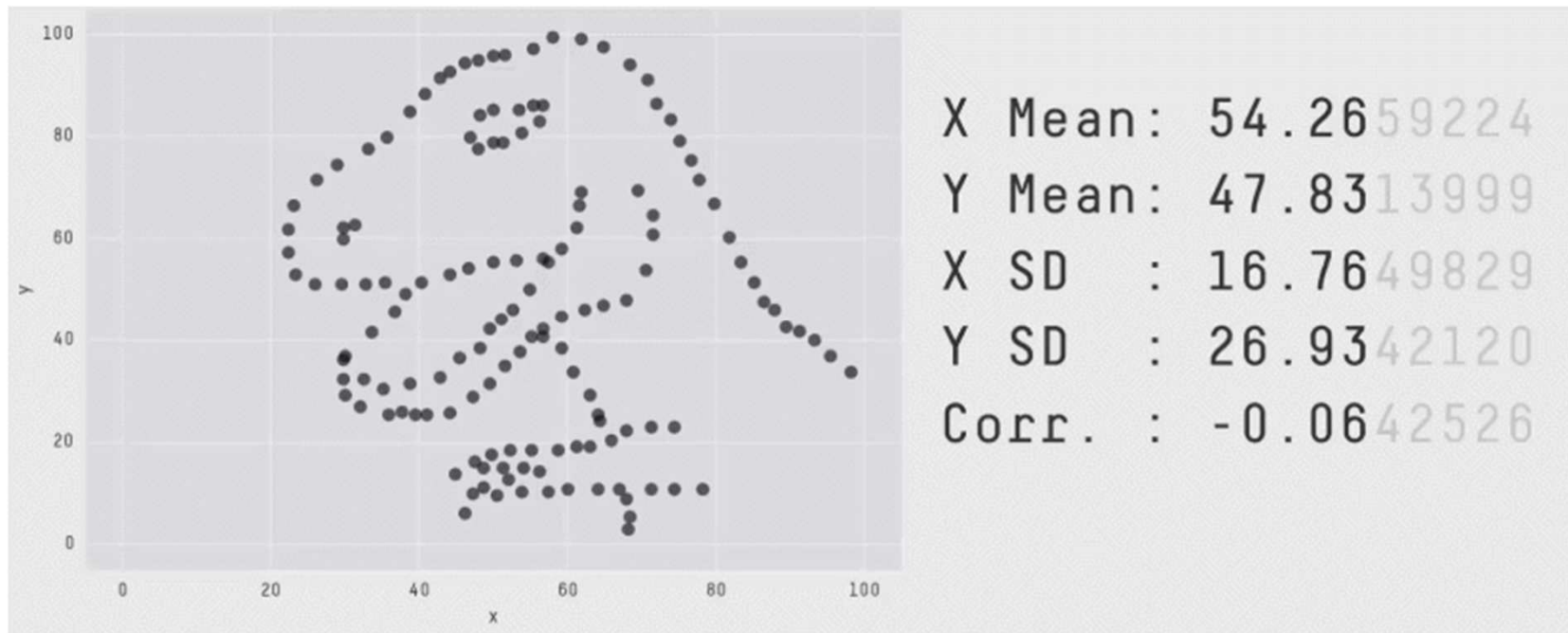

Seminarbaustein: Data Science mit

R ist das am häufigsten genutzte Werkzeug zur statistischen Datenanalyse und zur grafischen Darstellung der Ergebnisse. Es ist kostenfrei, flexibel einsetzbar und wird von einer großen Community ständig weiter entwickelt. Im Seminar erlernen Sie den Umgang damit und erhalten hilfreiche Unterstützung zur Anwendung.

Hier eine kurze Einführung.

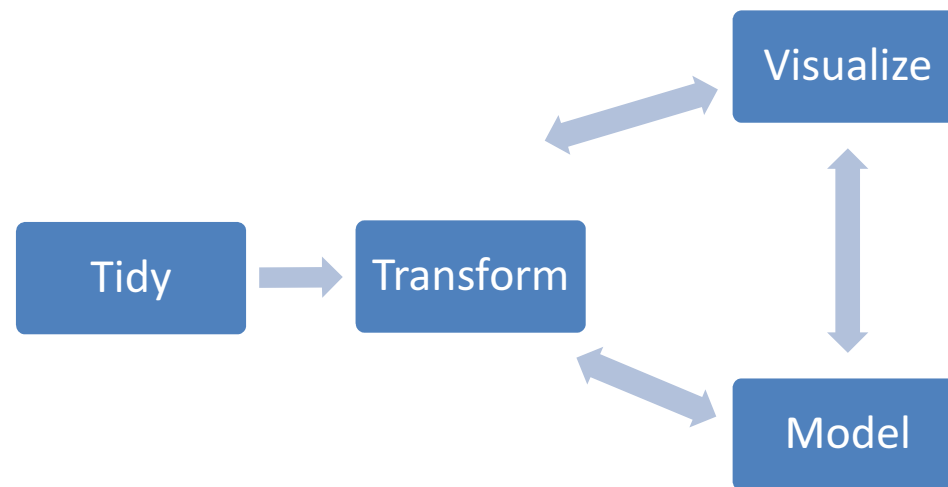
Dozent: Dr. Andreas Busjahn

Was versteckt sich in den Daten? (?Dinosaurier?)



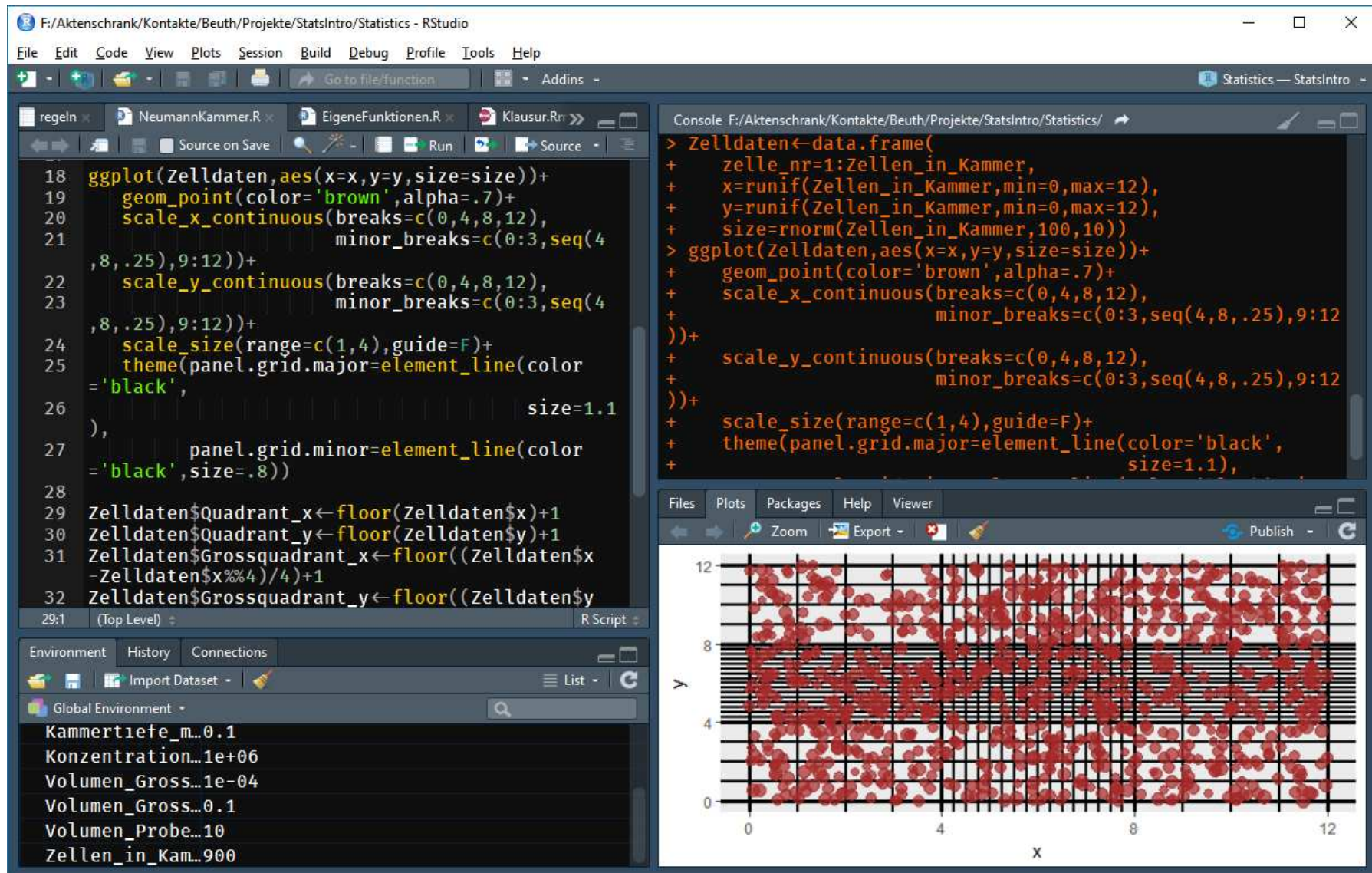
- Skriptsprache für statistische Analysen und Grafik
- Multiversum universeller und hochspezialisierter Lösungen
- Freie Software mit 13.741 Erweiterungspaketen
- Unterstützt durch IBM, Microsoft, Google, Oracle...
- Bioconductor: Teilprojekt für Bioinformatik, ...omics
- Grammar of Graphics: Flexibles Grafiksystem

- Import (1 bis 50 Zeilen)
- Datenvorbereitung (50 bis 1000 Zeilen)
- Visualisierungen (10 bis 100 Zeilen)
- Tests / Modelle (ca. 100 Zeilen, Tests: 1 bis 5)



Graphische Benutzeroberfläche RStudio

5



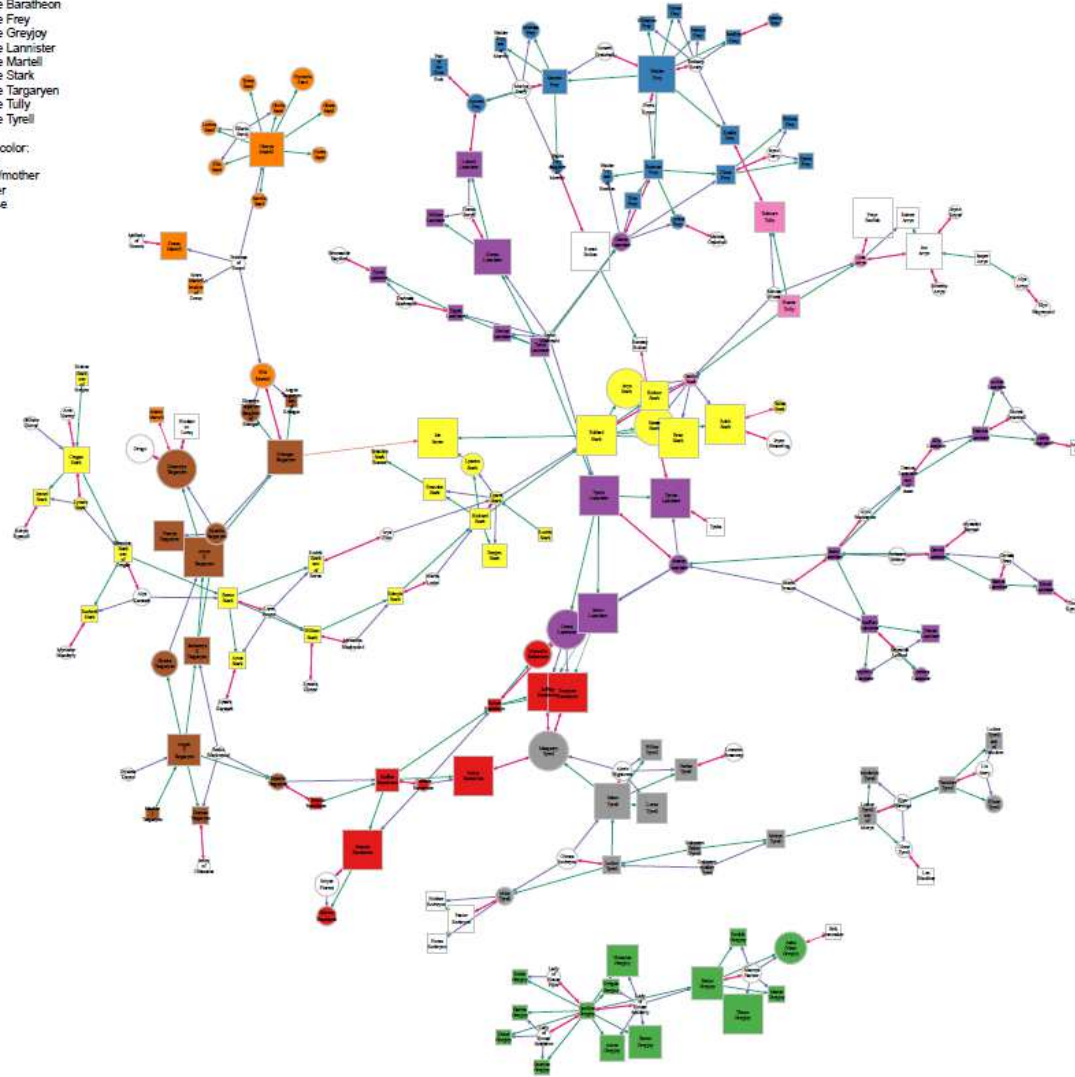
The screenshot displays the RStudio environment with the following components:

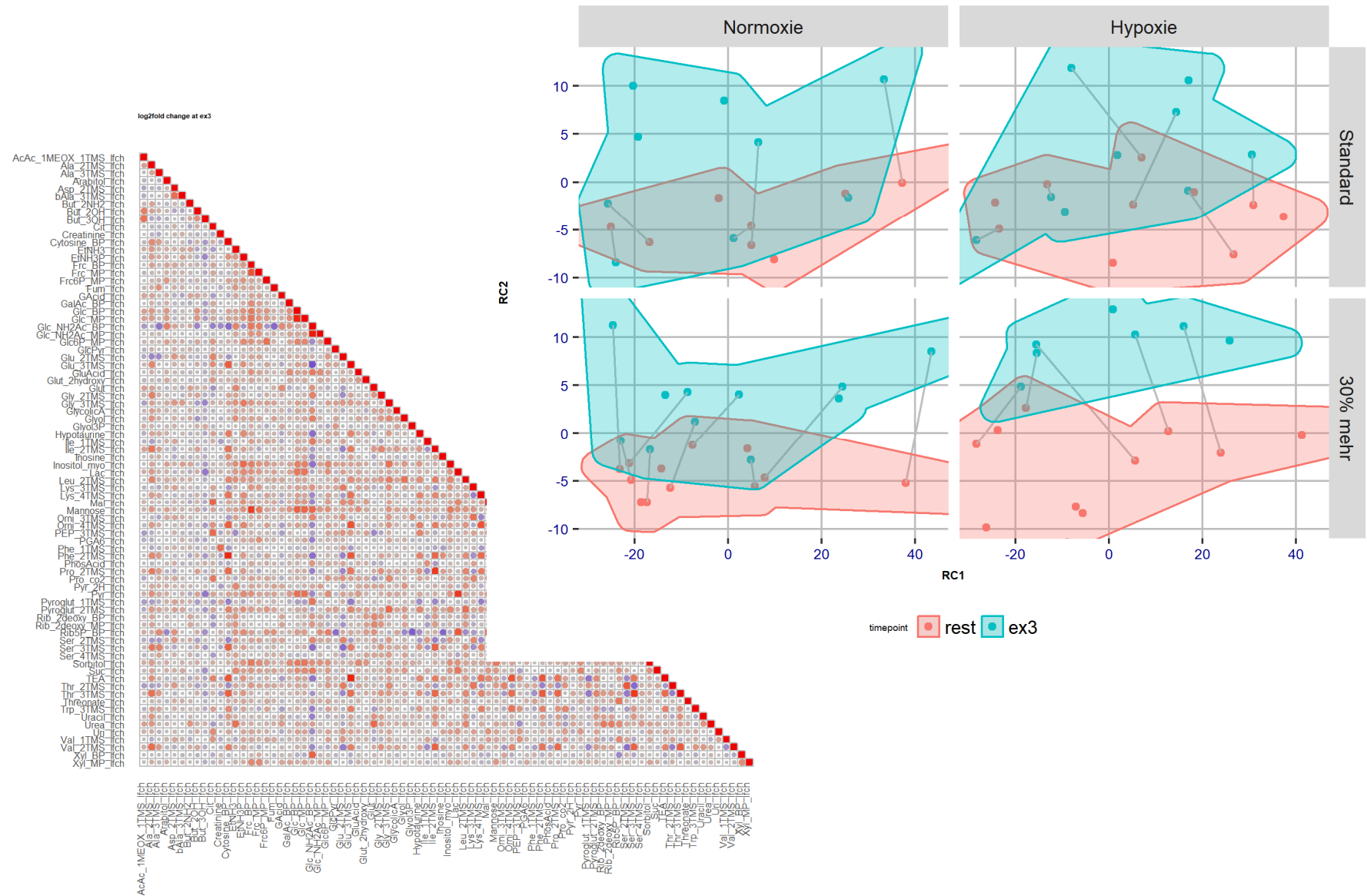
- Source Editor:** Contains R code for creating a data frame and plotting it. The code uses `geom_point` with a brown color and alpha of 0.7, and `scale_x_continuous` and `scale_y_continuous` to define a grid with major and minor breaks. It also includes `scale_size` and `theme` for styling the plot.
- Console:** Shows the execution of the R code, resulting in a scatter plot.
- Environment:** Lists variables in the global environment, including `Kammertiefe_m...0.1`, `Konzentration...1e+06`, `Volumen_Gross...1e-04`, `Volumen_Gross...0.1`, `Volumen_Probe...10`, and `Zellen_in_Kam...900`.
- Plots:** Displays a scatter plot of red points on a grid. The x-axis is labeled 'x' and the y-axis is labeled 'y', both ranging from 0 to 12. The points are distributed across the grid, with a higher density in the center.

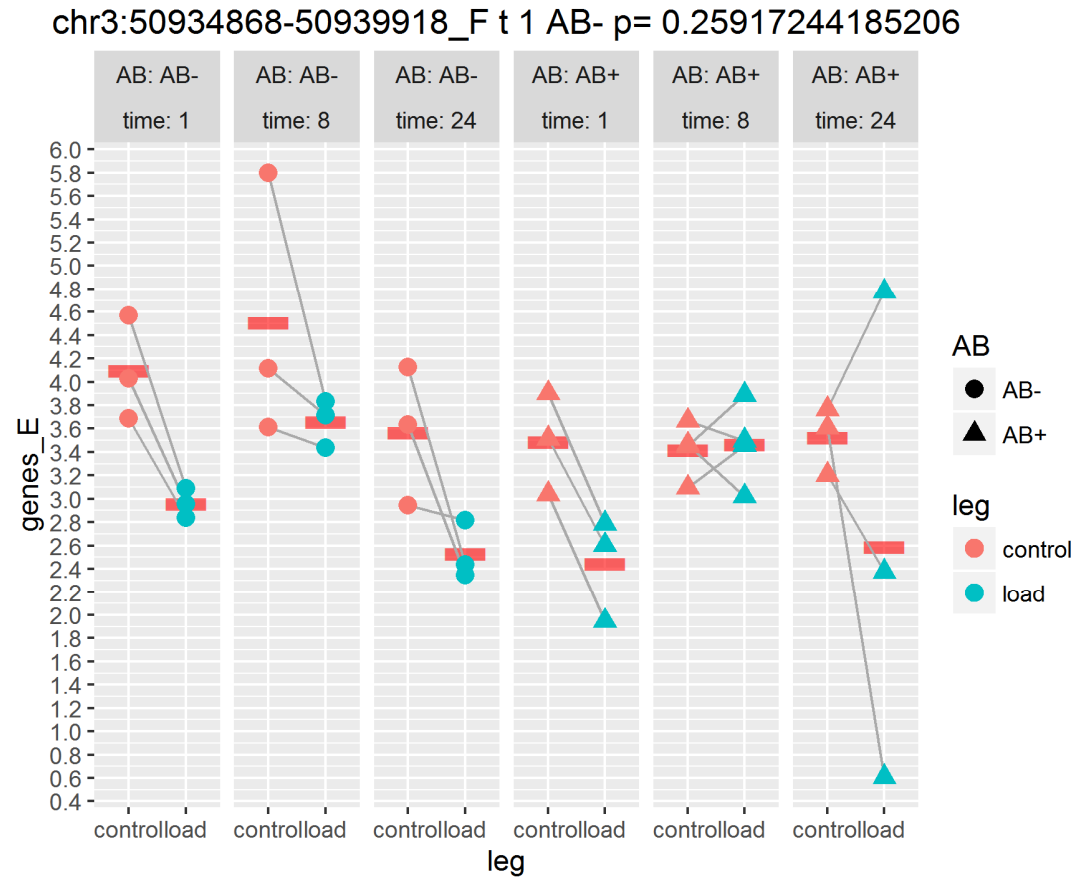
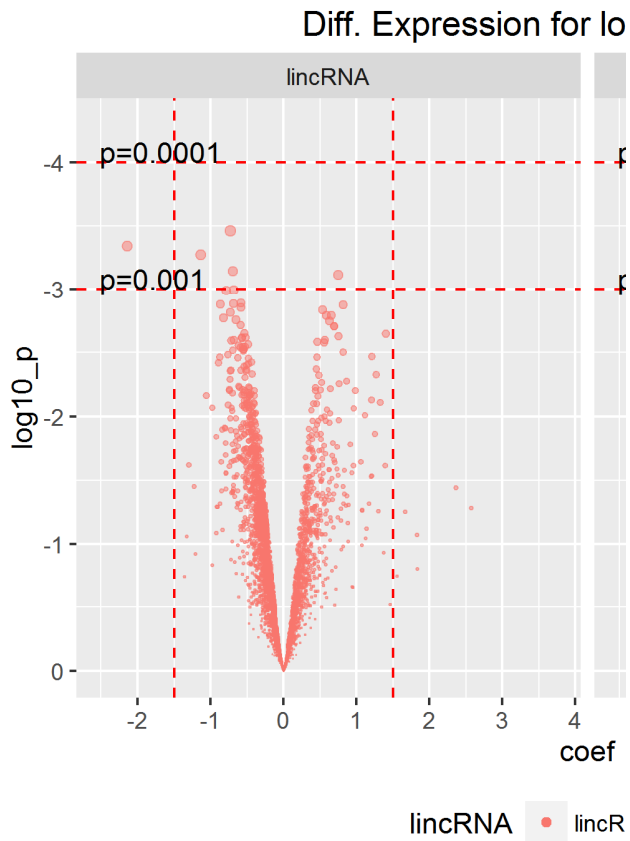
... Verlangen komplexe Analysen

Game of Thrones Family Ties


- Node color:
 - House Baratheon
 - House Frey
 - House Greyjoy
 - House Lannister
 - House Martell
 - House Stark
 - House Targaryen
 - House Tully
 - House Tyrell
- Edge color:
 - father
 - father/mother
 - mother
 - spouse







- Jeder Schritt einer wissenschaftlichen Studie muss nachvollziehbar sein
- Angabe statistischer Methoden (PCA, OPLS...) nicht ausreichend, da hunderte Variationen
- KEIN Schritt der Analyse darf manuell erfolgen (copy/paste in Excel...)

- Lösung:
 - ▶ Scriptbasierte Analysen mit 
 - ▶ R Markdown als Reportgenerator: Text, Tabellen und Abbildungen -> LaTeX / Worddokument