Big Data and Automated Content Analysis
Week 8 – Monday
»Looking back and froward«

Damian Trilling

d.c.trilling@uva.nl
@damian0604
www.damiantrilling.net

Afdeling Communicatiewetenschap
Universiteit van Amsterdam

26 March 2018

Today

1 Looking back
Putting the pieces together
Scaling up

2 Looking forward
rrkjnjrnkr

3 Final steps

Looking back
Putting the pieces together

Looking back | Looking forward | Final steps
○●○○○○○
○
Putting the pieces together

Computational Social Science

"It is an approach to social inquiry defined by (1) the use of large, complex datasets, often—though not always— measured in terabytes or petabytes; (2) the frequent involvement of "naturally occurring" social and digital media sources and other electronic databases; (3) the use of computational or algorithmic solutions to generate patterns and inferences from these data; and (4) the applicability to social theory in a variety of domains from the study of mass opinion to public health, from examinations of political events to social movements"

Shah, D. V., Cappella, J. N., & Neuman, W. R. (2015). Big Data, digital media, and computational social science: Possibilities and perils. The ANNALS of the American Academy of Political and Social Science, 659(1), 6–13. doi:10.1177/0002716215572084
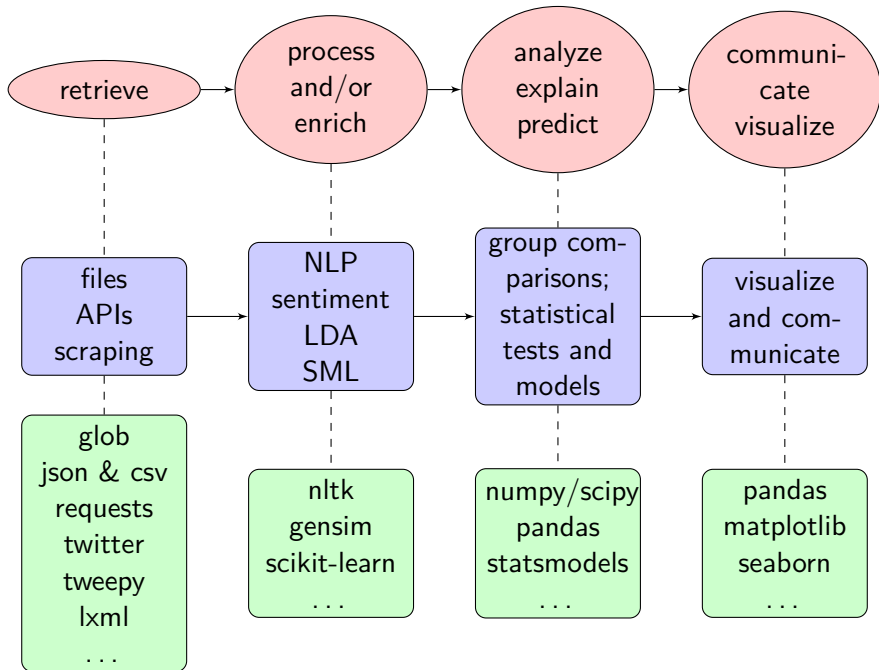
Computational Social Science

"[. . .] the computational social sciences employ the scientific method, complementing descriptive statistics with inferential statistics that seek to identify associations and causality. In other words, they are underpinned by an epistemology wherein the aim is to produce sophisticated statistical models that explain, simulate and predict human life."

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12. doi:10.1177/2053951714528481

Steps of a CSS project

We learned

retrieve → process and/or enrich → analyze explain predict → communicate visualize

| files APIs scraping | NLP sentiment LDA SML | group comparisons; statistical tests and models | visualize and communicate |

glob
json & csv
requests
twitter
tweepy
lxml
. . .

nltk
gensim
scikit-learn
. . .

numpy/scipy
pandas
statsmodels
. . .

pandas
matplotlib
seaborn
. . .

Looking back
Scaling up

rjbejn

functies schrijven
database
resusable

Looking forward
brnrjnkrnjrkn

HIER NOG EEN SLIDE OVER FINAL PROJECT

content...

Next meeting

Wednesday

Final chance for questions regarding final project (if you don't have any, you don't need to come.)

Deadline final exam

Hand in via filesender until Sunday evening, 23.59.