

**Corpus der Entscheidungen  
des  
Bundesgerichtshofs  
(CE-BGH)**

CODEBOOK

Version 2023-03-10



DOI: [10.5281/zenodo.7699032](https://doi.org/10.5281/zenodo.7699032)

<b>Titel</b>	Corpus der Entscheidungen des Bundesgerichtshofs
<b>Abkürzung</b>	CE-BGH
<b>Autor</b>	Seán Fobbe
<b>Version</b>	2023-03-10
<b>Download</b>	<a href="https://doi.org/10.5281/zenodo.7699032">https://doi.org/10.5281/zenodo.7699032</a>
<b>Lizenz</b>	CC0 1.0 Universal

### Zitiervorschlag

*Seán Fobbe* (2023). Corpus der Entscheidungen des Bundesgerichtshofs (CE-BGH). Version 2023-03-10. Zenodo. DOI: 10.5281/zenodo.7699032.

### 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 2023-03-10. 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.3942742. 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

<b>1</b>	<b>Einführung</b>	<b>4</b>
<b>2</b>	<b>Nutzung</b>	<b>5</b>
2.1	CSV-Dateien . . . . .	5
2.2	TXT-Dateien . . . . .	5
<b>3</b>	<b>Konstruktion</b>	<b>6</b>
3.1	Beschreibung des Datensatzes . . . . .	6
3.2	Datenquellen . . . . .	6
3.3	Sammlung der Daten . . . . .	6
3.4	Source Code und Compilation Report . . . . .	6
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 . . . . .	9
3.7.3	Beispiel eines Dateinamens . . . . .	9
3.8	Qualitätsprüfung . . . . .	9
3.9	Grafische Darstellung . . . . .	9
<b>4</b>	<b>Varianten und Zielgruppen</b>	<b>10</b>
<b>5</b>	<b>Variablen</b>	<b>12</b>
5.1	Datenstruktur . . . . .	12
5.2	Hinweise . . . . .	13
5.3	Erläuterung der Variablen . . . . .	13
<b>6</b>	<b>Registerzeichen</b>	<b>20</b>
<b>7</b>	<b>Präsident:innen</b>	<b>24</b>
7.1	Hinweise . . . . .	24
7.2	Lebensdaten . . . . .	24
7.3	Dienstalter und Lebensalter . . . . .	24
<b>8</b>	<b>Vize-Präsident:innen</b>	<b>25</b>
8.1	Hinweise . . . . .	25
8.2	Lebensdaten . . . . .	25
8.3	Dienstalter und Lebensalter . . . . .	26
<b>9</b>	<b>Linguistische Kennzahlen</b>	<b>27</b>
9.1	Erläuterung der Kennzahlen und Diagramme . . . . .	27
9.2	Werte der Kennzahlen . . . . .	27
9.3	Verteilung Zeichen . . . . .	28
9.4	Verteilung Tokens . . . . .	28
9.5	Verteilung Typen . . . . .	29
9.6	Verteilung Sätze . . . . .	29
<b>10</b>	<b>Zitations-Netzwerke</b>	<b>30</b>

<b>11 Inhalt des Korpus</b>	<b>32</b>
11.1 Zusammenfassung . . . . .	32
11.2 Nach Typ der Entscheidung . . . . .	32
11.3 Nach Spruchkörper (Aktenzeichen) . . . . .	33
11.4 Nach Spruchkörper (Datenbank) . . . . .	35
11.5 Nach Registerzeichen . . . . .	37
11.6 Nach Präsident:in . . . . .	40
11.7 Nach Vize-Präsident:in . . . . .	41
11.8 Nach Entscheidungsjahr . . . . .	42
11.9 Nach Eingangsjahr (ISO) . . . . .	44
<b>12 Dateigrößen</b>	<b>46</b>
<b>13 Kryptographische Signaturen</b>	<b>47</b>
13.1 Zwei-Phasen-Signatur . . . . .	47
13.2 Persönliche GPG-Signatur . . . . .	47
<b>14 Changelog</b>	<b>48</b>
14.1 Version 2023-03-10 . . . . .	48
14.2 Version 2022-08-16 . . . . .	48
14.3 Version 2022-02-12 . . . . .	48
14.4 Version 2021-04-27 . . . . .	49
14.5 Version 2020-07-09 . . . . .	49
<b>15 Parameter für strenge Replikationen</b>	<b>50</b>
<b>Literaturverzeichnis</b>	<b>52</b>

# 1 Einführung

Der **Bundesgerichtshof (BGH)** ist einer der fünf obersten Gerichtshöfe des Bundes und steht an der Spitze der ordentlichen Gerichtsbarkeit der Bundesrepublik Deutschland (Art. 95 Abs. 1 GG, §§ 12, 13 GVG).<sup>1</sup> Der BGH ist die höchste Instanz in Zivil- und Strafsachen, sowie in einigen ihm zugewiesenen Spezialgebieten. Er wurde am 1. Oktober 1950 errichtet und hat seinen Sitz in Karlsruhe (§ 123 GVG). Der 5. und 6. Strafsenat sind allerdings in Leipzig beheimatet.

Im Jahr 2022 am BGH eingerichtet sind 13 Zivilsenate, 6 Strafsenate, und 8 Spezialsenate (6 berufsrechtliche Senate, Kartellsenat und Landwirtschaftsenat), sowie ein Großer Zivilsenat, ein Großer Strafsenat und die Vereinigten Großen Senate.<sup>2</sup> Ein Senat hat grundsätzlich 7 bis 9 Mitglieder, entscheidet aber als Spruchkörper in Senatsgruppen von 5 Mitgliedern einschließlich des/der Vorsitzenden (§ 139 Abs. 1 GVG).

Die überwiegende Anzahl der Verfahren vor dem BGH sind Revisionen, d.h. die Überprüfung von Entscheidungen unterer Instanzen (Landgerichte oder Oberlandesgerichte, Amtsgericht nur bei Sprungrevision) auf Rechtsfehler ohne erneute Beweisaufnahme. In Zivilsachen ist der BGH zuständig für Revisionen, Sprungrevisionen, Rechtsbeschwerden und Sprungrechtsbeschwerden (§ 133 GVG). In Strafsachen ist er zuständig für Revisionen, Beschwerden gegen Beschlüsse und Verfügungen der Oberlandesgerichte, Beschwerden gegen Verfügungen des Ermittlungsrichters am BGH und Rügen der Besetzung eines Oberlandesgerichts (§ 135 GVG).

Wieso dieser Datensatz? Die quantitative Analyse von juristischen Texten, insbesondere denen des BGH, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.<sup>3</sup> 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*<sup>4</sup> und *openJur*<sup>5</sup> verbieten ausdrücklich das maschinelle Auslesen der Rohdaten. 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.<sup>6</sup>

---

<sup>1</sup> Die »ordentliche Gerichtsbarkeit« ist eine historische gewachsene Bezeichnung. Früher war die Verwaltungsgerichtsbarkeit nicht mit unabhängigen Richtern, sondern mit Verwaltungsbeamten besetzt und daher »außerordentlich«. Die mit unabhängigen Richtern besetzten Gerichte wurden als »ordentlich« bezeichnet.

<sup>2</sup> Geschäftsverteilungsplan des Bundesgerichtshofs für das Jahr 2021.

<sup>3</sup> Besonders positive Ausnahmen finden sich unter: <https://www.quantitative-rechtswissenschaft.de/>  
<sup>4</sup> <https://opinioiuris.de/>

<sup>5</sup> <https://openjur.de/>

<sup>6</sup> 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. CSV<sup>7</sup> 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.bgh <- 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.bgh <- readtext("./*.txt",
  docvarsfrom = "filenames",
  docvarnames = c("gericht",
                  "spruchkoerper_db",
                  "leitsatz",
                  "datum",
                  "spruchkoerper_az",
                  "registerzeichen",
                  "eingangsnummer",
                  "eingangsjahr_az",
                  "zusatz_az",
                  "name",
                  "kollision"),
  dvsep = "_",
  encoding = "UTF-8")
```

---

<sup>7</sup> 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 begründeten Entscheidungen, die auf der amtlichen Internetpräsenz des Bundesgerichtshofs (BGH) 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.

Speziell an Praktiker:innen richten sich die PDF-Sammlungen aller Leitsatzentscheidungen und aller mit Namen versehenen Entscheidungen.

### 3.2 Datenquellen

Datenquelle	Fundstelle
Primäre Datenquelle	<a href="https://www.bundesgerichtshof.de">https://www.bundesgerichtshof.de</a>
Source Code	<a href="https://doi.org/10.5281/zenodo.7699033">https://doi.org/10.5281/zenodo.7699033</a>
Personendaten	<a href="https://doi.org/10.5281/zenodo.4568682">https://doi.org/10.5281/zenodo.4568682</a>
Registerzeichen	<a href="https://doi.org/10.5281/zenodo.4569564">https://doi.org/10.5281/zenodo.4569564</a>

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.7699033>

Mit jeder Kompilierung des vollständigen Datensatzes wird auch ein umfangreicher **Compilation Report** in einem attraktiv designten PDF-Format erstellt (ähnlich diesem

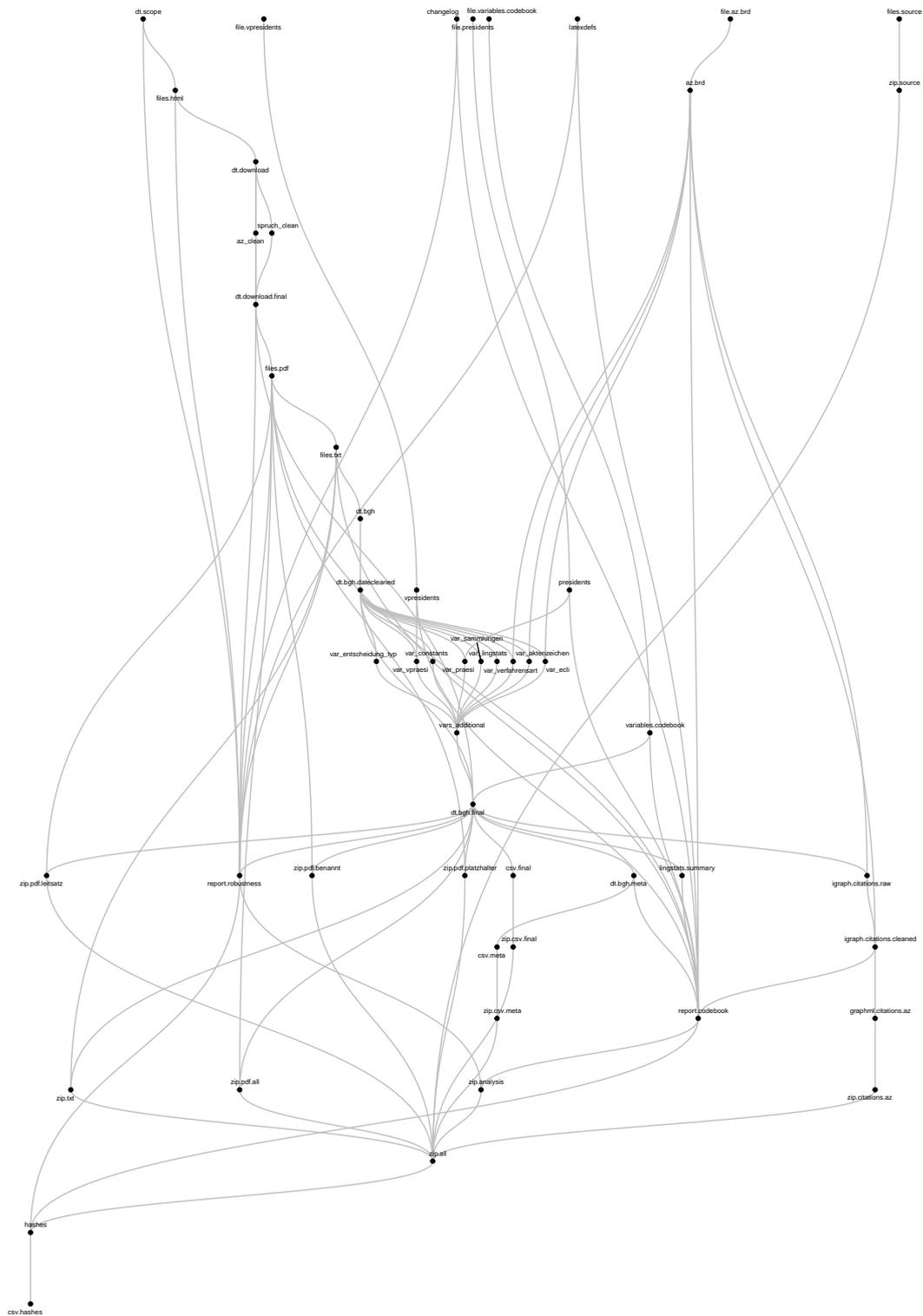


Abbildung 1: Der vollständige Prozess der Datensatz-Kompilierung.

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:

- Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*). Platzhalter für Nichtannahme- und Verwerfungsbeschlüsse ohne Begründung wurden automatisiert entfernt, sofern vorhanden.
- Es kann aufgrund technischer Grenzen bzw. Fehler sein, dass manche — im Grunde verfügbare — Entscheidungen nicht oder nicht korrekt abgerufen werden (*automation bias*).
- Es werden nur PDF-Dateien abgerufen (*file type bias*). Manche Entscheidungen sind nur als HTML verfügbar.
- Erst ab dem 1. Januar 2000 sind begründete Entscheidungen des Bundesgerichtshofs 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 in den Dateinamen sind größtenteils unverändert von den jeweiligen Datenbankeinträgen aus der amtlichen Datenbank des Bundesgerichtshofs entnommen. Berechnet und hinzugefügt wurden durch den Autor des Datensatzes 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, Spruchkörperangabe aus der Datenbank, die Angabe ob es sich um eine Leitsatzentscheidung handelt, Datum, das offizielle Aktenzeichen, einen Zusatz zum Aktenzeichen, ggf. einen besonderen Namen und eine durch den Autor des Datensatzes generierte Kollisions-ID.

### 3.7.2 Schema für die Dateinamen

[gericht]\_[spruchkoerper\_db]\_[leitsatz]\_[datum]\_[spruchkoerper\_az]\_  
[registerzeichen]\_[eingangsnummer]\_[eingangsjahr\_az]\_[zusatz\_az]\_  
[name]\_[kollision]

### 3.7.3 Beispiel eines Dateinamens

BGH\_Zivilsenat-1\_LE\_2007-10-04\_I\_ZR\_22\_05\_NA\_Umsatzsteuerhinweis\_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 Richter:innen des Bundesgerichtshofs ist »karmesinrot«. Der Hex-Wert hierfür ist vermutlich #7e0731. Das ist besonders bei der Erstellung thematisch passender Diagrammen hilfreich. Alle im Compilation Report und diesem Codebook präsentierten Diagramme sind in diesem karmesinrot gehalten.

## 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	<b>Traditionelle juristische Forschung.</b> Die PDF-Dokumente wie sie vom BGH 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. Neben dem vollen Datensatz sind für Praktiker:innen auch Varianten aufbereitet, die nur <i>Leitsatzentscheidungen</i> oder nur <i>besonders wichtige Entscheidungen mit Namen</i> enthalten.
CSV_Datensatz	<b>Legal Tech/Quantitative Forschung.</b> 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	<b>Legal Tech/Quantitative Forschung.</b> 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.
TXT	<b>Subsidiär für alle Zielgruppen.</b> 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
Netzwerke	<b>Alle Lehrenden und Forschenden. EXPERIMENTELL!</b> Das Zitationsnetzwerk des BGH. Umfasst sind nur Zitate von Aktenzeichen des BGH zu Aktenzeichen des BGH (d.h. zwischen Verfahren). Eine Auflösung auf Ebene einzelner Entscheidungen ist geplant. Kann mit statistischer Software oder aber auch GUI-Anwendungen wie Gephi analysiert werden.
ANALYSE	<b>Alle Lehrenden und Forschenden.</b> 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

### 5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 72517 obs. of 34 variables:
## $ doc_id : chr "BGH_Anwaltsenat_LE_2000-02-14_NA_AnwZB_12_99_NA_NA
_0.txt" "BGH_Anwaltsenat_LE_2000-02-14_NA_AnwZB_9_99_NA_NA_0.txt" "BGH_
Anwaltsenat_LE_2000-03-13_NA_AnwZB_28_99_NA_NA_0.txt" "BGH_Anwaltsenat_LE
_2000-05-29_NA_AnwZB_33_99_NA_NA_0.txt" ...
## $ url : chr "https://juris.bundesgerichtshof.de/cgi-bin/
rechtsprechung/document.py?Gericht=bgh&Art=en&Datum=2000&Sort=1&Seit"| __
truncated__ "https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/
document.py?Gericht=bgh&Art=en&Datum=2000&Sort=1&Seit"| __truncated__ "https
://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&
Art=en&Datum=2000&Sort=1&Seit"| __truncated__ "https://juris.
bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&
Datum=2000&Sort=1&Seit"| __truncated__ ...
## $ gericht : chr "BGH" "BGH" "BGH" "BGH" ...
## $ datum : IDate, format: "2000-02-14" "2000-02-14" ...
## $ entscheidungsjahr: int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000
...
## $ entscheidung_typ : chr "B" "B" "B" "B" ...
## $ leitsatz : chr "LE" "LE" "LE" "LE" ...
## $ spruchkoerper_db : chr "Anwaltsenat" "Anwaltsenat" "Anwaltsenat" "
Anwaltsenat" ...
## $ spruchkoerper_az : chr NA NA NA NA ...
## $ registerzeichen : chr "AnwZB" "AnwZB" "AnwZB" "AnwZB" ...
## $ verfahrensart : chr "Beschwerden gegen Entscheidungen eines
Anwaltsgerichtshofes" ...
## $ eingangsnummer : int 12 9 28 33 59 44 55 63 65 71 ...
## $ eingangsjahr_az : int 99 99 99 99 99 99 99 99 99 99 ...
## $ eingangsjahr_iso : num 1999 1999 1999 1999 1999 ...
## $ zusatz_az : chr NA NA NA NA ...
## $ name : chr NA NA NA NA ...
## $ kollision : int 0 0 0 0 0 0 0 0 0 0 ...
## $ aktenzeichen : chr "AnwZ(B) 12/99" "AnwZ(B) 9/99" "AnwZ(B) 28/99" "
AnwZ(B) 33/99" ...
## $ ecli : chr "ECLI:DE:BGH:2000:140200ANWZ.B.12.99.0" "ECLI:DE:
BGH:2000:140200ANWZ.B.9.99.0" "ECLI:DE:BGH:2000:130300ANWZ.B.28.99.0" "ECLI:
DE:BGH:2000:290500ANWZ.B.33.99.0" ...
## $ praesi : chr "Geiss" "Geiss" "Geiss" "Geiss" ...
## $ v_praesi : chr "Jähnke" "Jähnke" "Jähnke" "Jähnke" ...
## $ bemerkung : chr "Leitsatzentscheidung" "Leitsatzentscheidung" "
Leitsatzentscheidung" "Leitsatzentscheidung" ...
## $ bghst : num 0 0 0 0 0 0 0 0 0 0 ...
## $ bghr : num 0 0 0 0 0 0 0 0 0 0 ...
## $ nachschlagewerk : num 1 1 1 1 1 1 1 1 1 1 ...
## $ berichtigung : logi FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ zeichen : int 12566 13567 10306 12738 13120 10229 18019 9681
16189 11739 ...
```

```
## $ tokens      : int  1979 2210 1596 1884 1891 1564 2623 1481 2670 1741
...
## $ typen      : int  652 704 558 620 707 548 828 518 809 583 ...
## $ saetze    : int  107 161 95 86 89 102 141 85 165 94 ...
## $ version   : chr  "2023-03-10" "2023-03-10" "2023-03-10" "2023-03-10"
...
## $ doi_concept : chr  "10.5281/zenodo.3942742" "10.5281/zenodo.3942742"
"10.5281/zenodo.3942742" "10.5281/zenodo.3942742" ...
## $ doi_version : chr  "10.5281/zenodo.7699032" "10.5281/zenodo.7699032"
"10.5281/zenodo.7699032" "10.5281/zenodo.7699032" ...
## $ lizenz     : chr  "Creative Commons Zero 1.0 Universal" "Creative
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

- **Doppelte Codierung der Spruchkörper** — Für viele Urteile sind die Spruchkörper doppelt enthalten, einmal aus der Datenbank in menschenlesbarer Form (Variable »spruchkoerper\_db«), einmal durch das Aktenzeichen (Variable »spruchkoerper\_az«). Nicht wenige Aktenzeichen enthalten keine separate und einheitliche Angabe des Spruchkörpers, weil sich diese schon aus dem Registerzeichen ergibt, z.B. für Aktenkontrollen in Haftprüfungsverfahren oder Kartellsachen. Die Variable »spruchkoerper\_db« ermöglicht hier eine einfache Selektion bestimmter Spruchkörper anhand einer einheitlichen Variable, ohne im Detail erst jedes passende Registerzeichen suchen zu müssen.
- **Abweichende Codierung der Registerzeichen** — Die Registerzeichen wurden gekürzt und Sonderzeichen entfernt um die Arbeit mit ihnen zu vereinfachen. Beachten Sie bitte die Gegenüberstellungstabelle unter Punkt 6.
- **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. Die einzige editorische Änderung gegenüber den TXT-Dateien ist die Zusammenfügung von über Zeilengrenzen gebrochenen Wörtern.
url	String	(Nur CSV) Link zum Volltext der Entscheidung in der Datenbank.

(continued)

Variable	Typ	Erläuterung
gericht	String	Name des Gerichts. Es ist nur der Wert »BGH« vergeben. Dies ist der ECLI-Code für »Bundesgerichtshof«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Date	Datum der Entscheidung im Format YYYY-MM-DD (ISO-8601).
entscheidungsjahr	Integer	(Nur CSV) Jahr der Entscheidung im Format YYYY (ISO-8601).
entscheidung_typ	String	(Nur CSV-Datei) Der Typ der Entscheidung. Es sind die Werte »B« (Beschluss), »U« (Urteil) und »V« (Verfügung) vergeben. Wurde durch <i>regular expressions</i> aus der Variable »text« berechnet.
leitsatz	String	Ob es sich um eine Leitsatzentscheidung handelt. Wenn ja ist der Wert »LE«, ansonsten »NA«. Wurde für jede Entscheidung bejaht, die in der Variable »bemerkung« die Zeichenfolge »Leitsatz« enthält.
spruchkoerper_db	String	Der Spruchkörper wie er in der amtlichen Datenbank des Bundesgerichtshofs eingetragen ist. Gekürzt auf eine für Dateinamen angemessene Länge und konvertiert in arabische Zahlen. Die übrigen Senate haben eindeutige Namen.
spruchkoerper_az	String	Der im Aktenzeichen angegebene Spruchkörper. Arabische Ziffern stehen für die Strafsenate, römische Ziffern für die Zivilsenate. Die Spezialsenate und großen Senate sind in dieser Variable nicht codiert (weil über die Registerzeichen eindeutig identifiziert) und weisen daher immer den Wert »NA« auf. Nutzen Sie für Analysen aller Senate bitte die Variablen »spruchkoerper_db« oder »registerzeichen«.
registerzeichen	String	Das amtliche Registerzeichen. Eine Erläuterung der Abkürzungen findet sich im Abschnitt 6.
verfahrensart	String	Die Verfahrensart, auf die das Registerzeichen hinweist. Siehe auch Abschnitt 6.
eingangsnummer	Integer	Eingangsnummer. Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine fortlaufende Nummer (Ordinalzahl) in der Reihenfolge ihres Eingangs.

(continued)

Variable	Typ	Erläuterung
eingangsjahr_az	Integer	Eingangsjahr laut Aktenzeichen. Das Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Integer	Eingangsjahr im Format YYYY-MM-DD (ISO-8601).
zusatz_az	String	Der Zusatz zu einem Aktenzeichen, falls vorhanden.
name	String	Der Name der Entscheidung wie er in der amtlichen Datenbank angegeben ist. In den CSV-Dateien ist er identisch mit dem Datenbankeintrag, in den Dateinamen der PDF- und TXT-Dateien wurden bestimmte Sonderzeichen bereinigt um Kompatibilitätsprobleme zu vermeiden.
kollision	Integer	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. <b>ACHTUNG:</b> diese Variable ist von dem Autor des Datensatzes berechnet und keine amtliche Angabe. Die Nutzung für die Erstellung der ECLI ist daher problematisch.
aktenzeichen	String	Das amtliche Aktenzeichen im Format [senatsnummer] [registerzeichen] [eingangsnummer] / [eingangsjahr] [ggf. zusatz_az]

(continued)

Variable	Typ	Erläuterung
ecli	String	(Nur CSV-Datei) <b>Experimentell!</b> 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« und »kollision« durch den Autor des Datensatzes berechnet. <b>ACHTUNG:</b> Da keine amtlichen Angaben zur Kollisionsvariable vorliegen sind die ECLIs für alle Entscheidungen mit Variable »kollision« größer 0 und ihre verwandten Entscheidungen mit Variable »kollision« gleich 0 mit hoher Wahrscheinlichkeit fehlerhaft. Für die weit überwiegende Anzahl der Entscheidungen gibt es jedoch keine Kollisionen und die ECLIs sollten funktionstüchtig sein. Amtlich vergeben werden jedoch nur ECLIs für Entscheidungen ab dem 1. Januar 2016.
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.
bemerkung	String	Zusätzliche Bemerkungen aus der Datenbank des Bundesgerichtshofs. Enthält in der Regel Informationen dazu ob es sich um eine Leitsatzentscheidung handelt, welchen Eigennamen die Entscheidung trägt und wann welche Pressemitteilung darüber herausgegeben wurde.
bghst	Binary	(Nur CSV-Datei) Ob die Entscheidung in die amtliche Sammlung BGHSt aufgenommen wird. »1« falls ja, »0« falls nein.

(continued)

Variable	Typ	Erläuterung
bghr	Binary	(Nur CSV-Datei) Ob die Entscheidung in die Sammlung BGHR aufgenommen wird. »1« falls ja, »0« falls nein.
nachschlagewerk	Binary	(Nur CSV-Datei) Ob die Entscheidung in das Nachschlagewerk aufgenommen wird. »1« falls ja, »0« falls nein.
berichtigung	Logical	(Nur CSV-Datei) Eine Angabe ob es sich um einen Berichtigungsbeschluss handelt, TRUE wenn ja, sonst FALSE. Wurde durch den Autor des Datensatzes aus der Variable »bemerkung« extrahiert.
zeichen	Integer	(Nur CSV-Datei) Die Anzahl Zeichen eines Dokumentes.
tokens	Integer	(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	Integer	(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	Integer	(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	Date	(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 <a href="http://www.doi.org">www.doi.org</a> immer die <b>aktuellste Version</b> 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 <b>konkreten Version</b> des Datensatzes. Dieser ist langzeit-stabil (persistent). Über diese DOI kann via <a href="http://www.doi.org">www.doi.org</a> 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 Registerzeichen

Die Registerzeichen des BGH sind in der *Anweisung für die Verwaltung des Schriftguts in Rechtssachen bei der Geschäftsstelle des Bundesgerichtshofes und des Generalbundesanwalts beim Bundesgerichtshof - Aktenordnung Bundesgerichtshof* definiert. Die im Datensatz enthaltenen Registerzeichen wurden jeweils um die runden Klammern bereinigt, um Probleme bei der Nutzung unter Windows zu vermeiden.

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
AK	Aktenkontrolle für Haftprüfungsverfahren
AR	Allgemeines Register
AR(Anw)	Allgemeines Register (Anwaltssachen)
AR(Not)	Allgemeines Register (Notarsachen)
AR(Ri)	Allgemeines Register (Dienstgericht des Bundes)
AR(VS)	Entscheidungen über Justizverwaltungsakte (Strafsachen)
AR(VZ)	Entscheidungen über Justizverwaltungsakte (Zivilsachen)
AR(Vollz)	Vorlegungssachen (Strafvollzugssachen)
ARZ	Allgemeines Register und Gerichtsstandsbestimmungen (Zivilsachen)
ARs	Allgemeines Register und Gerichtsstandsbestimmungen (Strafsachen)
AnwSt	Anträge auf Einleitung des anwaltsgerichtlichen Verfahrens und Beschwerden in Ordnungsmittelverfahren gegen Rechtsanwälte beim Bundesgerichtshof
AnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwSt(R)	Revisionen gegen Urteile eines Anwaltsgerichtshofes
AnwZ	Klagen über Entscheidungen in Zulassungssachen oder gegen sonstige Verwaltungsakte betreffend Rechtsanwälte beim Bundesgerichtshof
AnwZ(B)	Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen eines Anwaltsgerichtshofes
AnwZ(P)	Klagen über die Anfechtung von Wahlen und Beschlüssen der Rechtsanwaltskammer beim BGH und der Bundesrechtsanwaltskammer
BGs	Einzelne richterliche Anordnungen des Ermittlungsrichters (Strafsachen)

BLw	Rechtsbeschwerden in Landwirtschaftssachen der freiwilligen Gerichtsbarkeit
EnRB	Rechtsbeschwerden in Bußgeldverfahren nach dem EnWG
EnVR	Rechtsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnVZ	Nichtzulassungsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Rechtsstreits in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
GSSt	Großer Senat (Strafsachen)
GSZ	Großer Senat (Zivilsachen)
GmSOGB	Gemeinsamer Senat der Obersten Gerichtshöfe des Bundes
KRB	Rechtsbeschwerden in Kartellbußgeldverfahren
KVR	Rechtsbeschwerden (Kartellverwaltungssachen)
KVZ	Nichtzulassungsbeschwerden (Kartellverwaltungssachen)
KZA	Anträge außerhalb eines in der Revisionsinstanz anhängigen Rechtsstreits (Kartellsachen)
KZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
KZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
LwZA	Anträge außerhalb eines in der Revisionsinstanz für Landwirtschaftssachen anhängigen Verfahrens
LwZB	Beschwerden und Rechtsbeschwerden in Landwirtschaftssachen der streitigen bürgerlichen Gerichtsbarkeit
LwZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision (Landwirtschaftssachen)
NotSt(B)	Beschwerden gegen Beschlüsse der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotSt(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Urteile der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotZ	Verwaltungsstreitverfahren in Notarsachen und Beschwerden gegen Entscheidungen der Oberlandesgerichte
NotZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen der Oberlandesgerichte in Notarsachen

PatAnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden nach der PatAO
PatAnwSt(R)	Revisionen nach der PatAO
PatAnwZ	Verwaltungsstreitverfahren in Patentanwaltssachen; Berufungen und Anträge auf Zulassung der Berufung und Beschwerden gegen Entscheidungen der Oberlandesgerichte
RiSt	Disziplinarsachen und Anträge auf Einleitung oder Einstellung des förmlichen Disziplinarverfahrens, auf vorläufige Dienstenthebung, auf Einbehaltung von Dienstbezügen sowie auf Aufhebung dieser Maßnahmen gegen Richter und Staatsanwälte im Bundesdienst sowie Mitglieder des Bundesrechnungshofes
RiSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden der Richter, Staatsanwälte und Mitglieder des Bundesrechnungshofes gegen Disziplinarverfügungen
RiSt(R)	Revisionen in Disziplinarsachen nach dem Deutschen Richtergesetz
RiZ	Anträge betreffend Richter im Bundesdienst und Mitglieder des Bundesrechnungshofes auf gerichtliche Entscheidung im Versetzungs- und Prüfungsverfahren sowie auf vorläufige Untersagung der Amtsgeschäfte
RiZ(B)	Beschwerden gegen die Nichtzulassung der Revision nach dem Deutschen Richtergesetz
RiZ(R)	Revisionen in Versetzungs- und Prüfungsverfahren nach dem Deutschen Richtergesetz
StB	Beschwerden (Strafsachen)
StR	Revisionen und Vorlegungssachen nach § 121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (Strafsachen)
StbSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigten-sachen)
StbSt(R)	Revisionen in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigten-sachen)
VGS	Vereinigte Große Senate
VRG	Ersuchen des Bundesverfassungsgerichts zu Vorlagen nach Art. 100 GG und zu Verfassungsbeschwerden
WpSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
WpSt(R)	Revisionen in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
ZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Verfahrens (Zivilsachen)
ZB	Beschwerden, Rechtsbeschwerden, weitere Beschwerden, Beschwerden gegen die Nichtzulassung der Revision nach dem BEG (Zivilsachen)

ZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision, Anträge auf Zulassung der Sprungrevision, Berufungen in Patent-sachen (Zivilsachen)
ZR(Ü)	Erstinstanzliche Klagen auf Entschädigung wegen überlanger Ge-richtsverfahren und strafrechtlicher Ermittlungsverfahren (Zivilsa-chen)

---

## 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
Weinkauff	Hermann	1950-10-01	1960-03-31	1894-02-10	1981-07-09
Heusinger	Bruno	1960-04-01	1968-03-31	1900-03-02	1987-08-03
Fischer	Robert	1968-04-01	1977-09-30	1911-08-22	1983-03-04
Pfeiffer	Gerd	1977-10-01	1987-12-31	1919-12-22	2007-02-01
Odersky	Walter	1988-01-01	1996-07-31	1931-07-17	NA
Geiss	Karlmann	1996-08-01	2000-05-31	1935-05-31	NA
VACANCY-1	VACANCY-1	2000-06-01	2000-07-14	NA	NA
Hirsch	Günter	2000-07-15	2008-01-31	1943-01-30	NA
Tolksdorf	Klaus	2008-02-01	2014-01-31	1948-11-14	NA
VACANCY-2	VACANCY-2	2014-02-01	2014-06-30	NA	NA
Limberg	Bettina	2014-07-01	NA	1960-04-05	NA

### 7.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtsantritt)	Alter (Amtsende)	Alter (Tod)
Weinkauff	Hermann	56	66	87
Heusinger	Bruno	60	68	87
Fischer	Robert	56	66	71
Pfeiffer	Gerd	57	68	87
Odersky	Walter	56	65	NA
Geiss	Karlmann	61	65	NA
Hirsch	Günter	57	65	NA
Tolksdorf	Klaus	59	65	NA
Limberg	Bettina	54	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
Glanzmann	Roderich	1965-05-17	1972-04-30	1904-04-15	1988-08-07
VACANCY-1	VACANCY-1	1972-05-01	1972-05-22	NA	NA
Hauß	Fritz	1972-05-23	1976-10-31	1908-10-27	2003-04-30
VACANCY-2	VACANCY-2	1976-11-01	1976-11-02	NA	NA
Pfeiffer	Gerd	1976-11-03	1977-09-30	1919-12-22	2007-02-01
Stimpel	Walter	1977-10-01	1985-11-30	1917-11-29	2008-01-13
VACANCY-3	VACANCY-3	1985-12-01	1985-12-01	NA	NA
Thumm	Ludwig	1985-12-02	1988-04-30	1920-04-13	2011-04-11
Salger	Hannskarl	1988-05-01	1994-11-30	1929-11-02	2010-09-17
Hagen	Horst	1994-12-01	1999-02-28	1934-01-05	2019-12-05
Jähnke	Burkhard	1999-03-01	2002-05-31	1937-05-14	NA
Wenzel	Joachim	2002-06-01	2005-06-30	1940-06-23	2009-08-29
Müller	Gerda	2005-07-01	2009-06-30	1944-06-26	NA
Schlick	Wolfgang	2009-07-01	2015-07-31	1950-03-29	NA
VACANCY-4	VACANCY-4	2015-08-01	2016-12-01	NA	NA
Ellenberger	Jürgen	2016-12-02	NA	1960-04-19	NA

### 8.3 Dienstalter und Lebensalter

Nachname	Vorname	Alter (Amtesantritt)	Alter (Amtesende)	Alter (Tod)
Glanzmann	Roderich	61	68	84
Hauß	Fritz	63	68	94
Pfeiffer	Gerd	56	57	87
Stimpel	Walter	59	68	90
Thumm	Ludwig	65	68	90
Salger	Hannskarl	58	65	80
Hagen	Horst	60	65	85
Jähnke	Burkhard	61	65	NA
Wenzel	Joachim	61	65	69
Müller	Gerda	61	65	NA
Schlick	Wolfgang	59	65	NA
Ellenberger	Jürgen	56	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.

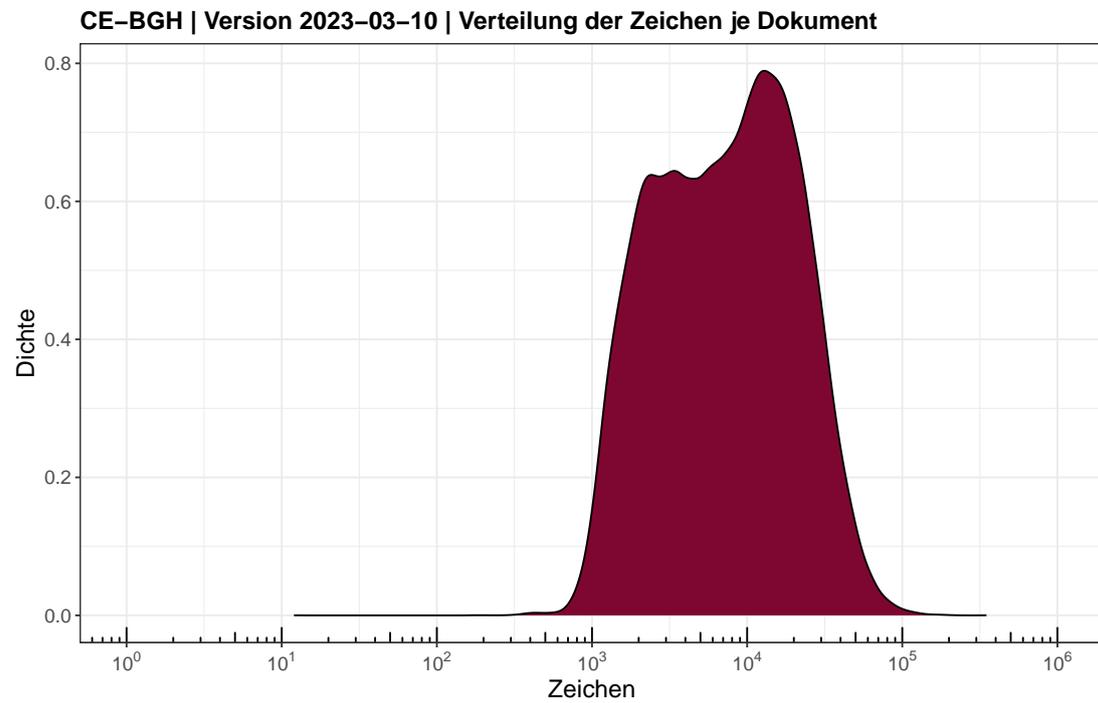
Es handelt sich bei den Diagrammen jeweils um »Density Charts«, die sich besonders dafür eignen die Schwerpunkte von Variablen mit stark schwankenden numerischen Werten zu visualisieren. Die Interpretation ist denkbar einfach: je höher die Kurve, desto dichter sind in diesem Bereich die Werte der Variable. 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.

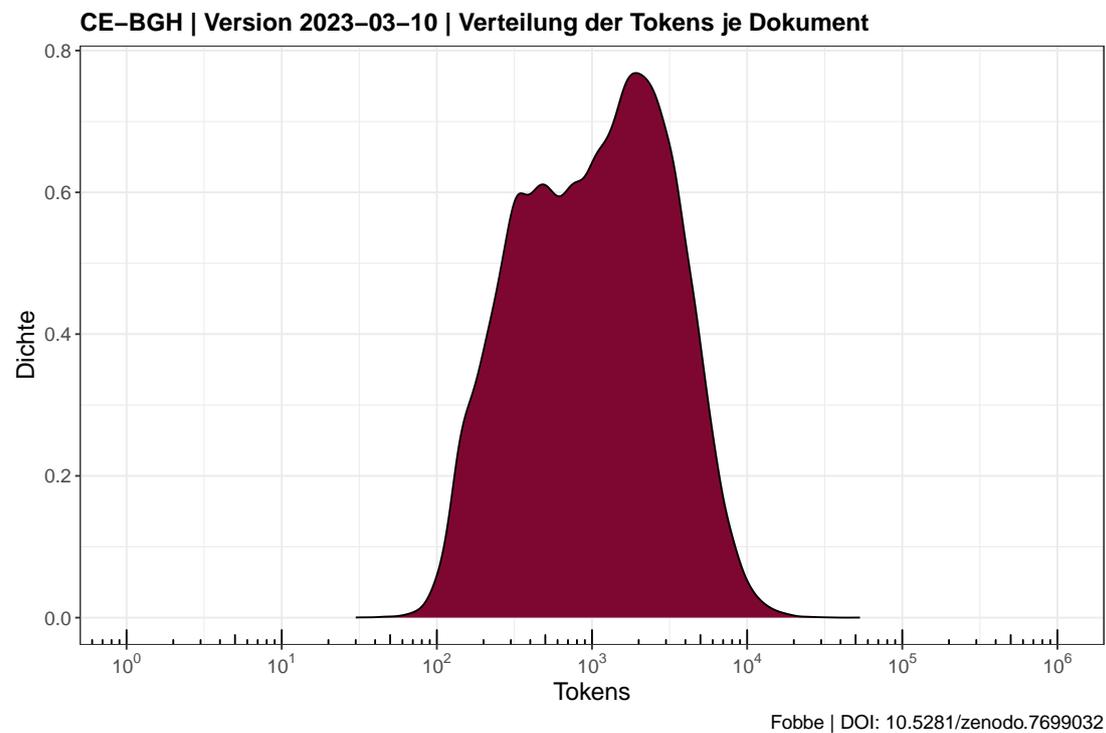
### 9.2 Werte der Kennzahlen

Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	839,547,003	12	3,050	7,459	11,577.24	16,085	348,611
tokens	126,394,005	0	432	1,103	1,742.96	2,421	53,371
typen	721,984	0	215	439	545.31	773	5,608
saetze	6,706,338	0	29	63	92.48	127	2,700

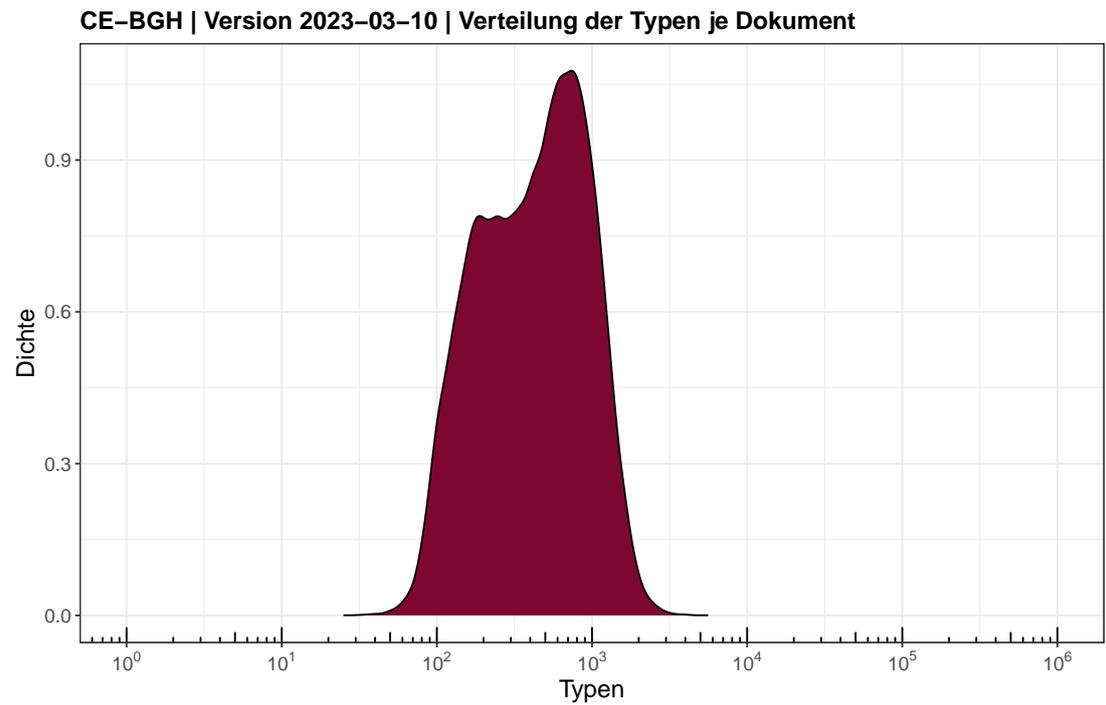
### 9.3 Verteilung Zeichen



### 9.4 Verteilung Tokens

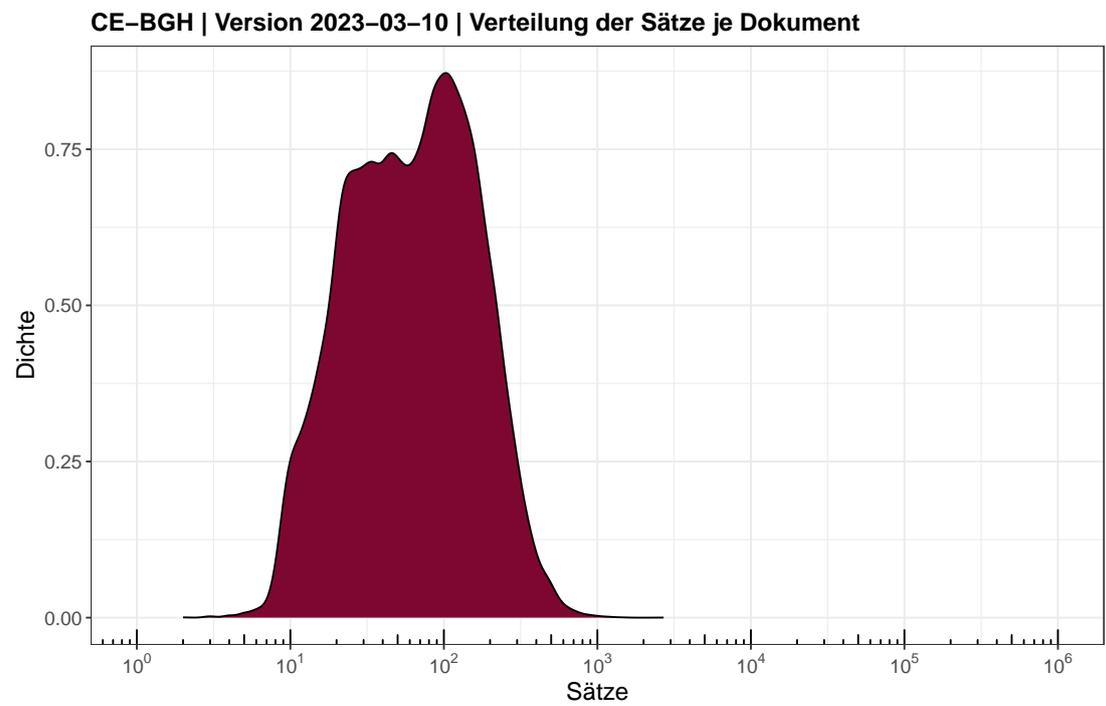


## 9.5 Verteilung Typen



Fobbe | DOI: 10.5281/zenodo.7699032

## 9.6 Verteilung Sätze



Fobbe | DOI: 10.5281/zenodo.7699032

## 10 Zitations-Netzwerke

Im Datensatz sind ab 2023 auch Zitate zwischen einzelnen Verfahren des BGH enthalten. Aus den Vollexten der BGH-Entscheidungen wurden alle Aktenzeichen des BGH extrahiert und mit dem Aktenzeichen der zitierenden Entscheidung verbunden.

**Achtung:** Es handelt sich nicht um Zitate von Entscheidung zu Entscheidung, sondern von Aktenzeichen zu Aktenzeichen. Die Auflösung der Zitate auf Entscheidungsebene ist geplant und wird in Zukunft mitveröffentlicht.

Beachten Sie auch bitte folgende Punkte:

- Das gesamte Netzwerk ist sehr groß (ca. 85.000 Knoten, ca. 300.000 Kanten) und die Analyse ist daher ohne weitere Einschränkungen rechenintensiv und anspruchsvoll. In der Regel sollten Sie das Netzwerk auf die für Sie interessanten Teile reduzieren.
- Zur Reduktion des Netzwerks auf eine handliche Größe stelle ich zwei Variablen bereit: den Senat und das Registerzeichen. Sie könnten beispielsweise über das Registerzeichen nur die Strafsenate betrachten oder über die Senatsangabe nur den Zivilsenat mit der für Sie interessanten Spezialmaterie
- Die Extraktion mit *regular expressions* ist nicht perfekt. Es kann daher sein, dass Zitate fehlen, wenn sie nicht als solche erkannt wurden, wegen Tippfehlern, ungewöhnlichem Textumfeld etc. Es ist aktuell unklar wieviele Zitate fehlen könnten, weil es keinen Goldstandard zum Abgleich gibt. Wenn Ihnen größere Fehlbestände auffallen, melden Sie sich bitte.
- Alle Knoten die nach der Extraktion kein gültiges Aktenzeichen des BGH enthalten wurden entfernt, das betrifft nur eine kleine zweistellige Zahl (siehe Compilation Report für den genauen Wert)

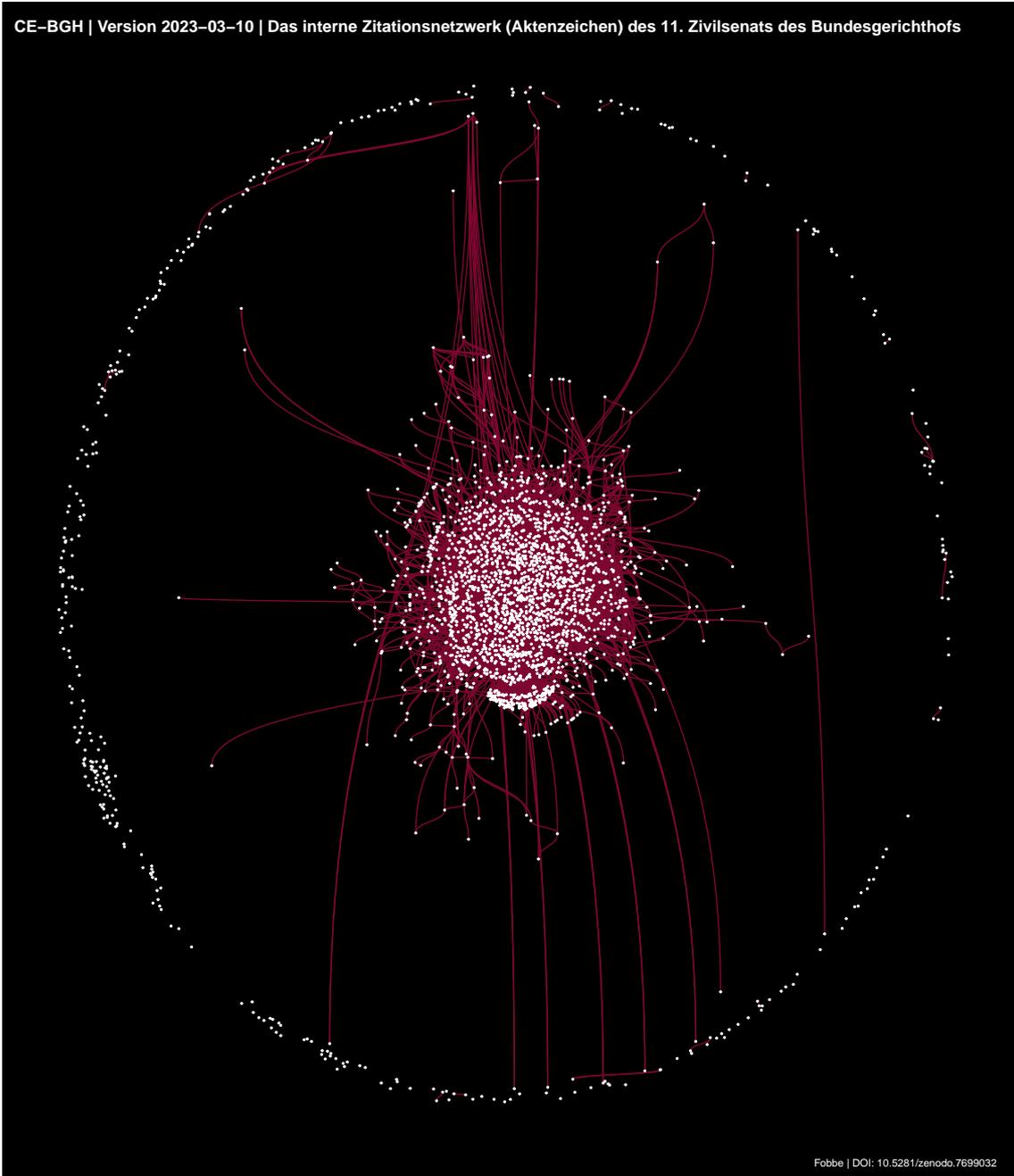


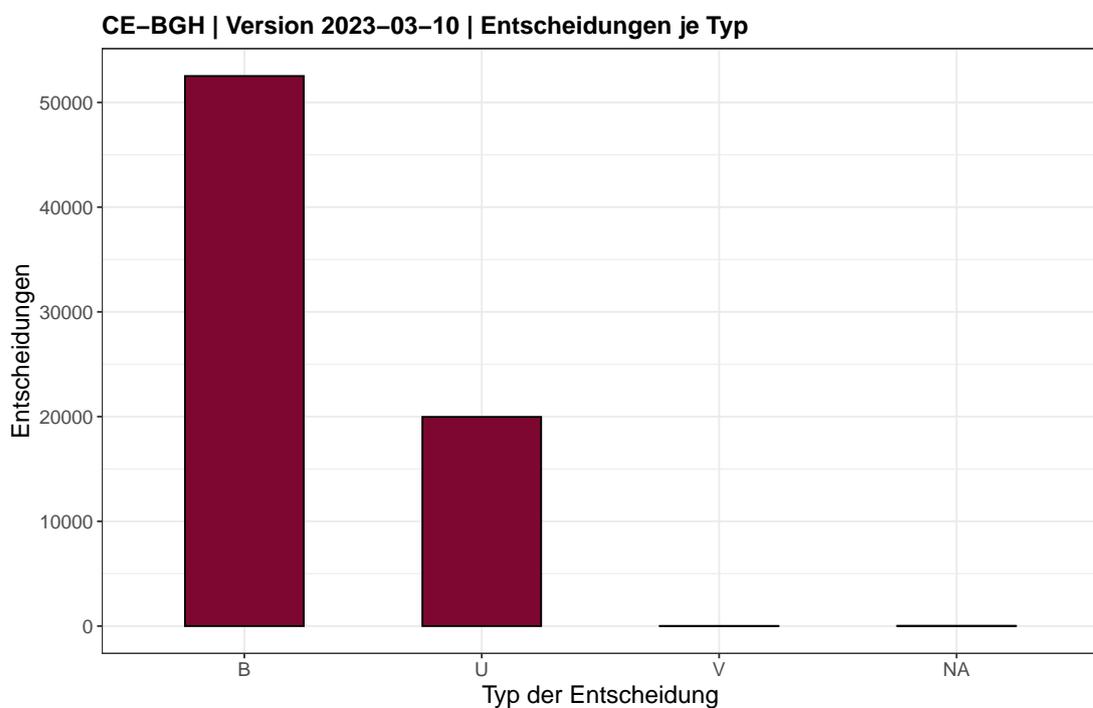
Abbildung 2: Das interne Zitationsnetzwerk (Aktenzeichen) des 11. Zivilsenats des Bundesgerichtshofs (Kamada Kawai-Layout)

## 11 Inhalt des Korpus

### 11.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mean	Quart3	Max
entscheidungsjahr	24	2000	2006	2011	2011.38	2017	2023
eingangsjahr_iso	35	1980	2005	2010	2010.52	2016	2023
eingangsnummer	832	1	55	156	198.93	304	1352

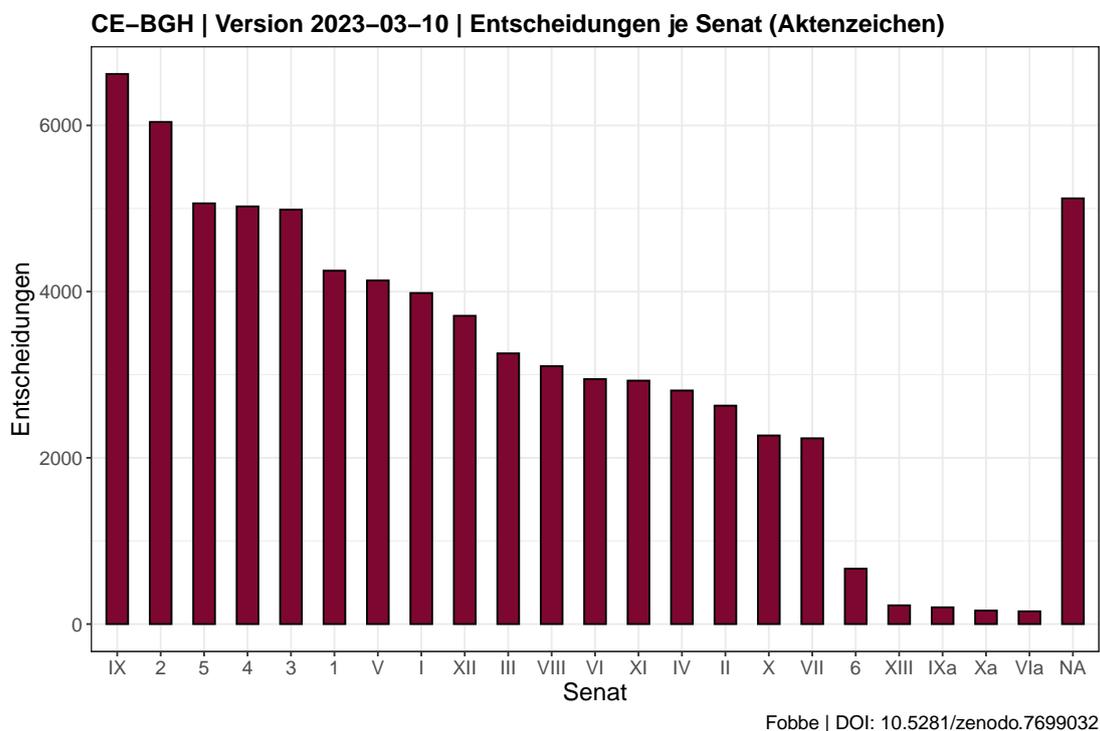
### 11.2 Nach Typ der Entscheidung



Fobbe | DOI: 10.5281/zenodo.7699032

Typ	Entscheidungen	% Gesamt	% Kumulativ
NA	20	0.03	0.03
B	52519	72.42	72.45
U	19969	27.54	99.99
V	9	0.01	100.00
Total	72517	100.00	100.00

### 11.3 Nach Spruchkörper (Aktenzeichen)

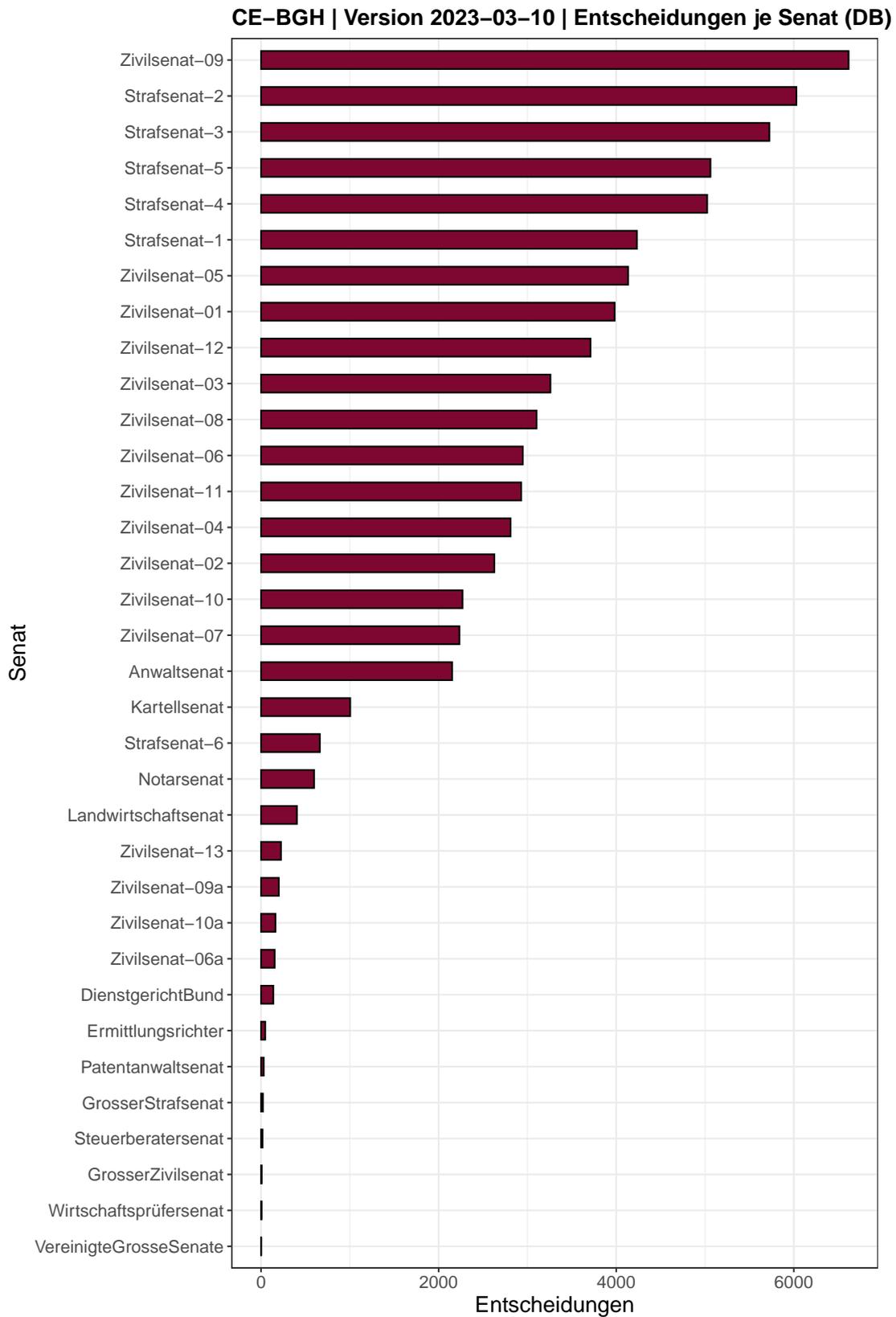


Senat	Entscheidungen	% Gesamt	% Kumulativ
NA	5122	7.06	7.06
1	4252	5.86	12.93
2	6041	8.33	21.26
3	4986	6.88	28.13
4	5025	6.93	35.06
5	5062	6.98	42.04
6	666	0.92	42.96
I	3983	5.49	48.45
II	2628	3.62	52.08
III	3257	4.49	56.57
IV	2810	3.87	60.44
IX	6618	9.13	69.57
IXa	201	0.28	69.85
V	4134	5.70	75.55
VI	2947	4.06	79.61
VII	2235	3.08	82.69
VIII	3103	4.28	86.97

*(continued)*

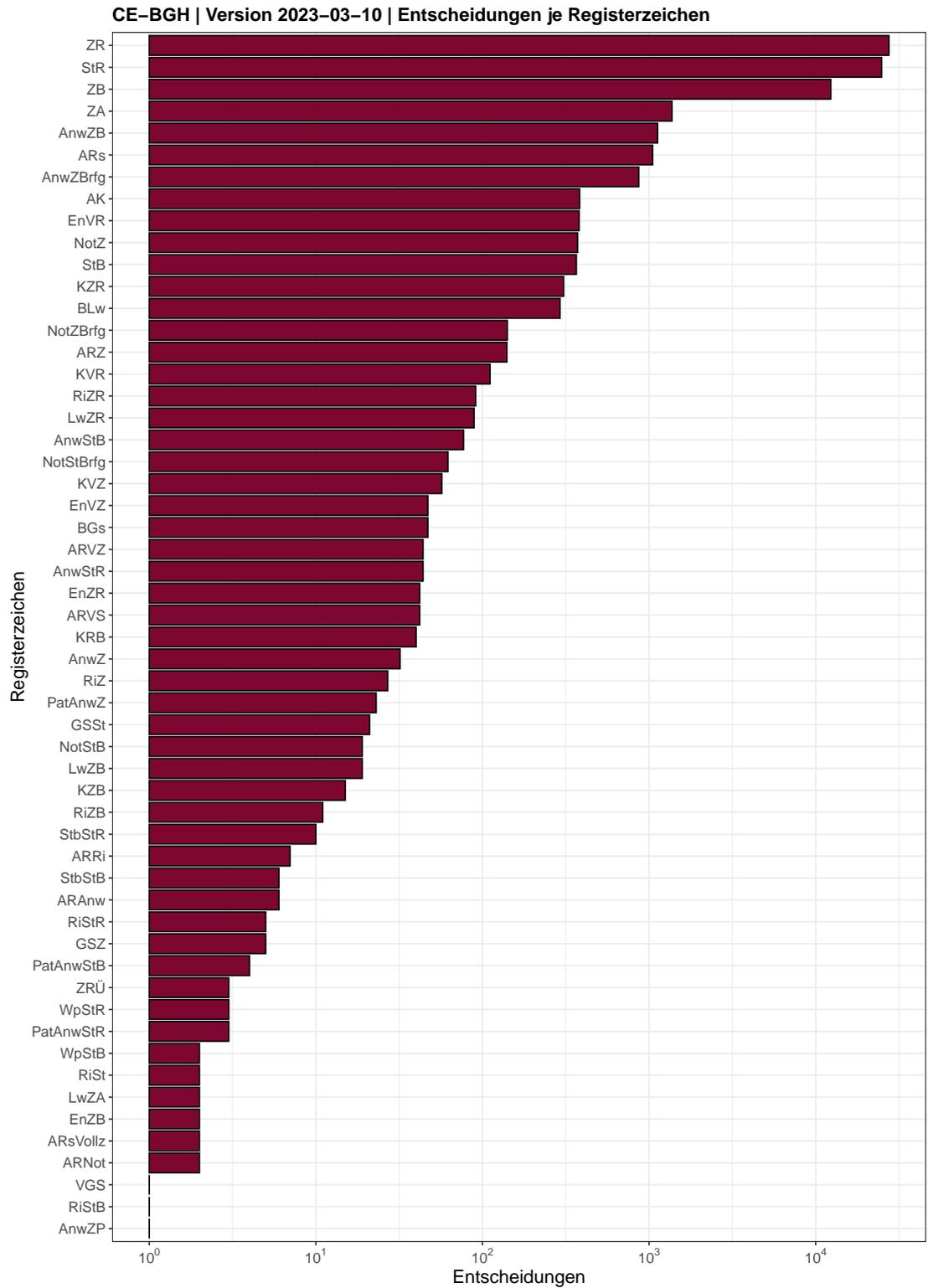
Senat	Entscheidungen	% Gesamt	% Kumulativ
VIa	153	0.21	87.18
X	2269	3.13	90.31
XI	2929	4.04	94.35
XII	3709	5.11	99.47
XIII	225	0.31	99.78
Xa	162	0.22	100.00
Total	72517	100.00	100.00

## 11.4 Nach Spruchkörper (Datenbank)



Senat	Entscheidungen	% Gesamt	% Kumulativ
Anwaltsenat	2152	2.97	2.97
DienstgerichtBund	138	0.19	3.16
Ermittlungsrichter	47	0.06	3.22
GrosserStrafsenat	21	0.03	3.25
GrosserZivilsenat	6	0.01	3.26
Kartellsenat	1004	1.38	4.64
Landwirtschaftsenat	404	0.56	5.20
Notarsenat	597	0.82	6.02
Patentanwaltsenat	30	0.04	6.07
Steuerberatersenat	16	0.02	6.09
Strafsenat-1	4233	5.84	11.93
Strafsenat-2	6029	8.31	20.24
Strafsenat-3	5725	7.89	28.13
Strafsenat-4	5024	6.93	35.06
Strafsenat-5	5061	6.98	42.04
Strafsenat-6	662	0.91	42.95
VereinigteGrosseSenate	1	0.00	42.96
Wirtschaftsprüfersenat	5	0.01	42.96
Zivilsenat-01	3982	5.49	48.45
Zivilsenat-02	2627	3.62	52.08
Zivilsenat-03	3259	4.49	56.57
Zivilsenat-04	2810	3.87	60.45
Zivilsenat-05	4134	5.70	66.15
Zivilsenat-06	2948	4.07	70.21
Zivilsenat-06a	153	0.21	70.42
Zivilsenat-07	2234	3.08	73.50
Zivilsenat-08	3103	4.28	77.78
Zivilsenat-09	6617	9.12	86.91
Zivilsenat-09a	200	0.28	87.18
Zivilsenat-10	2269	3.13	90.31
Zivilsenat-10a	162	0.22	90.53
Zivilsenat-11	2930	4.04	94.58
Zivilsenat-12	3710	5.12	99.69
Zivilsenat-13	224	0.31	100.00
Total	72517	100.00	100.00

## 11.5 Nach Registerzeichen

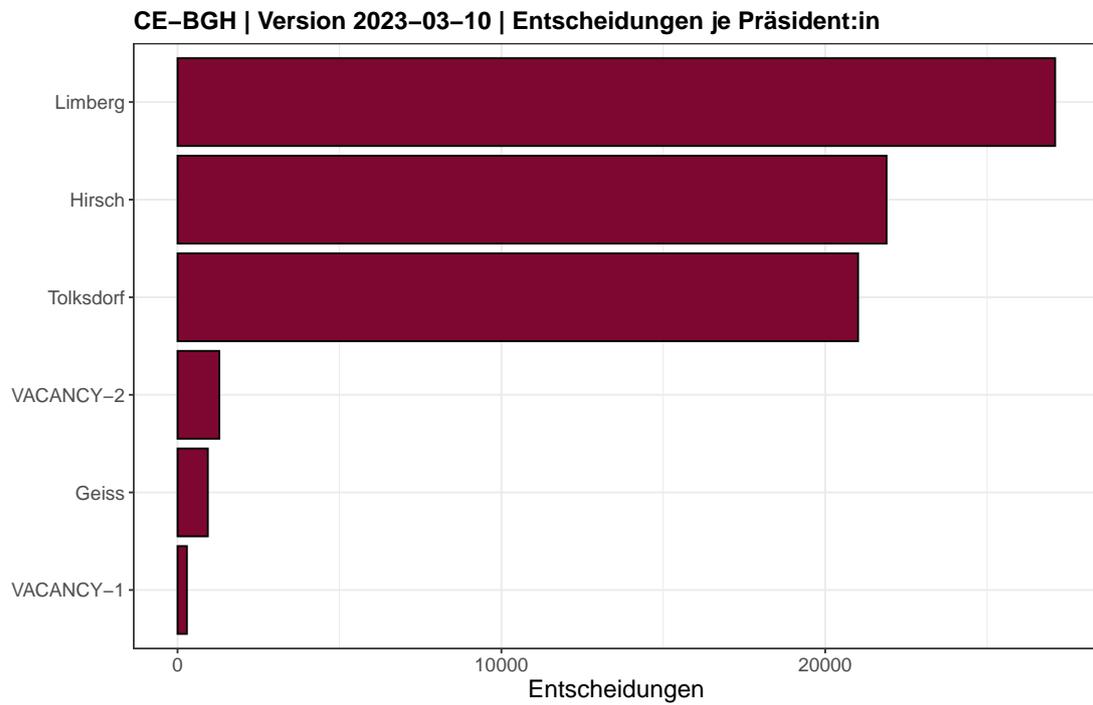


Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
AK	383	0.53	0.53
ARAnw	6	0.01	0.54
ARNot	2	0.00	0.54
ARRi	7	0.01	0.55
ARVS	42	0.06	0.61
ARVZ	44	0.06	0.67
ARZ	140	0.19	0.86
ARs	1051	1.45	2.31
ARsVollz	2	0.00	2.31
AnwStB	77	0.11	2.42
AnwStR	44	0.06	2.48
AnwZ	32	0.04	2.52
AnwZB	1125	1.55	4.07
AnwZBrfg	867	1.20	5.27
AnwZP	1	0.00	5.27
BGs	47	0.06	5.34
BLw	292	0.40	5.74
EnVR	380	0.52	6.26
EnVZ	47	0.06	6.33
EnZB	2	0.00	6.33
EnZR	42	0.06	6.39
GSSt	21	0.03	6.42
GSZ	5	0.01	6.42
KRB	40	0.06	6.48
KVR	111	0.15	6.63
KVZ	57	0.08	6.71
KZB	15	0.02	6.73
KZR	307	0.42	7.16
LwZA	2	0.00	7.16
LwZB	19	0.03	7.18
LwZR	89	0.12	7.31
NotStB	19	0.03	7.33
NotStBrfg	62	0.09	7.42
NotZ	372	0.51	7.93
NotZBrfg	141	0.19	8.13
PatAnwStB	4	0.01	8.13
PatAnwStR	3	0.00	8.14

(continued)

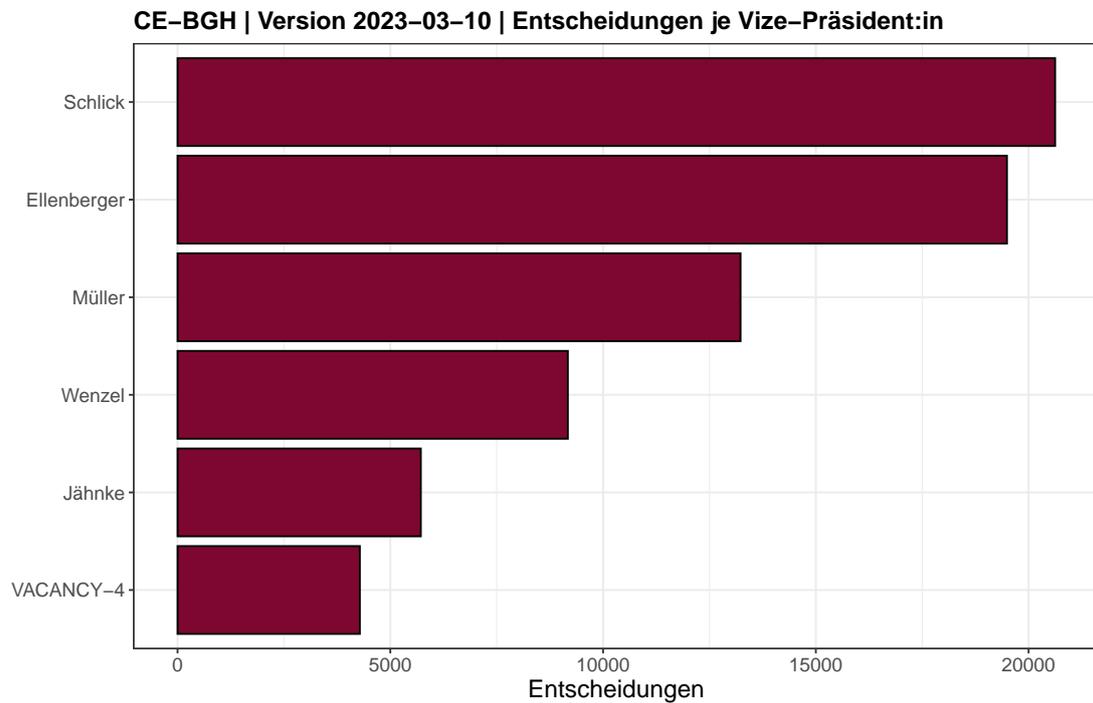
Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
PatAnwZ	23	0.03	8.17
RiSt	2	0.00	8.17
RiStB	1	0.00	8.17
RiStR	5	0.01	8.18
RiZ	27	0.04	8.22
RiZB	11	0.02	8.23
RiZR	91	0.13	8.36
StB	366	0.50	8.86
StR	24875	34.30	43.16
StbStB	6	0.01	43.17
StbStR	10	0.01	43.19
VGS	1	0.00	43.19
WpStB	2	0.00	43.19
WpStR	3	0.00	43.19
ZA	1373	1.89	45.09
ZB	12318	16.99	62.07
ZR	27500	37.92	100.00
ZRÜ	3	0.00	100.00
Total	72517	100.00	100.00

## 11.6 Nach Präsident:in



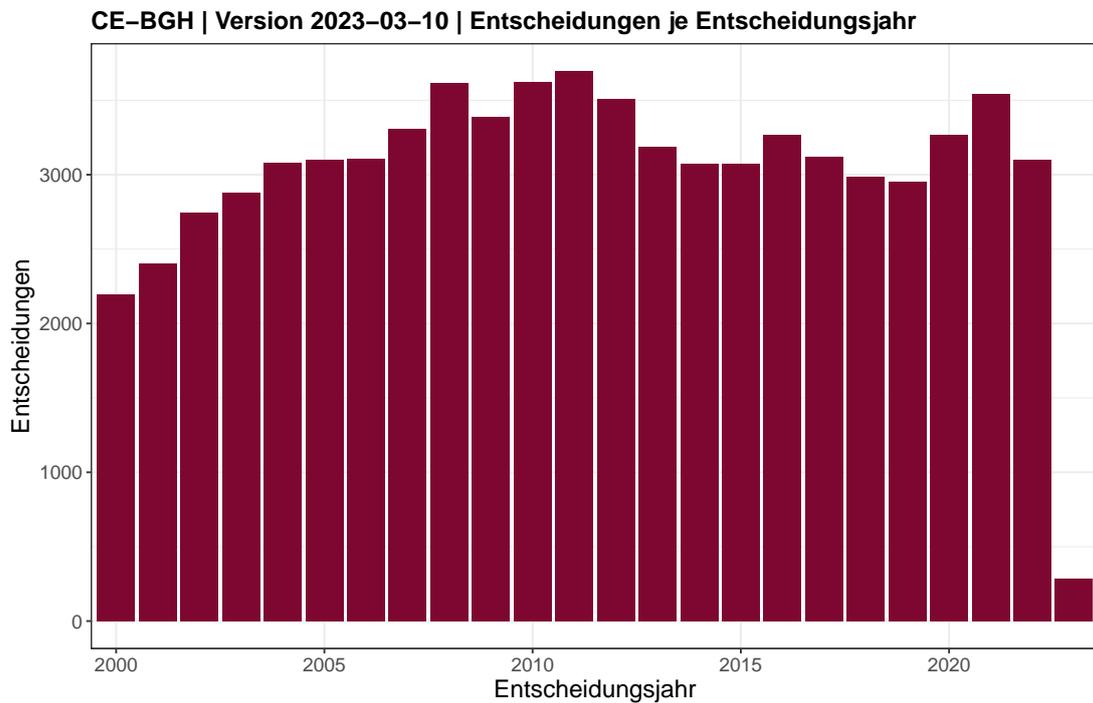
Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Geiss	935	1.29	1.29
Hirsch	21893	30.19	31.48
Limberg	27098	37.37	68.85
Tolksdorf	21014	28.98	97.83
VACANCY-1	288	0.40	98.22
VACANCY-2	1289	1.78	100.00
Total	72517	100.00	100.00

## 11.7 Nach Vize-Präsident:in



Vize-Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Ellenberger	19491	26.88	26.88
Jähnke	5717	7.88	34.76
Müller	13230	18.24	53.01
Schlick	20623	28.44	81.44
VACANCY-4	4284	5.91	87.35
Wenzel	9172	12.65	100.00
Total	72517	100.00	100.00

## 11.8 Nach Entscheidungsjahr



Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2000	2197	3.03	3.03
2001	2404	3.32	6.34
2002	2746	3.79	10.13
2003	2881	3.97	14.10
2004	3079	4.25	18.35
2005	3098	4.27	22.62
2006	3106	4.28	26.91
2007	3310	4.56	31.47
2008	3614	4.98	36.45
2009	3389	4.67	41.13
2010	3625	5.00	46.13
2011	3697	5.10	51.22
2012	3506	4.83	56.06
2013	3190	4.40	60.46
2014	3071	4.23	64.69
2015	3074	4.24	68.93
2016	3271	4.51	73.44

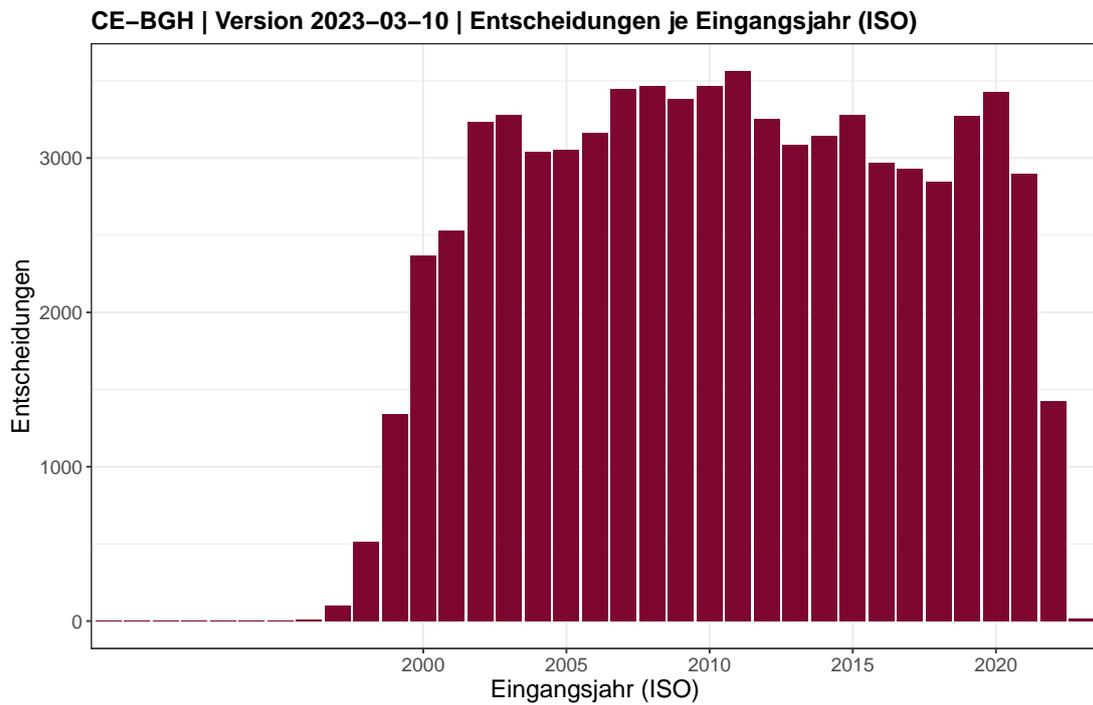
*(continued)*

---

Entscheidungsjahr	Entscheidungen	% Gesamt	% Kumulativ
2017	3120	4.30	77.74
2018	2988	4.12	81.86
2019	2950	4.07	85.93
2020	3265	4.50	90.44
2021	3546	4.89	95.33
2022	3102	4.28	99.60
2023	288	0.40	100.00
Total	72517	100.00	100.00

---

## 11.9 Nach Eingangsjahr (ISO)



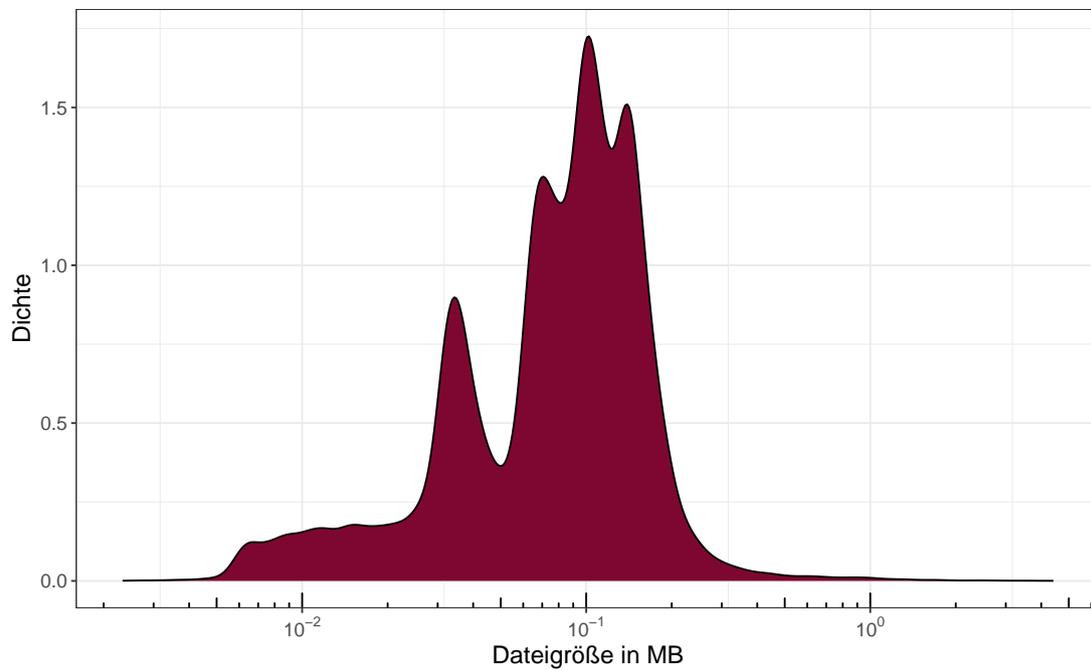
Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
1980	1	0.00	0.00
1986	1	0.00	0.00
1988	1	0.00	0.00
1991	1	0.00	0.01
1993	2	0.00	0.01
1994	2	0.00	0.01
1995	4	0.01	0.02
1996	9	0.01	0.03
1997	100	0.14	0.17
1998	515	0.71	0.88
1999	1343	1.85	2.73
2000	2368	3.27	5.99
2001	2530	3.49	9.48
2002	3234	4.46	13.94
2003	3280	4.52	18.47
2004	3038	4.19	22.66
2005	3055	4.21	26.87

*(continued)*

Eingangsjahr	Entscheidungen	% Gesamt	% Kumulativ
2006	3161	4.36	31.23
2007	3447	4.75	35.98
2008	3465	4.78	40.76
2009	3384	4.67	45.43
2010	3466	4.78	50.20
2011	3564	4.91	55.12
2012	3257	4.49	59.61
2013	3086	4.26	63.87
2014	3144	4.34	68.20
2015	3280	4.52	72.73
2016	2970	4.10	76.82
2017	2929	4.04	80.86
2018	2848	3.93	84.79
2019	3270	4.51	89.30
2020	3427	4.73	94.02
2021	2896	3.99	98.02
2022	1426	1.97	99.98
2023	13	0.02	100.00
Total	72517	100.00	100.00

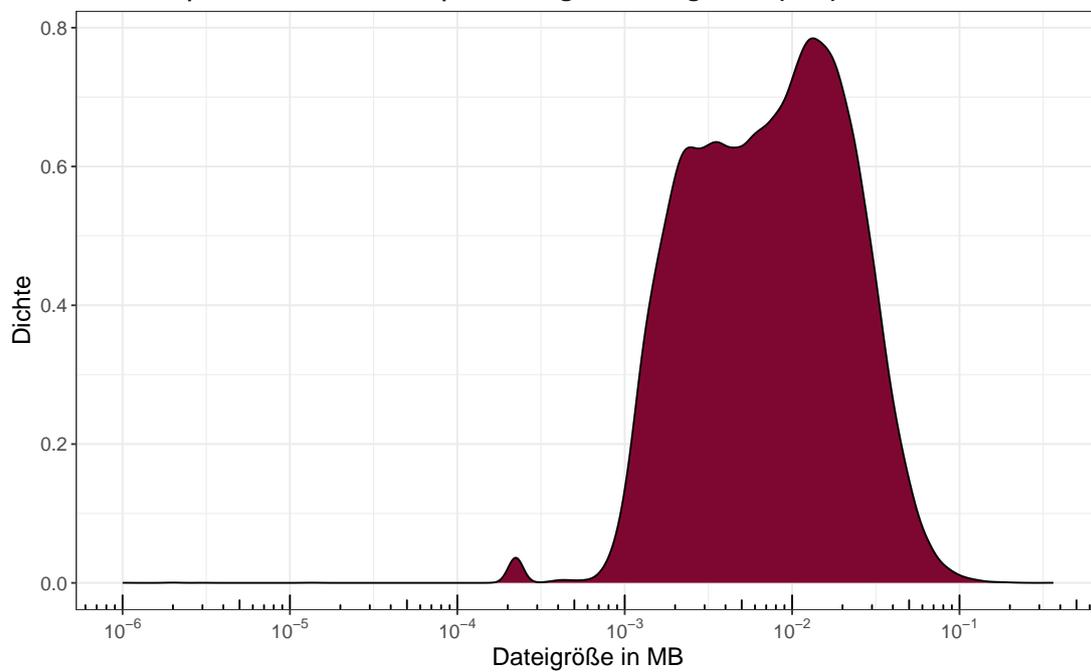
## 12 Dateigrößen

CE-BGH | Version 2023-03-10 | Verteilung der Dateigrößen (PDF)



Fobbe | DOI: 10.5281/zenodo.7699032

CE-BGH | Version 2023-03-10 | Verteilung der Dateigrößen (TXT)



Fobbe | DOI: 10.5281/zenodo.7699032

## 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 2023-03-10

- Vollständige Aktualisierung der Daten
- NEU Zitations-Netzwerk zwischen allen Aktenzeichen des Bundesgerichtshofs als GraphML zur freien Verwendung (EXPERIMENTELL!)
- Gesamte Laufzeitumgebung mit Docker versionskontrolliert
- Aktenzeichen aus dem Eingangszeitraum 2000 bis 2009 nun korrekt mit führender Null formatiert (z.B. 1 BvR 44/02 statt 1 BvR 44/2)
- Vereinfachung der Konfigurations-Datei
- Verbesserte Formatierung von Warnungen und Fehlermeldungen im Compilation Report
- Verbesserung des Robustness Check Reports
- 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
- Korrektur für RiSt-Aktenzeichen eingefügt

### 14.2 Version 2022-08-16

- Vollständige Aktualisierung der Daten
- Neuer Entwurf des gesamten Source Codes im `{targets}` framework
- Die Zivilsenate in der Variable “`spruchkoerper_db`” sind jetzt arabisch nummeriert, damit sie automatisch sortiert werden können. Die originale römische Nummerierung ist weiterhin in der Variable “`spruchkoerper_az`” zu finden.
- Variable “`comment`” wurde in “`bemerkung`” umbenannt
- Variable “`berichtigung`” ist jetzt entweder mit `TRUE` oder `FALSE` codiert
- Diagramme sind in neuer Reihenfolge nummeriert, um die Reihenfolge im Codebook abzubilden
- Umfang der Datenbankabfrage ist nun vollautomatisiert

### 14.3 Version 2022-02-12

- Vollständige Aktualisierung der Daten
- Strenge Kontrolle und semantische Sortierung aller Variablen (entsprechend der Reihenfolge im Codebook)
- Datenstruktur wird nicht mehr im Codebook angezeigt um Fehler mit der UTF8-Kodierung und *listings* für  $\LaTeX$  zu vermeiden
- Strenge Versionskontrolle aller R packages mit *renv*
- Der Prozess der Kompilierung ist jetzt detailliert konfigurierbar, insbesondere die Parallelisierung
- Parallelisierung nun vollständig mit *future* statt mit *foreach* und *doParallel*
- Fehlerhafte Kompilierungen werden beim vor der nächsten Kompilierung vollautomatisch aufgeräumt
- Alle Ergebnisse werden automatisch fertig verpackt in den Ordner ‘output’ sortiert
- README und CHANGELOG sind jetzt externe Markdown-Dateien, die bei der Kompilierung automatisiert eingebunden werden

- Issue #1 fixed: Senate normalisiert; die Variable “spruchkoerper\_db” enthält nun die Präfixe “Strafsenat” und “Zivilsenat” vor der jeweiligen Senatsnummer um in den Dateinamen eine einfachere Orientierung zu ermöglichen
- Issue #2 fixed: Variablen nicht mehr doppelt definiert
- Issue #3 fixed: Alle Dateinamen-Präfixe nun korrekt
- Source Code des Changelogs zu Markdown konvertiert
- In der Vergangenheit fälschlich als “Platzhalter” aussortierte drei Dokumente sind nun im Datensatz enthalten
- Das Diagramm “Entscheidungen je Registerzeichen” ist nun zu einer Log-Skala konvertiert um die Darstellung informativer zu gestalten

#### 14.4 Version 2021-04-27

- 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 Lizenz, Versionsnummer, Concept DOI, Version DOI, ECLI, Typ der Entscheidung, Präsident:in, Vize-Präsident:in, Verfahrensart, Name, Leitsatz, Bemerkungen, Berichtigungen, und linguistische Kennzahlen (Zeichen, Tokens, Typen, Sätze)
- Einführung von PDF-Varianten für Leitsatzentscheidungen und namentlich gekennzeichneten Entscheidungen.
- Zusammenfügung von über Zeilengrenzen getrennten Wörtern in der CSV-Variante
- Automatisierung und Erweiterung der Qualitätskontrolle
- Einführung von Diagrammen zur Visualisierung von Prüfergebnissen
- Einführung kryptographischer Signaturen
- Alle Variablen sind nun in Kleinschreibung und Snake Case gehalten
- Variable ‘Ordinalzahl’ in ‘eingangsnummer’ umbenannt

#### 14.5 Version 2020-07-09

- 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_1.31.0 quanteda_3.2.4
## [4] readtext_0.81 data.table_1.14.8 scales_1.2.1
## [7] ggraph_2.1.0 ggplot2_3.4.1 pdftools_3.3.3
## [10] kableExtra_1.3.4 knitr_1.42 rvest_1.0.3
## [13] httr_1.4.5 mgsub_1.7.3 zip_2.2.2
## [16] fs_1.6.1 testthat_3.1.6 RcppTOML_0.2.2
## [19] tarchetypes_0.7.4 targets_0.14.2
##
## loaded via a namespace (and not attached):
## [1] bit64_4.0.5 webshot_0.5.4 rprojroot_2.0.3
## [4] future.callr_0.8.1 tools_4.2.2 backports_1.4.1
## [7] utf8_1.2.3 R6_2.5.1 colorspace_2.1-0
## [10] withr_2.5.0 tidyselect_1.2.0 gridExtra_2.3
## [13] processx_3.8.0 bit_4.0.5 curl_5.0.0
## [16] compiler_4.2.2 cli_3.6.0 xml2_1.3.3
## [19] desc_1.4.2 labeling_0.4.2 stringfish_0.15.7
## [22] callr_3.7.3 askpass_1.1 systemfonts_1.0.4
## [25] stringr_1.5.0 digest_0.6.31 rmarkdown_2.20
## [28] svglite_2.1.1 pkgconfig_2.0.3 htmltools_0.5.4
## [31] parallelly_1.34.0 fastmap_1.1.1 rlang_1.0.6
## [34] rstudioapi_0.14 farver_2.1.1 generics_0.1.3
## [37] RApiSerialize_0.1.2 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 viridis_0.6.2 lifecycle_1.0.3
## [46] furrr_0.3.1 stringi_1.7.12 yaml_2.3.7
## [49] MASS_7.3-58.1 brio_1.1.3 grid_4.2.2
## [52] parallel_4.2.2 listenv_0.9.0 ggrepel_0.9.3
## [55] lattice_0.20-45 graphlayouts_0.8.4 ps_1.7.2
## [58] pillar_1.8.1 igraph_1.4.1 base64url_1.4
```

```
## [61] pkgload_1.3.2      codetools_0.2-18    stopwords_2.3
## [64] fastmatch_1.1-3    glue_1.6.2          evaluate_0.20
## [67] qpdf_1.3.0         RcppParallel_5.1.7 vctrs_0.5.2
## [70] tweenr_2.0.2       selectr_0.4-2       gtable_0.3.1
## [73] purrr_1.0.1        polyclip_1.10-4     tidyr_1.3.0
## [76] qs_0.25.5          xfun_0.37           ggforce_0.4.1
## [79] tidygraph_1.2.3    viridisLite_0.4.1  tibble_3.1.8
## [82] tinytex_0.44       globals_0.16.2
```

## 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*.
- Hester, Jim, Hadley Wickham, and Gábor Csárdi. 2023. *Fs: Cross-Platform File System Operations Based on Libuv*.
- 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. 2023. *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*.