

Corpus des Deutschen Bundesrechts (C-DBR)

COMPILATION REPORT

Version 2022-08-05

License MIT-0

DOI: [10.5281/zenodo.6959381](https://doi.org/10.5281/zenodo.6959381)

Titel	Source Code des »Corpus des Deutschen Bundesrechts«
Abkürzung	C-DBR-Source
Autor	Seán Fobbe
Version	2022-08-05
Download	https://doi.org/10.5281/zenodo.6959381
Lizenz	MIT No Attribution (MIT-0)

Zitiervorschlag

Seán Fobbe (2022). Source Code des »Corpus des Deutschen Bundesrechts« (C-DBR-Source). Version 2022-08-05. Zenodo. DOI: 10.5281/zenodo.6959381.

Digital Object Identifier (DOI): Concept DOI und Version DOI

Soweit nicht anders angegeben ist die DOI immer eine »Version DOI« und bezieht sich nur auf eine bestimmte Version der Software. Sie verlinkt daher nur Version 2022-08-05. Für das Gesamtkonzept der Software steht eine »Concept DOI« zur Verfügung, die auf der Zenodo-Seite jeder Version unter »Cite all versions?« zu finden ist. Sie lautet 10.5281/zenodo.4072934. Die »Concept DOI« verlinkt immer die aktuellste Version.

Lizenz: MIT No Attribution (MIT-0)

Copyright — 2022— Seán Fobbe

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the »Software«), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED »AS IS«, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Disclaimer

Dieser Datensatz ist eine private wissenschaftliche Initiative und steht in keiner Verbindung zu Behörden, Gerichten oder anderen amtlichen Stellen der Bundesrepublik Deutschland.

Inhaltsverzeichnis

1	Corpus des Deutschen Bundesrechts (C-DBR)	9
1.1	Überblick	9
1.2	Funktionsweise	9
1.3	Kompilierung	10
1.4	Systemanforderungen	10
1.4.1	Betriebssystem	10
1.4.2	Software	10
1.4.3	Parallelisierung	11
1.4.4	Speicherplatz	11
1.5	Weitere Open Access Veröffentlichungen (Fobbe)	11
1.6	Kontakt	11
2	Vorbereitung	12
2.1	Datumsstempel	12
2.2	Datum und Uhrzeit (Beginn)	12
2.3	Packages Laden	12
2.4	Zusätzliche Funktionen einlesen	14
2.5	Verzeichnis für Analyse-Ergebnisse und Diagramme definieren	15
2.6	Weitere Verzeichnisse definieren	15
2.7	Dateien aus vorherigen Runs bereinigen	15
2.8	Verzeichnisse anlegen	15
2.9	Vollzitate statistischer Software schreiben	16
2.10	Allgemeine Konfiguration	16
2.10.1	Konfiguration einlesen	16
2.10.2	Konfiguration anzeigen	16
2.10.3	Knitr Optionen setzen	17
2.10.4	Download Timeout setzen	17
2.10.5	Quellenangabe für Diagramme definieren	17
2.10.6	Präfix für Dateien definieren	18
2.10.7	Präfix für Diagramme definieren	18
2.10.8	Quanteda-Optionen setzen	18
2.11	LaTeX Konfiguration	18
2.11.1	LaTeX Parameter definieren	18
2.11.2	LaTeX Parameter schreiben	19
2.12	Parallelisierung aktivieren	19
2.12.1	Anzahl logischer Kerne festlegen	20
2.12.2	Quanteda	20
2.12.3	Data.table	20
3	Download vorbereiten	21
3.1	XML-Inhaltsverzeichnis einlesen	21
3.2	Links zu XML-Dateien aus XML-Inhaltsverzeichnis extrahieren	21
3.3	Debugging-Modus: Ausgewertete Links reduzieren	21
3.4	Links zu HTML Landing Pages generieren	21
3.5	Funktion anzeigen: f.linkextract	21
3.6	Links aus HTML Landing Pages extrahieren	22
3.7	Dateinamen von PDF und EPUB-Dateien in separate Vektoren sortieren	22

3.8	Test: Gleiche Länge der Dateinamen-Vektoren	22
3.9	Vektor der Langtitel erstellen	23
3.9.1	Namen bereinigen und kürzen	23
3.9.2	Indizes der AEG bestimmen	23
3.9.3	AEGs umbenennen	23
3.9.4	Indizes der VBVG bestimmen	23
3.9.5	VBVG umbenennen	23
3.10	Vektor der Kurztitel erstellen	23
3.11	Vektoren der Titel vereinigen	24
3.12	Prüfung auf Namens-Kollisionen	24
3.13	Bereinigung von Namens-Kollisionen	24
3.14	Dateierweiterungen hinzufügen	24
3.15	Links zu EPUB-Dateien erstellen	25
3.16	Links zu PDF-Dateien erstellen	25
3.17	Data Table für Download vorbereiten	25
3.18	Abkürzungsverzeichnis erstellen	25
3.19	Download Table als CSV speichern	26
3.20	Verzeichnis aller Rechtsakte als CSV speichern	26
3.21	Debugging-Modus: Anzahl der heruntergeladenen Dateien reduzieren	26
3.22	Anzahl herunterzuladender Dateien	26
3.22.1	Pro Format	26
3.22.2	Insgesamt	26
4	Verarbeitung der DTD und XML-Dateien mit Anlagen	27
4.1	Document Type Definition herunterladen	27
4.2	Download der XML-Dateien	27
4.3	Download-Ergebnis	27
4.3.1	Anzahl herunterzuladender Dateien	27
4.3.2	Anzahl heruntergeladener Dateien	27
4.3.3	Fehlbetrag	28
4.3.4	Fehlende Dateien	28
4.4	Extrahieren der XML-Dateien und ihrer Anlagen	28
4.5	XML Dateien auflisten und Dateigrößen speichern	28
4.6	Korpus erstellen: Einzelnormen	29
4.6.1	Beginn XML Parsing	32
4.6.2	Parallelisierung definieren	33
4.6.3	XML Parsen	33
4.6.4	Liste in Data Table umwandeln	33
4.6.5	Ende XML Parsing	33
4.6.6	Dauer XML Parsing	33
4.6.7	Variable “doc_id” erstellen	33
4.6.8	Variablen-Name für Ausfertigungsdatum anpassen	34
4.6.9	Variable “fundstellentyp” anpassen	34
4.6.10	Variable “builddate_iso” erstellen	34
4.6.11	Variable “aenderung_datum” erstellen	34
4.6.12	Variable “aufhebung_verkuendung_datum” erstellen	34
4.6.13	Variable “aufhebung_wirkung_datum” erstellen	34
4.6.14	Variable “neufassung_datum” erstellen	35
4.6.15	Variable “ausfertigung_jahr” hinzufügen	35

4.6.16	Variable “doi_concept” hinzufügen	35
4.6.17	Variable “doi_version” hinzufügen	35
4.6.18	Variable “version” hinzufügen	35
4.6.19	Variable “lizenz” hinzufügen	35
4.7	Stichprobe für Qualitätsprüfung ziehen	36
4.8	Korpus erstellen: Rechtsakte	37
4.8.1	Variablen definieren	37
4.8.2	Vollständiger Satz an Variablen	37
4.8.3	Einzelnormen zu Rechtsakten vereinigen	37
4.8.4	Variable “dateiname” in “doc_id” umbenennen	38
4.9	Datensatz erstellen: XML-Metadaten	39
4.9.1	Funktion für XML-Parsing definieren	39
4.9.2	Beginn XML Parsing	41
4.9.3	Parallelisierung definieren	41
4.9.4	XML Parsen	41
4.9.5	Liste in Data Table umwandeln	41
4.9.6	Ende XML Parsing	41
4.9.7	Dauer XML Parsing	41
4.9.8	Variablen-Name für Ausfertigungsdatum anpassen	42
4.9.9	Variable “fundstellentyp” anpassen	42
4.9.10	Variable “builddate_iso” erstellen	42
4.9.11	Variable “aenderung_datum” erstellen	42
4.9.12	Variable “aufhebung_verkuendung_datum” erstellen	42
4.9.13	Variable “aufhebung_wirkung_datum” erstellen	42
4.9.14	Variable “neufassung_datum” erstellen	43
4.9.15	Variable “ausfertigung_jahr” hinzufügen	43
4.9.16	Variable “doi_concept” hinzufügen	43
4.9.17	Variable “doi_version” hinzufügen	43
4.9.18	Variable “version” hinzufügen	43
4.9.19	Variable “lizenz” hinzufügen	43
4.10	Netzwerk-Analyse (experimentell!)	44
4.10.1	Funktion definieren: f.kennzahlen.search	44
4.10.2	Funktion definieren: f.kennzahlen.collapse	44
4.10.3	Funktion definieren: f.kennzahlen.edgelist	44
4.10.4	Funktion definieren: f.network.analysis	46
4.10.5	Netzwerk-Analyse durchführen	49
4.10.6	Beginn Network Analysis	50
4.10.7	Parallelisierung definieren	50
4.10.8	XML Parsen	50
4.10.9	XML-Dateien bei denen Fehler auftreten	50
4.10.10	Ende XML Parsing	51
4.10.11	Dauer XML Parsing	51
4.11	Wiederverpacken der XML-Dateien	51
4.11.1	XML-Dateien definieren	51
4.11.2	XML-Dateien verpacken	51
4.11.3	Anhänge zu XML-Dateien verpacken	51
5	Frequenztabellen erstellen: Einzelnormen	53
5.1	Funktion anzeigen: f.fast.freqtable	53

5.2	Liste zu prüfender Variablen	54
5.3	Frequenztabellen erstellen	55
6	Frequenztabellen erstellen: Rechtsakte	115
6.1	Variablen ignorieren	115
6.2	Liste zu prüfender Variablen	115
6.3	Frequenztabellen erstellen	115
7	Frequenztabellen erstellen: XML-Metadaten	124
7.1	Liste zu prüfender Variablen	124
7.2	Frequenztabellen erstellen	124
8	Frequenztabellen visualisieren	134
8.1	Präfixe erstellen	134
8.2	Tabellen für Einzelnormen einlesen	134
8.3	Tabellen für Rechtsakte einlesen	134
8.4	Tabellen für XML-Metadaten einlesen	134
8.5	Periodikum	135
8.5.1	Einzelnormen	135
8.5.2	Rechtsakte	137
8.5.3	XML-Metadaten	139
8.6	Ausfertigungsjahr	141
8.6.1	Einzelnormen	141
8.6.2	Rechtsakte	142
8.6.3	XML-Metadaten	143
9	Korpus-Analytik	144
9.1	Berechnung linguistischer Kennwerte	144
9.1.1	Funktion anzeigen: future_lingsummarize	144
9.1.2	Berechnung durchführen	145
9.2	Variablen-Namen anpassen	146
9.2.1	Einzelnormen	146
9.2.2	Rechtsakte	146
9.3	Kennwerte den Korpora hinzufügen	146
9.3.1	Einzelnormen	146
9.3.2	Rechtsakte	146
9.4	Varianten mit Metadaten erstellen	147
9.4.1	Einzelnormen	147
9.4.2	Rechtsakte	147
9.5	Linguistische Kennwerte: Einzelnormen	148
9.5.1	Zusammenfassungen berechnen	148
9.5.2	Zusammenfassungen anzeigen	150
9.5.3	Zusammenfassungen speichern	150
9.6	Linguistische Kennwerte: Rechtsakte	151
9.6.1	Zusammenfassungen berechnen	151
9.6.2	Zusammenfassungen anzeigen	153
9.6.3	Zusammenfassungen speichern	153
9.7	Verteilungen	154
9.7.1	Density (Zeichen)	154
9.7.2	Density (Tokens)	156

9.7.3	Density (Typen)	158
9.7.4	Density (Sätze)	160
9.8	Quantitative Variablen	161
9.8.1	Ausfertigungsdatum	161
9.8.2	Ausfertigungsjahr	162
10	Strenge Kontrolle der Variablen-Namen	164
10.1	Semantische Sortierung der Variablen	164
10.1.1	Variablen sortieren: Einzelnormen	164
10.1.2	Variablen sortieren: Rechtsakte	166
10.1.3	Variablen sortieren: XML-Metadaten	168
10.2	Anzahl Variablen der Datensätze	168
10.3	Alle Variablen-Namen der Datensätze	169
11	CSV-Dateien erstellen	172
11.1	Einzelnormen (Korpus)	172
11.1.1	Name für CSV definieren	172
11.1.2	Datensatz speichern	172
11.2	Einzelnormen (Metadaten)	172
11.2.1	Name für CSV definieren	172
11.2.2	Datensatz speichern	172
11.3	Rechtsakte (Korpus)	172
11.3.1	Name für CSV definieren	172
11.3.2	Datensatz speichern	172
11.4	Rechtsakte (Metadaten)	173
11.4.1	Name für CSV definieren	173
11.4.2	Datensatz speichern	173
11.5	XML-Metadaten	173
11.5.1	Name für CSV definieren	173
11.5.2	Datensatz speichern	173
12	Download der PDF-Dateien	174
12.1	Download durchführen	174
12.2	Download-Ergebnis	174
12.2.1	Anzahl herunterzuladender Dateien	174
12.2.2	Anzahl heruntergeladener Dateien	174
12.2.3	Fehlbetrag	175
12.2.4	Fehlende Dateien	175
13	TXT-Dateien erstellen	176
13.1	Anzahl zu extrahierender Dateien	176
13.2	Funktion anzeigen: future_pdf_to_txt	176
13.3	Text Extrahieren	177
14	Download der EPUB-Dateien	178
14.1	Download durchführen	178
14.2	Download-Ergebnis	178
14.2.1	Anzahl herunterzuladender Dateien	178
14.2.2	Anzahl heruntergeladener Dateien	178
14.2.3	Fehlbetrag	178

14.2.4 Fehlende Dateien	179
15 Dateigrößen analysieren	180
15.1 Gesamtgröße	180
15.1.1 PDF-Dateien (MB)	180
15.1.2 EPUB-Dateien (MB)	180
15.1.3 XML-Dateien (MB)	180
15.1.4 TXT-Dateien (MB)	181
15.1.5 Objekte in RAM (MB)	181
15.2 Verteilung der Dateigrößen (PDF)	182
15.3 Verteilung der Dateigrößen (EPUB)	183
15.4 Verteilung der Dateigrößen (XML)	184
15.5 Verteilung der Dateigrößen (TXT)	185
16 ZIP-Archive erstellen	186
16.1 Verpacken der CSV-Dateien	186
16.2 Verpacken der PDF-Dateien	186
16.3 Verpacken der TXT-Dateien	186
16.4 Verpacken der EPUB-Dateien	186
16.5 Verpacken der Netzwerk-Dateien	187
16.6 Verpacken der Analyse-Dateien	187
16.7 Verpacken der Source-Dateien	187
17 Roh-Dateien löschen	189
18 Kryptographische Hashes	190
18.1 Liste der ZIP-Archive erstellen	190
18.2 Funktion anzeigen: future_multihashes	190
18.3 Hashes berechnen	191
18.4 In Data Table umwandeln	191
18.5 Index hinzufügen	191
18.6 Hashes in CSV-Datei speichern	192
18.7 Leerzeichen hinzufügen um Zeilenumbruch zu ermöglichen	192
18.8 In Bericht anzeigen	193
19 Aufräumen	196
20 Abschluss	197
20.1 Datumsstempel	197
20.2 Datum und Uhrzeit (Anfang)	197
20.3 Datum und Uhrzeit (Ende)	197
20.4 Laufzeit des gesamten Skripts	197
20.5 Warnungen	197
21 Parameter für strenge Replikationen	198
Literaturverzeichnis	200

1 Corpus des Deutschen Bundesrechts (C-DBR)

1.1 Überblick

Das **Corpus des deutschen Bundesrechts (C-DBR)** ist eine möglichst vollständige Sammlung der konsolidierten Fassungen aller Gesetze und Verordnungen auf Bundesebene. Der Datensatz nutzt als seine Datenquelle das amtliche Internetangebot www.gesetze-im-internet.de des Bundesministeriums der Justiz und wertet dieses vollständig aus.

Alle mit diesem Skript erstellten Datensätze werden dauerhaft kostenlos und urheberrechtsfrei auf Zenodo, dem wissenschaftlichen Archiv des CERN, veröffentlicht. Alle Versionen sind mit einem separaten und langzeit-stabilen (persistenten) Digital Object Identifier (DOI) versehen.

Aktuellster, funktionaler und zitierfähiger Release des Datensatzes: <https://doi.org/10.5281/zenodo.3832111>

Aktuellster, funktionaler und zitierfähiger Release des Source Codes: <https://doi.org/10.5281/zenodo.4072934>

Lesen Sie bitte zuerst den *Compilation Report* auf Zenodo (via Source Code link)! Dieser enthält den gesamten R Source Code, relevante Rechenergebnisse, alle Diagramme, Zeitstempel, sowie ein detailliertes und klickbares Inhaltsverzeichnis. Sie werden sich auf diese Weise viel schneller im eigentlichen Source Code zurechtfinden.

1.2 Funktionsweise

Primäre Endprodukte des Skripts sind folgende ZIP-Archive:

1. Der volle Datensatz im CSV-Format, unterteilt in Einzelnormen; nur Rechtsakte mit veröffentlichtem Normtext sind erfasst
2. Die Metadaten aller Einzelnormen im CSV-Format (wie 1, nur ohne Normtexte)
3. Der volle Datensatz im CSV-Format, unterteilt in Rechtsakte; nur Rechtsakte mit veröffentlichtem Normtext sind erfasst
4. Die Metadaten aller Rechtsakte im CSV-Format (wie 3, nur ohne Normtexte)
5. Die Metadaten aller auf »Gesetze im Internet« als XML veröffentlichten Rechtsakte, im CSV-Format, unabhängig davon ob sie Normtext enthalten oder nicht
6. Der volle Datensatz im XML-Format, unterteilt in Rechtsakte; Grundlage für die CSV-Varianten
7. Alle Anlagen zu den XML-Dateien im jeweiligen Original-Format
8. Alle Rechtstexte im TXT-Format, unterteilt in Rechtsakte (deutlich reduzierter Umfang an Metadaten)
9. Alle Rechtstexte im PDF-Format, unterteilt in Rechtsakte (deutlich reduzierter Umfang an Metadaten)
10. Alle Rechtstexte im EPUB-Format, unterteilt in Gesetze (deutlich reduzierter Umfang an Metadaten)
11. Alle Analyse-Ergebnisse (Tabellen als CSV, Grafiken als PDF und PNG)
12. Netzwerk-Strukturen (Adjazenzmatrizen, Edgelisten, GraphML, und Netzwerk-Diagramme) für alle Rechtsakte (experimentell!)

Zusätzlich werden für alle ZIP-Archive kryptographische Signaturen (SHA2-256 und SHA3-512) berechnet und in einer CSV-Datei hinterlegt. Die Analyse-Ergebnisse werden zum Ende hin nicht gelöscht, damit sie für die Codebook-Erstellung verwendet werden können.

Weiterhin kann optional ein PDF-Bericht erstellt werden (siehe unter »Kompilierung«).

1.3 Kompilierung

Achtung: Verwenden Sie immer einen eigenständigen und *leeren* Ordner für die Kompilierung. Die Skripte löschen innerhalb des Ordners (working directory) vollautomatisch alle Dateien mit bestimmten Datei-Endungen (PDF, TXT, CSV usw.), die den Datensatz verunreinigen könnten — aber auch nur dort.

Alle Kommentare sind im roxygen2-Stil gehalten. Die beiden Skripte können daher auch ohne `render()` regulär als R-Skripte ausgeführt werden. Es wird in diesem Fall kein PDF-Bericht erstellt und Diagramme werden nicht abgespeichert.

Um den **vollständigen Datensatz** zu kompilieren, sowie Compilation Report und Codebook zu erstellen, kopieren Sie bitte alle im Source-Archiv bereitgestellten Dateien in einen leeren Ordner (!) und führen mit R diesen Befehl aus:

```
source("00_C-DBR_FullCompile.R")
```

Bei der Prüfung der GPG-Signatur im Codebook wird ein Fehler auftreten und im Codebook dokumentiert, weil die Daten nicht mit meiner Original-Signatur versehen sind. Dieser Fehler hat jedoch keine Auswirkungen auf die Funktionalität und hindert die Kompilierung nicht.

1.4 Systemanforderungen

1.4.1 Betriebssystem

Der Code in seiner veröffentlichten Form kann nur unter Linux ausgeführt werden, da er Linux-spezifische Optimierungen (z.B. Fork Cluster) und Shell-Kommandos (z.B. OpenSSL) nutzt. Der Code wurde unter Fedora Linux entwickelt und getestet. Die zur Kompilierung benutzte Version entnehmen Sie bitte dem `sessionInfo()`-Ausdruck am Ende des Compilation Reports im Zenodo-Archiv.

1.4.2 Software

Sie müssen die Programmiersprache R installiert haben. Starten Sie danach eine Session im Ordner des Projekts, Sie sollten automatisch zur Installation aller packages in der empfohlenen Version aufgefordert werden. Andernfalls führen Sie bitte folgenden Befehl aus:

```
renv::restore()
```

Um die PDF Reports zu kompilieren benötigen Sie eine LaTeX-Installation. Sie können diese auf Fedora wie folgt installieren:

```
sudo dnf install texlive-scheme-full
```

Alternativ können sie das R package **tinytex** installieren.

1.4.3 Parallelisierung

In der Standard-Einstellung wird das Skript vollautomatisch die maximale Anzahl an Rechenkernen/Threads auf dem System zu nutzen. Die Anzahl der verwendeten Kerne kann in der Konfigurationsdatei angepasst werden. Wenn die Anzahl Threads auf 1 gesetzt wird, ist die Parallelisierung deaktiviert.

1.4.4 Speicherplatz

Auf der Festplatte sollten 8 GB Speicherplatz vorhanden sein.

1.5 Weitere Open Access Veröffentlichungen (Fobbe)

Website — <https://www.seanfobbe.de>

Open Data — <https://zenodo.org/communities/sean-fobbe-data/>

Source Code — <https://zenodo.org/communities/sean-fobbe-code/>

Volltexte regulärer Publikationen — <https://zenodo.org/communities/sean-fobbe-publications/>

1.6 Kontakt

Fehler gefunden? Anregungen? Kommentieren Sie gerne im Issue Tracker auf GitHub oder schreiben Sie mir eine E-Mail an fobbe-data@posteo.de

2 Vorbereitung

2.1 Datumsstempel

Dieser Datumsstempel wird in alle Dateinamen eingefügt. Er wird am Anfang des Skripts gesetzt, für den den Fall, dass die Laufzeit die Datumsbarriere durchbricht.

```
datestamp <- Sys.Date()
print(datestamp)
```

```
## [1] "2022-08-05"
```

2.2 Datum und Uhrzeit (Beginn)

```
begin.script <- Sys.time()
print(begin.script)
```

```
## [1] "2022-08-05 21:45:43 CEST"
```

2.3 Packages Laden

```
library("zip")          # ZIP Files
```

```
##
## Attaching package: 'zip'
```

```
## The following objects are masked from 'package:utils':
##
##      unzip, zip
```

```
library("rvest")        # HTML/XML-Extraktion
library("xml2")          # Verarbeitung von XML-Format
library("RcppTOML")      # Verarbeitung von TOML-Format
library("knitr")         # Professionelles Reporting
library("kableExtra")    # Verbesserte Kable Tabellen
library("magick")        # Verarbeitung von Bild-Dateien
```

```
## Linking to ImageMagick 6.9.12.52
## Enabled features: cairo, fontconfig, freetype, ghostscript, lcms, pango, raw,
  rsvg, webp, x11
## Disabled features: fftw, heic
```

```
## Using 16 threads
```

```
library("pdftools")      # Extrahieren von PDF-Dateien
```

```
## Using poppler version 22.01.0
```

```
library("parallel")      # Parallelisierung
##library("doParallel")  # Parallelisierung
library("ggplot2")        # Fortgeschrittene Datenvisualisierung
library("data.table")     # Fortgeschrittene Datenverarbeitung
```

```
## data.table 1.14.2 using 8 threads (see ?getDTthreads). Latest news: r-
  datatable.com
```

```
library("quanteda")      # Fortgeschrittene Computerlinguistik
```

```
## Package version: 3.2.1
## Unicode version: 13.0
## ICU version: 69.1
```

```
## Parallel computing: 16 of 16 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
library("scales")        # Skalierung von Diagrammen
library("openssl")        # Kryptographische Signaturen; can possibly be deprecated
```

```
## Linking to: OpenSSL 3.0.5 5 Jul 2022
```

```
library("igraph")      # Analyse von Graphen
```

```
##  
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':  
##  
##      decompose, spectrum
```

```
## The following object is masked from 'package:base':  
##  
##      union
```

```
library("ggraph")      # Analyse von Graphen  
##library("qgraph")    # Analyse von Graphen  
library("future")      # Parallelisierung
```

```
##  
## Attaching package: 'future'
```

```
## The following objects are masked from 'package:igraph':  
##  
##      %>% , %<-%
```

```
library("future.apply") # Parallelisierung von base-r-Funktionen
```

2.4 Zusätzliche Funktionen einlesen

Hinweis: Die hieraus verwendeten Funktionen werden jeweils vor der ersten Benutzung in vollem Umfang angezeigt um den Lesefluss zu verbessern.

```
source("R-fobbe-proto-package/f.linkextract.R")  
source("R-fobbe-proto-package/f.fast.freqtable.R")  
#source("R-fobbe-proto-package/f.lingsummarize.iterator.R") # deprecated  
#source("R-fobbe-proto-package/f.dopar.pagenums.R") # deprecated  
#source("R-fobbe-proto-package/f.dopar.pdfextract.R") # deprecated  
#source("R-fobbe-proto-package/f.dopar.multihashes.R") # deprecated
```

```
source("functions/f.heading.transform.R")
source("functions/f.namechain.R")
source("functions/f.zero.NA.R")
source("functions/f.multihashes.R")
source("functions/f.future_multihashes.R")
source("functions/f.pdf_to_txt.R")
source("functions/f.future_pdf_to_txt.R")
source("functions/f.future_lingsummarize.R")
source("functions/f.download_robust.R")
source("functions/f.linkextract_regex.R")
```

2.5 Verzeichnis für Analyse-Ergebnisse und Diagramme definieren

Muss mit einem Schrägstrich enden!

```
dir.analysis <- paste0(getwd(),
                      "/analyse/")
```

2.6 Weitere Verzeichnisse definieren

```
dirs <- c("output",
         "temp",
         "netzwerke")
```

2.7 Dateien aus vorherigen Runs bereinigen

```
unlink(dir.analysis, recursive = TRUE)

unlink(dirs, recursive = TRUE)

files.delete <- list.files(pattern = "\\\\.zip|\\.xml|\\.jpe?g|\\.png|\\.gif|\\.pdf|\\.epub",
                           ignore.case = TRUE)

unlink(files.delete)
```

2.8 Verzeichnisse anlegen

```
dir.create(dir.analysis)

lapply(dirs, dir.create)
```

```
## [[1]]
## [1] TRUE
##
## [[2]]
## [1] TRUE
##
## [[3]]
## [1] TRUE
```

```
dir.create("netzwerke/Edgelisten")
dir.create("netzwerke/Adjazenzmatrizen")
dir.create("netzwerke/Netzwerkdiagramme")
dir.create("netzwerke/GraphML")
dir.create("netzwerke/Gliederungstabellen")
```

2.9 Vollzitate statistischer Software schreiben

```
knitr::write_bib(c(.packages()),
                 "temp/packages.bib")
```

2.10 Allgemeine Konfiguration

2.10.1 Konfiguration einlesen

```
config <- parseTOML("C-DBR_Config.toml")
```

2.10.2 Konfiguration anzeigen

```
print(config)
```

```
## List of 11
## $ cores      :List of 2
##   ..$ max     : logi TRUE
##   ..$ number: int 8
## $ debug      :List of 2
##   ..$ sample: int 500
##   ..$ toggle: logi FALSE
## $ doi        :List of 2
##   ..$ data     :List of 2
##   .. ..$ concept: chr "10.5281/zenodo.3832111"
##   .. ..$ version: chr "10.5281/zenodo.6959374"
##   ..$ software:List of 2
##   .. ..$ concept: chr "10.5281/zenodo.4072934"
```



```
## ..$ version: chr "10.5281/zenodo.6959381"
## $ download :List of 1
## ..$ timeout: int 600
## $ fig      :List of 3
## ..$ align : chr "center"
## ..$ dpi   : int 300
## ..$ format: chr [1:2] "pdf" "png"
## $ freqtable:List of 1
## ..$ ignore: chr [1:13] "periodikum" "fundstellentyp" "check_neuf" "check_
aufh" ...
## $ license  :List of 2
## ..$ code: chr "MIT-0"
## ..$ data: chr "Creative Commons Zero 1.0 Universal"
## $ parallel :List of 10
## ..$ downloadEPUB      : logi TRUE
## ..$ downloadPDF       : logi TRUE
## ..$ downloadXML       : logi TRUE
## ..$ extractPDF        : logi TRUE
## ..$ htmlLandingPages  : logi TRUE
## ..$ lingsummarize     : logi TRUE
## ..$ multihashes       : logi TRUE
## ..$ parseEinzelnormen: logi FALSE
## ..$ parseMeta         : logi FALSE
## ..$ parseNetworks    : logi FALSE
## $ project  :List of 3
## ..$ author   : chr "Seán Fobbe"
## ..$ fullname : chr "Corpus des Deutschen Bundesrechts"
## ..$ shortname: chr "C-DBR"
## $ qa       :List of 1
## ..$ sample: int 50
## $ quanteda :List of 1
## ..$ tokens_locale: chr "de_DE"
```

2.10.3 Knitr Optionen setzen

```
knitr::opts_chunk$set(fig.path = dir.analysis,
                      dev = config$fig$format,
                      dpi = config$fig$dpi,
                      fig.align = config$fig$align)
```

2.10.4 Download Timeout setzen

```
options(timeout = config$download$timeout)
```

2.10.5 Quellenangabe für Diagramme definieren

```
caption <- paste("Fobbe | DOI:",
                config$doi$data$version)
print(caption)
```

```
## [1] "Fobbe | DOI: 10.5281/zenodo.6959374"
```

2.10.6 Präfix für Dateien definieren

```
prefix.files <- paste0(config$project$shortname,  
                        "-",  
                        datestamp)  
print(prefix.files)
```

```
## [1] "C-DBR_2022-08-05"
```

2.10.7 Präfix für Diagramme definieren

```
prefix.figuretitle <- paste(config$project$shortname,  
                             "| Version",  
                             datestamp)
```

2.10.8 Quanteda-Optionen setzen

```
quanteda_options(tokens_locale = config$quanteda$tokens_locale)
```

2.11 LaTeX Konfiguration

2.11.1 LaTeX Parameter definieren

```
latexdefs <- c("%=====\\n% Definitionen\\n  
%=====  
            "\\n% NOTE: Diese Datei wurde während des Kompilierungs-Prozesses  
automatisch erstellt.\\n",  
            "\\n%-----Autor-----",  
            paste0("\\\\newcommand{\\projectauthor}{",  
                    config$project$author,  
                    "}"),  
            "\\n%-----Version-----",  
            paste0("\\\\newcommand{\\version}{",  
                    datestamp,  
                    "}"),  
            "\\n%-----Titles-----",  
            paste0("\\\\newcommand{\\datatitle}{",  
                    config$project$fullname,  
                    "}"),  
            paste0("\\\\newcommand{\\datashort}{",  
                    config$project$shortname,
```

```

    "}" ),
    paste0("\\newcommand{\\softwaretitle}{Source Code des \\enquote{",
            config$project$fullname,
            "}}"),
    paste0("\\newcommand{\\softwareshort}{",
            config$project$shortname,
            "-Source}"),
    "\\n%-----Data DOIs-----",
    paste0("\\newcommand{\\dataconceptdoi}{",
            config$doi$data$concept,
            "}" ),
    paste0("\\newcommand{\\dataversiondoi}{",
            config$doi$data$version,
            "}" ),
    paste0("\\newcommand{\\dataconcepturldoi}{https://doi.org/",
            config$doi$data$concept,
            "}" ),
    paste0("\\newcommand{\\dataversionurldoi}{https://doi.org/",
            config$doi$data$version,
            "}" ),
    "\\n%-----Software DOIs-----",
    paste0("\\newcommand{\\softwareconceptdoi}{",
            config$doi$software$concept,
            "}" ),
    paste0("\\newcommand{\\softwareversiondoi}{",
            config$doi$software$version,
            "}" ),

    paste0("\\newcommand{\\softwareconcepturldoi}{https://doi.org/",
            config$doi$software$concept,
            "}" ),
    paste0("\\newcommand{\\softwareversionurldoi}{https://doi.org/",
            config$doi$software$version,
            "}" ))

```

2.11.2 LaTeX Parameter schreiben

```

writeLines(latexdefs,
           paste0("temp/",
                  config$project$shortname,
                  "_Definitions.tex"))

```

2.12 Parallelisierung aktivieren

Parallelisierung wird zur Beschleunigung des XML-Parsings, der Konvertierung von PDF zu TXT und der Datenanalyse mittels **quanteda** und **data.table** verwendet. Die Anzahl threads wird automatisch auf das verfügbare Maximum des Systems gesetzt, kann aber auch nach Belieben auf das eigene System angepasst werden. Die Parallelisierung kann deaktiviert werden, indem die Variable **fullCores** auf 1 gesetzt wird.

Die hier verwendete Funktion **makeForkCluster()** ist viel schneller, funktioniert aber nur auf Unix-basierten Systemen (Linux, MacOS). Bei einer Ausführung unter Windows

sollten Sie `makecluster()` verwenden.

2.12.1 Anzahl logischer Kerne festlegen

```
if (config$cores$max == TRUE){  
  fullCores <- detectCores() - 1  
}  
  
if (config$cores$max == FALSE){  
  fullCores <- as.integer(config$cores$number)  
}  
  
print(fullCores)
```

```
## [1] 15
```

2.12.2 Quanteda

```
quanteda_options(threads = 1)
```

2.12.3 Data.table

```
setDTthreads(threads = fullCores)
```

3 Download vorbereiten

3.1 XML-Inhaltsverzeichnis einlesen

```
URL <- "https://www.gesetze-im-internet.de/gii-toc.xml"

XML <- read_xml(URL)
```

3.2 Links zu XML-Dateien aus XML-Inhaltsverzeichnis extrahieren

```
links <- html_elements(XML,
                        "link")

links.xml <- xml_text(links)
```

3.3 Debugging-Modus: Ausgewertete Links reduzieren

```
#if (config$debug$toggle == TRUE){

#   links.xml <- links.xml[sample(length(links.xml),
#                                 config$debug$sample)]

#}
```

3.4 Links zu HTML Landing Pages generieren

```
links.html <- gsub("/xml.zip",
                  "/index.html",
                  links.xml)
```

3.5 Funktion anzeigen: f.linkextract

```
print(f.linkextract)
```

```
## function(URL){
##   tryCatch({
##     read_html(URL) %>%
##       html_nodes("a")%>%
##       html_attr('href')},
##     error = function(cond) {
##       return(NA)}
##   )
## }
```

3.6 Links aus HTML Landing Pages extrahieren

```
if(config$parallel$htmlLandingPages == TRUE){  
  plan("multicore",  
        workers = fullCores)  
}else{  
  plan("sequential")  
}
```

```
names.list <- future_lapply(links.html,  
                             f.linkextract_regex,  
                             regex = "(.pdf$)|(.epub$)")
```

3.7 Dateinamen von PDF und EPUB-Dateien in separate Vektoren sortieren

```
# PDF  
names.pdf <- lapply(names.list, grep, pattern = "pdf", value = TRUE)  
names.pdf <- lapply(names.pdf, f.zero.NA)  
filenames.pdf <- unlist(names.pdf)  
  
# EPUB  
names.epub <- lapply(names.list, grep, pattern = "epub", value = TRUE)  
names.epub <- lapply(names.epub, f.zero.NA)  
filenames.epub <- unlist(names.epub)
```

3.8 Test: Gleiche Länge der Dateinamen-Vektoren

```
if (!length(links.html) == length(filenames.pdf)){  
  stop("PDF-Dateinamen sind fehlerhaft.")  
}  
  
if (!length(links.html) == length(filenames.epub)){  
  stop("EPUB-Dateinamen sind fehlerhaft.")  
}
```

3.9 Vektor der Langtitel erstellen

Hinweis: Es gibt zwei Rechtsakte mit dem Namen “Allgemeine Eisenbahngesetz,” obwohl es sich um zwei unterschiedliche Rechtsakte handelt. Die beiden Rechtsakte werden daher um ihr jeweiliges Ausfertigungsjahr ergänzt um die Dateinamen einzigartig zu machen.

```
longtitle.raw <- html_elements(XML, "title") %>% xml_text()
```

3.9.1 Namen bereinigen und kürzen

```
longtitle <- gsub("[[:punct:]]", "", longtitle.raw)
longtitle <- gsub(" ", "-", longtitle)
```

3.9.2 Indizes der AEG bestimmen

```
AEG.index <- grep("Allgemeines-Eisenbahngesetz", longtitle)
```

3.9.3 AEGs umbenennen

```
longtitle[AEG.index] <- c("Allgemeines-Eisenbahngesetz-1993",
                          "Allgemeines-Eisenbahngesetz-1951")
```

3.9.4 Indizes der VBVG bestimmen

```
vormund.index <- grep("Vormünder", longtitle)
```

3.9.5 VBVG umbenennen

```
longtitle[vormund.index] <- c("Gesetz-über-die-Vergütung-von-Vormündern-und-
  Betreuern-2005",
                             "Gesetz-über-die-Vergütung-von-Vormündern-und-Betreuern
  -2023")
```

3.10 Vektor der Kurztitel erstellen

```
shorttitle <- filenames.pdf
shorttitle <- gsub(".pdf",
                  "",
                  shorttitle)
```

```
shorttitle <- gsub("_",  
                  "",  
                  shorttitle)
```

3.11 Vektoren der Titel vereinigen

Die Kurz- und Langtitel werden zu einem Vektor zusammengefügt. Dieser wird dann auf maximal 200 Zeichen gekürzt, damit keine Probleme für Windows-User entstehen.

```
title <- paste(shorttitle,  
               longtitle,  
               sep="_")  
  
title <- strtrim(title,  
                 200)
```

3.12 Prüfung auf Namens-Kollisionen

Kollidierende Namen anzeigen. Wenn Namens-Kollisionen bestehen (wie oben beim AEG) müssen diese unbedingt bereinigt werden, weil ansonsten beim Herunterladen eine Datei alle anderen mit dem gleichen Namen überschreibt.

```
title[duplicated(title)]
```

```
## character(0)
```

3.13 Bereinigung von Namens-Kollisionen

Eine manuelle Bereinigung von Kollisionen ist bevorzugt. Falls keine manuelle Bereinigung stattgefunden hat wird in diesem Schritt eine automatische Bereinigung durchgeführt.

```
title <- make.unique(title,  
                     sep = "-")
```

3.14 Dateierweiterungen hinzufügen

```
title.xml <- paste0(title, ".zip")  
title.epub <- paste0(title, ".epub")  
title.pdf <- paste0(title, ".pdf")
```


3.15 Links zu EPUB-Dateien erstellen

```
prelinks.epub <- gsub("xml.zip",  
                      "",  
                      links.xml)  
  
links.epub <- paste0(prelinks.epub,  
                     filenames.epub)
```

3.16 Links zu PDF-Dateien erstellen

```
prelinks.pdf <- gsub("xml.zip",  
                    "",  
                    links.xml)  
  
links.pdf <- paste0(prelinks.pdf,  
                   filenames.pdf)
```

3.17 Data Table für Download vorbereiten

```
download <- data.table(title.xml,  
                       links.xml,  
                       title.epub,  
                       links.epub,  
                       title.pdf,  
                       links.pdf)
```

3.18 Abkürzungsverzeichnis erstellen

```
ID <- gsub("\\\\.epub",  
          "",  
          filenames.epub)  
  
conctable <- data.table(ID,  
                        shorttitle,  
                        longtitle.raw)  
  
colnames(conctable) <- c("ID",  
                          "Kurztitel",  
                          "Langtitel")
```

3.19 Download Table als CSV speichern

```
fwrite(download,
       paste0(dir.analysis,
              config$project$shortname,
              "_02_Links.csv"),
       na = "NA")
```

3.20 Verzeichnis aller Rechtsakte als CSV speichern

```
fwrite(conctable,
       paste0("output/",
              prefix.files,
              "_DE_AlleRechtsakteVerzeichnis.csv"),
       na = "NA")
```

3.21 Debugging-Modus: Anzahl der heruntergeladenen Dateien reduzieren

```
if (config$debug$toggle == TRUE){
  download <- download[sample(download[, .N],
                              config$debug$sample)]
}
```

3.22 Anzahl herunterzuladender Dateien

3.22.1 Pro Format

```
download[, .N]
```

```
## [1] 6659
```

3.22.2 Insgesamt

```
download[, .N] * 3
```

```
## [1] 19977
```

4 Verarbeitung der DTD und XML-Dateien mit Anlagen

4.1 Document Type Definition herunterladen

Die Document Type Definition (DTD) “definiert den Aufbau des XML-Formats zur Veröffentlichung der aktuellen Bundesgesetze und Rechtsverordnungen ueber www.gesetze-im-internet.de” (Zitat aus dem Inhalt der Datei).

```
download.file("https://www.gesetze-im-internet.de/dtd/1.01/gii-norm.dtd",
             paste0("output/",
                   prefix.files,
                   "_DE_XML_DocumentTypeDefinition_v1-01.dtd"))
```

4.2 Download der XML-Dateien

```
if(config$parallel$downloadXML == TRUE){
  plan("multicore",
      workers = fullCores)
}else{
  plan("sequential")
}
```

```
future_mapply(f.download_robust,
             url = download$links.xml,
             destfile = download$title.xml)
```

4.3 Download-Ergebnis

4.3.1 Anzahl herunterzuladender Dateien

```
download[, .N]
```

```
## [1] 6659
```

4.3.2 Anzahl heruntergeladener Dateien

```
files.zip <- list.files(pattern = "\\\\.zip")
length(files.zip)
```

```
## [1] 6659
```

4.3.3 Fehlbetrag

```
N.missing <- download[,.N] - length(files.zip)
print(N.missing)
```

```
## [1] 0
```

4.3.4 Fehlende Dateien

```
missing <- setdiff(download$title.xml,
                    files.zip)
print(missing)
```

```
## character(0)
```

4.4 Extrahieren der XML-Dateien und ihrer Anlagen

XML-Dateien und ihre Anlagen sind einzeln nach Rechtsakten in ZIP-Archiven verpackt. Diese werden nun extrahiert und die ZIP-Archive im Anschluss gelöscht.

```
files.zip <- list.files(pattern = "\\\\.zip",
                        ignore.case = TRUE)

for (file in files.zip){
  unzip(zipfile = file)
}

unlink(files.zip)
```

4.5 XML Dateien auflisten und Dateigrößen speichern

```
files.xml <- list.files(pattern = "\\\\.xml",
                        ignore.case = TRUE)

xml.MB <- file.size(files.xml) / 10^6
```

4.6 Korpus erstellen: Einzelnormen

Wichtiger Hinweis: Es werden für diese Variante nur Rechtsakte ausgewertet, bei denen mindestens eine Einzelnorm mit Text-Inhalt vorhanden ist!

Die XML-Daten enthalten keine Leerzeichen zwischen den XML-Tags, sowie zwischen den XML-Tags und ihrem Inhalt. Damit beim Entfernen der XML-Tags keine Inhalte zusammengefügt werden, wird die XML-Datei zunächst als Character-Vektor eingelesen, Leerzeichen hinzugefügt und im Anschluss erst die XML-Struktur eingelesen. Zwischen dem Anfang des Dokuments und dem ersten XML-Tag darf kein Leerzeichen sein, dieses wird einzeln nachkorrigiert. Zusätzlicher whitespace ist bei späterer Text-Verarbeitung unschädlich und wird im Rahmen der Tokenisierung praktisch immer entfernt.

Ohne diesen Schritt können Ergebnisse so aussehen: “Zollkodex,d)alle Verfahren” ###
Funktion für XML-Parsing definieren

```
xmlparse.einzelnormen <- function(file.xml){  
  
  ## XML als Character-Vektor einlesen  
  xml.char <- readChar(file.xml,  
                        file.info(file.xml)$size)  
  
  ## Leerzeichen einfügen  
  xml.char <- gsub(">", "> ", xml.char)  
  xml.char <- gsub("<", " <", xml.char)  
  xml.char <- sub(" <", "<", xml.char)  
  
  ## XML-Struktur lesen  
  XML <- read_xml(xml.char)  
  
  ## Schleife vorbereiten  
  nodes <- html_elements(XML, xpath = "//norm")  
  scope <- seq_along(nodes)  
  
  ## Inhaltsdaten extrahieren  
  text.temp <- vector("list", max(scope))  
  enbez.temp <- vector("list", max(scope))  
  g.kennzahl.temp <- vector("list", max(scope))  
  g.bez.temp <- vector("list", max(scope))  
  g.titel.temp <- vector("list", max(scope))  
  
  for (i in scope){  
  
    text.temp[[i]] <- html_elements(nodes[i],  
                                   xpath = "textdaten//text//Content") %>% xml_  
text(trim = TRUE)  
  
    enbez.temp[[i]] <- html_elements(nodes[i],  
                                   xpath = "metadaten//enbez") %>% xml_text(  
trim = TRUE)  
  
    g.kennzahl.temp[[i]] <- html_element(nodes[i],  
                                       xpath = "metadaten//gliederungseinheit/  
/gliederungskennzahl") %>% xml_text(trim = TRUE)
```

```

g.bez.temp[[i]] <- html_elements(nodes[i],
                                xpath = "metadaten//gliederungseinheit//
gliederungsbez") %>% xml_text(trim = TRUE)

g.titel.temp[[i]] <- html_elements(nodes[i],
                                   xpath = "metadaten//gliederungseinheit//
gliederungstitel") %>% xml_text(trim = TRUE)

}

## Leere Elemente mit NA kennzeichnen
enbez <- sapply(enbez.temp, f.zero.NA)
text <- sapply(text.temp, f.zero.NA)
g.kennzahl.pos <- sapply(g.kennzahl.temp, f.zero.NA)
g.bez.pos <- sapply(g.bez.temp, f.zero.NA)
g.titel.pos <- sapply(g.titel.temp, f.zero.NA)

## Gliederungsinformationen transformieren
gliederungskennzahl <- f.heading.transform(g.kennzahl.pos)
gliederungsbez <- f.heading.transform(g.bez.pos)
gliederungstitel <- f.heading.transform(g.titel.pos)

## Grundlage für Ketten extrahieren
g.kennzahl.vec <- html_elements(XML, xpath = "//norm//gliederungskennzahl")
%>% xml_text(trim = TRUE)
g.bez.vec <- html_elements(XML, xpath = "//norm//gliederungsbez") %>% xml_
text(trim = TRUE)
g.titel.vec <- html_elements(XML, xpath = "//norm//gliederungstitel") %>% xml
_text(trim = TRUE)

## Ketten anhand von Gliederungskennzahlen erstellen
chain.dt <- f.namechain(g.kennzahl.vec,
                       g.titel.vec,
                       g.bez.vec)

## Ketten einfügen
titelkette <- chain.dt$titelchain[match(gliederungskennzahl,
                                       chain.dt$einzelzahl)]

bezketten <- chain.dt$bezchain[match(gliederungskennzahl,
                                   chain.dt$einzelzahl)]

## Build Date extrahieren
builddate_original <- xml_attr(nodes, attr = "builddate")

## Content Data Table erstellen
content.out <- data.table(builddate_original,
                          gliederungskennzahl,
                          gliederungsbez,
                          bezkette,
                          gliederungstitel,
                          titelkette,
                          enbez,
                          text)

```

```

content.out <- content.out[text != ""]

## Allgemeine Metadaten extrahieren

varlist <- c("jurabk",
            "amtabk",
            "ausfertigung-datum",
            "periodikum",
            "zitstelle",
            "langue",
            "kurzue")

meta <- vector("list", length(varlist))

for (i in 1:length(varlist)){

  temp <- html_element(XML, varlist[i]) %>% xml_text(trim = TRUE)
  meta[[i]] <- rep(temp,
                  content.out[,.N])

}

setDT(meta)
setnames(meta, new = varlist)

meta$fundstellentyp <- rep(html_element(XML, "fundstelle") %>% xml_attr(attr
= "typ"),
                        content.out[,.N])

meta$dateiname <- rep(file.xml,
                    content.out[,.N])

## Standangaben extrahieren
standtyp <- html_elements(XML, "standtyp") %>% xml_text(trim = TRUE)
standkommentar <- html_elements(XML, "standkommentar") %>% xml_text(trim =
TRUE)
standcheck <- html_elements(XML, "standangabe") %>% xml_attr(attr = "checked"
)

dt.stand <- data.table(standtyp,
                      standkommentar,
                      standcheck)

if (dt.stand[,.N] > 0){

  ## Standkommentar
  dt.typ <- dt.stand[,
                    lapply(list(standkommentar),
                          function(x)paste(x, collapse = " | ")),
                    keyby = c("standtyp")]

  setnames(dt.typ,
            "V1",

```

```

        "standkommentar")

dt.typ <- transpose(dt.typ,
                    make.names = "standtyp")

setnames(dt.typ,
         names(dt.typ),
         tolower(names(dt.typ)))

## Standcheck
dt.check <- dt.stand[,lapply(.SD, as.factor)][, .(standtyp, standcheck)]
dt.check <- dt.check[, lapply(list(standtyp), unique), keyby = "
standcheck"]
setnames(dt.check,
        "V1",
        "standtyp")

dt.check <- transpose(dt.check, make.names = "standtyp")

setnames(dt.check,
        names(dt.check),
        paste0("check_",
               tolower(names(dt.check))))

dt.stand.all <- cbind(dt.typ, dt.check)

dt.stand.all.rep <- dt.stand.all[rep(dt.stand.all[, .I],
                                   content.out[, .N])]

out.dt <- cbind(meta,
                dt.stand.all.rep,
                content.out)
}else{
  out.dt <- cbind(meta,
                  content.out)
}

return(out.dt)
}

xmlparse.einzelnormen.robust <- function(file.xml){
  tryCatch({xmlparse.einzelnormen(file.xml)},
    error = function(cond) {
      return(NA)}
  )
}

```

4.6.1 Beginn XML Parsing

```
begin.parse <- Sys.time()
```


4.6.2 Parallelisierung definieren

```
if(config$parallel$parseEinzelnormen == TRUE){  
  plan("multicore",  
        workers = fullCores)  
}else{  
  plan("sequential")  
}
```

4.6.3 XML Parsen

```
out.einzelnormen <- future_lapply(files.xml,  
                                  xmlparse.einzelnormen.robust)
```

4.6.4 Liste in Data Table umwandeln

```
dt.normen <- rbindlist(out.einzelnormen,  
                       use.names = TRUE,  
                       fill = TRUE)
```

4.6.5 Ende XML Parsing

```
end.parse <- Sys.time()
```

4.6.6 Dauer XML Parsing

```
end.parse - begin.parse
```

```
## Time difference of 3.950533 mins
```

4.6.7 Variable “doc_id” erstellen

Eine einzigartige doc_id wird benötigt um z.B. einen Quanteda-Korpus erstellen zu können. Diese wird aus dem Dateinamen zusammen mit einer Kollisionsnummer gebildet.

```
dt.normen$doc_id <- make.unique(dt.normen$dateiname)
```

4.6.8 Variablen-Name für Ausfertigungsdatum anpassen

```
setnames(dt.normen,  
  "ausfertigung-datum",  
  "ausfertigung_datum")
```

4.6.9 Variable “fundstellentyp” anpassen

```
dt.normen[grepl("amtlich",  
  dt.normen$fundstellentyp,  
  invert = TRUE)]$fundstellentyp <- "nichtamtlich"
```

4.6.10 Variable “builddate_iso” erstellen

```
dt.normen$builddate_iso <- as.POSIXct(dt.normen$builddate_original,  
  format = "%Y%m%d%H%M%S")
```

4.6.11 Variable “aenderung_datum” erstellen

```
dt.normen$aenderung_datum <- as.Date(gsub(".*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4})  
  .*",  
    "\\1",  
    dt.normen$stand),  
  format = "%d.%m.%Y")
```

4.6.12 Variable “aufhebung_verkuendung_datum” erstellen

Das Textfeld mit Informationen zur Aufhebung enthält zwei Daten. Das erste ist das der Verkündung des aufhebenden Rechtsaktes, das zweite das der Wirkung des aufhebenden Rechtsaktes. Für diese Variable wird das erste Datum verwendet.

```
dt.normen$aufhebung_verkuendung_datum <- as.Date(sub(".*  
  ([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*",  
    "\\1",  
    dt.normen$aufh),  
  format = "%d.%m.%Y")
```

4.6.13 Variable “aufhebung_wirkung_datum” erstellen

Das Textfeld mit Informationen zur Aufhebung enthält zwei Daten. Das erste ist das der Verkündung des aufhebenden Rechtsaktes, das zweite das der Wirkung des aufhebenden Rechtsaktes. Für diese Variable wird das zweite Datum verwendet.

```
dt.normen$aufhebung_wirkung_datum <- as.Date(sub(".*
([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*"
,
                                "\\2",
                                dt.normen$aufh),
                                format = "%d.%m.%Y")
```

4.6.14 Variable “neufassung_datum” erstellen

```
dt.normen$neufassung_datum <- as.Date(gsub(".*
([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*",
                                "\\1",
                                dt.normen$neuf),
                                format = "%d.%m.%Y")
```

4.6.15 Variable “ausfertigung_jahr” hinzufügen

```
dt.normen$ausfertigung_jahr <- year(dt.normen$ausfertigung_datum)
```

4.6.16 Variable “doi_concept” hinzufügen

```
dt.normen$doi_concept <- rep(config$doi$data$concept,
                                dt.normen[,.N])
```

4.6.17 Variable “doi_version” hinzufügen

```
dt.normen$doi_version <- rep(config$doi$data$version,
                                dt.normen[,.N])
```

4.6.18 Variable “version” hinzufügen

```
dt.normen$version <- as.character(rep(datestamp,
                                dt.normen[,.N]))
```

4.6.19 Variable “lizenz” hinzufügen

```
dt.normen$lizenz <- as.character(rep(config$license$data,
                                dt.normen[,.N]))
```

4.7 Stichprobe für Qualitätsprüfung ziehen

```
print(config$qa$sample)
```

```
## [1] 50
```

```
idx <- sample(dt.normen[,.N],  
              config$qa$sample)  
  
check <- dt.normen[idx]  
  
fwrite(check,  
        paste0(dir.analysis,  
                prefix.files,  
                "_Stichprobe_Normen.csv"),  
        na = "NA")
```

4.8 Korpus erstellen: Rechtsakte

4.8.1 Variablen definieren

Zunächst der vordefinierte Satz an Metadaten.

```
varlist.r1 <- c("jurabk",  
               "amtabk",  
               "ausfertigung_datum",  
               "periodikum",  
               "zitstelle",  
               "langue",  
               "kurzue")
```

Die Stand-Variablen haben immer auch ein Pendant das mit "check_" beginnt.

```
standvars <- c("stand",  
              "aufh",  
              "neuf",  
              "hinweis",  
              "sonst")  
  
standvars <- c(standvars,  
              paste0("check_",  
                    standvars))
```

4.8.2 Vollständiger Satz an Variablen

```
varlist.r2 <- c(varlist.r1,  
               standvars,  
               "fundstellentyp",  
               "ausfertigung_jahr",  
               "aenderung_datum",  
               "aufhebung_verkuendung_datum",  
               "aufhebung_wirkung_datum",  
               "neufassung_datum",  
               "doi_concept",  
               "doi_version",  
               "version",  
               "lizenz")
```

4.8.3 Einzelnormen zu Rechtsakten vereinigen

```
text.rechtsakte <- dt.normen[,  
                             lapply(list(text),  
                                   function(x)paste(x, collapse = " ")),  
                             keyby = dateiname]
```

```

setnames(text.rechtsakte,
         "V1",
         "text")

meta.rechtsakte <- dt.normen[,
                             lapply(.SD, unique),
                             .SDcols = varlist.r2,
                             keyby = dateiname]

dt.rechtsakte <- text.rechtsakte[meta.rechtsakte,
                                on = "dateiname"]

```

4.8.4 Variable “dateiname” in “doc_id” umbenennen

```

setnames(dt.rechtsakte,
         "dateiname",
         "doc_id")

```

4.9 Datensatz erstellen: XML-Metadaten

An dieser Stelle werden Metadaten für alle Rechtsakte von “Gesetze im Internet” erhoben, unabhängig davon ob die Rechtsakte Text enthalten oder nur mit Überschrift nachgewiesen sind.

4.9.1 Funktion für XML-Parsing definieren

```
xmlparse.meta <- function(file.xml){  
  
  ## XML-Struktur lesen  
  XML <- read_xml(file.xml)  
  
  ## Schleife vorbereiten  
  nodes <- html_elements(XML, xpath = "//norm//metadaten")  
  scope <- 1:length(nodes)  
  
  ## Metadaten extrahieren  
  
  varlist <- c("jurabk",  
              "amtabk",  
              "ausfertigung-datum",  
              "periodikum",  
              "zitstelle",  
              "langue",  
              "kurzue")  
  
  meta <- vector("list", length(varlist))  
  
  for (i in 1:length(varlist)){  
    meta[[i]] <- html_element(XML, varlist[i]) %>% xml_text()  
  }  
  
  setDT(meta)  
  setnames(meta, new = varlist)  
  
  meta$fundstellentyp <- html_element(XML, "fundstelle") %>% xml_attr(attr = "  
  typ")  
  
  meta$doc_id <- file.xml  
  
  meta$bulddate_original <- xml_attr(XML, attr = "bulddate")  
  
  ## Standangaben extrahieren  
  standtyp <- html_elements(XML, "standtyp") %>% xml_text(trim = TRUE)  
  standkommentar <- html_elements(XML, "standkommentar") %>% xml_text(trim =  
  TRUE)  
  standcheck <- html_elements(XML, "standangabe") %>% xml_attr(attr = "checked"  
  )  
  
  dt.stand <- data.table(standtyp,  
                        standkommentar,  
                        standcheck)
```

```

if (dt.stand[,.N] > 0){

  ## Standkommentar
  dt.typ <- dt.stand[,
    lapply(list(standkommentar),
      function(x)paste(x, collapse = "  ")),
    keyby = c("standtyp")]

  setnames(dt.typ,
    "V1",
    "standkommentar")

  dt.typ <- transpose(dt.typ,
    make.names = "standtyp")

  setnames(dt.typ,
    names(dt.typ),
    tolower(names(dt.typ)))

  ## Standcheck
  dt.check <- dt.stand[,lapply(.SD, as.factor)][, .(standtyp, standcheck)]
  dt.check <- dt.check[, lapply(list(standtyp), unique), keyby = "
standcheck"]
  setnames(dt.check,
    "V1",
    "standtyp")

  dt.check <- transpose(dt.check,
    make.names = "standtyp")

  setnames(dt.check,
    names(dt.check),
    paste0("check_",
      tolower(names(dt.check))))

  dt.stand.all <- cbind(dt.typ, dt.check)

  meta <- cbind(meta,
    dt.stand.all)
}

return(meta)
}

xmlparse.meta.robust <- function(file.xml){
  tryCatch({xmlparse.meta(file.xml)},
    error = function(cond) {
      return(NA)}
  )
}

```


4.9.2 Beginn XML Parsing

```
begin.parse <- Sys.time()
```

4.9.3 Parallelisierung definieren

```
if(config$parallel$parseMeta == TRUE){  
  plan("multicore",  
        workers = fullCores)  
}else{  
  plan("sequential")  
}
```

4.9.4 XML Parsen

```
out.meta <- future_lapply(files.xml,  
                           xmlparse.meta.robust)
```

4.9.5 Liste in Data Table umwandeln

```
dt.meta <- rbindlist(out.meta,  
                     use.names = TRUE,  
                     fill = TRUE)
```

4.9.6 Ende XML Parsing

```
end.parse <- Sys.time()
```

4.9.7 Dauer XML Parsing

```
end.parse - begin.parse
```

```
## Time difference of 1.023563 mins
```

4.9.8 Variablen-Name für Ausfertigungsdatum anpassen

```
setnames(dt.meta,  
         "ausfertigung-datum",  
         "ausfertigung_datum")
```

4.9.9 Variable “fundstellentyp” anpassen

```
dt.meta[grep("amtlich", dt.meta$fundstellentyp, invert = TRUE)]$fundstellentyp <-  
  "nichtamtlich"
```

4.9.10 Variable “builddate_iso” erstellen

```
dt.meta$builddate_iso <- as.POSIXct(dt.meta$builddate_original,  
                                   format = "%Y%m%d%H%M%S")
```

4.9.11 Variable “aenderung_datum” erstellen

```
dt.meta$aenderung_datum <- as.Date(gsub(".*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*",  
                                       "\\1",  
                                       dt.meta$stand),  
                                format = "%d.%m.%Y")
```

4.9.12 Variable “aufhebung_verkuendung_datum” erstellen

Das Textfeld mit Informationen zur Aufhebung enthält zwei Daten. Das erste ist das der Verkündung des aufhebenden Rechtsaktes, das zweite das der Wirkung des aufhebenden Rechtsaktes. Für diese Variable wird das erste Datum verwendet.

```
dt.meta$aufhebung_verkuendung_datum <- as.Date(sub(".*  
  ([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*",  
  ,  
  "\\1",  
  dt.meta$aufh),  
  format = "%d.%m.%Y")
```

4.9.13 Variable “aufhebung_wirkung_datum” erstellen

Das Textfeld mit Informationen zur Aufhebung enthält zwei Daten. Das erste ist das der Verkündung des aufhebenden Rechtsaktes, das zweite das der Wirkung des aufhebenden Rechtsaktes. Für diese Variable wird das zweite Datum verwendet.

```
dt.meta$aufhebung_wirkung_datum <- as.Date(sub(".*
([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4}).*"
,
                                "\\2",
                                dt.meta$aufh),
                                format = "%d.%m.%Y")
```

4.9.14 Variable “neufassung_datum” erstellen

```
dt.meta$neufassung_datum <- as.Date(gsub(".*([0-9]{1,2}\\.[0-9]{1,2}\\.[0-9]{4})."
*,
                                "\\1",
                                dt.meta$neuf),
                                format = "%d.%m.%Y")
```

4.9.15 Variable “ausfertigung_jahr” hinzufügen

```
dt.meta$ausfertigung_jahr <- year(as.IDate(dt.meta$ausfertigung_datum))
```

4.9.16 Variable “doi_concept” hinzufügen

```
dt.meta$doi_concept <- rep(config$doi$data$concept, dt.meta[,.N])
```

4.9.17 Variable “doi_version” hinzufügen

```
dt.meta$doi_version <- rep(config$doi$data$version, dt.meta[,.N])
```

4.9.18 Variable “version” hinzufügen

```
dt.meta$version <- as.character(rep(datestamp, dt.meta[,.N]))
```

4.9.19 Variable “lizenz” hinzufügen

```
dt.meta$lizenz <- as.character(rep(config$license$data,
                                dt.meta[,.N]))
```

4.10 Netzwerk-Analyse (experimentell!)

4.10.1 Funktion definieren: f.kennzahlen.search

```
f.kennzahlen.search <- function(pattern, targetvec){  
  
  pattern.N <- nchar(pattern)  
  target <- substr(targetvec, 1, pattern.N)  
  targetvec[grepl(pattern, target, fixed = TRUE)]  
  
}
```

4.10.2 Funktion definieren: f.kennzahlen.collapse

```
f.kennzahlen.collapse <- function(lev.begin, targets.list){  
  
  out.list <- vector("list", length(targets.list))  
  
  for (i in 1:length(targets.list)){  
  
    targets.vector <- targets.list[[i]]  
  
    out.list[[i]] <- data.table(rep(lev.begin[i],  
                                  length(targets.vector)),  
                              targets.vector)  
  
  }  
  
  out.vec <- rbindlist(out.list)  
  return(out.vec)  
  
}
```

4.10.3 Funktion definieren: f.kennzahlen.edgelist

f.kennzahlen.edgelist: erstellt aus einem vektor an Gliederungskennzahlen und dem Gesetzesnamen ein Netzwerk-Diagramm der Inhaltsstruktur. Basiert auf f.kennzahlen.search und f.kennzahlen.collapse.

```
f.kennzahlen.edgelist <- function(kennzahl, name){  
  
  level <- nchar(kennzahl) / 3  
  
  level.unique <- sort(unique(level))  
  
  depth.begin <- head(seq_along(level.unique), -1)  
  depth.end <- depth.begin + 1  
  
  out.list <- vector("list", length(depth.begin))
```

```

for (i in seq_along(depth.begin)){

  lev.begin <- kennzahl[level == depth.begin[i]]
  lev.end <- kennzahl[level == depth.end[i]]

  targets.list <- lapply(lev.begin, f.kennzahlen.search, lev.end)
  out.list[[i]] <- f.kennzahlen.collapse(lev.begin, targets.list)

}

out.dt <- rbindlist(out.list)

## Add zero level

if (length(depth.begin != 0)){
  lev1 <- kennzahl[level == depth.begin[1]]

  zerolinks <- data.table(rep(name, length(lev1)),
                          lev1)

  out.dt <- rbind(zerolinks,
                  out.dt,
                  use.names = FALSE)
}else{
  lev1 <- kennzahl
  out.dt <- data.table(rep(name, length(lev1)),
                      lev1)

}

setnames(out.dt,
          new = c("from",
                  "to"))

return(out.dt)
}

f.split.gliederungseinheit <- function(gliederungseinheit){

  kennzahl <- html_elements(gliederungseinheit, xpath = "gliederungskennzahl")
  %>% xml_text()

  bez <- html_elements(gliederungseinheit, xpath = "gliederungsbez") %>% xml_text()

  # Newlines, damit Umbrüche in Diagrammen funktionieren
  bez <- gsub(" +",
              "\n",
              bez)

  titel <- html_elements(gliederungseinheit, xpath = "gliederungstitel") %>%
  xml_text()

```

```

titel <- gsub(" +",
             "\n",
             titel)

if(length(titel) == 0){
  titel <- NA
}

dt <- data.table(kennzahl,
                 bez,
                 titel)

return(dt)
}

#xml.name <- "XML/BJNR001950896.xml" # BGB

#xml.name <- "XML/BJNR335610017.xml" # problem
#f.network.analysis(xml.name)

```

4.10.4 Funktion definieren: f.network.analysis

f.network.analysis benötigt f.kennzahlen.search, f.kennzahlen.collapse und f.kennzahlen.edgelist.

```

f.network.analysis <- function(xml.name,
                               prefix.figuretitle,
                               caption){

#   message(xml.name) # remove when debugging done
  XML <- read_xml(xml.name)

  ## Gliederungseinheiten extrahieren
  gliederungseinheit <- html_elements(XML, xpath = "//norm//gliederungseinheit"
)

  ## Gliederungseinheit splitten
  gliederungseinheit.split <- lapply(gliederungseinheit,
                                     f.split.gliederungseinheit)
  gliederungseinheit.split <- rbindlist(gliederungseinheit.split)

  gliederungseinheit.split <- unique(gliederungseinheit.split, by = "kennzahl")

  if (gliederungseinheit.split[,.N] > 0){

    ## Abkürzung extrahieren
    jurabk <- html_element(XML, xpath = "//norm//jurabk") %>% xml_text()

    if (length(jurabk) == 0){
      jurabk <- "NA"
    }
  }
}

```

```

## Titel als Label priorieren, sonst Bezeichnung einsetzen
node.labels0 <- ifelse(gliederungseinheit.split$titel != "",
                      gliederungseinheit.split$titel,
                      gliederungseinheit.split$bez)

## Rechtsakt als Quelle des Netzwerks einfügen
node.labels <- c(jurabk,
                 node.labels0)

## Edgelist erstellen
edgelist <- tryCatch({f.kennzahlen.edgelist(kennzahl = gliederungseinheit
.split$kennzahl,
                                           name = jurabk)}},
                    error = function(cond) {
                      return(0)}
                    )

# to do: print errorfilename to disk

if (length(edgelist) != 0){

## Node Labels definieren
nodes.df <- gliederungseinheit.split[,.(kennzahl, titel)]

addname <- data.table(jurabk,
                     jurabk)

setnames(addname, new = c("kennzahl",
                          "titel"))

nodes.df <- rbind(addname,
                 nodes.df)

setnames(nodes.df, new = c("kennzahl",
                          "label"))

## Graph aus Edgelist erstellen
g <- graph.data.frame(edgelist,
                      directed = TRUE,
                      vertices = nodes.df)

## Adjazenz-Matrix erstellen
M.adjacency <- as.matrix(get.adjacency(g,
                                       edges = F))

## Dateiname definieren
filename <- paste0(gsub("( +)|(</)",
                      "-",
                      jurabk),
                  "-",
                  gsub("\\\\.xml",
                      "",
                      xml.name))

```

```

## Gliederungstabelle speichern
fwrite(gliederungseinheit.split,
       paste0("netzwerke/Gliederungstabellen/",
              filename,
              "_Gliederungstabelle.csv"))

## Edgelist speichern
fwrite(edgelist,
       paste0("netzwerke/Edgelisten/",
              filename,
              "_Edgelist.csv"))

## Adjazenz-Matrix speichern
fwrite(M.adjacency,
       paste0("netzwerke/Adjazenzmatrizen/",
              filename,
              "_AdjazenzMatrix.csv"))

## GraphML speichern
write_graph(g,
            file = paste0("netzwerke/GraphML/",
                           filename,
                           ".graphml"),
            format = "graphml")

## Diagramm erstellen und speichern
if (length(V(g)) > 1){

  networkplot <- ggraph(g,
                        'dendrogram',
                        circular = TRUE) +
    geom_edge_elbow(colour = "grey") +
    geom_node_text(aes(label = label),
                  size = 2,
                  repel = TRUE)+
  theme_void()+
  labs(
    title = paste(prefix.figuretitle,
                  "| Struktur des",
                  jurabk),
    caption = caption
  )+
  theme(
    plot.title = element_text(size = 50,
                              face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )

  ## may conflict with markdown save
  ggsave(
    filename = paste0("netzwerke/Netzwerkdiagramme/",
                      filename,
                      "_NetzwerkDiagramm.pdf"),

```



```

        plot = networkplot,
        device = "pdf",
        scale = 1,
        width = 50,
        height = 50,
        units = "in",
        dpi = 300,
        limitsize = FALSE
    )
}

}

}

}

f.network.analysis.robust <- function(xml.name,
                                     prefix.figuretitle,
                                     caption){

    tryCatch({f.network.analysis(xml.name,
                                prefix.figuretitle,
                                caption)},
            error = function(cond) {
                return(NA)}
    )

}

```

4.10.5 Netzwerk-Analyse durchführen

```

files.xml <- list.files(pattern = "\\\\.xml$")

errorfiles <- c("BJNR008810961.xml",
               "BJNR010599989.xml",
               "BJNR043410015.xml",
               "BJNR093000015.xml",
               "BJNR135410017.xml",
               "BJNR158720007.xml",
               "BJNR203210978.xml",
               "BJNR203220978.xml",
               "BJNR277700013.xml",
               "BJNR284600017.xml",
               "BJNR364800009.xml",
               "BJNR000939960.xml")

files.xml <- setdiff(files.xml, errorfiles)

length(files.xml)

```

```
## [1] 6647
```

```
#https://www.gesetze-im-internet.de/bgb/BJNR001950896.epub

#xml.name <- files.xml[205]

#xml.name <- "XML/BJNR002089971.xml" # problem
#xml.name <- "XML/BJNR001950896.xml" # BGB
```

4.10.6 Beginn Network Analysis

```
begin.netanalysis <- Sys.time()
```

4.10.7 Parallelisierung definieren

Parallele Berechnung funktioniert nicht mit errorfiles; sequentielle Berechnung schon

```
if(config$parallel$parseNetworks == TRUE){

  plan("multicore",
        workers = fullCores)

}else{

  plan("sequential")

  }

#f.network.analysis.robust(files.xml[35],
#                           prefix.figuretitle = prefix.figuretitle,
#                           caption = caption)
```

4.10.8 XML Parsen

```
out.netanalysis <- future_lapply(files.xml,
                                  f.network.analysis.robust,
                                  prefix.figuretitle = prefix.figuretitle,
                                  caption = caption,
                                  future.seed = TRUE)
```

4.10.9 XML-Dateien bei denen Fehler auftreten

```
files.xml[grep("error",
               out.netanalysis)]
```

```
## character(0)
```

4.10.10 Ende XML Parsing

```
end.netanalysis <- Sys.time()
```

4.10.11 Dauer XML Parsing

```
end.netanalysis - begin.netanalysis
```

```
## Time difference of 3.286686 mins
```

4.11 Wiederverpacken der XML-Dateien

Wiederverpacken der gesammelten XML-Dateien in ein einziges Archiv. Wiederverpacken der Anlagen in ein separates Archiv. Die Roh-Daten werden im Anschluss jeweils gelöscht.

4.11.1 XML-Dateien definieren

```
files.xml <- list.files(pattern = "\\\\.xml")
```

4.11.2 XML-Dateien verpacken

```
zip(paste0(prefix.files,  
           "_DE_XML_Datensatz.zip"),  
    files.xml,  
    mode = "cherry-pick")  
  
unlink(files.xml)
```

4.11.3 Anhänge zu XML-Dateien verpacken

```
attachments <- list.files(pattern = "(\\.jpg)|\\.gif)|\\.pdf)|\\.png)",  
                           ignore.case = TRUE)  
  
if (length(attachments) > 0){  
  zip(paste0(prefix.files,
```

```
        "_DE_XML_Anlagen.zip"),
    attachments,
    mode = "cherry-pick")

}

unlink(attachments)
```

5 Frequenztabellen erstellen: Einzelnormen

5.1 Funktion anzeigen: f.fast.freqtable

```
print(f.fast.freqtable)
```

```
## function(x,
##           varlist = names(x),
##           sumrow = TRUE,
##           output.list = TRUE,
##           output.kable = FALSE,
##           output.csv = FALSE,
##           outputdir = "./",
##           prefix = "",
##           align = "r"){
##
##   ## Begin List
##   freqtable.list <- vector("list", length(varlist))
##
##   ## Calculate Frequency Table
##   for (i in seq_along(varlist)){
##
##     varname <- varlist[i]
##
##     freqtable <- x[, .N, keyby=c(paste0(varname))]
##
##     freqtable[, c("exactpercent",
##                  "roundedpercent",
##                  "cumulpercent") := {
##       exactpercent <- N/sum(N)*100
##       roundedpercent <- round(exactpercent, 2)
##       cumulpercent <- round(cumsum(exactpercent), 2)
##       list(exactpercent,
##            roundedpercent,
##            cumulpercent)}]
##
##     ## Calculate Summary Row
##     if (sumrow == TRUE){
##       colsums <- cbind("Total",
##                       freqtable[, lapply(.SD, function(x){round(sum(x)
##
##                                     .SDcols = c("N",
##                                                  "exactpercent",
##                                                  "roundedpercent")
##                                     ], round(max(freqtable$cumulpercent)))
##
##       colnames(colsums)[c(1,5)] <- c(varname, "cumulpercent")
##       freqtable <- rbind(freqtable, colsums)
##     }
##
##     ## Add Frequency Table to List
##     freqtable.list[[i]] <- freqtable
```

```

##
##      ## Write CSV
##      if (output.csv == TRUE){
##
##          fwrite(freqtable,
##                file.path(outputdir,
##                          paste0(prefix,
##                                varname,
##                                ".csv")),
##                na = "NA")
##
##      }
##
##      ## Output Kable
##      if (output.kable == TRUE){
##
##          cat("\n-----\n")
##          cat(paste0("Frequency Table for Variable:  ", varname, "\n"))
##          cat("-----\n")
##          cat(paste0("\n ",
##                    x[, .N, keyby=c(paste0(varname))][, .N],
##                    " unique value(s) detected.\n\n"))
##
##          print(kable(freqtable,
##                    format = "latex",
##                    align = align,
##                    booktabs = TRUE,
##                    longtable = TRUE) %>% kable_styling(latex_options = "
repeat_header"))
##      }
##
##      ## Return List of Frequency Tables
##      if (output.list == TRUE){
##          return(freqtable.list)
##      }
## }

```

5.2 Liste zu prüfender Variablen

```
print(config$freqtable$ignore)
```

```

## [1] "periodikum"      "fundstellentyp"  "check_neuf"
## [4] "check_aufh"      "check_sonst"     "check_hinweis"
## [7] "check_stand"     "gliederungskennzahl" "ausfertigung_jahr"
## [10] "doi_concept"     "doi_version"     "version"
## [13] "lizenz"

```

5.3 Frequenztabellen erstellen

```
prefix.freqtable.einzelnormen <- paste0(config$project$shortname,
                                         "_01_Einzelnormen_Frequenztafel_var-")
```

```
f.fast.freqtable(dt.normen,
                 varlist = config$freqtable$ignore,
                 sumrow = TRUE,
                 output.list = FALSE,
                 output.kable = TRUE,
                 output.csv = TRUE,
                 outputdir = dir.analysis,
                 prefix = prefix.freqtable.einzelnormen)
```

Frequency Table for Variable: periodikum

21 unique value(s) detected.

periodikum	N	exactpercent	roundedpercent	cumulpercent
BAGVBl	1	0.0009867	0.00	0.00
BAnz	1230	1.2136761	1.21	1.21
BGBI	2305	2.2744092	2.27	3.49
BGBI I	85243	84.1116977	84.11	87.60
BGBI II	3253	3.2098278	3.21	90.81
GBl DDR	233	0.2299077	0.23	91.04
GBl DDR I	616	0.6078248	0.61	91.65
GBl DDR II	43	0.0424293	0.04	91.69
GVBl BE	9	0.0088806	0.01	91.70
NV	15	0.0148009	0.01	91.71
RAnz	38	0.0374957	0.04	91.75
RGBI	5718	5.6421136	5.64	97.39
RGBI I	1654	1.6320489	1.63	99.03
RGBI II	350	0.3453550	0.35	99.37
RMBI	229	0.2259608	0.23	99.60
VOBl BrZ	19	0.0187478	0.02	99.62

(continued)

periodikum	N	exactpercent	roundedpercent	cumulpercent
VkBl	111	0.1095269	0.11	99.73
WiGBI	122	0.1203809	0.12	99.85
ZBl	63	0.0621639	0.06	99.91
eBAanz	52	0.0513099	0.05	99.96
Öff Anz	41	0.0404559	0.04	100.00
Total	101345	100.0000000	100.00	100.00

Frequency Table for Variable: fundstellentyp

1 unique value(s) detected.

fundstellentyp	N	exactpercent	roundedpercent	cumulpercent
amtlich	101345	100	100	100
Total	101345	100	100	100

Frequency Table for Variable: check_neuf

2 unique value(s) detected.

check_neuf	N	exactpercent	roundedpercent	cumulpercent
NA	77143	76.1192	76.12	76.12
ja	24202	23.8808	23.88	100.00
Total	101345	100.0000	100.00	100.00

Frequency Table for Variable: check_aufh

2 unique value(s) detected.

check_aufh	N	exactpercent	roundedpercent	cumulpercent
NA	99698	98.374858	98.37	98.37
ja	1647	1.625142	1.63	100.00
Total	101345	100.000000	100.00	100.00

Frequency Table for Variable: check_sonst

2 unique value(s) detected.

check_sonst	N	exactpercent	roundedpercent	cumulpercent
NA	81171	80.09374	80.09	80.09
ja	20174	19.90626	19.91	100.00
Total	101345	100.00000	100.00	100.00

Frequency Table for Variable: check_hinweis

2 unique value(s) detected.

check_hinweis	N	exactpercent	roundedpercent	cumulpercent
NA	78751	77.70586	77.71	77.71
ja	22594	22.29414	22.29	100.00
Total	101345	100.00000	100.00	100.00

Frequency Table for Variable: check_stand

2 unique value(s) detected.

check_stand	N	exactpercent	roundedpercent	cumulpercent
NA	22272	21.97642	21.98	21.98
ja	79073	78.02358	78.02	100.00
Total	101345	100.00000	100.00	100.00

Frequency Table for Variable: gliederungskennzahl

1478 unique value(s) detected.

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
000	77	0.0759781	0.08	0.08
000010	12	0.0118407	0.01	0.09
000020	18	0.0177611	0.02	0.11
000030	17	0.0167744	0.02	0.12
000040	17	0.0167744	0.02	0.14
000050	18	0.0177611	0.02	0.16
000060	10	0.0098673	0.01	0.17
000070	3	0.0029602	0.00	0.17
000080	2	0.0019735	0.00	0.17
000090	1	0.0009867	0.00	0.17
000100	2	0.0019735	0.00	0.17
000110	3	0.0029602	0.00	0.18
000120	5	0.0049336	0.00	0.18
001000	2	0.0019735	0.00	0.18
001010	1	0.0009867	0.00	0.19
001020	3	0.0029602	0.00	0.19
001030	1	0.0009867	0.00	0.19
001040	1	0.0009867	0.00	0.19
001052	1	0.0009867	0.00	0.19
001053	1	0.0009867	0.00	0.19
001060	2	0.0019735	0.00	0.19
001080	1	0.0009867	0.00	0.20
001090	3	0.0029602	0.00	0.20
001100	1	0.0009867	0.00	0.20
001101	1	0.0009867	0.00	0.20
001110	1	0.0009867	0.00	0.20

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
001120	5	0.0049336	0.00	0.21
010	6244	6.1611328	6.16	6.37
010010	928	0.9156840	0.92	7.28
010010000010	1	0.0009867	0.00	7.28
010010000020	1	0.0009867	0.00	7.29
010010010	118	0.1164340	0.12	7.40
010010020	121	0.1193941	0.12	7.52
010010020010	3	0.0029602	0.00	7.52
010010020010010	37	0.0365090	0.04	7.56
010010020010020	24	0.0236815	0.02	7.58
010010020020	12	0.0118407	0.01	7.60
010010020030	1	0.0009867	0.00	7.60
010010030	28	0.0276284	0.03	7.62
010010030010	7	0.0069071	0.01	7.63
010010030020	16	0.0157877	0.02	7.65
010010030030	12	0.0118407	0.01	7.66
010010030040	12	0.0118407	0.01	7.67
010010040	35	0.0345355	0.03	7.71
010010050	11	0.0108540	0.01	7.72
010010060	6	0.0059204	0.01	7.72
010010070010	10	0.0098673	0.01	7.73
010010070020	5	0.0049336	0.00	7.74
010020	800	0.7893828	0.79	8.53
010020010	151	0.1489960	0.15	8.68
010020010000030	1	0.0009867	0.00	8.68
010020010000040	1	0.0009867	0.00	8.68
010020010000050	1	0.0009867	0.00	8.68
010020010000060	1	0.0009867	0.00	8.68
010020020	138	0.1361685	0.14	8.82

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010020020000070	1	0.0009867	0.00	8.82
010020020000080	1	0.0009867	0.00	8.82
010020020000090	1	0.0009867	0.00	8.82
010020020000100	1	0.0009867	0.00	8.82
010020020000110	1	0.0009867	0.00	8.82
010020020000120	1	0.0009867	0.00	8.82
010020020010	16	0.0157877	0.02	8.84
010020020020	19	0.0187478	0.02	8.86
010020020030	9	0.0088806	0.01	8.86
010020020040	14	0.0138142	0.01	8.88
010020020050	18	0.0177611	0.02	8.90
010020030	108	0.1065667	0.11	9.00
010020030000130	1	0.0009867	0.00	9.00
010020030000140	1	0.0009867	0.00	9.00
010020030000170	1	0.0009867	0.00	9.01
010020030000171	1	0.0009867	0.00	9.01
010020030000172	1	0.0009867	0.00	9.01
010020030000190	1	0.0009867	0.00	9.01
010020030000200	1	0.0009867	0.00	9.01
010020030000210	1	0.0009867	0.00	9.01
010020030000220	1	0.0009867	0.00	9.01
010020030000230	1	0.0009867	0.00	9.01
010020030000240	1	0.0009867	0.00	9.01
010020030010	3	0.0029602	0.00	9.02
010020030020	2	0.0019735	0.00	9.02
010020030030	2	0.0019735	0.00	9.02
010020040	75	0.0740046	0.07	9.09
010020040000250	1	0.0009867	0.00	9.10
010020040000260	1	0.0009867	0.00	9.10

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010020050	58	0.0572303	0.06	9.15
010020050020000380	1	0.0009867	0.00	9.15
010020050020000390	1	0.0009867	0.00	9.16
010020050020000400	1	0.0009867	0.00	9.16
010020050020000410	1	0.0009867	0.00	9.16
010020050020000420	1	0.0009867	0.00	9.16
010020060	44	0.0434161	0.04	9.20
010020060000430	1	0.0009867	0.00	9.20
010020060000440	1	0.0009867	0.00	9.20
010020060000450	1	0.0009867	0.00	9.21
010020060000460	1	0.0009867	0.00	9.21
010020070	31	0.0305886	0.03	9.24
010020070010000461	1	0.0009867	0.00	9.24
010020070020000462	1	0.0009867	0.00	9.24
010020070020000463	1	0.0009867	0.00	9.24
010020070030000464	1	0.0009867	0.00	9.24
010020070040000465	1	0.0009867	0.00	9.24
010020080	41	0.0404559	0.04	9.28
010020090	1	0.0009867	0.00	9.28
010030	532	0.5249396	0.52	9.81
010030000470	1	0.0009867	0.00	9.81
010030000480	1	0.0009867	0.00	9.81
010030010	133	0.1312349	0.13	9.94
010030010010	8	0.0078938	0.01	9.95
010030010020	14	0.0138142	0.01	9.96
010030010030	20	0.0197346	0.02	9.98
010030010040	12	0.0118407	0.01	9.99
010030010050	4	0.0039469	0.00	10.00
010030010060	8	0.0078938	0.01	10.01

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010030010070	8	0.0078938	0.01	10.01
010030010080	4	0.0039469	0.00	10.02
010030010380	2	0.0019735	0.00	10.02
010030010400	4	0.0039469	0.00	10.02
010030010440	1	0.0009867	0.00	10.03
010030010450	3	0.0029602	0.00	10.03
010030020	84	0.0828852	0.08	10.11
010030020010	86	0.0848587	0.08	10.20
010030020020	41	0.0404559	0.04	10.24
010030020030	12	0.0118407	0.01	10.25
010030020040	12	0.0118407	0.01	10.26
010030020050	9	0.0088806	0.01	10.27
010030020060	18	0.0177611	0.02	10.29
010030030	70	0.0690710	0.07	10.36
010030030010	9	0.0088806	0.01	10.36
010030030020	9	0.0088806	0.01	10.37
010030030030	6	0.0059204	0.01	10.38
010030030040	1	0.0009867	0.00	10.38
010030040	39	0.0384824	0.04	10.42
010030050	39	0.0384824	0.04	10.46
010030060	6	0.0059204	0.01	10.46
010030060010	16	0.0157877	0.02	10.48
010030060020	8	0.0078938	0.01	10.49
010030060030	3	0.0029602	0.00	10.49
010030060040	3	0.0029602	0.00	10.49
010030060050	2	0.0019735	0.00	10.49
010030070	17	0.0167744	0.02	10.51
010031	2	0.0019735	0.00	10.51
010040	343	0.3384479	0.34	10.85

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010040010	83	0.0818985	0.08	10.93
010040020	69	0.0680843	0.07	11.00
010040020010	3	0.0029602	0.00	11.00
010040020020	5	0.0049336	0.00	11.01
010040020030	2	0.0019735	0.00	11.01
010040030	42	0.0414426	0.04	11.05
010040030010	8	0.0078938	0.01	11.06
010040030020	3	0.0029602	0.00	11.06
010040040	21	0.0207213	0.02	11.08
010040040010	8	0.0078938	0.01	11.09
010040040020	11	0.0108540	0.01	11.10
010040040030	5	0.0049336	0.00	11.11
010040040040	9	0.0088806	0.01	11.12
010040040050	1	0.0009867	0.00	11.12
010040040060	3	0.0029602	0.00	11.12
010040040070	8	0.0078938	0.01	11.13
010040050	2	0.0019735	0.00	11.13
010040120010	1	0.0009867	0.00	11.13
010040120020	1	0.0009867	0.00	11.13
010040120030	1	0.0009867	0.00	11.13
010040120040	1	0.0009867	0.00	11.14
010040120050	1	0.0009867	0.00	11.14
010040120060	1	0.0009867	0.00	11.14
010041	2	0.0019735	0.00	11.14
010050	163	0.1608367	0.16	11.30
010050010	61	0.0601904	0.06	11.36
010050020	86	0.0848587	0.08	11.45
010050020020	2	0.0019735	0.00	11.45
010050030	43	0.0424293	0.04	11.49

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010050030010	2	0.0019735	0.00	11.49
010050030020	1	0.0009867	0.00	11.49
010050030030	2	0.0019735	0.00	11.49
010050040	2	0.0019735	0.00	11.50
010060	184	0.1815580	0.18	11.68
010060010	48	0.0473630	0.05	11.73
010060020	73	0.0720312	0.07	11.80
010060020010	11	0.0108540	0.01	11.81
010060020020	10	0.0098673	0.01	11.82
010060020030	5	0.0049336	0.00	11.82
010060020040	14	0.0138142	0.01	11.84
010060020050	1	0.0009867	0.00	11.84
010060030	57	0.0562435	0.06	11.89
010060030810	7	0.0069071	0.01	11.90
010060030820	8	0.0078938	0.01	11.91
010060030830	3	0.0029602	0.00	11.91
010060030840	1	0.0009867	0.00	11.91
010060040	14	0.0138142	0.01	11.93
010060050	33	0.0325620	0.03	11.96
010060051	4	0.0039469	0.00	11.96
010060060	23	0.0226948	0.02	11.99
010060070	5	0.0049336	0.00	11.99
010070	225	0.2220139	0.22	12.21
010070010	8	0.0078938	0.01	12.22
010070010010	11	0.0108540	0.01	12.23
010070010020	6	0.0059204	0.01	12.24
010070020	14	0.0138142	0.01	12.25
010070030	14	0.0138142	0.01	12.27
010080	134	0.1322216	0.13	12.40

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
010080010	9	0.0088806	0.01	12.41
010080010010	2	0.0019735	0.00	12.41
010080010020	11	0.0108540	0.01	12.42
010080020	17	0.0167744	0.02	12.44
010080030	8	0.0078938	0.01	12.44
010090	62	0.0611772	0.06	12.50
010090010	23	0.0226948	0.02	12.53
010090020	19	0.0187478	0.02	12.55
010090030	6	0.0059204	0.01	12.55
010091	5	0.0049336	0.00	12.56
010092	1	0.0009867	0.00	12.56
010100	29	0.0286151	0.03	12.59
010100010	2	0.0019735	0.00	12.59
010100020	1	0.0009867	0.00	12.59
010110	37	0.0365090	0.04	12.63
010110010	1	0.0009867	0.00	12.63
010110020	1	0.0009867	0.00	12.63
010110030	1	0.0009867	0.00	12.63
010120	4	0.0039469	0.00	12.63
010130	9	0.0088806	0.01	12.64
010140	5	0.0049336	0.00	12.65
011	11	0.0108540	0.01	12.66
012	2	0.0019735	0.00	12.66
014	2	0.0019735	0.00	12.66
020	6438	6.3525581	6.35	19.01
020000	1	0.0009867	0.00	19.02
020000010	2	0.0019735	0.00	19.02
020000020	8	0.0078938	0.01	19.03
020000030	8	0.0078938	0.01	19.03

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020000040	8	0.0078938	0.01	19.04
020000050	8	0.0078938	0.01	19.05
020000060	8	0.0078938	0.01	19.06
020000500	1	0.0009867	0.00	19.06
020000510	1	0.0009867	0.00	19.06
020000520	1	0.0009867	0.00	19.06
020000530	1	0.0009867	0.00	19.06
020000531	1	0.0009867	0.00	19.06
020010	1654	1.6320489	1.63	20.69
020010010	295	0.2910849	0.29	20.98
020010020	230	0.2269476	0.23	21.21
020010020010	25	0.0246682	0.02	21.24
020010020020	8	0.0078938	0.01	21.24
020010020020010	4	0.0039469	0.00	21.25
020010020020020	3	0.0029602	0.00	21.25
020010020020030	3	0.0029602	0.00	21.25
020010020020040	3	0.0029602	0.00	21.26
020010020020050	1	0.0009867	0.00	21.26
020010020030	10	0.0098673	0.01	21.27
020010020040	6	0.0059204	0.01	21.27
020010020050	5	0.0049336	0.00	21.28
020010020060	4	0.0039469	0.00	21.28
020010030	144	0.1420889	0.14	21.42
020010030010	11	0.0108540	0.01	21.44
020010030020	7	0.0069071	0.01	21.44
020010030030010	4	0.0039469	0.00	21.45
020010030030020010	5	0.0049336	0.00	21.45
020010030030020020	8	0.0078938	0.01	21.46
020010030040	4	0.0039469	0.00	21.46

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020010040	91	0.0897923	0.09	21.55
020010040010	5	0.0049336	0.00	21.56
020010040020	3	0.0029602	0.00	21.56
020010040030	6	0.0059204	0.01	21.57
020010041	5	0.0049336	0.00	21.57
020010050	93	0.0917658	0.09	21.66
020010060	32	0.0315753	0.03	21.70
020010070	38	0.0374957	0.04	21.73
020010070010	4	0.0039469	0.00	21.74
020010070020	7	0.0069071	0.01	21.74
020010070030	7	0.0069071	0.01	21.75
020010080	19	0.0187478	0.02	21.77
020010090	31	0.0305886	0.03	21.80
020010100	11	0.0108540	0.01	21.81
020010110	7	0.0069071	0.01	21.82
020010120	9	0.0088806	0.01	21.83
020011	3	0.0029602	0.00	21.83
020011042	1	0.0009867	0.00	21.83
020020	1564	1.5432434	1.54	23.37
020020010	160	0.1578766	0.16	23.53
020020010010	44	0.0434161	0.04	23.57
020020010010010	2	0.0019735	0.00	23.58
020020010010020	1	0.0009867	0.00	23.58
020020010010030	3	0.0029602	0.00	23.58
020020010010040	1	0.0009867	0.00	23.58
020020010010050	3	0.0029602	0.00	23.58
020020010020	12	0.0118407	0.01	23.60
020020010030	11	0.0108540	0.01	23.61
020020010040	15	0.0148009	0.01	23.62

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020020010050	3	0.0029602	0.00	23.63
020020010060	11	0.0108540	0.01	23.64
020020010070	6	0.0059204	0.01	23.64
020020010080	4	0.0039469	0.00	23.65
020020020	237	0.2338547	0.23	23.88
020020020010	33	0.0325620	0.03	23.91
020020020020	29	0.0286151	0.03	23.94
020020020030	25	0.0246682	0.02	23.97
020020020040	9	0.0088806	0.01	23.97
020020020050	19	0.0187478	0.02	23.99
020020020060	1	0.0009867	0.00	23.99
020020020070010	3	0.0029602	0.00	24.00
020020020070020	7	0.0069071	0.01	24.00
020020020070030	2	0.0019735	0.00	24.01
020020030	109	0.1075534	0.11	24.11
020020030010	19	0.0187478	0.02	24.13
020020030020	23	0.0226948	0.02	24.16
020020030030	19	0.0187478	0.02	24.17
020020030040	17	0.0167744	0.02	24.19
020020030050	5	0.0049336	0.00	24.20
020020030060	2	0.0019735	0.00	24.20
020020030070	16	0.0157877	0.02	24.21
020020030080	4	0.0039469	0.00	24.22
020020030090	3	0.0029602	0.00	24.22
020020030100	1	0.0009867	0.00	24.22
020020040	71	0.0700577	0.07	24.29
020020043	1	0.0009867	0.00	24.29
020020050	26	0.0256549	0.03	24.32
020020050010	24	0.0236815	0.02	24.34

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020020050020	9	0.0088806	0.01	24.35
020020060	50	0.0493364	0.05	24.40
020020060010	1	0.0009867	0.00	24.40
020020060020	6	0.0059204	0.01	24.41
020020060030	5	0.0049336	0.00	24.41
020020070	12	0.0118407	0.01	24.42
020020080	2	0.0019735	0.00	24.43
020020080010	6	0.0059204	0.01	24.43
020020080020	11	0.0108540	0.01	24.44
020020080030	4	0.0039469	0.00	24.45
020020080040	2	0.0019735	0.00	24.45
020020090	4	0.0039469	0.00	24.45
020020100	13	0.0128275	0.01	24.46
020021	23	0.0226948	0.02	24.49
020030	1113	1.0982288	1.10	25.59
020030010	161	0.1588633	0.16	25.74
020030010010	11	0.0108540	0.01	25.76
020030010020	5	0.0049336	0.00	25.76
020030010020010	2	0.0019735	0.00	25.76
020030010020020	7	0.0069071	0.01	25.77
020030010020030	2	0.0019735	0.00	25.77
020030010020040	3	0.0029602	0.00	25.77
020030010030	8	0.0078938	0.01	25.78
020030010040	6	0.0059204	0.01	25.79
020030020	214	0.2111599	0.21	26.00
020030020010	32	0.0315753	0.03	26.03
020030020020	9	0.0088806	0.01	26.04
020030021010	20	0.0197346	0.02	26.06
020030021020	2	0.0019735	0.00	26.06

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020030030	109	0.1075534	0.11	26.17
020030040	50	0.0493364	0.05	26.22
020030040010	6	0.0059204	0.01	26.22
020030040020	2	0.0019735	0.00	26.23
020030040030	5	0.0049336	0.00	26.23
020030040040	1	0.0009867	0.00	26.23
020030050	39	0.0384824	0.04	26.27
020030050010	9	0.0088806	0.01	26.28
020030050020	17	0.0167744	0.02	26.30
020030060	10	0.0098673	0.01	26.31
020030070	7	0.0069071	0.01	26.31
020030080	9	0.0088806	0.01	26.32
020030090	3	0.0029602	0.00	26.32
020030100	1	0.0009867	0.00	26.33
020030100010	1	0.0009867	0.00	26.33
020030100020	11	0.0108540	0.01	26.34
020030100030	1	0.0009867	0.00	26.34
020030110	3	0.0029602	0.00	26.34
020030120	2	0.0019735	0.00	26.34
020030130	7	0.0069071	0.01	26.35
020030140	2	0.0019735	0.00	26.35
020030150	8	0.0078938	0.01	26.36
020031	18	0.0177611	0.02	26.38
020040	759	0.7489269	0.75	27.13
020040010	66	0.0651241	0.07	27.19
020040020	64	0.0631506	0.06	27.26
020040030	38	0.0374957	0.04	27.29
020040030010	1	0.0009867	0.00	27.29
020040030020	6	0.0059204	0.01	27.30

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020040040	22	0.0217080	0.02	27.32
020040050	5	0.0049336	0.00	27.33
020040060	8	0.0078938	0.01	27.33
020040070	1	0.0009867	0.00	27.34
020040160	1	0.0009867	0.00	27.34
020041	6	0.0059204	0.01	27.34
020050	424	0.4183729	0.42	27.76
020050010	41	0.0404559	0.04	27.80
020050010010	5	0.0049336	0.00	27.81
020050010020	4	0.0039469	0.00	27.81
020050010030	3	0.0029602	0.00	27.81
020050020	39	0.0384824	0.04	27.85
020050030	5	0.0049336	0.00	27.86
020050030010	13	0.0128275	0.01	27.87
020050030020	13	0.0128275	0.01	27.88
020050030030	6	0.0059204	0.01	27.89
020050030040	2	0.0019735	0.00	27.89
020050030050	12	0.0118407	0.01	27.90
020050040	7	0.0069071	0.01	27.91
020050050	1	0.0009867	0.00	27.91
020050180	1	0.0009867	0.00	27.91
020050190	1	0.0009867	0.00	27.91
020050200	1	0.0009867	0.00	27.91
020050210	1	0.0009867	0.00	27.91
020051	7	0.0069071	0.01	27.92
020051211	1	0.0009867	0.00	27.92
020051212	1	0.0009867	0.00	27.92
020051213	1	0.0009867	0.00	27.92
020052	1	0.0009867	0.00	27.92

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020053	1	0.0009867	0.00	27.93
020060	309	0.3048991	0.30	28.23
020060010	31	0.0305886	0.03	28.26
020060020	18	0.0177611	0.02	28.28
020060020010	11	0.0108540	0.01	28.29
020060020020010	2	0.0019735	0.00	28.29
020060020020020	2	0.0019735	0.00	28.29
020060020020030	1	0.0009867	0.00	28.29
020060020020040	1	0.0009867	0.00	28.30
020060020020050	1	0.0009867	0.00	28.30
020060030	10	0.0098673	0.01	28.31
020060030010	2	0.0019735	0.00	28.31
020060030020	2	0.0019735	0.00	28.31
020060030030	12	0.0118407	0.01	28.32
020060030040	6	0.0059204	0.01	28.33
020060030050	1	0.0009867	0.00	28.33
020060040	9	0.0088806	0.01	28.34
020060050	4	0.0039469	0.00	28.34
020060050010	2	0.0019735	0.00	28.34
020060050020	14	0.0138142	0.01	28.36
020060060	6	0.0059204	0.01	28.36
020060070	9	0.0088806	0.01	28.37
020060080010	5	0.0049336	0.00	28.38
020060080020	14	0.0138142	0.01	28.39
020060090	11	0.0108540	0.01	28.40
020060100	1	0.0009867	0.00	28.40
020060110	3	0.0029602	0.00	28.41
020060220	1	0.0009867	0.00	28.41
020060230	2	0.0019735	0.00	28.41

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020061	3	0.0029602	0.00	28.41
020070	189	0.1864917	0.19	28.60
020070010	17	0.0167744	0.02	28.62
020070020	5	0.0049336	0.00	28.62
020070020010	10	0.0098673	0.01	28.63
020070020020	23	0.0226948	0.02	28.65
020070020020010	1	0.0009867	0.00	28.65
020070020020020	1	0.0009867	0.00	28.65
020070020030010	15	0.0148009	0.01	28.67
020070020030020010	3	0.0029602	0.00	28.67
020070020030020020	6	0.0059204	0.01	28.68
020070020030020030	1	0.0009867	0.00	28.68
020070020030020040	2	0.0019735	0.00	28.68
020070020030020050	1	0.0009867	0.00	28.68
020070020030030	1	0.0009867	0.00	28.68
020070020040	1	0.0009867	0.00	28.68
020070020050	2	0.0019735	0.00	28.69
020070020060	5	0.0049336	0.00	28.69
020070020069	1	0.0009867	0.00	28.69
020070020070	3	0.0029602	0.00	28.70
020070020080	1	0.0009867	0.00	28.70
020070020090	1	0.0009867	0.00	28.70
020070030	4	0.0039469	0.00	28.70
020070030010	10	0.0098673	0.01	28.71
020070030020	6	0.0059204	0.01	28.72
020070030030	14	0.0138142	0.01	28.73
020070030040	3	0.0029602	0.00	28.73
020070030050	1	0.0009867	0.00	28.73
020070040	5	0.0049336	0.00	28.74

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020070240	1	0.0009867	0.00	28.74
020080	139	0.1371553	0.14	28.88
020080010	5	0.0049336	0.00	28.88
020080010010	24	0.0236815	0.02	28.91
020080010020010	2	0.0019735	0.00	28.91
020080010020020	7	0.0069071	0.01	28.92
020080010020030	11	0.0108540	0.01	28.93
020080010030	11	0.0108540	0.01	28.94
020080010040	1	0.0009867	0.00	28.94
020080020	18	0.0177611	0.02	28.96
020080030	3	0.0029602	0.00	28.96
020080030010010	3	0.0029602	0.00	28.96
020080030010020	24	0.0236815	0.02	28.99
020080030020	3	0.0029602	0.00	28.99
020080030030	1	0.0009867	0.00	28.99
020080030040	1	0.0009867	0.00	28.99
020080030050	2	0.0019735	0.00	28.99
020080030060	2	0.0019735	0.00	28.99
020080040	22	0.0217080	0.02	29.02
020080050	1	0.0009867	0.00	29.02
020080050010	20	0.0197346	0.02	29.04
020080050020010	7	0.0069071	0.01	29.04
020080050020011	6	0.0059204	0.01	29.05
020080050020020010	4	0.0039469	0.00	29.05
020080050020020011	4	0.0039469	0.00	29.06
020080050020020020	16	0.0157877	0.02	29.07
020080050020030	5	0.0049336	0.00	29.08
020080050020040	14	0.0138142	0.01	29.09
020080050020050010	5	0.0049336	0.00	29.10

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020080050020050020	9	0.0088806	0.01	29.11
020080050020050030	2	0.0019735	0.00	29.11
020080050020050040	3	0.0029602	0.00	29.11
020080050020060	2	0.0019735	0.00	29.11
020080050030	6	0.0059204	0.01	29.12
020080050040	8	0.0078938	0.01	29.13
020080050050	31	0.0305886	0.03	29.16
020080060	10	0.0098673	0.01	29.17
020080070	7	0.0069071	0.01	29.17
020080080	3	0.0029602	0.00	29.18
020080080010	24	0.0236815	0.02	29.20
020080080020	8	0.0078938	0.01	29.21
020080090010010	25	0.0246682	0.02	29.23
020080090010020	8	0.0078938	0.01	29.24
020080090010030	6	0.0059204	0.01	29.25
020080090010040	1	0.0009867	0.00	29.25
020080090020	5	0.0049336	0.00	29.25
020080090030	2	0.0019735	0.00	29.25
020080090040	25	0.0246682	0.02	29.28
020080100010	4	0.0039469	0.00	29.28
020080100020	5	0.0049336	0.00	29.29
020080100030	1	0.0009867	0.00	29.29
020080100040	4	0.0039469	0.00	29.29
020080110	6	0.0059204	0.01	29.30
020080120010	13	0.0128275	0.01	29.31
020080120020	3	0.0029602	0.00	29.31
020080120030010	3	0.0029602	0.00	29.32
020080120030020	4	0.0039469	0.00	29.32
020080120030030010	4	0.0039469	0.00	29.33

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020080120030030020	7	0.0069071	0.01	29.33
020080120030030030	10	0.0098673	0.01	29.34
020080130	11	0.0108540	0.01	29.35
020080140	13	0.0128275	0.01	29.37
020080150	5	0.0049336	0.00	29.37
020080160	37	0.0365090	0.04	29.41
020080170	18	0.0177611	0.02	29.43
020080180	3	0.0029602	0.00	29.43
020080190	3	0.0029602	0.00	29.43
020080200	14	0.0138142	0.01	29.44
020080210	1	0.0009867	0.00	29.45
020080220	3	0.0029602	0.00	29.45
020080230	10	0.0098673	0.01	29.46
020080240	16	0.0157877	0.02	29.47
020080250	4	0.0039469	0.00	29.48
020080260	11	0.0108540	0.01	29.49
020080270	32	0.0315753	0.03	29.52
020090	107	0.1055799	0.11	29.63
020090010	15	0.0148009	0.01	29.64
020090020	3	0.0029602	0.00	29.64
020090030	12	0.0118407	0.01	29.66
020100	80	0.0789383	0.08	29.74
020100010	8	0.0078938	0.01	29.74
020100020	7	0.0069071	0.01	29.75
020100030	3	0.0029602	0.00	29.75
020100040	2	0.0019735	0.00	29.75
020110	106	0.1045932	0.10	29.86
020110010	1	0.0009867	0.00	29.86
020110020	2	0.0019735	0.00	29.86

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
020110030	2	0.0019735	0.00	29.86
020110040	2	0.0019735	0.00	29.87
020110050	3	0.0029602	0.00	29.87
020120	51	0.0503232	0.05	29.92
020120010	1	0.0009867	0.00	29.92
020120020	3	0.0029602	0.00	29.92
020120030	3	0.0029602	0.00	29.93
020130	73	0.0720312	0.07	30.00
020130010	1	0.0009867	0.00	30.00
020130020	1	0.0009867	0.00	30.00
020130030	7	0.0069071	0.01	30.01
020130040	4	0.0039469	0.00	30.01
020140	95	0.0937392	0.09	30.11
020150	42	0.0414426	0.04	30.15
020160	52	0.0513099	0.05	30.20
020170	39	0.0384824	0.04	30.24
020180	48	0.0473630	0.05	30.28
020190	40	0.0394691	0.04	30.32
020200	37	0.0365090	0.04	30.36
020210	37	0.0365090	0.04	30.40
020220	42	0.0414426	0.04	30.44
020230	46	0.0453895	0.05	30.48
020240	34	0.0335488	0.03	30.52
020250	43	0.0424293	0.04	30.56
020260	35	0.0345355	0.03	30.59
020270	36	0.0355222	0.04	30.63
020280	37	0.0365090	0.04	30.67
020290	13	0.0128275	0.01	30.68
020300	25	0.0246682	0.02	30.70

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
021	32	0.0315753	0.03	30.73
021010	5	0.0049336	0.00	30.74
021020	2	0.0019735	0.00	30.74
021030	4	0.0039469	0.00	30.75
022	3	0.0029602	0.00	30.75
030	5427	5.3549756	5.35	36.10
030000550	1	0.0009867	0.00	36.10
030000560	1	0.0009867	0.00	36.11
030000590	1	0.0009867	0.00	36.11
030000600	1	0.0009867	0.00	36.11
030000610	1	0.0009867	0.00	36.11
030000620	1	0.0009867	0.00	36.11
030000630	1	0.0009867	0.00	36.11
030000640	1	0.0009867	0.00	36.11
030000660	1	0.0009867	0.00	36.11
030000670	1	0.0009867	0.00	36.11
030000680	1	0.0009867	0.00	36.11
030000690	1	0.0009867	0.00	36.12
030000730	1	0.0009867	0.00	36.12
030000740	1	0.0009867	0.00	36.12
030000760	1	0.0009867	0.00	36.12
030000770	1	0.0009867	0.00	36.12
030000780	1	0.0009867	0.00	36.12
030000790	1	0.0009867	0.00	36.12
030000800	1	0.0009867	0.00	36.12
030000810	1	0.0009867	0.00	36.12
030000820	1	0.0009867	0.00	36.12
030000830	1	0.0009867	0.00	36.13
030000840	1	0.0009867	0.00	36.13

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030000850	1	0.0009867	0.00	36.13
030000860	1	0.0009867	0.00	36.13
030000890	1	0.0009867	0.00	36.13
030000900	1	0.0009867	0.00	36.13
030000910	1	0.0009867	0.00	36.13
030000930	1	0.0009867	0.00	36.13
030000940	1	0.0009867	0.00	36.13
030000960	1	0.0009867	0.00	36.13
030000970	1	0.0009867	0.00	36.13
030000980	1	0.0009867	0.00	36.14
030000990	1	0.0009867	0.00	36.14
030001000	1	0.0009867	0.00	36.14
030001010	1	0.0009867	0.00	36.14
030001020	1	0.0009867	0.00	36.14
030001040	1	0.0009867	0.00	36.14
030001050	1	0.0009867	0.00	36.14
030001060	1	0.0009867	0.00	36.14
030001070	1	0.0009867	0.00	36.14
030001080	1	0.0009867	0.00	36.14
030001090	1	0.0009867	0.00	36.15
030001100	1	0.0009867	0.00	36.15
030001110	1	0.0009867	0.00	36.15
030001120	1	0.0009867	0.00	36.15
030001130	1	0.0009867	0.00	36.15
030001140	1	0.0009867	0.00	36.15
030001150	1	0.0009867	0.00	36.15
030001160	1	0.0009867	0.00	36.15
030001170	1	0.0009867	0.00	36.15
030001180	1	0.0009867	0.00	36.15

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030001190	1	0.0009867	0.00	36.16
030001200	1	0.0009867	0.00	36.16
030001210	1	0.0009867	0.00	36.16
030001220	1	0.0009867	0.00	36.16
030001230	1	0.0009867	0.00	36.16
030001240	1	0.0009867	0.00	36.16
030001250	1	0.0009867	0.00	36.16
030001260	1	0.0009867	0.00	36.16
030001270	1	0.0009867	0.00	36.16
030001280	1	0.0009867	0.00	36.16
030001290	1	0.0009867	0.00	36.17
030001300	1	0.0009867	0.00	36.17
030001310	1	0.0009867	0.00	36.17
030001320	1	0.0009867	0.00	36.17
030001330	1	0.0009867	0.00	36.17
030001370	1	0.0009867	0.00	36.17
030001380	1	0.0009867	0.00	36.17
030001390	1	0.0009867	0.00	36.17
030001400	1	0.0009867	0.00	36.17
030001430	1	0.0009867	0.00	36.17
030001440	1	0.0009867	0.00	36.18
030001470	1	0.0009867	0.00	36.18
030001520	1	0.0009867	0.00	36.18
030010	1140	1.1248705	1.12	37.30
030010010	121	0.1193941	0.12	37.42
030010020	114	0.1124870	0.11	37.53
030010020010	4	0.0039469	0.00	37.54
030010020020	6	0.0059204	0.01	37.54
030010020030	6	0.0059204	0.01	37.55

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030010030	73	0.0720312	0.07	37.62
030010030010	17	0.0167744	0.02	37.64
030010030020	8	0.0078938	0.01	37.65
030010030030	4	0.0039469	0.00	37.65
030010030040	6	0.0059204	0.01	37.66
030010030050	1	0.0009867	0.00	37.66
030010031	5	0.0049336	0.00	37.66
030010040	35	0.0345355	0.03	37.70
030010040010	16	0.0157877	0.02	37.71
030010040020	5	0.0049336	0.00	37.72
030010050	19	0.0187478	0.02	37.74
030010060	10	0.0098673	0.01	37.75
030010070	3	0.0029602	0.00	37.75
030010080	3	0.0029602	0.00	37.75
030010150	3	0.0029602	0.00	37.76
030020	1044	1.0301446	1.03	38.79
030020010	106	0.1045932	0.10	38.89
030020010010	11	0.0108540	0.01	38.90
030020010020	13	0.0128275	0.01	38.91
030020010030	6	0.0059204	0.01	38.92
030020010050	4	0.0039469	0.00	38.92
030020010060	7	0.0069071	0.01	38.93
030020020	124	0.1223543	0.12	39.05
030020020010	4	0.0039469	0.00	39.06
030020020020	2	0.0019735	0.00	39.06
030020020030	3	0.0029602	0.00	39.06
030020020040	7	0.0069071	0.01	39.07
030020020050	3	0.0029602	0.00	39.07
030020020060	1	0.0009867	0.00	39.07

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030020020070	2	0.0019735	0.00	39.07
030020020080	2	0.0019735	0.00	39.08
030020020090	5	0.0049336	0.00	39.08
030020020100	1	0.0009867	0.00	39.08
030020030	87	0.0858454	0.09	39.17
030020040	30	0.0296019	0.03	39.20
030020050	21	0.0207213	0.02	39.22
030020060	10	0.0098673	0.01	39.23
030020060010	6	0.0059204	0.01	39.23
030020060020	4	0.0039469	0.00	39.24
030020070	4	0.0039469	0.00	39.24
030020070010	1	0.0009867	0.00	39.24
030020070020	5	0.0049336	0.00	39.25
030020070030	3	0.0029602	0.00	39.25
030020080	9	0.0088806	0.01	39.26
030020090	6	0.0059204	0.01	39.27
030020170	2	0.0019735	0.00	39.27
030020180	2	0.0019735	0.00	39.27
030020190	12	0.0118407	0.01	39.28
030021	2	0.0019735	0.00	39.28
030030	618	0.6097982	0.61	39.89
030030010	105	0.1036065	0.10	40.00
030030020	86	0.0848587	0.08	40.08
030030030	38	0.0374957	0.04	40.12
030030030010	10	0.0098673	0.01	40.13
030030030020	9	0.0088806	0.01	40.14
030030030030	7	0.0069071	0.01	40.15
030030030040	5	0.0049336	0.00	40.15
030030030050	7	0.0069071	0.01	40.16

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030030030060	20	0.0197346	0.02	40.18
030030040	35	0.0345355	0.03	40.21
030030050	17	0.0167744	0.02	40.23
030030060	4	0.0039469	0.00	40.23
030030070	4	0.0039469	0.00	40.24
030030270	2	0.0019735	0.00	40.24
030030280	2	0.0019735	0.00	40.24
030030290	5	0.0049336	0.00	40.24
030030330	4	0.0039469	0.00	40.25
030031	11	0.0108540	0.01	40.26
030032	8	0.0078938	0.01	40.27
030040	347	0.3423948	0.34	40.61
030040010	41	0.0404559	0.04	40.65
030040010010	1	0.0009867	0.00	40.65
030040010020	4	0.0039469	0.00	40.66
030040010030	3	0.0029602	0.00	40.66
030040010040	1	0.0009867	0.00	40.66
030040010050	2	0.0019735	0.00	40.66
030040010060	1	0.0009867	0.00	40.66
030040010070	1	0.0009867	0.00	40.66
030040010080	3	0.0029602	0.00	40.67
030040020	40	0.0394691	0.04	40.71
030040020010	44	0.0434161	0.04	40.75
030040020020	18	0.0177611	0.02	40.77
030040020030	8	0.0078938	0.01	40.77
030040020040	4	0.0039469	0.00	40.78
030040020050	2	0.0019735	0.00	40.78
030040020060	1	0.0009867	0.00	40.78
030040020070	1	0.0009867	0.00	40.78

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030040020080	4	0.0039469	0.00	40.79
030040030	8	0.0078938	0.01	40.79
030040030010	2	0.0019735	0.00	40.80
030040030020	5	0.0049336	0.00	40.80
030040030030	2	0.0019735	0.00	40.80
030040040	16	0.0157877	0.02	40.82
030040050	2	0.0019735	0.00	40.82
030040100	2	0.0019735	0.00	40.82
030041	2	0.0019735	0.00	40.82
030050	186	0.1835315	0.18	41.01
030050010	49	0.0483497	0.05	41.06
030050020	31	0.0305886	0.03	41.09
030050030	11	0.0108540	0.01	41.10
030050040	7	0.0069071	0.01	41.11
030050050	3	0.0029602	0.00	41.11
030060	131	0.1292614	0.13	41.24
030060010010	1	0.0009867	0.00	41.24
030060010020	3	0.0029602	0.00	41.24
030060010030	4	0.0039469	0.00	41.25
030060010040	1	0.0009867	0.00	41.25
030060010050	1	0.0009867	0.00	41.25
030060010060	1	0.0009867	0.00	41.25
030060020	3	0.0029602	0.00	41.25
030070	75	0.0740046	0.07	41.33
030070010	83	0.0818985	0.08	41.41
030070010010	4	0.0039469	0.00	41.41
030070010020	3	0.0029602	0.00	41.41
030070020	11	0.0108540	0.01	41.42
030070020010	8	0.0078938	0.01	41.43

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030070020020	5	0.0049336	0.00	41.44
030070030	3	0.0029602	0.00	41.44
030070030010	2	0.0019735	0.00	41.44
030070030020	8	0.0078938	0.01	41.45
030070030030	2	0.0019735	0.00	41.45
030070030040	1	0.0009867	0.00	41.45
030070040	6	0.0059204	0.01	41.46
030080	38	0.0374957	0.04	41.50
030080010	56	0.0552568	0.06	41.55
030080020	24	0.0236815	0.02	41.58
030081	2	0.0019735	0.00	41.58
030090	57	0.0562435	0.06	41.63
030100	47	0.0463762	0.05	41.68
030100010	6	0.0059204	0.01	41.69
030100020	4	0.0039469	0.00	41.69
030100030	4	0.0039469	0.00	41.69
030110	29	0.0286151	0.03	41.72
030120	36	0.0355222	0.04	41.76
030130	31	0.0305886	0.03	41.79
030130010	13	0.0128275	0.01	41.80
030130020	12	0.0118407	0.01	41.81
030130030	2	0.0019735	0.00	41.82
030130040	2	0.0019735	0.00	41.82
030140	26	0.0256549	0.03	41.84
030150	48	0.0473630	0.05	41.89
030160	1	0.0009867	0.00	41.89
030170	4	0.0039469	0.00	41.90
030220	1	0.0009867	0.00	41.90
030260	1	0.0009867	0.00	41.90

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
030280	15	0.0148009	0.01	41.91
030970	57	0.0562435	0.06	41.97
030971	5	0.0049336	0.00	41.97
030980	1	0.0009867	0.00	41.97
030990	1	0.0009867	0.00	41.98
030991	1	0.0009867	0.00	41.98
030992	1	0.0009867	0.00	41.98
030993	1	0.0009867	0.00	41.98
031	21	0.0207213	0.02	42.00
032	5	0.0049336	0.00	42.00
040	4226	4.1699146	4.17	46.17
040001570	1	0.0009867	0.00	46.17
040001630	1	0.0009867	0.00	46.18
040001640	1	0.0009867	0.00	46.18
040001650	1	0.0009867	0.00	46.18
040001660	1	0.0009867	0.00	46.18
040001670	1	0.0009867	0.00	46.18
040001680	1	0.0009867	0.00	46.18
040001690	1	0.0009867	0.00	46.18
040001700	1	0.0009867	0.00	46.18
040001710	1	0.0009867	0.00	46.18
040001720	1	0.0009867	0.00	46.18
040001730	1	0.0009867	0.00	46.19
040001740	1	0.0009867	0.00	46.19
040001750	1	0.0009867	0.00	46.19
040001760	1	0.0009867	0.00	46.19
040001770	1	0.0009867	0.00	46.19
040001780	1	0.0009867	0.00	46.19
040001790	1	0.0009867	0.00	46.19

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040001800	1	0.0009867	0.00	46.19
040001810	1	0.0009867	0.00	46.19
040001820	1	0.0009867	0.00	46.19
040001830	1	0.0009867	0.00	46.20
040001840	1	0.0009867	0.00	46.20
040001850	1	0.0009867	0.00	46.20
040001860	1	0.0009867	0.00	46.20
040001870	1	0.0009867	0.00	46.20
040001880	1	0.0009867	0.00	46.20
040001890	1	0.0009867	0.00	46.20
040001900	1	0.0009867	0.00	46.20
040001910	1	0.0009867	0.00	46.20
040001920	1	0.0009867	0.00	46.20
040001930	1	0.0009867	0.00	46.21
040001940	1	0.0009867	0.00	46.21
040001950	1	0.0009867	0.00	46.21
040001960	1	0.0009867	0.00	46.21
040001970	1	0.0009867	0.00	46.21
040001980	1	0.0009867	0.00	46.21
040001990	1	0.0009867	0.00	46.21
040002000	1	0.0009867	0.00	46.21
040002010	1	0.0009867	0.00	46.21
040002020	1	0.0009867	0.00	46.21
040002030	1	0.0009867	0.00	46.22
040002070	1	0.0009867	0.00	46.22
040002080	1	0.0009867	0.00	46.22
040002090	1	0.0009867	0.00	46.22
040002100	1	0.0009867	0.00	46.22
040002120	1	0.0009867	0.00	46.22

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040002130	1	0.0009867	0.00	46.22
040002140	1	0.0009867	0.00	46.22
040002150	1	0.0009867	0.00	46.22
040002170	1	0.0009867	0.00	46.22
040002180	1	0.0009867	0.00	46.23
040010	521	0.5140855	0.51	46.74
040010010	57	0.0562435	0.06	46.80
040010020	76	0.0749914	0.07	46.87
040010020010	7	0.0069071	0.01	46.88
040010020020	20	0.0197346	0.02	46.90
040010020030	1	0.0009867	0.00	46.90
040010020040	3	0.0029602	0.00	46.90
040010030	30	0.0296019	0.03	46.93
040010030010	8	0.0078938	0.01	46.94
040010030020	4	0.0039469	0.00	46.94
040010030030	7	0.0069071	0.01	46.95
040010040	19	0.0187478	0.02	46.97
040010050	27	0.0266417	0.03	46.99
040010060	1	0.0009867	0.00	47.00
040010060010	28	0.0276284	0.03	47.02
040010060020010	6	0.0059204	0.01	47.03
040010060020020	1	0.0009867	0.00	47.03
040010060020030010	7	0.0069071	0.01	47.04
040010060020030020	28	0.0276284	0.03	47.06
040010060020030030	20	0.0197346	0.02	47.08
040010060020030040	12	0.0118407	0.01	47.10
040010060020030050	36	0.0355222	0.04	47.13
040010060020040	1	0.0009867	0.00	47.13
040010060030	6	0.0059204	0.01	47.14

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040010070	1	0.0009867	0.00	47.14
040010070010	5	0.0049336	0.00	47.14
040010070011	2	0.0019735	0.00	47.15
040010070020010	1	0.0009867	0.00	47.15
040010070020020	13	0.0128275	0.01	47.16
040010070020030	4	0.0039469	0.00	47.16
040010070020040	4	0.0039469	0.00	47.17
040010070020050	3	0.0029602	0.00	47.17
040010070030	1	0.0009867	0.00	47.17
040010080	3	0.0029602	0.00	47.18
040020	432	0.4262667	0.43	47.60
040020010	141	0.1391287	0.14	47.74
040020010010	7	0.0069071	0.01	47.75
040020010020	13	0.0128275	0.01	47.76
040020010030	8	0.0078938	0.01	47.77
040020010040	11	0.0108540	0.01	47.78
040020010050	6	0.0059204	0.01	47.79
040020010060	7	0.0069071	0.01	47.79
040020010070	4	0.0039469	0.00	47.80
040020010080	1	0.0009867	0.00	47.80
040020020	117	0.1154472	0.12	47.91
040020020010	29	0.0286151	0.03	47.94
040020020020	8	0.0078938	0.01	47.95
040020020030	12	0.0118407	0.01	47.96
040020020040	15	0.0148009	0.01	47.98
040020020050	7	0.0069071	0.01	47.98
040020030	62	0.0611772	0.06	48.04
040020030010	28	0.0276284	0.03	48.07
040020030020	5	0.0049336	0.00	48.08

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040020030030	3	0.0029602	0.00	48.08
040020030040	1	0.0009867	0.00	48.08
040020040	37	0.0365090	0.04	48.12
040020050	82	0.0809117	0.08	48.20
040020060	16	0.0157877	0.02	48.21
040020070	12	0.0118407	0.01	48.23
040020070010	27	0.0266417	0.03	48.25
040020070020	6	0.0059204	0.01	48.26
040020080	7	0.0069071	0.01	48.26
040020090	5	0.0049336	0.00	48.27
040030	405	0.3996250	0.40	48.67
040030010	28	0.0276284	0.03	48.70
040030010010	38	0.0374957	0.04	48.73
040030010020	52	0.0513099	0.05	48.79
040030010030	21	0.0207213	0.02	48.81
040030010040	2	0.0019735	0.00	48.81
040030010050	7	0.0069071	0.01	48.82
040030010060	14	0.0138142	0.01	48.83
040030020	61	0.0601904	0.06	48.89
040030020010	9	0.0088806	0.01	48.90
040030020020	2	0.0019735	0.00	48.90
040030030	49	0.0483497	0.05	48.95
040030040	11	0.0108540	0.01	48.96
040030040010	1	0.0009867	0.00	48.96
040030040020	1	0.0009867	0.00	48.96
040030050	2	0.0019735	0.00	48.96
040030060	1	0.0009867	0.00	48.96
040040	205	0.2022793	0.20	49.17
040040010	71	0.0700577	0.07	49.24

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040040020	14	0.0138142	0.01	49.25
040040030	5	0.0049336	0.00	49.26
040050	102	0.1006463	0.10	49.36
040050010	5	0.0049336	0.00	49.36
040050020	3	0.0029602	0.00	49.36
040050030	7	0.0069071	0.01	49.37
040060	49	0.0483497	0.05	49.42
040060010	6	0.0059204	0.01	49.43
040060020	4	0.0039469	0.00	49.43
040060050	5	0.0049336	0.00	49.43
040070	12	0.0118407	0.01	49.45
040080	30	0.0296019	0.03	49.48
040090	35	0.0345355	0.03	49.51
040100	17	0.0167744	0.02	49.53
040110	12	0.0118407	0.01	49.54
040120	6	0.0059204	0.01	49.54
040130	5	0.0049336	0.00	49.55
040140	2	0.0019735	0.00	49.55
040160	14	0.0138142	0.01	49.57
040170	6	0.0059204	0.01	49.57
040190	1	0.0009867	0.00	49.57
040200	1	0.0009867	0.00	49.57
040390	1	0.0009867	0.00	49.57
040400	8	0.0078938	0.01	49.58
040410	1	0.0009867	0.00	49.58
040420	1	0.0009867	0.00	49.58
040430	1	0.0009867	0.00	49.59
040440	1	0.0009867	0.00	49.59
040450	1	0.0009867	0.00	49.59

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
040460	1	0.0009867	0.00	49.59
040470	1	0.0009867	0.00	49.59
040480	1	0.0009867	0.00	49.59
040500	1	0.0009867	0.00	49.59
040510	1	0.0009867	0.00	49.59
040520	1	0.0009867	0.00	49.59
041	38	0.0374957	0.04	49.63
041010	9	0.0088806	0.01	49.64
041020	5	0.0049336	0.00	49.64
041030	2	0.0019735	0.00	49.65
041040	10	0.0098673	0.01	49.66
041451	3	0.0029602	0.00	49.66
041452	3	0.0029602	0.00	49.66
041454	14	0.0138142	0.01	49.68
042	2	0.0019735	0.00	49.68
042010	1	0.0009867	0.00	49.68
042020	9	0.0088806	0.01	49.69
042030	1	0.0009867	0.00	49.69
042040	11	0.0108540	0.01	49.70
050	3143	3.1012877	3.10	52.80
050002190	1	0.0009867	0.00	52.80
050002200	1	0.0009867	0.00	52.80
050002210	1	0.0009867	0.00	52.80
050002220	1	0.0009867	0.00	52.80
050002230	1	0.0009867	0.00	52.81
050002231	1	0.0009867	0.00	52.81
050002240	3	0.0029602	0.00	52.81
050002250	1	0.0009867	0.00	52.81
050002260	1	0.0009867	0.00	52.81

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
050002270	1	0.0009867	0.00	52.81
050002280	1	0.0009867	0.00	52.81
050002290	60	0.0592037	0.06	52.87
050010	317	0.3127929	0.31	53.19
050010010	34	0.0335488	0.03	53.22
050010020	39	0.0384824	0.04	53.26
050010030	16	0.0157877	0.02	53.27
050010040	42	0.0414426	0.04	53.31
050010050	37	0.0365090	0.04	53.35
050010050010	1	0.0009867	0.00	53.35
050010050020	1	0.0009867	0.00	53.35
050010050030	2	0.0019735	0.00	53.36
050010050040	2	0.0019735	0.00	53.36
050010060	5	0.0049336	0.00	53.36
050010070	3	0.0029602	0.00	53.37
050010080	4	0.0039469	0.00	53.37
050010090	5	0.0049336	0.00	53.37
050010100010	3	0.0029602	0.00	53.38
050010100020	3	0.0029602	0.00	53.38
050010100030	1	0.0009867	0.00	53.38
050010110010	1	0.0009867	0.00	53.38
050010110020	19	0.0187478	0.02	53.40
050010110030	13	0.0128275	0.01	53.41
050010110040	8	0.0078938	0.01	53.42
050010110050	10	0.0098673	0.01	53.43
050010110060	1	0.0009867	0.00	53.43
050010120	9	0.0088806	0.01	53.44
050011	5	0.0049336	0.00	53.45
050020	291	0.2871380	0.29	53.73

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
050020010	72	0.0710445	0.07	53.80
050020010010	12	0.0118407	0.01	53.82
050020010010010	17	0.0167744	0.02	53.83
050020010010020	15	0.0148009	0.01	53.85
050020010010030	14	0.0138142	0.01	53.86
050020010020	12	0.0118407	0.01	53.87
050020020	38	0.0374957	0.04	53.91
050020020010	5	0.0049336	0.00	53.92
050020020020	15	0.0148009	0.01	53.93
050020020030	31	0.0305886	0.03	53.96
050020020040	28	0.0276284	0.03	53.99
050020020050	4	0.0039469	0.00	53.99
050020030	54	0.0532833	0.05	54.05
050020040	53	0.0522966	0.05	54.10
050020040010	28	0.0276284	0.03	54.13
050020040020	16	0.0157877	0.02	54.14
050020040030	8	0.0078938	0.01	54.15
050020050	17	0.0167744	0.02	54.17
050020060	11	0.0108540	0.01	54.18
050020070	4	0.0039469	0.00	54.18
050020080	2	0.0019735	0.00	54.18
050020090	2	0.0019735	0.00	54.19
050020100	2	0.0019735	0.00	54.19
050020110	9	0.0088806	0.01	54.20
050030	201	0.1983324	0.20	54.39
050030010	37	0.0365090	0.04	54.43
050030020	31	0.0305886	0.03	54.46
050030030	54	0.0532833	0.05	54.51
050030040	48	0.0473630	0.05	54.56

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
050030050	5	0.0049336	0.00	54.57
050030060	32	0.0315753	0.03	54.60
050030070	23	0.0226948	0.02	54.62
050030080	9	0.0088806	0.01	54.63
050040	110	0.1085401	0.11	54.74
050040010	9	0.0088806	0.01	54.75
050040020	17	0.0167744	0.02	54.76
050040030	8	0.0078938	0.01	54.77
050050	82	0.0809117	0.08	54.85
050060	20	0.0197346	0.02	54.87
050070	23	0.0226948	0.02	54.90
050080	22	0.0217080	0.02	54.92
050090	16	0.0157877	0.02	54.93
050460	6	0.0059204	0.01	54.94
050660	1	0.0009867	0.00	54.94
051	45	0.0444028	0.04	54.98
052	12	0.0118407	0.01	55.00
053	3	0.0029602	0.00	55.00
060	2380	2.3484138	2.35	57.35
060002300	1	0.0009867	0.00	57.35
060002310	10	0.0098673	0.01	57.36
060002320	12	0.0118407	0.01	57.37
060002330010	13	0.0128275	0.01	57.38
060002330020	7	0.0069071	0.01	57.39
060002340	12	0.0118407	0.01	57.40
060002350	2	0.0019735	0.00	57.40
060002360	3	0.0029602	0.00	57.41
060002370	2	0.0019735	0.00	57.41
060010	194	0.1914253	0.19	57.60

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
060010010	9	0.0088806	0.01	57.61
060010020	23	0.0226948	0.02	57.63
060010030	11	0.0108540	0.01	57.64
060010040	12	0.0118407	0.01	57.65
060010050	5	0.0049336	0.00	57.66
060010060	3	0.0029602	0.00	57.66
060010070	6	0.0059204	0.01	57.67
060020	196	0.1933988	0.19	57.86
060020010	26	0.0256549	0.03	57.89
060020020	23	0.0226948	0.02	57.91
060020030	2	0.0019735	0.00	57.91
060020030010	4	0.0039469	0.00	57.92
060020030020	24	0.0236815	0.02	57.94
060020030030	13	0.0128275	0.01	57.95
060020040	2	0.0019735	0.00	57.95
060020050	3	0.0029602	0.00	57.96
060020060	1	0.0009867	0.00	57.96
060021	4	0.0039469	0.00	57.96
060030	120	0.1184074	0.12	58.08
060030010	23	0.0226948	0.02	58.10
060030020	19	0.0187478	0.02	58.12
060030030	6	0.0059204	0.01	58.13
060030040	2	0.0019735	0.00	58.13
060030050	12	0.0118407	0.01	58.14
060030060	1	0.0009867	0.00	58.14
060040	58	0.0572303	0.06	58.20
060040010	7	0.0069071	0.01	58.21
060040020	6	0.0059204	0.01	58.21
060050	7	0.0069071	0.01	58.22

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
060101	6	0.0059204	0.01	58.23
060102	5	0.0049336	0.00	58.23
060110	3	0.0029602	0.00	58.23
060670	1	0.0009867	0.00	58.23
061	18	0.0177611	0.02	58.25
070	1622	1.6004736	1.60	59.85
070002380	4	0.0039469	0.00	59.86
070002390	1	0.0009867	0.00	59.86
070002400	7	0.0069071	0.01	59.86
070002420	2	0.0019735	0.00	59.87
070002430	1	0.0009867	0.00	59.87
070002440	1	0.0009867	0.00	59.87
070002460	1	0.0009867	0.00	59.87
070002461	4	0.0039469	0.00	59.87
070002462	2	0.0019735	0.00	59.88
070002463	1	0.0009867	0.00	59.88
070002464	2	0.0019735	0.00	59.88
070002465	2	0.0019735	0.00	59.88
070002470	19	0.0187478	0.02	59.90
070002471	2	0.0019735	0.00	59.90
070002480010	2	0.0019735	0.00	59.90
070002480020	9	0.0088806	0.01	59.91
070002480030	6	0.0059204	0.01	59.92
070002480040	4	0.0039469	0.00	59.92
070002490	3	0.0029602	0.00	59.93
070002500	10	0.0098673	0.01	59.93
070002510	2	0.0019735	0.00	59.94
070002520	1	0.0009867	0.00	59.94
070002530	3	0.0029602	0.00	59.94

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
070010	192	0.1894519	0.19	60.13
070010010	21	0.0207213	0.02	60.15
070010020	17	0.0167744	0.02	60.17
070010030	6	0.0059204	0.01	60.17
070010040	6	0.0059204	0.01	60.18
070010050	4	0.0039469	0.00	60.18
070011	1	0.0009867	0.00	60.18
070020	171	0.1687306	0.17	60.35
070020010	12	0.0118407	0.01	60.37
070020020	20	0.0197346	0.02	60.38
070020020010	3	0.0029602	0.00	60.39
070020020020	6	0.0059204	0.01	60.39
070020020030	1	0.0009867	0.00	60.39
070020030	11	0.0108540	0.01	60.41
070020040	1	0.0009867	0.00	60.41
070030	66	0.0651241	0.07	60.47
070030010	7	0.0069071	0.01	60.48
070030020	3	0.0029602	0.00	60.48
070030020010	8	0.0078938	0.01	60.49
070030020020	2	0.0019735	0.00	60.49
070030030	1	0.0009867	0.00	60.49
070030040	1	0.0009867	0.00	60.49
070040	34	0.0335488	0.03	60.53
070050	29	0.0286151	0.03	60.56
070060	6	0.0059204	0.01	60.56
070070	2	0.0019735	0.00	60.56
070080	5	0.0049336	0.00	60.57
070090	6	0.0059204	0.01	60.57
070680	1	0.0009867	0.00	60.58

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
070690	1	0.0009867	0.00	60.58
071	3	0.0029602	0.00	60.58
080	1031	1.0173171	1.02	61.60
080010	229	0.2259608	0.23	61.82
080010010	8	0.0078938	0.01	61.83
080010020	24	0.0236815	0.02	61.85
080010030	14	0.0138142	0.01	61.87
080010040	7	0.0069071	0.01	61.87
080010050	6	0.0059204	0.01	61.88
080020	157	0.1549164	0.15	62.04
080020010	12	0.0118407	0.01	62.05
080020020010	6	0.0059204	0.01	62.05
080020020020	23	0.0226948	0.02	62.08
080020020030	55	0.0542701	0.05	62.13
080020030	9	0.0088806	0.01	62.14
080020040	11	0.0108540	0.01	62.15
080020050	1	0.0009867	0.00	62.15
080020060	8	0.0078938	0.01	62.16
080030	69	0.0680843	0.07	62.23
080030010	15	0.0148009	0.01	62.24
080030020	1	0.0009867	0.00	62.24
080030020010	3	0.0029602	0.00	62.25
080030020020	3	0.0029602	0.00	62.25
080030020030	2	0.0019735	0.00	62.25
080030020040	1	0.0009867	0.00	62.25
080030020050	1	0.0009867	0.00	62.25
080030030	2	0.0019735	0.00	62.25
080030040	1	0.0009867	0.00	62.26
080040	67	0.0661108	0.07	62.32

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
080050	52	0.0513099	0.05	62.37
080060	26	0.0256549	0.03	62.40
080060010	4	0.0039469	0.00	62.40
080060020	3	0.0029602	0.00	62.41
080060030	5	0.0049336	0.00	62.41
080060040	2	0.0019735	0.00	62.41
080070	3	0.0029602	0.00	62.42
080080	7	0.0069071	0.01	62.42
081	5	0.0049336	0.00	62.43
090	1067	1.0528393	1.05	63.48
090010	43	0.0424293	0.04	63.52
090010010	25	0.0246682	0.02	63.55
090010020	12	0.0118407	0.01	63.56
090020	49	0.0483497	0.05	63.61
090020010	9	0.0088806	0.01	63.62
090020020	20	0.0197346	0.02	63.64
090020030	7	0.0069071	0.01	63.64
090030	68	0.0670975	0.07	63.71
090040	32	0.0315753	0.03	63.74
090040010	11	0.0108540	0.01	63.75
090040020	3	0.0029602	0.00	63.76
090050	20	0.0197346	0.02	63.78
090060	1	0.0009867	0.00	63.78
091	11	0.0108540	0.01	63.79
100	580	0.5723025	0.57	64.36
100010	39	0.0384824	0.04	64.40
100010010	5	0.0049336	0.00	64.40
100010020	10	0.0098673	0.01	64.41
100020	32	0.0315753	0.03	64.44

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
100020010	21	0.0207213	0.02	64.46
100020020	21	0.0207213	0.02	64.49
100020030	8	0.0078938	0.01	64.49
100020040	1	0.0009867	0.00	64.49
100020050	9	0.0088806	0.01	64.50
100030	27	0.0266417	0.03	64.53
100030010	4	0.0039469	0.00	64.53
100030020	4	0.0039469	0.00	64.54
100040	6	0.0059204	0.01	64.54
100050	14	0.0138142	0.01	64.56
100051	1	0.0009867	0.00	64.56
100060	8	0.0078938	0.01	64.57
100070	1	0.0009867	0.00	64.57
100080	2	0.0019735	0.00	64.57
100090	4	0.0039469	0.00	64.57
100100	1	0.0009867	0.00	64.57
101	11	0.0108540	0.01	64.59
102	11	0.0108540	0.01	64.60
102010	1	0.0009867	0.00	64.60
102020	2	0.0019735	0.00	64.60
102030010	10	0.0098673	0.01	64.61
102030020010	4	0.0039469	0.00	64.61
102030020020	6	0.0059204	0.01	64.62
102030020030	1	0.0009867	0.00	64.62
102030030	5	0.0049336	0.00	64.62
103	1	0.0009867	0.00	64.63
103010	1	0.0009867	0.00	64.63
103020	1	0.0009867	0.00	64.63
103030	1	0.0009867	0.00	64.63

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
103040	1	0.0009867	0.00	64.63
103050	1	0.0009867	0.00	64.63
103060	1	0.0009867	0.00	64.63
103070	1	0.0009867	0.00	64.63
103080	1	0.0009867	0.00	64.63
103090	1	0.0009867	0.00	64.63
103100	1	0.0009867	0.00	64.64
103110	1	0.0009867	0.00	64.64
103120	1	0.0009867	0.00	64.64
103130	1	0.0009867	0.00	64.64
104	1	0.0009867	0.00	64.64
105	1	0.0009867	0.00	64.64
105010	1	0.0009867	0.00	64.64
106	1	0.0009867	0.00	64.64
107	1	0.0009867	0.00	64.64
107010	1	0.0009867	0.00	64.64
108	1	0.0009867	0.00	64.65
109	1	0.0009867	0.00	64.65
110	431	0.4252800	0.43	65.07
110010	76	0.0749914	0.07	65.15
110010010	6	0.0059204	0.01	65.15
110010020	3	0.0029602	0.00	65.16
110010030	6	0.0059204	0.01	65.16
110020	32	0.0315753	0.03	65.19
110020010	13	0.0128275	0.01	65.21
110020020	5	0.0049336	0.00	65.21
110020030	5	0.0049336	0.00	65.22
110030	29	0.0286151	0.03	65.24
110030010	2	0.0019735	0.00	65.25

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
110030020	6	0.0059204	0.01	65.25
110030030	4	0.0039469	0.00	65.26
110030040	2	0.0019735	0.00	65.26
110040	7	0.0069071	0.01	65.27
110040010	3	0.0029602	0.00	65.27
110040020	5	0.0049336	0.00	65.27
110050	8	0.0078938	0.01	65.28
110050010	10	0.0098673	0.01	65.29
110050020	4	0.0039469	0.00	65.29
110050020010	4	0.0039469	0.00	65.30
110050020020	6	0.0059204	0.01	65.30
110050020030	2	0.0019735	0.00	65.31
110050020040	2	0.0019735	0.00	65.31
110050030	2	0.0019735	0.00	65.31
110050040	5	0.0049336	0.00	65.32
110050050	2	0.0019735	0.00	65.32
110050060	2	0.0019735	0.00	65.32
110050070	3	0.0029602	0.00	65.32
110050080	1	0.0009867	0.00	65.32
110060	10	0.0098673	0.01	65.33
110060010	9	0.0088806	0.01	65.34
110060020	5	0.0049336	0.00	65.35
110070	6	0.0059204	0.01	65.35
110070010	2	0.0019735	0.00	65.35
110070020	6	0.0059204	0.01	65.36
110080	12	0.0118407	0.01	65.37
111	1	0.0009867	0.00	65.37
112	2	0.0019735	0.00	65.38
120	195	0.1924121	0.19	65.57

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
120010	43	0.0424293	0.04	65.61
120020	24	0.0236815	0.02	65.63
120030	24	0.0236815	0.02	65.66
130	141	0.1391287	0.14	65.80
130010	13	0.0128275	0.01	65.81
130020	24	0.0236815	0.02	65.83
130030	10	0.0098673	0.01	65.84
130040	9	0.0088806	0.01	65.85
130050	29	0.0286151	0.03	65.88
130060	2	0.0019735	0.00	65.88
130070	3	0.0029602	0.00	65.89
130080	5	0.0049336	0.00	65.89
140	104	0.1026198	0.10	65.99
150	108	0.1065667	0.11	66.10
150010	8	0.0078938	0.01	66.11
150020	8	0.0078938	0.01	66.12
150030	1	0.0009867	0.00	66.12
160	111	0.1095269	0.11	66.23
160010	13	0.0128275	0.01	66.24
160020	8	0.0078938	0.01	66.25
160030	11	0.0108540	0.01	66.26
170	29	0.0286151	0.03	66.29
171	5	0.0049336	0.00	66.29
172	2	0.0019735	0.00	66.29
174	2	0.0019735	0.00	66.30
175	1	0.0009867	0.00	66.30
176	1	0.0009867	0.00	66.30
180	29	0.0286151	0.03	66.33
180010	34	0.0335488	0.03	66.36

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
180020	2	0.0019735	0.00	66.36
180030	5	0.0049336	0.00	66.37
180040	1	0.0009867	0.00	66.37
180060	1	0.0009867	0.00	66.37
180070	1	0.0009867	0.00	66.37
180080	1	0.0009867	0.00	66.37
180100	1	0.0009867	0.00	66.37
180110	1	0.0009867	0.00	66.37
180130	1	0.0009867	0.00	66.37
180140	3	0.0029602	0.00	66.38
180160	1	0.0009867	0.00	66.38
180170	1	0.0009867	0.00	66.38
180180	1	0.0009867	0.00	66.38
180190	1	0.0009867	0.00	66.38
180200	2	0.0019735	0.00	66.38
190	21	0.0207213	0.02	66.40
200	13	0.0128275	0.01	66.42
210	21	0.0207213	0.02	66.44
220	22	0.0217080	0.02	66.46
230	26	0.0256549	0.03	66.48
231	1	0.0009867	0.00	66.48
232	1	0.0009867	0.00	66.49
240	14	0.0138142	0.01	66.50
250	13	0.0128275	0.01	66.51
260	5	0.0049336	0.00	66.52
270	3	0.0029602	0.00	66.52
280	4	0.0039469	0.00	66.52
281	16	0.0157877	0.02	66.54
290	5	0.0049336	0.00	66.54

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
300	3	0.0029602	0.00	66.55
301	1	0.0009867	0.00	66.55
310	5	0.0049336	0.00	66.55
320	2	0.0019735	0.00	66.56
330	16	0.0157877	0.02	66.57
340	6	0.0059204	0.01	66.58
350	6	0.0059204	0.01	66.58
360	2	0.0019735	0.00	66.59
370	2	0.0019735	0.00	66.59
371	1	0.0009867	0.00	66.59
380	4	0.0039469	0.00	66.59
390	2	0.0019735	0.00	66.59
400	2	0.0019735	0.00	66.60
410	3	0.0029602	0.00	66.60
420	4	0.0039469	0.00	66.60
421	1	0.0009867	0.00	66.60
430	2	0.0019735	0.00	66.61
440	5	0.0049336	0.00	66.61
450	1	0.0009867	0.00	66.61
460	6	0.0059204	0.01	66.62
460460	3	0.0029602	0.00	66.62
460490	2	0.0019735	0.00	66.62
460510	8	0.0078938	0.01	66.63
460590	4	0.0039469	0.00	66.63
460620	1	0.0009867	0.00	66.64
470	3	0.0029602	0.00	66.64
480	4	0.0039469	0.00	66.64
490	1	0.0009867	0.00	66.64
491	4	0.0039469	0.00	66.65

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
510	2	0.0019735	0.00	66.65
520	6	0.0059204	0.01	66.66
530	2	0.0019735	0.00	66.66
550	1	0.0009867	0.00	66.66
560	2	0.0019735	0.00	66.66
580	1	0.0009867	0.00	66.66
590	1	0.0009867	0.00	66.66
600	2	0.0019735	0.00	66.66
610	2	0.0019735	0.00	66.67
620	8	0.0078938	0.01	66.67
630	7	0.0069071	0.01	66.68
640	1	0.0009867	0.00	66.68
650	1	0.0009867	0.00	66.68
680	1	0.0009867	0.00	66.68
681	2	0.0019735	0.00	66.69
690	2	0.0019735	0.00	66.69
700	2	0.0019735	0.00	66.69
700010	1	0.0009867	0.00	66.69
700020	1	0.0009867	0.00	66.69
700030	1	0.0009867	0.00	66.69
700031	1	0.0009867	0.00	66.69
700032	1	0.0009867	0.00	66.69
700040	1	0.0009867	0.00	66.70
700050	1	0.0009867	0.00	66.70
700060	1	0.0009867	0.00	66.70
700070	1	0.0009867	0.00	66.70
700080	1	0.0009867	0.00	66.70
700090	1	0.0009867	0.00	66.70
700100	1	0.0009867	0.00	66.70

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
700110	1	0.0009867	0.00	66.70
700120	1	0.0009867	0.00	66.70
700130	1	0.0009867	0.00	66.70
700140	1	0.0009867	0.00	66.71
700150	1	0.0009867	0.00	66.71
700160	1	0.0009867	0.00	66.71
700170	1	0.0009867	0.00	66.71
700180	1	0.0009867	0.00	66.71
710	2	0.0019735	0.00	66.71
720	7	0.0069071	0.01	66.72
730	1	0.0009867	0.00	66.72
730740	5	0.0049336	0.00	66.72
731	6	0.0059204	0.01	66.73
740	6	0.0059204	0.01	66.74
780	2	0.0019735	0.00	66.74
800	1	0.0009867	0.00	66.74
810	1	0.0009867	0.00	66.74
820	1	0.0009867	0.00	66.74
830	5	0.0049336	0.00	66.75
840	4	0.0039469	0.00	66.75
870	1	0.0009867	0.00	66.75
880	5	0.0049336	0.00	66.76
911	1	0.0009867	0.00	66.76
920	1	0.0009867	0.00	66.76
930	1	0.0009867	0.00	66.76
950	1	0.0009867	0.00	66.76
960	2	0.0019735	0.00	66.76
980	1	0.0009867	0.00	66.76
990	1	0.0009867	0.00	66.76

(continued)

gliederungskennzahl	N	exactpercent	roundedpercent	cumulpercent
991	3	0.0029602	0.00	66.77
NA	33680	33.2330159	33.23	100.00
Total	101345	100.0000000	99.00	100.00

Frequency Table for Variable: ausfertigung_jahr

124 unique value(s) detected.

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1869	149	0.1470225	0.15	0.15
1871	564	0.5565149	0.56	0.70
1877	98	0.0966994	0.10	0.80
1884	10	0.0098673	0.01	0.81
1886	3	0.0029602	0.00	0.81
1887	4	0.0039469	0.00	0.82
1889	218	0.2151068	0.22	1.03
1892	117	0.1154472	0.12	1.15
1894	4	0.0039469	0.00	1.15
1895	74	0.0730179	0.07	1.22
1896	2916	2.8773003	2.88	4.10
1897	1133	1.1179634	1.12	5.22
1898	3	0.0029602	0.00	5.22
1899	2	0.0019735	0.00	5.22
1901	51	0.0503232	0.05	5.28
1903	1	0.0009867	0.00	5.28
1906	1	0.0009867	0.00	5.28
1907	26	0.0256549	0.03	5.30
1908	8	0.0078938	0.01	5.31
1909	129	0.1272880	0.13	5.44

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1910	11	0.0108540	0.01	5.45
1911	47	0.0463762	0.05	5.50
1913	46	0.0453895	0.05	5.54
1919	99	0.0976861	0.10	5.64
1920	7	0.0069071	0.01	5.65
1921	40	0.0394691	0.04	5.68
1922	206	0.2032661	0.20	5.89
1923	15	0.0148009	0.01	5.90
1924	2	0.0019735	0.00	5.90
1925	2	0.0019735	0.00	5.91
1926	31	0.0305886	0.03	5.94
1927	32	0.0315753	0.03	5.97
1928	8	0.0078938	0.01	5.98
1929	8	0.0078938	0.01	5.98
1930	14	0.0138142	0.01	6.00
1931	46	0.0453895	0.05	6.04
1933	196	0.1933988	0.19	6.24
1934	526	0.5190192	0.52	6.76
1935	156	0.1539296	0.15	6.91
1936	287	0.2831911	0.28	7.19
1937	335	0.3305540	0.33	7.52
1938	41	0.0404559	0.04	7.56
1939	90	0.0888056	0.09	7.65
1940	215	0.2121466	0.21	7.87
1941	25	0.0246682	0.02	7.89
1942	20	0.0197346	0.02	7.91
1943	24	0.0236815	0.02	7.93
1944	12	0.0118407	0.01	7.95
1947	19	0.0187478	0.02	7.96

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1948	73	0.0720312	0.07	8.04
1949	314	0.3098327	0.31	8.35
1950	2106	2.0780502	2.08	10.42
1951	584	0.5762494	0.58	11.00
1952	445	0.4390942	0.44	11.44
1953	1806	1.7820317	1.78	13.22
1954	209	0.2062263	0.21	13.43
1955	303	0.2989787	0.30	13.73
1956	704	0.6946569	0.69	14.42
1957	975	0.9620603	0.96	15.38
1958	150	0.1480093	0.15	15.53
1959	766	0.7558340	0.76	16.29
1960	1106	1.0913217	1.09	17.38
1961	1310	1.2926143	1.29	18.67
1962	210	0.2072130	0.21	18.88
1963	336	0.3315408	0.33	19.21
1964	332	0.3275939	0.33	19.54
1965	1487	1.4672653	1.47	21.00
1966	579	0.5713158	0.57	21.58
1967	330	0.3256204	0.33	21.90
1968	477	0.4706695	0.47	22.37
1969	729	0.7193251	0.72	23.09
1970	425	0.4193596	0.42	23.51
1971	776	0.7657013	0.77	24.28
1972	831	0.8199714	0.82	25.10
1973	228	0.2249741	0.22	25.32
1974	788	0.7775421	0.78	26.10
1975	906	0.8939760	0.89	26.99
1976	2480	2.4470867	2.45	29.44

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1977	760	0.7499137	0.75	30.19
1978	399	0.3937047	0.39	30.58
1979	525	0.5180325	0.52	31.10
1980	1103	1.0883615	1.09	32.19
1981	568	0.5604618	0.56	32.75
1982	535	0.5278997	0.53	33.28
1983	243	0.2397750	0.24	33.52
1984	352	0.3473284	0.35	33.87
1985	580	0.5723025	0.57	34.44
1986	548	0.5407272	0.54	34.98
1987	455	0.4489615	0.45	35.43
1988	1284	1.2669594	1.27	36.69
1989	1143	1.1278307	1.13	37.82
1990	2445	2.4125512	2.41	40.23
1991	734	0.7242587	0.72	40.96
1992	994	0.9808081	0.98	41.94
1993	1000	0.9867285	0.99	42.93
1994	3766	3.7160195	3.72	46.64
1995	743	0.7331393	0.73	47.38
1996	914	0.9018699	0.90	48.28
1997	1789	1.7652573	1.77	50.04
1998	1416	1.3972076	1.40	51.44
1999	782	0.7716217	0.77	52.21
2000	705	0.6956436	0.70	52.91
2001	1604	1.5827125	1.58	54.49
2002	2150	2.1214663	2.12	56.61
2003	1182	1.1663131	1.17	57.78
2004	2465	2.4322858	2.43	60.21
2005	2032	2.0050323	2.01	62.22

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
2006	2204	2.1747496	2.17	64.39
2007	1648	1.6261286	1.63	66.02
2008	2166	2.1372539	2.14	68.15
2009	2661	2.6256845	2.63	70.78
2010	1416	1.3972076	1.40	72.18
2011	2354	2.3227589	2.32	74.50
2012	1491	1.4712122	1.47	75.97
2013	3396	3.3509300	3.35	79.32
2014	1601	1.5797523	1.58	80.90
2015	2047	2.0198332	2.02	82.92
2016	3102	3.0608318	3.06	85.98
2017	3507	3.4604569	3.46	89.44
2018	1596	1.5748187	1.57	91.02
2019	1763	1.7396023	1.74	92.76
2020	2909	2.8703932	2.87	95.63
2021	3575	3.5275544	3.53	99.15
2022	857	0.8456263	0.85	100.00
Total	101345	100.0000000	100.00	100.00

Frequency Table for Variable: doi_concept

1 unique value(s) detected.

doi_concept	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.3832111	101345	100	100	100
Total	101345	100	100	100

Frequency Table for Variable: doi_version

1 unique value(s) detected.

doi_version	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.6959374	101345	100	100	100
Total	101345	100	100	100

Frequency Table for Variable: version

1 unique value(s) detected.

version	N	exactpercent	roundedpercent	cumulpercent
2022-08-05	101345	100	100	100
Total	101345	100	100	100

Frequency Table for Variable: lizenz

1 unique value(s) detected.

lizenz	N	exactpercent	roundedpercent	cumulpercent
Creative Commons Zero 1.0 Universal	101345	100	100	100
Total	101345	100	100	100

6 Frequenztabellen erstellen: Rechtsakte

6.1 Variablen ignorieren

Folgende Variablen sind wegen der geringeren Auflösung der Metadaten (nur Rechtsaktebene, nicht Normebene) nicht mehr nutzbar:

```
varremove <- c("gliederungskennzahl")

vars.freqtable.rechtsakte <- grep(paste(varremove,
                                         collapse = "|"),
                                config$freqtable$ignore,
                                invert = TRUE,
                                value = TRUE)
```

6.2 Liste zu prüfender Variablen

```
print(vars.freqtable.rechtsakte)
```

```
## [1] "periodikum"      "fundstellentyp"  "check_neuf"
## [4] "check_aufh"      "check_sonst"     "check_hinweis"
## [7] "check_stand"     "ausfertigung_jahr" "doi_concept"
## [10] "doi_version"     "version"         "lizenz"
```

6.3 Frequenztabellen erstellen

```
prefix.freqtable.rechtsakte <- paste0(config$project$shortname,
                                       "_01_Rechtsakte_Frequenztable_var-")
```

```
f.fast.freqtable(dt.rechtsakte,
                 varlist = vars.freqtable.rechtsakte,
                 sumrow = TRUE,
                 output.list = FALSE,
                 output.kable = TRUE,
                 output.csv = TRUE,
                 outputdir = dir.analysis,
                 prefix = prefix.freqtable.rechtsakte)
```

Frequency Table for Variable: periodikum

21 unique value(s) detected.

periodikum	N	exactpercent	roundedpercent	cumulpercent
BAGVBl	1	0.0176398	0.02	0.02
BAnz	144	2.5401305	2.54	2.56
BGBI	21	0.3704357	0.37	2.93
BGBI I	4885	86.1704004	86.17	89.10
BGBI II	375	6.6149233	6.61	95.71
GBl DDR	15	0.2645969	0.26	95.98
GBl DDR I	43	0.7585112	0.76	96.74
GBl DDR II	3	0.0529194	0.05	96.79
GVBl BE	1	0.0176398	0.02	96.81
NV	1	0.0176398	0.02	96.82
RAnz	2	0.0352796	0.04	96.86
RGBI	51	0.8996296	0.90	97.76
RGBI I	47	0.8290704	0.83	98.59
RGBI II	45	0.7937908	0.79	99.38
RMBI	4	0.0705592	0.07	99.45
VOBl BrZ	1	0.0176398	0.02	99.47
VkBl	11	0.1940377	0.19	99.66
WiGBI	7	0.1234786	0.12	99.79
ZBl	2	0.0352796	0.04	99.82
eBAnz	6	0.1058388	0.11	99.93
Öff Anz	4	0.0705592	0.07	100.00
Total	5669	100.0000000	100.00	100.00

Frequency Table for Variable: fundstellentyp

1 unique value(s) detected.

fundstellentyp	N	exactpercent	roundedpercent	cumulpercent
amtlich	5669	100	100	100
Total	5669	100	100	100

Frequency Table for Variable: check_neuf

2 unique value(s) detected.

check_neuf	N	exactpercent	roundedpercent	cumulpercent
NA	5257	92.732404	92.73	92.73
ja	412	7.267596	7.27	100.00
Total	5669	100.000000	100.00	100.00

Frequency Table for Variable: check_aufh

2 unique value(s) detected.

check_aufh	N	exactpercent	roundedpercent	cumulpercent
NA	5536	97.653907	97.65	97.65
ja	133	2.346093	2.35	100.00
Total	5669	100.000000	100.00	100.00

Frequency Table for Variable: check_sonst

2 unique value(s) detected.

check_sonst	N	exactpercent	roundedpercent	cumulpercent
NA	5196	91.656377	91.66	91.66
ja	473	8.343623	8.34	100.00
Total	5669	100.000000	100.00	100.00

Frequency Table for Variable: check_hinweis

2 unique value(s) detected.

check_hinweis	N	exactpercent	roundedpercent	cumulpercent
NA	5285	93.226319	93.23	93.23
ja	384	6.773681	6.77	100.00
Total	5669	100.000000	100.00	100.00

Frequency Table for Variable: check_stand

2 unique value(s) detected.

check_stand	N	exactpercent	roundedpercent	cumulpercent
NA	2850	50.27342	50.27	50.27
ja	2819	49.72658	49.73	100.00
Total	5669	100.00000	100.00	100.00

Frequency Table for Variable: ausfertigung_jahr

124 unique value(s) detected.

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1869	1	0.0176398	0.02	0.02
1871	3	0.0529194	0.05	0.07
1877	3	0.0529194	0.05	0.12
1884	1	0.0176398	0.02	0.14
1886	2	0.0352796	0.04	0.18
1887	1	0.0176398	0.02	0.19
1889	2	0.0352796	0.04	0.23
1892	1	0.0176398	0.02	0.25
1894	2	0.0352796	0.04	0.28
1895	2	0.0352796	0.04	0.32
1896	2	0.0352796	0.04	0.35
1897	6	0.1058388	0.11	0.46

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1898	1	0.0176398	0.02	0.48
1899	2	0.0352796	0.04	0.51
1901	2	0.0352796	0.04	0.55
1903	1	0.0176398	0.02	0.56
1906	1	0.0176398	0.02	0.58
1907	1	0.0176398	0.02	0.60
1908	1	0.0176398	0.02	0.62
1909	3	0.0529194	0.05	0.67
1910	2	0.0352796	0.04	0.71
1911	2	0.0352796	0.04	0.74
1913	2	0.0352796	0.04	0.78
1919	4	0.0705592	0.07	0.85
1920	1	0.0176398	0.02	0.86
1921	4	0.0705592	0.07	0.93
1922	6	0.1058388	0.11	1.04
1923	2	0.0352796	0.04	1.08
1924	2	0.0352796	0.04	1.11
1925	2	0.0352796	0.04	1.15
1926	4	0.0705592	0.07	1.22
1927	5	0.0881990	0.09	1.31
1928	3	0.0529194	0.05	1.36
1929	3	0.0529194	0.05	1.41
1930	3	0.0529194	0.05	1.46
1931	5	0.0881990	0.09	1.55
1933	8	0.1411184	0.14	1.69
1934	7	0.1234786	0.12	1.82
1935	3	0.0529194	0.05	1.87
1936	7	0.1234786	0.12	1.99
1937	9	0.1587582	0.16	2.15

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1938	10	0.1763980	0.18	2.33
1939	6	0.1058388	0.11	2.43
1940	7	0.1234786	0.12	2.56
1941	2	0.0352796	0.04	2.59
1942	2	0.0352796	0.04	2.63
1943	1	0.0176398	0.02	2.65
1944	1	0.0176398	0.02	2.66
1947	1	0.0176398	0.02	2.68
1948	5	0.0881990	0.09	2.77
1949	12	0.2116775	0.21	2.98
1950	21	0.3704357	0.37	3.35
1951	32	0.5644735	0.56	3.92
1952	14	0.2469571	0.25	4.16
1953	40	0.7055918	0.71	4.87
1954	17	0.2998765	0.30	5.17
1955	26	0.4586347	0.46	5.63
1956	33	0.5821132	0.58	6.21
1957	45	0.7937908	0.79	7.00
1958	20	0.3527959	0.35	7.36
1959	27	0.4762745	0.48	7.83
1960	33	0.5821132	0.58	8.41
1961	52	0.9172694	0.92	9.33
1962	30	0.5291939	0.53	9.86
1963	26	0.4586347	0.46	10.32
1964	28	0.4939143	0.49	10.81
1965	45	0.7937908	0.79	11.61
1966	24	0.4233551	0.42	12.03
1967	39	0.6879520	0.69	12.72
1968	41	0.7232316	0.72	13.44

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1969	54	0.9525490	0.95	14.39
1970	44	0.7761510	0.78	15.17
1971	54	0.9525490	0.95	16.12
1972	53	0.9349092	0.93	17.06
1973	32	0.5644735	0.56	17.62
1974	51	0.8996296	0.90	18.52
1975	64	1.1289469	1.13	19.65
1976	73	1.2877051	1.29	20.94
1977	58	1.0231081	1.02	21.96
1978	43	0.7585112	0.76	22.72
1979	40	0.7055918	0.71	23.43
1980	52	0.9172694	0.92	24.34
1981	40	0.7055918	0.71	25.05
1982	41	0.7232316	0.72	25.77
1983	29	0.5115541	0.51	26.28
1984	32	0.5644735	0.56	26.85
1985	39	0.6879520	0.69	27.54
1986	50	0.8819898	0.88	28.42
1987	38	0.6703122	0.67	29.09
1988	48	0.8467102	0.85	29.93
1989	50	0.8819898	0.88	30.82
1990	157	2.7694479	2.77	33.59
1991	73	1.2877051	1.29	34.87
1992	75	1.3229847	1.32	36.20
1993	82	1.4464632	1.45	37.64
1994	159	2.8047275	2.80	40.45
1995	69	1.2171459	1.22	41.67
1996	84	1.4817428	1.48	43.15
1997	104	1.8345387	1.83	44.98

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1998	117	2.0638561	2.06	47.05
1999	77	1.3582642	1.36	48.40
2000	59	1.0407479	1.04	49.44
2001	108	1.9050979	1.91	51.35
2002	144	2.5401305	2.54	53.89
2003	95	1.6757806	1.68	55.57
2004	137	2.4166520	2.42	57.98
2005	126	2.2226142	2.22	60.20
2006	125	2.2049744	2.20	62.41
2007	102	1.7992591	1.80	64.21
2008	125	2.2049744	2.20	66.41
2009	151	2.6636091	2.66	69.08
2010	115	2.0285765	2.03	71.11
2011	119	2.0991357	2.10	73.21
2012	106	1.8698183	1.87	75.07
2013	175	3.0869642	3.09	78.16
2014	98	1.7286999	1.73	79.89
2015	129	2.2755336	2.28	82.17
2016	147	2.5930499	2.59	84.76
2017	174	3.0693244	3.07	87.83
2018	93	1.6405010	1.64	89.47
2019	114	2.0109367	2.01	91.48
2020	167	2.9458458	2.95	94.43
2021	240	4.2335509	4.23	98.66
2022	76	1.3406244	1.34	100.00
Total	5669	100.0000000	100.00	100.00

Frequency Table for Variable: doi_concept

1 unique value(s) detected.

doi_concept	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.3832111	5669	100	100	100
Total	5669	100	100	100

Frequency Table for Variable: doi_version

1 unique value(s) detected.

doi_version	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.6959374	5669	100	100	100
Total	5669	100	100	100

Frequency Table for Variable: version

1 unique value(s) detected.

version	N	exactpercent	roundedpercent	cumulpercent
2022-08-05	5669	100	100	100
Total	5669	100	100	100

Frequency Table for Variable: lizenz

1 unique value(s) detected.

lizenz	N	exactpercent	roundedpercent	cumulpercent
Creative Commons Zero 1.0 Universal	5669	100	100	100
Total	5669	100	100	100

7 Frequenztabellen erstellen: XML-Metadaten

7.1 Liste zu prüfender Variablen

```
print(vars.freqtable.rechtsakte)
```

```
## [1] "periodikum"      "fundstellentyp"  "check_neuf"
## [4] "check_aufh"      "check_sonst"     "check_hinweis"
## [7] "check_stand"     "ausfertigung_jahr" "doi_concept"
## [10] "doi_version"     "version"         "lizenz"
```

7.2 Frequenztabellen erstellen

```
prefix.freqtable.meta <- paste0(config$project$shortname,
                                "_01_Meta_Frequenztafel_var-")
```

```
f.fast.freqtable(dt.meta,
                 varlist = vars.freqtable.rechtsakte,
                 sumrow = TRUE,
                 output.list = FALSE,
                 output.kable = TRUE,
                 output.csv = TRUE,
                 outputdir = dir.analysis,
                 prefix = prefix.freqtable.meta)
```

Frequency Table for Variable: periodikum

35 unique value(s) detected.

periodikum	N	exactpercent	roundedpercent	cumulpercent
NA	4	0.0600691	0.06	0.06
BAGVBl	2	0.0300345	0.03	0.09
BAnz	255	3.8294038	3.83	3.92
BGBI	23	0.3453972	0.35	4.26
BGBI I	4903	73.6296741	73.63	77.89
BGBI II	1169	17.5551885	17.56	95.45
BGBI III	1	0.0150173	0.02	95.46

(continued)

periodikum	N	exactpercent	roundedpercent	cumulpercent
BremGBI	4	0.0600691	0.06	95.52
GBI BW	1	0.0150173	0.02	95.54
GBI DDR	15	0.2252590	0.23	95.77
GBI DDR I	44	0.6607599	0.66	96.43
GBI DDR II	3	0.0450518	0.05	96.47
GV NW	2	0.0300345	0.03	96.50
GVB I BB	1	0.0150173	0.02	96.52
GVB I BE	1	0.0150173	0.02	96.53
GVB I BY	4	0.0600691	0.06	96.59
GVB I HE	5	0.0750863	0.08	96.67
GVB I RP I	2	0.0300345	0.03	96.70
Mtbl BAA	1	0.0150173	0.02	96.71
NV	4	0.0600691	0.06	96.77
Nds GVB I	1	0.0150173	0.02	96.79
RAnz	2	0.0300345	0.03	96.82
RBesBI	1	0.0150173	0.02	96.83
RGB I	56	0.8409671	0.84	97.67
RGB I I	54	0.8109326	0.81	98.48
RGB I II	48	0.7208290	0.72	99.20
RMB I	4	0.0600691	0.06	99.26
RegBI WB	4	0.0600691	0.06	99.32
RegBI WH	1	0.0150173	0.02	99.34
VOBI BrZ	3	0.0450518	0.05	99.38
VkBI	22	0.3303799	0.33	99.71
WiGBI	7	0.1051209	0.11	99.82
ZBI	2	0.0300345	0.03	99.85
eBAnz	6	0.0901036	0.09	99.94
Öff Anz	4	0.0600691	0.06	100.00
Total	6659	100.0000000	100.00	100.00

Frequency Table for Variable: fundstellentyp

2 unique value(s) detected.

fundstellentyp	N	exactpercent	roundedpercent	cumulpercent
amtlich	6655	99.9399309	99.94	99.94
nichtamtlich	4	0.0600691	0.06	100.00
Total	6659	100.0000000	100.00	100.00

Frequency Table for Variable: check_neuf

2 unique value(s) detected.

check_neuf	N	exactpercent	roundedpercent	cumulpercent
NA	6247	93.812885	93.81	93.81
ja	412	6.187115	6.19	100.00
Total	6659	100.000000	100.00	100.00

Frequency Table for Variable: check_aufh

2 unique value(s) detected.

check_aufh	N	exactpercent	roundedpercent	cumulpercent
NA	6518	97.882565	97.88	97.88
ja	141	2.117435	2.12	100.00
Total	6659	100.000000	100.00	100.00

Frequency Table for Variable: check_sonst

2 unique value(s) detected.

check_sonst	N	exactpercent	roundedpercent	cumulpercent
NA	6176	92.746659	92.75	92.75
ja	483	7.253341	7.25	100.00
Total	6659	100.000000	100.00	100.00

Frequency Table for Variable: check_hinweis

2 unique value(s) detected.

check_hinweis	N	exactpercent	roundedpercent	cumulpercent
NA	6275	94.233368	94.23	94.23
ja	384	5.766632	5.77	100.00
Total	6659	100.000000	100.00	100.00

Frequency Table for Variable: check_stand

2 unique value(s) detected.

check_stand	N	exactpercent	roundedpercent	cumulpercent
NA	3838	57.63628	57.64	57.64
ja	2821	42.36372	42.36	100.00
Total	6659	100.00000	100.00	100.00

Frequency Table for Variable: ausfertigung_jahr

127 unique value(s) detected.

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
NA	1	0.0150173	0.02	0.02
1869	1	0.0150173	0.02	0.03
1871	3	0.0450518	0.05	0.08
1877	3	0.0450518	0.05	0.12

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1884	1	0.0150173	0.02	0.14
1886	2	0.0300345	0.03	0.17
1887	1	0.0150173	0.02	0.18
1889	2	0.0300345	0.03	0.21
1892	1	0.0150173	0.02	0.23
1894	2	0.0300345	0.03	0.26
1895	2	0.0300345	0.03	0.29
1896	2	0.0300345	0.03	0.32
1897	6	0.0901036	0.09	0.41
1898	1	0.0150173	0.02	0.42
1899	2	0.0300345	0.03	0.45
1901	2	0.0300345	0.03	0.48
1903	1	0.0150173	0.02	0.50
1906	1	0.0150173	0.02	0.51
1907	1	0.0150173	0.02	0.53
1908	1	0.0150173	0.02	0.54
1909	4	0.0600691	0.06	0.60
1910	2	0.0300345	0.03	0.63
1911	3	0.0450518	0.05	0.68
1913	2	0.0300345	0.03	0.71
1919	7	0.1051209	0.11	0.81
1920	1	0.0150173	0.02	0.83
1921	4	0.0600691	0.06	0.89
1922	6	0.0901036	0.09	0.98
1923	2	0.0300345	0.03	1.01
1924	2	0.0300345	0.03	1.04
1925	2	0.0300345	0.03	1.07
1926	4	0.0600691	0.06	1.13
1927	5	0.0750863	0.08	1.20

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1928	4	0.0600691	0.06	1.26
1929	4	0.0600691	0.06	1.32
1930	3	0.0450518	0.05	1.37
1931	5	0.0750863	0.08	1.44
1933	9	0.1351554	0.14	1.58
1934	7	0.1051209	0.11	1.68
1935	3	0.0450518	0.05	1.73
1936	9	0.1351554	0.14	1.86
1937	9	0.1351554	0.14	2.00
1938	11	0.1651900	0.17	2.16
1939	11	0.1651900	0.17	2.33
1940	8	0.1201382	0.12	2.45
1941	2	0.0300345	0.03	2.48
1942	2	0.0300345	0.03	2.51
1943	1	0.0150173	0.02	2.52
1944	4	0.0600691	0.06	2.58
1945	1	0.0150173	0.02	2.60
1946	12	0.1802072	0.18	2.78
1947	11	0.1651900	0.17	2.94
1948	13	0.1952245	0.20	3.14
1949	18	0.2703109	0.27	3.41
1950	23	0.3453972	0.35	3.75
1951	41	0.6157081	0.62	4.37
1952	24	0.3604145	0.36	4.73
1953	46	0.6907944	0.69	5.42
1954	26	0.3904490	0.39	5.81
1955	34	0.5105872	0.51	6.32
1956	54	0.8109326	0.81	7.13
1957	55	0.8259498	0.83	7.96

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1958	36	0.5406217	0.54	8.50
1959	48	0.7208290	0.72	9.22
1960	45	0.6757771	0.68	9.90
1961	69	1.0361916	1.04	10.93
1962	40	0.6006908	0.60	11.53
1963	36	0.5406217	0.54	12.07
1964	42	0.6307253	0.63	12.70
1965	61	0.9160535	0.92	13.62
1966	37	0.5556390	0.56	14.18
1967	50	0.7508635	0.75	14.93
1968	50	0.7508635	0.75	15.68
1969	71	1.0662262	1.07	16.74
1970	56	0.8409671	0.84	17.59
1971	67	1.0061571	1.01	18.59
1972	66	0.9911398	0.99	19.58
1973	68	1.0211744	1.02	20.60
1974	59	0.8860189	0.89	21.49
1975	74	1.1112780	1.11	22.60
1976	82	1.2314161	1.23	23.83
1977	71	1.0662262	1.07	24.90
1978	56	0.8409671	0.84	25.74
1979	56	0.8409671	0.84	26.58
1980	67	1.0061571	1.01	27.59
1981	47	0.7058117	0.71	28.29
1982	46	0.6907944	0.69	28.98
1983	40	0.6006908	0.60	29.58
1984	52	0.7808980	0.78	30.36
1985	51	0.7658808	0.77	31.13
1986	73	1.0962607	1.10	32.23

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
1987	46	0.6907944	0.69	32.92
1988	61	0.9160535	0.92	33.83
1989	59	0.8860189	0.89	34.72
1990	171	2.5679531	2.57	37.29
1991	80	1.2013816	1.20	38.49
1992	90	1.3515543	1.35	39.84
1993	99	1.4867097	1.49	41.33
1994	185	2.7781949	2.78	44.11
1995	112	1.6819342	1.68	45.79
1996	103	1.5467788	1.55	47.33
1997	135	2.0273314	2.03	49.36
1998	132	1.9822796	1.98	51.34
1999	93	1.3966061	1.40	52.74
2000	71	1.0662262	1.07	53.81
2001	123	1.8471242	1.85	55.65
2002	151	2.2676077	2.27	57.92
2003	114	1.7119688	1.71	59.63
2004	148	2.2225559	2.22	61.86
2005	147	2.2075387	2.21	64.06
2006	134	2.0123142	2.01	66.08
2007	115	1.7269860	1.73	67.80
2008	148	2.2225559	2.22	70.03
2009	164	2.4628323	2.46	72.49
2010	131	1.9672624	1.97	74.46
2011	132	1.9822796	1.98	76.44
2012	119	1.7870551	1.79	78.22
2013	179	2.6880913	2.69	80.91
2014	100	1.5017270	1.50	82.41
2015	131	1.9672624	1.97	84.38

(continued)

ausfertigung_jahr	N	exactpercent	roundedpercent	cumulpercent
2016	152	2.2826250	2.28	86.66
2017	178	2.6730740	2.67	89.34
2018	94	1.4116234	1.41	90.75
2019	117	1.7570206	1.76	92.51
2020	171	2.5679531	2.57	95.07
2021	250	3.7543175	3.75	98.83
2022	78	1.1713470	1.17	100.00
Total	6659	100.0000000	100.00	100.00

Frequency Table for Variable: doi_concept

1 unique value(s) detected.

doi_concept	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.3832111	6659	100	100	100
Total	6659	100	100	100

Frequency Table for Variable: doi_version

1 unique value(s) detected.

doi_version	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.6959374	6659	100	100	100
Total	6659	100	100	100

Frequency Table for Variable: version

1 unique value(s) detected.

version	N	exactpercent	roundedpercent	cumulpercent
2022-08-05	6659	100	100	100
Total	6659	100	100	100

Frequency Table for Variable: lizenz

1 unique value(s) detected.

	lizenz	N	exactpercent	roundedpercent	cumulpercent
Creative Commons Zero 1.0 Universal		6659	100	100	100
	Total	6659	100	100	100

8 Frequenztabellen visualisieren

8.1 Präfixe erstellen

```
prefix.normen <- paste0(basename(dir.analysis),
                        "/",
                        config$project$shortname,
                        "_01_Einzelnormen_Frequenztafel_var-")

prefix.rechtsakte <- paste0(basename(dir.analysis),
                           "/",
                           config$project$shortname,
                           "_01_Rechtsakte_Frequenztafel_var-")

prefix.meta <- paste0(basename(dir.analysis),
                     "/",
                     config$project$shortname,
                     "_01_Meta_Frequenztafel_var-")
```

8.2 Tabellen für Einzelnormen einlesen

```
table.normen.periodikum <- fread(paste0(prefix.normen,
                                          "periodikum.csv"))

table.normen.ausjahr <- fread(paste0(prefix.normen,
                                      "ausfertigung_jahr.csv"))
```

8.3 Tabellen für Rechtsakte einlesen

```
table.rechtsakte.periodikum <- fread(paste0(prefix.rechtsakte,
                                              "periodikum.csv"))

table.rechtsakte.ausjahr <- fread(paste0(prefix.rechtsakte,
                                          "ausfertigung_jahr.csv"))
```

8.4 Tabellen für XML-Metadaten einlesen

```
table.meta.periodikum <- fread(paste0(prefix.meta,
                                       "periodikum.csv"))

table.meta.ausjahr <- fread(paste0(prefix.meta,
                                    "ausfertigung_jahr.csv"))
```

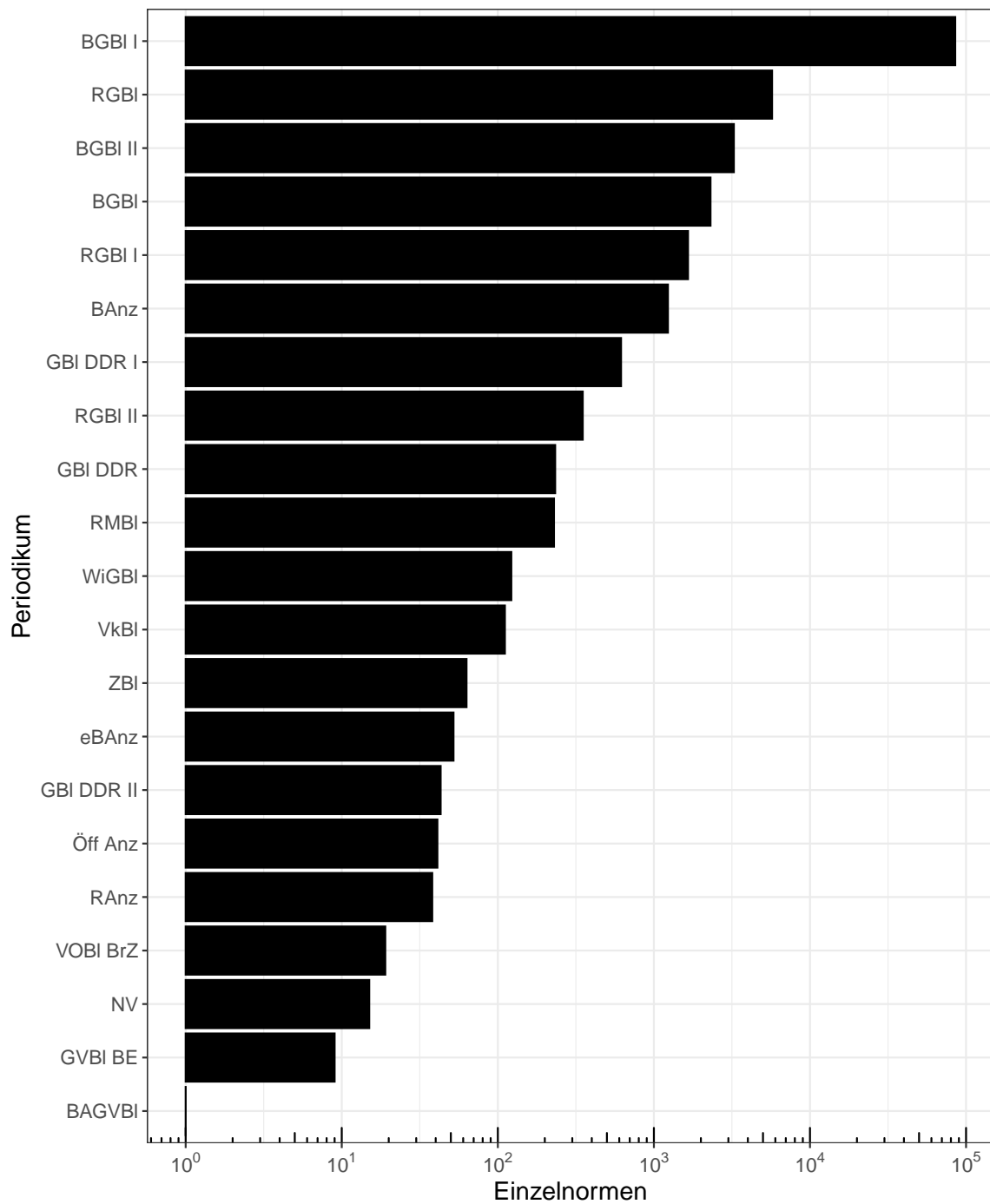
8.5 Periodikum

8.5.1 Einzelnormen

```
freqtable <- table.normen.periodikum[-.N]
```

```
ggplot(data = freqtable)+  
  geom_bar(aes(x = reorder(periodikum,  
                           N),  
              y = N),  
          stat = "identity",  
          fill = "black",  
          color = "black")+  
  scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),  
               labels = trans_format("log10", math_format(10^.x)))+  
  annotation_logticks(sides = "b")+  
  coord_flip()+  
  theme_bw()+  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Einzelnormen je Periodikum"),  
    caption = caption,  
    x = "Periodikum",  
    y = "Einzelnormen"  
  )+  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                              face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

C-DBR | Version 2022-08-05 | Einzelnormen je Periodikum



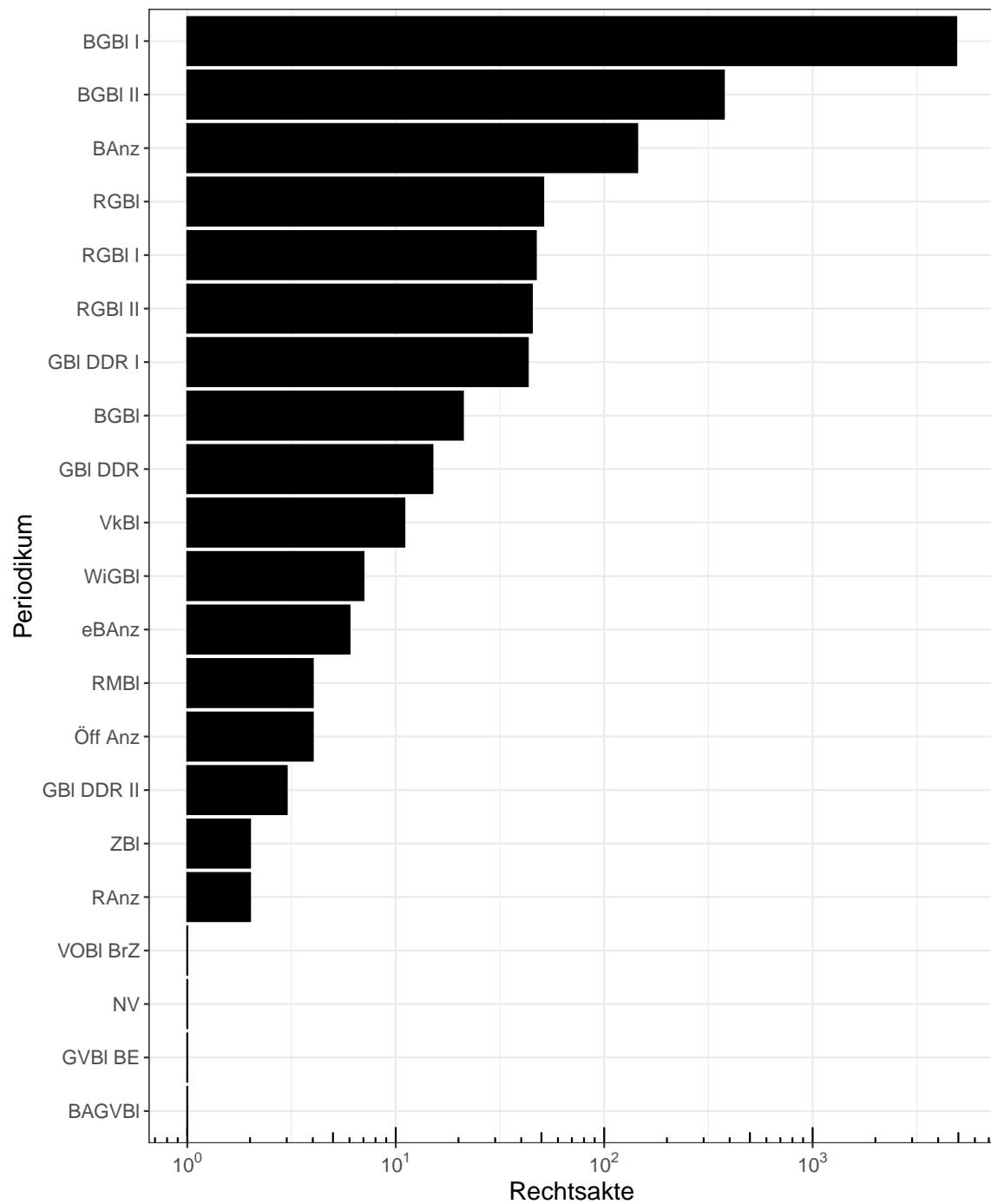
Fobbe | DOI: 10.5281/zenodo.6959374

8.5.2 Rechtsakte

```
freqtable <- table.rechtsakte.periodikum[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(periodikum,  
                           N),  
              y = N),  
          stat = "identity",  
          fill = "black",  
          color = "black") +  
  scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),  
               labels = trans_format("log10", math_format(10^.x)))+  
  annotation_logticks(sides = "b")+  
  coord_flip()+  
  theme_bw()+  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Rechtsakte mit Inhalt je Periodikum"),  
    caption = caption,  
    x = "Periodikum",  
    y = "Rechtsakte"  
  )+  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                              face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

C-DBR | Version 2022-08-05 | Rechtsakte mit Inhalt je Periodikum



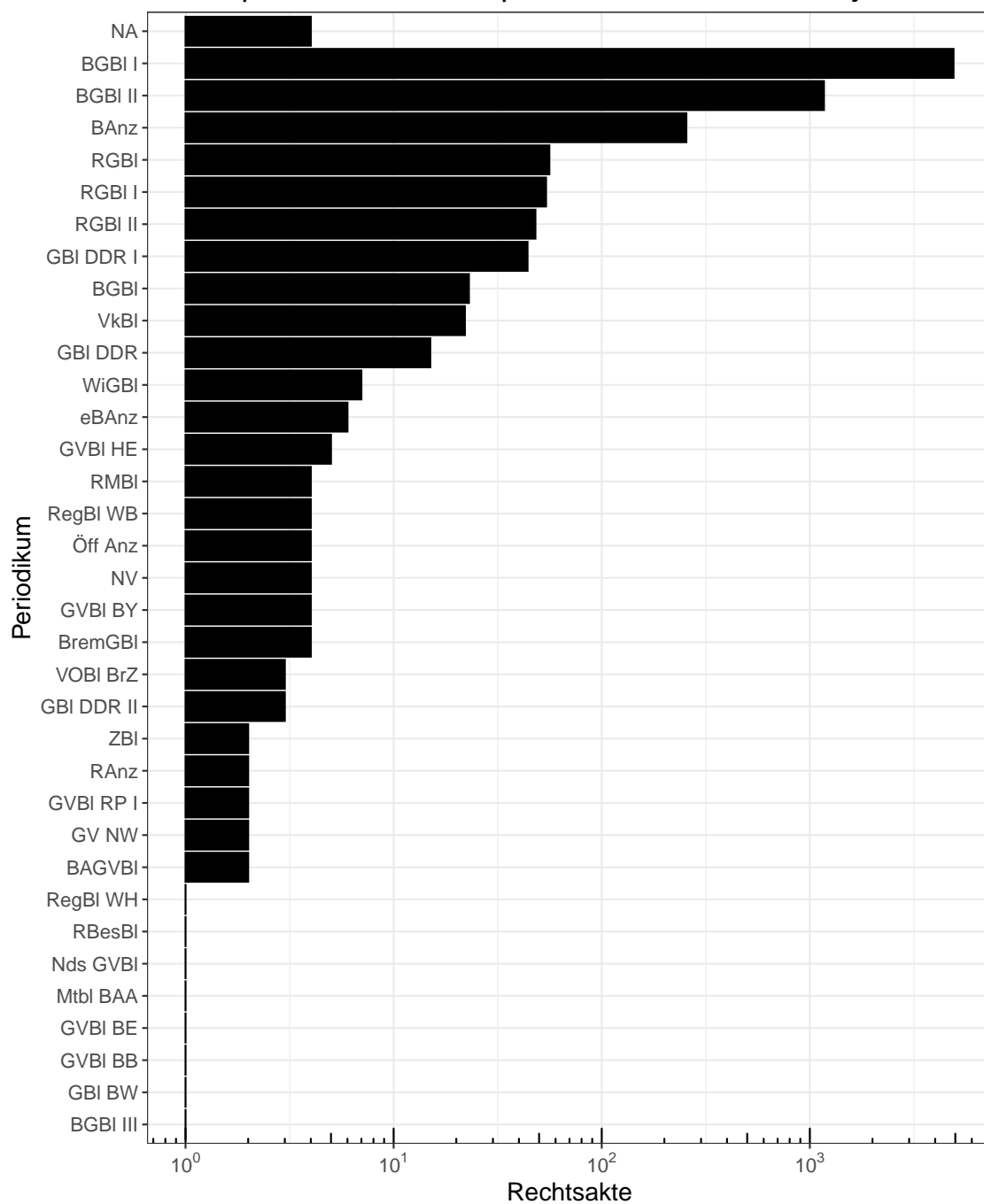
Fobbe | DOI: 10.5281/zenodo.6959374

8.5.3 XML-Metadaten

```
freqtable <- table.meta.periodikum[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(periodikum,  
                           N),  
              y = N),  
          stat = "identity",  
          fill = "black",  
          color = "black") +  
  scale_y_log10(breaks = trans_breaks("log10", function(x) 10^x),  
               labels = trans_format("log10", math_format(10^.x)))+  
  annotation_logticks(sides = "b")+  
  coord_flip()+  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Rechtsakte nach Metadaten je Periodikum"),  
    caption = caption,  
    x = "Periodikum",  
    y = "Rechtsakte"  
  )+  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                              face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

C-DBR | Version 2022-08-05 | Rechtsakte nach Metadaten je Periodikum



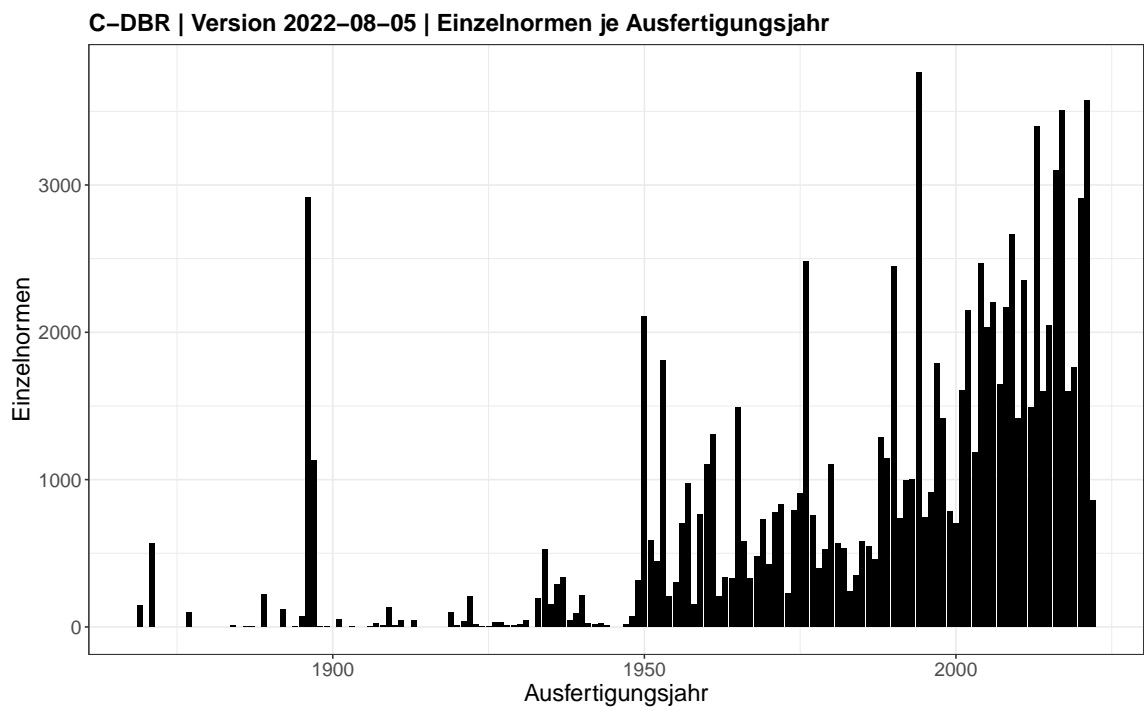
Fobbe | DOI: 10.5281/zenodo.6959374

8.6 Ausfertigungsjahr

8.6.1 Einzelnormen

```
frequetable <- table.normen.ausjahr[-.N][,lapply(.SD, as.numeric)]
```

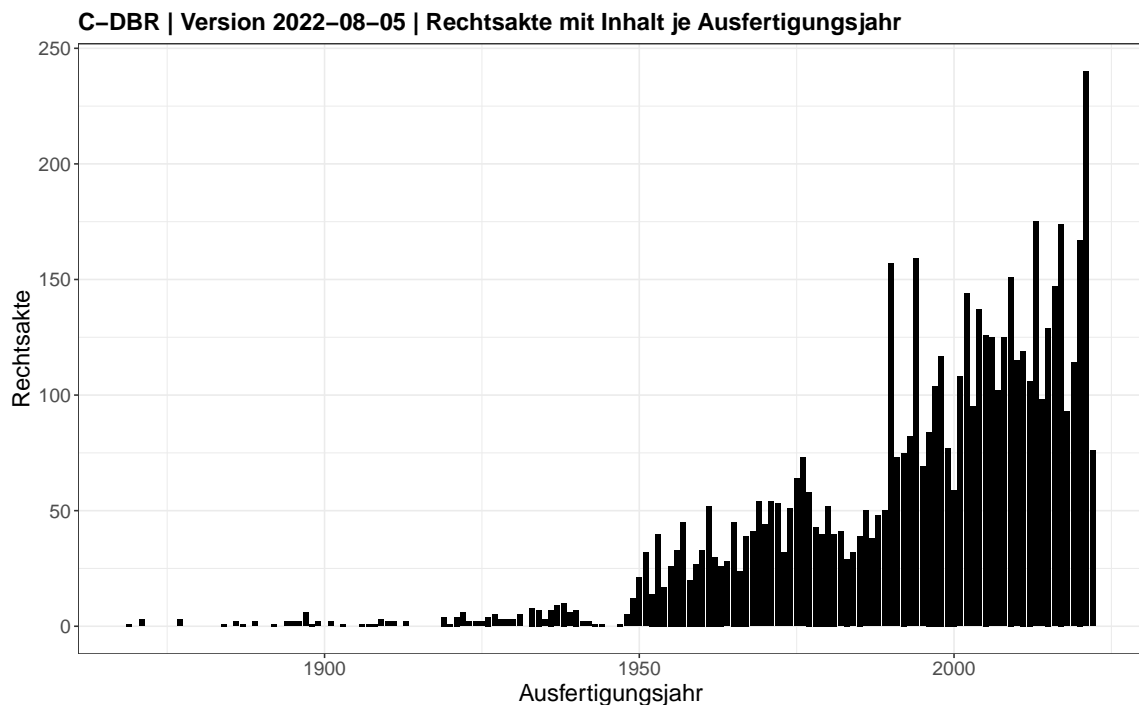
```
ggplot(data = frequetable) +  
  geom_bar(aes(x = ausfertigung_jahr,  
               y = N),  
           stat = "identity",  
           fill = "black")+  
  theme_bw()+  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Einzelnormen je Ausfertigungsjahr"),  
    caption = caption,  
    x = "Ausfertigungsjahr",  
    y = "Einzelnormen"  
  )+  
  theme(  
    text = element_text(size = 16),  
    plot.title = element_text(size = 16,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```



8.6.2 Rechtsakte

```
freqtable <- table.rechtsakte.ausjahr[-.N][,lapply(.SD, as.numeric)]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = ausfertigung_jahr,  
              y = N),  
          stat = "identity",  
          fill = "black") +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Rechtsakte mit Inhalt je Ausfertigungsjahr"),  
    caption = caption,  
    x = "Ausfertigungsjahr",  
    y = "Rechtsakte"  
  ) +  
  theme(  
    text = element_text(size = 16),  
    plot.title = element_text(size = 16,  
                              face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

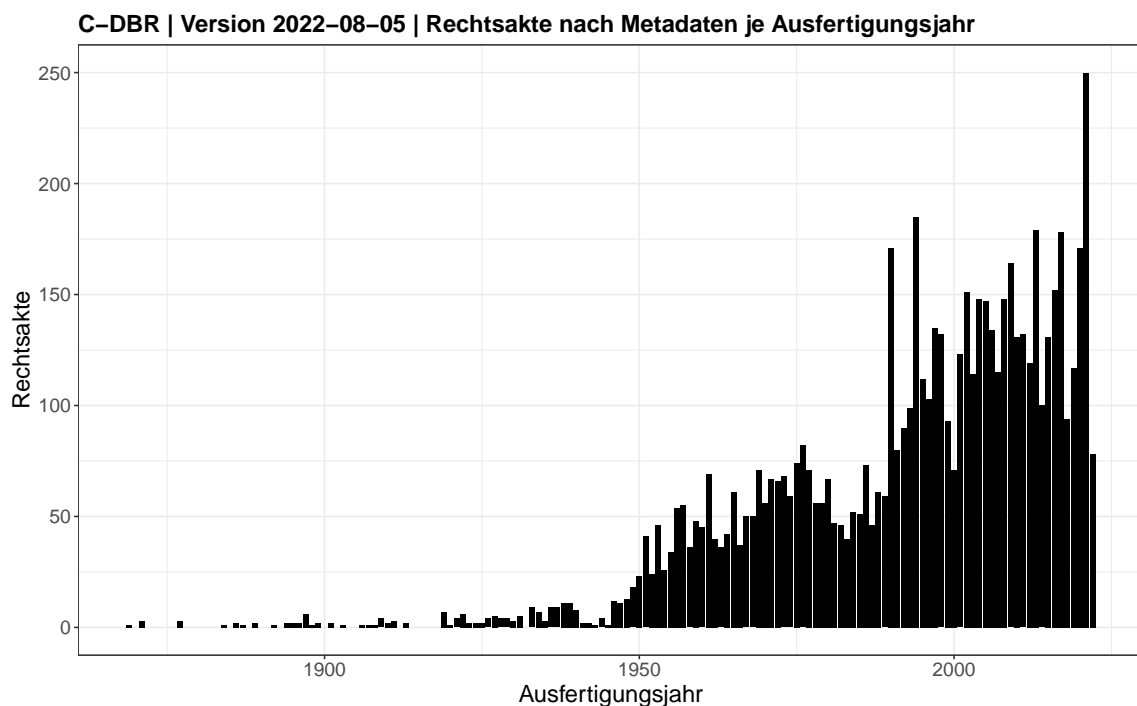


8.6.3 XML-Metadaten

```
freqtable <- table.meta.ausjahr[~.N][,lapply(.SD, as.numeric)]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = ausfertigung_jahr,  
               y = N),  
           stat = "identity",  
           fill = "black") +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Rechtsakte nach Metadaten je Ausfertigungsjahr"),  
    caption = caption,  
    x = "Ausfertigungsjahr",  
    y = "Rechtsakte"  
  ) +  
  theme(  
    text = element_text(size = 16),  
    plot.title = element_text(size = 16,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



Fobbe | DOI: 10.5281/zenodo.6959374

9 Korpus-Analytik

9.1 Berechnung linguistischer Kennwerte

An dieser Stelle werden für jedes Dokument die Anzahl Zeichen, Tokens, Typen und Sätze berechnet und mit den jeweiligen Metadaten verknüpft. Das Ergebnis ist grundsätzlich identisch mit dem eigentlichen Datensatz, nur ohne den Text der Entscheidungen.

9.1.1 Funktion anzeigen: `future_lingsummarize`

```
print(future_lingsummarize)
```

```
## function(dt,
##                                     chunksperworker = 1,
##                                     chunksize = NULL){
##
##   begin.dopar <- Sys.time()
##
##   dt <- dt[,.(doc_id, text)]
##
##   nchars <- dt[, lapply(.(text), nchar)]
##
##   print(paste0("Processing ",
##               dt[,.N],
##               " documents with a total length of ",
##               sum(nchars),
##               " characters."))
##
##   ord <- order(-nchars)
##   dt <- dt[ord]
##
##   raw.list <- split(dt, seq(nrow(dt)))
##
##   result.list <- future_lapply(raw.list,
##                                lingsummarize,
##                                future.seed = TRUE,
##                                future.scheduling = chunksperworker,
##                                future.chunk.size = chunksize)
##
##   result.dt <- rbindlist(result.list)
##
##
##   end.dopar <- Sys.time()
##   duration.dopar <- end.dopar - begin.dopar
##
##
##   summary.corpus <- cbind(nchars[ord],
##                           result.dt)
##
##   setnames(summary.corpus,
```



```
##          "V1",
##          "nchars")
##
##
##      if(dt["nchars" == 0, .N] > 0){
##
##          dt.charnull <- dt["nchars" == 0]
##          dt.charnull$text <- NULL
##          dt.charnull$ntokens <- rep(0, dt.charnull[,.N])
##          dt.charnull$ntypes <- rep(0, dt.charnull[,.N])
##          dt.charnull$nsentences <- rep(0, dt.charnull[,.N])
##
##          summary.corpus <- rbind(summary.corpus,
##                                  dt.charnull)
##      }
##
##
##      summary.corpus <- summary.corpus[order(ord)]
##
##
##      print(paste0("Runtime was ",
##                   round(duration.dopar,
##                           digits = 2),
##                   " ",
##                   attributes(duration.dopar)$units,
##                   ". Ended at ",
##                   end.dopar, "."))
##
##      return(summary.corpus)
##
## }
```

```
#deprecated; Parallelisierung jetzt mit futures
#lingstats.normen.raw <- f.lingsummarize.iterator(dt.normen,
#                                                  threads = fullCores,
#                                                  chunksize = 1)#

#deprecated; Parallelisierung jetzt mit futures
#lingstats.rechtsakte.raw <- f.lingsummarize.iterator(dt.rechtsakte,
#                                                      threads = fullCores,
#                                                      chunksize = 1)
```

9.1.2 Berechnung durchführen

```
if(config$parallel$lingsummarize == TRUE){

  plan("multicore",
        workers = fullCores)

}else{

  plan("sequential")

}
```

```
}
```

```
lingstats.normen.raw <- future_lingsummarize(dt.normen)
```

```
lingstats.rechtsakte.raw <- future_lingsummarize(dt.rechtsakte)
```

9.2 Variablen-Namen anpassen

9.2.1 Einzelnormen

```
setnames(lingstats.normen.raw,  
  old = c("nchars",  
          "ntokens",  
          "ntypes",  
          "nsentences"),  
  new = c("zeichen",  
          "tokens",  
          "typen",  
          "saetze"))
```

9.2.2 Rechtsakte

```
setnames(lingstats.rechtsakte.raw,  
  old = c("nchars",  
          "ntokens",  
          "ntypes",  
          "nsentences"),  
  new = c("zeichen",  
          "tokens",  
          "typen",  
          "saetze"))
```

9.3 Kennwerte den Korpora hinzufügen

9.3.1 Einzelnormen

```
dt.normen <- cbind(dt.normen,  
  lingstats.normen.raw)
```

9.3.2 Rechtsakte

```
dt.rechtsakte <- cbind(dt.rechtsakte,  
  lingstats.rechtsakte.raw)
```

9.4 Varianten mit Metadaten erstellen

9.4.1 Einzelnormen

```
meta.normen <- dt.normen[, !"text"]
```

9.4.2 Rechtsakte

```
meta.rechtsakte <- dt.rechtsakte[, !"text"]
```

9.5 Linguistische Kennwerte: Einzelnormen

Hinweis: Typen sind definiert als einzigartige Tokens und werden hier noch einmal bezogen auf den Gesamtkorpus berechnet, statt wie vorher bezogen auf jedes Dokument.

9.5.1 Zusammenfassungen berechnen

```
dt.summary.ling <- lingstats.normen.raw[, lapply(.SD,
                                              function(x) unclass(summary(x))),
                                          .SDcols = c("zeichen",
                                                    "tokens",
                                                    "typen",
                                                    "saetze")]

dt.sums.ling <- lingstats.normen.raw[,
                                      lapply(.SD, sum),
                                      .SDcols = c("zeichen",
                                                  "tokens",
                                                  "typen",
                                                  "saetze")]

tokens.normen <- tokens(corpus(dt.normen),
                        what = "word",
                        remove_punct = FALSE,
                        remove_symbols = FALSE,
                        remove_numbers = FALSE,
                        remove_url = FALSE,
                        remove_separators = TRUE,
                        split_hyphens = FALSE,
                        include_docvars = FALSE,
                        padding = FALSE
                        )

dt.sums.ling$typen <- nfeat(dfm(tokens.normen))

dt.stats.ling <- rbind(dt.sums.ling,
                      dt.summary.ling)

dt.stats.ling <- transpose(dt.stats.ling,
                           keep.names = "names")

setnames(dt.stats.ling, c("Variable",
                          "Sum",
                          "Min",
                          "Quart1",
                          "Median",
                          "Mean",
                          "Quart3",
```

```
"Max"))
```

9.5.2 Zusammenfassungen anzeigen

```
kable(dt.stats.ling,  
      format.args = list(big.mark = ","),  
      format = "latex",  
      booktabs = TRUE,  
      longtable = TRUE)
```

Variable	Sum	Min	Quart1	Median	Mean	Quart3	Max
zeichen	144,030,059	1	294	690	1,421.185643	1,493	398,079
tokens	22,529,761	1	47	109	222.307573	234	56,584
typen	421,806	1	36	66	93.364665	115	23,559
saetze	752,160	1	2	4	7.421777	8	2,666

9.5.3 Zusammenfassungen speichern

```
fwrite(dt.stats.ling,  
       paste0(dir.analysis,  
              config$project$shortname,  
              "_00_Einzelnormen_KorpusStatistik_ZusammenfassungLinguistisch.csv"),  
       ,  
       na = "NA")
```

9.6 Linguistische Kennwerte: Rechtsakte

Hinweis: Typen sind definiert als einzigartige Tokens und werden hier noch einmal bezogen auf den Gesamtkorpus berechnet, statt wie vorher bezogen auf jedes Dokument.

9.6.1 Zusammenfassungen berechnen

```
dt.summary.ling <- lingstats.rechtsakte.raw[, lapply(.SD,
                                                    function(x)unclass(summary(x))),
                                                    .SDcols = c("zeichen",
                                                                "tokens",
                                                                "typen",
                                                                "saetze")]

dt.sums.ling <- lingstats.rechtsakte.raw[,
                                           lapply(.SD, sum),
                                           .SDcols = c("zeichen",
                                                         "tokens",
                                                         "typen",
                                                         "saetze")]

tokens.rechtsakte <- tokens(corpus(dt.rechtsakte),
                             what = "word",
                             remove_punct = FALSE,
                             remove_symbols = FALSE,
                             remove_numbers = FALSE,
                             remove_url = FALSE,
                             remove_separators = TRUE,
                             split_hyphens = FALSE,
                             include_docvars = FALSE,
                             padding = FALSE
                             )

dt.sums.ling$typen <- nfeat(dfm(tokens.rechtsakte))

dt.stats.ling <- rbind(dt.sums.ling,
                      dt.summary.ling)

dt.stats.ling <- transpose(dt.stats.ling,
                           keep.names = "names")

setnames(dt.stats.ling, c("Variable",
                          "Sum",
                          "Min",
                          "Quart1",
                          "Median",
                          "Mean",
                          "Quart3",
```

```
"Max"))
```


9.6.2 Zusammenfassungen anzeigen

```
kable(dt.stats.ling,  
      format.args = list(big.mark = ","),  
      format = "latex",  
      booktabs = TRUE,  
      longtable = TRUE)
```

Variable	Sum	Min	Quart1	Median	Mean	Quart3	Max
zeichen	144,125,735	98	1,569	5,717	25,423.4847	22,801	2,087,193
tokens	22,529,761	17	265	925	3,974.2037	3,457	306,803
typen	421,806	15	130	305	646.1697	813	23,606
saetze	740,726	1	13	36	130.6626	115	14,306

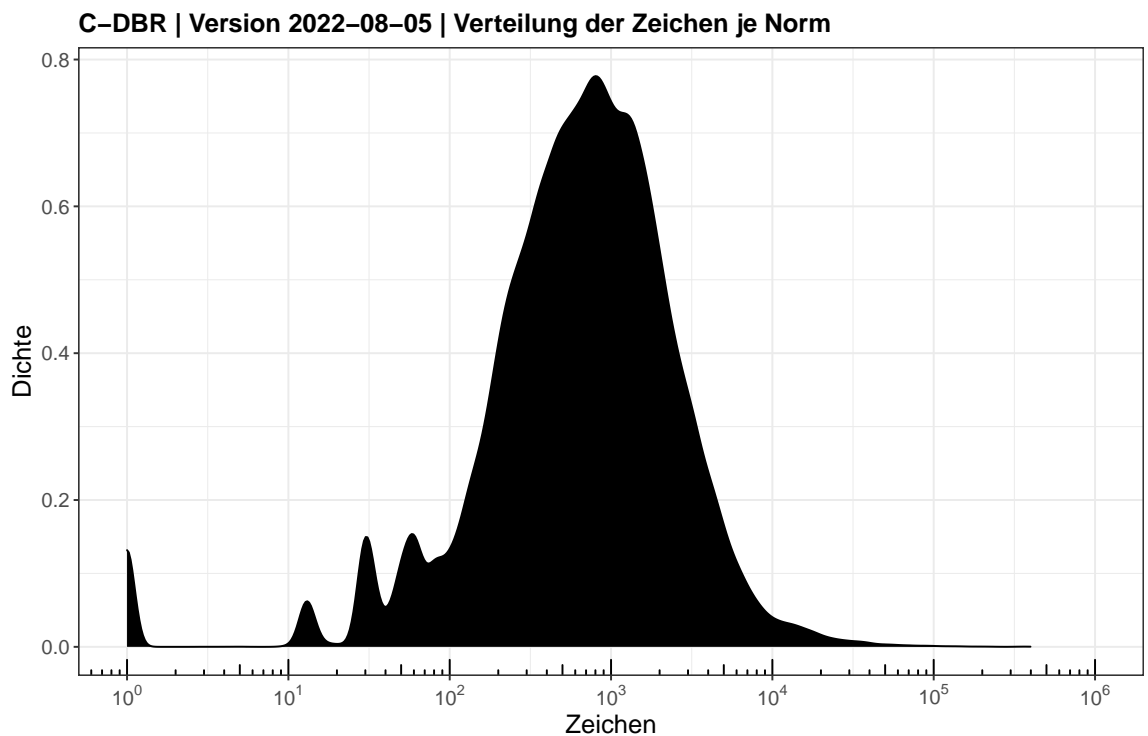
9.6.3 Zusammenfassungen speichern

```
fwrite(dt.stats.ling,  
       paste0(dir.analysis,  
              config$project$shortname,  
              "_00_Rechtsakte_KorpusStatistik_ZusammenfassungLinguistisch.csv"),  
       na = "NA")
```

9.7 Verteilungen

9.7.1 Density (Zeichen)

```
ggplot(data = meta.normen)+
  geom_density(aes(x = zeichen),
    fill = "black") +
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Zeichen je Norm"),
    caption = caption,
    x = "Zeichen",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```



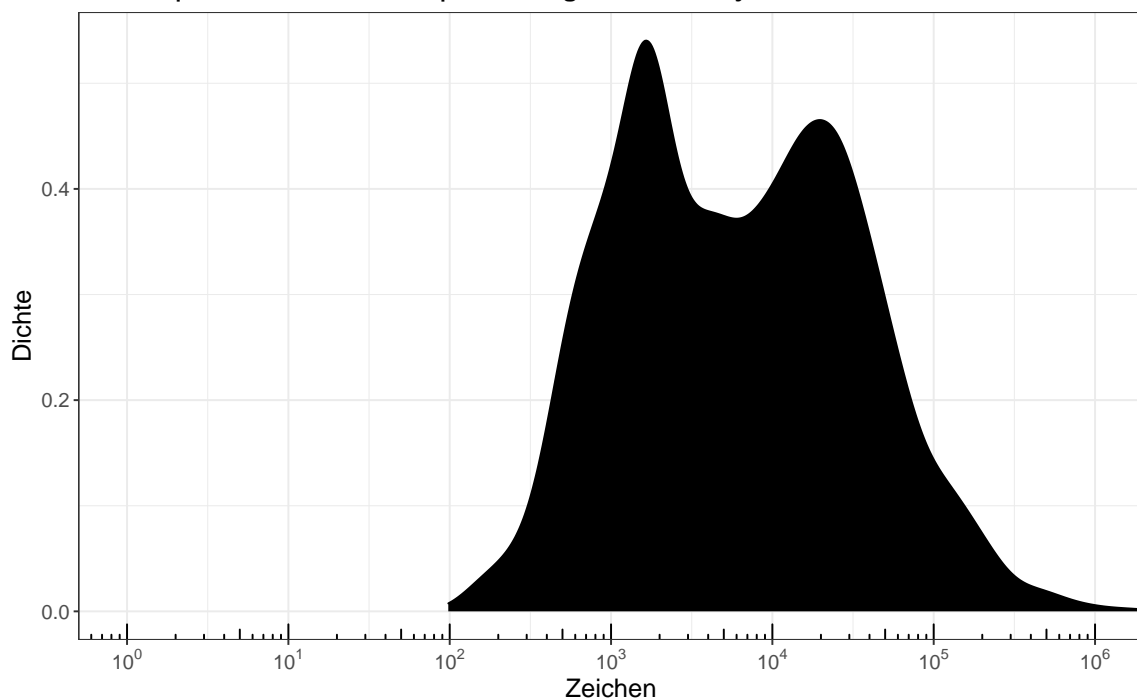
Fobbe | DOI: 10.5281/zenodo.6959374

```

ggplot(data = meta.rechtsakte)+
  geom_density(aes(x = zeichen),
    fill = "black") +
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Zeichen je Rechtsakt"),
    caption = caption,
    x = "Zeichen",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )

```

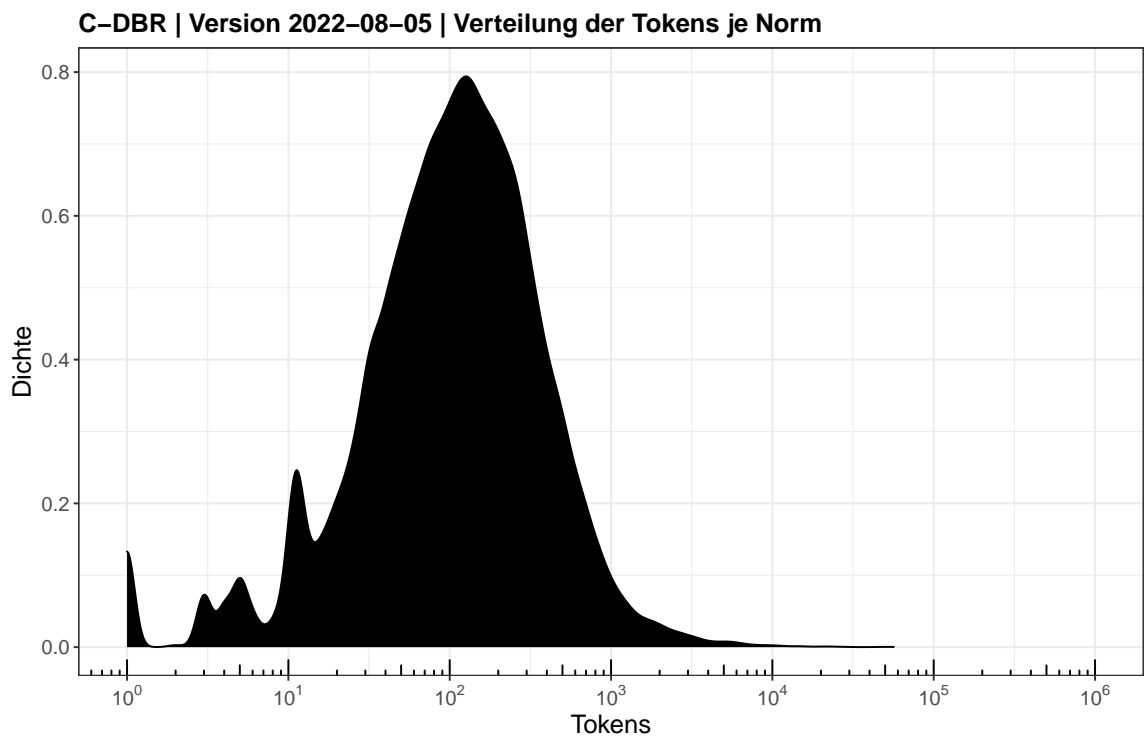
C-DBR | Version 2022-08-05 | Verteilung der Zeichen je Rechtsakt



Fobbe | DOI: 10.5281/zenodo.6959374

9.7.2 Density (Tokens)

```
ggplot(data = meta.normen)+
  geom_density(aes(x = tokens),
    fill = "black") +
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Tokens je Norm"),
    caption = caption,
    x = "Tokens",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```

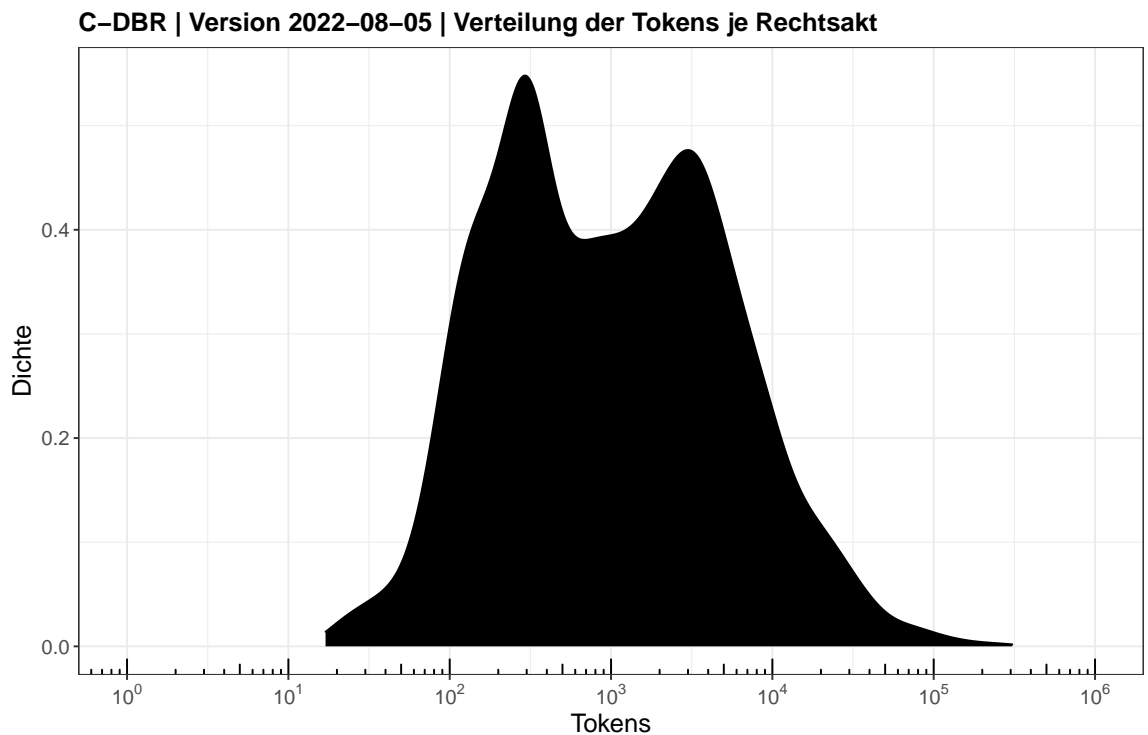


Fobbe | DOI: 10.5281/zenodo.6959374

```

ggplot(data = meta.rechtsakte)+
  geom_density(aes(x = tokens),
    fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw() +
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Tokens je Rechtsakt"),
    caption = caption,
    x = "Tokens",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )

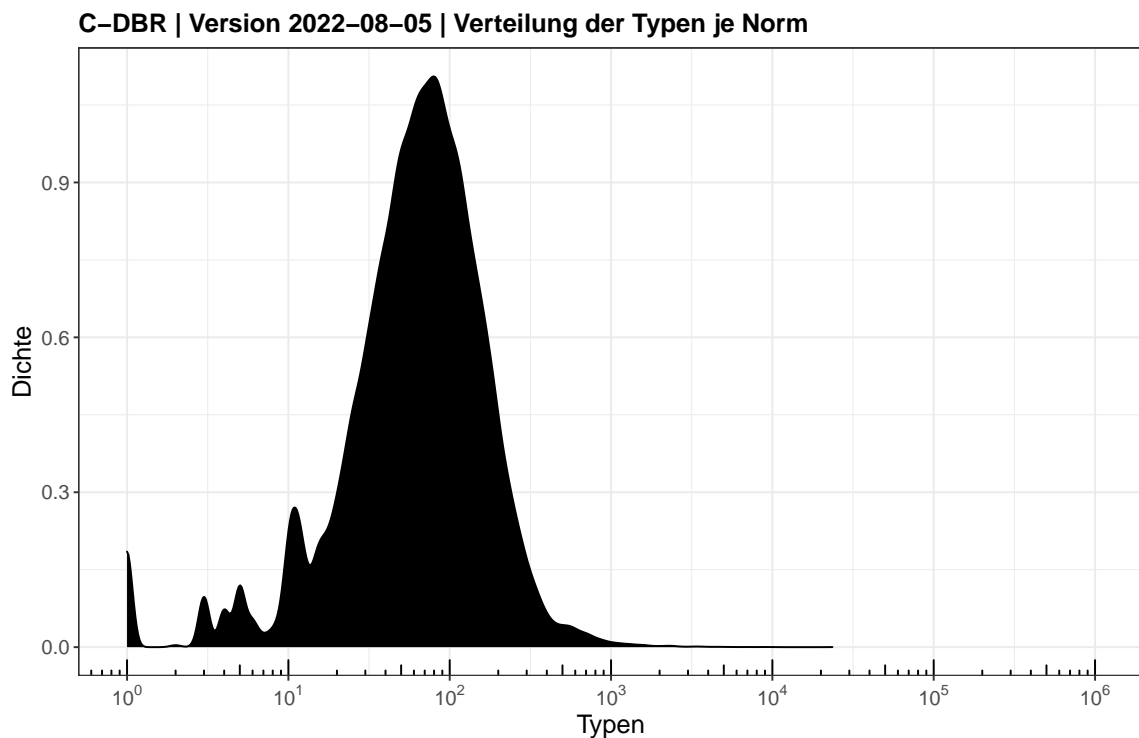
```



Fobbe | DOI: 10.5281/zenodo.6959374

9.7.3 Density (Typen)

```
ggplot(data = meta.normen)+
  geom_density(aes(x = typen),
    fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Typen je Norm"),
    caption = caption,
    x = "Typen",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```

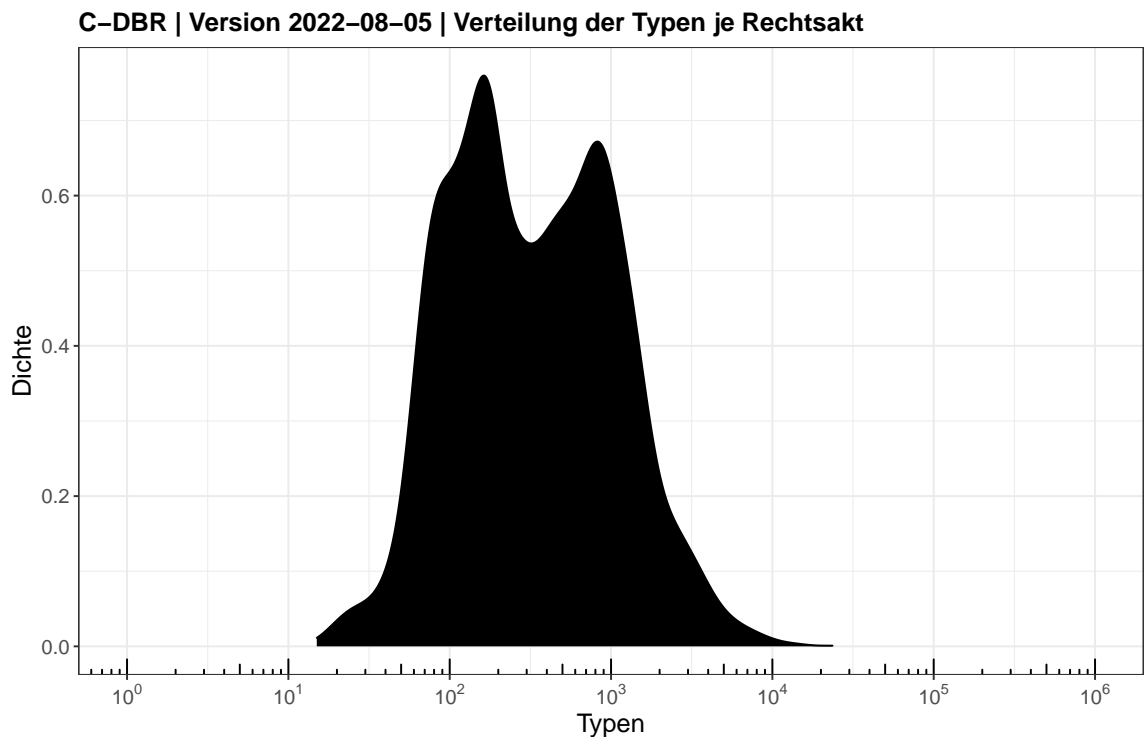


Fobbe | DOI: 10.5281/zenodo.6959374

```

ggplot(data = meta.rechtsakte)+
  geom_density(aes(x = typen),
    fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Typen je Rechtsakt"),
    caption = caption,
    x = "Typen",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )

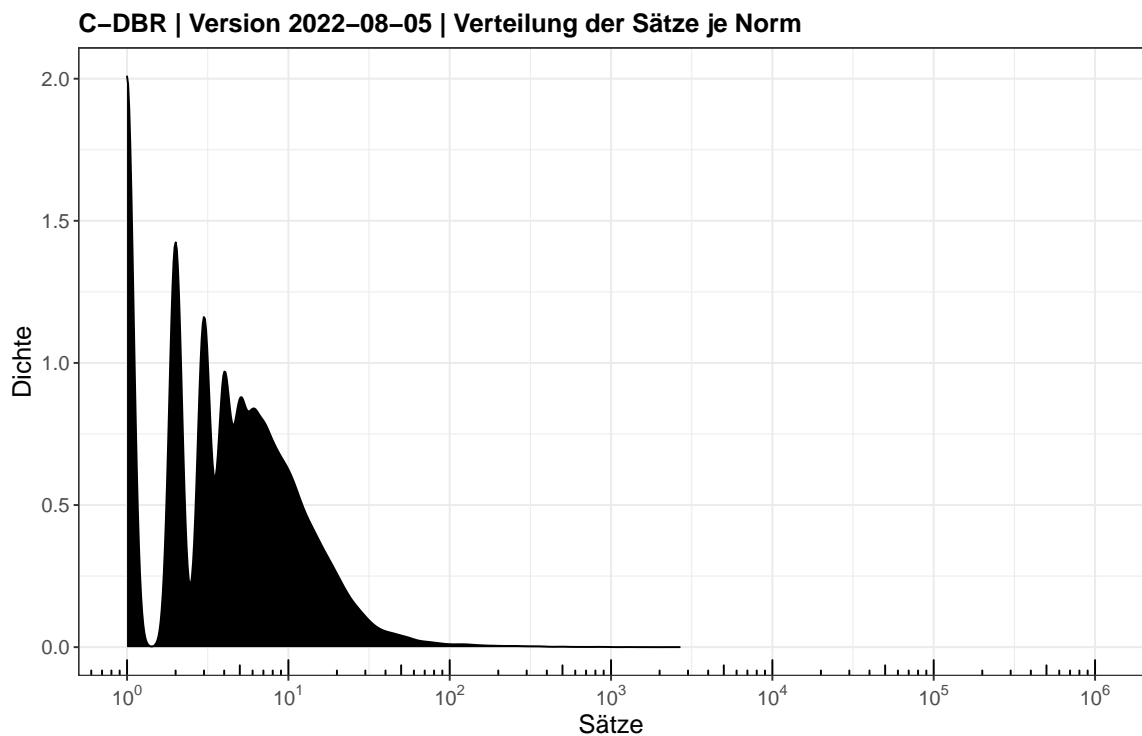
```



Fobbe | DOI: 10.5281/zenodo.6959374

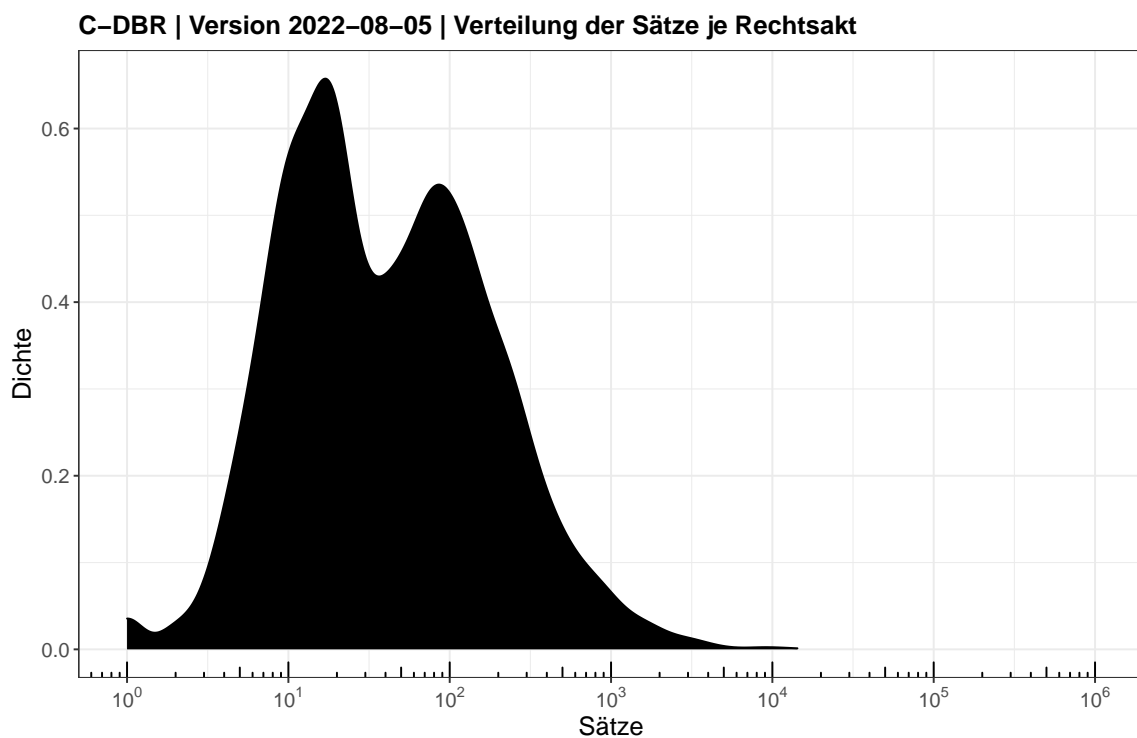
9.7.4 Density (Sätze)

```
ggplot(data = meta.normen)+
  geom_density(aes(x = saetze),
    fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Sätze je Norm"),
    caption = caption,
    x = "Sätze",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```



Fobbe | DOI: 10.5281/zenodo.6959374


```
ggplot(data = meta.rechtsakte)+
  geom_density(aes(x = saetze),
    fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Sätze je Rechtsakt"),
    caption = caption,
    x = "Sätze",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
)
```



Fobbe | DOI: 10.5281/zenodo.6959374

9.8 Quantitative Variablen

9.8.1 Ausfertigungsdatum

Einzelnormen

```
summary(as.IDate(dt.normen$ausfertigung_datum))
```

```
##           Min.         1st Qu.         Median         Mean         3rd Qu.         Max.
## "1869-06-21" "1972-10-04" "1997-12-22" "1988-06-23" "2012-05-02" "2022-07-27"
```

Rechtsakte

```
summary(as.IDate(dt.rechtsakte$ausfertigung_datum))
```

```
##           Min.         1st Qu.         Median         Mean         3rd Qu.         Max.
## "1869-06-21" "1981-12-22" "2001-06-25" "1995-05-15" "2012-12-18" "2022-07-27"
```

XML-Metadaten

```
summary(as.IDate(dt.meta$ausfertigung_datum))
```

```
##           Min.         1st Qu.         Median         Mean         3rd Qu.         Max.
## "1869-06-21" "1978-03-15" "1998-05-13" "1993-06-27" "2011-04-29" "2022-07-27"
##           NA's
##           "1"
```

9.8.2 Ausfertigungsjahr

Einzelnormen

```
summary(dt.normen$ausfertigung_jahr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1869   1972   1997   1988   2012   2022
```

Rechtsakte

```
summary(dt.rechtsakte$ausfertigung_jahr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1869   1981   2001   1995   2012   2022
```

XML-Metadaten

```
summary(dt.meta$ausfertigung_jahr)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	1869	1978	1998	1993	2011	2022	1

10 Strenge Kontrolle der Variablen-Namen

10.1 Semantische Sortierung der Variablen

10.1.1 Variablen sortieren: Einzelnormen

```
setcolorder(dt.normen,  
            c("doc_id",  
              "dateiname",  
              "text",  
              "jurabk",  
              "amtabk",  
              "langue",  
              "kurzue",  
              "gliederungskennzahl",  
              "gliederungsbez",  
              "gliederungstitel",  
              "enbez",  
              "bezketten",  
              "titelketten",  
              "ausfertigung_datum",  
              "ausfertigung_jahr",  
              "aenderung_datum",  
              "aufhebung_verkuendung_datum",  
              "aufhebung_wirkung_datum",  
              "neufassung_datum",  
              "fundstellentyp",  
              "periodikum",  
              "zitstelle",  
              "stand",  
              "aufh",  
              "neuf",  
              "hinweis",  
              "sonst",  
              "check_stand",  
              "check_aufh",  
              "check_neuf",  
              "check_hinweis",  
              "check_sonst",  
              "bulldate_original",  
              "bulldate_iso",  
              "zeichen",  
              "tokens",  
              "typen",  
              "saetze",  
              "doi_concept",  
              "doi_version",  
              "version",  
              "lizenz"))
```

```

setcolorder(meta.normen,
  c("doc_id",
    "dateiname",
    "jurabk",
    "amtabk",
    "langue",
    "kurzue",
    "gliederungskennzahl",
    "gliederungsbez",
    "gliederungstitel",
    "enbez",
    "bezketten",
    "titelkette",
    "ausfertigung_datum",
    "ausfertigung_jahr",
    "aenderung_datum",
    "aufhebung_verkuendung_datum",
    "aufhebung_wirkung_datum",
    "neufassung_datum",
    "fundstellentyp",
    "periodikum",
    "zitstelle",
    "stand",
    "aufh",
    "neuf",
    "hinweis",
    "sonst",
    "check_stand",
    "check_aufh",
    "check_neuf",
    "check_hinweis",
    "check_sonst",
    "bulldate_original",
    "bulldate_iso",
    "zeichen",
    "tokens",
    "typen",
    "saetze",
    "doi_concept",
    "doi_version",
    "version",
    "lizenz"))

```

10.1.2 Variablen sortieren: Rechtsakte

```
setcolorder(dt.rechtsakte,
  c("doc_id",
    "text",
    "jurabk",
    "amtabk",
    "langue",
    "kurzue",
    "ausfertigung_datum",
    "ausfertigung_jahr",
    "aenderung_datum",
    "aufhebung_verkuendung_datum",
    "aufhebung_wirkung_datum",
    "neufassung_datum",
    "fundstellentyp",
    "periodikum",
    "zitstelle",
    "stand",
    "aufh",
    "neuf",
    "hinweis",
    "sonst",
    "check_stand",
    "check_aufh",
    "check_neuf",
    "check_hinweis",
    "check_sonst",
    "zeichen",
    "tokens",
    "typen",
    "saetze",
    "doi_concept",
    "doi_version",
    "version",
    "lizenz")))
```

```
setcolorder(meta.rechtsakte,
  c("doc_id",
    "jurabk",
    "amtabk",
    "langue",
    "kurzue",
    "ausfertigung_datum",
    "ausfertigung_jahr",
    "aenderung_datum",
    "aufhebung_verkuendung_datum",
    "aufhebung_wirkung_datum",
    "neufassung_datum",
    "fundstellentyp",
    "periodikum",
    "zitstelle",
    "stand",
    "aufh",
    "neuf",
    "hinweis",
    "sonst",
    "check_stand",
    "check_aufh",
    "check_neuf",
    "check_hinweis",
    "check_sonst",
    "zeichen",
    "tokens",
    "typen",
    "saetze",
    "doi_concept",
    "doi_version",
    "version",
    "lizenz"))
```

10.1.3 Variablen sortieren: XML-Metadaten

```
setcolorder(dt.meta,
            c("doc_id",
              "jurabk",
              "amtabk",
              "langue",
              "kurzue",
              "ausfertigung_datum",
              "ausfertigung_jahr",
              "aenderung_datum",
              "aufhebung_verkuendung_datum",
              "aufhebung_wirkung_datum",
              "neufassung_datum",
              "fundstellentyp",
              "periodikum",
              "zitstelle",
              "stand",
              "aufh",
              "neuf",
              "hinweis",
              "sonst",
              "check_stand",
              "check_aufh",
              "check_neuf",
              "check_hinweis",
              "check_sonst",
              "bulddate_original",
              "bulddate_iso",
              "doi_concept",
              "doi_version",
              "version",
              "lizenz"))
```

10.2 Anzahl Variablen der Datensätze

```
length(dt.normen)
```

```
## [1] 42
```

```
length(meta.normen)
```

```
## [1] 41
```

```
length(dt.rechtsakte)
```



```
## [1] 33
```

```
length(meta.rechtsakte)
```

```
## [1] 32
```

```
length(dt.meta)
```

```
## [1] 30
```

10.3 Alle Variablen-Namen der Datensätze

```
names(dt.normen)
```

```
## [1] "doc_id" "dateiname"
## [3] "text" "jurabk"
## [5] "amtabk" "langue"
## [7] "kurzue" "gliederungskennzahl"
## [9] "gliederungsbez" "gliederungstitel"
## [11] "enbez" "bezkette"
## [13] "titelkette" "ausfertigung_datum"
## [15] "ausfertigung_jahr" "aenderung_datum"
## [17] "aufhebung_verkuendung_datum" "aufhebung_wirkung_datum"
## [19] "neufassung_datum" "fundstellentyp"
## [21] "periodikum" "zitstelle"
## [23] "stand" "aufh"
## [25] "neuf" "hinweis"
## [27] "sonst" "check_stand"
## [29] "check_aufh" "check_neuf"
## [31] "check_hinweis" "check_sonst"
## [33] "builddate_original" "builddate_iso"
## [35] "zeichen" "tokens"
## [37] "typen" "saetze"
## [39] "doi_concept" "doi_version"
## [41] "version" "lizenz"
```

```
names(meta.normen)
```

## [1]	"doc_id"	"dateiname"
## [3]	"jurabk"	"amtabk"
## [5]	"langue"	"kurzue"
## [7]	"gliederungskennzahl"	"gliederungsbez"
## [9]	"gliederungstitel"	"enbez"
## [11]	"bezkette"	"titelkette"
## [13]	"ausfertigung_datum"	"ausfertigung_jahr"
## [15]	"aenderung_datum"	"aufhebung_verkuendung_datum"
## [17]	"aufhebung_wirkung_datum"	"neufassung_datum"
## [19]	"fundstellentyp"	"periodikum"
## [21]	"zitstelle"	"stand"
## [23]	"aufh"	"neuf"
## [25]	"hinweis"	"sonst"
## [27]	"check_stand"	"check_aufh"
## [29]	"check_neuf"	"check_hinweis"
## [31]	"check_sonst"	"builddate_original"
## [33]	"builddate_iso"	"zeichen"
## [35]	"tokens"	"typen"
## [37]	"saetze"	"doi_concept"
## [39]	"doi_version"	"version"
## [41]	"lizenz"	

```
names(dt.rechtsakte)
```

## [1]	"doc_id"	"text"
## [3]	"jurabk"	"amtabk"
## [5]	"langue"	"kurzue"
## [7]	"ausfertigung_datum"	"ausfertigung_jahr"
## [9]	"aenderung_datum"	"aufhebung_verkuendung_datum"
## [11]	"aufhebung_wirkung_datum"	"neufassung_datum"
## [13]	"fundstellentyp"	"periodikum"
## [15]	"zitstelle"	"stand"
## [17]	"aufh"	"neuf"
## [19]	"hinweis"	"sonst"
## [21]	"check_stand"	"check_aufh"
## [23]	"check_neuf"	"check_hinweis"
## [25]	"check_sonst"	"zeichen"
## [27]	"tokens"	"typen"
## [29]	"saetze"	"doi_concept"
## [31]	"doi_version"	"version"
## [33]	"lizenz"	

```
names(meta.rechtsakte)
```

## [1]	"doc_id"	"jurabk"
## [3]	"amtabk"	"langue"
## [5]	"kurzue"	"ausfertigung_datum"

```
## [7] "ausfertigung_jahr"      "aenderung_datum"
## [9] "aufhebung_verkuendung_datum" "aufhebung_wirkung_datum"
## [11] "neufassung_datum"      "fundstellentyp"
## [13] "periodikum"            "zitstelle"
## [15] "stand"                 "aufh"
## [17] "neuf"                  "hinweis"
## [19] "sonst"                 "check_stand"
## [21] "check_aufh"            "check_neuf"
## [23] "check_hinweis"         "check_sonst"
## [25] "zeichen"               "tokens"
## [27] "typen"                 "saetze"
## [29] "doi_concept"           "doi_version"
## [31] "version"               "lizenz"
```

```
names(dt.meta)
```

```
## [1] "doc_id"                "jurabk"
## [3] "amtabk"                "langue"
## [5] "kurzue"                "ausfertigung_datum"
## [7] "ausfertigung_jahr"     "aenderung_datum"
## [9] "aufhebung_verkuendung_datum" "aufhebung_wirkung_datum"
## [11] "neufassung_datum"     "fundstellentyp"
## [13] "periodikum"           "zitstelle"
## [15] "stand"                "aufh"
## [17] "neuf"                 "hinweis"
## [19] "sonst"                "check_stand"
## [21] "check_aufh"           "check_neuf"
## [23] "check_hinweis"        "check_sonst"
## [25] "bulldate_original"    "bulldate_iso"
## [27] "doi_concept"          "doi_version"
## [29] "version"              "lizenz"
```

11 CSV-Dateien erstellen

11.1 Einzelnormen (Korpus)

11.1.1 Name für CSV definieren

```
csvname.normen.gesamt <- paste0(prefix.files,  
                                "_DE_CSV_Einzelnormen_Datensatz.csv")
```

11.1.2 Datensatz speichern

```
fwrite(dt.normen,  
       paste0(csvname.normen.gesamt),  
       na = "NA")
```

11.2 Einzelnormen (Metadaten)

11.2.1 Name für CSV definieren

```
csvname.normen.meta <- paste0(prefix.files,  
                               "_DE_CSV_Einzelnormen_Metadaten.csv")
```

11.2.2 Datensatz speichern

```
fwrite(meta.normen,  
       paste0(csvname.normen.meta),  
       na = "NA")
```

11.3 Rechtsakte (Korpus)

11.3.1 Name für CSV definieren

```
csvname.rechtsakte.gesamt <- paste0(prefix.files,  
                                    "_DE_CSV_Rechtsakte_Datensatz.csv")
```

11.3.2 Datensatz speichern

```
fwrite(dt.rechtsakte,  
       paste0(csvname.rechtsakte.gesamt),  
       na = "NA")
```

11.4 Rechtsakte (Metadaten)

11.4.1 Name für CSV definieren

```
csvname.rechtsakte.meta <- paste0(prefix.files,  
                                   "_DE_CSV_Rechtsakte_Metadaten.csv")
```

11.4.2 Datensatz speichern

```
fwrite(meta.rechtsakte,  
       paste0(csvname.rechtsakte.meta),  
       na = "NA")
```

11.5 XML-Metadaten

Diese Datei unterscheidet sich von der Variante “DE_CSV_Rechtsakte_Metadaten,” weil sie auch Rechtsakte enthält, die ohne Text veröffentlicht wurden. Die Differenz betrifft etwa 1000 Rechtsakte, ist also erheblich.

11.5.1 Name für CSV definieren

```
csvname.meta <- paste0(prefix.files,  
                       "_DE_CSV_MetadatenXML.csv")
```

11.5.2 Datensatz speichern

```
fwrite(dt.meta,  
       paste0(csvname.meta),  
       na = "NA")
```

12 Download der PDF-Dateien

12.1 Download durchführen

```
if(config$parallel$downloadPDF == TRUE){  
  plan("multicore",  
        workers = fullCores)  
}else{  
  plan("sequential")  
}
```

```
future_mapapply(f.download_robust,  
                url = download$links.pdf,  
                destfile = download$title.pdf)
```

```
## Warning in download.file(url = url, destfile = destfile): downloaded length 0  
!=  
## reported length 196
```

```
## Warning in download.file(url = url, destfile = destfile): cannot open URL  
## 'http://www.gesetze-im-internet.de/kochausbv_2022/NA': HTTP status was '404  
Not  
Found'
```

12.2 Download-Ergebnis

12.2.1 Anzahl herunterzuladender Dateien

```
download[, .N]
```

```
## [1] 6659
```

12.2.2 Anzahl heruntergeladener Dateien

```
files.pdf <- list.files(pattern = "\\..pdf")  
length(files.pdf)
```

```
## [1] 6658
```

12.2.3 Fehlbetrag

```
N.missing <- download[,.N] - length(files.pdf)
print(N.missing)
```

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

12.2.4 Fehlende Dateien

```
missing <- setdiff(download$title.pdf,
                    files.pdf)
print(missing)
```

```
## [1] "NA_Verordnung-über-die-Berufsausbildung-zum-Koch-und-zur-Köchin.pdf"
```

13 TXT-Dateien erstellen

An dieser Stelle wird der reine Text aus den PDF-Dateien extrahiert und ein zusätzliches Datei-Format (TXT) generiert. TXT-Dateien sind besonders für quantitative Analysten ohne XML-Kenntnisse ein lohnenswerter Einstieg und verringern die Hürde für die Arbeit mit dem Korpus.

```
files.pdf <- list.files(pattern = "\\\\.pdf",
                        ignore.case = TRUE)
```

13.1 Anzahl zu extrahierender Dateien

```
length(files.pdf)
```

```
## [1] 6658
```

13.2 Funktion anzeigen: future_pdf_to_txt

```
print(future_pdf_to_txt)
```

```
## function(x){
##
##   ## Timestamp: Begin
##   begin.extract <- Sys.time()
##
##   ## Intro message
##   message(paste("Processing",
##                 length(x),
##                 "files. Begin at:",
##                 begin.extract))
##
##   ## Perform conversion from PDF to TXT
##   invisible(future.apply::future_lapply(x,
##                                         pdf_to_txt,
##                                         future.seed = TRUE))
##
##   ## Construct full list of TXT names
##   txt.names <- gsub("\\\\.pdf",
##                    "\\\\.txt",
##                    x,
##                    ignore.case = TRUE)
##
##   ## Check list of TXT files in folder
##   txt.results <- list.files(pattern = "\\\\.txt")
## }
```



```
##
##   ## Compare full list to files in folder
##   txt.missing <- setdiff(txt.names,
##                           txt.results)
##
##   ## Timestamp: End
##   end.extract <- Sys.time()
##
##   ## Duration
##   duration.extract <- end.extract - begin.extract
##
##   ## Outro message
##   message(paste0("Successfully processed ",
##                   length(x) - length(txt.missing),
##                   " files. ",
##                   length(txt.missing),
##                   " files failed. Runtime was ",
##                   round(duration.extract,
##                           digits = 2),
##                   " ",
##                   attributes(duration.extract)$units,
##                   ". Ended at: ",
##                   end.extract))
##
##
## }
```

13.3 Text Extrahieren

```
if(config$parallel$extractPDF == TRUE){

  plan("multicore",
        workers = fullCores)

}else{

  plan("sequential")

}

future_pdf_to_txt(files.pdf)
```

```
## Processing 6658 files. Begin at: 2022-08-05 22:07:21
```

```
## Successfully processed 6658 files. 0 files failed. Runtime was 18.66 secs.
## Ended at: 2022-08-05 22:07:39
```

14 Download der EPUB-Dateien

14.1 Download durchführen

```
if(config$parallel$downloadEPUB == TRUE){  
  plan("multicore",  
        workers = fullCores)  
}else{  
  plan("sequential")  
}
```

```
future_mapapply(f.download_robust,  
               url = download$links.epub,  
               destfile = download$title.epub)
```

14.2 Download-Ergebnis

14.2.1 Anzahl herunterzuladender Dateien

```
download[,.N]
```

```
## [1] 6659
```

14.2.2 Anzahl heruntergeladener Dateien

```
files.epub <- list.files(pattern = "\\..epub")  
length(files.epub)
```

```
## [1] 6659
```

14.2.3 Fehlbetrag

```
N.missing <- download[,.N] - length(files.epub)  
print(N.missing)
```

```
## [1] 0
```

14.2.4 Fehlende Dateien

```
missing <- setdiff(download$title.epub, files.epub)
print(missing)
```

```
## character(0)
```

15 Dateigrößen analysieren

```
files.txt <- list.files(pattern = "\\..txt$",  
                        ignore.case = TRUE)  
  
files.pdf <- list.files(pattern = "\\..pdf$",  
                        ignore.case = TRUE)  
  
files.epub <- list.files(pattern = "\\..epub$",  
                         ignore.case = TRUE)  
  
txt.MB <- file.size(files.txt) / 10^6  
pdf.MB <- file.size(files.pdf) / 10^6  
epub.MB <- file.size(files.epub) / 10^6
```

15.1 Gesamtgröße

15.1.1 PDF-Dateien (MB)

```
sum(pdf.MB)
```

```
## [1] 643.7744
```

15.1.2 EPUB-Dateien (MB)

```
sum(epub.MB)
```

```
## [1] 449.8922
```

15.1.3 XML-Dateien (MB)

```
sum(xml.MB)
```

```
## [1] 269.1641
```

15.1.4 TXT-Dateien (MB)

```
sum(txt.MB)
```

```
## [1] 202.6076
```

15.1.5 Objekte in RAM (MB)

```
print(object.size(dt.normen),  
      standard = "SI",  
      humanReadable = TRUE,  
      units = "MB")
```

```
## 197 MB
```

```
print(object.size(dt.rechtsakte),  
      standard = "SI",  
      humanReadable = TRUE,  
      units = "MB")
```

```
## 151.9 MB
```

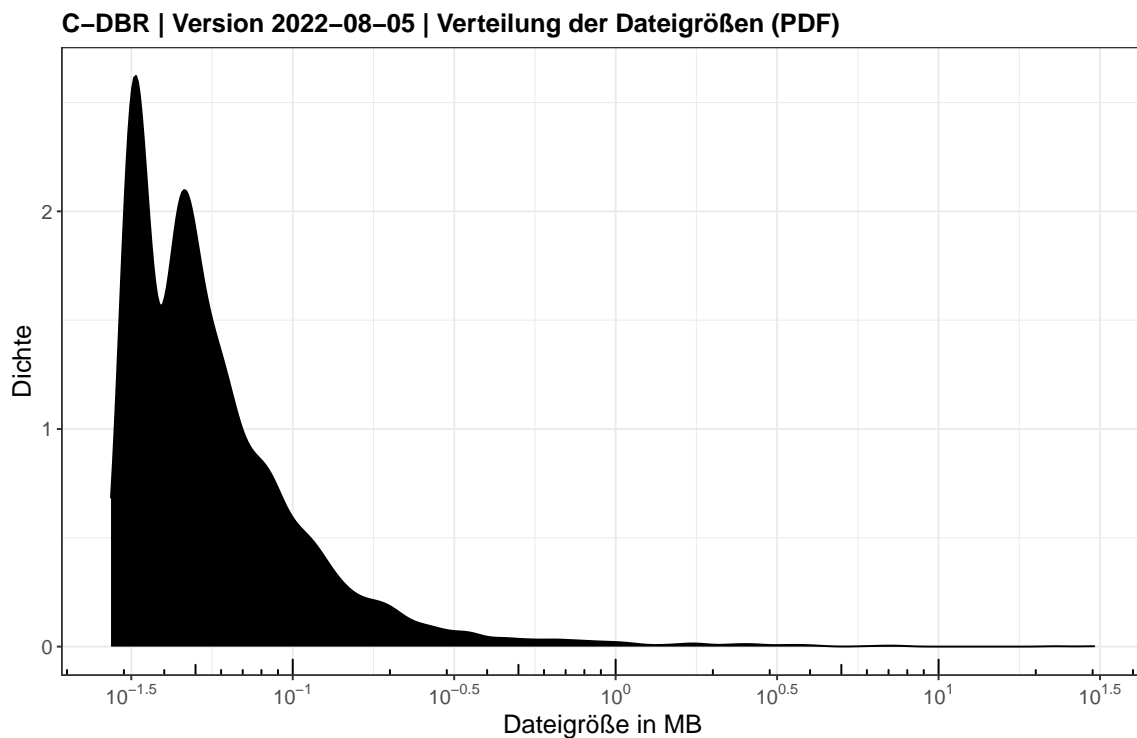
```
print(object.size(dt.meta),  
      standard = "SI",  
      humanReadable = TRUE,  
      units = "MB")
```

```
## 5.7 MB
```

15.2 Verteilung der Dateigrößen (PDF)

```
dt.plot <- data.table(pdf.MB)
```

```
ggplot(data = dt.plot,
       aes(x = pdf.MB)) +
  geom_density(fill = "black") +
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
               labels = trans_format("log10", math_format(10^.x))) +
  annotation_logticks(sides = "b") +
  theme_bw() +
  labs(
    title = paste(prefix.figuretitle,
                  "| Verteilung der Dateigrößen (PDF)"),
    caption = caption,
    x = "Dateigröße in MB",
    y = "Dichte"
  ) +
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
                              face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```

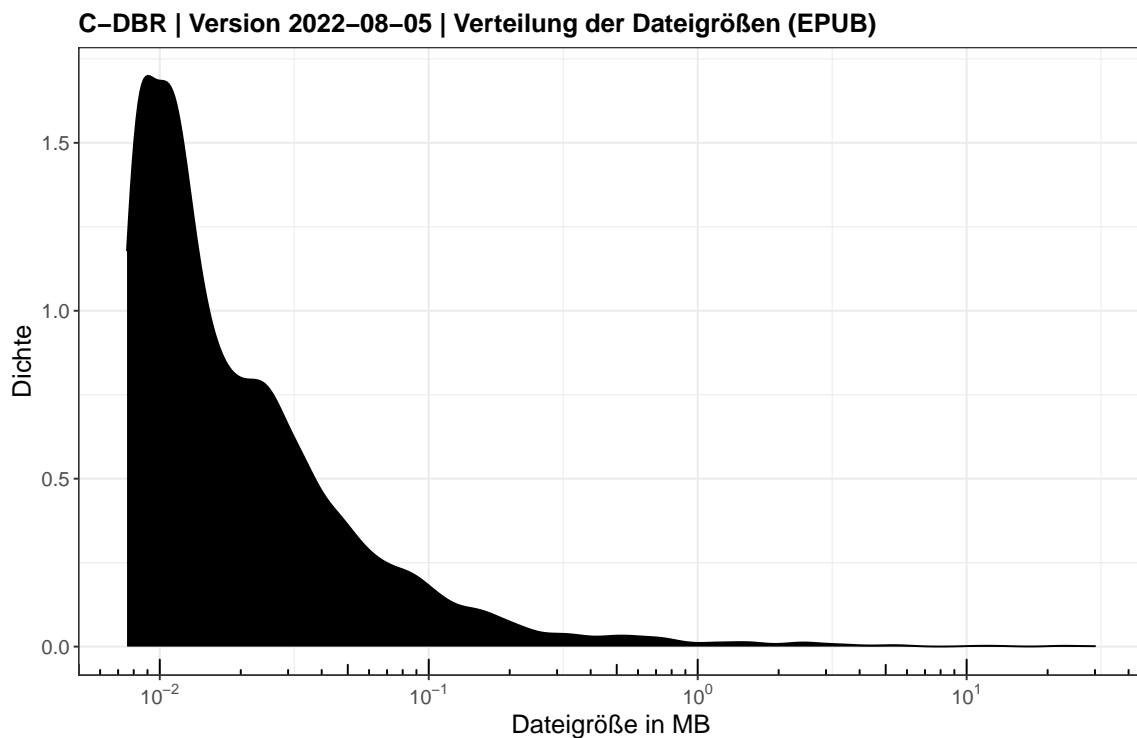


Fobbe | DOI: 10.5281/zenodo.6959374

15.3 Verteilung der Dateigrößen (EPUB)

```
dt.plot <- data.table(epub.MB)
```

```
ggplot(data = dt.plot,
  aes(x = epub.MB))+
  geom_density(fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Dateigrößen (EPUB)"),
    caption = caption,
    x = "Dateigröße in MB",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```

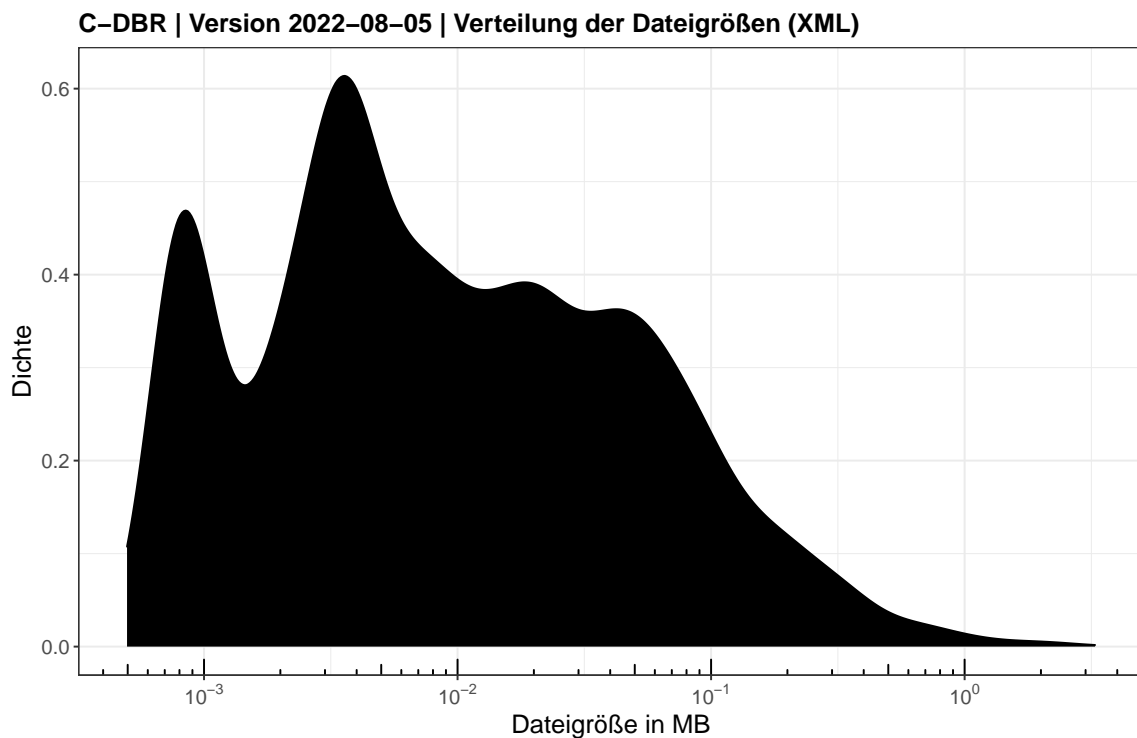


Fobbe | DOI: 10.5281/zenodo.6959374

15.4 Verteilung der Dateigrößen (XML)

```
dt.plot <- data.table(xml.MB)
```

```
ggplot(data = dt.plot,
  aes(x = xml.MB))+
  geom_density(fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Dateigrößen (XML)"),
    caption = caption,
    x = "Dateigröße in MB",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```

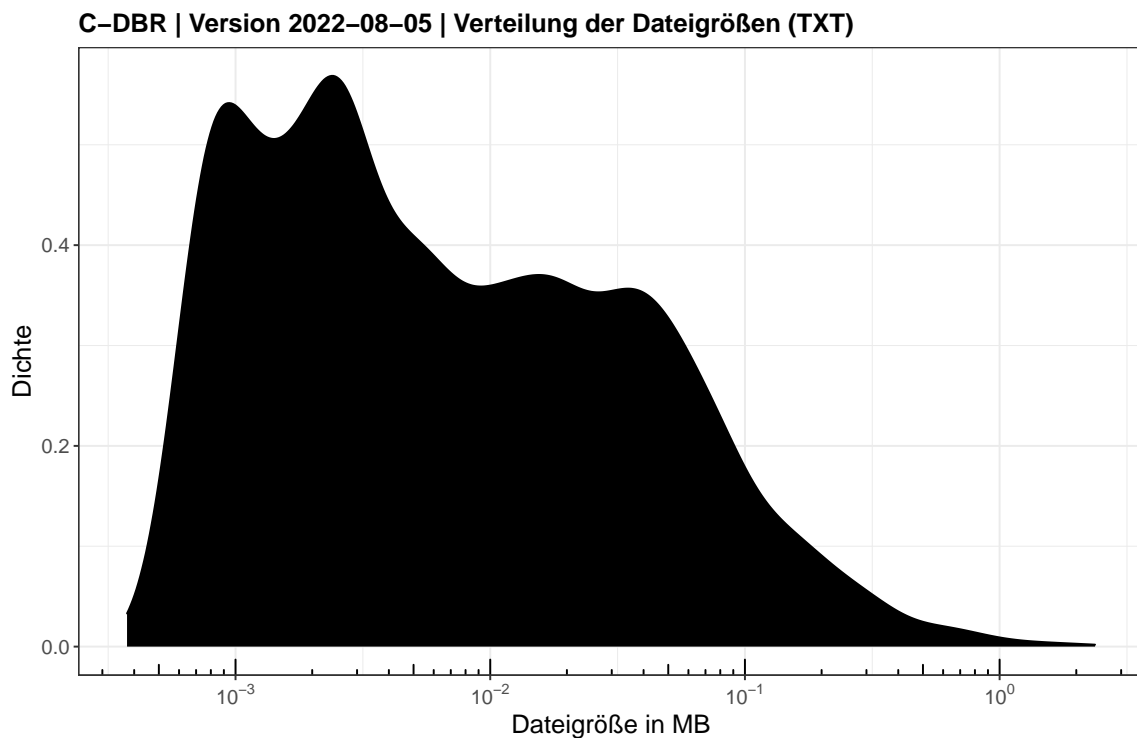


Fobbe | DOI: 10.5281/zenodo.6959374

15.5 Verteilung der Dateigrößen (TXT)

```
dt.plot <- data.table(txt.MB)
```

```
ggplot(data = dt.plot,
  aes(x = txt.MB))+
  geom_density(fill = "black")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
      "| Verteilung der Dateigrößen (TXT)"),
    caption = caption,
    x = "Dateigröße in MB",
    y = "Dichte"
  )+
  theme(
    text = element_text(size = 14),
    plot.title = element_text(size = 14,
      face = "bold"),
    legend.position = "none",
    plot.margin = margin(10, 20, 10, 10)
  )
```



Fobbe | DOI: 10.5281/zenodo.6959374

16 ZIP-Archive erstellen

16.1 Verpacken der CSV-Dateien

```
files.csv <- c(csvname.normen.gesamt,  
              csvname.normen.meta,  
              csvname.rechtsakte.gesamt,  
              csvname.rechtsakte.meta,  
              csvname.meta)  
  
csvnames.zip <- gsub(".csv",  
                    ".zip",  
                    files.csv)  
  
for (i in seq_along(files.csv)){  
  zip(csvnames.zip[i],  
      files.csv[i],  
      mode = "cherry-pick")  
}  
  
unlink(files.csv)
```

16.2 Verpacken der PDF-Dateien

```
files.pdf <- list.files(pattern = "\\..pdf$",  
                        ignore.case = TRUE)  
  
zip(paste0(prefix.files,  
            "_DE_PDF_Datensatz.zip"),  
    files.pdf,  
    mode = "cherry-pick")
```

16.3 Verpacken der TXT-Dateien

```
files.txt <- list.files(pattern = "\\..txt$",  
                        ignore.case = TRUE)  
  
zip(paste0(prefix.files,  
            "_DE_TXT_Datensatz.zip"),  
    files.txt,  
    mode = "cherry-pick")
```

16.4 Verpacken der EPUB-Dateien

```
files.epub <- list.files(pattern = "\\..epub$",
                        ignore.case = TRUE)

zip(paste0(prefix.files,
           "_DE_EPUB_Datensatz.zip"),
    files.epub,
    mode = "cherry-pick")
```

16.5 Verpacken der Netzwerk-Dateien

```
zip(paste0(prefix.files,
           "_DE_Netzwerke.zip"),
    "netzwerke",
    mode = "cherry-pick")
```

16.6 Verpacken der Analyse-Dateien

```
zip(paste0(prefix.files,
           "_DE_",
           toupper(basename(dir.analysis)),
           ".zip"),
    basename(dir.analysis),
    mode = "cherry-pick")
```

16.7 Verpacken der Source-Dateien

```
files.source <- c(list.files(pattern = "\\..R$|\\.toml$|\\.md$",
                            "R-fobbe-proto-package",
                            "functions",
                            "tex",
                            "gpg",
                            "buttons",
                            list.files(pattern = "renv\\.lock|\\.Rprofile",
                                        all.files = TRUE),
                            list.files("renv",
                                        pattern = "activate\\.R",
                                        full.names = TRUE)))

files.source <- grep("spin",
                    files.source,
                    value = TRUE,
                    ignore.case = TRUE,
                    invert = TRUE)

zip(paste0(prefix.files,
```

```
        "_Source_Code.zip"),  
files.source)
```

17 Roh-Dateien löschen

```
files.delete <- list.files(pattern = "\\..xml|\\..jpe?g|\\..png|\\..gif|\\..pdf|\\..txt  
|\\..epub",  
                           ignore.case = TRUE)  
  
unlink(files.delete)  
unlink("netzwerke", recursive = TRUE)  
unlink("Rplots.pdf", recursive = TRUE)
```

18 Kryptographische Hashes

Dieses Modul berechnet für jedes ZIP-Archiv zwei Arten von Hashes: SHA2-256 und SHA3-512. Mit diesen kann die Authentizität der Dateien geprüft werden und es wird dokumentiert, dass sie aus diesem Source Code hervorgegangen sind. Die SHA-2 und SHA-3 Algorithmen gelten derzeit als sicher und ein SHA3-Hash mit 512 bit Länge ist nach derzeitigem Wissen auch gegenüber quantenkryptoanalytischen Verfahren hinreichend resistent.

18.1 Liste der ZIP-Archive erstellen

```
files.zip <- list.files(pattern = "\\\\.zip$",  
                        ignore.case = TRUE)
```

18.2 Funktion anzeigen: future_multihashes

```
print(future_multihashes)
```

```
## function(x){  
##  
##   ## Timestamp: Begin  
##   begin <- Sys.time()  
##  
##   ## Intro Message  
##   message(paste("Processing",  
##                 length(x),  
##                 "files. Begin at:",  
##                 begin))  
##  
##   ## Compute Hashes  
##   hashes.list <- future.apply::future_lapply(x,  
##                                             multihashes)  
##  
##   ## Coerce List to data.table  
##   hashes.table <- data.table::rbindlist(hashes.list)  
##  
##   ## Coerce data.table to data.frame  
##   data.table::setDF(hashes.table)  
##  
##   ## Timestamp: End  
##   end <- Sys.time()  
##  
##   ## Duration  
##   duration <- end - begin  
##  
##   ## Result Message  
##   message(paste0("Processed ",
```

```
##           length(x),
##           " files. Runtime was ",
##           round(duration,
##             digits = 2),
##           " ",
##           attributes(duration)$units,
##           ".")
##
##   return(hashes.table)
##
## }
```

18.3 Hashes berechnen

```
if(config$parallel$multihashes == TRUE){
  plan("multicore",
        workers = fullCores)
}else{
  plan("sequential")
}

multihashes <- future_multihashes(files.zip)
```

```
## Processing 13 files. Begin at: 2022-08-05 22:10:43
```

```
## Processed 13 files. Runtime was 3.18 secs.
```

18.4 In Data Table umwandeln

```
setDT(multihashes)

setnames(multihashes,
         old = "x",
         new = "filename")
```

18.5 Index hinzufügen

```
multihashes$index <- seq_len(multihashes[,.N])
```

18.6 Hashes in CSV-Datei speichern

```
fwrite(multihashes,  
      paste0("output/",  
            prefix.files,  
            "_KryptographischeHashes.csv"),  
      na = "NA")
```

18.7 Leerzeichen hinzufügen um Zeilenumbruch zu ermöglichen

```
multihashes$sha3.512 <- paste(substr(multihashes$sha3.512, 1, 64),  
                             substr(multihashes$sha3.512, 65, 128))
```


18.8 In Bericht anzeigen

```
kable(multihashes[,.(index,filename)],  
      format = "latex",  
      align = c("p{1cm}", "p{13cm}"),  
      booktabs = TRUE,  
      longtable = TRUE)
```

index	filename
1	C-DBR_2022-08-05_DE_ANALYSE.zip
2	C-DBR_2022-08-05_DE_CSV_Einzelnormen_Datensatz.zip
3	C-DBR_2022-08-05_DE_CSV_Einzelnormen_Metadaten.zip
4	C-DBR_2022-08-05_DE_CSV_MetadatenXML.zip
5	C-DBR_2022-08-05_DE_CSV_Rechtsakte_Datensatz.zip
6	C-DBR_2022-08-05_DE_CSV_Rechtsakte_Metadaten.zip
7	C-DBR_2022-08-05_DE_EPUB_Datensatz.zip
8	C-DBR_2022-08-05_DE_Netzwerke.zip
9	C-DBR_2022-08-05_DE_PDF_Datensatz.zip
10	C-DBR_2022-08-05_DE_TXT_Datensatz.zip
11	C-DBR_2022-08-05_DE_XML_Anlagen.zip
12	C-DBR_2022-08-05_DE_XML_Datensatz.zip
13	C-DBR_2022-08-05_Source_Code.zip

```
kable(multihashes[,.(index,sha2.256)],
      format = "latex",
      align = c("c", "p{13cm}"),
      booktabs = TRUE,
      longtable = TRUE)
```

index	sha2.256
1	421e41291280c48b70e99613f5b4b46fbfa0251e314a6df88b475e83a3bda9e2
2	7885de723787d7e009dff6a787457102c754dd965f320db22bcf6d98c4ace209
3	59d72a842ec631b51111e6af5fdf03bfe9bc7dd5601156fb0e6577c4b5ba0684
4	3c3038ae09ee2a86ff0306b3e93872a0cd1e2faf250f6c6b7c8db4d2f24277e8
5	c917656f3168495157381a50b99669c002e3b51752c4b60492e1ba7944a09d6c
6	75cf987e1d1a4277e62ad094144be35b57a221b712c1efb75251fade0d4c044d
7	f30c812ebed900318e060b488b3998175b76f1534c737d90c8e2772bf8bf58f3
8	2752b26c7c27efde1e0e87a2bf6a38ceb9cb7d3819a547b16f974b592e09f043
9	a42ea6af1b5c4a5ad5dba24d1bddf58f830de21bd4d7354e5186074cdab9248a
10	5547d458bb94868189eea8200f63de568bb6bd3d64b455ae5ada1930b5ef2021
11	8ec95ceb778e3d928d98b12255a45cd9fb5e2d280b3a007402d1cc72c33fa063
12	5f372118f65f7bdf830863bbddb76e47ddd6f85ea2d21312c702f7e12ef653
13	2d435079958bd3deafdd0fd896fdf203ccebe86159822d306c0a0cd38a9743f6

```
kable(multihashes[,.(index,sha3.512)],
      format = "latex",
      align = c("c", "p{13cm}"),
      booktabs = TRUE,
      longtable = TRUE)
```

index	sha3.512
1	1ddc614b1e98e68995796d61be3adb7ec5296efde0745ff1c491586987a7c575ba801f44e80ea055ac3bd5077484aa9eefd9d2bb8c55c3d311caa94291e3fc4b
2	6e94935639b1ea97885884cdc09275915d3cbae8952a21d18258908a716cee6bc983ef7446735eab570385892637c94037ab87f20ebb9bd7ff545ce3d3408525
3	7482d60016ba86e3d7a56aa0cddc4ba539e5888d2078966c984d3af95c87f28b563619c32fbd81b9f7fdbaf012652147b7d9b2938b7dd69f91d7f7e6266e249f
4	03f3dab75a37aef5bd347b335a1559c79116584e221acaea07be42d5f0621da753898dd941c9fc4408d1626e3c2e32075cc94ae5df38750a6febc3b307fdb9e8
5	3cc4dd41585d18c0959375dfcdd5c267d998581d05ef66800a1ba3eb72e92d1109f44077788f7eb04fa9e6a44beb3f1fd2425a38b09dbe20dcfbfc7f82b94621
6	971b39eca3a6f4d927f7a0ea8e76d49f9289e0e09176457b34adf81fa968becbe8dce4946c187e09ea4bffaed2a216336db881d8fd42796ac522d80e4d309fcf
7	ace5f4aaccfb2909f58ed8b61edeacce83e381406df9af109f40bfc7f0c47a56245ef626537141b8a2be957c8b962e4ef09bfa521baef81d71c8708f309e6493
8	21e3150de6c0e6ec3765de11be9810ca8cf8498de1e54c684fa4c237c11938f8f88534872ec18b6c084ff8b82d070ff27b3ed021e2095783efba07fd704f6012
9	c143d06f801c393f63d9600f25df5334b8d159dd42c310e2cb209ea939d3698d86a2df61ad94d8a0c03fee80894f6190c8c1d9d505db8017e674e2c2d6d6d707
10	99331428517fc8571566a46aa259b19f1da6d5785faf171e2405e505a5ceedd8ca179449cdd09e048e5a1b8ef44d27f132b3b8db2d1f185ae9104c3c0fb8a8f9
11	2fa275ec0d84b06ea55c90ed2d368b5f9dbd512ba43707c6873209aba6e3ea29681aa9d68b6989b77ac223b967110f1f89b62e91284e6c05c671bfce3e741717
12	d484e8ce8b36c5b283493d3379ee16dce634f06a5ae8c4986f20c38448777ab681cf42e2f3d2deb7fb0a11142fc281151a20de077060bee847c9eec19a52d82a
13	c551ffe01bd354e8d8503157ca5c751566a75b485aed5dcb0b80f691c64f192b622296b291a8ad742ac93351f7bf166ad17a006cea0fd72ffade92fbc1c6c0c6

19 Aufräumen

```
cleanup <- list.files(pattern = "\\\\.zip")

cleanup.destination <- file.path("output",
                                cleanup)

print(cleanup)
```

```
## [1] "C-DBR_2022-08-05_DE_ANALYSE.zip"
## [2] "C-DBR_2022-08-05_DE_CSV_Einzelnormen_Datensatz.zip"
## [3] "C-DBR_2022-08-05_DE_CSV_Einzelnormen_Metadaten.zip"
## [4] "C-DBR_2022-08-05_DE_CSV_MetadatenXML.zip"
## [5] "C-DBR_2022-08-05_DE_CSV_Rechtsakte_Datensatz.zip"
## [6] "C-DBR_2022-08-05_DE_CSV_Rechtsakte_Metadaten.zip"
## [7] "C-DBR_2022-08-05_DE_EPUB_Datensatz.zip"
## [8] "C-DBR_2022-08-05_DE_Netzwerke.zip"
## [9] "C-DBR_2022-08-05_DE_PDF_Datensatz.zip"
## [10] "C-DBR_2022-08-05_DE_TXT_Datensatz.zip"
## [11] "C-DBR_2022-08-05_DE_XML_Anlagen.zip"
## [12] "C-DBR_2022-08-05_DE_XML_Datensatz.zip"
## [13] "C-DBR_2022-08-05_Source_Code.zip"
```

```
file.rename(cleanup,
            cleanup.destination)
```

```
## [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

20 Abschluss

20.1 Datumsstempel

```
print(datestamp)
```

```
## [1] "2022-08-05"
```

20.2 Datum und Uhrzeit (Anfang)

```
print(begin.script)
```

```
## [1] "2022-08-05 21:45:43 CEST"
```

20.3 Datum und Uhrzeit (Ende)

```
end.script <- Sys.time()  
print(end.script)
```

```
## [1] "2022-08-05 22:10:46 CEST"
```

20.4 Laufzeit des gesamten Skripts

```
print(end.script - begin.script)
```

```
## Time difference of 25.04827 mins
```

20.5 Warnungen

```
warnings()
```

21 Parameter für strenge Replikationen

```
system2("openssl", "version", stdout = TRUE)
```

```
## [1] "OpenSSL 3.0.5 5 Jul 2022 (Library: OpenSSL 3.0.5 5 Jul 2022)"
```

```
sessionInfo()
```

```
## R version 4.1.3 (2022-03-10)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora Linux 36 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libflexiblas.so.3.2
##
## locale:
##  [1] LC_CTYPE=en_US.utf8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.utf8      LC_COLLATE=en_US.utf8
##  [5] LC_MONETARY=en_US.utf8  LC_MESSAGES=en_US.utf8
##  [7] LC_PAPER=en_US.utf8     LC_NAME=C
##  [9] LC_ADDRESS=C            LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.utf8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats      graphics grDevices utils      datasets methods
## [8] base
##
## other attached packages:
##  [1] future.apply_1.9.0 future_1.27.0      ggraph_2.0.5      igraph_1.3.4
##  [5] openssl_2.0.2      scales_1.2.0      quanteda_3.2.1    data.table
##  [9] ggplot2_3.3.6      pdftools_3.3.0    magick_2.7.3      kableExtra_1.3.4
## [13] knitr_1.39         RcppTOML_0.1.7    xml2_1.3.3        rvest_1.0.2
## [17] zip_2.2.0
##
## loaded via a namespace (and not attached):
##  [1] viridis_0.6.2      httr_1.4.3        tidyr_1.2.0       tidygraph_1.2.1
##  [5] viridisLite_0.4.0 RcppParallel_5.1.5 askpass_1.1       highr_0.9
##  [9] selectr_0.4-2      renv_0.15.5       yaml_2.3.5        ggrepel_0.9.1
## [13] globals_0.16.0     qpdf_1.2.0        pillar_1.8.0      lattice_0.20-45
## [17] glue_1.6.2         digest_0.6.29     polyclip_1.10-0   colorspace_2.0-3
## [21] htmltools_0.5.3    Matrix_1.4-1      pkgconfig_2.0.3   listenv_0.8.0
## [25] purrr_0.3.4        webshot_0.5.3     svglite_2.1.0     tweenr_1.0.2
## [29] ggforce_0.3.3      tibble_3.1.8      generics_0.1.3    farver_2.1.1
## [33] withr_2.5.0        cli_3.3.0         magrittr_2.0.3    evaluate_0.15
## [37] stopwords_2.3      fansi_1.0.3       parallelly_1.32.1 MASS_7.3-58.1
## [41] tools_4.1.3        lifecycle_1.0.1   stringr_1.4.0     munsell_0.5.0
## [45] compiler_4.1.3     systemfonts_1.0.4 rlang_1.0.4       grid_4.1.3
## [49] rstudioapi_0.13    labeling_0.4.2    rmarkdown_2.14    gtable_0.3.0
```

## [53]	codetools_0.2-18	curl_4.3.2	graphlayouts_0.8.0	R6_2.5.1
## [57]	gridExtra_2.3	dplyr_1.0.9	fastmap_1.1.0	utf8_1.2.2
## [61]	fastmatch_1.1-3	stringi_1.7.8	Rcpp_1.0.9	vctrs_0.4.1
## [65]	tidyselect_1.1.2	xfun_0.31		

Literaturverzeichnis

- Bengtsson, Henrik. 2021. "A Unifying Framework for Parallel and Distributed Processing in r Using Futures." *The R Journal* 13 (2): 208–27. <https://doi.org/10.32614/RJ-2021-048>.
- . 2022a. *Future: Unified Parallel and Distributed Processing in r for Everyone*. <https://CRAN.R-project.org/package=future>.
- . 2022b. *Future.apply: Apply Function to Elements in Parallel Using Futures*. <https://CRAN.R-project.org/package=future.apply>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "Quanteda: An r Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3 (30): 774. <https://doi.org/10.21105/joss.00774>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe. 2022. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal Complex Systems*: 1695. <https://igraph.org>.
- Csárdi, Gábor, Kuba Podgórski, and Rich Geldreich. 2021. *Zip: Cross-Platform Zip Compression*. <https://github.com/r-lib/zip#readme>.
- Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of 'Data.frame'*. <https://CRAN.R-project.org/package=data.table>.
- Eddelbuettel, Dirk. 2020. *RcppTOML: Rcpp Bindings to Parser for Tom's Obvious Markup Language*. <http://dirk.eddelbuettel.com/code/rcpp.toml.html>.
- file., See AUTHORS. 2022. *Igraph: Network Analysis and Visualization*. <https://CRAN.R-project.org/package=igraph>.
- Ooms, Jeroen. 2021. *Magick: Advanced Graphics and Image-Processing in r*. <https://CRAN.R-project.org/package=magick>.
- . 2022a. *Openssl: Toolkit for Encryption, Signatures and Certificates Based on OpenSSL*. <https://github.com/jeroen/openssl>.
- . 2022b. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- Pedersen, Thomas Lin. 2021. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. <https://CRAN.R-project.org/package=ggraph>.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2021. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2022. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2021. *Xml2: Parse XML*. <https://CRAN.R-project.org/package=xml2>.
- Wickham, Hadley, and Dana Seidel. 2022. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/i>

- sbn/9781466561595.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2021. *kableExtra: Construct Complex Table with Kable and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.