

Wissenschaftliches Arbeiten mit R - Ausarbeitung des Lehr- Lernkonzepts

Sven Lesche¹

¹ Ruprecht-Karls-University Heidelberg

Author Note

Dieses Konzept wurde im Rahmen des Seminars: “Lehren lernen” erarbeitet und dient als Prüfungsleistung des Autors. Diese Arbeit wurde mit Hilfe von R-markdown verfasst, welches die Integration von Code und Text erlaubt. Alle Unterlagen zur exakten Replikation dieser Arbeit finden Sie unter: <https://github.com/SLesche/lehrkonzept>.

Wissenschaftliches Arbeiten mit R - Ausarbeitung des Lehr- Lernkonzepts

Struktur und Ziel des Seminars

Im jetzigen Lehrplan haben Studierende nur in ihren ersten beiden Studiensemestern Berührungspunkte mit R. Dann wird von ihnen allerdings erwartet, dass sie im Rahmen ihrer Abschlussarbeiten mit diesen begrenzten Ressourcen bereit sind Daten einzulesen, zu säubern, zu visualisieren, auszuwerten und zu interpretieren. Hierbei zeigt sich eine deutliche Lücke zwischen den Inhalten des Studiums und den Anforderungen der Abschlussarbeit. Dieses Seminar soll dazu dienen diese Lücke zu füllen und Studierenden die Möglichkeit zu geben einen vertieften Einblick in R zu gewinnen und praktische Anleitungen zur Wissenschaftlichen Arbeit mit R zu erhalten.

Hierzu lernen Studierende im Plenum und in Kleingruppen wöchentlich Grundlagen der Datenanalyse mit R kennen und üben anhand von praktischen Beispielen. Mit Hilfe dieser Kenntnisse sollen sie dann an einem Blocktermin einen vorgegebenen Datensatz in Kleingruppen analysieren und ihre Ergebnisse im Plenum präsentieren. Als Prüfungsleistung ist die Analyse und Präsentation eines eigenen Datensatzes vorgesehen. Die Abstimmung der Inhalte des Seminars auf das Prüfungsformat, sowie auf die Anforderungen von Abschlussarbeiten gewährleistet *constructive alignment* (Macke et al., 2016; Ulrich, 2020), eines der Voraussetzungen guter Lehre.

Voraussetzungen des Seminars

Interne Voraussetzungen

Dieses Seminar richtet sich besonders an Studierende mit wenig Erfahrung in R, Studierende mit fortgeschrittenen Kenntnissen sind allerdings herzlich willkommen. Wichtig für dieses Seminar ist die Bereitschaft neue Kenntnisse in R zu erlangen und gemeinsam mit anderen Studierenden an Problemen zu arbeiten. Programmieren ist eine Fähigkeit, die nicht ohne das Suchen von Hilfe funktioniert, KommilitonInnen können immer erste Ansprechpartner bei Problemen sein. Dieser Aspekt soll auch von den Rahmenbedingungen des Seminars unterstrichen werden. Daher werden zu Beginn des Seminars

“Selbsthilfegruppen” gebildet. Studierende sollen sich bei Fragen und Problemen in diesen Gruppen selbst helfen können (Battistich et al., 1993; Beach, 1974). Um den Zusammenhalt dieser Gruppen zu fördern werden Übungen und Aufgaben während der Seminarsitzungen häufig innerhalb dieser Gruppen durchgeführt. Gruppenarbeiten sollen soziale Interaktion unter den Teilnehmenden fördern (Hurst et al., 2013).

Neben der Bereitschaft zur Auseinandersetzung mit R und Zusammenarbeit mit KommilitonInnen ist es zwingend notwendig, dass Studierende einen eigenen Laptop mitbringen. Am besten kann man von diesem Seminar profitieren, wenn man schon eigene Datensätze aus den Abschlussarbeiten mitbringt und die Übungen anhand dieser Daten nachvollziehen kann. Praktische Relevanz der Inhalte soll die Motivation der Teilnehmenden fördern (Kember et al., 2008). Neue Konzepte und Übungen werden allerdings immer an vorgegebenen Beispieldatensätzen geübt, daher ist das mitbringen eigener Datensätze nicht zwingend notwendig.

Motivation der Teilnehmenden sollte besonders dadurch gestützt werden, dass sie in diesem Seminar praktisch relevante Inhalte für ihre Abschlussarbeit lernen. Inhalt des Seminars und Aufbau der Prüfungsleistung sind daher so gewählt, dass sie den Anforderungen der Datenanalyse einer Abschlussarbeit entsprechen (Biggs, 1996; Kandlbinder, 2014).

Externe Rahmenbedingungen

Das Seminar wird zu großen Teilen in Präsenz stattfinden, Aufgaben und vertiefende Informationen werden online bereitgestellt. Die Bereitstellung von online-ressourcen erhöht die Autonomie von Studierenden (Chen et al., 2010; Garcia & Pintrich, 1996), diese hybride Form von Präsenz- sowie Online-Lehre (*blended learning*) zeigte sich bereits vorteilhaft in Lernkontexten (Bernard et al., 2014; Means et al., 2013). Außerdem sollen wöchentliche Übungsaufgaben effektives verteiltes Lernen, eine besonders effektive Lerntechnik (Dunlosky et al., 2013), ermöglichen. Vorgesehen sind wöchentliche Sitzungen für die ersten 8 Wochen, in denen Grundlagen der Datenaufbereitung und Analyse vermittelt werden. Dies wird von

einem Blocktermin ergänzt, der die restlichen Sitzungstermine umfasst. Hier sollen Studierende in Kleingruppen die erworbenen Kenntnisse anwenden und anhand eines vorgegebenen Datensatzes eine Forschungsfrage untersuchen und ihre Ergebnisse sowie den benutzten Analysevorgang anschließend präsentieren. Anwesenheit in den wöchentlichen Sitzungen sowie Bearbeitung der Übungsaufgaben ist nicht verpflichtend (Macfarlane, 2013). Wie Studierende sich die Inhalte des Seminars aneignen ist ihnen freigestellt um Motivation zu fördern. Teilnahme am Blocktermin, Präsentation der Gruppenergebnisse des Blocktermins sowie Abgabe des abschließenden Projektes ist für eine Note zwingend erforderlich.

Ablaufplan - Semester

In Tabelle 1 ist der Ablaufplan des Semesters skizziert. Nach jeder Sitzung erhalten Studierende den Code der Sitzung sowie weiterführende Übungen, die sie freiwillig wahrnehmen können, um den Stoff zu vertiefen. Außerdem wird ein Fragenforum eingerichtet in dem Studierende Probleme und Nachfragen stellen und beantworten können. Ich werde an einem vorher bestimmten Tag selbst diese Fragen beantworten, ansonsten ist es angedacht, dass Studierende sich gegenseitig helfen.

Beispielsitzung - “neue Variablen erstellen”

Tabelle 2 enthält eine Übersicht über die Rahmenbedingungen, die für diese Sitzung relevant sind. Tabelle 3 skizziert den Ablauf der Sitzung. Ziel dieser Sitzung ist es, dass Studierende anhand einer Forschungsfrage und einem Datensatz begründen können, welche neuen Variablen erstellt werden sollen, damit sie später eigene Datensätze auf ihre Forschungsfrage hin manipulieren können. Außerdem sollen Studierende entscheiden können, welche Code-Umsetzung hierfür geeignet ist und die Befehle `mutate()` und `summarize()` anwenden können. Idealerweise verstehen die Studierenden die Funktionsweise von `group_by()`.

Hierzu werden zunächst Fragen und Inhalte der letzten Sitzung besprochen, dann soll sich gemeinsam anhand der Forschungsfrage, die zu Beginn des Seminars entwickelt wurde,

überlegt werden welche Variablen der Datensatz schon zur Verfügung hat und welche noch erstellt werden müssen. Im Anschluss wird die Seminarleitung zu 3 wichtigen Befehlen um neue Variablen zu erstellen kurze Impulse halten und ihre Funktionsweise vortragen. Studierende sollen diese Befehle dann direkt in einem interaktiven R-environment (**learnr** - Tutorials) (Aden-Buie et al., 2022) anwenden. Hier können sie basierend auf einem Datensatz verschiedene Anwendungen ausprobieren, und erhalten von der Software stets Feedback, über die Korrektheit ihrer Anwendung. In der Arbeitsphase werden Studierende in “Selbsthilfegruppen” aufgeteilt. Diese Gruppen sollen gemeinsam die Übungsaufgaben bearbeiten und erste Ansprechpartner bei Problemen sein. Das Geübte wird dann mit einem Kahoot-Quiz noch einmal überprüft. Am Ende wird in Form einer Blitzlicht-Runde kurz Feedback zum Sitzungsinhalt und zu noch offenen Fragen gegeben. Im Anschluss an die Sitzung werden vertiefende Aufgaben auf Moodle hochgeladen, die Studierenden ermöglichen sollen, die Konzepte des Seminars und Funktionsweise der Befehle noch weiter zu verstehen.

Die Sitzung legt großen Wert auf den theoretischen Aspekt der Anwendung der Forschungsfrage auf den Datensatz und besonders auf die technische Umsetzung dieser Überlegungen. Diese Inhalte werden in den Übungen innerhalb der Sitzung, am Blocktermin, in der Hausarbeit und in den jeweiligen Abschlussarbeiten der Studierenden relevant sein.

Fazit

Das Seminar soll Studierenden die Möglichkeit bieten Fähigkeiten zu erwerben, die Sie für ihre Abschlussarbeit benötigen werden. Daher sind die Seminarinhalte eng an den Vorgang einer Datenaufbereitung und Datenanalyse, wie sie in der Forschung durchgeführt wird, aufgebaut. Vom Beginn der Entwicklung einer Forschungsfrage, über das Einlesen, Aufbereiten, Verändern und Auswerten eines Datensatz bis zur anschließenden Präsentation der Ergebnisse sollen die Sitzungen sowie Prüfungsleistungen dieses Seminars den Prozess abbilden. Die Vorteile von *constructive alignment*, *blended learning*, Lerngruppen und praktischer Relevanz tragen zu besserer Lehrqualität des Seminars bei. Studierende erhalten stets die Möglichkeit die Inhalte vertiefend zu üben und anzuwenden.

Mögliche Probleme des Seminars sind:

- Software- oder Hardwareprobleme: Dies muss die Seminarleitung lösen können
- kein Wille zur Auseinandersetzung mit R: Hier besteht die Hoffnung, dass die hohe Relevanz des Seminars zu hoher Motivation führt. Außerdem werden voraussichtlich eher R-enthusiastische Menschen dieses Seminar wählen
- wenig Benutzung des Forums oder der Gruppen: Die Kleingruppen sowie das Selbsthilfeforum sind ein Angebot des Seminars, das es den Studierenden erleichtern soll Fragen zu stellen. Im Seminar wird mehrfach auf den Benefit dieses Angebots hingewiesen werden, wer es nicht nutzt ist sozusagen “selbst schuld”.
- zu viel Inhalt in einer Sitzung: Sitzungen sind absichtlich so geplant, dass sie tendenziell zu wenig Inhalt vermitteln. Das Seminar ist für R-Anfänger angedacht. Sollte eine Sitzung dann tatsächlich zu schnell abgearbeitet sein, kann Übungsmaterial in der Sitzung vorgezogen bearbeitet werden.

Ich halte das Seminar für eine sinnvolle Ergänzung zum jetzigen Lehrplan und würde es zwischen dem 4-6 Fachsemester ansiedeln. Meine Hoffnung ist, dass Studierende hier in einem lockeren Gruppensetting Vertrauen in ihre Fähigkeiten in R gewinnen können und langfristig von den Inhalten des Seminars profitieren.

References

- Aden-Buie, G., Schloerke, B., & Allaire, J. (2022). *Learnr: Interactive tutorials for r*.
<https://CRAN.R-project.org/package=learnr>
- Battistich, V., Solomon, D., & Delucchi, K. (1993). Interaction processes and student outcomes in cooperative learning groups. *The Elementary School Journal*, 94(1), 19–32.
- Beach, L. R. (1974). Self-directed student groups and college learning. *Higher Education*, 187–199.
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26, 87–122.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364.
- Chen, K.-C., Jang, S.-J., & Branch, R. M. (2010). Autonomy, affiliation, and ability: Relative salience of factors that influence online learner motivation and learning outcomes. *Knowledge Management & E-Learning: An International Journal*, 2(1), 30–50.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58.
- Garcia, T., & Pintrich, P. R. (1996). The effects of autonomy on motivation and performance in the college classroom. *Contemporary Educational Psychology*, 21(4), 477–486.
- Hurst, B., Wallace, R. R., & Nixon, S. B. (2013). The impact of social interaction on student learning. *Reading Horizons*.
- Kandlbinder, P. (2014). Constructive alignment in university teaching. *HERDSA News*, 36(3), 5–6.
- Kember, D., Ho, A., & Hong, C. (2008). The importance of establishing relevance in motivating student learning. *Active Learning in Higher Education*, 9(3), 249–263.

- Macfarlane, B. (2013). The surveillance of learning: A critical analysis of university attendance policies. *Higher Education Quarterly*, 67(4), 358–373.
- Macke, G., Hanke, U., Viehmann-Schweizer, P., & Raether, W. (2016). *Kompetenzorientierte hochschuldidaktik: Lehren-vortragen-prüfen-beraten*. Beltz Weinheim.
- Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record*, 115(3), 1–47.
- Ulrich, I. (2020). *Gute lehre in der hochschule - praxistipps zur planung und gestaltung von lehrveranstaltungen*. Springer. <https://doi.org/10.1007/978-3-658-31070-7>

Table 1*Sitzungsplan*

Sitzung	Titel	Beschreibung
1	Auftaktsitzung	Kennenlernen, Beschreibung der Seminarziele und Anforderungen, Sicherstellen der R-Installation, Aufsetzen der Ordnerstruktur
2	Datensätze einlesen	Einzelne Datensätze verschiedener Typen einlesen, mehrere Datensätze gleichzeitig einlesen, das rio-Paket, erstes Kennenlernen des Datensatzes
3	Datensatz strukturieren	Gute Variablennamen, Grundstrukturen von Datensätzen, Filtern von Variablen
4	Datensatz kennenlernen	Einträge zählen, Übersicht erhalten, erste grafische Checks
5	neue Variablen erstellen	Neue Zeilen mit mutate(), Übersichten mit summarize() erstellen
6	deskriptive Analyse	Forschungsfrage deskriptiv testen, Übersichten erstellen, erste plots mit ggplot()
7	Signifikanztests	Grundlagen von t-Tests, ANOVAs und linearen Modellen in R
8	Berichten von Ergebnissen	Wie beschreibe ich die Schritte meiner Datenanalyse? APA-konformes reporten der Ergebnisse
9-12	Blocktermin	Kleingruppen analysieren eine Forschungsfrage anhand eines Datensatzes und präsentieren ihre Ergebnisse dann im Plenum
Ende	Abgabe der Prüfungsleistung	Anhand eigens gewähltem Datensatz einen Forschungs- und Analysebericht erstellen. Code wird ebenfalls bewertet

Table 2*Rahmenbedingungen - Sitzung: neue Variablen erstellen*

Rahmenbedingungen	
Titel / Thema	neuen Variable mit mutate() erstellen
Externe Voraussetzungen	R, RStudio, Datensatz, tidyverse-Paket
Interne Voraussetzungen	Aufbau des Datensatzes verstanden, Forschungsfrage verstanden, minimale Vorkenntnisse im tidyverse
Angestrebte Kompetenzen	Entscheidungen bzgl. der Umsetzung einer Forschungsfrage in benötigte Variablen im Datensatz treffen. Fähigkeit den Vor- gang zum Erstellen einer neuen Variable zu begründen. Anwen- den der summarize()-Funktion zum Erstellen von Übersichten. Verstehen der group_by()-Funktion

Table 3*Ablaufplan - Sitzung: neue Variablen erstellen*

Zeit	Phase	Inhalt	Methode	Medium
5min	Einleitung	Ermitteln von Fragen zur letzten Sitzung, am Ende der Sitzung klären. (Ermitteln, Aktivieren, Orientieren)	Fragerunde	Konversation im Plenum
5min	Einleitung	Bereitstellung des aktuellen Stand des Datensatzes. Aktivieren des Wissens über den Datensatz, der in der letzten Sitzung erstellt wurde. (Motivieren, Aktivieren)	Fragerunde	Konversation im Plenum
5min	Einleitung	Was wollen wir mit unserer Fragestellung eigentlich überprüfen, welche Variablen brauchen wir dafür? (Aktivieren, Motivieren, Lenken)	Diskussion	Konversation im Plenum
10min	Arbeitsphase	Information - Grundkonzepte der Datensatzmanipulation - mutate()	Präsentation	Powerpoint
15min	Arbeitsphase	Arbeiten - Anwendung des mutate() Befehls	Übung	learnr-Tutorial
10min	Arbeitsphase	Information - Übersichten mit summarize()	Präsentation	Powerpoint
15min	Arbeitsphase	Arbeiten - Anwendung von summarize()	Übung	learnr-Tutorial
5min	Arbeitsphase	Exkurs - Gruppieren von Datensätzen mit group_by()	Präsentation	Powerpoint
10min	Arbeitsphase	Arbeiten - Anwendung von group_by() auf den vorher geschriebenen Code	Übung	learnr-Tutorial
5min	Abschluss	End-Quiz zu wie erstellt man eine Summe (Reihen, oder Zeilenweise) - wie z-standardisiert man eine Variable? (Überprüfen, Sichern)	Quiz	Kahoot
5min	Abschluss	Was hat gut funktioniert, wo sind noch Fragen offen? (Rückmelden/Rückblicken/Zusammenfassen)	Blitzlicht	Konversation im Plenum

Note. In Klammern sind jeweils die Verben, die das Ziel der jeweiligen Übung beschreiben angegeben (Macke et al., 2016).