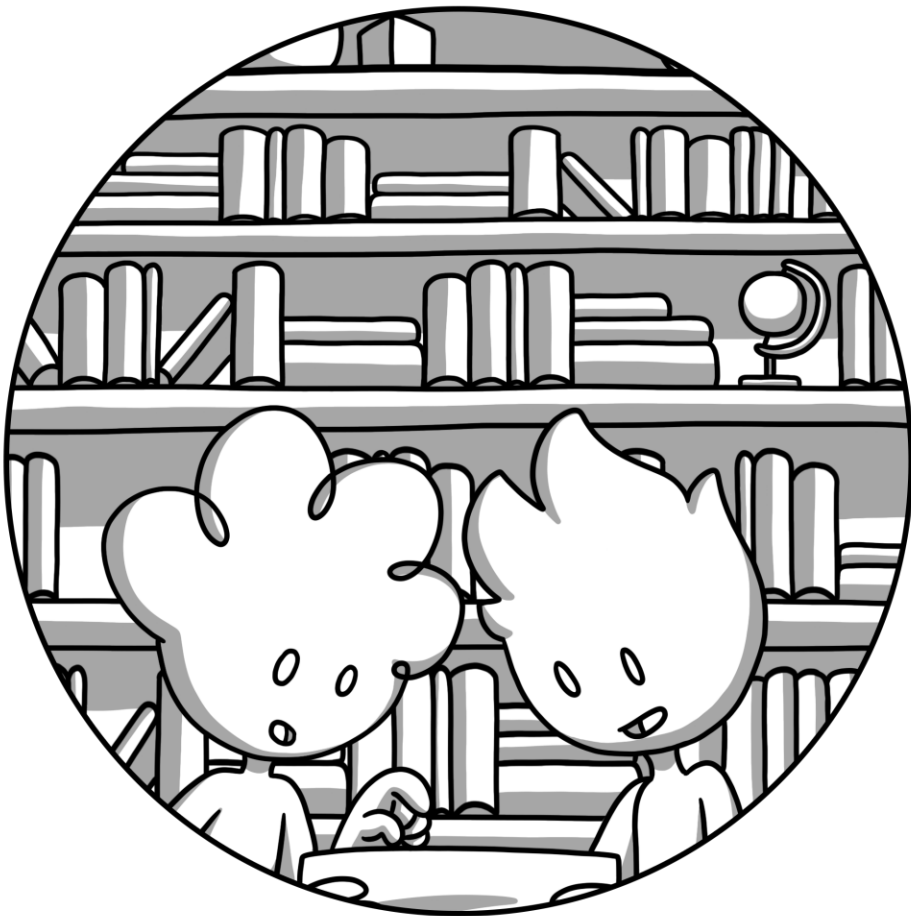


Datenattribute



COMIXPLAIN

Dieser Comic wurde im Zuge des Forschungsprojekts Comixplain, gefördert von der Fachhochschule St. Pölten im Rahmen des Innovation Call 2022, erstellt.

Projektteam:

Victor-Adriel De-Jesus-Oliveira
Hsiang-Yun Wu
Christina Stoiber
Magdalena Boucher
Alena Ertl

Kontakt:

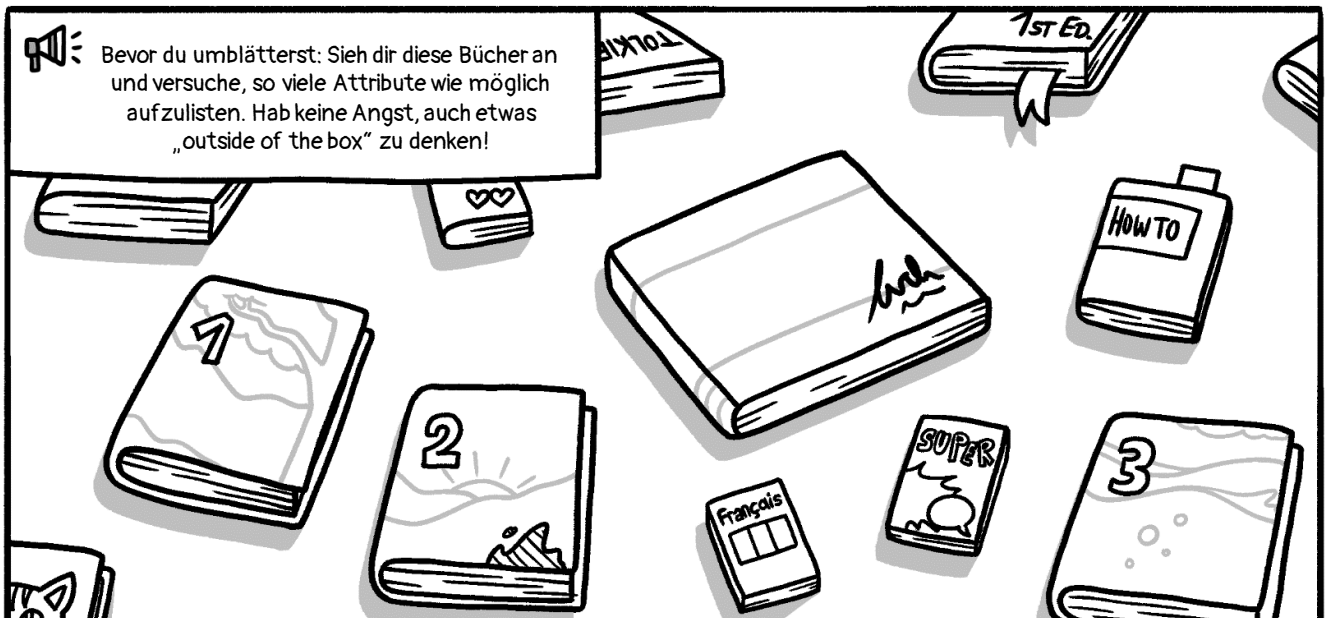
victor.oliveira@fhstp.ac.at

Illustrationen:

Magdalena Boucher & Alena Ertl



<https://fhstp.github.io/comixplain>



Farbe, Größe, Autor*innen, Material, Teil einer Serie, Jahr, Seiten, Sprache, Soft-/Hardcover, Genre, Gewicht, Ausgabe, signi...

Wow, das sind so viele Attribute... Wie wäre es denn mit der Farbe?

Zusatz

Klar! Wir könnten die Bücher nach Regenbogenfarben sortieren.

Oder alphabetisch nach Farbnamen!

Es sollte nicht so schwer sein, sich hier zu einigen...

Ja. Wieso ordnen wir anders, selbst wenn wir uns auf die Farbe geeinigt haben?

Ich schätze, es liegt an den Attributtypen...

Farbton ist ein **KATEGORISCHES** Attribut, genau wie Sprache, Genre, oder Autor*in. Wir können Elemente zwar nach Kategorien sortieren, aber innerhalb dieser Kategorien gibt es keine fixe Reihenfolge.

Ah, jetzt verstehe ich. Also, was wäre dann mit der Größe der Bücher? Da gibt es auf jeden Fall eine gewisse Reihenfolge, oder?

Genau. Attribute können kategorisch oder geordnet sein. Und geordnete Attribute können wiederum **ORDINAL** oder **QUANTITATIV** sein.

Wenn du die Bücher in Zentimetern abmisst, sind die Werte **QUANTITATIV**, da man sie mathematisch vergleichen kann.

20,5 cm

...ist um 8,4 cm kleiner als...

28,9 cm

Dasselbe gilt für das Gewicht in Gramm, oder die Anzahl der Seiten.

Aber wenn ich die Bücher einfach in Gruppen "Klein", "Mittel" und "Groß" einteile...

...dann wäre die Größe ein **ORDINALES** Attribut. Wir können mit den Werten zwar nicht rechnen, aber es gibt trotzdem eine fixe Ordnung.

Klein

Mittel

Groß

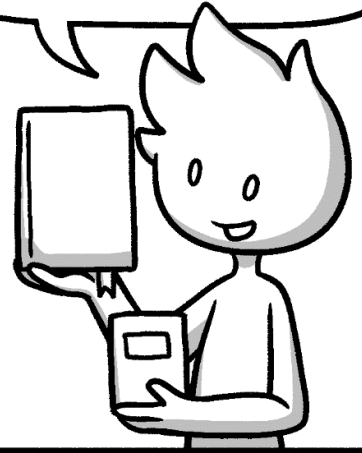
Der Attributtyp einer Variable kann sogar geändert werden, nachdem man die Daten gesammelt hat.



Wie wenn ich alle Bücher in **Zentimetern** messe, und nachher entscheide, dass jedes Buch unter 18 Zentimetern "**Klein**" ist? Dabei würde ich das **quantitative** Attribut in ein **ordinales** umwandeln.



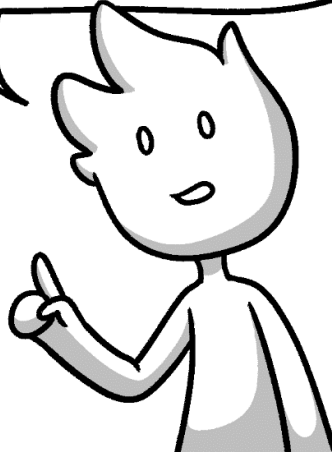
Du hast recht. Du könntest sogar noch weiter gehen und es in ein binäres kategorisches Attribut umwandeln: Taschenbuch oder normale Größe.



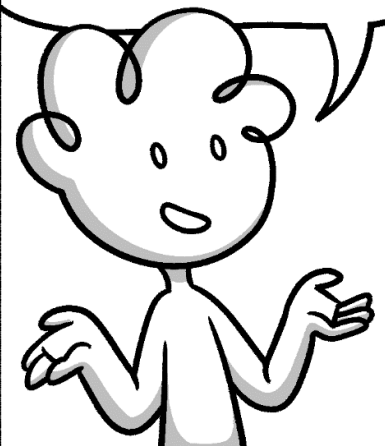
Oh, stimmt. Aber würden wir dabei nicht die meisten Details in den Daten verlieren?



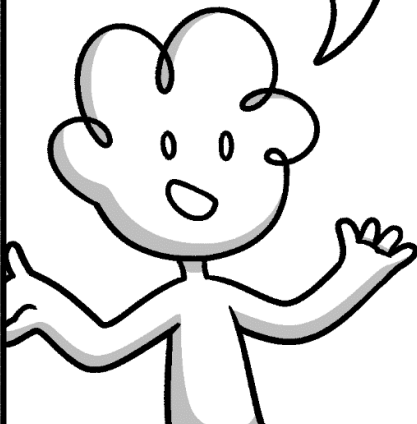
Ja, solche Umwandlungen sind verlustbehaftet. Deshalb ist es besser, Daten so detailliert wie möglich zu erheben, wenn man nicht sicher ist, wie man sie später verarbeitet.



Das ergibt Sinn. Man kann quantitative Werte jederzeit umwandeln, aber wenn man von Anfang an nur kategorische Werte hat, kann man später nicht zu quantitativen wechseln.



Toll, nun da wir alles über Datenattribute und Attributtypen wissen, können wir eine fundierte Entscheidung treffen!



Passt, dann TYPerlasse ich das mal dir und hol mir was zu essen. Die meisten Bücher gehören sowieso dir.



Nun da wir über Datenattribute und ihre Typen gesprochen haben, blättere zur nächsten Seite und versuche, dein Wissen in ein paar Übungen anzuwenden!

Zur Erinnerung: Attribute sind bestimmte Eigenschaften, die gemessen oder beobachtet werden können. Diese Grafik bietet einen Überblick über die verschiedenen Attributarten.

ATTRIBUTE

KATEGORISCH

z.B., Farbe, Genre, Autor*in, Sprache...

GEORDNET

ORDINAL

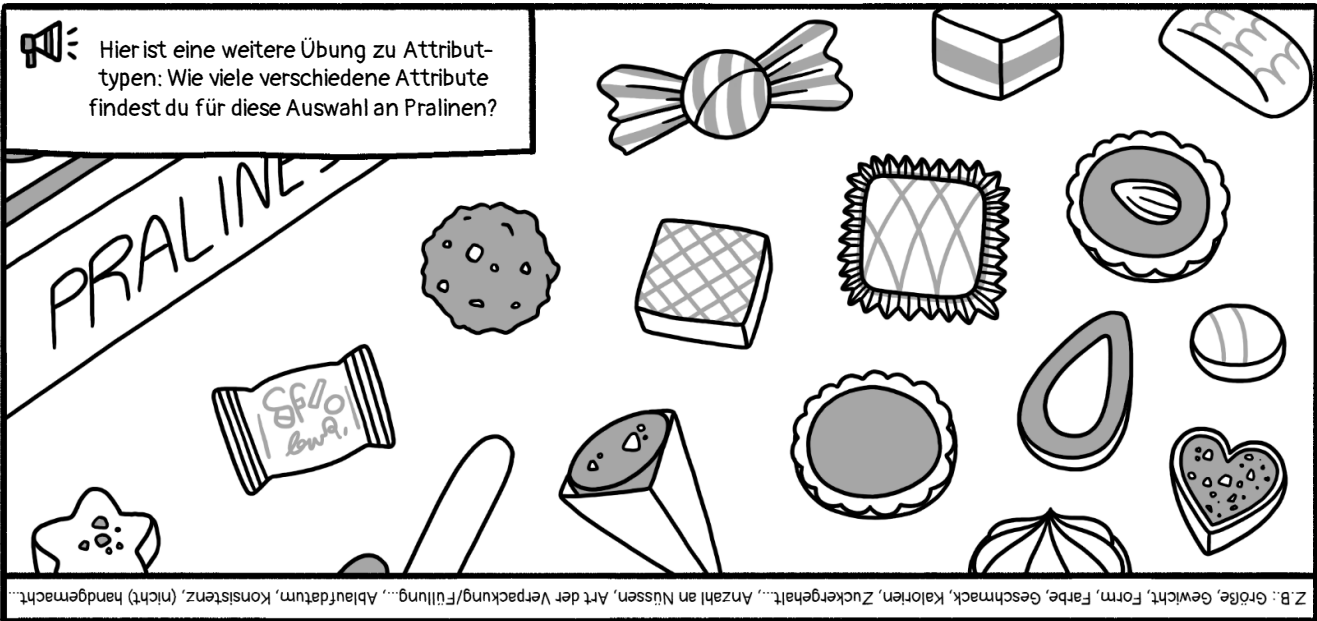
Bildungsgrad, Schuhgröße, Altersgruppe...

QUANTITATIV

Zentimeter, Kilogramm, Alter in Jahren, Temperatur...

Quellen:
- Munzner, T. (2014). Visualization analysis and design. CRC press.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research methods in human-computer interaction. Morgan Kaufmann.

Hier ist eine weitere Übung zu Attributtypen: Wie viele verschiedene Attribute findest du für diese Auswahl an Pralinen?



Z.B.: Größe, Gewicht, Form, Farbe, Geschmack, Kalorien, Zuckergehalt..., Anzahl an Nüssen, Art der Verpackung/Füllung..., Ablaufdatum, Konsistenz, (nicht) handgemacht...

Meistens sind Daten in Tabellenform aufbereitet. Kannst du die Attributtypen jeder Spalte erkennen?

Titel	Regisseur*in	Schauspieler*in	Länge	Jahr	Popularität	Genre
Goldfinger	Hamilton	Connery	112	1964	7.7	Action
Ben Hur	Wyler	Heston	212	1959	8.2	Action
Ben Hur	Niblo	Novarro	133	1926	7.4	Drama
Gladiator	Scott	Crowe	155	2000	8.5	History
Casablanca	Curtiz	Bogart	102	1942	9.5	Romance

Kategorisch: Titel, Regisseur*in, Schauspieler*in, Genre | Quantitativ: Jahr, Länge, Popularität

4

Quellen:

Munzner, T. (2014). Visualization analysis and design. CRC press.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research methods in human-computer interaction. Morgan Kaufmann.