A developer amongst (data) journalists

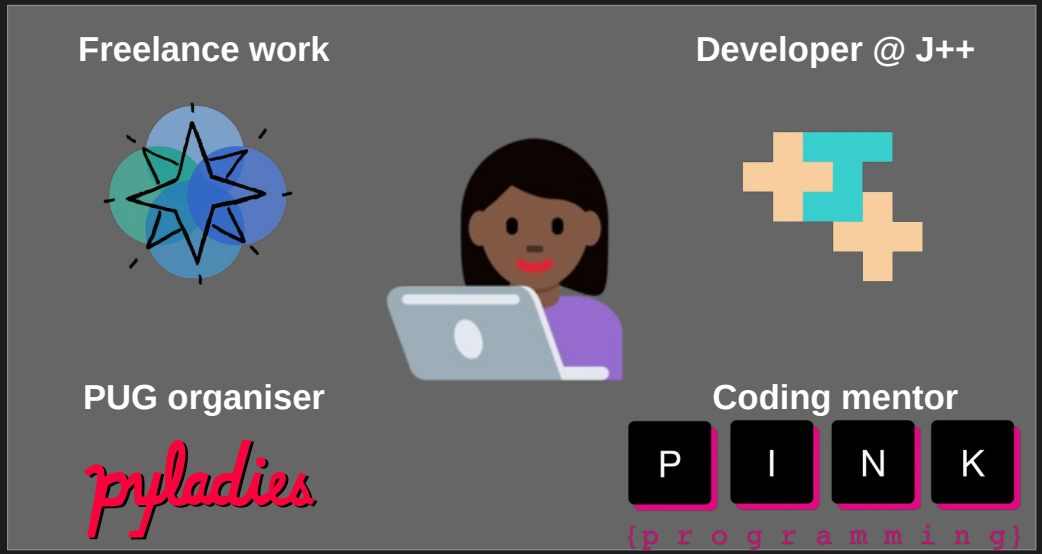
Éléonore Mayola

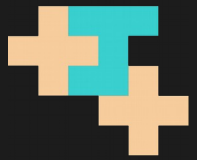
@EleonoreMayola
github.com/eleonore9
Elle-est-au-nord.com
jpluplus.org

Web developer + Data scientist

digital agency, early startup, consultancy

Python, Clojure

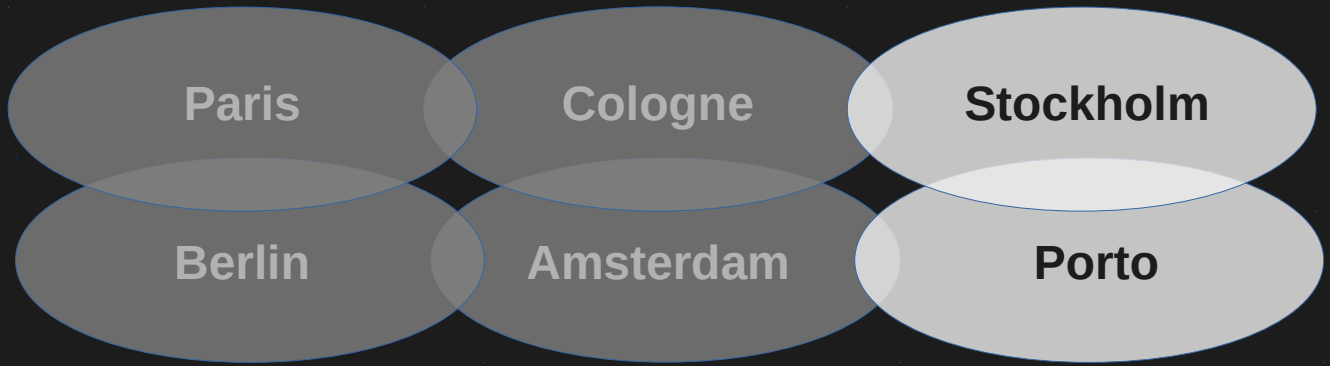




Journalism Plus Plus (J++)

jplusplus.org

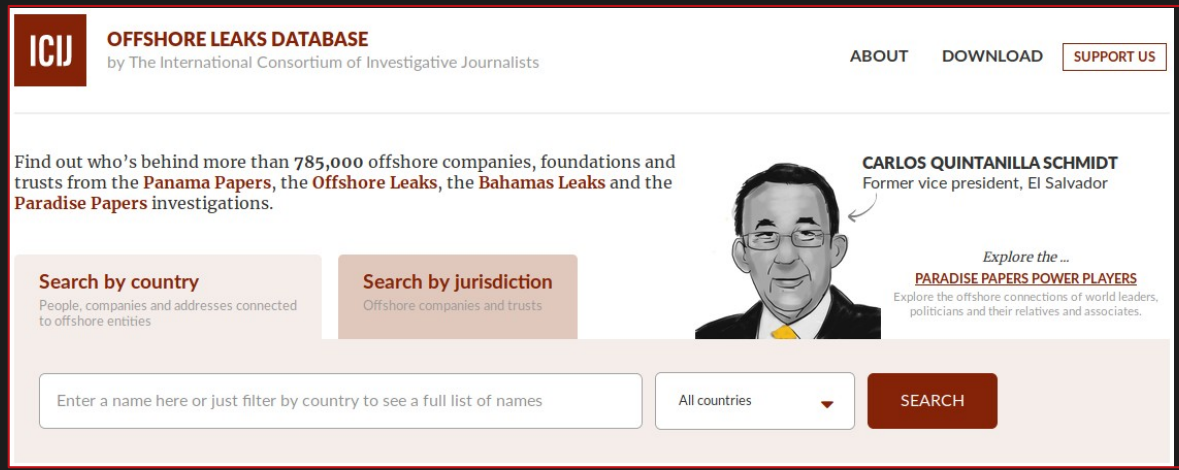
team of data journalism specialists



Data journalism

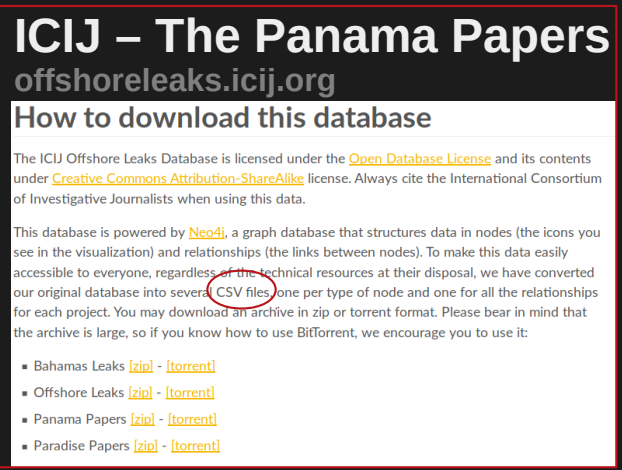
- **Identifying / investigating stories**
  web scraping, APIs, data analysis

- **Presenting / telling stories**
  web app, interactive data visualisation



ICIJ – The Panama Papers

offshoreleaks.icij.org

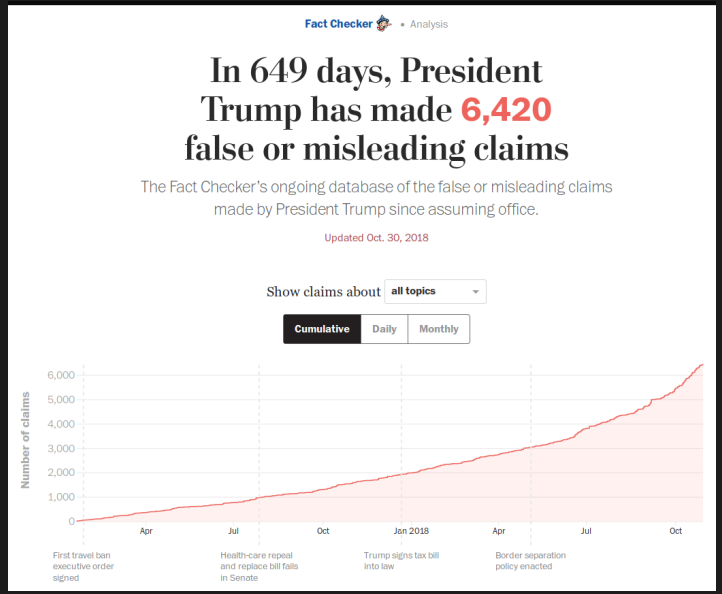


How to download this database

The ICIJ Offshore Leaks Database is licensed under the Open Database License and its contents under Creative Commons Attribution-ShareAlike license. Always cite the International Consortium of Investigative Journalists when using this data.

This database is powered by Neo4j, a graph database that structures data in nodes (the icons you see in the visualization) and relationships (the links between nodes). To make this data easily accessible to everyone, regardless of the technical resources at their disposal, we have converted our original database into several CSV files, one per type of node and one for all the relationships for each project. You may download an archive in zip or torrent format. Please bear in mind that the archive is large, so if you know how to use BitTorrent, we encourage you to use it:

- Bahamas Leaks [zip] - [torrent]
- Offshore Leaks [zip] - [torrent]
- Panama Papers [zip] - [torrent]
- Paradise Papers [zip] - [torrent]

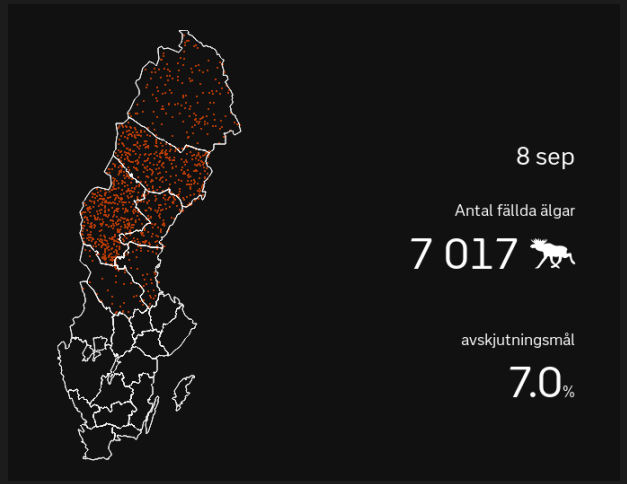


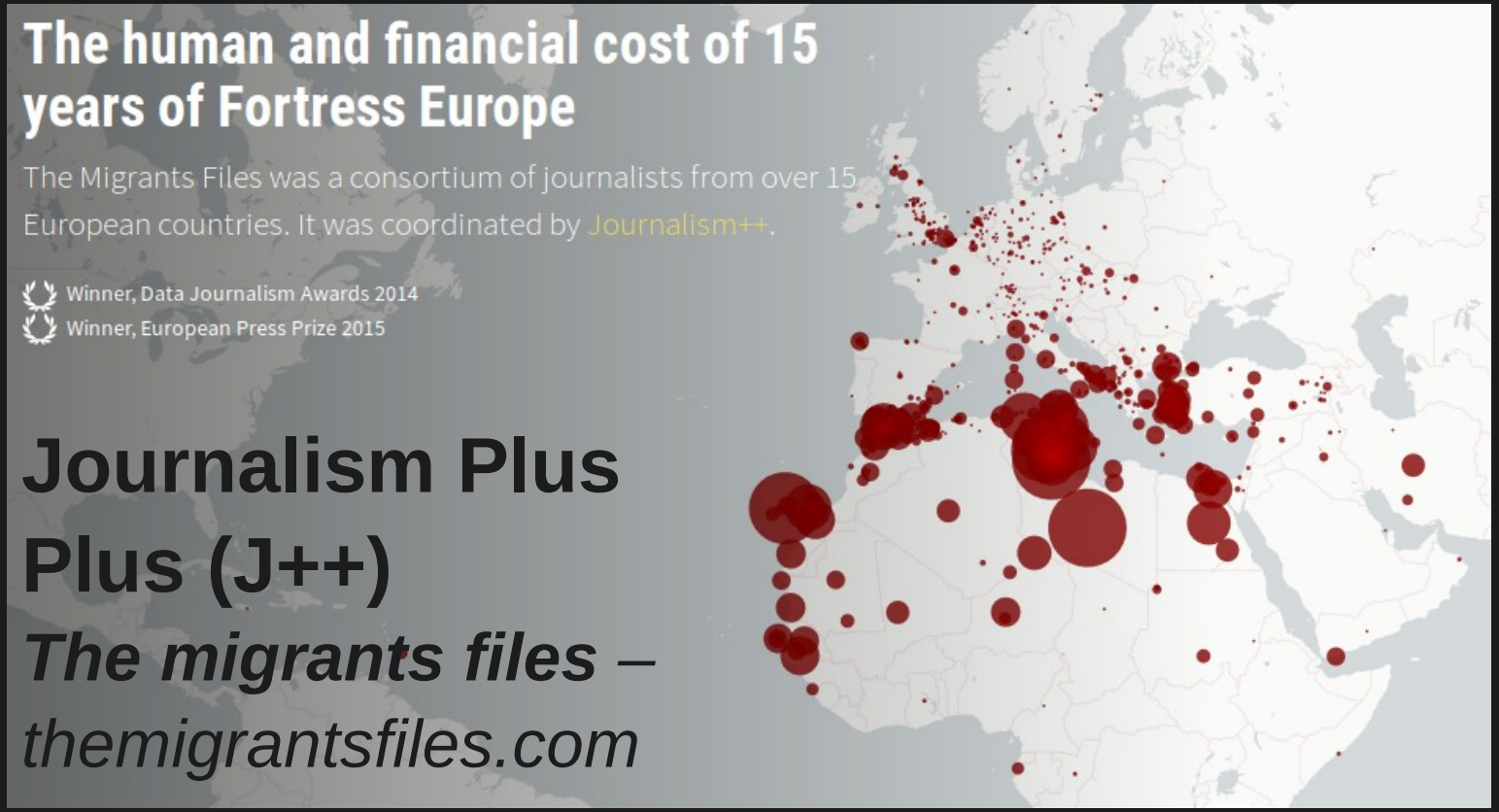
The Washington Post

washingtonpost.com/graphics/politics/trump-claims-database

SVT

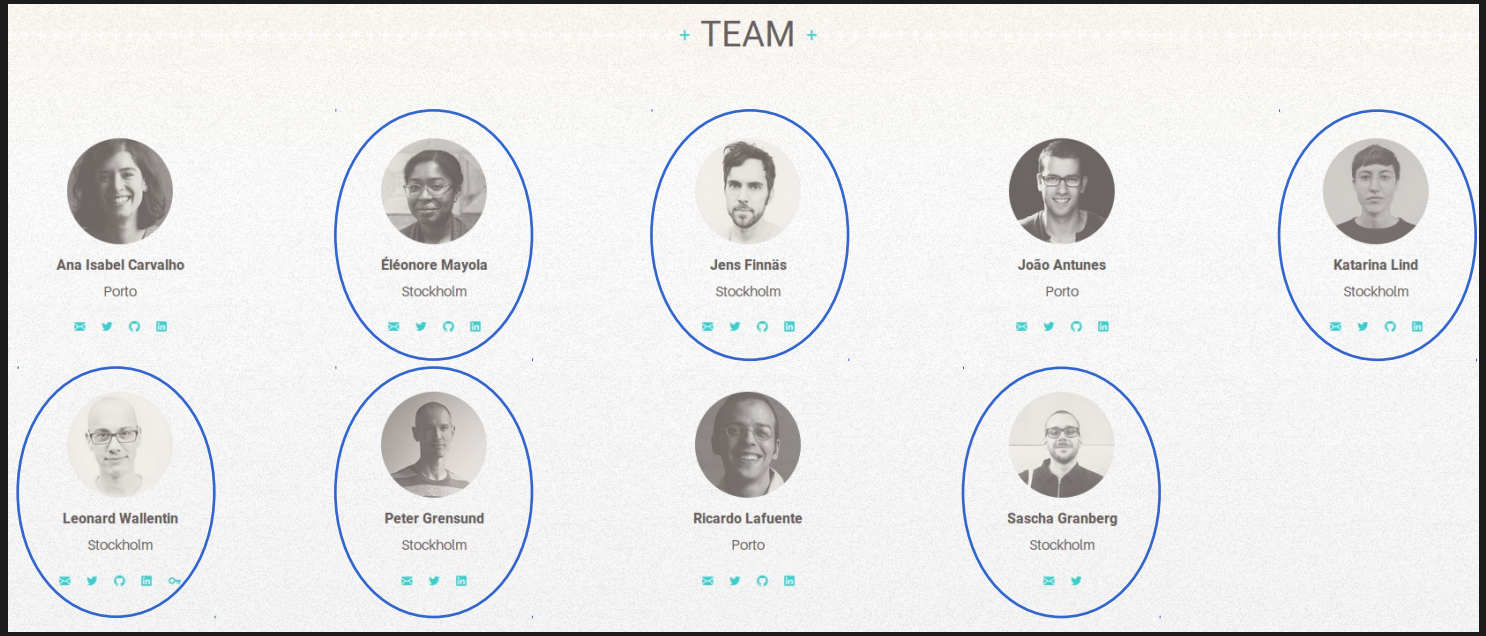
svt.se/special/algjakten-i-siffror




The human and financial cost of 15 years of Fortress Europe
The Migrants Files was a consortium of journalists from over 15 European countries. It was coordinated by Journalism++.
Winner, Data Journalism Awards 2014
Winner, European Press Prize 2015
Journalism Plus Plus (J++)
The migrants files – themigrantsfiles.com

Journalism Plus Plus (J++)

Stockholm and Porto offices



Journalism Plus Plus (J++)

Stockholm office

- **Programming for journalists** – Python course
- **Newsworthy** – automated story finding
- **Clients projects** – web applications, data analysis
- **Local/global news stories** – research, analysis, viz

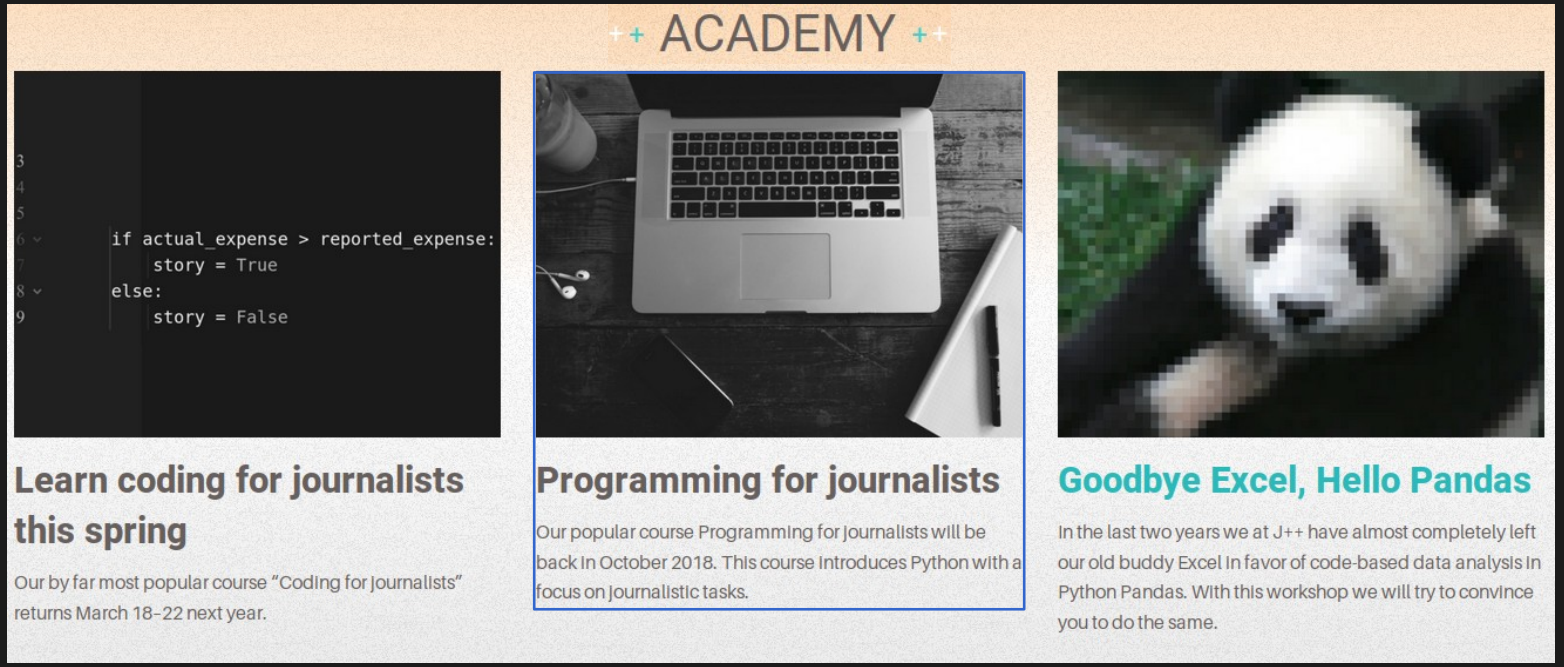
What I’m NOT going to talk about:

- Being a newsroom developer
- How to become a data journalist (although I’m happy to chat about it)

What I’m going to talk about:

- My experience at J++ (teaching Python to journalists, working on internal/client projects)
- My takeaways from that experience

Programming for journalists – Python course



jplusplus.org/en/academy

6 sessions already run → programme set but adapted to each class

Programming for journalists – Flexible scope

Classroom teaching

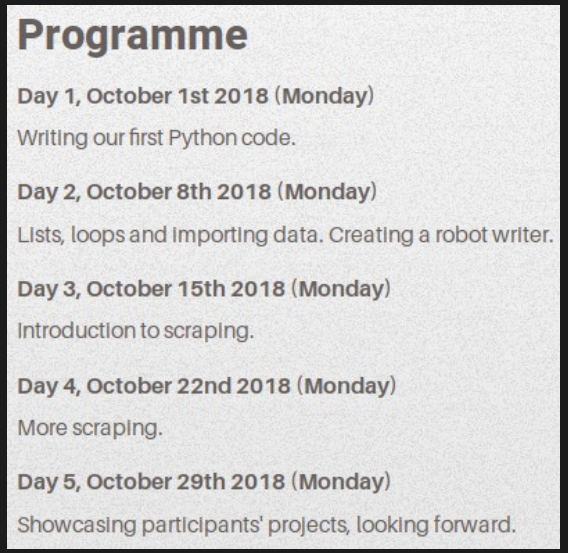
- CLI basics
- Python basic data structures
- *if* statements, *for* loops
- Work with Python libraries, APIs, reading/writing files

Support

- Slack group (24/7)

Applied to journalism

- Generating texts
- Web scraping
- Gather tweets
- Course project



Programme

Day 1, October 1st 2018 (Monday)

Writing our first Python code.

Day 2, October 8th 2018 (Monday)

Lists, loops and importing data. Creating a robot writer.

Day 3, October 15th 2018 (Monday)

Introduction to scraping.

Day 4, October 22nd 2018 (Monday)

More scraping.

Day 5, October 29th 2018 (Monday)

Showcasing participants' projects, looking forward.

jplusplus.org/en/academy/programming-for-journalists

Programming for journalists – Flexible scope

Day	2015/16	2017	2018
Day 1	Variables, data types, conditionals, functions	Variables, data types, conditionals, functions	Variables, data types, conditionals
Day 2	Lists, dicts, loops, more functions	Lists, dicts, loops, functions	Lists, dicts, loops
Day 3	HTML basics, intro to web scraping, read/write files	HTML basics, intro to web scraping, read/write files	HTML basics, intro to web scraping, read/write files
Day 4	Scrape data from a web page, save data in database	Scrape data from a web page	Scrape data from a web page
Day 5	Demo git / Github	Demo git / Github How the web works, HTTP requests, Python tools	Demo git / Github

Programming for journalists – Flexible scope

J++ | Python for Journalists

Day 1	Variables, data types, conditionals
Day 2	Lists, dicts, loops
Day 3	HTML basics, intro to web scraping, read/write files
Day 4	Scrape data from a web page
Day 5	Demo git / Github

MIT | Intro to CS and Programming in Python

1	What is computation?
2	Branching and Iteration
3	String Manipulation, Guess and Check, Approximations, Bisection
4	Decomposition, Abstractions, Functions
5	Tuples, Lists, Aliasing, Mutability, Cloning
6	Recursion, Dictionaries
7	Testing, Debugging, Exceptions, Assertions
8	Object Oriented Programming
9	Python Classes and Inheritance
10	Understanding Program Efficiency, Part 1
11	Understanding Program Efficiency, Part 2
12	Searching and Sorting

Udacity | Programming Foundations with Python

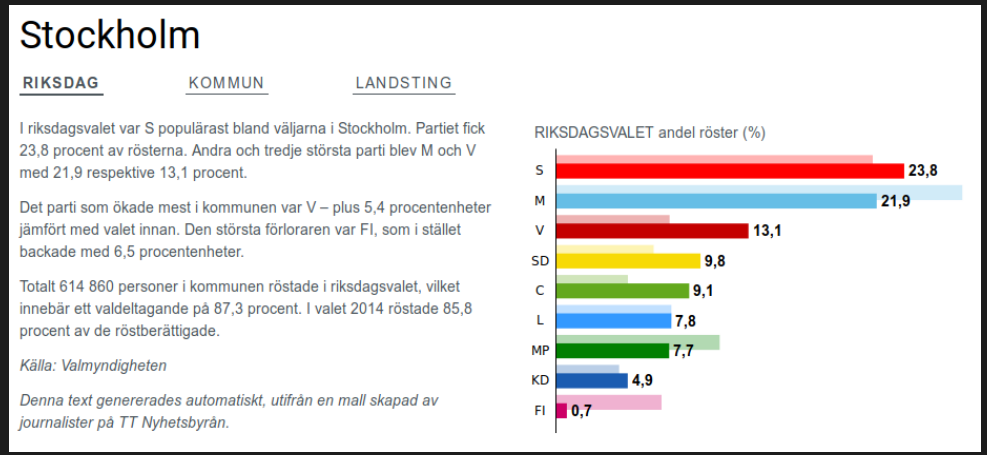
Lesson 1	Intro and installation
Lesson 2	Use functions, explore classes
Lesson 3	OOP concepts, use 3rd party packages
Lesson 4	Create classes
Lesson 5	Personal project

Programming for journalists – Text robots

Read CSV file → Use election results data

FOR loop → Go through cities/regions

Use a dictionary
IF statement
Print statement → Write an article About the results



Stockholm

RIKSDAG KOMMUN LANDSTING

I riksdagsvalet var S populärast bland väljarna i Stockholm. Partiet fick 23,8 procent av rösterna. Andra och tredje största parti blev M och V med 21,9 respektive 13,1 procent.

Det parti som ökade mest i kommunen var V – plus 5,4 procentenheter jämfört med valet innan. Den största förloraren var FI, som i stället backade med 6,5 procentenheter.

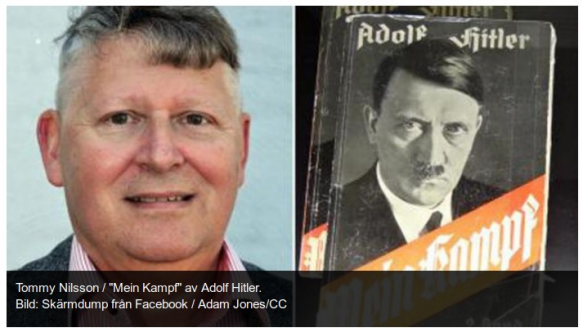
Totalt 614 860 personer i kommunen röstade i riksdagsvalet, vilket innebär ett valdeltagande på 87,3 procent. I valet 2014 röstade 85,8 procent av de röstberättigade.

Källa: Valmyndigheten

Denna text genererades automatiskt, utifrån en mall skapad av journalister på TT Nyhetsbyrån.

sverigesradio.se/sida/artikel.aspx?programid=4657&artikel=7037094

Programming for journalists – Students projects



https://www.etc.se/inrikes/sds-toppnamn-i-kyrkovalet-kopte-nazistbocker-mitt-i-valrorelsen

https://svenska.yle.fi/artikel/2016/10/23/kolla-hur-priserna-pa-bilforsakring-kan-se-ut-dar-du-bor-skillnader-pa-flera

It’s so exciting!

- Keeping things **simple** and **practical**
  - → implement what they learn
- We’ve actually made new members of the Python **community**
- Teaching new **skills**
  - → impact on their career
- We’re having a (indirect) **positive impact** on society

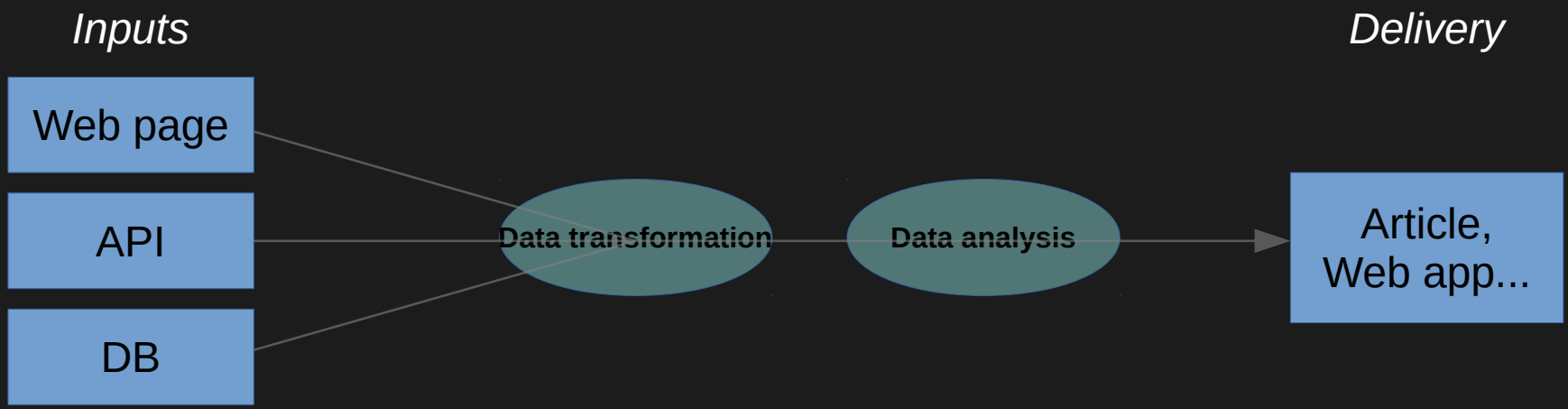
It’s also challenging

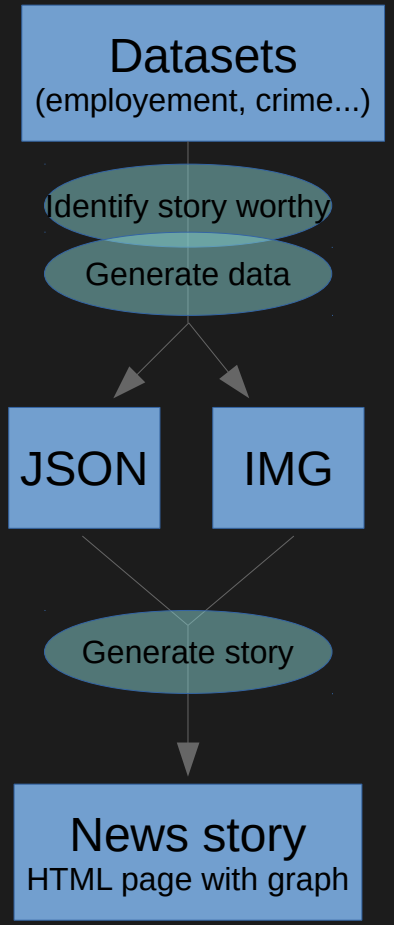
- **Adapting to the different students coding levels**
  - remove things if they need more time
  - give more to more advanced students

- **There are so many tools, notions, libraries**
  - don’t overwhelm them
  - develop a follow-up course

- **Get them to interact with other data journalists and other members of the Python community**
  Alumni network? Local meetup groups?

Data pipelines for internal projects and clients

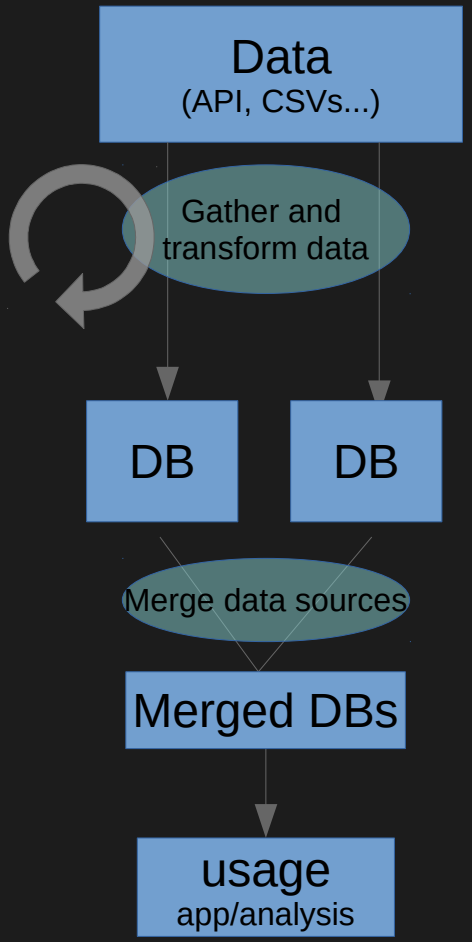




Newsworthy – automated story finding

Node.js, Amazon Web Services

- Internal project based on a statistical model
- Feeds on local or European level databases
- Generates JSON format data + graph images
- Query data to write news stories



Clients projects – web applications, data analysis

Node.js or Python, Postgres, AWS

- Use clients data and/or open data
- Feed a database + add data upon updates
- Query data for applications or analysis

Aim

- Data pipeline fully automated and reusable
- Include monitoring and alerting

Current state

- Data pipeline – not fully automated, differs / projects
- Code organised in a modular fashion
- Research and design for a new pipeline
- Identification of reusable bits

Plan and ideas [in progress]

- Lambda functions and serverless architecture
  → Make use of the modular structure
- Expand our use of AWS services
  → Quick set-up and for securely hosted services
- Independent app for monitoring and alerting system

Being a developer among (data) journalists

Skills involved

- Curiosity, general/domain knowledge
- Communication (document)
- Flexibility (new techs, ideas…)
- Deliver: be simple and practicle (cf Zen of Python :))
- Data work: Python (scripting, Jupyter Notebook, Pandas, Matplotlib….), QGIS
- Web work: JS (Node.js, React.js…), HTML, CSS
- Other: SQL, version control (Git), AWS, APIs...

Methodology and processes

Tools

- **Slack** (communication with Sthlm-Porto, Sthlm only, projects specific, ideas/cool stuffs)
- **Trello** (general planning, projects features)
- **Google docs**

Processes

- Weekly **planning** meeting
- Monthly **update** meeting Sthlm-Porto
- Meetings **project specific** within networks (EDJNET)

(Workplace) interactions

- Think your interactions with “non-technical” people
  - → Take on non-experienced coders
  - → Help non-coders eager to automate their work
  - → Be proactive on helping other professionals
- Balance between giving enough info, but not too much
- Favour knowledge exchange (code, domain knowkedge)

Work opportunities

Tech work != tech organisations

- **Widen your perspectives**
- **Widen your network(s)**
- **Widen your job search**

Science, art, education, travel, media...

Resources

- github.com/Eleonore9/pyconse18
- datajournalismhandbook.org

@EleonoreMayola
github.com/eleonore9
Elle-est-au-nord.com

jpluplus.org | @jplusplus_