

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

COMPILATION REPORT

Version 2021-09-19

License MIT-0

DOI: 10.5281/zenodo.5514084

Titel	Source Code des »Corpus der Entscheidungen des Bundesverfassungsgerichts«
Abkürzung	CE-BVerfG-Source
Autor	Seán Fobbe
Version	2021-09-19
Download	https://doi.org/10.5281/zenodo.5514084
Lizenz	MIT No Attribution (MIT-0)

Zitiervorschlag

Seán Fobbe (2021). Source Code des »Corpus der Entscheidungen des Bundesverfassungsgerichts« (CE-BVerfG-Source). Version 2021-09-19. Zenodo. DOI: 10.5281/zenodo.5514084.

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 2021-09-19. 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 — 2021 — 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	Einleitung	9
1.1	Überblick	9
1.2	Funktionsweise	9
1.3	Systemanforderungen	9
1.4	Kompilierung	10
1.4.1	Datensatz	10
1.4.2	Codebook	10
2	Parameter	11
2.1	Name des Datensatzes	11
2.2	DOI des Datensatz-Konzeptes	11
2.3	DOI der konkreten Version	11
2.4	Lizenz	11
2.5	Verzeichnis für Analyse-Ergebnisse	11
2.6	Modus: Debugging	11
2.7	Modus: Linguistische Annotationen	12
2.8	Optionen: Quanteda	12
2.9	Optionen: Knitr	12
2.9.1	Ausgabe-Formate	12
2.9.2	Auflösung der Raster-Grafiken	12
2.9.3	Ausrichtung von Grafiken im Compilation Report	12
2.10	Frequenztabellen: Ignorierte Variablen	12
3	Vorbereitung	13
3.1	Datumsstempel	13
3.2	Datum und Uhrzeit (Beginn)	13
3.3	Ordner für Analyse-Ergebnisse erstellen	13
3.4	Packages Laden	13
3.5	Zusätzliche Funktionen einlesen	14
3.6	Quanteda-Optionen setzen	14
3.7	Knitr Optionen setzen	15
3.8	Vollzitate statistischer Software	15
3.9	Parallelisierung aktivieren	15
3.9.1	Anzahl logischer Kerne bestimmen	15
3.9.2	Quanteda	15
3.9.3	Data.table	15
4	Download: Weitere Datensätze	16
4.1	Registerzeichen und Verfahrensarten	16
4.2	Personendaten zu Präsident:innen	16
4.3	Personendaten zu Vize-Präsident:innen	16
5	Funktionen definieren	17
5.1	f.bverfg.extract.meta	17
5.2	f.bverfg.extract.content	18
6	Download vorbereiten: Alle Entscheidungen des BVerfG	23
6.1	Zeitstempel: Linksammlung Beginn	23

6.2	Funktion zeigen	23
6.3	Maximale Seitenzahl auslesen	23
6.4	Auszuwertende Seiten	24
6.5	[Debugging Modus] Reduzierung der Seitenzahl	24
6.6	Linkliste erstellen: Erstversuch	24
6.7	Linkliste erstellen: Zweitversuch	26
6.7.1	Seiten für Zweitversuch definieren	26
6.7.2	Anzeigen der Seiten die noch einmal geprüft werden.	26
6.7.3	Seiten prüfen	26
6.7.4	Anzeigen der Links die beim Wiederholungsversuch gesammelt wurden.	27
6.8	Zeitstempel: Linksammlung Ende	27
6.9	Dauer: Linksammlung	27
6.10	Listen aus Erstversuch und Zweitversuch zusammenfügen	27
6.11	Bereinigte HTML-Links definieren	28
6.12	Bereinigte PDF-Links definieren	28
6.13	Dateinamen erstellen	28
6.13.1	Extrahieren relevanter Metadaten	28
6.13.2	Formatierung von Registerzeichen anpassen	29
6.13.3	Formatierung von Spruchkörper-Typ anpassen	29
6.13.4	Erste strenge REGEX-Validierung der Dateinamen	29
6.13.5	Ergebnis der ersten REGEX-Validierung	29
6.13.6	Skript stoppen falls erste REGEX-Validierung gescheitert	30
6.13.7	Zusätzliche Variablen einfügen	30
6.13.8	NAs einfügen für nicht benannte Entscheidungen	30
6.13.9	Zweite strenge REGEX-Validierung der Dateinamen	31
6.13.10	Ergebnis der zweiten REGEX-Validierung	31
6.13.11	Skript stoppen falls zweite REGEX-Validierung gescheitert	31
7	PDF-Download	32
7.1	Data Table für PDF-Download erstellen	32
7.2	Zeitstempel: PDF-Download Beginn	32
7.3	PDF-Download durchführen	32
7.4	Zeitstempel: PDF-Download Ende	33
7.5	Dauer: PDF-Download	33
7.6	PDF-Download: Ergebnis	33
7.6.1	Anzahl herunterzuladender Dateien	33
7.6.2	Anzahl heruntergeladener Dateien	33
7.6.3	Fehlbetrag	34
7.6.4	Fehlende Dateien	34
7.7	PDF-Wiederholungsversuch	34
7.8	PDF-Download: Gesamtergebnis	35
7.8.1	Anzahl herunterzuladender Dateien	35
7.8.2	Anzahl heruntergeladener Dateien	35
7.8.3	Fehlbetrag	35
7.8.4	Fehlende Dateien	36
7.8.5	Abschließende Hinweise	36
8	HTML-Download	37
8.1	Data Table für HTML-Download erstellen	37
8.2	Zeitstempel: HTML-Download Beginn	37

8.3	HTML-Download durchführen	37
8.4	Zeitstempel: HTML-Download Ende	37
8.5	Dauer: HTML-Download	38
8.6	HTML-Download: Ergebnis	38
8.6.1	Anzahl herunterzuladender Dateien	38
8.6.2	Anzahl heruntergeladener Dateien	38
8.6.3	Fehlbetrag	38
8.6.4	Fehlende Dateien	38
8.7	HTML-Wiederholungsversuch	39
8.8	HTML-Download: Gesamtergebnis	39
8.8.1	Anzahl herunterzuladender Dateien	39
8.8.2	Anzahl heruntergeladener Dateien	39
8.8.3	Fehlbetrag	39
8.8.4	Fehlende Dateien	40
9	HTML verarbeiten	41
9.1	HTML-Dateien definieren	41
9.2	HTML-Dateien einlesen	41
9.3	HTML-Dateien parsen	41
9.4	Data Table mit allen Metadaten (inkl. ECLI)	41
9.5	Data Table mit vollständiger segmentierter Variante	41
9.6	Special Character entfernen	41
9.7	Stichprobe Metadaten	42
9.8	Stichprobe Segmentierte Variante	42
10	Text-Extraktion aus PDF	43
10.1	Vektor der zu extrahierenden Dateien erstellen	43
10.2	Anzahl zu extrahierender Dateien	43
10.3	PDF extrahieren: Funktion anzeigen	43
10.4	Text Extrahieren	44
11	Korpus Erstellen	45
11.1	TXT-Dateien Einlesen	45
11.2	In Data Table umwandeln	45
11.3	Durch Zeilenumbruch getrennte Wörter zusammenfügen	45
11.3.1	Funktion anzeigen	45
11.3.2	Funktion ausführen	46
11.4	Variable "datum" als Datentyp "IDate" kennzeichnen	46
11.5	Variable "entscheidungsjahr" hinzufügen	46
11.6	Variable "eingangsjahr_iso" hinzufügen	46
11.7	Datensatz nach Datum sortieren	46
11.8	Variable "praesi" hinzufügen	47
11.8.1	Lebensdaten einlesen	47
11.8.2	Personaldaten anzeigen	47
11.8.3	Hypothetisches Amtsende für PräsidentIn	47
11.8.4	Schleife vorbereiten	48
11.8.5	Vektor erstellen	48
11.8.6	Vektor einfügen	48
11.9	Variable "v_praesi" hinzufügen	48
11.9.1	Personaldaten einlesen	48

11.9.2	Personaldaten anzeigen	48
11.9.3	Hypothetisches Amtsende für Vize-PräsidentIn	49
11.9.4	Schleife vorbereiten	49
11.9.5	Vektor erstellen	49
11.9.6	Vektor einfügen	50
11.10	Variable “verfahrensart” hinzufügen	50
11.10.1	Datensatz einlesen	50
11.10.2	Datensatz auf relevante Daten reduzieren	50
11.10.3	Indizes bestimmen	50
11.10.4	Vektor der Verfahrensarten erstellen und einfügen	50
11.11	Variable “aktenzeichen” hinzufügen	50
11.12	Variable “doi_concept” hinzufügen	51
11.13	Variable “doi_version” hinzufügen	51
11.14	Variable “version” hinzufügen	51
11.15	Variable “lizenz” hinzufügen	51
11.16	Variable “ecli” hinzufügen	51
11.16.1	Metadaten mit ECLI-Duplikaten entfernen	52
11.16.2	Fehlerhafte ECLI korrigieren (HTML Meta)	52
11.16.3	Fehlerhafte ECLI korrigieren (HTML Full)	53
11.16.4	ECLI-Test 1: ECLI die in PDF, aber nicht in HTML vorhanden sind	54
11.16.5	ECLI-Test 2: ECLI die in HTML, aber nicht in PDF vorhanden sind	54
11.16.6	Zum Vergleich: Beim Download fehlende PDF-Dateien	54
11.16.7	ECLI Merge: Metadaten aus Hauptdatensatz in segmentierte Variante mergen	54
11.16.8	ECLI Merge: Metadaten aus HTML-Extraktion in Hauptdatensatz mergen	55
11.17	Variable “entscheidung_typ” hinzufügen	55
11.17.1	Zitiervorschläge parsen	55
11.17.2	Kürzen	55
11.17.3	Vektor in Datensatz einfügen	55
12	Frequenztabellen erstellen	57
12.1	Funktion anzeigen	57
12.2	Ignorierte Variablen	58
12.3	Liste zu prüfender Variablen	58
12.4	Frequenztabellen erstellen	59
13	Frequenztabellen visualisieren	99
13.1	Präfix erstellen	99
13.2	Tabellen einlesen	99
13.3	Diagramm: Typ der Entscheidung	100
13.4	Diagramm: Typ des Spruchkörpers	102
13.5	Diagramm: Spruchkörper nach Aktenzeichen	104
13.6	Diagramm: Registerzeichen	106
13.7	Diagramm: Präsident:in	108
13.8	Diagramm: Vize-Präsident:in	110
13.9	Diagramm: Entscheidungsjahr	112
13.10	Diagramm: Eingangsjahr (ISO)	114
14	Korpus-Analytik	116

14.1	Berechnung linguistischer Kennwerte	116
14.1.1	Funktion anzeigen	116
14.1.2	Berechnung durchführen	118
14.2	Variablen-Namen anpassen	118
14.3	Kennwerte dem Korpus hinzufügen	118
14.4	Variante mit Metadaten erstellen	118
14.5	Linguistische Kennwerte	119
14.5.1	Zusammenfassungen berechnen	119
14.5.2	Zusammenfassungen anzeigen	120
14.5.3	Zusammenfassungen speichern	120
14.6	Quantitative Variablen	121
14.6.1	Entscheidungsdatum	121
14.6.2	Zusammenfassungen berechnen	121
14.6.3	Zusammenfassungen anzeigen	122
14.6.4	Zusammenfassungen speichern	122
14.7	Verteilungen linguistischer Kennwerte	123
14.7.1	Diagramm: Verteilung Zeichen	123
14.7.2	Diagramm: Verteilung Tokens	124
14.7.3	Diagramm: Verteilung Typen	126
14.7.4	Diagramm: Verteilung Sätze	127
15	Linguistische Annotationen berechnen	128
16	Kontrolle der Variablen	129
16.1	Semantische Sortierung der Variablen	129
16.1.1	Variablen sortieren: Hauptdatensatz	129
16.1.2	Variablen sortieren: Metadaten	130
16.1.3	Variablen sortieren: Segmentiert	131
16.2	Anzahl Variablen der Datensätze	132
16.3	Alle Variablen-Namen der Datensätze	132
17	CSV-Dateien erstellen	134
17.1	CSV mit vollem Datensatz speichern	134
17.2	CSV mit Metadaten speichern	134
17.3	CSV mit Segmenten speichern	134
17.4	CSV mit Annotationen speichern	134
18	Dateigrößen analysieren	136
18.1	Gesamtgröße	136
18.1.1	Korpus-Objekt in RAM (MB)	136
18.1.2	CSV Korpus (MB)	136
18.1.3	CSV Metadaten (MB)	136
18.1.4	CSV Annotiert (MB)	136
18.1.5	CSV Segmentiert (MB)	136
18.1.6	PDF-Dateien (MB)	137
18.1.7	TXT-Dateien (MB)	137
18.2	Diagramm: Verteilung der Dateigrößen (PDF)	138
18.3	Diagramm: Verteilung der Dateigrößen (TXT)	140
19	Erstellen der ZIP-Archive	142

19.1	Verpacken der CSV-Dateien	142
19.1.1	Vollständiger Datensatz	142
19.1.2	Metadaten	142
19.1.3	Segmentiert	142
19.1.4	Annotiert	142
19.2	Verpacken der PDF-Dateien	143
19.3	Verpacken der HTML-Dateien	143
19.4	Verpacken der TXT-Dateien	143
19.5	Verpacken der Analyse-Dateien	144
19.6	Verpacken der Source-Dateien	144
20	Kryptographische Hashes	145
20.1	Liste der ZIP-Archive erstellen	145
20.2	Funktion anzeigen	145
20.3	Hashes berechnen	146
20.4	In Data Table umwandeln	146
20.5	Index hinzufügen	146
20.6	In Datei schreiben	147
20.7	Leerzeichen hinzufügen um Zeilenumbruch zu ermöglichen	147
20.8	In Bericht anzeigen	147
21	Abschluss	150
21.1	Datumsstempel	150
21.2	Datum und Uhrzeit (Anfang)	150
21.3	Datum und Uhrzeit (Ende)	150
21.4	Laufzeit des gesamten Skriptes	150
21.5	Warnungen	150
22	Parameter für strenge Replikationen	151
	Literaturverzeichnis	153

1 Einleitung

1.1 Überblick

Dieses R-Skript 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 sind folgende ZIP-Archive:

1. Der volle Datensatz im CSV-Format
2. Die reinen Metadaten im CSV-Format (wie unter 1, nur ohne Entscheidungstexte)
3. (Optional) Tokenisierte Form aller Texte mit linguistischen Annotationen im CSV-Format
4. Alle Entscheidungen im TXT-Format (reduzierter Umfang an Metadaten)
5. Alle Entscheidungen im PDF-Format (reduzierter Umfang an Metadaten)
6. Alle Analyse-Ergebnisse (Tabellen als CSV, Grafiken als PDF und PNG)
7. 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. Weiterhin kann optional ein PDF-Bericht erstellt werden (siehe unter “Kompilierung”).

1.3 Systemanforderungen

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 dieses Berichts.

In der Standard-Einstellung wird das Skript vollautomatisch die maximale Anzahl an Rechenkernen/Threads auf dem System zu nutzen. Wenn die Anzahl Threads (Variable “fullCores”) auf 1 gesetzt wird, ist die Parallelisierung deaktiviert.

Auf der Festplatte sollten 8 GB Speicherplatz vorhanden sein.

Um die PDF-Berichte kompilieren zu können benötigen Sie das R package **rmarkdown**, eine vollständige Installation von L^AT_EX und alle in der Präambel-TEX-Datei angegebenen L^AT_EX Packages.

1.4 Kompilierung

Mit der Funktion `render()` von **rmarkdown** können der **vollständige Datensatz** und das **Codebook** kompiliert und die Skripte mitsamt ihrer Rechenergebnisse in ein gut lesbares PDF-Format überführt werden.

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.

1.4.1 Datensatz

Um den vollständigen Datensatz zu kompilieren und einen PDF-Bericht zu erstellen, kopieren Sie bitte alle im Source-Archiv bereitgestellten Dateien in einen leeren Ordner und führen mit R diesen Befehl aus:

```
rmarkdown::render(input = "CE-BVerfG_Source_CorpusCreation.R",
                  output_file = paste0("CE-BVerfG_",
                                       Sys.Date(),
                                       "_CompilationReport.pdf"),
                  envir = new.env())
```

1.4.2 Codebook

Um das **Codebook** zu kompilieren und einen PDF-Bericht zu erstellen führen Sie bitte im Anschluss an die Kompilierung des Datensatzes untenstehenden Befehl mit R aus.

Bei der Prüfung der GPG-Signatur 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.

```
rmarkdown::render(input = "CE-BVerfG_Source_CodebookCreation.R",
                  output_file = paste0("CE-BVerfG_",
                                       Sys.Date(),
                                       "_Codebook.pdf"),
                  envir = new.env())
```

2 Parameter

2.1 Name des Datensatzes

```
datasetname <- "CE-BVerfG"
```

2.2 DOI des Datensatz-Konzeptes

```
doi.concept <- "10.5281/zenodo.3902658" # checked
```

2.3 DOI der konkreten Version

```
doi.version <- "10.5281/zenodo.5514083" # checked
```

2.4 Lizenz

```
license <- "Creative Commons Zero 1.0 Universal"
```

2.5 Verzeichnis für Analyse-Ergebnisse

Hinweis: Muss mit einem Schrägstrich enden!

```
outputdir <- paste0(getwd(),  
                    "/ANALYSE/")
```

2.6 Modus: Debugging

Der Debugging-Modus reduziert den Such-Umfang auf den in der Variable “debug.scope” angegebenen Umfang Seiten (jede Seite enthält idR 10 Entscheidungen). Muss mindestens 10 betragen. Nur für Test- und Demonstrationszwecke. Kann mit anderen Modi kombiniert werden.

```
mode.debug <- FALSE  
debug.scope <- 10
```

2.7 Modus: Linguistische Annotationen

Wenn dieser Modus aktiviert ist wird mittels spacyr eine zusätzliche Variante des Datensatzes mit umfangreichen linguistischen Annotationen berechnet. Dieser Modus ist sehr rechenintensiv! Kann mit anderen Modi kombiniert werden.

```
mode.annotate <- TRUE
```

2.8 Optionen: Quanteda

```
tokens_locale <- "de_DE"
```

2.9 Optionen: Knitr

2.9.1 Ausgabe-Formate

```
dev <- c("pdf",  
        "png")
```

2.9.2 Auflösung der Raster-Grafiken

```
dpi <- 300
```

2.9.3 Ausrichtung von Grafiken im Compilation Report

```
fig.align <- "center"
```

2.10 Frequenztabellen: Ignorierte Variablen

Diese Variablen werden bei der Erstellung der Frequenztabellen nicht berücksichtigt.

```
varremove <- c("text",  
              "eingangsnummer",  
              "datum",  
              "doc_id",  
              "seite",  
              "name",  
              "ecli",  
              "aktenzeichen",  
              "pressemitteilung",  
              "zitiervorschlag",  
              "kurzbeschreibung")
```

3 Vorbereitung

3.1 Datumsstempel

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

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

```
## [1] "2021-09-19"
```

3.2 Datum und Uhrzeit (Beginn)

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

```
## [1] "2021-09-19 11:22:19 CEST"
```

3.3 Ordner für Analyse-Ergebnisse erstellen

```
dir.create(outputdir)
```

3.4 Packages Laden

```
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
```

```
## Loading required package: foreach
```

```
## Loading required package: iterators
```

```
## Loading required package: parallel
```

```
library(ggplot2)      # Fortgeschrittene Datenvisualisierung  
library(scales)       # Skalierung von Diagrammen  
library(data.table)   # Fortgeschrittene Datenverarbeitung
```

```
## data.table 1.14.0 using 8 threads (see ?getDTthreads). Latest news: r-  
datatable.com
```

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

```
## Package version: 3.1.0  
## 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
```

3.5 Zusätzliche Funktionen einlesen

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

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

3.6 Quanteda-Optionen setzen

```
quanteda_options(tokens_locale = tokens_locale)
```

3.7 Knitr Optionen setzen

```
knitr::opts_chunk$set(fig.path = outputdir,  
                      dev = dev,  
                      dpi = dpi,  
                      fig.align = fig.align)
```

3.8 Vollzitate statistischer Software

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

```
## tweaking foreach
```

3.9 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.

Der Download der Daten ist absichtlich nicht parallelisiert, damit das Skript nicht versehentlich als DoS-Tool verwendet wird.

Die hier verwendete Funktion **makeForkCluster()** ist viel schneller als die Alternativen, funktioniert aber nur auf Unix-basierten Systemen (Linux, MacOS).

3.9.1 Anzahl logischer Kerne bestimmen

```
fullCores <- detectCores()  
print(fullCores)
```

```
## [1] 16
```

3.9.2 Quanteda

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

3.9.3 Data.table

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

4 Download: Weitere Datensätze

4.1 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("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",  
  "AZ-BRD_1-0-1_DE_Registerzeichen_Datensatz.csv")  
}
```

4.2 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("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",  
  "PVP-FCG_2021-04-08_GermanFederalCourts_Presidents.csv")  
}
```

4.3 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("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",  
  "PVP-FCG_2021-04-08_GermanFederalCourts_VicePresidents.csv")  
}
```


5 Funktionen definieren

5.1 f.bverfg.extract.meta

Diese Funktion extrahiert relevante Metadaten aus einer auf einer HTML-Seite dargestellten Entscheidung auf www.bundesverfassungsgericht.de

```
f.bverfg.extract.meta <- function(HTML){

  ## ECLI
  eclli <- html_elements(HTML, "[class='eccli']") %>% 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"
##}

## Zitervorschlag
zitervorschlag <- 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(eccli,
                  pressemitteilung,
                  aktenzeichen_alle,
                  ##verkuendung,
                  kurzbeschreibung,
                  zitervorschlag,
                  richter)

return(meta)
```

5.2 f.bverfg.extract.content

Diese Funktion extrahiert relevante Teile des Entscheidungstextes aus einer auf einer HTML-Seite dargestellten Entscheidung auf www.bundesverfassungsgericht.de

```
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[-etstd.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 Langzitatü
berschriften

    }

    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-Ü
berschriften

    }

    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)
}

```

6 Download vorbereiten: Alle Entscheidungen des BVerfG

6.1 Zeitstempel: Linksammlung Beginn

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

```
## [1] "2021-09-19 11:22:22 CEST"
```

6.2 Funktion zeigen

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

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

6.3 Maximale Seitenzahl auslesen

```
temp <- f.linkextract("https://www.bundesverfassungsgericht.de/SiteGlobals/Forms/
  Suche/Entscheidungsuche_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] 813
```

6.4 Auszuwertende Seiten

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

6.5 [Debugging Modus] Reduzierung der Seitenzahl

```
if (mode.debug == TRUE){  
  pages <- sort(sample(pages,  
                        debug.scope))  
}
```

6.6 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
```

```
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))
```

6.7 Linkliste erstellen: Zweitversuch

6.7.1 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))
```

6.7.2 Anzeigen der Seiten die noch einmal geprüft werden.

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

```
## [1] 813
```

6.7.3 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))  
  
}
```

6.7.4 Anzeigen der Links die beim Wiederholungsversuch gesammelt wurden.

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

```
## [[1]]
## [1] "SharedDocs/Entscheidungen/DE/1983/12/rs19831215_1bvr020983.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [2] "SharedDocs/Entscheidungen/DE/1983/02/es19830216_2bve000183.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [3] "SharedDocs/Entscheidungen/DE/1977/10/qs19771016_1bvq000577.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [4] "SharedDocs/Entscheidungen/DE/1959/07/rs19590729_1bvr020558.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [5] "SharedDocs/Entscheidungen/DE/1958/01/rs19580115_1bvr040051.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [6] "SharedDocs/Entscheidungen/DE/1957/01/rs19570116_1bvr025356.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [7] "SharedDocs/Entscheidungen/DE/1952/10/rs19521010_1bvr051152.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [8] "SharedDocs/Entscheidungen/DE/1952/09/rs19520910_1bvr037952.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
## [9] "SharedDocs/Entscheidungen/DE/1951/09/qs19510909_2bvq000151.html;
      jsessionid=6E649B25AC6983903255387473BCF8EC.1_cid377"
```

6.8 Zeitstempel: Linksammlung Ende

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

```
## [1] "2021-09-19 11:54:07 CEST"
```

6.9 Dauer: Linksammlung

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

```
## Time difference of 31.73505 mins
```

6.10 Listen aus Erstversuch und Zweitversuch zusammenfügen

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

6.11 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)
```

6.12 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)
```

6.13 Dateinamen erstellen

6.13.1 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]*)",
                  "BVerfG_\\2-\\3-\\4_\\1_\\5_\\6_\\7_\\8_NA",
                  filenames)
```

Struktur von Verzögerungsrügen

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

```

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

```

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)

```

6.13.2 Formatierung von Registerzeichen anpassen

```

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

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

```

6.13.3 Formatierung von Spruchkörper-Typ anpassen

```

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

```

6.13.4 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-ZNA]+$",
                  filenames1,
                  invert = TRUE,
                  value = TRUE)

```

6.13.5 Ergebnis der ersten REGEX-Validierung

Das Ergebnis sollte ein leerer Vektor sein!

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

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

6.13.6 Skript stoppen falls erste REGEX-Validierung gescheitert

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

6.13.7 Zusätzliche Variablen einfügen

```
extravariablen <- fread("C-BVerfGE_Source_Variablen_NameBandSeite.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)  
}
```

6.13.8 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")

```

6.13.9 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\\-]+_[NA0-9]+_[NA0-9]+\\.pdf$",
                    filenames2,
                    value = TRUE,
                    invert = TRUE)

```

6.13.10 Ergebnis der zweiten REGEX-Validierung

Das Ergebnis sollte ein leerer Vektor sein!

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

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

6.13.11 Skript stoppen falls zweite REGEX-Validierung gescheitert

```

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

```

7 PDF-Download

7.1 Data Table für PDF-Download erstellen

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

7.2 Zeitstempel: PDF-Download Beginn

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

```
## [1] "2021-09-19 11:54:07 CEST"
```

7.3 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"
## [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"
```

7.4 Zeitstempel: PDF-Download Ende

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

```
## [1] "2021-09-19 15:22:17 CEST"
```

7.5 Dauer: PDF-Download

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

```
## Time difference of 3.469374 hours
```

7.6 PDF-Download: Ergebnis

7.6.1 Anzahl herunterzuladender Dateien

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

```
## [1] 8129
```

7.6.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8122
```

7.6.3 Fehlbetrag

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

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

7.6.4 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_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"
```

7.7 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"

```

7.8 PDF-Download: Gesamtergebnis

7.8.1 Anzahl herunterzuladender Dateien

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

```
## [1] 8129
```

7.8.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8122
```

7.8.3 Fehlbetrag

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

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

7.8.4 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"
```

7.8.5 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.

8 HTML-Download

8.1 Data Table für HTML-Download erstellen

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

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

8.2 Zeitstempel: HTML-Download Beginn

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

```
## [1] "2021-09-19 15:22:26 CEST"
```

8.3 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))

}
```

```
## Warning in download.file(dt.download.html$links.html[i],
## dt.download.html$names.html[i]): URL 'https://www.bundesverfassungsgericht.de/
## SharedDocs/Entscheidungen/DE/2013/06/rk20130619_1bvr066713.html': status was
## 'Server returned nothing (no headers, no data)'
```

8.4 Zeitstempel: HTML-Download Ende

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

```
## [1] "2021-09-19 18:00:48 CEST"
```

8.5 Dauer: HTML-Download

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

```
## Time difference of 2.639314 hours
```

8.6 HTML-Download: Ergebnis

8.6.1 Anzahl herunterzuladender Dateien

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

```
## [1] 8129
```

8.6.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8128
```

8.6.3 Fehlbetrag

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

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

8.6.4 Fehlende Dateien

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

```
## [1] "rk20130619_1bvr066713.html"
```

8.7 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))  
  }  
}
```

8.8 HTML-Download: Gesamtergebnis

8.8.1 Anzahl herunterzuladender Dateien

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

```
## [1] 8129
```

8.8.2 Anzahl heruntergeladener Dateien

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

```
## [1] 8129
```

8.8.3 Fehlbetrag

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

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

8.8.4 Fehlende Dateien

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

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


9 HTML verarbeiten

9.1 HTML-Dateien definieren

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

9.2 HTML-Dateien einlesen

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

9.3 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)  
}
```

9.4 Data Table mit allen Metadaten (inkl. ECLI)

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

9.5 Data Table mit vollständiger segmentierter Variante

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

9.6 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.

9.7 Stichprobe Metadaten

```
fwrite(dt.meta.html[sample(.N, 50)],  
       "QA_Stichprobe_HTML-Metadaten.csv")
```

9.8 Stichprobe Segmentierte Variante

```
fwrite(dt.segmented.full[sample(.N, 50)],  
       "QA_Stichprobe_HTML-SegmentierterVolltext.csv")
```

10 Text-Extraktion aus PDF

10.1 Vektor der zu extrahierenden Dateien erstellen

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

10.2 Anzahl zu extrahierender Dateien

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

```
## [1] 8122
```

10.3 PDF extrahieren: Funktion anzeigen

```
print(f.dopar.pdfextract)
```

```
function(x, threads = detectCores()){
```

```
begin.extract <- Sys.time()  
  
print(paste("Parallel processing using", threads, "threads. Begin at", begin.  
extract))  
  
cl <- makeForkCluster(threads)  
registerDoParallel(cl)  
  
newnames <- gsub("\\\\.pdf",  
                "\\\\.txt",  
                x)  
  
result <- foreach(i = seq_along(x),  
                  .errorhandling = 'pass') %dopar% {  
  
    ## Extract text layer from PDF  
    pdf.extracted <- pdf_text(x[i])  
  
    ## Write TXT to Disk  
    write.table(pdf.extracted,  
                newnames[i],  
                quote = FALSE,  
                row.names = FALSE,  
                col.names = FALSE)  
  
}
```

```

stopCluster(cl)

end.extract <- Sys.time()
duration.extract <- end.extract - begin.extract

print(paste0("Processed ",
             length(result),
             " files. Runtime was ",
             round(duration.extract,
                   digits = 2),
             " ",
             attributes(duration.extract)$units,
             ". Ended at ",
             end.extract, "."))

return(result)
}

```

10.4 Text Extrahieren

```

result <- f.dopar.pdfextract(files.pdf,
                             threads = fullCores)

```

11 Korpus Erstellen

11.1 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")
```

11.2 In Data Table umwandeln

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

11.3 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.

11.3.1 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-\n2",
##     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)
## }
```

11.3.2 Funktion ausführen

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

11.4 Variable “datum” als Datentyp “IDate” kennzeichnen

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

11.5 Variable “entscheidungsjahr” hinzufügen

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

11.6 Variable “eingangsjahr_iso” hinzufügen

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

11.7 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)
```

11.8 Variable “praesi” hinzufügen

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

11.8.1 Lebensdaten einlesen

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

11.8.2 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

11.8.3 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
```

11.8.4 Schleife vorbereiten

```
N <- praesi[,.N]

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

11.8.5 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(praesie$name_last[i],
                        praesi.N)
}
```

11.8.6 Vektor einfügen

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

11.9 Variable “v_praesi” hinzufügen

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

11.9.1 Personaldaten einlesen

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

11.9.2 Personaldaten anzeigen

```
kable(vpPraesi,
      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

11.9.3 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
```

11.9.4 Schleife vorbereiten

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

11.9.5 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)
}

```

11.9.6 Vektor einfügen

```

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

```

11.10 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.

11.10.1 Datensatz einlesen

```

az.source <- fread("AZ-BRD_1-0-1_DE_Registerzeichen_Datensatz.csv")

```

11.10.2 Datensatz auf relevante Daten reduzieren

```

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

```

11.10.3 Indizes bestimmen

```

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

```

11.10.4 Vektor der Verfahrensarten erstellen und einfügen

```

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

```

11.11 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)
```

11.12 Variable “doi_concept” hinzufügen

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

11.13 Variable “doi_version” hinzufügen

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

11.14 Variable “version” hinzufügen

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

11.15 Variable “lizenz” hinzufügen

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

11.16 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,
```

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

```

        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)

```

11.16.1 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)]

```

11.16.2 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](https://www.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)

```

11.16.3 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)

```

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)

```

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

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

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

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

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

11.16.5 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"
```

11.16.6 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.  
##      pdf"  
## [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"
```

11.16.7 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)
```

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

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

11.17 Variable “entscheidung_typ” hinzufügen

11.17.1 Zitervorschläge parsen

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

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

11.17.2 Kürzen

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

kurz.etyp <- c("U",
              "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)
```

11.17.3 Vektor in Datensatz einfügen

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


12 Frequenztabellen erstellen

12.1 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,
           paste0(outputdir,
                  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)
}

}

```

12.2 Ignorierte Variablen

```
print(varremove)
```

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

12.3 Liste zu prüfender Variablen

```
varlist <- names(txt.bverfg)
varlist <- grep(paste(varremove,
                      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"
```

12.4 Frequenztabellen erstellen

```
prefix <- paste0(datasetname,
  "_01_Frequenztafel_var-")
```

```
f.fast.freqtable(txt.bverfg,
  varlist = varlist,
  sumrow = TRUE,
  output.list = FALSE,
  output.kable = TRUE,
  output.csv = TRUE,
  outputdir = outputdir,
  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	8122	100	100	100
Total	8122	100	100	100

Frequency Table for Variable: spruchkoerper_typ

4 unique value(s) detected.

spruchkoerper_typ	N	exactpercent	roundedpercent	cumulpercent
B	7	0.0861857	0.09	0.09
K	6752	83.1322334	83.13	83.22

(continued)

spruchkoerper_typ	N	exactpercent	roundedpercent	cumulpercent
P	2	0.0246245	0.02	83.24
S	1361	16.7569564	16.76	100.00
Total	8122	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	0.0861857	0.09	0.09
1	4172	51.3666585	51.37	51.45
2	3943	48.5471559	48.55	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: registerzeichen

15 unique value(s) detected.

registerzeichen	N	exactpercent	roundedpercent	cumulpercent
BvB	23	0.2831815	0.28	0.28
BvC	256	3.1519330	3.15	3.44
BvE	113	1.3912829	1.39	4.83
BvF	69	0.8495444	0.85	5.68
BvG	11	0.1354346	0.14	5.81
BvH	8	0.0984979	0.10	5.91
BvK	15	0.1846836	0.18	6.09
BvL	345	4.2477222	4.25	10.34
BvM	4	0.0492490	0.05	10.39
BvN	1	0.0123122	0.01	10.40

(continued)

registerzeichen	N	exactpercent	roundedpercent	cumulpercent
BvP	3	0.0369367	0.04	10.44
BvQ	446	5.4912583	5.49	15.93
BvR	6819	83.9571534	83.96	99.89
PBvU	2	0.0246245	0.02	99.91
Vz	7	0.0861857	0.09	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: eingangsjahr_az

45 unique value(s) detected.

eingangsjahr_az	N	exactpercent	roundedpercent	cumulpercent
0	358	4.4077813	4.41	4.41
1	427	5.2573258	5.26	9.67
2	333	4.0999754	4.10	13.77
3	320	3.9399163	3.94	17.70
4	283	3.4843635	3.48	21.19
5	272	3.3489288	3.35	24.54
6	278	3.4228023	3.42	27.96
7	306	3.7675449	3.77	31.73
8	345	4.2477222	4.25	35.98
9	344	4.2354100	4.24	40.21
10	304	3.7429205	3.74	43.95
11	325	4.0014775	4.00	47.96
12	260	3.2011820	3.20	51.16
13	271	3.3366166	3.34	54.49
14	322	3.9645408	3.96	58.46
15	279	3.4351145	3.44	61.89
16	305	3.7552327	3.76	65.65

(continued)

eingangsjahr_az	N	exactpercent	roundedpercent	cumulpercent
17	328	4.0384142	4.04	69.69
18	221	2.7210047	2.72	72.41
19	307	3.7798572	3.78	76.19
20	310	3.8167939	3.82	80.00
21	101	1.2435361	1.24	81.25
51	2	0.0246245	0.02	81.27
52	2	0.0246245	0.02	81.30
56	1	0.0123122	0.01	81.31
58	1	0.0123122	0.01	81.32
77	1	0.0123122	0.01	81.33
81	1	0.0123122	0.01	81.35
83	3	0.0369367	0.04	81.38
84	3	0.0369367	0.04	81.42
85	4	0.0492490	0.05	81.47
86	7	0.0861857	0.09	81.56
87	9	0.1108101	0.11	81.67
88	18	0.2216203	0.22	81.89
89	14	0.1723713	0.17	82.06
90	36	0.4432406	0.44	82.50
91	52	0.6402364	0.64	83.14
92	63	0.7756710	0.78	83.92
93	84	1.0342280	1.03	84.95
94	143	1.7606501	1.76	86.72
95	137	1.6867767	1.69	88.40
96	179	2.2038907	2.20	90.61
97	189	2.3270131	2.33	92.93
98	253	3.1149963	3.11	96.05
99	321	3.9522285	3.95	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: kollision

5 unique value(s) detected.

kollision	N	exactpercent	roundedpercent	cumulpercent
NA	8082	99.5075105	99.51	99.51
1	1	0.0123122	0.01	99.52
a	34	0.4186161	0.42	99.94
b	4	0.0492490	0.05	99.99
c	1	0.0123122	0.01	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: band

68 unique value(s) detected.

band	N	exactpercent	roundedpercent	cumulpercent
NA	7391	90.9997538	91.00	91.00
1	2	0.0246245	0.02	91.02
7	1	0.0123122	0.01	91.04
62	1	0.0123122	0.01	91.05
65	1	0.0123122	0.01	91.06
90	1	0.0123122	0.01	91.07
91	1	0.0123122	0.01	91.09
93	1	0.0123122	0.01	91.10
96	5	0.0615612	0.06	91.16
97	18	0.2216203	0.22	91.38
98	14	0.1723713	0.17	91.55
99	26	0.3201182	0.32	91.87
100	14	0.1723713	0.17	92.05
101	15	0.1846836	0.18	92.23
102	17	0.2093081	0.21	92.44

(continued)

band	N	exactpercent	roundedpercent	cumulpercent
103	19	0.2339325	0.23	92.67
104	26	0.3201182	0.32	92.99
105	14	0.1723713	0.17	93.17
106	17	0.2093081	0.21	93.38
107	14	0.1723713	0.17	93.55
108	18	0.2216203	0.22	93.77
109	12	0.1477469	0.15	93.92
110	14	0.1723713	0.17	94.09
111	15	0.1846836	0.18	94.27
112	18	0.2216203	0.22	94.50
113	10	0.1231224	0.12	94.62
114	13	0.1600591	0.16	94.78
115	12	0.1477469	0.15	94.93
116	11	0.1354346	0.14	95.06
117	14	0.1723713	0.17	95.24
118	9	0.1108101	0.11	95.35
119	10	0.1231224	0.12	95.47
120	10	0.1231224	0.12	95.59
121	10	0.1231224	0.12	95.72
122	13	0.1600591	0.16	95.88
123	9	0.1108101	0.11	95.99
124	12	0.1477469	0.15	96.13
125	7	0.0861857	0.09	96.22
126	13	0.1600591	0.16	96.38
127	10	0.1231224	0.12	96.50
128	13	0.1600591	0.16	96.66
129	12	0.1477469	0.15	96.81
130	11	0.1354346	0.14	96.95
131	11	0.1354346	0.14	97.08

(continued)

band	N	exactpercent	roundedpercent	cumulpercent
132	13	0.1600591	0.16	97.24
133	14	0.1723713	0.17	97.41
134	14	0.1723713	0.17	97.59
135	11	0.1354346	0.14	97.72
136	11	0.1354346	0.14	97.86
137	9	0.1108101	0.11	97.97
138	12	0.1477469	0.15	98.12
139	10	0.1231224	0.12	98.24
140	12	0.1477469	0.15	98.39
141	8	0.0984979	0.10	98.49
142	13	0.1600591	0.16	98.65
143	8	0.0984979	0.10	98.74
144	4	0.0492490	0.05	98.79
145	9	0.1108101	0.11	98.90
146	7	0.0861857	0.09	98.99
147	9	0.1108101	0.11	99.10
148	10	0.1231224	0.12	99.22
149	13	0.1600591	0.16	99.38
150	8	0.0984979	0.10	99.48
151	9	0.1108101	0.11	99.59
152	12	0.1477469	0.15	99.74
153	6	0.0738734	0.07	99.82
154	7	0.0861857	0.09	99.90
155	8	0.0984979	0.10	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: entscheidungsjahr

38 unique value(s) detected.

entscheidungsjahr	N	exactpercent	roundedpercent	cumulpercent
1951	1	0.0123122	0.01	0.01
1952	2	0.0246245	0.02	0.04
1957	1	0.0123122	0.01	0.05
1958	1	0.0123122	0.01	0.06
1959	1	0.0123122	0.01	0.07
1977	1	0.0123122	0.01	0.09
1983	2	0.0246245	0.02	0.11
1991	1	0.0123122	0.01	0.12
1992	22	0.2708692	0.27	0.39
1993	22	0.2708692	0.27	0.66
1994	34	0.4186161	0.42	1.08
1995	31	0.3816794	0.38	1.47
1996	40	0.4924895	0.49	1.96
1997	74	0.9111056	0.91	2.87
1998	322	3.9645408	3.96	6.83
1999	367	4.5185915	4.52	11.35
2000	418	5.1465156	5.15	16.50
2001	442	5.4420094	5.44	21.94
2002	419	5.1588279	5.16	27.10
2003	312	3.8414184	3.84	30.94
2004	321	3.9522285	3.95	34.89
2005	266	3.2750554	3.28	38.17
2006	319	3.9276040	3.93	42.10
2007	247	3.0411229	3.04	45.14
2008	365	4.4939670	4.49	49.63
2009	363	4.4693425	4.47	54.10
2010	329	4.0507264	4.05	58.15
2011	298	3.6690470	3.67	61.82
2012	284	3.4966757	3.50	65.32
2013	208	2.5609456	2.56	67.88

(continued)

entscheidungsjahr	N	exactpercent	roundedpercent	cumulpercent
2014	229	2.8195026	2.82	70.70
2015	318	3.9152918	3.92	74.61
2016	397	4.8879586	4.89	79.50
2017	361	4.4447180	4.44	83.94
2018	276	3.3981778	3.40	87.34
2019	275	3.3858656	3.39	90.73
2020	510	6.2792416	6.28	97.01
2021	243	2.9918739	2.99	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: eingangsjahr_iso

45 unique value(s) detected.

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
1951	2	0.0246245	0.02	0.02
1952	2	0.0246245	0.02	0.05
1956	1	0.0123122	0.01	0.06
1958	1	0.0123122	0.01	0.07
1977	1	0.0123122	0.01	0.09
1981	1	0.0123122	0.01	0.10
1983	3	0.0369367	0.04	0.14
1984	3	0.0369367	0.04	0.17
1985	4	0.0492490	0.05	0.22
1986	7	0.0861857	0.09	0.31
1987	9	0.1108101	0.11	0.42
1988	18	0.2216203	0.22	0.64
1989	14	0.1723713	0.17	0.81
1990	36	0.4432406	0.44	1.26

(continued)

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
1991	52	0.6402364	0.64	1.90
1992	63	0.7756710	0.78	2.67
1993	84	1.0342280	1.03	3.71
1994	143	1.7606501	1.76	5.47
1995	137	1.6867767	1.69	7.15
1996	179	2.2038907	2.20	9.36
1997	189	2.3270131	2.33	11.68
1998	253	3.1149963	3.11	14.80
1999	321	3.9522285	3.95	18.75
2000	358	4.4077813	4.41	23.16
2001	427	5.2573258	5.26	28.42
2002	333	4.0999754	4.10	32.52
2003	320	3.9399163	3.94	36.46
2004	283	3.4843635	3.48	39.94
2005	272	3.3489288	3.35	43.29
2006	278	3.4228023	3.42	46.71
2007	306	3.7675449	3.77	50.48
2008	345	4.2477222	4.25	54.73
2009	344	4.2354100	4.24	58.96
2010	304	3.7429205	3.74	62.71
2011	325	4.0014775	4.00	66.71
2012	260	3.2011820	3.20	69.91
2013	271	3.3366166	3.34	73.25
2014	322	3.9645408	3.96	77.21
2015	279	3.4351145	3.44	80.65
2016	305	3.7552327	3.76	84.40
2017	328	4.0384142	4.04	88.44
2018	221	2.7210047	2.72	91.16
2019	307	3.7798572	3.78	94.94

(continued)

eingangsjahr_iso	N	exactpercent	roundedpercent	cumulpercent
2020	310	3.8167939	3.82	98.76
2021	101	1.2435361	1.24	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: praesi

10 unique value(s) detected.

praesi	N	exactpercent	roundedpercent	cumulpercent
Benda	3	0.0369367	0.04	0.04
Harbarth	496	6.1068702	6.11	6.14
Herzog	62	0.7633588	0.76	6.91
Höpker-Aschoff	3	0.0369367	0.04	6.94
Limbach	1845	22.7160798	22.72	29.66
Müller	1	0.0123122	0.01	29.67
Papier	2544	31.3223344	31.32	60.99
VACANCY-3	1	0.0123122	0.01	61.01
Voßkuhle	3165	38.9682344	38.97	99.98
Wintrich	2	0.0246245	0.02	100.00
Total	8122	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	6.7471066	6.75	6.75
Hassemer	1908	23.4917508	23.49	30.24
Henschel	44	0.5417385	0.54	30.78

(continued)

v__praesi	N	exactpercent	roundedpercent	cumulpercent
Katz	6	0.0738734	0.07	30.85
Kirchhof	2617	32.2211278	32.22	63.08
König	496	6.1068702	6.11	69.18
Limbach	8	0.0984979	0.10	69.28
Mahrenholz	55	0.6771731	0.68	69.96
Papier	1624	19.9950751	20.00	89.95
Seidl	177	2.1792662	2.18	92.13
Voßkuhle	636	7.8305836	7.83	99.96
Zeidler	3	0.0369367	0.04	100.00
Total	8122	100.0000000	100.00	100.00

Frequency Table for Variable: verfahrensart

15 unique value(s) detected.

verfahrensart	N	exactpercent	roundedpercent	cumulpercent
Abstrakte Normenkontrolle	69	0.8495444	0.85	0.85
Andere Streitigkeiten zwischen Bund und Ländern	8	0.0984979	0.10	0.95
Bund-Länder-Streitigkeiten	11	0.1354346	0.14	1.08
Divergenzvorlagen eines Landesverfassungsgerichts zur Auslegung des Grundgesetzes	1	0.0123122	0.01	1.10
Einstweilige Anordnungen	446	5.4912583	5.49	6.59
Feststellung der Anwendbar- keit einer Regel des Völkerge- wohnheitsrechts	4	0.0492490	0.05	6.64
Konkrete Normenkontrolle	345	4.2477222	4.25	10.88
Landesverfassungsstreitigkeiten	15	0.1846836	0.18	11.07
Organstreitverfahren	113	1.3912829	1.39	12.46
Plenarentscheidungen	2	0.0246245	0.02	12.48

(continued)

verfahrensart	N	exactpercent	roundedpercent	cumulpercent
Sonstige durch Bundesrecht zugewiesene Verfahren	3	0.0369367	0.04	12.52
Verfassungsbeschwerden; Kommunalverfassungsbeschwerden	6819	83.9571534	83.96	96.48
Verfassungswidrigkeit von Parteien	23	0.2831815	0.28	96.76
Verzögerungsbeschwerde	7	0.0861857	0.09	96.85
Wahlprüfungsverfahren	256	3.1519330	3.15	100.00
Total	8122	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	8122	100	100	100
Total	8122	100	100	100

Frequency Table for Variable: doi_version

1 unique value(s) detected.

doi_version	N	exactpercent	roundedpercent	cumulpercent
10.5281/zenodo.5514083	8122	100	100	100
Total	8122	100	100	100

Frequency Table for Variable: version

1 unique value(s) detected.

version	N	exactpercent	roundedpercent	cumulpercent
2021-09-19	8122	100	100	100
Total	8122	100	100	100

Frequency Table for Variable: lizenz

1 unique value(s) detected.

lizenz	N	exactpercent	roundedpercent	cumulpercent
Creative Commons Zero 1.0 Universal	8122	100	100	100
Total	8122	100	100	100

Frequency Table for Variable: richter

480 unique value(s) detected.

richter	N	exactpercent	roundedpercent	cumulpercent
	23	0.2831815	0.28	0.28
<us> Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	0.0123122	0.01	0.30	
Baer Britz Müller Kessal-Wulf	1	0.0123122	0.01	0.31
Baer Britz Radtke	11	0.1354346	0.14	0.44
Baer Ott Radtke	54	0.6648609	0.66	1.11
Benda Simon Hesse Katzenstein Niemeyer Ott Radtko Denschel Niedermöller	1.00123122	1.00	1.12	
Britz Ott Radtke	4	0.0492490	0.05	1.17
Broß Di Fabio Gerhardt	24	0.2954937	0.30	1.47
Broß Di Fabio Landau	117	1.4405319	1.44	2.91
Broß Di Fabio Lübbe-Wolff	11	0.1354346	0.14	3.04
Broß Di Fabio Osterloh	1	0.0123122	0.01	3.05
Broß Lübbe-Wolff Gerhardt	90	1.1081015	1.11	4.16

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Broß Lübbe-Wolff Landau	1	0.0123122	0.01	4.17
Broß Mellinghoff Gerhardt	1	0.0123122	0.01	4.19
Broß Mellinghoff Landau	2	0.0246245	0.02	4.21
Broß Osterloh Di Fabio Mellinghoff Lübbe- Wolff Gerhardt Landau	1	0.0123122	0.01	4.22
Broß Osterloh Gerhardt	1	0.0123122	0.01	4.24
Broß Osterloh Landau	2	0.0246245	0.02	4.26
Broß Osterloh Mellinghoff	100	1.2312238	1.23	5.49
Bryde Eichberger Schluckebier	37	0.4555528	0.46	5.95
Bryde Schluckebier Paulus	1	0.0123122	0.01	5.96
Bundesverfassungsrichter Lim- bach Kirchhof Sommer Jentsch Hassemer Broß Osterloh Winter ist an der Unter- schrift verhindert.	1	0.0123122	0.01	5.97
Böckenförde Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0123122	0.01	6.02
Böckenförde Kruis Sommer	2	0.0246245	0.02	6.05
Böckenförde Winter Sommer	1	0.0123122	0.01	6.06
Di Fabio Gerhardt Hermanns	17	0.2093081	0.21	6.27
Di Fabio Gerhardt Landau	1	0.0123122	0.01	6.28
Di Fabio Lübbe-Wolff Landau	2	0.0246245	0.02	6.30
Di Fabio Mellinghoff Gerhardt	1	0.0123122	0.01	6.32
Dr. Benda Dr. Böhmer Dr. Si- mon Dr. Faller Dr. Hesse Dr. Katzenstein	1	0.0123122	0.01	6.33
Dr. Dr. Höpker- Aschoff Ellinghaus Dr. Heiland Dr. Drath Dr. Stein Wessel Ritterspach Lehmann Dr. Zweigert	1	0.0123122	0.01	6.34

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Dr. Aschoff Dr. Höpker- Ellinghaus Dr. Scheffler Dr. Hei- land Dr. Scholtissek Dr. Drath Wessel Ritterspach Lehmann Dr. Zweigert	1	0.0123122	0.01	6.35
Dr. Katz Dr. Fröh- lich Wolff Dr. Roediger Dr. Klaas Henneka Dr. Friesen- hahn Dr. Rupp Dr. Gei- ger Leuser Dr. Federer	1	0.0123122	0.01	6.37
Dr. Müller Dr. Scheffler Dr. Heiland Dr. Heck Dr. Ber- ger Dr. Scholtissek Dr. Drath Dr. Stein Ritterspach	1	0.0123122	0.01	6.38
Dr. Wintrich Dr. Scheff- ler Dr. Heiland Dr. Heck Dr. Scholtissek Dr. Stein Wessel Ritterspach Lehmann	1	0.0123122	0.01	6.39
Eichberger Baer Britz	87	1.0711647	1.07	7.46
Eichberger Masing Baer	2	0.0246245	0.02	7.49
Eichberger Masing Britz	1	0.0123122	0.01	7.50
Eichberger Masing Paulus Baer Britz Ott Ott	1	0.0123122	0.01	7.51
Eichberger Paulus Britz	2	0.0246245	0.02	7.54
Eichberger Schluckebier Britz	1	0.0123122	0.01	7.55
Eichberger Schluckebier Masing	4	0.0492490	0.05	7.60
Gaier Eichberger Britz	2	0.0246245	0.02	7.62
Gaier Eichberger Kirchhof	1	0.0123122	0.01	7.63
Gaier Eichberger Paulus	1	0.0123122	0.01	7.65
Gaier Eichberger Schluckebier	2	0.0246245	0.02	7.67
Gaier Eichberger Schluckebier Masing Paulus Baer Britz	2	0.0246245	0.02	7.70
Gaier Eichberger Schluckebier Masing Paulus Henns Baer Britz	1	0.0123122	0.01	7.71
Gaier Masing Baer	13	0.1600591	0.16	7.87
Gaier Paulus Baer	1	0.0123122	0.01	7.88

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Gaier Paulus Britz	93	1.1450382	1.15	9.02
Gaier Schluckebier Baer	7	0.0861857	0.09	9.11
Gaier Schluckebier Britz	3	0.0369367	0.04	9.15
Gaier Schluckebier Kirchhof	1	0.0123122	0.01	9.16
Gaier Schluckebier Paulus	170	2.0930805	2.09	11.25
Gerhardt Gaier Eichberger Hermann	2	0.0246245	0.02	11.28
Gerhardt Hermanns Müller	62	0.7633588	0.76	12.04
Gerhardt Huber Hermanns	1	0.0123122	0.01	12.05
Graßhof Kirchhof Jentsch	1	0.0123122	0.01	12.07
Graßhof Kruis Kirchhof Winter Sommer Steiner Jentsch Hassemer	2	0.0246245	0.02	12.09
Graßhof Kruis Winter	1	0.0123122	0.01	12.10
Graßhof Winter Sommer	2	0.0246245	0.02	12.13
Grimm Haas Seidl	7	0.0861857	0.09	12.21
Grimm Haas Steiner	1	0.0123122	0.01	12.23
Grimm Hömig	2	0.0246245	0.02	12.25
Grimm Hömig Seidl	6	0.0738734	0.07	12.32
Grimm Kühling Die Richterin Haas ist an der Unterschrift gehindert	1	0.0123122	0.01	12.34
Grimm Kühling Jaeger Haas Hömig Steiner	1	0.0123122	0.01	12.35
Grimm Kühling Jaeger Haas Hömig Steiner Jentsch	1	0.0123122	0.01	12.40
Grimm Kühling Seibert Jaeger Hömig Steiner	1	0.0123122	0.01	12.41
Haas Bryde Eichberger	1	0.0123122	0.01	12.42
Haas Bryde Eichberger	14	0.1723713	0.17	12.60
Haas Hömig Bryde	26	0.3201182	0.32	12.92
Haas Hömig Hohmann- Dennhardt	1	0.0123122	0.01	12.93
Harbarth Baer Ott	65	0.8002955	0.80	13.73
Harbarth Britz Ott	4	0.0492490	0.05	13.78
Harbarth Britz Radtke	129	1.5882787	1.59	15.37

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Harbarth Christ Radtk	1	0.0123122	0.01	15.38
Harbarth Masing Paulus Baer Britz Ott Christ Radtk	10	0.123122	0.12	15.50
Harbarth Ott Radtk	1	0.0123122	0.01	15.51
Harbarth Paulus Baer Britz Ott Christ Radtk	10	0.123122	0.01	15.53
Harbarth Paulus Baer Britz Ott Christ Radtk	8	0.0984971	0.10	15.62
Hassemer Broß Di Fabio	11	0.1354346	0.14	15.76
Hassemer Broß Di Fabio Lübbe-Wolff Gerhardt	1	0.0123122	0.01	15.77
Hassemer Broß Lübbe-Wolff	1	0.0123122	0.01	15.78
Hassemer Broß Mellinghoff	3	0.0369367	0.04	15.82
Hassemer Broß Osterloh	1	0.0123122	0.01	15.83
Hassemer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	15.85
Hassemer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau	37	0.4555528	0.46	16.30
Hassemer Broß Osterloh Mellinghoff Lübbe-Wolff Gerhardt	0.0123122	0.01	16.31	
Hassemer Di Fabio Gerhardt	3	0.0369367	0.04	16.35
Hassemer Di Fabio Landau	167	2.0561438	2.06	18.41
Hassemer Di Fabio Mellinghoff	2	0.0246245	0.02	18.43
Hassemer Jaeger Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	18.44
Hassemer Jentsch Broß	1	0.0123122	0.01	18.46
Hassemer Jentsch Broß Di Fabio Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	18.47
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	1	0.0123122	0.01	18.48

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhard	1	0.0123122	0.01	18.49
Hassemer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	64	0.7879833	0.79	19.28
Hassemer Jentsch Broß Osterloh Mellinghoff Di Fabio Wolff Gerhardt	10	0.123122	0.01	19.29
Hassemer Jentsch Richter Broß ist an der Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Unterschrift verhindert.	1	0.0123122	0.01	19.31
Hassemer Jentsch Richter Broß ist an Osterloh Richter Di Fabio ist Mellinghoff Lübbe-Wolff Gerhardt der Unterschrift verhindert. an	1	0.0123122	0.01	19.32
Hassemer Osterloh Lübbe-Wolff	1	0.0123122	0.01	19.33
Hassemer Osterloh Mellinghoff	266	3.2750554	3.28	22.61
Hassemer Sommer Broß	1	0.0123122	0.01	22.62
Hassemer Sommer Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	22.63
Hassemer Sommer Jentsch	7	0.0861857	0.09	22.72
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff	40	0.0492490	0.05	22.77
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff	64	0.4186161	0.42	23.18
Hassemer Sommer Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Di Fabio	10	0.0123122	0.01	23.20
Hassemer Sommer Jentsch Osterloh Di Fabio Mellinghoff Lübbe-Wolff	10	0.0123122	0.01	23.21

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Hassemer Sommer Jentsch Richter Broß ist Osterloh Di Fabio Mellinghoff Lübbe- Wolff an der Unterschrift gehindert. Hassemer	1	0.0123122	0.01	23.22
Hassemer Sommer Lübbe- Wolff	1	0.0123122	0.01	23.23
Henschel Seidl Grimm Der Richter Söllner ist we- gen Erkrankung an der Unterschrift verhindert Hen- schel Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.25
Henschel Seidl Grimm Dieterich Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.26
Henschel Seidl Grimm Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.27
Henschel Seidl Grimm Richter Söllner ist wegen einer Auslandsreise an der Un- terschrift gehindert. Hen- schel Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.28
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.29
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.31
Henschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	16	0.0123122	0.20	23.50
Henschel Seidl Grimm Söllner Richter Dieterich ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehin- dert Kühling Seibert	1	0.0123122	0.01	23.52
Hentschel Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	23.53
Hermanns Kessal-Wulf König	1	0.0123122	0.01	23.54
Hermanns Kessal- Wulf Langenfeld	1	0.0123122	0.01	23.55
Hermanns König Maidowski	1	0.0123122	0.01	23.57
Hermanns Maidowski Langenfeld	39	0.4801773	0.48	24.05
Hermanns Müller Langenfeld	82	1.0096035	1.01	25.06
Hermanns Müller Maidowski	4	0.0492490	0.05	25.10
Herzog Henschel Böckenförde Seidl Grimm Dieterich Kühling Seibert Jaeger Haas	1	0.0123122	0.01	25.12

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Herzog Henschel Der Richter Seidl ist durch Krankheit an der Unter- schrift gehindert. Her- zog Grimm Söllner Dieterich Kühling Seibert	1	0.0123122	0.01	25.13
Herzog Henschel Seidl Grimm Dieterich Kühling Seibert	1	0.0123122	0.01	25.14
Herzog Henschel Seidl Grimm Söllner Dieterich Kühling Seibert	25	0.3076875	0.31	25.45
Herzog Henschel Seidl Grimm Söllner Dieterich Seibert	3	0.0369167	0.04	25.49
Herzog Henschel Seidl Grimm Söllner Kühling Seibert	6	0.0738334	0.07	25.56
Herzog Henschel Seidl Grimm Söllner Kühling Seibert Jaeger	5	0.0615385	0.06	25.62
Hoffmann- Riem Bryde Schluckebier	1	0.0123122	0.01	25.63
Hoffmann- Riem Eichberger Schluckebier	2	0.0246245	0.02	25.66
Hohmann- Dennhardt Bryde Kirchhof	1	0.0123122	0.01	25.67
Hohmann- Dennhardt Eichberger Kirchhof	1	0.0123122	0.01	25.68
Hohmann- Dennhardt Eichberger Schluckebier	1	0.0123122	0.01	25.70
Hohmann- Dennhardt Gaier Eichberger	2	0.0246245	0.02	25.72
Hohmann- Dennhardt Gaier Kirchhof	159	1.9576459	1.96	27.68
Hohmann- Dennhardt Gaier Masing	1	0.0123122	0.01	27.69
Hohmann- Dennhardt Gaier Paulus	57	0.7017976	0.70	28.39
Hohmann- Dennhardt Gaier Schluckebier	1	0.0123122	0.01	28.40
Hohmann- Dennhardt Hoffmann- Riem Gaier	2	0.0246245	0.02	28.43
Huber Hermanns Müller	1	0.0123122	0.01	28.44
Huber Kessal-Wulf König	267	3.2873676	3.29	31.73

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Huber Kessal- Wulf Langenfeld	1	0.0123122	0.01	31.74
Huber Kessal- Wulf Maidowski	3	0.0369367	0.04	31.78
Huber Kessal- Wulf Wallrabensein	1	0.0123122	0.01	31.79
Huber Kessal- Wulf Wallrabenstein	83	1.0219158	1.02	32.81
Huber König Langenfeld	1	0.0123122	0.01	32.82
Huber König Maidowski	1	0.0123122	0.01	32.84
Huber Langenfeld Wallrabenstein	3	0.0369367	0.04	32.87
Huber Müller Kessal-Wulf	2	0.0246245	0.02	32.90
Huber Müller Maidowski	100	1.2312238	1.23	34.13
Hömig Bryde Gaier	9	0.1108101	0.11	34.24
Hömig Haas Seidl	1	0.0123122	0.01	34.25
Hömig Hoffmann-Riem Bryde	2	0.0246245	0.02	34.28
Hömig Hoffmann-Riem Gaier	1	0.0123122	0.01	34.29
Hömig Jaeger Kühling	1	0.0123122	0.01	34.30
Hömig Kühling Seidl	1	0.0123122	0.01	34.31
Jaeger Frau Haas ist gehindert, ihre Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Lübbe-Wolff Unterschrift beizufügen. Jaeger	1	0.0123122	0.01	34.33
Jaeger Haas Bryde	1	0.0123122	0.01	34.34
Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde		0.0123122	0.01	34.35
Jaeger Hömig Bryde	235	2.8933760	2.89	37.24
Jaeger Hömig Hoffmann- Riem	4	0.0492490	0.05	37.29
Jaeger Hömig Steiner	1	0.0123122	0.01	37.31

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Jaeger Steiner Hohmann-Dennhardt	1	0.0123122	0.01	37.32
Jentsch Broß Gerhardt	13	0.1600591	0.16	37.48
Jentsch Broß Lübbe-Wolff	59	0.7264221	0.73	38.20
Jentsch Broß Mellinghoff	1	0.0123122	0.01	38.22
Jentsch Broß Osterloh	1	0.0123122	0.01	38.23
Jentsch Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt	1	0.0123122	0.01	38.24
Jentsch Hassemer Broß	2	0.0246245	0.02	38.27
Kessal-Wulf König Langenfeld	1	0.0123122	0.01	38.28
Kessal-Wulf König Maidowski	4	0.0492490	0.05	38.33
Kessal-Wulf Maidowski Wallrabenstein	1	0.0123122	0.01	38.34
Kirchhof Baer Britz	1	0.0123122	0.01	38.35
Kirchhof Bryde Schluckebier	55	0.6771731	0.68	39.03
Kirchhof Eichberger Baer	1	0.0123122	0.01	39.04
Kirchhof Eichberger Britz	137	1.6867767	1.69	40.73
Kirchhof Eichberger Masing	112	1.3789707	1.38	42.11
Kirchhof Eichberger Masing Paulus Baer Britz Ott	1	0.0123122	0.01	42.13
Kirchhof Eichberger Masing Paulus Baer Britz Ott Christ	1	0.0123122	0.01	42.23
Kirchhof Eichberger Schluckebier Masing Paulus Baer Britz	3	0.0369367	0.04	42.27
Kirchhof Eichberger Schluckebier Masing Paulus Baer Britz Ott	7	0.0861857	0.09	42.35
Kirchhof Gaier Eichberger Schluckebier Masing Paulus Baer Britz	59	0.7264221	0.73	43.20
Kirchhof Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Masing	2	0.0246245	0.02	43.23
Kirchhof Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Masing Paulus	19	0.2339325	0.23	43.46
Kirchhof Jentsch Graßhof	1	0.0123122	0.01	43.47
Kirchhof Jentsch Osterloh	3	0.0369367	0.04	43.51

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Kirchhof Masing Baer	178	2.1915784	2.19	45.70
Kirchhof Masing Gaier	1	0.0123122	0.01	45.72
Kirchhof Masing Ott	1	0.0123122	0.01	45.73
Kirchhof Masing Paulus	78	0.9603546	0.96	46.69
Kirchhof Masing Paulus Baer Britz Ott Christ	123123	0.0123122	0.01	46.70
Kirchhof Masing Schluckebier	1	0.0123122	0.01	46.71
Kirchhof Ott Christ	50	0.6156119	0.62	47.33
Kirchhof Paulus Baer	4	0.0492490	0.05	47.38
Kirchhof Paulus Britz	1	0.0123122	0.01	47.39
Kirchhof Schluckebier Baer	102	1.2558483	1.26	48.65
Kirchhof Schluckebier Masing	2	0.0246245	0.02	48.67
Kirchhof Schluckebier Ott	59	0.7264221	0.73	49.40
Klein Winter Sommer	1	0.0123122	0.01	49.41
Kruis	1	0.0123122	0.01	49.42
Kruis Winter Hassemer	2	0.0246245	0.02	49.45
Kruis Winter Jentsch	1	0.0123122	0.01	49.46
Kruis Winter Sommer	1	0.0123122	0.01	49.47
König Hermanns Müller	1	0.0123122	0.01	49.48
König Hermanns Müller Kessal- Wulf Maidowski Langenfeld Wallrabenstein	2	0.0246245	0.02	49.51
König Huber Hermanns Kessal- Wulf Langenfeld	1	0.0123122	0.01	49.52
König Huber Hermanns Kessal- Wulf Maidowski Langenfeld Wallrabenstein	2	0.0246245	0.02	49.54
König Huber Hermanns Müller Kessal- Wulf Maidowski Langenfeld	1	0.0492490	0.05	49.59
König Huber Hermanns Müller Kessal- Wulf Maidowski Langenfeld Wallrabenstein	51	0.8741689	0.87	50.47
König Huber Hermanns Müller Maidowski Langenfeld Härtel	10123122	0.0123122	0.01	50.48
König Huber Hermanns Müller Maidowski Langenfeld Wallrabenstein	369367	0.0369367	0.04	50.52

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
König Huber Müller Kessal-Wulf Maidowski Langenfeld	2	0.0246245	0.02	50.54
König Maidowski Langenfeld	2	0.0246245	0.02	50.57
König Maidowski Wallrabenstein	1	0.0123122	0.01	50.58
König Müller Maidowski	54	0.6648609	0.66	51.24
Kühling Jaeger Steiner	1	0.0123122	0.01	51.26
Kühling Hömig Hoffmann-Riem	1	0.0123122	0.01	51.27
Kühling Hömig Steiner	1	0.0123122	0.01	51.28
Kühling Jaeger Haas Hömig Steiner Hohmann Dennhardt Hoffmann-Riem	1	0.0123122	0.01	51.29
Kühling Jaeger Haas Hömig Steiner Hohmann Dennhardt Hoffmann-Riem Di Fabio	1	0.0123122	0.01	51.31
Kühling Jaeger Hömig	102	1.2558483	1.26	52.56
Kühling Jaeger Steiner	90	1.1081015	1.11	53.67
Landau Huber Hermanns Müller Kessal-Wulf König Maidowski	1	0.0123122	0.01	53.68
Landau Kessal-Wulf König	93	1.1450382	1.15	54.83
Landau Kessal-Wulf Maidowski	1	0.0123122	0.01	54.84
Landau König Maidowski	2	0.0246245	0.02	54.86
Landau Müller Kessal-Wulf	1	0.0123122	0.01	54.88
Landau Müller König	1	0.0123122	0.01	54.89
Landau Müller Maidowski	1	0.0123122	0.01	54.90
Landau Schluckebier Paulus Maidowski	2	0.0246245	0.02	54.92
Limbach Kirchhof Jentsch	1	0.0123122	0.01	54.94
Limbach Broß Di Fabio	1	0.0123122	0.01	54.95
Limbach Böckenförde Graßhof Kruis Kirchhof Winter Sommer	1	0.0123122	0.01	54.96
Limbach Böckenförde Klein Graßhof Kruis Kremer Winter Sommer	1	0.0123122	0.01	55.18
Limbach Böckenförde Klein Graßhof Kruis Winter Schubert Sommer	1	0.0123122	0.01	55.20

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Limbach Böckenförde Klein Graßhof Kruis Winter Sommer	1	0.0123122	0.01	55.21
Limbach Graßhof Kirchhof	26	0.3201182	0.32	55.53
Limbach Graßhof Kruis Kirchhof Winter Sommer Hassemer	1	0.0123122	0.01	55.54
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch	1	0.0246245	0.02	55.57
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Der Richter Hassemer ist an der Unterschrift verhindert. Limbach	1	0.0123122	0.01	55.58
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Der Richter Hassemer ist an der Unterzeichnung verhindert. Limbach	1	0.0123122	0.01	55.59
Limbach Graßhof Kruis Kirchhof Winter Sommer Jentsch Hassemer	1	0.49246245	0.49	56.08
Limbach Graßhof Sommer	1	0.0123122	0.01	56.09
Limbach Hassemer Broß	32	0.3939916	0.39	56.49
Limbach Hassemer Di Fabio	31	0.3816794	0.38	56.87
Limbach Hassemer Jentsch	2	0.0246245	0.02	56.89
Limbach Hassemer Mellinghoff	137	1.6867767	1.69	58.58
Limbach Hassemer Osterloh	1	0.0123122	0.01	58.59
Limbach Jentsch Di Fabio	52	0.6402364	0.64	59.23
Limbach Jentsch Hassemer	15	0.1846836	0.18	59.42
Limbach Jentsch Osterloh	1	0.0123122	0.01	59.43
Limbach Kirchhof Graßhof	3	0.0369367	0.04	59.47
Limbach Kirchhof Jentsch	24	0.2954937	0.30	59.76
Limbach Kirchhof Jentsch Hassemer Broß Osterloh	1	0.0246245	0.02	59.79
Limbach Kirchhof Sommer Jentsch Hassemer Broß Osterloh	1	0.0369367	0.11	59.90
Limbach Kirchhof Winter Sommer Jentsch Broß Osterloh	1	0.0123122	0.01	59.91
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß Osterloh	1	0.0123122	0.01	59.92
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß	1	0.49246245	0.05	59.97
Limbach Kirchhof Winter Sommer Jentsch Hassemer Broß Osterloh	1	0.1354246	0.14	60.11
Limbach Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0246245	0.02	60.13

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Limbach Kruis Graßhof Hassemer Kirchhof Sommer Jentsch Herr	50	0.123122	0.01	60.15
Limbach Kruis Kirchhof Winter Sommer Jentsch Herr Hassemer	101	0.0123122	0.01	60.16
Limbach Kruis Winter	26	0.3201182	0.32	60.48
Limbach Sommer Broß	2	0.0246245	0.02	60.50
Limbach Sommer Hassemer Broß Osterloh Di Fabio	101	0.0123122	0.01	60.51
Limbach Sommer Jentsch	2	0.0246245	0.02	60.54
Limbach Sommer Jentsch; Hassemer Broß Osterloh Di Fabio Mellinghoff	101	0.0123122	0.01	60.55
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio	101	0.0123122	0.01	60.56
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio	101	0.0123122	0.53	61.09
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio Mellinghoff	101	0.0123122	0.36	61.45
Limbach Sommer Jentsch Hassemer Broß Osterloh Di Fabio Mellinghoff	101	0.0123122	0.05	61.50
Limbach Sommer Jentsch Hassemer Herr	101	0.0123122	0.01	61.51
Broß ist an der Osterloh Di Fabio Unterschrift gehindert. Limbach				
Limbach Winter Hassemer	1	0.0123122	0.01	61.52
Limbach Winter Hassemer	62	0.7633588	0.76	62.29
Limbach Winter Kruis	2	0.0246245	0.02	62.31
Limbach Winter Sommer Jentsch Hassemer Broß Osterloh	101	0.0123122	0.01	62.32
Lübbe-Wolff Gerhardt Huber	1	0.0123122	0.01	62.34
Lübbe-Wolff Gerhardt Landau Huber Hermanns Müller Kessal-Wulf	13	0.1600591	0.16	62.50
Lübbe-Wolff Gerhardt Müller	1	0.0123122	0.01	62.51
Lübbe-Wolff Huber Kessal-Wulf	42	0.5171140	0.52	63.03
Lübbe-Wolff Landau Huber	14	0.1723713	0.17	63.20
Lübbe-Wolff Landau Kessal-Wulf	35	0.4309283	0.43	63.63

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Mahrenholz Böckenförde Klein Großhof Krusch Kieser Hof Winter Sommer	5	0.023122	0.06	63.69
Maidowski Langenfeld Wallrabenstein	2	0.0246245	0.02	63.72
Masing Baer Britz	1	0.0123122	0.01	63.73
Masing Britz Radtke	1	0.0123122	0.01	63.74
Masing Paulus Baer	1	0.0123122	0.01	63.75
Masing Paulus Baer Britz Ott Christ Radtke	3	0.0369367	0.04	63.79
Masing Paulus Christ	107	1.3174095	1.32	65.11
Masing Paulus Eichberger	1	0.0123122	0.01	65.12
Masing Paulus Huber Maidowski	1	0.0123122	0.01	65.13
Masing Paulus Ott	1	0.0123122	0.01	65.14
Mellinghoff Gerhardt Landau	1	0.0123122	0.01	65.16
Mellinghoff Landau Huber	1	0.0123122	0.01	65.17
Mellinghoff Lübbe-Wolff Gerhardt	32	0.3939916	0.39	65.56
Mellinghoff Lübbe-Wolff Huber	56	0.6894853	0.69	66.25
Mellinghoff Lübbe-Wolff Landau	1	0.0123122	0.01	66.26
Müller Kessal-Wulf Langenfeld	3	0.0369367	0.04	66.30
Müller Kessal-Wulf Maidowski	3	0.0369367	0.04	66.34
Müller König Langenfeld	4	0.0492490	0.05	66.39
Müller Maidowski Langenfeld	1	0.0123122	0.01	66.40
Osterloh Gerhardt Landau	2	0.0246245	0.02	66.42
Osterloh Lübbe-Wolff Gerhardt	1	0.0123122	0.01	66.44
Osterloh Mellinghoff Gerhardt	64	0.7879833	0.79	67.22
Osterloh Mellinghoff Landau	1	0.0123122	0.01	67.24
Osterloh Mellinghoff Lübbe-Wolff	2	0.0246245	0.02	67.26

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Osterloh Sommer Broß	3	0.0369367	0.04	67.30
Papier Haas Hoffmann-Riem	1	0.0123122	0.01	67.31
Papier Grimm Kühling Seibert Jaeger Haas Hömig Steiner Frau BVRin Seibert ist aus dem Amt ausgeschieden.	1	0.0123122	0.01	67.32
Papier Bryde Schluckebier	123	1.5144053	1.51	68.84
Papier Der Richter Rich- ter Grimm ist aus dem Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt Amt ausgeschie- den und deshalb an der Unterschrift gehindert. Papier	2	0.0246245	0.02	68.86
Papier Der Richter Kh- ling ist aus dem Amt geschieden und daher ge- hindert zu unterschreiben Papier Jaeger Haas Hmig Steiner Hohmann- Dennhardt Hoffmann-Riem	1	0.0123122	0.01	68.87
Papier Der Richter Kühling ist aus Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem dem Amt geschieden und daher gehindert zu unter- schreiben. Papier	1	0.0123122	0.01	68.89
Papier Der Richter Steine ist Hohmann- Dennhardt Hoffmann- Riem Bryde Gaier Eichberger Schluckebier ausgeschieden und daher der Unterschrift gehindert. Papier	1	0.0123122	0.01	68.90
Papier Die Richterin Haas ist aus dem Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Gaier Eichberger Amt ausgeschieden und des- halb an der Unterschrift gehindert. Papier	3	0.0369367	0.04	68.94

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
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	0.0123122	0.01	68.95
Papier Die Richterin Jaeger ist aus Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem dem Amt ausgeschieden und daher an der Unterschrift gehindert Papier	1	0.0123122	0.01	68.96
Papier Die Richterin Jaeger ist aus Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem dem Amt ausgeschieden und daher an der Unterschrift gehindert. Papier	1	0.0123122	0.01	68.97
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	0.0123122	0.01	68.99
Papier Die Richter Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Jaeger ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	1	0.0123122	0.01	69.00
Papier Die Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Richter Jaeger ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Papier	1	0.0123122	0.01	69.01
Papier Eichberger Masing	69	0.8495444	0.85	69.86

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Gaier Kirchhof	5	0.0615612	0.06	69.92
Papier Graßhof Grimm Kühling Fraul BVRin Seibert ist aus dem Amt ausgeschieden. Pa- pier Jaeger Haas Hömig		0.0123122	0.01	69.93
Papier Graßhof Grimm Kühling Jaeger Haas Hömig	6	0.023122	0.01	69.95
Papier Grimm Haas	1	0.0123122	0.01	69.96
Papier Grimm Hömig	171	2.1053928	2.11	72.06
Papier Grimm Khling Jaeger Haas Hömig Steiner Hohmann- Dennhardt	11	0.0423122	0.01	72.08
Papier Grimm Kühling Jaeger Haas Hömig Steiner	14	0.177469	0.15	72.22
Papier Grimm Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt	15	0.166591	0.16	72.38
Papier Grimm Kühling Jaeger Hömig Steiner Hohmann- Dennhardt	9	0.146245	0.02	72.41
Papier Grimm Kühling Seibert Jaeger Haas Hömig Steiner	7	0.086157	0.09	72.49
Papier Grimm Steiner	2	0.0246245	0.02	72.52
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	0.0123122	0.01	72.53
Papier Haas Hoffmann-Riem	55	0.6771731	0.68	73.21
Papier Haas Hohmann- Dennhardt	75	0.9234179	0.92	74.13
Papier Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Gaier	30	0.3693672	0.37	74.50
Papier Haas Steiner	26	0.3201182	0.32	74.82
Papier Haas Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde Gaier Eichberger	4	0.0492490	0.05	74.87

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Hassemer Sommer Jaeger Haas Hömig Steiner Hohmann Dennhardt Hoffmann-Riem Di Fabio Bryde Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	74.88
Papier Hoffmann-Riem Eichberger	11	0.1354346	0.14	75.02
Papier Hoffmann-Riem Gaier	12	0.1477469	0.15	75.17
Papier Hoffmann-Riem Kirchhof	1	0.0123122	0.01	75.18
Papier Hohmann-Dennhardt Bryde Gaier Eichberger Schluckebier Kirchhof Masing	27	0.3324304	0.33	75.51
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	0.0123122	0.01	75.52
Papier Hohmann-Dennhardt Gaier	1	0.0123122	0.01	75.54
Papier Hohmann-Dennhardt Hoffmann-Riem	142	1.7483378	1.75	77.28
Papier Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier	1	0.0123122	0.01	77.30
Papier Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier Kirchhof	8	0.0984979	0.10	77.39
Papier Hömig Bryde	1	0.0123122	0.01	77.41
Papier Hömig Hoffmann-Riem	2	0.0246245	0.02	77.43
Papier Hömig Hohmann-Dennhardt	2	0.0246245	0.02	77.46
Papier Jaeger Haas Hömig Steiner Hohmann-Dennhardt Bryde	2	0.0246245	0.02	77.48

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem	2	0.0246245	0.02	77.51
Papier Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde	168	0.8372322	0.84	78.34
Papier Jaeger Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem	1	0.0123122	0.01	78.36
Papier Jaeger Hömig Steiner Hohmann- Dennhardt Hoffmann- Riem Bryde	1	0.0123122	0.01	78.37
Papier Kühling Jaeger Haas Hömig Hohmann- Dennhardt Hoffmann- Riem Osterloh	1	0.0123122	0.01	78.38
Papier Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt	1	0.0123122	0.01	78.39
Papier Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt	1	0.0123122	0.01	78.40
Papier Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt	1	0.0123122	0.01	78.42
Papier Kühling Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem	18	0.0816203	0.22	78.64
Papier Kühling Jaeger Hömig Steiner Hohmann- Dennhardt	1	0.0123122	0.01	78.65
Papier Kühling Jaeger Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem	1	0.0123122	0.01	78.66
Papier Kühling Steiner	2	0.0246245	0.02	78.69
Papier Richter Kühling ist aus dem Jaeger Haas Hömig Steiner Hohmann- Dennhardt Hoffmann-Riem Amt geschieden und daher gehindert zu unterschreiben. Papier	1	0.0123122	0.01	78.70
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	0.0123122	0.01	78.71

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Papier Schluckebier Kirchhof	1	0.0123122	0.01	78.72
Papier Steiner Gaier	126	1.5513420	1.55	80.28
Papier Steiner Hoffmann-Riem	219	2.6963802	2.70	82.97
Papier Steiner Hohmann-Dennhardt	112	1.3789707	1.38	84.35
Papier Steiner Hohmann-Dennhardt Hoffmann-Riem Bryde Gaier Eichberger Schluckebier	16	0.1969958	0.20	84.55
Paulus Christ Härtel	81	0.9972913	1.00	85.55
Paulus Christ Radtke	1	0.0123122	0.01	85.56
Paulus Huber Baer Müller	1	0.0123122	0.01	85.57
Paulus Ott Christ	4	0.0492490	0.05	85.62
Präsidentin Limbach ist aus Sommer Jentsch Richter Broß ist an der Hassemer Osterloh Di Fabio Mellinghoff dem Amt ausgeschieden. Unterschrift gehindert. Hassemer	1	0.0123122	0.01	85.63
Schluckebier Baer Britz	1	0.0123122	0.01	85.64
Schluckebier Masing Baer	1	0.0123122	0.01	85.66
Schluckebier Masing Paulus	1	0.0123122	0.01	85.67
Schluckebier Paulus Baer	1	0.0123122	0.01	85.68
Schluckebier Paulus Ott	6	0.0738734	0.07	85.75
Seidl Graßhof Grimm Kühling Jaeger Haas Hömig	1	0.0123122	0.01	85.77
Seidl Graßhof Grimm Kühling Seibert Jaeger Haas Hömig	1	0.0123122	0.01	85.78
Seidl Grimm Haas	15	0.1846836	0.18	85.96
Seidl Grimm Hömig	6	0.0738734	0.07	86.04
Seidl Grimm Kühling Die Richter in Seibert ist an der Unterschrift gehindert. Seidl Jaeger Haas Hömig Steiner	2	0.0246245	0.02	86.06

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Seidl Grimm Kühling Die Richterin Seibert ist we- gen Erkrankung an der Unterschrift gehindert. Seidl Jaeger Haas Hömig Steiner	1	0.0123122	0.01	86.07
Seidl Grimm Kühling Jaeger Haas Hömig Steiner	2	0.0246245	0.02	86.22
Seidl Grimm Kühling Seibert Jaeger Haas Hömig Steiner	2	0.0246245	0.02	86.49
Seidl Grimm Richter Söllner ist wegen einer Auslandsreise an der Unterschrift verhindert. Seidl Kühling Seibert Jaeger Haas	1	0.0123122	0.01	86.51
Seidl Grimm Seibert	1	0.0123122	0.01	86.52
Seidl Grimm Söllner Dieterich Kühling Seibert	1	0.0123122	0.01	86.53
Seidl Grimm Söllner Kühling Seibert Jaeger	1	0.0123122	0.01	86.54
Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	86.56
Seidl Haas Hömig	6	0.0738734	0.07	86.63
Seidl Hömig Steiner	2	0.0246245	0.02	86.65
Sommer Broß Di Fabio	2	0.0246245	0.02	86.68
Sommer Broß Mellinghoff	68	0.8372322	0.84	87.52
Sommer Broß Osterloh	80	0.9849791	0.98	88.50
Sommer Di Fabio Lübbe-Wolff	49	0.6032997	0.60	89.10
Sommer Hassemer Broß	2	0.0246245	0.02	89.13
Sommer Hassemer Osterloh	1	0.0123122	0.01	89.14
Sommer Jentsch Broß	3	0.0369367	0.04	89.18
Sommer Jentsch Hassemer	23	0.2831815	0.28	89.46
Sommer Jentsch Osterloh	1	0.0123122	0.01	89.47
Sommer Osterloh Di Fabio	36	0.4432406	0.44	89.92
Sommer Osterloh Mellinghoff	1	0.0123122	0.01	89.93
Steiner Gaier Eichberger	1	0.0123122	0.01	89.94
Steiner Gaier Schluckebier	2	0.0246245	0.02	89.97
Steiner Hoffmann-Riem Gaier	1	0.0123122	0.01	89.98

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Steiner Hohmann-Dennhardt Bryde	3	0.0369367	0.04	90.01
Steiner Hohmann-Dennhardt Hoffmann-Riem	8	0.0984979	0.10	90.11
Steiner Jaeger Kühling	15	0.1846836	0.18	90.30
Vizepräsident Henschel ist wegen Urlaubs an der Unterschrift gehindert. Seidl Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	90.31
Vizepräsident Henschel ist wegen Urlaubs an der Unterschrift verhindert. Seidl Seidl Grimm Söllner Kühling Seibert Jaeger Haas	1	0.0123122	0.01	90.32
Vizepräsident Henschel und Richter Söllner sind aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Seidl Seidl Grimm Kühling Seibert Jaeger Haas	1	0.0123122	0.01	90.33
Vizepräsident Mahrenholz ist ausgeschieden. Böckenförde Böckenförde Klein Graßhof Kruis Kirchhof Winter Sommer	1	0.0123122	0.01	90.35
Vizepräsident Seidl Grimm Hömig ist wegen Ausscheidens aus dem Amt an der Unterschrift verhindert. Grimm	1	0.0123122	0.01	90.36
Vizepräsident Seidl Grimm Kühling Seibert Jaeger Haas Hömig Steiner ist aus dem Amt ausgeschieden und deshalb an der Unterschrift gehindert. Grimm	2	0.0246245	0.02	90.38
Vosskuhle Mellinghoff Lübbe-Wolff	1	0.0123122	0.01	90.40
Voßkuhle Broß Osterloh Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau	50	0.6156119	0.62	91.01

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Di Fabio Der Richter Mellinghoff Lübbe-Wolff Gerhardt Landau Huber Hermanns ist aus dem Amt aus- geschieden und deshalb an der Unterschrift gehindert. Voßkuhle	1	0.0123122	0.01	91.02
Voßkuhle Di Fabio Lübbe-Wolff Gerhardt Landau Huber Hermanns	6	0.0738734	0.07	91.10
Voßkuhle Di Fabio Mellinghoff	1	0.0123122	0.01	91.11
Voßkuhle Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau	1	0.0123122	0.01	91.12
Voßkuhle Di Fabio Mellinghoff Lübbe-Wolff Gerhardt Landau Huber Hermanns	23	0.2831815	0.28	91.41
Voßkuhle Di Fabio Mellinghoff Lübbe-Wolff Landau Huber Hermanns	1	0.0123122	0.01	91.42
Voßkuhle Die Richterin Lübbe- Wolff ist aus dem Amt Der Richter Gerhard ist aus dem Amt Landau Huber Herrmanns Müller Kessal-Wulf ausgeschieden und deshalb an der Unter- schrift gehindert Voßkuhle	1	0.0123122	0.01	91.43
Voßkuhle Gerhardt Hermanns	1	0.0123122	0.01	91.44
Voßkuhle Gerhardt Huber	25	0.3078060	0.31	91.75
Voßkuhle Gerhardt Landau	37	0.4555528	0.46	92.21
Voßkuhle Gerhardt Landau Huber Hermanns Müller Kessal-Wulf König	1	0.0123122	0.01	92.22
Voßkuhle Hermanns Müller	1	0.0123122	0.01	92.23
Voßkuhle Huber Hermanns Kessal-Wulf König Maidowski	1	0.0123122	0.01	92.24
Voßkuhle Huber Hermanns Kessal-Wulf König Maidowski Langenfeld	11	0.1354346	0.14	92.38

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Huber Hermanns Müller Kessal- Wulf König	1	0.0123122	0.01	92.39
Voßkuhle Huber Hermanns Müller Kessal- Wulf König Maidowski	1	0.1354346	0.14	92.53
Voßkuhle Huber Hermanns Müller Kessal- Wulf König Maidowski Langenfeld	130	1.6005910	1.60	94.13
Voßkuhle Huber Müller	1	0.0123122	0.01	94.14
Voßkuhle Kessal- Wulf Langenfeld	1	0.0123122	0.01	94.15
Voßkuhle Kessal- Wulf Maidowski	157	1.9330214	1.93	96.08
Voßkuhle Kirchhof Lübbe- Wolff Gerhardt Gaier Eichberger Schluckebier Masing Paulus Huber Hermanns Baer Britz Müller Kessal- Wulf	1	0.0123122	0.01	96.10
Voßkuhle König Langenfeld	1	0.0123122	0.01	96.11
Voßkuhle König Maidowski	2	0.0246245	0.02	96.13
Voßkuhle Landau Hermanns	57	0.7017976	0.70	96.84
Voßkuhle Landau Hermanns Müller Kessal- Wulf König Maidowski	1	0.0123122	0.01	96.85
Voßkuhle Landau Huber Hermanns Kessal- Wulf König Maidowski	15	0.0615612	0.06	96.91
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf	1094	0.0246245	0.02	96.93
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf König	1049	0.0492490	0.05	96.98
Voßkuhle Landau Huber Hermanns Müller Kessal- Wulf König Maidowski	67	0.8249200	0.82	97.81
Voßkuhle Landau Huber Müller Kessal- Wulf Maidowski	2	0.0246245	0.02	97.83
Voßkuhle Landau Kessal- Wulf	1	0.0123122	0.01	97.85
Voßkuhle Lübbe- Wolff Gerhardt	1	0.0123122	0.01	97.86

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
Voßkuhle Lübbe- Wolff Gerhardt Huber Hermanns Kessal- Wulf	1	0.0123122	0.01	97.87
Voßkuhle Lübbe- Wolff Gerhardt Huber Hermanns Müller Kessal- Wulf	1	0.0123122	0.01	97.88
Voßkuhle Lübbe- Wolff Gerhardt Landau Hermanns Müller Kessal- Wulf	1	0.0123122	0.01	97.89
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns	4	0.0492490	0.05	97.94
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Kessal- Wulf	1	0.0123122	0.01	97.96
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Müller Kessal- Wulf	47	0.5786752	0.58	98.53
Voßkuhle Lübbe- Wolff Gerhardt Landau Huber Hermanns Müller Kessal- Wulf	5	0.0615612	0.06	98.60
Voßkuhle Masing Huber Hermanns Kessal- Wulf König Maidowski Langenfeld	2	0.0246245	0.02	98.62
Voßkuhle Mellinghoff Gerhardt	3	0.0369367	0.04	98.66
Voßkuhle Mellinghoff Lübbe- Wolff	70	0.8618567	0.86	99.52
Voßkuhle Osterloh Gerhardt	2	0.0246245	0.02	99.54
Voßkuhle Osterloh Mellinghoff	29	0.3570549	0.36	99.90
Winter Jentsch Broß	1	0.0123122	0.01	99.91
Winter Jentsch Osterloh	1	0.0123122	0.01	99.93
Winter Sommer Hassemer	2	0.0246245	0.02	99.95
Winter Sommer Jentsch	1	0.0123122	0.01	99.96
Zeidler Rinck Wand Dr. Rottmann Dr. Dr. h. c. Niebler Steinberger Träger Mahrenholz H.- J. Rinck	1	0.0123122	0.01	99.98
gerichts Bryde Eichberger Schluckebier	1	0.0123122	0.01	99.99

(continued)

richter	N	exactpercent	roundedpercent	cumulpercent
unter Mitwirkung Voßkuhle Landau Huber Müller Kassal- Wulf Maidowski	8122	100.0000000	99.01	100.00
Total	8122	100.0000000	99.00	100.00

Frequency Table for Variable: entscheidung_typ

2 unique value(s) detected.

entscheidung_typ	N	exactpercent	roundedpercent	cumulpercent
B	7922	97.537552	97.54	97.54
U	200	2.462448	2.46	100.00
Total	8122	100.000000	100.00	100.00

13 Frequenztabellen visualisieren

13.1 Präfix erstellen

```
prefix <- paste0("ANALYSE/",  
                 datasetname,  
                 "_01_Frequenztabelle_var-")
```

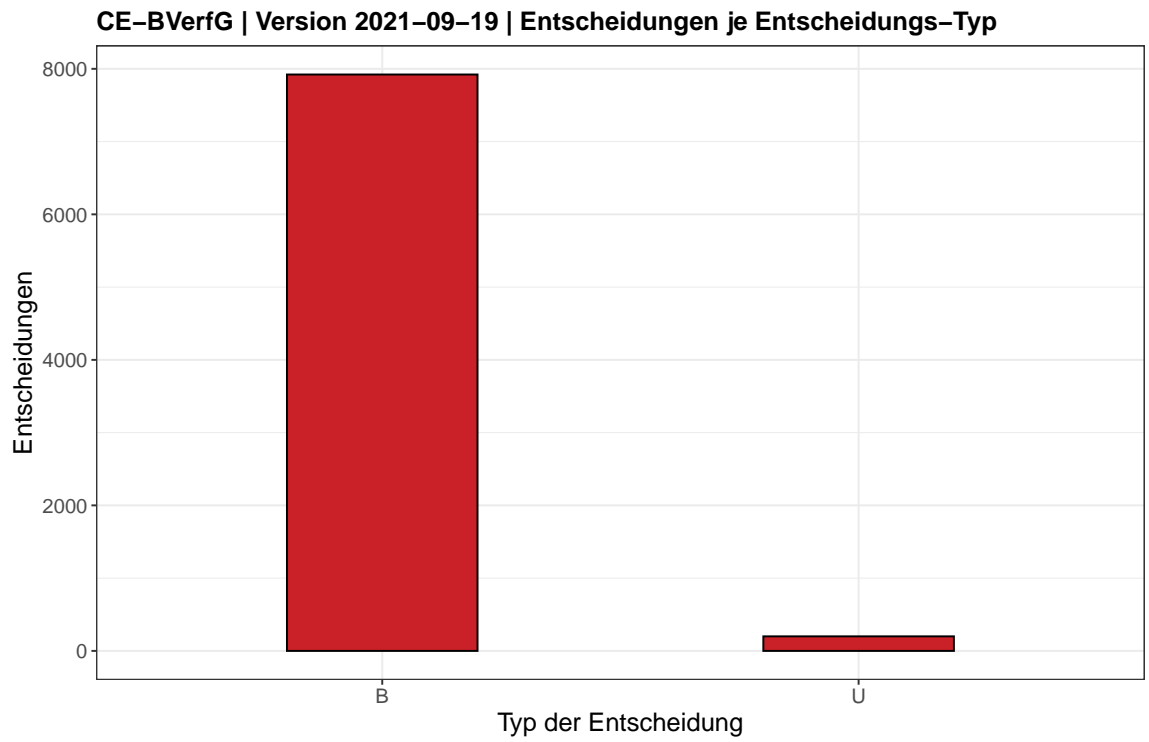
13.2 Tabellen einlesen

```
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"))
```

13.3 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Entscheidungs-Typ"),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```



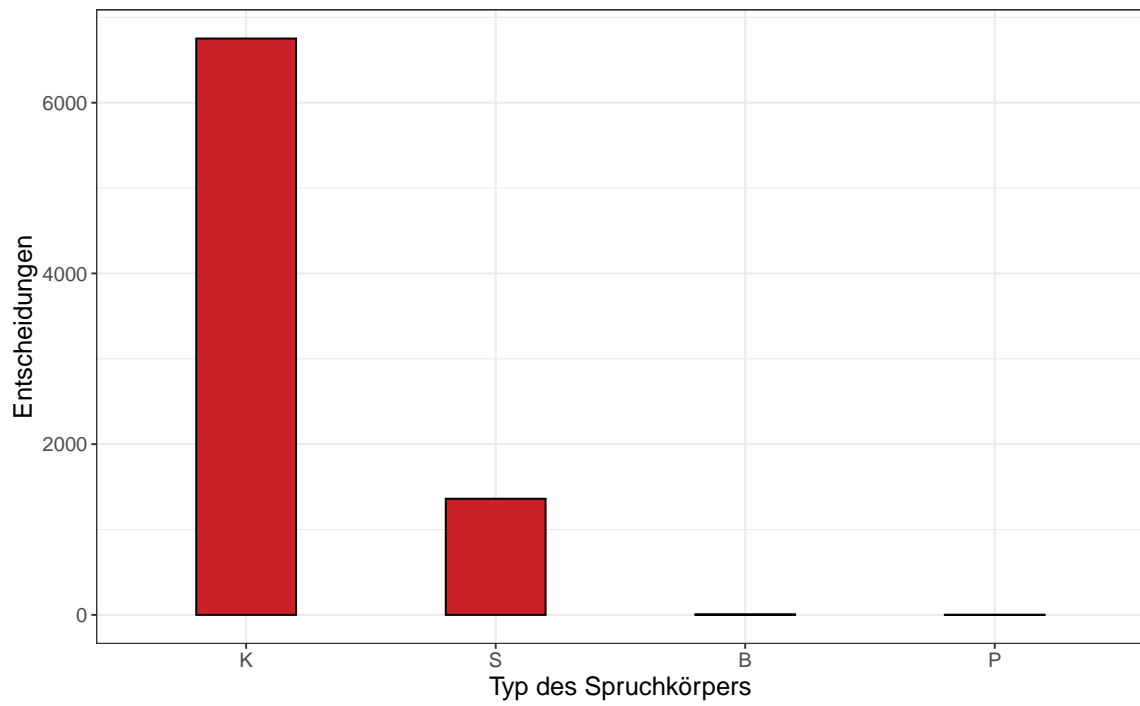
DOI: 10.5281/zenodo.5514083 | Fobbe

13.4 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Spruchkörper-Typ"),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```

CE-BVerfG | Version 2021-09-19 | Entscheidungen je Spruchkörper-Typ

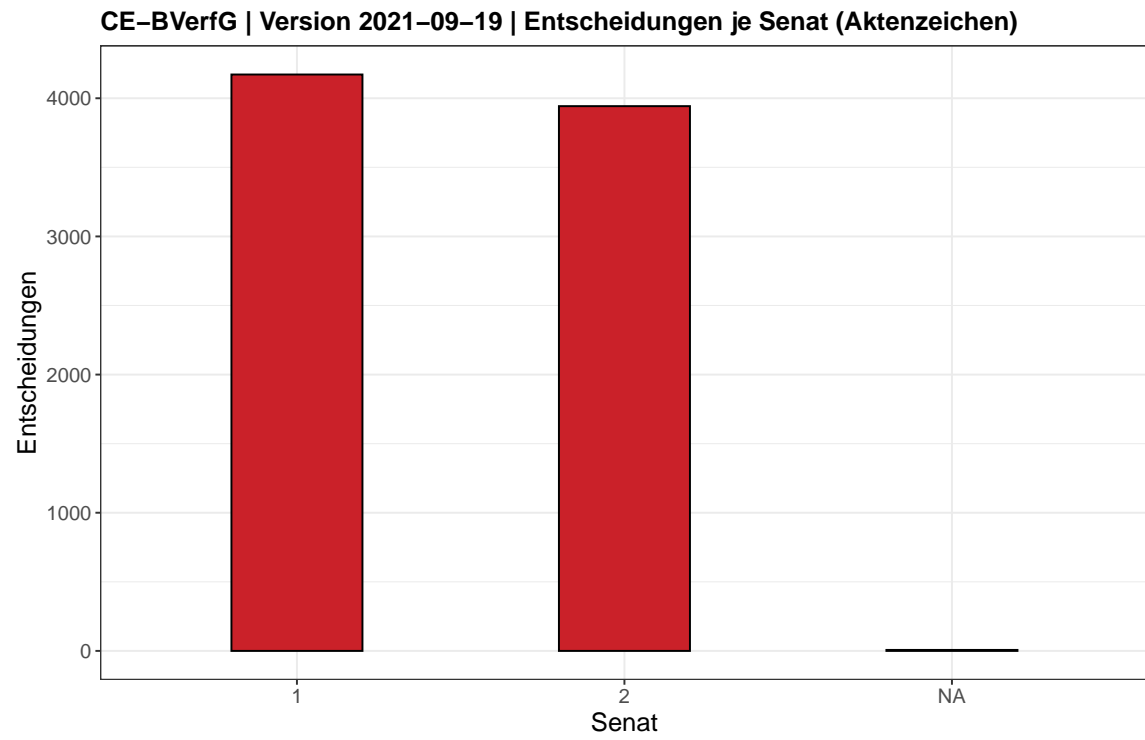


DOI: 10.5281/zenodo.5514083 | Fobbe

13.5 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Senat (Aktenzeichen)" ),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```

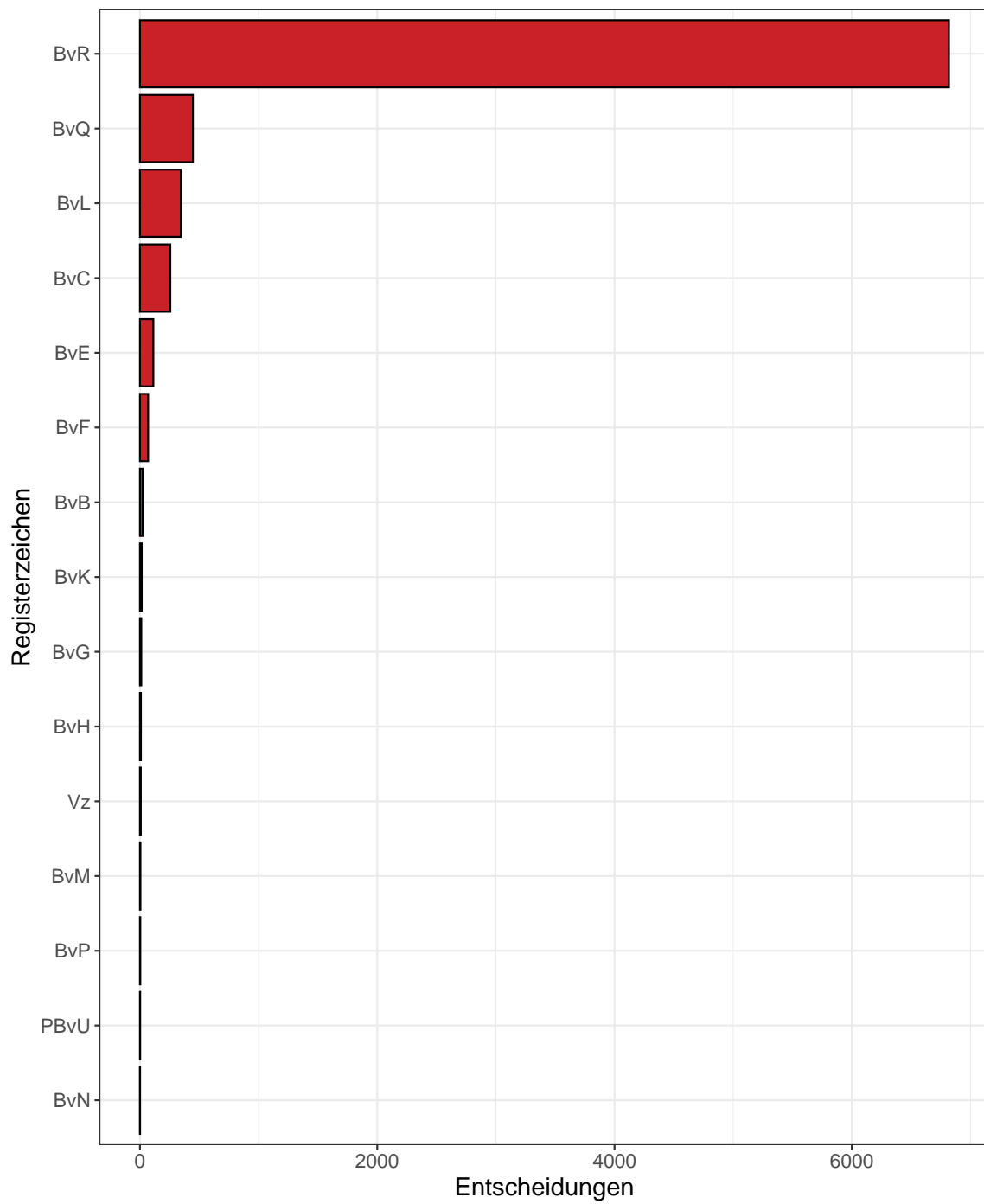
DOI: 10.5281/zenodo.5514083 | Fobbe

13.6 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Registerzeichen"),  
    caption = paste("DOI:",  
                   doi.version,  
                   "| Fobbe"),  
    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 2021-09-19 | Entscheidungen je Registerzeichen



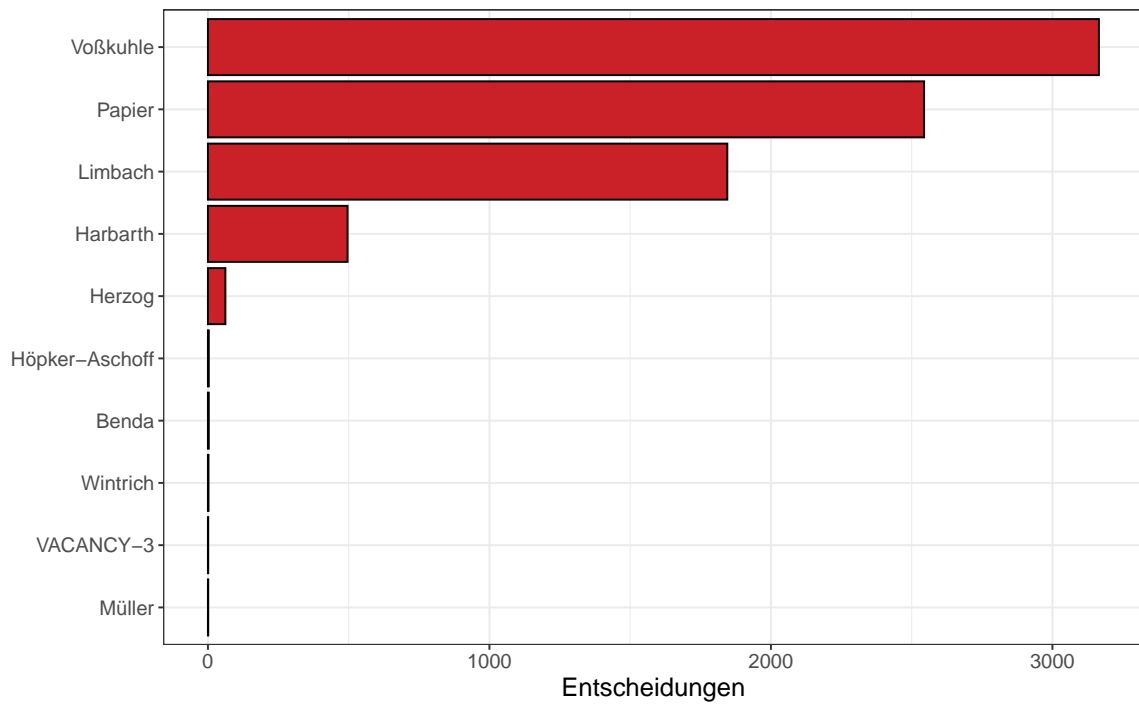
DOI: 10.5281/zenodo.5514083 | Fobbe

13.7 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Präsident:in"),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```

CE-BVerfG | Version 2021-09-19 | Entscheidungen je Präsident:in



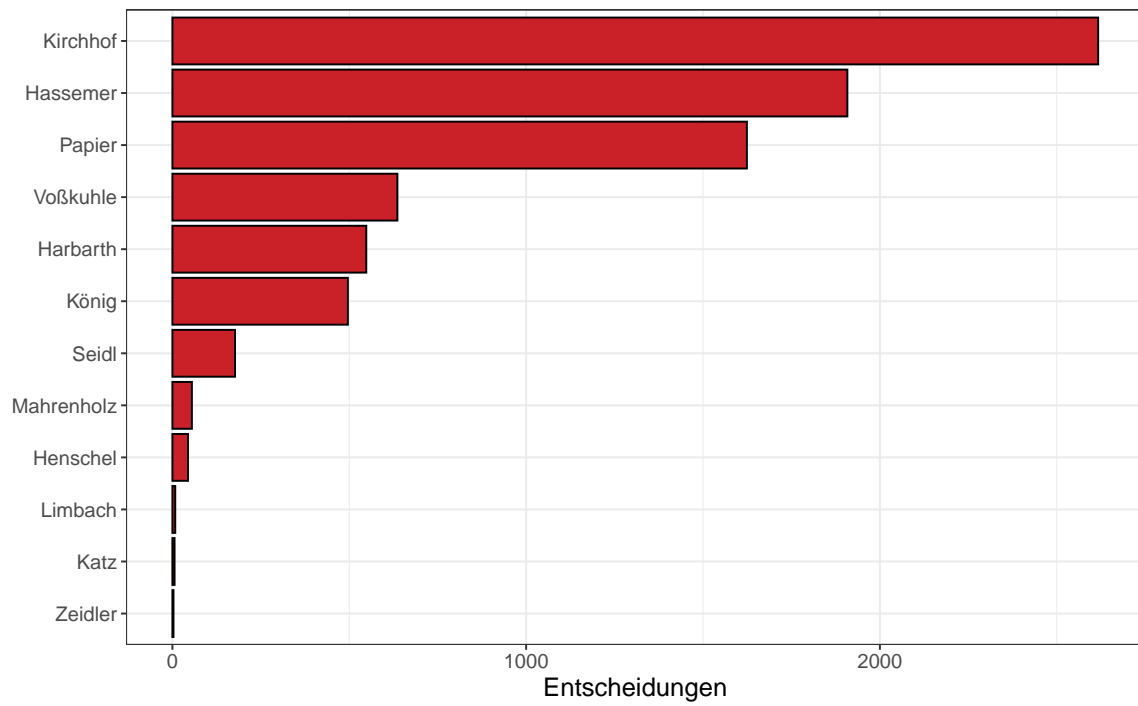
DOI: 10.5281/zenodo.5514083 | Fobbe

13.8 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Vize-Präsident:in"),  
    caption = paste("DOI:",  
                   doi.version,  
                   "| Fobbe"),  
    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)  
  )
```

CE-BVerfG | Version 2021-09-19 | Entscheidungen je Vize-Präsident:in



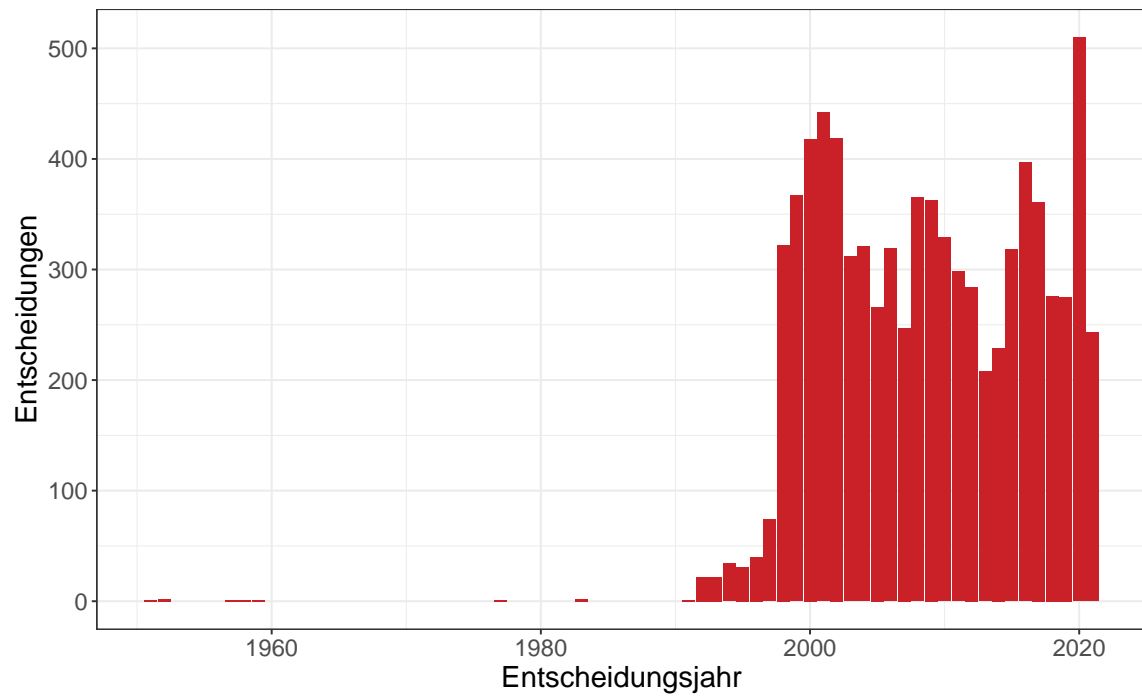
DOI: 10.5281/zenodo.5514083 | Fobbe

13.9 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(datasetsname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Entscheidungsjahr"),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```


CE-BVerfG | Version 2021-09-19 | Entscheidungen je Entscheidungsjahr



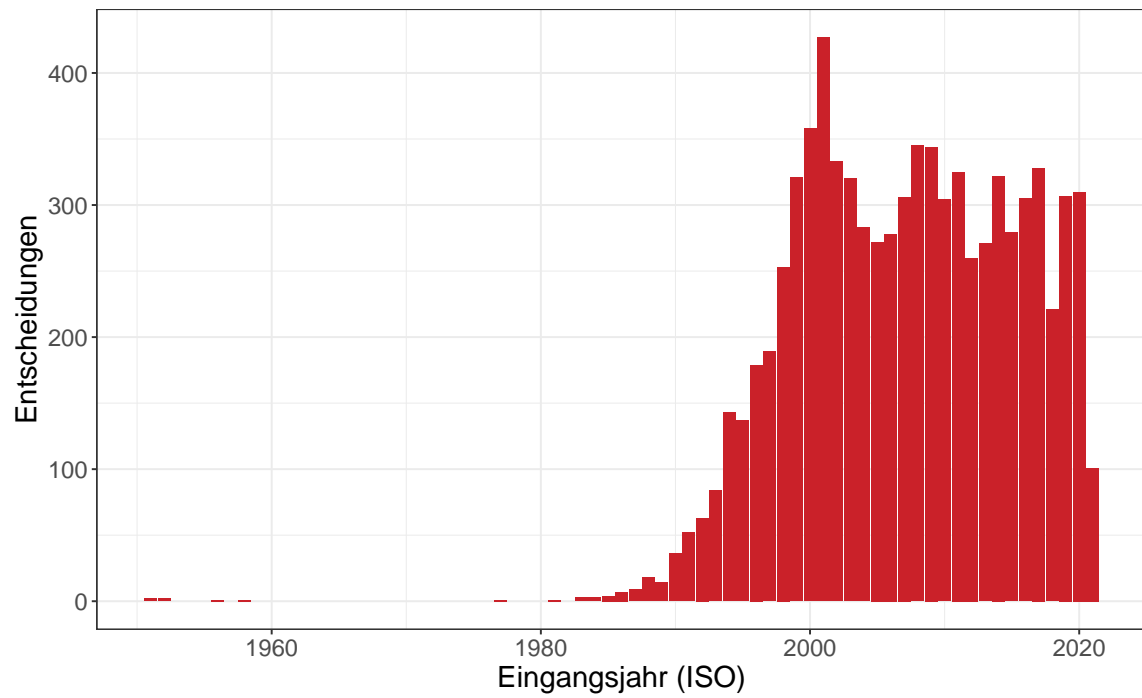
DOI: 10.5281/zenodo.5514083 | Fobbe

13.10 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(datasetname,  
                  "| Version",  
                  datestamp,  
                  "| Entscheidungen je Eingangsjahr (ISO)" ),  
    caption = paste("DOI:",  
                    doi.version,  
                    "| Fobbe"),  
    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)  
  )
```

CE-BVerfG | Version 2021-09-19 | Entscheidungen je Eingangsjahr (ISO)



DOI: 10.5281/zenodo.5514083 | Fobbe

14 Korpus-Analytik

14.1 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.

14.1.1 Funktion anzeigen

```
print(f.summarize.iterator)
```

```
## function(dt,
##                                     threads = detectCores(),
##                                     chunksize = 1){
##
##
##   begin.dopar <- Sys.time()
##   dt <- dt[,.(doc_id, text)]
##   nchars <- dt[, lapply(.(text), nchar)]
##   print(paste0("Parallel processing using ",
##               threads,
##               " threads. Begin at ",
##               begin.dopar,
##               ". Processing ",
##               dt[,.N],
##               " documents with a total length of ",
##               sum(nchars),
##               " characters."))
##
##   ord <- order(-nchars)
##   dt <- dt[ord]
##
##   cl <- makeForkCluster(threads)
##   registerDoParallel(cl)
##
##   itx <- iter(dt["nchars" > 0],
##              by = "row",
##              chunksize = chunksize)
##
##   result.list <- foreach(i = itx,
##                          .errorhandling = 'pass') %dopar% {
##
##               corpus <- corpus(i)
##
##               tokens <- tokens(corpus,
##                                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
##                                     )
##
##                                     ntokens <- unname(ntoken(tokens))
##                                     ntypes <- unname(ntype(tokens))
##                                     nsentences <- unname(nsentence(corpus))
##
##                                     temp <- data.table(ntokens,
##                                                         ntypes,
##                                                         nsentences)
##
##                                     return(temp)
##                                     }
##
## stopCluster(cl)
##
##
## end.dopar <- Sys.time()
## duration.dopar <- end.dopar - begin.dopar
##
## result.dt <- rbindlist(result.list)
##
## 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)
##
## }
```

14.1.2 Berechnung durchführen

```
summary.corpus <- f.summarize.iterator(txt.bverfg,
                                       threads = fullCores,
                                       chunksize = 1)
```

```
## [1] "Parallel processing using 16 threads. Begin at 2021-09-19 18:06:52.
      Processing 8122 documents with a total length of 158177907 characters."
## [1] "Runtime was 32.15 secs. Ended at 2021-09-19 18:07:24."
```

14.2 Variablen-Namen anpassen

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

14.3 Kennwerte dem Korpus hinzufügen

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

14.4 Variante mit Metadaten erstellen

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

14.5 Linguistische Kennwerte

14.5.1 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"))
```

14.5.2 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	158,177,907	1,255	4,967	11,237.5	19,475.2410	21,911.0	781,226
tokens	24,420,682	161	754	1,725.0	3,006.7326	3,390.5	115,540
typen	279,112	91	308	593.0	776.9042	981.0	13,491
saetze	1,392,240	6	46	105.0	171.4159	199.0	5,282

14.5.3 Zusammenfassungen speichern

```
fwrite(dt.stats.ling,  
       paste0(outputdir,  
              datasetname,  
              "_00_KorpusStatistik_ZusammenfassungLinguistisch.csv"),  
       na = "NA")
```


14.6 Quantitative Variablen

14.6.1 Entscheidungsdatum

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

```
##           Min.         1st Qu.         Median         Mean         3rd Qu.         Max.
## "1951-09-09" "2002-08-08" "2009-02-17" "2009-03-20" "2016-02-16" "2021-09-01"
```

14.6.2 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"))
```

14.6.3 Zusammenfassungen anzeigen

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

Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	38	1951	2002	2009	2008.7272	2016.00	2021
eingangsjahr_iso	45	1951	2001	2007	2007.1548	2014.00	2021
band	67	1	106	120	121.0999	136.00	155
eingangsnummer	2712	1	290	1054	1146.8772	1867.75	3634

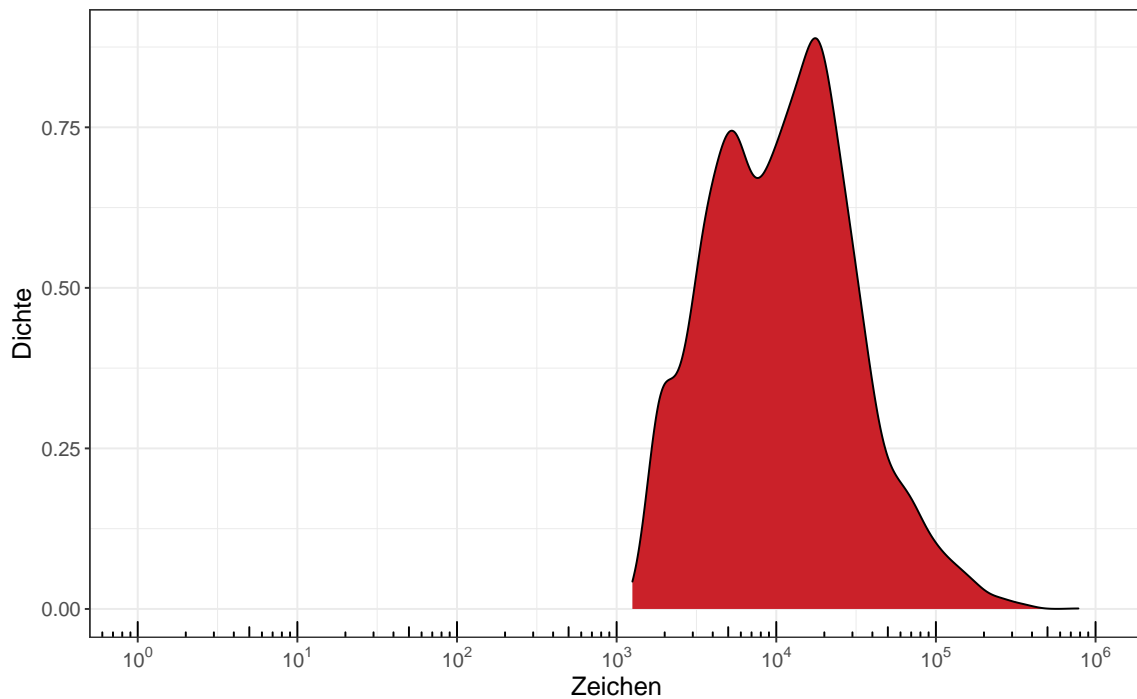
14.6.4 Zusammenfassungen speichern

```
fwrite(dt.stats.docvars,  
       paste0(outputdir,  
              datasetname,  
              "_00_KorpusStatistik_ZusammenfassungDocvarsQuantitativ.csv"),  
       na = "NA")
```

14.7 Verteilungen linguistischer Kennwerte

14.7.1 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(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Zeichen je Dokument"),
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
```

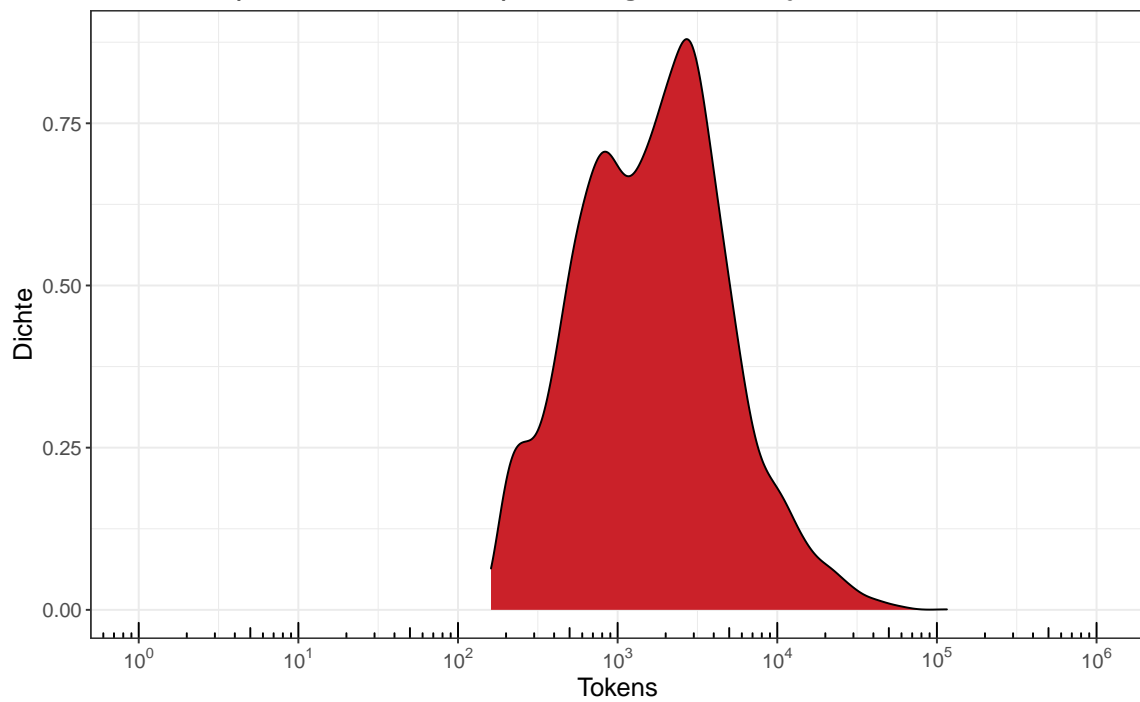


DOI: 10.5281/zenodo.5514083 | Fobbe

14.7.2 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(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Tokens je Dokument"),
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
```

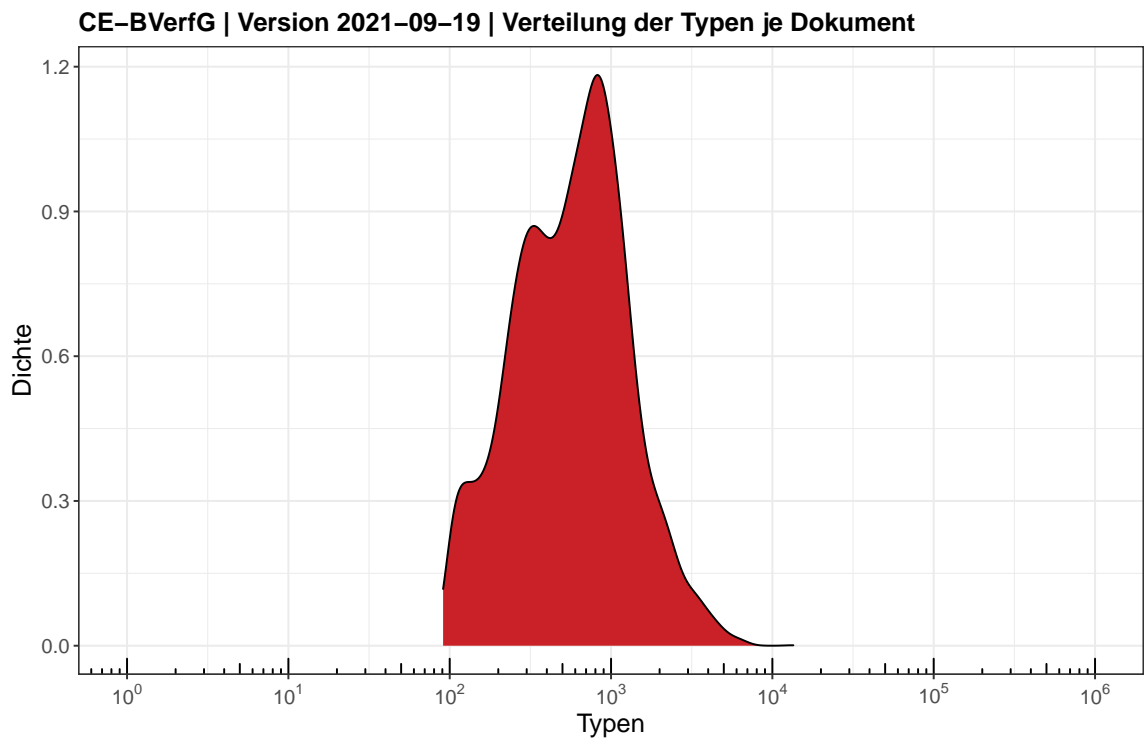
CE-BVerfG | Version 2021-09-19 | Verteilung der Tokens je Dokument



DOI: 10.5281/zenodo.5514083 | Fobbe

14.7.3 Diagramm: Verteilung Typen

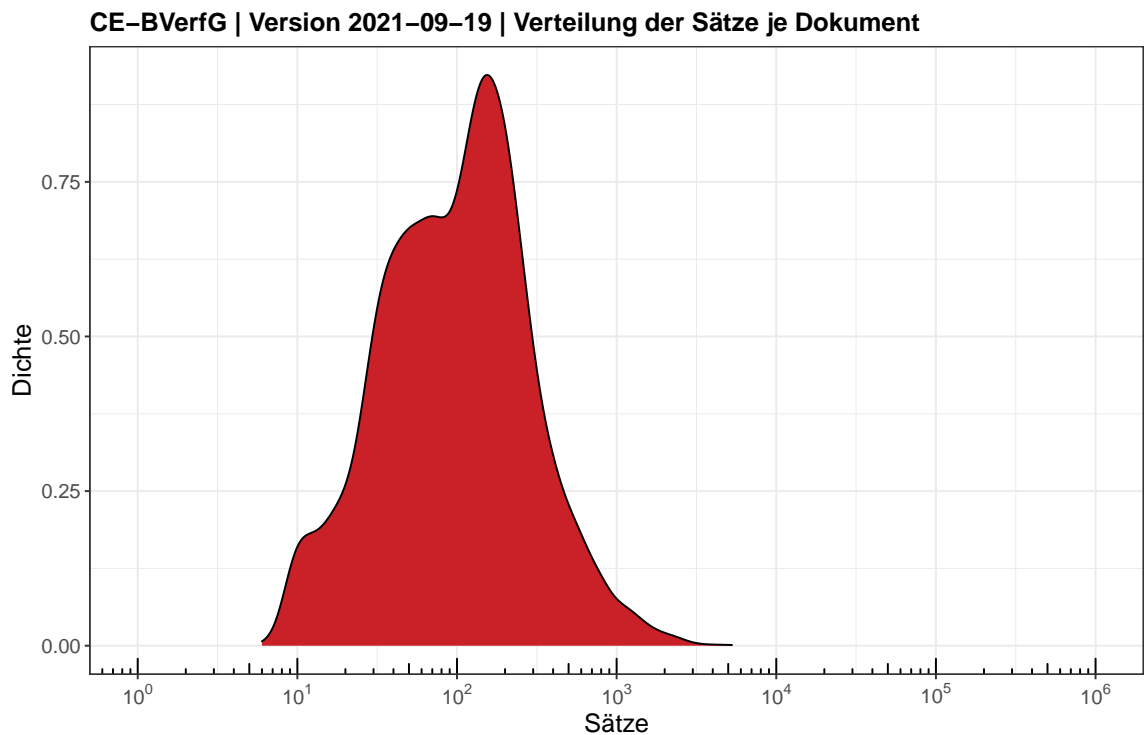
```
ggplot(data = meta.bverfg)+
  geom_density(aes(x = typen),
    fill = "#ca2129")+
  scale_x_log10(breaks = trans_breaks("log10", function(x) 10^x),
    labels = trans_format("log10", math_format(10^.x)))+
  annotation_logticks(sides = "b")+
  coord_cartesian(xlim = c(1, 10^6))+
  theme_bw()+
  labs(
    title = paste(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Typen je Dokument"),
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
)
```



DOI: 10.5281/zenodo.5514083 | Fobbe

14.7.4 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(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Sätze je Dokument"),
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
)
```



DOI: 10.5281/zenodo.5514083 | Fobbe

15 Linguistische Annotationen berechnen

```
if (mode.annotate == TRUE){  
  
  txt.annotated <- f.dopar.spacyparse(txt.bverfg,  
                                     threads = detectCores(),  
                                     chunksize = 1,  
                                     model = "de_core_news_sm",  
                                     pos = TRUE,  
                                     tag = TRUE,  
                                     lemma = TRUE,  
                                     entity = TRUE,  
                                     dependency = TRUE,  
                                     nounphrase = TRUE)  
  
}
```

```
## Found 'spacy_condaenv'. spacyr will use this environment
```

```
## successfully initialized (spaCy Version: 3.1.1, language model: de_core_news_  
sm)
```

```
## (python options: type = "condaenv", value = "spacy_condaenv")
```

```
## [1] "Parallel processing using 16 threads. Begin at 2021-09-19 18:08:14.  
Processing 8122 documents"  
## [1] "Runtime was 7.21 mins. Ended at 2021-09-19 18:15:27."
```


16 Kontrolle der Variablen

16.1 Semantische Sortierung der Variablen

16.1.1 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",  
    "zitativvorschlag",  
    "kurzbeschreibung",  
    "pressemitteilung",  
    "praesi",  
    "v_praesi",  
    "richter",  
    "zeichen",  
    "tokens",  
    "typen",  
    "saetze",  
    "version",  
    "doi_concept",  
    "doi_version",  
    "lizenz"))
```

16.1.2 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",  
    "zitativorschlag",  
    "kurzbeschreibung",  
    "pressemitteilung",  
    "praesi",  
    "v_praesi",  
    "richter",  
    "zeichen",  
    "tokens",  
    "typen",  
    "saetze",  
    "version",  
    "doi_concept",  
    "doi_version",  
    "lizenz"))
```

16.1.3 Variablen sortieren: Segmentiert

```
setcolororder(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",  
    "zitativorschlag",  
    "kurzbeschreibung",  
    "pressemitteilung",  
    "praesi",  
    "v_praesi",  
    "richter",  
    "version",  
    "doi_concept",  
    "doi_version",  
    "lizenz"))
```

16.2 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
```

16.3 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"         "zitativorschlag"
## [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"
```

17 CSV-Dateien erstellen

17.1 CSV mit vollem Datensatz speichern

```
csvname.full <- paste(datasetname,
                      datestamp,
                      "DE_CSV_Datensatz.csv",
                      sep = "_")

fwrite(txt.bverfg,
       csvname.full,
       na = "NA")
```

17.2 CSV mit Metadaten speichern

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

```
csvname.meta <- paste(datasetname,
                     datestamp,
                     "DE_CSV_Metadaten.csv",
                     sep = "_")

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

17.3 CSV mit Segmenten speichern

```
csvname.segmented <- paste(datasetname,
                          datestamp,
                          "DE_CSV_Segmentiert.csv",
                          sep = "_")

fwrite(dt.segmented.full,
       csvname.segmented,
       na = "NA")
```

17.4 CSV mit Annotationen speichern

```
if (mode.annotate == TRUE){

  csvname.annotated <- paste(datasetname,
                          datestamp,
                          "DE_CSV_Annotiert.csv",
                          sep = "_")
```

```
fwrite(txt.annotated,  
       csvname.annotated,  
       na = "NA")  
}
```

18 Dateigrößen analysieren

18.1 Gesamtgröße

18.1.1 Korpus-Objekt in RAM (MB)

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

```
## 168.7 MB
```

18.1.2 CSV Korpus (MB)

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

```
## [1] 166.2967
```

18.1.3 CSV Metadaten (MB)

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

```
## [1] 5.698702
```

18.1.4 CSV Annotiert (MB)

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

```
## [1] 3007.831
```

18.1.5 CSV Segmentiert (MB)

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

```
## [1] 332.8157
```


18.1.6 PDF-Dateien (MB)

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

```
## [1] 657.5543
```

18.1.7 TXT-Dateien (MB)

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

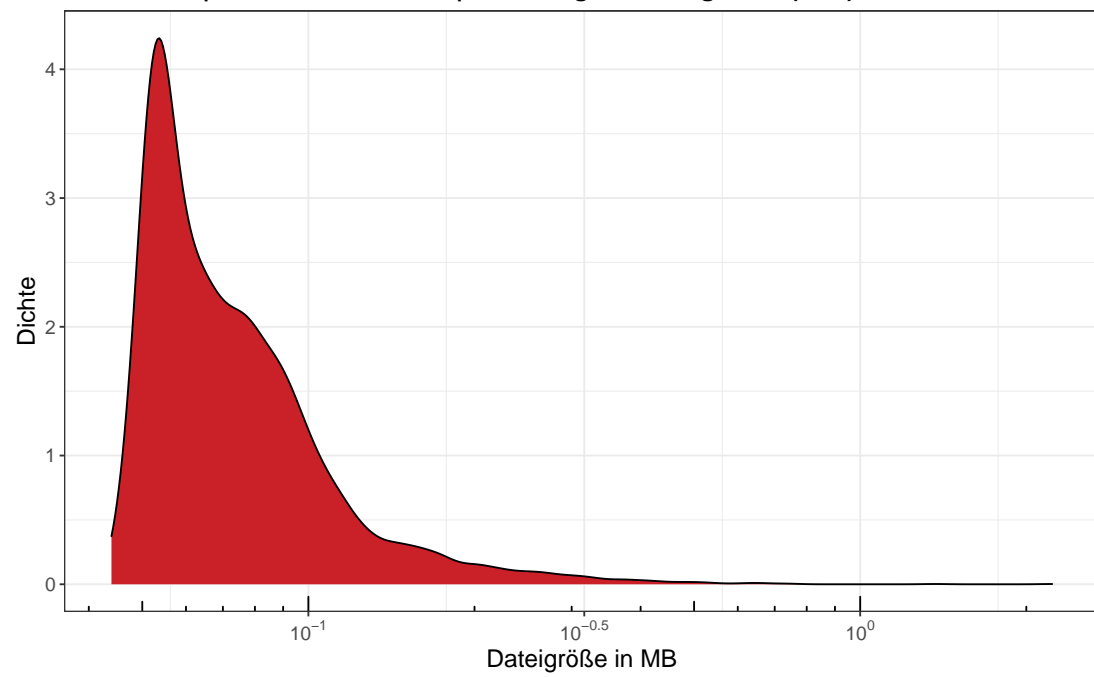
```
## [1] 163.0131
```

18.2 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(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Dateigrößen (PDF)" ),
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
```

CE-BVerfG | Version 2021-09-19 | Verteilung der Dateigrößen (PDF)

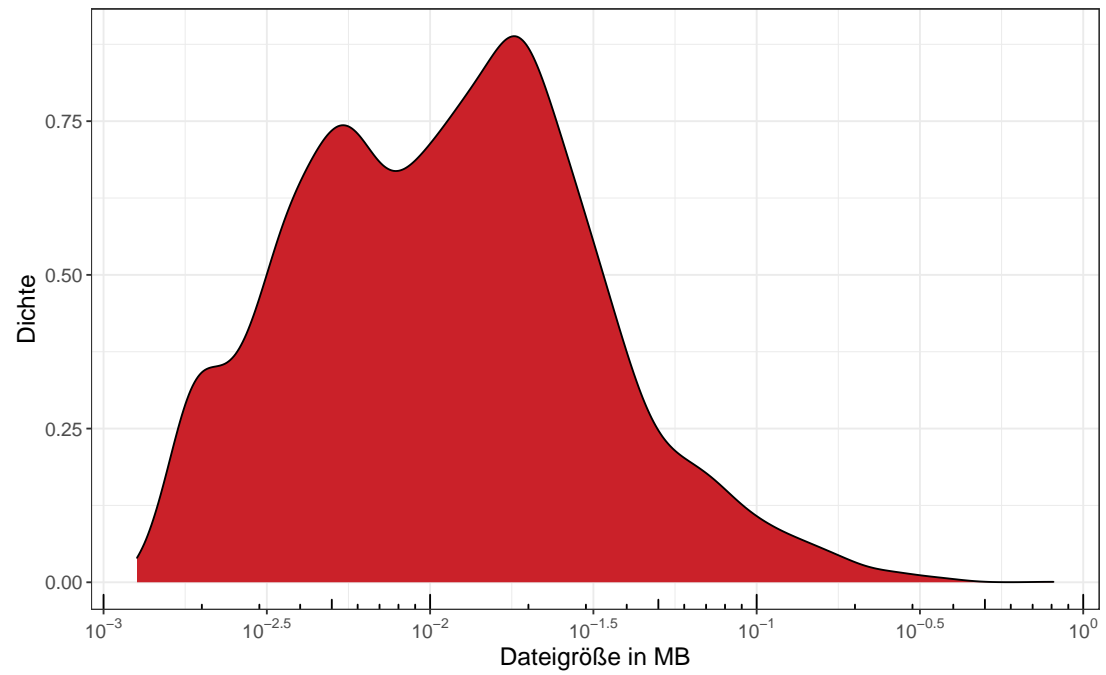


18.3 Diagramm: Verteilung der Dateigrößen (TXT)

```
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(datasetname,
      "| Version",
      datestamp,
      "| Verteilung der Dateigrößen (TXT)",
    caption = paste("DOI:",
      doi.version,
      "| Fobbe"),
    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)
  )
```

CE-BVerfG | Version 2021-09-19 | Verteilung der Dateigrößen (TXT)



19 Erstellen der ZIP-Archive

19.1 Verpacken der CSV-Dateien

19.1.1 Vollständiger Datensatz

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

19.1.2 Metadaten

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

19.1.3 Segmentiert

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

19.1.4 Annotiert

```
if (mode.annotate == TRUE){  
  
    csvname.annotated.zip <- gsub(".csv",  
                                ".zip",  
                                csvname.annotated)  
  
    zip(csvname.annotated.zip,  
        csvname.annotated)  
  
}
```

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

19.2 Verpacken der PDF-Dateien

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

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

unlink(files.pdf)
```

19.3 Verpacken der HTML-Dateien

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

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

unlink(files.html)
```

19.4 Verpacken der TXT-Dateien

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

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

unlink(files.txt)
```

19.5 Verpacken der Analyse-Dateien

```
zip(paste0(datasetname,  
           "_",  
           datestamp,  
           "_DE_",  
           basename(outputdir),  
           ".zip"),  
    basename(outputdir))
```

19.6 Verpacken der Source-Dateien

```
files.source <- c(list.files(pattern = "Source"),  
                  "buttons")  
  
files.source <- grep("spin",  
                    files.source,  
                    value = TRUE,  
                    ignore.case = TRUE,  
                    invert = TRUE)  
  
zip(paste(datasetname,  
          datestamp,  
          "Source_Files.zip",  
          sep = "_"),  
    files.source)
```


20 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.

20.1 Liste der ZIP-Archive erstellen

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

20.2 Funktion anzeigen

```
print(f.dopar.multihashes)
```

```
function(x, threads = detectCores()){
```

```
  print(paste("Parallel processing using", threads, "threads."))  
  
  begin <- Sys.time()  
  
  cl <- makeForkCluster(threads)  
  registerDoParallel(cl)  
  
  multihashes <- foreach(filename = x,  
                        .errorhandling = 'pass',  
                        .combine = 'rbind') %dopar% {  
  
    sha2.256 <- system2("openssl",  
                      paste("sha256",  
                            filename),  
                      stdout = TRUE)  
  
    sha2.256 <- gsub("^.*\\\\" = "  
    "",  
    sha2.256)  
  
    sha3.512 <- system2("openssl",  
                      paste("sha3-512",  
                            filename),  
                      stdout = TRUE)  
  
    sha3.512 <- gsub("^.*\\\\" = "  
    "",
```

```

                                sha3.512)

                                out <- data.frame(filename,
                                                    sha2.256,
                                                    sha3.512)
                                return(out)
                                }
stopCluster(cl)

end <- Sys.time()
duration <- end - begin

print(paste0("Processed ",
             length(x),
             " files. Runtime was ",
             round(duration,
                    digits = 2),
             " ",
             attributes(duration)$units,
             "."))

return(multihashes)

}

```

20.3 Hashes berechnen

```
multihashes <- f.dopar.multihashes(files.zip)
```

```
## [1] "Parallel processing using 16 threads."
## [1] "Processed 9 files. Runtime was 3.99 secs."
```

20.4 In Data Table umwandeln

```
setDT(multihashes)
```

20.5 Index hinzufügen

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

20.6 In Datei schreiben

```
fwrite(multihashes,
      paste(datasetname,
            datestamp,
            "KryptographischeHashes.csv",
            sep = "_"),
      na = "NA")
```

20.7 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))
```

20.8 In Bericht anzeigen

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

index	filename
1	CE-BVerfG_2021-09-19_DE_ANALYSE.zip
2	CE-BVerfG_2021-09-19_DE_CSV_Annotiert.zip
3	CE-BVerfG_2021-09-19_DE_CSV_Datensatz.zip
4	CE-BVerfG_2021-09-19_DE_CSV_Metadaten.zip
5	CE-BVerfG_2021-09-19_DE_CSV_Segmentiert.zip
6	CE-BVerfG_2021-09-19_DE_HTML_Datensatz.zip
7	CE-BVerfG_2021-09-19_DE_PDF_Datensatz.zip
8	CE-BVerfG_2021-09-19_DE_TXT_Datensatz.zip
9	CE-BVerfG_2021-09-19_Source_Files.zip

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

index	sha2.256
1	72064d20edc0383c1fcf3e7d718782f9ebaa6db9b8ccbaa3ce3b559b583883f4
2	26395d7cbb713ba74d80ccd94744d8cb4c10ef0c286d7cb5a19415ef35e9b2f4
3	53d3abab30339bcdd2a16014e5f9eb44d28df7040adb9b55a3976f596c097aff
4	ebe2fd4d5966b9a9c26e695520a6619fd5fadb9d467dbe61e3c1621c483915e1
5	29e00e80181fd44fa049382ac0f7a332e6c6a6ff1819950f8be1c5621fe6a58f
6	c87ef40e0a9be515c5c38497216ff9b78762a838246e8c0f7b4b556694c59716
7	1f8369f96894b1a428092ab4043545a6346f0dce7d41e1c9ed597cd822714a74
8	b1a3e033427f92e823f6725dcbccd5868a9419a98b60115fd1cbb1a317209310
9	645fae48f2f1525e1d7abba39a3eeef5f42e23bf3c7e45d778eedb2002b100ab

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

index	sha3.512
1	e65d98c5165219bc81719b87d6025181dc8df4c53cdbff5e4d5bb88afc44543d78bb32cad44f4dcf76a405230b3333df2758b5805a9498d36cb8f3d7ebc6b799
2	20e48066cd4e3bc39385ab4452a5449ee7abcb543d986dbc1fbb40f52effb1f43ce3a35723a5752eb45db71fe2fdc3f8fd932b3ce07273733cab279d3fba4157
3	8891eb6b2d01d02f4998a12da2431f4ba862c61465dd46a6d15695932ae2b54f02791291dfdd112812e69a638bc0a5454174c544b862c155b0ecd348875ee4b9
4	76c2784d3d849de966ae8691510e9735f2c864cc22d1b51c2b9e37609829e5840f3ee8240643c073edf16cc584c70860a64fb4bdb8647b51c63da9c1b3057913
5	a0d1819aec94268e710457a60d578238f14ef8f66358b55bfee6a1f07c0c146a279ea77fd310d98a492bbd8a63488865a441f4e4daa883920f66233eb2b2ab80

6	86378c7dc68e27ced6ba7a16c28e1568e7570dc1e1593f1f6b3c7671c0d4d87d d7434adeb494ee7448ddec1bb047aa4da7917aaa9ff135d10e08983f5b926c80
7	4d5d712d81aee802c52c59851679472317f93b931d4ac2da04a944a70814581d 8673c0a4e73b037aaa546f64ca0e51c86f3b757705ae482891589e97fa0b1f04
8	ad8f44ae559a19c4bc2237c2c81b3c95c3b3504187568663d7cc9b99134981c6 2f92bc2c4460f364edab831f71aaf178c947b653040f8b39b8f4b7410ab6fc4f
9	2fe26e087f1ee0471d22fc38eacac542880994f700cd9e95b6e7cd61ee3939b9 c55fda765a258e94402027caae5f45351c96c11eb76a4213f2564dca10ffc7a

21 Abschluss

21.1 Datumsstempel

```
print(datestamp)
```

```
## [1] "2021-09-19"
```

21.2 Datum und Uhrzeit (Anfang)

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

```
## [1] "2021-09-19 11:22:19 CEST"
```

21.3 Datum und Uhrzeit (Ende)

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

```
## [1] "2021-09-19 18:18:03 CEST"
```

21.4 Laufzeit des gesamten Skriptes

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

```
## Time difference of 6.928968 hours
```

21.5 Warnungen

```
warnings()
```

22 Parameter für strenge Replikationen

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

```
## [1] "OpenSSL 1.1.1k 25 Mar 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] parallel stats      graphics grDevices utils      datasets methods
## [8] base
##
## other attached packages:
##  [1] spacyr_1.2.1      quanteda_3.1.0    readtext_0.81     data.table_1.14.0
##  [5] scales_1.1.1      ggplot2_3.3.5     doParallel_1.0.16 iterators_1.0.13
##  [9] foreach_1.5.1     pdftools_3.0.1    kableExtra_1.3.4  knitr_1.34
## [13] rvest_1.0.1       httr_1.4.2        mgsub_1.7.3
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.7        svglite_2.0.0     lattice_0.20-44   png_0.1-7
##  [5] digest_0.6.27     utf8_1.2.2        R6_2.5.1          evaluate_0.14
##  [9] highr_0.9         pillar_1.6.2      rlang_0.4.11      curl_4.3.2
## [13] rstudioapi_0.13   magick_2.7.3      Matrix_1.3-4      reticulate_1.20
## [17] rmarkdown_2.10    qpdf_1.1          labeling_0.4.2     webshot_0.5.2
## [21] stringr_1.4.0     selectr_0.4-2     munsell_0.5.0     compiler_4.0.5
## [25] xfun_0.25         pkgconfig_2.0.3   askpass_1.1       systemfonts
## [29] htmltools_0.5.2   tidyselect_1.1.1  tibble_3.1.4      codetools_0.2-18
## [33] fansi_0.5.0       viridisLite_0.4.0 crayon_1.4.1      dplyr_1.0.7
## [37] withr_2.4.2       rappdirs_0.3.3    grid_4.0.5        jsonlite_1.7.2
## [41] gtable_0.3.0      lifecycle_1.0.0   magrittr_2.0.1    RcppParallel
## [45] stringi_1.7.4     farver_2.1.0      xml2_1.3.2        ellipsis_0.3.2
## [49] stopwords_2.2     generics_0.1.0    vctrs_0.3.8       fastmatch_1.1-3
```

## [53]	tools_4.0.5	glue_1.4.2	purrr_0.3.4	fastmap_1.1.0
## [57]	yaml_2.2.1	colorspace_2.0-2		

Literaturverzeichnis

- Analytics, Revolution, and Steve Weston. 2020. *Iterators: Provides Iterator Construct*. <https://github.com/RevolutionAnalytics/iterators>.
- 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>.
- Corporation, Microsoft, and Steve Weston. 2020. *DoParallel: Foreach Parallel Adaptor for the Parallel Package*. <https://CRAN.R-project.org/package=doParallel>.
- Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of 'Data.frame'*. <https://CRAN.R-project.org/package=data.table>.
- 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/>.
- Revolution Analytics, and Steve Weston. n.d. *Foreach: Provides Foreach Looping Construct*.
- 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>.