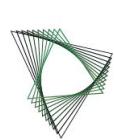


Modul 2: Umsetzung - Implementierung der Stichprobenziehung

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)
www.ada-oeffentliche-verwaltung.de



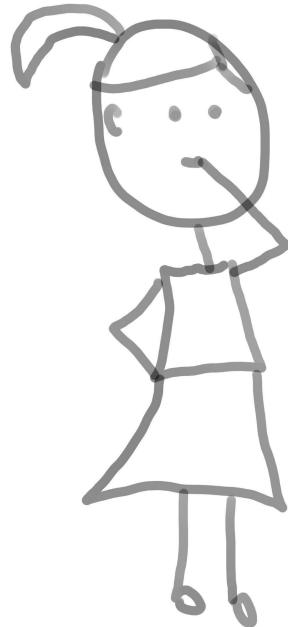
BERD
@NFDI



Bayerisches Staatsministerium
für Digitales

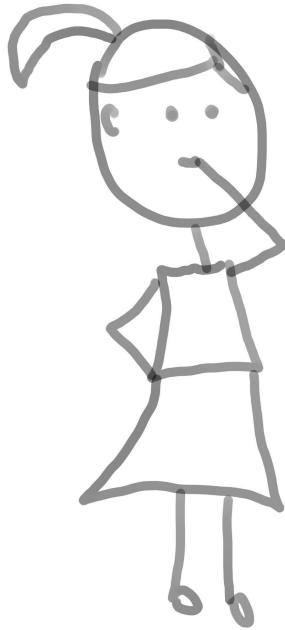


Rückblick



Was ist Ihnen von gestern besonders in Erinnerung geblieben?

Modul 2



Am Ende dieses Moduls können Sie...

- ... eine Entscheidung über eine gute Strategie für die Stichprobenziehung fällen.
- ... die Stichprobe analysieren und visuell darstellen.
- ... Schwächen der Strategie einschätzen und verteidigen.

Einführung	10:00 - 10:15
Überblick+Recap: Stichprobenziehung	10:15 - 10:45
Pause	10:45 - 11:00
Besprechung der Strategie	11:00 - 12:00
Mittagspause	12:00 - 13:00
Umsetzung: Prototyp Flowchart entwickeln	13:00 - 14:00
Pause	14:00 - 14:20
Umsetzung: Prototyp Notebook entwickeln	14:20 - 15:15
Wrap-up und Ausblick	15:15 - 15:30

Einführung

Überblick zur Stichprobenziehung

Im Zweifel Zufall

Terrorist Detektor: 99.9% korrekt

Terrorist klassifiziert als harmloser Passagier: 0.001

Harmloser Passagier klassifiziert als Terrorist: 0.001

Annahme: 1 Person in 1 Millionen ist ein Terrorist (0.000001).

Das heißt bei 999 von 1.000 Leuten wird der Detektor falschen Alarm schlagen.

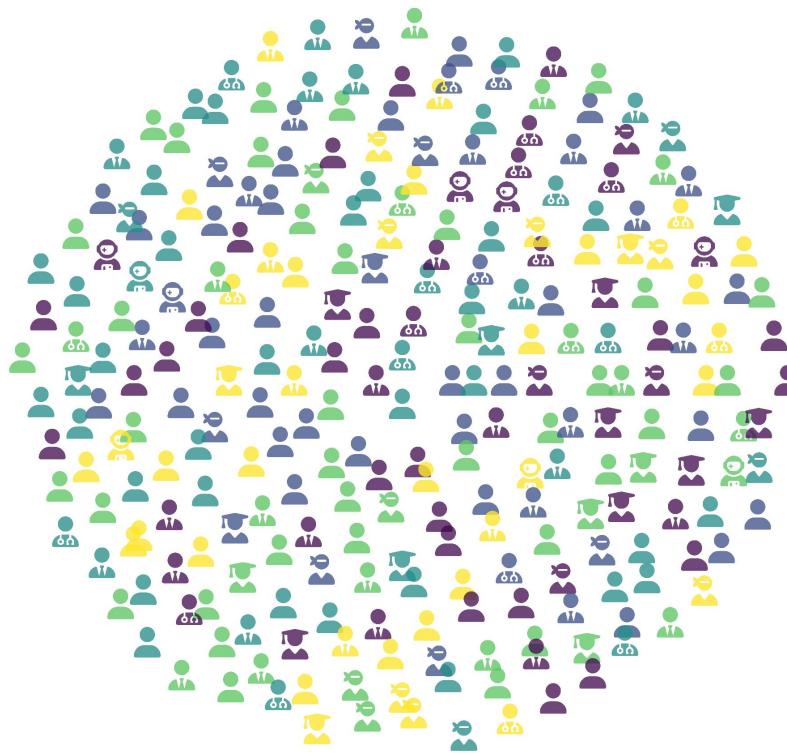


Wir kennen das Problem auch aus der Corona Zeit ...

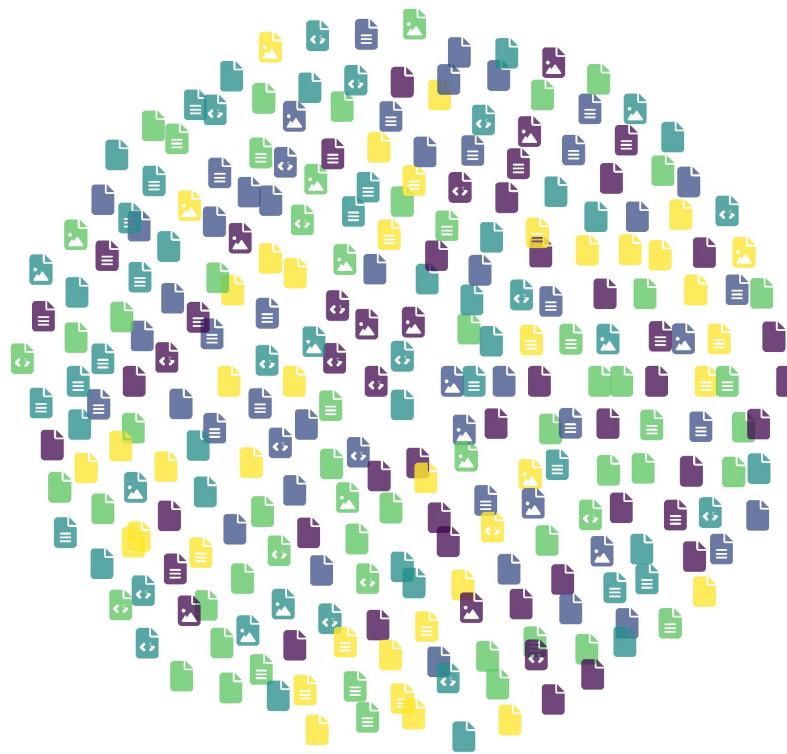


Die Grundgesamtheit ist die Menge aller Personen...

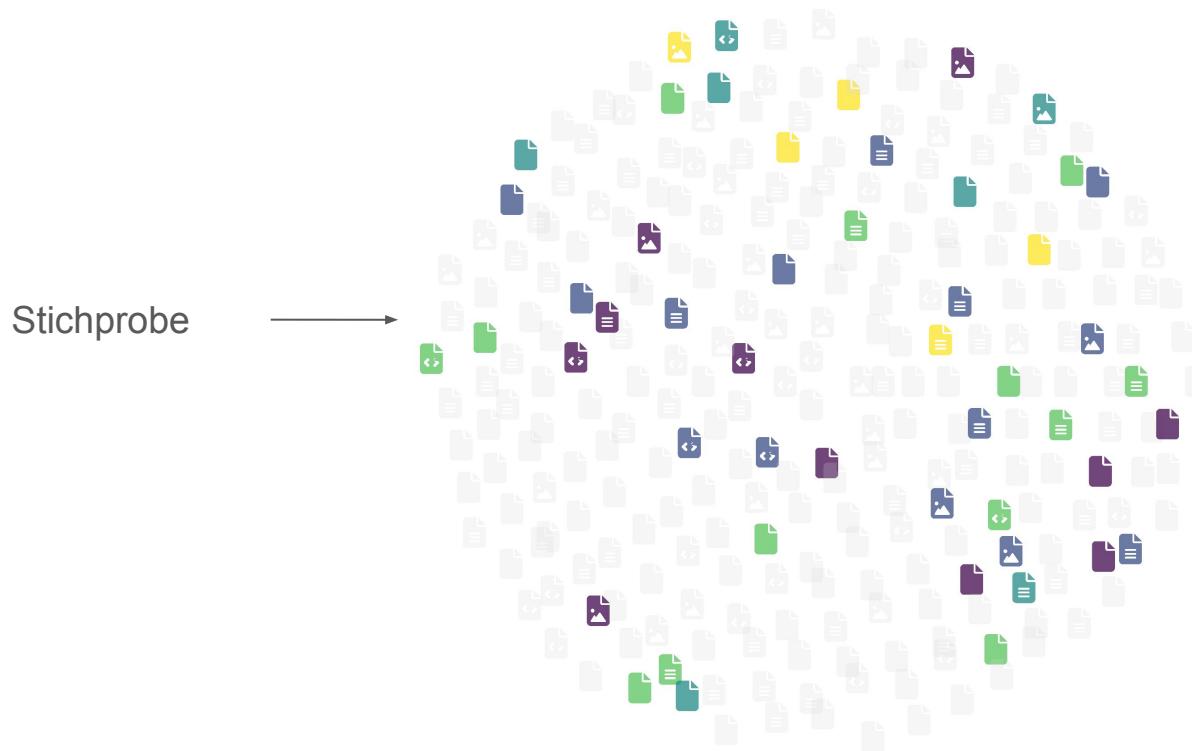
Grundgesamtheit →



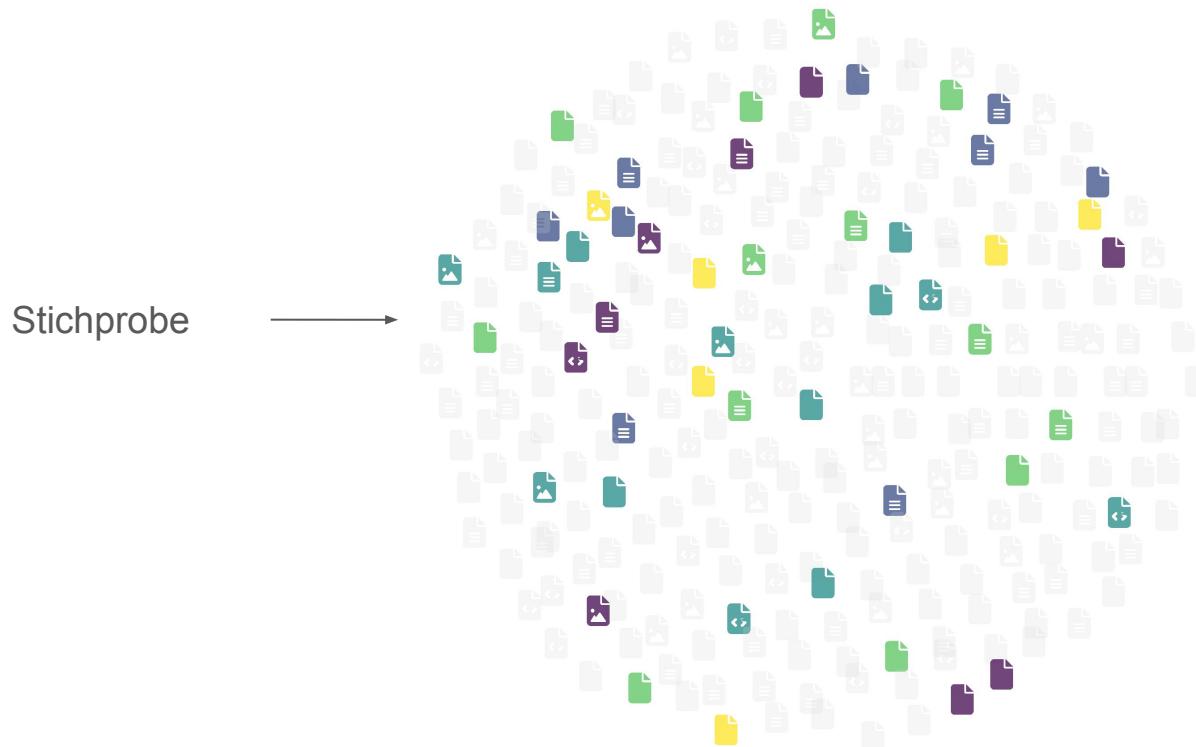
... oder Akten über die wir eine Aussage treffen wollen



Eine Stichprobe ist eine Teilmenge der Grundgesamtheit



Zufallsstichproben unterscheiden sich bei Wiederholung



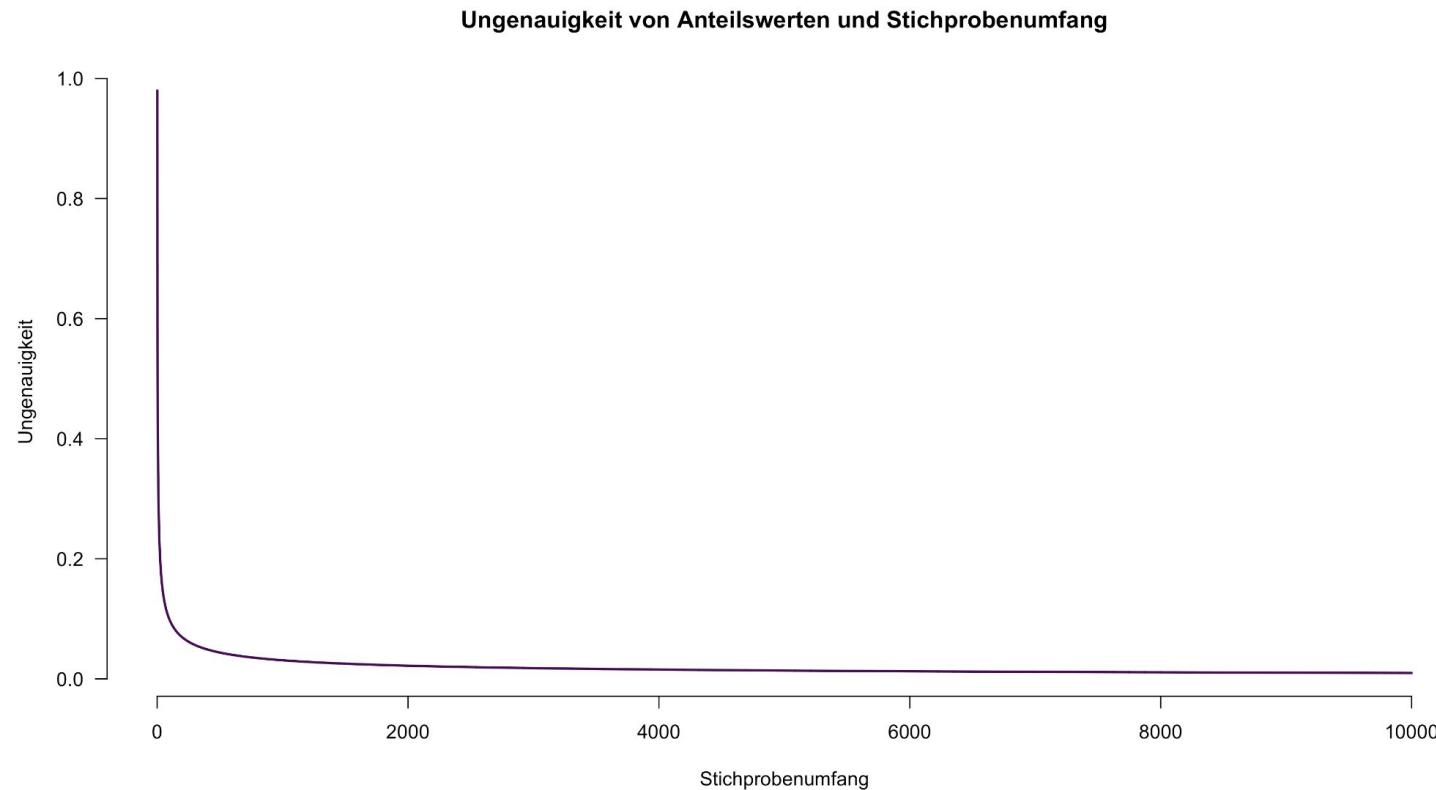
Wir nutzen die Stichprobe um Merkmale zu untersuchen

Erste Stichprobe

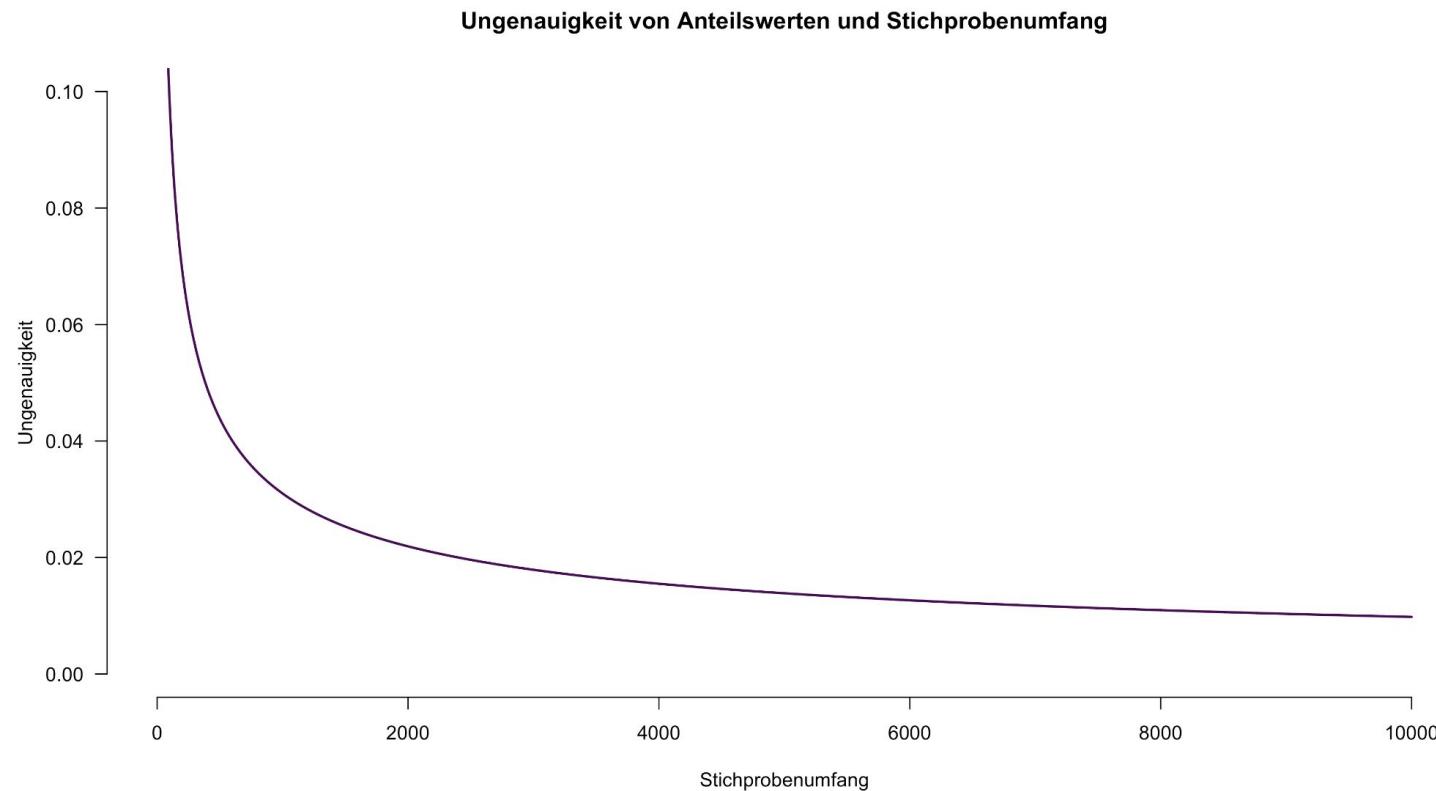


Zweite Stichprobe

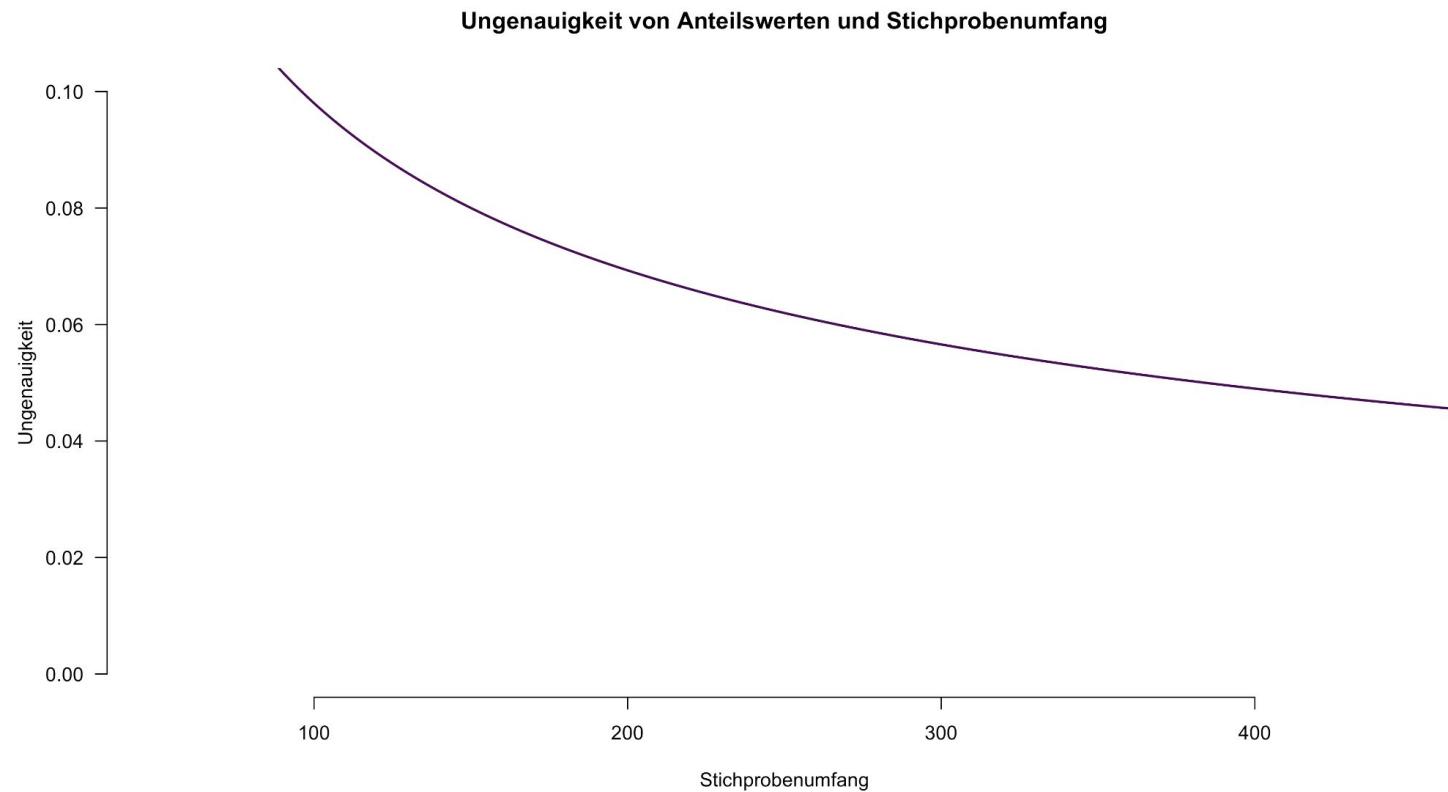
Wie genau können wir Anteilswerte bestimmen?



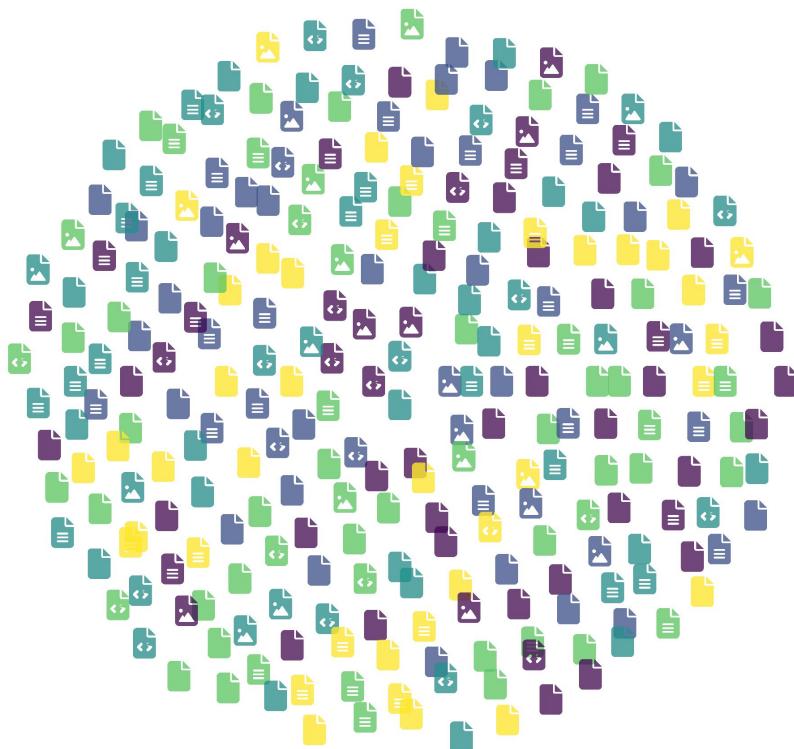
Wie genau können wir Anteilswerte bestimmen?



Wie genau können wir Anteilswerte bestimmen?



Wie können wir eine einfache Zufallsstichprobe ziehen?



Jede Akte soll die gleiche Chance haben!
(simple random sample, SRS)

Soll jede Akte die gleiche Chance haben?
(SRS with unequal probabilities)



Losverfahren
Urne

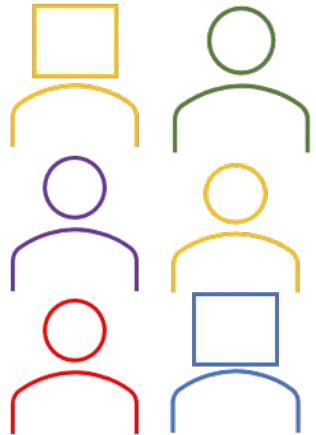
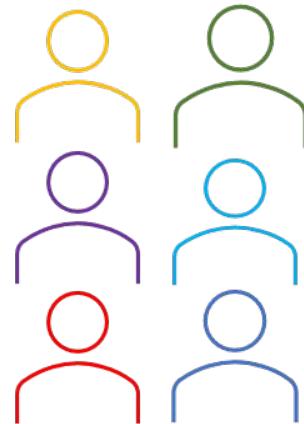
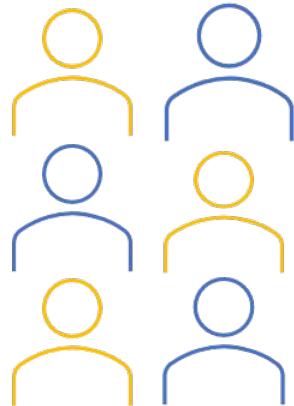
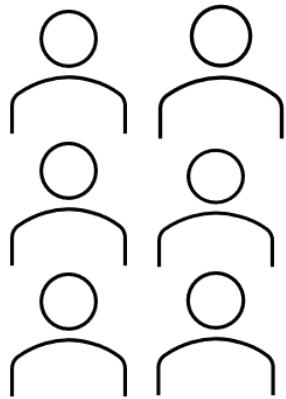


Systematische
Zufallsauswahl



Zufallszahlentabelle/
Generator

Wie groß soll unsere Stichprobe sein?



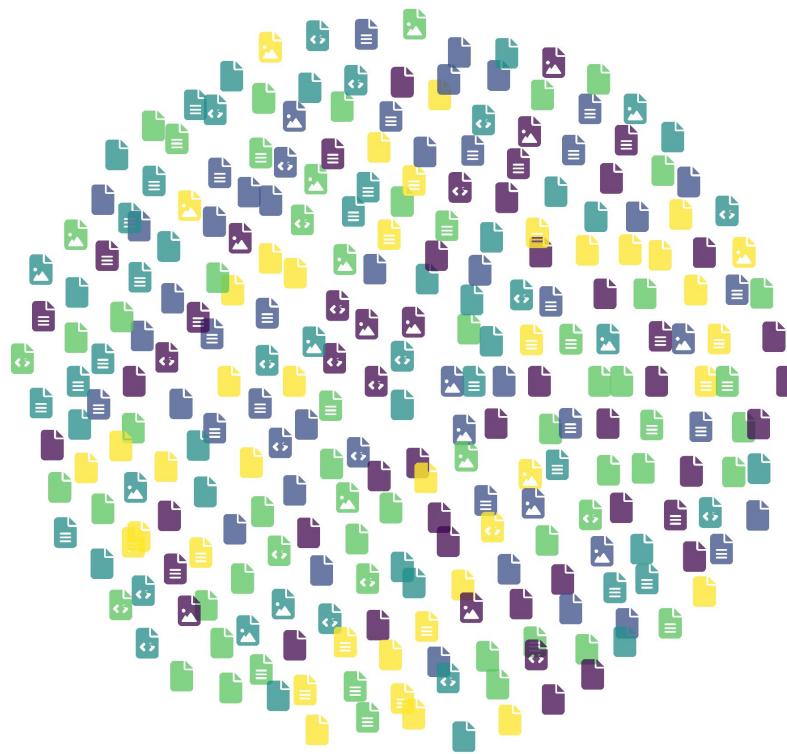
Vor- und Nachteile der einfachen Zufallsstichprobe

Alle Akten haben dieselbe
Auswahlwahrscheinlichkeit.

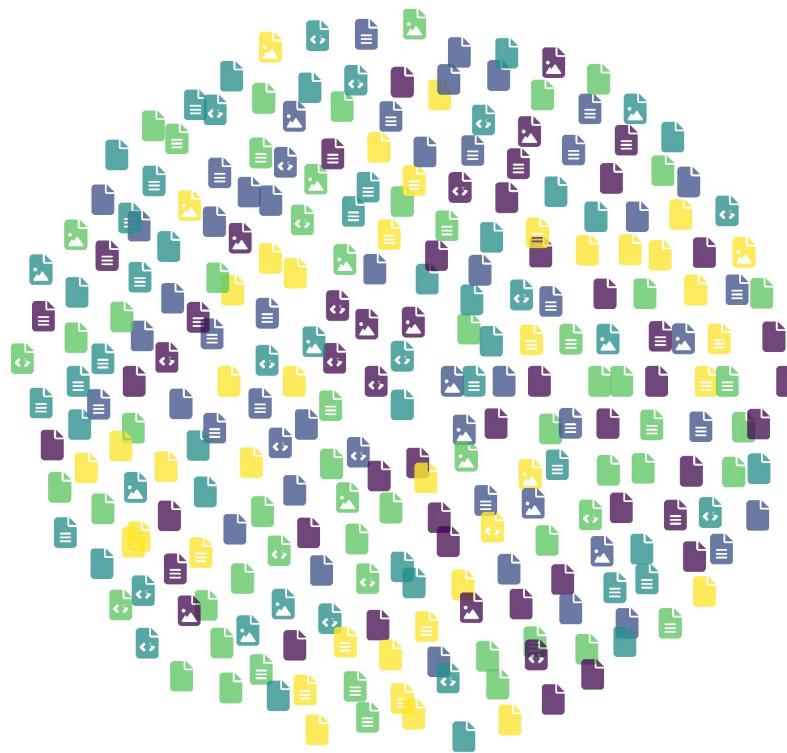
Alle Akten haben dieselbe
Auswahlwahrscheinlichkeit.

Ist eine einfache Zufallsstichprobe mit den
Kriterien der Aussonderungsbekanntmachung
Justiz kompatibel?

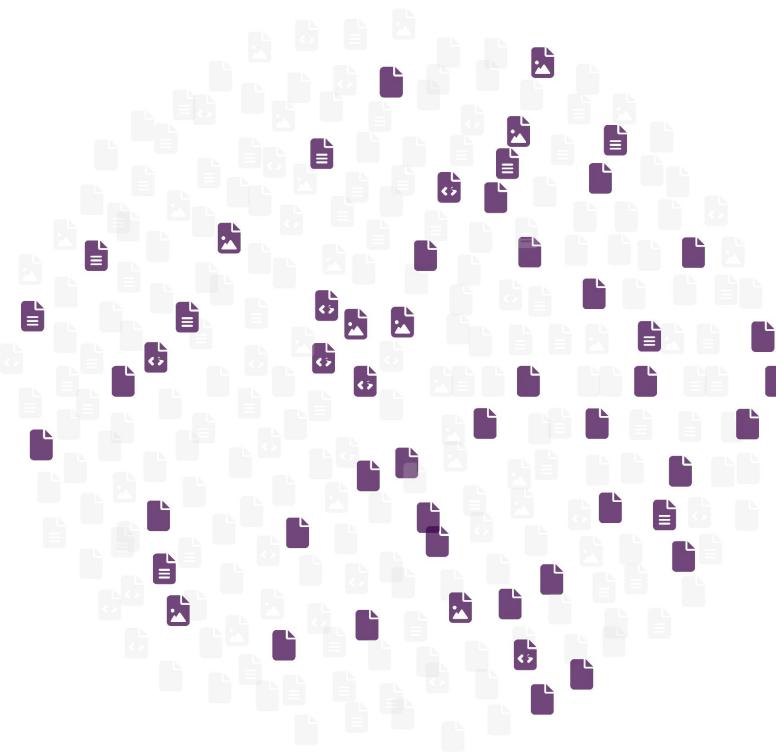
Mit geschichteten Zufallsstichproben können wir...



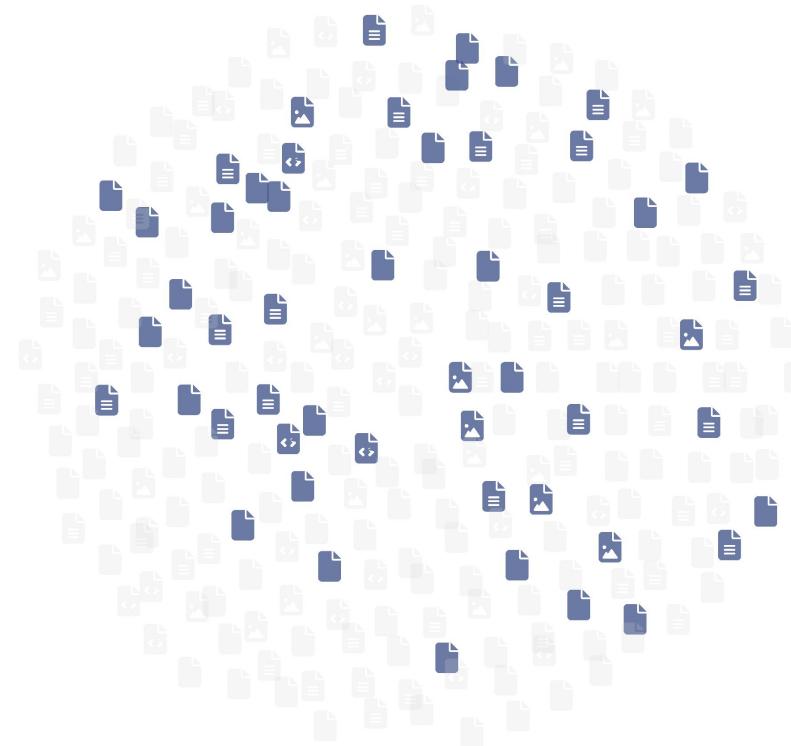
die Zufallsauswahl mit qualitativen Kriterien kombinieren



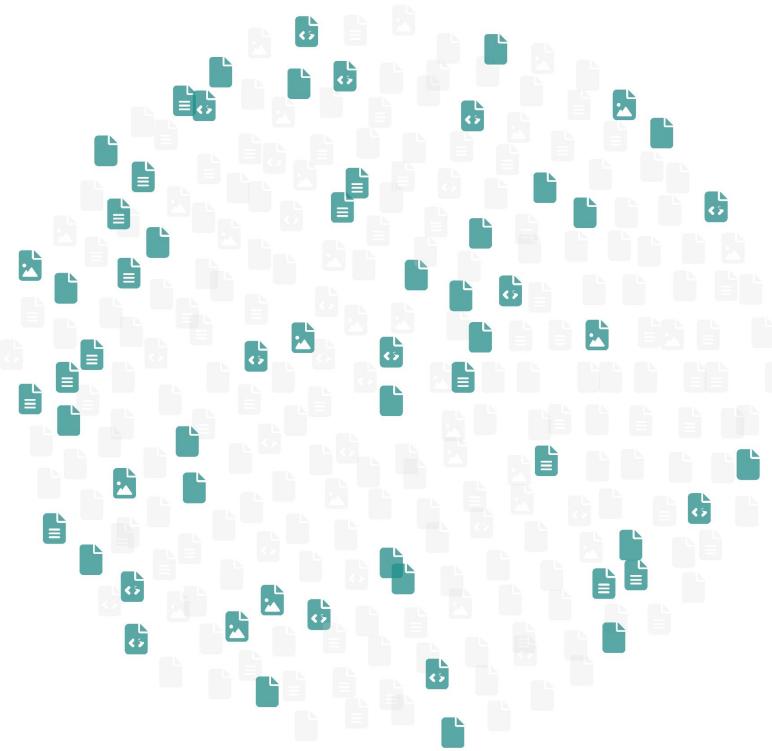
Unsere Grundgesamtheit geschichtet nach Farben



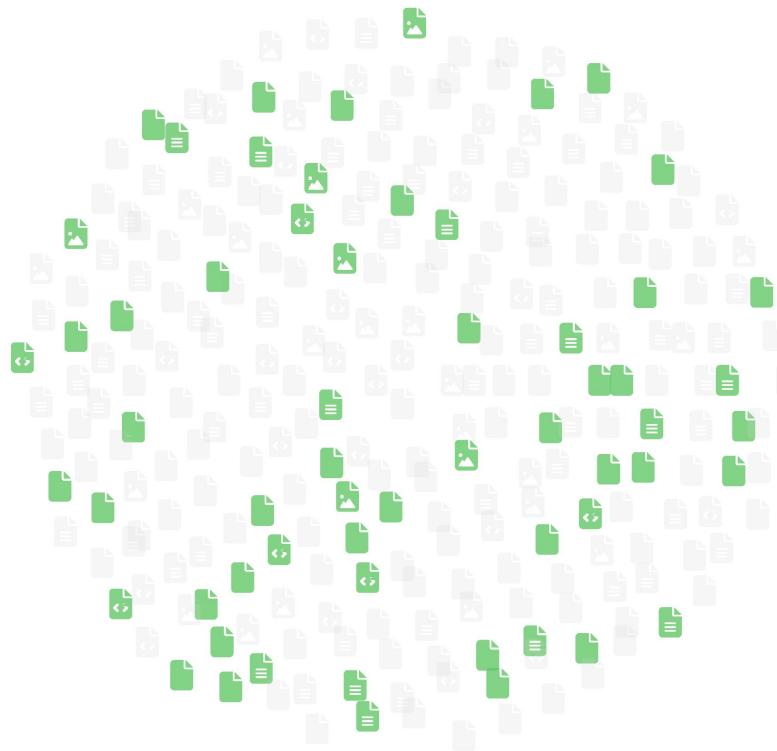
Unsere Grundgesamtheit geschichtet nach Farben



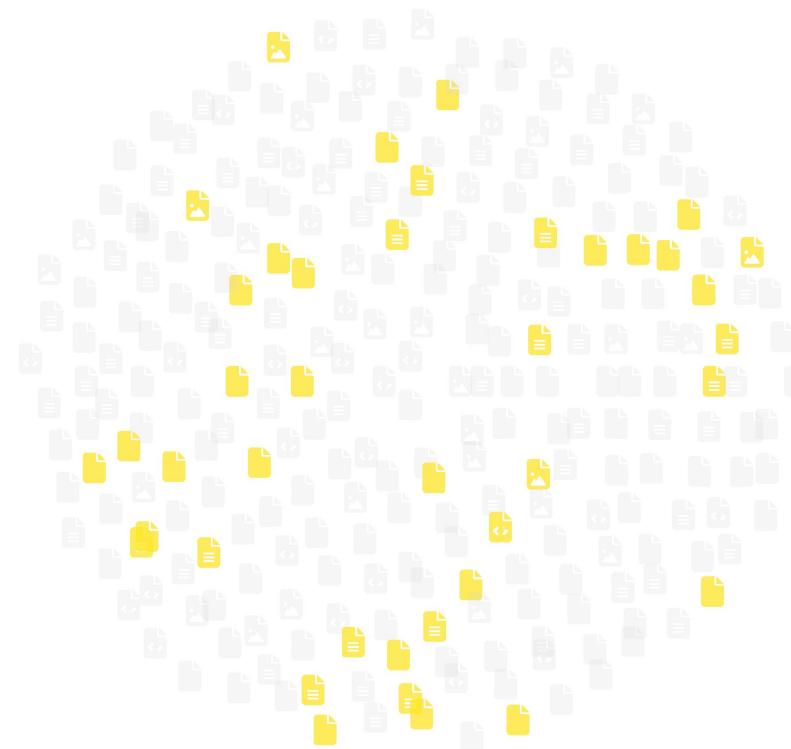
Unsere Grundgesamtheit geschichtet nach Farben



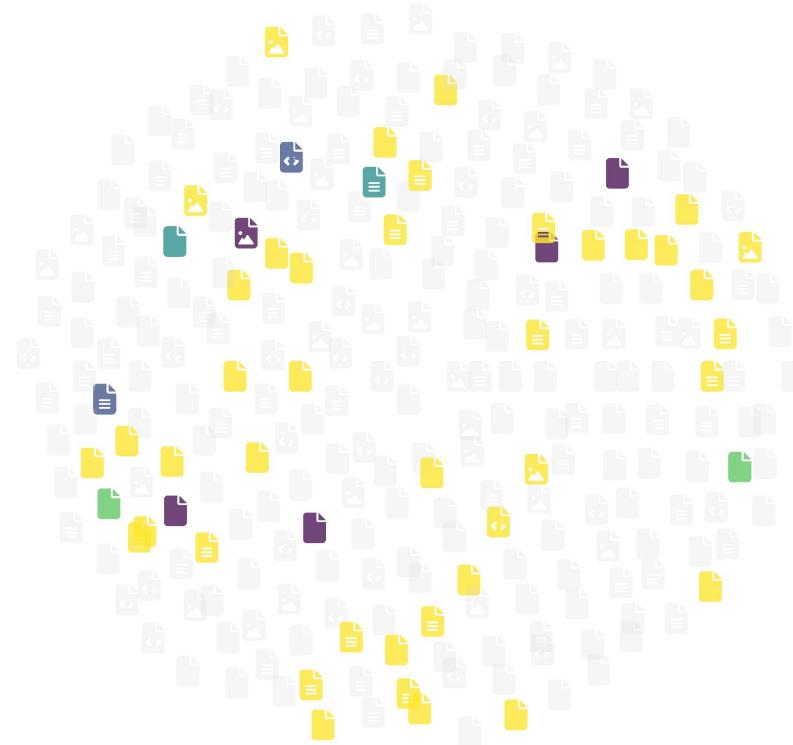
Unsere Grundgesamtheit geschichtet nach Farben



Unsere Grundgesamtheit geschichtet nach Farben

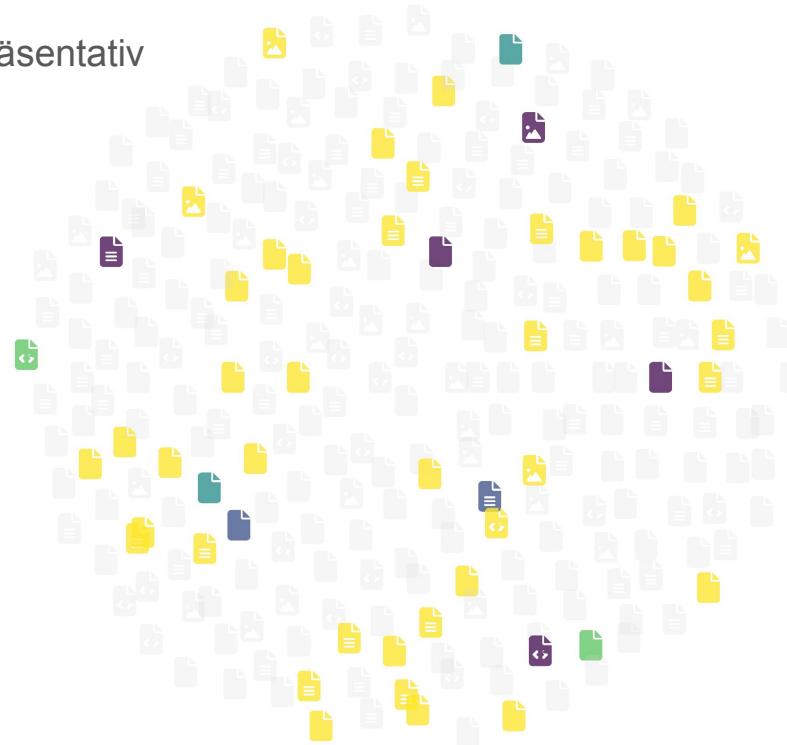


Wir wählen alle gelben Akten, fünf lila Akten und je zwei aus den anderen Schichten



Eine weitere geschichtete Zufallsstichprobe

Sind diese Stichproben repräsentativ
für die Grundgesamtheit?

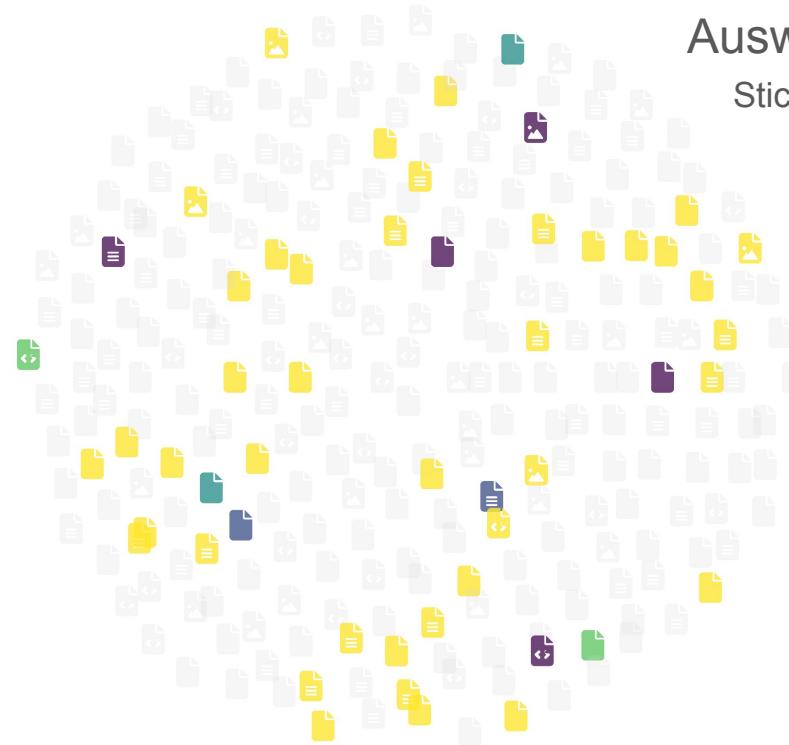


Bei Transparenz zur Auswahl können wir hochrechnen

Transparenz über:

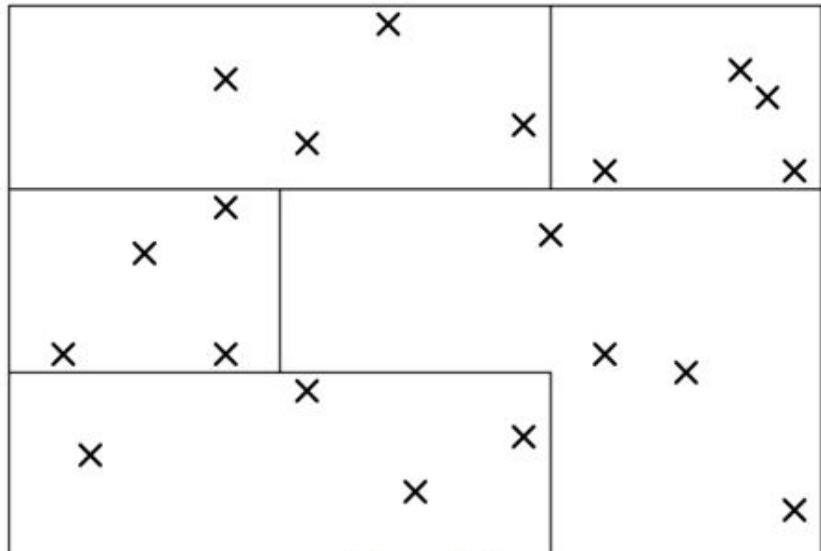
Größe der Schichten

Stichprobengröße pro
Schicht

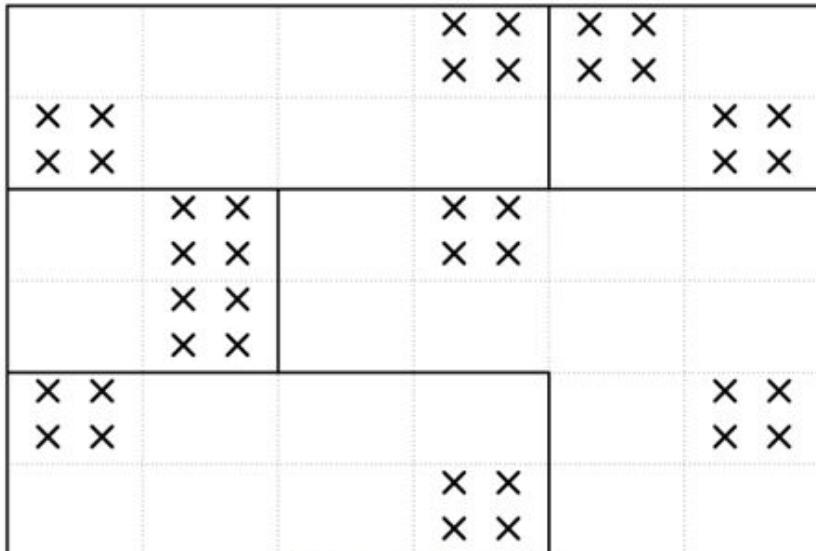


Auswahlwahrscheinlichkeit =
Stichprobengröße/Größe der Schicht

Hochrechnungsfaktor =
1/Auswahlwahrscheinlichkeit

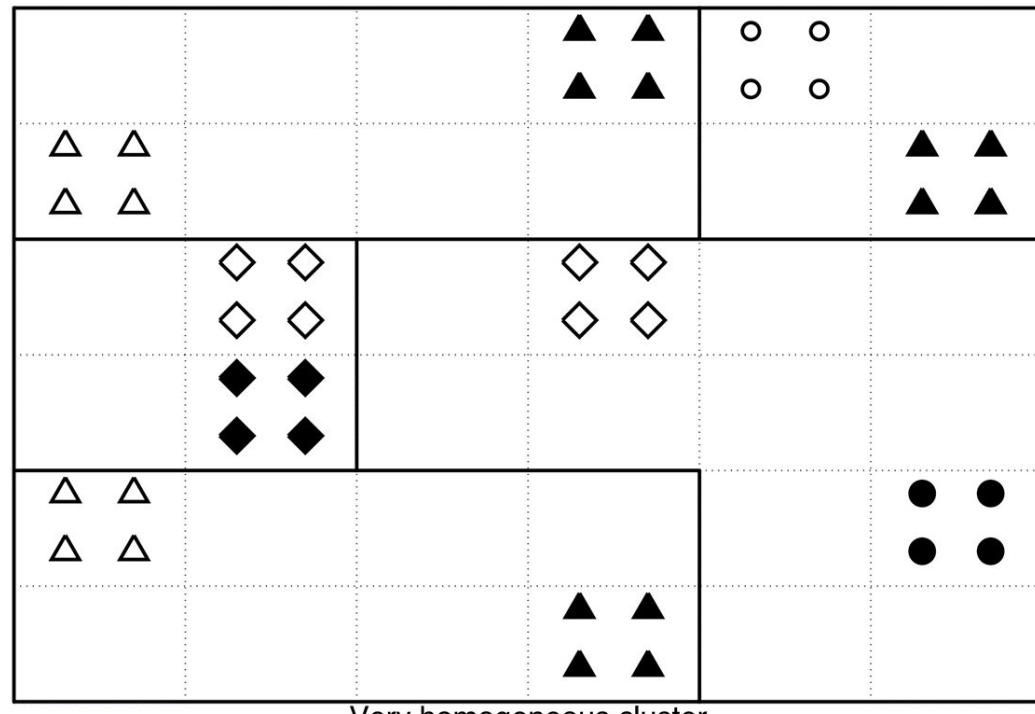


Stratified sample



Cluster sample within strata

Wie viele von wo?



Pause

Besprechung der Stichprobenstrategie

Experte: Prof. Helmut
Küchenhoff

- Welche Bedürfnisse gibt es?
- Was hat die GDA sich bereits überlegt?
(siehe Vortrag gestern + folgende Slides)
- Wie sollen wir es tatsächlich umsetzen?

Von GDA Vorgeschlagenes Auswahlmodell für C-Verfahren I

Auswahl des Besonderen, also z.B. Akten mit:

- Mediale Berichterstattung
- Bedeutende Persönlichkeiten, Firmen oder Stiftungen
- Kennzeichnung gem. § 10 Abs. 2 Nr. 4 AktO (Eignung für Ausbildungszwecke)
- Berücksichtigung von auffälligen Prozesshäufungen bzw. Prozesslawinen in bestimmten Sachgebieten als Indikator für zeittypische Phänomene
- Verfahren mit einer besonders langen Dauer (als Proxy für die intensive Auseinandersetzung mit dem Fall)
 - dies erfolgt mittels Stichprobenziehung

Durchführung:

- Mittels Archivsachenvermerk, durchgeführt vom sachbearbeitenden Richter bzw Geschäftsstelle

Von GDA Vorgeschlagenes Auswahlmodell für C-Verfahren II

Auswahl des Alltäglichen soll mittels Stichprobenziehung erfolgen:

Grundgesamtheit: Gerichtsverfahren, die mit einer Entscheidung in der Sache endeten, d.h.:

- streitiges Urteil
- gerichtlicher Vergleich
- Versäumnisurteil
- Anerkenntnis- oder Verzichtsurteil
- Beschluss über einstweilige Verfügung
- Beschluss nach § 91a ZPO

Von GDA Vorgeschlagenes Auswahlmodell für C-Verfahren II

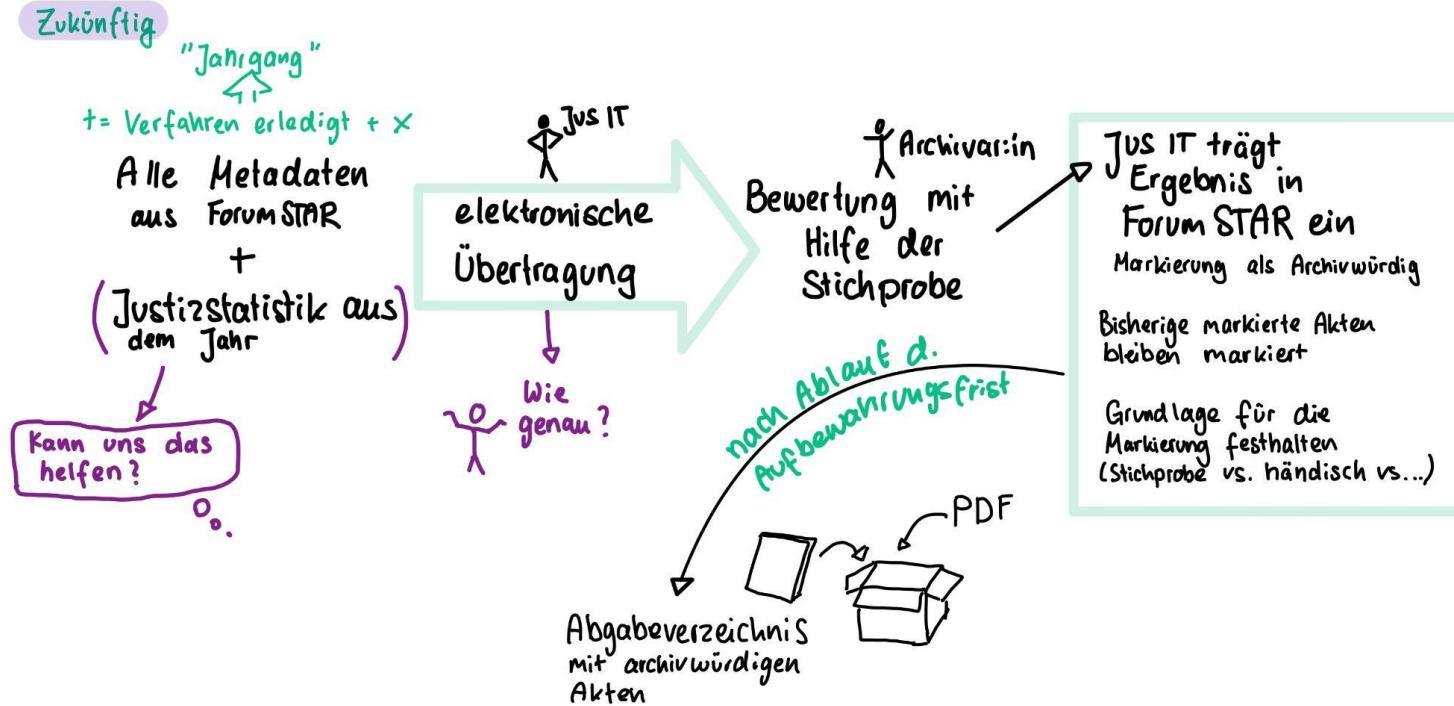
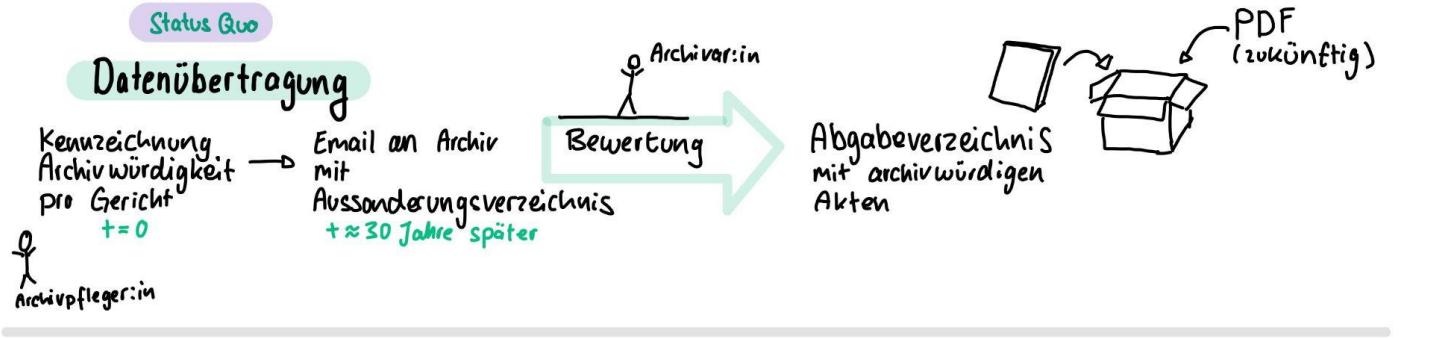
Vorstellungen für die Stichprobenziehung:

- Ziehung aus allen bayerischen Amtsgerichten
 - Einzelne Amtsgerichte dürfen unberücksichtigt bleiben
- Stichprobengröße: 100-1000 Akten oder $\sqrt{N_{\text{Grundgesamtheit}}}$
- Erhöhte Inklusionswahrscheinlichkeit von Akten mit langer Verfahrensdauer
 - Angelehnt an Justizstatistik?
 - Diskretisiert in Tagen ausgewertet?

Ergebnisse

Archivierungsmodell für Zivilgerichtsverfahren der Amtsgerichte

- Methodik – Typisches: Stichprobenziehung
 - Reduzierung auf Erledigungen mit Entscheidungen in der Sache (Aussagekraft)
 - Berücksichtigung aller Amtsgerichte
 - nicht Stichprobenziehung auf Ebene der einzelnen Amtsgerichte
 - Nachteil: einige Amtsgerichte werden immer wieder leer ausgehen
 - Stichprobengröße: Wurzel aus Grundgesamtheit (welcher?) – Alternativen?
 - Verteilung der Gewichtung der Stichprobenziehung nach Länge der Verfahrensdauer (in längeren Verfahren wir mehr über den Lebenssachverhalt)
 - Nicht: Stratifizierung nach Streitwert oder nach Sachgebieten
 - Zentralisierung der Bewertung bei der GDA



Grundgesamtheit

Alle C-Akten, die im Jahr X erledigt wurden und die mit einer Entscheidung in der Sache endeten

Stichprobenumfang

mindestens 300

Wie groß sind die Schichten?

Kreuztabelle \Rightarrow Entscheidung basierend auf Anzahl / Schicht

Abdeckung
↑

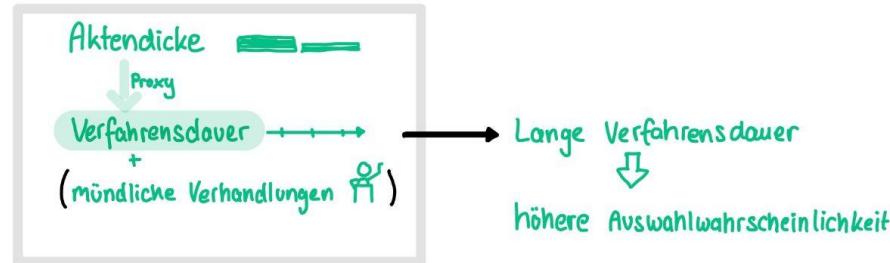
Wie dokumentiert man die Stichprobenziehung?

Verzeichnungsdatensatz
Archiv-Dokumentations-System

Sonderfall oder "Normal" fall?

Schichten definieren
↳ Kategorien wie in Justizstatistik

Schichten



Sachgebiete

Mindestens 1, ansonsten proportional

Regierungsbezirke ?

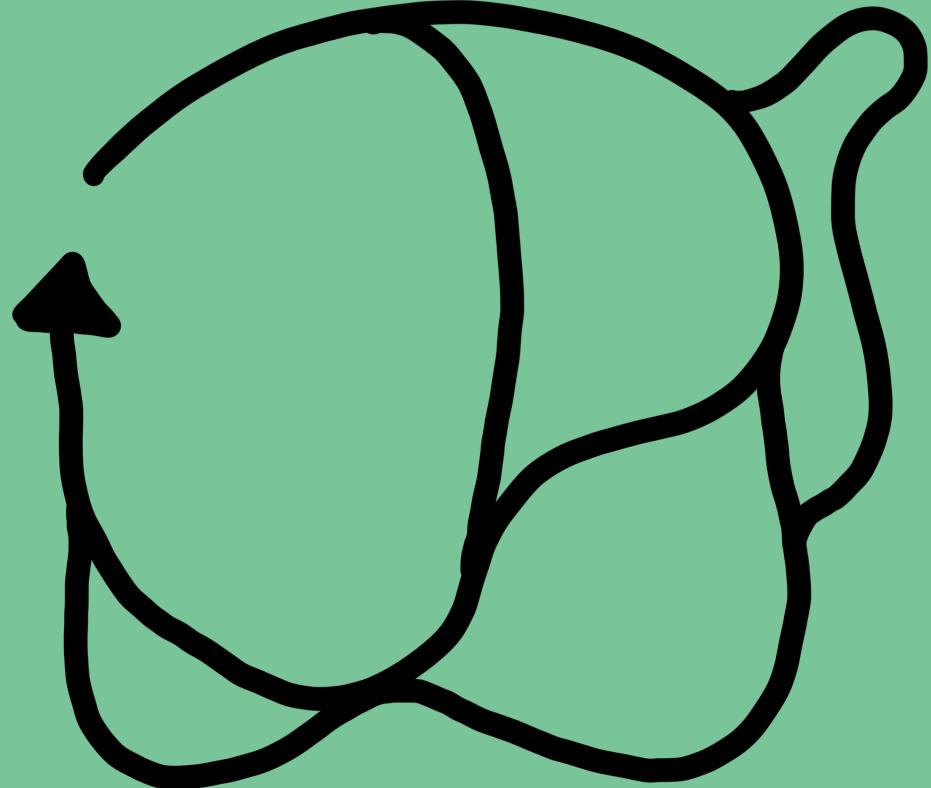
Mindestens 100, ansonsten Proportional

↳ noch zu diskutieren / auszuprobieren

Mittagspause

Umsetzung

Prototyp Flowchart entwickeln



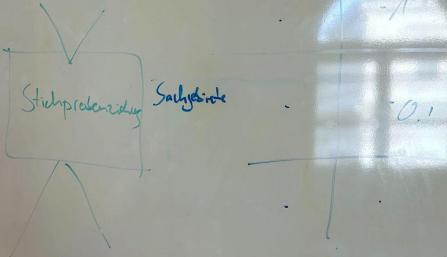
Umsetzung

Prototyp Flowchart entwickeln

- Welche Schritte braucht es, um die geplante Stichprobe zu ziehen?
- Welche Stolperfallen kann es in den einzelnen Schritten geben?
- Was ist jedes Jahr gleich, was verändert sich von Jahr zu Jahr

Input:
- Auszug aus Forum STAR
.csv, .xlsx, .xml
→ Jahrgang

- StichprobengröÙe $n = 0$
- Auswahlwahrscheinlichkeiten D_{ausw} $< 12 \quad \geq 12$
- Definition Grundgesamtheit G $| -1 \quad 0,1$



Output:

- neue Spalten Auswahl $\begin{array}{|c|c|} \hline & 1 \\ \hline 1 & 1 \\ \hline \end{array} \quad \geq 12$
- Tabelle Schichtung $\begin{array}{|c|c|} \hline \text{Auswahlwahrscheinlichkeit} & 1 \\ \hline 1 & 1 \\ \hline \end{array}$
- zusätzlich
- Verteilung über Gewölle $\begin{array}{|c|c|} \hline & 1 \\ \hline 1 & 1 \\ \hline \end{array}$

Grundgesamtheit

Alle C-Akten, die im Jahr X erledigt wurden und die mit einer Entscheidung in der Sache endeten

/ Intervall?
nur alle 5 Jahre?

Trade-off

Stichprobenumfang

mindestens 300 größer!

Wie groß sind die Schichten?

Abdeckung
↑

Kreuztabelle ⇒ Entscheidung basierend auf Anzahl / Schicht

Schichten definieren
↳ Kategorien wie in Justzistatistik

Schichten

Aktendicke ———
Proxy
Verfahrensdauer ←→
+ (mündliche Verhandlungen ♂)

Auswahlwahrscheinlichkeit

Lange Verfahrensdauer

höhere Auswahlwahrscheinlichkeit

Mindestens 1, ansonsten proportional

Mindestens 100, ansonsten Proportional

Sachgebiete

Staatsarchive



Wie dokumentiert man die Stichprobenziehung?
↓
Verzeichnungsdatensatz
Archiv-Dokumentations-System

Daten passend zu Grundgesamtheit anfordern

Alle C-Akten, die im Jahr X erledigt wurden und die mit einer Entscheidung in der Sache endeten

Form? Tabelle mit Zeile pro Akte oder anders?

Tabelle digital übersenden

Daten einlesen + prüfen

Gruppe Heidi

- ↳ Fehlende Werte
- ↳ Fehlende Spalten
- ↳ Falsche Formatierung
- ↳ Falsche Werte

Probleme besprechen

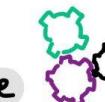
- Schichten
- Auswahlwsk
- Stichprobengröße diskutieren



Kreuztabelle (n) erstellen

	kurz	mittel	lang
Kaufsachen	100	50	5
Mietsachen	-	80	90
:	:	:	:

- Schichten
- Auswahlwsk
- Stichprobengröße festlegen



Wie hält man das zusammen?

Bewertetes Aussonderungsverzeichnis

= Liste mit Aktenzeichen

Stichprobe ziehen

>sample()

Dokumentation der Stichprobe

+ Metadaten der Grundgesamtheit

Markierung der gewählten Akten

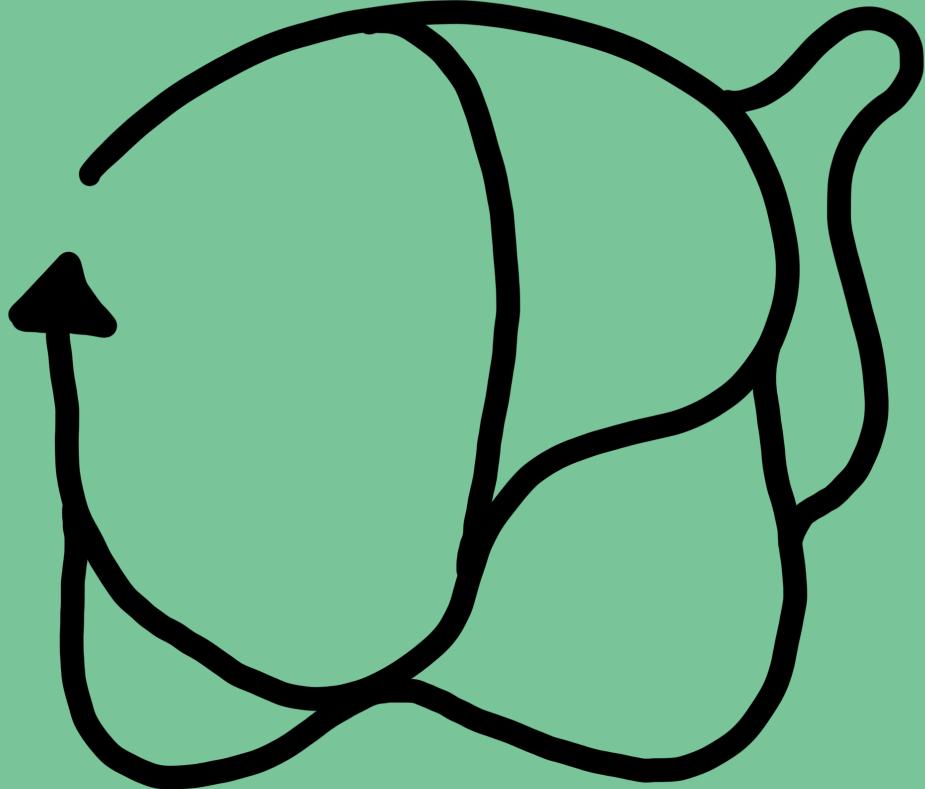
Übertragung an Justiz

=

Pause

Umsetzung

Prototyp Notebook entwickeln



Umsetzung

Prototyp Notebook entwickeln

Implementieren Sie in Ihrer Gruppe
die geplanten Schritte in R.

Zusammenfassung von Sachgebieten in der Justizstatistik

Angelehnt an die Sachgebiete in der Justizstatistik: https://www.verkuendung-bayern.de/files/baymbl/2019/284/anhang/Anhang_01.pdf

Verkehrsunfallsachen - sehr viele Fälle

13 . 11 Verkehrsunfallsachen 29 023 / 22,9 28 055 / 21,7

Körper und Person (Eingriff in die körperliche Unversehrtheit) - besonders schwerwiegende Fälle

15 . 13 Arzthaftungssachen 161 / 0,1 177 / 0,1

25 . 23 Schadensersatzansprüche aus vorsätzlicher Körperverletzung 836 / 0,7 999 / 0,8

Leben und Wohnen (Zusammenleben)

18 . 16 Nachbarschaftssachen 1 121 / 0,9 1 122 / 0,9

20 . 18 Wohnungsmietsachen 21 503 / 17,0 22 238 / 17,2

21 . 19 Sonstige Mietsachen 2 941 / 2,3 3 155 / 2,4

26 . 25 Wohnungseigentumssachen nach § 43 Nrn. 1 bis 4 WEG (Binnenstreitigkeiten) 3 263 / 2,6 3 500 / 2,7

26 . 26 Wohnungseigentumssachen nach § 43 Nr. 5 WEG (Klagen Dritter) 132 / 0,1 147 / 0,1

Zusammenfassung von Sachgebieten in der Justizstatistik

Angelehnt an die Sachgebiete in der Justizstatistik: https://www.verkuendung-bayern.de/files/baymbl/2019/284/anhang/Anhang_01.pdf

Wirtschaftsrecht

12 . 10 Bau-/ArchitektenSachen (ohne Architektenhonorarsachen) 1 265 / 1,0 1 198 / 0,9

14 . 12 Kaufsachen 16 715 / 13,2 16 373 / 12,6

16 . 14 Reisevertragssachen 6 758 / 5,3 5 282 / 4,1

17 . 15 Kredit-/Leasingsachen 2 297 / 1,8 2 571 / 2,0

22 . 20 Ansprüche aus Versicherungsverträgen (ohne Verkehrsunfallsachen) 4 551 / 3,6 5 109 / 3,9

23 . 21 Gesellschaftsrechtliche Streitigkeiten 1 265 / 1,0 738 / 0,6

26 . 24 Honorarforderungen von Personen, für die eine besondere Honorarordnung gilt 5 328 / 4,2 5 774 / 4,5

Sonstige

19 . 17 Schuldrechtsanpassungs- und Bodenrechtssachen der neuen Länder 6 / 0,0 2 / 0,0

27 . 39 Sonstiger Verfahrensgegenstand 29 487 / 23,3 33 025 / 25,5

ANFORDERUNGEN ANS PROGRAMM

Zuständigkeiten:

Datenanalyse: GDA

Wir brauchen
gute Defaults

Möglichkeiten zur
Einstellung bieten, aber
die jetzigen Entscheidungen
vorgeben.

- Auswahlkriterien &
-Wahrscheinlichkeiten
eingeben (pro Schicht)
 - ↳ min/max
 - ↳ Proportional
 - ↳ Wahrscheinlichkeiten
 - ↳ Multiplikator ?

Stichprobe
ziehen

>sample()

▪ Dokumentation
der
Stichprobe

Automatisch

+ Metadaten der Grundgesamtheit

Daten einlesen + prüfen

- Tabelle anzeigen
- Übersichten + Grafiken
- Filter

+

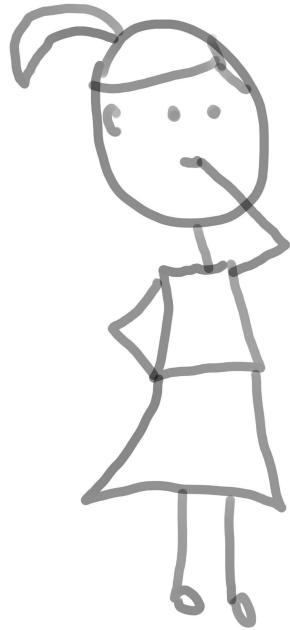
✓ Checkliste: was ist
zu prüfen?

- ↳ Fehlende Werte
- ↳ Fehlende Spalten
- ↳ Falsche Formatierung
- ↳ Falsche Werte

Gruppe
Heidi

Kreuztabelle (n)
erstellen

- Schichten auswählen
 - ↳ Dropdown
 - ↳ Kategorien anpassen
- Ansicht / Darstellung



Was war für Sie heute überraschend?