# Angewandte Statistik

Julian Hinz, Timo Adam — Universität Bielefeld

# Session 1

**Organisatorisches, Roadmap und R**

# Team

Julian Hinz

Timo Adam

Lars Abt

Marc Panhorst

Aaron Lohmann

# Was ist *angewandte* Statistik?

# Statistik

- … entwickelt Methoden zur Analyse von Daten
- … und unterstützt bei deren Anwendung auf echte Daten

# Was ist angewandte Statistik?

## Mathematische Statistik

- befasst sich primär mit **Methodenentwicklung**
- im Vordergrund stehen die mathematischen Eigenschaften
- oft abstrakt, praktischer Nutzen nicht immer sofort klar (Mathe halt…)

# Was ist angewandte Statistik?

## Angewandte Statistik

- soll mit geeigneten Methoden aus **echten Daten** Informationen gewinnen
- braucht Methodenkompetenz sowie Verständnis des inhaltlichen Problems
- in der Regel stark interdisziplinäre Zusammenarbeit

# Was zeichnet angewandte Statistik aus?

- Bedeutung der angewandten Statistik wächst im Datenzeitalter rasant
- hohes Maß an Interdisziplinarität — mit allen Vor- und Nachteilen: hoher Bedarf an effektiver Kommunikation
- oft aufwändige Datenaufbereitung und viel Programmiertätigkeit
- komplexe Daten, welche es eben nicht erlauben, Standardwerkzeuge anzuwenden

# Angewandte Statistik klingt langweilig?

# Data Science.

# Slides

https://statistik.julianhinz.com

# Plan

1. Organisatorisches
2. Roadmap
3. R

# Vorlesung

- **Dozenten**: Julian Hinz und Timo Adam (julian.hinz@uni-bielefeld.de und timo.adam@uni-bielefeld.de)
- 2 Gastvorlesungen: Aaron Lohmann
- Termin: Dienstags 16:15–17:45 Uhr in H7
- Theorie und Praxis: Computer mitbringen, wir werden programmieren!

# Tutorien

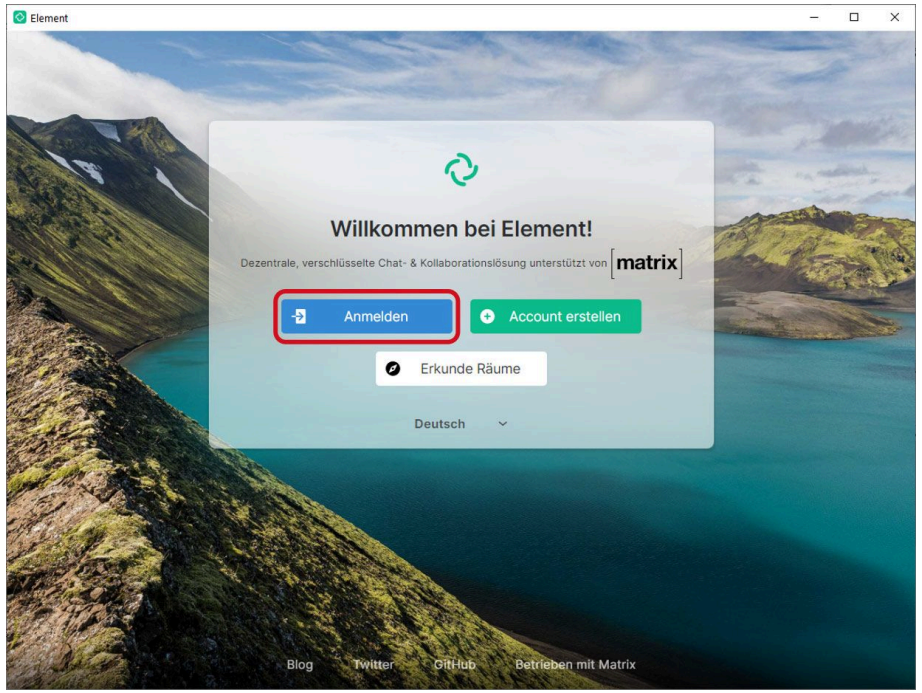
- **Tutoren**: Lars Abt und Marc Panhorst
- 4 Tutorien pro Woche — Donnerstags 12-14h, 14-16h (2 mal), 16-18h
- Platzvergabe morgen, also: **heute** anmelden im ekvv!
- 3 Tutorien in den eKVV-Stundenplan nehmen, inkl. Prioritäten
- Ziel: optimal auf die Projektaufgaben vorzubereiten
- Die Tutorien beginnen nächste Woche (17. April)

Platzvergabe morgen, also:
**jetzt** anmelden im ekvv!

# Material

- Kurs-Website: https://statistik.julianhinz.com
- Fahrmeir L., Heumann, C., Künstler, R., Pigeot, I. und Tutz, G. (2016), *Statistik: Der Weg zur Datenanalyse*, Springer
- Fahrmeir, L., Kneib, T. und Lang, S. (2009), *Regression*, Springer
- Bruce, P., Bruce A. und Gedeck, P. (2021), *Praktische Statistik für Data Scientists*, O'Reilly

# Kommunikation außerhalb der Vorlesung




Element-Raum für Fragen insb. zu den Aufgabenzetteln:

- https://matrix.to/#/#angstat2025:uni-bielefeld.de
- Registrierung mit …@uni-bielefeld.de E-mail-Adresse
- kann direkt im Browser oder per App genutzt werden

# Prüfungsleistung

**Ersttermin** im Sommersemester:

- Portfolioprüfung über Öko/CoMet/AngStat
- je 30 Punkte pro Veranstaltung
- Öko und CoMet liegt bereits hinter Ihnen
- die 30 Punkte in Angewandte Statistik werden über **Projektaufgaben** vergeben

**Zweittermin** im Sommersemester:

- normale Klausur (vermutlich im September) über alle 3 Vorlesungen
- kein Nachteil was den Schwierigkeitsgrad/die Bewertung betrifft

# Projektaufgaben

Formalia:

- **2 Aufgabenzettel** mit je 3-5 Aufgaben, 15 Punkte pro Zettel
- jeweils über einen Zeitraum von einer Woche zu bearbeiten
- Zettel 1: Mitte Mai, Zettel 2: Ende Juni
- **nur Einzelabgaben**

Inhaltlich:

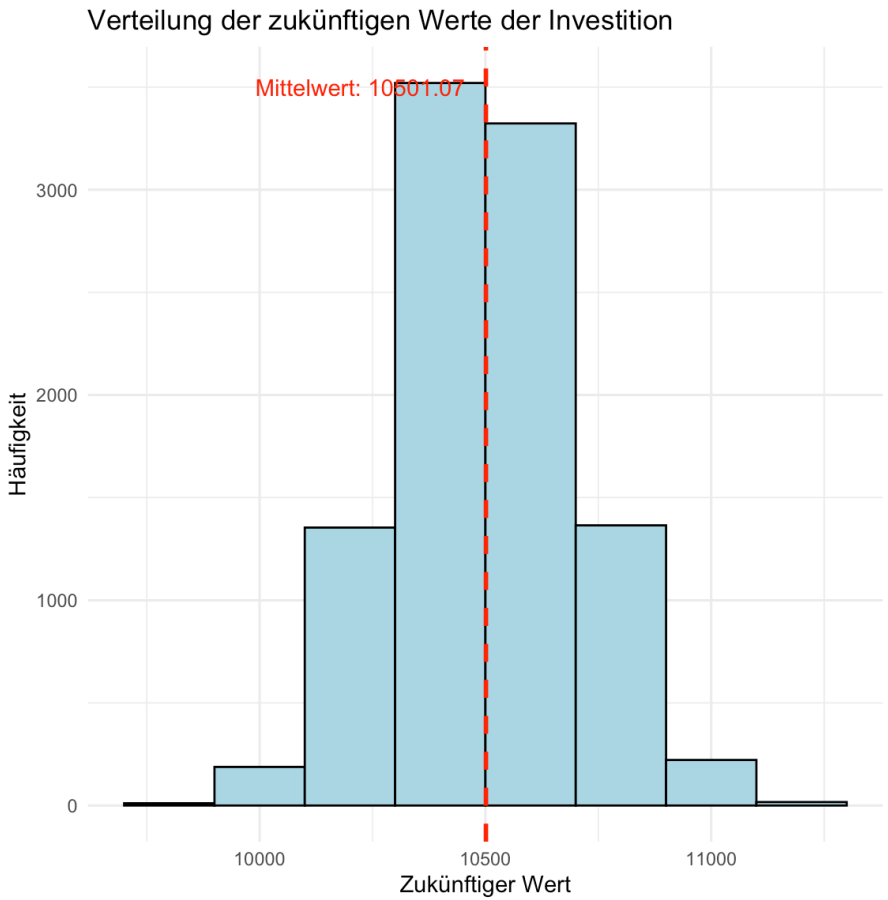
- jede*r von Ihnen bekommt **eigene Datensätze** zur Analyse
- Methoden die in Vorlesung und Tutorien geübt werden
- **Nutzung von R** ist essenziell

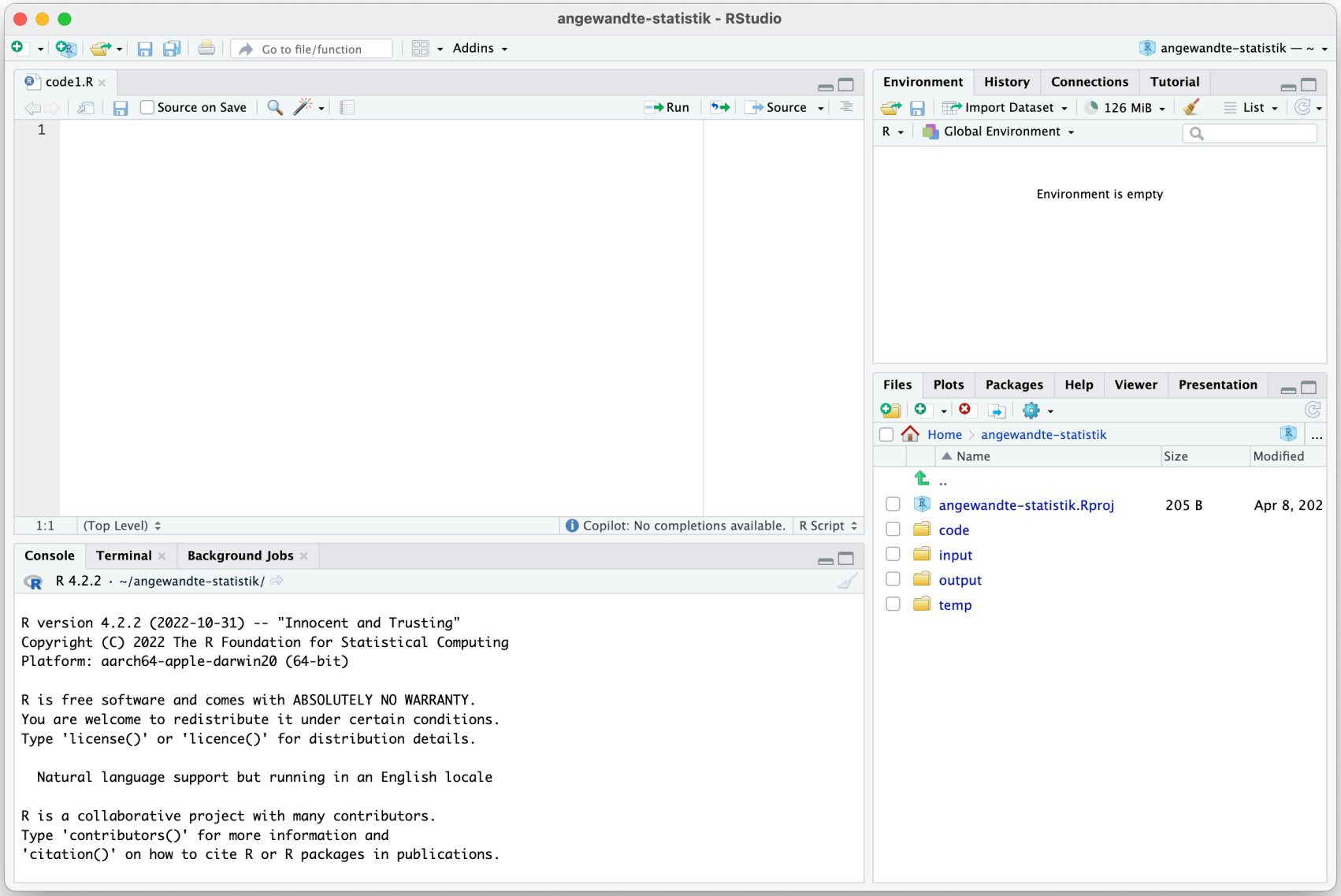
# Roadmap

# Session 1

*heute — Julian Hinz, Timo Adam*

- R
- Simulationen



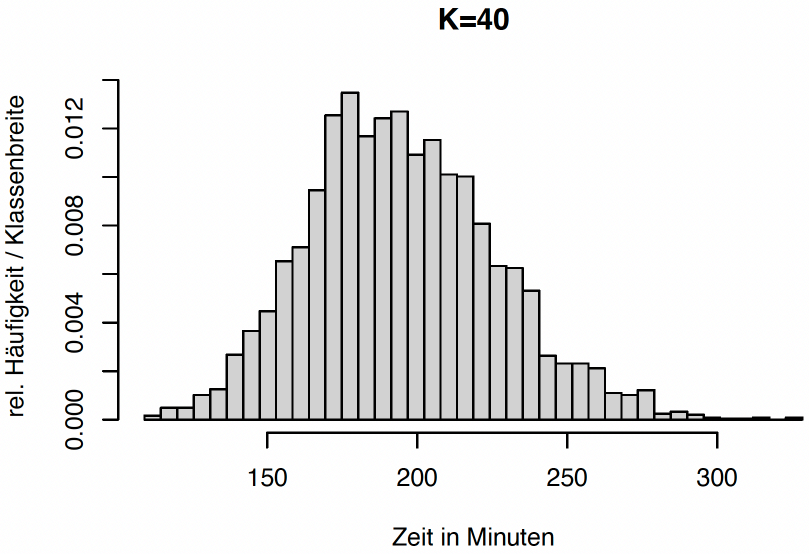

angewandte-statistik - RStudio
Go to file/function
Addins
angewandte-statistik — ~
code1.R
Source on Save
Run
Source
1
1:1
(Top Level)
Copilot: No completions available.
R Script
Console
Terminal
Background Jobs
R 4.2.2 · ~/angewandte-statistik/
R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Environment
History
Connections
Tutorial
Import Dataset
126 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Home
angewandte-statistik
Name
Size
Modified
..
angewandte-statistik.Rproj
205 B
Apr 8, 202
code
input
output
temp

# Session 2

*15. April — Timo Adam*

- Verteilungsschätzung
- empirische Verteilungsfunktionen
- Histogramme

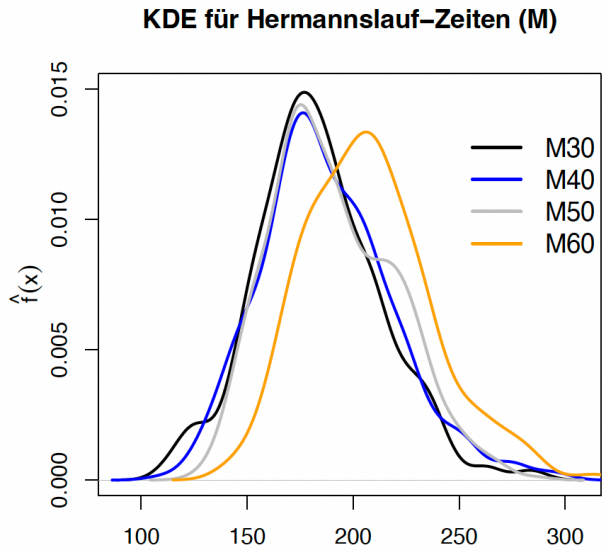
| Platz | Vorname | Nachname | Zeit (in Minuten) |
|---|---|---|---|
| 1 | Elias | Sansar | 109.1 |
| 2 | Patrick | Boehme | 110.4 |
| 3 | Niklas | Hänze | 110.8 |
| 4 | Erik | Peters | 113.4 |
| 5 | Markus | Scheller | 114.7 |
| 6 | Jan | Kaschura | 116.2 |
| .... | ... | ... | ... |

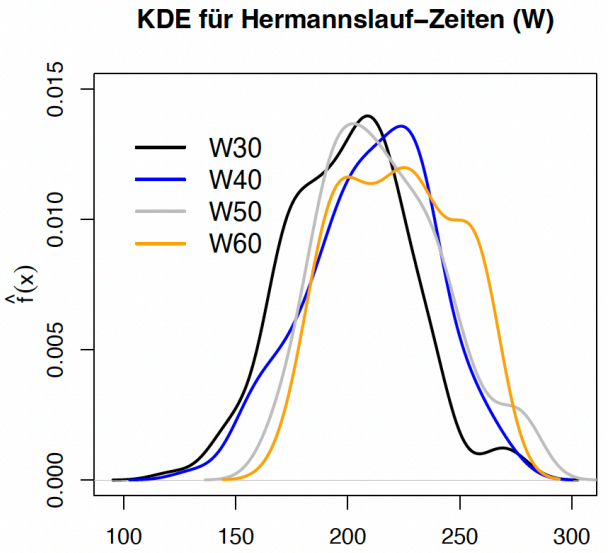


# Session 3

*22. April — Timo Adam*

- Verteilungsschätzung
- Kerndichteschätzer



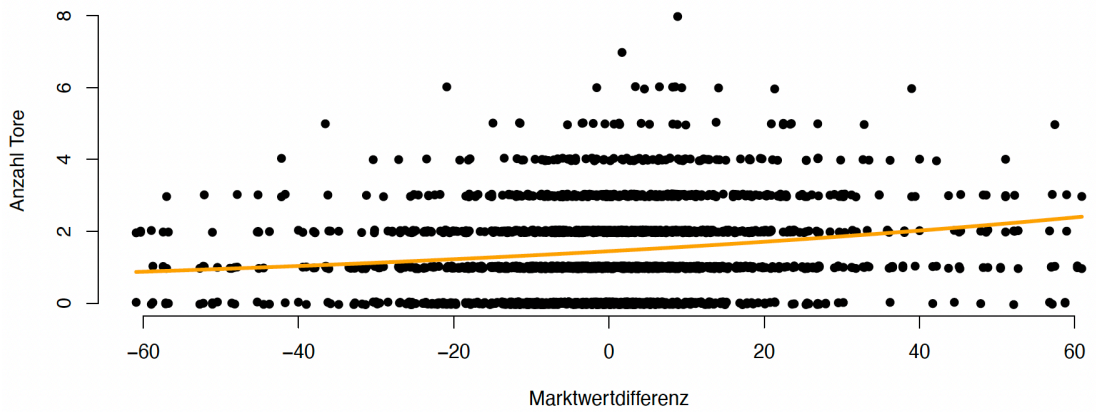
# Session 4

*29. April — Timo Adam*

- Lineare Regression
- Poissonregression (für Zähldaten)

DFB-Pokalsieger: Wie hoch ist die Quote darauf, dass Bielefeld den Pokal gewinnt?

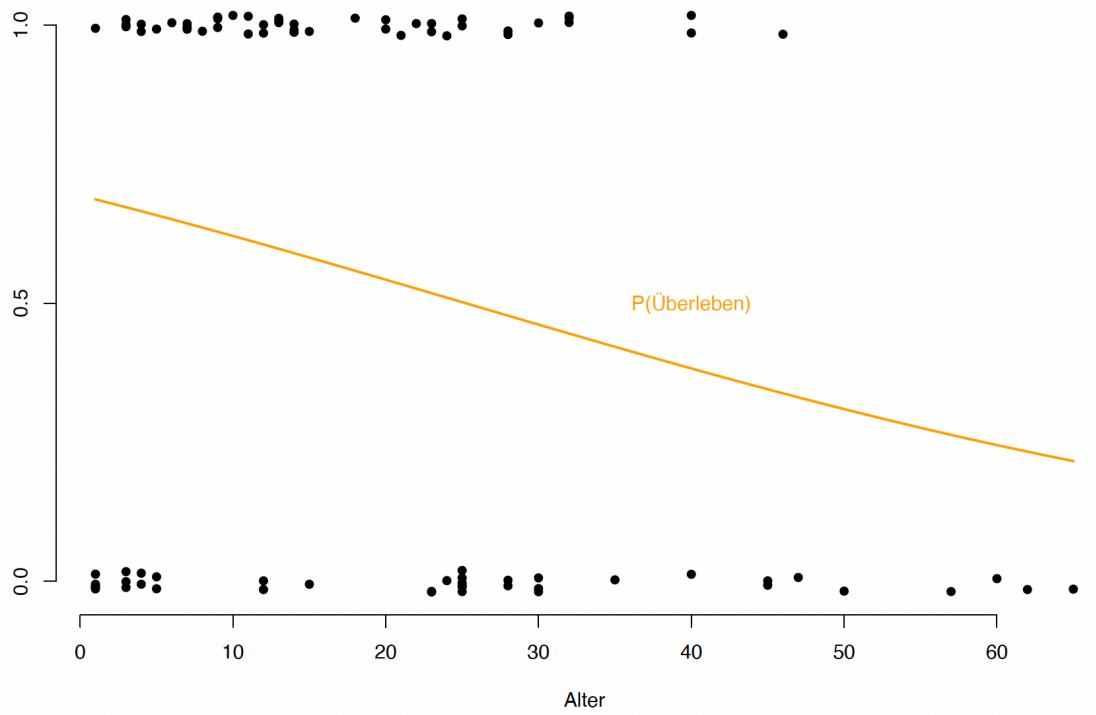
VfB Stuttgart **1,25** Arminia Bielefeld **3,80**



# Session 5

*6. Mai — Timo Adam*

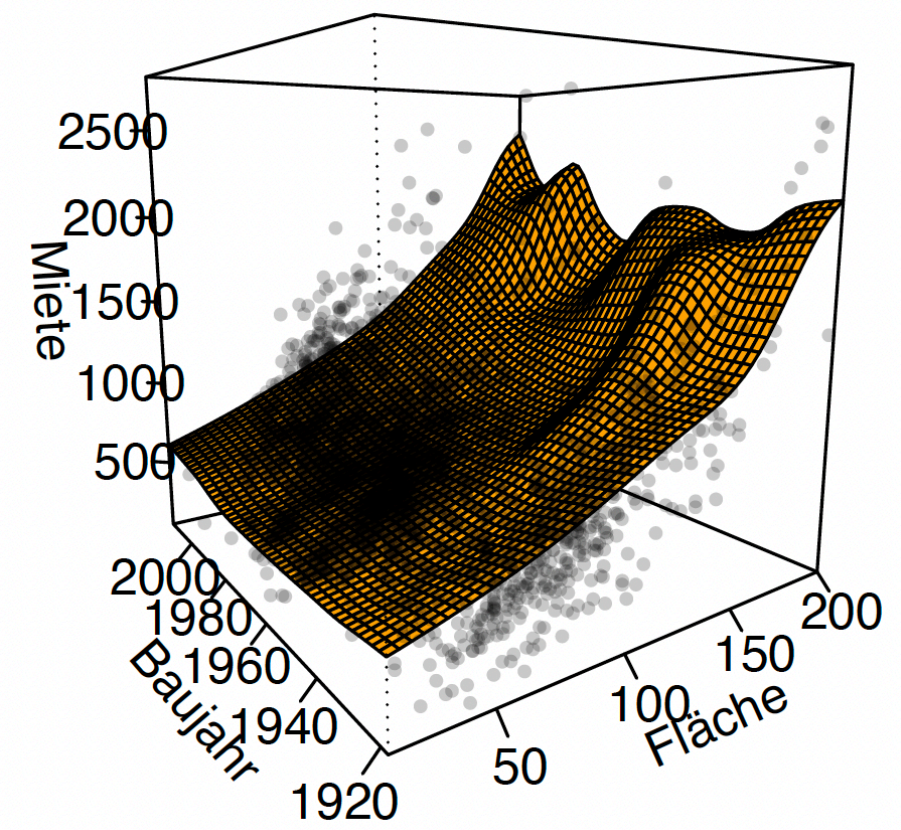
- Logistische Regression (für binäre Daten)



# Session 6

*13. Mai — Timo Adam*

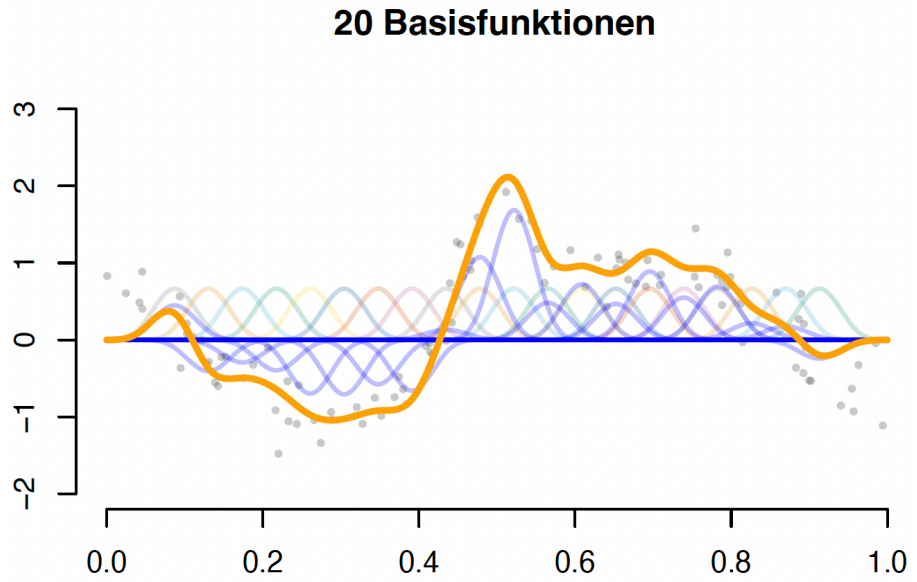
- Nicht-parametrische Regression (für nicht-lineare Zusammenhänge)
- Nadaraya-Watson



# Session 7

*20. Mai — Timo Adam*

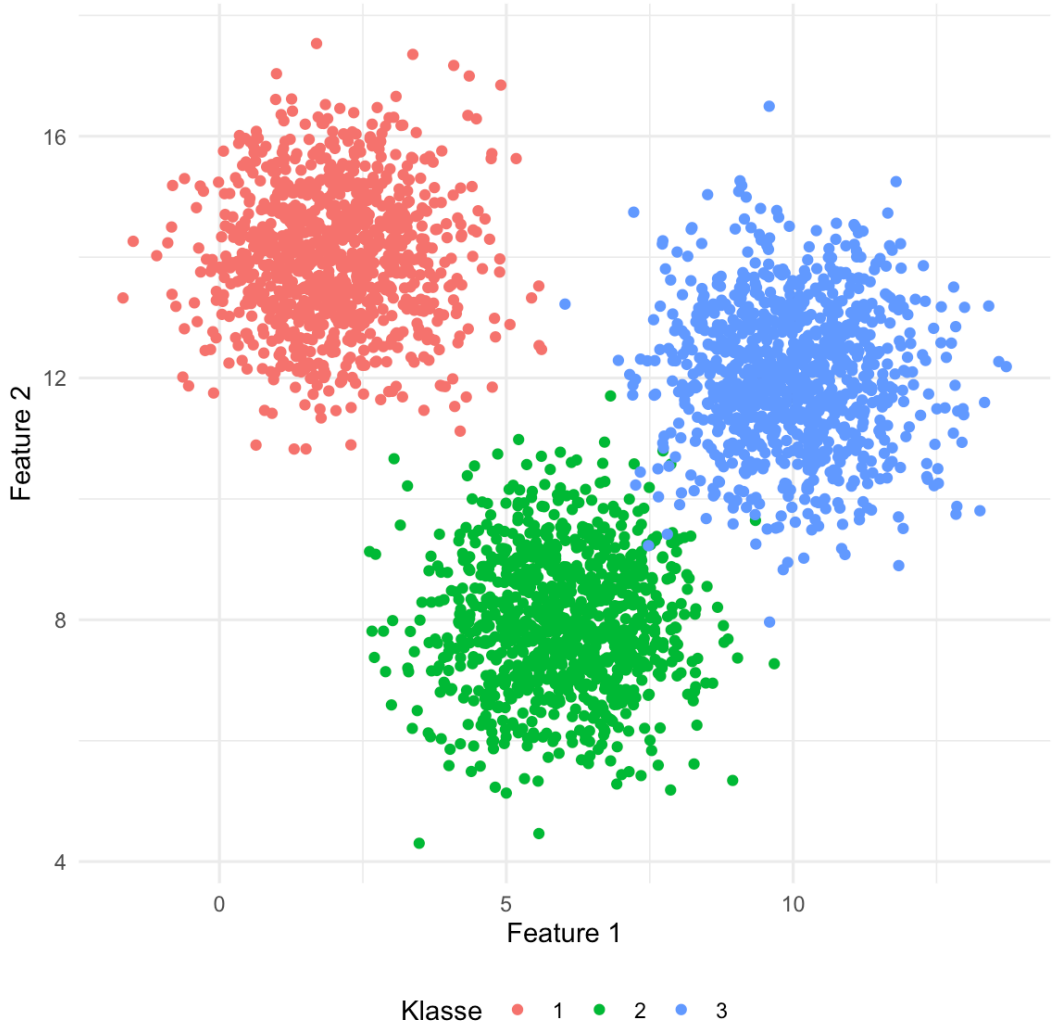
- Nicht-parametrische Regression (für nicht-lineare Zusammenhänge)
- Splines



# Session 8

*27. Mai — Julian Hinz*

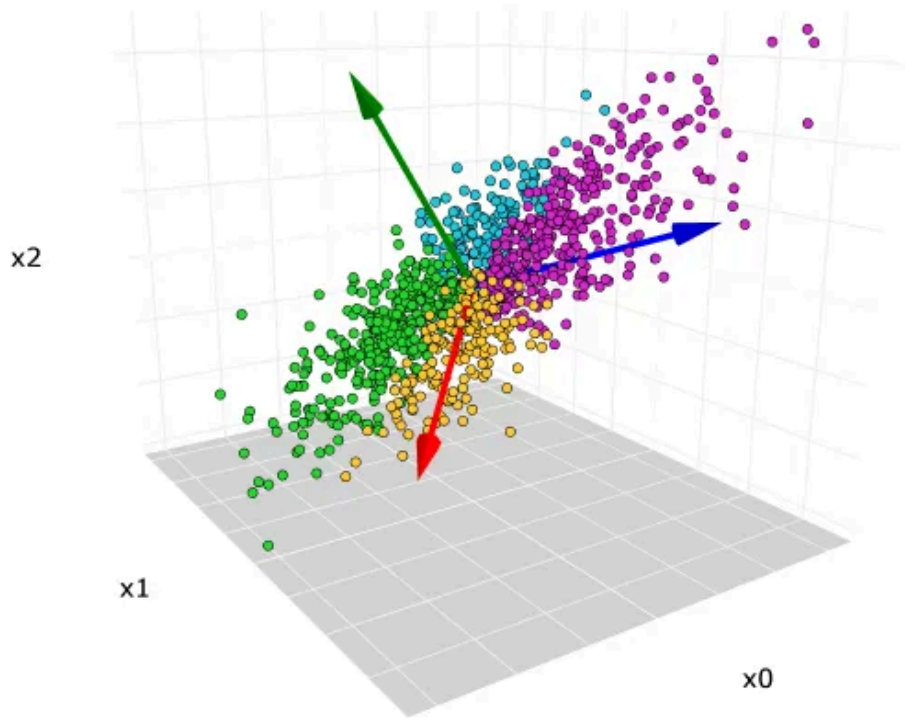
- Klassifikation



# Session 9

*3. Juni — Julian Hinz*

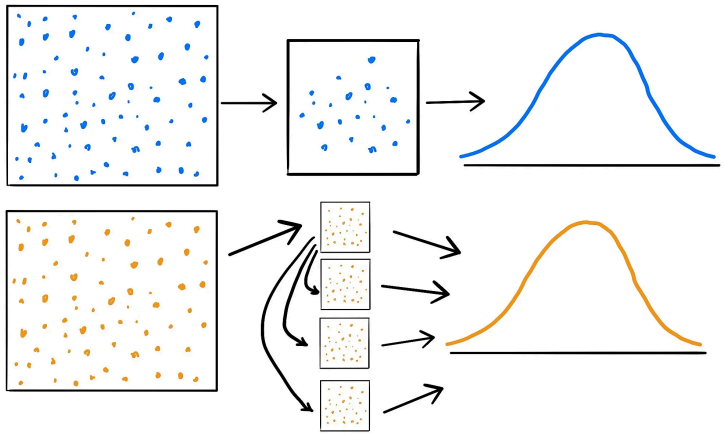
- Dimensionsreduktion und Clustering



# Session 10

*10. Juni — Julian Hinz*

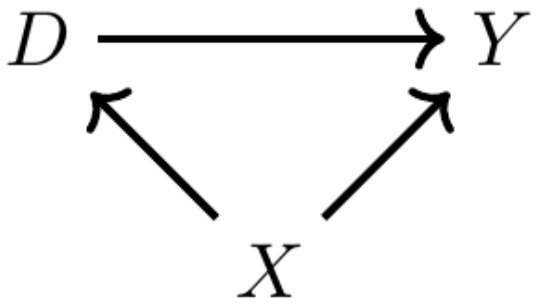
- Bootstrapping

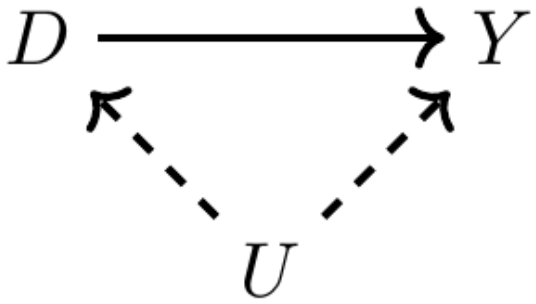


# Session 11

*17. Juni — Aaron Lohmann*

- Kausale Inferenz

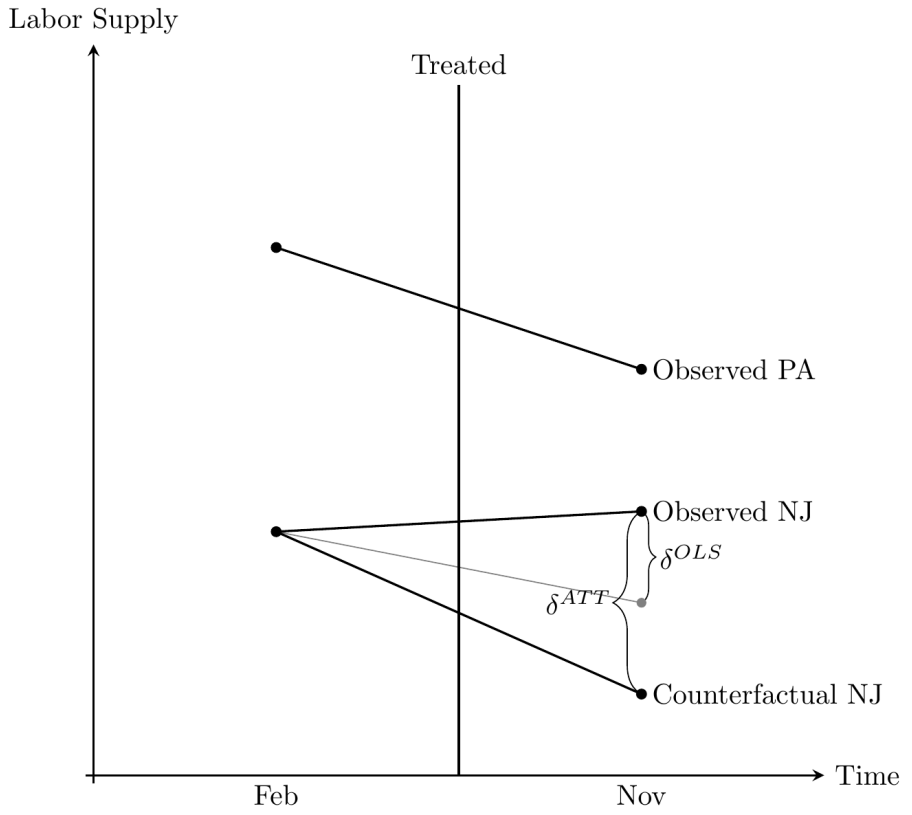


# Session 12

*24. Juni — Aaron Lohmann*

- Differences in Differences



# Session 13

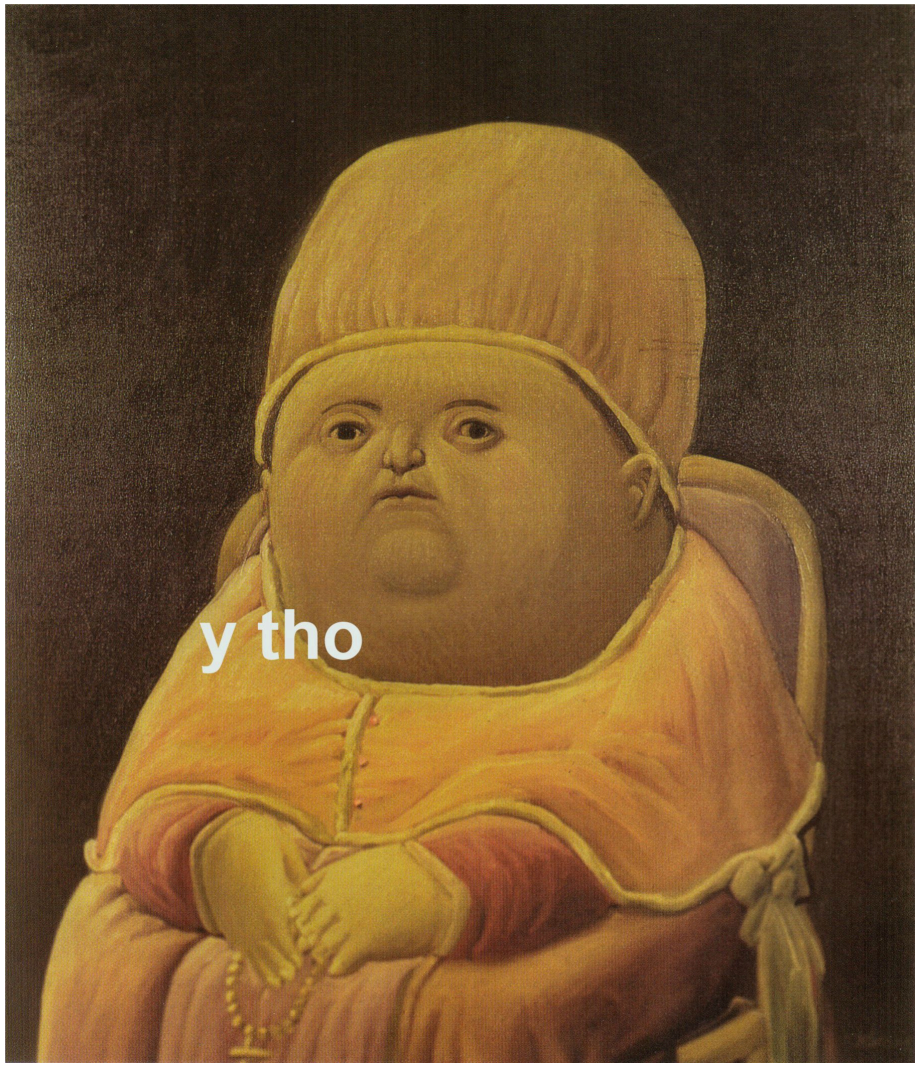
*1. Juli — Julian Hinz*

- LLMs

# Session 14

*8. Juli — Julian Hinz, Timo Adam*

- Wrap-up und Q & A

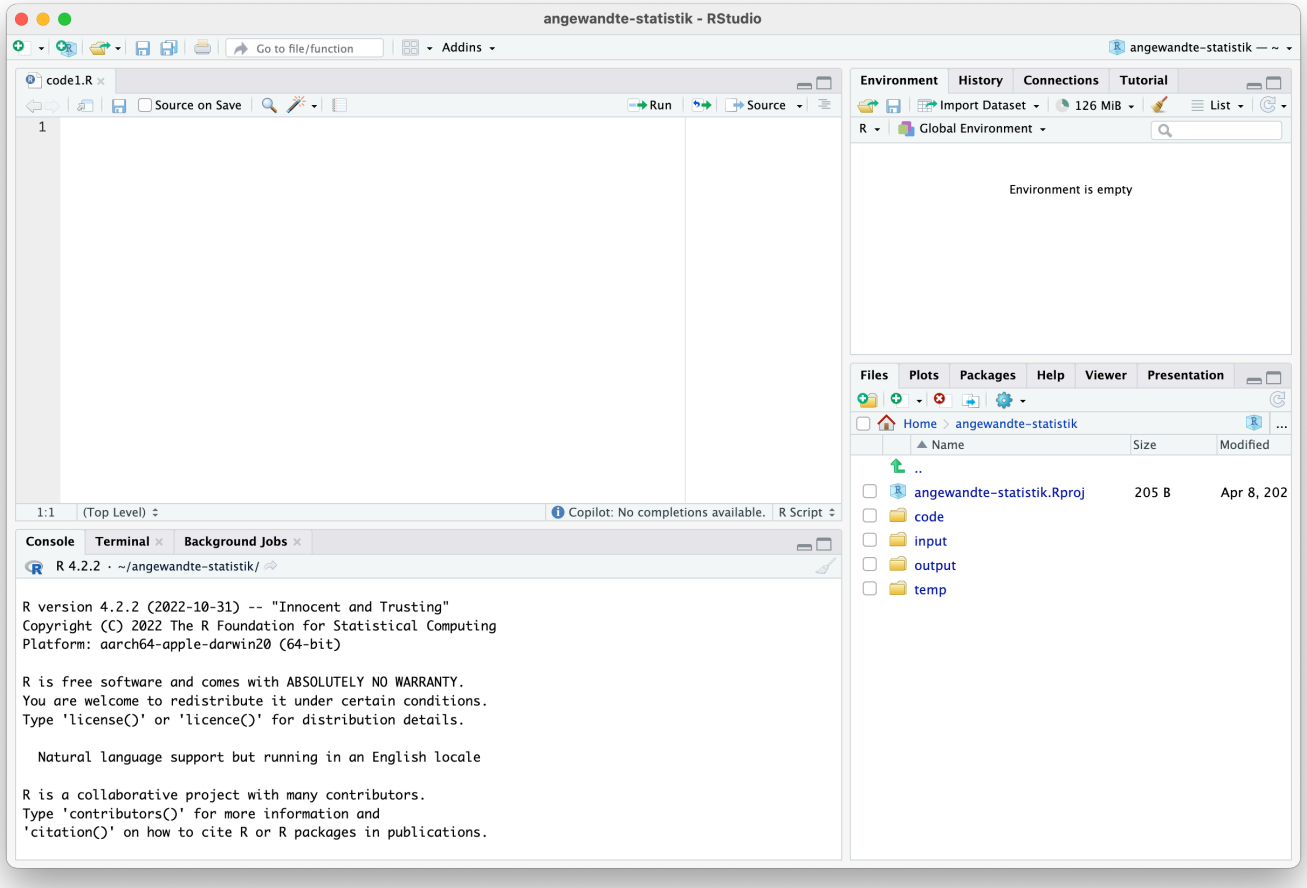


R

# R

- Angewandte Statistik ist ohne Programmierung undenkbar
- Statistiksoftware R: sehr mächtig, kostenlos, und open source → ideal!
- wir wissen: R Kenntnisse sehr heterogen über Hörer*innenschaft verteilt
  → Tutorien entsprechend gestaltet

# Installation von R und RStudio



# Installation von R und RStudio

## Schritt 1: R installieren

- Mac (Apple Silicon): https://cran.rstudio.com/bin/macosx/big-sur-arm64/base/R-4.4.3-arm64.pkg
- Mac (Intel): https://cran.rstudio.com/bin/macosx/big-sur-x86_64/base/R-4.4.3-x86_64.pkg
- Windows: https://cran.rstudio.com/bin/windows/base/R-4.4.3-win.exe
- Linux: `sudo apt install r-base r-base-dev`

# Installation von R und RStudio

## Schritt 2: RStudio installieren

- Mac: https://download1.rstudio.org/electron/macos/RStudio-2024.12.1-563.dmg
- Windows: https://download1.rstudio.org/electron/windows/RStudio-2024.12.1-563.exe
- Linux: …

# Fragen? → Element.