

# Modul 3: Nachvollziehbares Reporting, Visualisierung und Dashboards

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!

Online + hier vor Ort

Info: Gruppen bleiben, auch wenn manche heute online sind, die die letzten Male vor Ort waren.

Vortrag: Was bisher geschah + Visualisieren	10:00 - 10:45
Pause	10:45 - 10:55
Teamarbeit: Datenanalyse	10:55 - 12:00
Mittagspause	12:00 - 12:45
Vortrag: Reporting + Export aus der Cloud	12:45 - 13:10
Teamarbeit: Projekt festzurren, Export von Grafiken/Ergebnissen	13:10 - 14:45
Pause: Selbstbestimmt nach Bedarf der Teams	
Vortrag: Tagesabschluss	14:45 - 15:00

# Rückblick

- Was ist Ihnen vom letzten mal besonders in Erinnerung geblieben?
- Wo haben Sie letztes mal aufgehört? Wo müssen wir wieder einsteigen?
- Wie viel Zeit brauchen wir für die Fertigstellung der Aufgaben von Modul 2?

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

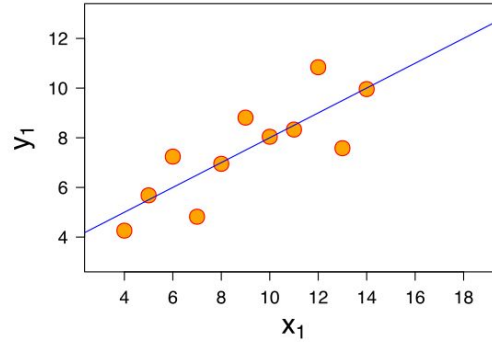
corr x2 y2 (obs=11): 0.816

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



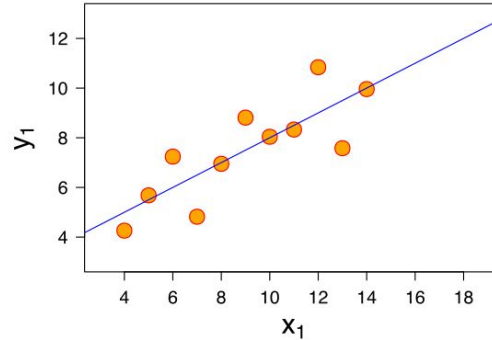
# Beispiel Explorative Grafik - Beispiel Korrelationen

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



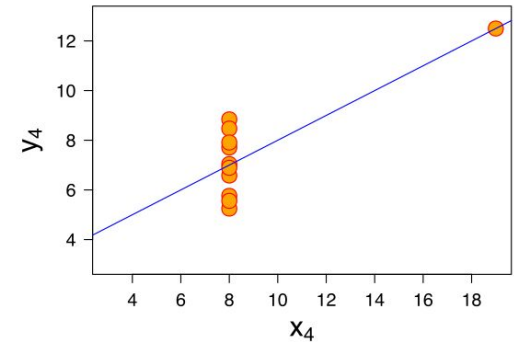
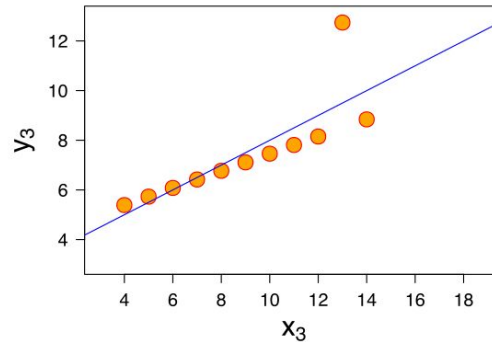
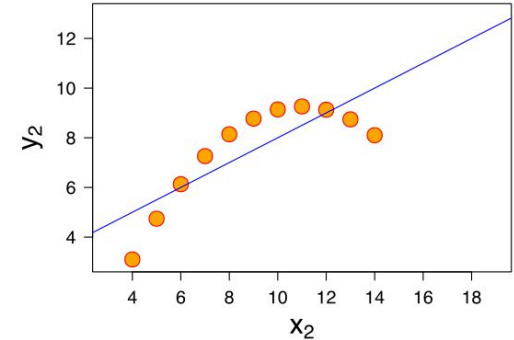
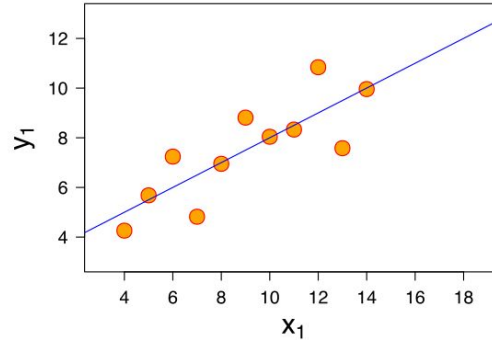
# Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816

corr x2 y2 (obs=11): 0.816

corr x3 y3 (obs=11): 0.816

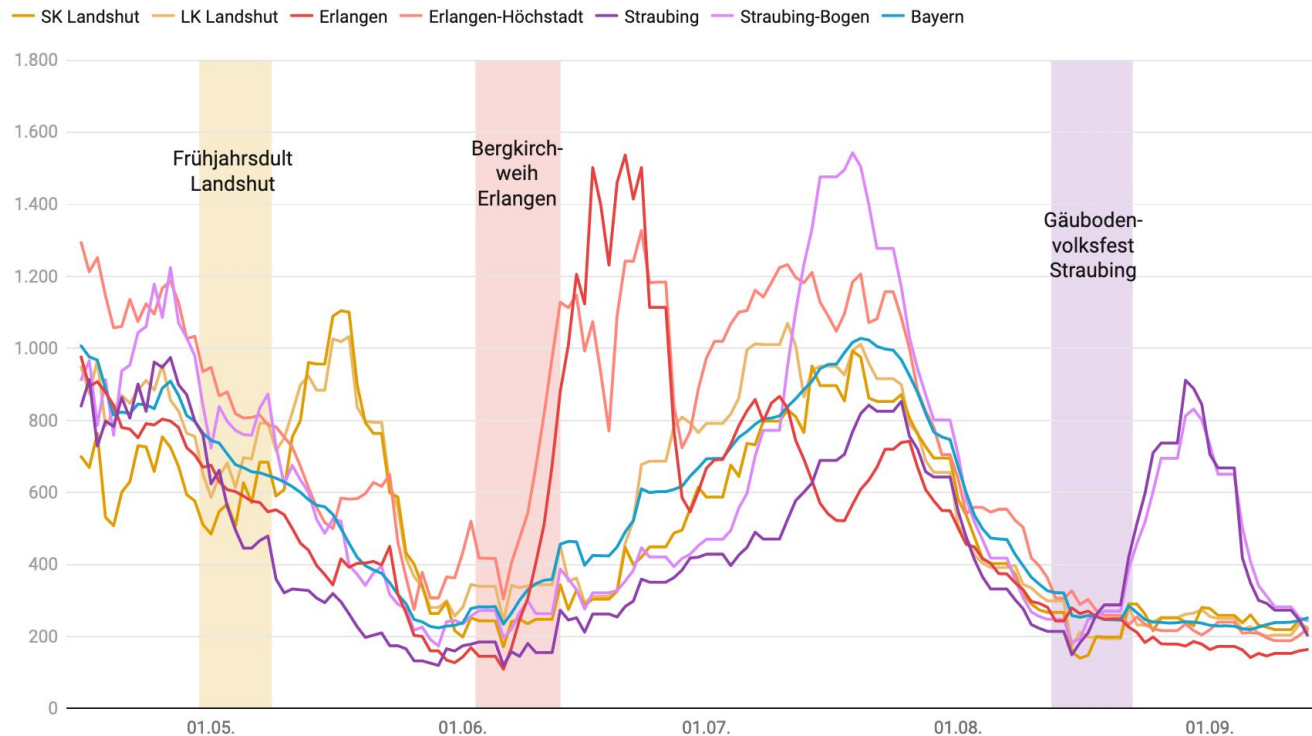
corr x4 y4 (obs=11): 0.816



# Beispiel erzählende Grafik

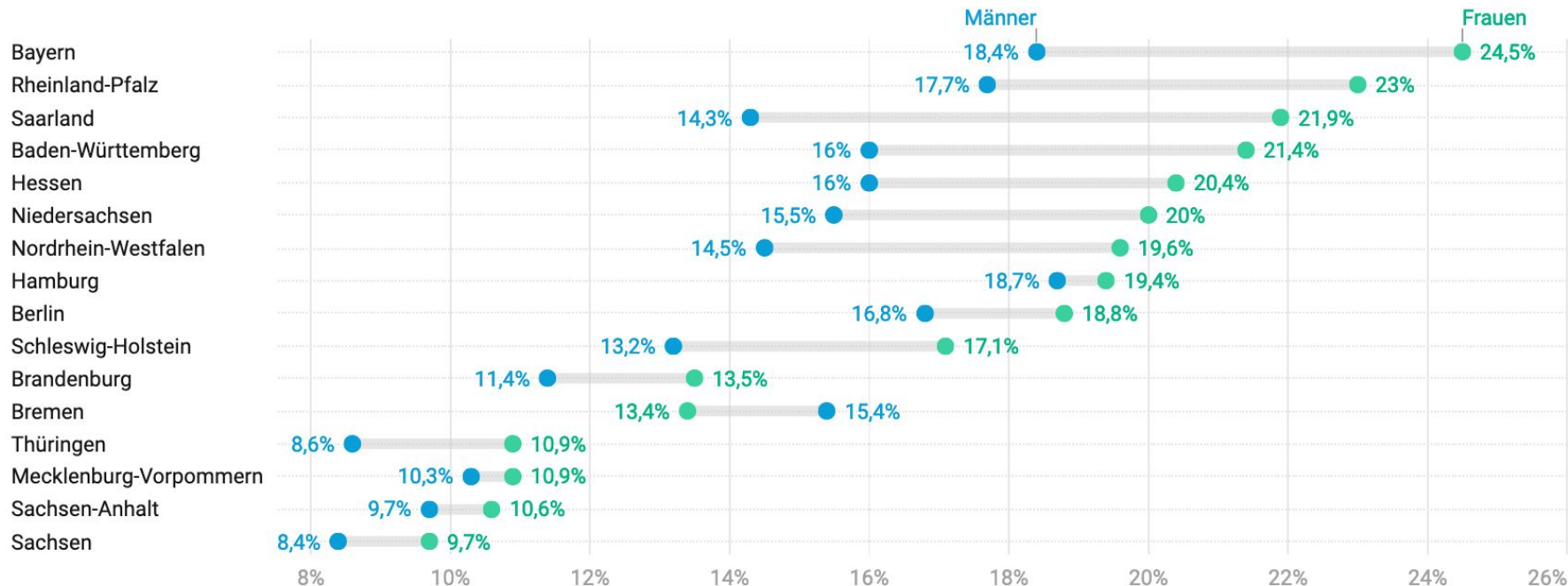
## 7-Tage-Inzidenzen nach Volksfesten: Deutlicher Anstieg

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

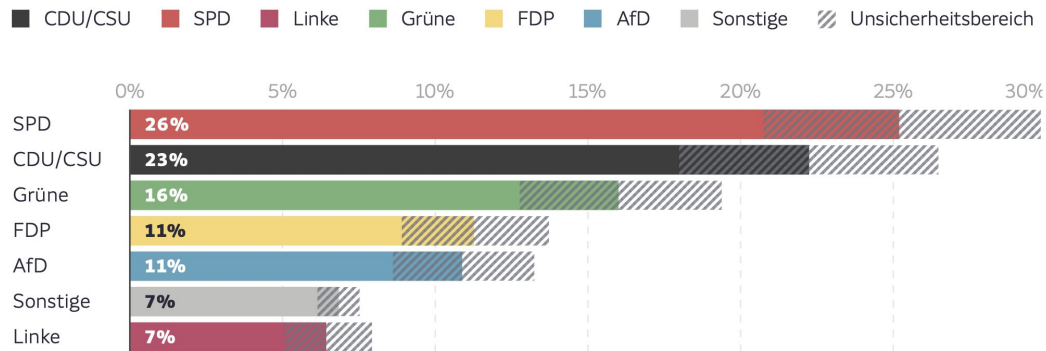


\*Nach EU-Standard: Anteil der Bevölkerung oder Bevölkerungsgruppe, dem weniger als 60% des mittleren Äquivalenzeinkommens zur Verfügung steht. Gemessen am Landesmedianeinkommen des jeweiligen Bundeslandes.

# 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](#)

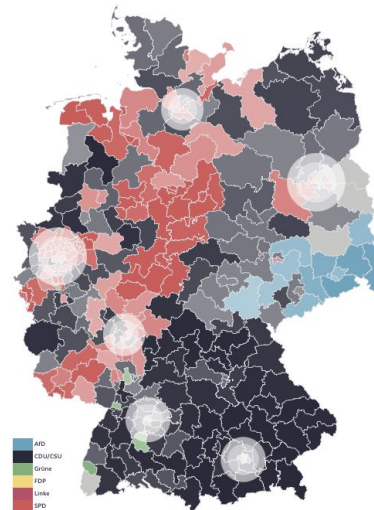
2021 Bundestagswahl

Umfragen Wahlkreise Koalitionen Kandidaten Themen

Prognose für Ihren Wahlkreis

Geben Sie Ihre PLZ ein

Übersicht über alle Wahlkreise

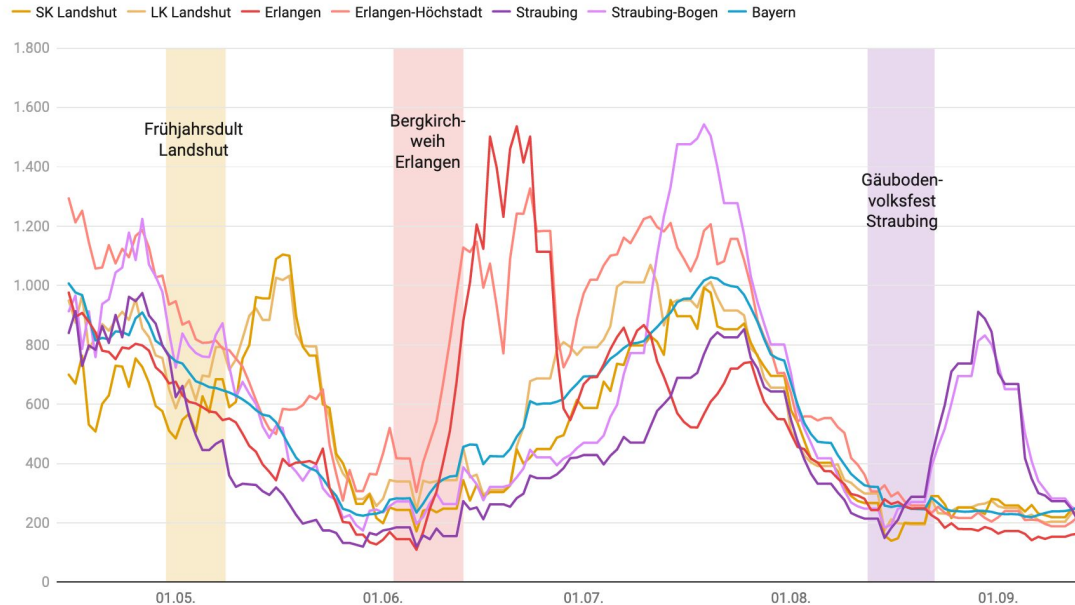


Umfragen geben einen guten Überblick, wie die politische Stimmung aktuell aussieht. Wahlprognosen dagegen fokussieren sich auf den Wahltag und betrachten neben Umfragen auch andere Variablen die das Wahlverhalten erklären. So können sie für einzelne Wahlkreise das Rennen vorhersagen. Mehr zu unserem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



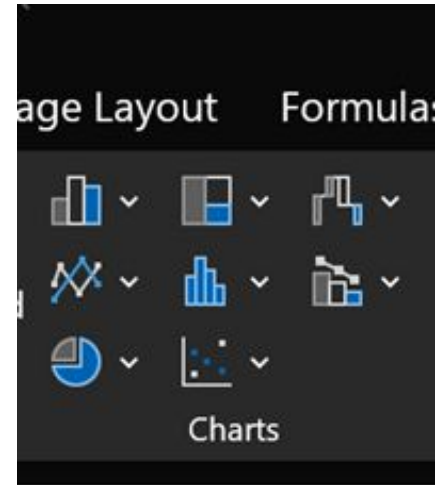
Welche Grafik?

Column or Bar-Chart?

Line Chart?

Pie or Doughnut Chart?

Histogram or Box-Wishker?



# Chart Doctor

Was wollen wir uns anschauen?

- Abweichung
- Korrelation
- Rangfolge
- Verteilung
- Veränderung über die Zeit
- Ganz-Teil Beziehung
- Größen
- Verteilung im Raum
- Ströme: Veränderung von einer Kategorie zur nächsten

**Deviation**  
Emphasizes variation. It's from a baseline or average. Typical of performance reports. Typical of the variance given over time. Can be a single or multiple categories. Can be used to compare. Can be used to highlight.

**Correlation**  
Shows the relationship between two or more variables. Can be positive or negative. Can be used to highlight the strength of the relationship. Can be used to highlight the direction of the relationship.

**Ranking**  
Use when an item needs to be ranked in a list. Typical of performance reports. Typical of the variance given over time. Can be a single or multiple categories. Can be used to compare. Can be used to highlight.

**Distribution**  
Shows the distribution of a variable. Can be used to highlight the range of the variable. Can be used to highlight the center of the distribution. Can be used to highlight the shape of the distribution.

**Change over Time**  
Use when you want to show how a variable changes over time. Can be used to highlight the direction of change. Can be used to highlight the rate of change. Can be used to highlight the magnitude of change.

**Part-to-whole**  
Shows how a single entity can be broken down into its constituent parts. Can be used to highlight the relative size of each part. Can be used to highlight the total size of the entity.

**Magnitude**  
Shows size comparisons. Can be used to highlight the relative size of each entity. Can be used to highlight the total size of the entity.

**Spatial**  
Use when you want to show the spatial distribution of a variable. Can be used to highlight the location of each entity. Can be used to highlight the shape of the distribution.

**Flow**  
Shows the flow of a variable. Can be used to highlight the direction of flow. Can be used to highlight the magnitude of flow. Can be used to highlight the shape of the distribution.

## Visual vocabulary

Designing with data  
There are so many ways to visualise data – how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

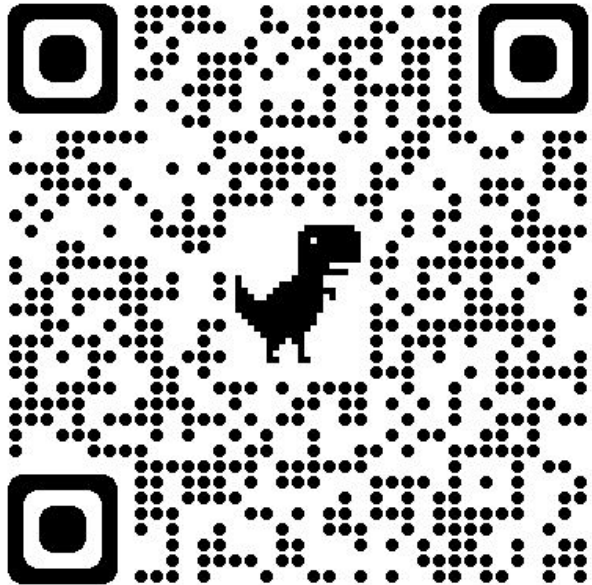
© 2014 Alan Watts, Chris Crossland, Ben L. Frost, Graham Furness, Mike Sheringham, Paul Hill, David Martin, Steve Hopper. All rights reserved. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

[ft.com/vocabulary](http://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



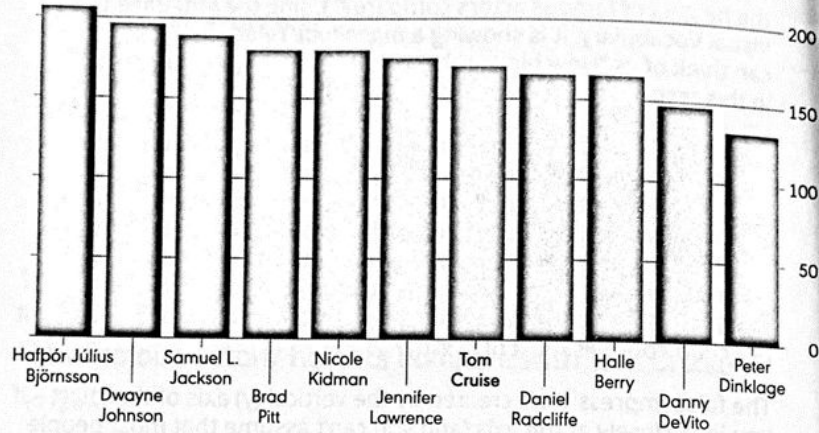
Ergebnisse kommunizieren



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

## How famous actors' heights compare with Tom Cruise

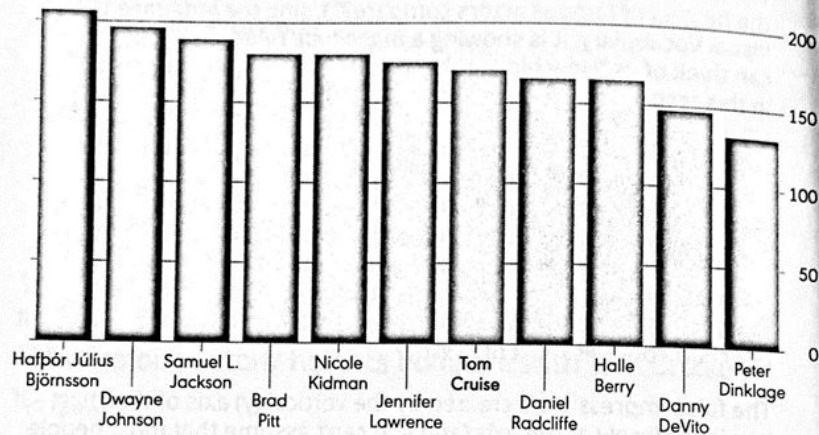
Height (cm)



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

## How famous actors' heights compare with Tom Cruise

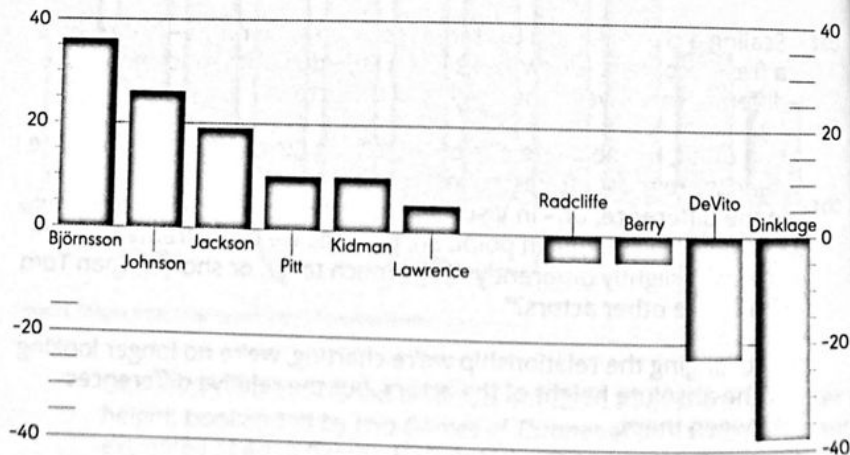
Height (cm)



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

## How famous actors' heights compare with Tom Cruise

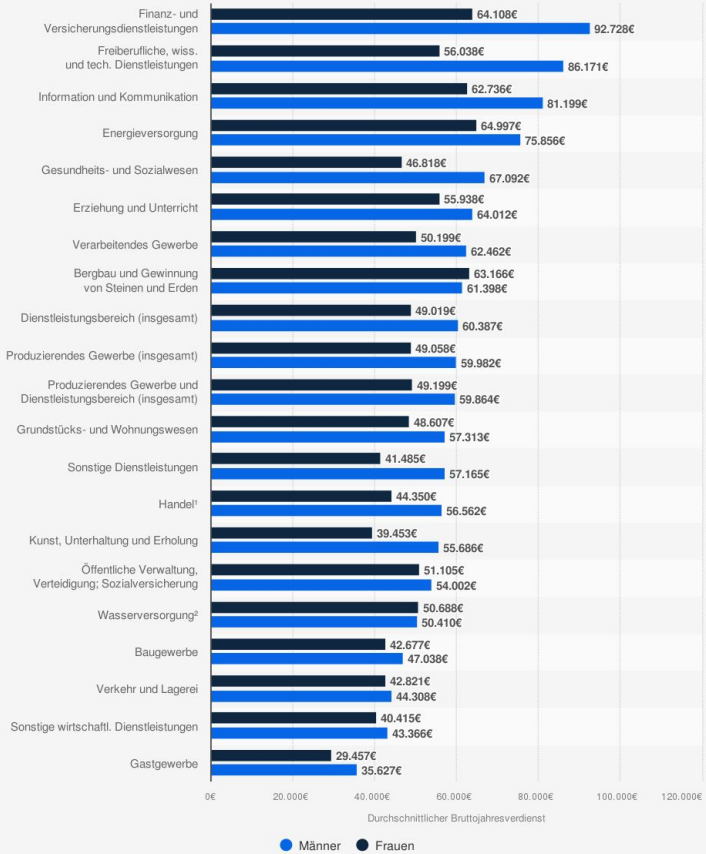
Amount taller/shorter than Cruise (cm)



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

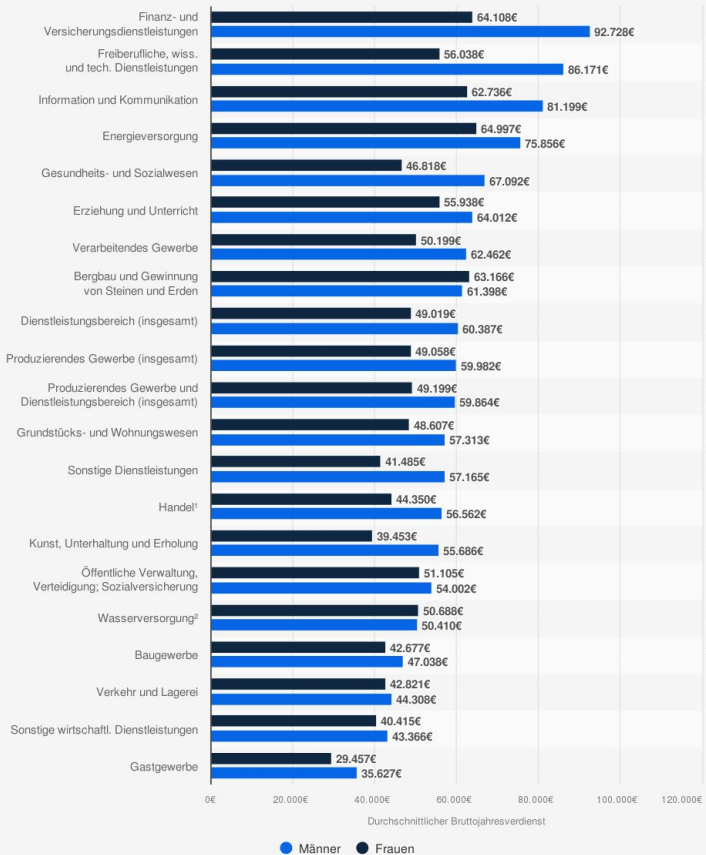


## Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen) nach Wirtschaftsbereichen und Geschlecht im Jahr 2022



Weitere Informationen:  
Deutschland; 2022; Mit Sonderzahlungen

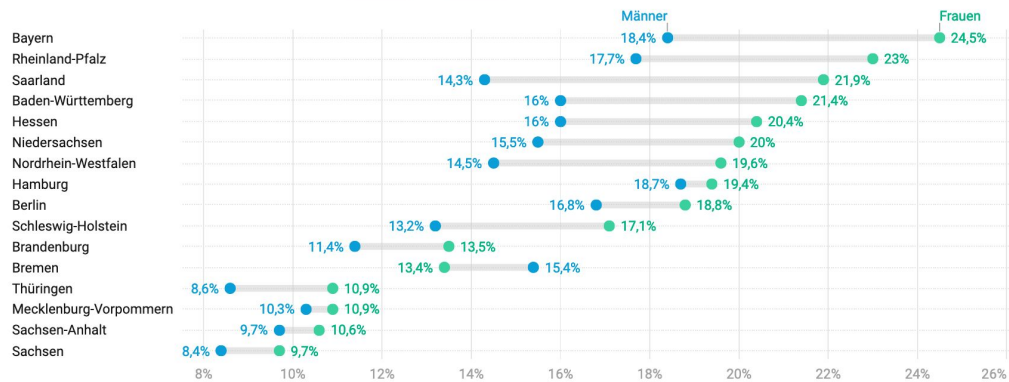
## Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen) nach Wirtschaftsbereichen und Geschlecht im Jahr 2022



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

# Farben und Intensität

<https://www.theguardian.com/world/interactive/2012/may/08/gay-rights-united-states>



Scale states equally

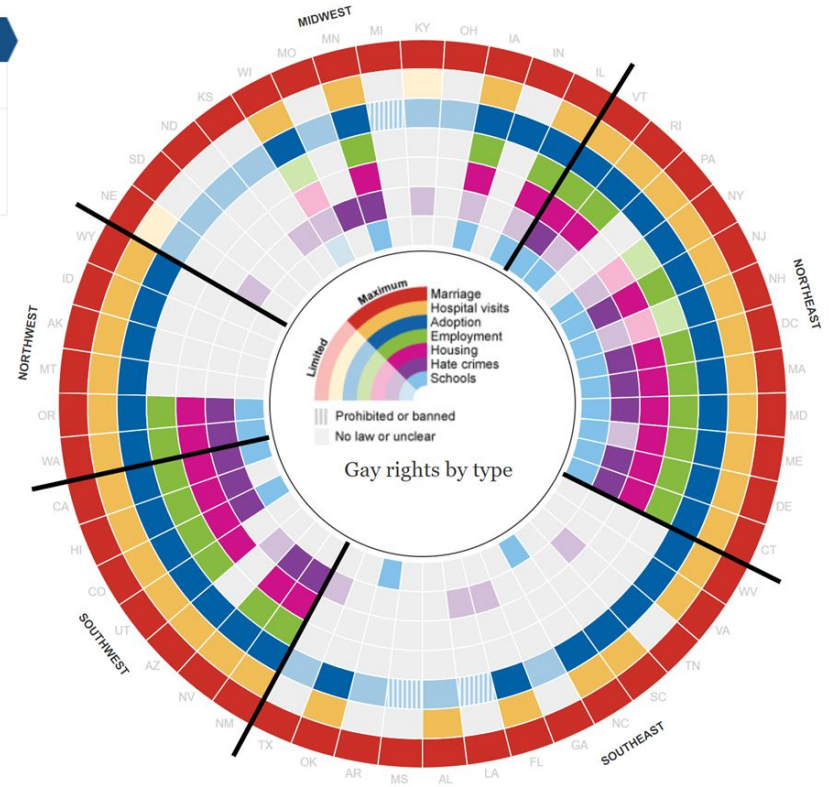
Scale states by population

Where your friends live

Connect to Facebook to see the rights of states where your friends live. Your information will not be saved.

Share your state on Facebook

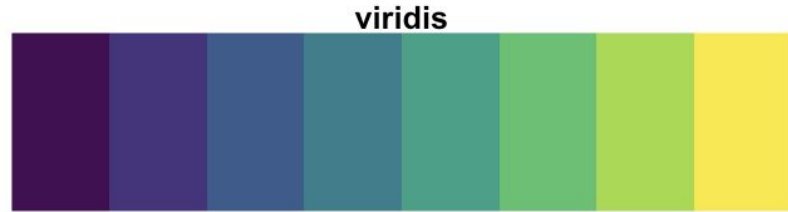
You are not connected to Facebook.



# Farben werden nicht immer gleich wahrgenommen



# Robuste Farbpaletten helfen bei der Kommunikation



Welche Software?



Flourish

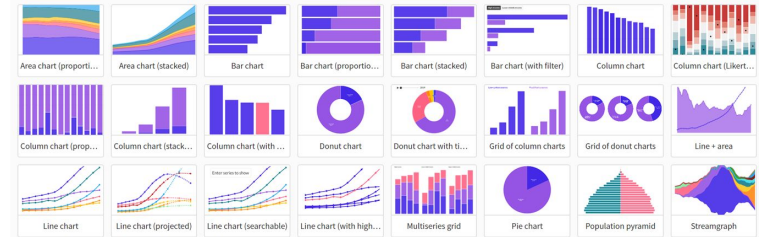
Choose a template

Show [Featured](#) Premium Favorites Mine Company

**Line, bar and pie charts**

Basic types of chart, single or in a grid

STARTING POINTS



**Projection map**

Make data maps with region-shading and/or points. Adapt an existing example or upload your own geographic boundary file.

... und viele mehr

# Teamarbeit - Datenanalyse

Die Teams arbeiten an einem Projekt, das am Ende in einer Präsentation vorgestellt werden soll:



3 Slides pro Team:

- Ergebnisse
- Plan für Umsetzung
- Zukunftsvision: Was bräuchte man noch für das "perfekte" Archivierungs-System?



# Reporting

# Das IDEEN Arbeitsblatt

## Wer ist unsere Zielgruppe?

Schreibe alle Gruppen oder Personen auf die adressiert werden sollen.

Was bewegt die Zielgruppe?

Was soll die Zielgruppe mit den Informationen tun?

## Um was geht es?

Was passiert wenn die Zielgruppe die gelieferten Informationen nicht nutzt?

Was ist das der positive Effekt wenn die Informationen genutzt werden?

## Formuliere deine Idee/Message

Ein Satz, der deinen Standpunkt verdeutlicht und vermittelt um was es geht.

# Storyboard



Daten



Analyse



Ergebnis



Kontext



Problem-  
stellung



Empfehlung

# Zahlen nie kopieren – “Powerpoint” direkt erstellen

```
--
title: "Stichprobenziehung"
author: "Andreas Nestl"
format: pptx


--
## Daten

- Beispieldatensatz aus Formustar
- Zeitraum Januar bis Juni

## Zivilklagen Streitwerte nach Streitgegenstand

@fig-streit zeigt die Verteilung der Streitwerte auf Bayern.
```{r} #| label: fig-streit
#| fig-cap: "Streitwert nach Streitgegenstand in Euro."
#| warning: false library(ggplot2)

ggplot(akten, aes(x = Streitgegenstand, y = `Streitwert in EURO`)) +
  geom_violin() + lims(y = c(0, 100000)) + coord_flip()
``
```





## Government Public Sector

Mostly open data



Animated NYC metro traffic



City Cycle Race (with STRAVA) - Compare the cycling speed of cities



Crime Watch



Dublin Transport Info



Freedom of Press Index



Locating Blood Banks in India

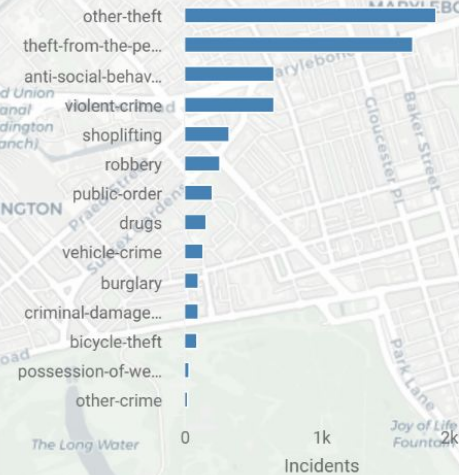
Type an address or location

Or use your current location?

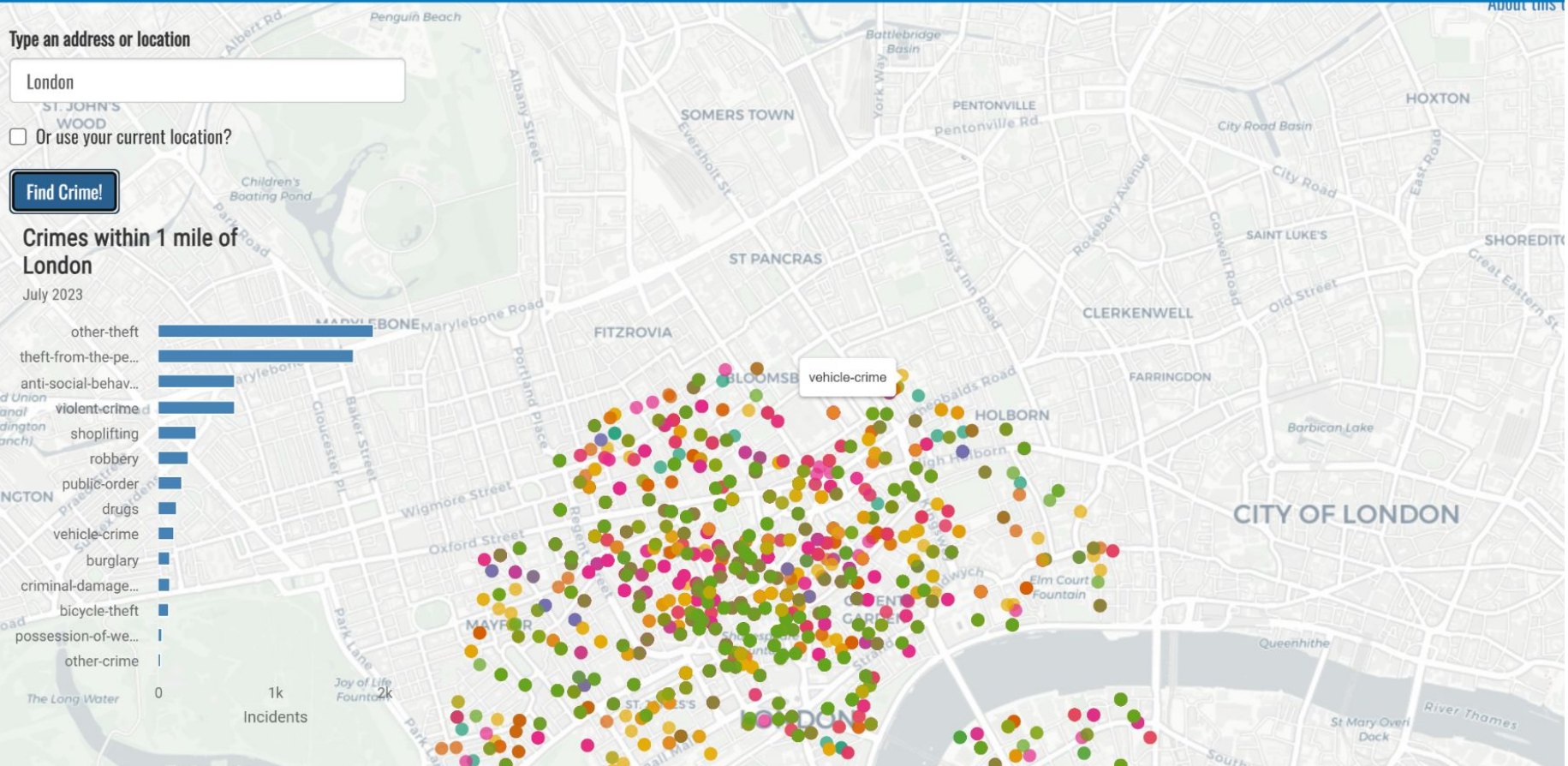
**Find Crime!**

## Crimes within 1 mile of London

July 2023



0 1k 2k  
Incidents



# ChatGPT: How I Made An R Shiny App In Under 10 Minutes

month	entertainment	groceries	rent	transportation	utilities
April	200.00	500.00	600.00	200.00	200.00
February	200.00	500.00	600.00	200.00	200.00
January	200.00	500.00	600.00	200.00	200.00
March	200.00	500.00	600.00	200.00	200.00

**What You Learn**  
How to Use ChatGPT to Develop Shiny Apps

**MORE VIDEOS**  
Pause (k)

**Making A Shiny App With ChatGPT**

R-Tip 60

0:06 / 10:09

YouTube

## ChatGPT: How I Made An R Shiny App In Under 10 Minutes

YouTube | Business Science | 13,9K views | 6 months ago



BAYSICS bringt Bürger\*innen, Schüler\*innen und Wissenschaftler\*innen aller Altersgruppen zusammen, um die Auswirkungen des Klimawandels auf **Pflanzen und Tiere in Bayern zu beobachten**, zu analysieren und auszuwerten. Mit BAYSICS erforschen, verstehen und handeln.

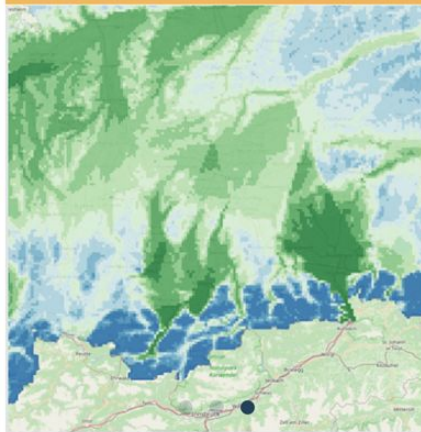
**mach mit!**

### PORTAL



**analysiere**

### NatureExplorer



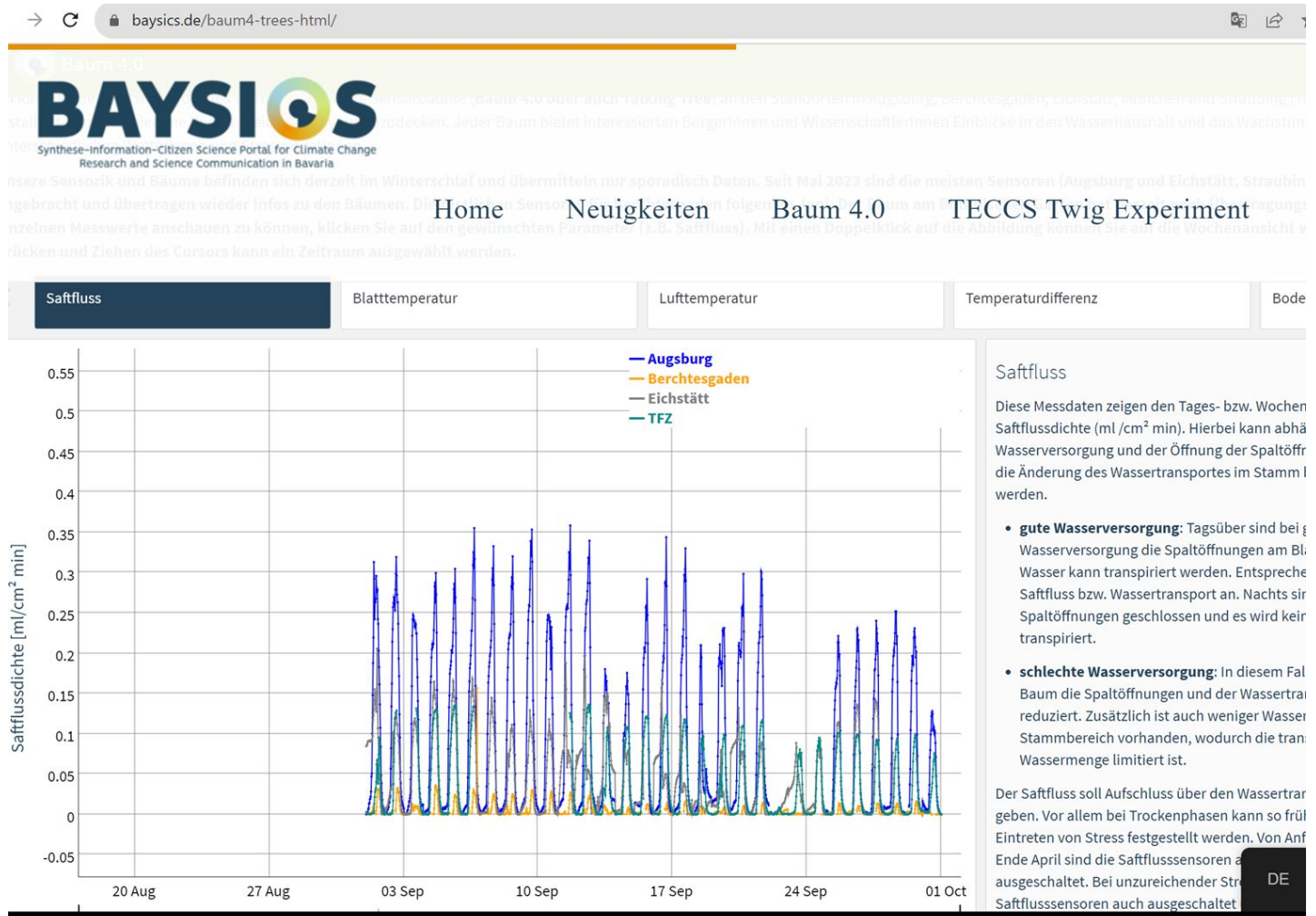
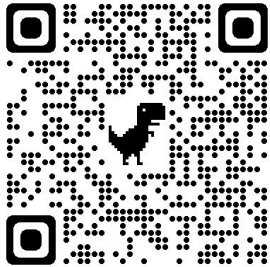
### Forschungsprojekt





# Shiny App

Zeigt auf der Webseite Safffluss, Temperatur etc. für Messstationen in Bayern an



**Export**

# Rahmenkonzept: **Five Safes**



Safe  
projects



Safe  
people



Safe  
settings



Safe  
data



Safe  
export

# Safe export

Informationen dürfen die Plattform nur nach Kontrolle verlassen

- Statistiken dürfen keinen Rückschluss auf individuelle Daten erlauben
- Ein 'export request' wird zunächst von Ihnen selbst, dann von den MitarbeiterInnen am ADRF geprüft



# Daten teilen – mit Bedacht

The logo for Netflix, consisting of the word "NETFLIX" in a bold, red, sans-serif font.

2006 hatte Netflix 100 Millionen anonymisierter Filmbewertungen bereitgestellt.

Jede mit ID, Titel, Erscheinungsjahr, Jahr der Nutzerbewertung.

16 Tage später hatten A. Narayanan (Princeton) und V. Shmatikov (Cornell Tech) Nutzer identifiziert und deren sonstigen (nicht öffentlich sichtbaren Filme) zuordnen können.

# Gute Praxis für den Export



1. Revisionsprozess ist manuell und zeitaufwendig.



2. Nur Endversionen zur Überprüfung einreichen.



3. Erstellen Sie Ausgabedateien (z.B. Tabellen, Grafiken) über den Code. Code-Dateien benötigen eine Beschreibung. Code mit Kommentaren.



4. Aussagekräftige Variablennamen verwenden (z.B. "Streitwert" statt "var1").

## Gute Praxis bei Tabellen:

**Zellen** mit <10 Beobachtungen unterdrücken.

**Zellwerte**: auf sinnvolle Einheiten runden (z.B. \$45,675 -> \$46,000).

**Gewichtung**: Gewichtete als auch nicht gewichtete Zählungen berichten.

**Verhältnisse**: Anzahl gültiger Fälle für Zähler und Nenner berichten.

**Perzentile**: Statt Perzentile "unscharfe Mediane" berechnen.

**Maxima und Minima unterdrücken**. Exaktes durch top-codiertes ersetzen.

# Gute Praxis bei Grafiken

Grafiken sind Tabellendarstellungen, deshalb Quelldaten der zugrundeliegenden Tabelle gemäß Tabellenrichtlinien bereitstellen.

Erwägen Sie, die Tabelle zu exportieren und die Grafik extern zu erstellen.

Bei Grafiken aus Einzeldaten, die Einzelwerte anzeigen (z.B. Streudiagramme): sicherstellen, dass keine Reidentifikation möglich ist und Werte nur mit hoher Unsicherheit geschätzt werden können.

Maßnahmen können sein:

- Verteilungsenden abschneiden,
- Ausreißer entfernen,
- Werte ruckeln,
- Achsenwerte ändern.



# Innerhalb der Datenumgebung



Please keep in mind that any removal of data is prohibited. This includes taking screenshots or writing down numbers on paper. All output has to go through disclosure review. Violations of this policy may result in penalties and revoking access from the system.

 <b>Jupyter Notebooks</b> Learn the basics of Notebooks, accessing data with DBeaver & reading data using Python and R.	 <b>User Documentation</b> Access the Corebridge Initiative ADRF User Documentation	 <b>Support Link</b> For questions about using the ADRF or to request applications, libraries or resources please email us at <a href="mailto:support@cokebridgeinitiative.org">support@cokebridgeinitiative.org</a>
 <b>Export Request</b> Initiate new or review the status of existing export requests	 <b>Data Catalog</b> Search and review the Data Catalog Coming Soon!	 <b>Usage Metrics</b> Review your usage of the ADRF platform Coming Soon!

#### Important Announcements:

- To log off of the system please click on the **Profile** icon in the upper right of your screen and select **End Session**.
- Maintenance Window:** Friday, Thursday 8/09/2024 - 1:30:00AM PST/PT

7. Click My Requests, or the top (person-shaped) icon, at the left side of the window as shown in the screenshot below.

The screenshot shows a web browser window with the URL `export.adrf.net/reviews`. The page header is "ADRF EXPORTS" with a user profile for "Josh Edelman". On the left sidebar, under the "EXPORT" section, the "My Requests" menu item is highlighted with a red box. The main content area is titled "My Requests" and features a "NEW ITEM" button. Below this is a table with the following data:

Date Request	Project Name	Primary Reviewer	Secondary Reviewer	Last Status	Last Status Date	Actions
2021/05/03 19:11:21	CI_Admin	Nathan Caplan		Rejected by Reviewer	2021/05/06 20:34:04	

At the bottom of the table, it indicates "Rows per page: 10" and "1-1 of 1".

8. Click New Item as shown below.

This screenshot is identical to the one above, but the "NEW ITEM" button in the top right corner of the main content area is highlighted with a red box.

Vortrag: Was bisher geschah + Visualisieren	10:00 - 10:45
Pause	10:45 - 10:55
Teamarbeit: Datenanalyse	10:55 - 12:00
Mittagspause	12:00 - 12:45
Vortrag: Reporting + Export aus der Cloud	12:45 - 13:10
Teamarbeit: Projekt festzurren, Export von Grafiken/Ergebnissen	13:10 - 14:45
Pause: Selbstbestimmt nach Bedarf der Teams	
Vortrag: Tagesabschluss	14:45 - 15:00

Was waren die Interessantesten Erkenntnisse bisher?



# Vorgehen bei der Beantwortung von Fragen mit Daten

Daten ergänzen

- record linkage
- Web scraping

Next steps

