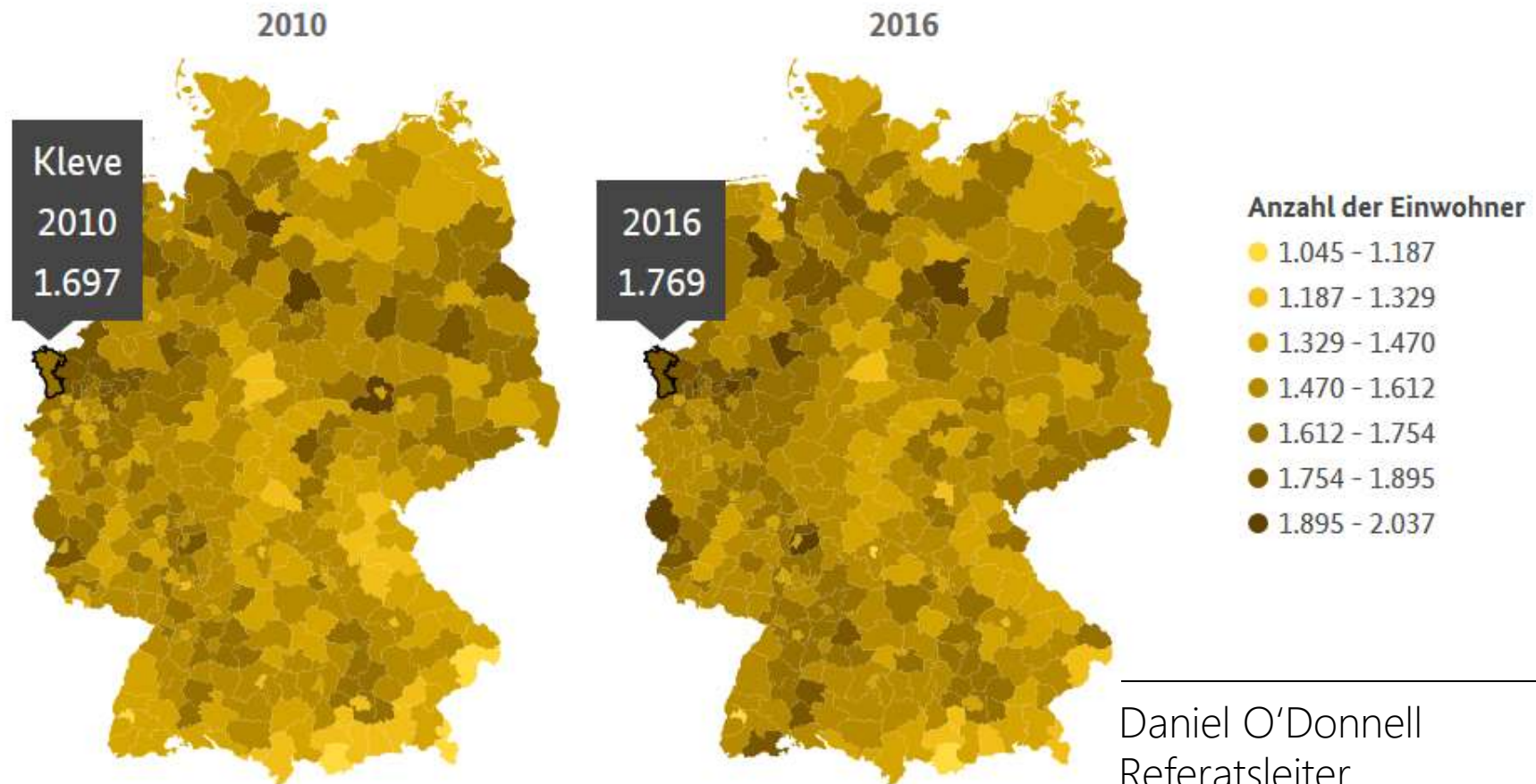


Interaktive Visualisierung statistischer Daten

Visualize | 4 Juni 2019 | Hochschule Rhein-Waal



Quelle: Destatis

Daniel O'Donnell
Referatsleiter
Internationale Statistik

Warum Datenvisualisierung?

Muster schneller erkennen, Kernaussagen leichter erfassen

Anscombe's quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Für alle Zahlenreihen gilt

Durchschnitt y 9,0

Varianz y 11,0

Durchschnitt x 7,5

Varianz x 4,1

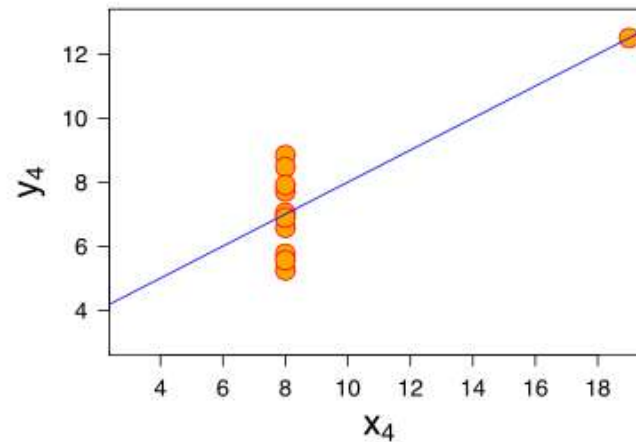
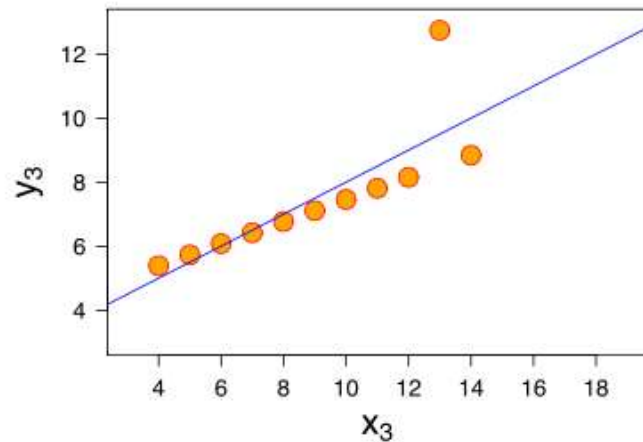
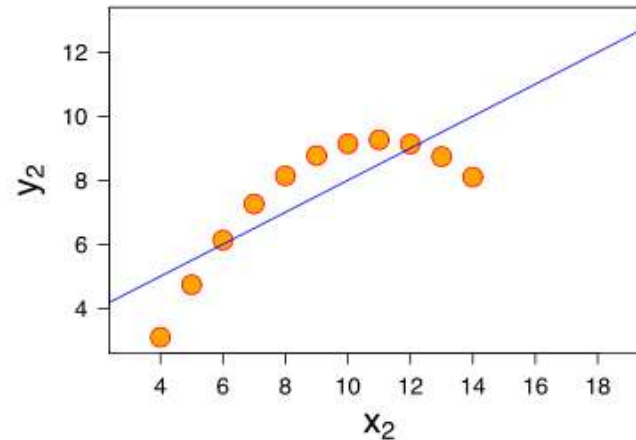
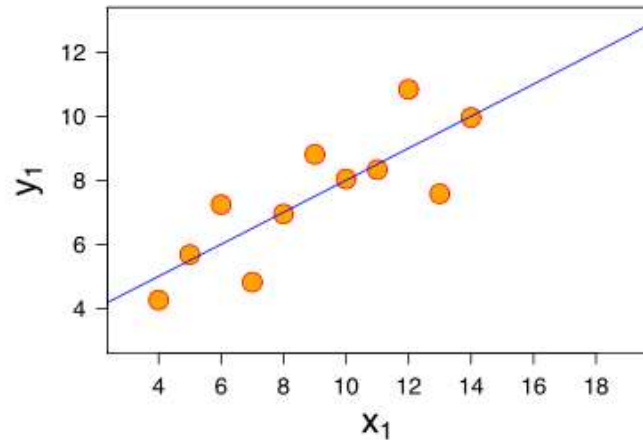
Korrelation 0,82

Lineare Regression $y = 3 + 0,5x$

Quelle: F. J. Anscombe: Graphs in Statistical Analysis.
In: American Statistician. 27, Nr. 1, 1973, S. 17–21

Warum Datenvisualisierung?

Muster schneller erkennen, Kernaussagen leichter erfassen



Quelle: F. J. Anscombe: Graphs in Statistical Analysis.
In: American Statistician. 27, Nr. 1, 1973, S. 17–21

Warum Datenvisualisierung?

Die Geschichten hinter den Zahlkolonnen sichtbar machen

 **eurostat** Wichtiger rechtlicher Hinweis
v3.4.11b-20190314-e2667-ESTAT_LINUX_PR
DATA-EXPLORER_PRODmanaged22

[lfsa_urgan]

Arbeitslosenquoten nach Geschlecht, Alter und Staatsangehörigkeit (%)
Letzte Aktualisierung: 23-05-2019

Tabellenanpassung [zeigen](#)

TIME +

+ Altersklasse

15 bis 24 Jahre +

GEO +

+ Land der Staatsangehörigkeit

Insgesamt +

+ Geschlecht

Insgesamt +

+ Maßeinheit

Prozent +

GEO	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Europäische Union - 28 Länder	20,1	21,2	21,7	23,2	23,7	22,2	20,3	18,7	16,8	15,2
Belgien	21,9	22,4	18,7 ^(b)	19,8	23,7	23,2	22,1	20,1	19,3 ^(b)	15,8
Bulgarien	16,2	21,9 ^(b)	25,0 ^(b)	28,1	28,4	23,8	21,6	17,2	12,9	12,7
Tschechien	16,6	18,3	18,1 ^(b)	19,5	19,0	15,9	12,6	10,5	7,9	6,7
Dänemark	11,8	14,0	14,2	14,1	13,1	12,6	10,8	12,0 ^(b)	11,0 ^(b)	9,4
Deutschland (bis 1990 frühere Länder)	11,2	9,8 ^(b)	8,5 ^(b)	8,0	7,8	7,7	7,2	7,1	6,8	6,2
Estland	27,4	32,9	22,4	20,9	18,7	15,0	13,1	13,4	12,1	11,8
Irland	24,5	28,1	29,6	30,8	26,7	23,4	20,2	16,8	14,4 ^(b)	13,8
Griechenland	25,7 ^(b)	33,0	44,7	55,3	58,3	52,4	49,8	47,3	43,6	39,9
Spanien	37,7	41,5	46,2	52,9	55,5	53,2	48,3	44,4	38,6	34,3
Frankreich	22,9	22,5	21,9	23,7	24,1	24,2 ^(b)	24,7	24,6	22,3	20,8
Kroatien	25,2	32,4	36,7	42,1	50,0	45,5	42,3	31,3	27,4	23,7
Italien	25,3	27,9	29,2	35,3	40,0	42,7	40,3	37,8	34,7	32,2
Zypern	13,8 ^(b)	16,6	22,4	27,7	38,9	36,0	32,8	29,1	24,7	20,2
Lettland	33,3	36,2	31,0	28,5	23,2	19,6	16,3	17,3	17,0	12,2
Litauen	29,6	35,7	32,6	26,7	21,9	19,3	16,3	14,5	13,3	11,1
Luxemburg	17,2 ^(b)	14,2	16,8	18,8	15,5	22,6	17,3 ^(b)	18,9	15,4	14,2
Ungarn	26,4	26,4	26,0	28,2	26,6	20,4	17,3	12,9	10,7	10,2
Malta	14,5	13,2	13,3	13,8	12,7	11,7	11,6	10,7	10,6	9,2
Niederlande	10,2	11,1	10,0	11,7	13,2	12,7	11,3	10,8	8,9	7,2
Österreich	10,7	9,5	8,9	9,4	9,7	10,3	10,6	11,2	9,8	9,4
Polen	20,6	23,7 ^(b)	25,8	26,5	27,3	23,9	20,8	17,7	14,8	11,7
Portugal	20,3	22,8	30,3 ^(b)	37,9	38,1	34,8	32,0	28,0	23,9	20,3
Rumänien	20,8	22,1 ^(b)	23,9	22,6	23,7	24,0	21,7	20,6	18,3	16,2
Slowenien	13,6	14,7	15,7	20,6	21,6	20,2	16,3	15,2	11,2	8,8
Slowakei	27,3	33,6	33,4 ^(b)	34,0	33,7	29,7	26,5	22,2	18,9	14,9
Finnland	21,5	21,4	20,1	19,0	19,9	20,5	22,4	20,1	20,1	17,0
Schweden	25,0	24,8	22,8	23,6	23,5	22,9	20,4	18,9	17,9	16,8
Vereinigtes Königreich	19,1	19,9	21,3	21,2	20,7	17,0	14,6	13,0	12,1	11,3

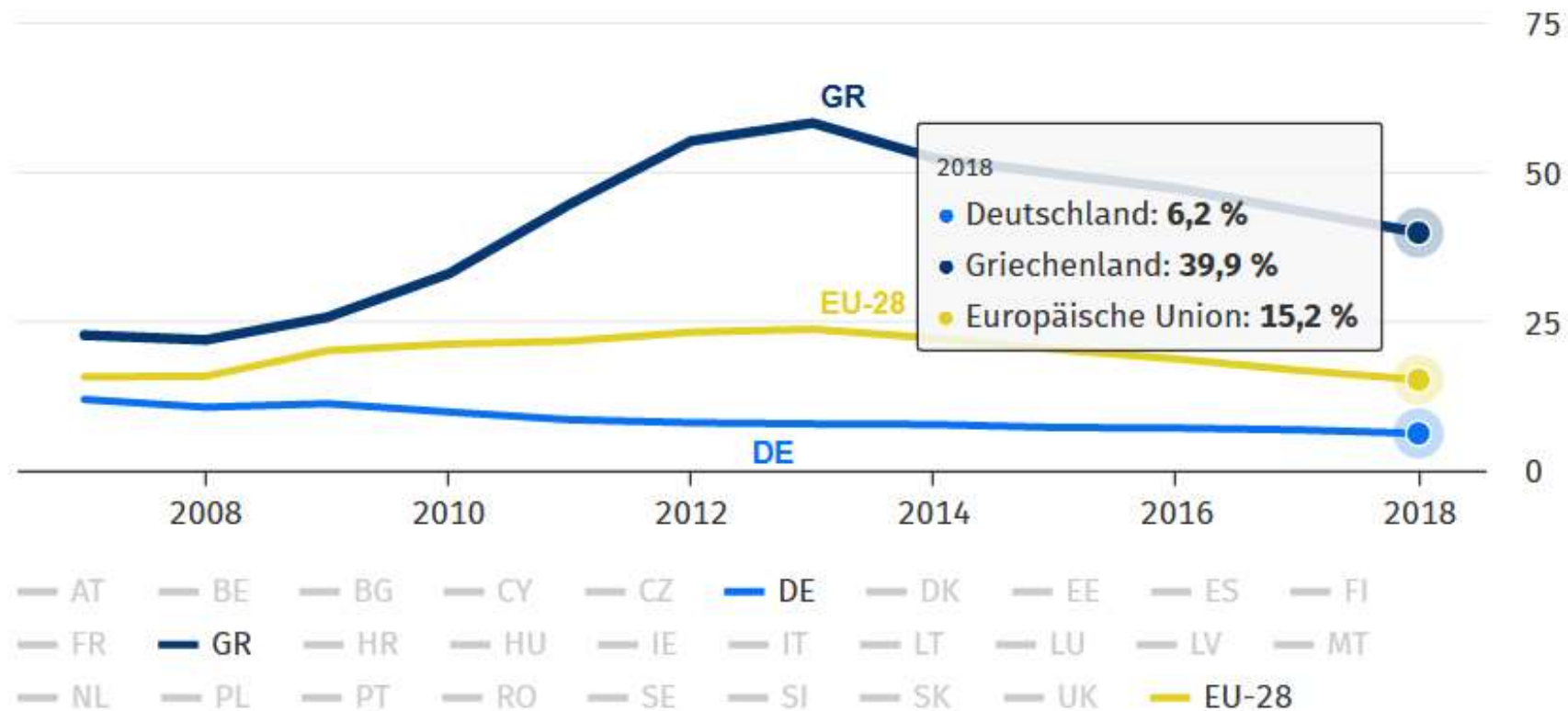
Quelle: Eurostat Datenbank. Tabelle lfsa_urgan

Warum Datenvisualisierung?

Visuell erfassen wir Daten einfacher

Jugenderwerbslosenquote

in % der 15- bis 24-jährigen Erwerbspersonen



Quelle: Eurostat

© Statistisches Bundesamt (Destatis), 2019

Warum interaktiv?

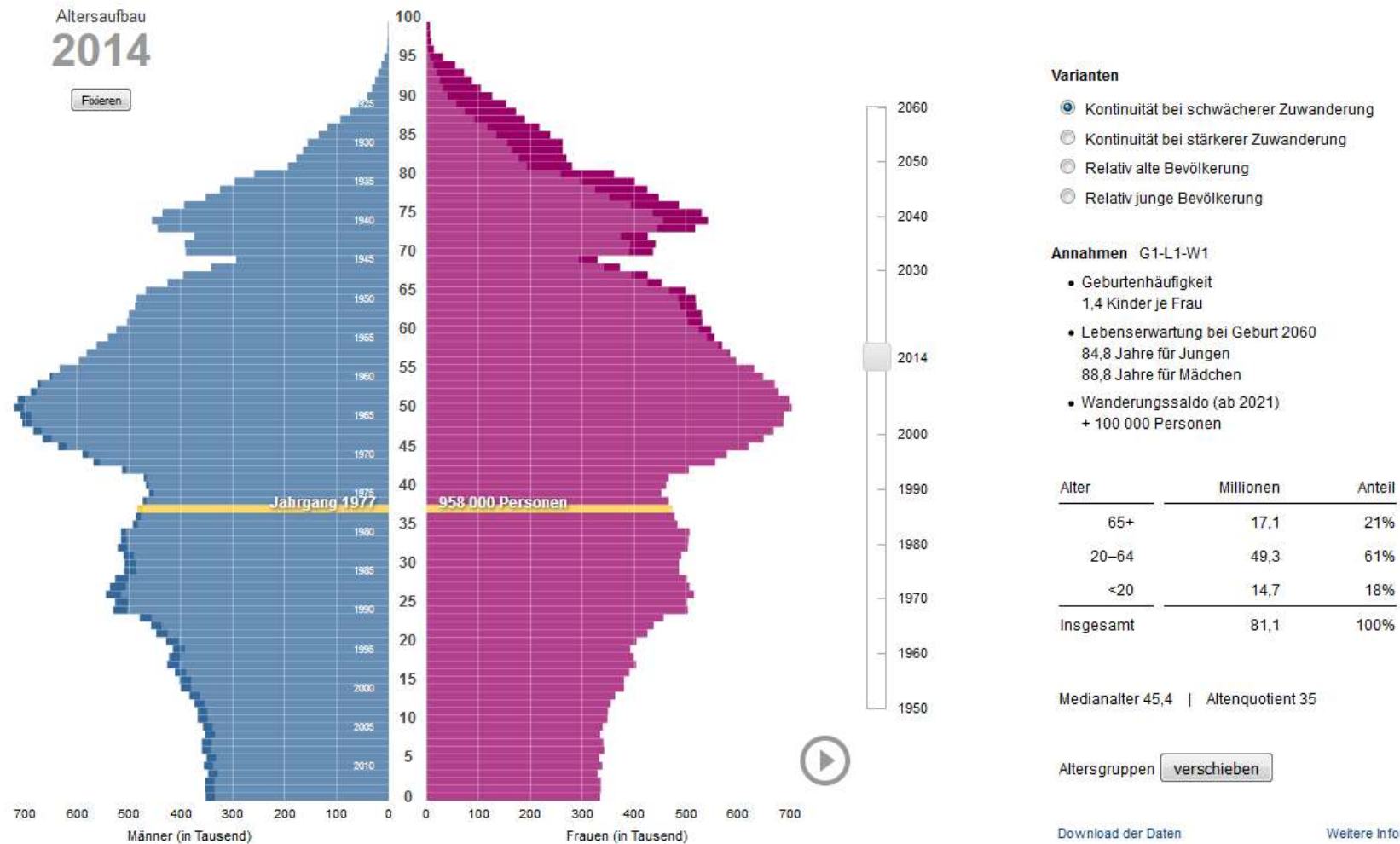
- Interaktion ermöglichen: Nutzer sollen selbst in den Daten stöbern und entscheiden was für Sie wichtig oder interessant ist
Wie sieht es in meiner Region aus?
Wie schneidet meine Altersgruppe ab?
- Und – es gibt Fälle wo man die Tragweite der Daten wirklich nur mit interaktiver Datenvisualisierung erkennt. Schlichte Tabelle, schlichtes Balkendiagramm nicht ausreichend
Perfektes Beispiel: Bevölkerungspyramide

Warum Datenvisualisierung?

Klassiker Bevölkerungspyramide

13. koordinierte Bevölkerungsvorausberechnung für Deutschland

STATIS
Statistisches Bundesamt



Quelle: Destatis

Interaktivität ermöglicht Fokussierung

Interaktiv wird Datenauswahl persönlicher

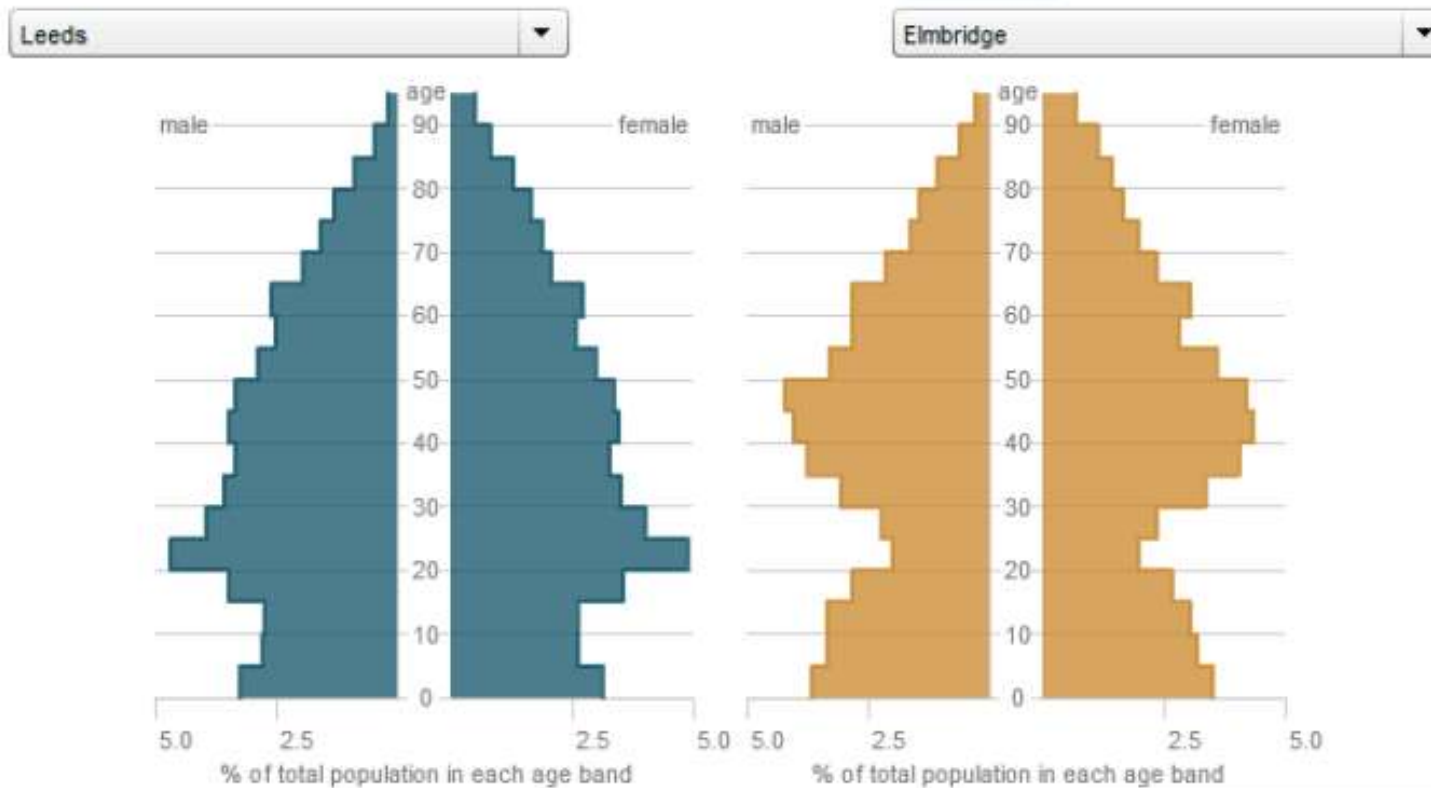


Georgina Orso @georginaorso

The Census 2011 tells me exactly why moving from Leeds to Elmbridge was a bad thing for making new friends:

pic.twitter.com/SR2MoY1V

16 Jul



Quelle: Twitter. Zitiert von Alan Smith (Office for National Statistics UK).

Interaktive Datenvisualisierung

Chancen und Gefahren des Trends

Datenvisualisierung hat viele Vorteile

1. Muster leichter erkennbar
2. Kernaussage schneller erfassbar
3. Mehr Informationen vermittelbar
4. Kann Komplexität reduzieren
5. Darstellung interessanter und zeitgemäßer

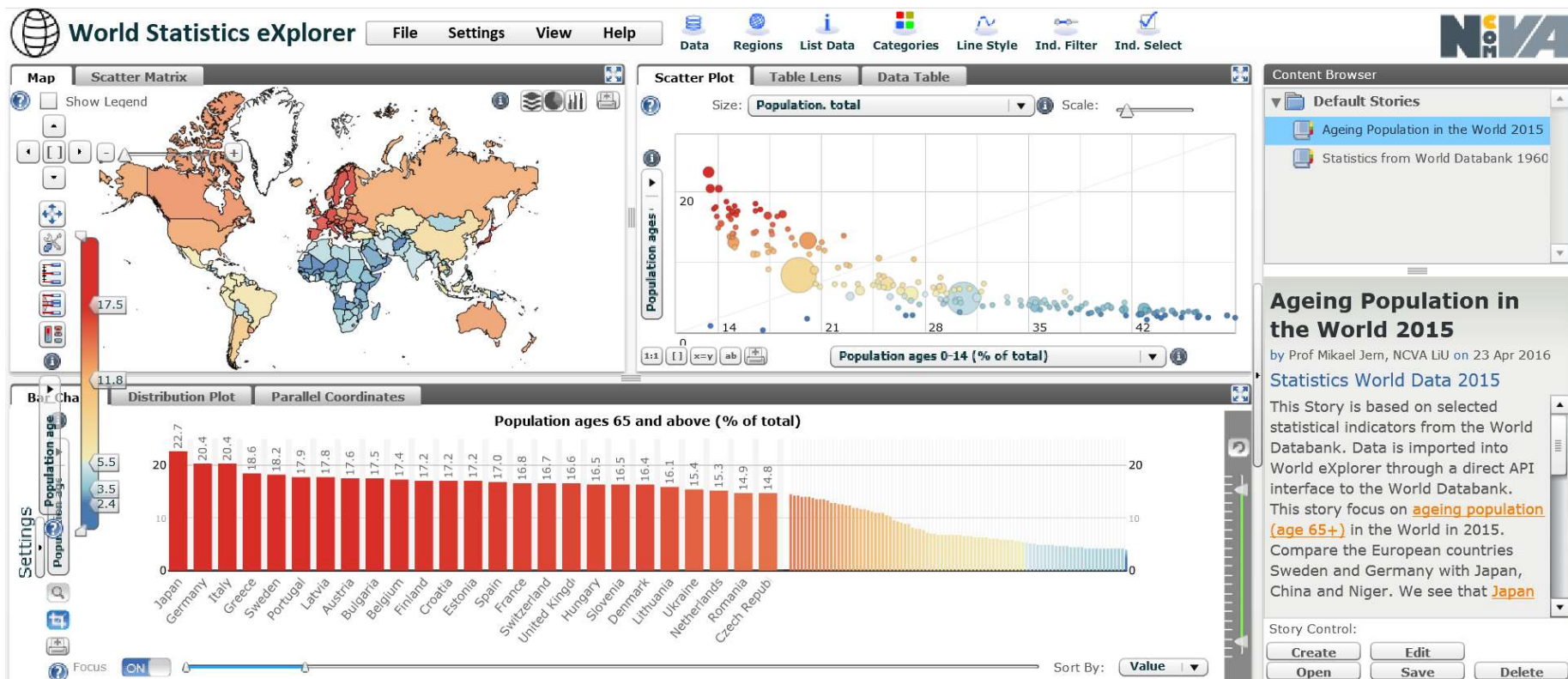
Aber es gibt auch Gefahren

1. Überforderung von Nutzer/-innen durch ein Überangebot an Optionen
2. Zu fokussiert auf bestimmte Entwicklung, Ausblenden des Kontext (Gefahr insb. bei Korrelationen)
3. Nichteinhaltung bestimmter Standards → Verzerrung der Aussage
4. Visualisierung weil modern, nicht weil sinnvoll
5. Darstellungsform nicht datengerecht

Datenvisualisierung: Grundregeln

Worauf gilt es zu achten?

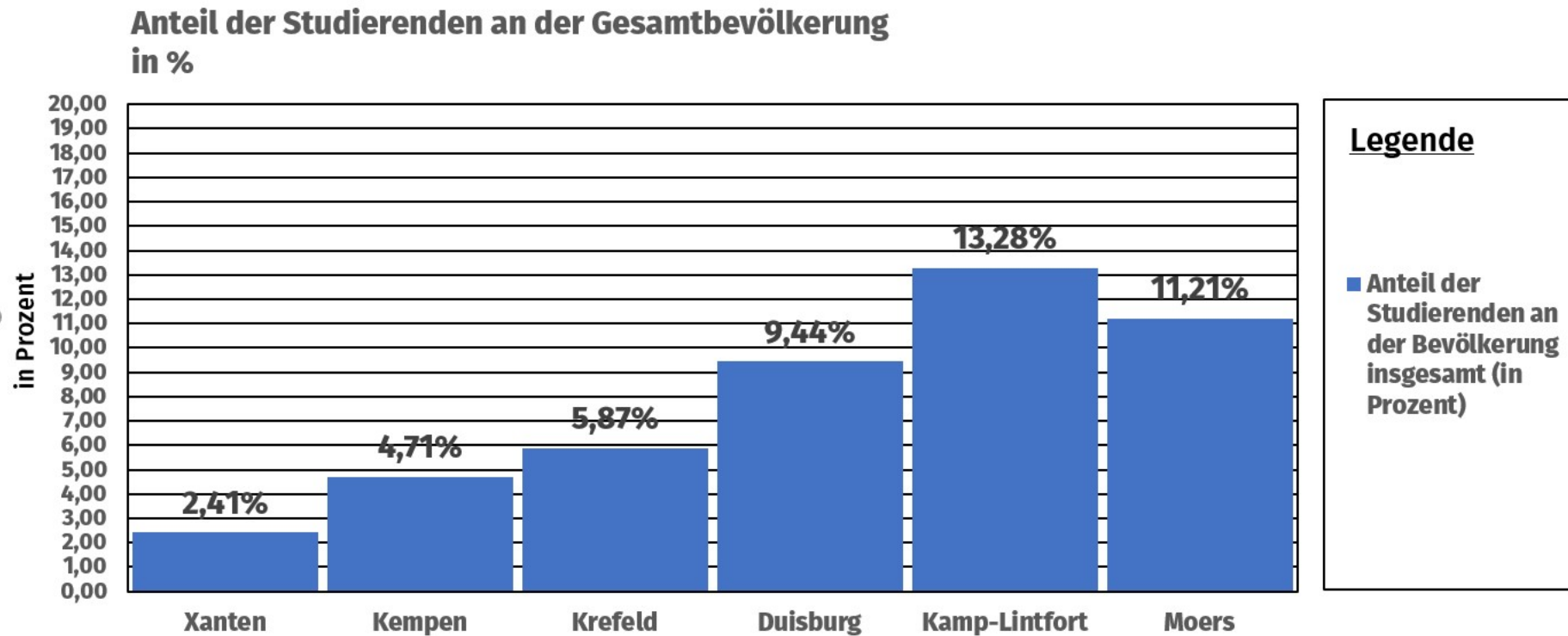
Regel 1: Weniger ist oft mehr – besseres Datenverständnis statt Wow-Effekt



Quelle: NComVA

Datenvisualisierung: Grundregeln

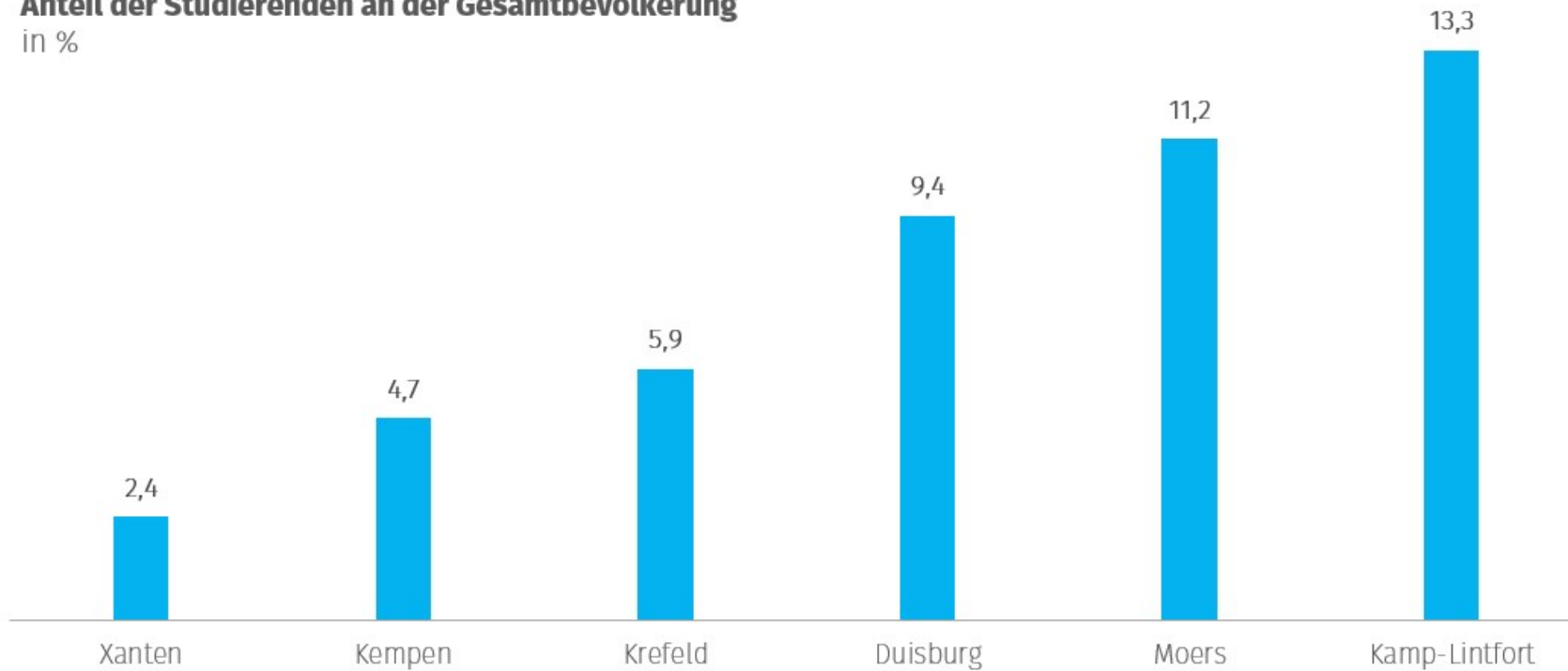
Regel 2: Gutes Verhältnis zwischen Daten und Druckerschwärze → Ed Tufte



Datenvisualisierung: Grundregeln

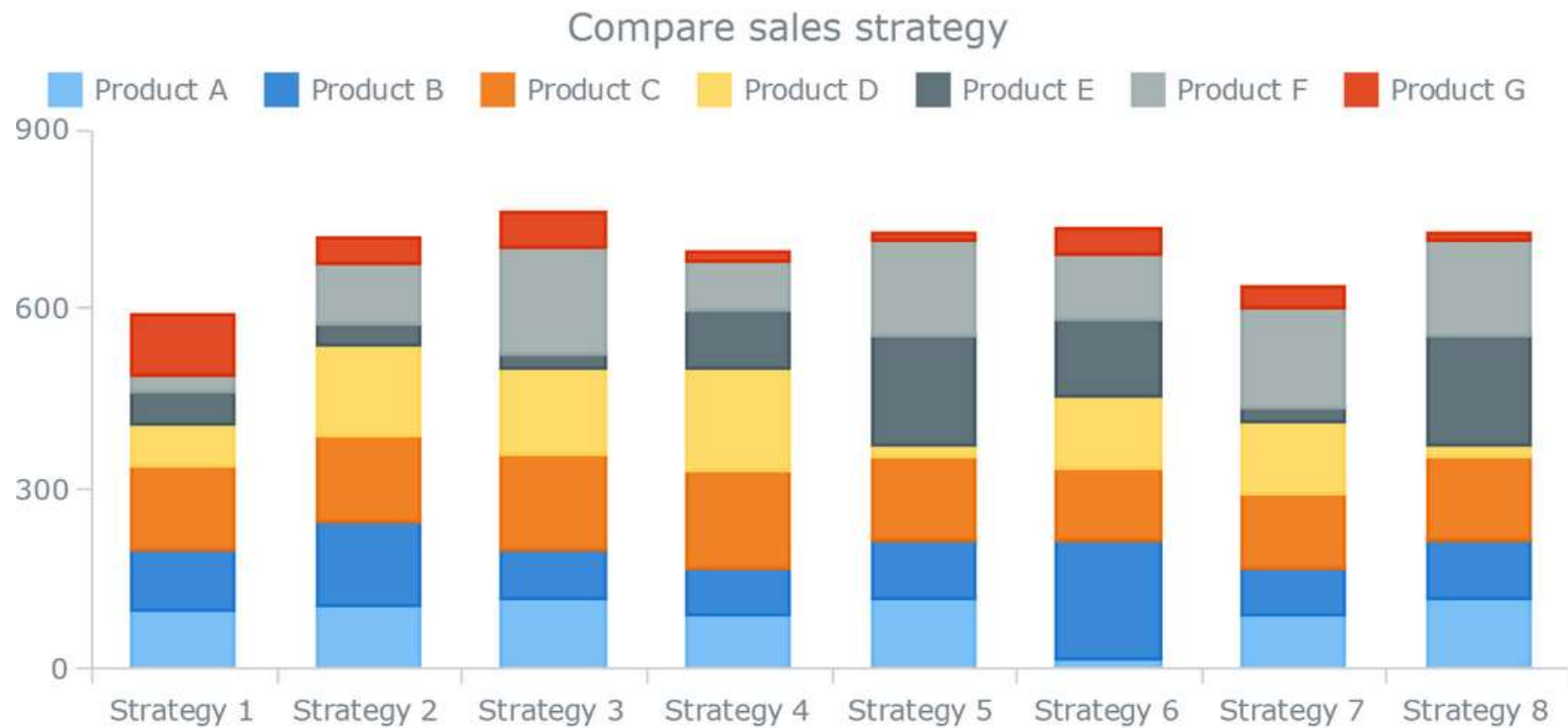
Regel 2: Gutes Verhältnis zwischen Daten und Druckerschwärze

Anteil der Studierenden an der Gesamtbevölkerung
in %



Datenvisualisierung: Grundregeln

Regel 3: Grafiktyp muss zu Daten und Aussage passen



Quelle: Vitaly Radionov. <https://codepen.io/Radionov/pen/ALWARJ/>

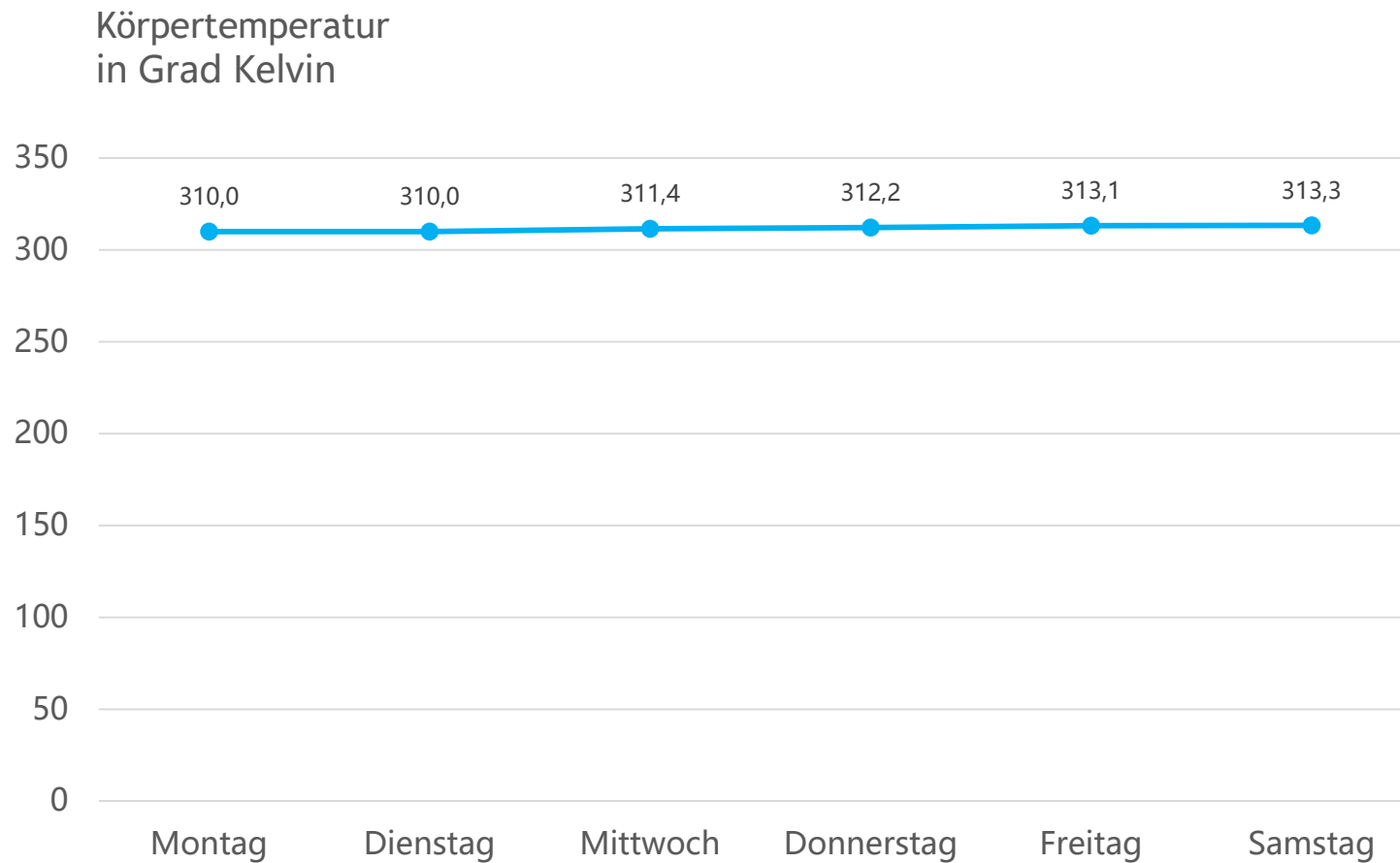
Regel 3: Grafiktyp muss zu Daten und Aussage passen

Entscheidend ist
Kernaussage: Zeitvergleich,
Verteilung, Korrelation,
Ranking etc.

ft.com/vocabulary

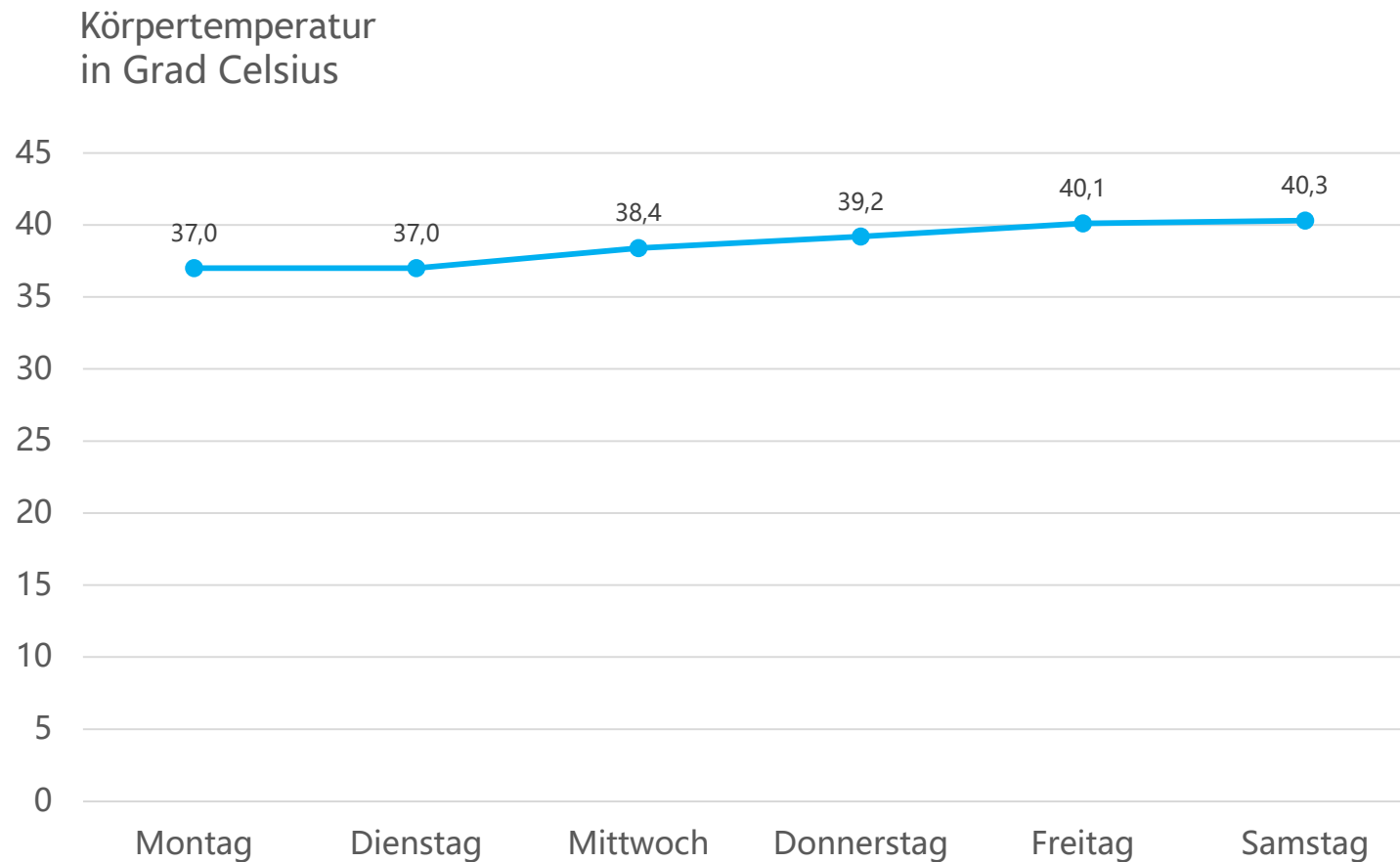
Datenvisualisierung: Grundregeln

Regel 4: Gestalte Achsen bewusst und sei Dir bewusst, dass Du sie gestaltest



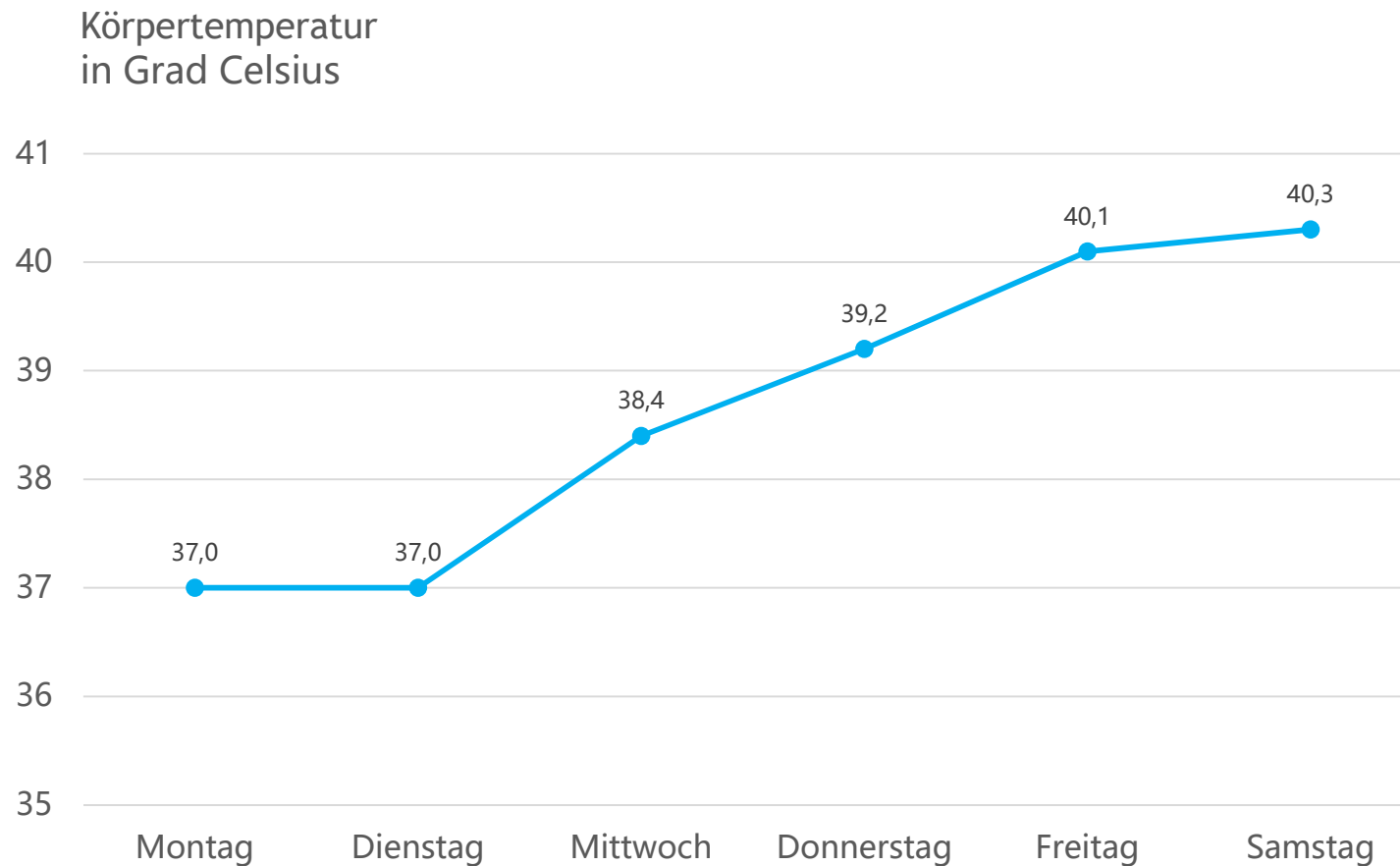
Datenvisualisierung: Grundregeln

Regel 4: Gestalte Achsen bewusst und sei Dir bewusst, dass Du sie gestaltest



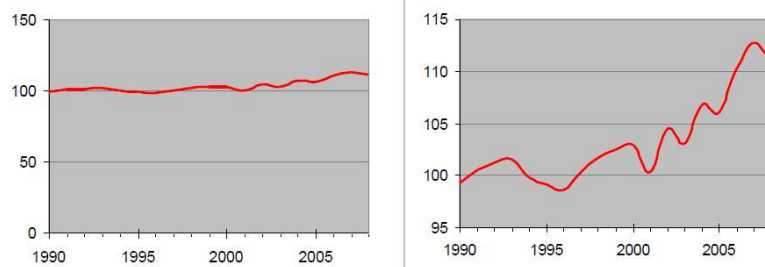
Datenvisualisierung: **Grundregeln**

Regel 4: Gestalte Achsen bewusst und sei Dir bewusst, dass Du sie gestaltest



Datenvisualisierung: Grundregeln

Regel 4: Gestalte Achsen bewusst und sei Dir bewusst, dass Du sie gestaltest



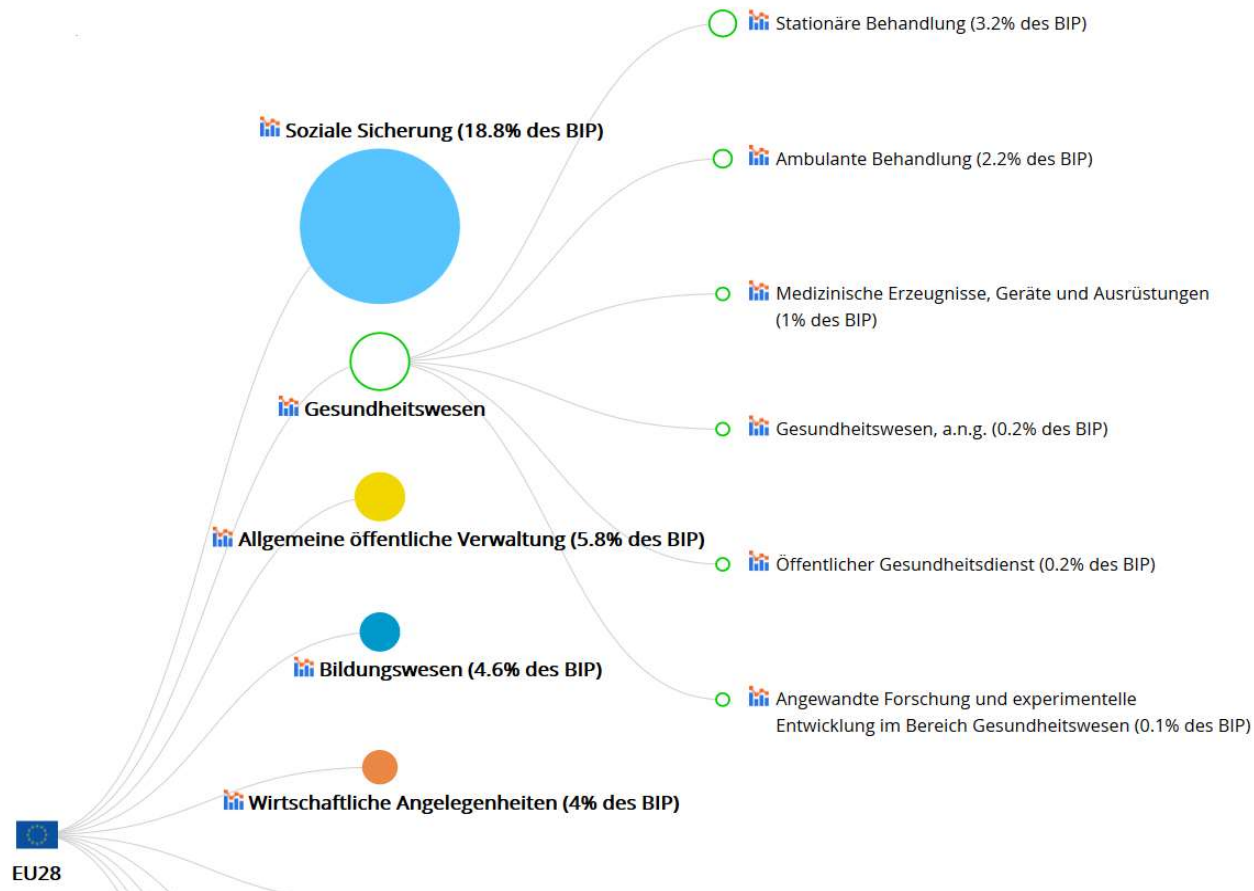
Die Frage ob die Y-Achse bei Null beginnen sollte oder nicht, ist komplex. Und die Antwort ist kontextabhängig. Wo die X-Achse beginnt ist bei Zeitverläufen eine mindestens ebenso relevante Frage. Wichtig ist, dass man Entscheidungen bewusst trifft und dass man Achsen auch klar kennzeichnet. Will man eine bestimmte Entwicklung hervorheben, ist ggf. Veränderungsrate bessere Lösung

www.learnviaweb.com/y-axis-zero-not-zero

[Office for National Statistics Blog](#)

Datenvisualisierung: Grundregeln

Regel 5: Deine Kernaussage sollte Dich leiten und Du solltest auch eine haben



Datenvisualisierung: Grundregeln

Regel 6: Wähle Farben vorsichtig

Angst Warnung Schulden Liebe	Sonne Helligkeit Positivität	Natur Ökologie Guthaben
Neutralität Dunkelheit Sachlichkeit Schwere	Wärme Freude Glück	Religion Träume

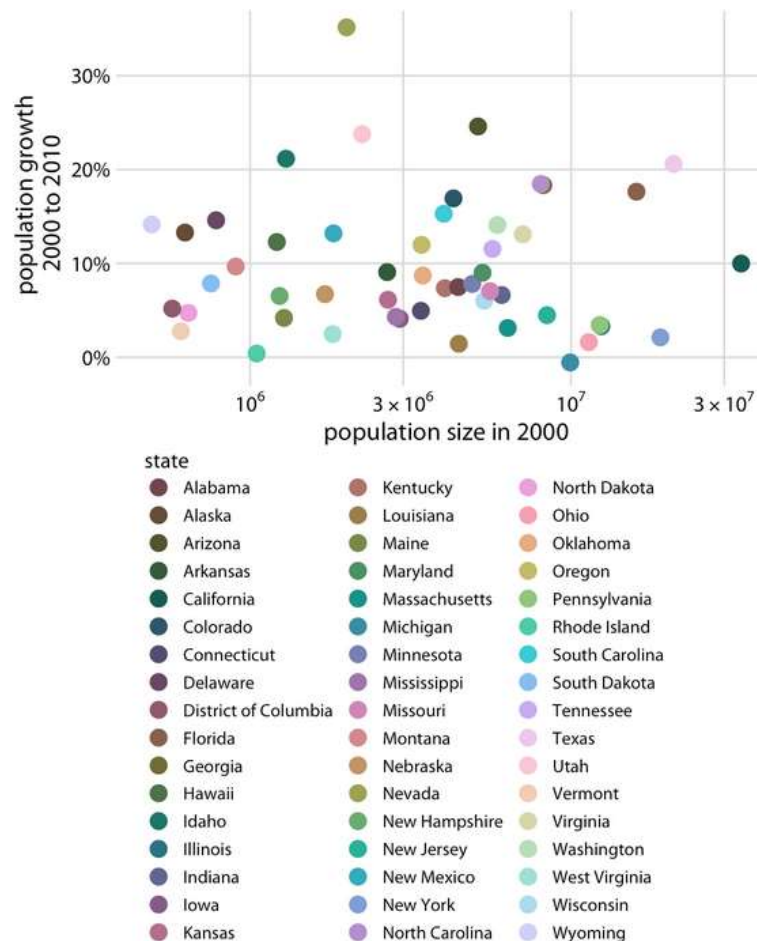
Farben verbindet man mit Eigenschaften. Zugleich zeigt dieses Bild ein weiteres Problem auf: Lesbarkeit von Schriften, Sichtbarkeit von Farben.

cambridge-intelligence.com/choosing-colors-for-your-data-visualization

colorbrewer2.org

Datenvisualisierung: Grundregeln

Regel 6: Wähle Farben vorsichtig



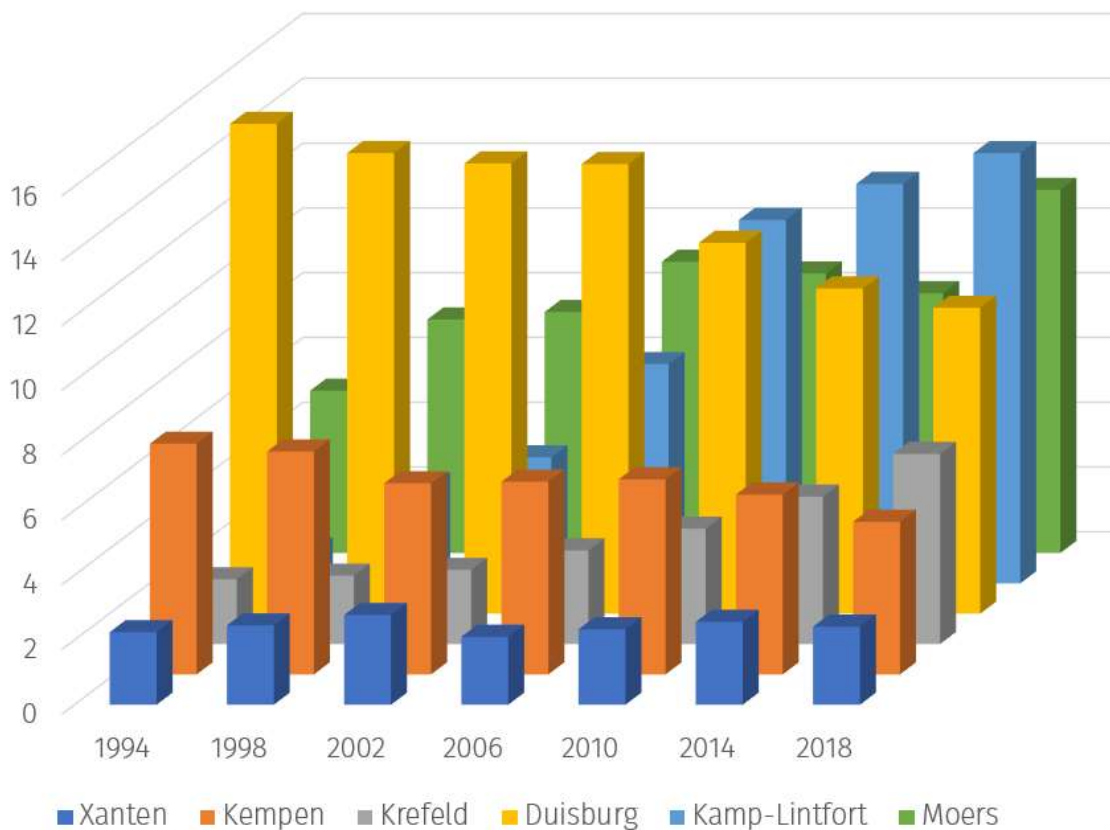
Überforderung: Man kann auch von Farben zu viel erwarten. Und man sollte mehrere Farben nur verwenden wenn es Mehrwert hat



Datenvisualisierung: Grundregeln

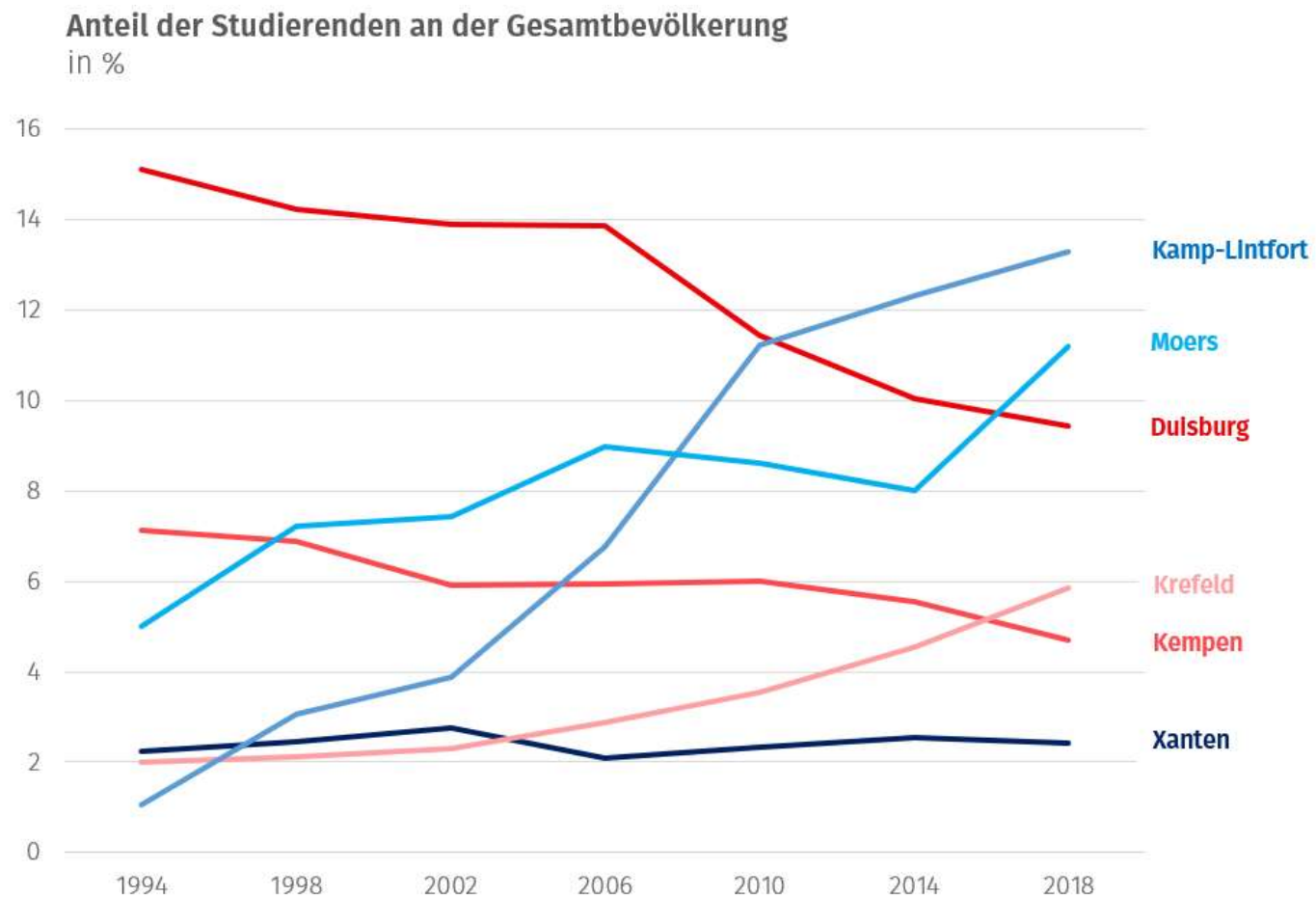
Regel 7: Bedenke die Grenzen menschlicher Perzeption – 3D vermeiden

Anteil der Studierenden an der Gesamtbevölkerung
in %



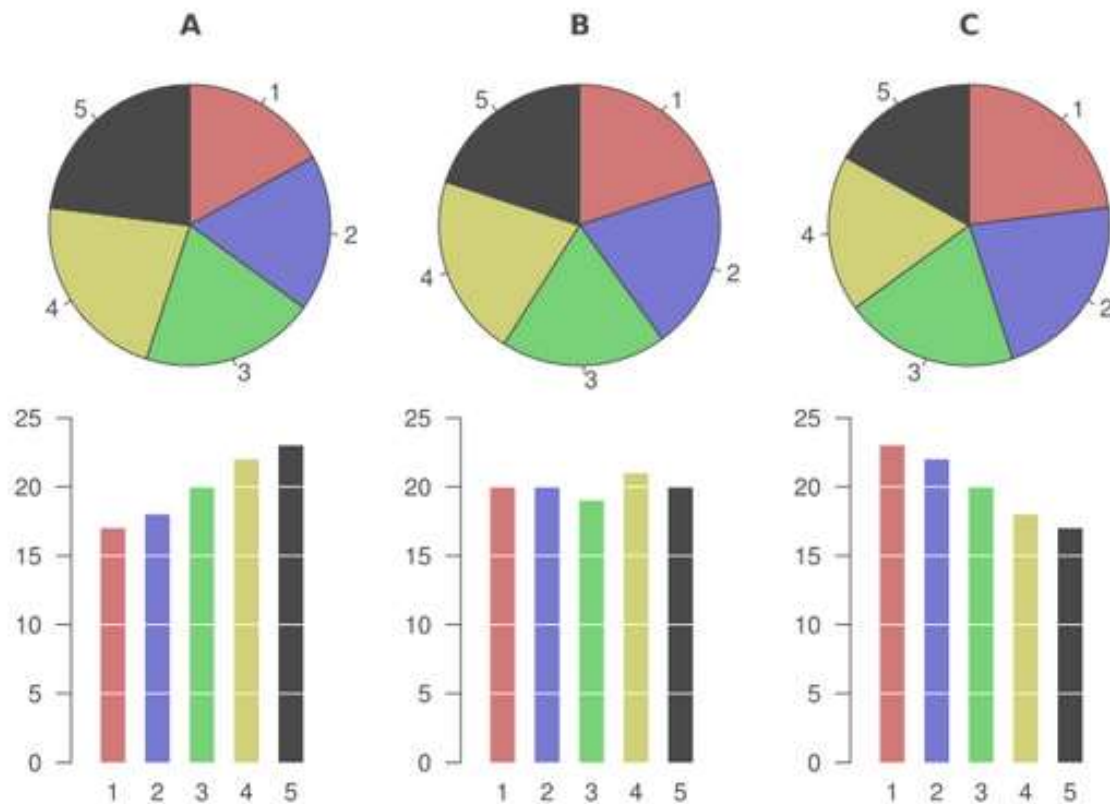
Datenvisualisierung: Grundregeln

Regel 7: Bedenke die Grenzen menschlicher Perzeption – 3D vermeiden



Datenvisualisierung: Grundregeln

Regel 7: Bedenke die Grenzen menschlicher Perzeption – Kreisdiagramme



Quelle: Frédéric Schütz
www.swissinfographics.com/archives/author/frederic-schutz

Die Interpretation von Winkeln fällt uns schwerer als die Interpretation von Linien oder Balken

Datenvisualisierung: Grundregeln

Regel 7: Es gibt auch Experten, die Kreisdiagramme verteidigen

Bevölkerung im Alter von 55 bis 64 Jahre

Nach Erwerbsstatus in %



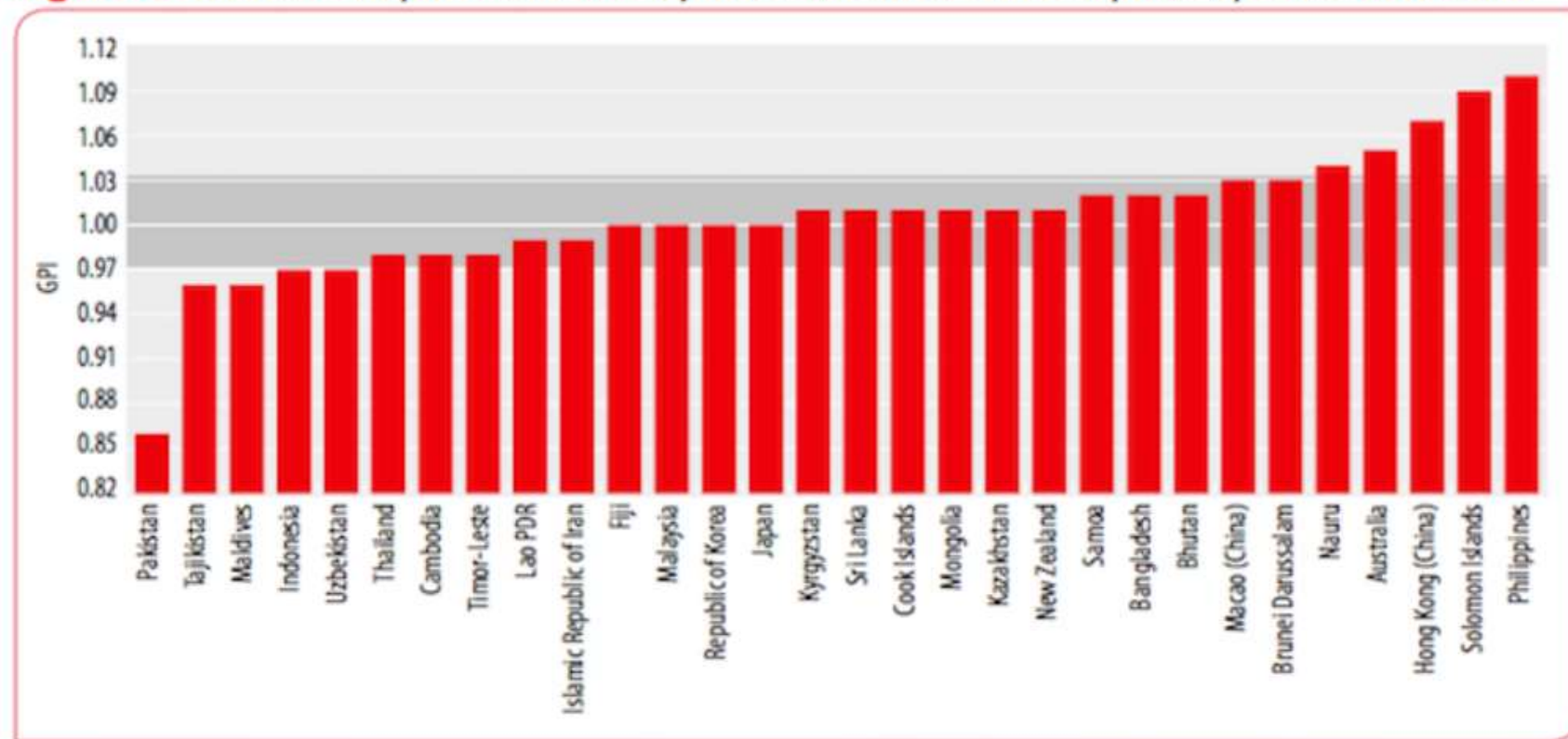
Bedingungen

1. Nur wenige Kategorien
2. Immer oben beginnen
3. Nach Größe sortieren
4. Vergleich Torte 1 mit Torte 2 steht nicht im Fokus
5. Segmente werden möglichst gut beschriftet
6. Es geht wirklich um den Anteil der Segmente am Ganzen

Datenvisualisierung: Grundregeln

Regel 8: Kenne Deine Zielgruppe und setze nicht zu viel Wissen voraus

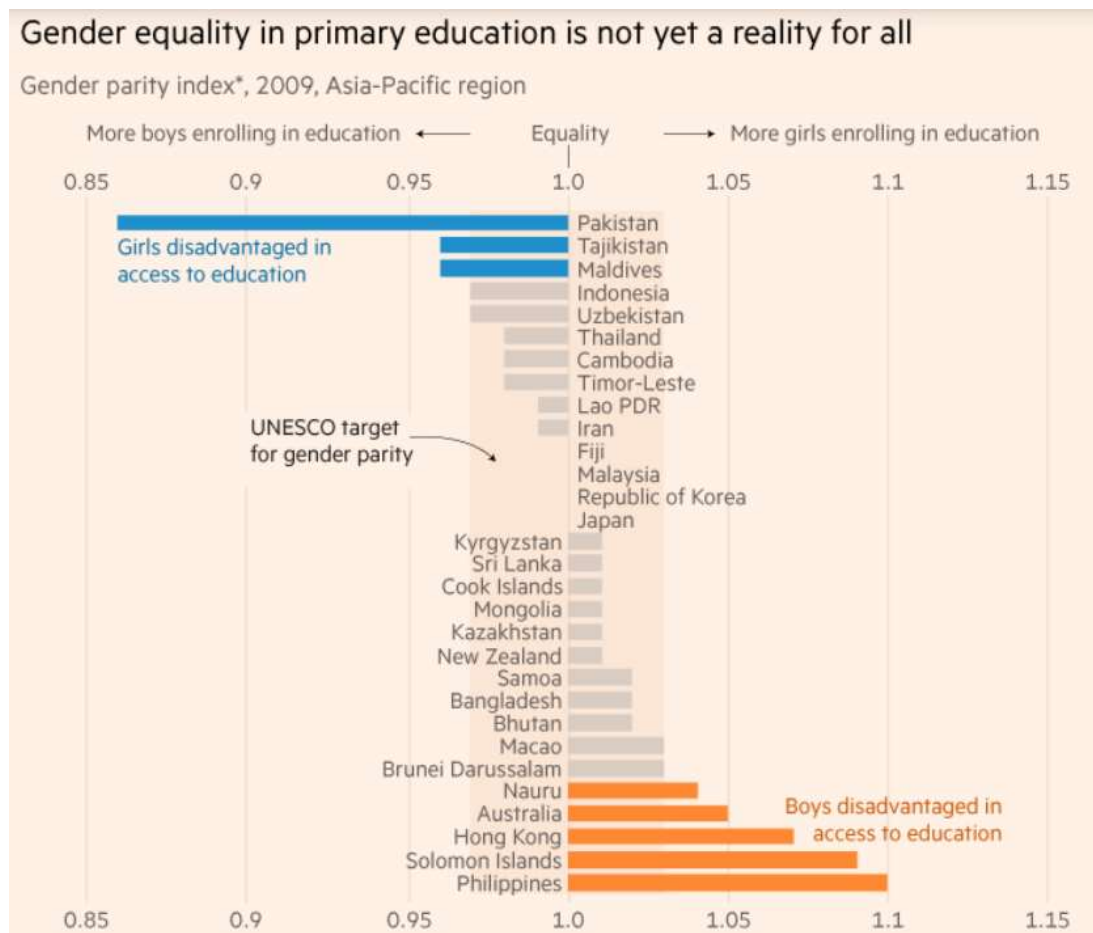
Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.

Datenvisualisierung: Grundregeln

Regel 8: Kenne Deine Zielgruppe und setze nicht zu viel Wissen voraus



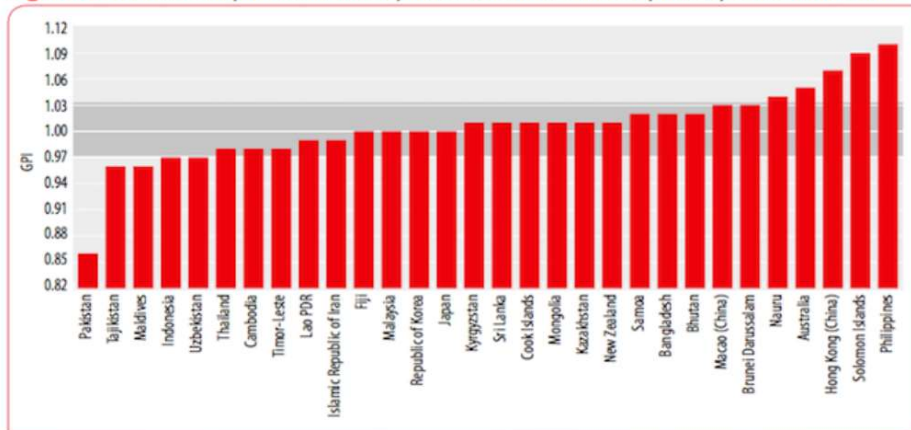
Quelle: UIS, 2011. Statistical Table 2. Graphic by Alan Smith. © FT.

Gute Balance:
Notwendige Zusatzinfos
ja aber Grafik nicht
überfrachten

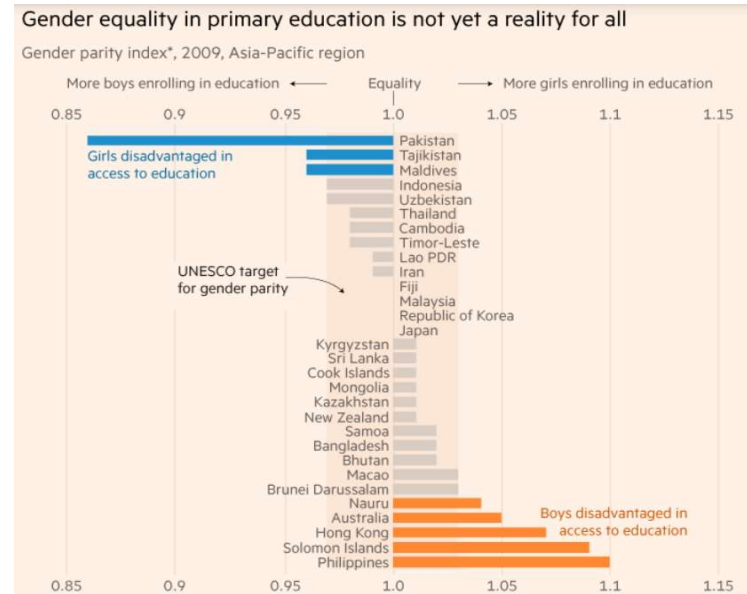
Datenvisualisierung: Grundregeln

Regel 9: Gestalte Deine Grafik lesefreundlich

Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.



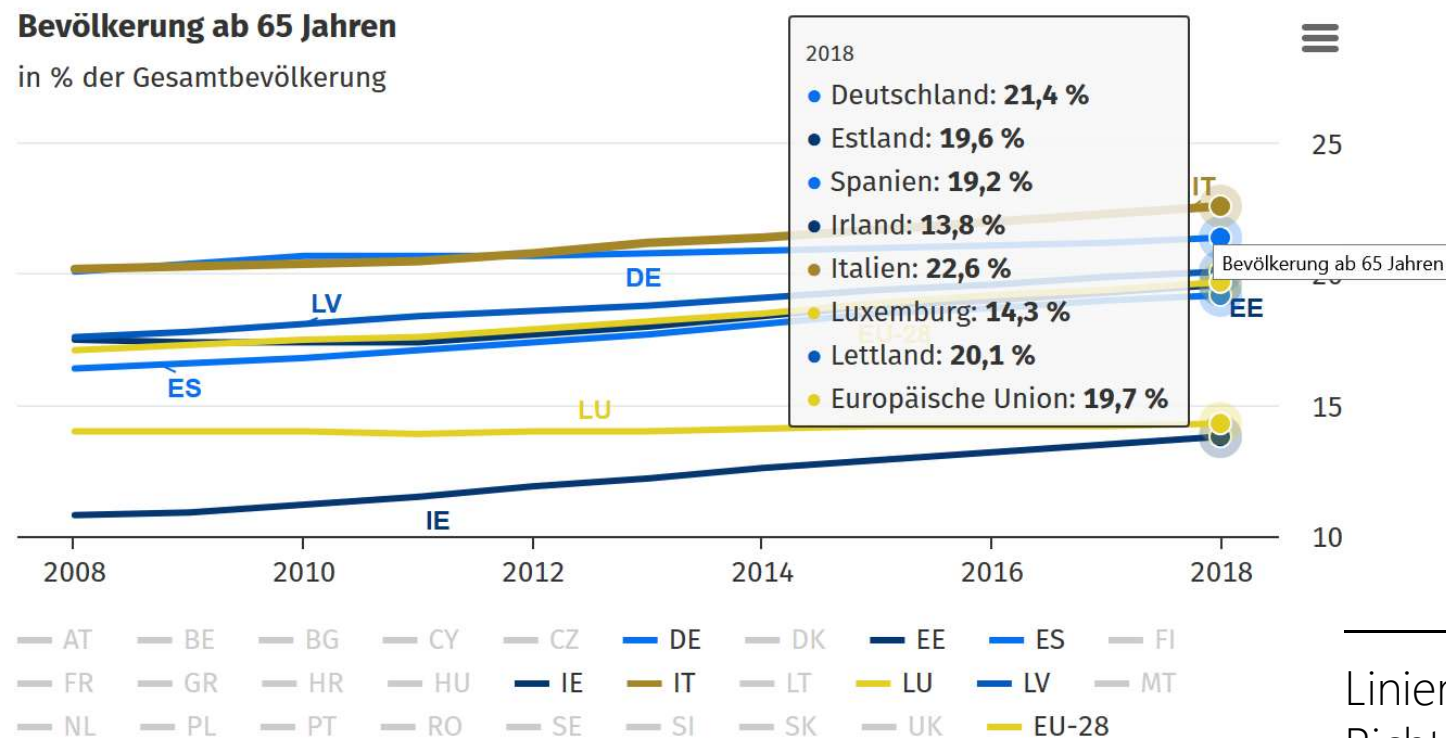
Quelle: UIS, 2011. Statistical Table 2. Graphic by Alan Smith. © FT.

Einige Beispiele

1. Balkendiagramm häufig besser als Säulendiagramm
2. Linien- und Kreisdiagramme direkt beschriftet statt Legende
3. Prüfen, dass Werte wirklich gut erkennbar sind

Datenvisualisierung: Grundregeln

Regel 10: Begrenze die Zahl der Kategorien bei Grafiken nach Möglichkeit



Quelle: Destatis.

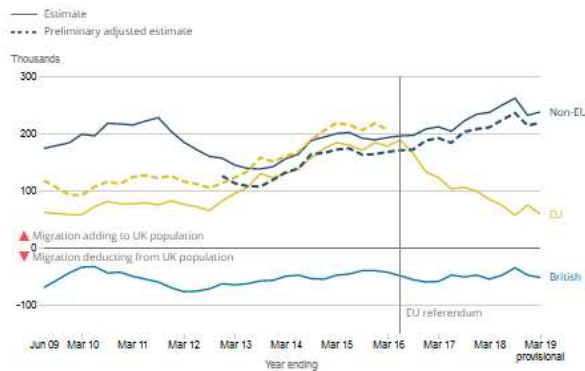
Liniendiagramm:
Richtwert max 6

Kreisdiagramm:
Richtwert max 4-5

Datenvisualisierung: Grundregeln

Regel 11: Verwende verbreitete Formate, achte auf Wiederverwendbarkeit und Nachvollziehbarkeit

Net migration by citizenship, UK, year ending June 2009 to year ending March 2019



Source: Office for National Statistics – Long-Term International Migration (LTIM), LTIM with preliminary adjustments based on DWP and Home Office data

Notes:

1. Preliminary adjusted estimates are shown as dashed lines on the chart.
2. Preliminary adjusted estimates for EU8 citizens have been produced for year ending June 2009 to year ending March 2016 and preliminary adjusted estimates for non-EU students have been produced for year ending December 2012 to year ending March 2019. Preliminary adjusted estimates for net migration have been produced for year ending March 2012 to year ending March 2019.
3. Confidence intervals for the unadjusted estimates are not shown in the chart, but are available in the accompanying dataset.
4. Confidence intervals for the preliminary adjusted estimates are not yet available. This does not mean that the estimates are more certain. As we refine our methods, we will develop our approach for displaying uncertainty around estimates based on multiple data sources.

Download this chart

[Image](#) [.xls](#) [.csv](#)

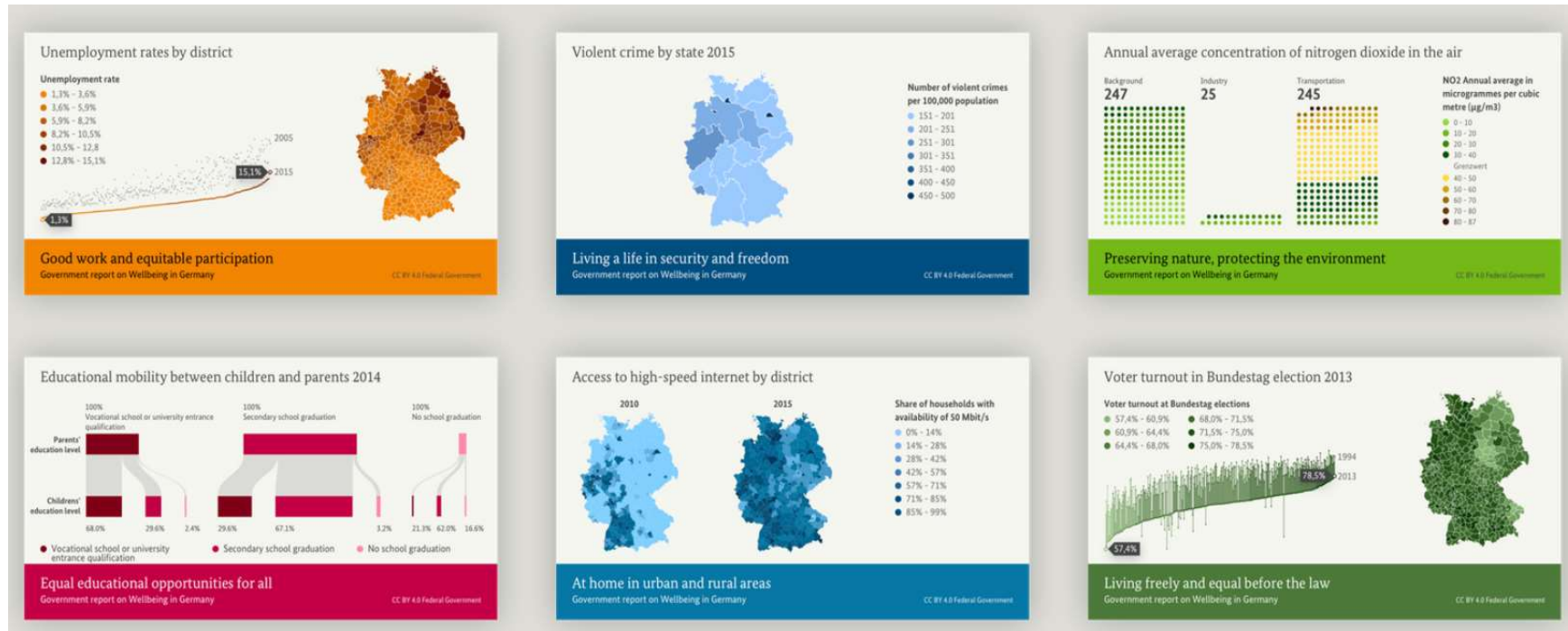
► [Embed code](#)

Quelle: Office for National Statistics (ONS)

-
1. Angabe von Quellen
 2. Nach Möglichkeit Primärquelle nutzen
 3. Code bereitstellen – zum Beispiel auf GitHub
 4. Daten und Grafik zum Download anbieten
 5. Einbetten der Inhalte ermöglichen

Gelungenes Beispiel

Gut leben in Deutschland: Interaktiver Bericht



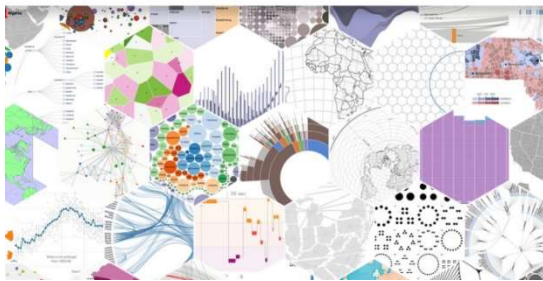
www.gut-leben-in-deutschland.de/static/LB

www.interactivethings.com/work/wellbeing-in-germany

Interessante Ressourcen | Online

Wo finde ich mehr Info?

 Data-Driven Documents



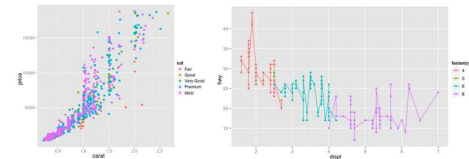
d3js.org



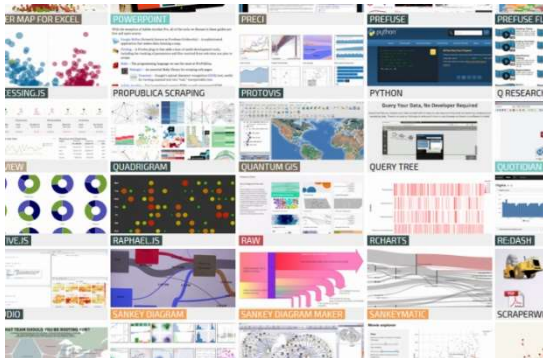
THE PYTHON
GRAPH GALLERY

python-graph-gallery.com

Data Visualization with 



rkabacoff.github.io/datavis/



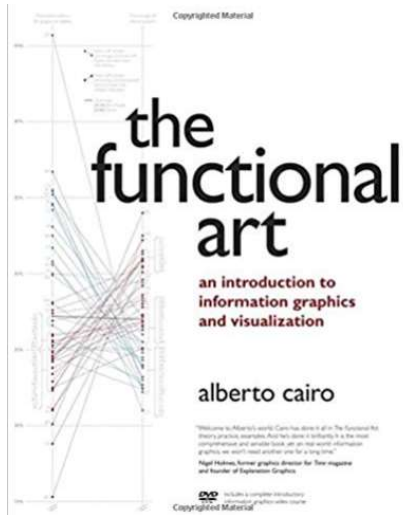
www.visualisingdata.com

 **stackoverflow**

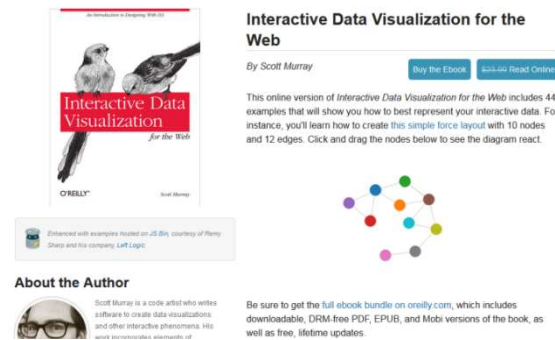
stackoverflow.com

Interessante Ressourcen | Bücher

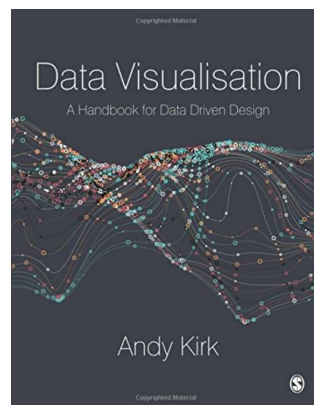
Wo finde ich mehr Info?



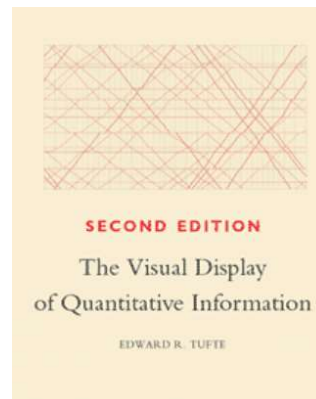
www.thefunctionalart.com



alignedleft.com/work/d3-book



www.visualisingdata.com/book



www.edwardtufte.com/tufte/books_vdqi