

Woche 05 Daten: Visualisierung

Skript

Erarbeitet von
Ann-Kathrin Selker

| | |
|--------------------------------|----|
| Lernziele | 1 |
| Inhalt | 1 |
| Einstieg..... | 1 |
| Diagrammarten..... | 2 |
| Tipps für Präsentationen | 6 |
| Abschluss | 10 |
| Quellen | 11 |
| Weiterführendes Material..... | 11 |
| Disclaimer | 11 |

Lernziele

- den Nutzen von Visualisierung erläutern können
- verschiedene Diagrammarten benennen und deren Anwendung mit Beispielen verdeutlichen können
- Grafiken mithilfe der gegebenen Tipps bewerten können

Inhalt

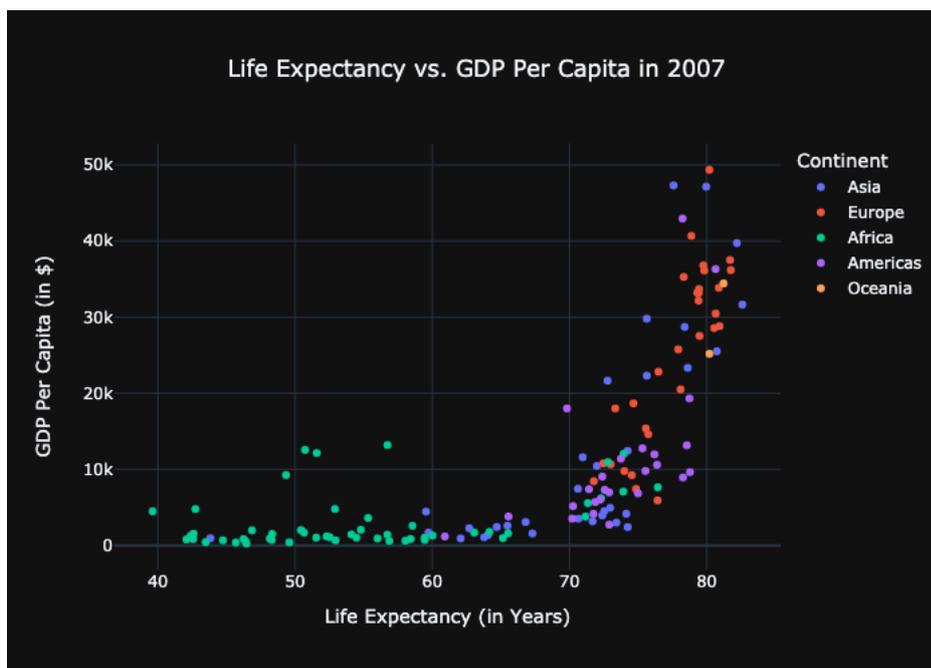
Es ist sehr wichtig, sich gut mit seinen Daten auszukennen. Doch was für Möglichkeiten gibt es hierfür, außer auf seine Daten zu starren? Und wie kann ich meine Erkenntnisse anderen mitteilen?

Einstieg

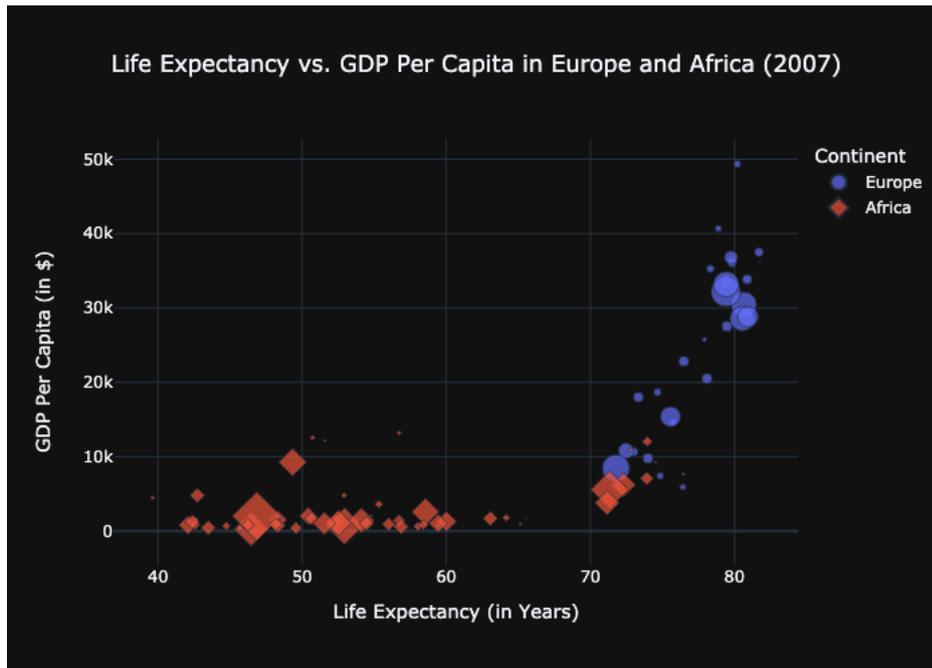
Die Beispieldaten aus diesem Video stammen von der Organisation Gapminder. In den Gapminder-Daten liegen unter anderem Werte für Bevölkerung, Lebenserwartung und BIP, also das Bruttoinlandsprodukt pro Kopf in den Ländern dieser Welt, vor. Wie verteilt sich die Bevölkerung auf die Kontinente? Wie hat sich die Lebenserwartung in Deutschland in den letzten Jahren entwickelt? Und wie hoch ist das BIP in Deutschland im Vergleich zu Nachbarländern? Die Antworten auf all diese Fragen liegen in den Daten, und wir können sie mit statistischen Mitteln herausfinden. Anschaulicher und häufig schneller können grobe Fragen wie diese aber mit Datenvisualisierung beantwortet werden. Je nachdem, was mit der Visualisierung gezeigt oder untersucht werden soll, sollten unterschiedliche Visualisierungstypen verwendet werden. Hier zeige ich kurz ein paar typische Anwendungsbeispiele. Natürlich kann mit jedem Grafiktyp noch viel mehr angezeigt werden, als ich hier vorstelle.

Diagrammarten

Um deine Datenpunkte anzuzeigen, eignet sich ein sogenanntes Streudiagramm (im englischen *scatter plot*). Hier werden alle deine Punkte auf gewählten Achsen angezeigt, wie hier z. B. alle Länder in Bezug auf Lebenserwartung und BIP.

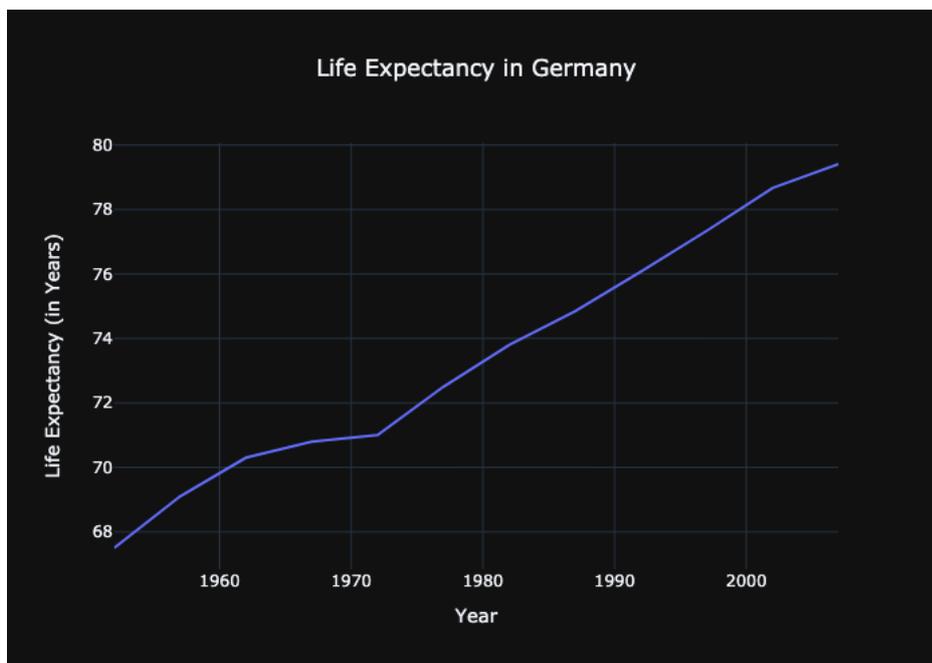


Du kannst abhängig von den Werten mancher Features die Markierungen, Farbe und Größe deiner Punkte ändern, um einen größeren Überblick zu bekommen. Das folgende Diagramm zeigt zum Beispiel, beschränkt auf die Kontinente Europa und Afrika, wie hoch die Lebenserwartung und BIP sind, während die Größe des Punktes auch noch die Bevölkerungszahl angibt. Mit dem Streudiagramm kannst du dir also auch einen ersten Überblick darüber verschaffen, wie deine Daten verteilt sind und was für Zusammenhänge bestehen könnten.



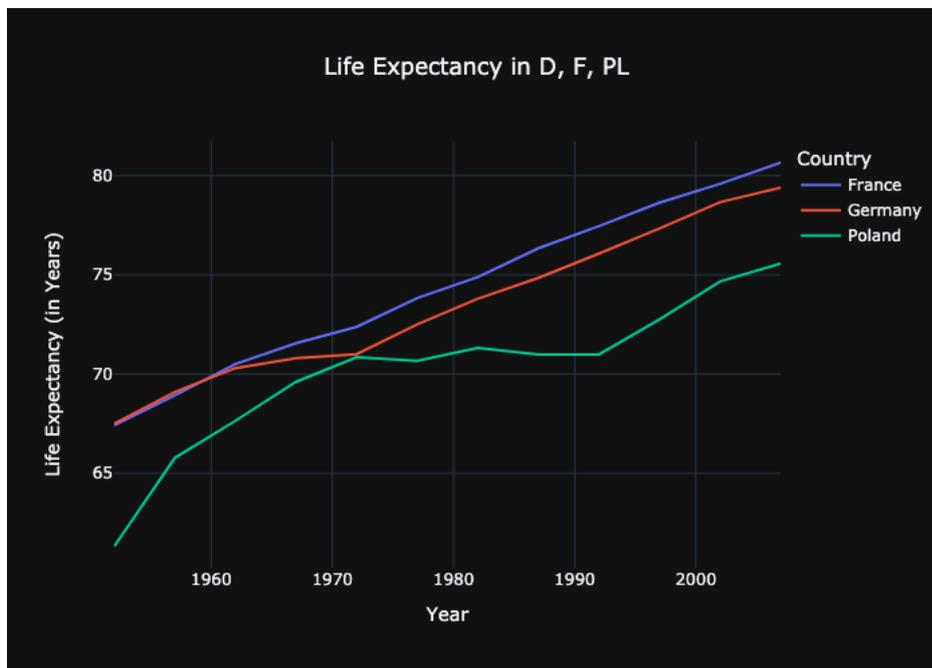
Einblendung Streudiagramme

Um zeitliche Verläufe und Trends darzustellen, eignen sich unter anderem Liniendiagramme sehr gut. Hier werden die Datenpunkte durch eine Linie verbunden, sodass man schnell Trends erkennen kann. Ein Beispiel hierfür ist das folgende Diagramm. Es wird die Entwicklung der Lebenserwartung in Deutschland dargestellt.



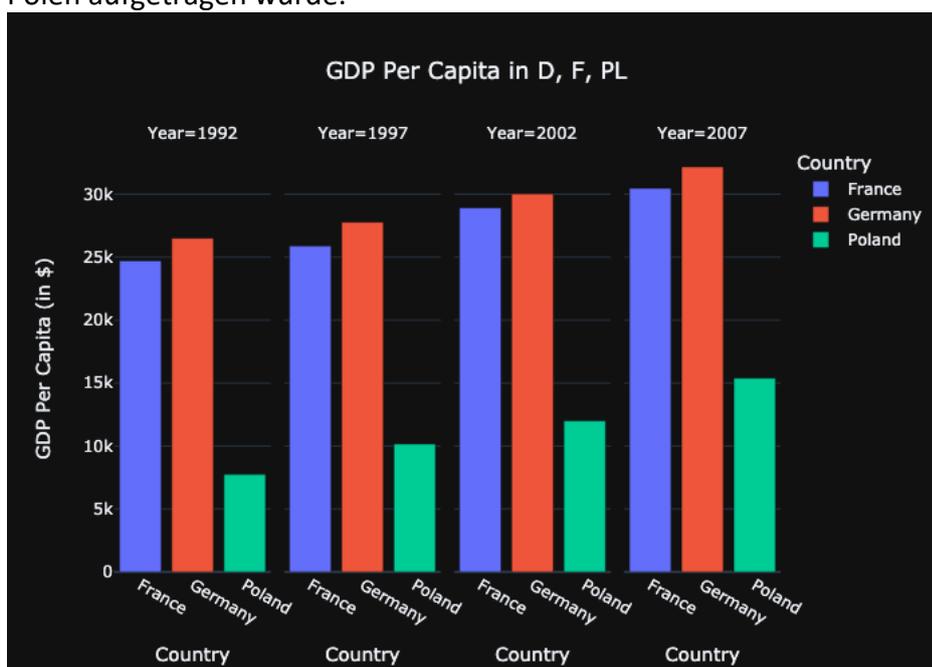
Man sieht schnell, dass sich die Lebenserwartung im Aufwärtstrend befindet. Wenn man diese Trends jetzt auch noch vergleichen möchte, können auch mehrere Liniendiagramme

zusammen dargestellt werden, wie hier mit den Ländern Deutschland, Frankreich und Polen.

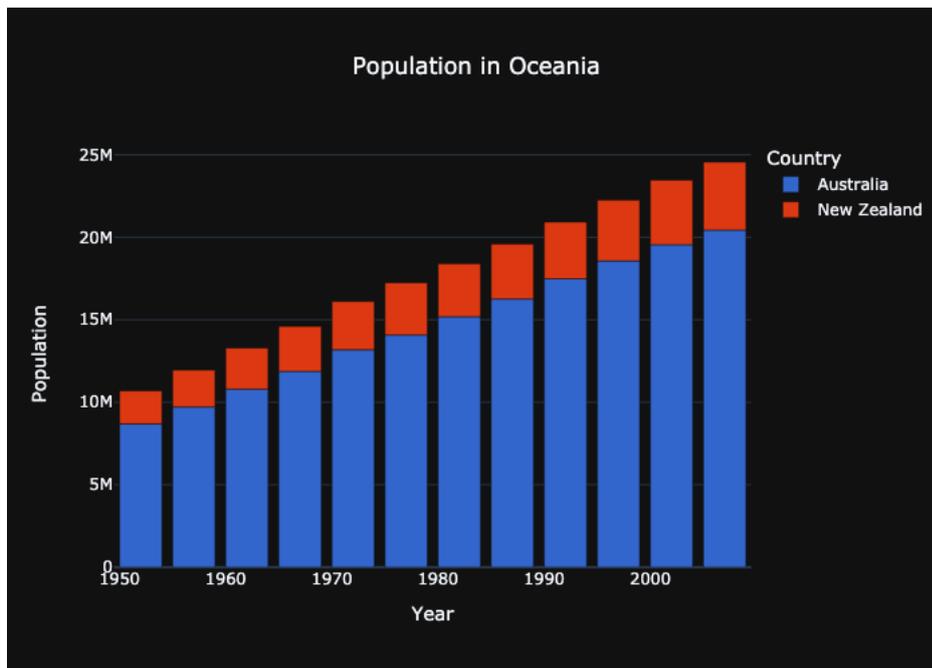


Einblendung Liniendiagramme

Zum Vergleichen von Daten eignen sich Balken- bzw. Säulendiagramme gut. Das Diagramm heißt Balkendiagramm, wenn die Balken horizontal verlaufen, und Säulendiagramm, wenn sie vertikal verlaufen. Die horizontale Variante bietet sich an, wenn Balken deutliche Größenunterschiede aufweisen. Ein Beispiel für ein Säulendiagramm ist dieses Diagramm, bei dem das Bruttoinlandsprodukt pro Kopf für die Länder Deutschland, Frankreich und Polen aufgetragen wurde.

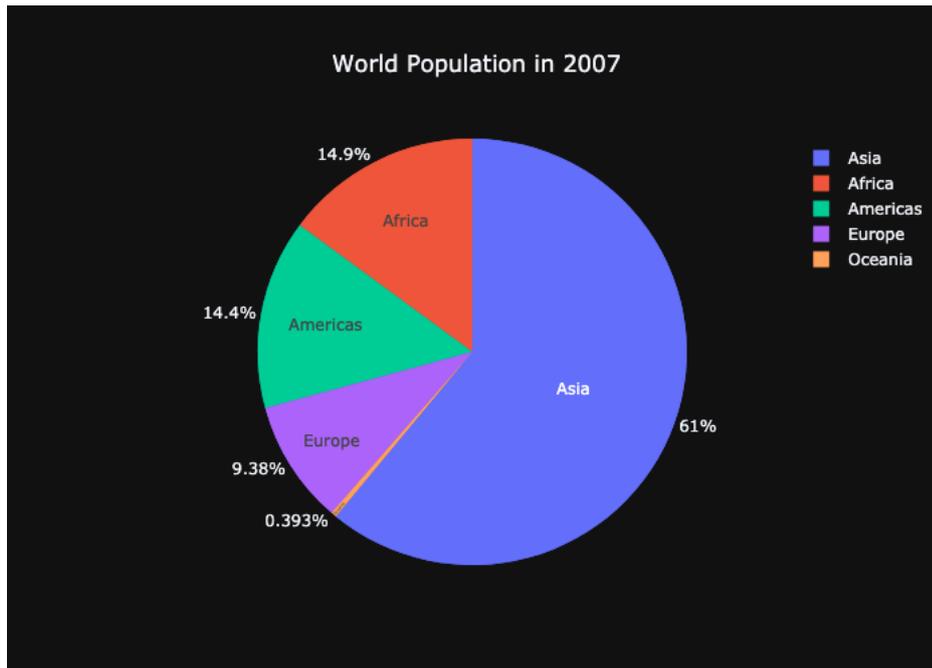


Balken- und Säulendiagramme gibt es auch in einer sogenannten „gestapelten“ Form, bei denen nicht nur der Gesamtwert, sondern auch die Zusammensetzung von Daten gezeigt wird. In diesem Diagramm kann man nicht nur die Bevölkerungszahl von Ozeanien im Verlauf der Jahre sehen, sondern auch, wie viel die beiden Länder Australien und Neuseeland jeweils dazu beigetragen haben.



Einblendung Säulendiagramme

Eine weitere Diagrammart, um die Zusammensetzung von Daten zu zeigen, ist zum Beispiel das Kuchendiagramm. Hier wird jedem Wert ein Kuchenstück proportional zum Anteil an den Gesamtdaten zugewiesen. Das folgende Diagramm zeigt den Bevölkerungsanteil der Kontinente an der Gesamtbevölkerung. Die Bevölkerung Asiens macht über die Hälfte der Erdbevölkerung aus, daher ist das entsprechende Kuchenstück auch dementsprechend groß. Da es sich bei den Kuchenstücken um die Anteile von einem ganzen Kuchen handelt, ist es ganz wichtig, dass sich die gezeigten Prozentzahlen zu 100 % aufaddieren.

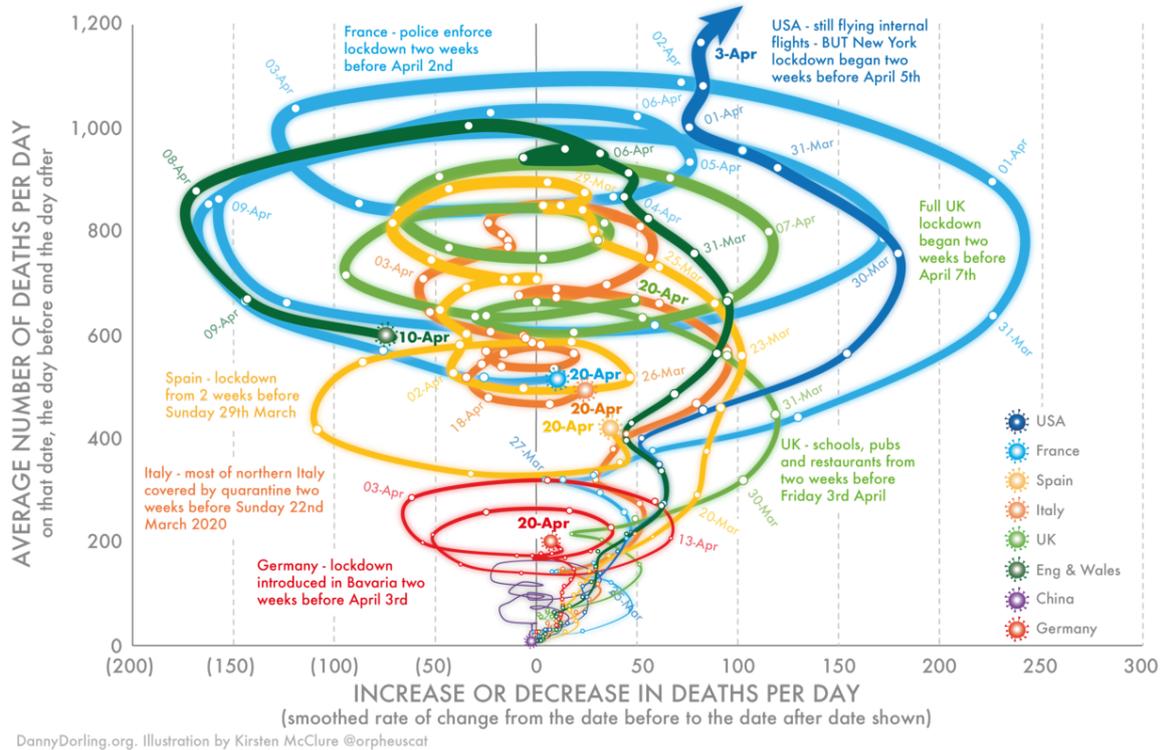


Einblendung Kuchendiagramm

Tipps für Präsentationen

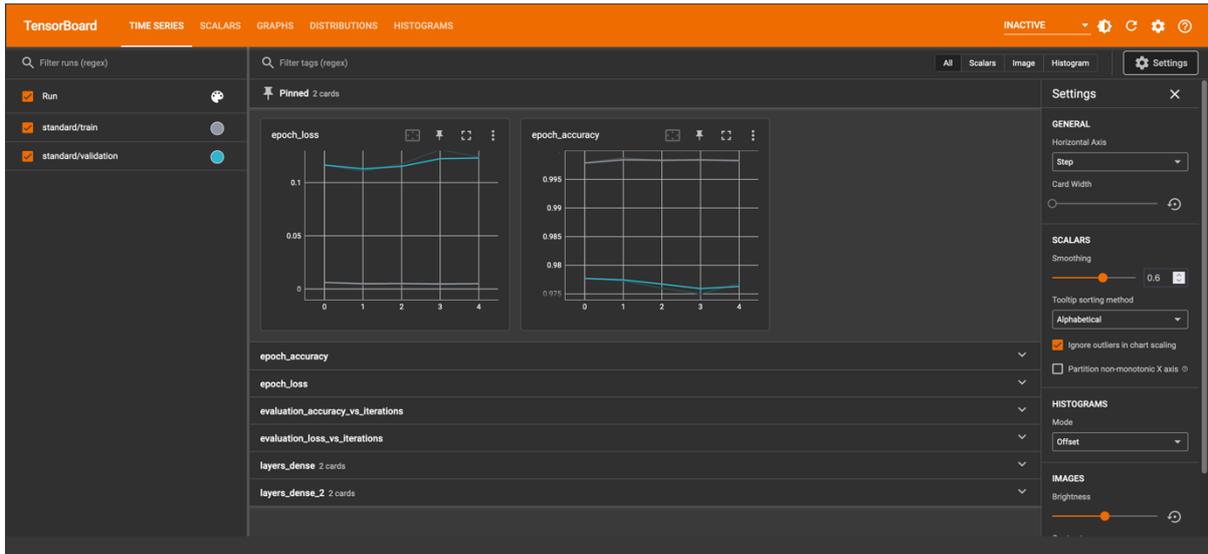
Diagramme sind aber natürlich nicht nur für die grobe Datenanalyse geeignet. Du willst deine Zwischen- oder Endergebnisse sicher auch anderen Personen präsentieren. Für diese Diagramme gibt es aber andere Anforderungen als für den Eigengebrauch. Richte die Art, wie du deine Informationen präsentierst, an deinem Publikum aus. Welche Vorkenntnisse und welchen Hintergrund hat dein Publikum? Informationen für Kinder müssen anders aufbereitet werden als Informationen für Fachfremde als Informationen für deine Kollegen und Kolleginnen.

Grafiken in Vorträgen müssen innerhalb von Sekunden verstanden werden und dürfen daher nicht überladen sein. Wenn ich dir jetzt diese Grafik vorsetze, liegt deine Aufmerksamkeit sicher für eine ganze Weile auf der Grafik und nicht mehr auf dem Sprecher, während du versuchst, die Aussage der Grafik zu verstehen. Wir können das ja kurz testen, hörst du mir gerade noch aufmerksam zu?



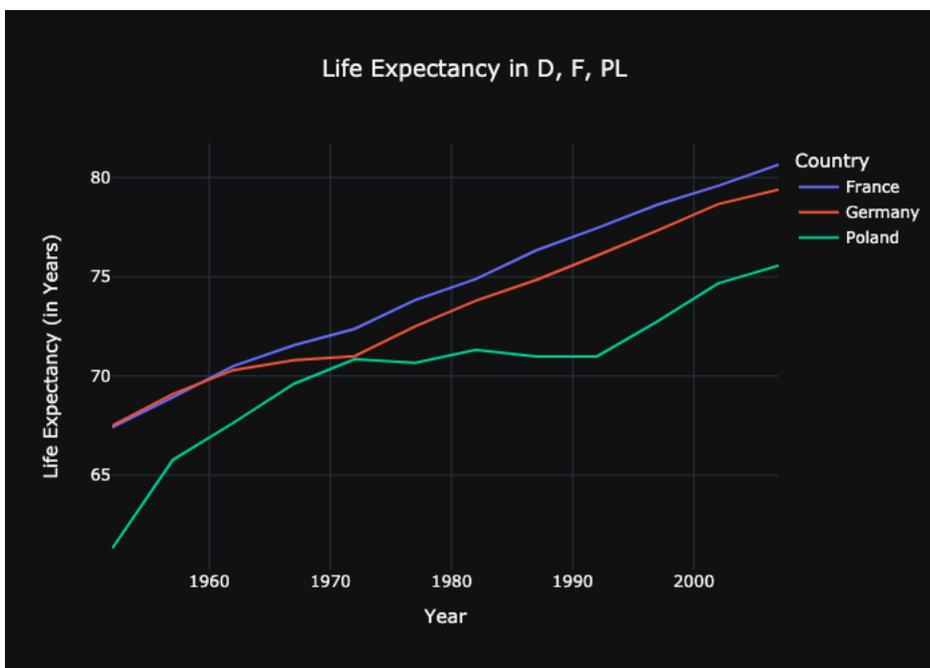
Einblendung Diagramm (Quelle [1])

Aber auch Grafiken, die du auf Dashboards veröffentlichst, sollten schnell zugänglich sein. Unter *Dashboard* verstehen wir eine grafische Benutzeroberfläche, auf der mehrere Visualisierungen zusammen angezeigt werden. Sinn der Sache ist es, die wichtigsten Informationen auf einen Blick erfassen zu können. Auf diesem Dashboard wird z. B. der Trainingsverlauf eines Machine-Learning-Modells visualisiert.



Einblendung Tensorboard-Dashboard

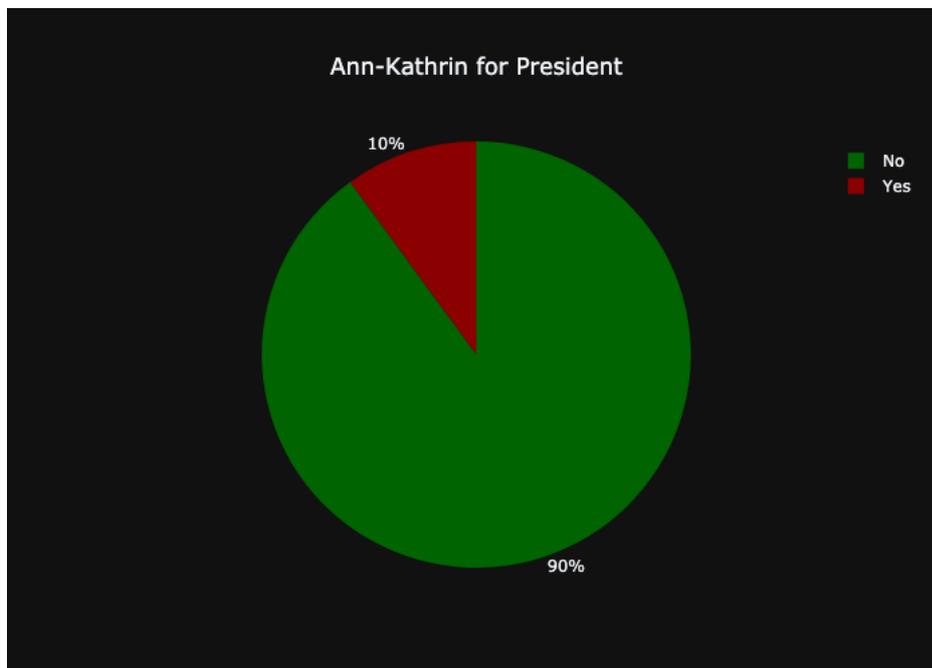
Wichtig ist es, dass jedem sofort klar ist, worum es in der Grafik geht, also immer an aussagekräftige Titel, Beschriftungen und Legenden denken. Zur Erinnerung: Eine Legende gibt an, welche Farbe und/oder Symbol welches Feature repräsentiert. In der Grafik hier findest du den Titel oben und die Legende rechts.



Einblendung Diagramm mit markiertem Titel und Legende

Außerdem möchtest du sicher nicht dein Publikum in die Irre führen: Pass also auf, dass du Größen proportional angibst, also zum Beispiel kein Kuchenstück als halben Kuchen darstellst, wenn es eigentlich nur ein Drittel des Kuchens ausmacht. Halte dich an gewisse Normen, was die Bedeutung von Farben und Symbolen angeht. So steht zum Beispiel grün

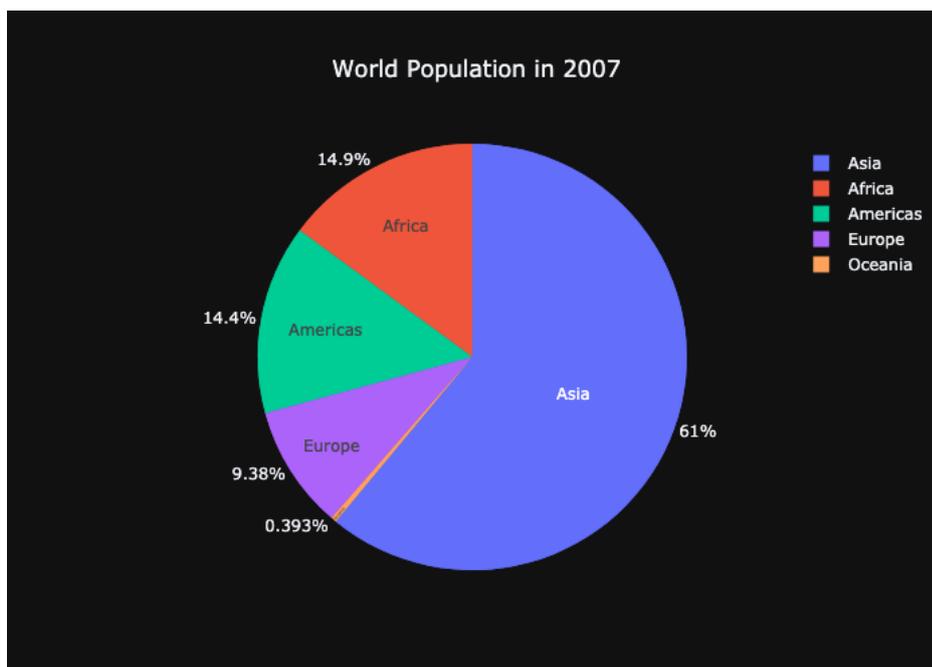
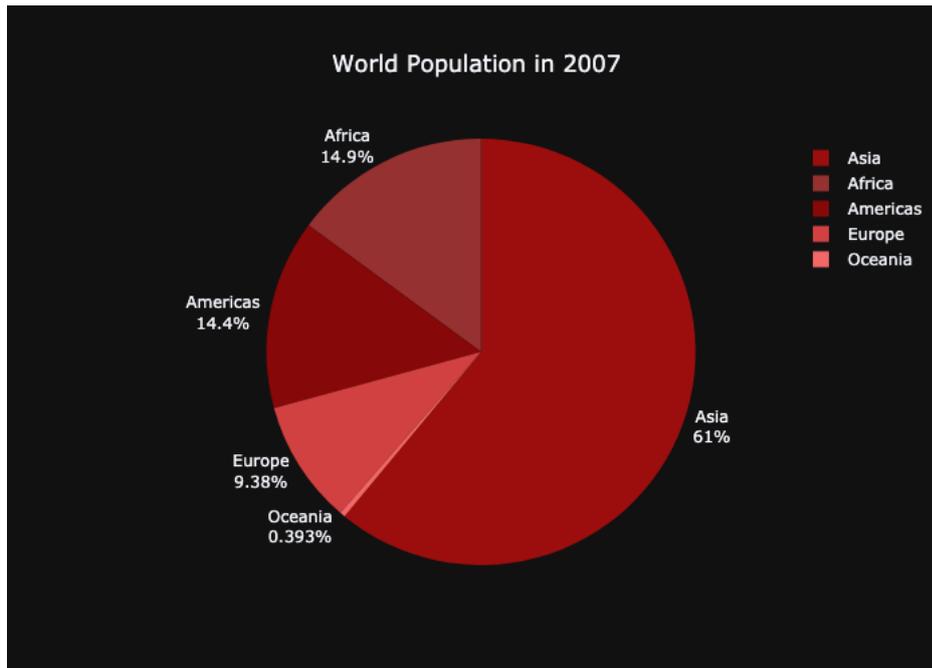
häufig für Zustimmung und rot für Ablehnung, sodass z. B. diese Grafik hier auf den ersten Blick missverstanden werden könnte.



Einblendung Kuchendiagramm

Pass auch auf deine Achsen auf. Die x-Achse wird von links nach rechts gelesen und die y-Achse von unten nach oben, daher sollten die Werte auch so angeordnet werden. Verzichte auf Sprünge in den Achsen, also z. B. 1, 2, 5, 9, 10, da dies deine Diagramme verzerrt. Diese und viele weitere Fehler bei der Gestaltung von Diagrammen können dein Publikum, ob jetzt gewollt oder ungewollt, die falschen Schlüsse über deine Daten ziehen lassen. Ein weiterer wichtiger Punkt für das schnelle Verständnis ist die Barrierefreiheit von Diagrammen. Dies betrifft z. B. die Farbwahl. Hier solltest du auf kontrastreiche Farben setzen. Sieh dir zum Beispiel den Unterschied zwischen diesen beiden Grafiken an, einmal mit kontrastarmen und einmal mit kontrastreichen Farben. Bei dem kontrastreichen Bild kann man die Legende klar zuordnen und auch Bereiche besser auseinanderhalten.

Einblendung Kuchendiagramme



Außerdem ist es hilfreich, wenn Farben nicht das einzige Unterscheidungsmerkmal zwischen zwei Teilen der Grafik sind. So bietet es sich z. B. an, in einem Streudiagramm zwei Punkteguppen nicht nur durch Farbe, sondern auch durch unterschiedliche Symbole unterscheidbar zu machen. Auch die Wahl der Schriftgröße- und art solltest du dir vorher überlegen.

Abschluss

In diesem Video hast du gelernt, wie dir Datenvisualisierung dabei helfen kann, dich besser mit deinen Daten auszukennen und erste Trends zu identifizieren. Du kannst jetzt

Diagrammarten benennen und ihre Anwendung mit Beispielen verdeutlichen. Außerdem hast du einige Tipps erhalten, wie du deine Grafiken präsentieren kannst.

Quellen

Quelle [1] DannyDorling.org. Illustration by Kirsten McClure @orpheuscat
Präsentiert von Lydia Hooper, <https://venngage.com/blog/bad-infographics/>

Weiterführendes Material

<https://www.kaggle.com/code/residentmario/welcome-to-data-visualization>

<https://www.kaggle.com/code/kanncaa1/plotly-tutorial-for-beginners>

Disclaimer

Transkript zu dem Video „Woche 05 Daten: Visualisierung“, Ann-Kathrin Selker.
Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY](https://creativecommons.org/licenses/by/4.0/) 4.0 veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.