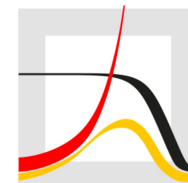




**EDSD 2022**  
ACCESSING AND ANALYSING WEB AND SOCIAL MEDIA  
DATA

Tom Theile

Lab of digital and computational  
demography



## ABOUT MYSELF

Software developer at the lab of digital  
and computational demography

[https://www.demogr.mpg.de/en/about\\_us\\_6113/staff\\_directory\\_1899/tom\\_theile\\_3933/](https://www.demogr.mpg.de/en/about_us_6113/staff_directory_1899/tom_theile_3933/)

<https://github.com/tomthe>

<https://twitter.com/TheileTom>

<https://btw21.deinwal.de/>



# INTRODUCE YOURSELF, PLEASE!

- What is your name?
- What is your background? (Do you have knowledge relevant to this course already (Scraping, APIs, etc. ?))
- What central question would you like to answer with your research?
- What data would be helpful to answer your question?



## AGENDA

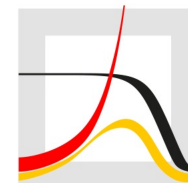
- Day 1
  - What is the Internet?
  - How to scrape data from websites?
- Day 2
  - Finding and working with good datasets
- Day 3
  - Working with APIs
  - Getting data from Twitter
  - Working with Twitter-data
- Day 4
  - Getting data from Facebook audience estimates

Less theory, more practice than other courses!  
Feel free to interrupt me at any time!



## ASSIGNMENTS

- Small assignments everyday
  - Group work possible
  - I will be here for help
  - We will discuss the results together
- Final assignment on Thursday
  - Deadline: Friday 4pm



# WHAT IS THE INTERNET

Continuation in RMarkdown: Day1-1-Introduction-Extracting\_data\_from\_the\_web.pdf

Materials are all available on Github:

[https://github.com/tomthe/EDSD22\\_web\\_and\\_social\\_media\\_data](https://github.com/tomthe/EDSD22_web_and_social_media_data)



THANK YOU FOR  
YOUR ATTENTION!

**Tom Theile**

Software developer at the lab of digital and  
computational demography

[theile@demogr.mpg.de](mailto:theile@demogr.mpg.de)

THANK YOU!

