

# The ABC of Computational Text Analysis

#1 Introduction +  
Where is the digital revolution?

Alex Flückiger

Faculty of Humanities and Social Sciences  
University of Lucerne

25 February 2021

# Lectures #COVID-19

Virtual lectures are not optimal, so ...

- interrupt me anytime!
- use all channels for interaction  
*chat + audio + video + emoji*
- mute mic unless you speak

We are here together! 😊

# Outline

1. digital revolution or hype?
2. about us
3. goals of this course

# AI: A non-standard Introduction

# The world has changed, hasn't it?



A symbolic image of artificial intelligence (HWZ)

# An Era of Big Data + AI

# Group Discussion

**What makes a computer looking intelligent?**

*AI is a moving target with respect to ...*

- *human capabilities*
- *technological abilities*

# Transfer Human Intelligence

from static machines to more flexible devices

- mimicking intelligent behavior  
*speaking + seeing*
- a sense of contextual perception
- many degrees of freedom

# Seeing like a Human?



An image segmentation with Facebook's Detectron2

# Speaking like a Human?

---

## Meena Conversation 4

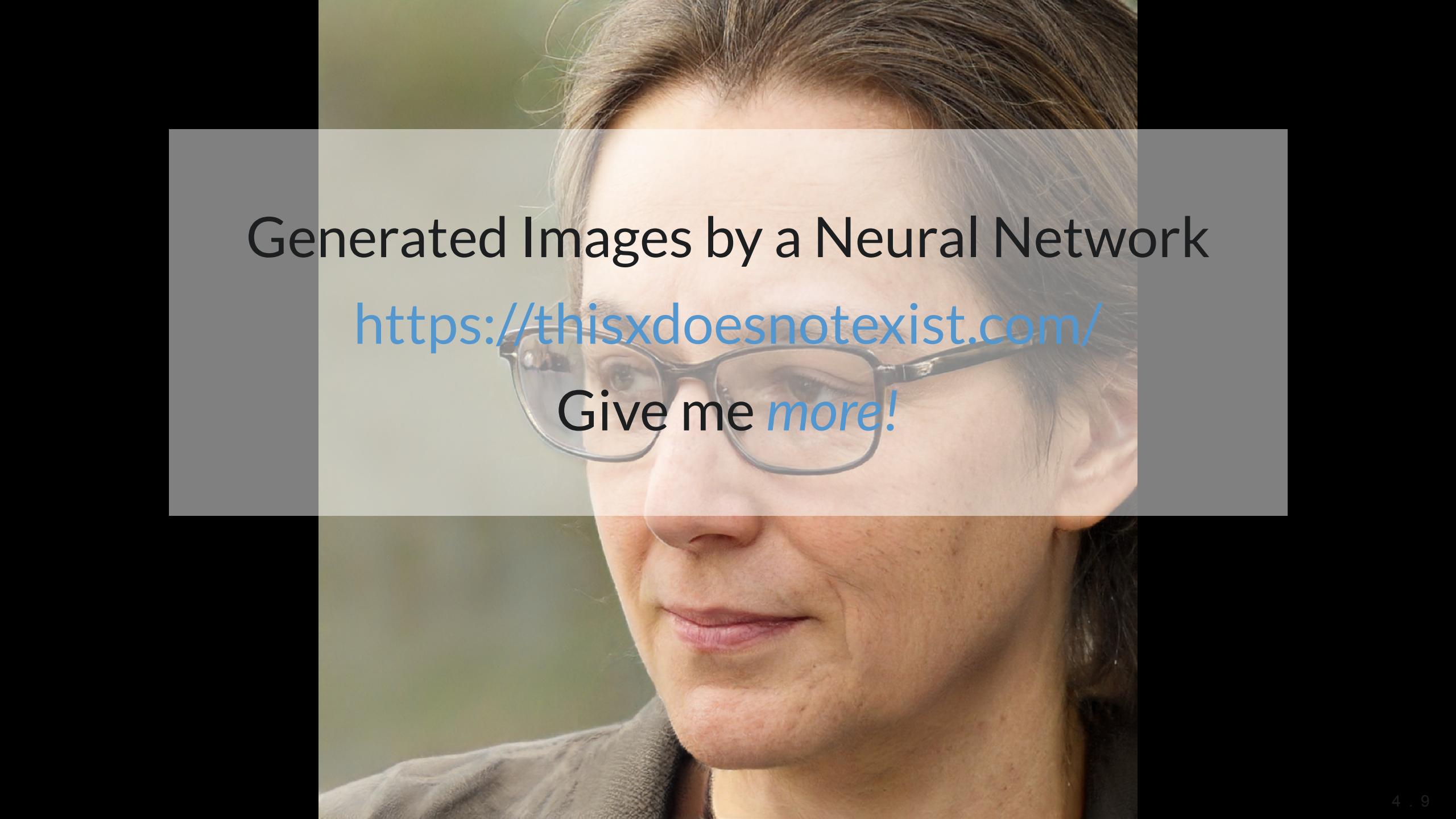
---

**Human:** *Hi!*  
**Meena:** *Hey! How's it going?*  
**Human:** *How's what going?*  
**Meena:** *Pretty good. How's your day been?*  
**Human:** *Excellent. I spent most of trying to learn to surf.*  
**Meena:** *Sweet! What kind of surfing?*  
**Human:** *The kind where I try to catch a wave and fail*  
**Meena:** *Haha. Do you have a surfboard?*  
**Human:** *I borrowed a friend's, to see if I like it. I'm still a beginner.*  
**Meena:** *Ah. I'm sure you'll get the hang of it.*  
**Human:** *Appreciate the vote of confidence :) Have you ever been surfing?*  
**Meena:** *I have! It's actually a ton of fun.*  
**Human:** *Did you find it hard at first?*  
**Meena:** *A little bit, yes. But once I got the hang of it, it was really fun.*  
**Human:** *That's good to know. Do you go surfing regularly?*  
**Meena:** *Yup. I live in southern Arizona, so there's plenty of surfing to be had.*

---

*Chatting with Google's Meena*

# Beyond Perception



Generated Images by a Neural Network

<https://thisxdoesnotexist.com/>

Give me *more!*

# AI Hype in a Nutshell

```
AI = from human_knowledge import solution
```

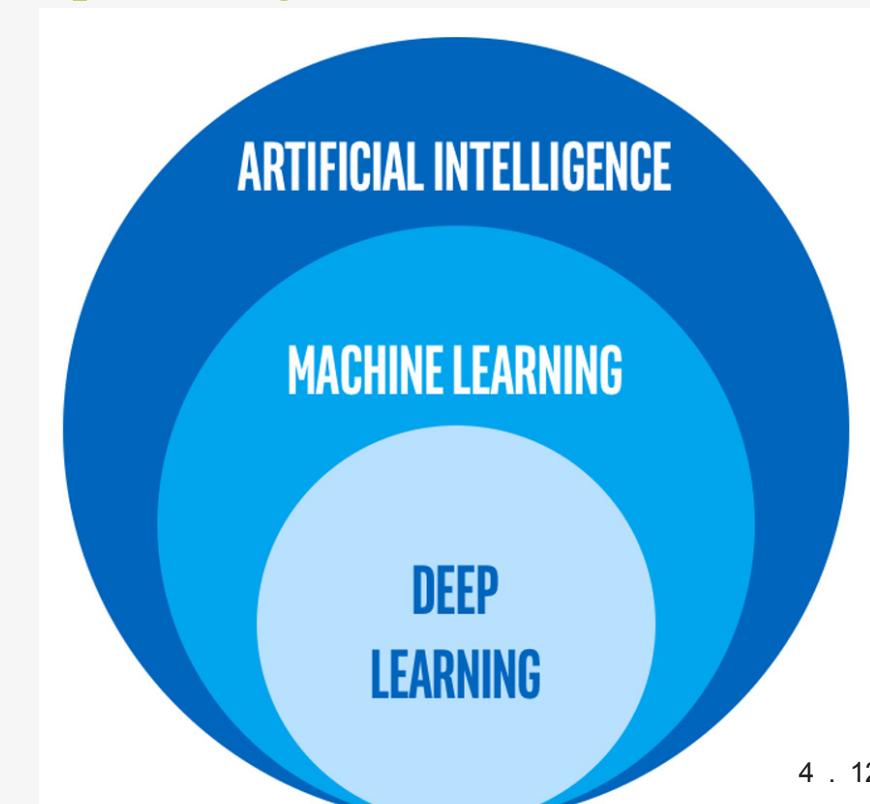
# Artificial Intelligence

## Subfields

- Natural Language Processing (NLP)
- Computer Vision (CV)
- Robotics

# How does Computer Intelligence work?

- interchangeably (?) used concepts  
*Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL)*
- generalize **patterns** from lots of data  
*more recycling than genuine intelligence  
theory agnostically*
- supervised **training**  
*input-output pairs*



# Demos of *intelligent* Text Processing

Can you disenchant them?

- sentiment analysis
- question answering
- creating images from text
- text generation
- chat bot
- machine translation

# Why this matters for Social Science

# Computational Social Science

data-driven research

- computational social science (*Lazer et al. 2009*)  
*Digital Humanities, Computational History, Data Science*
- highly interdisciplinary
- early computational history already in 1960s (*Graham, Milligan, and Weingart 2015*)

# Group Discussion

What kind of data are there?

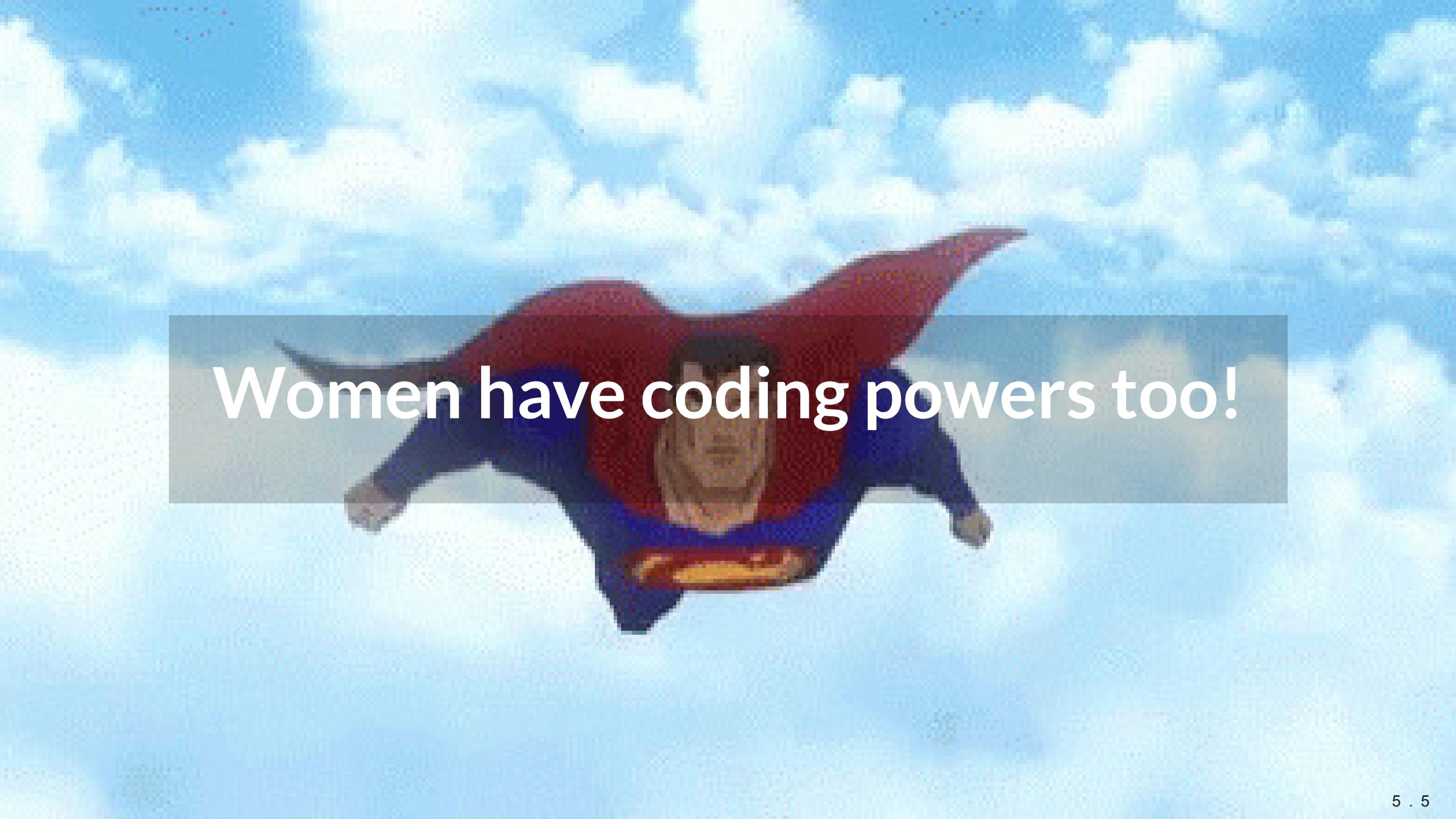
What data is relevant for social science?

- *data as traces of social behaviour*  
*tabular, text, image*
- *datafication*  
*sensors of smartphone, digital communication*
- *much of human knowledge compiled as text*

# About the Mystery of Coding

coding is like...

- cooking with recipes
- superpowers

A woman with short brown hair is flying through a bright blue sky filled with white clouds. She is wearing a superhero costume consisting of a red cape, a blue bodysuit with a red belt, and a gold skirt. Her arms are outstretched, and she has a determined expression. A large, semi-transparent rectangular box covers the middle portion of the image, containing the text.

Women have coding powers too!

# Where the actual Revolution is

Coding is a **superpower** ...

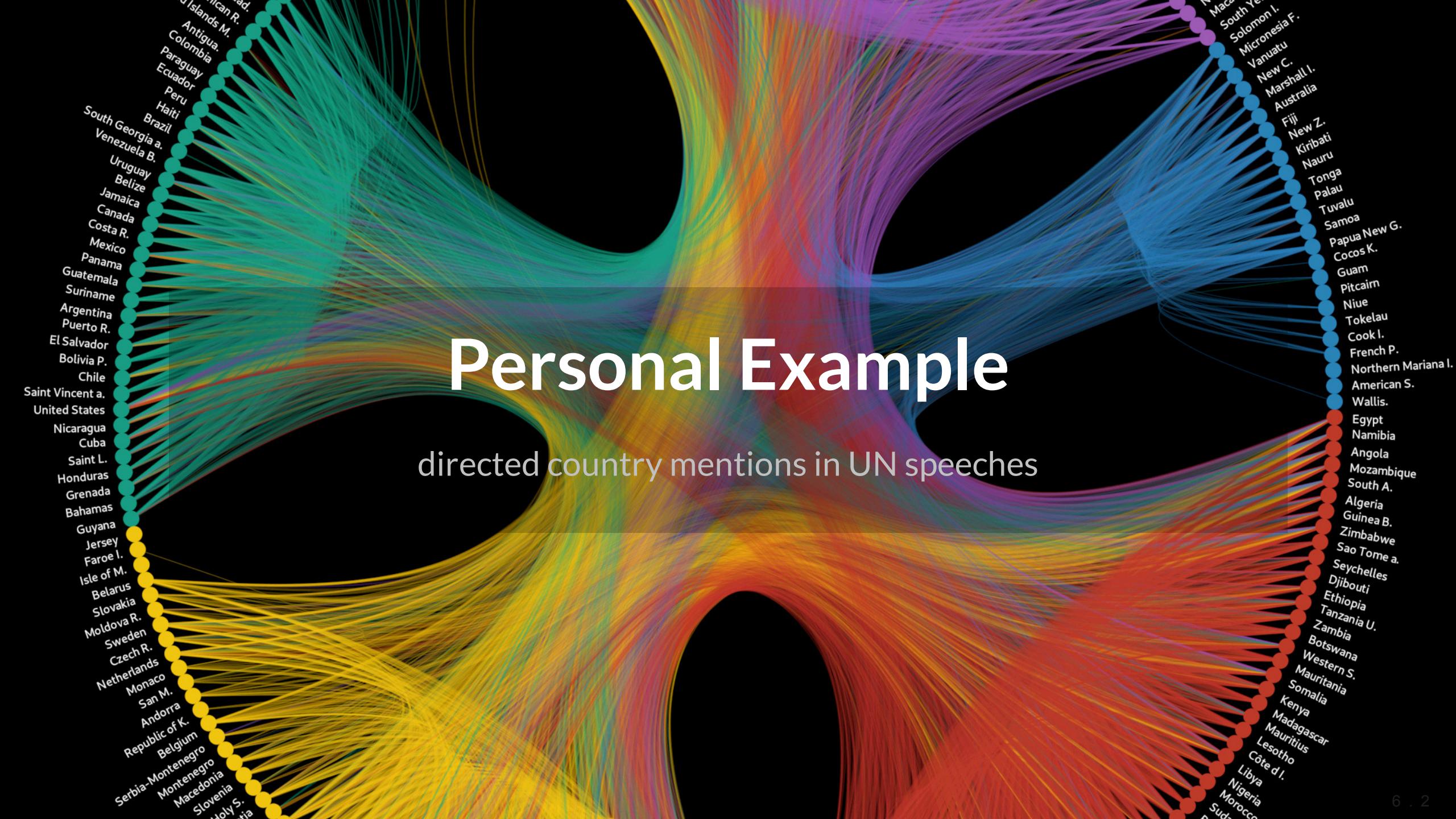
- flexible
- reusable
- reproducible
- inspectable
- collaborative

... to tackle complex problems

# About us

# Personal Example

# directed country mentions in UN speeches



# Goals of this Course

# What you learn

- computationally analyze, interpret, and visualize texts  
*command line + Python*
- digital literacy + scholarship
- problem-solving capacity

# Learnings from previous Courses

- too much content, too little **practice**
- programming can be overwhelming
- **learning by doing**, doing by **googling**

# Levels of Proficiency

1. **awareness** of today's computational potential
2. **analyzing** existing datasets
3. **creating** + analyzing new datasets
4. applying advanced **machine learning**

# What I teach

- computational **practises**
- **critical perspective** on technology
- lecture-style introductions
- hands-on coding sessions
- discussions + experiments in groups

# Topics

## techniques

- text processing
- extracting and aggregating information
- creating simple visualizations
- optical character recognition (OCR)
- scraping files

## data

- existing resources
- creating new resources



inputs are more than welcome!

# Provisional Schedule

Date	Topic
25 February 2021	Introduction + Where is the digital revolution?
04 March 2021	Text as Data
11 March 2021	Setting up your Development Environment
18 March 2021	Introduction to the Command-line
25 March 2021	Basic NLP with the Command-line
01 April 2021	Learning Regular Expressions
08 April 2021	<i>no lecture (Osterpause)</i>
15 April 2021	Advanced RegEx + Data Sources
22 April 2021	Creating new Data Sets + Ethics
29 April 2021	Introduction to Python
06 May 2021	NLP with Python
13 May 2021	<i>no lecture (Christi Himmelfahrt)</i>
20 May 2021	NLP with Python + Working Session
27 May 2021	Mini-Project Presentations + concluding Discussion
03 June 2021	<i>no lecture (Fronleichnam)</i>

# TL;DR



You will be tech-savvy...  
...yet no programmer applying fancy machine learning

# Requirements

- no technical skills required   
*self-contained course*
- laptop (macOS, Win10, Linux)   
*update system*  
*free up at least 15GB storage*  
*backup files*

# Grading



- 3 exercises during semester  
*no grades (pass/fail)*
- mini-project with presentation  
*backup claims with numbers*  
*work in teams*  
*data of your interest*
- optional: writing a seminar paper  
*in cooperation with Prof. Sophie Mützel*

# Organization

- Seminar on Thursday from 2.15pm - 4.00pm
  - weekly virtual live-lectures*
  - recorded sessions on [SWITCHtube](#)*
- slides + data on GitHub: [KED2021](#)
- readings on [OLAT](#)
- communication on [OLAT Forum](#)
  - forum for everything except personal*
  - direct: [alex.flueckiger@doz.unilu.ch](mailto:alex.flueckiger@doz.unilu.ch)*

# Who are you?

Please fill out this questionnaire





Questions?

# Reading

## Required

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323(5915):721-23.

(via OLAT)

## Optional

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Open Draft Version. Under contract with Imperial College Press.

[online](#)

# References

Graham, Shawn, Ian Milligan, and Scott Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Open Draft Version. Under contract with Imperial College Press.  
<http://themacroscope.org>.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. "Computational Social Science." *Science* 323 (5915): 721–23.  
<https://doi.org/10.1126/science.1167742>.