

**Corpus der Entscheidungen
des
Bundesverfassungsgerichts
(CE-BVerfG-Source)**

COMPILATION REPORT

Version 2022-02-01

License MIT-0

DOI: 10.5281/zenodo.5910155

Titel	Source Code des »Corpus der Entscheidungen des Bundesverfassungsgerichts«
Abkürzung	CE-BVerfG-Source
Autor	Seán Fobbe
Version	2022-02-01
Download	https://doi.org/10.5281/zenodo.5910155
Lizenz	MIT No Attribution (MIT-0)

Zitiervorschlag

Seán Fobbe (2022). Source Code des »Corpus der Entscheidungen des Bundesverfassungsgerichts« (CE-BVerfG-Source). Version 2022-02-01. Zenodo. DOI: 10.5281/zenodo.5910155.

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 2022-02-01. 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.4308216. 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 der Entscheidungen des Bundesverfassungsgerichts	9
1.1	Überblick	9
1.2	Funktionsweise	9
1.3	Kompilierung	9
1.4	Systemanforderungen	10
1.4.1	Betriebssystem	10
1.4.2	Software	10
1.4.3	Parallelisierung	10
1.4.4	Speicherplatz	10
1.5	Weitere Open Access Veröffentlichungen (Fobbe)	10
1.6	Kontakt	10
2	Vorbereitung	11
2.1	Datumsstempel	11
2.2	Datum und Uhrzeit (Beginn)	11
2.3	Packages Laden	11
2.4	Zusätzliche Funktionen einlesen	12
2.5	Verzeichnis für Analyse-Ergebnisse und Diagramme definieren	13
2.6	Weitere Verzeichnisse definieren	13
2.7	Dateien aus vorherigen Runs bereinigen	13
2.8	Verzeichnisse anlegen	13
2.9	Vollzitate statistischer Software schreiben	13
2.10	Allgemeine Konfiguration	14
2.10.1	Konfiguration einlesen	14
2.10.2	Konfiguration anzeigen	14
2.10.3	Knitr Optionen setzen	15
2.10.4	Download Timeout setzen	15
2.10.5	Quellenangabe für Diagramme definieren	15
2.10.6	Präfix für Dateien definieren	15
2.10.7	Präfix für Diagramme definieren	15
2.10.8	Quanteda-Optionen setzen	16
2.11	LaTeX Konfiguration	16
2.11.1	LaTeX Parameter definieren	16
2.11.2	LaTeX Parameter schreiben	17
2.12	Parallelisierung aktivieren	17
2.12.1	Anzahl logischer Kerne festlegen	17
2.12.2	Quanteda	18
2.12.3	Data.table	18
3	Download: Weitere Datensätze	19
3.1	Registerzeichen und Verfahrensarten	19
3.2	Personendaten zu Präsident:innen	19
3.3	Personendaten zu Vize-Präsident:innen	19
4	Download vorbereiten: Alle Entscheidungen des BVerfG	20
4.1	Zeitstempel: Linksammlung Beginn	20
4.2	Funktion zeigen	20
4.3	Maximale Seitenzahl auslesen	20

4.4	Auszuwertende Seiten	21
4.5	[Debugging Modus] Reduzierung der Seitenzahl	21
4.6	Linkliste erstellen: Erstversuch	21
4.7	Linkliste erstellen: Zweitversuch	23
4.7.1	Seiten für Zweitversuch definieren	23
4.7.2	Anzeigen der Seiten die noch einmal geprüft werden.	23
4.7.3	Seiten prüfen	23
4.7.4	Anzeigen der Links die beim Wiederholungsversuch gesammelt wurden.	24
4.8	Zeitstempel: Linksammlung Ende	24
4.9	Dauer: Linksammlung	24
4.10	Listen aus Erstversuch und Zweitversuch zusammenfügen	24
4.11	Bereinigte HTML-Links definieren	24
4.12	Bereinigte PDF-Links definieren	25
4.13	Dateinamen erstellen	25
4.13.1	Extrahieren relevanter Metadaten	25
4.13.2	Formatierung von Registerzeichen anpassen	26
4.13.3	Formatierung von Spruchkörper-Typ anpassen	26
4.13.4	Erste strenge REGEX-Validierung der Dateinamen	26
4.13.5	Ergebnis der ersten REGEX-Validierung	27
4.13.6	Skript stoppen falls erste REGEX-Validierung gescheitert	27
4.13.7	Zusätzliche Variablen einfügen	27
4.13.8	NAs einfügen für nicht benannte Entscheidungen	28
4.13.9	Zweite strenge REGEX-Validierung der Dateinamen	28
4.13.10	Ergebnis der zweiten REGEX-Validierung	29
4.13.11	Skript stoppen falls zweite REGEX-Validierung gescheitert	29
5	PDF-Download	30
5.1	Data Table für PDF-Download erstellen	30
5.2	Zeitstempel: PDF-Download Beginn	30
5.3	PDF-Download durchführen	30
5.4	Zeitstempel: PDF-Download Ende	31
5.5	Dauer: PDF-Download	31
5.6	PDF-Download: Ergebnis	31
5.6.1	Anzahl herunterzuladender Dateien	31
5.6.2	Anzahl heruntergeladener Dateien	31
5.6.3	Fehlbetrag	32
5.6.4	Fehlende Dateien	32
5.7	PDF-Wiederholungsversuch	32
5.8	PDF-Download: Gesamtergebnis	33
5.8.1	Anzahl herunterzuladender Dateien	33
5.8.2	Anzahl heruntergeladener Dateien	33
5.8.3	Fehlbetrag	33
5.8.4	Fehlende Dateien	34
5.8.5	Abschließende Hinweise	34
6	HTML-Download	35
6.1	Data Table für HTML-Download erstellen	35
6.2	Zeitstempel: HTML-Download Beginn	35
6.3	HTML-Download durchführen	35
6.4	Zeitstempel: HTML-Download Ende	35

6.5	Dauer: HTML-Download	36
6.6	HTML-Download: Ergebnis	36
6.6.1	Anzahl herunterzuladender Dateien	36
6.6.2	Anzahl heruntergeladener Dateien	36
6.6.3	Fehlbetrag	36
6.6.4	Fehlende Dateien	36
6.7	HTML-Wiederholungsversuch	37
6.8	HTML-Download: Gesamtergebnis	37
6.8.1	Anzahl herunterzuladender Dateien	37
6.8.2	Anzahl heruntergeladener Dateien	37
6.8.3	Fehlbetrag	37
6.8.4	Fehlende Dateien	38
7	HTML verarbeiten	39
7.1	Funktion anzeigen: f.bverfg.extract.meta	39
7.2	Funktion anzeigen: f.bverfg.extract.content	40
7.3	HTML-Dateien definieren	44
7.4	HTML-Dateien einlesen	44
7.5	HTML-Dateien parsen	44
7.6	Data Table mit allen Metadaten (inkl. ECLI)	45
7.7	Data Table mit vollständiger segmentierter Variante	45
7.8	Special Character entfernen	45
7.9	Stichprobe Metadaten	45
7.10	Stichprobe Segmentierte Variante	45
8	Text-Extraktion aus PDF	46
8.1	Vektor der zu extrahierenden Dateien erstellen	46
8.2	Anzahl zu extrahierender Dateien	46
8.3	PDF extrahieren: Funktion anzeigen	46
8.4	Text Extrahieren	47
9	Korpus Erstellen	48
9.1	TXT-Dateien Einlesen	48
9.2	In Data Table umwandeln	48
9.3	Durch Zeilenumbruch getrennte Wörter zusammenfügen	48
9.3.1	Funktion anzeigen	48
9.3.2	Funktion ausführen	49
9.4	Variable "datum" als Datentyp "IDate" kennzeichnen	49
9.5	Variable "entscheidungsjahr" hinzufügen	49
9.6	Variable "eingangsjahr_iso" hinzufügen	49
9.7	Datensatz nach Datum sortieren	49
9.8	Variable "praesi" hinzufügen	50
9.8.1	Lebensdaten einlesen	50
9.8.2	Personaldaten anzeigen	50
9.8.3	Hypothetisches Amtsende für PräsidentIn	51
9.8.4	Schleife vorbereiten	51
9.8.5	Vektor erstellen	51
9.8.6	Vektor einfügen	51
9.9	Variable "v_praesi" hinzufügen	51
9.9.1	Personaldaten einlesen	51

9.9.2	Personaldaten anzeigen	52
9.9.3	Hypothetisches Amtsende für Vize-PräsidentIn	52
9.9.4	Schleife vorbereiten	52
9.9.5	Vektor erstellen	53
9.9.6	Vektor einfügen	53
9.10	Variable “verfahrensart” hinzufügen	53
9.10.1	Datensatz einlesen	53
9.10.2	Datensatz auf relevante Daten reduzieren	53
9.10.3	Indizes bestimmen	53
9.10.4	Vektor der Verfahrensarten erstellen und einfügen	53
9.11	Variable “aktenzeichen” hinzufügen	54
9.12	Variable “doi_concept” hinzufügen	54
9.13	Variable “doi_version” hinzufügen	54
9.14	Variable “version” hinzufügen	54
9.15	Variable “lizenz” hinzufügen	54
9.16	Variable “ecli” hinzufügen	55
9.16.1	Metadaten mit ECLI-Duplikaten entfernen	56
9.16.2	Fehlerhafte ECLI korrigieren (HTML Meta)	56
9.16.3	Fehlerhafte ECLI korrigieren (HTML Full)	56
9.16.4	ECLI-Test 1: ECLI die in PDF, aber nicht in HTML vorhanden sind	57
9.16.5	ECLI-Test 2: ECLI die in HTML, aber nicht in PDF vorhanden sind	57
9.16.6	Zum Vergleich: Beim Download fehlende PDF-Dateien	57
9.16.7	ECLI Merge: Metadaten aus Hauptdatensatz in segmentierte Vari- ante mergen	58
9.16.8	ECLI Merge: Metadaten aus HTML-Extraktion in Hauptdatensatz mergen	58
9.17	Variable “entscheidung_typ” hinzufügen	58
9.17.1	Zitiervorschläge parsen	58
9.17.2	Kürzen	58
9.17.3	Vektor in Datensatz einfügen	59
10	Frequenztabellen erstellen	60
10.1	Funktion anzeigen	60
10.2	Ignorierte Variablen	61
10.3	Liste zu prüfender Variablen	61
10.4	Frequenztabellen erstellen	62
11	Frequenztabellen visualisieren	103
11.1	Präfix erstellen	103
11.2	Tabellen einlesen	103
11.3	Diagramm: Typ der Entscheidung	104
11.4	Diagramm: Typ des Spruchkörpers	105
11.5	Diagramm: Spruchkörper nach Aktenzeichen	106
11.6	Diagramm: Registerzeichen	107
11.7	Diagramm: Präsident:in	109
11.8	Diagramm: Vize-Präsident:in	110
11.9	Diagramm: Entscheidungsjahr	111
11.10	Diagramm: Eingangsjahr (ISO)	112
12	Korpus-Analytik	113

12.1	Berechnung linguistischer Kennwerte	113
12.1.1	Funktion anzeigen	113
12.1.2	Berechnung durchführen	114
12.2	Variablen-Namen anpassen	115
12.3	Kennwerte dem Korpus hinzufügen	115
12.4	Variante mit Metadaten erstellen	115
12.5	Linguistische Kennwerte	115
12.5.1	Zusammenfassungen berechnen	115
12.5.2	Zusammenfassungen anzeigen	116
12.5.3	Zusammenfassungen speichern	116
12.6	Quantitative Variablen	118
12.6.1	Entscheidungsdatum	118
12.6.2	Zusammenfassungen berechnen	118
12.6.3	Zusammenfassungen anzeigen	119
12.6.4	Zusammenfassungen speichern	119
12.7	Verteilungen linguistischer Kennwerte	120
12.7.1	Diagramm: Verteilung Zeichen	120
12.7.2	Diagramm: Verteilung Tokens	120
12.7.3	Diagramm: Verteilung Typen	122
12.7.4	Diagramm: Verteilung Sätze	123
13	Linguistische Annotationen berechnen	124
13.1	Funktion anzeigen: f.future_spacyparse	124
13.2	Berechnungen durchführen	125
14	Kontrolle der Variablen	126
14.1	Semantische Sortierung der Variablen	126
14.1.1	Variablen sortieren: Hauptdatensatz	126
14.1.2	Variablen sortieren: Metadaten	127
14.1.3	Variablen sortieren: Segmentiert	128
14.2	Anzahl Variablen der Datensätze	129
14.3	Alle Variablen-Namen der Datensätze	129
15	CSV-Dateien erstellen	131
15.1	CSV mit vollem Datensatz speichern	131
15.2	CSV mit Metadaten speichern	131
15.3	CSV mit Segmenten speichern	131
15.4	CSV mit Annotationen speichern	131
16	Dateigrößen analysieren	132
16.1	Gesamtgröße	132
16.1.1	Korpus-Objekt in RAM (MB)	132
16.1.2	CSV Korpus (MB)	132
16.1.3	CSV Metadaten (MB)	132
16.1.4	CSV Annotiert (MB)	132
16.1.5	CSV Segmentiert (MB)	132
16.1.6	PDF-Dateien (MB)	133
16.1.7	TXT-Dateien (MB)	133
16.2	Diagramm: Verteilung der Dateigrößen (PDF)	134
16.3	Diagramm: Verteilung der Dateigrößen (TXT)	135

17 Erstellen der ZIP-Archive	136
17.1 Verpacken der CSV-Dateien	136
17.1.1 Vollständiger Datensatz	136
17.1.2 Metadaten	136
17.1.3 Segmentiert	136
17.1.4 Annotiert	136
17.2 Verpacken der PDF-Dateien	137
17.3 Verpacken der HTML-Dateien	137
17.4 Verpacken der TXT-Dateien	137
17.5 Verpacken der Analyse-Dateien	138
17.6 Verpacken der Source-Dateien	138
18 Kryptographische Hashes	139
18.1 Liste der ZIP-Archive erstellen	139
18.2 Funktion anzeigen: future_multihashes	139
18.3 Hashes berechnen	140
18.4 In Data Table umwandeln	140
18.5 Index hinzufügen	140
18.6 In Datei schreiben	141
18.7 Leerzeichen hinzufügen um Zeilenumbruch zu ermöglichen	141
18.8 In Bericht anzeigen	141
19 Aufräumen	144
20 Abschluss	145
20.1 Datumsstempel	145
20.2 Datum und Uhrzeit (Anfang)	145
20.3 Datum und Uhrzeit (Ende)	145
20.4 Laufzeit des gesamten Skriptes	145
20.5 Warnungen	145
21 Parameter für strenge Replikationen	146
Literaturverzeichnis	148

1 Corpus der Entscheidungen des Bundesverfassungsgerichts

1.1 Überblick

Dieser Code lädt alle auf www.bundesverfassungsgericht.de verfügbaren Entscheidungen des Bundesverfassungsgerichts (BVerfG) herunter und verarbeitet sie in einen reichhaltigen menschen- und maschinenlesbaren Korpus. Es ist die Grundlage für den **Corpus der Entscheidungen des Bundesverfassungsgerichts (CE-BVerfG)**.

Alle mit diesem Skript erstellten Datensätze werden dauerhaft kostenlos und urheberrechtsfrei auf Zenodo, dem wissenschaftlichen Archiv des CERN, veröffentlicht. Alle Versionen sind mit einem persistenten Digital Object Identifier (DOI) versehen. Die neueste Version des Datensatzes ist immer über den Link der Concept DOI erreichbar: <https://doi.org/10.5281/zenodo.3902658>

1.2 Funktionsweise

Primäre Endprodukte des Skripts (im Ordner ‘output’) sind folgende ZIP-Archive:

- Der volle Datensatz im CSV-Format
- Die reinen Metadaten im CSV-Format (wie unter 1, nur ohne Entscheidungstexte)
- (Optional) Tokenisierte Form aller Texte mit linguistischen Annotationen im CSV-Format
- Alle Entscheidungen im TXT-Format (reduzierter Umfang an Metadaten)
- Alle Entscheidungen im PDF-Format (reduzierter Umfang an Metadaten)
- Alle Analyse-Ergebnisse (Tabellen als CSV, Grafiken als PDF und PNG)
- Der Source Code und alle weiteren Quelldaten

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

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_CE-BVerfG_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

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

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] "2022-02-01"
```

2.2 Datum und Uhrzeit (Beginn)

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

```
## [1] "2022-02-01 12:23:14 CET"
```

2.3 Packages Laden

```
library(RcppTOML) # Verarbeitung von TOML-Format
library(mgsub)   # Mehrfache simultane String-Substitutions
library(httr)    # HTTP-Werkzeuge
library(rvest)   # HTML/XML-Extraktion
library(knitr)   # Professionelles Reporting
library(kableExtra) # Verbesserte Kable Tabellen
library(pdftools) # Verarbeitung von PDF-Dateien
```

```
## Using poppler version 21.01.0
```

```
#library(doParallel) # Parallelisierung; to be deprecated
library(ggplot2)     # Fortgeschrittene Datenvisualisierung
library(scales)      # Skalierung von Diagrammen
library(data.table)  # Fortgeschrittene Datenverarbeitung
```

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

```
library(readtext)    # TXT-Dateien einlesen
library(quanteda)   # Fortgeschrittene Computerlinguistik
```

```
## Package version: 3.2.0
## Unicode version: 13.0
## ICU version: 67.1
```

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

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

```
library(spacyr)      # Linguistische Annotationen
library(future)      # Parallelisierung mit Futures
library(future.apply) # Apply-Funtionen für Futures
```

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("functions/f.bverfg.extract.content.R")
source("functions/f.bverfg.extract.meta.R")

source("R-fobbe-PROTO-package/f.remove.specialunderline.R")
source("R-fobbe-PROTO-package/f.linkextract.R")

source("R-fobbe-PROTO-package/f.hyphen.remove.R")
source("R-fobbe-PROTO-package/f.year.iso.R")
source("R-fobbe-PROTO-package/f.fast.freqtable.R")

#source("R-fobbe-PROTO-package/f.lingsummarize.iterator.R") # to be deprecated
#source("R-fobbe-PROTO-package/f.dopar.spacyparse.R") # to be deprecated
#source("R-fobbe-PROTO-package/f.dopar.multihashes.R") # to be deprecated
#source("R-fobbe-PROTO-package/f.dopar.pdfextract.R") # to be deprecated

source("R-fobbe-PROTO-package/f.future_lingsummarize.R")
source("R-fobbe-PROTO-package/f.future_multihashes.R")
source("R-fobbe-PROTO-package/f.future_pdf_to_txt.R")
source("R-fobbe-PROTO-package/f.future_spacyparse.R")

#source("General_Source_Functions.R") # deprecated
```

2.5 Verzeichnis für Analyse-Ergebnisse und Diagramme definieren

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

2.6 Weitere Verzeichnisse definieren

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

2.7 Dateien aus vorherigen Runs bereinigen

```
unlink(dir.analysis,  
  recursive = TRUE)  
  
unlink(dirs,  
  recursive = TRUE)  
  
files.delete <- list.files(pattern = "\\\\.zip|\\.pdf|\\.txt|\\.html",  
  ignore.case = TRUE)  
  
unlink(files.delete)
```

2.8 Verzeichnisse anlegen

```
dir.create(dir.analysis)  
  
lapply(dirs, dir.create)
```

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

2.9 Vollzitate statistischer Software schreiben

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

```
## Warning in as.vector(x): input string ' author = {Stéphane Dray and Anne-  
## Béatrice Dufour and Jean Thioulouse and Thibaut Jombart and Sandrine Pavoine  
and  
## Jean R. Lobry and Sébastien Ollier and Aurélie Siberchicot and Daniel Chessel  
},'  
## cannot be translated to UTF-8, is it valid in 'ANSI_X3.4-1968'?
```

2.10 Allgemeine Konfiguration

2.10.1 Konfiguration einlesen

```
config <- parseTOML("CE-BVerfG_Config.toml")
```

2.10.2 Konfiguration anzeigen

```
print(config)
```

```
## List of 11  
## $ annotate :List of 1  
## ..$ toggle: logi TRUE  
## $ cores :List of 2  
## ..$ max : logi TRUE  
## ..$ number: int 8  
## $ debug :List of 2  
## ..$ sample: int 10  
## ..$ toggle: logi FALSE  
## $ doi :List of 4  
## ..$ aktenzeichen : chr "10.5281/zenodo.4569564"  
## ..$ data :List of 2  
## .. ..$ concept: chr "10.5281/zenodo.3902658"  
## .. ..$ version: chr "10.5281/zenodo.5910152"  
## ..$ personendaten: chr "10.5281/zenodo.4568682"  
## ..$ software :List of 2  
## .. ..$ concept: chr "10.5281/zenodo.4308216"  
## .. ..$ version: chr "10.5281/zenodo.5910155"  
## $ 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 [1:11] "text" "eingangsnummer" "datum" "doc_id" ...  
## $ license :List of 2  
## ..$ code: chr "MIT-0"  
## ..$ data: chr "Creative Commons Zero 1.0 Universal"  
## $ parallel :List of 4  
## ..$ extractPDF : logi TRUE
```

```
## ..$ lingsummarize: logi TRUE
## ..$ multihashes : logi TRUE
## ..$ spacyparse : logi FALSE
## $ project :List of 3
## ..$ author : chr "Seán Fobbe"
## ..$ fullname : chr "Corpus der Entscheidungen des Bundesverfassungsgerichts"
## ..$ shortname: chr "CE-BVerfG"
## $ quanteda :List of 1
## ..$ tokens_locale: chr "de_DE"
```

2.10.3 Knitr Optionen setzen

```
knitr::opts_chunk$set(fig.path = paste0(dir.analysis, "/"),
  dev = config$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.5910152"
```

2.10.6 Präfix für Dateien definieren

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

```
## [1] "CE-BVerfG_2022-02-01"
```

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,
           "}" ),
    "\\n%-----Additional DOIs-----",
    paste0("\\newcommand{\\aktenzeichenurldoi}{https://doi.org/",
           config$doi$aktenzeichen,
           "}" ),
    paste0("\\newcommand{\\personendatenurldoi}{https://doi.org/",
           config$doi$personendaten,
           "}" ))

```

2.11.2 LaTeX Parameter schreiben

```

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

```

2.12 Parallelisierung aktivieren

Parallelisierung wird zur Beschleunigung 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.

2.12.1 Anzahl logischer Kerne festlegen

```

if (config$cores$max == TRUE){
  fullCores <- availableCores()
}

if (config$cores$max == FALSE){
  fullCores <- as.integer(config$cores$number)
}

print(fullCores)

```

```

## system
##      16

```

2.12.2 Quanteda

```
quanteda_options(threads = fullCores)
```

2.12.3 Data.table

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

3 Download: Weitere Datensätze

3.1 Registerzeichen und Verfahrensarten

Die Registerzeichen werden im Laufe des Skripts mit ihren detaillierten Bedeutungen aus dem folgenden Datensatz abgeglichen: "Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564." Das Ergebnis des Abgleichs wird in die Variable "verfahrensart" in den Datensatz eingefügt.

```
if (file.exists("data/AZ-BRD_1-0-1_DE_Registerzeichen_Datensatz.csv") == FALSE){
  download.file("https://zenodo.org/record/4569564/files/AZ-BRD_1-0-1_DE_
  Registerzeichen_Datensatz.csv?download=1",
  "data/AZ-BRD_1-0-1_DE_Registerzeichen_Datensatz.csv")
}
```

3.2 Personendaten zu Präsident:innen

Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2021). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2021-04-08. Zenodo. DOI: 10.5281/zenodo.4568682«.

```
if (file.exists("data/PVP-FCG_2021-04-08_GermanFederalCourts_Presidents.csv") ==
  FALSE){
  download.file("https://zenodo.org/record/4568682/files/PVP-FCG_2021-04-08_
  GermanFederalCourts_Presidents.csv?download=1",
  "data/PVP-FCG_2021-04-08_GermanFederalCourts_Presidents.csv")
}
```

3.3 Personendaten zu Vize-Präsident:innen

Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2021). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2021-04-08. Zenodo. DOI: 10.5281/zenodo.4568682«.

```
if (file.exists("data/PVP-FCG_2021-04-08_GermanFederalCourts_VicePresidents.csv")
  == FALSE){
  download.file("https://zenodo.org/record/4568682/files/PVP-FCG_2021-04-08_
  GermanFederalCourts_VicePresidents.csv?download=1",
  "data/PVP-FCG_2021-04-08_GermanFederalCourts_VicePresidents.csv
  ")
}
```

4 Download vorbereiten: Alle Entscheidungen des BVerfG

4.1 Zeitstempel: Linksammlung Beginn

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

```
## [1] "2022-02-01 12:23:16 CET"
```

4.2 Funktion zeigen

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

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

4.3 Maximale Seitenzahl auslesen

```
temp <- f.linkextract("https://www.bundesverfassungsgericht.de/SiteGlobals/Forms/
Suche/Entscheidungensuche_Formular.html?language_=de")

temp1 <- grep("list.*253D([0-9]+).*",
             temp,
             value = TRUE)

temp2 <- gsub(".*253D([0-9]+).*",
             "\\1",
             temp1)

maxpage <- max(as.numeric(temp2))

print(maxpage)
```

```
## [1] 822
```

4.4 Auszuwertende Seiten

```
pages <- seq_len(maxpage)
```

4.5 [Debugging Modus] Reduzierung der Seitenzahl

```
if (config$debug$toggle == TRUE){  
  pages <- sort(sample(pages,  
                      config$debug$sample))  
}
```

4.6 Linkliste erstellen: Erstversuch

```
breaks <- sort(sample(pages, 10))  
  
indices <- seq_along(pages)  
  
links.list <- vector("list",  
                    length(indices))  
  
print(pages)
```

```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18  
## [19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36  
## [37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54  
## [55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72  
## [73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90  
## [91] 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108  
## [109] 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126  
## [127] 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144  
## [145] 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162  
## [163] 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180  
## [181] 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198  
## [199] 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216  
## [217] 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234  
## [235] 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252  
## [253] 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270  
## [271] 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288  
## [289] 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306  
## [307] 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324  
## [325] 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342  
## [343] 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360  
## [361] 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378  
## [379] 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396  
## [397] 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
```

```

## [415] 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
## [433] 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
## [451] 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468
## [469] 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486
## [487] 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504
## [505] 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522
## [523] 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
## [541] 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558
## [559] 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576
## [577] 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594
## [595] 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612
## [613] 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630
## [631] 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648
## [649] 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666
## [667] 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684
## [685] 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702
## [703] 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
## [721] 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738
## [739] 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756
## [757] 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774
## [775] 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792
## [793] 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810
## [811] 811 812 813 814 815 816 817 818 819 820 821 822

```

```

for (i in indices){

  URL <- paste0("https://www.bundesverfassungsgericht.de/SiteGlobals/Forms/
Suche/Entscheidungensuche_Formular.html?gtp=5403124_list%253D",
               pages[i],
               "&language_=de")

  volatile <- f.linkextract(URL)

  links.l1 <- grep ("SharedDocs/Entscheidungen",
                  volatile,
                  ignore.case = TRUE,
                  value = TRUE)

  links.list[[i]] <- links.l1

  Sys.sleep(runif(1, 0.8, 2))

  if (i %in% breaks) Sys.sleep(runif(1, 2, 7))

}

Sys.sleep(runif(1, 20, 40))

```

4.7 Linkliste erstellen: Zweitversuch

4.7.1 Seiten für Zweitversuch definieren

Jede Seite der Datenbank sollte im Idealfall 10 Entscheidungen enthalten. Alle Seiten, die diese Bedingung nicht erfüllen werden noch einmal geprüft. Die letzte Seite enthält fast immer weniger als 10 Entscheidungen, wird sicherheitshalber aber trotzdem geprüft, statt sie auszusortieren. Dabei entstehende Duplikate werden später durch `unique()` entfernt.

```
retry.pages <- which(unlist(lapply(links.list,  
                                function(x)length(x)<10)))  
  
retry.indices <- seq_along(retry.pages)  
  
retry.links.list <- vector("list",  
                           length(retry.pages))
```

4.7.2 Anzeigen der Seiten die noch einmal geprüft werden.

```
print(retry.pages)
```

```
## integer(0)
```

4.7.3 Seiten prüfen

```
for (i in retry.indices){  
  
  URL <- paste0("https://www.bundesverfassungsgericht.de/SiteGlobals/Forms/  
Suche/Entscheidungensuche_Formular.html?gtp=5403124_list%253D",  
               retry.pages[i],  
               "&language_=de")  
  
  volatile <- f.linkextract(URL)  
  
  links.l1 <- grep ("SharedDocs/Entscheidungen",  
                  volatile,  
                  ignore.case = TRUE,  
                  value = TRUE)  
  
  retry.links.list[[i]] <- links.l1  
  
  Sys.sleep(runif(1, 0.8, 2))  
  
}
```

4.7.4 Anzeigen der Links die beim Wiederholungsversuch gesammelt wurden.

```
print(retry.links.list)
```

```
## list()
```

4.8 Zeitstempel: Linksammlung Ende

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

```
## [1] "2022-02-01 13:08:14 CET"
```

4.9 Dauer: Linksammlung

```
end.links - begin.links
```

```
## Time difference of 44.96777 mins
```

4.10 Listen aus Erstversuch und Zweitversuch zusammenfügen

```
links.relative <- c(unlist(links.list),  
                  unlist(retry.links.list))
```

4.11 Bereinigte HTML-Links definieren

```
links.raw <- paste0("https://www.bundesverfassungsgericht.de/",  
                  links.relative)  
  
links.html <- gsub("(.+\\.html).*",  
                 "\\1",  
                 links.raw)  
  
links.html <- unique(links.html)
```


4.12 Bereinigte PDF-Links definieren

```
links.pdf <- gsub("Entscheidungen",
                 "Downloads",
                 links.html)

links.pdf <- gsub("\\.html.*",
                 "\\..pdf\\?__blob=publicationFile\\&v\\=1",
                 links.pdf)

links.pdf <- unique(links.pdf)
```

4.13 Dateinamen erstellen

4.13.1 Extrahieren relevanter Metadaten

Die Links zu jeder Entscheidung enthalten das Ordinalzahl-Element ihres jeweiligen ECLI-Codes. Struktur und Inhalt der ECLI für deutsche Gerichte sind auf dem Europäischen Justizportal näher erläutert.¹

```
filenames <- basename(links.pdf)

filenames <- gsub("[?].*",
                 "",
                 filenames)
```

Normale Struktur

```
filenames1 <- gsub("[a-z]([a-z])([0-9]{4})([0-9]{2})([0-9]{2})_([0-9])([a-z]*)
                  ([0-9]{4})([0-9]{2}).*",
                  "BVerfG_\\2-\\3-\\4_\\1_\\5_\\6_\\7_\\8_NA",
                  filenames)
```

Struktur von Verzögerungsrügen

```
filenames1 <- gsub("_vz", "NA_Vz", filenames1)

filenames1 <- gsub("[a-z]([a-z])([0-9]{4})([0-9]{2})([0-9]{2})(NA)_([Vz])([0-9]{4})
                  ([0-9]{2}).*",
                  "BVerfG_\\2-\\3-\\4_\\1_\\5_\\6_\\7_\\8_NA",
                  filenames1)
```

Struktur von Entscheidungen mit Kollisions-Variable

¹ https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-de-de.do?member=1

```

filenames1 <- gsub("[a-z]([a-z])([0-9]{4})([0-9]{2})([0-9]{2})([a-z0-9])_([0-9])
([a-z]*)([0-9]{4})([0-9]{2}).*",
                  "BVerfG_\\2-\\3-\\4_\\1_\\6_\\7_\\8_\\9_\\5",
                  filenames1)

```

4.13.2 Formatierung von Registerzeichen anpassen

```

filenames1 <- gsub("_bv([a-z])_",
                  "_Bv\\U\\1_",
                  perl = TRUE,
                  filenames1)

filenames1 <- gsub("pbvu",
                  "PBvU",
                  filenames1)

```

4.13.3 Formatierung von Spruchkörper-Typ anpassen

```

filenames1 <- gsub("_([kpsb])_",
                  "_\\U\\1_",
                  perl = TRUE,
                  filenames1)

```

4.13.4 Erste strenge REGEX-Validierung der Dateinamen

```

#regex.test1 <- grep("BVerfG_[0-9]{4}-[0-9]{2}-[0-9]{2}_[A-Z]_[0-9NA]+_[A-Za-z]+_
[0-9]{4}_[0-9]{2}_[0-9a-zA]+$",
#
#           filenames1,
#           invert = TRUE,
#           value = TRUE)

regex.test1 <- grep(paste0("BVerfG", # gericht
                           "_",
                           "[0-9]{4}-[0-9]{2}-[0-9]{2}", # datum
                           "_",
                           "[A-Z]", # spruchkoerper_typ
                           "_",
                           "[0-9NA]+", # spruchkoerper_az
                           "_",
                           "[A-Za-z]+", # registerzeichen
                           "_",
                           "[0-9]{4}", # eingangsnummer
                           "_",
                           "[0-9]{2}", # eingangsjahr
                           "_",
                           "[0-9a-zA]+$"), # kollision

```

```
filenames1,  
invert = TRUE,  
value = TRUE)
```

4.13.5 Ergebnis der ersten REGEX-Validierung

Das Ergebnis sollte ein leerer Vektor sein!

```
print(regex.test1)
```

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

4.13.6 Skript stoppen falls erste REGEX-Validierung gescheitert

```
if (length(regex.test1) != 0){  
  stop("REGEX VALIDIERUNG 1 GESCHEITERT: AKTENZEICHEN ENTSPRECHEN NICHT DEM  
  CODEBOOK-SCHEMA!")  
}
```

4.13.7 Zusätzliche Variablen einfügen

```
extravariablen <- fread("data/BVerfGE_Variablen_NameBandSeite_2021_09_19.csv")  
  
extravariablen$newname <- paste(extravariablen$oldname,  
                               extravariablen$name,  
                               extravariablen$band,  
                               extravariablen$seite,  
                               sep = "_")  
  
extravariablen$newname <- paste0(extravariablen$newname,  
                                ".pdf")  
  
filenames2 <- filenames1  
  
targetindices <- match(extravariablen$oldname,  
                       filenames2)  
  
newname <- extravariablen$newname  
  
dt <- data.table(targetindices, newname)[complete.cases(targetindices)]  
  
if(dt[,.N] > 0){
```

```

filenames2 <- replace(filenames2,
                      dt$targetindices,
                      dt$newname)
}

```

4.13.8 NAs einfügen für nicht benannte Entscheidungen

```

values <- grep(".pdf",
              filenames2,
              invert = TRUE,
              value = TRUE)

indices <- grep(".pdf",
              filenames2,
              invert = TRUE)

filenames2[indices] <- paste0(values,
                              "_NA_NA_NA.pdf")

```

4.13.9 Zweite strenge REGEX-Validierung der Dateinamen

```

#regex.test2 <- grep("^BVerfG_[0-9]{4}-[0-9]{2}-[0-9]{2}_[SPKB]_[0-9NA]+_[A-Za-z
]+_[0-9]{4}_[0-9]{2}_[0-9a-zA]+_[0-9ÄÜÖäüöA-Za-z\\-]+_[NAO-9]+_[NAO-9]+\\.
pdf$",
#
#           filenames2,
#           value = TRUE,
#           invert = TRUE)

regex.test2 <- grep(paste0("^BVerfG", # gericht
                           "_",
                           "[0-9]{4}-[0-9]{2}-[0-9]{2}", # datum
                           "_",
                           "[SPKB]", # spruchkoerper_typ
                           "_",
                           "[0-9NA]+", # spruchkoerper_az
                           "_",
                           "[A-Za-z]+", # registerzeichen
                           "_",
                           "[0-9]{4}", # eingangsnummer
                           "_",
                           "[0-9]{2}", # eingangsjahr_az
                           "_",
                           "[0-9a-zA]+", # kollision
                           "_",
                           "[0-9ÄÜÖäüöA-Za-z\\-]+", # name
                           "_",
                           "[NAO-9]+", # band
                           "_",
                           "[NAO-9]+", # seite
                           "\\..pdf$"), # Dateiendung

```

```
filenames2,  
value = TRUE,  
invert = TRUE)
```

4.13.10 Ergebnis der zweiten REGEX-Validierung

Das Ergebnis sollte ein leerer Vektor sein!

```
print(regex.test2)
```

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

4.13.11 Skript stoppen falls zweite REGEX-Validierung gescheitert

```
if (length(regex.test2) != 0){  
  stop("REGEX VALIDIERUNG 2 GESCHEITERT: AKTENZEICHEN ENTSPRECHEN NICHT DEM  
  CODEBOOK-SCHEMA!")  
}
```

5 PDF-Download

5.1 Data Table für PDF-Download erstellen

```
dt <- data.table(links.pdf,  
                 filenames2)
```

5.2 Zeitstempel: PDF-Download Beginn

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

```
## [1] "2022-02-01 13:08:14 CET"
```

5.3 PDF-Download durchführen

Hinweis: Es ist nötig jeden Link auf das Vorhandensein einer PDF-Datei zu prüfen, weil für manche Entscheidungen zwar HTML-Seiten vorhanden sind, aber keine korrespondierende PDF-Datei.

```
for (i in seq_len(dt[,.N])){  
  
  response <- GET(dt$links.pdf[i])  
  
  Sys.sleep(runif(1, 0.25, 0.75))  
  
  if (response$headers$"content-type" == "application/pdf;charset=UTF-8" &  
      response$status_code == 200){  
    tryCatch({download.file(url = dt$links.pdf[i],  
                           destfile = dt$filenames2[i])  
            },  
            error=function(cond) {  
              return(NA)}  
            )  
  }else{  
    print(paste0(dt$filenames2[i],  
                 " : kein PDF vorhanden"))  
  }  
  Sys.sleep(runif(1, 0.3, 1))  
}
```

```
## [1] "BVerfG_2021-03-25_S_2_BvF_0001_20_NA_NA_NA_NA.pdf : kein PDF vorhanden"  
## [1] "BVerfG_2014-06-10_K_1_BvR_0669_14_NA_NA_NA_NA.pdf : kein PDF vorhanden"  
## [1] "BVerfG_2014-03-19_K_1_BvR_1417_10_NA_NA_NA_NA.pdf : kein PDF vorhanden"  
## [1] "BVerfG_2012-10-28_K_2_BvR_0737_11_NA_NA_NA_NA.pdf : kein PDF vorhanden"  
## [1] "BVerfG_2004-05-19_K_1_BvR_0711_04_NA_NA_NA_NA.pdf : kein PDF vorhanden"
```

```
## Warning in download.file(url = dt$links.pdf[i], destfile = dt$filenames2[i]):  
## URL 'https://www.bundesverfassungsgericht.de/SharedDocs/Downloads/DE/2003/08/  
## qk20030814b_1bvq003003.pdf?__blob=publicationFile&v=1': status was 'Server  
## returned nothing (no headers, no data)'
```

```
## [1] "BVerfG_2001-10-23_K_2_BvR_1236_01_NA_NA_NA_NA.pdf : kein PDF vorhanden"  
## [1] "BVerfG_2001-10-01_S_2_BvB_0001_01_NA_NPD-Verbot-EA-RückgabeEDV-2_104_61.  
pdf : kein PDF vorhanden"
```

5.4 Zeitstempel: PDF-Download Ende

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

```
## [1] "2022-02-01 17:10:36 CET"
```

5.5 Dauer: PDF-Download

```
end.download - begin.download
```

```
## Time difference of 4.039545 hours
```

5.6 PDF-Download: Ergebnis

5.6.1 Anzahl herunterzuladender Dateien

```
dt[, .N]
```

```
## [1] 8220
```

5.6.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8212
```

5.6.3 Fehlbetrag

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

```
## [1] 8
```

5.6.4 Fehlende Dateien

```
missing.pdf <- setdiff(dt$filenames2,
                      files.pdf)

print(missing.pdf)
```

```
## [1] "BVerfG_2021-03-25_S_2_BvF_0001_20_NA_NA_NA_NA.pdf"
## [2] "BVerfG_2014-06-10_K_1_BvR_0669_14_NA_NA_NA_NA.pdf"
## [3] "BVerfG_2014-03-19_K_1_BvR_1417_10_NA_NA_NA_NA.pdf"
## [4] "BVerfG_2004-05-19_K_1_BvR_0711_04_NA_NA_NA_NA.pdf"
## [5] "BVerfG_2003-08-14_K_1_BvQ_0030_03_b_NA_NA_NA.pdf"
## [6] "BVerfG_2001-10-23_K_2_BvR_1236_01_NA_NA_NA_NA.pdf"
## [7] "BVerfG_2001-10-01_S_2_BvB_0001_01_NA_NPD-Verbot-EA-RückgabeEDV-2_104_61.
pdf"
```

5.7 PDF-Wiederholungsversuch

Download für fehlende Dokumente wiederholen.

```
if(N.missing.pdf > 0){

  dt.retry <- dt[filenames2 %in% missing.pdf]

  for (i in seq_len(dt.retry[,.N])){

    response <- GET(dt.retry$links.pdf[i])

    Sys.sleep(runif(1, 0.25, 0.75))

    if (response$headers$"content-type" == "application/pdf;charset=UTF-8" &
        response$status_code == 200){

      tryCatch({download.file(url = dt.retry$links.pdf[i],
                             destfile = dt.retry$filenames2[i])

      },
      error = function(cond) {
        return(NA)}
    )
  }
}
```



```

    )
  }else{
    print(paste0(dt.retry$filenames2[i],
                " : kein PDF vorhanden"))
  }
  Sys.sleep(runif(1, 0.5, 1.5))
}
}

```

```

## [1] "BVerfG_2021-03-25_S_2_BvF_0001_20_NA_NA_NA_NA.pdf : kein PDF vorhanden"
## [1] "BVerfG_2014-06-10_K_1_BvR_0669_14_NA_NA_NA_NA.pdf : kein PDF vorhanden"
## [1] "BVerfG_2014-03-19_K_1_BvR_1417_10_NA_NA_NA_NA.pdf : kein PDF vorhanden"
## [1] "BVerfG_2004-05-19_K_1_BvR_0711_04_NA_NA_NA_NA.pdf : kein PDF vorhanden"
## [1] "BVerfG_2001-10-23_K_2_BvR_1236_01_NA_NA_NA_NA.pdf : kein PDF vorhanden"
## [1] "BVerfG_2001-10-01_S_2_BvB_0001_01_NA_NPD-Verbot-EA-RückgabeEDV-2_104_61.
pdf : kein PDF vorhanden"

```

5.8 PDF-Download: Gesamtergebnis

5.8.1 Anzahl herunterzuladender Dateien

```
dt[,.N]
```

```
## [1] 8220
```

5.8.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8213
```

5.8.3 Fehlbetrag

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

```
## [1] 7
```

5.8.4 Fehlende Dateien

```
missing.pdf <- setdiff(dt$filenames2, files.pdf)
print(missing.pdf)
```

```
## [1] "BVerfG_2021-03-25_S_2_BvF_0001_20_NA_NA_NA_NA.pdf"
## [2] "BVerfG_2014-06-10_K_1_BvR_0669_14_NA_NA_NA_NA.pdf"
## [3] "BVerfG_2014-03-19_K_1_BvR_1417_10_NA_NA_NA_NA.pdf"
## [4] "BVerfG_2004-05-19_K_1_BvR_0711_04_NA_NA_NA_NA.pdf"
## [5] "BVerfG_2001-10-23_K_2_BvR_1236_01_NA_NA_NA_NA.pdf"
## [6] "BVerfG_2001-10-01_S_2_BvB_0001_01_NA_NPD-Verbot-EA-RückgabeEDV-2_104_61.
pdf"
```

5.8.5 Abschließende Hinweise

Hinweis: Für die Entscheidung vom 1.10.2001 zur Rückgabe von EDV-Anlagen im Rahmen des NPD-Verfahrens war auch nach manueller Suche keine PDF-Datei auffindbar.

6 HTML-Download

6.1 Data Table für HTML-Download erstellen

```
names.html <- basename(links.html)

dt.download.html <- data.table(links.html,
                               names.html)
```

6.2 Zeitstempel: HTML-Download Beginn

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

```
## [1] "2022-02-01 17:10:50 CET"
```

6.3 HTML-Download durchführen

```
for (i in sample(dt.download.html[, .N])){

  tryCatch({download.file(dt.download.html$links.html[i],
                        dt.download.html$names.html[i])
            },
            error = function(cond) {
              return(NA)}
            )

  Sys.sleep(runif(1, 0.3, 1))

}
```

6.4 Zeitstempel: HTML-Download Ende

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

```
## [1] "2022-02-01 19:48:24 CET"
```

6.5 Dauer: HTML-Download

```
end.download - begin.download
```

```
## Time difference of 2.626071 hours
```

6.6 HTML-Download: Ergebnis

6.6.1 Anzahl herunterzuladender Dateien

```
dt.download.html[,.N]
```

```
## [1] 8220
```

6.6.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8220
```

6.6.3 Fehlbetrag

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

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

6.6.4 Fehlende Dateien

```
missing <- setdiff(dt.download.html$names.html,  
                  files.html)  
  
print(missing)
```

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

6.7 HTML-Wiederholungsversuch

Download für fehlende Dokumente wiederholen.

```
if(N.missing > 0){  
  dt.retry <- dt.download.html[names.html %in% missing]  
  for (i in seq_len(dt.retry[,.N])){  
    tryCatch({download.file(dt.retry$links.html[i],  
                           dt.retry$names.html[i])  
    },  
    error = function(cond) {  
      return(NA)}  
    )  
    Sys.sleep(runif(1, 0.5, 1.5))  
  }  
}
```

6.8 HTML-Download: Gesamtergebnis

6.8.1 Anzahl herunterzuladender Dateien

```
dt.download.html[,.N]
```

```
## [1] 8220
```

6.8.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8220
```

6.8.3 Fehlbetrag

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

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

6.8.4 Fehlende Dateien

```
missing <- setdiff(dt.download.html$names.html,  
                  files.html)  
  
print(missing)
```

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

7 HTML verarbeiten

7.1 Funktion anzeigen: f.bverfg.extract.meta

```
print(f.bverfg.extract.meta)
```

```
## function(HTML){
##
##   ## ECLI
##   ecli <- html_elements(HTML, "[class='ecli']") %>% html_text(trim = TRUE)
##
##   ## Alle Aktenzeichen
##   aktenzeichen_alle <- html_elements(HTML, "[class='az2']") %>% html_text(
trim = TRUE)
##   aktenzeichen_alle <- gsub("-", "|",
##                               "",
##                               aktenzeichen_alle)
##   aktenzeichen_alle <- gsub(" +",
##                               " ",
##                               aktenzeichen_alle)
##   aktenzeichen_alle <- tstrsplit(aktenzeichen_alle,
##                                   split = "\n")
##   aktenzeichen_alle <- unlist(aktenzeichen_alle)
##   aktenzeichen_alle <- na.omit(aktenzeichen_alle)
##   aktenzeichen_alle <- trimws(aktenzeichen_alle)
##   aktenzeichen_alle <- unique(aktenzeichen_alle)
##   aktenzeichen_alle <- paste(aktenzeichen_alle, collapse = "|")
##
##   ## Pressemitteilungen
##   pressemitteilung <- html_elements(HTML, "[class='pm']") %>% html_text(trim
= TRUE)
##
##   if(length(pressemitteilung) == 0){
##     pressemitteilung <- "NA"
##   }else{
##     pressemitteilung <- pressemitteilung[1]
##   }
##
##   ## Kurzbeschreibung
##   kurzbeschreibung <- HTML %>% html_elements("title") %>% html_text(trim =
TRUE)
##   kurzbeschreibung <- gsub(".*Bundesverfassungsgericht.*-.*Entscheidungen.*
-(.*)",
##                               "\\1",
##                               kurzbeschreibung)
##   kurzbeschreibung <- trimws(kurzbeschreibung)
##
##
##   ## Verkündung ## hier treten noch Probleme auf
##   ##verkuendung <- html_elements(HTML, "[class='vvm2']") %>% html_text(trim
= TRUE)
##}
```

```

##   ##if(length(verkuendung) == 0){
##   ##   verkuendung <- "NA"
##   ##}
##
##
##   ## Zitiervorschlag
##   zitiervorschlag <- html_elements(HTML, "[class='cite']") %>% html_text(
trim = TRUE)
##
##   ## Richter:innen
##   richter <- html_elements(HTML, "[class='st']") %>% html_elements("td") %>%
html_text(trim = TRUE)
##   richter <- tstrsplit(richter,
##                       split = "\n")
##   richter <- unlist(richter)
##   richter <- na.omit(richter)
##   richter <- unique(richter)
##   #richter <- richter[grep("ausgeschieden|gehindert|verhindert", richter,
invert = TRUE)] # optional Verhinderungen entfernen; Verhinderungen werden
zunächst nicht entfernt, da sie durchaus wertvolle Informationen enthalten.
##   richter <- paste(richter, collapse = "|")
##   richter <- gsub("\\|+",
##                  "\\|",
##                  richter)
##   richter <- gsub("(^\\..*\\|)|\\|$",
##                  "",
##                  richter)
##
##   ## Beschwerdeführer (idR anonymisiert, daher nicht aufgenommen)
##   ##beschwerdefuehrer <- html_elements(HTML, "[class='vb3']") %>% html_text(
trim = TRUE)
##
##   meta <- data.table(ecli,
##                     pressemitteilung,
##                     aktenzeichen_alle,
##                     ##verkuendung,
##                     kurzbeschreibung,
##                     zitiervorschlag,
##                     richter)
##
##   return(meta)
##
## }

```

7.2 Funktion anzeigen: f.bverfg.extract.content

```
print(f.bverfg.extract.content)
```

```
## function(HTML){
##
```



```

##      ## Leitsätze
##      leitsaetze <- unique(html_elements(HTML, ".ls") %>% html_text(trim = TRUE)
##    )
##
##      if (length(leitsaetze) != 0){
##        segment <- paste0("leitsatz-",
##                          1:length(leitsaetze))
##
##        dt.leitsaetze <- data.table(leitsaetze,
##                                   segment)
##
##      }else{
##        dt.leitsaetze <- data.table("NA",
##                                   "NA")
##
##      }
##
##      ## Gegenstand
##      gegenstand <- html_elements(HTML, ".vgt3") %>% html_text(trim = TRUE)
##
##      if (length(gegenstand) == 0){
##        gegenstand <- html_elements(HTML, ".vg1") %>% html_text(trim = TRUE)
##      }
##
##      if (length(gegenstand) != 0){
##        segment <- paste0("gegenstand-",
##                          1:length(gegenstand))
##
##        dt.gegenstand <- data.table(gegenstand,
##                                   segment)
##
##      }else{
##        dt.gegenstand <- data.table("NA",
##                                   "NA")
##
##      }
##
##      ## Formel
##      formel <- paste(html_elements(HTML, "[class='bf']") %>% html_text(trim =
##    TRUE),
##                      collapse = " ")
##
##      if (length(formel) != 0){
##        segment <- "formel"
##
##        dt.formel <- data.table(formel,
##                                segment)
##
##      }else{
##        dt.formel <- data.table("NA",
##                                "NA")
##

```

```

##
##   }
##
##   ## Tenor
##   tenor <- html_elements(HTML, "ol .bs, .hr") %>% html_text(trim = TRUE) #
Mehrere Tenorpunkte
##
##   if (length(tenor) == 0){
##
##       tenor <- html_elements(HTML, ".bs, .hr") %>% html_text(trim = TRUE) #
Ein Tenorpunkt
##
##   }
##
##
##
##   if (length(tenor) != 0){
##       segment <- paste0("tenor-",
##                         1:length(tenor))
##
##       dt.tenor <- data.table(tenor,
##                              segment)
##
##   }else{
##       dt.tenor <- data.table("NA",
##                              "NA")
##
##   }
##
##
##
##   ## Gründe der Entscheidung
##   gruende.nodes <- html_elements(HTML, ".std, .lszb")
##
##   if (length(gruende.nodes) != 0){
##
##       if(html_attr(gruende.nodes, "class")[1] == "lszb"){
##
##           ## Entferne Leitsatzüberschriften
##           gruende.classes <- html_attr(gruende.nodes, "class")
##           ueberschrift.indizes <- grep("lszb", gruende.classes[1:5])
##
##           gruende.nodes <- gruende.nodes[-ueberschrift.indizes]
##       }
##
##       inhaltsverzeichnis.indizes <- grep("inhvz", html_attr(gruende.nodes, "
class"))
##
##       if (length(inhaltsverzeichnis.indizes) > 0){
##           gruende.nodes <- gruende.nodes[-inhaltsverzeichnis.indizes] #
Entferne Inhaltsverzeichnis
##       }
##
##
##       etstd.indizes <- grep("et std", html_attr(gruende.nodes, "class"))
##
##       if (length(etstd.indizes) > 0){

```

```

##
##           gruende.nodes <- gruende.nodes[-etzstd.indizes] # Entferne
Langzitate
##
##           }
##
##           etzstd.indizes <- grep("(etz std)|(std etz)", html_attr(gruende.nodes,
"class"))
##
##           if (length(etzstd.indizes) > 0){
##
##           gruende.nodes <- gruende.nodes[-etzstd.indizes] # Entferne
Langzitatüberschriften
##
##           }
##
##
##
##           abw.vorhanden <- "lszb" %in% html_attr(gruende.nodes, "class")
##
##           if (abw.vorhanden == TRUE){
##
##           gruende.classes <- html_attr(gruende.nodes, "class")
##
##           abw.logical <- gruende.classes %in% "lszb"
##
##           abw.pos.all <- which(abw.logical)
##           abw.pos.first <- abw.pos.all[1]
##
##           gruende.nodes <- gruende.nodes[-abw.pos.all] # Entferne Abw-
Meinung-Überschriften
##
##           }
##
##           gruende <- html_text(gruende.nodes,
##                               trim = TRUE)
##
##           segment <- paste0("gruende-",
##                             1:length(gruende))
##
##           if (abw.vorhanden == TRUE){
##
##           abw.indizes <- abw.pos.first:length(segment)
##           segment[abw.indizes] <- paste0(segment[abw.indizes],
##                                           "-sondervotum")
##           }
##
##
##           dt.gruende <- data.table(gruende,
##                                   segment)
##
##           }else{
##           dt.gruende <- data.table("NA",
##                                   "NA")
##           }
##
##

```

```

##
##   ## Unterschriften
##   unterschriften <- paste(html_elements(HTML, "[class='st']") %>% html_
elements("td") %>% html_text(trim = TRUE),
##                               collapse = " ")
##
##   segment <- "unterschriften"
##   dt.unterschriften <- data.table(unterschriften,
##                                   segment)
##
##   ## Segmentierter Inhalt
##
##   inhalt <- rbind(dt.leitsaetze,
##                  dt.gegenstand,
##                  dt.formel,
##                  dt.tenor,
##                  dt.gruende,
##                  dt.unterschriften,
##                  use.names = FALSE)
##
##   setnames(inhalt,
##            new = c("text",
##                  "segment"))
##
##   inhalt <- inhalt[grep("NA", inhalt$segment, invert = TRUE)]
##
##   inhalt[, text := lapply(.(text), f.remove.specialunderline)]
##
##   return(inhalt)
## }

```

7.3 HTML-Dateien definieren

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

7.4 HTML-Dateien einlesen

```
html.list <- lapply(files.html,
                   read_html)
```

7.5 HTML-Dateien parsen

```
meta.list <- lapply(html.list,
                   f.bverfg.extract.meta)
```

```

content.list <- lapply(html.list,
                      f.bverfg.extract.content)

segmented.full.list <- vector("list",
                             length(meta.list))

for (i in 1:length(meta.list)){
  content.rows <- content.list[[i]][,.N]
  meta.replicate <- meta.list[[i]][rep(1, content.rows)]
  segmented.full.list[[i]] <- cbind(content.list[[i]],
                                   meta.replicate)
}

```

7.6 Data Table mit allen Metadaten (inkl. ECLI)

```
dt.meta.html <- rbindlist(meta.list)
```

7.7 Data Table mit vollständiger segmentierter Variante

```
dt.segmented.full <- rbindlist(segmented.full.list)
```

7.8 Special Character entfernen

An dieser Stelle wird ein mysteriöser Unterstrich entfernt, vermutlich ein non-breaking space. Es ist allerdings unklar wieso dieser in den Daten auftaucht. Der Code wird nicht im Compilation Report angezeigt, weil sich dieses Zeichen bei dem listings package zu Fehlern führt.

7.9 Stichprobe Metadaten

```

fwrite(dt.meta.html[sample(.N, 50)],
       file.path(dir.analysis,
                 "QA_Stichprobe_HTML-Metadaten.csv"))

```

7.10 Stichprobe Segmentierte Variante

```

fwrite(dt.segmented.full[sample(.N, 50)],
       file.path(dir.analysis,
                 "QA_Stichprobe_HTML-SegmentierterVolltext.csv"))

```

8 Text-Extraktion aus PDF

8.1 Vektor der zu extrahierenden Dateien erstellen

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

8.2 Anzahl zu extrahierender Dateien

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

```
## [1] 8213
```

8.3 PDF extrahieren: Funktion anzeigen

```
print(f.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,  
                                       f.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))

```

```

}

```

8.4 Text Extrahieren

```

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

f.future_pdf_to_txt(files.pdf)

```

```

## Processing 8213 files. Begin at: 2022-02-01 19:52:54

```

```

## Successfully processed 8213 files. 0 files failed. Runtime was 14.67 secs.
   Ended at: 2022-02-01 19:53:09

```

9 Korpus Erstellen

9.1 TXT-Dateien Einlesen

```
txt.bverfg <- readtext("./*.txt",
  docvarsfrom = "filenames",
  docvarnames = c("gericht",
    "datum",
    "spruchkoerper_typ",
    "spruchkoerper_az",
    "registerzeichen",
    "eingangsnummer",
    "eingangsjahr_az",
    "kollision",
    "name",
    "band",
    "seite"),
  dvsep = "_",
  encoding = "UTF-8")
```

9.2 In Data Table umwandeln

```
setDT(txt.bverfg)
```

9.3 Durch Zeilenumbruch getrennte Wörter zusammenfügen

Durch Zeilenumbrüche getrennte Wörter stellen bei aus PDF-Dateien gewonnene Text-Korpora ein erhebliches Problem dar. Wörter werden dadurch in zwei sinnentleerte Tokens getrennt, statt ein einzelnes und sinnvolles Token zu bilden. Dieser Schritt entfernt die Bindestriche, den Zeilenumbruch und ggf. dazwischenliegende Leerzeichen.

9.3.1 Funktion anzeigen

```
print(f.hyphen.remove)
```

```
## function(text){
##   ## Examples: Ham-\nburg, Mei-\n  nungsäußerung
##   text.out <- gsub("[a-zöäüß]-[[:blank:]]*\n[[:blank:]]*([a-zöäüß])",
##     "\\1\\2",
##     text)
##   ## Examples: SARS-CoV-\n2
##   text.out <- gsub("[a-zA-ZöäüÖÄÜß]-[[:blank:]]*\n[[:blank:]]*([A-Z0-9ÖÄÜß
##     ])",
##     "\\1-\\2",
##     text.out)
##   ## Example: hat- 2\nte, Unsterb- 6\nliche
```



```
##   text.out <- gsub("[a-zöäüß)-[[:blank:]]*[0-9]+[[:blank:]]*\n[[:blank:]]*"
##   ([a-zöäüß]),
##           "\\1\\2",
##   text.out)
##
##   ## Example: hat- \n 2 te, Unsterb- \n 6 liche
##   text.out <- gsub("[a-zöäüß)-[[:space:]]*[0-9]+[[:blank:]]*([a-zöäüß)",
##           "\\1\\2",
##   text.out)
##
##   return(text.out)
## }
```

9.3.2 Funktion ausführen

```
txt.bverfg[, text := lapply(.text), f.hyphen.remove]
dt.segmented.full[, text := lapply(.text), f.hyphen.remove]
```

9.4 Variable “datum” als Datentyp “IDate” kennzeichnen

```
txt.bverfg$datum <- as.IDate(txt.bverfg$datum)
```

9.5 Variable “entscheidungsjahr” hinzufügen

```
txt.bverfg$entscheidungsjahr <- year(txt.bverfg$datum)
```

9.6 Variable “eingangsjahr_iso” hinzufügen

```
txt.bverfg$eingangsjahr_iso <- f.year.iso(txt.bverfg$eingangsjahr_az)
```

9.7 Datensatz nach Datum sortieren

Aufgrund der Position der Datums-Variable ist der Datensatz vermutlich schon von Linux nach Datum sortiert worden. Die Erstellung der Variablen für Präsidenten und Vize-Präsidenten trifft allerdings die starke Annahme, dass eine aufsteigende Sortierung nach Datum besteht. Wäre das nicht der Fall, würden dort Fehler auftreten. Diese Sortierung ist als fail-safe gedacht.

```
setorder(txt.bverfg,
         datum)
```

9.8 Variable “praesi” hinzufügen

Diese Variable dokumentiert für jede Entscheidung welche/r Präsident:in am Tag der Entscheidung im Amt war.

9.8.1 Lebensdaten einlesen

```
praesi <- fread(file.path("data",
                           "PVP-FCG_2021-04-08_GermanFederalCourts_Presidents.csv"
                           ))
praesi <- praesi[court == "BVerfG", c(1:3, 5:6)]
```

9.8.2 Personaldaten anzeigen

```
kable(praesi,
      format = "latex",
      align = "r",
      booktabs = TRUE,
      longtable = TRUE) %>% kable_styling(latex_options = "repeat_header")
```

court	name_last	name_first	term_begin_date	term_end_date
BVerfG	Höpker-Aschoff	Hermann	1951-09-07	1954-01-15
BVerfG	VACANCY-1	VACANCY-1	1954-01-16	1954-03-22
BVerfG	Wintrich	Josef	1954-03-23	1958-10-19
BVerfG	VACANCY-2	VACANCY-2	1958-10-20	1959-01-07
BVerfG	Müller	Gebhard	1959-01-08	1971-12-07
BVerfG	Benda	Ernst	1971-12-08	1983-12-19
BVerfG	Zeidler	Wolfgang	1983-12-20	1987-11-15
BVerfG	Herzog	Roman	1987-11-16	1994-06-30
BVerfG	VACANCY-3	VACANCY-3	1994-07-01	1994-09-13
BVerfG	Limbach	Jutta	1994-09-14	2002-04-09
BVerfG	Papier	Hans-Jürgen	2002-04-10	2010-03-15
BVerfG	Voßkuhle	Andreas	2010-03-16	2020-06-21
BVerfG	Harbarth	Stephan	2020-06-22	NA

9.8.3 Hypothetisches Amtsende für PräsidentIn

Weil der/die aktuelle PräsidentIn noch im Amt ist, ist der Wert für das Amtsende "NA". Dieser ist aber für die verwendete Logik nicht greifbar, weshalb an dieser Stelle ein hypothetisches Amtsende in einem Jahr ab dem Tag der Datensatzerstellung fingiert wird. Es wird nur an dieser Stelle verwendet und danach verworfen.

```
praesi[is.na(term_end_date)]$term_end_date <- Sys.Date() + 365
```

9.8.4 Schleife vorbereiten

```
N <- praesi[,.N]
praesi.list <- vector("list", N)
```

9.8.5 Vektor erstellen

```
for (i in seq_len(N)){
  praesi.N <- txt.bverfg[datum >= praesi$term_begin_date[i] & datum <= praesi$
  term_end_date[i], .N]
  praesi.list[[i]] <- rep(praesiname_last[i],
    praesi.N)
}
```

9.8.6 Vektor einfügen

```
txt.bverfg$praesi <- unlist(praesiname.list)
```

9.9 Variable "v_praesi" hinzufügen

Diese Variable dokumentiert für jede Entscheidung welche/r Vize-PräsidentIn am Tag der Entscheidung im Amt war.

9.9.1 Personaldaten einlesen

```
vpraesi <- fread(file.path("data",
  "PVP-FCG_2021-04-08_GermanFederalCourts_VicePresidents.
  csv"))
vpraesi <- vpraesi[court == "BVerfG", c(1:3, 5:6)]
```

9.9.2 Personaldaten anzeigen

```
kable(vpraesi,  
      format = "latex",  
      align = "r",  
      booktabs = TRUE,  
      longtable = TRUE) %>% kable_styling(latex_options = "repeat_header")
```

court	name_last	name_first	term_begin_date	term_end_date
BVerfG	Katz	Rudolf	1951-09-07	1961-07-23
BVerfG	Wagner	Friedrich Wilhelm	1961-12-19	1967-10-18
BVerfG	Seuffert	Walter	1967-10-18	1975-11-07
BVerfG	Zeidler	Wolfgang	1975-11-07	1983-12-20
BVerfG	Herzog	Roman	1983-12-20	1987-11-16
BVerfG	Mahrenholz	Ernst Gottfried	1987-11-16	1994-03-24
BVerfG	Limbach	Jutta	1994-03-24	1994-09-14
BVerfG	Henschel	Johann Friedrich	1994-09-29	1995-10-13
BVerfG	Seidl	Otto	1995-10-13	1998-02-27
BVerfG	Papier	Hans-Jürgen	1998-02-27	2002-04-10
BVerfG	Hassemer	Winfried	2002-04-10	2008-05-07
BVerfG	Voßkuhle	Andreas	2008-05-07	2010-03-16
BVerfG	Kirchhof	Ferdinand	2010-03-16	2018-11-30
BVerfG	Harbarth	Stephan	2018-11-30	2020-06-22
BVerfG	König	Doris	2020-06-22	NA

9.9.3 Hypothetisches Amtsende für Vize-PräsidentIn

Weil der/die aktuelle Vize-PräsidentIn noch im Amt ist, ist der Wert für das Amtsende "NA". Dieser ist aber für die verwendete Logik nicht greifbar, weshalb an dieser Stelle ein hypothetisches Amtsende in einem Jahr ab dem Tag der Datensatzerstellung fingiert wird. Es wird nur an dieser Stelle verwendet und danach verworfen.

```
vpraesi[is.na(term_end_date)]$term_end_date <- Sys.Date() + 365
```

9.9.4 Schleife vorbereiten

```
N <- vpraesi[,.N]
vpraesi.list <- vector("list", N)
```

9.9.5 Vektor erstellen

```
for (i in seq_len(N)){
  vpraesi.N <- txt.bverfg[datum >= vpraesi$term_begin_date[i] & datum < vpraesi
    $term_end_date[i], .N]
  vpraesi.list[[i]] <- rep(vpraesi$name_last[i],
    vpraesi.N)
}
```

9.9.6 Vektor einfügen

```
txt.bverfg$v_praesi <- unlist(vpraesi.list)
```

9.10 Variable “verfahrensart” hinzufügen

Die Registerzeichen werden an dieser Stelle mit ihren detaillierten Bedeutungen aus dem folgenden Datensatz abgeglichen: “Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.” Das Ergebnis des Abgleichs wird in der Variable “verfahrensart” in den Datensatz eingefügt.

9.10.1 Datensatz einlesen

```
az.source <- fread(file.path("data",
  "AZ-BRD_1-0-1_DE_Registerzeichen_Datensatz.csv"))
```

9.10.2 Datensatz auf relevante Daten reduzieren

```
az.bverfg <- az.source[stelle == "BVerfG" & position == "hauptzeichen"]
```

9.10.3 Indizes bestimmen

```
targetindices <- match(txt.bverfg$registerzeichen,
  az.bverfg$zeichen_code)
```

9.10.4 Vektor der Verfahrensarten erstellen und einfügen

```
txt.bverfg$verfahrensart <- az.bverfg$bedeutung[targetindices]
```

9.11 Variable “aktenzeichen” hinzufügen

```
txt.bverfg$aktenzeichen <- paste0(txt.bverfg$spruchkoerper_az,  
    " ",  
    txt.bverfg$registerzeichen,  
    " ",  
    txt.bverfg$eingangsnummer,  
    "/",  
    txt.bverfg$eingangsjahr_az)
```

Bei Entscheidungen der Verzögerungskammer fehlt das Spruchkörper-Element des Aktenzeichens. Diese Zeile entfernt die “NA”-Angabe um ein korrektes Aktenzeichen herzustellen.

```
txt.bverfg$aktenzeichen <- gsub("NA ",  
    "",  
    txt.bverfg$aktenzeichen)
```

9.12 Variable “doi_concept” hinzufügen

```
txt.bverfg$doi_concept <- rep(config$doi$data$conzept,  
    txt.bverfg[,.N])
```

9.13 Variable “doi_version” hinzufügen

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

9.14 Variable “version” hinzufügen

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

9.15 Variable “lizenz” hinzufügen

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

9.16 Variable “ecli” hinzufügen

Struktur und Inhalt der ECLI für deutsche Gerichte sind auf dem Europäischen Justizportal näher erläutert.²

Sofern die Variablen korrekt extrahiert wurden lässt sich die ECLI vollständig rekonstruieren.

```
ecli.ordinalzahl <- paste0(gsub("Bv([A-Z])",
                              "\\1",
                              txt.bverfg$registerzeichen),
                          txt.bverfg$spruchkoerper_typ,
                          txt.bverfg$datum,
                          txt.bverfg$kollision,
                          ".",
                          txt.bverfg$spruchkoerper_az,
                          txt.bverfg$registerzeichen,
                          formatC(txt.bverfg$eingangsnummer,
                                   width = 4,
                                   flag = "0"),
                          formatC(txt.bverfg$eingangsjahr_az,
                                   width = 2,
                                   flag = "0"))

ecli.ordinalzahl <- gsub("NA",
                       "",
                       ecli.ordinalzahl)

ecli.ordinalzahl <- gsub("-",
                       "",
                       ecli.ordinalzahl)

ecli.ordinalzahl <- tolower(ecli.ordinalzahl)

ecli.ordinalzahl <- gsub("vzb",
                       "vb",
                       ecli.ordinalzahl)

ecli.ordinalzahl <- gsub("pup",
                       "up",
                       ecli.ordinalzahl)

txt.bverfg$ecli <- paste0("ECLI:DE:BVerfG:",
                          txt.bverfg$entscheidungsjahr,
                          ":",
                          ecli.ordinalzahl)
```

² https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-de-de.do?member=1

9.16.1 Metadaten mit ECLI-Duplikaten entfernen

Für eine Entscheidung wird versehentlich auch die englischsprachige Zusammenfassung abgerufen, diese wird hier entfernt.

```
dt.meta.html <- dt.meta.html[grep("Order",
  dt.meta.html$zitativvorschlag,
  invert = TRUE)]
```

9.16.2 Fehlerhafte ECLI korrigieren (HTML Meta)

```
dt.meta.html$ecli <- gsub("ECLI:DE:BVerfG:1951:rs19580115.1bvr040051",
  "ECLI:DE:BVerfG:1958:rs19580115.1bvr040051",
  dt.meta.html$ecli) ## Lüth

dt.meta.html$ecli <- gsub("ECLI:DE:BVerfG:2003:rk20030407.2bvr212902",
  "ECLI:DE:BVerfG:2003:rk20030407.1bvr212902",
  dt.meta.html$ecli)

dt.meta.html$ecli <- gsub("ECLI:DE:BVerfG:2007:rk20060529.1bvr043003",
  "ECLI:DE:BVerfG:2006:rk20060529.1bvr043003",
  dt.meta.html$ecli)
```

Die folgende ECLI ist auf der Homepage des BVerfG fehlerhaft. Sie betrifft das Vorverfahren statt die Verzögerungsbeschwerde. Auf rechtsprechung-im-internet.de ist sie korrekt nachgewiesen.³

```
dt.meta.html$ecli <- gsub("ECLI:DE:BVerfG:2015:rs20151208a.1bvr009911",
  "ECLI:DE:BVerfG:2015:vb20151208.vz000115",
  dt.meta.html$ecli)
```

9.16.3 Fehlerhafte ECLI korrigieren (HTML Full)

```
dt.segmented.full$ecli <- gsub("ECLI:DE:BVerfG:1951:rs19580115.1bvr040051",
  "ECLI:DE:BVerfG:1958:rs19580115.1bvr040051",
  dt.segmented.full$ecli) ## Lüth

dt.segmented.full$ecli <- gsub("ECLI:DE:BVerfG:2003:rk20030407.2bvr212902",
  "ECLI:DE:BVerfG:2003:rk20030407.1bvr212902",
  dt.segmented.full$ecli)

dt.segmented.full$ecli <- gsub("ECLI:DE:BVerfG:2007:rk20060529.1bvr043003",
  "ECLI:DE:BVerfG:2006:rk20060529.1bvr043003",
  dt.segmented.full$ecli)
```

³ <https://www.rechtsprechung-im-internet.de/jportal/?quelle=jlink&docid=KVRE412291501&psml=bsjrsprod.psml&max=true>

Die folgende ECLI ist auf der Homepage des BVerfG fehlerhaft. Sie betrifft das Vorverfahren statt die Verzögerungsbeschwerde. Auf [rechtsprechung-im-internet.de](https://www.rechtsprechung-im-internet.de) ist sie korrekt nachgewiesen.⁴

```
dt.segmented.full$ecli <- gsub("ECLI:DE:BVerfG:2015:rs20151208a.1bvr009911",  
                             "ECLI:DE:BVerfG:2015:vb20151208.vz000115",  
                             dt.segmented.full$ecli)
```

9.16.4 ECLI-Test 1: ECLI die in PDF, aber nicht in HTML vorhanden sind

```
sort(setdiff(txt.bverfg$ecli, dt.segmented.full$ecli))
```

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

9.16.5 ECLI-Test 2: ECLI die in HTML, aber nicht in PDF vorhanden sind

```
sort(setdiff(dt.segmented.full$ecli, txt.bverfg$ecli))
```

```
## [1] "ECLI:DE:BVerfG:2001:bs20011001.2bvb000101"  
## [2] "ECLI:DE:BVerfG:2001:rk20011023.2bvr123601"  
## [3] "ECLI:DE:BVerfG:2004:rk20040519.1bvr071104"  
## [4] "ECLI:DE:BVerfG:2014:rk20140319.1bvr141710"  
## [5] "ECLI:DE:BVerfG:2014:rk20140610.1bvr066914"  
## [6] "ECLI:DE:BVerfG:2021:fs20210325.2bvf000120"
```

9.16.6 Zum Vergleich: Beim Download fehlende PDF-Dateien

```
print(sort(missing.pdf))
```

```
## [1] "BVerfG_2001-10-01_S_2_BvB_0001_01_NA_NPD-Verbot-EA-RückgabeEDV-2_104_61.  
## [2] "BVerfG_2001-10-23_K_2_BvR_1236_01_NA_NA_NA_NA.pdf"  
## [3] "BVerfG_2004-05-19_K_1_BvR_0711_04_NA_NA_NA_NA.pdf"  
## [4] "BVerfG_2014-03-19_K_1_BvR_1417_10_NA_NA_NA_NA.pdf"  
## [5] "BVerfG_2014-06-10_K_1_BvR_0669_14_NA_NA_NA_NA.pdf"  
## [6] "BVerfG_2021-03-25_S_2_BvF_0001_20_NA_NA_NA_NA.pdf"
```

⁴ <https://www.rechtsprechung-im-internet.de/jportal/?quelle=jlink&docid=KVRE412291501&psml=bsjrsprod.psml&max=true>

9.16.7 ECLI Merge: Metadaten aus Hauptdatensatz in segmentierte Variante mergen

```
meta.bverfg <- txt.bverfg[,!"text"]

dt.segmented.full <- merge(dt.segmented.full,
                           meta.bverfg,
                           by = "ecli",
                           all.x = TRUE,
                           sort = FALSE)
```

9.16.8 ECLI Merge: Metadaten aus HTML-Extraktion in Hauptdatensatz mergen

```
txt.bverfg <- merge(txt.bverfg,
                   dt.meta.html,
                   by = "ecli",
                   all.x = TRUE,
                   sort = FALSE)
```

9.17 Variable “entscheidung_typ” hinzufügen

9.17.1 Zitiervorschläge parsen

```
entscheidung_typ.main <- gsub(".*(Beschluss|Urteil|Verfügung).*",
                             "\\1",
                             txt.bverfg$zitiervorschlag,
                             ignore.case = TRUE)

entscheidung_typ.segmented <- gsub(".*(Beschluss|Urteil|Verfügung).*",
                                   "\\1",
                                   dt.segmented.full$zitiervorschlag,
                                   ignore.case = TRUE)
```

9.17.2 Kürzen

```
lang.etyf <- c("Urteil",
              "Beschluss",
              "Verfügung")

kurz.etyf <- c("Ü",
              "B",
              "V")

entscheidung_typ.main <- mgsub(entscheidung_typ.main,
```

```
        lang.etyp,  
        kurz.etyp,  
        ignore.case = TRUE)  
entscheidung_typ.segmented <- mgsub(entscheidung_typ.segmented,  
        lang.etyp,  
        kurz.etyp,  
        ignore.case = TRUE)
```

9.17.3 Vektor in Datensatz einfügen

```
txt.bverfg$entscheidung_typ <- entscheidung_typ.main  
dt.segmented.full$entscheidung_typ <- entscheidung_typ.segmented
```

10 Frequenztabellen erstellen

10.1 Funktion anzeigen

```
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)
}

}

```

10.2 Ignorierte Variablen

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

```
## [1] "text"           "eingangsnummer" "datum"           "doc_id"
## [5] "seite"          "name"            "ecli"            "aktenzeichen"
## [9] "pressemitteilung" "zitiervorschlag" "kurzbeschreibung"
```

10.3 Liste zu prüfender Variablen

```
varlist <- names(txt.bverfg)
varlist <- grep(paste(config$freqtable$ignore,
                      collapse = "|"),
               varlist,
               invert = TRUE,
               value = TRUE)
print(varlist)
```

```
## [1] "gericht"          "spruchkoerper_typ" "spruchkoerper_az"
## [4] "registerzeichen"  "eingangsjahr_az"  "kollision"
## [7] "band"            "entscheidungsjahr" "eingangsjahr_iso"
## [10] "praesi"          "v_praesi"         "verfahrensart"
## [13] "doi_concept"     "doi_version"      "version"
## [16] "lizenz"          "richter"          "entscheidung_typ"
```

10.4 Frequenztabellen erstellen

```
prefix <- paste0(config$project$shortname,
                 "_01_Frequenztafel_var-")
```

```
f.fast.freqtable(txt.bverfg,
                 varlist = varlist,
                 sumrow = TRUE,
                 output.list = FALSE,
                 output.kable = TRUE,
                 output.csv = TRUE,
                 outputdir = dir.analysis,
                 prefix = prefix,
                 align = c("p{5cm}",
                          rep("r", 4)))
```

Frequency Table for Variable: gericht

1 unique value(s) detected.

gericht	N	exactpercent	roundedpercent	cumulpercent
BVerfG	8213	100	100	100
Total	8213	100	100	100

Frequency Table for Variable: spruchkoerper_typ

4 unique value(s) detected.

spruchkoerper_typ	N	exactpercent	roundedpercent	cumulpercent
B	7	0.0852307	0.09	0.09
K	6827	83.1243151	83.12	83.21

(continued)

spruchkoerper_typ	N	exactpercent	roundedpercent	cumulpercent
P	2	0.0243516	0.02	83.23
S	1377	16.7661025	16.77	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: spruchkoerper_az

3 unique value(s) detected.

spruchkoerper_az	N	exactpercent	roundedpercent	cumulpercent
NA	7	0.0852307	0.09	0.09
1	4210	51.2601972	51.26	51.35
2	3996	48.6545720	48.65	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: registerzeichen

15 unique value(s) detected.

registerzeichen	N	exactpercent	roundedpercent	cumulpercent
BvB	23	0.2800438	0.28	0.28
BvC	260	3.1657129	3.17	3.45
BvE	114	1.3880433	1.39	4.83
BvF	69	0.8401315	0.84	5.67
BvG	11	0.1339340	0.13	5.81
BvH	8	0.0974066	0.10	5.91
BvK	15	0.1826373	0.18	6.09
BvL	349	4.2493608	4.25	10.34
BvM	4	0.0487033	0.05	10.39
BvN	1	0.0121758	0.01	10.40

(continued)

registerzeichen	N	exactpercent	roundedpercent	cumulpercent
BvP	3	0.0365275	0.04	10.43
BvQ	456	5.5521734	5.55	15.99
BvR	6891	83.9035675	83.90	99.89
PBvU	2	0.0243516	0.02	99.91
Vz	7	0.0852307	0.09	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: eingangsjahr_az

46 unique value(s) detected.

eingangsjahr_az	N	exactpercent	roundedpercent	cumulpercent
0	358	4.3589431	4.36	4.36
1	427	5.1990746	5.20	9.56
2	333	4.0545477	4.05	13.61
3	320	3.8962620	3.90	17.51
4	283	3.4457567	3.45	20.95
5	272	3.3118227	3.31	24.27
6	278	3.3848776	3.38	27.65
7	306	3.7258006	3.73	31.38
8	345	4.2006575	4.20	35.58
9	344	4.1884817	4.19	39.77
10	304	3.7014489	3.70	43.47
11	326	3.9693169	3.97	47.44
12	260	3.1657129	3.17	50.60
13	272	3.3118227	3.31	53.91
14	322	3.9206137	3.92	57.84
15	280	3.4092293	3.41	61.24
16	307	3.7379764	3.74	64.98

(continued)

eingangsjahr_az	N	exactpercent	roundedpercent	cumulpercent
17	330	4.0180202	4.02	69.00
18	225	2.7395592	2.74	71.74
19	311	3.7866797	3.79	75.53
20	329	4.0058444	4.01	79.53
21	156	1.8994277	1.90	81.43
22	2	0.0243516	0.02	81.46
51	2	0.0243516	0.02	81.48
52	2	0.0243516	0.02	81.50
56	1	0.0121758	0.01	81.52
58	1	0.0121758	0.01	81.53
77	1	0.0121758	0.01	81.54
81	1	0.0121758	0.01	81.55
83	3	0.0365275	0.04	81.59
84	3	0.0365275	0.04	81.63
85	4	0.0487033	0.05	81.68
86	7	0.0852307	0.09	81.76
87	9	0.1095824	0.11	81.87
88	18	0.2191647	0.22	82.09
89	14	0.1704615	0.17	82.26
90	36	0.4383295	0.44	82.70
91	52	0.6331426	0.63	83.33
92	63	0.7670766	0.77	84.10
93	84	1.0227688	1.02	85.12
94	143	1.7411421	1.74	86.86
95	137	1.6680872	1.67	88.53
96	179	2.1794716	2.18	90.71
97	189	2.3012298	2.30	93.01
98	253	3.0804822	3.08	96.09
99	321	3.9084378	3.91	100.00

(continued)

eingangsjahr_az	N	exactpercent	roundedpercent	cumulpercent
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: kollision

5 unique value(s) detected.

kollision	N	exactpercent	roundedpercent	cumulpercent
NA	8171	99.4886156	99.49	99.49
1	1	0.0121758	0.01	99.50
a	36	0.4383295	0.44	99.94
b	4	0.0487033	0.05	99.99
c	1	0.0121758	0.01	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: band

68 unique value(s) detected.

band	N	exactpercent	roundedpercent	cumulpercent
NA	7482	91.0994764	91.10	91.10
1	2	0.0243516	0.02	91.12
7	1	0.0121758	0.01	91.14
62	1	0.0121758	0.01	91.15
65	1	0.0121758	0.01	91.16
90	1	0.0121758	0.01	91.17
91	1	0.0121758	0.01	91.18
93	1	0.0121758	0.01	91.20
96	5	0.0608791	0.06	91.26
97	18	0.2191647	0.22	91.48

(continued)

band	N	exactpercent	roundedpercent	cumulpercent
98	14	0.1704615	0.17	91.65
99	26	0.3165713	0.32	91.96
100	14	0.1704615	0.17	92.13
101	15	0.1826373	0.18	92.32
102	17	0.2069889	0.21	92.52
103	19	0.2313406	0.23	92.76
104	26	0.3165713	0.32	93.07
105	14	0.1704615	0.17	93.24
106	17	0.2069889	0.21	93.45
107	14	0.1704615	0.17	93.62
108	18	0.2191647	0.22	93.84
109	12	0.1461098	0.15	93.99
110	14	0.1704615	0.17	94.16
111	15	0.1826373	0.18	94.34
112	18	0.2191647	0.22	94.56
113	10	0.1217582	0.12	94.68
114	13	0.1582856	0.16	94.84
115	12	0.1461098	0.15	94.98
116	11	0.1339340	0.13	95.12
117	14	0.1704615	0.17	95.29
118	9	0.1095824	0.11	95.40
119	10	0.1217582	0.12	95.52
120	10	0.1217582	0.12	95.64
121	10	0.1217582	0.12	95.76
122	13	0.1582856	0.16	95.92
123	9	0.1095824	0.11	96.03
124	12	0.1461098	0.15	96.18
125	7	0.0852307	0.09	96.26
126	13	0.1582856	0.16	96.42

(continued)

band	N	exactpercent	roundedpercent	cumulpercent
127	10	0.1217582	0.12	96.54
128	13	0.1582856	0.16	96.70
129	12	0.1461098	0.15	96.85
130	11	0.1339340	0.13	96.98
131	11	0.1339340	0.13	97.11
132	13	0.1582856	0.16	97.27
133	14	0.1704615	0.17	97.44
134	14	0.1704615	0.17	97.61
135	11	0.1339340	0.13	97.75
136	11	0.1339340	0.13	97.88
137	9	0.1095824	0.11	97.99
138	12	0.1461098	0.15	98.14
139	10	0.1217582	0.12	98.26
140	12	0.1461098	0.15	98.40
141	8	0.0974066	0.10	98.50
142	13	0.1582856	0.16	98.66
143	8	0.0974066	0.10	98.76
144	4	0.0487033	0.05	98.81
145	9	0.1095824	0.11	98.92
146	7	0.0852307	0.09	99.00
147	9	0.1095824	0.11	99.11
148	10	0.1217582	0.12	99.23
149	13	0.1582856	0.16	99.39
150	8	0.0974066	0.10	99.49
151	9	0.1095824	0.11	99.60
152	12	0.1461098	0.15	99.74
153	6	0.0730549	0.07	99.82
154	7	0.0852307	0.09	99.90
155	8	0.0974066	0.10	100.00

(continued)

band	N	exactpercent	roundedpercent	cumulpercent
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: entscheidungsjahr

39 unique value(s) detected.

entscheidungsjahr	N	exactpercent	roundedpercent	cumulpercent
1951	1	0.0121758	0.01	0.01
1952	2	0.0243516	0.02	0.04
1957	1	0.0121758	0.01	0.05
1958	1	0.0121758	0.01	0.06
1959	1	0.0121758	0.01	0.07
1977	1	0.0121758	0.01	0.09
1983	2	0.0243516	0.02	0.11
1991	1	0.0121758	0.01	0.12
1992	22	0.2678680	0.27	0.39
1993	22	0.2678680	0.27	0.66
1994	34	0.4139778	0.41	1.07
1995	31	0.3774504	0.38	1.45
1996	40	0.4870328	0.49	1.94
1997	74	0.9010106	0.90	2.84
1998	322	3.9206137	3.92	6.76
1999	367	4.4685255	4.47	11.23
2000	418	5.0894923	5.09	16.32
2001	442	5.3817119	5.38	21.70
2002	419	5.1016681	5.10	26.80
2003	312	3.7988555	3.80	30.60
2004	321	3.9084378	3.91	34.51
2005	266	3.2387678	3.24	37.75

(continued)

entscheidungsjahr	N	exactpercent	roundedpercent	cumulpercent
2006	319	3.8840862	3.88	41.63
2007	247	3.0074272	3.01	44.64
2008	365	4.4441739	4.44	49.08
2009	363	4.4198222	4.42	53.50
2010	329	4.0058444	4.01	57.51
2011	298	3.6283940	3.63	61.13
2012	284	3.4579325	3.46	64.59
2013	208	2.5325703	2.53	67.13
2014	229	2.7882625	2.79	69.91
2015	318	3.8719104	3.87	73.79
2016	397	4.8338001	4.83	78.62
2017	361	4.3954706	4.40	83.01
2018	276	3.3605260	3.36	86.38
2019	275	3.3483502	3.35	89.72
2020	510	6.2096676	6.21	95.93
2021	327	3.9814928	3.98	99.91
2022	7	0.0852307	0.09	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: eingangsjahr_iso

46 unique value(s) detected.

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
1951	2	0.0243516	0.02	0.02
1952	2	0.0243516	0.02	0.05
1956	1	0.0121758	0.01	0.06
1958	1	0.0121758	0.01	0.07
1977	1	0.0121758	0.01	0.09

(continued)

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
1981	1	0.0121758	0.01	0.10
1983	3	0.0365275	0.04	0.13
1984	3	0.0365275	0.04	0.17
1985	4	0.0487033	0.05	0.22
1986	7	0.0852307	0.09	0.30
1987	9	0.1095824	0.11	0.41
1988	18	0.2191647	0.22	0.63
1989	14	0.1704615	0.17	0.80
1990	36	0.4383295	0.44	1.24
1991	52	0.6331426	0.63	1.88
1992	63	0.7670766	0.77	2.64
1993	84	1.0227688	1.02	3.66
1994	143	1.7411421	1.74	5.41
1995	137	1.6680872	1.67	7.07
1996	179	2.1794716	2.18	9.25
1997	189	2.3012298	2.30	11.55
1998	253	3.0804822	3.08	14.64
1999	321	3.9084378	3.91	18.54
2000	358	4.3589431	4.36	22.90
2001	427	5.1990746	5.20	28.10
2002	333	4.0545477	4.05	32.16
2003	320	3.8962620	3.90	36.05
2004	283	3.4457567	3.45	39.50
2005	272	3.3118227	3.31	42.81
2006	278	3.3848776	3.38	46.20
2007	306	3.7258006	3.73	49.92
2008	345	4.2006575	4.20	54.12
2009	344	4.1884817	4.19	58.31
2010	304	3.7014489	3.70	62.01

(continued)

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
2011	326	3.9693169	3.97	65.98
2012	260	3.1657129	3.17	69.15
2013	272	3.3118227	3.31	72.46
2014	322	3.9206137	3.92	76.38
2015	280	3.4092293	3.41	79.79
2016	307	3.7379764	3.74	83.53
2017	330	4.0180202	4.02	87.54
2018	225	2.7395592	2.74	90.28
2019	311	3.7866797	3.79	94.07
2020	329	4.0058444	4.01	98.08
2021	156	1.8994277	1.90	99.98
2022	2	0.0243516	0.02	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: praesi

10 unique value(s) detected.

praesi	N	exactpercent	roundedpercent	cumulpercent
Benda	3	0.0365275	0.04	0.04
Harbarth	587	7.1472056	7.15	7.18
Herzog	62	0.7549008	0.75	7.94
Höpker-Aschoff	3	0.0365275	0.04	7.98
Limbach	1845	22.4643857	22.46	30.44
Müller	1	0.0121758	0.01	30.45
Papier	2544	30.9752831	30.98	61.43
VACANCY-3	1	0.0121758	0.01	61.44
Voßkuhle	3165	38.5364666	38.54	99.98
Wintrich	2	0.0243516	0.02	100.00

(continued)

praesi	N	exactpercent	roundedpercent	cumulpercent
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: v_praesi

12 unique value(s) detected.

v_praesi	N	exactpercent	roundedpercent	cumulpercent
Harbarth	548	6.6723487	6.67	6.67
Hassemer	1908	23.2314623	23.23	29.90
Henschel	44	0.5357360	0.54	30.44
Katz	6	0.0730549	0.07	30.51
Kirchhof	2617	31.8641179	31.86	62.38
König	587	7.1472056	7.15	69.52
Limbach	8	0.0974066	0.10	69.62
Mahrenholz	55	0.6696700	0.67	70.29
Papier	1624	19.7735298	19.77	90.06
Seidl	177	2.1551199	2.16	92.22
Voßkuhle	636	7.7438208	7.74	99.96
Zeidler	3	0.0365275	0.04	100.00
Total	8213	100.0000000	100.00	100.00

Frequency Table for Variable: verfahrensart

15 unique value(s) detected.

verfahrensart	N	exactpercent	roundedpercent	cumulpercent
Abstrakte Normenkontrolle	69	0.8401315	0.84	0.84
Andere Streitigkeiten zwischen Bund und Ländern	8	0.0974066	0.10	0.94

(continued)

verfahrensart	N	exactpercent	roundedpercent	cumulpercent
Bund-Länder-Streitigkeiten	11	0.1339340	0.13	1.07
Divergenzvorlagen eines Landesverfassungsgerichts zur Auslegung des Grundgesetzes	1	0.0121758	0.01	1.08
Einstweilige Anordnungen	456	5.5521734	5.55	6.64
Feststellung der Anwendbarkeit einer Regel des Völkerge- wohnheitsrechts	4	0.0487033	0.05	6.68
Konkrete Normenkontrolle	349	4.2493608	4.25	10.93
Landesverfassungsstreitigkeiten	15	0.1826373	0.18	11.12
Organstreitverfahren	114	1.3880433	1.39	12.50
Plenarentscheidungen	2	0.0243516	0.02	12.53
Sonstige durch Bundesrecht zugewiesene Verfahren	3	0.0365275	0.04	12.57
Verfassungsbeschwerden; Kommunalverfassungsbe- schwerden	6891	83.9035675	83.90	96.47
Verfassungswidrigkeit von Par- teien	23	0.2800438	0.28	96.75
Verzögerungsbeschwerde	7	0.0852307	0.09	96.83
Wahlprüfungsverfahren	260	3.1657129	3.17	100.00
Total	8213	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.3902658	8213	100	100	100
Total	8213	100	100	100

Frequency Table for Variable: doi_version

1 unique value(s) detected.

doi_version	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.5910152	8213	100	100	100
Total	8213	100	100	100

Frequency Table for Variable: version

1 unique value(s) detected.

version	N	exactpercent	roundedpercent	cumulpercent
2022-02-01	8213	100	100	100
Total	8213	100	100	100

Frequency Table for Variable: lizenz

1 unique value(s) detected.

lizenz	N	exactpercent	roundedpercent	cumulpercent
Creative Commons Zero 1.0 Universal	8213	100	100	100
Total	8213	100	100	100

Frequency Table for Variable: richter

484 unique value(s) detected.

richter	N	exactpercent	roundedpercent	cumulpercent
	23	0.2800438	0.28	0.28

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
<us> Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt		0.0121758	0.01	0.29
Baer Britz Müller Kessal-Wulf	1	0.0121758	0.01	0.30
Baer Britz Radtke	11	0.1339340	0.13	0.44
Baer Ott Radtke	58	0.7061975	0.71	1.14
Benda Simon Hesse Katzenstein Niemeyer Hübner Henschel Niederhoffer Britz Ott Radtke	4	0.0487033	0.05	1.21
Broß Di Fabio Gerhardt	24	0.2922197	0.29	1.50
Broß Di Fabio Landau	117	1.4245708	1.42	2.92
Broß Di Fabio Lübbe-Wolff	11	0.1339340	0.13	3.06
Broß Di Fabio Osterloh	1	0.0121758	0.01	3.07
Broß Lübbe-Wolff Gerhardt	90	1.0958237	1.10	4.16
Broß Lübbe-Wolff Landau	1	0.0121758	0.01	4.18
Broß Mellinghoff Gerhardt	1	0.0121758	0.01	4.19
Broß Mellinghoff Landau	2	0.0243516	0.02	4.21
Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau	1	0.0121758	0.01	4.23
Broß Osterloh Gerhardt	1	0.0121758	0.01	4.24
Broß Osterloh Landau	2	0.0243516	0.02	4.26
Broß Osterloh Mellinghoff	100	1.2175819	1.22	5.48
Bryde Eichberger Schluckebier	37	0.4505053	0.45	5.93
Bryde Schluckebier Paulus	1	0.0121758	0.01	5.94
Bundesverfassungsrichter Lim- bach Kirchhof Sommer Jentsch Hassemer Broß Osterloh Winter ist an der Unter- schrift verhindert.	1	0.0121758	0.01	5.95
Böckenförde Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0121758	0.01	6.00
Böckenförde Kruis Sommer	2	0.0243516	0.02	6.03

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Böckenförde Winter Sommer	1	0.0121758	0.01	6.04
Di Fabio Gerhardt Hermanns	17	0.2069889	0.21	6.25
Di Fabio Gerhardt Landau	1	0.0121758	0.01	6.26
Di Fabio Lübbe-Wolff Landau	2	0.0243516	0.02	6.28
Di Fabio Mellinghoff Gerhardt	1	0.0121758	0.01	6.29
Dr. Benda Dr. Böhmer Dr. Simon Dr. Faller Dr. Hesse Dr. Katzenstein	1	0.0121758	0.01	6.31
Dr. Aschoff Ellinghaus Dr. Heiland Dr. Drath Dr. Stein Wessel Ritterspach Lehmann Dr. Zweigert	1	0.0121758	0.01	6.32
Dr. Aschoff Ellinghaus Dr. Scheffler Dr. Heiland Dr. Scholtissek Dr. Drath Wessel Ritterspach Lehmann Dr. Zweigert	1	0.0121758	0.01	6.33
Dr. Katz Dr. Fröhlich Wolff Dr. Roediger Dr. Klaas Henneka Dr. Friesenhahn Dr. Rupp Dr. Geiger Leuser Dr. Federer	1	0.0121758	0.01	6.34
Dr. Müller Dr. Scheffler Dr. Heiland Dr. Heck Dr. Berger Dr. Scholtissek Dr. Drath Dr. Stein Ritterspach	1	0.0121758	0.01	6.36
Dr. Wintrich Dr. Scheffler Dr. Heiland Dr. Heck Dr. Scholtissek Dr. Stein Wessel Ritterspach Lehmann	1	0.0121758	0.01	6.37
Eichberger Baer Britz	87	1.0592962	1.06	7.43
Eichberger Masing Baer	2	0.0243516	0.02	7.45
Eichberger Masing Britz	1	0.0121758	0.01	7.46
Eichberger Masing Paulus Baer Britz Ott Ott	4	0.0483516	0.01	7.48
Eichberger Paulus Britz	2	0.0243516	0.02	7.50

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Eichberger Schluckebier Britz	1	0.0121758	0.01	7.51
Eichberger Schluckebier Masing	4	0.0487033	0.05	7.56
Gaier Eichberger Britz	2	0.0243516	0.02	7.59
Gaier Eichberger Kirchhof	1	0.0121758	0.01	7.60
Gaier Eichberger Paulus	1	0.0121758	0.01	7.61
Gaier Eichberger Schluckebier	2	0.0243516	0.02	7.63
Gaier Eichberger Schluckebier Masing Paulus Baer Britz	2	0.0243516	0.02	7.66
Gaier Eichberger Schluckebier Masing Paulus Hermanns Baer Britz	1	0.0121758	0.01	7.67
Gaier Masing Baer	13	0.1582856	0.16	7.83
Gaier Paulus Baer	1	0.0121758	0.01	7.84
Gaier Paulus Britz	93	1.1323512	1.13	8.97
Gaier Schluckebier Baer	7	0.0852307	0.09	9.06
Gaier Schluckebier Britz	3	0.0365275	0.04	9.10
Gaier Schluckebier Kirchhof	1	0.0121758	0.01	9.11
Gaier Schluckebier Paulus	170	2.0698892	2.07	11.18
Gerhardt Gaier Eichberger Hermanns	2	0.0243516	0.02	11.20
Gerhardt Hermanns Müller	62	0.7549008	0.75	11.96
Gerhardt Huber Hermanns	1	0.0121758	0.01	11.97
Graßhof Kirchhof Jentsch	1	0.0121758	0.01	11.98
Graßhof Kruis Kirchhof Winter Sommer Steiner Jentsch Hassemer	2	0.0243516	0.02	12.01
Graßhof Kruis Winter	1	0.0121758	0.01	12.02
Graßhof Winter Sommer	2	0.0243516	0.02	12.04
Grimm Haas Seidl	7	0.0852307	0.09	12.13
Grimm Haas Steiner	1	0.0121758	0.01	12.14
Grimm Hömig	2	0.0243516	0.02	12.16
Grimm Hömig Seidl	6	0.0730549	0.07	12.24
Grimm Kühling Die Richterin Haas ist an der Unterschrift gehindert	1	0.0121758	0.01	12.25

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Grimm Kühling Jaeger Haas Hömig Steiner	1	0.0121758	0.01	12.26
Grimm Kühling Jaeger Haas Hömig Steiner Jochims	1	0.0487033	0.05	12.31
Grimm Kühling Seibert Jaeger Hömig Steiner	1	0.0121758	0.01	12.32
Haas Bryde Eichberger	1	0.0121758	0.01	12.33
Haas Bryde Eichberger	14	0.1704615	0.17	12.50
Haas Hömig Bryde	26	0.3165713	0.32	12.82
Haas Hömig Hohmann-Dennhardt	1	0.0121758	0.01	12.83
Harbarth Baer Ott	65	0.7914282	0.79	13.62
Harbarth Britz Ott	4	0.0487033	0.05	13.67
Harbarth Britz Radtko	139	1.6924388	1.69	15.37
Harbarth Christ Radtko	1	0.0121758	0.01	15.38
Harbarth Masing Paulus Baer Britz Ott Christ Radtko	1	0.0121758	0.12	15.50
Harbarth Ott Radtko	1	0.0121758	0.01	15.51
Harbarth Paulus Baer Britz Ott Christ Radtko	1	0.0121758	0.01	15.52
Harbarth Paulus Baer Britz Ott Christ Radtko Eckert	1	0.0121758	0.18	15.71
Hassemer Broß Di Fabio	11	0.1339340	0.13	15.84
Hassemer Broß Di Fabio Lübbe-Wolff Gerhardt	1	0.0121758	0.01	15.85
Hassemer Broß Lübbe-Wolff	1	0.0121758	0.01	15.87
Hassemer Broß Mellinghoff	3	0.0365275	0.04	15.90
Hassemer Broß Osterloh	1	0.0121758	0.01	15.91
Hassemer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0121758	0.01	15.93
Hassemer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau	37	0.4505053	0.45	16.38
Hassemer Broß Osterloh Mellinghoff Lübbe-Wolff Gerhardt	1	0.0121758	0.01	16.39
Hassemer Di Fabio Gerhardt	3	0.0365275	0.04	16.43

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hassemer Di Fabio Landau	167	2.0333617	2.03	18.46
Hassemer Di Fabio Mellinghoff	2	0.0243516	0.02	18.48
Hassemer Jaeger Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0121758	0.01	18.50
Hassemer Jentsch Broß	1	0.0121758	0.01	18.51
Hassemer Jentsch Broß Di Fabio Mellinghoff Lübbe-Wolff	1	0.0121758	0.01	18.52
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	1	0.0121758	0.01	18.53
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhard	1	0.0121758	0.01	18.54
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	64	0.7792524	0.78	19.32
Hassemer Jentsch Broß Osterloh Mellinghoff Di Fabio Lübbe-Wolff Gerhardt	1	0.0121758	0.01	19.34
Hassemer Jentsch Richter Broß ist an der Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Unterschrift verhindert.	1	0.0121758	0.01	19.35
Hassemer Jentsch Richter Broß ist an Osterloh Richter Di Fabio ist Mellinghoff Lübbe-Wolff Gerhardt der Unterschrift verhindert. an	1	0.0121758	0.01	19.36
Hassemer Osterloh Lübbe-Wolff	1	0.0121758	0.01	19.37
Hassemer Osterloh Mellinghoff	266	3.2387678	3.24	22.61
Hassemer Sommer Broß	1	0.0121758	0.01	22.62

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hassemer Sommer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0121758	0.01	22.63
Hassemer Sommer Jentsch	7	0.0852307	0.09	22.72
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff	1	0.0487033	0.05	22.77
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.4139778	0.41	23.18
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Di Fabio	1	0.0121758	0.01	23.19
Hassemer Sommer Jentsch Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0121758	0.01	23.21
Hassemer Sommer Jentsch Richter Broß ist Osterloh Di Fabio Mellinghoff Lübbe-Wolff an der Unterschrift gehindert. Hassemer	1	0.0121758	0.01	23.22
Hassemer Sommer Lübbe-Wolff	1	0.0121758	0.01	23.23
Henschel Seidl Grimm Der Richter Söllner ist wegen Erkrankung an der Unterschrift verhindert Henschel Kühling Seibert Jaeger Haas	1	0.0121758	0.01	23.24
Henschel Seidl Grimm Dieterich Kühling Seibert Jaeger	1	0.0121758	0.01	23.26
Henschel Seidl Grimm Kühling Seibert Jaeger	1	0.0121758	0.01	23.27
Henschel Seidl Grimm Richter Söllner ist wegen einer Auslandsreise an der Unterschrift gehindert. Henschel Kühling Seibert Jaeger Haas	1	0.0121758	0.01	23.28
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger	1	0.0121758	0.01	23.29
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger	1	0.0121758	0.01	23.30
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.1948331	0.19	23.50

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Henschel Seidl Grimm Söllner Richter Dieterich ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehin- dert Kühling Seibert	1	0.0121758	0.01	23.51
Hentschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	6	0.0121758	0.01	23.52
Hermanns Kessal-Wulf König	1	0.0121758	0.01	23.54
Hermanns Kessal- Wulf Langenfeld	1	0.0121758	0.01	23.55
Hermanns König Maidowski	1	0.0121758	0.01	23.56
Hermanns Maidowski Langenfeld	47	0.5722635	0.57	24.13
Hermanns Müller Langenfeld	82	0.9984171	1.00	25.13
Hermanns Müller Maidowski	4	0.0487033	0.05	25.18
Herzog Henschel Böckenförde Seidl Grimm Dieterich Kühling Seibert	1	0.0121758	0.01	25.19
Herzog Henschel Der Richter Seidl ist durch Krankheit an der Unter- schrift gehindert. Her- zog Grimm Söllner Dieterich Kühling Seibert	1	0.0121758	0.01	25.20
Herzog Henschel Seidl Grimm Dieterich Kühling Seibert	1	0.0121758	0.01	25.22
Herzog Henschel Seidl Grimm Söllner Dieterich Kühling Seibert	25	0.3043555	0.30	25.52
Herzog Henschel Seidl Grimm Söllner Dieterich Seibert	3	0.0365275	0.04	25.56
Herzog Henschel Seidl Grimm Söllner Kühling Seibert	6	0.0730549	0.07	25.63
Herzog Henschel Seidl Grimm Söllner Kühling Seibert Jaeger	5	0.0608701	0.06	25.69
Hoffmann- Riem Bryde Schluckebier	1	0.0121758	0.01	25.70
Hoffmann- Riem Eichberger Schluckebier	2	0.0243516	0.02	25.73
Hohmann- Dennhardt Bryde Kirchhof	1	0.0121758	0.01	25.74
Hohmann- Dennhardt Eichberger Kirchhof	1	0.0121758	0.01	25.75
Hohmann- Dennhardt Eichberger Schluckebier	1	0.0121758	0.01	25.76

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hohmann-Dennhardt Gaier Eichberger	2	0.0243516	0.02	25.79
Hohmann-Dennhardt Gaier Kirchhof	159	1.9359552	1.94	27.72
Hohmann-Dennhardt Gaier Masing	1	0.0121758	0.01	27.74
Hohmann-Dennhardt Gaier Paulus	57	0.6940217	0.69	28.43
Hohmann-Dennhardt Gaier Schluckebier	1	0.0121758	0.01	28.44
Hohmann-Dennhardt Hoffmann-Riem Gaier	2	0.0243516	0.02	28.47
Huber Hermanns Müller	1	0.0121758	0.01	28.48
Huber Kessal-Wulf König	267	3.2509436	3.25	31.73
Huber Kessal-Wulf Langenfeld	1	0.0121758	0.01	31.74
Huber Kessal-Wulf Maidowski	3	0.0365275	0.04	31.78
Huber Kessal-Wulf Wallrabensein	1	0.0121758	0.01	31.79
Huber Kessal-Wulf Wallrabenstein	103	1.2541093	1.25	33.05
Huber König Langenfeld	1	0.0121758	0.01	33.06
Huber König Maidowski	1	0.0121758	0.01	33.07
Huber Langenfeld Wallrabenstein	3	0.0365275	0.04	33.11
Huber Müller Kessal-Wulf	2	0.0243516	0.02	33.13
Huber Müller Maidowski	100	1.2175819	1.22	34.35
Hömig Bryde Gaier	9	0.1095824	0.11	34.46
Hömig Haas Seidl	1	0.0121758	0.01	34.47
Hömig Hoffmann-Riem Bryde	2	0.0243516	0.02	34.49
Hömig Hoffmann-Riem Gaier	1	0.0121758	0.01	34.51
Hömig Jaeger Kühling	1	0.0121758	0.01	34.52

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hömig Kühling Seidl	1	0.0121758	0.01	34.53
Jaeger Frau Haas ist gehindert, ihre Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Lübbe-Wolff Unterschrift beizufügen. Jaeger	1	0.0121758	0.01	34.54
Jaeger Haas Bryde	1	0.0121758	0.01	34.55
Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde		0.0121758	0.01	34.57
Jaeger Hömig Bryde	235	2.8613174	2.86	37.43
Jaeger Hömig Hoffmann- Riem	4	0.0487033	0.05	37.48
Jaeger Hömig Steiner	1	0.0121758	0.01	37.49
Jaeger Steiner Hohmann- Dennhardt	1	0.0121758	0.01	37.50
Jentsch Broß Gerhardt	13	0.1582856	0.16	37.66
Jentsch Broß Lübbe-Wolff	59	0.7183733	0.72	38.38
Jentsch Broß Mellinghoff	1	0.0121758	0.01	38.39
Jentsch Broß Osterloh	1	0.0121758	0.01	38.40
Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe- Wolff Gerhardt	1	0.0121758	0.01	38.41
Jentsch Hassemer Broß	2	0.0243516	0.02	38.44
Kessal- Wulf König Langenfeld	1	0.0121758	0.01	38.45
Kessal-Wulf König Maidowski	4	0.0487033	0.05	38.50
Kessal- Wulf Langenfeld Wallrabenstein	1	0.0121758	0.01	38.51
Kessal- Wulf Maidowski Wallrabenstein	1	0.0121758	0.01	38.52
Kirchhof Baer Britz	1	0.0121758	0.01	38.54
Kirchhof Bryde Schluckebier	55	0.6696700	0.67	39.21

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Kirchhof Eichberger Baer	1	0.0121758	0.01	39.22
Kirchhof Eichberger Britz	137	1.6680872	1.67	40.89
Kirchhof Eichberger Masing	112	1.3636917	1.36	42.25
Kirchhof Eichberger Masing Paulus Baer Britz Ott Christ	2	0.0243516	0.02	42.27
Kirchhof Eichberger Masing Paulus Baer Britz Ott Christ	19	0.2313406	0.10	42.37
Kirchhof Eichberger Schluckebier Masing Paulus Baer Britz	3	0.0365275	0.04	42.41
Kirchhof Eichberger Schluckebier Masing Paulus Baer Britz Ott	7	0.0852307	0.09	42.49
Kirchhof Gaier Eichberger Schluckebier Masing Paulus Baer Britz	69	0.840115	0.84	43.33
Kirchhof Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Masing	2	0.0243516	0.02	43.36
Kirchhof Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Masing Paulus	19	0.2313406	0.23	43.59
Kirchhof Jentsch Graßhof	1	0.0121758	0.01	43.60
Kirchhof Jentsch Osterloh	3	0.0365275	0.04	43.64
Kirchhof Masing Baer	178	2.1672958	2.17	45.81
Kirchhof Masing Gaier	1	0.0121758	0.01	45.82
Kirchhof Masing Ott	1	0.0121758	0.01	45.83
Kirchhof Masing Paulus	78	0.9497139	0.95	46.78
Kirchhof Masing Paulus Baer Britz Ott Christ Paule	1	0.0121758	0.01	46.79
Kirchhof Masing Schluckebier	1	0.0121758	0.01	46.80
Kirchhof Ott Christ	50	0.6087909	0.61	47.41
Kirchhof Paulus Baer	4	0.0487033	0.05	47.46
Kirchhof Paulus Britz	1	0.0121758	0.01	47.47
Kirchhof Schluckebier Baer	102	1.2419335	1.24	48.72
Kirchhof Schluckebier Masing	2	0.0243516	0.02	48.74
Kirchhof Schluckebier Ott	59	0.7183733	0.72	49.46
Klein Winter Sommer	1	0.0121758	0.01	49.47
Kruis	1	0.0121758	0.01	49.48
Kruis Winter Hassemer	2	0.0243516	0.02	49.51
Kruis Winter Jentsch	1	0.0121758	0.01	49.52

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Kruis Winter Sommer	1	0.0121758	0.01	49.53
König Hermanns Müller	1	0.0121758	0.01	49.54
König Hermanns Müller Kessal- Wulf Maidowski Langenfeld Wallrabenstein	2	0.0243516	0.02	49.57
König Huber Hermanns Kessal- Wulf Langenfeld	1	0.0121758	0.01	49.58
König Huber Hermanns Kessal- Wulf Maidowski Langenfeld Wallrabenstein	2	0.0243516	0.02	49.60
König Huber Hermanns Müller Kessal- Wulf Maidowski Langenfeld	1	0.0487033	0.05	49.65
König Huber Hermanns Müller Kessal- Wulf Maidowski Langenfeld Wallrabenstein	2	0.9618897	0.96	50.61
König Huber Hermanns Müller Maidowski Langenfeld Härtel	1	0.0121758	0.01	50.63
König Huber Hermanns Müller Maidowski Langenfeld Wallrabenstein	2	0.0365275	0.04	50.66
König Huber Müller Kessal- Wulf Maidowski Langenfeld	2	0.0243516	0.02	50.69
König Maidowski Langenfeld	2	0.0243516	0.02	50.71
König Maidowski Wallrabenstein	1	0.0121758	0.01	50.72
König Müller Maidowski	67	0.8157799	0.82	51.54
König Müller Wallrabenstein	1	0.0121758	0.01	51.55
Kühling Jaeger Steiner	1	0.0121758	0.01	51.56
Kühling Hömig Hoffmann- Riem	1	0.0121758	0.01	51.58
Kühling Hömig Steiner	1	0.0121758	0.01	51.59
Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem	1	0.0121758	0.01	51.60
Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Di Fabio	1	0.0121758	0.01	51.61
Kühling Jaeger Hömig	102	1.2419335	1.24	52.86
Kühling Jaeger Steiner	90	1.0958237	1.10	53.95
Landau Huber Hermanns Müller Kessal- Wulf König Maidowski	1	0.0121758	0.01	53.96

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Landau Kessal-Wulf König	93	1.1323512	1.13	55.10
Landau Kessal-Wulf Maidowski	1	0.0121758	0.01	55.11
Landau König Maidowski	2	0.0243516	0.02	55.13
Landau Müller Kessal-Wulf	1	0.0121758	0.01	55.14
Landau Müller König	1	0.0121758	0.01	55.16
Landau Müller Maidowski	1	0.0121758	0.01	55.17
Landau Schluckebier Paulus Maidowski	2	0.0243516	0.02	55.19
Limbach Kirchhof Jentsch	1	0.0121758	0.01	55.21
Limbach Broß Di Fabio	1	0.0121758	0.01	55.22
Limbach Böckenförde Graßhof Kruis Kirchhof Winter Sommer	1	0.0121758	0.01	55.23
Limbach Böckenförde Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0121758	0.01	55.45
Limbach Böckenförde Klein Graßhof Kruis Winter Seibert Sommer	1	0.0121758	0.01	55.46
Limbach Böckenförde Klein Graßhof Kruis Winter Sommer	1	0.0121758	0.01	55.47
Limbach Graßhof Kirchhof	26	0.3165713	0.32	55.79
Limbach Graßhof Kruis Kirchhof Winter Sommer Hassemer	1	0.0121758	0.01	55.80
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch	1	0.0121758	0.01	55.83
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Der Richter Hassemer ist an der Unterschrift verhindert. Limbach	1	0.0121758	0.01	55.84
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Der Richter Hassemer ist an der Unterzeichnung verhindert. Limbach	1	0.0121758	0.01	55.85
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Hassemer	1	0.0121758	0.01	56.34
Limbach Graßhof Sommer	1	0.0121758	0.01	56.35
Limbach Hassemer Broß	32	0.3896262	0.39	56.74
Limbach Hassemer Di Fabio	31	0.3774504	0.38	57.12
Limbach Hassemer Jentsch	2	0.0243516	0.02	57.14
Limbach Hassemer Mellinghoff	137	1.6680872	1.67	58.81

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Limbach Hassemer Osterloh	1	0.0121758	0.01	58.82
Limbach Jentsch Di Fabio	52	0.6331426	0.63	59.45
Limbach Jentsch Hassemer	15	0.1826373	0.18	59.64
Limbach Jentsch Osterloh	1	0.0121758	0.01	59.65
Limbach Kirchhof Graßhof	3	0.0365275	0.04	59.69
Limbach Kirchhof Jentsch	24	0.2922197	0.29	59.98
Limbach Kirchhof Jentsch Hassemer Broß Osterloh	2	0.0243516	0.02	60.00
Limbach Kirchhof Sommer Jentsch Hassemer Broß Osterloh	1	0.0121758	0.11	60.11
Limbach Kirchhof Winter Sommer Jentsch Broß Osterloh	1	0.0121758	0.01	60.12
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß Osterloh	1	0.0121758	0.01	60.14
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß	1	0.0121758	0.05	60.19
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß Osterloh	1	0.0121758	0.13	60.32
Limbach Klein Graßhof Kruis Kirchhof Winter	1	0.0121758	0.02	60.34
Limbach Kruis Graßhof Hassemer Kirchhof Sommer	1	0.0121758	0.01	60.36
Limbach Kruis Kirchhof Winter Sommer Jentsch Hassemer	1	0.0121758	0.01	60.37
Limbach Kruis Winter	26	0.3165713	0.32	60.68
Limbach Sommer Broß	2	0.0243516	0.02	60.71
Limbach Sommer Hassemer Broß Osterloh Di Fabio	1	0.0121758	0.01	60.72
Limbach Sommer Jentsch	2	0.0243516	0.02	60.75
Limbach Sommer Jentsch; Hassemer Broß Osterloh Di Fabio Mellinghoff	1	0.0121758	0.01	60.76
Limbach Sommer Jentsch Hassemer Broß Osterloh	1	0.0121758	0.01	60.77
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio	4	0.0487034	0.52	61.29
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio Mellinghoff	2	0.0243516	0.35	61.65
Limbach Sommer Jentsch Hassemer Broß Osterloh Mellinghoff	1	0.0121758	0.05	61.69

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Limbach Sommer Jentsch Hassemer Herr Broß ist an der Osterloh Di Fabio Unterschrift gehindert. Limbach	1	0.0121758	0.01	61.71
Limbach Winter Hasseme	1	0.0121758	0.01	61.72
Limbach Winter Hassemer	62	0.7549008	0.75	62.47
Limbach Winter Kruis	2	0.0243516	0.02	62.50
Limbach Winter Sommer Jentsch Hassemer Broß Osterloh	1	0.0121758	0.01	62.51
Lübbe-Wolff Gerhardt Huber	1	0.0121758	0.01	62.52
Lübbe- Wolff Gerhardt Landau Huber Hermanns Müller Kessal- Wulf	13	0.1582856	0.16	62.68
Lübbe-Wolff Gerhardt Müller	1	0.0121758	0.01	62.69
Lübbe-Wolff Huber Kessal- Wulf	42	0.5113844	0.51	63.20
Lübbe-Wolff Landau Huber	14	0.1704615	0.17	63.38
Lübbe-Wolff Landau Kessal- Wulf	35	0.4261537	0.43	63.80
Mahrenholz Böckenförde Klein Graßhof Kruis Kruis Hof Winter Sommer	1	0.0121758	0.01	63.86
Maidowski Langenfeld Wallrabenstein	2	0.0243516	0.02	63.89
Masing Baer Britz	1	0.0121758	0.01	63.90
Masing Britz Radtke	1	0.0121758	0.01	63.91
Masing Paulus Baer	1	0.0121758	0.01	63.92
Masing Paulus Baer Britz Ott Christ Radtke	1	0.0121758	0.01	63.96
Masing Paulus Christ	107	1.3028126	1.30	65.26
Masing Paulus Eichberger	1	0.0121758	0.01	65.27
Masing Paulus Huber Maidowski	1	0.0121758	0.01	65.29
Masing Paulus Ott	1	0.0121758	0.01	65.30
Mellinghoff Gerhardt Landau	1	0.0121758	0.01	65.31
Mellinghoff Landau Huber	1	0.0121758	0.01	65.32
Mellinghoff Lübbe- Wolff Gerhardt	32	0.3896262	0.39	65.71

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Mellinghoff Lübbe- Wolff Huber	56	0.6818459	0.68	66.39
Mellinghoff Lübbe- Wolff Landau	1	0.0121758	0.01	66.41
Müller Kessal- Wulf Langenfeld	3	0.0365275	0.04	66.44
Müller Kessal- Wulf Maidowski	3	0.0365275	0.04	66.48
Müller König Langenfeld	4	0.0487033	0.05	66.53
Müller Maidowski Langenfeld	1	0.0121758	0.01	66.54
Müller Maidowski Wallrabenstein	2	0.0243516	0.02	66.57
Osterloh Gerhardt Landau	2	0.0243516	0.02	66.59
Osterloh Lübbe- Wolff Gerhardt	1	0.0121758	0.01	66.60
Osterloh Mellinghoff Gerhardt	64	0.7792524	0.78	67.38
Osterloh Mellinghoff Landau	1	0.0121758	0.01	67.39
Osterloh Mellinghoff Lübbe- Wolff	2	0.0243516	0.02	67.42
Osterloh Sommer Broß	3	0.0365275	0.04	67.45
Papier Haas Hoffmann-Riem	1	0.0121758	0.01	67.47
Papier Grimm Kühling Seibert Jaeger Haas Hömig Steiner Frau BVRin Seibert ist aus dem Amt ausgeschieden.	1	0.0121758	0.01	67.48
Papier Bryde Schluckebier	123	1.4976257	1.50	68.98
Papier Der Richter Grimm ist aus dem Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt Amt ausgeschie- den und deshalb an der Unterschrift gehindert. Papier	2	0.0243516	0.02	69.00

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Der Richter Kühling ist aus dem Amt geschieden und daher gehindert zu unterschreiben Papier Jaeger Haas Hmig Steiner Hohmann-Dennhardt Hoffmann-Riem	1	0.0121758	0.01	69.01
Papier Der Richter Kühling ist aus Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem dem Amt geschieden und daher gehindert zu unterschreiben. Papier	1	0.0121758	0.01	69.02
Papier Der Richter Steine ist Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier ausgeschieden und daher der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.04
Papier Die Richterin Haas ist aus dem Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	3	0.0365275	0.04	69.07
Papier Die Richterin Haas Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger ist aus dem Amt ausge- schieden und deshalb an der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.09
Papier Die Richterin Jaeger ist aus Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem dem Amt ausgeschieden und daher an der Unterschrift gehindert Papier	1	0.0121758	0.01	69.10

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Die Richterin Jaeger ist aus Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem dem Amt ausgeschieden und daher an der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.11
Papier Die Richterin Jaeger ist aus Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.12
Papier Die Richterin Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Jaeger ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.13
Papier Die Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Richterin Jaeger ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	1	0.0121758	0.01	69.15
Papier Eichberger Masing	69	0.8401315	0.84	69.99
Papier Gaier Kirchhof	5	0.0608791	0.06	70.05
Papier Graßhof Grimm Kühling Fraul BVRin Seibert ist aus dem Amt ausgeschieden. Pa-pier Jaeger Haas Hömig	1	0.0121758	0.01	70.06
Papier Graßhof Grimm Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.01	70.07
Papier Grimm Haas	1	0.0121758	0.01	70.08
Papier Grimm Hömig	171	2.0820650	2.08	72.17
Papier Grimm Khling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.01	72.18
Papier Grimm Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.15	72.32

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Grimm Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	15	0.158267	0.16	72.48
Papier Grimm Kühling Jaeger Hömig Steiner Hohmann-Dennhardt	17	0.117647	0.02	72.51
Papier Grimm Kühling Seibert Jaeger Haas Hömig Steiner	7	0.128571	0.09	72.59
Papier Grimm Steiner	2	0.0243516	0.02	72.62
Papier Haas Der Richter Hömig ist aus dem Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Amt ausgeschieden und deshalb an der Unterschrift gehindert Papier	1	0.0121758	0.01	72.63
Papier Haas Hoffmann-Riem	55	0.6696700	0.67	73.30
Papier Haas Hohmann-Dennhardt	75	0.9131864	0.91	74.21
Papier Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier	30	0.3652746	0.37	74.58
Papier Haas Steiner	26	0.3165713	0.32	74.89
Papier Haas Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger	4	0.0487033	0.05	74.94
Papier Hassemer Sommer Jaeger Haas Hömig Steiner Jentsch Broß Osterloh Hohmann-Dennhardt Hoffmann-Riem Di Fabio Bryde Mellinghoff Lübbe-Wolff	17	0.117647	0.01	74.95
Papier Hoffmann-Riem Eichberger	11	0.1339340	0.13	75.09
Papier Hoffmann-Riem Gaier	12	0.1461098	0.15	75.23
Papier Hoffmann-Riem Kirchhof	1	0.0121758	0.01	75.25
Papier Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Kirchhof Masing	27	0.3287471	0.33	75.58

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Hohmann-Dennhardt Der Richter Hoffmann-Riem ist Bryde Gaier Eichberger Schluckebier Kirchhof aus dem Amt ausgeschieden und daher an der Unterschrift gehindert. Papier	1	0.0121758	0.01	75.59
Papier Hohmann-Dennhardt Gaier	1	0.0121758	0.01	75.60
Papier Hohmann-Dennhardt Hoffmann-Riem	142	1.7289663	1.73	77.33
Papier Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier	1	0.0121758	0.01	77.34
Papier Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier Kirchhof	8	0.0974066	0.10	77.44
Papier Hömig Bryde	1	0.0121758	0.01	77.45
Papier Hömig Hoffmann-Riem	2	0.0243516	0.02	77.47
Papier Hömig Hohmann-Dennhardt	2	0.0243516	0.02	77.50
Papier Jaeger Haas Hömig Steiner Hohmann-Dennhardt Bryde	2	0.0243516	0.02	77.52
Papier Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem	2	0.0243516	0.02	77.55
Papier Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde	168	0.8279557	0.83	78.38
Papier Jaeger Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem	1	0.0121758	0.01	78.39
Papier Jaeger Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde	1	0.0121758	0.01	78.40
Papier Kühling Jaeger Haas Hömig Hohmann-Dennhardt Hoffmann-Riem Osterloh	1	0.0121758	0.01	78.41

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Kühling Jaeger Haas Hömig Steiner	1	0.0121758	0.01	78.42
Papier Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.01	78.44
Papier Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.01	78.45
Papier Kühling Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem	1	0.0121758	0.22	78.67
Papier Kühling Jaeger Hömig Steiner Hohmann-Dennhardt	1	0.0121758	0.01	78.68
Papier Kühling Jaeger Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem	1	0.0121758	0.01	78.69
Papier Kühling Steiner	2	0.0243516	0.02	78.72
Papier Richter Kühling ist aus dem Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Amt geschieden und daher gehindert zu unterschreiben. Papier	1	0.0121758	0.01	78.73
Papier Richter Kühling ist aus dem Jaeger Haas Hömig Steiner Hohmann-Dennhardt Hoffmann-Riem Amt geschieden und daher gehindert, die Unterschrift zu leisten. Papier	1	0.0121758	0.01	78.74
Papier Schluckebier Kirchhof	1	0.0121758	0.01	78.75
Papier Steiner Gaier	126	1.5341532	1.53	80.29
Papier Steiner Hoffmann-Riem	219	2.6665043	2.67	82.95
Papier Steiner Hohmann-Dennhardt	112	1.3636917	1.36	84.32
Papier Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier	16	0.1948131	0.19	84.51
Paulus Britz Ott Christ Radtker Härtel	1	0.0121758	0.01	84.52
Paulus Christ Härtel	97	1.1810544	1.18	85.71
Paulus Christ Radtker	1	0.0121758	0.01	85.72

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Paulus Huber Baer Müller	1	0.0121758	0.01	85.73
Paulus Ott Christ	4	0.0487033	0.05	85.78
Präsidentin Limbach ist aus Sommer Jentsch Richter Broß ist an der Hassemer Osterloh Di Fabio Mellinghoff dem Amt ausgeschieden. Unterschrift gehindert. Hassemer	1	0.0121758	0.01	85.79
Schluckebier Baer Britz	1	0.0121758	0.01	85.80
Schluckebier Masing Baer	1	0.0121758	0.01	85.82
Schluckebier Masing Paulus	1	0.0121758	0.01	85.83
Schluckebier Paulus Baer	1	0.0121758	0.01	85.84
Schluckebier Paulus Ott	6	0.0730549	0.07	85.91
Seidl Graßhof Grimm Kühling Jaeger Haas Hömig	1	0.0121758	0.01	85.92
Seidl Graßhof Grimm Kühling Seibert Jaeger Haas Hömig	1	0.0121758	0.01	85.94
Seidl Grimm Haas	15	0.1826373	0.18	86.12
Seidl Grimm Hömig	6	0.0730549	0.07	86.19
Seidl Grimm Kühling Die Richterin Seibert ist an der Unterschrift gehindert. Seidl Jaeger Haas Hömig Steiner	2	0.0243516	0.02	86.22
Seidl Grimm Kühling Die Richterin Seibert ist we- gen Erkrankung an der Unterschrift gehindert. Seidl Jaeger Haas Hömig Steiner	1	0.0121758	0.01	86.23
Seidl Grimm Kühling Jaeger Haas Hömig Steiner	2	0.0243516	0.02	86.38
Seidl Grimm Kühling Seibert Jaeger Haas Hömig Steiner	2	0.0243516	0.02	86.64
Seidl Grimm Richter Söllner ist wegen einer Auslandsreise an der Unterschrift verhindert. Seidl Kühling Seibert Jaeger Haas	1	0.0121758	0.01	86.66
Seidl Grimm Seibert	1	0.0121758	0.01	86.67
Seidl Grimm Söllner Dieterich Kühling Seibert	1	0.0121758	0.01	86.68

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0121758	0.01	86.69
Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0121758	0.01	86.70
Seidl Haas Hömig	6	0.0730549	0.07	86.78
Seidl Hömig Steiner	2	0.0243516	0.02	86.80
Sommer Broß Di Fabio	2	0.0243516	0.02	86.83
Sommer Broß Mellinghoff	68	0.8279557	0.83	87.65
Sommer Broß Osterloh	80	0.9740655	0.97	88.63
Sommer Di Fabio Lübbe-Wolff	49	0.5966151	0.60	89.22
Sommer Hassemer Broß	2	0.0243516	0.02	89.25
Sommer Hassemer Osterloh	1	0.0121758	0.01	89.26
Sommer Jentsch Broß	3	0.0365275	0.04	89.30
Sommer Jentsch Hassemer	23	0.2800438	0.28	89.58
Sommer Jentsch Osterloh	1	0.0121758	0.01	89.59
Sommer Osterloh Di Fabio	36	0.4383295	0.44	90.03
Sommer Osterloh Mellinghoff	1	0.0121758	0.01	90.04
Steiner Gaier Eichberger	1	0.0121758	0.01	90.05
Steiner Gaier Schluckebier	2	0.0243516	0.02	90.08
Steiner Hoffmann-Riem Gaier	1	0.0121758	0.01	90.09
Steiner Hohmann-Dennhardt Bryde	3	0.0365275	0.04	90.13
Steiner Hohmann-Dennhardt Hoffmann-Riem	8	0.0974066	0.10	90.22
Steiner Jaeger Kühling	15	0.1826373	0.18	90.41
Vizepräsident Henschel ist wegen Urlaubs an der Unterschrift gehindert. Seidl Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0121758	0.01	90.42
Vizepräsident Henschel ist wegen Urlaubs an der Unterschrift verhindert. Seidl Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0121758	0.01	90.43

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Vizepräsident Henschel und Richter Söllner sind aus dem Amt aus- ge- schieden und deshalb an der Unterschrift gehindert. Seidl Seidl Grimm Kühling Seibert Jaeger Haas	1	0.0121758	0.01	90.44
Vizepräsident Mahrenholz ist ausgeschieden. Böckenför- de Böckenförde Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0121758	0.01	90.45
Vizepräsident Seidl Grimm Hömig ist we- gen Ausscheidens aus dem Amt an der Unterschrift verhindert. Grimm	1	0.0121758	0.01	90.47
Vizepräsident Seidl Grimm Kühling Seibert Jaeger Haas Hömig Steiner ist aus dem Amt ausgeschie- den und deshalb an der Unterschrift gehindert. Grimm	2	0.0243516	0.02	90.49
Vosskuhle Mellinghoff Lübbe- Wolff	1	0.0121758	0.01	90.50
Voßkuhle Broß Osterloh Di Fabio Mellinghoff Lübbe- Wolff Gerhardt Landau	50	0.6087909	0.61	91.11
Voßkuhle Di Fabio Der Richter Mellinghoff Lübbe- Wolff Gerhardt Landau Huber Hermanns ist aus dem Amt aus- ge- schieden und deshalb an der Unterschrift gehindert. Voßkuhle	1	0.0121758	0.01	91.12
Voßkuhle Di Fabio Lübbe- Wolff Gerhardt Landau Huber Hermanns	6	0.0730549	0.07	91.20
Voßkuhle Di Fabio Mellinghoff	1	0.0121758	0.01	91.21
Voßkuhle Di Fabio Mellinghoff Lübbe- Wolff Gerhardt Landau	1	0.0121758	0.01	91.22

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Di Fabio Mellinghoff Lübbe- Wolff Gerhardt Landau Huber Hermanns	23	0.2800438	0.28	91.50
Voßkuhle Di Fabio Mellinghoff Lübbe- Wolff Landau Huber Hermanns	1	0.0121758	0.01	91.51
Voßkuhle Die Richterin Lübbe- Wolff ist aus dem Amt Der Richter Ger- hardt ist aus dem Amt Landau Huber Herrmanns Müller Kessal- Wulf ausgeschieden und deshalb an der Unter- schrift gehindert Voßkuhle	1	0.0121758	0.01	91.53
Voßkuhle Gerhardt Hermanns	1	0.0121758	0.01	91.54
Voßkuhle Gerhardt Huber	25	0.3043955	0.30	91.84
Voßkuhle Gerhardt Landau	37	0.4505053	0.45	92.29
Voßkuhle Gerhardt Landau Huber Hermanns Müller Kessal- Wulf König	0.0121758	0.01	92.30	
Voßkuhle Hermanns Müller	1	0.0121758	0.01	92.32
Voßkuhle Huber Hermanns Kessal- Wulf König Maidowski	1	0.0121758	0.01	92.33
Voßkuhle Huber Hermanns Kessal- Wulf König Maidowski Langenfeld	11	0.1339340	0.13	92.46
Voßkuhle Huber Hermanns Müller Kessal- Wulf König	1	0.0121758	0.01	92.48
Voßkuhle Huber Hermanns Müller Kessal- Wulf König Maidowski	1	0.1339340	0.13	92.61
Voßkuhle Huber Hermanns Müller Kessal- Wulf König Maidowski Langenfeld	130	1.5828564	1.58	94.19
Voßkuhle Huber Müller	1	0.0121758	0.01	94.20
Voßkuhle Kessal- Wulf Langenfeld	1	0.0121758	0.01	94.22
Voßkuhle Kessal- Wulf Maidowski	157	1.9116036	1.91	96.13

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Kirchhof Lübbe- Wolff Gerhardt Gaier Eichberger Schluckebier Masing Paulus Huber Hermanns Baer Britz Müller Kessal- Wulf	1	0.0121758	0.01	96.14
Voßkuhle König Langenfeld	1	0.0121758	0.01	96.15
Voßkuhle König Maidowski	2	0.0243516	0.02	96.18
Voßkuhle Landau Hermanns	57	0.6940217	0.69	96.87
Voßkuhle Landau Hermanns Müller Kessal- Wulf König Maidowski	1	0.0121758	0.01	96.88
Voßkuhle Landau Huber Hermanns Kessal- Wulf König Maidowski	5	0.0608791	0.06	96.94
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf	1	0.0121758	0.02	96.97
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf König	1	0.0121758	0.05	97.02
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf König Maidowski	67	0.8157799	0.82	97.83
Voßkuhle Landau Huber Müller Kessal- Wulf Maidowski	21	0.0243516	0.02	97.86
Voßkuhle Landau Kessal- Wulf	1	0.0121758	0.01	97.87
Voßkuhle Lübbe- Wolff Gerhardt	1	0.0121758	0.01	97.88
Voßkuhle Lübbe- Wolff Gerhardt Huber Hermanns Kessal- Wulf	1	0.0121758	0.01	97.89
Voßkuhle Lübbe- Wolff Gerhardt Huber Hermanns Müller Kessal- Wulf	1	0.0121758	0.01	97.91
Voßkuhle Lübbe- Wolff Gerhardt Landau Hermanns Müller Kessal- Wulf	1	0.0121758	0.01	97.92
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns	4	0.0487033	0.05	97.97
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Kessal- Wulf	1	0.0121758	0.01	97.98

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Müller Kessal- Wulf	47	0.5722635	0.57	98.55
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Müller Kessal- Wulf	5	0.0608791	0.06	98.61
Voßkuhle Masing Huber Hermanns Kessal- Wulf König Maidowski Langenfeld	2	0.0243516	0.02	98.64
Voßkuhle Mellinghoff Gerhardt	3	0.0365275	0.04	98.67
Voßkuhle Mellinghoff Lübbe- Wolff	70	0.8523073	0.85	99.53
Voßkuhle Osterloh Gerhardt	2	0.0243516	0.02	99.55
Voßkuhle Osterloh Mellinghoff	29	0.3530987	0.35	99.90
Winter Jentsch Broß	1	0.0121758	0.01	99.91
Winter Jentsch Osterloh	1	0.0121758	0.01	99.93
Winter Sommer Hassemer	2	0.0243516	0.02	99.95
Winter Sommer Jentsch	1	0.0121758	0.01	99.96
Zeidler Rinck Wand Dr. Rottmann Dr. Dr. h. c. Niebler Steinberger Träger Mahrenholz H.- J. Rinck	1	0.0121758	0.01	99.98
gerichts Bryde Eichberger Schluckebier	1	0.0121758	0.01	99.99
unter Mitwirkung Voßkuhle Landau Huber Müller Kessal- Wulf Maidowski	1	0.0121758	0.01	100.00
Total	8213	100.0000000	99.00	100.00

Frequency Table for Variable: entscheidung_typ

2 unique value(s) detected.

entscheidung_typ	N	exactpercent	roundedpercent	cumulpercent
B	8013	97.564836	97.56	97.56
U	200	2.435164	2.44	100.00

(continued)

entscheidung_typ	N	exactpercent	roundedpercent	cumulpercent
Total	8213	100.000000	100.00	100.00

11 Frequenztabellen visualisieren

11.1 Präfix erstellen

```
prefix <- file.path(dir.analysis,  
                    paste0(config$project$shortname,  
                            "_01_Frequenztafel_var-"))
```

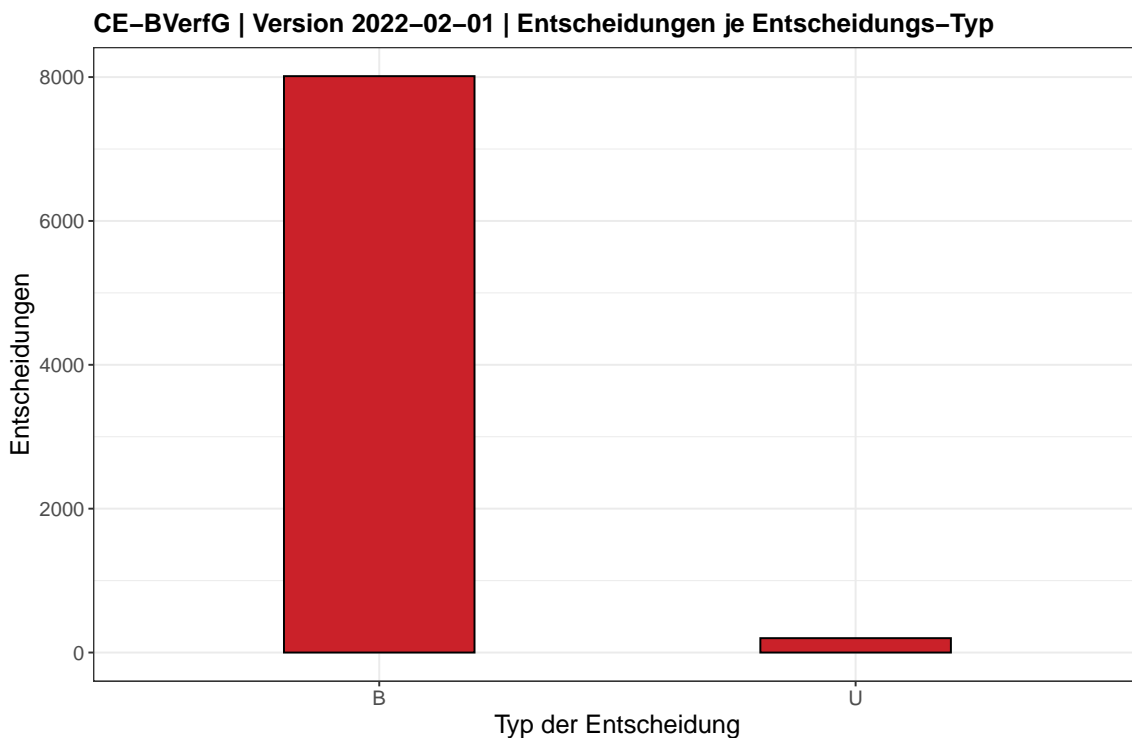
11.2 Tabellen einlesen

```
table.entsch.typ <- fread(paste0(prefix,  
                                 "entscheidung_typ.csv"))  
  
table.spruch.typ <- fread(paste0(prefix,  
                                 "spruchkoerper_typ.csv"))  
  
table.spruch.az <- fread(paste0(prefix,  
                                 "spruchkoerper_az.csv"))  
  
table.regz <- fread(paste0(prefix,  
                            "registerzeichen.csv"))  
  
table.jahr.eingangISO <- fread(paste0(prefix,  
                                     "eingangsjahr_iso.csv"))  
  
table.jahr.entscheid <- fread(paste0(prefix,  
                                    "entscheidungsjahr.csv"))  
  
table.output.praesi <- fread(paste0(prefix,  
                                   "praesi.csv"))  
  
table.output.vpraesi <- fread(paste0(prefix,  
                                   "v_praesi.csv"))
```

11.3 Diagramm: Typ der Entscheidung

```
freqtable <- table.entsch.typ[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(entscheidung_typ,  
                          -N),  
              y = N),  
          stat = "identity",  
          fill = "#ca2129",  
          color = "black",  
          width = 0.4) +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Entscheidungs-Typ"),  
    caption = caption,  
    x = "Typ der Entscheidung",  
    y = "Entscheidungen"  
  ) +  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

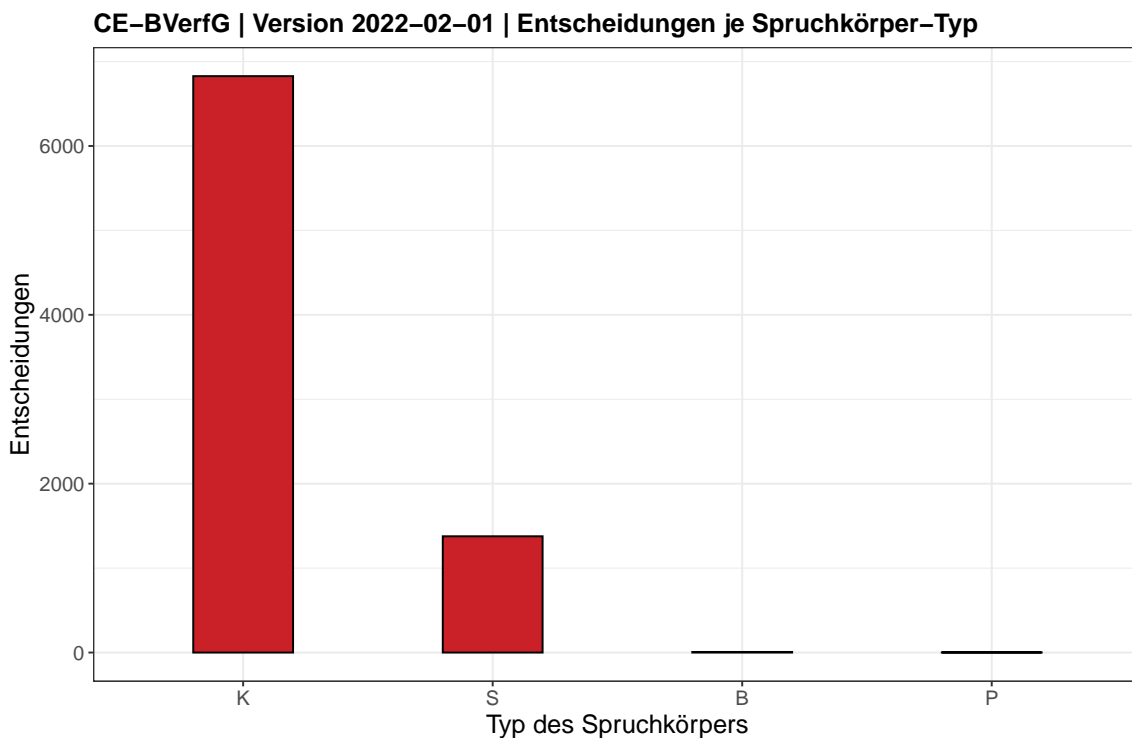


Fobbe | DOI: 10.5281/zenodo.5910152

11.4 Diagramm: Typ des Spruchkörpers

```
freqtable <- table.spruch.typ[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(spruchkoerper_typ,  
                          -N),  
              y = N),  
          stat = "identity",  
          fill = "#ca2129",  
          color = "black",  
          width = 0.4) +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Spruchkörper-Typ"),  
    caption = caption,  
    x = "Typ des Spruchkörpers",  
    y = "Entscheidungen"  
  ) +  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

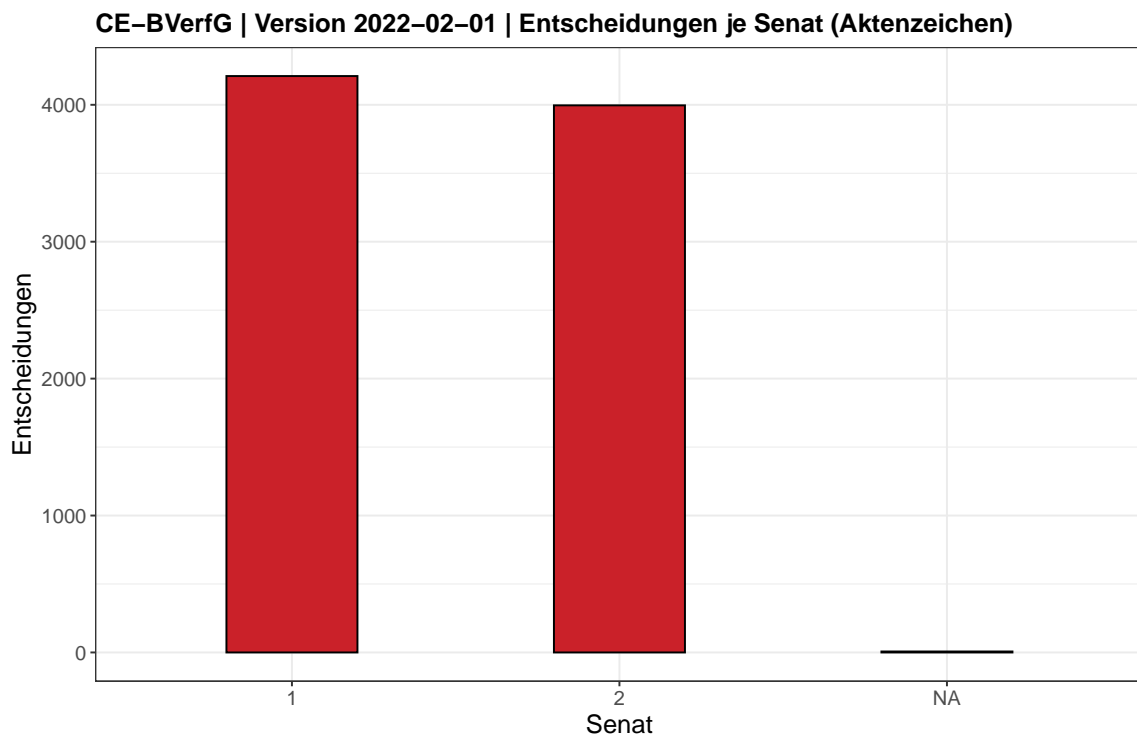


Fobbe | DOI: 10.5281/zenodo.5910152

11.5 Diagramm: Spruchkörper nach Aktenzeichen

```
freqtable <- table.spruch.az[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = spruchkoerper_az,  
              y = N),  
          stat = "identity",  
          fill = "#ca2129",  
          color = "black",  
          width = 0.4) +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Senat (Aktenzeichen)",  
    caption = caption,  
    x = "Senat",  
    y = "Entscheidungen"  
  )+  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```



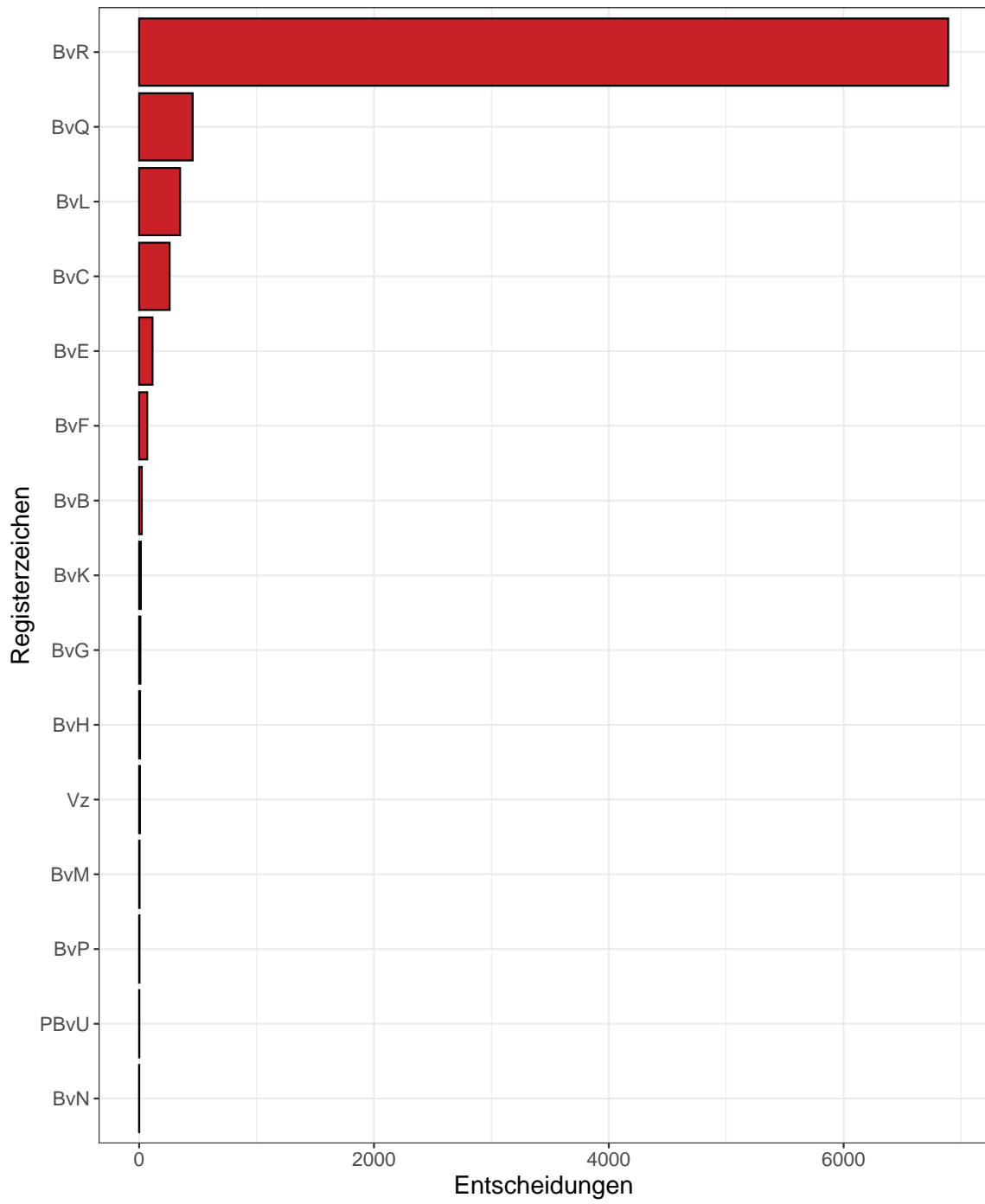
Fobbe | DOI: 10.5281/zenodo.5910152

11.6 Diagramm: Registerzeichen

```
freqtable <- table.regz[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(registerzeichen,  
                        N),  
              y = N),  
           stat = "identity",  
           fill = "#ca2129",  
           color = "black") +  
  coord_flip()+  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Registerzeichen"),  
    caption = caption,  
    x = "Registerzeichen",  
    y = "Entscheidungen"  
  )+  
  theme(  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

CE-BVerfG | Version 2022-02-01 | Entscheidungen je Registerzeichen

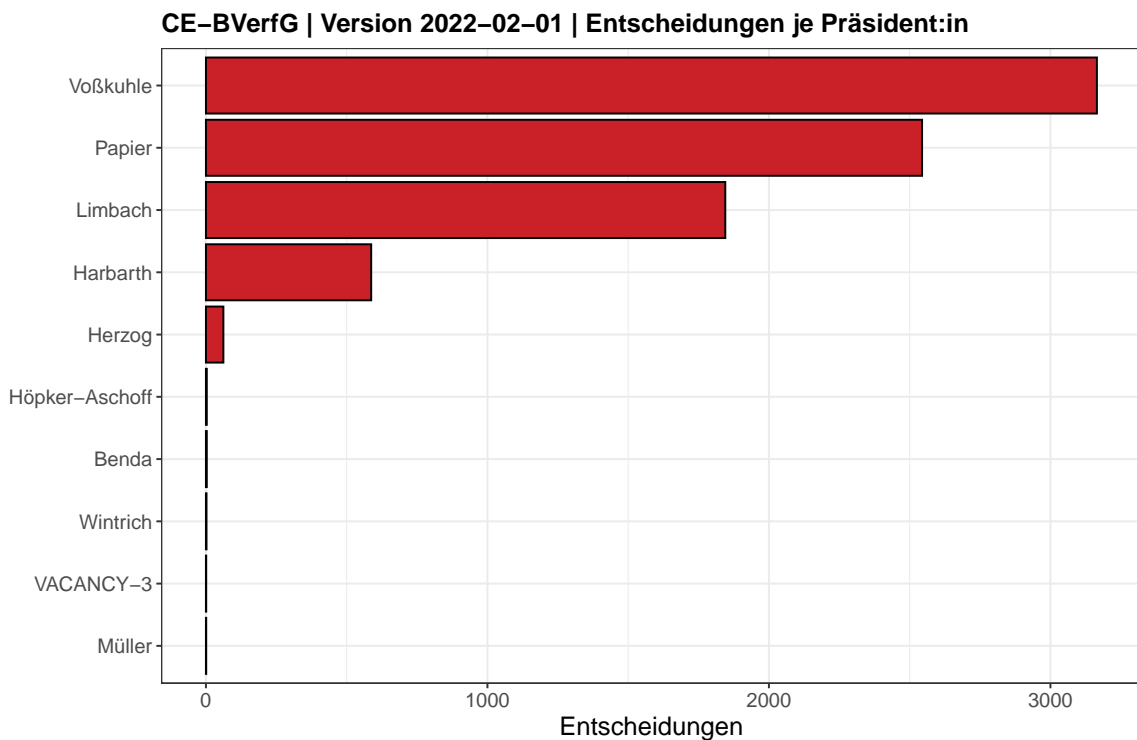


Fobbe | DOI: 10.5281/zenodo.5910152

11.7 Diagramm: Präsident:in

```
freqtable <- table.output.praesi[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(praesi,  
                          N),  
              y = N),  
          stat = "identity",  
          fill = "#ca2129",  
          color = "black") +  
  coord_flip()+  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Präsident:in"),  
    caption = caption,  
    x = "Präsident:in",  
    y = "Entscheidungen"  
  )+  
  theme(  
    axis.title.y = element_blank(),  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                              face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

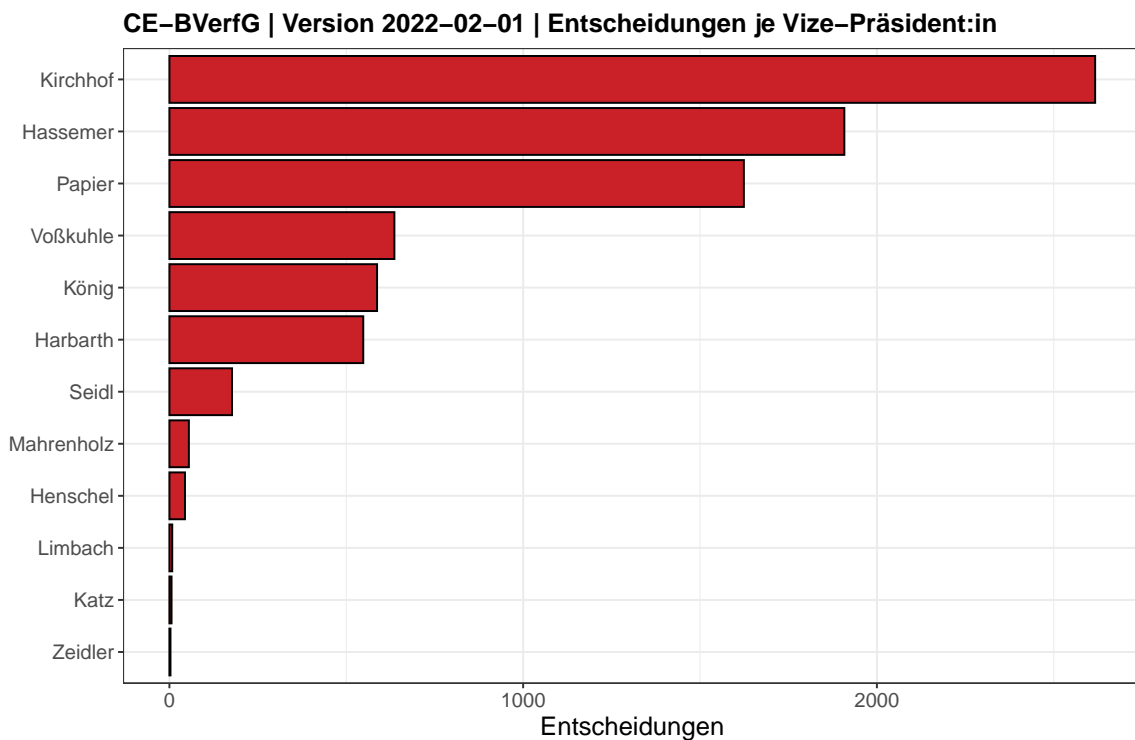


Fobbe | DOI: 10.5281/zenodo.5910152

11.8 Diagramm: Vize-Präsident:in

```
freqtable <- table.output.vpraesi[-.N]
```

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = reorder(v_praesi,  
                          N),  
              y = N),  
          stat = "identity",  
          fill = "#ca2129",  
          color = "black") +  
  coord_flip()+  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Vize-Präsident:in"),  
    caption = caption,  
    x = "Vize-Präsident:in",  
    y = "Entscheidungen"  
  )+  
  theme(  
    axis.title.y = element_blank(),  
    text = element_text(size = 14),  
    plot.title = element_text(size = 14,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

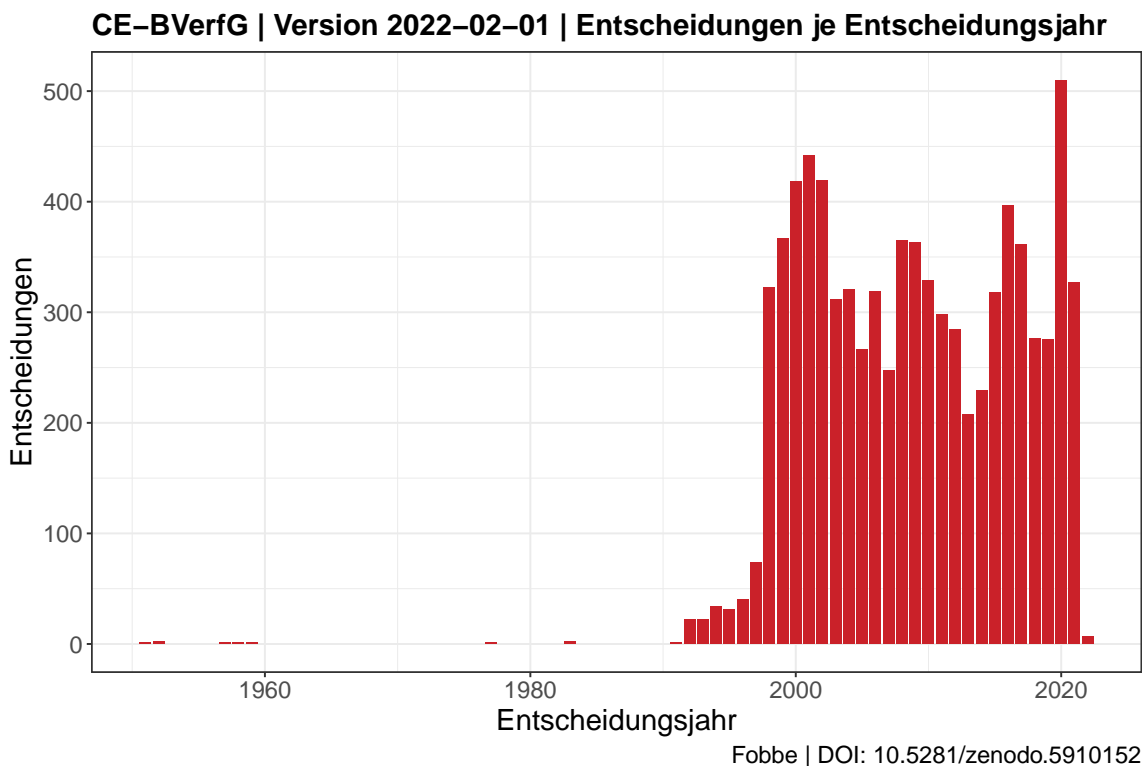


Fobbe | DOI: 10.5281/zenodo.5910152

11.9 Diagramm: Entscheidungsjahr

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

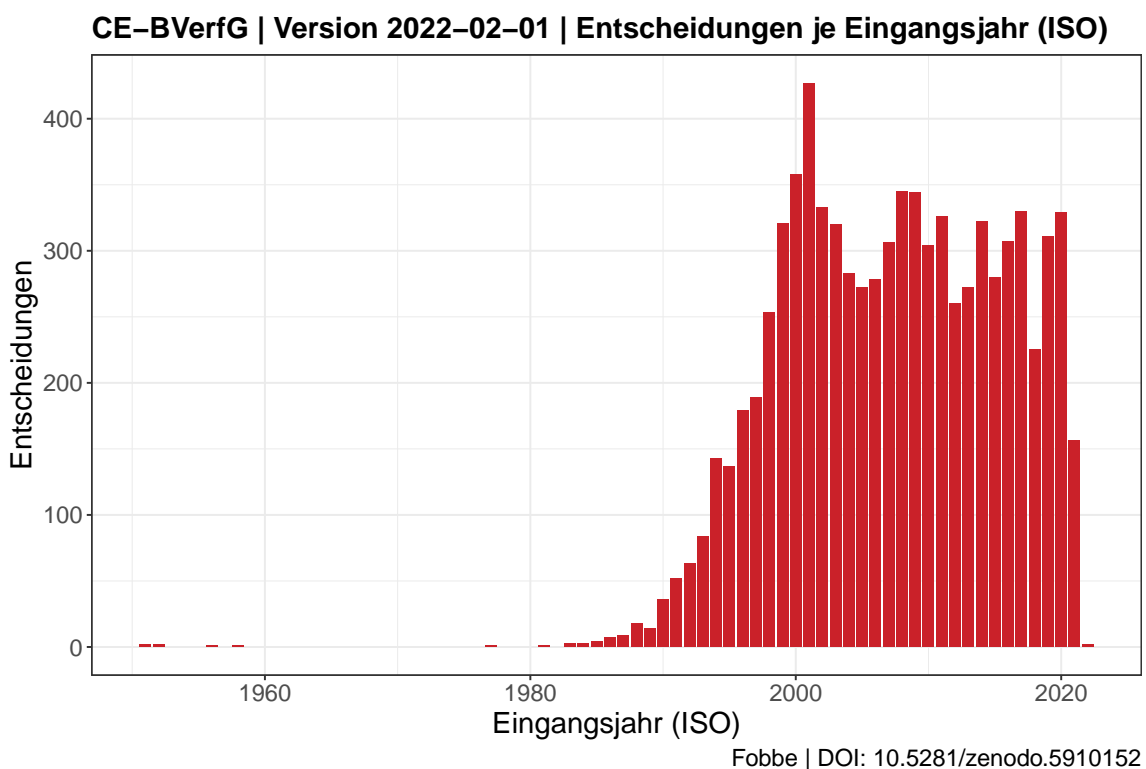
```
ggplot(data = freqtable) +  
  geom_bar(aes(x = entscheidungsjahr,  
              y = N),  
          stat = "identity",  
          fill = "#ca2129") +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Entscheidungsjahr"),  
    caption = caption,  
    x = "Entscheidungsjahr",  
    y = "Entscheidungen"  
  ) +  
  theme(  
    text = element_text(size = 16),  
    plot.title = element_text(size = 16,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```



11.10 Diagramm: Eingangsjahr (ISO)

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

```
ggplot(data = freqtable) +  
  geom_bar(aes(x = eingangsjahr_iso,  
              y = N),  
          stat = "identity",  
          fill = "#ca2129") +  
  theme_bw() +  
  labs(  
    title = paste(prefix.figuretitle,  
                  "| Entscheidungen je Eingangsjahr (ISO)"),  
    caption = caption,  
    x = "Eingangsjahr (ISO)",  
    y = "Entscheidungen"  
  ) +  
  theme(  
    text = element_text(size = 16),  
    plot.title = element_text(size = 16,  
                               face = "bold"),  
    legend.position = "none",  
    plot.margin = margin(10, 20, 10, 10)  
  )
```



12 Korpus-Analytik

12.1 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.

12.1.1 Funktion anzeigen

```
print(f.future_lingsummarize)
```

```
## function(dt,
##           chunkswerker = 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,
##                               f.lingsummarize,
##                               future.seed = TRUE,
##                               future.scheduling = chunkswerker,
##                               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)
## }

```

12.1.2 Berechnung durchführen

```

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

```

```

summary.corpus <- f.future_lingsummarize(txt.bverfg)

#deprecated; Parallelisierung jetzt mit futures
#summary.corpus <- f.lingsummarize.iterator(txt.bverfg,
#                                           threads = fullCores,
#                                           chunksize = 1)

```

12.2 Variablen-Namen anpassen

```
setnames(summary.corpus,  
  old = c("nchars",  
          "ntokens",  
          "ntypes",  
          "nsentences"),  
  new = c("zeichen",  
          "tokens",  
          "typen",  
          "saetze"))
```

12.3 Kennwerte dem Korpus hinzufügen

```
txt.bverfg <- cbind(txt.bverfg,  
  summary.corpus)
```

12.4 Variante mit Metadaten erstellen

```
meta.bverfg <- txt.bverfg[, !"text"]
```

12.5 Linguistische Kennwerte

12.5.1 Zusammenfassungen berechnen

```
dt.summary.ling <- meta.bverfg[, lapply(.SD,  
  function(x)unclass(summary(x))),  
  .SDcols = c("zeichen",  
             "tokens",  
             "typen",  
             "saetze")]  
  
dt.sums.ling <- meta.bverfg[,  
  lapply(.SD, sum),  
  .SDcols = c("zeichen",  
             "tokens",  
             "typen",  
             "saetze")]  
  
tokens.temp <- tokens(corpus(txt.bverfg),  
  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.temp))

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"))

```

12.5.2 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	160,301,985	1,255	4,951	11,198	19,518.0793	21,961	781,226
tokens	24,755,119	161	751	1,720	3,014.1384	3,395	115,540
typen	281,164	91	307	591	777.1955	983	13,491
saetze	1,411,651	6	45	105	171.8801	199	5,282

12.5.3 Zusammenfassungen speichern

```
fwrite(dt.stats.ling,  
      file.path(dir.analysis,  
                paste0(config$project$shortname,  
                        "_00_KorpusStatistik_ZusammenfassungLinguistisch.csv")),  
      na = "NA")
```

12.6 Quantitative Variablen

12.6.1 Entscheidungsdatum

```
summary(as.IDate(meta.bverfg$datum))
```

```
##           Min.         1st Qu.         Median         Mean         3rd Qu.         Max.
## "1951-09-09" "2002-09-11" "2009-03-17" "2009-05-10" "2016-04-19" "2022-01-31"
```

12.6.2 Zusammenfassungen berechnen

```
dt.summary.docvars <- meta.bverfg[,
                                lapply(.SD, function(x)unclass(summary(na.omit(
x))))),
                                .SDcols = c("entscheidungsjahr",
                                             "eingangsjahr_iso",
                                             "band",
                                             "eingangsnummer")]

dt.unique.docvars <- meta.bverfg[,
                                  lapply(.SD, function(x)length(unique(na.omit(x))
                                  )),
                                  .SDcols = c("entscheidungsjahr",
                                                "eingangsjahr_iso",
                                                "band",
                                                "eingangsnummer")]

dt.stats.docvars <- rbind(dt.unique.docvars,
                          dt.summary.docvars)

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

setnames(dt.stats.docvars, c("Variable",
                             "Anzahl",
                             "Min",
                             "Quart1",
                             "Median",
                             "Mean",
                             "Quart3",
                             "Max"))
```

12.6.3 Zusammenfassungen anzeigen

```
kable(dt.stats.docvars,  
      format = "latex",  
      booktabs = TRUE,  
      longtable = TRUE)
```

Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	39	1951	2002	2009	2008.8640	2016	2022
eingangsjahr_iso	46	1951	2001	2008	2007.2986	2014	2022
band	67	1	106	120	121.0999	136	155
eingangsnummer	2716	1	289	1054	1146.0886	1866	3634

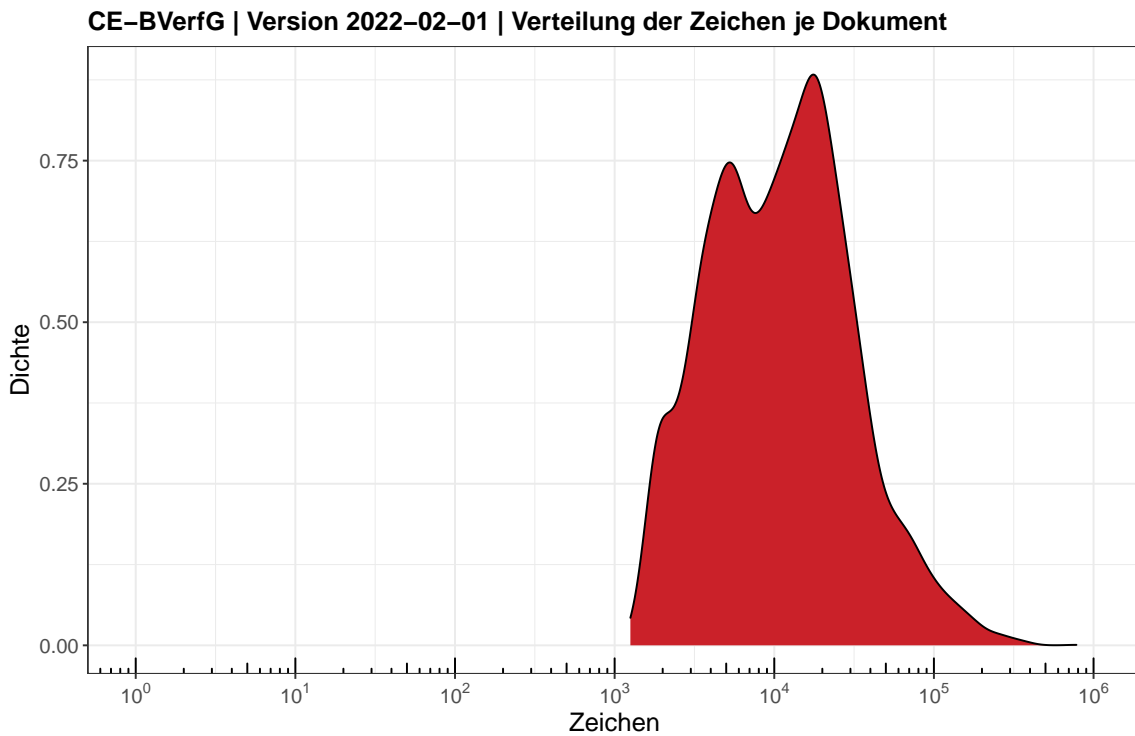
12.6.4 Zusammenfassungen speichern

```
fwrite(dt.stats.docvars,  
       file.path(dir.analysis,  
                 paste0(config$project$shortname,  
                         "_00_KorpusStatistik_ZusammenfassungDocvarsQuantitativ.  
csv")),  
       na = "NA")
```

12.7 Verteilungen linguistischer Kennwerte

12.7.1 Diagramm: Verteilung Zeichen

```
ggplot(data = meta.bverfg)+
  geom_density(aes(x = zeichen),
               fill = "#ca2129")+
  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 Dokument"),
    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.5910152

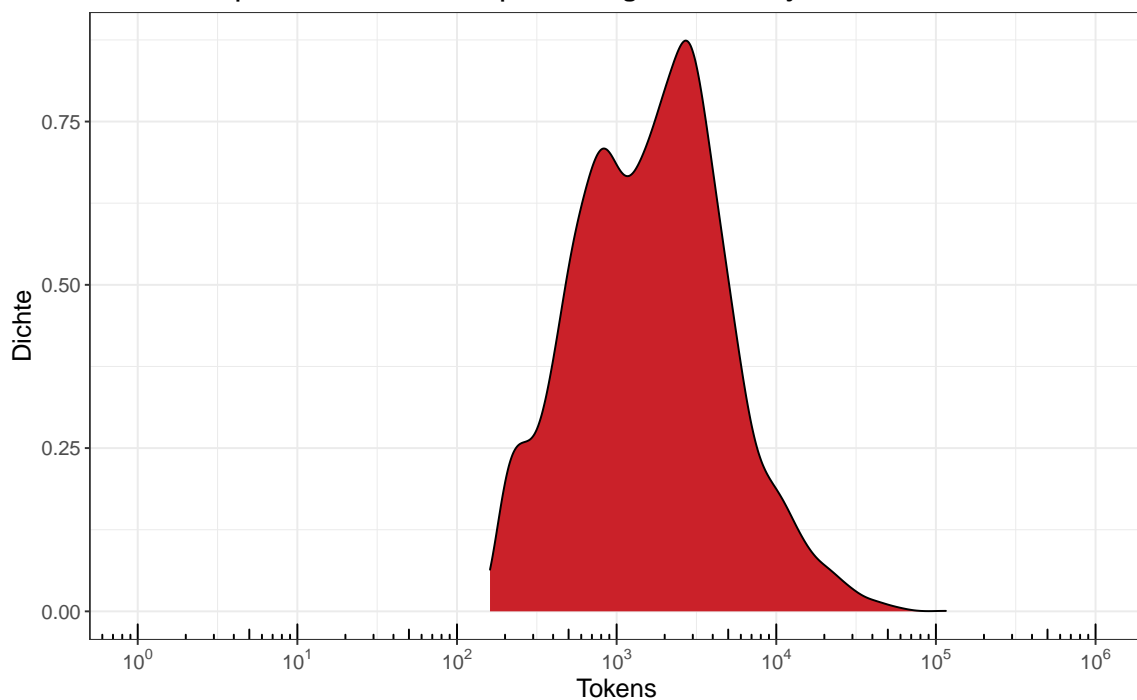
12.7.2 Diagramm: Verteilung Tokens


```

ggplot(data = meta.bverfg)+
  geom_density(aes(x = tokens),
               fill = "#ca2129")+
  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 Dokument"),
    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)
  )

```

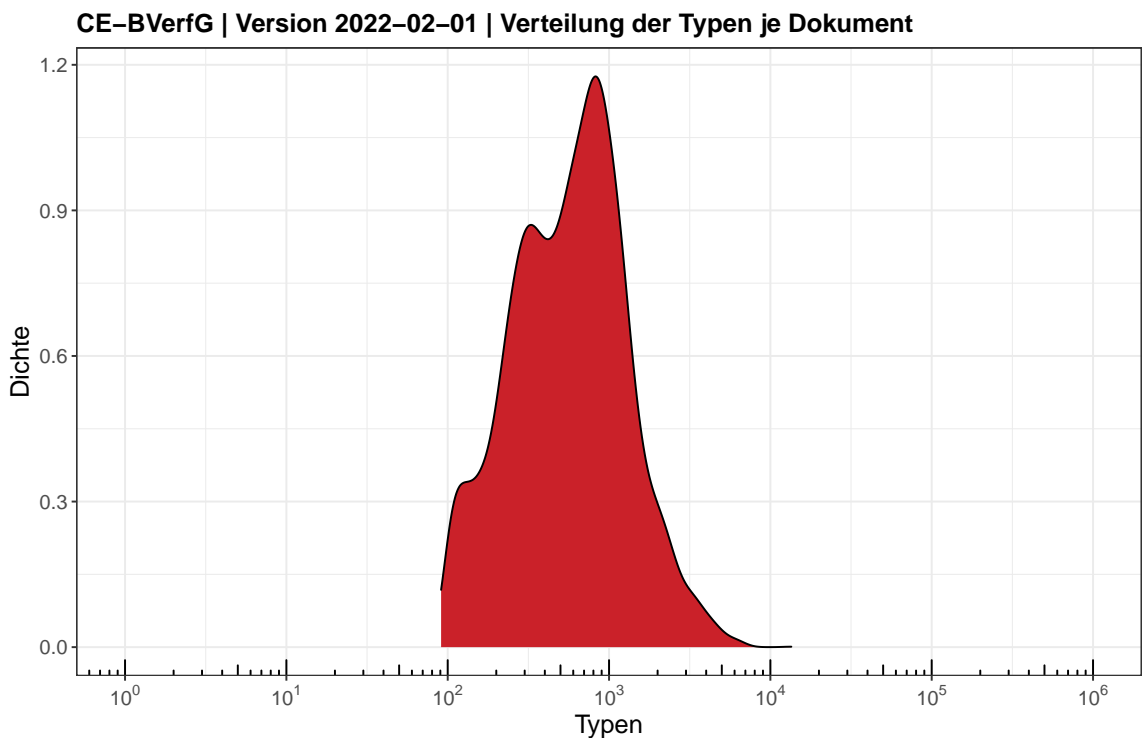
CE-BVerfG | Version 2022-02-01 | Verteilung der Tokens je Dokument



Fobbe | DOI: 10.5281/zenodo.5910152

12.7.3 Diagramm: Verteilung Typen

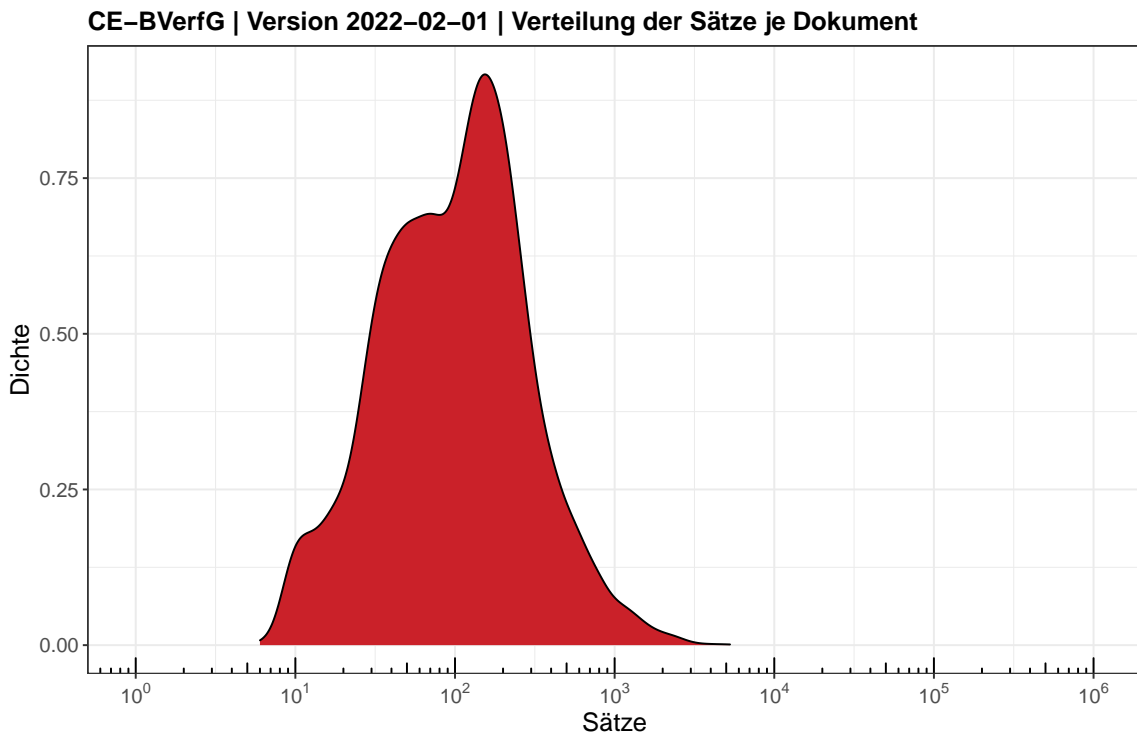
```
ggplot(data = meta.bverfg)+
  geom_density(aes(x = typen),
               fill = "#ca2129")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
               labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(prefix.figuretitle,
                  "| Verteilung der Typen je Dokument"),
    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.5910152

12.7.4 Diagramm: Verteilung Sätze

```
ggplot(data = meta.bverfg)+
  geom_density(aes(x = saetze),
               fill = "#ca2129")+
  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 Dokument"),
    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.5910152

13 Linguistische Annotationen berechnen

13.1 Funktion anzeigen: f.future_spacyparse

```
print(f.future_spacyparse)
```

```
## function(x,
##           chunkspartner = 1,
##           chunksize = NULL,
##           model = "en_core_web_sm",
##           pos = TRUE,
##           tag = FALSE,
##           lemma = FALSE,
##           entity = FALSE,
##           dependency = FALSE,
##           nounphrase = FALSE){
##
##   begin <- Sys.time()
##
##   spacy_initialize(model = model)
##
##   print(paste0("Begin at ",
##               begin,
##               ". Processing ",
##               x[,.N],
##               " documents"))
##
##
##   raw.list <- split(x, seq(nrow(x)))
##
##   result.list <- future_lapply(raw.list,
##                               spacy_parse,
##                               future.seed = TRUE,
##                               future.scheduling = chunkspartner,
##                               future.chunk.size = chunksize,
##                               pos = pos,
##                               tag = tag,
##                               lemma = lemma,
##                               entity = entity,
##                               dependency = dependency,
##                               nounphrase = nounphrase,
##                               multithread = FALSE)
##
##   result.dt <- rbindlist(result.list)
##
##
##   end <- Sys.time()
##   duration <- end - begin
```

```

##
##   print(paste0("Runtime was ",
##               round(duration,
##                     digits = 2),
##               " ",
##               attributes(duration)$units,
##               ". Ended at ",
##               end, "."))
##
##   spacy_finalize()
##
##   return(result.dt)
##
## }

```

13.2 Berechnungen durchführen

```

if (config$annotate$toggle == TRUE){

  if(config$parallel$spacyparse == TRUE){

    plan("multicore",
         workers = fullCores)

  }else{

    plan("sequential")

  }

  txt.annotated <- f.future_spacyparse(txt.bverfg,
                                       chunksperworker = 1,
                                       chunksize = NULL,
                                       model = "de_core_news_sm",
                                       pos = TRUE,
                                       tag = TRUE,
                                       lemma = TRUE,
                                       entity = TRUE,
                                       dependency = TRUE,
                                       nounphrase = TRUE)

}

```

14 Kontrolle der Variablen

14.1 Semantische Sortierung der Variablen

14.1.1 Variablen sortieren: Hauptdatensatz

```
setcolorder(txt.bverfg,  
            c("doc_id",  
              "text",  
              "gericht",  
              "datum",  
              "entscheidung_typ",  
              "spruchkoerper_typ",  
              "spruchkoerper_az",  
              "registerzeichen",  
              "verfahrensart",  
              "eingangsnummer",  
              "eingangsjahr_az",  
              "eingangsjahr_iso",  
              "entscheidungsjahr",  
              "kollision",  
              "name",  
              "band",  
              "seite",  
              "aktenzeichen",  
              "aktenzeichen_alle",  
              "ecli",  
              "zitiervorschlag",  
              "kurzbeschreibung",  
              "pressemitteilung",  
              "praesi",  
              "v_praesi",  
              "richter",  
              "zeichen",  
              "tokens",  
              "typen",  
              "saetze",  
              "version",  
              "doi_concept",  
              "doi_version",  
              "lizenz"))
```

14.1.2 Variablen sortieren: Metadaten

```
setcolorder(meta.bverfg,  
  c("doc_id",  
    "gericht",  
    "datum",  
    "entscheidung_typ",  
    "spruchkoerper_typ",  
    "spruchkoerper_az",  
    "registerzeichen",  
    "verfahrensart",  
    "eingangsnummer",  
    "eingangsjahr_az",  
    "eingangsjahr_iso",  
    "entscheidungsjahr",  
    "kollision",  
    "name",  
    "band",  
    "seite",  
    "aktenzeichen",  
    "aktenzeichen_alle",  
    "ecli",  
    "zitiervorschlag",  
    "kurzbeschreibung",  
    "pressemitteilung",  
    "praesi",  
    "v_praesi",  
    "richter",  
    "zeichen",  
    "tokens",  
    "typen",  
    "saetze",  
    "version",  
    "doi_concept",  
    "doi_version",  
    "lizenz"))
```

14.1.3 Variablen sortieren: Segmentiert

```
setcolorder(dt.segmented.full,  
  c("doc_id",  
    "text",  
    "segment",  
    "gericht",  
    "datum",  
    "entscheidung_typ",  
    "spruchkoerper_typ",  
    "spruchkoerper_az",  
    "registerzeichen",  
    "verfahrensart",  
    "eingangsnummer",  
    "eingangsjahr_az",  
    "eingangsjahr_iso",  
    "entscheidungsjahr",  
    "kollision",  
    "name",  
    "band",  
    "seite",  
    "aktenzeichen",  
    "aktenzeichen_alle",  
    "ecli",  
    "zitiervorschlag",  
    "kurzbeschreibung",  
    "pressemitteilung",  
    "praesi",  
    "v_praesi",  
    "richter",  
    "version",  
    "doi_concept",  
    "doi_version",  
    "lizenz"))
```


14.2 Anzahl Variablen der Datensätze

```
length(txt.bverfg)
```

```
## [1] 34
```

```
length(meta.bverfg)
```

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

```
length(txt.annotated)
```

```
## [1] 12
```

```
length(dt.segmented.full)
```

```
## [1] 31
```

14.3 Alle Variablen-Namen der Datensätze

```
names(txt.bverfg)
```

```
## [1] "doc_id"          "text"           "gericht"
## [4] "datum"          "entscheidung_typ" "spruchkoerper_typ"
## [7] "spruchkoerper_az" "registerzeichen" "verfahrensart"
## [10] "eingangsnummer" "eingangsjahr_az" "eingangsjahr_iso"
## [13] "entscheidungsjahr" "kollision"      "name"
## [16] "band"           "seite"          "aktenzeichen"
## [19] "aktenzeichen_alle" "ecli"           "zitiervorschlag"
## [22] "kurzbeschreibung" "pressemitteilung" "praesi"
## [25] "v_praesi"       "richter"        "zeichen"
## [28] "tokens"         "typen"          "saetze"
## [31] "version"        "doi_concept"    "doi_version"
## [34] "lizenz"
```

```
names(meta.bverfg)
```

```
## [1] "doc_id"          "gericht"         "datum"
## [4] "entscheidung_typ" "spruchkoerper_typ" "spruchkoerper_az"
## [7] "registerzeichen" "verfahrensart"    "eingangsnummer"
## [10] "eingangsjahr_az" "eingangsjahr_iso" "entscheidungsjahr"
## [13] "kollision"       "name"            "band"
## [16] "seite"           "aktenzeichen"   "aktenzeichen_alle"
## [19] "ecli"            "zitiervorschlag" "kurzbeschreibung"
## [22] "pressemitteilung" "praesi"          "v_praesi"
## [25] "richter"        "zeichen"         "tokens"
## [28] "typen"          "saetze"          "version"
## [31] "doi_concept"    "doi_version"     "lizenz"
```

```
names(txt.annotated)
```

```
## [1] "doc_id"          "sentence_id"     "token_id"        "token"
## [5] "lemma"          "pos"             "tag"             "head_token_id"
## [9] "dep_rel"        "entity"          "nounphrase"     "whitespace"
```

```
names(dt.segmented.full)
```

```
## [1] "doc_id"          "text"            "segment"
## [4] "gericht"         "datum"           "entscheidung_typ"
## [7] "spruchkoerper_typ" "spruchkoerper_az" "registerzeichen"
## [10] "verfahrensart"   "eingangsnummer"  "eingangsjahr_az"
## [13] "eingangsjahr_iso" "entscheidungsjahr" "kollision"
## [16] "name"           "band"            "seite"
## [19] "aktenzeichen"   "aktenzeichen_alle" "ecli"
## [22] "zitiervorschlag" "kurzbeschreibung" "pressemitteilung"
## [25] "praesi"         "v_praesi"        "richter"
## [28] "version"        "doi_concept"     "doi_version"
## [31] "lizenz"
```

15 CSV-Dateien erstellen

15.1 CSV mit vollem Datensatz speichern

```
csvname.full <- paste(prefix.files,  
                      "DE_CSV_Datensatz.csv",  
                      sep = "_")  
  
fwrite(txt.bverfg,  
       csvname.full,  
       na = "NA")
```

15.2 CSV mit Metadaten speichern

Diese Datei ist grundsätzlich identisch mit dem eigentlichen Datensatz, nur ohne den Text der Entscheidungen.

```
csvname.meta <- paste(prefix.files,  
                     "DE_CSV_Metadaten.csv",  
                     sep = "_")  
  
fwrite(meta.bverfg,  
       csvname.meta,  
       na = "NA")
```

15.3 CSV mit Segmenten speichern

```
csvname.segmented <- paste(prefix.files,  
                           "DE_CSV_Segmentiert.csv",  
                           sep = "_")  
  
fwrite(dt.segmented.full,  
       csvname.segmented,  
       na = "NA")
```

15.4 CSV mit Annotationen speichern

```
if (config$annotate$toggle == TRUE){  
  
  csvname.annotated <- paste(prefix.files,  
                             "DE_CSV_Annotiert.csv",  
                             sep = "_")  
  
  fwrite(txt.annotated,  
        csvname.annotated,  
        na = "NA")  
  
}
```

16 Dateigrößen analysieren

16.1 Gesamtgröße

16.1.1 Korpus-Objekt in RAM (MB)

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

```
## 171 MB
```

16.1.2 CSV Korpus (MB)

```
file.size(csvname.full) / 10 ^ 6
```

```
## [1] 168.5188
```

16.1.3 CSV Metadaten (MB)

```
file.size(csvname.meta) / 10 ^ 6
```

```
## [1] 5.762413
```

16.1.4 CSV Annotiert (MB)

```
file.size(csvname.annotated) / 10 ^ 6
```

```
## [1] 3046.186
```

16.1.5 CSV Segmentiert (MB)

```
file.size(csvname.segmented) / 10 ^ 6
```

```
## [1] 337.1686
```

16.1.6 PDF-Dateien (MB)

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

```
## [1] 665.7415
```

16.1.7 TXT-Dateien (MB)

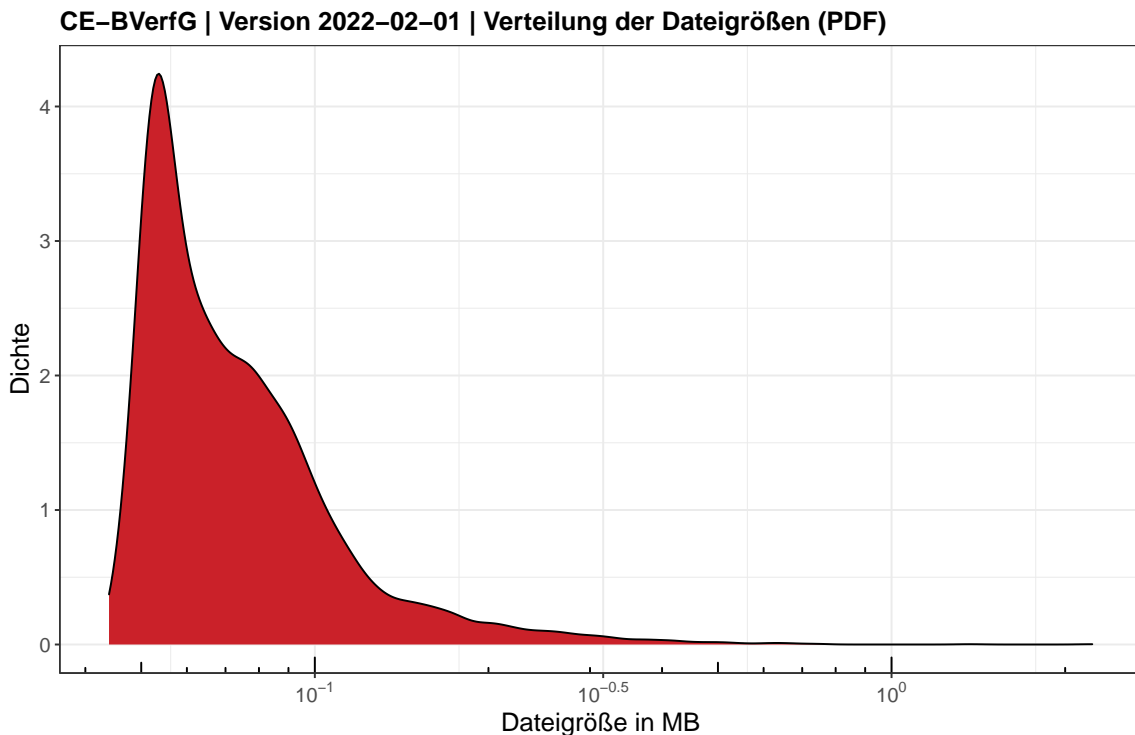
```
files.txt <- list.files(pattern = "\\..txt$",  
                       ignore.case = TRUE)  
  
txt.MB <- file.size(files.txt) / 10^6  
sum(txt.MB)
```

```
## [1] 165.2048
```

16.2 Diagramm: Verteilung der Dateigrößen (PDF)

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

```
ggplot(data = dt.plot,  
  aes(x = pdf.MB)) +  
  geom_density(fill = "#ca2129") +  
  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",  
    panel.spacing = unit(0.1, "lines"),  
    plot.margin = margin(10, 20, 10, 10)  
  )
```

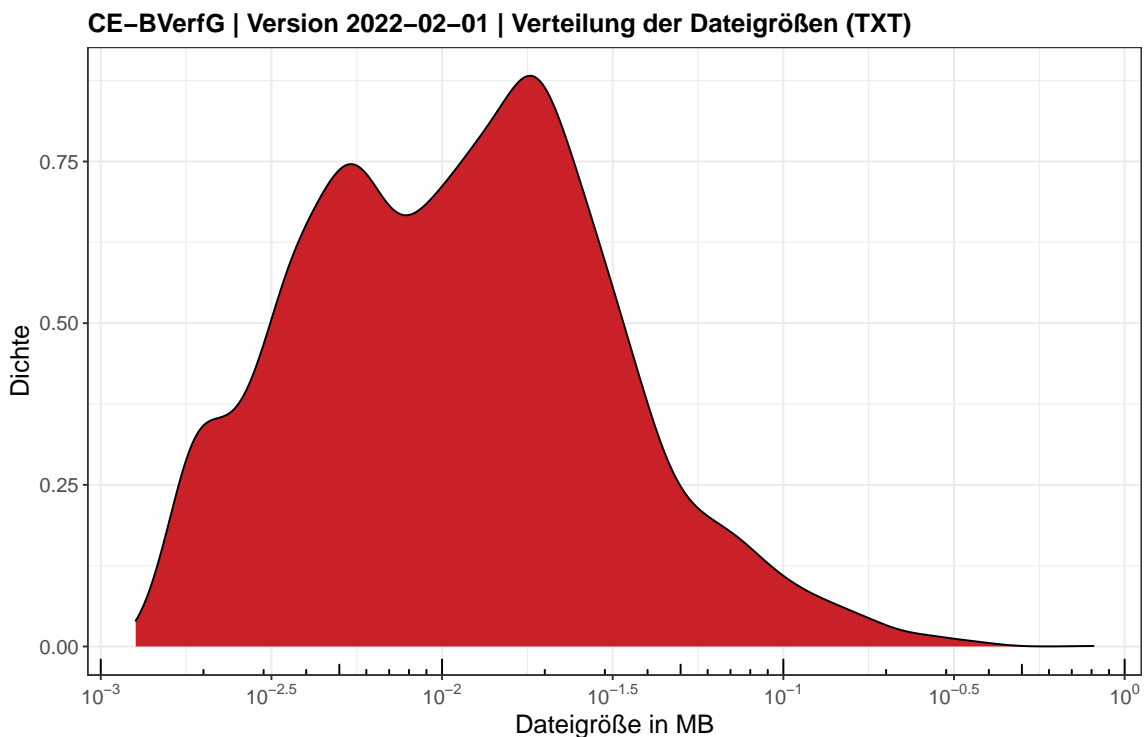


Fobbe | DOI: 10.5281/zenodo.5910152

16.3 Diagramm: Verteilung der Dateigrößen (TXT)

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

```
ggplot(data = dt.plot,  
  aes(x = txt.MB)) +  
  geom_density(fill = "#ca2129") +  
  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",  
    panel.spacing = unit(0.1, "lines"),  
    plot.margin = margin(10, 20, 10, 10)  
  )
```



Fobbe | DOI: 10.5281/zenodo.5910152

17 Erstellen der ZIP-Archive

17.1 Verpacken der CSV-Dateien

17.1.1 Vollständiger Datensatz

```
csvname.full.zip <- gsub(".csv",  
                        ".zip",  
                        csvname.full)  
  
zip(csvname.full.zip,  
    csvname.full)  
  
unlink(csvname.full)
```

17.1.2 Metadaten

```
csvname.meta.zip <- gsub(".csv",  
                        ".zip",  
                        csvname.meta)  
  
zip(csvname.meta.zip,  
    csvname.meta)  
  
unlink(csvname.meta)
```

17.1.3 Segmentiert

```
csvname.segmented.zip <- gsub(".csv",  
                              ".zip",  
                              csvname.segmented)  
  
zip(csvname.segmented.zip,  
    csvname.segmented)  
  
unlink(csvname.segmented)
```

17.1.4 Annotiert

```
if (config$annotate$toggle == TRUE){  
  
    csvname.annotated.zip <- gsub(".csv",  
                                  ".zip",  
                                  csvname.annotated)  
  
    zip(csvname.annotated.zip,  
        csvname.annotated)
```



```
    unlink(csvname.annotated)
}
```

17.2 Verpacken der PDF-Dateien

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

```
zip(paste(prefix.files,
           "DE_PDF_Datensatz.zip",
           sep = "_"),
    files.pdf)

unlink(files.pdf)
```

17.3 Verpacken der HTML-Dateien

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

```
zip(paste(prefix.files,
           "DE_HTML_Datensatz.zip",
           sep = "_"),
    files.html)

unlink(files.html)
```

17.4 Verpacken der TXT-Dateien

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

```
zip(paste(prefix.files,
           "DE_TXT_Datensatz.zip",
           sep = "_"),
    files.txt)

unlink(files.txt)
```

17.5 Verpacken der Analyse-Dateien

```
zip(paste0(prefix.files,  
           "_DE_ANALYSE.zip"),  
    basename(dir.analysis))
```

17.6 Verpacken der Source-Dateien

```
files.source <- c(list.files(pattern = "\\\\.R$|\\.toml$"),  
                 "CHANGELOG.md",  
                 "README.md",  
                 "R-fobbe-proto-package",  
                 "buttons",  
                 "data",  
                 "functions",  
                 "tex",  
                 "gpg",  
                 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(paste(prefix.files,  
          "Source_Files.zip",  
          sep = "_"),  
    files.source)
```

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 sind äußerst resistent gegenüber *collision* und *pre-imaging* Angriffen, sie gelten derzeit als kryptographisch sicher. Ein SHA3-Hash mit 512 bit Länge ist nach Stand von Wissenschaft und Technik auch gegenüber quantenkryptoanalytischen Verfahren unter Einsatz des *Grover-Algorithmus* 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(f.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,
##                                              f.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)$units,
##                 ".")
##
##   return(hash$hashes)
## }

```

18.3 Hashes berechnen

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

multihashes <- f.future_multihashes(files.zip)

```

```
## Processing 9 files. Begin at: 2022-02-01 20:46:07

```

```
## Processed 9 files. Runtime was 3.3 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 In Datei schreiben

```
fwrite(multihashes,
      file.path("output",
                paste(prefix.files,
                      "KryptographischeHashes.csv",
                      sep = "_")),
      na = "NA")
```

18.7 Leerzeichen hinzufügen um Zeilenumbruch zu ermöglichen

Hierbei handelt es sich lediglich um eine optische Notwendigkeit. Die normale 128 Zeichen lange Zeichenfolge wird ansonsten nicht umgebrochen und verschwindet über die Seiten-
grenze. Das Leerzeichen erlaubt den automatischen Zeilenumbruch und damit einen für
Menschen sinnvoll lesbaren Abdruck im Codebook. Diese Variante wird nur zur Anzeige
verwendet und danach verworfen.

```
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	CE-BVerfG_2022-02-01_DE_ANALYSE.zip
2	CE-BVerfG_2022-02-01_DE_CSV_Annotiert.zip
3	CE-BVerfG_2022-02-01_DE_CSV_Datensatz.zip
4	CE-BVerfG_2022-02-01_DE_CSV_Metadaten.zip
5	CE-BVerfG_2022-02-01_DE_CSV_Segmentiert.zip
6	CE-BVerfG_2022-02-01_DE_HTML_Datensatz.zip
7	CE-BVerfG_2022-02-01_DE_PDF_Datensatz.zip
8	CE-BVerfG_2022-02-01_DE_TXT_Datensatz.zip
9	CE-BVerfG_2022-02-01_Source_Files.zip

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

index	sha2.256
1	d28272e47975649c065bf961de504c88755399cdefd2764585e1343f16ff2e03
2	1a0750699de93aa67896e69c53720046d347aa0e54e734a2ecb89530355d5580
3	0f4979e17333e0ea440f871550fa6fa6c04bf718d89fa33aa0bbc36839fe9f6c
4	31cf8ad829f606618d8bb57c12065fe0899561e3e1be6fe410c0c83287006a70
5	79c990c3072207d1c72271b09a00cf833d317459919dae397022229b51cfaa6f
6	fe439eacb68fdd80ba93c41a842f8542b855ab555ae95d37d81b01f54be94a04
7	6c04373d829daa0a7292521e6c63e5e1cc50a5948983ab20dcdb3dc25d7e47aa
8	af690be76001aab0bce1c8084c61e42c6a9404b15b9c141e7dcf88ce7b0b2bde
9	163a748aee0b975adedd3986fe48e9a8fe2155677b317703a37ab15211222af0

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

index	sha3.512
1	f3f6c2a5cbf4891cddfb2366e71450088ce055b52b6789c0e028fbb93e7a9cb1 afa6e2ebaef566cc4fb19ae1bfc47ac8eeb6ef9148a8e44a42f3eee8489e6352
2	e920f6efc832ae20407cc8312b111d9cb349652e5caebd2fe1979054e06d432d 1d55235077384added723141419c015b683dd11567ff7a62dd01186aa9749c02
3	d5bff447fd17c03701958e269c47e341b4beac181ca3d37b52535b589f180acd 81b03a058d7f9cd9c470284d9da9b95db9d5713b2c77322bf93f7b9032e7c605
4	5585c111a7967af020b9513a0536eae69e6ddabb92d976fafdd9e4a004e3b27c d9495e59963cc17fc8545dd06bcf683a94f3a7e4dd91d2f759727438d0123086
5	60d3c822b300253a0f65ee32fa3ce939cf7b3fabb7a7dd5006074f60d5a05fe6 0347e6bf38ade3774cbceb389bfac2f38c16441abdb827d787d44b17c1b7f294

- 6 67f4b036e92e6336ba6ed06a823e79cd1985b80527c3e4d32c873f344e53338f
d45f4d0503a03490b4f1d1f8b3d7ee1afdfc89f474f29d5cd200fb3a0ba5c509
 - 7 cb00b57b72c1430fcdbb4f441fe3d415f46193b4dc25e5d480bce0d0f7404a76
f98818bad0c1de2900e71b82873b7d96092f26f652239224542ec70fd413475c
 - 8 b36ae6f21e8a2002d664999ab3f14544060c7c9703f4add711cd0f3b2713f1cb
5c3c98be21d77d38a2975a28c3900cc25eacb942ea8a11fef4f90fc758309978
 - 9 d45c25fceba7df50ff1cb0297ff321ea9b966d2e2ea70a507d3a0e89bf03aa2e
aad1372cbf4a621e8570958b201dc2c44a70d15f724618f12188b836c47a8a3d
-

19 Aufräumen

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

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

print(files.output)
```

```
## [1] "CE-BVerfG_2022-02-01_DE_ANALYSE.zip"
## [2] "CE-BVerfG_2022-02-01_DE_CSV_Annotiert.zip"
## [3] "CE-BVerfG_2022-02-01_DE_CSV_Datensatz.zip"
## [4] "CE-BVerfG_2022-02-01_DE_CSV_Metadaten.zip"
## [5] "CE-BVerfG_2022-02-01_DE_CSV_Segmentiert.zip"
## [6] "CE-BVerfG_2022-02-01_DE_HTML_Datensatz.zip"
## [7] "CE-BVerfG_2022-02-01_DE_PDF_Datensatz.zip"
## [8] "CE-BVerfG_2022-02-01_DE_TXT_Datensatz.zip"
## [9] "CE-BVerfG_2022-02-01_Source_Files.zip"
```

```
file.rename(files.output,
            output.destination)
```

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


20 Abschluss

20.1 Datumsstempel

```
print(datestamp)
```

```
## [1] "2022-02-01"
```

20.2 Datum und Uhrzeit (Anfang)

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

```
## [1] "2022-02-01 12:23:14 CET"
```

20.3 Datum und Uhrzeit (Ende)

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

```
## [1] "2022-02-01 20:46:10 CET"
```

20.4 Laufzeit des gesamten Skriptes

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

```
## Time difference of 8.382178 hours
```

20.5 Warnungen

```
warnings()
```

21 Parameter für strenge Replikationen

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

```
## [1] "OpenSSL 1.1.1l 24 Aug 2021"
```

```
sessionInfo()
```

```
## R version 4.0.5 (2021-03-31)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 34 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libflexiblas.so.3.0
##
## 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] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
##  [1] future.apply_1.8.1 future_1.23.0  spacyr_1.2.1  quanteda_3.2.0
##  [5] readtext_0.81 data.table_1.14.2 scales_1.1.1  ggplot2_3.3.5
##  [9] pdftools_3.0.1 kableExtra_1.3.4 knitr_1.33    rvest_1.0.2
## [13] httr_1.4.2      mgsub_1.7.3   RcppTOML_0.1.7
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.7      here_1.0.1    svglite_2.0.0  lattice_0.20-45
##  [5] listenv_0.8.0  png_0.1-7     rprojroot_2.0.2 digest_0.6.29
##  [9] utf8_1.2.2     parallelly_1.30.0 R6_2.5.1      evaluate_0.14
## [13] highr_0.9      pillar_1.6.4  rlang_0.4.12  curl_4.3.2
## [17] rstudioapi_0.13 magick_2.7.3  Matrix_1.4-0  reticulate_1.22
## [21] rmarkdown_2.11 qpdf_1.1      labeling_0.4.2 webshot_0.5.2
## [25] stringr_1.4.0  selectr_0.4-2 munsell_0.5.0 compiler_4.0.5
## [29] xfun_0.29      pkgconfig_2.0.3 askpass_1.1    systemfonts
##          _1.0.3
## [33] globals_0.14.0  htmltools_0.5.2  tibble_3.1.6  codetools_0.2-18
## [37] fansi_1.0.0     viridisLite_0.4.0 crayon_1.4.2  withr_2.4.3
## [41] rappdirs_0.3.3  grid_4.0.5     jsonlite_1.7.2  gtable_0.3.0
## [45] lifecycle_1.0.1 magrittr_2.0.1  RcppParallel_5.1.4 stringi_1.7.6
## [49] farver_2.1.0    renv_0.15.0     xml2_1.3.3     ellipsis_0.3.2
## [53] stopwords_2.3   vctrs_0.3.8    fastmatch_1.1-3 tools_4.0.5
## [57] glue_1.6.0      parallel_4.0.5  fastmap_1.1.0  yaml_2.2.1
```

```
## [61] colorspace_2.0-2
```

Literaturverzeichnis

- Bengtsson, Henrik. 2021a. “A Unifying Framework for Parallel and Distributed Processing in R Using Futures.” <https://journal.r-project.org/archive/2021/RJ-2021-048/index.html>.
- . 2021b. “A Unifying Framework for Parallel and Distributed Processing in R Using Futures.” <https://journal.r-project.org/archive/2021/RJ-2021-048/index.html>.
- . 2021c. *Future.apply: Apply Function to Elements in Parallel Using Futures*. <https://CRAN.R-project.org/package=future.apply>.
- . 2021d. *Future: Unified Parallel and Distributed Processing in R for Everyone*. <https://CRAN.R-project.org/package=future>.
- Benoit, Kenneth, and Akitaka Matsuo. 2020. *Spacyr: Wrapper to the spaCy 'Nlp' Library*. <https://spacyr.quanteda.io>.
- Benoit, Kenneth, and Adam Obeng. 2021. *Readtext: Import and Handling for Plain and Formatted Text Files*. <https://github.com/quanteda/readtext>.
- 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. 2021. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- 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>.
- Ewing, Mark. 2021. *Mgsub: Safe, Multiple, Simultaneous String Substitution*. <https://CRAN.R-project.org/package=mgsub>.
- Ooms, Jeroen. 2021. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2021. *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>.
- . 2020. *Httr: Tools for Working with Urls and Http*. <https://CRAN.R-project.org/package=httr>.
- . 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. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.

- Wickham, Hadley, and Dana Seidel. 2020. *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/isbn/9781466561595>.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2021. *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>.