

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

CODEBOOK

Version 2021-04-27



DOI: [10.5281/zenodo.4705855](https://doi.org/10.5281/zenodo.4705855)

Titel	Corpus der Entscheidungen des Bundesgerichtshofs
Abkürzung	CE-BGH
Autor	Seán Fobbe
Version	2021-04-27
Download	https://doi.org/10.5281/zenodo.4705855
Lizenz	CC0 1.0 Universal

Zitiervorschlag

Seán Fobbe (2021). Corpus der Entscheidungen des Bundesgerichtshofs (CE-BGH). Version 2021-04-27. Zenodo. DOI: 10.5281/zenodo.4705855.

Digital Object Identifier (DOI): Concept DOI und Version DOI

Soweit nicht anders angegeben ist die DOI immer eine »Version DOI« und bezieht sich nur auf eine bestimmte Version des Datensatzes. Sie verweist daher nur auf Version 2021-04-27. 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

1	Einführung	5
2	Nutzung	6
2.1	CSV-Dateien	6
2.2	TXT-Dateien	6
3	Konstruktion	7
3.1	Beschreibung des Datensatzes	7
3.2	Datenquellen	7
3.3	Sammlung der Daten	7
3.4	Source Code und Compilation Report	7
3.5	Grenzen des Datensatzes	8
3.6	Urheberrechtsfreiheit von Rohdaten und Datensatz	8
3.7	Metadaten	8
3.7.1	Allgemein	8
3.7.2	Schema für die Dateinamen	9
3.7.3	Beispiel eines Dateinamens	9
3.8	Qualitätsprüfung	9
3.9	Grafische Darstellung	9
4	Varianten und Zielgruppen	10
5	Variablen	12
5.1	Datenstruktur	12
5.2	Hinweise	13
5.3	Erläuterungen zu den einzelnen Variablen	14
6	Registerzeichen	20
7	Präsident:Innen	24
7.1	Hinweise	24
7.2	Lebensdaten	24
7.3	Dienstalter und Lebensalter	24
8	Vize-Präsident:innen	25
8.1	Hinweise	25
8.2	Lebensdaten	25
8.3	Dienstalter und Lebensalter	26
9	Linguistische Kennzahlen	27
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
10	Inhalt des Korpus	30
10.1	Zusammenfassung	30

10.2	Nach Typ der Entscheidung	30
10.3	Nach Spruchkörper (Aktenzeichen)	31
10.4	Nach Spruchkörper (Datenbank)	33
10.5	Nach Registerzeichen	36
10.6	Nach Präsident:in	39
10.7	Nach Vize-Präsident:in	40
10.8	Nach Entscheidungsjahr	41
10.9	Nach Eingangsjahr (ISO)	43
11	Dateigrößen	45
11.1	Verteilung PDF-Dateigrößen	45
11.2	Verteilung TXT-Dateigrößen	45
11.3	Gesamtgröße je ZIP-Archiv	46
12	Prüfung kryptographischer Signaturen	47
12.1	Allgemeines	47
12.2	Persönliche GPG-Signatur	47
12.3	Import: Public Key	47
12.4	Prüfung: GPG-Signatur der Hash-Datei	48
12.5	Prüfung: SHA3-512 Hashes der ZIP-Archive	49
13	Changelog	50
14	Parameter für strenge Replikationen	51
	Literaturverzeichnis	52

1 Einführung

Das **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).¹ Der Bundesgerichtshof 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 2021 am Bundesgerichtshof 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.² 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 Bundesgerichtshof 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 Bundesgerichtshof 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 Bundesgerichtshof und Rügen der Besetzung eines Oberlandesgerichts (§ 135 GVG).

Die quantitative Analyse von juristischen Texten, insbesondere denen des Bundesgerichtshofs, ist in den deutschen Rechtswissenschaften ein noch junges und kaum bearbeitetes Feld.³ Zu einem nicht unerheblichen Teil liegt dies auch daran, dass die Anzahl an freizugänglichen 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. Freizugängliche Datenbanken wie *Opinio Iuris*⁴ und *openJur*⁵ 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.⁶

¹ 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.

² Geschäftsverteilungsplan des Bundesgerichtshofs für das Jahr 2021.

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

⁵ <https://openjur.de/>

⁶ 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. CSV⁷ ist ein einfaches und maschinell gut lesbares Tabellen-Format. In diesem Datensatz sind die Werte komma-separiert. Jede Spalte entspricht einer Variable, jede Zeile einer Entscheidung. Die Variablen sind unter Punkt 5 genauer erläutert.

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

```
library(data.table)
dt.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")
```

⁷ 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	https://www.bundesgerichtshof.de
Source Code	https://doi.org/10.5281/zenodo.4705865
Personendaten	https://doi.org/10.5281/zenodo.4568682
Registerzeichen	https://doi.org/10.5281/zenodo.4569564

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

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

3.3 Sammlung der Daten

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

3.4 Source Code und Compilation Report

Der gesamte Source Code — sowohl für die Erstellung des Datensatzes, als auch für dieses Codebook — ist öffentlich einsehbar und dauerhaft erreichbar im wissenschaftlichen Archiv des CERN unter dieser Adresse hinterlegt: <https://doi.org/10.5281/zenodo.4705865>

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

debook). Der Compilation Report enthält den vollständigen und kommentierten Source Code, dokumentiert relevante Rechenergebnisse, gibt sekundengenaue Zeitstempel an und ist mit einem klickbaren Inhaltsverzeichnis versehen. Er ist zusammen mit dem Source Code hinterlegt. Wenn Sie sich für Details der Herstellung interessieren, lesen Sie diesen bitte zuerst.

3.5 Grenzen des Datensatzes

Nutzer:innen sollten folgende wichtige Grenzen beachten:

1. Der Datensatz enthält nur das, was das Gericht auch tatsächlich veröffentlicht, nämlich begründete Entscheidungen (*publication bias*). Platzhalter für Nichtannahme- und Verwerfungsbeschlüsse ohne Begründung wurden automatisiert entfernt, sofern vorhanden.
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 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_I_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 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	Traditionelle juristische Forschung. Die PDF-Dokumente wie sie vom Bundesgerichtshof 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	Legal Tech/Quantitative Forschung. Diese CSV-Datei ist die für statistische Analysen empfohlene Variante des Datensatzes. Sie enthält den Volltext aller Entscheidungen, sowie alle in diesem Codebook beschriebenen Metadaten. Über Zeilenumbrüche getrennte Wörter wurden zusammengefügt.
CSV_Metadaten	Legal Tech/Quantitative Forschung. Wie die andere CSV-Datei, nur ohne die Entscheidungstexte. Sinnvoll für Analyst:innen, die sich nur für die Metadaten interessieren und Speicherplatz sparen wollen.
TXT	Subsidiär für alle Zielgruppen. Diese Variante enthält die vollständigen aus den PDF-Dateien extrahierten Entscheidungstexte, aber nur einen reduzierten Umfang an Metadaten, der dem der PDF-Dateien entspricht. Die TXT-Dateien sind optisch an das Layout der PDF-Dateien angelehnt. Geeignet für qualitativ arbeitende Forscher:innen, die nur wenig Speicherplatz oder eine langsame Internetverbindung zur Verfügung haben oder für quantitativ arbeitende Forscher:innen, die beim Einlesen der CSV-Dateien Probleme haben. Über Zeilenumbrüche getrennte Wörter wurden <i>nicht</i> zusammengefügt.

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

5 Variablen

5.1 Datenstruktur

```
## Classes 'data.table' and 'data.frame': 66344 obs. of 31 variables:
## $ doc_id : chr "BGH_3_NA_2000-01-04_3_StR_561_99_NA_NA_0.txt" "BGH_
_3_NA_2000-01-04_3_StR_567_99_NA_NA_0.txt" "BGH_3_NA_2000-01-05_3_StR_273_99_
NA_NA_0.txt" "BGH_3_NA_2000-01-05_3_StR_465_99_NA_NA_0.txt" ...
## $ text : chr " BUNDESGERICHTSHOF\n
BESCHLUSS\n3 StR 561/99\n "| __truncated__
" BUNDESGERICHTSHOF\n BESCHLUSS\n3
StR 567/99\n "| __truncated__ "
BUNDESGERICHTSHOF\n BESCHLUSS\n3 StR 273/99\n
"| __truncated__ " BUNDESGERICHTSHOF\n
BESCHLUSS\n3 StR 465/99\n "|
__truncated__ ...
## $ gericht : chr "BGH" "BGH" "BGH" "BGH" ...
## $ spruchkoerper_db : chr "3" "3" "3" "3" ...
## $ leitsatz : chr "NA" "NA" "NA" "NA" ...
## $ datum : IDate, format: "2000-01-04" "2000-01-04" ...
## $ spruchkoerper_az : chr "3" "3" "3" "3" ...
## $ registerzeichen : chr "StR" "StR" "StR" "StR" ...
## $ eingangsnummer : int 561 567 273 465 473 560 651 651 638 640 ...
## $ eingangsjahr_az : int 99 99 99 99 99 99 99 99 99 99 ...
## $ zusatz_az : chr NA NA NA NA ...
## $ name : chr NA NA NA NA ...
## $ kollision : int 0 0 0 0 0 0 1 0 0 ...
## $ entscheidungsjahr: int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000
...
## $ eingangsjahr_iso : int 1999 1999 1999 1999 1999 1999 1999 1999 1999 1999
...
## $ praesi : chr "Geiss" "Geiss" "Geiss" "Geiss" ...
## $ v_praesi : chr "Jähnke" "Jähnke" "Jähnke" "Jähnke" ...
## $ verfahrensart : chr "Revisionen und Vorlegungssachen nach § 121 Abs.1
Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (Strafsachen)"
"Revisionen und Vorlegungssachen nach § 121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs.
3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (Strafsachen)" "Revisionen und
Vorlegungssachen nach § 121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13
Abs. 4, 25 StrRehaG (Strafsachen)" "Revisionen und Vorlegungssachen nach §
121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (
Strafsachen)" ...
## $ aktenzeichen : chr "3 StR 561/99" "3 StR 567/99" "3 StR 273/99" "3 StR
465/99" ...
## $ entscheidung_typ : chr "B" "B" "B" "B" ...
## $ ecli : chr "ECLI:DE:BGH:2000:040100B3STR561.99.0" "ECLI:DE:BGH
:2000:040100B3STR567.99.0" "ECLI:DE:BGH:2000:050100B3STR273.99.0" "ECLI:DE:
BGH:2000:050100B3STR465.99.0" ...
## $ bemerkung : chr "" "" "" "" ...
## $ berichtigung : chr NA NA NA NA ...
## $ 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.4705855" "10.5281/zenodo.4705855"
"10.5281/zenodo.4705855" "10.5281/zenodo.4705855" ...
## $ version : IDate, format: "2021-04-27" "2021-04-27" ...
```

```

## $ 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" ...
## $ zeichen        : int  4469 2406 2285 757 3333 8560 1841 1841 6733 3867
...
## $ tokens         : int  651 320 311 75 461 1339 247 247 953 552 ...
## $ typen          : int  313 203 165 52 244 541 134 134 462 288 ...
## $ saetze         : int  27 15 27 8 19 66 16 16 38 27 ...
## - attr(*, ".internal.selfref")=<externalptr>

```

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 Erläuterungen zu den einzelnen Variablen

Variable	Typ	Erläuterung
doc_id	String	(Nur CSV-Datei) Der Name der extrahierten TXT-Datei.
text	String	(Nur CSV-Datei) Der vollständige Inhalt der Entscheidung, so wie er in der heruntergeladenen PDF-Datei dokumentiert ist. Die einzige editorische Änderung gegenüber den TXT-Dateien ist die Zusammenfügung von über Zeilengrenzen gebrochenen Wörtern.
gericht	Alphabetisch	In diesem Datensatz ist nur der Wert »BGH« vergeben. Dies ist der ECLI-Gerichtscode für »Bundesgerichtshof«. Diese Variable dient vor allem zur einfachen und transparenten Verbindung der Daten mit anderen Datensätzen.
datum	Datum (ISO)	Das Datum der Entscheidung im Format YYYY-MM-DD (Langform nach ISO-8601). Die Langform ist für Menschen einfacher lesbar und wird maschinell auch öfter automatisch als Datumsformat erkannt.
entscheidung_typ	Alphabetisch	(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	Alphabetisch	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. Arabische Ziffern stehen für die Strafsenate, römische Ziffern für die Zivilsenate. Die übrigen Senate haben eindeutige Namen.
spruchkoerper_az	Alphanumerisch	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«.

Variable	Typ	Erläuterung
registerzeichen	String	Das amtliche Registerzeichen. Es gibt die Verfahrensart an, in der die Entscheidung ergangen ist. Eine Erläuterung der Registerzeichen und der zugehörigen Verfahrensarten findet sich unter Punkt 6.
verfahrensart	String	Die ausführliche Beschreibung der Verfahrensart, die dem Registerzeichen zugeordnet ist. Eine Erläuterung der Registerzeichen und der zugehörigen Verfahrensarten findet sich unter Punkt 6.
eingangsnummer	Natürliche Zahl	Verfahren des gleichen Eingangsjahres erhalten vom Gericht eine Nummer in der Reihenfolge ihres Eingangs.
eingangsjahr_az	Natürliche Zahl	Das im Aktenzeichen angegebene Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine zweistellige Jahreszahl (YY).
eingangsjahr_iso	Natürliche Zahl	(Nur CSV-Datei) Das nach ISO-8601 codierte Jahr in dem das Verfahren beim Gericht anhängig wurde. Das Format ist eine vierstellige Jahreszahl (YYYY), um eine maschinenlesbare und eindeutige Jahreszahl für den Eingang zur Verfügung zu stellen. Wurde aus der Variable »ingangsjahr_az« durch den Autor des Datensatzes berechnet, unter der Annahme, dass Jahreszahlen über 50 dem 20. Jahrhundert zugeordnet sind und andere Jahreszahlen dem 21. Jahrhundert.
entscheidungsjahr	Natürliche Zahl	(Nur CSV-Datei) Das Jahr in dem die Entscheidung ergangen ist. Das Format ist eine vierstellige Jahreszahl (YYYY). Wurde aus der Variable »datum« durch den Autor des Datensatzes berechnet.
zusatz_az	Alphabetisch	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.

Variable	Typ	Erläuterung
kollision	Numerisch	In wenigen Fällen sind am gleichen Tag mehrere Entscheidungen zum gleichen Aktenzeichen ergangen. Diese werden ab der zweiten Entscheidung pro Tag durch eine Kollisions-ID mit einer fortlaufenden Zahl ausdifferenziert. Für die erste Entscheidung ist der Wert der Variable »0«. Die zweite Entscheidung ist mit »1« identifiziert, die dritte mit »2« und so fort. ACHTUNG: diese Variable ist von dem Autor des Datensatzes berechnet und keine amtliche Angabe. Die Nutzung für die Erstellung der ECLI ist daher problematisch.
praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
v_praesi	Alphabetisch	(Nur CSV-Datei) Der Nachname des oder der Vize-Präsident:in in dessen/deren Amtszeit das Datum der Entscheidung fällt.
aktenzeichen	String	(Nur CSV-Datei) Das amtliche Aktenzeichen. Die Variable wurde aus den Variablen »spruchkoerper_az«, »registerzeichen«, »eingangsnummer« und »eingangsjahr_az« durch den Autor des Datensatzes berechnet.
bemerkung	String	(Nur CSV-Datei) 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.
berichtigung	Alphabetisch	(Nur CSV-Datei) Eine Angabe ob es sich um einen Berichtigungsbeschluss handelt (dann Wert »Berichtigung«), sonst »NA«. Wurde durch den Autor des Datensatzes aus der Variable »bemerkung« extrahiert.

Variable	Typ	Erläuterung
ecli	String	(Nur CSV-Datei) Experimentell! 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. ACHTUNG: 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.
zeichen	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Zeichen eines Dokumentes.
tokens	Natürliche Zahl	(Nur CSV-Datei) Die Anzahl Tokens (beliebige Zeichenfolge getrennt durch whitespace) eines Dokumentes. Diese Zahl kann je nach Tokenizer und verwendeten Einstellungen erheblich schwanken. Für diese Berechnung wurde eine reine Tokenisierung ohne Entfernung von Inhalten durchgeführt. Benutzen Sie diesen Wert eher als Anhaltspunkt für die Größenordnung denn als exakte Aussage und führen sie ggf. mit ihrer eigenen Software eine Kontroll-Rechnung durch.

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

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

Codierung	Registerzeichen	Verfahrensart
AK	AK	Aktenkontrolle für Haftprüfungsverfahren
AR	AR	Allgemeines Register
ARAnw	AR(Anw)	Allgemeines Register (Anwaltssachen)
ARNot	AR(Not)	Allgemeines Register (Notarsachen)
ARRi	AR(Ri)	Allgemeines Register (Dienstgericht des Bundes)
ARVS	AR(VS)	Entscheidungen über Justizverwaltungsakte (Strafsachen)
ARVZ	AR(VZ)	Entscheidungen über Justizverwaltungsakte (Zivilsachen)
ARsVollz	AR(Vollz)	Vorlegungssachen (Strafvollzugssachen)
ARZ	ARZ	Allgemeines Register und Gerichtsstandsbestimmungen (Zivilsachen)
ARs	ARs	Allgemeines Register und Gerichtsstandsbestimmungen (Strafsachen)
AnwSt	AnwSt	Anträge auf Einleitung des anwaltsgerichtlichen Verfahrens und Beschwerden in Ordnungsmittelverfahren gegen Rechtsanwälte beim Bundesgerichtshof
AnwStB	AnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwStR	AnwSt(R)	Revisionen gegen Urteile eines Anwaltsgerichtshofes
AnwZ	AnwZ	Klagen über Entscheidungen in Zulassungssachen oder gegen sonstige Verwaltungsakte betreffend Rechtsanwälte beim Bundesgerichtshof
AnwZB	AnwZ(B)	Beschwerden gegen Entscheidungen eines Anwaltsgerichtshofes
AnwZBrfg	AnwZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen eines Anwaltsgerichtshofes

(continued)

Codierung	Register- zeichen	Verfahrensart
AnwZP	AnwZ(P)	Klagen über die Anfechtung von Wahlen und Beschlüssen der Rechtsanwaltskammer beim BGH und der Bundesrechtsanwaltskammer
BGs	BGs	Einzelne richterliche Anordnungen des Ermittlungsrichters (Strafsachen)
BLw	BLw	Rechtsbeschwerden in Landwirtschaftssachen der freiwilligen Gerichtsbarkeit
EnRB	EnRB	Rechtsbeschwerden in Bußgeldverfahren nach dem EnWG
EnVR	EnVR	Rechtsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnVZ	EnVZ	Nichtzulassungsbeschwerden in energiewirtschaftsrechtlichen Verwaltungssachen nach dem EnWG
EnZA	EnZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Rechtsstreits in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZB	EnZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
EnZR	EnZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten nach dem EnWG
GSSt	GSSt	Großer Senat (Strafsachen)
GSZ	GSZ	Großer Senat (Zivilsachen)
GmsOGB	GmSOGB	Gemeinsamer Senat der Obersten Gerichtshöfe des Bundes
KRB	KRB	Rechtsbeschwerden in Kartellbußgeldverfahren
KVR	KVR	Rechtsbeschwerden (Kartellverwaltungssachen)
KVZ	KVZ	Nichtzulassungsbeschwerden (Kartellverwaltungssachen)
KZA	KZA	Anträge außerhalb eines in der Revisionsinstanz anhängigen Rechtsstreits (Kartellsachen)
KZB	KZB	Rechtsbeschwerden und Beschwerden in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
KZR	KZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision in bürgerlichen Rechtsstreitigkeiten (Kartellsachen)
LwZA	LwZA	Anträge außerhalb eines in der Revisionsinstanz für Landwirtschaftssachen anhängigen Verfahrens

(continued)

Codierung	Registerzeichen	Verfahrensart
LwZB	LwZB	Beschwerden und Rechtsbeschwerden in Landwirtschaftssachen der streitigen bürgerlichen Gerichtsbarkeit
LwZR	LwZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision und Anträge auf Zulassung der Sprungrevision (Landwirtschaftssachen)
NotStB	NotSt(B)	Beschwerden gegen Beschlüsse der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotStBrfg	NotSt(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Urteile der Oberlandesgerichte in Disziplinarsachen gegen Notare
NotZ	NotZ	Verwaltungsstreitverfahren in Notarsachen und Beschwerden gegen Entscheidungen der Oberlandesgerichte
NotZBrfg	NotZ(Brfg)	Berufungen und Anträge auf Zulassung der Berufung gegen Entscheidungen der Oberlandesgerichte in Notarsachen
PatAnwStB	PatAnwSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden nach der PatAO
PatAnwStR	PatAnwSt(R)	Revisionen nach der PatAO
PatAnwZ	PatAnwZ	Verwaltungsstreitverfahren in Patentanwaltssachen; Berufungen und Anträge auf Zulassung der Berufung und Beschwerden gegen Entscheidungen der Oberlandesgerichte
RiSt	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
RiStB	RiSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden der Richter, Staatsanwälte und Mitglieder des Bundesrechnungshofes gegen Disziplinarverfügungen
RiStR	RiSt(R)	Revisionen in Disziplinarsachen nach dem Deutschen Richtergesetz
RiZ	RiZ	Anträge betreffend Richter im Bundesdienst und Mitglieder des Bundesrechnungshofes auf gerichtliche Entscheidung im Versetzungs- und Prüfungsverfahren sowie auf vorläufige Untersagung der Amtsgeschäfte

(continued)

Codierung	Register- zeichen	Verfahrensart
RizB	RiZ(B)	Beschwerden gegen die Nichtzulassung der Revision nach dem Deutschen Richtergesetz
RiZR	RiZ(R)	Revisionen in Versetzungs- und Prüfungsverfahren nach dem Deutschen Richtergesetz
StB	StB	Beschwerden (Strafsachen)
StR	StR	Revisionen und Vorlegungssachen nach § 121 Abs.1 Nr.1, Abs. 2 GVG, § 79 Abs. 3 OWiG, §§ 13 Abs. 4, 25 StrRehaG (Strafsachen)
StbStB	StbSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigtensachen)
StbStR	StbSt(R)	Revisionen in berufsgerichtlichen Verfahren (Steuerberater- und Steuerbevollmächtigtensachen)
VGS	VGS	Vereinigte Große Senate
VRG	VRG	Ersuchen des Bundesverfassungsgerichts zu Vorlagen nach Art. 100 GG und zu Verfassungsbeschwerden
WpStB	WpSt(B)	Beschwerden gegen die Nichtzulassung der Revision und Beschwerden in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
WpStR	WpSt(R)	Revisionen in berufsgerichtlichen Verfahren (Wirtschaftsprüfersachen)
ZA	ZA	Anträge außerhalb eines in der Rechtsmittelinstanz anhängigen Verfahrens (Zivilsachen)
ZB	ZB	Beschwerden, Rechtsbeschwerden, weitere Beschwerden, Beschwerden gegen die Nichtzulassung der Revision nach dem BEG (Zivilsachen)
ZR	ZR	Revisionen, Beschwerden gegen die Nichtzulassung der Revision, Anträge auf Zulassung der Sprungrevision, Berufungen in Patentsachen (Zivilsachen)
ZRÜ	ZR(Ü)	Erstinstanzliche Klagen auf Entschädigung wegen überlanger Gerichtsverfahren und strafrechtlicher Ermittlungsverfahren (Zivilsachen)

7 Präsident:Innen

7.1 Hinweise

- Die Personaldaten 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 Personaldaten 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.

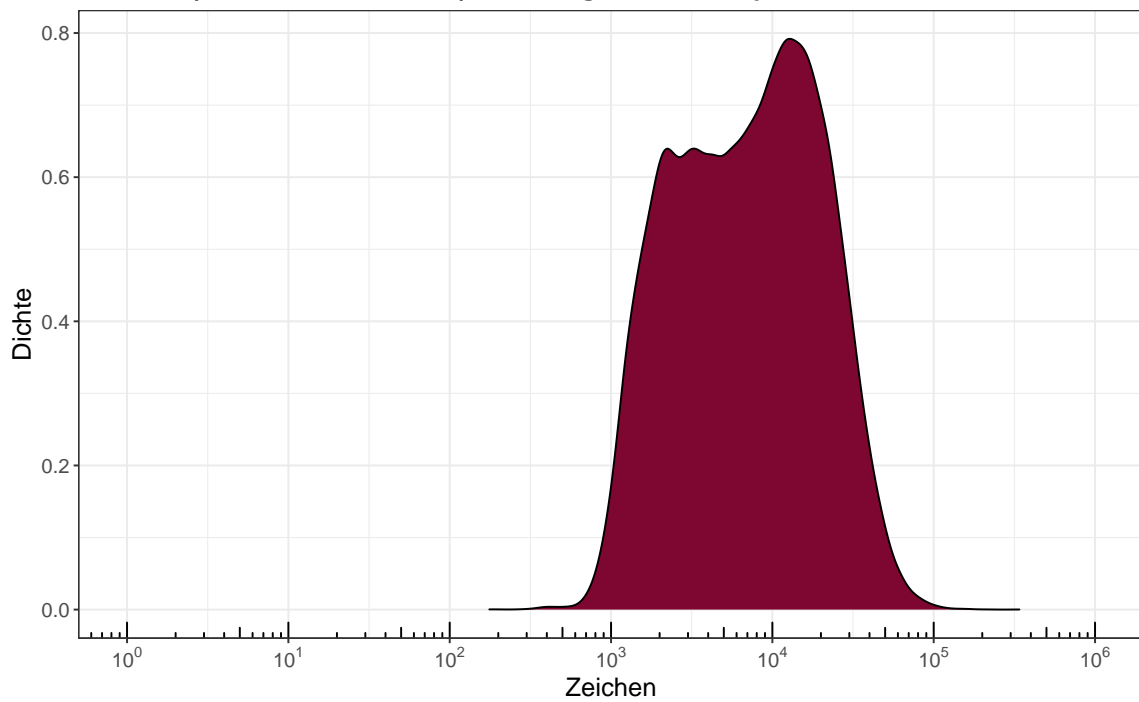
Vorsicht bei der Interpretation: Die x-Achse ist logarithmisch skaliert, d.h. in 10er-Potenzen und damit nicht-linear. Die kleinen Achsen-Markierungen zwischen den Schritten der Exponenten sind eine visuelle Hilfestellung um diese nicht-Linearität zu verstehen.

9.2 Werte der Kennzahlen

Variable	Summe	Min	Quart1	Median	Mittel	Quart3	Max
zeichen	742,090,149	176	2,962.75	7,283	11,185.49	15,664.75	339,765
tokens	113,513,695	30	427.00	1,096	1,710.99	2,395.25	53,108
saetze	6,033,607	2	28.00	63	90.94	126.00	2,683
typen	NA	25	214.00	437	540.69	769.00	5,594

9.3 Verteilung Zeichen

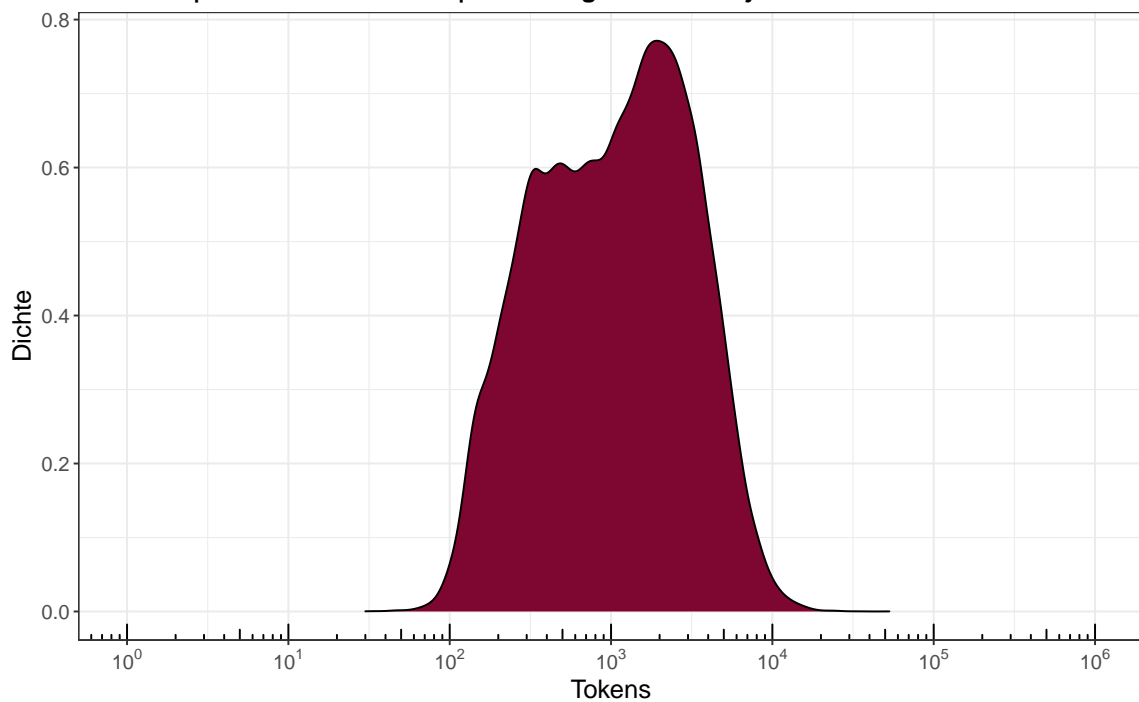
CE-BGH | Version 2021-04-27 | Verteilung der Zeichen je Dokument



DOI: 10.5281/zenodo.4705855

9.4 Verteilung Tokens

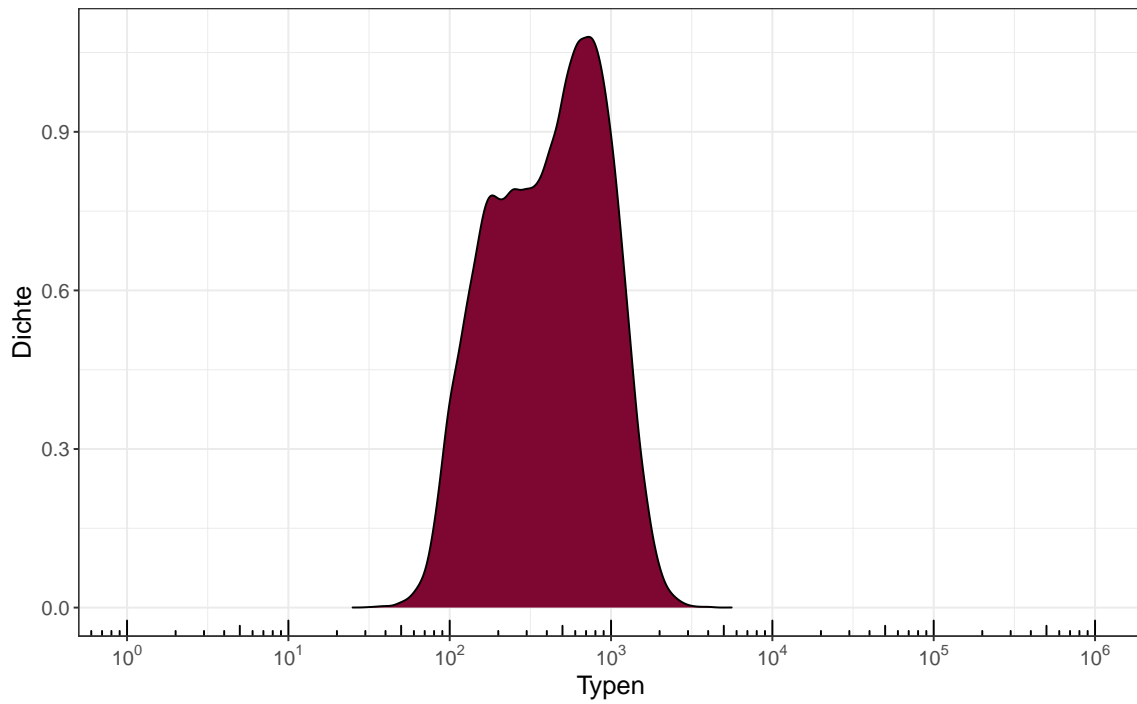
CE-BGH | Version 2021-04-27 | Verteilung der Tokens je Dokument



DOI: 10.5281/zenodo.4705855

9.5 Verteilung Typen

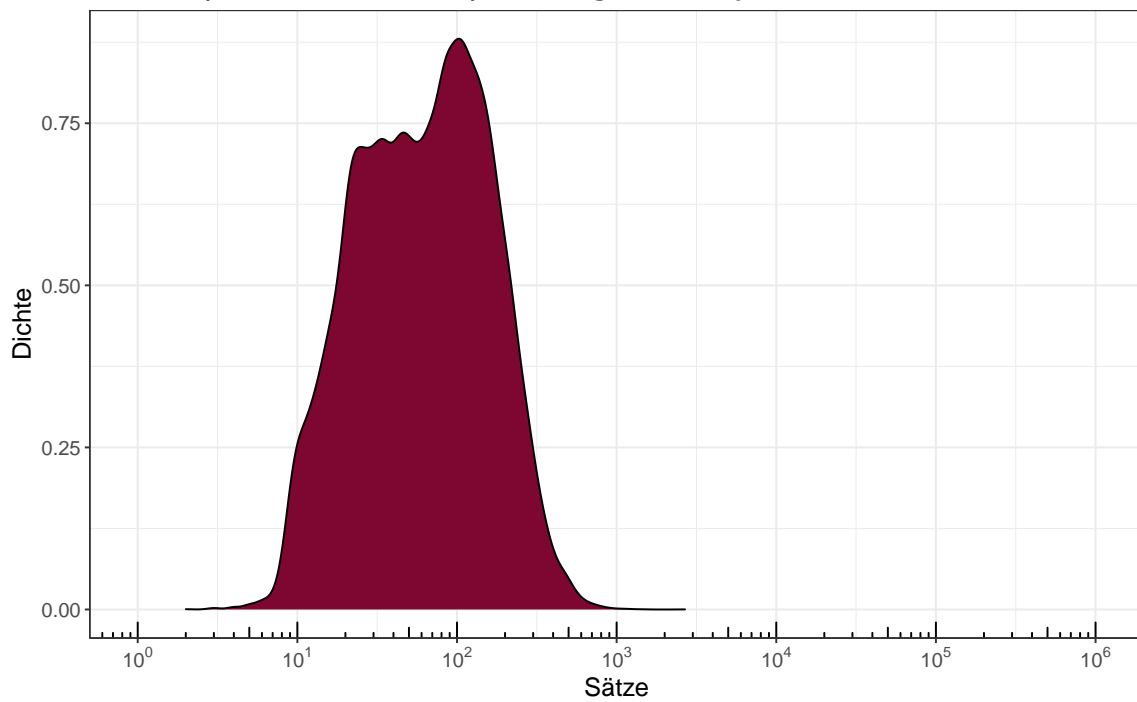
CE-BGH | Version 2021-04-27 | Verteilung der Typen je Dokument



DOI: 10.5281/zenodo.4705855

9.6 Verteilung Sätze

CE-BGH | Version 2021-04-27 | Verteilung der Sätze je Dokument



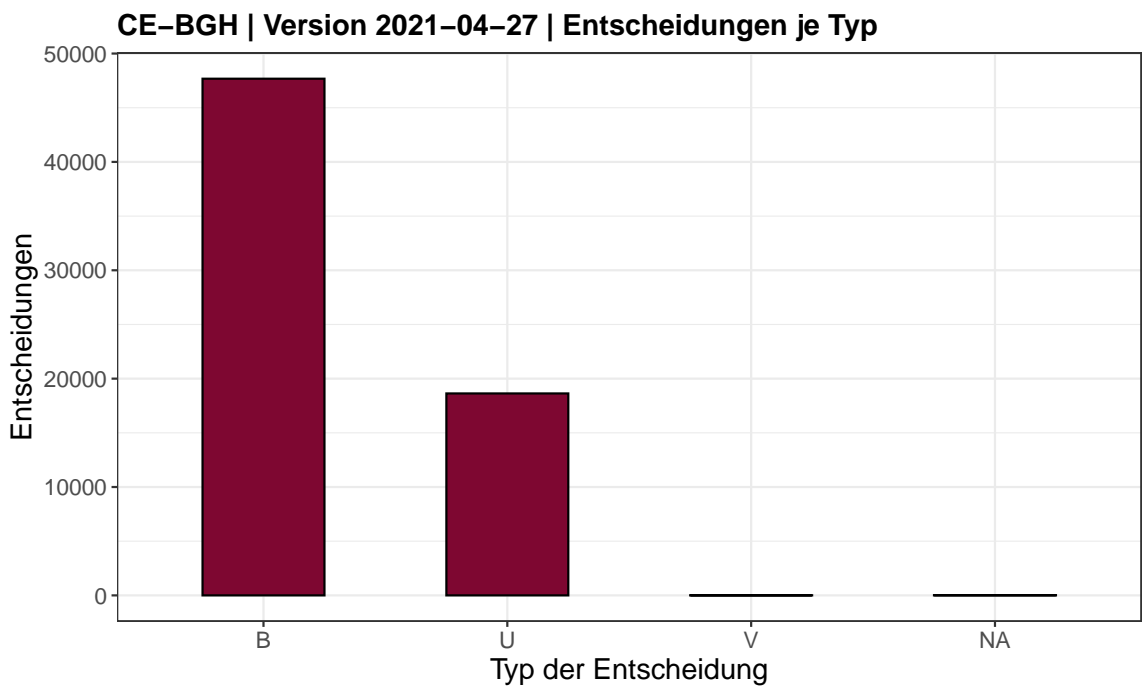
DOI: 10.5281/zenodo.4705855

10 Inhalt des Korpus

10.1 Zusammenfassung

Variable	Anzahl	Min	Quart1	Median	Mittel	Quart3	Max
entscheidungsjahr	22	2000	2006	2010	2010.43	2015	2021
eingangsjahr_iso	33	1980	2005	2010	2009.56	2015	2021
eingangsnummer	752	1	56	156	198.27	302	1304

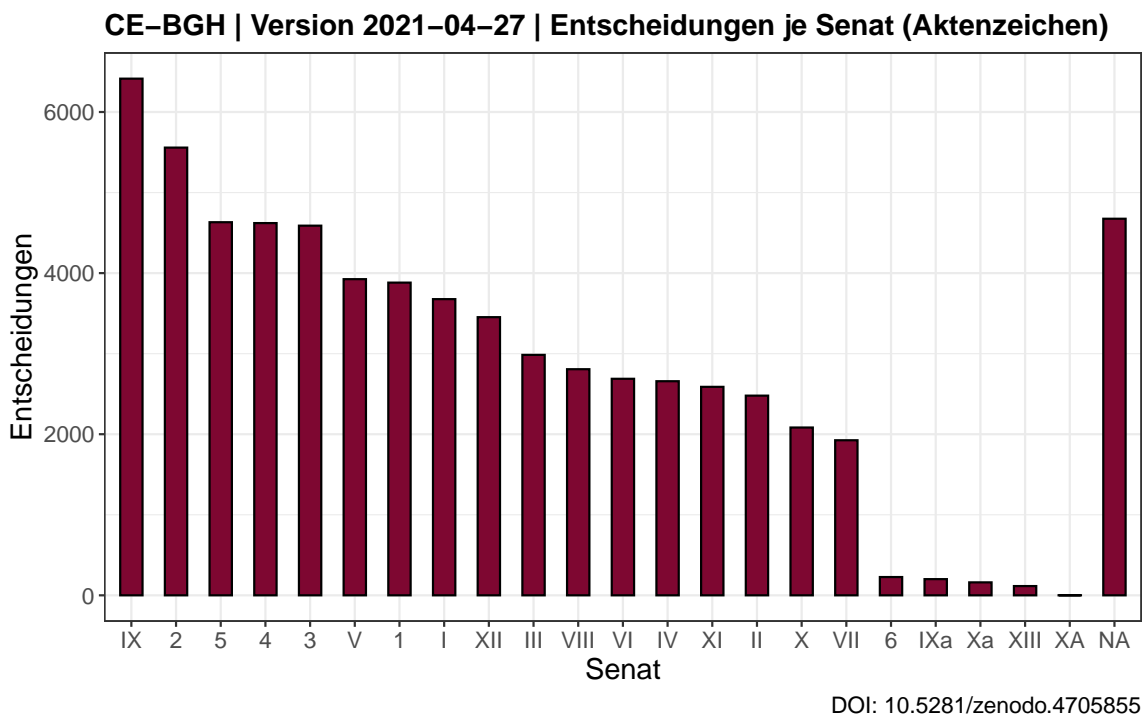
10.2 Nach Typ der Entscheidung



DOI: 10.5281/zenodo.4705855

Typ	Entscheidungen	% Gesamt	% Kumulativ
NA	22	0.03	0.03
B	47677	71.86	71.90
U	18634	28.09	99.98
V	11	0.02	100.00
Total	66344	100.00	100.00

10.3 Nach Spruchkörper (Aktenzeichen)



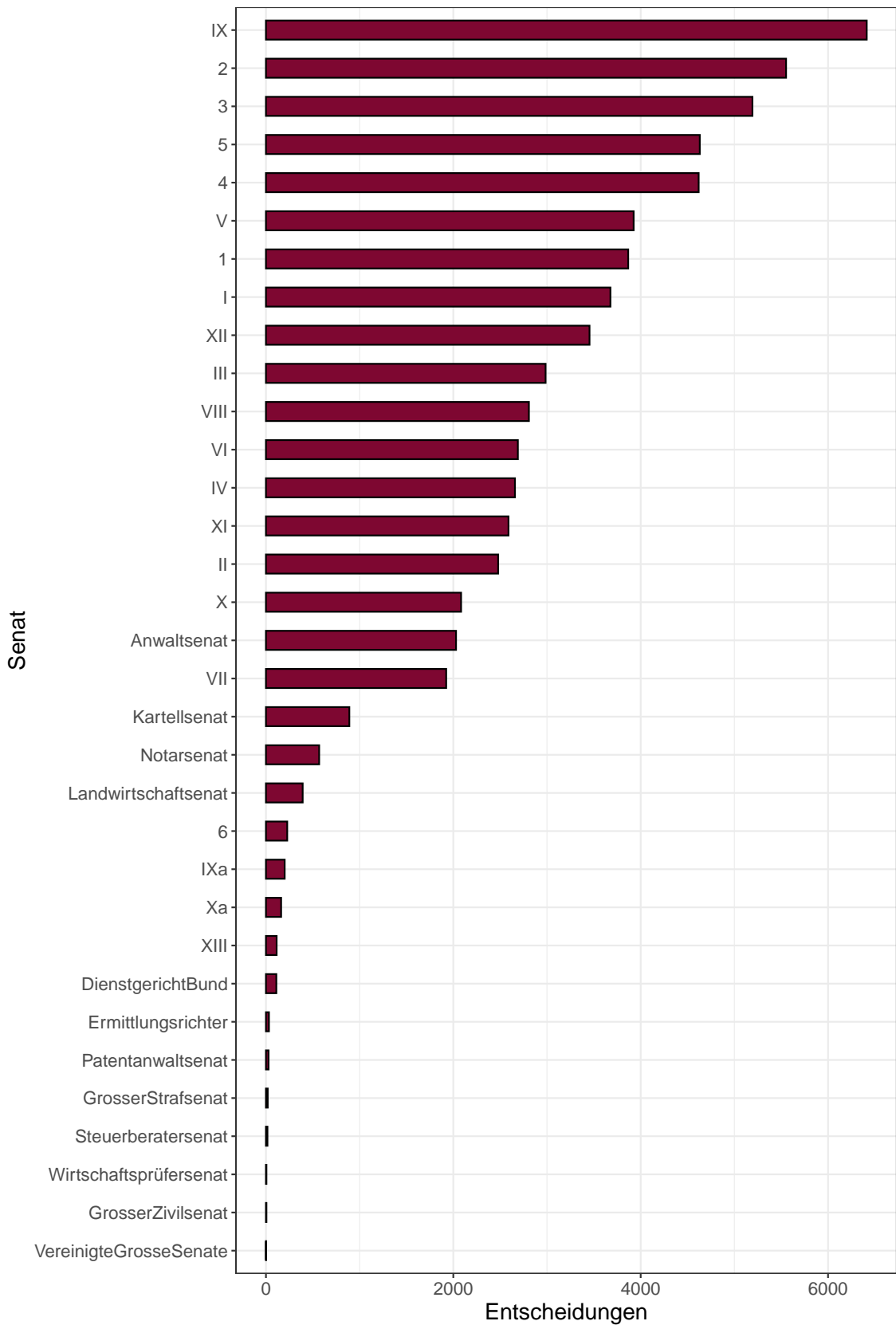
Senat	Entscheidungen	% Gesamt	% Kumulativ
NA	4675	7.05	7.05
1	3882	5.85	12.90
2	5558	8.38	21.28
3	4589	6.92	28.19
4	4621	6.97	35.16
5	4632	6.98	42.14
6	227	0.34	42.48
I	3677	5.54	48.02
II	2479	3.74	51.76
III	2985	4.50	56.26
IV	2658	4.01	60.27
IX	6414	9.67	69.93
IXa	201	0.30	70.24
V	3925	5.92	76.15
VI	2688	4.05	80.20

(continued)

Senat	Entscheidungen	% Gesamt	% Kumulativ
VII	1925	2.90	83.11
VIII	2807	4.23	87.34
X	2083	3.14	90.48
XA	1	0.00	90.48
XI	2588	3.90	94.38
XII	3453	5.20	99.58
XIII	115	0.17	99.76
Xa	161	0.24	100.00
Total	66344	100.00	100.00

10.4 Nach Spruchkörper (Datenbank)

CE-BGH | Version 2021-04-27 | Entscheidungen je Senat (DB)



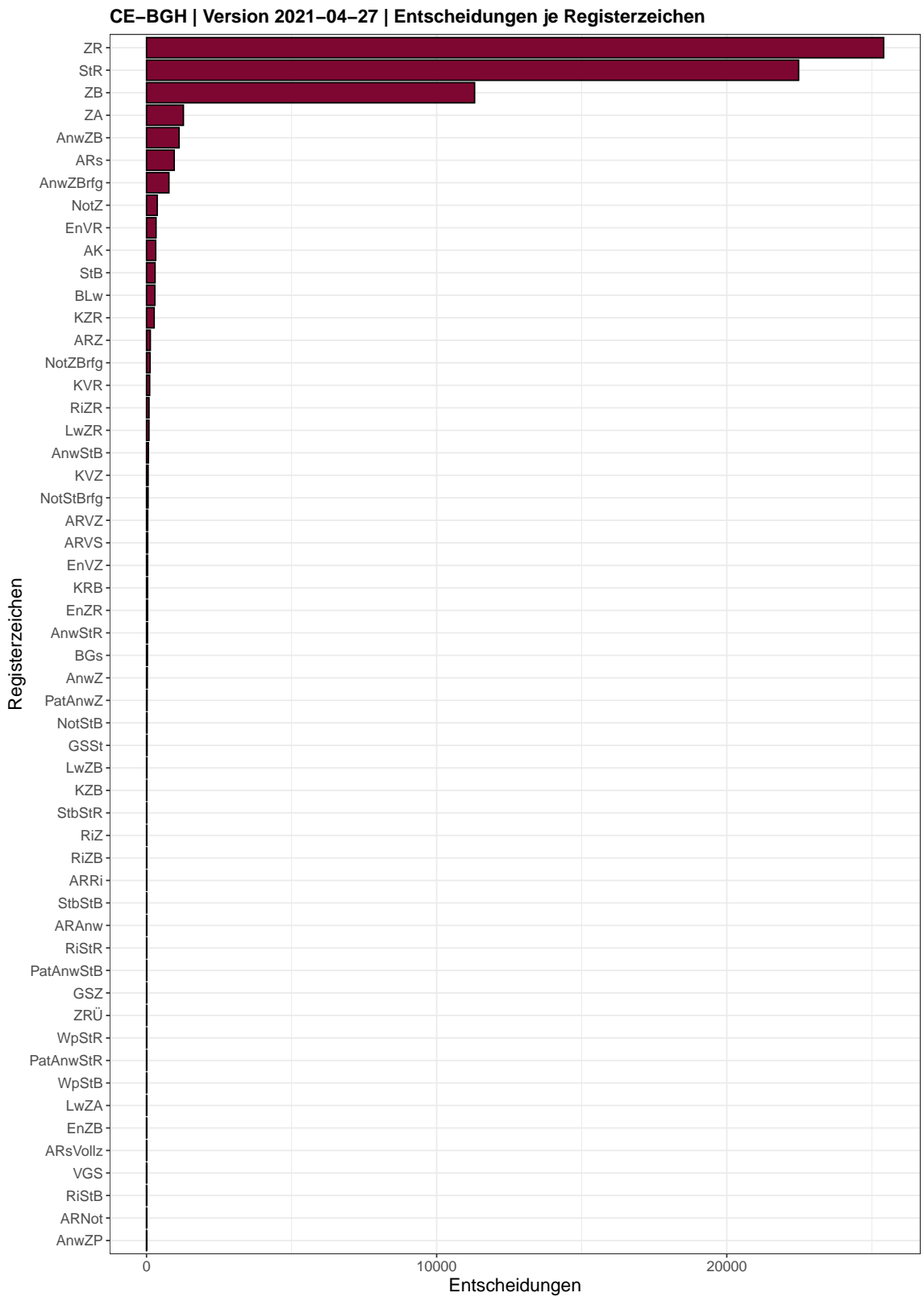
DOI: 10.5281/zenodo.4705855

Senat	Entscheidungen	% Gesamt	% Kumulativ
1	3867	5.83	5.83
2	5552	8.37	14.20
3	5193	7.83	22.02
4	4619	6.96	28.99
5	4631	6.98	35.97
6	227	0.34	36.31
Anwaltsenat	2029	3.06	39.37
DienstgerichtBund	111	0.17	39.53
Ermittlungsrichter	32	0.05	39.58
GrosserStrafsenat	19	0.03	39.61
GrosserZivilsenat	5	0.01	39.62
I	3677	5.54	45.16
II	2479	3.74	48.90
III	2985	4.50	53.40
IV	2658	4.01	57.40
IX	6413	9.67	67.07
IXa	200	0.30	67.37
Kartellsenat	890	1.34	68.71
Landwirtschaftsenat	392	0.59	69.30
Notarsenat	568	0.86	70.16
Patentanwaltsenat	28	0.04	70.20
Steuerberatersenat	16	0.02	70.23
V	3925	5.92	76.14
VI	2689	4.05	80.20
VII	1924	2.90	83.10
VIII	2807	4.23	87.33
VereinigteGrosseSenate	1	0.00	87.33
Wirtschaftsprüfersenat	5	0.01	87.34
X	2083	3.14	90.48
XI	2589	3.90	94.38
XII	3454	5.21	99.58
XIII	114	0.17	99.76

(continued)

Senat	Entscheidungen	% Gesamt	% Kumulativ
Xa	162	0.24	100.00
Total	66344	100.00	100.00

10.5 Nach Registerzeichen

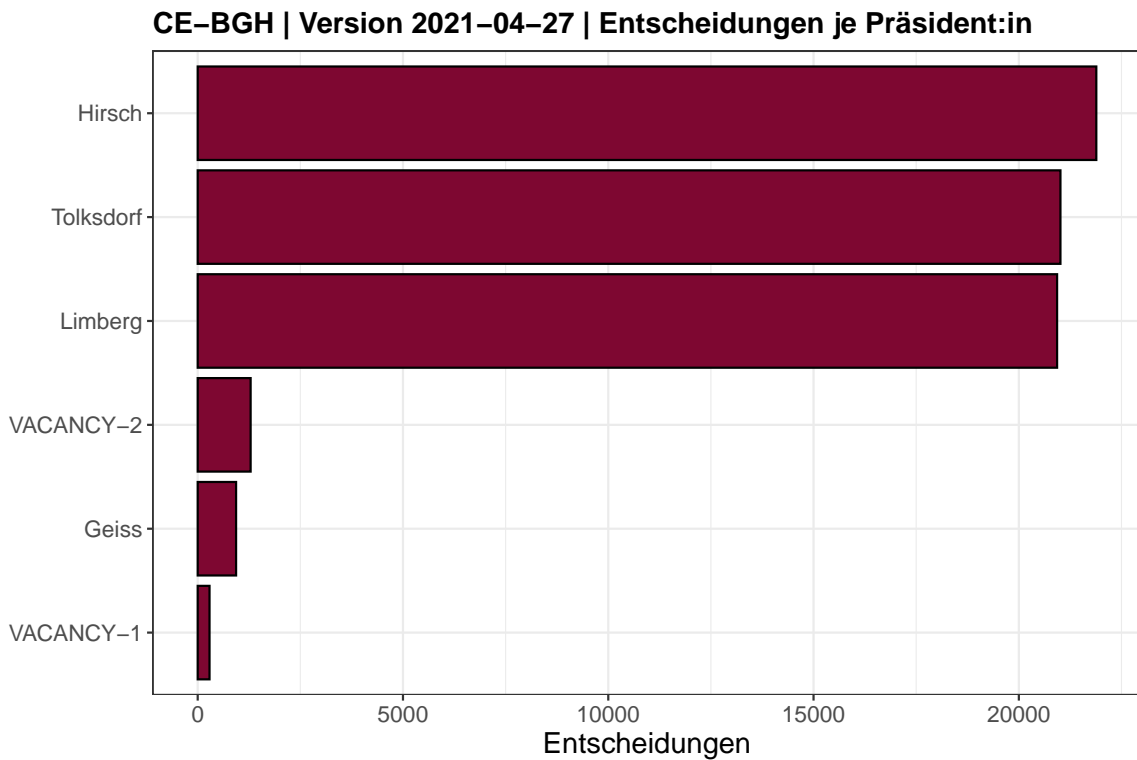


Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
AK	317	0.48	0.48
ARAnw	6	0.01	0.49
ARNot	1	0.00	0.49
ARRi	7	0.01	0.50
ARVS	42	0.06	0.56
ARVZ	42	0.06	0.63
ARZ	131	0.20	0.82
ARs	955	1.44	2.26
ARsVollz	2	0.00	2.27
AnwStB	65	0.10	2.36
AnwStR	35	0.05	2.42
AnwZ	27	0.04	2.46
AnwZB	1122	1.69	4.15
AnwZBrfg	773	1.17	5.31
AnwZP	1	0.00	5.31
BGs	32	0.05	5.36
BLw	288	0.43	5.80
EnVR	329	0.50	6.29
EnVZ	39	0.06	6.35
EnZB	2	0.00	6.35
EnZR	38	0.06	6.41
GSSt	19	0.03	6.44
GSZ	4	0.01	6.45
KRB	38	0.06	6.50
KVR	109	0.16	6.67
KVZ	56	0.08	6.75
KZB	12	0.02	6.77
KZR	264	0.40	7.17
LwZA	2	0.00	7.17
LwZB	15	0.02	7.19
LwZR	85	0.13	7.32
NotStB	19	0.03	7.35

(continued)

Registerzeichen	Entscheidungen	% Gesamt	% Kumulativ
NotStBrfg	54	0.08	7.43
NotZ	370	0.56	7.99
NotZBrfg	123	0.19	8.18
PatAnwStB	4	0.01	8.18
PatAnwStR	3	0.00	8.19
PatAnwZ	21	0.03	8.22
RiStB	1	0.00	8.22
RiStR	5	0.01	8.23
RiZ	10	0.02	8.24
RiZB	7	0.01	8.25
RiZR	87	0.13	8.38
StB	295	0.44	8.83
StR	22476	33.88	42.71
StbStB	6	0.01	42.72
StbStR	10	0.02	42.73
VGS	1	0.00	42.73
WpStB	2	0.00	42.73
WpStR	3	0.00	42.74
ZA	1272	1.92	44.66
ZB	11308	17.04	61.70
ZR	25406	38.29	100.00
ZRÜ	3	0.00	100.00
Total	66344	100.00	100.00

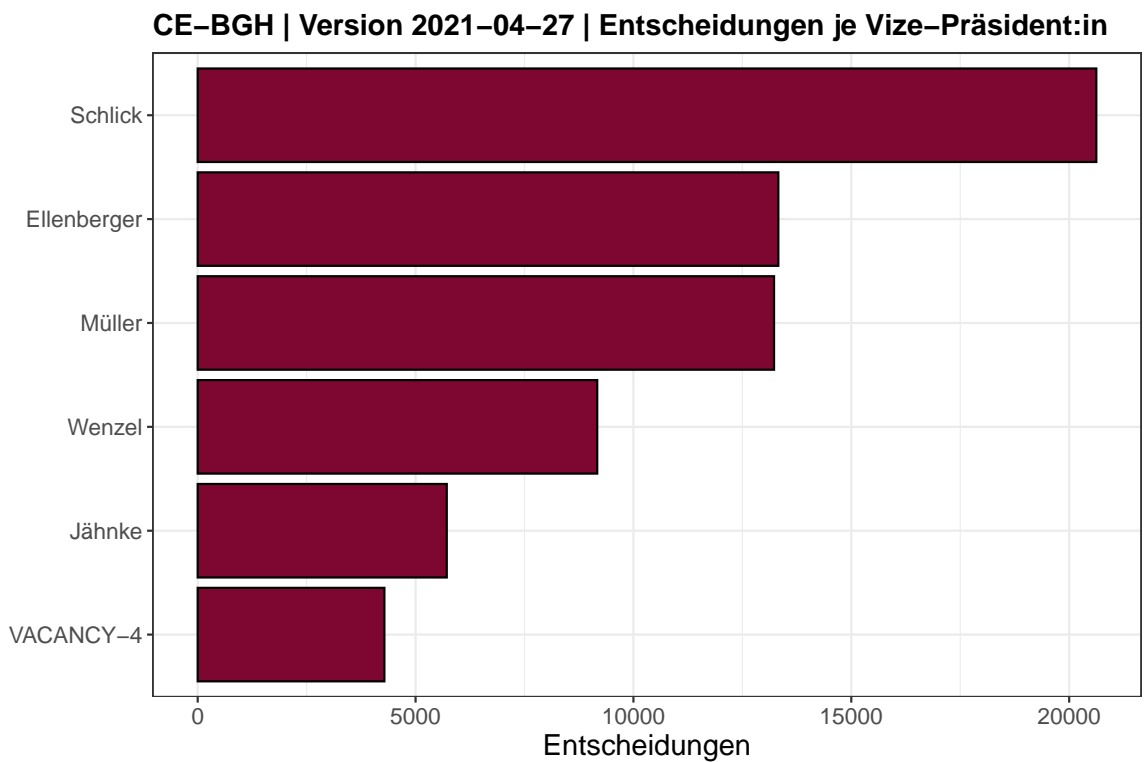
10.6 Nach Präsident:in



DOI: 10.5281/zenodo.4705855

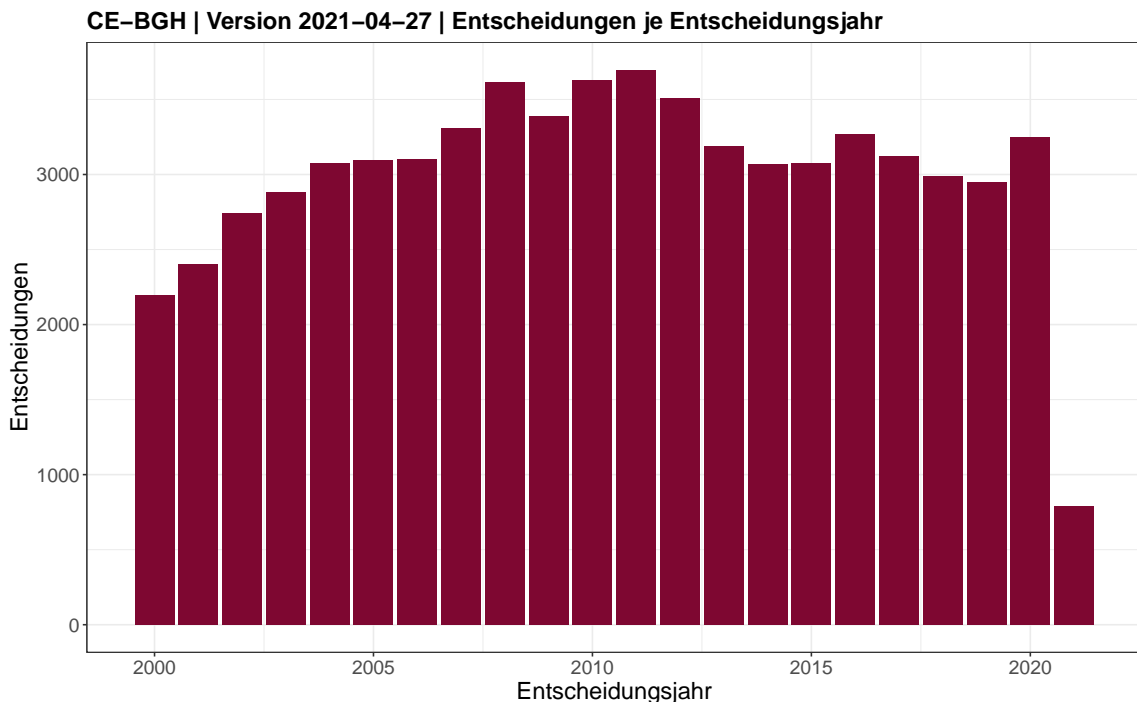
Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Geiss	935	1.41	1.41
Hirsch	21887	32.99	34.40
Limberg	20932	31.55	65.95
Tolksdorf	21013	31.67	97.62
VACANCY-1	288	0.43	98.06
VACANCY-2	1289	1.94	100.00
Total	66344	100.00	100.00

10.7 Nach Vize-Präsident:in



Vize-Präsident:in	Entscheidungen	% Gesamt	% Kumulativ
Ellenberger	13325	20.08	20.08
Jähnke	5716	8.62	28.70
Müller	13227	19.94	48.64
Schlick	20623	31.08	79.72
VACANCY-4	4284	6.46	86.18
Wenzel	9169	13.82	100.00
Total	66344	100.00	100.00

10.8 Nach Entscheidungsjahr



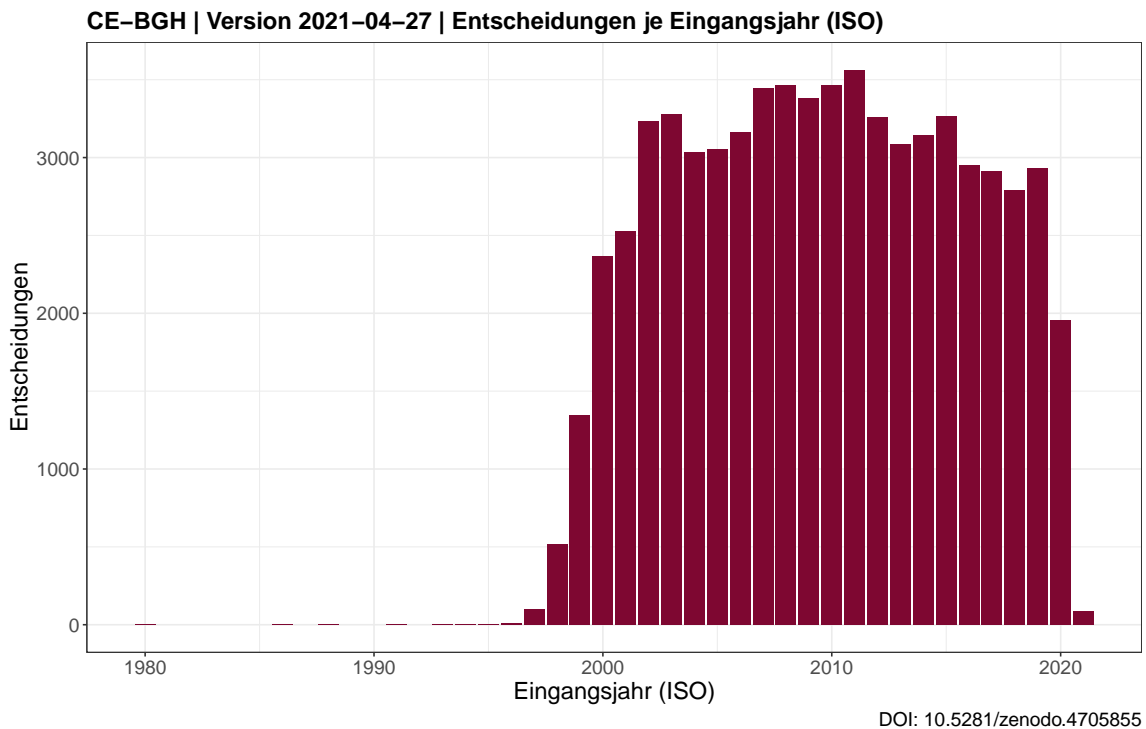
DOI: 10.5281/zenodo.4705855

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2000	2197	3.31	3.31
2001	2404	3.62	6.94
2002	2745	4.14	11.07
2003	2879	4.34	15.41
2004	3078	4.64	20.05
2005	3098	4.67	24.72
2006	3104	4.68	29.40
2007	3310	4.99	34.39
2008	3613	5.45	39.83
2009	3389	5.11	44.94
2010	3625	5.46	50.41
2011	3697	5.57	55.98
2012	3506	5.28	61.26
2013	3190	4.81	66.07
2014	3071	4.63	70.70

(continued)

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2015	3074	4.63	75.33
2016	3271	4.93	80.26
2017	3120	4.70	84.97
2018	2988	4.50	89.47
2019	2947	4.44	93.91
2020	3247	4.89	98.81
2021	791	1.19	100.00
Total	66344	100.00	100.00

10.9 Nach Eingangsjahr (ISO)



Jahr	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	0.15	0.18
1998	515	0.78	0.96
1999	1343	2.02	2.98
2000	2368	3.57	6.55
2001	2530	3.81	10.37
2002	3231	4.87	15.24
2003	3279	4.94	20.18

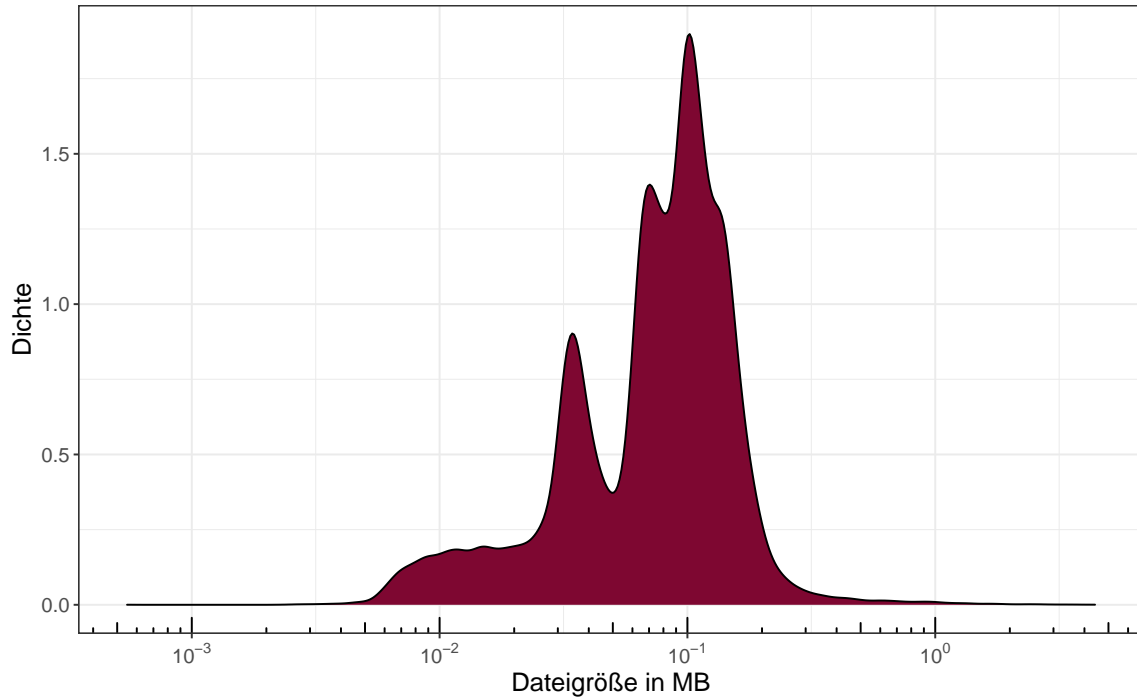
(continued)

Jahr	Entscheidungen	% Gesamt	% Kumulativ
2004	3037	4.58	24.76
2005	3054	4.60	29.36
2006	3161	4.76	34.12
2007	3446	5.19	39.32
2008	3464	5.22	44.54
2009	3384	5.10	49.64
2010	3465	5.22	54.86
2011	3563	5.37	60.23
2012	3257	4.91	65.14
2013	3086	4.65	69.79
2014	3143	4.74	74.53
2015	3268	4.93	79.46
2016	2954	4.45	83.91
2017	2912	4.39	88.30
2018	2788	4.20	92.50
2019	2933	4.42	96.92
2020	1953	2.94	99.87
2021	89	0.13	100.00
Total	66344	100.00	100.00

11 Dateigrößen

11.1 Verteilung PDF-Dateigrößen

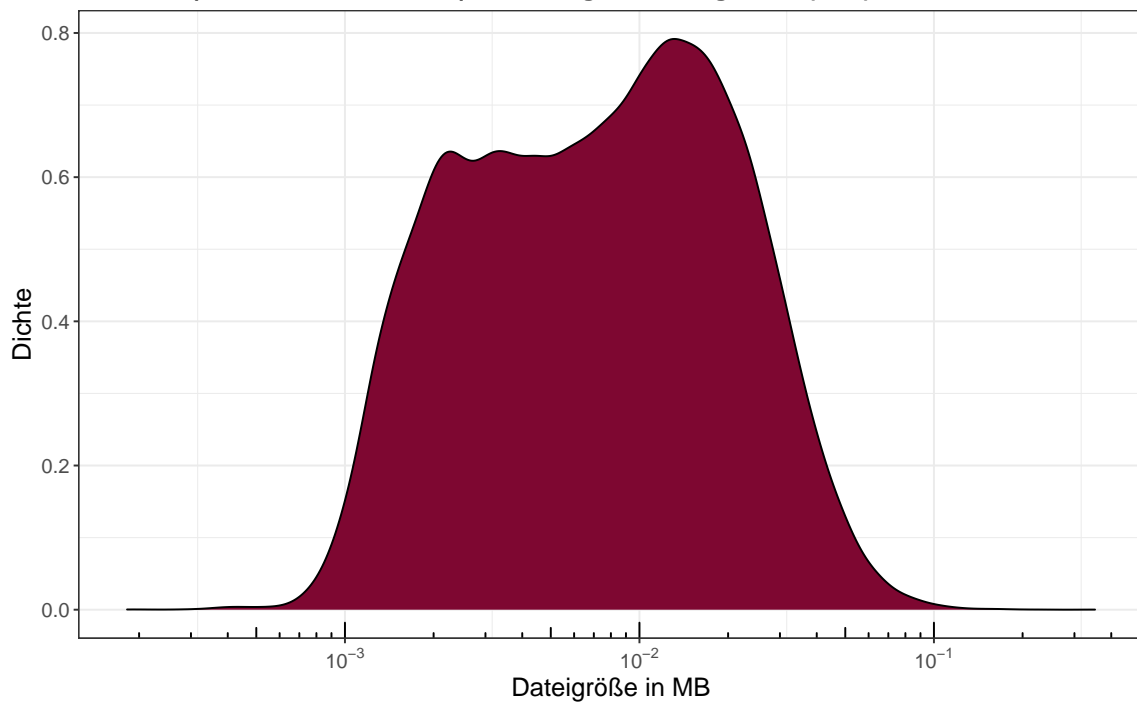
CE-BGH | Version 2021-04-27 | Verteilung der Dateigrößen (PDF)



DOI: 10.5281/zenodo.4705855

11.2 Verteilung TXT-Dateigrößen

CE-BGH | Version 2021-04-27 | Verteilung der Dateigrößen (TXT)



DOI: 10.5281/zenodo.4705855

11.3 Gesamtgröße je ZIP-Archiv

Datei	Größe in MB
CE-BGH_2021-04-27_DE_ANALYSE.zip	1.51
CE-BGH_2021-04-27_DE_CSV_Datensatz.zip	212.62
CE-BGH_2021-04-27_DE_CSV_Metadaten.zip	2.63
CE-BGH_2021-04-27_DE_PDF_Datensatz.zip	5,525.27
CE-BGH_2021-04-27_DE_PDF_Entscheidungen-mit-Namen.zip	562.04
CE-BGH_2021-04-27_DE_PDF_Leitsatz-Entscheidungen.zip	2,134.47
CE-BGH_2021-04-27_DE_TXT_Datensatz.zip	286.08
CE-BGH_2021-04-27_Source_Files.zip	0.06

12 Prüfung kryptographischer Signaturen

12.1 Allgemeines

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

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

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

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

12.2 Persönliche GPG-Signatur

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

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

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

12.3 Import: Public Key

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

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

12.4 Prüfung: GPG-Signatur der Hash-Datei

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

```
## [1] "CE-BGH_2021-04-27_KryptographischeHashes.csv"
```

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

```
## [1] "CE-BGH_2021-04-27_FobbeSignaturGPG_Hashes.gpg"
```

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

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

Ergebnis

gpg: Signature made Tue 27 Apr 2021 12:41:40 PM CEST

gpg: using RSA key FE6FB888F0E5656C1D253B9A50C41384F44A4E42

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

12.5 Prüfung: SHA3-512 Hashes der ZIP-Archive

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

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

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

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

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

Datei	Ergebnis
CE-BGH_2021-04-27_DE_ANALYSE.zip	Verifiziert!
CE-BGH_2021-04-27_DE_CSV_Datensatz.zip	Verifiziert!
CE-BGH_2021-04-27_DE_CSV_Metadaten.zip	Verifiziert!
CE-BGH_2021-04-27_DE_PDF_Datensatz.zip	Verifiziert!
CE-BGH_2021-04-27_DE_PDF_Entscheidungen-mit-Namen.zip	Verifiziert!
CE-BGH_2021-04-27_DE_PDF_Leitsatz-Entscheidungen.zip	Verifiziert!
CE-BGH_2021-04-27_DE_TXT_Datensatz.zip	Verifiziert!
CE-BGH_2021-04-27_Source_Files.zip	Verifiziert!

13 Changelog

Version	Details
2021-04-27	<ul style="list-style-type: none">• Vollständige Aktualisierung der Daten• Veröffentlichung des vollständigen Source Codes• Deutliche Erweiterung des inhaltlichen Umfangs des Codebooks• Einführung der vollautomatischen Erstellung von Datensatz und Codebook• Einführung von Compilation Reports um den Erstellungsprozess exakt zu dokumentieren• Einführung von Variablen für 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
2020-07-09	<ul style="list-style-type: none">• Erstveröffentlichung

14 Parameter für strenge Replikationen

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

```
## R version 4.0.4 (2021-02-15)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora 33 (Workstation Edition)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libopenblas-r0.3.12.so
##
## locale:
## [1] LC_CTYPE=en_US.utf8      LC_NUMERIC=C
## [3] LC_TIME=en_US.utf8       LC_COLLATE=en_US.utf8
## [5] LC_MONETARY=en_US.utf8   LC_MESSAGES=en_US.utf8
## [7] LC_PAPER=en_US.utf8      LC_NAME=C
## [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.utf8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] parallel stats graphics grDevices utils datasets methods
## [8] base
##
## other attached packages:
## [1] magick_2.7.1      quanteda_2.1.2    readtext_0.80     data.table_1.14.0
## [5] scales_1.1.1     ggplot2_3.3.3    doParallel_1.0.16 iterators_1.0.13
## [9] foreach_1.5.1    pdftools_2.3.1   kableExtra_1.3.4  knitr_1.31
## [13] rvest_1.0.0      httr_1.4.2       mgsub_1.7.2       fs_1.5.0
##
## loaded via a namespace (and not attached):
## [1] tinytex_0.30      qpdf_1.1          tidyselect_1.1.0  xfun_0.22
## [5] purrr_0.3.4      lattice_0.20-41   colorspace_2.0-0  vctrs_0.3.6
## [9] generics_0.1.0   htmltools_0.5.1.1 viridisLite_0.3.0 yaml_2.2.1
## [13] utf8_1.2.1       rlang_0.4.10     pillar_1.5.1     glue_1.4.2
## [17] withr_2.4.1      selectr_0.4-2    lifecycle_1.0.0  stringr_1.4.0
## [21] munsell_0.5.0    gtable_0.3.0     codetools_0.2-18 evaluate_0.14
## [25] labeling_0.4.2   curl_4.3          fansi_0.4.2       highr_0.8
## [29] Rcpp_1.0.6       RcppParallel_5.0.3 webshot_0.5.2     farver_2.1.0
## [33] systemfonts_1.0.1 fastmatch_1.1-0   stopwords_2.2     askpass_1.1
## [37] digest_0.6.27    stringi_1.5.3    dplyr_1.0.5       grid_4.0.4
## [41] tools_4.0.4      magrittr_2.0.1   tibble_3.1.0     crayon_1.4.1
## [45] pkgconfig_2.0.3  Matrix_1.3-2     ellipsis_0.3.1   xml2_1.3.2
## [49] rmarkdown_2.7    svglite_2.0.0    rstudioapi_0.13  R6_2.5.0
## [53] compiler_4.0.4
```

Literaturverzeichnis

- Analytics, Revolution, and Steve Weston. 2020. *Iterators: Provides Iterator Construct*. <https://github.com/RevolutionAnalytics/iterators>.
- Benoit, Kenneth, and Adam Obeng. 2020. *Readtext: Import and Handling for Plain and Formatted Text Files*. <https://github.com/quanteda/readtext>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. “Quanteda: An R Package for the Quantitative Analysis of Textual Data.” *Journal of Open Source Software* 3 (30): 774. <https://doi.org/10.21105/joss.00774>.
- Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, Jiong Wei Lua, Jouni Kuha, and William Lowe. 2020. *Quanteda: Quantitative Analysis of Textual Data*. <https://quanteda.io>.
- Corporation, Microsoft, and Steve Weston. 2020. *DoParallel: Foreach Parallel Adaptor for the Parallel Package*. <https://CRAN.R-project.org/package=doParallel>.
- Dowle, Matt, and Arun Srinivasan. 2021. *Data.table: Extension of ‘Data.frame’*. <https://CRAN.R-project.org/package=data.table>.
- Ewing, Mark. 2020. *Mgsub: Safe, Multiple, Simultaneous String Substitution*. <https://CRAN.R-project.org/package=mgsub>.
- Hester, Jim, and Hadley Wickham. 2020. *Fs: Cross-Platform File System Operations Based on Libuv*. <https://CRAN.R-project.org/package=fs>.
- Ooms, Jeroen. 2020. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Revolution Analytics, and Steve Weston. n.d. *Foreach: Provides Foreach Looping Construct*.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2020. *Httr: Tools for Working with Urls and Http*. <https://CRAN.R-project.org/package=httr>.
- . 2021. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2020. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization*. <https://CRAN.R-project.org/package=scales>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich

Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.

———. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with Kable and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.