FULogo_Ausdruck_WORD_RGB.wmfDr. Christoph Nguyen

Otto-Suhr-Institut für Politikwissenschaft

Politisches System Deutschlands

Ihnestraße 22, 14195 Berlin

Mail: [christoph.nguyen@fu-berlin.de](mailto:christoph.nguyen@fu-berlin.de)

**Techniken der Empirischen Parlamentsanalyse in R**

Dienstag, 10-12 Uhr

# Zusammenfassung

Wie können wir die Komplexität moderner politischer Prozesse empirisch erfassen? Das Ziel dieses Kurses ist, Studierenden grundlegende Techniken der quantitativen Analyse in der Statistiksoftware R (anhand von spezifischen Anwendungsbeispielen zu vermitteln. Am Beispiel der parlamentarischen Analyse betrachten wir so nicht nur Standardtechniken (Visualisierung, deskriptive und inferentielle Statistik) sondern auch fortgeschrittene Methoden (Text-as-Data). Der Kurs richtet sich explizit auch an Studierende, die bisher keinen Zugang zur quantitativen Politikwissenschaft gehabt haben. Statistische Vorkenntnisse werden nicht erwartet, und der Fokus des Kurses liegt in der Vermittlung von anwendungsbezogenen und allgemein-relevanten technischen Fähigkeiten. Das Seminar findet als Mischung aus synchronen und asynchronen Elementen statt.

Materialien im Kurs sind zum Teil in Englischer Sprache. Die Kurssprache ist Deutsch.

# Organisatorische Hinweise

## Literatur und Kursmaterial

Die Pflichtlektüre und weitere Informationen (Videos + R Code + Datensätze) werden auf Blackboard bereitgestellt.

## Allgemeine Sprechstunde

Dienstag 12-14 Uhr. Bitte vereinbaren Sie vorab einen Termin auf <https://www.polsoz.fu-berlin.de/polwiss/forschung/systeme/polsystem/Team/Christoph-Nguyen.html>

Leider können in diesem Semester Sprechstunden nur digital angeboten werden. Sie können direkt über <https://fu-berlin.webex.com/meet/christoph.nguyen> der Sprechstunde beitreten.

Bitte beachten Sie auch die gesonderten Sprechstunden die exklusiv für diesen Kurs angeboten werden (siehe Seminarplan). Diese Sprechstunden direkt auf Blackboard gebucht werden.

## Kommunikation

Die Kommunikation im Kurs findet via Blackboard und E-Mail statt. Bitte kontaktieren Sie mich über die E-Mail-Adresse [christoph.nguyen@fu-berlin.de](mailto:christoph.nguyen@fu-berlin.de). Inhaltliche Fragen sollten ausschließlich in den Sprechstunden besprochen werden.

**Bitte beachten Sie auch, dass es nicht in jeder Woche ein synchrones Seminar geben wird.** Ich stehe Ihnen aber auch in diesen Wochen für besondere Sprechstunden zur Verfügung, die Sie gerne in Anspruch nehmen sollten.

Spezifisch hat der Seminarplan die folgenden Elemente:

* Video: Vorab aufgezeichnete Videos, in denen technische oder theoretische

Inhalte vermittelt werden. Diese Videos sind die Basis der Online-Diskussion **und sollten vor dem jeweiligen Seminartermin bearbeitet werden**.

* ­Datacamp: Digitale Lerneinheiten, die die technischen Fähigkeiten vermitteln

direkt testen. Auch diese Inhalte sind Grundlage für die Online-Diskussionen, und sollten deshalb **vor dem jeweiligen Seminartermin abgeschlossen werden**.

Einheiten die als „Extra-Datacamp“ ausgewiesen werden sind als mögliche Erweiterungen gedacht, werden aber nicht zwangsläufig erwartet.

* Diskussionen Synchrone Diskussionen entweder über Webex oder in Person.

Idealerweise werden wir nicht die vollen 90 Minuten ausschöpfen, da die asynchronen Elemente vergleichsweise viel Zeit einnehmen werden.

* Sprechstunde In Wochen ohne synchrone Diskussionen stehe ich Ihnen für

kurspezifische Kleingruppensprechstunden (mit maximal 2 Teilnehmer\*Innen) zur Verfügung. Bitte nutzen Sie für die Anmeldung in dieser Sprechstunde den gesonderten Link, den Sie auf Blackboard finden.

* Problem-Sets Um kursspezifische Anwendungsbeispiele der Methoden zu

ermöglichen, werden im Laufe des Semesters vier Problem-Sets angeboten. Geben Sie diese bitte **bis zum Sonntagabend (23:59 Uhr) vor der Deadline auf Blackboard ab.**

Es geht hier explizit nicht darum, dass Sie schon alle Antworten beantworten können. Vielmehr ist es hilfreich, Verständnisprobleme so schnellstmöglich zu identifizieren und beheben zu können.

Weiterhin ist die Kollaboration zwischen Studierenden hier ausdrücklich erwünscht. Sie können in Kleingruppen (Max. 3 Personen) zusammenarbeiten. Allerdings sollte jede Person ein eigenes Problemset abgeben.

### Technische Voraussetzung

Für die Teilnahme am Kurs benötigen Sie einen PC/Mac auf dem das Statistikprogramm R und R-Studio installiert und genutzt werden können. Tablets und ähnliche Endgeräte sind leider nicht ausreichend.

### Datacamp Zugang

Dieser Kurs wird mit Hilfe des Online Dienstes Datacamp angeboten. Diese Unterlagen werden auf Englisch angeboten. Den Zugang zu Datacamp lasse ich Ihnen nach Beginn auf Blackboard zukommen.

# Leistungsanforderungen im Seminar

### Aktive Teilnahme:

Abschluss der Aufgaben in Datacamp + Problem Sets 1 - 4

### Leistungsschein:

1. Leistungen für die aktive Teilnahme + Vertiefung der Problem Sets / Forschungsdesign (3000 Worte, exkl. Bibliographie)

Oder

1. Leistung für die aktive Teilnahme + Hausarbeit (6000 Worte, exkl. Bibliographie)

**Deadlines:**

17.7.2023: Abgabe der Exposees + (nichtverbindliche) Anmeldung für die Hausarbeit bitte digital über Blackboard.

30.9.2023: Abgabetermin

Bitte reichen Sie Ihre Hausarbeiten digital über Blackboard ein. Abgabe des R Codes wird nicht erwartet.

## Hintergrundliteratur

Wickham, Hadley. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O’Reily Media. Kostenfrei online [hier](https://r4ds.had.co.nz/).

Healy, Kieran. 2018. *Data Visualization - A Practical Introduction*. Princeton University Press.

## Seminarplan

Sitzung 1: Einführung und Vorstellung (synchron) 18.4.2023

Seminar: Vorstellung und Q&A – Ablauf und Logistik

Sitzung 2: Theoretischer Hintergrund und Fragestellungen (synchron) 25.4.2023

Videos: R Installation + Projektorganisation in R

**Deadline: *R muss auf dem eigenen Rechner installiert sein***

Seminar: Parlamente und politikwissenschaftliche Fragestellungen

*Literatur – Einer der Vier Texte:*

Wenzelburger, G., & Fehrenz, S. (2018). Die Union und die „Ehe für Alle“. Bestimmungsfaktoren des Abstimmungsverhaltens in der CDU/CSU-Fraktion im Bundestag. *Zeitschrift Für Parlamentsfragen*, *49*(3), 512–530

Bäck, H., & Debus, M. (2020). Personalized versus partisan representation in the speeches of migrant members of parliament in the German Bundestag. *Ethnic and Racial Studies*, *43*(9), 1673–1691. <https://doi.org/10.1080/01419870.2019.1654112>

Ehlert, Niek, Annika Hennl, and André Kaiser. 2007. “Föderalismus, Dezentralisierung Und Performanz. Eine Makroquantitative Analyse Der Leistungsfähigkeit Territorialer Politikorganisation in Entwickelten Demokratien.” *Politische Vierteljahresschrift* 48(2): 243–68.

Stecker, Christian. 2016. “The Effects of Federalism Reform on the Legislative Process in Germany.” *Regional and Federal Studies* 26(5): 603–24.

Sitzung 3: Einführung in das Statistikprogram R Sitzung I 2.05.2023

Datacamp: Introduction to R

Onlinesprechstunde: Bitte um vorherige Anmeldung über Blackboard

Extra-Datacamp: Intermediate R

Sitzung 4: Wie lese & schreibe ich eine quant. Arbeit? (synchron) 9.05.2023

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences* 115, no. 11 (March 13, 2018): 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.

Rubenson, Daniel. “Tie My Hands Loosely: Pre-Analysis Plans in Political Science.” *Politics and the Life Sciences* 40, no. 2 (ed 2021): 142–51. <https://doi.org/10.1017/pls.2021.23>.

Sitzung 5: Einführung in das Statistikprogram R Sitzung II: Tidyverse 16.05.2023

Datacamp: Introduction to the tidyverse (Data wrangling + Grouping and summarizing)

Video: Deskriptive Daten am Beispiel Föderalismus, Dezentralisierung und

Performanz im internationalen Vergleich

Onlinesprechstunde: Bitte um vorherige Anmeldung über Blackboard

Sitzung 6: Anwendung I: Deskriptive Daten (synchron) 23.05.2023

**Deadline:**  ***Problem Set I – Deskriptive Daten***

Seminar: Diskussion: Problem Set I

Diskussion: Anwendung auf eigene Fragestellungen

Sitzung 7: Daten Visualisieren mit ggplot2 30.05.2023

Datacamp: Introduction to the tidyverse (Data Visuaization +

Types of Visualization)

Video: Daten visualisieren am Beispiel Legislatives Verhalten

Onlinesprechstunde: Bitte um vorherige Anmeldung über Blackboard

Extra-Datacamp: Interactive Data Visualization with plotly in R

Datavisualization with ggplot2 (Part 1 + Part 2)

Sitzung 8: Anwendung 2: Visualisierungen (synchron) 06.06.2023

**Deadline:**  ***Problem Set II – Visualisierungen***

Seminar: Diskussion der Visualisierungen + Diskussion Problem Set II

Sitzung 9: Regressions-Analyse 13.06.2023

Datacamp: Modeling with Data in the Tidyverse

Video: Regressionsanalysen am Beispiel

Video: Germaparl und der Deutsche Bundestag

Sitzung 10: Anwendung 3 – Regressionsanalysen (synchron) 20.06.2023

**Deadline:**  ***Problem Set III – Regression***

Seminar: Diskussion der Regressionsanalysen + Diskussion Problem Set III

Sitzung 11: Text-as Data I: Basics 27.06.2023

Datacamp: Introduction to text analysis in R (Chapter 1-2)

Video: Einführung in Quanteda

Onlinesprechstunde: Bitte um vorherige Anmeldung über Blackboard

Sitzung 12: Text-as Data III: Fortgeschritten Techniken 04.07.2023

Datacamp: Introduction to text analysis in R (Chapter 3-4)

Video: Topic Modeling und Word Scores

Video: Sentiment Analysen

Bräuninger T., Debus M., Müller J., Stecker C. (2020) Dimensionen des politischen Wettbewerbs. In: Parteienwettbewerb in den deutschen Bundesländern. Springer VS, Wiesbaden. <https://doi.org/10.1007/978-3-658-29222-5_3>

Onlinesprechstunde: Bitte um vorherige Anmeldung über Blackboard

Sitzung 12: Text-as Data II: Anwendung (synchron) 11.07.2023

**Deadline:**  ***Problem Set IV–Quanteda***

Seminar: Besprechung des Problem Sets

Workshop- Eigene Textdaten finden und in Quanteda nutzen

Sitzung 14: Besprechungen der Hausarbeitsthemen (Online-Synchron) 18.07.2023

**Deadline:**  ***Exposee für die Hausarbeit***

***Peer Feedback***