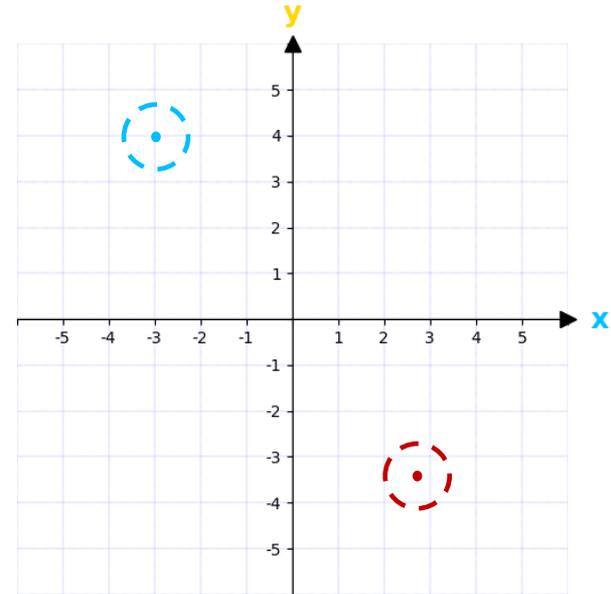


Streudiagramme

Der Blick in die Daten

Streudiagramme

- Stellen Datenpunkte in einem Koordinatensystem dar
- In einem zweidimensionalen Koordinatensystem:
 - 2 Dimensionen
 - jeder Punkt ein Wertepaar
 - zum Beispiel
 - $(-3|4)$
 - $(2,7|-3,3)$



Beispiel: die “Irisdaten”

- Bekannte gemeinfreie Datenbank mit Daten zur Klassifizierung von Schwertlilien: die “Irisdaten”
- Erhältlich z.B. hier <https://www.kaggle.com/datasets/arshid/iris-flower-dataset>
- Blattlängen und –breiten verschiedener Klassen (Arten) von Schwertlilien:



Iris setosa



Iris versicolor

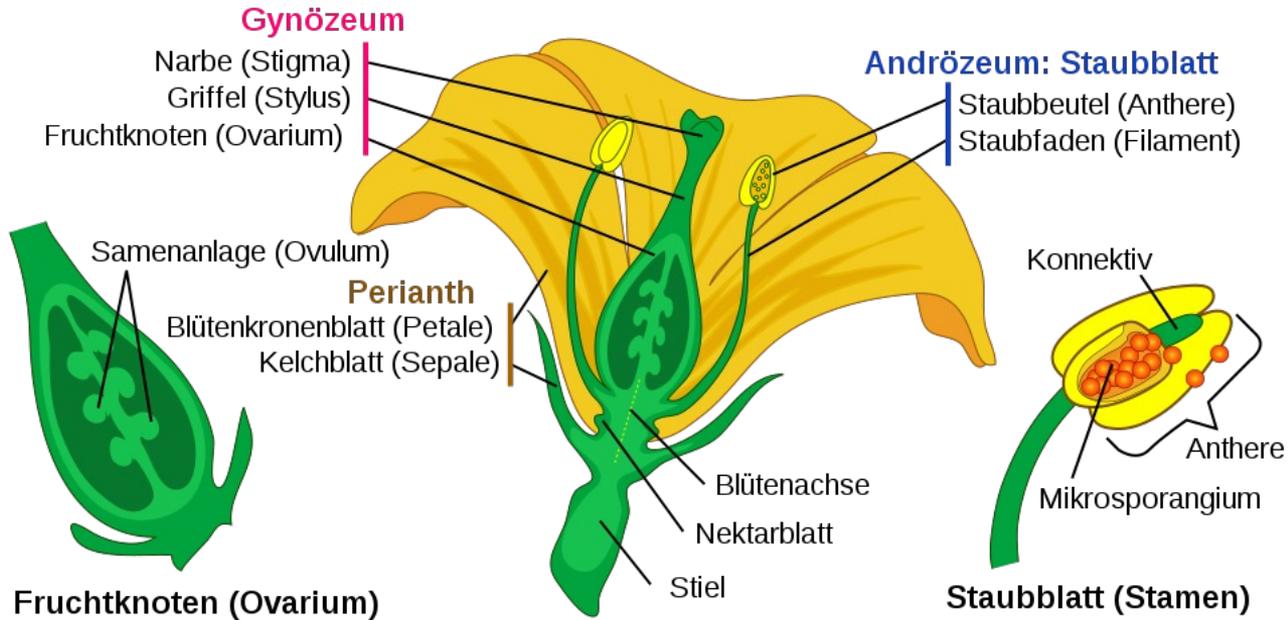


Iris virginica

Die Irisdaten

	Kelchblattlänge	Kelchblattbreite	Kronblattlänge	Kronblattbreite	Irisart
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

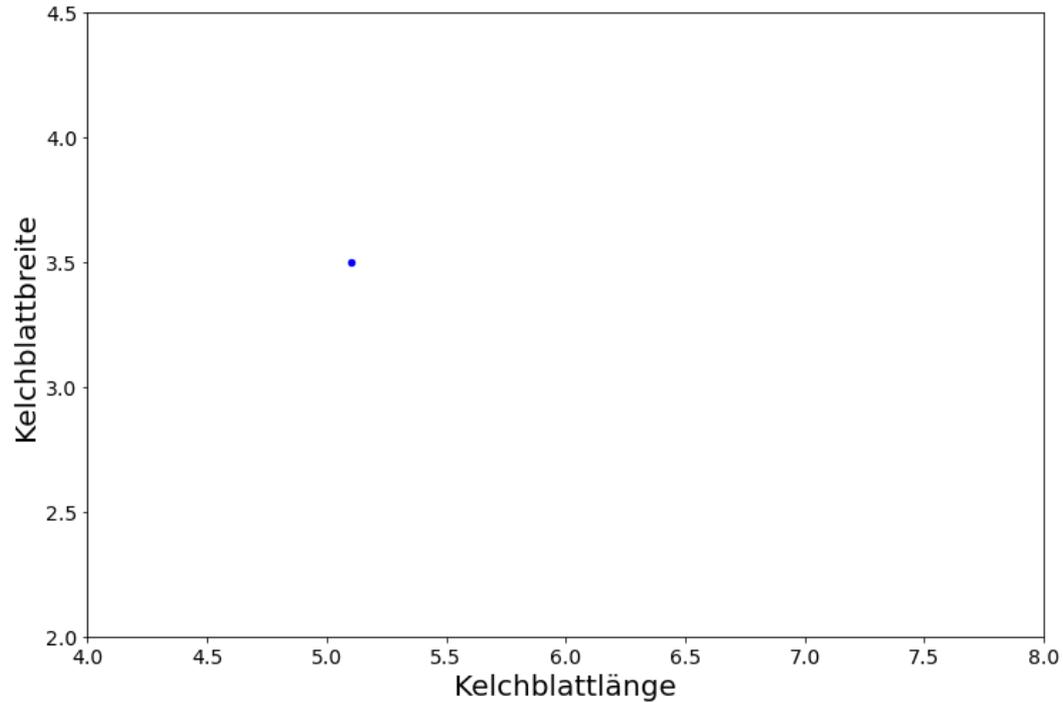
Was sind Kelchblätter und Kronblätter?



Das Wertepaar für den ersten Datenpunkt

	Kelchblattlänge	Kelchblattbreite	Kronblattlänge	Kronblattbreite	Irisart
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

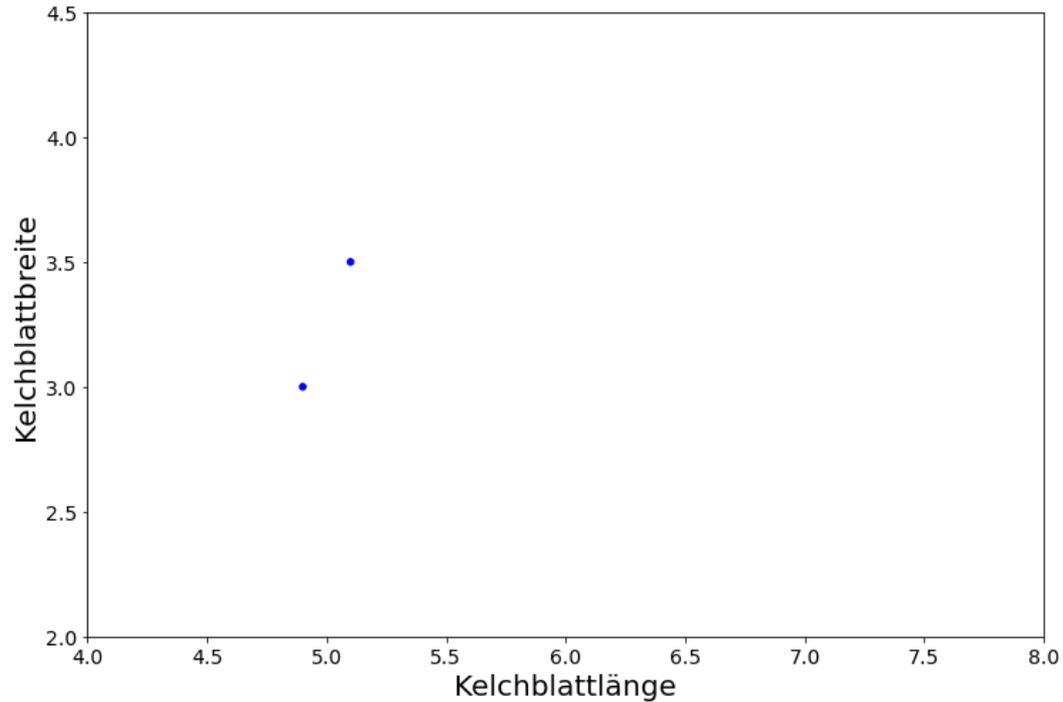
Der erste Punkt im Streudiagramm



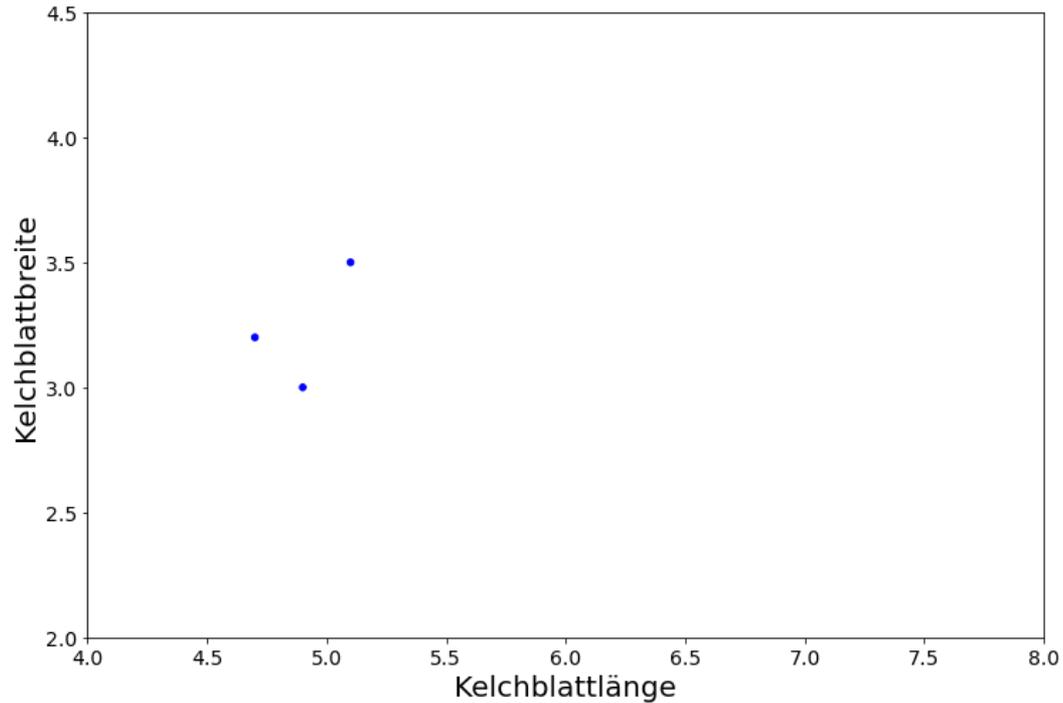
Das Wertepaar für den zweiten Datenpunkt

	Kelchblattlänge	Kelchblattbreite	Kronblattlänge	Kronblattbreite	Irisart
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

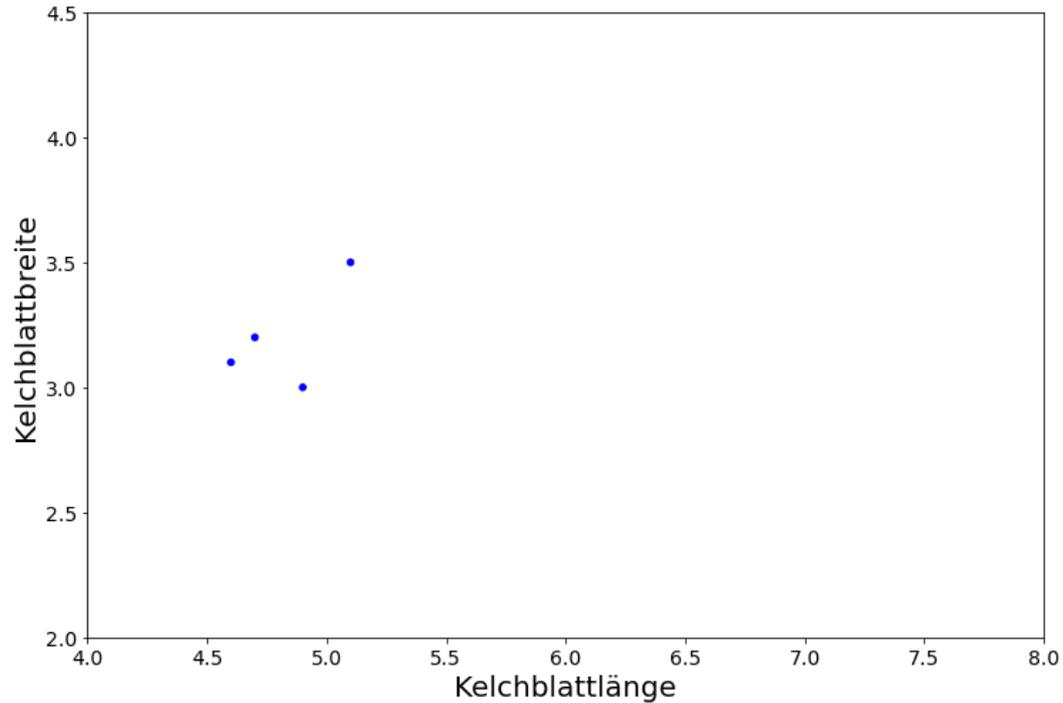
Der zweite Punkt im Streudiagramm



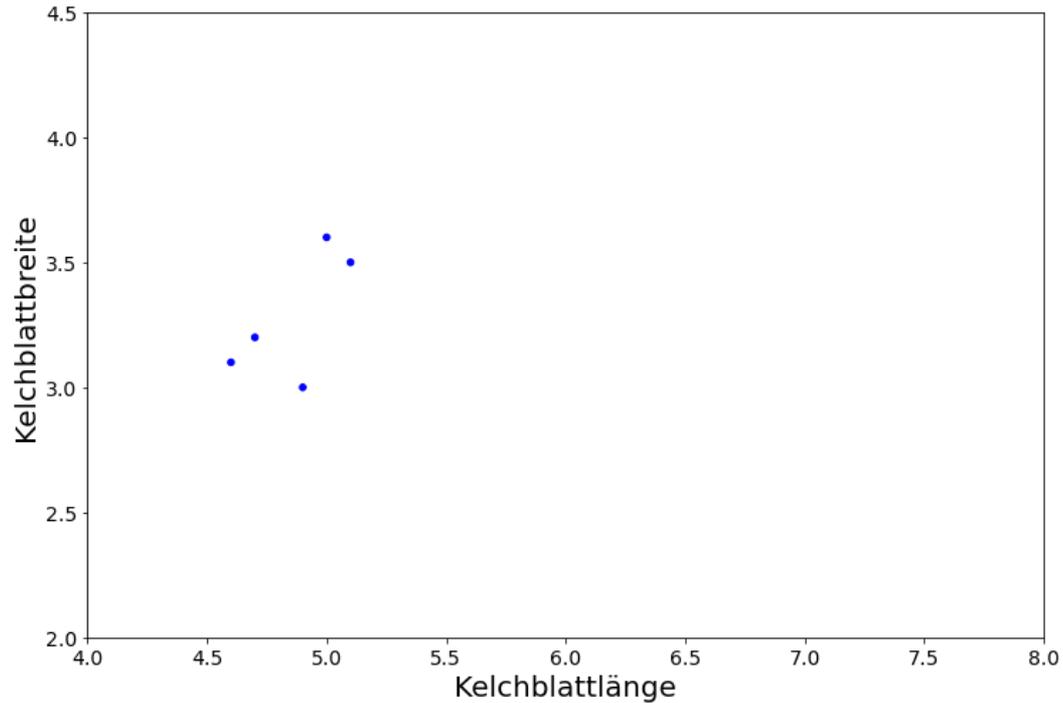
Weitere Punkte im Streudiagramm



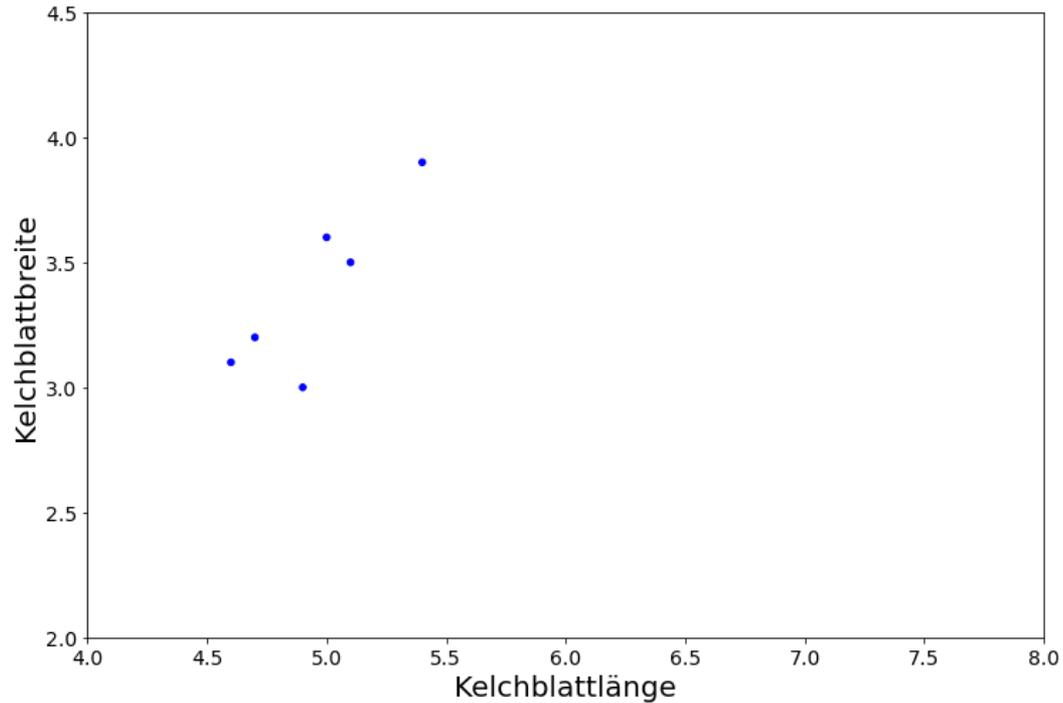
Weitere Punkte im Streudiagramm



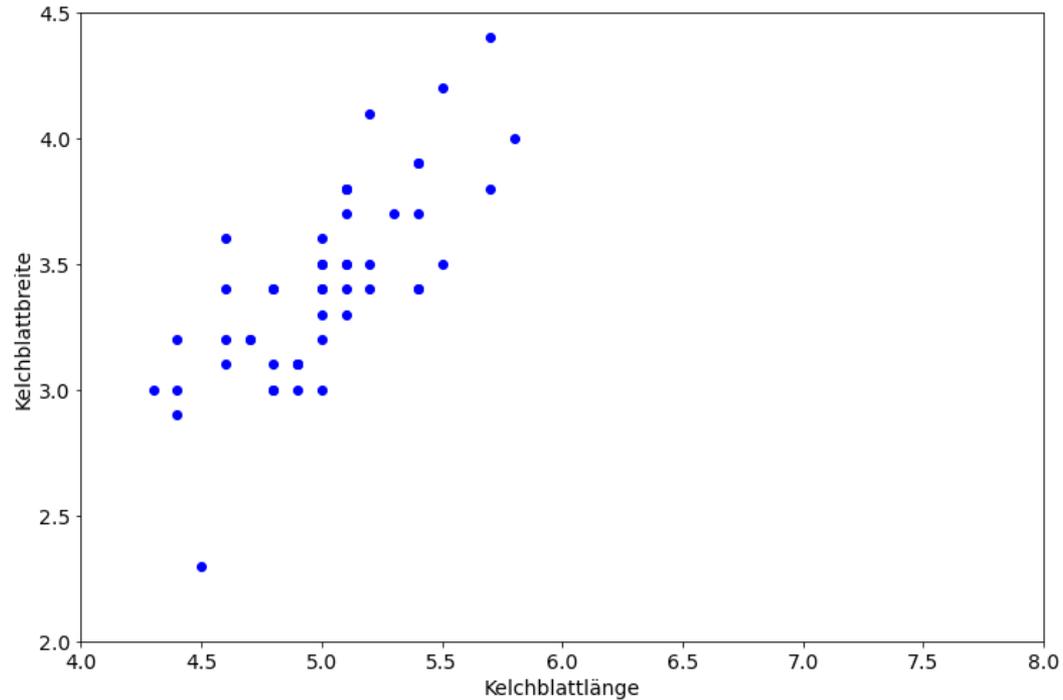
Weitere Punkte im Streudiagramm



Weitere Punkte im Streudiagramm



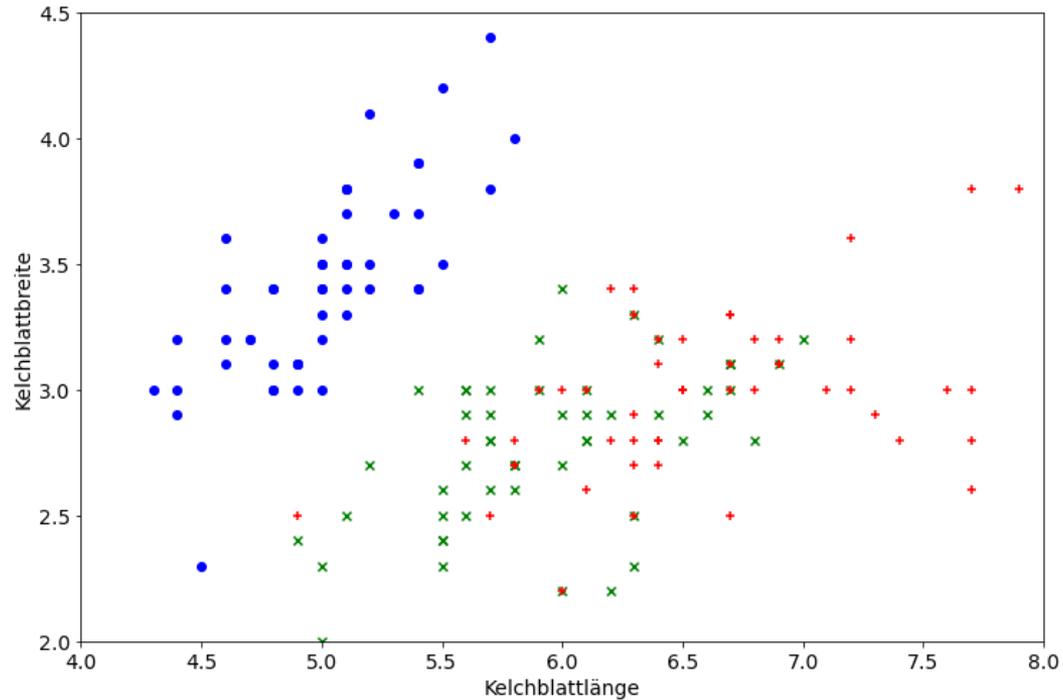
Iris Setosa im Streudiagramm



**Ein Streudiagramm visualisiert die
Beziehung zwischen Paaren von
Werten.**

**Streudiagramme werden oft auch
als Scatterplots bezeichnet.**

Alle Klassen im Streudiagramm



- Setosa
- × Versicolor
- + Virginica

Streudiagramme interpretieren: Erkenntnisse bisher

Iris Setosa



Kurze, breite Kelchblätter

Iris Versicolor



Mittlere Kelchblattlänge,
nicht allzu breit

Breite ähnlich zu Virginia

Iris Virginica



Längste Kelchblätter,
nicht allzu breit

Breite ähnlich zu Versicolor

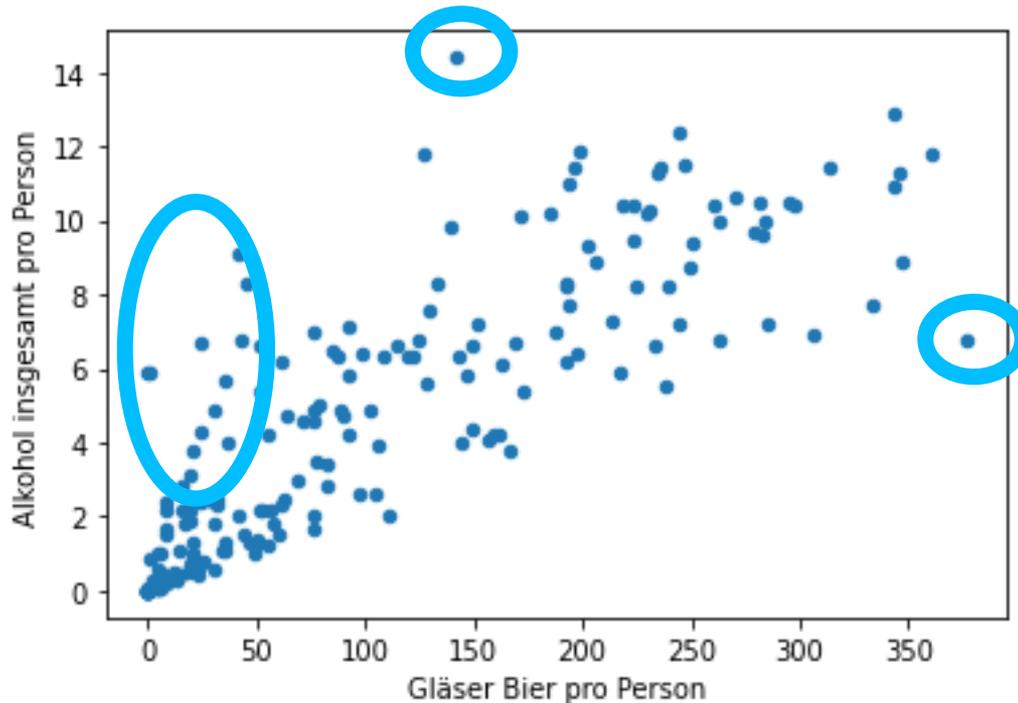
**Streudiagramme sind sehr gut
geeignet, um Muster in der
Verteilung zweier Variablen
erkennbar zu machen.**

Beispiel: die “Alkoholdaten”

- Datenbank mit Daten zum Alkoholverbrauch in verschiedenen Ländern
- Erhältlich z.B. hier <https://github.com/fivethirtyeight/data/tree/master/beer-consumption>

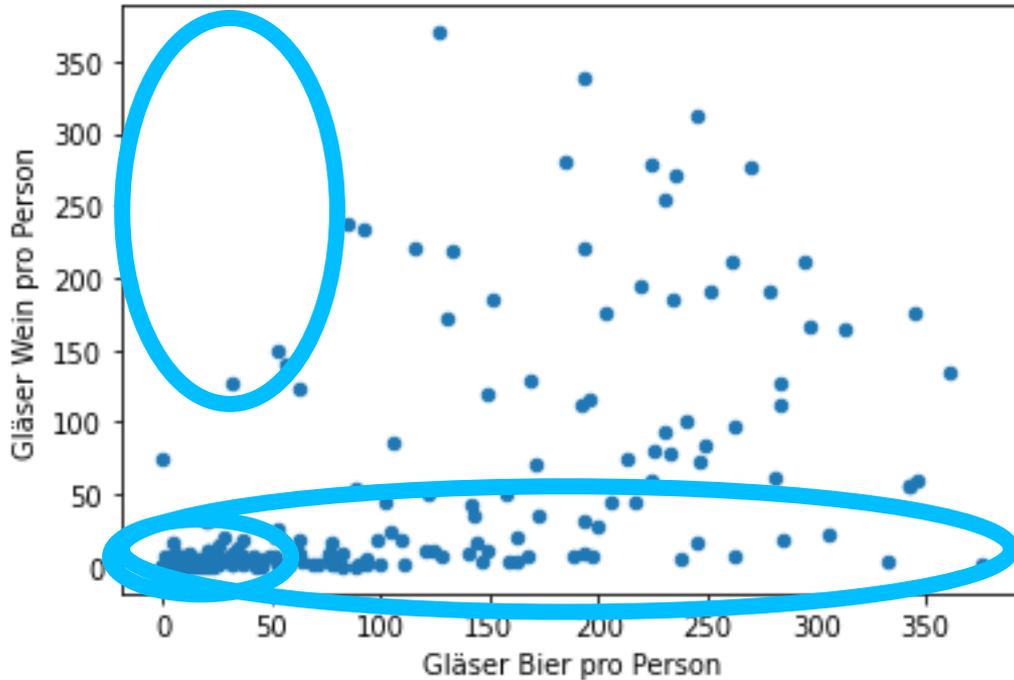
	Land	Gläser_Bier	Gläser_Schnaps	Gläser_Wein	reiner_Alkohol_insgesamt
0	Afghanistan	0	0	0	0.0
1	Albania	89	132	54	4.9
2	Algeria	25	0	14	0.7
3	Andorra	245	138	312	12.4
4	Angola	217	57	45	5.9

Bier und Alkohol pro Kopf in verschiedenen Ländern



- Zu erwarten: je mehr Bier getrunken wird, desto mehr Alkohol wird insgesamt getrunken
- Es gibt auch Länder, in denen wenig Bier, aber dennoch Alkohol getrunken wird
- Land mit dem höchsten Bierverbrauch ist bei weitem nicht das Land mit dem höchsten Alkoholverbrauch insgesamt, und umgekehrt

Bier und Wein pro Kopf in verschiedenen Ländern



- In vielen Ländern wird sowohl wenig Wein als auch wenig Bier getrunken
- Es gibt sehr viele Länder, in denen eher wenig Wein getrunken wird
- Es gibt wenig Länder, in denen viel Wein, aber wenig Bier getrunken wird

Jupyter Notebook zu Streudiagrammen

mittel

Dr. Antje Schweitzer

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung



Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung
Institut für Software Engineering



Industrie- und Handelskammer
Reutlingen

Reutlingen | Tübingen | Zollernalb



Region Stuttgart



Industrie- und Handelskammer
Karlsruhe



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Lizenzbestimmungen

“Streudiagramme” 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

Seiten 3 und 18: *Iris versicolor*, von terri bateman, gemeinfrei, via Wikimedia Commons. *Iris setosa*, von Денис Анисимов, gemeinfrei, via Wikimedia Commons.

Seiten 3 und 18: *Iris virginica*, von Eric Hunt, lizenziert unter CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0>), via Wikimedia Commons

Seite 5, Mature flower diagram, von Mariana Ruiz, bearbeitet von Matt, gemeinfrei, via Wikimedia Commons