

**Corpus der Entscheidungen
des
Bundesverwaltungsgerichts
(CE-BVerwG)**

CODEBOOK

Version 2021-04-15



DOI: [10.5281/zenodo.4625123](https://doi.org/10.5281/zenodo.4625123)

Titel	Corpus der Entscheidungen des Bundesverwaltungsgerichts
Abkürzung	CE-BVerwG
Autor	Seán Fobbe
Version	2021-04-15
Download	https://doi.org/10.5281/zenodo.4625123
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Seán Fobbe (2021). Corpus der Entscheidungen des Bundesverwaltungsgerichts (CE-BVerwG). Version 2021-04-15. Zenodo. DOI: 10.5281/zenodo.4625123.

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 des Datensatzes. Sie verweist daher nur auf Version 2021-04-15. Für das Gesamtkonzept dieses Datensatzes 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.3911067. Die »Concept DOI« verlinkt immer die aktuellste Version.

Urheberrecht

Der Datensatz und dieses Dokument sind unter einer **Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication Lizenz** veröffentlicht. Ich stelle den Datensatz und das Codebook vollständig gemeinfrei und verzichte weltweit auf alle damit verbundenen Urheberrechte, einschließlich aller ähnlichen Rechte, soweit dies gesetzlich möglich ist.

Sie können die Werke kopieren, modifizieren, verteilen und aufführen ohne um Erlaubnis bitten zu müssen, selbst für kommerzielle Zwecke. Patente und Markenschutzrechte bleiben von CC0 unberührt. CC0 hat auch keine Auswirkungen auf etwaige Datenschutz- oder Persönlichkeitsrechte. Jegliche Haftung für die Benutzung dieses Werkes ist ausgeschlossen, bis zu dem maximalen Umfang in dem dies gesetzlich möglich ist.

Wenn Sie diese Werke nutzen oder zitieren sollten Sie nicht den Eindruck erwecken, der Autor unterstütze ihre Nutzung.

Dies ist nur eine unverbindliche deutsche Zusammenfassung der Lizenz, den vollständigen und rechtsverbindlichen Lizenztext finden Sie hier: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>

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	Einführung	5
2	Nutzung	6
2.1	CSV-Dateien	6
2.2	TXT-Dateien	6
3	Konstruktion	7
3.1	Beschreibung des Datensatzes	7
3.2	Datenquellen	7
3.3	Sammlung der Daten	7
3.4	Source Code und Compilation Report	7
3.5	Grenzen des Datensatzes	8
3.6	Urheberrechtsfreiheit von Rohdaten und Datensatz	8
3.7	Metadaten	8
3.7.1	Allgemein	8
3.7.2	Schema für die Dateinamen	8
3.7.3	Beispiel eines Dateinamens	9
3.8	Qualitätsprüfung	9
3.9	Grafische Darstellung	9
4	Varianten und Zielgruppen	10
5	Variablen	12
5.1	Datenstruktur	12
5.2	Hinweise	13
5.3	Erläuterungen zu den einzelnen Variablen	13
6	Registerzeichen	18
7	Präsident:innen	20
7.1	Hinweise	20
7.2	Lebensdaten	20
7.3	Dienstalter und Lebensalter	21
8	Vize-Präsident:innen	22
8.1	Hinweise	22
8.2	Lebensdaten	22
8.3	Dienstalter und Lebensalter	23
9	Linguistische Kennzahlen	24
9.1	Erläuterung der Kennzahlen und Diagramme	24
9.2	Werte der Kennzahlen	24
9.3	Verteilung Zeichen	25
9.4	Verteilung Tokens	25
9.5	Verteilung Typen	26
9.6	Verteilung Sätze	26
10	Inhalt des Korpus	27
10.1	Zusammenfassung	27

10.2	Nach Typ der Entscheidung	27
10.3	Nach Spruchkörper (Aktenzeichen)	28
10.4	Nach Registerzeichen	29
10.5	Nach Präsident:in	31
10.6	Nach Vize-Präsident:in	32
10.7	Nach Entscheidungsjahr	33
10.8	Nach Eingangsjahr (ISO)	35
11	Dateigrößen	37
11.1	Verteilung PDF	37
11.2	Verteilung TXT	37
11.3	Gesamtgröße je ZIP-Archiv	38
12	Prüfung kryptographischer Signaturen	39
12.1	Allgemeines	39
12.2	Persönliche GPG-Signatur	39
12.3	Import: Public Key	39
12.4	Prüfung: GPG-Signatur der Hash-Datei	40
12.5	Prüfung: SHA3-512 Hashes der ZIP-Archive	41
13	Changelog	42
14	Parameter für strenge Replikationen	43
	Literaturverzeichnis	44

1 Einführung

Das **Bundesverwaltungsgericht (BVerwG)** ist einer der fünf obersten Gerichtshöfe des Bundes und steht an der Spitze der Verwaltungsgerichtsbarkeit der Bundesrepublik Deutschland. Im Rahmen der Verwaltungsgerichtsbarkeit ist es für alle öffentlich-rechtlichen Streitigkeiten nichtverfassungsrechtlicher Art zuständig (§ 40 Absatz 1 Satz 1 VwGO). Es wurde am 23. September 1952 durch Bundesgesetz errichtet.¹

Am Bundesverwaltungsgericht eingerichtet sind derzeit 13 reguläre Senate, davon zehn Revisionssenate, zwei Wehrdienstsenate und ein Fachsenat nach § 189 VwGO.² Zusätzlich besteht ein Großer Senat nach § 11 VwGO und es sind Güterichter nach § 173 Satz 1 VwGO benannt.³ Die drei Disziplinarsenate wurden 1971, 2004 und 2015 aufgelöst.

Die weit überwiegende Anzahl der Verfahren vor dem Bundesverwaltungsgericht sind Revisionen, d.h. die Überprüfung von Entscheidungen unterer Instanzen (Verwaltungsgericht, Oberverwaltungsgericht/Verwaltungsgerichtshof) auf Rechtsfehler ohne erneute Beweisaufnahme (§§ 49, 132 ff VwGO).⁴ In wenigen Fällen — beispielsweise Klagen zwischen Bund und Bundesländern, zwischen Bundesländern, Vereinsverboten und Angelegenheiten des Bundesnachrichtendienstes — ist das Bundesverwaltungsgericht erste und letzte Instanz (§ 50 VwGO).

Die quantitative Analyse von juristischen Texten, insbesondere denen des Bundesverwaltungsgerichts, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.⁵ Zu einem nicht unerheblichen Teil liegt dies auch daran, dass die Anzahl an frei nutzbaren Datensätzen außerordentlich gering ist.

Die meisten hochwertigen Datensätze lagern (fast) unerreichbar in kommerziellen Datenbanken und sind wissenschaftlich gar nicht oder nur gegen Entgelt zu nutzen. Frei verfügbare Datenbanken wie *Opinio Iuris*⁶ und *openJur*⁷ verbieten ausdrücklich das maschinelle Auslesen der Rohdaten.⁸ Wissenschaftliche Initiativen wie der Juristische Referenzkorpus (JuReKo) sind nach jahrelanger Arbeit hinter verschlossenen Türen verschwunden.

In einem funktionierenden Rechtsstaat muss die Rechtsprechung öffentlich, transparent und nachvollziehbar sein. Im 21. Jahrhundert bedeutet dies auch, dass sie systematischer Überprüfung mittels quantitativen Analysen zugänglich sein muss. Der Erstellung und Aufbereitung des Datensatzes liegen daher die Prinzipien der allgemeinen Verfügbarkeit durch Urheberrechtsfreiheit, strenge Transparenz und vollständige wissenschaftliche Reproduzierbarkeit zugrunde. Die FAIR-Prinzipien (Findable, Accessible, Interoperable and Reusable) für freie wissenschaftliche Daten inspirieren sowohl die Konstruktion, als auch die Art der Publikation.⁹

¹ Gesetz über das Bundesverwaltungsgericht, 23. September 1952, BGBl. I, S. 625.

² Geschäftsverteilungsplan des Bundesverwaltungsgerichts für das Geschäftsjahr 2021. <<https://www.bverw.de/rechtsprechung/geschaeftsverteilungsplan>>

³ Ibid.

⁴ Baden-Württemberg, Bayern und Hessen machen von der Option in § 184 VwGO Gebrauch und bezeichnen ihr Oberverwaltungsgericht als »Verwaltungsgerichtshof«.

⁵ Besonders positive Ausnahmen finden sich unter: <https://www.quantitative-rechtswissenschaft.de/>
<https://opinioiuris.de/>

⁷ <https://openjur.de/>

⁸ Openjur beabsichtigt eine API anzubieten, diese war 2021 aber immernoch nicht verfügbar. Openjur ist seit 2008 in Betrieb.

⁹ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

2 Nutzung

Die Daten sind in offenen, interoperablen und weit verbreiteten Formaten (CSV, TXT, PDF) veröffentlicht. Sie lassen sich grundsätzlich mit allen modernen Programmiersprachen (z.B. Python oder R), sowie mit grafischen Programmen nutzen.

Wichtig: Nicht vorhandene Werte sind sowohl in den Dateinamen als auch in der CSV-Datei mit »NA« codiert.

2.1 CSV-Dateien

Am einfachsten ist es die **CSV-Dateien** einzulesen. CSV¹⁰ ist ein einfaches und maschinell gut lesbares Tabellen-Format. In diesem Datensatz sind die Werte komma-separiert. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Die Variablen sind unter Punkt 5 genauer erläutert.

Zum Einlesen empfehle ich für **R** dringend das package **data.table** (via CRAN verfügbar). Dessen Funktion **fread()** ist etwa zehnmal so schnell wie die normale **read.csv()**-Funktion in Base-R. Sie erkennt auch den Datentyp von Variablen sicherer. Ein Vorschlag:

```
library(data.table)
dt.bverwg <- fread("filename.csv")
```

2.2 TXT-Dateien

Die **TXT-Dateien** inklusive Metadaten können zum Beispiel mit **R** und dem package **readtext** (via CRAN verfügbar) eingelesen werden. Ein Vorschlag:

```
library(readtext)
df.bverwg <- readtext("./*.txt",
  docvarsfrom = "filenames",
  docvarnames = c("gericht",
    "datum",
    "entscheidung_typ",
    "spruchkoerper_az",
    "registerzeichen",
    "eingangsnummer",
    "eingangsjahr_az",
    "verzoegerung",
    "kollision"),
  dvsep = "_",
  encoding = "UTF-8")
```

¹⁰ Das CSV-Format ist in RFC 4180 definiert, siehe <https://tools.ietf.org/html/rfc4180>

3 Konstruktion

3.1 Beschreibung des Datensatzes

Dieser Datensatz ist eine digitale Zusammenstellung von möglichst allen Entscheidungen, die auf der amtlichen Internetpräsenz des Bundesverwaltungsgerichts (BVerwG) am jeweiligen Stichtag veröffentlicht waren. Die Stichtage für jede Version entsprechen exakt der Versionsnummer.

Zusätzlich zu den aufbereiteten maschinenlesbaren Formaten (TXT und CSV) sind die PDF-Rohdaten enthalten, damit Analyst:innen gegebenenfalls ihre eigene Konvertierung vornehmen können. Die PDF-Rohdaten wurden inhaltlich nicht verändert und nur die Dateinamen angepasst um die Lesbarkeit für Mensch und Maschine zu verbessern.

3.2 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	https://www.bverwg.de
Source Code	https://doi.org/10.5281/zenodo.4625135
Personendaten	https://doi.org/10.5281/zenodo.4568682
Registerzeichen	https://doi.org/10.5281/zenodo.4569564

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«.

Die Tabelle der Registerzeichen und der ihnen zugeordneten Verfahrensarten stammt aus dem folgenden Datensatz: »Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.«

3.3 Sammlung der Daten

Die Daten wurden unter Beachtung des Robot Exclusion Standard (RES) gesammelt. Der Abruf geschieht ausschließlich über TLS-verschlüsselte Verbindungen. Die Entscheidungen sind laut dem Gericht anonymisiert, aber ungekürzt.

3.4 Source Code und Compilation Report

Der gesamte Source Code — sowohl für die Erstellung des Datensatzes, als auch für dieses Codebook — ist öffentlich einsehbar und dauerhaft erreichbar im wissenschaftlichen Archiv des CERN unter dieser Adresse hinterlegt: <https://doi.org/10.5281/zenodo.4625135>

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem Codebook). Der Compilation Report enthält den vollständigen und kommentierten Source Code, dokumentiert relevante Rechenergebnisse, gibt sekundengenaue Zeitstempel an und

ist mit einem klickbaren Inhaltsverzeichnis versehen. Er ist zusammen mit dem Source Code hinterlegt. Wenn Sie sich für Details der Herstellung interessieren, lesen Sie diesen bitte zuerst.

3.5 Grenzen des Datensatzes

Nutzer:innen sollten folgende wichtige Grenzen beachten:

1. Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*).
2. Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Entscheidungen nicht oder nicht korrekt abgerufen werden (*automation bias*).
3. Es werden nur PDF-Dateien abgerufen (*file type bias*). Manche Entscheidungen sind nur als HTML verfügbar.
4. Erst ab dem Jahr 2002 sind begründete Entscheidungen des Bundesverwaltungsgerichts einigermaßen vollständig veröffentlicht (*temporal bias*). Die Frequenztabellen geben hierzu genauer Auskunft.

3.6 Urheberrechtsfreiheit von Rohdaten und Datensatz

An den Entscheidungstexten und amtlichen Leitsätzen besteht gem. § 5 Abs. 1 UrhG kein Urheberrecht, da sie amtliche Werke sind. § 5 UrhG ist auf amtliche Datenbanken analog anzuwenden (BGH, Beschluss vom 28.09.2006, I ZR 261/03, »Sächsischer Ausschreibungsdienst«).

Alle eigenen Beiträge (z.B. durch Zusammenstellung und Anpassung der Metadaten) und damit den gesamten Datensatz stelle ich gemäß einer *CC0 1.0 Universal Public Domain Lizenz* vollständig urheberrechtsfrei.

3.7 Metadaten

3.7.1 Allgemein

Die Metadaten wurden größtenteils aus den Hyperlinks zur jeweiligen Datei extrahiert. Hinzugefügt wurden von mir eine Reihe weitere Variablen, sowie in den Dateinamen der PDF/TXT-Dateien Unter- und Trennstriche, um die Maschinenlesbarkeit zu erleichtern. Der volle Satz an Metadaten ist nur in den CSV-Dateien enthalten. Alle hinzugefügten Metadaten sind vollständig maschinenlesbar dokumentiert. Sie sind entweder im Source Code enthalten, mit dem Source Code zusammen dokumentiert oder über dauerhaft stabile Identifikatoren (z.B. DOI) zitiert.

Die Dateinamen der PDF- und TXT-Dateien enthalten Gerichtsname, Datum (Langform nach ISO-8601, d.h. YYYY-MM-DD), den Typ der Entscheidung, das offizielle Aktenzeichen, die Angabe ob es sich um eine Verzögerungsrüge handelt und eine Kollisions-ID.

3.7.2 Schema für die Dateinamen

```
[gericht]_[datum]_[entscheidung_typ]_[spruchkoerper_az]_[registerzeichen]_[eingangsnummer]_[eingangsjahr_az]_[verzoeigerung]_[kollision]
```


3.7.3 Beispiel eines Dateinamens

BVerwG_2004-08-12_U_5_C_58_03_NA_0.pdf

3.8 Qualitätsprüfung

Die Typen der Variablen wurden mit *regular expressions* strikt validiert. Die möglichen Werte der jeweiligen Variablen wurden zudem durch Frequenztabellen und Visualisierungen auf ihre Plausibilität geprüft. Insgesamt werden zusammen mit jeder Kompilierung Dutzende Tests zur Qualitätsprüfung durchgeführt. Alle Ergebnisse der Qualitätsprüfungen sind aggregiert im Compilation Report und einzeln im Archiv »ANALYSE« zusammen mit dem Datensatz veröffentlicht.

3.9 Grafische Darstellung

Die Robenfarbe der Bundesverwaltungsrichter ist »karmesinrot«.¹¹ Der Hex-Wert hierfür ist vermutlich #7e0731. Das ist besonders bei der Erstellung thematisch passender Graphen hilfreich. Alle im Compilation Report und diesem Codebook präsentierten Graphen sind in diesem karmesinrot gehalten.

¹¹ Artikel 1, Absatz II der Anordnung des Bundespräsidenten über die Amtstracht bei dem Bundesverwaltungsgericht in der im Bundesgesetzblatt Teil III, Gliederungsnummer 340-3, veröffentlichten bereinigten Fassung, die zuletzt durch die Anordnung vom 19. Dezember 2001 (BGBl. I S. 3777) geändert worden ist.

4 Varianten und Zielgruppen

Dieser Datensatz ist in verschiedenen Varianten verfügbar, die sich an unterschiedliche Zielgruppen richten. Zielgruppe sind nicht nur quantitativ forschende Rechtswissenschaftler:innen, sondern auch traditionell arbeitende Jurist:innen. Idealerweise müssen quantitative Methoden ohnehin immer durch qualitative Interpretation, Theoriebildung und kritische Auseinandersetzung verstärkt werden (*mixed methods approach*).

Lehrende werden von den vorbereiteten Tabellen und Diagrammen besonders profitieren, die bei der Erläuterung der Charakteristika der Daten hilfreich sein können und Zeit im universitären Alltag sparen. Alle Tabellen und Diagramme liegen auch als separate Dateien vor um sie einfach z.B. in Präsentations-Folien oder Handreichungen zu integrieren.

Variante	Zielgruppe und Beschreibung
PDF	Traditionelle juristische Forschung. Die PDF-Dokumente wie sie vom Bundesverwaltungsgericht auf der amtlichen Webseite bereitgestellt werden, jedoch verbessert durch semantisch hochwertige Dateinamen, die der leichteren Auffindbarkeit von Entscheidungen dienen. Die Dateinamen sind so konzipiert, dass sie auch für die traditionelle qualitative juristische Arbeit einen erheblichen Mehrwert bieten. Im Vergleich zu den CSV-Dateien enthalten die Dateinamen nur einen reduzierten Umfang an Metadaten, um Kompatibilitätsprobleme unter Windows zu vermeiden und die Lesbarkeit zu verbessern.
CSV_Datensatz	Legal Tech/Quantitative Forschung. Diese CSV-Datei ist die für statistische Analysen empfohlene Variante des Datensatzes. Sie enthält den Volltext aller Entscheidungen, sowie alle in diesem Codebook beschriebenen Metadaten. Über Zeilenumbrüche getrennte Wörter wurden zusammengefügt.
CSV_Metadaten	Legal Tech/Quantitative Forschung. Wie die andere CSV-Datei, nur ohne die Entscheidungstexte. Sinnvoll für Analyst:innen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen.
TXT	Subsidiär für alle Zielgruppen. Diese Variante enthält die vollständigen aus den PDF-Dateien extrahierten Entscheidungstexte, aber nur einen reduzierten Umfang an Metadaten, der dem der PDF-Dateien entspricht. Die TXT-Dateien sind optisch an das Layout der PDF-Dateien angelehnt. Geeignet für qualitativ arbeitende Forscher:innen, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben oder für quantitativ arbeitende Forscher:innen, die beim Einlesen der CSV-Dateien Probleme haben. Über Zeilenumbrüche getrennte Wörter wurden <i>nicht</i> zusammengefügt.

Variante	Zielgruppe und Beschreibung
ANALYSE	<p>Alle Lehrenden und Forschenden. Dieses Archiv enthält alle während dem Kompilierungs- und Prüfprozess erstellten Tabellen (CSV) und Diagramme (PDF, PNG) im Original. Sie sind inhaltsgleich mit den in diesem Codebook verwendeten Tabellen und Diagrammen. Das PDF-Format eignet sich besonders für die Verwendung in gedruckten Publikationen, das PNG-Format besonders für die Darstellung im Internet. Analyst:innen mit fortgeschrittenen Kenntnissen in R können auch auf den Source Code zurückgreifen. Empfohlen für Nutzer:innen die einzelne Inhalte aus dem Codebook für andere Zwecke (z.B. Präsentationen, eigene Publikationen) weiterverwenden möchten.</p>

5 Variablen

5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 24943 obs. of 25 variables:
## $ doc_id : chr "BVerwG_1997-02-26_U_6_C_3_96_NA_0.txt" "BVerwG
_2000-12-14_B_1_WB_107_00_NA_0.txt" "BVerwG_2002-01-10_B_9_A_9_02_NA_0.txt" "
BVerwG_2002-01-14_B_4_BN_1_02_NA_0.txt" ...
## $ text : chr " BUNDESVERWALTUNGSGERICHT\n
IM NAMEN DES VOLKES\n URTEIL\nBVer" | __truncated__
" BUNDESVERWALTUNGSGERICHT\n
Beschluss\nBVerwG 1 WB 107.00, 1 WB 113.00" | __truncated__ "
BUNDESVERWALTUNGSGERICHT\n BESCHLUSS\nBVerwG 9 A 9.02
(vormals 4 A 74.01)\n " | __truncated__ " BUNDESVERWALTUNGSGERICHT\n
n BESCHLUSS\nBVerwG 4 BN 1.02\nVGH 8 S 1853/01\n
" | __truncated__ ...
## $ gericht : chr "BVerwG" "BVerwG" "BVerwG" "BVerwG" ...
## $ datum : chr "1997-02-26" "2000-12-14" "2002-01-10" "2002-01-14"
...
## $ entscheidung_typ : chr "U" "B" "B" "B" ...
## $ spruchkoerper_az : chr "6" "1" "9" "4" ...
## $ registerzeichen : chr "C" "WB" "A" "BN" ...
## $ eingangsnummer : int 3 107 9 1 12 3 4 5 1 2 ...
## $ eingangsjahr_az : int 96 0 2 2 2 2 2 2 2 ...
## $ verzoegerung : chr NA NA NA NA ...
## $ kollision : int 0 0 0 0 0 0 0 0 0 ...
## $ entscheidungsjahr: int 1997 2000 2002 2002 2002 2002 2002 2002 2002
...
## $ eingangsjahr_iso : int 1996 2000 2002 2002 2002 2002 2002 2002 2002
...
## $ praesi : chr "Franßen" "Franßen" "Franßen" "Franßen" ...
## $ v_praesi : chr "Franke" "Hien" "Hien" "Hien" ...
## $ aktenzeichen : chr "BVerwG 6 C 3.96" "BVerwG 1 WB 107.0" "BVerwG 9 A
9.2" "BVerwG 4 BN 1.2" ...
## $ eclli : chr "ECLI:DE:BVerwG:1997:260297U6C3.96.0" "ECLI:DE:
BVerwG:2000:141200B1WB107.00.0" "ECLI:DE:BVerwG:2002:100102B9A9.02.0" "ECLI:
DE:BVerwG:2002:140102B4BN1.02.0" ...
## $ verfahrensart : chr "Revision (Verwaltungsstreitsachen)" "Verfahren
nach Wehrbeschwerdeordnung" "Erstinstanzliche Klage, inklusive
Wiederaufnahmeverfahren (Verwaltungsstreitsachen)" "Nichtzulassungsbeschwerde
(Normenkontrollsaachen)" ...
## $ doi_concept : chr "10.5281/zenodo.3911067" "10.5281/zenodo.3911067"
"10.5281/zenodo.3911067" "10.5281/zenodo.3911067" ...
## $ doi_version : chr "10.5281/zenodo.4625123" "10.5281/zenodo.4625123"
"10.5281/zenodo.4625123" "10.5281/zenodo.4625123" ...
## $ version : IDate, format: "2021-04-15" "2021-04-15" ...
## $ zeichen : int 37358 13741 805 3619 1384 1376 1382 1380 2445 2742
...
## $ tokens : int 5471 2336 128 557 213 213 213 213 377 418 ...
## $ typen : int 1470 721 84 277 128 128 126 128 224 227 ...
## $ saetze : int 267 143 15 48 18 18 18 18 21 27 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

5.2 Hinweise

- Fehlende Werte sind immer mit »NA« codiert
- Strings können grundsätzlich alle in UTF-8 definierten Zeichen (insbesondere Buchstaben, Zahlen und Sonderzeichen) enthalten.

5.3 Erläuterungen zu den einzelnen Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV-Datei) Der Name der extrahierten TXT-Datei.
text	String	(Nur CSV-Datei) Der vollständige Inhalt der Entscheidung, so wie er in der von www.bverwg.de heruntergeladenen PDF-Datei dokumentiert ist. Die einzige editorische Änderung gegenüber den TXT-Dateien ist die Zusammenfügung von über Zeilengrenzen gebrochenen Wörtern.
gericht	Alphabetisch	In diesem Datensatz ist nur der Wert »BVerwG« vergeben. Dies ist der ECLI-Gerichtscode für »Bundesverwaltungsgericht«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Datum (ISO)	Das Datum der Entscheidung im Format YYYY-MM-DD (Langform nach ISO-8601). Die Langform ist für Menschen einfacher lesbar und wird maschinell auch öfter automatisch als Datumsformat erkannt.
entscheidung_typ	Alphabetisch	Der Typ der Entscheidung. Es sind die Werte »B« (Beschluss), »U« (Urteil) und »G« (Gerichtsbescheid) vergeben.
spruchkoerper_az	Natürliche Zahl	Der im Aktenzeichen angegebene Spruchkörper. Es sind die Werte »1« bis »10« (Revisions- und Wehrdienstsenate) und »20« (Fachsenat) vergeben. Die Werte stehen für den jeweiligen Senat des Gerichts. Achtung: die ersten beiden Revisionssenate und die beiden Wehrdienstsenate sind inhaltsgleich mit »1« und »2« in den jeweiligen Aktenzeichen codiert. Um zwischen den Revisions- und Wehrdienstsenaten zu unterscheiden müssen Sie in einer Analyse zwingend auch nach dem Registerzeichen differenzieren. Die Registerzeichen der Wehrdienstsenate beginnen immer mit »W«.

Variable	Typ	Erläuterung
registerzeichen	String	Das amtliche Registerzeichen. Es gibt die Verfahrensart an, in der die Entscheidung ergangen ist. Eine Erläuterung der Registerzeichen und der zugehörigen Verfahrensarten findet sich unter Punkt 6.
verfahrensart	String	Die ausführliche Beschreibung der Verfahrensart, die dem Registerzeichen zugeordnet ist. Eine Erläuterung der Registerzeichen und der zugehörigen Verfahrensarten findet sich unter Punkt 6.
eingangsnummer	Natürliche Zahl	Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine Nummer in der Reihenfolge ihres Eingangs.
eingangsjahr_az	Natürliche Zahl	Das im Aktenzeichen angegebene Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Natürliche Zahl	(Nur CSV-Datei) Das nach ISO-8601 codierte Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine vierstellige Jahreszahl (YYYY), um eine maschinenlesbare und eindeutige Jahreszahl für den Eingang zur Verfügung zu stellen. Wurde aus der Variable »ingangsjahr_az« durch den Autor des Datensatzes berechnet, unter der Annahme, dass Jahreszahlen über 50 dem 20. Jahrhundert zugeordnet sind und andere Jahreszahlen dem 21. Jahrhundert.
entscheidungsjahr	Natürliche Zahl	(Nur CSV-Datei) Das Jahr in dem die Entscheidung ergangen ist. Das Format ist eine vierstellige Jahreszahl (YYYY). Wurde aus der Variable »datum« durch den Autor des Datensatzes berechnet.
verzoeigerung	Alphabetisch	Verzögerungsrügen sind mit »D« gekennzeichnet. Alle anderen Entscheidungen (der weit überwiegende Anteil) sind mit dem Wert »NA« codiert.
kollision	Numerisch	In wenigen Fällen sind am gleichen Tag mehrere Entscheidungen zum gleichen Aktenzeichen ergangen. Diese werden ab der zweiten Entscheidung pro Tag durch eine Kollisions-ID mit einer fortlaufenden Zahl ausdifferenziert. Für die erste Entscheidung ist der Wert der Variable »0«. Die zweite Entscheidung ist mit »1« identifiziert, die dritte mit »2« und so fort.
praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.

Variable	Typ	Erläuterung
v_praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der Vize-Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
aktenzeichen	String	(Nur CSV-Datei) Das amtliche Aktenzeichen. Die Variable wurde aus den Variablen »spruchkoerper_az«, »registerzeichen«, »eingangsnummer« und »eingangsjahr_az« durch den Autor des Datensatzes berechnet.
ecli	String	(Nur CSV-Datei) Der European Case Law Identifier (ECLI) der Entscheidung. Jeder Entscheidung ist eine einzigartige ECLI zugewiesen. Die ECLI ist vor allem dann hilfreich, wenn dieser Datensatz mit anderen Datensätzen zusammengeführt und Duplikate vermieden werden sollen. Alle inhaltlichen Bestandteile der ECLI sind in diesem Datensatz zusätzlich auch anderen und besser verständlichen Variablen zugewiesen. Nutzen Sie bevorzugt diese anderen Variablen, statt Informationen aus der ECLI zu extrahieren. Die Variable wurde aus den Variablen »gericht«, »entscheidungsjahr«, »datum«, »entscheidung_typ«, »spruchkoerper_az«, »registerzeichen«, »eingangsnummer«, »eingangsjahr_az«, »verzoegerung« und »kollision« durch den Autor des Datensatzes berechnet.
zeichen	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Zeichen eines Dokumentes.
tokens	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.

Variable	Typ	Erläuterung
typen	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl <i>einzigartiger</i> Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung und Typenzählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
saetze	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Sätze. Die Definition entspricht in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail allerdings sehr komplex und in »Unicode Standard: Annex No 29« beschrieben. Diese Zahl kann je nach Software und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Zählung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.
version	Datum (ISO)	(Nur CSV-Datei) Die Versionsnummer des Datensatzes im Format YYYY-MM-DD (Langform nach ISO-8601). Die Versionsnummer entspricht immer dem Datum an dem der Datensatz erstellt und die Daten von der Webseite des Gerichts abgerufen wurden.
doi_concept	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) des Gesamtkonzeptes des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer die aktuellste Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien (»data are assigned globally unique and persistent identifiers«) empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.

Variable	Typ	Erläuterung
doi_version	String	(Nur CSV-Datei) Der Digital Object Identifier (DOI) der konkreten Version des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via www.doi.org immer diese konkrete Version des Datensatzes abgerufen werden. Prinzip F1 der FAIR-Data Prinzipien («data are assigned globally unique and persistent identifiers») empfiehlt die Dokumentation jeder Messung mit einem persistenten Identifikator. Selbst wenn die CSV-Dateien ohne Kontext weitergegeben werden kann ihre Herkunft so immer zweifelsfrei und maschinenlesbar bestimmt werden.

6 Registerzeichen

Die Tabelle der Registerzeichen und der ihnen zugeordneten Verfahrensarten stammt aus dem folgenden Datensatz: »Seán Fobbe (2021). Aktenzeichen der Bundesrepublik Deutschland (AZ-BRD). Version 1.0.1. Zenodo. DOI: 10.5281/zenodo.4569564.«

Registerzeichen	Verfahrensart
A	Erstinstanzliche Klage, inklusive Wiederaufnahmeverfahren (Verwaltungsstreitsachen)
AR	Allgemeines Register (Vorverfahren oder sonstige Verfahrensarten)
AV	Allgemeine Verfahren (sonstige Anträge außerhalb eines schwebenden Verfahrens)
B	Nichtzulassungsbeschwerde oder Beschwerde (Verwaltungsstreitsachen)
BN	Nichtzulassungsbeschwerde (Normenkontrollsachen)
C	Revision (Verwaltungsstreitsachen)
CN	Revision (Normenkontrollsachen)
D	Revision (Disziplinarsachen)
D-AV	Allgemeines Register: Vorverfahren oder sonstige Verfahrensarten (Disziplinarsachen)
D-PKH	Prozesskostenhilfe (Disziplinarsachen)
D-St	Angeforderte Stellungnahmen der Senate des Bundesverwaltungsgerichts (Disziplinarsachen)
DB	Beschwerde (Disziplinarsachen)
DW	Wiederaufnahme (Disziplinarsachen)
ER	Sonstige Eingänge
F	Fachsenat (Verwaltungsstreitsachen)
GrSen	Großer Senat des Bundesverwaltungsgerichts
KSt	Kostensachen
P	Rechtsbeschwerde (Personalvertretungs- und Richtervertretungssachen)
PB	Nichtzulassungsbeschwerde (Personalvertretungs- und Richtervertretungssachen)
PKH	Prozesskostenhilfe
St	Angeforderte Stellungnahmen der Senate des Bundesverwaltungsgerichts
VR	Vorläufiger Rechtsschutz
WA	Erstinstanzliche Klage, inklusive Wiederaufnahmeverfahren (Wehrdienstsenat)
WB	Verfahren nach Wehrbeschwerdeordnung
WBW	Wiederaufnahme, Wehrbeschwerdeordnung

WD	Berufung (Wehrdisziplinarsachen)
WDB	Rechtsmittelverfahren über Anträge, Beschwerden und Vorlagen nach der Wehrdisziplinarordnung
WDS-AV	Allgemeine Verfahren (Wehrdienstsenate)
WDS-KSt	Kostensachen (Wehrdienstsenate)
WDS-PKH	Prozesskostenhilfe (Wehrdienstsenate)
WDS-St	Angeforderte Stellungnahmen der Senate des Bundesverwaltungsgerichts (Wehrdienstsenate)
WDS-VR	Vorläufiger Rechtsschutz (Wehrdienstsenate)
WDW	Wiederaufnahme (Wehrdisziplinarsachen)
WNB	Nichtzulassungsbeschwerde nach Wehrbeschwerdeordnung
WRB	Rechtsbeschwerde nach Wehrbeschwerdeordnung

7 Präsident:innen

7.1 Hinweise

- 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«.
- Das Datum bezieht sich jeweils auf das Amt als Präsident:in, nicht auf die Amtszeit als Richter:in.

7.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Frege	Ludwig	1953-03-28	1954-12-31	1884-08-28	1964-03-25
VACANCY-1	VACANCY-1	1955-01-01	1955-04-28	NA	NA
Egidi	Hans	1955-04-29	1958-06-30	1890-06-02	1970-12-03
VACANCY-2	VACANCY-2	1958-07-01	1958-07-17	NA	NA
Werner	Fritz	1958-07-18	1969-12-26	1906-05-04	1969-12-26
VACANCY-3	VACANCY-3	1969-12-27	1970-06-14	NA	NA
Zeidler	Wolfgang	1970-06-15	1975-11-07	1924-09-02	1987-12-31
VACANCY-4	VACANCY-4	1975-11-08	1976-08-18	NA	NA
Fürst	Walther	1976-08-19	1980-02-29	1912-02-10	2009-10-23
Sendler	Horst	1980-03-01	1991-06-30	1925-06-17	2006-01-13
Franßen	Everhardt	1991-07-01	2002-09-30	1937-10-01	NA
Hien	Eckart	2002-10-01	2007-05-31	1942-05-13	NA
Eckertz-Höfer	Marion	2007-06-01	2014-01-31	1948-11-23	NA
VACANCY-5	VACANCY-5	2014-02-01	2014-06-30	NA	NA
Rennert	Klaus	2014-07-01	NA	1955-09-24	NA

7.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtesantritt)	Alter (Amtesende)	Alter (Tod)
Frege	Ludwig	68	70	79
Egidi	Hans	64	68	80
Werner	Fritz	52	63	63
Zeidler	Wolfgang	45	51	63
Fürst	Walther	64	68	97
Sendler	Horst	54	66	80
Franßen	Everhardt	53	64	NA
Hien	Eckart	60	65	NA
Eckertz-Höfer	Marion	58	65	NA
Rennert	Klaus	58	NA	NA

8 Vize-Präsident:innen

8.1 Hinweise

- 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«.
- Das Datum bezieht sich jeweils auf das Amt als Vize-Präsident:in, nicht auf die Amtszeit als Richter:in.

8.2 Lebensdaten

Nachname	Vorname	Amtsantritt	Amtsende	Geboren	Gestorben
Külz	Helmut R.	1970-12-23	1971-07-31	1903-07-27	1985-09-24
VACANCY-1	VACANCY-1	1971-08-01	1971-11-15	NA	NA
Fürst	Walther	1971-11-16	1976-08-18	1912-02-10	2009-10-23
Sendler	Horst	1976-08-19	1980-02-29	1925-06-17	2006-01-13
Oppenheimer	Johannes	1980-03-01	1986-07-31	1918-07-10	2007-01-13
Zehner	Günter	1986-08-01	1990-08-31	1923-08-29	2002-07-21
Schlichter	Otto	1990-09-01	1993-09-30	1930-07-14	2011-03-19
Franke	Ingeborg	1993-10-01	2000-05-31	1935-05-21	NA
VACANCY-2	VACANCY-2	2000-06-01	2000-06-21	NA	NA
Hien	Eckart	2000-06-22	2002-09-30	1942-05-13	NA
Eckertz-Höfer	Marion	2002-10-01	2007-05-31	1948-11-23	NA
Hund	Michael	2007-06-01	2011-10-31	NA	NA
VACANCY-3	VACANCY-3	2011-11-01	2012-11-20	NA	NA
Rennert	Klaus	2012-11-21	2014-06-30	1955-09-24	NA
Christ	Josef	2014-07-01	2017-11-30	NA	NA
VACANCY-4	VACANCY-4	2017-12-01	2019-05-21	NA	NA
Korbmacher	Andreas	2019-05-22	NA	NA	NA

8.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtesantritt)	Alter (Amtesende)	Alter (Tod)
Külz	Helmut R.	67	68	82
Fürst	Walther	59	64	97
Sendler	Horst	51	54	80
Oppenheimer	Johannes	61	68	88
Zehner	Günter	62	67	78
Schlichter	Otto	60	63	80
Franke	Ingeborg	58	65	NA
Hien	Eckart	58	60	NA
Eckertz-Höfer	Marion	53	58	NA
Hund	Michael	NA	NA	NA
Rennert	Klaus	57	58	NA
Christ	Josef	NA	NA	NA
Korbmacher	Andreas	NA	NA	NA

9 Linguistische Kennzahlen

9.1 Erläuterung der Kennzahlen und Diagramme

Zur besseren Einschätzung des inhaltlichen Umfangs des Korpus dokumentiere ich an dieser Stelle die Verteilung der Werte für einige klassische linguistische Kennzahlen.

Kennzahl	Definition
Zeichen	Zeichen entsprechen grob den <i>Graphemen</i> , den kleinsten funktionalen Einheiten in einem Schriftsystem. Beispiel: das Wort »RichterIn« besteht aus 9 Zeichen.
Tokens	Eine beliebige Zeichenfolge, getrennt durch whitespace-Zeichen, d.h. ein Token entspricht in der Regel einem »Wort«, kann aber auch Zahlen, Sonderzeichen oder sinnlose Zeichenfolgen enthalten, weil es rein syntaktisch berechnet wird.
Typen	Einzigartige Tokens. Beispiel: wenn das Token »Verfassungsrecht« zehnmal in einer Entscheidung vorhanden ist, wird es als ein Typ gezählt.
Sätze	Entsprechen in etwa dem üblichen Verständnis eines Satzes. Die Regeln für die Bestimmung von Satzanfang und Satzende sind im Detail aber sehr komplex und in »Unicode Standard: Annex No 29« beschrieben.

Es handelt sich bei den Diagrammen jeweils um »Density Charts«, die sich besonders dafür eignen die Schwerpunkte von Variablen mit stark schwankenden numerischen Werten zu visualisieren. Die Interpretation ist denkbar einfach: je höher die Kurve, desto dichter sind in diesem Bereich die Werte der Variable. Der Wert der y-Achse kann außer Acht gelassen werden, wichtig sind nur die relativen Flächenverhältnisse und die x-Achse.

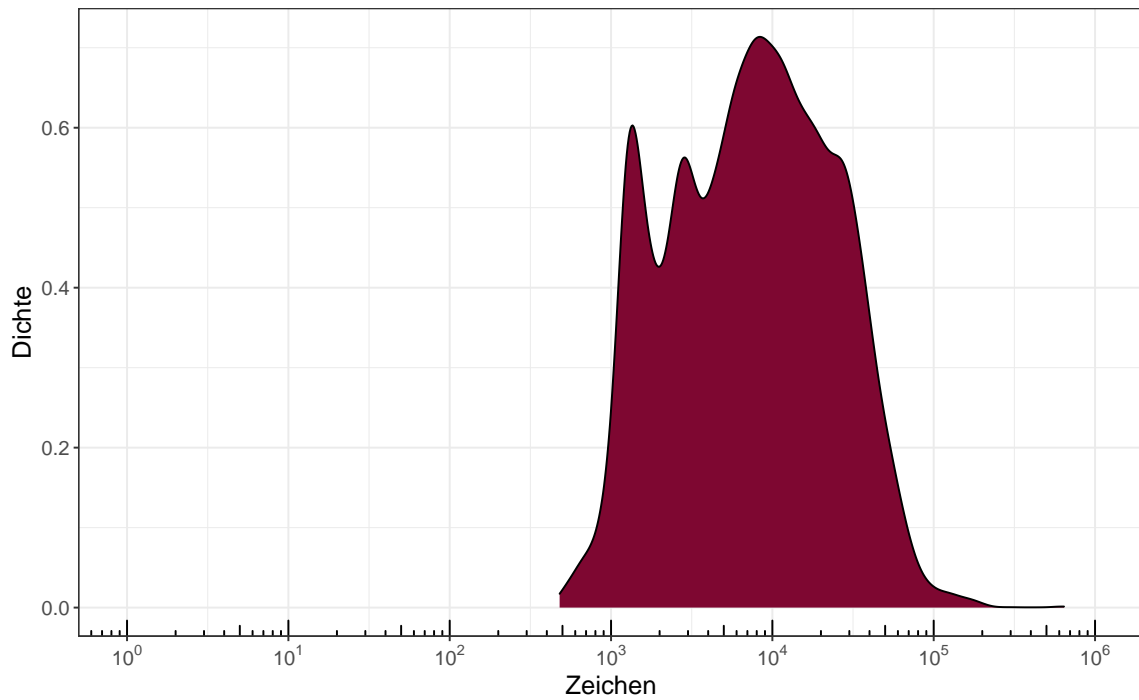
Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

9.2 Werte der Kennzahlen

Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	338,527,212	480	2,851	7,626	13,572.03	17,877	642,248
tokens	52,027,453	62	410	1,163	2,085.85	2,753	100,685
saetze	2,991,595	4	30	70	119.94	156	5,000
typen	NA	47	217	465	610.92	840	11,507

9.3 Verteilung Zeichen

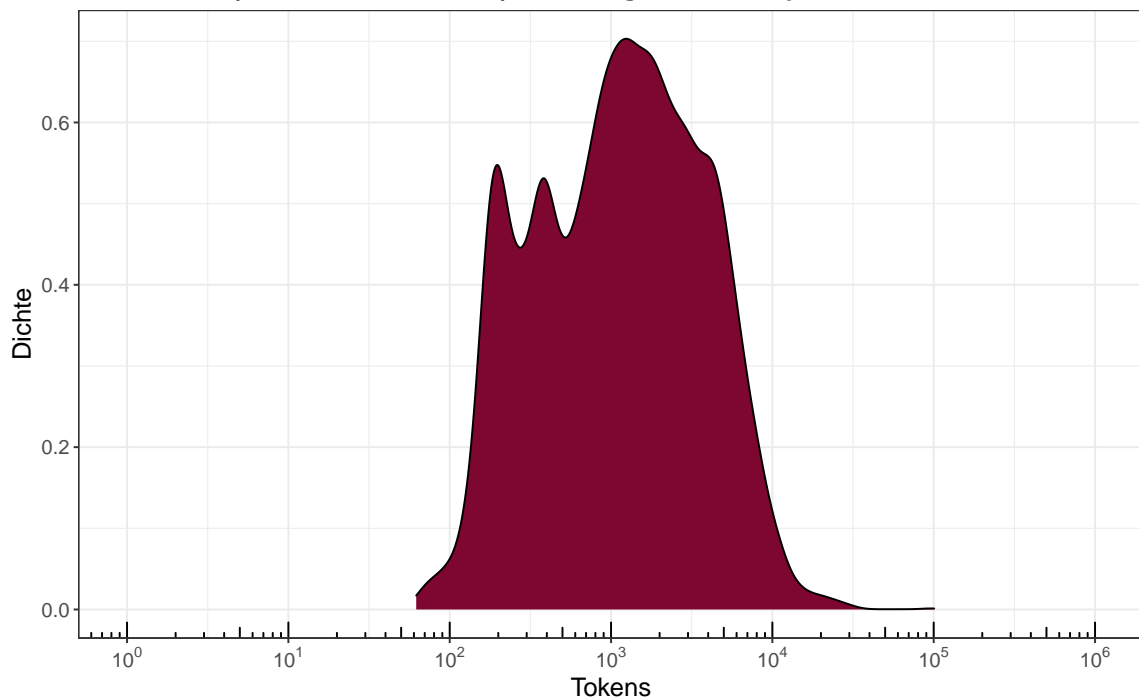
CE-BVerwG | Version 2021-04-15 | Verteilung der Zeichen je Dokument



DOI: 10.5281/zenodo.4625123

9.4 Verteilung Tokens

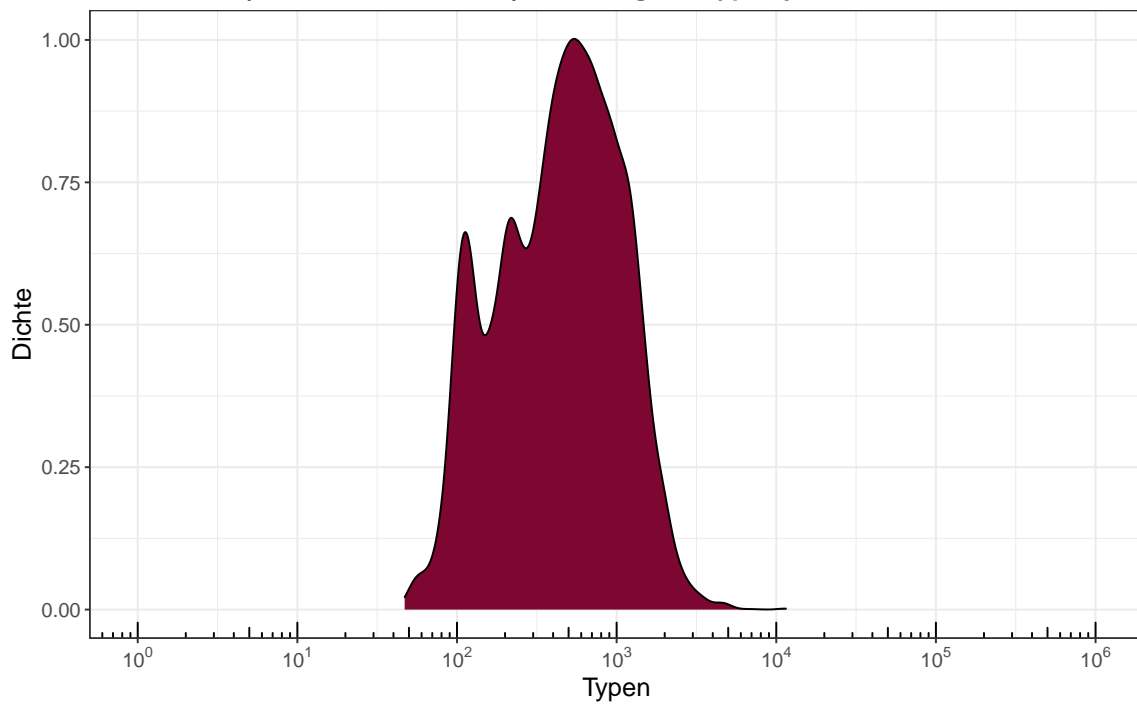
CE-BVerwG | Version 2021-04-15 | Verteilung der Tokens je Dokument



DOI: 10.5281/zenodo.4625123

9.5 Verteilung Typen

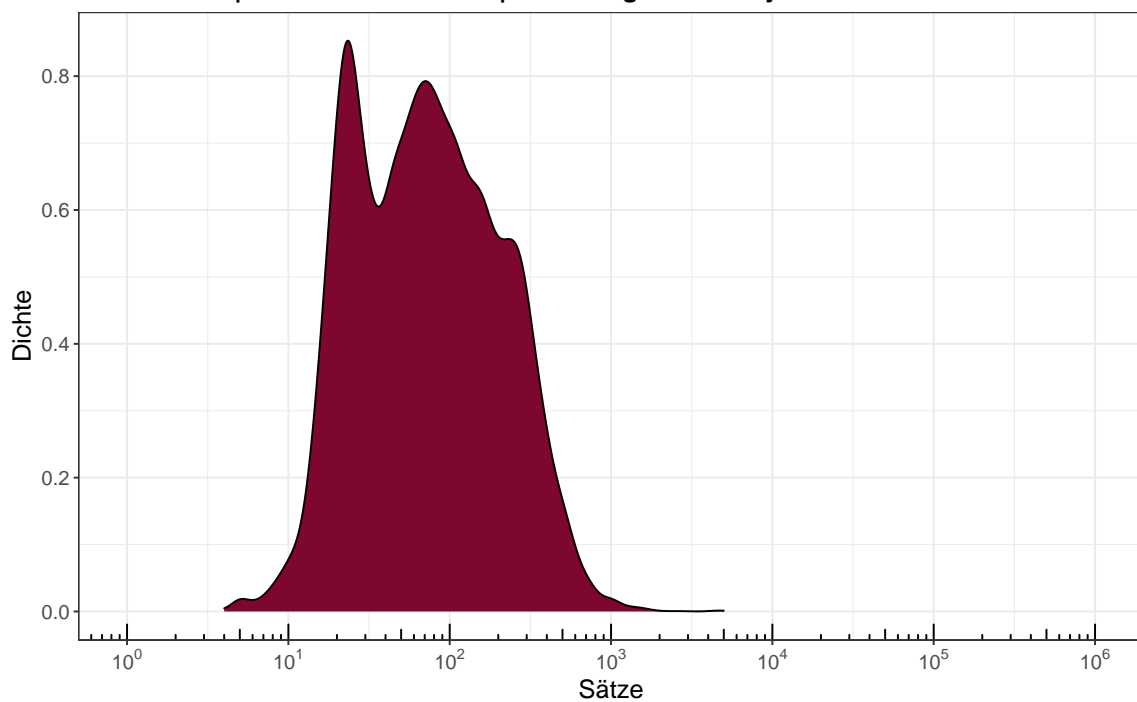
CE-BVerwG | Version 2021-04-15 | Verteilung der Typen je Dokument



DOI: 10.5281/zenodo.4625123

9.6 Verteilung Sätze

CE-BVerwG | Version 2021-04-15 | Verteilung der Sätze je Dokument



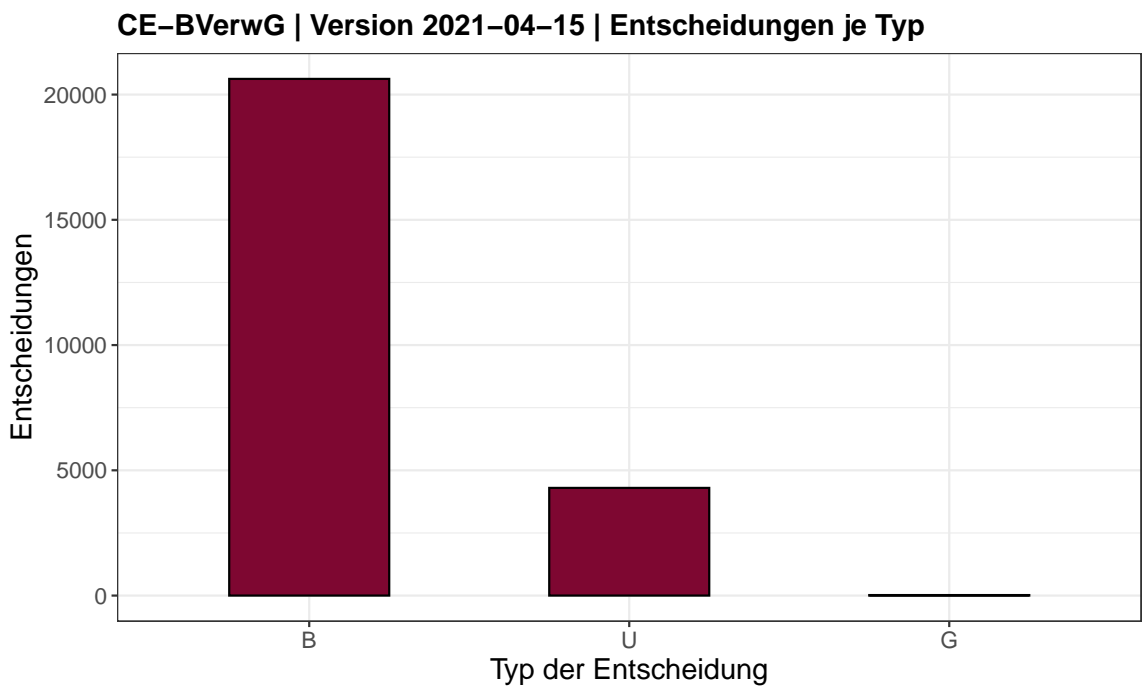
DOI: 10.5281/zenodo.4625123

10 Inhalt des Korpus

10.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mittel	Quart3	Max
entscheidungsjahr	22	1997	2005	2009	2009.62	2014	2021
eingangsjahr_iso	23	1996	2005	2008	2009.15	2013	2021
eingangsnummer	573	1	10	27	62.47	59	7005

10.2 Nach Typ der Entscheidung

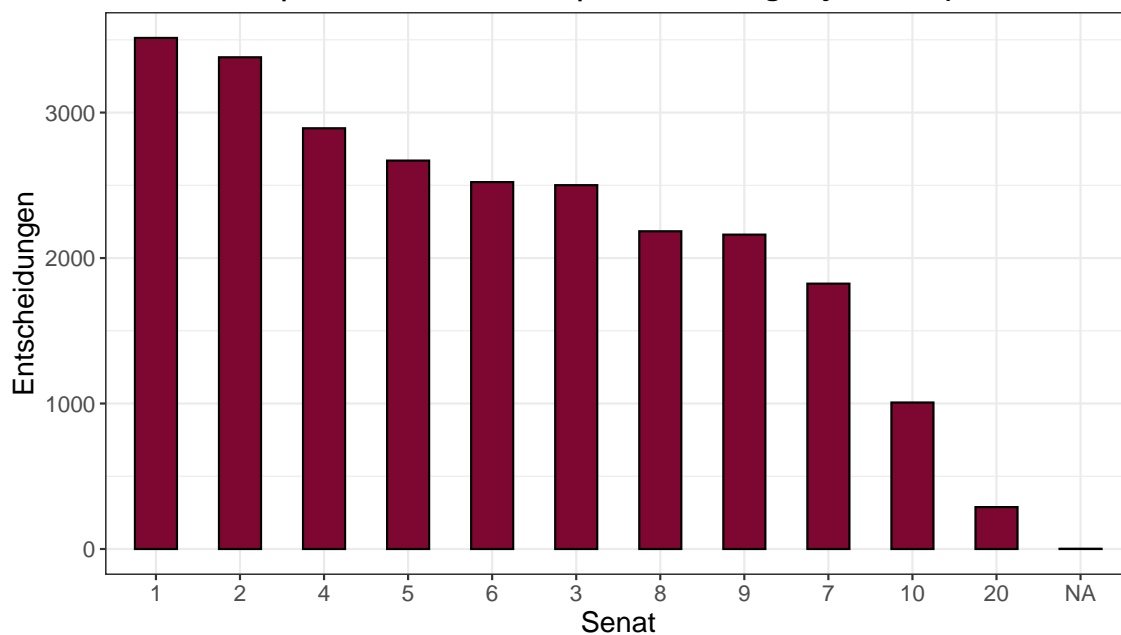


DOI: 10.5281/zenodo.4625123

Typ	Entscheidungen	% Gesamt	% Kumulativ
B	20625	82.69	82.69
G	20	0.08	82.77
U	4298	17.23	100.00
Total	24943	100.00	100.00

10.3 Nach Spruchkörper (Aktenzeichen)

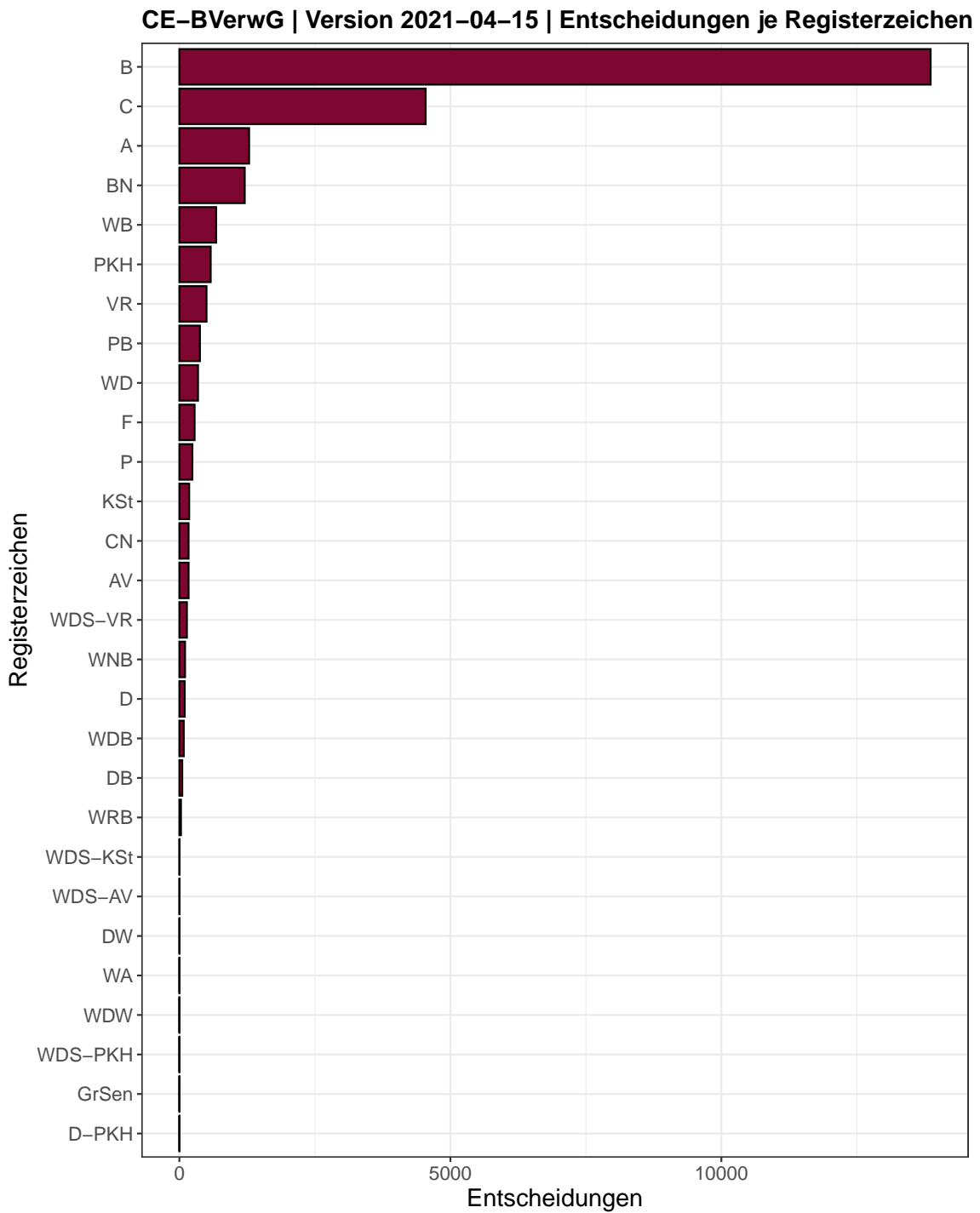
CE-BVerwG | Version 2021-04-15 | Entscheidungen je Senat (Aktenzeichen)



DOI: 10.5281/zenodo.4625123

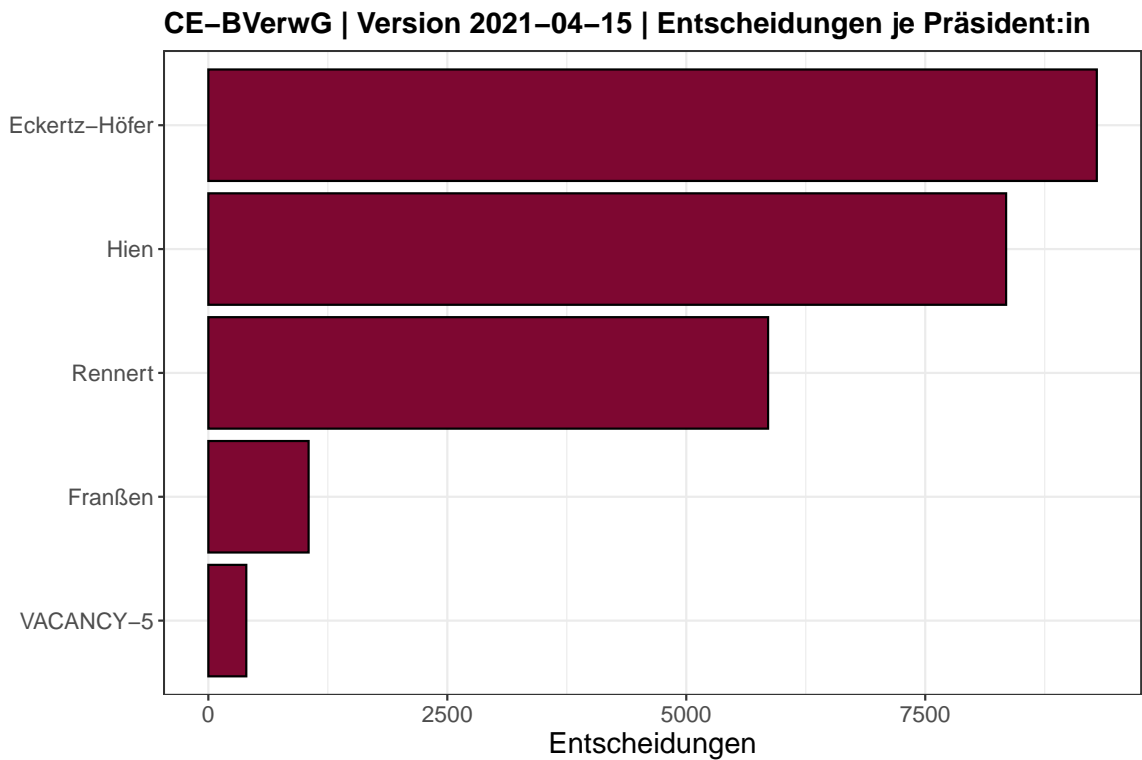
Senat	Entscheidungen	% Gesamt	% Kumulativ
1	3513	14.08	14.09
2	3380	13.55	27.64
3	2501	10.03	37.67
4	2892	11.59	49.26
5	2670	10.70	59.96
6	2522	10.11	70.08
7	1824	7.31	77.39
8	2184	8.76	86.14
9	2161	8.66	94.81
10	1007	4.04	98.85
20	288	1.15	100.00
NA	1	0.00	0.00
Total	24943	100.00	100.00

10.4 Nach Registerzeichen



Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
A	1287	5.16	5.16
AV	170	0.68	5.84
B	13861	55.57	61.41
BN	1206	4.84	66.25
C	4544	18.22	84.46
CN	170	0.68	85.15
D	99	0.40	85.54
D-PKH	1	0.00	85.55
DB	53	0.21	85.76
DW	3	0.01	85.77
F	281	1.13	86.90
GrSen	1	0.00	86.90
KSt	181	0.73	87.63
P	240	0.96	88.59
PB	379	1.52	90.11
PKH	578	2.32	92.43
VR	501	2.01	94.44
WA	2	0.01	94.44
WB	678	2.72	97.16
WD	344	1.38	98.54
WDB	81	0.32	98.87
WDS-AV	4	0.02	98.88
WDS-KSt	4	0.02	98.90
WDS-PKH	1	0.00	98.90
WDS-VR	139	0.56	99.46
WDW	1	0.00	99.46
WNB	105	0.42	99.88
WRB	29	0.12	100.00
Total	24943	100.00	100.00

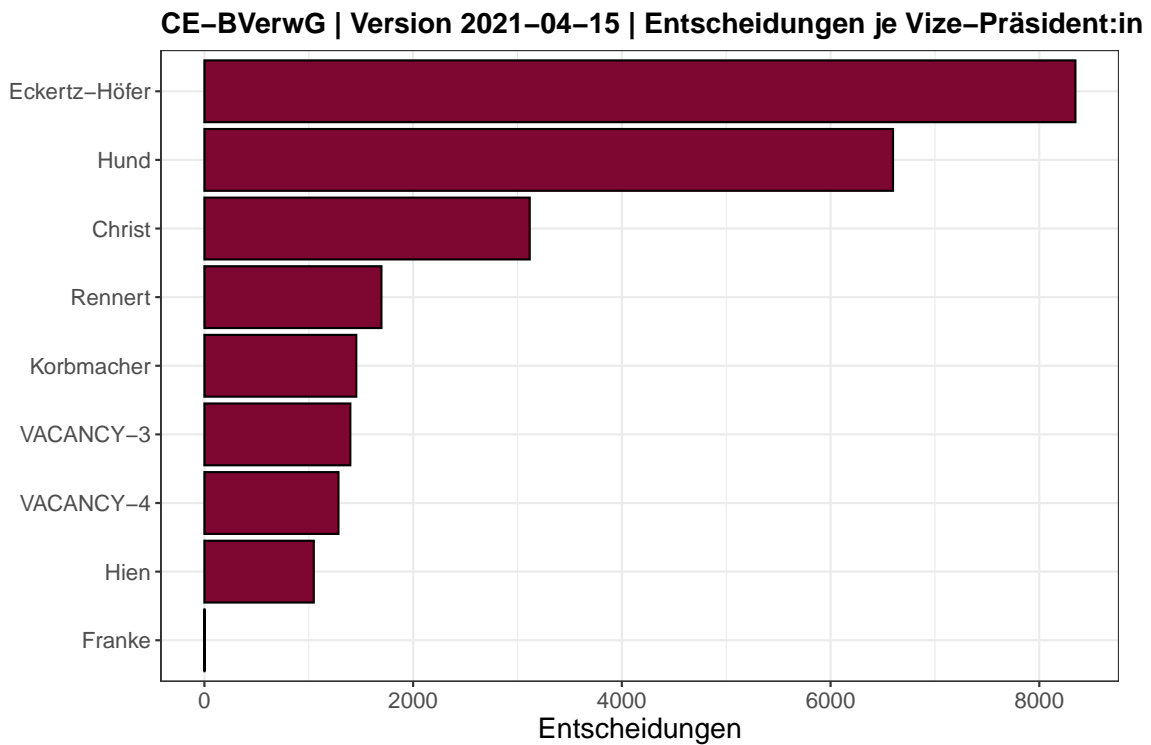
10.5 Nach Präsident:in



DOI: 10.5281/zenodo.4625123

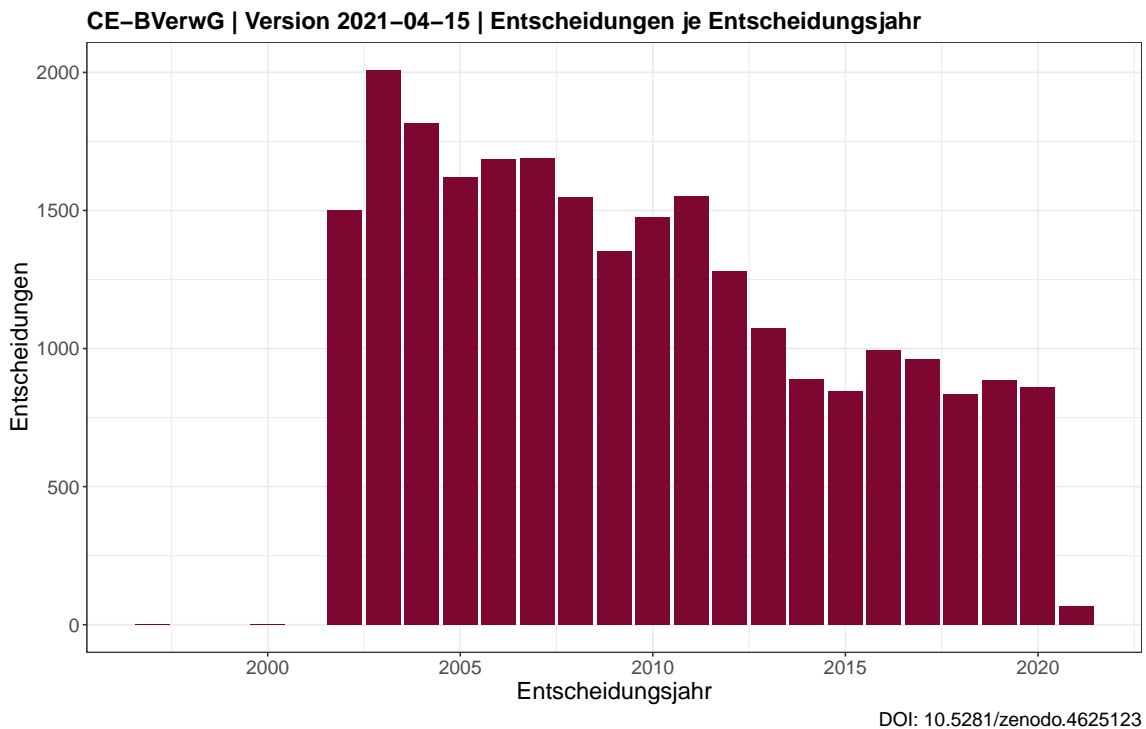
Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Eckertz-Höfer	9295	37.26	37.26
Franßen	1049	4.21	41.47
Hien	8346	33.46	74.93
Rennert	5856	23.48	98.41
VACANCY-5	397	1.59	100.00
Total	24943	100.00	100.00

10.6 Nach Vize-Präsident:in



Vize-Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Christ	3117	12.50	12.50
Eckertz-Höfer	8346	33.46	45.96
Franke	1	0.00	45.96
Hien	1048	4.20	50.16
Hund	6598	26.45	76.61
Korbmacher	1455	5.83	82.45
Rennert	1696	6.80	89.25
VACANCY-3	1398	5.60	94.85
VACANCY-4	1284	5.15	100.00
Total	24943	100.00	100.00

10.7 Nach Entscheidungsjahr

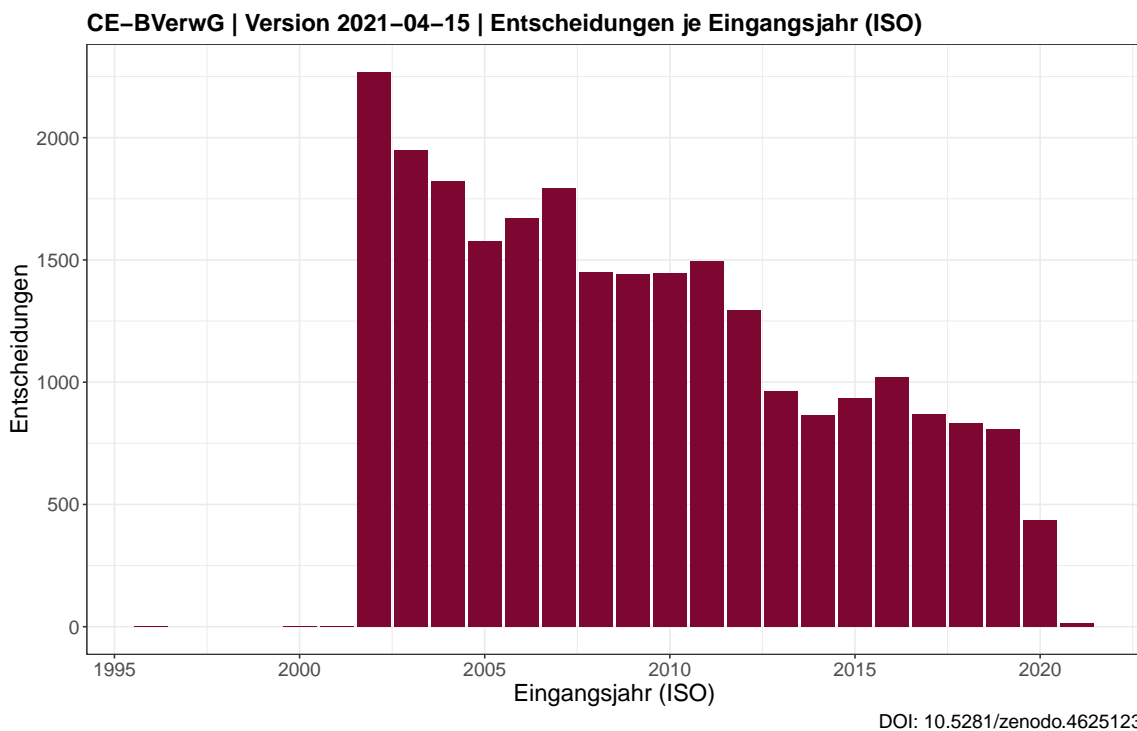


Jahr	Entscheidungen	% Gesamt	% Kumulativ
1997	1	0.00	0.00
2000	1	0.00	0.01
2002	1502	6.02	6.03
2003	2009	8.05	14.08
2004	1817	7.28	21.37
2005	1619	6.49	27.86
2006	1684	6.75	34.61
2007	1690	6.78	41.39
2008	1549	6.21	47.60
2009	1352	5.42	53.02
2010	1474	5.91	58.93
2011	1553	6.23	65.15
2012	1281	5.14	70.29
2013	1072	4.30	74.59
2014	889	3.56	78.15

(continued)

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2015	845	3.39	81.54
2016	995	3.99	85.53
2017	960	3.85	89.38
2018	836	3.35	92.73
2019	887	3.56	96.28
2020	859	3.44	99.73
2021	68	0.27	100.00
Total	24943	100.00	100.00

10.8 Nach Eingangsjahr (ISO)



Jahr	Entscheidungen	% Gesamt	% Kumulativ
1996	1	0.00	0.00
2000	1	0.00	0.01
2001	2	0.01	0.02
2002	2269	9.10	9.11
2003	1948	7.81	16.92
2004	1822	7.30	24.23
2005	1575	6.31	30.54
2006	1670	6.70	37.24
2007	1794	7.19	44.43
2008	1450	5.81	50.24
2009	1440	5.77	56.02
2010	1445	5.79	61.81
2011	1493	5.99	67.79
2012	1293	5.18	72.98
2013	964	3.86	76.84

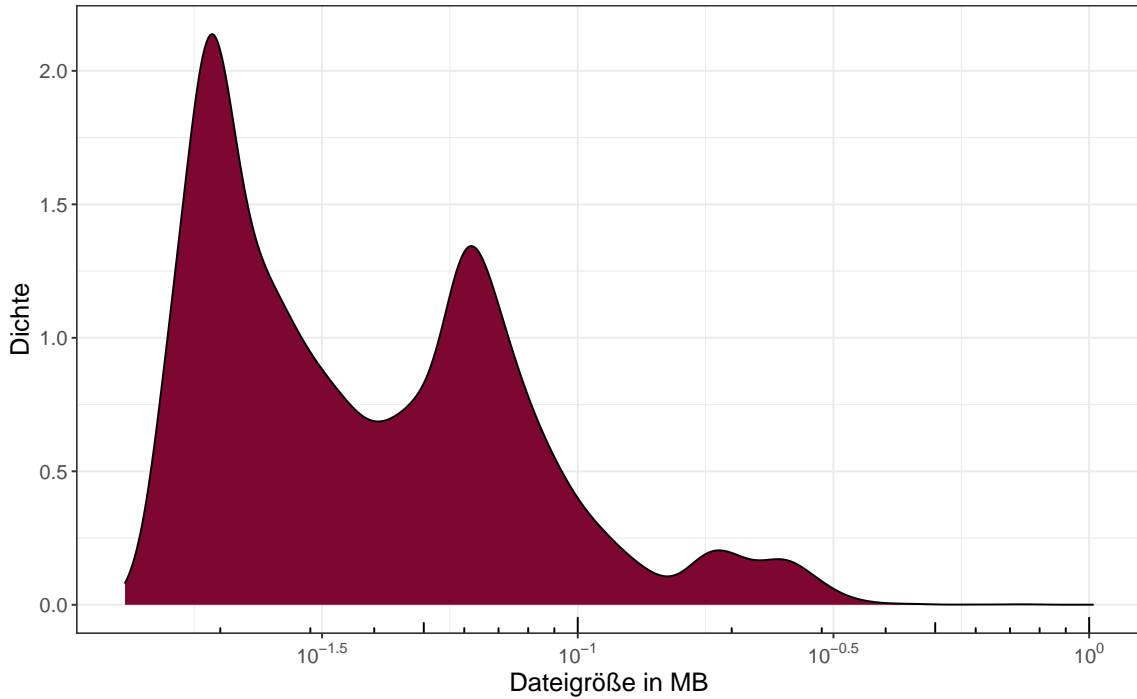
(continued)

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2014	864	3.46	80.31
2015	936	3.75	84.06
2016	1020	4.09	88.15
2017	869	3.48	91.63
2018	832	3.34	94.97
2019	806	3.23	98.20
2020	436	1.75	99.95
2021	13	0.05	100.00
Total	24943	100.00	100.00

11 Dateigrößen

11.1 Verteilung PDF

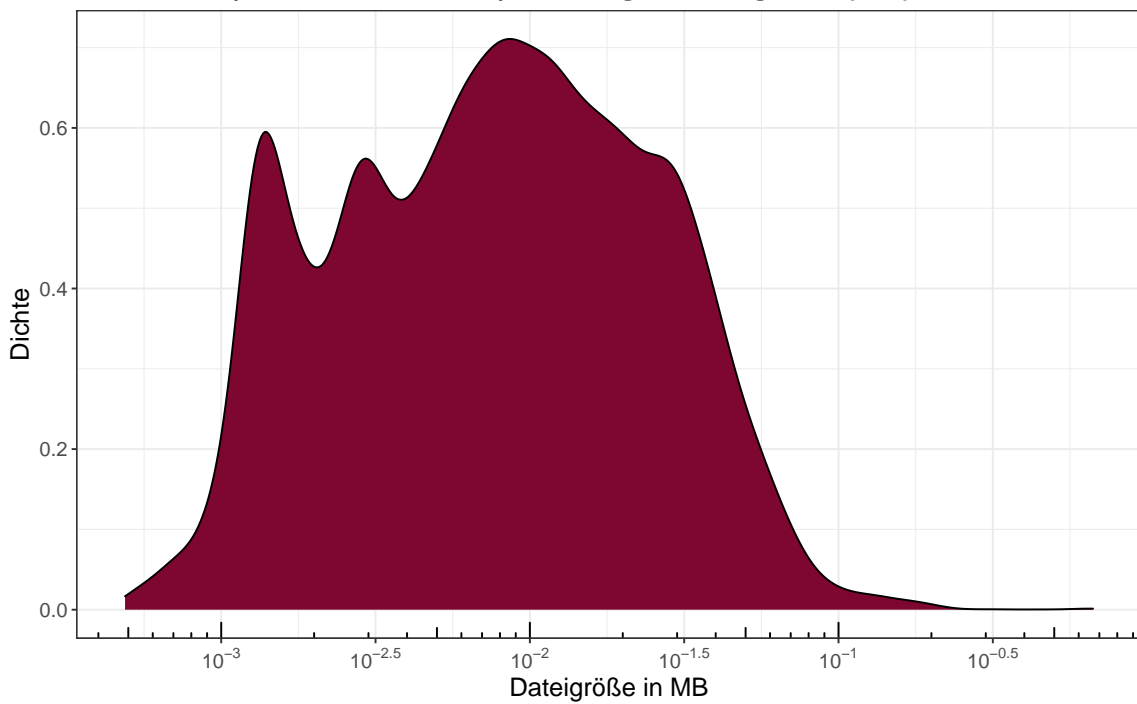
CE-BVerwG | Version 2021-04-15 | Verteilung der Dateigrößen (PDF)



DOI: 10.5281/zenodo.4625123

11.2 Verteilung TXT

CE-BVerwG | Version 2021-04-15 | Verteilung der Dateigrößen (TXT)



DOI: 10.5281/zenodo.4625123

11.3 Gesamtgröße je ZIP-Archiv

Datei	Größe in MB
CE-BVerwG_2021-04-15_DE_ANALYSE.zip	1.32
CE-BVerwG_2021-04-15_DE_CSV_Datensatz.zip	91.67
CE-BVerwG_2021-04-15_DE_CSV_Metadaten.zip	0.79
CE-BVerwG_2021-04-15_DE_PDF_Datensatz.zip	955.47
CE-BVerwG_2021-04-15_DE_TXT_Datensatz.zip	126.92
CE-BVerwG_2021-04-15_Source_Files.zip	0.05

12 Prüfung kryptographischer Signaturen

12.1 Allgemeines

Die Integrität und Echtheit der einzelnen Archive des Datensatzes sind durch eine Zwei-Phasen-Signatur sichergestellt.

In **Phase I** werden während der Kompilierung für jedes ZIP-Archiv Hash-Werte in zwei verschiedenen Verfahren (SHA2-256 und SHA3-512) berechnet und in einer CSV-Datei dokumentiert.

In **Phase II** wird diese CSV-Datei mit meinem persönlichen geheimen GPG-Schlüssel signiert. Dieses Verfahren stellt sicher, dass die Kompilierung von jedermann durchgeführt werden kann, insbesondere im Rahmen von Replikationen, die persönliche Gewähr für Ergebnisse aber dennoch vorhanden bleibt.

Dieses Codebook ist vollautomatisch erstellt und prüft die kryptographisch sicheren SHA3-512 Signaturen (»hashes«) aller ZIP-Archive, sowie die GPG-Signatur der CSV-Datei, welche die SHA3-512 Signaturen enthält. SHA3-512 Signaturen werden durch einen system call zur OpenSSL library auf Linux-Systemen berechnet. Eine erfolgreiche Prüfung meldet »Signatur verifiziert!«. Eine gescheiterte Prüfung meldet »FEHLER!«

12.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen ist mit meiner persönlichen GPG-Signatur versehen. Der mit dieser Version korrespondierende Public Key ist sowohl mit dem Datensatz als auch mit dem Source Code hinterlegt. Er hat folgende Kenndaten:

Name: Sean Fobbe (fobbe-data@posteo.de)

Fingerabdruck: FE6F B888 F0E5 656C 1D25 3B9A 50C4 1384 F44A 4E42

12.3 Import: Public Key

```
system2("gpg2", "--import GPG-Public-Key_Fobbe-Data.asc",
        stdout = TRUE,
        stderr = TRUE)
```

```
## [1] "gpg: key 50C41384F44A4E42: \"Sean Fobbe <fobbe-data@posteo.de>\" not
      changed"
## [2] "gpg: Total number processed: 1"
## [3] "gpg:                unchanged: 1"
```

12.4 Prüfung: GPG-Signatur der Hash-Datei

```
# CSV-Datei mit Hashes  
print(hashfile)
```

```
## [1] "CE-BVerwG_2021-04-15_KryptographischeHashes.csv"
```

```
# GPG-Signatur  
print(signaturefile)
```

```
## [1] "CE-BVerwG_2021-04-15_FobbeSignaturGPG_Hashes.gpg"
```

```
# GPG-Signatur prüfen  
testresult <- system2("gpg2",  
                      paste("--verify", signaturefile, hashfile),  
                      stdout = TRUE,  
                      stderr = TRUE)  
  
# Anführungsstriche entfernen um Anzeigefehler zu vermeiden  
testresult <- gsub("'", "", testresult)
```

```
kable(testresult, format = "latex", booktabs = TRUE,  
      longtable = TRUE, col.names = c("Ergebnis"))
```

Ergebnis

gpg: Signature made Thu 15 Apr 2021 12:19:05 PM CEST

gpg: using RSA key FE6FB888F0E5656C1D253B9A50C41384F44A4E42

gpg: Good signature from Sean Fobbe <fobbe-data@posteo.de> [full]

12.5 Prüfung: SHA3-512 Hashes der ZIP-Archive

```
# Prüf-Funktion definieren
sha3test <- function(filename, sig){
  sig.new <- system2("openssl",
                    paste("sha3-512", filename),
                    stdout = TRUE)
  sig.new <- gsub("^.*\\|= ", "", sig.new)
  if (sig == sig.new){
    return("Signatur verifiziert!")
  }else{
    return("FEHLER!")
  }
}

# Ursprüngliche Signaturen importieren
table.hashes <- fread(hashfile)
filename <- table.hashes$filename
sha3.512 <- table.hashes$sha3.512

# Signaturprüfung durchführen
sha3.512.result <- mcmapply(sha3test, filename, sha3.512, USE.NAMES = FALSE)

# Ergebnis anzeigen
testresult <- data.table(filename, sha3.512.result)
```

```
kable(testresult, format = "latex", booktabs = TRUE,
      longtable = TRUE, col.names = c("Datei", "Ergebnis"))
```

Datei	Ergebnis
CE-BVerwG_2021-04-15_DE_ANALYSE.zip	Signatur verifiziert!
CE-BVerwG_2021-04-15_DE_CSV_Datensatz.zip	Signatur verifiziert!
CE-BVerwG_2021-04-15_DE_CSV_Metadaten.zip	Signatur verifiziert!
CE-BVerwG_2021-04-15_DE_PDF_Datensatz.zip	Signatur verifiziert!
CE-BVerwG_2021-04-15_DE_TXT_Datensatz.zip	Signatur verifiziert!
CE-BVerwG_2021-04-15_Source_Files.zip	Signatur verifiziert!

13 Changelog

Version	Details
2021-04-15	<ul style="list-style-type: none">• Vollständige Aktualisierung der Daten• Veröffentlichung des vollständigen Source Codes• Deutliche Erweiterung des inhaltlichen Umfangs des Codebooks• Einführung der vollautomatischen Erstellung von Datensatz und Codebook• Einführung von Compilation Reports um den Erstellungsprozess exakt zu dokumentieren• Einführung von Variablen für Versionsnummer, Concept DOI, Version DOI, ECLI, Präsident:in, Vize-Präsident:in, Verfahrensart und linguistische Kennzahlen (Zeichen, Tokens, Typen, Sätze)• Zusammenfügung von über Zeilengrenzen getrennten Wörtern• Automatisierung und Erweiterung der Qualitätskontrolle• Einführung von Diagrammen zur Visualisierung von Prüfergebnissen• Einführung kryptographischer Signaturen• Variable »Suffix« in »kollision« umbenannt.• Variable »Ordinalzahl« in »eingangsnummer« umbenannt.• Variable »Entscheidungsart« in »entscheidung_typ« umbenannt.• Alle übrigen Variablen sind nun in Kleinschreibung und Snake Case gehalten
2020-06-23	<ul style="list-style-type: none">• Erstveröffentlichung

14 Parameter für strenge Replikationen

```
## [1] "OpenSSL 1.1.1k FIPS 25 Mar 2021"
```

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 33 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libopenblas-r0.3.12.so
##
## 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] magick_2.7.1      quanteda_2.1.2      readtext_0.80      data.table_1.14.0
## [5] scales_1.1.1      ggplot2_3.3.3      doParallel_1.0.16  iterators_1.0.13
## [9] foreach_1.5.1    pdftools_2.3.1     kableExtra_1.3.4   knitr_1.31
## [13] rvest_1.0.0       httr_1.4.2
##
## loaded via a namespace (and not attached):
## [1] tinytex_0.30      qpdf_1.1           tidyselect_1.1.0   xfun_0.22
## [5] purrr_0.3.4       lattice_0.20-41    colorspace_2.0-0   vctrs_0.3.6
## [9] generics_0.1.0    htmltools_0.5.1.1 viridisLite_0.3.0  yaml_2.2.1
## [13] utf8_1.2.1        rlang_0.4.10      pillar_1.5.1       glue_1.4.2
## [17] withr_2.4.1       selectr_0.4-2     lifecycle_1.0.0    stringr_1.4.0
## [21] munsell_0.5.0     gtable_0.3.0      codetools_0.2-18  evaluate_0.14
## [25] labeling_0.4.2    curl_4.3           fansi_0.4.2        highr_0.8
## [29] Rcpp_1.0.6        RcppParallel_5.0.3 webshot_0.5.2      farver_2.1.0
## [33] systemfonts_1.0.1 fastmatch_1.1-0    stopwords_2.2      askpass_1.1
## [37] digest_0.6.27     stringi_1.5.3     dplyr_1.0.5        grid_4.0.4
## [41] tools_4.0.4       magrittr_2.0.1    tibble_3.1.0       crayon_1.4.1
## [45] pkgconfig_2.0.3   Matrix_1.3-2      ellipsis_0.3.1     xml2_1.3.2
## [49] rmarkdown_2.7     svglite_2.0.0     rstudioapi_0.13    R6_2.5.0
## [53] compiler_4.0.4
```

Literaturverzeichnis

- Analytics, Revolution, and Steve Weston. 2020. *Iterators: Provides Iterator Construct*. <https://github.com/RevolutionAnalytics/iterators>.
- Benoit, Kenneth, and Adam Obeng. 2020. *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, Jiong Wei Lua, Jouni Kuha, and William Lowe. 2020. *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>.
- Ooms, Jeroen. 2020. *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. 2020. *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.crepress.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>.