# Using Digital Trace Data in the Social Sciences: Syllabus

Junprof. Dr. Andreas Jungherr

Room: D 307  
Postfach 85  
Universitätsstraße 10  
78464 Konstanz

Phone: +49-7531-882150  
Email: andreas.jungherr@gmail.com

Office hours: By appointment

Secretary:  
Karin Becker  
Room: D 312  
Phone: +49-7531-882602  
Email: office.cdm@uni-konstanz.de

Course website:  
http://ajungherr.github.io/dtd-konstanz-summer2018

## 1 Course 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.

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: Vertiefungsseminar—6 ECTS  
POL-19650: Seminar—7 ECTS  
POL-19630: Doktorandenseminar—4 ECTS / 6 ECTS (GSDS)

## 2 Housekeeping

The goal of the course is to enable students to independently design a research question based on current theory and field-specific debates, collect data on Twitter, and run a fitting data analysis on said data. The course itself will focus heavily on the practicalities of collecting data and running selected types of analyses on said data. Information,
readings, and example scripts for the respective sessions will be posted on the course website http://ajungherr.github.io/dtd-konstanz-summer2018.

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. The tutorial is freely available on the Social Science Research Network (SSRN) at https://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.


The articles in readings are linked to from the syllabus and should be available to you by using your Uni-Konstanz VPN-access. The books listed are not linked to from the document but should be available to you through the dedicated online services of the respective publishers. Books by O’Reilly, PACKT, and Apress are available through the Proquest Safari-Books Online-Shelf, at http://proquest.techbus.safaribooksonline.de. Access to Safari-Books Online is freely available by using your Uni-Konstanz VPN-access.

3 Requirements

3.1 Course Work

1. Regular and active participation.

2. Independent study of Python and R in order to follow the example scripts.

3. Regular independent follow-up of example scripts introduced in the respective sessions.

4. Presentation: In the last weeks of the course, you will be asked to present the research project forming the basis of your term paper. For this presentation please prepare a slide deck introducing your research question, your motivation, proposing a mechanism linking signals you hope to find on Twitter to your phenomenon of interest, your proposed approach to the data collection, and open questions. This presentation will take approximately 10 minutes and will be followed by a quick round of feedback from the other participants.

3.2 Term Paper

Your grade for the course will be based on a paper that you hand in following the course. For the paper, you will be asked to perform and report an independent data analysis based on data collected on Twitter by you on a research question chosen 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.
3.3 Reading Papers

Although the course will not feature the dedicated discussion of research papers, you are advised to read at least a selection of the papers listed for each session. This will allow you to familiarize with some of the most prominent research papers using digital trace data. This familiarity will allow you to formulate better research questions and facilitate the work on your term paper. Make sure to use the time during the semester to familiarize yourself with the papers listed in the syllabus.

In reading the paper keep the following questions in mind:

1. What is the research question?
2. What is the phenomenon under examination?
3. How explicitly does the author link their research question to concepts in the social sciences?
4. What signals in digital trace data does the author take as indicators of the phenomenon of interest?
5. What is the mechanism that leads the author to expect signals found in digital trace data to be linked to the phenomenon of interest?
6. To which literature does the article contribute?
7. What is the method of analysis? How are data collected? How appropriate are the method and data?
8. What are the empirical findings? How convincingly are they linked to the phenomenon ostensibly of interest? How transparent is the author in discussing limitations arising from the mediated nature of digital trace data?
Further, in assessing causal claims put forward in papers using digital trace data make sure to ask the following questions:\footnote{1}  

1. Is there a plausible mechanism for the effect?  
2. Does evidence come from peer-reviewed sources?  
3. Are all relevant studies considered?  
4. Are results of specific studies misrepresented?  
5. Are causal claims based on experiment, correlation or analogy?  
6. Is technical, scientific terminology used to obfuscate rather than clarify?  

You may find it useful to keep notes on the papers read by you.

4 Course Outline

Class will meet at the following times and locations:

Thursday 15:15-16:45 (C421)

4.1 Week 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science (April 19)
4.2 Week 2: Set Up and Introduction to Collecting Data on Twitter (April 26)
4.3 Week 3: Introduction to Python (May 3)
4.4 Week 4: Christi Himmelfahrt (May 10)—no meeting
4.5 Week 5: Collecting Data Through Twitter’s API (May 17)
4.6 Week 6: How to Find A Research Question? & Data Lab (May 24)
4.7 Week 7: Fronleichnam (May 31)—no meeting
4.8 Week 8: Loading Twitter Data Into a Database (June 7)
4.9 Week 9: Sample Analyses: Counts & Time Series (June 14)
4.10 Week 10: Sample Analyses: Networks (June 21)
4.11 Week 11: Independent Study and Preparation of Presentations (June 28)—no meeting
4.12 Week 12: Student Presentations I. (July 5)
4.13 Week 13: Student Presentations II. (July 12)
4.14 Week 14: Student Presentations III. & Where to take it from here? Discussion of Open Questions and Paper (July 19)
4.1 Week 1: Introduction and Conceptual Issues in the Use of Digital Trace Data in Social Science (April 19)

Required Readings:

Background Readings:
4.2 Week 2: Set Up and Introduction to Collecting Data on Twitter (April 26)

**Required Readings:**

**Background Readings:**

4.3 Week 3: Introduction to Python (May 3)

**Required Readings:**

**Background Readings:**
4.4 Week 4: Christi Himmelfahrt (May 10)—no meeting
This is an official holiday, so no course meeting on this day. Maybe use the time to further familiarize yourself with Python in independent study.

4.5 Week 5: Collecting Data Through Twitter’s API (May 17)
Required Readings:

Background Readings:

4.6 Week 6: How to Find A Research Question? & Data Lab (May 24)
Required Readings:

Background Readings:
How to Find a Research Question:

**Conceptual Issues in Working with Digital Trace Data:**

**Case Studies Illustrating Different Approaches to the Use of Twitter data:**


Public Datasets:

4.7 Week 7: Fronleichnam (May 31)—no meeting
This is an official holiday, so no course meeting on this day.

4.8 Week 8: Loading Twitter Data Into a Database (June 7)

Required Readings:

Background Readings:

4.9 Week 9: Sample Analyses: Counts & Time Series (June 14)

Required Readings:

Background Readings:
4.10 Week 10: Sample Analyses: Networks (June 21)

Required Readings:

Background Readings:

4.11 Week 11: Independent Study and Preparation of Presentations (June 28)—no meeting

4.12 Week 12: Student Presentations I. (July 5)

In Week 6, students commit to a date for their presentation.

4.13 Week 13: Student Presentations II. (July 12)

4.14 Week 14: Student Presentations III. & Where to take it from here? Discussion of Open Questions and Paper (July 19)

Using other data sources:

Extending your analytical skill set:
How might you employ these skills outside of academia: