

Empirisch-Wissenschaftliches Arbeiten

Übung zur computergestützten Datenanalyse

© Prof. Dr. Stephan Huber

4. Juni 2024

Inhaltsverzeichnis

Vorwort	1
Wikipedia sagt:	1
1 Die Programmiersprache R	3
2 Das Verfassen eines wissenschaftlichen Aufsatzes	4
2.1 Die traditionelle Weise	4
2.2 Vorteile von Quarto	5
2.3 Einführung in Quarto	5
2.4 APA konformes Manuscript erstellen mit Quarto	6
References	7

Abbildungsverzeichnis

Tabellenverzeichnis

Vorwort

Wikipedia sagt:

“Die Psychologie [...] ist eine empirische Wissenschaft”

Diese Unterlagen helfen...

- die Abfolge und die Inhalte der *Übung zur computergestützten Datenanalyse* zu überblicken,
- die Übungsaufgaben zu verstehen und zu bearbeiten und
- die Projektarbeit der Veranstaltung *Empirisch-Wissenschaftliches Arbeiten* erfolgreich zu gestalten.

Die Übung vermittelt...

- Kenntnisse der Programmiersprache R welche eine wissenschaftliche Datenbearbeitung und Datenanalyse ermöglichen.
- Kenntnisse zum programm-basierten Verfassen von wissenschaftlichen Arbeiten (Aufsätze, Bücher, Arbeitspapiere, Hausarbeiten).

Studierende lernen...

- Daten mit der Programmiersprache R und mit Hilfe der integrierten Entwicklungsumgebung RStudio einzulesen, zu bearbeiten und empirisch auszuwerten.
- Empirische Ergebnisse in ein publikationswürdiges Format zu übertragen.
- Einen APA konformes Manuskript mit Quarto, bzw. (R)Markdown, zu erstellen und dies entsprechend zu publizieren.
- Literatur entsprechend wählbaren Zitationsregeln unter Verwendung von Quarto und BibTeX in einen Aufsatz einzuarbeiten.

Studierende sollen...

- Die angeführte Literatur studieren: Ohne eigenständige Vor- und Nachbereitung lassen sich die Programmierkenntnisse nicht erlernen.
- Aktiv um Hilfe bitten: Wenn etwas unklar ist, kann ich individuell während des Kurses versuchen zu helfen. Für eine intensivere Betreuung, bitte ich mich zu kontaktieren, in die [Sprechstunde](#) zu kommen, oder eine außerordentliche Sprechstunde zu vereinbaren. Dies ist möglich und erwünscht.
- Inhaltliche Fragen und Wünsche jederzeit kommunizieren. Es besteht die Möglichkeit diese in das Curriculum aufzunehmen.

Liebe Studierende,

das Erlernen einer Programmiersprache in Verbindung mit empirischen Arbeiten ist eine Herausforderung die Vielen keinen Spaß macht. So ist es nur Verständlich, dass die Sinnhaftigkeit dieses Kurses teilweise von Studierenden angezweifelt wird. Tätigkeiten die keinen Spaß machen, sollten sinnstiftend sein oder zumindest ein monetäres Einkommen sichern. Da das Vorhandensein von empirischen Kenntnissen und einer Programmiersprache in einem Lebenslauf zweifelsfrei in der heutigen Zeit die Vermittlungsfähigkeit und die Verhandlungsposition am Arbeitsmarkt wesentlich verbessern, will ich mich hier kurz bemühen, die Sinnhaftigkeit zu thematisieren.

Ich verstehe die Abneigung gegenüber diesen Kurs: Viele haben sich nicht für ein Studium der Psychologie entschieden, um empirische Methoden und deren computergestützte Umsetzung zu erlernen. In der modernen Welt aber, insbesondere in der psychologischen Forschung, ist ein Verständnis von empirischen Methoden sowie deren computergestützten Umsetzung die praktische Voraussetzung zum Erkenntnisgewinn. Ohne dieses Verständnis verharret man bei rein theoretische und philosophische Überlegungen ohne jede Evidenz. Eine professionell agierende Psychologin und Psychologe, sollte die Fähigkeit besitzen die Literatur in seinem Fach zu begreifen sowie in der Lage ein die Ergebnisse kritisch zu hinterfragen und/oder zu überprüfen.

Ich bemühe mich, die Veranstaltung so attraktiv wie möglich zu gestalten. Ich biete...

- ein ausführliches [Skript zur Programmiersprache R](#) an, welches
 - eine Batterie an [Übungsaufgaben mit Lösungskripten](#) und
 - eine Vielzahl an [interaktive Übungen](#) zum eigenständigen bearbeiten enthält.
- dieses Skript, welches
 - Psychologie-spezifische empirische Inhalte aufgreift und
 - Software vorstellt, welche die Erstellung der Projektarbeit erleichtert.
- mündliche Erklärungen in der Veranstaltung.
- die Möglichkeit spezifische Fragen zu stellen und Unklarheiten anzusprechen.
- individuelle Betreuung während und außerhalb der [Sprechstunde](#).

Wenn sie Vorschläge und Wünsche bezüglich der Inhalte oder der didaktischen Aufbereitung haben, bitte ich diese auszusprechen. Konstruktive Kritik ist sehr willkommen. Ich nehme diese an und ernst. Ob Sie diesen Kurs letztendlich als gelungen betrachten, ist ihrer Wahrnehmung überlassen. Bevor Sie den Kurs aber schlecht evaluieren, bitte ich sie um Folgendes: Fragen Sie sich, ob ihr Wille und ihr Wunsch ausgeprägt genug waren, um sich ernsthaft mit den Inhalten und den Angeboten auseinanderzusetzen und ob sie mir evtl. die Gelegenheit gegeben haben auf Ihre Wünsche einzugehen.

Abschließend wünsche ich Ihnen viel Freude mit dem Kurs und den angebotenen Unterlagen und Inhalten. Ich freue mich, diesen Kurs halten zu dürfen und zu können. Es ist mir stets eine Freude, den anwesenden Studierenden R, Quarto, BibTeX und Co. erklären zu können. Ich wünsche mir, möglichst Viele mit den dargebotenen Inhalten, das Studium zu bereichern und die Bearbeitung der Projektarbeit sowie der Abschlusarbeit zu erleichtern.

Ihr
Stephan Huber

1 Die Programmiersprache R

Ich bitte Sie, studieren sie das Skript *How to use R for data science* [Huber, 2024].

In den ersten Wochen werden wir uns ausschließlich damit beschäftigen, die Programmiersprache R zu erlernen. Das ist ähnlich mühsam wie das Erlernen einer wirklichen Sprache. Wer keine Lust darauf hat, wird es schwer haben. Ich beispielsweise hatte in der Schule überhaupt keine Lust auf Englisch und Latein. Dementsprechend schlecht waren meine Noten. Ich musste die siebte Klasse wiederholen und bis zum Abitur waren Sprachen für mich ein nötiges Übel. Erst als ich im Studium sah, dass praktisch alle relevanten und für mich interessanten Artikel und Bücher in englischer Sprache verfasst sind, machte das Erlernen der Sprache einen Sinn für mich. Jetzt lehre ich abseits dieses Kurses ausschließlich auf Englisch und publiziere in englischer Sprache. Interesse und Freude sind mächtige Katalysatoren für Erfolg.

Das Schreiben von Code ist für die meisten Studierenden Neuland. Studierende im Jahr 2024 sind zumeist mit dem Smartphone aufgewachsen und demnach sind Sie es gewohnt, ihre Geräte (Smartphone, Tablet, Desktop-PC) ohne zur Hilfenahme einer Programmiersprache zu steuern. Das ist wunderbar: Die grafische Benutzeroberfläche heutzutage erlaubt eine effiziente und intuitive Art der Steuerung mit der Computermaus, durch Wischen, Tippen oder durch Spracheingabe. Leider hat diese Art der Steuerung massive Nachteile beim wissenschaftlich orientierten Arbeiten mit Daten. Insbesondere was die Reproduzierbarkeit der Ergebnisse und die Flexibilität des Arbeitsprozesses anbelangt, stößt man bei Applikationen ohne Code an Grenzen. Die Vor- und Nachteile von Script-basierten Arbeiten werden im Kapitel *The limitations of no-code applications* [Huber, 2024] ausführlich erläutert.

Zusammenfassend sollten Studierende nach den ersten 5-6 Unterrichtseinheiten folgendes getan haben beziehungsweise erlernt haben:

- Installation von
 - R,
 - RStudio und der
 - gängigsten Pakete.
- Wissen über...
 - den Aufbau von R Skripten.
 - die Verwendung von Funktionen, Objekten und Pakete in R.
 - die grundsätzlichen Eigenheiten der Programmiersprache R.
 - das Ausführen von Code (**Ctrl+Enter**, Klicken von **Run**, oder durch die Funktion `source()`).
 - die Verwendung von Pipes mit dem Pipe Operator (`|>`).
 - die Verwendung von logischen und relativen Operatoren.
 - die Funktionen des Pakets `dplyr` (`filter()`, `select()`, `mutate()`, `summarise()`, etc.)

2 Das Verfassen eines wissenschaftlichen Aufsatzes

Studierende die einen wissenschaftlichen Text verfassen müssen, verwenden zumeist Textverarbeitungsprogrammen wie *Microsoft Word* oder Apple's *Pages*. Diese sind bekannt und für Viele erscheinen diese Programme alternativlos. Das ist aber mitnichten so: Es gibt viele alternative Textsatz-Systeme wie LaTeX, Markdown, R Markdown und Quarto. Immer noch werden in der Wissenschaft eine Vielzahl an Doktorarbeiten mit LaTeX verfasst und fast alle Herausgeber arbeiten mit LaTeX.

Bei Office-Programmen wie MS Word und Pages wird das Dokumentlayout während des Schreibens angezeigt. Das wird auch das "What you see is what you get (WYSIWYG)" Prinzip genannt. Bei den genannten Alternativen hingegen, werden die Angaben für das Layout am Anfang vom Text oder direkt in den Fließtext geschrieben. Das fertige Dokument wird erst nach dem Umwandeln in ein Dokumentenformat wie PDF sichtbar. Das mag zunächst gewöhnungsbedürftig sein und ist sicherlich weniger intuitiv als die WYSIWYG Programme. Die intuitivste Lösung ist aber oft nicht die beste und einfachste Lösung. Die vielen Studienarbeiten die ich begleiten durfte, zeigen mir klar, dass die intuitiven Features von MS Word und Pages sich mittel- und langfristig oft als zeitaufwändig erweisen und die Anwender nur schlecht dabei unterstützen Fehler beim Verfassen von wissenschaftlichen Arbeiten zu vermeiden. Studierende die sich gegen eine "What you see is what you get (WYSIWYG)" Office Anwendung entschieden haben, sind meist glücklicher und erfolgreicher. Zumindest gilt das in den von mir begleiten Arbeiten. Code basierte Anwendungen helfen den Schreibenden, sich auf die Textarbeit zu konzentrieren. Die Ausgestaltung der Formatierung oder die Einhaltung von Zitationsregeln wird größtenteils von der Software übernommen. Die notwendige Anfangsinvestition, um bspw. Quarto zu lernen, amortisiert sich schnell und die Qualität der Aufsätze verbessert sich spürbar.

In den folgenden Unterabschnitten werde ich zunächst die typische Verwendung von WYSIWYG Anwendungen skizzieren. Danach werde ich die Vorteile von des Code basierten Verfassen von Texten am Beispiel von Quarto thematisieren, um sodann anzuleiten wie das Verfassen von Texten mit Quarto gelingen kann.

2.1 Die traditionelle Weise

Die Verwendung von klassischen Textverarbeitungsprogrammen wie Microsoft Word oder Apple Pages zum Verfassen wissenschaftlicher Texte ist gängige Praxis in der studentischen Welt. Trotz ihrer Benutzerfreundlichkeit für alltägliche Schreibprojekte, verursachen diese Programme jedoch erheblichen Mehraufwand, um den wissenschaftlichen Ansprüchen gerecht zu werden.

Eines der ersten Probleme, auf das Studenten stoßen, ist das Einbinden von Literatur. Die korrekte Formatierung gemäß verschiedenen Zitierrichtlinien ist oft nicht intuitiv und Fehler schleichen sich leicht ein. Insbesondere wenn die von der Software zur Verfügung gestellten Zitat- und Bibliographiefunktionen nicht oder nicht ordentlich benutzt werden. Anstatt einen externen Zitationsmanagern zu verwenden und sich einzuarbeiten, wie die Zitationen in der

jeweiligen Software einzuarbeiten sind, schreiben die Studierenden gerne die Zitationen und die Literaturliste händisch. Erfahrungsgemäß machen hier Studierende viele kleinere und manch größere Fehler. Allesamt sind diese Unregelmäßigkeiten unnötig.

Ein weiterer großer Schwachpunkt ist die Einhaltung spezifischer Formatierungsrichtlinien. Akademische Institutionen und Zeitschriften verlangen oft strenge Einhaltung von Formatierungsanweisungen, einschließlich Titelseite, Kopf- und Fußzeilen, Seitenränder und Überschriftenhierarchien. Word und Pages bieten zwar Vorlagen und Stile, diese müssen aber für jedes Dokument angepasst und oft während des Schreibens aufgrund kleiner Veränderungen im Text überarbeitet werden. Muss die Formatierung geändert werden, ist dies oft nur mit erheblichen Aufwand zu erledigen.

Das Einbetten von empirischen Ergebnissen wie statistischen Daten und Grafiken ist eine weitere Hürde. In Word und Pages ist der Prozess oft manuell: Forschungsergebnisse müssen aus Statistiksoftware exportiert, als Bilder gespeichert und dann in das Dokument eingefügt werden. Jede Änderung in den Daten erfordert eine Wiederholung dieses umständlichen Prozesses, was zu einer erheblichen Steigerung des Arbeitsaufwands und potenziellen Fehlern führen kann.

2.2 Vorteile von Quarto

Insgesamt kann der traditionelle Ansatz zum Verfassen wissenschaftlicher Aufsätze mit MS Word oder Pages für Studierende zu einer zeitaufwändigen und fehleranfälligen Angelegenheit werden. Im Folgenden Abschnitt werde ich Quarto (bzw. R Markdown) vorstellen. Das ist eine moderne Alternative die folgende Vorteile aufweist:

- Mit Quarto können mühelos verschiedene Output Formate erstellt werden. Beispielsweise kann der Text als Website, Manuscript und Buch dargestellt werden.
- Mit Quarto kann die Formatierung einfach verändert werden.
- Literatur kann einfach eingebunden werden und das Einhalten von Zitierregeln übernimmt die Software. Quarto's Integration mit Zitationsverwaltungssystemen erlaubt es, Literaturverweise und Bibliographien effizienter und konsistenter zu verwalten als bspw. in Word.
- Querverweise auf Abschnitte, Tabellen und Abbildungen, sind leicht zu erstellen.
- Die Datenanalyse und das Erstellen von Datenoutput geschieht integriert in Quarto. Daher sind die dargestellten Grafiken und Tabellen immer auf dem aktuellen Stand und eine händische Nachbearbeitung entfällt. Das Gebot der Reproduzierbarkeit der Ergebnisse ist erfüllt. Verfasser können ihre Datenvisualisierungen direkt in den Text einbetten, ohne dass manuelle Schritte wie die Erstellung von Grafiken, deren Export und das anschließende Einfügen in ein Word-Dokument nötig sind. Darüber hinaus erleichtert die Verwendung von Quarto in Kombination mit Versionskontrollsystemen wie Git die Zusammenarbeit an wissenschaftlichen Dokumenten, da Änderungen nachverfolgt, überprüft und integriert werden können, ohne sich auf komplexe Vergleichstools verlassen zu müssen, die in Word oft zu Konflikten führen.
- Mit ein und demselben Quarto-Dokument können eine Vielzahl von Formaten ausgegeben werden, darunter PDF, HTML, Word und sogar Präsentationsfolien. Diese Vielseitigkeit bedeutet, dass man sich weniger um das Format und mehr um den Inhalt kümmern kann.

2.3 Einführung in Quarto

Mit Quarto lassen sich APA konforme Texte verfassen. Hierzu bitte ich wie folgt vorzugehen:

- Installieren Sie R und R Studio.
- Installieren Sie Quarto mit

```
install.packages("quarto")
```

- Installieren Sie tinytex, um pdf Dateien zu erzeugen:

```
install.packages("tinytex")  
tinytex::install_tinytex()
```

- Lernen Sie Markdown. Markdown ist eine leichtgewichtige Auszeichnungssprache mit einer Syntax zur Formatierung von Klartext. Es ist eine wesentliche Fähigkeit, um Quarto effektiv zu nutzen. Beginnen Sie damit, genügend Markdown zu lernen, um Ihre Arbeit zu strukturieren, einschließlich Überschriften, Listen, Links und Codeblöcke. Sie können Markdown in 10 Minuten erlernen. Gehen Sie einfach auf www.markdowntutorial.com und arbeiten Sie sich durch die interaktiven Lektionen.
- Lernen Sie Quarto kennen. Lesen Sie hierzu Telford [2023]: [Enough Markdown to Write a Thesis](#). Er behandelt praktisch alles was für das akademische Schreiben nützlich sein kann. Alternativ können Sie direkt auf der quarto.org Website umfangreiche Informationen zur Arbeit mit Quarto finden, siehe quarto.org/docs/guide.

i Quarto und R markdown

Quarto ist ein relativ neues Werkzeug. Es kann als der Nachfolger von R Markdown angesehen werden. Fast alle R Markdown-Dokumente sind mit Quarto kompatibel. Quarto enthält jedoch einige Verbesserungen gegenüber R Markdown, die seine Benutzerfreundlichkeit erhöhen. Eine detaillierte Beschreibung aller Unterschiede und Gemeinsamkeiten zwischen den beiden Programmen finden Sie in [diesem Artikel] (<https://quarto.org/docs/faq/rmarkdown.html>).

2.4 APA konformes Manuscript erstellen mit Quarto

References

Stephan Huber. How to use R for data science, 2024. URL <https://hubchev.github.io/ds/>.

Richard J Telford. Enough markdown to write a thesis, 9 2023. URL <https://biostats-r.github.io/biostats/quarto/>.