

Modul 3: Nachvollziehbares Reporting, Visualisierung und Dashboards

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)
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

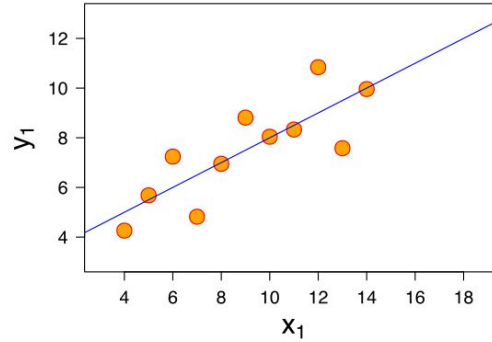
corr x2 y2 (obs=11): 0.816

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



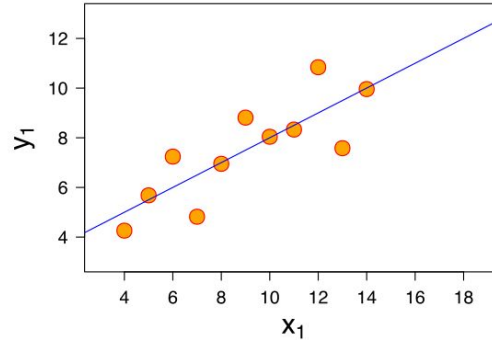
Beispiel Explorative Grafik - Beispiel Korrelationen

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



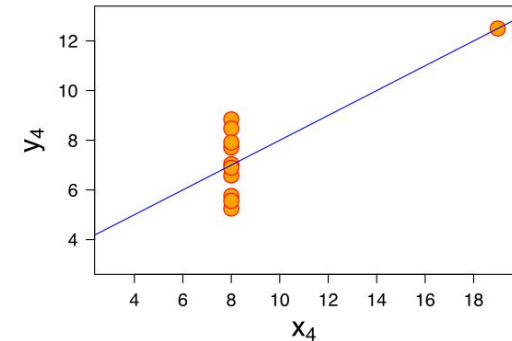
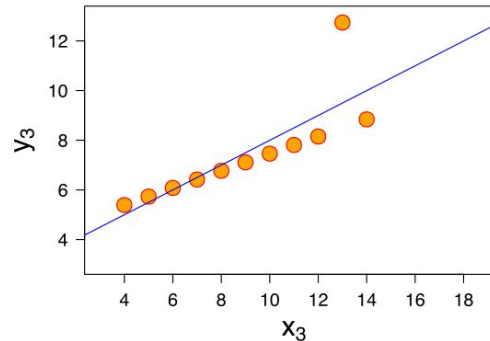
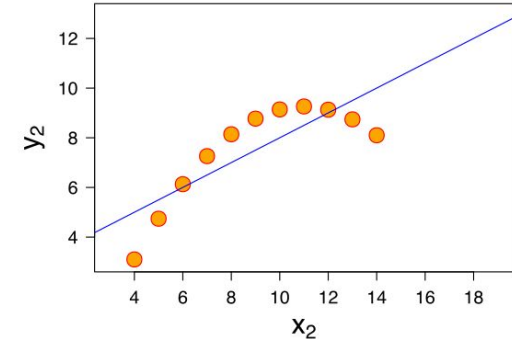
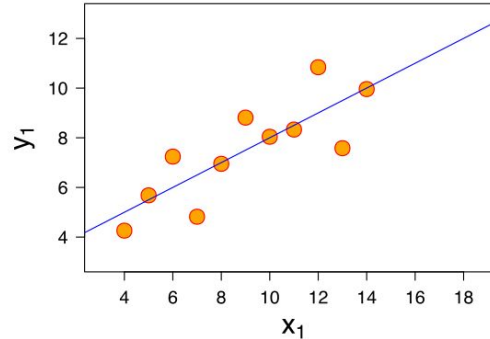
Beispiel Explorative Grafik - Beispiel Korrelationen

corr x1 y1 (obs=11): 0.816

corr x2 y2 (obs=11): 0.816

corr x3 y3 (obs=11): 0.816

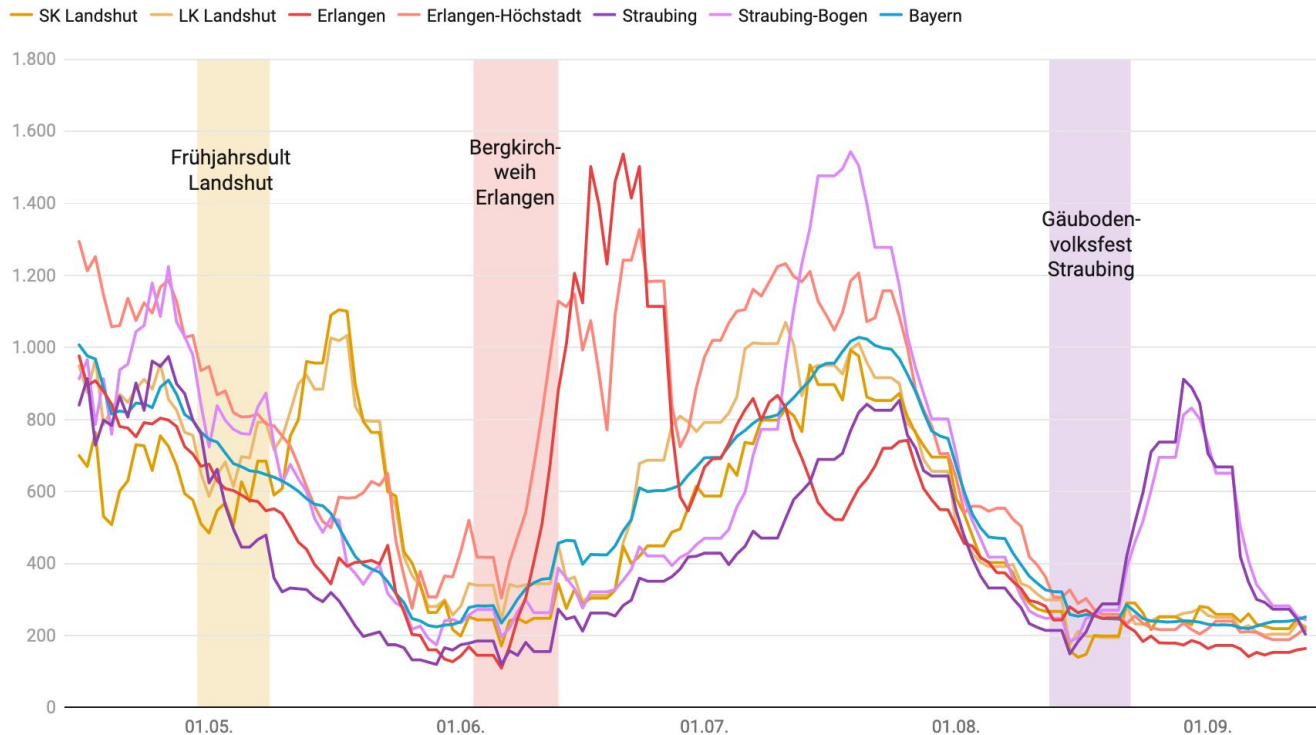
corr x4 y4 (obs=11): 0.816



Beispiel erzählende Grafik

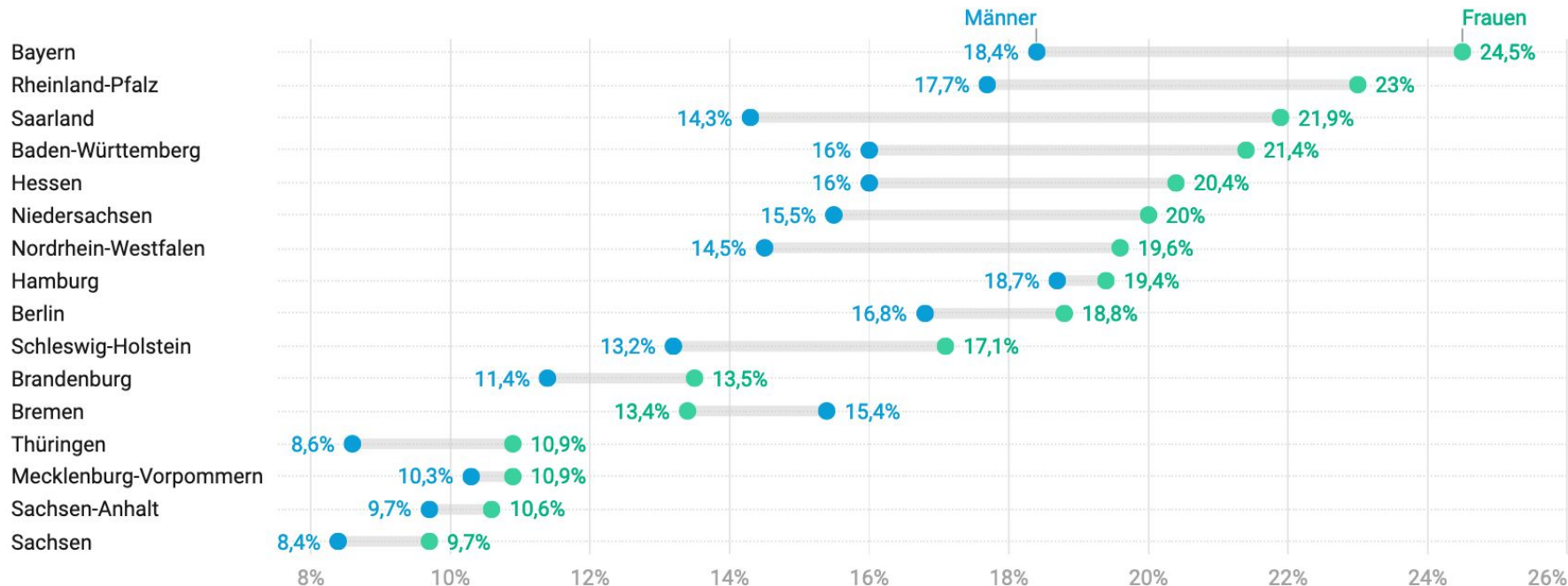
7-Tage-Inzidenzen nach Volksfesten: Deutlicher Anstieg

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

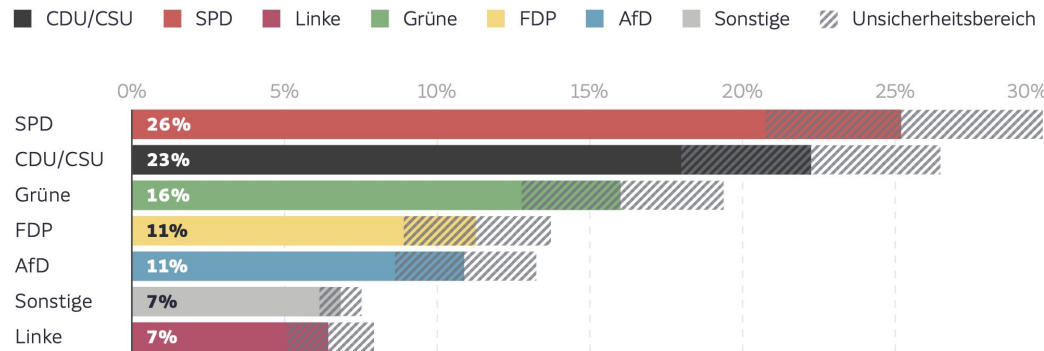


*Nach EU-Standard: Anteil der Bevölkerung oder Bevölkerungsgruppe, dem weniger als 60% des mittleren Äquivalenzeinkommens zur Verfügung steht. Gemessen am Landesmedianeinkommen des jeweiligen Bundeslandes.

Wahlvorhersagen sind immer unsicher

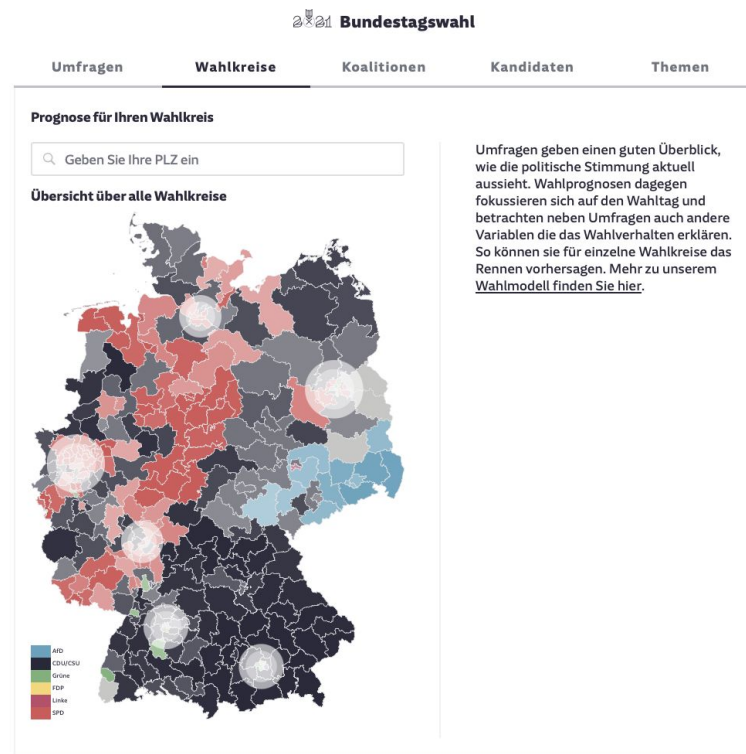
Wahlprognose für die Bundestagswahl

in diesen Bereichen landen die Parteien im Wahlmodell mit hoher Wahrscheinlichkeit



Stand: 25.09.21 00:17 Uhr

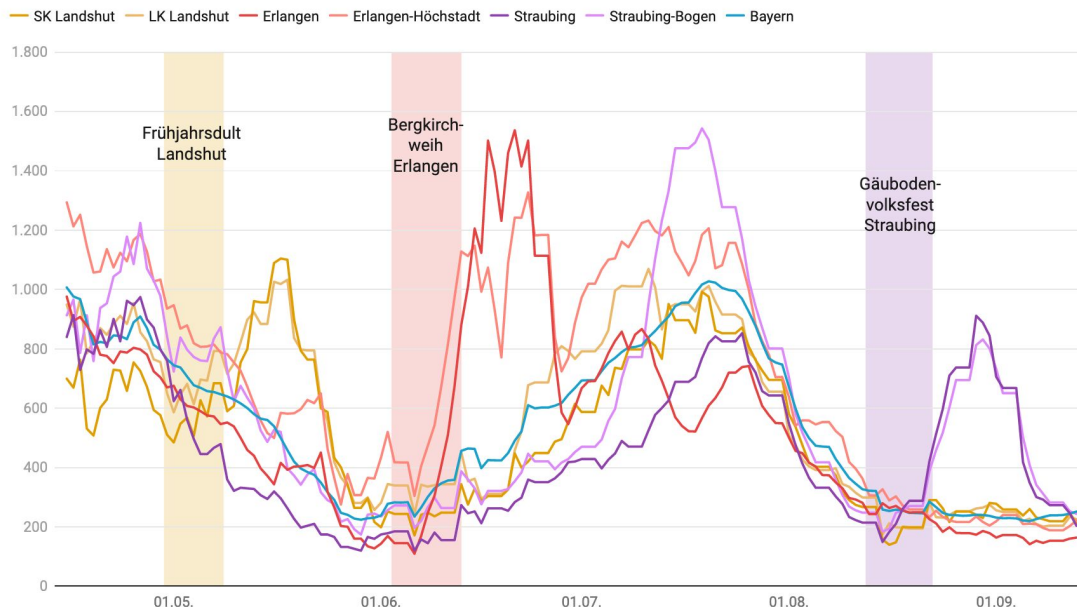
Quelle: [Zweitstimme.org](https://www.zweitstimme.org) • [Rohdaten herunterladen](#)



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-Whisker?

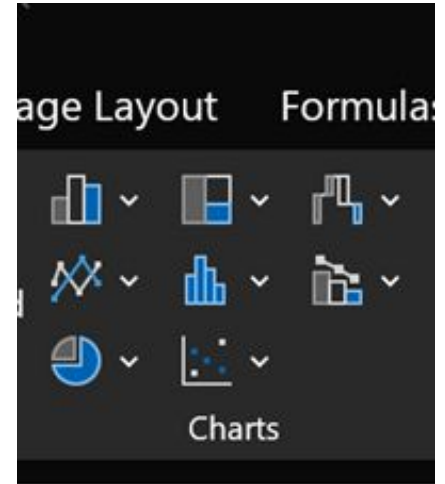
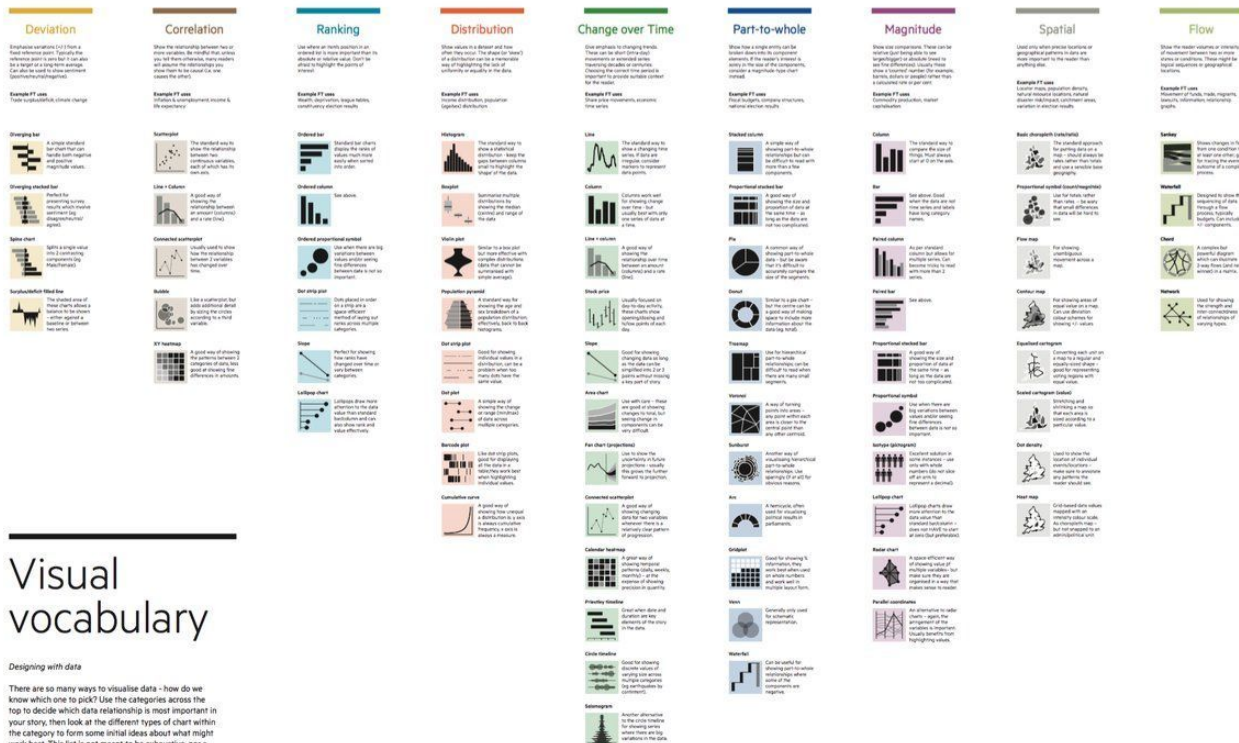


Chart Doctor

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



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 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



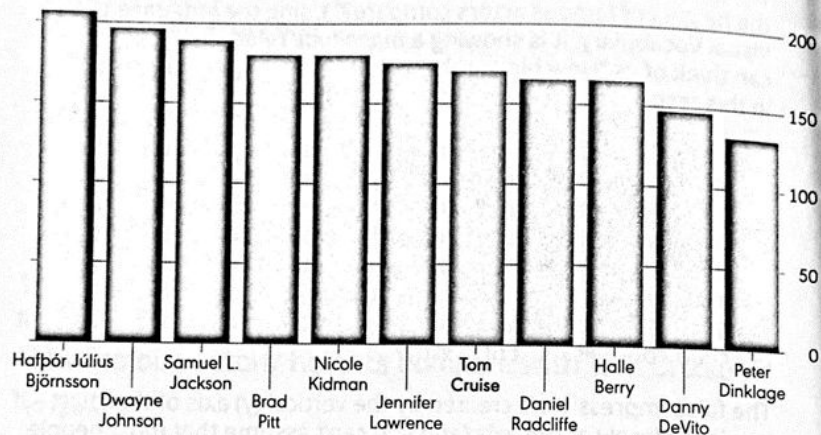
Ergebnisse kommunizieren



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

How famous actors' heights compare with Tom Cruise

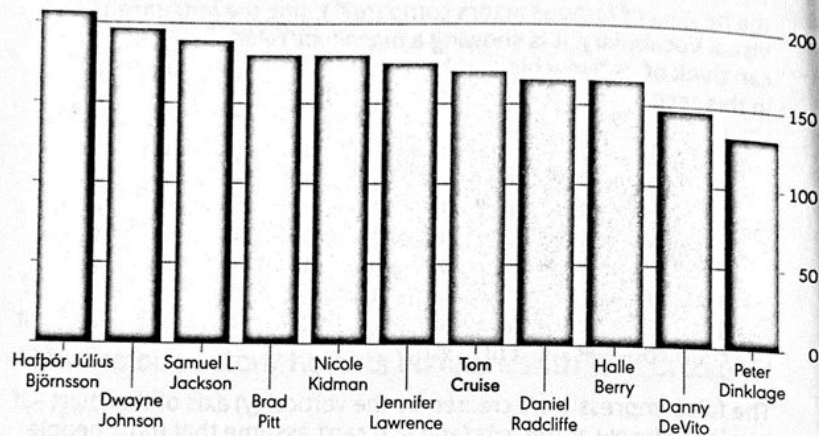
Height (cm)



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

How famous actors' heights compare with Tom Cruise

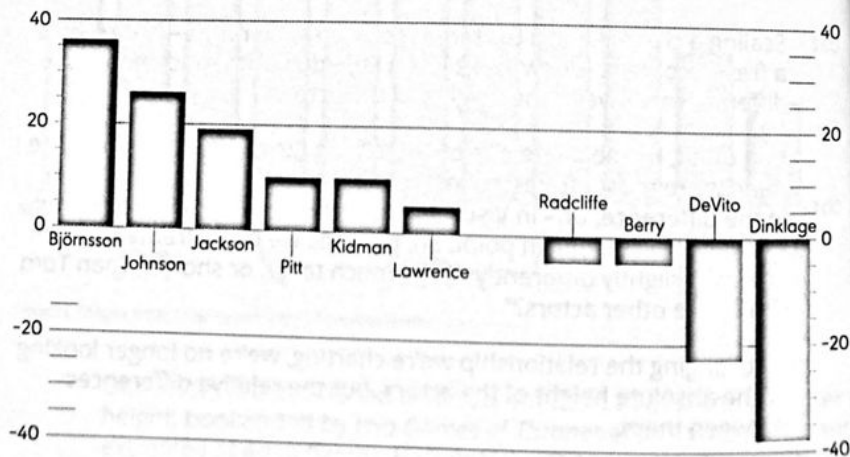
Height (cm)



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

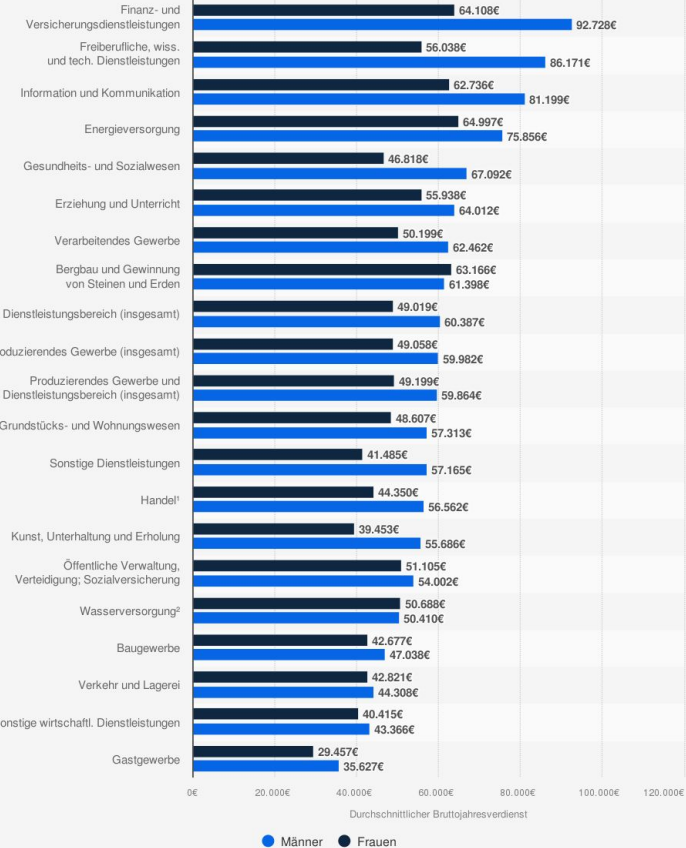
How famous actors' heights compare with Tom Cruise

Amount taller/shorter than Cruise (cm)



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

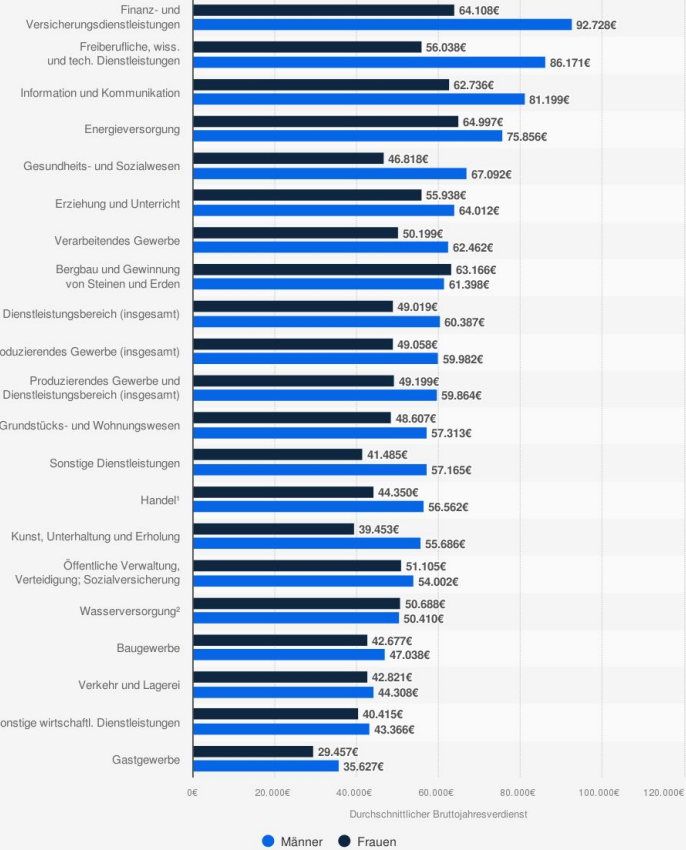
Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen)
nach Wirtschaftsbereichen und Geschlecht im Jahr 2022



Durchschnittlicher Bruttojahresverdienst

● Männer ● Frauen

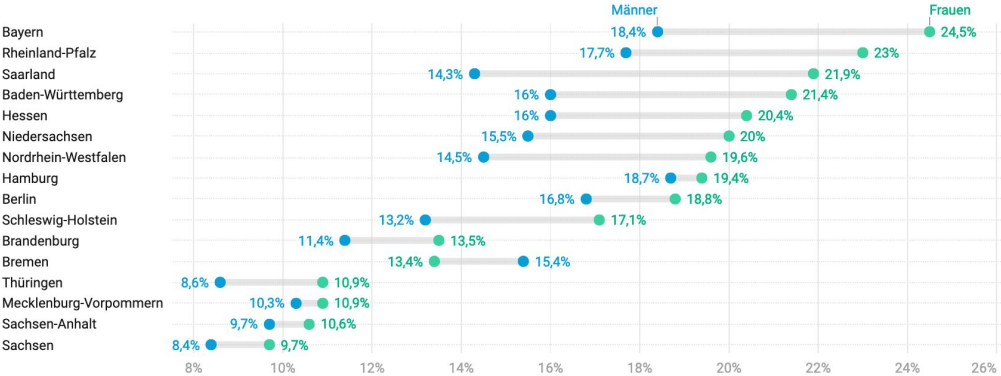
Durchschnittlicher Bruttojahresverdienst von Arbeitnehmern (mit Sonderzahlungen) nach Wirtschaftsbereichen und Geschlecht im Jahr 2022



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

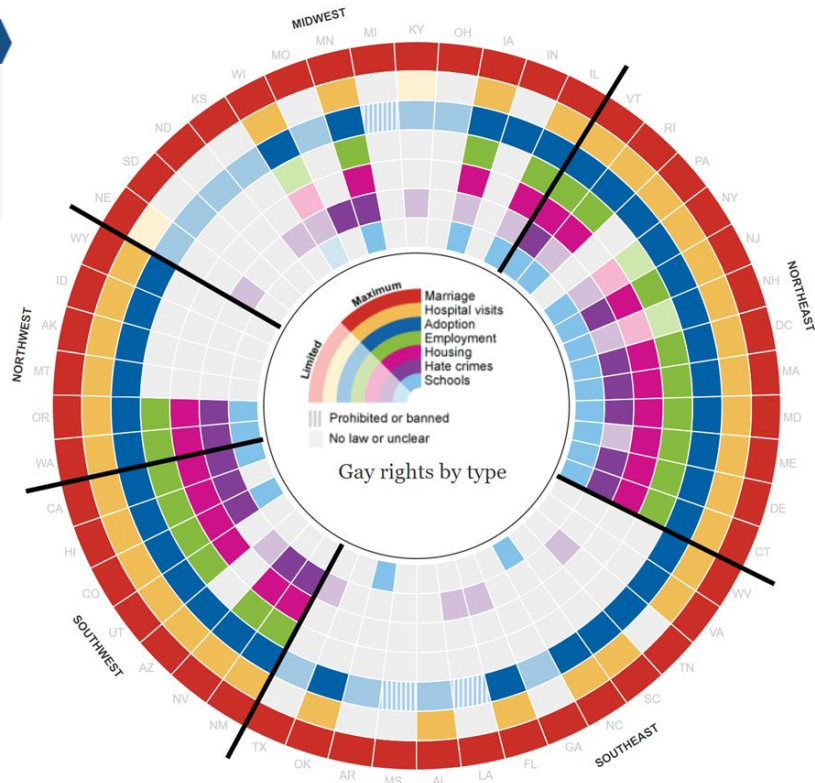


*Nach EU-Standard: Anteil der Bevölkerung oder Bevölkerungsgruppe, dem weniger als 60% des mittleren Äquivalenzeinkommens zur Verfügung steht. Gemessen am Landesmedianeinkommen des jeweiligen Bundeslandes.

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

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

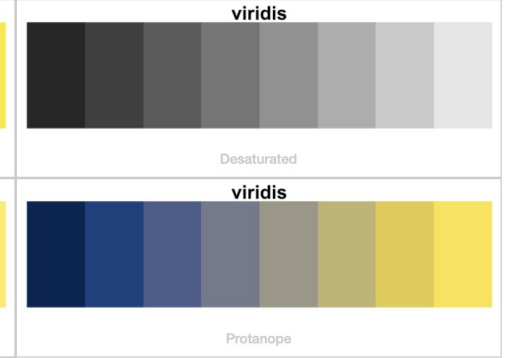
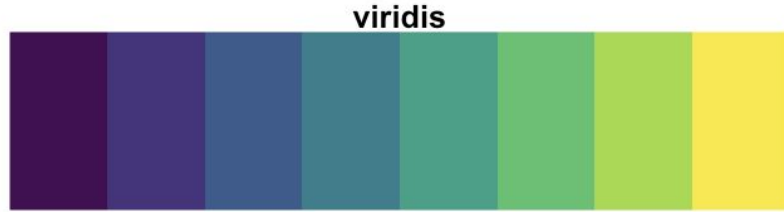
You are not connected to Facebook.



Farben werden nicht immer gleich wahrgenommen



Robuste Farbpaletten helfen bei der Kommunikation



Welche Software?



Flourish[®]

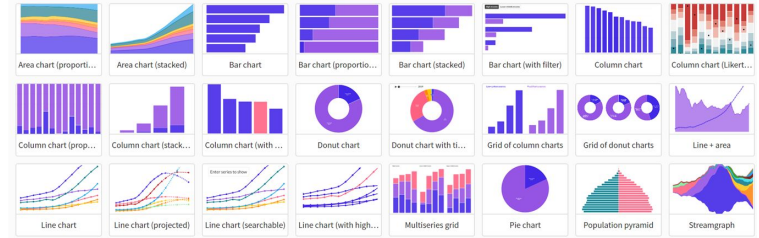
Choose a template

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

```
--  
title: "Stichprobenziehung"  
author: "Andreas Nestl"  
format: pptx  
  
--  
## Daten  
  
- Beispieldatensatz aus Formustar  
- Zeitraum Januar bis Juni  
  
## Zivilklagen verteilt auf Bayern  
  
@fig-geog zeigt die Verteilung der Zivilklagen auf Bayern.  
```{r} #| label: fig-airquality  
#| fig-cap: "Temperature and ozone level."
#| warning: false library(ggplot2)

ggplot(airquality, aes(Temp, Ozone))
 + geom_point()
 + geom_smooth(method = "loess") ```
```

**Export**

# Rahmenkonzept: **Five Safes**



Safe  
projects



Safe  
people



Safe  
settings



Safe  
data

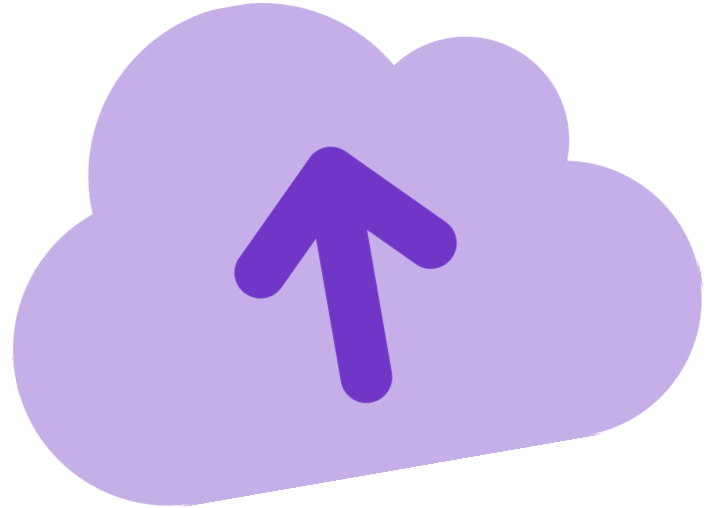


Safe  
export

# 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 Netflix logo, 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.

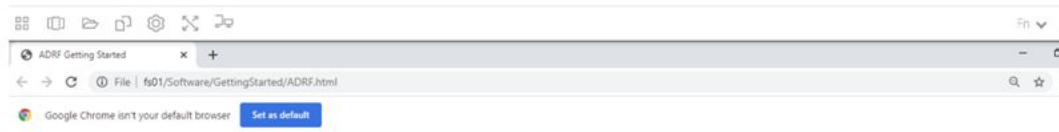
Erwägen Sie, die Tabelle zu exportieren und die Grafik extern zu erstellen.

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



## Attention

Unauthorized access or use of this computer system may subject violators to criminal, civil, and/or administrative action. All information on this computer system may be intercepted, recorded, used, copied, and disclosed by and to authorized personnel for official purposes, including criminal investigations. Such information includes sensitive data encrypted to comply with confidentiality and privacy requirements. Access or use of this computer system by any person, whether authorized or unauthorized, constitutes consent to these terms. There is no right of privacy in this system.

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.



### Jupyter Notebooks

Learn the basics of Notebooks, accessing data with Dilever & reading data using Python and R.



### User Documentation

Access the Coleridge Initiative ADRF User Documentation



### Support Link

For questions about using the ADRF or to request applications, libraries or resources please email us at [support@coleridgeinitiative.org](mailto:support@coleridgeinitiative.org)



### Export Request

Initiate new or review the status of existing export requests



### Data Catalog

Search and review the Data Catalog  
Coming Soon!



### Usage Metrics

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:** Every Thursday 8:00PM - 1:00AM PST/EST

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 title is "My Requests". In the left sidebar, under the "EXPORT" section, the "My Requests" link (which includes a person icon) is highlighted with a red box. Below this link is a "Pending Reviews" section. To the right of the sidebar is a table of requests. At the top right of the table area is a "NEW ITEM" button. The table has columns for Date Request, Project Name, Primary Reviewer, Secondary Reviewer, Last Status, Last Status Date, and Actions. There is one row of 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	

Rows per page: 10 1-1 of 1

8. Click New Item as shown below.

This screenshot is identical to the previous one, but the "NEW ITEM" button at the top right of the table area is highlighted with a red box.

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	

Rows per page: 10 1-1 of 1

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

