

# A developer amongst (data) journalists



Éléonore Mayola

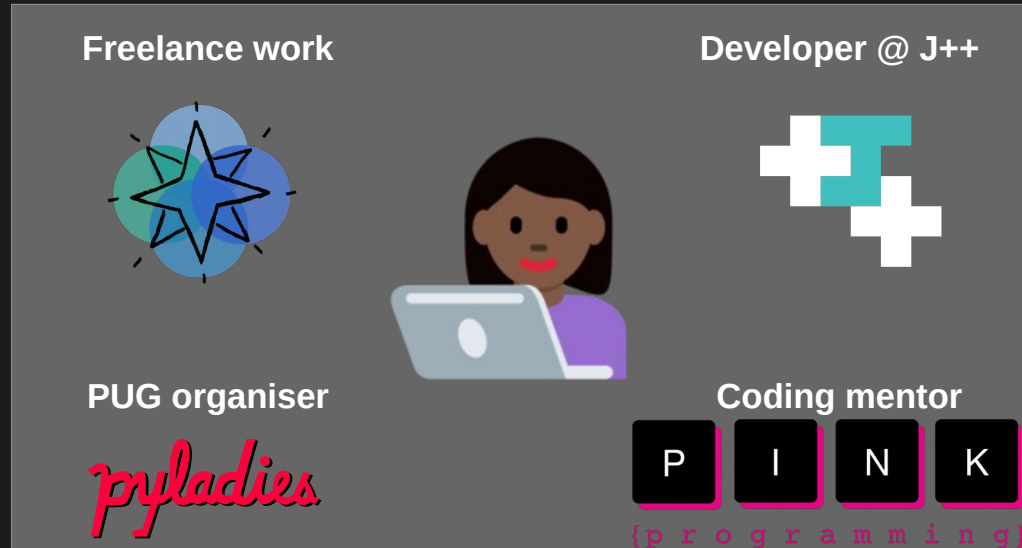


@EleonoreMayola  
github.com/eleonore9  
Elle-est-au-nord.com  
jpluplus.org

# Web developer + Data scientist

*digital agency, early startup, consultancy*

Python, Clojure

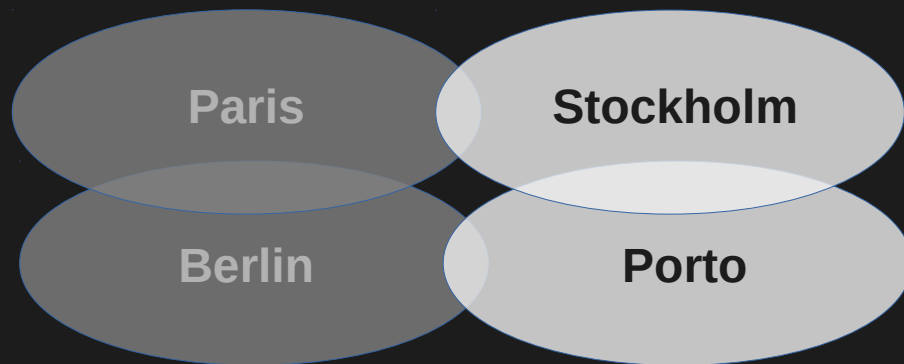




# Journalism Plus Plus (J++)

*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

**OFFSHORE LEAKS DATABASE**  
 by The International Consortium of Investigative Journalists

[ABOUT](#)
[DOWNLOAD](#)
[SUPPORT US](#)

Find out who's behind more than **785,000** offshore companies, foundations and trusts from the **Panama Papers**, the **Offshore Leaks**, the **Bahamas Leaks** and the **Paradise Papers** investigations.

**Search by country**  
 People, companies and addresses connected to offshore entities

**Search by jurisdiction**  
 Offshore companies and trusts to offshore entities

**CARLOS QUINTANILLA SCHMIDT**  
Former vice president, El Salvador

Explore the ...  
**PARADISE PAPERS POWER PLAYERS**  
 Explore the offshore connections of world leaders, politicians and their relatives and associates.

All countries ▼

## ICIJ – The Panama Papers

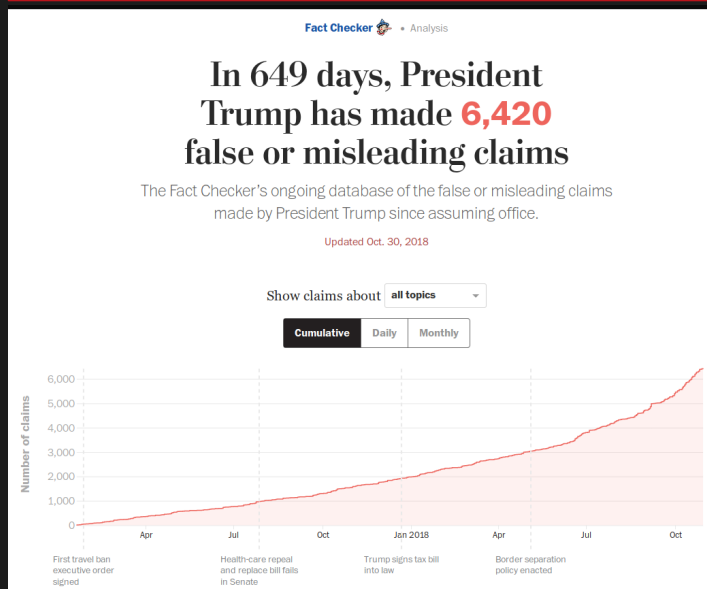
offshoreleaks.icij.org

### 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 [\[zip\]](#) - [\[torrent\]](#)

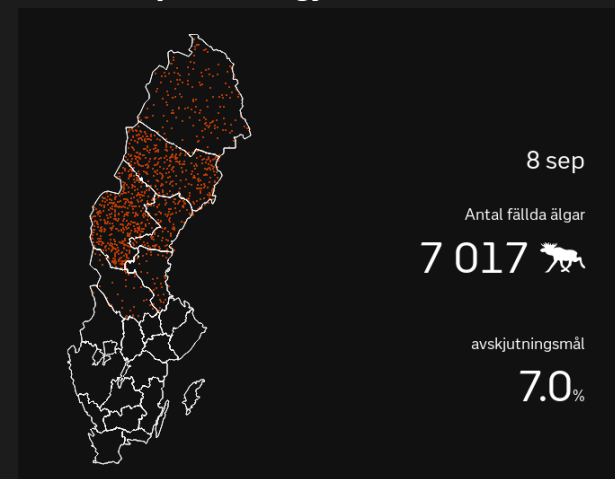


## The Washington Post

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

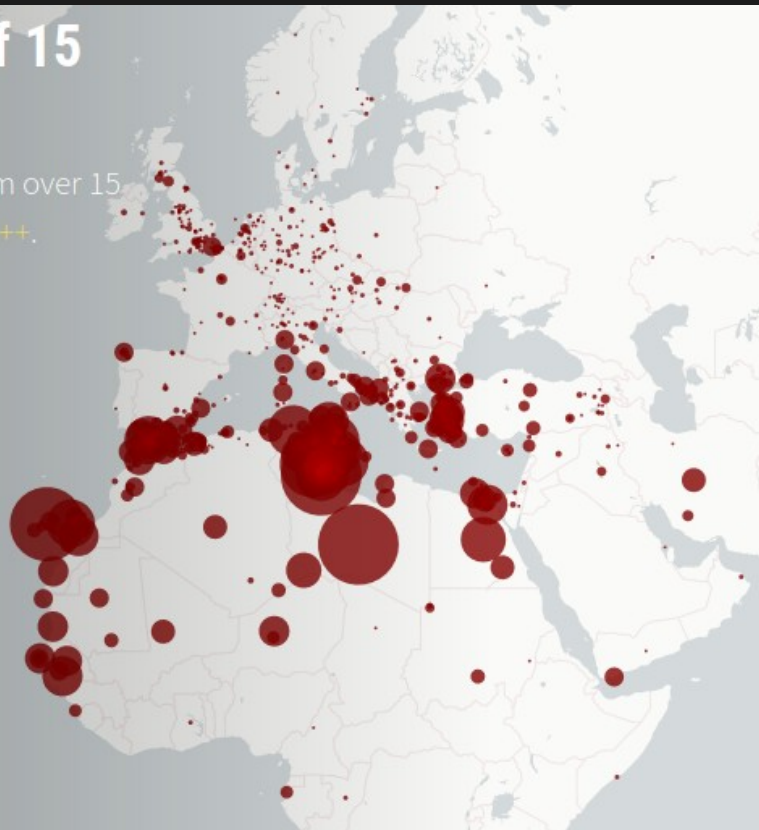
The Migrants Files was a consortium of journalists from over 15 European countries. It was coordinated by Journalism++.



Winner, Data Journalism Awards 2014

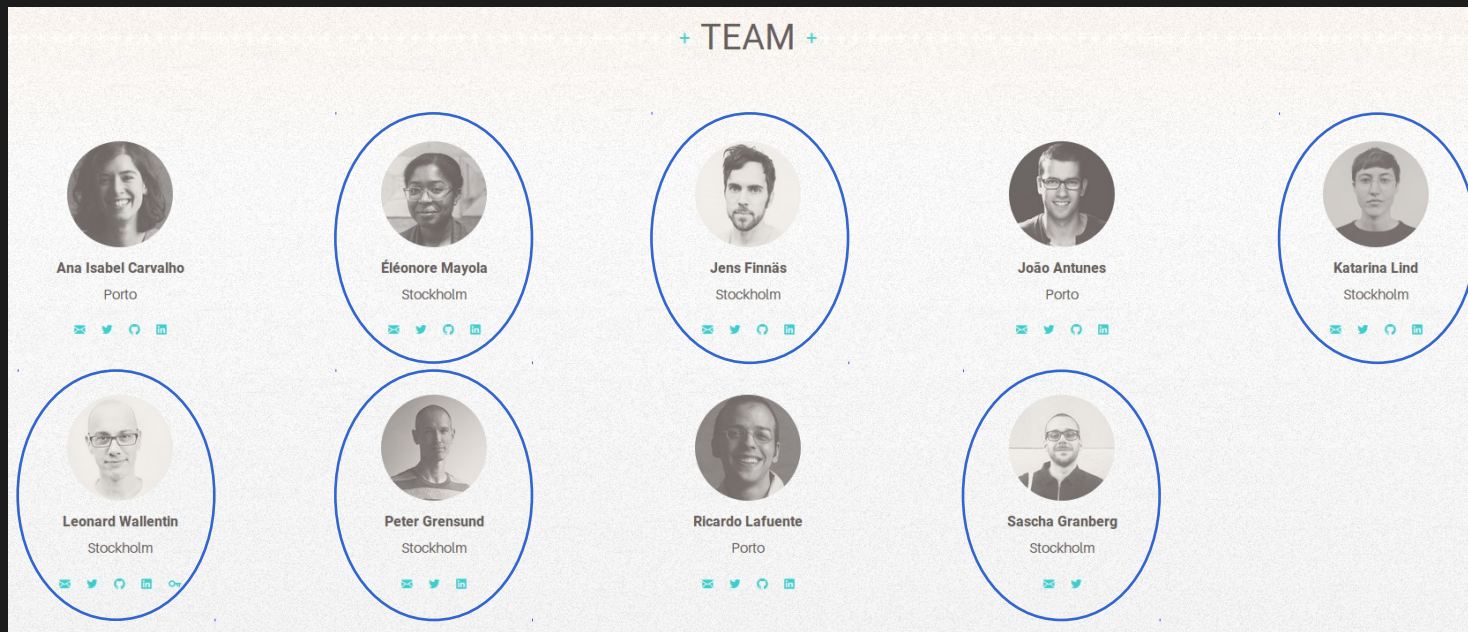
Winner, European Press Prize 2015

**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

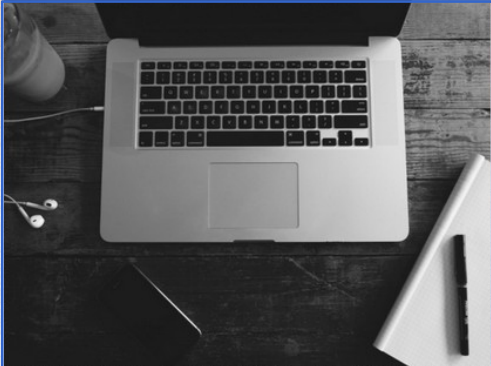
## Programming for journalists – Python course

++ ACADEMY ++




**Learn coding for journalists this spring**

Our by far most popular course "Coding for journalists" returns March 18-22 next year.



**Programming for journalists**

Our popular course Programming for journalists will be back in October 2018. This course introduces Python with a focus on journalistic tasks.



**Goodbye Excel, Hello Pandas**

In the last two years we at J++ have almost completely left our old buddy Excel in favor of code-based data analysis in Python Pandas. With this workshop we will try to convince you to do the same.

[jplusplus.org/en/academy](http://jplusplus.org/en/academy)

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

## 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.

Slide Leo Wallentin

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

How 'unicorns in the newsroom' change the storytelling process

Leading figures from the Guardian and Wall Street Journal detailed the thinking behind their digital projects at the International Journalism Festival in Perugia

Posted: 18 April 2015 By: **Alexander Peil**

Source: **By Omidia, via Flickr, some rights reserved.**

Coming from a print journalism background, Aaron Pilhofer, the Guardian's executive editor for digital, was sheepish "as the first thing" at the newsroom. But such a sentiment has resulted in him working at the very leading edge of digital storytelling in recent years.

After nearly ten years at the New York Times, Pilhofer joined the Guardian last summer, speaking at the International Journalism Festival in Perugia.



### The unicorns of the newsroom

25 April 2015

Share Tweet LinkedIn +1

On the first day of the International Journalism Festival in Perugia a panel discussion was held on the gap between traditional reporting and coding. The international guest speakers from world's leading media described the notion of the 'unicorn', a rare magical creature that lives in the newsroom on the intersection between text and code.

Jacqui Maher, interactive journalist at the BBC News Lab, started the discussion by telling about her journey of becoming a 'unicorn' in the newsroom. She first came into contact with an interactive form of journalism at the New York Times, while working on a project dedicated to finding missing people in the aftermath of the Haiti earthquake.

### THE UNICORNS OF THE NEWSROOM

JOHN CROWLEY | GARRETT GOODMAN | JACQUI MAHER | ARON PILHOFER

So you found a newsroom unicorn – now what?

News organisations need to consider how to integrate data developers and coders – who often come from a very different background – into their 'traditional' newsrooms if they want to survive.

Newsroom Unicorns: Behaviours, habitat and life cycle

10 April 2015, 10:00am - 12:00pm



There are magical creatures who live in the newsrooms at the intersection of journalism and coding.

This is discussed by Omidia, via Flickr, some rights reserved.

## Classroom teaching

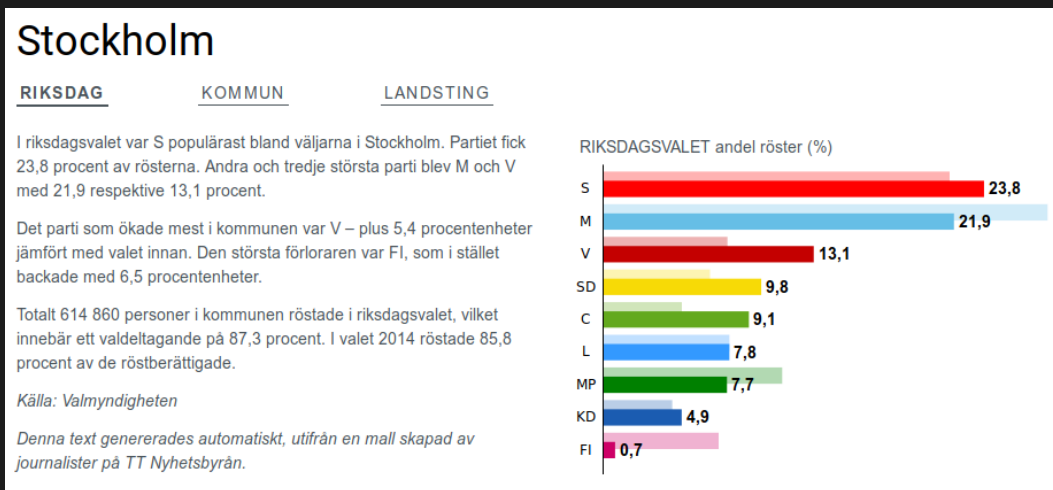
- CLI basics
- Python basic data structures
- *if* statements, *for* loops
- Work with Python libraries, APIs, reading/writing files

## Support

- Slack group (24/7)

## Applied to journalism

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



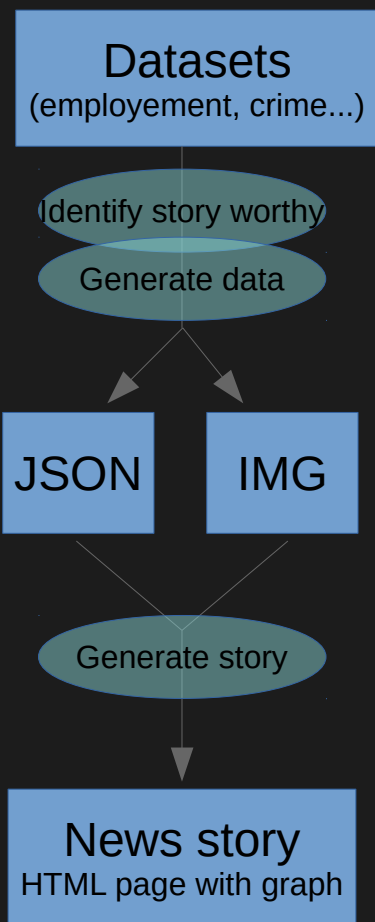
[sverigesradio.se/sida/artikel.aspx?programid=4657&artikel=7037094](https://sverigesradio.se/sida/artikel.aspx?programid=4657&artikel=7037094)

## 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**

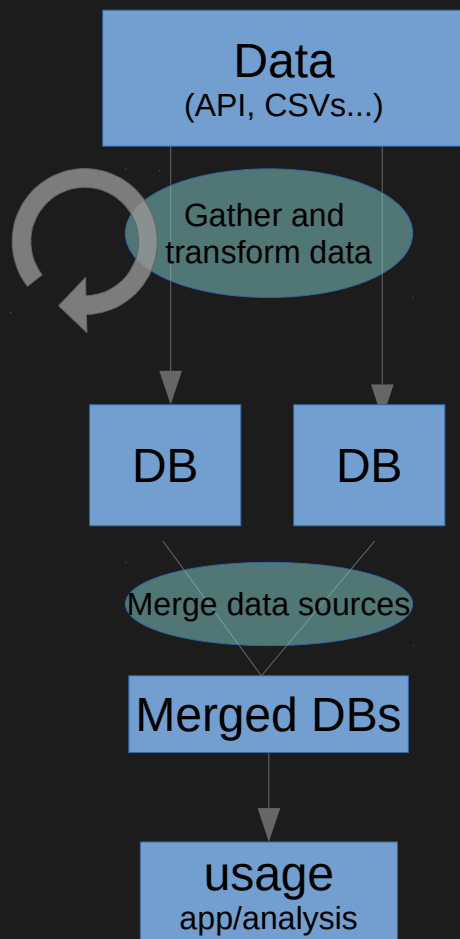
- **Adapting to the different students coding levels**
  - remove things if they need more time
  - give more to more advanced students
- **There are so many tools, notions, libraries**
  - don't overwhelm them
  - develop a follow-up course
- **Get them to interact with other data journalists and other members of the Python community**
  - Alumni network? Local meetup groups?



## 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

## Local/global news stories – research, analysis, viz

### Example: EDJNET - OneDegreeWarmer

- Investigating climate data
- Part of a European network of journalists
- First story (09/2018): temperatures in European cities
- Communication, coordination, data analysis/modelling, text generation, web application

# European Data Journalism Network - EDJNet

europeandatajournalism.eu | @EdjNet



Consortium of data journalists  
throughout Europe

Independent platform for data-  
driven news on European affairs

Open source content in several  
languages

### Europe is warming, rapidly

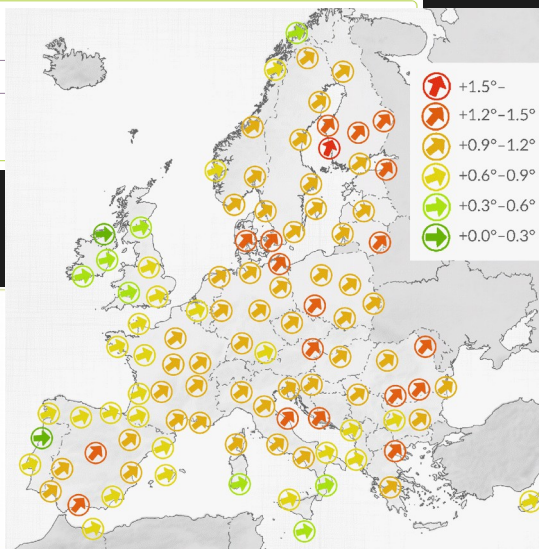
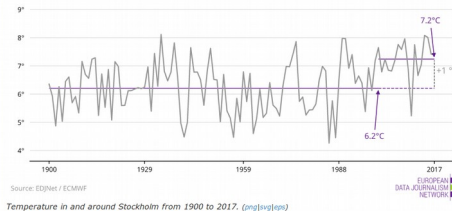
An analysis of over 100 million meteorological data points, from 117 years of weatherdata, shows that every major city in Europe is warmer in the 21st century than it was in the 20th, with the average city being **1°C warmer** compared to last century. Subarctic regions, Andalusia and southern Romania are most affected.

See how much temperatures increased in your city!

Select your city (558)

Filter list by country

Average temperature in and around Stockholm



## OneDegreeWarmer

onedegreewarmer.eu

Analysis of temperatures in 558 European cities from 1900 to 2017

Using 2 datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF)

Analysis by ECMWF scientists to make up for weather stations change of location, cities expansion...

Highlight temperature increases and consequences (health, roads/rails traffic, students focus...)

# OneDegreeWarmer – the story spread

## PUBLISHED ARTICLES

### English

- EDJNet: Europe is getting warmer, and it's not looking like it's going to cool down anytime soon
- EU Observer: Every major city in Europe is getting warmer
- Kosovo 2.0: Western Balkan's heating up
- OBC Transeuropa: South-East Europe is burning: how temperatures have risen for the last 117 years
- Vox Europ: Europe is getting warmer, and that is not going to change

### Albanian

- Kosovo 2.0: Po nxehet Ballkani Perëndimor

### Bosnian/Serbian/Croatian

- EurActive: U Jugoistočnoj Evropi sve toplije
- H-Alter: Toplo, toplije
- OBC Transeuropa: U Jugoistočnoj Evropi sve toplije

### Bulgarian

- Capital Bulgaria: 24 сеп 2018, 9:00, 794 прочитания Средната температура в 11 български града се е покачила с над 1

### French

- EDJNet: L'Europe se réchauffe, et c'est parti pour durer
- Rue 89 Bordeaux: La température moyenne a grimpé de près d'un degré depuis 1900 à Bordeaux
- Rue 89 Stasbourg: Voici comment Strasbourg s'est réchauffée d'un degré depuis 2000
- Rue 89 Lyon: Lyon, plus forte augmentation de Rhône-Alpes

### German

- EDJNet: In Europa wird es immer wärmer, und es sieht nicht so aus, als würde es sich bald abkühlen
- Spiegel Online: Europa: ein Grad wärmer
- Vox Europ: Europa wird unaufhaltsam wärmer

### Italian

- AskaNews: Surriscaldamento record porta a Belluno vino e olio d'alta quota
- EDJNet: L'Europa si riscalda, sempre di più
- Corriere delle Alpi: Cambiamenti climatici, a Belluno il record negativo in Italia
- Gazzetta di Mantova: Inchiesta sul clima, così il surriscaldamento globale mette a rischio il nostro futuro
- Il Piccolo: Clima più caldo in Fvg, a Trieste e Gorizia aumento di un grado in 20 anni
- Il Tirreno: Emergenza clima, le città che si stanno riscaldando di più: Piombino seconda in Italia
- La Nuova Ferrara: Clima sempre più bollente, Ferrara nella "top ten" in Italia
- La Provincia Pavese: Cambiamenti climatici. Pavia terza in Italia per surriscaldamento
- Messaggero Veneto: Clima, estati roventi ed inverni senza neve: così si sta surriscaldando il Friuli Venezia Giulia
- OBC Transeuropa: Sud-est Europa: sempre più caldo

### Polish

- EDJNet: Europa jest coraz cieplejsza i nie zapowiada się na ochłodzenie

### Portuguese

- EDJNet: Europa está a ficar mais quente e parece que não vai refrescar assim tão cedo
- Público: As cidades europeias estão mais quentes. Portugal está no fim da lista
- Zap Europa cada vez mais quente. Portugal é o país que menos aqueceu

### Slovenian

- Pod Črto: Analiza temperatur na stari celini: v evropskih mestih je vse bolj vroče
- RTV Slovenija: Analiza temperatur na stari celini: v evropskih mestih je vse bolj vroče

### Spanish

- EDJNet: Europa se está calentando, y no va a enfriarse pronto
- El Confidencial: ¿Cuántos días de calor sufren tus hijos en el colegio comparado con los que tú pasaste?

## OneDegreeWarmer – challenges

### Europe One Degree Warmer: How we got things wrong—and are working on fixing them

The European Data Journalism Network published [an analysis](#) of temperature trends in 558 cities and their surroundings in Europe on Monday, 24 September 2018, starting at 08:00 CEST, in partnership with several news outlets. [A few hours after publication, several commenters pointed out that the temperature data for some cities in Sweden, which had been described in the analysis as the fastest-warming in Europe, were probably erroneous.](#) Swedish blogger Göran Johnson, for instance, [used](#) the Swedish weather service's data to point out the problems in our data for the city of Kiruna.

[medium.com/european-data-journalism-network/europe-one-degree-warmer-how-we-got-things-wrong-and-are-working-on-fixing-them-67c9c892c13c](https://medium.com/european-data-journalism-network/europe-one-degree-warmer-how-we-got-things-wrong-and-are-working-on-fixing-them-67c9c892c13c)

The mistake was identified on Tuesday, a day after publication. [Data for 38 locations had been collected on a different machine from the rest and had not been properly checked.](#) The methodology for all cities was to use the [ERA-20C](#) database for the period 1900–1978, the [ERA-interim](#) database for the period 1979–2017 and to correct the first time series using the second one (the code used for the reconciliation is available in [a Jupyter Notebook](#)). [For the 38 locations with erroneous data, the reference data for the period 1979–1998 was ERA-20C's \(instead of ERA-interim's\) and no reconciliation had taken place.](#) The result was that some locations were shown to be warming much more rapidly than they actually were, sometimes by more than two degrees Celsius. (The data for the 520 other locations was correct).

# OneDegreeWarmer – challenges

We would like to apologise to all the media partners of EDJNet and their readers for the mistake in the data. The error was preventable, had we put in place more checks to ensure the integrity of the data.

The error was compounded by the fact that most of the erroneous data showed faster warming in sub-arctic regions, which was an outcome experts mentioned we could expect. A strong confirmation bias was at play. While we did check the results at the European level with climate experts, a talk with local (Swedish) experts would have allowed us to discover the error before publication.

## Technical challenges

→ automated checks (monitoring / alerting)

## Human challenges

→ domain knowledge checks / feedbacks

Methology and data at <https://gitlab.com/edjn/onedegwarmer-shared-data>

## 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

- <https://github.com/Eleonore9/-pyconse18>
- <http://datajournalismhandbook.org/1.0/en/>



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



jpluplus.org