



Johannes Kepler Universität Linz

# ***Abteilung für Datelyse 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.<sup>in</sup> Dipl.-Ing.<sup>in</sup> Dr.<sup>in</sup> Bettina Grün***

## Table of Content

Vorwort .....	1
1. Aufgabe .....	2
2. Aufgabe .....	1
3. Aufgabe .....	2
4. Aufgabe .....	1
5. Aufgabe .....	3
6. Aufgabe .....	4
7. Aufgabe .....	6
8. Aufgabe .....	7
9. Aufgabe .....	8
Session Info .....	9

## 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.

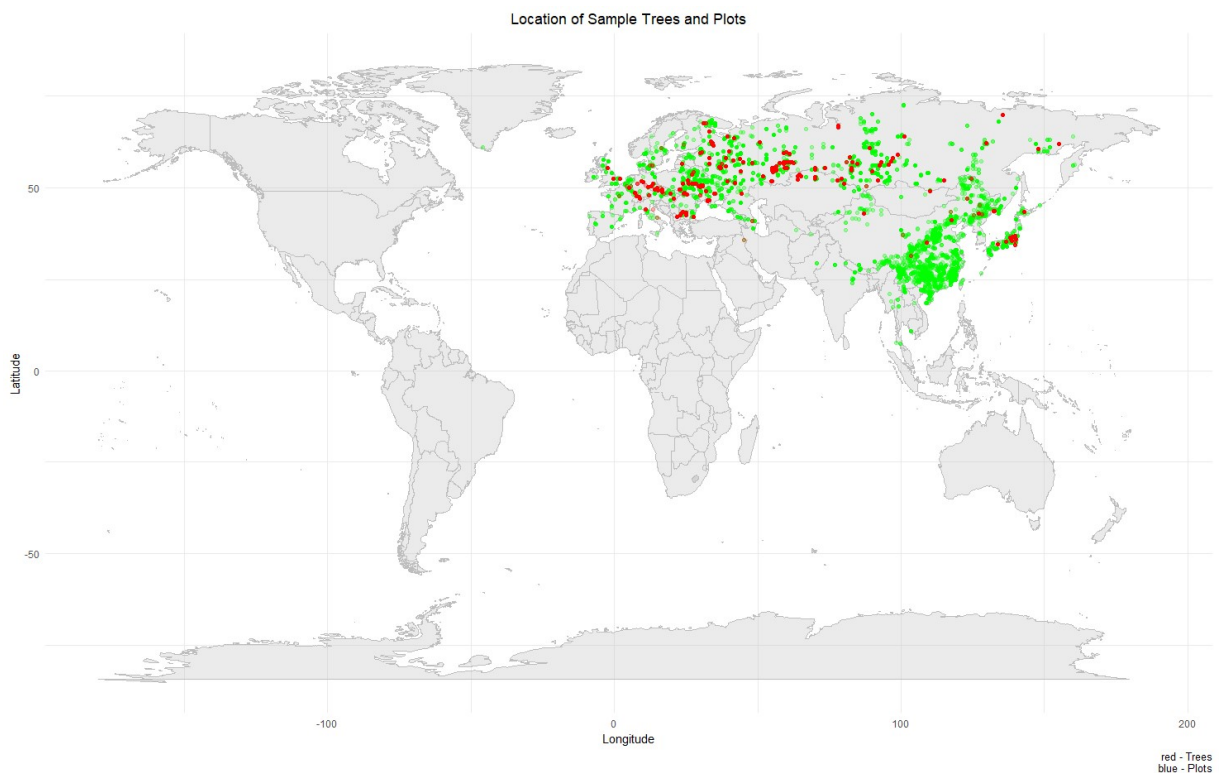
Für die Replikation der Aufgabenstellungen nutze ich neben den Lehrinhalten des Kurses zu R die Bibliotheken ggplot, dplyr und tidyr. Funktionen, welche für die Replikation zu implementieren waren befinden sich in einem Hilfsskript mit dem Titel util.R. Ein main.R Skript führt die Replikation als gesamtes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

## 1. Aufgabe

Plot enthält 44 Variablen und 10351 Observationen

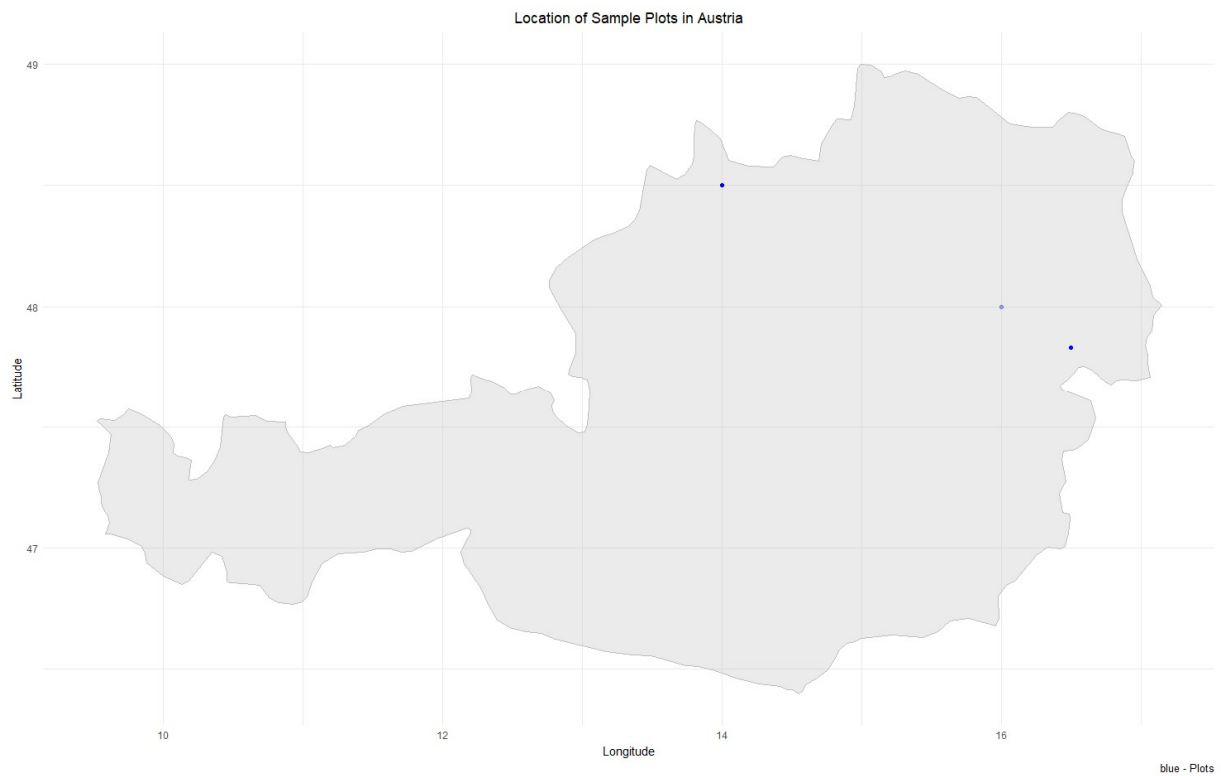
Tree enthält 27 Variablen und 9613 Observationen

```
> 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.

## 2. 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

### 3. Aufgabe

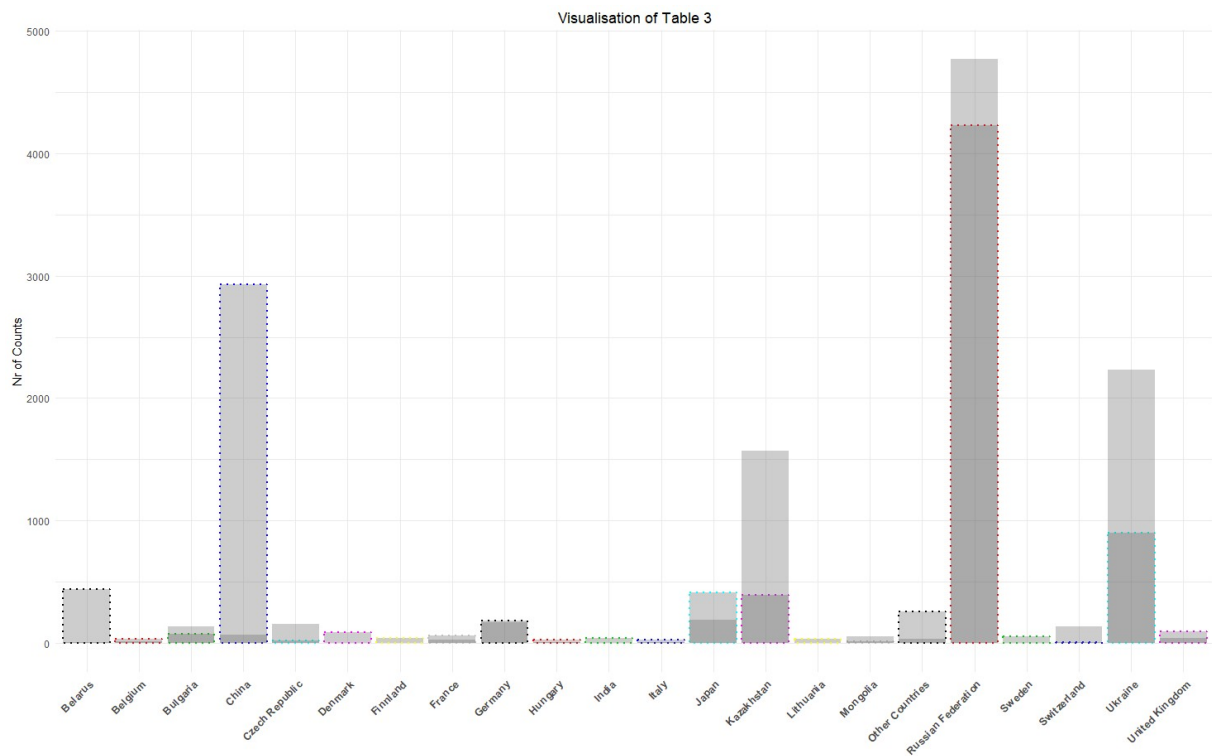
Field	Number	Unique	Median	Min	Max
Domint tree species (Latin me)	10351	465			
Domint tree species code	10351		104	101	1012
Tree species composition	9819	1016			
Age structure of stands (even or uneven aged)	700	5			
Origin (tural or planted)	9860	6			
Site index	7324	18			
Mean stand age (year)	10232	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

<b>Undergrowth &amp; shrubs including roots</b>	10351		0	0	73
<b>Green forest above ground</b>	10351		0	0	78.7
<b>Green forest including roots</b>	10351		0	0	65.1
<b>Dead organic matter</b>	10351		0	0	35.23
<b>Sgs</b>	10351		0	0	74.6
<b>Logs</b>	10351		0	0	74.4
<b>Dead branches of living trees</b>	10351		0	0	79.03
<b>Dead roots</b>	10351		0	0	35.25
<b>Litter</b>	10351		0	0	193.07
<b>Country code (ISO ALPHA 3)</b>	10351	43			
<b>Latitude (8 N to 72 N)</b>	10351		51.33	21367	72.47
<b>Longitude (8 W to 160 E)</b>	10351		70.17	-45.97	160.67
<b>Altitude (m a.s.l.)</b>	3880	915			
<b>Year of measurement</b>	10351		1982	0	2014
<b>Number of trees selected for destructive sampling</b>	1705	37			
<b>Leaf area index of trees</b>	10351		0	0	41.78
<b>Ecoregion</b>	10351		80436	40106	81330
<b>Reference</b>	10351	1482			



## 4. Aufgabe

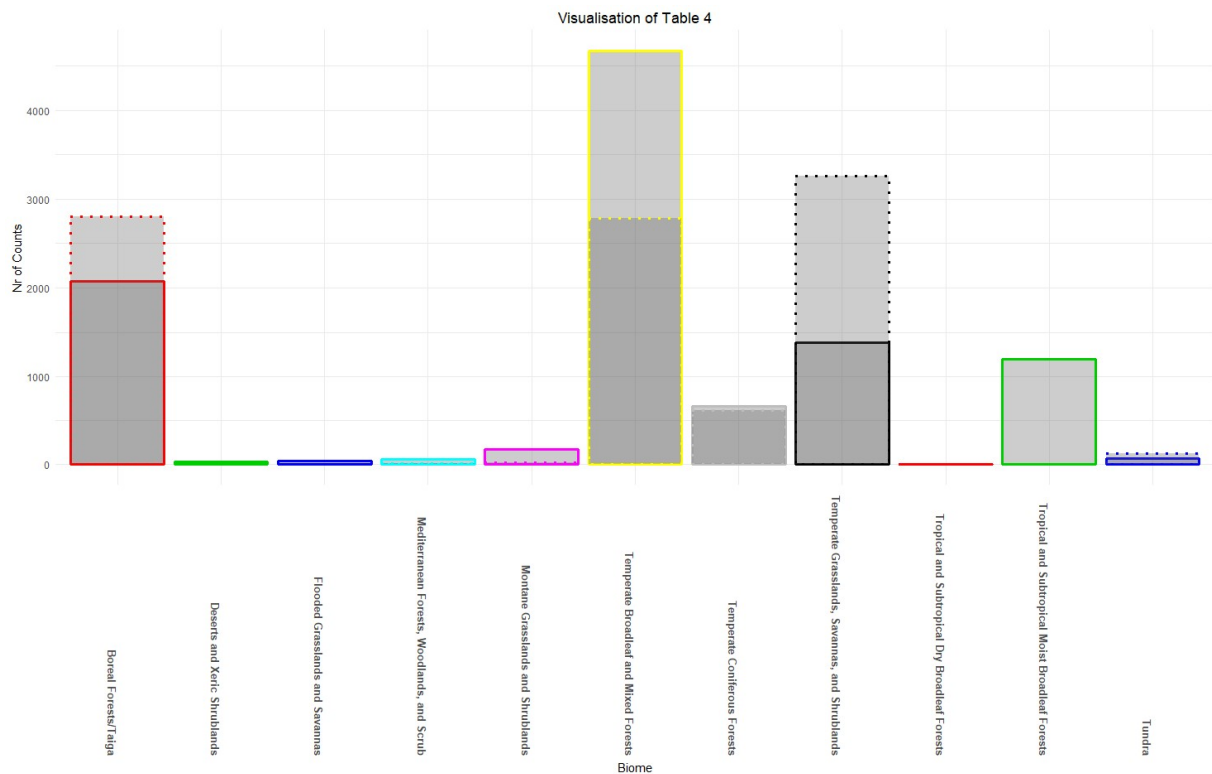
Country	Plots	Trees
Belarus	439	8
Belgium	34	14
Bulgaria	75	137
Chi	2933	65
Czech Republic	20	153
Denmark	85	1
Finnland	38	0
France	63	24
Germany	184	167
Hungary	28	9
India	37	0
Italy	29	9
Japan	412	186
Kazakhstan	393	1564
Lithuania	33	0
Mongolia	10	57
Other Countries	254	32
Russian Federation	4228	4771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



Visualisierung von Tabelle 4. Plots sind punktiert umrandet.

## 5. Aufgabe

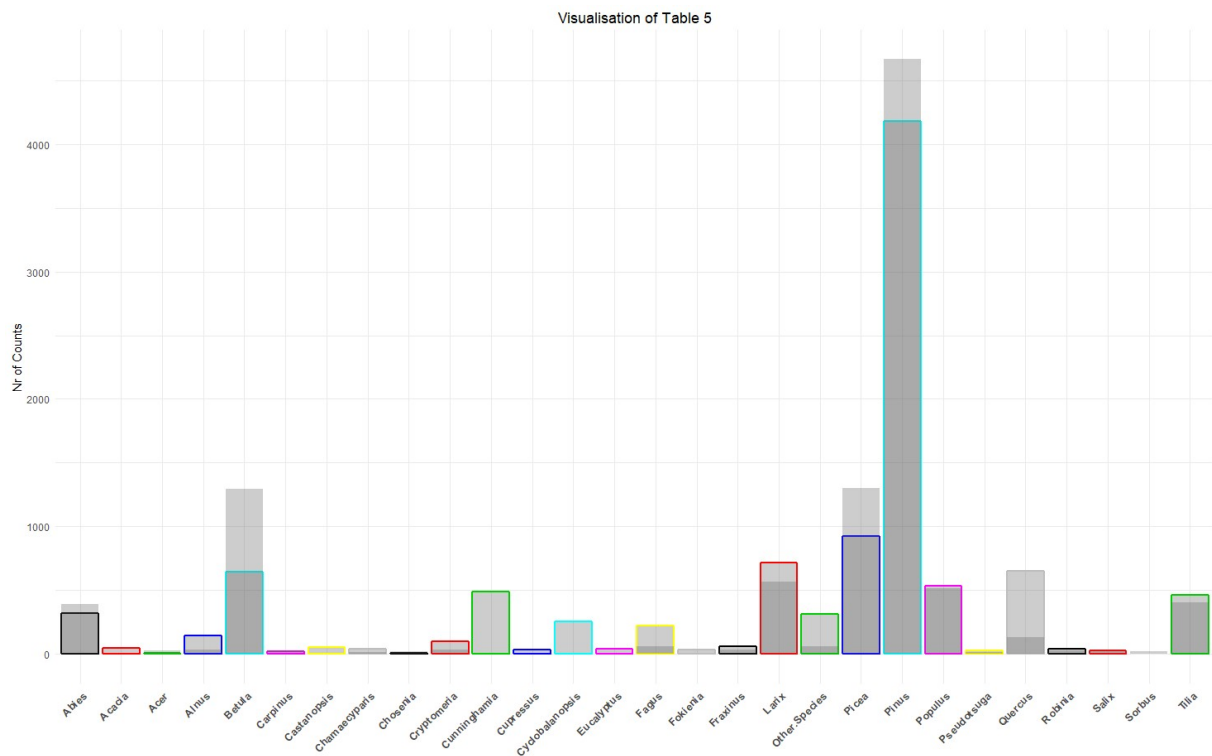
Biome	Plots	Trees
Boreal Forests/Taiga	2072	2805
Deserts and Xeric Shrublands	31	-
Flooded Grasslands and Savans	39	1
Mediterranean Forests, Woodlands, and Scrub	55	8
Montane Grasslands and Shrublands	171	18
Temperate Broadleaf and Mixed Forests	4673	2788
Temperate Coniferous Forests	655	606
Temperate 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.

## 6. 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
<b>Total</b>	<b>10351</b>	<b>9613</b>

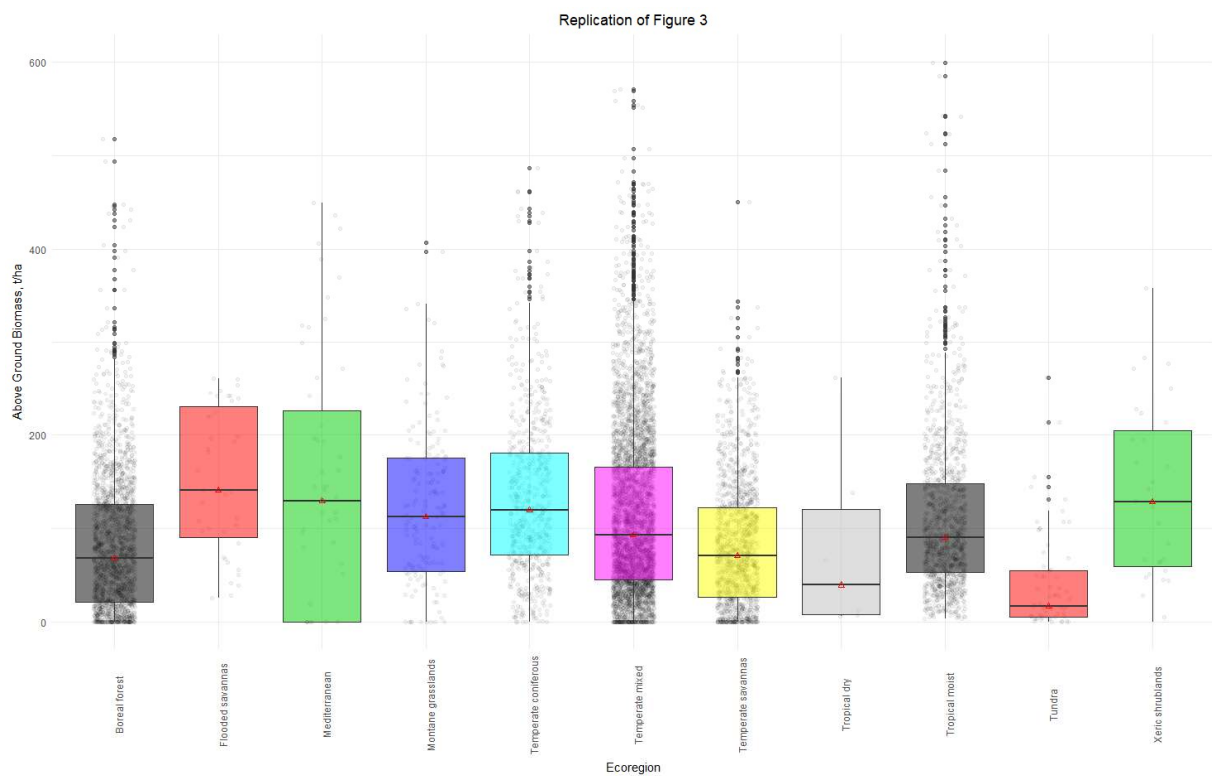


Vorhergehende Grafik visualisiert Tabelle 5. Die farbliche Umrandung von Balken identifiziert die Verteilung der Sample Plots.

**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.

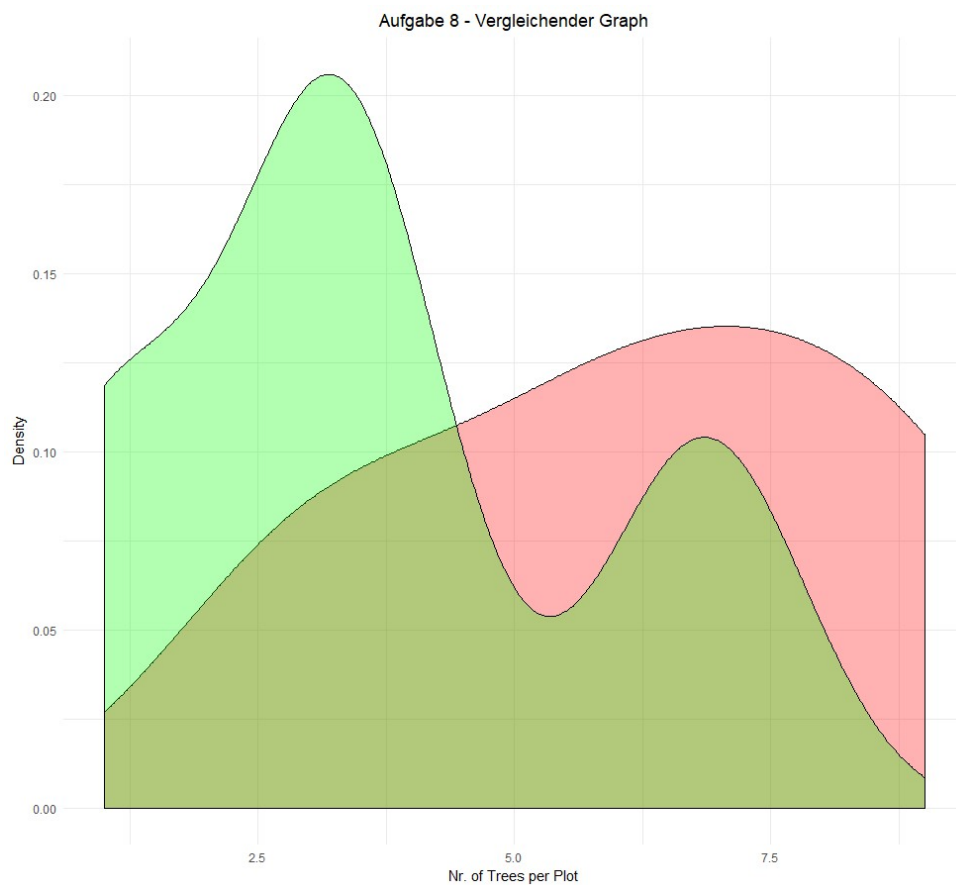
## 7. Aufgabe



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.

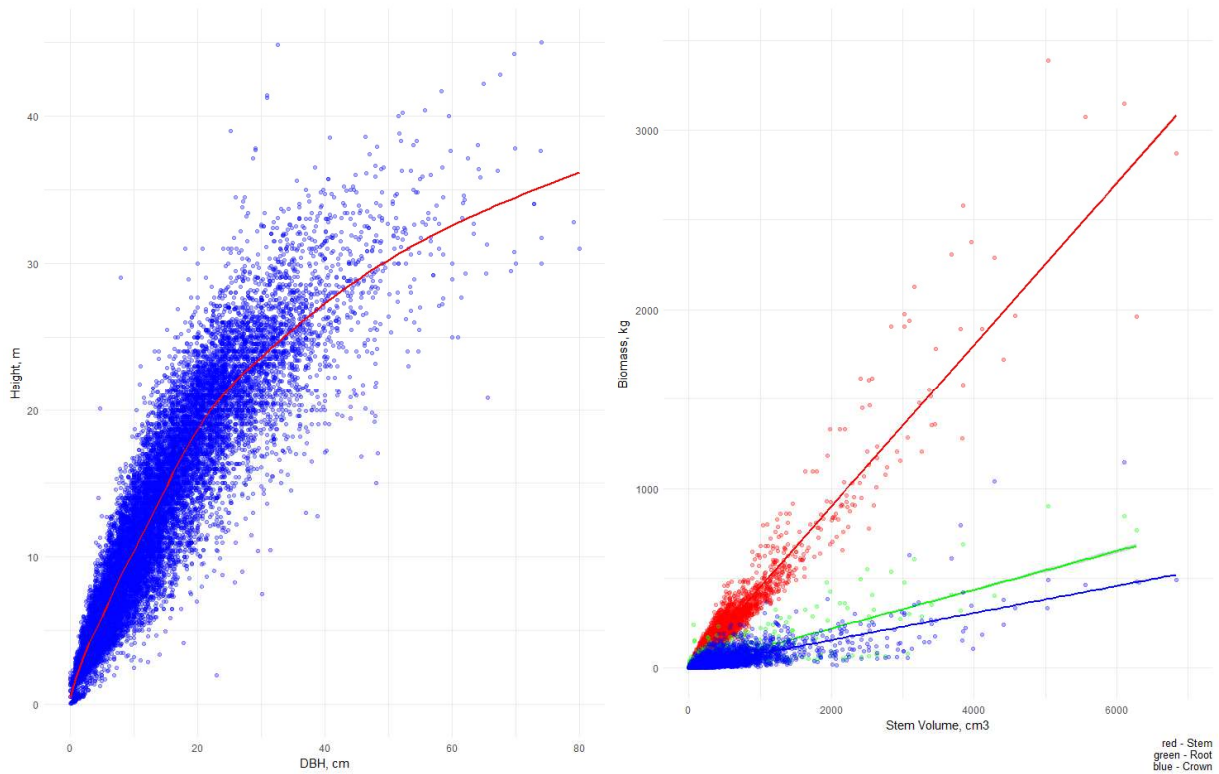
## 8. Aufgabe

	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.

## 9. 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

```

10. R version 3.4.1 (2017-06-30)
11. Platform: x86_64-w64-mingw32/x64 (64-bit)
12. Running under: windows >= 8 x64 (build 9200)
13.
14. Matrix products: default
15.
16. locale:
17. [1] LC_COLLATE=German_Germany.1252
18. [2] LC_CTYPE=German_Germany.1252
19. [3] LC_MONETARY=German_Germany.1252
20. [4] LC_NUMERIC=C
21. [5] LC_TIME=German_Germany.1252
22.
23. attached base packages:
24. [1] grid      stats    graphics  grDevices utils
25. [6] datasets  methods  base
26.
27. other attached packages:
28. [1] bindrcpp_0.2  maps_3.2.0    ggmap_2.6.1   ggplot2_2.2.1
29. [5] tidyr_0.6.3   dplyr_0.7.1
30.
31. loaded via a namespace (and not attached):
32. [1] Rcpp_0.12.12    compiler_3.4.1  plyr_1.8.4
33. [4] bindr_0.1       tools_3.4.1     digest_0.6.12
34. [7] nlme_3.1-131    tibble_1.3.3    gtable_0.2.0
35. [10] lattice_0.20-35 mgcv_1.8-17     pkgconfig_2.0.1
36. [13] png_0.1-7       rlang_0.1.1     Matrix_1.2-10
37. [16] mapproj_1.2-5   proto_1.0.0     stringr_1.2.0
38. [19] RgoogleMaps_1.4.1 glue_1.1.1      R6_2.2.2
39. [22] jpeg_0.1-8      sp_1.2-5        reshape2_1.4.2
40. [25] magrittr_1.5    scales_0.4.1    assertthat_0.2.0
41. [28] colorspace_1.3-2 geosphere_1.5-5 labeling_0.3
42. [31] stringi_1.1.5   lazyeval_0.2.0  munsell_0.4.3
43. [34] rjson_0.2.15

```