

# Angewandte Datenverarbeitung und Visualisierung

Kursübersicht - SoSe2024

Daniela Palleschi  
Humboldt-Universität zu Berlin

Di. den 16.04.2024

# Überblick

- 1 Moodle
- 2 Kurs-Website
- 3 Kursübersicht
- 4 Software
- 5 Ressourcen

# 1 Moodle

Moodle: “Angewandte Datenverarbeitung und Visualisierung: R für Linguistik und Sozialwissenschaften”

Schlüssel: r4ling

# Moodle: Allgemein

- Forum
  - Bitte stellen Sie alle Fragen/Probleme, die Sie haben, ins Forum, damit Ihre Kommilitonen von den Antworten profitieren können
- R Cheatsheets (Spickzettel)
  - normalerweise eine 1-seitige Zusammenfassung nützlicher Funktionen und Befehle, die für ein bestimmtes Paket oder eine bestimmte Aufgabe relevant sind
- Ressourcen
  - erforderliche und vorgeschlagene Materialien

# Moodle: Ressourcen

- Die Folien/Vortragsnotizen sind in drei Formate verfügbar:
  - Folien (endend mit `_folien.html`)
  - HTML (endend mit `_blatt.html`)
  - PDF (endend mit `.pdf`)

# Moodle: Thematische Überschriften

- erforderliche und vorgeschlagene Lektüre/Materialien sind unter jeder Woche aufgeführt
- Unterrichtsmaterialien (z. B. meine Folien/Vortragsnotizen) werden kurz vor Beginn des Unterrichts zur Verfügung stehen

# 2 Kurs-Website

- Ich habe eine Kurs-Website erstellt: [daniela-palleschi.github.io/r4ling](https://daniela-palleschi.github.io/r4ling)
- Diese ist ergänzend und dient dazu, den Unterricht über jedes Semester hinweg zu erleichtern
- im Moment enthält sie dieselben Informationen wie die Folien, die ich jede Woche präsentiere

# 3 Kursübersicht

- Studienleistungen
  - 3LP
    - 1LP: wöchentlichen Hausaufgaben
      - 4 über Moodle einzureichen, 4 Wochen nach Wahl (je 0,5LP)
    - 2LP: zwei “in-class” Berichte (je 1LP)
- Sprechstunden: Dienstags, 11.00-12.00 (nach Vereinbarung)



# Kursziele

- im Großen und Ganzen werden Sie lernen, wie man mit einem Datensatz arbeitet und ihn beschreibt
- es wird vorausgesetzt, dass Sie keine Erfahrung mit Programmierung oder Datenanalyse haben

# Was werden Sie lernen?

- Programmieren
- Datenimport
- Daten bereinigen
  - Daten angemessen zu strukturieren
- Datenvisualisierung
  - Wie man verschiedene Arten von Daten darstellt
- Datenkommunikation
  - beschreiben

## Was werden Sie *nicht* lernen?

- Wir werden nicht auf die Inferenzstatistik eingehen (d.h. Hypothesentests)
- Big Data
- andere Programmiersprachen wie Python, Julia
  - aber Sie können sie in RStudio verwenden

# 4 Software

- R: eine Programmiersprache, in der wir Code schreiben werden (grundlegende Software)
- RStudio: ein Programm, das uns die Arbeit in R erleichtert (zusätzliche Entwicklungsumgebung)
- LaTeX: ein Satzsystem, das Dokumente im PDF-Format erzeugt
- warum R?
  - R und RStudio sind quelloffene und kostenlose Software
  - sie sind in Wissenschaft und Wirtschaft weit verbreitet



# R installieren

- wir brauchen die kostenlose und quelloffene Statistiksoftware R, um unsere Daten zu analysieren
- herunterladen und installieren R: <https://www.r-project.org>

# RStudio installieren

- wir brauchen RStudio, um einfacher mit R arbeiten zu können
- RStudio herunterladen und installieren: <https://rstudio.com>
- es kann hilfreich sein, Englisch als Sprache in RStudio beizubehalten
  - Wir werden mehr hilfreiche Informationen finden, wenn wir Fehlermeldungen auf Englisch im Internet suchen
- Wenn Sie Probleme bei der Installation von R oder RStudio haben, sehen Sie sich diese Hilfeseite an (auf Deutsch): <http://methods-berlin.com/wp-content/uploads/Installation.html>

# LaTeX installieren

- wir werden nicht direkt mit LaTeX arbeiten, aber es wird im Hintergrund benötigt
- LaTeX herunterladen und installieren: <https://www.latex-project.org/get/>

# 5 Ressourcen

- viele Aspekte dieses Kurses sind inspiriert von Nordmann & DeBruine ([2022](#)) und Wickham et al. ([2023](#))
  - beide frei online verfügbar (in Englisch)
- für deutschsprachige Ressourcen besuchen Sie die Website der [Methodengruppe Berlin](#)



# Fehlersuche (EN: Troubleshooting)

- Fehlermeldungen sind in der Programmierung sehr häufig, und zwar auf allen Niveaus.
- Wie man Lösungen für diese Fehlermeldungen findet, ist eine Kunst für sich
- Google ist Ihr Freund! Wenn möglich, googeln auf Englisch, um mehr Informationen zu erhalten

# Literaturverzeichnis

Nordmann, E., & DeBruine, L. (2022). *Applied Data Skills* (Version 2.0). Zenodo. <https://doi.org/10.5281/zenodo.6365078>

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for Data Science* (2. Aufl.). <https://r4ds.hadley.nz/>

