



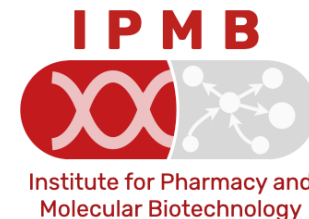
# Data Analysis Projects

## MoBi 4. FS - SoSe 2024

17.04.2024



GERMAN  
CANCER RESEARCH CENTER  
IN THE HELMHOLTZ ASSOCIATION



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

# Zeitplan

- Allgemeine Vorstellung des Moduls (ca. 15 Minuten)
- Vorstellung der 5 Themen (ca. 10 Minuten/Thema)
- Kurze Einführung in GitHub and (R)markdown

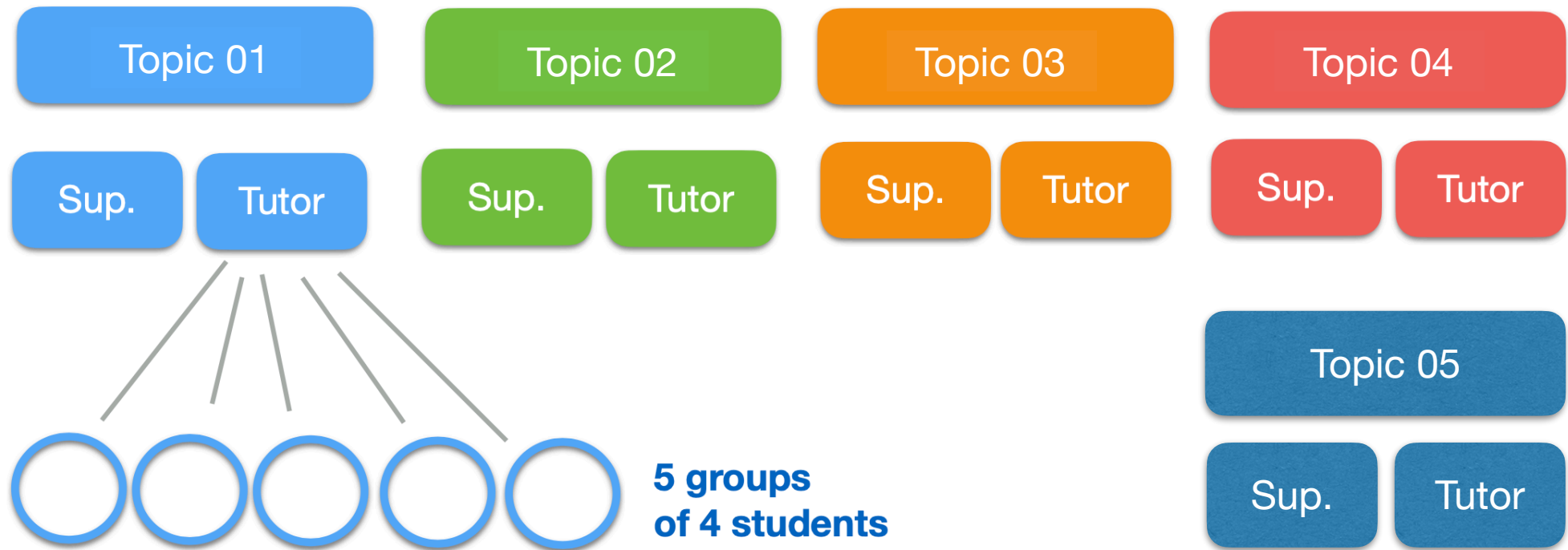
- Fortsetzung der **Vorlesung Datenanalyse** aus dem 3. FS
- **Projekt-orientiertes Lernen:** Erfahrung in der konkreten programmatischen Umsetzung der Methoden der Datenanalyse
- **Ziele**
  - Erfahrung in der **Anwendung der Methoden der Datenanalyse** anhand von reellen Datensätzen und wissenschaftlichen Fragen!
  - Erfahrung in der **Teamarbeit**
  - Erlernen des **Umgangs mit modernen Werkzeugen** der Datenanalyse (R / Python / Markdown / GitHub / ...)

***Learning by doing***



# Themen / Projekte

- **5 Forschungsthemen**
- Für jedes Thema gibt es bis zu 5 Sub-Projekte
- Jedes Projekt wird durch ein Team von 4 Studierenden bearbeitet
- **Pro Thema: wissenschaftliche Betreuer und Tutor/Tutorin**
- **Aufgabe der TutorInnen:**
  - wöchentliche Treffen mit den Gruppen die an den Projekten eines Themas arbeiten  
(z.B. Mittwochs 10-13h)
- Supervisors and **TutorInnen:**
  1. Biomedical image analysis - python (K. Rohr, L. Kostrykin) + **Leonie Thomas**
  2. Drug screens/Cancer Hallmarks - R (C. Herrmann) + **Anna von Bachmann**
  3. Proteome screens of RNA-binding proteins - R (M. Caudron-Herger) + **Malte Hermes**
  4. Protein thermostability - python (D. Niopek, B. Wolf) + **Maximilian Fidlin**
  5. DNA methylation - R (M. Scherer) + **Stefanie Mantz**



R Markdown

from R Studio

# Themen / Projekte

- **Topic 01: *Biomedical Image Analysis***  
(Karl Rohr / Leonid Kostykin; Tutor: Leonie Thomas)
  - Data types: MNIST images / cell nuclei images
- **Topic 02: *Cancer Hallmark and Metabolic Pathways in Cancer***  
(Carl Herrmann; Tutor: Anna von Bachmann)
  - Data types: Genetic data
- **Topic 03: *Proteome Screen of RNA-Binding Proteins***  
(Maiwen Caudron-Herger; Tutor: Malte Hermes)
  - Data types: mass spectrometry data
- **Topic 04: *Protein Thermostability***  
(Dominik Niopek / Benedict Wolf; Tutor: Maximilian Fidlin)
  - Data types: mutation data
- **Topic 05: *DNA Methylation***  
(Michael Scherer; Tutor: Stefanie Mantz)
  - Data types: genetic data

# Zeitplan - Modul

*Was wir machen...*

*Was ihr macht!*

17/04

Vorstellung der Projekte  
Intro zu GitHub

19/04

Auswahl der Projekte  
und Teams (Google)

24/04

Plenum lineare Regression

15/05

Präsentation des  
project proposal  
(10 + 10 min)

08/07 (8pm)

*Schließung der GitHub repos*

10/07

Finale Präsentation und  
Bericht  
(15+10 minutes)





# Project proposal (15/05)

- In der Präsentation des **project proposals** solltet ihr...
  - einige des angegebenen Referenzen im Kontext des Projektes vorstellen
  - Die allgemeine Fragestellung / Herausforderungen erklären
  - Die Daten beschreiben
  - Die Ziele eures Projektes definieren
  - Ungefährer Zeitplan
    - ▶ milestones = important steps in the analysis
    - ▶ when these milestones should be achieved
- Mündliche Präsentation vor Betreuer / Tutoren
  - 10 Minuten Präsentation
  - 10 Minuten Diskussion / Fragen
- ***Alle Mitglieder des Teams sollten aktiv beitragen!***

# Finale Präsentation und Bericht

- **Finale Präsentation am 10.07**
  - 15 Minuten Präsentation
  - 10 Minuten Fragen
  - Vorstellung der wichtigsten Ergebnisse
- **Bericht (auf Englisch!)**
  - pdf Bericht (im GitHub repo)
  - **10 Seiten max.**
  - Aufbau: Introduction / Material and methods / Results/ Discussion
  - Wichtig: sorgfältige Auswahl der Plots; Beschriftung der Plots wichtig!
- **Github repo**
  - Repo vor deadline aufräumen !!
  - Bitte ein klares README erstellen

***GitHub repo schließt am Montag 08.07 um 20h  
Bericht (PDF) sollte bis dahin fertig sein!***

# Benotung

- Project proposal = 30%
- Finale Präsentation = 30%
- Bericht = 40%
- **Kriterien**
  - Qualität der Einführung in die wissenschaftliche Frage (insbesondere Vorstellung der Literatur)
  - Klare Definition der Forschungsfragen
  - Qualität der Plots und Beschriftungen
  - Qualität der Besprechung der Ergebnisse

# Themen / Projekte

- **Topic 01: *Biomedical Image Analysis***  
(Karl Rohr / Leonid Kostrykin; Tutor: Leonie Thomas)
  - Data types: MNIST images / cell nuclei images
- **Topic 02: *Cancer Hallmark and Metabolic Pathways in Cancer***  
(Carl Herrmann; Tutor: Anna von Bachmann)
  - Data types: Genetic data
- **Topic 03: *Proteome screen of RNA-binding proteins***  
(Maiwen Caudron-Herger; Tutor: Malte Hermes)
  - Data types: mass spectrometry data
- **Topic 04: *Protein thermostability***  
(Dominik Niopek / Benedict Wolf; Tutor: Maximilian Fidlin)
  - Data types: mutation data
- **Topic 05: *DNA Methylation***  
(Michael Scherer; Tutor: Stefanie Mantz)
  - Data types: genetic data