

Geodatenanalyse I: Multivariate Statistik

Kathrin Menberg



Stundenplan

	08:30 – 12:30 Uhr	13:30 – 17:30 Uhr
Montag	Tag 1 / Block 1	Tag 1 / Block 2
Dienstag	Tag 2 / Block 1	Tag 2 / Block 2
Mittwoch	Tag 3 / Block 1	Tag 3 / Block 2
Donnerstag	Tag 4 / Block 1	Tag 4 / Block 2
Freitag	Tag 5 / Block 1	Tag 5 / Block 2

- ▶ 2.4 Bivariate Statistik
- ▶ **2.5 Multivariate Statistik**
- ▶ 2.6 Zeitreihenanalyse

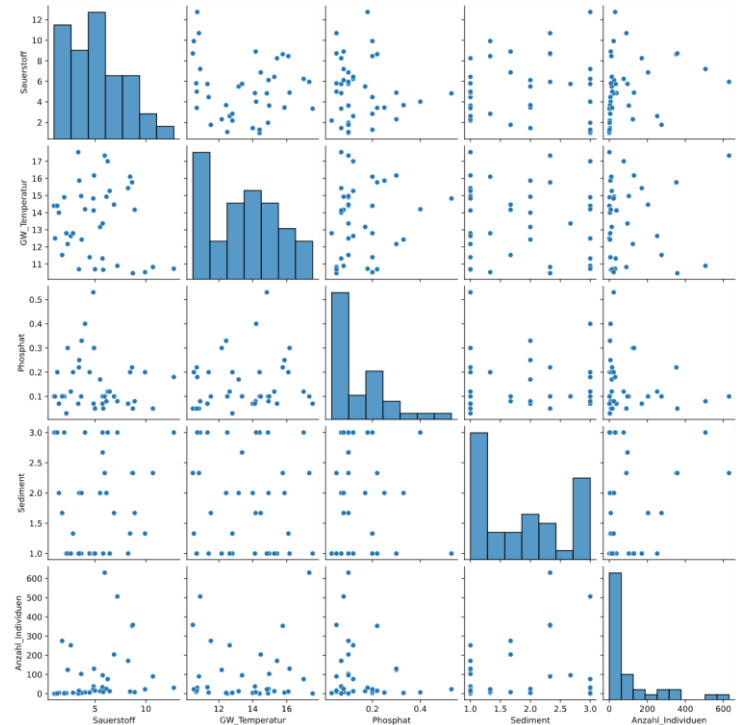
Lernziele Block 2.5

Am Ende der Stunde werden die Teilnehmer:


- ▶ ... mit den statistischen Konzepten der Datentransformation, Eigenvektoren und Eigenwerten vertraut sein.
- ▶ ... Methoden zur Reduzierung von Dimensionen auf Geodatenätze anwenden und die Ergebnisse graphisch darstellen können.

n-dimensionale Datensätze

- ▶ Beziehungen zwischen allen Parametern
- ▶ Parameterraum (parameter space)
- ▶ Gemeinsame graphische Darstellung von vielen Parametern schwierig
- ▶ Scatterplots auf drei Dimensionen beschränkt



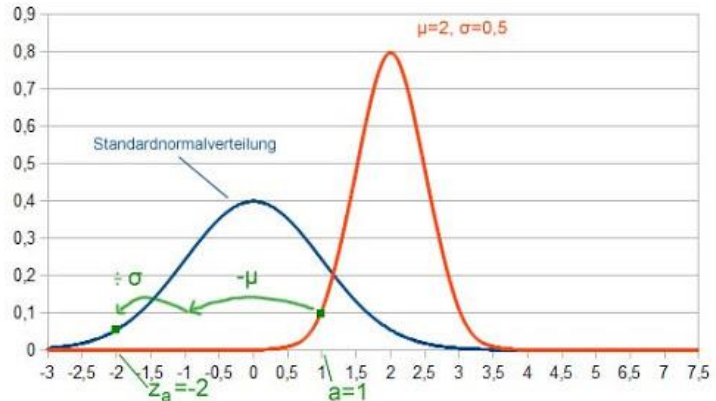
Multivariate Statistik

- ▶ Dimensionen von Datensätzen reduzieren ohne viel Information zu verlieren
- ▶ 2D – Visualisierung von komplexen Beziehungen
 - ▶ Hauptkomponentenanalyse (principal component analysis) 
 - ▶ Faktorenanalyse (factor analysis)
 - ▶ Unabhängigkeitsanalyse (Independent Component Analysis)
- ▶ Datenpunkte mit ähnlichen Eigenschaften identifizieren
 - ▶ Clusteranalyse (cluster analysis)

Transformieren von Datensätzen

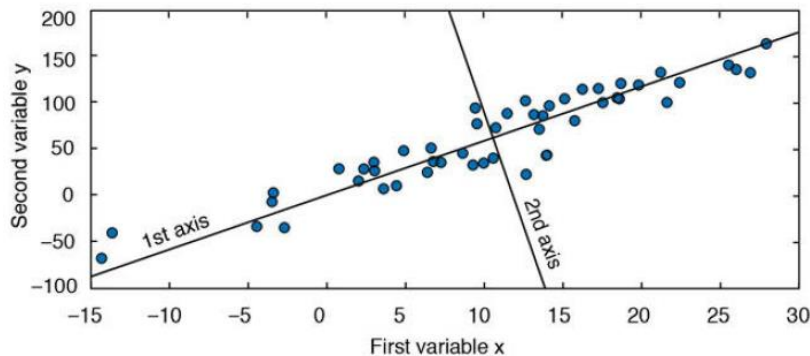
- ▶ Rohdaten oft nicht normalverteilt, Unterschiede in Varianzen zwischen einzelnen Parametern, usw.
- ▶ Standardisieren von Daten
- ▶ Standard-Normalverteilung
 - ▶ $X \sim N(0,1)$

$$\text{standardized } x_i = \frac{x_i - \bar{x}}{\text{std}(x)}$$



Hauptkomponentenanalyse

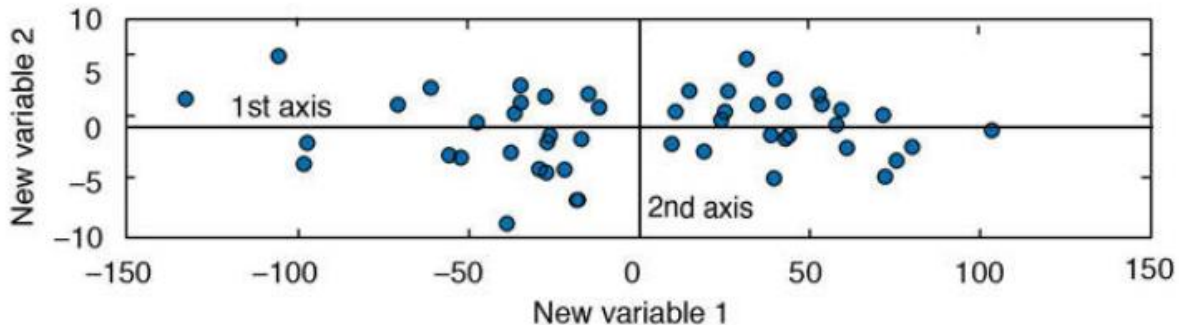
- ▶ Ziel: Reduzieren von Dimensionen
- ▶ Prinzip:
 - ▶ Bestimmung von linearen Beziehungen zwischen Parameter
 - ▶ Ersetzen von Gruppen korrelierender Parameter durch neue nicht korrelierte Variablen



Trauth (2015) Fig. 9.1

Hauptkomponentenanalyse

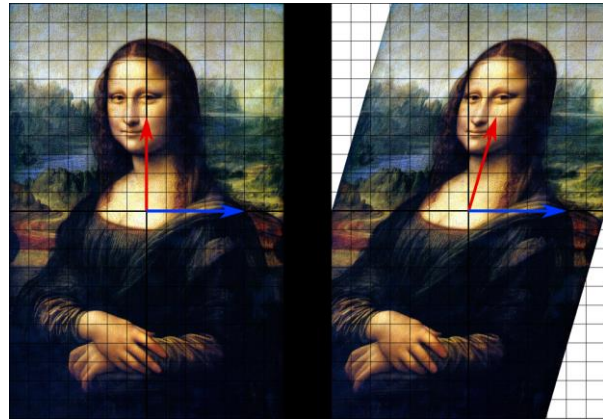
- ▶ Ziel: Reduzieren von Dimensionen
- ▶ Prinzip:
 - ▶ Bestimmung von linearen Beziehungen zwischen Parameter
 - ▶ Ersetzen von Gruppen korrelierender Parameter durch neue nicht korrelierte Variablen



Trauth (2015) Fig. 9.1

Hauptkomponentenanalyse

- ▶ Identifizieren der Achsen, d.h. der Hauptkomponenten über die Kovarianz der Parameter
 - ▶ Eigenvektoren ($n \times 1$)
 - ▶ Eigenwerte (Skalar)



wikipedia.org

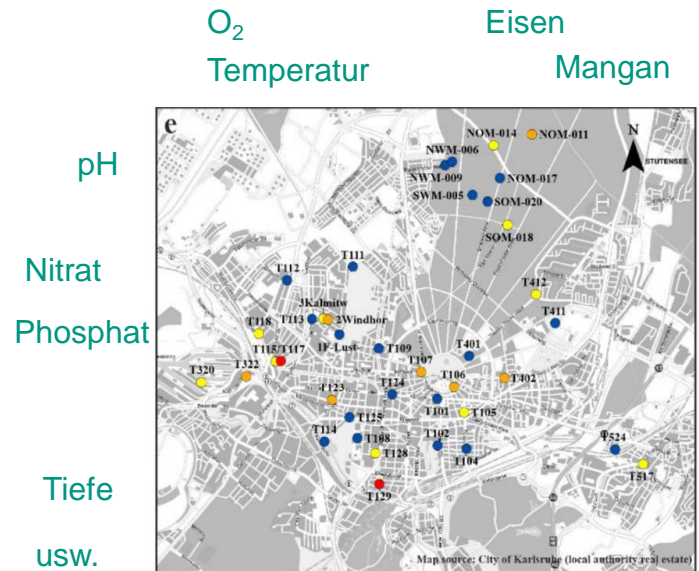
- ▶ Alle Hauptkomponenten zusammen enthalten die gesamte Varianz des Datensatzes
- ▶ Transformieren, bzw. Rotieren der Datenmatrix

Übung 2.5: Multivariate Statistik

► Grundwasserdatensatz Karlsruhe

- Hauptkomponentenanalyse
- Matrizenrechnung
- Visualisierung

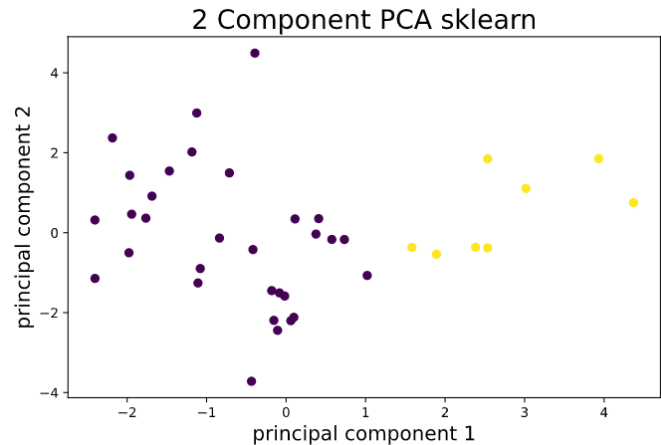
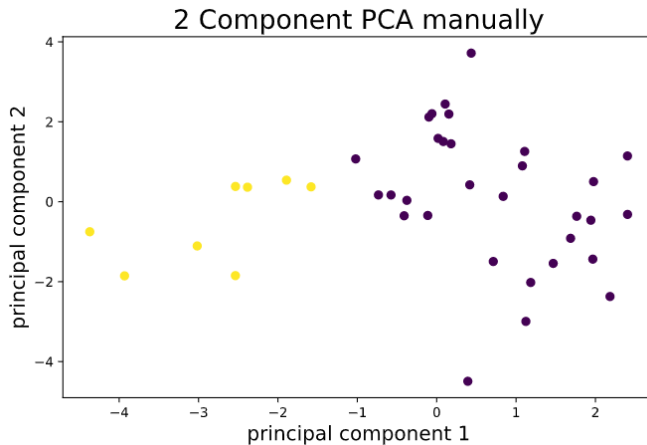
► Aufgaben in Jupyter Notebook: geodatenanalyse_1-2-5



Koch et al. (2020)

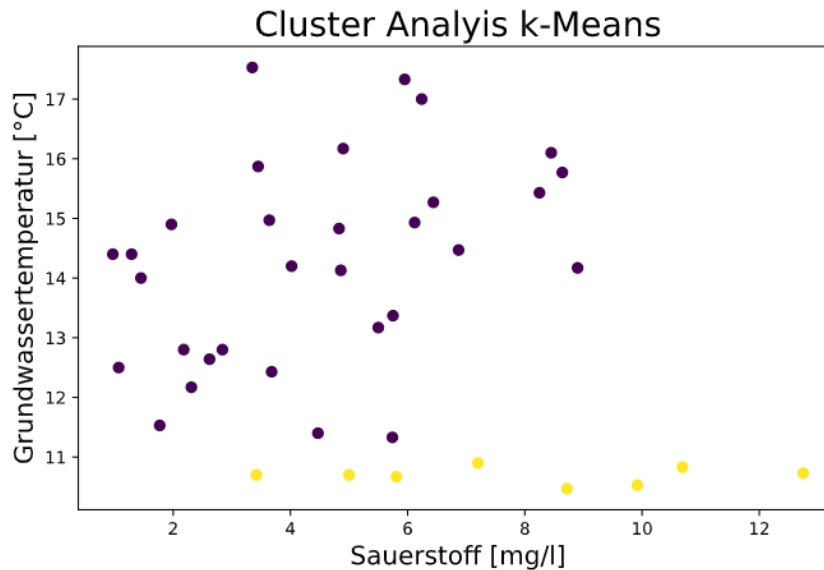
Aufgabenbesprechung

► Hauptkomponentenanalyse



$\text{var_explained} = [0.194, 0.175]$
 $\text{var_tot} = [0.370]$

► Clusteranalyse: K-Means



Literatur

- ▶ Trauth (2015) MATLAB Recipes for Earth Sciences (4th Ed.), Springer
- ▶ Koch et al. (2020) Groundwater fauna in an urban area: natural or affected?, Hydrology and Earth System Sciences Discussions
- ▶ Lever et al. (2017) Principal component analysis, Nature Methods 14(7), 641-642

Nützliche Weblinks:

- ▶ <https://towardsdatascience.com/a-complete-guide-to-principal-component-analysis-pca-in-machine-learning-664f34fc3e5a>

