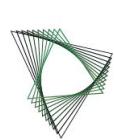


# Modul 1: Einführung, Problemerfassung und Arbeitsabläufe

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)  
[www.ada-oeffentliche-verwaltung.de](http://www.ada-oeffentliche-verwaltung.de)



**BERD**  
@NFDI



Bayerisches Staatsministerium  
für Digitales



# Über uns



Prof. Dr. Frauke Kreuter



Dr. Malte Schierholz



Dr. Marcel Neunhoeffer



Prof. Dr. Helmut Küchenhoff



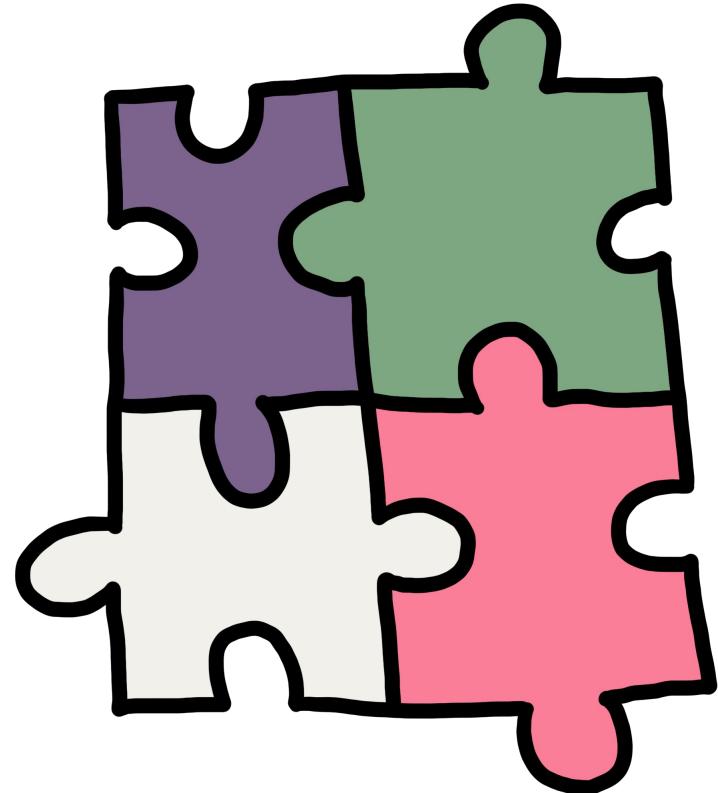
Dr. Heidi Seibold



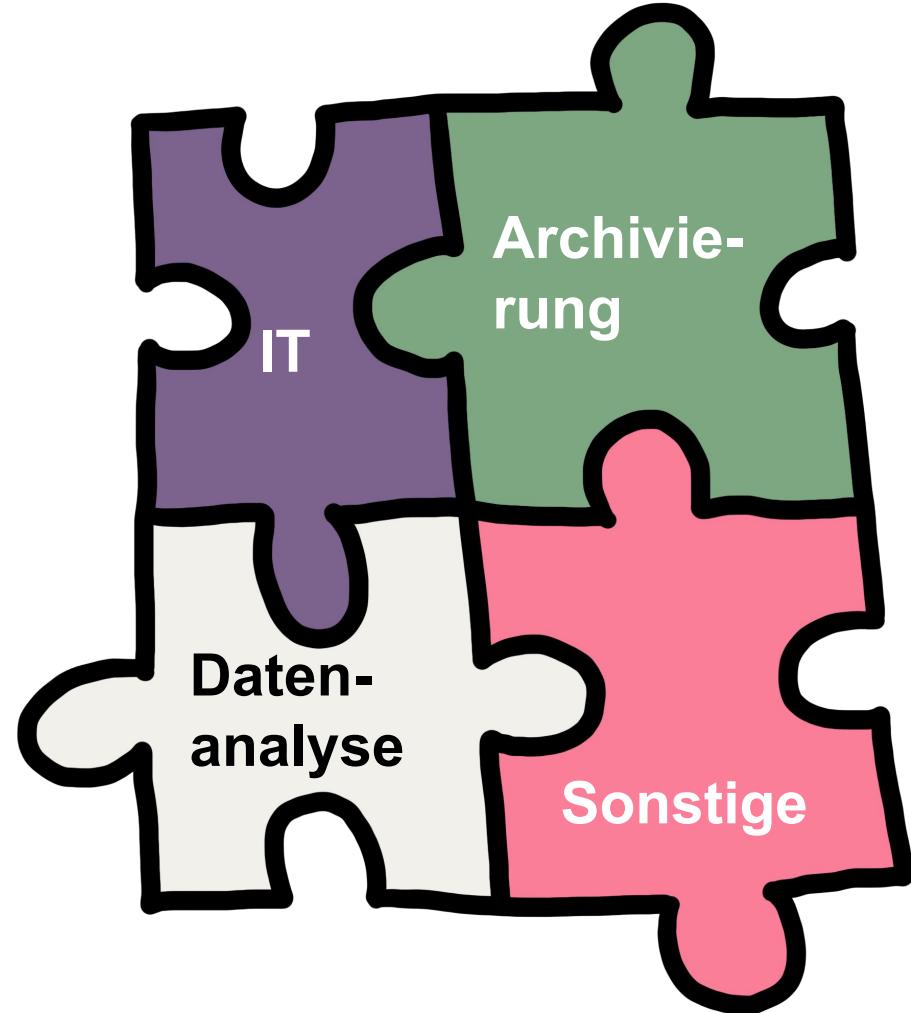
Felix Henninger

# Über euch/Sie

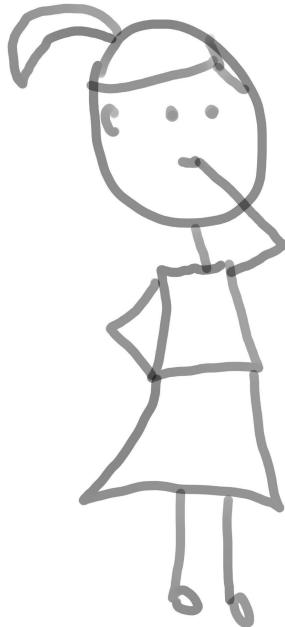
- Name
- Fachlicher Hintergrund (kurz!)
- Grüner Zettel: *Es war ein guter Workshop, wenn...*
- Roter Zettel: *Das soll hier auf gar keinen Fall passieren*



# Data Science ist ein Team Sport!



# Modul 1



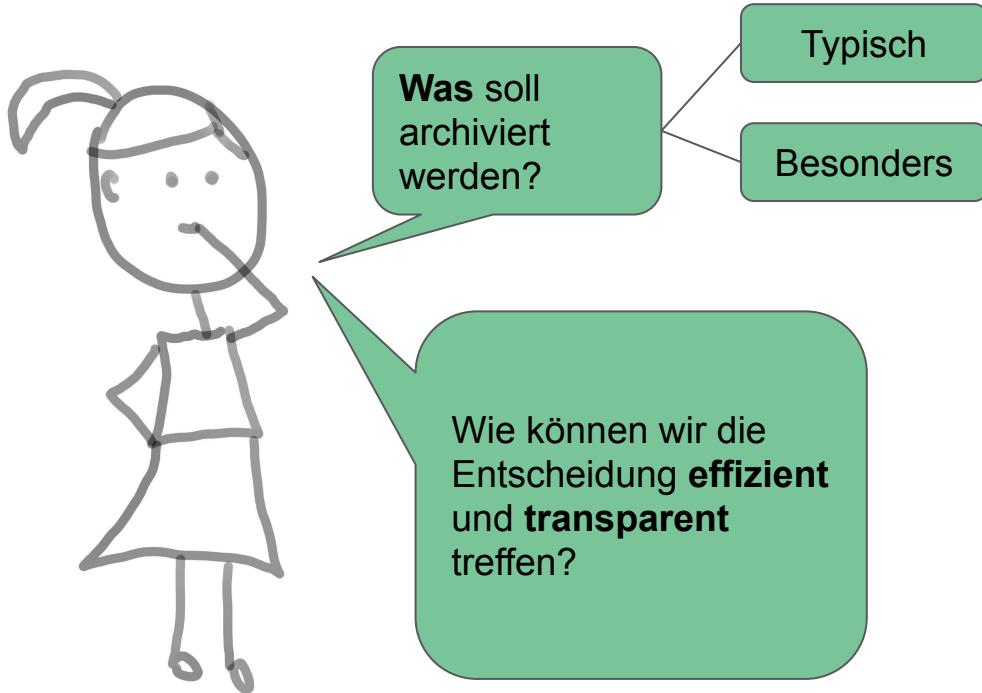
Am Ende dieses Moduls können Sie...

- ... Kontext und Vorarbeiten des Projekts nennen.
- ... verstehen, was das Ziel dieses Projekts ist.
- ... die Datenanalyse als Teil eines größeren Kontexts beschreiben.
- ... die einzelnen Schritte von der Vereinbarung der Datenübertragung bis hin zur Archivierung der einzelnen Akten beschreiben.
- ... mit den relevanten Personen in Kontakt treten und einen reibungsarmen Ablauf ermöglichen.

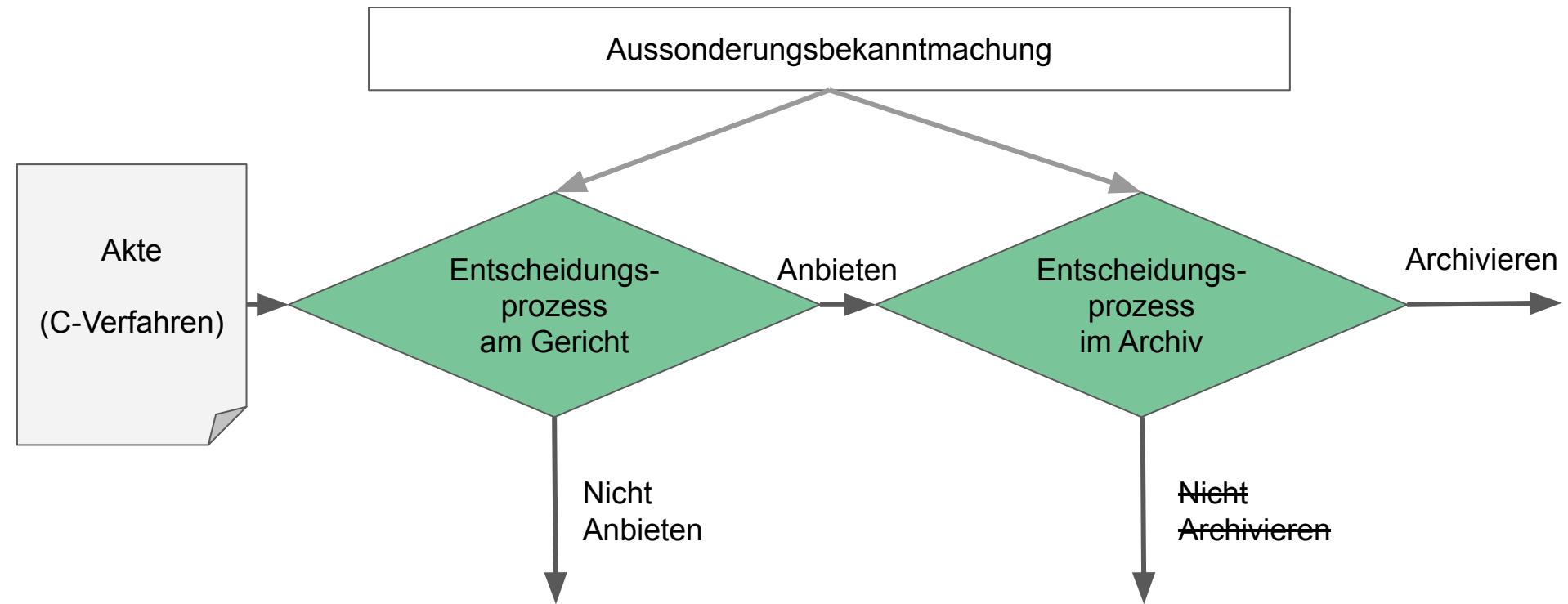
Einführung	10:00 - 11:05
Pause	11:05 - 11:20
Arbeitsabläufe: von der Datenübertragung zur Archivierung	11:20 - 11:50
Cloud	11:50 - 12:00
Mittagspause	12:00 - 13:00
Sichtung der bisherigen Notebooks	13:00 - 13:50
Pause	13:50 - 14:10
Sichtung der bisherigen Notebooks ctd.	14:25 - 15:00
Pause	15:00 - 15:10
Wrap-up und Ausblick	15:10 - 15:30

Es beginnt mit einer Frage...

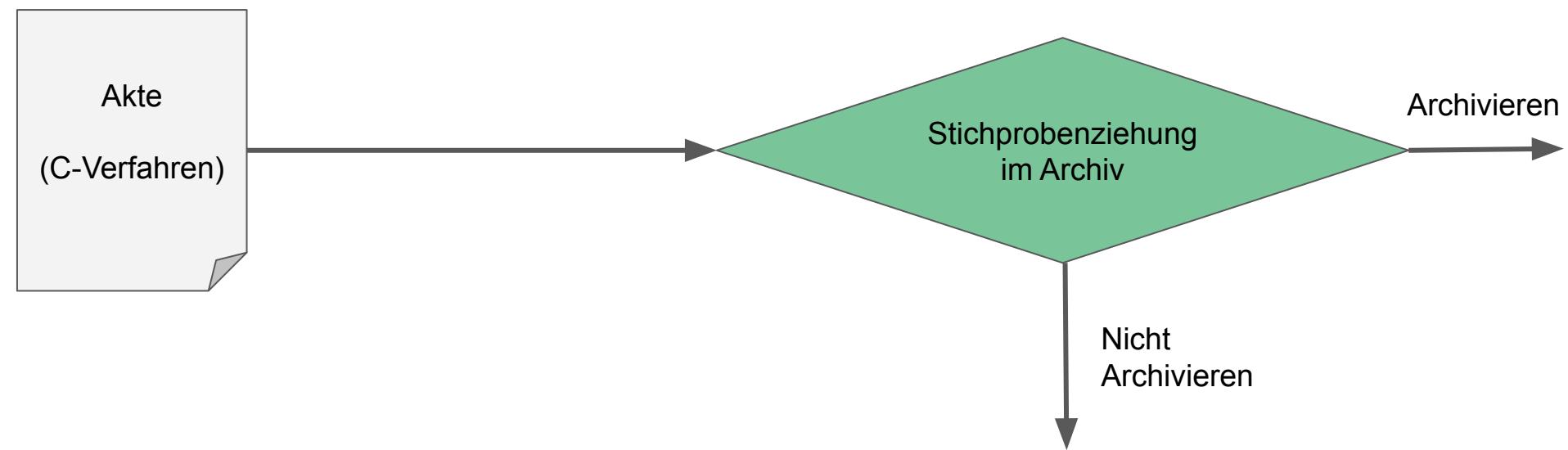
# Was ist eigentlich die Frage? Problemstellung definieren.



# Entscheidungsprozess heute

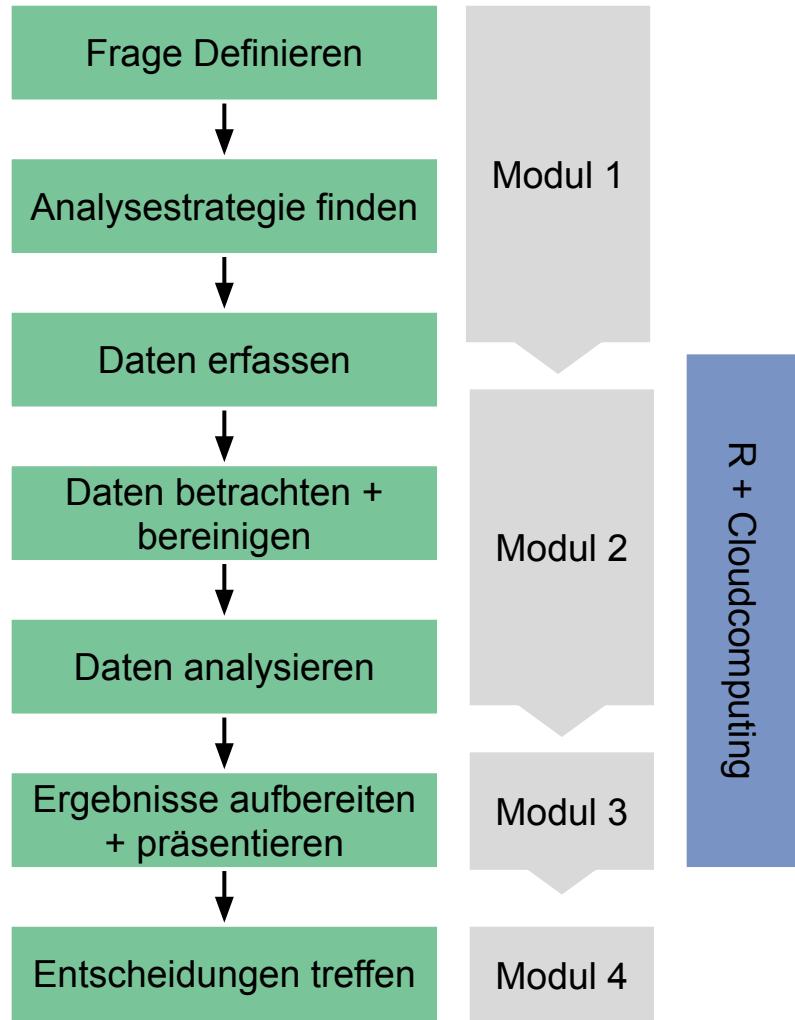


# Entscheidungsprozess mit Hilfe von Stichprobenziehung: V.a. für typische Fälle

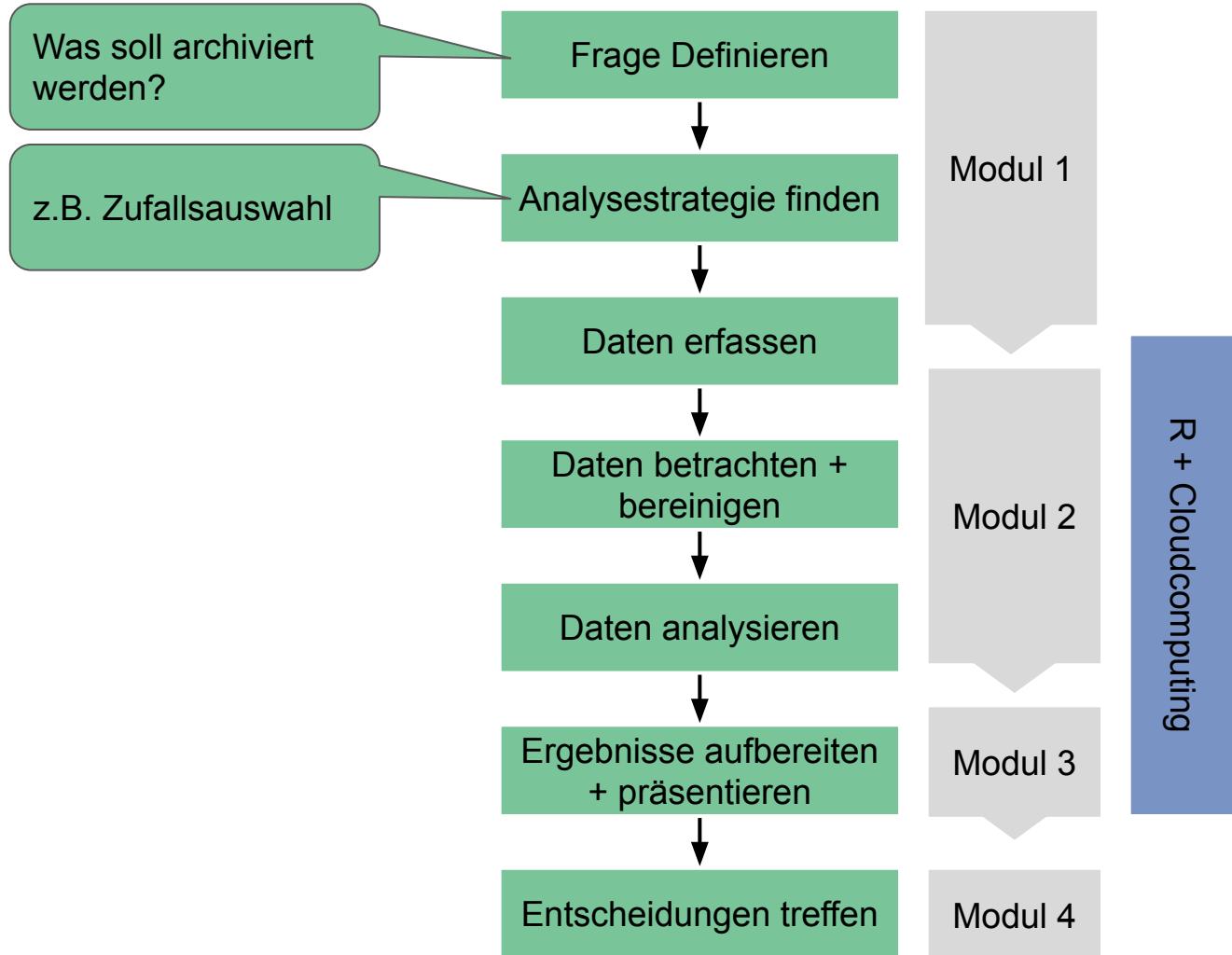


Was bisher geschah

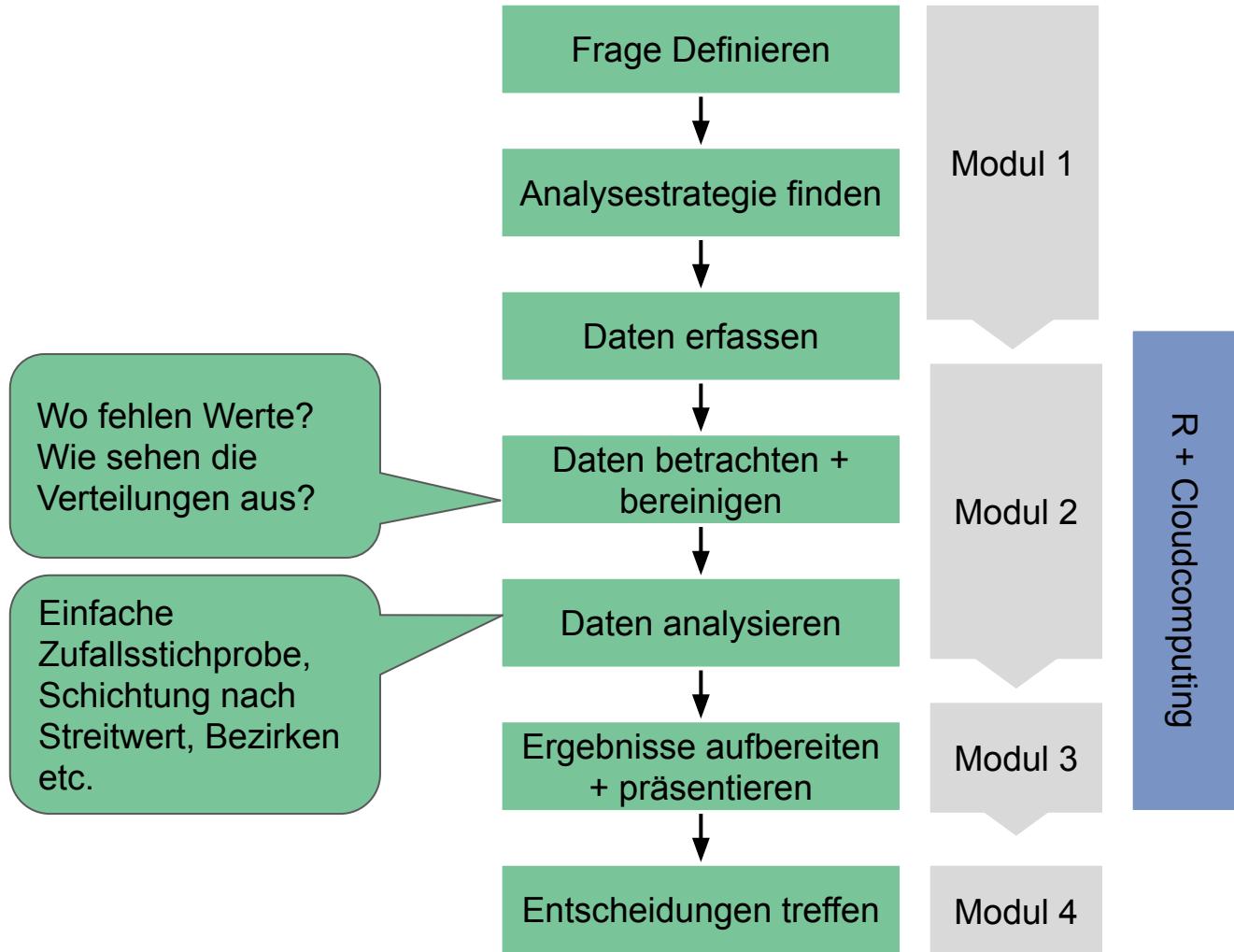
# Vorgehen bei der Beantwortung von Fragen mit Daten



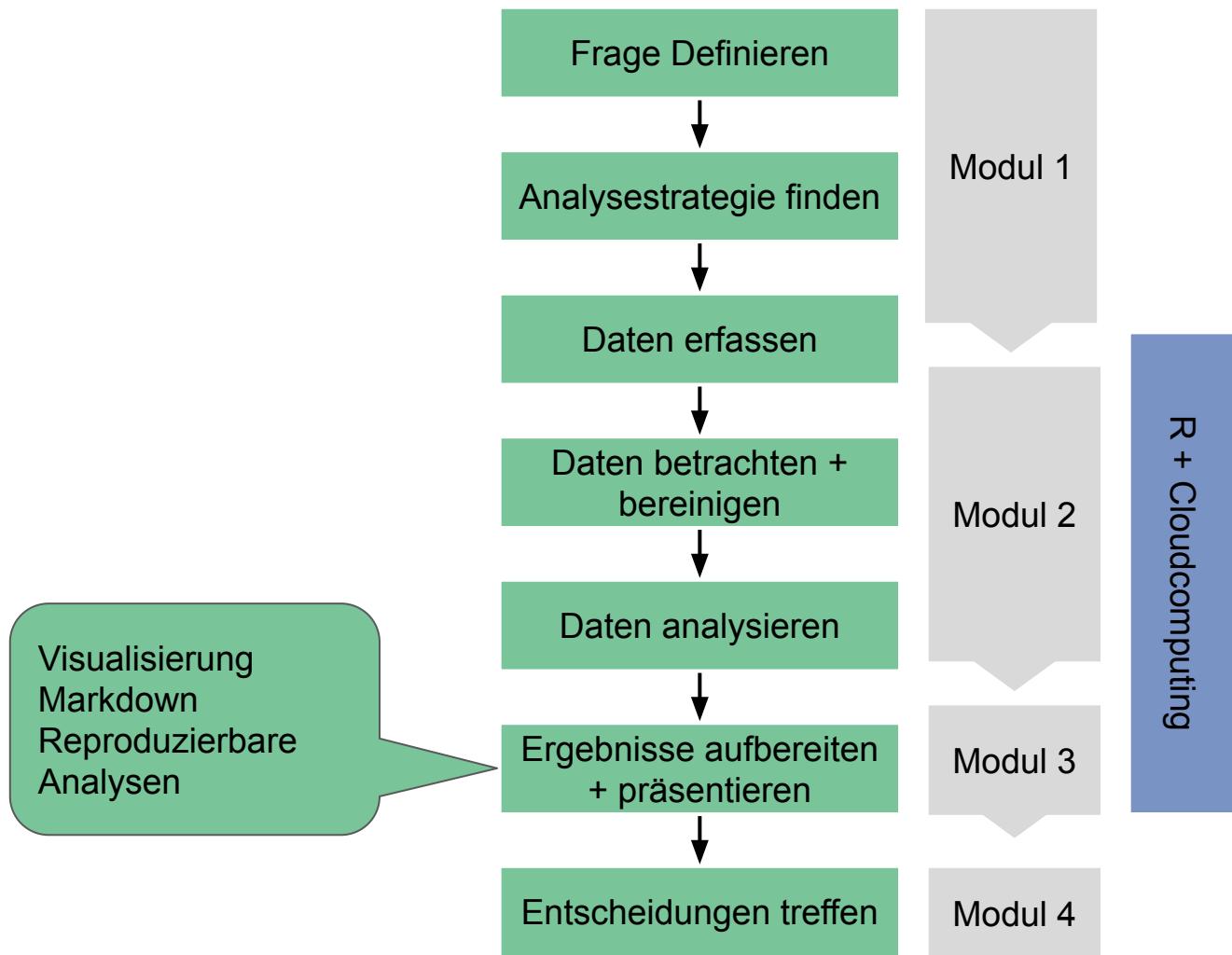
# Rückblick



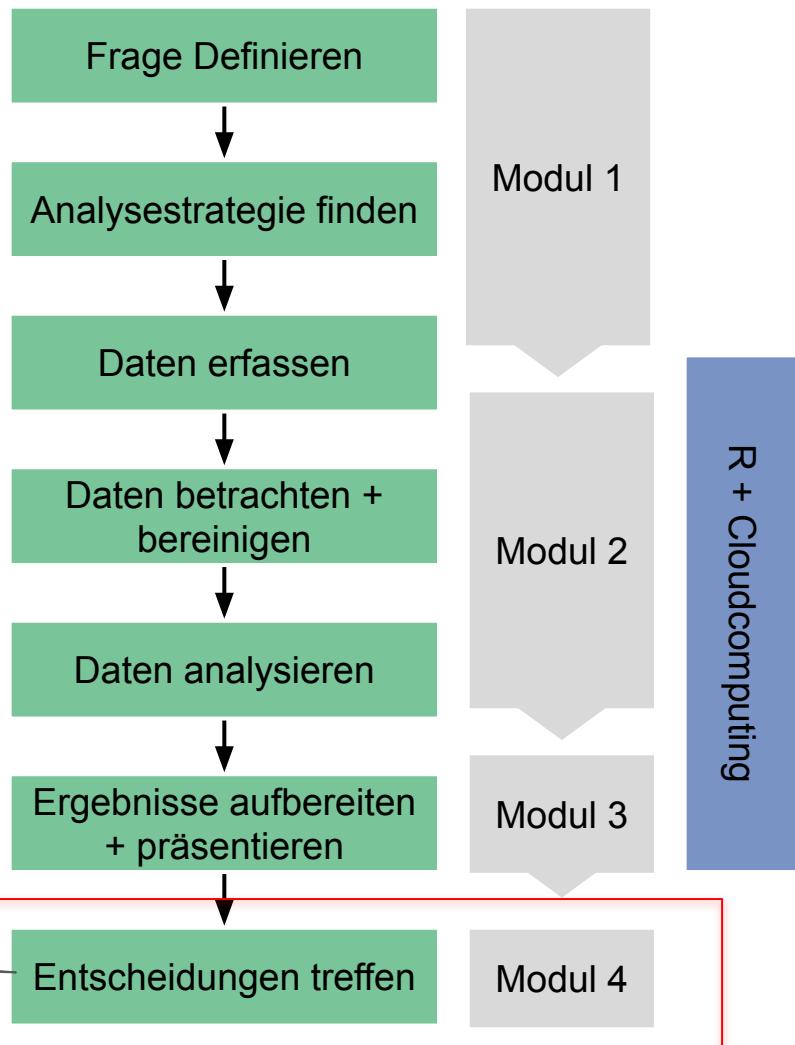
# Rückblick



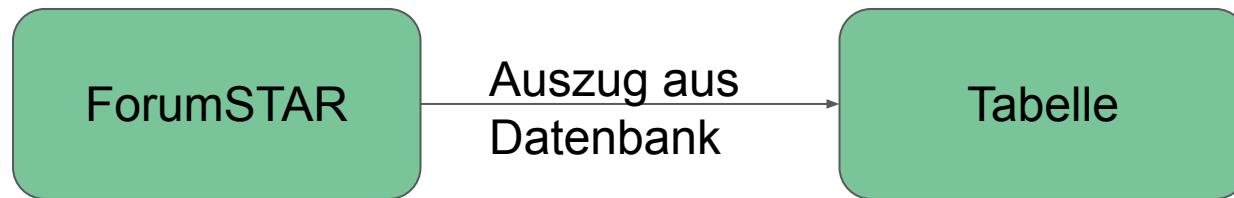
# Rückblick



# Rückblick



# Welche Daten gibt es in unserem Projekt?



AZ: 1 F 1/15

### Daten betr. Scheidung

Wurde dieses Verfahren nach dem 20.06.2012 eingeleitet?  ja  nein

Sind beide Ehegatten Iraner?  ja  nein

örtliche Zuständigkeit [?](#)

F\_Sch\_ROM III

**Heirat:** 15.08.1990 **Ehezeitbeginn:** 01.08.1990 [?](#) [§](#) **ZU SchAntrag:** 18.01.2015 **Ehezeitende:** 31.12.2014 [?](#)

vor dem Standesbeamten?  Ja  Nein

**Standesamt:** Rosenheim **Staat:** Deutschland [?](#)

**HeiratRegNr:** 15/1990 **(wenn Heiratsort im Ausland)**

Hatten die Ehegatten zwischen Rechtshängigkeit und letzter mtl. Verhandlung zu irgendeinem Zeitpunkt - gemeinsam ihren gew. Aufenthaltsort in Deutschland?  ja  nein [?](#)

Haben die Ehegatten eine Rechtswahl getroffen?  ja  nein

Hat es früher eine Trennungsentscheidung gegeben (Art. 9 ROM III-VO)?  ja  nein [?](#)

Hatten beide Ehegatten bei Anrufung des Gerichts ihren gew. Aufenthaltsort in Deutschland?  ja  nein

Staatsangehörigkeit bei ZU des Scheidungsantrags (wegen VA gem. Art. 17 Abs. 3 EGBGB):

<input checked="" type="checkbox"/> Ehemann: <b>Türkei</b>	<input checked="" type="checkbox"/> Ehefrau: <b>Türkei</b>
<input type="checkbox"/> außerdem - nicht "effektiv" -	<input type="checkbox"/> außerdem - nicht "effektiv" -
<input type="checkbox"/> Ehemann ist staatenlos	<input type="checkbox"/> Ehefrau ist staatenlos

War einer der Ehegatten bei ZU des Scheidungsantrags Flüchtling oder Asylberechtigter?

Antragsteller/in <input checked="" type="checkbox"/> Ehefrau <input type="checkbox"/> Ehemann	<input type="checkbox"/> ja <input checked="" type="checkbox"/> nein
VB Antragsgegner/in <input checked="" type="checkbox"/> kein VB oder VB stellt keinen Antrag <input type="checkbox"/> stellt Scheidungsantrag <input type="checkbox"/> beantragt Abweisung Scheidungsantrag	<input type="checkbox"/> widerspricht <input type="checkbox"/> widerspricht nicht <input type="checkbox"/> äußert sich nicht
Antragsgegner/in persönlich <input checked="" type="checkbox"/> stimmt zu <input type="checkbox"/> widerspricht	<input type="checkbox"/> ja <input type="checkbox"/> nein

Soll der Verfahrensteil Scheidung begründet werden?  ja  nein

Feststellungen des Gerichts zur Dauer des Getrenntlebens  weniger als 1 Jahr (unzumutbare Härte)  mindestens 1 Jahr

[Generelle Vorgaben ändern](#)

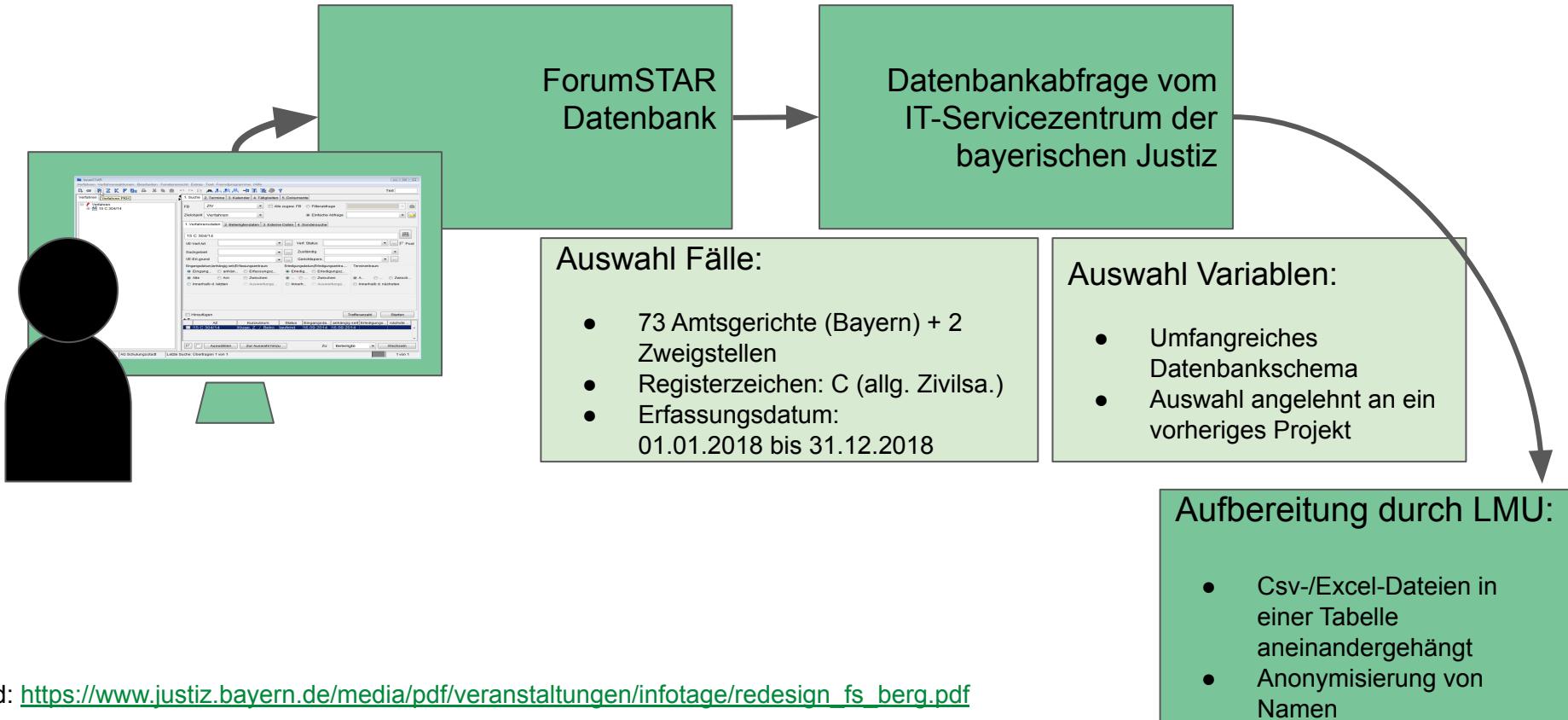
**Weiter** [E](#)

DE 15:46 18.01.2015

# Wie entstehen die verfügbaren Daten?



# Wie entstehen die verfügbaren Daten?



# Ergebnisse Workshop-Serie 1

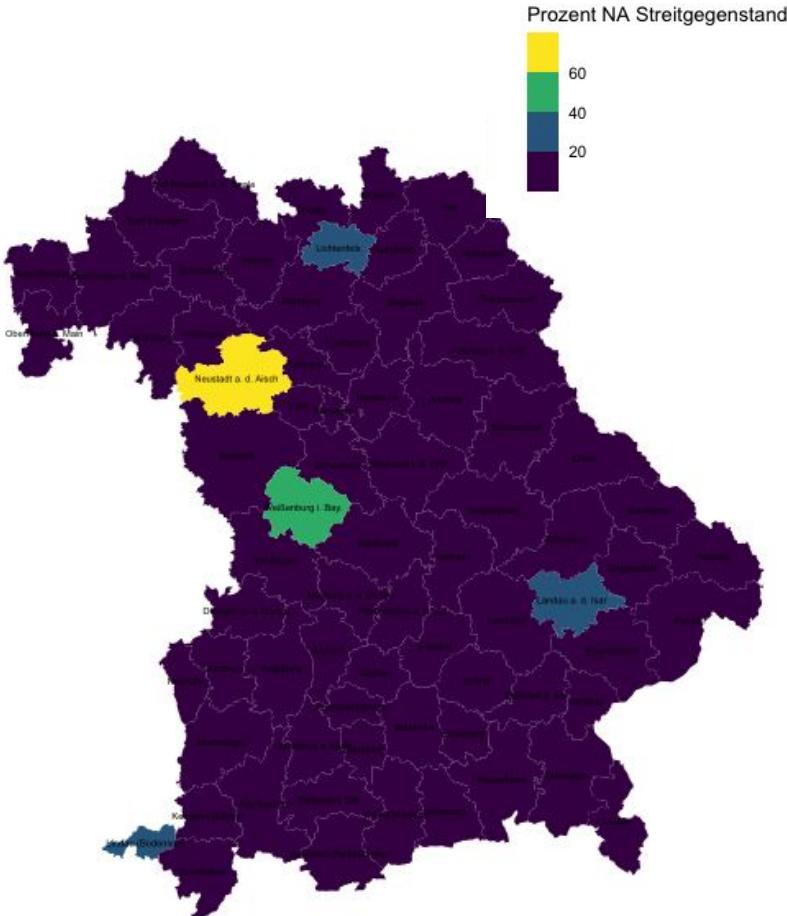
# Team 1

Fokus auf  
Datenqualität

Relevante Spalten:

- Gericht / Bezirk
- Streitwert
- Gesamtstreitgegenstand
- Streitgegenstand
- Sachgebiet
- Anzahl Termine
- Daten (Datum)
- Anzahl an Beteiligten
- Es ging voraus... (bisherige Verfahren)
- Informationen über die Person (Minderjährige, berühmte Personen, Beteiligungsart, etc.)
- ggf. weitere

Streitgegenstand	Anzahl
Vollstreckungsabwehrklage	305
Beseitigung	362
Abgabe einer Willenserklärung	388
Entschädigung	444
Vornahme einer Handlung	589
Urheberrecht	737
Herausgabe	784
Schmerzensgeld	801
Feststellung	826
Zustimmung zur Mieterhöhung	1015
Unterlassung	1069
Beschlussanfechtung	1269
einstweiliger Verfügung	1331
Räumung und Forderung	1998
Räumung	3126
Räumung und Herausgabe	3938
Duldung	4876
NA	10065
Schadensersatz	28051
Forderung	63277



# Team 2

**Ziel:** Interessensneutral die Tätigkeit des Staat abbilden ... womit sich Staat beschäftigt hat

**Grundgesamtheit** kennen  
und Populationswerte und **Ziehungswahrscheinlichkeiten** aufbewahren (Justizstatistik)

**Gesamtzahl der Akten:** 126.945  
“Certainty Strata” (Angeboten): 372

Reduktion der Population um Erledigungsgrund “Rücknahme” 16.520  
Quote etwa: 1-5%

Streitwert (>4000 JN) x Länge (>10, JN)  
**100% aus den besonderen**

Sachgebiet (17) x Erledigungsgrund (29) x  
6% alles was nicht besonders ist

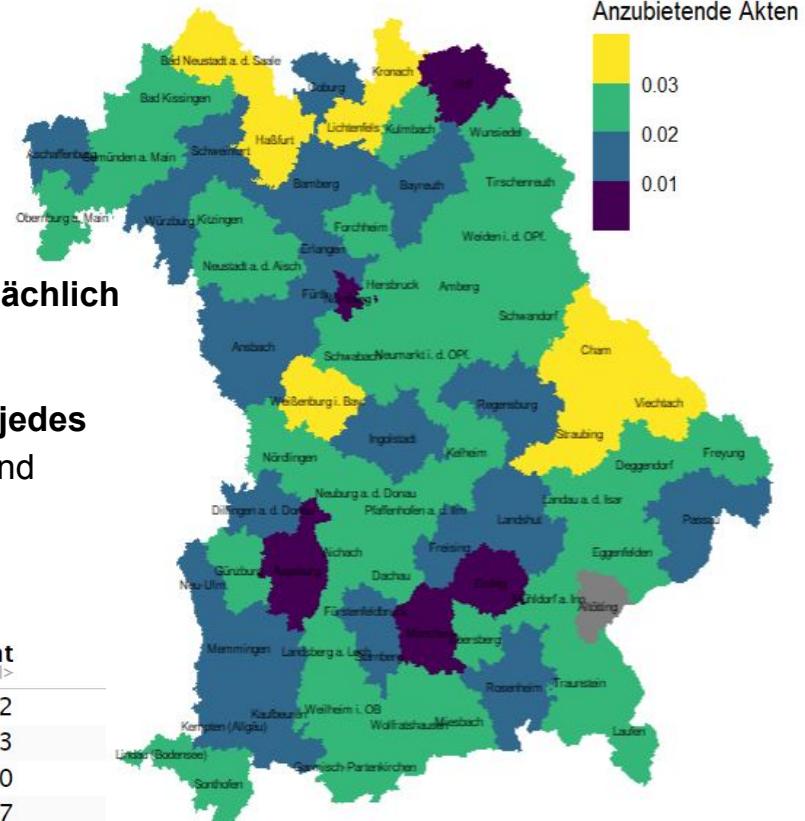
	<10 Tage	>10
<4000 Euro	71.745	199
> 4000 Euro	12.250	<b>103</b>

# Team 3

## Zufallsziehung (für typische Fälle)

- Berücksichtige nur Verfahren, bei denen ein **Prozess** tatsächlich stattgefunden hat
  - **Einfache Zufallsstichprobe getrennt pro Bezirk**
  - Stichprobenumfang: **Stichprobengröße = Wurzel(N)** für jedes **Gericht** (Kompromiss zw. gleich viele Akten pro Gericht und proportionale Anzahl pro Gericht)

strata	akten_pro_stratum	anzahl_akten_pro_schicht
<chr>	<int>	<dbl>
Aichach	457	22
Amberg	489	23
Ansbach	846	30
Aschaffenburg	690	27
Augsburg	2449	50
Bad Kissingen	297	18
Bad Neustadt a. d. Saale	244	16
Bamberg	796	29
Bayreuth	811	29
besonderes Verfahren	49	49



# Team 4

In den vorliegenden Daten haben wir **zu wenig Information** für die sinnvolle Auswahl von Akten zur Archivierung.

Eine **einfache Zufallsstichprobe ist nicht ausreichend** für die Auswahl im Einklang mit der Aussonderungsbekanntmachung.

## ABER

Wir können/wollen eine einfache Zufallsstichprobe nutzen um **besser zu verstehen was besondere Akten ausmacht**.

Mit den gelernten Kriterien können Akten **in Zukunft** mit Hilfe einer **geschichteten Zufallsstichprobe** ausgewählt werden.

- Akten die auf die Beschreibung in der Aussonderungsbekanntmachung passen werden immer (mit höherer Wahrscheinlichkeit) ausgewählt.
- Darüber hinaus werden weitere Akten mit geringerer Auswahlwahrscheinlichkeit zufällig gezogen.  
→ So können zeittypische Verfahren abgedeckt werden

Perspektive GDA

# **ADA Bayern – Erkenntnisse und Ergebnisse aus der ersten Workshopserie**

Dr. Michael Unger – Andreas Nestl  
Generaldirektion der Staatlichen Archive Bayerns

## Kurze Wiederholung

**Projektziel:** Ermittlung archivwürdiger Akten aus einer sehr großen Anzahl massenhaft gleichförmiger Unterlagen: Verfahrensakten der allgemeinen Zivilgerichtsbarkeit der Amtsgerichte (C-Verfahren; ca. 120.000 erledigte Verfahren/Jahr)

- anhand von zu jedem Einzelfall vorliegenden Metadaten
- durch datenanalytische Verfahren
- unter einer **qualitativen** Verbesserung bestehender Auswahlmethoden (Sampling)
- im Interesse einer möglichst auswertungsoffenen, transparenten Überlieferungsbildung
- und im Interesse von Effizienzgewinnen für Justiz und Archive

# Kurze Wiederholung - bisherige Bewertungspraxis

## BayArchivG

- Anbietetpflicht für nicht mehr benötigte Unterlagen
- Archivwürdigkeit: bleibender Wert für wissenschaftliche Forschung, für berechtigte Belange Betroffener und Dritter, für Gesetzgebung, Rechtsprechung und Verwaltung
- Bewertung durch Archiv im Benehmen mit der abgebenden Stelle
- Umfang anzubietender gleichförmiger Unterlagen kann durch Vereinbarung festgelegt werden

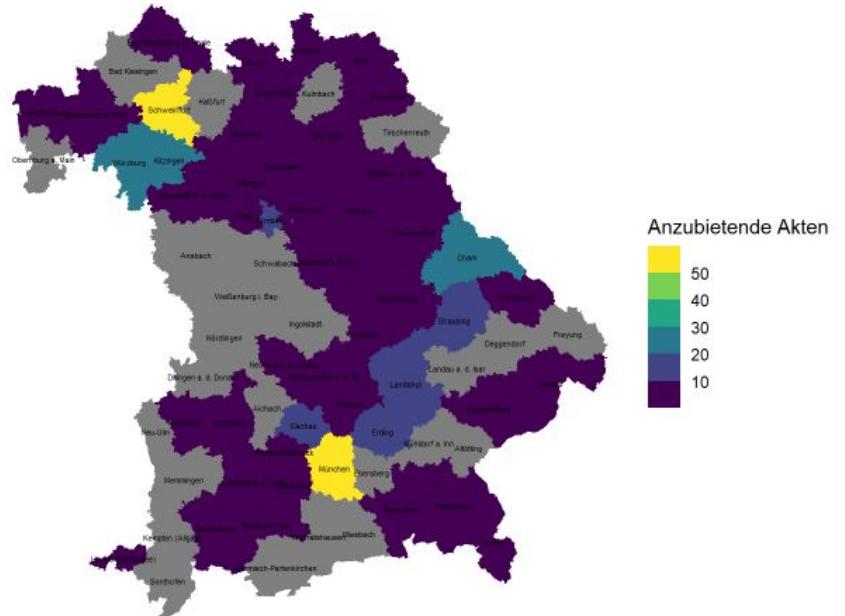
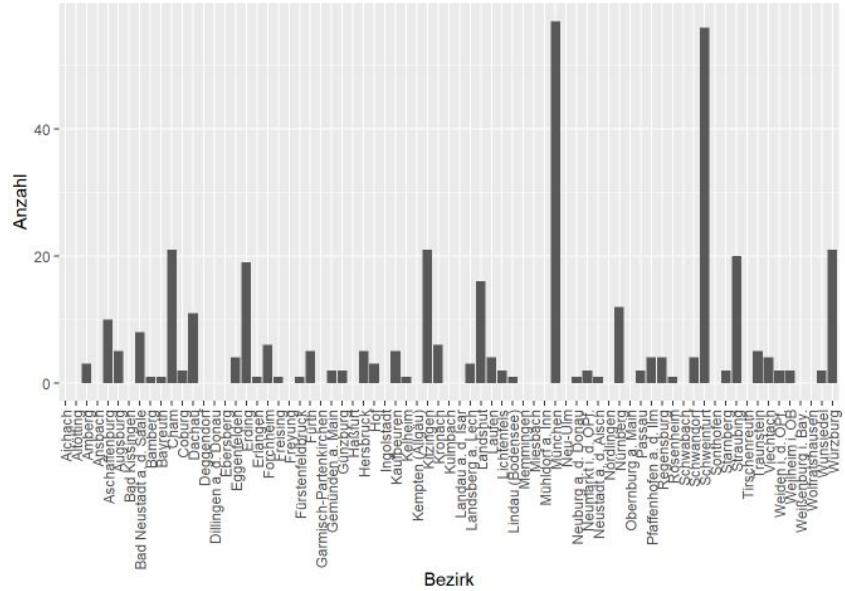
## Aussonderungsbekanntmachung Justiz

- Anbietung nach Registerzeichen als Positivkatalog - Verfahren der allgemeinen Zivilgerichtsbarkeit nicht berücksichtigt
- Ersichtlich bedeutsame Unterlagen nach Kennzeichnung mit **Archivsachenvermerk**
- Kriterien (insb. Nr. 10.2. Aussonderungsbekanntmachung Justiz) – ersichtlich **bedeutsame** Verfahren:
  - Akten über Verfahren, an denen bekannte Persönlichkeiten des öffentlichen Lebens (Politiker, Wissenschaftler, Künstler usw.), bedeutende Familien, Stiftungen, Firmen oder sonstige Unternehmen beteiligt waren;
  - Akten über Verfahren, in denen Landesrecht aus der Zeit vor 1900 angewendet wurde;
  - Akten über Verfahren mit wertvollen Hinweisen auf Volksstute und Brauchtum, auch wenn es sich um Bagatelfälle handelte;
  - Akten über **zeittypische** Zivilrechtsstreitigkeiten (Mietstreitigkeiten, Reisevertragssachen, Sachmängelprozesse usw.)

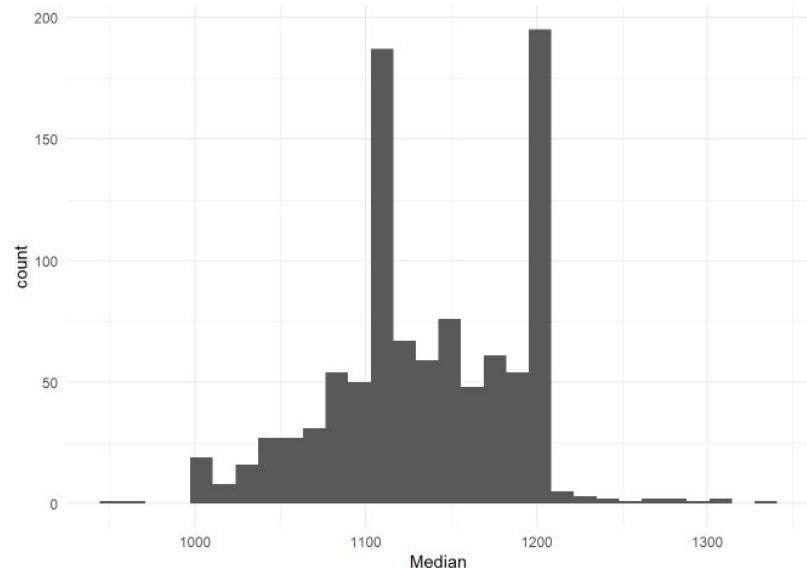
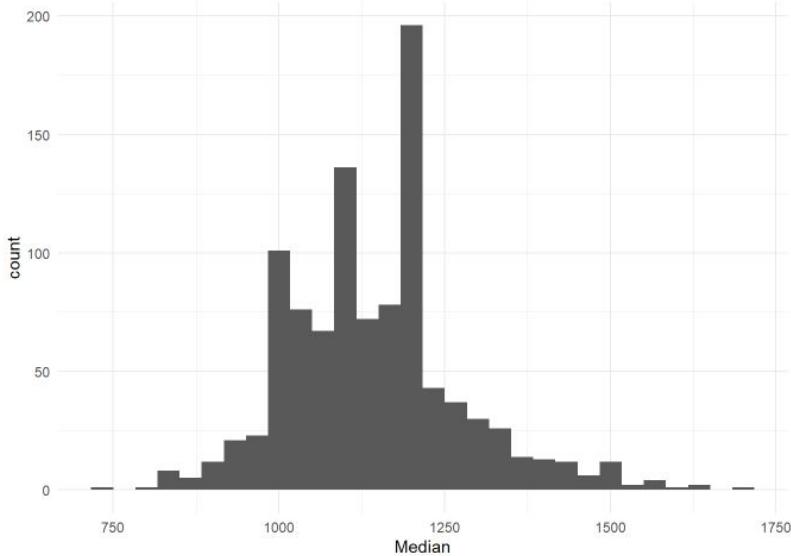
## Praxis: tendenziell aleatorisch

- Archivsachenvermerk durch den Sachbearbeiter bzw. die Geschäftsstelle (delegierte Vorbewertung)
- Auswahl der Papierakten am Fach nach Rücksprache mit dem Archiv (Kriterien: Pressestelle, Aktendicke, Rechtsmitteleinlegung, ...)

# Kurze Wiederholung – Daten visualisieren



# Kurze Wiederholung – einfache Stichprobenziehung



Ziehung einfacher Zufallsstichproben; Auswirkung der **Stichprobengröße** auf die Wiedergabe der Grundgesamtheit am Beispiel des Streitwertmedians (1133.49 €). Links n=200, rechts n=1000; jede Ziehung wurde 1000 mal durchgeführt

# Kurze Wiederholung – stratifizierte Stichprobenziehung

Bei einer stratifizierten Stichprobe werden weitere Informationen aus der Grundgesamtheit zur Stichprobenziehung verwendet.

Die Grundgesamtheit wird in Schichten unterteilt, die unterschiedlich gewichtet werden können, z. B. Berücksichtigung jedes Amtsgerichtsbezirk mit einem gleichen Anteil.

Akten aus kleinen Amtsgerichtsbezirken hatten in dieser Stichprobe eine deutlich höhere Wahrscheinlichkeit, in die Stichprobe zu gelangen. Wenn man dies unberücksichtigt lässt, können keine statistisch validen Aussagen über die Grundgesamtheit getroffen werden!

beschreibung_stichprobe		
Bezirk	Anzahl Akten	Stichprobengröße
<chr>	<int>	<dbl>
Aichach	927	10
Altötting	784	10
Amberg	1065	10
Ansbach	1689	10
Aschaffenburg	2678	10
Augsburg	5395	10
Bad Kissingen	700	10
Bad Neustadt a. d. Saale	500	10
Bamberg	1715	10
Bayreuth	1582	10

1-10 of 73 rows

Previous [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) ... [8](#) [Next](#)

# Kurze Wiederholung – stratifizierte Stichprobenziehung

Stratum Verfahrensdauer:

- kurz (< 100 Tage)
- mittel (100-1500 Tage)
- lang (> 1500 Tage)
- bei fehlenden Werten wurde von einem kurzen Verfahren ausgegangen

Stichprobengröße pro Stratum wurde manuell festgelegt

Alternativen:

- proportional zum Umfang der jeweiligen Schicht
- gleiche Stichprobengröße pro Stratum
- optimierte Genauigkeit der Schätzung für eine Variable

verfahrensdauer	akten_pro_stratum
<chr>	<int>
kurzes Verfahren	72474
langes Verfahren	178
mittleres Verfahren	54293

3 rows

```
beschreibung_stichprobe <- beschreibung_stichprobe %>% mutate(  
  anzahl_akten_pro_schicht = case_when(  
    verfahrensdauer == "kurzes Verfahren" ~ 50,  
    verfahrensdauer == "mittleres Verfahren" ~ 100,  
    verfahrensdauer == "langes Verfahren" ~ 178)  
)
```

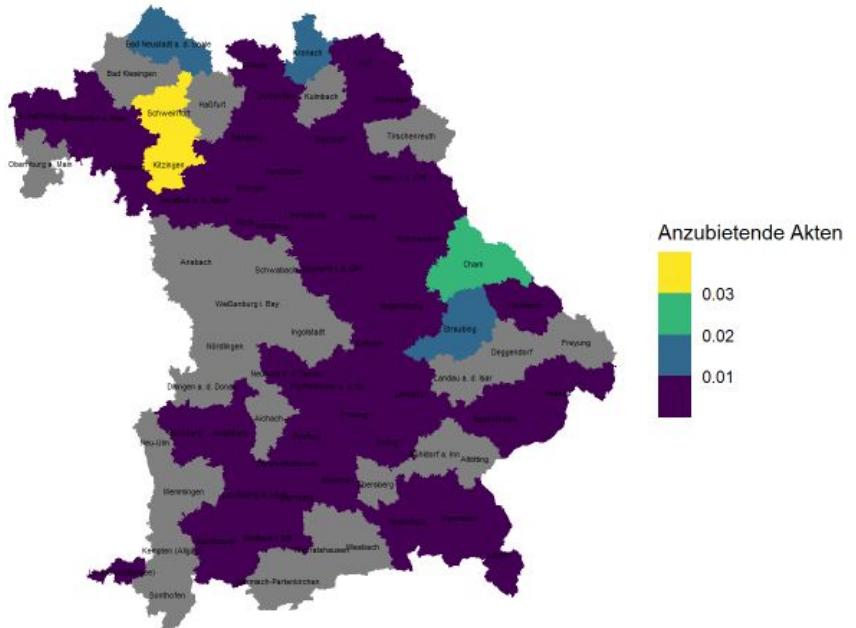
```
beschreibung_stichprobe
```

verfahrensdauer	akten_pro_stratum	anzahl_akten_pro_schicht
<chr>	<int>	<dbl>
kurzes Verfahren	72474	50
langes Verfahren	178	178
mittleres Verfahren	54293	100

3 rows

# Erkenntnisse

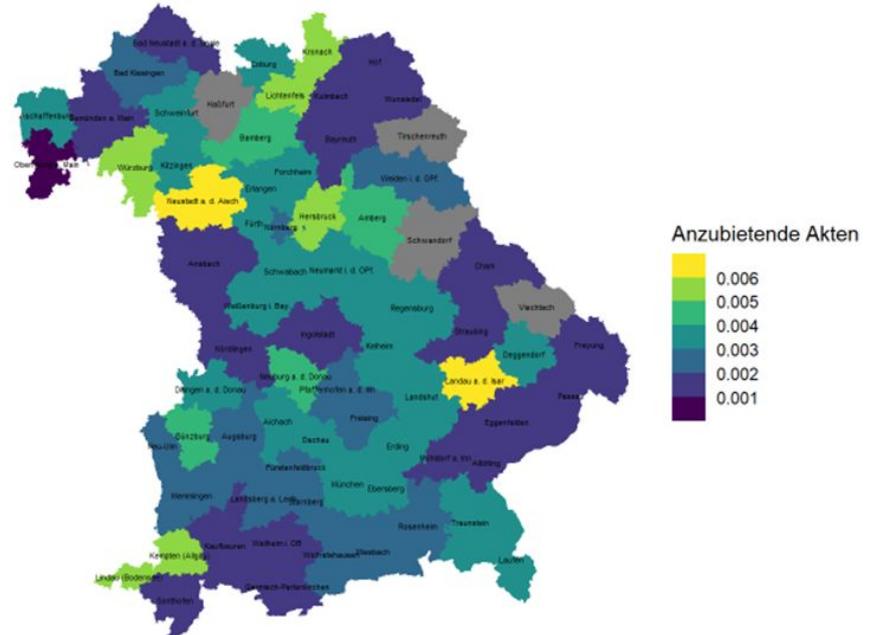
SO...



Verteilung einer Auswahl von 372 Akten (= tatsächlich als archivwürdig gekennzeichnete Akten)

# Erkenntnisse

oder so!



Verteilung einer Auswahl von 372 Akten (= einfache Stichprobenziehung)

# Erkenntnisse

- Erkennen, was in den Daten steckt – Analyse und Visualisierung der Grundgesamtheit im Vergleich zu den angebotenen Verfahren bzw. anzubietenden Verfahren nach Stichprobenziehung
- Ziehung echter Zufallsstichproben und stratifizierter Stichproben als Methode der Überlieferungsbildung (im Vergleich zur Auswahl nach Aussonderungsbekanntmachung Justiz bzw. mathematischen Stichproben) operationalisierbar
- Erhebliche Verbesserung der Bewertungsergebnisse gegenüber der bisherigen Auswahl auch und gerade im Hinblick auf künftige Auswertungen durch Nutzerinnen und Nutzer
- Reflexion auf die Zielsetzung der Überlieferungsbildung – was soll überliefert werden?
  - Nur das Besondere – welche Methoden stehen hierfür zur Verfügung?
  - auch das Allgemeine? – einfache Zufallsstichprobe
  - Oder eine Kombination aus beidem – stratifizierte Stichprobe?
  - Wie groß muss dann die Stichprobe sein? Welchen Rückschluss auf (welche) Grundgesamtheit muss die Stichprobe zulassen?

# Erkenntnisse

- Erkennen, was **nicht** in den Daten steckt
  - z. B. Beteiligung berühmter Persönlichkeiten - Verknüpfung mit Wikidata zur Identifizierung „berühmter“ Personen war mangels eindeutiger Identifikatoren in forumSTAR nicht möglich (Beispiel: Thomas Müller)
- Anpassungsbedarfe der Prozesse auf Justiz- und Archivseite
  - Datenverarbeitung (Exporte, Datentransfer, Datenauswertung)
  - Betrieb, Schulung und Anwendung eines neuen Software-Tools (R), langfristiges Wissensmanagement
  - Zeitnahe Bewertung, lange vor der tatsächlichen Aussonderung
  - Kommunikation und Umsetzung von Bewertungsergebnissen in Gerichten und Archiven
  - Anpassung der formalen und inhaltlichen Grundlage (Aussenderungsbekanntmachung Justiz)
  - Evidenz und Transparenz der Methodik für künftige Nutzerinnen und Nutzer (Erschließung)
- Fragen an die Zukunft
  - Ablösung von forumSTAR durch GeFa
  - Aussonderung von elektronischen Verfahrensakten der Justiz anhand des Standards XJustiz

# Ergebnisse

Archivierungsmodell für Zivilgerichtsverfahren der Amtsgerichte

- Zivilprozessakten: besonderer Quellenwert für Fragestellungen aus der Alltags- und Regionalgeschichte
- Überlieferungsziel: Alltägliches und Besonderes (Justizstatistik mit Leben füllen)
- Methodik – Besonderes
  - Festhalten am Archivsachenvermerk weiterhin erforderlich. Kriterien:
    - Einschaltung der Pressestelle, mediale Berichterstattung (ggf. vermerkt: Auskünfte nach Art. 4 BayPrG)
    - Bedeutende Persönlichkeiten, Firmen oder Stiftungen (ggf. wiederum mediale Berichterstattung – folgt die Bedeutung der Persönlichkeiten aus der Berichterstattung?)
    - Kennzeichnung gem. § 10 Abs. 2 Nr. 4 AktO (Eignung für Ausbildungs- und Prüfungszwecke)
    - Berücksichtigung von auffälligen Prozesshäufungen bzw. Prozesslawinen in bestimmten Sachgebieten als Indikator für zeittypische Phänomene
  - Verfahren mit einer besonders langen Dauer – Berücksichtigung im Rahmen einer Stichprobenziehung

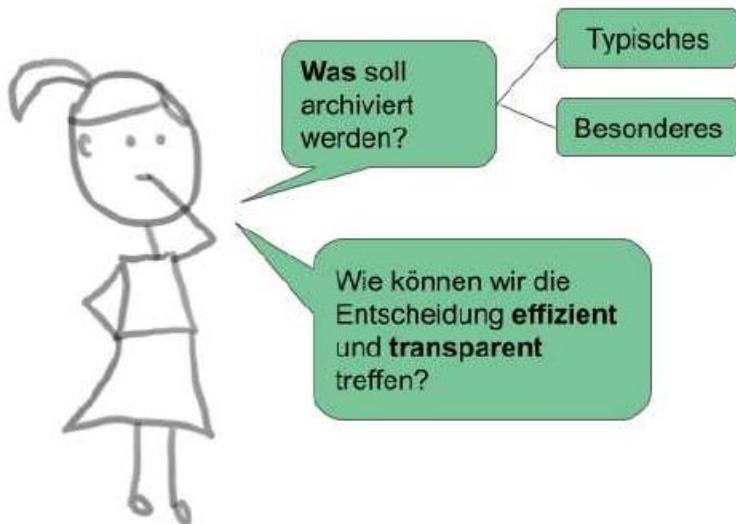
# Ergebnisse

Archivierungsmodell für Zivilgerichtsverfahren der Amtsgerichte

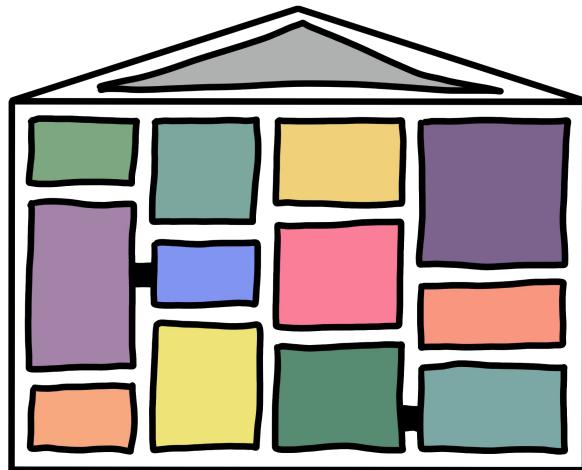
- Methodik – Typisches: Stichprobenziehung
  - Reduzierung auf Erledigungen mit Entscheidungen in der Sache (Aussagekraft)
  - Berücksichtigung aller Amtsgerichte
    - nicht Stichprobenziehung auf Ebene der einzelnen Amtsgerichte
    - Nachteil: einige Amtsgerichte werden immer wieder leer ausgehen
  - 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

# Vielen Dank für die Aufmerksamkeit!

Erprobung, Verfestigung und Erweiterung des Modells in Workshopserie 2!



# Pause



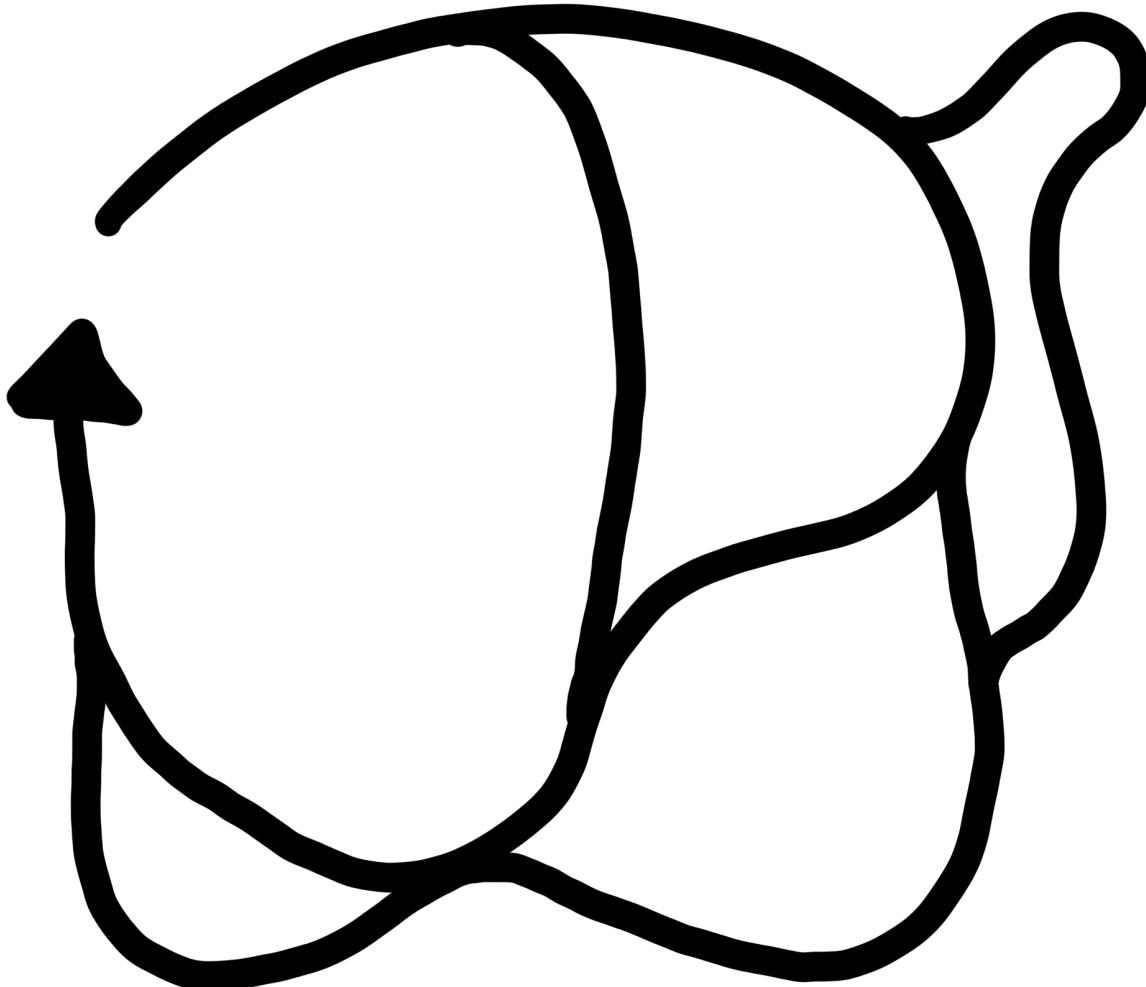
# Was wünschen Sie sich?

- Was braucht es, für einen erfolgreichen Ausgang dieses Workshops?
- Welche Gedanken haben Sie sich bereits zur Umsetzung der Stichprobenziehung gemacht?

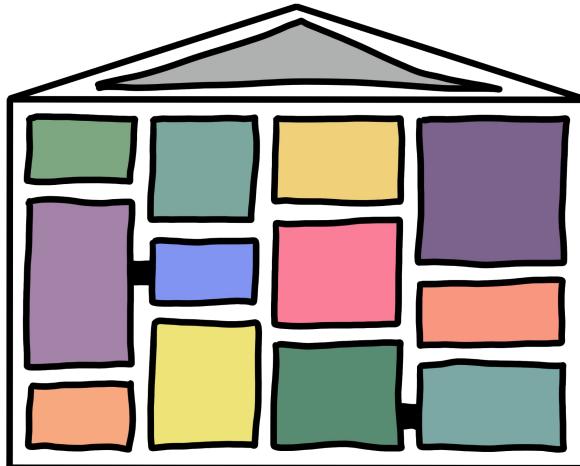
# Arbeitsabläufe:

*von der Datenübertragung  
bis zur Archivierung*

- Welche Schritte gibt es?
- Welche Personen sind in diese Schritte involviert?
- Wie ist der Stand der Dinge?



# Cloud-Plattform Check



[login.adrf.eu](http://login.adrf.eu)

- Können Sie auf die Website zugreifen?
- Können Sie sich einloggen?

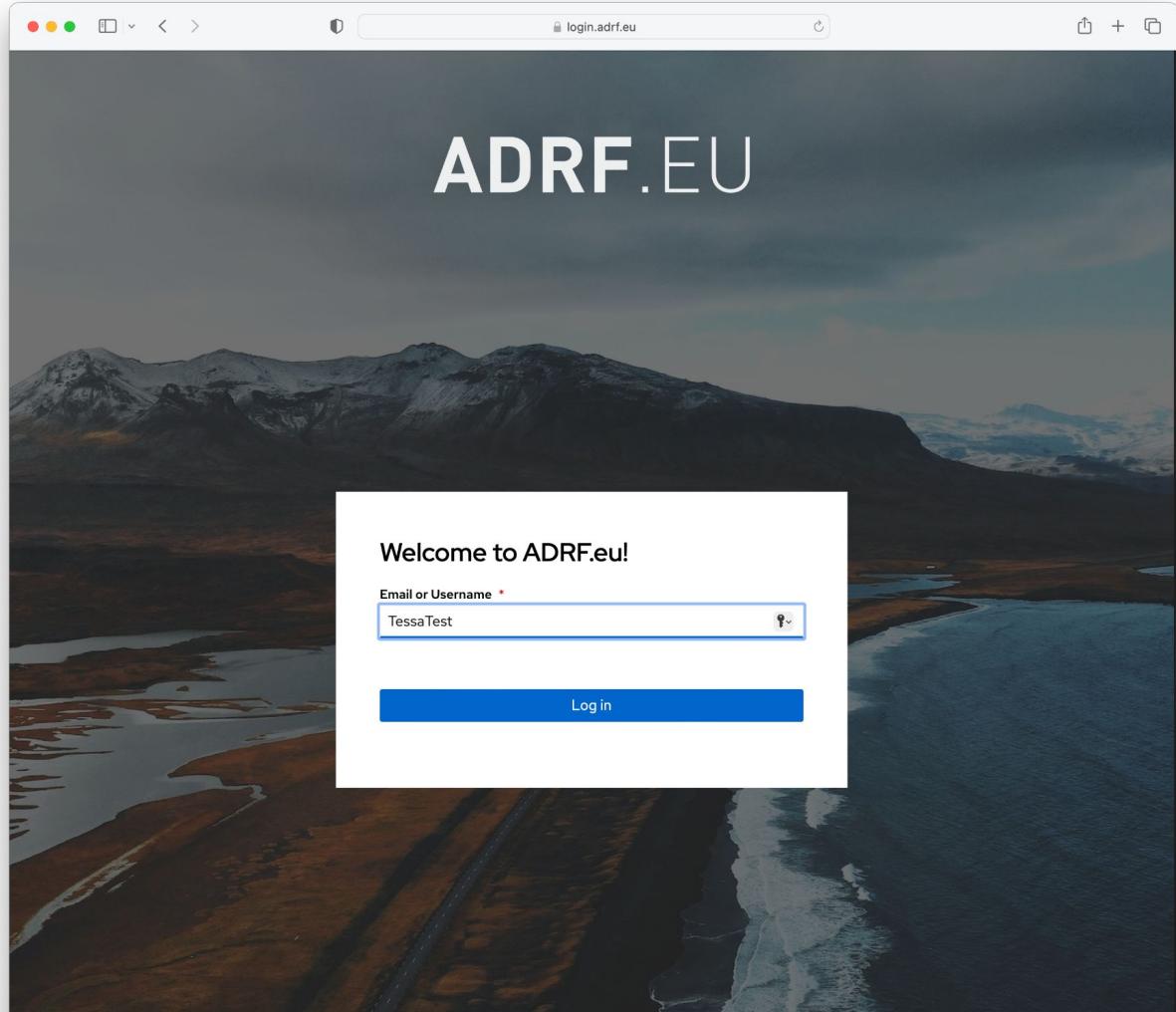
# Cloud-Plattform Check



# Login

Zugang erhalten Sie  
immer über die  
Adresse  
<https://login.adrf.eu>

Bitte loggen Sie sich  
mit dem  
**Benutzernamen**  
**VornameNachname**  
ein (TessaTestuserin)



# Zugang

Nach dem Login  
sehen Sie den  
Zugang für unseren  
Workshop.

My applications

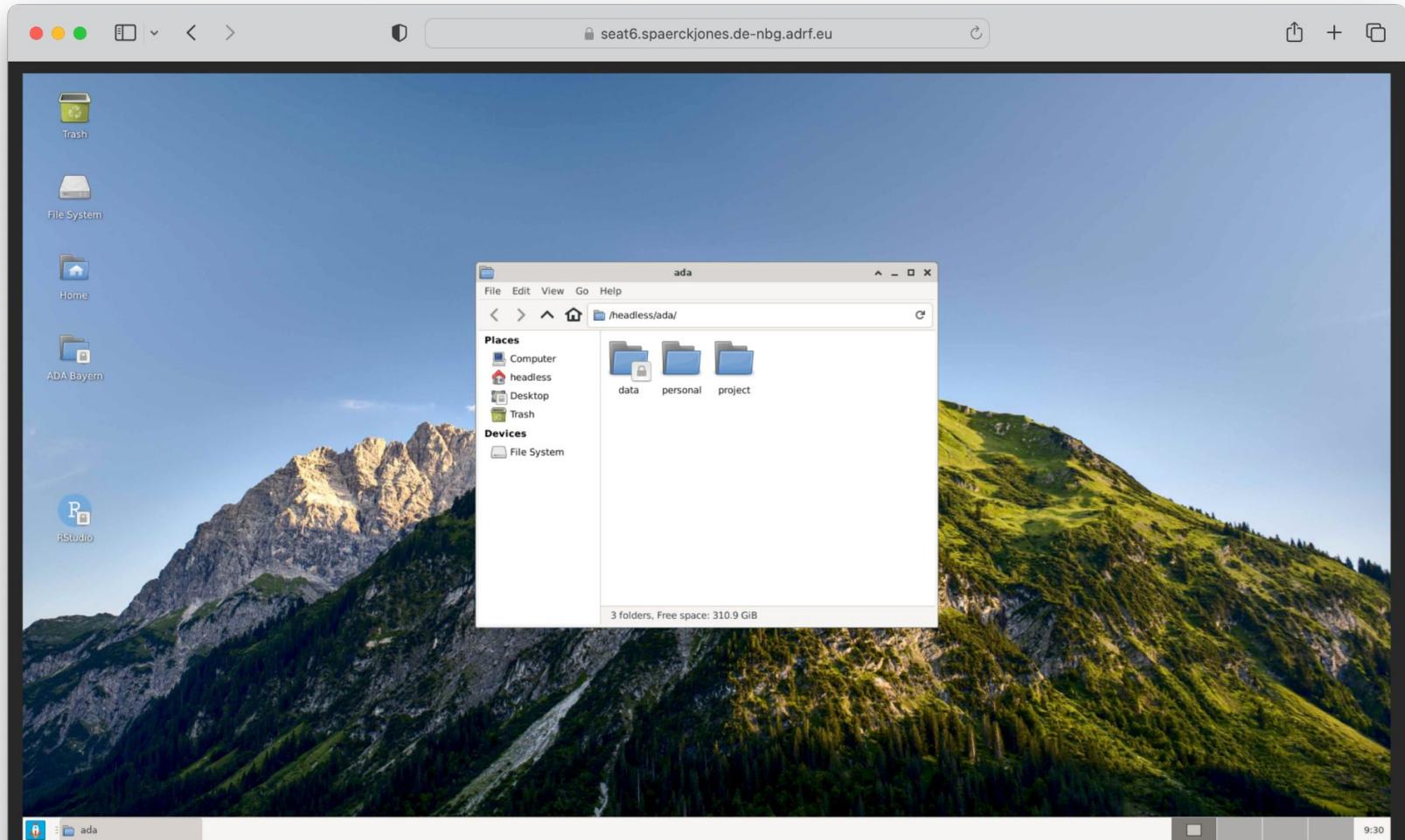
ADA Bayern – Workshop 2

A

ADA Cloudzugang

Search...





Mittags-  
Pause

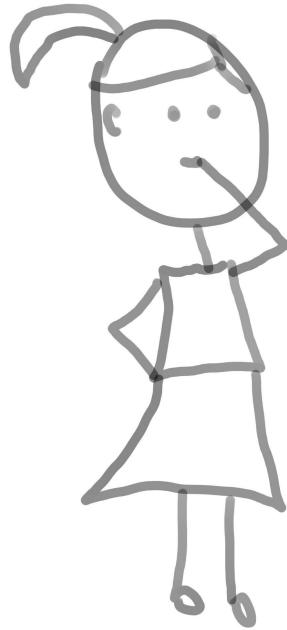


# Sichtung der bisherigen Arbeiten + Notebooks

Gruppenarbeit:

- Loggen Sie sich in die Cloud-Umgebung ein.
- Öffnen Sie den Ordner **ada** und den darin liegenden Ordner **personal**.
- Gehen Sie in Ihren persönlichen Arbeitsordner **personal** und klicken sie auf die Datei **Notebooks.Rproj** im Ordner **Notebooks**. Dadurch öffnet sich RStudio.
- Nun können Sie im Panel unten rechts auswählen, welche der Notebooks Sie sich ansehen möchten. Fokussieren Sie sich zunächst auf:
  - ADA\_Modul\_2\_1\_Erste-Analysen.qmd
  - ADA\_Modul\_3\_Best\_of\_Stichprobenziehung.qmd
- Probieren Sie gemeinsam, den Code auszuführen und die Inhalte *grob* nachzuvollziehen.
  - Data Scientist und erfahrene R-Nutzer:innen: unterstützt die anderen.
  - Alle: keine Frage ist dumm und es wird nicht erwartet, dass Sie jetzt schon alles verstehen. Es geht darum, ein erstes Gefühl für die Notebooks zu entwickeln.

Pause



*Was nehmen Sie von heute mit?*

*Welche Fragen / Punkte sollten wir morgen (mit  
unserem Stichproben-Expertin) klären?*

## **Modul 2 Umsetzung: Implementierung der Stichprobenziehung**

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.

## **Modul 3 Umsetzung: Implementierung der Stichprobenziehung im Kontext des Gesamtablaufs**

Am Ende dieses Moduls können Sie...

- ... die gewählte Strategie für Datensätze aus verschiedenen Jahren umsetzen.
- ... Lücken erkennen, die im Projekt zu schließen sind, um die Analysen in das Tagesgeschäft aufzunehmen.
- ... die Datenanalysen dokumentieren.

## **Modul 4 Nachhaltige Implementierung: Wissensweitergabe und Infrastrukturerhalt**

Am Ende dieses Moduls können Sie...

- ... einen nachhaltigen Prozess etablieren, wie die Daten übertragen, die Analysen durchgeführt und Archivierungsentscheidungen dokumentiert werden.
- ... ihr Wissen an andere weitergeben.
- ... einschätzen, wie die benötigte Infrastruktur erhalten werden kann.