

Syllabus

Projektseminar Konzeption quantitativer Forschung
„Big Data? Webbasierte Datenerhebung, exploratives Data Mining
und Critical Statistics für die Humangeographie“
Wintersemester 2017/18

*Institut für Humangeographie
Goethe Universität Frankfurt
Dozent: Till Straube <straube@geo.uni-frankfurt.de>*

Stand: 29.01.2018

Überblick

Unter dem Schlagwort „Big Data“ werden seit ca. einer Dekade im Allgemeinen die großen Datenmengen verstanden, die unter den Bedingungen vernetzt verlaufender Produktion, Distribution und Konsumtion sowie den dafür notwendigen Interaktionen (und hier insbesondere im Bereich der Telekommunikation, des elektronischen Zahlungsverkehrs, des Online-Handels etc.) produziert werden.

Das Seminar „Big Data!? — Webbasierte Datenerhebung, exploratives Data Mining und Critical Statistics für die Humangeographie“ stellt sich neuen Fragestellungen und Herausforderungen auf den Gebieten der Datenerhebung und -auswertung im Rahmen eines zweisemestrigen Projektseminars.

Im Zentrum steht der Kompetenzaufbau in der Extraktion von webbasierten Daten (Scraping), in der statistisch-explorativen Datenauswertung (Data Mining) sowie auf den Gebieten der kritischen Reflexion und der gesellschaftskritisch orientierten Anwendung techno-statistischer Komplexe (Critical Statistics).

Im ersten Teil des Projektseminars (Wintersemester 2017/18) erlernen die Studierenden den Umgang mit verschiedenen Datenformaten und Tools zur Datenerhebung. Im zweiten Teil (Sommersemester 2018) werden die Teilnehmenden ein eigenes Projekt entlang der Prozesskette Datenerhebung, Datenaufbereitung, -auswertung und Interpretation vor dem Hintergrund einer eigenständig entwickelten Problemstellung durchführen. Die Softwareumgebung R ist die zentrale Grundlage für beide Seminarteile. Eine begleitende kritische Reflexion der vermittelten Kenntnisse anhand von Textarbeit findet durchgehend statt.

Leistungsanforderungen

Individuell

- Lesetexte gründlich vorbereiten
- R-Aufgaben selbständig lösen

In der Projektgruppe

- Anfertigung Projektskizze (bis 7. Januar 2018)
- Datenerhebung + Bericht (bis 18. März 2018)
- Abschlussbericht (bis 31. August 2018)

Ressourcen

R Tutorials

- <https://www.tutorialspoint.com/r/index.htm>

- <https://www.statmethods.net/r-tutorial/index.html>

Ausführliches Skript

- <https://www.stat.berkeley.edu/classes/s133/schedule.html>

R für Fortgeschrittene

- <http://adv-r.had.co.nz/>

R Cheat Sheets

- <https://www.rstudio.com/resources/cheatsheets/>

HTML

- <https://www.w3schools.com/html/>

XPath

- https://www.w3schools.com/xml/xpath_intro.asp

R Markdown

- <http://rmarkdown.rstudio.com/lesson-1.html>

Reguläre Ausdrücke

- <https://danielfett.de/de/tutorials/tutorial-regulare-ausdrucke/>

Termine

Montag, 16. Oktober 2017

- Vorbereitungssitzung

Montag, 23. Oktober 2017

- Textbesprechung: Anderson (2008)
- Installation von R und RStudio auf Laptops
- Konsole
- Orientierung in RStudio

Montag, 30. Oktober 2017

- Datentypen
- Vektoren
- Vorgegebene Funktionen und Hilfe
- Vorbesprechung Aufgabenpaket 1

Montag, 6. November 2017

- Nachbesprechung Aufgabenpaket 1
- Data Frames
- Import von Daten
- Data Frame manipulieren
- Vorbesprechung Aufgabenpaket 2

Montag, 13. November 2017

- Nachbesprechung Aufgabenpaket 2
- Einführung: APIs
- Beispiel: API

Montag, 20. November 2017

- Textbesprechung: Kitchin (2014)
- Crashkurs: HTML
- Beispiel: Tabellen (Wikipedia)

Montag, 27. November 2017

- Vertiefung: HTML
- Beispiel: Scraping (marinetraffic.com)
- Diskussion: Datenquellen
- Auftrag: Ideen für Projekte

Montag, 4. Dezember 2017

- Ideensammlung Projekte
- Feedback

Montag, 11. Dezember 2017

- Verbindliche Zuordnung Projekte
- Kollaboration und Projektarbeit
- Funktionen

Montag, 18. Dezember 2017

- Textbesprechung: Breuer (2005)
- Übung zu Funktionen
- Recherche Datenquellen

Montag, 8. Januar 2018

- Vorstellung und Besprechung der Projektskizzen
- Zusammentragen von Problemlagen
- Planung weiterer Themen

Montag, 15. Januar 2018

- Beispiel API: Twitter

Montag, 22. Januar 2018

- Vertiefung explorative Statistik
- Projektarbeit

Montag, 29. Januar 2018

- Datensäuberung und -extraktion mit RegEx
- Projektarbeit

Montag, 5. Februar 2018

- Zusammentragen Stand der Projekte
- Projektarbeit

Literatur

Anderson, Chris. 2008. The End of Theory. The Data Deluge Makes the Scientific Method Obsolete. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory (zugegriffen: 8. Mai 2015).

Breuer, Ingo. 2005. Statistiken oder: Wie werden „Nomaden“ in Marokko gemacht? In: *Methoden als Aspekte der Wissenskonstruktion. Fallstudien zur Nomadismusforschung*, hg. von Jörg Gertel, 55–73. Halle: Mitteilungen des SFB „Differenz und Integration“ 8.

Kitchin, Rob. 2014. Big Data. In: *The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Consequences*, 67–79. London: Sage.