Using Digital Trace Data in the Social Sciences (SS 2016)

Time:
July 18, 2016 Monday—10:00 to 13:15
July 19, 2016 Tuesday—13:30 to 18:30
July 20, 2016 Wednesday—10:00 to 13:15
July 21, 2016 Thursday—08:15 to 13:15
July 22, 2016 Friday—08:15 to 11:30

Place: E E 405
Start: July 18, 2016
Office Hours: By appointment

Description: In the course, students will learn fundamental techniques of data collection preparation, and analysis with digital trace data in the social sciences. In this, we will focus on working with the microblogging-service Twitter. Over the course, students are expected to become proficient in the use of two programming languages, Python and R. The course will be offered as a Blockseminar after the end of the regular term.

Level: Create—Students are expected to independently perform theory-driven data collections on the microblogging-service Twitter and use these data in the context of a series of specified prototypical analyses.

Vst.-Nr. & ECTS-Punkte:
POL-19640-20161—Vertiefungsseminar—6 ECTS
POL-19650-20161—Seminar—7 ECTS
POL-19630-20161—Doktorandenseminar—4 ECTS

Kursseite: http://andreasjungherr.net

I will post information, readings, and example scripts for the sessions of this course on my website http://andreasjungherr.net. The course follows closely a tutorial written by Pascal Jürgens and me, A Tutorial for Using Twitter Data in the Social Sciences: Data Collection, Preparation, and Analysis (2016). The tutorial is freely available on the Social Science Research Network (SSRN) at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2710146. I recommend all participants in the course to download the tutorial and the accompanying set of scripts available at https://github.com/trifle/twitterresearch. You will very likely profit from preparing the respective sections of the tutorial before and after the corresponding session.

A good many of the recommended further readings is available through the Proquest Safari-Books Online-Shelf (http://proquest.techbus.safaribooksonline.de). Access to Safari-Books Online is freely available by using your Uni-Konstanz VPN-access.
Requirements:
1. Regular and active participation.
2. Independent data analysis:
   - Following the course, you will be asked to perform and report an independent data analysis based on data collected on Twitter by you.
   - The aim of this paper is for you to demonstrate that you are able to independently apply and adapt the techniques learned during the course in the context of a specific research question developed by you.
   - **Presentation**: You will be asked to present a research question of interest to you on Thursday, July 21. For this presentation please prepare a short statement introducing your research question, your motivation, your proposed approach, and open questions. This presentation will take approximately 5 minutes and will be followed by a quick round of feedback from the other participants.
   - **Paper**: Font—Times New Roman, 12pt; Line-separation—1.5; Page borders—2.5 cm left and right, 2cm above and below; Page set—Block; The first line of each paragraph is indented.
   - **Citation Style**: Please follow the citation convention of the *American Political Science Review (APSR)* as given here [http://www.apsanet.org/apsrsubmissions2016](http://www.apsanet.org/apsrsubmissions2016).
   - **Cover page**: University, department, course title, paper title, name, Matriknr., semester count, study program, and e-mail-address.
   - **Length**: ca. 4000 words +10%
   - **Deadline**: Please return the paper on the date specified by the department (September 15) electronically at andreas.jungherr@gmail.com and by hardcopy with Birgit Jacob (Raum D 312). The date is mandatory and can only be extended in case of officially certified illness.
Syllabus

Course Overview and Session Details:
http://andreasjungherr.net/

Textbook:

Code Repository:
https://github.com/trifle/twitterresearch

Background Readings:
*Using Digital Trace Data in the Social Sciences:*

Python:
- Swaroop Chitlur: *A Byte of Python*. Available at http://python.swaroopch.com

R:

Data Collection Online:
Sessions Overview

Session 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science, Computational Social Science, Digital Methods, and Big Data
July 18, 2016 Monday—10:00 to 11:30

Session 2: Set Up and Introduction to Collecting Data on Twitter
July 18, 2016 Monday—11:45 to 13:15

Session 3: Introduction to Python
July 19, 2016 Tuesday—13:30 to 15:00

Session 4: Collecting Data Through Twitter’s API
July 19, 2016 Tuesday—15:15 to 16:45

Session 5: Data Lab
July 19, 2016 Tuesday—17:00 to 18:30

Session 6: Loading Twitter Data Into a Database
July 20, 2016 Wednesday—10:00 to 11:30

Session 7: Extracting Data for Typical Analyses
July 20, 2016 Wednesday—11:45 to 13:15

Session 8: Presentation and Discussion of Students’ Research Projects Pt. 1
July 21, 2016 Thursday—08:15 to 09:45

Session 9: Presentation and Discussion of Students’ Research Projects Pt. 2
July 21, 2016 Thursday—10:00 to 11:30

Session 10: Presentation and Discussion of Students’ Research Projects Pt. 3
July 21, 2016 Thursday—11:45 to 13:15

Session 11: Data Lab
July 22, 2016 Friday—08:15 to 09:45

Session 12: Where to take it from here? Discussion of Open Questions and Paper
July 22, 2016 Friday—10:00 to 11:30
Detailed Session Plan and Suggested Readings

Session 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science, Computational Social Science, Digital Methods, and Big Data
July 18, 2016 Monday—10:00 to 11:30

Required Readings:

Background Readings:

Session 2: Set Up and Introduction to Collecting Data on Twitter
July 18, 2016 Monday—11:45 to 13:15

Required Readings:
- Jürgens & Jungherr (2016) (pp. 15-20)
Background Readings:

Session 3: Introduction to Python
July 19, 2016 Tuesday—13:30 to 15:00

Required Readings:

Background Readings:
- Swaroop Chitlur: *A Byte of Python*. Available at [http://python.swaroopch.com](http://python.swaroopch.com)

Session 4: Collecting Data Through Twitter’s API
July 19, 2016 Tuesday—15:15 to 16:45

Required Readings:
- Jürgens & Jungherr (2016) (pp. 21-28)

Background Readings:

Session 5: Data Lab
July 19, 2016 Tuesday—17:00 to 18:30

Session 6: Loading Twitter Data Into a Database
July 20, 2016 Wednesday—10:00 to 11:30

Required Readings:
- Jürgens & Jungherr (2016) (pp. 29-41)

Background Readings:

Session 7: Extracting Data for Typical Analyses
July 20, 2016 Wednesday—11:45 to 13:15

Required Readings:
- Jürgens & Jungherr (2016) (pp. 42-79)

Background Readings:
- Garrett Grolemund and Hadley Wickham (2016) *R for Data Science*. O'Reilly Media, Inc. Available at [http://r4ds.had.co.nz](http://r4ds.had.co.nz)
Session 8: Presentation and Discussion of Students’ Research Projects Pt. 1
July 21, 2016 Thursday—08:15 to 09:45

Required Readings:

Background Readings:  
How to Find a Research Questions:

Conceptual Issues in Working with Digital Trace Data:

Case Studies Illustrating Different Approaches to the Use of Twitter data

Session 9: Presentation and Discussion of Students’ Research Projects Pt. 2
July 21, 2016 Thursday—10:00 to 11:30

Session 10: Presentation and Discussion of Students’ Research Projects Pt. 3
July 21, 2016 Thursday—11:45 to 13:15

Session 11: Data Lab
July 22, 2016 Friday—08:15 to 09:45

Session 12: Where to take it from here? Discussion of Open Questions and Paper
July 22, 2016 Friday—10:00 to 11:30
Background Readings:
Using other data sources:

Extending your analytical skill set:

How might you employ these skills outside of academia:
• Fred Benenson. *On to the next 2,271 days...* January 12, 2016.
• Andrew Therriault. “Finding a Place in Political Data Science”. In: *PS: Political Science & Politics* 49.3. 2016. pp. 531-533. doi: 10.1017/S1049096516000925