

[zh.ch](#)

# Datenwerkstatt für Schulen

24–29 Minuten

---

- [Startseite](#)
- [Navigation](#)
- [Inhalt](#)
- [Kontakt](#)
- [Inhaltsverzeichnis](#)
- [Suche](#)
- [Fussbereich](#)

## Datenwerkstatt für Schulen

**Auf dieser Seite finden Lehrerinnen und Lehrer einen Einstieg ins Thema Daten. Dieser hilft den Lehrpersonen dabei, im Unterricht Data und Statistical Literacy zu vermitteln - eine Schlüsselkompetenz des 21. Jahrhunderts.**

## Inhaltsverzeichnis

## Grundlagen und Übungen

Den Themen Daten und Statistik nähert man sich am besten über einen konkreten Anwendungsfall. Dieser ermöglicht es, Daten selbst zu generieren, zu verstehen, bearbeiten und auszuwerten. Die Inhalte dieser Seite stehen in engem Zusammenhang mit dem Projekt «CO2 vs. CO19». Die Grundlagen und Übungen sind aber auch unabhängig davon nützlich, wenn Sie einen Einstieg in die Themen Daten und Statistik suchen.

## Was sind Daten?

Daten sind der maschinell verarbeitbare Teil einer Information. Damit aus einer Information Daten werden, muss sie in kleinere Bestandteile zersetzt werden. Daten sind objektiv und brauchen keinen Bezug. Dazu werden die Daten, die im Zusammenhang eine Information ergeben, aus ihrem Kontext genommen. Umgekehrt erhalten wir aus Daten erst eine Information, wenn sie in einen Kontext gesetzt werden.

Ein Beispiel: Wenn Ihnen Ihre Nachbarin die Information «Morgen scheint die Sonne» gibt, setzen Sie diese automatisch in einen Kontext (Ort, Zeit) und verstehen sie. Um Daten jedoch objektiv und maschinell nutzbar zu machen, müsste diesem Wetterdatum der Kontext in weiteren Daten zur Seite gestellt werden: Ort und Zeit. Menschen verbinden diese drei Datenpunkte automatisch wieder zu einer Information. Für die maschinelle Verwendung der Daten hingegen ist die Unterteilung von Information in einzelne, objektive und voneinander trennbare Datenpunkte nötig. Der Zusammenhang von Daten, Information und Wissen wird meistens hierarchisch in einer [Pyramide](#) dargestellt.

In einer analogen Welt müssen Daten meistens aus Information heruntergebrochen werden. Digitale Dienste basieren hingegen grundlegend auf Daten und generieren mit jeder Nutzung mehr davon. Wenn diese Daten zueinander in Kontext gesetzt werden, erhalten wir daraus neue Informationen. Damit diese Informationen zur weiteren Verwendung geeignet sind, müssen also sowohl die Daten als auch die vorgenommene Kontextualisierung richtig sein. Letzteres lässt sich beurteilen, indem wir die Information mit anderer Information abgleichen. Doch wie beurteilt man Datenqualität?

## Übungen

Benötigte Materialien: keine

Um die Wissenspyramide zu erklären, könnte man als Übung annehmen, dass jemand in der Klasse ein Geburtstagsfest organisieren möchte. Wir kommen praktisch von Daten zur Handlung:

**Daten:**

eine Liste mit Namen aller Leute, die ich kenne. Darunter Kinder, Nachbarn, Grosseltern usw. Aber wen lade ich zu meinem Geburtstagsfest ein?

**Information:**

Eine Liste mit dem Namen aller Leute, die ich einladen möchte. Zum Fest möchte ich beispielsweise nur Menschen einladen, die Kinder sind und ich mag. Indem ich ein Geburtstagsfest plane, werden die Daten in einen Kontext gesetzt.

**Wissen:**

Ich überlege mir, welche der Spiele, die ich besitze, den Gästen und mir am meisten Spass machen und sich für dieses Geburtstagsfest am besten eignen.

**Weisheit:**

Ich schätze das benötigte Material in der richtigen Menge ein, um die Spiele an meinem Geburtstagsfest spielen zu können und kaufe es ein (Handlung).

Benötigte Materialien: CO2-Ampel

Mit den Schülern und Schülerinnen kann die Wissenspyramide anhand der CO2-Ampel besprochen werden.

**Daten:**

Das Gerät misst Werte. Obwohl die Ampel uns in Farben sagt, wie es mit dem CO<sub>2</sub> im Raum steht, so misst sie in Wirklichkeit reine Werte: 300, 500, 700, 1000. Das sind Daten, mit denen wir ohne Kontext aber nichts anfangen können.

**Information:**

Was bedeuten diese Werte? Diese Zahlen stehen für Mikrogramm CO<sub>2</sub> pro Kilogramm Luft-Volumen. Das ist die Bedeutung dieser Rohdaten und so erschliesst sich uns ein Teil der Information.

**Wissen:**

Doch auch hier fehlt noch der Vergleich: was bedeutet nun 300, 1000 oder 2000 ppm (part per million) CO<sub>2</sub> in der Luft? Wenn man draussen ist, dann liegt der Anteil bei ca. 400 ppm (Anmerkung für Lehrpersonen: parts per Million ist zwar eine verwirrende aber die offizielle Bezeichnung dieser Grösse). Dieser kann zwar unterschiedlich sein, je nachdem, ob man beispielsweise in einem Stadtzentrum oder in der Peripherie ist, der Wert 400ppm ist aber ein brauchbarer Durchschnitt.

**Weisheit:**

Nun hat das BAG auch Grenzwerte für die Innenluft angegeben. Man sollte versuchen den Wert wenn möglich unter 2000 ppm zu behalten (maximal 10% der Unterrichtszeit darüber) und generell möglichst unter 1400 ppm, da ansonsten die Wahrscheinlichkeit steigt müde zu werden, da der CO<sub>2</sub>-Anteil in der Luft zu hoch ist. Gleichzeitig sagt das BAG, dass diese höheren Werte auch zu einer schnelleren Virenübertragung führen. Wenn die Ampel nun Gelb (über 1400 ppm) oder Rot (über 2000 ppm) leuchtet, können wir somit schlussfolgern, dass wir lüften sollten.

Benötigte Materialien: Sensordaten

- [Arbeitsdatensatz mit beispielhaften Sensordaten \(kann anstelle des eigenen CO2-Datensatzes genutzt werden\):](#)

Anhand der Sensordaten kann der Aufbau eines Datensatzes besprochen werden. Beispielsweise können die Bedeutung der Spalten und Zeilen diskutiert werden sowie unterschiedliche Skalenniveaus im Datensatz veranschaulicht werden.

Skalenniveaus werden weiter unten noch thematisiert.

Beispielfragen zum Datensatz allgemein:

- Wie viele Spalten hat der Datensatz?
- Wie viele Zeilen hat der Datensatz?
- Wie sind die Spalten im Datensatz benannt?
- Warum haben die Spalten einen Namen und die Zeilen nicht?
- In welchen Zeitabständen wurden die Datenpunkte erfasst?
- Sind die Zeitabstände immer dieselben?

Beispielfragen zu den Spalten (= Variablen) im Datensatz:

- Was bedeuten die Variablennamen?
- Welche Variablen sind teilbar und welche nicht, sprich mit welchen Variablen kann man Berechnungen durchführen?
- In der Spalte `sensor` steht immer dasselbe. Warum ist das so?
- Was sind die Skalenniveaus der Variablen?

### **Was sind gute Daten?**

Die grosse Menge verfügbarer Daten («Big Data») spielt in vielen

Anwendungen, gerade im Bereich der künstlichen Intelligenz, eine wichtige Rolle. Jedoch sind Daten auch in grosser Menge nichts wert, wenn ihre Qualität nicht gut ist. Man spricht hier auch vom «garbage in, garbage out»-Prinzip und meint damit: auf einer qualitativ schlechten Datengrundlage lässt sich keine gute Anwendung aufbauen.

In der Realität lässt die Datenqualität leider oft zu Wünschen übrig. Die Qualität eines Datensatzes wird durch technische, administrative und ethische Kriterien bestimmt. Die wichtigsten davon sind:

- Einheitlichkeit: Die Daten sind strukturiert (z.B. in Tabellenform), die Struktur und ihre Benennung ist einheitlich über alle Datenpunkte und im Zeitverlauf
- Zuverlässigkeit: Es ist nachvollziehbar, wie die Daten entstanden sind
- Eindeutigkeit: Der Datensatz und seine Variablen sind eindeutig und verständlich benannt
- Vollständigkeit: Alle für die Anwendung benötigten Datenpunkte sind durchgängig und in guter Genauigkeit vorhanden
- Die Daten sind aktuell und werden nachgeführt
- Korrekt (Data Veracity): Die Daten sind zutreffend und wurden nicht verfälscht.

Wenn die Datenqualität schlecht ist, kann aus diesen Daten keine Information gezogen werden. Es ist dennoch ein häufiger Fehler anzunehmen, dass eine hohe Quantität genüge. Für Big Data gibt es deswegen die 5-V-Kriterien (je nach Zählweise auch 4 oder 7 V), welche berücksichtigt werden müssen. Insbesondere wenn es um die Beurteilung der Data Veracity (Datenwahrhaftigkeit, Vertrauenswürdigkeit) geht, ist die Datenherkunft wichtig. Hier überschneidet sich die Frage der Datenqualität mit der Diskussion um Fake News und Deepfakes einerseits, und um jene um Datenschutz andererseits.

## Übungen

Benötigte Materialien: keine

Als Grundregel sollte man sich im Umgang mit Daten und Informationen stets folgende Fragen stellen: Woher stammt die Information? Wer hat diese Information generiert und wieso? Mit diesen Fragen, kann man eine Datenquelle gut einordnen.

Es gibt viele pädagogische Beispiele zu fake news und generell zum Thema Datenquellen im Internet. Einige finden Sie in den Links.

Benötigte Materialien: keine

Die Schüler und Schülerinnen geben an, wie viele Geschwister sie haben. Die Lehrperson verkündet dass sie 25 Geschwister hat. Vermutlich werden die Schüler und Schülerinnen ungläubig reagieren. Anhand dieses Beispiels kann die Wichtigkeit von Datenplausibilisierung diskutiert werden: was hat die Schüler und Schülerinnen dazu verleitet, ungläubig zu reagieren?

Wie könnte so ein Fehler im Datensatz entstehen?

Ein anderes Beispiel zur Plausibilisierung: Aus einer Tabelle lesen wir, dass jemand 1.78 kg schwer ist und 58 cm gross. In diesem Fall ist ein Fehler passiert, der sehr häufig passiert: Variablen (das Gewicht und die Grösse) wurden vertauscht.

Benötigte Materialien: keine, allenfalls eine CO2-Ampel

Es gibt verschiedene Störfaktoren, die dazu führen können, dass der CO2-Gehalt zu hoch oder zu tief angezeigt wird. Entsprechend wird die Datenqualität negativ beeinflusst. Beispielsweise sind die Messwerte laut Gerätehersteller ungenauer, wenn das Gerät im Wind steht. Gemeinsam mit den Schüler und Schülerinnen können solche Störfaktoren, die die Datenqualität negativ beeinflussen, besprochen werden. Zum Beispiel kann Folgendes ausprobiert und

diskutiert werden:

- Was passiert, wenn man in das Gerät haucht und warum?
- Wäre die Ampel brauchbar, wenn sie den CO<sub>2</sub>-Gehalt von vor 10 Minuten widerspiegeln würde? Wäre sie noch brauchbar, wenn sie den CO<sub>2</sub>-Gehalt von vor 12 Stunden abbildet? Was ist die maximale zeitliche Verzögerung, die ihr vertretbar findet und warum?

## **Was kann ich aus Daten lesen?**

Wenn wir aus Daten oder einer Datenvisualisierung etwas lesen wollen, möchten wir daraus eine Information ziehen. Um von Daten zur Information zu gelangen, brauchen wir Kontext zu diesen Daten. Dazu müssen wir grundsätzlich verstehen, aus welchen Elementen ein Datensatz besteht: Aus Datenpunkten (die erfassten Fälle oder Objekte) und aus Variablen (die zu den Datenpunkten erfassten Merkmale).

Es gibt mehrere Möglichkeiten, Information aus Daten zu gewinnen: mithilfe von Visualisierungen, Kennwerten aus der deskriptiven Statistik und mithilfe von inferentieller Statistik.

Die deskriptive Statistik beinhaltet Methoden anhand derer ein Datensatz mittels Visualisierungen oder Kennwerten beschrieben und zusammengefasst werden kann. Häufig werden dabei mehrere Kennwerte der Variablen eines Datensatzes in tabellarischer Form dargestellt. Üblicherweise unterscheidet man zwischen sogenannten Lage- und Streuungsmassen. Lagemasse zeigen die zentrale Lage einer Variable an und Streuungsmasse die Unterschiedlichkeit. Haben wir in unserem Datensatz beispielsweise die Variable «Körpergrösse», so geben Lagemasse an, welche Körpergrösse unsere Stichprobe am besten beschreibt. Die Streuungsmasse hingegen geben Aufschluss über die Unterschiedlichkeit der Personen bezüglich ihrer Körpergrösse. Zu den Lagemassen zählen mitunter der Mittelwert, der Median sowie der Modus, und zu den Streuungsmassen die



Standardabweichung, die Varianz und die Spannweite.

Mit inferenzstatistischen Methoden können anhand eines Datensatzes inhaltliche Hypothesen statistisch überprüft werden. Dabei möchte man idealerweise Rückschlüsse von der Stichprobe, von der die Daten stammen, auf die Grundgesamtheit ziehen. Die Themen Stichprobe und Grundgesamtheit werden weiter unten erläutert. Zudem finden Sie mehr Ressourcen in der Linkliste.

Um aus Daten mittels statistischer Methoden Information ziehen zu können, müssen wir daher nicht nur wissen, wie die Datenqualität ist, sondern auch, wie die Daten zueinander und zu anderen Daten stehen: wie gross und aktuell ist der Datensatz? Wie hoch ist der höchste und wie tief der niedrigste Wert (Streuung)? In Datenvisualisierungen lassen sich diese Zusammenhänge oft sehr leicht verständlich darstellen. Visualisierungen stellen Daten in einen Kontext und legen einen Fokus; wenn wir keine zusätzliche Information zum gesamten Datensatz erhalten, können Visualisierungen und andere Vereinfachungen daher sehr leicht in die Irre führen. Die Werbung nennt dann vielleicht «14 Prozent mehr Wimpernwachstum», ohne dass wir wissen, im Verhältnis wozu dieser Zuwachs von 14 Prozent steht oder wie sich Wimpernwachstum definiert: mehr Wimpern? Oder längere Wimpern? Oder wachsen sie einfach schneller und fallen auch schneller wieder aus?

## Übungen

Benötigte Materialien: Sensordaten, Statistikprogramm der Wahl (zB R Studio)

- [R-Skript um die Übungen mit R Studio ausführen zu können:](#)

Anhand des Datensatzes können Lagemasse (z. B. Mittelwert, Median, Modus) und Streuungsmasse (z. B. Standardabweichung, Varianz) berechnet und diskutiert werden. Für welche Variablen im Datensatz ist es sinnvoll, Lage- und Streuungsmasse zu berechnen und für welche nicht?

Benötigte Materialien: Schreibutensilien, Sensordaten, Statistikprogramm (zB R Studio)

- [R-Skript um die Übungen mit R Studio ausführen zu können:](#)

Es wird ein Zeitraum definiert (z. B. zwei Stunden), über den die Schüler und Schülerinnen festhalten, wann im Schulzimmer gelüftet wird.

Im Anschluss formulieren und visualisieren die Schüler und Schülerinnen ihre Erwartungen bezüglich der Entwicklung des CO<sub>2</sub>-Gehalts, der relativen Luftfeuchtigkeit und der Temperatur im Schulzimmer über diesen Zeitraum: Wie hoch waren die Werte vermutlich zu Beginn? Und wie haben sich die Werte verändert? Hierfür kann es auch sinnvoll sein, die Schüler und Schülerinnen selber recherchieren zu lassen, wie hoch der CO<sub>2</sub>-Gehalt, die relative Luftfeuchtigkeit und die Temperatur in Innenräumen typischerweise ist. Die Lüftungszeitpunkte sowie die erwarteten Daten können beispielsweise in einer Grafik mit vordefinierten Achsen ohne Inhalt eingezeichnet werden. Alternativ können die

Schüler und Schülerinnen die gesamte Grafik von Hand aufzeichnen und dabei die x- und y- Achse selber beschriften.

Zuletzt visualisieren die Schüler und Schülerinnen (oder die Lehrpersonen) die erhobenen Daten über den gewählten Zeitraum (mittels Colibird App oder einem Statistikprogramm), um die zuvor erarbeiteten Erwartungen mit den Daten zu vergleichen. Folgende Fragen können dabei diskutiert werden:

- Inwiefern stimmen die Erwartungen der Schüler und Schülerinnen mit den Sensordaten überein?
- Sind die Daten im Hinblick auf die Lüftungszeitpunkte plausibel? Gibt es sogenannte Ausreisser (i.e., Werte, die unerwartet klein oder gross sind)?
- Was ist die Beziehung zwischen dem CO<sub>2</sub>-Gehalt, der Feuchtigkeit und der Temperatur? Steigen respektive sinken die Werte gemeinsam

Es ist absehbar, dass die Vorhersagen der Schüler und Schülerinnen besser werden, wenn sie diese Übung mit neuen Daten wiederholen. Dadurch wird indirekt aufgezeigt, wie wir aus Daten lernen und wie uns bestehende Daten dabei helfen können, genauere Vorhersagen für zukünftige Ereignisse zu machen.

## **Statistik hilft bei Ungewissheit**

Mittels Statistik können wir Entscheidungen datenbasiert treffen. Dabei lassen sich durch statistische Verfahren sogenannte Schätzer ermitteln, inklusive damit verbundene Unsicherheit. Statistik hilft uns fundierte Entscheidungen zu treffen, wenn wir nur wenig Information haben. Um das zu erreichen, gibt die Statistik Annäherungswerte bekannt, sagt aber auch, welche Abweichungen von diesen Schätzwerten wahrscheinlich sind. Damit gibt sie in unsicheren Momenten die bestmöglichen Entscheidungsgrundlagen.

Wenn Daten zur Grundlage wichtiger Entscheidungen werden, wird es umso wichtiger, dass die Unsicherheiten klar benannt werden:

die Statistik weiss es selten ganz genau; aber man kann abschätzen, wie unsicher die Schätzung ist (z.B mittels Konfidenzintervall und Standardfehler).

## Übungen

Benötigte Materialien: keine

Statistik kann uns beispielsweise dabei helfen, eine Grösse, die wir kennen möchten (die «Bevölkerung»), durch eine «Stichprobe» zu ermitteln. Statistik hilft auch einzuschätzen, wie genau wir mit der Stichprobe die Bevölkerung etwa abbilden können.

Beispiel: Wenn man die mittlere Grösse der Fische in einem See kennen möchte, dann muss man nicht alle Fische aus dem See fischen, um sie zu vermessen. Man kann eine Stichprobe zu ziehen, um zu berechnen, wie sicher wir uns sind, dass wir den echten Mittelwert für die «Gesamtfischbevölkerung» berechnen können.

Wir Menschen tendieren dazu, unsere eigenen Erfahrungen zu verallgemeinern. Wenn alle in meiner Familie gern Fussball schauen, dann denke ich, dass alle Familien das so machen bis ich vom Gegenteil überzeugt werde. Wenn das in der Statistik geschieht, besteht die Gefahr der falschen Generalisierung. Deswegen befolgt die Statistik beim Ziehen einer Stichprobe gewisse Regeln: Die Zusammensetzung der Bevölkerung muss bereits in der Stichprobe richtig vertreten sein: man kann z.B. nicht nur Tessiner befragen und dann auf die ganze Schweiz verallgemeinern.

Eine weitere Variante um das Thema Stichprobe und Grundgesamtheit anhand der CO2-Ampel zu diskutieren:

Was sagt die Messung der Luftqualität eines Raumes über die Schule oder über alle Schulen im Land aus? Was sagen 10 Messungen in 10 verschiedenen Schulen aus? Und 100? Vielleicht

ergeben sich Muster beispielsweise bezüglich des Baustils: Sind sich Messungen von Gebäuden mit Minergie-Standards ähnlich? Dabei geht es darum zu zeigen, dass man sich womöglich nicht auf eine einzelne Messung verlassen kann und nicht ohne weiteres auf andere Schulklassen oder Schulhäuser generalisieren kann.

Benötigte Materialien: Keine. Allenfalls eine CO<sub>2</sub>-Ampel zur Veranschaulichung

CO<sub>2</sub> ist für das blosse Auge nicht sichtbar. Echtzeitdaten zum CO<sub>2</sub>-Gehalt in der Luft helfen dabei zu entscheiden, wann man in einem Innenraum lüften sollte. In diesem Kontext können folgende Fragen mit den Schüler und Schülerinnen diskutiert werden:

- Was ist die Bedeutung des Ampelsystems? Wann soll im Schulzimmer gelüftet werden? Wie lange soll gelüftet werden?
- Zur Info: Ab 1400 ppm (part per million) wird die Ampel gelb und ab 2000 ppm wird sie rot
- Wenn alle LED-Lampen rot leuchten, beträgt der gemessene CO<sub>2</sub>-Gehalt in der Luft (mindestens) 2000 ppm. Das Gerät ist so programmiert, dass der kleinstmögliche registrierte Wert 350 ist. Um wie viel ppm ist der geschätzte CO<sub>2</sub>-Gehalt gestiegen, wenn eine zusätzliche Lampe aufleuchtet?
- Zur Info: Dies kann folgendermassen ermittelt werden. Besteht die Ampel aus 12 LED-Lampen, dann ist der Anstieg pro Lampe  $(2000 \text{ ppm} - 350 \text{ ppm}) / (12 - 1) = 150 \text{ ppm}$ .
- Wie hoch ist die CO<sub>2</sub>-Konzentration, wenn die Hälfte der LED-Lampen leuchtet?
- Wo ist der Einsatz des Geräts am sinnvollsten, sprich wo wird es am dringendsten gebraucht? Diesbezüglich können die Schüler und Schülerinnen überlegen und ausprobieren, an welchem Ort im Schulhaus die Luftqualität am schlechtesten ist. Gibt es Orte, wo die Ampel besonders schnell gelb wird? Gibt es Orte, wo die Ampel nie grün ist/wird?

Benötigte Materialien: Schreibutensilien, Visualisierung der Daten über die Zeit

Der Sensor misst den CO<sub>2</sub>-Gehalt nicht eins zu eins. Die Messung ist eine Annäherung. Dabei gibt es eine Abweichung von +/- 40 ppm (part per million). Das heisst, dass der Sensor möglicherweise einen um 40 ppm zu hohen oder zu tiefen Wert anzeigt. Die Schüler und Schülerinnen können diese Messungenauigkeit in einer Visualisierung der CO<sub>2</sub>-Daten einzeichnen. Basierend darauf können folgende Fragen diskutiert werden:

- Welchen Effekt hat die Mess(un)genauigkeit auf die Visualisierung? Hier kann mit den Schüler und Schülerinnen darüber gesprochen werden, dass kleine Schwankungen des CO<sub>2</sub>-Werts über die Zeit nicht viel Bedeutung beigemessen werden kann, da die Messung fehlerbehaftet ist.
- Was wäre, wenn der Messfehler statt +/- 40 ppm +/-400 ppm wäre. Welche Implikationen hätte dies für das Lüftungsverhalten im Schulzimmer?
- Ab welchem Messfehler würde der Sensor unbrauchbar?

### **Themen und Werkzeuge gesucht!**

Hinweis: Diese Seite ist work in progress. Über Vorschläge für weitere Themenblöcke und ausgearbeitete Unterrichtsmaterialien freuen wir uns sehr.

### **Weiterführende Infos**

**Ist diese Seite verständlich?**

**Vielen Dank für Ihr Feedback!**

**Für dieses Thema zuständig:**

Filtern nach Amt, Fachstelle, Direktion

**Meist gesucht**

