Digitale Richterschaft
3. Mai 2023
15.00 Uhr

Legal Data Science

Der moderne Weg zur Wahrheit

Seán Fobbe

Legal Data Science

Der moderne Weg zur Wahrheit

Seán Fobbe

Ludwig-Maximilians-Universität München

3. Mai 2023

Digitale Richterschaft

Struktur der Präsentation

1. Wahrheit
2. Über Legal Data Science
3. Theorie: Verteilungen
4. Theorie: Repräsentativität
5. Praxisteil

Über diese Präsentation

Umfang

- Präsentation
- Notizen
- Source Code
- Vertiefungshinweise

Open Access verfügbar nächste Woche

Direkter Link	www.doi.org/10.5281/zenodo.7877804
Blog	www.seanfobbe.de

Praxis: Arbeit mit der Programmiersprache »R«

»R« und »Python« sind die dominanten Sprachen im Data Science-Bereich

- Wir arbeiten heute mit R, genauer mit der WebR-Variante für Browser
- Rufen Sie bitte https://webr.r-wasm.org/latest/ auf
- Herunterladen dauert 1–2 Minuten
- Wird lokal auf ihrem Computer in einer Sandbox ausgeführt

Teil I

Wahrheit

Erwartung

Wahrheit und Gerechtigkeit

Realität

Vertretbarkeit und Dogmatik

Rechtswissenschaft II

Methodenlehre oder Methodenleere?

- Rechtswissenschaft ist heute (fast) reine Literaturwissenschaft
- Logisches Denken? Nur implizit, moderne Logik wird kaum gelehrt oder genutzt
- Keine Empirie? Wie realitätsfremd darf Rechtswissenschaft sein?

Methodenleere in der Rechtswissenschaft

»Jurist:innen können sich in alles einarbeiten« ist die moderne Lebenslüge der Rechtswissenschaft.

Wahrheit in der Moderne

Eins, zwei, drei, viele? Allgemeine oder breit anwendbare Aussagen beruhen fast immer auf quantitativen Belegen mit Unsicherheiten

- Umfragen zu Wahlen, Präferenzen, Märkten
- Infektionsschutz (7-Tage-Inzidenz, Intensivbettenquote)
- Wirksamkeit von Medikamenten
- Klimawandel (Erderwärmung, Dürre, Extremwetter)
- Crime Mining (HessenData)
- Personalbedarf in der Justiz (PEBB§Y)

Moderne Wahrheit

Moderne Wahrheiten sind oft quantitativ und probabilistisch.

Rechtswissenschaft III

Sprachlosigkeit und ihre Konsequenzen

- Sprachlosigkeit der Dogmatik — es fehlen Vokabular und Methoden um Sachverhalte zu korrekt beschreiben
- Falscher Sachverhalt $\Rightarrow$ falsche Subsumtion
- Fehlende Kompetenz bei Jurist:innen führt nicht zu Zurückhaltung bei der Bewertung empirischer Sachverhalte
- Beispiele: Entscheidungszwang bei Gerichten, fehlendes Problembewusstein, Selbstüberschätzung

Konsequenzen

- Digitalisierung von oben und von außen
- Kaum Anschluss an hochwertige empirische Forschung und Methoden

Materielle Wahrheitstheorien I

Definitionen

Korrespondenztheorie

Wahrheit ist die Korrespondenz von Aussagen (allgemeiner: »Wahrheitsträger«) mit einer objektiven Wirklichkeit.

Kohärenztheorie

Wahrheit ist die Kohärenz von Aussagen mit einer bestimmten Menge anderer Aussagen.

Die Korrespondenztheorie ist das dominierende philosophische Paradigma der Naturwissenschaften, die Kohärenztheorie das der Geisteswissenschaften

Materielle Wahrheitstheorien II

Wahrheitsähnlichkeit

Matrix-Problem

Woher wissen wir, dass wir nicht in der Matrix (d.h. einer täuschend echten Computersimulation) leben?

- Wahrheitsähnlichkeit als Alternative?
- Wie bestimmen wir den Abstand zur »echten« Wahrheit, wenn wir die »echte« Wahrheit nicht kennen?

Materielle Wahrheitstheorien III

Wissenschaftliche Näherung an Wahrheit

- Wahrheit ist ein regulatives Ideal im Sinne Kants: unerreichbar, aber als normative Zielmarke überragend wichtig
- Die Vorhersagekraft von wissenschaftlichen Theorien ist der Gradmesser ihrer Wahrheit und Güte
- Empirie ist der Abgleich von Erwartungen (Hypothesen) mit Erfahrungen (Daten)
- »Wahr-schein-lichkeit« als rationale Annäherung an Wahrheit durch quantitative und qualitative Bestimmung der Unsicherheit

Wahrscheinlichkeit im Recht I

Aussagen mit Wahrscheinlichkeitsgehalt

Wahrscheinlichkeitsaussagen sind in Rechtswissenschaft und -praxis allgegenwärtig

- »Allgemeine Lebenserfahrung«, »lebensnahe Auslegung«, »lebensfremd«
- »Anstandsgefühl aller billig und gerecht Denkenden«
- herrschende Meinung, bzw. allgemeine Ansicht
- abstrakte/konkrete/gegenwärtige Gefahren (Sicherheitsrecht)
- Prognosen über zukünftiges Verhalten der Angeklagten (Strafrecht)

Wahrscheinlichkeit im Recht II

Begriffe der Wahrscheinlichkeit

- Hinreichende Wahrscheinlichkeit (z.B. Gefahrbegriffe)
- Überwiegende Wahrscheinlichkeit (z.B. Glaubhaftmachung)
- »mit an Sicherheit grenzender Wahrscheinlichkeit«
  (z.B. Art. 83 II 2 BayPAG, finaler Rettungsschuss)
- »für das praktische Leben brauchbare[r] Grad von Gewißheit, der den Zweifeln Schweigen gebietet, ohne sie völlig auszuschließen« (Vollbeweis ZPO, BGH in »Anastasia«)

Mathematische Wahrscheinlichkeitstheorien

Frequenztheorie (Fisher, h.M.)

Wahrscheinlichkeit ist die relative Frequenz eines Ereignisses bei unendlich vielen Wiederholungen eines Zufallsexperimentes.

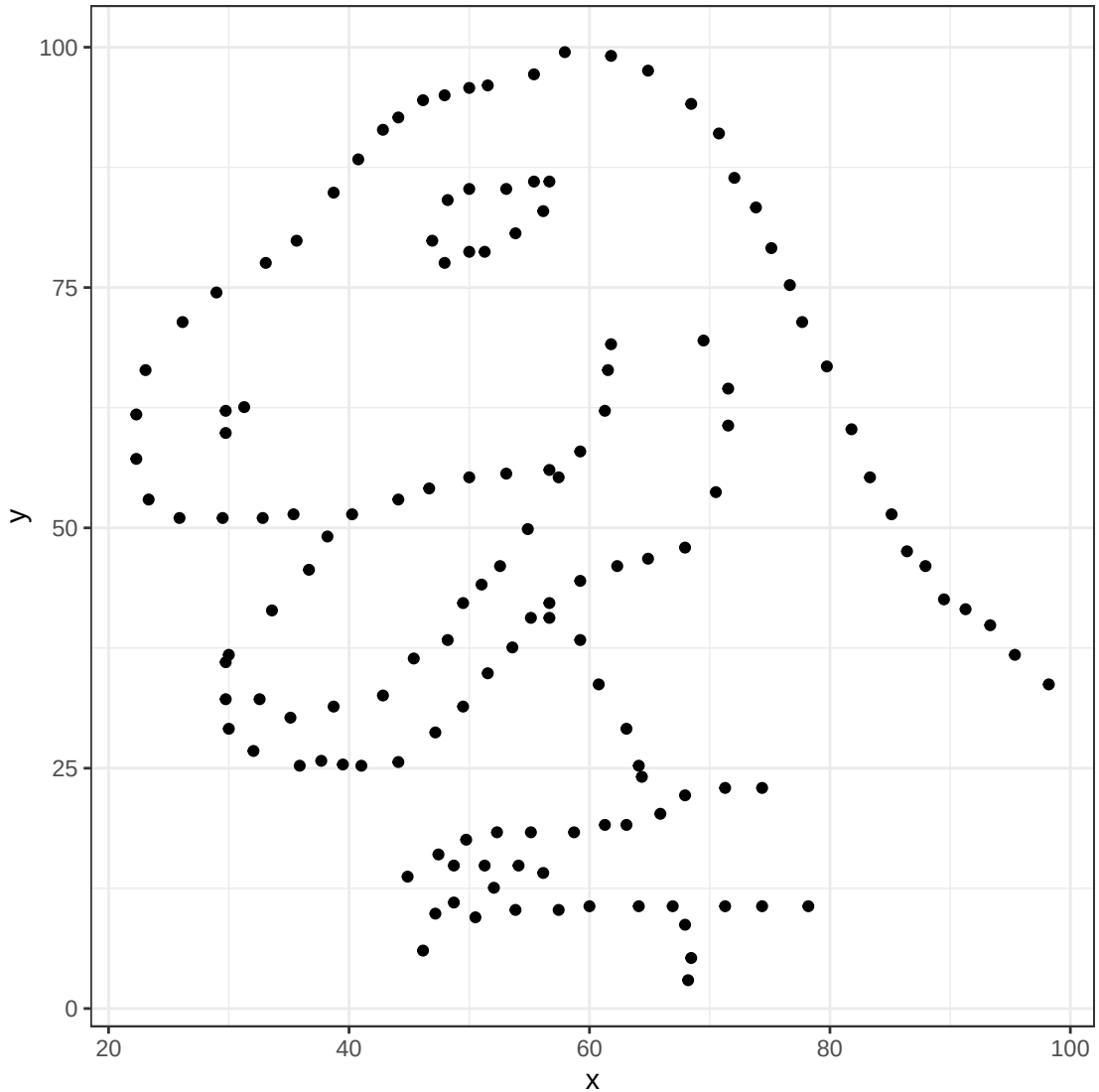
Logische Theorie (Laplace/Jeffreys/Cox/Jaynes)

Wahrscheinlichkeit ist der Anteil der tatsächlich möglichen Pfade zu einem Ereignis, geteilt durch die maximal möglichen Pfade zu allen möglichen Ereignissen.

Subjektive Theorie (de Morgan/de Finetti)

Wahrscheinlichkeit ist ein subjektiver Grad an Überzeugung.

Der Datasaurus



Quelle: Rhian Davies u. a., »datasauRus: Datasets from the Datasaurus Dozen« (*CRAN*, 4. Mai 2022) ⟨https://cran.r-project.org/web/packages/datasauRus/⟩. Der originale Datasaurus ist von Alberto Cairo, die konkrete Berechnungsmethode von Matejka und Fitzmaurice (2017).

Der Datasaurus

Eine Metapher für »künstliche Intelligenz«

Der Datasaurus bietet eine Metapher für »Künstliche Intelligenz«

- Menschen erkennen in den Punkten einen Dinosaurier
- Für die Maschine sind es nur Zahlen und Punkte ohne Bedeutung, sie »versteht« sie nicht
- Das ursprüngliche Input (das Bild eines Dinosauriers) und der modifizierende Algorithmus wurden durch Menschen erstellt

Künstliche Intelligenz ist nicht künstlich

Am Anfang und Ende von »künstlicher Intelligenz« stehen immer Menschen, es findet nur eine Art »Datenwäsche« statt. Entsprechende Modelle sind eher destilliertes menschliches Verhalten, inklusiver aller Fehlbarkeiten.

Künstliche Intelligenz ist eine Illusion

- Es gibt nur Statistik, die Sie verstehen und Statistik, die Sie nicht verstehen
- Es gibt auch Statistik, die ihre Erfinder:innen nicht verstehen (z.B. Deep Learning)
- Ein Werkzeug muss nicht denken können, um nützlich zu sein
- KI-Mystizismus ist eine Gefahr für den Rechtsstaat

Mathematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not cast a spell over him.

People v Collins 68 Cal 2d 319, 320 (Supreme Court of California 1968)

Teil II

Über Legal Data Science

Was ist Legal Data Science?

Überblick

Data Science Teildisziplinen (Conway)

1. Programmieren
2. Statistik
3. Domänenkenntnis (z.B. Jura)

Arbeitsdefinition »Legal Data Science«

Das Trennen von relevanten und irrelevanten Informationen (»Data«) durch reproduzierbare quantitative Analysen (»Science«), um juristische Fragestellungen zu beantworten oder Daten aus juristischen Kontexten zu analysieren.

Kenntnisstufen in Legal Data Science

1. Unwissenheit ($\approx$ reine Dogmatik)
2. Fähigkeit, Fragen zu stellen ($\approx$ Legal Data Analyst)
3. Fähigkeit, Antworten zu geben ($\approx$ Legal Data Scientist)

Warum Legal Data Science?

- Bottom-Up Entwicklung von Lösungen für Probleme der Justiz
- Verständnis beim Top-Down Einkauf und der Entwicklung von GovTech-Lösungen
- Verbesserte Beweisführung und dogmatische Bewertungen
- Nutzung komplexerer GovTech-Lösungen als Anwender:in
- Informierte Teilnahme an weitreichenden Justizreformen wie PEBB§Y
- Beschleunigung von Arbeitsabläufen durch Erkennen von Problemen und Gefühl für Lösungen

Natural Language Processing (NLP)

Oder Legal Language Processing?

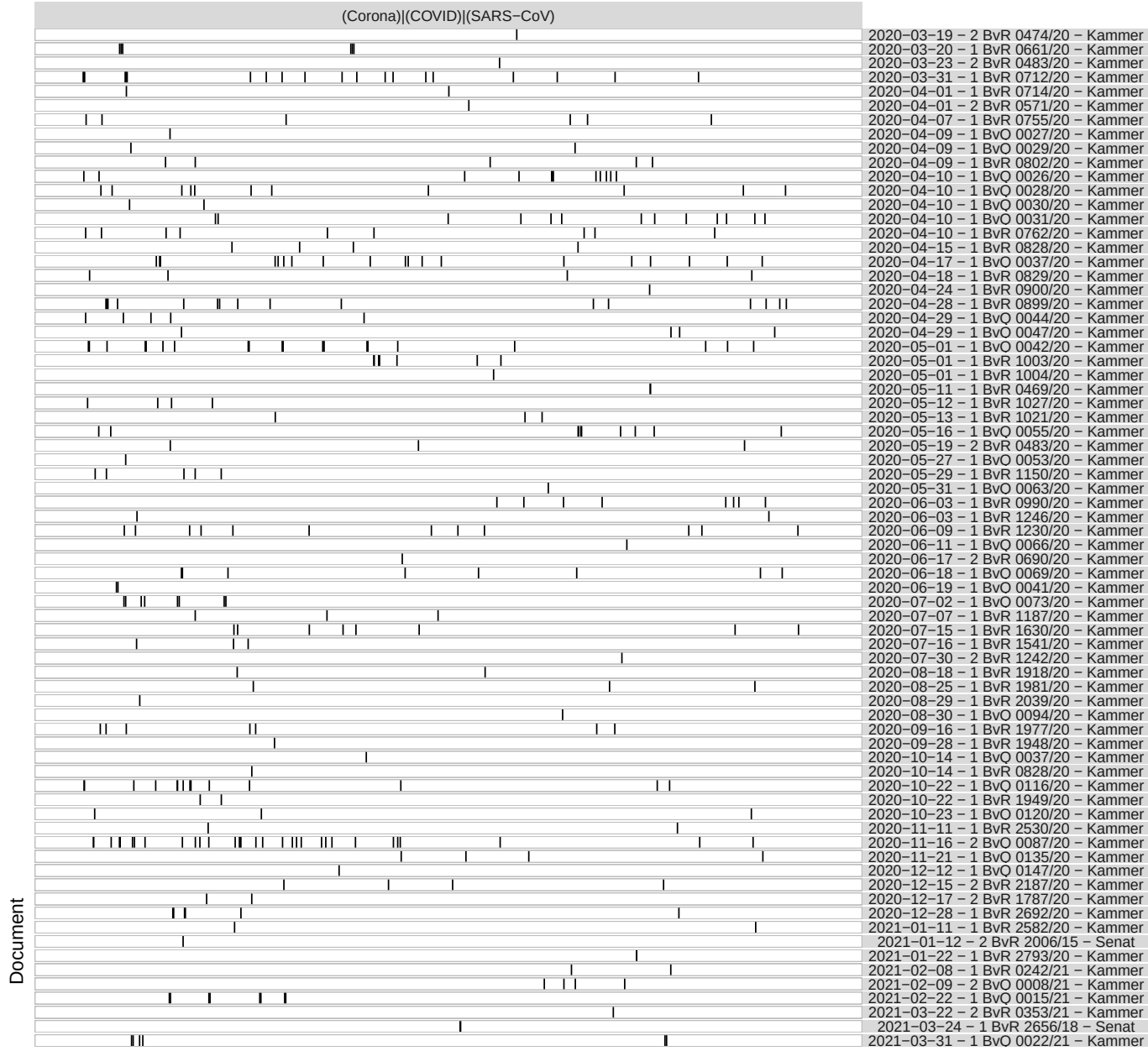
Natural Language Processing (NLP)

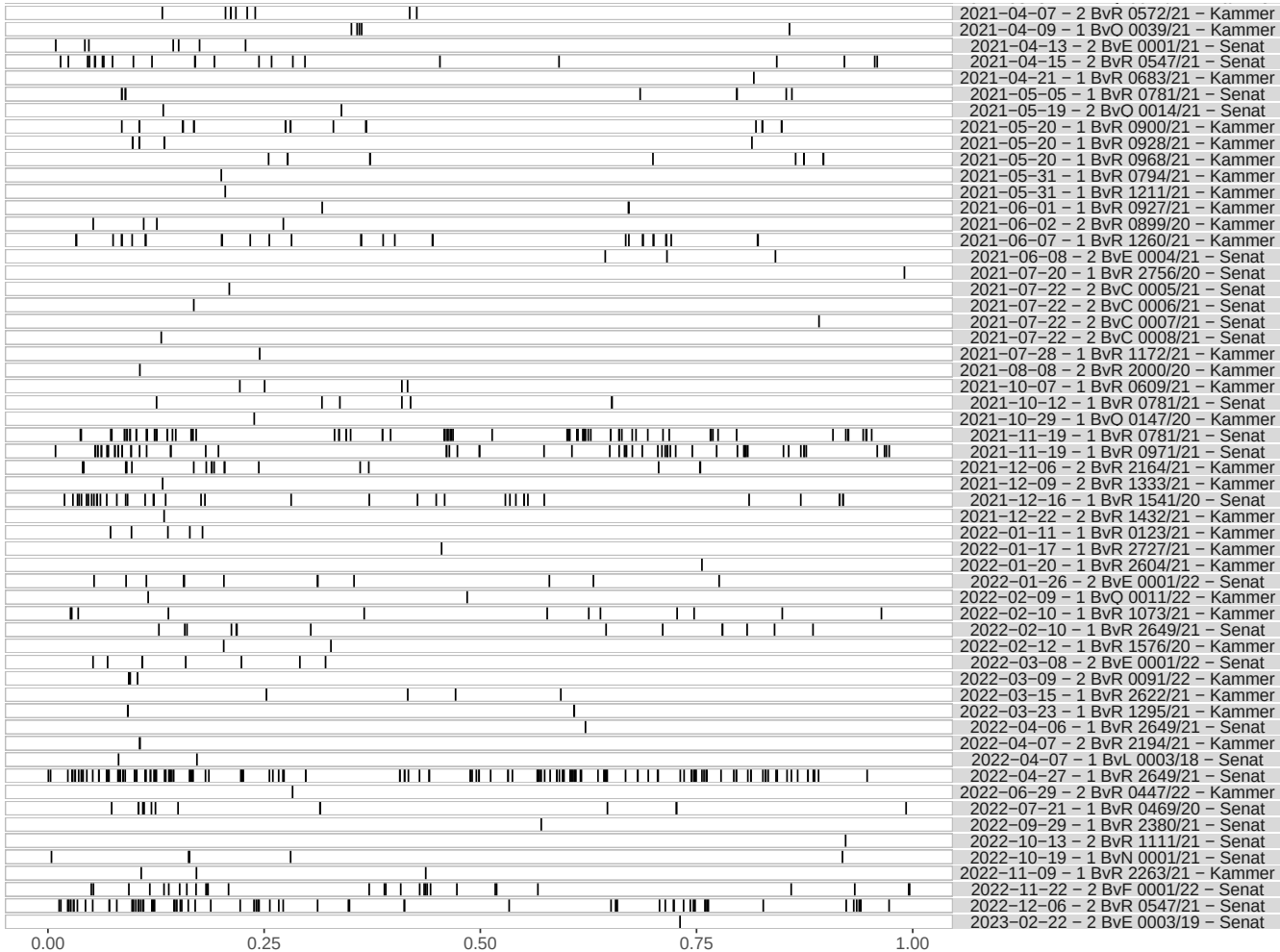
Natural Language Processing ist die maschinelle Verarbeitung und Analyse von Textdaten.

Anwendungsfälle

- Finden von relevanten Texten
- Zusammenfassung von langen Texten
- Extraktion von bestimmten Informationen
- Generierung passgenauer Texte (Massenklagen)

BVerfG−Corona | Version 2023−02−26 | Lexical Dispersion Plot





Fobbe | DOI: 10.5281/zenodo.7807020

Legal Judgment Prediction

Oder: die Bedeutung von Domänenkenntnis

Legal Judgment Prediction

Legal Judgment Prediction ist die datenbasierte Vorhersage des Ausgangs von Gerichtsverfahren (Vertiefung: Medvedeva 2022).

Beliebter Versuchsaufbau: Einlesen des Sachverhalts einer Entscheidung, Vorhersage des Ergebnisses mittels neuronaler Netzwerke

Probleme

1. Text der Entscheidung (inkl. Sachverhalt) ist vor der Verkündung nur dem Gericht bekannt
2. Entscheidung fällt oft vor Niederschrift des Sachverhalts (nur »relevante« Fakten im Sachverhalt)
3. Entscheidung ergeht oft auf Basis der Hauptverhandlung, nicht der Akte (Ausnahme: rein schriftliches Verfahren)

Entscheidungsbäume

Entscheidungsbäume bringen Transparenz in Entscheidungsprobleme und können mit Wahrscheinlichkeiten angereichert werden.

Anwendungsfälle

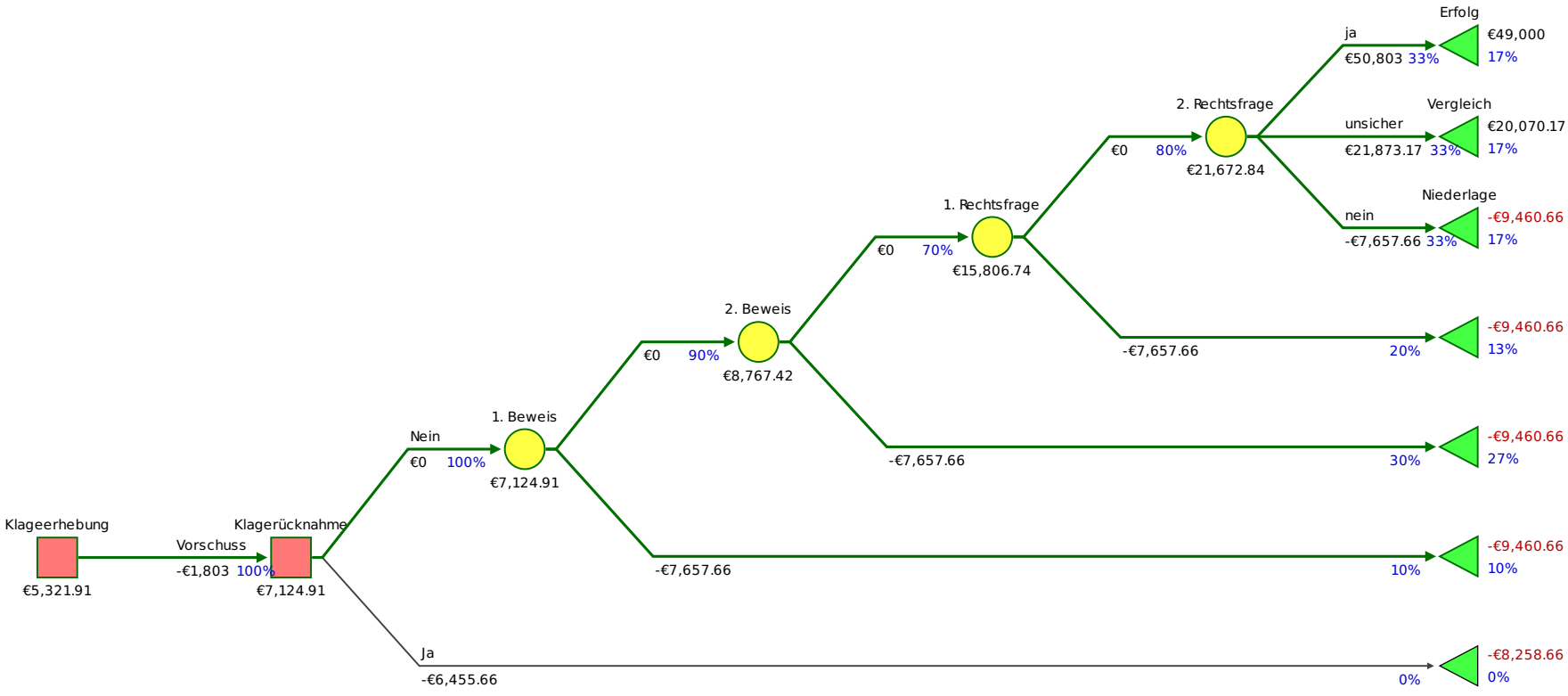
- Prozessrisikoanalyse / Legal Judgment Prediction
- Darstellung der Struktur einer Norm / eines Rechtsproblems
- Hilfe bei der Erstellung und Interpretation von Gesetzen

Empfehlung

Silver Decisions — http://silverdecisions.pl/ (Open Source)

Klage mit 49.000 € Streitwert

Zwei Beweise, zwei Rechtsfragen, mit Vergleichsoption



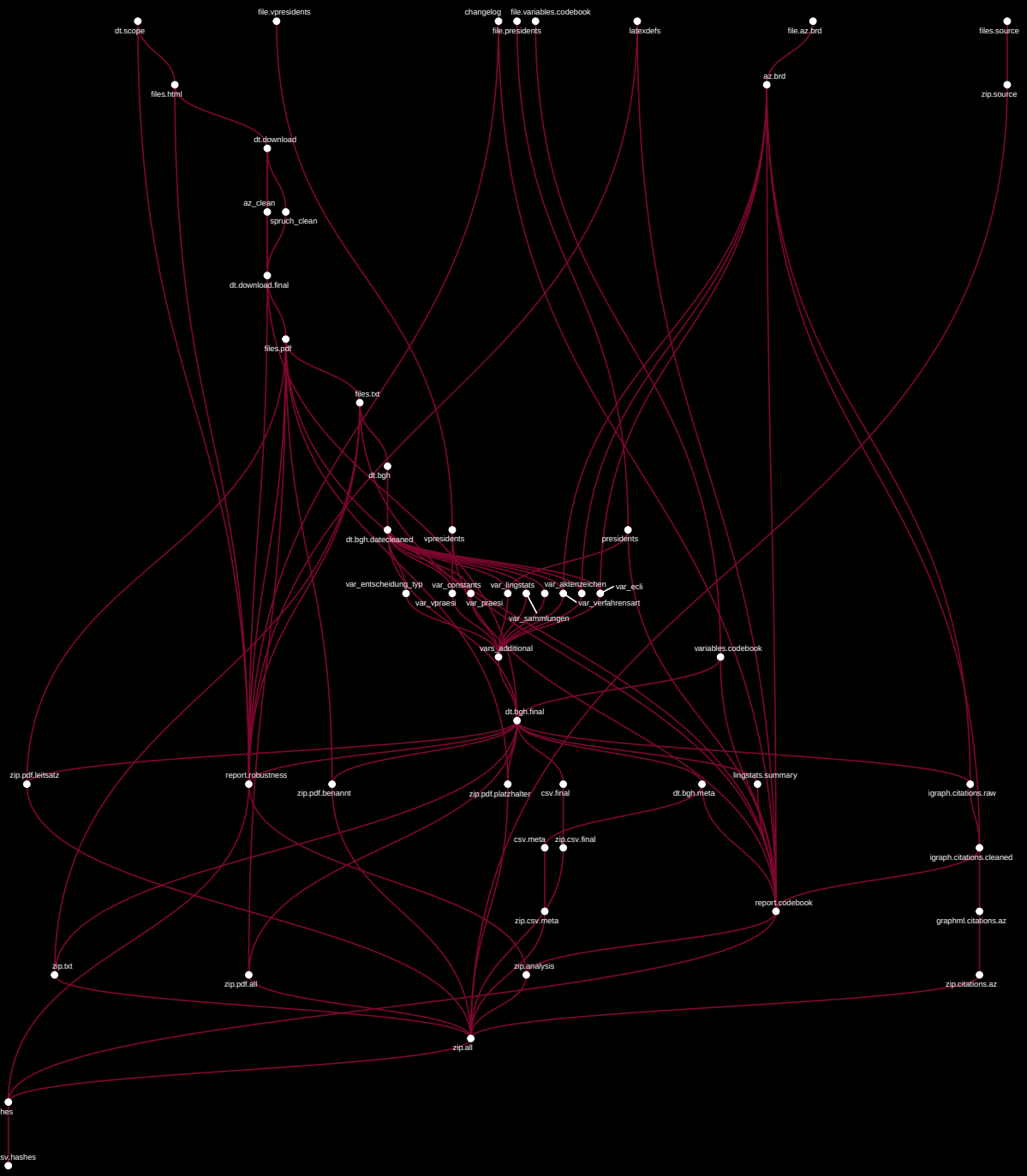
Legal Data Engineering

- Sammlung juristischer Daten
- Strukturierte Aufbereitung
- Qualitätskontrolle
- Dokumentation
- Auslieferung an Endnutzer:innen

ETL Pipeline

Die Abkürzung »ETL« im Data Engineering steht für »extract, transform, load«, d.h. die Sammlung, Aufbereitung und Auslieferung von Datensätzen/-strömen.

CE−BGH | Version 2023−03−10 | Vollständiger Prozess der Datensatz−Kompilierung



March 10, 2023

Dataset Open Access

Corpus der Entscheidungen des Bundesgerichtshofs (CE-BGH)

Fobbe, Sean

Überblick

Das **Corpus der Entscheidungen des Bundesgerichtshofs (CE-BGH)** ist der bislang größte, frei verfügbare Datensatz von Entscheidungen des Bundesgerichtshofs. Er ist eine Zusammenstellung aller Entscheidungen, die in der amtlichen Datenbank des Bundesgerichtshofs am jeweiligen Stichtag veröffentlicht waren.

Bitte beachten Sie das beiliegende Codebook! Es enthält wichtige Informationen zur korrekten Nutzung des Datensatzes. Es hilft auch bei der Entscheidung, welche Variante für Sie am besten geeignet ist. In der Regel empfehle ich für quantitative Forschung die CSV-Dateien und für traditionelle Forschung die PDF-Sammlung.

Für Praktiker:innen stelle ich zusätzlich nach Senat sortierte PDF-Sammlungen aller ***Leitsatzentscheidungen*** und aller ***Entscheidungen mit Namen*** (z.B. »Trabrennbahn«) zur Verfügung.

Ab Version 2023-03-10 ist auch das gesamte Zitationsnetzwerk des BGH (nur Aktenzeichen) für die einfache Nutzung mit graphischer Software wie Gephi oder für die maschinelle Weiterverarbeitung als GraphML verfügbar.

Aktualisierung

Dieser Datensatz wird ***1-2 mal im Jahr*** aktualisiert. Benachrichtigungen über neue und aktualisierte Datensätze veröffentliche ich immer zeitnah auf Mastodon unter @seanfobbe@fediscience.org

Bekannte Fehler

- Die Variable "bghr" ist in Version 2023-03-10 fehlerhaft und sollte nicht verwendet werden. Ein Bugfix ist für den nächsten Release vorgesehen

NEU in 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

Publication date:
March 10, 2023

DOI:
DOI 10.5281/zenodo.7699032

Keyword(s):
Bundesgerichtshof, BGH, Bundesgericht, Oberster Gerichtshof des Bundes, Ordentliche Gerichtsbarkeit, Revision, Urteil, Beschluss, Entscheidung, Richter, Gericht, Open Legal Data, Deutschland, Bundesrepublik Deutschland, Dienstgericht des Bundes, Zivilrecht, Strafrecht, Patentrecht, Kartellrecht, Urheberrecht, Gesellschaftsrecht, Erbrecht, Sachenrecht, Kaufrecht, Werkvertragsrecht, Insolvenzrecht, Bankrecht, Vergaberecht, Familienrecht, Decision, Judgment, Court, Text Mining, Text-as-data, Federal Court of Justice, Germany, Federal Republic of Germany

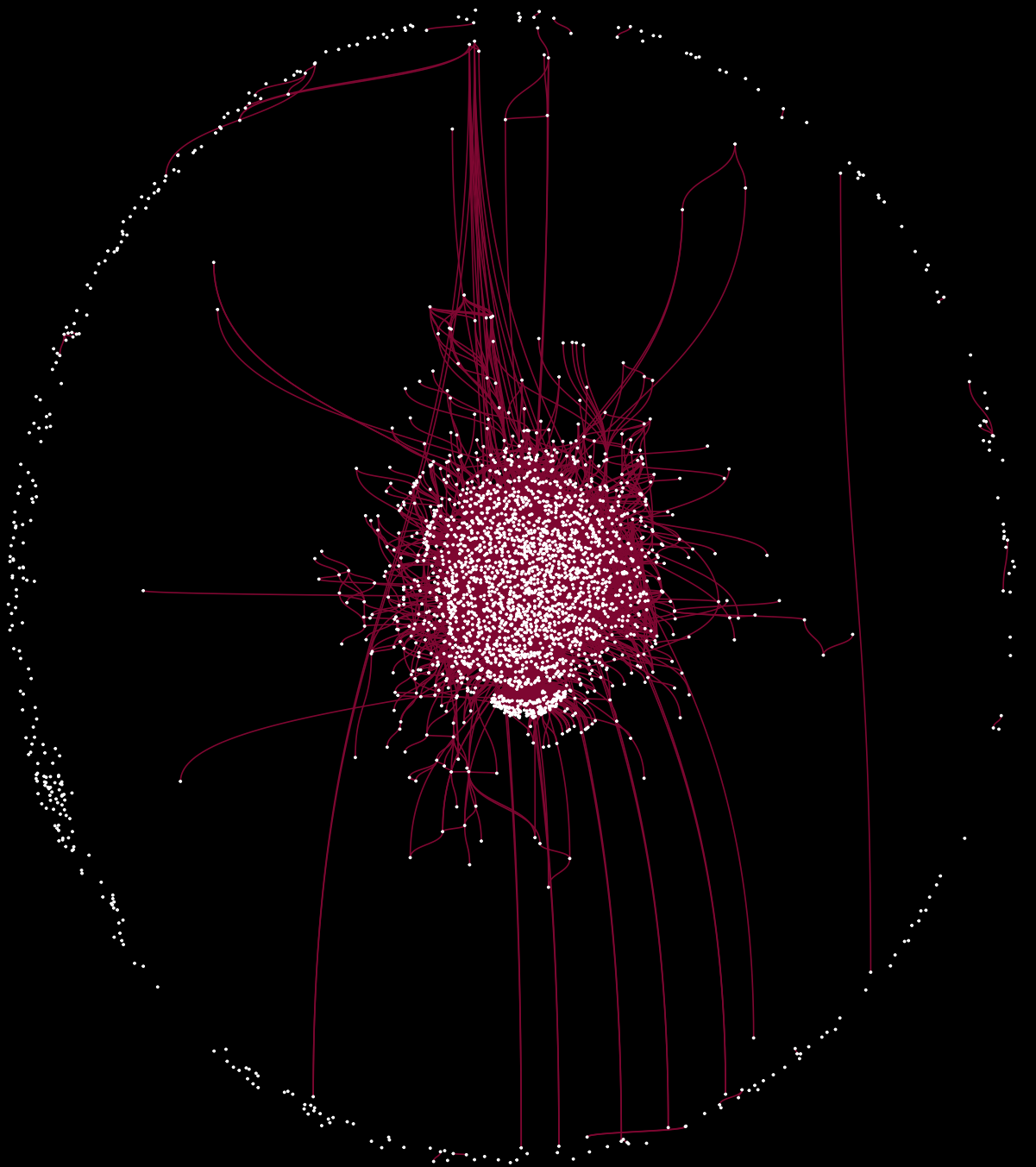
Related identifiers:
Compiled by
10.5281/zenodo.7699033 (Software)

Derived from
https://www.bundesgerichtshof.de

References
https://github.com/seanfobbe/ce-bgh (Software)

Communities:
Open Access Data Sets by Seán Fobbe

CE−BGH | Version 2023−03−10 | Das interne Zitationsnetzwerk (Aktenzeichen) des 11. Zivilsenats des Bundesgerichthofs



Teil III

Theorie: Verteilungen

Die Grenze von Abstraktion

The Law of Leaky Abstractions

All non-trivial abstractions, to some degree, are leaky.

Joel Spolsky

PEBB§Y

Personal**b**edarfs**b**erechnungs**sy**stem. System zur Ermittlung des Personalbedarfs in der Justiz seit 2001.

- Letzte Erhebung 2014, nächste Erhebung 2027
- Selbstaufschreibung durch Richter:innen, Staats- und Amtsanwält:innen, Rechtspfleger:innen und Service-Einheiten
- Ermittlung einer Basiszahl (durchschnittliche Bearbeitungszeit) je »Produktkategorie«, z.B. 137 Minuten pro allgemeine Jugendstrafsache am Amtsgericht (RA 210)

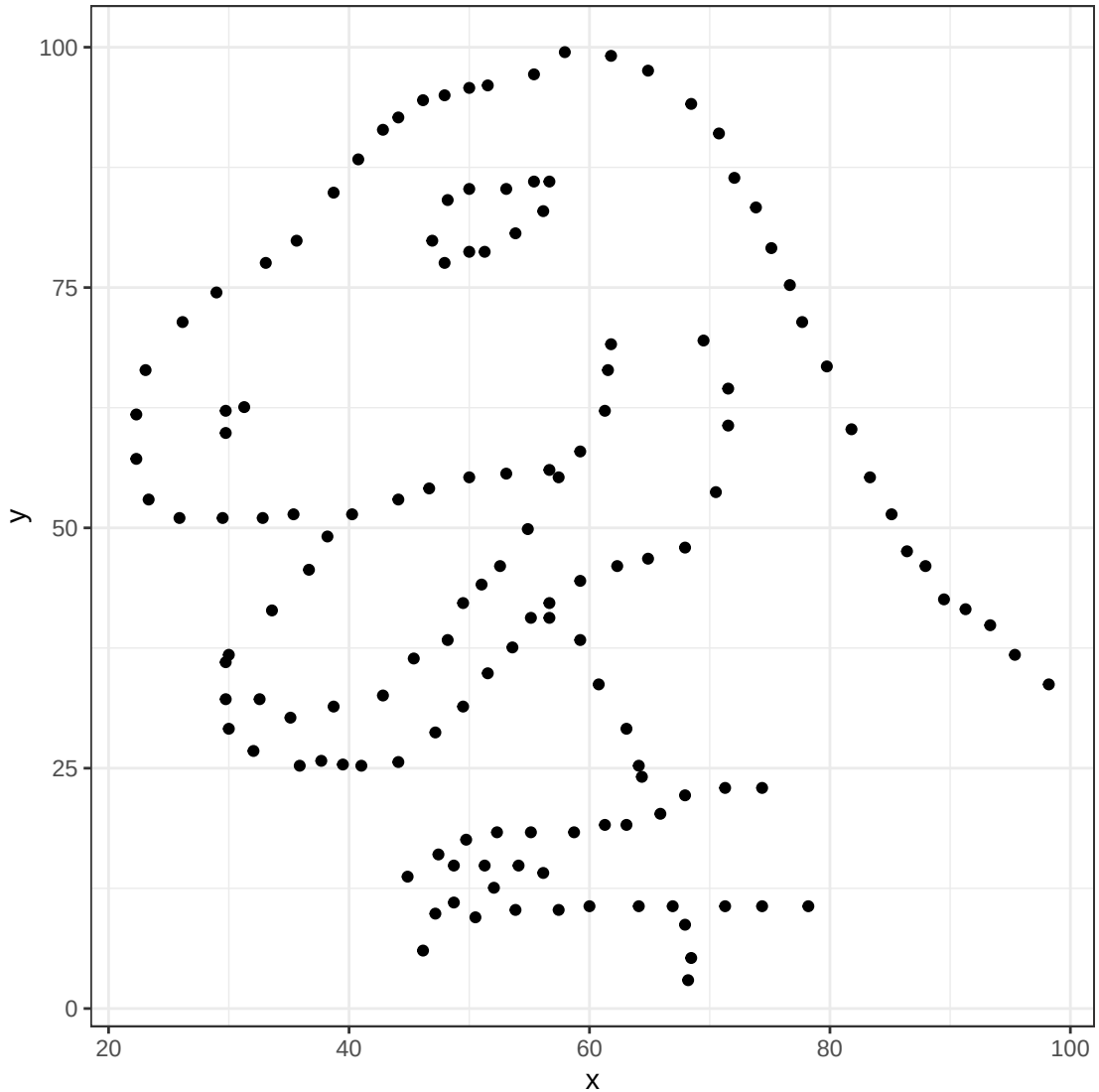
$$\text{Basiszahl} = \frac{\text{Bearbeitungszeiten}}{\textit{Verfahrensmenge}}$$

Verteilungen

- Unsere Informationen über die Realität sind fast immer mit Unsicherheit behaftet
- Auch bei perfekter Informationslage ist die empirische Realität praktisch nie homogen, es gibt immer Varianz
- Klassische Kennzahlen wie der Durchschnitt (arithmetisches Mittel) alleine sind irreführend

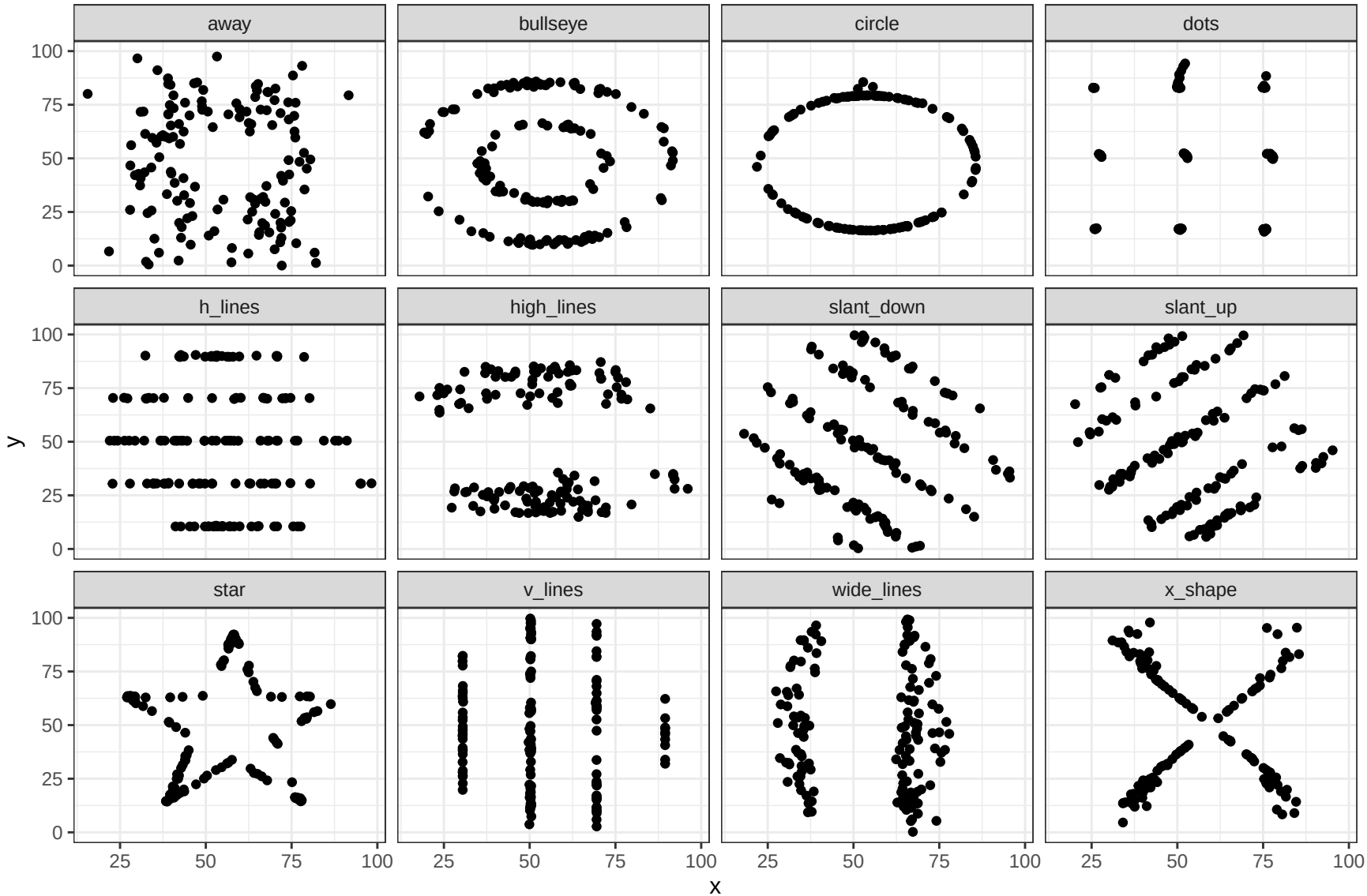
$\Rightarrow$ Die Analyse von Verteilungen bietet uns eine Möglichkeit, Unsicherheit und Varianz zu verstehen

Der Datasaurus



Quelle: Rhian Davies u. a., »datasauRus: Datasets from the Datasaurus Dozen« (*CRAN*, 4. Mai 2022) ⟨https://cran.r-project.org/web/packages/datasauRus/⟩. Der originale Datasaurus ist von Alberto Cairo, die konkrete Berechnungsmethode von Matejka und Fitzmaurice (2017).

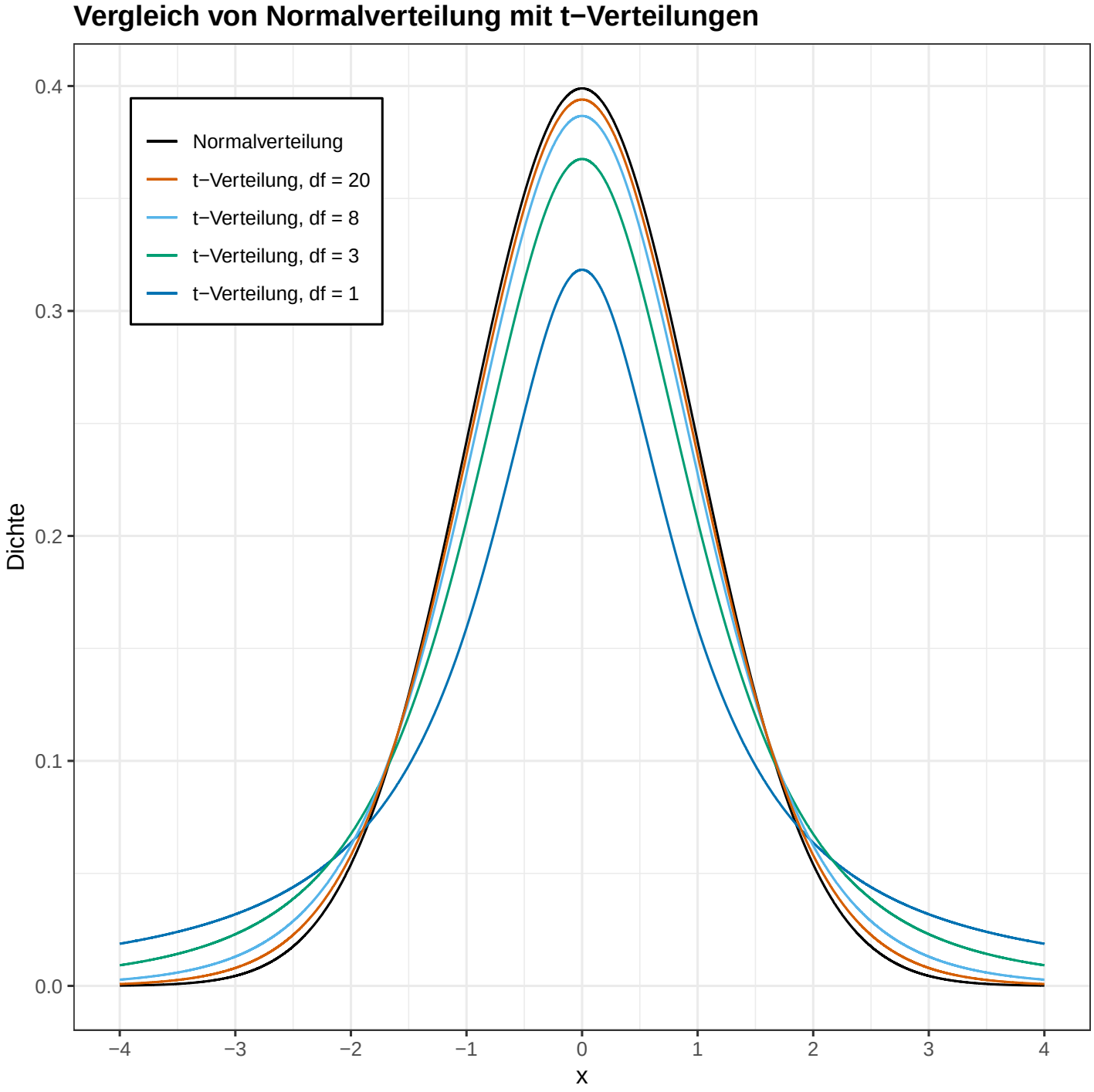
Datasaurus Dozen

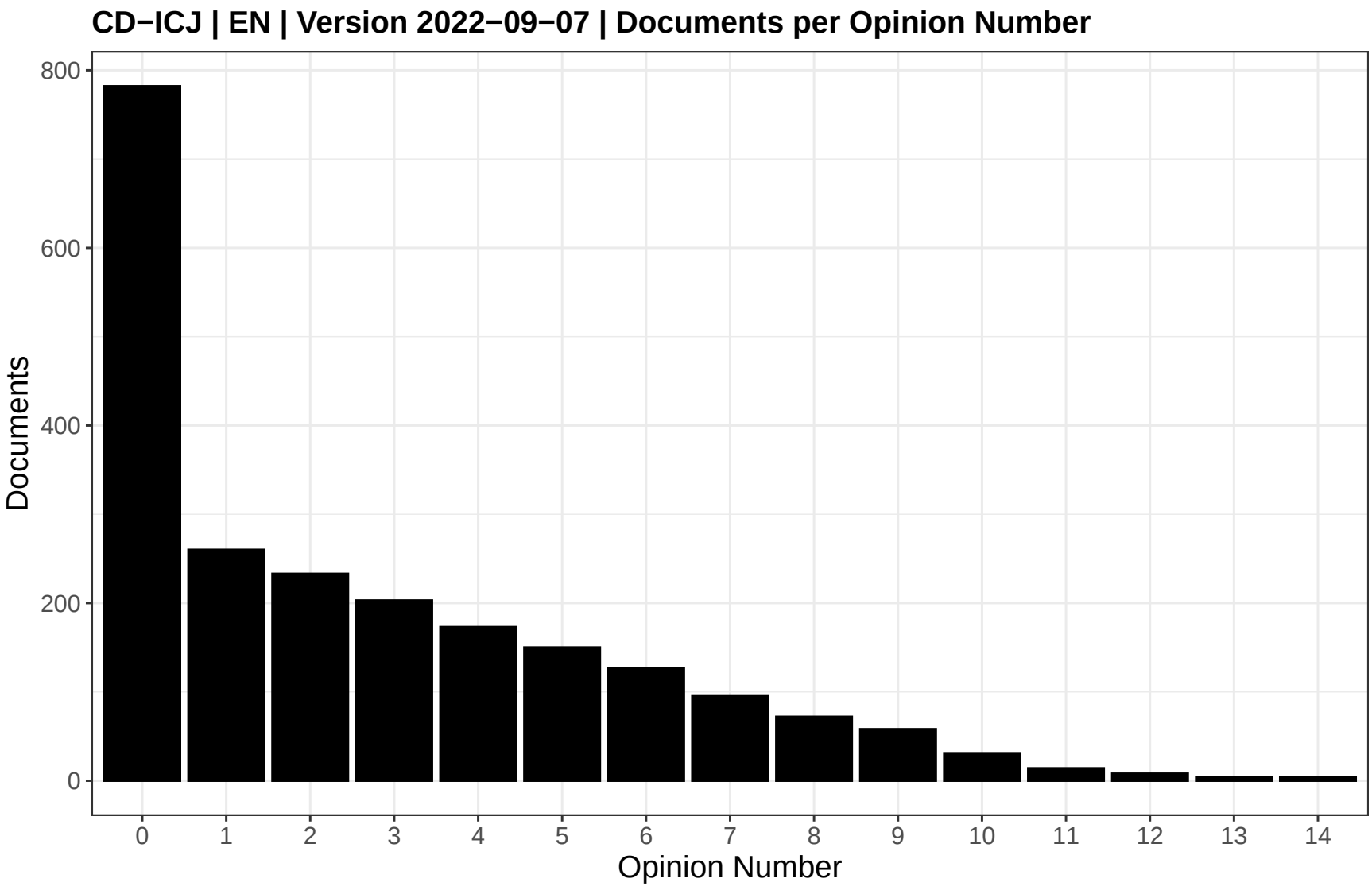


Quelle: Rhian Davies u. a., »datasauRus: Datasets from the Datasaurus Dozen« (*CRAN*, 4. Mai 2022) ⟨https://cran.r-project.org/web/packages/datasauRus/⟩. Die konkrete Berechnungsmethode ist von Matejka und Fitzmaurice (2017).

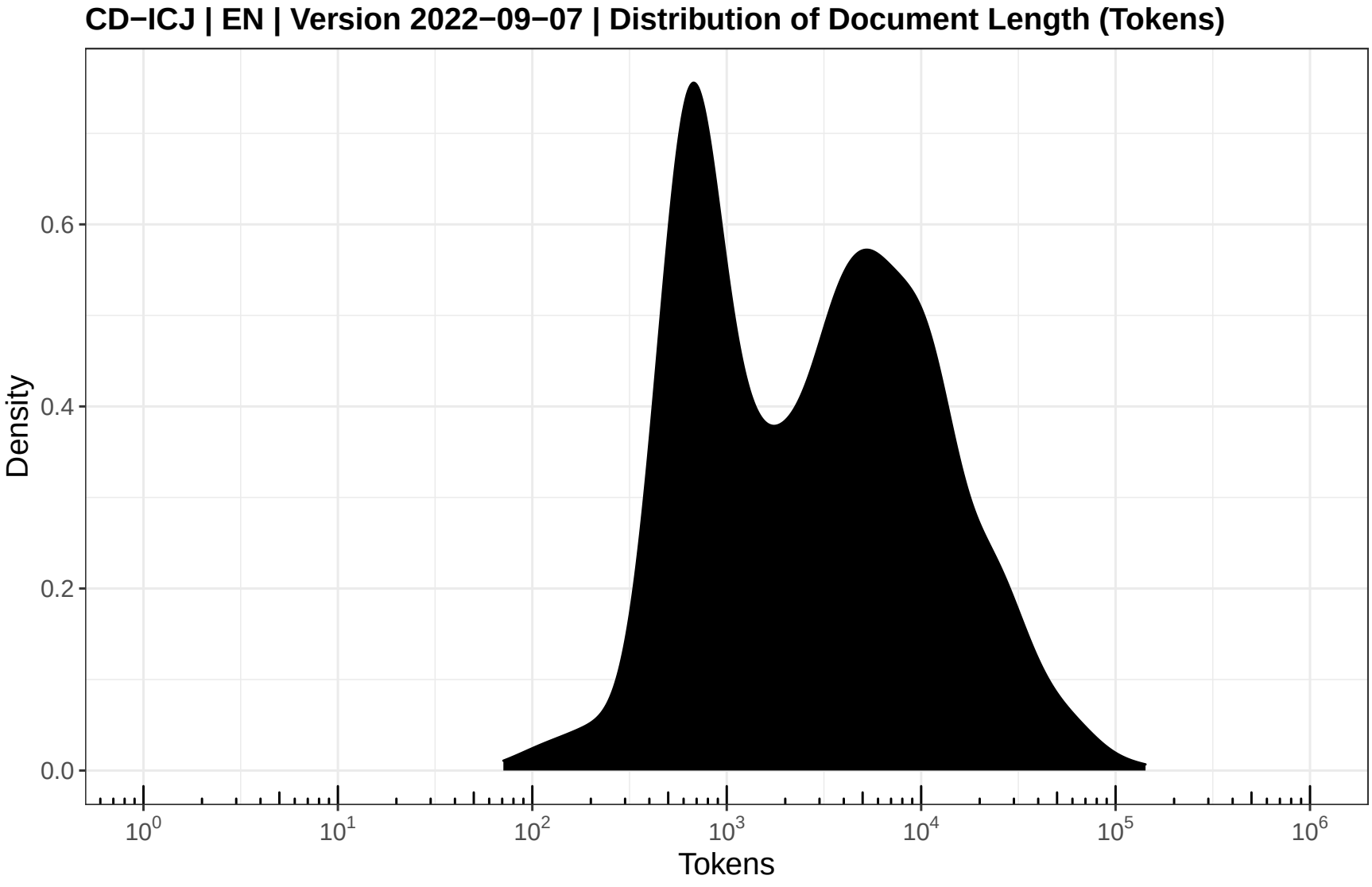
Datasaurus Dozen

dataset	x_mean	x_sd	y_mean	y_sd	corr
dino	54.26327	16.76514	47.83225	26.93540	-0.0644719
away	54.26610	16.76983	47.83472	26.93974	-0.0641284
h_lines	54.26144	16.76590	47.83025	26.93988	-0.0617148
v_lines	54.26993	16.76996	47.83699	26.93768	-0.0694456
x_shape	54.26015	16.76996	47.83972	26.93000	-0.0655833
star	54.26734	16.76896	47.83955	26.93027	-0.0629611
high_lines	54.26881	16.76670	47.83545	26.94000	-0.0685042
dots	54.26030	16.76774	47.83983	26.93019	-0.0603414
circle	54.26732	16.76001	47.83772	26.93004	-0.0683434
bullseye	54.26873	16.76924	47.83082	26.93573	-0.0685864
slant_up	54.26588	16.76885	47.83150	26.93861	-0.0686092
slant_down	54.26785	16.76676	47.83590	26.93610	-0.0689797
wide_lines	54.26692	16.77000	47.83160	26.93790	-0.0665752

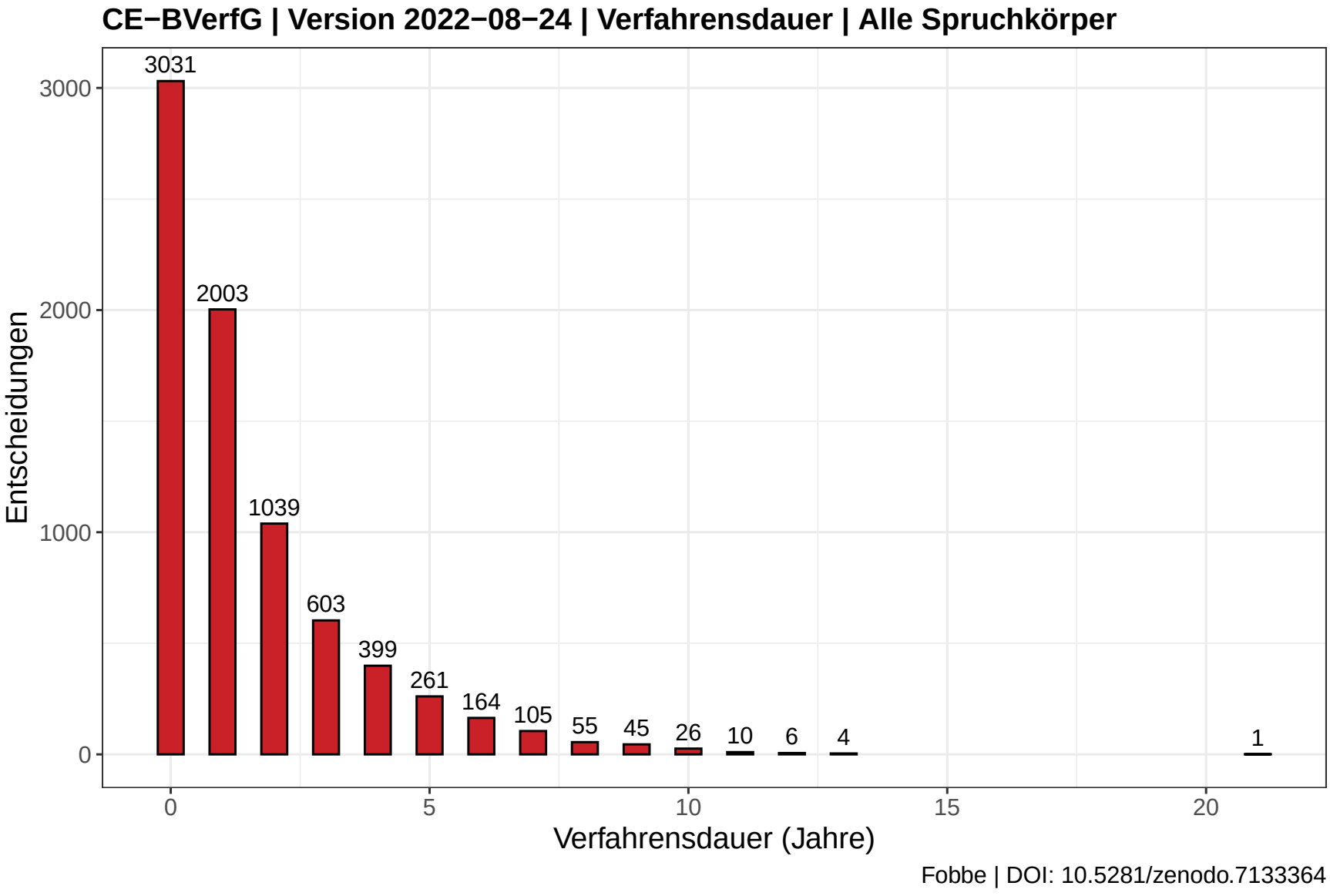

Vergleich von Normalverteilung mit t−Verteilungen
Normalverteilung
t−Verteilung, df = 20
t−Verteilung, df = 8
t−Verteilung, df = 3
t−Verteilung, df = 1
Dichte
0.0
0.1
0.2
0.3
0.4
−4
−3
−2
−1
0
1
2
3
4
x



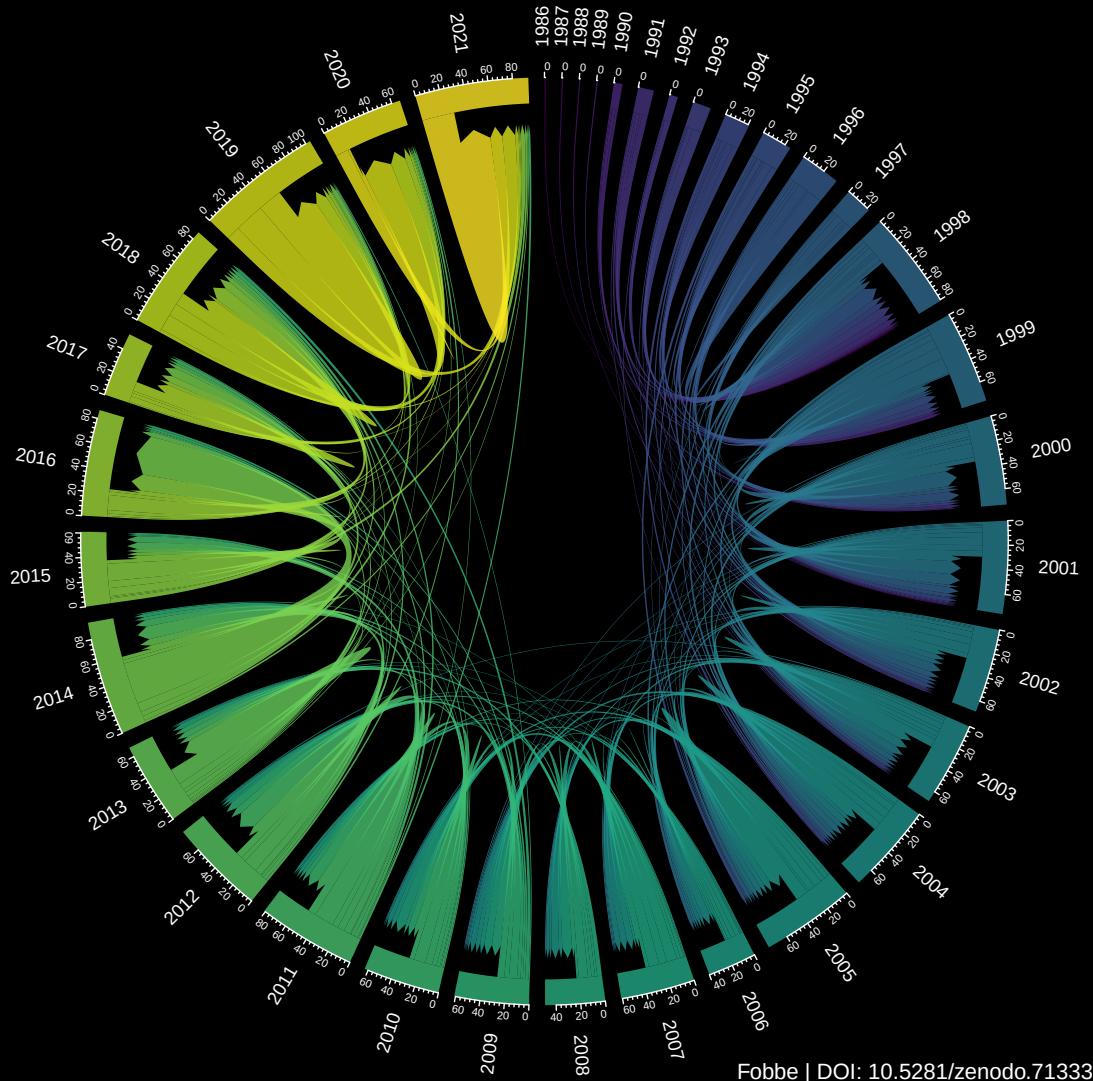
DOI: 10.5281/zenodo.7051929



DOI: 10.5281/zenodo.7051929

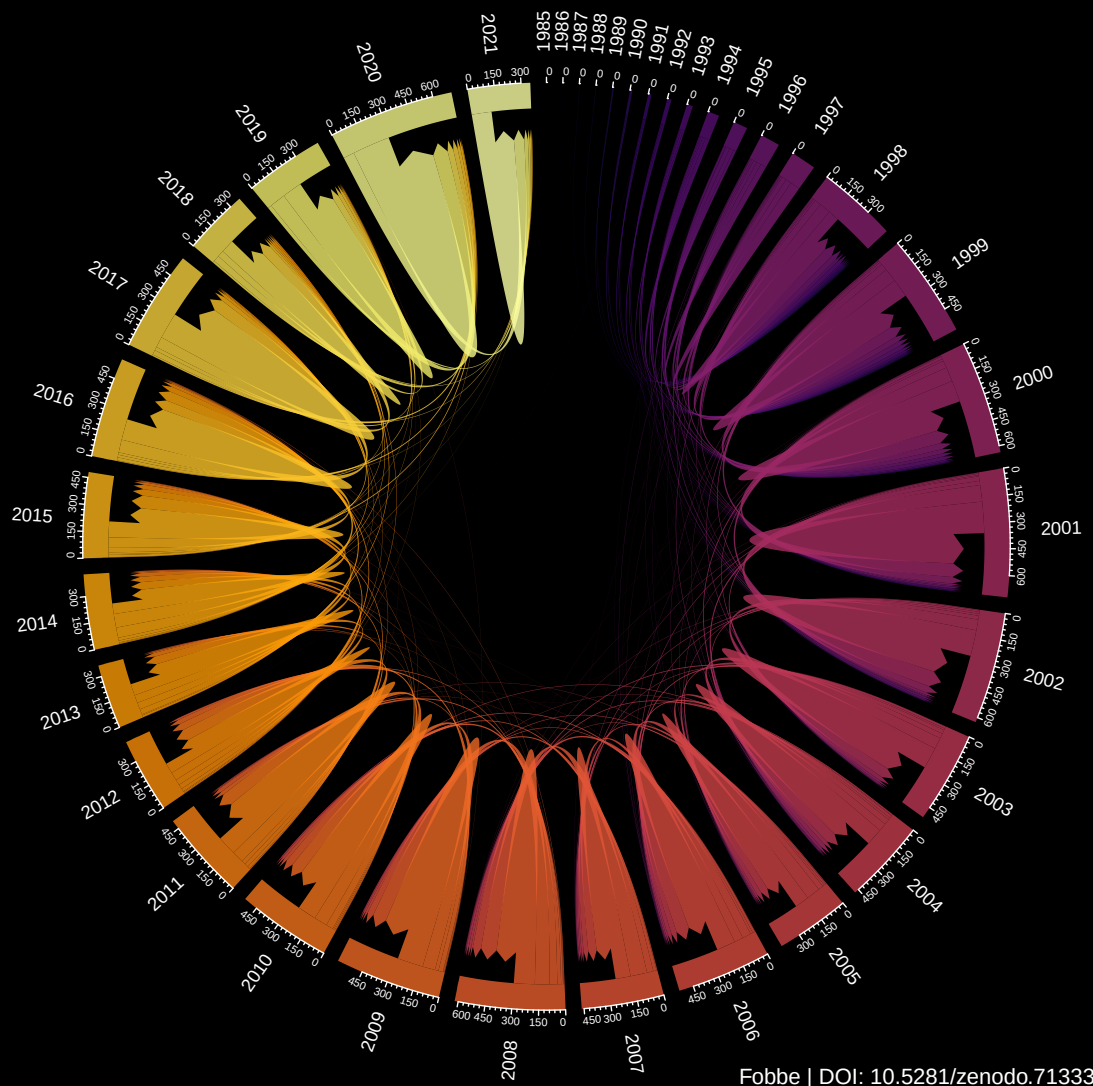

CE−BVerfG | Version 2022−08−24 | Verfahrensdauer | Alle Spruchkörper
Entscheidungen
3000
2000
1000
0
3031
2003
1039
603
399
261
164
105
55
45
26
10
6
4
1
0
5
10
15
20
Verfahrensdauer (Jahre)
Fobbe | DOI: 10.5281/zenodo.7133364

CE−BVerfG | Version 2022−08−24 | Verfahrensdauer 1998−2021 | Senate



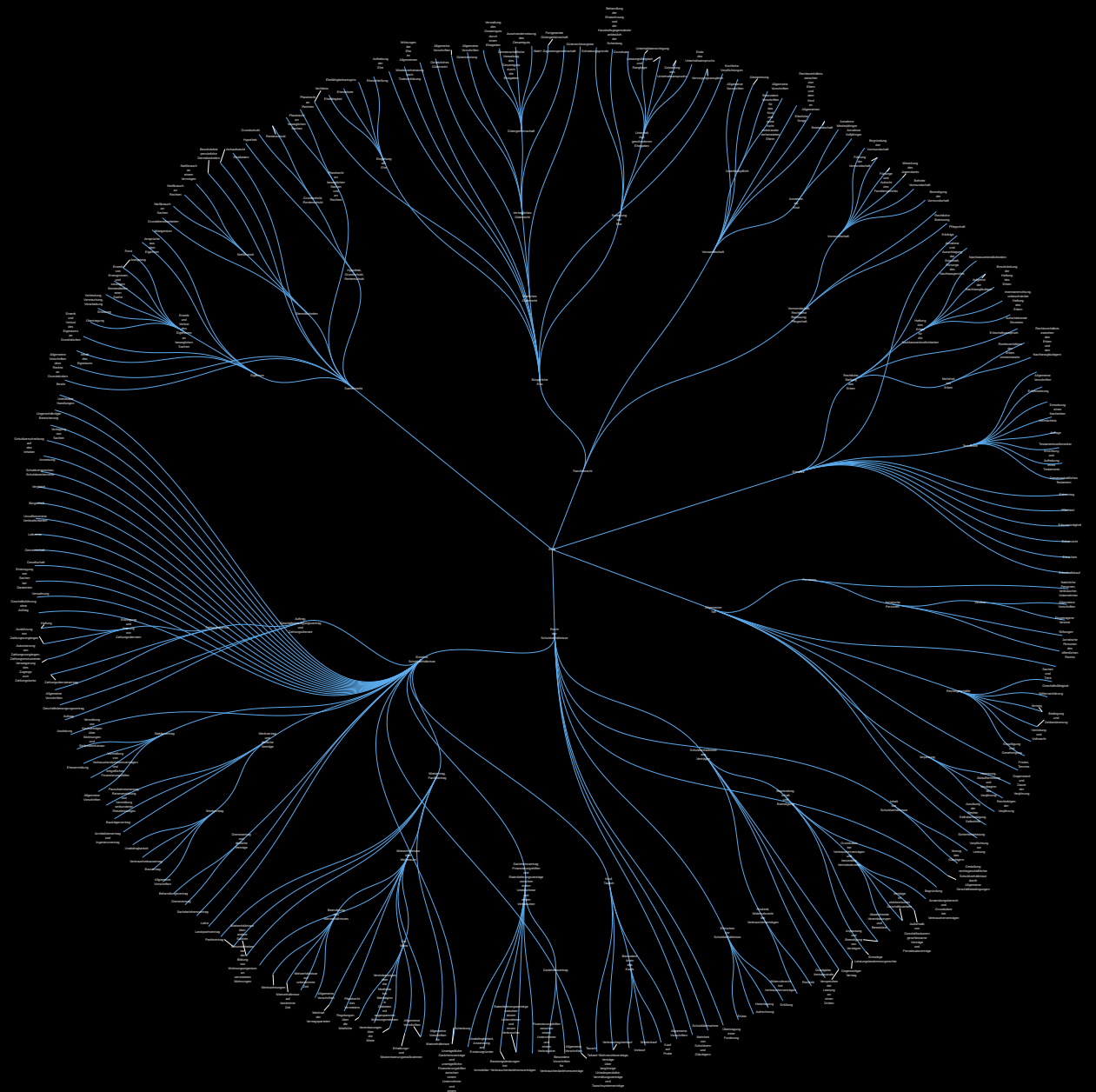
Fobbe | DOI: 10.5281/zenodo.7133364

CE−BVerfG 2022−08−24 | Verfahrensdauer 1998−2021 | Kammern



Fobbe | DOI: 10.5281/zenodo.7133364

C−DBR | Version 2021−09−16 | Struktur des BGB



DOI: 10.5281/zenodo.5468668 | Fobbe

Teil IV

Theorie: Repräsentativität

Grundgesamtheit und Vollerhebung

Grundgesamtheit

Alle statistischen Einheiten von Interesse in einer Untersuchung. Beispielsweise alle Deutschen, alle Krebskranken, alle Arbeitnehmer:innen, alle Produkte einer Fabrik.

Vollerhebung

Die vollständige Untersuchung der Grundgesamtheit ist meist zu teuer, unethisch (z.B. Medikamentenstudien) oder schlicht unmöglich (nicht alle Mitglieder der Grundgesamtheit sind bekannt).

Stichprobe und Teilerhebung

Stichprobe

Eine Teilmenge der Grundgesamtheit. Beispielsweise 1.000 Einwohner:innen Deutschlands in einer Wahlumfrage. Stichproben können zufällig, systematisch oder willkürlich gezogen werden.

Teilerhebung

Die Untersuchung einer Stichprobe. Eine repräsentative Stichprobe lässt mit einer gewissen Unsicherheit eine Hochrechnung auf die Grundgesamtheit zu.

Repräsentativität und die Bundesrechtsanwaltskammer

»Die Umfrageergebnisse zeichnen aufgrund der Durchmischung der [über 5.000] Teilnehmer (vom Einzelanwalt bis zum Partner in der Großkanzlei) ein repräsentatives Bild der aktuellen Situation der Anwaltschaft.«[1]

»Rechtsanwältin Ulrike Paul, Vizepräsidentin der BRAK, zeigt sich über die Ergebnisse besorgt: ›Unter Berücksichtigung der Teilnehmerzahl [8.125] können wir die Umfrage als repräsentativ bezeichnen und davon ausgehen, dass bereits jetzt über ein Fünftel der deutschen Anwaltschaft persönlich betroffen ist.‹«[2]

»Durch die erstmalig rein digital durchgeführte Umfrage konnte die Teilnahme niederschwellig angeboten werden, so dass von 165.587 Anwältinnen und Anwälten 4.757 an der Befragung teilnahmen, also 2,87%. Der Rücklauf war dank der großen Unterstützung der Kammern insgesamt repräsentativ.«[3]

---

[1] https://www.brak.de/anwaltschaft/tipps-und-leitfaeden/corona/#c8717

[2] https://www.brak.de/presse/presseerklaerungen/2022/ein-fuenftel-aller-anwaltlichen-anderkonten-gekuendigt/

[3] https://www.brak-mitteilungen.de/media/BRAK_2023_Heft01_komprimiert.pdf — Seite 3

Repräsentativität

- Echte Repräsentativität können nur echte Zufallsstichproben gewährleisten
- Selbst Zufallsstichproben haben einen Schätzfehler abhängig von ihrer Stichprobengröße
- Echte Zufallsstichproben sind schwierig, daher oft Clusterbildung oder Stratifizierung
- Gewichtung, Quotierung oder statistische Kontrolle von Merkmalen kann nur bei Kenntnis aller für das Ergebnis relevanten Merkmale zu einer Repräsentativität führen

Repräsentativität und PEBB§Y

- PEBB§Y soll eine Vollerhebung sein, ist tatsächlich aber eine Stichprobe
- 70 Erhebungsdienststellen (PEBB§Y 2014); aber: 777 Gerichte der ordentlichen Gerichtsbarkeit in Deutschland (2020)
- Auswahl der Erhebungsdienststellen nach Königsteiner Schlüssel, Ministerialentscheidung, Freiwilligkeit, Quotierung

Keine Repräsentativität

Teil V

Praxis zum Mitmachen

Praxis: Arbeit mit der Programmiersprache »R«

»R« und »Python« sind die dominanten Sprachen im Data Science-Bereich

- Wir arbeiten heute mit R, genauer mit der WebR-Variante für Browser
- Rufen Sie bitte https://webr.r-wasm.org/latest/ auf
- Herunterladen dauert 1–2 Minuten
- Wird lokal auf ihrem Computer in einer Sandbox ausgeführt

Diagramme

```r
#  Histogramm: Haeufigkeitsverteilung mit Stufen
hist(USJudgeRatings$INTG)

#  Density Chart: Haeufigkeitsverteilung mit Glaettung
plot(density(USJudgeRatings$INTG))

# Boxplot
# - Median (Linie in der Mitte)
# - mittlere Haelfte der Werte (Box)
# - 1,5 mal die Spannweite der Box (Whiskers)
# - Ausreisser (einzelne Punkte) an
boxplot(USJudgeRatings)
```

Beschreibende Statistik

```
# Datensatz mit Bewertungen von US-Richter:innen
print(USJudgeRatings)

# Darstellung der Struktur eines Datensatzes
str(USJudgeRatings)

# Berechnung von Minimum, Maximum, Mittelwert, Median
und Quartilen
summary(USJudgeRatings)
```

Kontakt

Haben Sie Fragen?

---

www.seanfobbe.de