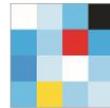


# Modul 3: Visualisieren, Entscheiden

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



**Willkommen zurück!**

# Heute

Was bisher geschah + Visualisieren	10:00 - 10:30
Besuch des Digitalministers Dr. Fabian Mehring	10:30 - 11:15
Visualisieren und Vorbereitung Nachmittag	11:15 - 11:50
Mittagspause	11:50 - 12:50
Teamarbeit	12:50 - 15:00
Pause: Selbstbestimmt nach Bedarf der Teams	
Abschluss	15:00 - 15:30

Ein Bild sagt mehr als 1000 Worte

Wie viele 3en können Sie hier zählen?

756395068473

658663037576

860372658602

846589107830

Wie viele 3en können Sie hier zählen?

756395068473

658663037576

860372658602

846589107830

# Zwei Arten von Visualisierung

1. *Explorative Grafiken* - werden verwendet um Informationen in (oft grossen) Datensätzen zu entdecken
2. *Erzählende Grafiken* - werden verwendet um Ergebnisse zu kommunizieren

Frage: Für welche dieser  muss man wissen wie es geht?

# Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816

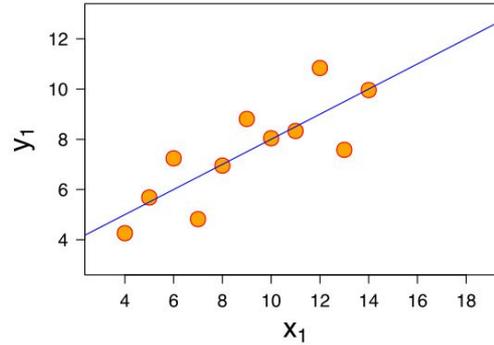
corr x2 y2 (obs=11): 0.816

corr x3 y3 (obs=11): 0.816

corr x4 y4 (obs=11): 0.816

# Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816



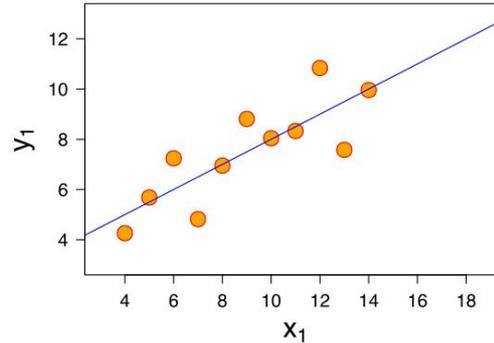
# Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816

corr x2 y2 (obs=11): 0.816

corr x3 y3 (obs=11): 0.816

corr x4 y4 (obs=11): 0.816



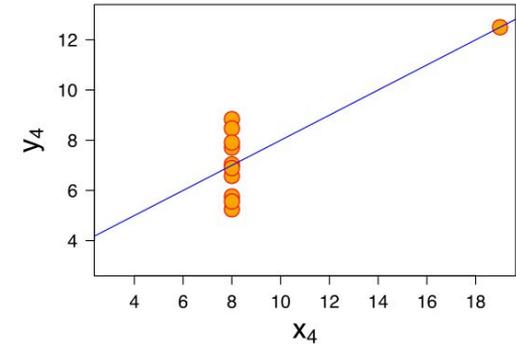
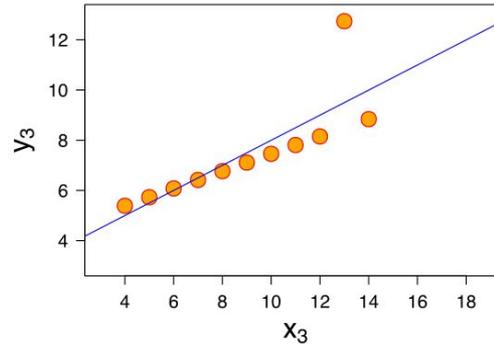
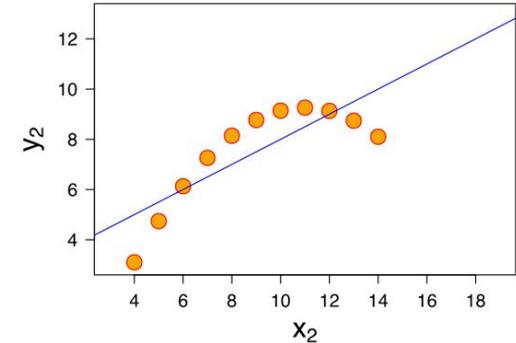
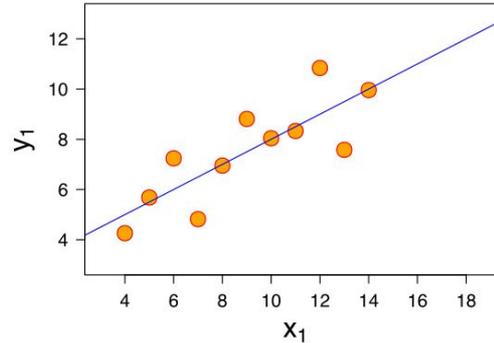
# Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816

corr x2 y2 (obs=11): 0.816

corr x3 y3 (obs=11): 0.816

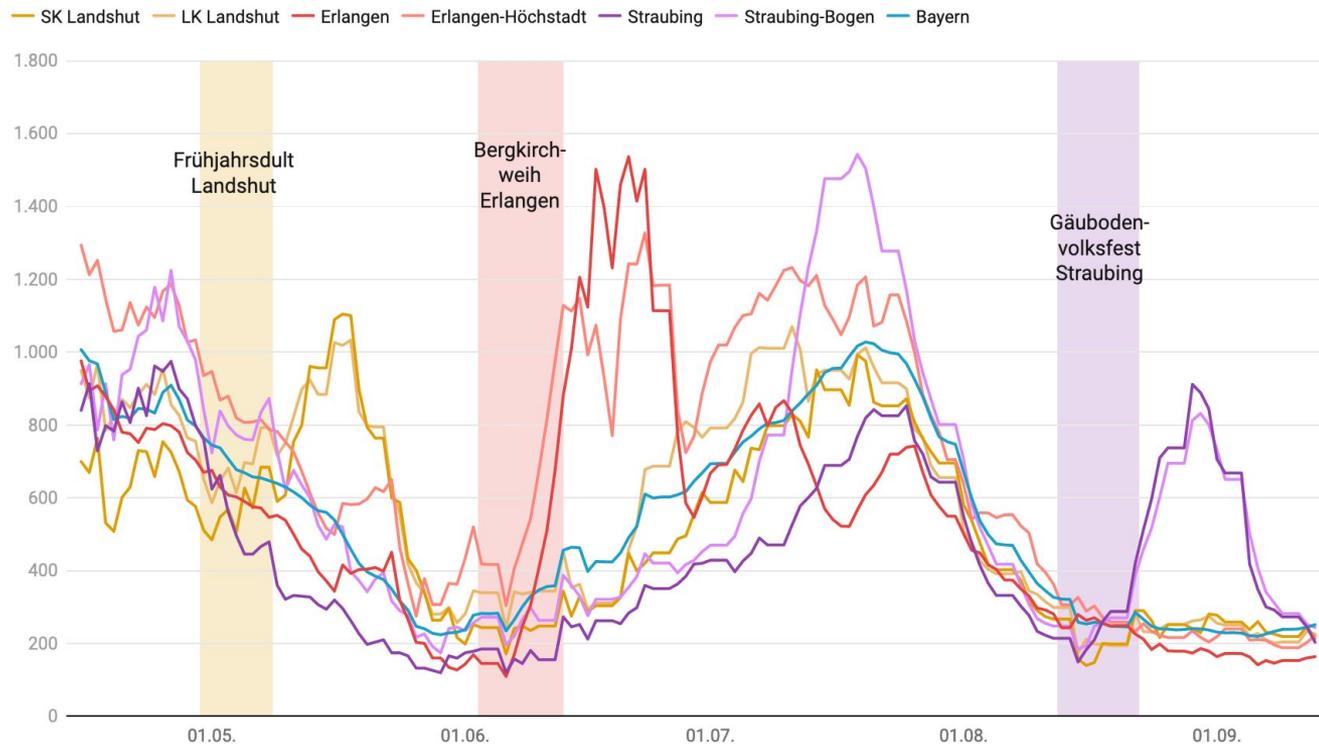
corr x4 y4 (obs=11): 0.816



# Beispiel erzählende Grafik

## 7-Tage-Inzidenzen nach Volksfesten: Deutlicher Anstieg

Gemeldete Corona-Neuinfektionen pro 100.000 Einwohner in den jeweils vergangenen 7 Tagen (Stand: 14.8.2022)



# Gefahr der Altersarmut ist für Frauen in Bayern am größten

Armutsgefährdungsquote\* der über-65-jährigen **Männer** und **Frauen**, aufgeschlüsselt nach Bundesländern im Jahr 2022

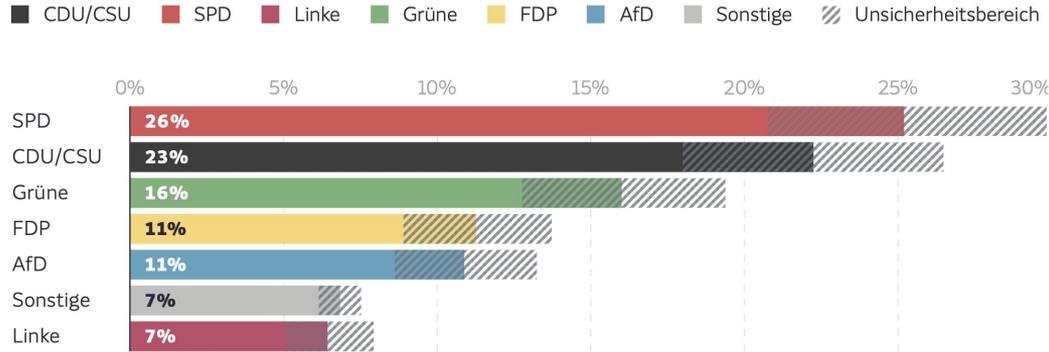


\*Nach EU-Standard: Anteil der Bevölkerung oder Bevölkerungsgruppe, dem weniger als 60% des mittleren Äquivalenzeinkommens zur Verfügung steht. Gemessen am Landesmedianeinkommen des jeweiligen Bundeslandes.

# Wahlvorhersagen sind immer unsicher

## Wahlprognose für die Bundestagswahl

in diesen Bereichen landen die Parteien im Wahlmodell mit hoher Wahrscheinlichkeit



Stand: 25.09.21 00:17 Uhr

Quelle: [Zweitstimme.org](https://www.zweitstimme.org) • [Rohdaten herunterladen](#)

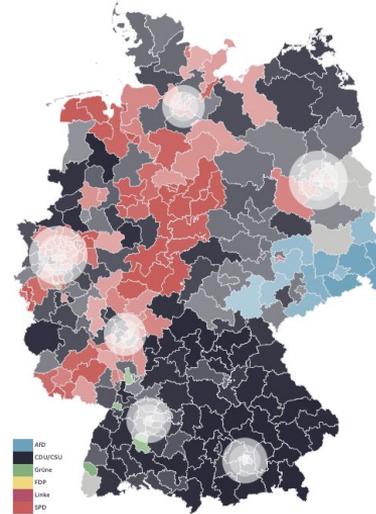
2021 Bundestagswahl

Umfragen Wahlkreise Koalitionen Kandidaten Themen

Prognose für Ihren Wahlkreis

Geben Sie Ihre PLZ ein

Übersicht über alle Wahlkreise

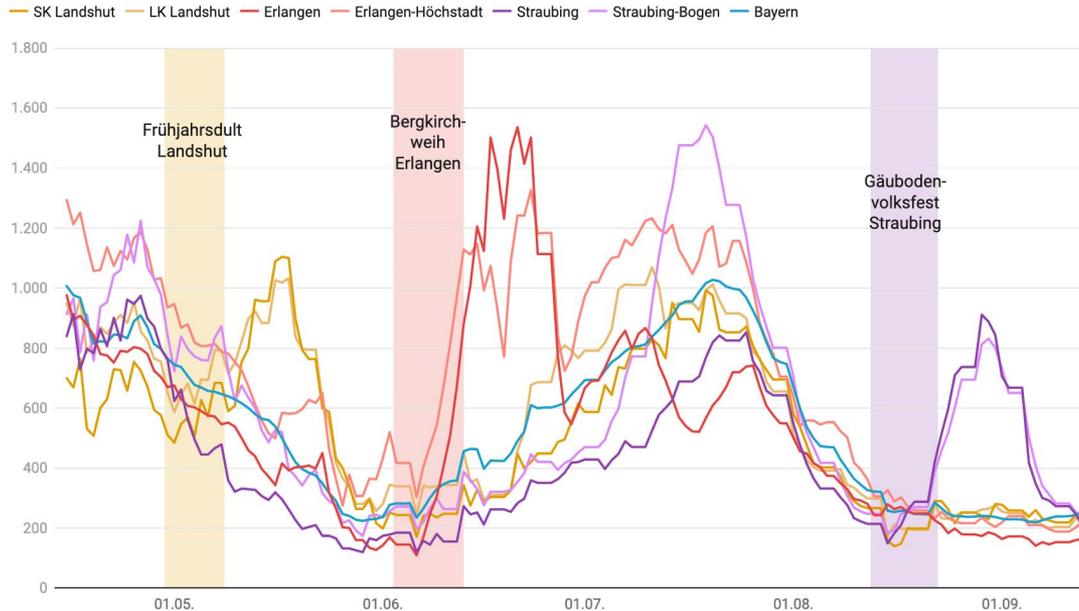


Umfragen geben einen guten Überblick, wie die politische Stimmung aktuell aussieht. Wahlprognosen dagegen fokussieren sich auf den Wahltag und betrachten neben Umfragen auch andere Variablen die das Wahlverhalten erklären. So können sie für einzelne Wahlkreise das Rennen vorhersagen. Mehr zu unserem [Wahlmodell finden Sie hier.](#)

# Was macht eine gute Grafik aus?

## 7-Tage-Inzidenzen nach Volksfesten: Deutlicher Anstieg

Gemeldete Corona-Neuinfektionen pro 100.000 Einwohner in den jeweils vergangenen 7 Tagen (Stand: 14.8.2022)



Grafik: BR • Quelle: Robert Koch-Institut (RKI)

Gutes Verhältnis: Tinte-Information

Daten verständlich machen

Gute Beschriftungen

Optimal: Erzählt eine Geschichte

Inspiziert von: Edward R. Tufte, The Visual Display of Quantitative Information

Welche Grafik?

## Abweichung

Darstellung von Abweichungen (z.B. von einem Referenzwert) ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Abweichung kann durch die Differenz der Werte oder durch die Differenz der Abweichungsdichte (z.B. die Differenz der Abweichungsdichte) dargestellt werden.

**FT Anwendungsbereiche:** Qualitätsmanagement, Medizin, Wirtschaft, Wetter, etc.



## Korrelation

Darstellung des Zusammenhangs zwischen zwei oder mehr Variablen. Die Richtung des Zusammenhangs kann durch die Steigung der Datenreihen dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Rangordnung

Die Darstellung der Reihenfolge der Datenpunkte ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Reihenfolge kann durch die Position der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Verteilung

Darstellung der Verteilung der Datenpunkte ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Verteilung kann durch die Form der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Veränderung über Zeit

Die Darstellung der Veränderung der Datenpunkte über die Zeit ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Veränderung kann durch die Steigung der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Größenordnung

Die Darstellung der Größenordnung der Datenpunkte ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Größenordnung kann durch die Größe der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Teil-zum-Ganzen

Die Darstellung des Zusammenhangs zwischen zwei oder mehreren Variablen ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Der Zusammenhang kann durch die Steigung der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Räumlich

Die Darstellung der räumlichen Verteilung der Datenpunkte ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die räumliche Verteilung kann durch die Position der Datenpunkte dargestellt werden.

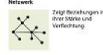
**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



## Strömung

Die Darstellung der Strömung der Datenpunkte ist die Regel für den Vergleich von zwei oder mehreren Datenreihen. Die Strömung kann durch die Richtung der Datenpunkte dargestellt werden.

**FT Anwendungsbereiche:** Medizin, Wirtschaft, Wetter, etc.



# Visueller Wortschatz

Gestalten mit Daten

Es gibt viele Möglichkeiten, Daten zu visualisieren. Aber wie entscheiden wir, welche wir wählen sollen? Verwenden Sie die Kategorien oben, um die wichtigste Datenbeziehung für Ihre Story auszuheben. Schauen Sie sich dann die verschiedenen Diagrammtypen in der Kategorie an, um erste Ideen zu entwickeln, was am besten funktionieren könnte. Diese Liste ist kein Assistent, sondern ein nützlicher Ausgangspunkt für aussagekräftige Datenvisualisierungen.

ft.com/vocabulary



# Chart Doctor auf Deutsch (von Max Lang für ADA Bayern)

<https://maxmlang.shinyapps.io/shiny-chart-doc/>



## Strömung

Zeigt dem Betrachter den Umfang oder die Intensität der Bewegung zwischen zwei oder mehreren Zuständen oder Be- und Abfolgen von Sequenzen oder geografische Orte handeln.

Anwendungsbeispiele

Geldverkehr, Handel, Migranten, Rechtsstreitigkeiten, Informationen; Beziehungsdiagramme.

Sankey Diagramm



Zeigt Veränderungen in den Strömen von einer Bedingung zu mindestens einer anderen; gut geeignet, um das Endergebnis eines komplexen Prozesses zu verfolgen.

Code Beispiele



Wasserfall Diagramm



Entwickelt, um die Abfolge von Daten durch einen Flussprozess zu zeigen, typischerweise Budgets. Kann +/- Komponenten enthalten.

Code Beispiele



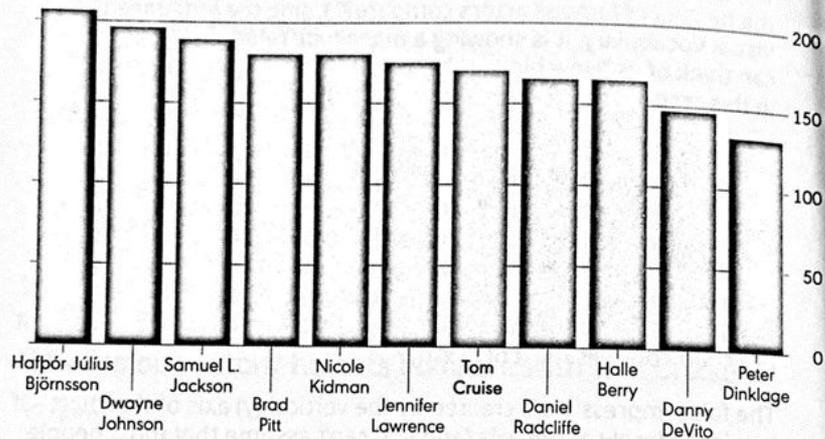
Ergebnisse kommunizieren



*The Audience*  
(Hand drawn digitally  
coloured print, 2018)  
Fil OK Art

## How famous actors' heights compare with Tom Cruise

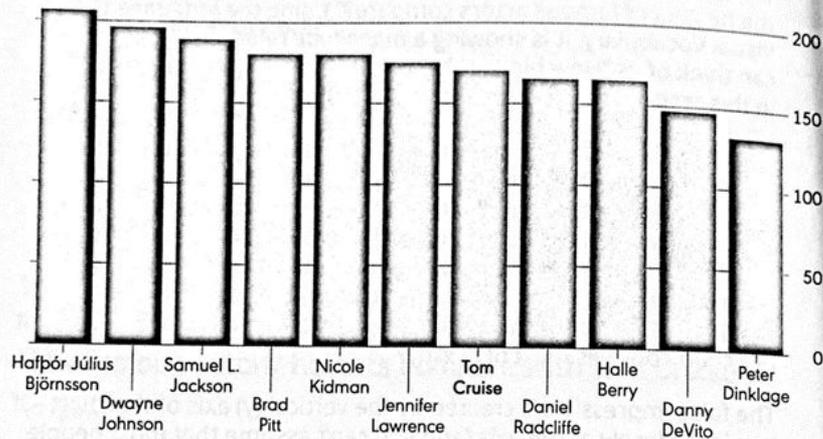
Height (cm)



Sources: Google, IMDB. Used by permission from Alan Smith.

## How famous actors' heights compare with Tom Cruise

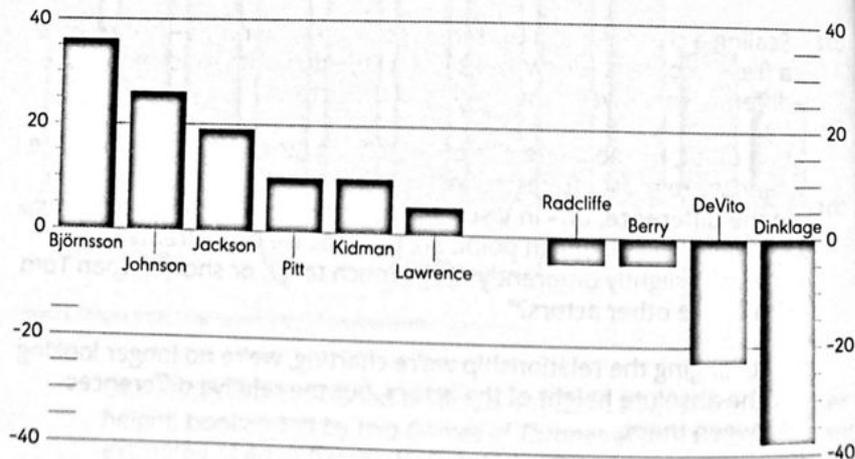
Height (cm)



Sources: Google, IMDB. Used by permission from Alan Smith.

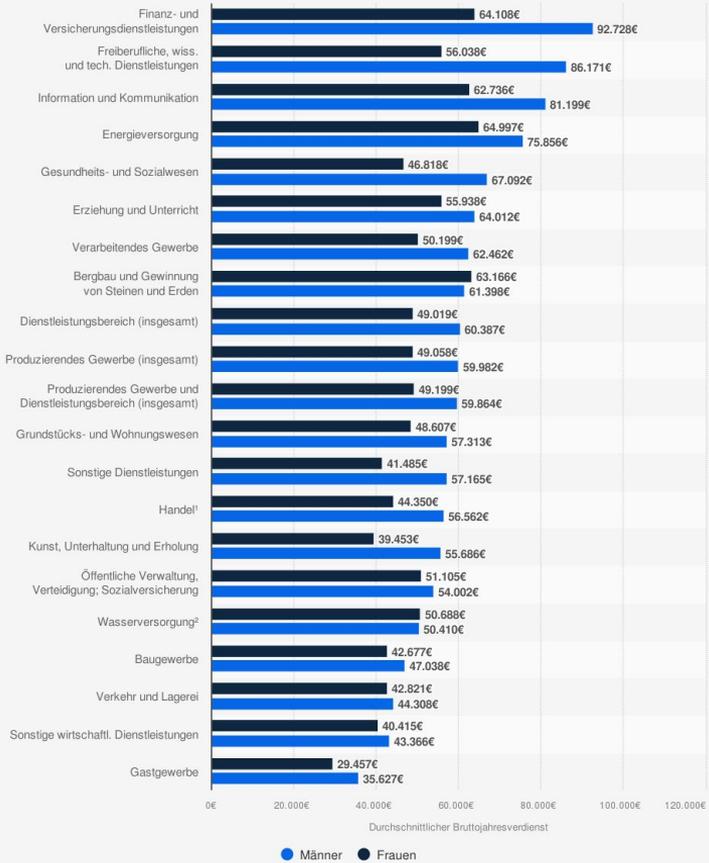
## How famous actors' heights compare with Tom Cruise

Amount taller/shorter than Cruise (cm)



Sources: Google, IMDB. Used by permission from Alan Smith.

## Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen) nach Wirtschaftsbereichen und Geschlecht im Jahr 2022

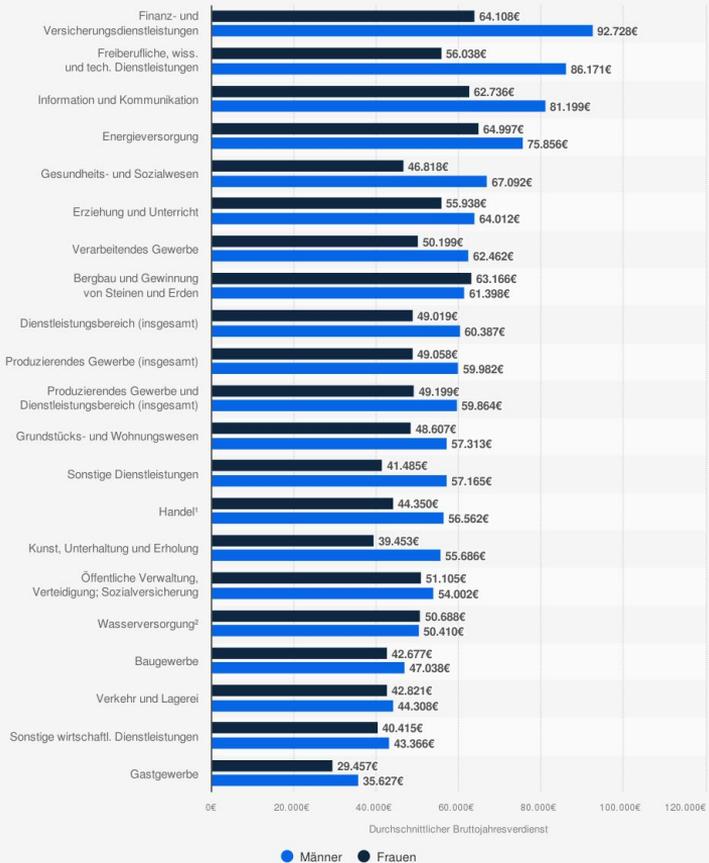


Quelle:  
Statistisches Bundesamt  
© Statista 2023

### Weitere Informationen:

Deutschland; 2022; Mit Sonderzahlungen

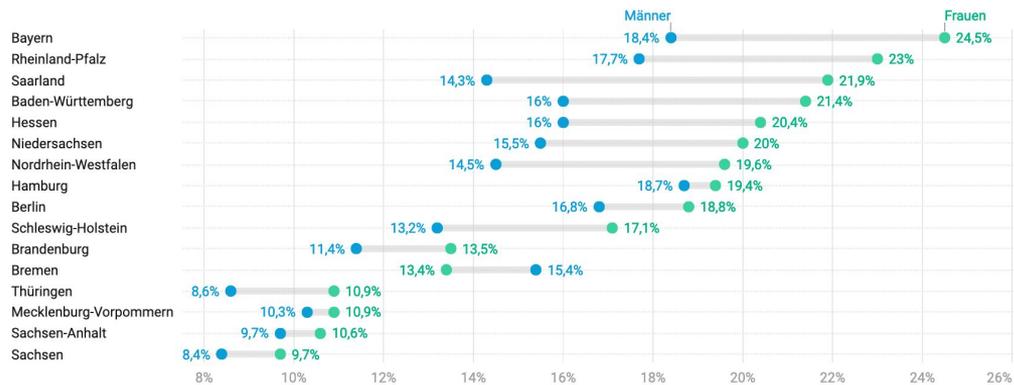
## Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen) nach Wirtschaftsbereichen und Geschlecht im Jahr 2022



Quelle: Statistisches Bundesamt © Statista 2023

Weitere Informationen:  
Deutschland; 2022; Mit Sonderzahlungen

## Armutsgefährdungsquote\* der über-65-jährigen Männer und Frauen, aufgeschlüsselt nach Bundesländern im Jahr 2022



\*Nach EU-Standard: Anteil der Bevölkerung oder Bevölkerungsgruppe, dem weniger als 60% des mittleren Äquivalenzeinkommens zur Verfügung steht. Gemessen am Landesmedianeinkommen des jeweiligen Bundeslandes.

Grafik: BR24 · Quelle: Statistische Ämter des Bundes und der Länder





# Robuste Farbpaletten helfen bei der Kommunikation



**Besuch des Staatsministers für Digitales**  
*Dr. Fabian Mehring*

**Vorstellung der Aktivitäten im Bereich Data Literacy  
und Digitalisierung der öffentlichen Verwaltung**

*Prof. Dr. Frauke Kreuter*

# Lauffeuer der Digitalisierung in der öffentlichen Verwaltung



COLERIDGE INITIATIVE

Capabilities ▾ Collaborations ▾ News & Events ▾

## Building Bridges, Breaking Barriers: Data Collaboration for the Public Good

Coleridge Initiative's 4th Annual National Convening was held on **March 18th - 20th, 2024** at the Crystal Gateway Marriott in Arlington, VA. Attendees gained insights into how leaders in the field are pushing the boundaries of traditional governance, forging data partnerships, and reshaping the future of public administration.

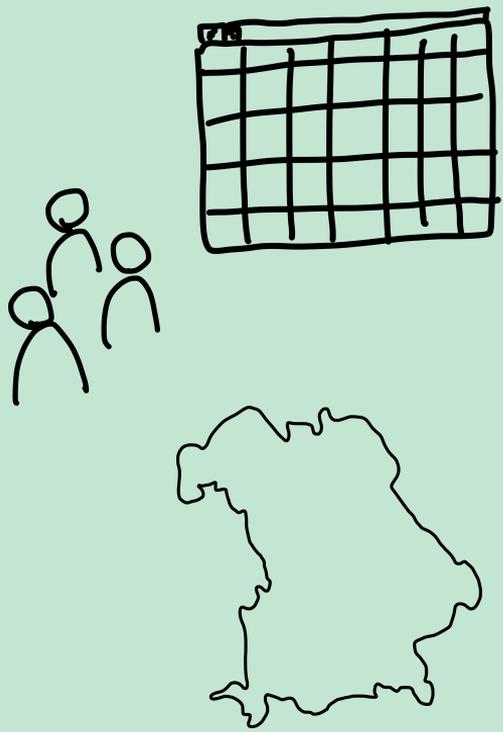
Quarter: 2016-Q4  
Total Organizations: 0  
Total Participants: 0



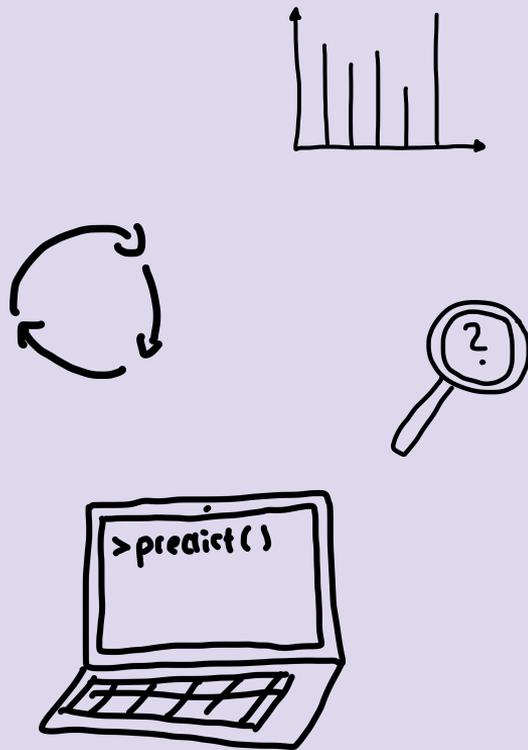
Number of Organizations  0  1-5  5-10  10+

Number of Participants  5  10  15  20

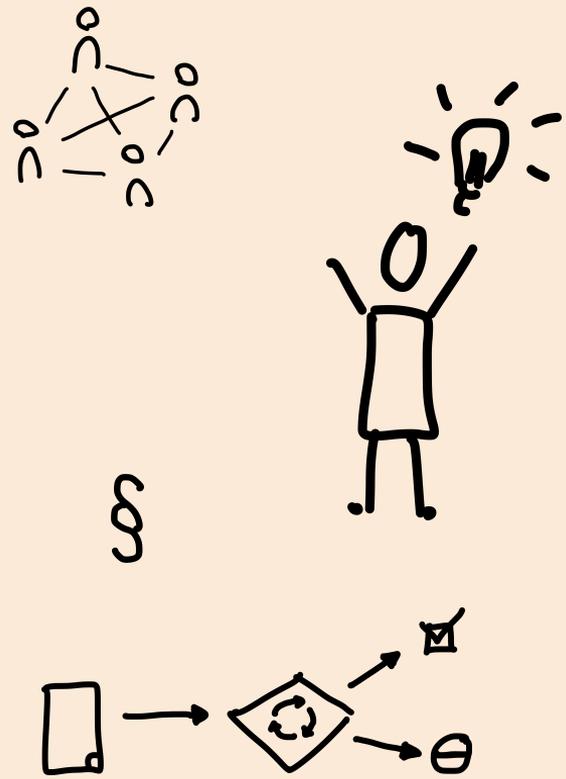
Daten



Inhalte



Ergebnisse





## Cloud-Plattform

Räume = Projekte

Gänge = Data-Sharing

- Rechenzentrum in D
- Virtuelle Desktops ohne Verbindung zum Internet, Zwei-Faktor-Authentifizierung
- Individuell und kollaborativ nutzbar
- Standort- und Anbieterunabhängige Technologie

ADA Bayern

Workshop Serie 1: ▾

Daten-basierte  
Archivierung von  
Gerichtsakten

- 1 Problemerkfassung  
und Daten als Lösung
- 2 Cloud-Computing  
und Datenanalyse
- 3 Visualisierung und  
Reporting
- 4  
Ergebnispräsentation  
und Umsetzung

Workshop Serie 2: ▾

Implementation  
Daten-basierter

## Angewandte Datenanalyse für die öffentliche Verwaltung

### ADA Bayern



Fragen beantworten mit  
Daten

Berührungslängste mit  
Cloud-Computing  
abbauen

Zusammenarbeit mit  
Data Scientists / IT  
vereinfachen

*Eine Weiterbildung für Personen in der öffentlichen Verwaltung,  
die Interesse an Digitalisierung haben.*



# Data Science for the Social Good Summer Program

- 2 Monate im Sommer
- 2 Teams mit 4-5 Fellows, alle arbeiten an einem Social Good Projekt
- Unterstützt von technischen Consultants und Projektmanagern



Bayerisches Staatsministerium  
für Digitales



**mcmml**  
Munich Center for Machine Learning





Die Feuerwehr benötigte eine Lösung um die Reichweite von Schläuchen verschiedener Längen abzuschätzen unter Einbezug von Komplexitäten wie Treppen, Gebäude und Straßenbreite.

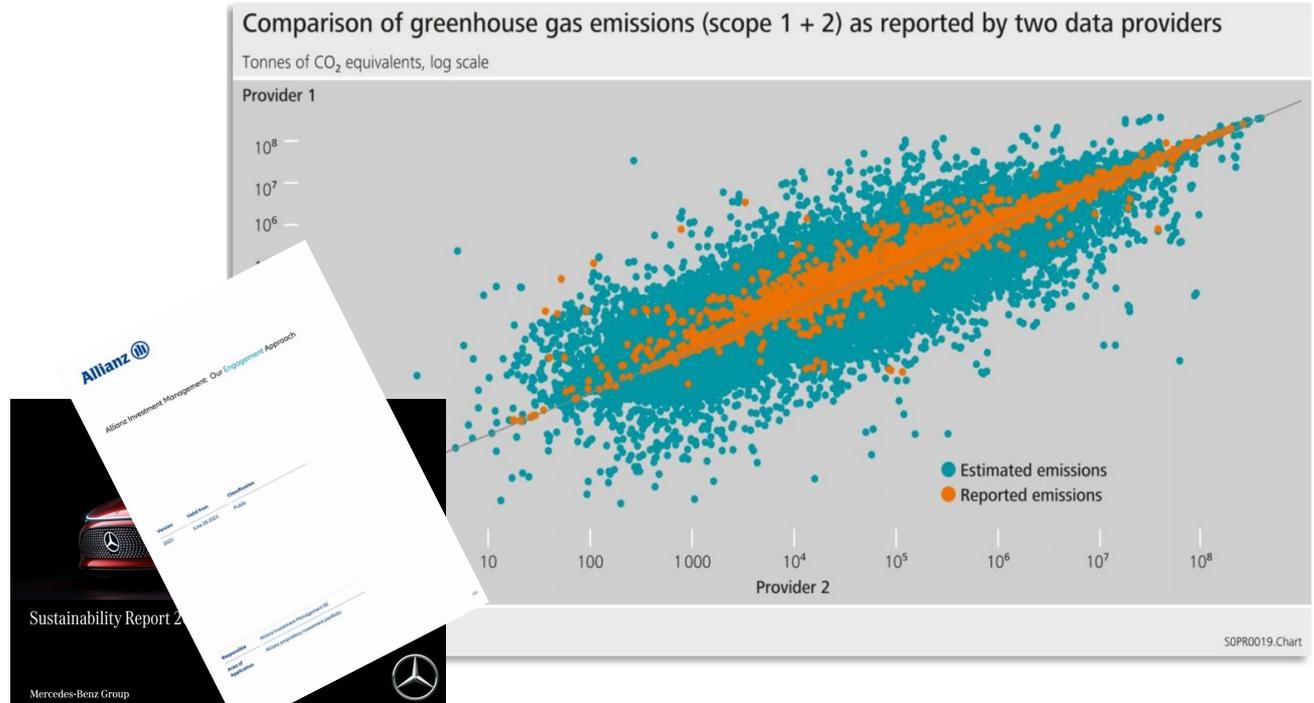
Die DSSGx Fellows haben eine Web-Applikation entwickelt, welche es ermöglicht die von einem Hydranten erreichbaren Regionen für variable Schlauchlängen und Transportmodi anzuzeigen.



Bundesamt  
für Bevölkerungsschutz  
und Katastrophenhilfe



# GIST: Greenhouse Gas Insights and Sustainability Tracking



**Vorstellung *ADA Bayern:*  
Angewandte Datenanalyse für die  
öffentliche Verwaltung**  
*Dr. Heidi Seibold*

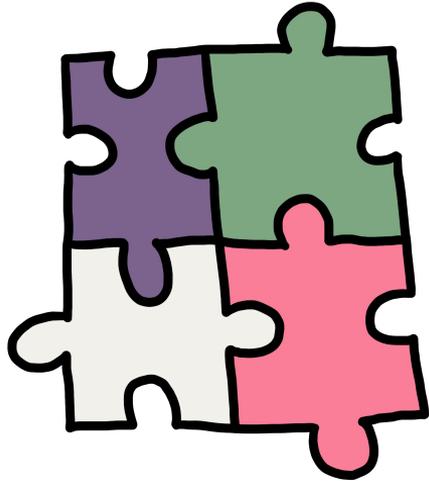


Probleme  
Lösen



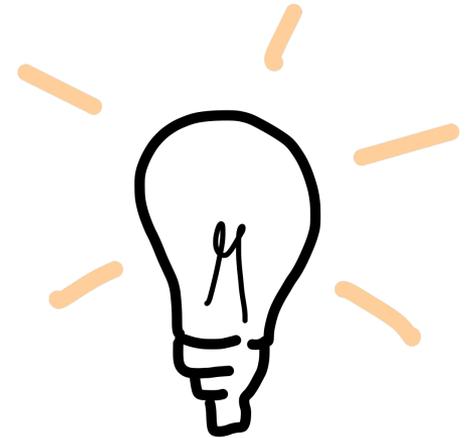
Cloud

# ADA Bayern



Teamarbeit

Gemeinsam  
Lernen



**Februar 2023**

Start der  
Gespräche

**Februar 2024**

Folgeworkshop  
mit Implementierung

**Zukunft**

Workshops mit  
anderen Behörden

**Oktober 2023**

Workshopserie  
mit Justiz und  
Archiven

**April 2024**

3. Workshopserie:  
Wir gehen in die  
Breite



# Es entstehen...

- ★ Netzwerke
- ★ Digitales Mindset
- ★ Weiternutzbare Lernmaterialien
- ★ App für Stichprobenziehung
- ★ Veröffentlichungen
- ★ ...

# Thema im laufenden Workshop



# Teamarbeit



**Team 1:** Digitale Dokumentation der Stichprobenziehung

**Team 2:** Datenqualität und Data Stewardship für digitale Akten

**Team 3:** Übergang zur digitalen Akte gestalten

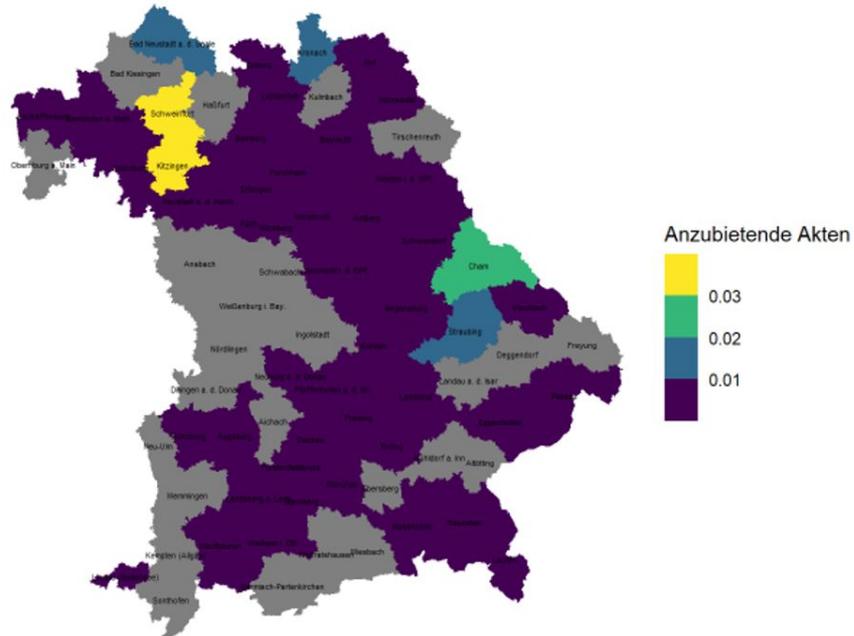
**Vorstellung der Ergebnisse aus den  
*ADA Bayern Workshops*  
*Dr. Michael Unger & Andreas Nestl***

# Workshopreihe ADA Bayern – Erkenntnisse und Ergebnisse

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

# Erkenntnisse

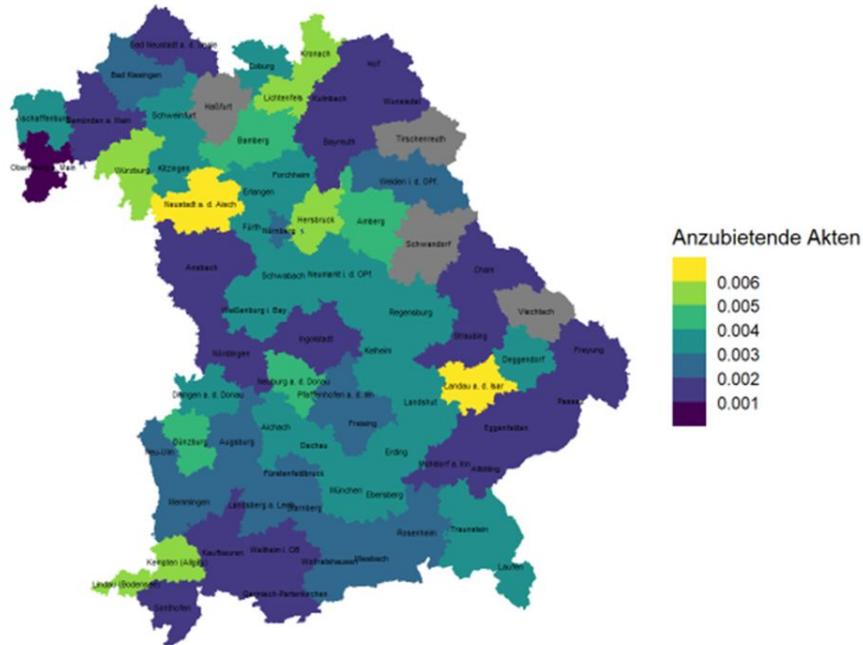
SO...



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

# Erkenntnisse

oder so!



Verteilung einer Auswahl von 372 Akten (= einfache Stichprobenziehung)

# Erkenntnisse

## **Interdisziplinärer Austausch** zwischen

- Data Scientists
- Datenspezialisten der Justiz
- Archivarinnen und Archivaren
- Archivreferendarinnen und -referendaren
- Studierenden

und Workshop-Format konstruktiv und produktiv

Verbesserte **Cloud-Umgebung** als geeignete Arbeitsumgebung

**Methoden der Datenanalyse** für archivische Zwecke praktikabel

# Erkenntnisse

**Erkennen, was in den Daten steckt** – Analyse und Visualisierung der Grundgesamtheit im Vergleich zu den angebotenen Verfahren bzw. anzubietenden Verfahren nach Stichprobenziehung

Echte Zufallsstichproben und stratifizierter **Stichproben sind operationalisierbar**

(im Vergleich zur Auswahl nach Aussonderungsbekanntmachung Justiz oder gängigen Stichprobenverfahren)

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?

Methodik, Prozessen und Kommunikationsszenarien **neu denken und konkretisieren**

**Verbesserung der Bewertungsergebnisse** gegenüber der bisherigen Auswahl auch und gerade im Hinblick auf künftige Auswertungen durch Nutzerinnen und Nutzer

# Erkenntnisse

Toolgestützte Lösung zur Anwendung von *R* ist ergonomisch und – in Grenzen – generisch

Ausweitung der Anwendung auf vergleichbare Nutzungsszenarien ist möglich und erwünscht

# Ergebnisse

Tool zur Bewertung auf der Grundlage von *R*



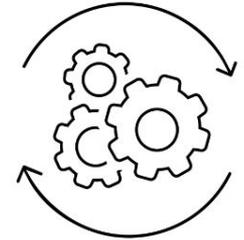
Nachhaltige Aufbereitung für Ausbildungs- und Fortbildungszwecke



# Ergebnisse

## Anpassungsbedarfe auf **Justiz- und Archivseite**

- **Datenverarbeitung** (Exporte, Datentransfer, Datenauswertung)
- Betrieb, Schulung und Anwendung eines neuen **Software-Tools (R)**, langfristiges **Wissensmanagement**
- **Zeitnahe und zentrale Bewertung**
- Kommunikation und **Umsetzung von Bewertungsergebnissen** in Gerichten und Archiven
- Anpassung der **formalen und inhaltlichen Grundlage** (Aussonderungsbekanntmachung Justiz)
- Transparenz für künftige **Nutzerinnen und Nutzer** (Erschließung)



**=> Digitale Transformation statt bloßer Digitalisierung!**

## Fragen an die Zukunft

- Andere Fachverfahren, andere Daten? (GeFa)
- Aussonderung von elektronischen Verfahrensakten der Justiz anhand des Standards XJustiz?
- Integration in unsere Standard-Aussonderungsclients?
- KI?

## Archivierungsmodell für Zivilgerichtsverfahren der bayerischen AGe

# Archivierungsmodell für Zivilgerichtsverfahren der AGE

Zivilprozessakten: Quellenwert für **Alltags- und Regionalgeschichte**

Überlieferungsziel: **Alltägliches** und Besonderes (Justizstatistik mit Leben füllen)

Grundgesamtheit: alle **erledigten** Verfahren in einem Jahr (Gleichlauf der Überlieferung mit der Justizstatistik)

## Methodik – **Besonderes**

- **Archivsachenvermerk** durch Mitarbeiter der Justiz. Allgemein über alle Aktengruppen der Justiz oder konkret für einzelne Aktengruppen, z.B. C-Verfahren?
- Verfahren mit hoher Anzahl an Terminen vor dem Richter – Berücksichtigung im Rahmen einer Stichprobenziehung (Aktendicke) – **keine** Kennzeichnung erforderlich!

# Archivierungsmodell für Zivilgerichtsverfahren der AGe

Methodik - **Typisches**: Stichprobenziehung

- **Reduzierung** der Grundgesamtheit auf Verfahren mit inhaltlichem Aussagegehalt  
=> Weglassen von Erledigungsarten, die keine Sachentscheidung erwarten lassen
- **Jährliche** Stichprobenziehung, verteilt über alle Amtsgerichte
  - **keine** Stichprobenziehung auf Ebene der **einzelnen Amtsgerichte**
  - **keine** Stichprobenziehung auf Ebene der **einzelnen Staatsarchive**
  - Nachteil: einige Amtsgerichte werden immer wieder leer ausgehen, die Stichprobe verteilt sich „ungleich“ über alle Staatsarchive (aber: Ausgleich mit Zeitablauf)
  - Vorteil: vertretbare Menge, jeder Jahrgang wird berücksichtigt
- Stichprobengröße: 400
- Berücksichtigung zweier Schichten (Anzahl an Terminen und Sachgebietsgruppen)

# Archivierungsmodell für Zivilgerichtsverfahren der AGe

**Erste Schicht:** Gewichtung der Stichprobenziehung nach **Anzahl der Termine** (in längeren Verfahren erfahren wir mehr über den Lebenssachverhalt)

- Auswahlwahrscheinlichkeit 1 bei **20 oder mehr Terminen** – da übernehmen wir alle Verfahren (insgesamt wohl nicht mehr als 10% der Stichprobengröße)
- Auswahlwahrscheinlichkeit bei **10 bis 20 Terminen** – sollen überproportional vertreten sein (mind. 20% der Stichprobengröße)
- **0 bis 10 Termine** – Rest der Stichprobengröße

# Archivierungsmodell für Zivilgerichtsverfahren der AGe

**Zweite Schicht:** Verteilung der Stichprobenziehung über **Sachgebietsgruppen** (gleiche Auswahlwahrscheinlichkeit für jede Gruppe)

- **Verkehrsunfallsachen** (sehr viele Verfahren)
- **Körper und Person** (Eingriff in die körperliche Unversehrtheit, besonders schwerwiegende Fälle)
  - Arzthaftungssachen
  - Körperverletzung
- **Leben und Wohnen** (Zusammenleben)
  - Nachbarschaftssachen
  - Wohnungsmietsachen
  - Sonstige Mietsachen
  - Wohnungseigentumssachen
- **Wirtschaftsrecht**
  - Bau-/Architektensachen (ohne Architektenhonorarsachen)
  - Kaufsachen
  - Reisevertragsachen
  - Kredit-/Leasingsachen
  - Ansprüche aus Versicherungsverträgen (ohne Verkehrsunfallsachen)
  - Gesellschaftsrechtliche Streitigkeiten
  - Honorarforderungen von Personen, für die eine besondere Honorarordnung gilt
- **Sonstiges**
  - Schuldrechtsanpassungs- und Bodenrechtssachen der neuen Länder (ganz wenig)
  - Sonstiger Verfahrensgegenstand

# Vielen Dank für die Aufmerksamkeit!



**Grußwort des Staatsministers für Digitales**  
*Dr. Fabian Mehring*

# Teamarbeit

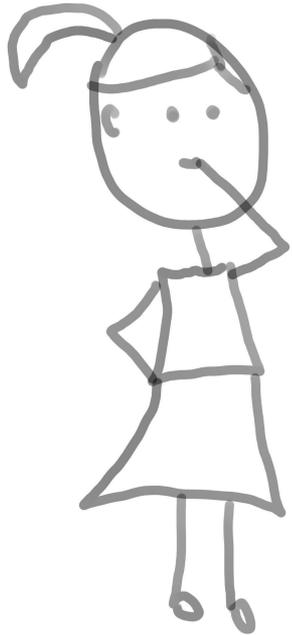


**Team 1:** Digitale Dokumentation der Stichprobenziehung

**Team 2:** Datenqualität für digitale Akten (*Datengestützte Autopsie von digitalen Akten*)

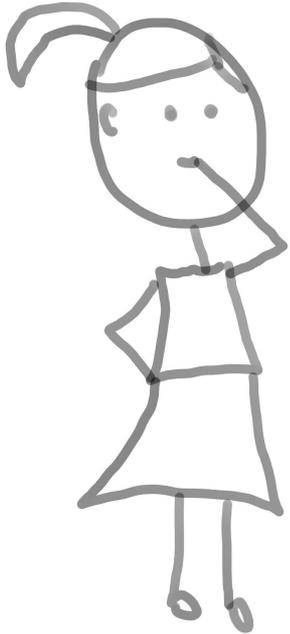
**Team 3:** Automatisierte Archivierungsentscheidung für “das Besondere” + geeignete Formatstandards

# Feedback



*Bitte fülle unser Feedback-Formular aus*

# Abschlussrunde



*Hat sich dein grüner Zettel erfüllt?  
Konnten wir deinen roten Zettel abwenden?*

*Was war **merk-würdig**?*

# Wie geht es weiter?

## **Gruppenarbeitstermine**

Online (Zoom)

- 17.04. 12:00 - 13:30
- 24.04. 12:00 - 13:30

## **Abschlusstermin**

Hybrid Zoom/LMU

30.04. 10:00 - 15:30

# Abschlussstermin 30.04.

Letzte Diskussionen der Teams (nur Studierende an der LMU)	10:00 - 12:00
Mittagspause (kein Catering)	12:00 - 13:30
Eröffnung der Abschlussvorträge (hybrid Zoom/LMU)	13:30 - 13:45
Vorträge der Gruppen (hybrid Zoom/LMU)	13:45 - 14:45
Abschluss	14:45 - 15:00