

---

**Vorlesung: Programmieren in Anwendungen**  
**Übungsblatt 5**

[http://softech.informatik.uni-kl.de/Homepage/PIA\\_SS15](http://softech.informatik.uni-kl.de/Homepage/PIA_SS15)

---

## Informationen zu R und Installation von RStudio

- Zur Bearbeitung der Übungen benötigen Sie die Software RStudio. Auf den Rechnern des SCI ist sie bereits installiert.  
Zur Installation auf Ihrem privaten Rechner / Laptop gehen Sie wie folgt vor:
  1. Laden Sie die neueste Version für Ihr Betriebssystem herunter  
<http://www.rstudio.com/products/rstudio/download/>  
und installieren Sie diese.
  2. Installieren Sie die Packages `plyr`, `reshape`, `ggplot2` und `knitr` inklusive aller abhängigen Pakete. Dies geht am einfachsten auf der R-Console mit dem Befehl  
`install.packages("plyr")` etc.
- Auf der Website <http://www.r-project.org> erhalten Sie Information und Material zur Statistiksoftware R.

## Movies!

Das `ggplot2`-Packet beinhaltet auch einen Datensatz zu Filmen, die auf der Webseite der Internet Movie Database, `IMDB.com`, bewertet wurden.

Aus `help(movies)`:

The internet movie database, <http://imdb.com/>, is a website devoted to collecting movie data supplied by studios and fans. It claims to be the biggest movie database on the web and is run by amazon. More about information `imdb.com` can be found online, [http://imdb.com/help/show\\_leaf?about](http://imdb.com/help/show_leaf?about), including information about the data collection process, [http://imdb.com/help/show\\_leaf?infosource](http://imdb.com/help/show_leaf?infosource).

1. Machen Sie sich mit der Struktur des Datensatzes bekannt! Was beinhalten die einzelnen Spalten? Wie viele Einträge gibt es?
2. Wie viele Star Wars Movies sind in dem Datensatz enthalten?
3. Wie heisst der längste Kurzfilm?
4. Was sind die TOP 10 Filme mit dem höchsten Rating?
5. Fügen Sie dem Datensatz einen Faktor hinzu, der das Filmgenre kodiert. Geben Sie eine Übersicht zur Verteilung der Filmgenre.
6. Visualisieren Sie mit `qplot()` oder `ggplot()` weitere Informationen, z.B. Entwicklung des Budgets im Laufe der Jahre, Altersbeschränkung pro Genre, Rating vs. Länge.

*Hinweis:* Filtern Sie dazu die Einträge heraus, deren Wert nicht angegeben ist (NA), um Warnungen zu unterdrücken.

## Sportangebot der TU Kaiserslautern

Im Rahmen des Seminars "Forschungslogik und Forschungsdesign" wurde im Wintersemester 2014/2015 eine Befragung zur Zufriedenheit der Studierenden bezüglich der Auswahl an Kursen des Sportangebotes an der TU durchgeführt.

Der erhobene Datensatz zu dieser Befragung wird Ihnen zusammen mit dem Fragebogen bereits per Email zugeschickt.

1. In welchem Format liegt der Datensatz vor? Lesen Sie den Datensatz entsprechend in einen Dataframe in R ein. Die letzten 12 Spalten sollten dabei entfernt werden, da sie keine Daten enthalten.
2. Wieviele Einträge hat der Datensatz und welche Struktur?
3. Benennen Sie die Spalten sinnvoll unter Zuhilfenahme des Fragebogens.
4. Transformieren Sie das Geschlecht der Befragten in einen Faktor und die Semesterzahl in einen Zahlenwert.
5. Geben Sie eine tabellarische und eine geeignete grafische Übersicht zum Geschlecht, Semester und zur Anfahrtzeit der Befragten.
6. Wie zufrieden sind die Befragten in Bezug auf die Trainer, die Breite des Angebots und die Ausfallhäufigkeit von Kursen?
7. Gibt es einen Zusammenhang zwischen den geplanten Besuchen von Kursen und der Zufriedenheit bzgl. der Kursauswahl?