

---

# Empirisch-experimentelle Forschungsmethoden in der Anwendung

Seminar

# Themen heute

---

- Fragen aus der Vorlesung
- Pitch zur Hausaufgabe
- Recherche
- Arbeiten mit R Studio

# Pitch zur Hausaufgabe

---

- Inhalt der Literatur?
- Forschungsfrage?
- max. 120 Sekunde pro Gruppe (hartes Limit)

---

# Recherche

# Was sind „Publikationen“

---

- Formen wissenschaftlicher Veröffentlichungen
  - Journal Artikel, Review-Artikel, Editorials, Konferenz-Beiträge, Poster, Monographien, Book-Chapter, Sammelbände, Patente, White-Paper, Industrie-Zeitschriften, Dissertationen, Master-Arbeiten, Gebäude, Kunstwerke
  - Zitierbar nach Thompson Reuters: Article, Review, Proceedings Paper
    - Innerhalb des Web of Science
- Disziplinäre Unterschiede in Bezug auf
  - Was zählt als „Publikation“
  - Welche Bedeutung haben Eigenschaften wie Outlet, Länge, Referenzen, Zitations-Halbwertszeit, # Ko-Autoren
  - Größe von Community, Disziplinen-Kultur, „Härte“ einer Disziplin

# Beispiel: Psychologie

---

- Ziel: High Impact Journals (>90% Ablehnungsquote)
  - Alternativ: Buch Kapitel
  - Kaum Monographien (Ausnahme Dissertation)
- Trend: kürzere Artikel, häufigere Publikationen
- 60% Internationale Outlets
- Erstautor: Haupt-Idee, -Beitrag; sonst alphabetisch oder nach Beitrag
- Inhalt: Schwerpunkt Methodologie, Daten-Analyse

# Beispiel: Teilchen-Physik

---

- Ziel: Physical Review, Journal of Physics
  - Alternativ: Elektronisch Open-Access, JINST, JHEP
  - Kein Nature/Science (verbieten arXiv -> <http://www.sherpa.ac.uk/romeo/>)
- Trend: Viele Autoren (>1000)
- Hauptsächlich regionale internationale Outlets
- Autoren in alphabetischer Reihenfolge
- Inhalt: Daten-Analyse

# Unterschiede bzgl. Autorenschaft

- **Biologie** viele Autoren (>5), kurze Artikel (11-15p), 14.72 avg. Zitationen, Halbwertszeit (HWZ): 6y
- **Physik** weniger Autoren(2-3), 7.71 avg. Zitationen, HWZ: 5.2y
- **Soziologie** hauptsächlich Alleinautoren, längere Artikel (41-45p), 3.55 avg. Zitationen, HWZ >10y

Anzahl Autoren		1	2	3	4	5	mehr als 5	Total
Cell 2006	Anzahl Artikel	21	58	39	38	36	177	369
	in % aller Artikel	6 %	16 %	11 %	10 %	10 %	48 %	100 %
Physical Review D 2005	Anzahl Artikel	400	869	592	290	93	212	2456
	in % aller Artikel	16 %	35 %	24 %	12 %	4 %	9 %	100 %
American Journal of Sociology 2006	Anzahl Artikel	19	11	5	2	0	0	37
	in % aller Artikel	51 %	30 %	14 %	5 %	0 %	0 %	100 %



# Unterschiede innerhalb von Disziplinen

- Abhängig von Community-Größe, Publikationshäufigkeit, Zitationshäufigkeit

physics subdisciplines as defined by ISI	# papers (2000)	# citations (2000-2006)	# citations per paper
Physics, Applied	24367	171611	7.04
Physics, Atomic, Mol. & Chem.	11097	111893	10.08
Physics, Condensed Matter	21332	155676	7.30
Physics, Fluids & Plasmas	4764	43507	9.13
Physics, Mathematical	6828	47913	7.02
Physics, Nuclear	5799	46190	7.97
Physics, Particles & Fields	7693	81216	10.56
<i>Physics, All Categories</i>			9.05*
* ESI (1996-2006) Average Citation Rates - Datum der Suche: 07.04.06			

# Problem für Quantitative Publikations-Analyse

---

- Qualität der Daten hängt von Datenbank-Coverage ab
  - (für Engineering z.B. ist Web of Science ca 40%)
- Interpretation von Zitationsdaten ist Disziplinspezifisch und Community-abhängig
- Individuelle Einflüsse, Arbeitsgruppenzusammensetzung, Wichtigkeit für Institut
- Kleine Stichproben vergrößern den Fehler

# Fragen

---

?

# Quantitative Publikationsanalyse – Äpfel und Birnen

---

- Was messen?
  - Wie viel wird publiziert/zitiert? Wie wird normiert (Disziplin/Autoren/Alter..) ?
- Welche Datenquelle?
  - Scopus/WOS/Scholar/BTH/Research Gate/etc.
- Vergleichbarkeit von Messwerten
  - Statistisch kaum relevant bei kleinen Stichproben (Power-Law-Distribution)
- In Bezug auf „was“ messen
  - Personen lassen sich schwer quantitativ erfassen (Stichprobenfehler)
  - Passiert leider doch häufig

# Typische Metriken (Journals)

---

- Impact-Factor (z.B. ISI-WOS)
  - Berechnung: Citations/Article (3year Window) within ESSI
  - Zweck: Bibliothekarisch „Was abonnieren?“
  - Eigenschaft eines Journals stark disziplinenabhängig
  - NICHT: Qualität einzelner Artikel, Qualität akzeptierter Autoren
    - streut PowerLawDistribution
  - Schlecht/gar nicht geeignet für Bewertung von Autoren
  - Eher Schlecht geeignet für Journal-Auswahl

# Typische Metriken (Personen)

---

- Hirsch-Index
  - Berechnung: Anzahl  $h$  Publikationen mit mind.  $h$  Zitationen
  - Zweck: Vergleichbarkeit von Autoren
  - Disziplinenabhängig (Koautoren, #Citations)
  - Nicht monotone Funktion bzgl. Parameter Publikationen/Zitationen
    - streut Power-Law-Distribution
  - Altersabhängig

# Was wollen wir wissen?

---

- Überblick über ein völlig neues Fachgebiet
  - Schneller Einstieg
  - Fundierte Wissensbasis
  - Ggfs. Review-Paper schreiben(?)
- Arbeitsschritte:
  - Terminologie-Sammlung (Onenote, Text-Datei etc.)
  - Suchen und Daten extrahieren
  - Daten säubern (Excel oder R)
  - Stalking 😊
  - Visualisieren mit draw.io

# Recherche Web of Science

Web of Science™ InCites® Journal Citation Reports® Essential Science Indicators™ EndNote®

WEB OF SCIENCE™

Search All Databases

Basic Search

Sheet metal forming

Topic

Search

Click here for tips to improve your search.

Timespan

All years

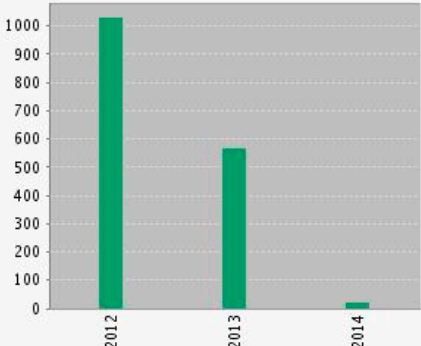
From 2012

Citation Report: 1,619

You searched for: TOPIC: (Sheet metal forming) ...More

This report reflects citations to source items indexed within All Databases.

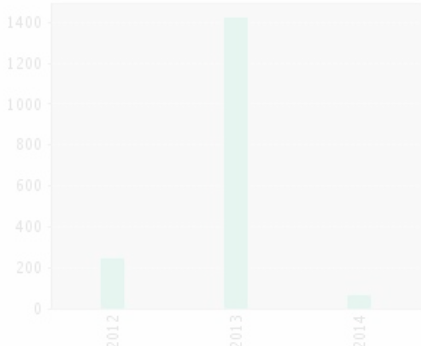
Published Items in Each Year



Year	Published Items
2012	1050
2013	550
2014	20

The latest 20 years are displayed.

Citations in Each Year



Year	Citations
2012	250
2013	1400
2014	100

The latest 20 years are displayed.

Results found: 1619

Sum of the Times Cited [?]: 1748

Sum of Times Cited without self-citations [?]: 1602

Citing Articles: [?]: 1606

Citing Articles without self-citations [?]: 1499

Average Citations per Item [?]: 1.08

h-index [?]: 14

- Literatursuche
  - Most recent vs most cited
  - Citation Report
- Journal Auswahl
- Citation Report für Institutionen



---

# Einführung in R und R Studio

# Einstieg in die Programmierung mit R

---

- Syntax, Semantik, Pragmatik, Guter Stil ?
- Welche Datentypen kennen Sie?
- R ist sehr einsteigerfreundlich...



...aber der Funktionsumfang ist mächtig und kann unübersichtlich wirken

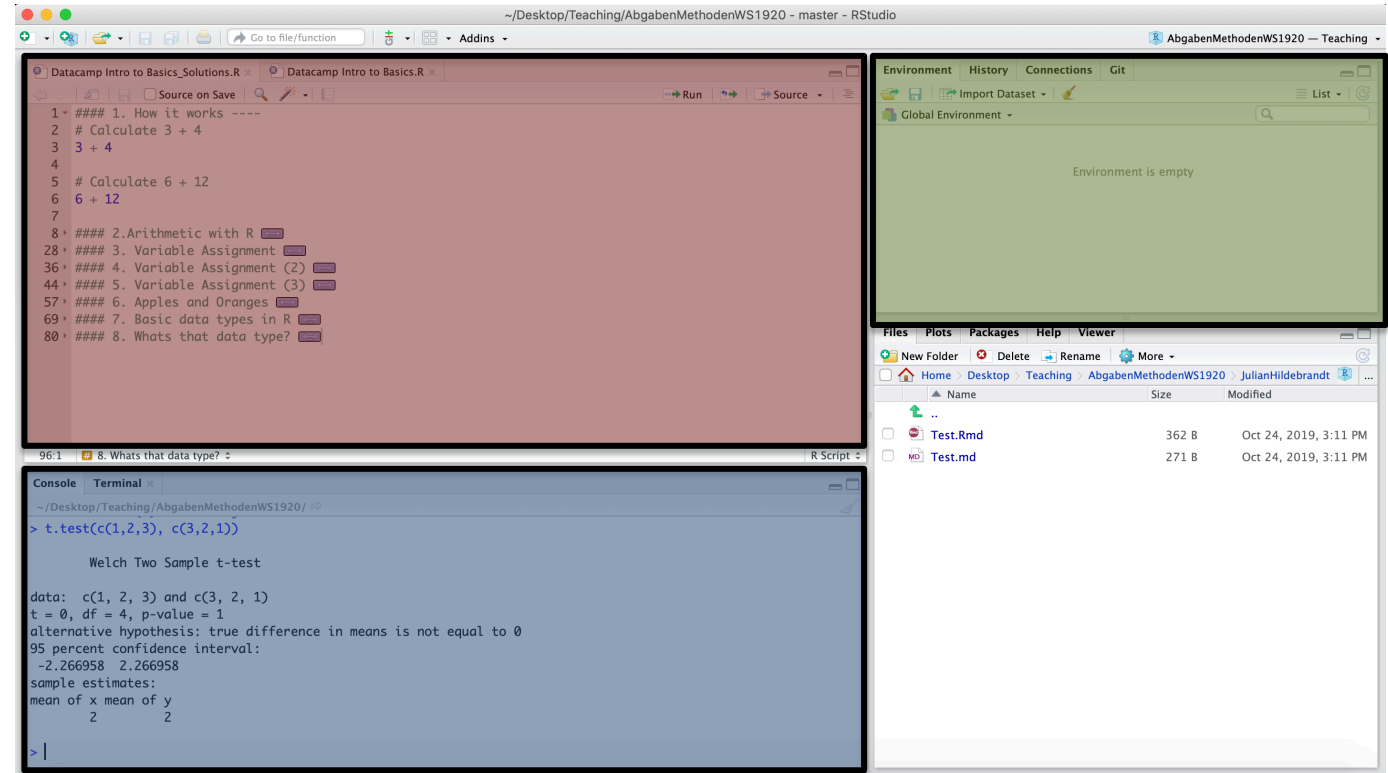
# 3 wesentliche Konzepte: Datentypen, Variablen und Funktionen

---

- Die für Sie wichtigsten atomare Datentypen in R
  - *numeric, character, logical*
  - + diverse komplexe Datentypen: *vector, matrix, list, data.frame, factor usw.*
- Variablenzuweisungen
  - Daten werden mit einem Namen versehen. In R mit *name <- value*  
`x <- 5`  
`name <- "Anna"`  
`muenzwurf <- TRUE`
- Funktionsaufrufe
  - Eine Kurzschreibweise für eine Programmsequenz. In R mit *function(arg1, arg2, etc.)*  
`class(x)`  
`mean(c(1, 2, 3))`  
`t.test(listA, listB)`

# R Studio

- Live Demo: **Environment**, **Texteditor** und **Console**



# Der nächste Termin

---

- Nächste Woche: 01.11.2018 Fragebogen (Hausaufgabe)
- Hausaufgaben über Slack
- Schönen Feiertag 😊

