

# **Tag 2: Big Data, Good Data? Die kritische Nutzung von Daten**

## Session 8: Kommunikation von und mit Daten

---

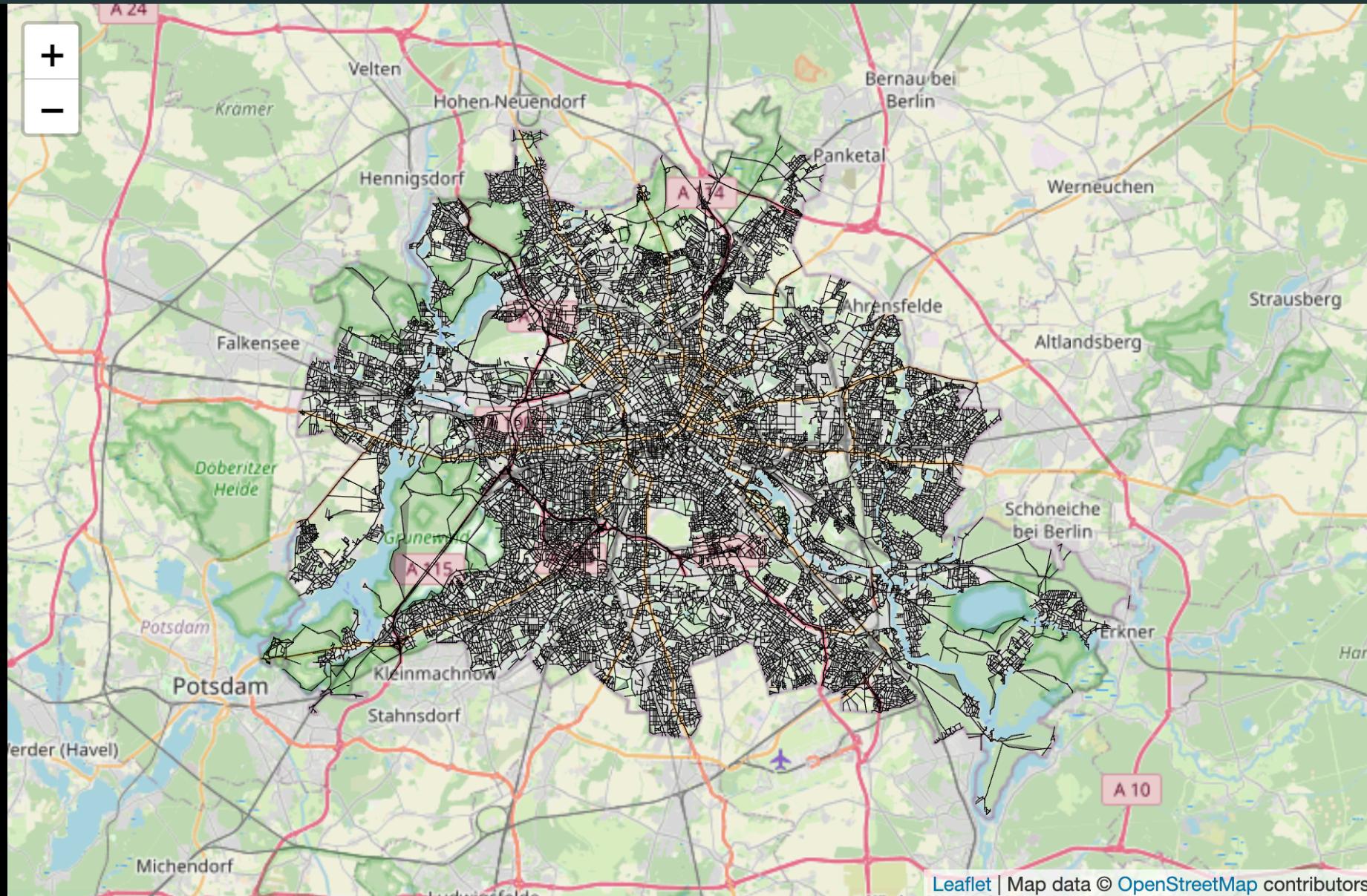
Simon Munzert  
Hertie School

1. Warum Datenvisualisierung?
2. Arten der Datenvisualisierung
3. Visualisierungstools
4. Übung: Best (and worst) practices
5. Interaktive Kommunikation und Monitoring mit Dashboards

# Was sehen Sie hier?



# So besser?



# **Warum Datenvisualisierung?**

---

# Warum Datenvisualisierung?

## Eine mächtige Methode in der Data-Science-Toolbox

- Datenvisualisierung ist eine Methode zur Exploration (und nicht nur Darstellung) von Daten.
- Dabei geht es um mehr als Datenvisualisierung im engeren Sinne, d.h. um die **Kodierung quantitativer Informationen** in visuellen Objekten.
- Entscheidungsträger sind meist an **Mustern** und **Regelmäßigkeiten** interessiert, nicht an einzelnen und genauen Werten.
- Zwei Möglichkeiten, quantitative Informationen sinnvoll zu nutzen:



## Das Argument für Visualisierung

- Visualisierung setzt **wenig oder keine Annahmen** über die Natur der Daten voraus.
- Visualisierung **lässt Sie Dinge sehen**, die sonst unsichtbar wären, insbesondere Beziehungen zwischen Daten (Muster, Trends, Ausreißer).
- Visualisierung erleichtert die Interaktion zwischen Entscheidungsträger und Daten - **sie ist ein Mittel** zur Hypothesenbildung und Entscheidungsfindung.

*"Die entscheidende Frage ist, wie man die Daten am besten in etwas umwandelt, das Menschen verstehen können, um optimale Entscheidungen zu treffen."* Colin Ware, 2013

# Abbildungen vs. Tabellen

- Die Tabelle auf der rechten Seite enthält die Datensätze I bis IV, die jeweils aus elf  $(x, y)$ -Punkten bestehen.
- Studieren Sie die Tabelle sorgfältig.** Wie verhalten sich  $x$  und  $y$  sowie ihr Verhältnis zueinander in den verschiedenen Datensätzen?

I.	II.	III.	IV.				
$y_1$	$x_1$	$y_2$	$x_2$	$y_3$	$x_3$	$y_4$	$x_4$
8.04	10	9.14	10	7.46	10	6.58	8
6.95	8	8.14	8	6.77	8	5.76	8
7.58	13	8.74	13	12.74	13	7.71	8
8.81	9	8.77	9	7.11	9	8.84	8
8.33	11	9.26	11	7.81	11	8.47	8
9.96	14	8.10	14	8.84	14	7.04	8
7.24	6	6.13	6	6.08	6	5.25	8
4.26	4	3.10	4	5.39	4	12.50	19
10.84	12	9.13	12	8.15	12	5.56	8
4.82	7	7.26	7	6.42	7	7.91	8
5.68	5	4.74	5	5.73	5	6.89	8

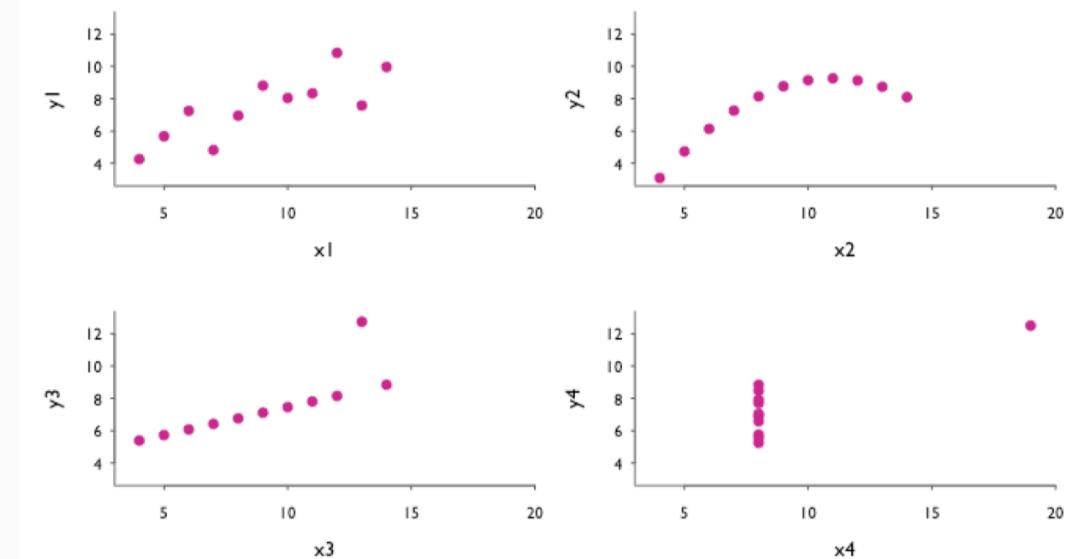
# Abbildungen vs. Tabellen

- Die Tabelle auf der rechten Seite enthält die Datensätze I bis IV, die jeweils aus elf  $(x, y)$ -Punkten bestehen.
- Studieren Sie die Tabelle sorgfältig.** Wie verhalten sich  $x$  und  $y$  sowie ihr Verhältnis zueinander in den verschiedenen Datensätzen?
- Es zeigt sich, dass alle Datensätze nahezu identische einfache deskriptive Statistiken in Bezug auf Mittelwert, Standardabweichung, Korrelation und lineare Anpassung aufweisen!

	I.	II.	III.	IV.				
	$y_1$	$x_1$	$y_2$	$x_2$	$y_3$	$x_3$	$y_4$	$x_4$
8.04	10	9.14	10	7.46	10	6.58	8	
6.95	8	8.14	8	6.77	8	5.76	8	
7.58	13	8.74	13	12.74	13	7.71	8	
8.81	9	8.77	9	7.11	9	8.84	8	
8.33	11	9.26	11	7.81	11	8.47	8	
9.96	14	8.10	14	8.84	14	7.04	8	
7.24	6	6.13	6	6.08	6	5.25	8	
4.26	4	3.10	4	5.39	4	12.50	19	
10.84	12	9.13	12	8.15	12	5.56	8	
4.82	7	7.26	7	6.42	7	7.91	8	
5.68	5	4.74	5	5.73	5	6.89	8	
Mean( $y$ )		7.50		7.5		7.50		7.5
Mean( $x$ )		9.0		9.0		9.0		9.0
SD( $y$ )		2.03		2.03		2.03		2.03
SD( $x$ )		3.32		3.32		3.32		3.32
$r(y, x)$		.82		.82		.82		.82
$y = a + bx$		$y = 3 + 0.5x$						
$R^2$		.67		.67		.67		.67

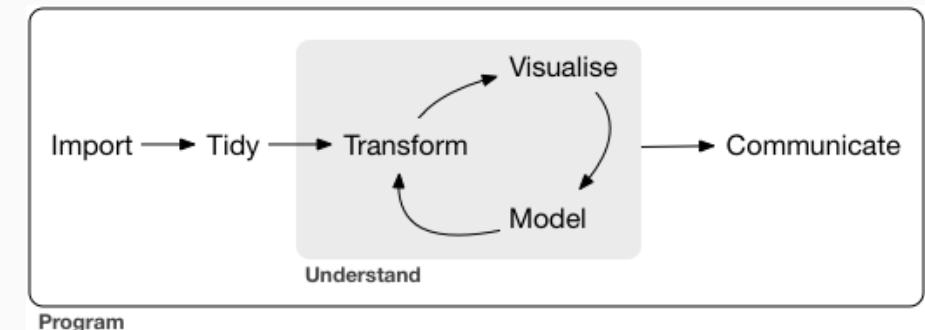
# Abbildungen vs. Tabellen

- Die Tabelle auf der rechten Seite enthält die Datensätze I bis IV, die jeweils aus elf  $(x, y)$ -Punkten bestehen.
- **Studieren Sie die Tabelle sorgfältig**. Wie verhalten sich  $x$  und  $y$  sowie ihr Verhältnis zueinander in den verschiedenen Datensätzen?
- Es zeigt sich, dass alle Datensätze nahezu identische einfache deskriptive Statistiken in Bezug auf Mittelwert, Standardabweichung, Korrelation und lineare Anpassung aufweisen!
- **Die grafische Darstellung der Daten zeigt sehr unterschiedliche Verteilungen** und widerspricht der Heuristik, dass "numerische Berechnungen genau sind, aber Grafiken grob sind". ([Anscombe 1973](#)).



# Visualisierung im Data-Science-Workflow

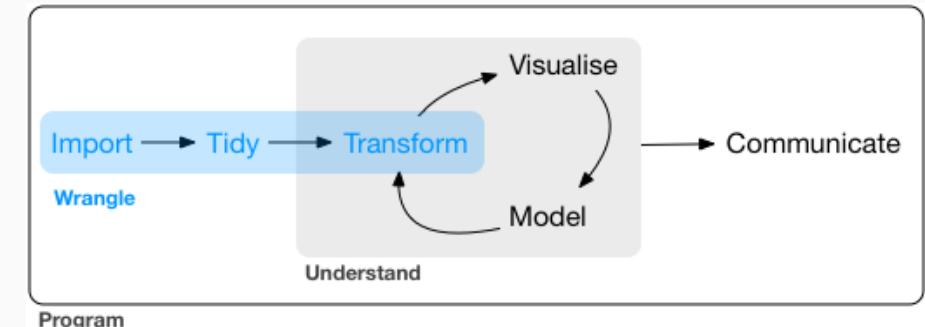
Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.



Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Datenaufbereitung

- Sanity checks
- Identifizierung von Ausreißern
- Unterstützung von Data cleaning

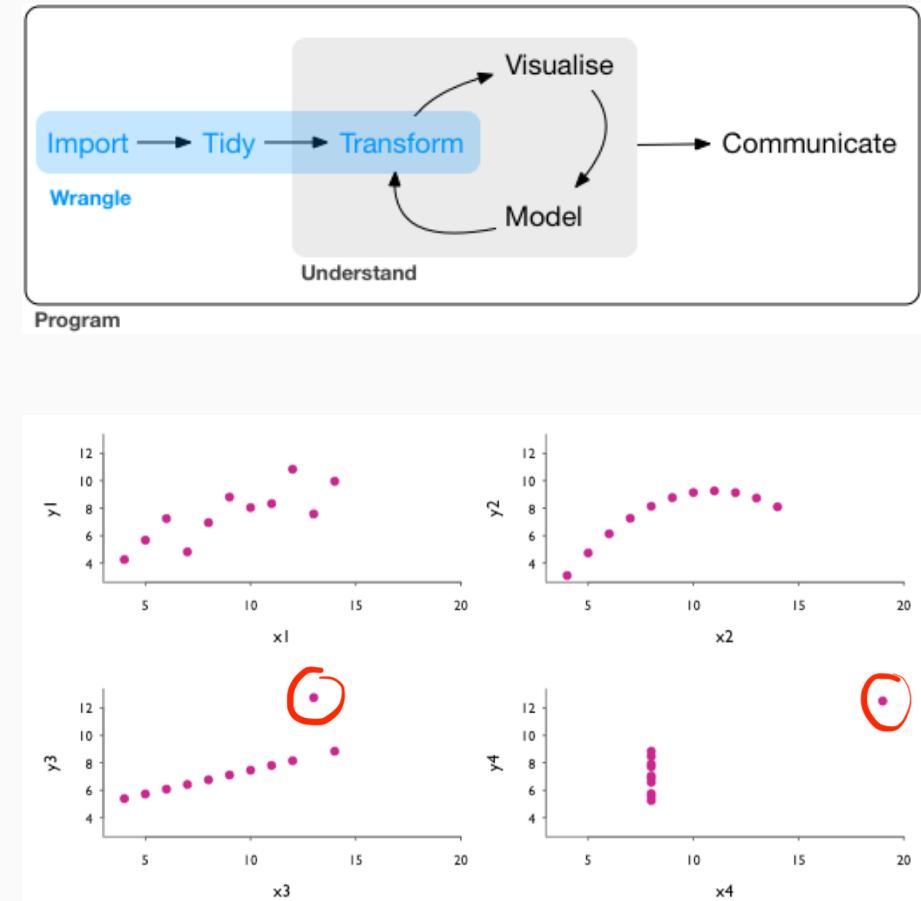


# Visualisierung im Data-Science-Workflow

Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Datenaufbereitung

- Sanity checks
- Identifizierung von Ausreißern
- Unterstützung von Data cleaning



Scatter plots to identify outliers in bivariate relationships.

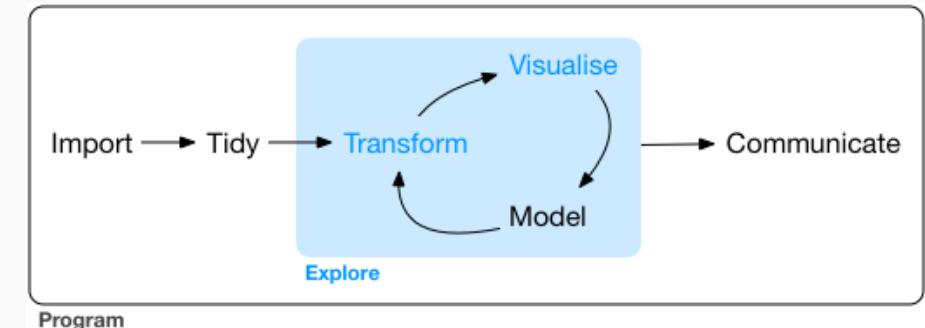
Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Datenaufbereitung

- Sanity checks
- Identifizierung von Ausreißern
- Unterstützung von Data cleaning

## Datenexploration

- Darstellung von Verteilungen
- Muster und Beziehungen entdecken



# Visualisierung im Data-Science-Workflow

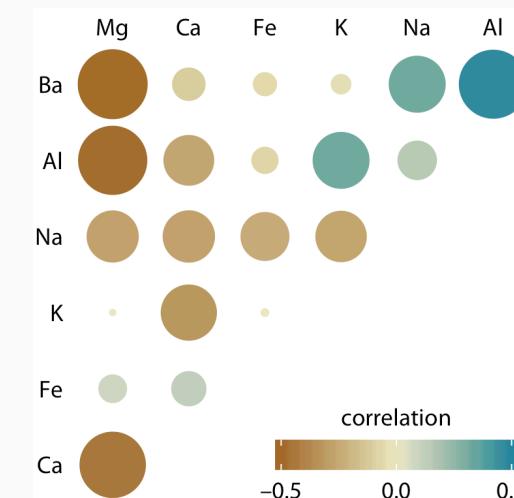
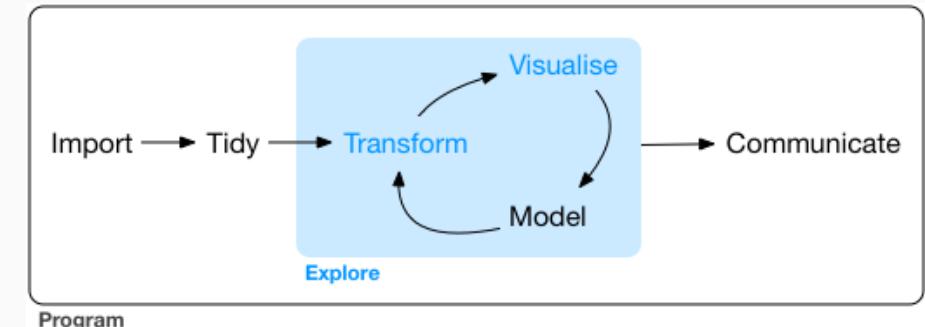
Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Datenaufbereitung

- Sanity checks
- Identifizierung von Ausreißern
- Unterstützung von Data cleaning

## Datenexploration

- Darstellung von Verteilungen
- Muster und Beziehungen entdecken

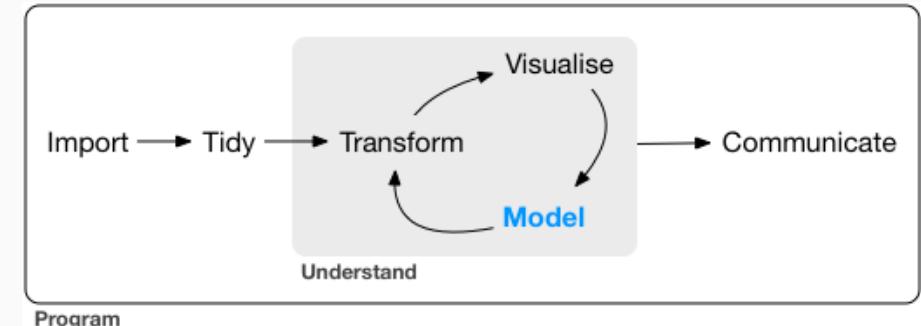


Correlogram to visualize amount of association  
between pairs of variables

Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Modellieren

- Hypothesen testen
- Zusammenfassen von (mehrfachen) Modellschätzungen
- Visualisierung der Unsicherheit
- Robustheits-/Sensitivitätsanalysen berichten

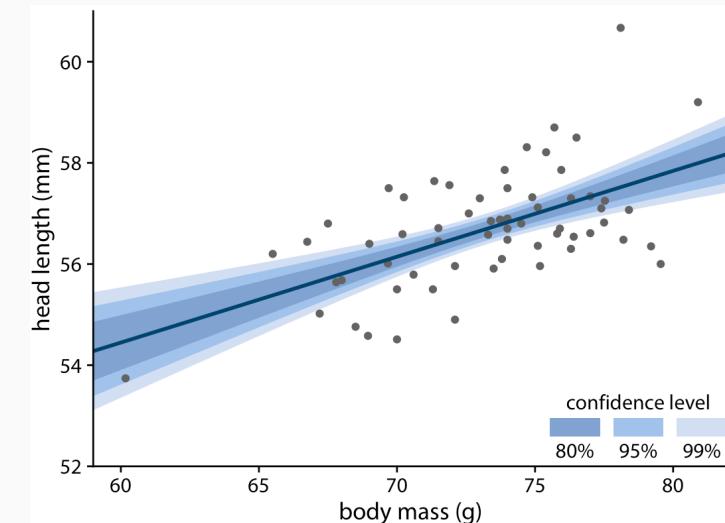
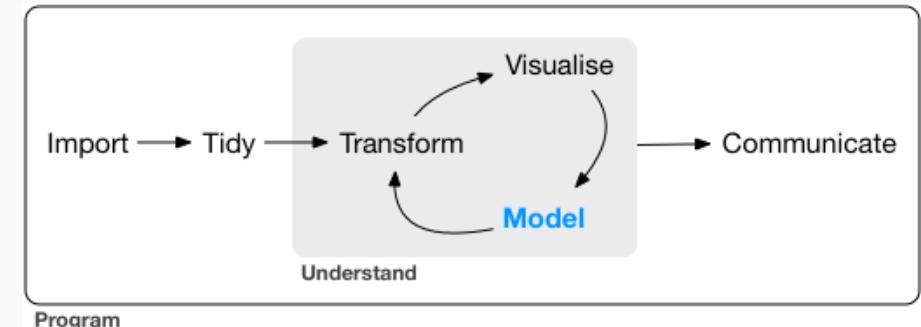


# Visualisierung im Data-Science-Workflow

Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Modellieren

- Hypothesen testen
- Zusammenfassen von (mehrfachen) Modellschätzungen
- Visualisierung der Unsicherheit
- Robustheits-/Sensitivitätsanalysen berichten

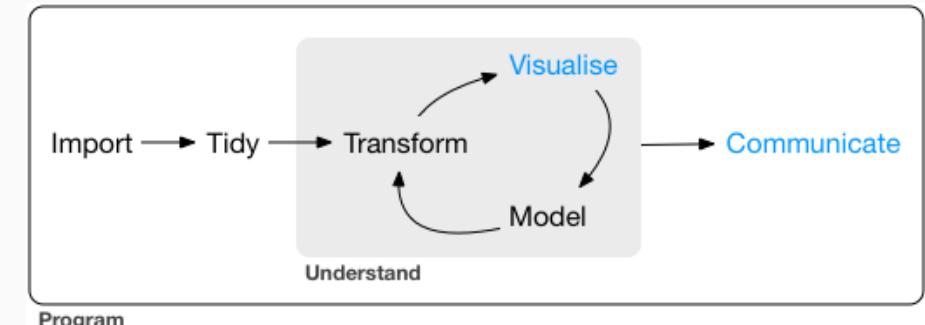


Raw data and trend line with confidence bands  
to visualize uncertainty of fit

Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Modellieren

- Hypothesen testen
- Zusammenfassen von (mehrfachen) Modellschätzungen
- Visualisierung der Unsicherheit
- Robustheits-/Sensitivitätsanalysen berichten



## Kommunizieren

- Präsentation von Rohdaten/aufbereiteten Daten
- Darstellung der Auswirkungen von Modellergebnissen

# Visualisierung im Data-Science-Workflow

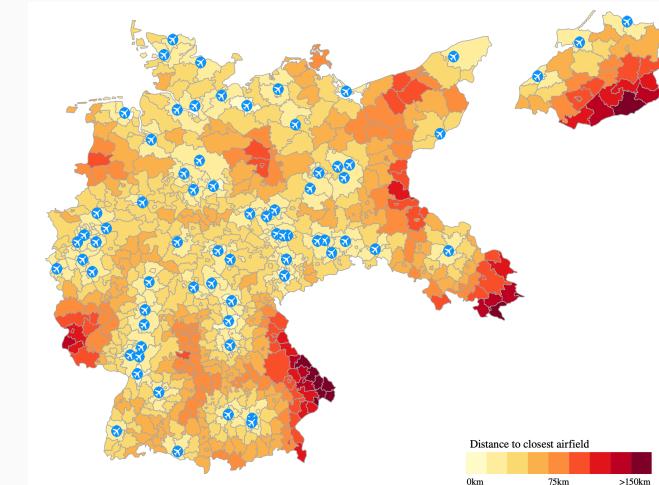
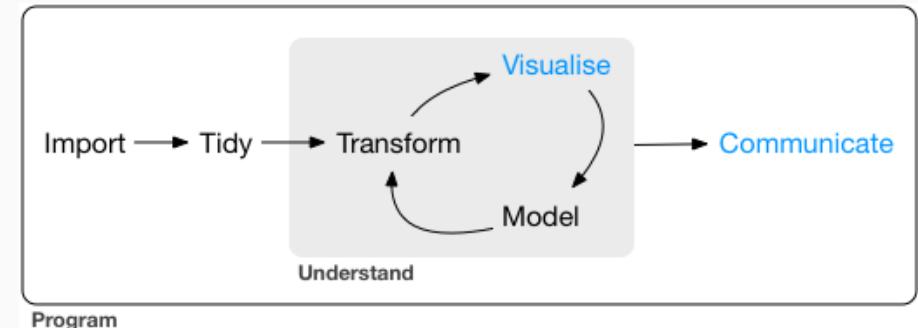
Datenvisualisierung ist eine Schlüsselkompetenz für die Vermittlung von Erkenntnissen aus Daten. Sie ist in jedem Schritt des Workflows von Bedeutung.

## Modellieren

- Hypothesen testen
- Zusammenfassen von (mehrfachen) Modellschätzungen
- Visualisierung der Unsicherheit
- Robustheits-/Sensitivitätsanalysen berichten

## Kommunizieren

- Präsentation von Rohdaten/aufbereiteten Daten
- Darstellung der Auswirkungen von Modellergebnissen



Choroplethenkarte mit der Lage der zivilen Flugplätze im Deutschen Reich, 1932. Die Verwaltungsbezirke sind entsprechend der Entfernung ihres Schwerpunkts zum nächstgelegenen Flugplatz schattiert 18 / 60

# Diskussion

Warum Datenvisualisierung in der / durch die Verwaltung?

1. (Wie) kann die Visualisierung von Daten in Ihrem Arbeitskontext nützlich sein?
2. Glauben Sie, dass Datenvisualisierung die Kommunikation zwischen verschiedenen Abteilungen und mit der Öffentlichkeit verbessern kann?

# Arten der Datenvisualisierung

---

Ein häufiger Fehler bei der Visualisierung ist, dass **Plot-Typen für Zwecke verwendet werden, für die sie nicht gedacht sind**. Bevor Sie mit dem Abbildungen erstellen, fragen Sie sich selbst:

**1. Was genau möchte ich visualisieren?**

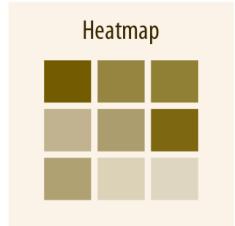
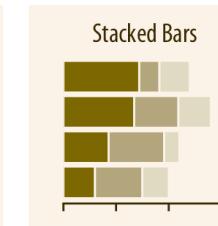
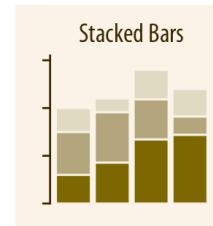
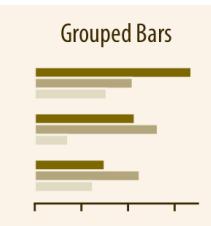
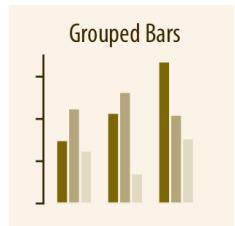
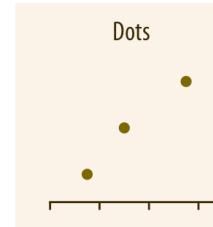
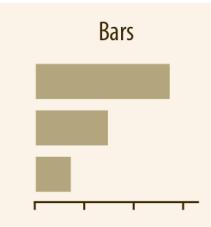
- Zahlen und Mengen
- Verteilungen
- Proportionen
- Assoziationen
- Strukturen
- Trends
- Schätzungen
- Vorhersagen
- Unsicherheit

**2. Welche Frage möchte ich beantworten?**

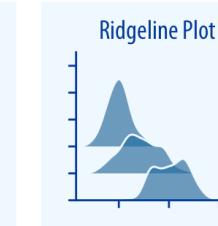
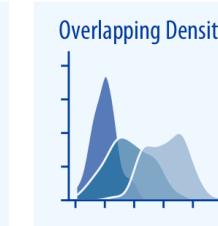
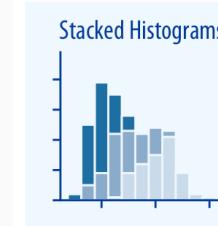
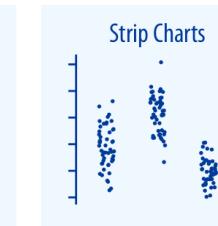
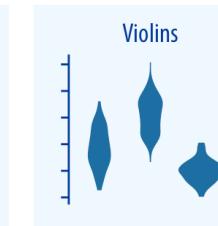
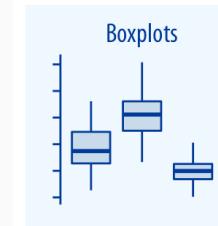
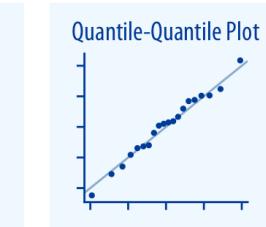
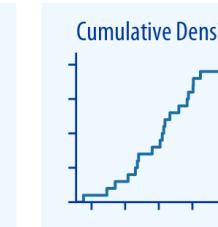
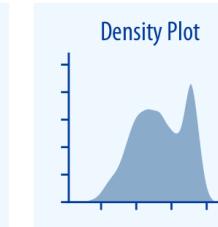
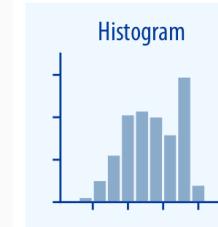
- "Ist die *Verteilung* normal (oder gleichmäßig oder...)??" → Histogramm, Dichteplot
- "Sind *Verteilungen* zwischen Gruppen unterschiedlich?" → Boxplots, Ridgelines
- "Wie unterscheiden sich die *Zahlen* zwischen den Gruppen?" → Balkenplot, Dotplot
- "Wie ist die *Beziehung* zwischen x und y?" → Scatterplot, Konturplot
- Wie *korrelieren* in einer Reihe von Variablen?" → Korrelogramm, small multiples
- "Wie hat sich ein *Trend* im Laufe der Zeit entwickelt?" → Liniendiagramm
- "Sind die Daten nach Untergruppen *geclustert*?" → Scatterplot mit Farbe
- "Gibt es ein *räumliches Muster*?" → Choropleth, Kartogramm-Heatmap
- "Wie groß sind die relativen und absoluten *Effekte*?" → Koeffizienten-Plot
- Wie unsicher sind die *Schätzungen*? → Fehlerbalken, Vertrauensbereiche

# Ein Verzeichnis von Visualisierungen

## Visualisierung von Mengen



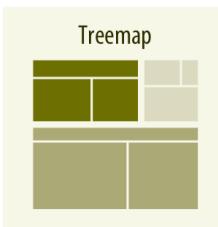
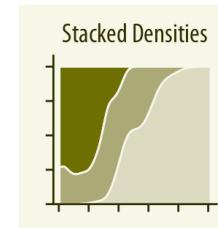
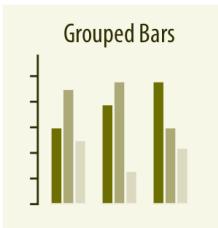
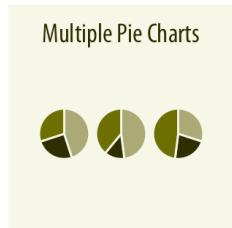
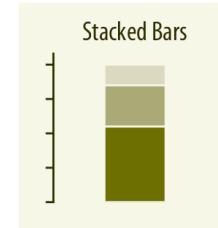
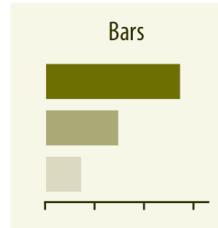
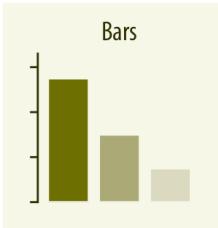
## Visualisierung von Verteilungen



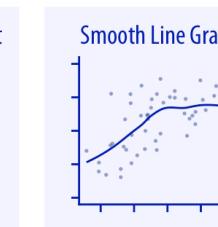
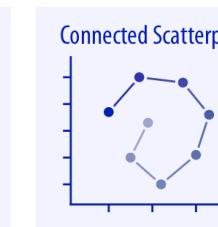
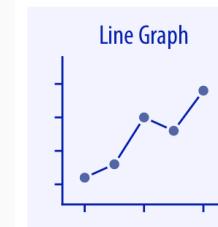
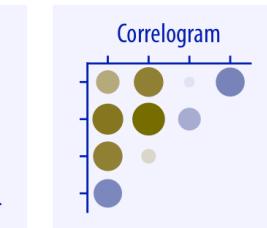
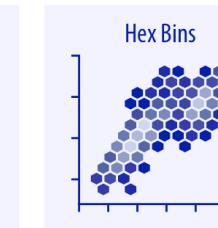
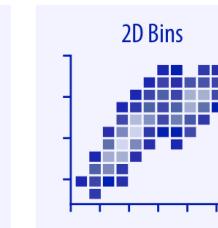
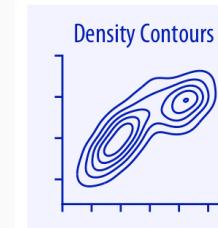
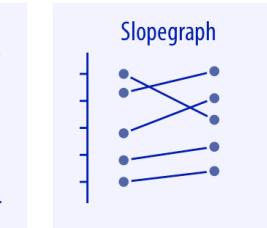
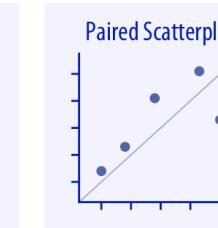
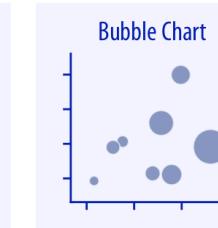
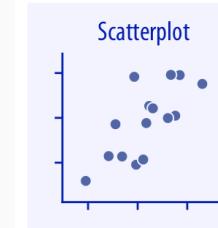
Quelle: Claus Wilke, Fundamentals of Data Visualization

# Ein Verzeichnis von Visualisierungen

## Visualisierung von Proportionen



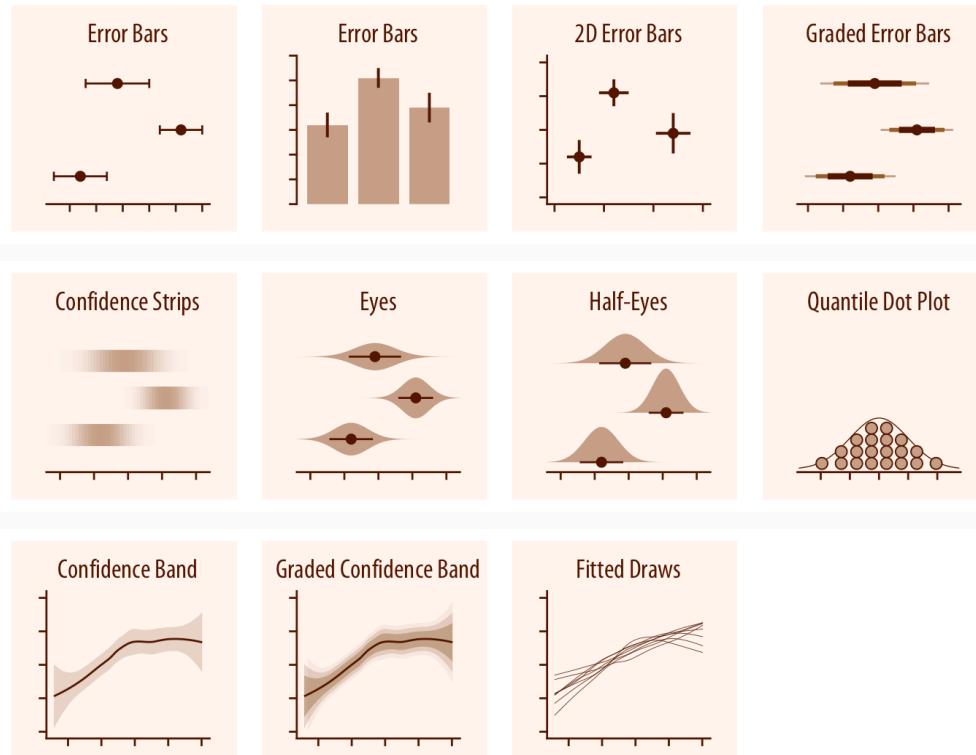
## Visualisierung von x-y-Beziehungen



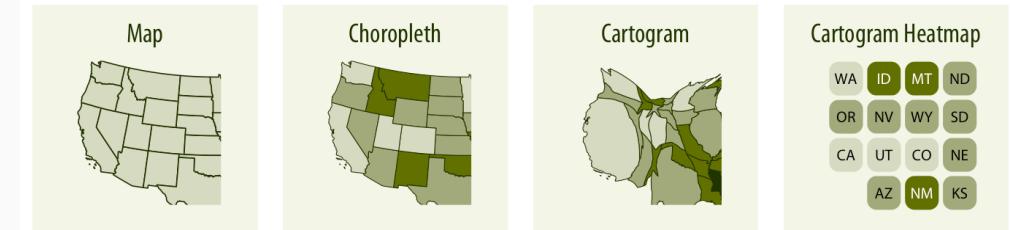
Quelle: Claus Wilke, Fundamentals of Data Visualization

# Ein Verzeichnis von Visualisierungen

## Visualisierung von Unsicherheit



## Visualisierung räumlicher Daten



Quelle: Claus Wilke, Fundamentals of Data Visualization

# Visualisierungstools

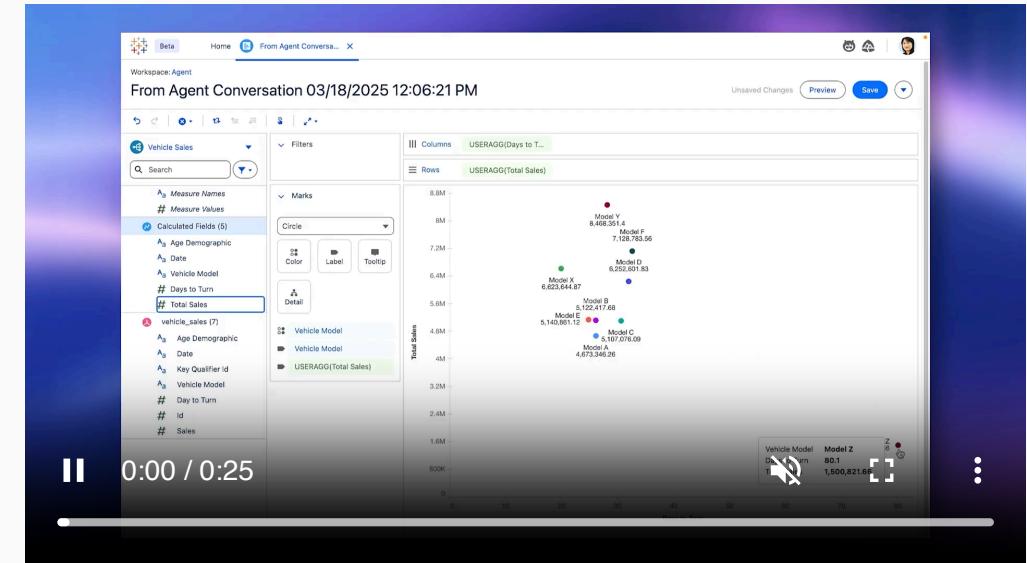
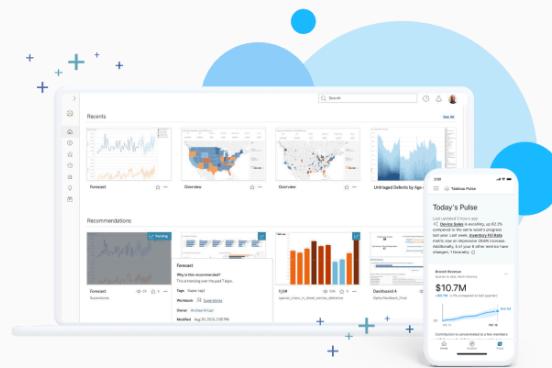
---

# Klick- vs. codebasierte Visualisierungstools

## Klick-basierte Tools

Diese Tools sind **benutzerfreundliche Anwendungen**, für die man keine Programmierkenntnisse braucht. Sie *klicken einfach auf Schaltflächen, ziehen Elemente und wählen Optionen aus, um Visualisierungen zu erstellen.*

Diese Anwendungen sind ideal, wenn Sie schnell Diagramme und Grafiken erstellen möchten, ohne sich in die Feinheiten der Programmierung einarbeiten zu müssen. "Malen-nach-Zahlen"-Tools.



Quelle [Tableau/YouTube](#)

# Klick- vs. codebasierte Visualisierungstools

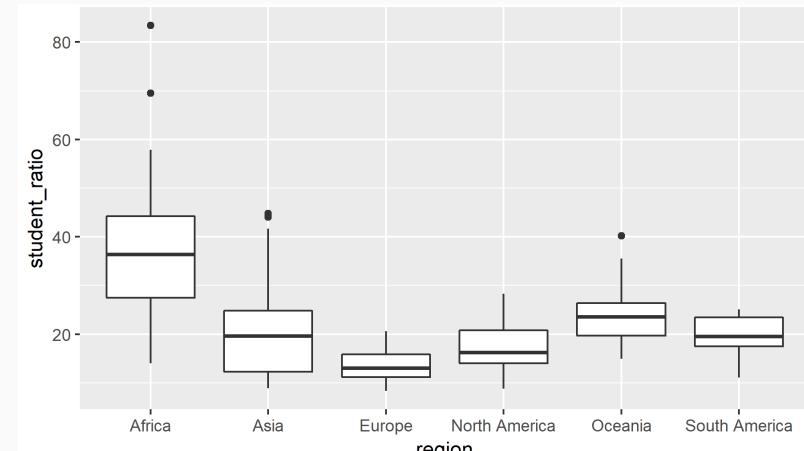
## Code-basierte Tools

Tools für Künstler, die lieber von Grund auf malen.

Anstatt auf Schaltflächen zu klicken, skripten Sie Grafiken *in einer Programmiersprache* wie Python oder R, um Ihre Visualisierungen zu erstellen. Das bedeutet große **Flexibilität** und **Kontrolle** über jedes Detail. Steile Programmierlernkurve - eher nur für Profis.



```
R> df_sorted <-  
+   df_ratios %>%  
+   mutate(region =  
+           fct_reorder(region,  
+                           -student_ratio_region))  
R> ggplot(df_sorted,  
+           aes(x = region, y = student_ratio)) +  
+   geom_boxplot()
```



Data: UNESCO Institute for Statistics  
Visualization by Cédric Scherer

Quelle und Code Cedric Scherer

## Vorteile

1. **Benutzerfreundlichkeit:** Klick-basierte Tools wie Tableau oder Power BI bieten **intuitive Oberflächen**, die es den Benutzern leicht machen, Visualisierungen ohne Programmierkenntnisse zu erstellen.
2. **Rapid Prototyping:** Benutzer können mit **schnell Prototypen erstellen** und Visualisierungen mit klickbasierten Tools überarbeiten, was eine schnellere Erkundung von Daten und Erkenntnissen ermöglicht.
3. **Interaktive Funktionen:** Klickbasierte Tools verfügen oft über **integrierte interaktive Funktionen**, die es den Benutzern ermöglichen, Daten dynamisch zu erforschen, z. B. durch Filtern, Aufschlüsseln oder Zoomen auf bestimmte Datenpunkte.
4. **Vorgefertigte Vorlagen:** Klickbasierte Tools bieten oft eine Vielzahl von **vorgefertigten Vorlagen** und Visualisierungen, die den Benutzern Zeit und Mühe bei der Gestaltung und Formatierung ersparen.

## Nachteile

1. **Eingeschränkte Anpassungsmöglichkeiten:** Klickbasierte Tools können im Vergleich zu codebasierten Tools Einschränkungen bei der Anpassung aufweisen, was die Möglichkeiten der Benutzer zur Erstellung von hochgradig maßgeschneiderten Visualisierungen einschränkt.
2. **Skalierbarkeitsprobleme:** Klick-basierte Tools können **Schwierigkeiten mit großen Datensätzen** oder komplexen Visualisierungen haben, was zu Leistungsproblemen oder Einschränkungen in der Funktionalität führt.
3. **Anbieterabhängigkeit:** Benutzer können von bestimmten klickbasierten Tools abhängig werden, was zu einer Anbieterabhängigkeit und potenziellen **Einschränkungen bei der Flexibilität oder Kompatibilität** mit anderen Tools führt.
4. **Kosten:** Klickbasierte Tools sind oft mit **Lizenzgebühren oder Abonnementkosten** verbunden, was für Einzelpersonen oder Organisationen mit begrenztem Budget unerschwinglich sein kann.

## Vorteile

1. **Flexibilität:** Mit codebasierten Werkzeugen wie "ggplot2" haben die Benutzer die volle Kontrolle über die Anpassung, so dass sie hochgradig individuelle und komplizierte Visualisierungen erstellen können.
2. **Reproduzierbarkeit:** Code-basierte Tools erleichtern die Reproduzierbarkeit, da der Code leicht weitergegeben und wiederverwendet werden kann, was **konsistente Ergebnisse in verschiedenen Umgebungen** gewährleistet.
3. **Skalierbarkeit:** Diese Tools eignen sich oft besser für den Umgang mit **großen Datensätzen und komplexen Visualisierungen**, da sie fortgeschrittenere Programmierungsmöglichkeiten bieten.
4. **Integration:** Codebasierte Tools können nahtlos in Datenanalyse-Workflows integriert werden, was eine direkte Manipulation von Daten und die Integration in statistische Analysen ermöglicht.

## Nachteile

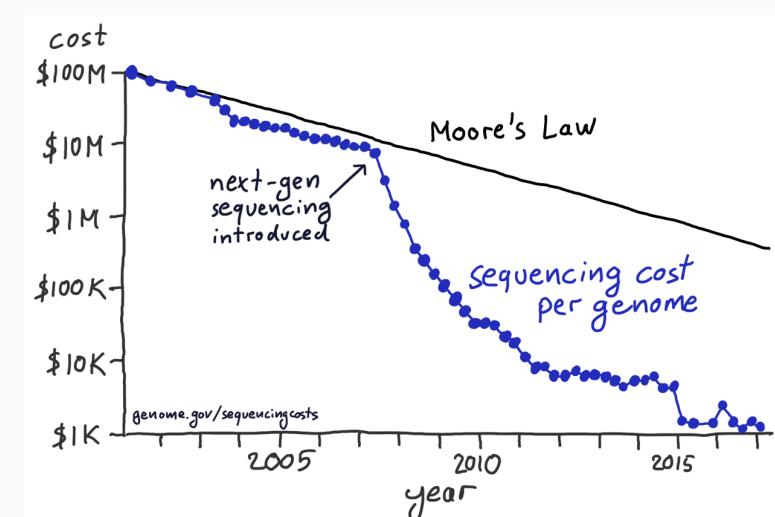
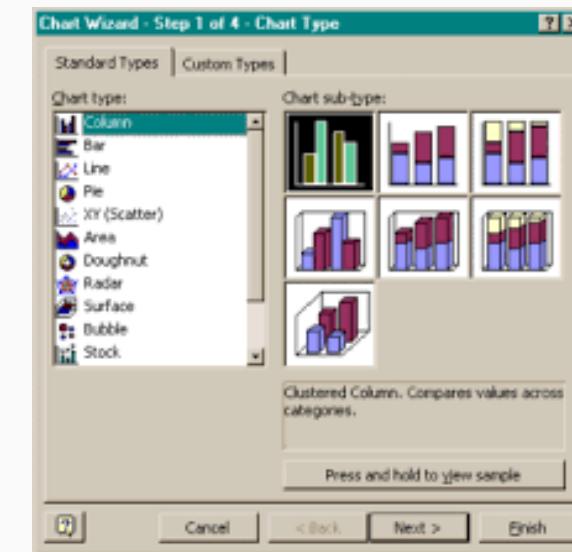
1. **Lernkurve:** Für Benutzer ohne Programmiererfahrung kann es schwierig sein, die Syntax und die Konzepte zu erlernen, die erforderlich sind, um codebasierte Visualisierungswerkzeuge effektiv zu nutzen.
2. **Zeitaufwändig:** Visualisierungen von Grund auf mit Code zu erstellen, kann **zeitaufwendig** sein, besonders für Benutzer, die nicht über Programmierkenntnisse verfügen.
3. **Komplexität bei der Fehlersuche:** Die Fehlersuche bei codebasierten Visualisierungen kann im Vergleich zu klickbasierten Tools komplexer sein, insbesondere wenn es um komplizierte Darstellungen oder Fehler im Code geht.
4. **Wartung:** Codebasierte Visualisierungen können im Laufe der Zeit mehr **Wartung und Updates** erfordern, insbesondere wenn sich Bibliotheken oder Abhängigkeiten ändern.

# Vergleich

Tool / Plattform	Stärken	Schwächen	Lizenz / Kosten	Eignung für Verwaltung / Behörden	Typische Einsatzszenarien
R (ggplot2, plotly, Shiny)	<ul style="list-style-type: none"> <li>• Hohe Flexibilität</li> <li>• Reproduzierbare Analysen</li> <li>• Automatisierung möglich</li> <li>• Interaktive Dashboards (Shiny)</li> </ul>	<ul style="list-style-type: none"> <li>• Einarbeitung erforderlich</li> <li>• Deployment kann komplex sein</li> </ul>	Open Source	Sehr gut für datenaffine Behörden/Statistikstellen	Forschung, Monitoring, Open-Data-Dashboards
Python (matplotlib, seaborn, plotly, Dash)	<ul style="list-style-type: none"> <li>• Breites Ökosystem</li> <li>• Gut für KI/ML</li> <li>• Dash für interaktive Reports</li> <li>• Gute Webintegration</li> </ul>	<ul style="list-style-type: none"> <li>• Programmierung nötig</li> <li>• Pflegeaufwand bei Dashboards</li> </ul>	Open Source	Gut für Digitalisierungs-/Daten-Teams	Automatisierte Berichte, Datenpipelines, Web-APIs
Tableau	<ul style="list-style-type: none"> <li>• Intuitive Bedienung</li> <li>• Hochwertige Visuals</li> <li>• Gute DB-Anbindung</li> <li>• Governance-Funktionen</li> </ul>	<ul style="list-style-type: none"> <li>• Lizenzkosten</li> <li>• Proprietär</li> <li>• Begrenzte algorithmische Erweiterbarkeit</li> </ul>	Kostenpflichtig (~70 €/User/Monat)	Sehr gut für Berichtswesen/Management	Open-Data-Visuals, Controlling, Dashboards
Power BI	<ul style="list-style-type: none"> <li>• MS-Integration (Excel/Teams/Azure)</li> <li>• Gute Governance/Security</li> <li>• Einfache Intranet-Publikation</li> </ul>	<ul style="list-style-type: none"> <li>• Eingeschränkte Individualisierung</li> <li>• Proprietär</li> <li>• Performance-Limits bei großen Daten</li> </ul>	Kostenpflichtig (ab ~10 €/User/Monat)	Ideal bei bestehender MS-Infrastruktur	KPI-Monitoring, Personal-/Finanzberichte
Excel (Diagramme, Power Query, Power Pivot)	<ul style="list-style-type: none"> <li>• Niedrige Einstiegshürde</li> <li>• Weit verbreitet</li> <li>• Power Query / Pivot</li> <li>• Gut für kleinere Datensätze</li> </ul>	<ul style="list-style-type: none"> <li>• Begrenzte Interaktivität</li> <li>• Fehleranfällig bei manueller Pflege</li> <li>• Wenig Automatisierung</li> </ul>	Teil von Microsoft 365	Sehr gut für Fachbereiche/Basisberichte	Ad-hoc-Analysen, Standardberichte, Haushalt
Highcharts / Highmaps	<ul style="list-style-type: none"> <li>• Interaktive Webcharts</li> <li>• Einfache Einbindung</li> <li>• Barrierearme Optionen</li> </ul>	<ul style="list-style-type: none"> <li>• Lizenzkosten für kommerzielle Nutzung</li> <li>• Programmierung nötig</li> </ul>	Kostenlos nicht-kommerziell, sonst Lizenz	Gut für Web-/Open-Data-Portale	Öffentliche Dashboards, Karten

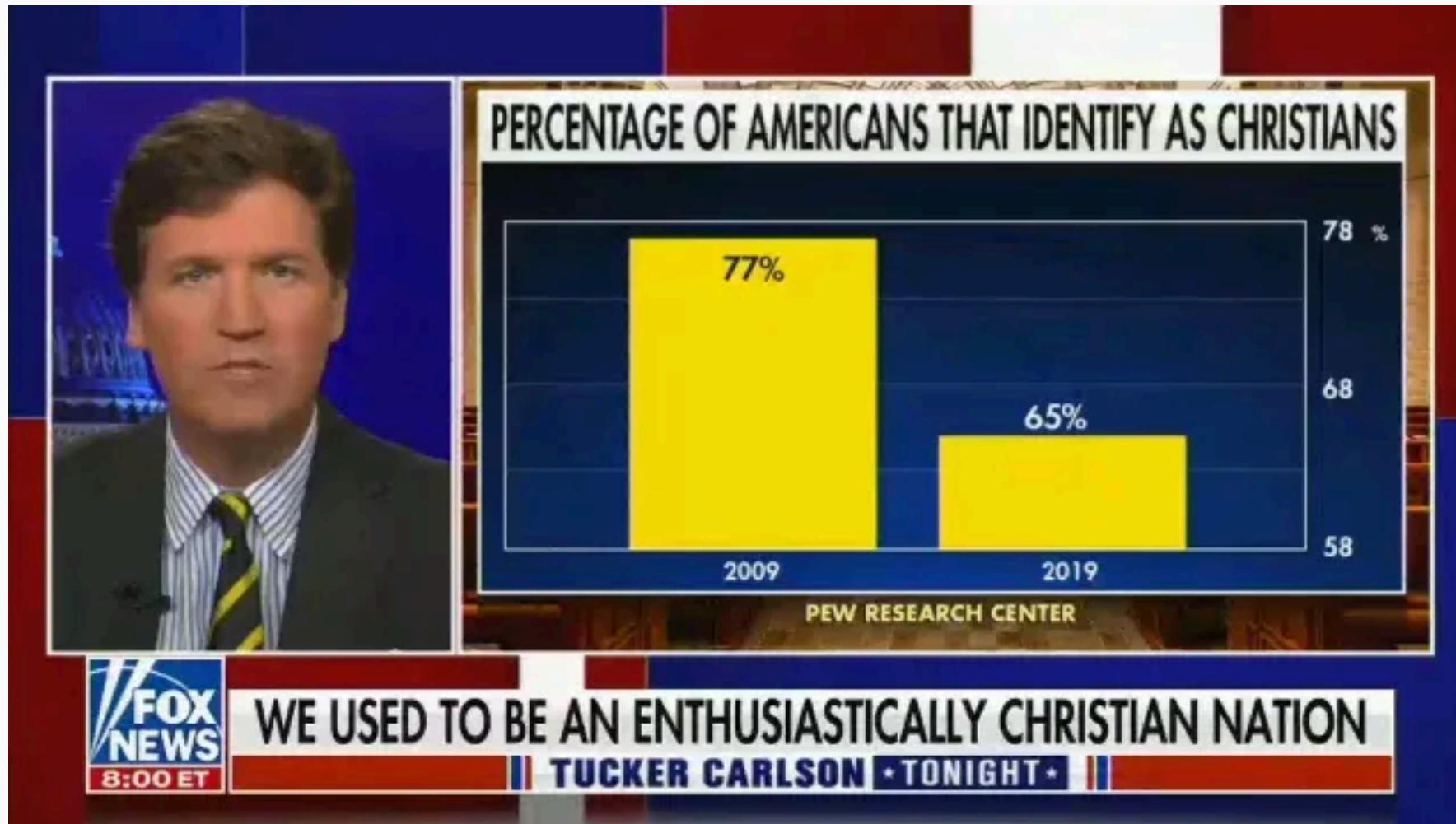
# Einige Hinweise für die richtige Visualisierung

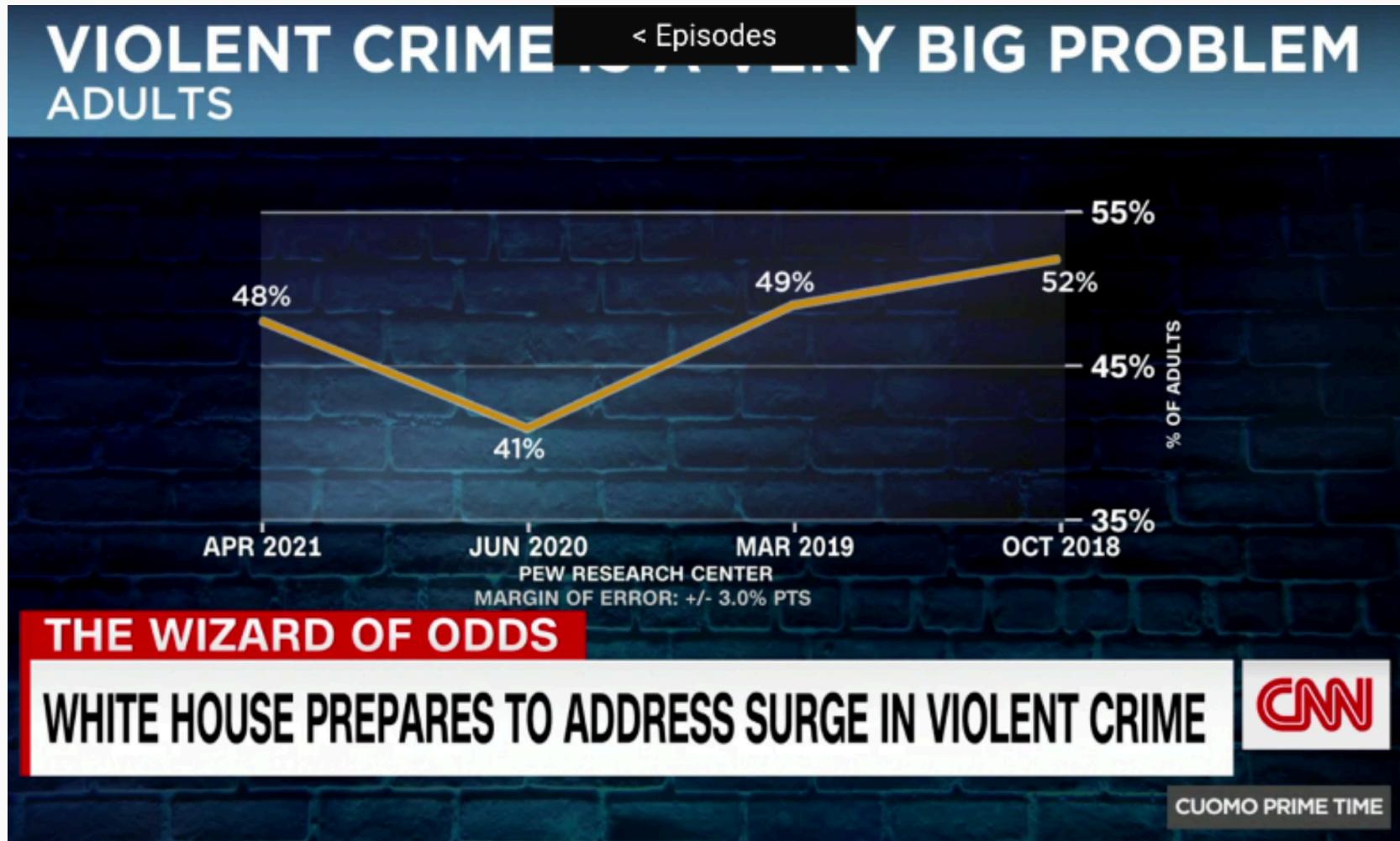
1. Immer zuerst einen Plan entwickeln.
2. Ziehen Sie auch code-basierte Lösungen in Betracht, um reproduzierbar zu bleiben (d.h. R/Python vor Excel).
3. Für konzeptionelle Diagramme (nicht: Datenvisualisierung!) können andere Werkzeuge gut geeignet sein (z.B. Powerpoint oder sogar handgezeichnete Figuren).
4. Lassen Sie sich nicht von interaktiven Elementen ablenken (wie sie z. B. von highcharts.com, Tableau und anderen angeboten werden).
5. Eine gute Visualisierung braucht Zeit, selbst wenn man erfahren ist.

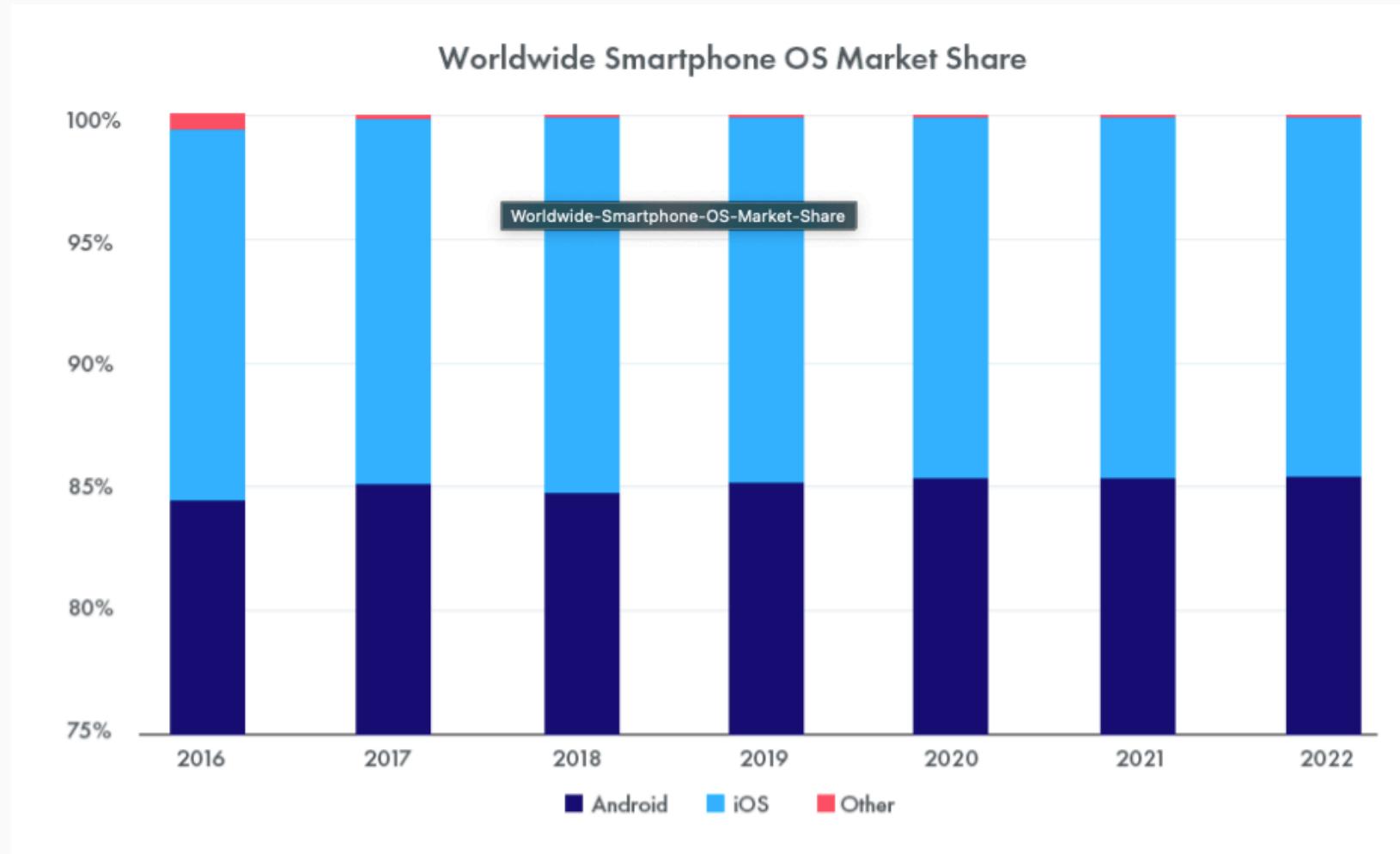


## **Übung: Best (and worst) practices**

---

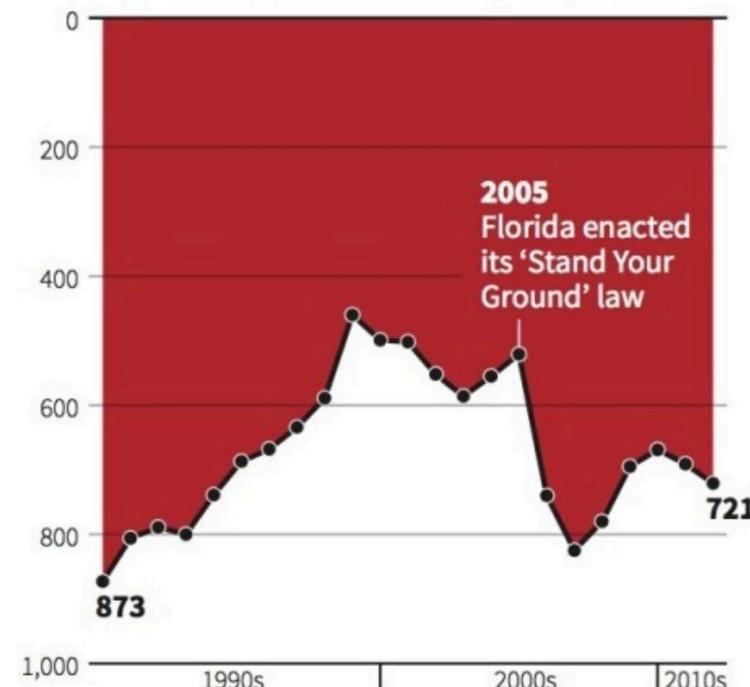






## Gun deaths in Florida

Number of murders committed using firearms

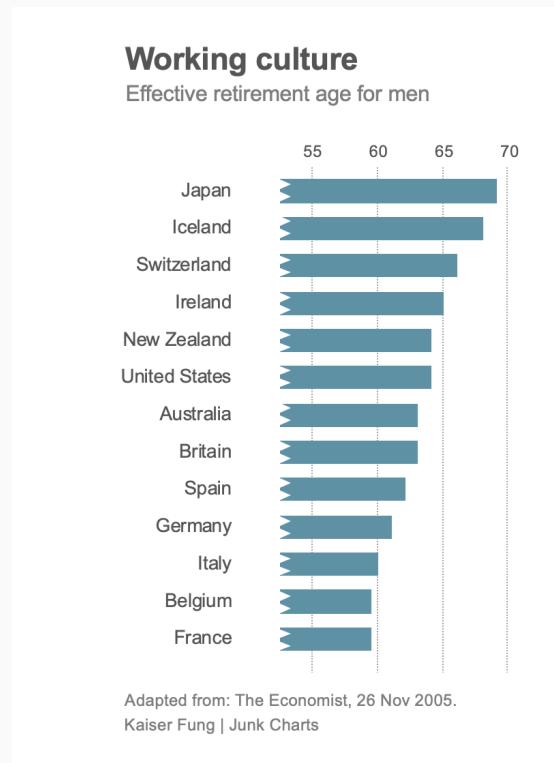


Source: Florida Department of Law Enforcement

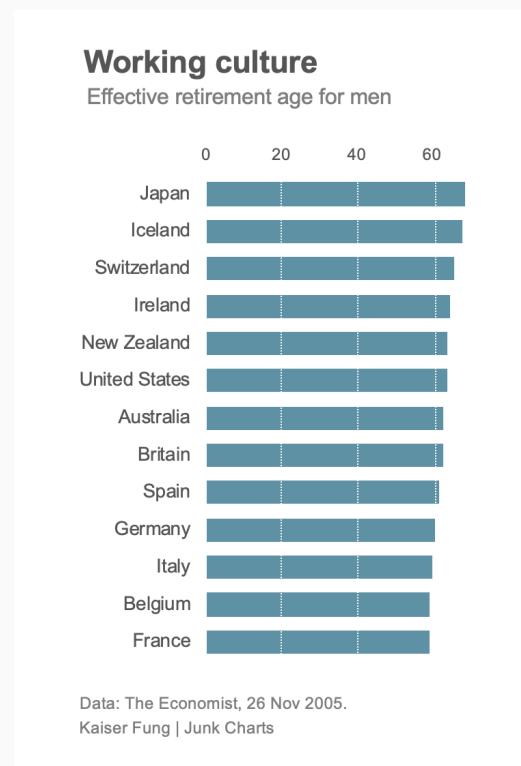
C. Chan 16/02/2014

REUTERS

# Todsünde der Datenvisualisierung: Achsentrunkierung

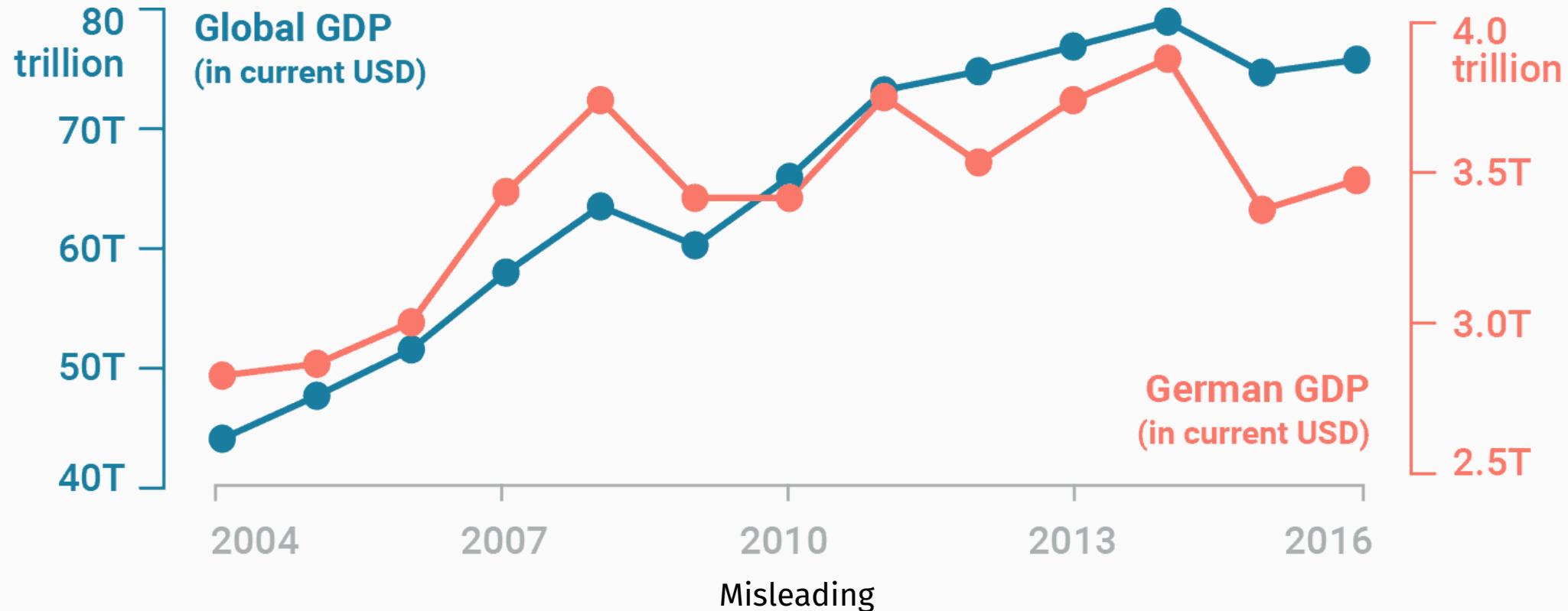


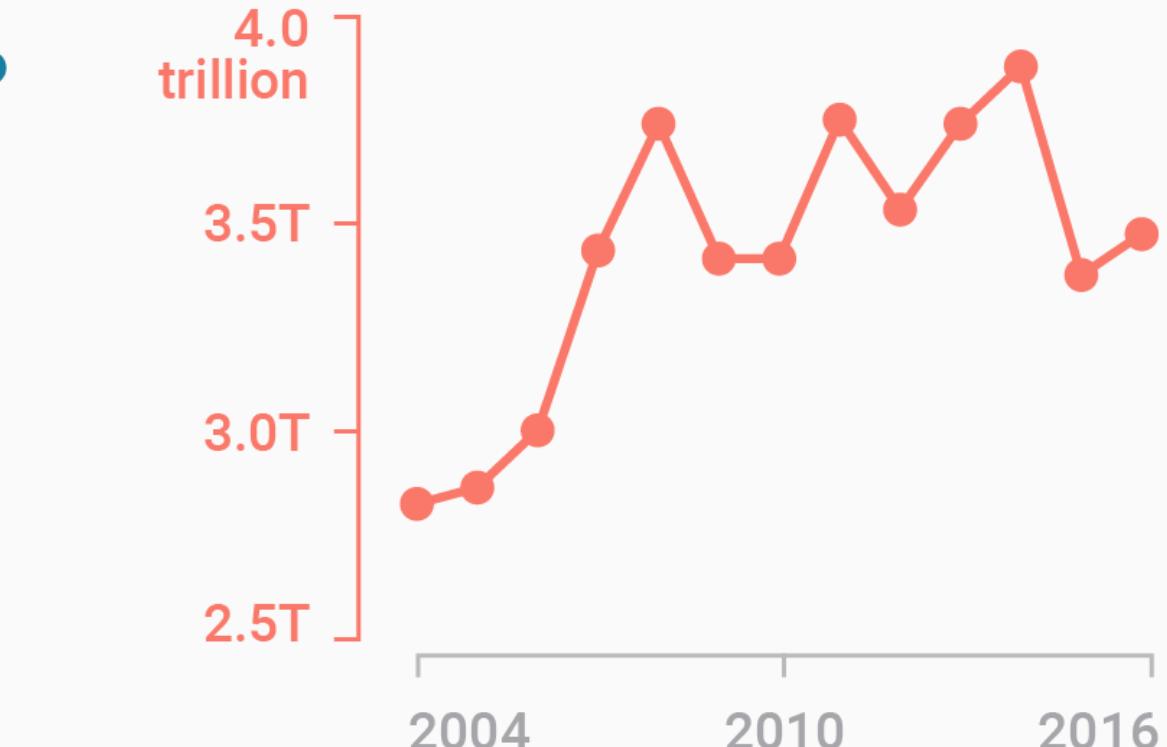
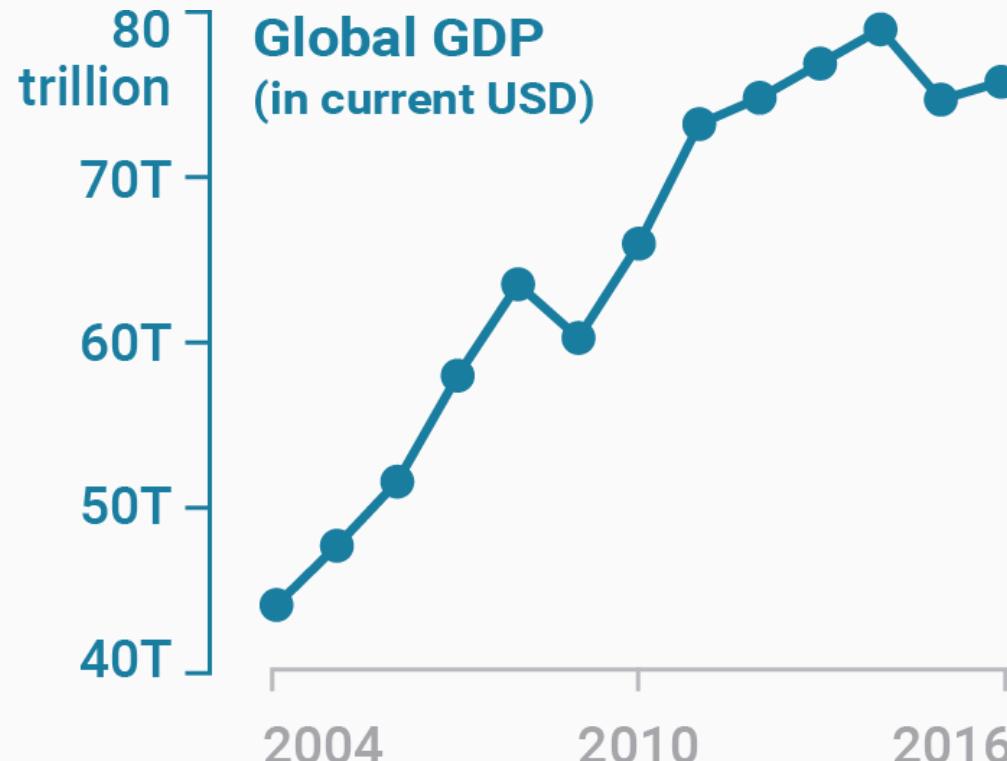
Misleading



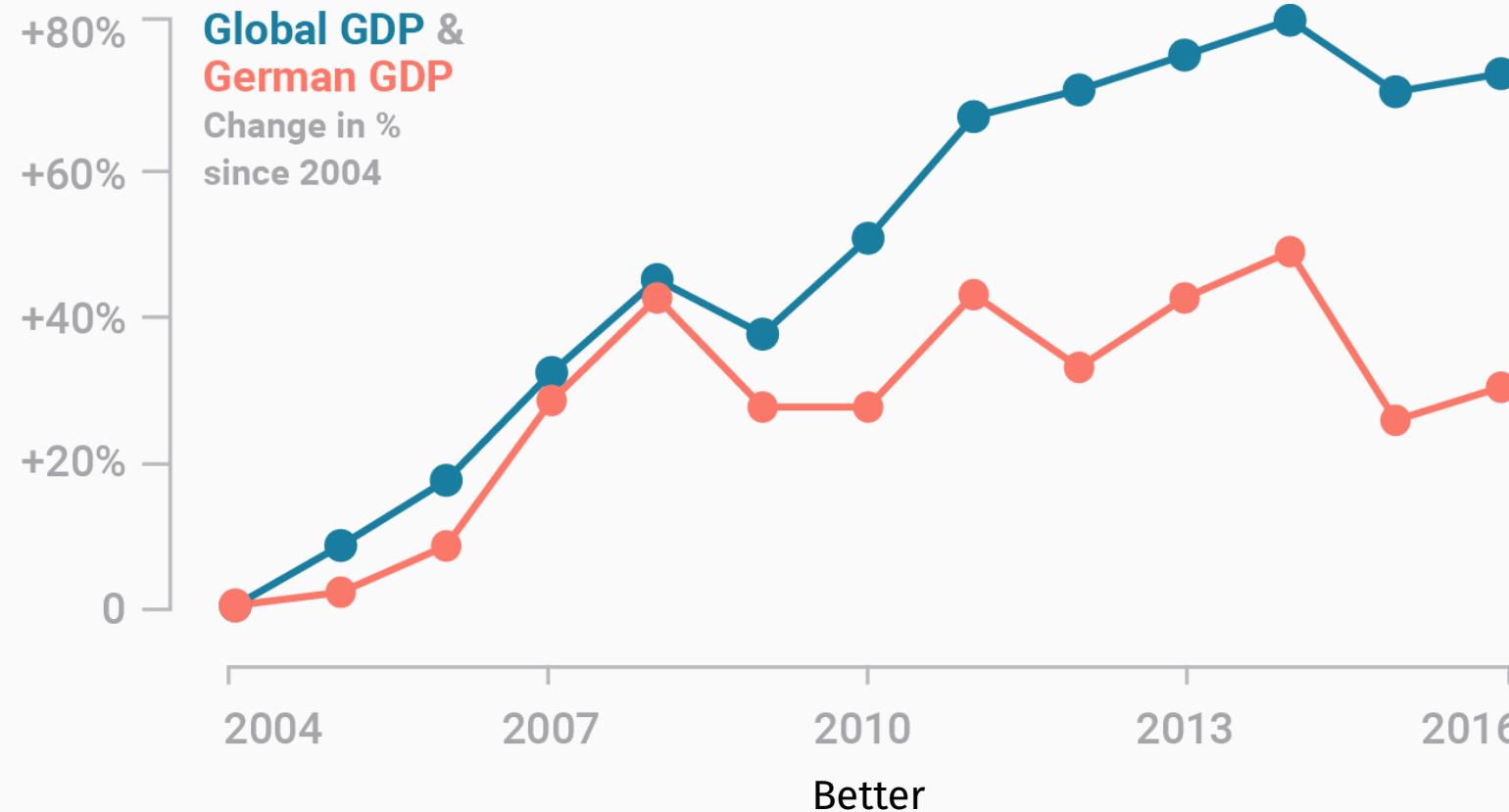
Better

# Todsünde der Datenvisualisierung: Doppelte Achsen

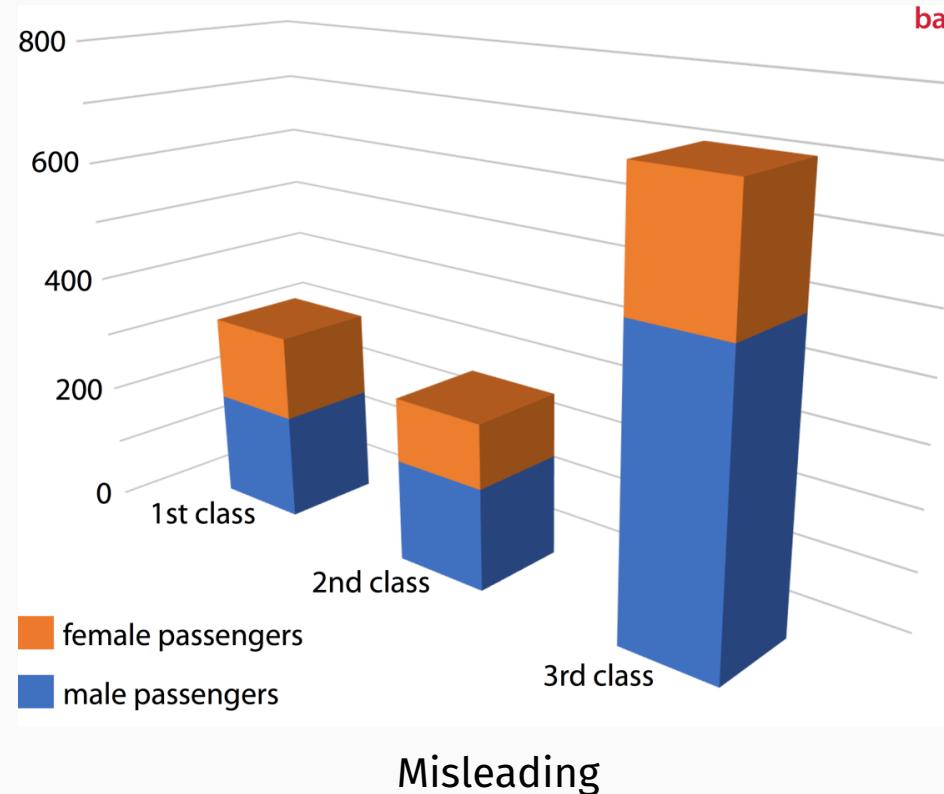


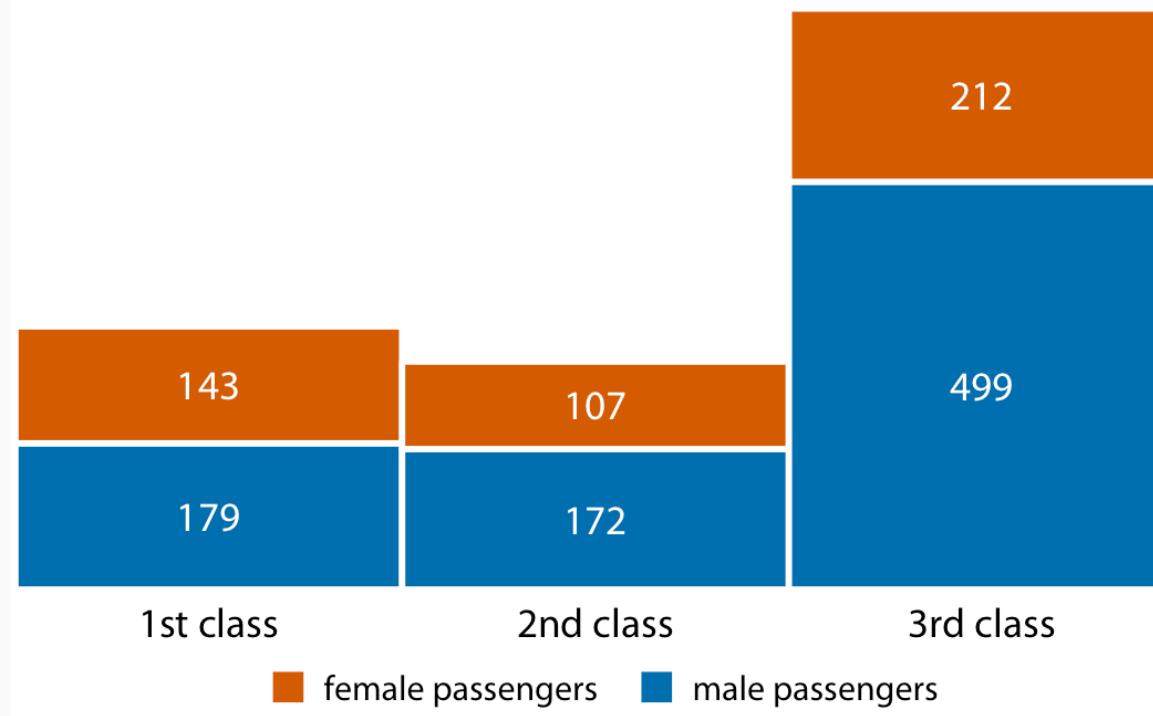


Better



# Todsünde der Datenvisualisierung: sinnlose 3D-Elemente

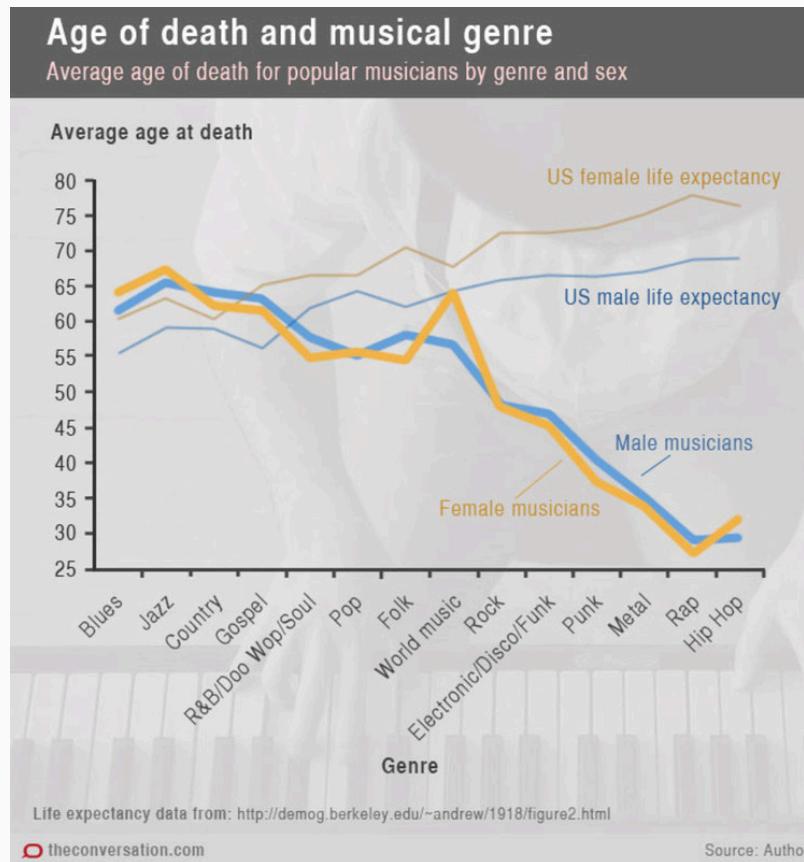




# Eine Liste von Dos und Don'ts

1. Folge dem Prinzip der proportionalen Tinte.
2. Maximiere den Anteil der Daten am Diagramm – aber mit Augenmaß.
3. Vermeide überlagerte, unsichtbare Datenpunkte.
4. Lass alles Unwichtige weg.
5. Überlade Grafiken nicht – lieber mehrere einfache statt einer komplexen.
6. Verwende Farbskalen, die zur Logik der Daten passen.
7. Nutze Farben, die auch bei Farbsehschwächen unterscheidbar sind.
8. Verwende eine Legende nur, wenn sie wirklich nötig ist.
9. Achte auf eine sinnvolle Reihenfolge in der Legende.
10. Beschriffe Achsen klar, aber ohne überflüssige Informationen.
11. Nutze Raster- und Hilfslinien sparsam und gezielt.
12. Sortiere nicht alphabetisch, sondern in einer natürlichen oder inhaltlich sinnvollen Reihenfolge.
13. Bei Balkendiagrammen sollte die Achse immer bei null beginnen.
14. Die erklärende Variable gehört auf die x-Achse, das Ergebnis auf die y-Achse.
15. Achsen haben eine natürliche Richtung: größere Werte rechts bzw. oben.
16. Vermeide unbedingt doppelte y-Achsen.
17. Kreisdiagramme besser meiden – oder sehr überlegt einsetzen.
18. Keine 3D-Diagramme verwenden.
19. Wähle gut lesbare Schriftarten und -größen.
20. Manchmal reicht auch einfach eine Tabelle.

# Daten-Tatort II: Sind bestimmte Musikgenres tödlicher als andere?



Quelle The Conversation, Hintergrund (mit Autoren-Replik) Calling Bullshit

[Study author Dianna Kenny] found that musicians from older genres – including blues, jazz, country and gospel – have similar lifespans to American people their own age. The life expectancy for R&B musicians is slightly lower, while the life expectancy for newer genres like rock, techno, punk, metal, rap and hip hop is significantly shorter."

Ana Swanson, Washington Post

"It's a cautionary tale to some degree," Kenny told the Washington Post. "People who go into rap music or hip hop or punk, they're in a much more occupational hazard profession compared to war. We don't lose half our army in a battle."

Dianna Kenny, quoted by Washington Post

Cause of death by genre					
Various causes of death for musicians of different genres					
	Accidental	Suicide	Homicide	Heart-related	Cancer
% deaths per cause	19.5%	6.8%	6.0%	17.4%	23.4%
Blues	9.2%	2.0%	3.5%	28.0%	24.2%
Jazz	10.6%	2.7%	1.9%	20.7%	30.6%
Country	15.8%	4.7%	1.6%	23.5%	25.1%
Gospel	13.3%	0.9%	3.6%	18.5%	23.0%
R&B	11.5%	1.6%	5.0%	23.2%	26.8%
Pop	19.0%	6.4%	2.9%	16.4%	26.7%
Folk	15.9%	5.5%	4.4%	15.3%	32.3%
World music	12.7%	3.4%	9.6%	17.8%	19.9%
Rock	24.4%	7.2%	3.6%	15.4%	24.7%
Electronic	16.7%	5.0%	10.0%	15.0%	25.0%
Punk	30.0%	11.0%	8.2%	12.6%	18.3%
Metal	36.2%	19.3%	5.9%	11.0%	14.1%
Rap	15.9%	6.2%	51.0%	6.9%	7.6%
Hip Hop	18.3%	7.4%	51.5%	6.1%	6.1%

Red: significantly above the overall average rate for cause of death  
 Blue: above the overall average rate for cause of death  
 Green: significantly below the overall average rate for cause of death

Note: not all causes shown

Source: Author

theconversation.com

## Some issues

- 1. Sanity check:** Sind manche Genres wirklich so tödlich? Sterben z. B. Rap-Musiker wirklich mit einem Durchschnittsalter von ~30 Jahren?
- 2. Datenzensierung:** Einige Genres sind jünger als andere, und die Auswertung setzt voraus, dass die Musiker bereits gestorben sind. Die meisten Rap- und Hip-Hop-Stars sind heute noch am Leben - Lebensdauer unbekannt!
- 3. Bedingte Wahrscheinlichkeiten:** Die Wahrscheinlichkeiten der einzelnen Todesursachen hängen davon ab, dass der Tod zum Zeitpunkt der Untersuchung bereits eingetreten ist. 4) (Kleineres Problem) Liniendiagramm für kategoriale Daten? Keine gute Idee.

# Lösung des Falles

## Evidenz und Schlüsse

Der Tatort	Musikgenre korreliert mit Todesalter
Die Verdächtigen	Lifestyle (z.B. Drogenkonsum), Alter des Genres, unvollständige Datenerhebung
Die Tatwaffe	Konfundierung durch Daten-Zensur

"In other words, it's not that rap stars will likely die young; it's that the rap stars who have died certainly died young because rap hasn't been around long enough for it to be otherwise."

Carl Bergstrom and Jevin West, [Calling Bullshit](#)<sup>1</sup>

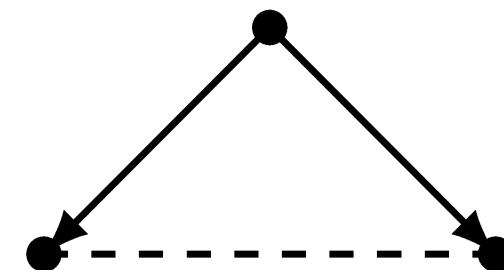
Inspector Lestrade:



Age of musicians in that genre

Sherlock:

Age of musicians in that genre

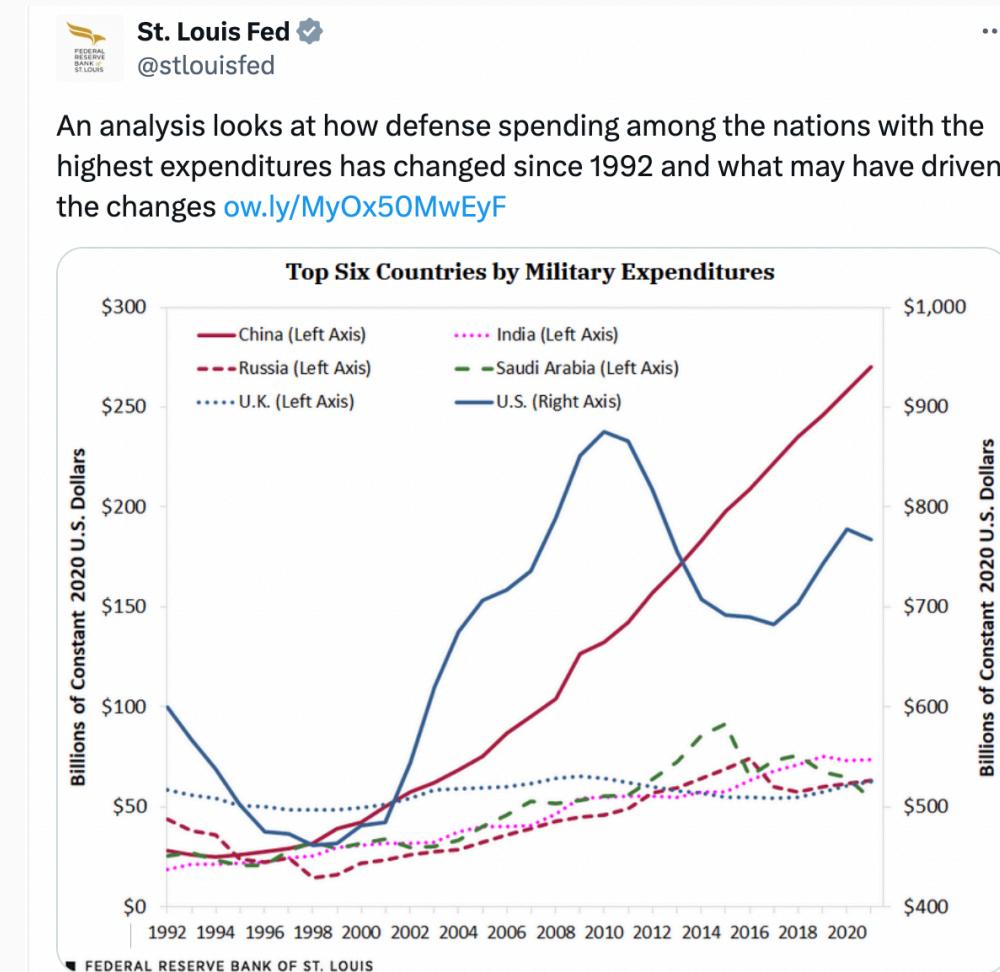


<sup>1</sup>Die Autorin Dianna Kenny hat mit einer plausiblen Antwort reagiert und auf eine [besser designete Studie verwiesen](#), deren Evidenz konsistent mit den ursprünglichen Mustern zu sein scheint.

# Diskussion in Gruppen

## 5 Minuten in 2er-Teams

- Was ist Ihrer Meinung nach das Ziel der Visualisierung?
  - Was wollte der Designer vermitteln?
  - Gelingt es ihm/ihr, die Botschaft zu vermitteln?
- Finden Sie irgendwelche grundlegenden Probleme mit der Visualisierung?
  - Ist dies die beste Art der Darstellung?
  - Gibt es irgendwelche Merkmale, die den Betrachter in die Irre führen könnten?
- Können Sie sich vorstellen, wie man die Visualisierung verbessern könnte?



# Interaktive Kommunikation mit Dashboards

---

# Was sind Dashboards?

## Überblick

- Ein (Unternehmens- oder Daten-) **Dashboard** ist eine grafische Benutzeroberfläche, die einen Überblick über Leistungsindikatoren oder andere Größen von Interesse bietet.
- Es ist ein **Monitoring-** (und weniger ein Analyse-) **Werkzeug**.
- **Verknüpfung von Datenvisualisierung und Bericht.**
- Beliebt, um **Informationen** aus operativen Einheiten (für strategische und analytische Zwecke) zu synthetisieren.
- Der Datenjournalismus hat begonnen, Dashboards im Zusammenhang mit **Wahlen**, der **COVID-19-Pandemie** und **Sport** zu nutzen.
- **Gemeinsame Merkmale** sind:
  - Zugänglichkeit über Webbrowser
  - Ausstattung mit interaktiven Elementen
  - Starker Fokus auf vergleichende Visualisierung
  - Bereitstellung von Trends zu wichtigen Leistungsindikatoren (KPIs)



Credit Tim Green



Credit HelicalInsight OpenSourceBI

# Warum gibt es Dashboards?

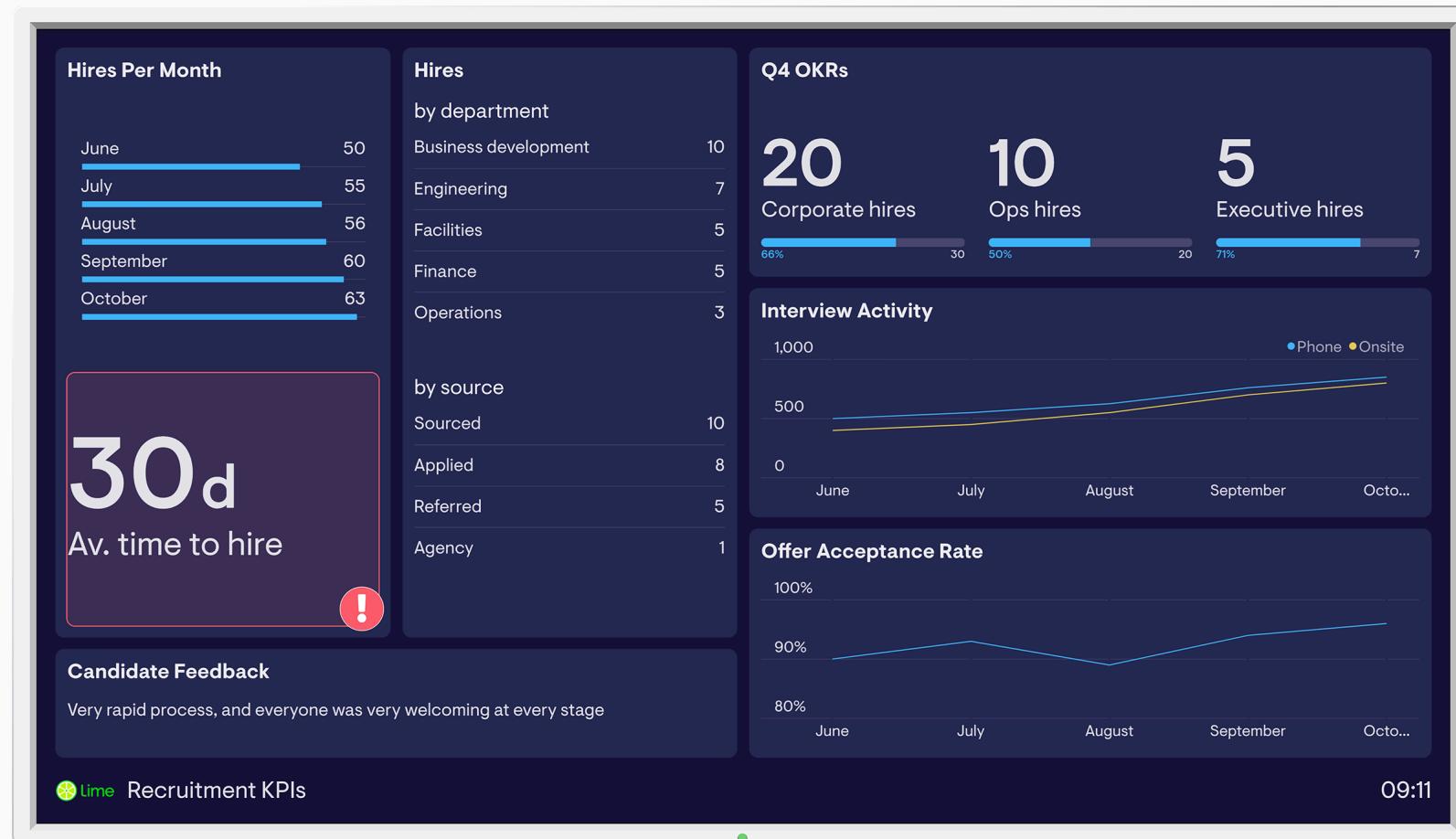
## Warum gibt es sie?

- Es gibt eine zunehmende **Fülle von Daten** (oft prozessbedingt), die nicht für sich selbst sprechen können.
- Bei kluger Nutzung können diese Daten einen **wichtigen Teil der Business Intelligence** und eine Grundlage für eine **evidenzbasierte Entscheidungsfindung** auf hohem Niveau bieten .
- Kontinuierliche Quantifizierung von Indikatoren von Interesse (→ **Monitoring**).
- **Verringerung des Informationsgefälles** zwischen Analysten und Beteiligten.
- Außerdem kann die **Messung der Gesundheit von Organisationen** dazu beitragen, die Kontrolle zu behalten (wenn auch nur als performativer Akt) und das Bedürfnis der Manager nach Mikromanagement zu befriedigen.



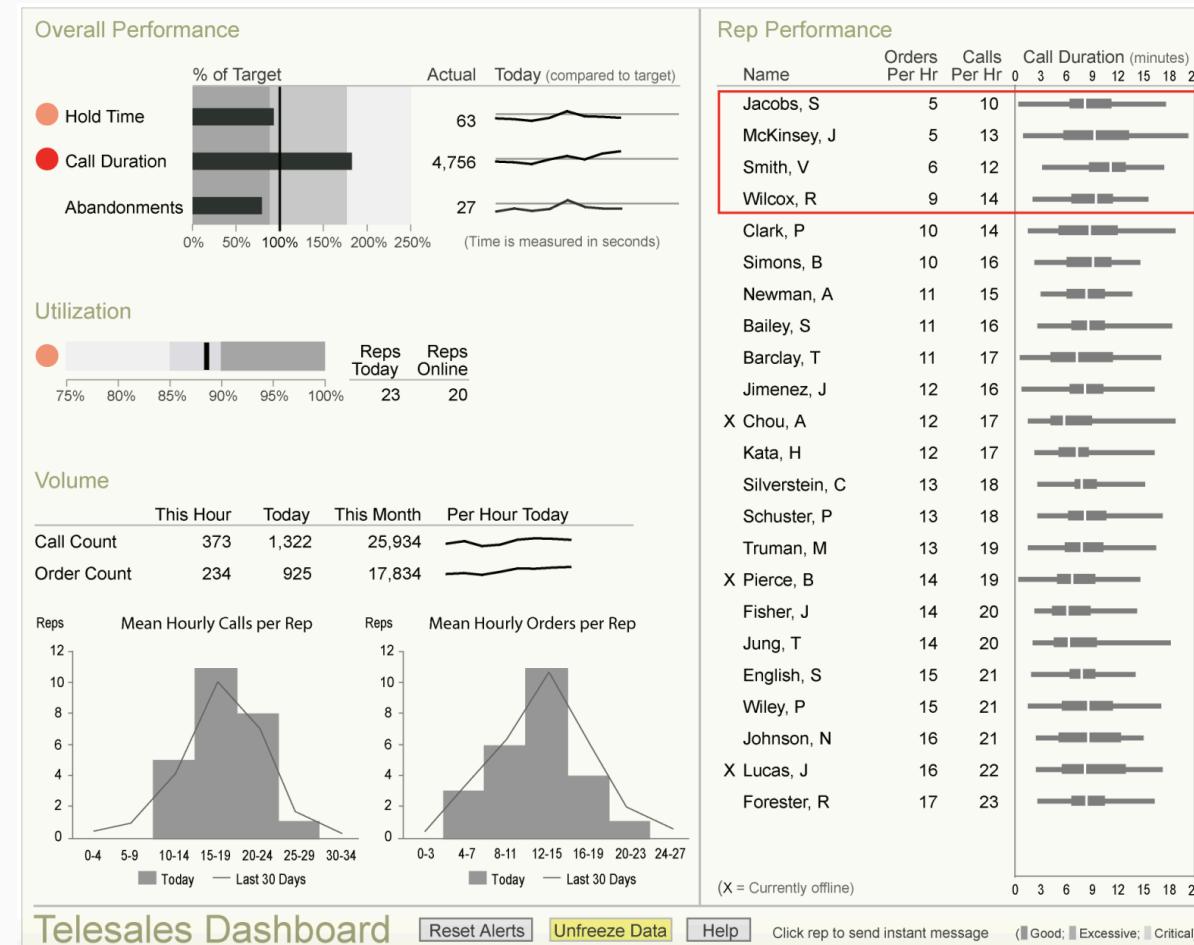
Credit [towardsdatascience.com](https://towardsdatascience.com/)

# Dashboards in freier Wildbahn



Credit geckoboard.com

# Dashboards in freier Wildbahn



Credit Stephen Few

# Dashboards in freier Wildbahn

# London

51.51 N, 0.13 W

Mon 9 Dec @ 16:02:42

[Go to Map](#) - [Go to Grid](#) - [Change City](#)

**WEATHER STATIONS (MULTIPLE SOURCES)**

STATION	Wind Speed	Wind Gusts	Direction	TEMPERATURE	HUMIDITY	Rain Today	Pressure	Forecast
CASA Office: Bloomsbury W1	8 mph	9 mph	SE ↘	11.5 °C	76%	0.0 mm	1027.9 mbar	Clear Night
Lambeth Meters: Brixton SW9	4.3 mph	4.3 mph	SW ↗	11.0 °C	83%	0.0 mm	1026.4 mbar	Clear Night
Hampstead NW3	3.6 mph	3.6 mph	S ↑	9.8 °C	84%	0.0 mm	1029.0 mbar	Clear Night

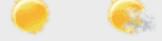
**WEATHER (METAR)** [848]

**London City Airport** [848]

Mostly clear	SW at 3 mph	11 °C
--------------	-------------	-------

**FORECAST (YAHOO! WTH)** [1748]

Mon	10 °C
Tue	9 °C



**TUBE LINE STATUS (TFL)** [39]

Bakerloo	Good Service
Central	Good Service
Circle	Good Service
District	Good Service
H & C	Good Service
Jubilee	Good Service
Metropolitan	Good Service
Northern	Good Service
Piccadilly	Good Service
Victoria	Good Service
W & C	Good Service
Overground	Good Service
DLR	Good Service

**BIKE SHARING (TFL)** [38]

4.3 %	4.9 %
Stations Full	Stations Empty
7354	430
Bikes Available	Bikes or Docks Faulty

Available Bikes (last 24h)



**IN SERVICE (TFL)** [8]

7197	London buses
378	Underground trains

**AIR POLLUTION (DEFRA)** [1748]

µg/m³	TIME AVERAGED	OZONE	NO <sub>2</sub>	SO <sub>2</sub>	PM <sub>2.5</sub>	PM <sub>10</sub>
Bloomsbury	13	38	4	9	10	
Marylebone Rd	9	16	26	22	34	
N Kensington	14	40	?	12	18	

**RADS (CASA)** [1]

CASA Office Desk ‡

6 cpm (uncalibrated)

**RIVER LEVEL (PLA)** [248]

Thames (Tower Pier)

4.13 metres

**STOCKS (YAHOO)** [7]

FTSE 100 Index

6552.34 +0.35 (0.01%)

**RANDOM TRAFFIC CAMERAS (TFL)** [10]

Old Kent Rd/Asylum Rd	High St/Grosvenor Rd W Wickham

**BBC LONDON NEWS (BBC)** [48]

Ricky killer 'a soldier of Allah' Mayor bike 'scaring' claim withdrawn. Murder police found grave in garden. Cameron praises towering Mandela. Police crackdown on pirate site ads. Why do we value gold?

**OPENSTREETMAP UPDATES (OSM)** [248]

Third attempt to name the terraced cottages around the Green. Revert my change to terraced cottages as they get rendered with wrong address. Added Tibet Foundation refining Name error.



**ELECTRICITY (N.GRID)** [29]

Demand (Great Britain)

48211 MW

**Mood (LSE MAPPINESS)** [38]

8% unhappier than the long term average for here	13% happier than the whole country right now
--	--

**TWITTER TRENDS FOR LONDON** [153]

MPs #NFL Christmas #Confident Xmas #ashes London #RIPAlexTurner #12DaysofJonesDAY9 Waca

Credit [idashboards.com](http://idashboards.com)

## Gestalterische Herausforderungen

- **Sie sagen zu wenig.** Der Verlust von Informationen durch Reduktion auf wenige KPIs ist **fatal** für eine gute Entscheidungsfindung.
- **Sie sagen zu viel** (irrelevante Dinge).
- Dashboards scheitern oft nicht an der Technologie, sondern an der Kommunikation (zurückzuführen auf schlechtes Design):
- "Dashboards sind nicht zur Show gedacht. Keine noch so niedliche technische Raffinesse kann eine klare Kommunikation ersetzen." **Stephen Few**, *Perceptual Edge*
- **Es alle Regeln für gute/schlechte Visualisierung.** **Hier** eine schöne Fallstudie zur Verbesserung der Gestaltung eines Dashboards.

## Analytische Herausforderungen

- Dashboards entsprechen dem Wunsch, auf der Grundlage einiger weniger ausgewählter Metriken gute Entscheidungen treffen zu können.
- **Realität wird stark vereinfacht.**
- Alle Herausforderungen, die sich analytischer Arbeit stellen - Selektion, Messung, Kausalität, Vorhersagbarkeit - sind nach wie vor gültig, werden aber bei der Aggregation von Daten verschleiert.
- Einfache Metriken können immer noch nützlich sein, aber oft **braucht man kontextuelles Wissen** (das in Dashboards schwer zu vermitteln ist).
- Eine weitere Folge des "Dashboarding" von Business Intelligence kann sein, dass Entscheidungen auf Basis von Metriken schnell die Metriken sabotiert (**Goodhart's Law**).

## Checkliste<sup>1</sup>

1. Geht es um ein Monitoring, bei dem Ihre Daten/Metriken häufig aktualisiert werden müssen?
2. Wer wird das Dashboard nutzen und zu welchem Zweck? Welche Fragen sollen damit beantwortet werden? Welche Maßnahmen werden sie als Reaktion auf diese Antworten ergreifen?
3. Welche spezifischen Informationen sollen angezeigt werden, und sind sie auch ohne viel Kontext aussagekräftig?
4. Was könnte dazu führen, dass die Metriken falsch/irreführend sind?

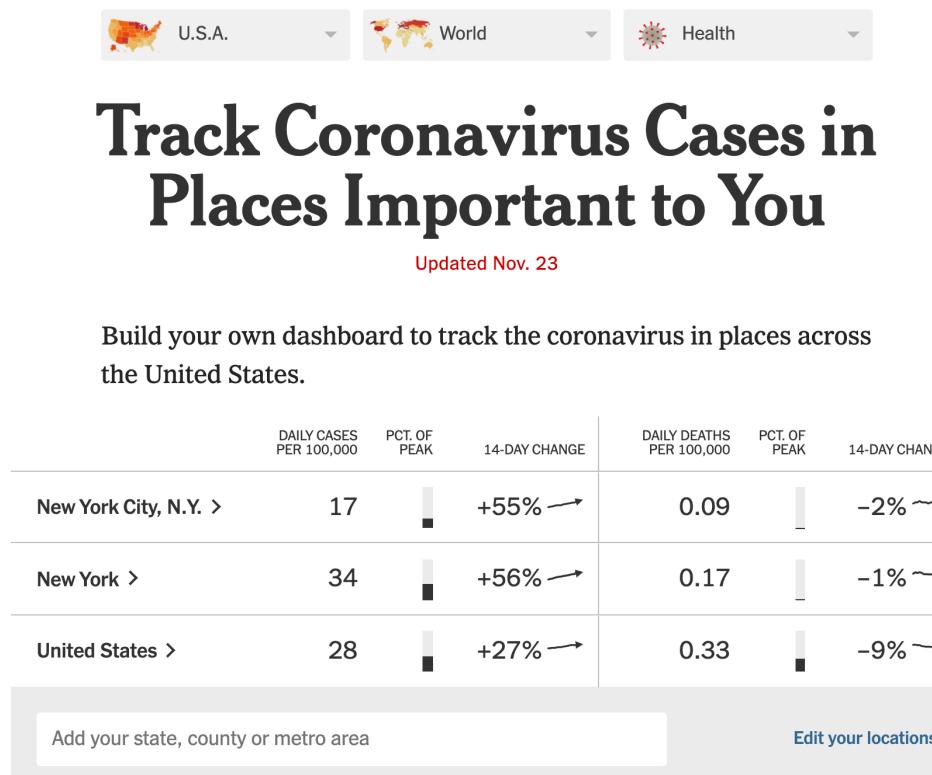
## Design-Hinweise

- Minimieren Sie Ablenkungen.
- Konzentrieren Sie sich auf sinnvolle KPIs, nicht auf solche, die interessant aussehen.
- Überladen Sie das Dashboard nicht mit Informationen.
- Wenden Sie alle Regeln guter Visualisierung an.
- Verwenden Sie interaktive Elemente mit Bedacht (z. B. um optionale Inhalte bedingt sichtbar zu machen).
- Idealerweise ist alles Wesentliche auf einem Bildschirm sichtbar (auch auf Smartphones?).

<sup>1</sup>Quelle: Stephen Few/Perceptual Edge

# Dashboards in freier Wildbahn: COVID-19-Edition

Hertie School

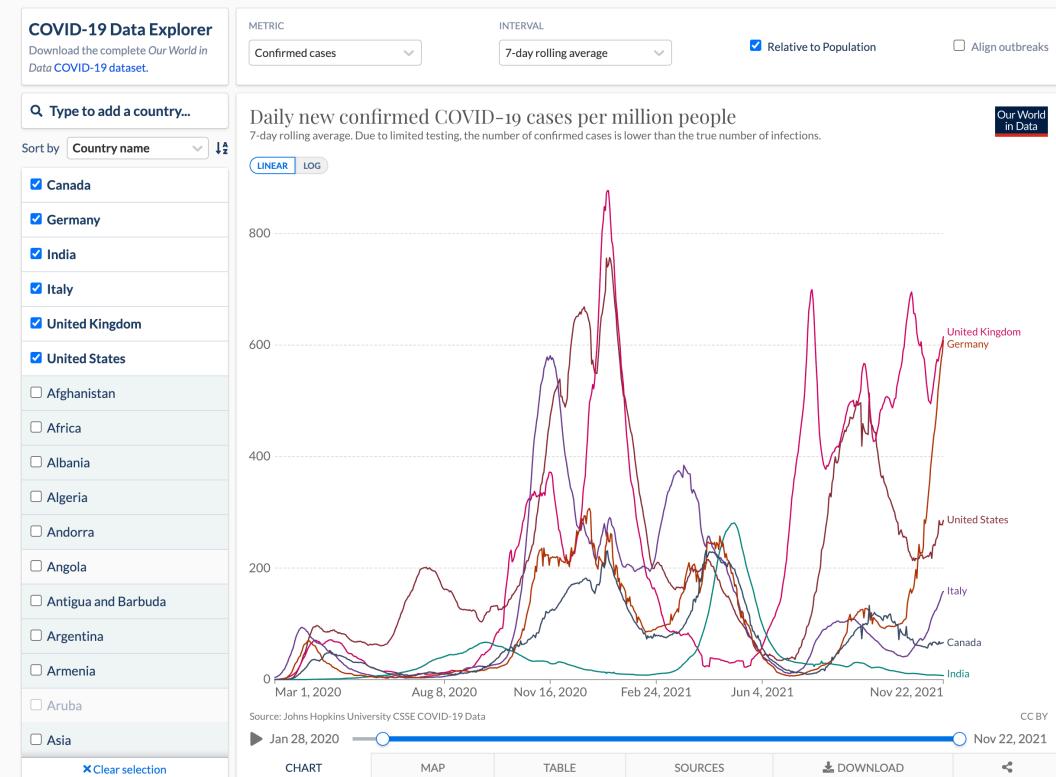


14-day change trends are calculated with 7-day averages. Pct of peak indicates how an area compares to when cases or deaths were at their highest.

Share this view of places with a friend: <https://www.nytimes.com/interactive/2021/covid-19/us.html>

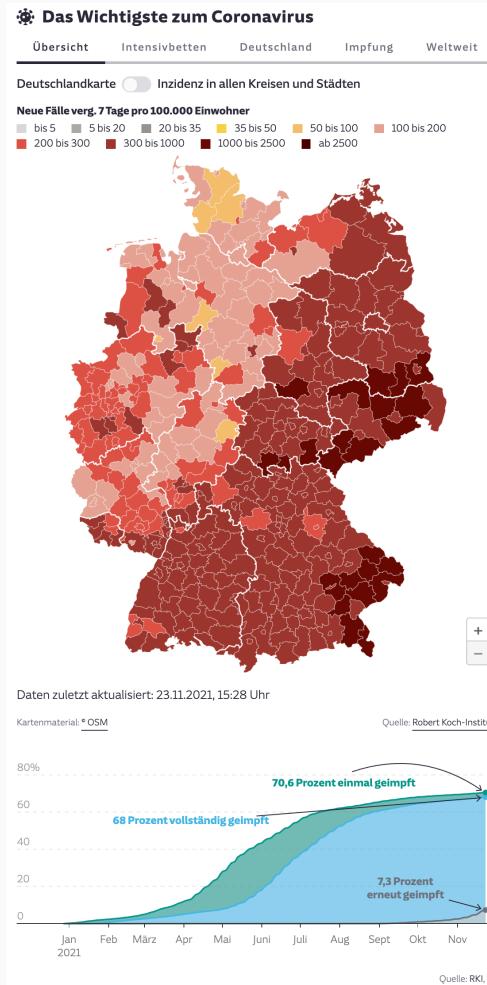
Save your dashboard across devices and get updates by email →

Quelle NY Times

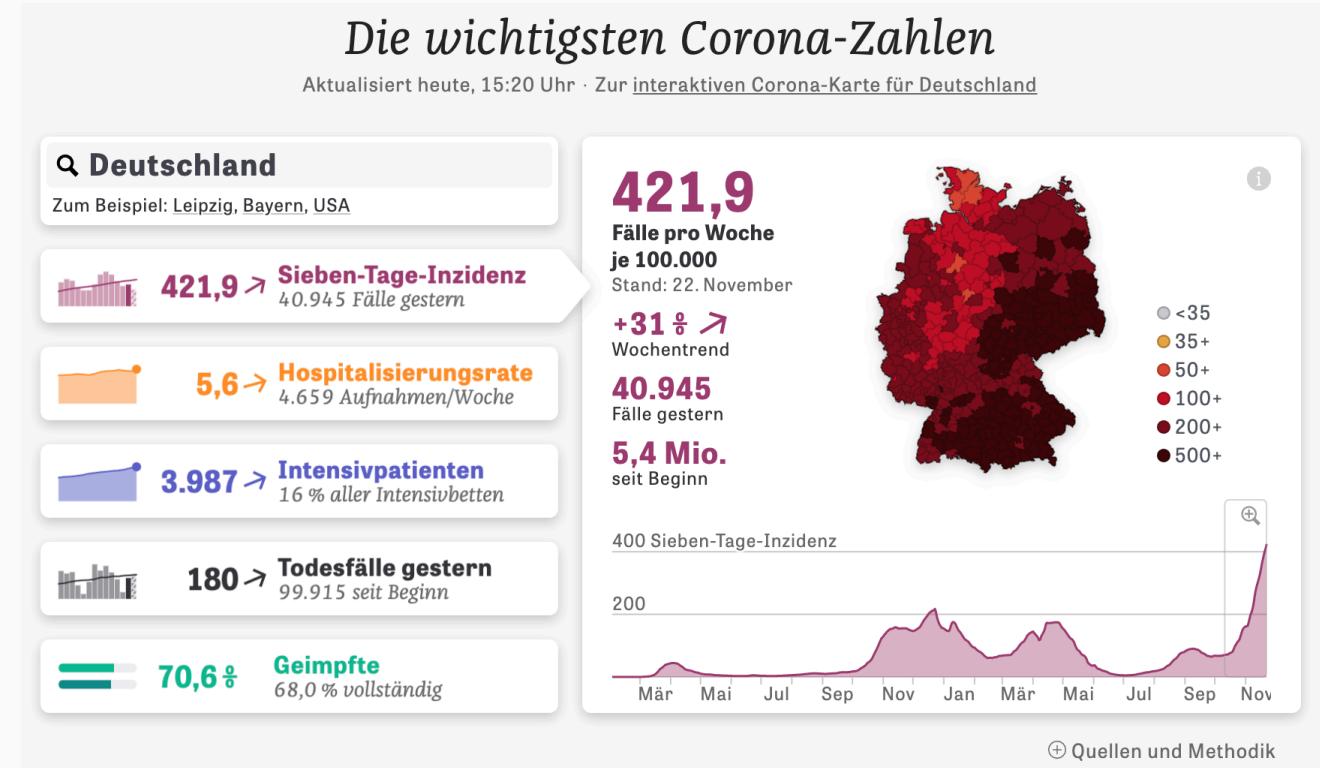


Quelle Our World in Data

# Dashboards in freier Wildbahn: COVID-19-Edition



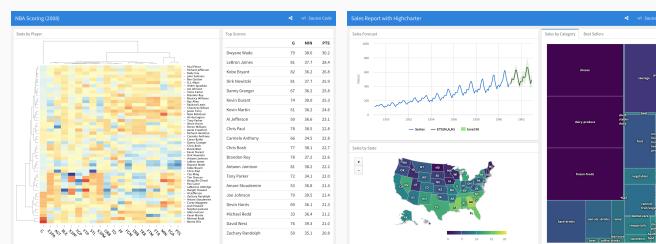
Quelle SZ Online



Quelle ZEIT Online

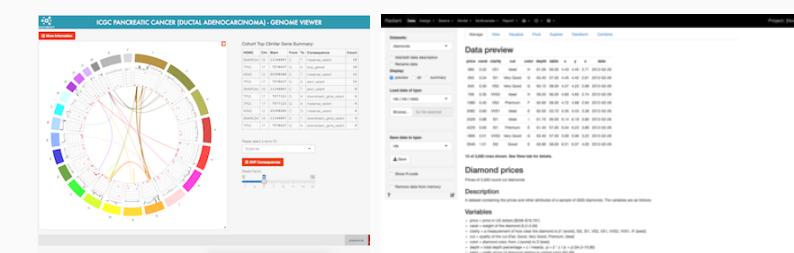
## flexdashboard package

- Übersicht [hier](#).
- Gut für die einfache Erstellung von Dashboards
- Einfach ein Dokument, das wie ein Dashboard aussieht
- Kann interaktiven Code nur clientseitig ausführen (in eingebettetem JavaScript)



## shiny package

- Übersicht [hier](#).
- Komplexer zu programmieren, aber die bessere Option für komplexe Anwendungen.
- Kann jedes Layout implementieren.
- Benötigt einen Server dahinter, um R-Code auf Benutzereingaben auszuführen.
- Das Paket [shinydashboard](#) bietet eine weitere Möglichkeit, Dashboards mit Shiny zu erstellen.



## Online-Ressourcen

- Shiny official website
- Shiny official tutorial
- Shiny cheatsheet
- Mastering Shiny, book by Hadley Wickham
- Many useful articles about different topics
- Publishing own Shiny apps for free with shinapps.io
- Hosting your Shiny app [on your own server](#)
- [Debugging Shiny](#)

## Ein Überblick über Shiny-Erweiterungen

- [awesome-shiny-extensions](#)

## Einige Highlights

- [shinythemes](#): Ändern der Themes von Shiny-Anwendungen
- [shinyjs](#): Apps mit JavaScript-Operationen anreichern
- [leaflet](#): Interaktive Karten
- [ggvis](#): Ähnlich wie ggplot2, aber mit Fokus auf Web und Interaktion
- [shinydashboard](#): Werkzeuge zur Erstellung visueller Dashboards