



Forschungsseminar Politischer Datenjournalismus

Universität Zürich, Frühjahr 2024

Dozierende: Lucas Leemann (leemann@ipz.uzh.ch), Karsten Donnay (donnay@ipz.uzh.ch),
Jacqueline Büchi (jacqueline.buechi@tamedia.ch), Reto Mitteregger (mitteregger@ipz.uzh.ch)

Kontakt: donnay@ipz.uzh.ch

Sprechstunden: Nach Vereinbarung

Kursnummer: 615-508a Seminar

Termin: Di. 10:15-12:00 (ab 20.02.2024)

Raum: AND-3-06

Webseite: [OLAT](#)

Inhalt und Ziel:

Das Forschungsseminar gibt Studierenden praktische Einblicke in die angewandte, datenbasierte Arbeit und zugehörige journalistische Aufarbeitung. Die Methodenkenntnisse aus dem ersten Semester werden zudem noch gezielt in einzelnen datenwissenschaftlichen Themenbereichen, z.B. im Bereich Textanalyse, Webscraping bzw. API-Nutzung, Verwendung räumlicher Daten, weiter vertieft. Neben der Vermittlung von fortgeschrittenen datenjournalistischen Methoden, ist das Seminar bewusst sehr Praxis-nah konzipiert. Insbesondere wird Jacqueline Büchi als erfahrener Journalistin das Seminar begleiten und Einblicke in den (daten)journalistischen Prozess geben.

Voraussetzungen:

Der Kurs setzt die Teilnahme am Kurs «Vorbereitung zum Forschungsseminar Politische Datenjournalismus» (im HS 2023) voraus. Auch dieser zweite Teil des 2-semesterigen Seminar-Zyklus erwartet für den erfolgreichen Abschluss des Kurses von den Studierenden eine Bereitschaft zur aktiven Teilnahme. Wir arbeiten auch in diesem Seminar wieder mit der freien Statistiksoftware [R](#) und [RStudio](#) und gehen davon aus, dass die Studierenden sich damit auskennen.

Bewertung:

Als Prüfungsleistung realisieren die Studierenden jeweils ein Forschungsprojekt in der Form eines journalistischen Beitrags. Konkret muss dazu ein datenwissenschaftlicher Blogbeitrag (inkl. zugrunde liegender Datenauswertung) erstellt werden; weitere Details zu den genauen Anforderungen finden sich am Ende des Syllabus. Im Verlauf des Kurses werden die Projektideen gemeinsam mit den Dozierenden entwickelt, reflektiert und verfeinert. Es ist vorgesehen, dass Blogbeiträge, die mit Note 5,0 oder besser bewertet wurden, auf dem [datenjournalistischen Blog](#) des IPZ veröffentlicht werden. Abgabefrist für die Forschungsprojekte bzw. Blogbeiträge ist der 30. Juni 2024.



Kursstruktur:

Das Seminar ist grob in vier Blöcke strukturiert, zwei mit mehr methodischem Fokus, zwei mit mehr (daten)journalistischen Inhalten. Durch den gesamten Kurs zieht sich zudem ein inhaltliches Leitthema; dies wird zu Anfang des 2. Blocks am 19.03.2024 präsentiert. Das Leitthema soll dazu dienen die Projekte thematisch zu fokussieren sowie nützliche Daten-Ressourcen zur Verfügung zu stellen. Die Erarbeitung der Blogbeiträge ist integraler Bestandteil des Kurses. Die Schritte bzw. Milestones sind hier in der Übersicht kurz dargestellt, detaillierte Informationen zum Ablauf und den Anforderungen finden sich auf der letzten Seite.

Block 1: Einführung und Vertiefung Methoden I

- 20.02. Einführung & Begrüssung (Karsten Donnay)
- 27.02. Was ist eine gute Geschichte? (Jacqueline Büchi)
- 05.03. APIs & Scraping I (Karsten Donnay)
- 12.03. APIs & Scraping II (Karsten Donnay)

Block 2: (Daten)Journalistisches Arbeiten

- 19.03. Vorstellung des inhaltlichen Leitthemas (Karsten Donnay, Lucas Leemann)
- 26.03. Ideenworkshop: «Redaktionssitzung» (Jacqueline Büchi, Karsten Donnay, Lucas Leemann)

Milestone 1: 2-min Pitch der Projektidee

Osterferien

- 09.04. Daten in Stories giessen (Jacqueline Büchi)

Block 3: Vertiefung Methoden II

- 16.04. Geographische Daten (Karsten Donnay, Aufzeichnung)
- 23.04. Textanalyse I (Gastdozent: Florian Schaffner)
- 30.04. Textanalyse II (Gastdozent: Florian Schaffner)

Block 4: Abschluss Projekte

- 07.05. Projektworkshop (Jacqueline Büchi, Reto Mitteregger)
Milestone 2: detaillierte Projektidee & erste Analysen
- 14.05. Redaktionsbesuch bei TA, 10:30-11:30 (Jacqueline Büchi, Reto Mitteregger, Lucas Leemann)
- 21.05. Präsentation der Blog-Projekte (Jacqueline Büchi, Reto Mitteregger, Lucas Leemann)
- 28.05. Präsentation der Blog-Projekte (Jacqueline Büchi, Reto Mitteregger, Lucas Leemann)
Milestone 3: endgültige Projektskizze mit genauer Planung (inkl. möglicher Schlagzeile)



Empfohlene Literatur:

Wir empfehlen einige eher grundlegende methodischen Ressourcen zu (angewandten) datenwissenschaftlicher Arbeit in R:

- Hadley Wickham, Mine Çetinkaya-Rundel & Garrett Golemund. (2023). *R for Data Science* (2nd Edition). O'Reilly Media. <https://r4ds.hadley.nz/>
- Edzer Pebesma & Roger Bivand. (2023). *Spatial Data Science*. CRC Press. <https://r-spatial.org/book/>
- Justin Grimmer, Margaret Roberts & Brandon Stewart. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.

Zusätzlich empfehlen wir für die (daten)journalistischen Aspekte die folgenden Lehr- bzw. Handbücher:

- Stephan Russ-Mohl & Tanjev Schultz. (2023). *Journalismus. Das Lehr- und Handbuch*. Herbert von Halem Verlag.
- Christoph Fasel. (2020). *Textsorten (3. Auflage)*. Herbert von Halem Verlag.
- Lorenz Matzat. (2016). *Datenjournalismus. Methode einer digitalen Welt*. Herbert von Halem Verlag

Und spezifisch für das Thema Textanalyse empfehlen wir die folgenden Quellen; weitere spezifische Literaturempfehlungen werden von den Dozierenden nach Bedarf für die einzelnen Themenbereiche in den jeweiligen Wochen zur Verfügung gestellt.

- Word embeddings: Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84 (5), 905–949. <https://doi.org/10.1177/0003122419877135>
- Topic Models: Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, 91, 1–40. <https://doi.org/10.18637/jss.v091.i02>
- Dictionary Methods: Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re)Claiming Our Expertise: Parsing Large Text Corpora With Manually Validated and Organic Dictionaries. *Political Communication*, 36(2), 214–226. <https://doi.org/10.1080/10584609.2018.1517843>
- Text Scaling: Slapin, J. B., & Proksch, S.-O. (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science*, 52(3), 705–722. <https://doi.org/10.1111/j.1540-5907.2008.00338.x>
- Supervised Machine Learning: Barberá, P., Boydston, A. E., Linn, S., McMahon, R., & Nagler, J. (2021). Automated Text Classification of News Articles: A Practical Guide. *Political Analysis*, 29(1), 19–42. <https://doi.org/10.1017/pan.2020.8>



Blogbeiträge – Format und Bewertung

Die Betreuung der Blogbeiträge erfolgt in aufeinander aufbauenden Schritten:

- Woche 6 (26.03.): 2-min Pitch der Projektidee; vorab muss eine Kurzbeschreibung (max. 300 Wörter) der Idee eingereicht werden
- Woche 11 (07.05.): gemeinsamer Projektworkshop; Projektidee für den Blogpost muss im Detail ausgearbeitet sein; erste Analysen sollten im Hintergrund bereits gemacht sein
- Woche 13 (21.05.) & 14 (28.05.): Präsentation der endgültigen Projektskizze mit genauer Planung (Daten, Analyse etc.) und einer möglichen Schlagzeile (Titel & Lead)

Im Anschluss an die Präsentationen zum Ende des Kurses bleibt ca. 1 Monat bis zur Abgabe des fertigen Beitrags am 30. Juni 2024. In der Zeit sollte hauptsächlich an der Ausarbeitung des eigentlichen Blogbeitrags gearbeitet werden.

Der datenjournalistische Blogbeitrag soll ein aktuelles und gesellschaftlich relevantes Thema aufgreifen. Der Blogbeitrag richtet sich an eine breite und allgemeingebildete Leserschaft. Wie ein guter journalistischer Beitrag soll er Interesse für die Thematik wecken und komplexe Zusammenhänge verständlich machen. Der Themenbereich soll klar eingegrenzt und die Forschungsfrage aktuell und eigenständig entwickelt sein. Die Wahl der Methode soll überzeugend sein. Datenbeschaffung und die Datenanalyse müssen transparent dargelegt werden.

Der Blogbeitrag zeichnet sich durch eine gute textliche und grafische Präsentation der sorgfältig und korrekt ausgewerteten und interpretierten Daten aus. Der Einstieg in den Text soll das Interesse am Thema wecken und kann kreativ gewählt werden. Der Beitrag ist einfach und verständlich, aber dabei sehr sorgfältig geschrieben und ist in kurze Abschnitte mit Zwischentiteln gegliedert. Der rote Faden ist hierbei immer erkennbar und die Argumentation ist überzeugend. An den Stellen, wo es Sinn macht, werden die ausgewerteten Daten mit einer Grafik, einer Tabelle oder einem anderen visuellen Element gezeigt. Die visuellen Elemente sind einfach verständlich und gut lesbar. Quellen müssen korrekt und einheitlich gekennzeichnet werden, sowohl für wörtliches und nichtwörtliches Zitieren und Bibliografieren.

Formale Anforderungen

- Titelinformationen: Titel des Blogs; Name, E-Mail und Matrikelnummer der/s Autorin/s; Abgabedatum des Blogs; Name und Dozierende der Vorlesung; Angabe der Anzahl Worte (excl. Lead, Anhänge und Literaturverzeichnis).
- Lead: In zwei bis drei möglichst kurzen Sätzen ins Thema einführen und das Interesse der Leser wecken. Im Titel und Lead soll noch nicht alles verraten werden.
- Länge: Maximal 1000 Wörter und mindestens zwei Auswertungen (z.B. Tabelle oder Grafik)
- Am Ende die Quelle(n) nennen und in separater Box Validität besprechen
- Eine unterschriebene Selbstständigkeitserklärung muss, wie bei allen schriftlichen Arbeiten, separat abgegeben werden.