

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

CODEBOOK

Version 2024-03-13



DOI: 10.5281/zenodo.10809039

Titel	Corpus der Entscheidungen des Bundesverwaltungsgerichts
Abkürzung	CE-BVerwG
Autor	Seán Fobbe
Version	2024-03-13
Download	https://doi.org/10.5281/zenodo.10809039
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Seán Fobbe (2024). Corpus der Entscheidungen des Bundesverwaltungsgerichts (CE-BVerwG). Version 2024-03-13. Zenodo. DOI: 10.5281/zenodo.10809039.

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 2024-03-13. 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	9
3.6	Urheberrechtsfreiheit von Rohdaten und Datensatz	9
3.7	Metadaten	9
3.7.1	Allgemein	9
3.7.2	Schema für die Dateinamen	10
3.7.3	Beispiel eines Dateinamens	10
3.8	Qualitätsprüfung	10
3.9	Grafische Darstellung	10
4	Varianten und Zielgruppen	11
5	Variablen (Allgemein)	13
5.1	Datenstruktur	13
5.2	Hinweise	14
5.3	Erläuterung der Variablen	14
6	Variablen (Linguistische Annotationen)	19
6.1	Datenstruktur	19
6.2	Hinweise	19
6.3	Erläuterung der Variablen	20
7	Registerzeichen	21
8	Präsident:innen	23
8.1	Hinweise	23
8.2	Lebensdaten	23
8.3	Dienstalter und Lebensalter	24
9	Vize-Präsident:innen	25
9.1	Hinweise	25
9.2	Lebensdaten	25
9.3	Dienstalter und Lebensalter	26
10	Linguistische Kennzahlen	27
10.1	Erläuterung der Kennzahlen und Diagramme	27
10.2	Werte der Kennzahlen	27
10.3	Verteilung Zeichen	28
10.4	Verteilung Tokens	28

10.5	Verteilung Typen	29
10.6	Verteilung Sätze	29
11	Inhalt des Korpus	30
11.1	Zusammenfassung	30
11.2	Nach Typ der Entscheidung	30
11.3	Nach Spruchkörper (Aktenzeichen)	31
11.4	Nach Registerzeichen	32
11.5	Nach Präsident:in	34
11.6	Nach Vize-Präsident:in	35
11.7	Nach Entscheidungsjahr	36
11.8	Nach Eingangsjahr (ISO)	38
12	Dateigrößen	40
13	Kryptographische Signaturen	41
13.1	Zwei-Phasen-Signatur	41
13.2	Persönliche GPG-Signatur	41
14	Changelog	42
14.1	Version 2024-03-13	42
14.2	Version 2023-03-21	42
14.3	Version 2022-08-07	42
14.4	Version 2021-10-19	43
14.5	Version 2021-04-15	43
14.6	Version 2020-06-23	43
15	Parameter für strenge Replikationen	44
	Literaturverzeichnis	46

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 BVerwG eingerichtet sind derzeit 13 reguläre Senate, davon zehn Revisionsenate, 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 BVerwG erste und letzte Instanz (§ 50 VwGO).

Wieso dieser Datensatz? Die quantitative Analyse von juristischen Texten, insbesondere denen des BVerwG, 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 2022. <<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/>
⁶ <https://opinioiuris.de/>

⁷ <https://openjur.de/>

⁸ Openjur beabsichtigt eine API anzubieten, diese war 2024 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** 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 Beispiel:

```
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.10809041
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 (2023). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2023-03-21. Zenodo. <https://github.com/SeanFobbe/pvp-fcg>«.

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](https://doi.org/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.10809041>

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 kommentierten Source Code für die Daten-Pipeline, 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 BVerwG einigermaßen vollständig veröffentlicht (*temporal bias*). Die Frequenztabellen geben hierzu genauer Auskunft.
5. Folgende Entscheidungen veröffentlicht das BVerwG in der Regel *nicht*: Einstellungsbeschlüsse, Ruhensbeschlüsse, Entscheidungen über Prozesskostenhilfe, Beiordnungsbeschlüsse, Streitwertbeschlüsse, Kostenfestsetzungsbeschlüsse, Beiladungen, Anhörungsrügen, Vergleiche, Verwerfungen von Beschwerden zum Bundesverwaltungsgericht nach § 152 VwGO, Entscheidungen, die dem Geheimschutz unterliegen oder die durch die gesetzlich vorgeschriebene Anonymisierung unverständlich oder verfälscht werden.

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]_[verzoegerung]_[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 BVerwG 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 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. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Über Zeilenumbrüche getrennte Wörter wurden zusammengefügt.
CSV_Metadaten	Legal Tech/Quantitative Forschung. Wie die vorige CSV-Variante, nur ohne die Entscheidungstexte. Sinnvoll für Analyst:innen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung.
CSV_Annotiert	Legal Tech/Quantitative Forschung. Alle Entscheidungen in tokenisierter Form mit linguistischen Annotationen. Beachten Sie bitte die besondere Variablen-Struktur unter Punkt 6. Jede Spalte entspricht einer Variable, jede Zeile einem Token.

Variante	Zielgruppe und Beschreibung
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.
ANALYSE	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.

5 Variablen (Allgemein)

5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 27200 obs. of 30 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" ...
## $ url : chr "https://www.bverwg.de/entscheidungen/pdf/260297
U6C3.96.0.pdf" "https://www.bverwg.de/entscheidungen/pdf/141200B1WB107.00.0.
pdf" "https://www.bverwg.de/entscheidungen/pdf/100102B9A9.02.0.pdf" "https://
www.bverwg.de/entscheidungen/pdf/140102B4BN1.02.0.pdf" ...
## $ gericht : chr "BVerwG" "BVerwG" "BVerwG" "BVerwG" ...
## $ datum : IDate, format: "1997-02-26" "2000-12-14" ...
## $ entscheidung_typ : chr "U" "B" "B" "B" ...
## $ entscheidungsjahr: int 1997 2000 2002 2002 2002 2002 2002 2002 2002 2002
...
## $ spruchkoerper_az : int 6 1 9 4 1 1 1 1 8 8 ...
## $ registerzeichen : chr "C" "WB" "A" "BN" ...
## $ verfahrensart : chr "Revision (Verwaltungsstreitsachen)" "Verfahren
nach Wehrbeschwerdeordnung" "Erstinstanzliche Klage, inklusive
Wiederaufnahmeverfahren (Verwaltungsstreitsachen)" "Nichtzulassungsbeschwerde
(Normenkontrollsachen)" ...
## $ eingangsnummer : int 3 107 9 1 12 3 4 5 1 2 ...
## $ eingangsjahr_az : int 96 0 2 2 2 2 2 2 2 ...
## $ eingangsjahr_iso : num 1996 2000 2002 2002 2002 ...
## $ verzoegerung : chr NA NA NA NA ...
## $ kollision : int 0 0 0 0 0 0 0 0 0 ...
## $ aktenzeichen : chr "6 C 3.96" "1 WB 107.00" "9 A 9.02" "4 BN 1.02" ...
## $ ecli : 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" ...
## $ praesi : chr "Franßen" "Franßen" "Franßen" "Franßen" ...
## $ v_praesi : chr "Franke" "Hien" "Hien" "Hien" ...
## $ bverwge : num 0 0 0 0 0 0 0 0 0 ...
## $ leitsatz : num 0 0 0 0 0 0 0 0 0 ...
## $ fachpresse : num 0 0 0 0 0 0 0 0 0 ...
## $ uebersetzung : num 0 0 0 0 0 0 0 0 0 ...
## $ zeichen : int 37496 13855 826 3646 1401 1395 1402 1404 2470 2760
...
## $ tokens : int 5479 2336 128 557 213 213 213 213 377 418 ...
## $ typen : int 1465 721 84 277 128 128 126 128 224 227 ...
## $ saetze : int 266 143 15 48 18 18 18 18 21 27 ...
## $ version : chr "2024-03-13" "2024-03-13" "2024-03-13" "2024-03-13"
...
## $ 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.10809039" "10.5281/zenodo.10809039"
"10.5281/zenodo.10809039" "10.5281/zenodo.10809039" ...
## $ lizenz : chr "Creative Commons Zero 1.0 Universal" "Creative
Commons Zero 1.0 Universal" "Creative Commons Zero 1.0 Universal" "Creative
Commons Zero 1.0 Universal" ...
## - attr(*, ".internal.selfref")=<externalptr>
## - attr(*, "sorted")= chr "doc_id"
```

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 Erläuterung der Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV) Dateiname der extrahierten TXT-Datei.
text	String	(Nur CSV) Volltext der Entscheidung. Über Zeilengrenzen getrennte Wörter sind zusammengefügt.
url	String	(Nur CSV) Link zum Volltext der Entscheidung in der Datenbank des BPatG.
gericht	String	Name des Gerichts. Es ist nur der Wert »BPatG« vergeben. Dies ist der ECLI-Code für »Bundespatentgericht«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Datum (ISO)	Datum der Entscheidung im Format YYYY-MM-DD (ISO-8601).
entscheidung_typ	String	Der Typ der Entscheidung. Es sind die Werte »B« (Beschluss), »U« (Urteil) und »G« (Gerichtsbescheid) vergeben.
entscheidungsjahr	Natürliche Zahl	(Nur CSV) Jahr der Entscheidung im Format YYYY (ISO-8601).
spruchkoerper_az	String	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«.
registerzeichen	String	Das amtliche Registerzeichen. Eine Erläuterung der Abkürzungen findet sich im Abschnitt 7.

(continued)

Variable	Typ	Erläuterung
verfahrensart	String	Die Verfahrensart, auf die das Registerzeichen hinweist. Siehe auch Abschnitt 7.
eingangsnummer	Natürliche Zahl	Eingangsnummer. Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine fortlaufende Nummer (Ordinalzahl) in der Reihenfolge ihres Eingangs, um sie voneinander abzugrenzen.
eingangsjahr_az	Natürliche Zahl	Eingangsjahr laut Aktenzeichen. Das Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Natürliche Zahl	Eingangsjahr im Format YYYY-MM-DD (ISO-8601).
verzoegerung	String	Verzögerungsrügen sind mit »D« gekennzeichnet. Alle anderen Entscheidungen (der weit überwiegende Anteil) sind mit dem Wert »NA« codiert.
kollision	Natürliche Zahl	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.
aktenzeichen	String	Das amtliche Aktenzeichen im Format [senatsnummer] [registerzeichen] [eingangsnummer] / [eingangsjahr] [ggf. zusatz_az]
ecli	String	European Case Law Identifier (ECLI) der Entscheidung. Die ECLI wurde mit den offiziellen ECLIs abgeglichen und sollte fehlerfrei sein.
praesi	String	(Nur CSV-Datei) Der Nachname des oder der Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
v_praesi	String	(Nur CSV-Datei) Der Nachname des oder der Vize-Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
bverwge	Binary	(Nur CSV-Datei) Ob die Entscheidung in die amtliche Sammlung BVerwGE aufgenommen wird. »1« falls ja, »0« falls nein.
leitsatz	Binary	(Nur CSV-Datei) Ob es sich um eine Leitsatzentscheidung handelt. »1« falls ja, »0« falls nein.

(continued)

Variable	Typ	Erläuterung
fachpresse	Binary	(Nur CSV-Datei) Ob die Entscheidung der Fachpresse zukommen soll. »1« falls ja, »0« falls nein. Hinweis: die Verwendung dieser Markierung scheint in Entscheidungen nicht konsequent verwendet zu werden, z.B. fehlt sie bei manchen BVerwGE-Entscheidungen.
uebersetzung	Binary	(Nur CSV-Datei) Ob die Entscheidung übersetzt wird. »1« falls ja, »0« falls nein.
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.
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.

(continued)

Variable	Typ	Erläuterung
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.
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.

(continued)

Variable	Typ	Erläuterung
lizenz	String	Die Lizenz für den Gesamtdatensatz. In diesem Datensatz immer »Creative Commons Zero 1.0 Universal«.

6 Variablen (Linguistische Annotationen)

6.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 61752550 obs. of 12 variables:
## $ doc_id : chr "BVerwG_1997-02-26_U_6_C_3_96_NA_0.txt" "BVerwG
  _1997-02-26_U_6_C_3_96_NA_0.txt" "BVerwG_1997-02-26_U_6_C_3_96_NA_0.txt" "
  BVerwG_1997-02-26_U_6_C_3_96_NA_0.txt" ...
## $ sentence_id : int 1 1 1 1 1 1 1 1 2 2 ...
## $ token_id : int 1 2 3 4 5 6 7 8 1 2 ...
## $ token : chr " " "BUNDESVERWALTUNGSGERICHT" "\n\n
  " "IM" ...
## $ lemma : chr " " "BUNDESVERWALTUNGSGERICHT" "\n\n
  " "in" ...
## $ pos : chr "SPACE" "NOUN" "SPACE" "ADP" ...
## $ tag : chr "_SP" "NE" "_SP" "APPR" ...
## $ head_token_id: num 2 2 2 2 4 7 5 7 1 1 ...
## $ dep_rel : chr "dep" "ROOT" "dep" "mnr" ...
## $ entity : chr "" "" "" "" ...
## $ nounphrase : chr "beg" "end_root" "" "" ...
## $ whitespace : logi FALSE FALSE FALSE TRUE TRUE TRUE ...
## - attr(*, ".internal.selfref")=<externalptr>
```

6.2 Hinweise

Diese Variante des Datensatzes beruht nur auf den Variablen »doc_id« und »text« des regulären Datensatzes, die tokenisiert und mittels der Software »spacy«¹² mit linguistischen Annotationen versehen wurden.

Die Metadaten des Gesamtdatensatzes sind nicht in der linguistische annotierten Fassung enthalten, weil die Bereitstellung von Metadaten für jedes Token die Dateigröße und damit auch den RAM-Bedarf für Analysen gewaltig steigern würde. Um anhand der Metadaten Teilmengen der linguistischen Annotationen zu bilden, gehen sie bitte wie folgt vor:

- CSV-Datei mit Metadaten einlesen
- Anhand der Metadaten die gewünschte Teilmenge der Dokumente bilden
- CSV-Datei mit Linguistischen Annotationen einlesen
- Die Werte der Variable »doc_id« der Teilmenge nutzen um aus den annotierten Daten nur diejenigen herauszufiltern, deren Variable »doc_id« mit der Teilmenge übereinstimmt

¹² Die den Annotationen zugrundeliegende Software ist *spacy*, die hier verfügbar ist <https://spacy.io/>. Diese wird in R mittels des *spacyr* packages integriert: <https://spacyr.quanteda.io/>.

6.3 Erläuterung der Variablen

Variable	Typ	Erläuterung
doc_id	String	Der Dateiname des Dokumentes, aus dem die Tokens stammen. Identische Werte wie im Hauptdatensatz. Geeignet um Metadaten mit den linguistischen Annotationen zu verbinden.
sentence_id	Natürliche Zahl	Die Ordinalzahl des Satzes in dem Dokument, dem das Token zugeordnet ist.
token_id	Natürliche Zahl	Die Nummer des Tokens in einem Dokument.
token	String	Einzelne Tokens in der Reihenfolge ihres Vorkommens im jeweiligen Dokument.
lemma	String	Die lemmatisierte Form des Tokens.
pos	String	Grobe Annotation des einzelnen Tokens nach dem universal dependency POS tagset, siehe auch http://universaldependencies.org/u/pos/ .
tag	String	Feine Annotation des einzelnen Tokens mit dem »de_core_news_sm«-Modell von spacy. Für eine detaillierte Erläuterung der Annotationen siehe: https://spacy.io/models/de
head_token_id	Natürliche Zahl	Das führende Token.
dep_rel	String	Die <i>dependency relation</i> zum head_token.
entity	String	Erkennung von benannten Entitäten (Named Entity Recognition).
nounphrase	String	Erkennung von Nominalphrasen.
whitespace	Logisch	Ob es sich bei dem Token um Whitespace handelt oder nicht.

7 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

8 Präsident:innen

8.1 Hinweise

- Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2023). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2023-03-21. Zenodo. <https://github.com/SeanFobbe/pvp-fcg>«.
- Das Datum bezieht sich jeweils auf das Amt als Präsident:in, nicht auf die Amtszeit als Richter:in.

8.2 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	2021-06-30	1955-09-24	NA
VACANCY-6	VACANCY-6	2021-07-01	2022-09-07	NA	NA
Korbmacher	Andreas	2022-09-08	NA	1960-01-12	NA

8.3 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	65	NA
Korbmacher	Andreas	62	NA	NA

9 Vize-Präsident:innen

9.1 Hinweise

- Die Personendaten stammen aus folgendem Datensatz: »Seán Fobbe and Tilko Swalve (2023). Presidents and Vice-Presidents of the Federal Courts of Germany (PVP-FCG). Version 2023-03-21. Zenodo. <https://github.com/SeanFobbe/pvp-fcg>«.
- Das Datum bezieht sich jeweils auf das Amt als Vize-Präsident:in, nicht auf die Amtszeit als Richter:in.

9.2 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	2022-09-07	1960-01-12	NA
VACANCY-4	VACANCY-4	2022-09-08	2023-01-30	NA	NA
Rublack	Susanne	2023-01-31	NA	1962-08-13	NA

9.3 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	59	62	NA
Rublack	Susanne	60	NA	NA

10 Linguistische Kennzahlen

10.1 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.

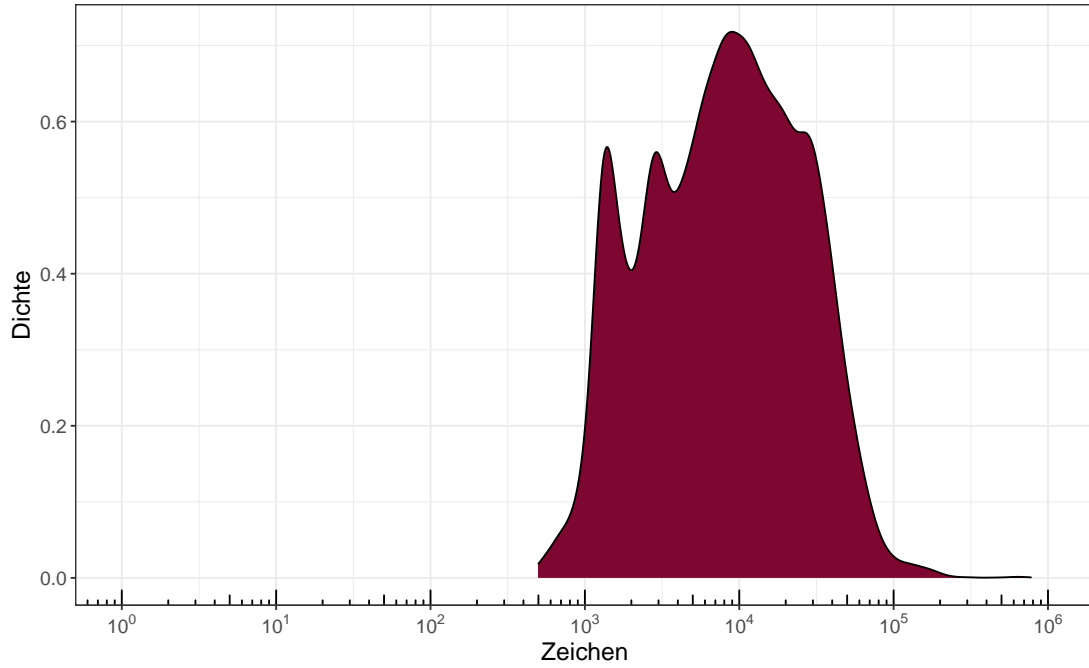
Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

10.2 Werte der Kennzahlen

Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	391,878,093	497	3,076	8,206.5	14,407.28	19,155.00	782,167
tokens	59,282,333	62	441	1,235.0	2,179.50	2,887.25	119,583
typen	383,132	47	226	484.0	629.18	866.00	12,955
saetze	3,424,057	4	32	74.0	125.88	165.00	5,127

10.3 Verteilung Zeichen

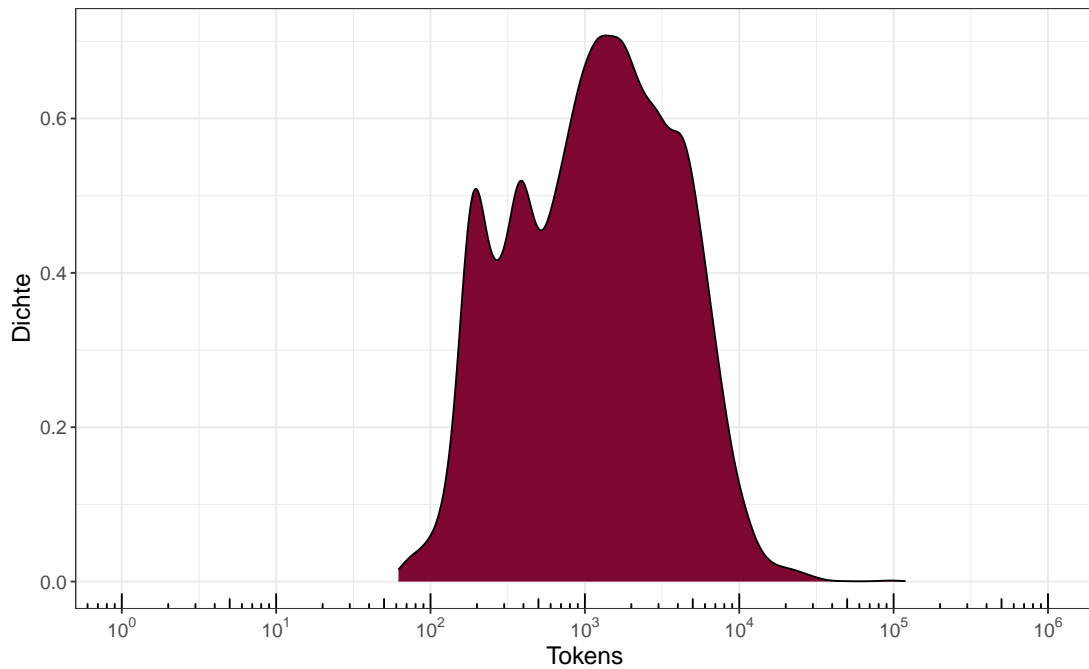
CE-BVerwG | Version 2024-03-13 | Verteilung der Zeichen je Dokument



Fobbe | DOI: 10.5281/zenodo.10809039

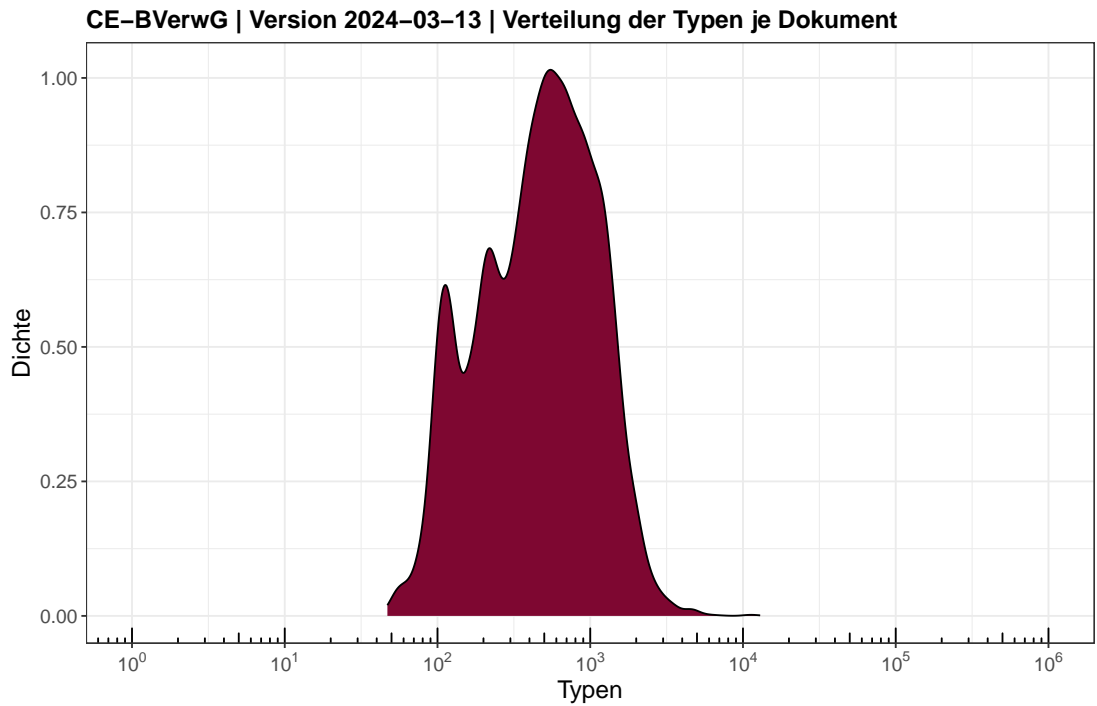
10.4 Verteilung Tokens

CE-BVerwG | Version 2024-03-13 | Verteilung der Tokens je Dokument



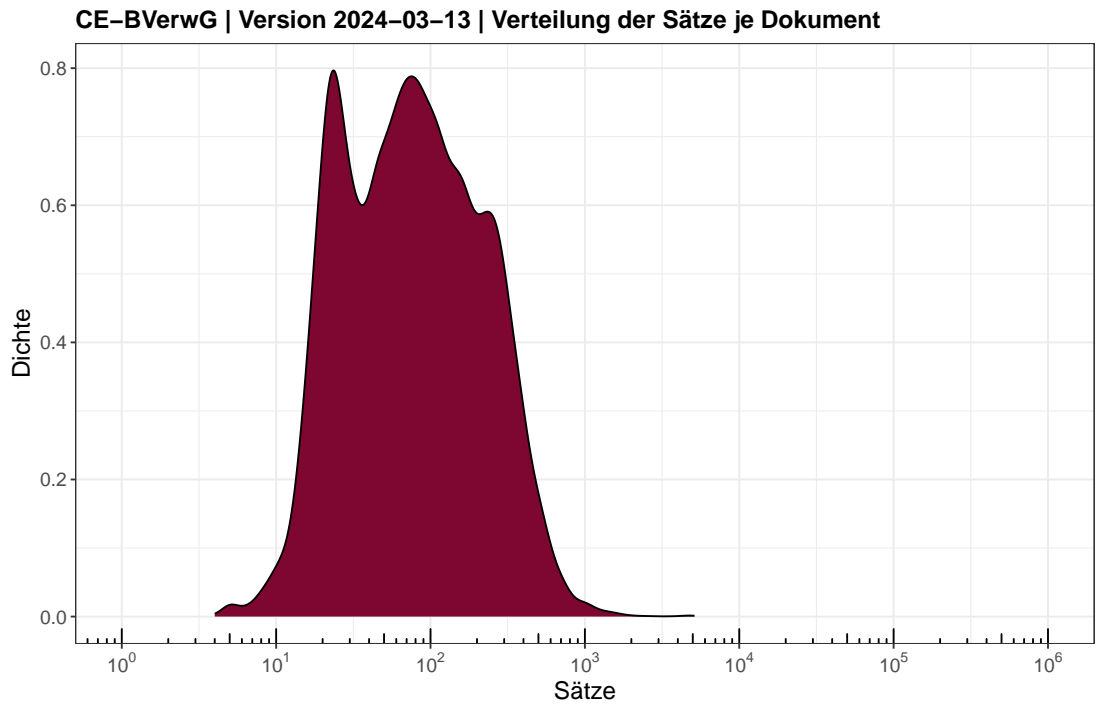
Fobbe | DOI: 10.5281/zenodo.10809039

10.5 Verteilung Typen



Fobbe | DOI: 10.5281/zenodo.10809039

10.6 Verteilung Sätze



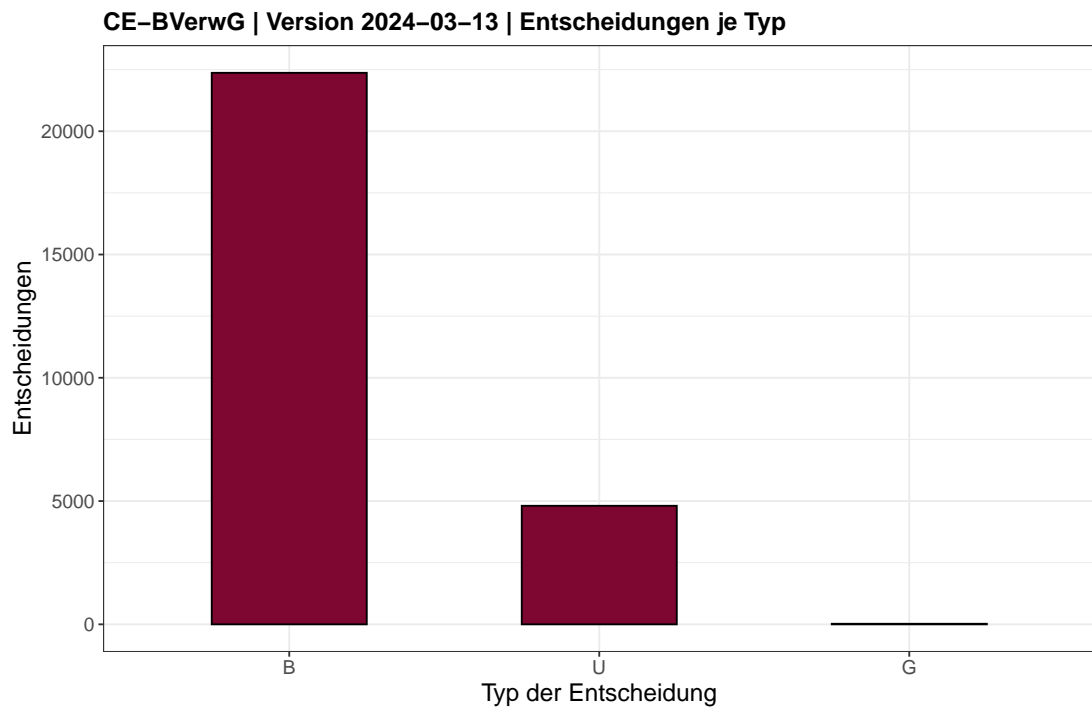
Fobbe | DOI: 10.5281/zenodo.10809039

11 Inhalt des Korpus

11.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	25	1997	2005	2010	2010.65	2016	2024
eingangsjahr_iso	26	1996	2005	2009	2010.16	2015	2024
eingangsnummer	573	1	9	25	58.81	56	7005

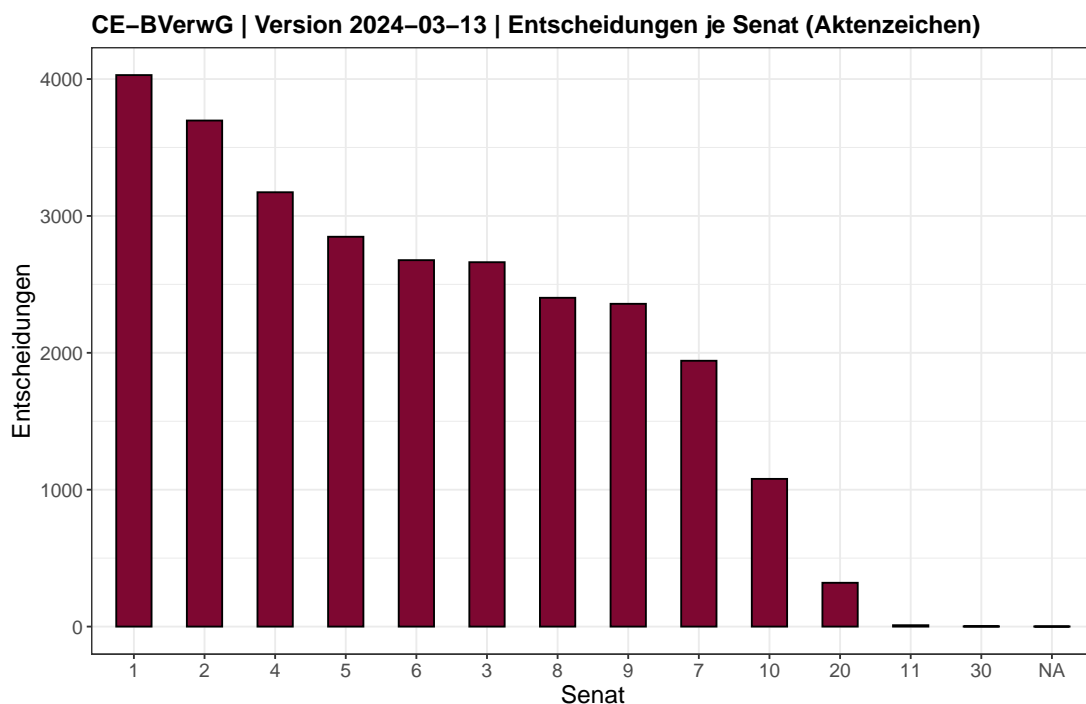
11.2 Nach Typ der Entscheidung



Fobbe | DOI: 10.5281/zenodo.10809039

Typ	Entscheidungen	% Gesamt	% Kumulativ
B	22370	82.24	82.24
G	22	0.08	82.32
U	4808	17.68	100.00
Total	27200	100.00	100.00

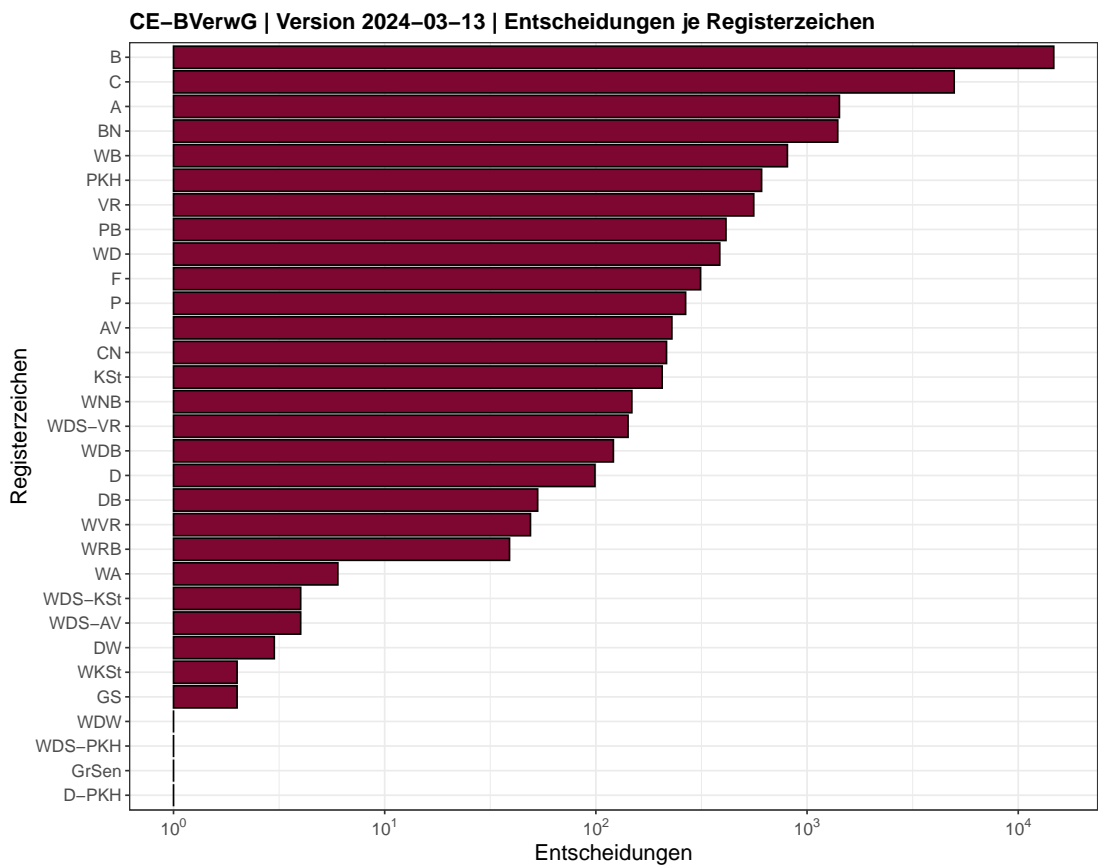
11.3 Nach Spruchkörper (Aktenzeichen)



Fobbe | DOI: 10.5281/zenodo.10809039

Senat	Entscheidungen	% Gesamt	% Kumulativ
NA	1	0.00	0.00
1	4029	14.81	14.82
2	3697	13.59	28.41
3	2662	9.79	38.19
4	3173	11.67	49.86
5	2848	10.47	60.33
6	2677	9.84	70.17
7	1942	7.14	77.31
8	2402	8.83	86.14
9	2358	8.67	94.81
10	1079	3.97	98.78
11	10	0.04	98.82
20	320	1.18	99.99
30	2	0.01	100.00
Total	27200	100.00	100.00

11.4 Nach Registerzeichen

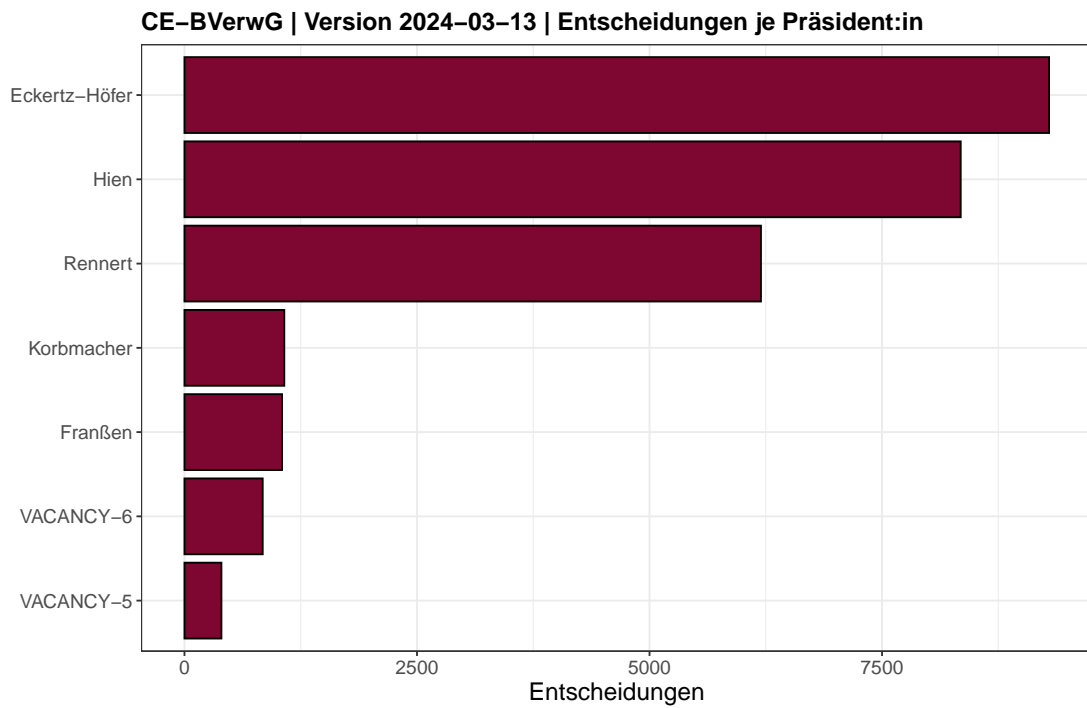


Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
A	1423	5.23	5.23
AV	229	0.84	6.07
B	14727	54.14	60.22
BN	1397	5.14	65.35
C	4973	18.28	83.64
CN	216	0.79	84.43
D	99	0.36	84.79
D-PKH	1	0.00	84.80
DB	53	0.19	84.99
DW	3	0.01	85.00
F	313	1.15	86.15
GS	2	0.01	86.16
GrSen	1	0.00	86.17

(continued)

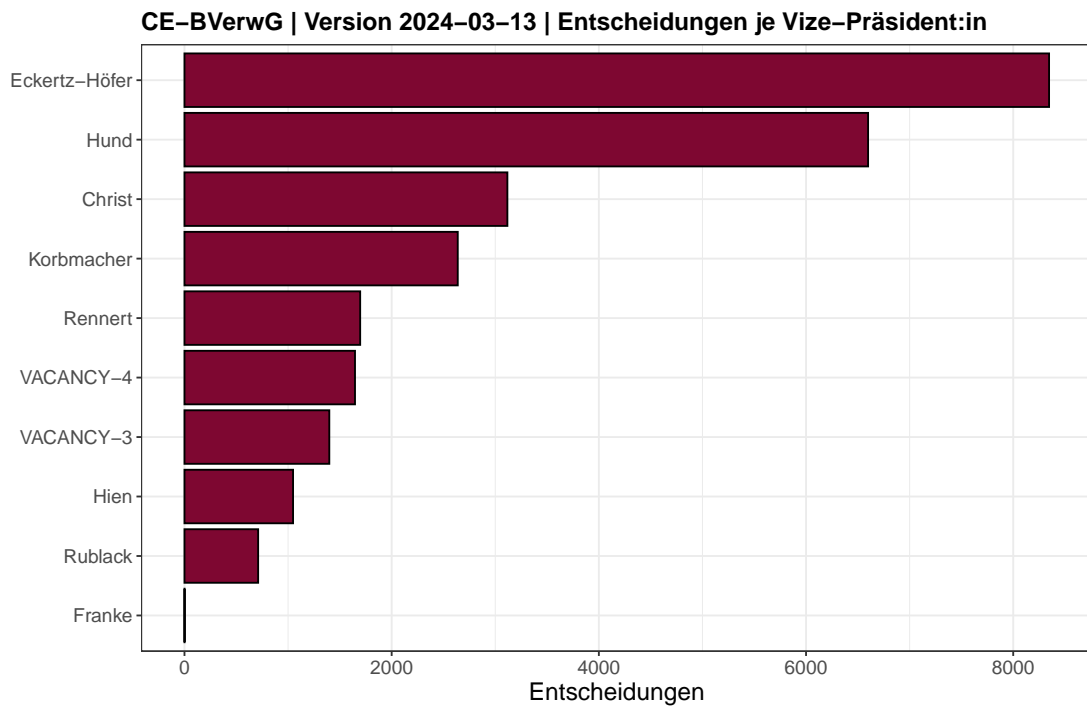
Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
KSt	206	0.76	86.92
P	266	0.98	87.90
PB	413	1.52	89.42
PKH	609	2.24	91.66
VR	559	2.06	93.71
WA	6	0.02	93.74
WB	807	2.97	96.70
WD	386	1.42	98.12
WDB	121	0.44	98.57
WDS-AV	4	0.01	98.58
WDS-KSt	4	0.01	98.60
WDS-PKH	1	0.00	98.60
WDS-VR	142	0.52	99.12
WDW	1	0.00	99.12
WKSt	2	0.01	99.13
WNB	148	0.54	99.68
WRB	39	0.14	99.82
WVR	49	0.18	100.00
Total	27200	100.00	100.00

11.5 Nach Präsident:in



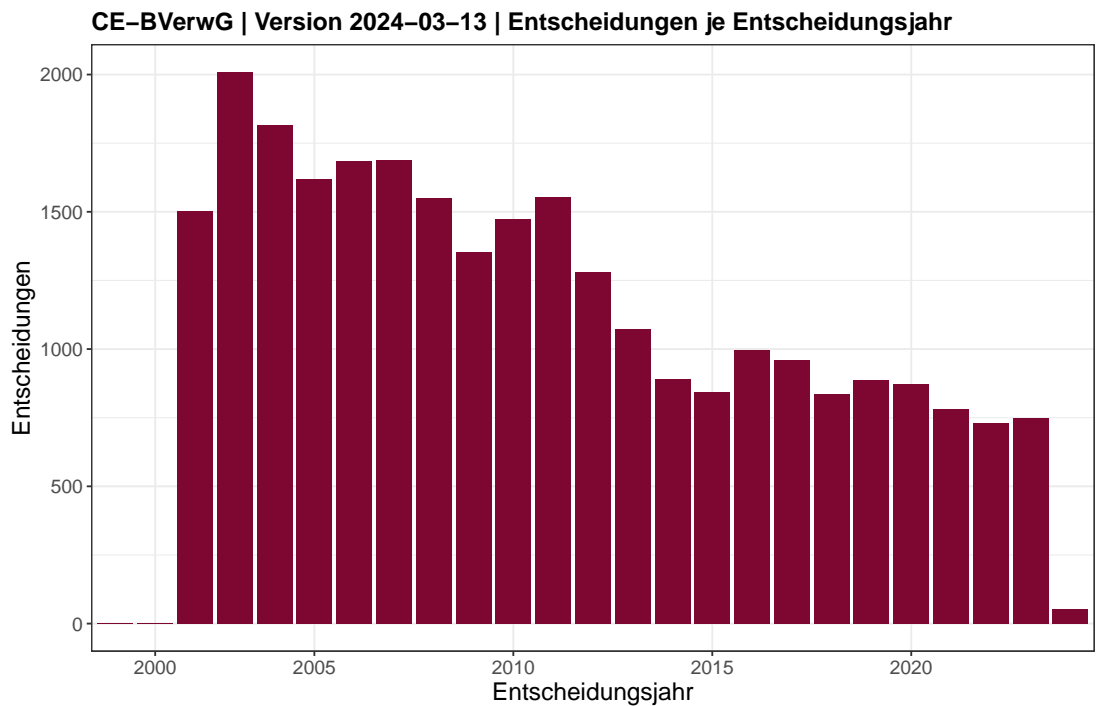
Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Eckertz-Höfer	9296	34.18	34.18
Franßen	1049	3.86	38.03
Hien	8346	30.68	68.72
Korbmacher	1073	3.94	72.66
Rennert	6198	22.79	95.45
VACANCY-5	397	1.46	96.91
VACANCY-6	841	3.09	100.00
Total	27200	100.00	100.00

11.6 Nach Vize-Präsident:in



Vize-Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Christ	3117	11.46	11.46
Eckertz-Höfer	8346	30.68	42.14
Franke	1	0.00	42.15
Hien	1048	3.85	46.00
Hund	6599	24.26	70.26
Korbmacher	2638	9.70	79.96
Rennert	1696	6.24	86.19
Rublack	711	2.61	88.81
VACANCY-3	1398	5.14	93.95
VACANCY-4	1646	6.05	100.00
Total	27200	100.00	100.00

11.7 Nach Entscheidungsjahr

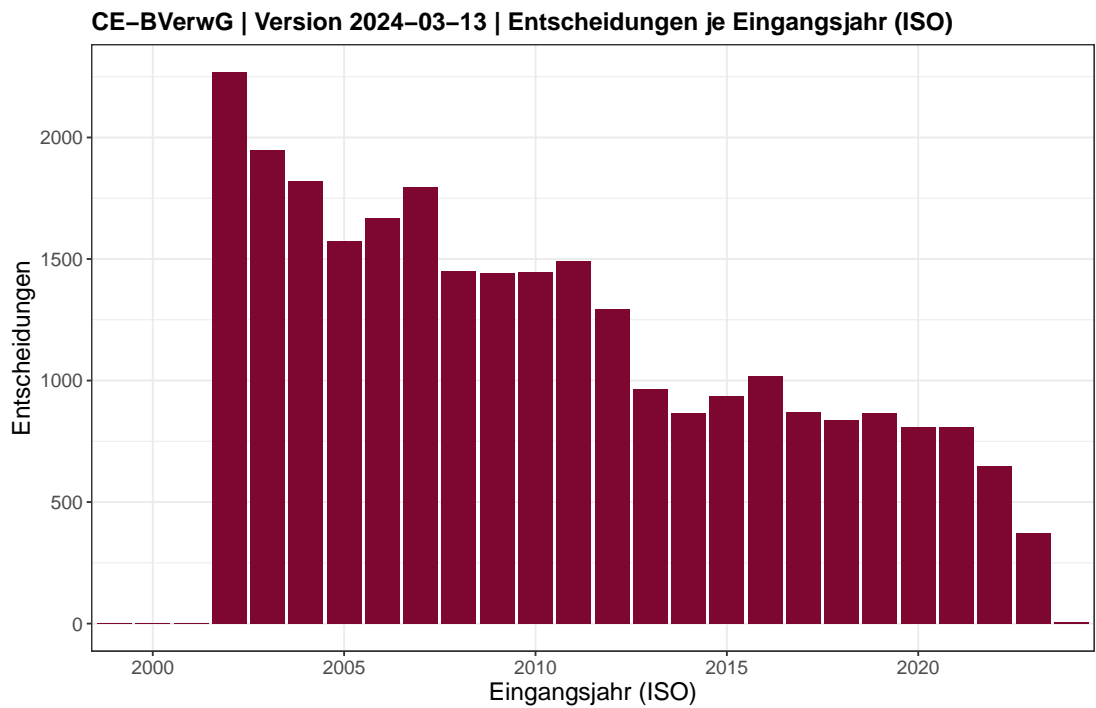


Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
1997	1	0.00	0.00
2000	1	0.00	0.01
2002	1502	5.52	5.53
2003	2009	7.39	12.92
2004	1817	6.68	19.60
2005	1619	5.95	25.55
2006	1684	6.19	31.74
2007	1690	6.21	37.95
2008	1549	5.69	43.65
2009	1353	4.97	48.62
2010	1474	5.42	54.04
2011	1553	5.71	59.75
2012	1281	4.71	64.46
2013	1072	3.94	68.40
2014	889	3.27	71.67
2015	845	3.11	74.78

(continued)

Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2016	995	3.66	78.43
2017	960	3.53	81.96
2018	836	3.07	85.04
2019	887	3.26	88.30
2020	871	3.20	91.50
2021	780	2.87	94.37
2022	731	2.69	97.06
2023	747	2.75	99.80
2024	54	0.20	100.00
Total	27200	100.00	100.00

11.8 Nach Eingangsjahr (ISO)



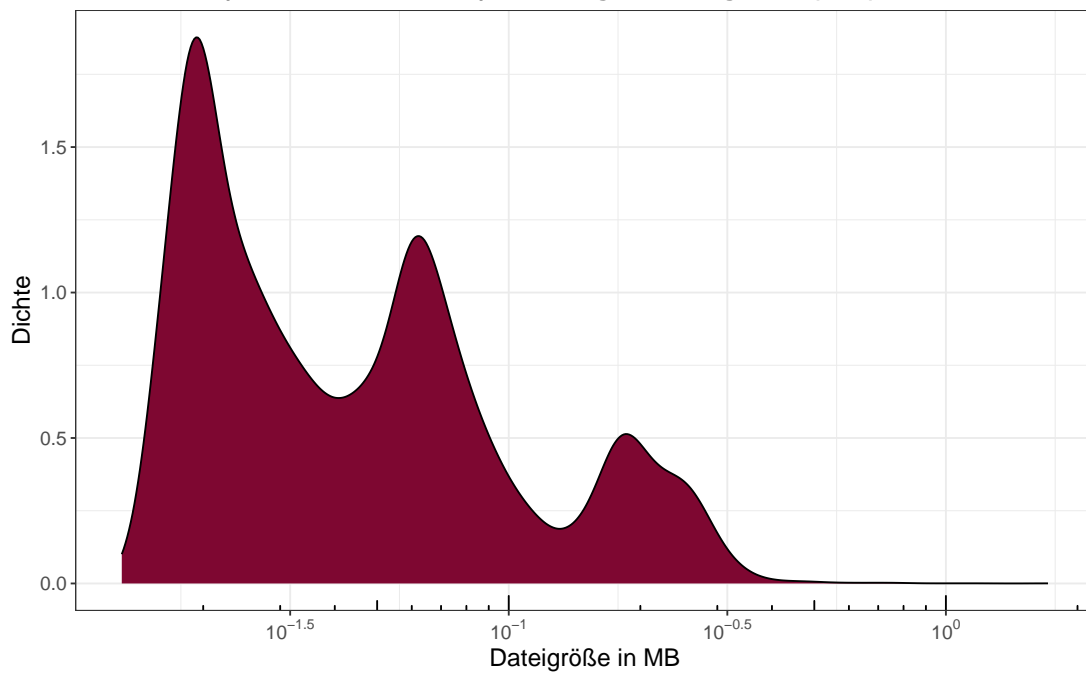
Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
1996	1	0.00	0.00
2000	1	0.00	0.01
2001	2	0.01	0.01
2002	2269	8.34	8.36
2003	1948	7.16	15.52
2004	1822	6.70	22.22
2005	1575	5.79	28.01
2006	1670	6.14	34.15
2007	1794	6.60	40.74
2008	1451	5.33	46.08
2009	1440	5.29	51.37
2010	1445	5.31	56.68
2011	1493	5.49	62.17
2012	1293	4.75	66.93
2013	964	3.54	70.47
2014	864	3.18	73.65

(continued)

Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
2015	936	3.44	77.09
2016	1020	3.75	80.84
2017	869	3.19	84.03
2018	836	3.07	87.11
2019	864	3.18	90.28
2020	808	2.97	93.25
2021	809	2.97	96.23
2022	646	2.38	98.60
2023	374	1.38	99.98
2024	6	0.02	100.00
Total	27200	100.00	100.00

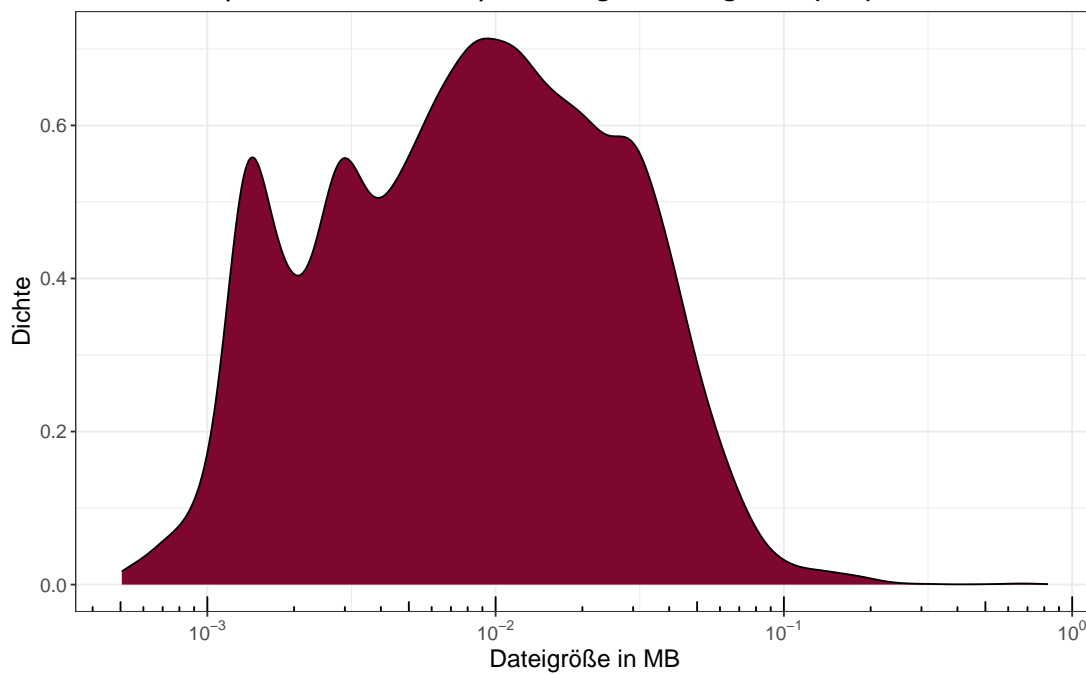
12 Dateigrößen

CE-BVerwG | Version 2024-03-13 | Verteilung der Dateigrößen (PDF)



Fobbe | DOI: 10.5281/zenodo.10809039

CE-BVerwG | Version 2024-03-13 | Verteilung der Dateigrößen (TXT)



Fobbe | DOI: 10.5281/zenodo.10809039

13 Kryptographische Signaturen

13.1 Zwei-Phasen-Signatur

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, das Codebook und die Robustness Checks Hash-Werte in zwei verschiedenen Verfahren (SHA2-256 und SHA3-512) berechnet und in einer CSV-Datei dokumentiert.

In **Phase II** werden diese CSV-Datei und der Compilation Report 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.

13.2 Persönliche GPG-Signatur

Die während der Kompilierung des Datensatzes erstellte CSV-Datei mit den Hash-Prüfsummen und der Compilation Report sind 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

14 Changelog

14.1 Version 2024-03-13

- Vollständige Aktualisierung der Daten
- Die Pipeline mit allen Zwischenergebnissen wird nun automatisch in “output/” archivierte
- Gruppierung von Dateinamen der Diagramme nach Typ
- Vereinfachung der Repository-Struktur
- Anpassung von Docker Compose an Debian 11
- Update GPG Public Key in Repository

14.2 Version 2023-03-21

- Vollständige Aktualisierung der Daten
- Gesamte Laufzeitumgebung mit Docker versionskontrolliert
- 4 neue Variablen: BVerwGE, Leitsatz, Fachpresse und Übersetzung (jeweils binär)
- Linguistisch annotierte Variante ist wieder verfügbar
- Aktenzeichen aus dem Eingangszeitraum 2000 bis 2009 nun korrekt mit führender Null formatiert (z.B. 1 BvR 44/02 statt 1 BvR 44/2)
- Aktenzeichen in Verzögerungsverfahren nun korrekt formatiert
- Aktenzeichen verwenden nun richtigerweise den Punkt als Trenner (statt den Schrägstrich wie bei anderen Gerichten)
- Berücksichtigung von Präsident Korbmacher und Vize-Präsidentin Rublack in den jeweiligen Variablen
- Update des Run-Skripts und des Delete-Skripts
- Proto-Package Mono-Repo entfernt, alle Funktionen nun fest projektbasiert versionskontrolliert
- Vereinfachung der Konfigurations-Datei
- Neue Funktion für automatischen clean run (Löschung aller Zwischenergebnisse)
- Update der Download-Funktion
- Überflüssige Warnung in `f.future_lingsummarize`-Funktion entfernt
- Alle Roh-Dateien werden nun im Unterordner “files” gespeichert
- Vom BVerwG nicht veröffentlichte Entscheidungstypen im Codebook genauer beschrieben
- Verbesserte Formatierung von Profiling, Warnungen und Fehlermeldungen im Compilation Report
- Zusätzliche Unit-Tests
- Verbesserung des Robustness Check Reports
- README im Hinblick auf Docker überarbeitet

14.3 Version 2022-08-07

- Vollständige Aktualisierung der Daten
- Neuentwurf des Source Codes im *targets* framework
- Einführung von separater Konfigurations-Datei
- Strenge Versionskontrolle aller R packages mit *renv*
- Variante mit linguistischen Annotationen ist temporär nicht mehr verfügbar
- Neue Visualisierung der gesamten Daten-Pipeline
- Vielzahl zusätzlicher Unit Tests (inklusive `type safety`) für wichtige Funktionen

14.4 Version 2021-10-19

- Vollständige Aktualisierung der Daten
- Neue Variante mit linguistische Annotationen
- Neue Variable für Lizenz
- Strenge Kontrolle und semantische Sortierung aller Variablen-Namen
- Source Code des Codebooks deutlich vereinfacht (insbes. Diagramme und Changelog)
- Erweiterung der Dokumentation
- In den linguistischen Kennzahlen werden jetzt auch die Anzahl Typen bezogen auf den Gesamtdatensatz berechnet
- Standardisierung der Diagramme auf 6:9 Zoll (Breite/Höhe)

14.5 Version 2021-04-15

- 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

14.6 Version 2020-06-23

- Erstveröffentlichung

15 Parameter für strenge Replikationen

```
## [1] "OpenSSL 3.0.2 15 Mar 2022 (Library: OpenSSL 3.0.2 15 Mar 2022)"
```

```
## R version 4.2.2 (2022-10-31)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.2 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p-r0.3.20.so
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8 LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## other attached packages:
## [1] future.apply_1.10.0 future_1.32.0 spacyr_1.2.1
## [4] quanteda_3.2.4 readtext_0.81 data.table_1.14.8
## [7] scales_1.2.1 ggraph_2.1.0 ggplot2_3.4.1
## [10] pdftools_3.3.3 kableExtra_1.3.4 knitr_1.42
## [13] rvest_1.0.3 httr_1.4.5 mgsub_1.7.3
## [16] zip_2.2.2 fs_1.6.1 testthat_3.1.7
## [19] RcppTOML_0.2.2 tarchetypes_0.7.5 targets_0.14.3
##
## loaded via a namespace (and not attached):
## [1] webshot_0.5.4 rprojroot_2.0.3 future.callr_0.8.1
## [4] tools_4.2.2 backports_1.4.1 utf8_1.2.3
## [7] R6_2.5.1 colorspace_2.1-0 withr_2.5.0
## [10] tidyselct_1.2.0 gridExtra_2.3 processx_3.8.0
## [13] curl_5.0.0 compiler_4.2.2 cli_3.6.0
## [16] xml2_1.3.3 desc_1.4.2 labeling_0.4.2
## [19] stringfish_0.15.7 callr_3.7.3 askpass_1.1
## [22] rappdirs_0.3.3 systemfonts_1.0.4 stringr_1.5.0
## [25] digest_0.6.31 rmarkdown_2.20 svglite_2.1.1
## [28] pkgconfig_2.0.3 htmltools_0.5.4 parallely_1.34.0
## [31] fastmap_1.1.1 rlang_1.0.6 rstudioapi_0.14
## [34] RApiSerialize_0.1.2 farver_2.1.1 generics_0.1.3
## [37] jsonlite_1.8.4 dplyr_1.1.0 magrittr_2.0.3
## [40] Matrix_1.5-1 Rcpp_1.0.10 munsell_0.5.0
## [43] fansi_1.0.4 reticulate_1.28 viridis_0.6.2
## [46] lifecycle_1.0.3 furrr_0.3.1 stringi_1.7.12
## [49] yaml_2.3.7 MASS_7.3-58.1 brio_1.1.3
## [52] grid_4.2.2 parallel_4.2.2 listenv_0.9.0
## [55] ggrepel_0.9.3 lattice_0.20-45 graphlayouts_0.8.4
## [58] ps_1.7.2 pillar_1.8.1 igraph_1.4.1
```

```
## [61] base64url_1.4      codetools_0.2-18  stopwords_2.3
## [64] pkgload_1.3.2      fastmatch_1.1-3  glue_1.6.2
## [67] evaluate_0.20      qpdf_1.3.0       RcppParallel_5.1.7
## [70] png_0.1-8          vctrs_0.5.2      tweenr_2.0.2
## [73] selectr_0.4-2     gtable_0.3.1     purrr_1.0.1
## [76] polyclip_1.10-4   tidyr_1.3.0      qs_0.25.5
## [79] xfun_0.37          ggforce_0.4.1    tidygraph_1.2.3
## [82] viridisLite_0.4.1 tibble_3.2.0     tinytex_0.44
## [85] globals_0.16.2    here_1.0.1
```

Literaturverzeichnis

- Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2023. *Rmarkdown: Dynamic Documents for R*.
- Bengtsson, Henrik. 2021. “A Unifying Framework for Parallel and Distributed Processing in R Using Futures.” *The R Journal* 13 (2): 208–27. <https://doi.org/10.32614/RJ-2021-048>.
- . 2022. *Future.apply: Apply Function to Elements in Parallel Using Futures*.
- . 2023. *Future: Unified Parallel and Distributed Processing in R for Everyone*.
- 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. 2022. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Csardi, Gabor, and Tamas Nepusz. 2006. “The Igraph Software Package for Complex Network Research.” *InterJournal Complex Systems*: 1695. <https://igraph.org>.
- Csárdi, Gábor, Kuba Podgórski, and Rich Geldreich. 2022. *Zip: Cross-Platform Zip Compression*. <https://github.com/r-lib/zip#readme>.
- Dowle, Matt, and Arun Srinivasan. 2023. *Data.table: Extension of 'Data.frame'*.
- Eddelbuettel, Dirk. 2023. *RcppTOML: Rcpp Bindings to Parser for "Tom's Obvious Markup Language"*. <http://dirk.eddelbuettel.com/code/rcpp.toml.html>.
- Ewing, Mark. 2021. *Mgsub: Safe, Multiple, Simultaneous String Substitution*.
- file., See AUTHORS. 2023. *Igraph: Network Analysis and Visualization*.
- Gagolewski, Marek. 2022. “stringi: Fast and Portable Character String Processing in R.” *Journal of Statistical Software* 103 (2): 1–59. <https://doi.org/10.18637/jss.v103.i02>.
- Gagolewski, Marek, Bartek Tartanus, others; Unicode, Inc., and others. 2023. *Stringi: Fast and Portable Character String Processing Facilities*.
- Landau, William Michael. 2021a. *Tarchetypes: Archetypes for Targets*.
- . 2021b. “The Targets R Package: A Dynamic Make-Like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing.” *Journal of Open Source Software* 6 (57): 2959. <https://doi.org/10.21105/joss.02959>.
- . 2023a. *Tarchetypes: Archetypes for Targets*.
- . 2023b. *Targets: Dynamic Function-Oriented Make-Like Declarative Pipelines*.

- Ooms, Jeroen. 2023a. *Magick: Advanced Graphics and Image-Processing in R*.
- . 2023b. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*.
- Pedersen, Thomas Lin. 2022. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*.
- Ushey, Kevin. 2023. *Renv: Project Environments*. <https://rstudio.github.io/renv/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2022. *Rvest: Easily Harvest (Scrape) Web Pages*.
- . 2023. *Httr: Tools for Working with Urls and Http*.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2023. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*.
- Wickham, Hadley, and Dana Seidel. 2022. *Scales: Scale Functions for Visualization*.
- 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.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.
- Xie, Yihui, J. J. Allaire, and Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.
- Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.
- Zhu, Hao. 2021. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*.