

Clusteranalyse - Einführung

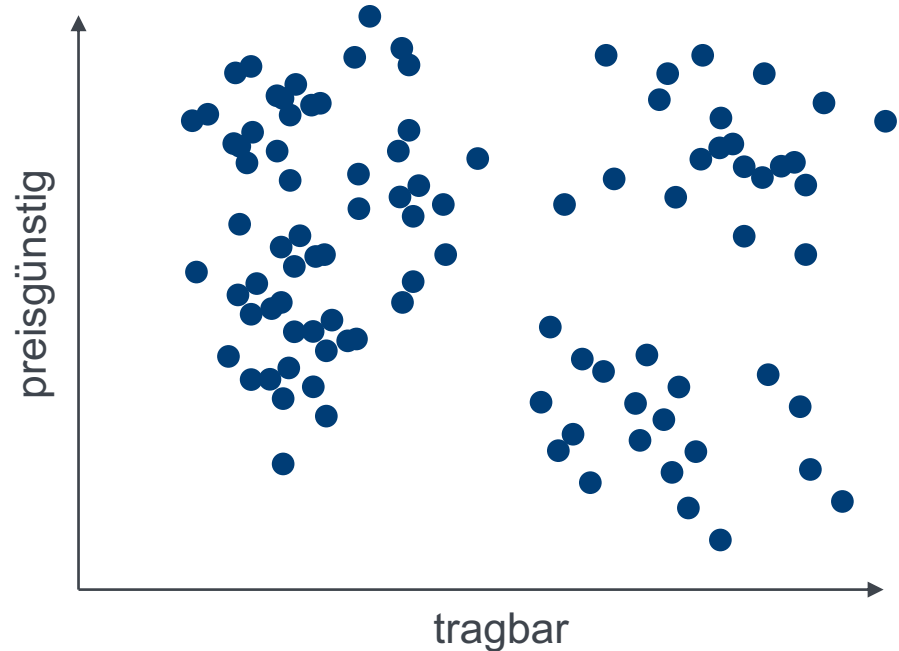
Ordnung ins Chaos bringen

Beispiel: Kundenbefragung

Befragung eines Herstellers von
Autokindersitzen

- Wie wichtig ist Ihnen, dass der Sitz tragbar ist?
- Wie wichtig ist Ihnen, dass der Sitz preisgünstig ist?

**Ziel: Produkte an
Kundentypen anpassen**

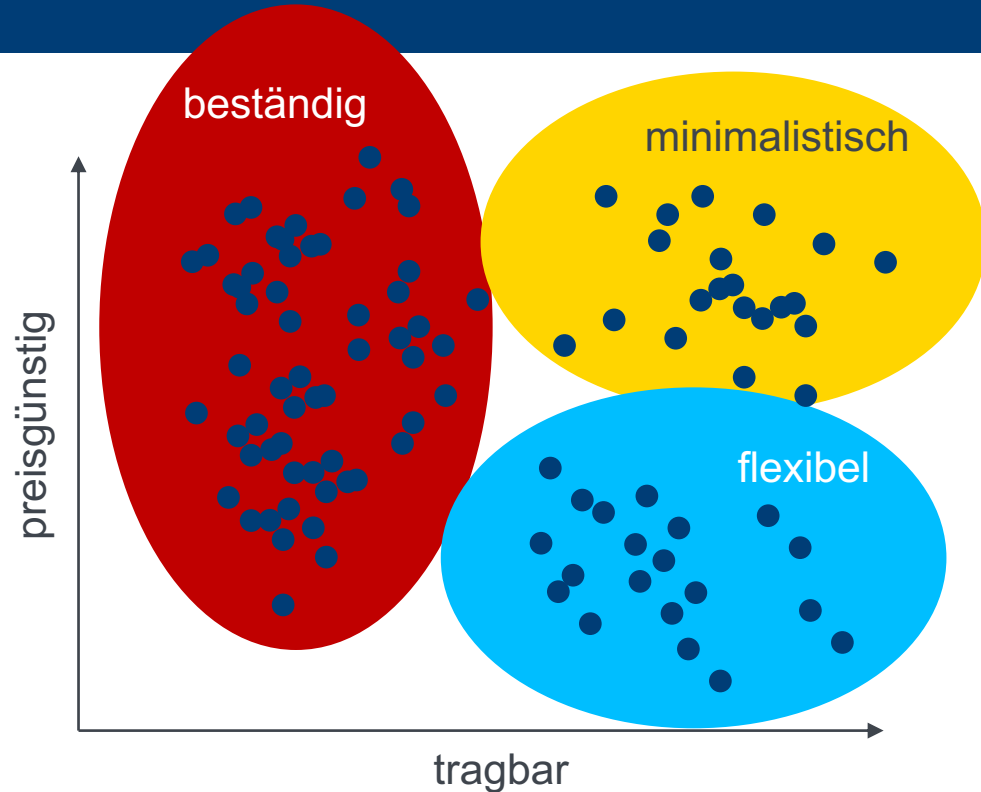


Beispiel: Kundenbefragung

Befragung eines Herstellers von Autokindersitzen

- Wie wichtig ist Ihnen, dass der Sitz tragbar ist?
- Wie wichtig ist Ihnen, dass der Sitz preisgünstig ist?

Ziel: Produkte an Kundentypen anpassen

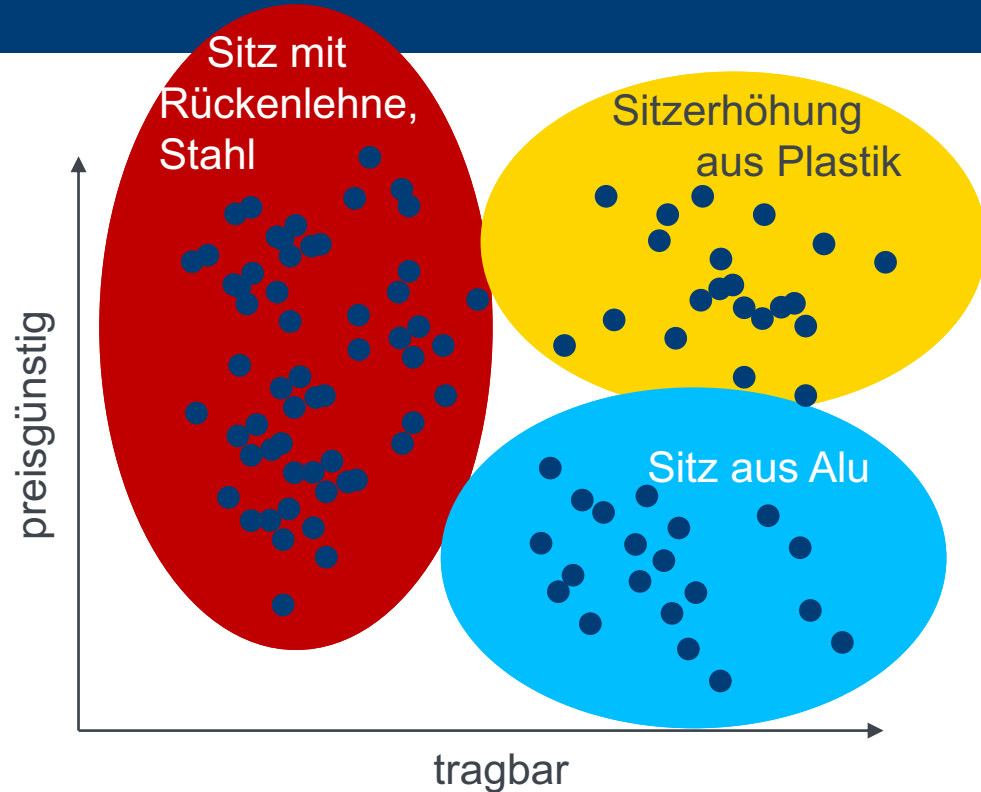


Beispiel: Kundenbefragung

Befragung eines Herstellers von Autokindersitzen

- Wie wichtig ist Ihnen, dass der Sitz tragbar ist?
- Wie wichtig ist Ihnen, dass der Sitz preisgünstig ist?

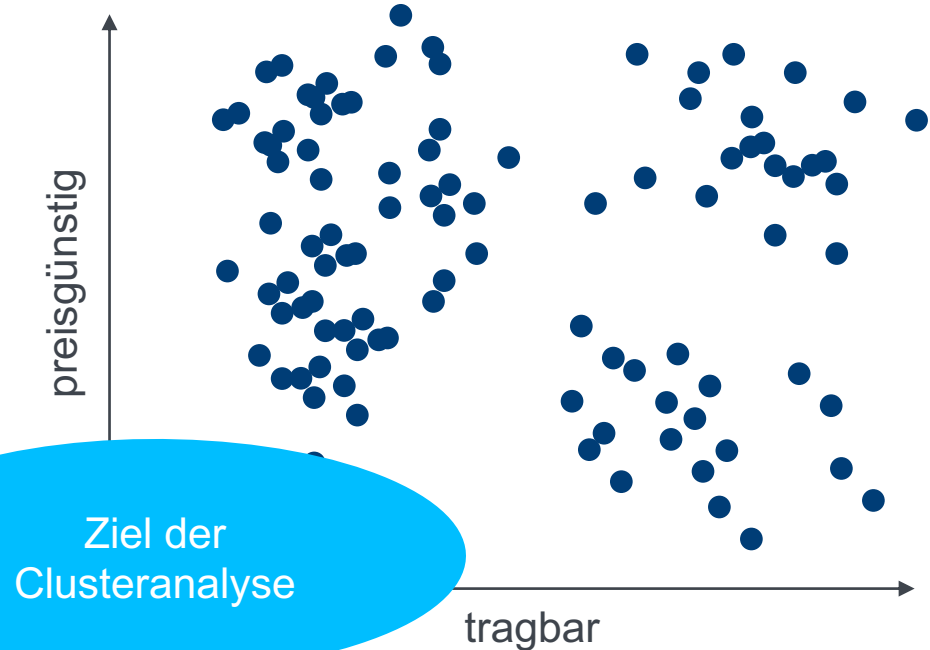
Ziel: Produkte an Kundentypen anpassen



Clusteranalyse

- Gruppen von ähnlichen Datenpunkten automatisch finden
- Meist in vielen Dimensionen

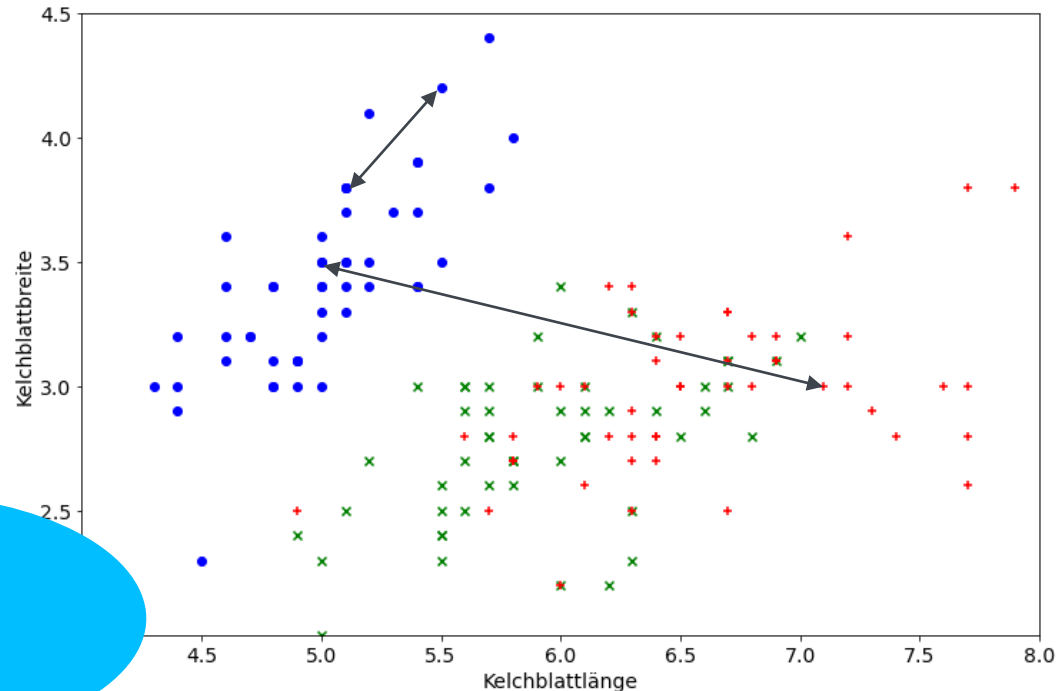
- Alle Punkte in einem Cluster möglichst ähnlich
- Punkte aus verschiedenen Clustern möglichst verschieden



Rückblick: Distanzfunktionen

- Irisdaten
- Nahe beieinander liegende Datenpunkte repräsentieren ähnliche Blattmaße
- Wie misst man die Abstände?

Distanz-
funktionen

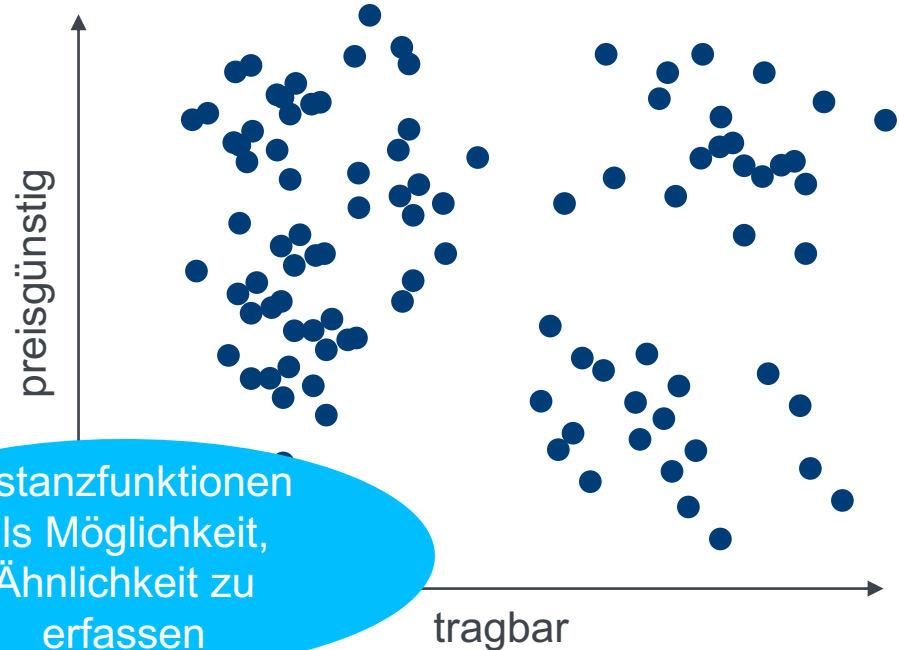


Clusteranalyse

- Gruppen von ähnlichen Datenpunkten automatisch finden
- Meist in vielen Dimensionen

- Abstände innerhalb der Cluster minimieren
- Abstände zwischen Clustern maximieren

Distanzfunktionen
als Möglichkeit,
Ähnlichkeit zu
erfassen



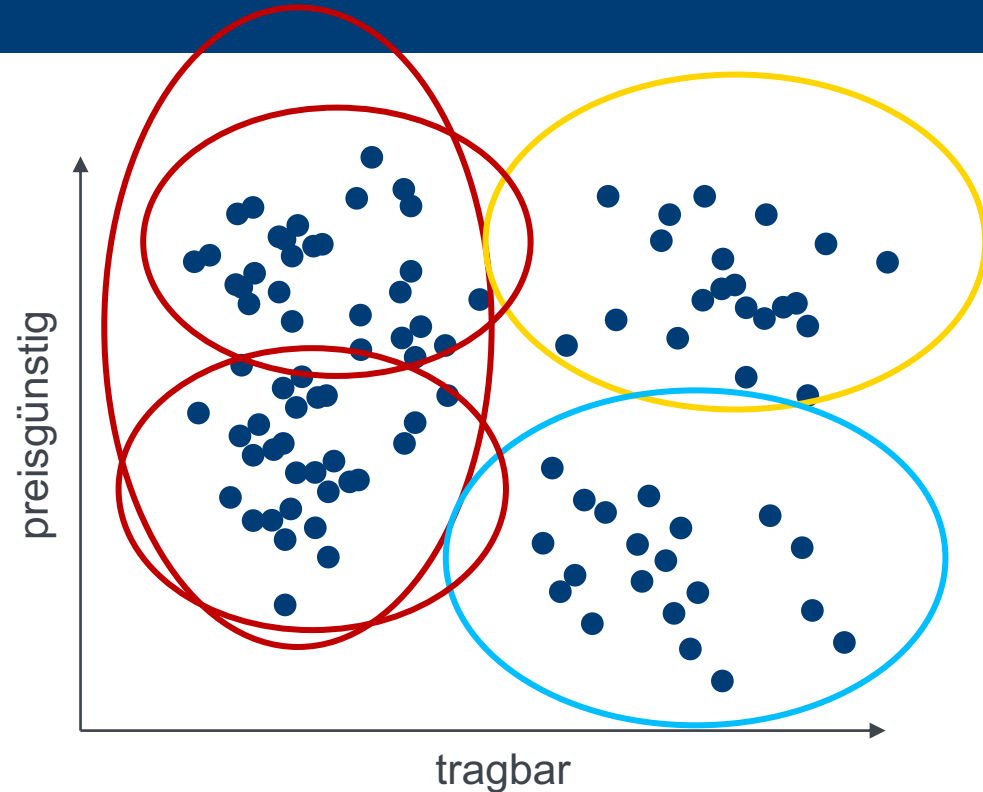
**Ziel der Clusteranalyse ist es,
Gruppen von ähnlichen
Datenpunkten zu identifizieren.**

**Häufig wird dabei Ähnlichkeit
mithilfe von Distanzmaßen
gemessen.**

Beispiel Kundenbefragung

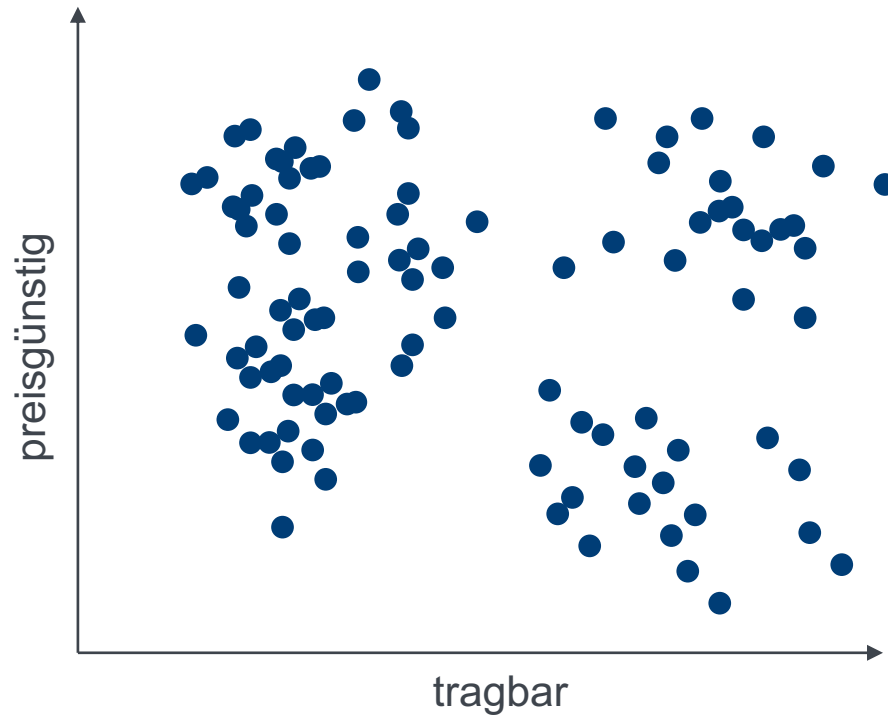
- Die korrekten Cluster sind vorher nicht bekannt
- Nicht einmal die korrekte Anzahl von Clustern

Unüberwachtes
Lernen: kein
korrektes Ergebnis
vorgegeben

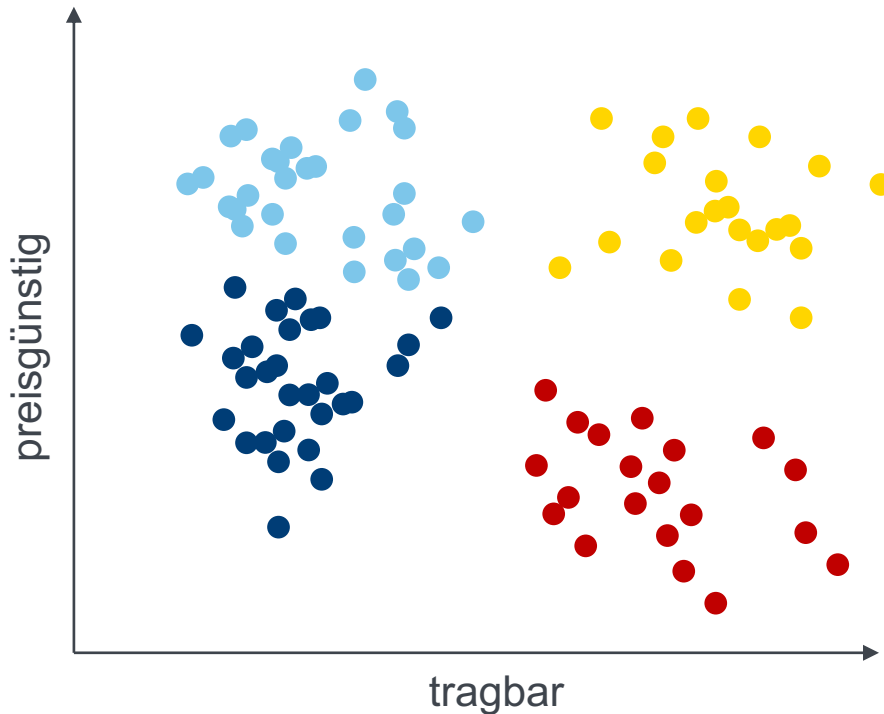


**Clusteranalyse gehört zu den
unüberwachten Lernverfahren: es
gibt kein „korrektes“ Ergebnis.**

Bedeutung von weiteren Dimensionen

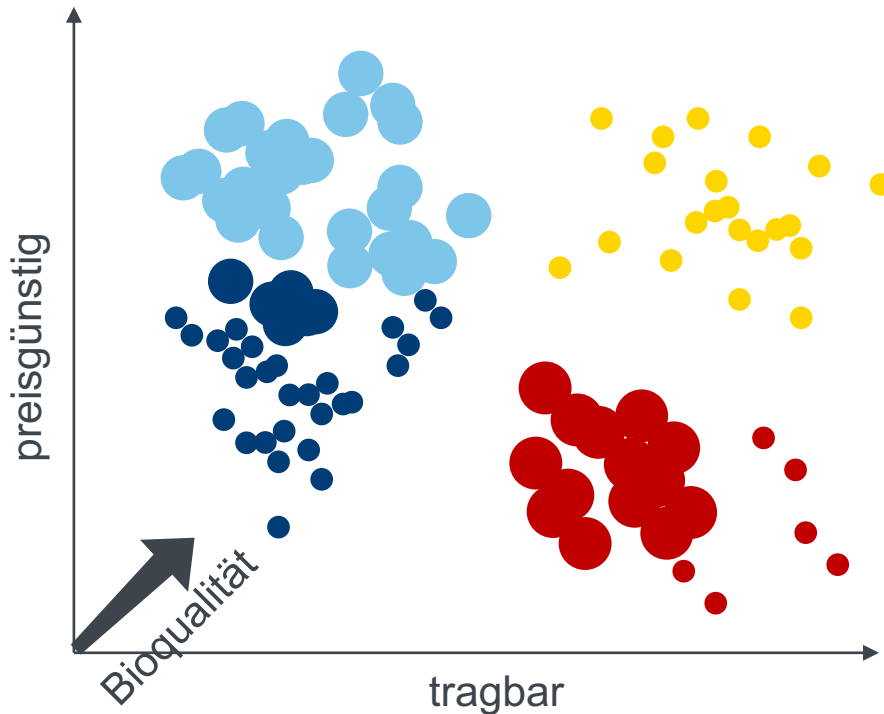


Bedeutung von weiteren Dimensionen



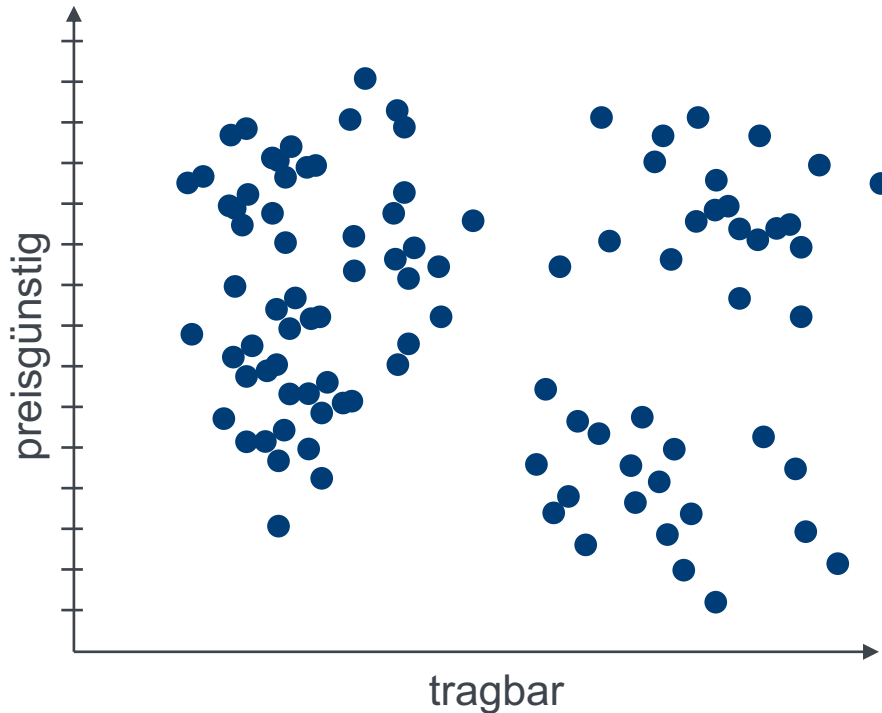
- Wenn weitere Dimensionen hinzukommen, ändert sich häufig die Aufteilung in Cluster

Bedeutung von weiteren Dimensionen



- Wenn weitere Dimensionen hinzukommen, ändert sich häufig die Aufteilung in Cluster

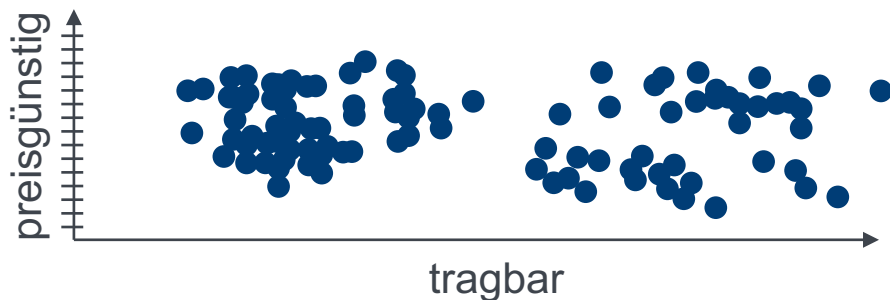
Bedeutung der Skalierung der Dimensionen



- Wenn weitere Dimensionen hinzukommen, ändert sich häufig die Aufteilung in Cluster

Bedeutung der Skalierung der Dimensionen

- Wenn weitere Dimensionen hinzukommen, ändert sich häufig die Aufteilung in Cluster
- Wenn anders skaliert wird, ändert sich häufig die Aufteilung in Cluster



**Die verwendeten Dimensionen und
deren Skalierung haben
(bei den meisten Verfahren)
Einfluss auf das Resultat der
Clusteranalyse.**

Dr. Antje Schweitzer

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung



Universität Stuttgart



Industrie- und Handelskammer
Reutlingen

Reutlingen | Tübingen | Zollernalb



Region Stuttgart



Industrie- und Handelskammer
Karlsruhe



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Lizenzbestimmungen

“Clusteranalyse – Einführung” von Antje Schweitzer, KI B³ / Uni Stuttgart

Das Werk - mit Ausnahme der folgenden Elemente:

- Logos der Verbundpartner und des Förderprogramms
- im Quellenverzeichnis aufgeführte Medien

ist lizenziert unter:

  [CC BY 4.0 \(https://creativecommons.org/licenses/by/4.0/deed.de\)](https://creativecommons.org/licenses/by/4.0/deed.de)

(Namensnennung 4.0 International)

Quellenverzeichnis

Titelfoto: [Rick Mason \(https://unsplash.com/de/@egnar0\)](https://unsplash.com/de/@egnar0), verschiedenfarbiges Plastikspielzeug, auf [Unsplash \(https://unsplash.com/de/fotos/verschiedenfarbiges-plastikspielzeug-2FaCKyEEtis\)](https://unsplash.com/de/fotos/verschiedenfarbiges-plastikspielzeug-2FaCKyEEtis), lizenziert unter [Unsplash-Lizenz \(https://unsplash.com/license\)](https://unsplash.com/license). Bildausschnitt verändert.