

# Programming, data visualization & AI for academic audiences across institutions and disciplines: lessons learned

**Andres Karjus, PhD** / ERA Chair for Cultural Data Analytics, Tallinn University; Datafigure Ltd

Talk at the "Cross-university collaboration in Digital Humanities & Social Science (DHSS) and Digital Humanities & Cultural Heritage (DHCH) Education" workshop of the DHNB2023 conference

academic: [andreskarjus.github.io](https://andreskarjus.github.io)

twitter&mastodon: [andreskarjus](#)

business: [datafigure.eu](https://datafigure.eu)

*Very quickly, who am I*

Cultural analytics postdoc at Uni Tallinn, CUDAN Lab

Instructor at Datafigure Ltd

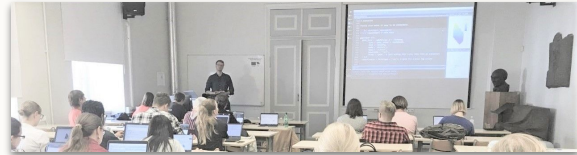
PhD (linguistics) Uni Edinburgh, MSc (AI) from KU Leuven

Digital skills workshops & courses since 2015

Actively since 2017; incorporated in 2021

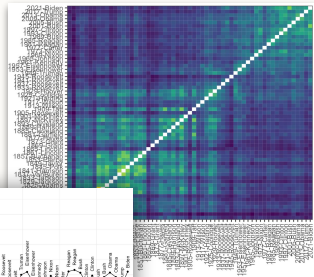
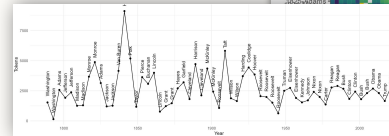
# Digital skills workshops

- Venues: summer schools, conferences, invited lectures, stand-alone events; countries across Europe (Scotland, Wales, Estonia, Latvia, Lithuania, Netherlands, Czech Republic)
- Audience: both (grad) students and academic staff; also private sector
- Disciplines: predominantly (digital) humanities, social sciences; linguistics, media studies



# Types of workshops

- R programming and dataviz, data wrangling (tidyverse, ggplot2)
- Corpus linguistics, basic NLP
- Hackathons with a quick programming workshops
- New: intro to ChatGPT and generative artificial intelligence

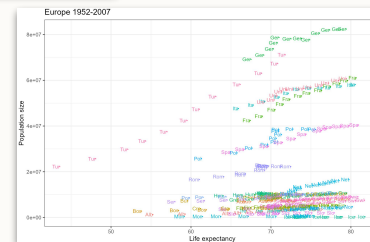
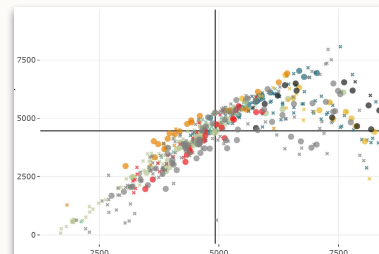
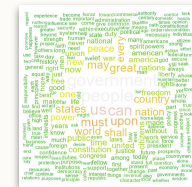
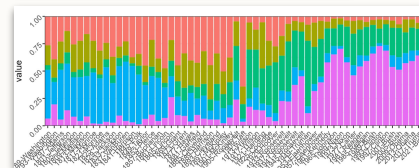
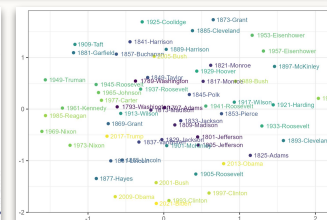
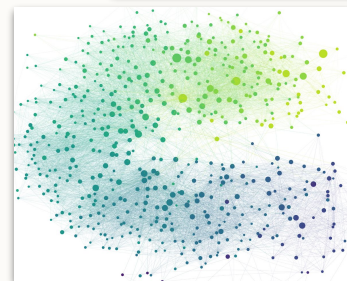


# What's in an R workshop?

- Typically: very short intro, focus on practical guided exercises
- Flexible: typically a mix of beginner to moderately skilled participants
- Students encouraged to share and discuss solutions
- From zero to knowing basic coding after 1-2 days of workshop is common outcome
- Materials open-source (cf. datafigure.eu)

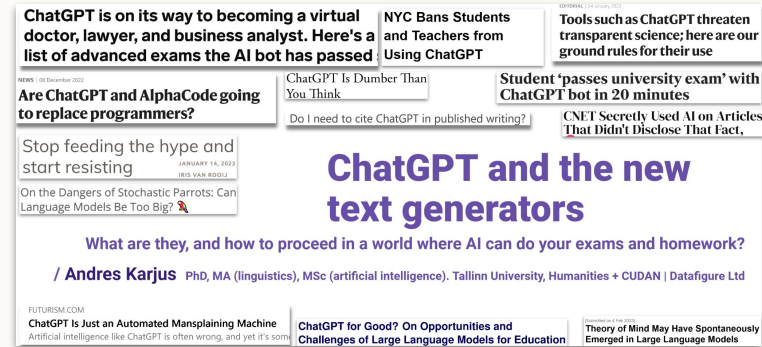
## Basics of R for data visualization & exploration

Andres Karjus



# Generative AI workshops

- Tech like ChatGPT, Bing AI etc. are going to shake the foundations of education and learning, and how people do writing.
- Teaching staff need imminent training



The collage features several news headlines related to ChatGPT: 'ChatGPT is on its way to becoming a virtual doctor, lawyer, and business analyst. Here's a list of advanced exams the AI bot has passed', 'NYC Bans Students and Teachers from Using ChatGPT', 'Tools such as ChatGPT threaten transparent science; here are our ground rules for their use', 'Are ChatGPT and AlphaCode going to replace programmers?', 'ChatGPT Is Dumber Than You Think', 'Student 'passes university exam' with ChatGPT bot in 20 minutes', 'Do I need to cite ChatGPT in published writing?', 'CNET Secretly Used AI on Articles That Didn't Disclose That Fact.', 'Stop feeding the hype and start resisting', 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', 'What are they, and how to proceed in a world where AI can do your exams and homework?' by Andres Karjus, and 'ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education'. The central article is titled 'ChatGPT and the new text generators'.

## The near future / now

- Students use these services anyway.
- No guidance; so students may erroneously use as a knowledge base/wikipedia, which it is not.
- Students can see the AI is often a faster (better?) writer than them. How will that affect their learning and wellbeing?
- The job market will require generative AI skills very soon across various disciplines (the productivity increase will likely drown out any concerns about ethics, bias, and misinformation)
- Increasing AI inequality (in the access & skill gap sense)



# Pros and cons of extracurricular private sector activities like workshops or consulting

Pros: Guaranteed to meet some interesting & very smart people. Some will be in your (or adjacent) fields, may offer academic collaborations too (but these often don't pan out). Can be an income boost.

Cons: Without careful balancing will siphon your research time. Can make you question why you do the same work for a uni for half the money and double the admin.

# Observations

- This is going to be (uncharacteristically) qualitative
- Obviously not completely unbiased, as the people who come to me are the ones in need of training



# Observations

- Humanities & socsci students (incl phd) and research staff in many European institutions still lack proper training in digital skills, basic coding, basic stats
- Yet many (most?) of them deal with some data, and need to engage in some form of data wrangling, cleaning and analytic tasks at some point of the research process

# Observations

- These tasks are *immeasurably* easier if one knows even basic programming.
- Knowing basic stats helps to at least know when to seek and how to ask for help
- Digital skills often seen as something extra, external to and not *really* necessary in hum/socsci
- Suboptimal practices of data, experimentation and writing handed down from supervisors to students

# Observations

- Unimaginable number of work-hours spent on entirely automatable tasks
- Although, good examples: Edinburgh Uni Writing Centre, CDCS training events; Baltic Digital Humanities summer school series
- But training should start in *undergrad*. Not a waste of time, all transferable skills.

# Basic example: reference managers, style

Do?

- Do references by hand for every paper + manually reformat
- Manually format paper style for journal
- Manually reference tables&figures (& change everything if adding a new one to the front)

# Basic example: reference managers, style

Do?

- Do references by hand for every paper + manually reformat
- Manually format paper style for journal
- Manually reference tables&figures (& change everything if adding a new one to the front)

Or:

- Use a ref manager (e.g. Zotero)
- Collect refs using a one-click plugin
- Cite and auto-generate ref list in Latex/Markdown or at least a Word/Docs plugin
- Automate, style, refs to tab/fig in Latex/markdown (or at least as much as you can in Word/Docs)

# Example: data cleaning & wrangling

Do?

- spend countless manual hours in Word/Excel
- leaving no trace of the process and decision making

# Example: data cleaning & wrangling

Do?

- spend countless manual hours in Word/Excel
- leaving no trace of the process and decision making

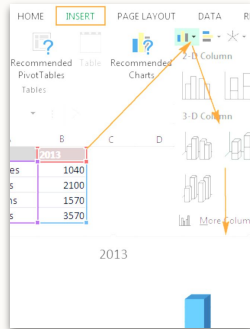
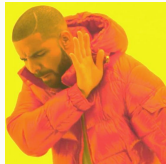
Or:

- use basic programming (e.g. R tidyverse) and automate repetitive tasks (e.g. regex for text)
- reproducible workflow, clear what steps were taken

# Example: data visualization

Do?

- click million buttons in Excel/Tableau etc
- click all the buttons again if you need to redo the plot

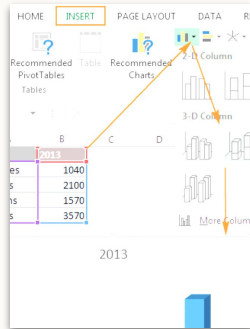
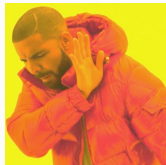




# Example: data visualization

Do?

- click million buttons in Excel/Tableau etc
- click all the buttons again if you need to redo the plot



Or:

- write a small block of code (e.g. R ggplot2)
- create a fully reproducible graph
- probably will look nicer too

```
ggplot(data, aes(x=event))+  
  geom_hist()+  
  theme_bw()
```



# Reasons to rethink your curriculum

- Why not teach people how to be efficient
- Why not teach young people transferable skills
- Digital skills are conducive to Open Science
- But also, the machines are coming for you.  
No, seriously.

# Reasons to rethink your curriculum, 2

- Analyzing text, writing and creativity have been one of the last safe bastions of humanities & social science
- ChatGPT etc is now approaching human level in those.
- Don't be lulled into false security by those making fun of generative AI's small yet to be ironed out flaws and mistakes. These are just early clunky beta versions.
- But also, learning programming has *never ever* been easier than now (free materials, Stackoverflow, ChatGPT/Copilot)

# Conclusions

Let's teach people skills that are useful & efficient

From undergrad to the tenured professor

If you do workshops, make sure they are flexible

Anybody can learn new things, in my experience

Thanks!

academic: [andreskarjus.github.io](https://github.com/andreskarjus)

twitter&mastodon: [andreskarjus](#)

hire me: [datafigure.eu](https://datafigure.eu)