

# Mehrdimensionale Daten visualisieren

## Übersicht

- **Autorin** Anna Lena Diehl ([Anna.Diehl@stud.uni-heidelberg.de](mailto:Anna.Diehl@stud.uni-heidelberg.de))
- **Lektorat** Daniel Wunderlich ([wunderlich@informatik.uni-heidelberg.de](mailto:wunderlich@informatik.uni-heidelberg.de))
- **Klassenstufen** Der Unterricht wurde für die Klassenstufe 10 im baden-württembergischen Profilfach IMP (Informatik, Mathematik, Physik) konzipiert, kann jedoch auch später durchgeführt werden. Zu beachten sind die fachlichen Voraussetzungen (s. u.).
- **Dauer** 90 Minuten
- **Lizenz** Das Material unterliegt der [Creative-Commons-By-Non-Commercial-Share-Alike-Lizenz 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) (Bearbeiten und Teilen unter folgenden Bedingungen: Namensnennung, nicht kommerziell, Weitergabe unter gleichen Bedingungen)
- **Zusammenfassung** Ein- und zweidimensionale Daten lassen sich häufig sehr gut visualisieren. So können z. B. Datentupel bestehend aus dem Alter und der Körpergröße befragter Personen einfach als Säulen- oder Balkendiagramm oder als Punkte in einem Koordinatensystem, dessen x-Achse dem Alter und dessen y-Achse der Körpergröße entspricht, dargestellt werden.

Deutlich anspruchsvoller wird das Unterfangen, Datentupel mit mehr als zwei Dimensionen zu visualisieren, da man hier mehr Achsen bräuchte und schnell an die Grenzen des „einfach machbaren“ stößt.

In der vorliegenden Doppelstunde werden die Schüler:innen für diese Problematik sensibilisiert und lernen eine Möglichkeit kennen, ihr zu begegnen. Hierzu wird ihnen das Konzept von Dimension Reduction Algorithmen vorgestellt. Anschließend implementieren sie einzelne Bestandteile eines entsprechenden Programms in Processing, wobei Aspekte im Umgang mit mehrdimensionalen Arrays wiederholt und gefestigt werden.

- **Lernziele** Die angestrebten fachlichen Lernziele der Doppelstunde sind:
  - Die Schüler:innen erklären, was man unter mehrdimensionalen Daten versteht.
  - Die Schüler:innen erklären, weshalb die Visualisierung solcher Daten schwierig, aber auch wichtig ist.
  - Die Schüler:innen können erklären, warum Dimensional Reduction Algorithmen zur Darstellung von Daten wichtig sind.
  - Die Schüler:innen können den groben Aufbau solcher Algorithmen beschreiben und auch die wichtigsten Schwierigkeiten bei der Umsetzung nennen.
  - Die Schüler:innen können mehrdimensionale Daten in Processing als Arrays von Arrays implementieren und diese nutzen, um einen Dimensional Reduction Algorithmus (teilweise) zu implementieren.
- **Voraussetzungen** In der ersten Phase des Unterrichts werden keine konkreten fachlichen Fähigkeiten und Kompetenzen vorausgesetzt. Die Inhalte basieren auf grundlegenden Aspekten des Mathematikunterrichts in Sekundarstufe I (Unterstufe).

Die ändert sich in der zweiten Phase des Unterrichts: Die Aufgabe zur Implementierung setzt grundlegende Fähigkeiten im Programmieren in [Processing](#) voraus. Der Umgang mit Variablen und Kontrollstrukturen (Verzweigung, Zählschleife/for-Schleife) wird ebenso vorausgesetzt, wie der Einsatz von Funktionen (Unterprogrammen/Methoden) und mehrdimensionalen Arrays.

Darüber hinaus beruht die Abstandsberechnung von bis zu dreidimensionalen Datentupeln auf dem Satz des Pythagoras im Raum.

Im baden-württembergischen Bildungsplan bietet sich eine Durchführung am Ende von Klassenstufe 10 an, da dann alle benötigten inhaltsbezogenen Kompetenzen durch den [Bildungsplan 2016](#) abgedeckt sein sollten.

- **Material** Das Material für den Unterricht umfasst folgende Dateien:

Übersicht (dieses Dokument)	Mehrdimensionale-Daten-Visualisieren-Übersicht.odt/.pdf
Tabellarischer Verlaufsplan	Mehrdimensionale-Daten-Visualisieren-Verlaufsplan.odt/.pdf
Datentupel zum Druck auf weißes Papier	Mehrdimensionale-Daten-Visualisieren-Datentupel-weißes-Papier.odt/.pdf
Datentupel zum Druck auf farbiges Papier	Mehrdimensionale-Daten-Visualisieren-Datentupel-farbiges-Papier.odt/.pdf
Präsentationsfolien	Mehrdimensionale-Daten-Visualisieren-Präsentation.pptx/.pdf
Arbeitsblatt	Mehrdimensionale-Daten-Visualisieren-Arbeitsblatt.docx/.pdf
Hinweiskärtchen	Mehrdimensionale-Daten-Visualisieren-Hinweiskärtchen.odt/.pdf
Vorlage und Musterlösung der Programmieraufgabe	code/visualisierung_aufgabe code/visualisierung_loesung

- **Stundenverlauf**

- **Phase 1:**

- Einstieg in das Thema und gleichzeitig die erste Erarbeitungsphase soll ein Spiel sein, in dem die Schüler:innen selbst die Aufgabe eines Dimension Reduction Algorithmus' (DR-Algorithmus) übernehmen. Nach einer kurzen Einführung durch die Lehrkraft bekommen die Schüler:innen „Datentupel“ in Form von Punkten zugeteilt – Papierkreise, auf denen die Werte dreier Attribute gedruckt sind. Im Folgenden legen die Schüler:innen ihre Punkte nacheinander so auf dem Boden ab, dass Punkte mit ähnlichen Attributwerten möglichst nah beisammen liegen und Punkte mit unterschiedlichen Attributwerten großen Abstand zueinander haben.
- Haben alle Schüler:innen ihren Datenpunkt abgelegt, folgt eine kurze Phase des Überprüfens, in der die Schüler:innen gemeinsam überlegen sollen, wie zufrieden sie mit ihrer Verteilung sind. Anschließend dürfen sie – wieder nacheinander – ihre Datenpunkte verschieben, um die Verteilung zu optimieren. Somit ist die Grundstruktur von DR-Algorithmen bereits in der ersten Einstiegsphase für die Schüler:innen auf einer sehr niederschweligen Ebene erlebbar.
- Durchgeführt wird dieses Spiel z. B. auf dem Flur vor dem Klassensaal. Die Schüler:innen nehmen dazu ihre Punkte und ein Handy oder ein vergleichbares Gerät pro Gruppe mit, um die Ergebnisse dokumentieren zu können.
- Die Klasse wird für dieses Spiel in drei Gruppen eingeteilt, jede Gruppe erhält dasselbe Set von Datentupeln. Dies hat den Vorteil, dass die verschiedenen Vorgehensweisen und Ergebnisse im Anschluss im Plenum verglichen und diskutiert werden können. Beim Vergleichen der Ergebnisse wird vor allem deutlich, dass es keine eindeutige und optimale Lösung für das Problem gibt. Gleichzeitig werden bereits Qualitätsmerkmale für Verteilungen diskutiert, welche in der zweiten Phase des Unterrichts eine Rolle spielen werden.
- Die Gruppeneinteilung wird durch die Lehrkraft vorgegeben und durch Symbole auf der Rückseite der Datenpunkte kommuniziert.
- Nach Beendigung des Spiel kehren die Schüler:innen in den Klassensaal zurück.
- Die anschließende Diskussion im Plenum dient vor allem dazu, das zuvor Erlebte einzuordnen und zu reflektieren. Dazu laden die Gruppen ihre Fotos auf der Schulcloud hoch und die Lehrkraft kopiert die Bilder auf eine Powerpointfolie, sodass die Ergebnisse aller Gruppen gleichzeitig über den Beamer sichtbar sind. Um das Vergleichen der Lösungen zu vereinfachen, haben die Datenpunkte innerhalb eines Sets unterschiedliche Farben, setübergreifend ist die Farbverteilung gleich. So kann man auf den ersten Blick Unterschiede in den einzelnen Lösungen erkennen.

○ **Phase 2:**

- Die zweite Unterrichtsphase beginnt mit einem Input durch die Lehrkraft, in der die bereits im Spiel erfahrenen Konzepte anhand einer Präsentation mit Folien noch einmal genauer definiert und mit festen Begriffen versehen werden. Außerdem wird thematisiert, in welchen Bereichen Dimension Reduction Algorithmen eingesetzt werden. Beispielshaft wird hier kurz der Projection Space Explorer demonstriert, der die DR-Algorithmen UMAP, t-SNE und ForceAtlas2 verwendet, um mehrdimensionale Daten als Scatterplot zu visualisieren.
- In der zweiten großen Erarbeitungsphase soll die Lerngruppe selbst Teile eines vereinfachten DR-Algorithmus implementieren. Dafür wird den Schüler:innen eine Programm-Vorlage zur Verfügung gestellt, die sie mithilfe eines Arbeitsblattes vervollständigen sollen. Hierzu werden für viele Aufgabe Hilfskärtchen zur Verfügung gestellt, um auch schwächere Schüler:innen bei der Lösungsfindung zu unterstützen.
- Nachdem die Schüler:innen den DR-Algorithmus umgesetzt haben, sollen sie ihre Verteilung aus der ersten Phase mithilfe des Programms evaluieren und bewerten, wie zufrieden sie mit der Evaluation sind.
- Leistungsstärkere Schüler:innen können sich darauf aufbauend mit der Thematik beschäftigen, wie Fragen mit (vorerst) nicht ohne Weiteres numerisch diskret quantifiziert werden können („Was ist deine Lieblingsfarbe?“) behandelt werden können.
- Abschließend stellen einzelne Schüler:innen ihre Ergebnisse im Plenum vor. Auch hier können wieder Lösungen verglichen und diskutiert werden.

- **Rechtliche Hinweise zu den Quellen** Alle verwendeten Grafiken sind eigenständig erstellt.

Sollte trotz sorgfältiger Prüfung ein urheberrechtliches Vergehen vorliegen, bitten wir um eine Meldung an [Anna.Diehl@stud.uni-heidelberg.de](mailto:Anna.Diehl@stud.uni-heidelberg.de) oder [wunderlich@informatik.uni-heidelberg.de](mailto:wunderlich@informatik.uni-heidelberg.de) mit Angabe der Unterrichtseinheit und des betroffenen Bildes bzw. der betroffenen Stelle. Wir werden dies dann umgehend beheben!