 27 | 13 |
| **Quercus** | 650 | 130 |
| **Robinia** | 40 | 24 |
| **Salix** | 22 | 23 |
| **Tilia** | 462 | 402 |
| **Sorbus** |  | 20 |
| **Other.Species** | 310 | 56 |
| **Total** | **10351** | **9613** |

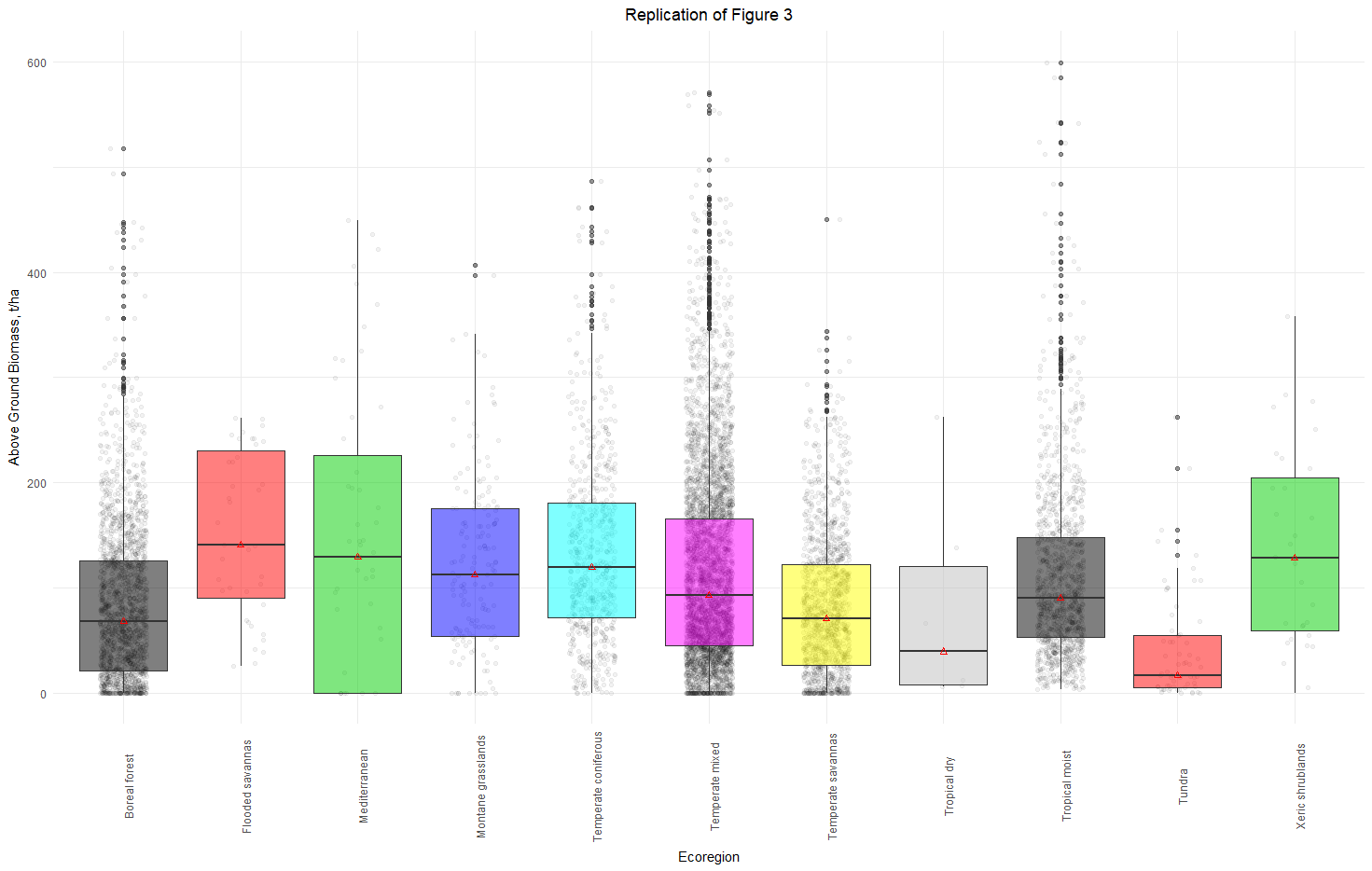


Vorhergehende Grafik visualisiert Tabelle 5. Die farbliche Umrandung von Balken identifiziert die Verteilung der Sample Plots.

**Tabelle 5** weicht teilweise in den Summen in der plot.db von der ursprünglichen Publikation ab. Abweichende Werte sind gelb hinterlegt. Im Gegensatz zur plot.db stimmen die Werte für die tree.db vollständig mit der Publikation überein. Ich nehme an, dass ein Fehler, in der Spezifikation des Regulären Ausdrucks, die unterschiedlichen Werte für Sample Plots verursacht.

Ein vorher begangener Fehler war der Parameter ignore.case = T in der grep Funktion. Dies führte dazu, das Carpinus von Pinus überschrieben wurde und somit keine Einträge für Carpinus in der Tabelle vorhanden waren.

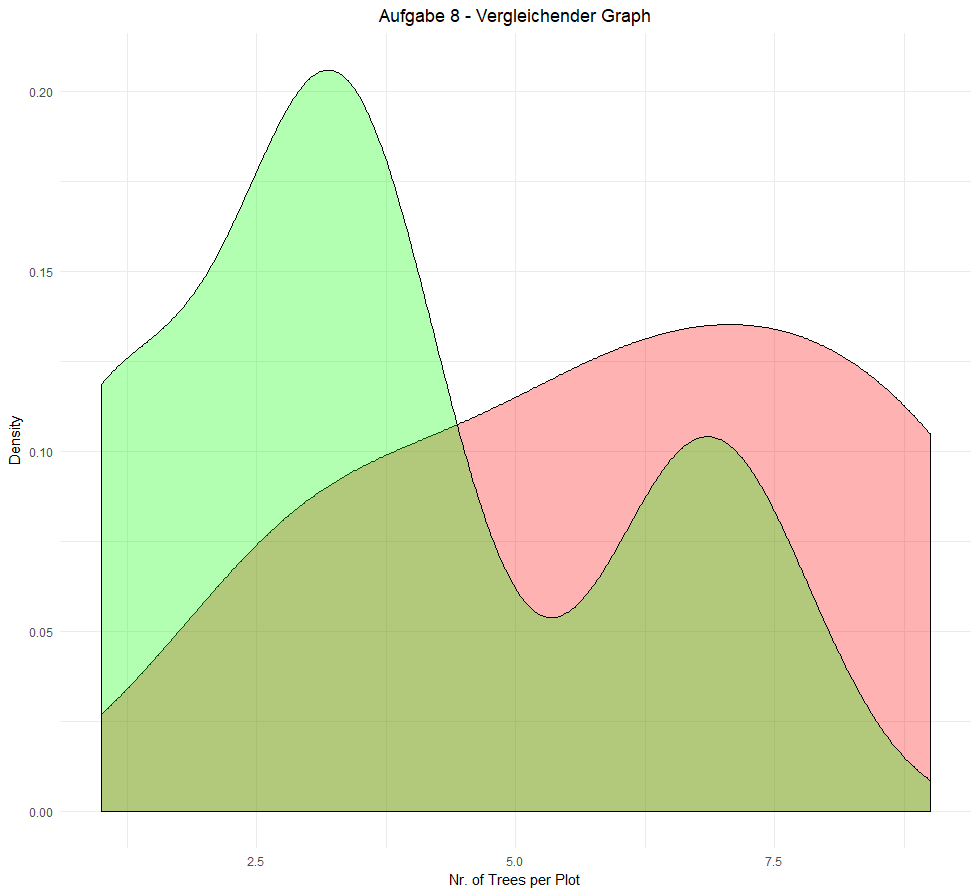
# Aufgabe



Folgende Abbildung repliziert Figure 3, wo die Verteilung der Above Ground Biomass dargestellt wird. Im Gegensatz zur originalen Publikation sind die Ecoregionen in dieser Darstellung alphabetisch sortiert.

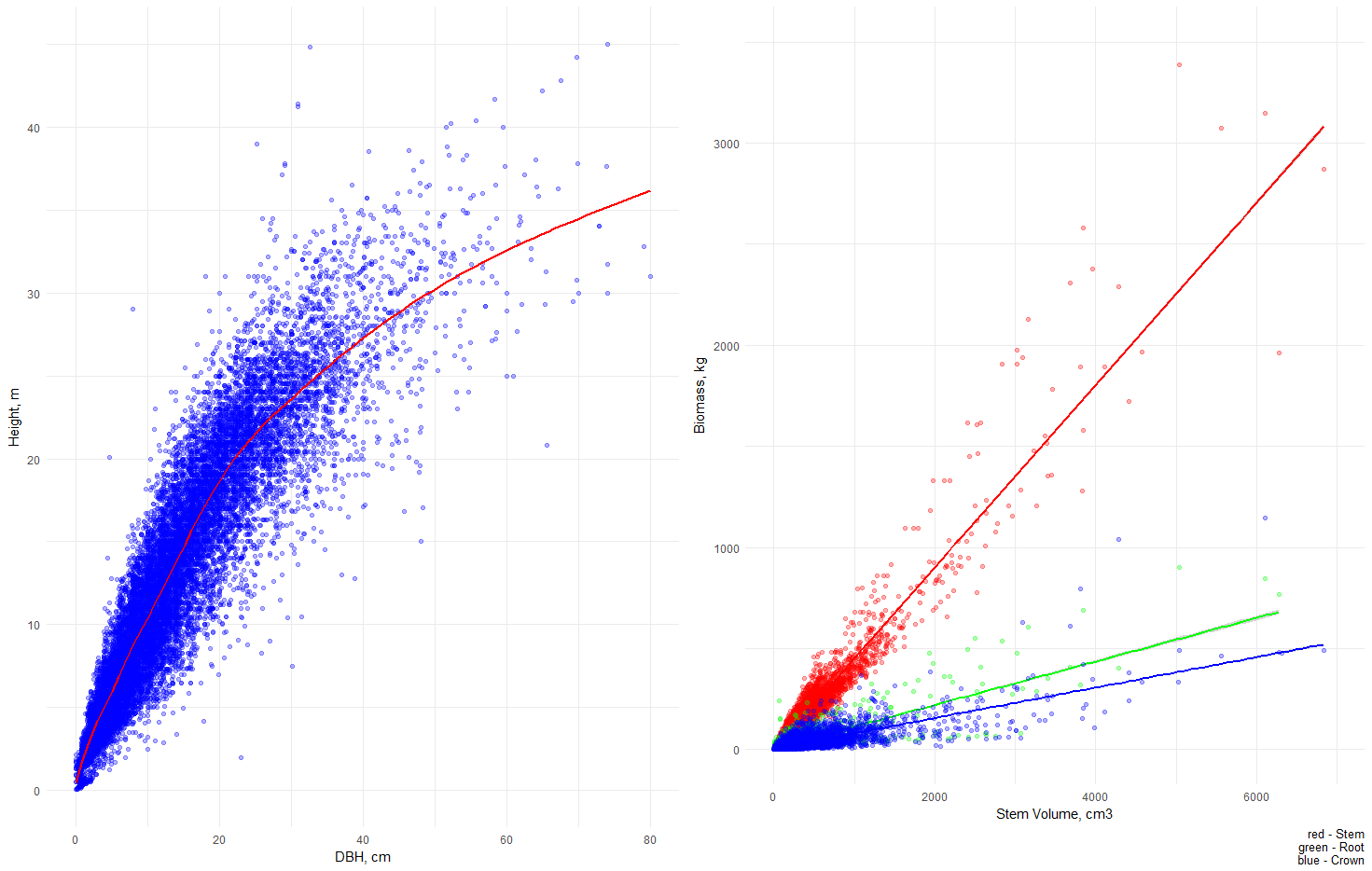
# Aufgabe

|  |  |  |
| --- | --- | --- |
|  | **Data with Representatives** | **Data without Representatives** |
| **mean** | 5.449.258 | 3.562.644 |
| **median** | 3.000.000 | 1.000.000 |
| **sd** | 7.218.975 | 6.861.607 |
| **iqr** | 9.000.000 | 3.000.000 |



Die Density Plots zeigen die Verteilung der Variable Number of Trees for destructive sampling, einmal mit Repräsentanten der Intervallsbereiche (rote Fläche) und einmal ohne Intervallbereiche (grüne Fläche), welche hier auf NA gesetzt wurden. Repräsentanten wurden für den Plot mit gleichverteilter Wahrscheinlichkeit ermittelt. Als erstes ist anzumerken, dass diese Plots unterschiedliche Datenmengen visualisieren. Intervallwerte wurden bei grün NA gesetzt wodurch die große Anzahl an fehlenden Werten mit geringer Dichte in der mittleren Baumanzahl sichtbar wird. Die gleichwahrscheinliche Wahl von Repräsentanten zeichnet sich in einer flacheren Dichte.

# Aufgabe



Smoothing der Beziehung zwischen DBH und Height mittels GAM (Generalized Additive Model).

Schätzen der Beziehung zwischen Stem Volume und Biomass mittels OLS.

# Session Info

1. R version 3.4.1 (2017-06-30)
2. Platform: x86\_64-w64-mingw32/x64 (64-bit)
3. Running under: Windows >= 8 x64 (build 9200)
4. Matrix products: default
5. locale:
6. [1] LC\_COLLATE=German\_Germany.1252
7. [2] LC\_CTYPE=German\_Germany.1252
8. [3] LC\_MONETARY=German\_Germany.1252
9. [4] LC\_NUMERIC=C
10. [5] LC\_TIME=German\_Germany.1252
11. attached base packages:
12. [1] grid stats graphics grDevices utils
13. [6] datasets methods base
14. other attached packages:
15. [1] bindrcpp\_0.2 maps\_3.2.0 ggmap\_2.6.1 ggplot2\_2.2.1
16. [5] tidyr\_0.6.3 dplyr\_0.7.1
17. loaded via a namespace (and not attached):
18. [1] Rcpp\_0.12.12 compiler\_3.4.1 plyr\_1.8.4
19. [4] bindr\_0.1 tools\_3.4.1 digest\_0.6.12
20. [7] nlme\_3.1-131 tibble\_1.3.3 gtable\_0.2.0
21. [10] lattice\_0.20-35 mgcv\_1.8-17 pkgconfig\_2.0.1
22. [13] png\_0.1-7 rlang\_0.1.1 Matrix\_1.2-10
23. [16] mapproj\_1.2-5 proto\_1.0.0 stringr\_1.2.0
24. [19] RgoogleMaps\_1.4.1 glue\_1.1.1 R6\_2.2.2
25. [22] jpeg\_0.1-8 sp\_1.2-5 reshape2\_1.4.2
26. [25] magrittr\_1.5 scales\_0.4.1 assertthat\_0.2.0
27. [28] colorspace\_1.3-2 geosphere\_1.5-5 labeling\_0.3
28. [31] stringi\_1.1.5 lazyeval\_0.2.0 munsell\_0.4.3
29. [34] rjson\_0.2.15