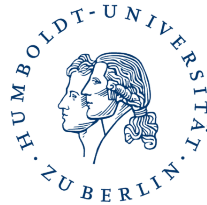


R for Political Science



Humboldt-Universität zu Berlin
Institut für Sozialwissenschaften
Politisches Verhalten im Vergleich

Wintersemester 2022/23

Infos

Mi 10:00 – 12:00
Universitätsstr. 3b, Raum 203

Tim Wappenhans
tim.wappenhans@hu-berlin.de

Moodle: moodle.hu-berlin.de
Kursnummer: 530664
Passwort: ThisIsNotAPipe

Inhalt und Ziel

Ziel des Kurses ist es, Wissen und Fähigkeiten zu vermitteln, die es euch ermöglichen, eigenständig Datenanalyseprojekte durchzuführen. Dazu zählt auch die Fähigkeit, neue Probleme anzugehen und kreative Lösungen zu finden. Konkret werden wir folgendes abdecken

- reproduzierbaren R code schreiben und debuggen
- Workflow und Programming Konventionen
- Data wrangling
- Data mining
- Data communication
- Tools für Kooperationen
- lernen, weiter zu lernen

Coden lernt man am besten durch ausprobieren. Dieses Seminar wird daher vor allem Workshop-Charakter haben. Das heißt, ich plane etwa die Hälfte der Zeit dafür ein, euch etwas vorzustellen und die andere Hälfte der Zeit dafür, dass ihr (interaktiv) coden könnt.

Teilnahmeleistungen

Die Pandemie stellt für uns alle eine Extremsituation dar. Solltet ihr Probleme mit dem Bewältigen der Anforderungen haben, meldet euch bitte bei mir.

In diesem Seminar werden 5 Leistungspunkte erworben. Das entspricht insgesamt 125 Stunden Aufwand, der sich dabei in folgende Punkte aufteilt:

1. Aktive Teilnahme

Der Kurs lebt durch eure regelmäßige und aktive Beteiligung. Das gilt zwar für jeden Kurs, aber ist ganz besonders in diesem Format wichtig. Wir werden viel hands-on programmieren, auf Probleme stoßen und versuchen, diese gemeinsam zu lösen. (1 LP)

2. Lektüre

Coden lernt man am besten durch ausprobieren, worauf der Fokus dieses Seminars liegt. Trotzdem gibt es sehr gute Literatur, die uns hilft, R besser zu verstehen. Lest die Lektüre also aufmerksam vor jeder Sitzung. Wir werden dabei vor allem auf das Standardwerk im Feld zurückgreifen:

- [R for Data Science](#)

Darüberhinaus gibt es noch viele weitere, kostenlose Ressourcen, die euch online zur Verfügung stehen. Nur zwei sehr gute Beispiele:

- [YaRrr! The Pirate's Guide to R](#)
- [Modern Dive into Tidyverse](#)

Viele Anwendungsbeispiele und politikwissenschaftlichen Daten kommen aus dem nicht kostenlosen Buch von Kosuke Imai und Nora Webb Williams

- [Quantitative Social Science: An Introduction in tidyverse](#)

(0.5 LP)

3. Problem Sets

Jede Woche werden wir das, was wir gelernt haben, in einem Problem Set anwenden. Ihr sollt in diesen Übungen kreative und unterschiedliche Lösungen finden, lernen zu googlen und im Moodle-Forum um Hilfe zu fragen. Die Problems Sets sollen bis Montag 23:59 auf Moodle hochgeladen werden. Ihr habt zwei Joker, das heißt, ihr könnt für zwei Sitzungen keine Problem Sets abgeben ohne das vorher abzusprechen. (2 LP)

4. Data Project Report

Zum Ende des Semester solltet ihr in der Lage sein, ein eigenständiges Datenanalyseprojekt durchzuführen, in dem ihr Programming und Workflow Konventionen anwendet. Das Projekt sollte folgende Punkte abdecken


- motivierende Forschungsfrage
- zwei unterschiedliche Datenquellen finden und verknüpfen
- Daten aufbereiten
- Deskriptive Statistiken (Graphiken und Tabellen)

- Analysen (Graphiken und Tabellen)
- Fazit

Der geschriebene Text kann sehr knapp ausfallen, es geht um die Datenarbeit. Findet euch in Gruppen bis zu 4 Personen dafür zusammen. Optimaler Weise hat dieses Projekt etwas mit euren Interessen zu tun, sodass ihr das als Grundlage für spätere Arbeiten wiederverwenden könnt.

Eure Ergebnisse werden in den letzten beiden Wochen des Semester präsentiert. Geht in der Präsentation (ca. 10 min) vor allem auf **Probleme** ein, auf die ihr bei der Bearbeitung gestoßen seid und welche **kreativen Lösungen** ihr gefunden habt. Abgabe des Dokuments ist der Dienstag vor den Präsentationen (7. Feb.) . Tragt euch für die Gruppenfindung bitte für ein Team im Pad auf Moodle ein. (1.5 LP)

Kursplan (vorläufig)

1. Willkommen	19.10.2022
2. R Basics	26.10.2022
3. Tidyverse	02.11.2022
4. Descriptives	09.11.2022
5. ggplot2	16.11.2022
6. Relational data	23.11.2022
7. Loops, Functions, Simulations	30.11.2022
8. Models	07.12.2022
9. Advanced Models	14.12.2022
	
10. APIs	04.01.2023
11. Textual Data	11.01.2023
12. Spatial Data	18.01.2023
13. L ^A T _E X und Overleaf	25.01.2023
14. Work on Project	01.02.2023
15. Präsentationen I	08.02.2023
16. Präsentationen II	15.02.2023